Humboldt-Universität zu Berlin
Fachbereich Mathematik und Informatik

Master thesis

to obtain the academic degree
Master of Science (M. Sc.)
in Mathematics

Mean-field theory of the predictive coding spiking neural network

Alexander Spokoinyi
aspokoinyi@gmail.com

1. Supervisor: Prof. Dr. Tilo Schwalger, Technische Universität Berlin
2. Supervisor: Prof. Dr. Markus Reiss, Humboldt-Universität Berlin

November 3, 2021

Contents

Acknowledgments

This thesis is dedicated to my grandfather Vadim Paley, who passed away while I was still working on it. His unparalleled kindness, uncompromising joy of life, and hard-work mentality driven by profound enthusiasm will always remain one of my greatest inspirations in life.

I would also like to take this opportunity to express my deepest gratitude to my primary supervisor, Professor Tilo Schwalger, who guided me throughout the thesis. His expertise and dedicated input were invaluable for the formulation of key research questions, methodology and crucial calculations.
At the same time, I would like to thank Dr. Veronika Koren, whose proficiency on the topic and whose insightful feedback were extremely helpful in various stages of my work.

Aside from academic support, I am extremely grateful to my family and friends for their continuous encouragement and help. First and foremost, I am deeply thankful to Karin Tröber for her passionate, unwavering support even through very hard times. I would also like to thank Ann Christin Vietor for both her moral and professional assistance, on which I could rely at all times.

Lastly, I cannot express how grateful I am to my grandmother Doroteya Paley, whose courage, virtue, wisdom and optimism have been a source of true strength and inspiration.

1 Introduction

Biological neurons are challenged with the enormous task to reliably process large amounts of information in a highly uncertain environment. Indeed, neuronal activity, encoded through sequences of action potentials (or spikes) - fast signals communicated through synaptic connections - is notoriously known for its ubiquitous, Poisson-like variability. Therefore, a central problem in computational neuroscience is to explain how efficient coding can be performed under such conditions.

One possibility could be that reliable outcomes are obtained through averages over larger populations of neurons carrying the same information. This could minimize the effects of biophysical noise and disturbances, which affect individual neurons. However, the overall effect of such microscopic fluctuations is too weak to account for the full extent of cortical variability [1, 2]. Moreover, one would reasonably hope that evolution came up with a more sophisticated way to perform calculations than to rely on a large amount of redundant information. Indeed, the associated error for a network of N neurons decreases only as $1/\sqrt{N}$, such that a significant increase in computational units would yield only a small improvement of the overall error.

This has led researchers to consider the global network activity itself as a possible source of fluctuations. Using mean-field theory, where stochastic variables that depend on population-level activity are replaced by their average values, the dynamics of large networks can be derived analytically [3]. Such an approximation becomes exact in the large N limit $N \to \infty$ and can yield important insights for realistic population sizes as well. In particular, it has been shown that the dynamics of a large network of leaky integrate-and-firing (LIF) neurons stabilize globally, while individual firing rates are still highly irregular [4, 5]. A crucial element here is the notion of balance. Each neuron is constantly swamped by excitatory and inhibitory currents, both strong and of order $O(\sqrt{N})$. However, the currents cancel each other at subthreshold values, such that the net input to each neuron is governed by the resulting fluctuations of order 1, yielding irregular spiking at low firing rates.

While these theoretical results successfully reconciled stable, reliable dynamics with the strong heterogeneity observed in cortical recordings, the precise mechanisms and reasons which would lead to such a regime remained unclear.

More recently, a powerful framework has been introduced by Boerlin et al [6], who demonstrate that efficient coding and balanced dynamics are signatures of networks that follow the paradigm of predictive coding [7]. In particular, the neural network generates real-time predictions of some dynamic sensory input, capable of implementing arbitrary linear dynamical systems. From the biological point of view, the resulting prediction errors can be propagated to other cortical areas and have been shown to play an important role in various contexts such as audiovisual speech [8], motor control [9] or learning of causal relationships in general [10].

The predictive coding model gives a compelling account of neural computations. Biologically plausible leaky integrate-and-fire dynamics arise automatically from the optimization of a given objective function. This ensures that spikes are used optimally to minimize the prediction error. Moreover, neurons exhibit high Poisson-like variability, while still ensuring high coding accuracy with error decreasing as $1/N$. This feat is achieved through a meticulous balance of feedforward excitation and recurrent inhibition, characterized by extremely strong levels of both antagonists, with each current growing as $O(N)$ - a regime which is known as 'tight' balance.

As a consequence, the results have spurred an impressive amount of consecutive research [11, 12, 13, 14, 15] (among others). Recently, first theoretical results have also been established for the model using mean-field theory [16]. However, the approach is based entirely on firing rates, with artificially imposed constraints to fit the predictive coding framework. Therefore, results cannot be applied directly to the original framework, nor realistic spiking models (as

they are observed in the brain) in general.
In this work, we bridge the gap between theoretical rate models and realistic spiking networks by deriving mean-field results for a Linear-Nonlinear Poisson approximation of the original model developed by Boerlin et al [6]. Specifically, we match instantaneous firing rates as well as structured connectivity directly to the LIF dynamics of the functional predictive coding network. By producing spikes through an inhomogeneous Poisson point process driven by the derived instantaneous firing rates, the obtained results can be applied directly to the LIF network. Moreover, incorporating spiking into our analysis reveals important differences in dynamics compared to pure rate models.
Lastly, a particular focus is made regarding the level of balance required for the efficient functioning of the network. Indeed, the question of balance employed in the brain is a topic of intense debate in the non-scientific community [17, 18]. While theoretical researchers focused mostly on tightly balanced networks, a recent study [18] argues that loose balance, where excitation and inhibition are merely of order 1, is a much more realistic regime based on physiological evidence. Therefore, instead of imposing tightly balanced dynamics as in [6], we introduce a generalized, unifying scaling framework where the level of balance can be adjusted independently (similar to the analysis of [16]).
All results are supplemented with empirical simulations. The corresponding code is available under the link github.com/spokworks/thesis_predictive_coding.

2 Preliminaries

Neurons convey information via sequences of action potentials (or 'spikes'), which are characterized by a fast rise and subsequent fall in voltage recorded intracellularly. within the membrane potential of a neuron. Due to the fast and stereotyped trajectories of action potentials, a sequence $(t_1, t_2, \dots)$ of spikes is typically modeled as the sum y of instantaneous events in the form of Dirac δ functions [19]:

$$y(t) = \sum_k \delta(t - t_k).$$

Given $y(t)$, which we refer to as the spike train of the neuron, one can easily formalize and perform computations that depend on the neuronal firing activity. The central question, therefore, lies in the modeling of the spike arrivals, i.e. the underlying mechanism which causes the neuron to fire.
In this section, we give a brief introduction to two standard concepts in the context of spike generation, which we will use throughout this work. Hereby, we closely follow the descriptions of Dayan and Abbott [19] and Gerstner et al. [20]. For the unfamiliar reader, we can highly recommend either of these books for an excellent introduction to computational neuroscience.

2.1 Leaky integrate-and-fire models

A widely used model in computational neuroscience is the leaky integrate-and-fire (LIF) model. Introduced by Lapique in 1907 [21], it serves to this day as a simple yet powerful approximation of neuronal dynamics. Moreover, it has been shown to reproduce activity recorded in several parts of the neocortex [22]. Generally, the state of a neuron can be described by the electric potential in its cell membrane relative to its exterior. In its resting state, the potential is typically around -70mV. When the membrane potential is sufficiently depolarized through incoming electrical currents, the neuron emits a spike in form of an action potential. The idea behind the leaky integrate-and-fire model is to describe these main physiological characteristics, which govern the behavior of a typical neuron while ignoring the underlying biophysical details [19]. In particular, the distinct conductances and mechanisms involved in spike generation (e.g. as described by the Hodgkin-Huxley model [23])

are replaced by a single overall leak current, which controls the intrinsic dynamics of the membrane potential and maintains its resting state. Moreover, the model markedly reduces the complexity in the representation of action potentials by postulating a simple mechanistic rule: Whenever the membrane potential V reaches a critical threshold value ϑ, the neuron emits a spike and the membrane potential is instantaneously reset to a value $V_R < \vartheta$. Thus, as motivated above, only the event of firing itself matters disregarding its exact electrophysiological evolution. In analytical terms, the model can be expressed as follows: [20]: Firstly, the subthreshold dynamics of the membrane potential V of a given neuron are described by a simple dynamical equation:

$$\tau_m \dot{V}(t) = -(V(t) - V_{rest}) + RI_{syn}(t), \quad V < \vartheta. \tag{1}$$

Here, $-V$ on the right-hand side accounts for the name-giving leak current, V_{rest} is the resting potential, I_{syn} is a synaptic input current, τ_m denotes the membrane time constant and R is the membrane resistance. If the external input I_{syn} is zero, V decays exponentially to the resting potential, with the duration of decay determined by the time constant τ_m. The resistance R describes how strong the neuron responds to given stimuli.
Secondly, as explained above, action potentials occur at times t_k when the membrane potential crosses the threshold ϑ:

$$t_k = \inf\{t : t > t_{k-1}, V(t) \geq \vartheta\}, \quad k > 2,$$

with first spike time

$$t_1 = \inf\{t : t > 0, V(t) = \vartheta\}.$$

At any such crossing time, the membrane potential is instantaneously reset to a value V_R below the threshold, i.e.

$$V(t_k^+) = V_R < \vartheta,$$

where $t_k^+ = \lim_{t \downarrow t_k} t$ denotes the moment immediately after the spike. Without loss of generality, we will consider a model with membrane resistance $R = 1$, resting potential of $V_{rest} = 0$ and threshold $\vartheta = 1$. This corresponds to a simple rescaling of the membrane potential V. Indeed, writing $V^* = (V - V_{rest})/(\vartheta R)$ and $I^*_{syn} = I_{syn}/\vartheta$ it follows

$$\tau_m \dot{V}^* = \frac{1}{\vartheta R}\left(\tau_m \dot{V}\right) = -V^* + I^*_{syn}.$$

Moreover, the firing condition $V(t) = \vartheta$ becomes equivalent to $V^*(t) = 1$. Thus, redefining both membrane potential and external input as $V := V^*$ and $I_{syn} := I^*_{syn}$, respectively, yields the desired properties for our model. Note that due to the scaling by $1/\vartheta$, both quantities (as well as any other current in the model) become dimensionless.
To account for the highly irregular nature of neurophysiological data, it is common to add a stochastic noise term to the dynamic equation. It can be interpreted as the combined effect of the vast amount of interactions and physiological processes which take place at the microscopic scale and appear in the membrane potential in the form of small, unpredictable fluctuations. Here, we write

$$\tau_m \dot{V}(t) = -V + I_{syn}(t) + \sqrt{2\tau_m}\sigma_V \eta(t), \tag{2}$$

where η corresponds to Gaussian white noise with autocorrelation $\langle \eta(t)\eta(t')\rangle = \delta(t - t')$. This stochastic dynamic equation corresponds to an Ornstein-Uhlenbeck process with noise intensity (accounting for typical amplitudes of the noise) σ_V. For an insightful, yet concise introduction to stochastic process in the context of computational neuroscience we refer to the book chapter by Lindner [24].

In the remainder of this work, we will mainly use a slightly modified version of this classical notation, described by the dynamic equation

$$\dot{V} = -\lambda V + I_{syn} + \sqrt{2\lambda}\sigma_V \eta, \tag{3}$$

where $\lambda = \tau_m^{-1}$ denotes the inverse of the membrane potential. Note that the model is still mathematically equivalent to the one described by Equation (2), the only difference being that the external input is implicitly scaled by τ_m.

2.2 Firing rate models and LIF transfer function

As spiking sequences vary from trial to trial, neuronal activity is commonly described in a statistical manner using firing rates. Formally, the firing rate r of a neuron can be defined as the expected number of spikes per unit time step,

$$r(t) = \langle y(t) \rangle.$$

Here, the 'expectation' brackets $\langle \cdot \rangle$ can be understood as an average over infinitely many equivalent trials. Firing rate models work under the assumption that the mean response r to a given stimulus is sufficient to explain the main characteristics of a neural network. Thus, instead of modeling detailed spike sequences for each individual neuron, one only needs to approximate the corresponding rates. The main advantage of this approach is that it often provides the means for an analytical treatment of network dynamics, which are intractable otherwise. On the other hand, rate models might fall short of reliably explaining neural behavior, if the approximation turns out to be too coarse. Therefore, a significant challenge in computational neuroscience is to relate analytically tractable rate models to detailed, microscopic dynamics such as provided by the spike trains of the interacting neurons. In the following, we will briefly sketch how such a mapping can be constructed.
In particular, we are interested in developing a rate model which can qualitatively match the behavior of a LIF model. Firstly, however, consider the case that a rate model has already been established for individual neurons. In this case, sample spikes can easily be generated as realizations of a stochastic point process. Indeed, for sufficiently small time intervals Δ (such that no two spikes can occur within that time frame), the probability that the neuron fires is given by $r \cdot \Delta$. Combined with the (classical) assumption that spikes occur independently from each other, this leads to a simple Poisson process with firing rate $r(t)$. In particular, for a homogeneous Poisson process for which the firing rate r is stationary, the probability of n spikes in any time interval $[0, T]$ is given by

$$P_T(n) = \frac{(rT)^n}{n!} e^{-rT}.$$

Numerous empirical evidence has been collected from brain recordings which indicates that neurons indeed exhibit Poisson-like spiking behavior [25, 26]. Similarly, before fitting a LIF model by a Poisson rate model, one should check that this approximation is justified. A standard statistic to verify (though by no means prove) the Poisson assumption is the coefficient of variation (CV), which describes the irregularity of spike timings. The CV is defined as the ratio of standard deviation and mean of the interspike interval distribution D_{ISI} (i.e. the distribution of the durations between any two subsequent spikes):

$$CV = \frac{\sqrt{\mathrm{Var}(D_{ISI})}}{\langle D_{ISI} \rangle}$$

Characteristically, the CV of a homogeneous Poisson process is equal to one.
If the LIF model satisfies the Poisson assumption, then it can be approximated using a Linear-Nonlinear Poisson (or cascade) model [27, 28, 29]. The cascade model consists of three

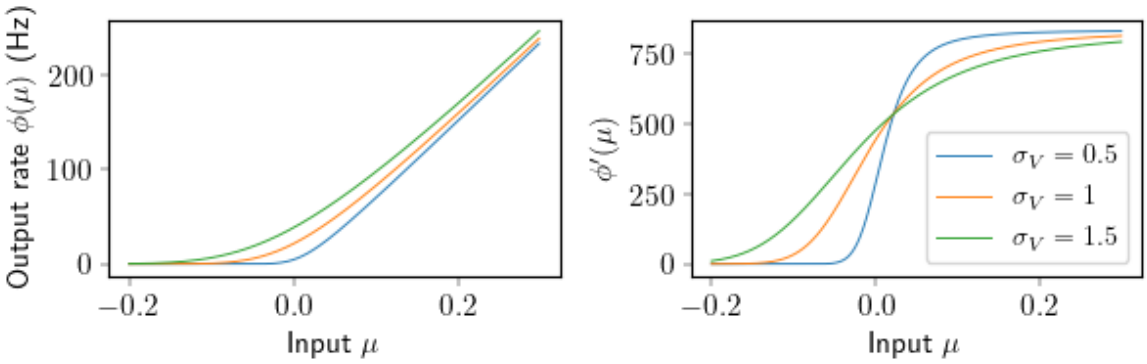

Figure 1: Transfer function ϕ (left panel) and its derivative (right panel) for different values of noise strength σ_V in the LIF model. Parameters: $\vartheta = 1,\ V_{rest} = 0,\ V_R = -0.2$.

consecutive steps [20]. First, a filtered version of the input determines the input potential h. Here, we will use a simple low-pass filter, analogous to the form of the LIF model in Equation (2):

$$\tau_h \dot{h}(t) = -h(t) + I_{syn}(t),$$

where τ_h is the time-scale of the model. Second, input potential h is put through a nonlinearity ϕ, which maps the input potential to a corresponding firing rate $r(t)$. Lastly, as presented above, spikes are generated via an inhomogeneous Poisson process.
The non-linearity ϕ is typically chosen such that for stationary input $I_{syn}(t) = \mu$, the corresponding (steady-state) input-output rate $r = \phi(h) = \phi(\mu)$ correctly describes the firing rate in the original population. In the case of the LIF model, the input-output (or transfer function) ϕ can be derived analytically using the Fokker-Planck equation and is well known from literature [30]. In recent decades, it has been used intensively to study network dynamics in a variety of settings [31, 32]. For the LIF model in the form of Equation (3), the transfer function is given by

$$\phi(\mu) := \phi(\mu; \sigma_V) := \phi_{LIF}(\mu, \sigma_V) := \left(\tau_m \sqrt{\pi} \int_{\frac{\tau_m \mu - \vartheta}{\sqrt{2}\sigma_V}}^{\frac{\tau_m \mu - V_R}{\sqrt{2}\sigma_V}} e^{x^2} \operatorname{erfc}(x) \mathrm{d}x \right)^{-1}. \tag{4}$$

The graph of the function together with its derivative is plotted in Figure 1. The derivative, which will be used later on, is computed as follows.
Let $\psi(x) = e^{x^2} \operatorname{erfc}(x)$ and $\Psi = \int \psi(x) \mathrm{d}x$. Write the integration limits in (4) as $a(\mu) = \frac{\tau_m \mu - \vartheta}{\sqrt{2}\sigma_V}$ and $b(\mu) = \frac{\tau_m \mu - V_R}{\sqrt{2}\sigma_V}$. Then

$$\int_{a(\mu)}^{b(\mu)} \psi(x) dx = \Psi(b(\mu)) - \Psi(a(\mu)).$$

Thus

$$\begin{aligned} \frac{\partial}{\partial \mu} \int_{a(\mu)}^{b(\mu)} \psi(x) dx &= \frac{\partial}{\partial \mu} \Psi(b(\mu)) - \frac{\partial}{\partial \mu} \Psi(a(\mu)) \\ &= b'(\mu)\psi(b(\mu)) - a'(\mu)\psi(a(\mu)) \\ &= \frac{\tau_m}{\sqrt{2}\sigma_V} \left(\psi\left(\frac{\tau_m \mu - V_R}{\sqrt{2}\sigma_V} \right) - \psi\left(\frac{\tau_m \mu - \vartheta}{\sqrt{2}\sigma_V} \right) \right). \end{aligned}$$

Using the chain rule, we obtain

$$\frac{\partial}{\partial\mu}\left(\phi_\sigma(\mu)\right) = -\frac{1}{\sqrt{2\pi}\sigma_V}\left(\psi\left(\frac{\tau_m\mu - V_R}{\sqrt{2}\sigma_V}\right) - \psi\left(\frac{\tau_m\mu - \vartheta}{\sqrt{2}\sigma_V}\right)\right)\left(\int_{\frac{\tau_m\mu-\vartheta}{\sqrt{2}\sigma_V}}^{\frac{\tau_m\mu-V_R}{\sqrt{2}\sigma_V}} \psi(x)\mathrm{d}x\right)^{-2}.$$

3 A functional model for predictive coding

For our analysis, we adopt the predictive coding model introduced by Boerlin and al. [6], capable of approximating an arbitrary linear dynamical system in real-time. This means that the neurons continuously encode an incoming signal, which is then used to perform a dynamic computation. Finally, the processed input results in neuronal spikes, from which the desired output is decoded (or 'predicted').
A key property of the network is that the dynamics are derived from a single functional principle, namely the minimization of a given objective function. In particular, this approach automatically leads to the well-known integrate-and-fire model, without imposing such dynamics artificially.
In this section, we introduce the general model conceptually (using the modified version introduced by Koren et al. [13]) and show how the spiking rule arises from the functional objective. We conclude by illustrating the main characteristics of the network and presenting a simple experimental setup, that we will use throughout this work.

3.1 Encoder-decoder structure and the functional principle of efficient coding

Here and in the remainder of the work, we use bold symbols to distinguish vectors (or matrices), from scalars. Let $\mathbf{s}(t) = (s_1(t), ..., s_M(t))$ be an arbitrary continuous M-dimensional input function. Given $\mathbf{s}$, the target signal $\mathbf{x} = (x_1, \ldots, x_M)$ can be computed by the dynamical equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{s}(t), \tag{5}$$

with A an arbitrary $M \times M$ state transition matrix. Similar to recent works [13, 16], we assume that $\mathbf{A} = -\lambda\mathbf{I_M}$ (where $\mathbf{I_M}$ is the identity matrix) for simplicity, i.e. the signal is obtained by the leaky integration of the input s with $\lambda > 0$:

$$\dot{\mathbf{x}}(t) = -\lambda\mathbf{x}(t) + \mathbf{s}(t), \tag{6}$$

Further, consider a neural network consisting of N neurons, receiving $\mathbf{s}$ as input and generating a readout $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_M)$, which serves as an approximation (or internal representation) of the target signal $\mathbf{x}$. The activity of each neuron i is described by its spike train $y_i(t) = \sum_k \delta\left(t - t_{i,k}\right)$, where δ is the Dirac delta function and $t_{i,k}$ is the time of the k-th spike of neuron i. Based on the spike trains of all neurons combined, i.e. the total activity of the network, the readout is defined (component-wise) as

$$\dot{\hat{x}}_\alpha(t) = -\lambda\hat{x}_\alpha(t) + \sum_{i=1}^{N} w_{\alpha i} y_i(t),\ \alpha = 1, \ldots, M. \tag{7}$$

Here, both the decoding weights $w_{\alpha i}$ and the decay rate $\lambda > 0$ are fixed, such that the output $\hat{x}$ of the network at time t is indeed determined solely by the spike trains of the neurons. In particular, the form of the readout describes a simple leaky integration of the weighted spike trains.

