Firing rates and representational error in efficient spiking networks are bounded by design

Matin Urdu[1], Gabriel Matías Lorenz[1][0009−0006−3129−906X], Ching-Peng Huang[1][0009−0007−9561−7431], Stefano Panzeri[1][0000−0003−1700−8909], and Veronika Koren[1][0000−0003−2920−2717]

University Medical Center Hamburg-Eppendorf, Hamburg, Germany
✉ vkoren@uke.de

Abstract. Recent studies have demonstrated that spiking neural networks designed under the principle of efficient coding process information efficiently with properties close to that of real networks. However, there is no analytical study of whether efficient spiking networks impose a finite bound on neural firing rates and on the representational error, and whether the objective of improving the neural representation with every spike is realized in biophysically realistic neural implementations. Here, we demonstrate that such networks are guaranteed to keep firing rates finite, and to achieve a finite representational error. Further, we show that in a model with optimal parameters, the vast majority of spikes improve the representation and carry the information about the stimulus. Thus, efficient spiking networks are well-defined mathematically and realize their objectives with biologically realistic spiking dynamics.

Keywords: Efficient coding · Spiking Networks · Recurrent Networks.

1 Introduction

Biological brains form neural representations of sensory stimuli [17]. Sensory systems in the brain extract information about time-dependent sensory features with receptor neurons (photoreceptors in the retina, mechanoreceptors in the skin, etc.), and then process this input information with recurrently connected network of spiking neurons to create neural signals that are useful for guiding animal behavior [19,10]. The principles that dictate these computations are currently unknown. An approach to address this question and gain insights about brain computations is to model information processing in biophysically realistic networks [1]. An influential hypothesis has been that neurons process the information *efficiently*, e.g., without wasting more resources than necessary to form neural representations that are useful in guiding the behavioral responses of animals [2]. This hypothesis has been quantitatively developed by formulating neural networks guided by a computational objective to encode natural images with sparse activity patterns [15]. This has been implemented in dynamical systems, firing rate models of recurrent neural networks [18,22,8], and finally also in recurrent spiking networks [3,14], with progressively more and more biologically plausible equations and features [7,9]. The computational objective in efficient spiking networks is expressed as a moment-by-moment minimization of a loss function with spikes.

However, several fundamental mathematical aspects of these models are still unaddressed. By design, efficient spiking networks explicitly guarantee an upper bound on the encoding error at any moment in time [6,16], but it has remained unclear if this bound is finite. This is because the encoding error depends on the firing rates of neurons, and it is yet unclear if the efficient coding theory guarantees finite firing rates. Very high firing rates and excessive across-neuron synchronization have indeed been reported in simulations of efficient spiking networks [11,5,20] that operate in absence of a saturating input-output function. A second question that has so far remained only partially addressed is how well the model objective (to generate spikes only when this decreases the loss function) translates on biophysical implementations with a spiking neural network.

Here, we first introduce the notion of a biologically plausible neural spike train and show that efficient spiking models have plausible spike trains and finite firing rates (in any finite time interval). We then show that finite firing rates imply finite encoding error. Finally, we show that a biophysical implementation of efficient coding with optimized parameters decreases the loss and the encoding error with the majority of its spikes and analyze numerically the dependency of efficient spiking on model and stimulus parameters.

2 Results

2.1 Biologically plausible spike train has no accumulation points

A spike train of a neuron is a point process defined as a sum of Dirac δ functions:

$$f(t) := \sum_n \delta(t - t_n), \quad t_n \in \mathcal{S}, \quad \forall \epsilon > 0 : \int_{t_n-\epsilon}^{t_n+\epsilon} \delta(t - t_n) = 1, \tag{1}$$

where $\mathcal{S}$ is a countable subset of $\mathbb{R}$. Every element of $\mathcal{S}$ is called a spiking time of the neuron, $\mathcal{S} = \{t_1 < t_2 < \cdots\} \subset \mathbb{R}_{t \geq 0}$. The spike train consists of a monotonically increasing sequence of spike times $(t_n)_{n \in \mathbb{N}}$.

For the subset $\mathcal{S}$, an *accumulation point* is a point $t_{acc} \in \mathbb{R}$ such that one can find a sequence $\{t_n | n \in \mathbb{N}\} \subset \mathcal{S}$ that converges to it, i.e. $\lim_{n\to\infty} t_n = t_{acc}$. A spike train is **biologically plausible** if $\mathcal{S}$ has no accumulation points.

An example of a spike train with an accumulation point is the sequence $\{t_n = 1 - \frac{1}{n} | n \in \mathbb{N}\}$, which has an accumulation point 1. Any (finite) subset of the form $\{t_n = 1 - \frac{1}{n} | n = 1, 2, \cdots, N, \ N < \infty\}$ has no accumulation point. For examples of an implausible spike train with an accumulation point and a plausible spike train generated by an efficient spiking neuron, see Fig. 1.

