Organic Convolution in The Ventral Visual Pathway Can Explain the Variety of Shape Tuning in Area V4

Carl Gold[1]

[1]Independent Researcher, Berkeley, CA, USA.

Contributing authors: carl.steven.gold@gmail.com;

Abstract

This modeling study proposes a novel theory for how V4 neurons embody selectivity to the varying degrees of curvature in diverse receptive field arrangements reported in previous recording studies. Cortex is modeled as having neurons with simple, fixed arrangements of connections to topologically organized afferent inputs repeating in a rotational tiling: V1 aggregates points into selectivity for lines and edges at a variety of rotations; V2 aggregates orientated segments into selectivity for corners at a variety of angles and rotations; V4 aggregates corners into curvature selectivity. It is demonstrated that a simple, systematic approach to curvature selectivity can explain a wide variety of past observations. The model is implemented in approximately 900,000 integrate and fire units obeying Dale's law in standard cortical micro-circuits. The simple patterns of afferent connections repeat in topologically organized layers. and a novel method for simulating large numbers of integrate and fire units with tensor programming and GPU hardware is employed. By combining convolution with the time course of post-synaptic potentials (PSP), computation occurs in a feedforward cascade of synchronized spikes. This leaves open the question of connection formation as the model has relatively few parameters which are tuned with the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) plus manual fine tuning. This is sufficient to qualitatively reproduce a wide variety of V4 recordings and demonstrates that a single mechanism of curve selectivity can give rise to the diverse repertoire of behaviors observed in V4. A synthesis of hierarchical and convolutional network models, this study adds novel elements to both: In comparison to previous hierarchical models there is a novel V4 mechanism and a computation based on single spikes using tensor convolution as the computational engine. In comparison to previous convolution models there is a much higher degree of biological realism and a novel unsupervised approach to connection formation.

Keywords: Spiking, Convolution, Neural Network, V2, V4

1 Introduction

1.1 Organic Convolution

This study demonstrates a novel theory for how V4 pyramidal cells implement rapid recognition of curved and straight image patches by simple patterns of connections with afferent inputs from V2. A model based on the theory can reproduce diverse sets of recordings made in V4 [Pasupathy and Connor, 2001, Nandy et al., 2013]: These studies found that there are V4 neurons that are selective to either high, medium or low curvature. There seem to be some neurons selective to a constant degree of curvature in a certain position and orientation

in its receptive field, and other neurons selective to a constant degree of curvature around the entire receptive field with orientation selectivity that rotates with position. Still other V4 neurons are selective to straight or nearly straight contours with an aligned orientation. The inputs to V4 were recently described in [Liu et al., 2016]: V2 contains units with receptive fields that are "complex-shaped" because they combine two Gabor-like receptive fields from V1 input sub-units with each sub-unit at a different orientation. A typical "complex-shaped" V2 unit is selective to a corner where two orientations meet. Curvature selectivity in a V4 unit may be built up from afferent corner selective inputs that have a characteristic range of separation angles and aligned orientations that rotate systematically around the center of the receptive field. The arrangement is akin to the simple arrangement of inputs that leads to motion detection described in [Kim et al., 2014, Borst and Helmstaedter, 2015]: In motion detection afferent inputs, each having a different characteristic delay, connect at different preferred distance along each dendrite of the receiving cell. In this theory for V4 the proposed V4 arrangements of afferent inputs are selective for the form of objects rather than motion, and also have simple yet purposeful arrangements along the dendrites.

If simple connection patterns lead to selectivity in diverse neurons across cortex it relies on the fact that the same connection patterns are repeated by neurons at different locations in topographically organized cortical layers. Convolution is the mathematical model for such a computation: Convolution refers to the mathematical operation common in signal processing in which one function, a filter, is applied to every point in an input signal. (see e.g. [Smith et al., 1997].) This theory for V4 curvature selectivity and the associated model are referred to as Organic Convolution (OC) because it proposes that computation equivalent to convolution can be implemented biologically, and in a realistic biophysical model, not just in V1 but also in V2 and V4.

In OC the primary mechanism for synaptic connection pattern formation is assumed to be a genetically controlled process that guides local connections during healthy development without external stimuli; it is "unsupervised" in the parlance of machine learning. The connection patterns assumed by OC are intended to meet the criteria proposed in [Zador, 2019] in that they are simple enough to be created through mechanisms controlled by the genome. The theory is that simple connection patterns form a foundation for computation and it remains an open question what forms of learning may complete our understanding of cortical processing - biologically plausible learning mechanisms are beyond the scope of this study. As will be detailed in the discussion of background literature, it is assumed that the concerted action of local development rules can enable neurons distributed in a topographically organized layer to have approximately the same patterns of connections. This approach satisfies the primary objection to convolution as a result of a learning process [Bartunov et al., 2018], which is that neurons at diverse locations could not learn the same connection strengths from experience.

All modeling in this study refers to cascades of individual spikes, defined as a computation in which every neuron in every layer fires at most once. Cascades are of interest to biophysicists seeking to explain neural computation for several reasons: As described in [Thorpe and Imbert, 1989, VanRullen et al., 2005], object recognition can be achieved by a single cascade. The fastest possible neural computation, a cascade is an elementary unit of computation from which higher level computations may be built. Computation involving rates or bursts and lateral feedback within a brain area or long range feedback between brain areas take place on longer time scales. These more complex modes of computation must build on an initial feedforward cascade that occurs when a stimulus is first presented. While bursts, rates, lateral feedback and long range feedback may be requirements of a fully realistic model of visual recognition, according to [Thorpe and Imbert, 1989, VanRullen et al., 2005] they are not necessary for the fastest form of visual recognition. For these reasons this study focuses on cascades in the belief that it will be a useful building block in understanding those other more complex phenomena.[1]

[1] Another element necessary for fully realistic spiking computation absent from the model of single cascades is a mechanism of synchronization that enables rapid, successive cascade computations. As will be described in detail in the Results and Methods, the decay time constants of PSP's at

To make a biophysically realistic model based on a convolution calculation, the implementation of OC uses tensor based computation technologies from Deep Learning and Deep Convolutional Neural Networks (DCNN) [LeCun et al., 2015] and adapts it to make a of model individual spikes. The algorithm, named Spike Time Convolution (STC), adds the time course of post synaptic potentials (PSP's) onto a standard DCNN model, as described in detail in the Methods. At the same time, the selectivity of the convolution filters in the V1 and V2 models of this study take the usual form for a Hierarchical Model (HM) of cortex [Fukushima, 1980, Riesenhuber and Poggio, 1999]: V1 includes orientation selective units, there are layers of invariance operations, and V2 units have selectivities combining orientations.

The model in this study is biophysically realistic to an intermediate level of neuronal detail: It includes the time course of post-synaptic potentials at the soma of excitatory and inhibitory model neurons obeying Dale's Law [Dale, 1935] but not details of dendritic compartments or different types of ionic channels. The excitatory and inhibitory units are arranged into standard cortical microcircuits [Douglas and Martin, 2004]. Biophysically realistic time constants are used for the arrival time distribution of spikes in V1, the time course of PSP's, and transmittal latency between layers and areas. These result in simulations of about 30 ms, the time necessary for a cascade of spikes arriving in V1 to propagate from V1 to V4, as detailed in the Results and Methods.

There are also multiple sources of randomness in the model, detailed in the Methods: The probability of spikes arriving in V1 is proportional to the input intensity, the arrival time of spikes in V1 and the spike transmission time between cortical layers and areas are modeled using Gamma distributions, and the size of PSP's are drawn from Normal distributions. However, the model in this study has far fewer neurons than are present in the ventral visual pathway of mammals, and it does not include a model of transmission failure; for these reasons the model units in this study are proposed as models for averages of redundant faulty units in real cortex.

The network architecture is summarized in Figure 1. As detailed in the Methods, the model starts with V1 L4 units (Figure 1C.) which spike with a probability proportional to the input image intensity at times which are randomly, but tightly synchronized as in real cortex [Singer, 1999]. This input layer serves as a lumped model of the visual pathway from the retina to V1 L4. These spikes are the input to the V1 L2/3 pyramidal units (Figure 1 E,F) and also L2/3 inhibitory units (D), which also send their spikes to the L2/3 pyramidal units. The inhibitory units match the excitatory units one to one and each has enough excitatory input to fire shortly after receiving an input spike, its output matching the input with a delay introduced by its own PSP timing properties - these properties are common to the inhibitory units throughout the model.

The L2/3 pyramidal units have convolution filters created to resemble the receptive fields of V1 simple cells [Hubel and Wiesel, 1962, Ferster and Miller, 2000]: excitation juxtaposed with inhibition to form either a band (Figure 1E) or an edge (Figure 1F). The band filters have 8 versions at 8 orientations (every 22.5° up to 180°) and the edge filters at 16 orientations (every 22.5° up to 360°.) Each orientation of edge or band forms a "channel" in the terminology or DCNN: a layer of units having identical input/output properties - in this case, the oriented simple convolution filters. Figure 1T illustrates the preferred stimuli of the V1 L2/3 pyramidal cells.

The spikes from the V1 L2/3 units are the inputs to the V2 L4 units (Figure 1G). In the model, the L4 units perform a downsampling function known as "pooling" in the DCNN or "invariance" in the terminology of HM: These units have 2×2 receptive field and sufficient excitation to fire whenever any input fires. Matching orientations of edge selective and pool selective V1 L2/3 units are combined into eight orientated channels in the V2 L4 model (one channel every 22.5°.)

The output of the V2 L4 units travels to both the V4 L2/3 inhibitory units (Figure 1H) and the V4 L2/3 pyramidal units. The convolution filters for the V4 L2/3 pyramidal units are based on "complex-shaped" units described

the soma are long relative to the time scale of spike integration. It is hypothesized that synchronized inhibition associated with inhibitory networks responsible for the gamma oscillation [Buzsáki and Wang, 2012] act as a "reset" mechanism for repolarization faster than passive decay implied by the time constants. Modeling such mechanisms are also beyond the scope of the present study.

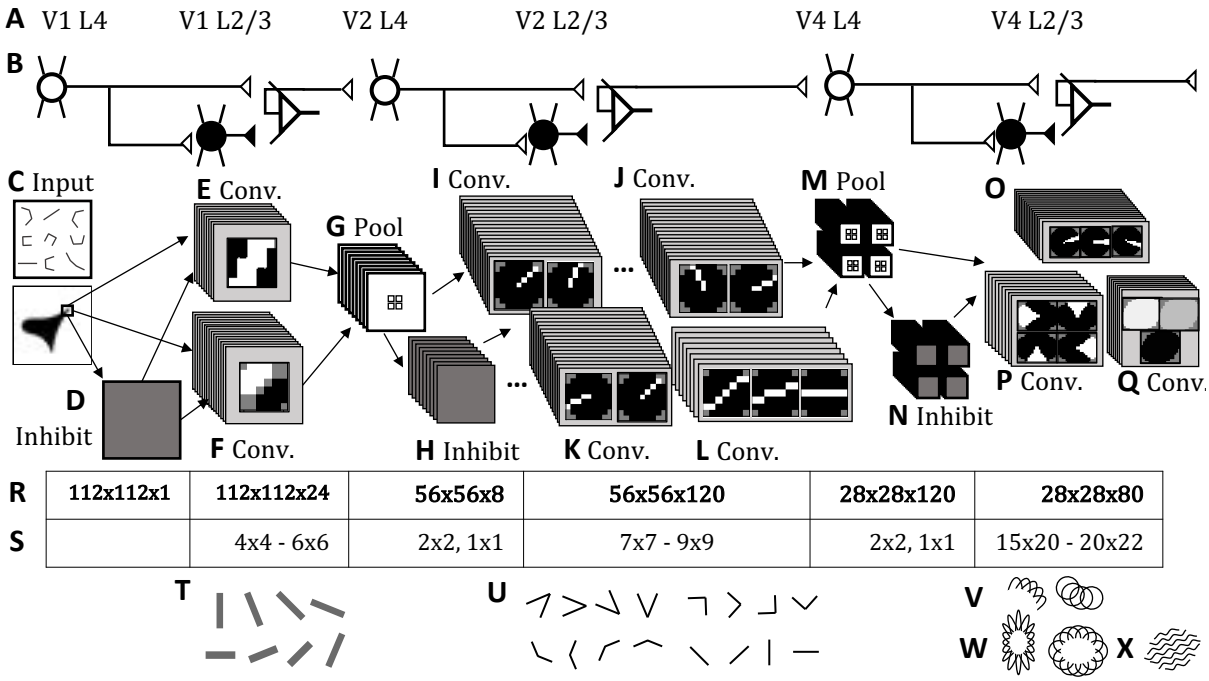

Fig. 1 Organic Convolution Model Architecture - see the accompanying text for a detailed explanation. **A** Labeling of network parts by cortical area. **B** Identification of cell types represented by different parts of the network: Stellate cells in L4, Inhibitory and Pyramidal and cells in L2/3. L2/3 pyramidal cells are modeled with a convolution layer (Conv.) while L4 stellate are modeled with a pooling (Pool) layer. Inhibition are modeled as an inversion layer (Inhibit). **C** Input layer, corresponding to V1 L4 stellate cells. **D** Inhibition of input layer corresponding to V1 L2/3 inhibitory interneurons. **E** V1 L2/3 pyramidal neurons selective to oriented lines represented by eight channels of convolution. **F** V1 L2/3 pyramial neurons selective to oriented edges represented by 16 channels of convolution. **G** The V2 L4 stellate model is 24 channels of pool operations corresponding to the V1 L2/3 channels. **H** Inhibition layer representing V2 L2/3 inhibitory interneurons. **I-L** The V2 L2/3 pyramidal neurons are represented by convolutional channels that combine inputs at a range of angles: acute (I), right (J), obtuse (K), fully straight (L); additional angles are omitted for brevity. **H** The V4 L4 stellate model is 120 channels of pool operations corresponding to the V2 L2/3 channels. **N** Inhibition layer representing V4 L2/3 inhibitory interneurons.**O** V4 L2/3 neurons sensitive to a medium or high degree of curvature over an asymmetric portion of the receptive field represented by 48 channels of convolution. **P** V4 L2/3 neurons sensitive to a medium or high degree of curvature over an a a fully symmetric receptive field represented by 24 channels of convolution. **Q** V4 L2/3 neurons sensitive to low curvature are represented by 8 channels of convolution. **R** The size and number of channels for the layers above. Note that L4 stellate pooling layers and L2/3 inhibition layers share common dimensions and numbers of channels and are presented together. **S** The sizes of the convolution filters for the layers above. **T** Preferred stimuli for the V1 L2/3 units: short, oriented images segments. **U** Preferred stimuli for V2 L2/3 units: small corners at a range of angles, from acute to fully straight lines. **V** Preferred stimuli for asymmetric V4 medium and high curvature selective units. **W** Preferred stimuli for fully symmetric V4 medium and high curvature selective units. **X** Preferred stimuli of the V4 low curvature selective units.

[Liu et al., 2016] in that there are units that combine two orientations at a particular angle (Figure 1I-K). In the model there are 6 separation angles, every 22.5° from 45° to 180° (45°, 67.5°, 90°, 112.5°, 135°, 157.5°.). There are also units that have selectivity to long straight line segments, "ultra-long" units in the terminology of [Liu et al., 2016]. The corner selective units have 16 channels per separation angle, with the pattern rotated by 22.5° in each channel. The units selective to extended straight inputs have 8 rotations rotated by 22.5°. Figure 1U illustrates the preferred stimuli of the V2 L2/3 pyramidal cells. This definition of V2 selectivity is substantially similar to that proposed in [Riesenhuber and Poggio, 1999], albeit in a spiking convolutional form.

The models for V4 L4 (Figure 1 M) and V4 L2/3 inhibitory units (Figure 1 N) are the same as those for V2 L4 and V2 L2/3 inhibition: The V4 L4 units pool and downsample the V2 L2/3 outputs. In the V4 model each separation angle

and orientation of V2 L2/3 unit has its own separate channel of units performing pooling and inhibition.

The V4 L2/3 pyramidal cell model uses convolution filters that embody a novel proposal for how V4 units may create curvature selectivity from V2 afferents (Figure 1O-Q): Selectivity to curvature of a certain degree is formed by pooling afferent inputs having a degree of curvature in a particular range: High curvature corresponds to pooling inputs having selectivity to a separating angle less than 90°; medium curvature selectivity corresponds to pooling afferent units with selectivity to a separating angle from 90° to 135°; low curvature selectivity corresponds to pooling 158° selective units along with straight selective units (Figure 1Q). To achieve curvature selectivity, only similarly oriented V2 corner selective units are pooled in any part of the V4 L2/3 units receptive field. Importantly, the selected orientation of afferent V2 inputs shows a systematic rotation around the receptive field of V4 L2/3 pyramidal neurons - this is more apparent in V4 units with curvature selectivity throughout their receptive field (Figure 1P), and also matches the results for V4 units with curvature selectivity in a particular portion of the receptive field (Figure 1O). Illustrations of the preferred stimuli of the V4 L2/3 pyramidal cells with curvature selectivity in a particular region of the receptive field are shown in Figure 1 V; the preferred stimuli of V4 L2/3 pyramidal cells with curvature selectivity that is symmetric around the receptive field are shown in Figure 1 W; and the preferred stimuli of V4 L2/3 pyramidal cells with selectivity for straight or mostly straight stimuli is shown in Figure 1 X.

The remaining sections of the Introduction provide further details about the prior research that was introduced briefly in this section. The Results section shows the reproduction of the recordings by the model, provides further details about the V4 unit connection patterns that reproduce the recordings, presents data on the propagation and timing of spike cascades in the network, and demonstrates predictions that the model makes about novel stimuli. The Methods section provides details about the spiking neuron model, the Spike Time Convolution computation, the constants of the biophysical model, and the construction and tuning of the convolution filters. The Appendix contains illustrations of the propagation of spikes in every part of the network and details of methods omitted from the main Methods section for brevity.

1.2 Visual Area V4

Visual object identification is processed in the ventral visual pathway through primary visual areas V1 and V2, and an intermediate area V4, before object recognition is processed in area IT, or inferior temporal cortex (IT) [Tsao et al., 2006, Kobatake and Tanaka, 1994]. It has been known for many years that there are neurons in area V1 that encode local orientation [Hubel and Wiesel, 1962, Ferster and Miller, 2000].

More recent studies [Liu et al., 2016] have found that V2 contains neurons in which V1 receptive fields converge into two main types: The first type of V2 selectivity is described as "ultra-long" (aligned) which combines afferent inputs from orientation selective V1 neurons that all share the same orientation, effectively extending the short oriented selective patches into selectivity for longer straight contours in V2. The second type of selectivity seen in V2 neurons is "complex-shaped" which combines multiple oriented sub-unit at an angle, forming a corner where two orientations meet.

The functional properties of V4 are less well understood: studies suggest that area V4 occupies an intermediate level in the hierarchy of visual processing, responding to moderately complex stimuli, but there is not consensus on the precise nature of V4's processing [Roe et al., 2012]. In this last two decades, multiple streams of evidence suggest that V4 neurons are selective to different degrees of curvature in a variety of configurations. [Pasupathy and Connor, 2001, Nandy et al., 2013, Hu et al., 2020]

Two examples of stimuli used in studies recording from individual V4 neurons are shown in Figure 2. The stimuli from [Pasupathy and Connor, 2001] are shown in Figure 2 A: The stimuli consist of systematic combinations of convex and concave boundary elements, yielding 366 total stimuli. [Pasupathy and Connor, 2001] found that individual cells showed consistency to preferring one type of boundary element, for example an

acute convex curvature, in a specific position. (See Figures 3 and 7) Later studies using similar methods found that some V4 neuron seem to prefer texture [El-Shamayleh and Pasupathy, 2016]. More recently, it was found that a given convex or concave shape preference tends to be selective for similar stimuli with a range of sizes [Kim et al., 2019].

The stimuli from [Nandy et al., 2013] are shown in Figure 2B: the stimuli consisted of oriented bars combined to form approximate curves ranging from straight to tightly curved at a range of orientations. These elements were sized to be smaller than the receptive field of a single V4 neuron and presented at multiple locations throughout the receptive field. This approach allowed the detailed investigation of shape tuning throughout the receptive field. [Nandy et al., 2013] found that the V4 neurons responsive to these stimuli can be categorized into those selective for low curvature, medium curvature and high curvature. Only low curvature selective V4 neurons were found to have spatially invariant shape tuning: V4 neurons selective to medium and high curvatures had spatially variable tuning. [Nandy et al., 2013] showed units that were selective to a consistent degree of curvature but at varying orientations throughout the receptive field. (See Figures 5, 9 and 11.)

Recently [Srinath et al., 2021] measured V4 neuron responses with a dataset generated on the fly by a genetic algorithm designed to find shapes that had the highest neuronal response. This dataset was more complex than those of [Pasupathy and Connor, 2001] and [Nandy et al., 2013] in that it included not only flat imagines (against dark and light backgrounds), but also three dimensional shapes. However, the stimuli of [Srinath et al., 2021] did not include lines or outlines of shapes. [Srinath et al., 2021] found that many V4 neurons seem to be selective to three dimensional shapes more than a flat projection of a similar shape. This suggests that the form of complex objects is already represented to some degree in V4 and not only in IT. However, [Srinath et al., 2021] did not characterize the shape selectivity of V4 neurons by categories like the acuteness or orientation of the selected shapes. Applying the model of this study to the data of [Srinath et al., 2021] is an attractive area for future research.

