

# Distinct recurrent versus afferent dynamics in cortical visual processing

Kimberly Reinhold<sup>1-4</sup>, Anthony D Lien<sup>1-4</sup> & Massimo Scanziani<sup>1-4</sup>

**How intracortical recurrent circuits in mammalian sensory cortex influence dynamics of sensory representation is not understood. Previous methods could not distinguish the relative contributions of recurrent circuits and thalamic afferents to cortical dynamics. We accomplish this by optogenetically manipulating thalamus and cortex. Over the initial 40 ms of visual stimulation, excitation from recurrent circuits in visual cortex progressively increased to exceed direct thalamocortical excitation. Even when recurrent excitation exceeded thalamic excitation, upon silencing thalamus, sensory-evoked activity in cortex decayed rapidly, with a time constant of 10 ms, which is similar to a neuron's integration time window. In awake mice, this cortical decay function predicted the time-locking of cortical activity to thalamic input at frequencies <15 Hz and attenuation of the cortical response to higher frequencies. Under anesthesia, depression at thalamocortical synapses disrupted the fidelity of sensory transmission. Thus, we determine dynamics intrinsic to cortical recurrent circuits that transform afferent input in time.**

Recurrent excitatory circuits are a hallmark of mammalian sensory cortex. How these recurrent circuits within cortex contribute to processing sensory information is still a matter of debate. While progress has been made in understanding how intracortical recurrent circuits combine sensory information in space<sup>1-4</sup>, still very little is known about how they influence sensory-evoked activity in time. Revealing how cortical circuits affect the temporal representation of neural activity has implications for our mechanistic understanding of perception and cognition. The impact of cortical recurrent circuits is difficult to predict. If one population of cortical neurons drives a second population of cortical neurons and that second population also drives the first, self-sustaining activity may result, prolonging the sensory-evoked response in cortex. However, different levels of cortical inhibition, which counteracts cortical recurrent excitation, may lead to different durations of the sensory response in cortex<sup>5-7</sup>. Thus, the following question must be addressed empirically: does the recruitment of cortical recurrent circuits by a sensory stimulus affect the representation of that stimulus in time, and, if so, by what mechanism?

Answering this question has been technically challenging. The cortex receives its primary source of sensory input from the thalamus via thalamic afferents to cortex, and the cortex projects back to thalamus. This reciprocal interaction between thalamus and cortex could alter the cortical representation of sensory activity in time. Hence, to determine how intracortical recurrent circuits themselves affect sensory activity in time, we must factor out the thalamic contribution to the duration and dynamics of the sensory response. Previous attempts to determine how cortical circuits affect sensory activity in time compared the spiking response in thalamus with the spiking response in cortex<sup>8-16</sup>. However, the synapses of thalamic afferents connecting thalamus to cortex exhibit their own dynamic

properties, like frequency-dependent short-term depression, which may affect the sensory response of the cortex<sup>17-20</sup>.

But how can we determine the dynamics of intracortical recurrent circuits in response to sensory stimulation in isolation from the thalamic afferents, given that activity in these thalamic afferents is needed to trigger a response in cortical recurrent circuits? Using the thalamocortical visual system of the mouse as a model, we have developed an optogenetic method to silence, with millisecond precision, visual thalamus at any arbitrary delay following the presentation of a visual stimulus. This approach allows us to follow the duration of visually evoked activity continuing in intracortical recurrent circuits in the absence of thalamic input and thus to directly measure sensory-evoked dynamics intrinsic to recurrent circuits of cortex. To test whether our measurement can predict the temporal transformation of sensory activity by cortex, we then compare the spiking response of cortex with the output of thalamic afferent synapses detected in isolation from the cortical recurrent circuits by silencing these recurrent circuits using a second, complementary optogenetic technique. This multifocal and multifunctional optogenetic approach thus identifies the relative contributions of thalamic afferent versus cortical recurrent connections to the temporal processing of visual information (**Fig. 1a**).

We found that the intrinsic dynamics of cortical recurrent circuits in response to sensory stimulation are on the order of the integration time window of a single cortical neuron. Furthermore, through experiments engaging different optogenetic actuators in different brain areas within the same mouse to silence thalamus while also altering cortical inhibition, we found that cortical inhibition affects these cortical recurrent circuit dynamics. Moreover, we show that, in awake mice, these dynamics of cortical recurrent circuits predict how visual cortex transforms sensory input in time. Thus, the cortical

<sup>1</sup>Neurosciences Graduate Program, University of California San Diego, La Jolla, California, USA. <sup>2</sup>Center for Neural Circuits and Behavior, Neurobiology Section, University of California San Diego, La Jolla, California, USA. <sup>3</sup>Department of Neuroscience, University of California San Diego, La Jolla, California, USA. <sup>4</sup>Howard Hughes Medical Institute, University of California San Diego, La Jolla, California, USA. Correspondence should be addressed to K.R. ([ucsdkimerein@gmail.com](mailto:ucsdkimerein@gmail.com)) or M.S. ([massimo@ucsd.edu](mailto:massimo@ucsd.edu)).

Received 2 July; accepted 30 September; published online 26 October 2015; doi:10.1038/nn.4153

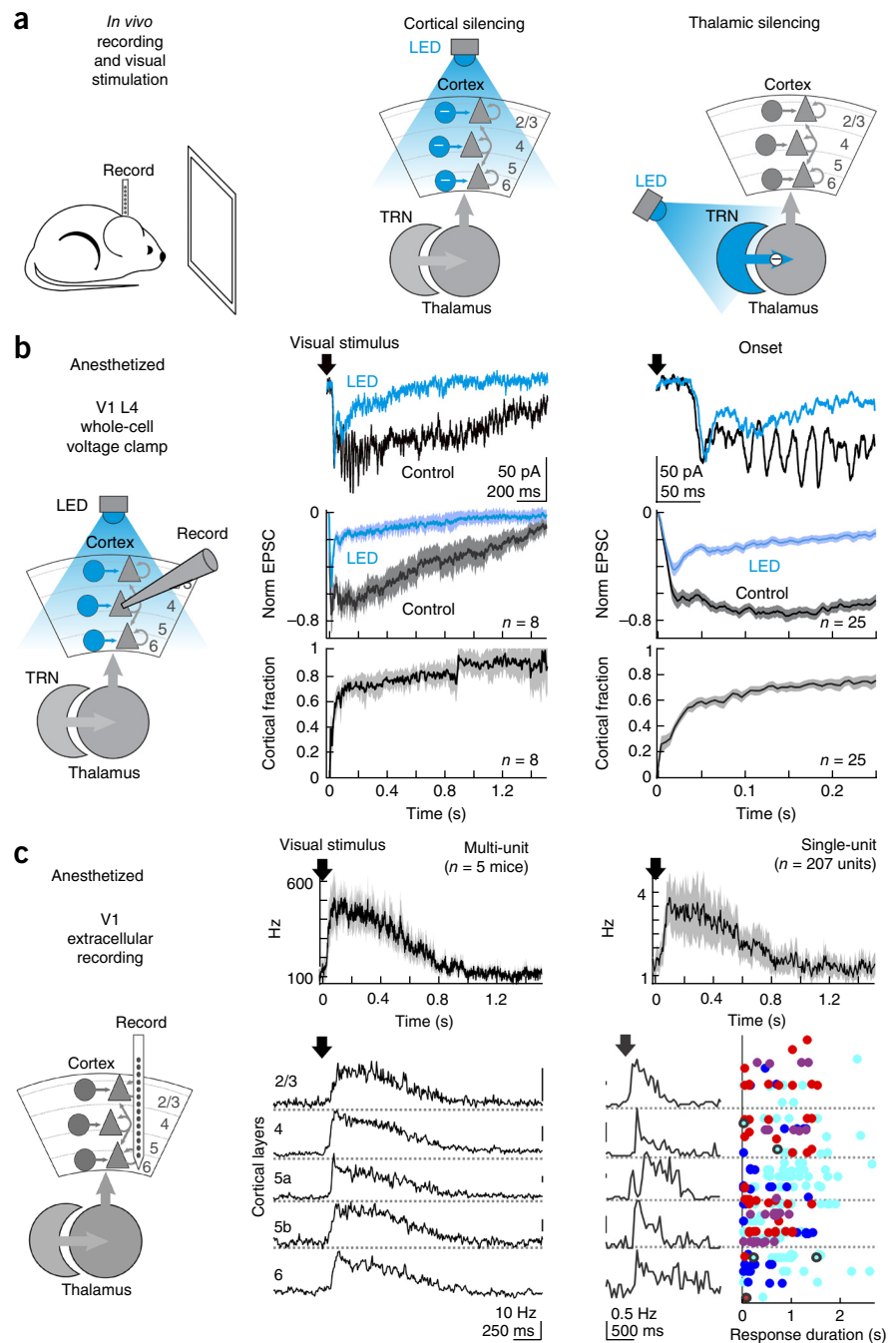
recurrent circuit dynamics limit the rate of the sensory response in the awake state. Notably, our data also indicate that, under anesthesia, the dynamics of synaptic depression at thalamic afferents, rather than the dynamics of cortical recurrent circuits, are rate-limiting in sensory processing, leading to a pronounced disruption of the timing of the sensory response in the nonconscious state. We suggest that the interplay of these mechanisms is a general principle of thalamocortical interaction.

## RESULTS

### Time course of intracortical versus thalamic excitation

To study the impact of intracortical recurrent circuits on the temporal structure of the sensory response, we needed to establish recurrent circuits' contributions to the sensory response over time. To do this, we recorded visually evoked excitation in whole-cell voltage-clamp mode from layer 4 neurons of primary visual cortex (V1) in response to the appearance of a static visual stimulus (for example, an oriented grating) in anesthetized mice. We quantified the contribution of visually evoked excitation mediated by cortical recurrent circuits (cortical fraction) by comparing the excitatory postsynaptic current (EPSC) under control conditions with the EPSC observed on interleaved

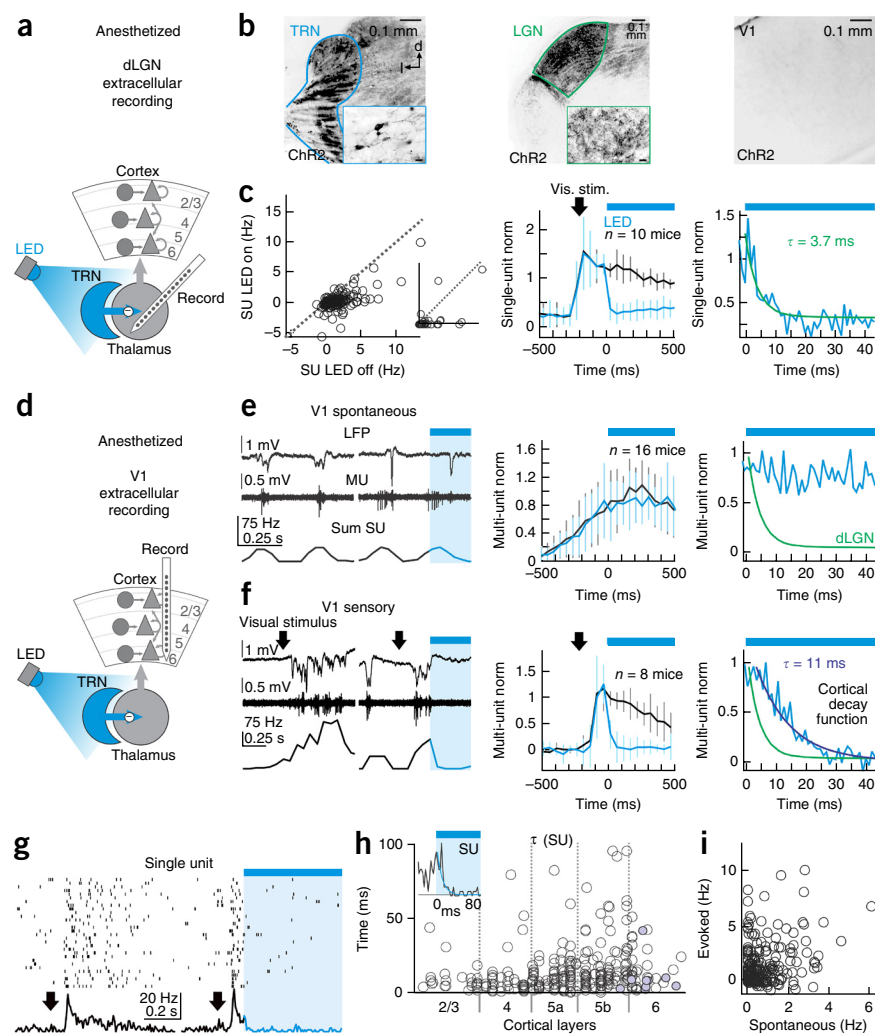
trials in which recurrent excitatory circuits were silenced (Fig. 1b). We silenced recurrent excitatory circuits by optogenetically activating cortical inhibitory interneurons, as described previously<sup>21</sup>. Hence, during this manipulation, the remaining excitation to cortical neurons is direct, feedforward input from thalamus<sup>21</sup>. Silencing recurrent circuits abolished most (79 ± 9%, mean ± s.e.m., as a fraction of the total integrated response,  $n = 8$  cells, 1.7-s-long stimuli; Fig. 1b and Supplementary Fig. 1) visually evoked synaptic excitation to cortical neurons in layer 4, the main thalamic afferent recipient layer<sup>22</sup>, consistent with previous work in rodents and carnivores<sup>21,23,24</sup>. The total synaptic excitatory current, including both cortical recurrent and direct thalamic components, peaked 150 ms after the onset of visually evoked excitation and decayed slowly over the next 1.35 s.



**Figure 1** Time course of cortical recurrent excitation versus thalamic afferent excitation in visual cortex. **(a)** Approach. Center: silencing cortical recurrent circuits by activating ChR2-expressing cortical inhibitory interneurons to isolate thalamic input. Right: silencing thalamus by activating ChR2-expressing TRN to isolate cortical recurrent circuits. **(b)** Cortical recurrent versus thalamic afferent excitation to V1 layer 4 (L4) neurons. Center: top, example EPSC in response to appearance of static grating (arrow, 1.7 s duration). Control (black) versus cortical silencing to isolate thalamic EPSC (blue). Middle, normalized (norm) EPSC mean ± s.e.m. of 8 similar experiments (8 cells, 6 mice). Bottom, fraction of EPSC mediated by cortical excitation (black minus blue, divided by black; mean ± s.e.m.). Right: as middle but expanded timescale at LED onset. Includes shorter stimuli ( $n = 25$  cells from 19 mice, including 8 cells from left and 17 cells for 250-ms static stimuli). **(c)** Left, extracellular multi-channel recordings spanning all V1 layers. Center: multi-unit (MU) response to appearance of grating (static for 3 s). Top, mean ± s.e.m. over 1.5 s; bottom, MU mean subdivided by layer (Supplementary Fig. 1d, dLGN response). Right: top, mean ± s.e.m. single-unit (SU) response to same stimulus. Bottom, example SUs (left) and scatter plot (right) of response duration of SU subdivided by layer and visual stimulus type. Points are units ( $n = 207$  from 5 mice). Fast-spiking outlined ( $n = 5$ ). Cyan, 3-s-long static grating (mean ± s.d. of unit response duration:  $0.96 \pm 0.28$  s,  $n = 120$  from 5 mice); dark blue, 10–100 ms luminance step ( $0.38 \pm 0.25$  s,  $n = 73$  from 8 mice); red, 3-s-long luminance step ( $0.46 \pm 0.24$  s,  $n = 48$  from 3 mice); purple, checkerboard reversal every 3 s ( $0.62 \pm 0.18$  s,  $n = 30$  from 1 mouse).

**Figure 2** Time course of shut-off of visually evoked cortical activity during thalamic silencing. Blue bar and shading, LED on to photo-activate TRN and silence relay thalamus; blue curves, with TRN photo-activation; black, no optogenetic manipulation. Error bars, s.d. across mice.

(a) Experimental configuration for c. (b) TdTomato-ChR2 fluorescence, coronal sections. TRN (blue)  $-1.58$  mm; dLGN (green)  $-2.46$  mm; V1  $-3.08$  mm A–P; d, dorsal; l, lateral. Insets: close-ups (scale bars,  $5 \mu\text{m}$ ). Histology quantified in 16 mice (Supplementary Fig. 3). (c) dLGN activity. Left: single-unit (SU) spiking (baseline-subtracted,  $n = 157$  from 12 mice) during visual stimulation (3-s-long grating) in control (LED off) or plus TRN photo-activation (LED on) (other visual stimuli, Supplementary Fig. 2a). Points below broken line (unity) are suppressed. Inset, spontaneous activity (not baseline-subtracted; x, y axes 0 to 7 Hz;  $n = 38$  from 3 mice). Center: mean normalized (norm; Online Methods) dLGN visual response followed by TRN photo-activation (blue) versus control (black). Right: expanded timescale at LED onset. Green, exponential fit. (d) Experimental configuration for e–h. (e) V1 spontaneous Up states. Left: example single-trial LFP, multi-unit (MU) and SU with (blue shading) or without (unshaded) thalamic silencing. Center: mean MU Up state beginning within 300 ms of thalamic silencing (blue) versus control (black). Right: expanded timescale; green from c. (f) Data presented as in e, but for V1 response to appearance of 3-s-long static visual stimulus (arrows). Right: cortical decay function (light blue) and exponential fit (dark blue); green from c. (g) Example SU. (h) Time constants of SU shut-offs ( $n = 297$  from 18 mice). Filled circles, fast-spiking units ( $n = 9$ ; rank-sum, fast-spiking versus regular-spiking,  $P = 0.16$ ). (i) Spontaneous versus sensory-evoked V1 SU firing rates ( $n = 208$  from 15 mice).



