Biologically plausible solutions for spiking networks with efficient coding

Anonymous Author(s)
Affiliation
Address
email

Abstract

Understanding how dynamics of neural networks is shaped by the computations the networks perform is a fundamental topic in neuroscience. Recently, the framework of efficient coding proposed a theory on how spiking neural networks can compute low-dimensional stimulus signals with high efficiency. Efficient spiking networks are based on time-dependent minimization of a cost function related to information coding with spikes. To inform the understanding of the function and dynamics of biological networks in the brain, however, the mathematical models have to obey constraints of biological networks. Currently, spiking network models of efficient coding have been extended to include some features of biological plausibility, such as the inclusion of architectures with excitatory and inhibitory neurons. However, biological realism of efficient coding theories is still limited to simple cases and does not include single neuron and network properties that are known to be important in biological circuits. Here, we revisit the theory of efficient coding with spikes to develop spiking neural networks that are closer to biological circuits. Namely, we find a biologically plausible spiking model realizing efficient coding in the case of a generalized leaky integrate-and-fire network with excitatory and inhibitory units, equipped with fast and slow synaptic currents, local homeostatic currents such as spike-triggered adaptation, hyperpolarization-activated rebound current, heterogeneous firing thresholds and resets, heterogeneous postsynaptic potentials, and structured, low-rank connectivity. We show how the complexity of E-E connectivity matrix shapes network responses.

1 Introduction

Complex biological systems are thought to be shaped through evolution to optimally perform a set of desired functions within the constraints of biological building blocks they are made of. A bulk of previous influential work [30, 21] has proposed that the organization of the mammalian cortex, the most sophisticated biological organ known to humanity, makes no exception and follows too the general principles of evolution towards optimality under biological constraints.

One prominent theory is that the function of the cerebral cortex is organized for efficient coding, in that it optimizes the encoding of information under a number of biological constraints such as the metabolic expenditure, the speed and the timing of information coding. Imposed constraints can be formalized with a cost function.

To understand biological neural networks that operate in animal's brains, the mathematical description of networks optimizing functionally relevant cost functions has to include as many features of biological realism as possible, so that the constraints satisfied by the model network will be informative of those obeyed by biological neural circuits. Thus, all elements of the model should have a counterpart

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

in biological circuits [13] and be directly relatable to them. It is still unclear how to derive efficient coding theories of biologically plausible neuronal networks.

An influential proposal of efficient coding in cortical circuits has been on extracting statistical regularities from static natural images with a sparse set of basis functions, which yielded basis functions that resemble receptive fields in V1 [30]. A related proposal in the domain of auditory stimuli yielded tuning properties of auditory nerve fibers [21]. In these studies, the goal of efficient coding is to find sparse spatial or temporal filters that enables optimal information transmission about the stimulus by the activity of each single neuron. Analyses of cortical networks, however, show that information processing tends to be distributed across neurons [15] and neglecting neuron-to-neuron interactions typically reduces the information encoded in the spiking activity of neural circuits [37] or the ability of use this information to produce relevant behavioral outputs [38, 43]. Moreover, analyses of local neural circuits show that millisecond precision of spike timing of sensory neurons carries information about both static [34, 31] and dynamic stimuli [16, 28], which suggests that the timing of spiking activity is highly relevant for the information processing in the cortex. Analyses of neural data suggest that meaningful theories of coding in the real cerebral cortex must rely on spiking neuron models and on population codes.

To address this issue, models of how to optimally perform a given coding function have been extended to spiking dynamics. The cost function includes a quadratic error between the desired signal and a linear read-out of population activity and the distance between the signal and the read-out is minimized with every spike [1, 4, 2]. However, the efficient coding with spikes, and the related theory of predictive coding [46], have been so far developed as abstract mathematical frameworks that lack biological realism. Namely, these models are not consistent with the observation that a single neuron can either be excitatory or inhibitory, but not both (Dale's law). Very recently, progress has been made in this direction by deriving efficient coding with spiking networks with a more biologically plausible neural circuit architecture with interacting excitatory (E) and inhibitory (I) neurons [22], opening the possibility to model neural computations with biologically relevant models.

Here, we make major further progress by deriving a mathematically well defined extension of efficient coding for considerably more biologically plausible spiking neural network models of cortical circuits. We develop a recurrently connected neural circuit with E and I cell type that performs a generic linear transformation between its inputs and its outputs. We impose the E-I architecture and recover a network with low-rank and structured connectivity that maximizes the information transmission about the stimulus with every spike. Our proposed solution recovers a generalized leaky integrate-and-fire (LIF) neuron model under the constraints of Dale's law. Our solution also takes into account faster membrane time constant of I compared to E neurons, compatible with the biophysics of pyramidal neurons and fast-spiking inhibitory interneurons. Our derivations prescribe subthreshold dynamics, firing thresholds and resets after spiking, thus giving a complete description of a spiking neural network along with network's computation. Importantly, all elements of the model are biologically plausible and directly relatable to measurable currents in biological networks.

We start by introducing the cost functions and explaining how the spike timing is conditioned on a minimization of cost functions of E and I neurons in every time step. We continue with an outline of the analytical treatment of cost functions that leads to expressions of membrane potentials and firing thresholds of a generalized LIF network model with structured, low-rank connectivity. Further on, we provide a description and biologically plausible interpretation of derived currents. Finally, we simulate the network with biologically plausible parameters and analyze the behavior of the network with respect to the complexity of E-E synaptic connections.

2 Results

2.1 Minimization of the loss function with spikes

From a spiking network of N^E excitatory and N^I inhibitory neurons, we define a linear population-read-out of the spiking activity of E and I neurons:

$$\dot{\hat{\boldsymbol{x}}}_Y(t) = -\lambda_Y \hat{\boldsymbol{x}}_Y(t) + W_Y \boldsymbol{f}_Y(t), \qquad Y \in \{E, I\}, \tag{1}$$

with $\lambda_Y > 0$ the inverse time constant of the read-out. In the population read-out, the decoding matrix W_Y weights the vector of spikes $\boldsymbol{f}_Y(t) = [f_1^Y(t), \ldots, f_{N^Y}^Y(t)]^\intercal$. The M-by-N^Y matrix W_Y associates each of the N^E excitatory (N^I inhibitory) neurons with M weights, $W_Y = (w_{mi}^Y); m =$

$1 \ldots, M,\ i = 1, \ldots, N^Y$, where N^Y is the number of neurons of the cell type Y, and M is the number of stimulus features estimated by the network. The population read-out in eq. (1) is an estimate of the desired signal $\boldsymbol{x}(t)$, that is an abstract signal, unknown to the network.

