PREPRINT

# Deep-learning-assisted simulation of a cortical circuit: integrating anatomy, physiology and function

Shinya Ito[1], Darrell Haufler[1], Javier Galván Fraile[2], Kael Dai[1], Joseph Aman[1], Guozhang Chen[3], Claudio Mirasso[2], Wolfgang Maass[4], and Anton Arkhipov[1*]

[1]Allen Institute, Seattle WA, USA
[2]IFISC, Universitat de les Illes Balears, Palma de Mallorca, Spain
[3]Peking University, Beijing, China
[4]Graz University of Technology, Graz, Austria

*Correspondence: antona@alleninstitute.org

**Abstract**

Mechanistic understanding of the brain requires models constrained by anatomy, physiology, and functional activity. We present a differentiable simulator and a ∼67,000-neuron model of mouse primary visual cortex that integrates multimodal data, including electron-microscopy connectomics, multipatch synaptic physiology, cell-type-resolved electrophysiology, and large-scale Neuropixels recordings across diverse cell types. End-to-end training completes on a single GPU in ∼10 hours while preserving biological constraints. Networks trained only on brief drifting-grating responses reproduce cell-type-specific benchmarks and generalize to new contrasts and natural scenes. We uncover heterogeneous cell-type- and tuning-dependent synaptic organization and show that training preferentially sculpts inhibitory connectivity into distinct cohorts with strong causal influence on network activity. Targeted ablations show that removing biological priors on synaptic weight distributions can preserve functional activity yet disrupts emergent wiring rules. Our freely shared models and code facilitate differentiable simulations as a computationally practical framework for studying brain circuit function and mechanisms under biological constraints.

## Introduction

Biorealistic models of brain circuits can connect synapse-level organization to cellular-resolution neural dynamics and function by systematically integrating multi-modal data on circuit composition, connectivity, and activity, with the potential to transform our understanding of brain mechanisms, computations, and diseases[1–6]. However, constructing such models remains challenging due to limitations in the availability of the necessary data. Moreover, they remain difficult to constrain at scale. Even when a model reproduces selected neural activity statistics, multiple parameterizations can yield similar outputs[7–9], limiting identifiability and mechanistic interpretability. Reducing this degeneracy requires combining complementary constraints that restrict admissible parameter regimes and neural dynamics within a given circuit architecture.

Recent datasets are starting to provide such constraints with cellular resolution across modalities and in multiple species and brain areas. In particular, extensive characterization of neuronal intrinsic electrophysiology, synaptic connectivity, and in-vivo neural activity with unprecedented cell type resolution in the mouse visual cortex creates unique opportunities for biorealistic modeling[6]. Systematic multimodal characterization of cortical cell types—through transcriptomics, electrophysiology,

and morphological reconstruction—has yielded detailed taxonomies that serve as essential building blocks for constructing biorealistic models of mouse cortical circuits[10–16]. Large-scale synaptic physiology surveys have quantified cell type-specific synaptic strength and temporal dynamics[17,18]. Surveys of in vivo activity across cortical layers and cell types provide robust targets for model evaluation and fitting[19,20]. More recently, electron microscopy-based functional connectomics has begun to link synapse-level anatomy to functional tuning within the same tissue, enabling direct tests of predicted structure–function relationships[21]. Although full-brain connectomes have been obtained for other species—C. elegans[22,23] and Drosophila[24,25]—the abundance of cell type-resolved multi-modal data about composition, synaptic properties, and in vivo activity for the mouse visual cortex makes it an excellent system for integrative biorealistic circuit modeling.

Data-driven cortical models have started integrating subsets of these constraints. For example, a model of the mouse primary visual cortex (V1) by Billeh et al.[26] combined diverse—though rather incomplete—anatomical and physiological measurements available at the time to build a large recurrent spiking circuit that reproduced multiple response benchmarks and yielded testable hypotheses about circuit organization[27–31]. A remaining key bottleneck is optimization: highly recurrent spiking networks can be unstable and historically required laborious and extremely time-consuming manual tuning to maintain realistic activity regimes[26]. Recently, differentiable simulation frameworks started emerging that enable gradient-based optimization of biological spiking networks against empirical objectives, shifting functional datasets from post hoc validation to active constraints[32–34]. Nevertheless, efficient automatic optimization of large-scale, heterogeneous, biorealistic brain circuit models remains challenging, both in terms of the computational resources required and in terms of creating suitable optimization objectives or loss functions.

Here we present a deep learning-assisted simulation framework and a point-neuron recurrent spiking model of mouse V1 that integrates extensive new data from EM-derived connection statistics (MICrONS[21] and V1 DeepDive[35]), multipatch-derived synaptic strength and temporal dynamics[18], and large-scale in vivo recordings[19,20], resolving all cortical layers and diverse cell types. We provide a differentiable, GPU-accelerated simulator to enable efficient BPTT at scale, and we train the V1 network to match cell-type-specific firing-rate and selectivity statistics measured in vivo while preserving biological architecture and the experimentally determined synaptic weight distributions. This approach yields trained circuits that generalize well beyond the training stimuli and expose synaptic organization that can be compared directly to emerging functional connectomics.

Our GPU-accelerated framework trains the full ∼67,000-neuron V1 model end-to-end in approximately half a day on a single mainstream GPU, and the resulting network can be simulated faster than real time, making large-scale biorealistic circuit modeling accessible without supercomputing infrastructure. We show that models trained only on gray screen and a narrow set of drifting grating responses generalize to natural images and previously unseen contrast regimes, reproducing cell-type-specific firing rates, selectivity, and gain modulation consistent with experiment. Analysis of the trained circuits reveals rich, cell type-specific synaptic organization, including like-to-like excitatory wiring rules consistent with emerging connectomic data[36–38], as well as other diverse relationships. It also uncovers a prominent role for inhibitory interneurons: training preferentially sculpts inhibitory synaptic structure into functionally distinct outgoing-weight cohorts, which, despite small population size, exert outsized causal control over network-wide activity and selectivity. Furthermore, targeted constraint ablations demonstrate that biological priors on synaptic weight distributions, although not required to match aggregate activity statistics, are essential for shaping the detailed circuit organization and cell-type-specific computations. All model code, trained network parameters, and the simulation framework are made publicly available, and the underlying experimental datasets are likewise openly accessible, providing a community resource for reproducible,

data-integrated cortical simulation.

## Results

### *Model overview and data integration*

The V1 model integrates (see Methods) multiple categories of data, such as Electron Microscopy (EM) connectomics[21,35], neuron intrinsic electrophysiology[12,13], multipatch synaptic physiology[18], and large-scale in vivo recordings from the Neuropixels Visual Coding dataset[20] (Fig. 1a). Categories like neuron properties, connectivity, and synaptic parameters inform network initialization, producing a fully specified network in the SONATA format[39]. A central goal of the initialization was to preserve the biological heterogeneity revealed by the datasets, including broad distributions of unitary synaptic strengths and total input synapse counts that vary both across cell types and among individual neurons within a cell type. The initialized network serves as the starting point for the deep-learning-style gradient optimization. Neuropixels data provide cell-type-resolved firing-rate and selectivity targets for this optimization.

We constructed 10 instances of recurrent spiking networks comprising ~66,670 ± 240 (STD) neurons drawn from 19 cell types—excitatory populations spanning layers 2/3 through 6 (L2/3-6), including distinct L5 ET, IT, and NP subtypes, and inhibitory populations (PV, SST, and VIP in each of L2/3, 4, 5, and 6, plus an L1 inhibitory population)—represented by 201 point-neuron models fit to experimental data[12,26]. I wonder whether we should include, at least, the full names of the excitatory cell subtypes (ET, IT, NP). These acronyms do not mean much to me and although I can find their meaning with a quick Google search, it would make the text clearer and easier to follow if the names were written out. It's just a suggestion. Neurons were placed within a cylindrical cortical column (400 μm radius) at layer- and type-appropriate densities (Fig. 1b).

To incorporate the experimentally observed cell type-dependent diversity (Fig. 1c) of synaptic transmission properties[18], we analyzed PSP amplitude distributions, well described by log-normal fits, and synaptic current kinetics, captured with a double alpha function that simultaneously fits the rapid rise and slower decay of the synaptic current better than a single alpha (see examples in Fig. 1d,e). Connection probabilities, spatial spread parameters, and per-neuron target in-degrees were derived from EM data[21,35]. Recurrent connections were then assigned using distance-dependent, orientation-dependent[40], and like-to-like connectivity rules, with each neuron's total input scaled to match the broad, log-normal in-degree distributions observed in the EM data (see Methods). The resulting peak connection probabilities and median PSP amplitudes across all pathway types are summarized in Fig. 1c. The total number of connections in the networks is ~24.3 ± 0.2 million.

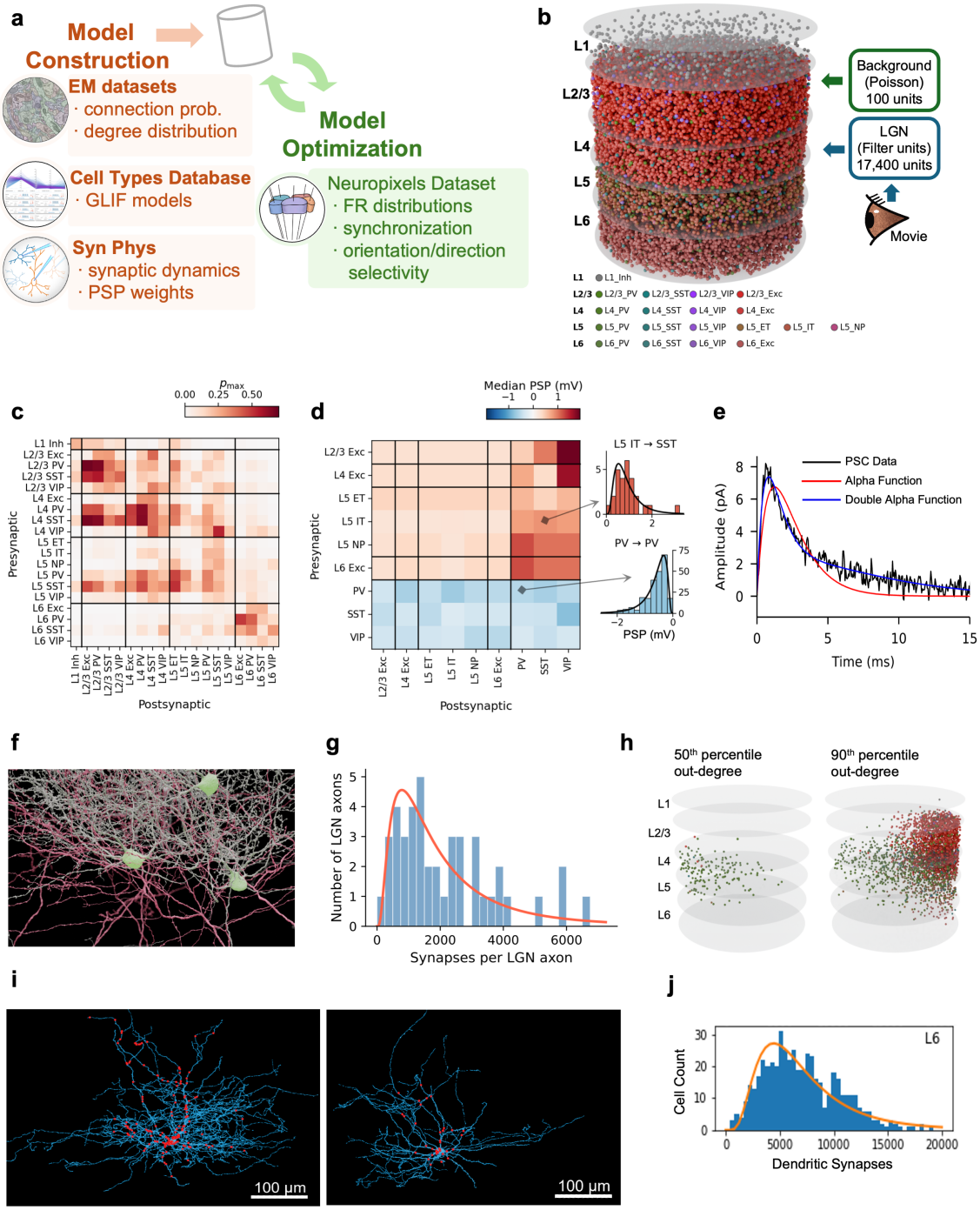



**Figure 1: Model overview and data integration.**

**Figure 1: Model overview and data integration (continued).** Multimodal experimental data are integrated to parameterize a $\sim 67{,}000$-neuron spiking network model of mouse V1. **a,** Schematic of the major data sources integrated into the model. EM connectomics, intrinsic neuronal electrophysiology, and multipatch synaptic physiology inform model construction; large-scale Neuropixels recordings provide targets for model optimization. **b,** Three-dimensional rendering of the model. Individual neurons are colored by cell type; LGN (thalamocortical) and background inputs are indicated. **c,** Connection matrix across cell-type pairs showing the peak connection probability ($p_{\max}$). **d,** Postsynaptic potential (PSP) amplitudes across cell-type pairs derived from the Synaptic Physiology Dataset[18]. Left, median PSP amplitudes; right, example PSP distributions and their maximum-likelihood log-normal fits. Top: L5-IT to SST ($N$ = $24$ connected pairs); bottom: PV$\rightarrow$PV ($N$ = $240$). Amplitudes are 90th-percentile values for each pair. See Extended Data Fig X for full data sources. **e,** An example of the postsynaptic current recorded using voltage clamp from a PV$\rightarrow$L5 IT connection. Black, data; red, single alpha function fit; blue, double alpha function fit, which captures both fast and slow decay components. **f,** EM reconstruction showing three L4 basket (putatively PV) interneurons (green) contacted by two heavily branching LGN axons (red). **g,** Distribution of total synapse counts across 43 reconstructed LGN axons in EM data[21]. Bin width: 250 synapses. The red line is a log-normal fit. **h,** V1 target neurons receiving synapses from two example LGN projections in the constructed model. Left, a projection for the median of the log-normal synapse count distribution; right, a projection at the 90th percentile, illustrating the heavy-tailed connectivity. **i,** EM reconstructions of two L6 basket (putative PV) interneurons (dendrites, blue) with synapses from other L6 basket cells (red). Left, a neuron with extensive dendritic arborization and numerous inputs from other L6 basket cells; right, a neuron with more compact morphology and fewer inputs. **j,** Distribution of total input synapse counts for L6 basket interneurons in EM data. Orange curve, maximum-likelihood log-normal fit.

The MICrONS EM data[21] also contained 43 LGN axons, from which we extracted properties of the thalamocortical projections in V1 (see example in Fig. 1f). Across the reconstructed LGN axons, per-axon synapse counts spanned a broad range (Fig. 1g), and the per-connection synapse count distribution was well fit by a Yule–Simon distribution (Extended Data Fig. 1). In the model, a population of 17,400 LGN units generates spike trains through linear–nonlinear receptive-field filters[41,42]. Each LGN unit contacts a variable number of V1 target neurons, with per-connection synapse counts drawn from the fitted Yule–Simon distribution for each cell type and stronger connections preferentially assigned to spatially closer sources. LGN projections near the median of the distribution contact a moderate number of V1 targets, whereas projections on the heavy tail form denser, more convergent input patterns (Fig. 1h).

To assign synaptic weights, we needed to account for the substantial within-cell type variability in dendritic extent and total synaptic input observed in EM data. For example, L6 PV interneurons (putatively represented by basket cells in the EM data) range from cells with compact dendritic arbors receiving relatively few recurrent PV synapses to cells with more extensive arborizations and correspondingly greater input (Fig. 1i). The distribution of total input synapses per cell follows a log-normal distribution (Fig. 1j), which we used to assign each model neuron a target in-degree. Synaptic weights were then constructed so that, across the population, per-connection strengths recover the experimentally measured log-normal PSP distribution for each connection type[43,44], while accounting for neuron-to-neuron variability in total input count and in neuron model intrinsic properties, following the dendritic constancy rule[45]. The resulting model networks reproduce this heterogeneity: neurons assigned larger in-degrees receive more, individually weaker connections, while those with smaller in-degrees receive fewer, stronger ones (Fig. 1k).

### *Biologically constrained, scalable training of the cortical circuit model*

After initialization, a brain circuit model requires optimization that ideally (i) preserves biological structure, (ii) fits multiple experimentally measured activity statistics simultaneously, and (iii) remains computationally practical (Fig. 2a). Typically, large recurrent spiking network models are either biologically detailed but difficult to optimize[34,46–48] or trainable[49–52] but omit substantial biological realism. To address this, we designed a GPU-native differentiable implementation and a multi-objective optimization framework, briefly described below (see Methods for details).

Our simulator preserves sparse delayed connectivity, detailed synaptic dynamics, and a fixed excitatory/inhibitory sign structure (Dale's law[53]). Training recurrent spiking neural networks (RSNNs) is intrinsically more challenging than in conventional artificial neural networks because spikes arise from thresholding operations with an almost-everywhere zero derivative, which blocks direct gradient flow[51,54–57]. We therefore use BPTT with surrogate gradients (see Methods), which allows gradient propagation through long recurrent trajectories without changing the forward spiking dynamics.

We also adapted the optimizer to the biological constraints of the model. Standard gradient descent methods such as Adam[58] assume a Euclidean parameter geometry, where additive updates define the natural notion of distance[59,60]. However, this assumption is biologically agnostic: additive updates neither preserve the fixed sign structure imposed by excitatory and inhibitory neurons nor respect the heavy-tailed distribution of synaptic strengths observed across cortical circuits unless additional constraints (for example, clipping) are imposed[44,61–64]. We therefore use exponentiated gradient descent (EGD), a mirror-descent formulation with multiplicative updates that naturally preserves weight sign and better aligns with multiplicative changes in synaptic strength[59,60,63,64]. Building on this idea, we introduce an *Exponentiated Adam* optimizer that combines Adam-style adaptive

moments[58] with EGD-style sign-preserving multiplicative updates[63,64] (see Methods). This choice is motivated by the need to optimize efficiently while respecting structural synaptic constraints.

Training the V1 model at this scale required implementation-level acceleration in addition to algorithmic choices. We optimized the simulator for sparse event-driven computation on GPUs with techniques including mixed precision[65], selective JIT compilation for stable tensor-heavy components, gradient checkpointing[66], and parallel execution (see Methods). This enabled the use of BPTT on a large, sparse, heterogeneous cortical RSNNs without reducing the biological detail represented in the model state.

What target should one use for training? While our framework allows for a great variety of loss functions, including neural activity or task performance, or both, in practice, the choice must depend on the quantity and quality of available data and account for the substantial computational expense involved in optimizing these large and complex RSNNs. One approach is to assume a specific function for the trained network, such as motion detection[67]. However, it is unlikely that motion detection or another single computation is the sole function of mouse V1. We therefore chose to match neural responses experimentally recorded in vivo. In the Neuropixels data we use[20], responses to drifting gratings (DGs) are well represented. Though an artificial stimulus, DGs are also associated with well-established metrics like orientation and direction selectivity indices (OSI/DSI), whereas responses to stimuli like natural images or movies (also present in the dataset) may require more sophisticated analyses. For these reasons, and to control computational cost, we used a subset of DG responses as a target training set, an approach justified post-hoc by a good generalization of the trained V1 networks across stimulus classes (see below). An avenue for future investigation—though, requiring vast amounts of computation—is to explore whether other choices may lead to similar or better results.

At each optimization step, visual stimuli were converted to LGN spikes and propagated through the recurrent V1 circuit, and the resulting activity was compared to Neuropixels-derived benchmarks[20] (Fig. 2a; see Methods). Each step used two 500 ms branches with matched batch structure (five gray-screen and five DG trials; 80% contrast, 0.04 cycles/degree, 2 Hz, random orientation and phase). We measured agreement with in vivo data using multiple complementary statistics: both firing-rate and OSI/DSI distributions for each cell type, and excitatory population multi-scale synchrony (via Fano factor). In parallel, we regularized membrane voltages and synaptic-weight distributions for each presynaptic–postsynaptic cell-type pair. The latter is particularly important for maintaining biological realism, as it preserves the distributions of synaptic weights[18] from data while permitting changes in individual synapses to achieve the training targets. For training, synaptic weights were optimized under these constraints, whereas all other model parameters were fixed.

Obtaining stable dynamics in highly recurrent network models, like those of the cortex, has historically been one of the most difficult aspects of model optimization. A common experience in the field is that model networks exhibit either runaway excitation or almost complete suppression of responses, and finding a stable regime of sustainable controlled activity between these extremes via manual optimization of synaptic parameters is extremely challenging. The training trajectories in Fig. 2b show coordinated reduction of the total loss and its component terms across epochs, with no evidence of runaway instability (Extended Data Fig. 2). The firing-rate, OSI/DSI, and synchronization terms decrease throughout training, whereas the voltage and weight regularizers decrease more gradually as weights are refined. This suggests that the regularizers act primarily as constraints on admissible solutions, rather than as the dominant driver of optimization. This is relevant because recurrent spiking models are often susceptible to pathological regimes during gradient-based training[68,69]. Here, the initialization is already stable, and training progressively retunes the circuit toward experimentally matched operating points. Representative raster plots

(Fig. 2c) illustrate this transition: activity remains stable at initialization and shifts during training toward more realistic firing levels and selectivity structure across layers and cell types.

Quantification of computational scaling (Fig. 2d) shows a fixed-overhead regime for small and an approximately linear scaling for larger networks (see Methods). Under the testing configuration (duration $T$ = 4000 ms, batch size $B$ = 10), inference remained faster than real time for networks below 100,000 neurons. A complementary batch-size sweep on an NVIDIA L40S system shows sublinear growth in step time and near-linear growth in VRAM with increasing batch size (Extended Data Fig. 3). In production training on a single A100 GPU, each 1 s training step required $\sim$ 17.6 s wall-clock time, and full optimization took $\sim$ 10 h (see Methods). Thus, end-to-end training of complex biologically realistic RSNNs with > 60,000 neurons and tens of millions of synapses is practical on a single GPU.

To control BPTT memory cost without discarding postsynaptic current (PSC) diversity, we approximated the fitted double-alpha kernels using a shared 4-function alpha basis (Fig. 2e,f). We fit the basis globally across all 121 connection types (determined by combinations of pre- and post-synaptic cell types, see Methods) using shared basis time constants and connection-type-specific coefficients. Approximation error decreased rapidly with basis size (Fig. 2f,e), and with $B$ = 4 all connection-types achieved mean squared error (MSE) < $10^{-3}$ while preserving integrated charge transfer within 1.7%. This compression reduced PSC state storage from 44 states/neuron to 8 states/neuron, corresponding to a 5.5$\times$ reduction in PSC activation memory under BPTT.

Together, these developments facilitate differentiable and scalable simulations of large biorealistic circuits, efficient training constrained by multimodal data, and preservation of stable dynamics from a biologically grounded initialization throughout training.

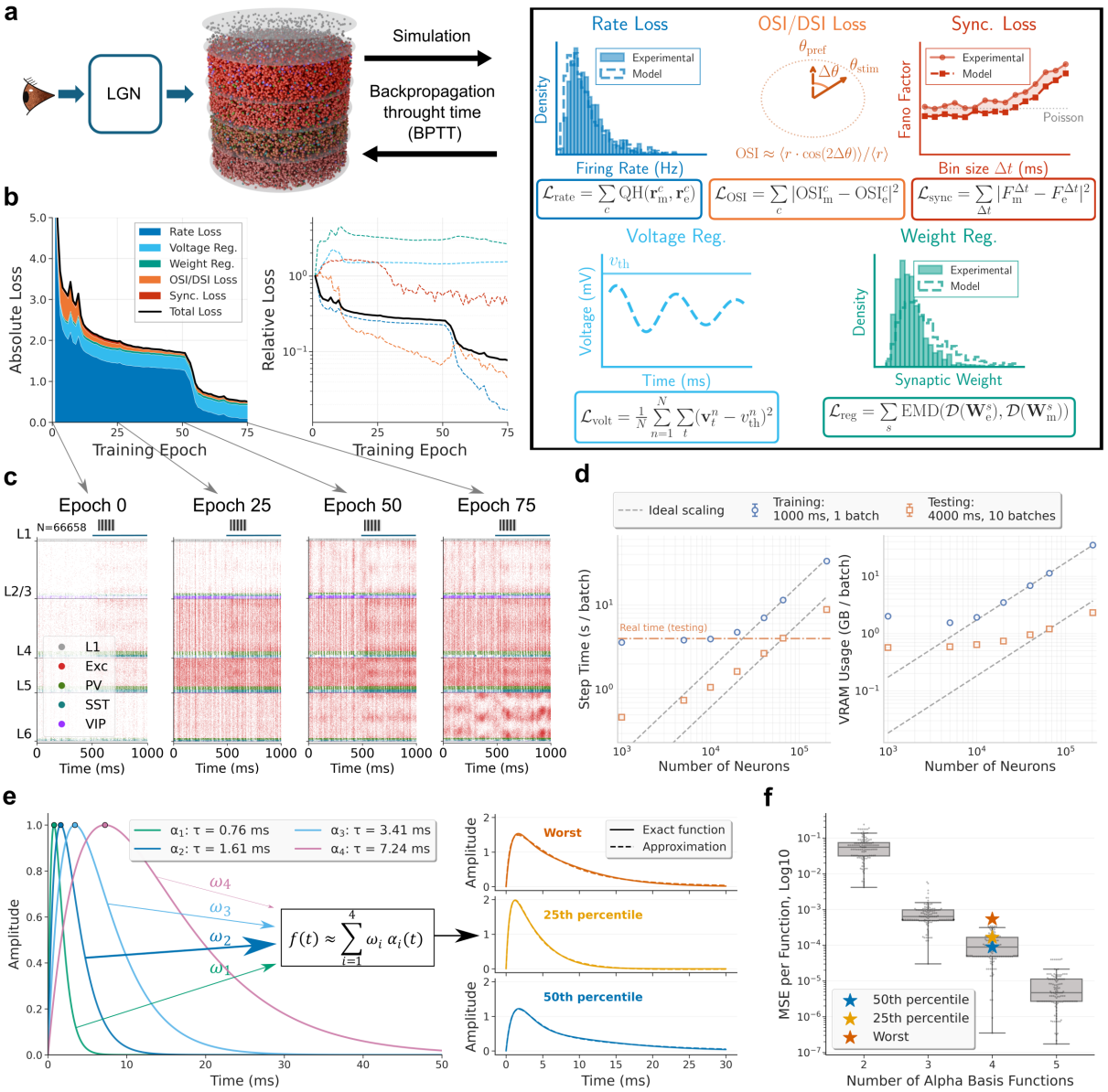



**Figure 2: Biologically constrained, scalable training of the cortical circuit model.**

**Figure 2: Biologically constrained, scalable training of the cortical circuit model (continued). a,** End-to-end training schematic. Visual input is processed by the LGN model and drives the recurrent V1 network. Network activity is compared with experimentally derived constraints through a multi-objective loss, including firing-rate distributions (evoked and spontaneous), orientation/direction selectivity, synchronization/variability statistics, voltage regularization, and synaptic-weight regularization. Gradients are propagated through time (BPTT) using surrogate-gradient approximations. See Methods for details. **b,** Training trajectories. Left: absolute loss components and total loss across epochs. Right: corresponding relative loss scales (log axis), showing coordinated optimization across objectives. **c,** Representative population rasters at initialization and during training (epochs 0, 25, 50, 75), organized by cortical layer and cell type, with neurons of each cell type vertically ordered by preferred tuning angle. **d,** Computational scaling with network size under training and testing settings. The gray dashed lines show an ideal linear-scaling reference, anchored at $N = 40,000$ neurons. Real-time performance for inference mode ("testing"; 4 real-world seconds for a 4 s-long simulation) is highlighted with a dashed orange line. **e,** Multi-alpha postsynaptic current (PSC) kernel approximation. Left: four shared basis functions ($\tau = \{0.76\ ms, 1.61\ ms, 3.41\ ms, 7.24\ ms\}$). Right: reconstructed double-alpha kernels for representative median, 25th-percentile, and worst-case fits. **f,** Basis size versus approximation error. Box plots show MSE distributions across 121 fitted double-alpha kernels. Error decreases sharply with basis size; with four basis functions, all fits satisfy $\mathrm{MSE} < 10^{-3}$. Stars indicate the median, 25th-percentile, and worst-case examples shown in panel e.