At any time t, the neurons follow a single computational objective - the greedy minimization of the cost function E (also referred to as loss or objective function),

$$E(t) = \sum_{\alpha=1}^{M} \left(\hat{x}_\alpha(t) - x_\alpha(t)\right)^2 + \nu \sum_{i=1}^{N} f_i(t) + \mu \sum_{i=1}^{N} f_i^2(t).$$

We speak of 'greedy' minimization, as only the immediate effect on the cost is evaluated, regardless of the longer-term effects of an emitted spike. The function itself encodes two main principles of efficient coding: Firstly, the 'prediction error' of the readout, described by its squared error $\left(\hat{x}_\alpha(t) - x_\alpha(t)\right)^2$ over all input dimensions, should be as small as possible. At the same time, the total spiking cost of the network should remain minimal as well. This is ensured by the two penalty (or cost) terms $\nu \sum_{i=1}^{N} f_i(t)$ (linear penalty) and $\mu \sum_{i=1}^{N} f_i^2(t)$ (quadratic penalty), with $\mu, \nu > 0$. Here, f_i denotes the low-pass filtered version of the spike train of neuron i:

$$\dot{f}_i(t) = -\lambda f_i(t) + y_i(t), \tag{8}$$

with the same decay rate as for the readout $\hat{\mathbf{x}}$. Thus, the readout components can be written as

$$\hat{x}_\alpha = \sum_i w_{\alpha i} f_i, \tag{9}$$

which is easily verified by differentiating the right-hand side: $\sum_i w_{\alpha i} \dot{f}_i = -\lambda \left(\sum_i w_{\alpha i} f_i\right) + \sum_i w_{\alpha i} y_i$, exactly corresponding to the definition of $\hat{x}$ in (7). Furthermore, f_i is proportional to the firing rate of neuron i, which can be estimated via $\tau_m \dot{r}_i = -r_i + y_i$ with τ_m the membrane time constant of the neuron. Thus, if $\tau_m = 1/\lambda$ (as will be shown in the following section), then

$$r_i = \lambda f_i.$$

In the following, we will use the term firing rate interchangeably for f_i as well, where the factor of proportionality will not play a role.

Both cost terms penalize high firing rates, forcing the network to use spikes in a sparse manner. Additionally, the quadratic cost specifically penalizes high firing rates in individual neurons (relative to all others), forcing the network to distribute the spikes homogeneously across the population [6]. For illustration, consider the case of two equivalent neurons with a fixed total spike amount $f_1 + f_2 = C$. Then the spiking cost is described by $f_1^2 + f_2^2 = f_1^2 + (C - f_1)^2 = 2f_1^2 - 2Cf_1 + C^2$, which is a quadratic function in f_1 with minimum at $f_1 = C/2$, immediately implying $f_2 = C/2$ as well. In other words, the quadratic cost is minimal if the contributions are distributed equally among the neurons.

The overall trade-off between the prediction accuracy of the readout and the spiking cost is controlled by the cost parameters $\mu \geq 0$ and $\nu \geq 0$: For very small parameters, only the mean squared error produced by the network approximation of x will play a significant role in the objective function. In turn, the prediction error can be expected to decrease at the cost of arbitrarily high firing rates. On the other hand, for very high parameters the network will hardly spike at all and therefore the readout will stay close to 0 instead of tracking the signal x. A study of the optimal choice of these model 'hyperparameters' (note the analogy to other optimization problems, e.g. in Machine Learning) was done by Koren et al [13].

3.2 Derivation of the spiking rule

The functional principle of loss minimization gives rise to specific dynamics for each neuron in the network. At any spike time t, the objective function makes a jump of size $\mathrm{d}E(t^+) = E(t^+) - E(t)$, where $E(t^+) = \lim_{s \downarrow t} E(s)$ describes the value of the cost function immediately after the spike. Here, we assume that $E(t)$ is cadlag. Consider the case where a single neuron

k fires at time t. The corresponding effect in the readout $\hat{x}_\alpha$ occurs through the additional spike in the spike train of neuron k, namely $\hat{x}_\alpha(t^+) = \hat{x}_\alpha(t) + \sum_i w_{\alpha i}\delta_{ik} = \hat{x}_\alpha(t) + w_{\alpha k}$, with δ_{ik} the Kronecker delta. Similarly, the filtered spike train of neuron i is updated to $f_i(t^+) = f_i(t) + \delta_{ik}$. Consequently, the jump size is

$$\begin{aligned}
\mathrm{d}E(t^+) &= E(t^+) - E\left(t\right) \\
&= \sum_{\alpha=1}^{M} \left(\hat{x}_\alpha\left(t\right) + w_{\alpha k} - x_\alpha(t)\right)^2 + \nu \sum_{i=1}^{N} \left(f_i\left(t\right) + \delta_{ik}\right) + \mu \sum_{i=1}^{N} \left(f_i\left(t\right) + \delta_{ik}\right)^2 \\
&\quad - \sum_{\alpha=1}^{M} \left(\hat{x}_\alpha\left(t\right) - x_\alpha(t)\right)^2 + \nu \sum_{i=1}^{N} f_i\left(t\right) + \mu \sum_{i=1}^{N} f_i^2\left(t\right) \\
&= 2\sum_{\alpha=1}^{M} w_{\alpha k}\left(\hat{x}_\alpha\left(t\right) - x_\alpha(t)\right) + \sum_{\alpha=1}^{M} w_{\alpha k}^2 + \nu + \mu + 2\mu f_k\left(t\right).
\end{aligned}$$

The greedy minimization rule requires that the spike of neuron k instantaneously decreases the objective function, implying $dE(t) < 0$. Inserting this condition and rearranging the terms, it follows

$$\sum_{\alpha=1}^{M} w_{\alpha k}\left(x_\alpha\left(t\right) - \hat{x}_\alpha(t)\right) - \mu f_k\left(t\right) > \frac{1}{2}\left(\sum_{\alpha=1}^{M} w_{\alpha k}^2 + \nu + \mu\right), \tag{10}$$

where all the time-varying terms (depending on either the signal or the network activity) were collected on the left-hand side, while the right-hand side contains those model parameters which are constant. In turn, the left-hand side can be interpreted as a dynamic current associated with neuron k, which emits a spike if and only if the current surpasses the fixed threshold value given on the right-hand side. Accordingly, we define the membrane potential V_k and threshold ϑ_k of a given neuron k as

$$V_k(t) = \sum_{\alpha=1}^{M} \left(x_\alpha(t) - \hat{x}_\alpha(t)\right) w_{\alpha k} - \mu f_k(t) \tag{11}$$

and

$$\vartheta_k = \frac{1}{2}\left(\sum_{\alpha=1}^{M} w_{\alpha k}^2 + \nu + \mu\right). \tag{12}$$

Thus, Equation (10) can be rewritten as

$$V_k\left(t\right) > \vartheta_k,$$

which holds at spike time of the respective neuron. Using the definitions of x, $\hat{x}$ and f given by the respective differential equations (6), (7) and (8), the time derivative of the membrane

potential is computed as

$$\begin{aligned}
\dot{V}_k &= \sum_{\alpha=1}^{M} \left(\dot{x}_\alpha(t) - \dot{\hat{x}}_\alpha(t)\right) w_{\alpha k} - \mu \dot{f}_k(t) \\
&= \sum_{\alpha=1}^{M} \left(-\lambda x_\alpha(t) + s_\alpha(t) + \lambda \hat{x}_\alpha(t) - \sum_{i=1}^{N} w_{\alpha i} y_i(t)\right) w_{\alpha k} + \mu\lambda f_k(t) - \mu y_k(t) \\
&= -\lambda \sum_{\alpha=1}^{M} \left(x_\alpha(t) - \hat{x}_\alpha(t)\right) w_{\alpha i} - \mu f_k(t) \\
&\qquad + \sum_{\alpha=1}^{M} w_{\alpha k} s_\alpha(t) - \sum_{i=1}^{N}\sum_{\alpha=1}^{M} w_{\alpha k} w_{\alpha i} y_i(t) - \mu y_k(t)) \\
&= -\lambda V_k + \sum_{\alpha=1}^{M} w_{\alpha k} s_\alpha(t) - \sum_{i=1}^{N}\sum_{\alpha=1}^{M} w_{\alpha k} w_{\alpha i} y_i(t) - \mu y_k(t). \qquad (13)
\end{aligned}$$

Combined with the spiking rule (and an intrinsic, 'hidden' reset, as explained below), this constitutes a leaky integrate-and-fire model and each term in (13) has a distinct, easily interpretable role. Clearly, $-\lambda V_k$ is the leakage, with the inverse of λ determining the membrane time constant: $\tau_m = 1/\lambda$. The second term, $\sum_\alpha w_{\alpha k} s_\alpha$, is the external input (or stimulus), which is integrated differently by each neuron k depending on its weights $\mathbf{w_k} = (w_{1k}, \ldots, w_{Mk})$. These determine whether the effective current into the neuron is excitatory or inhibitory, as well as its strength.

At the same time, the weights determine both structure of the synaptic connectivity and the magnitude of recurrent feedback, described by the third term in (13). Indeed, as long as none of the neurons fire, all spike trains (and thus the whole term) remain at zero. On the other hand, if neuron i spikes, the instantaneous postsynaptic effect to neuron k is given by $\mathbf{w_i}^T \mathbf{w_k} = \sum_\alpha w_{\alpha k} w_{\alpha i}$. Note that this leads to an all-to-all connected network, including an inhibitory autapse of strength $\|\mathbf{w_i}\|^2$. The structure is summarized in the connectivity matrix $-\mathbf{w^T w}$, where $\mathbf{w}$ is the $M \times N$ matrix of all weights $w_{\alpha i}$. Each entry $-(\mathbf{w^T w})_{ij}$ of the connectivity matrix corresponds to the postsynaptic current which neuron i exerts on neuron j at spike time:

$$-\mathbf{w^T w} = \begin{pmatrix} -\|\mathbf{w_1}\|^2 & \cdots & -\mathbf{w_1}^T \mathbf{w_N} \\ \vdots & \ddots & \vdots \\ -\mathbf{w_N}^T \mathbf{w_1} & \cdots & -\|\mathbf{w_N}\|^2 \end{pmatrix}.$$

Similarly, the last term describes a 'hidden' self-reset mechanism: Suppose that neuron k evolves according to the dynamic equation (13) and eventually hits the threshold ϑ. Due to the spiking rule, it immediately fires a spike and in turn, triggers a delta spike and corresponding jump of $-\mu$ downward in its membrane potential. Combined with the inhibitory self-connection from the recurrent term, this yields the resulting reset potential

$$V_R^{(k)} = \vartheta - \|\mathbf{w_k}\|^2 - \mu. \qquad (14)$$

Finally, we add a stochastic noise term for biological plausibility. Physiologically, it accounts for multiple sources of perturbations such as synaptic noise, failures, background activity transmitted from other brain areas, etc. The noise is modeled in a standard way via additive Gaussian white noise η, with autocorrelation $\langle \eta(t)\eta(t')\rangle = \delta(t-t)$ and corresponding intensity σ_V:

$$\dot{V}_k = -\lambda V_k + \sum_{\alpha=1}^{M} w_{\alpha k} s_\alpha(t) - \sum_{i=1}^{N}\sum_{\alpha=1}^{M} w_{\alpha k} w_{\alpha i} y_i(t) - \mu y_k(t) + \sqrt{2\lambda}\sigma_V \eta_k. \qquad (15)$$

This completes the leaky integrate-and-fire dynamics in their usual form. Using vector notation, the full model can be summarized as

$$\dot{\mathbf{V}} = -\lambda \mathbf{V} + \mathbf{w}^T \mathbf{s} - \mathbf{w}^T \mathbf{w} \mathbf{y} - \mu \mathbf{y} + \sqrt{2\lambda} \sigma_V \eta. \tag{16}$$

Note that the weights, although arising from a purely functional role in the decoder $\hat{x}$, play a crucial role in the intrinsic structure of the model and correspond to (largely) immutable morphological and physiological features of the biological neuron. In particular, both the number and magnitude of the weights, determining the synaptic effects for each neuron, are fixed and independent of any given stimulus.

As a consequence, the dimensionality M can be interpreted as the internal capacity of the network - the maximum number of scalar signals (or equivalently, maximal target dimensionality) which can be tracked by the network at any given time [3]. The case where the number of relevant features M' is smaller than the capacity M can be seamlessly incorporated into the general model. Indeed, the absence of any stimulus corresponds to zero external input in the LIF dynamics. Thus, any M'-dimensional input is equivalent to an M-dimensional signal with $M - M'$ input features set to zero. In Section 6, we will show that the dynamics of the model are largely unaffected by these 'dummy' inputs, yielding the same performance for the relevant M' features for sufficiently large networks.

Therefore, without loss of generality, we can assume $M' = M$, which will be used for convenience except where stated otherwise.

3.3 Main characteristics of the model

The specifics of the network behavior can readily be understood using a toy model with just two neurons n1 and n2, no noise, and a single (i.e. one-dimensional) external input drive s. This illustrative approach was inspired by a similar example used by Koren and Denève [13]. Corresponding to the input s, both the signal x and the readout $\hat{x}$ are one-dimensional and each of the neurons has a single readout weight associated with it. Suppose that n1 has positive weight $w_1 = J > 0$, while n2 has the opposite weight $w_2 = -J$.

We consider how the network evolves in the presence of a simple step current: First, the network is at rest, i.e. $s = 0$, before being stimulated by a constant positive input. The results are shown in Figure 2 (left panel). During the time where there is no external input (first 30ms), the network activity remains silent: the membrane potentials of the two neurons, as well as the readout, stay at 0. However, as soon as the exciting stimulus current appears, it drives the membrane potentials into opposite directions, prescribed by the sign (or 'direction') of their weight.

In our example, 'plus'-neuron n1 gets excited while 'minus'-neuron n2 is inhibited due to the positive input signal. This illustrates a central feature of the model, prescribed in the definition of the membrane equation (11). Namely, all neurons track (in their membrane potential) the prediction error $x - \hat{x}$ produced by the readout [6]. More precisely, they track the projection of the prediction error onto the direction of their weights. In general, a neuron possesses a separate weight for each dimension of the encoded signal, which allows for arbitrary directions in the M-dimensional space. In the toy model, the weights are simply given by $+J$ and $-J$, leading to projections $+J(x - \hat{x})$ for n1 and $-J(x - \hat{x})$ for n2.

As soon as the projected prediction error for n1 gets large enough (this is the moment when the threshold is reached), the neuron fires a spike, driving the readout towards the true signal. Indeed, the readout makes an instantaneous jump of size $w_1 = J$ in the direction of the true signal. Conversely, in the case of a negative signal, the neurons would switch roles and n2 would get excited, driving the estimate down. Thus, each neuron possesses a selectivity based on its weights, and - as prescribed by the functional objective - only the units best suited for optimal coding are stimulated for any given input. Moreover, the neurons whose spike would increase the error are automatically driven away from the threshold.

Each spike is transmitted via instantaneous delta connections to all neurons in the network, including the neuron which fired the spike itself. This constitutes the recurrent feedback in the predictive coding model. The structure and strength of the synaptic connections are described by the connectivity matrix $-\mathbf{w^T w}$. In the one-dimensional case, it is given by $(-\mathbf{w^T w})_{ij} = -w_i \cdot w_j$, yielding

$$-\mathbf{w^T w} = \begin{pmatrix} -w_1^2 & -w_1 w_2 \\ -w_2 w_1 & -w_2^2 \end{pmatrix} = \begin{pmatrix} -J^2 & J^2 \\ J^2 & -J^2 \end{pmatrix}.$$

Each entry Ω_{ij} describes how a spike of the neuron j affects neuron i. Consequently, the recurrent connections are determined only by the individual weights of the pre- and postsynaptic neurons. The spike of the 'plus'-neuron n1 causes an upward jump in the membrane potential of the 'minus'-neuron n2 while contributing a decrease of the same amount to its own potential. Indeed, the reset value after any spike time t_1 of n1 is given by $V_1(t_1^+) = \vartheta - \mu - w_1^2$, where the quadratic cost parameter μ determines the fixed part of the reset for all neurons, while the additional decrease by w_1^2 comes from the inhibitory self-connection.
Due to the negation of the pairwise products in the connectivity matrix, neurons with similar selectivity generally inhibit each other, while oppositely weighted neurons excite each other. Analogously, in the general multi-dimensional case $M > 1$ the matrix consists of pairwise scalar products

$$-\mathbf{w^T w} = \begin{pmatrix} -\|\mathbf{w_1}\|^2 & \cdots & -\mathbf{w_1}^T\mathbf{w_N} \\ \vdots & \ddots & \vdots \\ -\mathbf{w_N}^T\mathbf{w_1} & \cdots & -\|\mathbf{w_N}\|^2 \end{pmatrix}$$

and two neurons are said to have similar selectivity if their scalar product is positive. Assuming that the number of neurons N significantly exceeds the dimensionality of the target signal M, many neurons will necessarily share a similar selectivity. In that case, the recurrent mechanism implements a competition for each population of similar neurons. As soon as the first neuron i reaches the threshold and spikes, the other 'competing' neurons with weights $w_j \approx w_i$ are 'set back' by an amount $\mathbf{w_i}^T\mathbf{w_j} \approx \|\mathbf{w_i}\|^2$ [6]. Simultaneously, the spike exerts an effect on the prediction error, pushing the readout into the direction of its weight. In turn, the recurrent connections realize a form of 'communication' [13] about the state of computation: The instantaneous, inhibitory feedback prevents that too many similar neurons spike at once - all pushing into the same direction and causing the readout to overshoot the signal.
Similarly, as described in detail by Koren and Denève [13], the recurrent connections incorporate a mechanism of 'error correction'. For instance, suppose the same setting as in the toy model, but taking into account some level of noise as described by equation (15). If the noise level is high enough, it may eventually lead to an erroneous spike of n2. Due to its excitatory postsynaptic effect on n1, however, this coding error could quickly be corrected by a spike of the latter. In the same manner, the network can swiftly respond to sudden changes in the evolution of the signal (e.g. if it's suddenly decreasing) [6]. From the biological point of view, this duality in the postsynaptic effect of each neuron seems implausible. In particular, it violates Dale's law by which neurons are either purely excitatory or purely inhibitory [33]. However, this issue can be bypassed by introducing two separate objective functions for excitatory and inhibitory neurons [6].
After the first spike, the model continues to evolve the same manner (as long as the external input stays the same), with n1 spiking as soon as the prediction error again hits the critical threshold. After an initial transient period, the signal x, as well as the activity of the network, becomes stationary, with periodic spiking of neuron n1. This corresponds to the leaky integrate-and-fire model without noise in the stationary regime. From the computational point of view, this leads (averaging over time after the initial transient) to a fixed bias in the estimate of the true signal: $\langle x - \hat{x} \rangle = x - \langle \hat{x} \rangle = const.$

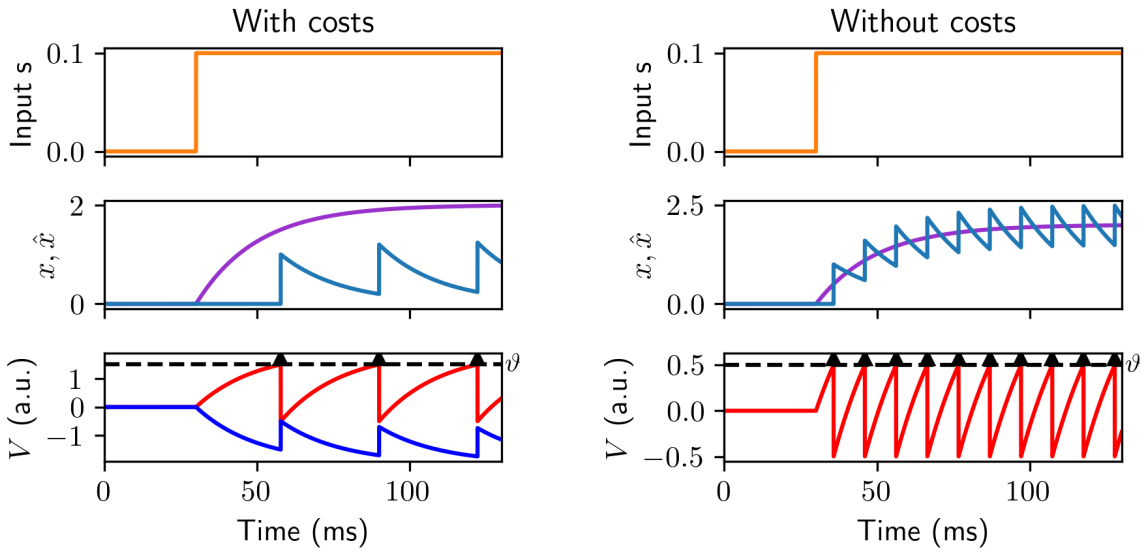

Figure 2: Illustration of a toy model with and without costs. Left: Toy model with two neurons and non-zero costs $\mu = \nu = 1$. Right: Model with one neuron and zero costs. In the middle panels, the readout (blue) is plotted together with the true signal (purple). The lower plots display the temporal evolution of the membrane potentials (red: positive weight, blue: negative weight). Black triangles indicate spikes. In the case without costs, the readout fluctuates around the true signal, spiking as soon as the error becomes larger than half of its weight $w = 1$. On the other hand, for $\mu = \nu = 1$ the threshold is increased by $(\mu + \nu)/2 = 1$, yielding a bias of the same magnitude. Simulation parameters: A Euler method with a time step Δ=0.01ms is used throughout our simulations. Moreover, the time constant is fixed to $\tau_m = 20$ms. Except where stated otherwise, the costs are set to $\mu = \nu = 1$. Other parameters here: $J = 1$.

