bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Analysis of visual processing capabilities and neural coding strategies of a detailed model for laminar cortical microcircuits in mouse V1

Guozhang Chen, Franz Scherr, and Wolfgang Maass*

Institute of Theoretical Computer Science, Graz University of Technology, Graz, Austria

December 7, 2021

Abstract

The neocortex is a network of rather stereotypical cortical microcircuits that share an exquisite genetically encoded architecture: Neurons of a fairly large number of different types are distributed over several layers (laminae), with specific probabilities of synaptic connections that depend on the neuron types involved and their spatial locations. Most available knowledge about this structure has been compiled into a detailed model [Billeh et al., 2020] for a generic cortical microcircuit in the primary visual cortex, consisting of 51,978 neurons of 111 different types. We add a noise model to the network that is based on experimental data, and analyze the results of network computations that can be extracted by projection neurons on layer 5. We show that the resulting model acquires through alignment of its synaptic weights via gradient descent training the capability to carry out a number of demanding visual processing tasks. Furthermore, this weight-alignment induces specific neural coding features in the microcircuit model that match those found in the living brain: High dimensional neural codes with an arguably close to optimal power-law decay of explained variance of PCA components, specific relations between signal- and noise-coding dimensions, and network dynamics in a critical regime. Hence these important features of neural coding and dynamics of cortical microcircuits in the brain are likely to emerge from aspects of their genetically encoded architecture that are captured by this data-based model in combination with learning processes. In addition, the model throws new light on the relation between visual processing capabilities and details of neural coding.

1 Introduction

The mammalian neocortex is a two-dimensional sheet of rather stereotypical local cortical microcircuits with diverse types of neurons distributed over 6 laminae with largely genetically encoded interconnection profiles. It is commonly conjectured that understanding the organization of information representation and computation in such a generic cortical microcircuit holds the key for understanding brain function. Intense research during the past decades has provided a rich set of data on the diverse types of neurons and their specific laminar locations and interconnectivity [Mountcastle, 1998, Thomson and Lamy, 2007, Markram et al., 2015], culminating in the model [Billeh et al., 2020] for a patch of area V1 in mice, see Fig. 1. We will use this model, more precisely its point neuron version that consists of 51,978 neurons of

*Corresponding author: maass@igi.tugraz.at

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

111 different types on 6 laminae, as the basis for our work. We will refer to this model simply as the Billeh model. More precisely, we complement this model with a noise model that is based on experimental data from [Stringer et al., 2019a]. Furthermore, we propose and analyze biologically plausible rules for reading out results of network computations from the Billeh model.

The Billeh model with the parameters provided in [Billeh et al., 2020] is not able to solve nontrivial computing tasks since its synaptic weights have not been properly aligned. They were chosen largely on the basis of experimental data about the mean and variance of synaptic weights. But it is well-known that such data do not provide salient insight into computational capabilities of a network, since they do not reflect correlations and other higher-order relations between the weights of different synaptic connections. These higher-order dependencies arise in the living brain largely through synaptic plasticity, and are essential for endowing the network with computing capabilities. The relevance of these higher-order relations for understanding computational capabilities of a network can be illustrated through an analogy: Knowledge of the mean and variance of synaptic weights in a trained artificial neural network provides hardly any information about its computing capabilities.

However, the data of [Billeh et al., 2020] allow us to investigate the question of what computing tasks this cortical microcircuit model can solve if its weights are properly aligned for those, in other words, if it has been trained for those tasks. We aligned its synaptic weights through a variation of backpropagation-through-time (BPTT) from [Scherr and Maass, 2021] that works quite well for the Billeh model. In particular, it can be applied to networks consisting of generalized leaky integrate-and-fire (GLIF) neuron models, out of which the point neuron version of the Billeh model is composed. These neuron models, which are more precisely GLIF_3 neurons with two additional hidden variables besides the membrane potential, have been fitted to individual neurons of the cell data-based in the Allen Brain Atlas, see [Teeter et al., 2018]. Instead of training the model for just one computing task as in [Scherr and Maass, 2021], we train it for 5 different visual processing tasks that have commonly been considered in biological experiments. We find that the Billeh model can simultaneously achieve high accuracy on all 5 tasks. In addition, its task performance turns out to be highly resilient also to new forms of noise that were not present during training.

We next address the dimensionality of input presentations in the Billeh model, both before and after training. Lots of data had pointed to low-dimension representations of visual inputs due to stereotypical inherent dynamics of neural ensembles [Pérez-Ortega et al., 2021]. Low-dimension representations also offer the promise of higher noise robustness. However, [Stringer et al., 2019a] found that representations of visual inputs in area V1 of mouse are very high dimensional, and the explained variance of PCA components follows a power law that is optimal from a theoretical perspective because it combines large representational capacity with good generalization capabilities. We found that the Billeh model in conjunction with our data-based default noise model reproduces these neural coding properties with remarkable precision. However, we also found that this power law is not necessary for the visual processing capabilities that we examined, since a high level of noise can destroy these coding properties without affecting task performance.

We examined in the Billeh model also to what extent it reproduces other experimental data for mouse V1 that address limitations of neural coding due to noise correlations [Rumyantsev et al., 2020]. This study found that V1 reduces the impact of correlated noise on neural coding through extensive use of coding dimensions that are orthogonal to the strongest noise component. However, noise was nevertheless found to limit the separability of neural codes for different visual stimuli because lower noise components are entangled with signaling dimensions. We show that both of these experimental findings are reproduced by the Billeh model in combination with the added noise model. We can also confirm in this model a conjecture of [Stringer et al., 2021], that some results on information-limiting correlations in visual cortex are affected by extrapolating results of recordings from smaller sets of neurons to larger neural systems. One clearly sees in the Billeh model that the separability of neural codes for images keeps increasing

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

when one increases the number of neurons beyond the number considered in [Rumyantsev et al., 2020]. Our model offers, in addition, an answer to the puzzle that was addressed in [Stringer et al., 2021]: That the behavioral discrimination threshold for orientations of grating is much higher than suggested by the measured joint information in large ensembles of neurons in V1. We found that an idealized linear readout from all neurons of the Billeh model, as commonly used in previous models, does support a much lower discrimination threshold. However, if one takes into account how much of this information can be extracted and communicated by projection neurons of the circuit, one sees that not all information that is represented by the neurons can be used by downstream networks.

Finally, we addressed the question in which dynamic regime the trained and untrained Billeh models operate, in particular, to what extent they operate in a critical regime that is favored by some theoretical studies [Chen and Gong, 2019, Wilting and Priesemann, 2019]. We used the branching factor of neural activity, as suggested for example by [Wilting and Priesemann, 2018a], to evaluate the criticality of these models. We found that their branching factors are in the same range as those of neural networks in the living brain.

The research method that we used in this work, training of large-scale models, capitalizes on significant improvements in software and computing hardware that resulted from efforts to accelerate the training of deep neural networks in machine learning and AI. We show that some of these improvements can be recruited to also speed up simulation and training of large-scale data-based models for neural networks of the brain.

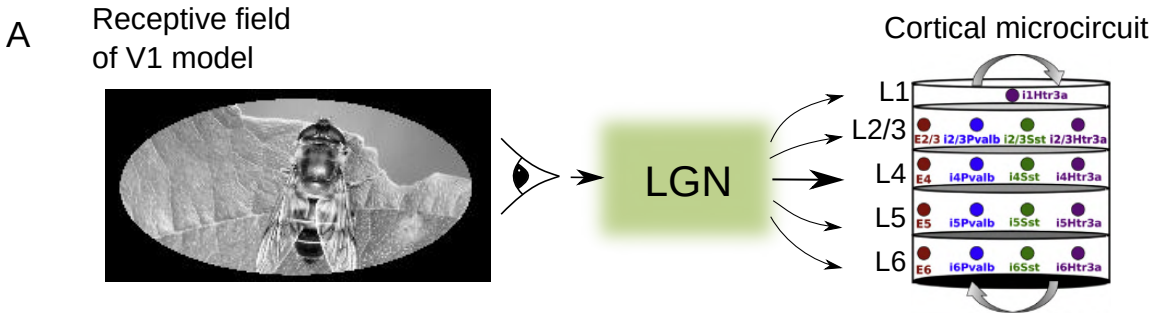
2 Results

2.1 Preparing the ground for investigating computational properties of a data-based laminar cortical microcircuit model.

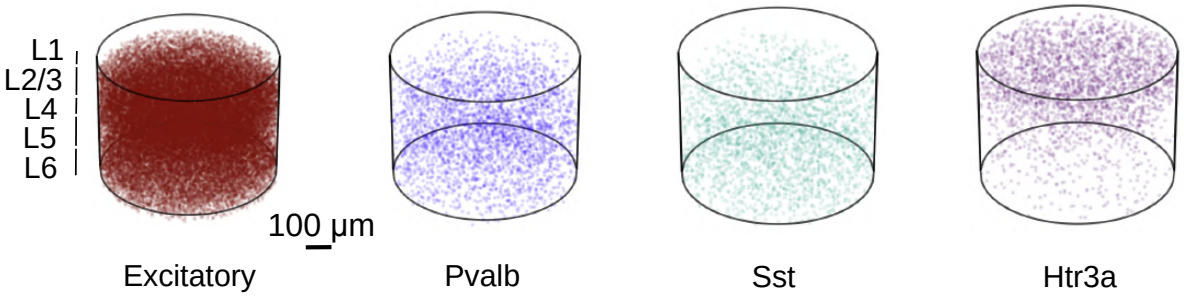
We build on the new compilation and drastic expansion of anatomical and neurophysiological data on the cortical microcircuit that is reported in [Billeh et al., 2020]. We will refer in the following to the point-neuron version of their model, and to our small modification of it regarding noise, as the Billeh model. This point-neuron version of the model in [Billeh et al., 2020] is a large-scale model of a 400 µm-radius patch, also commonly referred to as cortical microcircuit or column, of area V1 in mouse. It consists of 51,978 neurons and a simplified model of LGN (Fig. 1A). This cortical microcircuit model distinguishes 17 neuron classes, resulting from 4 main types of neurons, one excitatory type and 3 inhibitory types, located on 5 different layers (laminae), denoted L1, L2/3, L4, L5, L6, see Fig. 1B. Excitatory neurons (pyramidal cells) are distributed over L2/3, L4, L5, L6 (labeled E2/3, E4, E5, and E6 in Fig. 1A). The 3 inhibitory neuron types are Htr3a neurons (subsuming VIP neurons), that are present in all layers, as well as parvalbumin (Pvalb) and somatostatin (Sst) neurons in L2/3 to L6. The fraction of each type of neuron on each layer is based on experimental data. The resulting 17 cell classes are further split into 111 neuron types according to their dynamics (firing response). They are modeled as generalized leaky integrate-and-fire 3 ($GLIF_3$) neurons that have besides the membrane potential two types of spike-triggered internal currents as dynamic variables. The hard reset of membrane potential in [Teeter et al., 2018] was replaced by a voltage reduction (last term in the first line of Eq. 1). Their diverse dynamics are specified by synaptic weights and the parameters which are colored red in the following equations

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A



B

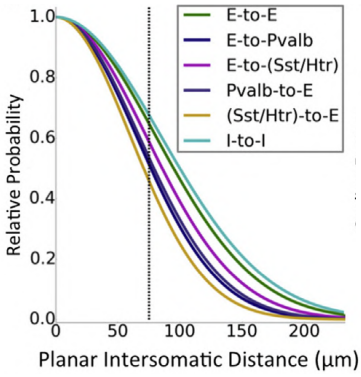


C

Connection probabilities at 75µm

Trg / Src	i1Htr3a	E2/3	i2/3 Pvalb	i2/3 Sst	i2/3 Htr3a	E4	i4 Pvalb	i4 Sst	i4 Htr3a	E5	i5 Pvalb	i5 Ssst	i5 Htr3a	E6	i6 Pvalb	i6 Sst	i6 Htr3a
i1Htr3a	0.656	0.356	0.093	0.068	0.464	0.148				0.148	0	0		0.148			
E2/3	0	0.16	0.395	0.182	0.105	0.016	0.083	0.083	0.083	0.083	0.081	0.102		0			
i2/3Pva	0.024	0.411	0.451	0.03	0.22	0.05	0.05	0.05	0.05	0.07	0.073						
i2/3Sst	0.279	0.424	0.857	0.082	0.77	0.05	0.05	0.05	0.05	0.021		0					
i2/3Htr	0	0.087	0.02	0.625	0.028	0.05	0.05	0.05	0.05	0							
E4		0.14	0.1	0.1	0.1	0.243	0.43	0.571	0.571	0.104	0.101	0.128	0.05	0.032			
i4Pvalb		0.25	0.05	0.05	0.05	0.437	0.451	0.03	0.22	0.088	0.091	0.03	0.03				
i4Sst	0.241	0.25	0.05	0.05	0.05	0.351	0.857	0.082	0.77	0.026	0.03	0	0.03				
i4Htr3a		0.25	0.05	0.05	0.05	0.351	0.02	0.625	0.028	0	0.03	0.03	0.03				
E5	0.017	0.021	0.05	0.05	0.05	0.007	0.05	0.05	0.05	0.116	0.083	0.063	0.105	0.047	0.03	0.03	0.03
i5Pvalb	0	0	0.102			0	0.034	0.03	0.03	0.455	0.361	0.03	0.22	0.03	0.01	0.01	0.01
i5Sst	0.203	0.169		0.017		0.056	0.03	0.006	0.03	0.317	0.857	0.04	0.77	0.03	0.01	0.01	0.01
i5Htr3a						0.03	0.03	0.03	0.03	0.125	0.02	0.625	0.02	0.03	0.01	0.01	0.01
E6		0				0				0.012	0.01	0.01	0.01	0.026	0.145	0.1	0.1
i6Pvalb		0.1				0.1				0.1	0.03	0.03	0.03	0.1	0.08	0.1	0.08
i6Sst										0.03	0.03	0.03	0.03	0.1	0.05	0.05	0.05
i6Htr3a										0.03	0.03	0.03	0.03	0.1	0.05	0.05	0.03

D



(caption next page)

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 1 ***(previous page)*****: Cortical microcircuit model of [Billeh et al., 2020]. (A)** The model consists of 4 types of neurons on 5 different layers, together with a model for LGN that transforms visual inputs into input currents to neurons in the microcircuit. The cortical microcircuit model receives visual input from an oval in the image space; note that we consider only the detailed"core" part of the model in [Billeh et al., 2020]. **(B)** The model contains one excitatory and three inhibitory types of neurons. Each color point labels the position of one neuron. **(C)** Base connection probabilities from [Billeh et al., 2020] depend on the cell class to which the presynaptic (row labels) and postsynaptic neuron (column labels) belong. **(D)** The base connection probability from **(C)** is multiplied according to [Billeh et al., 2020]) by an exponentially decaying factor that depends on the horizontal distance between the neurons involved, and their type. Several panels of this figure are reproduced from [Billeh et al., 2020].

(see Methods for details):

$$
\begin{aligned}
v_j(t+1) &= \alpha v_j(t) + \frac{1-\alpha\tau}{C}\left[I_j^e(t+1) + \sum_m I_j^m(t+1) + gE_{\mathrm{L}} + I_j^{\mathrm{syn}}(t)\right] - z_j(t)v_{\mathrm{th}} \\
z_j(t) &= H\left(v_j(t) - v_{\mathrm{th}}\right) \\
\alpha &= e^{-\delta t/\tau} \\
I_j^e(t) &= \sum_i W_{ji}^{\mathrm{in}} x_i(t) + qK_j^{\mathrm{quick}}(t) + sK_j^{\mathrm{slow}} \\
I^m(t+1) &= f^m I^m(t) + z(t)\delta I^m;\ m = 1, 2.
\end{aligned}
\tag{1}
$$

These parameters can be interpreted as neuron capacity, resistance, resting potential, as well as amplitudes and decay time constants of two types of after-spike currents; the length of refractory period and synaptic time constant were also determined by the neuron type. These neuron parameters have been fitted to experimental data from 111 selected neurons according to the cell database of the Allen Brain Atlas [Allen Institute, 2018], see [Teeter et al., 2018, Billeh et al., 2020]. In addition, the Billeh model specifies the connection probability between neurons, based on experimental data. The base connection probability for any pair of neurons from the 17 cell classes is provided in [Billeh et al., 2020] by a table (reproduced in Fig. 1C); white grid cells denote unknown values. The entries in this table are based on measured frequencies of synaptic connections for neurons at maximal 75 µm horizontal intersomatic distance. This base connection probability was scaled by an exponentially decaying factor in terms of the horizontal distance of the somata of the two neurons (Fig. 1D). This distance-dependent scaling is also based on statistical data from experiments (leaving aside finer details of connection probabilities). This microcircuit model receives synaptic inputs from a simple LGN model according to [Billeh et al., 2020] (Fig. 1A), which we model as input currents. This LGN model transforms visual inputs, i.e., images and movies, into retinotopic and lamina-specific inputs to neurons in the microcircuit model.

The neurons j in the cortical microcircuit model of [Billeh et al., 2020] received, besides external inputs, inputs from other neurons also a noise current K_j. This noise current consisted of a single Poisson source that was injected into all neurons and resulted in a global noise correlation that was not based on experimental data. We replace it with a noise model that is based on experimental data from [Stringer et al., 2019a], more precisely on measured responses of large numbers of neurons in mouse V1 to repeated presentations of 2,800 natural images (Methods). There an additive form of noise was extracted from experimental data by subtracting from neural signals their mean over two presentations of the same natural images (see Methods for details). This resulted in the empirical noise distribution shown in Fig. 2A. One sees that this data-based noise distribution is heavy-tailed. We define our noise model by drawing the noise current K_j from this distribution, independently for each neuron j and each time step t, i.e., every ms.