We note that in biophysical models of neural dynamics such as leaky integrate-and-fire models [4], spike times $\mathcal{S}$ also correspond to times when the membrane potential of the neuron, $u(t)$, reaches or crosses the firing threshold ϑ,

$$\mathcal{S}(t) \equiv \mathcal{S} = \left\{ t_n \in [t_0, t] \mid u\left(t_n^-\right) \geq \vartheta \right\}, \tag{2}$$

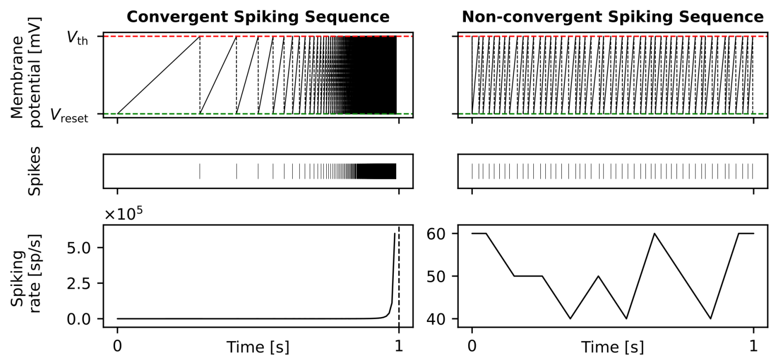

Fig. 1. Examples of a convergent and a non-convergent sequence of spike times. Top: The membrane potential (black) with the reset potential (green) and the firing threshold (red) of a neuron generating a sequence of spike times. Middle: The spike train. The neuron on the left generates an implausible spike train, because its sequence of spike times converges at time $t = 1$. The neuron on the right instead generates a biologically plausible spiking sequence that has no accumulation points. Bottom: The time-averaged firing rate. The firing rate of the neuron on the left diverges (to $+\infty$). The neuron on the right, on the contrary, has a finite firing rate.

with t_n^- the left-side limit of the spike time at t_n. The goal of the efficient spiking model is to determine spike times of neurons such that these spike times enable the neural network to encode neural signals efficiently, i.e., accurately and with the minimal number of spikes.

The spike train is transformed into a *low-pass filtered spike train*, defined by the inhomogeneous linear ordinary differential equation (ODE):

$$\dot{r}(t) := -\frac{1}{\tau_r} r(t) + f(t), \qquad r(t_0) = 0, \tag{3}$$

for a constant $\tau_r > 0$. As we begin observing the neural system at $t = t_0$, the set $\mathcal{S}$ is initially empty and then updated each time the neuron spikes. Without loss of generality, we assume that at $t = t_0$, the neuron is silent. Using the initial condition $r(t_0) = 0$, the specific solution for the low-pass filter is:

$$r(t) = \sum_{t_n \in \mathcal{S}} \exp\left(-\frac{t - t_n}{\tau_r}\right), \qquad t \geq 0. \tag{4}$$

We now consider a neuronal ensemble composed of N_E excitatory (E) neurons and N_I inhibitory (I) neurons. Each neuron belongs to exactly one ensemble (population), which we refer to with $y \in \{E, I\}$. For each neuron $i = 1, \cdots, N_y$, we denote their spike trains and low-pass filters as $\mathcal{S}_i^y$ and $r_i^y(t)$. We define the *linear population readout* $\hat{\boldsymbol{x}}_y(t) \in \mathbb{R}^M$ of the spiking activity for each population $y \in \{E, I\}$ as a linear combination of the low-pass filtered spike trains:

$$\dot{\hat{\boldsymbol{x}}}_y(t) := -\frac{1}{\tau_y} \hat{\boldsymbol{x}}_y(t) + W_Y \boldsymbol{f}_y(t), \qquad \hat{\boldsymbol{x}}_y(t_0) = 0 \;\forall m = 1, \cdots, M, \tag{5}$$

with time constant $\tau_y > 0$ and the matrix of decoding weights $W_y = [w^y_{mi}]_{m=1,\ldots,M,i=1,\ldots,N_y}$. The Eq. 5 defines how neural signals are decoded from neural spiking activity. These population readouts take the role of estimators of the target signals.

The population readout of population y from the Eq. 5 for a specific output dimension m is an inhomogeneous linear ODE:

$$\dot{\hat{x}}^y_m(t) := -\frac{1}{\tau_y}\hat{x}^y_m(t) + \sum_{i=1}^{N_y} w^y_{mi} f^y_i(t), \qquad \hat{x}^E_m(t_0) = 0. \tag{6}$$

The solution of Eq. 6 for $t > 0$ is:

$$\hat{x}^y_m(t) = \sum_{i=1}^{N_y} \sum_{t^{i,y}_n \in \mathcal{S}^y_i} w^y_{mi} \exp\left(-\frac{t - t^{i,y}_n}{\tau_y}\right). \tag{7}$$

Note that solutions to the differential equations are always interpreted as distributional solutions to the integral equations.