***Trained models reproduce physiological responses across cell types***

We first evaluated trained networks using DGs—the same stimulus set as that used for training, but longer (2 s individual DG presentation instead of 0.5 s) and with drift along 8 canonical directions (Fig. 3a, b). This stimulus set was chosen to evaluate stability to longer stimuli (with duration matching that in experimental recordings[20] used here for comparison) and the basic visual response properties typically used in the field.

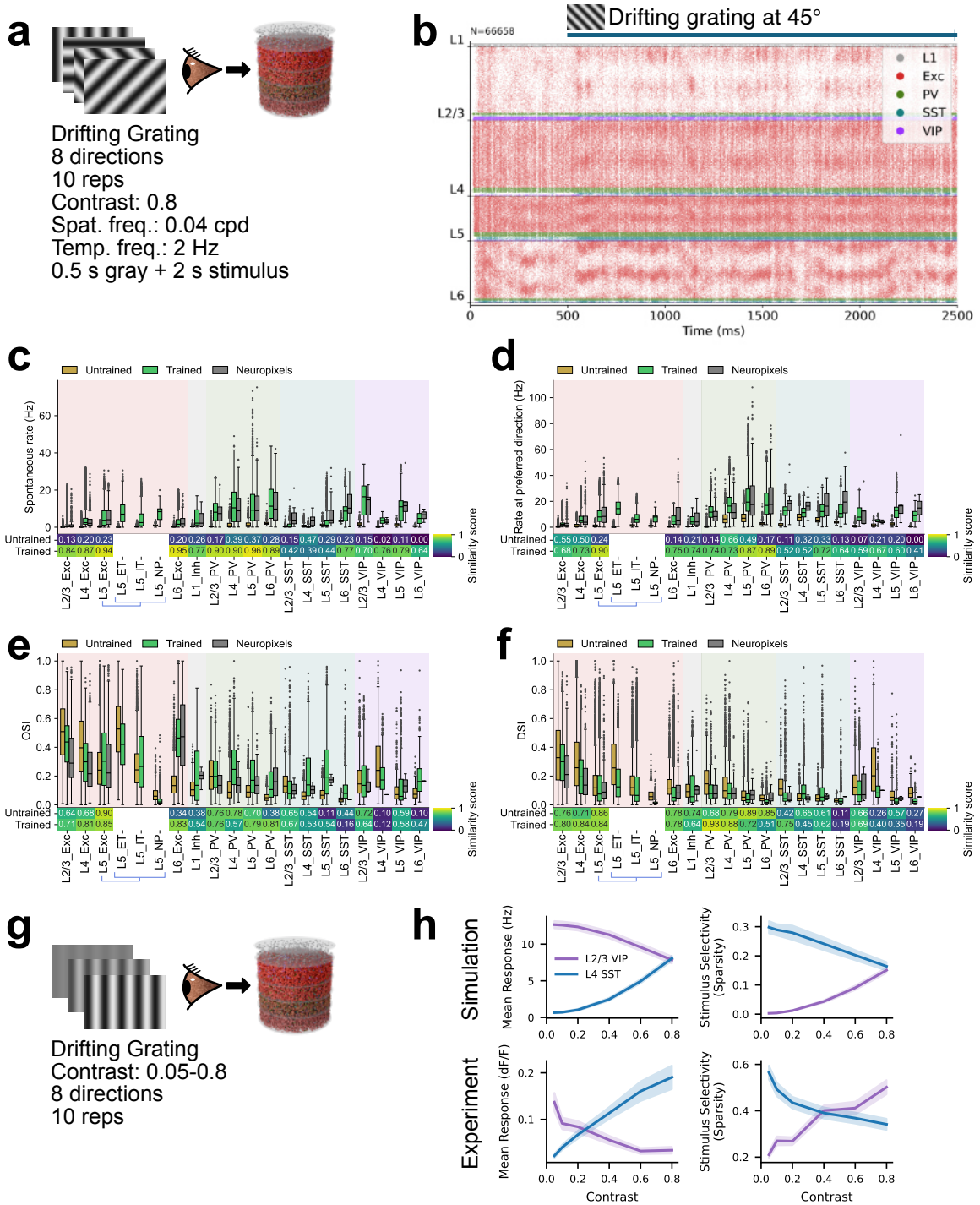



**Figure 3: Trained models reproduce physiological responses across cell types. a,** Schematic of the simulations with the DG stimuli. **b,** Raster plot of the activity of a trained network (sorted by cell type, then by the preferred angle). **c-f,** Quantification of response properties, computed for each cell and grouped by cell type. Trained and untrained models (N=10 individually trained models each) are compared with experimental Neuropixels recordings in vivo. Similarity scores between the models and experiments are shown below the plots. Note that both aggregated and segregated results for L5 excitatory types are shown, but only aggregated results are available in the Neuropixels data. **c,** Spontaneous rates (i.e., gray screen). **d,** Firing rates to preferred direction. **e,** Orientation selectivity index (OSI). **f,** Direction selectivity index (DSI). **g,** Contrast-response simulations. **h,** Mean response and stimulus selectivity as a function of contrast for L2/3 VIP and L4 SST.

As described above, the biological structure priors yield stable network dynamics already at initialization, and stability is maintained throughout training (Fig. 3b; Extended Data Fig. 2). Notably, this robust operating regime generalizes to the longer 2 s stimuli used here for evaluation, compared to the 0.5 s trials used during training.

Fig. 3c,d shows that training greatly improved the match with experimental data for both spontaneous and evoked firing rates across cell types—for both excitatory and inhibitory neurons—with median similarity scores increasing from 0.21 to 0.80 for spontaneous and from 0.21 to 0.70 for evoked firing rates. Orientation selectivity index (OSI) distributions (Fig. 3e) were also better matched after training (median OSI similarity score: 0.56 to 0.64). Direction selectivity index (DSI) (Fig. 3f) was already close to experimental distributions before training (median DSI similarity score: 0.70), but had been achieved at much lower firing rates. Training brought firing rates to physiological levels while preserving selectivity, placing the networks in a more realistic operating regime. This validates the selectivity loss function described above, which used only cell-type-level approximations of OSI and DSI computed from 0.5 s training trials rather than from individual neurons. These results also demonstrate successful generalization to the longer 2 s stimuli used in evaluation.

Furthermore, while training used only high-contrast (0.8) DGs, the trained models generalized well to low contrast DGs. In particular, we observed (Fig. 3g,h) a contrast-dependent dichotomy between the SST and VIP populations in the upper layers, previously demonstrated in experiment[70]: the L2/3 VIP activity decreases and L4 SST activity increases with increasing contrast. Another noteworthy observation was a decrease in selectivity for SST neurons with increasing contrast—observed both in the simulation and the experiment—while all other cell types increased their selectivity (Extended Data Fig. 4). Thus, even with a relatively limited set of response statistics used in training, our biologically realistic network models correctly capture phenomena occurring in the previously unseen stimulus regime, achieving correct stimulus-dependent gain and selectivity modulation.

***Trained models generalize to natural images.***

We next tested whether the models trained on the artificial DG stimuli generalized to the more ethologically relevant and complex natural-image inputs. We first used the natural scenes from the "Brain Observatory" (BO) set (Fig. 4a). While relatively small (118 images), this set is important in that it was used for extensive Neuropixels recordings in mouse V1[20], offering an excellent quantitative benchmark. The BO natural images evoked neural responses that were overall consistent in their magnitude between the trained models and Neuropixels data across cell types, showing an improvement over the untrained models (median similarity score: 0.24 for untrained vs. 0.65 for trained; Fig. 4b). Selectivity of neurons to image presentations was consistent as well (median similarity score: 0.32 for untrained vs. 0.58 for trained; Fig. 4c).

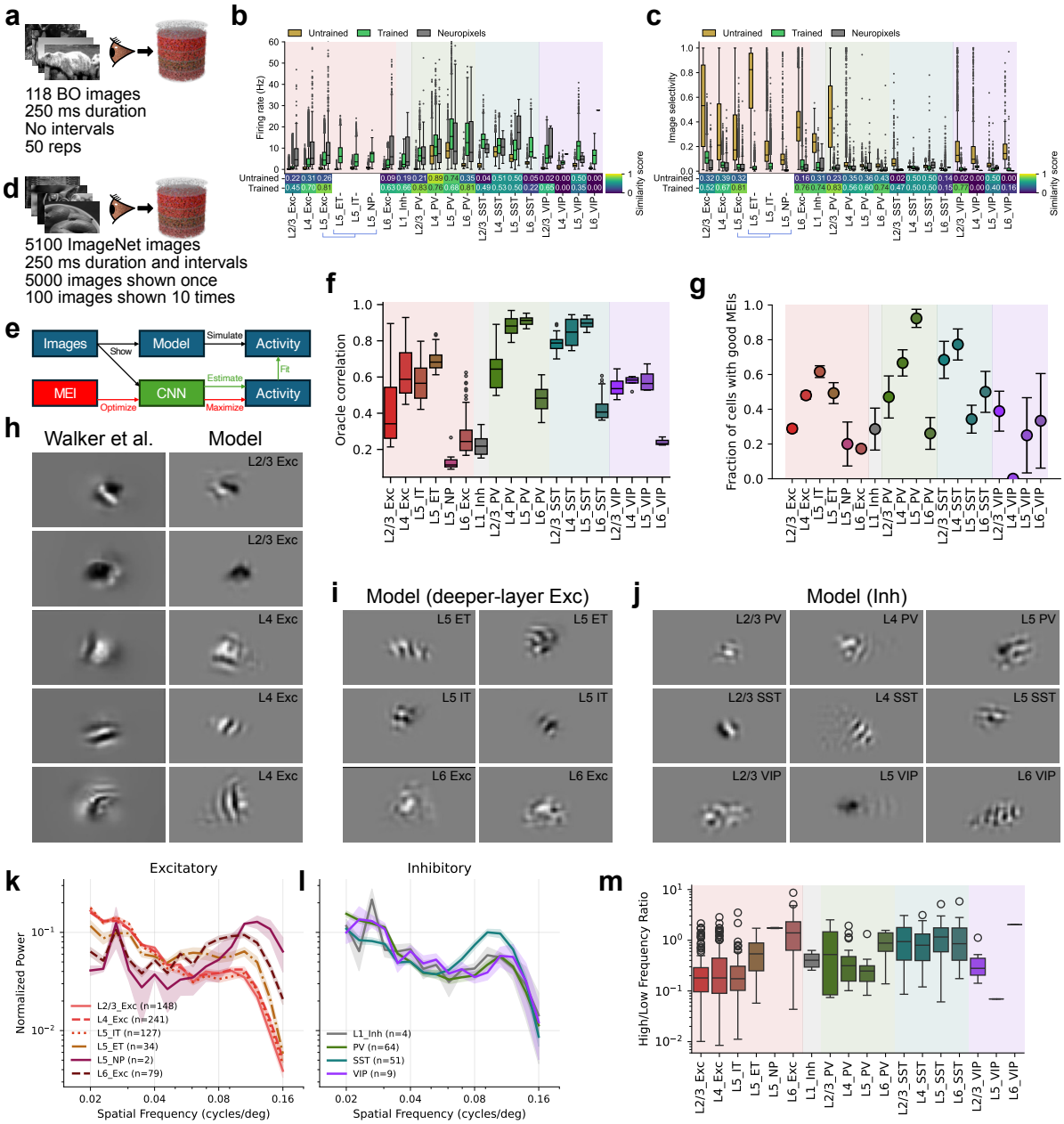



**Figure 4: Trained models generalize to natural images. a,** Brain Observatory (BO) natural scenes stimulus, consisting of 118 images. **b–c,** Firing rates (b) and image selectivity (c) for the BO stimulus, compared between trained/untrained models and experiment. **d,** ImageNet stimulus. Stimuli with 250 ms duration were intervened by 250 ms gray screen. 5000 images were shown once while 100 images were shown 10 times to calculate response reliability. **e,** MEI optimization schematic. **f,** Oracle correlation, a measure of response reliability, for each cell type. **g,** Fraction of cells that produced good MEIs. **h–j,** Example MEIs for upper-layer excitatory neurons (including comparisons with MEIs obtained using experimental data[71] available for these populations), deeper-layer excitatory, and inhibitory neurons. **k–l,** Normalized spatial power spectrum of the MEIs for excitatory (k) and inhibitory (l) cell types. **m,** Distributions of the ratio between the high frequency power and the low frequency power of the MEI spatial power spectrum for each cell type.

To probe the features that drive individual model neurons, we additionally presented images from ImageNet (5000 images presented once and 100 images repeated 10 times, with a random order) (Fig. 4d) and generated most exciting inputs (MEIs)[71] by optimizing stimuli to maximize each neuron's predicted response (Fig. 4e). This approach creates a convolutional neural network (CNN)-based model of each neuron's responses by fitting its activity (here, applied to the activity simulated in our network models), and uses iterative optimization to discover synthetic stimuli that drive neurons to their maximal firing rates, revealing nonlinear feature selectivity that goes beyond traditional linear receptive field analysis. For each neuron in all our network models, we assessed the reliability of the cell's response to the same repeated stimuli using an oracle-correlation metric (Fig. 4f). We then randomly chose a total of 2000 neurons among neurons with the highest 30% oracle-correlation within each cell type, across all 10 trained models. We quantified the fraction of neurons for which the maximization of the CNN-response produced a stable, interpretable MEI (Fig. 4g). A substantial subset of neurons yielded high-quality MEIs, enabling qualitative and quantitative comparisons across cell types.

The MEI analysis provides a compact characterization of what the trained circuit is responsive to at the single-neuron level, in a way that goes beyond selectivity indices. It also offers a diagnostic for whether a neuron's response function is smooth and optimizable under the natural image statistics: neurons with low oracle correlation can be interpreted as either having weak stimulus drive under the MEI-optimization regime or exhibiting response landscapes that are difficult to optimize reliably. Importantly, the distribution of oracle correlations and fraction of cells with good MEIs differed across populations (Fig. 4f–g), implying that the ability to obtain stable MEIs is itself cell-type dependent and reflects underlying differences in cell models and circuit organization.

In particular, L5 NP cells had the lowest oracle score, which indicates that they do not exhibit reliable image-dependent activity patterns. Consequently, we found fewer good spatially confined MEIs for L5 NP cells. Interestingly, this is different from other L5 excitatory populations, L5 ET and IT (Fig. 4f–g), even though our training targets used the same DG activity metrics for all L5 excitatory cells because the Neuropixels data do not distinguish L5 ET, IT, and NP. This suggests that response differences between these L5 cell types are due to the distinct ways in which they are integrated in the V1 circuit, which was captured by our model building procedure—in particular, reflecting cell-type-specific wiring in EM data—and translated in different activity patterns in the trained models. To test this prediction of our models, we computed oracle scores of neural responses in the MICrONS dataset and found that indeed L5 NP cell show lower scores than L5 ET and IT (Extended Data Fig. 5).

We found that MEIs for upper-layer excitatory neurons exhibited localized structure that often included smooth stripe patterns, with many examples closely resembling MEIs previously reported for experimentally recorded neurons from the same populations in mouse V1 (Fig. 4h)[71]. In deeper excitatory populations, MEIs from our network models showed more complex spatial patterns (Fig. 4i) along with the stripes; inhibitory neurons exhibited similarly rich structure across all layers (Fig. 4j). To quantify these differences, we computed the spatial power spectra of the MEIs for excitatory (Fig. 4k) and inhibitory (Fig. 4l) classes and summarized each neuron's spectral profile by a high (0.08–0.16 cycles per degree (cpd)) vs low (0.02–0.04 cpd) spatial-frequency ratio (Fig. 4m). This analysis revealed a systematic shift toward higher-frequency content in deeper-layer and inhibitory neurons, consistent with increased spatial complexity in the MEIs for these populations.

Thus, the statistics of trained models' responses to natural images matched experimental data and their MEI-derived response characteristics were consistent with experiment where data were available (i.e. upper-layer excitatory neurons). Beyond these validated areas, the MEIs reveal layer- and cell-type-dependent differences in preferred features, which serve as predictions for future

experimental testing.

***Training yields tuning-dependent synaptic structure and reveals inhibitory cohorts with distinct causal roles.***

Having established that the trained models exhibit realistic response statistics across stimulus families, we next asked what synaptic patterns and circuit-level mechanisms emerge in these models. We characterized synaptic weight structure using two complementary functional axes: similarity of stimulus-evoked responses and similarity of the preferred direction of motion in a connected pair of neurons. Note that presence or absence of connections is not modified in our training method (see Methods), meaning that the effects described below are in addition to the existing connectivity structure.

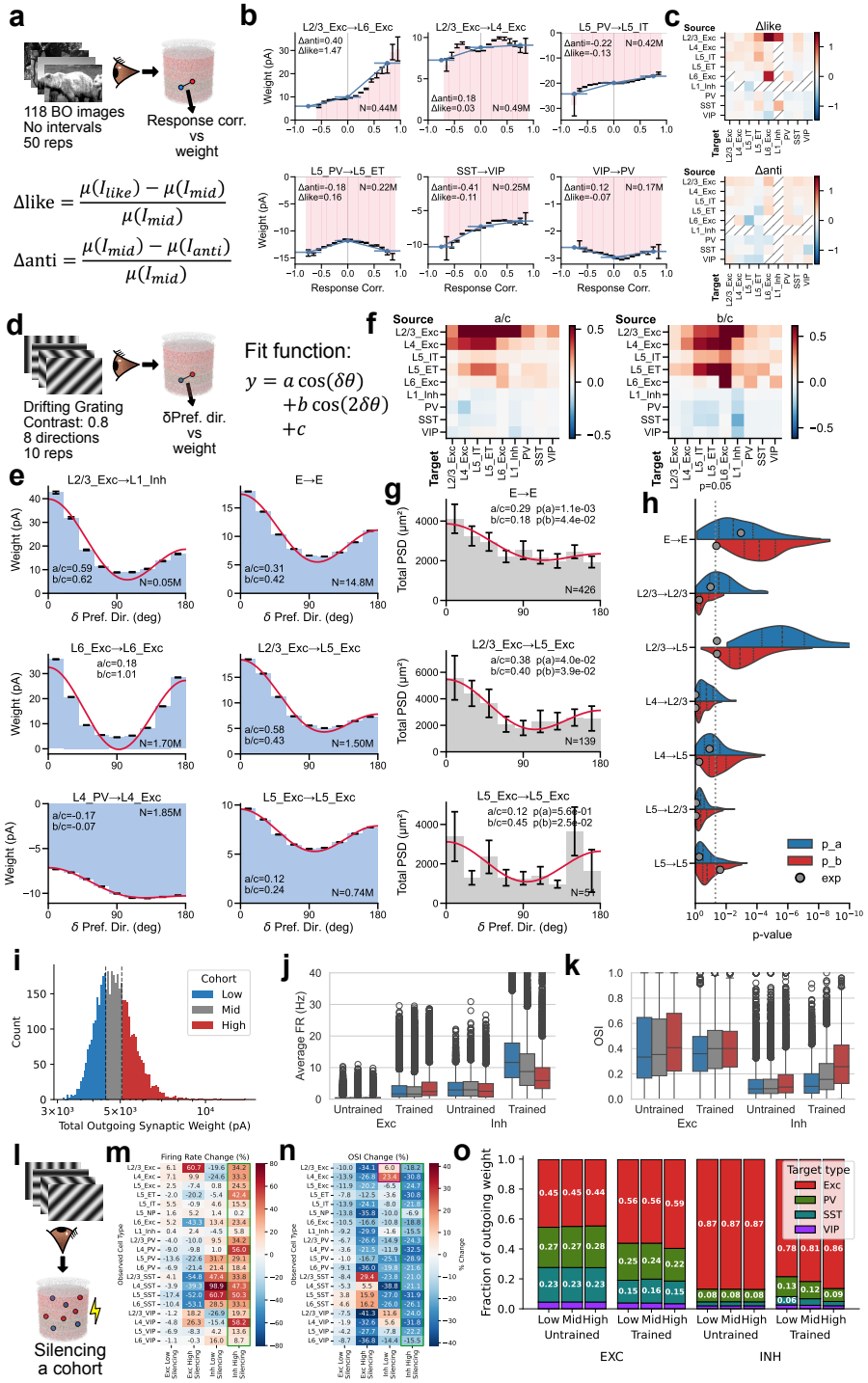



**Figure 5: Training yields tuning-dependent synaptic structure and reveals inhibitory cohorts with distinct causal roles.**

**Figure 5: Training yields tuning-dependent synaptic structure and reveals inhibitory cohorts with distinct causal roles (continued). a,** Schematics of the stimulus and analysis for weight changes as a function of response correlation. **b,** Examples of diverse relations between synaptic weight and response correlation of connected neuron pairs in the trained network models. Number (N) of pairs analyzed is indicated. **c,** Heatmaps of $\Delta_{\mathrm{like}}$ and $\Delta_{\mathrm{anti}}$ for select cell types. Matrix elements involving PV, SST, and VIP types exclude inter-layer connections. (See full heatmaps including intra- and inter-layer connections in Extended Data Fig. 6.) For both heatmaps, positive values correspond to like-to-like and negative to anti-like-to-like relations. When neuron pairs with sufficient correlation ($< -0.5$ or $> 0.5$) were not found, these metrics were not calculated and a diagonal line is placed in the heatmap element. **d,** Schematics of the stimulus and analysis for weight changes as a function of preferred-direction difference. **e,** Examples of relations between synaptic weight and preferred-direction difference. **f,** Heatmaps of cosine-fit effect sizes (a/c: direction like-to-like component, b/c: orientation like-to-like component) for select cell types. Matrix elements involving PV, SST, and VIP types exclude inter-layer connections. (See full heatmaps including intra- and inter-layer connections in Extended Data Fig. 7 and 8.) **g,** Electron microscopy experimental data (MICrONS) showing the relation between preferred-direction difference and post-synaptic density (PSD). **h,** Statistical comparison between MICrONS data (gray) and Monte Carlo sampling from the trained models when the same number of connections as available in MICrONS data were chosen (red and blue). **i,** Example synaptic weight distribution (L2/3 Exc to L2/3 Exc), split into low, mid and high outgoing-weight cohorts. **j,** Firing rates (pre vs. post training) for each cohort. **k,** OSI (pre vs. post training) for each cohort. **l,** Schematic for silencing simulations. **m,** Change in firing rates during silencing. The green rectangle highlights cell-type agnostic broad increase in firing rate when high-outgoing-weight inhibitory cohorts are silenced. **n,** Change in OSI during silencing. The green rectangle highlights a global suppression of the selectivity when high-outgoing-weight inhibitory neurons are silenced. The purple rectangle highlights paradoxical disinhibition of the L2/3 and L2 excitatory populations when low-output-weight inhibitory neurons are silenced. **o,** Fraction of outgoing synaptic weights to different cell types for each cohort.

**Relation of synaptic weights and response correlations.**
The logic of like-to-like connectivity has been a central theme in cortical circuit studies, with multiple reports of such structure especially for L2/3 excitatory neurons[36–38,72–76]. However, findings are more nuanced across layers; for instance, while L2/3 pyramidal neurons receive functionally matched inputs, L5 pyramidal neurons have been reported to receive inputs with a broader range of functional properties[72]. The organization principles for inhibitory connections also remain a subject of active debate[73,77,78]. Our models offer an opportunity to study these functional wiring rules for all cell types at scale.

We computed response correlations (see Methods) under the BO natural scenes (Fig. 5a) and observed a substantial diversity of weight-vs.-correlation relations across source–target type pairs (Fig. 5b,c). While prior studies focused primarily on connection probability rather than weights, we found that synaptic strengths in our models also exhibit clear functional logic. Generally, excitatory connections are dominated by like-to-like relations, where pairs with similar stimulus responses have stronger weights (e.g., L2/3 Exc to L6 Exc; Fig. 5b). Inhibitory connections are notably more diverse: some pathways follow like-to-like rules, while others exhibit anti-like-to-like relations (stronger weights for negatively correlated pairs, such as L5 PV to L5 IT and SST to VIP) or more complex non-monotonic patterns (Fig. 5b; see full matrix in Extended Data Fig. 6).

