Recurrent cortical networks encode natural sensory statistics via sequence filtering

Ciana E. Deveau*[1,2,3], Zhishang Zhou*[1], Paul K. LaFosse[1,2,4], Yanting Deng[1], Saghar Mirbagheri[5], Nicholas Steinmetz[5], Mark H. Histed[1]
* equal contributions

1 Intramural Program, National Institute of Mental Health, National Institutes of Health; Bethesda, MD USA. 2 NIH Graduate Partnership Program, Bethesda, MD USA, 3 Department of Neuroscience, Brown University, Providence RI USA, 4 Neuroscience and Cognitive Science Program, University of Maryland, College Park MD USA 5 Department of Biological Structure, University of Washington, Seattle, WA USA

Email: mark.histed@nih.gov

Abstract

Recurrent neural networks can generate dynamics, but in sensory cortex it has been unclear if any dynamic processing is supported by the dense recurrent excitatory-excitatory network. Here we show a new role for recurrent connections in mouse visual cortex: they support powerful dynamical computations, but by filtering sequences of input instead of generating sequences. Using two-photon optogenetics, we measure neural responses to natural images and play them back, finding inputs are amplified when played back during the correct movie dynamic context— when the preceding sequence corresponds to natural vision. This sequence selectivity depends on a network mechanism: earlier input patterns produce responses in other local neurons, which interact with later input patterns. We confirm this mechanism by designing sequences of inputs that are amplified or suppressed by the network. These data suggest recurrent cortical connections perform predictive processing, encoding the statistics of the natural world in input-output transformations.

Main text

A defining feature of all cerebral cortical areas is their extensive local recurrent connectivity. Excitatory cortical cells make up approximately 80% of the neurons in all cortical areas, and each excitatory cell receives hundreds or thousands of inputs[1]. A majority of those input synapses come from other local neurons, within a distance of a few hundred microns[2,3]. The resulting massive local excitatory-excitatory connectivity is a common feature of all cortical areas (Fig. 1a-d). In several cortical regions this dense recurrent connectivity is thought to generate ongoing or slow dynamics, for example, dynamics in motor cortex associated with muscle movement, and in prefrontal cortex, delay dynamics associated with short-term memory[4,5].

In sensory areas, however, the role of recurrent connections has been less clear (Fig. 1e). In visual cortex, for example, responses to visual input end within a few tens of milliseconds after visual stimuli are extinguished[6], suggesting that sustained or generative responses are not robustly produced by sensory recurrent connections[7]. The computational role of the dense excitatory connectivity in the sensory cortex has thus been unclear. What computations do these local excitatory connections perform during sensation?

Here we study whether these extensive recurrent connections in sensory cortical regions mediate processing of dynamic stimuli. Sensory stimuli activate thousands or more neurons in the brain, and therefore many kinds of input — feedforward, feedback, and recurrent — arrive together to cortical neurons during visual stimulation. Because recurrent inputs co-vary with other kinds of input, it has historically been difficult to isolate and study how neurons' firing affects other nearby neurons via recurrent connections. Two-photon holographic stimulation (Fig. 1f-g)[8,9] allows us to deliver direct inputs to populations of cells in the cortex and directly study the influences of local neural activation on other local neurons (Fig. 1g). This approach delivers optogenetic inputs that are precise spatially (Fig. 1g), and also precise temporally, with one input pattern changing to the next within tens of milliseconds (Fig. 1j; 30 ms pattern duration). We use a viral expression strategy[10] that co-expresses an opsin and a calcium indicator in excitatory neurons in layer 2/3 of mouse primary visual cortex (V1), stably over days to weeks, with minimal optical crosstalk between imaging laser and opsin[10].

Two-photon optogenetics can mirror some features of natural visual input by driving population activity in the cortical network. When animals are shown natural visual movies, many neurons in the visual cortex respond (Fig. 1h). Using two-photon stimulation we can simulate these activity patterns by giving input to many neurons at once (Fig. 1i; in this work 15-30 targeted cells). This all-optical approach allows us to measure sensory responses, select neurons that respond, and replay activity patterns.

Sequential order affects V1 neurons' responses through non-target suppression

If recurrent effects influence V1 dynamics, then changing the order of patterns of input might be expected to produce differences in neurons' responses. To study this, we chose patterns of neurons at random (three patterns; A, B, and C, each a fixed duration, 30 or 60 ms in different experiments, spanning a range of the temporal frequencies often seen in vision: e.g. frame times used in videos[11], flicker fusion frequencies[12,13], and photoreceptor dynamics[14]) and presented these patterns in two different orders (ABC or CBA, Fig. 2a).

We find the sequential order of these population input patterns does indeed affect neurons' responses (Fig. 2b-e). The first pattern produces a greater response than the same pattern when it is delivered later in the sequence (Fig. 2b-e). This sequential context dependence occurs in the cells stimulated as part of the A, B, or C patterns. On average, the cells stimulated first in the sequence show stronger responses (Fig. 2c-e); in particular, the median of the A and C

distributions differ in Fig. 2e. Larger effects are seen in individual neurons, which are strongly modulated by sequential position in both the positive and negative direction (e.g. there is significant variance in the distributions in Fig. 2e, 37% of A neurons are significantly different from zero, 45% of C neurons, K-S test $p < 0.05$).

To quantify the effects of sequential context — that is, the effect of the patterns that precede a given input pattern — we analyzed responses in the B cells (Fig. 2f). The B input pattern is preceded by one pattern in each sequential order: the A pattern in the ABC sequence, and the C pattern in the reverse sequence. Thus, if there is a specific effect of sequential context on individual neurons, the firing of some B cells should be modulated not just by the B pattern input, but also by the pattern that precedes it. Indeed, we found that many of these cells were significantly modulated by sequential context (Fig. 2g-l). Some B cell responses were larger in the ABC sequence (Fig. 2g-i) and some were larger in the CBA sequence (Fig. 2j-l; variance differences via K-S test, Fig. 2n-o).

In sum, stimulation with sequences of random population input patterns shows that V1 neurons' responses to fixed inputs are highly dependent on sequential context. That is, responses to an input pattern are affected by prior patterns of input to different neurons. This suggests that the visual cortex recurrent network may perform a time-based computation for sensory stimuli. To understand this, we next investigated how sequential modulation relates to natural vision.

Patterns corresponding to natural visual inputs are amplified

To study sequential modulation in the visual context, we turned to dynamic natural visual inputs. One of the most common kinds of sequential input that the visual cortex receives comes in response to natural motion. Because of the size and scatter of cortical receptive fields[15,16], and because axons carrying inputs contact many cells, the cortical network receives a changing set of patterned inputs during dynamic vision. As an example, when we see a person or animal running, the cortex "sees" a sequence of patterns of input.

In principle, the cortical network could store, in its recurrent connections, information about the temporal structure of the natural visual world. This could allow the network to preferentially respond to sequences corresponding to natural vision. In that case, sequences of input arising from natural vision would be expected to produce different responses in the network than sequences that do not correspond to natural vision. And this means that earlier parts of a dynamic natural stimulus should influence responses to input arriving at later times.

We tested this hypothesis by using two-photon stimulation to mimic responses from one part of a natural movie while we changed the prior visual context.

We first measured responses to a single frame taken from a movie (Fig. 3a). We constructed a two-photon input pattern from the cells activated by that single frame. We then showed animals the movie and played back the response to the frame with two-photon stimulation, replacing the

visual frame with the input pattern either at the correct time in the movie (Fig. 3a,b: matched context), or at a different time, when the preceding movie frames were not matched to the stimulation pattern (Fig. 3a: unmatched context).

If V1 responses were dependent on the dynamic context of natural vision, neurons should produce different responses to the same, fixed input pattern when presented in different contexts.

That is what we found (Fig. 3d-i). The input pattern produced a larger response in the sequential context matched to the visual movie than when presented in the unmatched context. Responses were not just amplified in the matched context, but were attenuated in the unmatched context relative to the response to the frame when it was not preceded by a dynamic visual stimulus (Fig. 3g-i; Supp Fig. 2a).

An immediate question is how responses to one pattern of input might influence responses to later input. For this to happen — for a recurrent network interaction to modify responses to later inputs — it would seem that responses from one frame should be sustained somewhat in time, to allow earlier responses to interact with responses to later frames. To determine if this were true, we turned to Neuropixels electrophysiological recordings due to their fidelity in measuring the times of spikes. We recorded V1 neuron responses to flashed single image frames (as in Fig. 3a), and sorted neurons into groups based on whether they showed activated or sustained responses. We found that suppressed responses were sustained for several hundred milliseconds (225 ms, Fig. 3j) after a single image frame, long enough to interact with inputs from later frames. For comparison, a movie at 25 frames per second, a rate comfortably within the range where humans see smooth motion, has a 40 ms frame time. The mouse visual system likely supports even faster processing than the human, and yet here we found that natural images flashed for short times produced responses 75 ms long for activated neurons (Fig. 3j) and 225 ms for suppressed cells (Fig. 3j, inset). These data suggest that the dynamics of input during natural vision is sufficient to allow overlap between patterns, allowing recurrent influences from earlier input patterns to affect responses to later inputs.

A recurrent neural network model that amplifies sequences produces the context-dependent responses we see in our data

The observation that responses to single input patterns are modulated by context — by earlier patterns or frames (Fig. 3a-i) — implies that longer sequences of input are also selectively filtered as each movie frame response interacts with later movie frames. That is, the pairwise sequential interactions we found, where single frames of input are amplified based on prior input, can also enable processing of longer sequences of dynamic input where longer natural sequences of input are amplified. Biological motion, like observing a predator moving to attack, can be ongoing for some time and generate changing sequences of input patterns to the cortex through that time. Ethologically, it makes sense that the organism would wish to treat natural input

sequences preferentially compared to all other sequences, such as sequences generated by random fluctuations or spontaneous activity.