Previous work [3,13] proposed that neural activity is such that population readouts $\hat{\boldsymbol{x}}_y(t)$ approximate time-dependent *target representation* $\boldsymbol{x}(t) \in \mathbb{R}^M$ with the same dimension while using the minimal number of spikes. This can be achieved by using the following *instantaneous loss function*:

$$\begin{aligned} \mathrm{L}_E(t; \boldsymbol{f}_E) &= \sum_{m=1}^{M} \left(x_m(t) - \hat{x}^E_m(t)\right)^2 + \mu_E \sum_{i=1}^{N_E} \left(r^E_i(t)\right)^2 \\ \mathrm{L}_I(t; \boldsymbol{f}_I) &= \sum_{m=1}^{M} \left(\hat{x}^E_m(t) - \hat{x}^I_m(t)\right)^2 + \mu_I \sum_{i=1}^{N_I} \left(r^I_i(t)\right)^2 \end{aligned} \tag{8}$$

for metabolic constants $\mu_E, \mu_I > 0$. The first terms describe the objective of the readout to approximate the target representation, while the second terms, weighted by the metabolic constants, penalize the spiking activity to ensure efficiency.

We now model the firing rule as a minimization of the instantaneous loss and a noise term with each spike. That is, we assume a spike in neuron i of type $y \in \{E, I\}$ is fired at time $t^{i,y}_n$ if and only if this spike decreases the loss of the neuron's population.

Definition 1 (Condition for neuronal firing). *Let $t^{i,y}_n$ be a firing time for neuron $i \in \{1, \ldots, N_y\}$, $y \in \{E, I\}$ if for $\hat{\boldsymbol{f}}_y(t) := \boldsymbol{f}_y(t) + [0, \ldots, \delta(t - t^{i,y}_n), \ldots 0] \in \mathbb{R}^{N_y}$, the following is true:*

$$\mathrm{L}_y(t^+; \hat{\boldsymbol{f}}_y) < \mathrm{L}_y(t^-; \boldsymbol{f}_y) + \xi^y_i(t^-), \tag{9}$$

with t^- and t^+ the left- and right-side limit of the spike time and $\xi^y_i(t)$ a continuous noise process. If the condition in Eq. 9 is true, neuron i of type y fired at time $t^{i,y}_n$ and we register a spike, e.g., $t^{i,y}_n \in \mathcal{S}^y_i$. The continuity of the noise is sufficient for the model to generate plausible spike trains.

The optimization goal of efficient coding is to generate spikes at a precise time when they minimize the loss through the Eq. 9. The loss efficiently estimates a time-dependent target signal $\boldsymbol{x}(t)$ by tracking the target signal with the population readouts $\hat{\boldsymbol{x}}_E(t)$ (E neurons) and $\hat{\boldsymbol{x}}_I(t)$ (I neurons).

By analytically developing the firing rule in the Eq. 9, we express the rule for spike generation on an algorithmic level. We define the time-dependent encoding error for each neuron i from the population E and I as:

$$\epsilon_i^E(t) := \sum_{m=1}^{M} w_{mi}^E \left(x_m(t) - \hat{x}_m^E(t)\right), \quad \epsilon_i^I(t) := \sum_{m=1}^{M} w_{mi}^I \left(\hat{x}_m^E(t) - \hat{x}_m^I(t)\right). \quad (10)$$

Theorem 1. *With the above setups, if $t_n^{i,y} \in S_i^y$ is a spiking time, and t^- is its left-side limit, then*

$$\epsilon_i^y(t^-) - \mu_y r_i^y(t^-) + \frac{1}{2}\xi_i^y(t^-) \geq \frac{1}{2}(||\boldsymbol{w}_i^y||^2 + \mu_y), \quad (11)$$

where $\boldsymbol{w_i^y}$ is the i-th row of W_y, i.e. the decoding vector for neuron i of type y and $||\boldsymbol{w}_i^y||^2$ is the squared length of the decoding vector, $||\boldsymbol{w}_i^y||^2 = \sum_m^M (w_{mi}^y)^2$.

For the proof, see [3,13,12].

We define the left-hand side of Eq. 11 as the membrane potential $u_i^y(t)$ and the right-hand side as the firing threshold ϑ_i^y,

$$u_i^y(t) := \epsilon_i^y(t) - \mu_y r_i^y(t) + \frac{1}{2}\xi_i^y(t), \qquad \vartheta_i^y := \frac{1}{2}(||\boldsymbol{w}_i^y||^2 + \mu_y). \quad (12)$$

We now state the firing rule in the Eq. 9 in terms of the inequalities in the Eq. 11. Neuron i fires a spike at time $t_n^{i,y}$ whenever its membrane potential reaches or crosses the firing threshold, $u_i^y(t^-) \geq \vartheta_i^y$ which corresponds to the condition for firing a spike in biological neurons. Up to noise, a spike from neuron i always lowers its potential to the same reset value, $u_i^y(t^+) = u_{reset,i}^y + \xi_i^y(t)$.