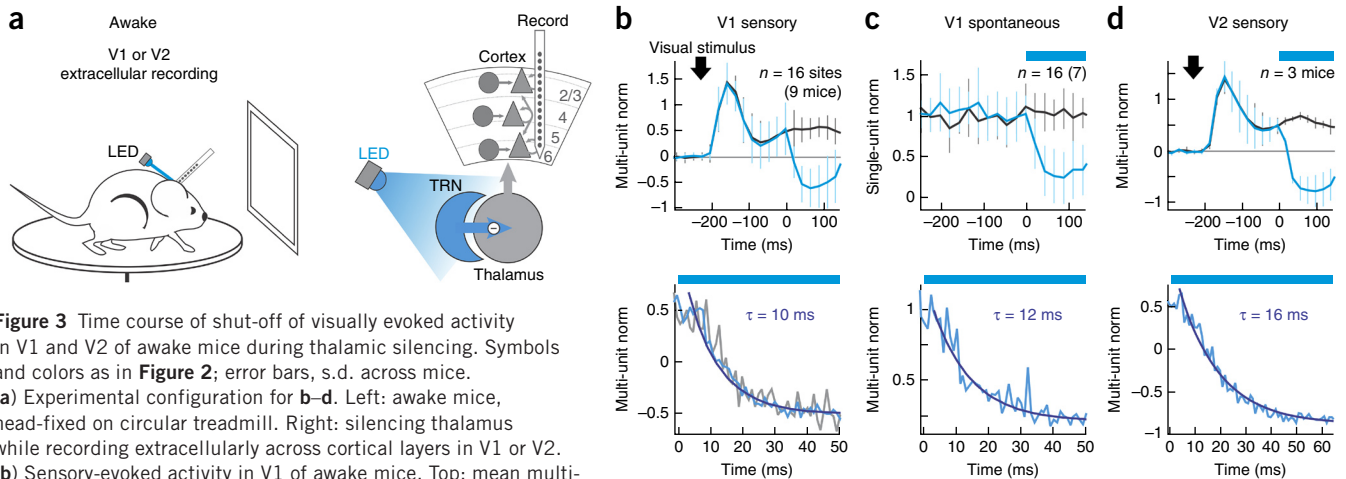
The direct thalamic excitation to layer 4 was dominant over the first several tens of milliseconds of the response (time until thalamic fraction was less than 50% of total instantaneous EPSC amplitude:  $43 \pm 6.5$  ms, mean  $\pm$  s.e.m.,  $n = 25$ ; over the first 10 ms,  $82 \pm 4\%$  of excitation was of direct thalamic origin, as a fraction of the total integrated response; Fig. 1b). The fraction of excitation dependent on cortical recurrent circuits (cortical fraction) grew progressively from 0% to  $72 \pm 6\%$  (as a fraction of the instantaneous EPSC amplitude,  $n = 25$ , including 1.7-s-long and 250-ms-long stimuli) over the first 250 ms of the response, overwhelming the direct thalamic component. Thus, by 250 ms, cortical circuits amplified the thalamic input more than threefold. Therefore, in response to visual stimulation, the contribution of cortical recurrent excitation progressively builds up to overwhelm the contribution of direct thalamic excitation to visual cortex.

If the cortical neurons in recurrent circuits are the source of the cortical component of the excitatory synaptic current, the spiking of neurons in visual cortex should follow a time course similar to that of cortical synaptic excitation. Over the first 800 ms of the response, multi-unit and single-unit activity across all cortical layers detected with extracellular linear probes (Online Methods) had a time course comparable to that of recurrent synaptic excitation (Fig. 1c). Thus, powerful recurrent excitation from other spiking cortical neurons is larger in magnitude than the direct thalamic input after the first several tens of ms of the evoked response in V1.

### Silencing thalamic input to visual cortex

Once recurrent excitation builds up in V1, over what timescale do cortical recurrent circuits alone sustain the sensory response in the absence of continued thalamic input? To answer this, we needed to precisely gate thalamic input to cortex. We developed an optogenetic approach to rapidly silence thalamic input at any arbitrary time following the onset of a sensory response in visual cortex. Unlike electrically or optogenetically<sup>25</sup> stimulating thalamic afferents, which recruits a random population of afferents, our method allowed us to measure the duration of activity reverberating through cortical recurrent circuits in response to actual visual stimuli.

To gate thalamic input, we exploited the inhibitory projection from the thalamic reticular nucleus (TRN), which does not itself project to cortex, onto relay nuclei of thalamus (such as the dorsal lateral geniculate nucleus (dLGN)), which transmit sensory activity to cortex. We conditionally expressed Channelrhodopsin 2 (ChR2), a light-activated cation channel, conditionally on Cre recombinase expression in the TRN<sup>26</sup> using Gad2-Cre mice (expressing Cre in GABAergic cells; Fig. 2a,b and Online Methods). Photo-activating the TRN (step pulse, 470 nm) rapidly and powerfully suppressed visually evoked (>96%) and spontaneous (>88%) activity in the dLGN of anesthetized mice (Fig. 2c; 3.7 ms time constant of dLGN silencing; for duration of silencing, see Online Methods). TRN photo-activation also suppressed activity



**Figure 3** Time course of shut-off of visually evoked activity in V1 and V2 of awake mice during thalamic silencing. Symbols and colors as in **Figure 2**; error bars, s.d. across mice.

(a) Experimental configuration for **b–d**. Left: awake mice, head-fixed on circular treadmill. Right: silencing thalamus while recording extracellularly across cortical layers in V1 or V2.

(b) Sensory-evoked activity in V1 of awake mice. Top: mean multi-unit (MU) activity in V1 in response to appearance of visual stimulus (arrow, 3 s duration) in control (black) or followed by thalamic silencing (blue). MU is baseline-subtracted and normalized (norm) to first 150 ms of control evoked response (Online Methods for stimuli). Gray line, pre-stimulus baseline. Bottom: expanded timescale showing the decay of cortical activity at LED onset (the CDF) during static (gray) or moving (light blue) gratings. Dark blue, single exponential. Data include both running and non-running epochs (**Supplementary Fig. 4c** for breakdown).

(c) As in **b**, but for spontaneous activity in V1 (not baseline-subtracted). (d) As in **b**, but for sensory-evoked activity recorded in V2.

in the lateral posterior nucleus, the second-order visual thalamic nucleus in rodents (**Supplementary Fig. 2a**).

### Decay of evoked cortical activity after silencing thalamus

We measured multi-unit activity in visual cortex of anesthetized mice after rapidly silencing thalamus. Silencing thalamus in the absence of visual stimulation had no effect on cortical spontaneous active states, called Up states (**Fig. 2d,e** and **Supplementary Fig. 2b**)—bouts of high activity occurring spontaneously in the anesthetized cortex—indicating that cortical recurrent circuits can generate and sustain spiking activity for several hundreds of milliseconds in the absence of thalamic input, consistent with previous work<sup>27,28</sup>. We triggered sensory-evoked active states in cortex using the static visual stimuli described above (as in **Fig. 1**). These sensory-evoked active states decayed slowly after removal of the visual stimulus, continuing for several hundreds of milliseconds (**Fig. 1** and **Supplementary Fig. 1d**). If cortical recurrent circuits sustain this slowly decaying activity, then silencing thalamus should have little effect on the time course of this decay. Notably, silencing thalamus 250 ms into the sensory-evoked response led to a decay of visually evoked activity in V1 that was more than two orders of magnitude faster than the decay observed following simply the removal of the visual stimulus (**Fig. 2f** and **Supplementary Fig. 2**; compare **Fig. 1**). We call the time course of this decay of cortical activity after silencing thalamus the cortical decay function (CDF). The CDF was fit by an exponential time constant of  $9 \pm 3$  ms (mean  $\pm$  s.d.;  $n = 26$  mice; **Fig. 2f**), with 3 ms delay before the start of the exponential to account for the time it takes thalamus to shut off (Online Methods, **Supplementary Fig. 2c**). Thus, even after the build-up of cortical recurrent excitation in response to sensory stimulation, cortical recurrent circuits sustain sensory-evoked activity for only a few tens of milliseconds without thalamic input.

We verified that this fast CDF was not influenced by off-target expression of ChR2 in visual cortex interneurons and that ChR2 expression in TRN axons to dLGN was the main histologic correlate of the shut-off of visually evoked activity in V1 (**Supplementary Fig. 3**, Online Methods). The CDF was fast across all V1 layers, even those that do not receive the main thalamocortical input (**Supplementary Fig. 2d** and **Supplementary Table 1**), although we did observe subtle but significant differences across layers (mean fit to CDF

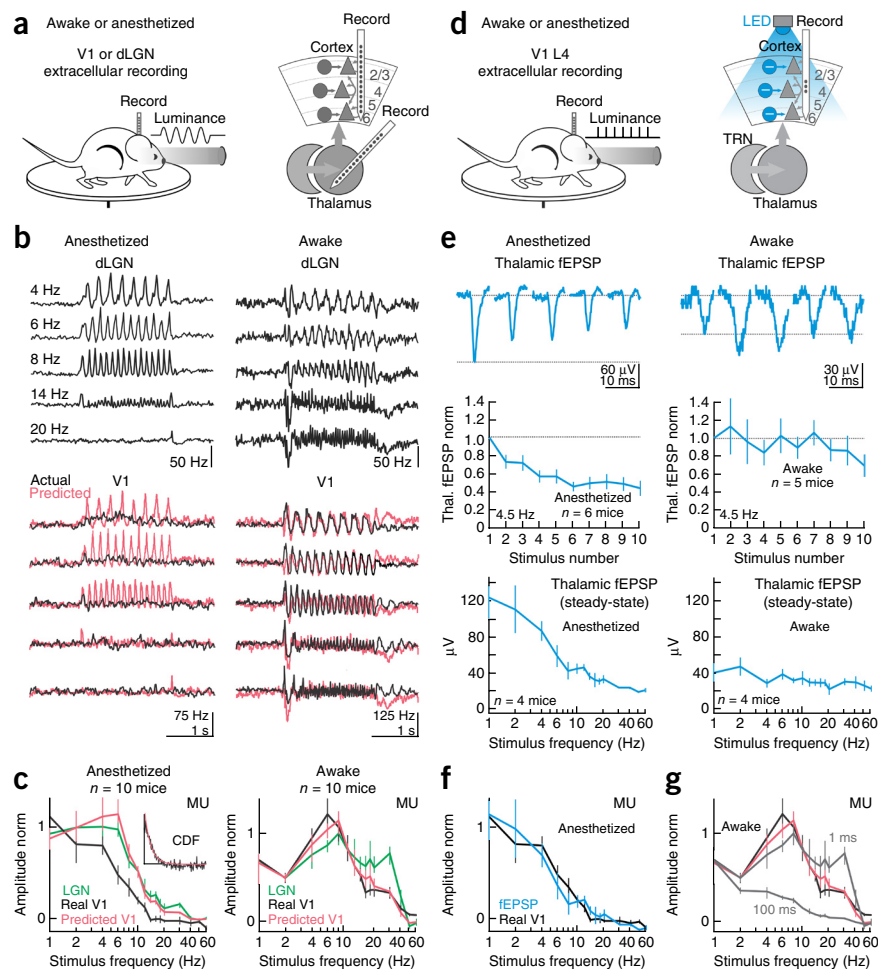
$\pm$  s.e.m. subdivided by layer, L2/3:  $9.8 \pm 1.7$  ms, L4:  $9.0 \pm 2.2$  ms, L5a:  $8.9 \pm 1.3$  ms, L5b:  $15.7 \pm 2.5$  ms, L6:  $7.6 \pm 1.5$  ms;  $P$  values in **Supplementary Table 1**). Furthermore, although it has been suggested that slow cortical sensory-evoked dynamics emerge at low contrast<sup>29</sup>, we found that the CDF was independent of stimulus contrast (**Supplementary Fig. 1e**). The CDF was also the same at later time points in the visual response (300, 600 and 900 ms after stimulus onset; **Supplementary Fig. 1e**), demonstrating that cortical sensory-evoked activity locks to thalamic input over the full time course of the sensory response. Moreover, the CDF was independent of strength of the thalamic silencing (**Supplementary Fig. 1e**), indicating that cortex exhibits the same fast sensory-evoked dynamics even when partial thalamic drive remains.

Consistent with the multi-unit data, the average CDF of visually responsive, isolated single units in V1 was fit by a  $12 \pm 1$  ms time constant (mean  $\pm$  s.e.m.,  $n = 297$  units from all cortical layers; **Supplementary Fig. 1f,g**). Furthermore, most (>90% of) single units considered individually had fast CDFs (time constant < 20 ms; **Fig. 2g,h** and **Supplementary Fig. 2**), and only a small minority in layer 5 were a few tens of ms slower. Are the neurons that exhibit spontaneous Up states part of a different subnetwork than neurons responsive to visual stimulation? No, many of the visually responsive units also participated in Up states, indicating that thalamus-locked sensory activity and thalamus-independent Up states engage many of the same neurons in the cortical circuits (**Fig. 2i**), despite marked differences in the dynamics of these two types of cortical activity.

Taken together, these results show that, even at time points when cortical recurrent circuits generate the majority of visually evoked synaptic excitation (that is, more than 43 ms after the onset of the visually evoked response; **Fig. 1**), activity in cortical recurrent circuits time-locks to thalamic input. Hence, cortical recurrent circuits have fast dynamics.

Previous work suggests that dynamics in recurrent circuits might be very different between the anesthetized and awake brain states<sup>27,30,31</sup> and between primary and higher order cortical areas<sup>32</sup>. We found that the CDF was similar in awake mice (**Fig. 3a,b** and **Supplementary Fig. 4**; awake:  $10 \pm 3$  ms, mean  $\pm$  s.d.,  $n = 9$  mice; anesthesia:  $9 \pm 3$  ms,  $n = 26$  mice; unpaired  $P = 0.15$ —all CDF comparisons in text are two-tailed  $t$ -tests; see **Supplementary Fig. 4d** for CDF distribution

**Figure 4** Relative contributions of thalamocortical synaptic depression and recurrent circuit dynamics to V1's response in anesthetized versus awake mice. In **b,c,e,f**: left, anesthetized; right, awake. See Online Methods for curve calculation, normalization and alignment. **(a)** Experimental configuration for **b,c**. **(b)** Multi-unit (MU) response in dLGN (top) or V1 (bottom) to visual flicker stimuli of different temporal frequencies (from different example mice at top and bottom; see **c** for summary across all 10 mice). Black, MU response; pink (bottom), prediction of V1's response by convolving CDF with dLGN responses (Online Methods). **(c)** Normalized (norm) response amplitude as function of stimulus frequency. Green, dLGN median SU (anesthetized  $n = 131$  from 4 mice; awake  $n = 80$  from 3 mice; error bars, 45th–55th percentiles; **Supplementary Fig. 5b** for full distributions). Black, V1 MU mean  $\pm$  s.e.m. Pink, CDF-based prediction of V1 response. Pink error bars calculated from green error bars. **(d)** Experimental configuration for **e**. Cortical silencing to isolate thalamic (Thal.) fEPSP detected in V1 L4. Visual stimulus, train of full-field flashes. **(e)** Top: average (30 trials) thalamic fEPSP response to 4.5-Hz visual stimulus train (each stimulus 10 ms duration) under anesthesia (left) and awake (right; same animal, same recording site). Middle: mean  $\pm$  s.e.m. thalamic fEPSP amplitude versus stimulus number. Amplitudes normalized to first response and spike rate adaptation in thalamus (Online Methods). Bottom: mean  $\pm$  s.e.m. steady-state thalamic fEPSP amplitude as function of visual stimulus train frequency ( $n = 4$  same mice across brain states). **(f)** Baseline-subtracted, normalized steady-state thalamic fEPSP (blue from **e**, bottom left) overlaid on V1's frequency response under anesthesia (black from **c**, left). Error bars from **c** and **e**. **(g)** Black and pink from **c**, right. Gray: failed predictions of V1 response using 1-ms or 100-ms time constant instead of real CDF. Black and pink error bars from **c**. Gray error bars calculated from green error bars in **c**, right.



and discussion of normality; *t*-tests assumed that CDF distribution was normal and that variances were equal across groups, but this was not formally tested), even when mice were running, a behavior that increases the gain of V1's response<sup>33</sup> (running:  $9 \pm 6$  ms, stationary:  $10 \pm 6$  ms,  $n = 4$  mice; paired  $P = 0.16$ ; **Supplementary Fig. 4c**). Furthermore, in contrast to Up states under anesthesia, spontaneous cortical activity in awake mice decreased after silencing thalamus ( $73 \pm 7\%$  decrease in single-unit baseline, mean  $\pm$  s.e.m.; **Fig. 3c** and **Supplementary Fig. 4b**) at a rapid rate consistent with the CDF (**Fig. 3c**; awake spontaneous CDF:  $14 \pm 4$  ms,  $n = 7$  mice), indicating that spontaneous cortical activity in awake mice relies on ongoing thalamic input. Also, the CDF was only a few milliseconds slower in a higher order area of visual cortex (**Fig. 3d**; awake V2 lateral, Online Methods; V2 CDF:  $16 \pm 5$  ms,  $n = 3$  mice, unpaired versus V1  $P = 0.003$ ). Thus, visually evoked activity in cortical recurrent circuits locks to the timing of the thalamic input even in awake conditions and secondary visual areas (**Supplementary Fig. 4** and **Supplementary Table 1**).

#### Thalamo-cortical depression filters anesthetized response

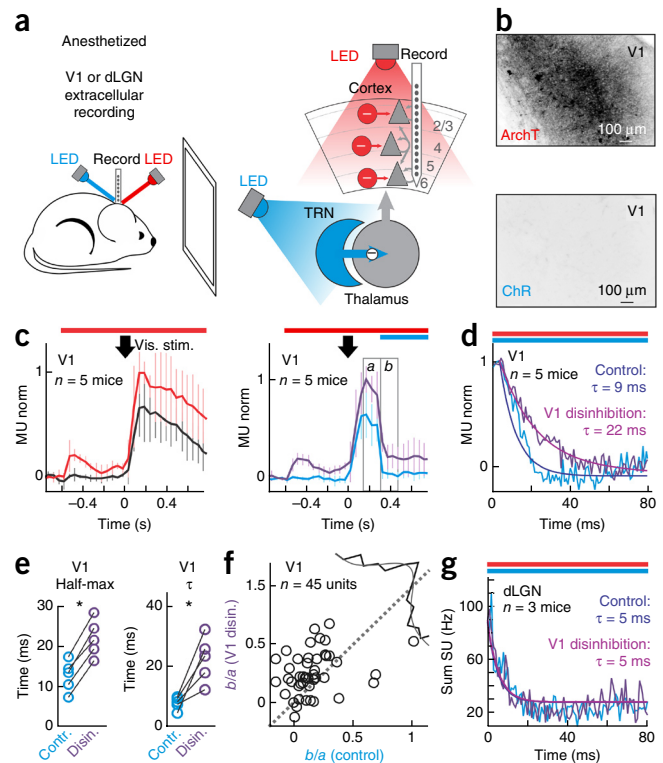
What are the implications of such a consistent  $\sim 10$  ms CDF for how cortical recurrent circuits influence the timing of sensory activity in visual cortex? If this fast CDF sets the rate of the cortical

sensory response, then cortex should precisely follow fast fluctuations in sensory activity up to  $\sim 15$  Hz but increasingly attenuate frequencies above  $\sim 15$  Hz, the frequency cutoff predicted by the filtering properties of the  $\sim 10$  ms CDF (Online Methods).