The desired signal $\boldsymbol{x}(t)$ is a linear function of stimulus features $\boldsymbol{s}(t) = [s_1(t), \ldots, s_m(t), \ldots, s_M(t)]^\intercal$ and is defined as follows:

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + \boldsymbol{s}(t), \tag{2}$$

where $A = (a_{mn});\ m, n = 1, ..., M$ is a square matrix determining the linear transformation between the input features $\boldsymbol{s}(t)$ and the desired signal $\boldsymbol{x}(t)$. We will refer to the transformation between the features and the desired signal as network's computation. If A is a diagonal matrix, network's computation operates independently across features, while a non-diagonal matrix A implements linear mixing of features.

We pose that the objective of the excitatory network is to minimize the distance between the desired signal and the population read-out of E neurons, while the objective of the inhibitory population is to minimizes the distance between the read-out of the E and I populations [2, 6]. Besides the coding error, we also take into account the metabolic cost on spiking, since spiking activity is energetically expensive for animal agents [29]. These objectives are formalized by loss functions with a quadratic coding error and a quadratic regularizer, relative to E and I cell type:

$$\mathrm{L}_E(t) = \sum_{m=1}^{M} (x_m(t) - \hat{x}_m^E(t))^2 + \mu_E \sum_{i=1}^{N_E} (r_i^E(t))^2 \tag{3a}$$

$$\mathrm{L}_I(t) = \sum_{m=1}^{M} (\hat{x}_m^E(t) - \hat{x}_m^I(t))^2 + \mu_I \sum_{i=1}^{N_I} (r_i^I(t))^2, \tag{3b}$$

where constants $\mu_Y > 0$ are weighting the relative contribution of the metabolic cost over the coding error. An alternative to the cost function in eq. (3b) would be to minimize the distance between the read-out of I population and the signal, however, such a formulation, without further approximations, leads to a network where E and I neurons are unconnected. The variable $r_i^Y(t)$ is defined as a low-pass filtered spike train of the neuron i of cell type $Y \in \{E, I\}$,

$$\dot{r}_i^Y = -\alpha_i^Y r_i^Y(t) + f_i^Y(t), \tag{4}$$

with $\alpha_i^Y > 0$ the inverse time constant of the low-pass filter. As we gather the variable $r_i^Y(t)$ across neurons, we get a vector of time-dependent neural activities $\boldsymbol{r}_Y(t) = [r_1^Y(t), \ldots, r_{N^Y}^Y]^\intercal$.

We pose that there will be a spike of neuron i of cell type Y at time $t = t^+$ if and only if this decreases the cost function, i.e.

$$\mathrm{L}_Y\left(t^+ | (f_i^Y(t^+) = 1)\right) + \eta_i^Y(t^+) < \mathrm{L}_Y\left(t^- | (f_i^Y(t^-) = 0)\right), \tag{5}$$

where the noise term $\eta_i^Y(t)$ is added to account for randomness of the spike generation in biological neurons. The noise term $\eta_i^Y(t) = \sigma_i^Y \xi_i^Y(t)$ is defined as a Gaussian random variable $\xi_i^Y(t)$ with zero mean and covariance $\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t - t')$, and σ_i^Y is the noise intensity. Contrary to previous approaches where the noise is added in the membrane potential [2, 19], we here assume noise in the condition for spiking. While the noise in the membrane potential models synaptic inputs that are unrelated to the coding function of the network, the noise in spike generation captures a source of noise that is intrinsic to neurons, namely the noise in the fraction of open ionic channels for a given membrane potential [9]. Opening and closing of sodium and potassium channels is responsible for the generation of spikes, but the fraction of open channels for a given membrane potential is probabilistic.

Assuming that a spike is only fired if this decreases the cost function as in eq. (5), we apply an analytical treatment of eqs. 3a-3b in parts similar to [2], and obtain the following expressions:

$$W_E^\intercal\left(\boldsymbol{x}(t) - \hat{\boldsymbol{x}}_E(t)\right) - \mu_E \boldsymbol{r}_E(t) < \boldsymbol{\theta}_E + \boldsymbol{\eta}_E(t) \tag{6a}$$

$$W_I^\intercal\left(\hat{\boldsymbol{x}}_E(t) - \hat{\boldsymbol{x}}_I(t)\right) - \mu_I \boldsymbol{r}_I(t) < \boldsymbol{\theta}_I + \boldsymbol{\eta}_I(t), \tag{6b}$$

where vectors of firing thresholds, $\boldsymbol{\theta}_Y = [\theta_1^Y, \ldots, \theta_{N^Y}^Y]^\intercal$ for $Y \in \{E, I\}$, have elements

$$\theta_i^Y = \frac{1}{2}\left(\|\boldsymbol{w}_i^Y\|_2^2 + \mu_Y\right), \tag{7}$$

with $\|\boldsymbol{w}_i^Y\|_2^2$ the squared length of decoding vector of the local neuron, $\|\boldsymbol{w}_i^Y\|_2^2 = \sum_{m=1}^M (w_{mi}^Y)^2$. Meanwhile, $\boldsymbol{\eta}_Y(t)$ is the vector of noise processes, $\boldsymbol{\eta}_Y(t) = [\sigma_1^Y \xi_1^Y(t), \ldots, \sigma_{N_Y}^Y \xi_{N_Y}^Y(t)]$. While the right-hand side of inequalities in (6a) and (6b) are firing thresholds, we assume that the left-hand sides are proportional to the membrane potentials of E and I neurons, $\boldsymbol{u}_E(t)$ and $\boldsymbol{u}_I(t)$, respectively, with $\boldsymbol{u}_Y(t) = [u_1^Y(1), \ldots, u_{N^Y}^Y(t)]^\intercal$.

