

Temporal prediction captures key differences between spiking excitatory and inhibitory V1 neurons

Luke Taylor^{1*}, Friedemann Zenke^{2, 3}, Andrew J. King¹,
Nicol S. Harper^{1*}

¹Department of Physiology, Anatomy, and Genetics, University of Oxford, Sherrington Rd, Oxford, OX1 3PT, UK.

²Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, Basel, 4058, Switzerland.

³Faculty of Science, University of Basel, Peterspl. 1, Basel, 4001, Switzerland.

*Corresponding author(s). E-mail(s): luke.taylor@hertford.ox.ac.uk;
nicol.harper@dpag.ox.ac.uk;

Contributing authors: friedemann.zenke@fmi.ch;
andrew.king@dpag.ox.ac.uk;

Abstract

Neurons in primary visual cortex (V1) respond to natural scenes with a sparse and irregular spike code that is carefully balanced by an interplay between excitatory and inhibitory neurons. These neuron classes differ in their spike statistics, tuning preferences, connectivity statistics and temporal dynamics. To date, no single computational principle has been able to account for these properties. We developed a recurrently connected spiking network of excitatory and inhibitory units trained for efficient temporal prediction of natural movie clips. We found that the model exhibited simple and complex cell-like tuning, V1-like spike statistics, and, notably, also captured key differences between excitatory and inhibitory V1 neurons. This suggests that these properties collectively serve to facilitate efficient prediction of the sensory future.

Introduction

Our visual system allows us to navigate the world and interpret the most unlikely of scenes, like a toy brachiosaurus sitting atop a brick wall. The primary visual cortex (V1) plays a central role in processing diverse and complex stimuli, with neurons tuned to the orientation and motion of moving edges and shapes [1–3]. These features are communicated by a sparse and irregular spike code [4] that is carefully balanced by interactions between the cortex’s excitatory and inhibitory neural subpopulations [5–8]. There is growing evidence for the idea that visual perception arises from the brain engaging in predictive processing [9]. In particular, recent research suggests that neurons in sensory brain areas, such as V1, may have evolved to represent features in natural stimuli that are predictive of the immediate sensory future [10, 11]. Such representations are evolutionary advantageous, since they allow inferences about the underlying causes of sensory signals, which may help to guide future action, and eliminate unnecessary sensory information [12].

Recent computational studies have demonstrated that feedforward artificial neural networks (ANNs) trained for temporal prediction using natural movies develop units with receptive field properties that resemble those of V1 neurons [10, 11]. Specifically, these units replicate the spatial tuning of V1 neurons and notably also their temporal dynamics - a property not readily captured by other models [10]. However, while feedforward ANNs loosely resemble V1’s biological circuit [13], they omit several key features, including the recurrent connectivity of the cortex; they do not conform to Dale’s law [14], where neurons are either excitatory or inhibitory; and they omit the brain’s key physiological unit of computation: the spike. It therefore remains unclear to what extent - if at all - a biologically detailed model of V1 optimized for temporal prediction would capture the physiological properties of its neurons.

To bridge this shortcoming, we developed a spiking model of recurrently-connected excitatory and inhibitory units trained on natural movies under metabolic-like constraints to predict the sensory future from recent spike activity. We found the units in our model to exhibit simple and complex cell-like tuning. Extending beyond prior results, the spike statistics of our model resembled those of V1 neural responses to natural stimuli. Notably, the excitatory and inhibitory units in our model captured the differences in spike statistics, tuning preferences, connectivity statistics and temporal dynamics between excitatory and inhibitory V1 neurons.

Results

The spiking V1 model

We implemented our model as a population of recurrently-connected spiking units and adopted the leaky integrate-and-fire (LIF) model to mimic real neural spiking dynamics [15] (Figure 1A). Similar to the ratio of excitatory-to-inhibitory neurons in V1 [5–8], we set 85% of these units as excitatory and the remaining 15% as inhibitory. The model’s input consists of a sequence of 20×20 px patches obtained from retina-filtered natural movies recorded at 120 frames-per-second. At each time step, the model linearly translates a 126ms (15 frames) span of stimulus history into a distinct input

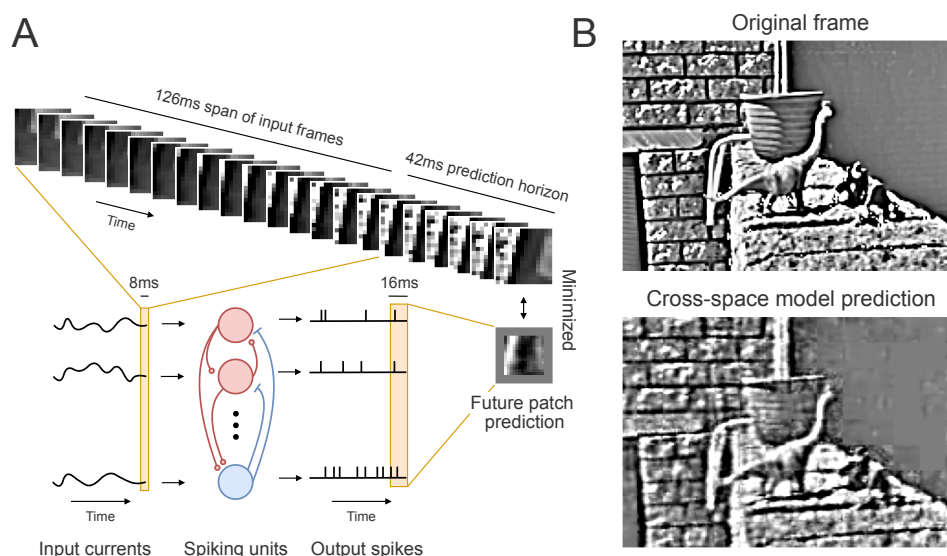


Fig. 1: The spiking temporal prediction model with recurrently connected excitatory and inhibitory units. **A.** Schematic of our model, which at every time step linearly maps a 126ms span of natural movie frame patches of the past into a distinct input current value to every spiking unit. A 16ms span of population output spikes is linearly mapped to create a prediction of the spatial movie patch 42ms into the future. Red units are excitatory and blue units are inhibitory. **B.** An example movie frame of a toy dinosaur sitting on a brick wall, which was not included in the training data (top), with the model's reconstruction using multiple patch predictions across different spatial locations (bottom).

current to each unit. The 126ms duration covers the potential integration span and spread of latencies of the incoming thalamocortical input [16]. The model was trained to linearly predict the spatial movie patch 42ms (5 frames) into the future using a 16ms (2 frames) span of past population spike activity (Figure 1B). We chose the 42ms prediction target to be of similar duration to the minimum latency of visual input to V1 [17–20] and of similar duration to the response latency of the fastest visuomotor actions [21].

To emulate the metabolic energy constraints of the brain, we also trained our model to reduce a metabolic-like cost whilst performing the frame prediction (see Methods). This approximated the activity-induced cost at the model weights, similarly to the energy expenditure of neural synapses. We made an assumption about the brain in our metabolic-like cost function, namely that inhibitory V1 neurons consume less energy than their excitatory counterparts per spike. We found this assumption to be important for obtaining all of the V1-like tuning differences between the excitatory and inhibitory model units. Lastly, training was performed using a variant of the backpropagation algorithm [22], known as surrogate gradient learning [23], which permits gradient descent in non-differentiable spiking networks. This optimization process

determined the model's connectivity and the membrane time constant of each spiking unit. We remain agnostic about the biological plausibility of this training method, and how cortex might optimize for the objective of temporal prediction.

Emergence of a V1-like spike code

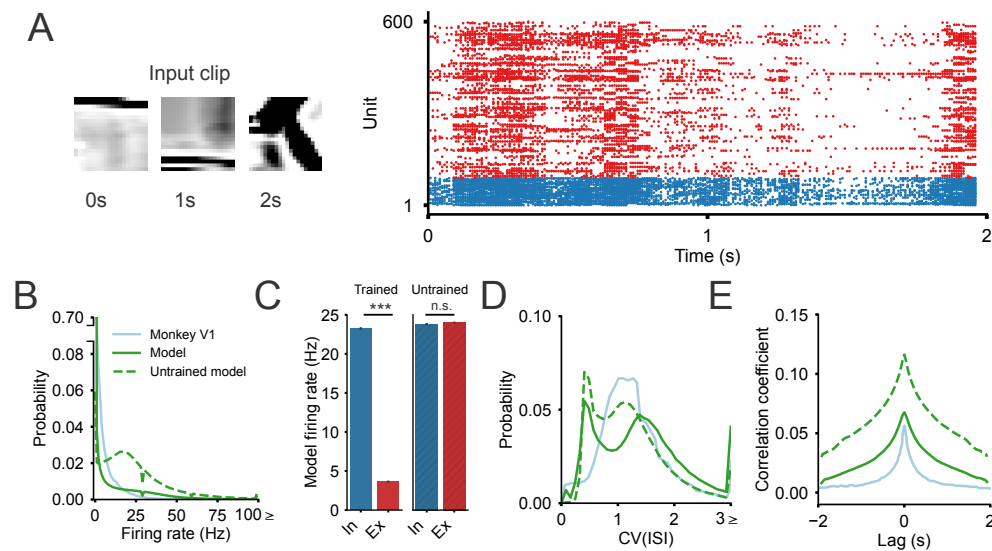


Fig. 2: Spiking statistics of the model resemble those of V1. **A.** Input movie clip (left) and resulting model spike raster (right). Red and blue dots correspond to the excitatory and inhibitory units, respectively. **B.** Firing rate distribution of the trained and untrained model units and monkey V1 neurons. **C.** Difference in the model's firing rates between the inhibitory (In) and excitatory (Ex) units in the trained (left) and untrained model (right). Bars plot the mean and standard error over units and input clips, with statistical significance assessed using the Mann-Whitney U test ($***p < 0.001$). **D.** Coefficient of variation of the interspike intervals CV(ISI) distribution of the trained and untrained model units and monkey V1 neurons. **E.** Cross-correlogram of the trained and untrained model units and monkey V1 neurons. Monkey V1 data in all plots were obtained from [4] in response to monkey's watching segments of the Star Wars movies.