We determined V1's ability to follow the timing of sensory activity from thalamus by measuring the spike rate response to spatially uniform, full-field fluctuations of luminance (flicker) as a function of flicker temporal frequency between 1 and 60 Hz, called the frequency response, in dLGN and V1 (**Fig. 4a**; Online Methods). In contrast to the prediction by the CDF, however, under anesthesia V1 could not follow visually evoked thalamic activity at frequencies above  $\sim 6$  Hz (**Fig. 4b,c**, left). This low-pass filtering of the sensory response by V1 is consistent with previous measurements of V1's frequency response<sup>34</sup> but implies a rate-limiting step between thalamus and cortex markedly slower than our  $\sim 10$  ms CDF. This low-pass filtering was already pronounced in layer 4 (**Supplementary Fig. 5a**), V1's main thalamo-recipient layer, indicating that the filtering of the sensory response occurs very early, as visually evoked activity enters the cortex.

A candidate source of low-pass filtering of the sensory response under anesthesia is frequency-dependent depression of thalamo-cortical (TC) afferent synapses<sup>17,18,20,35,36</sup>, a progressive depression of synaptic strength as the frequency of thalamic activity increases.

**Figure 5** Cortical decay function is modulated by inhibition. Blue bar, thalamic silencing; red bar, cortical disinhibition. (a) Simultaneous thalamic silencing by blue LED and suppression of ArchT-expressing cortical inhibitory interneurons by amber (red in graphics) LED. (b) ArchT-GFP expression in V1. No V1 ChR2-TdTomato expression. Histology of all mice in this figure matches this profile. We included the 5 mice with the highest expression of ArchT in visual cortex in this figure. See **Supplementary Figure 9** for results from mice with weak ArchT expression. (c) Left: V1 mean  $\pm$  s.d. multi-unit (MU) response to appearance of visual stimulus (3-s-long grating, arrow) with (red) or without (black) cortical disinhibition (disin.). Right: During the same experiment, on interleaved trials, thalamus was silenced while cortex was (purple) or was not (blue) disinhibited. MU normalized to window surrounding peak response in red. *a* and *b* refer to *f*. (d) CDF during normal (blue) versus reduced (purple) cortical inhibition, with exponential fits. (e) Prolonged CDF quantification. Time to-half-max (left) or time constant of fit (right). Each pair of points is one mouse. Two-sample *t*-tests (here and in *f*, we assume normally distributed data and equal variances across groups, unless otherwise specified, but these assumptions were not formally tested): time to half-max,  $*P = 0.0002$ , assuming equal variances;  $\tau$ ,  $*P = 0.01$ , assuming unequal variances. (f) Prolonged shut-off of single-unit (SU) population ( $n = 45$  from 5 mice) during cortical disinhibition as ratio of mean SU firing rate in window *b* over window *a* (c, right) (**Supplementary Fig. 10b** for SU examples). Bias toward points above broken line (unity) indicates significant prolonged SU shut-off ( $P = 0.0068$ , two-sided paired *t*-test). Black curve is histogram of deviation from unity (shortest distance between each point and broken line), with Gaussian fit in gray. (g) As in *d*, but dLGN shut-off. dLGN recordings performed in only these three of the five mice in *d*.



If present, this process would preferentially attenuate the cortical response to high-frequency thalamic activity. We measured TC depression in our preparation *in vivo* under anesthesia (**Fig. 4d**). To do this, we optogenetically silenced cortical recurrent excitatory circuits, as in **Figure 1**, while recording field excitatory postsynaptic potentials (fEPSPs) in layer 4, which under these conditions represent TC transmission (**Fig. 1** and described previously<sup>21</sup>). These TC fEPSPs showed strong frequency-dependent depression, quantified as the decrease in response amplitude over a train of visual stimuli (each stimulus a flash of 10 ms duration, train of 10 stimuli) relative to the amplitude of the first response (normalized by the response amplitude in thalamus; see Online Methods for details; **Fig. 4e**, left). Notably, the measured frequency dependence of the TC fEPSPs accounted quantitatively for V1's low-pass filtering of thalamic input under anesthesia (**Fig. 4f**).

If this TC depression underlies the low-pass filtering of V1's response under anesthesia, conditions reducing TC depression should improve V1's ability to follow high-frequency thalamic input. We observed that frequency-dependent TC depression in layer 4 disappeared as animals woke up from anesthesia (**Fig. 4e**, right; for a potential mechanism, see **Supplementary Fig. 6**) (refs. 17,18,20,36). Thus, if TC depression impairs V1's ability to follow high-frequency stimulation under anesthesia, then in awake mice, relief from TC depression should enable V1 to better follow high-frequency thalamic activity, up to the cutoff predicted by the CDF.

### CDF predicts amplitude of visual responses in awake mice

Waking the animals led to a markedly stronger V1 response to thalamic activity at temporal frequencies above  $\sim 6$  Hz (**Fig. 4b,c**, right). Furthermore, consistent with our previous measurements of recurrent circuit dynamics in visual cortex, the response attenuation in V1 in awake mice precisely matched the attenuation between thalamus and cortex predicted by the CDF (**Fig. 4b,c**, right). Combining (convolving) the CDF with dLGN's frequency response successfully

predicted V1's frequency response measured in terms of amplitude during both high (100%) and low (20%) contrast visual flicker (**Fig. 4c**, right, and **Supplementary Fig. 5**; for additional filtering properties of CDF, including subtle theta-band amplification<sup>37</sup>, also observed in cats<sup>31</sup>, see **Supplementary Fig. 7**). In awake mice, the cortex exhibited increasing attenuation of sensory-evoked thalamic activity above  $\sim 15$  Hz, attenuating two-thirds of the amplitude of the thalamic input at 30 Hz, in accordance with the CDF prediction. Notably, combining dLGN's frequency response with a time constant ten times faster or slower than the 10 ms fit to the CDF failed to predict V1's frequency response (**Fig. 4g**).

Thus, in awake mice, the dynamics of recurrent circuits predict how V1 attenuates the amplitude of a temporally modulated thalamic input. But what fraction of V1's total response in awake mice is actually explained by this simple model of input attenuation? To answer this precisely, we performed dual simultaneous extracellular recordings in dLGN and V1 to measure response coherence between these structures. We found that dLGN's average multi-unit response convolved with the CDF explained 53% of V1's average multi-unit response between 1 and 30 Hz (**Supplementary Fig. 5f**). Furthermore, both fundamental (at the temporal frequency of the stimulus, 65% of total V1 response amplitude) and first-harmonic (twice the temporal frequency of the stimulus, 35% of total) components of the visual response were attenuated at high frequencies in V1 regular-spiking units in a manner quantitatively consistent with filtering by the CDF (**Supplementary Fig. 8**). Thus, these data suggest that, in awake mice and therefore in the absence of low-pass filtering by thalamic afferent synapses, dynamics of cortical recurrent circuits are the key constraint on V1's response to sensory activity from thalamus. As further validation of how brain state affects the temporal transformation of sensory activity between thalamus and cortex, deconvolution of dLGN's spiking response to the flicker stimulus from V1's response, an analysis that measures the combined effects of both TC afferent synapse and cortical recurrent circuit dynamics, showed slow

dynamics under anesthesia but faster dynamics in the awake state (**Supplementary Fig. 5c**).

### Cortical decay function depends on inhibition

What mechanism regulates the dynamics of recurrent circuits to prevent prolonged reverberation of cortical activity after thalamic input ends? Certain models<sup>5,7</sup> propose that cortical inhibition balances strong recurrent excitation to damp sensory activity. To test the role of cortical inhibitory interneurons in setting the fast CDF, we expressed the light-activated proton pump ArchT<sup>38</sup> in V1's inhibitory interneurons (Cre-dependent ArchT, Gad2-Cre mice) to suppress their activity by application of amber (595 nm) light (**Fig. 5a,b**). During the suppression of cortical GABAergic interneurons, we photo-activated the ChR2-expressing TRN in the same mice to silence the thalamus (473 nm light from the fiber above TRN). We performed these experiments in anesthetized mice because disinhibition of cortex in awake mice led to runaway activity<sup>39</sup>.

Suppressing cortical interneurons on interleaved trials increased the gain of V1's evoked response by 47% (**Fig. 5c**), consistent with previous reports<sup>39</sup>, and slowed the CDF by 263% (**Fig. 5d–f** and **Supplementary Figs. 9** and **10a–c**; single exponential fit changes from  $8 \pm 2$  ms to  $21 \pm 7$  ms;  $n = 5$  mice, paired  $P = 0.01$ ). This lengthening of the CDF originated in cortex, because suppressing cortical interneurons did not affect the rate of dLGN silencing (paired  $P = 0.78$ ; **Fig. 5g**) or cortical sensitivity to thalamic input differentially at low levels of input (**Supplementary Fig. 10d**). Furthermore, this lengthening of the CDF was not due to increased gain of the V1 response *per se*, because under conditions of normal cortical inhibition the CDF is not affected by the gain of the V1 response (that is, across stimulus contrasts or behavioral states; **Supplementary Figs. 2e** and **4** and **Supplementary Table 1**). Hence, cortical inhibition contributes to enforcing fast recurrent circuit dynamics, which are essential to lock the duration of recurrent excitation to the timing of thalamic input.

### Prolonging the CDF predictably alters V1's visual response

The CDF predicts that V1 can readily follow thalamic frequencies up to ~15 Hz but attenuates thalamic frequencies above ~15 Hz. If V1's ability to rapidly follow high-frequency thalamic input is constrained by the CDF, prolonging the CDF should reduce V1's response to high-frequency fluctuations of thalamic activity. To test a causal role for the CDF in rate-limiting V1's response to high-frequency input, we prolonged the CDF by suppressing cortical interneurons in visual cortex of anesthetized mice, as described above. Because in anesthetized mice TC depression attenuates the cortical response to repeated high-frequency visual stimulation (**Fig. 4**), we tested the effects of prolonging the CDF at the onset of high-contrast visual stimulation, when thalamic activity increases rapidly and before TC depression develops. Prolonging the CDF lengthened V1's onset response to an extent predicted by the measured change in the CDF (**Supplementary Fig. 10e**). Control experiments to change the gain of V1's response without prolonging the CDF did not lengthen V1's onset response (**Supplementary Fig. 10f**). Thus, changing the CDF changes cortical dynamics. Taken together, these results show that the CDF of visual cortex is fast enough to temporally lock the cortical response to the timing of sensory-evoked thalamic activity.

### DISCUSSION

Cortical dynamics are fundamental to sensation and cognition. These dynamics are sculpted by intracortical recurrent circuits, which dominate the anatomy of both associative and sensory cortical areas<sup>22,40</sup>.

In sensory areas, these recurrent circuits provide the majority of sensory-evoked excitation to cortical neurons<sup>21,23,24</sup>. How these recurrent circuits influence the duration and dynamics of the response in sensory areas is still not known. Here we measure directly the dynamics intrinsic to recurrent circuits in visual cortex in isolation from dynamics at thalamic afferents. We demonstrate that the dynamics of these cortical recurrent circuits are on the order of the integration time window of a single cortical neuron (~10 ms)<sup>41</sup>. Furthermore, we establish that these dynamics are a fundamental constraint on the temporal response of cortex to visual stimulation (**Supplementary Fig. 11**) and precisely predict the amplitude attenuation of the cortical response in awake mice over a wide range of stimulus frequencies.

By optogenetically dissecting intracortical and thalamic contributions to the sensory response, we found that the prolonged timescale of continuing sensory-evoked activity in visual cortex after removal of the sensory stimulus is a product, not of cortical recurrent circuits, but of thalamocortical communication. Therefore, thalamocortical interactions are necessary for long-lasting responses in visual cortex. In higher cortical areas, even longer-lasting sensory responses have been observed<sup>32,42</sup> and are often attributed to intrinsic dynamics of cortical recurrent circuits<sup>43,44</sup>. Our work thus raises the possibility that long-lasting responses in higher cortical areas arise from thalamocortical interactions rather than cortical recurrent circuits<sup>45</sup>. However, consistent with our suggestion that inhibition may modulate intrinsic dynamics of cortical recurrent circuits, it is also plausible that different relative properties of inhibition and excitation lead to different intrinsic dynamics of the recurrent circuits in these higher cortical areas. Our approach will allow researchers to measure these intrinsic dynamics across cortical areas.

Our results reveal that thalamocortical communication is necessary to sustain sensory-evoked activity in visual cortex over timescales longer than a few tens of milliseconds. The duration of the response in thalamus may be set by input from peripheral sensory structures<sup>46</sup> or by corticothalamic feedback<sup>45</sup>.

We found that cortical recurrent circuits provide the majority of synaptic excitation to cortical neurons only after an initial transient excitation from thalamus, which dominates the first 43 ms of the response, at least under anesthesia. If this also holds true in the awake state, this initial transient excitation from thalamus might be particularly relevant to natural viewing conditions producing rapid transitions of visual input<sup>47,48</sup>.

What mechanism(s) ensure a rapid decay of sensory-evoked activity in intracortical-recurrent circuits after silencing thalamus? Notably, the CDF was invariant to changes in stimulus strength (for example, stimulus contrast or size), consistent with the CDF's ability to predict the amplitude of the cortical response across stimulus contrasts and temporal frequencies. Existing models of cortex achieve dynamics invariant to stimulus strength by assuming that local inhibition counteracts recurrent excitation to preserve the speed of the response<sup>5,7</sup>. Here we empirically test this assumption and find evidence that an appropriate ratio of excitation and inhibition contributes to the invariance of the CDF. However, additional mechanisms, including the biophysical membrane properties of cortical neurons, the dynamics of cortico-cortical synapses and these synapses' specific connectivity patterns may also critically contribute to the fast CDF. Future experiments that modulate specific types of interneurons in cortex will be required to understand exactly how inhibition contributes to the CDF.

Our approach also provides a mechanism for how anesthesia disrupts the timing of the cortical response, despite the CDF's invariance to anesthesia. We find that marked depression of thalamocortical

synapses under anesthesia disrupts the transmission of high-frequency sensory activity to cortex. In the awake state, however, thalamocortical synapses do not depress over a train of repeated visual stimuli and thus transmit both low and high frequencies of sensory activity equally well. Our data suggest a possible explanation for this difference between the anesthetized and awake states. We observed increased spontaneous activity in thalamus in the awake state, which is likely to 'pre-depress' the thalamocortical synapses before visual stimulation so no further depression is observed during visual stimulation<sup>19</sup>. If this brain state-specific switch in thalamocortical communication generalizes to other cortical areas, it might underlie the distinction between the awake and non-conscious states. Consistent with the effects of anesthesia, nonalert awake states show higher thalamocortical synaptic depression than alert awake states<sup>20,36</sup>, suggesting that reduced thalamocortical transmission of high temporal frequencies is not just specific to anesthesia. It is likely to represent a normal mode of brain function correlated with reduced alertness or inattention.

Moreover, the increase in spontaneous thalamic activity as animals wake up from anesthesia is likely to also explain our finding that the majority of spontaneous cortical activity in the awake state depends on thalamic input, from either dLGN or higher order thalamic nuclei, such as the lateral posterior, which was also silenced by our approach.

Thus, pre-depression of thalamocortical synapses in the awake state, along with an appropriate balance of excitation and inhibition in recurrent circuits, seems to enable the sensitivity of visual cortex to stimulus frequencies up to 30 Hz, consistent with the perceptual cutoff in mice<sup>49</sup> and humans. Many neurologic disorders with strong perceptual alterations, such as schizophrenia and autism, have been considered to be, in part, disruptions of the balance between excitation and inhibition in cortex<sup>50</sup>. The resulting change in the intrinsic dynamics of cortical recurrent circuits may account, in part, for the observed perceptual alterations. Understanding the mechanistic source of such alterations in brain dynamics should allow us to better target interventions. More generally, a mechanistic understanding of cortical dynamics will be vital to understanding various processes underlying cognition.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank J. Evora for help with genotyping and mouse husbandry, as well as J. Isaacson, B. Bloodgood, C. Reinhold, S. Larson and members of the Scanziani and Isaacson labs for discussions about the project and comments on the manuscript. This project was supported by the National Science Foundation Graduate Research Program Fellowship, the Gatsby charitable foundation and the Howard Hughes Medical Institute.

## AUTHOR CONTRIBUTIONS

K.R. and M.S. designed the study. K.R. conducted all experiments and analyses, except the whole-cell recordings. A.D.L. performed and analyzed the whole-cell recordings. K.R. and M.S. wrote the paper.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).
- Hubel, D.H. & Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Haider, B. *et al.* Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron* **65**, 107–121 (2010).
- Cossell, L. *et al.* Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* (2015).
- Douglas, R.J., Koch, C., Mahowald, M., Martin, K.A. & Suarez, H.H. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).
- Druckmann, S. & Chklovskii, D.B. Neuronal circuits underlying persistent representations despite time varying activity. *Curr. Biol.* **22**, 2095–2103 (2012).
- van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
- Hawken, M.J., Shapley, R.M. & Grosof, D.H. Temporal-frequency selectivity in monkey visual cortex. *Vis. Neurosci.* **13**, 477–492 (1996).
- Movshon, J.A., Thompson, I.D. & Tolhurst, D.J. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J. Physiol. (Lond.)* **283**, 101–120 (1978).
- DeAngelis, G.C., Ohzawa, I. & Freeman, R.D. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J. Neurophysiol.* **69**, 1091–1117 (1993).
- Holub, R.A. & Morton-Gibson, M. Response of visual cortical neurons of the cat to moving sinusoidal gratings: response-contrast functions and spatiotemporal interactions. *J. Neurophysiol.* **46**, 1244–1259 (1981).
- Saul, A.B. & Humphrey, A.L. Temporal-frequency tuning of direction selectivity in cat visual cortex. *Vis. Neurosci.* **8**, 365–372 (1992).
- O'Keefe, L.P., Levitt, J.B., Kiper, D.C., Shapley, R.M. & Movshon, J.A. Functional organization of owl monkey lateral geniculate nucleus and visual cortex. *J. Neurophysiol.* **80**, 594–609 (1998).
- Foster, K.H., Gaska, J.P., Nagler, M. & Pollen, D.A. Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J. Physiol. (Lond.)* **365**, 331–363 (1985).
- MacLean, J.N., Watson, B.O., Aaron, G.B. & Yuste, R. Internal dynamics determine the cortical response to thalamic stimulation. *Neuron* **48**, 811–823 (2005).
- Miller, J.E., Ayzenshtat, I., Carrillo-Reid, L. & Yuste, R. Visual stimuli recruit intrinsically generated cortical ensembles. *Proc. Natl. Acad. Sci. USA* **111**, E4053–E4061 (2014).
- Kloc, M. & Maffei, A. Target-specific properties of thalamocortical synapses onto layer 4 of mouse primary visual cortex. *J. Neurosci.* **34**, 15455–15465 (2014).
- Castro-Alamancos, M.A. & Oldford, E. Cortical sensory suppression during arousal is due to the activity-dependent depression of thalamocortical synapses. *J. Physiol. (Lond.)* **541**, 319–331 (2002).
- Borst, J.G. The low synaptic release probability *in vivo*. *Trends Neurosci.* **33**, 259–266 (2010).
- Swadlow, H.A., Gusev, A.G. & Bezdudnaya, T. Activation of a cortical column by a thalamocortical impulse. *J. Neurosci.* **22**, 7766–7773 (2002).
- Lien, A.D. & Scanziani, M. Tuned thalamic excitation is amplified by visual cortical circuits. *Nat. Neurosci.* **16**, 1315–1323 (2013).
- da Costa, N.M. & Martin, K.A. The proportion of synapses formed by the axons of the lateral geniculate nucleus in layer 4 of area 17 of the cat. *J. Comp. Neurol.* **516**, 264–276 (2009).
- Li, Y.T., Ibrahim, L.A., Liu, B.H., Zhang, L.I. & Tao, H.W. Linear transformation of thalamocortical input by intracortical excitation. *Nat. Neurosci.* **16**, 1324–1330 (2013).
- Ferster, D., Chung, S. & Wheat, H. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **380**, 249–252 (1996).
- Poulet, J.F., Fernandez, L.M., Crochet, S. & Petersen, C.C. Thalamic control of cortical states. *Nat. Neurosci.* **15**, 370–372 (2012).
- Halassa, M.M. *et al.* Selective optical drive of thalamic reticular nucleus generates thalamic bursts and cortical spindles. *Nat. Neurosci.* **14**, 1118–1120 (2011).
- Constantinople, C.M. & Bruno, R.M. Effects and mechanisms of wakefulness on local cortical networks. *Neuron* **69**, 1061–1068 (2011).
- Timofeev, I., Grenier, F., Bazhenov, M., Sejnowski, T.J. & Steriade, M. Origin of slow cortical oscillations in deafferented cortical slabs. *Cereb. Cortex* **10**, 1185–1199 (2000).
- Nauhaus, I., Busse, L., Carandini, M. & Ringach, D.L. Stimulus contrast modulates functional connectivity in visual cortex. *Nat. Neurosci.* **12**, 70–76 (2009).
- Steriade, M., Timofeev, I. & Grenier, F. Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* **85**, 1969–1985 (2001).
- Rager, G. & Singer, W. The response of cat visual cortex to flicker stimuli of variable frequency. *Eur. J. Neurosci.* **10**, 1856–1877 (1998).
- Murray, J.D. *et al.* A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661–1663 (2014).
- Niell, C.M. & Stryker, M.P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
- Niell, C.M. & Stryker, M.P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
- Boudreau, C.E. & Ferster, D. Short-term depression in thalamocortical synapses of cat primary visual cortex. *J. Neurosci.* **25**, 7179–7190 (2005).



