Supplementary text to "Biologically plausible solutions for spiking networks with efficient coding"

Veronika Koren
Department of Excellence for Neural Information Processing
Center for Molecular Neurobiology (ZMNH)
University Medical Center Hamburg-Eppendorf (UKE)
Falkenried 94, 20251 Hamburg, Germany
v.koren@uke.de

Stefano Panzeri
Department of Excellence for Neural Information Processing
Center for Molecular Neurobiology (ZMNH)
University Medical Center Hamburg-Eppendorf (UKE)
Falkenried 94, 20251 Hamburg, Germany
Istituto Italiano di Tecnologia
Genoa, Italy
s.panzeri@uke.de

Here we provide further details about biologically relevant solutions for a spiking neural network with efficient coding. The analytical derivation of the spiking neural network is split in three parts,

1. Definitions and analytical derivations of the loss function.
2. Analytical derivation of the temporal dynamic of the membrane potentials.
3. Expressing the efficient spiking network as a generalized leaky integrate-and-fire neuron model.

Part 1 of the derivation follows closely previous works of efficient coding with spikes (1; 5), however, with important conceptual differences. The first part of the derivation of our framework therefore differs from the previous work by imposing the E-I network architecture. We summarize the first part of the derivation in the section 1. The second step of the derivation deviates in several ways from the one in (1), and we provide an overview of it in section 2. In section 3, we examine derived expression of membrane currents about their biological plausibility and express the network as a generalized leaky integrate-and-fire network model. In last section (section 4) we present and comment on alternative solutions to the ones described in section 3.

1 From the loss function to the membrane potentials

In (1), a recurrently connected spiking network is developed from a single loss function of the form:

$$\mathrm{L}(t) = \sum_{m=1}^{M} (x_m(t) - \hat{x}_m(t))^2 + \nu \sum_{i=1}^{N} r_i(t) + \mu \sum_{i=1}^{N} r_i^2(t) \tag{1}$$

with $\nu, \mu > 0$ and where $x_m(t)$, $\hat{x}_m(t)$ are the signal and the estimate of the $m-$th feature of the stimulus, respectively, and $r_i(t)$ is the low-pass filtered spike train of the neuron i.

In the present work, we introduced two loss functions, one for the excitatory (E) and one for the inhibitory (I) neurons, which allowed us to define biologically plausible membrane equations for E and I neurons. We consider a sensory stimulus with M independent features, encoded by N_E excitatory (E) and N_I inhibitory (I) neurons. We define the loss functions related to the activity of E and I neurons as follows:

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

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

with $\mu_E, \mu_I > 0$ and where $\hat{x}_m^E(t)$ ($\hat{x}_m^I(t)$) is the estimate of the desired signal $x_m(t)$, formed by the population read-out of the spiking activity of E (I) neurons. The variable $r_i^y(t)$ is the low-pass filtered spike train of neuron i,

$$\dot{r}_i^y(t) = -\alpha_i^y r_i^y(t) + f_i^y(t), \qquad y \in \{E, I\} \tag{3}$$

with $\alpha_i^y > 0 \; \forall i$ the inverse time constant of the single neuron read-out, $(\alpha_i^y)^{-1} = \tau_i^{r,y}$. Note that $r_i^y(t)$ is proportional to the instantaneous firing rate of the neuron i.

We write definitions for the signal $\boldsymbol{x}(t) = [x_1(1), \ldots, x_M(t)]^\intercal$ and the estimates $\hat{\boldsymbol{x}}_y(t) = [x_1^y(t), \ldots, x_M^y(t)]^\intercal, y \in \{E, I\}$, as follows:

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + \boldsymbol{s}(t) \tag{4a}$$

$$\dot{\hat{\boldsymbol{x}}}_E(t) = -\lambda_E \hat{\boldsymbol{x}}_E(t) + W_E \boldsymbol{f}_E(t) \tag{4b}$$

$$\dot{\hat{\boldsymbol{x}}}_I(t) = -\lambda_I \hat{\boldsymbol{x}}_I(t) + W_I \boldsymbol{f}_I(t) \tag{4c}$$

where $A \in \mathbb{R}^{M\mathrm{x}M}$ is the mixing matrix of the features of sensory stimuli $\boldsymbol{s}(t) = [s_1(t), \ldots, s_M(t)]^\intercal$. Scalars $\lambda_E, \lambda_I > 0$ are the inverse time constants of the population read-out of E and I neurons, respectively, with $(\lambda_y)^{-1} = \tau_y$. Vector of spike trains, $\boldsymbol{f}_y(t) = [f_1^y(t), \ldots, f_{N_y}^y(t)]^\intercal$, assembles spike trains across excitatory ($y = E$) and inhibitory ($y = I$) neurons, where the spike train of the neuron i is defined as the sum of Dirac delta distributions, $f_i^y(t) = \sum_k \delta(t - t_{i,y}^k)$, with $t_{i,y}^k$ the $k-$th spike time of neuron i.

Every neuron is assigned a decoding vector, $\boldsymbol{w}_i^y = [w_{1i}^y, \ldots, w_{Mi}^y]^\intercal$, as $m-$th element of the decoding vector relates the spike train of the neuron i to the $m-$th dimension of the estimate $\hat{x}_m^y(t)$. The weighting matrix $W_y \in \mathbb{R}^{M\mathrm{x}N_y}$ assembles decoding vectors across neurons, $W_y = [\boldsymbol{w}_1^y, \ldots, \boldsymbol{w}_N^y]$.