To express the temporal dynamics of the membrane potentials, $\dot{\boldsymbol{u}}_E(t)$ and $\dot{\boldsymbol{u}}_I(t)$, we take the time-derivative of all time-dependent terms on the left-hand side of eqs. (6a) and 6b. Contrary to previous approaches and without loss of generality, we express the mixing matrix A as a difference of a square matrix B and a diagonal matrix $\lambda_E \mathbf{I}^{MxM}$,

$$A = B - \lambda_E \mathbf{I}^{MxM}, \tag{8}$$

with $\mathbf{I}$ the identity matrix. Seen that cost functions minimize the distance between the signal and the E estimate (eq. 3a), and between the E and I estimates (eq. 3b), we use the approximations $\boldsymbol{x}(t) \approx \hat{\boldsymbol{x}}_E(t)$ and $\hat{\boldsymbol{x}}_E(t) \approx \hat{\boldsymbol{x}}_I(t)$. Moreover, we remove synaptic connections that violate Dale's law (see the Supplementary material). After re-arranging the terms and without further approximations, interestingly, the network model can be expressed as a generalized LIF model, i.e., the LIF model with spike-triggered adaptation [23].

2.2 Generalized integrate-and-fire network with low-rank structured connectivity

Due to aforementioned approximations, we write the membrane potential as $u_i^Y(t) \approx C_m V_i^Y(t)$, with C_m the capacitance of the cell membrane. The subthreshold dynamics of E and I neurons is then dependent on the following currents:

$$C_m \dot{V}_i^E(t) = I_i^{\text{leak } E}(t) + I_i^{\text{ff}}(t) + I_i^{EI}(t) + I_i^{\text{slow } EE}(t) + I_i^{\text{local } E}(t) + I_i^{\text{rebound}}(t), \tag{9a}$$

$$C_m \dot{V}_i^I(t) = I_i^{\text{leak } I}(t) + I_i^{IE}(t) + I_i^{II}(t) + I_i^{\text{slow } EI}(t) + I_i^{\text{local } I}(t), \tag{9b}$$

and complemented with fire-and-reset rule: if $V_i^Y(t^-) \geq \vartheta_i^Y(t^-) \rightarrow V_i^Y(t^+) = V_i^{\text{reset } Y}$. For simplicity, we set the resting potential at 0, but a biologically plausible resting potential can be introduced without affecting network's dynamics. From eqs.(6a), (6b) and (7), the effective threshold that takes into account the noise in the condition for spiking (eq. 5) is the following:

$$\vartheta_I^Y(t) = \frac{1}{2}(\mu_Y + \|\boldsymbol{w}_i^Y\|_2^2) + \sigma_i^Y \xi_i^Y(t), \qquad Y \in \{E, I\} \tag{10}$$

while the reset potentials for E and I cell types are

$$V_i^{\text{reset } E} = -\frac{1}{2}(\mu_E - \|\boldsymbol{w}_i^E\|_2^2), \tag{11a}$$

$$V_i^{\text{reset } I} = -\frac{1}{2}(\mu_I + \|\boldsymbol{w}_i^I\|_2^2) \tag{11b}$$

The firing thresholds in eq. 10 and reset potentials in eqs. 11a-11b are thus proportional to the length of decoding vector of the local neuron $\|\boldsymbol{w}_i^Y\|_2$, and to the regularizer μ_Y that affects equally all the neurons from the same cell type. If decoding weights w_{mi}^Y are heterogeneous across neurons, firing thresholds and resets are heterogeneous. The noise at threshold in eq. (10) implements the escape noise that has been shown to be relevant for modeling of randomness in spike generation in cortical neurons [11].

The membrane equations in eqs. 9a-9b depend on a number of currents that, interestingly, all have a straightforward counterpart in biological networks. The first currents on the right-hand site of eqs. 9a-9b are leak currents,

$$I_i^{\text{leak } Y}(t) = -\lambda_Y V_i^Y(t), \qquad Y \in \{E, I\} \tag{12}$$

that result from absorbing of several terms, among others the diagonal matrix $\lambda_E \mathbf{I}^{MxM}$ from eq. 8. The feedforward current to E neurons is proportional to the weighted sum of external inputs $s_k(t)$,

$$I_i^{\text{ff}}(t) = \sum_{m=1}^M w_{mi}^E s_m(t). \tag{13}$$

Fast synaptic currents between E-to-I, I-to-I and I-to-E cells are given by a weighted sum of spikes of presynaptic neurons,

$$I_i^{IE}(t) = \sum_{j=1}^{N_E} C_{ij}^{IE} f_j^E(t), \qquad I_i^{II}(t) = -\sum_{\substack{j=1 \\ j\neq i}}^{N_I} C_{ij}^{II} f_j^I(t) \qquad I_i^{EI}(t) = -\sum_{j=1}^{N_I} C_{ij}^{EI} f_j^I(t), \quad (14)$$

with structured connectivity matrices C_{ij}^{YZ}, where the strength of synaptic connection is proportional to the similarity of decoding vectors,

$$C_{ij}^{YZ} = \begin{cases} (\boldsymbol{w}_i^Y)^\intercal \boldsymbol{w}_j^Z, & \text{if } (\boldsymbol{w}_i^Y)^\intercal \boldsymbol{w}_j^Z > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

Note that synaptic connections between neurons with different selectivity (i.e., neuronal pairs with negative dot product of decoding vectors, $(\boldsymbol{w}_i^Y)^\intercal \boldsymbol{w}_j^Z < 0$), have been set to zero to make the network consistent with Dale's law (see the Supplementary material). In particular, we had to ensure that a particular neuron can only send excitatory or inhibitory currents to other neurons, but not both. Even though there is no direct fast synaptic connections between E-to-E neurons, effectively, fast synaptic connectivity implements lateral inhibition between E neurons with similar selectivity. Lateral inhibition (or competition) between E neurons with similar selectivity is a dynamical effect that is essential for an efficient neural code [17] and has been demonstrated in biological circuits [7].

