

Interactions of optimal balance and recurrent plasticity for compensating neuronal death

Machens lab rotation report

DENNIS GOLDSCHMIDT

Champalimaud Neuroscience Programme,
dennis.goldschmidt@neuro.fchampalimaud.org

October 24, 2016

Abstract

Information processing in the brain is surprisingly robust, even when facing neuronal death due to damaging environmental influences, neuronal aging, or neurodegenerative diseases. Many studies have shown that the robustness is due to adaptive molecular mechanisms and synaptic plasticity, which reorganize network connectivity. However, it is also hypothesized that neurons are able to self-regulate their firing properties to compensate for such perturbations. The theoretical framework of optimal balanced networks recently proposed a mechanism for such neuronal self-regulation based on quadratic optimization. This report describes simulations of optimal balanced networks to examine the interactions of firing rate compensation and recurrent plasticity during neuronal death. We find that after the rapid compensation by locally increased firing rates, recurrent plasticity can further recover the network's balance by decreasing the variance of the population rate. This leads to a lower coding error, especially for severe loss of more than 60% of the population.

I. INTRODUCTION

The brain encodes and processes sensory stimuli in order to extract relevant information from the environment to act upon. Computational neuroscience is interested in finding out about the computational and algorithmic implementations of sensory information processing in the brain [1]. Given that neurons transmit action potentials, it is apparent to ask how information might be represented in such a spike code. Early seminal works by Adrian [2] have shown that the firing rate of single neurons conveys information about the amplitude of a stimulus. However, the world contains an enormous amount of information continuously sensed by the sensory afferents and it also possesses a certain amount of uncertainty not only in stimuli, but also in actions generated by the brain. Nonetheless, brains are capable of processing this uncertain information and make appropriate decisions [3–5].

Natural stimuli are spatially and temporally correlated, and therefore are statistically redund-

dant. The *efficient coding* hypothesis states that brains have evolved to perform information transfer to reduce redundancy efficiently in the presence of noise [6, 7]. This idea stems from the concept of information theory [8], which presents the necessary basis for efficiently coding and compressing various types of information such as text, images, and sound. In vision, efficient coding has successfully explained receptive fields properties of primary visual cortex derived from natural image statistics [9, 10], and evidence for efficient coding has also been found in lower visual systems like the retina in vertebrates [11, 12] and invertebrates [13, 14]. Besides vision, auditory [15] and somatosensory systems [16] have been shown to also exhibit efficient codes.

How is efficient coding implemented in a network of neurons? While even single neurons can efficiently adapt their firing rate properties to match stimuli statistics [13, 17], encoding information in a populations of neurons is useful for redundancy reduction in several ways. First, using joint activities of many neurons may ef-

fectively decorrelate the sensory input. Second, information coding in populations allows for synergy, in which the population activity carries more information than the mere sum of its individual neurons [18]. Finally, it can be shown that neural populations can utilize noise correlations in some cases to reduce sensory uncertainty [19, 20]. Tuning curves of individual neurons are typically given by a bell-shaped function with respect to a stimulus, while maxima of these tuning curves in the population spread across the stimulus domain. Prominent examples are the orientation tuning in visual cortex [21], and movement direction tuning in motor cortex [22].

Many studies on cortical spiking activity have noticed that, while *in-vitro* spike trains of isolated neurons are reliable and regular for same stimuli [23], *in-vivo* network spike trains are significantly variable and irregular across trials with same stimuli (most notably, [24]). For efficient coding, this seems to be a problem, since noise or variability often leads to inaccurate codes. *Balanced networks* have been shown to provide a mechanistic explanation for asynchronous and irregular spike trains in cortical networks [25, 26]. By constructing a specific recurrent connectivity, each neuron will on average receive a same amount of excitatory and inhibitory postsynaptic potentials (i.e., synaptic inputs are balanced). Besides forming this balance, such networks exhibit suprathreshold or chaotic dynamics which allows for enhanced information processing and coding. Therefore, asynchronous and irregular spike trains do not necessarily deteriorate coding performance, because the activity across the population is reliable and describes computationally powerful dynamics.

The framework of optimal balanced networks was developed by Denève and Machens [27, 28] in an effort to unify the concepts of *efficient coding* and *balanced networks*. The model is derived from two assumptions: a linear readout and a spiking mechanism that minimizes a well-defined loss function. As such, these networks are able to perform coding and computation of continuous variables [28], and account

for tuning properties of various biological systems, such as the goldfish oculomotor system [29]. Additionally, optimal balanced networks have shown to compensate for neuronal death in a robust way [30]. Up to a certain recovery boundary, compensation of eliminated neurons has been achieved by adjusting the firing rates of the remaining neurons of the network. However, beyond this recovery boundary the network is not able to perform properly. In this report, we will combine this acute compensation with learning rules previously developed for optimal balanced networks [31], in order to test whether this can further improve compensation for neuronal death on longer time scales.

In the following section, we will derive the dynamical equations for optimal balanced networks and formulate learning equations for adapting recurrent connections optimally. Section III shows results from simulated networks with learning for compensating neuronal death. Finally, section IV discusses the findings of this report.

II. DERIVATIONS

The goal of this section is to derive a framework for efficient coding by a neural network. Efficient coding states that the network is optimized to represent a time-varying signal $\mathbf{x}(t)$ efficiently. This signal $\mathbf{x}(t) \in \mathbb{R}^M$ is filtered by a linear dynamical equation

$$\tau \dot{\mathbf{x}}(t) = -\mathbf{x}(t) + \mathbf{c}(t) \quad (1)$$

of a dynamic variable $\mathbf{c}(t)$ with a time constant τ . Note that $M \gg 1$ is exhibited for many real-life sensory systems, such as vision.

i. Assumptions

The framework of optimal balanced networks [27, 28] is derived from two assumptions.

