Encoding symbolic sequences with spiking neural reservoirs

Renato Duarte[1,*], Marvin Uhlmann[2], Dick van den Broek[2], Hartmut Fitz[2,3], Karl Magnus Petersson[2,4], Abigail Morrison[1,5]

[1]INM-6 / IAS-6 / INM-10, Jülich Research Centre, Jülich, Germany
[2]Max Planck Institute for Psycholinguistics Nijmegen, the Netherlands
[3]Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, the Netherlands
[4]Centre for Biomedical Research (CBMR), University of Algarve, Portugal
[5]Institute of Cognitive Neuroscience, Ruhr-University Bochum, Germany
[*]E-mail: r.duarte@fz-juelich.de

Abstract—**Biologically inspired spiking networks are an important tool to study the nature of computation and cognition in neural systems. In this work, we investigate the representational capacity of spiking networks engaged in an identity mapping task. We compare two schemes for encoding symbolic input, one in which input is injected as a direct current and one where input is delivered as a spatio-temporal spike pattern. We test the ability of networks to discriminate their input as a function of the number of distinct input symbols. We also compare performance using either membrane potentials or filtered spike trains as state variable. Furthermore, we investigate how the circuit behavior depends on the balance between excitation and inhibition, and the degree of synchrony and regularity in its internal dynamics. Finally, we compare different linear methods of decoding population activity onto desired target labels. Overall, our results suggest that even this simple mapping task is strongly influenced by design choices on input encoding, state-variables, circuit characteristics and decoding methods, and these factors can interact in complex ways. This work highlights the importance of constraining computational network models of behavior by available neurobiological evidence.**

Index Terms—**stimulus encoding, reservoir computing, balanced random networks**

I. Introduction

Spiking neural networks are the computational substrate of neurobiological systems. The properties of observed population responses and their suitability for processing is dependent on the complexity of the underlying neuronal biophysics, the details of excitatory and inhibitory synaptic transmission and the recurrent interactions that they mediate. These local features influence the responsiveness of the circuit and determine how specific spatio-temporal activity patterns emerge from the interaction between ongoing, internally generated activity and external inputs. Input sensitivity and selectivity determine whether the circuit state can be transferred from one subspace to another at any point in time (real-time computing) in response to relevant input features.

Apart from the nature of input signals, the manner in which these are delivered to the circuit might play a critical role. In real brains, only the peripheral systems have access to direct external input as a mechanical or chemical signal from the environment. These systems are endowed with highly specialized, modality-specific adaptations, that allow them to convert external signals into sequences of spike trains. Any subsequent processing stage then receives, processes and transmits signals as a complex spatio-temporal sequence of spike trains. Thus, spatio-temporal spike patterns are the language of the system, i.e., its natural inputs and outputs. In computational studies, however, input signals are often represented as direct, somatically injected currents. This approach simplifies the encoding process given that the input signal has a direct impact on the circuit dynamics. This is particularly true if there is no recurrence in the system which then becomes a passive non-linear encoder [7]. Recurrence is a ubiquitous design feature of neurobiological networks (particularly cortical microcircuits), which can lead to complex and unpredictable responses, and its functional role is not fully understood. It complicates the encoding process, and has made the computational analysis of recurrent spiking networks difficult (but see, e.g. [1], [4], [11]). Ideally, for biological compatibility, the encoding process should be mediated by spatiotemporal spike patterns and should drive the circuit to operate in a physiologically plausible regime.

Additionally, the choice of state variables to characterize the representational properties of spiking networks is also important. These states can then be linearly combined to extract relevant information and solve a given computational task. It is common to calibrate the readouts on an instantaneous rate estimate, obtained by convolving each neuron's spike trains with an adequate kernel (typically exponential). However, as we have argued in [13] and set out to quantify in this study, this approach may introduce biases in that the parameters of the filter kernel are arbitrarily chosen and may influence the results. Using membrane potentials instead provides access to the information contained in the circuit in a more natural and unbiased way.

In order to probe the computational capabilities of a circuit in relation to its intrinsic properties, it is important not to introduce confounding factors. Thus, the decoder itself should be memoryless and only able to access traces of past inputs that persist in the system's dynamics. The choice of a spike filter kernel and its associated time constant can distort results

as it imposes its own timescale. Training a readout directly on neuronal state variables, such as the membrane potential, avoids this issue.

The main goal of this study is to better understand how different encoding strategies influence the circuit's response properties, interpreted in light of known physiological constraints, and how these strategies influence the characteristics of the high-dimensional population dynamics in response to multiple input stimuli.

II. METHODS

In order to focus primarily on the encoding and decoding processes and isolate their effects from the circuit properties, we used relatively simple, homogeneous and static neuronal circuits, tuned to operate in a balanced state under stationary conditions. To establish suitable benchmarks and to ensure that variations in computational performance were solely related to the parameters of encoding and decoding, no form of long-term adaptation or learning was introduced in any of the circuit's components.

A. Task specification

We define a computational task as the problem of learning a mapping between an input sequence $\mathbf{u}[n] \in \{0,1\}^{N_u}$ and a target output sequence $\mathbf{y}[n] \in \{0,1\}^{N_y}$, both viewed as discrete N^u- or N^y-dimensional time series of binary feature vectors (orthogonal stimulus representations), with finite length T ($n = 1, ..., T$). In this study, we focus on the simplest mapping where inputs and targets were identical, i.e. $\mathbf{y} = \mathbf{u}$. This task measures the network's ability to map N^u distinct input stimuli onto linearly separable internal states and this ability is the basis for more complex cognitive functions that can involve sequential dependencies. In language processing, for example, stimuli would correspond to word representations activated by auditory or visual input. The capacity to separate these discrete units of information in representational space is critical in order to compute the compositional meaning of an utterance. Hence, our approach provides a first step towards systematically relating circuit features to a fundamental property of cognitive systems.