To summarize these effects, we computed a metric ($\Delta$like and $\Delta$anti) that captures how strongly synaptic weight increases on the positively and negatively correlated sides, respectively (Fig. 5a,c; see Methods). This analysis reveals a general trend for like-to-like connectivity: excitatory neurons tend to form like-to-like connections regardless of the recipient cell type, and this pattern holds for both positively and negatively correlated pairs. This tendency is more pronounced in projections from superficial to deeper layers. In contrast, inhibitory connections are more diverse, exhibiting both like-to-like and anti-like-to-like motifs, and frequently switching signs between positive and negative correlation regimes. These results demonstrate that the mapping from correlation to inhibitory synaptic strength is complex, non-monotonic, and pathway-dependent, reflecting the intricate interplay between interneuron types. Furthermore, these findings highlight how single averaged summaries (for example, an average like-to-like effect across all inhibitory connections) can obscure critical layer- and type-specific structure.

**Relation of synaptic weights and preferred-direction differences.**
In the visual cortex, many cells show preference for the direction of motion, suggesting such direction tuning as a potential major axis for circuit organization. Previous experimental studies established that, whereas similarity in preferred direction between two cells in V1 is a good predictor of connection probability[36–38,72,75], it does not appear to predict well the connection strength—at least, for L2/3 E-to-E connections that have been studied the most[37]. To investigate this phenomenon, we measured the dependence of synaptic weight on the difference in preferred direction between pre- and post-synaptic neurons in our trained models and quantified these effects using a cosine-series fit (Fig. 5d–f, see full matrices in Extended Data Fig. 7). In this fit, the coefficients $a$ and $b$ represent the strength of direction and orientation like-to-like rules, respectively, normalized by the offset $c$ (where negative values indicate anti-like-to-like relations). This analysis revealed widespread like-to-like organization of synaptic weights relative to preferred direction similarity for excitatory and anti-like-to-like for inhibitory connections. Fig. 5e shows prime examples of both cases (like-to-like: L6 Exc $\rightarrow$ L6 Exc; anti-like-to-like: L4 PV $\rightarrow$ L4 Exc). Excitatory connections exhibited a laminar organization, with generally stronger like-to-like tendencies from superficial to deeper layers.

To directly compare model predictions to anatomy, we used the V1DD EM data[35] linking functional tuning differences to post-synaptic density (PSD), used as a putative synaptic strength. Because the MICrONS dataset provides a relatively small and fixed set of reconstructed and proofread

connections (with physiology available only for excitatory neurons), obtaining conclusive observations is challenging. Nevertheless, we found pathways with like-to-like relations between synaptic weights and preferred direction similarity reaching statistical significance (for all excitatory cells in aggregate, L2/3 Exc → L5 Exc, and L5 Exc → L5 Exc connections, Fig. 5g), as predicted by our modeling results above. We then matched sample sizes between models and experiment using a Monte Carlo resampling of our models (Fig. 5h): for each pathway, we repeatedly redrew the same number of connections as available in the MICrONS data and recomputed the tuning-dependent statistic and its associated $p$-values. Repeating this procedure for 100 redraws yields a distribution of model $p$-values under matched sample sizes (the red and blue distributions in Fig. 5h), which we compare against the single $p$-value obtained from the experimental measurements. This resampling-based comparison illustrates that the MICrONS dataset allows sampling of only a limited number of pathways (those shown in Fig. 5h). For these, our models suggest that while most connection types should be like-to-like, the effect may not reach significance if sampling is limited. L2/3 E-to-E is a notable example where connections are generally weak and a like-to-like relation is difficult to discern even in models; with low sampling, previous studies concluded there is no like-to-like relation for weights vs. direction similarity[37]. In conclusion, the models show specific structure not only for correlation but also for direction similarity, which is generally like-to-like for excitatory and anti-like-to-like for inhibitory connections. Experimental data support this where sampling is sufficient, and future experiments should resolve this further.

### Synaptic weight heterogeneity and its functional correlates.

One of the key features of the organization of brain circuits is the heavy-tailed distribution of synaptic weights[44,61], which appears to be essential for nonrandom microcircuit topology emphasizing hub- or driver-like neurons—small subsets of neurons that can be disproportionately connected or strongly coupled, with outsized influence on network dynamics[79–81]. We set out to investigate this property in our trained models. We observed approximately log-normal distributions of outgoing synaptic weights in the trained circuits (Fig. 5i). We then asked whether neurons with large outgoing synaptic weights form a cohort with distinct functional roles in the circuit relative to other neurons. To address this, we split neurons within each cell type into low-, mid- and high-outgoing-weight cohorts (1/3 each) and compared their firing rates, selectivity, downstream targeting, and causal impact under cohort-specific silencing (Fig. 5j–n,o).

We did not observe a strong cohort-specific differences of the firing rates or OSI in the excitatory populations, suggesting that the functional roles of the excitatory cells are not segregated by the amount of their output even after training. On the other hand, the inhibitory populations showed prominent cohort-specific differences after training. In the trained networks, firing rates differed systematically across inhibitory cohorts (Fig. 5j), and the same cohorts showed structured differences in OSI (Fig. 5k), consistent with the idea that subsets of inhibitory neurons can acquire distinct functional profiles. These cohort differences were attenuated or absent in the untrained network, suggesting that high-outgoing-weight inhibitory phenotypes emerge as part of the learned circuit solution rather than being a trivial consequence of initialization. Such training-dependent changes in functional properties were not observed for cohorts separated by the incoming connections (Extended Data Fig. 9). Therefore, we focused on characterizing the relation between functional properties and outgoing weights.

These cohort effects provide a mechanistic interpretation for how training can satisfy global distributional constraints while retaining strong functional diversity. In particular, the high-outgoing-weight cohort behaves as a small set of strong units: they can act as broadly targeting suppressors, whereas low-outgoing-weight neurons can act more selectively or indirectly through their targeting patterns.

Cohort-specific silencing reveals network-wide causal impacts.
We performed cohort-specific silencing simulations (Fig. 5l, see Methods) and quantified the resulting changes in firing rate and OSI across the remainder of the network. Silencing inhibitory cohorts produced strong network-wide effects (Fig. 5m–n) despite the cohorts' small size relative to the whole network size ( 1% of all neurons), which is dominated by excitatory neurons. In particular, silencing high-outgoing-weight inhibitory neurons led to a widespread increase of the firing rates and suppression of OSI across populations (green rectangles), whereas silencing low-outgoing-weight inhibitory neurons produced paradoxical suppression[82,83] of firing in L2/3 and L4 excitatory populations (purple rectangle). These results suggest that inhibitory heterogeneity, when organized along outgoing-weight axes, can create distinct causal regimes of circuit control.

We also found that inhibitory cohorts differed in which populations they targeted (Fig. 5o). The high-outgoing-weight cohort preferentially allocated a larger fraction of their synaptic weight to excitatory targets and low-outgoing-weight cohort—to inhibitory ones, consistent with a more disinhibitory role. This difference provides a concrete circuit mechanism by which high-outgoing-weight inhibitory cohorts can exert disproportionate control over recurrent excitation and OSI, and may also help explain the paradoxical suppression observed when silencing low-outgoing-weight inhibitory cohorts: silencing a disinhibitory cohort can release other inhibitory populations from inhibition, increasing net inhibition onto excitatory neurons.

These cohort-specific training effects on firing rates and OSI were consistent across inhibitory subtypes (Extended Data Fig. 10). Silencing cohorts within individual subtypes produced partial versions of the network-wide effects. PV cohorts had the strongest impact, qualitatively reproducing the effects of silencing all inhibitory cohorts for both high- and low-outgoing-weight groups. Silencing high-weight SST cohorts also caused global rate increases and OSI suppression; however, silencing low-weight SST cohorts induced paradoxical suppression in L4, but not L2/3, reflecting the lower abundance of SST neurons in L2/3. Finally, silencing VIP cohorts produced much weaker effects than the other two subtypes. Taken together, these results show that cohort identity—defined purely by outgoing synaptic weight rather than by cell-type labels—captures a continuous axis of inhibitory heterogeneity that is in principle orthogonal to transcriptomic or morphological types. Consistent with this, a recent study found that inhibitory, but not excitatory, neurons in the piriform cortex exhibited experience-dependent correlation between stimulus selectivity and the degree of connection of a neuron[74]. Future similar experiments in V1 can test our model predictions.

***Effect of biological weight constraints.***
The preceding analyses show that training arrives at a fine-tuned synaptic structure of the circuit, with a specific organization conserved across our N=10 trained models, driven by the biological constraints we use. We therefore asked how strongly the emerging circuit organization depends on such biological constraints. Among numerous types of constraints used in this work, we focused on characterizing the effects of synaptic weight distributions which were constrained during training to data recorded experimentally in multi-patch synaptic physiology measurements[18]. On the one hand, such tests are not prohibitively difficult to implement: one trains models with the part of the loss function constraining weight distributions removed. On the other, since it might be easier to find solutions without rather than with weights constrained, this test may be less prone to exhibiting trivial large differences than some other, more drastic perturbations, which may result in lower-quality solutions. Furthermore, the experimentally determined weight distributions represent a case of highly detailed data that constrain the models not just as an overall conceptual framework but specifically and quantitatively, suggesting their potential importance in determining mechanisms of the circuit function.

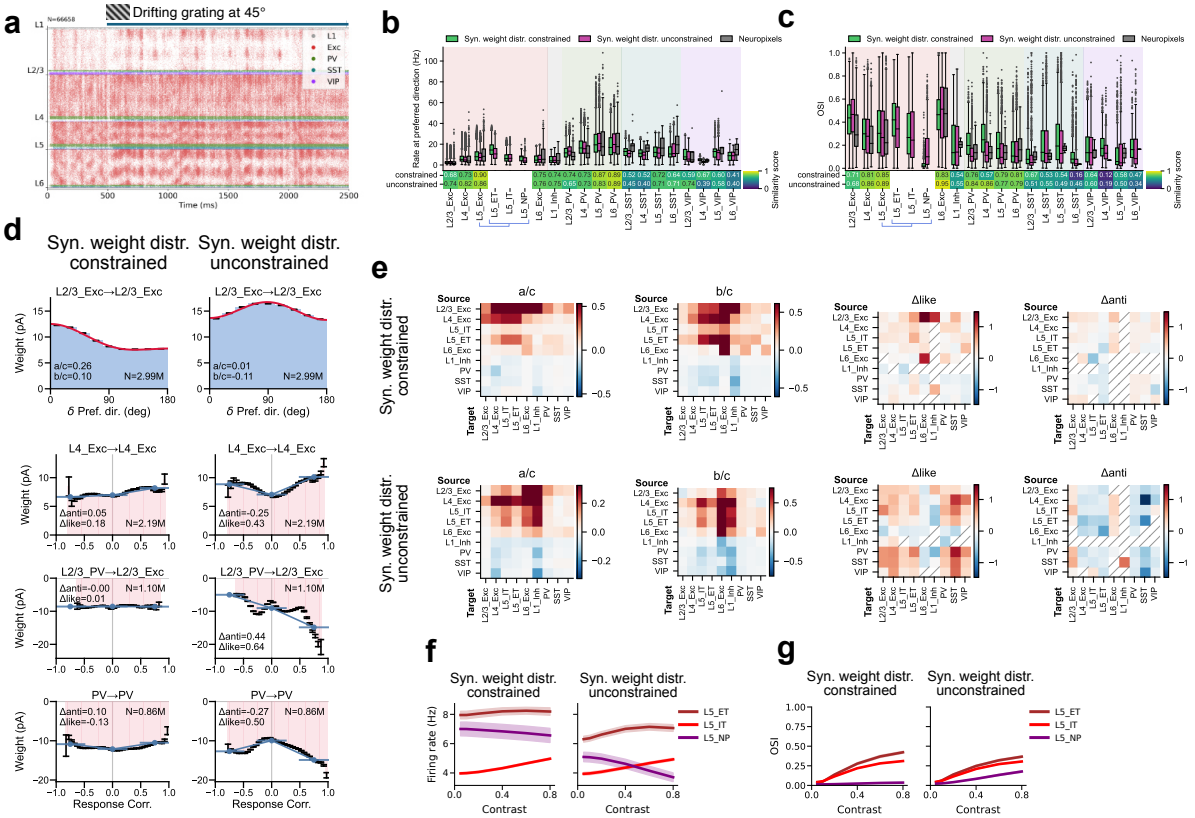



**Figure 6: Effect of biological weight constraints. a,** Raster plot of the network trained with synaptic weight distributions free to change. **b, c,** Firing rates to preferred direction and OSI for networks trained with and without synaptic weight distributions constrained to the experimental data. **d,** Examples of synaptic weight vs. preferred-direction difference or response correlation for networks trained with and without synaptic weight distribution constraints. **e,** Heatmaps summarizing effect sizes and correlation-derived $\Delta$ metrics across cell-type pairs for networks trained with and without synaptic weight distribution constraints. **f, g** Firing rate and OSI of the three L5 excitatory cell types as a function of DG contrast for networks trained with and without synaptic weight distribution constraints.

The N=10 models trained without weight constraints produced activity patterns (see example in Fig. 6a) that were highly consistent with our previous models trained with such constraints. The unconstrained models remained stable throughout training and reproduced the firing-rate and selectivity patterns from the Neuropixels data (Fig. 6b–c; more metrics in Extended Data Fig. 11), with similarity scores consistent with those seen for the constrained models (median peak firing rate similarity score: 0.73 for unconstrained vs. 0.70 for constrained; OSI: 0.68 for unconstrained vs. 0.64 for constrained). This indicates that the functional objectives without weight constraints (though with all other biological details used in constructing the model) can guide the network to a satisfactory operating regime. At the same time, the fact that the unconstrained model converged to the same target metrics is itself informative: it implies that the optimization landscape contains multiple different synaptic solutions that satisfy the same high-level activity constraints. This multiplicity motivates explicitly comparing constrained and unconstrained solutions, because differences between them can identify which aspects of synaptic organization are determined by functional objectives alone and which require additional biological priors.

Remarkably, despite activity patterns were well captured, removing weight-distribution constraints altered the detailed structure of the emergent synaptic rules—even though these constraints are agnostic of any functional logic of the circuit, such as like-to-like rules. Yet, we find multiple cases where such rules were dramatically affected (Fig. 6d,e). For the dependence of synaptic weight on preferred-direction difference (Fig. 6d), the L2/3 Exc→L2/3 Exc pathway shifted from a direction-like-to-like pattern to an orientation anti-like-to-like pattern. While most other cell-type pairs did not switch the sign of the effect between the constrained and unconstrained training, relative magnitudes of the effect were substantially altered, such that, e.g., the strongest like-to-like effect in *a/c* switched from the connections originating from L2/3 Exc in the constrained case to those originating in L4 Exc in the unconstrained case. Like-to-like patterns for response correlation differed more substantially between constrained and unconstrained training runs, with many cell-type pairs exhibiting more exaggerated and/or complex structures (Fig. 6d). The general trend of like-to-like excitatory and anti-like-to-like inhibitory connections were no longer the case in the unconstrained models. Rather, most of the positive correlation side (Δlike) became like-to-like and most of the negative correlation side (Δanti) became anti-like-to-like, making both strongly correlated and anti-correlated neuron pairs have strong synaptic weights. These differences are summarized in the effect-size heatmaps across cell-type pairs (Fig. 6e).

Finally, we observed some deviations in activity patterns when considering specific cell types. In particular, models trained with weight distribution constraints uncovered marked distinctions between the three L5 excitatory cell types in terms of their contrast responses. All three types had different levels of firing rates, with L5 ET and IT exhibiting increases as a function of the stimulus contrast, while L5 NP exhibited the opposite. In addition, the three types of neurons had very different levels of orientation selectivity with ET the highest and NP the lowest. Because the Neuropixels data used for training did not distinguish these three types of L5 Exc cells, these functional differences must have emerged from the circuit architecture differences between these types obtained from the EM data. The relative firing rates of these cell types are largely preserved regardless of synaptic weight constraints (Fig. 6f); however, the orientation selectivity, which differed significantly in the constrained network, collapsed to a similar level for all three L5 excitatory types in the unconstrained models (Fig. 6g).

These results emphasize that biological constraints can be valuable even when they are not strictly necessary to fit functional response statistics. By limiting the space of allowable parameter distributions, constraints can guide training toward solutions that better approximate real brain circuit mechanisms. This suggests a practical strategy for future model development: use functional

objectives to ensure that neural activity is realistic, incorporate the available biological constraints to facilitate realistic circuit mechanisms, and use targeted constraint ablations to identify which constraints support specific wiring motifs and cell type-level organization.

## Discussion

This work provides advances in several areas crucial for facilitating future biologically realistic simulations of the brain. First, we developed models of mouse V1 that integrate multi-modal data of unprecedented breadth and depth constraining the model network composition, connectivity, and activity. These include population sizes and electrophysiological properties of 19 cell types spanning all cortical layers, detailed cell-type-resolved characterization of connectivity, synaptic weights, and synaptic kinetics from both EM connectomics and multi-patch synaptic physiology measurements[18,21], and a representation of visual thalamocortical inputs firmly rooted in functional and connectomic data. Extensive Neuropixels recordings in vivo[20] were used for training and testing of the neural activity generated by the models. Second, we developed fast and scalable implementations of gradient descent training for these highly recurrent, heterogeneous, large-scale models of brain circuits, resutling in end-to-end training of the $\sim 67,000$-neuron V1 model within half a day on a single mainstream GPU. In inference mode, simulations run faster than real time. Third, we designed and tested loss functions and regularizations that support efficient training with limited available neurophysiology data and synaptic weight constraints. This enables training on a small subset of 500 ms-long presentations of gray screen and DGs, with the resulting model nevertheless generalizing well to a variety of diverse longer stimuli.

It is worth noting that, while full-brain connectomic reconstructions have been achieved in other model species—C. Elegans[22,23] and Drosophila[24,25]—the electrophysiological properties of their cell types and synaptic connections are currently less well resolved than those in the mouse cortex. Furthermore, in vivo recordings in those species mostly rely on slower optical imaging methods as opposed to high temporal resolution Neuropixels electrophysiology available for the mouse[84]. The broad range of data available for mouse V1, integrated in our models, constitute some of the most complete cell type, connectomic, and physiological characterization of brain circuits available to date, supporting substantial level of biological realism in simulations[6].

These technical capabilities translate into a set of scientific findings about cortical circuit organization. The generalization noted above is remarkably broad: from the narrow training set, the models reproduce cell-type-specific firing rates, selectivity, and gain modulation across 2 s-long DGs at multiple contrasts, 118 Brain Observatory natural scenes, and thousands of ImageNet images. Non-trivial phenomena that were not a part of the training objective, such as the contrast-dependent dichotomy between SST and VIP populations[70], emerge naturally. This suggests that biological architecture, combined with cell-type-resolved activity constraints, captures general principles of cortical processing rather than fitting stimulus-specific patterns—complementing earlier large-scale models that relied on forward simulation for evaluation[26,46,85].

Analysis of the trained synaptic weights reveals functional wiring rules that extend well beyond a uniform like-to-like motif. Excitatory connections generally follow like-to-like organization[36–38], but inhibitory pathways exhibit a far more diverse repertoire—anti-like-to-like and non-monotonic relationships that vary by cell type and layer[73,78]. These predictions span the full set of cell types we use, including deep-layer populations (e.g., L6) not yet probed experimentally. Where comparison with connectomic data is possible, the predictions are confirmed: we find significant like-to-like structure in synaptic weight versus preferred-direction similarity for pathways where previous studies with limited sample sizes concluded no such relationship exists[21,37]. The emergence of functional connectomics[21,38] now provides a route that may soon help evaluate such model-derived predictions

directly at synapse resolution across the full set of cell types.

Perhaps the most unexpected finding concerns inhibitory interneurons. Training preferentially reshapes inhibitory—but not excitatory—synaptic structure, giving rise to functionally distinct cohorts defined by outgoing synaptic weight that differ in firing rate, selectivity, and downstream targeting. These differences emerge through training rather than from initialization, indicating that they reflect functional requirements of the circuit. Cohort-specific silencing shows that these small subpopulations—each roughly 1% of the network—exert outsized causal control: silencing high-outgoing-weight inhibitory neurons produces widespread disinhibition and selectivity loss, whereas silencing low-outgoing-weight neurons triggers paradoxical suppression[82,83,86] via disinhibition. These results point to an axis of inhibitory heterogeneity, defined by connectivity strength rather than transcriptomic identity[87,88], that may constitute a fundamental organizing principle for cortical computation.

A persistent challenge in mechanistic circuit modeling is identifiability: multiple parameterizations can satisfy the same activity constraints, limiting interpretability[7–9,89]. Constraint ablations offer a practical handle on this degeneracy. We find that removing biological priors on synaptic weight distributions does not prevent the network from matching aggregate firing-rate and selectivity targets, yet it alters the emergent synaptic rules and collapses cell-type-specific differences—most strikingly, the three L5 excitatory types (IT, ET, and NP) converge to similar selectivity profiles. Thus, constraints from synaptic physiology act not merely as regularizers but as selectors among functionally equivalent solutions, guiding the circuit toward organizations that preserve cell-type-level computational specialization.

More broadly, this work aligns with a growing effort to construct integrative circuit models that are both biologically grounded and quantitatively testable[4,26,48,67,90,91]. As multimodal datasets expand—with richer perturbation protocols, improved interneuron coverage, and synapse-resolution connectomics[21]—differentiable circuit models[32,34] should enable ever tighter linking of biological priors, learned parameters, and experimentally falsifiable predictions. Moreover, while large networks can now be simulated at the brain scale [92, 93], mechanistic insight requires models whose synapses can be optimized against functional data. Many emergent circuit phenomena only arise in sufficiently large and heterogeneous networks, making scalable training essential for linking anatomy, dynamics, and computation within a unified framework.

All models, code, and trained parameters from this work are freely shared to support such efforts.

## Data and code availability

The code and data used in this study are available at https://github.com/AllenInstitute/biorealistic-v1-model. Please read the instructions in the repository for more details.

## References


1. D'Angelo, E., Solinas, S., Garrido, J., Casellato, C., Pedrocchi, A., Mapelli, J., Gandolfi, D., & Prestori, F. "Realistic modeling of neurons and networks: towards brain simulation". *Functional Neurology 28*, 153–166 (2013).
2. Einevoll, G. T. *et al.* "The Scientific Case for Brain Simulations". *Neuron 102*, 735–744 (2019).
3. D'Angelo, E. & Jirsa, V. "The quest for multiscale brain modeling". *Trends in Neurosciences 45*, 777–790 (2022).
4. Haufler, D., Ito, S., Koch, C., & Arkhipov, A. "Simulations of cortical networks using spatially extended conductance-based neuronal models". *The Journal of Physiology 601*, 3123–3139 (2023).
5. Zanichelli, N., Schons, M., Freeman, I., Shiu, P., & Arkhipov, A. *State of Brain Emulation Report 2025*. 2025.
6. Arkhipov, A. *et al.* "Integrating multimodal data to understand cortical circuit architecture and function". *Nature Neuroscience 28*, 717–730 (2025).
7. Prinz, A. A., Bucher, D., & Marder, E. "Similar network activity from disparate circuit parameters". *Nature Neuroscience 7*, 1345–1352 (2004).
8. Goaillard, J.-M., Taylor, A. L., Schulz, D. J., & Marder, E. "Functional consequences of animal-to-animal variation in circuit parameters". *Nature Neuroscience 12*, 1424–1430 (2009).
9. Eriksson, O. *et al.* "Combining hypothesis- and data-driven neuroscience modeling in FAIR workflows". *eLife 11*, e69013 (2022).
10. Gouwens, N. W., Berg, J., Feng, D., Sorensen, S. A., Zeng, H., Hawrylycz, M. J., Koch, C., & Arkhipov, A. "Systematic generation of biophysically detailed models for diverse cortical neuron types". *Nature Communications 9*, 710 (2018).
11. Tasic, B. *et al.* "Shared and distinct transcriptomic cell types across neocortical areas". *Nature 563*, 72–78 (2018).
12. Teeter, C. *et al.* "Generalized leaky integrate-and-fire models classify multiple neuron types". *Nature Communications 9*, 709 (2018).
13. Gouwens, N. W. *et al.* "Classification of electrophysiological and morphological neuron types in the mouse visual cortex". *Nature Neuroscience 22*, 1182–1195 (2019).
14. Gouwens, N. W., Sorensen, S. A., Baftizadeh, F., Budzillo, A., Lee, B. R., Jarsky, T., *et al.* "Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells". *Cell 183*, 935–953.e19 (2020).
15. Scala, F. *et al.* "Phenotypic variation of transcriptomic cell types in mouse motor cortex". *Nature 598*, 144–150 (2021).
16. Yao, Z., van Velthoven, C. T. J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., *et al.* "A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain". *Nature 624*, 317–332 (2023).
17. Seeman, S. C. *et al.* "Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex". *eLife 7*, e37349 (2018).
18. Campagnola, L. *et al.* "Local connectivity and synaptic dynamics in mouse and human neocortex". *Science 375*, eabj5861 (2022).
19. De Vries, S. E. J. *et al.* "A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex". *Nature Neuroscience 23*, 138–151 (2020).
20. Siegle, J. H. *et al.* "Survey of spiking in the mouse visual system reveals functional hierarchy". *Nature 592*, 86–92 (2021).
21. Bae, J. A. *et al.* "Functional connectomics spanning multiple areas of mouse visual cortex". *Nature 640*, 435–447 (2025).
22. White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. "The structure of the nervous system of the nematode Caenorhabditis elegans". *Philosophical Transactions of the Royal Society of London. B, Biological Sciences 314*, 1–340 (1986).
23. Cook, S. J. *et al.* "Whole-animal connectomes of both Caenorhabditis elegans sexes". *Nature 571*, 63–71 (2019).
24. Winding, M. *et al.* "The connectome of an insect brain". *Science 379*, eadd9330 (2023).
25. Dorkenwald, S. *et al.* "Neuronal wiring diagram of an adult brain". *Nature 634*, 124–138 (2024).
26. Billeh, Y. N. *et al.* "Systematic Integration of Structural and Functional Data into Multi-scale Models of Mouse Primary Visual Cortex". *Neuron 106*, 388–403.e18 (2020).
27. Rimehaug, A. E. *et al.* "Uncovering circuit mechanisms of current sinks and sources with biophysical simulations of primary visual cortex". *eLife 12*, e87169 (2023).
28. Fraile, J. G., Scherr, F., Ramasco, J. J., Arkhipov, A., Maass, W., & Mirasso, C. R. "Modeling circuit mechanisms of opposing cortical responses to visual flow perturbations". *PLOS Computational Biology 20*, e1011921 (2024).
29. Rimehaug, A. E., Dale, A. M., Arkhipov, A., & Einevoll, G. T. "Uncovering population contributions to the extracellular potential in the mouse visual system using Laminar Population Analysis". *PLOS Computational Biology 20*, e1011830 (2024).
30. Bodor, A. L. *et al.* "The synaptic architecture of layer 5 thick tufted excitatory neurons in mouse visual cortex". *Nature Neuroscience 28*, 1704–1715 (2025).