Next, we have currents that are triggered by spiking of the local neuron,

$$\dot{I}_i^{\text{local } Y}(t) = -\alpha_i^Y I_i^{\text{local } Y}(t) - \mu_Y \left(\lambda_Y - \alpha_i^Y\right) f_i^Y(t), \qquad Y \in \{E, I\}. \quad (16)$$

The kinetics of the local current is determined by the time constant of low-pass filtered spike train, α_i^Y, while the sign and the strength of the current depend on the relation of time constants of the population read-out λ_Y, and the single neuron read-out α_i^Y. To allow the population read-out $\hat{\boldsymbol{x}}_Y(t)$ to track fast changes of the signal $\boldsymbol{x}(t)$, we assume that the population read-out is fast, while the single neuron read-out, related to a homeostatic read-out of single neuron's firing frequency [8], is assumed to be a slower process. Such a relation of time constants, $\lambda_Y > \alpha_i^Y$, constrains the local current in eq. 16 to spike-triggered adaptation. If, on the contrary, $\lambda_Y < \alpha_i^Y$, we get spike-triggered facilitation, a current that might be relevant during learning and restructuring of synaptic connections [33]. Besides the difference of time constants between the population and single neuron read-out, the strength of the local current depends on the constant of the regularizer μ_Y. The stronger the penalty of the metabolic cost in the objective function in eq. (3a) -(3b), the stronger the current in eq. (16). This makes most sense in case when the local current is spike-triggered adaptation, where a stronger regularizer increases the amplitude of the hyperpolarizing jump of the adaptation current and thus reduces the firing frequency of the spiking neuron.

Besides fast synaptic currents (eq. 14), there are slower synaptic currents between E-to-E and E-to-I neurons:

$$\dot{I}_i^{\text{slow } EE}(t) = -\alpha_i^E I_i^{\text{slow } EE}(t) + \sum_{\substack{j=1 \\ j\neq i}}^{N_E} D_{ij}^{EE} f_j^E(t), \quad (17a)$$

$$\dot{I}_i^{\text{slow } IE}(t) = -\alpha_i^E I_i^{\text{slow } IE}(t) + (\lambda_I - \lambda_E) \sum_{j=1}^{N_E} D_{ij}^{IE} f_j^E(t), \quad \lambda_I > \lambda_E, \quad (17b)$$

with matrices of synaptic interactions:

$$D_{ij}^{YE} = \begin{cases} (\boldsymbol{w}_i^Y)^\intercal B \boldsymbol{w}_j^E, & \text{if } (\boldsymbol{w}_i^Y)^\intercal \boldsymbol{w}_j^Z > 0, \qquad B \text{ positive semi-def.} \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

where $Y = E$ for E-to-E synapses and $Y = I$ for E-to-I synapses. The synaptic time constant α_i^Y in eqs. 17a-17b is given by the time constant of low-pass filtered spike trains (eq. 4). Same as with fast synaptic currents in eq. (14), the strength of the synapse between the presynaptic neuron j and the postsynaptic neuron i in eq. (18) depends on the similarity of decoding vectors of the presynaptic and postsynaptic neurons, however, here the synaptic current depends also on the matrix B that was expressed from the mixing matrix A (see eq. 8). To make slower synaptic currents consistent with

Dale's law, we removed connections between neurons with different selectivity and constrained the matrix B to be positive semi-definite. This ensures that presynaptic spikes of E neurons always cause an excitatory current in the postsynaptic neurons. The postsynaptic current in the neuron i caused by a spike of the presynaptic neuron j is: $\text{PSC}_{ij}^{\text{slow } EY} = D_{ij}^{EY} \exp\left(-\alpha_i^E (t - t'_{j,E})\right)$ with $t'_{j,E}$ the spike time of the excitatory presynaptic neuron j. Note that with heterogeneous parameters α_i^Y, the duration of the PSC is heterogeneous across neurons. Contrary to fast connections that implement lateral inhibition among E neurons with similar selectivity (see eq. 14), slower E-to-E synaptic connections implement cooperation among E units with similar selectivity. E-to-E connections are important for the description of biological networks in the cortex [18].

Slower E-to-E synapses are complemented by slower E-to-I synapses (eq. 17b), where the excitatory effect of an E-to-I synapse constrains the following relation of time constants: $\lambda_I > \lambda_E$. Since λ_E and λ_I are membrane time constants of the E and I cell type, respectively (see eq. 12), this relation constrains the membrane time constant in E neurons to be slower than in I neurons, consistent with E cell type modeling pyramidal cells while the I cell type models fast-spiking interneurons [26]. Fast-spiking (somatostatin) interneurons are a prevalent class of inhibitory neurons in cortical circuits [42]. As pyramidal neurons account for about 80 $\%$ of cortical neurons, and fast-spiking interneurons account for at least a half of inhibitory neurons in the cortex [44], our model altogether provides a description for about 90 $\%$ of cortical neurons. Note that alternative solutions for slow currents exist (see the Supplementary Material). These alternative solutions are, however, of lesser biological relevance since they either result in a network without E-to-E connections, or induce global imbalance of E and I currents in the network.

The efficient E-I network has low-rank connectivity. The rank of fast synaptic connectivity is, by definition of the connectivity matrices C (eq. 15) and D (eq. 18), equivalent or smaller than the number of features encoded by the network, M. It is natural to assume that the number of active features $s_m(t)$ is much smaller than the number of neurons in the network, $M << N^E$, which gives low-rank connectivity matrices. Low-rank connectivity forces the neural activity to live on a low-dimensional manifold [25], and the relevance of the latter for the description of the dynamics of biological neural ensembles has been strongly suggested by the data [10]. Note also that using heterogeneous and sparse decoding weights w_i^Y gives heterogeneous and sparse synaptic connectivity in case of both fast and slower synapses.