B. Signal transduction and encoding

The signal transduction process (illustrated in Figure 1) converts a discrete, binary input sequence $\mathbf{u}$ into a continuous input signal that is then delivered to the circuit through some appropriate encoding strategy. Each input stimulus is active for a fixed duration of $d^u = 200$ ms with no added inter-stimulus interval. Although variations in stimulus duration and inter-stimulus intervals, as well as their amplitudes and kinetics, can play an important role in modulating population responses and computational performance, we did not explore their impact in this study. Instead, we focused on the encoding process *per se* by which the input signal was mapped into the circuit. The features and parameters of input transduction were thus fixed and chosen to maximize the information content in the input space while avoiding the introduction of additional

variation. For simplicity, we also disregarded potential sources of top-down information and focused on the characteristics of stimulus representation and processing in purely input-driven circuits.
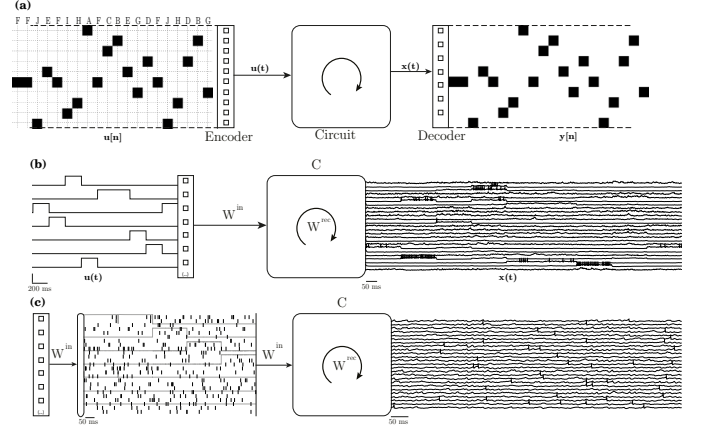

Fig. 1. Schematic representation of the set-up used throughout this manuscript. **(a)** A symbolic, discrete time series $\mathbf{u}$ is converted to a continuous signal $u(t)$, which is used to drive the circuit C. Population activity, $x(t)$ in response to input $u(t)$ recorded (see example traces in **(b, c)**), sampled and used by a linear decoder to classify the identity of each stimulus, producing the discrete output time series $\mathbf{y}$. **(b)** - Example of direct encoding, where $N^u = 8$ stimuli are injected into the circuit as a direct input current. **(c)** - Example of indirect encoding, where each stimulus is converted to a unique spatio-temporal spike pattern. Example traces of population responses (the circuit state, $x(t)$) are depicted on the right-hand side.

We examined two different encoding schemes:

a) Direct encoding: the signal was scaled by ρ_u, which specifies the current amplitude to be injected into the target neurons. Each input channel was then mapped to a random subset of neurons (determined by the structure of the input weights $W^{\text{in}} \in \mathbb{R}_+^{N_u \times N}$). In this way, each neuron's membrane received one additional input source (see Figure 1**b** and Equation 2):

$$I_{\text{in}}(t) = \rho_u W^{\text{in}} u(t) \tag{1}$$

b) Indirect encoding: the input signal was converted into a set of distributed spike trains across a population of N_{in} input neurons. We assumed that the network receives a large number of external input spike trains, conveyed through excitatory synapses. Thus, we fixed the number of encoding neurons to be equal to the number of recurrent, excitatory neurons, i.e. $N_{\text{in}} = N_E$ and treated all input synapses as excitatory. The scaling parameter ρ_u was used to control the average firing rate across the input population. Each stimulus was then converted into a fixed, spatio-temporal spike pattern across N_{in} neurons. In other words, each stimulus was an instance of 'frozen noise', composed of N_{in} independent Poisson processes, at a rate ρ_u and with a duration d^u ms.

For both encoding schemes, the input weight matrix W^{in} was sparse and random, with density $p^{\text{in}} = p^{\alpha E} = 0.1$ (see below). To add some variability to the input mapping, each input weight w_{ij}^{in} was randomly drawn from a normal

distribution with $\mu^{\text{in}} = 1$ and $\sigma = 0.5\mu^{\text{in}}$, truncated to the interval $[0.0001, 10\mu^{\text{in}}]$.

C. Circuit specification

Each circuit C, consisted of a random, sparse and recurrently coupled network of N = 10000 homogeneous, adaptive exponential integrate-and-fire neurons, sub-divided into $N_E = 0.8$ N excitatory and $N_I = 0.2$ N inhibitory populations. All input and recurrent connections, W^{in} and W^{rec}, were sparse, with density p^{syn} and random, with values drawn independently from truncated normal distributions: $w_{ij}^{\text{syn}} \sim \mathcal{N}(\mu^{\text{syn}}, 0.5\mu^{\text{syn}})$, with $w_{ij}^{\text{syn}} \in (0.0001, 10\mu^{\text{syn}})$. The same was true for recurrent connection delays (d^{syn}): $d_{ij}^{\text{syn}} \sim \mathcal{N}(d^{\text{syn}}, 0.5d^{\text{syn}})$, with $d_{ij}^{\text{syn}} \in (0.1, 10d^{\text{syn}})$. Connection parameters were fixed and determined according to the nature of the presynaptic population, i.e., whether the connection was excitatory or inhibitory (see Table I).

