ARTICLE Communicated by Anthony Neville Burkitt

Single Circuit in V1 Capable of Switching Contexts During Movement Using an Inhibitory Population as a Switch

Doris Voina
dvoina@uw.edu
Applied Mathematics, University of Washington, Seattle, WA 98195 U.S.A.

Stefano Recanatesi
stefano.recanatesi@gmail.com
Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, U.S.A.

Brian Hu
brian.hu@kitware.com
Allen Institute for Brain Science, Seattle, WA 98109 U.S.A.

Eric Shea-Brown
etsb@uw.edu
Applied Mathematics, University of Washington, Seattle, WA 98195, U.S.A.

Stefan Mihalas
stefanm@alleninstitute.org
Applied Mathematics, University of Washington, Seattle, WA 98195, U.S.A., and Allen Institute for Brain Science, Seattle, WA 98109, U.S.A.

As animals adapt to their environments, their brains are tasked with processing stimuli in different sensory contexts. Whether these computations are context dependent or independent, they are all implemented in the same neural tissue. A crucial question is what neural architectures can respond flexibly to a range of stimulus conditions and switch between them. This is a particular case of flexible architecture that permits multiple related computations within a single circuit.

Here, we address this question in the specific case of the visual system circuitry, focusing on context integration, defined as the integration of feedforward and surround information across visual space. We show that a biologically inspired microcircuit with multiple inhibitory cell types can switch between visual processing of the static context and the moving context. In our model, the VIP population acts as the switch and modulates the visual circuit through a disinhibitory motif. Moreover, the VIP population is efficient, requiring only a relatively small number of neurons to switch contexts. This circuit eliminates noise in videos by using

Neural Computation **34, 541–594** (2022)
https://doi.org/10.1162/neco_a_01472

appropriate lateral connections for contextual spatiotemporal surround modulation, having superior denoising performance compared to circuits where only one context is learned. Our findings shed light on a minimally complex architecture that is capable of switching between two naturalistic contexts using few switching units.

1 Introduction

Our brains are unique in their ability to adapt to the context in which stimuli appear. Animals face the problem of processing visual stimuli rapidly and efficiently while adapting to different contexts every time they transition to a new environment (e.g., from jungle to savanna, from the shores of a river to underwater). A classic example of adaptation to different contexts is discussed in Barlow's "efficient coding hypothesis" (Barlow, 1961), which proposes that sensory systems encode maximal information about environments with different statistics (Olshausen & Field, 1996a, 1996b). In this and other cases, when context changes, neural circuits switch from previous strategies of feature representation to new ones that are better adapted to the statistical properties of the new context. How the neuronal circuitry of the brain is organized to account for the multitude of contexts animals may encounter has not been established (Yang, Cole, & Rajan, 2019). In particular, when do we need separate circuits for different contexts, and when can single circuits be modulated to switch among multiple contexts (Gozzi et al., 2010; Koganezawa, Kimura, & Yamamoto, 2016; Zhou et al., 2017; Cardin, 2019; Mante, Sussillo, Shenoy, & Newsome, 2013; Cohen, Dunbar, & McClelland, 1990; Yang et al., 2019)? Our aim is to identify a biologically constrained network that is capable of switching contexts and to infer the building blocks required for such switching. In constructing such a network, we will only discuss and include the structural and functional detail needed for the switching of contexts.

We focus on a concrete setting in which rapid context switching is apparent. This is mouse V1, which responds differently to inputs when the animal is running (moving condition) compared to when it is stationary (static condition) (Niell & Stryker, 2010; Fu et al., 2014). When the animal transitions from standing still to running, visually evoked firing rates significantly increase. For example, in one experimental setting, the firing rate of neurons in layers II/III of area V1 more than doubled (Niell & Stryker, 2010), while in layer V of V1, noise correlations between pairs of neurons were substantially reduced (Dadarlat & Stryker, 2017).

While an enormous diversity of cell types has been characterized (Tasic et al., 2018), in this work we focus on the three primary classes of inhibitory interneurons—vasoactive intestinal peptide (VIP), somatostatin (SST), and parvalbumin (PV)—and one class of long-range projecting excitatory neurons: the pyramidal neurons (PYR) as shown in Figure 1a (Fu et al., 2014;

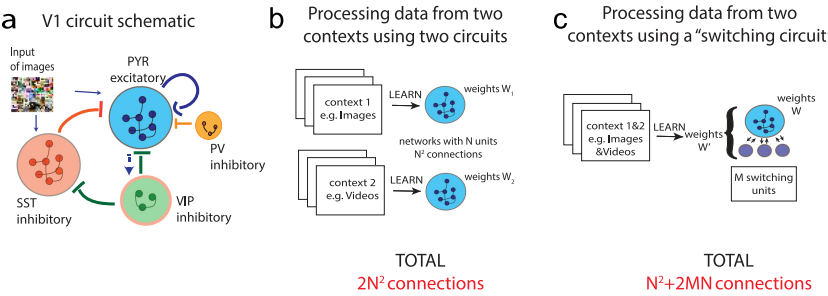

Figure 1: (a) Schematic of circuit involving VIP, SST, PV, and PYR groups of neurons. When VIP are silent, PYR are self-excitatory, while SST and PV inhibit PYR. When VIP are active, they inhibit the PYR while also creating a disinhibitory motif given by VIP-SST-PYR. The potential connection from PYR to VIP explored in this article is marked with a dotted arrow. (b) Processing of two input types (e.g., images, videos) happens using two separate networks for each type of input, each having N units with $2N^2$ weights in total to learn. (c) Processing of two input types can be done with one circuit: a switching circuit with N units adapted to one of the contexts and M switching units that turn on when the other context is presented. We may want $M << N$, with $N^2 + 2NM$ connections to learn (assuming switching units are not interconnected). When the number of switching units required in a switching circuit is small, fewer connections need to be learned; more specifically, if $M < N/2 \Rightarrow N^2 + 2MN < 2N^2$. This generalizes well to a range of circuits, including in the case of sparse connectivities, as often presented throughout the article.

Cardin, 2018; Rudy, 2011; Pfeffer, Xue, He, Huang, & Scanziani, 2013). VIP is an inhibitory population of neurons that is strongly modulated by running (Fu et al., 2014). In our simplified model of the circuit, VIP neurons act in a switch-like manner: they are silent when animals are static but start firing when animals are running, inhibiting SST cells and hence releasing PYR cells from SST inhibition. The disinhibition of PYR cells is not uniform, but rather a complex pattern that is dependent on the particular PYR cell response. We will show that the switch can be effective only if PYR cells provide input information to the VIP cells. Although this simple model does not capture all the physiological responses of VIP neurons, we believe the model captures the crux of the disinhibitory switching computation at the expense of biological realism.

We study this circuit using a model in which the contextual information is stored in the lateral connections between neurons (Iyer, Hu, & Mihalas, 2020). Each neuron receives information about the visual scene from feedforward connections (which can be arbitrary in this model) and complements this with surround information provided by nearby neurons. The connections are dependent on the statistics of the environment; more

precisely, they depend on the frequency of co-occurrence in the environment of the features which the neurons represent. These connections are most useful if the information from the feedforward connections is corrupted (e.g., by occlusions).

Importantly, the contextual information via lateral connections comes not only from the spatial surround but also from the past. Synaptic delays introduce a constraint on the available information each neuron gets. During the static condition, past surround information matches present information, and thus there is no temporal variability of the context. During movement, this no longer holds; neighboring features now also vary temporally, which changes the co-occurrence frequency; hence, the statistics of the moving context are different. We aim to find connection strengths from the switching VIP units that, during movement, modulate firing rates and neuronal correlation structure to adapt and enhance the encoding of visual stimuli when the moving context is turned on. Although throughout the article, we focus on the visual circuit and the switching role of the VIP neural population, these results can be generalized to circuits processing multiple contexts, and thus their applicability has broader scope. In section 3, we list several other biological examples of circuits processing multiple contexts.

Understanding switching circuits may also further aid efforts to design both flexible and efficient artificial neural architectures. This research area has benefited from bio-inspired architectures and algorithms like elastic weight consolidation (Kirkpatrick, Pascanu, & Hadsel, 2017), intelligent synapses (Zenke, Poole, & Ganguli, 2017), iterative pruning (Mallya & Lazebnik, 2018), leveraging prior knowledge through lateral connections (Rusu et al., 2016), task-based hard attention mechanism (Serra, Suris, Miron, & Karatzoglou, 2018), and block-modular architecture (Terekhov, Montone, & O'Regan, 2015), for example, to enable sequential learning by eliminating "catastrophic forgetting" (where previously acquired memories are overwritten once new tasks are learned). We hypothesize that a few switching units akin to VIP can be incorporated as part of the hidden layers to enable context modulation. This makes such a switching circuit architecture (see Figure 1c) more efficient than employing separate circuits for the different contexts (see Figure 1b) because switching circuits have fewer connections to learn.[1] We hope such a circuit architecture will inspire next-generation flexible artificial nets that can process stimuli in changing contexts.

[1] In general, if N is the number of neurons per location, L is the number of locations, and C is the number of connections per neuron, then the total number of connections in a circuit is NLC. Two identical circuits have $2NLC$ connectivities, while a switching circuit has $NLC + LM(c_{in} + c_{out})$, where M is the number of switching (VIP) units and c_{in}, c_{out} are the number of connections to and from the switching units, respectively. When $M \ll N$ and $c_{in}, c_{out} < C$, then $2NLC > NLC + LM(c_{in} + c_{out}) \Leftrightarrow NC > M(c_{in} + c_{out})$ which is true for circuits with small M, c_{in}, c_{out}.

1.1 Article Outline. In section 2.1, we first detail a model introduced in Iyer et al. (2020) that describes neuronal connections and firing rates of a circuit adapted to static visual scenes (images). We next extend this model to the case of circuits adapted to moving visual scenes (videos). These circuits are attuned to the statistical regularities of movement and take into account constraints of biological networks, like synaptic delay. We are able to map these two circuit models to the V1 circuit, consisting of PYR, SST, and PV neuron populations. We thus obtain two different networks with full cell-type specifications achieving optimal context integration for static and moving contexts, respectively. In section 2.2 we detail the data sets and procedures used to quantify connectivities and firing rates in these two circuits. In section 2.3, we go on to describe a circuit that can switch between neuronal activity in static circuit and neuronal activity in the moving circuit by virtue of adding a single population, the VIP. We find that VIP projections to SST and PYR are not enough to shift activity during movement, but that we need a feedback connection from the PYR to the VIP (section 2.4). The resulting circuit is the minimally complex circuit resembling V1 we have found to switch contexts. In section 2.5, we describe how this circuit switches using only a small number of VIP units. We follow up on these results in section 2.6, where we use this switching circuit to obtain better reconstructions of videos in conditions of high noise. Finally, we evaluate the new switching circuit architecture with data from V1 that confirms some of the model's predictions (see section 2.7).

2 Results

2.1 Theoretical Models of Processing Visual Information in Static and Moving Contexts. We first introduce two models of visual processing in the V1 in the static and moving contexts where the circuits implementing the computations perform optimal inference and are adapted to the statistical regularities of the contexts through the lateral connections between neurons.

2.1.1 Model of Visual Processing in the Static Context. To study optimal context integration in the static condition (where the visual input is static images), we take as a starting point a model proposed by Iyer et al. (2020) where model neurons respond to a patch in the visual space—the classical receptive field—but this response is modulated by a larger region of space—the extraclassical receptive field. The extraclassical receptive field contribution is determined by nearby local receptive fields providing indirect input from a larger area of visual space (see Figure 2a). Specifically, interneuron interactions providing extraclassical information from the surround via lateral connections (see section 4.1) complement intrinsic neuronal responses to classical receptive fields to determine firing rates.

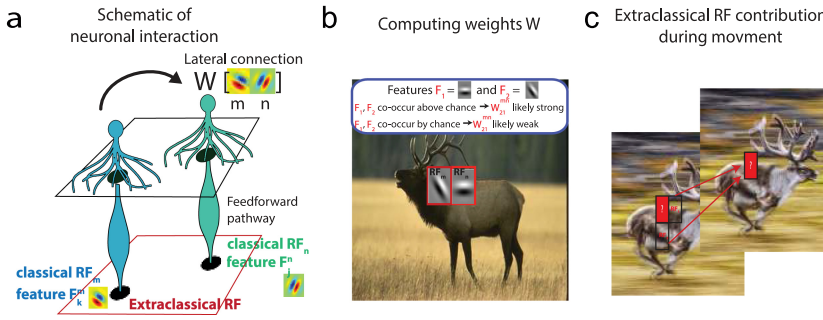

Figure 2: (a) Neurons receive stimulus input from a patch in space at position n, their classical receptive field (RF_n), but also from surrounding patches in space (e.g., the patch at position m) through interactions with other neurons. These neurons are connected by weights $\mathbf{W}_{jk}^{mn}$ that depend on the statistical regularities of natural scenes. (b) When features $\mathbf{F}_1$ and $\mathbf{F}_2$ at positions m, n occur together often in natural scenes, then $\mathbf{W}_{21}^{mn}$ is strong; when $\mathbf{F}_1$ and $\mathbf{F}_2$ occur together by chance, without significant correlation, $\mathbf{W}_{21}^{mn}$ is close to 0. (c) Spatiotemporal surround for motion processing. Due to synaptic delay, context integration uses surrounding patches that are also Δt ms in the past to assess the features in the present frame.

Starting from the assumption that firing rates of a population of neurons encode the probability of specific features being present in a given location of the image, we consider a probabilistic framework that includes probability of feature occurrence and feature co-occurrence, which we can then map to an equation involving firing rates of neurons and weights (see section 4.1). In general, a feature j, denoted by $\mathbf{F}_j$, describes a specific pattern that neurons are most attuned to, which can vary from simplistic, like Gabor filters, to complex, like faces or objects that are robust to stimulus transformations such as scale and position changes. In more detail, for neurons responding to $\mathbf{F}_j^n$ (feature j at patch n in visual space), we define $\mathbf{f}_j^n$ to be the steady-state firing rate due to the classical receptive field and $\mathbf{r}_j^n$ to be the (overall) steady-state firing rate taking into account the extraclassical receptive field contribution. The probabilistic assumption stated above is such that $\mathbf{f}_j^n$ relates to the probability $p(\mathbf{F}_j^n|i^n)$ by the following relation,

$$\mathbf{f}_j^n = g(p(\mathbf{F}_j^n|i^n)), \tag{2.1}$$

where g is a monotonically increasing function, i^n is a patch n in visual space, and $\sum_j p(\mathbf{F}_j^n|i^n) = 1$. For simplicity, we fix g to be the identity, leaving the relaxation of this linear assumption for future work. With $\mathbf{f}_j^n = p(\mathbf{F}_j^n|i^n)$, neurons tuned for distinct features respond differently to the same patch i^n in

visual space depending on how well its corresponding feature is represented. Operationally, to compute $\mathbf{f}_j^n$ in response to an image, we first chose a basis of features, for example, features obtained by approximating spatial receptive fields from recorded neurons in V1. We then preprocessed the image (see section 4.2), convolved the image with feature j and normalized the result such that the sum over all features is 1 at each spatial position, and finally considered the patch i^n of the normalized convolution.

Once $\mathbf{f}_j^n$ is computed, we can continue assuming that neuronal firing rates contain information about feature occurrence in the surround, so that $\mathbf{r}_j^n = p(\mathbf{F}_j^n) = p(\mathbf{F}_j^n | i^1, i^2, \ldots, i^n, \ldots)$, where $i^1, i^2, \ldots, i^m$ are surrounding patches of i^n. We can then use Bayes' rule to express this probability in terms of feature probability at patch i^n and at surrounding locations i^m (see section 4.1 for a detailed calculation) and finally map the resulting equations to neurobiological quantities (see section 4.1). In summary, these operations yield that the firing rates $\mathbf{r}_j^n$ of neurons are the result of modulating the classical receptive field firing rate $\mathbf{f}_j^n$ by extraclassical receptive field information from the surround, which is a linear function of other neurons' classical receptive field firing rates, $\mathbf{f}_k^m$. These firing rates are weighed by the lateral connections $\mathbf{W}^{\text{static}}$, representing the prior information about the statistical regularities of natural images. After ignoring terms that are due to higher-order modulation of the surround (see section 4.1), specifically neurons from the surround having surround modulation of their own, we obtain the following firing rates (see Figure 2a) as explained in detail in section 4.1:

$$\mathbf{r}_j^n \approx \mathbf{f}_j^n \circ \left(1 + \sum_{m,k} \mathbf{W}_{kj}^{mn} \mathbf{f}_k^m\right), \tag{2.2}$$

with the weights expressed as

$$\mathbf{W}_{kj}^{mn} = \frac{p(\mathbf{F}_k^m \cap \mathbf{F}_j^n)}{p(\mathbf{F}_k^m) p(\mathbf{F}_j^n)} - 1 = \frac{\langle \mathbf{f}_k^m, \mathbf{f}_j^n \rangle_{\text{all images}}}{\langle \mathbf{f}_k^m \rangle_{\text{all images}} \langle \mathbf{f}_j^n \rangle_{\text{all images}}} - 1, \tag{2.3}$$

where $\mathbf{F}_j^n$ is a Gabor-like feature n at location j that we will illustrate shortly, the symbol $\cap$ denotes the co-occurrence of two features, and $\circ$ is the Hadamard product, the element-wise multiplication between tensors $\mathbf{f}_j^n$ and $1 + \sum_{m,k} \mathbf{W}_{kj}^{mn} \mathbf{f}_k^m$. Further, $\mathbf{f}_j^n$ is the evoked firing rate due to the classical receptive field of neurons firing for feature $\mathbf{F}_j^n$, and $\mathbf{r}_j^n$ is the firing rate of neurons firing for feature $\mathbf{F}_j^n$ using information from classical and extraclassical receptive fields. The sum $\sum_{m,k} \mathbf{W}_{kj}^{mn} \mathbf{f}_k^m$ is over neurons with receptive fields at different locations m, responsive to features k. Finally, $\mathbf{W}_{kj}^{mn}$ is the connectivity in the static context between neurons responsive to

features $\mathbf{F}_k^m$ and $\mathbf{F}_j^n$. We define $\mathbf{W}^{\text{static}} \equiv \{\mathbf{W}_{kj}^{mn}\}_{m,n,k,j}$ as the connectivity applied to static visual scenes. Assuming that weights only connect neurons with nonoverlapping receptive fields, the resulting weights are sparse (see section 4.2).

From a computational perspective, the organism cannot measure the feature probabilities and joint probabilities in equations 2.1 and 2.3 directly, but these can be estimated given our defined neural code as the convolutions between image and feature, $p(\mathbf{F}_j^n|i^n) = \mathbf{f}_j^n = i^n * \mathbf{F}_j$, and as the cross-correlations between classical receptive field firing rates, $p(\mathbf{F}_k \cap \mathbf{F}_j) = \mathbf{f}_k * \mathbf{f}_j$. By mapping these probabilistic statements on feature occurrence to neurobiological quantities that capture firing rates and weights, we have obtained a circuit that does approximate context integration, extracting information through priors embedded in the neural connectivities. While the start of the model is Bayes optimal via equations 4.12 and 4.14, a set of approximations is needed to keep the circuit simple.