Finally, we derived a depolarizing current triggered by the spike of the local excitatory neuron, that we termed $I_i^{\text{rebound}}(t)$. In the efficient spiking network, the diagonal of the E-to-E recurrent connectivity matrix D^{EE} (eq. 17a) implements a positive "self-connection" - as the neuron spikes, it generates a local depolarizing current with slower kinetics of the low-pass filtered spike train. We write this effect as follows:

$$\dot{I}_i^{\text{rebound}}(t) = -\alpha_i^E I_i^{\text{rebound}}(t) + D_{ii}^{EE} f_i^E(t). \tag{19}$$

Since the matrix B is constrained to be positive semi-definite (see eq. 18), the current in eq. 19 is always depolarizing. After a spike, the local neuron is reset to its reset potential, which, in E neurons, is necessarily below the resting potential (see eq. 11a; the resting potential is here at 0 mV). Right after the spike, the local neuron is therefore strongly hyperpolarized, and precisely at this moment, the rebound current activates. The current $I_i^{\text{rebound}}(t)$ thus creates a rebound of the membrane potential after a strong hyperpolarization. In biology, a current with these properties is the hyperpolarization-activated cation current (h-current [32]), a current with slower kinetics that is important in the generation of network oscillations, spontaneous firing and in controlling the excitability of cortical neurons [24, 36]. A schematic of the network is illustrated on Figure 1A.

2.3 Effect of complexity of E-E connectivity on network responses

In the following, we analyze the effect of the complexity of E-E connections on network responses with analytical considerations and simulations. We start by considering a simplified network model where we assume the matrix A in eq. (8) to be a diagonal matrix. A simple way to do so is to set the elements of the matrix B to 0, i.e., $B = (b_{mn})$; $b_{mn} = 0 \; \forall m, n = 1, \ldots, M$, which constrains the desired signal to a leaky integration of input features that is independent across features,

$$\dot{\boldsymbol{x}}(t) = -\lambda_E \boldsymbol{x}(t) + \boldsymbol{s}(t), \tag{20}$$

The time constant of the desired signal $\boldsymbol{x}(t)$ is now equal to the time constant of the population read-out of the E cell type eq. (1), and such a network is devoid of E-E (eq. 17a) and slower E-I

(eq. 17b) synaptic currents, as well as of the rebound current (eq. 19). We further simplify the network by considering the special case where the time constants of the population and the single neuron read-out are equal, $(\alpha_i^Y)^{-1} = (\lambda_Y)^{-1} \forall i; Y \in \{E, I\}$, which sets the local currents in eq.(16) to zero.

To simulate such a simplified network, as well as more complex ones, we first have to determine decoding weights w_{mi}^Y, since these are free parameters. We use unstructured decoding weights that we draw from the normal distribution with zero mean and standard deviation σ_w^Y, $w_{mi}^Y \sim \mathcal{N}(0, \sigma_w^Y)$, with $\sigma_w^E = 1$ and the ratio $\sigma_w^I : \sigma_w^E = 3 : 1$. We chose decoding weights in the I cell type to be stronger than in the E cell type, but the number of I neurons is smaller then the number of E neurons, with a biologically plausible ratio $N^E : N^I = 4 : 1$. Decoding weights then determine fast synaptic currents $I_i^{EI}(t)$, $I_i^{II}(t)$ and $I_i^{IE}(t)$ through eqs. 14-15, as well as slower synaptic currents $I_i^{\text{slow } EE}(t)$ and $I_i^{\text{slow } IE}(t)$ through eqs. 17a-18.

As we stimulate the simplified network with a step input in one of the input features, the network responds with asynchronous spiking of neurons that are aligned to the active feature (Figure 1B). The feedforward current to neuron i depends on the alignment of the input $\boldsymbol{s}(t)$ with neuron's decoding vector $\boldsymbol{w}_i^E$ (see eq. 13). When a single feature is activated, e.g. $s_n(t) \neq 0$, a positive input, $s_n(t) > 0$, drives neurons with positive decoding weight for the $n-$th feature (i.e., neurons with $w_{ni}^E > 0$), while a negative input, $s_n(t) < 0$, drives neurons with negative decoding weights (neurons with $w_{ni}^E < 0$). If several features are active simultaneously, the feedforward current to E neurons is given by a linear mixture of features. The simplified network is purely feedforward-driven and the network response never outlasts the stimulus (Figure 1B). The recurrent connectivity of such a network consists of structured fast connections between neurons with similar decoding selectivity (eq. 15), and the stronger the similarity of decoding vectors of neurons i and j, the stronger their synapse.

We now consider the network with E-E and slower E-I connections by choosing the matrix A to be non-diagonal. The response of the network now critically depends on the strength and the structure of recurrent excitatory (E-E) connections. The synaptic weight between excitatory neurons i and j depends not only on the similarity of decoding vectors $\boldsymbol{w}_i^E$ and $\boldsymbol{w}_j^E$ (as with fast connections), but also on the square matrix B (eq. 18). Following the eq. 18, the matrix B has to be positive semi-definite to ensure that the network obeys Dale's law. Positive semi-definiteness in turn constrains the matrix B to be symmetric and to have non-negative eigenvalues. To ensure these properties, we write the matrix B as follows:

$$B = a\Gamma\Gamma^\intercal, \qquad \Gamma = [\boldsymbol{b_1}, \dots, \boldsymbol{b_{M'}}] \tag{21}$$

with $a > 0$ a positive constant influencing the strength of E-E connections and $\boldsymbol{b_1}, \dots, \boldsymbol{b_{M'}}$, a set of linearly independent column vectors with entries $\boldsymbol{b_n} = [b_{n1}, \dots, b_{nM}]^\intercal, n = 1, \dots, M'$, and where the number of vectors is constrained as follows: $1 \leq M' \leq M$. The matrix Γ determines the rank of the matrix B, as M' linearly independent column vectors set the rank of B to $\text{rank}(B) = M'$. The rank of the matrix B in turn determines the complexity of connectivity matrices D^{EE} and D^{IE} through eq. (18).