36. Swadlow, H.A., Bezdudnaya, T. & Gusev, A.G. Spike timing and synaptic dynamics at the awake thalamocortical synapse. *Prog. Brain Res.* **149**, 91–105 (2005).
37. Stark, E. *et al.* Inhibition-induced theta resonance in cortical circuits. *Neuron* **80**, 1263–1276 (2013).
38. Han, X. *et al.* A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. *Front. Syst. Neurosci.* **5**, 18 (2011).
39. Atallah, B.V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**, 159–170 (2012).
40. Fuster, J.M. Memory networks in the prefrontal cortex. *Prog. Brain Res.* **122**, 309–316 (2000).
41. Gabernet, L., Jadhav, S.P., Feldman, D.E., Carandini, M. & Scanziani, M. Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron* **48**, 315–327 (2005).
42. Fuster, J.M. & Alexander, G.E. Neuron activity related to short-term memory. *Science* **173**, 652–654 (1971).
43. Camperi, M. & Wang, X.J. A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J. Comput. Neurosci.* **5**, 383–405 (1998).
44. Gold, J.I. & Shadlen, M.N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
45. Sherman, S.M. Thalamocortical interactions. *Curr. Opin. Neurobiol.* **22**, 575–579 (2012).
46. Bignall, K.E. & Rutledge, L.T. Origin of a photically evoked afterdischarge in cat visual cortex. *J. Neurophysiol.* **27**, 1048–1062 (1964).
47. David, S.V., Vinje, W.E. & Gallant, J.L. Natural stimulus statistics alter the receptive field structure of v1 neurons. *J. Neurosci.* **24**, 6991–7006 (2004).
48. Muller, J.R., Metha, A.B., Krauskopf, J. & Lennie, P. Information conveyed by onset transients in responses of striate cortical neurons. *J. Neurosci.* **21**, 6978–6990 (2001).
49. Nathan, J. *et al.* Scotopic and photopic visual thresholds and spatial and temporal discrimination evaluated by behavior of mice in a water maze. *Photochem. Photobiol.* **82**, 1489–1494 (2006).
50. Murray, J.D. *et al.* Linking microcircuit dysfunction to cognitive impairment: effects of disinhibition associated with schizophrenia in a cortical working memory model. *Cereb. Cortex* **24**, 859–872 (2014).

## ONLINE METHODS

**Animal protocols.** All procedures were conducted in accordance with the National Institutes of Health guidelines and with the approval of the Committee on Animal Care at UCSD (protocol S02160M). Animals were housed on a reverse light cycle in cages of four mice or less. At the time of electrophysiology, all animals were older than 3.5 weeks. Both male and female animals were used in an approximately equal ratio.

**Mouse lines.** Gad2-Cre (Jackson Labs stock number: 010802)  $\times$  C57Bl6 (ref. 51), vGat-ChR2 (Jackson Labs stock number: 014548)<sup>52</sup>, Scnn1a-Tg3-Cre (Jackson Labs stock number: 009613)<sup>53</sup> and PV-Cre (Jackson Labs stock number: 008069)<sup>54</sup>.

**Viruses.** AAV2/1.CAGGS.flex.ChR2.tdTomato.SV40 (Addgene 18917)<sup>55</sup> from the University of Pennsylvania viral vector core. AAV2/9.CAG.flex.Arch.GFP<sup>38</sup> from the University of North Carolina viral vector core.

**Animal surgery: viral injections.** *Stereotactic viral injections of TRN.* We stereotactically injected AAV2/1.CAGGS.flex.ChR2.tdTomato.SV40 into the thalamic reticular nucleus (TRN) of adult Gad2-Cre  $\times$  C57Bl6 transgenic mice. Mice were anesthetized with 2% isoflurane and placed into a Kopf stereotax. Core body temperature was maintained at just above 35 °C for the duration of the surgery with an FHC rectal probe and a heating pad. Lubricating ointment (Artificial Tears) was applied to the eyes, the head was shaved, and the skin was sterilized with alcohol and povidone-iodine before exposing and stereotactically flattening the skull. We made a small craniotomy (approx. 50  $\mu$ m in diameter above the somatosensory cortex, 1,540  $\mu$ m posterior and 2,235  $\mu$ m lateral to the bregma), inserted a thin pipette containing virus to 400  $\mu$ m beneath the final target injection site (final target site: 1,540  $\mu$ m posterior, 2,235  $\mu$ m lateral, 3,158  $\mu$ m ventral to the bregma), returned the pipette tip to the final target injection site, and pressure-injected 200 nL of the virus (titer:  $6.86 \times 10^{12}$  genome copies/mL) into the TRN, at an injection speed of 30 nL/min. We removed the pipette only 10–15 min after completion of the injection to reduce the spread of virus along the pipette track. We administered a single dose of 10% buprenorphine as a postoperative analgesic, sutured the skin, applied povidone-iodine, and waited 2 weeks before recording.

*Viral injections into V1 of Gad2-Cre mice.* AAV2/9.CAG.flex.Arch.GFP or AAV2/1.CAGGS.flex.ChR2.tdTomato.SV40 was injected into V1 of adult Gad2-Cre  $\times$  C57Bl6 transgenic mice. The virus injection surgery was the same as above. However, for the AAV2/9.CAG.flex.Arch.GFP injections, we made three small craniotomies (<50  $\mu$ m each in diameter) in a triangular pattern tiling the extent of mouse V1 and performed three pressure injections of 125 nL each (titer:  $6 \times 10^{12}$  genome copies/mL), at a depth of between 400 and 600  $\mu$ m, at a speed of 30 nL/min, and waiting 10–15 min between injection sites. We recorded 4–6 weeks after injection. For AAV2/1.CAGGS.flex.ChR2.tdTomato.SV40 injections, we made a single craniotomy over the center of V1 (<50  $\mu$ m in diameter) and injected 200 nL (titer:  $6.86 \times 10^{12}$  genome copies/ml) of the virus at a depth of 400–600  $\mu$ m. We recorded 2 weeks after injection.

*Viral injections into V1 of PV-Cre mice.* PV-Cre animals were injected at postnatal day 0–1 with AAV2/1.CAGGS.flex.ChR2.tdTomato.SV40 (see ref. 21 for pup injection methods).

**Animal surgery: terminal experiments under anesthesia.** During our surgical preparation of a mouse for electrophysiological recording, we maintained an anesthetic depth such that the mouse did not respond to a toe pinch, using either 2% gas isoflurane plus 2 mg/kg chlorprothixene (intraperitoneal injection), 2% gas isoflurane alone, or 0.5–1% isoflurane plus 1.5 g/kg urethane diluted to 0.1 g/ml in phosphate-buffered saline (PBS; i.p. injection). We covered the animal's eyes in transparent seed oil. After shaving the head and sterilizing the skin, we removed the skin and fascia, scored the skull with a bone scraper, and secured a metal head-frame to the skull using dental cement (Ortho-Jet) mixed with black paint, which prevents light during optogenetic stimulation from penetrating the dental cement. We allowed the cement to set for at least 30 min before drilling the skull.

*Craniotomies and thinned skulls in anesthetized preparations.* For photo-activation of TRN, we made a large circular craniotomy (diameter approx. 1.25 mm, 1,540  $\mu$ m posterior and 2,235  $\mu$ m lateral to the bregma) above the TRN.

We then used a cautery to make a hole in the brain tissue above the TRN and inserted an optical fiber (1 mm diameter) to an approximate depth of 1.25 mm, through the somatosensory cortex and part of the hippocampus at 1.54 mm posterior, 2.235 mm lateral and 1.5 mm ventral to the bregma. We verified that this acute insertion of the fiber optic did not affect our results by (i) confirming all results in a subset of anesthetized mice in which we used a smaller, 250- $\mu$ m-diameter fiber optic inserted more than a week earlier (see “Animal surgery: awake recordings”) and (ii) verifying that the CDF and extent of shut-off of V1 activity did not change over the course of long anesthetized recordings (>1 h).

For V1 whole-cell recordings, we made a circular craniotomy of 1–2 mm diameter over V1 and removed the dura using the bent tip of a 27- or 30-gauge needle. The cortical surface was covered in 1.5–2% agarose dissolved in artificial cerebrospinal fluid (ACSF).

For V1 extracellular recordings, we made a small craniotomy (diameter approx. 50  $\mu$ m) above V1.

For thalamic extracellular recordings, we made a craniotomy above dLGN and the lateral posterior nucleus (diameter: 200–300  $\mu$ m; approx. 1.8 mm posterior and 2 mm lateral to the bregma).

For photo-activation of ArchT in V1, we thinned the skull in the approximately 1.5  $\times$  1.5 mm square overlying visual cortex. We covered the brain in artificial cerebrospinal fluid (ACSF: 142 mM NaCl, 5 mM KCl, 10 mM D-glucose, 10 mM HEPES sodium salt, 3.1 mM CaCl<sub>2</sub>, 1.3 mM MgCl<sub>2</sub>, pH 7.4) and inserted the recording electrode. During the recording, we reduced the level of isoflurane to 0.9–1% (or less when combined with urethane).

**Animal surgery: awake recordings.** We fitted animals with a head frame for awake recordings more than a week before the day of the recording.

*Head-frame surgery.* We anesthetized animals with 2% isoflurane. After we shaved and sterilized the head, we cut off the skin and scored the skull using a bone scraper and surgical blade, which removes the soft upper layer of bone. We glued the edge of the skin to the skull using VetBond. We then inserted two bone screws bilaterally at 0.75 mm anterior and 2 mm lateral to the bregma. We used black dental cement to attach a metal head frame to the skull.

*Cannulation.* For optogenetic silencing of the thalamus, during the head-frame surgery, we also inserted a metal cannula (a guide for the optical fiber) with an outer diameter of 460  $\mu$ m through the skull and brain to a depth of 2,375  $\mu$ m beneath the bregma, just above the TRN at coordinates 1.54 mm posterior and 2.235 mm lateral to the bregma. The bottom of this cannula was angled 45 degrees and beveled to provide a clean, sharp edge for penetrating the brain tissue. Also, the bottom of the cannula was sealed with a clear window of Kwik-Sil that permanently separated the inside of the cannula from the brain. At the time of the recording, we inserted an optical fiber into this cannula and optically stimulated through this clear window, without re-damaging the brain. We cemented this windowed cannula to the skull and capped it with Kwik-Cast.

*Recording well.* During the head-frame surgery, we used the black dental cement to build up a recording well surrounding visual cortex. Finally, we applied a thin layer of clear cement to the remaining exposed skull to prevent infection. We sterilized the edges of the skin with povidone-iodine again before allowing the animals to wake up and administered a single dose of 10% buprenorphine as a postoperative analgesic. We checked on the mice daily after the head-frame surgery.

*Day of recording.* On the day of recording, we anesthetized the mice using 1.8–2% isoflurane, drilled off the clear dental cement covering the recording site, and made a small craniotomy (diameter approx. 50  $\mu$ m) over either V1 or visual thalamus. We covered the animal's eyes in Artificial Tears lubricant. We fixed the head-frame to a post, positioned the body of the mouse on the circular treadmill, and inserted the recording electrode into the brain before allowing the animal to wake up from anesthesia. Immediately after waking up, the mouse groomed to remove the lubricant from its eyes. When we recorded in the thalamus both before and after the animals woke up from anesthesia, we applied only a thin layer of the relatively transparent eye lubricant to be able to provide visual stimulation through this protective layer.

**Whole-cell recordings in visual cortex under anesthesia.** Glass pipettes with tip resistance of 3–5 M $\Omega$  were filled with either cesium-based (in mM: 135 cesium methanesulfonate, 8 NaCl, 10 HEPES, 0.3 Na-GTP, 4 Mg-ATP, 0.3 EGTA, 2 QX-314 chloride; pH 7.4 with CsOH, 285–290 mOsm) or potassium-based (in mM:

135 potassium gluconate, 8 NaCl, 10 HEPES, 0.3 Na-GTP, 4 Mg-ATP, 0.3 EGTA; pH 7.4 with KOH, 285–290 mOsm) internal solution and guided to the pial surface, which was visualized with a microscope using oblique infrared LED illumination. Recordings were made using the blind whole-cell patch clamp technique<sup>56</sup>. The depth of the recording was determined using the distance the pipette traveled from the pial surface as indicated by the micromanipulator reading. Neurons recorded at 300–550  $\mu\text{m}$  depth were considered to be in L4. Neurons were recorded in the voltage-clamp configuration and held close to the reversal potential of inhibition at  $-70$  mV to isolate excitatory postsynaptic currents.

**Post-mortem histology.** Animals were not perfused. We fixed the brains in 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS). After tissue fixation, we rinsed each brain in PBS overnight, placed the brain into 30% refrigerated sucrose solution for 2 d, and then cryosectioned the tissue using a freezing microtome. We mounted the 50- $\mu\text{m}$ -thick sections on slides with mounting medium including DAPI. For V2 track reconstructions, we stained the slices with blue NeuroTrace (fluorescent Nissl) before mounting. To stain with NeuroTrace, we rinsed each slice in PBS plus 0.1% Triton X-100 for 20 min, washed twice for 5 min in PBS, incubated the slices in 1:200 of NeuroTrace:PBS for 30 min at room temperature, and washed the slices overnight at 4 °C. When using the NeuroTrace stain, we did not include DAPI in the mounting medium. Viral expression (ChR2-TdTomato or ArchT-GFP) and DiI tracks were imaged on a fluorescent microscope.

**Excluding animals with ChR2 expression in V1.** All mice with stereotactic injections of AAV2/1.CAGGS.flex.ChR2.tdTomato.SV40 into the TRN were tested post-mortem for off-target expression of ChR2 in visual cortex. We excluded all the data from animals showing ChR2 expression of >25% fractional coverage (fraction of pixels presenting any detectable ChR2-TdTomato reporter fluorescence) in any slice of visual cortex. We verified that <25% fractional coverage of V1 by ChR2 had no effect on V1 activity during TRN photoactivation (**Supplementary Fig. 3**). Such off-target expression of ChR2 was observed in approximately 5–10% of the animals. In addition, when we pointed an LED-coupled fiber directly at V1 rather than at the TRN, we found that direct V1 illumination did not affect V1 activity.

**Excluding ChR2 expression in dLGN interneurons.** In about 20% of the mice, we observed some ChR2-TdTomato reporter fluorescence in local GABAergic interneurons of the dLGN. Excluding these mice had no effects on the results (that is, there was no change in the strength of V1 shut-off or its time course, the CDF). Furthermore, ChR2 expression in local dLGN interneurons, if anything, should lead to stronger and faster silencing of the dLGN, not slower and weaker silencing of dLGN. Therefore, we included all mice in the final figures.

**Labeling of electrophysiology recording track.** We verified recording sites in the thalamus and V2 by post-mortem inspection of the electrophysiology recording tracks. To label a track, at the time of the experiment but after conclusion of the electrophysiological recordings, we retracted the recording electrode along the axis of entry, applied a drop of 1 mM DiI to the probe, quickly wicked away the artificial cerebrospinal fluid (ACSF) on the surface of the brain, and then reinserted the recording electrode to the site of data collection. We allowed the DiI-coated electrode to remain in the brain for at least 10 min before removing the electrode, sacrificing the animal, and fixing the brain.

**Verification of V2 recording sites.** After slicing the brain (50- $\mu\text{m}$  sections), we identified the section containing the fluorescent DiI recording track in lateral secondary visual cortex (V2L). We identified the boundary between V1 and V2L, first, by imaging the section in bright-field before mounting the tissue, which reveals the V1–V2L border as a difference in darkness between the middle layers of V1 and V2L (likely due to differences between V1 and V2L in the density of myelinated thalamic afferents terminating in these layers); second, by Nissl staining and mounting to reveal the dense cytoarchitectonic band of layer 4 in V1; and, finally, by registering the images from these two methods to each other and to the Paxinos mouse atlas<sup>57</sup>. We excluded data obtained from animals in which the boundary between V1 and V2L<sup>58</sup> appeared inconsistent or ambiguous.

**Habituating awake mice to the recording set-up.** We habituated mice to the head-fixed electrophysiological set-up before recording in the awake state. Four or more days after the head-frame surgery, we began a habituation protocol that exposed a mouse to the recording set-up for 30–60 min each day for at least 3 d. During each habituation session, the mouse was head-fixed (by fixing the

head-frame to a metal post) and allowed to run freely during visual and mock optogenetic stimulation (LED pulses near the head but pointed away from the brain). We recorded using the same set-up.