This bias arises automatically for non-zero cost parameters. Indeed, if both μ and ν are zero, the membrane potential is driven purely by the prediction error $w_i(x - \hat{x})$ and $\hat{x}$ jumps by the amount J each time when n1 reaches the threshold: $V_1(t) = J(x - \hat{x}) = \vartheta = J^2/2$, or equivalently $x - \hat{x} > J/2$. This is illustrated in the right panel of Figure 2 for a network consisting of just a single neuron $n1'$ with positive weight $J = 1$ and the same input as before. The readout fluctuates around the true signal, spiking as soon as the error reaches half of its decoding weight. However, as soon as additional costs are introduced, the threshold gets higher without any change in the evolution of the membrane potential. In effect,
The neuron spikes much more often than in the case regularized by non-zero cost parameters $\mu = \nu = 1$. Moreover, adding a neuron $n2'$ with negative weight of the same strength would lead to the so-called 'ping-pong' effect: As soon as $n1'$ reaches the threshold $\vartheta' = J^2/2$ and spikes, it initiates an instantaneous jump of $n2'$ by $J^2 = 1$ upward. As the evolution of the membrane potentials is symmetric for both neurons (i.e. immediately before the spike, $n2'$ had a potential of $-J^2/2$), they switch roles and $n2'$ is a threshold and spikes, again triggering a spike of $n1'$, etc. This shows that the costs are an essential part of the model not only to ensure biological plausibility but also from a computational point of view.

3.4 Sample case: One-dimensional model with binary weights

In theory, a single neuron would suffice to track the signal with arbitrary precision. Indeed, if the cost parameters are set to 0, the readout in the toy model presented in 3.3 fluctuates around the true signal with a maximal deviation equal to half of the weight size. Thus, to achieve arbitrarily high precision, it suffices to make the weights sufficiently small. However,

as one might expect, the toy model (or very small networks in general) is far too limited to give a meaningful account of the computations performed in the brain. Fast increments of small magnitude, which minimize the mean-squared error of $\hat{x}$, rely on a large number of consecutive spikes from the same population. In sparse models, this necessarily leads to excessively high firing rates of the individual neurons, irreconcilable with biology. At the same time, the spiking activity is much more predictable compared to the highly irregular, Poisson-like nature of neuronal spikes recorded in the brain. Furthermore, very small networks are highly susceptible to failures and malfunctions of the few involved neurons.
In the following, we generalize the toy model to a more realistic setting where larger populations of neurons interact to efficiently encode a signal in the presence of disruptive noise. This will serve as the standard example for illustration throughout this work. Similar to the toy model, we consider the one-dimensional case and binary weights

$$w_i = J\xi_i, \quad \xi_i \in \{-1, 1\}, J > 0$$

in which half of the weights are positive and the other half negative: $P(w_i = J) = P(w_i = -J) = 0.5$. This implies two commensurate populations of equivalent neurons: one with uniform weight J and corresponding positive selectivity, and one which is exactly opposite. Where the distinction might be of importance, we will refer to the stochastic realizations ξ_i as patterns, opposed to the weights $w_i = J\xi_i$. Further, as described in Equation (15), each neuron receives external white noise with intensity $\sigma_V > 0$.
The cost parameters are set such that the fixed (i.e. weight-independent) part of the threshold, given by $(\mu + \nu)/2$ is equal to one. Further, we impose that the linear and quadratic cost parameters should be equal, which has been shown to yield optimal results in previous studies [13]. This implies $\mu = \nu = 1$. In conclusion, the model is specified by membrane equation

$$\dot{V}_i = -\lambda V_i + J\xi_i s_\alpha(t) - \sum_{i=1}^{N} J^2 \xi_i \xi_i y_i(t) - y_i(t) + \sqrt{2\lambda}\sigma_V$$

and uniform threshold

$$\vartheta = \vartheta_i = 1 + \frac{J^2}{2}. \tag{17}$$

Interestingly, for larger network sizes the noise is not only a plausible physiological addition but also a necessity from the computational perspective: Otherwise, in any quiescent period with absent stimuli, the membrane voltage of all neurons would quickly decay to 0, sharing the exact same state. This is equivalent to a uniform initial condition in the defining equation (13), leading to perfect synchronization within each distinct population. As discussed in depth by Koren et al. [13], a significant amount of synchronization poses one of the main computational problems in the given predictive coding model and can lead to chaotic behavior [16].
The result in the case of $N = 100$ neurons (which can still be considered as a very small network) for both stationary and time-varying input are displayed in Figure 3. In both cases, the network exhibits asynchronous spiking with the readout closely trailing the true signal, albeit with a bias due to the non-zero cost parameters.
The time-varying input provides a good illustration of how the two populations are activated in alternation, depending on whether the signal is increasing or decreasing (note, however, the small time lag in the readout). Despite only one population being driven at each given time and the mean readout remaining very stable close to the true signal, spiking activity is highly irregular. Firing is asynchronous and homogeneously distributed across the network. The CV, measured over a long period of constant activity and averaged over the population, is found to be very close to 1 (0.96), suggesting Poisson-like behavior similar to experimental recordings.

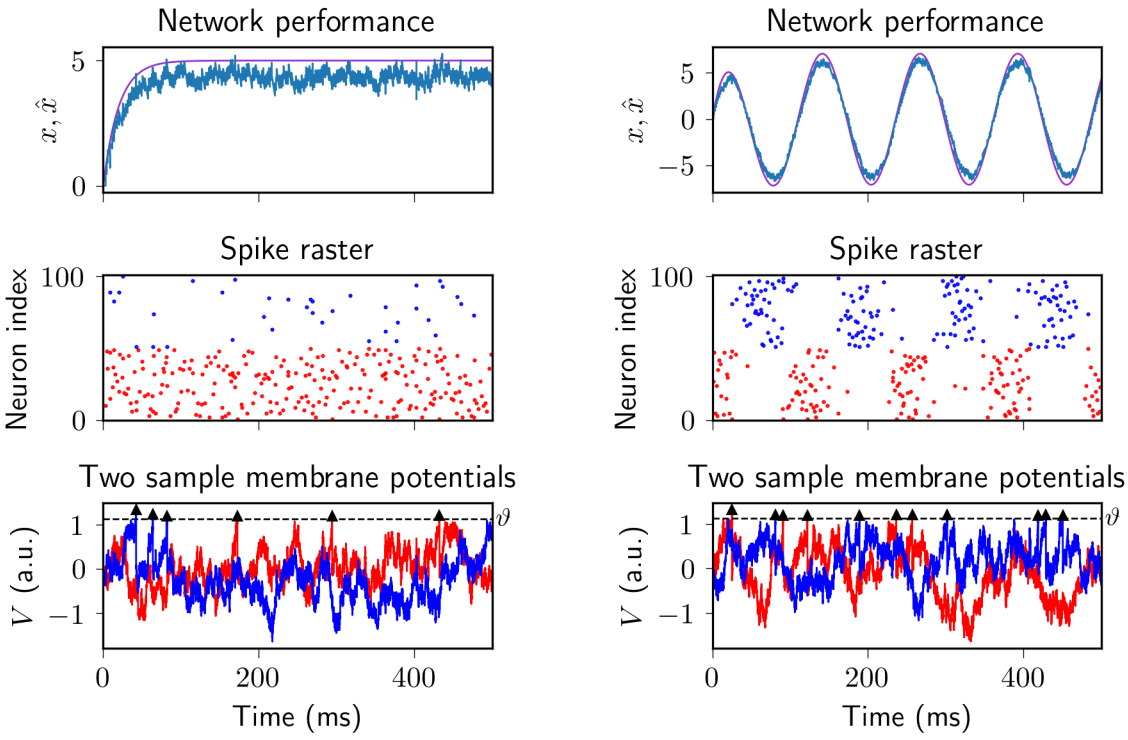

Figure 3: Visualizing the behavior of the model using a one-dimensional model with binary weights. Left: results for a constant input $s = 0.25$, leading to stationary activity. Right: Network performance for a sinusoidal input. In the middle panels (spike rasters), blue dots correspond to neurons with negative weight, while red dots visualize positive weighted neurons. The same color scheme is used in the lower plot, where one neuron of each population was picked randomly. Triangles correspond to spikes of the respective neuron, triggered as the membrane potential passes the threshold. Note that due to instantaneous delta connections, the membrane potential can jump over the threshold, immediately releasing a spike. This can also lead to synchronization if multiple neurons are close to the threshold potential and then jump simultaneously above it due to excitatory postsynaptic feedback. Parameters: $N = 100,\ J = 0.5,\ \sigma_V = 0.5$.

This interplay of high, reliable performance and strong variability is achieved through a tight balance of excitation and inhibition - a characteristic qualitative feature of the predictive coding model [6] supported by biological evidence [34]. Strong driving inputs are counteracted by an equal (or greater) amount of lateral inhibition. Any surplus in excitation leading to excessive spikes immediately triggers negative feedback (as discussed above), yielding a tight temporal correlation that controls the output. The dominance of lateral inhibition manifests itself in the mean membrane potential, which is pushed downward to concentrate at subthreshold values. As a consequence, spikes are triggered by random noise - a state which is known as fluctuation-driven regime. In effect, although membrane potentials are correlated due to identical feed-forward drive across the populations, the resulting spikes are uncorrelated and highly variable.
In conclusion, the simple one-dimensional model encapsulates both the power and the main properties of the network. The results can easily be extended, as the same behavior was demonstrated for more intricate and biologically relevant examples such as multidimensional sensory tracking or 2D arm control [6, 13]. In the following, we will study the model in a generalized setting and provide an analytical treatment, while continually comparing the results with empirical simulations.

4 General framework with free scaling parameters

4.1 Generalized model

Ultimately, our goal is to study the dynamics in the mean-field limit of infinitely large networks. Hence we have to ensure that the main characteristics and behavior of the model do not change for a growing number of neurons N. On the other hand, the question about how networks should be scaled with N is not clear in general. For instance, a highly debated question in computational neuroscience is the level of balance in which neural circuits operate. Here, 'balance' refers to the mutual level of excitation and inhibition which drives network activity. Typically, one distinguishes between two main types of cortical balance [17, 18]: Loose balance, where excitation and inhibition currents are of the same order as the net resulting input, and tight balance where both excitation and inhibition are large but cancel each other out to yield a similarly moderate current.
Assume the classical setting where excitation is mainly produced by external feedforward input while being stabilized by lateral inhibition (as is the case in the predictive coding model). In this case, loose balance corresponds to both external and recurrent inputs scale with $O(1)$, while tight balance refers to the case where both scale as $O(N)$. Additionally, an intermediate scaling of $O(\sqrt{N})$ is known under the term 'classical' [35, 16] scaling, honoring the pioneering theoretical works on the problem [4, 5].
So far, the spiking predictive coding was only analyzed under the assumption of tight balance [6]. However, recent findings suggest that the cortex might much rather operate under a loosely balanced regime [18]. Here, inspired by the work of Kadmon et al. in the context of rate models [16], we present a general framework that can be universally applied in the context of the predictive coding model and which is consistent with previously formulated scaling rules. Moreover, it allows to dynamically incorporate and compare different assumptions, such as the level of balance between excitation and inhibition in the model.
For full generality, we consider all network parameters $w_{\alpha k}, \nu, \mu$ to have independently adjustable scaling factors. For that matter, we write

$$w_{\alpha k} = b_N \xi_{\alpha k}, \quad \nu = c_N \hat{\nu}, \quad \mu = d_N \hat{\mu},$$

where the scaling factors b_N, c_N and d_N can depend on N while $\xi_{\alpha k}, \hat{\nu}$ and $\hat{\mu}$ are of order 1 (i.e. are constant). Moreover, the threshold rule (10) itself can be scaled by a factor a_N on both sides. Incorporating the factors into Equations (11) and (12) of membrane potential and threshold, respectively, yields

$$V_k(t) = a_N \left(b_N \sum_{\alpha=1}^{M} \left(x_\alpha(t) - \hat{x}_\alpha(t) \right) \xi_{\alpha k} - d_N \hat{\mu} f_k(t) \right),$$
$$\vartheta_k = \frac{a_N}{2} \left(b_N^2 \sum_{\alpha} \xi_{\alpha k}^2 + c_N \hat{\nu} + d_N \hat{\mu} \right).$$

Accordingly, the time derivative of the membrane potential is

$$\dot{V} = -\lambda V_k + a_N \left(b_N \sum_{\alpha} \xi_{\alpha k} s_\alpha(t) - b_N^2 \sum_{i=1} \sum_{\alpha} \xi_{\alpha k} y_i(t) - d_N \hat{\mu} y_k(t) \right).$$

In the original model introduced by Boerlin et al. [6], both the firing rates and the readout are assumed to remain constant on average in the limit of large networks. This was achieved by scaling the synaptic weights w_i as 1/N, while simultaneously scaling cost parameters with $1/N^2$. Scaling of the firing rule itself was not considered explicitly but occurs implicitly

through normalization over the norm of the weights to obtain realistic values for the membrane potential (i.e. of order 1). In the terms introduced above, this setup can be expressed by setting

$$b_N = \frac{1}{N}, \quad c_N = d_N = \frac{1}{N^2}, \quad a_N = \frac{1}{b_N^2} = N^2.$$

While the approach of Boerlin et al. is certainly reasonable in its own right, there are a few limitations that arise as a consequence: Firstly, the authors assume that the costs are negligibly small, such that the resulting bias in the readout prediction is hardly noticeable. As a result, the main contribution to both threshold and reset is made up by the sum of squared weights $\sum_\alpha w_{\alpha i}^2$, which simultaneously determines the strength of the postsynaptic potential emitted by the corresponding neuron. Although mean amplitudes of latent transmissions vary largely in cortical recordings, a typical postsynaptic current is far too weak to trigger an action potential on its own [36]. A more reasonable assumption is that both threshold and reset are determined mainly by a fixed amount (i.e. by the cost parameters) which is significantly higher than a regular postsynaptic impulse. This also ensures that threshold and reset are comparable for all neurons, such that they can easily be rescaled to realistic physiological values.
Moreover, the scaling imposes a tight balance as both feedforward input $a_N b_N \sum_\alpha \xi_{\alpha k} s_\alpha(t)$ and lateral feedback $a_N b_N^2 \sum_{i=1}^N \sum_\alpha \xi_{\alpha k} y_i(t)$ to each neuron are of order N. In the following, we derive scaling conditions that resolve these issues by allowing a more flexible approach for both the coding of the error and the level of synaptic balance.

4.2 Derivation of scaling rules for the large N limit

To find the appropriate scaling as the number of neurons N tends to infinity, we demand three requirements. Firstly, the threshold should remain (roughly) constant as it corresponds to a fixed biological feature in the given LIF model: The threshold equation thus implies that

$$a_N = O\left(\frac{1}{b_N^2 + c_N + d_N}\right), \tag{18}$$

as $\xi_{\alpha k}, \hat{\nu}$ and $\hat{\mu}$ are of order 1 and thus $b_N^2 \sum_\alpha \xi_{\alpha k}^2 + c_N \hat{\nu} + d_N \hat{\mu} = O(b_N^2 + c_N + d_N)$. Secondly, the mean membrane potential should remain of order 1 as well. This implies

$$b_N \sum_{\alpha=1}^M \left(x_\alpha(t) - \hat{x}_\alpha(t)\right) \xi_{\alpha k} - d_N \hat{\mu} f_k(t) = O(a_N^{-1}), \tag{19}$$

which is satisfied in general if

$$\sum_{\alpha=1}^M \left(x_\alpha(t) - \hat{x}_\alpha(t)\right) \xi_{\alpha k} = O\left(b_N^{-1} a_N^{-1}\right) \quad \text{and} \quad d_N f_k = O\left(a_N^{-1}\right). \tag{20}$$

In the stationary case, where the target signal x is constant, we can reasonably assume that the mean readout $\langle \hat{x} \rangle$ is constant as well. Correspondingly, we assume a well-behaved firing activity with the mean rate $\langle f_k \rangle$ remaining constant. This immediately implies $d_N = O\left(a_N^{-1}\right)$, i.e. without loss of generality we can set $d_N = a_N^{-1}$.
Moreover, the mean prediction error $\langle \hat{x}_\alpha \rangle - x_\alpha$ must be of order 1 or below: Either the readout will perfectly track the signal in the limit, or there will be a fixed bias. Thus, from the two conditions in (20) we conclude

$$b_N^{-1} a_N^{-1} = O(1) \quad \text{and} \quad d_N = a_N^{-1}. \tag{21}$$

Note that the big O here is a slight abuse of notation (though technically correct by definition), as we still allow the case that both terms vanish in the limit (i.e. that both scale as $o(1)$). See below for a more conclusive discussion.
Without loss of generality, we can choose the parameters such that Relation (18) holds as a direct equality. Thus,

$$b_N^{-1}a_N^{-1} = \frac{b_N^2 + c_N + d_N}{b_N} = O(1),$$

yielding the conditions

$$b_N = O(1) \quad \text{and} \quad c_N + d_N = O(b_N). \tag{22}$$

The third requirement is that the variance of the intrinsic fluctuations (i.e. which are not induced by external noise (described by Equation (11)) in the network vanishes, which we denote as $Var(V_k) = o(1)$, slightly abusing notation. Note that using the explicit form of the readout given by Equation (9), the membrane potential can be written as

$$\begin{aligned} V_k(t) &= a_N \left(b_N \sum_\alpha \left(x_\alpha - \hat{x}_\alpha(t) \right) \xi_{\alpha k} - d_N \hat{\mu} f_k(t) \right) \\ &= a_N \left(b_N \sum_\alpha \left(x_\alpha - b_N \sum_i f_i(t) \right) \xi_{\alpha k} - d_N \hat{\mu} f_k(t) \right) \end{aligned}$$