TABLE I
RECURRENT CONNECTIVITY PARAMETERS.

Parameter	Value	Description
$p^{\alpha E}$	0.1	excitatory connection probability
$p^{\alpha I}$	0.2	inhibitory connection probability
$\mu^{\alpha E}$	1.2 nS	mean excitatory synaptic strength
$\mu^{\alpha I}$	$\gamma\mu^{\alpha E}$ nS	mean inhibitory synaptic strength
γ	15, 11	E/I balance parameter (direct, indirect)
$d^{\alpha E}$	1.2 ms	mean excitatory connection delay
$d^{\alpha I}$	0.8 ms	mean inhibitory connection delay

As a compromise between model complexity and biophysical realism, we used the adaptive exponential integrate-and-fire (AdEx) neuron [3], a reduced, two-dimensional point neuron model that combines an exponential spiking mechanism with an intrinsic adaptation current that has a sub-threshold and a spike-triggered component:

$$C_m \frac{dV_i}{dt} = -G_L (V_i - E_L) + G_L \Delta_T \exp\left(\frac{V_i - V_T}{\Delta_T}\right) - w_i - \sum_{j \in \text{pre}} I_{ij}^{\text{syn}} \quad (2)$$

$$\tau_w \frac{dw_i}{dt} = -w_i + a (V_i - E_L) + \tau_w b \sum_{t_f \in F(i)} \delta(t - t_f) \quad (3)$$

Neuron i emits a spike at times t_f when the membrane potential V_i diverges towards infinity, capturing the upswing of the action potential (parameterized by V_T and Δ_T). The corresponding spike train is defined as $S_i(t) = \sum_{t_f} \delta(t - t_f)$. At these times, the membrane potential is reset to a constant reset potential V_{reset} for $t \in (t_f, t_f + t_{\text{refr}}]$, after which integration is resumed as above. The parameters $a = 4$ nS and $b = 80.5$ mV determine the relative contribution of sub-threshold and spike-triggered adaptive processes, respectively. The time constant of intrinsic adaptation is set to $\tau_w = 144$ ms, following [8] (see Table II).

TABLE II
NEURON PARAMETERS.

Parameter	Value	Description
C_m	250.0 pF	membrane capacitance
E_L	−70.0 mV	resting membrane potential
V_{th}	−50.0 mV	fixed firing threshold
V_{reset}	−60.0 mV	reset potential
G_L	16.7 nS	leak conductance
Δ_T	2.0	threshold sharpness parameter
a	4.0 nS	sub-threshold intrinsic adaptation
b	80.5 mV	spike-triggered intrinsic adaptation
t_{ref}	2.0 ms	absolute refractory time
τ_w	144.0 ms	adaptation time constant

Synaptic transmission is modelled as a conductance-based process, where each presynaptic spike triggers an instantaneous rise of amplitude w_{ij}^{syn} in the corresponding conductance, followed by an exponential decay with time constant τ_{syn}:

$$\tau_{\text{syn}} \frac{dg_{ij}^{\text{syn}}(t)}{dt} = -g_{ij}^{\text{syn}}(t) + w_{ij}^{\text{syn}} \sum_{t_f \in F(j)} \delta\left(t - t_f - d_{ij}^{\text{syn}}\right) \quad (4)$$

The total synaptic current I_{ij}^{syn}, generated by inputs from presynaptic neuron j and mediated by synapse type $\text{syn} \in \{E, I\}$, is thus given by:

$$I_{ij}^{\text{syn}}(t) = g_{ij}^{\text{syn}}(t) [V_i(t) - E_{\text{syn}}] \quad (5)$$

where $E_E = 0$ mV and $E_I = -75.0$ mV are the reversal potentials for excitatory and inhibitory synapses and $\tau_E = 2$ ms, $\tau_I = 6$ ms are the corresponding time constants. All remaining synaptic and connectivity parameters are specified in Table I).

D. Characterizing population activity

In order to quantify and compare the system's operating point in various conditions, we employ a set of metrics that, taken together, provide a statistical characterization of the macroscopic features of population activity in response to input. Of particular interest is the degree of regularity and synchrony at the population level, which we quantify in the form of a summarized, numerical description of the system's operating point and how it varies during stimulus processing. For this purpose, we use several complementary parameter-free metrics, grouped by the feature they are meant to quantify. These provide global measures of regularity and synchrony, computed as the mean absolute distance between the measured population results ($\hat{\theta}$) and the values that would be obtained by N stationary Poisson processes at the same rate, which we consider as the 'ground truth' for the desired asynchronous irregular state:

$$I^{\alpha} = \frac{1}{c} \sum_{i=1}^{c} (\theta_i - \hat{\theta}_i) \quad (6)$$

where I^{α} represents the global index summarizing the results obtained for metric set α (e.g., synchrony, regularity, etc.), composed of c different metrics:

a) Regularity (I^{reg}): to determine the (ir)regularity of population activity, we employ a set of metrics consisting of reg $= \{\text{CV}_{\text{ISI}}, \text{LV}_{\text{ISI}}\}$, whose expected values for a pure Poisson process are $\theta_{\text{reg}} = [1, 1]$.

b) Synchrony (I^{sync}): we consider the set of metrics sync $= \{\text{D}^{\text{ISI}}, \text{D}^{\text{SPIKE}}, \text{C}^{\text{SPIKE}}\}$, whose expected values are $\theta_{\text{sync}} = [0.5, 0.3, 0.25]$, respectively (see [12] for details about these metrics and their implementation).

c) E/I Ratio: to quantify the degree of balance between excitation and inhibition, we record the difference in the mean amplitude of each neuron's excitatory and inhibitory input currents $\langle I_E - I_I \rangle$ and the average correlation coefficient between them. The set of metrics is, in this case, $\text{EI} = \{\langle I_E - I_I \rangle, \text{CC}(I_E, I_I)\}$ and the target values correspond to a purely balanced condition, $\theta_{\text{EI}} = [0, 1]$.