31. Kuriyama, R. *et al.* *Microscopic-Level Mouse Whole Cortex Simulation Composed of 9 Million Biophysical Neurons and 26 Billion Synapses on the Supercomputer Fugaku.* in (Association for Computing Machinery, 2025), 2158–2171.

32. Chen, G., Scherr, F., & Maass, W. "A data-based large-scale model for primary visual cortex enables brain-like robust and versatile visual processing". *Science Advances 8*, eabq7592 (2022).

33. Wang, C., Zhang, T., Chen, X., He, S., Li, S., & Wu, S. "BrainPy, a flexible, integrative, efficient, and extensible framework for general-purpose brain dynamics programming". *eLife 12*, e86365 (2023).

34. Deistler, M. *et al.* "JAXLEY: differentiable simulation enables large-scale training of detailed biophysical models of neural dynamics". *Nature Methods 22*, 2649–2657 (2025).

35. Allen Institute. *V1 Deep Dive dataset.* Data available at https://github.com/AllenInstitute/v1dd_physiology. 2024.

36. Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjöström, P. J., & Mrsic-Flogel, T. D. "Functional specificity of local synaptic connections in neocortical networks". *Nature 473*, 87–91 (2011).

37. Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., & Mrsic-Flogel, T. D. "Functional organization of excitatory synaptic strength in primary visual cortex". *Nature 518*, 399–403 (2015).

38. Ding, Z. *et al.* "Functional connectomics reveals general wiring rule in mouse visual cortex". *Nature 640*, 459–469 (2025).

39. Dai, K. *et al.* "The SONATA data format for efficient description of large-scale network models". *PLoS Computational Biology 16*, e1007696 (2020).

40. Rossi, L. F., Harris, K. D., & Carandini, M. "Spatial connectivity matches direction selectivity in visual cortex". *Nature 588*, 648–652 (2020).

41. Dai, K. *et al.* "Brain Modeling ToolKit: An open source software suite for multiscale modeling of brain circuits". *PLoS Computational Biology 16*, e1008386 (2020).

42. Durand, S., Iyer, R., Mizuseki, K., de Vries, S., Mihalas, S., & Reid, R. C. "A comparison of visual response properties in the lateral geniculate nucleus and primary visual cortex of awake and anesthetized mice". *Journal of Neuroscience 36*, 12144–12156 (2016).

43. Song, S., Sjöström, P. J., Reigl, M., Nelson, S., & Chklovskii, D. B. "Highly nonrandom features of synaptic connectivity in local cortical circuits". *PLoS Biology 3*, e68 (2005).

44. Buzsáki, G. & Mizuseki, K. "The log-dynamic brain: how skewed distributions affect network operations". *Nature Reviews Neuroscience 15*, 264–278 (2014).

45. Cuntz, H., Bird, A. D., Mittag, M., Beining, M., Schneider, M., Mediavilla, L., Hoffmann, F. Z., Deller, T., & Jedlicka, P. "A general principle of dendritic constancy: A neuron's size-and shape-invariant excitability". *Neuron 109*, 3647–3662 (2021).

46. Markram, H. *et al.* "Reconstruction and Simulation of Neocortical Microcircuitry". *Cell 163*, 456–492 (2015).

47. Reimann, M. W., Gevaert, M., Shi, Y., Lu, H., Markram, H., & Muller, E. "A null model of the mouse whole-neocortex micro-connectome". *Nature communications 10*, 3903 (2019).

48. Dura-Bernal, S. *et al.* "Large-Scale Mechanistic Models of Brain Circuits with Biophysically and Morphologically Detailed Neurons". *Journal of Neuroscience 44* (2024).

49. Nicola, W. & Clopath, C. "Supervised learning in spiking neural networks with FORCE training". *Nature communications 8*, 2208 (2017).

50. Huh, D. & Sejnowski, T. J. *Gradient Descent for Spiking Neural Networks.* in *Advances in Neural Information Processing Systems* (eds Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., & Garnett, R.) **31** (Curran Associates, Inc., 2018).

51. Bellec, G., Salaj, D., Subramoney, A., Legenstein, R., & Maass, W. *Long short-term memory and Learning-to-learn in networks of spiking neurons.* in *Advances in Neural Information Processing Systems* **31** (2018).

52. Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., & Maass, W. "A solution to the learning dilemma for recurrent networks of spiking neurons". *Nature Communications 11*, 3625 (2020).

53. Strata, P. & Harvey, R. "Dale's principle". *Brain Research Bulletin 50*, 349–350 (1999).

54. Taherkhani, A., Belatreche, A., Li, Y., Cosma, G., Maguire, L. P., & McGinnity, T. M. "A review of learning in biologically plausible spiking neural networks". *Neural Networks 122*, 253–272 (2020).

55. Neftci, E. O., Mostafa, H., & Zenke, F. "Surrogate Gradient Learning in Spiking Neural Networks: Bringing the Power of Gradient-Based Optimization to Spiking Neural Networks". *IEEE Signal Processing Magazine 36*, 51–63 (2019).

56. Zenke, F. & Vogels, T. P. "The Remarkable Robustness of Surrogate Gradient Learning for Instilling Complex Function in Spiking Neural Networks". *Neural Computation 33*, 899–925 (2021).

57. Eshraghian, J. K., Ward, M., Neftci, E. O., Wang, X., Lenz, G., Dwivedi, G., Bennamoun, M., Jeong, D. S., & Lu, W. D. "Training spiking neural networks using lessons from deep learning". *Proceedings of the IEEE* (2023).

58. Kingma, D. P. & Ba, J. *Adam: A Method for Stochastic Optimization.* 2014.

59. Nemirovskij, A. S. & Yudin, D. B. *Problem Complexity and Method Efficiency in Optimization.* A Wiley-Interscience publication; translated by E. R. Dawson, 388. ISBN: 9780471103455 (John Wiley & Sons, Chichester and New York, 1983).
60. Surace, S. C., Pfister, J.-P., Gerstner, W., & Brea, J. "On the choice of metric in gradient-based theories of brain function". *PLOS Computational Biology 16*, e1007640 (2020).
61. Loewenstein, Y., Kuras, A., & Rumpel, S. "Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo". *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 31*, 9481–9488 (2011).
62. Dorkenwald, S. *et al.* "Binary and analog variation of synapses between cortical pyramidal neurons". *eLife 11*, e76120 (2022).
63. Pogodin, R., Cornford, J., Ghosh, A., Gidel, G., Lajoie, G., & Richards, B. *Synaptic Weight Distributions Depend on the Geometry of Plasticity*. 2023.
64. Cornford, J., Pogodin, R., Ghosh, A., Sheng, K., Bicknell, B. A., Codol, O., Clark, B. A., Lajoie, G., & Richards, B. A. *Brain-like learning with exponentiated gradients*. 2024.
65. Micikevicius, P. *et al.* "Mixed precision training". *arXiv preprint arXiv:1710.03740* (2017).
66. Chen, T., Xu, B., Zhang, C., & Guestrin, C. "Training deep nets with sublinear memory cost". *arXiv preprint arXiv:1604.06174* (2016).
67. Lappalainen, J. K. *et al.* "Connectome-constrained networks predict neural activity across the fly visual system". *Nature 634*, 1132–1140 (2024).
68. Kim, C. M. & Chow, C. C. "Training Spiking Neural Networks in the Strong Coupling Regime". *Neural computation 33*, 1199–1233 (2021).
69. Herranz-Celotti, L. & Rouat, J. *Stabilizing Spiking Neuron Training*. 2024.
70. Millman, D. J. *et al.* "VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli". *eLife 9*, e55130 (2020).
71. Walker, E. Y. *et al.* "Inception loops discover what excites neurons most using deep predictive models". *Nature Neuroscience 22*, 2060–2065 (2019).
72. Wertz, A. *et al.* "Single-cell–initiated monosynaptic tracing reveals layer-specific cortical network modules". *Science 349*, 70–74 (2015).
73. Ogando, M. B., Abdeladim, L., Sit, K. K., Shin, H., Sridharan, S., Gopakumar, K., & Adesnik, H. "Feature-specific inhibitory connectivity augments the accuracy of cortical representations". *bioRxiv*, 2025.08.02.668307 (2025).
74. Fink, A. J. P., Muscinelli, S. P., Wang, S., Hogan, M. I., English, D. F., Axel, R., Litwin-Kumar, A., & Schoonover, C. E. *Experience-dependent reorganization of inhibitory neuron synaptic connectivity*. 2025.
75. Lee, W.-C. A., Bonin, V., Reed, M., Graham, B. J., Hood, G., Glattfelder, K., & Reid, R. C. "Anatomy and function of an excitatory network in the visual cortex". *Nature 532*, 370–374 (2016).
76. Finkelstein, A., Daie, K., Rózsa, M., Darshan, R., & Svoboda, K. "Connectivity underlying motor cortex activity during goal-directed behaviour". *Nature* (2025).
77. Kuan, A. T. *et al.* "Synaptic wiring motifs in posterior parietal cortex support decision-making". *Nature 627*, 367–373 (2024).
78. Znamenskiy, P., Kim, M.-H., Muir, D. R., Iacaruso, M. F., Hofer, S. B., & Mrsic-Flogel, T. D. "Functional specificity of recurrent inhibition in visual cortex". *Neuron 112*, 991–1000.e8 (2024).
79. Gal, E., London, M., Globerson, A., Ramaswamy, S., Reimann, M. W., Muller, E., Markram, H., & Segev, I. "Rich cell-type-specific network topology in neocortical microcircuitry". *Nature Neuroscience 20*, 1004–1013 (2017).
80. Bonifazi, P., Goldin, M., Picardo, M. A., Jorquera, I., Cattani, A., Bianconi, G., Represa, A., Ben-Ari, Y., & Cossart, R. "GABAergic hub neurons orchestrate synchrony in developing hippocampal networks". *Science (New York, N.Y.) 326*, 1419–1424 (2009).
81. Luccioli, S., Angulo-Garcia, D., Cossart, R., Malvache, A., Módol, L., Sousa, V. H., Bonifazi, P., & Torcini, A. "Modeling driver cells in developing neuronal networks". *PLOS Computational Biology 14*, e1006551 (2018).
82. Sanzeni, A., Akitake, B., Goldbach, H. C., Leedy, C. E., Brunel, N., & Histed, M. H. "Inhibition stabilization is a widespread property of cortical networks". *eLife 9*, e54875 (2020).
83. Mahrach, A., Chen, G., Li, N., van Vreeswijk, C., & Hansel, D. "Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation". *eLife 9*, e49967 (2020).
84. Jun, J. J. *et al.* "Fully integrated silicon probes for high-density recording of neural activity". *Nature 551*, 232–236 (2017).
85. Potjans, T. C. & Diesmann, M. "The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model". *Cerebral Cortex 24*, 785–806 (2014).
86. Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., & McNaughton, B. L. "Paradoxical effects of external modulation of inhibitory interneurons". *The Journal of Neuroscience 17*, 4382–4388 (1997).
87. Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., & Scanziani, M. "Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons". *Nature Neuroscience 16*, 1068–1076 (2013).

88. Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., & Kepecs, A. "Cortical interneurons that specialize in disinhibitory control". *Nature 503*, 521–524 (2013).
89. Marder, E. & Taylor, A. L. "Multiple models to capture the variability in biological neurons and networks". *Nature Neuroscience 14*, 133–138 (2011).
90. Dura-Bernal, S. *et al.* "Multiscale model of primary motor cortex circuits predicts in vivo cell-type-specific, behavioral state-dependent dynamics". *Cell Reports 42* (2023).
91. Plesser, H. E. *et al.* "Building on models—a perspective for computational neuroscience". *Cerebral Cortex 35*, bhaf295 (2025).
92. Igarashi, J., Yamaura, H., & Yamazaki, T. "Large-Scale Simulation of a Layered Cortical Sheet of Spiking Network Model Using a Tile Partitioning Method". *Frontiers in Neuroinformatics 13*, 71 (2019).
93. Lu, W. *et al.* *Digital Twin Brain: a simulation and assimilation platform for whole human brain*. 2023.

## Methods

### *V1 Network Construction*

**Data sources.** The V1 model integrates data from several Allen Institute datasets. Synaptic physiology data—paired whole-cell recordings providing PSP/PSC amplitudes and kinetics across cell-type combinations—were obtained from the Allen Institute Synaptic Physiology dataset[18]. Cell types were identified using Cre-driver lines targeting Pvalb, Sst, and Vip interneurons as well as excitatory populations, consistently with the earlier modeling approach[26]. Electron microscopy connectomics data were drawn from the MICrONS dataset[21], an EM volume spanning multiple visual cortical areas with matched 2-photon calcium imaging measurements of neural activity from the same neurons, and from the V1 Deep Dive (V1DD) dataset[35], which provides similar calcium imaging and EM data centered in V1. Due to its sole focus on V1, the V1DD dataset was used to derive connection probabilities and spatial spread parameters. Interneurons in V1DD were identified by innervation pattern—proximal-targeting, distal-targeting, and inhibitory-cell-targeting—and assigned to model PV, SST, and VIP types, respectively; interneurons with somata in L1 were assigned to the L1 Inh type. MICrONS data were used to characterize LGN axon synapse statistics, which has not been yet mapped out in V1DD, as well as in-degree distributions by cell class and per-connection synapse count distributions. For the latter two, we used MICrONS due to a more extensive proofreading (as of now) than in V1DD, therefore allowing for better statistics—the assumption being that in-degree and synapse count distributions do not differ substantially between V1 and nearby higher visual areas (whereas connection probabilities within these different areas may be reasonably expected to differ). Intrinsic electrophysiology and cell-type characterization data from the Allen Cell Types Database[12,13] provided the GLIF neuron models, and in vivo recordings from the Allen Brain Observatory[19] and Neuropixels Visual Coding[20] datasets provided firing-rate and selectivity targets for model evaluation and training (see Training Objective).

**Network composition.** The network contains 19 cell types organized by cortical layer and transcriptomic subclass, with the limitation that little data is available to differentiate certain subclasses in terms of their connectivity and other properties. This resulted in using six excitatory populations (L2/3 excitatory, L4 excitatory, L5 ET excitatory, L5 IT excitatory, L5 NP excitatory, and L6 excitatory) and thirteen inhibitory populations (PV, SST, and VIP in each of L2/3, L4, L5, and L6, plus L1 Inh). Neurons are represented as 201 distinct GLIF Type-3 point models (LIF with after-spike currents), each fit to an individual mouse V1 neuron from the Allen Cell Types Database[12] and parameterized by experimentally measured membrane capacitance, input resistance, resting potential, spike threshold, after-spike current amplitudes and time constants, and refractory period. Only models achieving an explained variance ratio $\geq 0.7$ on a held-out fixed-noise stimulus were retained. Multiple GLIF models per cell type capture within-class electrophysiological diversity, and, consequently, most cell types in our V1 model are represented by multiple GLIF models.

The densities of the excitatory and inhibitory neurons are estimated for each layer from the MICrONS EM data. Within inhibitory neurons, relative populations of PV, SST, and VIP neurons are set for each layer based on the data by Lee et al[94]. For a cylindrical cortical column with 400 μm radius, the total number of neurons is $\sim 67,000$. At construction, each V1 neuron is assigned a random position within the cylindrical column, with depth drawn from layer-appropriate ranges. Each neuron is also assigned a tuning angle, which is used to guide both network construction (e.g., LGN inputs, orientation-dependent connectivity rules) and model-training objectives. Finally, each neuron is assigned a target in-degree $N_{\text{syn}}^{\text{target}}$, representing its total dendritic input synapse count, drawn from a cell-type-specific log-normal distribution. These distributions were fit by maximum likelihood to total input synapse counts quantified for each cell class in the MICrONS volume[21],

using only neurons with proofread dendritic reconstructions and excluding those with fewer than 700 synapses as likely incompletely reconstructed. Cortical layer assignments were determined from soma positions[21]. The target in-degree serves as a proxy for dendritic extent and is used during both connectivity assignment (Target-size modulation) and synaptic weight construction (Weight decomposition).

V1 recurrent connectivity. The baseline connectivity between V1 cell types is modeled as a distance-dependent Gaussian: for each pre–post cell-type pair $a \to b$, the connection probability decays as a Gaussian function of lateral distance with peak probability $p^0_{ab}$ (at zero separation) and spatial spread $\sigma_{ab}$. These two parameters were fit for each cell-type pair by maximum likelihood estimation (MLE) applied to connection probability data from the V1DD dataset[35], supplemented by synaptic physiology measurements[18] and literature values where V1DD data were insufficient. This distance-dependent Gaussian profile serves as the foundation for recurrent connectivity, which is then refined by three additional rules: an orientation-dependent spatial modulation, a like-to-like modulation of excitatory connections, and a target-size modulation based on EM-derived in-degree distributions.

Spatial connectivity rule. The baseline Gaussian connectivity profile is extended to incorporate orientation-dependent spatial structure[26,40]: excitatory presynaptic ensembles are elongated—in terms of the arrangement of the receptive fields of the presynaptic neurons—along the postsynaptic neuron's preferred orientation, while inhibitory ensembles are displaced ahead of the preferred direction. We implement this as an anisotropic, orientation-dependent bivariate normal connectivity profile:

$$p_{ab}(\Delta\mathbf{r}) = p^0_{ab} \cdot \frac{\mathcal{N}(\Delta\mathbf{r};\, \boldsymbol{\mu}(\theta),\, \boldsymbol{\Sigma}(\theta))}{\mathcal{N}(\boldsymbol{\mu};\, \boldsymbol{\mu}(\theta),\, \boldsymbol{\Sigma}(\theta))}, \tag{1}$$

where $\Delta\mathbf{r} = \mathbf{r}_{\text{pre}} - \mathbf{r}_{\text{post}}$ is the lateral displacement (excluding cortical depth), $p^0_{ab}$ is the peak connection probability, and the denominator normalizes so that $p = p^0_{ab}$ at the mode. For **inhibitory** presynaptic neurons, the centroid is displaced by $d_R = 50$ µm along the postsynaptic neuron's preferred orientation, with isotropic spread: $\boldsymbol{\mu}_I = d_R[\cos\theta, \sin\theta]^T$, $\boldsymbol{\Sigma}_I = R_\theta \operatorname{diag}(\sigma^2_{ab}, \sigma^2_{ab})\, R^T_\theta$. For **excitatory** presynaptic neurons, the centroid is displaced by $d_R$ in the opposite direction, with anisotropic covariance ($\kappa = 1.5$): $\boldsymbol{\mu}_E = -d_R[\cos\theta, \sin\theta]^T$, $\boldsymbol{\Sigma}_E = R_\theta \operatorname{diag}((\sigma_{ab}/\kappa)^2, (\sigma_{ab} \cdot \kappa)^2)\, R^T_\theta$. The displacement is applied only within 1.5 times the core radius (default 200 µm) from the network center; outside this region, it is set to zero, and the profile reduces to a centered Gaussian.

Orientation-dependent (like-to-like) modulation. For excitatory-to-excitatory connection types, connection probability is additionally modulated by the orientation difference between pre- and postsynaptic neurons:

$$p_{ab} \to p_{ab} \cdot (B_1 + G\,\Delta\phi), \tag{2}$$

where $\Delta\phi = |\phi_{\text{pre}} - \phi_{\text{post}}|$ (folded to 0–90°), $B_1 = 2/(1+\beta)$, $B_2 = \beta\, B_1$, and $G = (B_2 - B_1)/90$. When $\beta < 1$, the rule elevates probability for similarly tuned neurons and reduces it for orthogonally tuned neurons. The modulation averages to unity over all orientation differences, preserving $p^0_{ab}$.

Target-size modulation. As described above (Network composition), each V1 neuron is assigned a target in-degree $N^{\text{target}}_{\text{syn}}$ representing its total dendritic input synapse count. Connection probability is scaled by the ratio of the neuron's target in-degree to its population mean:

$$p_{ab} \to p_{ab} \cdot \frac{N^{\text{target}}_{\text{syn}}}{\langle N_{\text{syn}} \rangle_b}, \tag{3}$$

so that neurons with larger dendritic arbors receive proportionally more connections, preserving the log-normal variability in total input observed in EM data. The same size parameter is reused during synaptic weight assignment (see below).

**Synaptic kinetics.** Synaptic currents are modeled with a double alpha function,

$$I(t) = A_1 \frac{t}{\tau_1} e^{-t/\tau_1} + A_2 \frac{t}{\tau_2} e^{-t/\tau_2}, \tag{4}$$

to capture distinct fast and slow components observed in voltage-clamp traces from the Synaptic Physiology dataset[18]. Averaged traces were aligned to response onset and fit by nonlinear least squares; the double alpha function substantially improved fits, relative to single-alpha function, across connection types (e.g., $R^2$ increased from 0.79 to 0.92 for Pvalb→Tlx3 L5; Fig. 1e). The effective synaptic time constant $\tau_{\text{syn}}$ for each pre–post combination was calibrated by simulating single spikes onto GLIF models and minimizing the difference between simulated and experimental rise times, averaged across models within each postsynaptic population.

**PSP amplitude distributions.** Synaptic strengths were drawn from log-normal distributions[43,44] fitted to experimental PSP amplitudes. For each connected pair in the Synaptic Physiology dataset[18], the 90th-percentile pulse response amplitude was used as a robust estimate of synaptic strength under elevated release probability. Log-normal distributions were fit by maximum likelihood for each connection type with $N \geq 5$ pairs. Because individual GLIF models differ in intrinsic properties, a per-model calibration step converts desired PSP amplitudes (in mV) to the simulator's dimensionless weight scale. For each of the 201 GLIF models and each connection type, a single presynaptic spike is delivered at unit weight using the calibrated double-alpha kernel in a minimal BMTK/NEST network, and the resulting peak membrane potential deflection is recorded as the unitary PSP amplitude. The final simulator weight is then obtained by dividing the desired PSP amplitude by this unitary value.

**Synaptic weight distributions.** Because neurons differ in their total number of input connections (in-degree), per-connection synaptic weights must be adjusted so that neurons with more inputs receive individually weaker connections, preserving the experimentally observed distribution of PSP amplitudes across the population[45]. This is achieved by exploiting the closure of the log-normal family of distributions under quotients: dividing one log-normal variable by an independent log-normal variable yields a log-normal result whose log-variance equals the sum of the two individual log-variances.

For each connection type $a \to b$, the experimentally fitted PSP distribution $\text{LogNormal}(\sigma_{\text{PSP}}, \mu_{\text{PSP}})$ and the target in-degree distribution $\text{LogNormal}(\sigma_N, \mu_N)$ together determine a per-connection weight distribution $\text{LogNormal}(\sigma_W, \mu_W)$ with

$$\sigma_W = \sqrt{\sigma_{\text{PSP}}^2 - \sigma_N^2}, \qquad \mu_W = \frac{\mu_{\text{PSP}} \cdot \mu_N}{\langle N_{\text{syn}} \rangle_b}. \tag{5}$$

A weight $W$ is drawn independently for each connection, and the final weight is scaled by the ratio of the population-mean in-degree to the target neuron's in-degree:

$$W_{\text{final}} = W \cdot \frac{\langle N_{\text{syn}} \rangle_b}{N_{\text{syn}}^{\text{target}}}. \tag{6}$$

Because $W$ and $N_{\text{syn}}^{\text{target}}$ are independent by construction, the quotient $W/N_{\text{syn}}^{\text{target}}$ is log-normal with variance $\sigma_W^2 + \sigma_N^2 = \sigma_{\text{PSP}}^2$ and median $\mu_{\text{PSP}}$, exactly recovering the experimental PSP distribution.

When $\sigma_{\mathrm{PSP}} < \sigma_N$, $\sigma_W$ is set to a small positive value, effectively removing weight variability for that connection type.

**LGN inputs.** Visual input is provided through an LGN model comprising 17,400 units, implemented with BMTK's FilterNet module[41], which generates spike trains by passing stimuli through spatiotemporally separable linear–nonlinear filters approximating LGN receptive field properties. The LGN model comprises sustained-ON, sustained-OFF, and transient-OFF functional classes[26], characterized by Durand et al.[42]. Direction selectivity in V1 arises from convergent input of spatially offset sustained and transient afferents, following the mechanism of Lien & Scanziani[95].

Thalamocortical connection statistics were constrained using anatomically identified LGN axons from the MICrONS volume[21]. Forty-three LGN axons were reconstructed and their synapses onto V1 neurons quantified (Fig. 1g). Based on the number of synapses from LGN axons to each cell type, the relative probability each V1 neuron receives from LGN is calculated (relative to L4 Exc cells; Extened Data Fig. 1). Per-connection synapse counts were modeled with the Yule–Simon distribution[96],

$$P(k) = \rho\, B(k, \rho + 1), \tag{7}$$

where $\rho > 0$ is the shape parameter and $B(\cdot, \cdot)$ the beta function. The Yule–Simon distribution arises from preferential attachment[97] and produces a power-law tail with exponent $\rho + 1$ without requiring a minimum-degree threshold. The shape parameter was fit by maximum likelihood to the MICrONS data for each of 6 excitatory types (layer segregated) and 3 inhibitory types (layer aggregated).

For each V1 target neuron, the total LGN synapse count is estimated from the neuron's in-degree (with LGN synapses comprising a fixed fraction of the total, e.g. 20% for L4 excitatory neurons), and the number of distinct connections is derived as $N_{\mathrm{conn}} = N_{\mathrm{syn}} \cdot (\rho - 1)/\rho$. Source LGN neurons are selected probabilistically based on receptive-field proximity and functional subtype, and each connection is assigned a synapse count drawn from the Yule–Simon distribution. Connections with higher synapse counts are preferentially assigned to spatially closer sources (Gaussian kernel with $\sigma = \sigma_{\mathrm{LGN}}/\sqrt{n_{\mathrm{syn}}}$, $\sigma_{\mathrm{LGN}} = 5°$), so that stronger connections arise from more precisely aligned receptive fields.