We examined the model's spike activity in response to retina-filtered clips taken from the Star Wars movies that were not used for training, and found a number of V1-like response characteristics to emerge. We observed the model's spike responses to be sparse and diverse, with variability in the number and timing of spikes across units (Figure 2A). Further quantifying the spike responses revealed a firing rate distribution with an exponential fall-off (Figure 2B), as has similarly been reported in monkey [4] and cat V1 [24]. The average model's firing rate was also comparable to mean

V1 firing rates observed across different animal species in response to natural stimuli (model $6.57 \pm 14.68\text{Hz}$; anesthetized monkey V1: $5.06 \pm 0.75\text{Hz}$ [4]; anesthetized cat V1: $3.96 \pm 3.61\text{Hz}$ [24] and awake cat V1: $8.9 \pm 7\text{Hz}$ [25]). Additionally, the inhibitory model units exhibited a higher mean firing rate compared to the excitatory model units (Figure 2C). Such differences in firing rates between excitatory and inhibitory neurons to visual stimulation have also been reported in mouse V1 [26].

The model’s spike responses were highly variable, exhibiting a low degree of correlation between the units. We examined the variability in the spike responses using the coefficient of variation of the interspike intervals $\text{CV}(\text{ISI})$ (Figure 2D), which is defined as the ratio of the ISI standard deviation to the ISI mean. A spike train with a $\text{CV}(\text{ISI})$ below unity is more regular than a spike train with a $\text{CV}(\text{ISI})$ above unity. We observed a large range of variability in the spike trains, with sampled spike trains exhibiting more irregular (65% with $\text{CV}(\text{ISI}) \geq 1$) than regular (35% with $\text{CV}(\text{ISI}) < 1$) firing patterns, reminiscent of the spike variability observed in cat [27] and monkey V1 [4, 28].

We quantified the correlation between the different unit spike trains by calculating the spike train cross-correlogram, which measures the correlation between randomly-sampled unit pairs at different time lags (Figure 2E). We found the model units to have a low peak correlation ($CC = 0.07$ at lag 0s), with the correlation symmetrically decaying for increasing and decreasing lag. Similar correlation statistics have also been reported in anesthetized monkey [4] and in awake mouse V1 [29]. Lastly, the resemblance of the model’s spiking statistics to the biology appeared to be a direct consequence of optimizing the model for efficient temporal prediction, as evident by the untrained model’s firing rates (Figure 2B and C) and spike correlations (Figure 2E) being less similar to V1 than the trained model. However, we did notice a slight divergence in the trained model’s spike variability compared to V1 (Figure 2D).

Emergence of simple and complex cell-like tuning

We obtained model unit receptive fields (RFs) using a spike-triggered average to spatiotemporal white-noise input clips, and observed a considerable number of similarities between the model unit RFs and the RFs of V1 neurons across different animal species. Qualitatively, we found many units that resemble simple cells [1, 3], with stereotyped RFs consisting of varying number of excitatory and inhibitory subfields tuned to a particular orientation and spatial frequency [32, 33] (Figure 3A). The temporal structure of these RFs also resemble those of V1 neurons, as evident in their decay into the past (Figure 3B); their polarity profile, with RF polarity either changing (Figure 3B units 1, 4 and 5) or remaining fixed over time (Figure 3B units 2 and 3) [35–37]; and their spatiotemporal structure, with RFs either being space-time separable (Figure 3B units 1, 2 and 3) or space-time inseparable (Figure 3B units 4 and 5) [35].

We fit each model unit’s spatial RF at the time with the highest power using a Gabor function to further quantify its shape, orientation and spatial frequency. The distribution over RF shape characteristics, as defined by the width n_x and height n_y of the RF relative to its spatial frequency, matches the neural data (Figure 3C). We found the RF shapes to extend along a 1D-curve, from blob-like RFs at the origin (Figure 3A first row) to elongated bars with multiple subfields that are located further

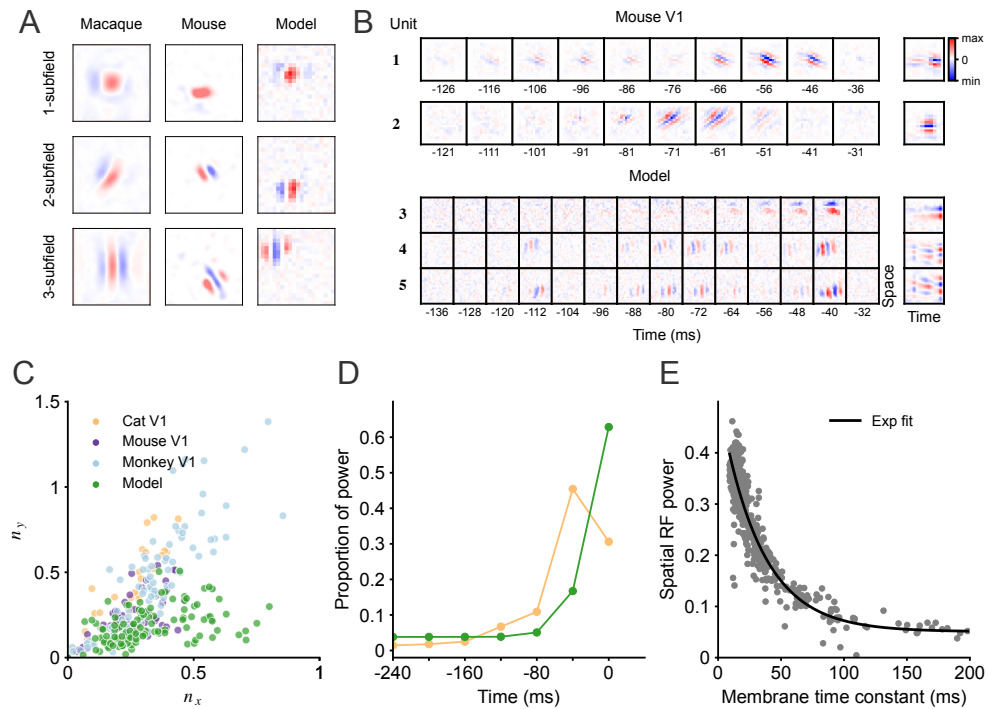


Fig. 3: Spatial and temporal receptive field properties of the model resemble V1. **A.** Example spatial RFs with varying number of subfields of macaque V1 neurons [30], mouse V1 neurons [26] and model units. **B.** 1 – 2: example space-time separable spatiotemporal RFs of mouse V1 neurons [31]. 3 – 5: example spatiotemporal RFs of model units (3: space-time separable and 4 – 5: space-time inseparable). Right column: corresponding spatiotemporal RFs obtained by summing along the RFs' axis of orientation (see Methods). **C.** Distribution of spatial RF shapes from V1 neurons in cats [32], mice [26], monkeys [33] and model units. Each model unit's spatial RF was taken at its time point of highest power. The x-axis (n_x) and y-axis (n_y) are a measure proportional to the number of RF subfields and their respective lengths [33]. **D.** Temporal RF power profile of model units and V1 cat neurons [34], quantified as the sum of squared weights over space and averaged across the population over ~ 40 ms time bins. **E.** Scatter plot of the spatial RF power (largest mean squared values of a spatial RF in time) and membrane time constant of the model units. Dark line is the fitted negative exponential curve ($R^2 = 0.85$).

away from the origin (Figure 3A last row). Our model captures a large portion of the blob-like RFs, which constitute a substantial part of the neural data (with n_x and $n_y < \sim 0.25$). It also captures units with multiple subfields ($n_x > \sim 0.25$), although the subfields are typically less elongated than in the biology. We found the model RFs to be most similar to the mouse V1 RFs, as measured by the Euclidean distance between the model RF-shape centroid and the different animal RF-shape centroids

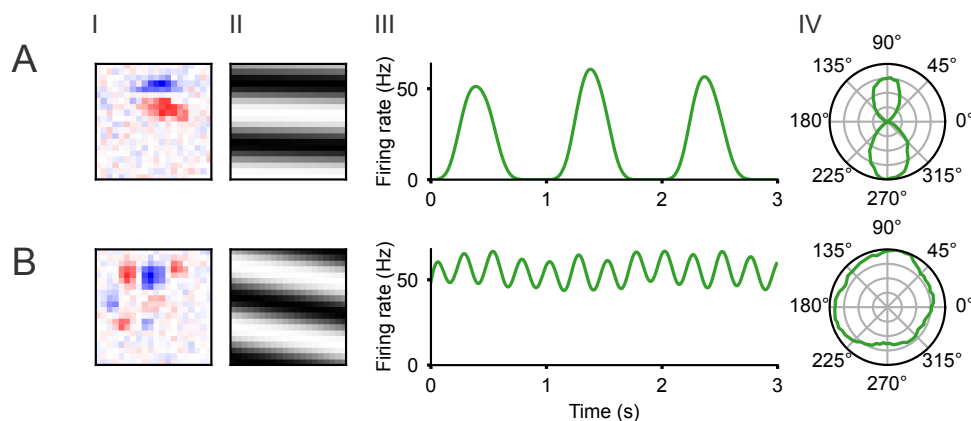


Fig. 4: Units resembling simple and complex cells are found in the model. **A.** Example simple cell-like unit, including its RF (I); optimal grating (II); response to its optimal grating (III); and the average response of the unit over time plotted in relation to grating orientation (in degrees) for gratings presented at the unit's preferred spatial and temporal frequency (IV). The optimal grating is defined as the full-field moving grating with an orientation and spatial and temporal frequency that give the highest average spike rate. **B.** Example complex cell-like unit.

(cat [32] centroid (0.26, 0.46) with distance 0.31; mouse [26] centroid (0.25, 0.24) with distance 0.09; monkey [33] centroid (0.33, 0.43) with distance 0.29; model centroid (0.26, 0.15)). The temporal profile across all model RFs also exhibited clear similarities to the neural data, as characterized by a decaying power profile (mean squared RF values over space, units and ~ 40 ms time bins) (Figure 3D).