First, the representation $\mathbf{y}(t) \in \mathbb{R}^M$ is the linear decoding of neural firing rates $\mathbf{r}(t) \in \mathbb{R}^N$ given by

$$\mathbf{y}(t) = \mathbf{Dr}(t), \quad (2)$$

where $\mathbf{D} \in \mathbb{R}^{(M \times N)}$ is a decoding matrix. Firing rates $\mathbf{r}(t)$ are defined as the filtered spike trains $\mathbf{o}(t)$ following a linear differential equation

$$\tau \dot{\mathbf{r}}(t) = -\mathbf{r}(t) + \mathbf{o}(t), \quad (3)$$

where spikes of the i th neuron are defined by the sum of delta function

$$o_i(t) = \sum_k \delta(t - t_k^i) \quad (4)$$

of spike times $\{t_k^i\}$. Thus, spikes are convoluted with an exponential kernel $\exp(t/\tau)/\tau$ to form the firing rate $\mathbf{r}(t)$. This leads to a differential equation for the linear readout $\mathbf{y} = \mathbf{Dr}$ given by

$$\tau \dot{\mathbf{y}}(t) = \tau \mathbf{D} \dot{\mathbf{r}}(t) = -\mathbf{Dr}(t) + \mathbf{Do}(t) \quad (5)$$

$$= -\mathbf{y}(t) + \mathbf{Do}(t) \quad (6)$$

Second, we assume that the coding error L is given by the squared L_2 norm

$$L(t) = \|\mathbf{x}(t) - \mathbf{y}(t)\|^2 + C(\mathbf{r}(t)), \quad (7)$$

describing a loss function with an additional cost term $C(\mathbf{r}(t))$, which we will define later. In our assumption, a neuron spikes only if the emitted spike reduces the loss function L , such that at time t

$$L(t|\text{spike in neuron } i) < L(t|\text{no spike}), \quad (8)$$

and the spike increases $r_i \rightarrow r_i + 1$ and thus leads to a change in \mathbf{y} of \mathbf{D}_i , the i th column of the decoder matrix. The spiking mechanism is therefore defined by

$$\begin{aligned} & \|\mathbf{x}(t) - \mathbf{y}(t) - \mathbf{D}_i\|^2 + C(\mathbf{r}(t) + \mathbf{e}_i) \\ & < \|\mathbf{x}(t) - \mathbf{y}(t)\|^2 + C(\mathbf{r}(t)), \end{aligned} \quad (9)$$

where $\mathbf{e}_i \equiv \mathbb{1}_i^N$ is the i th unit base vector. If we define the cost term to be a quadratic cost of firing rates with $C(\mathbf{r}(t)) = \mu \|\mathbf{r}(t)\|^2$, we obtain the following inequality

$$\mathbf{D}_i^T (\mathbf{x}(t) - \mathbf{y}(t)) - \mu r_i(t) > \frac{\mathbf{D}_i^T \mathbf{D}_i + \mu}{2}. \quad (10)$$

Note that the vector \mathbf{D}_i^T means the transpose of column vector \mathbf{D}_i , and not the i th column vector of the transposed matrix \mathbf{D}^T . In Eqn. (10),

we define the left side to describe the membrane potential

$$V_i(t) \equiv \mathbf{D}_i^T (\mathbf{x}(t) - \mathbf{y}(t)) - \mu r_i(t) \quad (11)$$

of the i th neuron, which means that the neuron fires a spike if $V_i(t)$ surpasses a certain threshold value

$$\Theta_i \equiv \frac{\mathbf{D}_i^T \mathbf{D}_i + \mu}{2}. \quad (12)$$

Since each spike increases \mathbf{y} by \mathbf{D}_i and r_i by 1, V_i decreases by $\mathbf{D}_i^T \mathbf{D}_i + \mu = 2\Theta_i$. The vectorized form of the membrane potential of N neurons is given by

$$\begin{aligned} \mathbf{V}(t) &= \mathbf{D}^T (\mathbf{x}(t) - \mathbf{y}(t)) - \mu \mathbf{r}(t) \\ &= \mathbf{D}^T \mathbf{x}(t) - \mathbf{D}^T \mathbf{Dr}(t) - \mu \mathbf{r}(t) \\ &= \mathbf{D}^T \mathbf{x}(t) - (\mathbf{D}^T \mathbf{D} + \mu \mathbb{1}^N) \mathbf{r}(t) \\ &= \mathbf{F}^* \mathbf{x}(t) - \mathbf{W}^* \mathbf{r}(t) \end{aligned} \quad (13)$$

Thus, each neuron receives input through feed-forward connections $\mathbf{F}^* \equiv \mathbf{D}^T$ and from the network's rates through recurrent connections $\mathbf{W}^* \equiv \mathbf{D}^T \mathbf{D} + \mu \mathbb{1}^N = \mathbf{F}^* \mathbf{D} + \mu \mathbb{1}^N$, where μ acts as a self-inhibitory term (i.e., autapse). Note that the asterisks at the weight matrices indicate the optimal solution for our loss function. In general, the network may be defined by arbitrary feed-forward and recurrent connections \mathbf{F} and \mathbf{W} forming a suboptimal solution.

ii. Network dynamics

In order to describe the time-varying membrane potential $\mathbf{V}(t)$, we take its temporal derivative

$$\begin{aligned} \tau \dot{\mathbf{V}}(t) &= \tau (\mathbf{F} \dot{\mathbf{x}}(t) - \mathbf{W} \dot{\mathbf{r}}(t)) \\ &= -\mathbf{F}(\mathbf{x}(t) - \mathbf{c}(t)) + \mathbf{W}(\mathbf{r}(t) - \mathbf{o}(t)) \\ &= -(\mathbf{F}\mathbf{x}(t) - \mathbf{W}\mathbf{r}(t)) + \mathbf{Fc}(t) - \mathbf{Wo}(t) \\ &= -\mathbf{V}(t) + \mathbf{Fc}(t) - \mathbf{Wo}(t), \end{aligned} \quad (14)$$

using Eqn. (13) for the membrane potential. Therefore, we derived a network model equivalent to leaky integrate-and-fire neurons [32] with threshold Θ and reset 2Θ . Given that the loss function is minimized, membrane potentials fluctuate around zero, and therefore the