To demonstrate that amplification of longer sequences results when each frame in a sequence is amplified by the input patterns that came before it, we examined a recurrent neural network (RNN) trained to preferentially amplify some input sequences (Fig. 3k-m). We asked if it also showed the context-dependent effects seen in our experiment. We found that it did. Indeed, when an RNN was trained to amplify a sequence of input, a single pattern extracted from the sequence and played back in the correct context produced an amplified response, compared to when it was presented in the incorrect context (Fig. 3m).

Thus, an RNN trained to amplify a long sequence of input (Fig. 3m) mirrors the context-dependent stimulation effect seen in our experiments (Fig. 3h). Together with the finding that neurons' responses to input are dependent on earlier inputs (Fig. 2), these observations support the idea that the visual cortex recurrent network is filtering extended sequences of input, specifically amplifying input sequences corresponding to natural vision.

Recurrent network mechanism: one input pattern creates responses in other, non-targeted local neurons, which interact with later input patterns

If sequence modulation was created by local, recurrent interactions, we should be able to see signs of these interactions in the network activity. To look for specific influences — to see how particular input patterns filter later patterns — we separated stimulation patterns in time, inserting delays of several seconds (8 sec, Methods) between one or more patterns in a sequence (Fig. 4a). We first inserted a delay between pattern B and C in an ABC sequence (A,B,C patterns randomly chosen, Methods).

The AB sequence produces striking non-target responses (Fig. 4a-c) in cells not receiving stimulation, with some non-stimulated cells activated, and others suppressed. These non-target neural responses are likely to be primarily due to local recurrent interactions, not due to axonal or dendrite activation, as here we used a somatically-targeted opsin (stChrimsonR[10]).

These non-target responses induced by the AB sequence can occur in neurons that will later be stimulated in the C pattern. This could be a mechanism for sequential modulation: that is, a single input pattern produces non-target responses in other local neurons, and these responses interact with later patterns of input to modify responses to those later patterns.

To see how earlier patterns impacted later patterns, we examined how C cells changed their responses when preceded by the AB pattern, and whether this change was predictable from the AB pattern responses. In prior work we have found attenuation-by-suppression[17], where V1 neurons, when suppressed, produce smaller (sublinear) responses to input. To understand if sequence selectivity could depend on such nonlinearities, we examined non-target responses in C cells generated by the AB sequence. We compared the ABC response to a prediction computed

by summing the C response and the AB response. We found that excited non-target cells did produce linear responses when they were stimulated next in sequence, with ABC responses well-predicted by the sum of non-target responses to AB and C (Fig. 4d,f, Supp. Fig. 3a). However, suppressed non-target cells showed nonlinear responses (Fig. 4e,f, Supp. Fig. 3a). That is, the response of the C pattern embedded in the full sequence was attenuated compared to the response to the C pattern alone. This nonlinearity can support sequential modulation.

These responses suggest that the sequential modulation we observed can be generated via a network mechanism. Non-target responses, induced by the local recurrent network via earlier patterns of input, interact with later patterns. And in some cases this summation can be nonlinear, supporting nonlinear transformation of inputs, such as amplification of natural sequences.

Designed input sequences produce predicted effects, confirming the network mechanism

To provide an explicit test of this network mechanism, we constructed sequences that should be either amplified or suppressed according to the mechanism described above. If our constructed sequences, when delivered to the network, produce the expected amplification or suppression, this would be strong evidence that this recurrent mechanism underlies sequence modulation.

To construct such sequences, we first measured responses to a single input pattern (Fig. 4g). We then created later stimulation patterns based on the non-target responses to those single patterns. We created ‘amplified’ sequences by adding a pattern of stimulated cells that were activated by the first pattern, and ‘attenuated’ sequences by adding a pattern of stimulated cells suppressed by the first pattern (Supp. Fig. 3g-j). We repeated this step a second time to create amplified and attenuated sequences that were three patterns long.

We then asked whether the putative amplified and attenuated sequences created different responses. Confirming our hypothesis, we found that the sequence designed to be amplified produced a much larger response than the sequence designed to be attenuated (Fig. 4h-j).

An additional finding was that amplified patterns produced responses that were sparser than suppressed sequences — a small number of outlier cells produced large responses that elevated the mean (Fig. 4i,k). This sparsity is a hallmark of responses to natural movies[18–20], and supports the idea that the cortical network can amplify sequences of responses associated with natural vision.

In sum, experiments separating sequences in time (Fig. 4a-f), and constructing sequences designed to be amplified or suppressed (Fig. 4g-k), together support a recurrent network mechanism for sequence tuning. Responses to an earlier pattern in a sequence create non-target responses in other local neurons that influence responses to later patterns.

Discussion

Our results, taken together, show a computational role for the dense recurrent network in layer 2/3 of sensory (visual) cortex: to amplify some sequences of input and attenuate others. While V1 does not generate complex sustained or ongoing dynamics, we find the V1 network does in fact support complex temporal processing — via filtering sequences of input. This computation is a form of predictive processing[21,22], where responses to a given visual input are influenced by previous inputs, so that the natural, expected sequence of inputs is amplified. This predictive processing is achieved by altering the network's input-output transformation, without explicit error or mismatch signals that might be expected from traditional predictive coding theories[22–26].

Our data suggest the large number of excitatory-excitatory recurrent connections in the cortex are used for learning the structure of the natural world, to amplify inputs corresponding to natural vision. No specific inhibitory population is needed to drive our effects, as the suppression we observed could be generated as withdrawal of excitation with broadly tuned inhibition, given the cortical network operates in a balanced state[27–32]. We find that is the details of the non-target responses — the excitatory neurons *not* receiving direct input, but influenced by recurrent interactions — that act to choose which sequences to amplify. That is, one pattern of input falls on a set of neurons, and those neurons' local connections lead to activation or suppression in subsets of other local neurons. The activation or suppression influences responses to later inputs. The overall effect is not a new dynamic sequence of activity *generated* by the recurrent network, but a selective response to particular dynamic input patterns — that is, active filtering of sequential input.

Our data suggest the numerous excitatory recurrent connections in the cortex give the network great capacity to selectively process the high-dimensional space of natural visual inputs[33]. Because a sequence corresponding to a few seconds of natural vision can affect the firing of many thousands of neurons in the cortex, the patterns of population responses to visual input are high-dimensional. Inputs can fall on the set of neurons in L2/3 in many different ways, and the potential non-target patterns can also occur in many different ways. To support selective processing of high-dimensional activity would seem to require a large number of synapses. The excitatory-excitatory network in the cortex has this large number of synapses, as there are hundreds of millions of excitatory-excitatory recurrent synapses in a cubic millimeter of cortex[1]. Therefore recurrent networks of the cortex are well-placed to support the transformations required for sequence filtering.

Prior observations that an echo of natural input can be seen in spontaneous activity[34–37] are consistent with our results. During spontaneous activity, neurons in the cortex fluctuate[31,32,38] with measurable, but weak, connection to natural visual responses. Our data suggest that while sequences of neural responses during spontaneous activity states are largely suppressed by the cortical network, some spontaneous patterns which partially match natural patterns can be weakly amplified by the recurrent network. However, because responses during spontaneous

activity lack the feedforward driving inputs that occur during vision, spontaneous patterns are not amplified to the same extent as visual inputs. Thus, active filtering via recurrent connections produces large effects on input-output transformations, and this leads to smaller effects in spontaneous activity.

Another related observation is the straightening of perceptual and neural responses seen in natural vision[39,40]. Straightening refers to the relationship between response patterns at different moments of a natural movie. Prior work has found that cortical responses at one moment are more geometrically similar to later responses than is seen in the cortical inputs. The sequence filtering by the recurrent network we find seems likely to be the mechanism for the straightening effect.

A longstanding question in cortical neuroscience is, as framed by Olshausen and Field in 2006, "What does the other 85% of V1 do?" That is, when population responses to natural scenes are constructed from estimates of the receptive fields of single neurons, why is our ability to predict network responses limited to a fraction of the total variance, even when trial-to-trial fluctuations are removed? Our present results suggest an answer to this question: that the recurrent network in V1 modulates responses, including temporal responses, to boost natural inputs. That is, a big part of the "other" responses in sensory cortex reflect active filtering — changes in the input-output function of the network — via recurrent processing. These input-output changes are difficult to see with recording methods alone. The causal population input modulations we use here allow us to measure input-output changes.

While we have made these measurements in visual cortex, we might speculate that recurrent connections in other sensory regions — auditory, somatosensory, etc — also are used for active filtering. Those sensory modalities also process natural inputs that vary in time, though often with faster changes than in vision, predicting different timescales of sequence filtering in other sensory regions. And associative brain areas downstream of primary sensory cortex can have tuning for complex sequences[41] and may filter at longer timescales[42]. Beyond biology, recurrent networks in artificial systems also often are used to create temporally-structured computations. Our model data aligns with the understanding that recurrent artificial networks can learn temporal statistical structure[43], as seen also in transformers like ChatGPT[44] that have been used to generate highly complex natural language sequences. Thus, densely-connected recurrent networks seem to be useful for sequence processing both in artificial systems and in biological brains.

In summary, these data show a new and powerful purpose for recurrent connectivity in the sensory cerebral cortex: to confer sensitivity to sequential input. The visual cortical L2/3 network is sensitive to dynamic visual context, boosting responses to sequences of input corresponding to dynamic natural sensation, and suppressing others.