We also gather the low-pass filtered spike trains across neurons, $\boldsymbol{r}_y(t) = [r_1^y(t), \ldots, r_{N_y}^y]^\intercal$, and express their definition in vector notation:

$$\begin{aligned}\dot{\boldsymbol{r}}_E(t) &= -\Lambda_E^r \boldsymbol{r}_E(t) + \boldsymbol{f}_E(t)\\ \dot{\boldsymbol{r}}_I(t) &= -\Lambda_I^r \boldsymbol{r}_I(t) + \boldsymbol{f}_I(t)\end{aligned} \tag{5}$$

with $\Lambda_y^r = \mathrm{diag}(\boldsymbol{\alpha}_y)$ a square diagonal matrix with diagonal $\boldsymbol{\alpha}_y = [\alpha_1^y, \ldots, \alpha_{N_y}^y]^\intercal$. Note that the number of features encoded by the network, M, determines the dimensionality of the desired signal $\boldsymbol{x}(t)$ and of the estimates $\hat{\boldsymbol{x}}_E(t)$ and $\hat{\boldsymbol{x}}_I(t)$. The number of neurons N_y is typically larger than the number of features M.

Similar to refence (1), we assume that a spike of the neuron i at time t will be fired only if this minimizes the loss function. Additionally, we assume that in a biological network, the condition on spiking is subjected to noise. It is unlikely that biological circuits could implement spiking as an entirely noiseless process. The condition to have a spike in the neuron i of cell type y is formulated as:

$$\mathrm{L}_y\left(t^+ | [f_i^y(t^+) = 1] + \eta_i^y(t^+)\right) < \mathrm{L}_y\left(t^- | [f_i^y(t^-) = 0]\right), \tag{6}$$

where $\eta_i^y(t^+) = \sigma_i^y \xi_i^y(t)$ models the noise at threshold crossing. The noise at threshold crossing has intensity σ_i^y while $\xi_i^y(t)$ is a Gaussian random process with zero mean and unit standard deviation, $\xi_i^y(t) \sim \mathcal{N}(0,1)$, with $\xi_i^y(t)$ independent and identically distributed over time, across neurons and across the two cell types.

Taking into account the effect of a spike on the estimates (eq. 4b-4c) and on the low-pass filtered spike trains (eq. 3) and applying those in the condition on spiking (eq. 6), we arrive to the following condition for the spiking neuron i:

$$\begin{aligned}\boldsymbol{w}_E^\intercal\left(\boldsymbol{x}(t) - \hat{\boldsymbol{x}}_E(t)\right) - \mu_E r_i^E(t) &> \frac{1}{2}\left(\|\boldsymbol{w}_i^E\|_2^2 + \mu_E\right) + \sigma_i^E \xi_i^E(t)\\ \boldsymbol{w}_I^\intercal\left(\hat{\boldsymbol{x}}_E(t) - \hat{\boldsymbol{x}}_I(t)\right) - \mu_I r_i^I(t) &> \frac{1}{2}\left(\|\boldsymbol{w}_i^I\|_2^2 + \mu_I\right) + \sigma_i^I \xi_i^I(t)\end{aligned} \tag{7}$$

with $\|\boldsymbol{w}_i^y\|_2^2 = \sum_{m=1}^M (w_{mi}^y)^2$ the squared length of decoding vector of the neuron i. As in (2; 1), we interpret the left-hand side of eq. 7 as the membrane potential of neuron i and the right-hand side as the firing threshold,

$$\begin{aligned}u_i^E(t) &\equiv \boldsymbol{w}_E^\intercal\left(\boldsymbol{x}(t) - \hat{\boldsymbol{x}}_E(t)\right) - \mu_E r_i^E(t)\\ u_i^I(t) &\equiv \boldsymbol{w}_I^\intercal\left(\hat{\boldsymbol{x}}_E(t) - \hat{\boldsymbol{x}}_I(t)\right) - \mu_I r_i^I(t)\\ \vartheta_i^y &\equiv \frac{1}{2}\left(\|\boldsymbol{w}_i^I\|_2^2 + \mu_y\right) + \sigma_1^y \xi_1^y(t), \qquad y \in \{E, I\}.\end{aligned} \tag{8}$$

Note that the firing threshold of the neuron i is proportional to the squared length of the decoding vector, $||\boldsymbol{w}_i^y||_2^2 = \sum_m (w_{mi}^y)^2$, and the constant of the regularizer μ_y (eq. 8). The vector of the membrane potentials for N_E excitatory and N_I inhibitory neurons, can now be written in vector notation as follows:

$$\begin{aligned}\boldsymbol{u}_E(t) &= W_E^\intercal\left(\boldsymbol{x}(t) - \hat{\boldsymbol{x}}_E(t)\right) - \mu_E \boldsymbol{r}_E(t)\\ \boldsymbol{u}_I(t) &= W_I^\intercal\left(\hat{\boldsymbol{x}}_E(t) - \hat{\boldsymbol{x}}_I(t)\right) - \mu_I \boldsymbol{r}_I(t).\end{aligned} \tag{9}$$

The membrane potentials $\boldsymbol{u}_y(t)$ are thus given by the projection of the coding error on the matrix of decoding weights, and in addition depend on the spiking frequency of the local neuron (eq. 9).