Thus, to obtain a summary statistic of the population state, we use a compound distance metric that quantifies the amount of asynchrony and irregularity in population activity (as the distance between the data vector and that obtained for a collection of pure Poisson processes), as well as the global balance of excitation and inhibition.

E. State decoding and readout

The state of the circuit, which we generically refer to as $x(t)$, consists of any of a number of adequate dynamical variables that describe the system. As mentioned in section I, it is common to use a continuous rate estimate obtained by low-pass filtering the spike trains (henceforth denoted by $S(t)$), but this option may introduce biases through the parameters of the filter kernel. A natural, unbiased alternative is to consider $x(t) = V(t)$, i.e., to readout directly from the membrane potentials (Equation 2).

Since input and target output correspond to symbolic labels and/or binary feature vectors, the continuous state variable $x(t)$ was downsampled. For each stimulus, a representative of the population response was gathered in a vector $\mathbf{x}[n]$ that corresponded to the population activity $x(t^*)$ sampled at stimulus offset t^*. These vectors were then combined into a state matrix $X \in \mathbb{R}^{N \times T}$. Unless otherwise stated, the total number of stimulus samples used was fixed at 10000.

To asses the quality of the input-state mappings for each experiment, we determined the capacity to linearly combine the input-driven population responses to approximate the desired output [10]. Since readouts are linear and feedforward, this is equivalent to:

$$\hat{Y} = W^{\text{out}} X \tag{7}$$

where $\hat{Y} \in \mathbb{R}^{N^y \times T}$ and $X \in \mathbb{R}^{N \times T}$ are the collection of all outputs and corresponding states over all time steps T, W^{out} is a $N \times N^y$ matrix of output weights from the network to the readout units, chosen to minimize the quadratic error between the readout's output $\hat{y}$ and the desired target values $\mathbf{y}$. Finding the optimal W^{out} that minimizes the error (training the readout) amounts to solving:

$$W^{\text{out}} = Y X^T \left(X X^T + \beta I \right)^{-1} \tag{8}$$

where standard linear regression is a special case ($\beta = 0$). The regularization parameter β penalizes solutions with large norm, which would correspond to a selective amplification of specific dimensions of $\mathbf{x}[n]$ which increases the risk of overfitting and large out-of-sample error. The readout norm $\|W^{\text{out}}\|$ thus also provides a metric to quantify the stability and generalizability of the solution found on the training data. Throughout this manuscript, we used ridge regression where β was chosen by leave-one-out cross-validation on the training set. For comparison, we also tested the solutions found by standard linear regression, using the Moore-Penrose pseudo-inverse method.

The readout was typically trained on a representative set comprising $T_{train} = 0.8T$ input-output samples in order to generate testable input-output relations for the remaining data set $T_{test} = 0.2T$. The accuracy of the solution was quantified by measuring readout error on the test data as:

$$E(\hat{y}, y) = \sum_{i=1}^{N_y} \left(\sum_{n=1}^{T} (y_i[n] - \hat{y}_i[n])^2 \right) \tag{9}$$

In addition, given the symbolic nature of the tasks, we also determined the accuracy in terms of a categorical decision, using a winner-takes-all function over the N^y readout dimensions.

F. Numerical Simulations and Analyses

All numerical simulations were run using NMSAT[1] [6], a high-level wrapper for NEST customized for the application of reservoir computing principles to complex microcircuits. NEST version 2.12.0 [9] was used for all the numerical simulations. A complete code and data package is available upon request.

III. RESULTS

Based on previous work, we expected that different stimulus encoding schemes would lead to substantial differences in the characteristics of population responses which, in turn, might influence the circuit's computational properties [5]. Likewise, post-processing steps commonly used to acquire population responses (e.g., filtering of spiking activity), might introduce biases that modulate performance, without reflecting the processes occurring in the circuit.

Throughout this study, we followed a generic procedure: generate microcircuit C (subsection II-C), inputs and target outputs ($\mathbf{u}[n], \mathbf{y}[n]$) (subsection II-A); build and connect the input layers (subsection II-B); set circuit parameters that allow it to operate in a suitable regime (see below); drive the system with the input for the entire length $T = 10000$ (corresponding to a total simulated time of 2000 s), discarding an added initial transient phase ($T_{\text{transient}} = 100$); gather all internal state vectors ($\mathbf{x}[n]$) generated by the population in response to each stimulus; use a subset of the data ($T_{\text{train}} = 0.8T$) to train the readouts and a disjoint subset ($T_{\text{test}} = 0.2T$) for testing (subsection II-E).