**Extracellular electrophysiology.** Extracellular recordings from cortex and thalamus were performed using a NeuroNexus silicon probe (A series) with 16 linear recording sites. In cortex, the 50- $\mu\text{m}$  spacing between these sites provided an array that spanned 800  $\mu\text{m}$  in total, enabling simultaneous recording from all cortical layers. In three of our thalamic recording experiments, we instead used an array with 25- $\mu\text{m}$  spacing to improve unit isolation. The recording electrode was connected to an AM Systems preamplifying head-stage (20 $\times$ ) through a Plexon adaptor, and then the voltage signals were further amplified 500 $\times$  and filtered between 0.1 Hz and 10 kHz by the AM Systems model 3600 extracellular amplifier. Data were digitized with a National Instruments Data Acquisition card and acquired with custom Matlab data acquisition software written by S. Olsen (University of California San Diego).

**V1 recordings.** We targeted the center of the monocular zone when recording from V1.

**Recordings in thalamus.** We targeted the dLGN by advancing the electrode straight down at coordinates 1.8 mm posterior and 2 mm lateral to the bregma. When post-mortem reconstruction of this thalamic recording track revealed penetration of both dorsal and ventral LGN (in about 10% of the animals), we included only the topmost visually responsive channels on the recording array to exclude from the analysis visually responsive units in ventral LGN. We targeted the lateral posterior nucleus at a site 300–400  $\mu\text{m}$  medial to this dLGN site.

**V2 recordings.** We stereotactically targeted V2 at >3.25 mm lateral of bregma. We only included V2L data sets showing an obvious visually evoked response and for which we were able to confirm by post-mortem histology a recording site more than 50  $\mu\text{m}$  lateral to a clear V1–V2L border (see “Verification of V2 recording sites”).

**Visual stimulation.** Visual stimuli in figures:

**Figure 1b** (center): 1.7-s-long static oriented gratings and luminance steps.

**Figure 1b** (right): 1.7-s-long static oriented gratings and luminance steps, and 250-ms long static oriented gratings.

**Figure 1c** (center): 3-s-long static oriented gratings.

**Figure 1c** (right, top): 3-s-long static oriented gratings.

**Figure 1c** (right, bottom, scatter plot): cyan, 3-s-long static oriented grating; red, 3-s-long luminance step (black to white); dark blue, 10- to 100-ms-long flash of light; purple, reversal of checkerboard pattern every 3 s.

**Figure 2c:** 3-s-long moving oriented gratings.

**Figure 2f,g:** 3-s-long static oriented gratings and luminance steps.

**Figure 2h:** a combination of all stimulus types.

**Figure 2i:** 3-s-long moving oriented gratings versus spontaneous activity.

**Figure 3b** (top): 3-s-long moving oriented gratings.

**Figure 3b** (bottom): 3-s-long moving oriented gratings or 3-s-long static gratings.

**Figure 3d:** 3-s-long moving oriented gratings.

**Figure 4b,c:** full-field flicker lasting for 2 s.

**Figure 4e:** repetitive train of full-field flashes (10 ms duration for each flash).

**Figure 4f,g:** full-field flicker lasting 2 s.

**Figure 5:** 3-s-long moving oriented gratings.

**Supplementary Figure 1b** (single-cell examples): 1.7-s- or 250-ms-long static oriented gratings.

**Supplementary Figure 1b** (average, left): 1.7-s-long static oriented gratings and 1.7-s-long luminance steps.

**Supplementary Figure 1b** (average, right): 250-ms-long static oriented gratings.

**Supplementary Figure 1d** (top): 3-s-long static oriented gratings.

**Supplementary Figure 1d** (2nd row): 3-s-long moving oriented gratings.

**Supplementary Figure 1d** (3rd row): 3-s-long static oriented gratings and 3-s-long luminance steps.

**Supplementary Figure 1d** (bottom, left): 3-s-long luminance step.

**Supplementary Figure 1d** (bottom, right): 10-ms-long luminance step.

**Supplementary Figure 2a:** 3-s-long moving oriented gratings.

**Supplementary Figure 2b:** no visual stimulus.

**Supplementary Figure 2c:** 3-s-long moving oriented gratings.

**Supplementary Figure 2d:** 3-s-long moving oriented gratings.

**Supplementary Figure 2e:** 3-s-long moving oriented gratings.

**Supplementary Figure 2f:** 3-s-long moving oriented gratings.

**Supplementary Figure 2g:** 3-s-long moving oriented gratings, 3-s-long static oriented gratings, and 3-s-long luminance steps.

**Supplementary Figure 2h:** 3-s-long moving oriented gratings or no stimulus.

**Supplementary Figure 2i:** offset of 3-s-long moving oriented grating.

**Supplementary Figure 4b (left, right):** 3-s-long moving oriented gratings.

**Supplementary Figure 4b (center):** no visual stimulus.

**Supplementary Figure 4c:** 3-s-long moving oriented gratings.

**Supplementary Figure 4d (left):** 3-s-long moving oriented gratings.

**Supplementary Figure 5 (a–c,e):** high-contrast (contrast = 1) full-field luminance flicker lasting 2 s.

**Supplementary Figure 5d:** low-contrast (contrast = 0.2) full-field luminance logarithmic frequency chirp; see **Supplementary Figure 5f** “Vis. stim.”

**Supplementary Figure 5f,g:** high-contrast (contrast = 1) full-field luminance logarithmic frequency chirp; see **Supplementary Figure 5f** “Vis. stim.”

**Supplementary Figure 8:** full-field high-contrast (contrast = 1) luminance flicker lasting 2 s.

**Supplementary Figure 9:** 3-s-long moving oriented gratings.

**Supplementary Figure 10:** 3-s-long moving oriented gratings.

**Visual stimulation.** We used an LCD computer monitor for visual stimulation (gamma-corrected, mean luminance 50 cd/m<sup>2</sup>, refresh rate 75 Hz, 16 × 24 inches) at a distance of 25 cm from the eye contralateral to the V1 recording site. We began the recording with the computer monitor at about a 60-degree angle from the anterior-posterior axis of the head. We then repositioned the monitor as necessary to place the apparent spatial receptive field of active units on the center of the screen. The Psychophysics Toolbox in Matlab supported our presentation of various stimuli. Unless otherwise noted, stimulus contrast was 1, and the mean luminance of each full-contrast stimulus matched that of the interleaved blank gray screen.

**Visual stimuli during V1 extracellular recordings.** Visual stimuli consisted of static gratings (duration: 3 s) of eight interleaved and randomized orientations. Static gratings (spatial frequency 0.04 cyc/deg) appeared following a gray screen (6 s inter-trial interval of gray screen) of mean luminance matching the mean luminance of the static grating. We also used as visual stimuli, as specified in the text, full-field luminance changes (transition from black to white screen: duration of white screen, 3 s; duration of inter-trial black screen, 5 s), reversing checkerboards (one-time contrast reversal of checkerboard pattern, each square side length 3 degrees, inter-reversal interval 5.5 s), moving gratings (transition from a gray screen to a pattern of black and white oriented bars, eight interleaved and randomized orientations, moving across the screen at a constant speed: spatial frequency, 0.04 cyc/deg; temporal frequency, 2 to 3 cyc/s) and full-field flashes of white or blue light lasting between 10 and 100 ms.

**Visual stimuli during V1 whole-cell recordings.** Visual stimuli consisted of static gratings (as above; duration either 250 ms or 1.7 s) at the cell’s preferred orientation (spatial frequency 0.04 cyc/deg) or a full-field luminance step lasting 1.7 s ( $n = 4$  cells). A given stimulus was presented at least four times and stimulus order was randomized. LED and non-LED blocks were interleaved.

**Visual stimulation to measure the frequency response.** We used the light of a blue light-emitting diode (LED) collimated through an objective lens (5×, 0.15 NA). The back aperture of this objective was pointed at the eye of the mouse to obtain a spatially unstructured, illuminated circular field at the eye (diameter ~3 cm at eye). The LED light source allowed us to produce precisely time-varying visual stimuli (frequencies between 1 and 60 Hz, or logarithmically modulated frequency sweeps, called chirps), uniformly illuminating much of the visual field of the eye contralateral to the recording site. We modulated the LED with a computer-generated voltage signal. The power output of the LED varied sinusoidally between 0 and 25 mW (3.5 mW/cm<sup>2</sup>) as the input signal varied in time.

**Visual stimulation to measure the depression of thalamo-cortical synapses.** We used the same LED visual stimulation set-up described above (see “Visual stimulation to measure the frequency response”) to present pulse trains of brief (10 ms duration, 3.5 mW/cm<sup>2</sup>) flashes to the contralateral eye. The frequency of the pulse train varied between 1 and 60 Hz.

**Photo-activation of TRN.** LED: LEDP-B\_PF960-0.50-1m-FC\_R2 from Doric Lenses, 470 nm. Laser: Omicron, 80 mW maximum output, 470 nm.

**Anesthetized animals.** The optical fiber coupled to a blue LED (1 mm diameter; maximum power output 80 mW) was inserted and targeted as described above

(“Animal surgery”). We were able to use <10-mW step functions to drive maximal silencing of the dLGN in most mice, but in some animals we increased the power output of the LED to achieve maximal thalamic silencing. In some experiments, we used various LED intensities below 10 mW to partially silence thalamus. In anesthetized animals, in the presence of continuous TRN illumination, dLGN silencing lasted for more than 1 s.

The ability of LED illumination to drive sustained spiking of the TRN units wore off over the course of about 1–2 h of recording under anesthesia when we frequently photo-activated the TRN (LED on for 1 s every 7 s). (We recorded the TRN population activity as LED-driven axonal signals in the dLGN—see below, “Sorting single units”.) Less frequent TRN photo-activation eliminated this run-down effect (LED on for <0.5 s every 10 s). This run-down was independent of anesthetic depth and time from penetration of the electrode. Thus, it was a function of illumination frequency and not of brain state or tissue recovery. For strong, reliable and sustained TRN photo-activation leading to strong thalamic silencing, we increased the inter-trial interval to eliminate this run-down effect and chose to include only data from the first hour of each anesthetized recording session.

**Awake animals.** We inserted an optical fiber (diameter 200 μm) coupled to a blue laser (power output 10 mW for photo-activation) into a cannula chronically implanted above the TRN, as described above (“Animal surgery”). In contrast to what was observed in anesthetized animals (see above), continuous illumination of the TRN in awake animals provided sustained silencing of the dLGN for only about 250 ms. In fact, dLGN silencing in awake animals was invariably followed by a rebound of activity (recorded in both dLGN and V1) approximately 250 ms after laser onset, although the laser remained on. This rebound was the first cycle of a thalamo-cortical oscillation (4–8 Hz) lasting for a second or more. Thus, in awake mice, we only considered the thalamus to be silenced during the first 125 ms immediately following TRN photo-activation. Given the ~10-ms CDF, this 125-ms time window is more than long enough for a complete characterization of the time course of V1 shut-off following TRN photo-stimulation in awake mice.

**Photo-activation of cortical interneurons to silence V1 recurrent excitatory circuits.** A blue LED (455 nm) coupled to a 1-mm fiber was positioned several millimeters above the pial surface. Total power from the fiber was 25 mW. In LED blocks, the LED was turned on 645 ms before the onset of each block of visual stimulus and lasted the duration of the stimulus block.

**Photo-activation of ArchT in V1 inhibitory interneurons.** We used an amber LED (1 mm diameter, approximately 10–30 mW output; see explanation below) positioned above visual cortex to photo-activate ArchT expressed in cortical inhibitory interneurons (see section on viral injections, above). We titrated the intensity of the amber LED to maximally disinhibit the visually evoked response without triggering ictal events. We performed this experiment in anesthetized mice because even slight disinhibition of the visual cortex in awake animals triggered ictal-like activity. To determine the CDF, we presented a visual stimulus (full-field moving grating, orientation varied at random) and silenced relay thalamus by activating the TRN with a blue LED (see above). We used four interleaved LED conditions: (1) No LED, (2) Blue LED, (3) Amber LED and (4) Amber plus Blue LED. These LED conditions were presented in the order “1, 2, 3, 4, 3, 2, 1, 4” to verify that any observed change in the CDF was not affected by the LED condition on the previous trial. The effect of the amber LED on the CDF was consistent over the course of the V1 recording. We observed the greatest slowing of the CDF in mice showing widespread expression of ArchT across visual cortex. We quantified the spread of ArchT expression as the ‘fractional coverage’ of V1: that is, the fraction of V1 pixels presenting any detectable ArchT-GFP reporter fluorescence in post-mortem sections. We selected the five mice with the greatest ArchT fractional coverage of V1 for **Figure 5**. When we included all the experiments in which we illuminated ArchT in V1, irrespective of fractional coverage ( $n = 19$  mice), suppression of inhibitory interneurons still produced a significant slowing of the cortical CDF (**Supplementary Fig. 9**). We recorded in V1 for an hour before moving the electrode to the dLGN of the same mouse to verify no change in shut-off in the dLGN.

**Data analysis: whole-cell (Fig. 1b).** The responses to repeated presentations of the stimulus were averaged, and the average value in a baseline window spanning

the time window from the onset of the stimulus to 25 ms after stimulus onset was subtracted from control (non-LED) and LED conditions for each cell. All traces were shifted in time such that the onset of the visual response in the LED condition occurred at time 0. Onset was defined as the time point at which the response grew to beyond 3 s.d. of the baseline. Excitatory postsynaptic currents (EPSCs) during control (non-LED) and LED conditions were normalized across cells by normalizing each cell's LED and non-LED trace to the peak amplitude of the non-LED response (Fig. 1b, middle row). Peak amplitude was determined from traces that were down-sampled with 5-ms binning. The population averages of non-LED and LED responses were calculated by averaging responses across all cells.

**Quantifying the fraction of response mediated by recurrent cortical circuits.** To determine the fraction of the response dependent on cortical recurrent circuits (Fig. 1b, bottom row), responses were processed as above but were not normalized. The responses were binned into 5-ms windows, and the LED response was shifted so that the first bin (0–5 ms) was equal to that of the non-LED response. For each cell, we calculated the difference between the non-LED (black) and LED (blue) traces, and then divided this quantity by the non-LED (black) trace at each time point. For cells in which the longer, 1.7-s-duration stimulation was employed, this calculation was performed only up to the time point at which the non-LED response was greater than 10% of its peak amplitude.

**Data analysis: single-unit response durations (Fig. 1c).** To measure the response duration for each single unit individually, we plotted the trial-averaged peri-stimulus time histogram (PSTH) of the unit's response and then measured the time it took for this PSTH to re-enter the noise (that is, when the unit's firing rate after the stimulus returned to within 1 s.d. of the mean of the unit's activity in the absence of a visual stimulus). We report this time as the response duration.

**Data analysis: sorting single units.** We used UltraMegaSort from D.N. Hill, S.B. Mehta and D. Kleinfeld<sup>59</sup> to cluster spike waveforms into putative units and then manual sorting to verify the quality of the isolated units. Our units were well-isolated clusters with large spikes (Gaussian fit to the distribution of spike amplitudes suggested >85% of the spikes in every cluster were greater than 4 × the s.d. of the high-frequency noise), contained fewer refractory period violations than 1% of the total spikes, and did not appear similar in waveform shape or amplitude to any neighboring clusters, thus excluding units that could be confused.

**Sorting thalamic relay units.** We separated the signals obtained from our dLGN recordings into spikes originating from putative relay cells and spikes originating from putative TRN axons. Spikes from putative relay cells had larger, broader waveforms and exhibited significantly higher F1 modulation. Furthermore, these spikes were suppressed by TRN photo-activation, and they showed a clear postinhibitory rebound. Activity from the putative TRN axons, in contrast, represented a very thin, low-amplitude spike population (hash). The spike rate of this hash was increased by TRN photo-stimulation with very low latency. Thus, we separated the spikes of the putative relay cells from the LED-driven hash by isolating single units and then excluding any units with an average spike waveform width at half-maximum of less than 0.22 ms.

**Separating regular-spiking and fast-spiking units in the cortex.** We separated the regular-spiking (putative excitatory) and fast-spiking (putative inhibitory) units in the cortex by evaluating the average waveform of each unit by eye. The mean width at half-maximum of the spike waveform for regular-spiking units was 0.44 ms (s.d. 0.16 ms). The mean width at half-maximum of the spike waveform for fast-spiking units was 0.25 ms (s.d. 0.067 ms).

**Registration of recording depth by current source density and alignment relative to Scnn1a-Tg3 Cre mouse line.** To correct for potential variation in the position of the 16 linear electrode channels relative to cortical layers across experiments, we used the current source density (CSD) to identify the site of the L4 current sink at stimulus onset and used this site as a reference. We verified the stability of this CSD-defined L4 sink relative to the cortical anatomy using conditional expression of virally injected ChR2 in the Scnn1a-Tg3-Cre mouse line, which expresses Cre recombinase mainly in L4 excitatory neurons<sup>60</sup>. We first measured the CSD-defined L4 sink at stimulus onset and then photo-stimulated the ChR2-expressing L4 population on interleaved trials to determine the relationship between these two reference points (functionally defined L4 sink

and anatomically defined Cre<sup>+</sup> neuronal population). We found that the depth relationship was consistent across awake and anesthetized mice, thus validating the utility of the CSD-defined L4 sink as an alignment point. Finally, we assigned putative cortical layers (notably, these are only rough estimates of the cortical layer boundaries) to the aligned average by stretching a template of putative cortical layer widths from slice experiments (quantification by D. Bortone, University of California San Diego) across the depth of cortex, with L4 at the L4 CSD sink.

**Data analysis: V1 active states under anesthesia.** We identified V1 active states according to the local field potential (LFP) power ratio (PR), a method that quantifies the ratio of the power in the LFP between 5 and 30 Hz over power between 30 and 100 Hz. Increases in this LFP PR, as calculated from single-trial spectrograms (Gabor-Morlet wavelets, as in ref. 61 or Chronux<sup>62</sup>) are indicative of Up states (cortical active states under anesthesia). We found a threshold for the PR able to separate Up (high activity) and Down (low activity) states by plotting the bimodal distribution of cortical multi-unit (MU) firing rates across 500 ms time bins, using receiver operating characteristic (ROC) analysis to identify the MU threshold best separating the two modes in this histogram (Up and Down states), and then converting this MU threshold to a PR threshold using the least-squares linear fit to the relationship between MU firing rate and PR. Thus, consistent with the identification of Up states in previous work, we selected time periods of cortical activity characterized by a high LFP PR. We applied this LFP PR threshold to all the cortical LFP data to pick out the Up states beginning within 300 ms of silencing of the thalamus, aligned the Up states to their onsets and to the LED onset, and calculated the average firing rate over the course of the Up state (Fig. 2e).