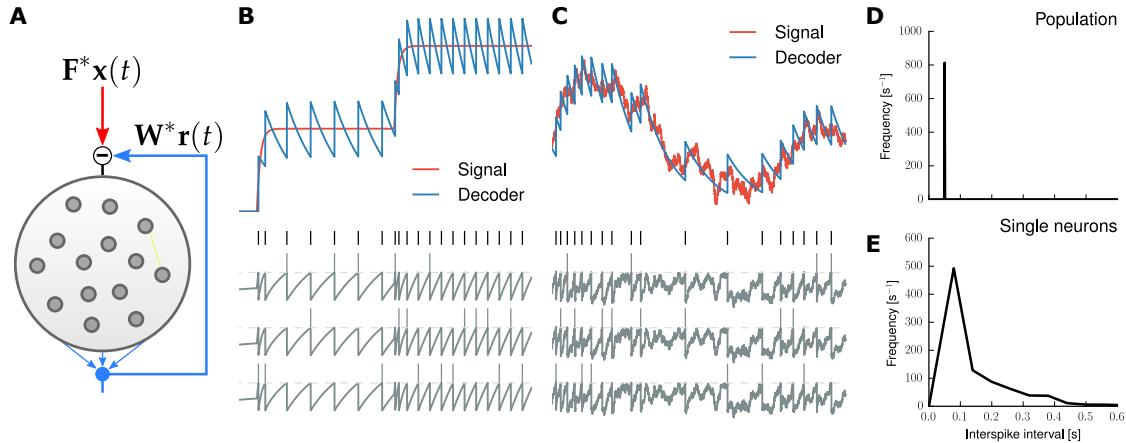


Figure 1: Encoding of a one-dimensional signal in a three-neuron network. (A) Schematic of optimal balanced networks. Steady-state membrane potentials of the neurons in the network receive balanced input from the encoding $F^*x(t)$ and decoding signal $W^*r(t)$. Thus, membrane potentials represent prediction errors of the reconstructed signal. (B) Step function (red) and decoding signal (blue) in a single trial. The membrane potential and spike trains for each of the three neurons are shown in separate colors and rows. (C) Random signal as input, otherwise same as (B). Note the trade-off between decoding accuracy and overall numbers of spikes emitted. (D–E) Interspike interval (ISI) distribution for population (D) and single-neuron (E) spike trains.

balance of excitatory and inhibitory currents of $\mathbf{Fc}(t)$ and $\mathbf{Wo}(t)$, respectively. In Fig. 1, we illustrate the network dynamics in a three-neuron example encoding a one-dimensional signal. While single neuron spike trains exhibit poisson-like statistics (Fig. 1D), the overall spiking of the population is regular given a constant input (Fig. 1E).

iii. Learning of recurrent connectivity

How can a randomly connected network learn to form optimal recurrent connections? From the leaky integrate-and-fire dynamics in Eqn. 14, we recall that an optimal connectivity achieves the balance of excitation and inhibition. For steady-state, optimal solutions, the membrane potentials $\mathbf{V}(t)$ will fluctuate around zero:

$$0 = \langle \mathbf{V}(t) \rangle = \dot{\mathbf{V}}(t) - \mathbf{V}(t). \quad (15)$$

A spike from neuron i at time t_k will reduce the membrane potential by \mathbf{We}_i , so that

$$\mathbf{V}(t_k) \rightarrow \mathbf{V}(t_k) - \mathbf{We}_i. \quad (16)$$

For a tight balance (i.e., $\langle \mathbf{V}(t) \rangle = 0$), we demand the steady-state equality

$$(\mathbf{We}_i)_\infty = -2\mathbf{V}(t_k) \quad (17)$$

which can be formed by a linear differential equation given by

$$\frac{1}{\eta} \dot{\mathbf{W}}_i = 2\mathbf{V}(t) - \mathbf{W}_i, \quad \text{if } i \text{ spikes}, \quad (18)$$

where $\mathbf{W}_i \equiv \mathbf{We}_i$ and η is a constant learning rate. This is the voltage-based learning rule without the quadratic cost term.

Considering the cost term, each spike by neuron i leads to $\mathbf{V}(t_k) \rightarrow \mathbf{V}(t_k) - \mathbf{We}_i - \mu \mathbf{e}_i$. Tight balance is achieved for the steady state, if

$$\mathbf{W}_i = \mathbf{FD}_i + \mu \mathbf{e}_i \quad (19)$$

Therefore, we construct a learning rule based on a linear differential equation given by

$$\frac{1}{\eta} \dot{\mathbf{W}}_i = 2(\mathbf{V}(t) + \mu \mathbf{r}(t)) - \mathbf{W}_i + \mu \mathbf{e}_i, \quad (20)$$

if neuron i spikes.