Based on the results of Boerlin et al [6], the spike trains y_i of the predictive coding model can be assumed to be uncorrelated and follow a Poisson point process with some intensity r_i. In turn, the variance of V_k can be approximated as

$$Var(V_k) \approx a_N^2 b_N^4 \sum_\alpha \sum_i \xi_{\alpha k}^2 Var(f_i) + a_N^2 d_N^2 \hat{\mu} Var(f_k) - 2 \sum_\alpha a_N^2 b_N^2 d_N \xi_{\alpha k} \hat{\mu} Var(f_k),$$

which can readily be seen to be of order $O(a_N^2 b_N^4 \sum_i Var(f_i))$ using the conditions (21) and (22) derived above and neglecting any contributions of order 1. For simplicity, consider again the stationary case. Then it is clear that the variance of the filtered Poisson spike train f_i is bounded (more precisely, proportional to the respective intensity r_i), yielding $Var(V_i) = O(a_N^2 b_N^4 N)$. Thus, the variance vanishes in the limit of large networks if and only if $a_N^2 b_N^4 N = o(1)$. In turn, $b_N^4 N$ has to scale as $o(a_N^{-2})$. Due to $d_N = a_N^{-1}$, it follows $b_N^4 N = o\left(a_N^{-2}\right) = o(d_N^2)$ and thus

$$b_N^2 \sqrt{N} = o(d_N). \tag{23}$$

Furthermore, due to the second equation in (22) it holds $d_N = O(b_N)$. Thus, $b_N^2 \sqrt{N} = o\left(d_N\right) = o(b_N)$, yielding that the weights have to converge to 0 faster than $1/\sqrt{N}$, further tightening the restriction given in (22):

$$b_N = o\left(\frac{1}{\sqrt{N}} \right). \tag{24}$$

Equation (23) also implies $b_N^2 = o(d_N)$, from which it follows

$$\begin{aligned} a_N &= O\left(\frac{1}{b_N^2 + c_N + d_N} \right) \\ &= O\left(\frac{1}{c_N + d_N} \right) \end{aligned}$$

which is fulfilled for $c_N = O(d_N)$.
Without loss of generality (except for some degree of freedom in decreasing c_N) we can thus set

$$c_N = d_N = a_N^{-1}. \tag{25}$$

To obtain the corresponding scaling factor b_N for the weights, we distinguish two cases. Firstly, consider the case where the mean prediction error $x - \langle \hat{x} \rangle$ converges to 0. Then the first relation in (21) reads $a_N^{-1} = o(b_N)$. In that case, the necessary scaling conditions are given by

$$c_N = d_N = a_N^{-1} = o(b_N) \quad \text{and} \quad b_N^2 \sqrt{N} = o(a_N^{-1}), \tag{26}$$

which implies (but is not equivalent to) $b_N = O(1/\sqrt{N})$, as we have seen above.
On the other hand, if there is a nonzero bias in the readout error, it follows that $a_N = O(b_N)$, i.e. we can set $a_N = b_N^{-1}$ and the scaling requirements are

$$c_N = d_N = a_N^{-1} = b_N \quad \text{and} \quad b_N = o\left(\frac{1}{\sqrt{N}}\right), \tag{27}$$

which automatically ensures $b_N^2 \sqrt{N} = o(a_N^{-1})$ as well.
Leaving room for freedom in the relation between a_N and b_N, the resulting LIF dynamics (including additive noise) are given by

$$\dot{V}_i = -\lambda V_i + a_N b_N \sum_\alpha \xi_{\alpha i} s_\alpha(t) - a_N b_N^2 \sum_{i=1} \sum_\alpha \xi_{\alpha i} y_i(t) - \mu y_i(t) + \sqrt{2\lambda} \sigma_V \eta$$

and

$$\vartheta_i = \frac{a_N b_N^2}{2} \sum_\alpha \xi_{\alpha i}^2 + \nu + \mu.$$

If b_N is chosen as $1/N$, then (taking into account the normalization by $1/N$ for the sum over all neurons in the recurrent term) the level of balance is fully specified by $a_N b_N = a_N/N$. The first scenario (26) leads to the condition $a_N = N^\chi$, with lower bound $\chi > 1$ due to $b_N^{-1} = o(a_N)$ and upper bound $\chi < 1.5$ due to $a_N = o((b_N^2 \sqrt{N})^{-1})$. Thus, the level of balance a_N/N lies above $O(1)$ but below $O(\sqrt{N})$, exactly corresponding to the range where balance is looser than classical. Similarly, the second scenario (27) with a fixed bias leads to $a_N b_N = 1$, again yielding a loosely balanced network.
The third scaling requirement can be replaced by a similar condition that the variance should remain of order 1. Then $a_N^2 b_N^4 N$ has to be of order 1 (instead of being $o(1)$, as derived above). Retracing the same steps as before but replacing the little Landau o by a big O, leads to $b_N^2 \sqrt{N} = O(d_N)$. In the case of a fixed bias, where d_N is of the same order as b_N (see Equation 21), this implies $b_N = O(1/\sqrt{N})$. As a consequence, the feedforward input would scale as $a_N \sqrt{N}$, while the lateral term would only scale with a_N. The result is an unbalanced regime, which might corrupt both readout performance and the characteristic variability of the network. Thus, we only consider the case where the readout converges to the true signal, $d_N = o(b_N)$. This yields $b_N = o(1/\sqrt{N})$ as in Equation (24). Using the same subsequent steps as before, the resulting dynamics are

$$c_N = d_N = a_N^{-1} = b_N^2 \sqrt{N}.$$

Setting $b_N = 1/N$, it follows $a_N b_N = O(\sqrt{N})$, yielding classical balance. Analogously, allowing the variance to scale (formally, by ignoring the threshold) as $O(N)$ yields $a_N b_N^2 = O(1)$, corresponding to tight balance for $b_N = 1/N$ (again only considering the case with vanishing readout bias).
In conclusion, all three regimes can be easily incorporated and compared within the introduced framework. Moreover, we established an explicit connection to the behavior of the readout bias. While for loose balance, this behavior can be freely adjusted, the other regimes dictate a vanishing bias with increasing network size. An illustration is given in Figure 4, where the vanishing bias in a regime with classical balance (left panels) is contrasted to a

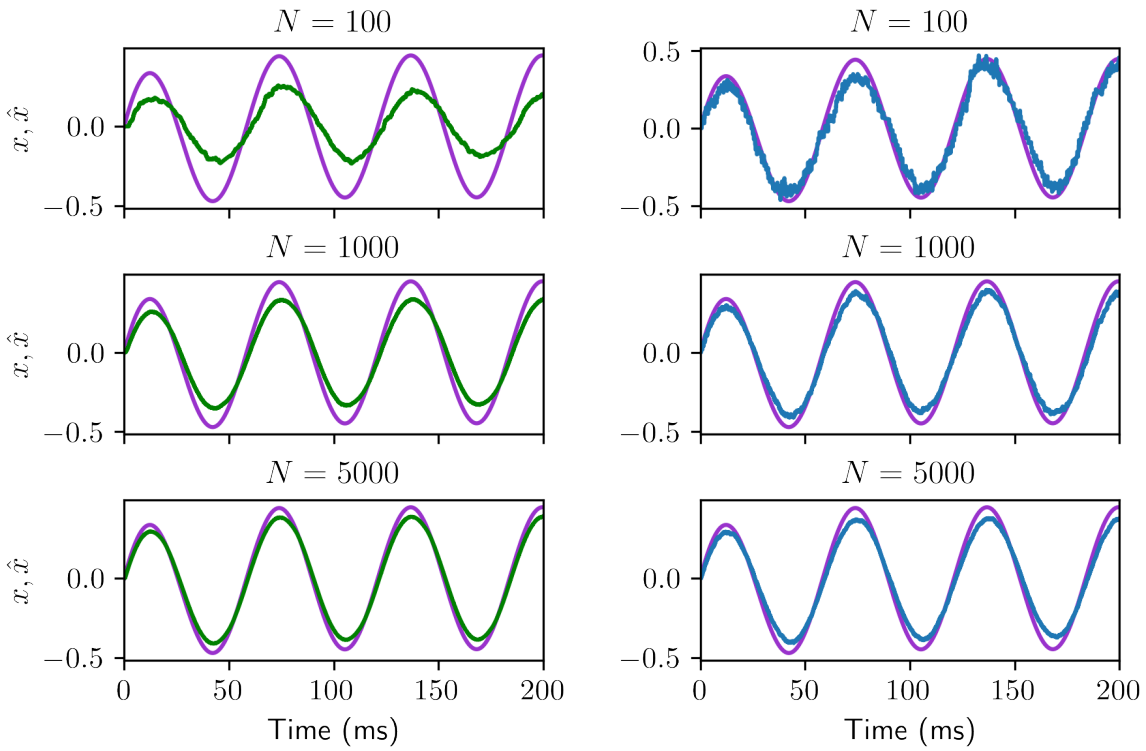

Figure 4: Evolution of readout performance with increasing network size in two different scaling regimes derived from the generalized framework. Left panels: Classical balance and vanishing readout bias, obtained by scaling weights with $b_N = 1/N$ and spiking rule with $a_N = N\sqrt{N}$. Despite steady convergence to the true signal, a bias remains even for larger network sizes. Right panels: Loose balance and stationary bias, where $a_N = N = b_N^{-1}$. The overall performance remains stable, with vanishing fluctuations for higher network sizes. Parameters: $J = 0.5,\ \sigma_V = 1$.

fixed bias in a loosely balanced regime (right panels). In both cases, readout fluctuations decrease with network size - a consequence of the decreasing decoder weights w_i (recall that we use the one-dimensional model presented in Section 3.4).
However, the effect is clearly stronger in the case of the fixed bias, suiting to the decreasing intrinsic variance imposed through our scaling assumption. While a non-improving bias might seem like a highly suboptimal computational feature at first, this setting offers a number of advantages. Firstly, any realistic network with finite size will necessarily have a bias in the estimate due to the non-zero costs (as is the case in Figure 4 for classical balance). In contrast to the regime of constantly decreasing bias, a fixed mean error ensures a robust performance independent of network size. Most importantly, we argue that as long as the bias is deterministic and tractable, it can be easily accounted for. In turn, removing the systematic but well-known bias leads to a perfect prediction [16]. In effect, a biased signal tracking might be of computational advantage due to a reduced number of required spikes, ensured both by the lower magnitude of the incrementally 'assembled' readout and the loosely balanced regime in general [18, 17].
In the following chapters, we will mainly focus on this regime of loose balance combined with a fixed readout bias. As will be shown in Section 6, the scaling lends itself naturally to a mean-field approach, leading to tractable results for both bias and variance.

5 Introduction of the Poisson spiking model

5.1 Approximation of LIF dynamics using a Linear-Nonlinear cascade model

The threshold and reset mechanisms (and the resulting refractoriness effects) render the theoretical analysis of the LIF model very difficult. A standard approach to make the dynamics tractable is to replace explicit spiking mechanisms with stochastic firing rates, for which accurate analytical approximations can be made on macroscopic levels. Recently, a first theoretical treatment of a predictive coding network based on a classical rate model [37] has been proposed by Kadmon et al. [16]. They replaced the spike trains y in membrane equation (15) by corresponding firing rates r (the same is done for the readout) and dropped threshold mechanism together with the corresponding reset. Instead, the rates themselves were determined by a nonlinear transformation of the membrane potential, i.e. $r = \tilde{\phi}(V)$. Here, $\tilde{\phi}$ was chosen as an arbitrary function that restricts the output values for the firing rates to be between 0 and 1. For instance, classical choices could be the logistic function or the *tanh*, which was used by the authors for illustration.
We adopt this approach for the original LIF model by modifying it in several ways. Firstly, instead of the membrane potential itself we model an input potential h to the network, comprising both the external and the recurrent inputs:

$$\tau_h \dot{h}_k(t) = -h_k(t) + a_N b_N \sum_\alpha \xi_{\alpha k} s_\alpha(t) - a_N b_N^2 \sum_{i=1}^{N} \sum_\alpha \xi_{\alpha k} y_i(t) \tag{28}$$

Here, τ_h is a specific time-scale that is different from the membrane time constant τ_m in general and can be regarded as a free parameter of the model. Secondly, to map the input potential directly to a corresponding instantaneous firing rate in the original model, we put it through the LIF transfer function (parameterized by the noise strength σ_V), acting as nonlinearity:

$$r_i(t) = \phi(h_i(t)) := \phi_{LIF}(h_i(t); \sigma_V). \tag{29}$$

Note that while the cost parameters do not appear in the dynamics anymore, they are absorbed by the LIF transfer function, such that the network can still be studied with all of its original characteristics. As a third and last step, spikes are generated stochastically with the probability of spiking given by the instantaneous firing rate, i.e. for small time intervals Δt

$$P(\text{Neuron k spikes in } (t, t + \Delta t)) = r_k(t)\Delta t. \tag{30}$$

In effect, the spiking mechanism obeys an inhomogeneous Poisson process based on the time-dependent firing rate $r(t)$.
The resulting model is a linear-nonlinear Poisson cascade model [27], referred to simply as 'Poisson model' in the scope of this work. By definition of the LIF transfer function, the mapping is exact in the case of unconnected leaky integrate-and-fire neurons receiving constant synaptic input. Moreover, such a linear-nonlinear approach using the LIF transfer function has been shown to reliably reproduce the dynamics of general spiking leaky integrate-and-fire networks [29]. This result was demonstrated both in the presence of time-varying inputs and different noise strengths, as long as the inputs to the neurons were not exceedingly low.
Since the model still generates individual spikes (opposed to the model of Kadmon et al. [16]), we can compare the performance and detailed characteristics of the Poisson network directly with its original LIF counterpart.
In Figure 5, we see that for an exemplary time-varying input in form of a sinus wave, the readout of the Poisson model excellently matches the main evolution of the LIF estimate already at a small network size of 100 neurons. It is able to capture even nuances such as

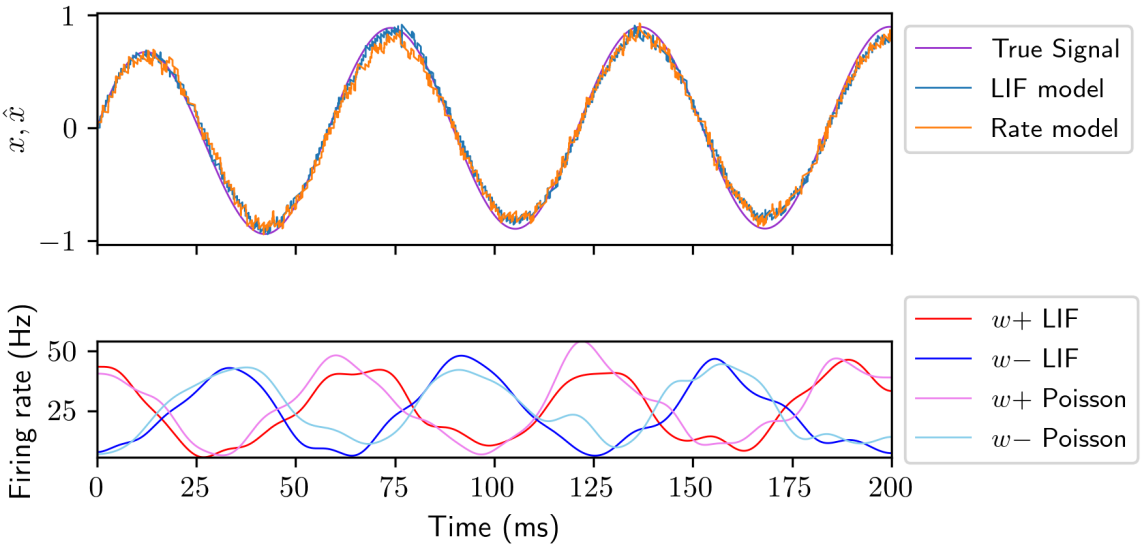

Figure 5: Comparison of the Poisson and LIF models for a sinusoidal input function. The Poisson network produces a good qualitative fit of the both readout and (mean) firing rates of the LIF model, albeit with slightly higher fluctuations. Note that the firing rates in both models are subject to a high degree of variability. The instantaneous rates are smoothed by a temporal Gaussian filter with a standard deviation of 4 ms. Parameters: $N = 100,\ J = 5,\ \sigma_V = 1,\ \tau_h = 5\text{ms}$

the small deviations from the true signal in the crests and troughs, as well the slight time-lag present throughout the simulation time. Similarly, the firing rates (averaged over the two populations of neurons with positive and negative selectivity and smoothed by a Gaussian filter) remain of the same magnitude and indicate similar dynamics in both models. Note that the exact firing rates, based on spiking averages from the relatively small network, underlie a high amount of variation due to both external noise and intrinsic network dynamics. Averaged over a large number of identical simulations with different realizations of noise, only the mean evolution, and magnitude of the rates would play a role. However, the fluctuations in spiking make up one of the main features influencing both the performance and the overall dynamics of the network.

However, one can see that the Poisson readout tends to deviate slightly more than the LIF estimate. This phenomenon is confirmed in the simulations for varying input signals in Figure 6. While the mean performance of the LIF model is accurately tracked by the Poisson estimate throughout the different scenarios (without specifically adjusting any model parameters), the Poisson model brings about sharp, irregular oscillations in the readout exceeding those produced by the LIF model.

We also note that there is a small but consistent mismatch at times where a strong input is rapidly changing (see the top right panel in Figure 6). The speed (and intensity) by which the input potential h evolves depends on the free parameter tau_h. In the dynamic range of the LIF transfer function ϕ (i.e. where it does not saturate with vanishing derivative), this directly affects the response in firing rates as well. In the seminal work of Ostojic and Brunel [29], the mapping between LIF networks and linear-nonlinear cascade models was considerably improved by introducing an adaptive, input-dependent time-scale, corresponding to the role of τ_h in our model. Similarly, we can explore the effect of varying tau_h for the predictive coding network. Using the aforementioned example of a fast cosine input stimulus from Figure 6, we observe a clear impact of the time-scale on both readout $\hat{x}$ and underlying firing rates r (see Figure 7). Small values of τ_h lead to high firing rates and considerably better

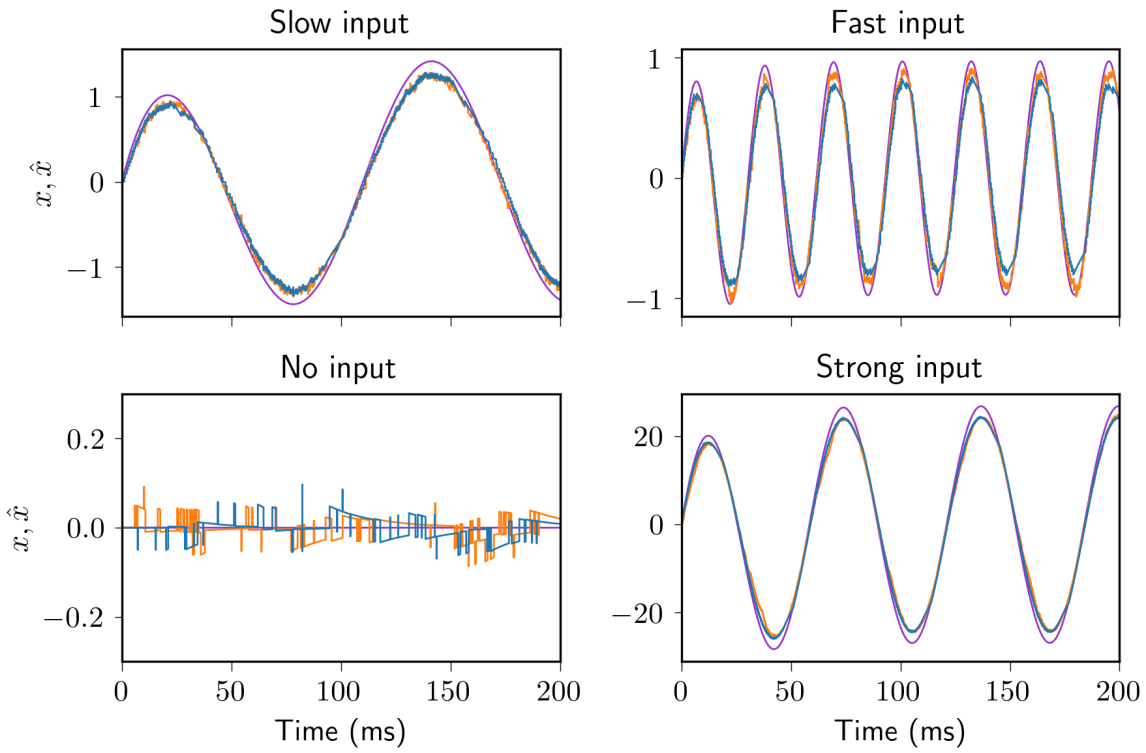

Figure 6: Comparison of the LIF (orange) and Poisson (green) models for different kinds of input. The true signal is shown in blue. When there is no external input (bottom left panel), both readouts fluctuate around zero driven solely by random noise. Different realizations of the noise yield the difference in trajectories. The Poisson model is able to closely replicate the behavior of the LIF model, except for notably faster variance in the quiescent state (bottom left panel). A consistent mismatch also arises in the crests and troughs of a rapidly evolving signal (top right panel). Parameters: $N = 100,\ , J = 5, \sigma_V = 0.5, \tau_h = 5ms$.

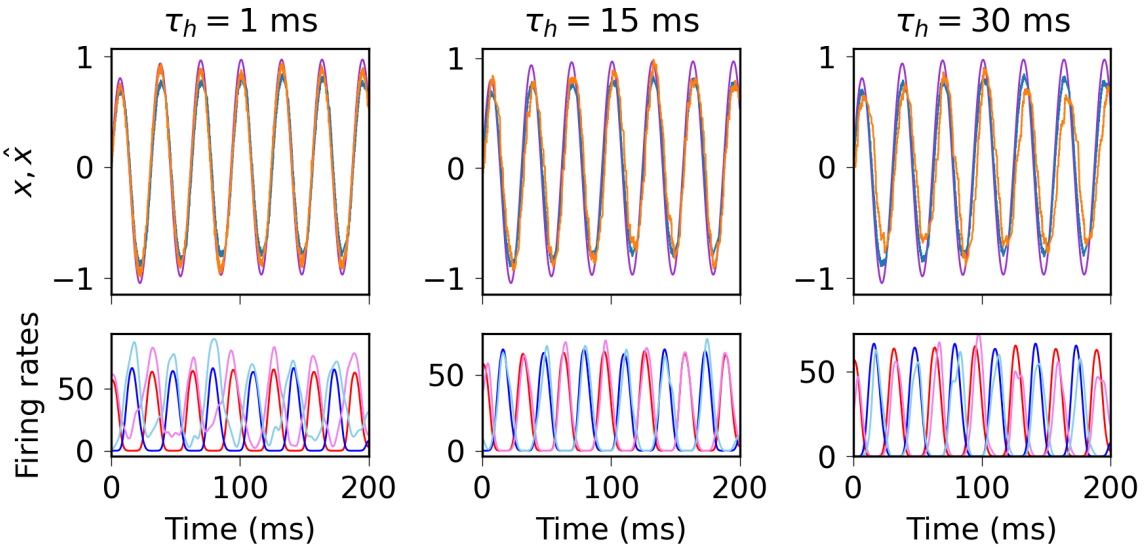

Figure 7: The performance of the Poisson model depends on the parameter τ_h. Color scheme as in Figure 5. The input corresponds to the fast signal from Figure 6. For small τ_h (left panel), the readout tends to overestimate the accuracy of the LIF model, tracking the signal with more precision at the expense of overly high firing rates. On the other hand, a large τ_h performs worse than the LIF model, with significantly lower firing rates (on average), lower amplitude, and a rougher path of $\hat{x}$. Moreover, a high time-scale of the input potential leads to an additional, consistent time lag in tracking the signal. A better match of both estimate $\hat{x}$ and corresponding firing rates for the given input is achieved by a moderate value of τ_h (middle panel). Parameters as in Figure 6.

coding accuracy than in the LIF model. Markedly, in the peaks and troughs, where the wave-like signal is strongest (in absolute value) and sharply changes direction, the Poisson output often matches the true signal better than the LIF readout. Note that while high accuracy with respect to the true signal is a desirable property in itself for a predictive coding network, the main motivation of introducing the Poisson model is to approximate the LIF dynamics. Therefore, a low τ_h is suboptimal in the case of the given signal. With larger values for the time-scale, the firing rates and the amplitude successively decrease. Moreover, as displayed in Figure 7 (right panel, $t_h = 30$), a significant time-lag emerges for overly high time-scales. Depending on the characteristics of both network and input, an optimal intermediate value can be fitted such that the firing rates and readout performance are close to the LIF network (middle panel in Figure 7).