There are multiple possible mappings from the probabilistic framework to the neurobiological circuit (Iyer et al., 2020), but the current correspondence is straightforward and yields successful predictions from data, such as like-to-like connectivity, as detailed below. When a pair of features is frequently co-occurring, weights between neurons preferential for these features are strong and positive (see Figure 2b). In contrast, when two features are unlikely to co-occur in the same image, the connectivity is strong and negative. Overall occurrence probabilities of individual features normalize the co-occurrence probabilities so that the weights express the co-occurrence of features over and above chance. Co-occurrence probabilities of features are then averaged over many natural scenes so that the corresponding weights $\mathbf{W}^{\text{static}}$ capture the statistical regularities of natural environments.

2.1.2 Model of Visual Processing in the Moving Context. We next show how the framework above can be applied to the moving context. While equations 2.2 and 2.3 show how connectivity and firing rates can be optimized to account for spatially co-occurring features—features that appear at the same moment in time but in different locations of the visual field—we now extend these equations to account for temporal co-occurring features—features that occur at nearby moments in time at different locations of the visual field.

In more detail, context is generally integrated from Δt in the past due to synaptic delay (see Figure 2c), and weights are proportional to co-occurrence probabilities of neighboring features that are also separated by a time window Δt. This is a direct generalization of the model in Iyer et al. (2020) to the time domain and includes synaptic delay as a biologically motivated constraint. The extended model can capture how local circuit connectivity is shaped by spatiotemporal correlations across receptive

fields and across time windows characteristic of biological processes like synaptic delay. The firing rate during the moving context is (see sections 4.1 and 4.2):

$$\mathbf{r}_j^{n,t} \approx \mathbf{f}_j^{n,t} \circ \left(1 + \sum_{m,k} \mathbf{W}_{kj}^{mn,\Delta t} \mathbf{f}_k^{m,t-\Delta t}\right), \tag{2.4}$$

with the weights expressed as

$$\mathbf{W}_{kj}^{mn,\Delta t} = \frac{p(\mathbf{F}_k^{m,t} \cap \mathbf{F}_j^{n,t-\Delta t})}{p(\mathbf{F}_k^{m,t}) p(\mathbf{F}_j^{n,t-\Delta t})} - 1 = \frac{\langle \mathbf{f}_k^{m,t}, \mathbf{f}_j^{n,t-\Delta t} \rangle_{\text{all videos}}}{\langle \mathbf{f}_k^{m,t} \rangle_{\text{all videos}} \langle \mathbf{f}_j^{n,t-\Delta t} \rangle_{\text{all videos}}} - 1, \tag{2.5}$$

where we apply an analogous notation as for equations 2.2 and 2.3, the only difference being the additional $t, \Delta t, t - \Delta t$ superscripts that denote the time coordinate for the features, firing rates, and weights. $\mathbf{W}^{\text{moving}} \equiv \mathbf{W}_{kj}^{nm,\Delta t}$ is the connectivity in the moving context between neurons responsive to features $\mathbf{F}_k^{m,t}$ and $\mathbf{F}_j^{n,t-\Delta t}$ whose activation is separated by a time delay Δt. Note that the expression for $\mathbf{W}_{kj}^{nm,\Delta t}$ as shown in equation 2.5 also holds for the static context when we use static visual input to compute the weights, such that $\mathbf{f}^t = \mathbf{f}^{t-\Delta t}$ for all $t, \Delta t$.

We have introduced a model of visual processing where feedforward and lateral connections between neurons serve different roles. The lateral connections between neurons perform unsupervised learning of the probability of co-occurrence of visual features that the neurons represent. For the purpose of this study, the feedforward connections can be arbitrary, and the microcircuit described here can be at any level of processing. This separation of the roles for the feedforward and lateral connections allows for an easy implementation of both supervised and unsupervised learning in deep networks (Hu & Mihalas, 2018).

Here, we show how this model can integrate information from the surround using these within-layer connectivities in both static and moving states. However, integration of these two contexts results in two distinct circuits needed to perform visual processing under different conditions (static versus moving). The model optimally integrates context in the Bayes sense, meaning it uses priors on the co-occurrence of features in natural scenes when integrating information from the surround. These priors reflect the known statistical regularities of the environment (Simoncelli, 2003; Barlow, 1961; Marr, 1982) and weigh the surround contributions appropriately. We are then able to map this model formalism to the circuit architecture in V1 described above while specifying steady-state network weights and activations, as well as cell type functionality. This model emphasizes robust

coding and applies best in conditions of high noise, where parts of the visual scene are missing due to occlusions or are corrupted, and thus where context information may play a critical role. We next describe our model of visual processing in detail.

2.2 Modeling Firing Rates and Weights in Networks Responding to Images and Videos. We next describe two separate circuits capable of doing optimal context integration in each of the moving and static contexts. We characterize these two circuits through the connectivities $\mathbf{W}^{\text{static}}$ and $\mathbf{W}^{\text{moving}}$, computed by using images and videos in training data sets and applying formulas 2.3 and 2.5. Once the corresponding connectivities are specified, we can further characterize the static and moving circuits by their neural activations. In the following, we elaborate, section by section, on the algorithm we implemented to compute the static and the moving weights.

2.2.1 Data Set and Feature Preparation. We applied our framework for processing static images and videos to different benchmark data sets, chosen to address differences in the statistics of visual features across conditions: during viewing of static images (static condition) and during viewing of videos that contain motion (moving condition). For the static condition, we used 300 selected grayscale images of the BSDS data set (Martin, Fowlkes, Tal, & Malik, 2001) (see Figure 3a) while for videos, the BSDS data set is preprocessed through a smaller sliding window that travels along the image to reproduce motion (see Figure 3b and section 4.4). Although in general the sliding window can move in any direction (see Figures S1 to S2 for results in this case), here we constrained it to move solely in the horizontal direction to roughly approximate flow of images across the (sideways-facing) eyes of mice during forward movement. We have not used a generic data set of natural videos since most videos in such data sets contain limited movement of objects, humans, or animals rather than movement of sections of an environment that would mimic the visual experience of a running animal.

We generated a dictionary of features (filters) based on a parameterized set of models derived from recordings in V1 (Durand et al., 2016). This contains 18 filters with gaussian subfields (see Figure 3d) at different relative intensities and orientations. We added filters containing a temporal dimension—*spatiotemporal filters*—to obtain a set of 34 filters. Our spatiotemporal filters consist of two frames (see Figure 3e) and represent a temporal shift by several pixels in the horizontal direction, corresponding to the direction of movement and amount of displacement of the sliding window in the videos described above.

To more easily illustrate and interpret our model, we first tested our framework on a different, synthetic context. We analyzed a simplified 9×9 world of horizontal and vertical bars moving up and down as well as

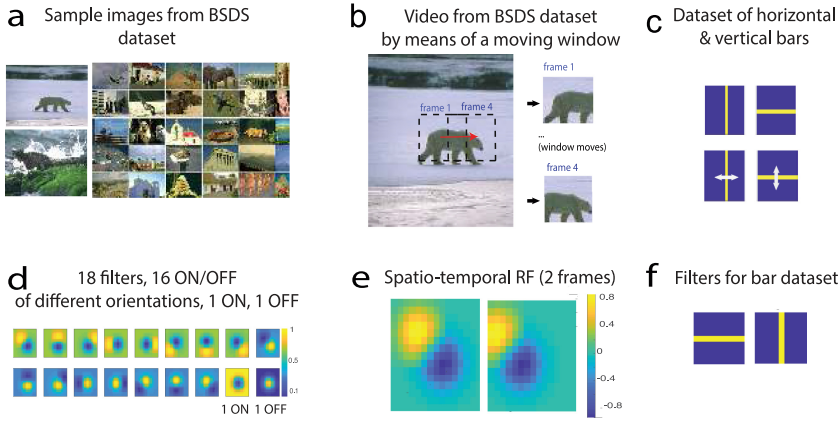

Figure 3: (a) Sample images from the BSDS data set. Images of animals, human faces, landscapes, buildings, and so on are used. (b) Sliding window on images from the BSDS data set so that the appearance of movement is achieved. Shown by the red arrow is how much the window has moved from frame 1 to frame 4. In general, movement of sliding window is random and in any direction, but we focus on horizontal movement in the case of natural videos. (c) Images of horizontal and vertical bars (above) and how the bars move in videos (below). (d) Eighteen filters: ON, OFF, ON/OFF with two gaussian subfields, different subfields dominating, at different intensities and orientations. Color bars show the different intensities of pixels. (e) Example of a spatiotemporal filter comprising two frames. Spatiotemporal filters are added to the 18 original filters to make up a total of 34 filters. The filter shown here over two frames captures a 45 deg bar moving to the left and is obtained by translating the original filter by three pixels. Color bars show the different intensities of pixels to the left. (f) Two filters for the simplistic "bar world" comprising a horizontal and a vertical bar, respectively.

left and right (see Figure 3c). This simple data set has only two features, horizontal bars and vertical bars (see Figure 3f), but movement can be in any of the four orthogonal directions.

2.2.2 Computing the Weights $\mathbf{W}^{static}$, $\mathbf{W}^{moving}$. The firing rates $\mathbf{f}$ due to the classical receptive field represent feature probabilities (see equation 2.1 with $g(x) = x$) and were computed by the following sequence of operations: preprocessing inputs and filters (see section 4.2), convolving the image or video frames with the respective sets of filters, rectifying, and then normalizing so that all firing rates $\mathbf{f}_k^m$ lie in the interval between 0 and 1 and sum up to 1 across all features k. To find the weights for static and moving contexts, $\mathbf{W}^{\text{static}}$ and $\mathbf{W}^{\text{moving}}$, we fixed Δt. After convolving $\mathbf{f}_k^t$ and $\mathbf{f}_j^{t-\Delta t}$ in accordance

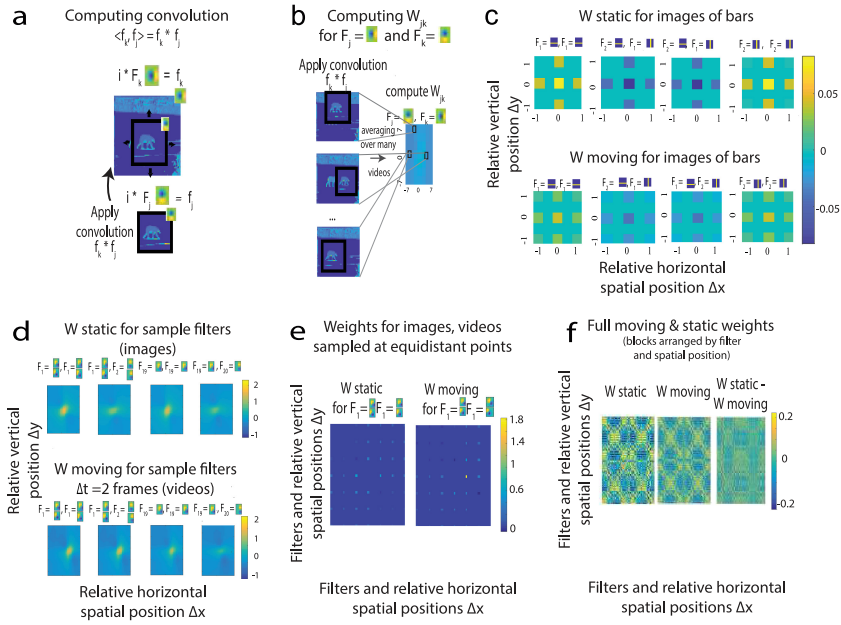

Figure 4: (a) To obtain the weight matrix, we first take the convolution of video frames with features from the feature basis (e.g., $i * F_k$, $i * F_j$). We then consider the convolution of these convolved image frames to detect feature co-occurrence (e.g., $f_k * f_j$). (b) Schematic of how weights are represented. Normalized convolutions between patches separated by the same spatial and temporal distances are averaged and stored in the corresponding entry of the weight matrix. (c) Top: Static weights for the data set of images of bars. Bottom: Moving weights for the data set of videos of bars. (d) Static weights (above) and moving weights (below) for the data set of natural images/videos during horizontal motion only. (e) Sparse versions of slices from the static and moving weights for the data sets of natural images/videos during horizontal motion. Weights between neurons whose receptive fields are not at certain preselected, sufficiently far apart locations in the visual space were discarded to satisfy the constraint that patches are independent. (f) The full (nonsparse) tensors $\mathbf{W}^{\text{static}}$, $\mathbf{W}^{\text{moving}}$, and $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$, ordered first by spatial position, then by filter.

with equations 2.3 and 2.5 and following the procedure outlined in Figures 4a and 4b, we obtained a high-dimensional tensor that characterizes the connections between each pair of cell types (k, j) at each position in the image. Using the feature $\mathbf{F}_j^k$ as a proxy for a cell "type," the resulting tensor is four-dimensional, with dimensions: cell type of the source, cell type of the target, and relative spatial position of the source and target in x and y directions.

2.2.3 Simplifications to Weights. We make three simplifications to reduce the number of parameters in this tensor (see section 4.2): (1) we assume translational invariance so that only the relative position of two filters is relevant ($\mathbf{W}^{n_1,n_2}_{j_1,j_2} = \mathbf{W}^{n_3,n_4}_{j_1,j_2}$ when $\vec{n_1} - \vec{n_2} = \vec{n_3} - \vec{n_4}$); (2) the model is designed to compute connections to neurons that receive independent observations; thus, we only consider connections between neurons whose receptive fields are sufficiently far apart (i.e., at least half a receptive field apart); (3) as statistical dependencies in natural images decay with distance, we limit the spatial extent of connectivity to three times the size of the classical receptive field. Figures 4c and 4d show several 2D slices through this tensor, corresponding to a specific cell source and target, as well as the full static and moving weights (see Figure 4f) ordered by spatial position and feature type (see also Figure S1a). Figure 4c serves to provide some intuition as to what these weights represent and how they are structured: in the data set of bars, horizontal feature $\mathbf{F}_1$ frequently occurs or is absent together with other horizontal features $\mathbf{F}_1$ at neighboring locations, which leads $\mathbf{W}^{\text{static}}_{11}$ to have positive values. Conversely, horizontal feature $\mathbf{F}_1$ occurs always when vertical feature $\mathbf{F}_2$ is absent, and vice versa, leading to negative weights $\mathbf{W}^{\text{static}}_{12}$, $\mathbf{W}^{\text{static}}_{21}$ (see Figure 4c).

2.2.4 Characterizing $\mathbf{W}^{moving}$ in the Case of Two Different Video Statistics. In the generation of the video data set we use a sliding window to enforce controlled and comparable statistics between the moving and static contexts. When the sliding window is free to move in all directions, the moving weights tend to be weaker in absolute value, which holds for the simple data set of bars (see Figure 4c), and the weights generated from the data set of natural images and videos (see Figures S1a to S1b). This effect is due to the weaker statistical dependence of features separated by the time window Δt. Feature co-occurrence, and thus connectivity, is affected by the distortions during movement, like change of orientation of objects or appearance or disappearance of objects in the visual scene. Moving weights in this case are approximately smoothed-out versions of the static weights (see Figures S1a to S1b). In these conditions, as the information from surround is less reliable, the feedforward input plays a more important role during movement.

In the case when the sliding window moves s pixels horizontally in Δt time steps, $\mathbf{F}^{n,t}_k$ and $\mathbf{F}^{n+(s,0),t-\Delta t}_k$ actually coincide so that their probability of co-occurrence is maximized. This means that for horizontal movement, $\mathbf{W}^{\text{moving}}_{kk}$ peaks s pixels from the center for any feature $\mathbf{F}_k$ and $\mathbf{W}^{n,n+(s,0),\Delta t}_{kk}$ is strong (see Figures 4d to 4e). Results for natural videos below are for horizontal movement, although the same general conclusions hold when movement is allowed in any direction (see Figure S2).

Finally, using $\mathbf{W}^{\text{static}}$, $\mathbf{W}^{\text{moving}}$ and applying equations 2.2 and 2.4 we obtain the corresponding firing rates $\mathbf{r}$ in both static and moving contexts.

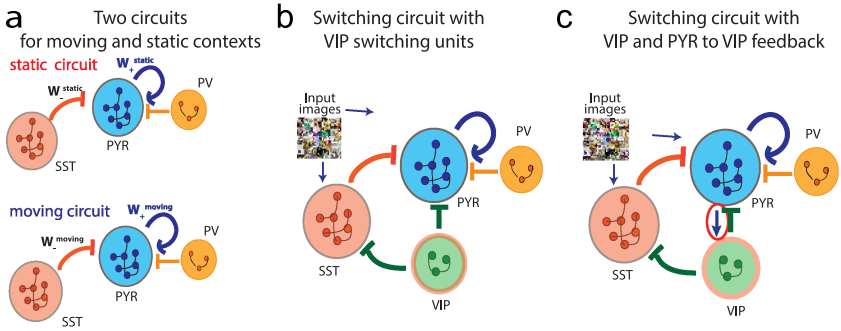

Figure 5: (a) Two separate circuits for optimal visual processing of static (top) and moving contexts (bottom), respectively. (b) The proposed switching circuit with the VIP population approximates the static circuit when the VIP are silent and the animal is static, and approximates the moving circuit when the VIP are active and the animal is moving. (c) Previous circuit, but with a feedback connection added from the PYR population to the VIP.

2.3 Implementing a Switching Circuit. Having two just defined optimal connectivities, $\mathbf{W}^{\text{static}}$ and $\mathbf{W}^{\text{moving}}$, for the static and moving contexts, we next consider whether a single circuit involving the cell types described above (VIP, PYR, SST, and PV) can respond optimally in these two contexts and switch between them. We additionally seek the computational principles behind the minimally complex circuit (i.e., the circuit with fewest connections) for such a switching circuit. Specifically, we ask whether a circuit with optimal weights for the static context can switch to produce nearly optimal activities in the moving context, via projections from a set of switching units. In such a circuit, every PYR neuron approximates Bayesian inference, combining classical receptive field information with information from the surround to estimate feature probability.

We start by rewriting the model described by equations 2.2 and 2.4 in vector form to obtain the following firing rates:

$$\mathbf{r}^{t,\text{static}} = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^t), \tag{2.6}$$

$$\mathbf{r}^{t,\text{moving}} = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{moving}}\mathbf{f}^t). \tag{2.7}$$

Assuming, as discussed above, that the activation of the VIP neural population implements the switch between contexts, we want the switching circuit to reproduce the firing rates given by equation 2.6 when the VIP neurons are silent in the static context, and the firing rates given by equation 2.7 when the VIP neurons are active in the moving context (see Figures 5a and 5b). We next explain how $\mathbf{r}^{\text{static}}$, $\mathbf{r}^{\text{moving}}$ above can be modeled as the firing rates of the PYR neurons.