We will refer to the previously described form of noise as quick noise, and denote its scaling factor by q.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

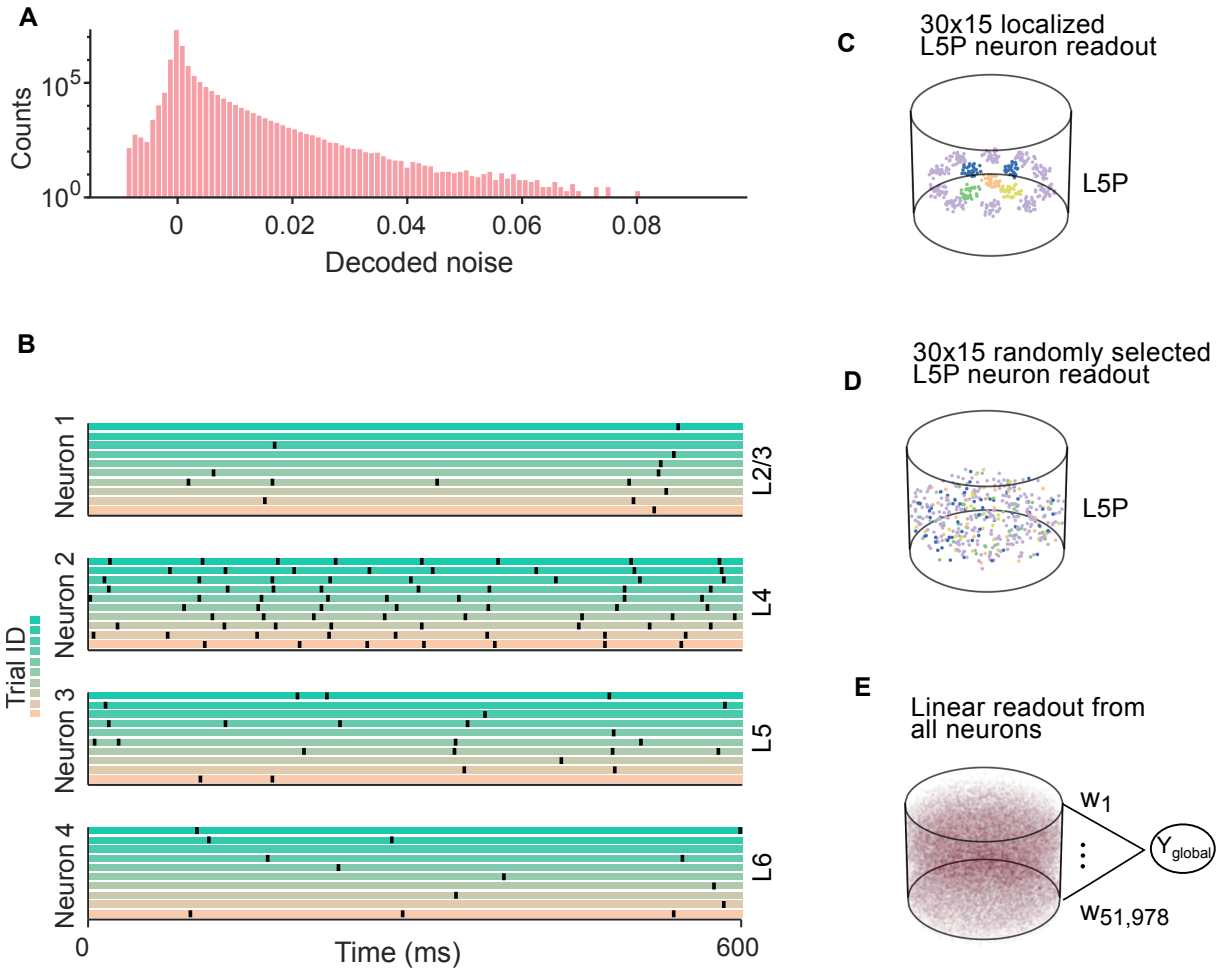


Figure 2: Adding biologically realistic noise and network readouts to the microcircuit model. (A) Histogram of noise values for individual neurons based on experimental data from [Stringer et al., 2019a]. We draw in our noise model the noise values from the corresponding empirical distribution. **(B)** Resulting trial-to-trial variability of 4 randomly selected neurons from L2/3-L6, respectively (those have spiking activities) in the microcircuit model for 10 trials with the same input image. The resulting variability is substantial, posing a nontrivial challenge for computational processing in the microcircuit so that it can produce at a desired time firing activity in its readout populations that are not affected by this response variability. **(C)** 15 spatially separated groups of 30 pyramidal neurons in L5 were selected to signal specific network outputs for 5 different tasks (see Fig. 3). **(D)** Alternative selection of these 15 populations in L5 without spatial clustering leads to very similar performance. Color scheme is the same in Fig. 3A, representing different output values/decisions for 5 chosen tasks. **(E)** Schematic diagram of a commonly considered readout convention that has less biological support: A linear readout receives synaptic input from all 51,978 neurons in the microcircuit model, using a corresponding number of weights that can all be optimized for one particular task.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

It appears to be suitable for modeling qualitatively the impact of stochastic release of vesicles at synaptic connections from pre-synaptic neurons. But neural networks of the brain are in addition subject to noise on a slower time-scale, commonly referred to as trial-to-trial variability. Whereas quick noise mainly shifts spike times but does not affect firing rates in a major way, trial-to-trial variability affects the firing rates of neurons. In other words, the firing rates of neurons tend to be different from trial to trial, often in a way that can be related to the network state at the beginning of a trial [Arieli et al., 1996]. We model this form of slow noise by adding to each neuron another input current whose amplitude is drawn from the same distribution as the quick noise, but is only drawn once at the beginning of a trial. We call the scaling factor of this slow noise s, so that each noise model that we consider is characterized by the two scaling factors q and s. In our default setting, both have the value 2. To illustrate the effect of this noise model in our modification of the Billeh model, we input the same natural image to LGN in 10 trials. Figure 2B shows the response of 4 randomly selected neurons in the microcircuit model in these 10 trials. One sees a substantial trial-to-trial variability that qualitatively matches generic experimental data (see e.g. Fig. 1B of [Nikolić et al., 2009]). The Fano factor of spike counts in 10 ms was 1.46, quantitatively comparable with experimental data from mouse V1 (1.39) [Montijn et al., 2014].

In order to compare the impact of our noise model on network responses to natural images with that found in the experiments of [Stringer et al., 2019a] we analyzed the correlations between the responses of large numbers of neurons to two presentations of the same image, projected onto selected principal components, as in Extended Data Fig. 5 of [Stringer et al., 2019a]. We found that our noise model places these correlations into a similar range, although resulting correlations are somewhat higher according to Fig. 4B. However, the measured correlations in [Stringer et al., 2019a] were not only reduced by noise in V1 neurons, but also by measurement noise. According to [Stringer et al., 2019a] several factors contribute to measurement noise in Calcium imaging from very large numbers of neurons, but the exact contribution of that appears to be unknown.

In order to investigate the computational capabilities of a cortical microcircuit model, one needs to make an assumption regarding which neurons project computational results of the microcircuit to downstream areas. One common convention for reading out binary computational results from microcircuit models (see e.g. [Haeusler et al., 2009]) is the postulate that a linear downstream "decision neuron" receives synaptic input from all neurons in the microcircuit, that a correspondingly large number (in our case 51,978) of weights to this global readout neuron can be trained for a specific computing task, and that this global linear readout neuron signals the binary result of a computation by assuming a value above/below a given threshold, see Fig. 2E for a schema. Such readout convention is problematic from a biological perspective. One issue is that experimental data suggest that there does not exist such global readout neurons from all neurons in a cortical microcircuit. Furthermore, theoretical results for Support Vector Machines suggest adding such global readout neuron with task-dependent readout weights gives to a quite powerful computational model even if the computational contribution of the microcircuit itself is rather benign, and the readout neuron is modeled as a simple linear neuron [Haeusler et al., 2009]. Hence, to understand the potential computational contribution of a microcircuit model it appears to be more adequate to assume that like in the brain, specific neurons in the microcircuit project their computational results to downstream networks, and that their projection content results from the dynamics in the microcircuit model and the weights of synaptic connections from other neurons in the microcircuit model (but no additional "readout weights" to a global external readout). It is well-known that only specific neuron classes project to other neocortical or subcortical areas [Harris and Shepherd, 2015]. In particular, many pyramidal cells on L5 project to subcortical areas, and can therefore provide the basis for a motor response that is triggered there. Therefore we selected for each computational task and each possible output value of this computation a population of 30 excitatory neurons on L5 that triggered a corresponding downstream response if their firing activity was during a specific time window larger than the activity of the other selected populations of 30 neurons on L5. 15 such populations were needed for the 5 computational tasks that we considered. In our default model, each such population is a co-localized set

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of pyramidal neurons in L5P that were randomly sampled within disjoint spheres of radius 55 µm, see Fig. 2C and Fig. 3A. But we considered also the option that each of these 15 populations of excitatory neurons was randomly distributed on L5 see Fig. 2D. It leads to very similar results: an average accuracy of 91.56% for all 5 tasks (instead of 89.10% for co-localized pools of readout neurons on L5). we also compared these results with the performance of a global linear readout (see Fig. 2E): it achieved a higher accuracy: 97.73%.

2.2 A data-based laminar circuit can be trained to solve diverse visual processing tasks

We trained the Billeh model to carry out simultaneously five visual processing tasks that have also been used as bases for recordings from trained mice. These are the fine orientation discrimination task (Fig. 3B) of [Rumyantsev et al., 2020, Stringer et al., 2021], a generic image classification task (Fig. 3C), visual change detection task for natural images and static gratings (Fig. 3D) [Garrett et al., 2020, Siegle et al., 2021], and evidence accumulation from a sequence of signals on the left and right (Fig. 3E) [Morcos and Harvey, 2016, Engelhard et al., 2019] (Methods). We designed details of these tasks to be as close as possible to these biological experiments, but otherwise we made them as simple as possible. For every task, the cortical microcircuit model received visual inputs from the LGN model of [Billeh et al., 2020] (Fig. 3A). Pairs of visual inputs and target outputs are collected in separate batches for each task and these batches are interlaced during training (Methods).

Before training, with the weights given by [Billeh et al., 2020], the network model is incapable of performing any task; the accuracy is close to the chance level. The model was trained through BPTT with the help of the pseudo derivative for handling the discontinuous dynamics of spiking neurons suggested by[Bellec et al., 2018] (Methods). We made sure that during training synaptic weights cannot change their sign, i.e., synaptic plasticity followed Dale's law (Methods). BPTT was allowed to change all synaptic weights of the microcircuit model (W^{in}, and W^{rec}) in order to minimize a loss function (Methods). The chosen loss function aimed at minimizing the cross-entropy between the network output and a desired (target) output and simultaneously keeping firing rates and membrane potentials of all neurons in a biologically plausible regime range. After training, the Billeh model achieved on all 5 tasks a performance that was in the same range as reported behavioral data (Table 1), with an average accuracy of 89.10%. Simultaneously neural activity in the model stayed in a biologically realistic sparse firing regime of about 4 Hz, see also the spike rasters in Figs. S1-S5. Also, the weight statistics stayed after training in a range that is qualitatively consistent with biological data. The mean strength of inhibitory synapses increased from 0.04 to 0.24; the mean weight of excitatory synapses increased from 0.14 to 0.17, see Fig. S6 for resulting changes in the distributions of all excitatory and all inhibitory weights, and (Fig. S7 for more detailed analysis in terms of pre- and postsynaptic neuron type. Overall we arrived at the conclusion that the Billeh model achieved after training a high computational performance that is in the same range as corresponding behavioral data for mice, while staying with regard to the firing activity and values of synaptic weights within a biologically realistic regime.

2.3 The trained model creates similar quasi-optimal high-dimensional neural representations as observed in the brain

One of the most controversial questions about neural coding is whether the brain employs low- or high-dimensional neural representations of sensory inputs. It was recently shown through large-scale recordings [Stringer et al., 2019a] that area V1 of the mouse employs high-dimensional neural codes. More precisely, the explained variance of the nth principal component of network representations of images follows a

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

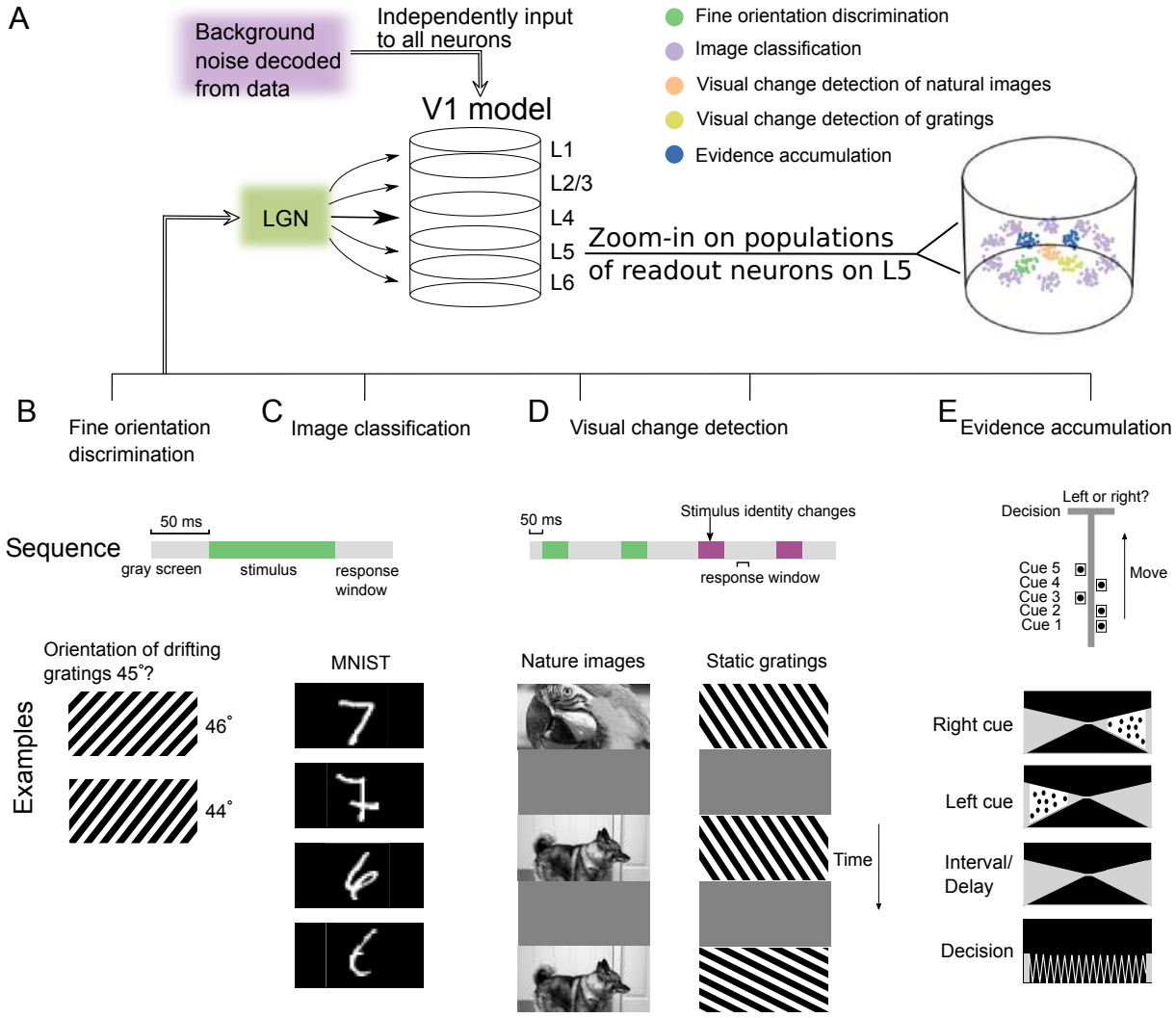


Figure 3: Illustration of the 5 visual processing tasks for which the model was trained. (A) The same visual stimuli as used in the corresponding biological experiments were input to the LGN. Separate pools of readout neurons on L5 were chosen for each task, one for each possible outcome of the network decision. The resulting 15 populations are shown on the right. **(B)** In the fine orientation discrimination task, the network received a drifting grating for 100 ms and a readout population had to report by producing at least one spike during the subsequent response window of 50 ms whether the orientation was smaller or larger than 45°. **(C)** In the image classification task, the network received a sample of a handwritten digit from 0 to 9 from the MNIST dataset (timing of input images and response windows as in the preceding task). The task was to decide which digit was denoted by the handwritten sample (a nontrivial decision in many real-world samples; two samples for 7 and 6 are shown). That one of 10 populations of readout neurons dedicated to this task that fired most strongly during the subsequent response window signaled the network decision. **(D)** For the visual change detection task a long sequence of images was presented, each for (100 ms) with (200 ms) gaps in between (while the input image represented a gray screen). The task was to report by at least one spike from the readout population within the response window if that image was different from the preceding one. Both natural images and static gratings were used. **(E)** For the evidence accumulation task, the mouse was running along a corridor in a virtual environment, where 5 transient visual cues were presented each for 50 ms at the left or right side of the corridor, with a gap of 10 ms between cues. The task was to turn at the end of the corridor to that side from which the majority of the cues had been received, independently of their order (250 ms after the offset of the last cue). The decision had to be indicated by activating one of two readout pools more than the other during the response window of 50 ms length.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 1: The Billeh model achieves high accuracy in all 5 tasks, consistent with the behavior performance of mice in similar tasks, after 6 training epochs.