5.2 Detailed comparison of LIF and Poisson models in the stationary regime

While an in-depth analysis for general time-dependent inputs lies out of the scope of this work, we will study the fit of the Poisson model as well as the effect of τ_h in the fundamental case of a single, constant input s. In this setting, the signal quickly evolves towards the steady-state solution $x = 1/\lambda s$. In such an adiabatic limit of slowly evolving signals, neuronal activity and key characteristics of the network become time-independent, such that mean statistics can be equivalently computed as averages over time or different stochastic realizations. In particular, the mean-squared error and the firing rates, constituting the central computational features encoded in the objective function, are stationary as well.
As depicted in Figure 9, the value of τ_h still has an impact on the readout. In particular,

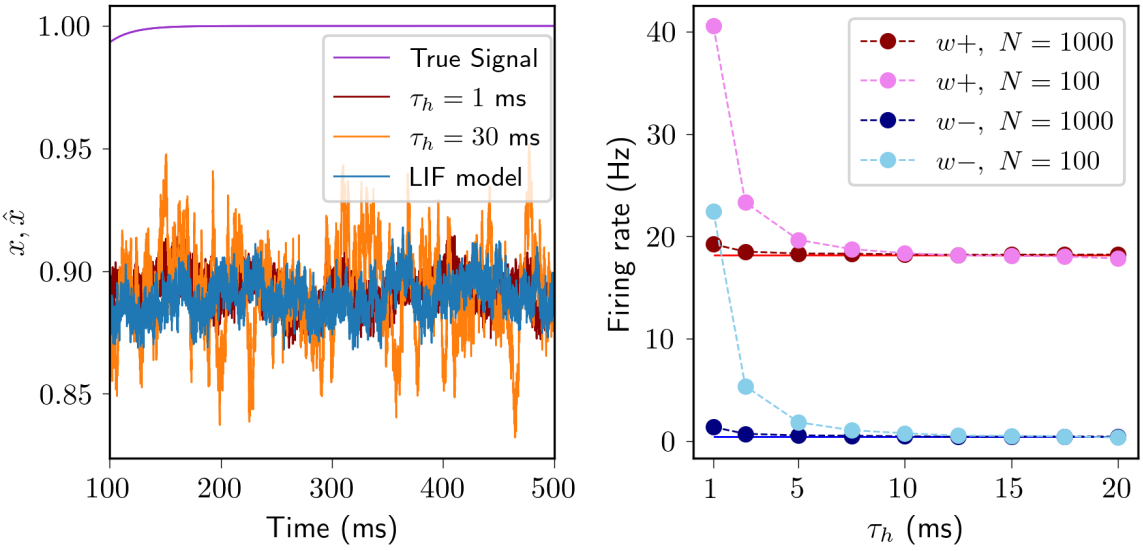

Figure 8: The fit of the Poisson model in the stationary regime. Left Panel: The readout of the Poisson network depends on the additional time-scale parameter τ_h. Larger values of the time-scale lead to larger excursions from the stationary mean. Parameters: $N = 1000,\ J = 5,\ \sigma_V = 0.5$. Right panel: Averaged firing rates over the two populations with positive and negative weights for increasing τ_h and varying network size. The solid blue and red lines describe the stationary firing rate of the LIF network, which remains stable independent of the number of neurons. The Poisson network fires with significantly higher rates for low values of the time-scale in the smaller network of $N = 100$ neurons. The surge occurs equally strongly in both populations. The mismatch decreases notably if the number of neurons increases to $N = 1000$. In both cases, firing rates align with those of the LIF network for larger time-scales. Parameters: $J = 5,\ \sigma_V = 0.5$

larger values of the time-scale lead to larger excursions from the stationary mean. On the other hand, smaller values tend to give higher firing rates than in the LIF model. This surplus appears in equal degrees for both populations of positive and negative weights, leading to a tight balance of the readout without shifting its mean. However, the mismatch quickly vanishes for higher time-scales, with the firing rates exponentially decaying to match those of the LIF model. Moreover, even for small time-scales the difference in rates becomes negligible with increasing network size (Figure 9, right panel). Note that the mean firing rate per neuron remains the same regardless of the number of neurons - a consequence of our scaling assumptions. The network response intensifies with increasing signal strength x, which is directly proportional to the stimulus $s = \lambda x$ in the stationary case. More precisely, the firing rates of the selective neurons (i.e. with positive weight in the case of a positive stimulus, and vice versa) grow linearly with x in the LIF model. As shown in Figure 9, the Poisson model accurately reproduces the same rates, with accuracy rapidly increasing not only for larger networks and higher time-scales (as discussed above) but also with stronger input current. Regardless of the time-scale τ_h, the Poisson model perfectly matches the mean value of the readout $\langle \hat{x} \rangle$ (Figure 9, lower left panel). This holds even for strong inputs, where the readout becomes increasingly biased with respect to the true signal. Conversely, the variance reveals a fundamental difference between the two models. While the fluctuations in the LIF model do not change with input strength, they increase steadily for the Poisson network (Figure 9, lower right panel). To explain that phenomenon, it is crucial to understand the source of the fluctuations in the two networks. In the LIF model, apart from the constant contribution of external noise, fluctuations can arise due to strong recurrent activity in the all-to-all connected network with fast synapses. The fluctuation-driven regime is characterized by a strong balance of excitation and inhibition, dominated by the network-intrinsic lateral activity. In effect, the mean membrane potential remains stably below (though possibly near) the threshold, and spikes occur due to random deviations from this mean activity. The dynamics are shaped by a high degree of variability, a key property emphasized by Boerlin et al. in their seminal paper [6]. In the Poisson model, these dynamics are postulated a priori by the name-giving spiking rule, leading to the characteristic CV of 1 independent of signal, noise strength, or time-scale. Small deviations in variability arise only due to the doubly stochastic nature of the model, with random spikes entering the dynamics of the driving input potential h, which in turn forms their probability distribution. Here, we can also see a small effect due to the time-scale τ_h, with smaller time-scales giving slightly higher CV values.

In the fluctuation-driven regime, high firing rates induced by strong input currents naturally lead to higher lateral activity and therefore greater amounts of intrinsic variability. As a consequence, readout variance is systematically amplified by stronger external stimuli in the Poisson model. Moreover, the variance of the readout is additionally magnified by the time-scale τ_h. As observed in Figure 8, a slow time-scale leads to larger excursions from the stationary mean, such that variance grows proportionally not only to the firing rates but also to τ_h.

In contrast to the Poisson network, the LIF network does not sustain a fluctuation-driven regime for increasing inputs. This can be deduced from the fastly decreasing CV values (Figure 9, upper right panel). Exceedingly strong external inputs swamp the membrane potential and drive it either directly towards the threshold or, for oppositely tuned neurons, far away from it. In turn, spiking becomes increasingly deterministic, with an accordingly smaller degree of fluctuations and a lower CV. The dynamics of the network switch from a fluctuations-driven to a mean-driven regime. This evolution remains unchanged in the case of higher external noise (see Figure 10). Even though the CV can significantly exceed 1 for lower input strengths (i.e. there is more variability in the LIF than in the Poisson model), it quickly drops for stronger stimuli. Interestingly, the total amount of variance in the readout remains

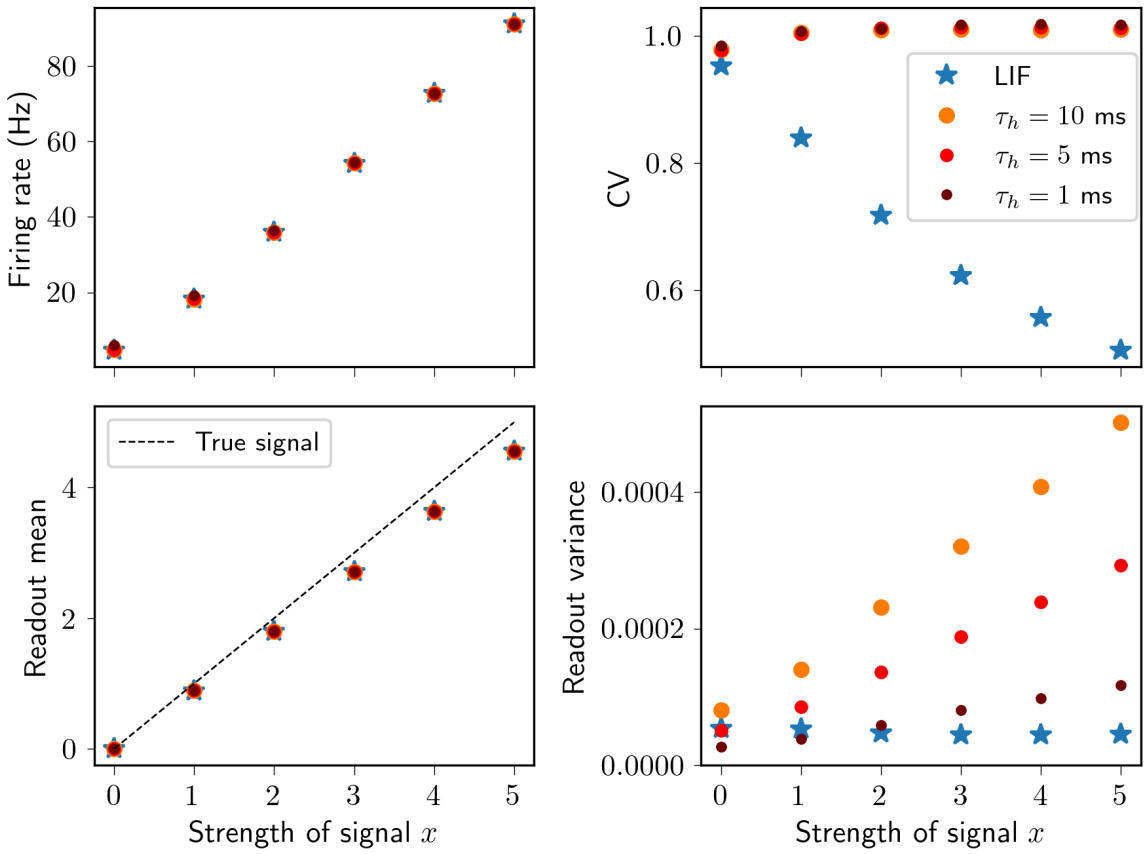

Figure 9: Dependence of LIF and Poisson dynamics on increasing input strength. Three different values of τ_h are considered (orange, red, and dark-red dots) for the Poisson model. Upper left panel: Firing rates of the population with positive weights for a stationary, positive signal x. The rates grow linearly with the signal strength. Except for a small mismatch in the case of $\tau_h = 1$ ms and smaller input strengths, the Poisson network accurately matches the firing rates of the LIF network. Upper right panel: Coefficient of variation (CV). The CV reveals an important systematic difference between the two models. While it remains very close to 1 in the Poisson model regardless of input strength, the CV drops significantly for stronger stimuli. This indicates a switch from the fluctuations-driven to the mean-driven regime in the dynamics of the LIF network. Lower left panel: Readout mean $\langle \hat{x} \rangle$. The means of the LIF network show an increasing bias with a higher input magnitude. The Poisson model produces an excellent match for the mean regardless of τ_h. Lower right panel: Variance of the readout $\hat{x}$. While the variance of the LIF readout remains unchanged for higher inputs, the variance of the Poisson model continuously increases. Both absolute value and growth with signal strength are higher for larger values of τ_h. The mismatch occurs due to the difference in variability shown in the upper right panel. Parameters: $N = 1000,\ J = 5,\ \sigma_V = 0.5$.

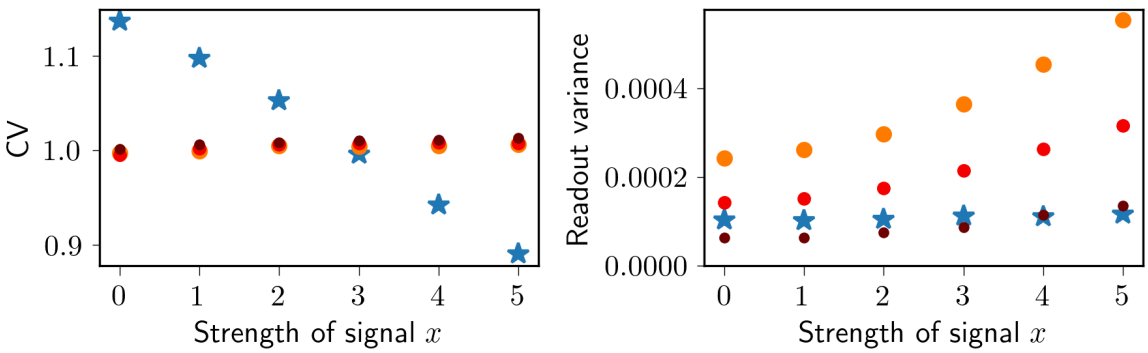

Figure 10: Variability in the LIF and Poisson networks for increased external noise. While the CV of the LIF network becomes larger than 1 for weak input stimuli, the overall pattern of decreasing variability in network activity and constant readout variance remains unchanged. Parameters: $N = 1000,\ J = 5,\ \sigma_V = 1$.

perfectly stable at very low levels relative to the mean. This indicates that the performance is tightly controlled by the error-correcting mechanism, a central idea of the computational model. Indeed, larger deviations from the mean are instantaneously propagated throughout the network, provoking a reverse response that draws the readout back to the stationary solution [6].

In conclusion, high firing rates do not yield higher fluctuations in the LIF model, as they are counter-balanced by the decrease in the network-induced variability. This qualitative difference cannot be fully reconciled even by tuning τ_h as a function of the input potential. As the variance grows continuously with x in the Poisson model, eventually it will overestimate the unchanging variance of the LIF readout even for small time-scales. On the other hand, extremely low values of τ_h lead to excessive fluctuations in the firing rates, as any change in the input is magnified by $1/\tau_h$.

Moreover, we observe that the match depends on other parameters such as noise parameter σ_V as well. For instance, while in the case of $\sigma_V = 0.5$ (Figure 9) a good match for the variance in the quiescent regime $x = 0$ is provided by $\tau_h = 5ms$, a lower value lower time-scale (e.g. $3ms$) would be more optimal for larger noise level $\sigma_V = 1$. Still, we note that the overall magnitude of the variance is predicted correctly for reasonable choices of the input strength and realistic CV values, suiting the high variability observed in vivo. In particular, in regimes where the CV of the LIF network is close to 1, a good fit is provided by the Poisson model with $\tau_h = 1ms$. Furthermore, the variance quickly becomes negligible for the total coding error even for larger inputs, as it remains at very low values compared to the growing bias.

6 Mean-field approach for the spiking predictive coding network

6.1 General mesoscopic dynamics

A major advantage of the Poisson model is that its stochastic properties are well-understood and allow for analytical tractability. The mean and variance of a neuron's spike count are given by the dynamic rate $r(t)$, such that for sufficiently large and homogeneous (with respect to selectivity) networks we can approximate the spike trains by using a Gaussian process as follows:

$$y_i(t) = r_i(t) + \sqrt{r_i(t)}\eta_i(t) = \phi(h_i(t)) + \sqrt{\phi(h_i(t))}\eta_i(t), \tag{31}$$

where η_i corresponds to Gaussian white noise with autocorrelation function $\langle \eta_i(t), \eta_j(t') \rangle = \delta_{ij}\delta(t-t')$. Indeed, for sufficiently high N the central limit theorem guarantees that the total activity, described by the sum of all spike trains, will have Gaussian statistics with approximately the same first and second moments.
Using this approximation, we can adapt the theoretical analysis of Kadmon et al. [16] to derive the mean-field dynamics of the Poisson model. In turn, this can yield instructive insights into the dynamics of the LIF network as well.
Assume that the patterns $\xi_{\alpha i}$ are i.i.d. samples generated from a symmetrical, zero mean probability distribution $\mathcal{P}(\xi)$ such that for each coding dimension α the vector $\boldsymbol{\xi_\alpha}$ is concentrated at $\boldsymbol{\xi_\alpha}^T\boldsymbol{\xi_\alpha} = N$. This is in line with our previous assumption that the patterns remain of order 1 independent of network size. An example are the binary patterns of our sample model introduced in Section 3.4. The choice of the patterns corresponds to the structured connectivity used by Kadmon et al. in their analysis. The main motivation of using the normalized weights is that it allows to easily decompose h into two orthogonal components $h = h^{\parallel} + h^{\perp}$, with

$$\mathbf{h}^{\parallel} = \mathbf{P}\mathbf{h}, \tag{32}$$

$$\mathbf{h}^{\perp} = (\mathbf{I} - \mathbf{P})\mathbf{h}. \tag{33}$$

Here, $\mathbf{P}$ denotes the $N \times N$ orthogonal projection matrix onto the direction of the patterns ξ, with entries $\mathbf{P}_{ij} = \frac{1}{N}\sum_\alpha \xi_{\alpha i}\xi_{\alpha j} = \frac{1}{N}(\boldsymbol{\xi^{(i)}})^T\boldsymbol{\xi^{(j)}}$ and $\boldsymbol{\xi^{(i)}} = (\xi_{1i}, \ldots, \xi_{Mi})$.
Recall that a symmetric matrix $\mathbf{B}$ is an orthogonal projection matrix if it is equal to its square, i.e. $\mathbf{B}^2 = \mathbf{I}$. Denote $\boldsymbol{\xi} = (\xi_{\alpha i})_{i=1,\ldots,N}^{\alpha=1,\ldots,M}$ the $N \times M$ matrix of all patterns with columns $\boldsymbol{\xi^{(i)}}$, $i = 1, \ldots, N$. Then the projection matrix can be written as $\mathbf{P} = \frac{1}{N}\boldsymbol{\xi}^T\boldsymbol{\xi}$. Due to the assumption of normalized patterns it holds $\boldsymbol{\xi}\boldsymbol{\xi}^T = N$, which immediately implies that $\mathbf{P}^2 = \frac{1}{N^2}\boldsymbol{\xi}^T\boldsymbol{\xi}\boldsymbol{\xi}^T\boldsymbol{\xi} = \frac{1}{N}\boldsymbol{\xi}^T\boldsymbol{\xi} = \mathbf{P}$, i.e. $\mathbf{P}$ is indeed a projection matrix. In particular, $\mathbf{P}$ projects h onto the M-dimensional subspace spanned by the pattern vectors $\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_M}$. Conversely, $\mathbf{I} - \mathbf{P}$ projects on the subspace which is perpendicular to the patterns: $(\mathbf{I} - \mathbf{P})^T\boldsymbol{\xi} = (\mathbf{I} - \mathbf{P})\boldsymbol{\xi}^T = \boldsymbol{\xi}^T - \frac{1}{N}\boldsymbol{\xi}^T\boldsymbol{\xi}\boldsymbol{\xi}^T = 0$.
Multiplying both sides of the vectorized dynamic equation (analogously as for the original model (16))

$$\tau_h \dot{h}(t) = -h(t) - a_N b_N^2 \xi^T\xi y + a_N b_N \xi^T s \tag{34}$$

by the projection $I - P$ onto the orthogonal subspace gives the simple form

$$\tau_h \dot{h}^{\perp} = -\dot{h}^{\perp}. \tag{35}$$

In other words, the orthogonal component decays to 0 and can be safely neglected, yielding

$$h_i(t) \approx h_i^{\parallel}(t). \tag{36}$$

Note that the approximation is exact if the orthogonal component is initialized as zero. The parallel component $h_i^{\parallel}(t)$ can be written as

$$h_i^{\parallel}(t) = \frac{1}{N}\sum_j \sum_\alpha \xi_{\alpha i}\xi_{\alpha j}h_j = \sum_\alpha \xi_{\alpha i} q_\alpha(t),$$

where

$$q_\alpha(t) = \frac{1}{N}\sum_j \xi_{\alpha j} h_j(t) \tag{37}$$

is a mesoscopic variable that can readily be analyzed using a mean-field approach. Indeed, as $\mathbf{q}$ incorporates the total activity of all neurons combined, it allows to take the limit of

infinitely large networks and to make use of corresponding approximations. However, in contrast to models based solely on macroscopic variables (such as population rates), $\mathbf{q}$ is still made up of the individual contributions of microscopic neuronal units. This allows to directly connect the mean-field results to smaller models as well. For instance, we can evaluate the theoretical insights for networks with a mesoscopic size of 100-5000 neurons.
The dynamics of $\mathbf{q}$ can be derived directly from those of the input potential h:

$$\begin{aligned}\tau_h \dot{q}_\alpha &= \frac{1}{N}\sum_i \xi_{\alpha i}(\tau_h \dot{h}_i) \\ &= \frac{1}{N}\sum_i \xi_{\alpha i}\left(-h_i + a_N b_N \sum_\nu \xi_{\nu i} s_\nu - a_N b_N^2 \sum_\nu \sum_j \xi_{\nu i}\xi_{\nu j} y_j\right) \\ &= -q_\alpha + \frac{a_N b_N}{N}\sum_i \sum_\nu \xi_{\alpha i}\xi_{\nu i} s_\nu - \frac{a_N b_N^2}{N}\sum_i \sum_\nu \sum_j \xi_{\alpha i}\xi_{\nu i}\xi_{\nu j} y_j.\end{aligned}$$