**Background inputs.** In addition to LGN-driven visual input, each V1 neuron receives excitatory background drive representing possible inputs from the rest of the brain. The background population consists of 100 virtual Poisson spike-generator units with a frequency of 250 Hz. Each V1 neuron is connected to 4 randomly selected background units, with synaptic weights specified per target cell model. All background connections use a delay of 1.0 ms[98].

### *Implementation of the TensorFlow-based simulator*

While the model can be run on the NEST simulator using CPUs, gradient-based fitting requires a differentiable implementation that supports backpropagation through time (BPTT). We therefore implemented a TensorFlow-based simulator following the general approach of Chen et al.[32], and substantially optimized it for the present study to improve GPU memory efficiency and throughput.

**Single-neuron dynamics ($\mathrm{GLIF}_3$).** All neurons were modeled with the $\mathrm{GLIF}_3$ formalism[12], which extends the leaky integrate-and-fire model with after-spike currents (ASCs) that capture refractory/adaptive effects from ionic channel dynamics. Simulations used forward Euler integration with time step $\Delta t = 1$ ms.

For neuron $j$, let $V_j[n]$ denote membrane potential at discrete step $n$, and let $I_j^{(q)}[n]$ ($q \in \{1, 2\}$)

denote the two ASC states. Defining

$$\alpha_j = e^{-\Delta t/\tau_{m,j}}, \qquad \beta_{j,q} = e^{-k_{j,q}\Delta t},$$

where $\tau_{m,j}$ is the membrane time constant and $k_{j,q}$ the decay rates of the ASCs, the dynamics are

$$\begin{aligned} V_j[n+1] &= (1-\alpha_j)E_{\mathrm{L},j} + \alpha_j V_j[n] + (1-\alpha_j)R_{m,j}\left(I_j^{\mathrm{syn}}[n] + \sum_{q=1}^{2} I_j^{(q)}[n]\right) - S_j[n]\Delta V_j, \\ \Delta V_j &= V_{\mathrm{th},j} - E_{\mathrm{L},j}, \\ S_j[n] &= \mathbb{1}[\rho_j[n]=0]\, H\left(V_j[n] - V_{\mathrm{th},j}\right), \\ I_j^{(q)}[n+1] &= \beta_{j,q} I_j^{(q)}[n] + S_j[n]\Delta I_{j,q}, \qquad q \in \{1,2\}, \end{aligned} \tag{8}$$

where $R_{m,j}$ is membrane resistance, $I_j^{\mathrm{syn}}[n]$ is the total synaptic current (recurrent V1 + LGN + BKG), $E_{L,j}$ the resting membrane potential, $V_{\mathrm{th},j}$ the firing threshold, $S_j[n] \in \{0,1\}$ is the spike indicator, $H$ is the Heaviside function, and $\rho_j[n]$ is an absolute-refractory countdown state (in ms steps). The term $S_j[n]\Delta I_{j,q}$ implements the spike-triggered increment of the $q$-th ASC. We used a *soft reset* (subtracting $\Delta V_j$ upon spiking) rather than a hard reset, following prior surrogate-gradient RSNN work and Chen et al.[32,99]. This preserves forward spiking behavior while improving gradient propagation.

The refractory countdown is updated as

$$\rho_j[n+1] = \begin{cases} \tau_{\mathrm{ref},j}/\Delta t, & \text{if } S_j[n]=1, \\ \max(\rho_j[n]-1, 0), & \text{otherwise}, \end{cases} \tag{9}$$

so that spike emission is disabled during the cell-type-specific absolute refractory period $\tau_{\mathrm{ref},j}$ (2–8 ms in our fitted parameter set).

Neuron electrophysiological parameters were taken from the fitted GLIF models in the Allen Cell Types Database[12] (201 selected neurons).

**Double-alpha synaptic current model.** Synaptic interactions were modeled with a current-based double-alpha formulation, capturing both fast and slow PSC components. For clarity, we write the dynamics for one postsynaptic neuron $j$ and one synaptic receptor family $r$ (the total current is the sum over families and presynaptic sources). The receptor-family contribution is

$$I_{j,r}^{\mathrm{syn}}[n] = I_{j,r}^{(f)}[n] + \lambda_r I_{j,r}^{(s)}[n], \tag{10}$$

where $\lambda_r$ scales the relative contribution of the slow timescale, and the full synaptic current is

$$I_j^{\mathrm{syn}}[n] = \sum_r I_{j,r}^{\mathrm{syn}}[n]. \tag{11}$$

Rather than explicitly convolving each presynaptic spike train with a kernel (which requires storing spike histories), we use an equivalent linear-state implementation[98,100]. Let $I_{j,r}^{(\ell)}[n]$ and $C_{j,r}^{(\ell)}[n]$ denote the two state variables of alpha component $\ell \in \{f,s\}$ (fast/slow), with decay time constants

$\tau_{r,\ell}$. For each component $\ell \in \{f, s\}$,

$$\begin{aligned} I_{j,r}^{(\ell)}[n+1] &= e^{-\Delta t/\tau_{r,\ell}}\, I_{j,r}^{(\ell)}[n] + \Delta t\, e^{-\Delta t/\tau_{r,\ell}}\, C_{j,r}^{(\ell)}[n], \\ C_{j,r}^{(\ell)}[n+1] &= e^{-\Delta t/\tau_{r,\ell}}\, C_{j,r}^{(\ell)}[n] + \frac{e}{\tau_{r,\ell}} \sum_{\text{src}} \sum_{i \in \text{pre}_{r,\text{src}}(j)} W_{ji,r}^{\text{src}}\, S_i^{\text{src}}[n - d_{ji,r}^{\text{src}}], \end{aligned} \tag{12}$$

where $W_{ji,r}^{\text{src}}$ is the signed synaptic weight, $d_{ji,r}^{\text{src}}$ is the axonal delay (in simulation steps), and $\text{pre}_{r,\text{src}}(j)$ is the presynaptic set for neuron $j$, receptor family $r$, and source $\text{src} \in \{\text{rec}, \text{lgn}, \text{bkg}\}$. This formulation yields constant-cost updates per time step and is well-suited to parallel implementation. Pair-specific parameters $(\tau_{r,f}, \tau_{r,s}, \lambda_r)$ were fit to synaptic physiology recordings[18], enabling cell-type-resolved synaptic kinetics consistent with experimental measurements.

The current implementation did not include synaptic plasticity or creation/elimination of connections due to a substantially higher computational expense associated with such functionality and a lack of extensive data to parameterize diverse connection types. This choice is appropriate for simulations of short-timescale (e.g., seconds) network dynamics studied here, while leaving the framework compatible with future extensions that may include plasticity mechanisms[101,102].

State normalization and numerical stability. To improve numerical stability in large-scale mixed-precision simulation[103], we normalize voltages and synaptic weights using the neuron-specific threshold distance $\Delta V_j = V_{\text{th},j} - E_{\text{L},j}$. We define

$$\tilde{v}_j[n] = \frac{V_j[n] - E_{\text{L},j}}{\Delta V_j}, \qquad w_{ji,r}^{\text{src}} = \frac{W_{ji,r}^{\text{src}}}{\Delta V_j}. \tag{13}$$

Under this transformation, reset is at $\tilde{v} = 0$ and threshold is at $\tilde{v} = 1$.

Using normalized currents (same scaling by $\Delta V_j$) and suppressing receptor/source indices for readability, the membrane update becomes

$$\tilde{v}_j[n+1] = \alpha_j \tilde{v}_j[n] + (1 - \alpha_j) R_{m,j}\, \tilde{i}_j[n] - S_j[n], \qquad S_j[n] = \mathbb{1}[\rho_j[n] = 0]\, H(\tilde{v}_j[n] - 1), \tag{14}$$

with $\tilde{i}_j[n]$ the normalized total current (ASC + synaptic contributions). The corresponding ASC and PSC state updates retain the same linear form as Eqs. (8) and (12), with amplitudes/weights expressed in normalized units. This scaling keeps principal state variables in comparable numerical ranges while preserving the underlying biophysical dynamics.

### *Gradient-based learning and biologically constrained optimization in RSNNs*

Training the recurrent V1 circuit requires solving temporal credit assignment under non-differentiable spike generation while preserving structural synaptic constraints (e.g., fixed sign by cell class). In the Results (Fig. 2a,b), this corresponds to the differentiable fitting stage that maps a biologically grounded initialization to a constrained physiological operating regime.

BPTT factorization. Backpropagation through time (BPTT)[104] unfolds the recurrent network across time and applies gradient descent on the resulting acyclic computational graph. Let $n = 1, \ldots, T$ index simulation steps within a training branch (here $T = 500$ for a 500 ms branch with $\Delta t = 1$ ms). We denote by $\mathbf{h}_n$ the full simulator hidden state at time step $n$, which comprises all dynamical variables of the network, including membrane potentials, ASC states, synaptic states, refractory counters, and delay-buffer states. Inputs $\mathbf{x}_n$ include stimulus-dependent LGN drive and background spikes.

The vector $\mathbf{z}_n \in \{0,1\}^N$ denotes the binary spike output of the network at step $n$, where $(\mathbf{z}_n)_j = 1$ indicates that neuron $j$ emits a spike and 0 otherwise. We use bold notation for vectors and matrices throughout. The recurrent dynamics can be written as

$$\mathbf{h}_n = f(\mathbf{x}_n, \mathbf{h}_{n-1}; \theta), \qquad \mathbf{z}_n = g(\mathbf{h}_n) \in \{0,1\}^N, \tag{15}$$

where $\mathbf{h}_0$ denotes the initial network state and $\theta$ the collection of trainable synaptic parameters (recurrent and background weights in normalized coordinates), and $g$ implements spike generation applied to the hidden state.

For a scalar loss $\mathcal{L}$, the exact BPTT gradient is

$$\nabla_\theta \mathcal{L} = \sum_{n=1}^{T} \frac{d\mathcal{L}}{d\mathbf{h}_n} \frac{d\mathbf{h}_n}{d\theta} = \sum_{n=1}^{T} \mathbf{E}_n^\top \delta_n, \tag{16}$$

with backward learning signals $\delta_n$ and forward eligibility traces $\mathbf{E}_n$[98,104,105]:

$$\delta_n = \mathbf{D}_n^\top \frac{\partial \mathcal{L}}{\partial \mathbf{z}_n} + \mathbf{J}_{n+1}^\top \delta_{n+1}, \qquad \delta_{T+1} = \mathbf{0}, \tag{17}$$

$$\mathbf{E}_n = \frac{\partial f_n}{\partial \theta} + \mathbf{J}_n \mathbf{E}_{n-1}, \qquad \mathbf{E}_0 = \mathbf{0}, \tag{18}$$

where $\mathbf{J}_n = \partial \mathbf{h}_n / \partial \mathbf{h}_{n-1}$ and $\mathbf{D}_n = \partial \mathbf{z}_n / \partial \mathbf{h}_n$. In practice, these derivatives are computed automatically by TensorFlow[106].

**Surrogate gradient for spike generation.** Because $g$ contains a Heaviside threshold, $\partial \mathbf{z}_n / \partial \mathbf{h}_n$ is zero almost everywhere under the exact derivative. We therefore use BPTT with surrogate gradients[55,56], preserving exact spiking in the forward pass while replacing the threshold derivative during back-propagation.

Let $u_{j,n} = \tilde{v}_j[n] - 1$ denote normalized distance to threshold. We use a triangular surrogate

$$\frac{\partial S_j[n]}{\partial u_{j,n}} \approx \gamma \max(1 - |u_{j,n}|, 0), \qquad \gamma = 0.1, \tag{19}$$

which has support in a unit-width window around the threshold and a damped slope $\gamma$[51]. In our setting, $\gamma$ was chosen to stabilize long-horizon BPTT in the large recurrent network (reducing the risk of gradient explosion) while maintaining useful gradient flow.

### *Biologically constrained optimization*

Temporal credit assignment alone does not intrinsically incorporate a biologically consistent geometry of synaptic change. Standard gradient descent optimizers such as Adam[58] operate in a Euclidean geometry with additive parameter updates. Although computationally effective[32,51,52], they are biologically agnostic: they do not intrinsically preserve the synaptic sign (which requires explicit constraints to enforce Dale's law) and do not respect the approximately lognormal distribution of cortical synaptic strengths[44,63,64].

In contrast, exponentiated gradient descent (EGD), a mirror-descent method that performs additive updates in log-space, is intrinsically sign-preserving and is more compatible with the heavy-tailed, lognormal-like statistics observed in cortex[59,60,63,64]. Indeed, EGD not only retains the standard gradient descent optimization power while embedding biologically meaningful constraints,

but also excels at biologically relevant tasks with sparse signals and confers robustness to synaptic pruning[64].

For each trainable synapse, we decompose the normalized weight $w_k$ into a fixed sign mask $s_k$ and a positive magnitude $a_k$:

$$w_k = s_k a_k, \qquad s_k \in \{-1, +1\} \text{ fixed}, \qquad a_k > 0. \tag{20}$$

The sign $s_k$ is determined by presynaptic cell class(Dale-consistent) and remains fixed during training. Optimization is performed on the positive magnitudes $\mathbf{a}$ (equivalently, in log-magnitude space). To make the multiplicative update unambiguous, we write it componentwise as

$$a_k^{(u+1)} = a_k^{(u)} \exp\left(-\eta \frac{\partial \mathcal{L}}{\partial a_k}\left(\mathbf{a}^{(u)}\right)\right), \qquad k = 1, \ldots, P, \tag{21}$$

where $P$ is the number of trainable synapses, $u$ indexes optimizer updates, and $\eta > 0$ is the learning rate. This update preserves positivity of every $a_k$ by construction and therefore preserves the synaptic signs when mapped back to $w_k$.

Exponentiated Adam optimizer. RSNN training exhibits heterogeneous gradient magnitudes across time, synapse types, and loss terms. To improve optimization stability while retaining multiplicative geometry, we use an *Exponentiated Adam* optimizer: Adam-style adaptive moments[58] combined with the multiplicative, sign-preserving structure of EGD[63,64].

For each trainable synapse $k = 1, \ldots, P$, let

$$g_k^{(u)} = \frac{\partial \mathcal{L}}{\partial a_k}\left(\mathbf{a}^{(u)}\right),$$

where $u$ indexes optimizer updates. Following Adam, we maintain two exponential moving averages of the gradient: a first-moment estimate $m_k^{(u)}$, which tracks the mean gradient, and a second-moment estimate $v_k^{(u)}$, which tracks the mean squared gradient. These quantities are updated componentwise as

$$m_k^{(u)} = \beta_1 m_k^{(u-1)} + (1 - \beta_1) g_k^{(u)}, \tag{22}$$

$$v_k^{(u)} = \beta_2 v_k^{(u-1)} + (1 - \beta_2)\left(g_k^{(u)}\right)^2, \tag{23}$$

with bias-corrected forms

$$\hat{m}_k^{(u)} = \frac{m_k^{(u)}}{1 - \beta_1^u}, \qquad \hat{v}_k^{(u)} = \frac{v_k^{(u)}}{1 - \beta_2^u}. \tag{24}$$

The exponentiated update is then

$$a_k^{(u+1)} = a_k^{(u)} \exp\left(-\eta \frac{\hat{m}_k^{(u)}}{\sqrt{\hat{v}_k^{(u)}} + \epsilon}\right), \qquad k = 1, \ldots, P, \tag{25}$$

and the signed synaptic weights are recovered as

$$w_k^{(u+1)} = s_k a_k^{(u+1)}. \tag{26}$$

Equivalently, in vector form,

$$\mathbf{w}^{(u+1)} = \mathbf{s} \odot \mathbf{a}^{(u+1)},$$

where $\odot$ denotes elementwise multiplication. This adaptive multiplicative rule preserves positivity of each $a_k$ by construction and therefore preserves the synaptic sign pattern throughout training. Because updates are applied in log-magnitude space, the optimizer is more compatible with multiplicative changes in synaptic strength than standard additive updates and is consistent with the broad, heavy-tailed synaptic weight distributions reported in cortex[60,61,63,64]. In practice, it provided stable optimization in our large recurrent spiking model while retaining the efficiency of Adam-style adaptive moment estimation. To our knowledge, this is the first application of exponentiated adaptive optimization to a large-scale, biophysically grounded cortical spiking model. Unless stated otherwise, we used $\eta = 5 \times 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-11}$. A detailed derivation is provided in Fraile et al.[98].

### *Multi-objective training loss*

Choosing the training objective is as important as choosing the circuit model itself: if the loss only matches one statistic (for example, firing-rate histograms), the optimizer can satisfy that target through biologically implausible mechanisms, such as excessive population synchrony or unrealistic redistribution of synaptic strength. To avoid such undesirable solutions, the objective must jointly constrain multiple dimensions of activity while remaining numerically well-behaved for gradient-based optimization (positive, smooth, and balanced across terms). The total loss therefore combines multiple activity-matching terms (derived from Neuropixels recordings[20]) with physiological regularizers on membrane voltage and recurrent-weight distributions. The default formulation uses a *core/annulus* readout: the core provides the high-fidelity functional readout, and the annulus provides contextual support to reduce boundary artifacts.

**Neuropixels target curation.** Targets were derived from the Allen Institute's Visual Coding Neuropixels resource[20] after stimulus and unit curation, as described below. We restricted training targets to a subset of the "Brain Observatory 1.1" conditions in the Neuropixels dataset: gray-screen spontaneous activity and full-field drifting gratings at 8 directions ($0° : 315° : 45°$), 80% contrast, 2 Hz temporal frequency, and 0.04 cycles/degree. OSI/DSI targets were derived from the 2 Hz drifting-grating condition.

Cell-type labels in the Neuropixels dataset were aligned to model populations with layer-aware pooling. Excitatory and PV neurons were identified by the waveform (broad spikes: Exc, narrow spikes: PV); their response statistics were taken from only V1. SST and VIP cells were identified in the dataset using optotagging. For SST and VIP, response statistics were pooled across V1 and the other five visual cortical areas (LM, AL, RL, AM, and PM) in the dataset, since V1 counts alone were sparse. L5 excitatory types were merged to match Neuropixels granularity. Units with receptive fields larger than 100° in azimuth or elevation were excluded. OSI/DSI target computation included only neurons with preferred-condition firing rate $\geq 0.5$ Hz, to avoid spuriously high selectivity estimates from ultra-sparse firing. For firing-rate distributions, an explicit 0 Hz sample was appended per population to preserve the silent-neuron fraction in short windows. The curated target set comprised 3,156 neurons from 48 mice.

**Training branches and global loss structure.** Each optimizer update uses two simulated branches of equal duration:

- spontaneous branch (gray screen): $T_{\mathrm{br}} = 500$ ms, $B_{\mathrm{spt}} = 5$ trials,
- evoked branch (drifting gratings): $T_{\mathrm{br}} = 500$ ms, $B_{\mathrm{evk}} = 5$ trials.

The evoked branch uses full-field drifting gratings (80% contrast, 0.04 cycles/degree, 2 Hz) with random orientation and phase across trials. The two branches are computed within the same training step and contribute jointly to the loss.

The total objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{core}} + \lambda_{\text{ann}}\mathcal{L}_{\text{ann}} + \mathcal{L}_{\text{volt}} + \mathcal{L}_{\text{w}}, \tag{27}$$

with

$$\mathcal{L}_{\text{core}} = \mathcal{L}^{\text{spt}}_{\text{rate,core}} + \mathcal{L}^{\text{evk}}_{\text{rate,core}} + \mathcal{L}^{\text{spt}}_{\text{sync,core}} + \mathcal{L}^{\text{evk}}_{\text{sync,core}} + \mathcal{L}^{\text{evk}}_{\text{OSI/DSI,core}}, \tag{28}$$

$$\mathcal{L}_{\text{ann}} = \mathcal{L}^{\text{spt}}_{\text{rate,ann}} + \mathcal{L}^{\text{evk}}_{\text{rate,ann}} + \mathcal{L}^{\text{evk}}_{\text{OSI/DSI,ann}}. \tag{29}$$

Rate and OSI/DSI terms use the full 500-ms window of each branch. Synchrony is computed on the 200–500 ms window. These loss blocks correspond directly to the schematic in Fig. 2a.

**Firing-rate distribution loss.** For each branch $\text{br} \in \{\text{spt}, \text{evk}\}$, we match model firing-rate distributions to empirical Neuropixels distributions using a quantile-Huber loss[107], as in prior RSNN training studies[32,98,108]. Let $\mathcal{C}$ denote the set of cell types, and let $N_c$ denote the number of model neurons of cell type $c$. For each cell type, we first sample a fixed target rate vector of length $N_c$ from the corresponding empirical Neuropixels distribution by inverse-CDF sampling, sort it, and keep it fixed throughout training:

$$r^{*,(\text{br})}_{c,(1)} \le \cdots \le r^{*,(\text{br})}_{c,(N_c)}.$$

Branch-specific model firing rates are computed from the 500 ms window:

$$\hat{r}^{(\text{br})}_i = \frac{1}{B_{\text{br}}\,T_{\text{br}}} \sum_{b=1}^{B_{\text{br}}} \sum_{n=0}^{T_{\text{br}}} S_{b,n,i}, \qquad T_{\text{br}} = 500 \text{ ms}, \tag{30}$$

where $B_{\text{br}}$ is the number of trials in branch br. Within each cell type $c$, model rates are sorted to obtain

$$\hat{r}^{(\text{br})}_{c,(1)} \le \cdots \le \hat{r}^{(\text{br})}_{c,(N_c)}.$$

Using $q = 1, \ldots, N_c$ to index sorted rank positions within cell type $c$, the firing-rate loss is

$$\tilde{\mathcal{L}}^{(\text{br})}_{\text{rate}} = \frac{1}{N} \sum_{c\in\mathcal{C}} \sum_{q=1}^{N_c} \rho^{\kappa}_{\tau_{c,q}} \left( \hat{r}^{(\text{br})}_{c,(q)} - r^{*,(\text{br})}_{c,(q)} \right), \qquad \tau_{c,q} = \frac{q}{N_c}, \quad N = \sum_{c\in\mathcal{C}} N_c \tag{31}$$

with quantile-Huber penalty

$$\rho^{\kappa}_{\tau}(u) = \left| \tau - \mathbb{1}[u \le 0] \right| \begin{cases} \dfrac{u^2}{2\kappa}, & |u| \le \kappa, \\ |u| - \dfrac{\kappa}{2}, & |u| > \kappa. \end{cases} \tag{32}$$

Finally, the branch-specific firing-rate loss is

$$\mathcal{L}^{(\text{br})}_{\text{rate}} = \lambda_{\text{rate}}\, \tilde{\mathcal{L}}^{(\text{br})}_{\text{rate}}. \tag{33}$$

**OSI/DSI loss (crowd surrogate).** Directly optimizing classical single-neuron OSI/DSI definitions inside BPTT is unstable in our training regime because each optimizer update contains only short 500 ms evoked trials and limited directional sampling. Therefore, to obtain a lower-variance differentiable signal, we use a population-level (*crowd*) surrogate that is computed separately within each cell type.

Let $\mathcal{C}$ denote the set of cell types, and let $N_c$ denote the number of model neurons of cell type $c$. For evoked trial $b$, let $\varphi_b$ denote the presented grating direction, and let $\theta_j^{\text{pref}}$ be the assigned preferred direction of neuron $j$. We define

$$\Delta\theta_{b,j} = \varphi_b - \theta_j^{\text{pref}}, \qquad r_{b,j} = \frac{1}{T_{\text{br}}} \sum_{n=0}^{T_{\text{br}}} S_{b,n,j}, \qquad T_{\text{br}} = 500 \text{ ms},$$

where $r_{b,j}$ is the evoked firing rate of neuron $j$ in trial $b$. We then normalize rates by an evoked exponential moving average (EMA) with floor $R_{\min} = 0.5$ Hz:

$$\tilde{r}_{b,j} = \frac{r_{b,j}}{\max(\text{EMA}_{\beta,j}, R_{\min})}, \qquad \beta_{\text{EMA}} = 0.95,$$

where $\tilde{r}_{b,j}$ denotes the normalized firing rate.

Within each model cell type, preferred directions are assigned uniformly by construction. Let $\Theta$ denote a random preferred direction within cell type $c$. We then use the population-averaging approximation

$$\frac{1}{N_c} \sum_{j \in \mathcal{N}_c} \tilde{r}_{b,j}\, h(\Delta\theta_{b,j}) \;\approx\; \mathbb{E}_\Theta\left[\tilde{r}_b(\Theta)\, h(\varphi_b - \Theta)\right],$$

where $\mathbb{E}_\Theta[\cdot]$ denotes expectation over preferred directions $\Theta$ within cell type $c$, and $h$ is the angular kernel of interest. In particular, we use $h(\delta) = \cos(2\delta)$ for orientation selectivity and $h(\delta) = \cos(\delta)$ for direction selectivity. This approximation replaces averaging across stimulus directions for a single neuron with averaging across neurons of the same cell type in a single trial.

Using these definitions, the crowd surrogate for trial $b$ and cell type $c$ is

$$D_{b,c} = \max\left(\frac{1}{N_c} \sum_{j \in \mathcal{N}_c} \tilde{r}_{b,j},\, R_{\min}\right), \tag{34}$$

$$\widehat{\text{OSI}}_{b,c} = \frac{\frac{1}{N_c} \sum_{j \in \mathcal{N}_c} \tilde{r}_{b,j} \cos(2\Delta\theta_{b,j})}{D_{b,c}}, \tag{35}$$

$$\widehat{\text{DSI}}_{b,c} = \frac{\frac{1}{N_c} \sum_{j \in \mathcal{N}_c} \tilde{r}_{b,j} \cos(\Delta\theta_{b,j})}{D_{b,c}}. \tag{36}$$

We then average these trial-level quantities over the evoked batch:

$$\widehat{\text{OSI}}_c = \frac{1}{B_{\text{evk}}} \sum_{b=1}^{B_{\text{evk}}} \widehat{\text{OSI}}_{b,c}, \qquad \widehat{\text{DSI}}_c = \frac{1}{B_{\text{evk}}} \sum_{b=1}^{B_{\text{evk}}} \widehat{\text{DSI}}_{b,c}, \tag{37}$$

where $B_{\text{evk}}$ is the number of evoked trials in the batch.