	Test accuracy	Behavior accuracy	Mean firing rate (Hz)	Spike raster
Fine orientation discrimination	93.15%	$\sim 83\%$[a]	3.97	Fig. S1
Image classification	88.92%	N/A	4.11	Fig. S2
Visual change detection of natural images	84.13%/83.25%[b]	$\sim 73\%/77\%$[c]	3.97	Fig. S3
Visual change detection of gratings	89.25%	$\sim 60\%$[d]	3.90	Fig. S4
Evidence accumulation	90.92%	$\sim 85\%$[e]	3.96	Fig. S5

(a) Estimated from Fig. 4C of [Lee et al., 2012] when the orientation difference was 90°.
(b) In the visual change detection task for natural images, the two values refer to testing with familiar and novel images, respectively.
(c) Estimated from Fig. 1I of [Garrett et al., 2020] when familiar and novel images were presented to mice, respectively.
(d) Estimated from Fig. 3A of [Glickfeld et al., 2013] when the orientation difference was 5°.
(e) Estimated from Fig. 1C of [Morcos and Harvey, 2016] when the number of cues was 6.
The behavior experiments had longer time delays that made tasks more difficult.

power-law, with an exponent that represents a quasi-optimal compromise between presentations that support fast learning of downstream decisions and presentations that support the continual refinement of downstream decisions. A major surprise of the Billeh model is that it reproduces these experimental findings with remarkable precision, although the model had not been designed or trained for that.

We analyze the response of neurons in the trained Billeh model in the same way as responses of V1 neurons were analyzed in [Stringer et al., 2019a]: A set of 2800 natural images are presented twice in the same order. The neural signal for an input image is defined as the correlated neural response for the two representations, and the noise in each presentation as the residual after subtraction of the signal. An unbiased estimate of the explained variance of the neural signal along with the nth principal component (computed from the first presentation) is achieved by computing the covariance of the projections of neural responses for the two repeats onto this component. We refer to Fig. 4A for an illustration and to the Suppl. of [Stringer et al., 2019a] for a detailed mathematical analysis of this method, referred to as cross-validated PCA, abbreviated cvPCA.

We demonstrate that the neural response of the trained Billeh model to natural images does not lie on any low-dimensional plane within the space of possible firing patterns. We chose 2800 natural images from the ImageNet database (same images as in [Stringer et al., 2019a]) and analyzed the cumulative variance in Fig. 4C. We found that, like in the recordings from area V1 in mouse, the amount of explained variance continued to increase as further dimensions were included, without saturating at any dimensionality below the maximum possible for this stimulus set. As a control, we carried out the same analysis for a set of 32 images that were each shown 96 times. We found that the variance saturated after 32 dimensions (Fig. 4C), thereby reproducing in the neural codes again the full dimensionality of the input space. We refer to the function that gives the variance of the neural signal projected onto the nth principal component as eigenspectrum, following the terminology of [Stringer et al., 2019a] In Fig. 4D, we compare the eigenspectrum of the Billeh model before and after training with the measured eigenspectrum of V1 responses from [Stringer et al., 2019a]. One sees that the training process for 5 unrelated visual processing tasks moves the eigenspectrum of the model close to the eigenspectrum of V1 responses for the same image set. The latter was found to follow a power-law with an exponent close to $\alpha = 1 + 2/d$, where d is the dimension of the stimulus space. This particular exponent was argued to be optimal from two perspectives: for creating maximally informative codes and for creating codes that are both easy to learn for downstream networks but still support the continual refinement of learning precision.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The eigenspectrum of the trained Billeh model can be fit to a power-law function with exponent $\alpha = 1.17$, see Fig. 5A. This value is somewhat higher than the theoretical optimum, but the power-law exponent of eigenspectrum approximates the formula $\alpha = 1 + 2/d$ also for 3 other stimulus sets with lower dimensionality Fig. 5C-F.

We also verified that the power law of the trained Billeh model can not be explained by the well-known power-law structure of natural images [Ruderman and Bialek, 1994]. To show this, we removed the image power law by spatially whitening the images, and presented the whitened stimuli to the model. Although the power law in the image pixels is abolished, the power law in the neural responses of the model remains valid (Fig. 5B). These results are similar to the experimental data for the mouse V1 [Stringer et al., 2019a].

We also analyzed the eigenspectrum for the untrained Billeh model, partially trained version of the Billeh model (in terms of the number of training epochs), and control models with the same number of neurons and synapses as the Billeh model but where all data-based structure was removed (Methods). To quantify how close the actual power-law exponent α is to the theoretically ideal values $1 + 2/d$, we calculated the coefficient of determination, R^2 in fitting α to $1 + 2/d$ (Methods). The larger value of R^2 indicates a better fitting. Fig. 5G shows that this R^2 value increases during training, both for the Billeh model and for the equal-sized control model without data-based structure. However, the R^2 value stays in a much lower range for the latter. Hence, our results support the hypotheses that the data-based structure of the Billeh model strongly enhances the quality of neural representations in the model, but this quality only becomes apparent once it is trained for visual processing tasks.

2.4 Noise correlations reduce but do not limit coding fidelity in the data-based model

Fundamental doubts about the capability of neural networks in area V1 of the mouse were raised by [Rumyantsev et al., 2020]. It was argued there, on the basis of simultaneous recordings from up to 1,300 neurons, that correlated noise reduces their capability to encode visual stimuli by sufficiently succinct neural codes that can be robustly separated by downstream networks. More precisely, they measured the discriminability index d', see (Fig. 6A for an illustration, where the distance between the mean of neural codes for two stimuli is divided by the standard deviation of the neural codes. The square $(d\prime)^2$ of this value is related to the Fisher information [Cover, 1999]. They found that the normalized $(d\prime)^2$ appears to reach a plateau when neural codes from the increasingly large number of neurons were integrated, thereby suggesting a fundamental limit on information about a visual stimulus that can be conveyed by neurons in V1 to downstream networks. These findings appear to be in contradiction to the results of [Stringer et al., 2019a], which suggest that the amount of information that is jointly conveyed by increasing number of neurons or PCA components does not reach a plateau at values around 1,000. Hence the question arises whether the Billeh model, which we have shown to reproduce the high-dimensional neural codes found by [Stringer et al., 2019a], also reproduces the information-limiting results due to noise correlations reported by [Rumyantsev et al., 2020]. As in their experiments, we used static gratings as visual stimuli, and the task was to judge whether their orientation was smaller or larger than 45°. We presented stimuli with orientations (43° or 47°), an orientation difference that was in the range of the reported behavioral discrimination threshold of the mouse [Glickfeld et al., 2013].

We examined how d' varied with the number of neurons from which we "record" in the model, using between 200 to 51,978 randomly sampled neurons. Figure 6B shows that the increase of d'^2 becomes substantially smaller when more than a few thousand neurons are taken into account, but keeps increasing when the number of neurons grows further, up to the number of neurons in the Billeh model. At the same time, we could also verify a strong information-limiting impact of noise correlation in the model,

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

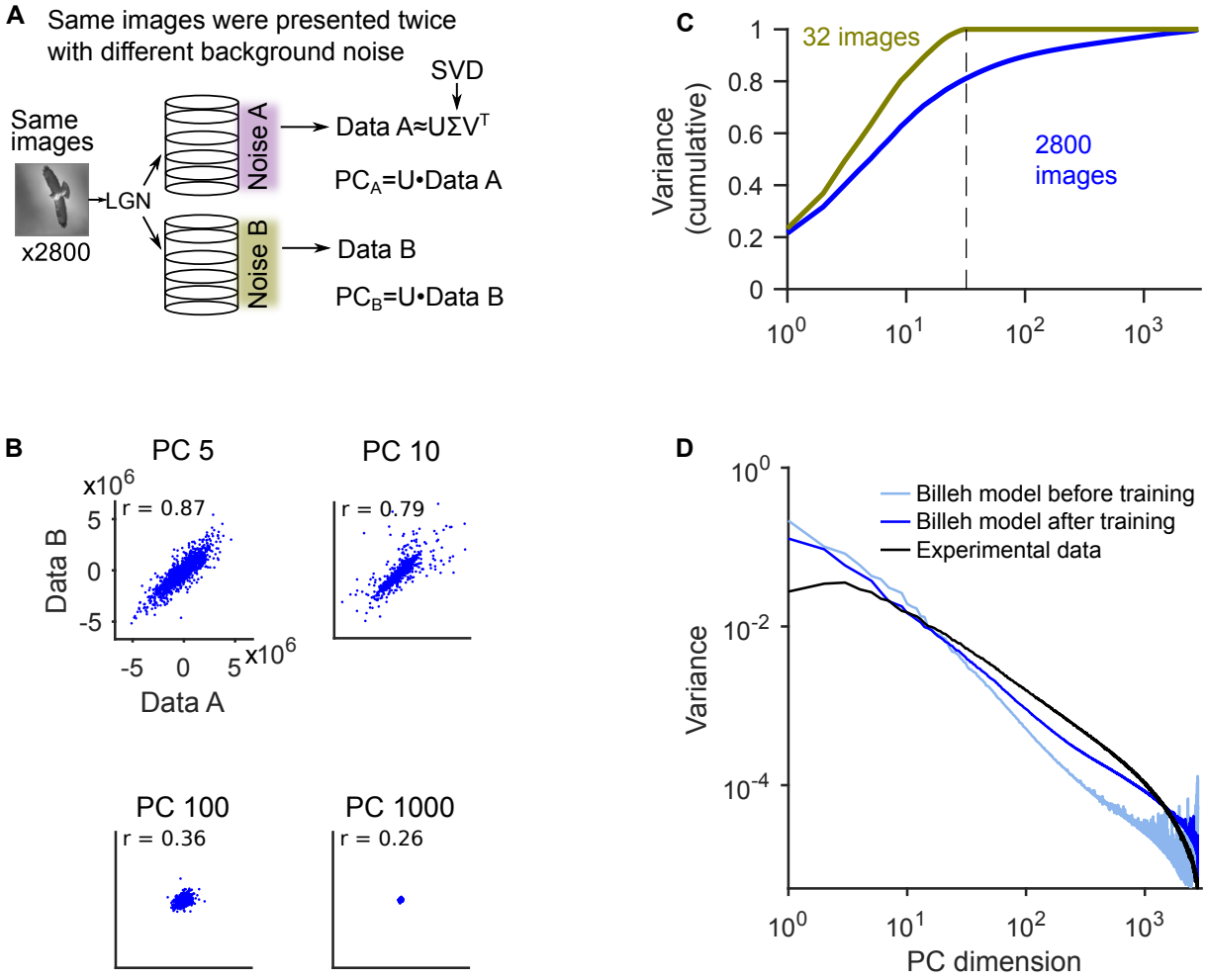


Figure 4: Dimensionality of neural codes of the model matches detailed experimental data. (A) Schematic diagram of cvPCA (cross-validated PCA) according to [Stringer et al., 2019a]. A set of images was presented twice. Noise was independently drawn from the noise distribution for each image (Noise A and Noise B). Eigenspectrum of visual stimulus responses was estimated by cvPCA; neural responses to the first presentation (Data A) were factorized by singular value decomposition (SVD). **(B)** Correlations of neural responses for two presentations of the same natural image, projected to selected principal components as in Extended Data Fig. 5 of [Stringer et al., 2019a]. Each plot shows the responses of all neurons in the trained Billeh model, projected onto the specified PC, for the first repeat (x axis) and second repeat (y axis). Thus each point represents the responses of the network to two presentations of the same stimulus. Altogether there are 2800 points in each panel, for the same set of 2800 natural images as used in [Stringer et al., 2019a]. One sees that the trial-to-trial variability in our modification of the Billeh model is substantial, although in the absence of measurement noise somewhat smaller as in the experiments. **(C)** In the trained Billeh model, the cumulative fraction of variance is shown for increasing numbers of principal components, for an ensemble of 2800 stimuli presented twice (blue) and for 96 presentations of 32 stimuli (green). The dashed line indicates 32 dimensions. One sees that the dimension of neural presentations that the network employs matches the number of images that are presented, similarly as in [Stringer et al., 2019a]. **(D)** Eigenspectra of the untrained/trained Billeh models, and the mouse V1 [Stringer et al., 2019a] responding to an ensemble of 2800 nature images. One sees that the eigenspectrum of the Billeh model is already at the beginning similar to that of the data, and moves during training even closer.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

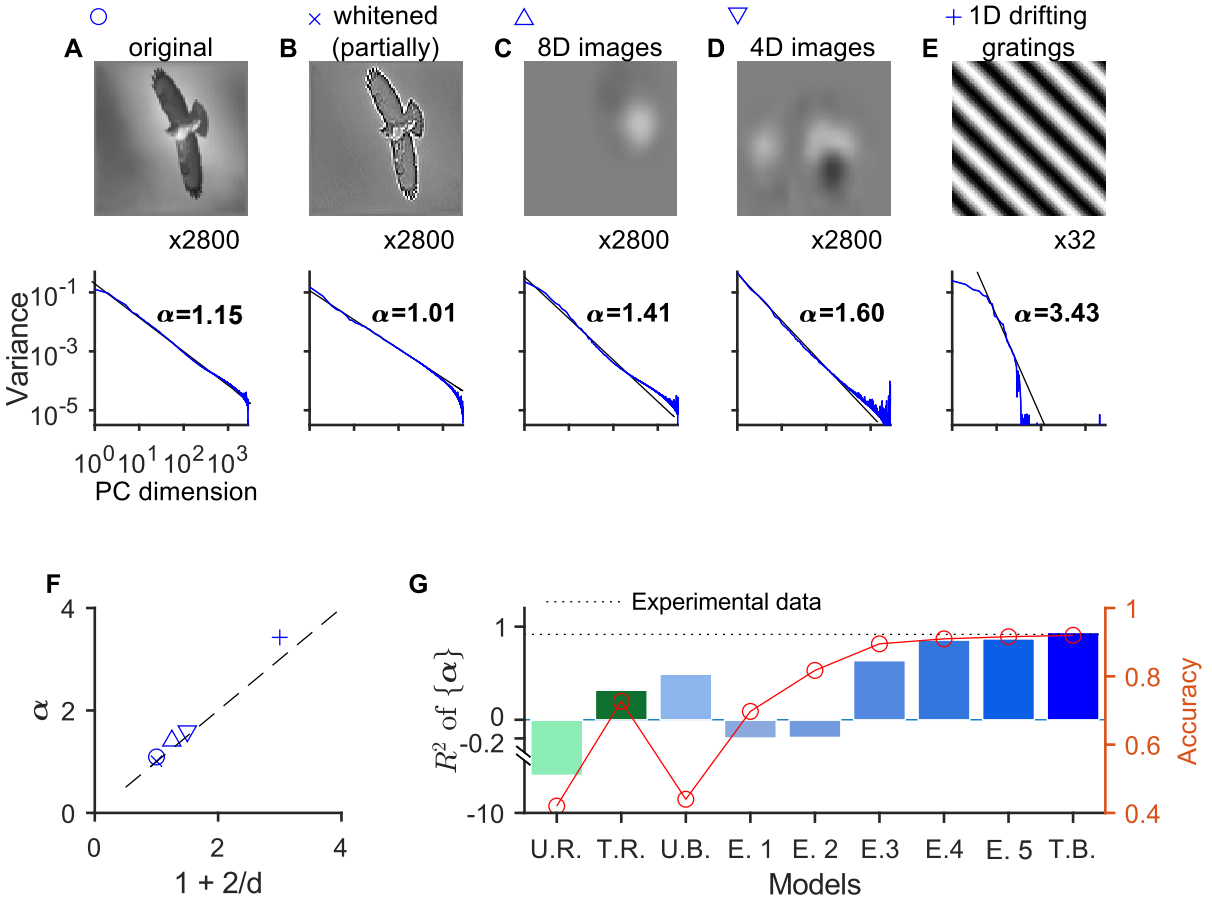


Figure 5: Power-law of the eigenspectrum of neural codes has in the trained Billeh model an exponent that is not far from the theoretical optimum. **(A)**-**(E)** Examples of presented images (top), and eigenspectra of neural codes in the model (bottom), for the original natural images **(A)**, spatially whitened images lacking $1/n$ image spectrum **(B)**, images projected into eight dimensions **(C)**, images projected into four dimensions **(D)**, drifting gratings, one-dimensional stimulus ensemble **(E)**. **(F)** Comparison of power-law exponents α in the model with the theoretical ideal blues $\alpha = 1 + 2/d$ (dashed line). The blue marker for each input ensemble is defined at the top left in **(A)**-**(E)**. **(G)** Quality of fits of power-law exponents (measured by R^2) to ideal values $1 + 2/d$ both for the Billeh model (B), untrained (U), after training episode i (E.i) and fully trained (T), as well as for a control network without biological structure (R). The black dashed line represents the R^2 for the experimental data of [Stringer et al., 2019a]. The red line with circles represents the averaged test accuracy on the 5 tasks. One sees that training improves for both network models the fit of their eigenspectrum to experimental data, but the Billeh model comes substantially closer.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

since the normalized d'^2 assumed for trial-shuffled data values that were by several orders of magnitude larger (Fig. 6C), similarly as in the recorded data [Rumyantsev et al., 2020].

According to [Rumyantsev et al., 2020], there was one interesting factor that limited the impact of correlated noise in their recorded data: They reported that the visual signaling dimensions were nearly orthogonal to the largest noise mode, which therefore had almost no effect on coding fidelity. Interestingly, we find that also in the Billeh model the projection of the signal difference $\Delta\mu$ onto the eigenvector for the largest noise eigenvalue is relatively small (Fig. 6D). Furthermore, compared with the untrained Billeh model (Fig. 6E), training had moved signal dimension to become more orthogonal to the largest noise dimension, similar but less pronounced than in the data of [Rumyantsev et al., 2020].