In the simplest case (beyond the trivial case where all elements of B are set to zero), Γ is given by a single column vector, $\Gamma = \boldsymbol{b_1}$, which sets the matrix B to the following: $B = a\boldsymbol{b_1}\boldsymbol{b_1}^\intercal$, and the rank of the matrix B to $\text{rank}(B) = 1$. We simulate such network, setting the membrane time constants of E and I cell type to biologically plausible values, with $\lambda_E^{-1} = 10$ ms for pyramidal neurons, and $\lambda_I^{-1} = 5$ ms for inhibitory interneurons, as found in biological data [42]. In addition, we assumed the time constants of the population read-out to be shorter (specifically, twice as short) than the time constant of the single neuron read-out, as single-neuron read-out is a slower process of homeostatic regulation of the firing rate of single neurons (see eq. 4). According to eq. (16), this gives spike-triggered adaptation in E and I neurons. We found a population-wide oscillation and a network response that outlast the external stimulation (Figure 1C), in contrast to the response of the simpler network ((Figure 1B)). Similarly to the simpler network, we still have that only neurons aligned with the input feature activate (e.g., neurons with $w_{1i}^Y > 0$ with $s_1(t) > 0$).

By increasing the rank of the matrix B, also neurons that are not aligned with the input feature participate in the network response. We now define the matrix B with $M' > 1$ linearly independent vectors, where $a_1\boldsymbol{b_1} + \dots + a_{M'}\boldsymbol{b_{M'}} = \boldsymbol{0}$ if and only if $a_n = 0 \, \forall n$. Linear independence of column vectors $\boldsymbol{b_n}$ implies that the matrix B has rank $M' > 1$. We find that with $M' > 1$, both neurons with

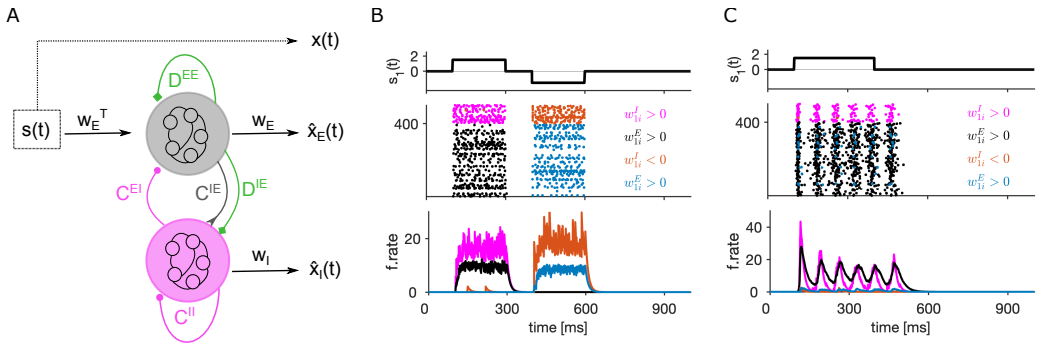

Figure 1: (A) Schema of the E-I network. E neurons (gray) receive a linear sum of external inputs $\boldsymbol{s}(t)$. E and I (magenta) neurons with similar selectivity are connected with fast synapses C^{YZ}, C^{II} and slower synapses D^{EE} and D^{IE} (green). Linear population read-out of the spiking activity of E and I cell types define estimates $\hat{\boldsymbol{x}}_E(t)$ and $\hat{\boldsymbol{x}}_I(t)$, respectively. (B) Activity of the simplified network without slower synapses and without the adaptation current. Top: The stimulus in the first feature $s_1(t)$. The remaining features are inactive. Middle: The spike trains of E and I neurons aligned with the positive input in the first feature (black and magenta, respectively), and aligned with the negative input in the first feature (blue and red, respectively). Bottom: Neuron-average firing rate, with the same color code as in the middle plot. Parameters for B: $M = 3$, $N^E = 400$, $N^E : N^I = 4 : 1$, $\lambda_E^{-1} = \lambda_I^{-1} = (\alpha_i^E)^{-1} = (\alpha_i^I)^{-1} = 10$ ms $\forall i$, $b_{mn} = 0\ \forall m, n = 1, \ldots, M$, $\mu_E = \mu_I = 6$, $\sigma_i^E = \sigma_i^I = 0.167\ \forall i$, $\sigma_w^E = 1$, $\sigma_w^I : \sigma_w^E = 3 : 1$, $dt = 0.02\ ms$. (C) Same as in (B) but showing the network with slower synapses and with adaptation current. Parameters for C: $\lambda_E^{-1} = 10$ ms, $\lambda_I^{-1} = 5$ ms, $(\alpha_i^E)^{-1} = 20$ ms $\forall i$, $(\alpha_i^I)^{-1} = 10$ ms $\forall i$, $B = a\boldsymbol{b_1}^\intercal \boldsymbol{b_1}$ with $a = 0.035$ and $\boldsymbol{b_1} = [0.5, 0, 0.3]^\intercal$, $\mu_E = \mu_I = 15$, $\sigma_i^E = \sigma_i^I = 0.25\ \forall i$. Other parameters as in B.

decoding vector aligned as well as misaligned with the input feature respond (Figure 2A-B; with $s_1(t) > 0$, neurons with $w_{1i}^Y > 0$ as well as with $w_{1i}^Y > 0$ are active). Duration of network's response depends on the strength of E-E synaptic connectivity, here controlled by the parameter a in eq. 21. With sufficiently strong parameter a, the network maintains sustained activity (Figure 2B). In the absence of external drive, but with sufficient noise at threshold that by itself occasionally leads to a spike (see eq. 10), such network responds with Up and Down states that are triggered by the noise and amplified by the recurrent excitatory synaptic activity (Figure 2C).