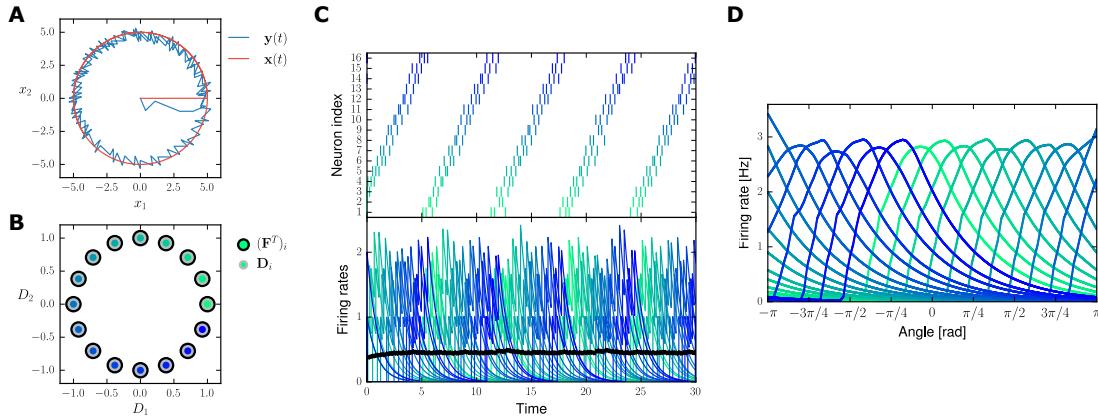


Figure 2: Encoding of a two-dimensional, circular signal in a network of $N = 16$ neurons using a quadratic cost factor of $\mu = 0.1$. (A) Input signal $x(t)$ and decoding signal $y(t)$ for $T = 10$ s. After only a few time steps, the coding error $\|(x(t) - y(t))^2\|$ between input and decoder is bounded by $\Theta = \frac{\|D\|}{2}$. (B) Feed-forward encoding and decoding weight vectors $(\mathbf{F}^T)_i$ and \mathbf{D}_i , respectively, in signal space. The edgecolor of the encoding vectors are black, while the decoding vectors are indicated by a gray edgecolor. Each neuron is color-coded using the Matplotlib colormap "WINTER". The same color code is used in (C–D). (C) Spike raster and firing rates for simulating the network for $T = 30$ s. The spike pattern reflects the periodic nature of the incoming signal. The population rate $R(t) = \frac{\sum_i^N r_i(t)}{N}$ is indicated by the black solid line. Note that, while single neuron spiking is variable, the population rate is reliably constant and is proportional to the amplitude of the input signal. (D) Numerical estimation of tuning curves of the network using univariate spline interpolation of firing rates over $T = 500$ s.

III. SIMULATIONS

In order to examine the interactions between optimal balanced compensation and recurrent plasticity during neuronal death, we carried out numerical simulations of optimal balanced networks with $N = 16$ neurons encoding a two-dimensional signal $\mathbf{x}(t)$ given by

$$\mathbf{x}(t) = \begin{pmatrix} 5 \cos(t) \\ 5 \sin(t) \end{pmatrix}, \quad (21)$$

which describes a circle in the two-dimensional space. The differential equations from section II are numerically solved using linear Euler integration $\mathbf{x}(t + dt) = \mathbf{x}(t) + dt \cdot \dot{\mathbf{x}}(t)$. Therefore, spike trains have to be scaled by $\frac{1}{dt}$ in order to mathematically be equivalent to a delta distribution as defined above. For all simulation described in this report, we applied an integration step size of $dt = 0.001$. Since this is approximately the time scale of spikes (≈ 1 ms), we have all temporal variables defined in seconds (s). Note that spike rates therefore are

described in Hertz (Hz). All other variables are assigned with arbitrary units, but it is possible to assign physical units by applying appropriate scalings and offsets. The time constants of the differential equations are set to $\tau = 1$ s. Note that individual time scales can be applied for model fitting.

Fig. 2 shows the results of an optimal network with 16 neurons encoding $\mathbf{x}(t)$ for a total duration of $T = 30$ s. Fig. 2A shows the signal $\mathbf{x}(t)$ and the decoder readout $\mathbf{y}(t)$ for the first ten seconds of the simulation, showing that the accurate representation of the signal is achieved rapidly by fast, repeated spiking. Almost instantaneously the membrane potential is bounded and balanced between $-\Theta$ and Θ , effectively bounding \mathbf{y} around \mathbf{x} by $\frac{\|D\|}{2}$. Fig. 2B shows the 16 encoding vectors $(\mathbf{F}^T)_i$ and decoding vectors \mathbf{D}_i in the signal space. The weights are set to

$$(\mathbf{F}^T)_i = \mathbf{D}_i = \begin{pmatrix} \cos(2\pi i) \\ \sin(2\pi i) \end{pmatrix}, i \in [0, N - 1] \quad (22)$$

which is shown by the exact alignment and evenly spaced vectors. Furthermore, tight balance in the network is achieved for the case of $\mathbf{W} = \mathbf{FD} - \mu \mathbf{1}^N$. In this simulation, we applied a quadratic cost factor of $\mu = 0.1$, which increases self-inhibition and therefore reduces firing rates. In Fig. 2C, spike raster and firing rates are shown, where the population rate $R(t) = \frac{\sum_i r_i(t)}{N}$ is indicated by the black solid line. The spike pattern describes population-wide the oscillatory course of the encoded signal, while individual spikes are variable. The tuning curves of each neuron is shown in Fig. 2D, which describe bell-shaped function similar to the orientation tuning in the visual cortex or directional tuning in the cricket cercal system. Note that the tuning curves have been calculated numerically from a longer simulation ($T = 500$ s). Therefore, we can explain the long tails in the curves by the way we estimate the rates using an exponential kernel.

In order to measure the effect of different quadratic cost factor μ on the firing statistics and coding accuracy, we carried out ten simulations ($T = 100$ s) for each of the values $\mu = \{0.0, 0.01, 0.02, 0.1, 0.2, 0.5, 1, 2, 5, 10\}$. We measured for each of the simulation the mean population rate $\langle R(t) \rangle$, mean population rate variance σ_R^2 , and mean squared coding errors $\langle \epsilon \rangle = \langle \| \mathbf{x} - \mathbf{y} \|^2 \rangle$. Note that the mean is calculated across time of individual trials. The results are shown in Fig. 3, where the mean of the ten repetitions of each value for μ is indicated by the dark-gray colored solid line.

i. Learning optimal recurrent weights

In order to show that Eqns. 19 & 20 lead to stable convergence of $\mathbf{W}(t)$ towards optimal recurrent weights $\mathbf{W}^* = \mathbf{FD} + \mu \mathbf{1}^N$, we carried out a simulation for $T = 200$ s with a learning rate of $\eta = 0.01$. The results are shown in Fig. 4. We initialize the recurrent weight matrix as a Gaussian random matrix, where each entry is drawn from a Gaussian distribution.