We now prove that the firing rate of each neuron is finite in any finite time interval, implying that the spike trains generated by the efficient coding model are biologically plausible and have no accumulation points. To prove the finiteness of the firing rate, we show that for any continuous and bounded target signal $\boldsymbol{x}(t)$ and non-negative weights w_{mi}^y, the low-pass filter $r_i^y(t)$ (Eq. 4) and the population readout $\hat{\boldsymbol{x}}_y(t)$ (Eq. 7) are necessarily finite. This also guarantees the existence of the solution for the membrane potential $u_i^y(t)$ for $t \in [t_0, \infty)$.

Specifically, we are interested in the problem of finding the largest $b \in [t_0, \infty]$ for which the solution in Eq. 12 exists on the interval $[t_0, b)$. When $b \in [t_0, \infty)$, we say that the solution exists locally on $[t_0, b)$ and when $b = \infty$, we say that the solution exists globally on $[t_0, \infty)$. We demonstrate that the only way to have the firing rates diverge is to have the sequence of spike times converge to a finite limit. Finally, we prove by contradiction that convergence of spike times to a finite limit is not possible.

We consider the membrane potential of excitatory neurons $u_i^E(t)$ in the deterministic case of the definition in the Eq. 12, $u_i^E(t) := \epsilon_i^E(t) - \mu_E r_i^E(t)$, with the additional assumption that all decoding weights in the matrix W_E are non-negative, $w^E_{mi} \geq 0 \; \forall m, i$. We also assume that the firing times are unique for each neuron. By registering a spike every time the membrane potential $u^E_i(t)$ reaches the firing threshold in Eq. 12, we fill the monotone sequence of spike times $\mathcal{S}^E_i(t)$ over time.

Because spike times can only increase over time, the spiking sequence $\mathcal{S}^E_i(t)$ is monotonically increasing (see Lemma 1). As the sequence is monotonically increasing, it either converges or diverges (to $+\infty$). If the sequence of spike times diverges, then the solution exists on all of $[t_0, +\infty)$. On the contrary, if the sequence of spike times converges to a real number $t' \in (t, +\infty), t > t_0$, then the solution is only defined in the time interval $[t_0, t')$, and no global solution exists.

To illustrate a hypothetical situation where no global solution exists, we define a hypothetical spiking sequence $(t_n)_{n \in \mathbb{N}}$, with $t_n := 1 - \frac{1}{\log_2(n+1)}$. Such sequence of spike times converges as the time approaches 1, i.e. $n \to \infty$, and $t_n \uparrow t' = 1$. Convergence of the spiking sequence to a finite limit causes a divergence of the firing rate of the neuron (Figure 1) and consequently a convergence of the inter-spike interval to 0. In this case, no solution exists at 1, because we cannot find a neighborhood around $t' = 1$ at which the membrane potential $u^E_i(t)$ is continuous. In this hypothetical case, it is therefore not possible to find a global solution for the membrane potential $u^E_i(t)$, because any such solution would not be differentiable around any open neighborhood of $t' = 1$. The divergence of the spiking sequence also implies the divergence of the low-pass filter $r^E_i(t)$ and of the population readout $\hat{\boldsymbol{x}}^E(t)$, because these quantities are proportional to spiking frequency of neurons.

We now prove that if $\boldsymbol{x}(t)$ is continuous or bounded and the decoding weights w^y_{mi} are non-negative, then a solution exists globally on $[t_0, \infty)$. To prove this, we use a lemma showing that if the sequence of spike times $\mathcal{S}^E_i(t)$ converges to a finite limit t', the neuron fires infinitely many spikes, causing a divergence of the low-pass filtered spike train and of the estimate, e.g., $r^E_i(t), \hat{x}^E_m(t) \to \infty$. We then show that divergence of these variables is in contradiction with our model of threshold crossing, and it therefore cannot happen.

Lemma 1. *Let $(t_n)_{n \in \mathbb{N}}$ be a monotonically increasing sequence. Then the function*

$$g(t) = \sum_{k=1}^{\sup\{n \in \mathbb{N} | t_n \leq t\}} \exp(-\lambda(t - t_k)), \quad \lambda > 0.$$

is well-defined on all of $\mathbb{R}$ if and only if $(t_n)_{n \in \mathbb{N}}$ diverges to ∞.

Proof. Let us first assume that we have $t_n \uparrow \infty$ for $n \to \infty$. Then for all $t \in \mathbb{R}$ the set $\sup\{n \in \mathbb{N} \mid t_n \leq t\}$ is finite and $g(\cdot)$, as a finite sum of well-defined functions, is also well-defined. Further, we assume that the sequence does not diverge and converges to a limit $t' \in \mathbb{R}$. Then, because $\sup\{n \in \mathbb{N} \mid t_n \leq t'\}$ contains (countably) infinite amount of elements and g is not real-valued at t':

$$g(t') = \lim_{n \to \infty} \sum_{k=1}^{n} \exp\left(-\lambda(t' - t_k)\right) = \exp(-\lambda t') \cdot \lim_{n \to \infty} \sum_{k=1}^{n} \exp(\lambda t_k) = \infty.$$

Thus, because $g(t')$ is not real-valued, g is not well-defined on all $\mathbb{R}$.