Figures

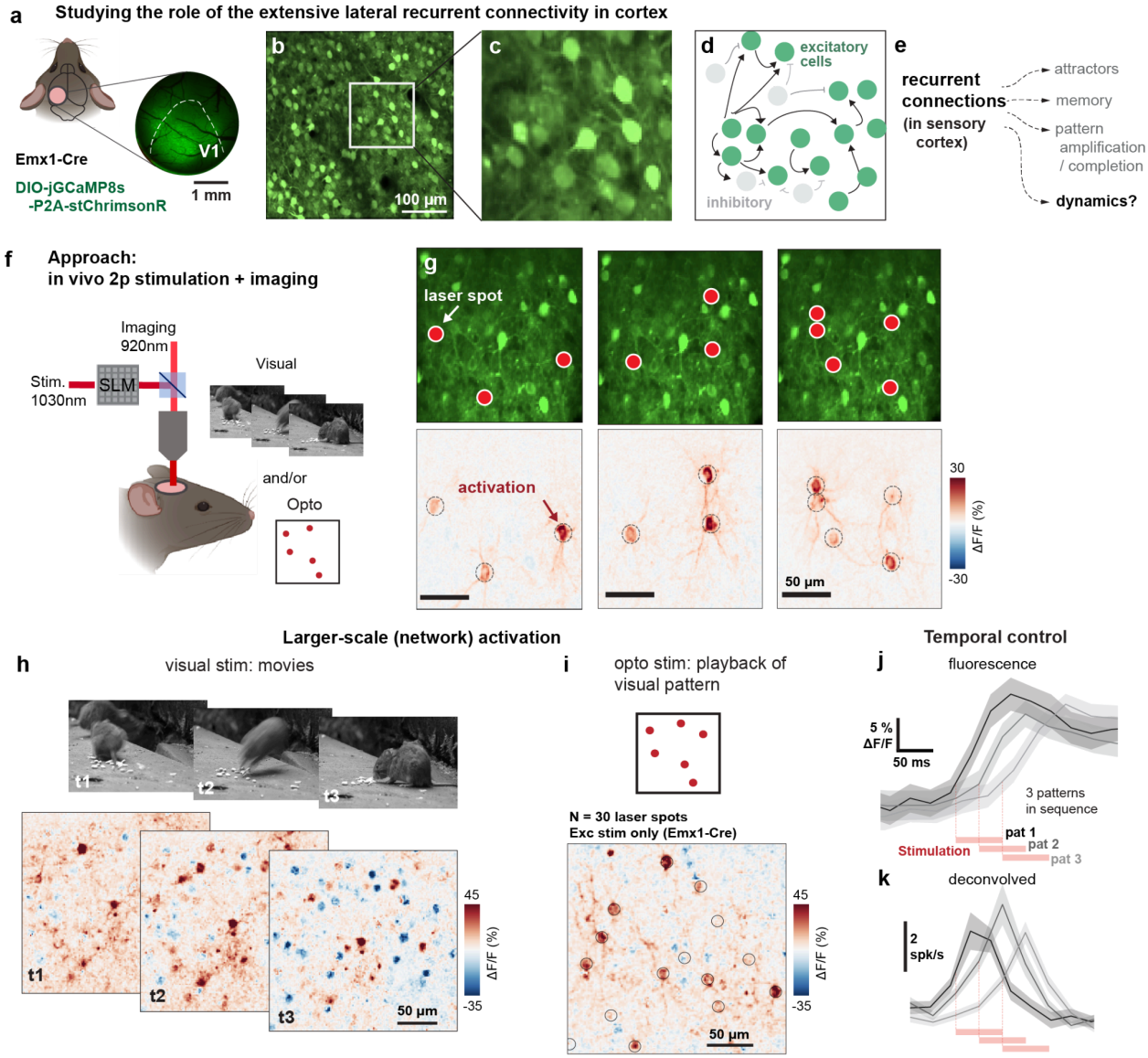


Figure 1: Using two-photon holographic stimulation and imaging to probe the role of excitatory recurrent connections.

a, 3mm window surgical implant over V1. **b**, FOV image of excitatory cells expressing both GCaMP8s and opsin stChrimsonR (via AAV-DIO-jGCaMP8s-P2A-stChrimsonR in Emx1-Cre line). **c**, Enlarged region of (b). **d**, Schematic of the ubiquitous recurrent connections between excitatory cortical neurons. **e**, Possible functions for cortical recurrent connections in V1. **f**, Schematic of two-photon imaging and stimulation experiments, allowing pairing of visual and patterned optogenetic stimuli. **g**, Example of spatial precision of 2p stimulation; left to right: three different patterns, with 3, 3, and 5 laser spots, each spot targeted to a neuron. **h**, Network activation from natural movie input at three timepoints. **i**, Network activation, with non-target changes in non-stimulated neurons, also results from optogenetic patterns (black circles: laser spots, 30 total, a subset of image shown here for visual clarity). **j-k**, Example of temporal precision of sequential 2p stimulation: 3 patterns, 15 cells each, 60 ms stim per pattern with 30

ms between pattern onset. **j**, fluorescence traces, reflecting calcium responses; **k**, corresponding deconvolved traces using OASIS (Methods). These data show that even with some overlap, imaging can resolve distinct temporal peaks with stimulation onsets separated by 30 ms.

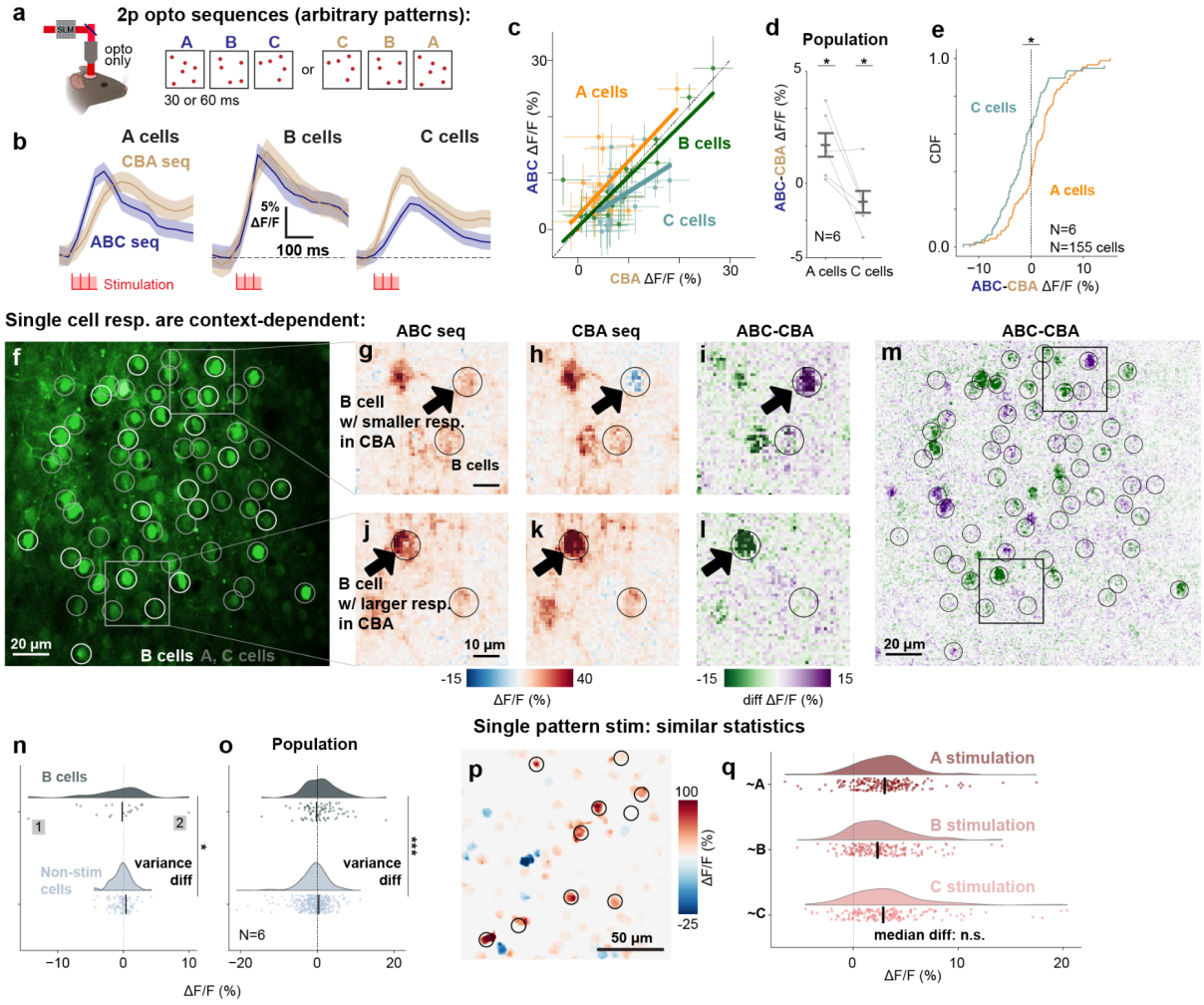


Figure 2: Patterns of input produce different responses in the V1 network depending on the sequential context

a, Experiment schematic: stimulation of three distinct patterns (e.g. N=30 cells per pattern, cells chosen at random for each pattern). Two sequences: forward (ABC) and reverse order (CBA). **b**, Responses depend on sequential order. Mean responses averaged across cells of A (left), B (middle), and C (right) cells in the ABC (blue) or CBA (beige) trials (example expt, pattern duration: 30 ms). **c**, Points: average response (across trials) of individual neurons, same data as in B. Change in slope for A, B, C indicates differences in response based on sequential order. **d**, Population data (N=6 expts). gray lines: ABC - CBA responses of A (left) and C (right) cells; averaged over A, C cells per experiment. Error bars: SEM. (t-test, A cells (N=79): p<0.0001, C cells (N=76): p=0.001). **e**, ABC and CBA responses differ for A cells (N=79) and C cells (N=76) (K-S test, p<0.001). **f-m**, B cell responses reveal that effects of sequential context vary from cell to cell. **f**, Anatomical image (green: bicistronic GCaMP8s and stChrimsonR expression), white circles: cells stimulated in B pattern, gray circles: A and C pattern cells. **g-l**, Neurons' (B cells, black circles) responses to stimulation change based on which other neurons (A, C) are stimulated before them. B cells. **g-i**: example B cell with stronger activation in the ABC sequence compared to CBA. **j-l**: example