**Data analysis: PSTH normalization. dLGN PSTH under anesthesia.** To show the effects of TRN photo-activation on thalamic relay cell activity (Fig. 2c), we isolated putative relay cells as described above ("Sorting thalamic relay units") and plotted, for each mouse, a PSTH of summed single-unit activity, analogous to multi-unit activity but excluding high-frequency noise and putative axonal hash (see above). We then normalized the PSTH for each mouse by the mean summed single-unit activity over the full duration of the moving grating (3 s) during control conditions (black). Error bars are s.d. of the normalized PSTH across mice. Summed single-unit activity in the thalamus was not baseline-subtracted.

**V1 PSTH under anesthesia.** For visually evoked activity in V1 (Fig. 2f), we show the baseline-subtracted multi-unit PSTH in response to brief visual stimuli, normalized by the mean multi-unit firing rate over the first 500 ms of the control evoked response. This PSTH includes responses to brief flashes, reversing checkerboards, full-field luminance transitions and static gratings. The results were the same for all these brief stimuli and also for moving gratings. We show the cortical decay function (CDF) during brief stimuli in Figure 2f. Again, there was no difference in the CDF between brief visual stimuli and moving gratings (Supplementary Table 1). The PSTH showing spontaneous activity in V1 during Up states (Fig. 2e) is multi-unit activity normalized by the mean firing rate in the control condition over the 500-ms window surrounding the peak of the Up state, similar to the PSTH showing visually evoked activity. In all cases, error bars are s.d. of the normalized PSTH across mice.

**Awake.** We also show the visually evoked PSTH in awake mice (Fig. 3b). In this case, we show the PSTH for visual stimuli including moving gratings and the CDF curves separately for brief visual stimuli (brief flashes, reversing checkerboards, full-field luminance transitions and static gratings) and moving gratings. As under anesthesia, there was no change in the CDF across any of the visual stimuli tested (Supplementary Table 1).

**Data analysis: accounting for shut-off delay in dLGN.** To calculate the exact time course of sensory activity decay intrinsic to the cortical recurrent circuits (in linear systems terms, the cortical impulse response function), we needed to remove (deconvolve) the time course of shut-off in the thalamus from the time course of shut-off in the cortex. However, deconvolution of real neural signals with noise introduces error. Therefore we decided to study the shape of the decay curve in V1 only after the thalamus was nearly off. The single exponential time constant fit to the shut-off of activity in dLGN was 3.7 ms. Thus, the dLGN activity has already decreased substantially (>60%) by about 3 ms after the LED onset. Furthermore, 3 ms is the average delay before activity begins to obviously decay in V1. Beyond 3 ms, a single exponential fit well approximated the

result of the deconvolution (Supplementary Fig. 2c). Furthermore, the V1 decay time course approximated an exponential function, and an exponential function has the same shape at all scales, meaning that theoretically we can measure any window of the exponential decay and get the same fit. Finally, considering this time window beyond 3 ms after LED onset is more than sufficient to allow us to study response frequencies between 1 and 60 Hz. For these reasons, in all cases, we fit the single exponential decay to cortical shut-off starting 3 ms after LED onset. This accounts for the time it takes the dLGN to shut off.

**Data analysis: single exponential fits to cortical shut-off.** We applied the following protocol to fit the CDF. (i) We used the baseline-subtracted PSTH of V1 shut-off between 3 and 50 ms after the onset of the LED (or longer; see time window in each figure). The 3-ms delay from LED onset was excluded from the shut-off curve to account for the time it takes the dLGN to shut off (see section above). (ii) Avoiding any assumptions about the expected fractional suppression of the evoked response, we used Matlab to fit a single exponential plus a constant offset to this PSTH. This method optimized the fit to all points in the trace, without forcing the initial or final values. (iii) If this method failed to return a fit, we forced the constant offset to equal the mean value of the last 5 ms of the shut-off PSTH. (iv) In the case of noisy data, this latter method could also fail to return a fit, in which case we forced both the initial value of the exponential fit to equal the mean of the PSTH over the 20 ms preceding the LED onset and the final value of the exponential fit to zero. (v) If all of these methods failed to fit the shut-off curve, we excluded the data. (vi) All fits were verified by eye.

**Data analysis: frequency response.** *Frequency response: heat map.* To compare the responses of dLGN and V1 to high-contrast, sinusoidally modulated, full-field luminance stimuli (called flicker stimuli) presented at various temporal frequencies (15 stimulus frequencies between 1 and 60 Hz), we used the frequency response. We computed the frequency response as follows.

We used Chronux<sup>62</sup> to compute the amplitude spectrum (or power spectrum) of the trial-averaged steady-state PSTH (between 0.4 and 2 s after stimulus onset) in response to each stimulus frequency. With this method, we picked out the frequency component of the response to a given stimulus as long as that component was phase-locked to the onset of the stimulus (that is, apparent in the trial-averaged PSTH). We created heat-map matrices of the power at each frequency to summarize the neural responses in dLGN and V1 as a function of stimulus frequency (Supplementary Fig. 8b). When combining single units, each single unit's heat map was first normalized to its peak along the diagonal, to equally weight all units in the combined heat map and thus avoid a larger contribution from cells with higher firing rates. We also excluded units that were not visually responsive. For dLGN, we show the median rather than the mean heat map and frequency response to reduce biases from very high firing rate outlier units observed in dLGN (about 3% of total units). For V1, we show the mean (in Fig. 4c,f,g) and the median (in Supplementary Fig. 5b). For V1 units, the mean and median are very similar. However, we also show the full distributions for dLGN and V1 amplitudes as percentiles in Supplementary Figure 5b to show that the full distributions are consistent with filtering by the CDF.

*Frequency response: two-dimensional plot.* To obtain a simple two-dimensional plot that captures how well the dLGN and V1 followed the stimulus input frequencies (Fig. 4c,f,g), we plotted the amplitude of the neural response at each frequency of the stimulus (Fig. 4c,f,g). To obtain a meaningful zero value for this two-dimensional plot, we subtracted the average amplitude of the non-specific response at all response frequencies above 70 Hz. This 'non-specific response' was equivalent to the amplitude at frequencies other than the frequency of the input, when we did not observe harmonics. In the power heat maps, for example, the nonspecific power corresponds to the dark blue background. This nonspecific response subtraction eliminates noise contributions to the final frequency response.

*Frequency response: prediction in the time domain.* We predicted the V1 PSTH (Fig. 4b, pink traces) by convolving the dLGN PSTH traces (Fig. 4b, top) with the regular-spiking empirical CDF of V1 measured during moving gratings (Fig. 4c, inset, and Supplementary Figs. 2g and 7, pink). The CDF did not change with stimulus type for all stimulus types tested, so we chose this very clean measurement of the regular-spiking units' CDF to attempt to predict the regular-spiking

units' frequency response in V1. For the example data in Figure 4b, to maintain the relative amplitudes of the predicted responses across temporal frequencies of the visual stimulus, we scaled all the pink traces by the same amount to best match, by eye, the actual V1 PSTH responses (black traces). We then aligned each pink trace to each black trace by eye. The pink prediction curves in Figure 4b for 4 and 6 Hz visual stimuli in the awake mouse have been flipped upside down with respect to the curve in dLGN to account for the fact that, in this example, units sampled in V1 in response to these stimulus temporal frequencies have a 90-degree phase offset with respect to the units sampled in dLGN. This does not affect the ability of the CDF to predict the amplitude of the V1 response (Fig. 4b). Sometimes our recordings sampled more On-responsive units; other times our recordings sampled more Off-responsive units. Therefore we focused on testing the ability of the CDF to predict V1's multi-unit response amplitude, not its phase.

*Prediction in the frequency domain.* We predicted V1's frequency response by multiplying the dLGN's frequency response (Fig. 4c, green) by the amplitude spectrum of the CDF (Supplementary Fig. 7b, pink). For the summary two-dimensional single-line graphs (Fig. 4c), we plotted the cortical prediction (pink) relative to the thalamic frequency response (green) such that an ideal frequency response curve representing 'perfect following' of the thalamic input (that is, filtered by an infinitesimally fast CDF) would lie atop the plotted thalamic response. Because the proportionality factor relating firing rates in the dLGN to firing rates in V1 is not known, the absolute amplitude of the actual V1 frequency response (black) relative to the predicted V1 frequency response (pink) is arbitrary. Thus, in Figure 4 we aligned the prediction for V1 (pink) to V1's actual frequency response (black) by scaling the heights of the curves to minimize the difference between the curves by eye, while also aligning these curves at the nonspecific, baseline-subtracted zero (see "Frequency response: 2D plot" above for explanation of zero in the frequency response). In Figure 4g, we use the same method to predict the frequency response curve that would result from replacing the real CDF with a time constant of 1 ms or 100 ms.

*Models of fundamental (F1) and first harmonic (F2) responses in V1 (Supplementary Fig. 8).* To test whether a simple transformation of spatial receptive field (SRF) structures between dLGN and V1 could account for the decrease in V1 response amplitudes to high temporal frequencies with respect to the response in dLGN, we considered an alternate simple model: responses at the fundamental (F1) in dLGN can become responses at the first harmonic (F2) in V1. We found that this model alone, in the absence of low-pass filtering by the cortical decay function (CDF), was not able to account for the observed F1 and F2 responses in V1 (Supplementary Fig. 8).

First, we measured the fundamental (response frequency matches stimulus frequency, called F1 modulation) and first harmonic (response frequency is twice the stimulus frequency, called F2 modulation) amplitudes in dLGN and V1 (Supplementary Fig. 8b,c). We then attempted to fit three models, inspired by previous work on spatial receptive fields in the visual system<sup>2</sup> (Supplementary Fig. 8d), to this data.

*Model 1.* Model 1 allows F1 activity in dLGN to drive F1 activity in V1. F2 activity is not considered. Model 1 is the simple model presented in Figure 4. For discussion of model 1 and filtering of the F1 response by the cortical decay function (CDF), see above.

*Model 2A.* Model 2A allows F1 activity in dLGN to drive either F1 or F2 activity in V1. F2 activity in dLGN can also drive F2 activity in V1. In this model, there is no filtering by the cortical decay function (CDF).

*Model 2B.* Model 2B allows the same F1-to-F2 transformations as model 2A but also includes filtering by the CDF.

*Equations for models of F1 and F2 responses in dLGN and V1.* Let  $d_{F1}(f)$  be the amplitude of dLGN's F1 response as a function of the visual stimulus temporal frequency  $f$ . Let  $d_{F2}(f)$  be the amplitude of dLGN's F2 response as a function of stimulus frequency. Similarly, let  $V_{F1}(f)$  and  $V_{F2}(f)$  be the amplitudes of V1's F1 and F2 responses, respectively, as functions of stimulus frequency. Let  $c_0$  be a constant scaling factor between dLGN spiking and V1 spiking that is independent of stimulus frequency. Let  $F_{NTC}(f_{resp})$  be the amplitude scaling produced by filtering thalamic activity at frequency  $f_{resp}$  with the CDF. Finally, let  $D(f)$  be the fraction of the F1 amplitude in the dLGN at stimulus frequency  $f$  that is transformed to an F2 response in V1, without a loss of amplitude during this transformation. This is the simplest model of an F1-to-F2 transformation between dLGN and V1.

Let  $D(f)$  vary as a function of stimulus frequency. Then we can write models 2A and 2B as two systems of equations:

Model 2A: no filtering by CDF:

$$V_{F1}(f) = c_0(1 - D(f))d_{F1}(f)$$

$$V_{F2}(f) = c_0(d_{F2}(f) + D(f)d_{F1}(f))$$

Model 2B: includes filtering by CDF:

$$V_{F1}(f) = c_0F_{NTC}(f)(1 - D(f))d_{F1}(f)$$

$$V_{F2}(f) = c_0F_{NTC}(2f)(d_{F2}(f) + D(f)d_{F1}(f))$$

To test models 2A and 2B, we solved each system of equations above for the fraction of the F1 response in dLGN that is frequency-doubled to produce F2 in V1 (a parameter we call  $D(f)$ ; see above). We took the average estimate of  $D(f)$  from each equation, forcing  $D(f) \geq 0$  (Supplementary Fig. 8e). We then used this estimate of  $D(f)$  to predict the F1 and F2 amplitudes for V1 and compared these predictions to the real F1 and F2 responses in V1 (Supplementary Fig. 8f). If a simple solution exists to model 2A or 2B, we should be able to find a value of  $D(f)$  that accurately predicts both V1's F1 response and V1's F2 response. We were only able to find such a value for  $D(f)$  in the context of model 2B, which includes filtering by the CDF, indicating the following: first, model 2A does not explain V1's response, and, second, V1's F1 and F2 responses are better explained by filtering with the CDF than by a simple, lossless (no loss of amplitude) F1-to-F2 transformation between dLGN and V1.

**Data analysis: deconvolution of dLGN PSTH from V1 PSTH.** In Supplementary Figure 5c, we approximate the temporal transformation of the sensory response between dLGN and V1 directly by computing the deconvolution of the dLGN response from the V1 response. We used Matlab to compute this deconvolution for each pair of responses in dLGN and V1 to the same visual stimulus (same temporal frequency of the flicker stimulus), across all possible pairs of recordings across all mice. We then computed the average and standard error of this deconvolution across all the pairwise dLGN-V1 comparisons. The result of the deconvolution is noisy because the deconvolution of real neural signals is not ideal (Supplementary Fig. 5c). Therefore, this deconvolution should be viewed as a rough approximation. To summarize the difference in this result between anesthesia and awake, we fit a single exponential decay function to each deconvolution (black fits in Supplementary Fig. 5c).

**Measuring thalamo-cortical synaptic depression *in vivo*.** We isolated the thalamic component of the field excitatory postsynaptic potential (fEPSP<sub>thal</sub>) recorded in layer 4 as described in detail previously<sup>21</sup>. Briefly, we optogenetically silenced visual cortex by photo-activating GABAergic cortical interneurons conditionally expressing Chr2 (in Gad2-Cre mice with viral injections—see above—or in vGat-Chr2 reporter mice) with a blue LED positioned above visual cortex while recording the fEPSP<sub>thal</sub> with a 16-channel linear extracellular array (NeuroNexus, A series). Visual stimuli consisted of brief pulse trains of blue light flashed at the contralateral eye ranging in frequency between 1 and 60 Hz (see “Visual stimulation to measure the depression of thalamo-cortical synapses”). To separate the depression occurring at the thalamo-cortical synapse<sup>63,64</sup> from any potential frequency-dependent changes occurring upstream of the dLGN-V1 synapse (for example, adaptation of the spike rate response in the dLGN), we also measured the spike rate response in the dLGN to the same visual pulse trains and normalized the fEPSP<sub>thal</sub> in V1 at each response number by the amplitude of that response in the dLGN multi-unit response. We also plotted the raw steady-state amplitude of the fEPSP<sub>thal</sub> response to the visual stimulus pulse train (Fig. 4e,f;

see “Visual stimulation to measure the depression of thalamo-cortical synapses”), comparing anesthetized and awake recordings.

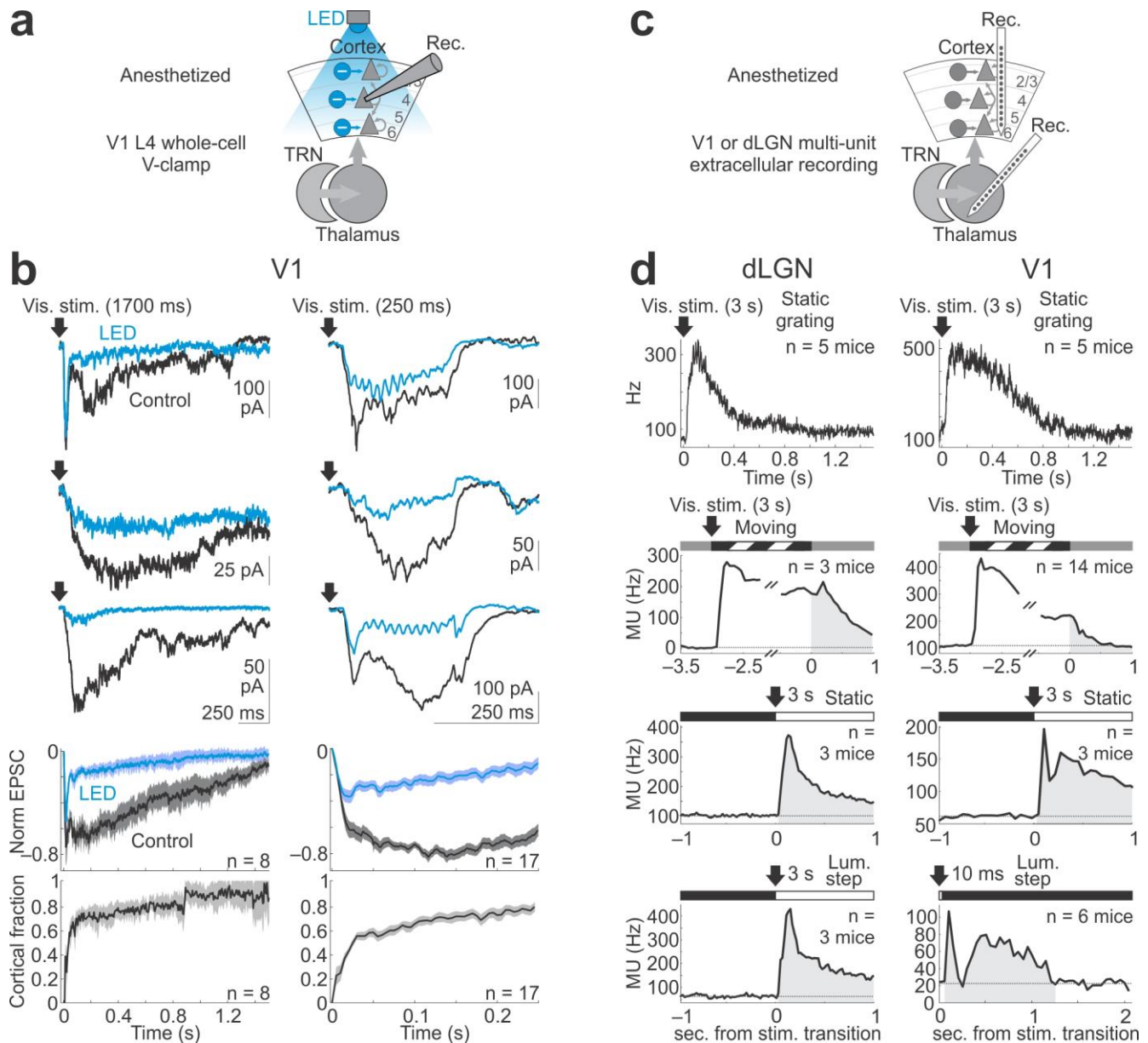
**Data analysis: predicting slowed cortical onset response when CDF is slowed.** In Supplementary Figure 10e, we predict the change in the cortical response to the onset of a visual stimulus (moving grating) when the CDF is slowed by suppressing cortical inhibitory interneurons. To do this, we first consider the onset of the sensory-evoked response in V1 (blue line in Supplementary Fig. 10e marked “Control”) to be the combination (convolution) of the time course of the response in the thalamus with the CDF under control conditions. Then we can ask: how does changing the CDF without changing the time course of response in the thalamus change the time course of response in the cortex? To analyze this, we first deconvolve the CDF under control conditions from the onset response in V1 under control conditions to arrive at the approximate onset response in the dLGN. Then we recombine (convolve) this onset response in the dLGN with the slower CDF under conditions of reduced cortical inhibition (dark purple line in Fig. 5d) to predict V1's onset response under conditions of reduced cortical inhibition. This convolution gives the prediction (lighter purple line in Supplementary Fig. 10e ii, marked “Prediction for disinhibition”). This prediction matches the actual time course (darker purple line in Supplementary Fig. 10e ii, marked “V1 disinh.”) of the V1 response to the onset of the visual stimulus under conditions of reduced cortical inhibition. We then plot how the onset response is slowed by V1 disinhibition as the time delay between the blue and dark purple onset curves. If these curves are the same and not shifted in time, the time delay should be centered around 0. However, we see that the time delay is centered at a few milliseconds lag, consistent with the prediction by the CDF.