Another recent study supports the notion of orientation and curvature tuning in V4: [Hu et al., 2020] used intrinsic optical imaging of V4 and found a pattern of curvature domains that were tuned to respond to either straight, medium curvature or high curvature. In [Hu et al., 2020] the stimuli consisted of conventional straight line gratings along with gratings having different degrees of curvature, all presented at different orientations. The results in [Hu et al., 2020] are complimentary to the results in [Nandy et al., 2013] and [Pasupathy and Connor, 2001], suggesting that neurons sensitive to different degrees of curvature are in fact central to the functioning of V4.

1.3 Hierarchical Models of Visual Cortex

Hierarchical models (HM) of visual cortex have a long history of explaining how cells that respond to complex visual stimuli may be built from cells that respond to simple stimuli. The most well known early example was the Neocognitron [Fukushima, 1980] which used an alternation between feature extracting cells, modeled after V1 simple cells, and cells modeled after complex units that pool together the input from multiple feature recognizing cells. The feature selective units in the first layer of this and later models correspond in some form to orientation selective units first identified in [Hubel and Wiesel, 1962]. The idea was further developed in [Perrett and Oram, 1993] which extended the idea to achieve invariance over a wider range of visual features. [Wallis and Rolls, 1997] used a four level network named Visnet and showed that a Hebbian inspired learning rule could train units to have invariance to the position and size of faces and other simple stimuli.

Another example of a mathematical model using alternating selectivity and invariance operations was the HMAX model [Riesenhuber and Poggio, 1999]. The HMAX model had two layers of simple (feature recognizing) and complex (feature pooling) cells, followed by a learned layer of view tuned cells. The view tuned cells were trained to have invariance to moderately complex shapes. The HMAX model was later extended specifically to match the V4 shape tuning data of [Pasupathy and Connor, 2001] in

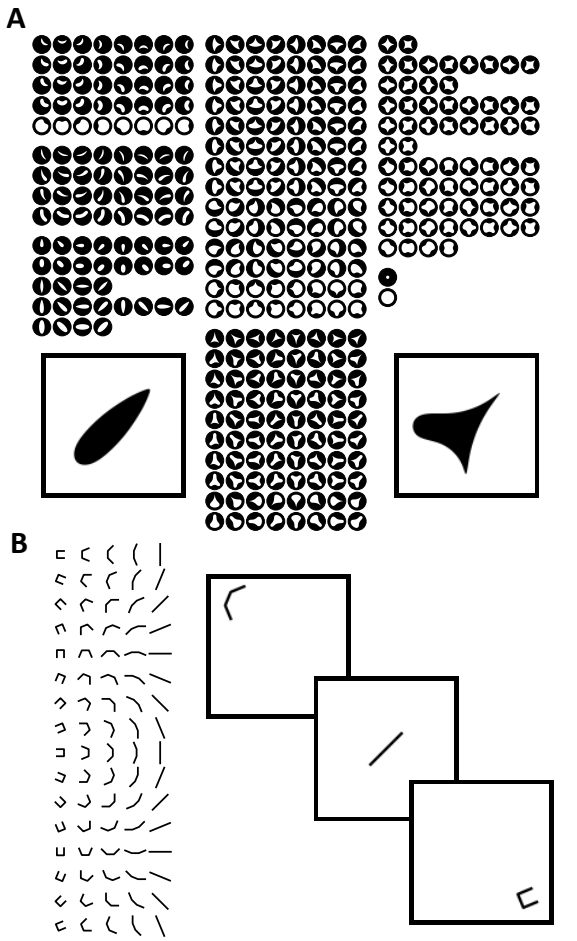

Fig. 2 Stimuli used to probe the response of V4 neurons. **A** [Pasupathy and Connor, 2001] used systemtatic configurations of convex and concave boundary elements included sharp convex angles, medium and broad convex curves, and medium and broad concave curves. Each stimulus is represented by a white icon positioned within a black disk that represents the cell's RF. The stimuli are arranged into blocks according to the number and configuration of convex projections. **B** [Nandy et al., 2013] used composites of the 3 bars at five different conjunction angles (0°, 22.5°, 45°, 67.5° and 90°). The five conjunction angles are presented at 16 orientations for a total of 72 unique stimuli. The composite shapes are presented on a grid that spans the receptive field. Insets: Examples of stimuli as presented.

[Cadieu et al., 2007]: this version of the model includes orientation selective units at a variety of orientations and scales to represent V1, followed by a pooling layer creating invariance. Another layer of features selective to combinations of orientations represents V2, and a second layer of invariance generating combinations is identified with V4. In [Cadieu et al., 2007], the inputs to the V4 combination layer were fit using a selection algorithm to match the data from [Pasupathy and Connor, 2001].

In comparison to [Cadieu et al., 2007], the most similar prior work, the Organic Convolution model replaces learning of the V4 selectivity details with an unsupervised mechanism that assumes a large number of generic V4 selectivities for different curvatures and rotations are present. Specifically, the results in this study suggest that an important unsupervised connection pattern in creating curvature selectivity in V4 is that the V2 afferents orientations preferred by the V4 neuron rotates with the topographical position around a central position in the receptive field. This model also embodies a novel biologically plausible cascade spike computation and a simulation implemented using tensor computational tools.

The HM's described here use a variety of mathematical descriptions for the activity of the neuronal representations. The models are generally interpreted as representing neuronal firing rates in response to a constant stimuli. From [Fukushima, 1980] to[Cadieu et al., 2007], the typical approach is some form of normalized dot product to represent the activity of a feature selective units, and then a non-linearity such as a hinge or a sigmoid function to determine the resulting activity level (firing rate) of the unit. None of the HM's surveyed here used a representation of individual spikes, although it has been shown that object recognition in humans and animals is too fast to rely on a rate code [Thorpe and Imbert, 1989]: rapid recognition happens so fast that there is only time for a single spike per layer/brain area. This means a single cascade is sufficient to perform rapid recognition. The issue of recognition speed and implications for spiking computation was also described in [Wallis and Rolls, 1997].

1.4 Deep Neural Networks

In the last decade deep convolution neural networks (DCNN's) [LeCun et al., 2015] were developed by artificial intelligence researchers that dramatically improved performance in visual recognition tasks. In fact, the idea of convolution in a neural network is inspired by the same observations about cortex as HM's : First, there is

the observation that selectivity to simple oriented stimuli in earlier layers seem to repeat more or less interchangeably over positions in topographically organized visual space. That makes it well represented by convolution, and even HM's have used convolution to implement their selectivity schemes [Fukushima, 1980]. Also the alternation of filter operations with non-linear pooling operations is taken from the idea of complex cells which became "pooling" layers in DCNN [Carandini et al., 2005]. However, modern DCNN use stochastic gradient descent training algorithms broadly known as deep learning (DL) to determine the selective unit properties rather than design them to match neurophysiological data.

There has been considerable interest among computational neuroscientists in whether DCNN and DL can be the basis for explaining the object recognition pathway [Yamins et al., 2014, Kietzmann et al., 2019]. However, there are questions as to whether DL models can serve as useful models for neuroscience: DL requires backpropagation of error signals [Rumelhart et al., 1986] carrying gradient information with the same weight as the feedforward signal, but such signals have not been observed [Bengio et al., 2015]; as a result DL in the form originally proposed is biologically implausible. It has also been noted, even by DL researchers, that no learning system can guarantee that the same weights will be learned by neurons at different locations [Bartunov et al., 2018] raising questions about the plausibility of the convolution model. Like hierarchical cortical models, DL models do not model spikes but instead rely on non-linear transfer functions, which are described as representing firing rates. DL models also ignore the fact that real neurons are constrained by their type to have entirely excitatory or entirely inhibitory influence on efferent connections [Cornford et al., 2020]. At the same time, there have recently been some serious attempts to investigate whether biological learning mechanisms like spike timing dependent plasticity (STDP) in a rate coded model can implement biological deep learning [Bengio et al., 2017, Kheradpisheh et al., 2018].

A few studies have attempted to use DL techniques to understand neurophysiological data from V4: [Pospisil et al., 2018] examined a state of the art DL model and found units that respond to the shapes used in [Pasupathy and Connor, 2001]. [Yamins et al., 2014] proposed a goal driven approach and trained a DL model to match the response of IT neurons; the network trained to match IT also matched V4 response properties. These studies are described by their authors as "top-down" in the sense that they create models to perform the functions of higher cortical areas and then try inspect the resulting learned connections to gain insight into lower cortical areas. [Cadena et al., 2019] also took the goal driven approach by analyzing a pre-trained DCNN trained with DL and added output non-linearities and noise to the unit activations to come up with spike counts. They found that intermediate convolution layers in the network best match V1, but did not consider visual areas beyond V1.

Despite high performance on AI vision benchmarks, there are several ways in which DL models are a poor match for visual processing in the brain: DL networks seem to rerquire much more data to learn concepts than biological brains, DL networks suffers from so-called "adversarial" attacks (that degrade performance with almost imperceptible changes in an image) and failure to generalize that are inconsistent with human behavior[Serre, 2019]. [Jacob et al., 2021] found that DCNN trained with DL can reproduce some qualitative properties of natural vision (e.g. Thatcher effect), but others were absent (e.g. surface invariance). [Feather et al., 2023] looked at what images are seen as invariant to humans and DL networks (termed metamers) and found distinctly different patterns of invariance in DL and biological networks.

1.5 Learning vs. Structure

A more meta problem with not only DL but also HM's is the assumption that the neural network is initially in a random state and useful behavior is learned from experience. While the requirement to "learn" connections through an active process has long been assumed in neural modeling, [Zador, 2019] argues that the competitive demands on newborns require brains to be "initialized" to perform significant functions. While true in mammals, this is even more the case in non-mammalian species with complex behavioral repertoires. [Zador, 2019] concludes that there

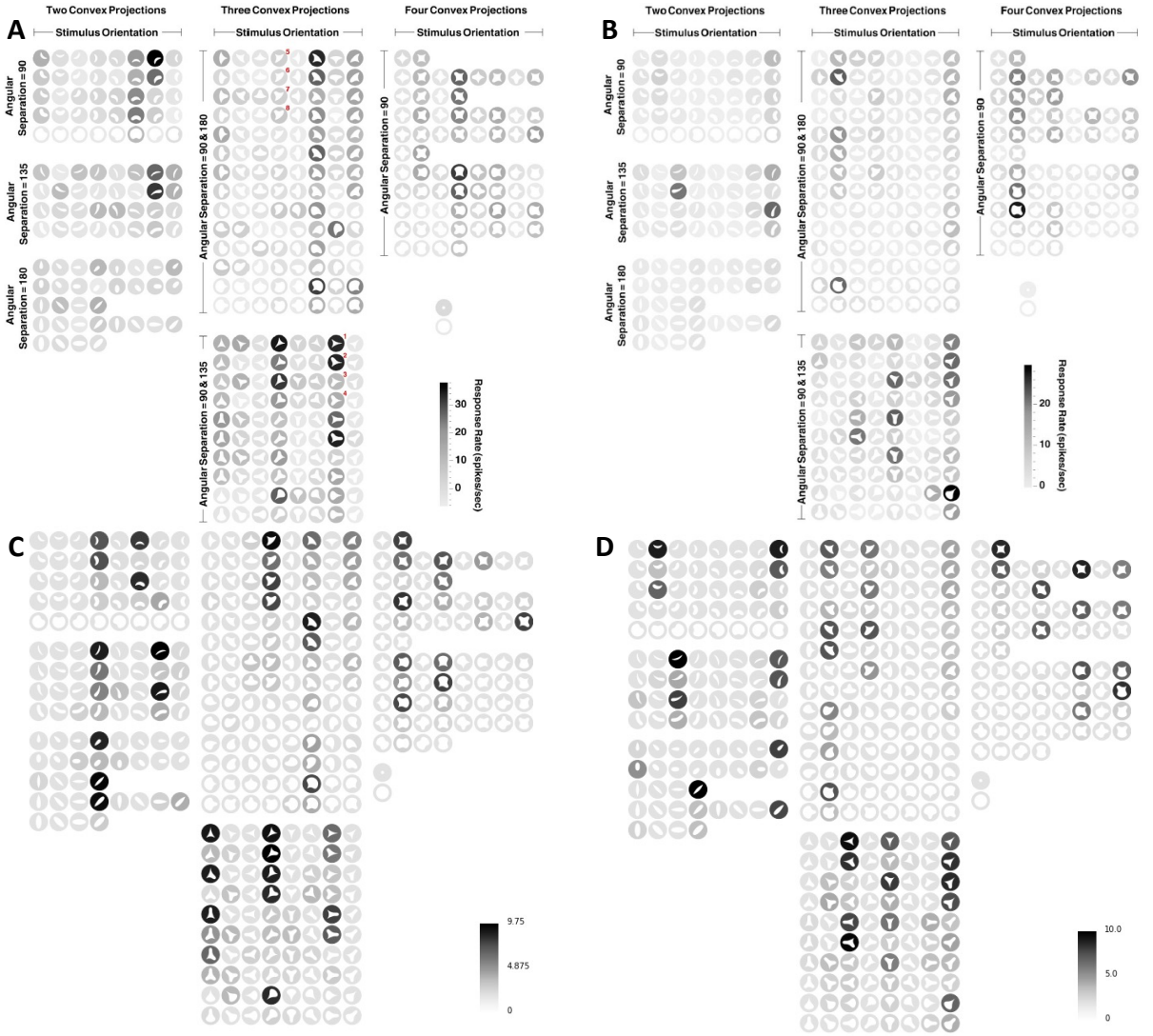

Fig. 3 Recording and simulation of Acute convex asymmetric selectivity in V4: The white shape in each location represents the the stimuli shape, which is actually presented as a black silhouette on a white background. The shading of the disc around the shape represents the intensity of the response as indicated by the legend. **A** Recording reproduced from [Pasupathy and Connor, 2001] showing a neuron tuned for acute convex curvature at the bottom left. **B** Recording reproduced from [Pasupathy and Connor, 2001] showing a neuron tuned for acute convex curvature at the top right. **C** Result of a model simulation qualitatively matching **A**. The results shown are the number of spikes out of 10 simulations. **D** Result of a model qualitatively matching **B**. (TODO: More trials for final version.)

must be an efficient encoding for the generation of such structure in the genome because the genome is not large enough to specify individual connections in complex brains.

Studies have already demonstrated this kind of structure : For example, pyramidal cells in mouse V1 display orientation selectivity at the time of *first eye opening* [Ko et al., 2013]. If orientation selectivity is present at that time its development must have been largely innate, or "unsupervised" in the parlance of machine learning. Another example is that both mammalian [Kim et al., 2014] and fly [Borst and Helmstaedter, 2015] visual systems detect motion by having specific sub-types of bipolar cells, each having a different characteristic delay, connect at different preferred distance along each dendrite of the receiving amacrine

cell. This is an example of a "pre-programmed" computation defined by simple arrangements of connections from different types of inputs at different distances from the soma. A variety of pathways and signals control the shape of axonal structures and even the specific connections between neurons of different types[Mueller, 1999] as well as the specific subtypes making up topographically mapped neuronal structures [Shirasaki and Pfaff, 2002].

2 Results

2.1 Acute Convex Selectivity

2.1.1 Composite Boundary Stimuli

[Pasupathy and Connor, 2001] showed that some V4 neurons are selective to acute convex curvature at different positions relative to the center of the receptive field; Figure 3A and 3B are reproduced from [Pasupathy and Connor, 2001]. The stimuli were previous shown in Figure 2A and consist of 366 combinations of convex and concave boundary elements. While the actual stimuli are presented as a silhouette against a light background, in Figure 3 the shape is shown in white and the shading represents the intensity of the neuron response as detailed in the legend. [Pasupathy and Connor, 2001] described the neuron in Figure 3A as "V4 neuron tuned for acute convex curvature at the bottom left" based on the consistent element among all of the stimuli to which the unit has a high response. Similarly, the neuron in Figure 3B is described as a "V4 neuron tuned for acute convex curvature at the top right" on the basis of the fairly obvious common elements among all of those stimuli to which the neuron responds.

Figure 3C and 3D are results for simulated V4 neurons with similar response properties to Figure 3A and 3B respectively. The simulated neuron in Figure 3C is qualitatively similar to Figure 3A in that both consistently respond to shapes with an acute convex curvature located in the bottom left. The simulation does not perfectly match the intensity, because the model at this stage was manually tuned based on the published image and without the data used to produce Figure 3A and 3B, as described in the Methods. Even so the model matches the primary feature of the recording; the main weakness of the model is in being a bit more consistent and firing to some acute convex curvatures that the neuron missed. The same qualities also apply to the simulated recreation of the recording in Figure 3B with the simulation in 3D which is the neuron tuned for acute convex curvature in the top right.

The connection pattern resulting in the selectivity of Figure 3B are shown in Figure 4. The figure shows convolution filters which implement the synaptic weights to the topographically arranged inputs, with the center of the unit receptive field represented at the center of the filters. The convolution filters are the model embodiment of the connection patterns that give rise to selectivity for acute convexity in a particular position. For further explanation of how connection patterns to topographically organized afferents manifest and how they are represented by the convolution filters in Figure 4, see the Supplementary Methods Figure B4 and the accompanying discussion.

The afferent inputs relevant to the V4 L2/3 neurons selective to acute convex curvature are represented by pictograms Figure 4A,C, E and G: These are the outputs of V2 neurons selective to corners having interior angles 45° (A,C) and 67.5° (E,G). The corresponding convolution filter weights are shown below each pictogram, Figure 4 B,D, F and H.The actual inputs to V4 L2/3 have passed through the V4 L4 neurons which reduced the dimension of the representation by two, and the V4 inhibitory neurons which provide the inhibitory version of the same selectivity slightly delayed. So areas shown as excitatory in Figure 4 are actually connections to the V4 L4 neurons, and those that are inhibitory are connections to the V4 inhibitory neurons. As detailed in the methods, these are actually implemented as two separate channels and two separate convolution filters but for simplicity and conciseness the strength of input connections are shown together and described by the selectivity of the V2 L2/3 neuron from which the afferents inputs originate.

The pattern of excitatory connections is simple: convex angles selective inputs on the bottom left are strongly excitatory and to an equal degree; all other angles are inhibited with a weaker and more diffuse strength. The primary connection is to convex 45° and 67.5° interior angle corners pointing at a 225° angle clockwise from horizontal. At the same time, there are connections to corners

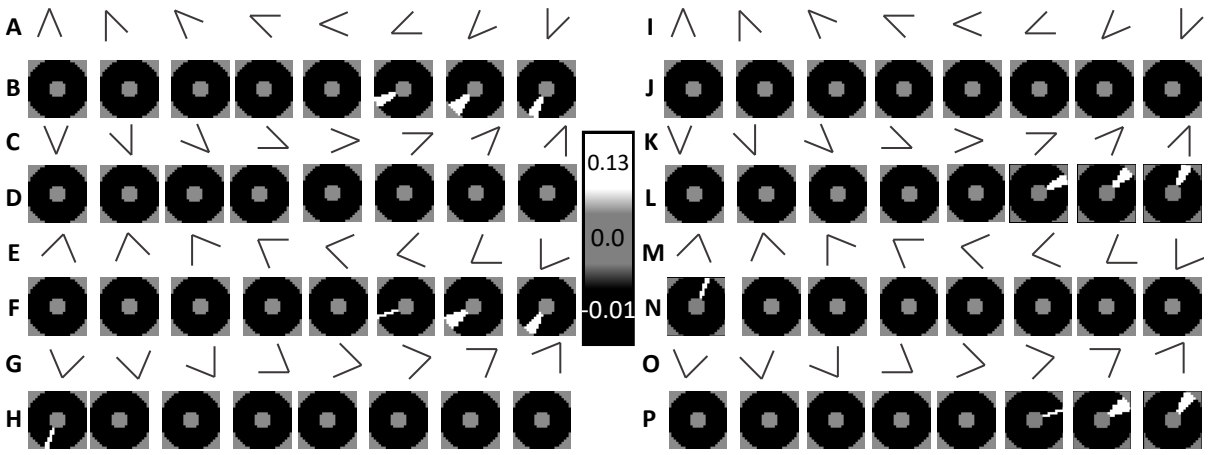

Fig. 4 Afferent Connections for model V4 L2/3 neurons selective to high curvatures in asymmetric zones. **A - H**: Connections for the model V4 neuron selective to acute convex curvature at the lower left. **I - P**: Connections for the model V4 neuron selective to acute convex curvature at the top right. **A & C**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 45° interior angle at a complete range of rotations every 22.5°. Details of V2 neuron selectivity are described in the Supplementary Methods B.3. **E & G**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 67.5° interior angle at a complete range of rotations every 22.5°. **B, D, F, H**: The synaptic connection strength versus position in topographic space relative to the soma at the center for a V4 neuron selective to acute convex curvature at the bottom left. Each square is a single 9×9 convolution filter implementing the weights for the V2 afferent inputs indicated by the pictogram above it. White=excitatory, Black=Inhibitory, Gray=No connection. All of the weights in B-G together are the connections to the topographically arranged inputs for one V4 neuron. The selectivity is due to excitatory inputs on afferents from corner selective V2 units that are convex and in the bottom left portion of the receptive field. There are also excitatory inputs from slightly rotated orientations of the V2 unit, at positions rotated around the receptive field so that the preferred stimuli remain convex. Other connections to afferent inputs are weakly inhibitory. **I, K, M, O**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 as in A/C/E/G. **J, L, N, P**: The synaptic connection strength versus position in topographic space relative to the soma at the center for a V4 neuron selective to acute convex curvature at the top right as in B/D/F/H. There are excitatory inputs on afferents from corner selective V2 units that are convex and in the top right portion of the receptive field. There are also excitatory inputs from slightly rotated orientations of the V2 unit, at positions rotated around the receptive field so that the preferred stimuli remain convex. Other weights on afferent inputs are weakly inhibitory.

angled 202.5° and 247.5° - these connections are in strips that rotate around the center of the receptive field so that the afferent inputs always have a selectivity that is convex. The center of the receptive field lacks excitatory connections - excitatory connections near the center of the V4 L2/3 pyramidal model neurons lead to poor matches with almost all the V4 recordings and this rule was followed to varying degrees in all of the models. This pattern of afferent connection is meant to represent a simplified or average pattern that might be seen in many neurons, each of which would have subtly different selectivity resulting from its own particular dendritic structure and modifications of synaptic weights acquired from experience. The connection pattern for acute convex curvature at the top right is shown in Figure 3D and is qualitatively similar to the neuron tuned for acute convex curvature at the bottom left. The overall preferred stimuli set for the pattern were illustrated in Figure 1V: it is a set of high curvature overlapping curves in a particular region of the receptive field.