Finally, we match the resulting cell-type means to the Neuropixels targets with a population-weighted mean-squared error (MSE):

$$\tilde{\mathcal{L}}_{\text{OSI/DSI}} = \frac{\sum_{c \in \mathcal{C}} N_c \left[(\widehat{\text{OSI}}_c - \text{OSI}_c^{\text{exp}})^2 + (\widehat{\text{DSI}}_c - \text{DSI}_c^{\text{exp}})^2\right]}{\sum_{c \in \mathcal{C}} N_c}, \tag{38}$$

and

$$\mathcal{L}^{\text{evk}}_{\text{OSI/DSI}} = \lambda_{\text{OSI/DSI}}\, \tilde{\mathcal{L}}_{\text{OSI/DSI}}. \tag{39}$$

Synchronization (multi-scale Fano factor) loss. Firing-rate distributions alone do not constrain shared variability: a network can reproduce correct rates while exhibiting unrealistic population synchronization. We therefore include a synchrony term based on multi-scale Fano factors (variance-to-mean spike-count ratio) computed from excitatory populations in the core. Neuropixels references were precomputed from recordings (separated by the animal state: running vs. not) and, on average across animals, display the expected cortical signatures across timescales: supra-Poisson variability, monotonic growth with bin width, and stronger fluctuations in evoked drifting-grating epochs than in spontaneous gray-screen periods (Extended Data Fig. 12)[20].

Each loss evaluation samples $N_{\text{pool}}$ random excitatory pools (default $N_{\text{pool}} = 500$); pool sizes are drawn from a clipped Gaussian $n_{\text{pool}} \sim \mathcal{N}(70, 30^2)$, clipped to $[15, N_E]$ to match experimental sample sizes. Spikes are summed within each pool and binned at log-spaced bin widths $\mathcal{W}_{\text{bin}} \subset [1, 150]$ ms over the 200–500 ms analysis window, retaining only bin widths that yield at least two bins in that window. For sampled pool $p$ and bin width $w \in \mathcal{W}_{\text{bin}}$,

$$F_w^{(p)} = \frac{\text{Var}(n_w^{(p)})}{\mathbb{E}(n_w^{(p)})}, \qquad \hat{F}_w = \frac{1}{N_{\text{pool}}} \sum_{p=1}^{N_{\text{pool}}} F_w^{(p)}. \tag{40}$$

Pools are drawn in quasi-without-replacement shuffled blocks to increase neuron coverage per step and reduce Monte Carlo noise in the gradient signal. The branch-specific synchrony loss is

$$\mathcal{L}^{(\cdot)}_{\text{sync}} = \lambda_{\text{sync}} \frac{1}{|\mathcal{W}_{\text{bin}}|} \sum_{w \in \mathcal{W}_{\text{bin}}} \left(\hat{F}_w - F_w^{\text{exp}}\right)^2, \qquad (\cdot) \in \{\text{spt}, \text{evk}\}. \tag{41}$$

Voltage regularization. To discourage drift into biologically implausible hidden-state regimes during BPTT, we regularize membrane voltage in normalized coordinates ($\tilde{v} = 0$ reset, $\tilde{v} = 1$ threshold):

$$\tilde{\mathcal{L}}_{\text{volt}} = \frac{1}{B_{\text{tot}}\, T_{\text{tot}}\, N} \sum_{b,n,j} (\tilde{v}_{b,n,j} - 1)^2, \qquad \mathcal{L}_{\text{volt}} = \lambda_{\text{volt}} \tilde{\mathcal{L}}_{\text{volt}}, \tag{42}$$

where $B_{\text{tot}} = 10$ is the total number of trials across the two branches in one update and $T_{\text{tot}}$ is the corresponding number of simulated steps included in this regularizer. This term acts as a soft guardrail; functional losses remain responsible for matching activity statistics[32,98,108].

Recurrent-weight regularization (EMD / Wasserstein-1). A major goal of our work was to train the circuit models while preserving experimentally observed values of synaptic weights. These are not known for individual connections, but the Synaptic Physiology dataset[18] provides distributions per connection type (defined based on the pre- and postsynaptic cell type). Our implementation of this constraint took advantage of the fact that we initialized network weights at model construction from these experimentally determined distributions. Therefore, we constrain cell type-wise distributional drift during optimization using Earth-Mover distance (Wasserstein-1) between current and initial recurrent weights. For recurrent connection class $m$ (out of $M$ classes), let $n_m$ be the number of synapses and let $w_{m,(q)}$ and $w^0_{m,(q)}$ denote sorted current and initial normalized weights (same sign class):

$$\tilde{\mathcal{L}}_{\text{w}} = \frac{1}{M} \sum_{m=1}^{M} \left( \frac{1}{n_m} \sum_{q=1}^{n_m} \left| w_{m,(q)} - w^0_{m,(q)} \right| \right), \qquad \mathcal{L}_{\text{w}} = \lambda_{\text{w}} \tilde{\mathcal{L}}_{\text{w}}. \tag{43}$$

This regularizer penalizes full distributional drift and therefore preserves the experimentally shaped heavy-tailed synaptic landscape while still allowing functionally useful adaptation.

**Loss coefficients used for reported training runs.** Unless stated otherwise, the loss weights were

$$\lambda_{\text{rate}} = 10^4, \quad \lambda_{\text{OSI/DSI}} = 20, \quad \lambda_{\text{sync}} = 1.5, \quad \lambda_{\text{volt}} = 1, \quad \lambda_{\text{w}} = 10, \quad \lambda_{\text{ann}} = 0.1.$$

Additional defaults were: quantile-Huber threshold $\kappa = 0.002$, evoked-rate EMA decay $\beta_{\text{EMA}} = 0.95$, and $N_{\text{pool}} = 500$ sampled excitatory pools for the synchrony loss.

### *TensorFlow implementation for efficient training*

The simulator is differentiable by construction, but practical BPTT at the scale of our V1 models is limited by memory traffic and sparse recurrent computation. At each 1 ms step, the simulator updates membrane, ASC, PSC, and delayed-spike states for a large recurrent network while preserving the trajectories needed for gradient propagation. In a naive implementation, this leads to $\mathcal{O}(T \cdot N)$ activation storage and repeated large recurrent products, which rapidly becomes prohibitive in VRAM and wall-clock time. The implementation therefore combines sparsity-aware kernels, compact state representations, and graph-level optimizations to make the biology-constrained fitting stage in Fig. 2 computationally practical.

**Exploiting structural and activity sparsity.** The recurrent pathway is sparse in two complementary senses: (i) *structural sparsity* in connectivity (about $2.35 \times 10^7$ recurrent connections across $\sim 6.7 \times 10^4$ neurons, that is, $\sim 0.5\%$ occupancy of all possible directed pairs), and (ii) *activity sparsity* in time (typical firing rates 3–5 Hz, so only $\sim 0.3$–$0.5\%$ of neurons spike in a 1 ms step). We therefore compute recurrent synaptic drive with an event-driven sparse kernel rather than a dense recurrent matrix multiply.

At each step, only active presynaptic events are expanded through Compressed Sparse Row (CSR) format lookup tables[109]; their contributions are accumulated by segmented reductions over postsynaptic targets and synaptic basis channels (Algorithm 1). Axonal delays are handled through an augmented delay buffer: each presynaptic neuron is expanded into $D_{\max}$ delay lanes, and synapses with delay $d$ index the corresponding lane directly. This preserves delayed transmission while keeping first-order per-step state updates.

The presynaptic lookup table is built on CPU (NumPy/Numba), where irregular indexing/bucketing is efficient, while per-step event expansion and segmented reductions run on GPU. Batches are processed jointly using flattened segment IDs[93]. TensorFlow does not provide a native sparse–sparse matmul path for this update, so we use (i) event-driven expansion + segmented sums in the forward path and (ii) sparse-to-dense products in the backward path to obtain exact Jacobian-vector products under autodiff.

**Gradient checkpointing.** All reported training runs enabled gradient checkpointing on the RSNN forward path (`tf.recompute_grad`). This reduces BPTT activation memory by recomputing selected intermediates during the backward pass instead of storing them, at the cost of additional computation[66].

**Mixed precision.** We used mixed precision for training (`float16` compute dtype, `float32` master variables) and `float32` for testing/evaluation. The optimizer was wrapped in `LossScaleOptimizer` to reduce `float16` gradient underflow[65]. Most recurrent state tensors and per-step arithmetic run in compute dtype, while trainable weight variables are maintained in `float32` and mirrored into compute shadows after each optimizer update.

**Algorithm 1** Event-driven recurrent synaptic current and custom gradient calculation

**Require:** Delayed spikes $\mathbf{z} \in \mathbb{R}^{B_{\text{alg}} \times N_{\text{pre,delay}}}$, with $N_{\text{pre,delay}} = D_{\max} N_{\text{pre}}$; batch size $B_{\text{alg}}$; number of postsynaptic neurons $N_{\text{post}}$; sparse edges (`synapse_indices`, $\mathbf{w}$, `syn_ids`); CSR-like map `pre_ind_table`; synaptic basis table $\mathbf{\Phi} \in \mathbb{R}^{S \times N_{\text{basis}}}$

**Forward path (event-driven sparse accumulation)**

1: `non_zero` $\leftarrow$ `tf.where`$(\mathbf{z} > 0)$ ▷ active events $(b, \text{pre})$
2: `batch_id` $\leftarrow$ `non_zero`$[:, 0]$, `pre_id` $\leftarrow$ `non_zero`$[:, 1]$
3: `selected_rows` $\leftarrow$ `tf.gather`(`pre_ind_table`, `pre_id`)
4: `active_syn` $\leftarrow$ `selected_rows.flat_values`, `deg` $\leftarrow$ `selected_rows.row_lengths`()
5: `new_indices` $\leftarrow$ `tf.gather`(`synapse_indices`, `active_syn`), `post_id` $\leftarrow$ `new_indices`$[:, 0]$
6: `new_weights` $\leftarrow$ `tf.gather`($\mathbf{w}$, `active_syn`), `new_syn_ids` $\leftarrow$ `tf.gather`(`syn_ids`, `active_syn`)
7: `basis_active` $\leftarrow$ `tf.gather`($\mathbf{\Phi}$, `new_syn_ids`, `axis` = 0)
8: `segment_id` $\leftarrow$ `tf.repeat`(`batch_id`, `deg`) $\cdot N_{\text{post}}$ + `post_id`
9: $\mathbf{I}_{\text{rec}} \leftarrow$ `tf.math.unsorted_segment_sum`$\Big($`new_weights`$[:, \mathit{None}] \cdot$ `basis_active`, `segment_id`, $B_{\text{alg}} N_{\text{post}}\Big)$

**Backward path A (gradient w.r.t. presynaptic spikes)**

10: $\partial \mathcal{L} / \partial \mathbf{z} \leftarrow \mathbf{0}$
11: **for** $b = 0, \ldots, N_{\text{basis}} - 1$ **do**
12: $\quad \mathbf{w}^{(b)} \leftarrow \mathbf{w} \cdot$ `tf.gather`$(\mathbf{\Phi}[:, b]$, `syn_ids`, `axis` = 0)
13: $\quad \mathbf{W}^{(b)} \leftarrow$ `tf.sparse.SparseTensor`(`synapse_indices`, $\mathbf{w}^{(b)}$, $[N_{\text{post}}, N_{\text{pre,delay}}]$)
14: $\quad \partial \mathcal{L} / \partial \mathbf{z}$ += `tf.sparse.sparse_dense_matmul`$\Big(\partial \mathcal{L} / \partial \mathbf{I}_{\text{rec}}[:, :, b], \mathbf{W}^{(b)}\Big)$
15: **end for**

**Backward path B (gradient w.r.t. recurrent weights)**

16: `dactive` $\leftarrow$ `tf.gather`$\Big(\partial \mathcal{L} / \partial \mathbf{I}_{\text{rec}}$, `segment_id`$\Big)$
17: $(\partial \mathcal{L} / \partial \mathbf{w})_{\text{active}} \leftarrow$ `tf.einsum`("cr,cr->c", `dactive`, `basis_active`)
18: $\partial \mathcal{L} / \partial \mathbf{w} \leftarrow$ `tf.math.unsorted_segment_sum`$\Big((\partial \mathcal{L} / \partial \mathbf{w})_{\text{active}}$, `active_syn`, $E\Big)$

Graph mode and selective JIT compilation. Training and evaluation loops were implemented in TensorFlow graph mode (`@tf.function`)[106]. JIT compilation (`jit_compile=True`) was applied selectively to stable tensor-heavy components (e.g., stimulus construction and several loss/regularizer kernels). The recurrent sparse/ragged custom-gradient path was kept without explicit JIT compilation, and global XLA autoclustering was not enabled, because this path was not consistently robust under our sparse custom operators.

Global alpha-basis approximation for memory-efficient BPTT. The receptor-specific double-alpha dynamics in Eq. (12) preserve synaptic-kinetic diversity, but their naive receptor-wise implementation is memory-intensive under BPTT because all PSC states must be retained (or recomputed) across time. In the Results (Fig. 2e,f), this appears as the final memory/compression component of the training framework.

A naive receptor-wise implementation stores four PSC states per receptor channel and neuron:

$$\{I_{j,r}^{(f)}, C_{j,r}^{(f)}, I_{j,r}^{(s)}, C_{j,r}^{(s)}\}.$$

Across the fitted synaptic physiology dataset, we obtained 121 distinct receptor families; in simulation this maps to up to 11 receptor channels per neuron, yielding 44 PSC states per neuron in the naive representation. BPTT memory therefore prohibitively scales as

$$\mathcal{O}\Big(4\, N_{\text{rec,ch}}\, N\, T\Big), \tag{44}$$

where $N_{\text{rec,ch}}$ is the number of receptor channels per neuron (up to 11), $N$ is neuron count, and $T$ is sequence length in steps. To reduce this cost, we approximate each fitted double-alpha kernel with a

shared basis of $N_{\text{basis}}$ single-alpha kernels with global time constants $\{\tau_b\}_{b=1}^{N_{\text{basis}}}$. For receptor family $r$,

$$I_{j,r}^{\text{syn}}[n] \approx \sum_{b=1}^{N_{\text{basis}}} \phi_{r,b}\, I_j^{(b)}[n], \tag{45}$$

where $I_j^{(b)}[n]$ are basis PSC states shared across receptor families and $\phi_{r,b}$ are family-specific coefficients. The basis coefficients and time constants were fit by least squares across all fitted receptor families. The error distribution in Fig. 2f decreases steeply with basis size. We selected $N_{\text{basis}} = 4$, for which all fitted receptor families achieved MSE $< 10^{-3}$ over 30 ms of kernel evolution and integrated charge transfer error $< 1.7\%$. The selected basis time constants were

$$\tau_1 = 0.76 \text{ ms}, \quad \tau_2 = 1.61 \text{ ms}, \quad \tau_3 = 3.41 \text{ ms}, \quad \tau_4 = 7.24 \text{ ms}.$$

Because basis states are shared across receptor families, each neuron stores only $2N_{\text{basis}}$ PSC states ($I_j^{(b)}$ and $C_j^{(b)}$), i.e. 8 states for $N_{\text{basis}} = 4$, instead of 44 in the receptor-wise implementation. PSC memory under BPTT then scales as

$$\mathcal{O}(2\, N_{\text{basis}}\, N\, T)\,. \tag{46}$$

For a 200,000-neuron simulation of 2 s duration at $\Delta t = 1$ ms ($T = 2000$) in `float32`, this corresponds to $\sim$ 11.9 GB for $N_{\text{basis}} = 4$, compared with $\sim$ 66 GB for the naive 11-channel receptor-wise double-alpha representation (5.5$\times$ reduction). This reduction is a key enabler of single-GPU training at circuit scale while preserving receptor-dependent PSC diversity.

### *Training pipeline and reporting conventions*

Gray-screen initialization. At the start of each training epoch, the network state was re-initialized from a spontaneous (gray-screen) state. Specifically, we simulated 500 ms of spontaneous LGN drive starting from the zero state and used the resulting RSNN state as the epoch start state. This reduces state drift across long training runs and improves consistency of recurrent initial conditions.

Per-update simulation protocol. In each optimizer update, gradients for the combined loss (Eq. (27)) across the spontaneous and evoked branches were computed by BPTT. Only recurrent and background synapses were optimized; LGN→V1 weights were fixed throughout.

Typical training configuration. Unless otherwise specified, reported training runs used:

$$\eta = 5 \times 10^{-3}, \quad 75 \text{ epochs}, \quad 25 \text{ optimizer updates/epoch}, \quad T_{\text{br}} = 500 \text{ ms}, \quad B_{\text{spt}} = B_{\text{evk}} = 5.$$

The learning rate was selected by logarithmic grid search over $[10^{-4}, 10^{-1}]$. Batch size was chosen to maximize GPU occupancy without exceeding available VRAM.

Training-curve metrics shown in Fig. 2b. The absolute-loss panel reports epoch means of validation metrics for each component $\{\mathcal{L}_{\text{rate}}, \mathcal{L}_{\text{OSI/DSI}}, \mathcal{L}_{\text{sync}}, \mathcal{L}_{\text{volt}}, \mathcal{L}_{\text{w}}\}$ and their total. Relative-loss traces are normalized by the first validation epoch:

$$\mathcal{L}_q^{\text{rel}}(e) = \frac{\mathcal{L}_q(e)}{\mathcal{L}_q(1)},$$

and plotted on a log scale to compare convergence rates across terms with different magnitudes.

### *Computational scaling benchmarks*

Reference training runtime (reported model). For the reported V1-scale models, one optimizer update (1 simulated second total: 500 ms spontaneous + 500 ms evoked) required $\sim 17.6$ s wall-clock time on a single NVIDIA A100 GPU, including stimulus generation, forward simulation, gradient computation, and parameter update. During training, the network operated at an average firing rate of $\sim 4$ Hz and used $\sim 37$ GB of GPU VRAM. Full optimization required $\sim 10$ h on one A100 GPU plus 4 Intel Xeon Gold 6330N CPU cores.

Scaling measurements used in Fig. 2d. To characterize scalability, we measured step time and VRAM as a function of network size under:

- training benchmark: $T = 1000$ ms total simulated time per update, batch size $B = 1$,
- testing benchmark: $T = 4000$ ms, batch size $B = 10$,

on one NVIDIA RTX PRO 6000 GPU and 16 AMD EPYC 9754 CPU cores. The testing configuration remained below the real-time reference line for networks < $100,000$ neurons (faster-than-biological-time inference).

Two regimes are visible (Fig. 2d): (i) a fixed-overhead regime for small networks ($N \leq 20$k), where framework/kernel-launch overhead dominates, and (ii) an approximately linear regime beyond $\sim 20,000$ neurons. For small networks, GPU utilization is low; increasing batch size improves effective throughput. To demonstrate this, we additionally benchmarked batch-size scaling on a single NVIDIA L40S GPU (48 GB VRAM) with 8 AMD EPYC 9754 CPU cores (Extended Data Fig. 3), observing sublinear growth in both step time and VRAM with increasing batch size.

Linear-regime extrapolation (single-GPU capacity; first-order estimate). To estimate approximate single-GPU capacity in the linear regime ($N \geq 20,000$), we fit

$$\text{StepTime}_{\text{train}}\ [\text{s}] \approx 7.77 \times 10^{-5} N + 3.14, \qquad \text{VRAM}_{\text{train}}\ [\text{GB}] \approx 2.12 \times 10^{-4} N + 0.36.$$

These fits provide first-order extrapolations (Table 1), but absolute numbers are hardware- and software-stack-dependent, and secondary bottlenecks (e.g., host–device transfer, kernel-launch overheads) may become important at very large $N$, where model partitioning across multiple GPUs may be preferable[93].

| GPU | VRAM | Max $N$ that fits | Approximate step time |
|---|---|---|---|
| NVIDIA L40S | 48 GB | $\sim 225,000$ | 21 s |
| NVIDIA H100 | 80 GB | $\sim 375,000$ | 32 s |
| NVIDIA H200 | 141 GB | $\sim 663,000$ | 55 s |

**Table 1:** Single-GPU capacity extrapolated from the linear fit in Fig. 2d. Coefficients were estimated from benchmarks of this implementation on an NVIDIA RTX PRO 6000 (96 GB) under the same training protocol ($T = 1000$ ms, $B = 1$). Absolute step times depend on hardware generation, memory bandwidth, drivers, CUDA, and TensorFlow versions.

### *Stimulus And Data Generation Methods*

Drifting gratings. Training gratings were generated directly in TensorFlow and transformed by an explicit LGN model. For spatial position $(x, y)$ and time $t$,

$$S(x, y, t) = C \sin\left[2\pi\left(\text{cpd}(x\cos\theta + y\sin\theta) + f_t t\right) + \phi\right], \tag{47}$$

with contrast $C$, orientation $\theta$, phase $\phi$, temporal frequency $f_t$, and spatial frequency cpd. In the default generator, orientation and phase are sampled per trial (or cycled deterministically for regular-angle evaluation), then padded by pre- and post-gray periods.

**LGN transformation and spike sampling.** Stimuli are passed through spatial and temporal LGN filters to obtain firing rates in Hz. LGN spikes are sampled as Bernoulli events with

$$p_t = 1 - \exp\left(-r_t/1000\right), \tag{48}$$

for 1 ms bins. Equivalent stateless generation is used for spontaneous gray-screen LGN activity. Data pipelines are configured as deterministic with per-pipeline folded seeds to ensure reproducibility across distributed workers.

### *Evaluation And Analysis Methods*

Unless otherwise specified, all evaluation metrics, statistical analyses, and summary distributions (e.g., aggregated box plots) were computed across an ensemble of 10 independently trained network models initialized with different random seeds. Validation used 2 s gray + 2 s gratings across 8 directions with 10 repeats/direction; natural images and contrast sweeps were held out for out-of-distribution evaluation[19,20,98].

**Drifting-grating response metrics.** Model responses were evaluated with drifting gratings across eight directions (0° to 315° in 45° increments), with repeated trials per angle. For each neuron, preferred-direction response, OSI, and DSI were estimated from trial-averaged activity over the evoked window; spontaneous rates were estimated from pre-stimulus gray periods.

**Comparison to experimental distributions.** Population-level statistics (spontaneous rate, evoked rate, OSI, DSI, and synchronization/Fano factor profiles) were compared against Neuropixels-derived references using matched cell-type groupings. The same grouping definitions used in the loss functions were applied for evaluation summaries. To quantify the agreement between model and experimental distributions, we computed a similarity score defined as $1 - D_{\mathrm{KS}}$, where $D_{\mathrm{KS}}$ is the Kolmogorov-Smirnov statistic between the two distributions.

**State and activity diagnostics.** During training and evaluation we tracked component losses, mean firing rates, raster summaries, and optional membrane-voltage traces. Power-spectrum and Fano factor analyses were computed on designated spontaneous and evoked windows for core populations. For Fig. 2c, representative rasters were rendered from checkpoints at epochs $0, 25, 50, 75$ using the same validation stimulus pair, with neurons grouped by layer/class and plotted over the analysis core.

**Orientation and direction selectivity indices.** For responses to drifting gratings, the Orientation Selectivity Index (OSI) and Direction Selectivity Index (DSI) were computed using vector summation. Note that these indices are evaluated for individual neurons, which is distinct from the group-level OSI defined earlier for the loss function. Let $R(\theta)$ be the trial-averaged firing rate of a given neuron in response to a grating drifting at angle $\theta$ (in radians). The DSI is defined as:

$$\mathrm{DSI} = \frac{\left|\sum_\theta R(\theta) e^{i\theta}\right|}{\sum_\theta R(\theta)}, \tag{49}$$

and the OSI is calculated using double-angle vectors:

$$\mathrm{OSI} = \frac{\left|\sum_\theta R(\theta) e^{2i\theta}\right|}{\sum_\theta R(\theta)}. \tag{50}$$

Image selectivity (response sparsity). Selectivity to natural images was quantified using response sparsity across the 118 Brain Observatory images[19]. For a given neuron, let $r_i$ be the trial-averaged response to the $i$-th image and $N = 118$ the total number of images. The image selectivity $S$ is defined as:

$$S = \frac{1 - \frac{\left(\frac{1}{N}\sum_{i=1}^{N} r_i\right)^2}{\frac{1}{N}\sum_{i=1}^{N} r_i^2}}{1 - \frac{1}{N}}. \tag{51}$$

A value of 0 indicates equal responses to all images, while a value of 1 indicates a response to only a single image.

Contrast response simulations. To evaluate contrast-dependent modulation, trained networks were presented with drifting gratings at varying contrast levels ($C \in \{0.05, 0.1, 0.2, 0.4, 0.6, 0.8\}$) while keeping spatial and temporal frequency fixed. Firing rates, OSI, and DSI were computed at each contrast level per neuron and then aggregated by cell type. These simulations were not included in training (which used only high-contrast gratings at $C = 0.8$) and thus serve as an out-of-distribution generalization test.

Network stability and participation ratio. To monitor network stability during optimization and detect pathological regimes—such as runaway excitation, extreme suppression, or transient synchronized events—we tracked the mean population firing rate and the maximum participation ratio at the end of each epoch (Extended Data Fig. 2). Both metrics were computed from 5 independent 500 ms trials for both spontaneous and evoked conditions.

The participation ratio is defined as the fraction of active neurons within a discrete time window. To avoid obscuring brief bursts of activity, we analyzed the maximum participation ratio across non-overlapping time bins of width $\Delta t \in \{10, 50, 100\}$ ms. For bin $b$, the fraction of active neurons is:

$$f_b = \frac{1}{N}\sum_{i=1}^{N} \Theta\left(\sum_{t \in b} s_i(t)\right), \tag{52}$$

where $N$ is the total number of neurons, $s_i(t) \in \{0, 1\}$ is the binarized spike train of neuron $i$ at 1 ms resolution, and $\Theta(x)$ is the Heaviside step function ($\Theta(x) = 1$ for $x > 0$, and 0 otherwise). We report the mean and standard deviation of the maximum $f_b$ over all bins, computed across the 5 trials.

Most exciting inputs (MEIs). To probe the features driving individual model neurons, we generated most exciting inputs (MEIs) following the approach and CNN architecture of Walker et al.[71]. For each neuron, a convolutional neural network was first trained to predict the neuron's trial-averaged responses to natural images. The CNN was then used as a differentiable surrogate, and gradient ascent in the stimulus space was performed to discover synthetic images that maximally excite each neuron.