B cell with stronger CBA activation. **m**, Larger view of cells shown in g-i. Image matches anatomical FOV in f. Black squares: regions shown in g-i. **n**, Some stimulated cells firing is increased and some decreased by previous stimulation pattern. x-axis: B cell response difference between ABC and CBA pattern; if responses were not modulated by sequence, variability would be similar to non-stimulated neurons (bottom, gray); instead it is larger (A-D test for variance diff, N=19 B cells, N=95 non-stim cells, expt shown in f-j, $p<0.01$). **o**, Same as k, across N=6 expts (A-D test, N=81 B cells, N=339 non-stim cells, $p<0.001$). **p-q**, Sequence modulation is not due to response differences to single pattern stim: A,B,C patterns have similar statistics (data from b-e; Mood's median test, N=3 stim, $p=0.15$).

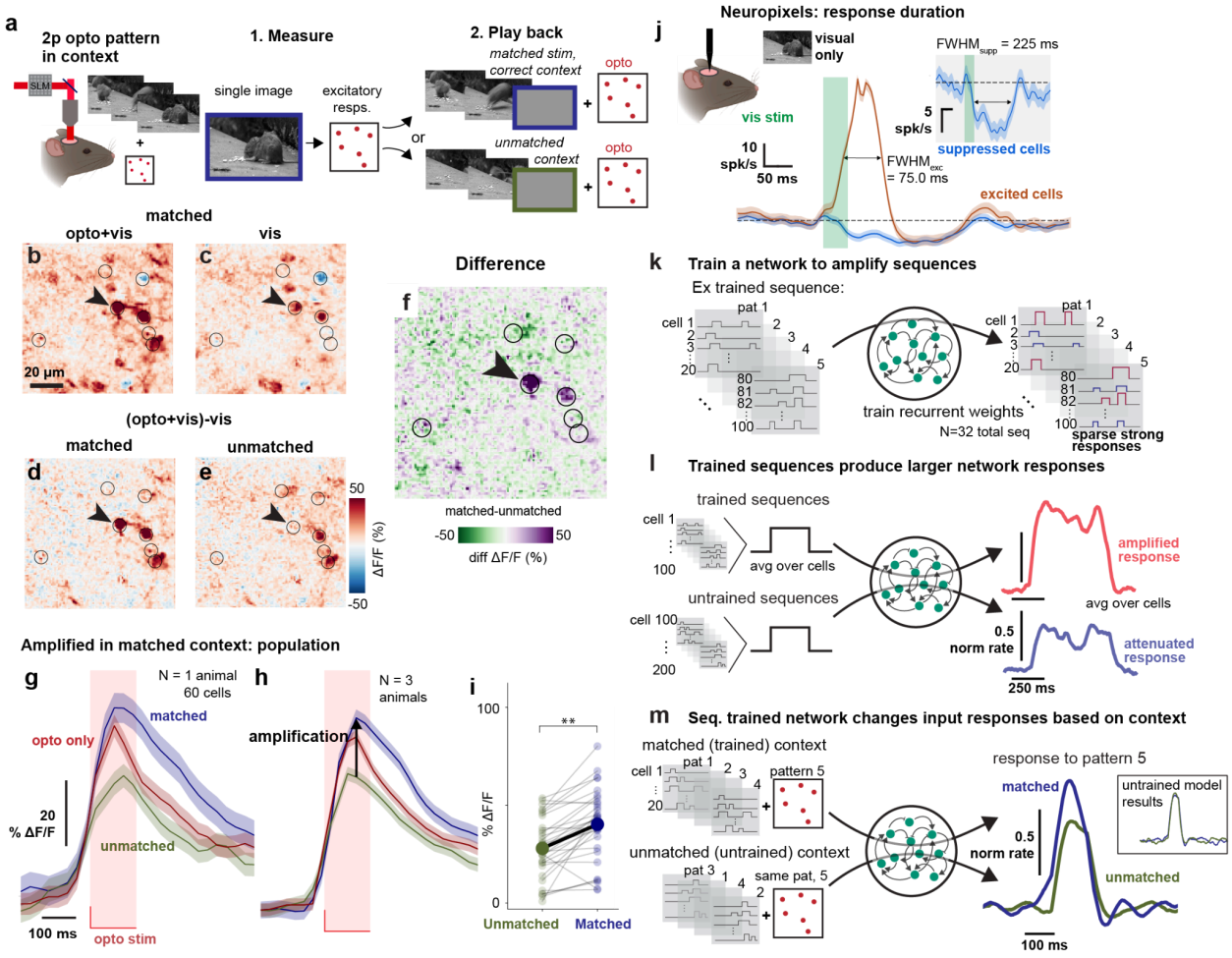


Figure 3: Sequence amplification for visual input, observed by playing back responses to a single frame of a movie.

a, Schematic of dynamic context playback experiment: We recorded responses to a frame from a natural movie, and played back that pattern using stimulation during a visual movie. We did this at the correct time in the movie (matched context) or an incorrect time (unmatched). **b-c**, Example cell responses to combined opto and visual stimulation (b) and visual stim alone (c). **d-e**, Responses in matched and unmatched contexts, computed by subtracting vis response from opto+vis combined response. Black circles: optogenetic stimulation targets. **f**, Example larger response in matched context (arrow), difference of responses shown in (d) and (e). **g**, Mean responses averaged across cells in matched (blue), unmatched (green), and opto only (red) trials (N=1 expt, pattern duration: 120 ms). **h**, Population time courses, N=3 animals, same conventions as (c). Responses are larger in matched context. **i**, Points are cell responses in matched (blue) and unmatched (green) contexts from data in h. Mean optogenetic responses are significantly larger in the matched context (t-test, N=32 cells, p=0.004). **j**, Electrophysiological responses to a flashed natural image show responses sustained for up to hundreds of milliseconds. Excited cells (orange; errorbar, SEM, largely hidden by mean). Suppressed cells (blue). FWHM: Full width at half

max. **k-m**, RNN model trained to amplify specific sequences shows responses that match our data. **k**, Network is trained to amplify sequences of input, resulting in (**l**) amplified output timecourses for trained input sequences. **m**, Simulating experiment of a-i recapitulates the data: single patterns taken from a natural sequence produce larger responses in the correct (blue) context. Inset: untrained model does not show context-dependent responses. Error bars: SEM.

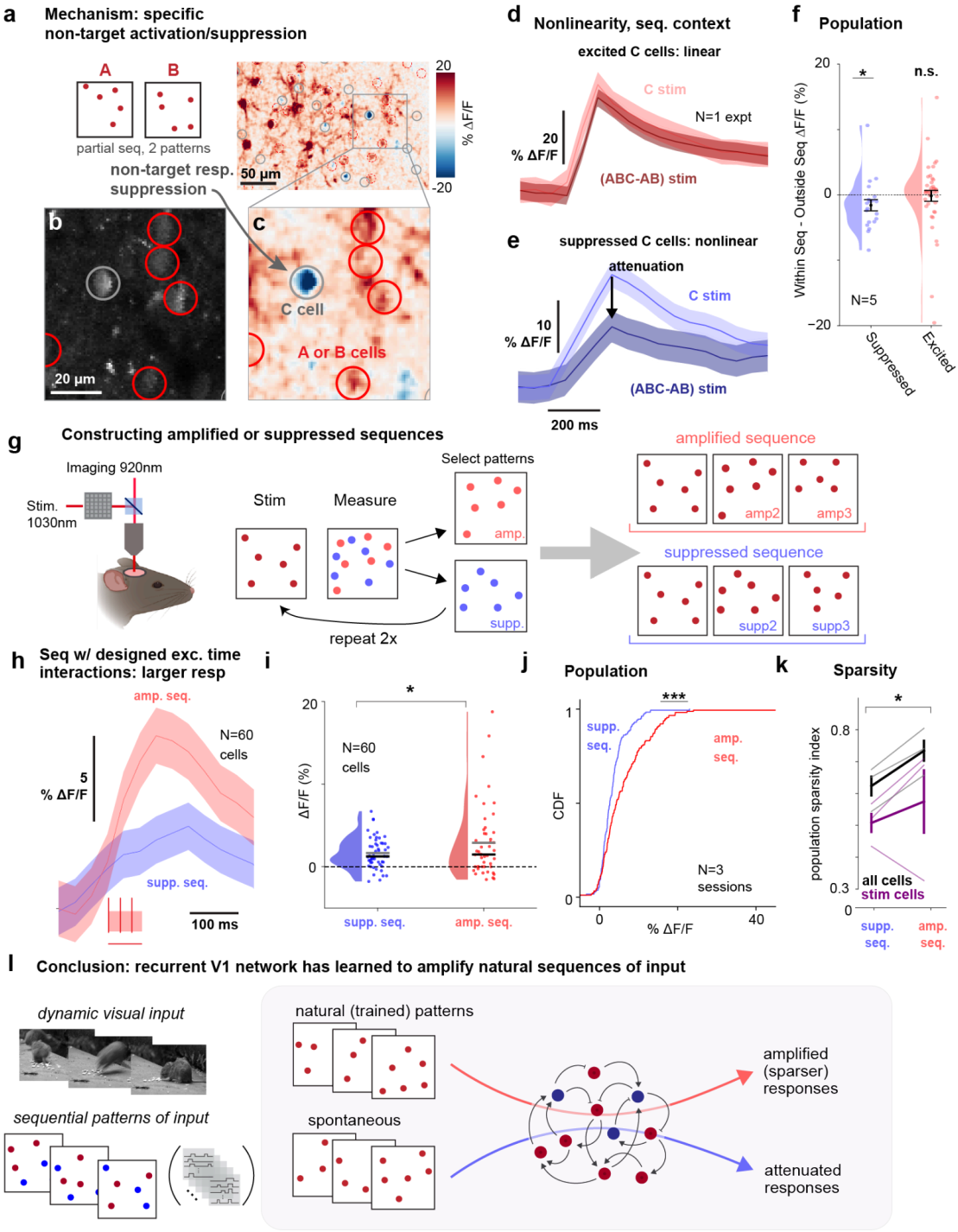


Figure 4: Sequences designed to be amplified produce strong and sparse responses, validating a recurrent network mechanism.