2.1.2 Grid of Composite Bar Stimuli

[Nandy et al., 2013] also found that some neurons are selective for high degrees of curvature, and also that some units were selective to high degree of curvature throughout the receptive field with an orientation preference varied with position. Figure 5A and B are reproduced from [Nandy et al., 2013] and represent the output of a V4 L2/3 neuron that is selective to a high degree of curvature throughout its receptive field. Figure 5A shows the intensity of the response at each location in the grid and the numbered positions identify the highest response; the numbers index the position from top left to bottom right. The

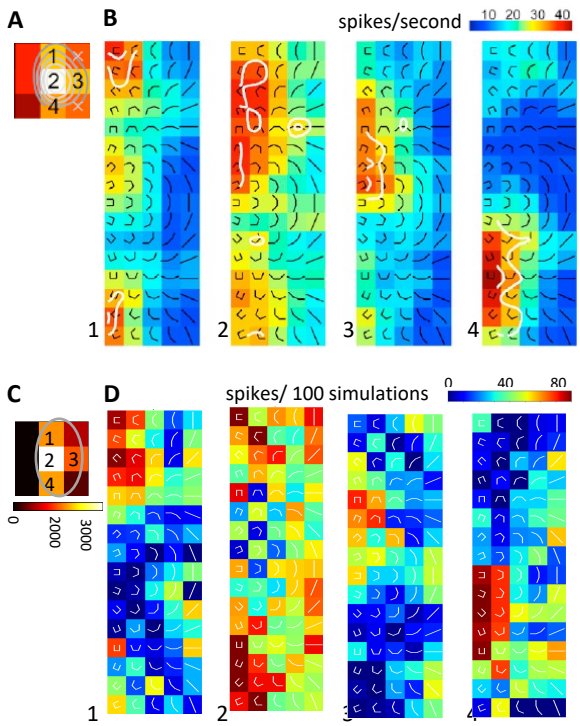

Fig. 5 High curvature selective unit recording (top) and simulation (bottom) **A** Intensity of response at 9 positions in the stimuli grid. The numbers index the four positions with the highest response. The shading indicates the overall response intensity. **B** Shape tuning for recording, 1-4 show the shape tuning at the indexed positions in the receptive field. Each pictogram represents a composite bar stimuli that varies by curvature (x-axis) and orientation (y-axis). The color coded scale shows the response firing rate. **C** Receptive field for simulation. The approximate location of the simulated neurons dendritic tree in the topographic visual field is shown. **D** Shape tuning for simulation, as in B, at positions in the simulated receptive field (C).

color in 5A represents the overall intensity of the response. Figure 5B details the response at each of the four indexed positions from Figure 5A: Each set of icons represents one position, and each icon represents a single stimuli. The color shows the intensity of the response as indicated by the legend.

For this neuron, the stimuli locations triggering the highest response lie along a vertical strip of three locations (Fig 5B1, 5B2 and 5B4), with one additional high response receptive field location on the right side of the main strip (Fig 5B3). The neuron responds primarily to highly curved stimuli, with broad tuning that includes moderate responses to low curvature and straight lines, especially for the most responsive stimuli locations (Figure 5B2). The orientation of the preferred stimuli varies by stimuli location. The response tuning curve at each stimuli location is broad but it peaks around curves oriented convexly towards the center of the receptive field: Location 1 (top) is more strongly tuned to curves open downward, location 2 (center) is tuned strongly for both upward and downward facing curves, location 3 (on the right) is responsive to left opening curves and location 4 (on the bottom) is responsive to upward facing curves. But the convex orientation of the most preferred stimuli does not appear consistent because of the broad tuning of the response.

The output of a V4 L2/3 model neuron with similar high curvature selective response is shown in Figure 5C and Figure 5D. The model response is qualitatively similar to that of the recording in the preference for high curvatures and it approximately matches the orientation selectivity of the recording as well. The main mismatch between the model and recording is that the real neuron shows broader tuning to orientations with high curvature. Also the model neuron has somewhat higher responses to straight line stimuli. The model at this stage was manually tuned based on the published image without the data used to produce Figure 5A and 5B, as described in the Methods.

The connection pattern to afferent inputs for for the [Nandy et al., 2013] high curvature selective V4 L2/3 cell model is shown in Figure 6. The afferent inputs to the V4 L2/3 neuron are represented by pictograms Figure 6 A,C, E, G, I and K: These are the outputs of V2 neurons selective to corners having interior angles 67.5° (A,C), 90° (E,G), and 112.5° (I,K). The corresponding convolution filter weights are shown below each pictogram, Figure 6 B,D, F, H, J and L. For a detailed explanation, see the text accompanying Figure 4 in Section 2.1.1. The pattern is that for every type of afferent input, half of the receptive field is excitatory and half is inhibitory (as small region at the center is neutral). Afferent inputs that are approximately convex to the center of the receptive field are excitatory, and afferent inputs that are approximately concave to the center of the receptive field are inhibited. To achieve this, the excitatory/inhibitory boundary rotates around the receptive field, depending on the orientation of the afferent inputs. The boundary is not

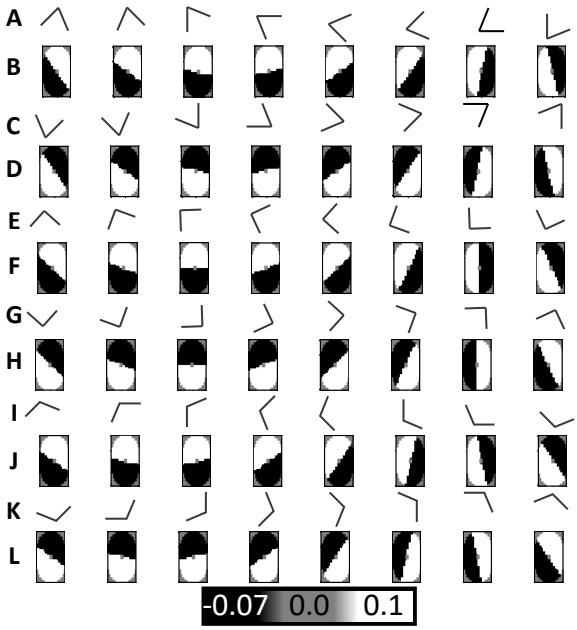

Fig. 6 Afferent Connections for model V4 L2/3 neurons selective to mostly convex high curvatures throughout the receptive field. **A & C**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 67.6° interior angle at a complete range of rotations every 22.5°. Details of V2 neuron selectivity are described in the Supplementary Methods B.3. **E & G**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 90° interior angle at a complete range of rotations every 22.5°. **I & K**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 112.5° interior angle at a complete range of rotations every 22.5°. **B, D, F, H, J, L**: The synaptic connection strength versus position in topographic space relative to the soma at the center for a V4 neuron selective to high curvature throughout the receptive field. Each square is a single 16×28 convolution filter implementing the weights for the V2 afferent inputs indicated by the pictogram above. White=excitatory, Black=Inhibitory, Gray=No connection. All of the weights in B-L together are the connections to the topographically arranged inputs for one V4 neuron. There are excitatory connections to every orientation of the V2 afferents, and the position of the excitation rotates around the receptive field with the orientation of the inputs so that convexly oriented inputs are excited and other inputs are inhibited.

exactly even - to better match the recording, there is a small rotational offset from perfectly convex orientation of the excitatory inputs. Non-preferred orientations are inhibited and afferent inputs with higher interior angle complex shape preferences are ignored. The overall preferred stimuli set for the pattern were illustrated in Figure 1W: it is a set of high curvature overlapping curves all around the receptive field.

2.2 Medium Curvature Selectivity

2.2.1 Composite Boundary Stimuli

[Pasupathy and Connor, 2001] found that some V4 neurons are selective to medium degree curvature at different positions relative to the center of the receptive field, and that while some units were selective to curvature that was convex to the center of the receptive field others were selective for curvature that is concave to the center of the receptive field. Figure 7A and 7B are reproduced from [Pasupathy and Connor, 2001] and each shows the response of a different identified unit to the boundary based stimuli set. See the beginning of section 2.1.1 for a detailed description. Figure 7A represents the recorded selectivity of a neuron described as a "V4 neuron tuned for concave curvature at the right" which describes the common element among those composite boundaries to which it responds. Figure 7B represents the recorded selectivity of a neuron described as "V4 neuron tuned for broad convex curvature at the top.". The corresponding model simulations reproducing these recordings are shown in Figures 7C and 7D. The model simulation matches the selectivities of the real neurons qualitatively. For the concave curvature selective unit (C), the selectivity is very close but the magnitude of responses are somewhat unequal.[1] For the convex curvature selective unit, the model response is stronger at a broader range of stimuli with medium convex curvature at the top and has somewhat less response to other less related stimuli. The model at this stage was manually tuned based on the published image and without the data used to produce Figure 7A and 7B, as described in the Methods.

The convolution filters resulting in the selectivity of Figures 7C and 7D are shown in Figure 8. The convolution filters are the model embodiment of the connection patterns that give rise to selectivity for medium curvature in a particular

[1] Note that the position of the stimuli with Four Convex Projections is not the same same in all of the Figures published in [Pasupathy and Connor, 2001]. This study adopts the most common configuration, not the specific arrangement in every original figure and as a result some of the stimuli positions in Figure 7 do not match between the recording and simulation.

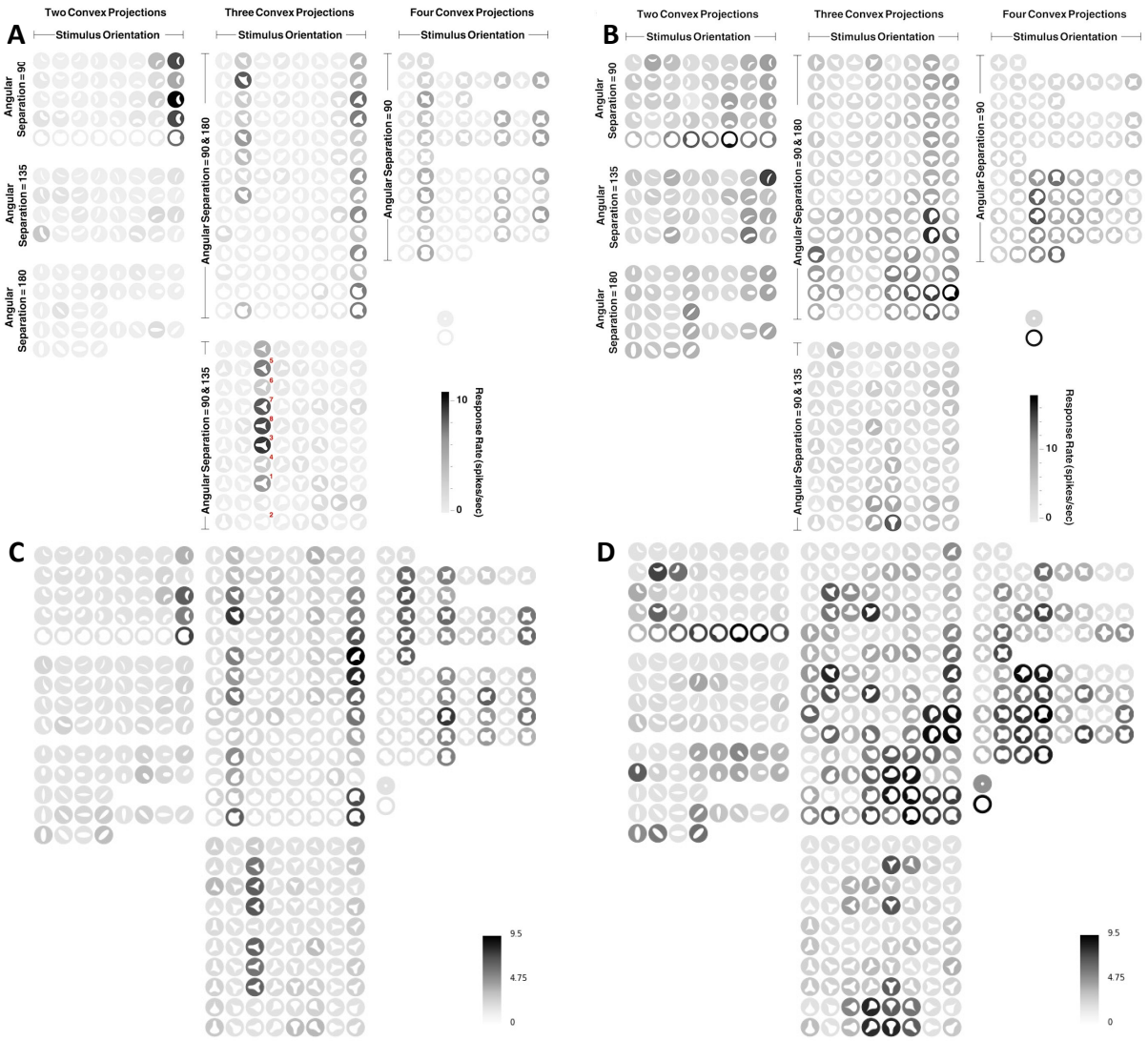

Fig. 7 Recording and simulation of Medium curvature selectivity in V4: The white shape in each location represents the the stimuli shape, which is actually presented as a black silhouette on a white background. The shading of the disc around the shape represents the intensity of the response as indicated by the legend. **A**Recording reproduced from [Pasupathy and Connor, 2001] showing a V4 neuron tuned for concave curvature at the right. **B** Recording reproduced from [Pasupathy and Connor, 2001] showing a V4 neuron tuned for broad convex curvature at the top. **C** Result of a model simulation qualitatively matching **A**. The results shown are the number of spikes out of N simulations. **D** Result of a model qualitatively matching **B**.

position. Figure 8 A-F show the connection pattern for the neuron tuned to concave curvature at the right, and Figure 8 G-L show the connection pattern for the neuron tuned to broad convex curvature at the top. For a complete explanation of the meaning of each part of the Figure, see the text accompanying Figure 4. The selectivity of the afferent inputs to the V4 L2/3 neurons are represented by pictograms in Figure 8: These are the outputs of V2 neurons selective to corners having interior angles 90° (A,G), 112.5° (C,I) and 145° (E,K). The corresponding convolution filter weights are shown below each pictogram, Figure 8 B,D, F (concave at the right) and H, J and L (broad convex at the top). For the neuron selective to concave curvature at right, the excitatory connections are to right opening corner selective afferents. For the neuron selective to

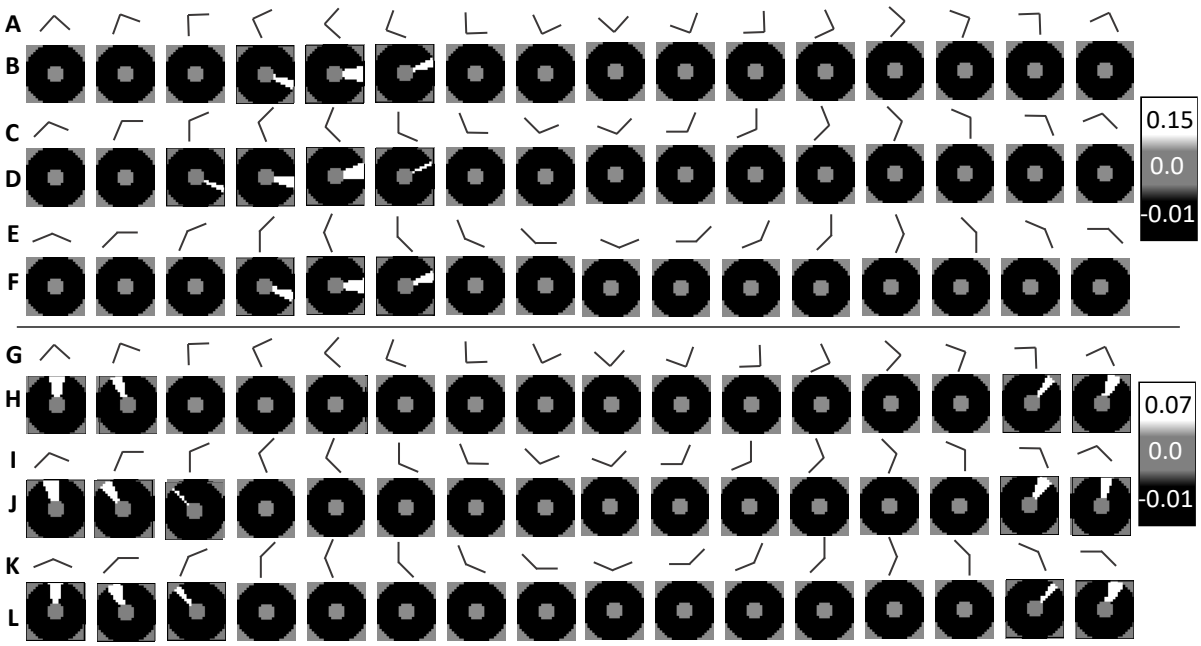

Fig. 8 Afferent Connections for V4 neurons selective to medium curvatures in asymmetric zones. **A - F**: Connections for the model V4 neuron selective to broad convex curvature at the top. **G - L**: Connections for the model V4 neuron selective to concave curvature at the right. **A, C & E**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 90° (A), 112.5° (C) and 145° (D) interior angles at a complete range of rotations every 22.5°. Details of V2 neuron selectivity are described in the Supplementary Methods B.3. **B, D & F**: The synaptic connection strength versus position in topographic space relative to the soma at the center for a V4 neuron selective to broad convex curvature at the top. Each square is a single 9×9 convolution filter implementing the weights for the V2 afferent inputs indicated by the pictogram above. White=excitatory, Black=Inhibitory, Gray=No connection. All of the weights in A-F together are the connections to the topographically arranged inputs for one V4 neuron. There are excitatory weights on the inputs that are convexly oriented in the top part of the receptive field; other afferent inputs are weakly inhibitory. The orientation of inputs that are excitatory rotates with position around the receptive field. **G, I & K**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 90° (G), 112.5° (I) and 145° (K) interior angles at a complete range of rotations every 22.5°. **H, J & L**: The synaptic connection strength versus position in topographic space relative to the soma at the center for a V4 neuron selective to concave curvature at the right, as in B/D/F. There are excitatory weights on the inputs that are concavely oriented in the right part of the receptive field; other afferent inputs are weakly inhibitory. The orientation of inputs that are excitatory rotates with position around the receptive field.

broad convex curvature at the top, the excitatory connections are to downward opening corner selective afferents. In both cases, there are excitatory connections to a small range of orientation of afferents in a wedge shaped zone outside the center of the receptive field. The precise orientation selectivity of the afferent inputs rotates as the position of the inputs rotates around the receptive field. The overall preferred stimuli set for the pattern were illustrated in Figure 1V: it is a set of medium curvature overlapping curves in a particular region of the receptive field.

2.2.2 Grid of Composite Bar Stimuli

[Nandy et al., 2013] also found that some V4 units are selective exclusively to medium curvature. And [Nandy et al., 2013] found that some units selective to medium curvature had such selectivity over a broad range of positions, and the orientation of the preferred curvature varied with position around the receptive field. The recording from [Nandy et al., 2013] of a medium curve selective V4 L2/3 unit with selectivity through its receptive field is shown in Figure 9 and compared to a model. For a detailed explanation of the parts of Figure 9 see the text accompanying Figure 5 at the beginning of Section 2.1.2. For both the model and recording the receptive field

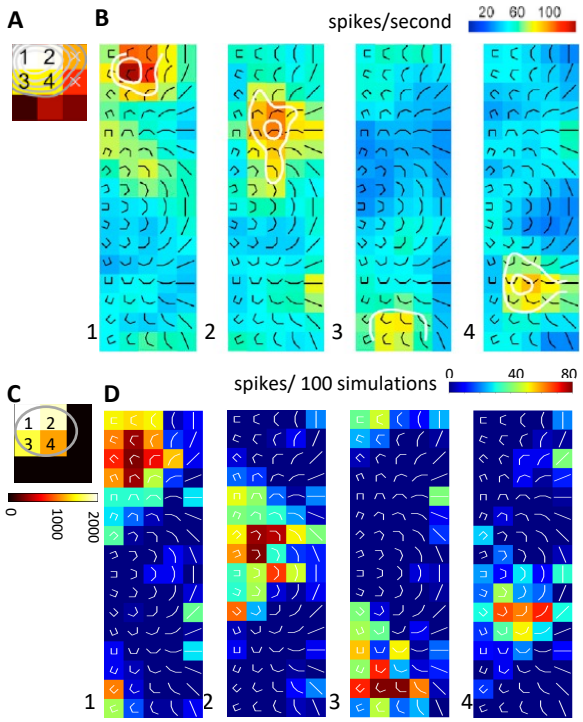

Fig. 9 Medium curvature selective unit recording (top) and simulation (bottom) **A** Receptive field for recording. **B** Shape tuning for recording, 1-4 show the shape tuning for curvature (x-axis) and orientation (y-axis) at indicated positions in the receptive field (A). **C** Receptive field for simulation. The approximate location and extent of the simulated neurons dendritic tree in the topographic space of the afferent inputs is shown. **D** Shape tuning for simulation, as in B, at positions in the simulated receptive field (C). The simulation qualitatively reproduces the most significant features of the medium curvature selective neuron recording.

is highly varied over multiple stimuli locations. There is strong response to medium curvatures, though also broad tuning that includes activation to straight and highly curved stimuli. The orientation of the most preferred stimuli are convex around the center of the receptive field. The main shortcoming of the model to reproduce the recording is that the orientations of some of the preferred stimuli are different by around 10°-20°, and also the model lacks background firing to the less preferred stimuli found in the recordings.