ImageNet stimuli consisted of 5000 images presented once and 100 images repeated 10 times, with each image shown for 250 ms followed by a 250 ms gray screen inter-stimulus interval, matching the protocol of Walker et al.[71]. Response reliability was quantified by an oracle-correlation metric computed from the repeated ImageNet presentations.

For a given neuron $i$, let $r_{i,g,k}$ denote its response (spike count) to the $k$-th presentation ($k \in \{1, \ldots, K\}$) of the $g$-th unique repeated image ($g \in \{1, \ldots, G\}$), where $K = 10$. We adopted a leave-one-out approach to compute the mean response to the same image, excluding the current trial. For

each trial $k$, the leave-one-out average response is given by:

$$\bar{r}_{i,g,\backslash k} = \frac{1}{K-1} \sum_{j \neq k} r_{i,g,j}. \tag{53}$$

The single-trial responses and their corresponding leave-one-out averages are then vectorized across all repeated images and trials, yielding two vectors $\mathbf{r}_i$ and $\bar{\mathbf{r}}_i$, each of length $G \times K$. The oracle correlation $\rho_i$ for neuron $i$ is defined as the Pearson correlation coefficient between these two vectors:

$$\rho_i = \frac{\mathrm{Cov}(\mathbf{r}_i, \bar{\mathbf{r}}_i)}{\sigma_{\mathbf{r}_i}\, \sigma_{\bar{\mathbf{r}}_i}}, \tag{54}$$

where $\mathrm{Cov}(\cdot,\cdot)$ denotes the sample covariance and $\sigma$ denotes the sample standard deviation.

For MEI generation, we selected a total of 2000 neurons among those with the highest 30% oracle correlation within each cell type, pooled across all 10 trained model networks.

To characterize the spatial structure of MEIs, we computed the two-dimensional spatial power spectrum for each MEI and summarized each neuron's spectral profile by the ratio of high-frequency power (0.08–0.16 cycles per degree) to low-frequency power (0.02–0.04 cycles per degree).

### *Synaptic Weight Analysis*

To explore the emergence of functional microcircuits following optimization, we systematically analyzed the relationship between trained synaptic weights and the functional tuning properties of pre- and post-synaptic neuron pairs. Note that the presence or absence of connections is not modified by our training procedure; only synaptic weights are adjusted. The analyses were conducted across the $19 \times 19$ cell-type connectivity matrix, as well as in aggregated form (layer segregated excitatory types and layer aggregated inhibitory types).

**Response correlation analysis.** For each connected neuron pair in the network, we calculated the Pearson correlation coefficient of their trial-averaged responses to the 118 Brain Observatory natural images[19]. Synaptic weights were then grouped by these response correlations into three functional domains: anti-correlated ($[-1, -0.5]$), uncorrelated (or very weakly correlated) ($[-0.25, 0.25]$), and highly correlated ($[0.5, 1.0]$). To quantify the strength of like-to-like connectivity, we computed the normalized weight differences relative to the "uncorrelated" baseline:

$$\Delta_{\mathrm{like}} = \frac{\mu(I_{\mathrm{like}}) - \mu(I_{\mathrm{mid}})}{\mu(I_{\mathrm{mid}})}, \qquad \Delta_{\mathrm{anti}} = \frac{\mu(I_{\mathrm{mid}}) - \mu(I_{\mathrm{anti}})}{\mu(I_{\mathrm{mid}})}, \tag{55}$$

where $\mu(I_{\mathrm{like}})$, $\mu(I_{\mathrm{anti}})$, and $\mu(I_{\mathrm{mid}})$ denote the mean synaptic weight in the like-correlated, anti-correlated, and uncorrelated domains, respectively. By this definition, positive values for both $\Delta_{\mathrm{like}}$ and $\Delta_{\mathrm{anti}}$ indicate a like-to-like connectivity structure (stronger connections for similarly responding neurons and weaker connections for oppositely responding neurons, relative to the uncorrelated baseline), whereas negative values indicate the opposite—that is, an anti-like-to-like structure. (Note that the above is true for both excitatory and inhibitory connections. The latter have negative weights. Thus, e.g., stronger inhibition for highly correlated vs. uncorrelated neurons would result in dividing a negative $\mu(I_{\mathrm{like}}) - \mu(I_{\mathrm{mid}})$ by a negative $\mu(I_{\mathrm{mid}})$, giving a positive result for $\Delta_{\mathrm{like}}$ and indicating like-to-like inhibition.)

For inhibitory cell types (PV, SST, VIP), the analysis was restricted to within-layer connections (i.e., inhibitory neurons were paired only with excitatory neurons from the corresponding cortical layer) to avoid conflating laminar targeting preferences with functional tuning relationships. An

exception was made for L1 Inh neurons, which were allowed to pair with excitatory neurons from all layers, as L1 contains no excitatory cells. When neuron pairs with sufficient correlation ($< -0.5$ or $> 0.5$) were not found for a given cell-type pair, the corresponding $\Delta$ metrics were not calculated.

Preferred-direction difference analysis. For drifting-grating responses, the absolute difference in preferred direction ($\delta\theta$, ranging from $0°$ to $180°$) was computed between pre- and post-synaptic neurons for each connection. Synaptic weights were binned by $\delta\theta$, and the relationship was quantified using a cosine-series fit of the form:

$$w(\delta\theta) = a\cos(\delta\theta) + b\cos(2 \cdot \delta\theta) + c, \tag{56}$$

where $a$ captures direction-specific like-to-like structure, $b$ captures orientation-specific like-to-like structure, and $c$ is the baseline offset. The normalized effect sizes $a/c$ and $b/c$ are reported, with positive values indicating like-to-like and negative values indicating anti-like-to-like relations (again, this applies to both excitatory and inhibitory connections).

Comparison to electron microscopy data. To directly compare model predictions to anatomy, we used the V1DD electron microscopy dataset[35], which links functional tuning differences (including responses to drifting gratings) to the size of the post-synaptic density (PSD) as a proxy for synaptic strength. Because the V1DD dataset provides a relatively small and fixed set of reconstructed connections (with functional tuning data available only for excitatory neurons), we employed a Monte Carlo resampling approach to match sample sizes between model and experiment. For each pathway, we repeatedly redrew the same number of connections as available in the V1DD data from our model and recomputed the tuning-dependent statistic and its associated $p$-values. This procedure was repeated 100 times to yield a distribution of model $p$-values under matched sample sizes, which was compared against the single $p$-value obtained from the experimental measurements.

### *Outgoing Weight Cohort Analysis*

To investigate functional heterogeneity within projection pathways, neurons from each of the 19 cell types were grouped into three cohorts—low (bottom tertile), mid (middle tertile), and high (top tertile)—based on their total absolute outgoing synaptic weight summed across all postsynaptic targets in the entire network (both core and periphery). Cohort assignment was performed independently within each cell type (e.g., L4_Exc, L5_ET, L6_VIP), ensuring that each cell type contributed approximately one-third of its neurons to each cohort. To avoid boundary artifacts, the selection pool for the cohorts was strictly limited to source neurons residing within the 200 μm analysis core ($\sqrt{x^2 + z^2} < 200$ μm).

Following grouping, the firing rates and OSI were compared between 10 instances of the trained and untrained networks. The target preferences of each cohort were characterized by computing the proportional distribution of outgoing synaptic weight across postsynaptic cell types (Exc, PV, SST, VIP).

We focused on characterizing the relation between functional properties and outgoing weights rather than incoming weights, because we did not observe a strong change between the total incoming synaptic weights and the functional metrics that is shaped by the training. An analogous cohort analysis based on total incoming synaptic weights was also performed and is reported in the Extended Data Figure 9.

### *Simulated Perturbations*

To establish causal links between specific network sub-populations and large-scale network dynamics, we performed targeted silencing simulations. Silencing was implemented by injecting a constant

hyperpolarizing current (−1000 pA) into the targeted neurons for the entire duration of the visual stimulus presentation.

Cohort-specific silencing. To dissect the distinct roles of high-impact versus low-impact projection streams, we selectively silenced either the high-outgoing-weight or low-outgoing-weight cohorts of excitatory or inhibitory populations within the core network. Cohort definitions followed the same tertile-based grouping described above. For each suppression condition, the full stimulus set of DGs with 8 directions × 10 trials was presented.

Inhibitory cell type-specific cohort suppression. We also evaluated the distinct, network-wide regulatory impacts of specific inhibitory interneuron types. We selectively silenced either the high-outgoing-weight or low-outgoing-weight cohorts of one specific inhibitory type (PV, SST, or VIP) at a time.

Perturbation analysis. The functional consequences of these perturbations were evaluated by computing the percentage change in firing rate and OSI relative to the unperturbed baseline. Neurons directly targeted for suppression were excluded from the post-perturbation metric calculations. This exclusion ensured that the resulting heatmaps represent indirect, trans-synaptic, network-wide effects rather than the trivial suppression of the targeted neurons. The unperturbed baseline was computed from the same non-targeted population to ensure a matched comparison.

Synaptic weight distribution constraints. In the constrained run, synaptic weights were initialized by sampling from the experimentally derived distribution as described above, and training was performed with the full multi-objective loss including the weight regularization term ($\mathcal{L}_w$) based on the Earth-Mover's Distance. In the unconstrained run, we initialized weights uniformly (all excitatory connections at the same value and all inhibitory connections at the same value), and we removed the weight regularization term ($\mathcal{L}_w$) from the loss function, allowing synaptic weights to evolve without distribution constraints.

Oracle score calculation in MICrONS data. We calculated the response reliability of layer 5 (L5) neurons from the MICrONS electron microscopy dataset[21]. Co-registered neurons were selected by excluding the top 25% of residuals and the bottom 25% of separation scores, as defined in the MICrONS flagship paper. Oracle scores were computed as described therein.

### *Software and hardware*

The model was constructed using the Brain Modeling ToolKit (BMTK)[41] with the NEST simulator, employing the SONATA data format[39]. MICrONS data were accessed via the CAVEclient interface. The SONATA network files produced by this construction pipeline serve as input to the TensorFlow-based[106] differentiable simulator.

All differentiable simulations and training were implemented in TensorFlow 2.15[106]. The TensorFlow implementation operates on the SONATA-derived connectivity and cell-parameter files generated by the BMTK/NEST construction pipeline, enabling direct reuse of the biologically grounded network specification in the gradient-based training workflow.

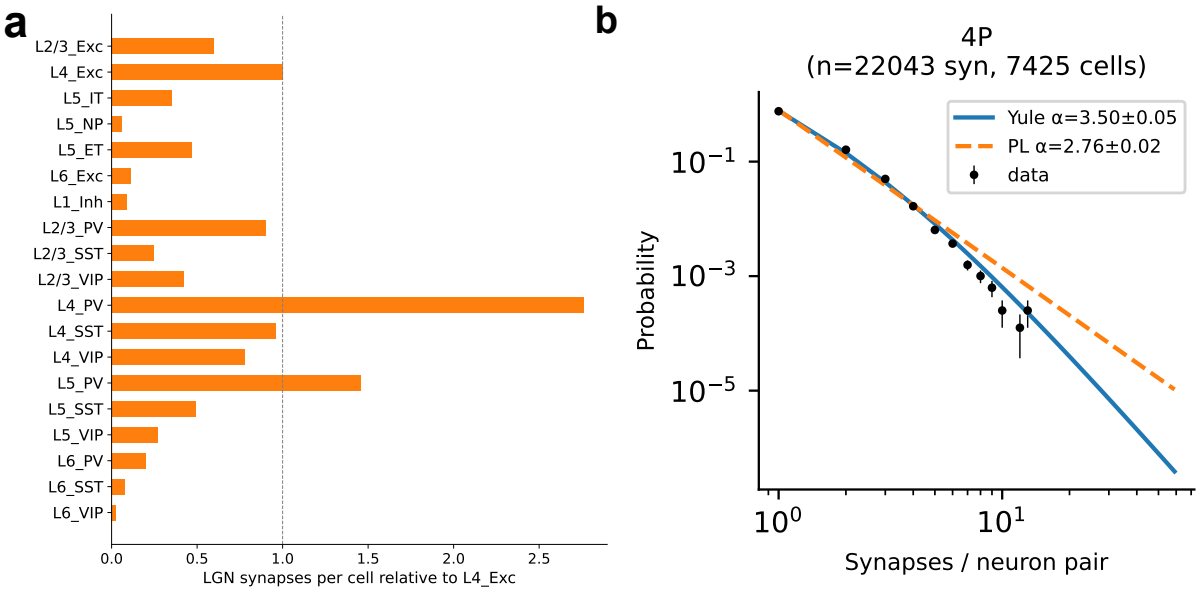
Code and trained model parameters are available at https://github.com/AllenInstitute/biorealistic-v1-model.

Benchmarks reported in Fig. 2d were run on a single NVIDIA RTX PRO 6000 GPU (96 GB VRAM) with 16 AMD EPYC 9754 CPU cores. Final production training runs were performed on a high-performance computing cluster using a single NVIDIA A100 GPU (40 GB VRAM) and Intel Xeon Gold 6330N processors.

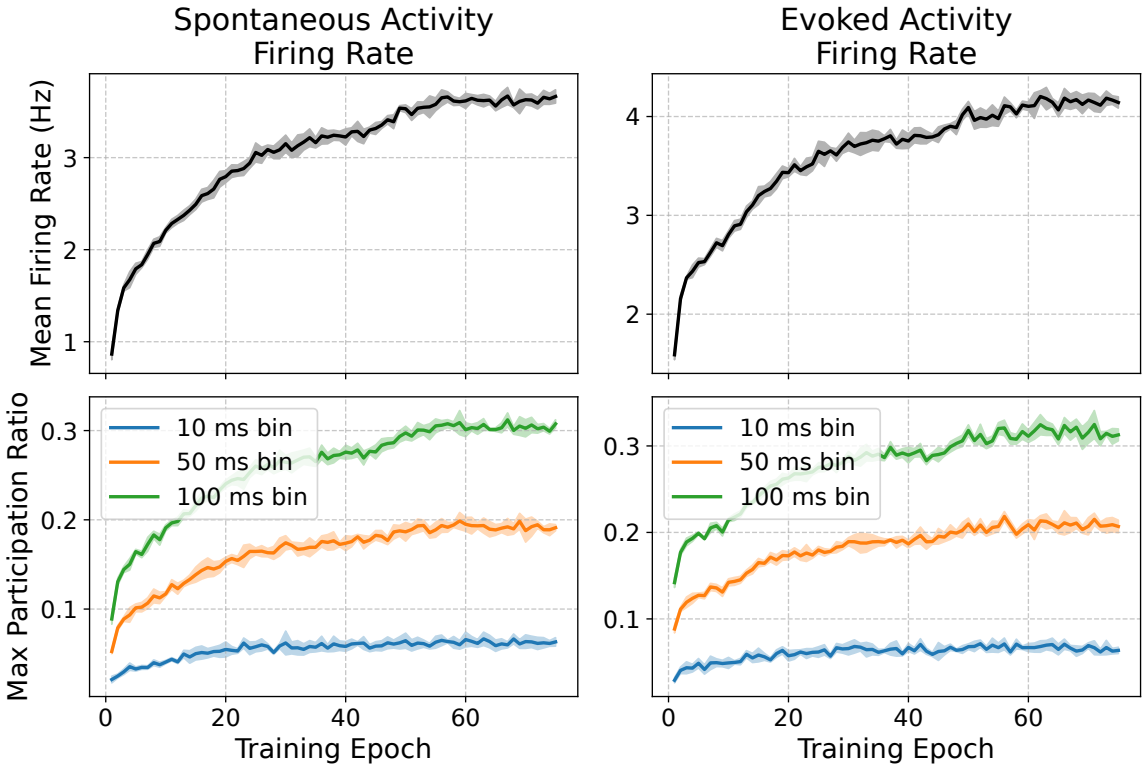
Simulations for model characterization (Figs. 3–6) were done with BMTK FilterNet and PointNet (with NEST simulator 3.6[110]).

## Methods References

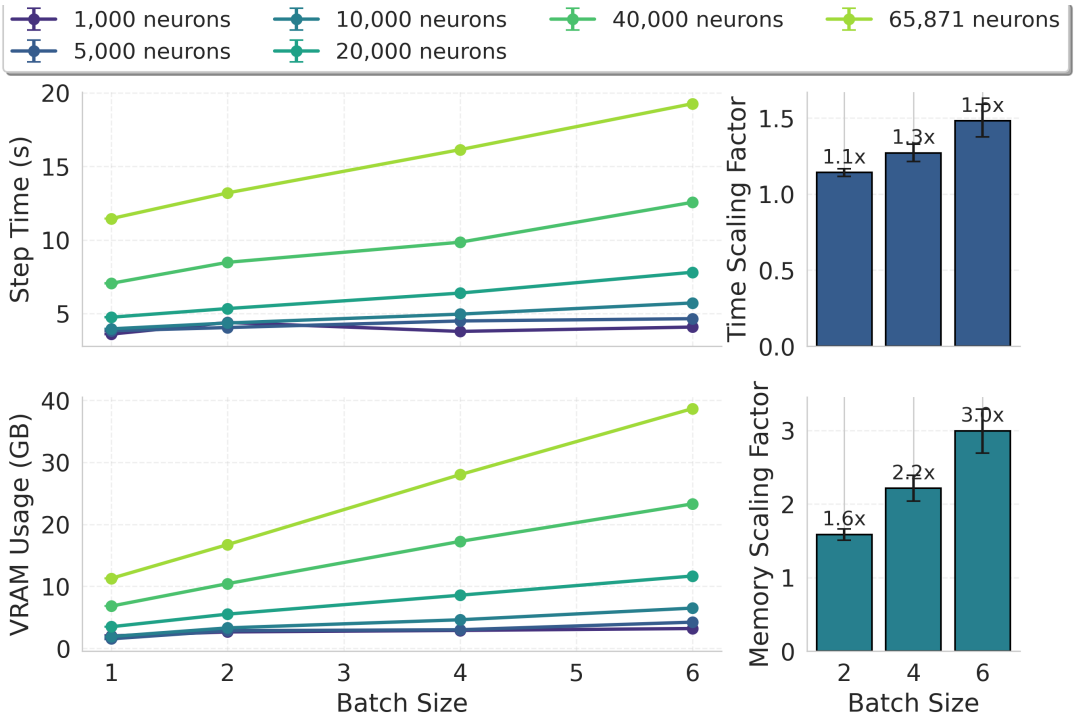

94. Lee, S., Hjerling-Leffler, J., Zagha, E., Fishell, G., & Rudy, B. "The Largest Group of Superficial Neocortical GABAergic Interneurons Expresses Ionotropic Serotonin Receptors". *The Journal of Neuroscience 30*, 16796–16808 (2010).
95. Lien, A. D. & Scanziani, M. "Cortical direction selectivity emerges at convergence of thalamic synapses". *Nature 558*, 80–86 (2018).
96. Simon, H. A. "On a class of skew distribution functions". *Biometrika 42*, 425–440 (1955).
97. Pachón, A., Polito, F., & Sacerdote, L. "Random Graphs Associated to some Discrete and Continuous Time Preferential Attachment Models". *Journal of Statistical Physics 162*, 1608–1638 (2016).
98. Fraile, J. G. *Neocortical dynamics and computational mechanisms: Integrating sensory and higher-order inputs.* Available at https://dspace.uib.es/xmlui/handle/11201/173343. PhD thesis (Universitat de les Illes Balears, 2025).
99. Guo, Y., Chen, Y., Zhang, L., Wang, Y., Liu, X., Tong, X., Ou, Y., Huang, X., & Ma, Z. *Reducing information loss for spiking neural networks.* in *European Conference on Computer Vision* (2022), 36–52.
100. Rotter, S. & Diesmann, M. "Exact digital simulation of time-invariant linear systems with applications to neuronal modeling". *Biological cybernetics 81*, 381–402 (1999).
101. Abbott, L. F. & Nelson, S. B. "Synaptic plasticity: taming the beast". *Nature Neuroscience 3*, 1178–1183 (2000).
102. Arkhipov, A. *et al.* "Visual physiology of the layer 4 cortical circuit in silico". *PLoS computational biology 14*, e1006535 (2018).
103. Fernandez-Hart, T., Knight, J. C., & Kalganova, T. "Posit and floating-point based Izhikevich neuron: A Comparison of arithmetic". *Neurocomputing 597*, 127903 (2024).
104. Werbos, P. J. "Backpropagation through time: what it does and how to do it". *Proceedings of the IEEE 78*, 1550–1560 (1990).
105. Lillicrap, T. P. & Santoro, A. "Backpropagation through time and the brain". *Current opinion in neurobiology 55*, 82–89 (2019).
106. Martin Abadi *et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* Software available from: https://www.tensorflow.org/. 2015.
107. Dabney, W., Rowland, M., Bellemare, M., & Munos, R. *Distributional reinforcement learning with quantile regression.* in *Proceedings of the AAAI conference on artificial intelligence* **32** (2018).
108. Scherr, F. & Maass, W. "Analysis of the computational strategy of a detailed laminar cortical microcircuit model for solving the image-change-detection task". *bioRxiv,* 2021–11 (2021).
109. Buluç, A., Fineman, J. T., Frigo, M., Gilbert, J. R., & Leiserson, C. E. *Parallel sparse matrix-vector and matrix-transpose-vector multiplication using compressed sparse blocks.* in *Proceedings of the twenty-first annual symposium on Parallelism in algorithms and architectures* (2009), 233–244.
110. Villamar, J. *et al. NEST 3.6.* 2023.



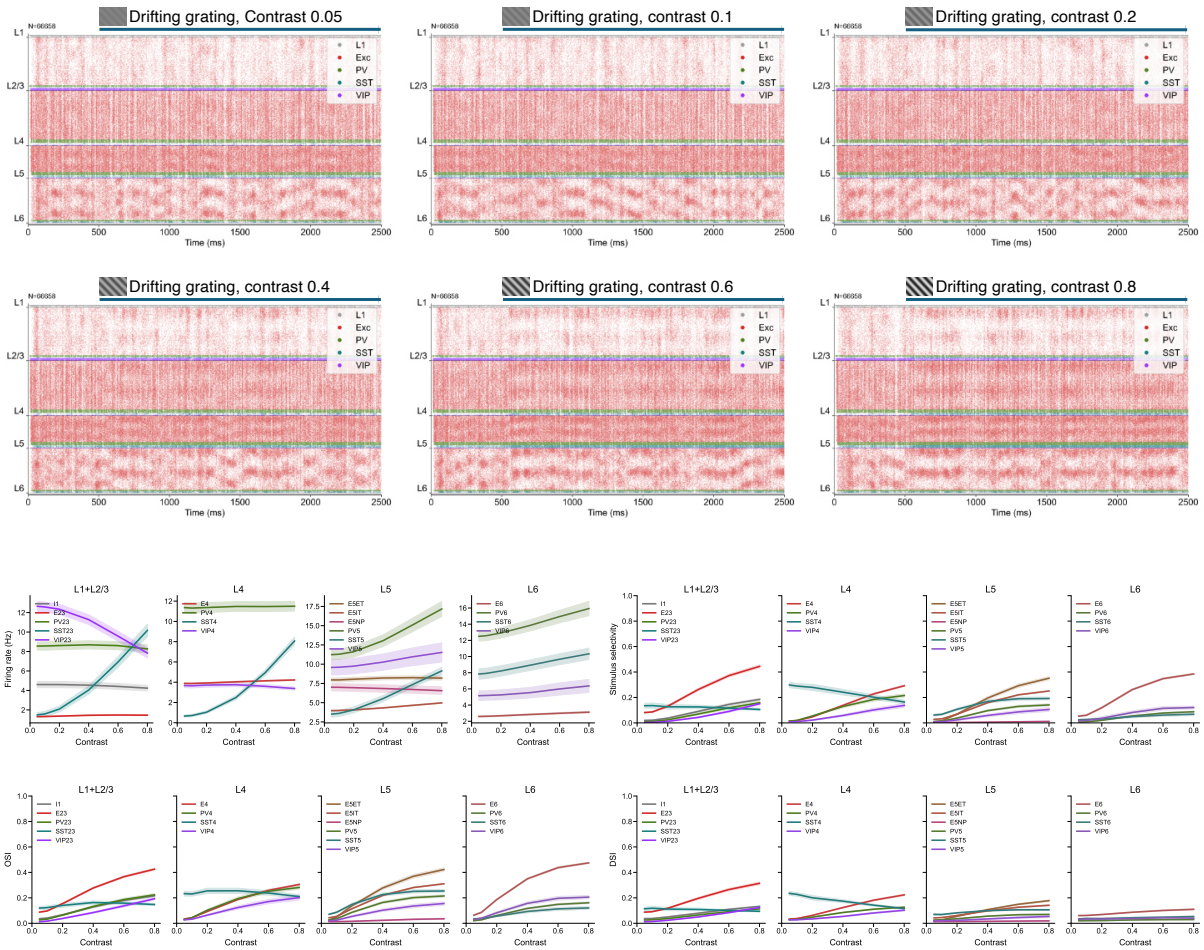

**Extended Data Figure 1: a**: Relative synapse count per cell for each cell type relative to L4 Exc population. **b**: Per-connection synapse count distribution follow a Yule–Simon distribution. An example Yule–Simon distribution fit (blue solid line) to a per-connection synapse count distribution between LGN axons and L4 pyramidal neurons. Yule–Simon distribution fits better than the power-law distribution (orange dashed line).



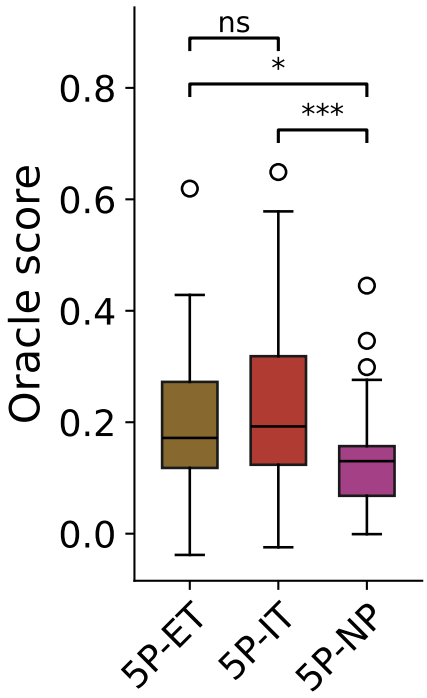

**Extended Data Figure 2: Firing rate and participation ratio during training.** Average firing rate (top) and maximum participation ratio (fraction of neurons that fired in a bin) across 500 ms simulations are shown for spontaneous (left) and evoked activity (right).