Splitting the double sum into two components,

$$\sum_i \sum_\nu \xi_{\alpha i}\xi_{\nu i} s_\nu = \sum_i \xi_{\alpha i}^2 s_\alpha + \sum_i \xi_{\alpha i} \sum_{\nu\neq\alpha} \xi_{\nu i} s_\nu, \tag{38}$$

one can once again make use of the pattern structure: The first sum becomes simply $N s_\alpha$, while the second term sums over products of the two independent, zero mean variables $\xi_{\alpha i}$ and $\sum_{\nu\neq\alpha} \xi_{\nu i} s_\nu$. Thus, the mean of the products is zero as well and their sum over all neurons vanishes for sufficiently high N. Similarly, we can approximate

$$\begin{aligned}\sum_i \sum_\nu \sum_j \xi_{\alpha i}\xi_{\nu i}\xi_{\nu j} y_j &= \sum_i \xi_{\alpha i}^2 \sum_j \xi_{\alpha j} y_j + \sum_i \xi_{\alpha i} \sum_{\nu\neq\alpha} \sum_j \xi_{\nu i}\xi_{\nu j} y_j \\ &= N \sum_j \xi_{\alpha j} y_j,\end{aligned} \tag{39}$$

which is exact in the mean-field limit $N \to \infty$ due to the mutual orthogonality of the patterns. Thus, the dynamics simplify to

$$\tau_h \dot{q}_\alpha = -q_\alpha + a_N b_N s_\alpha - a_N b_N^2 \sum_i \xi_{\alpha i} y_i. \tag{40}$$

An interesting special case is where the capacity M of the network exceeds the number M' of relevant features x_α to encode (as introduced in Section 3.2). In particular, assume that $s_{M'+1}, s_{M'+2}, \ldots, s_M$ are zero for some $M' < M$. Note that the above approximations (38) and (39) remain valid in the same way as before, leading to dynamics (40) for all coding dimensions α. In particular, any superfluous feature $\beta > M'$ is governed by

$$\tau_h \dot{q}_\beta = -q_\beta - a_N b_N^2 \sum_i \xi_{\beta i} y_i. \tag{41}$$

Moreover, the firing activity is driven by the non-zero inputs (or by pure noise), such that patterns $\xi_{\beta i}$ are equally likely to be positive or negative for any 'active' spike train y_i. Consequently, the sum $\sum_i \xi_{\beta i} y_i$ contributes a mean-zero random variable to the dynamics, with negligible variance in the case of large N and appropriate regularization by $a_N b_N^2 = o(\sqrt{N}^{-1})$ which is ensured by our scaling conditions (see Section 4.2). As a result, q_β vanishes in the large N limit and does not affect the dynamics of the relevant features (see the mean-field results for q_α below, e.g. Equation (46)). In other words, a low-dimensional signal yields equally low-dimensional dynamics in the network activity, guaranteeing the same degree of coding efficiency as in the idealized case $M = M'$.

Continuing the derivation towards a mean-field equation for general q_α, we use the Gaussian approximation for the spike trains presented in Equation (31):

$$\tau_h \dot{q}_\alpha = -q_\alpha + a_N b_N s_\alpha - a_N b_N^2 \sum_i \xi_{\alpha i} \left(\phi\left(h_i\right) + \sqrt{\phi\left(h_i\right)} \eta_i \right). \tag{42}$$

Using the orthogonal composition of h and the corresponding approximation (36), we obtain a closed-from equation in $\mathbf{q}$,

$$\tau_h \dot{q}_\alpha = -q_\alpha + a_N b_N s_\alpha - a_N b_N^2 \sum_i \xi_{\alpha i} \left(\phi\left(h_i^\parallel\right) + \sqrt{\phi\left(h_i^\parallel\right)} \eta_i \right) \tag{43}$$

$$= -q_\alpha + a_N b_N s_\alpha - a_N b_N^2 \sum_i \xi_{\alpha i} \left(\phi\left(\sum_\nu \xi_{\nu i} q_\nu\right) + \sqrt{\phi\left(\sum_\nu \xi_{\nu i} q_\nu\right)} \eta_i \right). \tag{44}$$

To obtain a stochastic differential equation with fixed drift and diffusion terms independent of individual realizations of the patterns ξ_i, we take the mean-field limit and replace the sums with integrals. In particular,

$$\frac{1}{N} \sum_i \xi_{\alpha i} \phi\left(\sum_\nu \xi_{\nu i} q_\nu(t)\right) \xrightarrow{N\to\infty} \int \xi_1 \phi\left(\sum_\nu \xi_\nu q_\nu\right) \prod_{\nu=1}^{M} \mathcal{P}(\xi_\nu)\, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_M. \tag{45}$$

Note that the integral is well-defined for bounded patterns $\xi = O(1)$ and any smooth transfer function ϕ, such as in our work. For the noise term

$$\eta^\parallel := \sum_i \xi_{\alpha i} \sqrt{\phi\left(h_i^\parallel\right)} \eta_i = \sum_i \xi_{\alpha i} \sqrt{\phi\left(\sum_\nu \xi_{\nu i} q_\nu\right)} \eta_i,$$

it suffices to compute the autocorrelation function, equal to

$$\left\langle \eta^\parallel(t) \eta^\parallel(t') \right\rangle = \sum_i \sum_j \left\langle \xi_{\alpha i} \xi_{\alpha j} \phi\left(\sum_\nu \xi_{\nu i} q_\nu\right) \eta_i(t) \eta_j(t') \right\rangle.$$

In the limit of large networks, we can reasonably assume that the contributions of individual noise components η_i to the variable $\mathbf{q}$ become negligibly small. In other words, we assume that $\mathbf{q}$ becomes deterministic in the mean-field limit, a standard approach in the analysis of stochastic computational models (in particular in the context of Hopfield networks) [38, 3]. As a consequence, $\mathbf{q}$ and the white noise components can be separated,

$$\begin{aligned}
\left\langle \eta^\parallel(t) \eta^\parallel(t') \right\rangle &= \sum_i \sum_j \left\langle \xi_{\alpha i} \xi_{\alpha j} \sqrt{\phi\left(\sum_\nu \xi_{\nu i} q_\nu(t)\right) \phi\left(\sum_\nu \xi_{\nu i} q_\nu(t')\right)} \right\rangle \left\langle \eta_i(t) \eta_j(t) \right\rangle \\
&= \sum_i \sum_j \left\langle \xi_{\alpha i} \xi_{\alpha j} \sqrt{\phi\left(\sum_\nu \xi_{\nu i} q_\nu(t')\right) \phi\left(\sum_\nu \xi_{\nu i} q_\nu(t')\right)} \right\rangle \delta_{ij} \delta(t - t') \\
&= \sum_i \left\langle \xi_{\alpha i}^2 \phi\left(\sum_\nu \xi_{\nu i} q_\nu(t)\right) \right\rangle \delta(t - t')
\end{aligned}$$

Once again averaging over the population, it follows

$$\frac{1}{N} \left\langle \eta^\parallel(t) \eta^\parallel(t') \right\rangle = \left(\int \xi_1^2 \phi\left(\sum_\nu \xi_\nu q_\nu(t)\right) \prod_{\nu=1}^{M} \mathcal{P}(\xi_\nu)\, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_M \right) \delta(t - t').$$

Thus, we again recover white noise but with rescaled autocorrelation. Denoting

$$F(\mathbf{q}) := F(q_1, \dots, q_M) := \int \xi_1 \phi \left(\sum_\nu \xi_\nu q_\nu \right) \prod_{\nu=1}^{M} \mathcal{P}(\xi_\nu) \, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_M,$$

$$R(\mathbf{q}) := R(q_1, \dots, q_M) := \int \xi_1^2 \phi \left(\sum_\nu \xi_\nu q_\nu \right) \prod_{\nu=1}^{M} \mathcal{P}(\xi_\nu) \, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_M,$$

we conclude the mean-field dynamics of $\mathbf{q}$:

$$\begin{aligned} \tau \dot{q}_\alpha &= -q_\alpha + a_N b_N s_\alpha - N a_N b_N^2 F(\mathbf{q}) - a_N b_N^2 \sqrt{N} \eta^{\parallel} \\ &= -q_\alpha + a_N b_N s_\alpha - N a_N b_N^2 F(\mathbf{q}) - a_N b_N^2 \sqrt{N R(\mathbf{q})} \eta, \end{aligned} \tag{46}$$

with standard Gaussian white noise η. Thus, we have derived a standard diffusion process - a closed-form stochastic differential equation, yielding a convenient basis for in-depth theoretical analysis.

Note that for both the dynamic drift term $F(\mathbf{q})$ and the autocorrelation of $\eta^{\parallel}$ we required an additional scaling by $1/N$ as normalization. In the dynamic equation, the resulting prefactors N and $\sqrt{N}$, respectively, suggest that the normalization should be incorporated a priori into the scaling requirements. In our scaling assumptions, we imposed that the intrinsic variance in the membrane potential V_i vanishes for $N \to \infty$. This has led to the condition $a_N b_N^2 = o(1/\sqrt{N})$, which exactly corresponds to the requirement needed here to eliminate the fluctuations in the dynamics of $\mathbf{q}$. At the same time, the scaling assumption yields that both external input and recurrent feedback to each neuron are of order $a_N b_N = O(N^\epsilon)$, $0 \leq \epsilon < 0.5$, corresponding to loose balance (see Section 3.2). As a consequence, loose balance is not only possible but necessary to obtain a well-defined, deterministic mean-field limit of the dynamics. In contrast, the prevalent view is that the predictive coding model requires a tight balance of excitation and inhibition with synaptic inputs scaled as $O(N)$ to operate efficiently [6, 16]. This raises the question of how the error scales with N and in particular, whether superclassical error scaling - with fluctuations decreasing faster than $1/\sqrt{N}$ - can be achieved by a network working in a loosely balanced regime. This problem is addressed in the following section.

6.2 Mean-field theory for bias and variance in the stationary case

In the following, we show how the mesoscopic variable q can be used to derive analytical results for the mean and variance of the readout $\hat{x}$. Hereby, we continue to adopt the main steps of the analysis of Kadmon et al. [16], extending their results to spiking models. Analogously to [16], we work in the adiabatic limit of very slow signals compared to the time-scale of the input potential, such that effectively the signal x can be assumed to remain constant over time. This corresponds to the case of a stationary external input over long periods. For simplicity, we will also assume the one-dimensional case of just a single constant input $s(t) \equiv s$. However, as shown by Kadmon et al., the same calculation can readily be applied to the multi-dimensional case $M > 1$ as well.

As discussed above, the balance needs to be loose for the fluctuations of q to vanish in the limit. For this purpose, let $b_N = J/N$, where J is a simple scalar. Writing $\hat{b} = a_N b_N / J = a_N / N$ for the level of balance, with $\hat{b} = O(N^\epsilon)$, $0 \leq \epsilon < 0.5$, the mean-field equation (46) becomes

$$\tau \dot{q} = -q + J\hat{b}s - J^2 \hat{b} F(q) - \frac{J^2 \hat{b}}{\sqrt{N}} \sqrt{R(q)} \eta, \tag{47}$$

with

$$F(q) = \int \xi\phi\left(\xi q\right)\mathcal{P}(\xi)\mathrm{d}\xi,$$
$$R(q) = \int \xi^2\phi\left(\xi q\right)\mathcal{P}(\xi)\mathrm{d}\xi.$$

As argued above, the input potential can be well approximated by

$$h_i(t) \approx h_i^{\parallel}(t) = \xi_i q(t). \tag{48}$$

Hence, to obtain the mean and variance of key model variables depending on h, it indeed suffices to compute the corresponding features for q. As a first step, we decompose $q(t) = \langle q\rangle + \delta q(t)$ of q into its temporal mean $\langle q\rangle$ and fluctuations $\delta q = q - \langle q\rangle$. Recall that in the stationary case the temporal mean can equivalently be interpreted as the ensemble mean over a large number of identical neurons or as a trial average. Inserting the decomposition into the dynamic equation gives

$$\tau_h\dot{\langle q\rangle} + \tau_h\delta\dot{q}(t) = \tau_h\delta\dot{q}(t) = -\langle q\rangle - \delta q(t) + J\hat{b}s - J^2\hat{b}F(q) - \frac{J^2\hat{b}}{\sqrt{N}}R(q)\eta, \tag{49}$$

where only the derivative $\delta\dot{q}(t)$ of the fluctuations play a role due to stationarity. In the two mean terms $F(q)$ and $R(t)$, q appears inside the non-linearity ϕ, which can make the analysis significantly more difficult in general. However, as the fluctuations δq diminish with increasing N due to our scaling assumptions, we can expand ϕ about $\delta q = 0$, i.e. around the mean $\langle q\rangle$, using Taylor. In particular, for the term $\phi(\xi q)$ inside the integrals, we obtain

$$\phi(\xi q(t)) = \phi\left(\xi\langle q\rangle + \xi\delta q(t)\right) = \phi\left(\xi\,\langle q\rangle\right) + \phi'\left(\xi\langle q\rangle\right)\xi\delta q(t). \tag{50}$$

Inserting into the definitions of F and R yields

$$F(q) = \int \xi\phi\left(\xi\,\langle q\rangle\right) + \xi^2\phi'\left(\xi\langle q\rangle\right)\delta q(t))\mathcal{P}(\xi)\mathrm{d}\xi,$$
$$R(q) = \int \xi^2\phi\left(\xi\,\langle q\rangle\right) + \xi^3\phi'\left(\xi\langle q\rangle\right)\delta q(t))\mathcal{P}(\xi)\mathrm{d}\xi.$$

Writing

$$F^{\langle\phi\rangle} = \int \xi\phi\left(\xi\langle q\rangle\right)\mathcal{P}(\xi)\mathrm{d}\xi, \qquad F^{\langle\phi'\rangle} = \int \xi^2\phi'\left(\xi\langle q\rangle\right)\mathcal{P}(\xi)\mathrm{d}\xi,$$
$$R^{\langle\phi\rangle} = \int \xi^2\phi\left(\xi\langle q\rangle\right)\mathcal{P}(\xi)\mathrm{d}\xi, \qquad R^{\langle\phi'\rangle} = \int \xi^3\phi'\left(\xi\langle q\rangle\right)\mathcal{P}(\xi)\mathrm{d}\xi,$$

we obtain a first order approximation of the dynamics in Equation (49):

$$\tau_h\delta\dot{q}(t) = -\langle q\rangle - \delta q(t) + J\hat{b}s - J^2\hat{b}F^{\langle\phi\rangle} - J^2\hat{b}F^{\langle\phi'\rangle}\delta q(t) - \frac{J^2\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle} + R^{\langle\phi'\rangle}\delta q(t)}\eta. \tag{51}$$

By definition, the temporal mean of the fluctuations δq is 0 in the stationary case. The same holds for the corresponding derivative $\delta\dot{q}$ and the diffusion term involving Gaussian white noise η. Thus, taking the mean on both sides of Equation (51) yields

$$0 = -\langle q\rangle + J\hat{b}s - J^2\hat{b}F^{\langle\phi\rangle}, \tag{52}$$

or equivalently $\langle q\rangle = J\hat{b}s - J^2\hat{b}F^{\langle\phi\rangle}$. Note that this relation does not give a closed-form solution for the mean of q directly, as $F^{\langle\phi\rangle}$ still depends on $\langle q\rangle$ as well. However, the relation

can be directly exploited to simplify Equation (51). Indeed, as the term $-\langle q\rangle + J\hat{b}s - J^2\hat{b}F^{\langle\phi\rangle}$ adds up to zero by Equation (52), it can be disregarded in the dynamics:

$$\begin{aligned}\tau_h\delta\dot{q}(t) &= -\delta q(t) - J^2\hat{b}F^{\langle\phi'\rangle}\delta q(t) - \frac{J^2\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle} + R^{\langle\phi'\rangle}\delta q(t)}\eta. \\ &= -(1 + J^2\hat{b}F^{\langle\phi'\rangle})\delta q(t) - \frac{J^2\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle} + R^{\langle\phi'\rangle}\delta q(t)}\eta. \end{aligned} \tag{53}$$

This constitutes a dynamic equation of the fluctuations $\delta q(t)$. Note that for a monotonously increasing nonlinearity ϕ, such as the LIF transfer function, the 'effective gain' (borrowing the terminology of Kadmon et al. [16]) $F^{\langle\phi'\rangle}$ is nonnegative, such that the drift coefficient $-(1 + J^2\hat{b}F^{\langle\phi'\rangle}) < 0$ implements negative feedback confining the extent of the fluctuations. This regularizing effect is further increased by a stronger balance $\hat{b}$, which simultaneously magnifies the noise term for moderate network sizes. Characteristic for the model, the dynamics incorporate multiplicative noise, influenced not only by the magnitude of the fluctuations themselves but also by the coefficients $R^{\langle\phi\rangle}$ and $R^{\langle\phi'\rangle}$, depending on the nonlinearity ϕ as well as the mean $\langle q\rangle$. Similarly, as for $\hat{b}$, this effect by definition disappears in the mean-field limit. However, it may play a larger role at the mesoscopic scale, which is our principal target of interest.
Using the first-order approximation of the dynamics of q, we can derive the mean and the variance of the readout $\hat{x}$.

Two relations between $\langle q\rangle$ and $\langle x\rangle$

Firstly, we connect the means of q and $\hat{x}$ by two relations, which in turn can be solved to obtain both simultaneously. For the first relation, it suffices to consider the initial form of the dynamics of q without the mean-field approximations described by Equation (40). Incorporating our scaling assumptions, the equation states

$$\tau_h\dot{q}(t) = -q(t) + J\hat{b}s - \frac{J\hat{b}}{N}\sum_i \xi_i y_i(t).$$

Recall that the readout is defined by the dynamic equation

$$\dot{\hat{x}}(t) = -\lambda\hat{x}(t) + \frac{J\hat{b}}{N}\sum_{i=1}\xi_i y_i(t),$$

implying $\frac{J\hat{b}}{N}\sum_i \xi_i y_i = \lambda\hat{x} + \dot{\hat{x}}$. Thus, the dynamics of q can be expressed as

$$\tau_h\dot{q} = -q + J\hat{b}\left(s - \lambda\hat{x} - \dot{\hat{x}}\right). \tag{54}$$

Taking the mean on both sides, it follows

$$0 = -\langle q\rangle + J\hat{b}\left(s - \lambda\langle\hat{x}\rangle - \langle\dot{\hat{x}}\rangle\right) \tag{55}$$

$$= -\langle q\rangle + J\hat{b}\left(s - \lambda\langle\hat{x}\rangle\right) \tag{56}$$

due to stationarity. This gives the first relation between the means of $\hat{x}$ and q, namely $\langle q\rangle = J\hat{b}\left(s - \lambda\langle\hat{x}\rangle\right)$ or equivalently

$$\lambda\langle\hat{x}\rangle = \left(s - \frac{\langle q\rangle}{J\hat{b}}\right). \tag{57}$$

For the second relation, we make use of mean-field approximations. In particular, by the same arguments as in the derivation of the mean-field dynamics of q, the following approximation holds in the large N limit:

$$\begin{aligned}\lambda\hat{x}(t) + \dot{\hat{x}}(t) &= \frac{J\hat{b}}{N}\sum_i \xi_i y_i(t) \\ &= \frac{J\hat{b}}{N}\sum_i \xi_i \left(\phi\left(\xi_i q(t)\right) + \sqrt{\phi\left(\xi_i q(t)\right)}\eta_i\right) \\ &\xrightarrow{N\to\infty} J\hat{b}F(q) + \frac{J\hat{b}}{\sqrt{N}}\sqrt{R(q)}\eta(t).\end{aligned}$$

Moreover, applying the first-order approximations in the same way as demonstrated above for the dynamics of q, it follows

$$\lambda\hat{x}(t) + \dot{\hat{x}}(t) = J\hat{b}F^{\langle\phi\rangle} - J\hat{b}F^{\langle\phi'\rangle}\delta q(t) - \frac{J\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle} + R^{\langle\phi'\rangle}\delta q(t)}\eta.$$

Once again taking the mean on both sides, we obtain the second relation

$$\lambda\langle\hat{x}\rangle = J\hat{b}F^{\langle\phi\rangle} = J\hat{b}\int \xi\phi\left(\xi\langle q\rangle\right)\mathcal{P}(\xi)\mathrm{d}\xi. \tag{58}$$

Using the two resulting equations, we can numerically compute the means of x and q. The solution is shown graphically in Figure 11 (upper left panel) for the case of binary weights. The first equation (57) describes a linear function in $\langle q\rangle$, with negative slope $-J^{-1} < 0$ and the intercept determined by the strength of the input s. On the other hand, the second relation (58), as a function of $\langle q\rangle$, can be viewed effectively as the odd version of the LIF transfer function (on the positive domain): the negative domain, where the LIF transfer function saturates at zero, is discarded and the remaining graph is mirrored diagonally(corresponding to a rotation of 180 degrees about the origin). Note that for a general, odd nonlinearityϕ such as the *tanh* function, the result would correspond to a smoothed version of the original function (see Kadmon et al. [16]).
The solution corresponds to the intersection of the two lines in the $\langle\hat{x}\rangle$-$\langle q\rangle$ plane. Note that the transfer function, as well as its counterpart described by the second relation for the means, is monotonously increasing. Thus, due to the negative slope of the linear function in the first relation, a unique solution for both means $\langle q\rangle$ and $\langle\hat{x}\rangle$ is guaranteed. For fixed network parameters, the point of intersection depends only on the input s. As the second line remains unchanged, the line for the first relation shifts vertically with s, ensuring that the readout adapts to the signal $x = \lambda^{-1}s$. As displayed in the upper right panel of Figure 11, the theory perfectly predicts the readout means of the LIF model for $N = 1000$ neurons. For smaller network sizes, the theoretical prediction slightly underestimates the true LIF mean (Figure 11, bottom panel), with the relative error remaining well below the 5% margin.
From the dynamical point of view, the solution corresponds to a unique equilibrium point, indicating that the system is governed by a single fixed point attractor. While this might be expected for a simple setting such as depicted here, more complex dynamic regimes can arise in higher dimensions, as well as in the presence of additional sources of irregularity (such as synaptic delays, failures, or chaos [13, 16]). The mean-field theory can be directly extended to investigate the dynamic landscapes of these general settings, as has been shown by Kadmon et al. in the case of rate models [16].