When the VIP are silent, the only groups of neurons active are PV, SST, and PYR. This circuit is equivalent to one without any VIP connections, reproducing firing rates of PYR given by equation 2.6 when the animal is static. PYR neurons contribute to integrating surround information through excitatory projections and receive inhibitory feedback from SST interneurons (Braitenberg & Schuz, 1991). PV implements a normalization of the PYR population in our model, consistent with data on their connectivity (Jiang et al., 2015; Pfeffer et al., 2013). Empirically it has been shown these neurons receive the average inputs of the PYR neurons whose receptive fields overlap with their classical receptive fields and project back equally (Pfeffer et al., 2013). In our model, this normalization applies to the classical receptive field $\mathbf{f}$, as described in section 4.1. As for the role of PYR and SST, given that PYR are excitatory and SST are inhibitory and that $\mathbf{W}^{\text{static}} = \mathbf{W}_{+}^{\text{static}} + \mathbf{W}_{-}^{\text{static}}$, it is natural to map the positive component of the static weights, $\mathbf{W}_{+}^{\text{static}}$, to the connections within the PYR population, and the negative component of the static weights, $\mathbf{W}_{-}^{\text{static}}$, to the inhibitory connections from SST to PYR. Hence, we obtain the following:

$$\mathbf{r}^{t,\text{static}} = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^t) = \mathbf{f}^t \circ (1 + \mathbf{W}_{+}^{\text{static}}\mathbf{f}^t + \mathbf{W}_{-}^{\text{static}}\mathbf{f}^t) \tag{2.8}$$

can be mapped to

$$\mathbf{r}^{t,\text{static}} = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^t) = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{PYR}\to\text{PYR}}\mathbf{f}^t + \mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{f}^t), \tag{2.9}$$

where $\mathbf{W}^{X\to Y}$ denotes the weights that connect neuronal populations $\mathbf{X}$ (the source) and $\mathbf{Y}$ (the target).

On the other hand when VIP are active, PYR firing rates ought to reproduce the activity given by equation 2.7. We make the simplifying assumptions that the switch from static to moving can happen instantaneously and that the VIP switch is binary. When the animal initiates movement and the VIP turns on, the model circuit should approximate the optimal response of PYR neurons resulting from the $\mathbf{W}^{\text{moving}}$ connectivities, within a circuit where the four neuronal populations interact (see Figure 5b). For VIP modulation of PYR (which is either direct or through the SST) that gives rise to the optimal firing rates in the moving context, we have that

$$\mathbf{r}^{t,\text{moving}} = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{moving}}\mathbf{f}^{t-\Delta t}) \tag{2.10}$$

is mapped to

$$\mathbf{r}^{t,\text{moving}} = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t} + \text{VIP contribution}). \tag{2.11}$$

Thus, the switch in the circuit occurs as VIP neurons modulate SST and PYR neurons and make PYR switch firing rates from $\mathbf{r}^{\text{static}}$ to $\mathbf{r}^{\text{moving}}$. We

now proceed to find the unknown connectivities from VIP to PYR and from VIP to SST that cause this to occur within the circuit (see Figures 5b and 5c).

2.4 In the Absence of Feedback to VIP Neurons, the Circuit Is Unable to Switch from Static to Moving Conditions. We attempt to describe the computational principles of the minimal switching circuit inspired by the V1 circuitry whose main structure and logic were described in Fu et al., (2014). After adding the switching population VIP, the goal is to find connectivities from VIP to the other two neuronal populations (PYR, SST) that would account for the PYR firing rates that yield optimal representation in the moving context. With the VIP contribution, the firing rate of PYR neurons can be expressed as (see section 4.5)

$$\begin{aligned}\mathbf{r}^{t,\text{moving}} = \mathbf{f}^{t} \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t} + \mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{W}^{\text{VIP}\to\text{SST}}\mathbf{f}^{t-\Delta t,VIP} \\ + \mathbf{W}^{\text{VIP}\to\text{PYR}}\mathbf{f}^{t-\Delta t,VIP}),\end{aligned} \tag{2.12}$$

where $\mathbf{f}^{t}$, $\mathbf{f}^{t-\Delta t}$ are firing rates due to the classical receptive field at times t and $t-\Delta t$ and inferred from the data set of natural videos as outlined in sections 2.1 and 4.2, $\mathbf{f}^{\text{t},VIP}$ are the intrinsic firing rates of the VIP at time t, and $\mathbf{r}^{t,\text{moving}}$ is the firing rate during the moving context with the extra-classical receptive field contribution. Here, $\mathbf{W}^{\text{SST}\to\text{PYR}}$ are weights from SST to PYR, $\mathbf{W}^{\text{VIP}\to\text{SST}}$ are weights from VIP to SST, and $\mathbf{W}^{\text{VIP}\to\text{PYR}}$ are weights from VIP to PYR. VIP neurons project to PYR neurons directly via weights $\mathbf{W}^{VIP\to PYR}$ and indirectly via the SST population. The effects of the indirect pathway VIP-SST-PYR can be captured by taking the product of connectivities, yielding $\mathbf{W}^{SST\to PYR}\mathbf{W}^{VIP\to SST}$. The three unknown variables are then $\mathbf{f}^{\text{t,VIP}}$, $\mathbf{W}^{\text{VIP}\to\text{SST}}$, and $\mathbf{W}^{\text{VIP}\to\text{PYR}}$, but since we assume $\mathbf{f}^{t,VIP}$ is constant in time t, this tensor can be combined with the connectivities to form the effective parameters

$$\mathbf{f}^{\alpha} = \mathbf{W}^{VIP\to SST}\mathbf{f}^{t-\Delta t,VIP} \tag{2.13}$$

and

$$\mathbf{f}^{\beta} = \mathbf{W}^{VIP\to PYR}\mathbf{f}^{t-\Delta t,VIP} \tag{2.14}$$

and hence reduce the number of unknowns and simplify notation. Our objective is to have firing rates in the switching circuit be as closely matched as possible to the firing rates in the separate moving circuit with $\mathbf{W}^{\text{moving}}$:

$$\begin{aligned}\mathbf{r}^{\text{moving},t} &= \mathbf{f}^{t} \circ (1 + \mathbf{W}^{\text{moving}}\mathbf{f}^{t-\Delta t}) \\ &\approx \mathbf{f}^{t} \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t} + \mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{f}^{\alpha} + \mathbf{f}^{\beta}).\end{aligned} \tag{2.15}$$

This amounts to minimizing the loss function defined by the approximation error $E_{\text{switch},1}$ over the variables $\mathbf{f}^{\alpha}$, $\mathbf{f}^{\beta}$:

$$\min_{\mathbf{f}^{\alpha},\mathbf{f}^{\beta}} E_{\text{switch},1} = \min_{\mathbf{f}^{\alpha},\mathbf{f}^{\beta}} \frac{1}{N} \sum_{f} \|(\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}})\mathbf{f} - \mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{f}^{\alpha} - \mathbf{f}^{\beta}\|_F, \tag{2.16}$$

where $\|\cdot\|_F$ is the Frobenius norm of a tensor, for all $\mathbf{f}$ (firing rates due to classical receptive fields) corresponding to video frames, and N is a normalization factor, the number of video frames in our data set. $\mathbf{f}$ is inferred through our model from the data sets of video frames and features using $\mathbf{f}_j^n = p(\mathbf{F}_j^n | i^n) = i^n * \mathbf{F}_j$ and thus is a known quantity throughout the optimization. Importantly, since $\mathbf{f}^{t,VIP}$ are firing rates and hence $\mathbf{f}^{t,VIP} \geq 0$, while $\mathbf{W}^{\text{SST}\to\text{PYR}} \leq 0$, $\mathbf{W}^{\text{VIP}\to\text{SST}} \leq 0$, and $\mathbf{W}^{\text{VIP}\to\text{PYR}} \leq 0$, we have that $\mathbf{f}^{\alpha}, \mathbf{f}^{\beta} \leq 0$, and $\mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{f}^{\alpha} \geq 0$.

This is a high-dimensional constrained optimization problem with the loss function defined as in equation 2.16, which we solved by means of a gradient descent method using the gradient-based Adam optimizer, implemented in PyTorch.[2] The weights $\mathbf{f}^{\alpha}$ and $\mathbf{f}^{\beta}$ as defined in equations 2.13 and 2.14 are unknown and learned by stochastic gradient descent (SGD), while $\mathbf{W}^{\text{moving}}$, $\mathbf{W}^{\text{static}}$, $\mathbf{W}^{SST\to PYR} \equiv [\mathbf{W}^{\text{static}}]_{-}$ are fixed. Finding the global minimum of the loss function is difficult, but the main goal is to find weights that give a small enough error $E_{switch,1}$ instead and later test these on a specific task to demonstrate that the optimal moving circuit can be approximated successfully (see section 2.6). We assessed the stability of our optimization by modifying several learning parameters—for example, learning rate (ranging from 0.001 to 0.1) and optimization algorithm (SGD, AdaGrad, RMSProp, Adam)—and checking the generalization error on a small number of frames (50) that were not used during training.

Regardless of hyperparameters, our optimization procedure did not find weights that together approximate the moving circuit significantly better than the static circuit. In other words, adding VIP neurons in an attempt to switch contexts does not lead to a significantly better approximation of the moving circuit than having no VIPs. This result holds for both the simple data set of horizontal and vertical bars and for the more complex data set of natural images and videos (see Figures 6b and 6c).

In order to understand the origin of this failure, we mathematically analyzed the circuit at hand. Analytically, if the loss is small $\text{E}_{\text{switch},1} \approx 0$, then $(\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}})\mathbf{f} \approx \mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{f}^{\alpha} + \mathbf{f}^{\beta}$, where $\mathbf{f}$ is unique to each image in the data. The left side becomes a term that varies across a wide range

[2] The tensor weights are very high-dimensional so that the least-squares method and variations thereof have failed due to the high memory requirements.

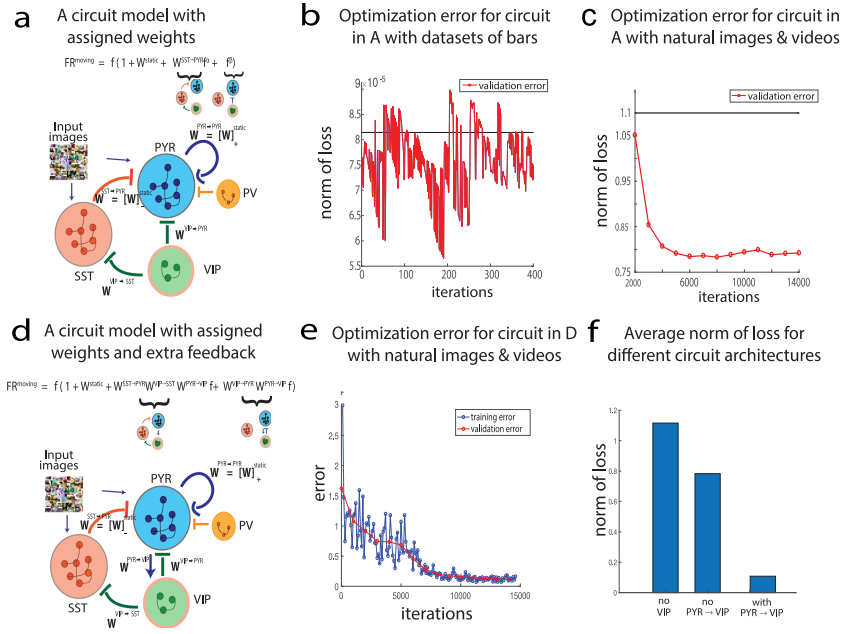

Figure 6: (a) Goal: Instead of two separate circuits for visual processing of static and moving contexts, the proposed circuit approximates the static circuit when the VIP are silent and the animal is static and the moving circuit when the VIP are active and the animal is moving. (b) Generalization/validation error found during the optimization to minimize the functional $E_{switch,1}$ for the data sets of static and moving bars does not converge. (c) Generalization/validation error found during the optimization to minimize the functional $E_{switch,1}$ for the data sets of natural images and videos converges, but the norm of the loss function decreases by only $\approx 25\%$. (d) Circuit as in panel a, but with a feedback connection added from the PYR population to the VIP. (e) Training error (blue) and generalization/validation error (red) found during the optimization to minimize the functional $E_{switch,2}$ (movement approximation error) for the data sets of natural images and videos converges to yield a relatively small error. (f) The movement approximation error for various circuit architectures: the static circuit with no VIP switching units, the circuit depicted in panel a without PYR to VIP feedback, and the circuit depicted in panel d.

of video frames, while the right side is a constant term incorporating the weights we are solving for: $\mathbf{f}^{\alpha}$, $\mathbf{f}^{\beta}$. This suggests that the failure of our optimization procedure to yield weights that approximate the moving circuit results from the VIP having no stimulus dependence.

We conclude that the circuit switching between static and moving contexts must be more complex than the simple circuit here, which has only

outgoing projections from VIP. Below, we introduce recurrent connections that make the VIP input dependent and overcome the limitations above.

2.5 VIP Circuit with Feedback from the PYR Cells Can Switch Context Integration from Static to Moving Conditions. Above we showed that a minimal switching circuit with only outgoing projections from the VIP units is insufficient to switch between the two contexts. Hence, we added a connection between PYR and VIP, such that the VIP group of neurons has access to information about the visual input through PYR (see Figure 5c). In this case we can approximate the firing rate of PYR during movement as follows, using the same conventions and assumptions as before (see section 4.5):

$$\begin{aligned}\mathbf{r}^{\text{moving},t} &= \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t} + \mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{W}^{\text{VIP}\to\text{SST}}\mathbf{W}^{\text{PYR}\to\text{VIP}}\mathbf{f}^{t-\Delta t} + \\ &\quad + \mathbf{W}^{\text{VIP}\to\text{PYR}}\mathbf{W}^{\text{PYR}\to\text{VIP}}\mathbf{f}^{t-\Delta t}). \qquad (2.17)\end{aligned}$$

We remind the reader that $\mathbf{f}$ is the contribution to the firing rate of the classical receptive field, and $\mathbf{W}^{X\to Y}$ are the weights from population X of neurons to population Y of neurons, where X, Y are the PYR, SST, VIP neurons. In addition to the fixed $\mathbf{W}^{\text{static}}$ and $\mathbf{W}^{\text{moving}}$, we also fix $\mathbf{W}^{\text{SST}\to\text{PYR}} = [\mathbf{W}^{\text{static}}]_-$. A schematic of the underlying circuit model, along with the corresponding formula for the firing rate of PYR, is shown in Figure 6d.

We would like to find the three unknown weights $\mathbf{W}^{VIP\to PYR}$, $\mathbf{W}^{VIP\to SST}$, and $\mathbf{W}^{PYR\to VIP}$ to best achieve the approximation:

$$\begin{aligned}\mathbf{r}^{\text{moving},t} &= \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{moving}}\mathbf{f}^{t-\Delta t}) \\ &\approx \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t} + \mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{W}^{\text{VIP}\to\text{SST}}\mathbf{W}^{\text{PYR}\to\text{VIP}}\mathbf{f}^{t-\Delta t} + \\ &\quad + \mathbf{W}^{\text{VIP}\to\text{PYR}}\mathbf{W}^{\text{PYR}\to\text{VIP}}\mathbf{f}^{t-\Delta t}). \qquad (2.18)\end{aligned}$$

We denote the approximated expression of equation 2.18 by $\mathbf{r}^{\text{approx}}$. This approximation $\mathbf{r}^{\text{approx}} \approx \mathbf{r}^{\text{moving}}$ amounts to minimizing the loss function defining the movement approximation error $E_{\text{switch},2}$:

$$\begin{aligned}E_{\text{switch},2} = \frac{1}{N}\sum_f \|(\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}})\mathbf{f} - \mathbf{W}^{\text{SST}\to\text{PYR}}\mathbf{W}^{\text{VIP}\to\text{SST}}\mathbf{W}^{\text{PYR}\to\text{VIP}}\mathbf{f} - \\ - \mathbf{W}^{\text{VIP}\to\text{PYR}}\mathbf{W}^{\text{PYR}\to\text{VIP}}\mathbf{f}\|_F, \qquad (2.19)\end{aligned}$$

for all N frames whose corresponding classical receptive field firing rate is $\mathbf{f}$. In the case of simple images and videos of bars, we consider $\mathbf{W}\cdot\mathbf{f}$ to be the regular matrix vector multiplication, while in the case of natural scenes, we perform the convolution operation $\mathbf{W} * \mathbf{f}$. Applying convolution for natural images and videos fits with the assumption we have applied for the PYR,

SST populations that weights between neurons are translationally invariant, and further reduces the number of parameters.

To solve this high-dimensional optimization problem, we set up, as in section 2.4, an optimization problem with the loss function being the average Frobenius norm as defined in equation 2.19. Weights to and from VIP are unknown ($\mathbf{W}^{\mathrm{VIP}\to\mathrm{SST}}$, $\mathbf{W}^{\mathrm{VIP}\to\mathrm{PYR}}$, and $\mathbf{W}^{\mathrm{PYR}\to\mathrm{VIP}}$) and learned by SGD, while $\mathbf{W}^{\mathrm{moving}} - \mathbf{W}^{\mathrm{static}}$, $\mathbf{W}^{\mathrm{SST}\to\mathrm{PYR}}$ are fixed. Importantly, Dale's law is enforced ($\mathbf{W}^{\mathrm{VIP}\to\mathrm{SST}}$, $\mathbf{W}^{\mathrm{VIP}\to\mathrm{PYR}} \leq 0$, $\mathbf{W}^{\mathrm{PYR}\to\mathrm{VIP}} \geq 0$) for biological realism.

To find how many switching units are needed, we varied the number of VIP neurons, which was equivalent to varying the dimensionality of tensors $\mathbf{W}^{\mathrm{VIP}\to\mathrm{SST}}$, $\mathbf{W}^{\mathrm{VIP}\to\mathrm{PYR}}$, and $\mathbf{W}^{\mathrm{PYR}\to\mathrm{VIP}}$. We found the smallest number of switching neurons VIP that enabled the loss (see equation 2.19) to be minimized. We first considered the simple image/video data set, which was 9×9 with horizontal and vertical bars. In this case, the loss was minimized with at least 20 VIP neurons (see Figure 7a). For comparison, there are 162 PYR and SST neurons, one for each filter and pixel in the image or frame. As increasing the number of VIP units further does not decrease the loss function, we conclude that for the case of bar-like images, having 20 switching units is enough.

Second, in the distinct case of more complex stimuli like images and videos of natural scenes, the movement approximation error in equation 2.19 was minimized when the number of VIP units is 34 per unit space, which matches the number of units in the PYR and SST population. However, the approximation error was already significantly minimized with only 5 VIP units per unit space, without any significant improvement after adding more units (see Figure 7b). Varying the dimensionality of spatial components of the tensors (see Figure S4) we were solving for ($\mathbf{W}^{VIP\to SST}$, $\mathbf{W}^{VIP\to PYR}$, $\mathbf{W}^{PYR\to VIP}$) and the synaptic delay Δt for sparse weights $\mathbf{W}$ that account for patch independence, we obtained the same qualitative results. Our results also hold for nonsparse weights, as shown in Figure S5a. Fixing the number of VIP units to 5 per unit space, we find that the approximated firing rate of equation 2.18 matches $\mathbf{r}^{\mathrm{moving}}$ compared to the $\mathbf{r}^{\mathrm{static}}$ firing rates of a circuit without VIP units (see Figure 7c). We conclude that for the specific parameters chosen in Figure 7b, the ratio of PYR to switching VIP units is $34/5 = 6.9$, so that the switching operation requires relatively few units, a fact we return to in the context of the underlying biology below.