2 Dynamics of the membrane potential

The second part consists in calculating the difference equation for membrane potentials. To obtain a difference equation, we take derivatives with respect to time of $\boldsymbol{u}_E(t)$ and $\boldsymbol{u}_I(t)$,

$$\begin{aligned}\dot{\boldsymbol{u}}_E(t) &= W_E^\intercal\left(\dot{\boldsymbol{x}}(t) - \dot{\hat{\boldsymbol{x}}}_E(t)\right) - \mu_E\dot{\boldsymbol{r}}_E(t)\\ \dot{\boldsymbol{u}}_I(t) &= W_I^\intercal\left(\dot{\hat{\boldsymbol{x}}}_E(t) - \dot{\hat{\boldsymbol{x}}}_I(t)\right) - \mu_I\dot{\boldsymbol{r}}_I(t).\end{aligned} \tag{10}$$

In eq. (10) we use definitions of the temporal derivatives of the signal (eq. 4a), the estimate by E neurons (eq. 4b), the estimate by I neurons (eq. 4c), and the definition of low-pass filtered spike trains (eq. 3). Without loss of generality, we also use the following substitutions:

$$A = B - \lambda_E\mathbf{I}^{M\mathrm{x}M} \tag{11a}$$

$$\Delta_E^r = \lambda_E\mathbf{I}^{[N_E\mathrm{x}N_E]} - \Lambda_E^r \tag{11b}$$

$$\Delta_I^r = \lambda_I\mathbf{I}^{[N_I\mathrm{x}N_I]} - \Lambda_I^r \tag{11c}$$

where $\mathbf{I}$ is an identity matrix, and $\Delta_y^r \in \mathbb{R}^{N_y\mathrm{x}N_y}$ are square diagonal matrices with diagonal elements $\delta_i^{r,y} = \lambda_y - \alpha_i^y$, for $i = 1, \dots, N_y$. Diagonal elements of Δ_E^r and Δ_I^r therefore evaluate the difference of the time constants of the population read-out (eq. 4b-4c) and the single neuron read-out (eq. 5) in E and I neurons, respectively. Using substitutions in eq. (11a)-11c, the exact solutions for the time-derivative of the membrane potentials are:

$$\dot{\boldsymbol{u}}_E(t) \;=\; -\lambda_E\boldsymbol{u}_E(t) + W_E^\intercal\boldsymbol{s}(t) - W_E^\intercal W_E\boldsymbol{f}_E(t) + W_E^\intercal B\boldsymbol{x}(t) - \mu_E\Delta_E^r\boldsymbol{r}_E(t) - \mu_E\boldsymbol{f}_E(t) \tag{12a}$$

$$\dot{\boldsymbol{u}}_I(t) = -\lambda_I\boldsymbol{u}_I(t) + W_I^\intercal W_E\boldsymbol{f}_E(t) - W_I^\intercal W_I\boldsymbol{f}_I(t) + \delta_I W_I^\intercal B\hat{\boldsymbol{x}}_E(t) - \mu_I\Delta_I^r\boldsymbol{r}_I(t) - \mu_I\boldsymbol{f}_I(t) \tag{12b}$$

with $\delta_I = \lambda_I - \lambda_E$. Right-hand side of eqs. 12a-12b comprise a term proportional to the leak current, synaptic terms $W_y^\intercal W_z\boldsymbol{f}_z(t)$ for $y, z \in \{E, I\}$, terms involving the signal $\boldsymbol{x}(t)$ and the estimate $\hat{\boldsymbol{x}}_E(t)$, terms with local currents with slower dynamics, $\mu_y\Delta_y^r\boldsymbol{r}_y(t)$, and local terms with faster dynamics $\mu_y\boldsymbol{f}_y(t)$. Excitatory neurons in addition have a term proportional to the feedforward current, $W_E^\intercal\boldsymbol{s}(t)$. Eqs. 12a-12b do not yet express a biologically plausible membrane equation, and several of the terms have to be constrained in order to obtain a framework that is consistent with know properties of biological networks.

3 Efficient spiking network as a generalized leaky integrate-and-fire neuron model.

The following section considers biologically plausible and computationally efficient solutions derived from eqs. 12a-12b. We examine the terms one by one, and express a biologically plausible solution in the form of an E-I network of generalized leaky integrate-and-fire neurons.

Leak current The terms $-\lambda_y\boldsymbol{u}_y(t)$ for $y \in \{E, I\}$ define the leak current in E and I cell type. In neuron i of cell type $y \in \{E, I\}$, the leak current is:

$$I_i^{\text{leak } y} \propto -\lambda_y u_i^y(t) = -\frac{1}{\tau_y}u_i^y(t), \qquad y \in \{E, I\} \tag{13}$$

with $\tau_y = (\lambda_y)^{-1}$ the membrane time constant of E ($y = E$) and I ($y = I$) neurons. Leak currents in eq. (13) result from absorbing terms that define the membrane potential, as in eq. (9), and are, contrary to the procedure in (1), calculated without approximations. In the E cell type, in particular, inserting the eq. (11a) in eq. (4a), we get

$$A\boldsymbol{x}(t) = B\boldsymbol{x}(t) - \lambda_E \mathbf{I}^{M \times M}\boldsymbol{x}(t). \tag{14}$$

The leak current in the E cell type (eq. 13 with $y = E$) absorbs, among others, the term $-\lambda_E \mathbf{I}^{M \times M}\boldsymbol{x}(t)$, while the remaining term, $B\boldsymbol{x}(t)$, is part of a synaptic current that is discussed further on. Similar leak term has been obtained in a previous work on efficient spiking networks (5), where the leak also emerged from analytical treatment of the loss function. In a previous work (5), the same leak current has been analytically derived in a simplified network with diagonal matrix $A = -\lambda_E \mathbf{I}^{M \times M}$, where it has also been assumed that the time constant of the signal $\boldsymbol{x}(t)$ is equivalent to the time constant of the neural membrane τ.