2.5 The Billeh model suggests a further reason why behavioral performance lags behind neural coding fidelity in area V1

[Stringer et al., 2021] reported that the behavioral discrimination threshold for orientations in the mouse V1 was almost 100 times larger than the discrimination threshold which they inferred from neural coding fidelity of populations of 50,000 neurons in area V1 of the mouse. They conjectured that this difference was caused by the limitations of downstream decoders. The Billeh model suggests a slightly more refined explanation. Direct measurements of coding fidelity on the basis of simultaneous recordings from 50,000 neurons do not account for the fact that their information content has to be extracted by neurons in V1 that project to downstream areas. They are conceptually similar to the postulate of having a global readout neuron that receives synaptic input from all 50,000 neurons, see Fig. 2E. However, one can demonstrate in the Billeh model that such global linear readout attains for the fine orientation discrimination task an accuracy of 98.81%. On the other hand, a pool of 30 projection neurons on L5 could only achieve an accuracy of 93.15% if one assumed that they were localized closely together (Fig. 2C)), and of 93.61% if they were assumed to be randomly distributed in L5 (Fig. 2D). These results suggest that the means by which information from area V1 is extracted and projected to downstream areas is a limiting factor that is likely to contribute to the gap between the performance of an ideal observer of neural activity in V1 and the behavioral performance of mice.

2.6 Relating data on criticality of brain networks to the dynamic regime of cortical microcircuit models

The critical brain hypothesis has inspired numerous theoretical and experimental studies [Mora and Bialek, 2011, Wilting and Priesemann, 2019]. Nonetheless, experimental results are still contradictory. Assessing criticality of brain networks is more intricate than first thought due to undersampling issues. We therefore focus here on a subsampling-invariant measurement of the branching ratio m as discussed in [Wilting and Priesemann, 2019]. It is intended to measure how many other spikes a spike causes on average (Methods). $m < 1$ indicates a subcritical regime, $m > 1$ a supercritical regime, and $m = 1$ a critical regime (Fig. 7A). We calculated the branching ratios of untrained/trained RSNN and Billeh models based on all neurons, and alternatively based just on all excitatory neurons (Fig. 7). Resulting values are in the range of *in vivo* spike recordings in many brain areas and species [Wilting and Priesemann, 2018a], although the number of neurons in the models is far larger than those in experimental recordings. In particular, both the untrained and trained Billeh model, containing 51,978 neurons operate according to our measurements in the same subcritical regime as neural networks of the brain.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

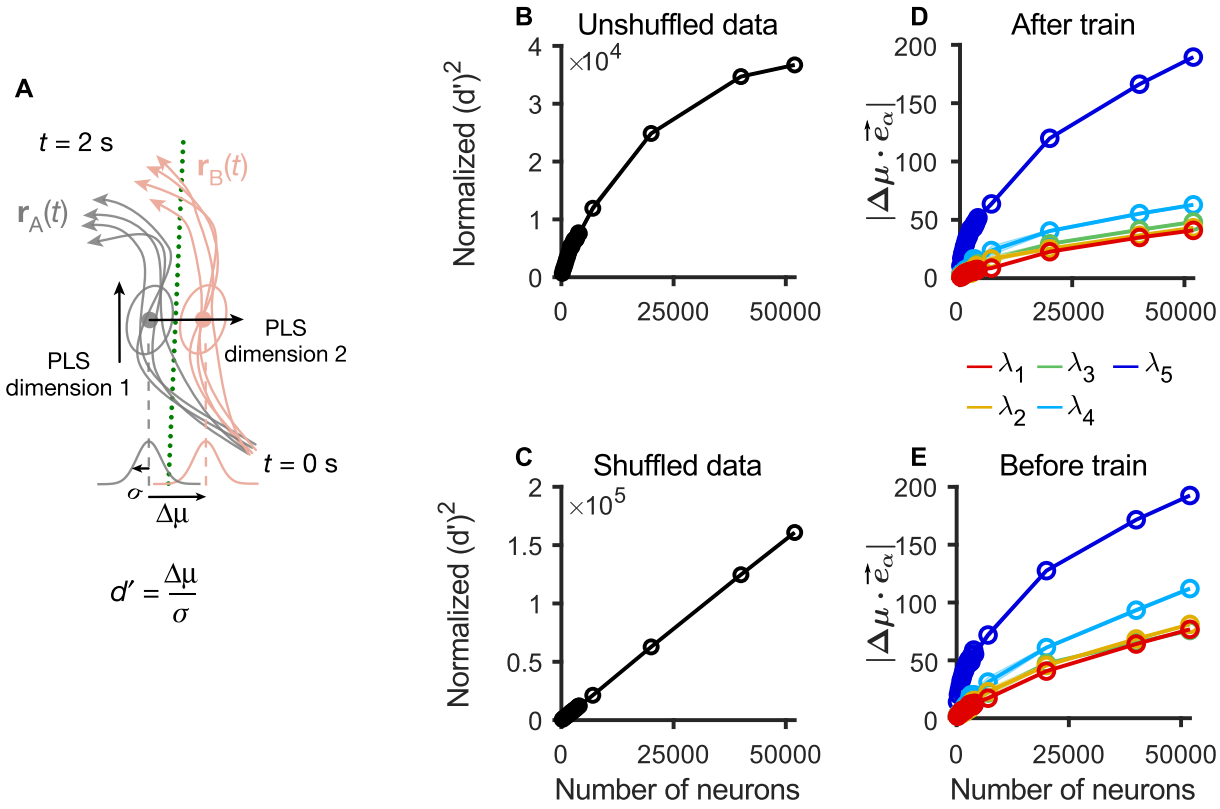


Figure 6: The impact of correlated noise on neural coding fidelity in the Billeh model qualitatively matches experimental data. (A) Schematic of neural ensemble dynamics in a population vector representation of reduced dimensionality. Trajectories, $\mathbf{r}_A(t)$ and $\mathbf{r}_B(t)$, depict single-trial responses to different (gray, red) stimuli. At a fixed time after stimulus onset, the sets of responses to the two stimuli form two distributions of points (ellipses). At the bottom left are projections of these distributions onto a subspace, found by PLS (partial least square) analysis, in which responses to the two stimuli are most distinct. The green line indicates the optimal linear boundary for classifying stimuli in this subspace. The stimulus discriminability, d', is defined as the separation, $\Delta\mu$, of the two distributions along the dimension orthogonal to this boundary, divided by the s.d., σ, of each distribution along this dimension. **(B)** $(d')^2$ values in the trained Billeh model during the interval 1 s from stimulation onset, plotted against the number of neurons used for analysis. The shading area indicates the standard deviation calculated over 100 different subsets of neurons. d' values are normalized by those obtained for trial-shuffled data (averaged across 1 s). **(C)** Same as in **(B)** but for trial-shuffled datasets in which the activity traces of each neuron are randomly permuted across all trials with the same stimulus. The greater value of d'^2 indicates that correlated noise degrades stimulus representations in the unshuffled data. **(D)**, **(E)** Training moves the neural coding dimensions of the Billeh model so that they become more orthogonal to the dominant noise dimension. Each color denotes a different eigenvector of the noise covariance matrix, numbered in decreasing order. The largest eigenvalue λ_1 is 3.34 times larger than the 2nd largest one λ_2, but training drastically reduced the projection of the neural coding dimensions on its eigenvector.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

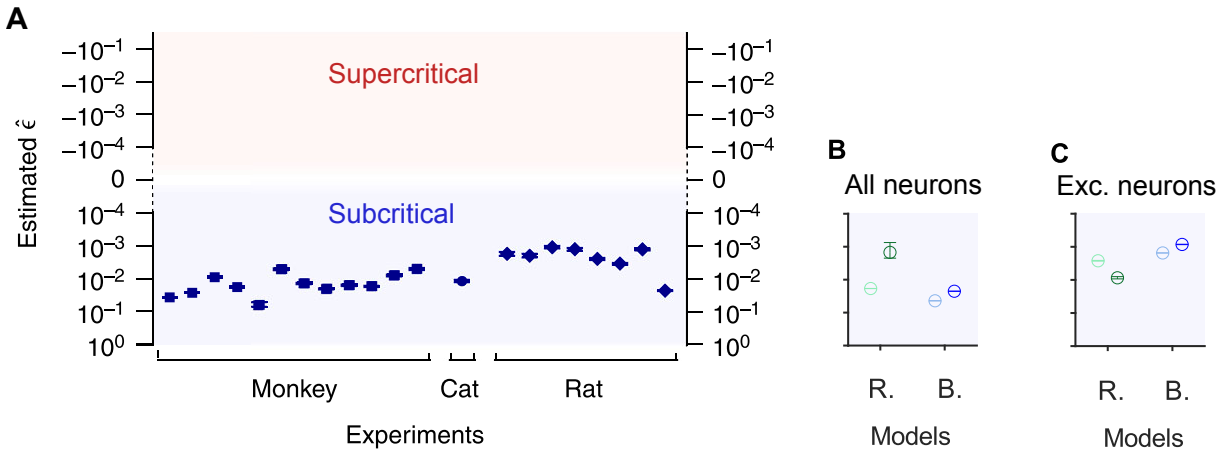


Figure 7: Comparison of the dynamic regime of the Billeh model with that of neural networks in the brain. (A) Positive values of $\epsilon = 1 - m$, where m is the branching ratio indicate a subcritical regime of a dynamical system. Data from spike recordings from rats, cats, and monkeys according to [Wilting and Priesemann, 2018a] are shown, their median $\hat{m} = 0.98$ indicating subcriticality. Error bars: 16 to 84% confidence intervals, note that some confidence intervals are too small to be resolved). **(B)** $\hat{\epsilon}$ estimated from all neurons in untrained (light color) and trained (dark color) RSNN (R.) and Billeh model (B.). **(C)** Same as in **(B)** but only for the excitatory (Exc.) neurons in the models. Scale as in **(A)**. One clearly sees that the models operate with regard to their branching ratio in the same dynamical regime as the brain networks.

2.7 The impact of noise on neural coding and task performance

We analyzed in the Billeh model also how high-dimensional neural coding properties according to [Stringer et al., 2019a] and task performance depend on the type of noise in the model. More precisely, we varied during this analysis the scaling factors q and s for the independent quick and slow noise currents in each neuron, for the fixed setting of synaptic weights that resulted from training with a low amplitude quick noise, and no slow noise. The results in (Table 2) show that task performance of the Billeh model generalizes very well to different noise settings, and tolerates in particular also substantial amplitudes of slow noise. In addition, the fitted power-law exponent of eigenspectrum remains invariant over a wide range of noise amplitudes. However, a remarkable effect occurs for the case of large scaling factors 10 for quick and slow noise: While task performance remains very high for all 5 tasks, the power law of the PCA eigenspectrum breaks down (Fig. S8). This type of noise may be stronger than that found in area V1, but this result for the Billeh model makes an important and unexpected point for the theory of neural networks in the brain, more precisely for establishing links between their neural coding properties and computational capability: It shows that the power law of the PCA eigenspectrum found in [Stringer et al., 2019a] is not necessary for achieving high accuracy in the 5 visual processing tasks that we are considering.

2.8 Salient features of cortical microcircuits speed up gradient descent learning of the model

We applied the same training procedure also to control models that lacked salient structural features of the Billeh model, and found that their task performance advances substantially slower. In particular, we applied the same training procedure that we had applied to the Billeh model also to control models where its laminar connectivity structure with topographic maps between laminae was replaced by random

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 2: Impact of different amplitudes of our noise model on the eigenspectrum of neural codes, the mean firing rate, and average performance for the 5 visual processing tasks. properties. The light cyan highlights the results for the default noise model used in this study.

	Power-law[a] eigenspectrum	Mean firing rate[b] (Hz)	Average acc.
$q = 1,\ s = 0$	$\alpha = 1.11,\ I > 100$	3.77	90.51%
$q = 1,\ s = 2$	$\alpha = 1.09,\ I > 100$	3.79	88.87%
$q = 1,\ s = 3$	$\alpha = 1.31,\ I > 100$	3.81	88.77%
$q = 1,\ s = 3.5$	$\alpha = 1.25,\ I > 100$	3.88	88.58%
$q = 1,\ s = 4$	$\alpha = 1.30,\ I > 100$	3.94	88.71%
$q = 2,\ s = 2$	$\alpha = 1.15,\ I > 100$	3.94	89.10%
$q = 2,\ s = 3.5$	$\alpha = 1.4,\ I > 100$	3.92	88.68%
$q = 3,\ s = 3$	$\alpha = 1.19,\ I > 100$	3.94	89.31%
$q = 3,\ s = 4$	$\alpha = 1.31,\ I > 100$	3.99	88.62%
$q = 4,\ s = 3$	$\alpha = 1.21,\ I > 100$	3.97	88.97%
$q = 4,\ s = 4$	$\alpha = 1.58,\ 0 < I < 100$	4.03	88.21%
$q = 5,\ s = 4$	$\alpha = 1.38,\ 0 < I < 100$	4.07	88.58%
$q = 10,\ s = 10$	$\alpha = N/A,\ I = 0$	4.18	87.89%
$q = 20,\ s = 20$	$\alpha = N/A,\ I = 0$	4.28	83.75%
$q = 200,\ s = 200$	$\alpha = N/A,\ I = 0$	10.19	50.77%

(a) α is the power-law exponent of eigenspectrum when the trained Billeh model response to 2,800 nature images. I is the interval of the best fitted power-law function ($R^2 > 0.8$); $I = 0$ if the best fitting $R^2 < 0.8$.
(b) The mean firing rate measured in the visual change detection task for nature images.

choice but keep the same number of synaptic connections, or $GLIF_3$ model was replaced by standard LIF neuron model, or all data-based details of the Billeh model were deleted (RSNN). We found that during the training time, the Billeh model achieved an approximately saturating task performance, whereas these control models were only able to reach a substantially lower task performance level, see Fig. 8A. The same effect could also be observed for a smaller version of the Billeh model with just 5,000 neurons, which we compared with RSNNs containing the same initial weights, numbers of neurons and synaptic connections (Fig. 8B). In this case, we found that the performance level of the generic RSNN did not even saturate when it was trained much longer, and remained at a performance level significantly below that of the Billeh model. Substantially larger computing resources will be needed to determine the performance levels that these control models can eventually reach after sufficiently long training.

It had already been shown that neuron models with slower changing internal variables tend to enhance BPTT training, see Fig. 2D of [Plank et al., 2021], and Fig. 3C and Supplementary Movie of [Bellec et al., 2020]. Fig. 3C of the latter reference also shows that a similar training advantage holds for a biologically more plausible variant of gradient descent learning. But Fig. 8A shows that also the laminar connectivity structure of the Billeh model contributes to its learning speed, even it the neuron models remain unchanged (yellow curve). One possible explanation is that the laminar structure enforces topographic maps between different layers, and hence tends to keep information more local within the network. The data-based rapid spatial decay of connection probabilities within a layer has a similar effect. This locality of information processing may facilitate learning through local learning mechanisms in such network. In contrast, in a generic randomly connected network without this connectivity structure, all information is continuously dissipated throughout the network, which is likely to impede the localization of processing errors and their correction.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

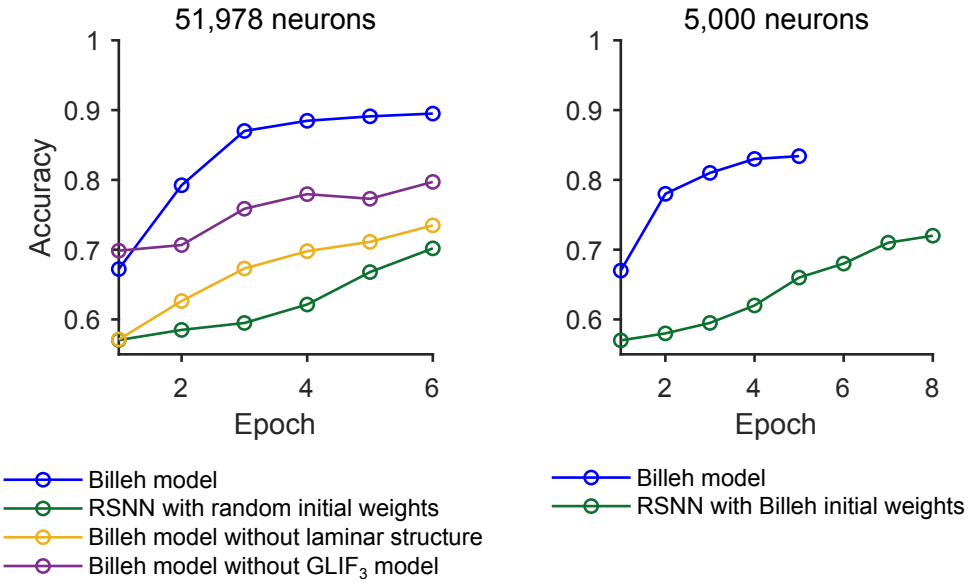


Figure 8: Data-based neural network models learn substantially faster. (A) Average test accuracy of five tasks as a function of training epoch. One epoch contains 781 iterations of weight updates (Methods), costing around 3.2 hours of wall time on 160 GPUs (A100). All models contain 51,978 neurons. **(B)** Same as in **(A)** but for RSNN and Billeh models with 5,000 neurons (column radius:124 µm). The input images and movies to LGN model are resized to 44×52, depending on the receptive fields of the 5,000 neurons. The number of readout neurons in each group of L5P is reduced to 20.

3 Discussion

We have demonstrated the feasibility of a research method where neural coding, the impact of noise, dynamical network regimes, and computational capabilities are investigated in a detailed large-scale model of a patch of a particular neocortical area in a particular species, and compared with recordings from many neurons in the same area, the same species, for the same visual inputs, and for the same computational tasks. This method has become feasible because of substantial advances in computer science and computing technology. It promises to substantially advance our insight into the link between structure and function of neural networks of the brain.

Specifically, we analyzed the detailed large-scale model for a patch of area V1 in mouse from [Billeh et al., 2020], and compared results from this model with recordings from large numbers of neurons in area V1 of mouse, for the same behavioral (computational) tasks, using the same visual stimuli. The results surprised us in several aspects. First, we found that one can assign values to those parameters of the model that are not constrained by experimental data, the synaptic weights, so that the model can carry out -with the same weight-setting- 5 different visual processing tasks that have already been used as behavioral tasks in biological experiments. Furthermore, the trained model achieved for these 5 tasks about the same performance as trained mice. It was not obvious that training of the model could achieve that, since we used a substantially more constrained and arguable biologically more realistic rule for extracting the computational result from the model: We demanded that the firing of concrete populations of projection neurons on L5 encode and transmit the result of the network computation rather than an abstract global readout neuron. In addition, we employed a biologically more realistic model for noise in the network, which additionally constrained the computing capability of the network.