We made an additional observation in our model, for which we found no prior reports in V1, by identifying a negative exponential relationship between the maximum spatial RF power and the membrane time constant of a unit (Figure 3E). The membrane time constant determines the operational timescale of a neuron, where neurons with smaller membrane time constants operate on shorter timescales than neurons with larger membrane time constants. Thus, translating our finding to the biology suggests that V1 neurons with a greater sensitivity to visual input operate on shorter timescales than neurons that exhibit a weaker stimulus sensitivity.

In addition to simple cells, V1 is also distinguished by another significant category of cells known as complex cells [2]. Like simple cells, complex cells are tuned for oriented edges and gratings. Unlike simple cells, however, complex cells are spatially invariant and respond to an oriented pattern regardless of its precise location in the neuron's RF. Using full-field drifting sinusoidal gratings, we qualitatively identified units with response characteristics reminiscent of both simple and complex cells. We observed unit responses to oscillate over time to the movement of their optimal grating. Like simple cells, some model units displayed a high phase dependence with pronounced variations in their response to grating movement [38] (Figure 4A). In contrast, other

units exhibited a larger mean response over time and were less influenced by the grating phase, consistent with the description of complex cells [39] (Figure 4B).

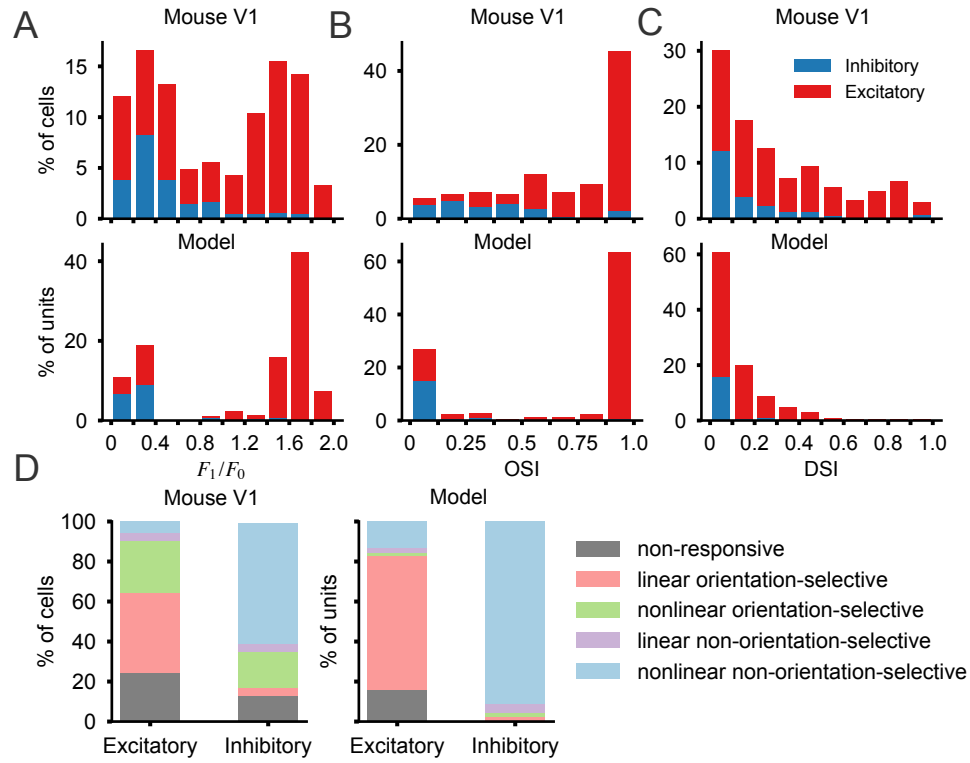


Fig. 5: Tuning measures of the model bear resemblance to those in V1. **A.** Distribution of modulation ratios F_1/F_0 . **B.** Distribution of orientation selectivity. **C.** Distribution of direction selectivity. **A-C** showing tuning properties of mouse V1 neurons [26] (top) and model units (bottom) to drifting gratings. **D.** Distribution of functional response categories for excitatory and inhibitory neurons in mouse V1 [26] (left) and for excitatory and inhibitory model units (right).

We further quantified the distribution of simple and complex cell-like units in our model by calculating their modulation ratio F_1/F_0 . This ratio is defined as the quotient between the amplitude of the first harmonic (F_1) and the mean value (F_0) of a unit's response to its optimal grating, with a value below one classifying a response as complex cell-like [40, 41]. The distribution of modulation ratios in our model is bimodal (Figure 5A), which has similarly been reported in V1 neurons of monkeys [42] and mice [26, 43]. Intriguingly, we found the inhibitory units (median $F_1/F_0 = 0.20$) to exhibit more complex cell-like responses compared to the excitatory units (median $F_1/F_0 = 1.62$) ($p = 3.59 \times 10^{-37}$; Mann-Whitney U test), a difference also found between inhibitory and excitatory V1 neurons in mice [26].

The model units exhibited similar orientation and direction preferences to those observed in V1 neurons in mice, as measured by the orientation selectivity index (OSI) and direction selectivity index (DSI), respectively [26, 44] (see Methods). Both measures range between zero and one, with a value of one corresponding to maximum selectivity. We found most units to have a strong preference for particular orientations ($OSI > 0.5$: 68%), with many units exhibiting near complete selectivity, as has similarly been reported in mouse V1 neurons ($OSI > 0.5$: 74%) [26] (Figure 5B). Capturing another difference between inhibitory and excitatory mouse V1 neurons [26], we found the inhibitory units (median $OSI = 0.03$) to be more broadly tuned than the excitatory units (median $OSI = 1.00$) ($p = 2.14 \times 10^{-43}$; Mann-Whitney U test). Additionally, we found the model units to have limited selectivity for direction ($DSI < 0.5$: 97%), as has similarly been reported for mouse V1 neurons ($DSI < 0.5$: 77%) [26] (Figure 5C).

We classified units into different functional response categories, classifying a unit as linear ($F_1/F_0 \geq 1$) or non-linear ($F_1/F_0 < 1$), and as orientation selective ($OSI \geq 0.5$) or non-orientation selective ($OSI < 0.5$) using the modulation ratio and OSI, respectively. The largest functional response categories in the model for both the inhibitory and excitatory units were the same as the largest functional response categories for inhibitory and excitatory neurons in mouse V1 [26] (Figure 5D). Specifically, most inhibitory model units are non-linear and non-orientation selective (91%), as are most inhibitory neurons in mouse V1 (60%), while most excitatory model units are linear and orientation selective (66%), as are most excitatory neurons in mouse V1 (40%).

Mirroring V1-cell physiology: membrane dynamics, connectivity and EI balance

V1 is characterized by distinct physiological differences between inhibitory and excitatory neurons, such as the timescales they operate at [45] and their connectivity statistics [8]. We found our model units to mirror these key differences between the inhibitory and excitatory neurons. Inhibitory neurons tend to operate at shorter timescales than excitatory neurons, as captured by differences in their membrane time constants. Mouse V1 inhibitory neurons exhibit a median membrane time constant of 9.76ms, whilst excitatory mouse V1 neurons have a notably longer median membrane time constant of 21.10ms [45] ($p = 3.43 \times 10^{-47}$; Mann-Whitney U test). Similar to mouse V1, we found the inhibitory and excitatory model units to operate at different timescales, with the inhibitory and excitatory units having a median membrane time constant of 12.64ms and 22.33ms, respectively ($p = 1.46 \times 10^{-30}$; Mann-Whitney U test) (Figure 6A).

Excitatory V1 neurons have been shown to synapse more strongly onto inhibitory V1 neurons than other excitatory neurons in mice [8] and rats [46], as measured by their excitatory postsynaptic potential (EPSP) amplitude. We made a similar observation in our model, with the excitatory units connecting more strongly with the inhibitory units than with other excitatory units, as measured by their absolute connection weight ($p = 0$; Mann-Whitney U test) (Figure 6B).

A fundamental aspect in neurophysiology is the concept of cortical neurons receiving a balance of excitatory and inhibitory (EI) synaptic currents, as reported in

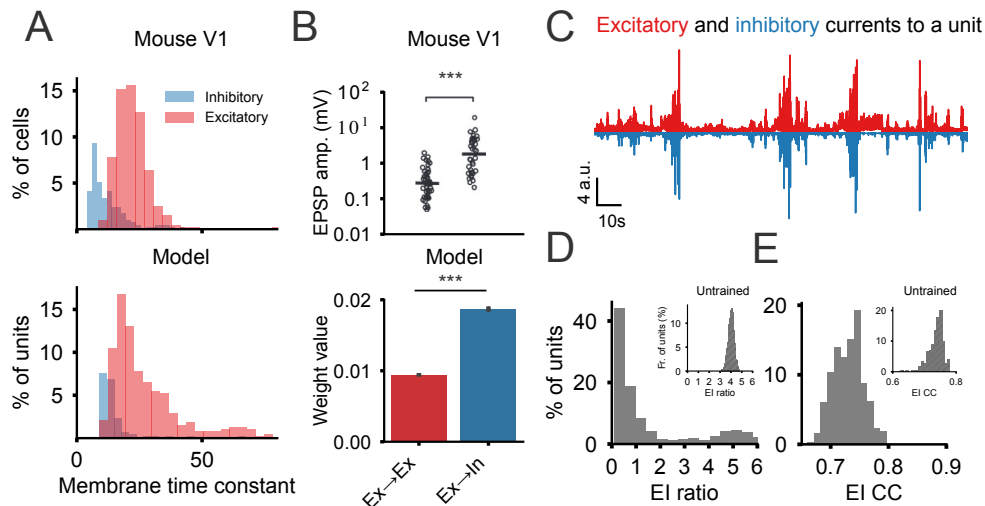


Fig. 6: Inhibitory and excitatory model units operate at different temporal scales, with distinct connectivity statistics and balanced excitatory and inhibitory currents, like in V1. **A.** Membrane time constant distribution of inhibitory and excitatory neurons in mouse V1 (layer 4) [45] (top), and of inhibitory and excitatory model units (bottom). Membrane time constants of 122 excitatory model units with a value above 80ms are not shown. **B.** Top: EPSP amplitudes between excitatory neurons (Ex→Ex) and from excitatory neurons to inhibitory neurons (Ex→In) in mouse V1 (black lines plot median amplitudes) [8]. Bottom: model weight values between excitatory units and from excitatory to inhibitory units. Bars plot the median and bootstrapped standard error, with statistical significance assessed using the Mann-Whitney U test (*** $p < 0.001$). **C.** Net-excitatory (top) and net-inhibitory (bottom) currents of an example model unit whilst inputting natural movie clips into the model. **D.** Distribution of global EI balance in the model. **E.** Distribution of precise EI balance in the model. Insets in **D.** and **E.** showcase the EI balances in the untrained model.