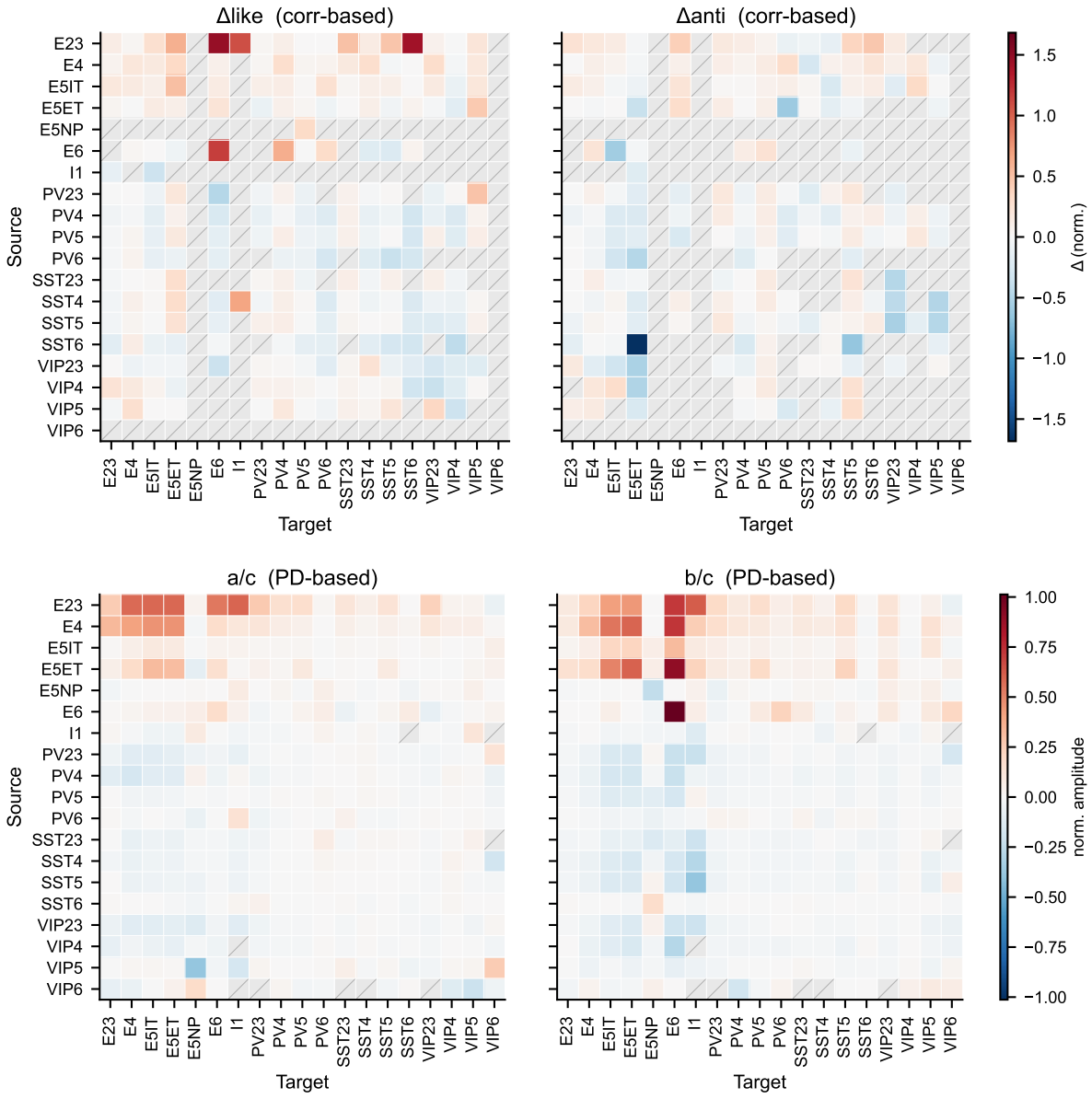

**Extended Data Figure 3: Batch-size scaling of training time and memory on an NVIDIA L40S platform.** Benchmarks were run with the TensorFlow simulator on a single NVIDIA L40S GPU (48 GB VRAM) using 8 AMD EPYC 9754 CPU cores for a 1000 ms (500 ms of gray screen and 500 ms of drifting gratings) training step. Left panels show absolute step time and GPU memory versus batch size for multiple network sizes (mean $\pm$ SEM over 3,000 runs). Right panels report scaling factors normalized to batch size $B = 1$ (mean $\pm$ SEM over 3,000 runs). Increasing batch size from $1$ to $6$ produced sublinear step-time growth (about $1.6\times$ slower) while VRAM also increased sublinearly, defining the efficient batch size processing in our implementation.



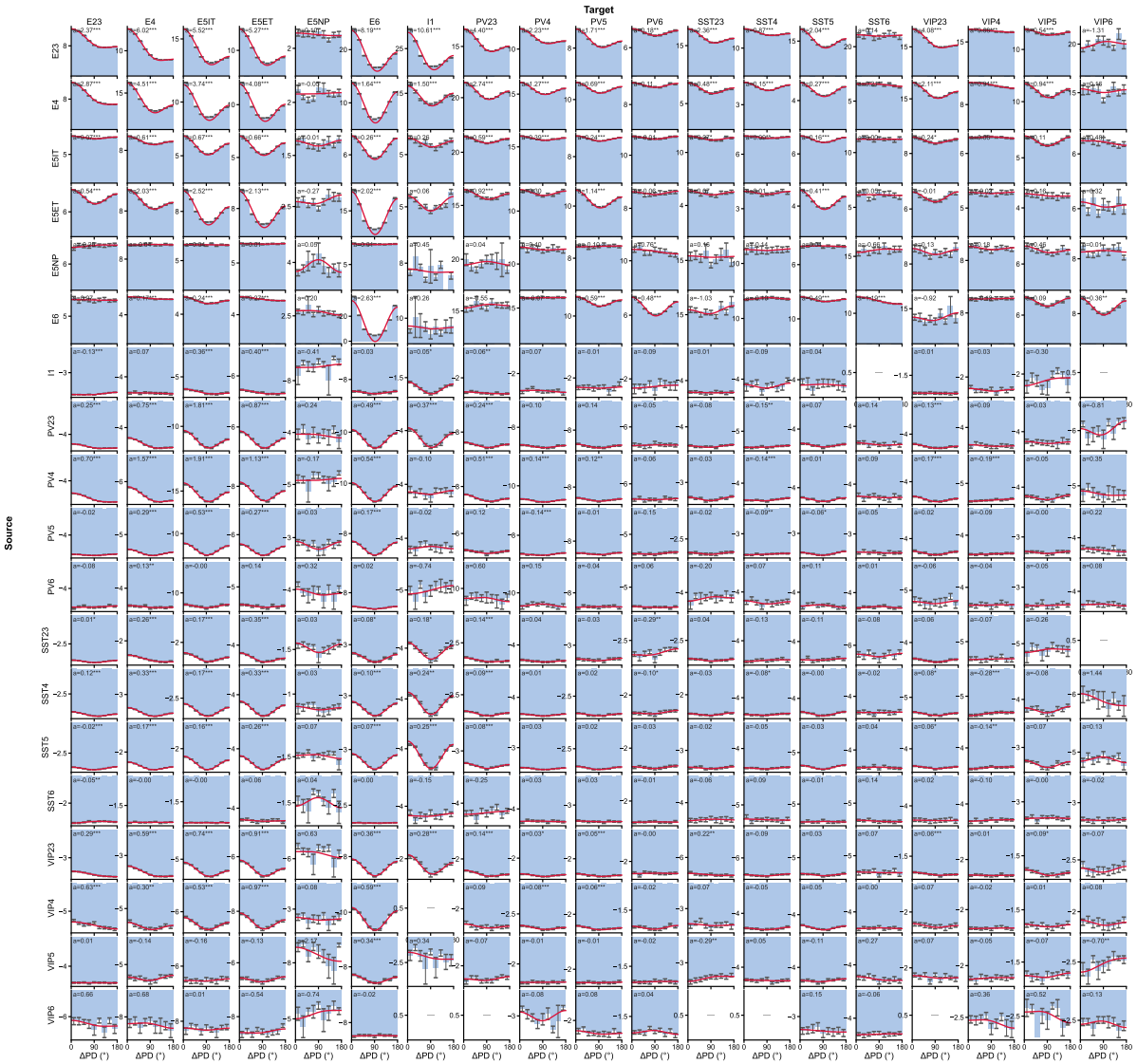

**Extended Data Figure 4: Contrast response.** Raster plots at different contrast levels are shown for $45^\circ$ drifting gratings (top). Line plots show firing rates, stimulus selectivity, orientation selectivity index (OSI), and direction selectivity index (DSI) (bottom) for all cell types.



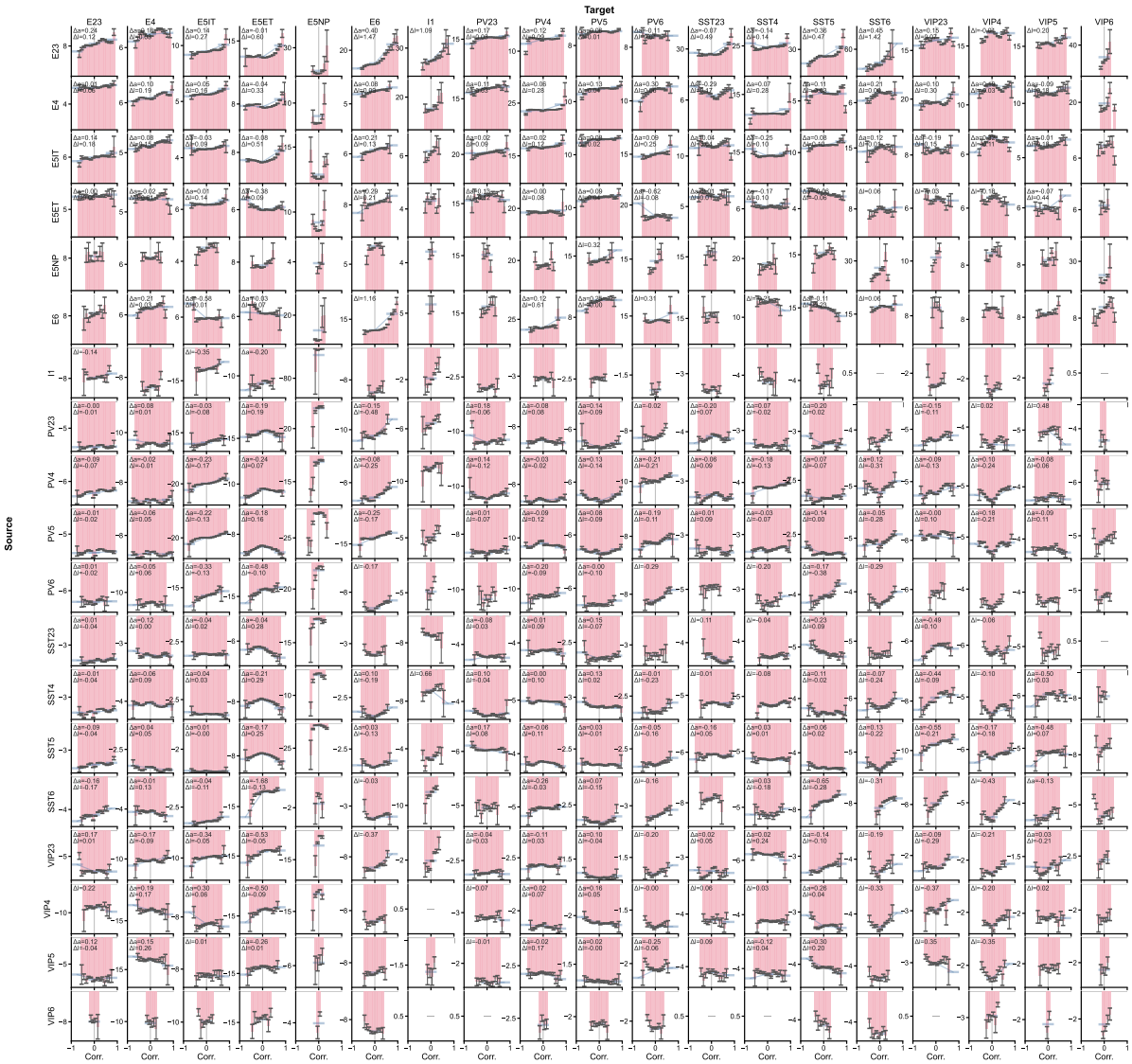

**Extended Data Figure 5: Oracle scores of layer 5 neurons in MICrONS.** Oracle scores of layer 5 neurons in MICrONS.



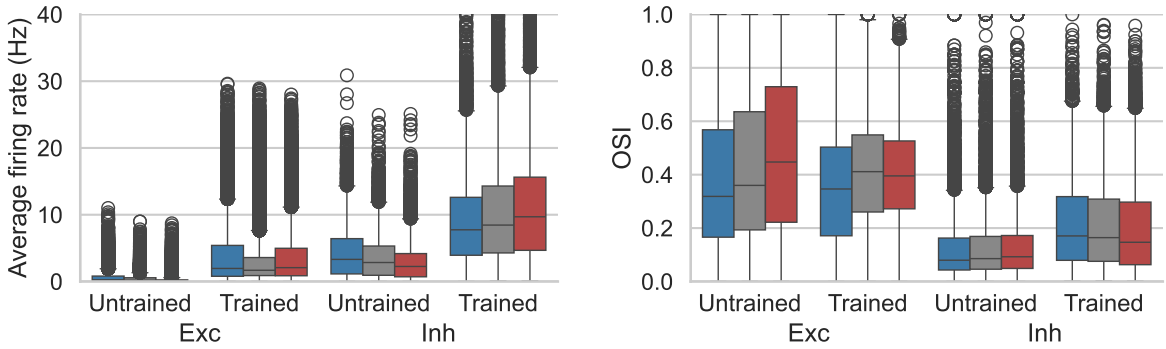

**Extended Data Figure 6: Like-to-like metrics.** Full heatmaps of the like-to-like metrics ($\Delta_{\text{like}}$, $\Delta_{\text{anti}}$, $a/c$, and $b/c$) are shown for all cell type pairs.



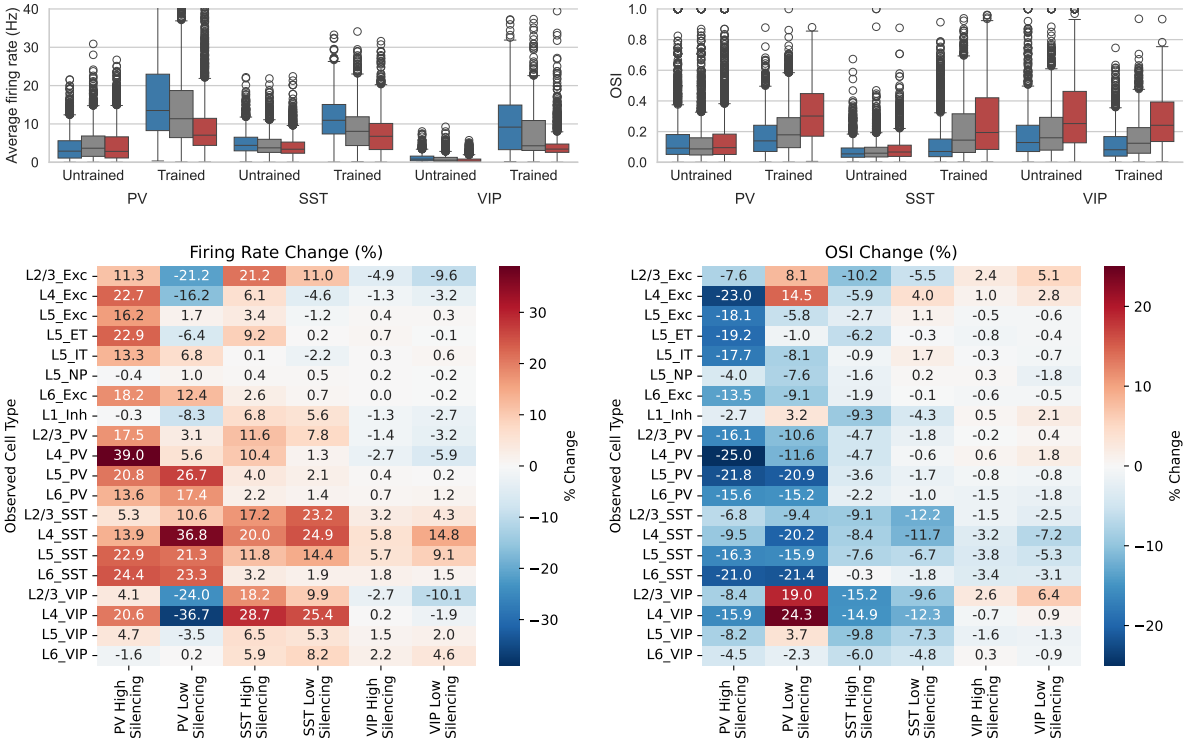

**Extended Data Figure 7: Like-to-like characterization with DGs.** Full histograms of synaptic weight as a function of difference in preferred-direction for DG responses are shown for all cell type pairs.



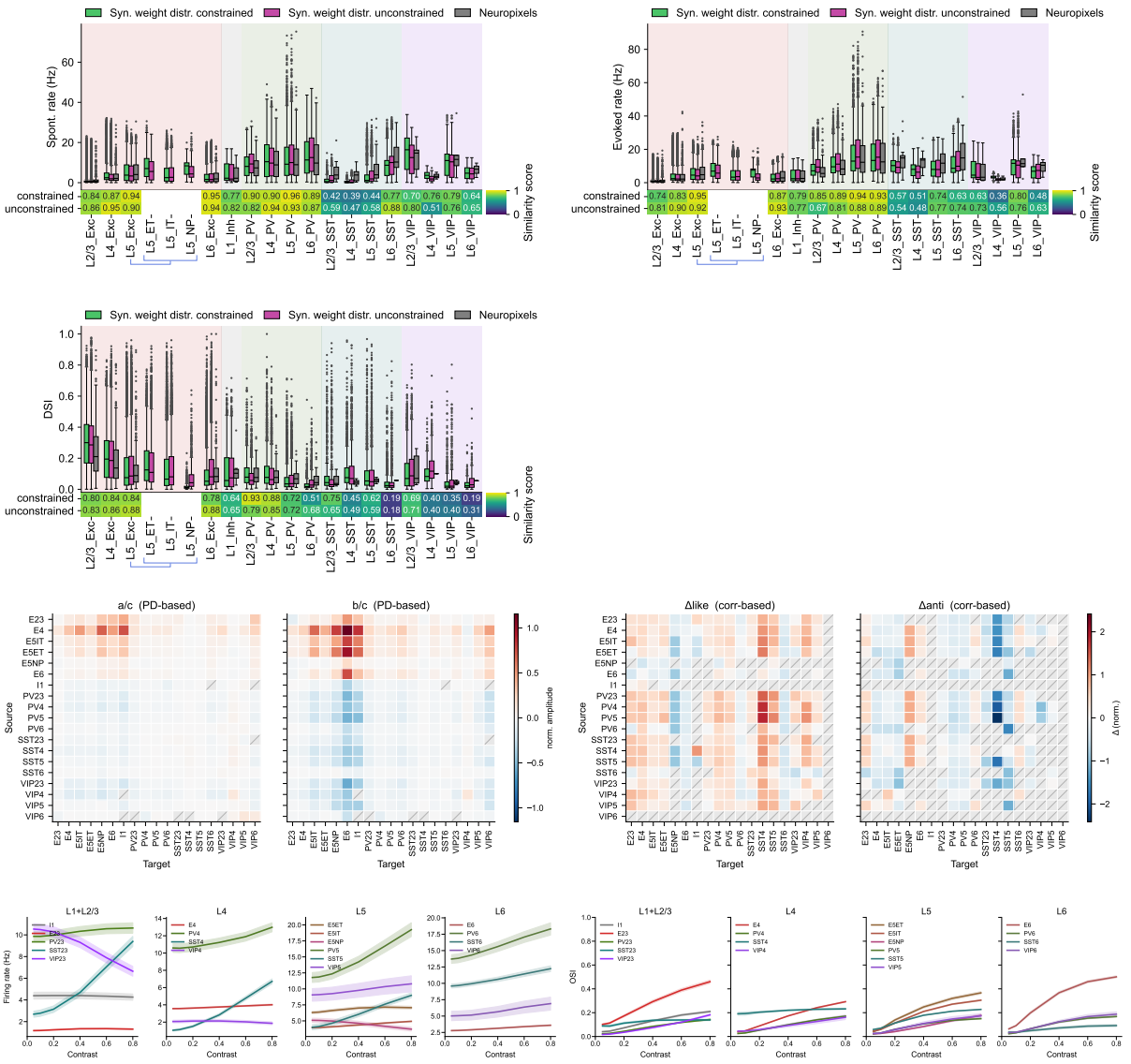

**Extended Data Figure 8: Like-to-like characterization with correlation of responses to natural images.** Full histograms of synaptic weight as a function of correlation of responses to Brain Observatory natural images are shown for all cell type pairs.



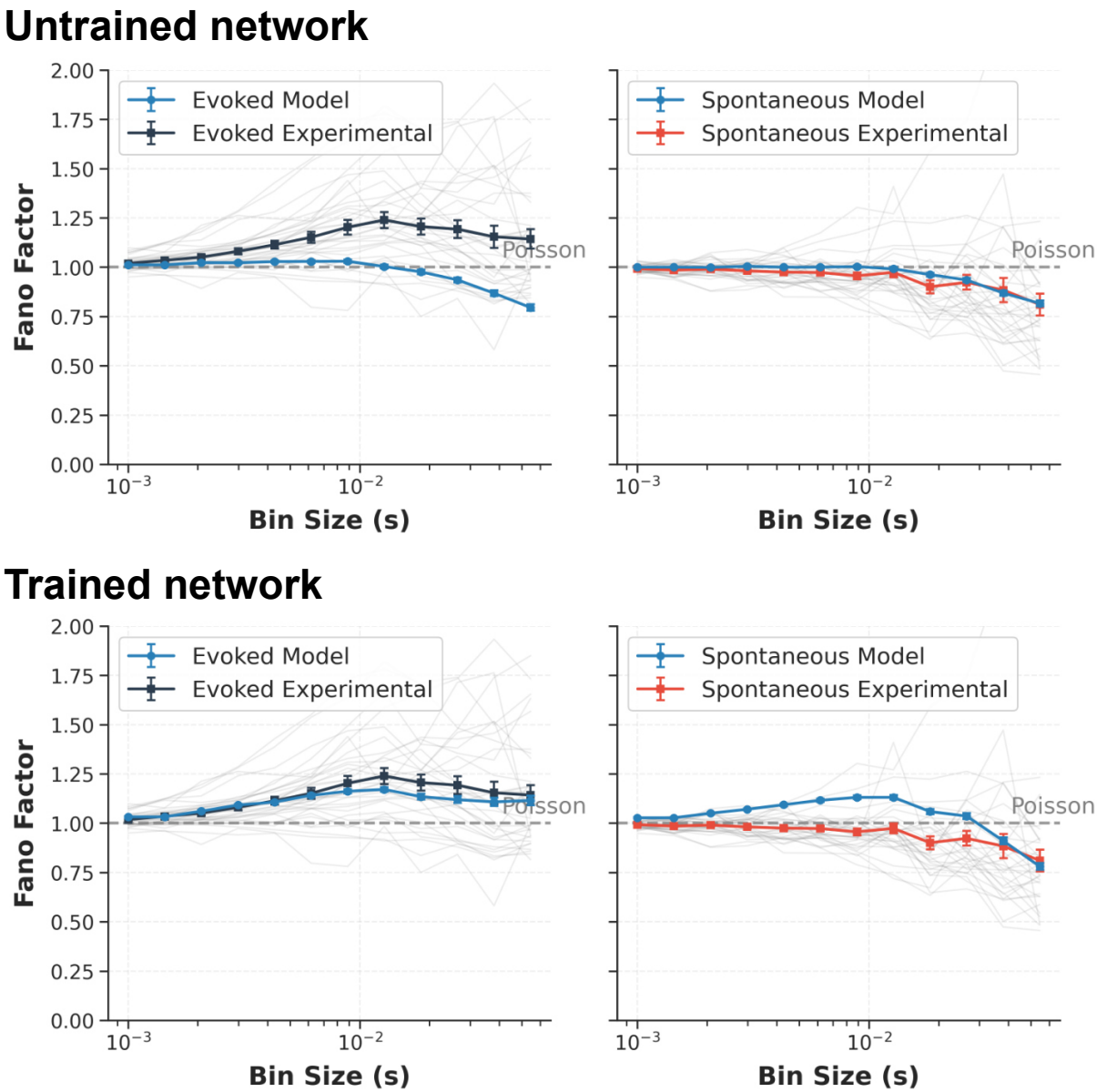

**Extended Data Figure 9: Cohorts by incoming synaptic weights.** Average firing rates and orientation selectivity index (OSI) for cohorts separated by total incoming synaptic weight.




**Extended Data Figure 10: Cell-type cohort suppression analysis.** Top: functional features (left: firing rates, right: OSI) of the cohorts stratified by outgoing weights for inhibitory subtypes. Bottom: The effect of suppression of the subtype cohorts on the firing rates and OSI.




**Extended Data Figure 11: Additional metrics for networks trained without synaptic weight distribution constraints. a,** Additional metrics (spontaneous firing rates, evoked firing rates, and direction selectivity index) for unconstrained networks. **b,** Full heatmaps of like-to-like metrics for unconstrained networks. **c,** Contrast responses for unconstrained networks.




**Extended Data Figure 12: Multi-scale Fano factors.** Fano factors computed across multiple bin sizes for evoked (left) and spontaneous (right) activity. The model results are shown for a single representative trained network (blue line; mean $\pm$ SEM from multiple pools of neurons). The experimental results are derived from the Brain Observatory Neuropixels dataset[20]. Black and red lines indicate the mean and SEM of 32 mice; thin gray lines indicate individual mice.