**Statistics.** No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those generally employed in the field. Data collection and analyses were not performed blind to the conditions of the experiment.

A Supplementary Methods Checklist is available.

**Code availability.** Custom code is available upon request.

51. Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).
52. Zhao, S. *et al.* Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nat. Methods* **8**, 745–752 (2011).
53. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
54. Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
55. Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
56. Margrie, T.W., Brecht, M. & Sakmann, B. *In vivo*, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *PLoS Arch.* **444**, 491–498 (2002).
57. Paxinos, G. & Franklin, K.B.J. *The Mouse Brain in Stereotaxic Coordinates* compact 2nd edn. (Elsevier Academic, 2004).
58. Wang, Q., Gao, E. & Burkhalter, A. *In vivo* transcranial imaging of connections in mouse visual cortex. *J. Neurosci. Methods* **159**, 268–276 (2007).
59. Fee, M.S., Mitra, P.P. & Kleinfeld, D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J. Neurosci. Methods* **69**, 175–188 (1996).
60. Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z.J. & Scanziani, M. A neural circuit for spatial summation in visual cortex. *Nature* **490**, 226–231 (2012).
61. Goard, M. & Dan, Y. Basal forebrain activation enhances cortical coding of natural scenes. *Nat. Neurosci.* **12**, 1444–1449 (2009).
62. Mitra, P. & Bokil, H. *Observed Brain Dynamics* (Oxford Univ. Press, 2008).
63. Abbott, L.F., Varela, J.A., Sen, K. & Nelson, S.B. Synaptic depression and cortical gain control. *Science* **275**, 220–224 (1997).
64. Fuhrmann, G., Cowan, A., Segev, I., Tsodyks, M. & Stricker, C. Multiple mechanisms govern the dynamics of depression at neocortical synapses of young rats. *J. Physiol. (Lond.)* **557**, 415–438 (2004).



**Supplementary Figure 1**

**Time course of thalamic and cortical responses to visual stimuli.**

**(a)** Experimental configuration: whole-cell voltage-clamp of layer 4 (L4) neurons to record visually evoked excitatory post-synaptic currents (EPSCs) with or without optogenetic silencing of cortical excitatory recurrent circuits, as in Fig. 1b. Refers to (b).

**(b)** Time course of synaptic excitation, comparing thalamic and cortical components.

**Left:** Visual stimulus is static for 1.7 s.

**Right:** Visual stimulus is static for 250 ms.

**Top:** EPSCs from 6 example cells (arranged by thalamic fraction) in response to appearance of static grating (arrow), comparing control (black) versus cortical silencing with LED to isolate the thalamic EPSC (blue).

**Middle:** Mean EPSC and s.e.m. across all cells presented with same-duration visual stimulus. Left average (n=8) includes both 1.7 s-long static gratings and 1.7 s-long luminance steps. Right average is response to 250



ms-long static gratings. Time until thalamic component is < 50% of total EPSC amplitude: (1.7 s-long stimuli)  $45 \pm 13$  ms, (250 ms-long stimuli)  $42 \pm 8$  ms (mean  $\pm$  s.e.m.).

**Bottom:** Fraction of EPSC mediated by cortical excitation (cortical fraction, i.e., difference between black and blue, divided by black). Cortical fraction of the total integrated response: (1.7 s-long stimuli)  $73 \pm 23\%$ , mean  $\pm$  s.d., (250 ms-long stimuli)  $62 \pm 23\%$ .

(c) Schematic of extracellular recordings in dLGN or V1 of anesthetized mice.

(d) Time course of extracellular multi-unit activity in response to visual stimuli.

**Left column:** Multi-unit activity in dLGN.

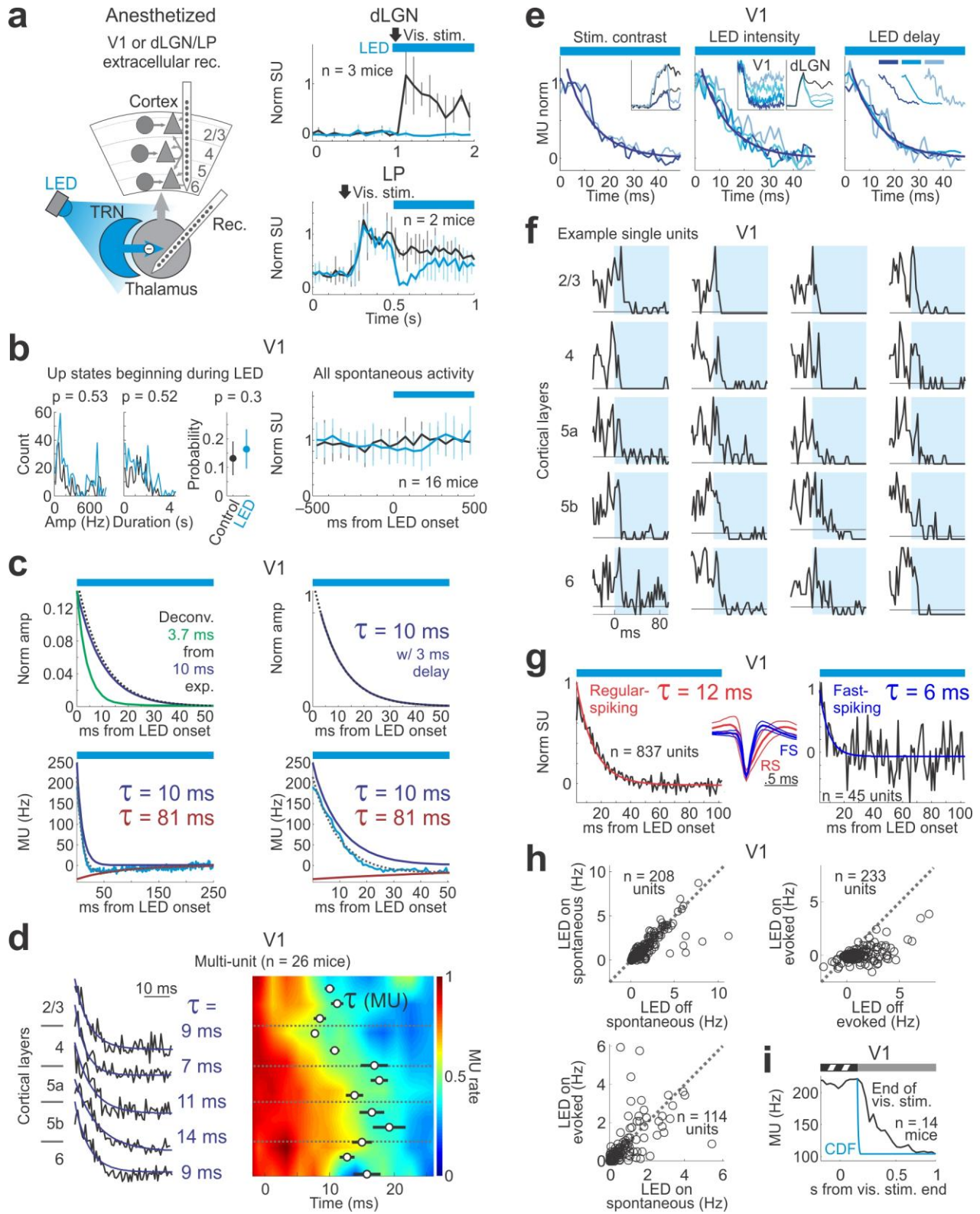
**Right column:** Multi-unit activity in V1 (different animals than dLGN recordings at left).

**Top row:** Average peri-stimulus time histogram (PSTH) of response to appearance of static grating (static for 3 s). V1 PSTH (right) is from Fig. 1c. Note that spike rate in V1 in response to the appearance of a static grating decreases more rapidly than the EPSC magnitude, possibly because intra-cortical synapses may depress more at the peak than later in the response<sup>63,64</sup> due to the higher spike frequency of cortical neurons at the peak response, because the multi-unit activity includes both excitatory and inhibitory neurons whereas the EPSC measures only excitation, because some component of the late-stage EPSC may originate outside of V1 (e.g., top-down input from other cortical areas), or for some combination of these reasons.

**Second row:** PSTH response to onset and offset of moving grating (moving for 3 s).

**Third row:** PSTH response to appearance of static stimulus (both 3 s-long static gratings and 3 s-long luminance step).

**Bottom row:** PSTH response to luminance step only (3 s duration, left) in dLGN or 10 ms-long luminance step (10 ms duration, occurring at 0 s) in V1 (right). Note that the response to even this very brief (10 ms-long) stimulus is long-lasting.



Supplementary Figure 2

Fitting the cortical decay function (CDF) in V1 of anesthetized mice.

For all parts, blue bar indicates LED illumination of TRN. Arrow indicates onset of visual stimulus.

**(a) Left:** Schematic of thalamic silencing by photo-activation of TRN. **Right:** (Top) TRN photo-activation prior to onset of visual stimulus prevents onset of a visual response in dLGN. (Bottom) TRN photo-activation suppresses ongoing visual response in LP. In these PSTHs, visual responses are to moving gratings (3 s), but silencing of the evoked response in dLGN was not different between responses to static or moving visual stimuli. At 300 ms into the response, as in Fig. 2c, the percentage of the evoked response suppressed in dLGN for static versus moving stimuli was: (static,  $n = 2$ )  $93 \pm 57\%$  (mean  $\pm$  s.e.m.), (moving,  $n = 12$ )  $95 \pm 28\%$ .

**(b)** Effect of thalamic silencing on spontaneous activity in V1 under anesthesia. Left to right: Effect of thalamic silencing on initiation of V1 Up states detected by LFP power ratio<sup>63</sup>. (Left) Histograms of Up state multi-unit (MU) amplitude and duration when Up states begin during period of thalamic silencing; (Middle) probability of Up state initiation over 1 s window during thalamic silencing (blue) or control (black) (Up states detected in  $n = 14$  mice); all p-values are from Wilcoxon rank-sum tests. (Right) Single-unit (SU) spontaneous activity from Up and Down states is summed within each recording session and normalized to 1. SU activity verifies that all spikes are from V1 neurons. Lines are mean, and error bars are s.d. of this spontaneous activity across mice. Spontaneous activity in V1 is not affected by silencing thalamus.

**(c) Fitting V1 CDF. Top Row:** To determine the exact time course of activity decay in V1 following thalamic silencing, i.e., the time course of the cortical decay function (CDF), we need to remove (deconvolve) the time course of shut-off in thalamus from the time course of shut-off in cortex. (Left) Deconvolution (dotted black) over 2-50 ms (using fits to dLGN = green and V1 = blue decays, see Methods). (Right) The deconvolution (dotted black) closely matches a 10 ms single exponential decay (blue) that begins at 3 ms after onset of the LED to silence the thalamus. Because the deconvolution introduces error in real, noisy data, we instead use a single exponential function, beginning at 3 ms after LED onset, to fit the CDF. **Bottom Row:** Fitting the CDF time course with two exponentials over longer timescales. (Left) Time course of average baseline-subtracted MU activity in V1 after silencing thalamus with LED. Dark blue: exponential fit component 1, a fast decay. Red: exponential fit component 2, a slow recovery component (red amplitude is 13% of dark blue). Gray dotted: sum of dark blue and red fits. (Right) As left panel but expanded timescale. Elsewhere in text we report a single exponential fit over initial 50-80 ms of the CDF, which is dominated by the fast component.

**(d) Left:** CDF across putative cortical layers (MU,  $n = 26$  mice) and fits (dark blue). **Right:** Heat-map normalized across recording depths. 1 (red) is MU rate before LED onset; 0 (blue) is baseline rate. LED onset at 0 ms. Dotted lines show putative layer boundaries based on current source density (Methods). Black overlay: Mean  $\pm$  s.e.m. of time constant of single exponential fit to CDF. Note slightly slower shut-off in deeper layers.

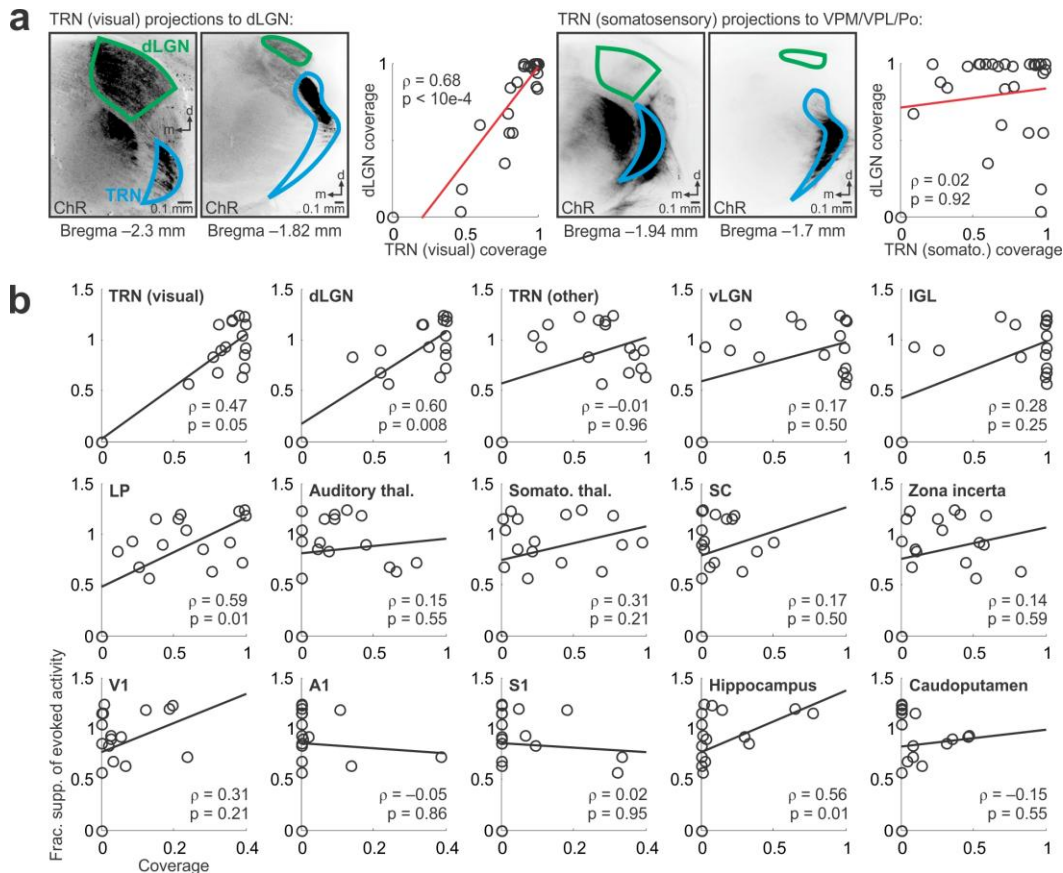
**(e)** The CDF (mean norm. MU) is not affected by stimulus contrast, LED intensity or LED onset delay. Traces scaled and superimposed. Dark blue is 11 ms fit from Fig. 2f. **Left:** Contrast 1: light blue. Contrast 0.25: dark blue ( $n = 5$  mice). Inset: Zoomed out MU evoked response in example mouse (thalamic silencing not complete) to low- (gray/dark blue) or high-contrast (light blue/black) moving grating with (light blue/dark blue) or without (black/gray) thalamic silencing; y axis: 200 Hz, x axis: 800 ms. **Middle:** As left but varying LED intensity to achieve 30-50%, 50-70%, 70-90%, or > 90% suppression of V1 evoked response ( $n = 2$  mice). Inset: (left inset) V1 CDFs un-normalized, (right inset) different levels of dLGN silencing during LED intensities 0.5, 3.6 and 6.5 mW; y axis: 150 Hz, x axis: 1.5 s, PSTHs smoothed with 250 ms window. **Right:** As left but for LED onset delays 200 ms, 300 ms or 900 ms relative to stimulus onset ( $n = 3$  mice). Inset: CDFs separated by onset delay, left to right: 200, 300, 900 ms. Supplementary Table 1 for more conditions.

**(f)** Example SU during visual stimulation at LED onset. Light blue shading: LED on. Gray line: mean baseline of each unit. Unit examples arranged by putative layer (see labels at left). All traces normalized to peak.

**(g) Left:** Mean norm. shut-off of regular-spiking (RS) units across cortical layers (red is fit). Inset shows mean and s.d. of waveforms for RS and fast-spiking (FS) units (Methods for separation of these unit types). **Right:** As at (left) but for fast-spiking (FS) units (blue is fit).

**(h)** SU activity at moment of silencing thalamus. Each point is a unit. **Top-Left:** Spontaneous Up states in V1 are unaffected. **Top-Right:** Visually evoked activity is suppressed (3 s-long moving visual stimulus). **Bottom-Left:** Analysis of whether each single unit in V1 returns to its own spontaneous baseline after thalamus is silenced during a sensory response. X axis is each unit's spontaneous rate (no visual stimulus, blank screen) during the LED, and Y axis is that unit's evoked response (moving grating present) during the LED. No systematic change from pre-visual stimulus spontaneous activity when thalamus is off during visual stimulus.

**(i)** Comparison of V1 CDF with V1 activity at offset of moving grating (from Supplementary Fig. 1d) on expanded timescale. Black and white angled bars: moving grating stimulus (3 s duration), followed by transition to gray screen. Blue curve: fit to shut-off of visually evoked activity following optogenetic silencing of thalamus (scaled to match level of activity before shut-off onset). Note much faster shut-off of visually evoked activity after silencing thalamus, with respect to shut-off in V1 after just turning off the visual stimulus.