[1]github.com/rcfduarte/nmsat

Circuit dynamics

The quality of the input-state mappings and the underlying circuit dynamics may be dependent on the circuit's specifications but also on how the input is delivered and on how the circuit interacts with it. In order to control for these factors and obtain comparable dynamics in the different encoding regimes, we independently tuned the input amplitude parameter ρ_u, along with the global ratio of inhibition to excitation in the network γ, for each of the two encoding schemes (see Figure 2).
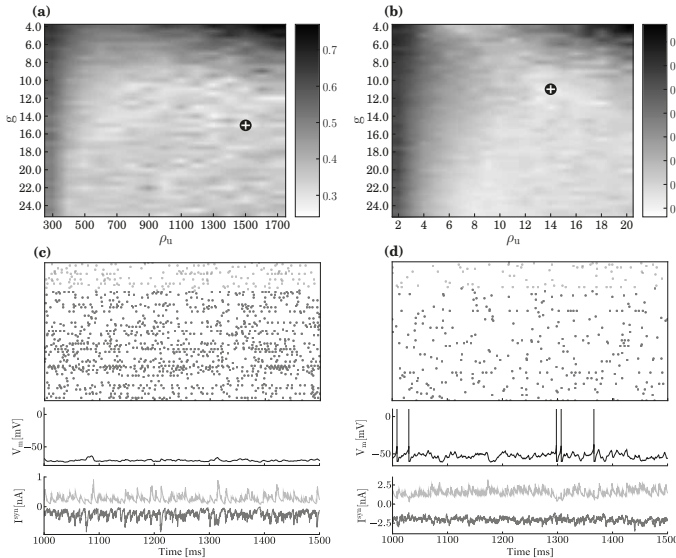
Fig. 2. Noise-driven circuit dynamics for direct (left) and indirect (right) encoding. **(a, b)** - The input amplitude parameter ρ_u was varied in combination with the E/I balance parameter g to obtain comparable macroscopic statistics across the population (low-rate asynchronous irregular activity with globally balanced excitation and inhibition). **(c, d)** - Illustrative examples of population activity (over a period of 500 ms) using the chosen parameter combination indicated as plus marker in (a, b): spiking activity of 100 randomly chosen neurons (top), membrane potential of one randomly chosen excitatory neuron (middle) and total excitatory and inhibitory synaptic currents onto the same neuron (bottom).

To tune the systems to operate in a low-rate asynchronous irregular state with balanced excitation and inhibition (subsection II-D), we explored the characteristics of population responses when driven by stationary, unspecific and stochastic input. For the direct encoding condition, the input in these experiments corresponded to Gaussian white noise, with mean 1 and standard deviation 0.5, whose amplitude was scaled by ρ_u. For the indirect encoding condition, the input in these experiments consisted of N^{in} independent realizations of a stationary Poisson process with rate ρ_u. The target operating point was achieved for values of $\rho_u = 1.5\text{nA}, 14.0\,\text{spikes/s}$ and $\gamma = 15, 11$, for the direct and indirect encoding, respectively (see Figure 2**a**, **b**).

In line with the activity traces illustrated in Figure 1, driving the circuit with spiking input leads to a more 'natural' response, a sparser and more distributed activity pattern where the degree of synchrony and, particularly, regularity is significantly minimized (Figure 2**c** and **d**). Importantly, despite firing

at a lower mean rate, the indirect encoding condition engages a larger fraction of the population. In addition, the magnitude of the synaptic currents is smaller in the direct encoding condition (where synaptic currents are only generated internally, through the recurrent synapses). Even though it is straightforward to place the system in the desired operating point in both encoding schemes (Figure 2**a**, **b**), driving the system with spiking input leads to physiologically more realistic responses, both at the single neuron and population levels, as it provides the necessary synaptic drive to increase the circuit's overall responsiveness.

Encoding capacity

In all the cases that we explored, representational capacity was high and the circuit achieved perfect classification accuracy for $N_u \leq 1000$. However, if we evaluate the error between the raw readout output $\hat{\mathbf{y}}$ and the binary target $\mathbf{y}$, the difference between encoding and decoding schemes becomes more apparent (Figure 3). Error increased steadily with the number of input stimuli but appeared to saturate in the indirect encoding condition (Figure 3**b**). This suggests that the total capacity was smaller than in the direct encoding condition (Figure 3**a**). This is not surprising given the larger impact this encoding regime has on the characteristics of population activity.
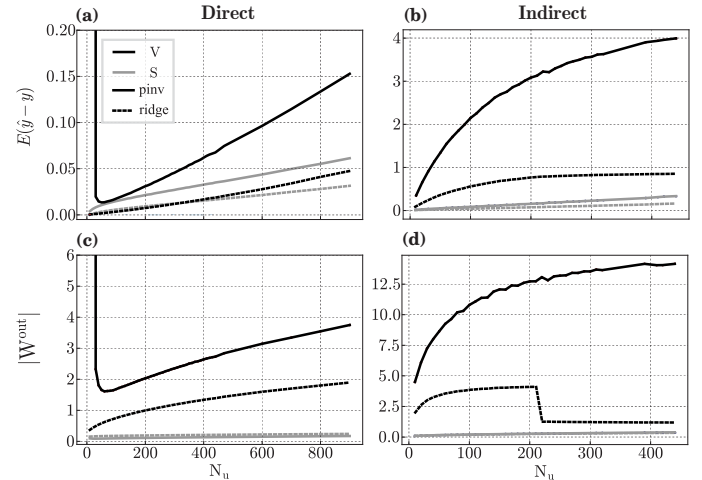

Fig. 3. Error and stability of readouts as a function of the number of stimuli for direct (left) and indirect (right) encoding. **(a, b)** Squared error; **(c, d)** Readout norm. In both conditions the classification accuracy was 1 for the range of stimuli considered. The plots depict the results obtained when decoding was performed either on the membrane potentials (V, black) or the filtered spike trains (S, gray); additionally, the impact of regularization (see subsection II-E) in the decoding process is shown by comparing the results obtained with $\beta = 0$ (solid curves, pseudo-inverse method) and $\beta > 0$ (dashed curves, Ridge regularization). Note that the scales differ across graphs.