$$W_{ij} \sim \mathcal{N} \left(0, \frac{1}{N^2} \right). \quad (23)$$

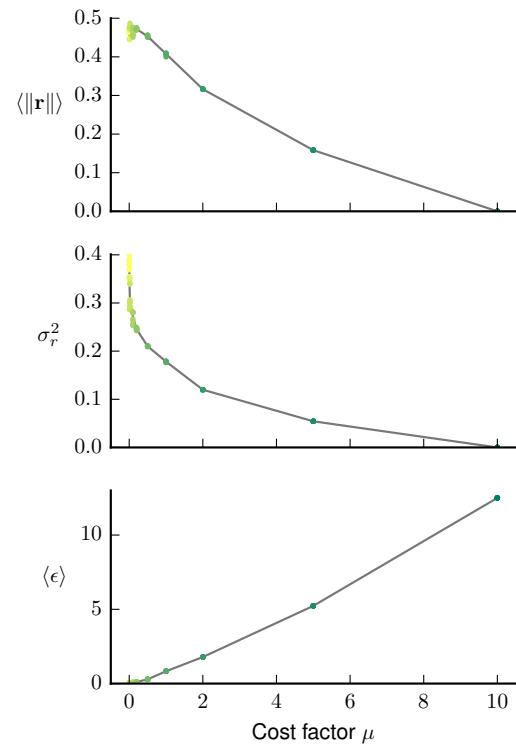


Figure 3: Effects of different quadratic cost factors μ on mean population firing rates, their variances, and mean squared coding errors. We varied μ with values $0.0, 0.01, 0.02, 0.1, 0.2, 0.5, 1, 2, 5, 10$, which are individually color-coded. The gray solid line indicates the mean values for each value of μ .

The encoding weight vectors $(\mathbf{F}^T)_i$ are assumed to be fixed and optimal for the circular input signal as described above. The decoding weight matrix \mathbf{D} is calculated from the final recurrent weights using

$$\begin{aligned} \mathbf{W}(T) &\approx \mathbf{FD} \\ \Leftrightarrow \mathbf{D} &\approx \mathbf{F}^+ \mathbf{W}(T), \end{aligned}$$

where \mathbf{F}^+ is the Moore-Penrose inverse matrix of the encoding weights. Fig. 4A shows the input and decoder signal for the first (top) and final 50 (bottom) seconds, respectively. After learning, the decoder reconstructs the input signal accurately up to $\frac{\|\mathbf{D}\|}{2}$. In Fig. 4B, the encoding and decoding weight vectors are shown. Since the cost factor $\mu = 0.01$ is rela-

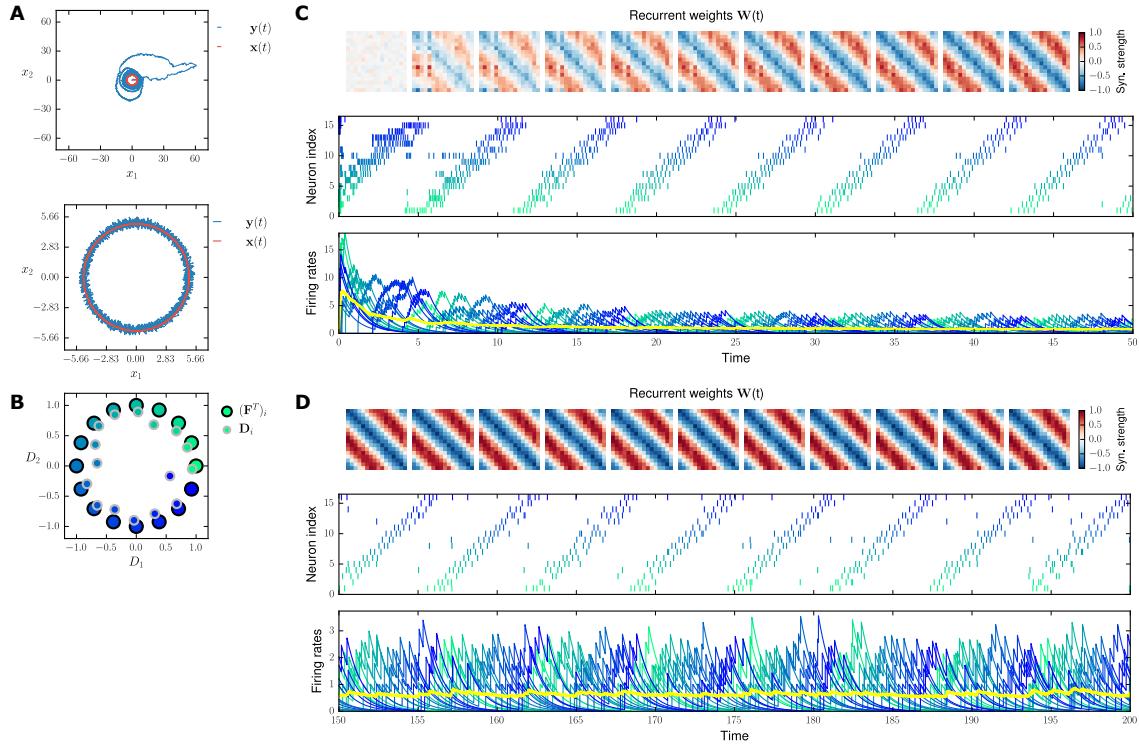


Figure 4: Learning optimal recurrent connectivity in a network of $N = 16$ neurons using a quadratic cost factor of $\mu = 0.01$. (A) Input signal $\mathbf{x}(t)$ and decoding signal $\mathbf{y}(t)$ during the initial phase of learning at $t = 0 - 50$ s (top) and final phase of learning at $t = 150 - 200$ s (bottom). First, the decoder is overshooting the signal reflecting suboptimal recurrent weights, which favor excitation in the network. After learning, the coding error is bounded like in the optimal network in Fig. 2. (B) Feed-forward encoding and decoding weight vectors $(\mathbf{F}^T)_i$ and \mathbf{D}_i , respectively, in signal space. The edgecolor of the encoding vectors are black, while the decoding vectors are indicated by a gray edgecolor. We used the same color code as in Fig. 2. (C–D) Recurrent weight matrix, spike trains and firing rates with respect to time for (C) $t = 0 - 50$ s and (D) $t = 450 - 500$ s. The population rate is shown as a yellow solid line.

tively small, the optimal solution for the firing rates allow for additional firing, which are indicated by near-optimal recurrent and therefore also decoding weights. Fig. 4C & D shows the time evolution of recurrent weights $\mathbf{W}(t)$, spike trains, and firing rates for the first (C) and the final (D) 50 seconds. During learning, the recurrent weights approach towards the optimal solution, which reduce the firing rates given by the unbalanced, random initial weights. Note that the population rate is shown as a yellow solid line. After learning, the spike pattern exhibits a similar periodic spike pattern and constant-mean fluctuating firing rate as in Fig. 2.