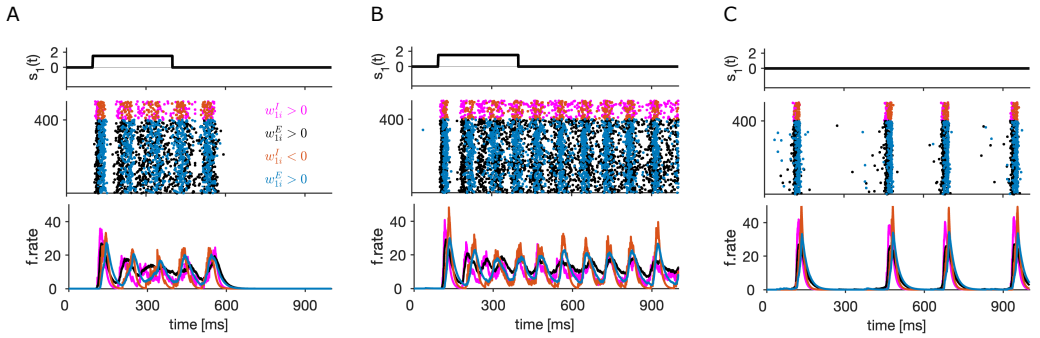

Figure 2: (A) Same as in Figure 1B, but using the matrix B with rank $M' = M = 3$. To construct the matrix B, we use the following column vectors: $\boldsymbol{b_1} = [0.4, 0.1, 0.1]^\intercal$, $\boldsymbol{b_2} = [0.1, 0.4, 0.1]^\intercal$, $\boldsymbol{b_3} = [0.1, 0.1, -0.4]^\intercal$. Other parameters as in Figure 1C. (B) As in (A), but with increased strength of E-E connectivity; $a = 0.04$. (C) Same as in (B), but without the external stimulus, $s_m(t) = 0\ \forall m \forall t$, and with increased noise; $\sigma_i^E = \sigma_i^I = 0.66\ \forall i$.

3 Discussion

We developed an E-I spiking network that optimizes information coding. The optimization is based on minimizing two cost functions comprising a quadratic estimation error and a quadratic regularizer and assuming an arbitrary linear input/output transformation. The general solution of an on-line minimization of cost functions with spikes is a generalized LIF model with a specific set of structural and dynamical features. Structural features include low-rank structured connectivity [40, 18, 20], where fast synapses between E and I neurons implement competition between neurons with similar decoding selectivity and slower synapses implement cooperation. Moreover, structural features resulting from optimization with spikes are external inputs that are shared across neurons, spike-triggered adaptation currents in single neurons and hyperpolarization-triggered rebound current in E neurons. Assuming heterogeneous decoding weights, all currents with the exception of leak currents are heterogeneous across neurons, increasing the biological plausibility of the model.

Similarly to the seminal paper by Boerlin and colleagues [2], the spiking network proposed here assumes that spikes are fired only if they minimize the quadratic error between the signal and the population read-out. As the network model is derived from such s cost function, efficient coding is a build-in property of the network. In earlier works on efficient coding [4], however, the leak current is not derived exactly but only added to the membrane equation for biological plausibility. Our derivations define an exact expression for the leak current, similarly as in [19]. Moreover, in most previous works on efficient coding with spikes, a single cost function is used to derive the membrane equations. Here, we propose two separate cost functions that relate to the activity of E and I neurons. A previous proposal that already used separate cost functions for E and I neurons [6] yielded a network similar to our simplified version of the network, but the behavior or their network is markedly different than our simplified model, due to very long and unrealistic synaptic time constants they use.

Most previous proposals on efficient coding [5, 6, 19, 4] assume a specific computation, namely a leaky integration of external inputs (but see [2]). We here developed the case of the general linear transformation between the external inputs and the internal signals of the network with a biologically plausible spiking network. This generalization allowed us to formulate a biologically plausible expression for E-E connections and to gain insights on how the structure of the mixing matrix qualitatively changes network responses. Without mixing of features, only neurons that are aligned to the external stimulus, and consequently driven by the feedforward current, participate in network responses. With increased complexity of E-E connections, we also observe activation of neurons that are not driven by the feedforward current, but are instead engaged through recurrent E-E connections. Moreover, in the absence of the external stimulus and with sufficient background noise, such a complex network produces spontaneous Up and Down states that could describe spontaneous Up and Down states observed in biological networks [39].

As it follows from our results, the minimization of the quadratic cost function with spikes results in a generalized LIF model, a spiking neuron model that has been shown to provide a good fit to the spiking dynamics of biological networks [12, 14]. An important advantage of the present computational framework over previously proposed generalized LIF models is the access to a functional interpretation of network parameters. In general, a change in a network parameter can have a strong influence on network's dynamics. Besides characterizing the effect on neural dynamics, it is insightful to also have the interpretation of the role of a specific parameter on the functional/computational level. In the present setting, parameters of the spiking network are directly related to the variables minimized in cost functions, which allows to connect their effect on dynamics with their role for network's computation.

In the present work, we emphasized the theoretical development of an optimization-based approach towards a biologically plausible spiking neural network with complex responses. Future work could introduce performance measures, for example by measuring the distance between the signal and the population read-out, or by using information-theoretic measures [3, 27, 35]. Moreover, future work could perform an analysis of network parameters and assess how are particular dynamical phenomena, such as spontaneous Up and Down states, related to the information processing in the spiking network. Finally, the framework can possibly be extended to include further biophysical properties of cortical networks, such as spatial extension of tuning curves, several types of inhibitory neurons [45], or non-linear computations [41].

References

[1] A. Alemi, C. Machens, S. Deneve, and J.-J. Slotine. Learning nonlinear dynamics in efficient, balanced spiking networks using local plasticity rules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[2] M. Boerlin, C. K. Machens, and S. Denève. Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comput Biol*, 9(11):e1003258, 2013.

[3] A. Borst and F. E. Theunissen. Information theory and neural coding. *Nature neuroscience*, 2(11):947–957, 1999.

[4] R. Bourdoukan, D. Barrett, S. Deneve, and C. K. Machens. Learning optimal spike-based representations. *Advances in neural information processing systems*, 25, 2012.

[5] W. Brendel, R. Bourdoukan, P. Vertechi, C. K. Machens, and S. Denéve. Learning to represent signals spike by spike. *PLoS computational biology*, 16(3):e1007692, 2020.

[6] M. Chalk, B. Gutkin, and S. Deneve. Neural oscillations as a signature of efficient coding in the presence of synaptic delays. *Elife*, 5:e13824, 2016.

[7] S. N. Chettih and C. D. Harvey. Single-neuron perturbations reveal feature-specific competition in v1. *Nature*, 567(7748):334–340, 2019.

[8] N. S. Desai, L. C. Rutherford, and G. G. Turrigiano. Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature neuroscience*, 2(6):515–520, 1999.

[9] A. A. Faisal, L. P. Selen, and D. M. Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9 (4):292–303, 2008.