Feedforward current The term $W_E^\intercal \boldsymbol{s}(t)$ in the E cell type (eq. 12a) defines a feedforward current and has been proposed in previous works (1; 2). The feedforward current to the neuron i is proportional to the sum of feedforward inputs $s_m(t)$, weighted by decoding weights of the neuron,

$$I_i^{\mathrm{ff}}(t) \propto (\boldsymbol{w}_i^E)^\intercal \boldsymbol{s}(t) = \sum_{m=1}^{M} w_{mi}^E s_m(t). \tag{15}$$

In case we assume the variables $s_m(t)$, for $m = 1 \ldots, M$, to correspond to M features of an external stimulus that the network is receptive to (i.e., sensory features of an image such as the orientation, the spatial frequency, the color, etc.), the eq. (15) is a plausible expression of the feedforward current. This is also the interpretation that we follow in the main paper.

Fast synaptic currents In eqs. (12a)-(12b), terms of the form $W_y^\intercal W_z \boldsymbol{f}_z(t)$ with $\{yz\} \in \{EE, IE, II\}$ define fast synaptic interactions between E-to-E, E-to-I and I-to-I neurons. We write the absolute value of fast synaptic currents at a postsynaptic neuron i as the sum of presynaptic inputs as follows:

$$|\tilde{I}_i^{yz}(t)| \propto \sum_{j=1}^{N_z} (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z f_j^z(t), \qquad \{yz\} \in \{IE, II, EE\}, \tag{16}$$

with $\boldsymbol{w}_i^y$ the decoding vector of the postsynaptic neuron, and $\boldsymbol{w}_j^z$, $f_j^z(t)$ the decoding vector and the spike train of the presynaptic neuron, respectively. These currents are in general not biologically plausible and have to be constrained. The sign of the synaptic interaction from the presynaptic neuron j of cell type z to the postsynaptic neuron i of cell type y depends on the similarity of decoding vectors between the presynaptic and the postsynaptic neuron:

$$\begin{aligned} &(\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z f_j^z(t) > 0 \text{ if } (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z > 0 \\ &(\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z f_j^z(t) < 0 \text{ if } (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z < 0. \end{aligned} \tag{17}$$

If the two neurons have similar decoding vectors, dot product of their decoding vectors is positive, while neuronal pairs with dissimilar decoding vectors have a negative dot product of their decoding vectors. Irrespectively of the sign in front of the synaptic current (see eq. 12a-12b), therefore, the same presynaptic neuron j sends positive (excitatory) and negative (inhibitory) synaptic currents to other neurons, depending on the similarity of decoding vectors of the presynaptic and the postsynaptic neuron. This is inconsistent with Dale's law that constrains a particular neuron to only send either excitatory or inhibitory currents to the postsynaptic neuron, but not both. A simple solution that enforces Dale's law consists in removing connections between neurons with dissimilar selectivity (e.g., neuronal pairs with negative dot product of weight vectors; see the second line in eq. 17). We get:

$$I_i^{\text{fast } IE}(t) \propto \sum_{j=1}^{N_E} C_{ij}^{IE} f_j^E(t), \qquad I_i^{\text{fast } II}(t) \propto -\sum_{\substack{j=1\\ j\neq i}}^{N_I} C_{ij}^{II} f_j^I(t), \qquad I_i^{\text{fast } EE}(t) \propto -\sum_{j=1}^{N_E} C_{ij}^{EE} f_j^E(t), \tag{18}$$

with C^{yz}, $\{yz\} \in \{IE, II, EE\}$ the connectivity matrix between the presynaptic population z and the postsynaptic population y,

$$C_{ij}^{yz} = \begin{cases} (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z, & \text{if } (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Currents received by I neurons, $I_i^{\text{fast } IE}(t)$ and $I_i^{\text{fast } II}(t)$, as prescribed by eqs. 18-19, obey Dale's law, while while the current $I_i^{\text{fast } EE}(t)$ is inconsistent with Dale's law. Elements of the matrix C^{yz} in eq. 19 are always positive and the sign of the synaptic current is given by the sign in front of the synaptic term in eq. 18. In currents received by I neurons, we get a positive (excitatory) current originating from E neurons ($I_i^{IE}(t)$), and a negative (inhibitory) current originating from I neurons ($I_i^{II}(t)$), which is consistent with Dale's law. The current $I_i^{\text{fast } EE}(t)$, on the contrary, originate from E neurons, but is negative (inhibitory). Since in biological networks, E neurons cannot send inhibitory currents, we make the following replacement: $-W_E^\intercal W_E \boldsymbol{f}_E(t) \approx -W_E^\intercal W_I \boldsymbol{f}_I(t)$, and get the following fast synaptic currents:

$$I_i^{\text{fast } IE}(t) \propto \sum_{j=1}^{N_E} C_{ij}^{IE} f_j^E(t), \qquad I_i^{\text{fast } II}(t) \propto -\sum_{\substack{j=1\\ j\neq i}}^{N_I} C_{ij}^{II} f_j^I(t), \qquad I_i^{\text{fast } EI}(t) \propto -\sum_{j=1}^{N_I} C_{ij}^{EI} f_j^I(t), \tag{20}$$

with the matrix of fast synaptic connections as in eq. 19. The current $I_i^{\text{fast } EI}(t)$ is now an inhibitory synaptic current and it originates from I neurons (right-most term in eq. 20), thus contributing fast inhibition to E neurons that is consistent with Dale's law.