If $(t_n)_{n\in\mathbb{N}}$ is a sequence of spike times, the function $g(n)$ corresponds to the estimate $\hat{x}_m^y(t)$, in the sense that $\forall n \in \mathbb{N} : g(n) \propto \hat{x}_m^y(t_n)$ and $g(n) \propto r_i^y(t_n)$. We have proven that, if there is an infinite sequence of convergent spikes $(t_n)_{n\in\mathbb{N}}$, then the estimate and the low-pass filtered spike train also diverge to ∞.

Theorem 2 (Existence of a solution). *For a continuous target signal $\boldsymbol{x}(t)$ and non-negative decoding matrices W_E, as $t \uparrow \infty$, the sequence of excitatory spike times $\left(t_n^{i,E}\right)_{n\in\mathbb{N}} \equiv (t_n)_{n\in\mathbb{N}}$ of each excitatory neuron diverges to ∞.*

Proof. We note that the excitatory estimate $\hat{x}_m^E(t)$ and the low-pass filtered spike train $r_i^E(t)$ are proportional to a continuous function $g(n)$ in Lemma 1. Now let $\vartheta \in \mathbb{R}$ be the firing threshold and we assume that we have a monotonically increasing convergent sequence of spike times $(t_n)_{n\in\mathbb{N}}$ with a limit $t' \in (t_0, \infty)$. Then the membrane potential $u_i^E(t)$ has to repeatedly cross the firing threshold ϑ as $t \uparrow t'$. In particular, we cannot have the following:

$$\lim_{t\uparrow t'} u_i^E(t) = \lim_{t\uparrow t'} x_m(t) - \hat{x}_m^E(t) - r_i^E(t) = -\infty,$$

because in this case $u_i^E(t)$ does not repeatedly cross the firing threshold ϑ. As $x_m(t)$ is a continuous function on all of $[t_0, \infty)$, it is also bounded on every closed or open interval of the domain. Combining this argument with the fact that all weights are non-negative, we conclude that the only way we obtain a divergence to $-\infty$ is if we have

$$\lim_{t\uparrow t'} -\hat{x}_m^E(t) = -\infty \quad \textbf{or} \quad \lim_{t\uparrow t'} -r_i^E(t) = -\infty.$$

However, both cases cannot happen (otherwise we have $u_i^E(t) \uparrow -\infty$ for $t \uparrow t'$). Therefore, for any non-negative constant weight vector $\boldsymbol{w}_i^E$, we obtain:

$$\lim_{t\uparrow t'} u_i^E(t) \equiv \lim_{t\uparrow t'} \left(\boldsymbol{w}_i^E\right)^\top (\boldsymbol{x}(t) - \hat{\boldsymbol{x}}_E(t)) - \mu_E r_i^E(t) > -\infty,$$

which by Lemma 1 implies that the spike train diverges to ∞.

Theorem 2 also proves that the individual components $\hat{x}_m^E, m \in \{1, \ldots, M\}$ of the population readout $\hat{\boldsymbol{x}}_E(t)$ are well-defined continuous functions in the open interval between two excitatory spike times, and are left-continuous at the spike times. While the Theorem 2 proved the existence of a solution for E neurons, same follows for I neurons. In the case of I neurons, the excitatory population readout $\hat{\boldsymbol{x}}_E(t)$ takes the role of the target, while the inhibitory population readout takes the role of the estimate (see Eq. 8).

By showing the divergence of the spiking sequence of all neurons, we proved that firing rates for all neurons in an efficient spiking model are finite. Theorem 2 also guarantees the existence of a global solution for any efficient spiking neural network with a given initial membrane potential at t_0. The existence of local solution $u_{i,\text{loc}}^y : [t_0, b) \to \mathbb{R}$ for any $b \in (t_0, \infty)$, implies the existence of a global solution on $u_i^y : [b, \infty) \to \mathbb{R}$ with $u_i^y|_{[t_0,b)} \equiv u_{i,\text{loc}}^y$ and $u_i^y(b) = \lim_{t\uparrow b} u_{i,\text{loc}}^y(t)$. If no neuron fires at time b, then we also have $u_i^y(b) = \lim_{t\to b} u_{i,\text{loc}}^y(t)$.

2.2 Efficient spiking models have finite encoding error

An important consequence of Theorems 1 and 2 is that the encoding error manifested by each spiking neuron is always finite.