ii. Optimal balance and recurrent plasticity for compensating neuronal death at constant rate

We carried out several simulations with a network described in the section above for $T = 2000$ s, in which we ablate neurons sequentially in random order. An ablation of neuron i at $t_{\text{abl},i}$ is given by setting

$$\mathbf{F}_i^T(t) = \mathbf{W}_i(t) = \mathbf{W}_i^T(t) = V_i(t) = 0, \quad \forall t \in [t_{\text{abl},i}, T]. \quad (24)$$

Ablations are at a constant rate of $f_{\text{abl}} = 0.01 \text{ s}^{-1}$ (i.e., one ablation per 100 s) and start after an initial phase of 10 seconds, so that the net-

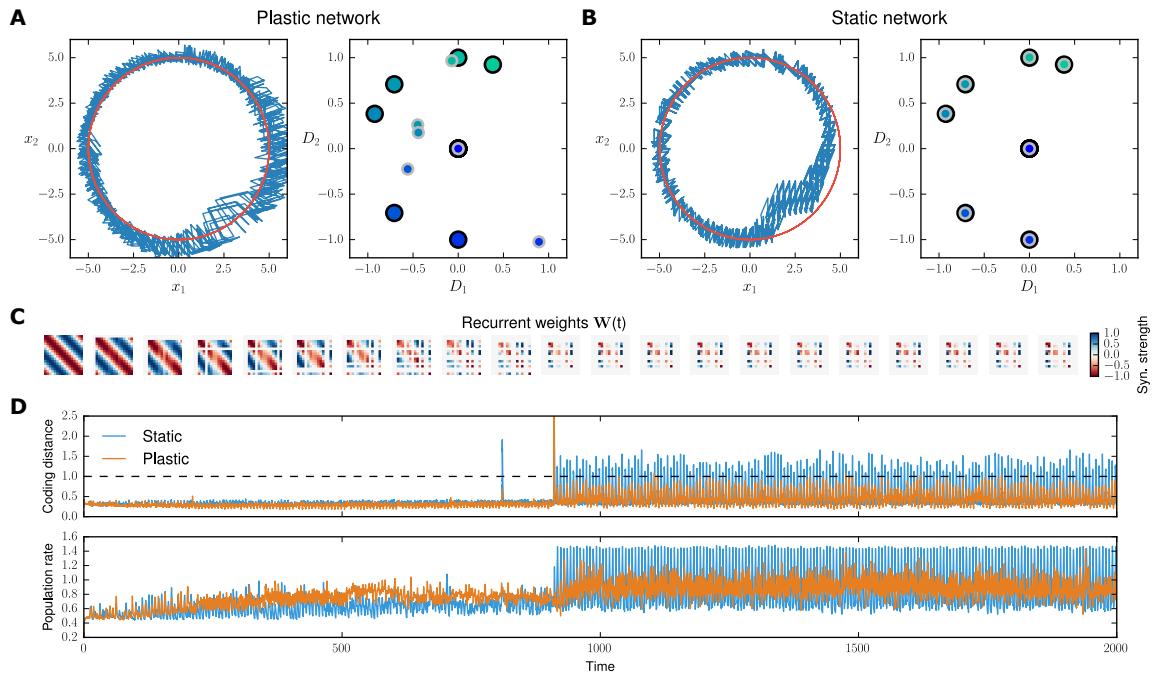


Figure 5: Comparison of networks with and without recurrent plasticity during constant-rate neuronal death. **(A–B)** Input signal $\mathbf{x}(t)$ and decoding signal $\mathbf{y}(t)$ during the final 50 s, as well as the final encoding and decoding weight vectors for the **(A)** plastic and **(B)** static network. The color code is similar to the previous figures. **(C)** Recurrent weight matrix with respect to time. The time difference between each display is $\Delta t = 100$ s to show effects of individual ablations. **(D)** Coding distance and population rates for the plastic and static network. We measure coding accuracy by using the distance $\|\mathbf{x} - \mathbf{y}\|$ between input and decoding signal, which is lower for the plastic network (i.e., better coding accuracy). We show that recurrent plasticity reduces the variance of the population rate by equalizing the increase of certain firing rates due to acute optimal balance compensation.

work is not driven by transients. We initialize optimal recurrent weights $\mathbf{W} = \mathbf{F}\mathbf{F}^T$. All other parameters are defined as described above. We compare a network with recurrent plasticity to one with static connectivity. Since the feed-forward and recurrent weights are changed with each ablation, we constantly recalculate the decoding matrix \mathbf{D} using the pseudoinverse relation from Eqn. 24.

Fig. 5 shows the results of comparing the plastic and the static network during neuronal death. The order of neurons that are ablated is random, but same for both networks: $\{0, 1, 5, 14, 13, 11, 8, 9, 2, 15\}$. After these ten ablations ($\cong 62.5\%$ of the population), the network is further simulated until it reaches $t = T$. Fig. 5A & B show the signal (input and decoder) and weight space (feed-forward and de-

coding) of the plastic (A) and static (B) network of the final 50 seconds, respectively. While the static network has a decreased coding accuracy for the lower right quadrant of the input signal, the plastic network adapts its recurrent weights to compensate for this error. Fig. 5C shows the recurrent connectivity of the plastic network with respect to time. Note that each matrix are spaced in time by $\Delta t = 100$ s, in order to show individual ablations. It is apparent, that the amplitude of recurrent weights is decreased as neurons are ablated. Fig. 5D shows the functional consequences of recurrent plasticity in terms of coding. We measured the coding distance $\|\mathbf{x} - \mathbf{y}\|$ and show that the plastic network has a distance below 1 for most of the time. The static network shows periodically high coding inaccuracies corresponding to the

ablated neurons. Furthermore, the firing rates of both networks increase as neurons are ablated. However, the plastic network exhibits a lower variance of the population rates. It shows that in the plastic network, the increase of firing rates is homogenous across the population.