For both encodings, as discussed in [13], reading out from membrane potentials is a more natural way to decode the population activity, leading to smaller classification errors (Figure 3**a**, **b**) and more stable solutions (Figure 3**c**, **d**). Regularization is also shown to benefit the decoding process in that it substantially reduces the error and the magnitude of

the norm of the estimated weights. This result highlights the importance of high-dimensionality and distributed population activity. Solutions that privilege a distributed readout which does not amplify a small number of dimensions (due to the imposed penalty on the norm, see subsection II-E) significantly reduce the error, particularly when reading out from the filtered spike trains.

It is worth noting that, in the direct encoding condition, only the neurons that are directly stimulated depolarize sufficiently to spike. This feature explains the large error obtained for small N_u (Figure 3**a**, **c**). When the number of stimuli is very small, so is the total number of stimulated neurons, which in turn leads to reduced spiking activity in the circuit and introduces numerical instability.

In summary, a direct encoding scheme leads to more discernible responses. It improves the circuit's representational capacity and enables accurate stimulus classification. When the input stimuli are encoded indirectly, population responses are 'healthier', but classifying stimulus identity is harder, leading to larger errors as well as relatively unstable solutions, especially when no regularization penalty is introduced.

Internal representations

The characteristics of the circuit's macroscopic state in response to its input are a key determinant of its computational performance. They determine how the system exploits its high-dimensional state space to project distinct input stimuli onto separable activity clusters. Consequently, they influence the circuit's representational and processing capacity. In this section, we explore how the choice of encoding scheme affects the system's effective dimensionality, i.e., the subspace on which neural activity predominantly lies (see [2]) and how these effects are reflected on different state variables.
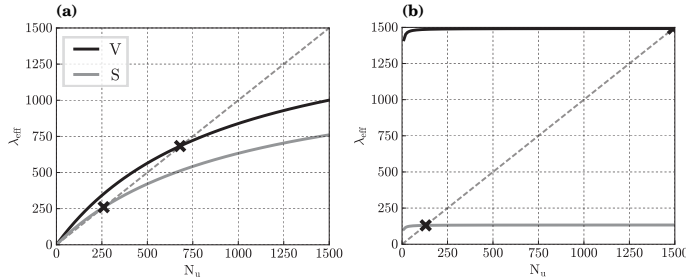

Fig. 4. Effective dimensionality of the state matrix X as a function of the number of input stimuli N_u for the direct **(a)** and indirect **(b)** encoding schemes and for conditions where the state matrix is composed of samples of either the population membrane potentials (V, black) or the filtered spike trains (S, gray). Dashed line denotes the diagonal.

Effective dimensionality (λ_{eff}) quantifies how explained variance is distributed among the principal components (PC). If all PCs capture equal amounts of variance in the data, λ_{eff} will be high. If some PCs explain a large amount of variance relative to other PCs, λ_{eff} will be low. The effective dimensionality in the direct encoding condition grows supralinearly with the number of stimuli up to a maximum of $N_u \approx 250$ and $N_u \approx 650$ stimuli, when the decoded variables are the filtered
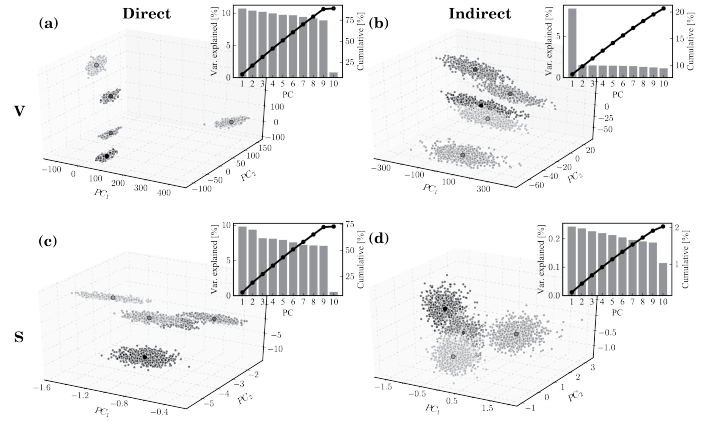

Fig. 5. Impact of encoding strategy and state variable on the characteristics of stimulus representations for direct (left column, **(a,c)**) and indirect (right column, **(b,d)**) encoding schemes. Depicted in each panel is a projection of stimulus-specific state vectors to the space spanned by the first three principal components when the state comprises samples of the population membrane potentials (V, **(a,b)**), or the filtered spike trains (S, **(c,d)**). The insets depict the amount of variance explained by the first 10 principal components.

spike trains S or the membrane potentials V, respectively (Figure 4**a**). Using the indirect encoding scheme, the effective dimensionality no longer depends on the number of unique input stimuli, given that the number of actual input channels is constant (N_{in}, see Figure 4**b**). Nevertheless, both conditions exhibit a marked decline when the analysis is performed on the filtered spiking activity, suggesting that this commonly used post-processing step may incur information loss as it underestimates the richness of the system's dynamics.

The complexity of the stimulus representations (and indirectly, the circuit's capacity) can also be evaluated when the stimulus-specific activity vectors are projected onto the space spanned by the first three principal components. The direct encoding scheme gives rise to a clearer state separation, where activity vectors in response to different stimuli lie in distinct and well-separated clusters (see Figure 5**a**), and this facilitates readout classification. Indirect encoding also leads to unique stimulus-specific representations, but with larger variance, as illustrated by the spread of the activity vectors in Figure 5**b**. This explains the lower capacity (and higher classification error) observed in this encoding regime and the smaller sensitivity of effective dimensionality to the number of input stimuli.