a-c, Separating A, B, and C patterns in time shows non-target responses induced by earlier patterns. Red circles: stimulated cells (from A or B pattern). Gray circle: C cells, here unstimulated. **a**, Responses to AB sequence. Left, schematic; right, GCaMP responses. **b**, Anatomical image (FOV; averaged over time). **c**, Zoomed view shows suppression in unstimulated neuron (gray circle). **d-f**, Excited non-target cells show

linear summation, but suppressed cells show sublinear summation. **d**, Excited cells from one example experiment (N=5 cells). Light red: response to C stimulus alone. Dark red: prediction, response to ABC sequence minus response to AB sequence. Response and prediction overlap. **e**, Suppressed cells, with sublinear summation (N=5 cells; same expt). Same conventions as d. Response to C within the ABC sequence (ABC-AB, dark blue) is smaller than resp. to C alone (light blue; t-test, 300 ms period after stim onset, p=0.013). **f**, Population data (N=5 expts; N=62 cells, C cells only; expt compares C stim, ABC stim, and AB+C stim as described in d-e). Blue: suppressed cells (response to AB stim<0; t-test vs linear prediction, N=20 cells, p=0.04). Red: excited cells (response to AB stim>0; t-test, N=42 cells, p=0.86). **g**, Designing sequences to be amplified or suppressed based on non-target interaction mechanism: experimental schematic. We image responses to stimulation, and then select cells based on response to form the next pattern in the sequence. **h**, Mean responses of amplified (pink) and suppressed (blue) sequences (N=20 cells per pattern, 3 patterns in the sequence; 30 ms pattern duration). **i**, Distribution of responses of stimulated cells. Mean (gray horizontal line), median (black), means are sig. different (t-test, p=0.033; N=60 cells from N=3 patterns). **j**, Population data (K-S test, p<0.0001, N=189 exc cells, N=179 supp cells, N=3 expts). **k**, Sparsity is greater for amplified patterns. Each point: average across neurons. Sparsity in entire population is significantly greater in amplified sequences (t-test, p=0.011; N=3 expts; all cells, gray). **l**, Mechanism and conceptual model from this work: the recurrent network filters sequential inputs, amplifying some and attenuating others. Errorbars: SEM.

Methods

Animals

All experiments were conducted in adherence to NIH and Department of Health and Human Services (HHS) guidelines for animal research and were approved by the Institutional Animal Care and Use Committee (IACUC) at the relevant institutions (optogenetic experiments: NIMH; Neuropixels experiments: University of Washington.)

Optogenetic experiments

Emx1-Cre animals (N=11, https://www.jax.org/strain/005628[45]) of both sexes were used. No systematic differences were observed between males and females. Viral injection and window implants were done at ages 2-7 months. After procedures, animals were singly housed on a reverse 12-hour dark/light cycle.

Neuropixels experiments

Male C57BL/6J mice (N=2), aged two to three months, were used. After procedures, animals were singly housed on a standard 12-hour dark/light cycle.

All animals were put on water schedule five or more days after head-plate surgery, and their weights were carefully monitored to ensure they remained above 85% of their baseline body weight.

Viral injection and cranial window implants

We performed injections and implants as described in [10] (optogenetic experiments; NIMH) or [46] (2023; Neuropixels; University of Washington). Minor differences in the two sets of approaches are not expected to impact results.

Optogenetic experiments

Briefly, mice were given dexamethasone (3.2 mg/kg) 30 minutes before surgery and anesthetized with 1-3% isoflurane (in 100% O_2). A titanium headplate was affixed using C&B Metabond (Parkell) and a 3 mm craniotomy was made over the primary visual cortex (-3.1 mm ML, +1.5 mm AP from lambda) and a glass optical window chronically inserted. Mice were injected with either a mixture of two AAV viruses to induce GCaMP and stChrimsonR into the cells (mixed virus injections), or a bicistronic virus expressing both GCaMP and stChrimsonR in each cell[10]. For the mixed injection, the fluorescent calcium indicator, (AAV9-syn-jGCaMP8s; titers: 5.0-10x10^{12} genome copies (GC)/ml) and the soma-targeted opsin (AAV9-hsyn-DIO-stChrimsonR-mRuby2; titer: 3.0x10^{12} GC/ml) were mixed in phosphate-buffered saline. In the case of the bicistronic virus injections, we used AAV9-hsyn-DIO-jGCaMP8s-p2a-stChrimsonR (titers: 2.9-5.9x10^{12} GC/mL). For each mouse, 3 to 5 injections (100 nL/min, 200 µm depth) were made to cover a wider area of the cortex. A custom made light-blocking cap was fixed onto the implant to limit ambient light

exposure and prevent debris from contacting the window. Experiments began three weeks after these procedures.

Neuropixel experiments

Briefly, mice were induced into anesthesia with 5% isoflurane, and maintained at 2-3%. Carprofen (5 mg/kg) and Lidocaine (2 mg/kg) were administered for analgesia. A titanium head plate and a 3D-printed chamber were affixed using Metabond. Carprofen was administered at 0.05 mg/ml in the water post-surgery for three days. Five days following recovery, mice were put on water restriction. After two days, they were habituated to head fixation for two days, during which they received up to thirty random 5µL sucrose water 10% rewards. Subsequently, animals were exposed to a set of 20 natural images presented over 1000 trials, followed by a 5µL sucrose water reward for five days. This step was performed to provide a control training condition for another experiment. None of the 20 images in this step were used as part of the experiment mentioned in this work. One or two days before recordings, a 2 x 2 mm craniotomy was performed over the primary visual cortex using a dental drill and stereotaxic techniques. The area was then covered with silicone gel. This process was conducted under anesthesia procedures used for the head plate implant, and Carprofen (5 mg/kg) was given for analgesia. A cap was placed on the chamber to prevent debris from getting into the area around the craniotomy.

Retinotopic mapping

Before optogenetic experiments, we determined the location of V1 in the cranial window using a hemodynamic intrinsic imaging protocol previously described in [47]. Briefly, small visual stimuli were presented while 530 nm light was delivered using a fiber-coupled LED (M350F2; Thorlabs, Newton NJ). Hemodynamic response was calculated as the change in reflectance of the cortical surface between the baseline period and a response window starting 3 ms after stimulus onset. Imaging was done on a Zeiss Discovery stereo microscope with a 1x widefield objective through a green long-pass emission filter, acquired at 2 Hz. An average retinotopic map was fit to the cortical responses based on the centroids of the hemodynamic response for each stimulus location.

Two-photon holographic imaging and stimulation

Two-photon imaging and stimulation procedures are described in detail in [10]. Briefly, animals were awake and alert under a 16x water-immersion objective (Nikon; Tokyo, Japan; NA=0.8) manually positioned over the implanted optical window. Imaging was done with a custom-built microscope and controlled by ScanImage software in MATLAB (The Mathworks, Natick, MA). Calcium responses were measured ~100-200 µm below the surface of the pia (L2/3 of V1) with an imaging field of view of 414 x 414 µm. Imaging was performed using 920 nm wavelength light at 15-20 mW and frames acquired at 30 Hz.

Holographic stimulation was performed using a femtosecond pulsed laser (1030 nm, Satsuma, Amplitude Laser, or 1040nm, Monaco, Coherent Inc.) A spatial light modulator (SLM) shaped the laser wavefront to create stimulation patterns (10 µm diameter disks; 30, 60, 90, or 120 ms; 8-16 mW/target, 500 kHz pulse rate). The radial point-spread function (PSF) of diffraction limited spots generated by the SLM was 9.4 µm and the axial PSF was 54 µm.The laser was gated on during horizontal flyback periods and off during the imaging pixel acquisition to allow for approximately simultaneous stimulation and imaging. Reported stimulation power is the average power over on and off periods (i.e. reduced by a factor of 0.3 from the laser power measured without this gating.)

Visual responses were measured in awake, head-fixed mice viewing flashed single natural image frames (40° circular mask with neutral gray background; 120 ms) or full natural movie stimuli (full-field, 2 seconds) presented on an LCD monitor. Mice were given small volumes of water on 20% of trials (3 µL).