Corollary 1 (Finite encoding error of single neurons). *The encoding error of an efficient spiking neuron i of type $y \in \{E, I\}$ accumulated over any finite time interval $[a, b]$, $t_0 < a < b < \infty$, is bounded and the bound is finite:*

$$\int_a^b \epsilon_i^y(t)\; dt \leq \int_a^b \left[\vartheta_i^y + \mu_y r_i^y(t) - \frac{1}{2}\xi_i^y(t) \right] dt, \qquad y \in \{E, I\}. \tag{13}$$

Proof. Inserting the definition of the firing threshold and using the definition of the time-dependent encoding error in Eq. 10, we rewrite the Eq. 13 as:

$$\int_a^b \epsilon_i^E(t)\; dt < \frac{1}{2} \left(\mu_E + \sum_{m=1}^{M} (w_{mi}^E)^2 \right) (b-a) - \frac{1}{2} \int_a^b \xi_i^E(t)\; dt + \mu_E \int_a^b r_i^E(t)\; dt.$$

With a finite metabolic parameter μ_E, finite noise intensity and finite decoding weights w_{mi}^y for all neurons, as expected in any biologically relevant case, and considering that the low-pass filter $r_i^y(t)$ cannot diverge in a limited time interval due to the Theorem 2, we have a guarantee that the integral of the encoding error in any finite time interval is finite.

Further, we consider the *instantaneous encoding error of the population*, defined as: $E^y(t) := \sum_{i=1}^{N_y} \epsilon_i^y(t)$. The instantaneous error of the population is bounded by spiking of individual neurons. The neuron with the largest encoding error at a particular time, i.e., neuron i for which it holds: $\arg\max_i \epsilon_i^y(t)$, is also the neuron closest to the firing threshold. As soon as the neuron with the largest encoding error reaches its firing threshold, its firing decreases the loss of neuron's population because of the firing rule in Eq. 9, and another neuron now has the largest encoding error. In case no two neurons fire at exactly the same time, the instantaneous encoding error of the population $E^y(t)$ is a sum of finite elements and therefore finite. This gives the following:

Corollary 2 (Finite instantaneous encoding error of the population). *Following previous definitions of an efficient spiking network with finite network size and assuming no neurons spike at exactly the same time, the encoding error of the population, $E^y(t), y \in \{E, I\}$, is finite at all times.*

Proof follows immediately from the Corollary 1. From here, it follows that the integral of the encoding error of the population accumulated over any finite time interval is also finite.

Corollary 3 (Finite accumulated encoding error of the population). *Following previous definitions of an efficient spiking network, the encoding error of the population, $y \in \{E, I\}$, accumulated over any finite time interval $[a, b]$, $t_0 < a < b < \infty$, is finite.*

2.3 Biophysical implementation of efficient spiking decreases the encoding error and loss with the majority of fired spikes

We finally used a numerical simulation of a biophysical implementation of efficient coding (code at zenodo.org/records/15722416) to verify if neurons indeed spike to decrease the loss, as imposed by the Eq. 8. We used the time-derivative of Eq. 12 developed in Refs. [3,13], which gives a biophysical neural model for the case where the target signal is defined as a leaky integrator of stimulus features:

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{dt} = -\lambda \boldsymbol{x}(t) + \boldsymbol{s}(t). \tag{14}$$

The stimulus $\boldsymbol{s}(t) \in \mathbb{R}^M$ is defined as a set of M independent Ornstein-Uhlenbeck (OU) processes,

$$\tau_s \frac{\mathrm{d}s_m(t)}{dt} = -s_m(t) + \sqrt{2\tau_s}\sigma_s \eta_m(t), \tag{15}$$

with standard deviation $\sigma_s = 2$ (mV)$^{1/2}$, correlation time $\tau_s = 10$ milliseconds (ms) and with $\eta_m(t)$ a white noise. It assumes encoding of a continuous and bounded set of stimuli, such as those filtered and passed to the brain by a set of peripheral receptors.

Following the definition of model's objectives in Eq. 8, we measured the time-dependent Squared error between targets and estimates:

$$\epsilon^E(t) = \sum_{m=1}^{M} \left(x_m(t) - \hat{x}_m^E(t)\right)^2, \qquad \epsilon^I(t) = \sum_{m=1}^{M} \left(\hat{x}_m^E(t) - \hat{x}_m^I(t)\right)^2. \tag{16}$$

and the time-dependent Metabolic cost, $\kappa^y(t) = \sum_{i=1}^{N_y} \left(r_i^y(t)\right)^2$ of population $y \in \{E, I\}$. Finally, similarly as in a previous work [12], we evaluated a weighted sum of the Squared error and the Metabolic cost: $Loss^y(t) = g\epsilon^y(t) + (1-g)\kappa^y(t)$ with $0 < g < 1$, a quantity that we refer to as the (empirical) Loss. As shown in [13], the biophysical implementation of the network follows the equations of a generalized leaky integrate-and-fire spiking network with recurrently connected E and I neurons. We used the parameter set that optimizes the time-average of the weighted sum of the Squared error and Metabolic cost reported in [12].