The connection pattern to afferent inputs for for the [Nandy et al., 2013] medium curvature selective V4 L2/3 cell model is shown in Figure 10. The afferent inputs to the V4 L2/3 neuron are represented by pictograms Figure 10 A,C, E, G, I

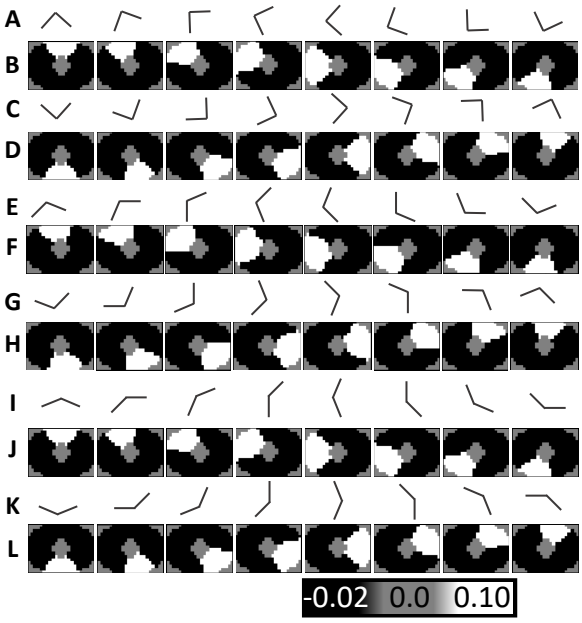

Fig. 10 Afferent Connections for model V4 L2/3 neurons selective to convex medium curvatures throughout the receptive field. **A & C**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 90° interior angle and a complete range of rotations every 22.5°. Details of V2 neuron selectivity are described in the Supplementary Methods B.3. **E & G**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 112.5° interior angle and a complete range of rotations every 22.5°. **I & K**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 145° interior angle and a complete range of rotations every 22.5°. **B, D, F, H, J, L**: The synaptic connection strength versus position in topographic space relative to the soma at the center for a V4 neuron selective to medium convex curvature throughout the receptive field. Each square is a single 20×24 convolution filter implementing the weights for the V2 afferent inputs indicated by the pictogram above. White=excitatory, Black=Inhibitory, Gray=No connection. All of the weights in B-L together are the connections to the topographically arranged inputs for one V4 neuron. There are excitatory connections to every orientation of the V2 afferents, and the position of the excitation rotates around the receptive field with the orientation of the inputs so that convexly oriented inputs are excited and other inputs are inhibited.

and K: These are the outputs of V2 neurons selective to corners having interior angles 90° (A,C), 112.5° (E,G), and 145° (I,K). The corresponding convolution filter weights are shown below each pictogram, Figure 10 B,D, F, H, J and L. For a detailed explanation of the meaning of each part of the Figure, see the text accompanying Figure 4 in Section 2.1.1. The pattern is that for every type

of afferent input, a wedge covering around 20% the receptive field is excitatory and the remainder is inhibitory, and a small region at the center is neutral. The direction of the excitatory wedge is such that the excitatory afferent inputs are approximately concave to the center of the receptive field. To achieve this, the excitatory wedge boundary rotates around the receptive field, depending on the orientation of the afferent inputs. Non-preferred orientations are inhibited and afferent inputs with both lower interior angle (more acute) and higher interior angle (more straight) complex shape preferences are ignored. The overall preferred stimuli set for the pattern were illustrated in Figure 1W: it is a set of medium curvature overlapping curves, convex to the center all around the receptive field.

2.3 V4 Low Curvature Selective Model

[Nandy et al., 2013] also found that some V4 units are selective to low curvature or straight stimuli with a consistent orientation tuning throughout the receptive field - for these V4 neurons, the selectivity was noted to be invariant with position in the receptive field. Figure 11 compares the shape tuning of the V4 L2/3 neuron model with the recording. For a detailed explanation of Figure 9 see the text accompanying Figure 5 at the beginning of Section 2.1.2. For both recording and simulation, the most significant activity for the unit occurs at 3 stimuli locations. The most preferred stimuli are straight or minimally curved and at a small angle counter clockwise from horizontal. However, there is broad tuning around the most preferred stimulus and non-zero activity as high as 50% of the maximum for stimuli that are at orientations as much as 22.5° away and at medium curvatures.

The main deficiency of the model in reproducing the recording is that the model does not have very weak responses (25% of the maximum) to as wide a range of orientation and curvatures as is seen in the recording. Also, the model neuron does not have any responses in the stimuli locations outside of top 4, while the recordings show weak responses.

The connect patterns embodied by convolution filters for the unit that produces the simulation result in Figure 11 is shown in Figure 12. The

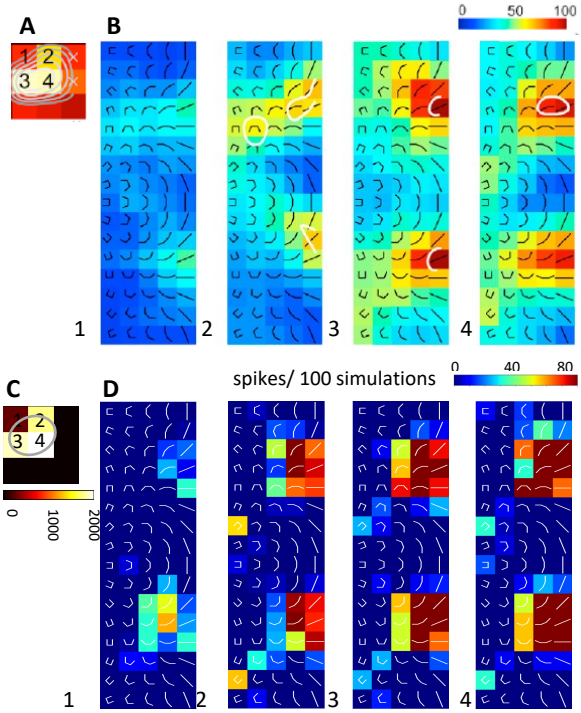

Fig. 11 Low curvature selective unit recording (top) and simulation (bottom) **A** Receptive field for recording. **B** Shape tuning for recording, 1-4 show the shape tuning for curvature (x-axis) and orientation (y-axis) at indicated positions in the receptive field (A). **C** Receptive field for simulation. The approximate location and extent of the simulated neurons dendritic tree in the topographic space of the afferent inputs is shown. **D** Shape tuning for simulation, as in B, at positions in the simulated receptive field (C). The simulation qualitatively reproduces the main features of the low curvature selective neuron recording.

afferent inputs to the V4 L2/3 neuron are represented by pictograms Figure 10 A,C and E: These are the outputs of V2 neurons that have "ultra-long" selectivity (A) and complex selectivity at very high interior angles of 157.5° (nearly straight). The corresponding convolution filter weights are shown below each pictogram, Figure 10 B,D, and F. The connection pattern is to have the strongest connections to afferent inputs from ultra-long units with an orientation of 22.5° from horizontal, but also with broad tuning including orientations from 0° to 45°. Afferents for nearly straight complex units make excitatory connections at when the orientation of the selectivity is close to 22.5° from horizontal only. The connections to the nearly straight selective V2 afferents are at a lower strength than the most preferred

orientation ultra-long units. Afferents from inputs selective for other orientations of nearly straight stimuli are accepted from inhibitory L4 units, as shown in Figure 12D, while afferents from inputs selective for more acute angled complex-shaped units are ignored.

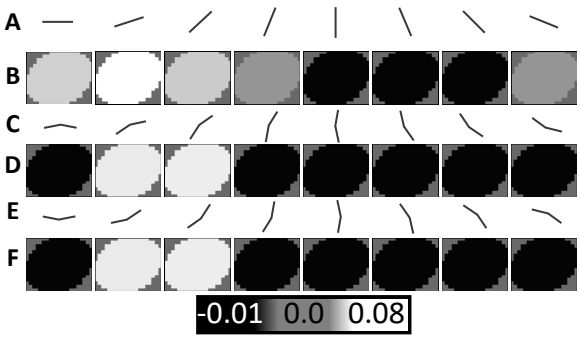

Fig. 12 Afferent Connections for model V4 L2/3 neurons selective to low curvatures throughout the receptive field. **A**: Pictorial representation of the preferred stimuli of V2 afferent inputs to V4 having selectivity to "ultra-long" straight lines at a range of orientations. **C & E**: Pictorial representation of the preferred stimuli of V2 afferents inputs to V4 having a 157.5° interior angle at a complete range of rotations every 22.5°. Details of V2 neuron selectivity are described in the Supplementary Methods B.3. **B, D, F**: The synaptic connection strength versus position in topographic space relative to the soma at the center for a V4 neuron selective to low curvature throughout the receptive field. Each square is a single 15×20 convolution filter implementing the weights for the V2 afferent inputs indicated by the pictogram above. White=excitatory, Black=Inhibitory, Gray=No connection. All of the weights in B-F together are the connections to the topographically arranged inputs for one V4 neuron. The strongest excitation comes from afferents selective to ultralong configurations at an angle 22.5° above horizontal, and weaker excitation from afferents selective to ultralong inputs at nearby angles. There are also nearly as strong weights on the afferents selective to 157.5° interior angle corners at an angle of 22.5° above horizontal.

In comparison to the model neurons for V1 and V2, the V4 model neuron has a relatively large receptive field: 15x20 in the reduced frame of V4, equivalent to 60x80 in the original 112×112 pixel visual field model. (In comparison, the V1 neurons have receptive fields of 6×9 and the V2 neurons have fields of 18×18 in the frame of V1.) This large receptive field is required to match the recordings, as shown in Figure 11 and the model does not match the weak secondary stimulation.

Due to the large receptive field, the total excitation in the V4 L2/3 units is much larger than in the V1 and V2 connection patterns. The total excitation for the low curvature selective V4 L2/3 neuron model is approximately 164, in units of the neurons threshold, while the total inhibition is 52. The high amount of excitation is required to match the recordings: As can be seen from Figure 11, a small stimulus in a fraction of the neurons receptive field is sufficient to cause a spike. Because the response is relatively uniform in the receptive field, it follows that there must be many times more excitative inputs available than what is required to respond to a small stimulus. The preferred stimuli were illustrated in Figure 1I: it is a classic parallel line grating, but with sone caveats: it allows imperfectly straight lines, and it requires a matching stimuli in only a small portion of the receptive field to trigger a spike.

2.4 PSP Simulation Dynamics

Figure 13 illustrates the dynamics of PSP summation at the soma for the model in several illustrative neurons. Figure 13A shows examples of Pyramidal model neurons in V1. The time constants for the model are based on the measurements of [Holmgren et al., 2003] and shown in Table 1. [Holmgren et al., 2003] found that PSP's in Pyramidal neurons have duration that is relatively long compared to the time scale of spike synchronization : The 10-90% rise time averages 4.5 ms and the width at half amplitude is 30 ms. In comparison, spikes arrive in V1 from the LGN in synchronized bursts on the order of a few ms [Singer, 1999]: As described in detail in the Methods, these arrival times are modeled with a gamma distribttion with a standard deviation of 2 ms. As a result, the PSP summation trajectories illustrated in 13A (from V1 model pyramidal neurons) appear fairly contiguous as summation and threshold crossings occur before any substantial decay of the individual PSP's. Note that according to [Holmgren et al., 2003] the PSP time constants of inhibitory to pyramidal connections are substantially similar to those of pyramidal-pyramidal connections; for the model of this paper these time constants were made the same.

Figure 13B illustrates the PSP's for the time constants of the PSP's in a model inhibitory neuron, received from a pyramidal neuron, in accordance with the result in [Holmgren et al., 2003]. These are "fast spiking" neurons and have significantly shorter PSP time constants: the 10-90% rise time is just 2.2 ms and the width at half amplitude is 14 ms. For the inhibitory neuron time constants the inflections in the trajectory due to the addition of excitation or subtraction of inhibition are somewhat more apparent but the PSP time scale is still long compared to the summation. Note that for illustrative purposes these model neurons of Figure 13B include inhibitory inputs: The actual inhibitory neurons in the model have only excitatory inputs and the trajectories are less able to illustrate the time constant properties.

In order to confirm that the model could reproduce more dynamic PSP trajectories, additional simulations were performed using time constants faster than those reported in [Holmgren et al., 2003]. The results of these simulations are shown in Figure 13C; for these simulations the 10-90% rise time was decreased to 0.2 ms and the width at half amplitude reduced to 2.6 ms. Also, for illustrative purposes the synchronization of the input spikes was decreased: the arrival time gamma distribution had a 3 ms standard deviation instead of 2 ms. With these unnatural time constants it is clear that the model can reproduce much more rapid PSP dynamics. However, this is for illustrative purposes and not applicable to the cortical model described in this paper.

While there is no question of the softwares ability to compute the time course of faster PSP dynamics, [Holmgren et al., 2003] suggests that the reality is slow relative to the tight input time synchronization. The PSP time constants are also slow relative to the presumed frequency of cascades: As summarized in [Buzsáki and Wang, 2012], cortical neurons engaged in gamma timescale activity typically have active bursts every 10-30ms. With such a slow time constant of decay, width at half amplitude 30ms for the excitatory and 14ms for the inhibitory neurons, there is not enough time for passive decays to return the neuron to a resting state. It is therefore assumed that the networks responsible for gamma oscillations include inhibition to return the cortical neurons back to their resting state more quickly. However, this level of detail is beyond the scope of this modeling study, focused on the feedforward cascade as it is.

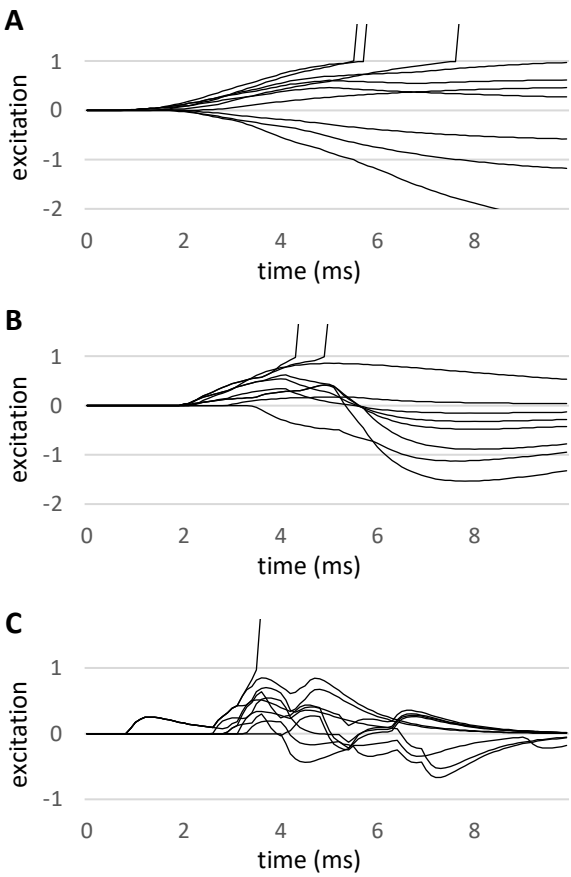

Fig. 13 Depolarization Dynamics **A** Parameters of the model L2/3 Pyramidal and L4 neurons **B** Parameters of the model inhibitory neurons, which are modeled as "Fast Spiking". **C** Test of faster dynamics.

2.5 Spike Propagation

The distribution of spike times in each layer of the model is illustrated in Figure 14**A**, in response to one of the stimuli composed of convex and concave boundaries from [Pasupathy and Connor, 2001] (Figure 14**B**). As described in the Methods, the spike transmission time between layers are drawn from gamma distributions designed to match typical intra- and inter-cortical propagation delays (See Methods Section 4.1.1 and Table 5.) Under these assumptions it takes around 25 ms for a spike wave to propagate from the

Table 1 PSP Time Constants (ms) and properties for Figure 13.

Sub-Figure	τ_{rise}	τ_{fall}	10-90% rise time	Width at half-amplitude	Source
A	4.5	30	5.3	36	[Holmgren et al., 2003] Pyr-Pyr
B	1.9	13	2.2	14.6	[Holmgren et al., 2003] Pyr-FS
C	0.25	1	0.2	2.6	NA

model V1 L4 input layer to V4 L2/3. The results are in agreement with [Girard et al., 2001] which shows orthodromic latency between V1 and V2 neurons of just around 2-3 ms and [Plomp et al., 2017] which shows inter-laminar latencies between L4 and L2/3 of around 1 ms. It is also consistent with the overall picture presented in [Thorpe and Imbert, 1989] and [VanRullen et al., 2005] that each cortical area takes around 10 ms to perform its role in a rapid feedforward cascade that leads to visual recognition.

The simulations of Figure 14 were run for 36 ms, although it was found that more than 99% of spikes that will arrive have arrived by around 27.5ms. The time limit for the simulations shown in Figures 3 though 11 was set to 30 ms to capture the vast majority of spikes in a reasonable length simulation.

Figure 14**C** shows the percentage of neurons spiking in each layer in response to the stimuli. While around 10% of the model units in the first V1 and final V4 layers spike, in the interior layers spiking is much sparser: only 1% or fewer units fire. Note that for this test the final layer includes both the model units created to match the results of [Pasupathy and Connor, 2001] (Figures 4 and 8) and [Nandy et al., 2013] (Figures 6, 10 and 12), and further includes all of these units at a complete set of rotations as described in the Methods.

Figure 14**D** through **F** shows the same tests as Figure 14**A** through **C**, but in this case one of the convex and concave boundaries stimuli from [Pasupathy and Connor, 2001] is rendered with an outline. This stimuli shows significantly less spikes in the first layer and slightly faster propagation of the signal to later layers. Otherwise the results are substantially the same. More details of tests on outlines of convex and concave boundary stimuli are shown in Section 2.6.2.

2.6 Simulated Responses to Novel Stimuli

With the model tuned to reproduce results from [Pasupathy and Connor, 2001] and [Nandy et al., 2013], it can be used to predict results for novel stimuli. One small example of this was already presented in Figure 14**E**: While [Pasupathy and Connor, 2001] used silhouette contours, the stimuli of [Nandy et al., 2013] are lines - Figure 14**E** is a blend of the two approaches, making a contour boundary stimuli as an outline. Model experiments like this can be used to verify the model in future neurophysiological experiments, improve future models with more specific constraints, suggest useful stimuli for future experiments, and help inform hypotheses about the function of V4.

2.6.1 Swapping Units and Stimuli from Past Studies

The V4 units selected for presentation in [Pasupathy and Connor, 2001] and [Nandy et al., 2013] were different: [Pasupathy and Connor, 2001] focused on neurons selective for particular degrees of curvature in particular zones of the receptive field, while [Nandy et al., 2013] focused on neurons with selectivity for a degree of curvature throughout the receptive field. In each case, the stimuli presented are well suited to illustrate the properties of the neurons they selected. How the neurons of one study would have responded to the stimuli of the other is an interesting question.