Synaptic currents with kinetics of low-pass filtered spikes Next, we address synaptic terms $W_E^\intercal B\boldsymbol{x}(t)$ in the E cell type and $\delta_I W_I^\intercal B\hat{\boldsymbol{x}}_E(t)$ in the I cell type (see eqs. 12a-12b). These terms will give synaptic currents with kinetics of low-pass filtered spikes (see eqs. 4a and 4b), thus describing synaptic transmission with slower dynamics.

In the E cell type (eq. 12a), the term $W_E^\intercal B\boldsymbol{x}(t)$ contains the signal $\boldsymbol{x}(t)$, a variable that does not by itself define a biologically plausible current to single neurons. By construction of the loss function of E neurons (eq. 2a), the signal $\boldsymbol{x}(t)$ is approximated by the E estimate $\hat{\boldsymbol{x}}_E(t)$, allowing us to make the substitution $\boldsymbol{x}(t) \approx \hat{\boldsymbol{x}}_E(t)$. Using the definition of the E estimate (eq. 4b), we get the following E-to-E synaptic current to the postsynaptic neuron i:

$$\begin{aligned}
&\dot{I}_i^{EE}(t) \propto -\frac{1}{\tau_E^{\text{syn}}} I_i^{EE}(t) + \sum_{\substack{j=1\\ j\neq i}}^{N_E} D_{ij}^{EE} f_j^E(t), \qquad \tau_E^{\text{syn}} = \tau_E \\
&D_{ij}^{EE} = \begin{cases} (\boldsymbol{w}_i^E)^\intercal B \boldsymbol{w}_j^E, & \text{if } (\boldsymbol{w}_i^E)^\intercal \boldsymbol{w}_j^E > 0, \qquad \text{B positive semi-def.} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{21}$$

To ensure that synaptic interactions are consistently excitatory, we only allowed connections between neurons with similar selectivity and constrained the matrix $B = (b_{mn});\ m, n = 1, \dots, M$, to be positive semi-definite.

In the I cell type in eq. (12b), we have the term $\delta_I W_I^\intercal B \hat{\boldsymbol{x}}_E(t)$ with $\delta_I = \lambda_I - \lambda_E$ and $\lambda_y = (\tau_y)^{-1}$ for $y \in \{E, I\}$. We again use the definition of the E estimate (eq. 4b), and get the following E-to-I synaptic current in I neurons:

$$\begin{aligned}
&\dot{I}_i^{\text{slow}IE}(t) \propto -\frac{1}{\tau_E^{\text{syn}}} I_i^{\text{slow }IE}(t) + (\frac{1}{\tau_I} - \frac{1}{\tau_E}) \sum_{j=1}^{N_E} D_{ij}^{IE} f_j^E(t), \qquad \tau_I < \tau_E, \\
&D_{ij}^{IE} = \begin{cases} (\boldsymbol{w}_i^I)^\intercal B \boldsymbol{w}_j^E, & \text{if } (\boldsymbol{w}_i^I)^\intercal \boldsymbol{w}_j^E > 0, \quad \text{B positive semi-def.} \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{22}$$

Since E-to-I currents originate from E neurons, they have to be excitatory. To ensure that E-to-I synapses are consistently excitatory (and taking into account that the matrix B has been constrained to be positive semi-definite in eq. 21), we get the following constraint on time constants: $\tau_I < \tau_E$, constraining the membrane time constant in I neurons to be faster than in E neurons. In summary, the strength of slower synaptic currents is proportional to the similarity of decoding vectors of the presynaptic and the postsynaptic neuron, similarly as with fast synaptic currents (eq. 20). Moreover, slower synaptic currents in addition depend on the matrix B (eqs. 21-22).

We note that E-to-E synaptic currents in eq. 21 as well as E-to-I synaptic currents in eq. 22 have kinetics of low-pass filtered spike trains of excitatory presynaptic neurons. Defining a low-pass filtered spike train with synaptic time constant τ_E^{syn}, we can simplify the notation and write these synaptic currents as follows:

$$\begin{aligned}
&I_i^{EE}(t) \propto \sum_{\substack{j=1\\ j\neq i}}^{N_E} D_{ij}^{EE} z_j^E(t), \qquad \tau_E^{\text{syn}} = \tau_E \\
&I_i^{\text{slow }IE}(t) \propto (\frac{1}{\tau_I} - \frac{1}{\tau_E}) \sum_{j=1}^{N_E} D_{ij}^{IE} z_j^E(t), \qquad \tau_I < \tau_E, \\
&\dot{z}_i^E(t) = -\frac{1}{\tau_E^{\text{syn}}} z_i^E(t) + f_i^E(t),
\end{aligned} \tag{23}$$

with the matrix D^{EE} and D^{IE} as in eqs. 21-22.