IV. CONCLUSIONS

In summary, we showed that optimal balanced networks do not only offer acute compensation for neuronal death, but when combined with learning rules, it further restores balance of excitation and inhibition. This is due to a decrease in population rate variance. For future work, it would be interesting to include feed-forward plasticity [31], because this way the network could be perturbed to the theoretical limit of possible ablations without losing coding accuracy (see Appendix). Furthermore, the network could be fed by a higher-dimensional signals, which are usually given in sensory systems. Nonetheless, this project was an important step towards understanding interactions of neuronal and synaptic compensations during neuronal death.

REFERENCES

- [1] David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press, 1982.
- [2] Edgar Douglas Adrian and Yngve Zotterman. The impulses produced by sensory nerve-endings: Part ii. the response of a single end-organ. *The Journal of physiology*, 61(2):151–171, 1926.
- [3] Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.
- [4] Timothy EJ Behrens, Mark W Woolrich, Mark E Walton, and Matthew FS Rushworth. Learning the value of information in an uncertain world. *Nature neuroscience*, 10(9):1214–1221, 2007.
- [5] Jan Drugowitsch, Gregory C DeAngelis, Eliana M Klier, Dora E Angelaki, and Alexandre Pouget. Optimal multisensory decision-making in a reaction-time task. *eLife*, 3:e03005, 2014. ISSN 2050-084X. doi: 10.7554/eLife.03005. URL <https://dx.doi.org/10.7554/eLife.03005>.
- [6] Fred Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183, 1954.
- [7] H. B. Barlow. *Possible principles underlying the transformations of sensory messages*, pages 217–234. MIT Press, Cambridge, MA, 1961.
- [8] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. ISSN 1538-7305. doi: 10.1002/j.1538-7305.1948.tb01338.x. URL <http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- [9] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:13, 1996.
- [10] Anthony J Bell and Terrence J Sejnowski. The “independent components” of natural scenes are edge filters. *Vision research*, 37(23):3327–3338, 1997.
- [11] Joseph J Atick and A Norman Redlich. What does the retina know about natural scenes? *Neural computation*, 4(2):196–210, 1992.
- [12] Xaq Pitkow and Markus Meister. Decorrelation and efficient coding by retinal ganglion cells. *Nature neuroscience*, 15(4):628–635, 2012.
- [13] M. V. Srinivasan, S. B. Laughlin, and A. Dubs. Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B: Biological Sciences*, 216(1205):427–459, 1982. ISSN 0080-4649. doi: 10.1098/rspb.1982.0085. URL <http://rspb.royalsocietypublishing.org/content/216/1205/427>.

- [14] JH Van Hateren. Real and optimal neural images in early vision. *Nature*, 1992.
- [15] Evan C Smith and Michael S Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–982, 2006.
- [16] Michael R Bale, Robin AA Ince, Greta Santagata, and Rasmus S Petersen. Efficient population coding of naturalistic whisker motion in the ventro-posterior medial thalamus based on precise spike timing. *Frontiers in neural circuits*, 9, 2015.
- [17] Naama Brenner, William Bialek, and Rob de Ruyter Van Steveninck. Adaptive rescaling maximizes information transmission. *Neuron*, 26(3):695–702, 2000.
- [18] Naama Brenner, Steven P Strong, Roland Koberle, William Bialek, and Rob R De Ruyter Van Steveninck. Synergy in a neural code. *Neural computation*, 12(7):1531–1552, 2000.
- [19] Emilio Salinas and Terrence J Sejnowski. Correlated neuronal activity and the flow of neural information. *Nature reviews neuroscience*, 2(8):539–550, 2001.
- [20] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366, 2006.
- [21] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.
- [22] Apostolos P Georgopoulos, Andrew B Schwartz, Ronald E Kettner, et al. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, 1986.
- [23] Zachary F Mainen and Terrence J Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503, 1995.
- [24] Michael N Shadlen and William T Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of neuroscience*, 18(10):3870–3896, 1998.
- [25] Carl van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724, 1996.
- [26] Alfonso Renart, Jaime De La Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex Reyes, and Kenneth D Harris. The asynchronous state in cortical circuits. *science*, 327(5965):587–590, 2010.
- [27] Sophie Denève and Christian K Machens. Efficient codes and balanced networks. *Nature neuroscience*, 19(3):375–382, 2016.
- [28] Martin Boerlin, Christian K Machens, and Sophie Denève. Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comput Biol*, 9(11):e1003258, 2013.
- [29] David G Barrett, Sophie Denève, and Christian K Machens. Firing rate predictions in optimal balanced networks. In *Advances in Neural Information Processing Systems*, pages 1538–1546, 2013.
- [30] David Barrett, Sophie Deneve, and Christian Machens. Optimal compensation for neuron death. *bioRxiv (preprint)*, 2015.
- [31] Wieland Brendel, Ralph Bourdoukan, Pietro Vertechi, David G. T. Barrett, Christian K. Machens, and Sophie Denève. Optimal spike-timing-dependent plasticity for maximizing representational efficiency. submitted, 2016.
- [32] Richard B Stein. Some models of neuronal variability. *Biophysical journal*, 7(1):37, 1967.

APPENDIX

Geometry of a spike

In the last section, we defined the spike mechanism of a neuron as a form of loss reduction for a given error and cost. Since \mathbf{x} , \mathbf{y} and \mathbf{D}_i lie in the same vector space \mathbb{R}^M , we are able to apply geometry to understand, when a spike is triggered and how a spike affects the loss function. In order to simplify this analysis, we neglect the cost term $C(\mathbf{r}(t))$ for now. Thus, we examine the trajectory of the error $\epsilon(t) = \mathbf{x}(t) - \mathbf{y}(t)$, for which the spiking mechanism in Eqn. (10) can be rewritten as

$$\mathbf{D}_i^T \epsilon(t) > \frac{\mathbf{D}_i^T \mathbf{D}_i}{2}, \quad (25)$$

where $\mathbf{D}_i^T \mathbf{D}_i = \|\mathbf{D}_i\|^2 = \|\mathbf{D}_i^T\|^2$ is the squared vector norm of \mathbf{D}_i . This leads to