Figure 15 illustrates what response could be expected if the units presented in [Pasupathy and Connor, 2001] were stimulated with the approach of [Nandy et al., 2013]. For the neurons selective to acute curvature the response, shown in Figure 15A and 15B, was weak and mainly to straight line stimuli and not to curves. This is because the stimuli of [Nandy et al., 2013] do not contain angles more

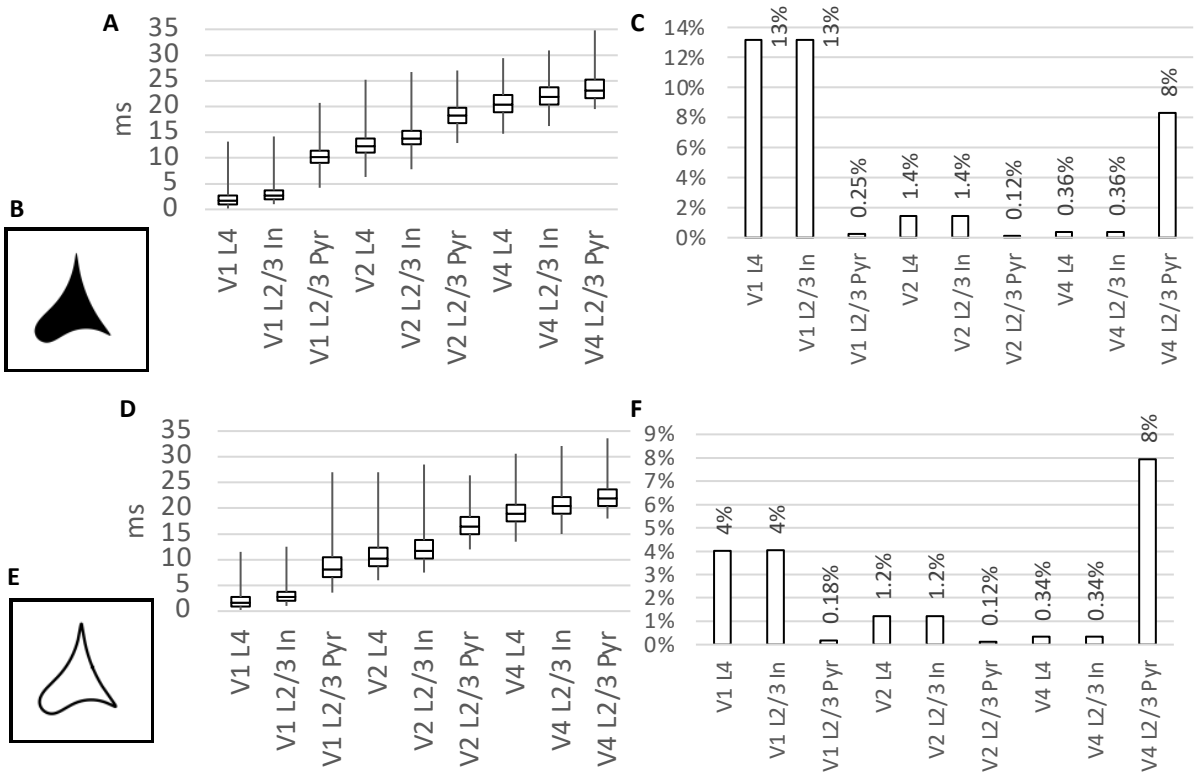

Fig. 14 Spike Propagation **A** Spike time distribution in the model for a single stimulus from [Pasupathy and Connor, 2001] (shown as **B**) Box-plots show the median and upper/lower quantiles, and the whiskers show the 1st/99th percentiles. **C** Percentage of neurons spiking in each layer for the stimulus **B**. **D** Spike time distribution in the model for a single outline stimulus based on [Pasupathy and Connor, 2001] (shown as **E**) **F** Percentage of neurons spiking in each layer for the stimulus **E**.

acute than 90°, but as illustrated in Figure 4 the model neuron selective for acute convex curvature are optimally stimulated by stimuli at an angle of 45°. So the responses in Figure 15 A and B are a secondary response to nearly straight stimuli that are aligned with the excitable portion of the receptive field. In this case the stimuli of [Nandy et al., 2013] seems inadequate to capture the full variability of V4 neurons.

The neurons selective to medium curvatures in [Pasupathy and Connor, 2001] respond well to the stimuli of [Nandy et al., 2013] in a manner matching the selectivity to the stimuli of [Pasupathy and Connor, 2001]. This is illustrated in Figure 15C for the V4 neuron tuned for concave curvature at the right, and Figure 15D for the V4 neuron tuned for broad convex curvature at the top.

Figure 16 illustrates what response could have been expected if the neurons presented in [Nandy et al., 2013] were stimulated with the approach of [Pasupathy and Connor, 2001]. In this case, all of the neurons responded very strongly to the stimuli. With the parameters tuned to match the results of [Nandy et al., 2013], the model neurons responded so vigorously to almost all stimuli that there was no differentiation in the shape tuning. To continue the experiment, the excitation threshold in the model the neurons was raised to the point where there was a varied range of responses across stimuli - this can be viewed as a form of adaptation by the model neurons albeit manual.

However, even after raising the threshold of the model neurons based on [Nandy et al., 2013] to the point where the stimuli of [Pasupathy and Connor, 2001] exhibit a full range of responses, so many of the stimuli elicited strong responses that it is still difficult to discern clear patterns. The response of the high curvature selective unit to the boundary contour stimuli is illustrated in Figure 16A: It tends to respond to

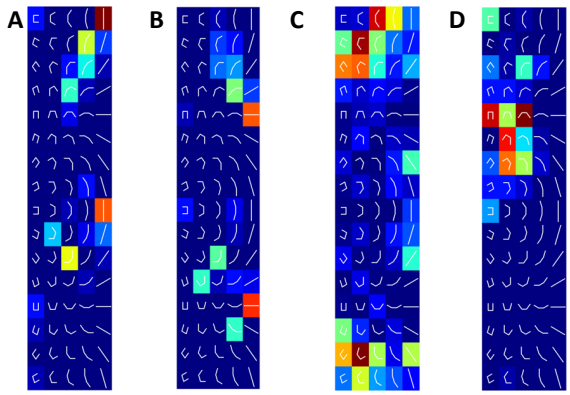

Fig. 15 Simulated response of neurons from [Pasupathy and Connor, 2001] to stimuli of [Nandy et al., 2013]. Each pictogram represents a composite bar stimuli that varies by curvature (x-axis) and orientation (y-axis); the color coded scale shows the response firing rate. For a complete description see Figures 2 and the text accompanying Figure 5. **A** The neuron of Figures 3A and 4 selective to for acute convex curvature at the bottom left. **B** The neuron of Figure 3B selective for acute convex curvature at the top right. **C** The neuron of Figure 7A tuned for concave curvature at the right. **D** The neuron of Figure 7B tuned for broad convex curvature at the top.

shapes that have at least some acute angles, but the pattern is irregular and inconsistent - all the shapes triggering strong response have some acute angle, but not all shapes with acute angles trigger a strong response. This is due to the fact that the shape of the receptive field (Figure 6) is elongated on the vertical dimension and some stimuli do not lie evenly inside the receptive field. The response of the medium curvature selective unit is shown in 16B and a similar characterization applies: It tends to favor shapes with some broad, non-acute curves in some part of the shape. But the response is uneven at different orientations. The response of the low curvature selective unit is illustrated in Figure 16C. Knowing that the neuron is selective to nearly straight lines at an angle that is approximately 22.5° above horizontal (as illustrated in Figure 12) it is fairly apparent that every shape to which the neuron has at least some boundary that matches. But the selectivity to straight and nearly straight lines at a particular orientation is not as apparent in the contour boundary stimuli of [Pasupathy and Connor, 2001] as in the grid of curves stimuli from [Nandy et al., 2013].

This comparison suggests that both the contour boundary stimuli of

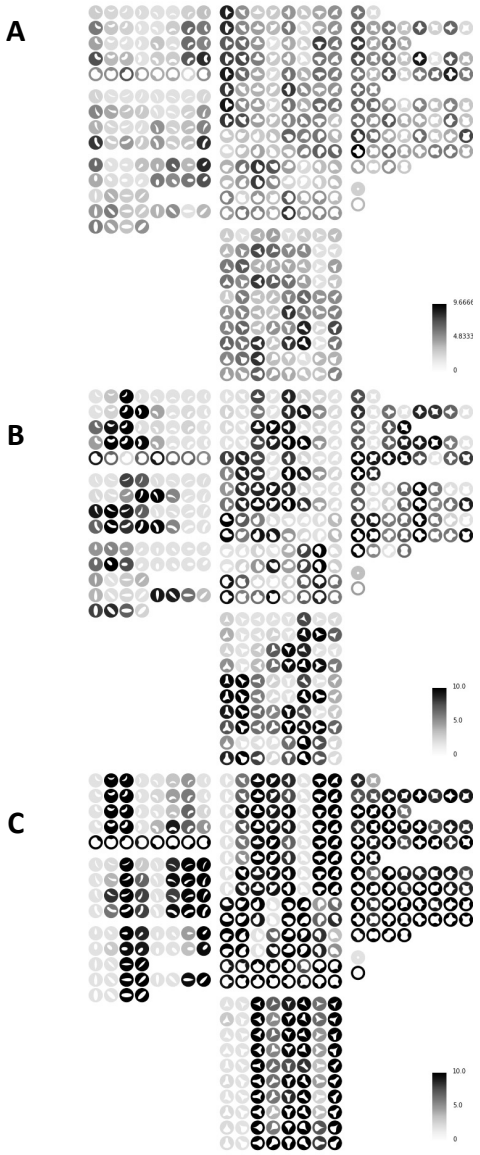

Fig. 16 Simulated response of neurons from [Nandy et al., 2013] to stimuli of [Pasupathy and Connor, 2001]. See Figure 3 for details. **A** The neuron of Figure 5 and 6 selective for high curvature. **B** The neuron of Figure 9 and Figure 10 selective for medium curvature. **C** The neuron of Figure 11 and Figure 12 selective for low curvature.

[Pasupathy and Connor, 2001] and the grid of curves stimuli of [Nandy et al., 2013] have certain blind spots in characterizing the response of V4 units that the other covers well. A better stimuli set for characterizing all of these units may be one that maps the orientations and acuteness of the selectivity on a grid like in [Nandy et al., 2013]

but includes more acute angled stimuli. For example, each stimuli could consist of just two bars at a full range of angles from fully straight to highly acute. (TODO: Simulate such a stimuli set and show the results for the model neurons of both studies?)

2.6.2 Response to Outlines of Convex and Concave Boundary Stimuli

As already illustrated in Figure 14, a novel stimuli set can be created by blending the characteristics of the stimuli in [Pasupathy and Connor, 2001] and [Nandy et al., 2013]: Contour boundary stimuli that are presented as outlines instead of silhouettes. For another example, see Figure 18A and B. Simulations were performed for the complete data set of [Pasupathy and Connor, 2001] presented as outlines, instead of as silhouettes. The results of this experiment are illustrated in Figure 17: the responses of the simulated V4 neurons to the outline are qualitatively similar to the standard silhouette. In every case the main shapes responded to are similar, but are not entirely the same. The neuron selective to concave shape at the right (Figure 17C) responds somewhat more to the outline stimuli (right) than the silhouette. Also, the neuron selective to broad concave curvature at the top (Figure 17D) responds somewhat more to the silhouette stimuli (left). For the neurons responsive to acute convex (Figure 17A and B) it appears that the responses are more equal in overall response but still not precisely the same.

A quantitative comparison of the response to the outline boundary stimuli as compared to the standard silhouette stimuli is shown in Table 2. The first measurements is the mean of the ratio of responses to corresponding shape outline stimuli and silhouette stimuli denoted simply $|\bar{Outline}|$ where the silhouette is assumed as the baseline. The second metric is the Pearson correlation coefficient between the responses, denoted $r(Outline)$, treating the response as a vector over all the stimuli. These results confirm that the simulated concave selective neuron has a higher average response to outlines, while the simulated board convex selective neuron has a lower average response to the outline stimuli. The acute selective neurons have very nearly the same response to both outline and silhouette stimuli. At the same time, the concave selective neuron has the least

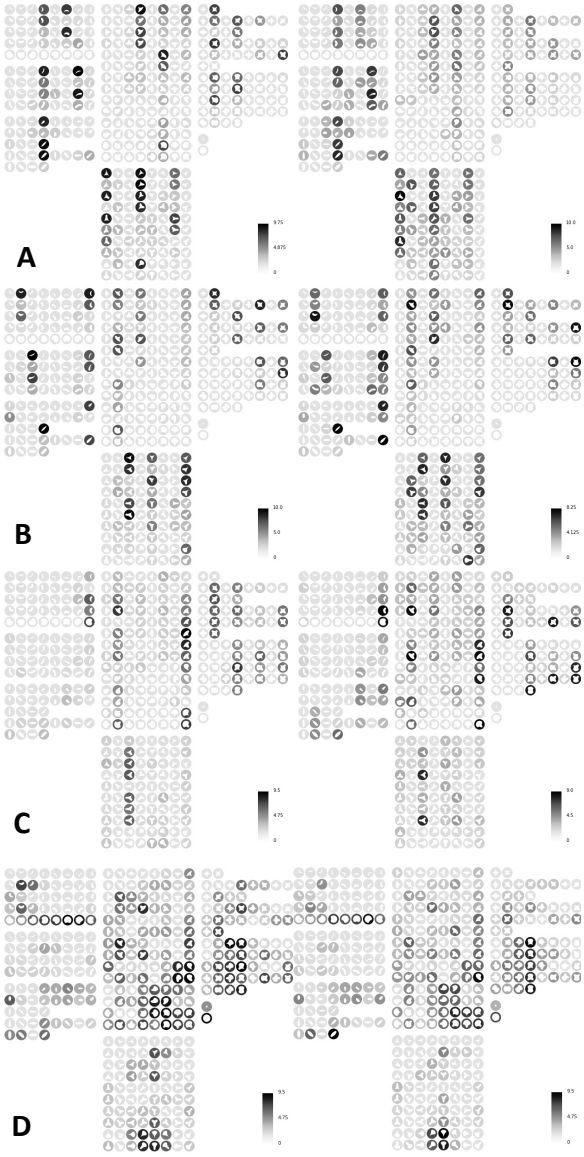

Fig. 17 Simulated response to the stimuli of [Pasupathy and Connor, 2001] in silhouette (Left) vs. outline (Right). **A** The neuron of Figure 3A, acute convex curvature at the bottom left. **B** The neuron of Figure 3B, acute convex curvature at the top right. **C** The neuron of Figure 7B, concave curvature at the right. **D** The neuron of Figure 7B, broad convex curvature at the top.

correlation between the responses to outline and silhouette versions of the same stimuli.

The reasons for the different response to outlined versions of the medium curvature convex and concave boundary based stimuli are striking because both neurons are selective for curves of

Table 2 Comparison of Response to Silhouette and Outlines Boundary Stimuli. $|\bar{Outline}|$: Average magnitude of responses for outline stimuli relative to the response of the standard silhouette stimuli. $r(Outline)$: Correlation of of outline and silhouette stimuli response vectors.

Neuron	$\|\bar{Outline}\|$	$r(Outline)$
Acute Convex Left	1.01	0.90
Acute Convex Right	1.05	0.87
Concave Right	1.21	0.66
Broad Convex Top	0.73	0.90
Average	1.00	0.83

the same degree, the only difference being the convex versus concave orientation. The reason for the differences has to do with the slightly unequal response to V1 neurons selective for edges and lines. These differences will be detailed in the Method section on tuning the response properties of the model. This subtle variation between the responses of neurons selective for convex and concave medium curvature is an phenomena which could validate the model, or help to refine it.

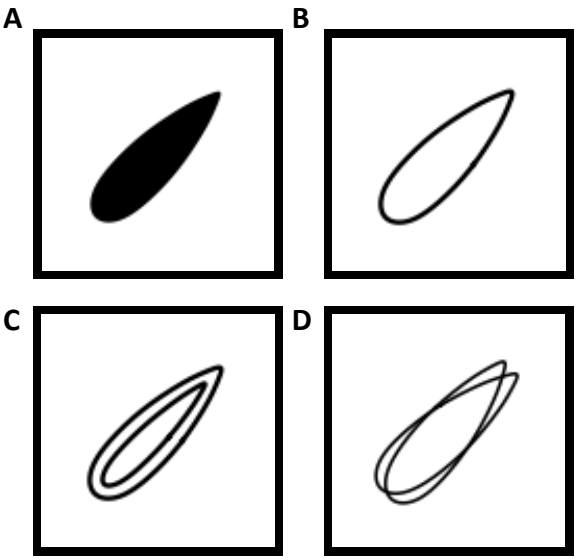

Fig. 18 Examples of alternate stimuli based on [Pasupathy and Connor, 2001]. **A** One of the original stimuli. **B** Outline. **C** Contour pair. **D** Rotated pair.

2.6.3 Response to Concentric and Rotated Outline Boundary Stimuli

Another set of experiments was performed to compare the response of the simulated V4 neurons when an outline of a boundary based stimuli is presented in two altered versions: A concentric version of the same shape as compared to a pair of outlines slightly rotated relative to each other. The concept is illustrated in Figure 18: The original stimuli is A, and an outline stimuli like the ones used in Section 2.6.2 is shown as B. Figure 18C is the concentric outline pair, and D is the rotated outline pair. In the rotated pair, two slightly rotated versions offset by +/-$\pi/16$ radians replace the original. As will be detailed in the Methods, some shapes were omitted from this data set because they were either too narrow to contain a concentric shape, or did not have meaningful rotations. The resulting set of stimuli consisted of 324 boundary configurations, out of the original 366. This experiment was chosen because it may help explain the results of [Srinath et al., 2021], which found that many V4 neurons are most responsive to 3-D objects. The concentric shape pairs (Figure 18C) were chosen because they resemble contour lines of shadows and luminescence that occur naturally on 3-D stimuli. The rotated shapes (18D) acts as a control that has approximately the same amount of total intensity in the receptive field and in almost the same configuration, but it does not have similarity to a 3-D object.

The results in Table 3 show that overall the V4 neurons are more selective for the concentric outline pairs than the rotated outline pairs. This is particularly true for the neuron selective to medium curvature at the right, and also true to a lesser degree for the neurons selective to acute convex curvature. In contrast, the rotated pairs are less effective at stimulating these neurons; this is particularly true for the concave and medium curvature selective neurons but also to some degree for the neurons selective for acute convex curvature. The one exception is the neuron selective to concave curvature - for that neuron, the model suggests that concentric pair stimuli will have a slightly reduced response in comparison so the simple outline, while the rotated pair stimuli will have a moderately increased response. These results suggest that comparison of response for

Table 3 Response to Paired Concentric and Rotated Outline Boundary Stimuli: $|\bar{...}|$: Average magnitude of response compared with the outline boundary stimuli. $r(...)$: Correlation of responses with response to outline boundary stimuli.

Neuron	$\lvert\overline{Concentric}\rvert$	$\lvert\overline{Rotated}\rvert$	$r(Concentric)$	$r(Rotated)$
Acute Convex Left	1.24	0.97	0.89	0.83
Acute Convex Right	1.21	0.88	0.82	0.69
Concave Right	0.91	1.14	0.74	0.76
Broad Convex Top	1.67	0.79	0.82	0.75
Average	1.26	0.94	0.82	0.76

concentric shape pairs stimuli and rotated shape pairs would be a useful experiment for validating and improving the OC model and understanding selectivity of V4 neuron to 3-D objects described in [Srinath et al., 2021].

3 Discussion

3.1 Conclusions

The Organic Convolution (OC) model is a simplified but biophysically accurate model for the feed forward propagation of spikes in cortex-like layers. The biophysical parameters for the time constants governing the spiking model PSP equations (Equation 1 and Table 4) are within plausible ranges as are the time constants for the distributions modeling the arrival and transmission of spikes (Table 5). Therefore OC meets the speed requirement in visual recognition [Thorpe et al., 2001] and shows how processing from V1 to V4 can occur with a cascade of single spikes.

The debatable part of this model is whether axons and dendrites can assemble into the proposed patterns of afferent connections throughout cortex. The connections patterns seem biologically plausible in that they define local receptive fields and are quite simple. The evidence from V1 [Ko et al., 2013] and V2 [Kim et al., 2014] is that neurons do pre-wire into computationally useful patterns. The proposed functional connection patterns in this study are suggestions for the types of connections are sufficient to match V4 recording data but are not definitive.

It also is not realistic to assume that all neurons of a given type will be regularly shaped nor is it plausible that all connection patterns of similar units at different locations are exactly identical. The OC model in this study makes more sense when the micro-architectural patterns are viewed as a model for the "average" or "ideal" version of each unit type. Cortex may contains many units that follow a micro-architectural pattern in any area of the receptive field, but each individual unit must have its own irregular shape determined by its dendritic growth pattern around the other elements in the neuropil. When combined such units may functionally achieve a robust version of the average micro-architectural pattern. Investigating OC models with realistic numbers of units, more variety in unit shape and connections, and transmission failures is an attractive area for further research.

For understanding the nature of visual processing in V4, the OC model may suggest a synthesis of the points of view in different neurophysiological approaches. Studies suggesting V4 responds in a place invariant manner to shapes defined by closed paths [Pasupathy and Connor, 2001, Kim et al., 2019] may be reconciled with the fine shape and orientation tuning shown in [Nandy et al., 2013]: The same underlying curvature selective connection patterns are present in both types of units. Selectivity to a curvature at a specific position in the receptive field is a partial implementation of a rotational connection pattern that may also wrap around the entire receptive field.

The recent results in [Hu et al., 2020] seem in agreement with[Pasupathy and Connor, 2001] and [Nandy et al., 2013] and the OC model in suggesting that an important function of V4 may be to represent local variations in orientation and curvature. The best way to resolve questions of which stimuli best characterize V4 neurons would be to use multiple types of stimuli in a single study, and see if a single model can reproduce results collected under uniform conditions. As suggested, the best way to reconcile earlier studies with the recent findings of [Srinath et al., 2021] that V4 neurons may prefer 3-D shapes is to

compare the response to similar shaded, outlined, and contoured versions of the same shape. With the greatest respect for the painstaking work of neurophysiological recording in awake subjects, and in particular the difficulty of recording from visual area V4, an outside observer is left with the impression that neurophysiologists studying V4 resemble the parable of the blind men and an elephant: Each study shows a different kind of selectivity depending on the nature of the stimuli presented. The hope is that the present modeling study may point to common underlying mechanisms with which all of these behaviors can be explained.

This study does not include a model of an unsupervised connection learning algorithm (e.g. [Miller, 1994]) that can generate the connection patterns. This may seem like a weakness, but to defuse this criticism consider an analogy: Biologists often observe stripes of pigment on the bodies of animals or plants and reach conclusions about the function of those stripes for camouflage, signals, etc. But no one expects the biological mechanisms of pigment stripe formation to involve an unsupervised learning algorithm. In the same way, neuroscientists can reason about the functions of simple strips of a certain types of preferred connections. The function of such simple arrangements may be understood without before we understand the mechanism of their generation. For example the connection arrangements for motion detection [Ko et al., 2013] are understood functionally but the precise mechanism of their generation is still a matter of debate.