All in all, we have shown that a switching circuit with relatively few numbers of switching VIP units and appropriate feedback connections can be implemented to achieve visual processing during the static and moving contexts, and for both a simple synthetic data set of bars and a biologically relevant data set of natural images and videos.

2.6 Context-Dependent Visual Processing with Extraclassical Receptive Fields Leads to Denoising. According to our theory (see section 4.1), the moving circuit achieves optimality of visual processing for videos, the

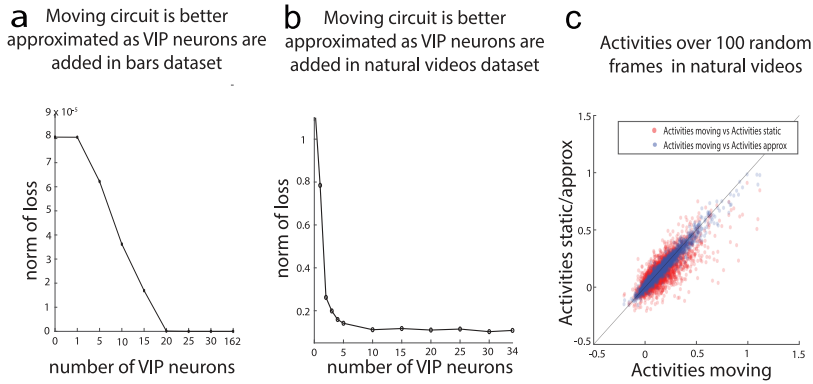

Figure 7: (a) Adding VIP switching units to the circuit processing videos of bars approximates the activity to that of the optimal circuit for moving context for this simple data set. However, no more than 20 VIPs are needed in practice, compared to the 162 PYR and SST cells. (b) Adding VIP switching units to the circuit processing natural videos approximates the activity to that of the optimal circuit for moving context for the naturalistic data set. However, no more than 5 VIPs per unit space are needed in practice, compared to the 34 PYR and SST cells per unit space. The parameters chosen for this optimization are $\Delta t = 2$ and $\dim(\mathbf{W}^{VIP\to SST}) = \dim(\mathbf{W}^{VIP\to PYR}) = 34 \times Nf_2 \times 3 \times 3$, $\dim(\mathbf{W}^{PYR\to VIP}) = Nf_2 \times 34 \times 3 \times 3$, where Nf_2 is the variable number of VIP units. (c) A random subset of activities corresponding to different video frames, filters, spatial positions for the static, moving, and approximated moving circuit. Red dots for activities for moving circuit ($\mathbf{r}^{\text{moving}}$) versus activities for static circuit ($\mathbf{r}^{\text{static}}$); blue dots for activities for moving circuit versus activities for approximated switching circuit ($\mathbf{r}^{\text{approx}}$). Activities are computed using weights with 5 VIP units/unit space. Activities chosen for the approximated switching circuit are able to better estimate the activities in the moving circuit in comparison to the ability of the activities in the static circuit to estimate the activities in the moving circuit.

static circuit achieves optimality of processing for static images, and we have found appropriate connectivities to and from a population of switching units—VIP—that can approximate either circuit in a model of V1, the *switching circuit*. We have, however, not yet assessed the performance of these circuits on specific visual processing tasks. We pursue this here for the task of denoising. Specifically, we ask how well (1) extra-classical receptive field contributions from the static or moving circuits (see Figure 5a) can improve reconstructions of noisy images and videos and (2) whether the switching circuit can achieve the same level of performance as the separately optimized moving circuit when processing videos. We focus on reconstructions of video frames and the superior performance of the moving

and switching circuits for processing moving contexts, although we also mention the comparably high performance of the static circuit and implicitly that of the switching circuit responding to static scenes, for processing static contexts.

To reconstruct a visual scene during movement, our brain uses information from the present but also time-delayed surround information, both of which can be inaccurate or incomplete. We use $\mathbf{W}^{\text{moving}}$ to weigh the past surround information, as these weights encapsulate the cross-correlational structure between features of the past and the present, thereby informing which features are more or less likely. We note that during motion, using $\mathbf{W}^{\text{static}}$ to weigh surround information may still be better than using no surround information at all: if movement in the videos is slow enough or Δt is small, features are smooth and $\mathbf{W}^{\text{static}}$ and $\mathbf{W}^{\text{moving}}$ are highly correlated.

To apply our models to the task of denoising, we apply gaussian white noise or salt and pepper noise ξ to the original frames X of the videos (see Figure 8a) and compute firing rates in the circuits in response to the noisy frames $X + \xi$. The firing rates are expressed as

$$\mathbf{r}^{\text{no EXC}}(t) = \mathbf{f}^t, \tag{2.20}$$

$$\mathbf{r}^{\text{static}}(t) = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t}), \tag{2.21}$$

$$\mathbf{r}^{\text{moving}}(t) = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{moving}}\mathbf{f}^{t-\Delta t}), \tag{2.22}$$

$$\begin{aligned}\mathbf{r}^{\text{approx}}(t) = \mathbf{f}^t \circ (1 &+ \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t} + \mathbf{W}^{SST\to PYR}\mathbf{W}^{VIP\to SST}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-\Delta t} + \\ &+ \mathbf{W}^{VIP\to PYR}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-\Delta t}).\end{aligned} \tag{2.23}$$

We denote "EXC" throughout the figures and text to represent the extraclassical receptive field contribution. Hence, $\mathbf{r}^{\text{no EXC}}$ is the firing rate due to only the feedforward pathway, with no lateral connections, and thus without any extraclassical, surround modulation. In the case of $\mathbf{r}^{\text{static}}$ ($\mathbf{r}^{\text{moving}}$), $\mathbf{W}^{\text{static}}$ ($\mathbf{W}^{\text{moving}}$) weights are the lateral connections applied that weigh the extraclassical receptive field information from the past surround. While $\mathbf{W}^{\text{static}}$ are nonoptimal weights to compute the firing rate, $\mathbf{W}^{\text{moving}}$ are optimal for inferring features in noisy conditions as described below (see section 4.1). Finally, $\mathbf{r}^{\text{approx}}$ results from lateral connections from our switching circuit with connections to and from VIP.

For each image frame X, we computed the corresponding firing rate $\mathbf{r}$ via equations 2.20 to 2.23 to obtain a tensor with entries for every filter and spatial position of X. We then deconvolved $\mathbf{r}$ for each filter $\mathbf{F}_j$ (see section 4.6) along its corresponding dimension to obtain the "reconstructed" frame X':

$$X + \xi \to \mathbf{r} \to X'. \tag{2.24}$$

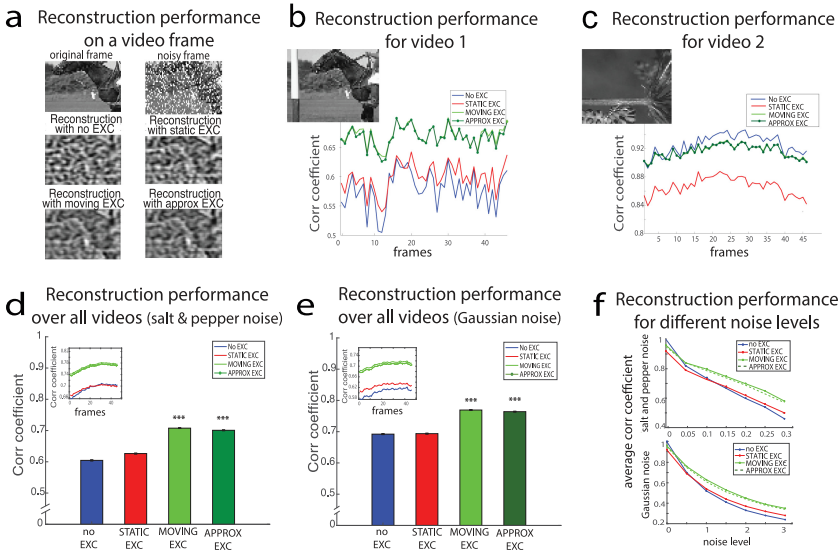

Figure 8: (a) Example of a reconstructed frame for each condition/circuit architecture: no EXC, static EXC, moving EXC, approximated EXC. (b) Average correlation coefficients between reconstructed noisy frames and reconstructed noiseless frames for one video in our data set. Here, reconstruction benefits from surround contextual information. (c) Same as panel a but in this case, the general inequality that holds on average $\rho(\mathbf{r}^{\text{no EXC}}), \rho(\mathbf{r}^{\text{static}}) < \rho(\mathbf{r}^{\text{moving}}) \approx \rho(\mathbf{r}^{\text{approx}})$ breaks down and $r(\mathbf{r}^{\text{no EXC}}) \approx r(\mathbf{r}^{\text{moving}})$. (d) Average correlation coefficient over all frames and all videos after salt and pepper noise was added to the video frames. The probability is 0.2 each pixel is changed to white and 0.2 each pixel is changed to black, and $\Delta t = 2$ (frames). The moving and approximated EXC average correlation coefficients are higher than for static EXC or no EXC (p-value < 0.05 using the Wilcoxon rank-sum test for all relevant comparisons). Inset: Correlation coefficients in time, averaged across videos. (e) Same as panel d for gaussian white noise with 0.5 standard deviation. $\Delta t = 2$ (frames). $p < 0.05$ for all relevant comparisons, Wilcoxon rank-sum test. (f) Average correlation coefficient over frames and videos as noise level is varied. Top: Salt and pepper noise is varied; Down: Gaussian white noise SD is varied.

Although there are ways for a biological circuit to do more accurate reconstructions (e.g., via learning weights), we have chosen a simple reconstruction approach that does not require additional assumptions here (e.g., the circuit does not know the structure of the noise or the input), as described in section 4.6.

We compare the quality of reconstructions from the four circuit models above. The baseline for these comparisons is the reconstruction of a noiseless image frame ($\xi = 0$), where the extraclassical contribution does not

provide any additional information. (Note that this reconstruction X' is not the same as the original frame X, as all feature information not included in the filters is lost in the initial convolution of the image frame to get $\mathbf{r}$). We denote by $\rho(\cdot)$ a metric of the quality of the reconstruction. This takes the firing rate $\mathbf{r}$ as input and generates the Pearson correlation coefficient between the reconstruction X' and the baseline reconstruction described above as output. The metric ρ for a video frame with noise ξ is

$$\rho(\mathbf{r}) = Corr(X'_{\xi}, X'_{\xi=0}) = \frac{(X'_{\xi} - \bar{X}'_{\xi}) \cdot (X'_{\xi=0} - \bar{X}'_{\xi=0})}{\|X_{\xi} - \bar{X}_{\xi}\|_2 \|X_{\xi=0} - \bar{X}_{\xi=0}\|_2}, \tag{2.25}$$

where $\cdot$ is the dot product and $\bar{X}$, $\bar{X}'$ are the means of the image and reconstruction, respectively. The upper limit for correlation coefficient is 1 when there is no noise in the image or frame (see Figure 8f).

Thus equipped, we ask which circuit architecture gives rise to neural activity best suited for decoding visual scenes in noisy conditions. Figure 8a shows reconstructions of a video frame using different such circuit architectures. We expect $\rho(\mathbf{r}^{\text{no EXC}}), \rho(\mathbf{r}^{\text{static}}) < \rho(\mathbf{r}^{\text{moving}}), \rho(\mathbf{r}^{\text{approx}})$ on average, as $\mathbf{W}^{\text{moving}}$ are the optimal lateral connections as defined above. However, the exact relationship between $\rho(\mathbf{r}^{\text{no EXC}}), \rho(\mathbf{r}^{\text{static}}), \rho(\mathbf{r}^{\text{moving}}), \rho(\mathbf{r}^{\text{approx}})$ depends on the exact correlational structure of the frames for each video. Some videos match our prediction that $\rho(\mathbf{r}^{\text{moving}})$ is maximized (see Figure 8b), while other videos do not (see Figure 8c). Specifically, there are videos where surround modulation is not effective, which appears to be due to the presence of independent features where the information in the extraclassical receptive field does not aid image reconstruction.

On average throughout the videos, $\mathbf{r}^{\text{moving}}$ and $\mathbf{r}^{\text{approx}}$ yield the best reconstructions (dark and light green bars), displaying the highest cross-correlation coefficients ρ between the noiseless reconstruction (the baseline) and the reconstructed frames (see Figure 8d). Figures 8d and 8e show this holds true when adding to the original frames either salt and pepper noise, when we varied the proportion of pixels occluded, or gaussian white noise, when we varied the standard deviation of the normal distribution of noise. The relation $\rho(\mathbf{r}^{\text{no EXC}}), \rho(\mathbf{r}^{\text{static}}) < \rho(\mathbf{r}^{\text{moving}}) \approx \rho(\mathbf{r}^{\text{approx}})$ is robust to the amount of noise added to the frames (see Figure 8f), whether for salt and pepper noise or gaussian noise. This holds true both when the complete set of 34 spatiotemporal filters is used (see Figure S10a) and when only the set of 18 filters with no temporal component is used (see Figure S10b). As expected, the addition of filters with a temporal component improves the reconstruction performance in all four circuit architectures presented (see Figure S10c). Furthermore, reconstruction performance for images in the static condition is maximized on average using $\mathbf{W}^{\text{static}}$ to weigh the surround so that $\rho(\mathbf{r}^{\text{no EXC}}), \rho(\mathbf{r}^{\text{moving}}), \rho(\mathbf{r}^{\text{approx}}) < \rho(\mathbf{r}^{\text{static}})$ on average (see Figure S9). This shows that the moving circuit is best used for processing noisy

video frames and that the static circuit (or switching circuit with VIP silent) is ideally used for processing images at the highest performance.

Thus, the switching circuit provides reconstruction performance comparable to a dedicated moving circuit for videos and comparable to a dedicated static circuit for images. In the case of videos, this is because the switching circuit reproduces firing rates that are close enough to $\mathbf{r}^{\text{moving}}$ to improve reconstruction fidelity. The correlation coefficients found between noiseless baseline reconstructions and reconstructions due to the moving and switching circuits, respectively, present almost perfect overlap (light and dark green curves in Figures S10a and S10b). In sum, we conclude that the extraclassical receptive field contribution in the moving circuit and approximated switching circuit generates neural activity that can be decoded to produce more accurate frame reconstructions in videos. To produce the most accurate image reconstructions, the VIP neurons in the switching circuit must be silent so that the network implements the static circuit.

2.7 Experimental Evidence of VIP Role in Movement-Related Visual Coding. When we examine the weights W to and from the VIP we have inferred in our model, we find that there are a few equally correct solutions for the optimization problem, equation 2.19, due to the multiple local minima of the movement approximation error. One of the possible solutions we found matched experimental data showing that in various layers of V1, the VIP-to-SST connection is strong compared to other connections, specifically the VIP-to-PYR connection (see Figure S6). Interestingly, this property arose only when including weights from SST to VIP in the circuit, consistent with experiments (Pfeffer et al., 2013, found the connection probability/strength from SST to VIP to be strong). Including this connection in our circuit and rewriting the circuit equations as in equation 4.24, we obtain a new set of connectivity patterns and activities so that we can now compare predictions of our model switching circuit to the extensive empirical evidence from the literature.

Importantly, we have not meticulously explored the set of all possible solutions from the optimization problem, equation 2.19, and, further, the optimization may allow additional constraints to the switching circuit while still admitting solutions. Acknowledging this, we now study both the connectivity and activity of the switching circuit with an additional SST to VIP connection.

2.7.1 Connectivity. We find that our model produces connectivity patterns that are largely consistent with empirical findings, as we describe next. Connection weights in the model can be interpreted as corresponding to a combination of connection probabilities and connection strengths in the data, as these have been shown to correlate well (Cossell et al., 2015). Regarding the connection from the VIP to SST, experimental data on connectivity in the visual cortex from Pfeffer et al. (2013) has shown that in

layer 4 of V1, the average connection probability from VIP to SST is double the connection probability from VIP to PYR (0.625 compared to 0.351), while in layer 5, VIP to SST is five times more probable (0.625 compared to 0.125) (Pfeffer et al., 2013). A recent study by Campagnola et al. (2021) has confirmed the relative paucity of VIP-to-PYR connections as compared to VIP-to-SST connections throughout all layers, for example finding 3 out of 52 VIP-to-PYR versus 5 out of 33 VIP to SST interarea L2/3 connections (Campagnola et al., 2021). VIP-to-SST connections are also stronger than VIP-to-PYR throughout all the layers: 0.32 compared to 0.28 as found by Jiang et al. (2015) and 0.3 compared to 0.21 as found by Campagnola et al. (2021).

We next examine the distribution of connectivity patterns in our computational model, as displayed in Figures 9a and 9b, and compare these model findings with experimental results. As found empirically, VIP-to-PYR connections in our model are sparser than VIP-to-SST connections, with a large peak at 0 in the connectivity histogram, in addition to being on average weaker (0.38 versus 0.47 for average weights in our model). Despite their sparsity, our model predicts a long tail to the distribution of VIP-to-PYR connection strengths. In addition, our model also predicts very high variability of $\mathbf{W}^{VIP\to PYR}$ connection strengths averaged with respect to the filter (which represents the postsynaptic cell type) as shown in Figure 9c. The strong connections correspond to the vertically oriented filters, as detailed below. We conclude that our model agrees with previous measurements and makes further predictions on the V1 microcircuit connectivity when including weights from SST to VIP.

We next inquire whether the synapses encode the contextual statistics by probing like-to-like connectivity both between PYR neurons and between VIP and PYR populations of neurons (see section 4.9). We find that while there is like-to-like connectivity between PYR neurons as found by Iyer et al. (2020), this effect is largely absent between the VIP and PYR. To further examine the pattern of connectivity from the VIP, we correlate both $\mathbf{W}^{VIP\to PYR}$ and $\mathbf{W}^{VIP\to SST}$ to $\mathbf{W}^{\text{moving/static}}$ and $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$, because these latter weights reflect the statistical regularities of the static and moving contexts. We obtain that after averaging over presynaptic filters ($N f_2$) and the spatial receptive fields, $\mathbf{W}^{VIP\to PYR}$ correlates positively with $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$ (0.41, p-value < 0.02, two-sided t-test); while $\mathbf{W}^{VIP\to SST}$ also correlates positively, the correlation coefficient is weaker and not statistically significant. Similarly, the convolution $\mathbf{W}^{PYR\to VIP\to PYR} \equiv \mathbf{W}^{VIP\to PYR} * \mathbf{W}^{PYR\to VIP}$ also correlates positively with $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$ (0.15, p-value < 0.01, two-sided t-test).