somatosensory cortex [47], primary auditory cortex [48] and V1 [49–51]. This equilibrium is crucial for ensuring stability in neural activity [52] and proper cortical function [53]. We qualitatively observed a balance between the excitatory and inhibitory currents in the model units to held-out natural input movie clips (Figure 6C). EI balance can be quantified as global when the average excitatory and inhibitory currents are equal over time, and as precise, when the excitatory and inhibitory currents are coupled in time, in which each excitatory input to a neuron arrives simultaneously with an inhibitory counterpart [54]. We found the model units to exhibit a median global EI ratio of 0.67 over the held-out natural input movie clips (Figure 6D), where an EI ratio below unity has similarly been reported in mouse V1 [51, 55, 56]. Furthermore, we found the model units to exhibit a precise EI balance, as indicated by a notable correlation between the net-excitatory and net-inhibitory currents over clips and time across the model units (median: 0.73) (Figure 6E). Lastly, we found the excitatory

and inhibitory currents in the model became more balanced during training for efficient temporal prediction, as evident in the untrained model having a notably larger median EI ratio of 4.04 (Figure 6D).

Discussion

Temporal prediction has been shown to capture the spatial and temporal tuning characteristics of V1 simple [10] and complex cells [11]. However, these prior modeling studies used feedforward ANNs that omitted recurrence, Dale’s law and all-or-nothing spiking. We explored whether a recurrently-connected, temporal-prediction spiking model incorporating Dale’s law would better replicate the functional properties of V1 neurons, particularly the differences known to exist between excitatory and inhibitory neuron subclasses. Our model captured simple and complex cell-like tuning and V1-like spike statistics to natural stimuli. Notably, the excitatory and inhibitory units in the model exhibited several properties that match their biological counterparts in V1. The much greater explanatory power of the recurrent, spiking model provides strong evidence that temporal prediction may be critical to the function of V1 and potentially other sensory brain areas.

A major goal of computational neuroscience is to develop models that help to interpret the vast amount of experimental data obtained from studies of regions such as V1. This requires creating biologically-realistic models with a functional objective that stipulates how neurons transform visual input [57]. Some previous studies have developed spiking models of V1 with distinct subpopulations of excitatory and inhibitory units. These models have demonstrated V1-like spike statistics and motion tuning [4, 58–61]. However, their parameters were hard-coded to produce these phenomena, rather than being trained to achieve a particular goal like temporal prediction. Consequently, they have provided limited insight into the functional principles underlying the transformation of visual signals in the cortex.

Other studies have shown how specific functional objectives can account for simple cell-like and in some cases complex cell-like properties. These objectives include efficient [62, 63] and sparse coding [64–66], which assume that V1 efficiently encodes sensory input under certain metabolic constraints; independent component analysis [67], which stipulates that V1 reduces input redundancies by encoding independent features within sensory stimuli [68, 69]; predictive coding, which also posits that neurons remove statistical redundancies but in this case by signalling unpredictable features within sensory input [70–73]; temporal coherence [74], slow subspace analysis [75] and slow feature analysis [76, 77], which hypothesize that V1 encodes slowly-varying features within the stimulus; and the information bottleneck hypothesis [78], which suggests that V1 encodes sensory features with maximum mutual information about the future, whilst minimizing information about the past [79]. However, these studies all used non-spiking models and did not include Dale’s law. An exception are sparse coding models that have been extended to be spiking [30, 80, 81] and to include Dale’s law [81]. Although these spiking sparse coding models successfully capture V1 spatial RF properties, they ignore their temporal tuning since the models were trained on

images. Temporal prediction has also previously been explored in a spiking form [82], but this study focused on the retina and did not explore Dale’s law.

Whilst the inclusion of spikes and Dale’s law increased the biological realism of our model over prior studies, there remains scope to include additional cortical circuit motifs. The cortex is a 6-layered structure, with layer 4 being the main target for sensory inputs from the thalamus [83, 84]. We currently envisage the units of our model comprising those over all layers. Sub-dividing our model units into a similarly multi-layered structure would allow us to examine further phenomena, like the layer-specific tuning characteristics found in V1 [26]. Additionally, our model could be extended to operate over an entire visual scene instead of patches, by jointly training multiple versions of the model at different spatial locations within the visual frame. Incorporating cross-spatial connectivity may segregate the tuning statistics in this model to capture the canonical columnar organization of V1 [2], where neurons with similar orientation preference cluster together [85, 86], or the lateral connectivity statistics, where different neurons are more likely to be connected if they share similar tuning preferences [87, 88]. The optimization of cortex for temporal prediction, at least for the non-hard-wired aspects of cortical tuning, could be guided by the error-sensitive neurons found in cortex [9], or possibly by hypothesized error-sensitive dendrites [89]. These error-sensitive components could compare long (such as from layer 2/3) and short (such as from layers 4 or 5/6) latency neural activity [90] to provide a temporal-prediction error signal [11, 91]. Modeling the learning using more biologically-based local learning mechanisms [92] and structures that are congruent with these phenomena would be another valuable future avenue.

Although our model exhibits V1-like firing dynamics, modelling the units using the adaptive leaky integrate-and-fire model would better match the adaptive spiking dynamics of V1 neurons [15], for example, the decrease in firing observed in response to a constant input stimulus [93]. This would also allow differences in adaptation between excitatory and inhibitory V1 neurons [94, 95] to be explored. In our model, we assumed the effect of input current from one unit to another to be instantaneous. However, synapses conduct current with a temporally-decaying profile, where different types of synapses are governed by a fast or slow decay [96]. Extending our model to include learnable synaptic-like temporal profiles may capture these synaptic phenomena. This would ideally require training our model at a finer temporal resolution than $\sim 8.33\text{ms}$ per time step, which remains a challenge for spiking networks due to speed and memory constraints (although see [97, 98]).

Our temporal prediction model makes several biological predictions. First, based on our finding of a negative exponential relationship between the highest spatial RF power and the membrane time constant across the model units, V1 neurons that are more sensitive to visual input may operate at faster timescales. Second, our findings have implications for calculating the exact energy demands of neurons, which remains an ongoing area of research [99–102]. Specifically, the model predicts that inhibitory V1 neurons may consume less energy per spike than their excitatory counterparts, as we found this assumption to be important for obtaining our results. Third, our results are relevant to the long-standing question of the extent to which visual processing is hierarchical [103, 104] - transforming through sequential layers - or shallow - emerging

through recurrent processes. Our model is consistent with the shallow brain hypothesis [105], as it showcases how a single layer of recurrently-connected spiking units captures non-linear phenomena like complex cell-like responses, which have otherwise been argued to emerge from hierarchical processing [106]. This may inspire new computer vision models that favor recurrence over standard hierarchical architectures [107].

In conclusion, we have shown how efficient temporal prediction in a spiking model accounts for simple and complex cell-like tuning, V1-like spike statistics and, notably, key differences between excitatory and inhibitory V1 neurons. Our findings contribute to the growing evidence that visual cortex supports prediction of the sensory future [10, 11]. Although predictive processing is increasingly widely recognized [9], it is perhaps surprising that this single computational principle accounts for several key differences between excitatory and inhibitory V1 neurons. Perhaps many more of the brain’s perplexing phenomena may be explained by the principle of temporal prediction.

Methods

Spiking V1 model

Training data

We used a publicly available dataset of natural movie recordings to train and test our model [108]. This dataset consists of 39 clips, of 9.6 second duration, of everyday objects and scenes (*e.g.* a dog walking, a fish swimming, or a human’s view walking through nature), recorded using different camera movements (*e.g.* moving forwards or sideways, panning or still). We used 31 clips for training and 8 clips for testing. All clips are grayscale, with a temporal resolution of 120Hz and a spatial resolution of 140×240 px. To emulate the transformation of the retina and thalamus, we bandpass filtered all clips, as done in other studies [10, 109]. We also normalized the training and test data by subtracting the pixel mean and dividing by the pixel standard deviation of the training data, and finally clipped the absolute pixel values to be no larger than 3.5 standard deviations. Training was performed using clips consisting of 20×20 px spatial patches of 350ms duration (*i.e.* 42 frames). The spatial position of each patch were randomly sampled on each training batch. Clips were also randomly flipped around the vertical axis to artificially increase the training data [110].