Variance of the fluctuations δq and $\delta\hat{x}$

Similar to the mean, we can derive the variance of the readout $\hat{x}$ using the explicit mean-field dynamics of the fluctuations of q. For completeness, we will first calculate the variance of δq

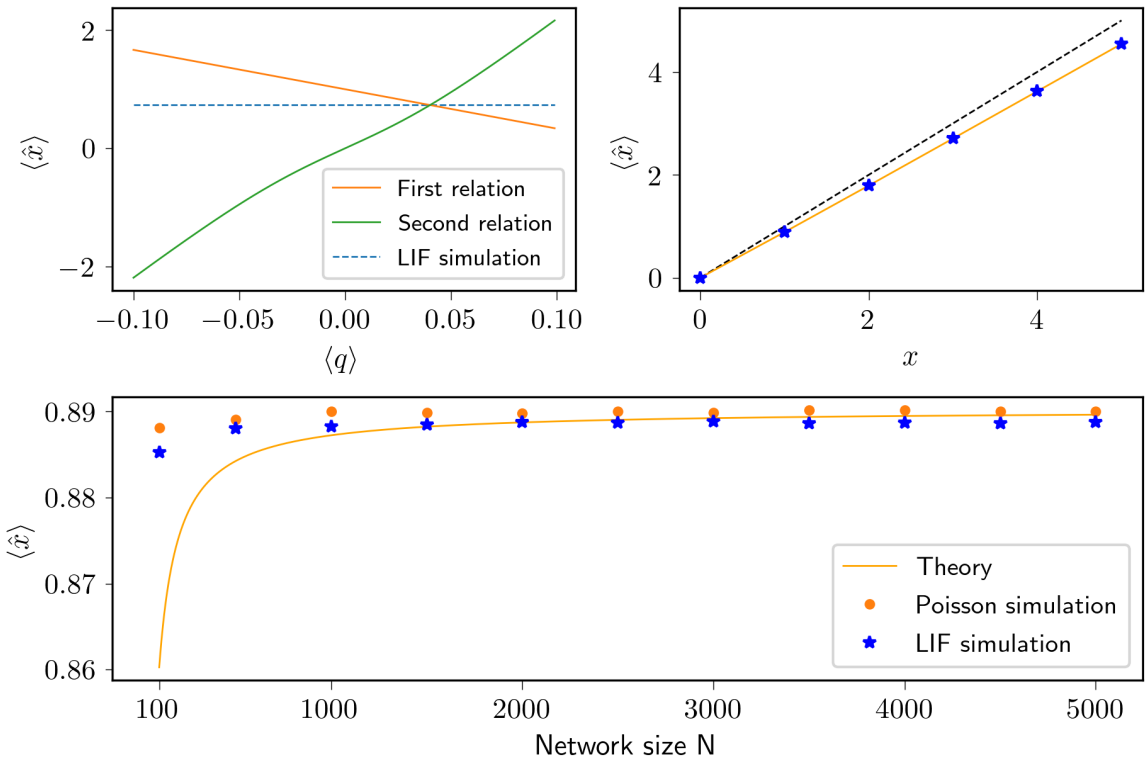

Figure 11: Analytical prediction of mean readout. Upper left plot: The mean can be obtained from 2 relations between the means of readout and mesoscopic variable q. Upper right plot: Analytical predictions excellently match the LIF results. Lower plot: For small networks, theory underestimates both Poisson and LIF readouts, though overall error remains within a 5% error margin. Accuracy is increased with the number of neurons N, where interpolates between slightly lower LIF results compared to Poisson. Parameters: $N = 1000$, $J = 5$, $\sigma_V = 0.5$.

itself. Recall that the dynamics of δq can be approximated as

$$\tau_h \delta\dot{q}(t) = -(1 + J^2\hat{b}F^{\langle\phi'\rangle})\delta q(t) - \frac{J^2\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle} + R^{\langle\phi'\rangle}\delta q(t)}\eta. \tag{59}$$

For convenience, consider a general SDE of the form

$$\tau \mathrm{d}X_t = -aX_t\mathrm{d}t + b\sqrt{c + kX_t}\mathrm{d}W_t, \tag{60}$$

which incorporates Equation (59) with $\tau = \tau_h$ and

$$a = 1 + J^2\hat{b}F^{\langle\phi'\rangle}, \qquad b = \frac{J^2\hat{b}}{\sqrt{N}},$$
$$c = R^{\langle\phi'\rangle}, \qquad k = R^{\langle\phi'\rangle}.$$

To calculate the mean and variance, we use the following "integrating factor" approach: Firstly, both sides of the equation are multiplied by an exponential factor $e^{a\tau^{-1}t}$, which yields (after some rearranging)

$$\tau(e^{a\tau^{-1}t}\mathrm{d}X_t + \tau^{-1}e^{a\tau^{-1}t}X_t\mathrm{d}t) = be^{a\tau^{-1}t}\sqrt{c + kX}\mathrm{d}W_t.$$

This implies

$$\tau \mathrm{d}\left(e^{a\tau^{-1}t}X_t\right) = be^{a\tau^{-1}t}\sqrt{c + kX}\mathrm{d}W_t,$$

which can be integrated to

$$\tau(e^{a\tau^{-1}t}X_t - X_0) = \int_0^t be^{a\tau^{-1}s}\sqrt{c + kX_s}\mathrm{d}W_s.$$

Thus, we obtain the general solution

$$X_t = \frac{e^{-a\tau^{-1}t}}{\tau}X_0 + \frac{1}{\tau}\int_0^t be^{-a\tau^{-1}(t-s)}\sqrt{c + kX_s}\mathrm{d}W_s.$$

Assume the initial condition $X_0 = 0$. Then due to the zero mean property of Itô integrals the mean of X_t is zero as well. Consequently, the second moment can be obtained as

$$\begin{aligned}
\left\langle X_t^2\right\rangle &= \left\langle \left(\frac{1}{\tau}\int_0^t be^{-a\tau^{-1}(t-s)}\sqrt{c + kX_s}\mathrm{d}W_s\right)^2\right\rangle \\
&= \frac{b^2}{\tau^2}\left\langle \int_0^t e^{-2a\tau^{-1}(t-s)}(c + kX_s)\mathrm{d}s\right\rangle, \\
&= \frac{b^2}{\tau^2}\int_0^t e^{-2a\tau^{-1}(t-s)}(c + k\left\langle X_s\right\rangle)\mathrm{d}s \\
&= \frac{b^2c}{\tau^2}\int_0^t e^{-2a\tau^{-1}(t-s)}\mathrm{d}s \\
&= \frac{b^2c}{2\tau a}(1 - e^{-2a\tau^{-1}t}),
\end{aligned}$$

where we used Itô's isometry in the second step.
In our case of stationary dynamics, we can neglect the exponential correction for a finite time in (61). Thus, we conclude the following variance for the fluctuations of q:

$$\mathrm{Var}(\delta q) = \frac{\hat{b}^2J^4R^{\langle\phi'\rangle}}{2\tau_h N(1 + J^2\hat{b}\ F^{\langle\phi'\rangle})}. \tag{61}$$

As a next step, we establish a connection between the fluctuations of q and those of $\hat{x}$. For that matter, we can re-employ the relation already used for the second identity between the means of the two variables, namely

$$\lambda\hat{x} + \dot{\hat{x}} = J\hat{b}F^{\langle\phi\rangle} - J\hat{b}F^{\langle\phi'\rangle}\delta q(t) - \frac{J\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle} + R^{\langle\phi'\rangle}\delta q(t)}\eta.$$

This time, we subtract the means from both sides of the equation which gives

$$\lambda\delta\hat{x}(t) + \delta\dot{\hat{x}}(t) = -J\hat{b}F^{\langle\phi'\rangle}\delta q(t) - \frac{J\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle} + R^{\langle\phi'\rangle}\delta q(t)}\eta. \tag{62}$$

The dynamics of δq (see Equation (59)) take a very similar form, the only two differences being an additional negative feedback of $-\delta q(t)$ as well as compared to the right-hand side of (62). This observation leads to the desired relation

$$\lambda\delta\hat{x} + \delta\dot{\hat{x}} = -\frac{1}{J\hat{b}}(\tau_h\delta\dot{q} + \delta q) \tag{63}$$

between the fluctuations of mesoscopic variable q and network prediction $\hat{x}$. To get rid of the derivatives, we make use of the Fourier theory. To this end, denote $\widetilde{y}$ the Fourier transform of a function $y(t)$. Applying the transform to both sides of (63), one obtains

$$\lambda\widetilde{\delta\hat{x}} + i\omega\widetilde{\delta\hat{x}} = -\frac{1}{J\hat{b}}\left(\widetilde{\delta q} + i\omega\tau_h\widetilde{\delta q}\right),$$

or equivalently

$$\begin{aligned}\widetilde{\delta\hat{x}}(\omega) &= -\frac{(\hat{b}J)^{-1}(1 + i\omega\tau_h)}{\lambda + i\omega}\widetilde{\delta q} \\ &= \tau_m(\hat{b}J)^{-1}\left(\frac{1 + i\omega\tau_h}{1 + i\omega\tau_m}\right)\widetilde{\delta q},\end{aligned}$$

where we used that λ is defined as the inverse of the membrane time constant τ_m. Consequently, the readout power spectral density $S_{\delta\hat{x}}(\omega)$ can be deduced from the power spectrum $S_{\delta\hat{q}}(\omega)$:

$$\begin{aligned}S_{\delta\hat{x}}(\omega) &= (2\pi)^{-1}\left|\widetilde{\delta\hat{x}}(\omega)\right| \\ &= \frac{\tau_m^2}{2\pi J^2\hat{b}}\left|\frac{1 + i\omega\tau_h}{1 + i\omega\tau_m}\right| \cdot \left|\widetilde{\delta q}(\omega)\right| \\ &= \frac{\tau_m^2}{2\pi J^2\hat{b}} \cdot \frac{1 + \omega^2\tau_h^2}{1 + \omega^2\tau_m^2}\left|\widetilde{\delta q}(\omega)\right| \\ &= \frac{\tau_m^2}{J^2\hat{b}} \cdot \frac{1 + \omega^2\tau_h^2}{1 + \omega^2\tau_m^2}S_{\delta\hat{q}}(\omega).\end{aligned}$$

Here $|z|$ denotes the absolute value of a number z, given by the product of z with its complex conjugate z^*. The variance of the readout fluctuations can be obtained from its power spectrum using Plancherel's theorem:

$$\mathrm{Var}(\delta\hat{x}) = \int_{-\infty}^{\infty} S_{\delta\hat{x}}(\omega)\frac{d\omega}{2\pi}.$$

In the special case when the time constants τ_m and τ_h are equal, the solution becomes straightforward once the variance of δq is known (see calculation above):

$$\int_{-\infty}^{\infty} S_{\delta\hat{x}}(\omega)\frac{d\omega}{2\pi} = \frac{\tau_m^2}{J^2\hat{b}}\int_{-\infty}^{\infty} S_{\delta q}(\omega)\frac{d\omega}{2\pi} = \frac{\tau_m^2}{J^2\hat{b}}\,\mathrm{Var}(\delta q).$$

Even though the calculation becomes more difficult in the general case $\tau_h \neq \tau_m$, the main idea remains the same. To exploit Plancherel's theorem for the variance of the readout fluctuations, it suffices to compute the power spectrum of δq. To this end, recall that the diffusion term in stochastic differential equation for δq (see Equation (59)) is given by $\frac{J\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle} + R^{\langle\phi'\rangle}\delta q(t)}\eta$. Here, it can be exploited that the white noise η is flat over the entire power spectrum. Moreover, as the fluctuations δq diminish with larger network size, the magnitude of the diffusion coefficient under the squared root is dominated by $R^{\langle\phi'\rangle}$. Therefore, we can approximate δq as

$$\tau_h \delta\dot{q}(t) = -(1 + J^2\hat{b}F^{\langle\phi'\rangle})\delta q(t) - \frac{J^2\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle}}\eta.$$

This describes a simple Ornstein Uhlenbeck process, characterized by a linear drift term and a constant diffusion coefficient which we denote as $\mu_q = 1 + J^2\hat{b}F^{\langle\phi'\rangle}$ and $\sigma_q := -\frac{J^2\hat{b}}{\sqrt{N}}\sqrt{R^{\langle\phi\rangle}}$, respectively. Applying the Fourier transform on both sides, it follows

$$\tau_h i\omega\widetilde{\delta q} = -\mu_q\widetilde{\delta_q}(\omega) + \sigma_q\tilde{\eta}(\omega).$$

Thus, the power spectral density of δq is given by

$$S_{\delta\hat{x}}(\omega) = \frac{1}{2\pi}\left|\widetilde{\delta q}(\omega)\right| = \left|\frac{\sigma_q\tilde{\eta}(\omega)}{\tau_h i\omega + \mu_q}\right| = \frac{\sigma_q^2}{\tau_h^2\omega^2 + \mu_q^2},$$

where we used that the power spectrum of white noise is determined by $|\widetilde{\eta}(\omega)| = \int \delta(z)e^{-i\omega z}dz = 1$. As described above, it remains to compute the variance by integrating over the power spectrum:

$$\begin{aligned}
\mathrm{Var}(\delta\hat{x}) &= \int_{-\infty}^{\infty} S_{\delta\hat{x}}(\omega)\frac{d\omega}{2\pi} \\
&= \frac{\tau_m^2}{2\pi J^2\hat{b}}\int_{-\infty}^{\infty}\frac{1+\omega^2\tau_h^2}{1+\omega^2\tau_m^2}S_{\delta\hat{q}}(\omega)\mathrm{d}\omega \\
&= \frac{\tau_m^2\sigma_q^2}{2\pi J^2\hat{b}}\int_{-\infty}^{\infty}\frac{1+\tau_h^2\omega^2}{(1+\tau_m^2\omega^2)(\mu_q^2+\tau_h^2\omega^2)}\mathrm{d}\omega.
\end{aligned}$$

The term under the integral sign can be simplified using partial fraction decomposition, with the ansatz

$$\frac{1+\tau_h^2\omega^2}{(1+\tau_m^2\omega^2)(\mu_q^2+\tau_h^2\omega^2)} = \frac{A}{1+\tau_m^2\omega^2} + \frac{B}{\mu_q^2+\tau_h^2\omega^2}.$$

The terms A and B have to satisfy

$$A(\mu_q^2 + \tau_h^2\omega^2) + B(1 + \tau_m^2\omega^2) = 1 + \tau_h^2\omega^2$$

for all ω. Canceling B by inserting $\omega = i/\tau_m$ yields

$$A = \frac{1 - \tau_h^2/\tau_m^2}{\mu_q^2 - \tau_h^2/\tau_m^2} = \frac{\tau_m^2 - \tau_h^2}{\tau_m^2\mu_q^2 - \tau_h^2}.$$

Analagously, using $\omega = i\mu_q/\tau_h$ implies

$$B = \frac{1 - \mu_q^2}{1 - \mu_q^2\tau_m^2/\tau_h^2} = \frac{(1-\mu_q^2)\tau_h^2}{\tau_h^2 - \mu_q^2\tau_m^2}.$$

Consequently, the integral can be rewritten and split as

$$\int \frac{1+\tau_h^2\omega^2}{(1+\tau_m^2\omega^2)(\mu_q^2+\tau_h^2\omega^2)}\mathrm{d}\omega = \int \left(\frac{\tau_m^2-\tau_h^2}{(\tau_m^2\mu_q^2-\tau_h^2)(1+\tau_m^2\omega^2)} + \frac{(1-\mu_q^2)\tau_h^2}{(\tau_h^2-\mu_q^2\tau_m^2)(\mu_q^2+\tau_h^2\omega^2)} \right) \mathrm{d}\omega$$
$$= \frac{\tau_m^2-\tau_h^2}{\tau_m^2\mu_q^2-\tau_h^2} \int \frac{1}{1+\tau_m^2\omega^2}\mathrm{d}\omega + \frac{(1-\mu_q^2)\tau_h^2}{\tau_h^2-\mu_q^2\tau_m^2} \int \frac{1}{\mu_q^2+\tau_h^2\omega^2}\mathrm{d}\omega.$$

The remaining two integrals can be easily solved by simple transformations. For the first integral, substituting $u = \tau_m\omega$ yields

$$\int \frac{1}{1+\tau_m^2\omega^2}\mathrm{d}\omega = \frac{1}{\tau_m}\int \frac{1}{1+u^2}\mathrm{d}u = \frac{\arctan(u)}{\tau_m} = \frac{\arctan(\tau_m\omega)}{\tau_m}.$$

Similarly, the second integral is solved via the substitution $u = \tau_h\omega/\mu_q$:

$$\int \frac{1}{\mu_q^2+\tau_h^2\omega^2}\mathrm{d}\omega = \frac{1}{\tau_h\mu_q}\int \frac{1}{1+u^2}\mathrm{d}u = \frac{\arctan(u)}{\tau_h\mu_q} = \frac{\arctan(\tau_h\omega/\mu_q)}{\tau_h\mu_q}.$$

Applying the integral limits, the solutions take the form π/τ_m and $\pi/(\tau_h\mu_q)$, respectively. Finally, we can conclude:

$$\begin{aligned}
\mathrm{Var}(\delta\hat{x}) &= \frac{\tau_m^2\sigma_q^2}{2\pi J^2\hat{b}} \left(\frac{(\tau_m^2-\tau_h^2)\pi}{(\tau_m^2\mu_q^2-\tau_h^2)\tau_m} + \frac{(1-\mu_q^2)\tau_h\pi}{(\tau_h^2-\mu_q^2\tau_m^2)\mu_q} \right) \\
&= \frac{\tau_m^2\sigma_q^2}{2J^2\hat{b}} \cdot \frac{(\tau_m^2-\tau_h^2)\mu_q + (1-\mu_q^2)\tau_h\tau_m}{\tau_m\mu_q(\tau_m^2\mu_q^2-\tau_h^2)} \\
&= \frac{\tau_m\sigma_q^2}{2J^2\hat{b}} \cdot \frac{\tau_m^2\mu_q - \tau_h^2\mu_q + \tau_h\tau_m - \mu_q^2\tau_h\tau_m}{\mu_q(\tau_m^2\mu_q^2-\tau_h^2)} \\
&= \frac{\tau_m\sigma_q^2}{2J^2\hat{b}} \cdot \frac{(\tau_m\mu_q-\tau_h)(\tau_m+\tau_h\mu_q)}{\mu_q(\tau_m\mu_q-\tau_h)(\tau_m\mu_q+\tau_h)} \\
&= \frac{\tau_m\sigma_q^2(\tau_m+\tau_h\mu_q)}{2J^2\hat{b}\mu_q(\tau_m\mu_q+\tau_h)}.
\end{aligned}$$