Local currents The terms $-\mu_y \Delta_y^r \boldsymbol{r}_y(t)$ in eq. 12a-12b define local, spike-triggered currents with dynamics of the low-pass filtered spike train $r_i^y(t)$. We defined Δ_y^r as a diagonal matrix with $i-$th diagonal element $(\Delta_y^r)_{ii} = \lambda_y - \alpha_i^y$ (eq. 5). Using that $\lambda_y = (\tau_y)^{-1}$ and $\alpha_i^y = (\tau_i^{r,y})^{-1}$ are inverse time constants, we can write the local current in neuron i as:

$$I_i^{\text{local } y}(t) \propto -\mu_y \left(\frac{1}{\tau_y} - \frac{1}{\tau_i^{r,y}} \right) r_i^y(t), \qquad y \in \{E, I\}. \tag{24}$$

Using the definition of the low-pass filtered spike train $r_i^y(t)$ in eq. 3, we can rewrite eq. 24 as leaky integration of spike trains $f_i^y(t)$:

$$\dot{I}_i^{\text{local } y}(t) \propto -\frac{1}{\tau_i^{r,y}} I_i^{\text{local } y}(t) - \mu_y \left(\frac{1}{\tau_y} - \frac{1}{\tau_i^{r,y}} \right) f_i^y(t), \qquad y \in \{E, I\}. \tag{25}$$

Local currents as in eq. (25) are biologically plausible, however, different solutions are obtained depending on the relation of time constants between the population read-out τ_y and the single neuron read-out $\tau_i^{r,y}$. The regularizer μ_y is non-negative by definition (see eqs.2a-2b) and does not influence the sign of the local current in eq. (25). If we constrains the time constant of the single neuron read-out to be longer than the time constant of the population read-out: $\tau_i^{r,y} > \tau_y$, current in eq. 25 is negative (hyperpolarizing), and we interpret is as spike-triggered adaptation. If, on the contrary, we have the following relation of inverse time constants: $\tau_i^{r,y} < \tau_y$, the local current in eq. 25 is positive (depolarizing), and we interpret it as spike-triggered facilitation. In the special case when the two time constants are equal, $\tau_i^{r,y} = \tau_y$, the local current vanishes. Note that this special case has been assumed in previous works (2; 1; 5; 4; 3), while we here developed a more general solution. Note that the kinetics as well as the strength of local currents is heterogeneous across neurons due to the heterogeneity of the time constant $\tau_i^{r,y}$ across neurons (see eq. 25).

Reset current The last terms on the right-hand side of eq. 12a-12b is of the following form: $-\mu_y \boldsymbol{f}_y(t)$, and defines resetting of the local neuron after a spike. The reset current depends on the constant of the quadratic regularizer μ_y:

$$\begin{aligned} I_i^{\text{reset } E}(t) &= -\mu_E f_i^E(t) \\ I_i^{\text{reset } I}(t) &= -(\mu_I + C_{ii}^{II}) f_i^E(t) \end{aligned} \tag{26}$$

and in I neurons, we also have the contribution of the negative self-connection C_{ii}^{II} with C^{II} the matrix of recurrent inhibitory connections as in eq. 19. Since the regularizer μ_y is by definition positive, the reset current in eq. 26 is always a negative (hyperpolarizing) current, which ensures its biological plausibility as a current that resets the membrane potential after the neuron has reached the firing threshold.

Spike-triggered rebound current In the definition of the recurrent E-to-E synaptic current (eq. 23), we omitted the self-connection, since a self-connection is not a synaptic current. The self connection is activated by the spike of the local neuron and has the dynamics of the low-pass filtered spike train. We interpret this contribution as the local rebound current:

$$\begin{aligned} \dot{I}_i^{\text{rebound}}(t) &= -\frac{1}{\tau_h} I_i^{\text{rebound}}(t) + D_{ii}^{EE} f_i^E(t), \qquad \tau_h = \tau_E \\ D_{ii}^{EE} &= (\boldsymbol{w}_i^E)^\intercal B \boldsymbol{w}_i^E, \qquad B \text{ positive semi-def.} \end{aligned} \tag{27}$$

Rebound current is always a positive (depolarizing) current, since the coefficient D_{ii}^{EE} is given by the product of the decoding vector of the spiking neuron, $\boldsymbol{w}_i^E$, with the positive semi-definite matrix B. Immediately after a spike of the neuron i, the neuron is strongly hyperpolarized by the reset current (eq. 26). Hence, the rebound current in eq. 27 activates when the neuron is strongly hyperpolarized and counteracts the strong hyperpolarization with a depolarizing rebound current that brings the membrane potential towards the firing threshold.

Integrate-and-fire formulation Finally, we gather results and express the efficient spiking network as an E-I network of generalized LIF neurons. We use the fact that the activation of the reset current is instantaneous (eq. 26) and creates a jump in the membrane potential as the neuron reaches the threshold. The jump in the membrane potential corresponds to the amplitude of the reset current, and the dynamics of E and I neurons can be expressed as a generalized LIF neuron model:

$$\begin{aligned} \tau_E \dot{V}_i^E(t) &= -V_i^E(t) + I_i^{\text{ff}}(t) + I_i^{EI}(t) + I_i^{EE}(t) + I_i^{\text{local }E}(t) + I_i^{\text{rebound}}(t) \\ \tau_I \dot{V}_i^I(t) &= -V_i^I(t) + I_i^{IE}(t) + I_i^{II}(t) + I_i^{\text{local }I}(t) \\ \text{if } V_i^y(t^-) &\geq \vartheta_i^y(t^-) \rightarrow V_i^y(t^+) = V_i^{\text{reset }y}, \qquad y \in \{E, I\}, \end{aligned} \tag{28a}$$