One should, however, also be aware of two differences between the 5 visual processing tasks in the model

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and in the animal experiments: In order to reduce computing time for training the Billeh model each visual processing task was implemented within trials that lasted just 600 ms, whereas the tasks in the experiments often contained longer delays. In addition, we were not able to model the contribution of other brain areas than LGN and V1 for solving these tasks, although other brain areas are likely to be salient -especially for the case of longer delays within a trial. On the side, we would like to emphasize that we tested the trained model -whenever possible- for new visual stimuli that had not been shown during training (this was not possible for the gratings that occurred in some of the experiments). In that sense, we evaluated the generalization capability of the trained Billeh model, rather than its capability to handle a fixed set of stimuli correctly (which often suffices to solve behavioral tasks in experiments).

A rather surprising result was that the trained Billeh model automatically reproduced a number of details of experimental recordings for which it had not been trained. In particular, high-dimensional neural codes for natural images emerge in this model, as reported in [Stringer et al., 2019a], with an exponent in the power law of the PCA spectrum that is very close to biological measurements. To be precise, the untrained Billeh model had already a PCA spectrum that was in the range of the recorded data (see Fig. 4D). But the training process moved it substantially closer to the experimental data, see Fig. 4D and Fig. 5G. Hence this PCA spectrum appears to be helpful for solving the 5 visual processing tasks. This is especially remarkable in view of the theoretical analysis of [Stringer et al., 2019a], which suggests that this particular exponent is close to optimal for combining a maximum of information about the stimulus with good generalization capability.

In addition, we found that correlated noise does increase the difficulty to discriminate neural responses to slightly different pairs of visual stimuli in the Billeh model, as found for mouse V1 in [Rumyantsev et al., 2020]. But since we could carry out this analysis for much larger number of neurons than those recorded in [Rumyantsev et al., 2020], our modeling result produced a concrete and somewhat surprising prediction (see Figure 6B): That recordings from larger numbers of neurons in V1 will find that the discrimination capability does not saturate, but will continue to increase when larger neural populations are taken into account. We also would like to mention that conceptually there is no contradiction between the high-dimensional neural codes found by [Stringer et al., 2019a] and the limiting impact of correlated noise found by [Rumyantsev et al., 2020]: The latter focused on the discrimination of two very similar visual stimuli, whereas the former analyzed neural codes for a large repertoire for natural images.

We also examined the dynamic regime of the model, using the subsampling-invariant estimator proposed by [Wilting and Priesemann, 2018a] based on the branching ratio. We found that according to this estimator the Billeh model operates in the same slightly subcritical regime as neural networks in the brain.

Another unexpected result of the model was that its computational performance (accuracy) remained, without adaptation of synaptic weights or other parameters, within the range of observed behavioral performance even when we drastically increased the noise during testing. In fact, we identified a high noise regime where the power law behavior of the PCA spectrum that was found in area V1 of mouse [Stringer et al., 2019a] was destroyed, but not the task performance of the model. This modeling result rise to an immediate question for further experimental work: Can one also destroy in the living brain this power law without significantly reducing behavioural performance for visual processing tasks?

But even before this prediction can be tested experimentally, our result has an immediate consequence for the theory in the Supplement of [Stringer et al., 2019a], where it had been shown that this power law is in a precise sense essential for effective neural coding and learning. Our result implies that this theory does not apply to the 5 visual processing tasks that we considered (although the experimentally found power law appears to be helpful for solving them, since training for these tasks caused the model to approximate this power law). Since these 5 tasks were among the most challenging visual processing tasks that mice appear to be able to solve, our result gives rise to the question of whether more challenging

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

visual processing tasks can be found that can be accomplished by mice, and for which the found power law of the PCA spectrum is essential.

Another suggestion for further research that results from our results is that it would be good to take a closer look at the amplitude and statistical structure of noise both in the brain. Noise has often also been neglected in models of neural circuits in the brain, and if noise was included, often little attention has been placed on calibrating its amplitude and distribution on the basis of experimental data. Data-based large-scale models offer an excellent venue for studying the impact of specific forms of noise on concrete neural coding and computing properties, and for comparing these results. As we have seen, such analysis appears to be relevant for testing the validity of theories on neural coding and computation.

Computational properties of neural networks in the brain have so far been primarily investigated in generic networks of recurrently connected networks of artificial neurons or generic spiking neurons. We have demonstrated that one can expand such investigations now to include also rather detailed data-based models, whose properties appear to be in several fundamental aspects quite different. In addition, these more detailed models make it much easier to compare modeling results with biological data, since these models are able to make specific predictions for the spike output of specific types of neurons on specific layers of cortical microcircuits.

Finally, an unexpected result of training the data-based model of [Billeh et al., 2020] was that this model was easier to train, i.e., it learned faster, than control models where essential features of this data-based model, such as laminar structure or generalizations of LIF neuron models that also model slower internal processes, were deleted (Fig. 8). It will be interesting to see whether this effect also arises for biologically more realistic learning methods. We did not try to train the Billeh model through a method that can be argued to be biologically realistic, such as [Bellec et al., 2020], because training with such methods tends to take substantially more trials, and therefore exceeded our computing resources. But since we now know that the Billeh model is in fact able to carry out the 5 visual processing tasks that we considered, one can now look for various biologically more realistic ways in the network initialization and/or learning algorithm -also for smaller instances of the Billeh model- that can induce a similar computing capability or neural coding features.

An interesting side result is that the trained Billeh model is able to solve the 5 visual processing tasks in an energy-efficient sparse firing regime of around 4 Hz. This demonstrates that visual processing is possible in large networks of spiking neurons with spike- rather than rate-coding, hence in an operating regime that is desirable for the design of architectures and algorithms for highly energy-efficient spike-based neuromorphic chips [Davies et al., 2021].

4 Methods

4.1 Neuron model

We base our study on the "core" part of the point-neuron version of the realistic V1 model introduced by [Billeh et al., 2020], containing 51,978 neurons from 111 different data-based neuron types. Each neuron type was modeled by the GLIF$_3$ model [Teeter et al., 2018]. To make it gradient-friendly, we replaced the hard reset of membrane potential after a spike emerges with the reduction of membrane potential $z_j(t)v_{th}$, where $z_j(t) = 1$ when neuron j fires at time t and $z_j(t) = 0$ otherwise. v_{th} is the firing threshold of membrane potential. We simulated each trial for 600 ms. The dynamics of the modified GLIF$_3$ model

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

was defined as

$$\begin{aligned} v_j(t+1) &= \alpha v_j(t) + \frac{1-\alpha\tau}{C}\left(I_j^e(t+1) + \sum_m I_j^m(t+1) + gE_{\mathrm{L}} + I_j^{\mathrm{syn}}(t)\right) - z_j(t)v_{\mathrm{th}} \\ z_j(t) &= H\left(v_j(t) - v_{\mathrm{th}}\right) \\ I_j^e(t) &= \sum_i W_{ji}^{\mathrm{in}} x_i(t) + qK_j^{\mathrm{quick}}(t) + sK_j^{\mathrm{slow}} \end{aligned} \tag{2}$$

where C represents the neuron's capacitance, E_L, the resting membrane potential, I^e, the external current, I^{syn}, the synaptic current, Θ, the spiking threshold. W_{ji}^{in} is the synaptic weight from LGN neuron i to V1 neuron j. The scales of the quick noise $K_j^{\mathrm{quick}}(t)$ and the slow noise K_j^{slow} to neuron j are $q = 2$ and $s = 2$, respectively. K_j was randomly drawn from the empirical noise distribution on which will be elaborated later. The decay factor α is given by $e^{-\delta t/\tau}$, where τ is the membrane time constant. δt denotes the discrete-time step size, which is set to 1 ms in our simulations. H denotes the Heaviside step function. Owing to the term $-z_j^t v_{\mathrm{th}}$, the neuron membrane potential is reduced by a constant value after an output spike, which relates our model to the spike response model [Gerstner et al., 2014]. To introduce a simple model of neuronal refractoriness, we further assumed that z_j^t is fixed to 0 after each spike of neuron j for a short refractory period depending on the neuron type. The after-spike current $I^m(t)$ was modeled as

$$I^m(t+1) = f^m I^m(t) + z(t)\delta I^m;\ m = 1, \ldots, N_{\mathrm{asc}}, \tag{3}$$

where the multiplicative constant $f^m = \exp\left(-k^m \delta t\right)$ and an additive constant, δI^m. In our study, $N_{\mathrm{asc}} = 2$. All neuron parameters depended on the neuron type as in [Billeh et al., 2020].

4.2 Synaptic Characteristics

Postsynaptic current-based synaptic mechanisms were used with dynamics described by an alpha-function:

$$I_{\mathrm{syn}}(t) = \frac{eW^{\mathrm{rec}}}{\tau_{syn}} t e^{-\frac{t}{\tau_{syn}}} \tag{4}$$

where I_{syn} is the postsynaptic current, τ_{syn} is the synaptic port time constant, and W^{rec} is the recurrent connection weight. The τ_{syn} constants for the mechanisms were 5.5 ms for excitatory-to-excitatory synapses, 8.5 ms for inhibitory-to-excitatory synapses, 2.8 ms for excitatory-to-inhibitory synapses, and 5.8 ms for inhibitory-to-inhibitory connections, which were extracted from [Billeh et al., 2020]. The synaptic delay is spread in $[1, 4]$ ms, which was extracted from the Fig. 4E of [Billeh et al., 2020] and converted to integers as the integration step is 1 ms.

4.3 Initial conditions

The initial conditions of spikes and membrane potentials were zero. The initial conditions of W^{in} and W^{rec} were given by the values in [Billeh et al., 2020] unless stated otherwise.

4.4 Noise models

The noises currents $K_j^{\mathrm{quick}}(t)$ and K_j^{slow} in Eq. 2 were randomly drawn from an empirical noise distribution. The quick noise $K_j^{\mathrm{quick}}(t)$ was drawn independently for all neurons in every 1 ms; the slow noise K_j^{slow}

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

was drawn independently for all neurons once 600 ms. The empirical noise distribution (Fig. 2A) was from the additive noise decoded from experimental data of mice response to 2,800 nature images [Stringer et al., 2019a] (https://figshare.com/articles/Recordings_of_ten_thousand_neurons_in_visual_cortex_in_response_to_2_800_natural_images/6845348). The decoding method was cross-validation principal component analysis (cvPCA) [Stringer et al., 2019b]; more precisely, the code in https://github.com/MouseLand/stringer-pachitariu-et-al-2018b.

4.5 Readout populations

In total, there were 15 readout populations in the V1 model, encoding the network decisions for the 5 visual processing tasks. Each population was 30 randomly selected excitatory neurons in layer 5, within a sphere of a radius of 55 µm. They were spatially located far away from each other (Fig. 3A).

4.6 Visual tasks

To demonstrate that the realistic V1 model is able to perform multiple tasks, we trained the model with five tasks together; each task was distributed in different 64 bathes. All stimuli of these tasks were given via the LGN model (Fig. 3A). As the requirement of the LGN model input, the original image was converted to gray-scale images and scaled the gray value of the image pixels to $[-Int, Int]$, $Int > 0$. The LGN model output was sent to the neural network model as external current, i.e.,

$$I_{\text{sti}} = W_{\text{input}} \cdot \text{LGN}(G_{Int}), \tag{5}$$

where G_{Int} the scaled gray images in $[-Int, Int]$.

Fine orientation discrimination task We set up the fine orientation discrimination task as in [Stringer et al., 2021] and [Rumyantsev et al., 2020] (Fig. 3C). Stimuli were sinusoidal drifting gratings (spatial frequency, 0.05 cycles per degree, drifting rate, 2 Hz). The intensity of the grating was 2 (Int in Eq. 5). The initial phase was randomly sampled. The inputting sequence consisted of 50 ms delay, 100 ms drifting gratings, and 50 ms delay. In the post-delay period, one readout population needed to report if the orientation in the drifting grating is larger than 45°. In the training set, the orientation was drawn from $[43, 47]^\circ$ (i.e., 45 ± 2) with the precision of 0.1°. The orientation difference was the same as in [Stringer et al., 2021].

Image classification task To isolate the classifying ability of the neural network, we trained the model to classify MNIST dataset (Fig. 3D). MNIST is made of simple curves without complex features and structures, which provides a suitable task that the V1 can potentially recognize without other association areas. The intensity of the image input to LGN was 2 (Int in Eq. 5). The inputting sequence consisted of 50 ms delay, 100 ms static image, and 50 ms delay. The neural network reported the classification in a window of 50 ms, 100 ms after the onset of the image. In the post-delay period, ten readout populations needed to report the image class.

Visual change detection task with natural images We trained the model to perform the visual change detection task considered in [Garrett et al., 2020, Siegle et al., 2021]. In this task, the model was presented with a sequence of images, interleaved by delays of gray screen, and has to report whenever the newly presented image differs from the previously shown one (Fig. 3B). More precisely, images were presented for 100 ms each, with the gray delays between them lasting for 200 ms. Note that the first image was presented after 50 ms. All images were selected from a set of 40 possible images from the ImageNet dataset. The changing probability of image identities is 50%. In case of a changed image

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

identity, the model had to perform a report within a time window of 50 ms length after 150 ms of image onset (response window).

Visual change detection task with drifting gratings We also replaced the natural images above with static gratings which have different orientations and kept the input sequence the same (Fig. 3B). The setting of the static grating is the same as in the fine orientation discrimination task except it is static. The changing probability of orientation is 50%; the orientation of static gratings was drawn in $[120, 150]$ (i.e., 135 ± 15) with the precision of 0.1°.

Evidence accumulation task A hallmark of cognitive computations in the brain is the capability to go beyond a purely reactive mode: to integrate diverse sensory cues over time, and to wait until the right moment arrives for an action. A large number of experiments in neuroscience analyze neural coding after learning such tasks (see e.g., [Morcos and Harvey, 2016, Engelhard et al., 2019]). We considered the same task that was studied in the experiments of [Morcos and Harvey, 2016, Engelhard et al., 2019]. There a rodent moved along a linear track in a virtual environment, where it encountered several visual cues on the left and right (Fig. 3E). Later, when it arrived at a T-junction, it had to decide whether to turn left or right. The network should report the direction from which it had previously received the majority of visual cues. The right (left) cue was represented by 50 ms of cue image in which the black dots on the right (left) side of the maze. Visual cues were separated by 10 ms, represented by the gray wall of the maze. After a delay of 250 ms, it has to decide if more cues were received on the left or right via two readout populations.

For the fine orientation discrimination task and the image classification task, 400 ms delay was added after the response window (Fig. 3B, C) to make the simulation time equal with other tasks. Besides the visual detection tasks, the spikes and membrane potentials were reset to 0 after a trial.

4.7 Loss function

The loss function was defined as

$$L = L_{\text{cross-entropy}} + \lambda_f L_{\text{rate reg.}} + \lambda_v L_{\text{v reg.}}, \tag{6}$$

where $L_{\text{cross-entropy}}$ represents the cross-entropy loss, λ_f and λ_v represent the weights of firing-rate regularization $L_{\text{rate reg.}}$ and voltage regularization $L_{\text{v reg.}}$, respectively. As an example, the cross-entropy loss of visual change detection tasks was given by

$$L_{\text{cross-entropy}} = -\sum_n \left[t^{(n)} \log \sigma \left(\theta \left(r^{(n)} - r_0 \right) \right) + \left(1 - t^{(n)} \right) \log \sigma \left(\theta \left(r^{(n)} - r_0 \right) \right) \right], \tag{7}$$

where the sum over n is organized into chunks of 50 ms and $r^{(n)}$ denotes the population firing rate of the readout neurons in that time interval. Similarly, $t^{(n)}$ denotes the target output in that time window, being 1 if a change in image identity should be reported and otherwise 0. r_0 denotes a baseline firing rate. σ represents the sigmoid function. θ is trainable scale ($\theta > 0$) of firing rate.

We also used regularization terms to penalize unrealistic firing rates as well as unrealistic membrane voltages. Their weights, $\lambda_{\text{rate reg.}} = 0.1$ and $\lambda_{\text{v reg.}} = 10^{-5}$. The rate regularization is given by the Huber loss [Huber, 1992] between the target firing rates, y, calculated from the model in Ref. [Billeh et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2020], and the firing rates, r, sampled the same number of neurons from the network model:

$$L_{\text{rate reg.}} = \sum_{j}^{N} \left|\tau_j - \mathbb{I}\left\{\delta_j < 0\right\}\right| \frac{\mathcal{L}_\kappa\left(\delta_j\right)}{\kappa}, \quad \text{with}$$
$$\mathcal{L}_\kappa\left(\delta_j\right) = \begin{cases} \frac{1}{2}\delta_j^2, & \text{if } |\delta_j| \le \kappa \\ \kappa\left(|\delta_j| - \frac{1}{2}\kappa\right), & \text{otherwise} \end{cases} \tag{8}$$

where j represents neuron j, N the number of neurons, $\tau_j = j/N$, $\delta = 0.002$, $\delta_j = r_j - r_j^{\text{target}}$. $\mathbb{I}(x) = 1$ when x is true; $\mathbb{I}(x) = 0$ when x is false.

The voltage regularization was given by

$$L_{\text{v reg.}} = \frac{1}{N} \sum_{j=0}^{j=N} \left(\left[\frac{v_j - E_{\text{L}}}{E_{\text{L}}} - 1 \right]^+ \right)^2 + \left(\left[-\frac{v_j - E_{\text{L}}}{E_{\text{L}}} + 1 \right]^+ \right)^2, \tag{9}$$

where N represents the total number of neurons, v_j, the membrane potential of neuron j, E_{L}, the resting membrane potential, $[]^+$, rectifier function.