To evaluate the extent to which the network actually decreases the Loss with every spike, we evaluated the performance metrics (the Squared error, the Metabolic cost and the Loss) around spike times. Spikes led almost always to negative jumps of the Squared error and of the Loss (Fig. 2A top and bottom) and always to positive jumps of the Metabolic cost (Fig. 2A middle). We quantified this by computing a spike-triggered average for each performance metrics. We found that spikes reliably trigger a negative jump of the Squared error and of the Loss for both cell types (Fig. 2B, left and right), and that 92 (85) % of spikes in E (I) neurons decreased the Loss, while 93 (85) % of spikes of E (I) neurons decreased the Squared error. Thus, the vast majority of spikes are both efficient and error-correcting. Whenever a spike decreases the representational error, it carries information about the target signal $\boldsymbol{x}(t)$ and thus about the

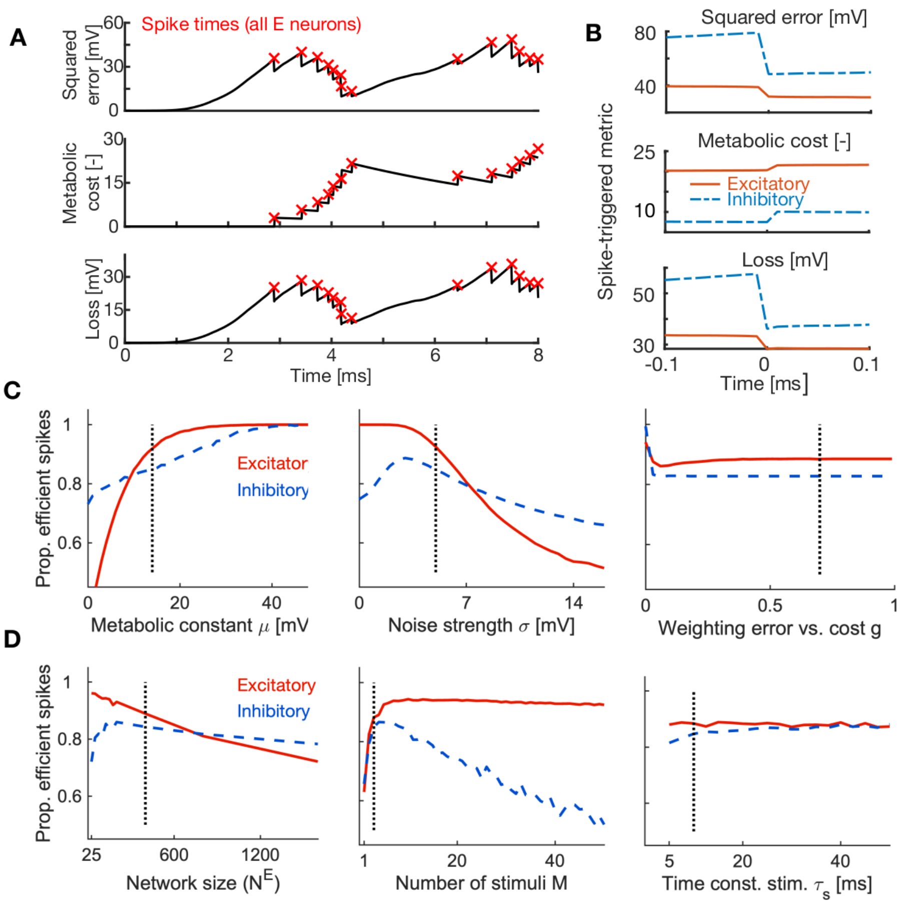

Fig. 2. Biophysical model decreases the error and the loss with most spikes **(A)** Close-up view of Squared error (top), Metabolic cost (middle), and Loss (bottom) of the E population. Spike times of E neurons (marked with red crosses) typically trigger a negative jump of the Squared error and the Loss, and always trigger a positive jump of the Metabolic cost. **(B)** Spike-triggered Squared error (top), Metabolic cost (middle), and Loss (bottom). Each metric is shown for E (solid red) and I (dashed blue) population. **(C)** Left: Proportion of efficient spikes (decreasing the Loss) in E (full red) and I (dashed blue) population as a function of the metabolic constant $\mu = \mu^E = \mu^I$ (default value marked with dotted black vertical line). Middle: Same as on the left, as function of the noise intensity σ for the uncorrelated Gaussian noise in the membrane potential of neurons. Right: Same as on the left, as a function of weighting of the Encoding error and the Metabolic cost g. **(D)** Same as in (C), varying the number of E neurons N^E (and with constant ratio of the number of E to I neurons at 4:1; left), the number of stimuli M (middle) and the stimulus time constant τ_s (right). Trial duration was 1 second (A,B) or 50 seconds (C,D) with timestep of 0.01 ms. The Encoding error vs Metabolic cost weighting was $g = 0.7$. Other parameters were as in [12].

stimulus $\boldsymbol{s}(t)$. Finally, we analyzed the dependence of efficient spiking on model parameters and stimulus statistics. The proportion of efficient E and I spikes increased with the metabolic constant μ. It peaked at intermediate noise levels in I neurons and was large for low to intermediate noise levels for E neurons. It did not change with g and with the stimulus time constant (Fig. 2C-D). Spiking of E neurons was more efficient at small network sizes and equally efficient over a range of the number of encoded stimuli, whereas spiking of I neurons was equally efficient over a range of network sizes but was more efficient when encoding a handful of stimuli (Fig. 2D). In sum, the spiking is highly efficient over a wide range of parameters and particularly for networks of moderate size encoding a handful of stimuli in the presence of noise.