Spiking neurons

We modeled the V1 cortical neurons as a population of $N = 600$ recurrently-connected excitatory and inhibitory spiking units. The input current $I_i[t] \in \mathbb{R}$ to unit i at time step t is derived from the bandpass-filtered input movie stimulus $x \in \mathbb{R}^{T \times H \times W}$ (of $T = 42$ frames, and spatial height $H = 20$ px and spatial width $W = 20$ px); model output spikes $S[t - 1] \in \mathbb{R}^N$ from the previous time step; and a bias term $b_i^{\text{in}} \in \mathbb{R}$.

$$I_i[t] = b_i^{\text{in}} + \underbrace{\sum_{t'=1}^{T_E} \sum_{h=1}^H \sum_{w=1}^W W_{it'hw}^{\text{in}} x_{hw}[t - t' + 1]}_{\text{Feedforward current}} + \underbrace{\sum_{j=1, j \neq i}^N W_{ij}^{\text{rec}} S_j[t - 1]}_{\text{Recurrent current}} \quad (1)$$

The feedforward connectivity $W^{\text{in}} \in \mathbb{R}^{N \times T_E \times H \times W}$ (of $T_E = 15$ temporal frame span) linearly maps the input movie stimulus into a feedforward current contribution (capturing the transformation of the retina and thalamus); the recurrent connectivity $W^{\text{rec}} \in \mathbb{R}^{N \times N}$ maps previous output spikes into a recurrent current contribution. We implemented all spiking units using the normalized and discretized leaky integrate-and-fire (LIF) model, which evolves the model membrane potential $V_i[t] \in \mathbb{R}$ of unit i at time step t using the following difference equation:

$$V_i[t] = \left(\beta_i V_i[t-1] + (1 - \beta_i) I_i[t] \right) (1 - S_i[t]) \quad (2)$$

During each time step, the membrane potential undergoes a decay governed by a factor β_i . We let each neuron learn its decay factor, and bounded values into the range $(0, 1)$ to enforce correct LIF dynamics [111]. $I_i[t]$ was replaced with $\tilde{I}_i[t]$ when synaptic noise was included. The membrane potential resets to zero in the event of a spike occurring in the preceding time step, which happens when the membrane potential reaches the firing threshold (equal to one).

$$S_i[t] = \begin{cases} 1 & \text{if } V_i[t] > 1 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Finally, at every time step t , a prediction of the future spatial movie frame $\hat{y}[t] \in \mathbb{R}^{H \times W}$ was generated (with $\hat{y}_{hw}[t] \in \mathbb{R}$ denoting the predicted pixel value). This was achieved by linearly mapping a span of $T_D = 2$ temporal frames of proceeding spike activity using projection weights $W^{\text{out}} \in \mathbb{R}^{N \times T_D \times H \times W}$, plus a bias $b^{\text{out}} \in \mathbb{R}$.

$$\hat{y}_{hw}[t] = b^{\text{out}} + \sum_{i=1}^N \sum_{t'=1}^{T_D} W_{it'hw}^{\text{out}} S_i[t - t' + 1] \quad (4)$$

V1 latencies

V1 neurons have a response latency of at least $\sim 30 - 40$ ms in mice and monkeys [17–20], and these latencies can vary substantially between neurons [112]. We modeled the transduction and transmission latency to V1, by zero-masking the first 5 spatial frames (*i.e.* 42ms) within the feedforward connectivity tensor W^{in} . This length is the temporal prediction span.

Neural noise

Cortical V1 neurons do not respond with identical spike trains to repeated visual stimulus presentations. One of the reasons for this is neural noise. We include two neural noise sources within our model. First, we modeled the noisy photoreceptors in the retina [113], by adding Gaussian noise to the bandpass-filtered movie clips $\epsilon_p \sim \mathcal{N}(0, 0.2^2)$. Second, we modeled the synaptic noise of cortical V1 neurons [114], by multiplicatively perturbing the input current $I_i[t]$ of every unit i at time step t using Gaussian sampled noise $\epsilon_g \sim \mathcal{N}(0, 0.6^2)$.

$$\tilde{I}_i[t] = I_i[t](1 + \epsilon_g) \quad (5)$$

Dale’s law: Enforcing units to be inhibitory or excitatory

Approximately 15 – 20% of cortical V1 neurons are inhibitory in mice, cats, and monkeys [5–8]. We assigned 15% of the model unit population to be inhibitory, resulting in $N_I = 90$ inhibitory units and $N_E = 510$ excitatory units. We enforced a neuron to be excitatory (or inhibitory), by taking the absolute value of every weight in the unconstrained recurrent weight matrix $\tilde{W}^{\text{rec}}$, and appropriately negating connection weights from unit j to i . All outwards connection weights from unit j are either positive (*i.e.* excitatory) or negative (*i.e.* inhibitory).

$$W_{ij}^{\text{rec}} = \begin{cases} -|\tilde{W}_{ij}^{\text{rec}}| & \text{if } j < N_I \\ |\tilde{W}_{ij}^{\text{rec}}| & \text{otherwise} \end{cases} \quad (6)$$

Weight initialization

The initial weights for feedforward (W^{in}), recurrent (W^{rec}), and projection (W^{out} - *i.e.* the readout weights) connections were randomly initialized from a uniform distribution $U(-k, k)$, where k was determined as $k = \frac{1}{\sqrt{T_E H W}}$ for the feedforward weights, $k = \frac{0.1}{\sqrt{N}}$ for the recurrent weights, and $k = \frac{1}{\sqrt{N T_D}}$ for the projection weights. Additionally, all initial membrane time constants were uniformly initialized to 20ms (*i.e.* $\beta_i \approx 0.81$ for all units). Lastly, the input bias terms were initialized to 0.2 and the output bias term to 0.

Loss function

We trained the model by minimizing loss $\mathcal{L}_{\text{total}}$, comprised of prediction loss $\mathcal{L}_{\text{normative}}$ and metabolic loss $\mathcal{L}_{\text{metabolic}}$, with respect to W^{in} , W^{rec} , W^{out} , b^{in} , b^{out} , and β . The prediction loss measures the similarity between the predicted and target future movie frames, and the metabolic loss measures the approximated physiological energy use of the network. We weighted this loss by hyperparameter $\lambda = 10^{-2.75}$, which we selected to produce the most V1-like receptive fields.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{prediction}} + \lambda \underbrace{\left(\gamma_{\text{transmission}} \mathcal{L}_{\text{input}} + (1 - \gamma_{\text{transmission}}) \mathcal{L}_{\text{spiking}} \right)}_{\text{Metabolic loss } \mathcal{L}_{\text{metabolic}}} \quad (7)$$

We computed the prediction loss $\mathcal{L}_{\text{prediction}}$ as the mean squared error (MSE) between all predicted spatial movie frames $\hat{y}$ and target spatial movie frames $y \in \mathbb{R}^{H \times W}$. The calculation was performed by taking the average over all batch samples B (omitted here for brevity), simulation steps T (starting from simulation step $t_0 = 5$ to allow unit membrane potentials to depolarize sufficiently, *i.e.* to enable network warm-up), and spatial dimensions H (height) and W (width), both which were cropped by c pixels on all sides.

$$\mathcal{L}_{\text{prediction}} = \underbrace{\frac{1}{(T - t_0)(H - c)(W - c)} \sum_{t=t_0}^T \sum_{h=c}^{H-c} \sum_{w=c}^{W-c}}_{\text{Avg. over time and space}} \underbrace{(\hat{y}_{hw}[t] - y_{hw}[t])^2}_{\text{MSE}} \quad (8)$$

We defined our metabolic-like loss $\mathcal{L}_{\text{metabolic}}$ as the average synaptic-like transmission over all the model weights, where synaptic transmission has been noted to require significant energy within cerebral cortex [102]. We defined the energy cost of the synaptic-like transmission of a single model weight as its absolute value, multiplied by the absolute value of every input value, summed over all time steps. We weighted the recurrent transmission ($\mathcal{L}_{\text{spiking}}$) relative to the input transmission ($\mathcal{L}_{\text{input}}$) using hyperparameter $\gamma_{\text{transmission}} = 0.3$. We also assumed the metabolic-like cost of the excitatory and inhibitory units to differ, with the inhibitory units consuming less energy than excitatory units, captured by hyperparameter $\gamma_{\text{type}} = 0.1$.

$$\mathcal{L}_{\text{input}} = \gamma_{\text{type}} \mathcal{L}_{\text{inhibitory-input}} + (1 - \gamma_{\text{type}}) \mathcal{L}_{\text{excitatory-input}} \quad (9)$$

$$\mathcal{L}_{\text{spiking}} = \gamma_{\text{type}} \mathcal{L}_{\text{inhibitory-spiking}} + (1 - \gamma_{\text{type}}) \mathcal{L}_{\text{excitatory-spiking}} \quad (10)$$

$$\mathcal{L}_{\text{inhibitory-input}} = \underbrace{\frac{1}{N_I T} \sum_{i=1}^{N_I} \sum_{t=1}^T}_{\text{Avg. over inhibitory units and time}} \underbrace{\left(|b_i| + \sum_{t'=1}^{T_E} \sum_{h=1}^H \sum_{w=1}^W |W_{it'hw}| |x_{hw}[t - t' + 1]| \right)}_{\text{Afferent synaptic cost}} \quad (11)$$

$$\mathcal{L}_{\text{excitatory-input}} = \underbrace{\frac{1}{N_E T} \sum_{i=N_I}^N \sum_{t=1}^T}_{\text{Avg. over excitatory units and time}} \underbrace{\left(|b_i| + \sum_{t'=1}^{T_E} \sum_{h=1}^H \sum_{w=1}^W |W_{it'hw}| |x_{hw}[t - t' + 1]| \right)}_{\text{Afferent synaptic cost}} \quad (12)$$

$$\mathcal{L}_{\text{inhibitory-spiking}} = \underbrace{\frac{1}{N_I T} \sum_{i=1}^{N_I} \sum_{t=1}^T}_{\text{Avg. over inhibitory units and time}} \underbrace{\left(\sum_{j=1, j \neq i}^N |W_{ij}^{\text{rec}}| |S_j[t - 1]| \right)}_{\text{Recurrent synaptic cost}} \quad (13)$$

$$+ \underbrace{\frac{1}{T H W} \sum_{t=1}^T \sum_{h=1}^H \sum_{w=1}^W}_{\text{Avg. over time and space}} \underbrace{\left(\sum_{i=1}^{N_I} \sum_{t'=1}^{T_D} |P_{it'hw}| |S_i[t - t' + 1]| \right)}_{\text{Efferent synaptic cost}}$$

$$\mathcal{L}_{\text{excitatory-spiking}} = \underbrace{\frac{1}{N_E T} \sum_{i=N_I}^N \sum_{t=1}^T}_{\text{Avg. over excitatory units and time}} \underbrace{\left(\sum_{j=1, j \neq i}^N |W_{ij}^{\text{rec}}| |S_j[t - 1]| \right)}_{\text{Recurrent synaptic cost}} \quad (14)$$

$$+ \underbrace{\frac{1}{THW} \sum_{t=1}^T \sum_{h=1}^H \sum_{w=1}^W}_{\text{Avg. over time and space}} \underbrace{\left(\sum_{i=N_I}^N \sum_{t'=1}^{T_D} |P_{it'hw}| S_i[t - t' + 1] \right)}_{\text{Efferent synaptic cost}}$$

Surrogate gradient descent

We trained the spiking models using a variation of the backpropagation algorithm [22] called surrogate gradient descent [23]. Here, to enable smooth gradient flow during training, a well-behaved function was employed as a replacement for the gradient of the non-differentiable Heaviside step function. We adopted the fast sigmoid function, which has been shown to perform well in practice [115, 116].

$$\frac{\partial S_i[t]}{\partial V_i[t]} = (10|V_i[t] - 1| + 1)^{-2} \quad (15)$$

We conducted all training using the Adam optimizer [117], employing its default parameters. The training spanned 1200 epochs with an initial learning rate set to 10^{-4} , which was subsequently reduced by a factor of 0.2 at epoch count 200, 600, and 800. We performed training using a batch size of 1024 and saved model weights whenever a new minimum training loss was achieved.