The firing thresholds and resets are proportional to the regularizer μ_y and the squared length of the decoding vector $||\boldsymbol{w}_i^y||_2^2$:

$$\begin{aligned} \vartheta_I^y(t) &= \frac{1}{2}(\mu_y + ||\boldsymbol{w}_i^y||_2^2 + \sigma_i^y \xi_i^y(t), \qquad y \in \{E, I\} \\ V_i^{\text{reset }E} &= -\frac{1}{2}(\mu_E - ||\boldsymbol{w}_i^E||_2^2) \\ V_i^{\text{reset }I} &= -\frac{1}{2}(\mu_I + ||\boldsymbol{w}_i^I||_2^2), \end{aligned} \tag{28b}$$

and the currents are:

$$
\begin{aligned}
I_i^{\text{ff}}(t) &= \tau_E \sum_{m=1}^{M} w_{mi}^E s_m(t) \\
I_i^{EE}(t) &= \tau_E \sum_{\substack{j=1 \\ j\neq i}}^{N_E} D_{ij}^{EE} z_j^E(t) \\
I_i^{IE}(t) &= \tau_I \sum_{j=1}^{N_E} C_{ij}^{IE} f_j^E(t) + \left(1 - \frac{\tau_I}{\tau_E}\right) \sum_{j=1}^{N_E} D_{ij}^{IE} z_j^E(t), \qquad \tau_E > \tau_I \\
I_i^{II}(t) &= -\tau_I \sum_{\substack{j=1 \\ j\neq i}}^{N_I} C_{ij}^{II} f_j^I(t) \\
I_i^{EI}(t) &= -\tau_E \sum_{j=1}^{N_I} C_{ij}^{EI} f_j^I(t) \\
I_i^{\text{local } y}(t) &= -\mu_y \left(1 - \frac{\tau_y}{\tau_i^{r,y}}\right) r_i^y(t), \qquad y \in \{E, I\} \\
I_i^{\text{rebound}}(t) &= \tau_E D_{ii}^{EE} z_i^E(t),
\end{aligned}
\tag{28c}
$$

Matrices C_{ij}^{yz} and D_{ij}^{yE} determine the strength of the fast and slow channels in the synapse, respectively:

$$
\begin{aligned}
C_{ij}^{yz} &= \begin{cases} (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z, & \text{if } (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z > 0 \qquad \{yz\} \in \{IE, II, EI\} \\ 0 & \text{otherwise} \end{cases} \\
D_{ij}^{yE} &= \begin{cases} (\boldsymbol{w}_i^y)^\intercal B \boldsymbol{w}_j^E, & \text{if } (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^E > 0, \quad \text{B positive semi-def.}, \qquad y \in \{E, I\} \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
\tag{28d}
$$

while $z_i^E(t)$ and $r_i^y(t)$ are low-pass filtered spike trains,

$$
\begin{aligned}
\dot{z}_i^E(t) &= -\frac{1}{\tau_E^{\text{syn}}} z_i^E(t) + f_i^E(t) \\
\dot{r}_i^y(t) &= -\frac{1}{\tau_i^{r,y}} r_i^y(t) + f_i^y(t), \qquad y \in \{E, I\}.
\end{aligned}
\tag{28e}
$$

With eqs.28a-28e, we obtained a complete description of a biologically plausible spiking network model, where all elements obey constraints of biological neurons and networks and describe the set of membrane currents that are highly relevant for the function and dynamics of networks in cerebral cortex.

4 Alternative solutions for slow synaptic currents

While the recurrent excitatory synaptic currents suggested in eq. 23 seem the most biologically plausible solution, we here, for completeness, present alternative solutions for excitatory synaptic currents with slower kinetics. These alternative solutions are mathematically well defined and obey Dale's law, but are less likely to describe biological neural networks because they lack global balance of excitation and inhibition, and/or because they describe an E-I network without E-to-E connections. While the function of recurrent E-E connections in biological networks is still unclear, and in some instances, the probability of E-to-E connections in the local network can be very low (6), recurrent E-to-E connections are presumably still relevant for the dynamics of the cortical circuitry, and the lack thereof only gives an incomplete description of cortical networks.

We so far defined recurrent excitatory synapses (eq. 21) by substituting the signal $\boldsymbol{x}(t)$ with the excitatory estimate $\hat{\boldsymbol{x}}_E(t)$ (eq. 21), and justified the substitution by the fact that the loss function of E neurons minimizes the distance between these two variables (eq. 2a). Seen that the loss function of I neurons minimizes the distance between the E and the I estimates (eq. 2b), we can further assume the following: $\boldsymbol{x}(t) \approx \hat{\boldsymbol{x}}_E(t) \approx \hat{\boldsymbol{x}}_I(t)$. With this assumption, several alternative solutions emerge.