Two-photon data analysis

We performed motion correction using the CaImAn toolbox[48] and cell segmentation using Suite2p[49]. All data analysis was done in Python (https://www.python.org). For pixel based analyses, we computed $\Delta F/F_0$ ($\Delta F/F_0$; F: raw fluorescence intensity; F_0: average fluorescence across the 45 imaging timepoints, 1.5 s, prior to stimulus presentation) at every pixel of the image stacks for display across the FOV. The $\Delta F/F_0$ for all time courses was calculated from fluorescent traces output from Suite2p. For full sequence trials (ABC, CBA), averaging windows for each pattern (A, B, C) started at stimulation onset of that pattern and lasted 300 ms; shifted for pattern order so the response window always began on the first frame of stimulation of that pattern (raw time course data without shift is shown in Fig. 2b). To isolate the optogenetic response following visual input, we subtracted off the response to the visual stimulus alone, [(vis+opto)-vis] (Fig. 3). Averaged response interval for individual cells began at optogenetic stimulus onset and ended at stimulus offset (120 ms, Fig. 3i). Activity of the C pattern in ABC sequence was calculated as [ABC - AB] (Fig. 4d,e). In Figure 4i-k, the averaging interval for individual cell responses started at stimulus onset and lasted 200 ms. Response intervals for all pixel based analyses started at stimulus onset and lasted 300 ms (Fig. 3b-e, 4a). Deconvolution (Fig. 1k) was done using OASIS with an autoregressive constant of 1[50]. Population sparsity (Fig. 4k) was calculated as[18] :

$$S = \{1 - [(\sum r_i/n)^2/\sum(r_i^2/n]\}/[1 - (1/n)]$$

Statistics

All statistical analyses were performed using Python. Statistical tests used include t-tests (Fig. 2d, Fig. 3i, Fig. 4d-f,i,k), Kolmogorov-Smirnov and Anderson-Darling tests (Fig. 2e,n-o, Fig. 4j), and Mood's median test (Fig. 2q). Significance threshold was held at α=0.05; n.s., not significant ($p>0.05$); *$p\leq0.05$, **$p\leq0.01$, ***$p\leq0.001$. All experiments were replicated in multiple animals.

Cell pattern selection

For analysis of visual response to flashed natural images, individual cell responses were averaged over a 300 ms interval starting from the stimulus onset. Cells activated by the visual stimulus were defined by responses exceeding a 5% $\Delta F/F_0$ threshold. All cells that met this threshold were included in the optogenetic pattern (11-15 cells per pattern, Supp. Fig. 2d,e).

To construct the amplified and attenuated patterns (Fig. 4), we first selected at random (not based on visual responses) the first 15-20 cell pattern in the sequence. Average $\Delta F/F_0$ response to stimulation of that pattern was calculated across a 300 ms time interval starting at stimulation onset. Excited non-target cells meeting a minimum threshold of 5% $\Delta F/F_0$ were included in the next pattern for the amplified sequence (15-20 cells per pattern). Conversely, suppressed non-target cell responses below -5% $\Delta F/F_0$ (15-20 cells per pattern) were selected for the attenuated sequence pattern. If more than 20 cells met these criteria, the 20 cells with the largest magnitude response were used. This process was repeated for the two sets of first and second patterns to generate the third pattern for each sequence. Each pattern in both sequences had an equal number of cells. Any cells included in previous patterns were excluded from selection for the subsequent patterns (Supp. Fig. 3g-j).

Electrophysiology

For Neuropixels electrophysiology the mice were head-fixed for the recordings while seated in plastic support with forepaws on a wheel[51], with the ability to move body parts other than the head but no ability to locomote. The behavioral state of the animals was monitored using two Basler acA2440-75um cameras at 560x560 resolution. One camera focused on the eye area was used to monitor the pupil diameter, and the other focused on the face was used to detect licks and ensure the animal was not in distress. Animals were awake during the duration of the experiment. Recordings were conducted from all layers of the primary visual cortex using Neuropixel 2.4 probes[52] with a sampling rate of 30 kHz. The recording sites were targeted based on the stereotaxic coordinates in the CCF Allen atlas using the Pinpoint system[46]. The reference for the recordings was set at the tip of the electrode. Three recordings were performed: one from the first animal and two (on subsequent days) from the second animal. Spike sorting was executed using Kilosort 2.5 (https://github.com/MouseLand/Kilosort; RRID: SCR_016422; [52]) on each of the four shanks separately.

Visual stimuli were presented on three 60 Hz screens (LG LP097QX1), surrounding the mice and covering 270 x 70 degrees (azimuth x elevation) visual angle. Each trial featured a natural image from a set of two, sourced from the Allen Institute Brain Observatory, and repeated across all screens. These images were equiluminant and underwent histogram equalization for contrast consistency. Images were warped to appear rectilinear from the animals' viewpoint. Each image was displayed 30 times for 50 ms. A reward of 5 µL of 10% sucrose water was administered at 1150 ms post-stimulus offset to maintain engagement. The trials were randomly interleaved with each other and with additional trials from another experiment, which also included natural image presentations and sucrose rewards. The inter-trial interval followed an exponential distribution with a minimum and mean of 2 and 2.6 seconds, respectively.

All electrophysiology analysis was done in Python. Baseline value computed in a 40 ms window ending 5 ms prior to stimulus presentation. Response value was calculated across a 50 ms interval beginning 5 ms after stimulus presentation. Excited cells are those with average response greater than baseline; suppressed cells those with response less than baseline.

Modeling

We trained a recurrent neural network (RNN) consisting of N=500 units, whose input dynamics for the *i*-th neuron are given by:

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} W_{ij}^{rec} \phi(x_i) + I_{im}(t) + \eta_i$$

The input structure is defined by:

.

$$I_{im}(t) = P_{km}^{wf}(t) \cdot P_{ik}^{id} \cdot w_i^{in}$$

Where P^{wf} is the input waveform, $P^{id} = \{0, 1\}$, k is the pattern, and m is the sequence.

The readout of the network is defined as:

$$Z(t) = \sum_{i=1}^{N} w_i^{out} \phi(x_i)$$

The transfer function of single units is $\phi(x) = tanh(x)$. The weights of the input pattern w_{in} are positive and exponentially distributed for a fraction $p = 0.3$ of units, and zero otherwise. The readout weights are an Identity matrix. The initial recurrent weights W_{ij}^{rec}, before any training, are independently sampled from a random Gaussian distribution with a mean zero and

standard deviation $g_0/\sqrt{N}$. The noise term η_i is randomly sampled from a zero mean distribution with standard deviation 0.0005 at every time step.

We trained the recurrent weights W_{ij}^{rec} of the RNN using backpropagation-through-time (ADAM optimizer[53] in PyTorch[54]) such that the network readout Z for designated "trained sequences" matches a sparse scaled version of the time-varying input $I_{im}(t)$ where 30% of the cells have a target that is 2x their input and the other 70% have a target that is 0.5x their input. The network readout Z for "untrained sequences" was set to match the time-varying input without any scaling. Input structure consisted of 5 distinct patterns of units with independently sampled weights from the exponential distribution. Sequences were generated by selecting from a set of 18 distinct input patterns. Untrained sequences were scrambled versions of the numerically ordered trained sequences. The input and output weights remained fixed. We trained the network on 500 epochs to produce these amplified sparse responses. Control training had a target readout that exactly matched the time-varying input for all sequences. Parameters: $\tau = 60\,ms$, $g_0 = 0.8$, Euler integration timestep $\Delta t = 1\,ms$, learning rate 0.01.

References

1. Braitenberg, V. & Schüz, A. *Anatomy of the Cortex: Statistics and Geometry*. (Springer Science & Business Media, 2013).
2. Liu, Y.-J. *et al.* Tracing inputs to inhibitory or excitatory neurons of mouse and cat visual cortex with a targeted rabies virus. *Curr. Biol.* **23**, 1746–1755 (2013).
3. Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. & Suarez, H. H. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).
4. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* **18**, 1025–1033 (2015).
5. Fujisawa, S., Amarasingham, A., Harrison, M. T. & Buzsáki, G. Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* **11**, 823–833 (2008).
6. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
7. Reinhold, K., Lien, A. D. & Scanziani, M. Distinct recurrent versus afferent dynamics in cortical visual processing. *Nat. Neurosci.* **18**, 1789–1797 (2015).
8. Emiliani, V., Cohen, A. E., Deisseroth, K. & Häusser, M. All-Optical Interrogation of Neural Circuits. *J. Neurosci.* **35**, 13917–13926 (2015).
9. Packer, A. M. *et al.* Two-photon optogenetics of dendritic spines and neural circuits. *Nat. Methods* **9**, 1202–1205 (2012).
10. LaFosse, P. K. *et al.* Bicistronic Expression of a High-Performance Calcium Indicator and Opsin for All-Optical Stimulation and Imaging at Cellular Resolution. *eNeuro* **10**, (2023).

11. Bauer, J., Margrie, T. W. & Clopath, C. Movie reconstruction from mouse visual cortex activity. *bioRxiv* 2024.06.19.599691 (2024) doi:10.1101/2024.06.19.599691.

12. Hecht, S. & Verrijp, C. D. The Influence of Intensity, Color and Retinal Location on the Fusion Frequency of Intermittent Illumination. *Proc. Natl. Acad. Sci. U. S. A.* **19**, 522–535 (1933).

13. Evaluation of critical flicker-fusion frequency measurement methods using a touchscreen-based visual temporal discrimination task in the behaving mouse. *Neurosci. Res.* **148**, 28–33 (2019).

14. Umino, Y., Solessio, E. & Barlow, R. B. Speed, Spatial, and Temporal Tuning of Rod and Cone Vision in Mouse. *J. Neurosci.* **28**, 189 (2008).

15. Bonin, V., Histed, M. H., Yurgenson, S. & Reid, R. C. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J. Neurosci.* **31**, 18506–18521 (2011).

16. Smith, S. L. & Häusser, M. Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nat. Neurosci.* **13**, 1144–1149 (2010).