Spike analysis

For all the spike analysis, we used the model output spike trains resulting from 8000 different retina-filtered movie patches taken from the Star Wars movies. We used these natural movies in our spiking analyses to be comparable to [4], where they calculated V1 spiking statistics in monkeys watching segments of the Star Wars movies. For additional comparability with the analysis of [4], we calculated all spike statistics using a time window of 2s with a step size of 0.2s.

Coefficient of variation

We quantified the variability of the spike pattern in each spike train by calculating the coefficient of variation of the interspike interval $CV(ISI)$, which is defined as the ratio between the standard deviation and the mean of the interspike intervals [15]. A $CV(ISI) < 1$ corresponds to a more regular firing pattern than a $CV(ISI) > 1$, and a Poisson process has $CV(ISI) = 1$. As done in [4], we only calculated $CV(ISI)$ values for a given spike train if it had at least three spikes.

Neuron synchronization

We quantified the correlation between different unit spike trains by calculating the mean cross-correlogram between the spike trains of randomly-sampled unit pairs. This measures the average Pearson correlation coefficient $\hat{\rho}[\tau]$ between the spike trains s_1 and s_2 of different units for varying temporal lag τ , defined as:

$$\hat{\rho}[\tau] = \mathbb{E}_{s_1, s_2} \left[\frac{\mathbb{E}[(s_1[t] - \mu_{s_1})(s_2[t - \tau] - \mu_{s_2})]}{\sigma_{s_1} \sigma_{s_2}} \right] \quad (16)$$

We sampled 400 random unit pairs for this calculation. Furthermore, we binned the spike trains with a bin size of 25ms before calculating the correlations, in order for our analyses to be comparable to [4].

Receptive field analysis

Spike-triggered average

We estimated the spatiotemporal receptive field of every unit using a spike-triggered average [96], using responses to 1000 spatiotemporal white-noise input clips of 100-frame duration. Clips were generated by sampling each pixel from a Gaussian distribution with a standard deviation of $\sigma = 10$.

Fitting Gabors

We fitted a Gabor function [118] to the RF of every unit to quantify its tuning properties. The Gabor function has been shown to provide a good approximation for most RF spatial aspects [32, 33] and provides a quantitative measure to compare the RFs of the model with the RFs of V1 neurons. The 2D Gabor function is defined as:

$$G(\hat{x}, \hat{y}) = A \exp \left(- \left(\frac{\hat{x}}{\sqrt{2}\sigma_x} \right)^2 - \left(\frac{\hat{y}}{\sqrt{2}\sigma_y} \right)^2 \right) \cos(2\pi f \hat{x} + \phi) \quad (17)$$

where spatial coordinates $(\hat{x}, \hat{y})$ are obtained by translating the center of the RF (x_0, y_0) to $(0, 0)$ and rotating the RF by its orientation θ :

$$\hat{x} = (x - x_0) \cos(\theta) + (y - y_0) \sin(\theta) \quad (18)$$

$$\hat{y} = -(x - x_0) \sin(\theta) + (y - y_0) \cos(\theta) \quad (19)$$

The amplitude of the Gaussian envelope is defined by A , whereas parameters σ_x and σ_y define the Gaussian envelope width along the $\hat{x}$ and $\hat{y}$ axes, respectively. Parameters f and ϕ define the spatial frequency (in cycles/pixel) and phase of the sinusoidal grating along $\hat{x}$. We optimized the Gabor parameters for each model unit RF by minimizing the mean squared error between the corresponding Gabor function and its respective model RF. To avoid local minima during fitting, we performed each Gabor fit in the spectral domain, by transforming the model RFs and Gabors using a 2D Fourier transform [10, 32]. We then used the resulting parameters to continue the fitting procedure in the spatial domain. We excluded units in Figure 3C which had a poor fit (correlation coefficient < 0.6 , 381 units), whose fit was based on very few pixel values (Gaussian envelopes $\sigma_x < 0.5$ or $\sigma_y < 0.5$, 382 units), and whose center position (x_0, y_0) of the Gabor was outside the RF (9 units). This resulted in 141 out of 519 active units with a sufficient fit, having a median pixel-wise correlation coefficient of 0.83 (81 out of the 600 model units were deemed inactive, see Motion tuning section).

Spatial receptive field properties

Power. The power of a RF is defined as the mean of the squared values of the RF over a span of time (as in Figure 3D) or at a point in time (as in Figure 3E) [10].

Structure. The structure of a RF is characterized by two variables $n_x = \sigma_x f$ and $n_y = \sigma_y f$, which give a measure of the number of oscillations of its sinusoidal component and the length of the bars in the RF, respectively [33]. Blob-like RFs lie close to the origin in the n_x by n_y plane, the number of bars increases along the n_x axis, and the RF bar-length increases along the n_y axis.

Estimating space-time separability

Spatiotemporal RFs are either space-time separable or inseparable, with space-time separable RFs tuned to phenomena such as bars flashing or moving regardless of direction, and space-time inseparable RFs tuned to direction of movement. Space-time separable spatiotemporal RFs can be represented as a product of a spatial and temporal function while space-time inseparable spatiotemporal RFs cannot [96]. We classified the space-time separability of model unit RFs using the singular values of a singular value decomposition (SVD), as done in [10]. We flattened the spatial dimensions and removed the transmission latency time steps for each model unit RF, and took the SVD of the resulting matrix. If the ratio between the second and first singular value was < 0.5 , the RF was deemed space-time separable; otherwise, the RF was deemed space-time inseparable.

2D spatiotemporal receptive fields

The 2D spatiotemporal RF illustrates the temporal evolution of the spatiotemporal RF [35]. This view was constructed by translating and rotating every spatial RF by the position and rotation that centers the spatial RF of the largest power, placing all oriented bars parallel to the y-axis. Next, all spatial RFs were flattened along the y-axis and concatenated, resulting in a 2D spatiotemporal RF. This new view can qualitatively inform us about the temporal phase (through polarity changes in the flattened spatial structure) and separability (through shifts in the flattened spatial structure) of a spatiotemporal RF.

Motion tuning

Virtual physiology

We constructed model unit responses to drifting full-field sinusoidal gratings, varying in orientation (from 0° to 360° in 5° increments), spatial frequency (ten values evenly spaced from 0.01 to 0.2 cycles per pixel), and temporal frequency (1, 2, 4, and 8Hz), each with an amplitude of one. Each grating clip, lasting 3 seconds, was presented to the model four times. We calculated the firing rate of each model unit by averaging responses over clip repeats and convolving with a Gaussian kernel ($\sigma = 72\text{ms}$) [96]. For each model unit, we computed its mean firing rate over each grating clip, and constructed a 3D mean-firing tuning space (over orientation, spatial, and temporal frequency), with the optimal grating (defined by its orientation, spatial and temporal

frequency) eliciting the largest mean firing rate. We only included responsive units in the tuning analysis of Figure 5 (519 out of 600 units) that had an optimal response no less than 10% of the mean optimal response over all model units.

Modulation ratio

We measured the modulation ratio F_1/F_0 of each model unit by taking the Fast Fourier Transform (FFT) of each unit's response to its optimal grating and calculating the ratio between the amplitude of the first harmonic (F_1) and DC component (F_0). A unit with a modulation ratio below one is considered complex cell-like, whereas a unit with a modulation above one is considered simple cell-like [26, 40, 41].

Orientation and direction selectivity

We calculated the orientation selectivity index (OSI) and direction selectivity index (DSI) for each unit, which quantify a unit's preference for stimulus orientation and direction of motion, respectively [26, 44]:

$$\text{OSI} = \frac{R_{\text{pref}}^{\text{orient}} - R_{\text{orth}}^{\text{orient}}}{R_{\text{pref}}^{\text{orient}} + R_{\text{orth}}^{\text{orient}}} \quad (20)$$

$$\text{DSI} = \frac{R_{\text{pref}}^{\text{dir}} - R_{\text{non-pref}}^{\text{dir}}}{R_{\text{pref}}^{\text{dir}} + R_{\text{non-pref}}^{\text{dir}}} \quad (21)$$

Both metrics are calculated using the mean unit responses to gratings at a unit's preferred spatial and temporal frequency. $R_{\text{pref}}^{\text{orient}} = R_{\text{pref}}^{\text{dir}}$ is a unit's mean response to a stimulus in its preferred orientation and direction, and $R_{\text{orth}}^{\text{orient}}$ and $R_{\text{non-pref}}^{\text{dir}}$ are the mean responses in the orthogonal orientation and opposite direction to the preferred direction, respectively. A value close to zero corresponds to less selective responses, and a value closer to one corresponds to greater selectivity for both measures.