For practical reasons, AI studies of neural networks have focused on algorithms to learn the large number of neural network parameters from labeled data with Deep Learning algorithms. If the ideas behind the OC model are correct, then evolution has found developmental mechanisms to create simple patterns that perform some of the same function without learning from experience. The approach in this study is to use models with as few parameters as possible, so that a non-gradient based optimization and a small amount of deliberate manual tuning sufficed to find adequate parameterization to match the data. In contrast, a recent proposal is that "goal driven" computational neuroscience studies should use Deep Learning [Yamins et al., 2014, Kietzmann et al., 2019]. These studies can be seen as implying that models with designed parameters are retrograde for neuroscience, just as it is for artificial intelligence. But this study shows that biophysicists seeking mechanistic explanations for cognitive processes should favor parsimony in the underlying model over convenience in finding useful parameterizations. The model in this study uses far fewer parameters than DL networks and still reproduces neurophysiological results very well. Whether or not this particular embodiment of Organic Convolution is exactly right, this is evidence for the "critique of pure learning" of [Zador, 2019], that much of brain function must come from simple structures encoded in the genome.

This study is a demonstration of a novel method based on GPU hardware for simulating waves of individual spikes in neural networks with connection patterns that can be modeled by convolution. While most of the issues raised in this discussion are not resolved definitively, it is clear that convolutional neural networks and tensor technologies can be used to experiment with models of afferent connection patterns in spiking neural networks at scale. Such models may be quite simple and still help explain neurophysiological data.

3.2 Ideas and Speculation

Proof of the Organic Convolution model could come from direct observation using connectomics techniques, for review see e.g. [Behrens and Sporns, 2012]. With the connectomics technologies that have become available in the past decade, the details of the V1, V2 and V4 feedforward connections could be mapped at different stages of development to provide direct evidence for the number and kind of connections that exist. The results in this study suggest it may be especially interesting to analyze the geometric arrangements of connections in topographically organized cortical afferents. Also, additional confirmation of Organic Convolution would require understanding of the mechanisms that guide such pattern formation, as discussed in [Shirasaki and Pfaff, 2002]).

The coding used in the convolution model is a population place code: each unit responds to the presence of a particular pattern of inputs at a particular location in the visual field, albeit with

broad tuning, and there are functionally identical units tiling the receptive field. Given the place code in the OC model, the functioning of the model is essentially the same whether considered as a binary network or a spiking network, due to the long time scale of integration and tight synchronization of the input. If the OC model is an accurate representation of brain activity, it suggests that the venerable McCullough-Pitts neuron model [McCulloch and Pitts, 1943] may go a long way in explaining important classes of brain activity. Also note that the OC model is compatible with earliest arrival spike time coding proposed in [Thorpe et al., 2001]: If the inputs to area V1 in the OC model are ordered so that the earliest arrivals are from the most intense stimuli, it may enhance the place code defined by the connection patterns. Models combining OC with spike time coding is an attractive future area of research.

The experiments in this study model a single cascade of spikes as the basic unit of computation, and such a system will manifest rate coding as well as place coding: If there are repeated waves of activity in response to a single stimuli, the place coded neurons best matching a stimuli would fire with the fewest failures and at a higher overall rate. But there is an important distinction between the meaning of "coding" in these two usages: The topographic place code is a mechanistic element of the computation because the patterns of connections rely on locality - without the place coding, the computation would not work. The term rate coding used here means that information can be read out from the network by the rate. But the rate is not a part of the computation: the biomechanics of the computation relies on the synchronized cascade. If each neuron spikes at most once, there is no rate code at all in the elementary computation. This raises an interesting new challenge to the biological plausibility of a common DCNN methodology: the RelU activation function. The Relu function is typically justified as biologically plausible because it is taken to represent a neural unit which has rates as inputs and a rate as the output - the RelU hinge is taken as increasing output rate once the stimuli match to the convolution filters is above the threshold. But if it is true that the biological mechanism of neuronal computation does not take rates as inputs and produces a rate as an output, a RelU activation neural network is not a biologically plausible neural network.

An important related question is precisely how Organic Convolution is used to achieve object recognition in IT. The convolution filters used in the OC model in this study are not exactly like those discovered in trained machine learning models, but they may still serve a similar purpose. Initial experiments suggest a network like the OC model may be used to perform machine learning tasks by feeding the final convolution output into a fully connected network that is trained with gradient descent [LeCun et al., 2015]. If OC is found to be a good representation of biology, and it can also function in object recognition applications, it may serve as a bridge for understanding the relationship between the current generation of artificial neural networks and the brain.

4 Materials and Methods

4.1 Spiking Neuron Model

The spiking neuron model used for the Spike Time Convolution (STC) model is the Spike Response Model (SRM) described in [Gerstner and Kistler, 2002]. The SRM model is an abstraction of the integrate and fire neuronal biophysics but it captures the key properties of spike integration and timing for single action potentials. Each model neuron has a potential level given by a the sum of the excitatory and inhibitory post synaptic potentials (PSP). The time course of the PSP's are described by a difference of decaying exponentials. For any given synapse (indexed by i) and given an input spike arrival time s the influence of the PSP is given by the kernel:

$$\epsilon_i(t \mid s) = \frac{w_i}{Z}[e^{(\frac{t-s}{\tau_{fall}})} - e^{(\frac{t-s}{\tau_{rise}})}]\Theta(t-s) \qquad (1)$$

In equation 1 w_i represents the synaptic input sign and weight. Weights are positive for excitatory and negative for inhibitory, depending on the type of the afferent input and the weight is determined by the micro-architectural connection pattern, as described below; Z is a normalization constant chosen so that the maximum value of the PSP is w_i; $\Theta(x)$ denotes the Heaviside step

Table 4 PSP Time Constants (ms) and properties for Equation 1; Source: [Holmgren et al., 2003]

Type	τ_{rise}	τ_{fall}	10-90% rise time	Width at half-amplitude
Excitatory Pyramidal / Stellate	4.5	30	5.3	36
Inhibitory Fast Spiking	1.9	13	2.2	15

Table 5 Arrival & Transmission Gamma Distribution Time Constants and properties (ms)

Type	α (shape)	β (scale)	Mean	Std. Dev.	90th Pcntl.	Source
Arrival	2	1	2	1.41	3.9	[Singer, 1999]
Inter-Area	3	0.5	1.5	0.87	2.7	[Plomp et al., 2017]
L4 - L2/3	1.2	0.7	0.8	0.77	1.9	[Girard et al., 2001]

function. The simulations of each wave of cortical activity lasted 10 ms, with a discretization step of 0.2 ms in the simulated PSP dynamics.

Given the PSP kernel function in equation 1 and a set of input spike times on a set S of afferent inputs indexed by i the neuron spikes at the time

$$\arg\min_{t} \sum_{i \in S} \epsilon_i(t \mid s_i) >= 1.0 \tag{2}$$

where the the threshold in equation 2 is set to 1.0 for model interpretability: with this definition, all synaptic weights w in the convolution kernels (Figures B5, B6, 12, 10 and 6) are expressed in terms of the positive excitatory threshold. The constant Z in equation 1 is determined by first calculating the time that PSP achieves its maximum, as described in the Appendix.

Time constants for the PSP's in the model are based on [Holmgren et al., 2003] and shown in Table 4. [Holmgren et al., 2003] showed that the time constant of both EPSP and IPSP in pyramidal neurons are approximately the same in rise time and differ primarily in decay time (these findings had large error bars around the mean). The decay times are so long that they do not really affect the spike time in a feedforward cascade like those modeled in this study, so for simplicity this model used the same time constants shown labeled "Excitatory" for EPSP and IPSP in all of the excitatory neurons in the model ("Excitatory" in 4 refers to the neuron, not the synaptic input.) [Holmgren et al., 2003] found that the Excitatory connections onto FS spiking interneurons had a much faster rise time and decay and they are approximated by the parameters labeled "Inhibitory" in table 4 ("Inhibitory" refers to the excitatory synapse in the FS inhibitory cells, not the inhibitory synapse on the pyramidal cell.)

Variability in the amplitude of individual cortical pyramidal cells is high: [Seeman et al., 2018] found the coefficient of variation (CV) to be approximately 0.6 in individual cells. However the units in this study represent idealized "average" units in which one model unit represents an average of a number of real neurons in mammalian visual cortex. For that reason the OC model uses a more conservative value of 0.25 CV on all PSP, which is embodied by multiplying the PSP value by a random value drawn from $N(1, 0.25)$. See Section 4.3 for details of the PSP implementation as a tensor calculation.

4.1.1 Input Model and Spike Transmission

The first layer of the network represents V1 L4 receiving input from LGN but these neurons are not modeled explicitly: V1 L4 units spike with a probability equal to the image intensity. The full PSP model of equation 2 begins with the models for V1 inhibitory and L2/3 pyramidal neurons. Because the stimuli (Figure 2) are black and white, in practice this means most V1 L4 neurons in the model spike deterministically with respect to a single image - only units representing an anti-aliased boundary in the stimuli have mid range intensity that lead to significant trial to trial variability in the response to the same image.

Although the occurrence of spikes in V1 L4 of the model is largely deterministic to the study stimuli, the timing of the V1 L4 spikes are random in every trial. The V1 L4 unit spike times are drawn from a gamma distribution with parameters in table 5. The arrival times are based on

Table 6 Network Summary

Model Layer	Model For	Channel Size	Connection Patterns	Orients.	Inhibitory Units?	Total Channels	Equivalent Neurons
1	V1 L4	112×112	1	1	Y	1	12,544
2	V1 L2/3 Inhib	112×112	1	1	Y	1	12,544
3	V1 L2/3 Pyr	112×112	2	16	N	24	301,560
4	V2 L4	56×56	1	8	Y	8	25,088
5	V2 L2/3 Inhib	56×56	1	8	Y	8	25,088
6	V2 L2/3 Pyr	56×56	7	16	N	104	326,144
7	V4 L4	28×28	1	16	Y	104	81,536
8	V4 L2/3 Inhib	28×28	1	16	Y	104	81,536
9	V4 L2/3 Pyr	28×28	7	16	N	88	37,632
						Total	903,168

findings from [Singer, 1999] showing that inputs from LGN to V1 L4 arrive with a synchronization window of around 5ms.

The delays for transmission time between layers in the same cortical area and between cortical areas are also modeled by gamma distributions with time constants in table 5. The L4 to L2/3 timing is based on [Girard et al., 2001] which show a mean of less than 1 ms transmission delay on average. The L4 - L2/3 delay distribution is used for V4 to Inhibitory, V4 to L2/3 Excitatory and V4 L2/3 Inhibitory to Excitatory delays. The transmission time between areas of cortex was studied in [Plomp et al., 2017] which found around 1.5 ms transmission delay. A distribution based on [Plomp et al., 2017] is used for transmission between L2/3 excitatory neurons and L4 neurons in the next cortical area model.

4.2 Network Model

The model network consisted of nine layers, representing L4 and L2/3 in cortical areas V1, V2, and V4 (inhibitory and excitatory L2/3 neurons are considered separate layers). As shown in Table 6, each layer represents a number of model neurons that is the product of the layer area and the number of different types of units at each position. The layer area starts as the number of pixels in the input layer, and at deeper layers the dimension shrinks due to pooling operations. The number of units at each position in the model is given by the number of connection patterns and the number of orientations at which the pattern is instantiated. Each unit type is modeled as a channel in the tensor simulation as described below. The total number of equivalent neurons represented by the model is approximately 900,000.

As explained in the Introduction, the model focuses on a simple feedforward wave of spikes and does not include other details of the cortical circuitry, such as lateral connections within cortical layers [Kisvárday et al., 1997] or cortical microcircuits to L5/6 [Douglas and Martin, 2004]. The model also does not include inter-cortical circuits giving feedback from V4 to V2 and V1 and from V2 to V1, nor the interaction between feedback and lateral circuits [Liang et al., 2017].

Conceptually, the model assumes that the feedforward dynamics in the ventral pathway is driven by episodic arrival of inputs in V1 from the lateral geniculate nucleus (LGN), synchronized with precision on the order of a few milliseconds, which has been observed in multiple studies, as reviewed in [Singer, 1999]. The model as implemented is memoryless in that it only models single episodes of activity.

4.3 Tensor Simulations

4.3.1 Convolution With a Binary Neuron Model

As an introduction to the Spike Time Convolution (STC) model, listing 4.2 is the Python code for a convolutional neural network layer in which the model neuron is a simple binary threshold unit with no time dependence [McCulloch and Pitts, 1943]. Listing 4.2 uses the `tensorflow` and `keras` packages: The class `BinaryLayer` descends from `tf.keras.layers.Layer`. `BinaryLayer` has convolution filter weights (and a stride) defined at

Listing 1 Binary Convolution

```
import numpy as np
import tensorflow as tf

class BinaryLayer(tf.keras.layers.Layer):
    def __init__(self, conv_weights : np.array,
                        stride : int):
        self.conv_weights= tf.constant(conv_weights)
        self.stride = stride

    def call(self, inputs : tf.Tensor) -> tf.Tensor:
        activation = tf.nn.conv2d(input=inputs,
                                  filters=self.conv_weights,
                                  padding='SAME',
                                  strides=[1, self.stride, self.stride, 1])
        spiked = tf.greater_equal(activation, 1.0)
        output = tf.cast(spiked, tf.float32)
        return output


def binary_simulation(binary_layers : list[BinaryLayer],
                      stimuli : tf.data.Dataset):
    model = tf.keras.Sequential()
    for layer in binary_layers:
        model.add(layer)
    model.compile()
    output = model.predict(stimuli)
```

initialization. The `call` function in Listing 4.2 defines the transformation from input tensors to output tensors for the layer. The convolution itself is performed with the `tensorflow` model function `nn.conv2d`. To perform a simulation, multiple `BinaryLayer`'s are added to a `tf.keras.Sequential` model object, as shown on lines 20-26 of Listing 4.2. After adding the layers, the model is compiled and outputs can be generated from a `tf.data.Dataset` holding the stimuli.

The code shown in Listings 4.2 is a simplification of that used for the simulations shown in the Results section in one important respect: Listing 4.2 perform the convolution using a single convolution filter size (lines 11-14). It is conventional to use a single filter size in DCNN applications for machine learning. For the Organic Convolution experiments described here, the filter sizes are determined by the receptive fields observed in recordings and they vary depending on the unit, particularly in the V1 and V4 models.

4.3.2 Spike Time Convolution With the Spike Response Model

Python code to model one layer of a Spike Time Convolutional neural network implementing the Spike Response Model is shown in Listing 2. The layer class `SpikingLayer` inherits from the `BinaryLayer` in Listing 4.2 and shares the same code for the convolution operation. The `SpikingLayer` has additional parameters in the initializer for the PSP time constants (τ_{rise} and τ_{fall} in Equation 1) and a flag indicating whether the inputs include inhibitory units. Note that the `SpikingLayer` has a class variable that is used to hold a reference to the current simulation time (Listing 2 line 6). The full details of the `SpikingLayer` initialization are omitted from Listing 2 for brevity; these are copying the parameters to member variables, and calculating the PSP scaling constant given by Equations B1 and B2.

Neural network layers in the `keras` framework have the option to define a `build` function

Listing 2 Spike Time Convolution with the Spike Response Model - TODO Update with separate inhibitory & delays

```python
import numpy as np
import tensorflow as tf
from binary_convolution import BinaryLayer

class SpikingLayer(BinaryLayer):
    current_time = None # Class variable for simulation time

    def __init__(self, psp_trise : float, psp_tfall : float,
                 inhibitory_inputs : bool, **kwargs):
        BinaryLayer.__init__(self, **kwargs)
        # Save scalar variables and calculate scale (not shown)

    # input_shape : batch size, y dimension, x dimension, number of channels
    def build(self, input_shape : tuple[int, int, int, int]):
        # Tensor of -1.0/time constants and scale for use in the PSP calculation
        self.psp_fall_z = tf.scalar_mul(-1.0/self.psp_tfall, tf.ones(input_shape))
        self.psp_rise_z = tf.scalar_mul(-1.0/self.psp_trise, tf.ones(input_shape))
        self.psp_scale = tf.scalar_mul(self.psp_scale, tf.ones(input_shape))
        # Initialize a tensor holding the input sign (excitatory or inhibitory)
        if not self.inhibitory_inputs:
            self.input_signs = tf.ones(input_shape)
        else:
            excite_signs = tf.ones(shape=(input_shape[0],input_shape[1],
                                          input_shape[2], input_shape[3]/2))
            inhib_signs = tf.scalar_mul(-1.0, excite_signs)
            self.input_signs = tf.concat([excite_signs, inhib_signs], axis=3)
        # Initialize a tensor for output spike times; 0 indicates no spike
        output_shape = (input_shape[0], input_shape[1]/self.stride,
                        input_shape[2]/self.stride, self.conv_weights.shape[3])
        self.spike_times = tf.zeros(shape=output_shape)

    def call(self, inputs : tf.Tensor) -> tf.Tensor:
        # Calculate time delta since spike arrival
        delta_t = tf.subtract(SpikingLayer.current_time, inputs)
        # if past the spike time, psp=scale*exp(-dt/tfall)-exp(-dt/trise)
        psp_unsigned = tf.where(tf.logical_and(tf.greater(delta_t, 0.0),
                                               tf.greater(inputs, 0.0))
                                tf.multiply(self.psp_scale, tf.subtract(
                                    tf.exp(tf.multiply(self.psp_fall_z, delta_t)),
                                    tf.exp(tf.multiply(self.psp_rise_z, delta_t)))),
                                tf.zeros(input_shape))
        # Apply the signs of the input layer to the PSP's
        psp_signed = tf.multiply(psp_unsigned,self.input_signs)
        # Call to the base class binary convolution using the PSP's as inputs
        conv_output = super(SpikingLayer,self).call(psp_signed)
        # Find newly spiking neurons
        new_spike = tf.logical_and(tf.equal(conv_output, 1.0),
                                   tf.equal(self.spike_times, 0.0))
        # Add new spike times to the outputs
        current_t = tf.scalar_mul(SpikingLayer.current_time,
            tf.ones_like(self.spike_times))
        self.spike_times = tf.where(new_spike, current_t, self.spike_times)
        return self.spike_times
```

Listing 3 Spike Time Convolution Simulation

```
import tensorflow as tf
from spiking_convolution import SpikingLayer


def spiking_simulation(spiking_layers : list[SpikingLayer],
                       stimuli : tf.data.Dataset,
                       input_shape : tuple[int,int,int,int]
                       dt : float, tmax : float):
    # Set input layer spike times
    spiking_layers[0].spike_times = tf.random.gamma(shape=input_shape,
                                                    alpha=gamma_shape,
                                                    beta=1.0 / gamma_scale)
    # Define first input and output reference
    first_input = tf.keras.Input(shape=input_shape[1:])
    output = None
    # Loop incrementing current time and calling layers in sequence
    for tidx, t in enumerate(np.arange(dt, tmax, dt)):
        # Set current time static variable to the tensor constant
        SpikingLayer.current_time = tf.constant(t)
        for layer_num in range(0, self.numLayers):
            input_for_layer = output if layer_num > 0 else first_input
            # Call the layer using the `()` operator
            output = spiking_layers[layer_num](input_for_layer)
    # Define the model from first input to final output
    model = tf.keras.Model(inputs=first_input, outputs=output)
    # Compile the model
    model.compile()
    # Generate output
    output = model.predict(stimuli)
```

that performs additional setup after initialization, given the input shape. (The input shape is unknown at initialization.) The `SpikingLayer` defines a `build` function to perform the following setup tasks:

1. Define tensor versions of the PSP time constants having the same shape as the input (lines 16-18)
2. For layers with both excitatory and inhibitory input, create a tensor of signs to apply to the PSP summation so that the inhibitory units reduce activation (lines 19-26).
3. Create a tensor variable to hold the spike times for every unit in the layer (lines 28-30).

The `call` function of the `SpikingLayer` takes an input tensor that represents the spike times of the units in the previous layer. The spike time tensor stores zero to indicate no spike. As will be explained below, the simulation starts at a time one discretization step after 0, and the class variable `current_time` is set before the `call` method is invoked (Listing 3).

In `call` for the `SpikingLayer`, the main steps are as follows:

1. Calculate the time since the spike of each unit that has fired (line 34); negative values result for inputs that have not spiked, but these are screened out at the next step.
2. Calculate the current PSP activations with the SRM model (line 36). This includes logic equivalent to the step function in equation 1.
3. For those inputs that are inhibitory, apply the required sign to the PSP activation (line 42)
4. Invoke the base class `call` function which performs the convolution with the current PSP values as the inputs (line 44). This step converts the input dimension to the output dimension, and emits a 1 for every neuron that has fired.

5. Find what inputs have newly fired by comparing the current convolution outputs to the previous spike times and store it in another tensor (line 46).
6. Store the current time in the member variable `spike_times` for all of the units that have newly spiked (lines 49-52)
7. Return the newly updated member variable of spike times as the new output vector (line 53)

After defining the layers, the tensor graph for a complete spiking simulation must be created before passing in inputs and executing the graph, which is shown in Listing 3. The first step in the simulation graph is to set the input times in the first `SpikingLayer` (the input layer) to random times drawn from a gamma distribution. The main tensor graph is created in a double loop over time, and layer (lines 17-20.) The first step in the simulation loop for a point in time is to set the `current_time` class variable to a tensor constant for use in all the layer calculations (Listing 3 line 19.) Then the layers are called in succession with the output of each layer as the input to the next (lines 20-23); note that for `keras.Layer` the `call` function is overloaded with the `()` operator and does not need to be named explicitly (line 23).