4.8 Training methods

We applied back-propagation through time (BPTT) [Scherr and Maass, 2021] to minimize the loss function. The non-existing derivative $\frac{\partial z_j^t}{\partial v_j^t}$ was replaced in simulations by a simple nonlinear function of the membrane potential that is called the pseudo-derivative. Outside of the refractory period, we chose a pseudo-derivative of the form

$$\psi_j^t = \gamma_{\text{pd}} \exp\left(\frac{-\left(v_{sc}^t\right)^2}{\sigma_p^2} \right),$$
$$v_{sc}^t = \frac{v^t - v_{\text{th}}}{v_{\text{th}} - E_{\text{L}}}, \tag{10}$$

where the dampening factor $\gamma_{\text{pd}} = 0.5$, the Gaussian kernel width $\sigma_p = 0.28$. During the refractory period, the pseudo derivative is set to 0. During the training, we added the sign constraint on the weights of the neural network to keep the Dale's law. Specifically, if an excitatory weight was updated to a negative value, it would be set to 0; vice versa.

4.9 Other simulation details

The BPTT training algorithm was coded in TensorFlow, which runs very efficiently on multiple GPUs. For 5 tasks, we run independent simulations in parallel by distributing each task in a batch with 64 instances. In every instance of each task, a simulation of the Billeh model for 600 ms of biological time was run, taking around 5 s on a NVIDIA A100 GPU. Once all batches finished their tasks (one step), gradients were calculated and averaged to update the weights by BPTT. We define an epoch as 781 iterations/steps because it is one cycle through the full training dataset of MNIST dataset. This computation had to be iterated 6 epochs in order to achieve high computational performance for the chosen tasks, which took 20 h of wall clock time on 160 GPUs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

4.10 Control model

The control model is a randomly connected recurrent network of spiking neurons with the same numbers of neurons and connections, referred to as RSNN. In RSNN, all data-based features of Billeh model were removed: $GLIF_3$ neuron model was replaced by standard LIF model; neural connectivity representing laminar structure was replaced by random connectivity; diverse neuron types were replaced by a single neuron type (the excitatory neuron on L2/3, node type id in Allen brain atlas: 487661754); Dale's law was removed; initial weights were replaced by random values drawn from the Gaussian distribution with the same mean and variance with the Billeh model.

4.11 High-dimensional population responses

To explore the dimensionality of internal representation in the Billeh model, we input the same stimuli to the V1 model as Stringer et al showed to the mice: nature images, whitened images, 4D images, 8D images, sparse noise, 1D drifting gratings [Stringer et al., 2019a]. Briefly, 2800 nature image stimuli were selected from the ImageNet database. The images were uniformly contrast normalized. Stringer et al presented stimulus replicated across all three screens, but at a different rotation on each screen (see their Fig. 1c). We chose the one shown on the left screen and resized it for the requirement of the LGN model. Additionally, we input a smaller set of 32 images, repeating 90 times, to enable estimation of trial-averaged responses. To break the power-law spectrum of natural images, all images were whitened by dividing their Fourier transform by the averaged Fourier spectrum across all images with a small constant value added for regularization purposes.

The 4D and 8D stimuli were constructed using a reduced-rank regression model. We first used reduced-rank regression to predict the neuronal population responses R from the natural images I ($N_{\text{pixels}} \times N_{\text{stimuli}}$) via a d-dimensional bottleneck: $R = A^T BI$, where A is a matrix of size $d \times N_{\text{neurons}}$ and B is a matrix of size $d \times N_{\text{pixels}}$. The dimensionality d was either eight or four depending on the set of stimuli being constructed. The columns of B represent the image dimensions that linearly explain the most variance in the neural population responses. The stimuli were the original 2800 natural images projected onto the reduced-rank subspace $B : I_{\text{low}-d} = B^T BI$. In addition to natural image stimuli, we also presented drifting gratings of 32 directions. They were presented 90 times each. Their spatial frequency was 0.05 cycles per degree and their temporal frequency was 2 Hz. To simulate the spontaneous activity, we input the image whose pixel values are zero to the LGN model.

All stimuli were input 50 ms after the simulation and sustained for 500 ms in each trial to be the same with experimental procedures, unless stated otherwise. They were played twice to allow cross-validated analysis. The initial condition of membrane potential and spike was set to zeros, unless otherwise stated. We input the 2800-nature-image stimuli 5 times with different random seeds to draw the noise and initial membrane potential from uniform distribution between resting potential and firing threshold, i.e., $U \sim [E_{\text{L}}, v_{\text{th}}]$. The results were not sensitive to the initial condition and noise.

To reproduce the high-dimensional geometry of population responses in the visual cortex [Stringer et al., 2019a], we used cross-validation principal component analysis (cvPCA) to estimate the stimulus-related variance. cvPCA provides a way to analyze the dimensionality of the image encoding in neural system. cvPCA measures the reliable variance of stimulus-related dimensions, excluding trial-to-trial variability from unrelated cognitive and/or behavioral variables or noise. It accomplishes this by computing the covariance of responses between two presentations of an identical stimulus ensemble (Fig. 4A). Because only stimulus-related activity will be correlated across presentations, cvPCA provides an unbiased estimate of the stimulus-related variance. To be consistent with [Stringer et al., 2019a], we sum the spikes of 500 ms in response to stimuli. We ran cvPCA ten times on the response of the neural network fed with the same

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

images that are used in [Stringer et al., 2019a]. On each iteration randomly sampling the population responses of each stimulus from the two repeats without replacement. We ran ten different runs and found they were very similar to each other, i.e, the s.d. was close to 0. For the trained Billeh model, we calculated the eigenspectra in three models trained with different noise and randomly generated data, and found the s.d. is 5.95×10^{-5}. The displayed eigenspectra of the trained Billeh model were averaged over these three models.

Power-law fitting of eigenspectra. Using the least-squares method, we fit power laws to the eigenspectra, $f(n)$, against PC dimension, n. The fitting function is $f(n) = n^{-\alpha}(n \in [n_{\min}, n_{\max}])$, where $n_{\min}$ and $n_{\max}$ are lower and higher bounds, respectively. For each possible pair of $n_{\min}$ and $n_{\max}$, we estimated the exponent α and its goodness-of-fit by the coefficient of determination (R^2). We then selected as our estimate of $n_{\min}$, $n_{\max}$, and α that gave the maximum R^2 (> 0.99) over all possibilities. For example, the $n_{\min} = 2$, $n_{\max} = 2253$, and $\alpha = 1.17$ in the eigenspectrum of the trained Billeh model response to 2,800 nature images.

Quantifying the goodness-of-fit of power-law exponent of eigenspectra to theoretical prediction. The power-law exponent α describes the decaying speed of eigenspectrum; the kernel theory predicted that the power-law exponent α characterizes the smoothness of the neural responses [Stringer et al., 2019a]. If the sensory stimuli presented can be characterized by d parameters, and if the mapping from these parameters to (noise-free) neural population responses is differentiable, then the population eigenspectrum must decay asymptotically faster than a power law of exponent $\alpha = 1 + 2/d$ (low dimension coding). Conversely, if the eigenspectrum decays slower than this (high dimension coding), a smooth neural code (similar stimuli give rise to similar responses) is impossible, allowing small changes in input to dominate population activity. Therefore, $1 + 2/d$ indicates the best coding scheme, at the balance between high-dimension (uncorrelated, efficient) and low-dimension (correlated, smooth) representation. We thus used the coefficient of determination (R^2) to quantify the power-law exponent (α) of eigenspectra fitting to the theoretical prediction $1 + 2/d$ where d is the dimension of input images. The larger value of R^2, the better fitting. We denoted the α fitted with input image type i as α_i and the dimension of input image type i as d_i. If $\overline{\alpha}$ was the mean of all the observed $\{\alpha_i\}$, then the variability of $\{\alpha_i\}$ can be measured with two sums of squares formulas: the residual sum of squares:

$$SS_{\text{res}} = \sum_i \left[\alpha_i - \left(1 + \frac{2}{d_i}\right)\right]^2, \tag{11}$$

and the total sum of squares (proportional to the variance of the data)

$$SS_{\text{tot}} = \sum_i [\alpha_i - \overline{\alpha}]^2. \tag{12}$$

The definition of R^2 is

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{13}$$

4.12 d' for neural responses to visual stimuli

To estimate how much information the neural activity conveyed about the stimulus identity, following [Rumyantsev et al., 2020], we used the metric d', which characterizes how readily the distributions of the neural responses to the two different sensory stimuli can be distinguished [Bishop, 2007]. The quantity $(d')^2$ is the discrete analog of Fisher information [Averbeck and Lee, 2006]. To be consistent with the experimental study [Rumyantsev et al., 2020], we calculated the neural response as the spike counts in each bin of 200 ms and evaluated two different approaches to compute d' values for the discrimination of

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the two different visual stimuli (gratings in the fine orientation discrimination task; the difference between two gratings is 2 °.

In the first approach, i.e., instantaneous decoding, we chose for analysis a specific time bin relative to the onset of visual stimulation. To examine the time-dependence of d', we used the instantaneous decoding approach and varied the selected time bin from $t = 0$ s to $t = 2$ s relative to the start of the trial. The number of dimensions of the neural ensemble activity evoked in response to the visual stimulus was N_0, the number of recorded neurons (N_0). Said differently, the set of estimated spike traces provided an N_0-dimensional population vector response to each stimulus presentation.

In the second approach, termed cumulative decoding, we concatenated the responses of each neuron over time, from the start of the trial up to a chosen time, t. In this case, the dimensionality of the population activity vector was $N_0 \times N_t$, where N_t is the number of time bins spanning the interval $[0\ \mathrm{s}, t]$.

In each of the decoding approaches, we arranged the traces of estimated spike counts into three-dimensional data structures (number of neurons × number of time bins × number of trials), for each of the two visual stimuli. To be consistent with [Rumyantsev et al., 2020], we used the neural response in 2 s, the bin size of 200 ms, and 4500 trials for each stimulus.

To determine d' accurately despite having about fewer trials than neuron number in the Billeh model, we reduced dimensional by using partial least squares (PLS) analysis [Geladi and Kowalski, 1986] to identify and retain only five population vector dimensions in which the stimuli were highly distinguishable as in [Rumyantsev et al., 2020]. In this five-dimensional representation, the neural dynamics evoked by the two stimuli become distinguishable over the first 200 ms of stimulus presentation. In the reduced space with $N_{\mathrm{R}=5}$ dimensions, we calculated the $(d')^2$ value of the optimal linear discrimination strategy as:

$$\left(d'_{\mathrm{opt}}\right)^2 = \Delta\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\Delta\boldsymbol{\mu} = \Delta\boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}_{\mathrm{opt}} \tag{14}$$

where $\Sigma = \frac{1}{2}\left(\Sigma_{\mathrm{A}} + \Sigma_{\mathrm{B}}\right)$ the noise covariance matrix averaged across two stimulation conditions, $\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathrm{A}} - \boldsymbol{\mu}_{\mathrm{B}}$ is the vector difference between the mean ensemble neural responses to the two stimuli and $\mathbf{w}_{\mathrm{opt}} = \Sigma^{-1}\Delta\boldsymbol{\mu}$, which is normal to the optimal linear discrimination hyperplane in the response space [Averbeck and Lee, 2006].

We also calculated $(d'_{\mathrm{shuffled}})^2$, the optimal linear discrimination performance using trial-shuffled datasets, which we created by shuffling the responses of each cell across stimulation trials of the same type. Owing to this shuffling procedure, the off-diagonal elements of Σ_{A} and Σ_{B} became near zero.

Eigenvalues of the noise covariance matrix To examine how the statistical structure of neural noise affects the ability to discriminate neural responses to the two different visual stimuli, we expressed $(d')^2$ in terms of the eigenvalues λ_α and eigenvectors $\boldsymbol{e}_\alpha$ of the noise covariance matrix Σ:

$$(d')^2 = \Delta\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\Delta\boldsymbol{\mu} = \sum_\alpha \left(\frac{\left|\Delta\boldsymbol{\mu}\cdot\boldsymbol{e}_\alpha\right|^2}{\lambda_\alpha}\right) \tag{15}$$

which can be viewed as a sum of signal-to-noise ratios, one for each eigenvector. Clearly, the eigenvectors well aligned with $\Delta\boldsymbol{\mu}$ are the most important for discriminating between the two distributions of neural responses.

4.13 Branching ratio

Based on the work of [Wilting and Priesemann, 2018b], we examined the branching ratio of the networks sampled in our model. Summarily, the branching ratio is the ratio of the number of neurons spiking at

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

time $t+1$ to the number of spiking neurons at time t. Critical regimes, by their nature, are balanced and avoid runaway gain (positive or negative) and have a branching ratio of 1.0. The methods introduced by [Wilting and Priesemann, 2018b] are robust to severe subsampling and thus provide an effective alternative approach to assessing critical dynamics in our recordings.

In a network with A active neurons at time t, if the branching ratio is a fixed value then $\langle A_{t+1} \mid A_t \rangle = mA_t + h$ where $<|>$ denotes the conditional expectation, m is the branching ratio and h is a mean rate of an external drive/stimulus. Considering subsampling, a_t is proportional to A_t on average $< a_t \,|A_t\rangle = \eta A_t + \xi$, where η and ξ are constants. This subsampling leads to a bias: $m\left(\eta^2 \operatorname{Var}[At]/\operatorname{Var}[at] - 1\right)$. Instead of using time t and $t+1$, this method focuses on times t and $t+k$ with different time lags $k = 1, \ldots, k_{\text{maximum}}$. With this, the branching ratio m_k is $< a_{t+k} \mid a_t >= m_k = \eta^2 \operatorname{Var}[A_t] / \operatorname{Var}[a_t]\, m^k = bm^k$, where b is a constant. To compute m_k with different k, we obtained an exponential curve and extracted m from this curve. For details see [Wilting and Priesemann, 2018a], we examined a range of k and searched different results using k_{maximum} from tens to one thousand. We then chose k_{maximum} for each animal by checking the baseline period. We selected k_{maximum} as the k that returned m closest to 1.0 during baseline. For the seven animals used in this study, k_{maximum} values were selected for each animal during the recording baseline and maintained throughout the remainder of the experiment.

We input a series of nature images as in the visual change detection task of nature image to stimulate the neural network for 15 s.

Acknowledgements

We would like to thank Eben Kadile, Oleg Kolner, Christoph Stöckl, and Yujie Wu for helpful comments on an earlier version of the manuscript. Also, we would like to thank Sandra Diaz for advice and help regarding large-scale computations. Most of computations were carried out on the Human Brain Project PCP Pilot Systems at the Jülich Supercomputing Centre, which received co-funding from the European Union (Grant Agreement number 604102). We also would like to thank the Beijing Academy of Artificial Intelligence for providing computational resources. This research was partially supported by the Human Brain Project (Grant Agreement number 785907) of the European Union and a grant from Intel.

References

[Allen Institute, 2018] Allen Institute (2018). © 2018 Allen Institute for Brain Science. Allen Cell Types Database, cell feature search. Available from: celltypes.brain-map.org/data.

[Arieli et al., 1996] Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868–1871.

[Averbeck and Lee, 2006] Averbeck, B. B. and Lee, D. (2006). Effects of noise correlations on information encoding and decoding. *Journal of neurophysiology*, 95(6):3633–3644.

[Bellec et al., 2018] Bellec, G., Salaj, D., Subramoney, A., Legenstein, R., and Maass, W. J. a. p. a. (2018). Long short-term memory and learning-to-learn in networks of spiking neurons.

[Bellec et al., 2020] Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., and Maass, W. (2020). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 11(1):3625.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[Billeh et al., 2020] Billeh, Y. N., Cai, B., Gratiy, S. L., Dai, K., Iyer, R., Gouwens, N. W., Abbasi-Asl, R., Jia, X., Siegle, J. H., Olsen, S. R., et al. (2020). Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron*, 106(3):388–403.

[Bishop, 2007] Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition.

[Chen and Gong, 2019] Chen, G. and Gong, P. (2019). Computing by modulating spontaneous cortical activity patterns as a mechanism of active visual processing. *Nature Communications*, 10(1):4915.

[Cover, 1999] Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.

[Davies et al., 2021] Davies, M., Wild, A., Orchard, G., Sandamirskaya, Y., Guerra, G. A. F., Joshi, P., Plank, P., and Risbud, S. R. (2021). Advancing neuromorphic computing with loihi: A survey of results and outlook. *Proceedings of the IEEE*, 109(5):911–934.

[Engelhard et al., 2019] Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H. J., Ornelas, S., Koay, S. A., Thiberge, S. Y., Daw, N. D., Tank, D. W., et al. (2019). Specialized coding of sensory, motor and cognitive variables in vta dopamine neurons. *Nature*, 570(7762):509–513.

[Garrett et al., 2020] Garrett, M., Manavi, S., Roll, K., Ollerenshaw, D. R., Groblewski, P. A., Ponvert, N. D., Kiggins, J. T., Casal, L., Mace, K., and Williford, A. (2020). Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. *eLife*, 9:e50340.

[Geladi and Kowalski, 1986] Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185:1–17.

[Gerstner et al., 2014] Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.

[Glickfeld et al., 2013] Glickfeld, L. L., Histed, M. H., and Maunsell, J. H. (2013). Mouse primary visual cortex is used to detect both orientation and contrast changes. *Journal of Neuroscience*, 33(50):19416–19422.

[Haeusler et al., 2009] Haeusler, S., Schuch, K., and Maass, W. (2009). Motif distribution, dynamical properties, and computational performance of two data-based cortical microcircuit templates. *Journal of Physiology-Paris*, 103(1-2):73–87.

[Harris and Shepherd, 2015] Harris, K. D. and Shepherd, G. M. G. (2015). The neocortical circuit: themes and variations. *Nature neuroscience*, 18(2):170.

[Huber, 1992] Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer.