17. LaFosse, P. K., Zhou, Z., O'Rawe, J. F., Friedman, N. G., Scott, V. M., Deng, Y. & Histed, M. H. Single-cell optogenetics reveals attenuation-by-suppression in visual cortical neurons. *bioRxiv* (2023) doi:10.1101/2023.09.13.557650.

18. Vinje, W. E. & Gallant, J. L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).

19. Froudarakis, E. *et al.* Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci.* **17**, 851–857 (2014).

20. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **14**, 481–487 (2004).

21. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).

22. Keller, G. B. & Mrsic-Flogel, T. D. Predictive Processing: A Canonical Cortical Computation. *Neuron* **100**, 424–435 (2018).

23. den Ouden, H. E. M., Kok, P. & de Lange, F. P. How prediction errors shape perception, attention, and motivation. *Front. Psychol.* **3**, 548 (2012).

24. Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).

25. Lindsay, G. The Challenge of Proving Predictive Coding. *Simons Foundation Global Brain* (2021).

26. Furutachi, S., Franklin, A. D., Aldea, A. M., Mrsic-Flogel, T. D. & Hofer, S. B. Cooperative thalamocortical circuit mechanism for sensory prediction errors. *Nature* **633**, 398–406 (2024).

27. O'Rawe, J. F. *et al.* Excitation creates a distributed pattern of cortical suppression due to varied recurrent input. *Neuron* **111**, 4086–4101.e5 (2023).

28. Sanzeni, A. *et al.* Mechanisms underlying reshuffling of visual responses by optogenetic stimulation in mice and monkeys. *Neuron* **111**, 4102–4115.e9 (2023).

29. Ahmadian, Y. & Miller, K. D. What is the dynamical regime of cerebral cortex? *Neuron* **109**, 3373–3391 (2021).

30. Sanzeni, A. *et al.* Inhibition stabilization is a widespread property of cortical networks. *Elife* **9**, (2020).

31. Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* **8**, 183–208 (2000).

32. van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).

33. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature* **571**, 361–365 (2019).

34. Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. Spontaneously emerging cortical representations of visual attributes. *Nature* **425**, 954–956 (2003).

35. Luczak, A., Barthó, P. & Harris, K. D. Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* **62**, 413–425 (2009).

36. Tsodyks, M., Kenet, T., Grinvald, A. & Arieli, A. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* **286**, 1943–1946 (1999).

37. Nguyen, N. D. *et al.* Cortical reactivations predict future sensory responses. *Nature* **625**, 110–118 (2024).

38. Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).

39. Hénaff, O. J., Goris, R. L. T. & Simoncelli, E. P. Perceptual straightening of natural videos. *Nat. Neurosci.* **22**, 984–991 (2019).

40. Hénaff, O. J. *et al.* Primary visual cortex straightens natural video trajectories. *Nat. Commun.* **12**, 5982 (2021).

41. Purandare, C. & Mehta, M. Mega-scale movie-fields in the mouse visuo-hippocampal network. *Elife* **12**, (2023).

42. Murray, J. D. *et al.* A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661–1663 (2014).

43. Graves, A. Generating Sequences With Recurrent Neural Networks. *arXiv [cs.NE]* (2013).

44. Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei. Learning Models are Few-Shot Learners. *Neural Information Processing Systems (NeurIPS)* (2020) doi:10.48550/arXiv.2005.14165 .

45. Gorski, J. A. *et al.* Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).

46. Ye, Z. *et al.* Brain-wide topographic coordination of traveling spiral waves. *bioRxiv* 2023.12.07.570517 (2023) doi:10.1101/2023.12.07.570517.

47. Goldbach, H. C., Akitake, B., Leedy, C. E. & Histed, M. H. Performance in even a simple perceptual task depends on mouse secondary visual areas. *Elife* **10**, (2021).

48. Giovannucci, A. *et al.* CaImAn an open source tool for scalable calcium imaging data analysis. *Elife* **8**, (2019).

49. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* 30 (2017) doi:10.1101/061507.

50. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13**, e1005423 (2017).

51. Burgess, C. P. *et al.* High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Rep.* **20**, 2513–2524 (2017).

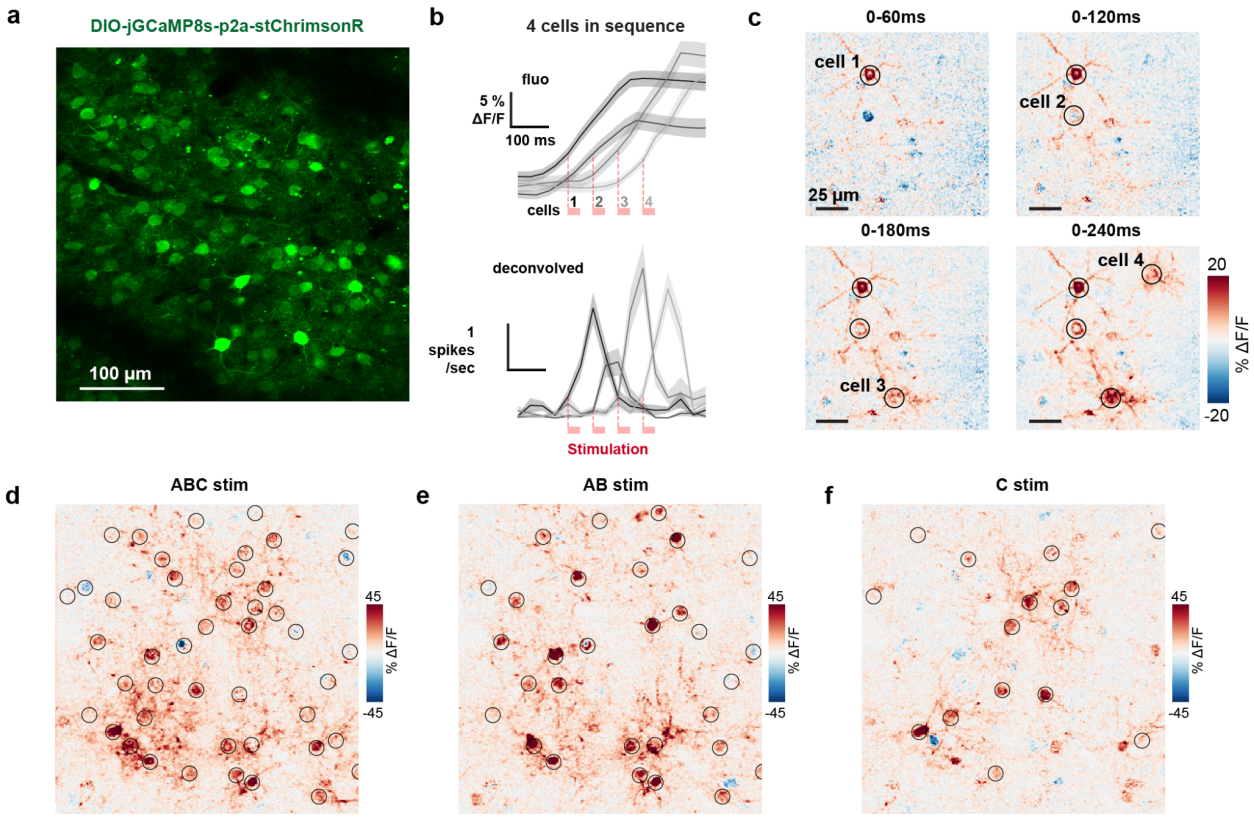
52. Steinmetz, N. A. *et al.* Neuropixels 2.0: A miniaturized high-density probe for stable,

long-term brain recordings. *Science* **372**, (2021).

53. Kingma, D. P. *et al*. Adam: A Method for Stochastic Optimization. *arXiv* (2014).

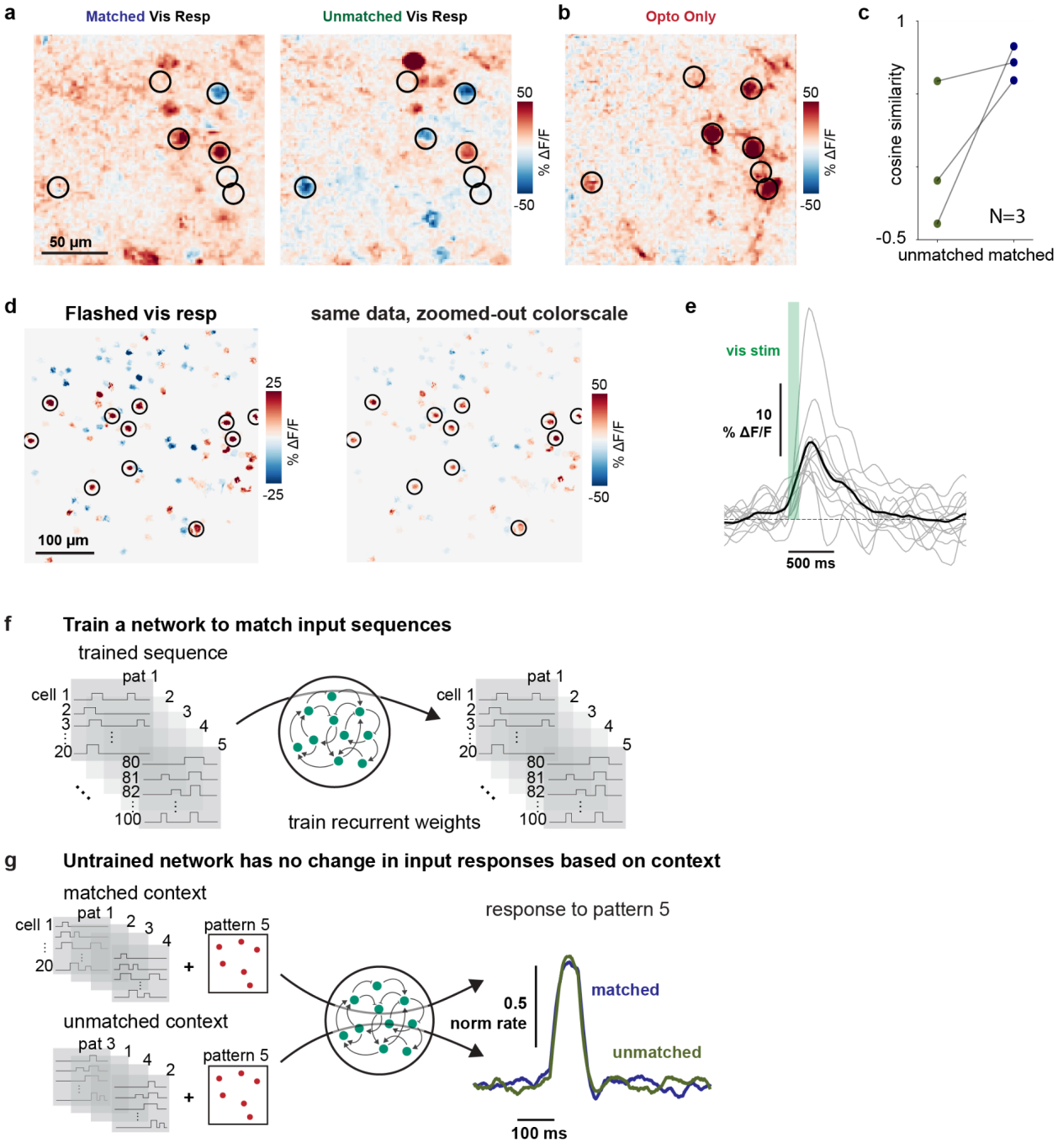
54. Paszke, A. *et al*. PyTorch: An imperative style, high-performance deep learning library. *arXiv* (2019).