Analyzing the average postsynaptic weights $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$, $\mathbf{W}^{VIP\to PYR}$, $\mathbf{W}^{PYR\to VIP\to PYR}$ more specifically, we find that the strongest connections are for inhibited postsynaptic units corresponding to vertical or diagonal filters. Looking at the strongest inhibitory weights for $\mathbf{W}^{VIP\to PYR}$, for example (see Figure 9c), we find that 6/10 correspond to

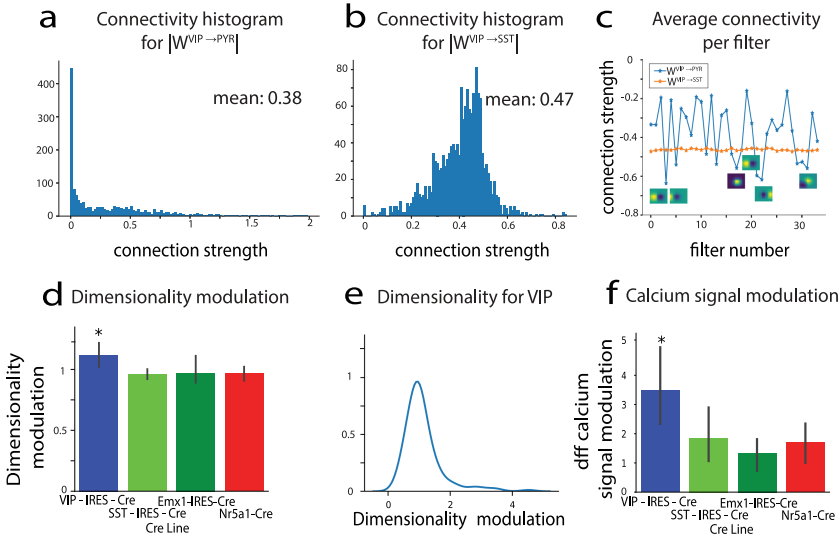

Figure 9: (a–c) Analysis of model connectivities $\mathbf{W}^{VIP\to SST}$, $\mathbf{W}^{VIP\to PYR}$. (a) Histogram of the absolute value of connectivities for $\mathbf{W}^{VIP\to PYR}$ showing a mean of 0.31. (b) Histogram of the absolute value of connectivities for $\mathbf{W}^{VIP\to SST}$ showing a mean of 0.4. (c) Average connectivity per filter, corresponding to the postsynaptic cell type, for $\mathbf{W}^{VIP\to PYR}$ (blue) and $\mathbf{W}^{VIP\to SST}$ (orange). Filters for postsynaptic units corresponding to the strongest conectivities are displayed to show what units are strongly inhibited during movement. (d–f) Data analysis of VIP population activity in calcium imaging data. (d) Dimensionality ratio (participation ratio measure) during periods of spontaneous activity between movement and static conditions across CRE lines. (e) Histogram of the modulation of dimensionality (statistics relative to the blue bar in panel d). (f) Activity (dff signal) ratio during periods of natural images viewing between movement and static conditions across CRE lines.

postsynaptic vertical filters and 9/10 to either vertical or diagonal postsynaptic filters. For $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$ and $\mathbf{W}^{PYR\to VIP\to PYR}$, 7 out of 10 such filters are vertical or diagonal (see Figure S11). We note also that the average connection strength of $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$ for postsynaptic vertical filters is negative and stronger ($-5.4 \cdot 10^{-5}$) compared to that for horizontal filters ($2 \cdot 10^{-6}$). This can be interpreted as follows: our videos feature horizontal movement, hence the spatiotemporal co-occurrence for vertical features in particular will be distorted during the moving context; this results in weaker $\mathbf{W}^{\text{moving}}$ weights overall when the postsynaptic cell responds to vertical filters and thus $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$ weights are strongly negative on average for such filters. The overall positive correlation of $\mathbf{W}^{\text{moving}} - \mathbf{W}^{\text{static}}$ with $\mathbf{W}^{VIP\to PYR}$, $\mathbf{W}^{PYR\to VIP\to PYR}$ determines that postsynaptic units tuned

for vertical features will be more strongly inhibited through these connections when switching from static to moving contexts. This phenomenon is more prevalent for $\mathbf{W}^{VIP\to PYR}$ where more of the strongest connections (9/10) are driven at least in part by inhibition of vertically tuned units, and in contrast with $\mathbf{W}^{\text{static}}$ and even $\mathbf{W}^{\text{moving}}$, where the strongest inhibitory connections are mostly for horizontally tuned units (5/5 and 5/5, respectively, of top inhibitory filters are either horizontal or diagonal) while $\mathbf{W}^{\text{static}}$, $\mathbf{W}^{\text{moving}}$ connections for vertically tuned units are mostly excitatory (on average, $1.4 \cdot 10^{-4}$ and $8.7 \cdot 10^{-5}$, respectively).

2.7.2 Activity. We next study the consistency of activity patterns produced by our model with respect to empirical data. Published experimental findings provide strong evidence that the VIP inhibitory population acts to modulate the visual circuitry in a movement-dependent manner (Niell & Stryker, 2010; Pfeffer et al., 2013; Fu et al., 2014). Very recent results show that VIP neurons respond synergistically to stimuli moving front to back during locomotion, a conjunction expected during locomotion in a natural environment for mice, with a preference for low but nonzero contrasts (Millman et al., 2020). Such movement-modulated activity matches the one required in our models, although we have not endowed the VIP units with specific feature selectivity. Additionally, we perform a set of new analyses of experimental data in the context of our model. These draw both on the literature and on the Allen Brain Observatory (http://observatory.brain-map.org/visualcoding, 2016), which contains in vivo physiological activity in the mouse visual cortex, featuring representations of visually evoked calcium responses from GCaMP6-expressing neurons in selected cortical layers, visual areas, and Cre lines. The data set contains calcium activations across multiple experimental conditions, and here we focus on periods of spontaneous activity, natural images, and drifting gratings.

Our model of the switching circuit shows that the relative number of VIP neurons required to switch between moving and static contexts is low when compared to the number of PYR or SST neurons (see Figures 7a and 7b). This number qualitatively matches the relative abundance of neurons in the three populations. Excitatory neurons PYR are more abundant than inhibitory ones (roughly 80% to 20%), and VIP are a minority of inhibitory cells. Moreover, the existing VIP cells recorded in the Allen Observatory do not appear to exploit substantially more degrees of freedom (as measured by their relative dimensionality) than other cell populations (see Figure S10a), consistent with a small number of effective VIP "units."

We now highlight two aspects of VIP neural activity that are directly related to our model and justify the choice of VIP as switching units whose activities are modulated by the locomotion state of the animal. First, VIP activity dimensionality is significantly modulated across the moving and static conditions during periods of spontaneous activity, as shown in

Figures 9d and 9e. To extract such dimensionality modulation, we considered periods of spontaneous activity in the recordings and divided the statistical distribution of the animal's speed, for each experimental session, into four quartiles. We then computed the average dimensionality, or participation ratio (PR; see section 4.8) for each recording in each quartile, which we define here as the (lower) dimension of a subspace where the data of activations can be represented while retaining some meaningful properties of the original data. We define the dimensionality modulation to be the ratio between the average dimensionality distribution within the highest quartile (movement condition) and the average within the first quartile (static condition). Such ratio is displayed in Figure 9e. The dimensionality of the VIP population is significantly modulated by movement, while in other populations, the same quantity was not significantly different across moving and static conditions (see Figure 9d). The histogram of such statistics is shown in Figure 9e.

Second, we analyzed evoked activity during the animals' viewing of natural scenes. We performed a calcium signal modulation analysis and found that for this stimulus set, the activity was strongly modulated for the VIP population and less so for other neural populations (see Figure 9f) across moving and static conditions assessed via the quartile method just described. This further confirms the stronger VIP modulation across the moving-static conditions. Further pieces of experimental evidence are presented in Figure S12.

Finally, we analyzed the activities of VIP and PYR neuron populations. Similar to Niell and Stryker (2010), we find the activity of the PYR during the moving condition to be higher than the stationary condition on average (0.066 versus 0.074, p-value < 0.01). However, our PYR population activity does not double during locomotion compared to periods of stationarity, as in Niell and Stryker (2010). More recent studies, however, have reproduced the relation between excitatory neuronal activity in mouse visual cortex and running but have observed a much weaker relation (Millman et al., 2020, Figure 5e).

We conducted further analysis to infer the tuning properties of the PYR and the VIP. This was achieved by considering a wavelet family (e.g., Daubechies), taking the two-dimensional discrete wavelet transforms of the video frames in our data, regarding the corresponding average wavelet transforms as features, and finally performing a linear regression or GLM against VIP or PYR activities with the average wavelet transforms as the independent variables (see section 4.9). We find that most PYR neurons are tuned to horizontal features, and much less to vertical features. Because VIP neurons in our model only get input from the PYR, while the top-down input activating VIP is described simply by the binary term s_t, VIP acquires the same preferential selectivity to horizontal features over and above that to vertical features (see Figure S13 and section 4.9 for details). This is counter to what we would expect if the VIP were capable of detecting the horizontal

movement in our data set by exhibiting preferential selectivity toward vertical features within their receptive fields instead of through the ad hoc bulit in switch term s_t. We conclude the simplification used by employing a binary term s_t in equation 4.27 prevents us from observing a more realistic VIP activation pattern that would deviate from the PYR pattern and provide further insight. This points to an important direction for future and to more detailed modeling expanding on our current simplified model.

Altogether these comparisons provide further support for our modeling assumptions and for the role of VIP neurons in visual coding across static and moving conditions. We conclude that our switching circuit model reproduces the global pattern of interactions via VIP that we expect, approximating the static and moving circuits, synchronal with VIP activation. Further analysis of future data sets, as examined in section 3, will guide next steps of circuit modeling.

3 Discussion

We have introduced a computational model for V1 circuitry that uses multiple cell types to integrate contextual information into local visual processing, during two different—static and moving—contexts. We have identified a need for recurrence, leading to the architecture of a switching circuit with bidirectional, learned connections to a switching population (here, the VIP cell class). Beyond V1 and biological circuit modeling, this circuit may be useful in searching for artificial neural network (ANN) architectures that can operate in different contexts and switch effectively between them.

Our model connects to a body of recent empirical studies elucidating V1 neural cell types and network logic. First, Niell and Stryker (2010) have established that as the speed of mice increases, the circuit increases spiking overall and changes the frequency content of local field potentials. Potentially, distinct activity patterns during locomotion could be attributed to effects from eye movements; however, Niell and Stryker provide evidence against this hypothesis. These findings prompt us to model the network as a switching circuit that adapts its activity as the state of the animal changes from static to moving. Later studies have focused on the connection strengths for excitatory and inhibitory neurons: neurons display like-to-like connectivity (Cossell et al., 2015; Ko et al., 2011), whereby neurons with similar orientation tuning have a higher probability of connecting and display stronger connections on average. Pfeffer et al. (2013) describe the V1 circuit logic by using transgenic mouse lines expressing fluorescent proteins or Cre-recombinase, providing a consistent classification of cell populations across experiments. Three large nonoverlapping classes of molecularly distinct interneurons that interact via a simple connectivity scheme were identified: PV, SST, and VIP inhibitory neurons. In particular, PV inhibit one another, SST avoid one another and inhibit all other types of interneurons, and VIP preferentially inhibit SST cells.

Another important development made by Fu et al. (2014) has established that locomotion activates VIP neurons independent of visual stimulation and predominantly through nicotinic inputs from basal forebrain. This study was the first to propose the existence of a cortical circuit for the enhancement of visual response by locomotion, describing a modulation of sensory processing by behavioral state. These studies motivate us to choose VIP as switching units and to map the positive and negative weights of our model to connectivities between different neuronal populations. Finally, another study suggests that differentiated network response during locomotion can be advantageous for visual processing (Dadarlat & Stryker, 2017): an increase in firing rates can enhance the mutual information between visual stimuli and single neuron responses over a fixed window of time, while noise correlations decrease across the population, which further improves stimulus discrimination. The authors hypothesize that cortical state modulation due to locomotion likely increases visually pertinent information encoded in the V1 population during times when visual information changes rapidly, such as during movement.

At least one study (Dipoppa et al., 2018) has disputed the findings of Neill and Stryker and of Fu et al., finding contrary evidence to the disinhibitory model. Experiments with the light on and visual stimuli present showed that locomotion increased both SST responses to large stimuli and VIP responses to small stimuli. However, the authors note that rerunning the measurements in darkness reproduced results from Fu et al., reinforcing the assumption that our model operates in conditions of poor visibility and high noise.

There is a vast literature on models of efficient coding starting with Barlow (1961) and Attneave (1954). (For a great description of this literature, see Chalk, Marre, & Tkačik, 2018.) On one extreme, if the signal-to-noise ratio is high and additional constraints (e.g., sparsity) are introduced, such models emphasize redundancy reduction (Olshausen & Field, 1996a; Rao & Ballard, 1999; Harpur & Prager, 1996; Comon, 1994; Bell & Sejnowski, 1995; Zemel, 1993; Dayan, Hinton, Neal, & Zemel, 1995). At the other extreme, if the signal-to-noise ratio is low, such models emphasize robust coding (Karklin & Simoncelli, 2011; Doi & Lewicki, 2014). We use a theoretical framework that emphasizes robust coding and that we have selected because of its generality. It starts with an assumption on neuronal activation functionality (i.e., firing rates of neurons encode the probability of specific features being present in a given location of the image). This model describes local circuit interactions needed for integration of information from surrounding visual stimuli in noisy conditions for an arbitrary representation. The model matches multiple empirical findings—for example that statistical regularities of natural images give rise to like-to-like local circuit connectivities, as observed experimentally (Cossell et al., 2015; Ko et al., 2011). However, in different contexts the model predicts different functional

lateral interactions. Therefore, we looked at circuits that can implement multiple functional interactions in one circuit.

Our model also relates to other switching circuits reported in the experimental literature. For example, selective inhibition of a subset of neurons in central nucleus of the amygdala (CeA) led to decreased conditioned freezing behavior and increased cortical arousal as visualized by fMRI (Gozzi et al., 2010). This therefore identifies a circuit that can shift fear reactions from passive to active. Another study has unraveled the cellular identity of the neural switch that governs the alternative activation of aggression and courtship in Drosophila fruit flies (Koganezawa et al., 2016). While these studies detail circuits responsible for switching behaviors, there are circuits switching between contexts: from detection of weak visual stimuli to discrimination after adaptation in mice (Ollerenshaw, Zheng, Millard, Wang, & Stanley, 2014); from high-response firing during active whisker movement, to low response when no tactile processing is initiated (Zhou et al., 2017); from odor attraction in food-deprived larva switching to odor aversion in well-fed larva (Vogt et al., 2020).

In contrast to this rich body of experimental studies, there are relatively few computational models proposed so far that explain switching of circuits (Yang et al., 2019). We may compare our V1 circuit to the recurrent circuits using FORCE learning, where a single unit or a few units project their feedback onto a recurrent neural net and momentarily disrupt chaotic activity to enable training. VIP units in our model precisely resemble such output units providing feedback in the FORCE framework, but it is unclear how far this analogy goes and to what extent the framework in Sussillo and Abbott (2009) is helpful in understanding V1 circuitry.

Another interesting example of a circuit with flexible, context-dependent behavior has been proposed by Mante et al. (2013), where prefrontal cortex (PFC) activity is modulated by the presence of a visual cue signaling which feature (color versus direction) the animals must integrate in a random-dots decision task. PFC functionality in this task has been modeled using a recurrent neural network (RNN) that takes the direction of motion, color of random dots, and visual cue as input and outputs the appropriate, reward-generating direction to saccade. This suggests the RNN enacts a potentially new mechanism for selection and integration of context-dependent inputs, with gating possible because the representations of the inputs and the upcoming choice are separable at the population level, even though they are deeply entangled at the single neuron level. The architecture of the model RNN proposed in this study is simpler than what we have laid out while also attaining high flexibility. There are important differences between the framework outlined in this article and our work. First, it is unclear what the number of weights in the network might be for the circuit in Mante et al. (2013) to be multitasking. One of our main motivations has been to achieve a switching circuit with few added units and weights, so that the circuit has fewer weights to learn than two separate circuits processing the two

contexts independently. It is unclear if this potential advantage holds in the case of Mante et al. Second, our circuit adapts to the statistics of both static and moving scenes and yields firing rates that are optimal for visual processing in either context. In the case of Mante et al., the circuit does not change momentary input processing when the context changes; it simply adapts its dynamics to integrate the appropriate feature and initiate the action that will be rewarded. Context takes on different meanings in these two instances: in our model, context is given by the statistical regularities of a certain environment, static or moving and in Mante et al., context refers to an input cue that changes the goals and reward dependencies of actions within the task. Importantly, we have focused on switching circuits that modulate their responses to different sensory contexts, as opposed to different input cues and behaviors. It is unclear whether identical or different mechanisms for switching apply in the case of sensory processing or action selection, when the animal changes scene statistics or behaviors, respectively.

Although our model is faithful to some aspects of the biology of V1 circuits, it has several limitations. First, it has been reported that during animal locomotion, firing rates of neurons more than double, at least in layers II/III of V1. Our firing rates are normalized to sum to one across features and cannot reproduce a doubling occurring uniformly over features. Second, the model does not reproduce a few experimental findings as reported in Ayaz, Saleem, Scholvinc, and Carandini (2013) and Keller et al. (2020). For instance, locomotion does not increase spontaneous activity as found by sequentially showing, to the static and moving circuits, images where every pixel takes on a constant value or images with gaussian white noise (0.13 versus 0.11 mean static, moving activity for constant pixel images; 0.063 versus 0.53 mean static, moving activity for gaussian white noise images/videos). Similar to Keller et al., the firing rate due to the cross-oriented surround is only slightly higher than the firing rate due to the iso-oriented surround (0.087 versus 0.085, p-value < 0.015; see Figure S14 for stimuli shown to the circuits). However, locomotion does weaken signals conveying surround suppression as reported in Ayaz et al. through the inhibition of the SST population by the VIP.

Moreover, another study (Dadarlat & Stryker, 2017) reported that noise correlations are reduced during motion, but this does not occur in our model. Further, we model VIP as a switch that is off during the static condition and has an activation during locomotion dependent on input images, whereas data show VIP activity is modulated at a finer scale and correlates strongly with speed (Fu et al., 2014). In addition, VIP switching units in our model turn on based on perfect knowledge of whether the animal is static or moving, rather than based on more subtle time-varying visual or motor features. Furthermore, data from Ko et al. (2011), Pfeffer et al. (2013), Jiang et al. (2015), Hofer et al. (2011), Lefort, Tomm, Floyd Sarria, and Petersen (2009), Thomson, West, Wang, and Bannister (2002), and Cauli et al.

(1997) on connection probabilities and strengths between neuron populations present a richer, more complex picture than our simplified circuit. There is wide-ranging connectivity to and from PV, there are strong connections from PYR to SST in most layers, and the weights from SST to VIP are strong (in terms of both connection probability and strength across layers), details that our simplified model cannot describe. Enabling weights from SST to VIP showed that we can similarly infer weights to and from VIP so that we are able to approximate the circuit during the moving condition (see Figures S6a and S6b). However, there are still many more potential connectivity structures between neuron populations our model does not describe.