[Lee et al., 2012] Lee, S.-H., Kwan, A. C., Zhang, S., Phoumthipphavong, V., Flannery, J. G., Masmanidis, S. C., Taniguchi, H., Huang, Z. J., Zhang, F., Boyden, E. S., et al. (2012). Activation of specific interneurons improves v1 feature selectivity and visual perception. *Nature*, 488(7411):379–383.

[Markram et al., 2015] Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2):456–492.

[Montijn et al., 2014] Montijn, J. S., Vinck, M., and Pennartz, C. (2014). Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation. *Frontiers in computational neuroscience*, 8:58.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[Mora and Bialek, 2011] Mora, T. and Bialek, W. (2011). Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302.

[Morcos and Harvey, 2016] Morcos, A. S. and Harvey, C. D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. *Nature neuroscience*, 19(12):1672–1681.

[Mountcastle, 1998] Mountcastle, V. B. (1998). *Perceptual neuroscience: The cerebral cortex*. Harvard University Press.

[Nikolić et al., 2009] Nikolić, D., Häusler, S., Singer, W., and Maass, W. (2009). Distributed fading memory for stimulus properties in the primary visual cortex. *PLoS biology*, 7(12):e1000260.

[Pérez-Ortega et al., 2021] Pérez-Ortega, J., Alejandre-García, T., and Yuste, R. (2021). Long-term stability of cortical ensembles. *Elife*, 10:e64449.

[Plank et al., 2021] Plank, P., Rao, A., Wild, A., and Maass, W. (2021). A long short-term memory for ai applications in spike-based neuromorphic hardware. *arXiv preprint arXiv:2107.03992*.

[Ruderman and Bialek, 1994] Ruderman, D. L. and Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical review letters*, 73(6):814.

[Rumyantsev et al., 2020] Rumyantsev, O. I., Lecoq, J. A., Hernandez, O., Zhang, Y., Savall, J., Chrapkiewicz, R., Li, J., Zeng, H., Ganguli, S., and Schnitzer, M. J. (2020). Fundamental bounds on the fidelity of sensory cortical coding. *Nature*, 580(7801):100–105.

[Scherr and Maass, 2021] Scherr, F. and Maass, W. (2021). Analysis of the computational strategy of a detailed laminar cortical microcircuit model for solving the image-change-detection task. *bioRxiv*.

[Siegle et al., 2021] Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., Groblewski, P. A., Ahmed, R., Arkhipov, A., Bernard, A., Billeh, Y. N., Brown, D., Buice, M. A., Cain, N., Caldejon, S., Casal, L., Cho, A., Chvilicek, M., Cox, T. C., Dai, K., Denman, D. J., de Vries, S. E. J., Dietzman, R., Esposito, L., Farrell, C., Feng, D., Galbraith, J., Garrett, M., Gelfand, E. C., Hancock, N., Harris, J. A., Howard, R., Hu, B., Hytnen, R., Iyer, R., Jessett, E., Johnson, K., Kato, I., Kiggins, J., Lambert, S., Lecoq, J., Ledochowitsch, P., Lee, J. H., Leon, A., Li, Y., Liang, E., Long, F., Mace, K., Melchior, J., Millman, D., Mollenkopf, T., Nayan, C., Ng, L., Ngo, K., Nguyen, T., Nicovich, P. R., North, K., Ocker, G. K., Ollerenshaw, D., Oliver, M., Pachitariu, M., Perkins, J., Reding, M., Reid, D., Robertson, M., Ronellenfitch, K., Seid, S., Slaughterbeck, C., Stoecklin, M., Sullivan, D., Sutton, B., Swapp, J., Thompson, C., Turner, K., Wakeman, W., Whitesell, J. D., Williams, D., Williford, A., Young, R., Zeng, H., Naylor, S., Phillips, J. W., Reid, R. C., Mihalas, S., Olsen, S. R., and Koch, C. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*.

[Stringer et al., 2021] Stringer, C., Michaelos, M., Tsyboulski, D., Lindo, S. E., and Pachitariu, M. (2021). High-precision coding in visual cortex. *Cell*, 184(10):2767–2778.

[Stringer et al., 2019a] Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019a). High-dimensional geometry of population responses in visual cortex. *Nature*, 571:361–365.

[Stringer et al., 2019b] Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., and Harris, K. D. (2019b). Spontaneous behaviors drive multidimensional, brainwide activity. *science*, 364(6437):eaav7893.

[Teeter et al., 2018] Teeter, C., Iyer, R., Menon, V., Gouwens, N., Feng, D., Berg, J., Szafer, A., Cain, N., Zeng, H., and Hawrylycz, M. (2018). Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature communications*, 9(1):1–15.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[Thomson and Lamy, 2007] Thomson, A. M. and Lamy, C. (2007). Functional maps of neocortical local circuitry. *Frontiers in neuroscience*, 1:2.

[Wilting and Priesemann, 2018a] Wilting, J. and Priesemann, V. (2018a). Inferring collective dynamical states from widely unobserved systems. *Nature communications*, 9(1):2325.

[Wilting and Priesemann, 2018b] Wilting, J. and Priesemann, V. (2018b). On the ground state of spiking network activity in mammalian cortex. *arXiv preprint arXiv:1804.07864*.

[Wilting and Priesemann, 2019] Wilting, J. and Priesemann, V. (2019). 25 years of criticality in neuroscience—established results, open controversies, novel concepts. *Current opinion in neurobiology*, 58:105–111.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Supplementary materials

Table S1: Performances on 5 tasks are not sensitive to the noise level. The light cyan highlights the noise level used in this study unless otherwise stated. VCDN is tested on novel images. FOD, fine orientation discrimination; IC, image classification; VCDN, visual change detection of nature images; VCDG, visual change detection of gratings; EA, evidence accumulation.

	FOD	IC	VCDN	VCDG	EA
	acc.	acc.	acc.	acc.	acc.
$q = 1,\ s = 0$	93.67%	89.14%	83.61%	92.37%	93.75%
$q = 1,\ s = 2$	93.09%	88.49%	83.42%	89.22%	92.92%
$q = 1,\ s = 3$	93.47%	88.97%	82.52%	88.73%	90.17%
$q = 1,\ s = 3.5$	93.73%	88.49%	81.11%	88.58%	91.00%
$q = 1,\ s = 4$	93.06%	88.81%	82.30%	87.86%	91.50%
$q = 2,\ s = 2$	93.15%	88.92%	83.25%	89.25%	90.92%
$q = 2,\ s = 3.5$	93.93%	89.19%	82.20%	87.51%	90.58%
$q = 3,\ s = 3$	94.10%	88.81%	81.87%	89.52%	92.25%
$q = 3,\ s = 4$	93.53%	89.02%	83.06%	88.14%	89.33%
$q = 4,\ s = 3$	93.91%	88.91%	82.53%	88.16%	91.33%
$q = 4,\ s = 4$	93.47%	88.16%	82.16%	88.17%	89.08%
$q = 5,\ s = 4$	93.32%	88.49%	81.41%	89.12%	90.58%
$q = 10,\ s = 10$	94.03%	88.54%	82.97%	89.56%	84.33%
$q = 20,\ s = 20$	91.01%	86.63%	77.41%	85.95%	77.75%
$q = 200,\ s = 200$	54.58%	39.56%	53.09%	53.96%	52.67%

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

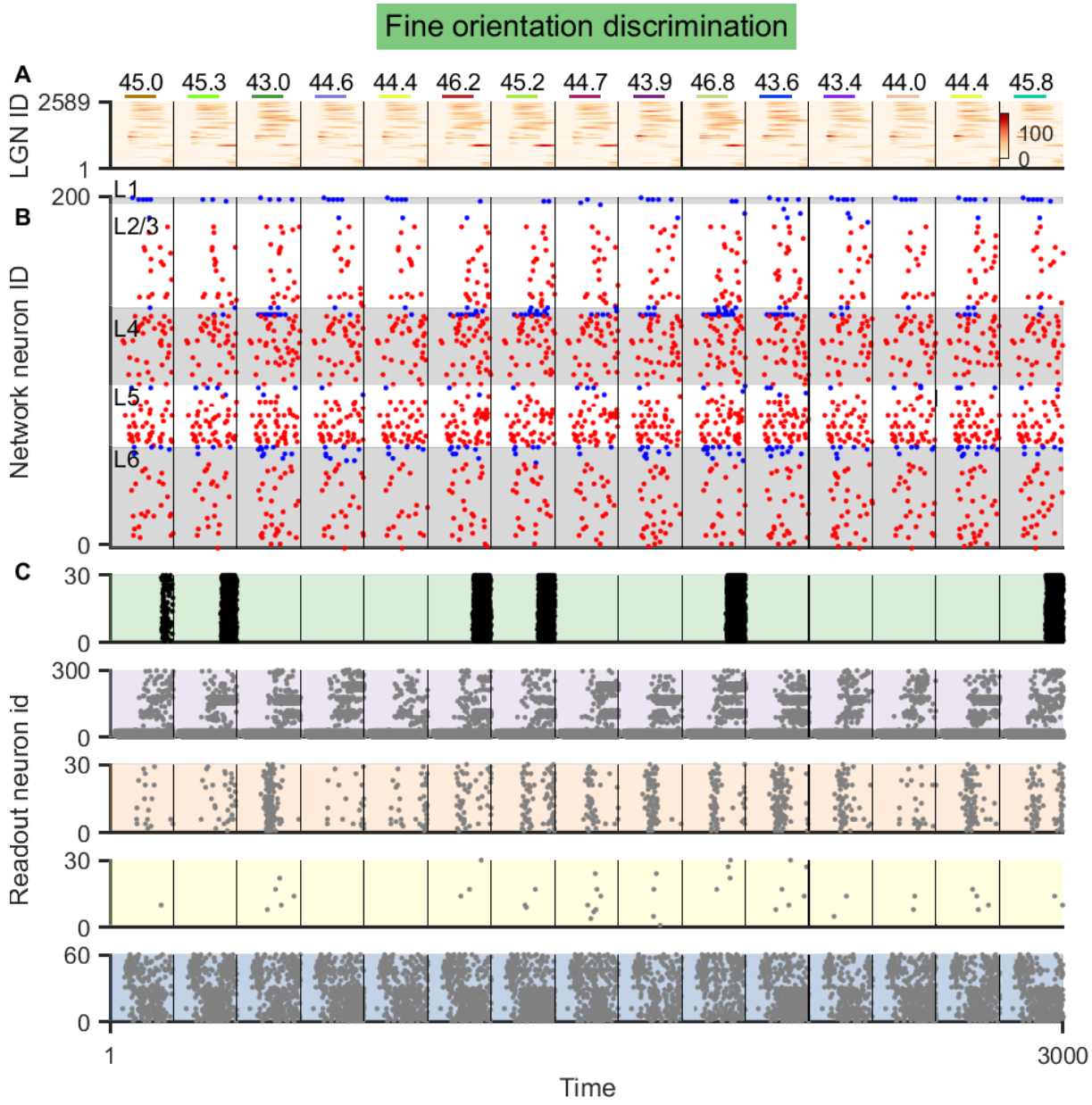


Figure S1: After learning, the model reliably distinguishes the orientation of gratings in the fine orientation discrimination task. **(A)** Colorful lines represent the timing of input images. Numbers on them represent the orientations of input gratings. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. Note that the spike and membrane potential of the model was reset to 0 after one classification was done (separated by the think black line). **(C)** Spike raster of readout neurons. Color codes of panels are the same as in Fig. 3A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

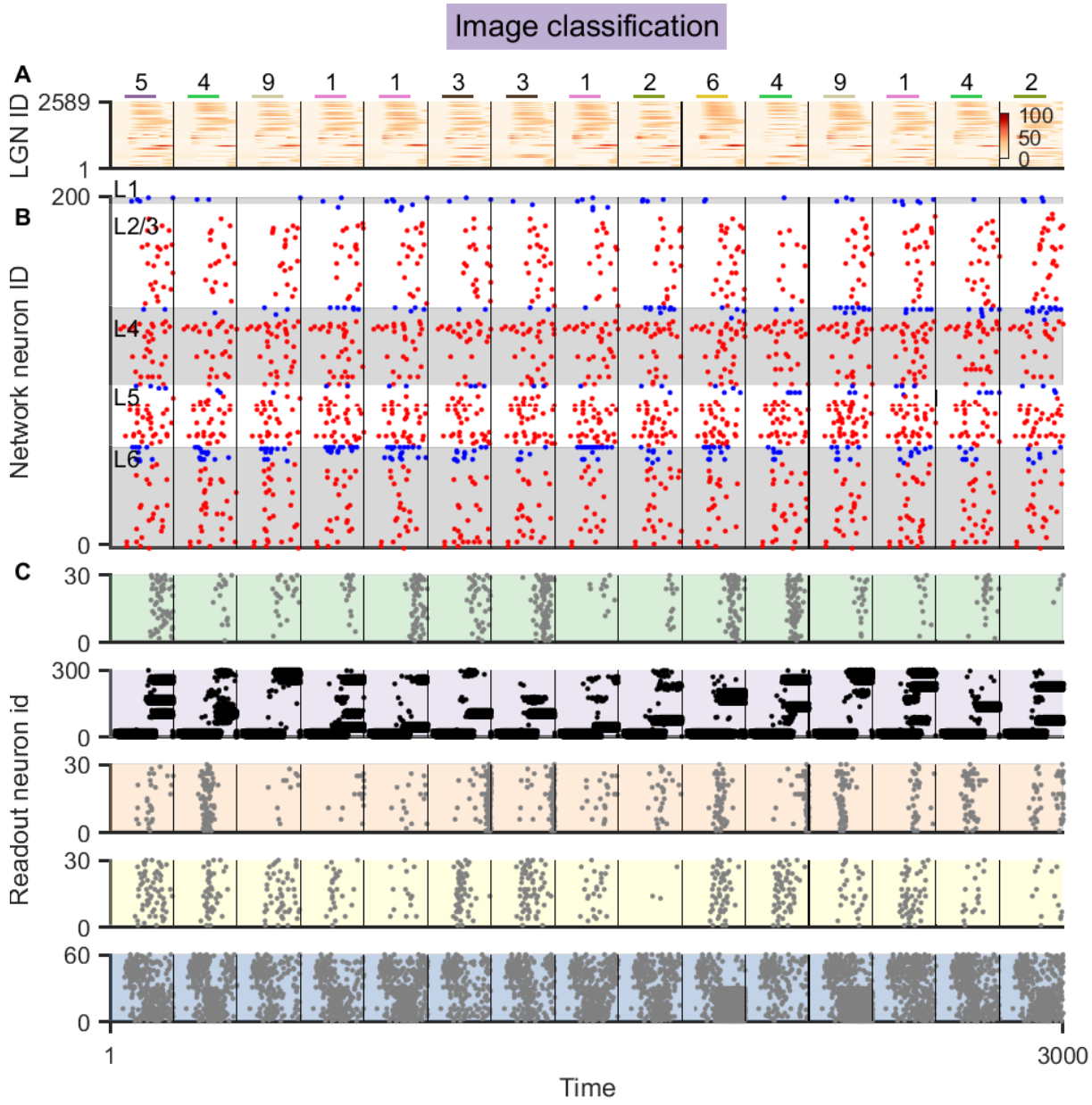


Figure S2: After learning, the model reliably classify the MNIST images. **(A)** Colorful lines represent the timing of input images. Numbers on them represent the digits in the input images. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. Note that the spike and membrane potential of the model was reset to 0 after one classification was done (separated by the think black line). **(C)** Spike raster of readout neurons. Color codes of panels are the same as in Fig. 3A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

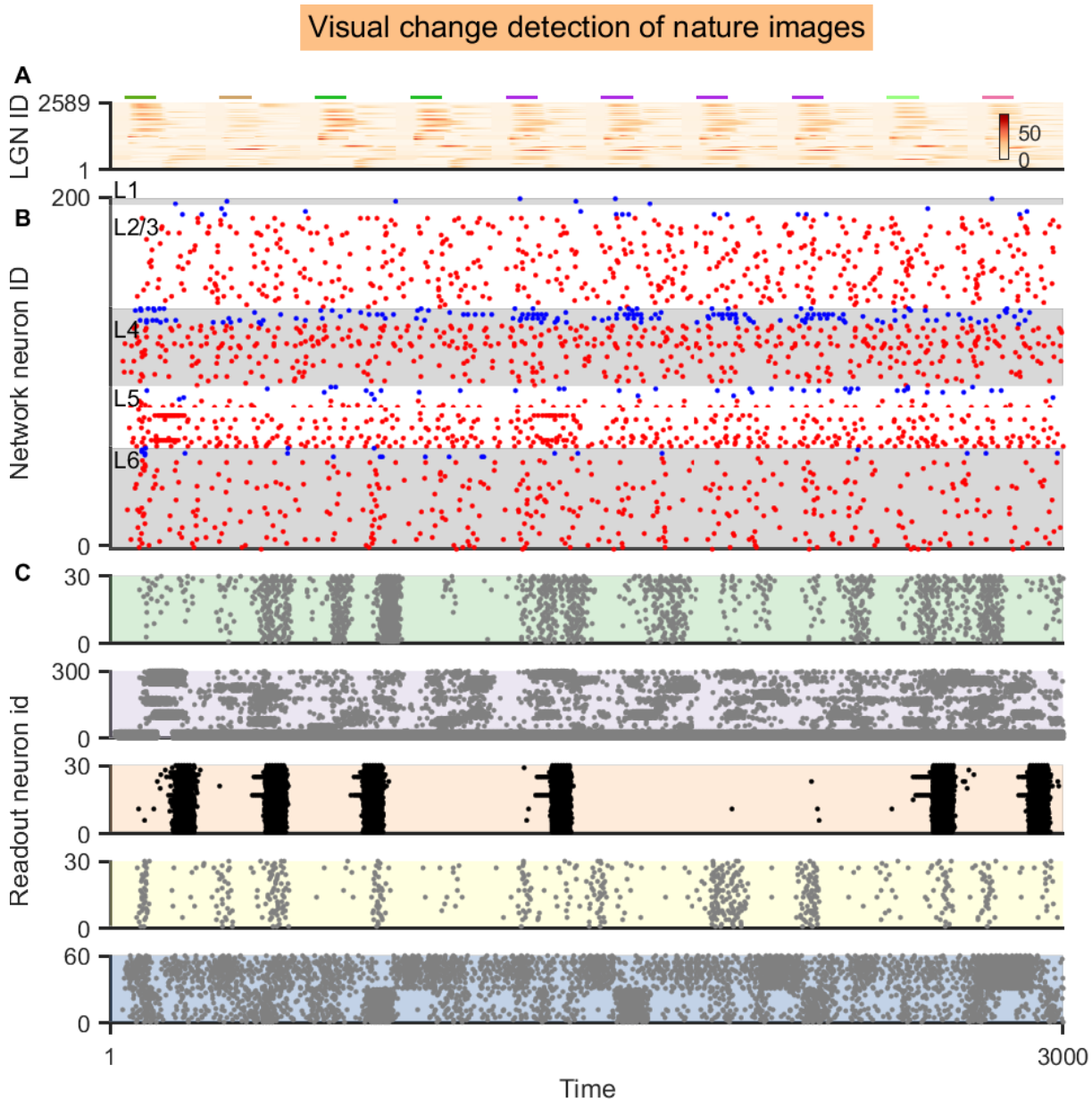


Figure S3: After learning, the model reliably reported the image identity change in the visual change detection task. **(A)** Colorful lines represent the timing of input images and the colors code the image identity. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. The slow noise was resampled every 600 ms. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. **(C)** Spike raster of readout neurons. Color codes of panels are the same as in Fig. 3A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