Let us first consider the solution that maintains slow recurrent excitation in E neurons and with that imposes positive semi-definiteness of the matrix B (as in eq. 21). In the membrane equation for the I cell type (eq. 12b), where we have the term $(\lambda_I - \lambda_E)W_I^\intercal B\hat{\boldsymbol{x}}_E(t)$, we now replace the E estimate with the I estimate, $\hat{\boldsymbol{x}}_E(t) \approx \hat{\boldsymbol{x}}_I(t)$. Using the definition of the I estimate (eq. 4c), we get the following solution for the slower component of synaptic currents:

$$
\begin{aligned}
&I_i^{EE}(t) \propto \sum_{\substack{j=1\\ j\neq i}}^{N_E} D_{ij}^{EE} z_j^E(t) \\
&I_i^{\text{slow}II}(t) \propto (\frac{1}{\tau_I} - \frac{1}{\tau_E}) \sum_{j=1}^{N_I} D_{ij}^{II} z_j^I(t), \quad \tau_I > \tau_E \\
&D_{ij}^{yy} = \begin{cases} (\boldsymbol{w}_i^y)^\intercal B \boldsymbol{w}_j^y, & \text{if } (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^y > 0, \quad \text{B positive semi-def.,} \\ 0 & \text{otherwise} \end{cases} \qquad \{yy\} \in \{EE, II\}.
\end{aligned}
\tag{29}
$$

The current $I_I^{\text{slow }II}(t)$ originates from I neurons and must therefore be inhibitory. To ensure the consistency of inhibitory connections, the membrane time constant of I neurons is slower than the membrane time constant of E neurons.

Moreover, slower E-to-E synaptic connections together with slower I-to-I synapses lead to a global imbalance of E-I currents. The network without slower synapses balances excitatory and inhibitory currents on its own. As we add slower E-to-E synapses, these bring additional excitation to the network that has to be counterbalanced by inhibition to maintain the global E-I balance. In the eq. 29, we instead have a slower inhibitory current in I neurons. Since E-to-E and I-to-I synaptic currents both promote the excitation at the network level, such a network is imbalanced and risks runaway excitation.

Two other alternative solutions describe networks without E-to-E connections. As we replace the signal $\boldsymbol{x}(t)$ in $W_E^\intercal B\boldsymbol{x}(t)$ with the estimate by I neurons, $\boldsymbol{x}(t) \approx \hat{\boldsymbol{x}}_I(t)$, this constrains the slow synaptic current in E neurons to originate from I neurons and the current in question is now constrained to be inhibitory. To ensure the synaptic current to E neurons to be inhibitory, the matrix B to has to be negative semi-definite. Assuming that the slow synaptic current in the I cell type is excitatory, the negative semi-definite matrix B now imposes the constant $(\lambda_I - \lambda_E)$ to be negative. This solutions reads as follows:

$$
\begin{aligned}
&I_i^{\text{slow}EI}(t) \propto \sum_{j=1}^{N_I} D_{ij}^{EI} r_j^I(t) \\
&I_i^{\text{slow}IE}(t) \propto (\frac{1}{\tau_I} - \frac{1}{\tau_E}) \sum_{j=1}^{N_E} D_{ij}^{IE} z_j^E(t), \quad \tau_I > \tau_E \\
&D_{ij}^{yz} = \begin{cases} (\boldsymbol{w}_i^y)^\intercal B \boldsymbol{w}_j^z, & \text{if } (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z > 0, \quad \text{B negative semi-def.,} \qquad \{yz\} \in \{EI, IE\} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
\tag{30}
$$

Constraint $\tau_I > \tau_E$ imposes that the membrane time constant of I neuron is longer than in E neurons. Moreover, such a network again risks a global imbalance of E and I currents. Slow inhibition in E neurons is accompanied by slow excitation in I neurons, and both currents globally promote inhibition. The latter solution seems of lesser biological relevance also because the network does not have E-to-E connections that are known to exist among excitatory neurons.

To prevent the imbalance of E and I currents, we can replace the excitatory estimate in $(\lambda_I - \lambda_E) W_I^\intercal B \hat{\boldsymbol{x}}_E(t)$ with the inhibitory estimate, $\hat{\boldsymbol{x}}_E(t) \approx \hat{\boldsymbol{x}}_I(t)$. This gives slow recurrent inhibition in I neurons, and the following set of solutions:

$$
\begin{aligned}
&I_i^{\text{slow}EI}(t) \propto \sum_{j=1}^{N_I} D_{ij}^{EI} r_j^I(t) \\
&I_i^{\text{slow}II}(t) \propto (\frac{1}{\tau_I} - \frac{1}{\tau_E}) \sum_{j=1}^{N_I} D_{ij}^{II} r_j^I(t), \quad \tau_I < \tau_E \\
&D_{ij}^{yz} = \begin{cases} (\boldsymbol{w}_i^y)^\intercal B \boldsymbol{w}_j^z, & \text{if } (\boldsymbol{w}_i^y)^\intercal \boldsymbol{w}_j^z > 0, \quad \text{B negative semi-def.,} \qquad \{yz\} \in \{EI, II\}. \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{31}
$$

In the solution as in eq. (31), we added an inhibitory current to E and to I neurons on top of a balanced network. Such a solution is expected to globally balance E and I currents in the network. However, the solution in eq. (31) leads to an incomplete description of cortical networks because the network lacks E-to-E connections.

5 Computational resources

Spiking network has been implemented with own computer code in Matlab, Mathworks, version 2021b. The complete computer code used to generate figures is part of the Supplementary material of the present paper. Integration of the membrane potential is done with Euler integration scheme. A network of 400 E and 100 I units is computed within seconds on a standard laptop.

References

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

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

[3] 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.

[4] 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.

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

[6] S. C. Seeman, L. Campagnola, P. A. Davoudian, A. Hoggarth, T. A. Hage, A. Bosma-Moody, C. A. Baker, J. H. Lee, S. Mihalas, C. Teeter, et al. Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *Elife*, 7:e37349, 2018.