To perform a simulation, a base class `keras.Model` is created from the first input and the final output; the standard `tf.keras.Sequential` model cannot be used for this specialized use case. After creation, the `Model` can be used to generate outputs as usual by first compiling the model, and then generating outputs with the `Model.predict` function.

Note that the `BinaryLayer` shown in listing 4.2 is not just an illustrative abstraction: During research and development, connection patterns were first tested with a network of binary neurons just like in listing 4.2. The behavior of the binary model is a non-random version of the behavior of the spiking model, and the speed of the binary simulations makes it useful for prototyping. After prototyping with a binary model, the weights were increased so the units would spike under noisy conditions.

4.4 Model tuning

The form of convolution filters are determined by the model theory of visual processing, but the precise value for excitatory and inhibitory weights had to be tuned. Tuning was performed using a recent open source hyper-parameter tuning package [Akiba et al., 2019] that implemented the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [Hansen, 2006]. TODO: Details on exactly what was optimized and how, and what was hand tuned.

Acknowledgments. Thanks to John Reynolds and Anirvan Nandy for allowing reproduction of their figures and for helpful discussions and suggestions.

Competing Interests. Some machine learning applications of the techniques described here may be covered by filed patent applications U.S. Pat. App. Ser. No. 16/125,818 and provisional application 63337083. The code and methods will always be freely available for academic research as described below.

Code Availability. Upon publication the source code from this study will be available at http://github.com/carl24k; during the review process the source code is available by request.

Appendix A Network Response Details

Figure A1, A2 and A3 illustrates the response of the entire network to low, medium and high curvature stimuli in the central position in the visual field. The sub-plots show the responses of the L4 pool units corresponding to each type of connection pattern in the model. The plots for V1 (row B) and V2 (rows E through O) are zoomed by 2.3X and 2X respectively. The following annotation key serves for all three figures:

- **A** Heatmap of average input intensity.
- **B** Pictogram representation of the preferred stimulus of the V1 units. See Figure B5 for details.
- **C** Heatmap of firing rate of the V2 L4 bar selective unit response measured in spikes per 100 simulations.
- **D** Pictogram for the preferred stimulus of the ultralong units. See Figure B6 for details.
- **E** Heatmap of the V4 L4 ultralong selective pool unit responses.

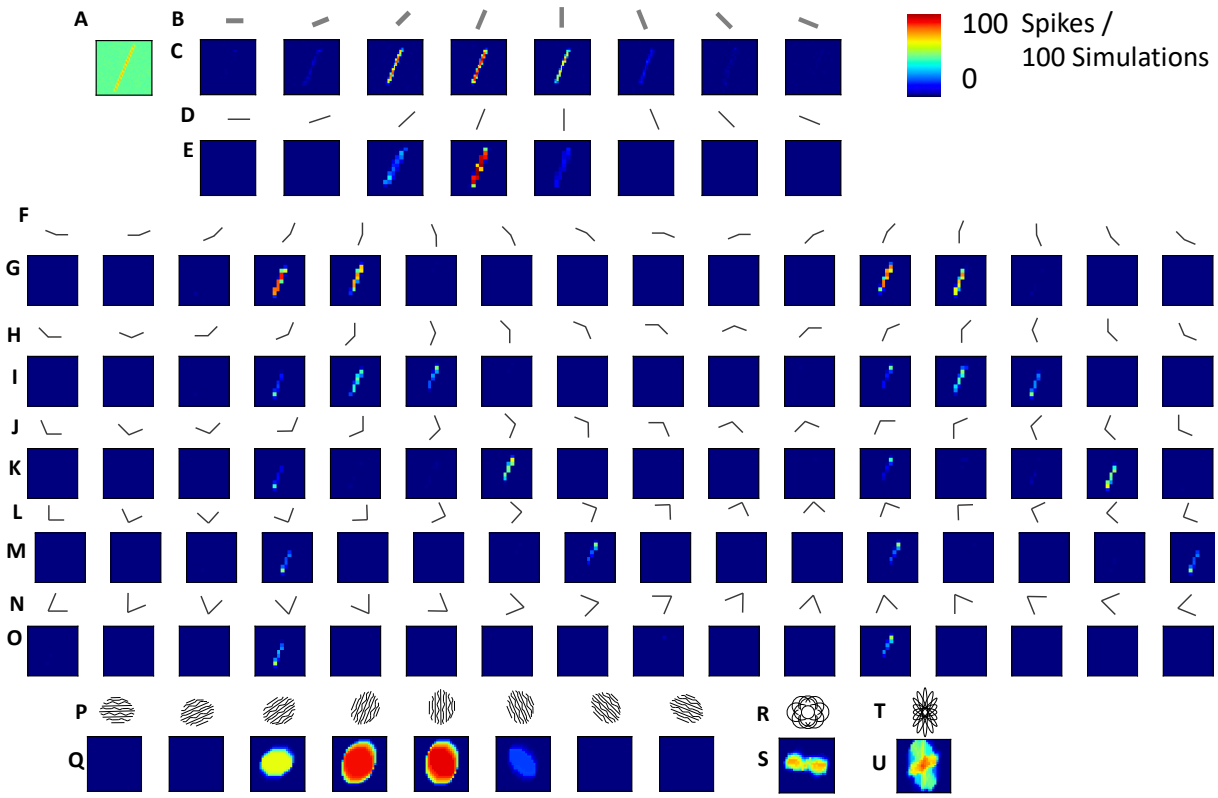

Fig. A1 Illustration of Network Response to Low Curvature Stimuli. See text for annotation key.

- **F** Pictograms for the preferred stimulus of complex-shaped units having 22.5° angle between sub-units.
- **G** Heatmap of the V4 L4 22.5° complex-shaped selective pool unit responses.
- **H** Pictograms for the preferred stimulus of complex-shaped units having 45° angle between sub-units.
- **I** Heatmap of the V4 L4 45° complex-shaped selective pool unit responses.
- **J** Pictograms for the preferred stimulus of complex-shaped units having 67.5° angle between sub-units.
- **K** Heatmap of the V4 L4 67.5° complex-shaped selective pool unit responses.
- **L** Pictograms for the preferred stimulus of complex-shaped units having 90° angle between sub-units.
- **M** Heatmap of the V4 L4 90° complex-shaped selective pool unit responses.
- **N** Pictograms for the preferred stimulus of complex-shaped units having 112.5° angle between sub-units.
- **O** Heatmap of the V4 L4 112.5° complex-shaped selective pool unit responses.
- **P** Pictogram representation of the preferred stimulus of the V4 low curvature selective unit with 8 different preferred orientations. These units are defined by the connection patterns and convolution kernels illustrated in Figure 12.
- **Q** Heatmap of the V4 low curvature selective unit responses.
- **R** Pictogram representation of the preferred stimulus of the V4 medium curvature selective unit. These units are defined by the connection patterns and convolution kernels illustrated in Figured 10.
- **S** Heatmap of the V4 medium curvature selective unit responses.
- **T** Pictogram representation of the preferred stimulus of the V4 high curvature selective units. These units are defined by the connection

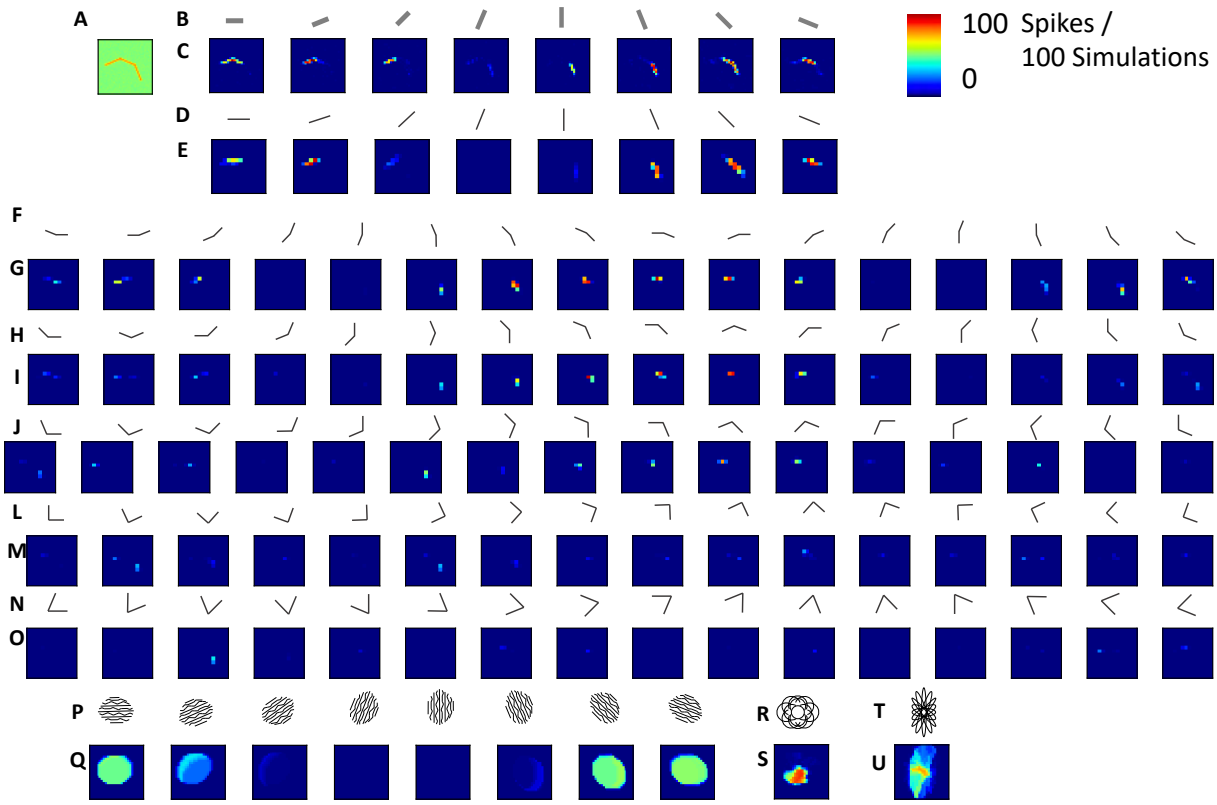

Fig. A2 Illustration of Network Response to Medium Curvature Stimuli. See text for annotation key.

patterns and convolution kernels illustrated in Figured 6.

- **U** Heatmap of the V4 high curvature selective unit responses.

Appendix B Supplementary Methods

B.1 Convolution Models for Patterns of Afferent Connections

Topographically organized connection patterns patterns are used to determine the convolution filter weights in the model. The patterns are implemented as rules giving the input preference for dendritic locations as a function of the position relative to the soma, using trigonometric functions. It is assumed that individual neurons have a baseline orientation property referring to a direction in the topographic map. The dendrite's preferences are determined relative to the neuron's baseline, which facilitates instantiating multiple versions of a pattern with different orientation preferences.

Given a rule defining the pattern, the following process was used to creates the convolution filters for the model:

1. Pick an ellipse defining the receptive field at the baseline orientation; points within the rectangular convolution filter dimensions but outside of the ellipse boundary are set to zero.
2. Dendrite paths are determined with a 2-dimensional[4] ball and stick model, traveling from the center of the receptive field to the edge of the ellipse.
3. For each rotated version of the pattern, the ellipse and all of the dendritic points are rotated to the new orientation. The smallest rectangle bounding the rotated ellipse defines the convolution filters for that channel.
4. The functional micro-architectural rule is applied at sample points along the dendrites to

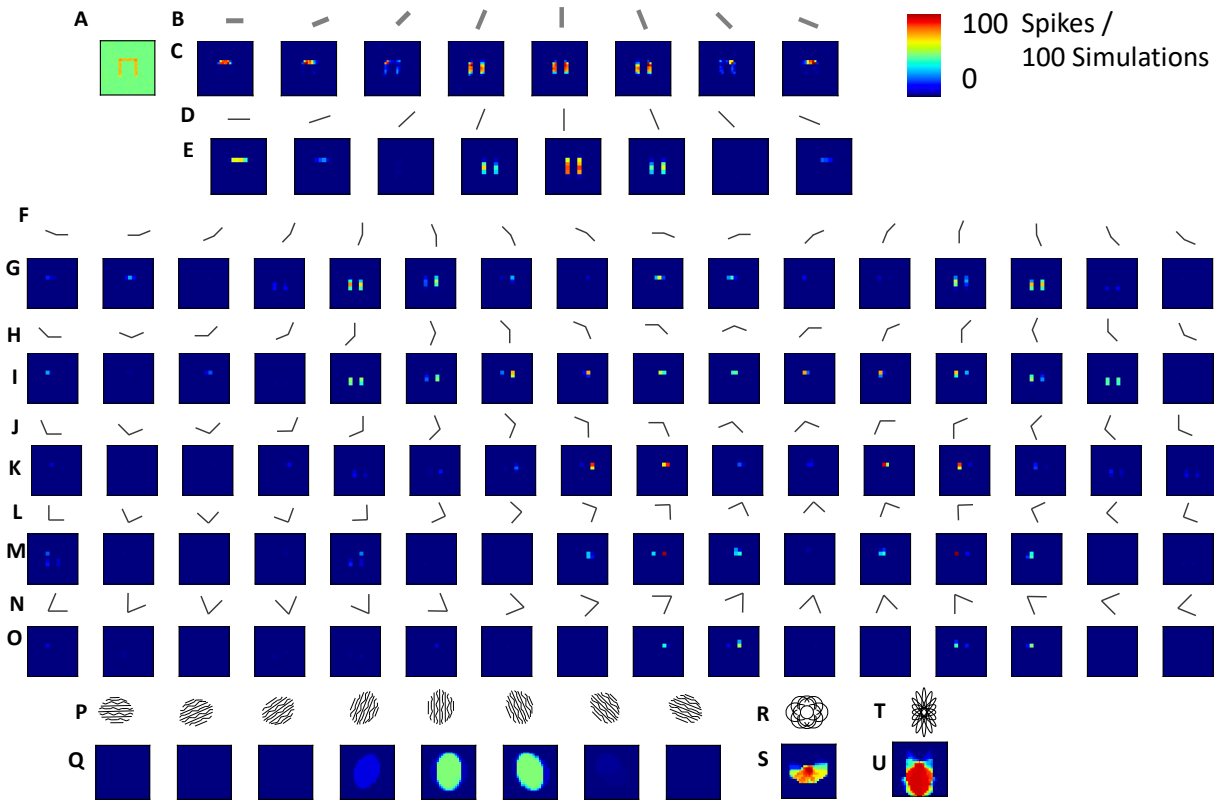

Fig. A3 Illustration of Network Response to High Curvature Stimuli. See text for annotation key.

Table B1 Connection Pattern Summary

Type	Receptive Field	# Dendrite	Excit-ation	Inhib-ition	Pattern Description
V1 L2/3	6×9	32	2.4	1.2	Gabor-like
V2 L4	2×2	8	4	0	Pool (Or)
V2 L2/3 Ultralong	9×9	32	3.2	7.7	Long Gabor-like
V2 L2/3 Complex	9×9	32	2.4	5.3	Gabor-like sub-units
V4 UL/C Pool	2×2	8	4	0	Pool (Or)
V4 Low Curve	15×20	96	164	52	Constant, low angle
V4 Medium Curve	16×22	96	507	185	Convex, medium angle
V4 High Curve	16×32	96	1769	868	Convex, high angle

determine the weights for every point in every available input channel.

The size and strength of the input connections for the functional micro-architectural patterns are summarized in Table B1.

B.2 V1 Model

The connection patterns for V1 L2/3 units can be extremely simple (no pun intended) and still reproduce the complex shape tuning of V4 recordings: The connection pattern is represented by a

[4]For all connection patterns discussed in this study the arrangement of the inputs in the direction of lower and higher cortical layers is not relevant to the operation of the micro-architecture. Only the layout of afferent inputs in the plane of the topographic map, or parallel to the layers of cortex, is relevant to the functioning described and the model only represents input layout in that plane.

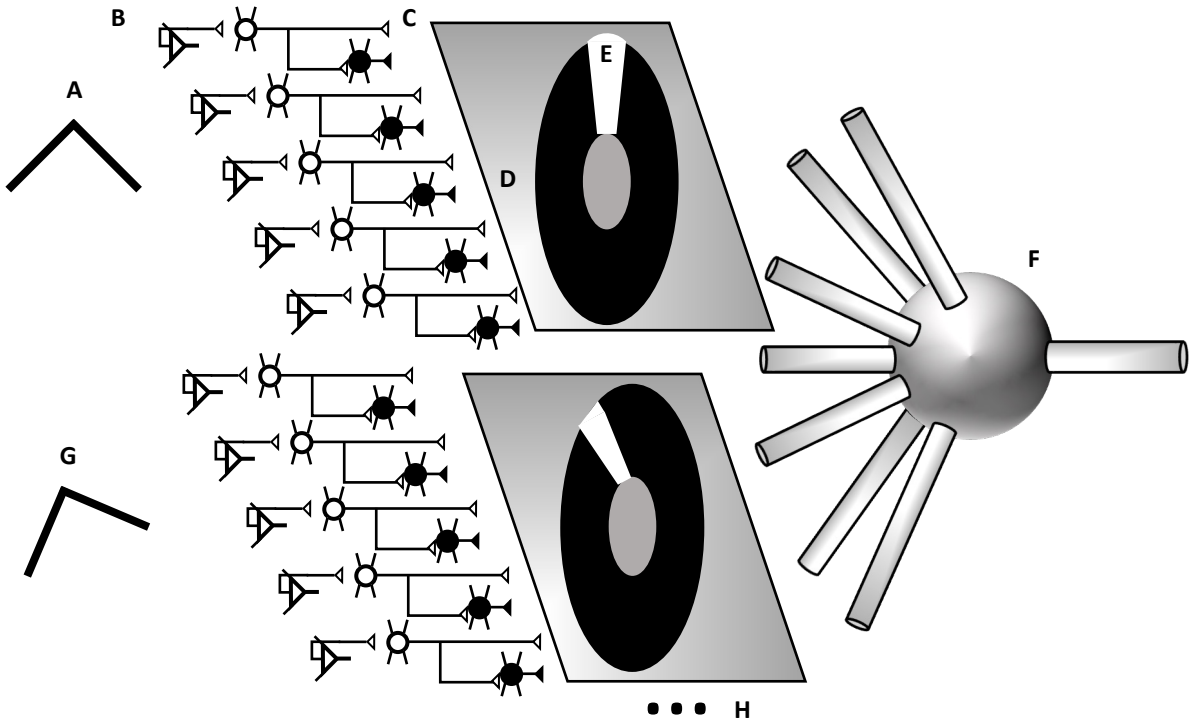

Fig. B4 Afferent Connection pattern illustration for a V4 curvature selective neuron. **A** V2 contains neurons selective to corners with a variety of interior angles and orientations in visual space, e.g. 90° corners oriented up. **B** Each type of selectivity is repeated in units covering the visual space in a topographically organized manner. Such relations are maintained as the signal is passed to V4 L4 neurons, and V4 fast spiking interneurons. **C** The topographically organized excitatory and inhibitory afferents are available to V4 L2/3 neurons. Afferents are selected to form patterns of inhibition **D** and excitation **E** by the basal dendrites and proximal apical dendrites which extend radially in the topographic plain from the V4 L2/3 pyramidal neuron soma. **G** The same systematic arrangements repeat for the other interior angles and rotations of selectivity available from V2. The same framework also applies to afferent inputs from the LGN in V1 and afferents from V1 in V2.

6×9 convolution filter with a 3 pixel wide band of excitation in the center surrounded by inhibition, as illustrated in figure B5. Figure B5**A** also illustrates how this pattern could be created by having a few dendrites lying close to an axis in the topographic space make (primarily) excitatory connections to afferent inputs from L4, while dendrites away from the axis are (primarily) inhibitory. The model includes 8 different orientations of the same pattern, 22.5° degrees apart in the orientation of the optimal stimulus.

To reproduce the curvature and orientation tuning found in V4, the V1 L2/3 units were excitable enough to have broad tuning. The ratio of excitatory to inhibitory weight was 60% to 40%. This is illustrated in Figure B5, which shows weights in units of the neuron threshold: Defining the neuron's threshold as 1, the total excitatory weight in 6×9 convolution filter was 3.3 and the total inhibitory weight was 2.1. The excitatory weight was high enough that if only 40% of the inputs were active, it was enough to cause the unit to spike.