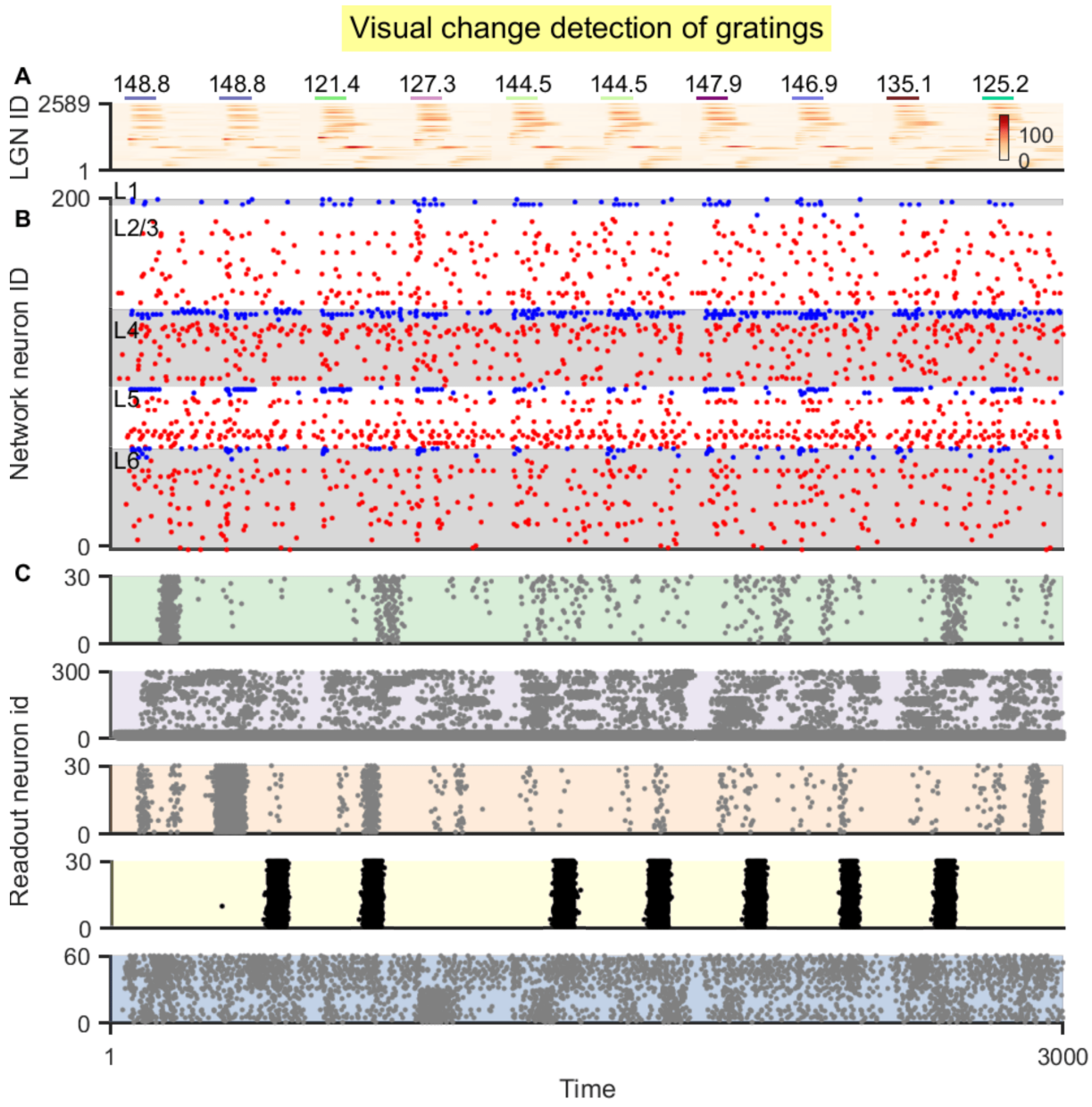


Figure S4: After learning, the model reliably reported the image identity change in the visual change detection task. **(A)** Colorful lines represent the timing of input images and the colors code the image identity. Numbers on them represent the orientations of input gratings. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. The slow noise was resampled every 600 ms. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. **(C)** Spike raster of readout neurons. Color codes of panels are the same as in Fig. 3A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

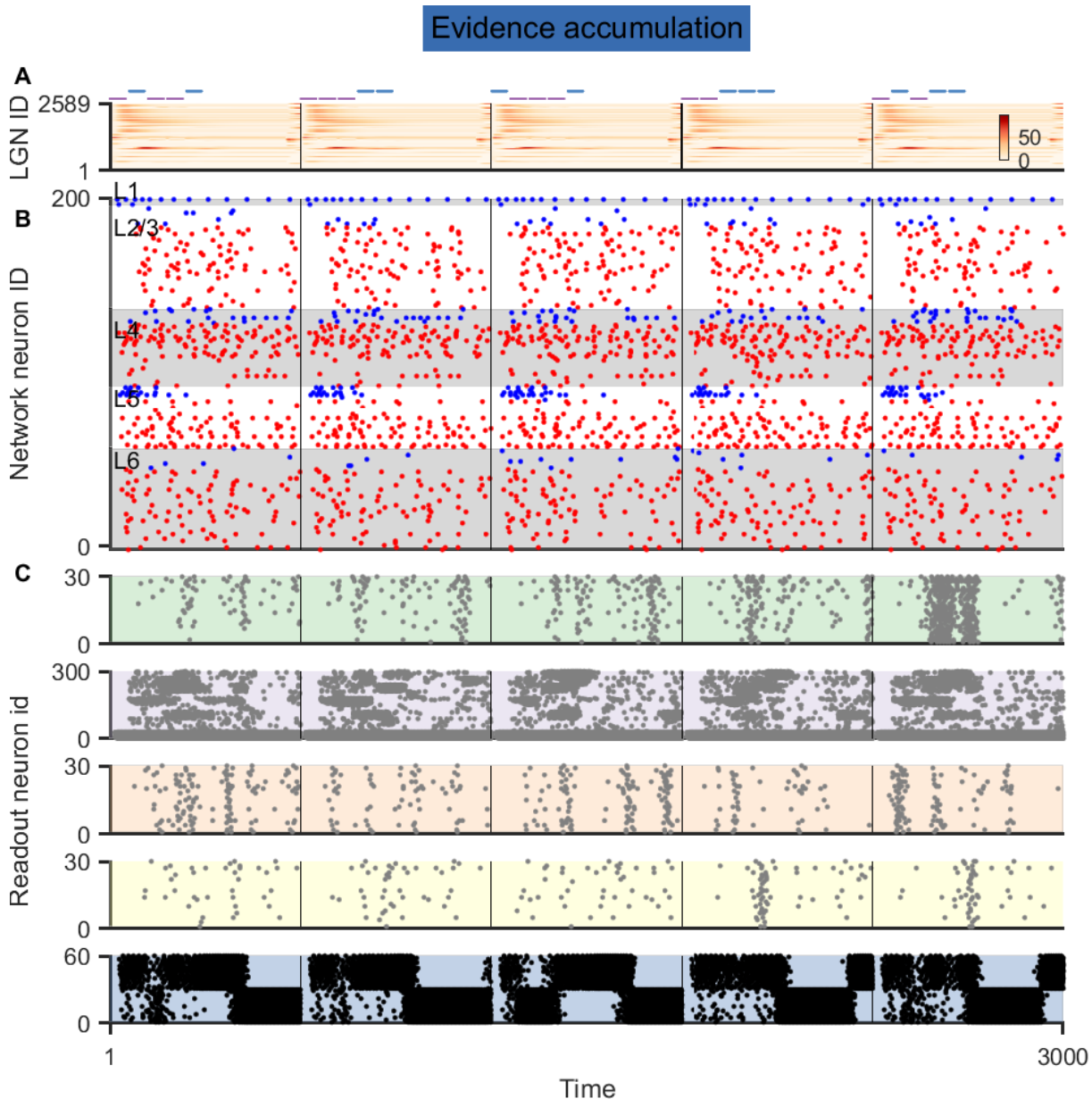


Figure S5: After learning, the spiking activity in the evidence accumulation task. **(A)** Colorful lines represent the timing of input left/right cues. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. Note that the spike and membrane potential of the model was reset to 0 after one classification was done (separated by the think black line). **(C)** Spike raster of readout neurons. Color codes of panels are the same as in Fig. 3A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

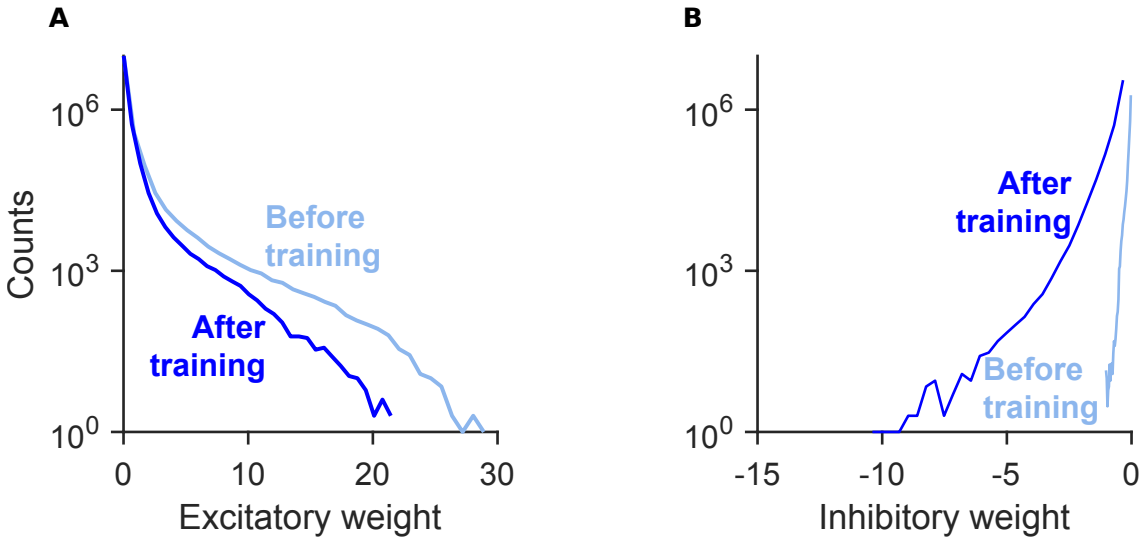


Figure S6: Distribution of recurrent weights before and after learning the 5 tasks. **(A)** Distribution of excitatory weights. **(A)** Distribution of inhibitory weights.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

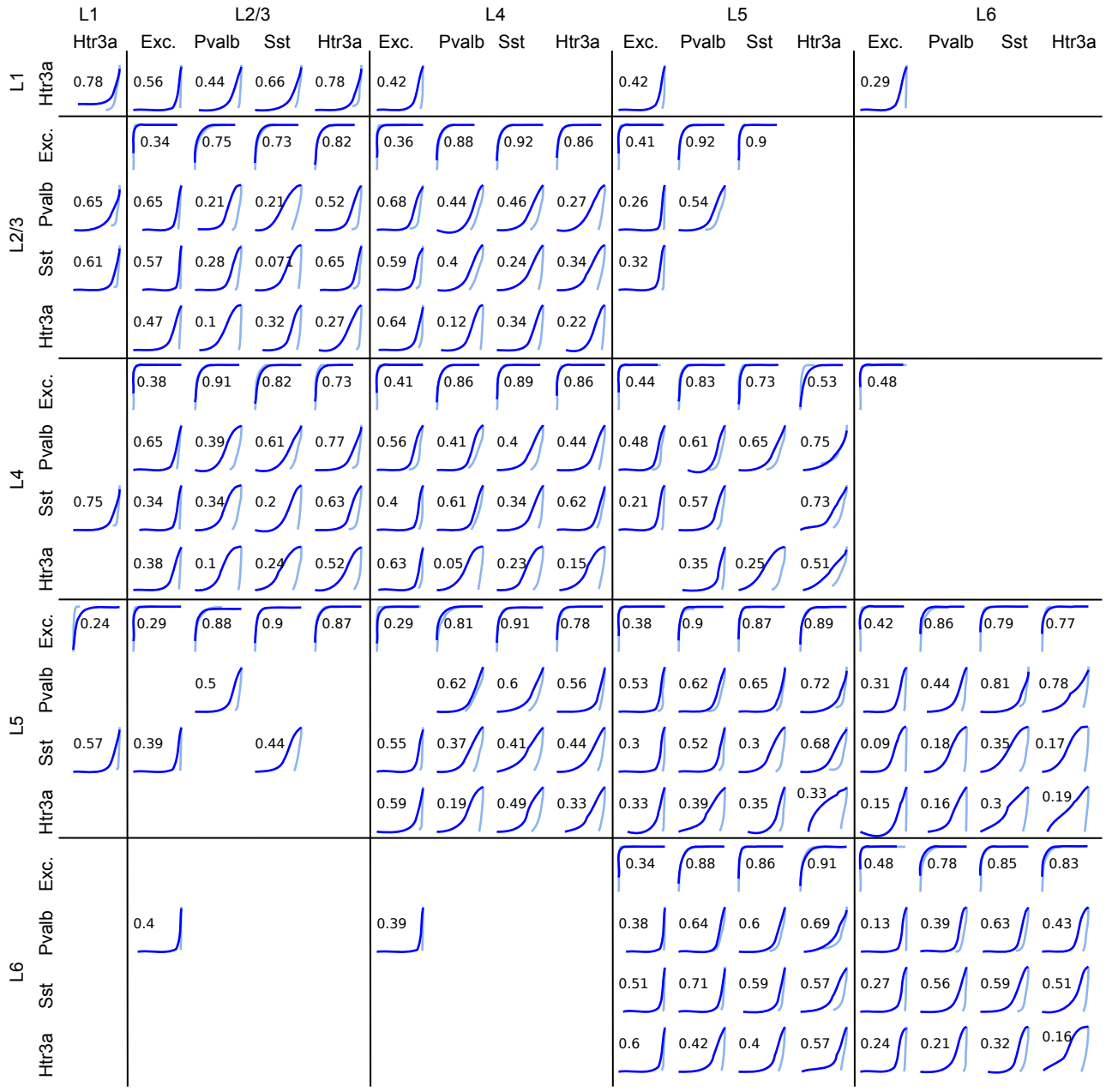


Figure S7: Distribution of recurrent weights between each population before (light blue) and after learning (dark blue) the task. Each row represents a pre-synaptic neuron population, and each column represents a post-synaptic neuron population. The histogram represents the distribution of synaptic weights of all synaptic connections that share the same pre-synaptic and post-synaptic neuron population. Vertical axis in each panel is log-scale. Horizontal axis is linear scale and horizontal range is from the smallest value to the largest value of each population. The number is $1 - D$ where D is from the Kolmogorov–Smirnov test, quantifying the similarity between distributions [Billeh et al., 2020]. Exc., excitatory neurons.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.07.471653; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

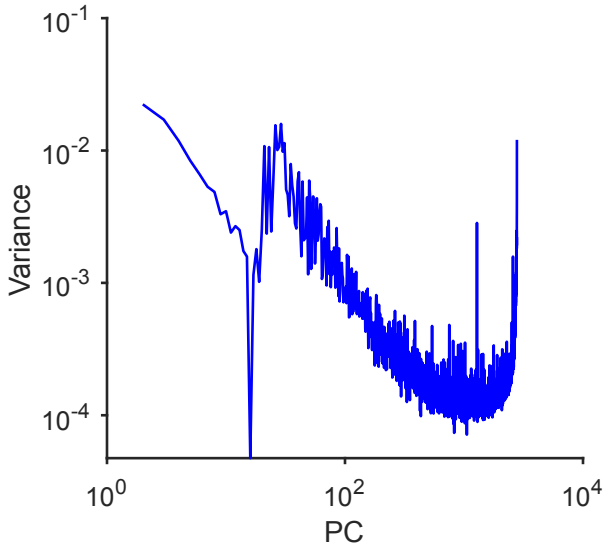


Figure S8: Strong noise breaks the power-law eigenspectrum. When $s = 10$, $q = 10$, the eigenspectrum of the trained Billeh model response to 2,800 nature images is not power law.