From a computational perspective, our model makes several simplifications in describing context integration in circuits tuned to the statistical regularities of natural scenes. These include approximating a product with a sum in equation 4.12 and ignoring higher-order surround modulation going from equation 4.6 to 4.8. Furthermore, our equations have omitted terms explicitly describing feedback from higher-order areas. Top-down input to the VIP that mediates increase of local PYR activity has been reported, for example, in Zhang et al. (2014), Wilmes and Clopath (2019), Hertäg and Sprekeler (2019), Batista-Brito, Zagha, Ratliff, & Vinck (2018), and Wall et al. (2016). In our model, terms modulating the VIP firing rate causing the neuronal population to have a switch-like behavior have been essentially encapsulated into the binary s_t variable in equation 4.27. Despite the fact that incorporating cell type-specific contributions of top-down feedback in our model is an avenue of clear importance to relate to recent experimental findings, we leave this to future work. For simplicity, we have also limited the basis set of filters to one that extracts information about oriented edges in natural scenes. However, the computation of the extraclassical receptive fields need not be intrinsically limited to simple cells responding to Gabor-like filters but can be extended to encompass neurons responding to more complex features in areas beyond V1. Switching circuits can occur more generally, including in somatosensory and auditory cortices, where some of the same neuronal populations interact using similar circuit logic (Niell & Stryker, 2010; Bigelow, Morrill, Dekloe, & Hasenstaub, 2019). Populations of neurons in general switching circuits can respond to diverse stimuli (e.g., the VIP in auditory cortex are activated by punishment in Pi et al., 2013).

The theoretical framework here did not make assumptions regarding the completeness of the basis. Instead, it focuses specifically on interactions outside the classical receptive field. Prior work of Olshausen (2013), Olshausen and Field (1996a, 1996b, 1997), and Lewicki and Sejnowski (2000) have discussed extensively the benefits of overcomplete bases. The key feature in our model is the normalization of the activity of the neurons in patch, not the orthogonality or completeness of the basis (indeed, 34 filters used here are not orthogonal). In our model, the interactions outside the classical receptive field of a cell are expressed exclusively on the representations by the cells with classical receptive fields in that location. As such, features

not represented in an incomplete basis will be ignored in the context calculations. We use a relatively simple linear model for the classical receptive field formation. If there are nonlinear interactions in the classical receptive field, the model can be expanded to represent covariance of neuronal activities rather than covariance of projections on a linear filter; however, the analysis of such an extension is beyond the scope of the current study.

Here, we showed how a biologically inspired switching mechanism can enable a network to efficiently process stimuli in two different conditions. Most artificial neural networks (ANNs) suffer from what has been termed catastrophic forgetting, by which previously acquired memories are overwritten once new tasks are learned. Conversely, humans and other animals are capable of transfer learning—the ability to use past information without overwriting previous knowledge. Proposed solutions to this problem, like elastic weight consolidation or intelligent synapses, are discussed in Kirkpatrick et al. (2017), Zenke et al. (2017), and Mallya, Davis, and Lazebnik (2018). When applied to a narrow condition of learning new contexts, our work adds a switching mechanism based on the connections among different cell types in V1. This may open new doors to artificial neural networks with analogous switching architectures.

4 Methods

4.1 A Theory of Optimal Integration of Static Context in Images. A theory of optimal context integration, first outlined in Iyer et al. (2020), describes a probabilistic framework for inferring features at particular locations of an image given the features at surrounding locations. The probabilities of these feature occurring and co-occurring are then mapped to elements of a biological circuit (firing rates, weights).

4.1.1 Neuronal Code. We assume the firing rate of neurons to be a function of the probability of a feature being present at a specific location of the image:

$$\mathbf{f}^m_{k,X} = g(p(\mathbf{F}^m_k | i_X)), \tag{4.1}$$

where $\mathbf{f}^m_{k,X}$ represents the firing rate due to the classical receptive field of a neuron coding for feature $\mathbf{F}_k$ at location m in response to image i_X and g is a monotonic function. For every image and every location, we impose a normalization over features:

$$\sum_k p(\mathbf{F}^m_k | i_X) = \sum_k g^{-1}(\mathbf{f}^m_{k,X}) = 1. \tag{4.2}$$

Thus, the sum over probabilities of features adds up to 1. Throughout the article, we assume $g(y) = y$, although the model may be applied with other monotonic functions as well.

4.1.2 Probabilistic Framework. We subdivide the image X into N patches that correspond to the classical receptive fields of neurons. Thus, we have

$$p(\mathbf{F}_k^m | i_X) = p(\mathbf{F}_k^m | i_X^1, i_X^2, \ldots, i_X^N). \tag{4.3}$$

We will assume from this point forward that the firing rates are in response to an image X (i_X) but omit the subscript X to simplify the notation.

We first look at the simple case where there are only two patches: the classical receptive field (patch i^m) and the surround, which is part of the extraclassical receptive field (patch i^n). We will take into account other surrounding patches later when we perform an order expansion from $p(\mathbf{F}_k^m | i^m, i^n)$ to $p(\mathbf{F}_k^m | i^1, i^2, \ldots, i^N)$. The aim in the simple case with two patches is to infer to what extent feature $\mathbf{F}_k$ at patch i^m, denoted by $\mathbf{F}_k^m$, is present given information from both the classical receptive field and the surrounding extraclassical receptive field. Using Bayes' rule and simple probabilistic relations, we sum over all possible features $\mathbf{F}_j^m$ in patch i^m to get

$$p(\mathbf{F}_k^m | i^m, i^n) = \sum_j p(\mathbf{F}_k^m | i^m, i^n, \mathbf{F}_j^n) p(\mathbf{F}_j^n | i^m, i^n). \tag{4.4}$$

We can simplify the above relation by assuming the surround contribution from i^n does not contain higher-order surround information; instead, it includes only data from the classical receptive field: $p(\mathbf{F}_k^m | i^m, i^n, \mathbf{F}_j^n) \approx p(\mathbf{F}_k^m | i^m, \mathbf{F}_j^n)$. Our previous probabilistic statement, equation 4.4, thus becomes

$$p(\mathbf{F}_k^m | i^m, i^n) = \sum_j p(\mathbf{F}_k^m | i^m, \mathbf{F}_j^n) p(\mathbf{F}_j^n | i^m, i^n). \tag{4.5}$$

Using Bayes' rule for the first term,

$$p(\mathbf{F}_k^m | i^m, \mathbf{F}_j^n) = \frac{p(\mathbf{F}_j^n | \mathbf{F}_k^m, i^m) p(\mathbf{F}_k^m | i^m)}{p(\mathbf{F}_j^n | i^m)}, \tag{4.6}$$

equation 4.5 becomes

$$p(\mathbf{F}_k^m | i^m, i^n) = p(\mathbf{F}_k^m | i^m) \sum_j \frac{p(\mathbf{F}_j^n | i^m, \mathbf{F}_k^m)}{p(\mathbf{F}_j^n | i^m)} p(\mathbf{F}_j^n | i^m, i^n). \tag{4.7}$$

Assuming that we can ignore higher-order contributions due to surround modulation (i.e., the surround modulation of the surround), we can make the following simplifications: $p(\mathbf{F}_j^n|i^m, \mathbf{F}_k^m) \approx p(\mathbf{F}_j^n|\mathbf{F}_k^m)$, $p(\mathbf{F}_j^n|i^m) \approx p(\mathbf{F}_j^n)$, and $p(\mathbf{F}_j^n|i^m, i^n) \approx p(\mathbf{F}_j^n|i^n)$. This way, patch i^n is in the surround of patch i^m and modulates the firing rate due to i^m, but we are not concerned about the further effect i^n has on i^m. Then equation 4.6 thus becomes

$$p(\mathbf{F}_k^m|i^m, \mathbf{F}_j^n) = \frac{p(\mathbf{F}_j^n \cap \mathbf{F}_k^m)p(\mathbf{F}_k^m|i^m)}{p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)}. \tag{4.8}$$

The original equation, 4.4, becomes

$$p(\mathbf{F}_k^m|i^m, i^n) = p(\mathbf{F}_k^m|i^m)\sum_j \left(1 + \frac{p(\mathbf{F}_j^n \cap \mathbf{F}_k^m) - p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)}{p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)}\right) p(\mathbf{F}_j^n|i^n) \Leftrightarrow \tag{4.9}$$

$$p(\mathbf{F}_k^m|i^m, i^n) = p(\mathbf{F}_k^m|i^m)\left(1 + \sum_j \frac{p(\mathbf{F}_j^n \cap \mathbf{F}_k^m) - p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)}{p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)} p(\mathbf{F}_j^n|i^n)\right). \tag{4.10}$$

The last equivalence holds because we have assumed in equation 4.2 that all probabilities sum to 1.

We can now go from two patches to N patches that cover the entire image: $i^1, i^2, \ldots, i^N$. We further assume that each patch provides independent information to a neuron coding for $\mathbf{F}_k^m$ so that we obtain

$$\begin{aligned} p(\mathbf{F}_k^m|i) &= p(\mathbf{F}_k^m|i^1, i^2, \ldots, i^N) \\ &= p(\mathbf{F}_k^m|i^m) \cdot \prod_{n\neq m}^{N} \left(1 + \sum_j \frac{p(\mathbf{F}_j^n \cap \mathbf{F}_k^m) - p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)}{p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)} p(\mathbf{F}_j^n|i^n)\right). \end{aligned} \tag{4.11}$$

If the contribution from each patch is very small, we can ignore the higher-order terms in equation 4.11 and apply the approximation $\prod_i(1 + x_i) \approx 1 + \sum_i x_i$ for $x_i \ll 1$:

$$\begin{aligned} p(\mathbf{F}_k^m|i) &= p(\mathbf{F}_k^m|i^1, i^2, \ldots, i^N) \\ &= p(\mathbf{F}_k^m|i^m) \cdot \left(1 + \sum_{n,n\neq m}\sum_j \frac{p(\mathbf{F}_j^n \cap \mathbf{F}_k^m) - p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)}{p(\mathbf{F}_j^n)p(\mathbf{F}_k^m)} p(\mathbf{F}_j^n|i^n)\right). \end{aligned} \tag{4.12}$$

4.1.3 Mapping from the Probabilistic Framework to a Neural Network. Using a simple neural code with $g(x) = x$, so that the firing rate represents the

probability of feature presence, we obtain a simple mapping to a network of neurons. We denote

$$\mathbf{W}_{kj}^{mn} = \frac{p(\mathbf{F}_k^m \cap \mathbf{F}_j^n) - p(\mathbf{F}_k^m)p(\mathbf{F}_j^n)}{p(\mathbf{F}_k^m)p(\mathbf{F}_j^n)} = \frac{p(\mathbf{F}_k^m \cap \mathbf{F}_j^n)}{p(\mathbf{F}_k^m)p(\mathbf{F}_j^n)} - 1 \tag{4.13}$$

and map $\mathbf{W}_{kj}^{mn}$ to the synaptic weight between neurons responding preferentially to features $\mathbf{F}_k^m$ and $\mathbf{F}_j^n$, respectively. Then equation 4.12 becomes

$$p(\mathbf{F}_k^m|i) = p(\mathbf{F}_k^m|i^m) \cdot \left(1 + \sum_{n,n\neq m} \sum_j \mathbf{W}_{kj}^{mn} p(\mathbf{F}_j^n|i^n)\right). \tag{4.14}$$

We can also map firing rates to probabilities: $\mathbf{r}_k^m = p(\mathbf{F}_k^m|i)$ and $\mathbf{f}_k^m = p(\mathbf{F}_k^m|i^m)$, where $\mathbf{r}_k^m$ is the firing of the neuron with receptive field at patch m and most responsive to feature $\mathbf{F}_k$, and $\mathbf{f}_k^m$ is the firing rate of the same neuron due to just the classical receptive field i^m. As we recognize below, inferring these firing rates from our image and video data sets requires rectification and normalization so that $\mathbf{f}$ and $\mathbf{r}$ can be interpreted as probabilities.

The formula for synaptic weight can be expressed based on average activities of cells, when X spans a comprehensive set of natural images:

$$\mathbf{W}_{kj}^{mn} = \frac{\langle \mathbf{r}_k^m \mathbf{r}_j^n \rangle_X}{\langle \mathbf{r}_k^m \rangle_X \langle \mathbf{r}_j^n \rangle_X} - 1. \tag{4.15}$$

These weights can be achieved using Hebbian learning in an unsupervised manner. To avoid writing implicit equations for the firing rates, which are difficult to solve, and to make the computation tractable in practice without requiring learning, we use an approximation that requires only $\mathbf{f}$, the firing rates due to the classical receptive fields:

$$\mathbf{W}_{kj}^{mn} \approx \frac{\langle \mathbf{f}_k^m \mathbf{f}_j^n \rangle_X}{\langle \mathbf{f}_k^m \rangle_X \langle \mathbf{f}_j^n \rangle_X} - 1. \tag{4.16}$$

Finally, the probabilistic equations, 4.12 to 4.14, outlined above can be rewritten in terms of biologically relevant quantities like firing rates and synaptic weights by applying the appropriate mappings,

$$\mathbf{r}_k^m = \frac{1}{\mathbf{L}^m} \mathbf{f}_k^m \prod_{n,n\neq 1} \left(1 + \sum_j \mathbf{W}_{kj}^{mn} \mathbf{f}_j^n\right), \tag{4.17}$$

or, more simply,

$$\mathbf{r}_k^m \approx \frac{1}{\mathbf{L}^m}\mathbf{f}_k^m\left(1+\sum_{n,n\neq m}\sum_j \mathbf{W}_{kj}^{mn}\mathbf{f}_j^n\right), \tag{4.18}$$

when lateral connections given by $\mathbf{W}_{kj}^{mn}$ all sum up together to have a multiplicative effect. Here $\mathbf{L}^m$ is a normalization coefficient for patch i^m since we require

$$\sum_k \mathbf{r}_k^m = 1 \tag{4.19}$$

and thus denote

$$\mathbf{L}^m = \sum_k \mathbf{f}_k^m \cdot \prod_{n\neq m}^{N}\left(1+\sum_j \mathbf{W}_{kj}^{mn}\mathbf{f}_j^n\right). \tag{4.20}$$

As outlined in Iyer et al. (2020), this can be implemented in a network in which a set of neurons responsible for normalization have a divisive effect on the neurons, are patch-specific (have a classical receptive field of similar size to the neurons), inhibit equally all the neurons in their image patch, are untuned to features in the visual space, and receive inputs equal to the average of the inputs of the neurons in the patch.

4.2 Computing the Synaptic Weights. To compute weights according to equation 4.16, we first compute $\mathbf{f}_k^n$, the firing rates due to the classical receptive field for every image X in a large data set. Initially, we preprocess the image: we convert the image to grayscale, subtract the mean, and normalize the image to have a maximum value of 1. Similarly, we preprocess the filters so the mean of each is 0. $\mathbf{f}_k$ is the result of convolving X with feature k, rectifying and then normalizing so that at each location n, the sum over features k of firing rates $\mathbf{f}_k^n$ is equal to 1. Rectification ensures that firing rates are nonnegative, while normalization further ensures we can interpret $\mathbf{f}$ as probabilities. We average these firing rates over all images X in the data set to obtain $\langle \mathbf{f}_k^n \rangle_X$ for each feature k. The feature co-occurrence probability given by $\langle \mathbf{f}_k^m, \mathbf{f}_j^n \rangle_X$ in the numerator for the synaptic weight formula is then computed by further pairwise convolution of firing rates due to the classical receptive field for each possible pair of filters in the basis set and each image in the data set and then averaged over all images.

For a data set of videos, formula 4.16 becomes

$$\mathbf{W}_{k_1,k_2}^{m,n,\Delta t} = \frac{\langle \mathbf{f}_k^{m,t}, \mathbf{f}_j^{n,t-\Delta t} \rangle_{\text{frames}}}{\langle \mathbf{f}_k^{m,t} \rangle_{\text{frames}} \langle \mathbf{f}_j^{n,t-\Delta t} \rangle_{\text{frames}}} - 1. \tag{4.21}$$

The feature co-occurrence probability given by $\langle \mathbf{f}_k^{m,t}, \mathbf{f}_j^{n,t-\Delta t} \rangle_{\text{frames}}$ is computed by convolution of firing rates due to the classical receptive field at different frames (t and $t - \Delta t$) for each video and averaged over all videos and video frames. The assumption here is that extraclassical effects are delayed by a time Δt that corresponds to the time between movie frames or, biologically, corresponds to the synaptic delay.

We first assume translational invariance so that only the relative position of two filters is relevant: $\mathbf{W}_{j_1,j_2}^{n_1,n_2} = \mathbf{W}_{j_1,j_2}^{n_3,n_4}$ when $\vec{n_1} - \vec{n_2} = \vec{n_3} - \vec{n_4}$. The assumption that weights act with translational invariance allows rewriting the connectivities as simply a function of the distance, in image space, between the receptive field centers of the two neurons. Second, the mathematical validity of our probabilistic framework relies on the assumption that patches in the visual space, representing receptive fields of neurons, contain independent information. To reconcile this assumption with our empirically derived weights, we only consider connections between neurons whose receptive fields are sufficiently far apart, regardless of their corresponding feature identity. This leads to the use of sparse weights for moving and static contexts (see Figure 4e), where the only nonzero weights we allow in W are spatially farther apart than a minimum distance, which is half of the receptive field size. More precisely, for every feature k, synaptic weights from target filters were sampled in steps of $0.5\times$ the receptive field size at three distances in each direction around (0,0), so that we have synaptic weights on a (7×7) grid (three connections to the left/up + 3 connections to the right/down + self-connection = 7). Instead of using these sparse weights after sampling, we could have also rescaled the original, nonsparse weights by a scalar α so that $\|\mathbf{W}^{\text{static/moving (sparse)}} - \alpha \mathbf{W}^{\text{static/moving}}\| \approx 0$. Searching over possible values of α, we find $\alpha \approx 1/50$. We choose, however, to work with sparse weights or test our results on the original nonsparse weights without worrying about the rescaling by α. Although results presented in this study are largely for sparse weights, we have checked that the main results also hold when using full connectivity, at least for small $\Delta t \in \{1, 2\}$ (see Figure S6a). Further, assuming that the contribution due to context integration decays as the filters are spatially farther and farther apart, we can limit the weights in space to three times the size of the classical receptive field. Sample synaptic weights obtained using this procedure are shown in Figure 4e (and Figures 4d and 4f without the sampling of weights).

4.3 Constructing the Feature Space for Natural Images and Videos. We chose a basis of spatial filters that was constructed as outlined in Iyer et al. (2020). This is done by averaging approximations of spatial receptive field sizes from 212 recorded neurons in V1 (Durand et al., 2016). This set of filters is our first feature space and consists of four classes of spatial RFs observed experimentally: ON (1 feature), OFF (1 feature), and two versions of ON/OFF neurons (8 features each, for a total of 16), with the first version having a stronger ON subfield and the second a stronger OFF subfield. Each subfield was modeled as a 2D gaussian with a standard deviation of $\sigma = 0.5\times$ average subfield size, which was measured to be 4.8 degrees for the OFF subfield and 4.2 degrees for the ON subfield. The relative orientation between two subfields for each ON/OFF class was varied uniformly in steps of 45 degrees, from 0 to 315 degrees. Also for the ON/OFF class, the relative distance between the centers of the ON and OFF subfields was chosen to be 5 degrees, which equates to roughly 2σ. According to the data, the amplitude of the weaker subfield is chosen to be half that of the stronger subfield, whose highest amplitude was chosen to be unity. These two subfields are then combined additively to form a receptive field whose size is 7 degrees (the distance between the two subfields plus σ). The set of 18 features is shown in Figure 3d.