It is also worth noting that in both encoding schemes, reading out from the filtered spike trains has detrimental effects as it 'distorts' the internal representations and reduces their uniqueness (Figure 5**c**, **d**). This increases the error in that it makes it harder for the readout to extract the relevant information even though it is present in the system.

How important is the balanced state?

An important assumption underlying the chosen baseline conditions (see section III) is that the macroscopic state of the circuit determines its computational performance. Even though physiological and computational evidence points to

the relevance of a balanced, asynchronous irregular state (AI state) for active processing, an explicit quantification of how functionally relevant these properties are is still lacking (but see, e.g. [5]).

Driving the circuit with unspecific and stochastic input (used to tune the population state, see subsection II-D) differs from the stimulus-driven condition, where specific random subsets of neurons were stimulated. Even though the AI state is retained in both cases, due to the maintenance of the chosen stimulus amplitudes, stimulus onset causes a noticeable modulation of the circuit's macroscopic statistics (Figure 6). For both encoding schemes, stimulus responses are characterized by a marked desynchronization, whereby the overall synchrony (I^{sync}) is substantially reduced. This effect is more noticeable in the indirect encoding condition. The responses also become more regular with a slight deviation away from the Poisson statistics.
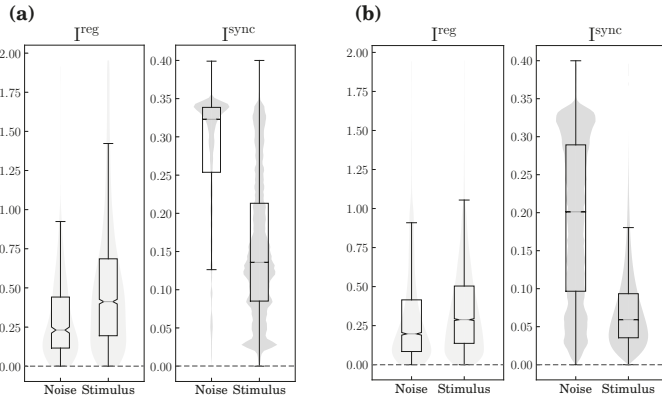

Fig. 6. Stimulus-evoked modulation of population state in comparison with the input noise condition for direct **(a)** and indirect **(b)** encoding schemes. For each condition the figure displays the distributions of regularity (I^{reg}) and synchrony (I^{sync}) indices in noisy vs evoked conditions. Note that these indices reflect the difference to a pure Poisson process, so smaller values indicate activity closer to the AI state.

Apart from these differences observed in the stimulus-vs noise-driven scenarios, a more relevant question is how these values affect computational capacity, i.e., how important is the balanced state? We thus proceeded to measure the classification error in the different encoding schemes, while varying the circuit state by systematically manipulating ρ_u and γ. For each parameter combination, we assess the degree of population-wide synchrony (I^{sync}) and regularity (I^{reg}) as well as the overall balance of excitation and inhibition (I^{EI}) relative to the circuit's classification performance in the direct and indirect encoding schemes (Figure 7).

The results show a non-trivial relation between the observed population state and the average classification error. Overall, the single-most important feature appears to be EI balance in that, regardless of the encoding condition, the error is minimized when the average difference between E and I currents is closest to 0, i.e., when the circuits are closest to a globally balanced regime.
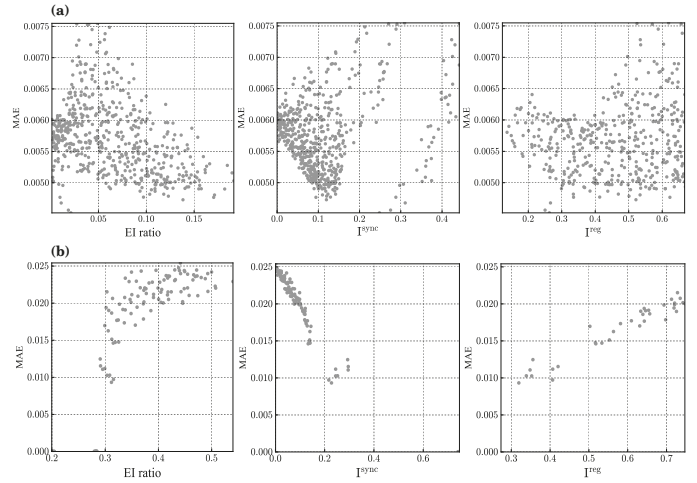

Fig. 7. How the population state and global EI balance affect the circuit's performance in the stimulus classification task with $N_u = 50$ under the direct **(a)** and indirect **(b)** encoding conditions. Mean absolute error (MAE) of the readout output ($\hat{\mathbf{y}}$) is displayed as a function of: difference of mean inhibitory and excitatory current amplitudes (for 1000 randomly sampled excitatory neurons, see subsection II-D), recorded during the test phase (left); and the summary indices for synchrony I^{sync} (middle) and regularity I^{reg} (right). Note that, due to computational constraints, **(b)** contains fewer data points than **(a)**.

Population-wide synchrony appears to affect performance in a more complex manner, whereby the minimum error is obtained at an intermediate level of synchronization, with $I^{\text{sync}} \approx 0.25$. In any case, both conditions demonstrate that complete asynchrony ($I^{\text{sync}} = 0$) does not lead to lower error. The functional impact of irregularity likewise appears to be dependent on the encoding scheme, and minimal error is also achieved at an intermediate value of $I^{\text{reg}} \approx 0.3$. More regular responses tend to lead to higher error, but a smaller I^{reg} does not correspond to the lowest error.