$$\frac{\mathbf{D}_i^T}{\|\mathbf{D}_i^T\|} \epsilon(t) > \frac{\|\mathbf{D}_i\|}{2}, \quad (26)$$

and therefore neuron i spikes, when the scalar projection of the error $\epsilon(t)$ onto the weight vector \mathbf{D}_i exceeds more than half of its vector length in the signal space \mathbb{R}^M . Therefore, the spiking thresholds Θ_i can be seen as hyperplanes

$$\mathcal{H}_i = \left\{ \mathbf{v} : \mathbf{D}_i^T \mathbf{v} = \frac{\Theta_i}{\|\mathbf{D}_i\|} \right\} \quad (27)$$

spanning in \mathbb{R}^M . As neuron i elicits a spike, we recall that \mathbf{y} increases by \mathbf{D}_i , and thus $\epsilon \rightarrow \epsilon - \mathbf{D}_i$ that moves $\epsilon(t)$ inside the threshold boundary. The intersections of hyperplanes $\{\mathcal{H}_i\}_{i=1}^N$ which form a so-called *polytope*, a convex hull of a bounded, finite set of points. This is given, if $N \geq M + 1$, and $\left| \frac{\mathbf{D}_i^T \mathbf{D}_j}{\|\mathbf{D}_i\| \|\mathbf{D}_j\|} \right| < 1 \quad \forall i \wedge (j \neq i)$ for the special case of $N = M + 1$. The error $\epsilon(t)$ is bounded for $t \rightarrow \infty$, if and only if the polytope of hyperplane intersections contains the origin $\mathbf{0}$, where the squared error is zero.

Examples in two dimensions

Here, we illustrate two simple examples in a low-dimensional ($M = 2$) space. Using these

examples, we aim to describe spiking properties of the network for error boundedness and convexity.

Example 1: $N = 3$ (2-simplex)

In two dimensions, the minimum number of neurons in order to bound the error is $N = 3$. In this extreme case, we also demand that all pairs of threshold boundaries are not parallel to each other. In Fig. 6, we show examples of different sets of decoding vectors and their resulting thresholds boundaries. In the left-most case, the decoding vectors are span in \mathbb{R}^2 evenly with equal length. We then varied the angle $\alpha_{\mathbf{D}_0}$ to the x-axis increasing it by 30° each time, while keeping the other vectors constant. The red-colored area indicates the polygon formed by the intersections of the threshold boundaries. For $\alpha_{\mathbf{D}_0} = 0^\circ$ and 30° , the polygon contains the origin, and therefore the error ϵ is bounded. For $\alpha_{\mathbf{D}_0} = 90^\circ$, two of the threshold boundaries are parallel, such that the error cannot be bounded. As the vector is further rotated, the polygon of the intersection does not contain the origin, and therefore error boundedness is not given. Note that it is equivalent to say that the error is bounded, if the vector span of the decoding vectors is equal to \mathbb{R}^N .

As described above, each spike leads to $\epsilon \rightarrow \epsilon - \mathbf{D}_i$, which may lead to an error outside another threshold boundary under certain conditions. Therefore the network would elicit multiple spikes in consecutive time steps. The 3-neuron network is a special case, where multispiking occurs very likely. In Fig. 7, we illustrate this phenomenon by several examples.

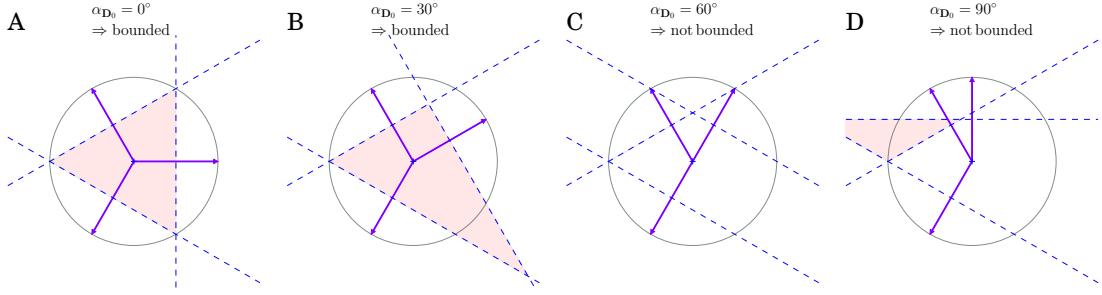


Figure 6: Examples of (A-B) bounded and (C-D) unbounded error conditions for the 3-neuron network in two dimensions.

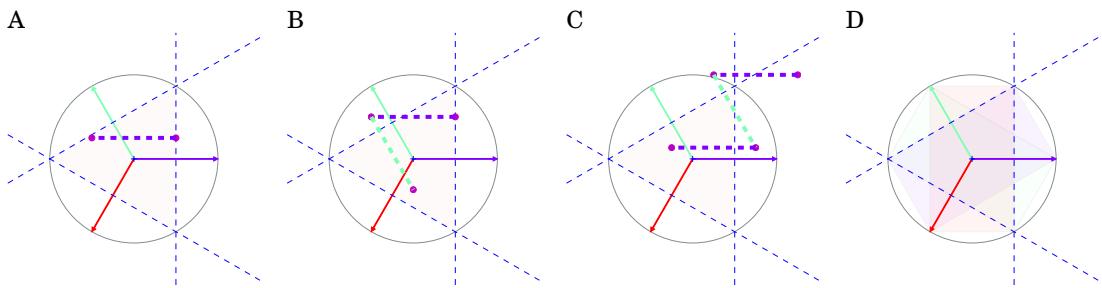


Figure 7: Examples of multiple spikes in the 3-neuron network for two dimensions. Depending on a given error, the neuron can spike (A) one, (B) two, or (C) more times to reach inside the error bound. (D) Assuming that the error dynamics are slow, i.e., the error slowly reaches the threshold, one can illustrate possible effects of a spike by a rectangle spanned by $-D_i$. The probability of exhibiting two spikes is therefore given by the ratio of areas outside vs. inside the error bound.