We then added 16 more filters with a temporal component, for a total of 34 filters. These filters have two frames with the first frame being one of the ON/OFF filters. The second frame is the ON/OFF filter in the previous frame shifted 3 pixels to the left, which matches the distance the sliding window moves every frame to generate the video. Such a spatiotemporal filter is shown in Figure 3e.

4.4 Data Sets of Natural and Synthetic Images and Videos.

4.4.1 Natural Images and Videos. For the data set of images, we used the Berkeley Segmentation Dataset (BSDS) training and test data sets (Martin et al., 2001). The training data set consists of 200 images of animals, human faces, landscapes, buildings, and so on and is used compute the weights $\mathbf{W}^{\text{static}}$. This same training set is then employed to construct the data set of 200 videos where a sliding window moves across the image for each frame of the video. In the simple case, the sliding window (167×167) moves 3 pixels per frame in the horizontal direction across the image (321×481 or 481×321), from left to right for 50 frames (see Figure 3b). The sliding window may also move in any random direction, resulting in different statistics of the video data set and hence different $\mathbf{W}^{\text{moving}}$. This different data set of videos is generated by choosing any pixel in the image and moving the sliding window toward it in smaller increments until that pixel is reached; a new pixel is then chosen from the image until the maximum limit of frames in the video (50 frames). Results from this different data set are shown in Figures S1 and S2. We further get 100 images from the BSDS test set to

generate the corresponding 100 videos and use in the optimization problem. These video frames are provided as input to the optimizer that minimizes the loss functions $E_{switch,1}$ and $E_{switch,2}$ to find $\mathbf{f}^{\alpha}$, $\mathbf{f}^{\beta}$ for $E_{switch,1}$ and $\mathbf{W}^{VIP\to SST}$, $\mathbf{W}^{VIP\to PYR}$, and $\mathbf{W}^{PYR\to VIP}$ for $E_{switch,2}$. For both optimization problems, we set 50 frames aside from these 100 videos to compute the generalization error during the minimization procedure.

In order to generate the numbers in Figure 8, another set of 100 videos generated from BSDS testing data set is altered by adding gaussian and salt-and-pepper noise of different parameters to each frame. The resulting noisy video frames are used to establish the ability of the switching circuit to do visual processing of stimuli with better reconstruction capability than the circuit implementing the static extraclassical receptive field or without extraclassical receptive field (see section 2.6). Gaussian white noise has standard deviation $\sigma = 0.5$ for reconstructions in Figure 8e, while salt-and-pepper noise turns pixels black or white with probability $p = 0.2$ each, for reconstructions in Figures 8d and 8f. Parameters σ and p are varied ($\sigma \in [0.5, 3]$, $p \in [0.05, 0.3]$) in Figure 8f.

4.4.2 Synthetic Data Sets of Images and Videos of Horizontal and Vertical Bars. This simple synthetic data set consists of 18 images of horizontal and vertical bars (9 horizontal, 9 vertical). Images are 9×9, each image having a bar at a different location. Videos consist of bars moving in any direction one pixel at a time: left or right (for horizontal bars) and up or down (for vertical bars).

4.5 Deriving an Equation for PYR Firing Rate Consistent with V1 Circuit Architecture. Let $\mathbf{f}$ be the firing rate due to the classical receptive field, $\mathbf{r}$ the firing rate incorporating extraclassical receptive field information, and $W^{X\to Y}$ the weights between neuronal populations X, Y. We can write approximated expressions for firing rates of PYR, SST, VIP neurons at time t:

Case a: When there is no feedback connection from PYR to VIP:

$$\mathbf{r}^t_{PYR} = \mathbf{f}^t_{PYR} \circ (1+\mathbf{W}^{PYR\to PYR}\mathbf{r}^{t-1}_{PYR} + \mathbf{W}^{SST\to PYR}\mathbf{r}^{t-1}_{SST} + \mathbf{W}^{VIP\to PYR}\mathbf{r}^{t-1}_{VIP}) \tag{4.22}$$

$$\mathbf{r}^t_{SST} = \mathbf{f}^t_{SST} + \mathbf{W}^{VIP\to SST}\mathbf{r}^t_{VIP} \tag{4.23}$$

$$\mathbf{r}^t_{VIP} = s_t \cdot \mathbf{f}^t_{VIP}. \tag{4.24}$$

Case b: When there is feedback from PYR to VIP:

$$\mathbf{r}^t_{PYR} = \mathbf{f}^t_{PYR} \cdot (1 + \mathbf{W}^{PYR\to PYR}\mathbf{r}^{t-1}_{PYR} + \mathbf{W}^{SST\to PYR}\mathbf{r}^{t-1}_{SST} + \mathbf{W}^{VIP\to PYR}\mathbf{r}^{t-1}_{VIP}) \tag{4.25}$$

$$\mathbf{r}^t_{SST} = \mathbf{f}^t_{SST} + \mathbf{W}^{VIP\to SST}\mathbf{r}^t_{VIP} \tag{4.26}$$

$$\mathbf{r}^t_{VIP} = s_t \cdot \mathbf{W}^{PYR\to VIP}\mathbf{r}^t_{PYR}, \tag{4.27}$$

where $\mathbf{f}^t_{VIP}$ of equation 4.24 is the intrinsic firing rate of VIP and s_t is a binary variable that takes the value 1 during the moving condition and 0 during the static condition. For the analysis of the firing rate during movement, we assume $s_t = 1$. Equations 4.22 and 4.25, expressing the firing rate $\mathbf{r}^t_{PYR}$ of the PYR population, assume the extraclassical receptive field contribution given by lateral connections has a multiplicative effect on the feedforward activities $\mathbf{f}_{PYR}$. This multiplicative gain is the result of mapping from the probabilistic framework of equations 4.14 and 4.18 and their analogs for the moving circuit activities and weights. This results in the network doing optimal inference of visual features via PYR firing rates as expressed in equations 4.22 and 4.25, and as detailed in section 2.1. The VIP firing rate $\mathbf{r}_{VIP}$ expression involves a binary gating term that switches based on state (static or moving), a simplification of what has been found empirically. The model could incorporate a term $\mathbf{f}_{VIP}$ into expression 4.27 describing VIP firing rates driven independently from PYR such that $\mathbf{r}^t_{VIP} = s_t \cdot \mathbf{W}^{PYR\to VIP}\mathbf{r}^t_{PYR} + \mathbf{f}_{VIP}$, but this change would not alter our main results. Finally, only the interneuron connections with the longest synaptic delay are assumed to be noninstantaneous (connections to and from PYR), while other connections are presumed to occur at a much faster timescale (connections between inhibitor neurons). Biologically, PYR are assumed to carry out computations by using dendritic trees, as outlined in Poirazi, Brannon, and Mel (2003), while SST and VIP are more spatially compact than PYR (Gouwens et al., 2019). Hence, synaptic delays between PYR and other neuron populations are longer than between other populations.

Making the appropriate substitutions in equations 4.22 and 4.25, we get the PYR firing rates, for case a,

$$\begin{aligned}\mathbf{r}^t_{PYR} = \mathbf{f}^t_{PYR} \circ [1 + \mathbf{W}^{PYR\to PYR}\mathbf{r}^{t-1}_{PYR} + \mathbf{W}^{SST\to PYR}(\mathbf{f}^{t-1}_{SST} + \mathbf{W}^{VIP\to SST}\mathbf{f}^{t-1}_{VIP}) + \\ + \mathbf{W}^{VIP\to PYR}\mathbf{f}^{t-1}_{VIP}]\end{aligned} \tag{4.28}$$

and for case b,

$$\begin{aligned}\mathbf{r}^t_{PYR} = \mathbf{f}^t_{PYR} \circ [1 + \mathbf{W}^{PYR\to PYR}\mathbf{r}^{t-1}_{PYR} \\ + \mathbf{W}^{SST\to PYR}(\mathbf{f}^{t-1}_{SST} + \mathbf{W}^{VIP\to SST}\mathbf{W}^{PYR\to VIP}\mathbf{r}^{t-1}_{PYR}) + \\ + \mathbf{W}^{VIP\to PYR}\mathbf{W}^{PYR\to VIP}\mathbf{r}^{t-1}_{PYR}].\end{aligned} \tag{4.29}$$

We can ignore further recurrence due to additional extraclassical receptive field contributions by making the approximation $\mathbf{r}_{PYR}^{t-1} = \mathbf{f}_{PYR}^{t-1}$. We are thus ignoring contextual surround modulation that is itself subject to surround influence—a "higher order" surround modulation—and instead consider only the classical receptive field response from surround neurons. These terms are small since this additional contribution is a linear combination of $\mathbf{f}_i\mathbf{f}_j$, $\mathbf{f}_i\mathbf{f}_j\mathbf{f}_k$, . . . where $\mathbf{f}_i$ are classical receptive field firing rates of neuron i and $0 \leq \mathbf{f}_i \leq 1$.

Additionally, we assume PYR and SST receive the same input so that $\mathbf{f}_{PYR}^t = \mathbf{f}_{SST}^t$. With these simplifications and dropping the subscript PYR for clarity, the equations for $\mathbf{r}_{PYR}^t$ become, for case a,

$$\begin{aligned}\mathbf{r}^t = \mathbf{f}^t \circ (1 &+ \mathbf{W}^{PYR\to PYR}\mathbf{f}^{t-1} + \mathbf{W}^{SST\to PYR}\mathbf{f}^{t-1} \\ &+ \mathbf{W}^{SST\to PYR}\mathbf{W}^{VIP\to SST}\mathbf{f}_{VIP} + \mathbf{W}^{VIP\to PYR}\mathbf{f}_{VIP}),\end{aligned} \tag{4.30}$$

which leads to

$$\mathbf{r}^t = \mathbf{f}^t \circ (1 + \mathbf{W}^{PYR\to PYR}\mathbf{f}^{t-1} + \mathbf{W}^{SST\to PYR}\mathbf{f}^{t-1} + \mathbf{W}^{SST\to PYR}\mathbf{f}^{\alpha} + \mathbf{f}^{\beta}), \tag{4.31}$$

where

$$\mathbf{f}^{\alpha} \equiv \mathbf{W}^{VIP\to SST}\mathbf{f}_{VIP} \tag{4.32}$$

and

$$\mathbf{f}^{\beta} \equiv \mathbf{W}^{VIP\to PYR}\mathbf{f}_{VIP} \tag{4.33}$$

while for case b,

$$\begin{aligned}\mathbf{r}^t = \mathbf{f}^t \circ (1 &+ \mathbf{W}^{PYR\to PYR}\mathbf{f}^{t-1} + \mathbf{W}^{SST\to PYR}\mathbf{f}^{t-1} \\ &+ \mathbf{W}^{SST\to PYR}\mathbf{W}^{VIP\to SST}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-1} \\ &+ \mathbf{W}^{VIP\to PYR}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-1}).\end{aligned} \tag{4.34}$$

During the static condition, there is no contribution from the VIP and $\mathbf{f}^t = \mathbf{f}^{t-1}$ so the firing rate becomes

$$\mathbf{r}^{\text{static}} = \mathbf{f} \circ (1 + \mathbf{W}^{PYR\to PYR}\mathbf{f} + \mathbf{W}^{SST\to PYR}\mathbf{f}). \tag{4.35}$$

However, we know from our theoretical framework that the firing rate during the static context can be written as

$$\mathbf{r}^{\text{static}} = \mathbf{f} \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}), \tag{4.36}$$

where $\mathbf{W}^{\text{static}}$ has been computed from the data set(s) of images and is a function of the average feature co-occurrence probability for pairs of spatial features. Therefore, we can consider a simple mapping that assigns $\mathbf{W}^{PYR\to PYR}$ and $\mathbf{W}^{SST\to PYR}$ to known weights: $\mathbf{W}^{PYR\to PYR} = \mathbf{W}_{+}^{\text{static}}$ and $\mathbf{W}^{SST\to PYR} = \mathbf{W}_{-}^{\text{static}}$, where $\mathbf{W}_{+}^{\text{static}}$ is the positive and $\mathbf{W}_{-}^{\text{static}}$ is the negative component of $\mathbf{W}^{\text{static}}$. The unknowns of equation 4.37 corresponding to the V1 circuit model with PYR-to-VIP connections are thus only three sets of weights to and from VIP: $\mathbf{W}^{VIP\to SST}$, $\mathbf{W}^{VIP\to PYR}$, $\mathbf{W}^{PYR\to VIP}$.

Finally, the equation for the firing rate of PYR neurons during the moving condition that we focus on throughout the article (with PYR projecting to VIP) becomes

$$\begin{aligned}\mathbf{r}^t &= \mathbf{f}^t \circ (1 + \mathbf{W}_{+}^{\text{static}}\mathbf{f}^{t-1} + \mathbf{W}_{-}^{\text{static}}\mathbf{f}^{t-1} \\ &\quad + \mathbf{W}_{-}^{\text{static}}\mathbf{W}^{VIP\to SST}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-1} + \mathbf{W}^{VIP\to PYR}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-1}) \\ &= \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-1} + \\ &\quad + \mathbf{W}^{SST\to PYR}\mathbf{W}^{VIP\to SST}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-1} + \mathbf{W}^{VIP\to PYR}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-1}). \quad (4.37)\end{aligned}$$

4.6 Reconstructions from Noisy Videos Using Firing Rates and Optimal Synaptic Weights of Different Circuit Architectures. To gain insight into how optimal synaptic weights can facilitate decoding of information present in the neuronal activity, we reconstructed natural image frames from videos using four distinct circuits. The firing rates in these circuits are described by the following equations:

$$\mathbf{r}^{\text{no EXC}}(t) = \mathbf{f}^t, \quad (4.38)$$

$$\mathbf{r}^{\text{static}}(t) = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t}), \quad (4.39)$$

$$\mathbf{r}^{\text{moving}}(t) = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{moving}}\mathbf{f}^{t-\Delta t}), \quad (4.40)$$

$$\mathbf{r}^{\text{approx}}(t) = \mathbf{f}^t \circ (1 + \mathbf{W}^{\text{static}}\mathbf{f}^{t-\Delta t} + \mathbf{W}^{SST\to PYR}\mathbf{W}^{VIP\to SST}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-\Delta t} \quad (4.41)$$

$$+ \mathbf{W}^{VIP\to PYR}\mathbf{W}^{PYR\to VIP}\mathbf{f}^{t-\Delta t}). \quad (4.42)$$

The first equation above describing $\mathbf{r}^{\text{no EXC}}$ relies solely on the feedforward information where no extraclassical receptive field contribution is included. The next two expressions restate how the firing rates for the static and moving circuits require contributions from the extraclassical receptive fields through lateral connections $\mathbf{W}^{\text{static}}$, $\mathbf{W}^{\text{moving}}$, reflective of the statistical regularities of images/videos. Equation 4.42 describes the switching circuit we have implemented and characterized above and should approximate the firing rate in the moving circuit when VIP are active: $\mathbf{r}^{\text{moving}} \approx \mathbf{r}^{\text{approx}}$.

The reconstruction was performed as follows. For any noisy input image $X + \xi$, where ξ is some random variable representing a noisy process, we calculated the effective firing rate (activity) $\mathbf{r}$ of neuron/feature k at location

n using equations 4.38 to 4.42. To reconstruct image frames from firing rates, we convolved the firing rates computed with the inverses of the filters in our basis set. More specifically, the activity $\mathbf{r}_k$ corresponding to filter k was convolved with the inverse of k, which was obtained by flipping k about the horizontal and vertical axes. These convolutions for all filters were then averaged to obtain the final reconstruction.

We then performed the reconstruction for the same image frame X without any noise added. We assessed the denoising capability of our circuits by computing the Pearson correlation coefficient ρ between the reconstruction of $X + \xi$ and the reconstruction of X. The latter is a baseline for our comparisons, as there is no noise to remove from the image frame through extraclassical surround modulation. The Pearson correlation coefficient ρ is a function of the activity $\mathbf{r}$ of different circuit architectures and is discussed and compared across circuits in section 2.6.

Two further issues merit further discussion. First, if the spectral content of the noise and image frame is known, a Wiener deconvolution can be applied, which minimizes the mean square error between the estimated reconstruction and the original frame. Such a Wiener deconvolution would minimize the impact of deconvolved noise at frequencies with poor signal-to-noise ratio. However, we assume here that interpretation of signals is done without access to knowledge of this spectral content, but rather implementing a naive reconstruction as would be optimal in the noise-free limit. Second, given the presence of extraclassical surround contribution, the deconvolution operation may be more complex than the simple, filter by filter, convolution with the inverse filter $\mathbf{F}^T$. Specifically, the inverse may contain information about the cross-correlation of features. Again we work in the simplifying limit in which this is not the case. We do not exclude, however, the possibility that the biological circuit may apply a more complex reconstruction (e.g., via learning weights), an interesting avenue to explore in future work.

4.7 Like-to-Like Connectivity for PYR and VIP Populations. In addition to interneuron connectivity discussed in section 2.7, PYR connection probability as a function of the difference in orientation tuning (see Figures S2c and S2d) qualitatively matches the same graph reported experimentally (Ko et al., 2011). This like-to-like connectivity, with neurons responding to similar features (orientations) more strongly connected, holds true for both static (shown in Iyer et al., 2020) and moving weights (shown in Figures S2c, S2d, and S3). Another feature concerns the amplitude of static and moving weights, which decreases with distance from the classical receptive field, with lower weights on average between neurons whose classical receptive fields are far away. Figure S2 shows the dependence of the maximum, minimum, and average positive and negative synaptic weights on the distance between neuronal receptive fields. Assuming an exponential spatial decay of weights with distance and using the first two points in the plot displaying

decreasing distance dependence in the mean positive static weights curve (see Figure S2a), we computed the spatial constants $D_{\text{static/moving}} = 0.8\times$ the classical receptive field size. This is in accordance with past findings (Angelucci and Bressloff, 2006; Iyer et al., 2020), suggesting that the near surround extends over a range similar in size to the classical receptive field.

We further study the inferred connections to and from the VIP to establish whether these weights reflect the contextual statistics of static and moving states. We first inferred whether there is like-to-like connectivity between VIP and PYR populations by building a similarity matrix of dimension number of VIP neurons $\times$ number of PYR neurons that measures response similarity between VIP and PYR neuronal populations. Each entry of this response similarity matrix is computed by taking the Pearson correlation between the GLM coefficients found above (see section 2.7) for each VIP neuron and each PYR neuron, respectively. We next built, from our $Nf_2 \times 34 \times 3 \times 3$ tensor $\mathbf{W}^{VIP\to PYR}$ used for convolution, a matrix of connectivities of dimension number of VIP neurons $\times$ number of PYR neurons. Finally, taking the Pearson correlation coefficient between the response similarity matrix and the matrix of connectivities yields a statistically significant but very low correlation coefficient (-0.01, p-value < 0.01, Kolmogorov-Smirnov test). We conclude that while like-to-like connectivity is present between PYR neurons, this phenomenon is not prevalent between VIP and PYR populations.