[10] J. A. Gallego, M. G. Perich, R. H. Chowdhury, S. A. Solla, and L. E. Miller. Long-term stability of cortical population dynamics underlying consistent behavior. *Nature neuroscience*, 23(2):260–270, 2020.

[11] W. Gerstner and W. M. Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.

[12] W. Gerstner and R. Naud. How good are neuron models? *Science*, 326(5951):379–380, 2009.

[13] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[14] R. Jolivet, F. Schürmann, T. K. Berger, R. Naud, W. Gerstner, and A. Roth. The quantitative single-neuron modeling competition. *Biological cybernetics*, 99(4):417–426, 2008.

[15] M. Kafashan, A. W. Jaffe, S. N. Chettih, R. Nogueira, I. Arandia-Romero, C. D. Harvey, R. Moreno-Bote, and J. Drugowitsch. Scaling of sensory information in large neural populations shows signatures of information-limiting correlations. *Nature communications*, 12(1):1–16, 2021.

[16] C. Kayser, N. K. Logothetis, and S. Panzeri. Millisecond encoding precision of auditory cortex neurons. *Proceedings of the National Academy of Sciences*, 107(39):16976–16981, 2010.

[17] P. D. King, J. Zylberberg, and M. R. DeWeese. Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of v1. *Journal of Neuroscience*, 33(13):5475–5485, 2013.

[18] H. Ko, S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjöström, and T. D. Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345):87–91, 2011.

[19] V. Koren and S. Denève. Computational account of spontaneous activity as a signature of predictive coding. *PLoS computational biology*, 13(1):e1005355, 2017.

[20] V. Koren, A. R. Andrei, M. Hu, V. Dragoi, and K. Obermayer. Pairwise synchrony and correlations depend on the structure of the population code in visual cortex. *Cell Reports*, 33(6):108367, 2020.

[21] M. S. Lewicki. Efficient coding of natural sounds. *Nature neuroscience*, 5(4):356–363, 2002.

[22] Q. Li and C. Pehlevan. Minimax dynamics of optimally balanced spiking networks of excitatory and inhibitory neurons. *Advances in Neural Information Processing Systems*, 33:4894–4904, 2020.

[23] Y.-H. Liu and X.-J. Wang. Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. *Journal of computational neuroscience*, 10(1):25–45, 2001.

[24] A. Lüthi and D. A. McCormick. H-current: properties of a neuronal and network pacemaker. *Neuron*, 21 (1):9–12, 1998.

[25] F. Mastrogiuseppe and S. Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623, 2018.

[26] D. A. McCormick, B. W. Connors, J. W. Lighthall, and D. A. Prince. Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of neurophysiology*, 54(4):782–806, 1985.

[27] M. Montemurro and S. Panzeri. Optimal information decoding from neuronal populations with specific stimulus selectivity. *Advances in Neural Information Processing Systems*, 17, 2004.

[28] I. Nemenman, G. D. Lewen, W. Bialek, and R. R. de Ruyter van Steveninck. Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS computational biology*, 4(3):e1000025, 2008.

[29] J. E. Niven. Neuronal energy consumption: biophysics, efficiency and evolution. *Current opinion in neurobiology*, 41:129–135, 2016.

[30] B. A. Olshausen and D. J. Field. Natural image statistics and efficient coding. *Network: computation in neural systems*, 7(2):333, 1996.

[31] S. Panzeri, R. S. Petersen, S. R. Schultz, M. Lebedev, and M. E. Diamond. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29(3):769–777, 2001.

[32] H.-C. Pape. Queer current and pacemaker: the hyperpolarization-activated cation current in neurons. *Annual review of physiology*, 58(1):299–327, 1996.

[33] A. Payeur, J. Guerguiev, F. Zenke, B. A. Richards, and R. Naud. Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nature neuroscience*, 24(7):1010–1019, 2021.

[34] R. S. Petersen, S. Panzeri, and M. E. Diamond. Population coding of stimulus location in rat somatosensory cortex. *Neuron*, 32(3):503–514, 2001.

[35] G. Pica, E. Piasini, H. Safaai, C. Runyan, C. Harvey, M. Diamond, C. Kayser, T. Fellin, and S. Panzeri. Quantifying how much sensory information in a neural code is relevant for behavior. *Advances in Neural Information Processing Systems*, 30, 2017.

[36] A. Pignatelli, M. Borin, A. Fogli Iseppe, C. Gambardella, and O. Belluzzi. The h-current in periglomerular dopaminergic neurons of the mouse olfactory bulb. *PLoS One*, 8(2):e56571, 2013.

[37] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[38] C. A. Runyan, E. Piasini, S. Panzeri, and C. D. Harvey. Distinct timescales of population coding across cortex. *Nature*, 548(7665):92–96, 2017.

[39] M. V. Sanchez-Vives and D. A. McCormick. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature neuroscience*, 3(10):1027–1034, 2000.

[40] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS biology*, 3(3):e68, 2005.

[41] D. Thalmeier, M. Uhlmann, H. J. Kappen, and R.-M. Memmesheimer. Learning universal computations with spikes. *PLoS computational biology*, 12(6):e1004895, 2016.

[42] R. Tremblay, S. Lee, and B. Rudy. Gabaergic interneurons in the neocortex: from cellular properties to circuits. *Neuron*, 91(2):260–292, 2016.

[43] M. Valente, G. Pica, G. Bondanelli, M. Moroni, C. A. Runyan, A. S. Morcos, C. D. Harvey, and S. Panzeri. Correlations enhance the behavioral readout of neural population activity in association cortex. *Nature neuroscience*, 24(7):975–986, 2021.

[44] C. P. Wonders and S. A. Anderson. The origin and specification of cortical interneurons. *Nature Reviews Neuroscience*, 7(9):687–696, 2006.

[45] J. Yu, H. Hu, A. Agmon, and K. Svoboda. Recruitment of gabaergic interneurons in the barrel cortex during active tactile behavior. *Neuron*, 104(2):412–427, 2019.

[46] M. Zhu and C. J. Rozell. Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. *PLoS computational biology*, 9(8):e1003191, 2013.