B.3 V2 Model

The V2 model is inspired by the findings in [Liu et al., 2016] and the V2 L2/3 units in the OC model implements both ultralong and complex-shaped selectivities described therein. The OC model further hypothesizes one underlying connection pattern pattern may lead to both selectivities: It is assumed that these V2 units have a small number of dendrites receiving feedforward excitation, those dendrite extend radially outward in the topographically organized cortical plane, and prefer to connect to excitatory V1 afferents that have the same preferred stimulus orientation as the direction of the dendrite itself. If a unit has two such dendrites that extend in a direction from the soma that is substantially opposite to each

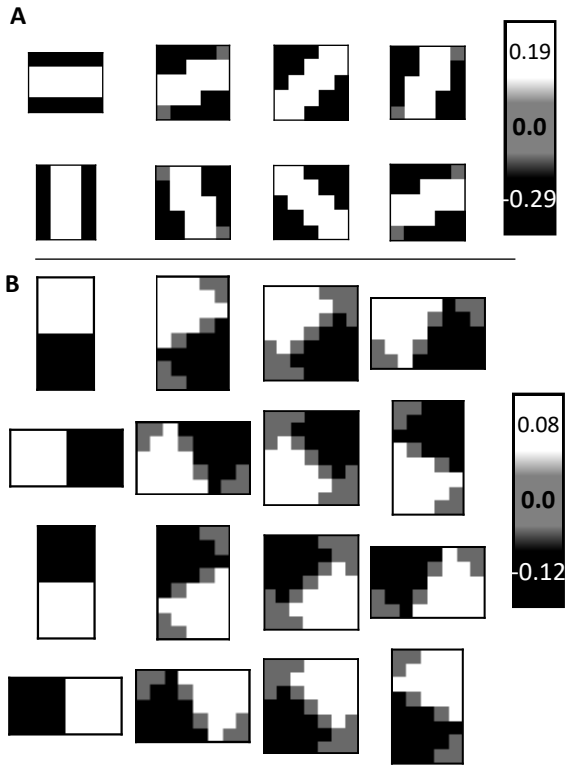

Fig. B5 Afferent Connection Patterns for V1 model neurons. The convolution pattern represents the afferent connections to excitatory and inhibitory inputs as illustrated in Figure B4. **A** The synaptic connection strength versus position in topographic space relative to the soma at the center for V1 neurons selective to lines/bands of intensity. There are 8 different orientations, oriented 22.5° apart. The filters are 4x5 (horizontal/vertical) pixels or 5x5 pixels (others). **B** The synaptic connection strength versus position in topographic space relative to the soma at the center for V1 neurons selective to edges in the intensity. There are 16 different orientations oriented 22.5° apart. The filters are 4x8 (horizontal/verical), 7x7 pixels (45° rotations) and 5x8 (others).

other, they will form an ultralong receptive field; this arrangement is illustrated in Figure B6A. Inhibitory inputs are connected for orientations other than the ones preferred by the excitatory dendrites, and there are other dendrites whose role in the feedforward circuit is predominantly inhibitory.

Using the same approach, if a unit has two dendrites with the same connection preference, but in the case where those dendrites have some angle to each other the result will be a complex-shaped receptive field; this arrangement is illustrated for two dendrites at 90° to each other in Figure B6F. These units prefer stimuli that resembles two lines meeting in a corner, and the angle between the two excitatory dendrites sets the angle preference for the unit. The 9×9 convolution filters resulting from some instantiations of the V2 L2/3 connection pattern are also illustrated in Figures B6: Excitatory connections are concentrated on a few orientations in simple straight or angular arrangements. Inhibition is diffuse and spread over the receptive field.

The model include 8 different orientation preferences for ultralong units, oriented 22.5° apart. The model also includes a range of angles defining the separation between the sub-units of the complex-shaped units, from 22.5° to 112.5° in 22.5° increments for a total of 5 complex-shaped sub-unit angles. Because the complex-shaped units are not symmetric to 180° rotations, 16 different orientations (22.5° apart) of each sub-unit separating angle are defined to respond to the full range of possible stimuli.

The best fit for V4 shape selectivity required the V2 neurons to have strong excitation, but in these cases the total inhibition was greater: The total excitatory weight in the 9×9 receptive fields for V2 ultralong units was 3.2× the threshold, while the total inhibition was 7.7× the threshold. For the complex-shaped units the excitation was 2.3 times the threshold, and the total inhibitory weight is 5.2.

B.4 Layer 4 Pooling Models

The first layer of the model serves as a real valued input layer to the neural network and is used as a lumped model for the eye/retina, LGN, V1 L4α and V1 L4B [Vanni et al., 2020]. An unusual property of the input layer in this model in comparison to a standard DCNN is that it contains a redundant set of excitatory and inhibitory units - this is modeled after V1 L4B which makes both excitatory and inhibitory projections to V1 L2/3. The excitatory and inhibitory units are redundant in the input intensity they receive from the stimuli, but have independent random spike arrival times. With this redundant structure, the V1 L2/3 layer receives both excitatory and inhibitory input from every position in the input while still obeying Dale's Law. In the convolutional model, this translates to two separate channels. The model ignores that there should be different time constants for

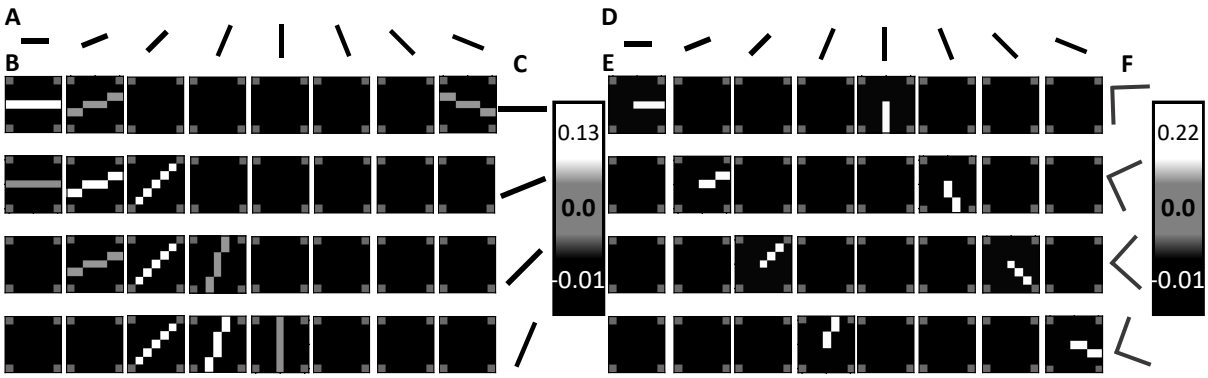

Fig. B6 Afferent Connection Patterns for model V2 L2/3 units. The convolution pattern represents the afferent connections to excitatory and inhibitory inputs as illustrated in Figure B4. **A-C** Illustration of a connection pattern for an ultra-long type V2 unit. **D-F** Illustration of a connection pattern for a complex type V2 unit having an interior angle of 90°. **A** & **D** Pictorial representation of the preferred stimuli orientation of V1 afferents inputs to V2. **B** The synaptic connection strength versus position in topographic space relative to the soma at the center for four V2 neurons selective to long, straight edges or lines; each row are the 7x7 convolution filters for one type of V2 unit. Four orientations are illustrated, ranging from horizontal to an upward angle of 67.5° in steps of 22.5°, and the model has a total of 8 orientations. The strongest excitation are for inputs aligned to preferred orientation of the unit in a line down the center of the receptive field. There are weaker excitatory inputs having orientations 22.5° from the preferred orientation. **C** Pictorial representation of the preferred selectivity for the unit having the convolution filters in the same row. **E** The synaptic connection strength versus position in topographic space relative to the soma at the center for four V2 neurons selective to corners having a 90° interior angle, and four different orientations 22.5° apart. The model contains similar units having interior angles ranging from 45° to 157.5° in steps of 22.5°. Each such unit has 16 different preferred orientations, for a total of 96 different types of units. The pattern is formed by having two excitatory strips running from the center to the edge of the receptive field, each coming from an afferent with an appropriate orientation for one arm of a corner. The positions of the excitatory strip rotate with the preferred orientation of the unit. **F** Pictorial representation of the preferred selectivity for the unit having the convolution filters in the same row.

the excitatory and inhibitory stellate cells in L4 [Kandel et al., 2000], leaving that as a subject for future study.

As in V1, V2 L4 plays the role of receiving excitatory input and replicating it so that it includes both excitatory and inhibitory afferents that obey Dale's Law. The L4 model for V2 also performs a pooling function: The units have 2×2 receptive fields and are spaced 2 positions apart. This is known as a stride of 2 in the terminology of DCNN, and results in the dimension of the output being reduced by one half along each axis: The input to model V2 L4 is 112×112, and the output is 56×56.

Each of the 8 channels of V1 L2/3 connection patterns has a corresponding unit type (channel) in V2 L4: each type accepts inputs from one channel of the V1 L2/3 outputs, a total of 4 inputs for each V2 L4 unit. The inputs are weighted so that any one of the 4 inputs from V1 to an area V2 L4 unit will trigger an output spike. The result is a spiking version of the maximum pooling operation in a DCNN. As in the area V1 L4 model, there are excitatory and inhibitory channels representing separate populations of excitatory and inhibitory neurons.

The model for V4 includes a model of L4 that is identical to the V2 L4 model, adapted to the V2 afferents: V4 L4 units take afferents from a 2×2 pixel receptive field, and are staggered 2 pixels apart and the dimension of the visual maps are reduced from 56×56 to 28×28. Each unit in the V4 L4 model accepts afferent inputs from a single type (channel) of the V2 connection pattern in terms of both sub-unit separation angle and the unit's overall orientation. The L4 units have sufficient excitation so that they are guaranteed to spike if any of their four inputs from V2 units spike.

B.5 Input Stimuli

Input images were created following the approach described in [Nandy et al., 2013] and adapted to the neural network modeling framework. Line segments 12 pixels long were drawn in black on a white 112×112 pixel background. Each model stimuli consisted of 3 line segments joined at the

end and forming 5 different degrees of curvature using 5 different angles: 0° for straight lines, 22.5° for low curvature, 45° for moderate curvature, 77.5° for high curvature, and 90° for "C" shaped. Each stimuli was used in 16 rotated versions separated by angles of 22.5°. The stimuli were placed in the image on a 3×3 grid, 38 pixels apart, beginning at pixel (18,18). This created a total of $5 \times 16 \times 9 = 720$ stimuli. Note that as in [Nandy et al., 2013], the straight line stimuli are redundant when rotated by 180° but these duplicates were left in the stimuli set for simplicity.

After plotting the lines and rendering the image as a bitmap, gaussian noise was added during each simulation: Each pixel was adjusted by zero mean gaussian noise with a standard deviation set to 40% of the maximum image intensity. The resulting image was then re-normalized. Examples of the resulting input images from a single simulation are shown in Figure **??**. The fairly high level of background noise was not used in the stimuli in [Nandy et al., 2013], but is included in the model to help match the broad background firing seen in the recordings. As noted in the Discussion, other sources of noise are ignored for computational expediency.

B.6 Hardware and Runtime

Assuming atemporal binary inputs and outputs as in Listing 4.2, the tensor graph for 6 layers (3 cortical areas, 2 layers per area) has around 150 operations. This is more than suggested by Listing 4.2, as the actual model code must handle convolution filters of varied sizes and the input layer. The spike time simulation using the PSP equations given by Listing 2 and Listing 3 was run for 10 ms discretized at 0.2 ms intervals; the resulting tensor simulation is an approximately 50,000-operation graph representing the layers at each time, plus additional operations for the PSP calculations and logic relating to the temporal simulation.

These large tensor graphs run on standard hardware: a simulation of a 525,000-neuron network for 10 ms takes 45 minutes to run on a MacBook Pro without a GPU and 3.5 minutes on an AWS p3.2xlarge GPU instance having a Tesla V100 GPU with 16 GB of RAM. As previously mentioned, the GPU memory is the key constraining factor when it comes to running longer or more detailed simulations. The simulations shown in the Results section run in just under the memory available.

B.7 PSP Maximum

The constant Z in equation 1 is determined by noting that the difference of exponentials in equation 1 takes its maximum at a time after the input arrival given by:

$$t_{max} = \tau_r \tau_f \frac{\ln(\tau_r/\tau_f)}{\tau_r - \tau_f} \tag{B1}$$

and the constant Z is given by

$$\frac{1}{Z} = [e^{(\frac{t_{max}}{\tau_{fall}})} - e^{(\frac{t_{max}}{\tau_{rise}})}] \tag{B2}$$

References

[Akiba et al., 2019] Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[Bartunov et al., 2018] Bartunov, S., Santoro, A., Richards, B. A., Marris, L., Hinton, G. E., and Lillicrap, T. (2018). Assessing the scalability of biologically-motivated deep learning algorithms and architectures. *arXiv preprint arXiv:1807.04587*.

[Behrens and Sporns, 2012] Behrens, T. E. and Sporns, O. (2012). Human connectomics. *Current opinion in neurobiology*, 22(1):144–153.

[Bengio et al., 2015] Bengio, Y., Lee, D.-H., Bornschein, J., Mesnard, T., and Lin, Z. (2015). Towards biologically plausible deep learning. *arXiv preprint arXiv:1502.04156*.

[Bengio et al., 2017] Bengio, Y., Mesnard, T., Fischer, A., Zhang, S., and Wu, Y. (2017). Stdp-compatible approximation of backpropagation in an energy-based model. *Neural computation*, 29(3):555–577.

[Borst and Helmstaedter, 2015] Borst, A. and Helmstaedter, M. (2015). Common circuit design in fly and mammalian motion vision. *Nature neuroscience*, 18(8):1067–1076.

[Buzsáki and Wang, 2012] Buzsáki, G. and Wang, X.-J. (2012). Mechanisms of gamma oscillations. *Annual review of neuroscience*, 35(1):203–225.

[Cadena et al., 2019] Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., and Ecker, A. S. (2019). Deep convolutional models improve predictions of macaque v1 responses to natural images. *PLoS computational biology*, 15(4):e1006897.

[Cadieu et al., 2007] Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., and Poggio, T. (2007). A model of v4 shape selectivity and invariance. *Journal of neurophysiology*, 98(3):1733–1750.

[Carandini et al., 2005] Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., Gallant, J. L., and Rust, N. C. (2005). Do we know what the early visual system does? *Journal of Neuroscience*, 25(46):10577–10597.

[Cornford et al., 2020] Cornford, J., Kalajdzievski, D., Leite, M., Lamarquette, A., Kullmann, D. M., and Richards, B. A. (2020). Learning to live with dale's principle: Anns with separate excitatory and inhibitory units. In *International Conference on Learning Representations*.

[Dale, 1935] Dale, H. (1935). Pharmacology and nerve-endings.

[Douglas and Martin, 2004] Douglas, R. J. and Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27:419–451.

[El-Shamayleh and Pasupathy, 2016] El-Shamayleh, Y. and Pasupathy, A. (2016). Contour curvature as an invariant code for objects in visual area v4. *Journal of Neuroscience*, 36(20):5532–5543.

[Feather et al., 2023] Feather, J., Leclerc, G., Mkadry, A., and McDermott, J. H. (2023). Model metamers reveal divergent invariances between biological and artificial neural networks. *Nature Neuroscience*, 26(11):2017–2034.

[Ferster and Miller, 2000] Ferster, D. and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual review of neuroscience*, 23(1):441–471.

[Fukushima, 1980] Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202.

[Gerstner and Kistler, 2002] Gerstner, W. and Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press.

[Girard et al., 2001] Girard, P., Hupé, J., and Bullier, J. (2001). Feedforward and feedback connections between areas v1 and v2 of the monkey have similar rapid conduction velocities. *Journal of neurophysiology*, 85(3):1328–1331.

[Hansen, 2006] Hansen, N. (2006). The cma evolution strategy: a comparing review. *Towards a new evolutionary computation: Advances in the estimation of distribution algorithms*, pages 75–102.

[Holmgren et al., 2003] Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *The Journal of physiology*, 551(1):139–153.

[Hu et al., 2020] Hu, J. M., Song, X. M., Wang, Q., and Roe, A. W. (2020). Curvature domains in v4 of macaque monkey. *Elife*, 9:e57261.

[Hubel and Wiesel, 1962] Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154.

[Jacob et al., 2021] Jacob, G., Pramod, R., Katti, H., and Arun, S. (2021). Qualitative similarities and differences in visual object representations between brains and deep networks. *Nature communications*, 12(1):1–14.

[Kandel et al., 2000] Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth, A. J., Mack, S., et al. (2000). *Principles of neural science*, volume 4. McGraw-hill New York.

[Kheradpisheh et al., 2018] Kheradpisheh, S. R., Ganjtabesh, M., Thorpe, S. J., and Masquelier, T. (2018). Stdp-based spiking deep convolutional neural networks for object recognition. *Neural Networks*, 99:56–67.

[Kietzmann et al., 2019] Kietzmann, T. C., McClure, P., and Kriegeskorte, N. (2019). Deep neural networks in computational neuroscience. In *Oxford research encyclopedia of neuroscience*.

[Kim et al., 2014] Kim, J. S., Greene, M. J., Zlateski, A., Lee, K., Richardson, M., Turaga, S. C., Purcaro, M., Balkam, M., Robinson, A., Behabadi, B. F., Campos, M., Denk, W., Seung,

S. H., and the EyeWireres (2014). Space-time wiring specificity supports direction selectivity in the retina. *Nature*, 509(7500):331.

[Kim et al., 2019] Kim, T., Bair, W., and Pasupathy, A. (2019). Neural coding for shape and texture in macaque area v4. *Journal of Neuroscience*, 39(24):4760–4774.

[Kisvárday et al., 1997] Kisvárday, Z. F., Toth, E., Rausch, M., and Eysel, U. T. (1997). Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cerebral Cortex (New York, NY: 1991)*, 7(7):605–618.

[Ko et al., 2013] Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S. B., and Mrsic-Flogel, T. D. (2013). The emergence of functional microcircuits in visual cortex. *Nature*, 496(7443):96–100.

[Kobatake and Tanaka, 1994] Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of neurophysiology*, 71(3):856–867.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

[Liang et al., 2017] Liang, H., Gong, X., Chen, M., Yan, Y., Li, W., and Gilbert, C. D. (2017). Interactions between feedback and lateral connections in the primary visual cortex. *Proceedings of the National Academy of Sciences*, 114(32):8637–8642.

[Liu et al., 2016] Liu, L., She, L., Chen, M., Liu, T., Lu, H. D., Dan, Y., and Poo, M.-m. (2016). Spatial structure of neuronal receptive field in awake monkey secondary visual cortex (v2). *Proceedings of the National Academy of Sciences*, 113(7):1913–1918.

[McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

[Miller, 1994] Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on-and off-center inputs. *Journal of Neuroscience*, 14(1):409–441.

[Mueller, 1999] Mueller, B. K. (1999). Growth cone guidance: first steps towards a deeper understanding. *Annual review of neuroscience*, 22(1):351–388.

[Nandy et al., 2013] Nandy, A. S., Sharpee, T. O., Reynolds, J. H., and Mitchell, J. F. (2013). The fine structure of shape tuning in area v4. *Neuron*, 78(6):1102–1115.

[Pasupathy and Connor, 2001] Pasupathy, A. and Connor, C. E. (2001). Shape representation in area v4: position-specific tuning for boundary conformation. *Journal of neurophysiology*, 86(5):2505–2519.

[Perrett and Oram, 1993] Perrett, D. I. and Oram, M. W. (1993). Neurophysiology of shape processing. *Image and Vision Computing*, 11(6):317–333.

[Plomp et al., 2017] Plomp, G., Michel, C. M., and Quairiaux, C. (2017). Systematic population spike delays across cortical layers within and between primary sensory areas. *Scientific reports*, 7(1):1–14.

[Pospisil et al., 2018] Pospisil, D. A., Pasupathy, A., and Bair, W. (2018). 'artiphysiology'reveals v4-like shape tuning in a deep network trained for image classification. *Elife*, 7:e38242.

[Riesenhuber and Poggio, 1999] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025.

[Roe et al., 2012] Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., and Vanduffel, W. (2012). Toward a unified theory of visual area v4. *Neuron*, 74(1):12–29.

[Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.

[Seeman et al., 2018] Seeman, S. C., Campagnola, L., Davoudian, P. A., Hoggarth, A., Hage, T. A., Bosma-Moody, A., Baker, C. A., Lee, J. H., Mihalas, S., Teeter, C., et al. (2018). Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *elife*, 7:e37349.

[Serre, 2019] Serre, T. (2019). Deep learning: the good, the bad, and the ugly. *Annual review of vision science*, 5(1):399–426.

[Shirasaki and Pfaff, 2002] Shirasaki, R. and Pfaff, S. L. (2002). Transcriptional codes and the control of neuronal identity. *Annual review of neuroscience*, 25(1):251–281.

[Singer, 1999] Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, 24(1):49–65.

[Smith et al., 1997] Smith, S. W. et al. (1997). The scientist and engineer's guide to digital signal processing.

[Srinath et al., 2021] Srinath, R., Emonds, A., Wang, Q., Lempel, A. A., Dunn-Weiss, E., Connor, C. E., and Nielsen, K. J. (2021). Early emergence of solid shape coding in natural and deep network vision. *Current Biology*, 31(1):51–65.

[Thorpe et al., 2001] Thorpe, S., Delorme, A., and Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural networks*, 14(6-7):715–725.

[Thorpe and Imbert, 1989] Thorpe, S. J. and Imbert, M. (1989). Biological constraints on connectionist modelling. *Connectionism in perspective*, pages 63–92.

[Tsao et al., 2006] Tsao, D. Y., Freiwald, W. A., Tootell, R. B., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311(5761):670–674.

[Vanni et al., 2020] Vanni, S., Hokkanen, H., Werner, F., and Angelucci, A. (2020). Anatomy and physiology of macaque visual cortical areas v1, v2, and v5/mt: bases for biologically realistic models. *Cerebral Cortex*, 30(6):3483–3517.

[VanRullen et al., 2005] VanRullen, R., Guyonneau, R., and Thorpe, S. J. (2005). Spike times make sense. *Trends in neurosciences*, 28(1):1–4.

[Wallis and Rolls, 1997] Wallis, G. and Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in neurobiology*, 51(2):167–194.

[Yamins et al., 2014] Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, 111(23):8619–8624.

[Zador, 2019] Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature communications*, 10(1):1–7.

[Zhang et al., 2019] Zhang, Y., Lee, T. S., Li, M., Liu, F., and Tang, S. (2019). Convolutional neural network models of v1 responses to complex patterns. *Journal of computational neuroscience*, 46(1):33–54.