4.8 Measuring Dimensionality with the Participation Ratio. We aim to characterize the dimensionality of the distribution of population vector responses representing neural activity. Across many trials, these population vectors populate a cloud of points. The dimensionality is a weighted measure of the number of axes explored by that cloud,

$$\text{Dim}(C) = \frac{(\text{Tr}C)^2}{\text{Tr}C^2} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2}, \tag{4.43}$$

where C is the covariance matrix of the matrix of neural activations and λ_i is the ith eigenvalue of the covariance matrix C. Dim(C) measures the dimensionality of neural activity of our network and is termed the *participation ratio*. The eigenvectors of the covariance matrix C are the axes of our cloud of points representing activity in neural space. If the neural activities are independent and all have equal variance, all the eigenvalues of the covariance matrix have the same value and $\text{Dim}(C) = N$. Alternatively, if the components are correlated so that the variance is evenly spread across M dimensions, only M eigenvalues would be nonzero and $\text{Dim}(C) = M$. For other correlation structures, this measure interpolates between these two regimes and, as a rule of thumb, the dimensionality can be thought as corresponding to the number of dimensions required to explain about 80% of

the total population variance in many settings (Mazzucato, Fontanini, & La Camera, 2016; Gao et al., 2020; Litwin-Kumar, Harris, Axel, Sompolinsky, & Abbott, 2017).

4.9 Inferring the Tuning Properties of VIP and PYR Neurons. We further study the activation patterns of units in our switching circuit model by inferring the tuning properties of VIP and PYR units. To achieve this, we first choose a wavelet family that will determine our features and differs from the basis approximating spatial receptive fields in V1 from section 4.3. We chose the Daubechies 4 wavelet family with a mother wavelet of length 15 pixels, as shown in Figure S13a. We then consider the 2D discrete wavelet transforms of our video frames to obtain the approximation, horizontal detail, vertical detail, and diagonal detail coefficients (wavelet transforms), respectively, for each frame. The goal is to use the averages of these coefficients as the independent variables of a linear regression or GLM that models VIP or PYR activations.

To achieve this, we first reduce the dimensionality of the wavelet transforms obtained above by considering $100 = 10 \times 10$ patches of size 5×5 that tile wavelet transforms of each video frame. Averaging over the spatial component of these patches, we obtain three sets of 10×10 coefficients (for the horizontal, vertical, and diagonal detail, respectively) that will be the independent variables of the linear regression or GLM. For each PYR/VIP neuron, we can regress its activity for every video frame against the 300 $(= 3 \times 10 \times 10)$ coefficients we have inferred,

$$a = C \cdot x, \tag{4.44}$$

where $a \in \mathbb{R}^{\text{no. frames}=4700}$ is the activity of a neuron (for every 4700 frames), $C \in \mathbb{R}^{\text{no. frames} \times \text{no. regressors}(300)}$ is the matrix of regressors, and $x \in \mathbb{R}^{\text{no. coefficients=no. regressors}}$ contains the unknowns that will determine the tuning of each neuron.

We obtain that most PYR neurons are tuned to horizontal features, and much less so to vertical features (data not shown). Using either a linear regression or a GLM with a Poisson distribution yields qualitatively similar results. Because VIP neurons in our model only get input from the PYR, while the top-down input activating VIP is described simply by the binary term s_t, we obtain that VIP acquires the same preferential selectivity to horizontal features to the detriment of vertical features (see Figure S13d). VIP neurons are tuned to horizontal features with an average regression coefficient of 0.65, while they are tuned to vertical features with an average regression coefficient of 0.015 (using the results from the linear regression). This runs counter to our expectation that VIP is capable of detecting horizontal movement in our data set by exhibiting preferential selectivity toward vertical features within their receptive fields, analogous to empirical

results in Millman et al. (2020). Clearly, the simplification we have made by employing a binary term s_t in equation 4.27 prevents us from observing a more realistic VIP activation pattern that would deviate from the PYR pattern and provide further insights. We leave the more detailed modeling expanding our current simplified model in this direction to future work.

Code

Source code is available in ModelDB (McDougal et al., 2017) at http://modeldb.yale.edu/267120.

Acknowledgments

We gratefully acknowledge the support of the Swartz Foundation Center for Theoretical Neuroscience at the University of Washington, and of NIH training grant 5 R90 DA 033461-08. We are grateful to Matthew Farrell and Kameron Harris for their helpful comments in producing the final manuscript. We thank Paul G. Allen, the founder of the Allen Institute for Brain Science, for his vision, encouragement, and support.

References

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feed-back connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. In S. Martinez-Conde (Ed.), *Progress in brain research*, vol. 2006 (pp. 93–120). Amsterdam: Elsevier. 10.1016/S0079-6123(06)54005-1

Ayaz, A., Saleem, A. B., Scholvinc, M. L., & Carandini, M. (2013). Locomotion controls spatial integration in mouse visual cortex. *Current Biology*, *23*, 890–894. 10.1016/j.cub.2013.04.012, PubMed: 23664971

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3). 10.1037/h0054663, PubMed: 13167245

Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

Batista-Brito, R., Zagha, E., Ratliff, J. M., & Vinck, M., (2018). Modulation of cortical circuits by top-down processing and arousal state in health and disease. *Current Opinion in Neurobiology*, *52*, 172–181. 10.1016/j.conb.2018.06.008, PubMed: 30064117

Bell A. J., & Sejnowski T. J., (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*, 1129–1159. 10.1162/neco.1995.7.6.1129, PubMed: 7584893

Bigelow, J., Morrill, R. J., Dekloe, J., & Hasenstaub, A. R. (2019). Movement and VIP interneuron activation differentially modulate encoding in mouse auditory cortex. *eNeuro*, *6*(5). 10.1523/ENEURO.0164-19.2019, PubMed: 31481397

Braitenberg V., & Schüz, A. (1991). *Anatomy of the cortex: Statistics and geometry*. Berlin: Springer-Verlag.

Campagnola, L., Seeman, S. C., Chartrand, T., Kim, L., Hoggarth, A., Gamlin, C., . . . Jarsky, T. (2021). *Local connectivity and synaptic dynamics in mouse and human neocortex*. https://www.biorxiv.org/content/10.1101/2021.03.31.437553v2

Cardin J., (2018). Inhibitory interneurons regulate temporal precision and correlations in cortical circuits. *Trends Neurosci.*, *41*, 689–700. 10.1016/j.tins.2018.07.015, PubMed: 30274604

Cardin J., (2019). Functional flexibility in cortical circuits. *Current Opinion in Neurobiology*, *58*, 175–180. 10.1016/j.conb.2019.09.008, PubMed: 31585330

Cauli, B., Audinat, E., Lambolez, B., Angulo, M. C., Ropert, N., Tsuzuki, M., . . . Rossier, J. (1997). Molecular and physiological diversity of cortical nonpyramidal cells. *J. Neurosci.*, *17*(10), 3894–3906. 10.1523/JNEUROSCI.17-10-03894.1997, PubMed: 9133407

Chalk, M., Marre, O., &Tkačik, G. (2018). Toward a unified theory of efficient, predictive, and sparse coding. In *Proceedings of the National Academy of Science USA*, *115*(1), 186–191. 10.1073/pnas.1711114115

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361. 10.1037/0033-295X.97.3.332, PubMed: 2200075

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, *36*, 287–314. 10.1016/0165-1684(94)90029-9

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., . . . Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, *518*(7539), 399–403. 10.1038/nature14182, PubMed: 25652823

Dadarlat, M. C., & Stryker, M. P. (2017). Locomotion enhances neural encoding of visual stimuli in mouse V1. *Journal of Neuroscience*, *37*(14), 3764–3775. 10.1523/JNEUROSCI.2728-16.2017, PubMed: 28264980

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*, 889–904. 10.1162/neco.1995.7.5.889, PubMed: 7584891

Dipoppa, M., Ranson, A., Krumin, M., Pachitariu, M., Carandini, M., & Harris, K. D. (2018). Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *Neuron*, *98*, 602–615. 10.1016/j.neuron.2018.03.037, PubMed: 29656873

Doi, E., & Lewicki, M. S. (2014). A simple model of optimal population coding for sensory systems. *PLOS Comput. Biol.*, *10*(8), e1003761. 10.1371/journal.pcbi.1003761, PubMed: 25121492

Durand, S., Iyer, R., Mizuseki, K., de Vries, S., Mihalas, S., & Reid, R. C. (2016). A comparison of visual response properties in the lateral geniculate nucleus and primary visual cortex of awake and anesthetized mice. *J. Neurosci.*, *36*(48), 12144–12156. 10.1523/JNEUROSCI.1741-16.2016, PubMed: 27903724

Fu, Y., Tucciarone, J. M., Espinosa, J. S., Sheng, N., Darcy, D. P., Nicoll R. A., . . . Stryker, M. P. (2014). A cortical circuit for gain control by behavioral state. *Cell*, *156*, 1139–1152.

Gao, P., Trautmann, E., Yu, B., Santhanam, G., Ryu, S., Shenoy, K., & Ganguli, S. (2020). *A theory of multineuronal dimensionality, dynamics and measurement*. https://www.biorxiv.org/content/early/2017/11/05/214262

Gouwens, N. W., Sorensen, S. A., Berg, J., Lee, C., Jarsky, T., Ting, J., . . . Koch, C. (2019). Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat. Neurosci.*, *22*(7), 1182–1195. 10.1038/s41593-019-0417-0, PubMed: 31209381

Gozzi, A., Jain, A., Giovanelli, A., Bertollini, C., Crestan, V., Schwarz, A. J., . . . Bifone, A. (2010). A neural switch for active and passive fear. *Neuron*, *67*, 656–666. 10.1016/j.neuron.2010.07.008, PubMed: 20797541

Harpur, G. F., & Prager, R. W. (1996). Development of low entropy coding in a recurrent network. *Network*, *7*, 277–284. 10.1088/0954-898X_7_2_007, PubMed: 16754387

Hertäg, L., & Sprekeler, H. (2019). Amplifying the redistribution of somatodendritic inhibition by the interplay of three interneuron types. *PLOS Comput. Biol.*, *15*(5), e1006999. doi:10.1371/journal.pcbi.1006999

Hofer S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., . . . Mrsic-Flogel, T. D. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, *14*, 1045–1052. 10.1038/nn.2876, PubMed: 21765421

Hu, B., & Mihalas, S. (2018). *Convolutional neural networks with extra-classical receptive fields*. https://arxiv.org/abs/1810.11594v1

Iyer, R., Hu, B., & Mihalas, S. (2020). Contextual integration in cortical and convolutional neural networks. *Front. Comput. Neurosci.*, April 23, 31.

Jiang, X., Shen, S., Cadwell, C. R., Berens, P., Sinz, F., Ecker, A. S., . . . Tolias, A. S. (2015). Principles of connectivity among morphologically defined cell types in adult neo-cortex. *Science*, *350* (6264), aac9462. 10.1126/science.aac9462

Karklin, Y., & Simoncelli, E. P. (2011). Efficient coding of natural images with a population of noisy linear-nonlinear neurons. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, *24* (pp. 999–1007). Red Hook, NY: Curran. 26273180

Keller, A. J., Dipoppa, M., Roth, M. M., Caudill, M. S., Ingrosso, A., K. D. Miller, & Scanziani M. A. (2020). *Disinhibitory circuit for contextual modulation in primary visual cortex*. https://www.biorxiv.org/content/10.1101/2020.01.31.929166v2.

Kirkpatrick, J., Pascanu, R., & Hadsel, R. (2017). Overcoming catastrophic forgetting in neural networks. *PNAS*, *114*(13), 3521–3526. 10.1073/pnas.1611835114, PubMed: 28292907

Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjostro, P. J., & Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, *473*(7345), 87–91, 5. 10.1038/nature09880, PubMed: 21478872

Koganezawa, M., Kimura, K., & Yamamoto, D. (2016). The neural circuitry that functions as a switch for courtship versus aggression in drosophila males. *Current Biology*, *26*, 1395–1403. 10.1016/j.cub.2016.04.017, PubMed: 27185554

Lefort, S., Tomm, C., Floyd Sarria, J. C., & Petersen, C. C. H. (2009). The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron*, *61*, 301–316. 10.1016/j.neuron.2008.12.020, PubMed: 19186171

Lewicki, M. S., & Sejnowski, T. J. (2000). Learning overcomplete representations, *Neural Computation*, *12*, 337–365. 10.1162/089976600300015826, PubMed: 10636946

Litwin-Kumar, A., Harris, K. D., Axel, R., Sompolinsky, H., & Abbott, L. F. (2017). Optimal degrees of synaptic connectivity. *Neuron*, *93*, 1153–1164.e7. 10.1016/j.neuron.2017.01.030, PubMed: 28215558

Mallya, A., Davis, D., & Lazebnik, S. (2018). Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In *Proceedings of the European Conference on Computer Vision*. Berlin: Springer.

Mallya, A., & Lazebnik, S. (2018). PackNet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*, 78–84. 10.1038/nature12742, PubMed: 24201281

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE.

Mazzucato, L., Fontanini, A., & La Camera, G. (2016). Stimuli reduce the dimensionality of cortical activity. *Front. Syst. Neurosci.*, *10*, 11. 10.3389/fnsys.2016.00011, PubMed: 26924968

McDougal, R. A., Morse, T. M., Carnevale, T., Marenco, L., Wang, R., Migliore, M., . . . Hines, M. L. (2017). Twenty years of ModelDB and beyond: Building essential modeling tools for the future of neuroscience. *J. Comput. Neurosci.*, *42*(1), 1–10. 10.1007/s10827-016-0623-7, PubMed: 27629590

Millman, D. J., Ocker, G. K., Caldejon, S., Kato, I., Larkin, J. D., Lee, E. K., . . . de Vries, E. J. (2020). VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli *eLife*, *9*, e55130. 10.7554/eLife.55130, PubMed: 33108272

Niell, C. M., & Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, *65*(4), 472–479. 10.1016/j.neuron.2010.01.033, PubMed: 20188652

Ollerenshaw, D. R., Zheng, H. J. V., Millard, D. C., Wang, Q., & Stanley, G. B. (2014). The adaptive trade-off between detection and discrimination in cortical representations and behavior. *Neuron*, *81*, 1152–1164. 10.1016/j.neuron.2014.01.025, PubMed: 24607233

Olshausen, B. A. (2013). Highly overcomplete sparse coding. In *Proceedings of SPIE* (vol. 8651). Bellingham, WA: SPIE.

Olshausen, B. A., & Field, D. J. (1996a). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609. 10.1038/381607a0

Olshausen, B. A., & Field, D. J. (1996b). Natural image statistics and efficient coding. *Network*, *7*(2), 333–339, 5. 10.1088/0954-898X_7_2_014

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.*, *37*, 3311–3325. 10.1016/S0042-6989(97)00169-7, PubMed: 9425546

Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., & Scanziani, M. (2013). Inhibition of inhi bition in visual cortex: The logic of connections between molecularly distinct interneurons. *Nat. Neurosci.*, *16*, 1068–1076. 10.1038/nn.3446, PubMed: 23817549

Pi, H. J., Hangya, B., Kvitsiani, D., Sanders, J., Huang, Z. J., & Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature*, *503*, 521–524. 10.1038/nature12676, PubMed: 24097352

Poirazi, P., Brannon, T., & Mel, B. W. (2003). Pyramidal neuron as two-layer neural network. *Neuron*, *37*, 989999. 10.1016/s0896-6273(03)00149-1

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1). 10.1038/4580, PubMed: 10195184

Rudy, B. (2011). Three groups of interneurons account for nearly 100% of neocor tical GABAergic neurons. *Dev. Neurobiol.*, *71*(1), 45–61. 10.1002/dneu.20853, PubMed: 21154909

Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., . . . Hadsell, R. (2016). *Progressive neural networks*. arXiv:1606.04671.

Serra, J., Suris, D., Miron, M., & Karatzoglou, A. (2018). Overcoming catastrophic forgetting with hard attention to the task. In *Proceedings of the 35th International Conference on Machine Learning* (pp. 4548–4557).

Simoncelli, E. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, *13*, 144–149. 10.1016/S0959-4388(03)00047-3, PubMed: 12744966

Sussillo, D., & Abbott, L. F., (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron*, *63*, 544–557. 10.1016/j.neuron.2009.07.018, PubMed: 19709635

Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., & Bertagnolli, D. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, *563*, 72–78. 10.1038/s41586-018-0654-5, PubMed: 30382198

Terekhov, A. V., Montone, G., & O'Regan, J. K. (2015). Knowledge transfer in deep block-modular neural networks. In S. Wilson, P. Verschure, A. Mura, & T. Prescott (Eds.), *Lecture Notes in Computer Science: Vol. 9222. Biomimetic and Biohybrid Systems. Living Machines*. Cham: Springer. 10.1007/978-3-319-22979-927

Thomson, A. M., West, D. C., Wang, Y., & Bannister, A. P. (2002). Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2–5 of adult rat and cat neocortex: Triple intracellular recordings and biocytin labelling in vitro. *Cerebral Cortex*, *12*(9), 936–953. 10.1093/cercor/12.9.936, PubMed: 12183393

Vogt, K., Zimmerman, D. M., Schlichting, M., Hernandez-Nuñez, L., Qin, S., Malacon, K., . . . Samuel, D. T. (2020). *Internal state configures olfactory behavior and early sensory processing in Drosophila larva*. 10.1101/2020.03.02.973941.

Wall, N. R., De La Parra, M., Sorokin, J. M., Taniguchi, H., Huang, Z. J., & Callaway, E. M. (2016). Brain-wide maps of synaptic input to cortical interneurons. *Journal of Neuroscience*, *36*(14), 4000–4009. 10.1523/JNEUROSCI.3967-15, PubMed: 27053207

Wilmes, K. A., & Clopath, C. (2019). Inhibitory microcircuits for top-down plasticity of sensory representations. *Nature Communications*, *10*, art. 5055. 10.1038/s41467-019-12972-2, PubMed: 31699994

Yang, G. R., Cole, M. W., & Rajan, K. (2019). How to study the neural mechanisms of multiple tasks. *Current Opinion in Behavioral Sciences*, *29*, 134–143. 10.1016/j.cobeha.2019.07.001, PubMed: 32490053

Zemel, R. S. (1993). *A minimum description length framework for unsupervised learning*, PhD thesis, University of Toronto.

Zenke, F., Poole, B., & Ganguli, S. (2017). Continual learning through synaptic intelligence. In *Proceedings of the International Conference on Machine Learning Research*, *70*, (pp. 3987–3995).

Zhou, T., Zhu, H., Fan, Z., Wang, F., Chen, Y., Liang, H., . . . Hu, H. (2017). History of winning remodels thalamo-PFC circuit to reinforce social dominance. *Science*, *357*, 162–168. 10.1126/science.aak9726, PubMed: 28706064

Received January 6, 2021; accepted September 21, 2021.