Physiology analysis

Calculating membrane time constants

The membrane time constant τ of every model unit is obtained from the decay factor β in the LIF Equation using the following relation:

$$\tau = \frac{-\Delta t}{\ln(\beta)} \quad (22)$$

EI balance

Using the held-out test dataset of natural movie stimuli, we calculated the net-excitatory (Ex_i) and net-inhibitory (In_i) input current to each model unit i at time

step t as:

$$\text{Ex}_i[t] = \max(0, I_i[t]) + \sum_{j=N_I, j \neq i}^N W_{ij}^{\text{rec}} S_j[t-1] \quad (23)$$

$$\text{In}_i[t] = \min(0, I_i[t]) + \sum_{j=1, j \neq i}^{N_I} W_{ij}^{\text{rec}} S_j[t-1] \quad (24)$$

The net-excitatory and net-inhibitory currents were derived from the excitatory and inhibitory units, respectively, with the feedforward current $I_i[t]$ being classed as either excitatory or inhibitory at a given time step, depending on its sign value. From this, we calculated the global EI balance of a model unit as:

$$\frac{\sum_t \text{Ex}_i[t]}{\sum_t |\text{In}_i[t]|} \quad (25)$$

with the excitatory and inhibitory current time series for each input clip concatenated over time.

We calculated the precise EI balance of every model unit by first convolving each current time series using a Gaussian kernel ($\sigma = 72\text{ms}$) [96] and then calculating the resulting correlation between the smoothed (negative) inhibitory and (positive) excitatory current time series (with both time series concatenated over time for each input clip). We smoothed the current time series to emulate synaptic conductance [15], finding the non-smoothed excitatory and inhibitory current time series of units to otherwise be uncorrelated (median CC of 0.04).

Acknowledgments. Luke Taylor was supported by the Clarendon Fund of the University of Oxford. Andrew King and Nicol Harper were supported by the Wellcome Trust (WT108369/Z/2015/Z). Friedemann Zenke was supported by Swiss National Science Foundation (grant no. PCEFP3.202981) and the Novartis Research Foundation.

Data availability

The natural movie training data can be downloaded from https://figshare.com/articles/dataset/Natural_movies/24265498 and movie test data used in the spiking analyses can be downloaded at <https://github.com/webstorms/V1Model>. The untrained and pre-trained spiking models can be found at <https://github.com/webstorms/V1Model>. Experimental data in Fig. 5 was extracted using an online tool available at <https://apps.automeris.io/wpd/>.

Code availability

The code of the spiking V1 model and the code for reproducing the results can be found at <https://github.com/webstorms/V1Model>.

References

- [1] Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology* **148**, 574–591 (1959).
- [2] Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology* **160**, 106 (1962).
- [3] Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology* **195**, 215–243 (1968).
- [4] Rasch, M. J., Schuch, K., Logothetis, N. K. & Maass, W. Statistical comparison of spike responses to natural stimuli in monkey area v1 with simulated responses of a detailed laminar network model for a patch of v1. *Journal of neurophysiology* **105**, 757–778 (2011).
- [5] Sillito, A. The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. *The Journal of physiology* **250**, 305–329 (1975).
- [6] Fitzpatrick, D., Lund, J., Schmechel, D. & Towles, A. Distribution of gabaergic neurons and axon terminals in the macaque striate cortex. *Journal of Comparative Neurology* **264**, 73–91 (1987).
- [7] Gonchar, Y., Wang, Q. & Burkhalter, A. H. Multiple distinct subtypes of gabaergic neurons in mouse visual cortex identified by triple immunostaining. *Frontiers in neuroanatomy* **2**, 93 (2008).
- [8] Hofer, S. B. *et al.* Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature neuroscience* **14**, 1045–1052 (2011).
- [9] Keller, G. B. & Mrsic-Flogel, T. D. Predictive processing: a canonical cortical computation. *Neuron* **100**, 424–435 (2018).
- [10] Singer, Y. *et al.* Sensory cortex is optimized for prediction of future input. *elife* **7**, e31557 (2018).
- [11] Singer, Y., Taylor, L., Willmore, B. D., King, A. J. & Harper, N. S. Hierarchical temporal prediction captures motion processing along the visual pathway. *Elife* **12**, e52599 (2023).

- [12] Bialek, W., Nemenman, I. & Tishby, N. Predictability, complexity, and learning. *Neural computation* **13**, 2409–2463 (2001).
- [13] Richards, B. A. *et al.* A deep learning framework for neuroscience. *Nature neuroscience* **22**, 1761–1770 (2019).
- [14] Eccles, J. C. From electrical to chemical transmission in the central nervous system: the closing address of the sir henry dale centennial symposium cambridge, 19 september 1975. *Notes and records of the Royal Society of London* **30**, 219–230 (1976).
- [15] Gerstner, W., Kistler, W. M., Naud, R. & Paninski, L. *Neuronal dynamics: From single neurons to networks and models of cognition* (Cambridge University Press, 2014).
- [16] De Valois, R. L., Cottaris, N. P., Mahon, L. E., Elfar, S. D. & Wilson, J. A. Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision research* **40**, 3685–3702 (2000).
- [17] Raiguel, S. E., Lagae, L., Gulyàs, B. & Orban, G. A. Response latencies of visual cells in macaque areas v1, v2 and v5. *Brain research* **493**, 155–159 (1989).
- [18] Nowak, L., Munk, M., Girard, P. & Bullier, J. Visual latencies in areas v1 and v2 of the macaque monkey. *Visual neuroscience* **12**, 371–384 (1995).
- [19] Land, R., Engler, G., Kral, A. & Engel, A. Response properties of local field potentials and multiunit activity in the mouse visual cortex. *Neuroscience* **254**, 141–151 (2013).
- [20] Kirchberger, L., Mukherjee, S., Self, M. W. & Roelfsema, P. R. Contextual drive of neuronal responses in mouse v1 in the absence of feedforward input. *Science advances* **9**, eadd2498 (2023).
- [21] Bixler, E. O., Bartlett, N. R. & Lansing, R. W. Latency of the blink reflex and stimulus intensity j. *Perception & Psychophysics* **2**, 559–560 (1967).
- [22] Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *nature* **323**, 533–536 (1986).
- [23] Neftci, E. O., Mostafa, H. & Zenke, F. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine* **36**, 51–63 (2019).
- [24] Baddeley, R. *et al.* Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **264**, 1775–1783 (1997).

- [25] Legendy, C. & Salzman, M. Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons. *Journal of neurophysiology* **53**, 926–939 (1985).
- [26] Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *Journal of Neuroscience* **28**, 7520–7536 (2008).
- [27] Holt, G. R., Softky, W. R., Koch, C. & Douglas, R. J. Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of neurophysiology* **75**, 1806–1814 (1996).
- [28] Softky, W. R. & Koch, C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of neuroscience* **13**, 334–350 (1993).
- [29] Arroyo, S., Bennett, C. & Hestrin, S. Correlation of synaptic inputs in the visual cortex of awake, behaving mice. *Neuron* **99**, 1289–1301 (2018).
- [30] Zylberberg, J. & DeWeese, M. R. Sparse coding models can exhibit decreasing sparseness while learning sparse codes for natural images. *PLoS computational biology* **9**, e1003182 (2013).
- [31] Williams, P. E. & Shapley, R. M. A dynamic nonlinearity and spatial phase specificity in macaque v1 neurons. *Journal of Neuroscience* **27**, 5706–5718 (2007).
- [32] Jones, J. P. & Palmer, L. A. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology* **58**, 1233–1258 (1987).
- [33] Ringach, D. L. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of neurophysiology* (2002).
- [34] Ohzawa, I., DeAngelis, G. C. & Freeman, R. D. Encoding of binocular disparity by simple cells in the cat’s visual cortex. *Journal of Neurophysiology* **75**, 1779–1805 (1996).
- [35] DeAngelis, G. C., Ohzawa, I. & Freeman, R. Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. i. general characteristics and postnatal development. *Journal of neurophysiology* **69**, 1091–1117 (1993).
- [36] DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. Receptive-field dynamics in the central visual pathways. *Trends in neurosciences* **18**, 451–458 (1995).
- [37] McLean, J., Raab, S. & Palmer, L. Contribution of linear mechanisms to the specification of local motion by simple cells in areas 17 and 18 of the cat. *Visual neuroscience* **11**, 271–294 (1994).

- [38] Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. Receptive field organization of complex cells in the cat's striate cortex. *The Journal of physiology* **283**, 79–99 (1978).
- [39] Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of physiology* **283**, 53–77 (1978).
- [40] Skottun, B. C. *et al.* Classifying simple and complex cells on the basis of response modulation. *Vision research* **31**, 1078–1086 (1991).
- [41] Mechler, F. & Ringach, D. L. On the classification of simple and complex cells. *Vision research* **42**, 1017–1033 (2002).
- [42] Ringach, D. L., Shapley, R. M. & Hawken, M. J. Orientation selectivity in macaque v1: diversity and laminar dependence. *Journal of neuroscience* **22**, 5639–5651 (2002).
- [43] Van den Bergh, G., Zhang, B., Arckens, L. & Chino, Y. M. Receptive-field properties of v1 and v2 neurons in mice and macaque monkeys. *Journal of Comparative Neurology* **518**, 2051–2070 (2010).
- [44] Rochefort, N. L. *et al.* Development of direction selectivity in mouse cortical neurons. *Neuron* **71**, 425–432 (2011).
- [45] Gouwens, N. W. *et al.* Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature neuroscience* **22**, 1182–1195 (2019).
- [46] Holmgren, C., Harkany, T., Svennenfors, B. & Zilberter, Y. Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *The Journal of physiology* **551**, 139–153 (2003).
- [47] Okun, M. & Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature neuroscience* **11**, 535–537 (2008).
- [48] Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**, 442–446 (2003).
- [49] Anderson, J. S., Carandini, M. & Ferster, D. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *Journal of neurophysiology* **84**, 909–926 (2000).
- [50] Monier, C., Chavane, F., Baudot, P., Graham, L. J. & Frégnac, Y. Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning. *Neuron* **37**, 663–680 (2003).