Note that the correction term $(\tau_m+\tau_h\mu_q)/(\tau_h+\tau_m\mu_q)$ is of order 1, i.e. a simple scalar, and disappears in the case $\tau_m = \tau_h$. Inserting the definitions of μ_q and σ_q, it follows

$$\mathrm{Var}(\delta\hat{x}) = \frac{J^2\hat{b}\tau_m R^{\langle\phi\rangle}}{2N\left(1+J^2\hat{b}F^{\langle\phi'\rangle}\right)} \cdot \frac{\tau_m+\tau_h\left(1+J^2\hat{b}F^{\langle\phi'\rangle}\right)}{\tau_h+\tau_m\left(1+J^2\hat{b}F^{\langle\phi'\rangle}\right)}. \tag{64}$$

As both $R^{\langle\phi\rangle}$ and $F^{\langle\phi'\rangle}$ are of order 1, the variance decreases as $O(1/N)$, ensuring superclassical error scaling as a direct consequence of scaling the weights as $1/N$. This is confirmed in our empirical simulations (see Figure 12, left panel), with theory closely tracking the evolution of the LIF readout variance in the case of $\sigma_V = 0.5$ and $x = 1$. Interestingly, the theoretical result predicts the LIF model with higher accuracy than the Poisson model in the case of $N = 100$ (see discussion below). In general, however, our derivations are strongly grounded on the Poisson assumption, implying that the results can only be generalized to the LIF network where this condition is reasonably satisfied. Indeed, the mean-field approximations yield the same mismatch for increasing signal strength as was observed in Section 5. While the variance of the LIF network remains roughly constant, the theory predicts a steadily increasing variance for larger stimuli (see right panel of Figure 12 as well as Figure

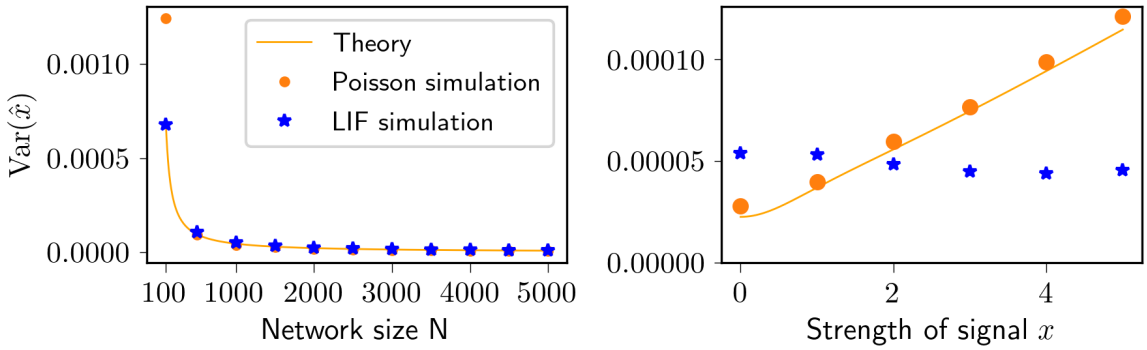

Figure 12: Mean-field predictions for the readout variance in the stationary case. Left panel: Decrease of readout variance with increasing network size for fixed target signal $x = 1$. Theory, Poisson, and LIF simulations coincide, except for the case of $N = 100$ neurons where the Poisson network exhibits larger fluctuations. Right panel: Dependence of variance on signal strength, shown for $N = 1000$ neurons. Theoretical results are marginally underestimating the variance in the Poisson model. The overall magnitude of the fluctuations roughly corresponds to the LIF results. However, the characteristically different behavior with increasing signal strength leads to a mismatch in variance with respect to both the Poisson approximation and corresponding mean-field theory. Parameters: $N = 1000$, $J = 5$, $\sigma_V = 0.5$, $\tau_h = 1$ms.

13). More precisely, the result in Equation (64) expresses a indirect dependence on the (absolute) mean through $R^{\langle\phi\rangle} = \int \xi^2 \phi\left(\xi\langle q\rangle\right) \mathcal{P}(\xi) \mathrm{d}\xi$ due to the facts that the LIF transfer function is non-saturating and $\langle q\rangle \propto \langle\hat{x}\rangle$. Consequently, for higher input strengths the fluctuations are dominated by intrinsic activity, whereas the contribution of external noise diminishes (see lower panels of Figure 13).

As a result, the analytical predictions excellently reproduce the variance of the Poisson spiking model for larger network sizes (Figure 12 and right-hand side panels in Figure 13). By contrast, the approximation becomes less reliable for small networks (e.g. for 100 neurons), where the theory tends to underestimate the variance of the Poisson network (left-hand side panels in Figure 13). Partially, this can be explained by the corresponding error produced in the means. Indeed, as the variance is very small compared to the mean (O(1/N) as opposed to O(1)), even small deviations in the mean can have a large relative effect in the estimation of the variance. On the other hand, as illustrated in Figure 13, the mismatch gets stronger at low time-scales τ_h. Opposed to that, the stationary mean is independent of the time-scale (see corresponding results above). A more plausible explanation is that the theory does not account for the surplus in firing rates observed for low N in and low time-scales (see Figure 8). Indeed, when adjusting for network size by considering $N \cdot \mathrm{Var}(\hat{x})$, the theoretical results show only a marginal change with increasing network size (Figure 13). Nuanced differences specific to small networks are not accounted for. This induces an interesting side effect: Recall that the Poisson approximation for the LIF readout is reasonably accurate in the regimes where the CV of the LIF network is close to 1 (see Section 5). As the error in firing rates introduces an additional error in this approximation for small Poisson networks, the theoretical result gives a more reliable estimation in these cases. On the other hand, the theory for the Poisson model requires an additional correction for the fluctuations which arise due to finite size effects in small networks.

Subtle changes in the variance for increasing N arise through differences in the LIF transfer function ϕ. Note that both the threshold $\vartheta_i = \frac{1}{2}\left(\frac{J^2\hat{b}}{N}\xi_i^2 + \nu + \mu\right)$ and the reset $V_R = \vartheta - \mu$, for instance, depend on the scaled weights. Consequently, the LIF transfer function and

the corresponding terms $R^{\langle\phi\rangle}$ and $F^{\langle\phi'\rangle}$ change implicitly with N. Similarly, other changes to model parameters are automatically adjusted for through the LIF transfer function. For instance, while the strength of external noise σ_V does not appear explicitly in Equation (64), the analytical results reliably describe its effect on the readout variance in the Poisson network (see lower panels in Figure 13). This illustrates how the mean-field theory can in principle be used for a systematic analysis of the Poisson model behavior in different environments. Specifically, it eliminates the need for extensive empirical studies and grid search for parameter optimization, as well as descriptive analysis.

7 Conclusion and Outlook

In summary, we introduce a new theoretical framework, leading to several insights for spiking neural networks operating in a functional regime of predictive coding.
To this end, we proceed in several steps. First, we propose several extensions for the renowned predictive coding model. Most notably, we present a general framework that allows to freely adjust the level of balance between excitation and inhibition.
As a result, we can study the network behavior within a loosely balanced regime, which has not been previously considered for the given network. In contrast to more restrictive regimes such as classical or tight balance, loose balance allows for an interesting alternative approach in the coding of the prediction error. In particular, the network can produce a systematic, stable bias in the prediction, while diminishing variance for larger network sizes. Such a strategy might be advantageous, as a well-known bias can easily be removed, leading to unvaryingly high precision at lower metabolic costs.
Secondly, we introduced a spiking Poisson model which closely replicates both functional structure and firing patterns of the original integrate-and-fire dynamics. By producing spikes from instantaneous firing rates, the Poisson model combines both analytical tractability and realistic output in form of neuronal spike trains. Thus, the results are directly comparable to the LIF model, closing the gap between purely theoretical rate models and spiking models, based on experimental frameworks closer to biology.
In simulations, the Poisson model shows a good approximation of the LIF network for various time-dependent inputs. The quality of the fit can be further increased by tuning the time-scale of the Poisson model, which arises as an additional free parameter.
However, for strong inputs the LIF model gradually loses its variability, leading to systematically lower readout variance than produced by the Poisson counterpart. Still, even in those cases the mean readout is perfectly tracked.
Using a Gaussian approximation of the Poisson dynamics and an orthogonal projection on the low-dimensional subspace spanned by the structured connectivity, we derive general mean-field dynamics of the predictive coding network. These reveal a strong multiplicative dependency of the overall fluctuations on the total recurrent input. In particular, the noise intensity is proportional to the scaling of the weights, given by the chosen amount of balance. Only the loosely balanced regime, where the magnitude of both external drive and lateral responses is small, ensures that fluctuations vanish in the large N limit.
In that case, we show that the mean-field approximation can be directly applied to compute the mean-squared error of the readout prediction for realistic, mesoscopic network sizes. The analytical results are confirmed in empirical simulations of the Poisson network, with accuracy quickly increasing with the number of neurons. In effect, the mean readout is perfectly predicted for the LIF network, and variance is accurately approximated as long as spiking variability is close to Poisson. As a concrete result, we demonstrate that the loosely balanced regime indeed leads to superclassical scaling of the error, in contrast to previous results based on theoretical firing rates only. This result rehabilitates loose dynamics as a credible and promising alternative to the predominant use of tight balance in the context of

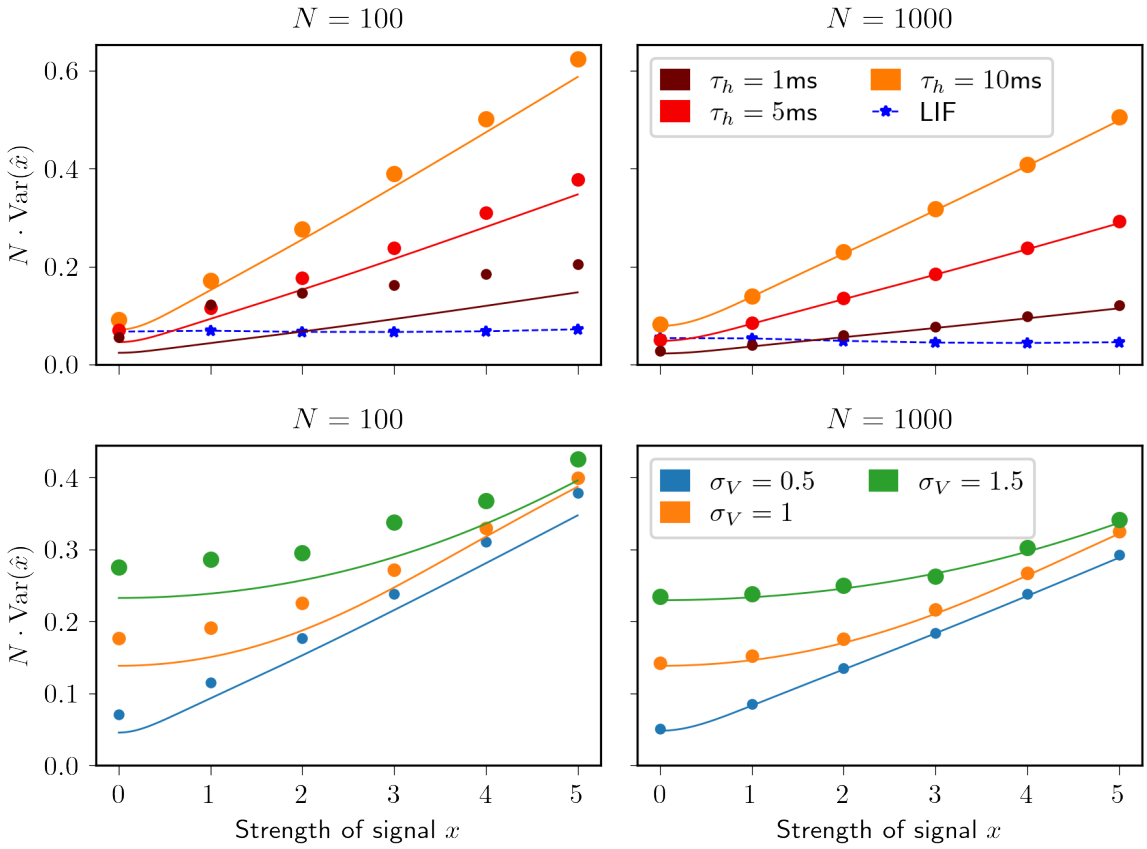

Figure 13: Dependence of readout variance on network size N, time-scale τ_h and noise strength σ_V in theory and Poisson simulations. In all plots, the variance was scaled with N, leading to results of the same magnitude (shared y-axis in each row). Upper panels: Dependence of τ_h. For a small network size of $N = 100$ neurons, the Poisson readout shows larger fluctuations than predicted by the theory (left panel). The mismatch is most pronounced for $\tau_h = 1$ms, decreasing at larger time-scales. For a larger network of $N = 1000$ neurons (right panel), the mismatch between theory and simulation disappears. In both cases, the variance increases linearly after a certain threshold near $x = 0$. The slope becomes slightly smaller for the larger network. Results are shown for $\sigma_V = 0.5$. Lower panels: Same for varying external noise σ_V, with timescale fixed to $\tau_h = 5$ms. The mismatch between theory and simulations appears in the same way as before. Otherwise, the noise strength has mainly an influence on the variance for low and intermediate input strengths. Larger σ_V leads to larger fluctuations and more robustness with respect to increasing stimuli. Eventually, however, the variance starts to increase linearly and the differences in fluctuations due to external noise diminish. As before, the growth rates of the (normalized) variance become slightly smaller for $N = 100$. Parameters: $J = 5$.

highly efficient coding in the brain.
Various directions can be explored in future research. From the theoretical point of view, an important extension would be to incorporate the general case where the network can represent an arbitrary linear dynamical system (in particular, with non-diagonal state transition matrix A in the defining Equation (5)). Similarly, an in-depth analysis of the multi-dimensional case might reveal richer dynamics such as multiple steady-states, limit cycles, or chaos. In the two-dimensional case, this can be done by phase plot analysis as performed by [39] for memory engrams. Specifically, one could investigate how the dynamics change in the case where the number of inputs M is large and comparable with the number of neurons.
Even in the one-dimensional case, numerous interesting theoretical questions can be addressed. Similar to Kadmon et al. [16], one could analyze the effects of synaptic delays, weight disorder and other biological disturbances. Theory and simulations can be extended to general time-varying and noisy inputs, for which the optimal choice of time-scale τ_h could be of significant importance. Moreover, explicit results for bias and variance could be derived through perturbation methods [40]. Lastly, alternative approaches towards mean-field results can be explored, in particular with respect to different scaling and balance regimes that are not incorporated in the current framework.
On the other hand, to obtain more valuable insights about real neural networks, it is crucial to reconcile simplistic computational models with biology. A major limitation of our approach is that we impose Poisson-like spiking dynamics, which can prove too restrictive to fully explain not only the LIF characteristics in the original model but biologically plausible neuronal activity in general. Our work shows promising results using a spiking network based on firing rates on a microscopic scale. Possibly, more sophisticated approaches could be found using a similar idea but without explicitly imposing Poisson firing statistics. Furthermore, various modifications can be incorporated into the model, transforming it towards biological realism. For instance, weights should be either excitatory or inhibitory, which can be achieved functionally by introducing separate objective functions for the two populations. Recently, a new idea has been proposed to loosen the rigidity in the decoder weights by introducing synaptic plasticity and Hebbian learning [14]. Moreover, connections should be sparse and respect biologically plausible synaptic timings (e.g. instead of instantaneous delta synapses). Ultimately, combining theory and simulation-driven computational models could lead to powerful insights and testable predictions, which can be evaluated on real data.

References

[1] Zachary F Mainen and Terrence J Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503–1506, 1995.

[2] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4):292–303, 2008.

[3] John Hertz, Anders Krogh, and Richard G Palmer. *Introduction to the theory of neural computation*. CRC Press, 2018.

[4] Carl Van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.

[5] Nicolas Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8(3):183–208, 2000.

[6] Martin Boerlin, Christian K. Machens, and Sophie Denève. Predictive coding of dynamical variables in balanced spiking networks. *PLOS Computational Biology*, 9(11):1–16, 2013.

[7] Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

[8] Luc H Arnal, Valentin Wyart, and Anne-Lise Giraud. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, 14(6):797–801, 2011.

[9] Georg B Keller, Tobias Bonhoeffer, and Mark Hübener. Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron*, 74(5):809–815, 2012.

[10] Nicolas Giret, Joergen Kornfeld, Surya Ganguli, and Richard HR Hahnloser. Evidence for a causal inverse model in an avian cortico-basal ganglia circuit. *Proceedings of the National Academy of Sciences*, 111(16):6063–6068, 2014.

[11] Matthew Chalk, Boris Gutkin, and Sophie Deneve. Neural oscillations as a signature of efficient coding in the presence of synaptic delays. *Elife*, 5:e13824, 2016.

[12] David GT Barrett, Sophie Deneve, and Christian K Machens. Optimal compensation for neuron loss. *Elife*, 5:e12454, 2016.

[13] Veronika Koren and Sophie Denève. Computational account of spontaneous activity as a signature of predictive coding. *PLOS Computational Biology*, 13(1):1–34, 2017.

[14] Wieland Brendel, Ralph Bourdoukan, Pietro Vertechi, Christian K Machens, and Sophie Denéve. Learning to represent signals spike by spike. *PLOS Computational Biology*, 16(3):e1007692, 2020.

[15] Camille E Rullán Buxó and Jonathan W Pillow. Poisson balanced spiking networks. *bioRxiv*, page 836601, 2019.

[16] Jonathan Kadmon, Jonathan Timcheck, and Surya Ganguli. Predictive coding in balanced neural networks with noise, chaos and delays. *Advances in Neural Information Processing Systems*, 33, 2020.

[17] Sophie Denève and Christian K Machens. Efficient codes and balanced networks. *Nature Neuroscience*, 19(3):375–382, 2016.

[18] Yashar Ahmadian and Kenneth D Miller. What is the dynamical regime of cerebral cortex? *Neuron*, 2021.

[19] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Computational Neuroscience Series, 2001.

[20] Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[21] Louis Lapique. Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization. *Journal of Physiology and Pathololgy*, 9:620–635, 1907.

[22] Alexander Rauch, Giancarlo La Camera, Hans-Rudolf Luscher, Walter Senn, and Stefano Fusi. Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo-like input currents. *Journal of Neurophysiology*, 90(3):1598–1612, 2003.

[23] AL Hodgkin and AF Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bulletin of Mathematical Biology*, 52(1-2):25–71, 1990.

[24] Benjamin Lindner. A Brief Introduction to Some Simple Stochastic Processes. In *Stochastic Methods in Neuroscience*. Oxford University Press, Oxford, 2009.

[25] W R Softky and C Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of Neuroscience*, 13(1):334–350, 1993.

[26] W Bair, C Koch, W Newsome, and K Britten. Power spectrum analysis of bursting cells in area mt in the behaving monkey. *Journal of Neuroscience*, 14(5 Pt 1):2870–2892, 1994.

[27] EJ Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12(2):199, 2001.

[28] Eero P Simoncelli, Liam Paninski, Jonathan Pillow, Odelia Schwartz, et al. Characterization of neural responses with stochastic stimuli. *The Cognitive Neurosciences*, 3(327-338):1, 2004.

[29] Srdjan Ostojic and Nicolas Brunel. From spiking neuron models to linear-nonlinear models. *PLOS Computational Biology*, 7(1):e1001056, 2011.

[30] Arnold JF Siegert. On the first passage time probability problem. *Physical Review*, 81(4):617, 1951.

[31] Nicolas Brunel and Vincent Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, 11(7):1621–1671, 1999.

[32] Nicolas Fourcaud and Nicolas Brunel. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Computation*, 14(9):2057–2110, 2002.

[33] John C Eccles, P Fatt, and K Koketsu. Cholinergic and inhibitory synapses in a pathway from motor-axon collaterals to motoneurones. *The Journal of Physiology*, 126(3):524–562, 1954.

[34] Bilal Haider, Alvaro Duque, Andrea R Hasenstaub, and David A McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, 26(17):4535–4545, 2006.

[35] Itamar Daniel Landau and Haim Sompolinsky. Coherent chaos in a recurrent neural network with structured connectivity. *PLOS Computational Biology*, 14(12):e1006309, 2018.

[36] Constance Hammond. Chapter 19 - the adult hippocampal network. In Constance Hammond, editor, *Cellular and Molecular Neurophysiology (Fourth Edition)*, pages 393–409. Academic Press, Boston, fourth edition edition, 2015.

[37] Shun-Ichi Amari. Characteristics of random nets of analog neuron-like elements. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-2(5):643–657, 1972.

[38] Daniel J Amit, Hanoch Gutfreund, and Haim Sompolinsky. Spin-glass models of neural networks. *Physical Review A*, 32(2):1007, 1985.

[39] Chiara Gastaldi, Tilo Schwalger, Emanuela De Falco, Rodrigo Quian Quiroga, and Wulfram Gerstner. When shared concept cells support associations: theory of overlapping memory engrams. *bioRxiv*, 2021.

[40] Carl M Bender and Steven A Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*. Springer Science & Business Media, 2013.

Selbstständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbstständig verfasst und noch nicht für andere Prüfungen eingereicht habe. Sämtliche Quellen, einschließlich Internetquellen, die unverändert oder abgewandelt wiedergegeben werden, insbesondere Quellen für Texte, Grafiken, Tabellen und Bilder, sind als solche kenntlich gemacht. Mir ist bekannt, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung eingeleitet wird.

Berlin, den 28. Oktober 2021.

Alexander Spokoinyi