3 Discussion

We proved, under minimal assumptions, that neural firing rates and the encoding error in efficient spiking networks are guaranteed to be finite. By simulating the efficient spiking model, we showed that in such networks the encoding error and the weighted sum of the encoding error and the metabolic cost decrease with the majority of spikes fired by the network, as prescribed by the efficient firing rule. Thus, most spikes convey stimulus information. These results are significant for computational neuroscience. They show that efficient spiking networks that were previously reported to describe well properties of real cortical networks [12] (including realistic spiking statistics, spike-triggered adaptation, and E-I balance), are well-defined mathematically, operate within biological plausible regimes and realize their objectives well. This is essential to support their biological plausibility and their use to understand if neural properties are driven by optimality principles. Because they reach energy efficiency for small network sizes, and because their analytical solutions provide explicitly and readily implementable efficient E-I connectivity rules, efficient spiking models are suitable for application in neuromorphic circuits. Architectures similar to that arising from optimal solutions of the efficient coding model studied here have been implemented in neuromoprhic systems to produce sparse and decorrelated activity [21]. Our results suggest that such architecture can operate robustly without incurring into pathological firing of in-silico neurons.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abbott, L.F., DePasquale, B., Memmesheimer, R.M.: Building functional networks of spiking model neurons. Nature Neuroscience **19**(3), 350–355 (2016)
2. Barlow, H.B., et al.: Possible principles underlying the transformation of sensory messages. Sensory communication **1**(1), 217–233 (1961)

3. Boerlin, M., Machens, C.K., Denève, S.: Predictive coding of dynamical variables in balanced spiking networks. PLoS Computational Biology **9**(11), e1003258 (2013)
4. Brunel, N.: Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. Journal of Computational Neuroscience **8**(3), 183–208 (2000)
5. Buxó, C.E.R., Pillow, J.W.: Poisson balanced spiking networks. PLoS Computational Biology **16**(11), e1008261 (2020)
6. Calaim, N., Dehmelt, F.A., Gonçalves, P.J., Machens, C.K.: The geometry of robustness in spiking neural networks. ELife **11**, e73276 (2022)
7. Denève, S., Machens, C.K.: Efficient codes and balanced networks. Nature Neuroscience **19**(3), 375–382 (2016)
8. Engelken, R., Goedeke, S.: A time-resolved theory of information encoding in recurrent neural networks. Advances in Neural Information Processing Systems **35**, 35490–35503 (2022)
9. Gutierrez, G.J., Denève, S.: Population adaptation in efficient balanced networks. ELife **8**, e46926 (2019)
10. Koren, V., Bondanelli, G., Panzeri, S.: Computational methods to study information processing in neural circuits. Computational and Structural Biotechnology Journal **21**, 910–922 (2023)
11. Koren, V., Denève, S.: Computational account of spontaneous activity as a signature of predictive coding. PLOS Computational Biology **13**(1), e1005355 (2017)
12. Koren, V., Malerba, S.B., Schwalger, T., Panzeri, S.: Efficient coding in biophysically realistic excitatory-inhibitory spiking networks. eLife **13**, RP99545 (2025)
13. Koren, V., Panzeri, S.: Biologically plausible solutions for spiking networks with efficient coding. Advances in Neural Information Processing Systems **35**, 20607–20620 (2022)
14. Moreno-Bote, R., Drugowitsch, J.: Causal inference and explaining away in a spiking network. Scientific Reports **5**(1), 17531 (2015)
15. Olshausen, B.A., Field, D.J.: Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Research **37**(23), 3311–3325 (1997)
16. Podlaski, W.F., Machens, C.K.: Approximating nonlinear functions with latent boundaries in low-rank excitatory-inhibitory spiking networks. Neural Computation **36**(5), 803–857 (2024)
17. Quian Quiroga, R., Panzeri, S.: Extracting information from neuronal populations: information theory and decoding approaches. Nature Reviews Neuroscience **10**(3), 173–185 (2009)
18. Rozell, C.J., Johnson, D.H., Baraniuk, R.G., Olshausen, B.A.: Sparse coding via thresholding and local competition in neural circuits. Neural computation **20**(10), 2526–2563 (2008)
19. Saxena, S., Cunningham, J.P.: Towards the neural population doctrine. Current Opinion in Neurobiology **55**, 103–111 (2019)
20. Timcheck, J., Kadmon, J., Boahen, K.: Optimal noise level for coding with tightly balanced networks of spiking neurons in the presence of transmission delays. PLOS Computational Biology **18**(4), e1009534 (2022)
21. Zendrikov, D., Solinas, S., Indiveri, G.: Brain-inspired methods for achieving robust computation in heterogeneous mixed-signal neuromorphic processing systems. Neuromorphic Computing and Engineering **3**(3), 034002 (2023)
22. Zhu, M., Rozell, C.J.: Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. PLOS Computational Biology **9**(8), e1003191 (2013)