- [51] Xue, M., Atallah, B. V. & Scanziani, M. Equalizing excitation–inhibition ratios across visual cortical neurons. *Nature* **511**, 596–600 (2014).
- [52] Dichter, M. A. & Ayala, G. Cellular mechanisms of epilepsy: a status report. *Science* **237**, 157–164 (1987).
- [53] Van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
- [54] Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**, 1569–1573 (2011).
- [55] Banerjee, A. *et al.* Jointly reduced inhibition and excitation underlies circuit-wide changes in cortical processing in rett syndrome. *Proceedings of the National Academy of Sciences* **113**, E7287–E7296 (2016).
- [56] Adesnik, H. Layer-specific excitation/inhibition balances during neuronal synchronization in the visual cortex. *The Journal of physiology* **596**, 1639–1657 (2018).
- [57] Levenstein, D. *et al.* On the role of theory and modeling in neuroscience. *arXiv preprint arXiv:2003.13825* (2020).
- [58] Zhu, W., Shelley, M. & Shapley, R. A neuronal network model of primary visual cortex explains spatial frequency selectivity. *Journal of computational neuroscience* **26**, 271–287 (2009).
- [59] McLaughlin, D., Shapley, R., Shelley, M. & Wiesel, D. J. A neuronal network model of macaque primary visual cortex (v1): Orientation selectivity and dynamics in the input layer 4cα. *Proceedings of the National Academy of Sciences* **97**, 8087–8092 (2000).
- [60] Cai, D., Tao, L., Shelley, M. & McLaughlin, D. W. An effective kinetic representation of fluctuation-driven neuronal networks with application to simple and complex cells in visual cortex. *Proceedings of the National Academy of Sciences* **101**, 7757–7762 (2004).
- [61] Chariker, L., Shapley, R., Hawken, M. & Young, L.-S. A computational model of direction selectivity in macaque v1 cortex based on dynamic differences between on and off pathways. *Journal of Neuroscience* **42**, 3365–3380 (2022).
- [62] Attneave, F. Some informational aspects of visual perception. *Psychological review* **61**, 183 (1954).
- [63] Barlow, H. B. *et al.* Possible principles underlying the transformation of sensory messages. *Sensory communication* **1**, 217–233 (1961).

- [64] Field, D. J. What is the goal of sensory coding? *Neural computation* **6**, 559–601 (1994).
- [65] Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
- [66] Willmore, B. D., Mazer, J. A. & Gallant, J. L. Sparse coding in striate and extrastriate visual cortex. *Journal of neurophysiology* **105**, 2907–2919 (2011).
- [67] Comon, P. Independent component analysis, a new concept? *Signal processing* **36**, 287–314 (1994).
- [68] Bell, A. J. & Sejnowski, T. J. The “independent components” of natural scenes are edge filters. *Vision research* **37**, 3327–3338 (1997).
- [69] van Hateren, J. H. & Ruderman, D. L. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 2315–2320 (1998).
- [70] Srinivasan, M. V., Laughlin, S. B. & Dubs, A. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **216**, 427–459 (1982).
- [71] Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* **2**, 79–87 (1999).
- [72] Spratling, M. W. Predictive coding as a model of response properties in cortical area v1. *Journal of neuroscience* **30**, 3531–3543 (2010).
- [73] Lotter, W., Kreiman, G. & Cox, D. A neural network trained for prediction mimics diverse features of biological neurons and perception. *Nature Machine Intelligence* **2**, 210–219 (2020).
- [74] Földiák, P. Learning invariance from transformation sequences. *Neural computation* **3**, 194–200 (1991).
- [75] Kayser, C., Einhäuser, W., Dümmer, O., König, P. & Körding, K. Extracting slow subspaces from natural videos leads to complex cells (2001).
- [76] Wiskott, L. & Sejnowski, T. J. Slow feature analysis: Unsupervised learning of invariances. *Neural computation* **14**, 715–770 (2002).
- [77] Weghenkel, B. & Wiskott, L. Slowness as a proxy for temporal predictability: An empirical comparison. *Neural computation* **30**, 1151–1179 (2018).

- [78] Tishby, N., Pereira, F. C. & Bialek, W. The information bottleneck method. *arXiv preprint physics/0004057* (2000).
- [79] Bialek, W., Van Steveninck, R. R. D. R. & Tishby, N. Efficient representation as a design principle for neural coding and computation (2006).
- [80] Zylberberg, J., Murphy, J. T. & DeWeese, M. R. A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of v1 simple cell receptive fields. *PLoS computational biology* **7**, e1002250 (2011).
- [81] King, P. D., Zylberberg, J. & DeWeese, M. R. Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of v1. *Journal of Neuroscience* **33**, 5475–5485 (2013).
- [82] Taylor, L., Zenke, F., King, A. J. & Harper, N. S. Temporal prediction captures retinal spiking responses across animal species. *bioRxiv* 2024–03 (2024).
- [83] Douglas, R. J. & Martin, K. A functional microcircuit for cat visual cortex. *The Journal of physiology* **440**, 735–769 (1991).
- [84] Douglas, R. J. & Martin, K. A. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27**, 419–451 (2004).
- [85] Bonhoeffer, T. & Grinvald, A. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature* **353**, 429–431 (1991).
- [86] Maldonado, P. E., Godecke, I., Gray, C. M. & Bonhoeffer, T. Orientation selectivity in pinwheel centers in cat striate cortex. *Science* **276**, 1551–1555 (1997).
- [87] Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).
- [88] Cossell, L. *et al.* Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* **518**, 399–403 (2015).
- [89] Auksztulewicz, R., Rajendran, V. G., Peng, F., Schnupp, J. W. H. & Harper, N. S. Omission responses in local field potentials in rat auditory cortex. *BMC biology* **21**, 130 (2023).
- [90] Maunsell, J. H. & Gibson, J. R. Visual response latencies in striate cortex of the macaque monkey. *journal of Neurophysiology* **68**, 1332–1344 (1992).
- [91] Nejad, K. K., Anastasiades, P., Hertag, L. & Costa, R. P. Self-supervised predictive learning accounts for layer-specific cortical observations. *BioRxiv* (2024).

- [92] Millidge, B., Tang, M., Osanlouy, M., Harper, N. S. & Bogacz, R. Predictive coding networks for temporal prediction. *PLOS Computational Biology* **20**, e1011183 (2024).
- [93] Heintz, T. G., Hinojosa, A. J., Dominiak, S. E. & Lagnado, L. Opposite forms of adaptation in mouse visual cortex are controlled by distinct inhibitory microcircuits. *Nature communications* **13**, 1031 (2022).
- [94] Nowak, L. G., Azouz, R., Sanchez-Vives, M. V., Gray, C. M. & McCormick, D. A. Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *Journal of neurophysiology* **89**, 1541–1566 (2003).
- [95] Cardin, J. A., Palmer, L. A. & Contreras, D. Stimulus feature selectivity in excitatory and inhibitory neurons in primary visual cortex. *Journal of Neuroscience* **27**, 10333–10344 (2007).
- [96] Dayan, P. & Abbott, L. F. *Theoretical neuroscience: computational and mathematical modeling of neural systems* (MIT press, 2005).
- [97] Taylor, L., King, A. J. & Harper, N. S. Addressing the speed-accuracy simulation trade-off for adaptive spiking neurons. *arXiv preprint arXiv:2311.11390* (2023).
- [98] Perez-Nieves, N. & Goodman, D. Sparse spiking gradient descent. *Advances in Neural Information Processing Systems* **34**, 11795–11808 (2021).
- [99] Attwell, D. & Laughlin, S. B. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism* **21**, 1133–1145 (2001).
- [100] Lennie, P. The cost of cortical computation. *Current biology* **13**, 493–497 (2003).
- [101] Niven, J. E. Neuronal energy consumption: biophysics, efficiency and evolution. *Current opinion in neurobiology* **41**, 129–135 (2016).
- [102] Howarth, C., Gleeson, P. & Attwell, D. Updated energy budgets for neural computation in the neocortex and cerebellum. *Journal of Cerebral Blood Flow & Metabolism* **32**, 1222–1232 (2012).
- [103] Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)* **1**, 1–47 (1991).
- [104] Young, M. P. & Yamane, S. Sparse population coding of faces in the inferotemporal cortex. *Science* **256**, 1327–1331 (1992).
- [105] Suzuki, M., Pennartz, C. M. & Aru, J. How deep is the brain? the shallow brain hypothesis. *Nature Reviews Neuroscience* 1–14 (2023).

- [106] Van Kleef, J. P., Cloherty, S. L. & Ibbotson, M. R. Complex cell receptive fields: evidence for a hierarchical mechanism. *The Journal of physiology* **588**, 3457–3470 (2010).
- [107] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition (2016).
- [108] Taylor, L. Natural movies (2023). URL https://figshare.com/articles/dataset/Natural_movies/24265498.
- [109] Olshausen, B. A. & Field, D. J. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research* **37**, 3311–3325 (1997).
- [110] Taylor, L. & Nitschke, G. Improving deep learning with generic data augmentation (2018).
- [111] Perez-Nieves, N., Leung, V. C., Dragotti, P. L. & Goodman, D. F. Neural heterogeneity promotes robust learning. *Nature communications* **12**, 5791 (2021).
- [112] Gao, E., DeAngelis, G. C. & Burkhalter, A. Parallel input channels to mouse primary visual cortex. *Journal of Neuroscience* **30**, 5912–5926 (2010).
- [113] Baylor, D., Matthews, G. & Yau, K. Two components of electrical dark noise in toad retinal rod outer segments. *The Journal of physiology* **309**, 591–621 (1980).
- [114] Rusakov, D. A., Savtchenko, L. P. & Latham, P. E. Noisy synaptic conductance: Bug or a feature? *Trends in Neurosciences* **43**, 363–372 (2020).
- [115] Zenke, F. & Ganguli, S. Superspike: Supervised learning in multilayer spiking neural networks. *Neural computation* **30**, 1514–1541 (2018).
- [116] Zenke, F. & Vogels, T. P. The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. *Neural computation* **33**, 899–925 (2021).
- [117] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [118] Gabor, D. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering* **93**, 429–441 (1946).