DISCUSSION

Using the reservoir computing (RC) approach as a reference conceptual frame, we can analyse the dynamics of information processing in biologically inspired architectures, where multiple features of the system's structural and functional components can be directly constrained by empirical data from multiple scientific domains. Using cognitively plausible computational tasks implemented in neurobiologically plausible systems, we can test relevant functional hypotheses.

However, as demonstrated throughout this study, the quality of the input-state mappings required to solve a particular task and the specificity with which the system is driven by the external input is not only determined by the specifications of the circuit considered, but also by how the input is delivered to it, how it encodes the relevant information and how population activity is decoded from it. These choices introduce non-negligible biases, which may limit the generalizability of results. Ideally, we ought to converge onto solutions that minimize or completely remove such biases, by applying strict neurobiological constraints.

Choosing an adequate encoding strategy depends on the nature of the computational problem one seeks to address but has been shown to drive the circuits in a very specific manner (Figures 2, 6), leading to specific macroscopic activity patterns and thus strongly imposing particular interpretations of the underlying processes that may not match relevant physiological observations. For example, it is common to drive the circuit with direct somatic current representing the input variables to be encoded (what we have termed direct encoding). However, as mentioned in section I, biological microcircuits communicate via complex spatio-temporal spike patterns and thus, any external input to be encoded and processed must be contained in complex spatio-temporal spike sequences. Failure to do so, leads to unrealistic population dynamics that is inconsistent with physiological observations.

Our results demonstrate that spiking input naturally leads to physiologically more realistic responses (Figure 2), where the majority of neurons operates close to their firing threshold and population activity is sparser and more evenly distributed, consistent with cortical activity during active processing states. The indirect encoding scheme thus provides the necessary synaptic drive that is missing in the direct encoding condition. In that case, smaller sets of neurons are sufficiently depolarized to ever reach the threshold and the only synaptic drive comes from internal, recurrent connections.

Unsurprisingly, and despite leading to population responses that are not as physiologically realistic, the direct encoding scheme is computationally beneficial (Figure 3), leading to more accurate classification and higher representational capacity. Different stimuli map onto compact and clearly separable clusters (Figure 5) and the dimensionality of the sub-space where population activity generally lies is more constrained and stimulus-dependent (Figure 4). The system's responses are enslaved by the input, facilitating the readout classification. On the other hand, the larger variability and higher-dimensional responses observed in the indirect encoding condition are likely to play an important functional role for more complex computations, but restrict the circuit's representational capacity.

An equally important issue that needs to be addressed when attempting to probe the characteristics of functional neurodynamics and the impact of different circuit features on computational performance, lies in the specification of the relevant state variables and how information is readout (the decoding problem). Ideally, the system's dynamics should be assessed with a direct, unbiased internal variable, but this is typically not the case in the context of SNNs. Instead, the most common approach is to extract information from the circuits using a continuous rate variable, obtained by convolving the population's spike trains with an exponential kernel. As discussed in [13] and further validated in this study, this procedure may incur information loss (Figure 3), by under-representing the richness of the system's dynamics (Figure 4).

Overall, the results presented in this study provide some preliminary constraints for studies applying RC principles to biologically inspired spiking neural network models of cognitive function and highlight the importance of constraining the properties of these systems and the experimental design by relevant biophysical considerations. Failure to account for many of these constraints may introduce significant biases that limit the generalizability of any conclusions drawn, in that the systems' behavior and computational performance may come to reflect these biases and not necessarily the circuit's properties.

References

[1] Larry F. Abbott, Brian DePasquale, and Raoul-Martin Memmesheimer. Building functional networks of spiking model neurons. *Nature Neuroscience*, 19:1–16, 2016.

[2] Larry F. Abbott, Kanaka Rajan, and Haim Sompolinsky. Interactions between intrinsic and stimulus-evoked activity in recurrent neural networks. In *The Dynamic Brain: An Exploration of Neuronal Variability and Its Functional Significance*, pages 1–16. 2011.

[3] Romain Brette and Wulfram Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, (94):3637–3642, 2005.

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

[5] Renato Duarte and Abigail Morrison. Dynamic stability of sequential stimulus representations in adapting neuronal networks. *Frontiers in Computational Neuroscience*, 8:124, 2014.

[6] Renato Duarte, Barna Zajzon, and Abigail Morrison. Neural Microcircuit Simulation And Analysis Toolkit. *Zenodo*, jan 2017.

[7] Chris Eliasmith and Charles H. Anderson. *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. MIT Press, 2003.

[8] 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.

[9] Susanne Kunkel et al. Nest 2.12.0, March 2017.

[10] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[11] Raoul-Martin Memmesheimer and Marc Timme. Designing the dynamics of spiking neural networks. *Physical Review Letters*, 97(18):1881011–4, 2006.

[12] Mario Mulansky, Nebojsa Bozanic, Andreea Sburlea, and Thomas Kreuz. A guide to time-resolved and parameter-free measures of spike train synchrony. *Proceedings of the 1st International Conference on Event-Based Control, Communication and Signal Processing, EBCCSP 2015*, (1):1–8, 2015.

[13] Dick van den Broek, Marvin Uhlmann, Hartmut Fitz, Renato Duarte, Peter Hagoort, and Karl Magnus Petersson. The best spike filter kernel is a neuron, 2017. Extended abstract, Cognitive Computational Neuroscience conference (CCN 2017), New York.