Supplemental figures



Supplementary Figure 1: Two-photon stimulation responses: timing and individual patterns.

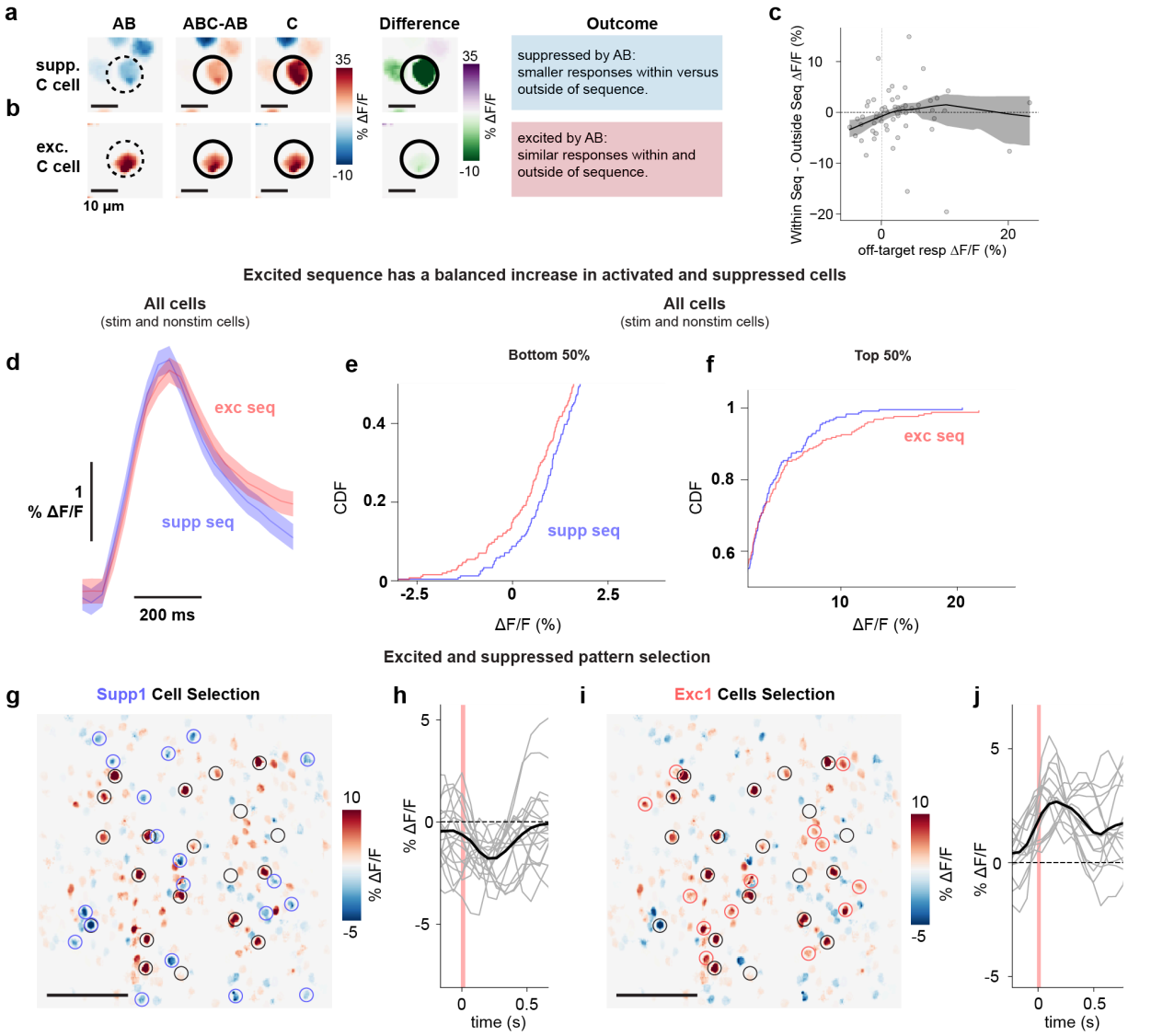
a, Example FOV image in L2/3 V1 with single virus injection, DIO-jGCaMP8s-p2a-stChrimsonR; this uses GCaMP8s which has even faster dynamics than GCaMP7s; more characterization in [10]. **b**, Sequential stimulation of 4 individual cells (30 ms per stim, 30 ms between stim). Fluorescent calcium traces above; deconvolved traces below. **c**, Response maps from data in A. Stimulation pattern lasts 30 ms each in sequence triggered at times: 0 ms, 60 ms, 120 ms, 180 ms. b-c use preparation with expression of indicator and opsin in two viruses, AAV9-syn-jGCaMP7s and AAV9-syn-DIO-stChrimsonR. **d-f**, Response maps of ABC, AB, and C pattern stimulation, using single bicistronic virus (jGCaMP8s, stChrimsonR, Methods). Associated with Figure 1 and 2.



Supplementary Figure 2: Amplification is not due to spike or indicator saturation.

a, Visual response alone of the same cells as shown in Fig. 3 for both matched and unmatched contexts at the time the optogenetic stim would be presented. **b**, Optogenetic response of same cells without prior visual stimulus. **c**, Greater cosine similarity between the optogenetic response alone and the visual responses alone at the time the optogenetic stim would be presented in the matched context than unmatched across experiments. **d-e**, Cell selection for the natural-image-derived optogenetic pattern. **d**, Response to flashed single frame; stimulated cells circled in black. Right: same data as left with

un-zoomed color scale to show the larger cell responses that were chosen for optogenetic stimulation. **e**, Timecourses of the circled cells in **d**. Only cells with a mean response above 5% are included in the pattern (Methods). **f-g**, Control training for the artificial recurrent neural network. **f**, Target output matched the input dynamics, no amplification. **g**, Context dependent changes are eliminated when the model is not trained to selectively amplify sequences. Associated with Figure 3.



Supplementary Figure 3: Constructed amplified and attenuated sequences have similar mean population responses.

a, Prior suppression attenuates the stimulated response. Example non-stimulated suppressed cell in response to AB stim. Dotted black circle: nonstimulated cell; Solid black circle: stimulated cell. **b**, Prior excitation leaves the incremental response largely unchanged. Example non-stimulated excited cell in response to AB stim with same comparison within versus outside the sequence. **c,** More suppression from prior stim leads to stronger attenuation of the cell response within the sequence. Lowess fit of C cell mean responses within vs. outside of the sequence ((ABC-AB) - C) depending on the non-target modulation from prior stimulation (AB). **d**, While the stimulated cells show amplification/suppression (Fig. 4) as expected based on how they were selected, the overall population response mean does not appreciably change. Both stimulated and non-target (non-stimulated) cells. Average response (mean +/- SEM) of all cells in the FOV for the attenuated sequence (blue) and amplified sequence (pink). The mean population

response for the amplified and attenuated sequences is similar. **e-f**, Increased responses in the amplified sequence are balanced by an increased number of suppressed cell responses. Bottom 50% of cell responses for all cells in the FOV (d) for the attenuated sequence (blue) and amplified sequence (pink). Top 50% of cell responses (e). **g-j**, Nonf-target patterns of suppression and excitation are selected for subsequent sequence patterns. **g**, First arbitrary pattern (N=20 cells; pattern A) that was used to initiate sequence design circled in black. Selected suppressed (attenuated) cells circled in blue. This pattern contains cells that are suppressed by prior stimulation. **h**, Timecourses of individual suppressed cells (from f) to A pattern stimulation. Mean trace (black) shows the mean is suppressed; suppressed cells were selected. **i**, Same data as in C with amplified pattern cells shown. Amplified pattern: selected cells (red circles) excited by A stimulation. This pattern contains cells that are activated by prior stim. **j,** Response timecourses of excited cells (f) in response to A pattern stimulation. Mean trace in black. Associated with Figure 4.