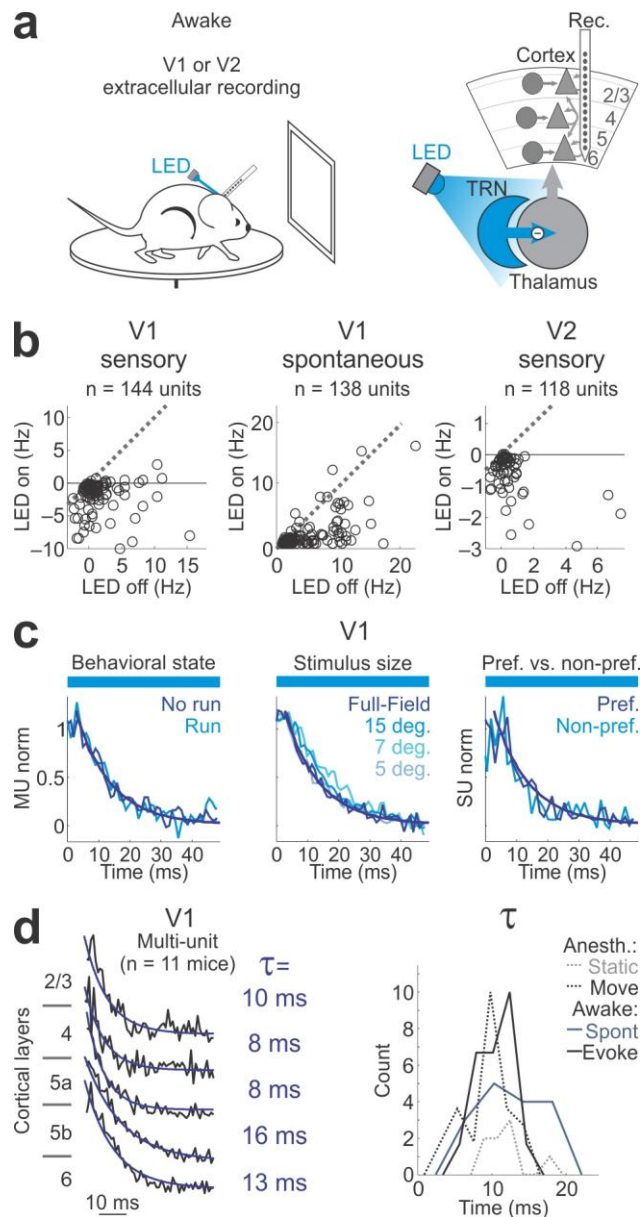


**Supplementary Figure 3**

### dLGN inhibition correlates with shut-off of visually evoked response in V1.

**(a)** Expression of fluorescent reporter TdTomato fused to Flexed-ChR2 in Gad2-Cre mice in two sets of coronal sections from two mice (first mouse: left two sections and left scatter, second mouse: right two sections and right scatter). Coverage is the fraction of pixels expressing TdTomato within region-of-interest (ROI) outlining given brain area. Scatter plots compare reporter coverage of TRN region with reporter coverage of dLGN. For brain structures with  $> 1$  mm anterior-posterior extent, coverage is average of coverage across three coronal sections spanning extent of structure. Coverage of TRN's caudo-dorsal sector (putative visual sector) but not TRN's somatosensory sector correlates with coverage of dLGN.  $\rho$  is correlation coefficient;  $p$  is  $p$ -value of correlation.

**(b)** Coverage, same as in (a), of different brain areas plotted against fractional suppression of V1 evoked response (frac. supp.) following photo-activation of TRN in anesthetized animals. Each point is one mouse.  $\rho$  is correlation coefficient;  $p$  is  $p$ -value of correlation. Coverage of TRN (visual sector) and dLGN are the only expression patterns necessary to observe shut-off of V1 evoked response (i.e., linear fit passes near the origin and has a positive slope). Note that expression of ChR2 in TRN axons to LP is not necessary to observe shut-off of V1 evoked response; even when ChR2 expression in axons to LP is weak, we may observe a strong shut-off of V1 evoked response. **Abbreviations:** V1 = primary visual cortex, A1 = primary auditory cortex, TRN = thalamic reticular nucleus, dLGN = dorsal lateral geniculate nucleus, vLGN = ventral lateral geniculate nucleus, IGL = intergeniculate leaflet, LP = lateral posterior nucleus, SC = superior colliculus, Auditory thal. = medial geniculate nucleus, S1 = primary somatosensory cortex, Somato. thal. = ventral posterior medial/lateral nuclei.



Supplementary Figure 4

### Fitting the CDF in awake mice.

**(a)** Awake extracellular recordings in V1 or V2 and optogenetic configuration to silence thalamus by photo-activating TRN.

**(b)** Single-unit (SU) activity at moment of silencing thalamus. Each point is a unit. Dotted lines are unity. Points below unity line are suppressed.

**Left:** Visually evoked activity (pre-stimulus baseline-subtracted) in V1 of awake mice is suppressed. Points below zero (solid horizontal gray line) are suppressed below baseline.

**Middle:** Spontaneous activity in V1 of awake mice is also suppressed by silencing thalamus. Not baseline-subtracted.

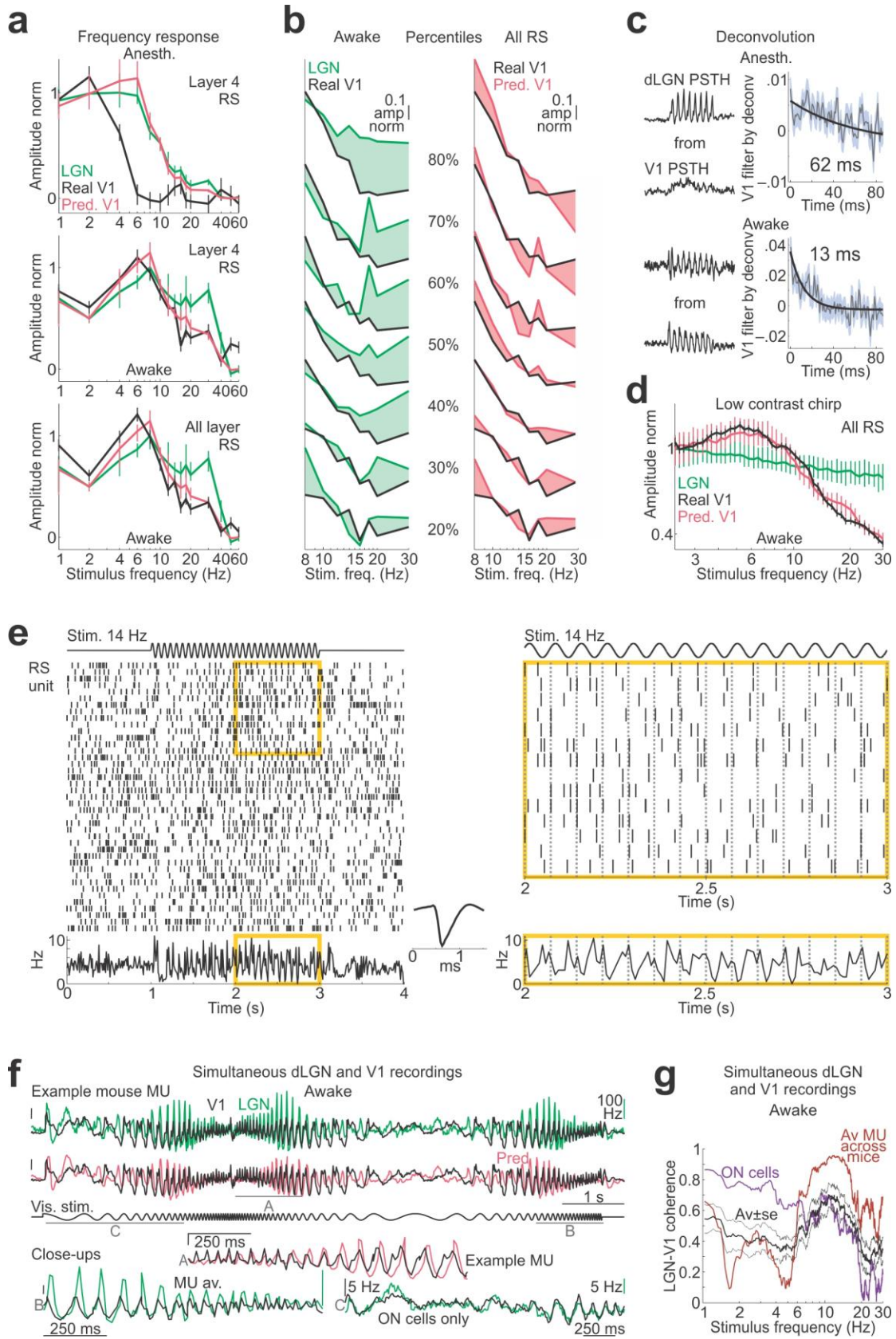
**Right:** Visually evoked activity (baseline-subtracted) in V2 of awake mice is suppressed. Points below zero (solid horizontal gray line) indicate suppression below baseline.

**(c)** V1 recording. Cortical decay function (CDF) is independent of animal's behavioral state (left; superimposed MU shut-off for running and non-running trials;  $n = 4$  mice), stimulus diameter (middle; as MU; 5, 7, 15 degrees or full-field;  $n = 3$  mice), and single unit's stimulus preference (i.e., grating orientation; mean SU during stimulation with preferred versus non-preferred orientation of moving grating;  $n = 127$  SU). Dark blue fit is 11 ms, as in Fig. 2f. See Supplementary Table 1 for more conditions.

**(d)** V1 recording.

**Left:** Single exponential fit (blue) to cortical decay function (MU, black) across putative cortical layers ( $n = 11$  awake mice).

**Right:** Histogram of time constants of single exponential fits to CDF (MU) across conditions. This shows that CDF distribution is approximately normal. For parametric statistical tests of the CDF across conditions reported in paper, CDF distribution was assumed to be normal and variances were assumed to be equal across groups, but this was not formally tested. Count is mice. See Supplementary Table 1 for further break-down.





## Supplementary Figure 5

### CDF predicts amplitude of V1 response to high- and low-contrast flicker stimuli while recording sequentially or simultaneously in dLGN and V1.

**(a) Top:** Anesthetized. Frequency response under anesthesia of regular-spiking (RS) units in Layer 4 (black;  $n = 30$  units; mean and s.e.m.). dLGN (green) and predicted V1 (pink) frequency response from left panel of Fig. 4c. **Middle:** Awake. Frequency response ( $n = 30$  units) of RS units in Layer 4 (black;  $n = 30$  units; mean and s.e.m.). dLGN (green) and predicted V1 (pink) frequency response from right panel of Fig. 4c. **Bottom:** Awake. Frequency response of RS throughout all layers (black;  $n = 270$  units; mean and s.e.m.). dLGN (green) and predicted V1 (pink) frequency response from right panel of Fig. 4c. Note that, consistent with V1 response curves in Fig. 4, we observe theta-band amplification (4-11 Hz)<sup>37</sup> in V1 response (see Supplementary Fig. 7).

**(b)** Attenuation predicted by CDF over full response amplitude distribution. Frequency response as percentiles of single-unit (SU) amplitudes (peak-norm., see Methods for alignment, normalization and scaling). **Left:** Percentiles of dLGN frequency response (green) vs. same percentiles of V1 RS SU frequency response (black). Note that V1 amp. is below dLGN amp. at high frequencies for all percentiles of response amplitude distributions. **Right:** Black as left. Pink is percentiles of prediction for V1 by convolving dLGN frequency response with CDF. Shaded areas are differences between curves. At all percentiles, prediction for V1 after filtering by CDF is much better match to real V1 response than dLGN response without filtering by CDF.

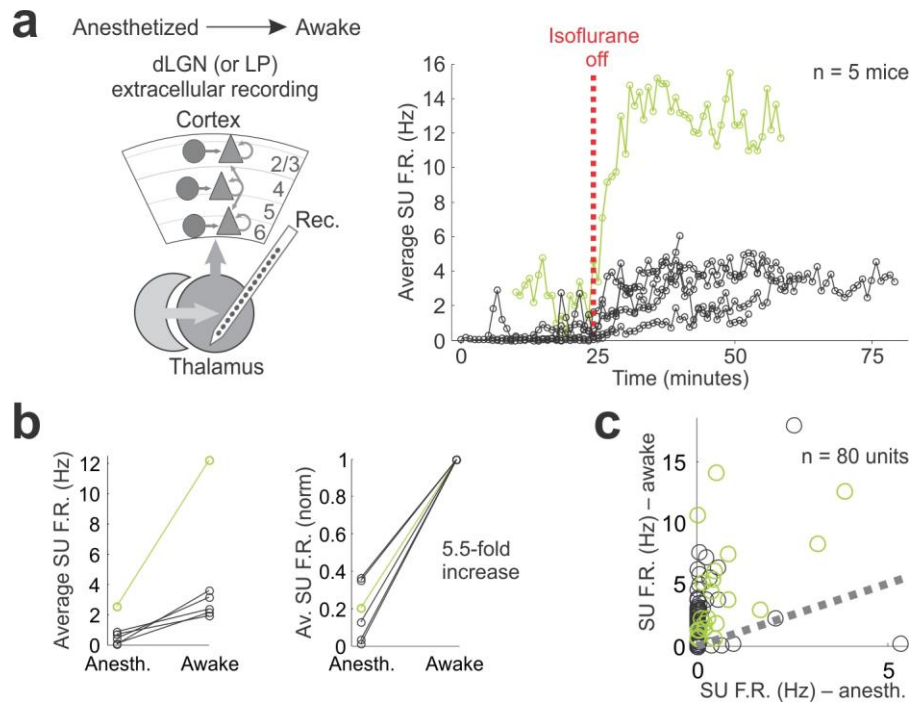
**(c)** Deconvolution of dLGN PSTH response to flicker stimulus from V1 PSTH response to same stimulus (across all temporal frequencies, Methods) in anesthetized (**Top**) and awake (**Bottom**) mice. Gray line: mean deconvolution across all thalamus-V1 pairs with s.e.m. (gray-blue error). Black: single exponential fit and time constant value. Note much slower dynamics under anesthesia as compared to awake.

**(d)** Frequency response of dLGN and V1 to low contrast (20%) frequency chirps (vis. stimuli are frequency-modulated sweeps, shown below in (f)). Simultaneous dLGN and V1 recordings. Green = median frequency response of dLGN units and 45<sup>th</sup> to 55<sup>th</sup> percentile error bars, as in Fig. 4c. Black = mean, s.e.m. of regular-spiking (RS) cells across layers. Pink = prediction for V1, i.e., green filtered by CDF. Zero is noise (for alignment and scaling, see Methods).

**(e)** Example single RS unit response to 14 Hz flicker. **Top:** raster, **Bottom:** PSTH. Inset: unit waveform. Yellow: close-ups.

**(f)** Simultaneous extracellular recordings in dLGN and V1. Visual stimulus is logarithmic frequency modulation of full-field luminance (chirp). Unless specified, vertical scale bars are 100 Hz. **Top traces:** Superimposed MU PSTH from dLGN (green) and V1 (black) of example mouse in response to luminance modulation shown below (Vis. stim.). **Middle traces:** V1 response prediction (pink; by convolution of dLGN response with CDF) superimposed on actual V1 response (black, from top). Note that pink predicts extent of amplitude filtering at higher frequencies. **Bottom traces:** Pink, black: Expanded timescale of segment A (above). Green, black: Averaging MU across mice (left) or including only ON-responsive units (right) improves coherence between dLGN and V1 traces. Responses are to chirp sections marked B and C.

**(g)** Coherence between dLGN and V1 recorded simultaneously. Black: mean and s.e.m. of coherence (between mean MU dLGN vs. V1 within each mouse) across mice. Red: Coherence between averaged-across-mice MU responses in dLGN and V1. Purple: Coherence between ON-responsive dLGN and V1 units in example mouse.



Supplementary Figure 6

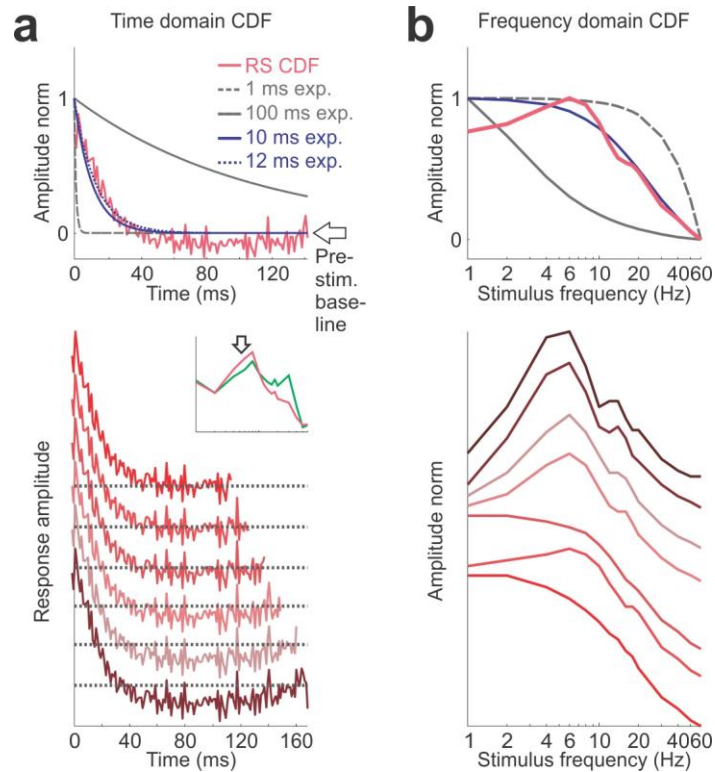
**Waking up from anesthesia increases baseline spiking in the thalamus.**

(a) Average thalamic single-unit (SU) spontaneous firing rate over time as mice wake up from isoflurane anesthesia (5 curves are 5 example recordings; anesth. average SU rate = 0.8 Hz, awake average SU rate = 4.2 Hz). Light green: recording site in LP; black: 4 sites in dLGN.

(b) **Left:** Average spontaneous SU firing rate under anesthesia and awake.

**Right:** Normalized to final rate in awake. Colors as in (a).

(c) Average spontaneous SU firing rate under anesthesia (isoflurane) versus SU average rate in awake state. Dotted gray line is unity. Points above dotted line have higher firing rates in the awake state. Light green dots are units in LP. Black dots are units in dLGN.



Supplementary Figure 7

### Theta-band amplification by the CDF.

The slow, small recovery component of the CDF (dip below baseline, see Supplementary Fig. 2c) predicts theta-band amplification in V1's frequency response.

#### (a) CDF in time domain.

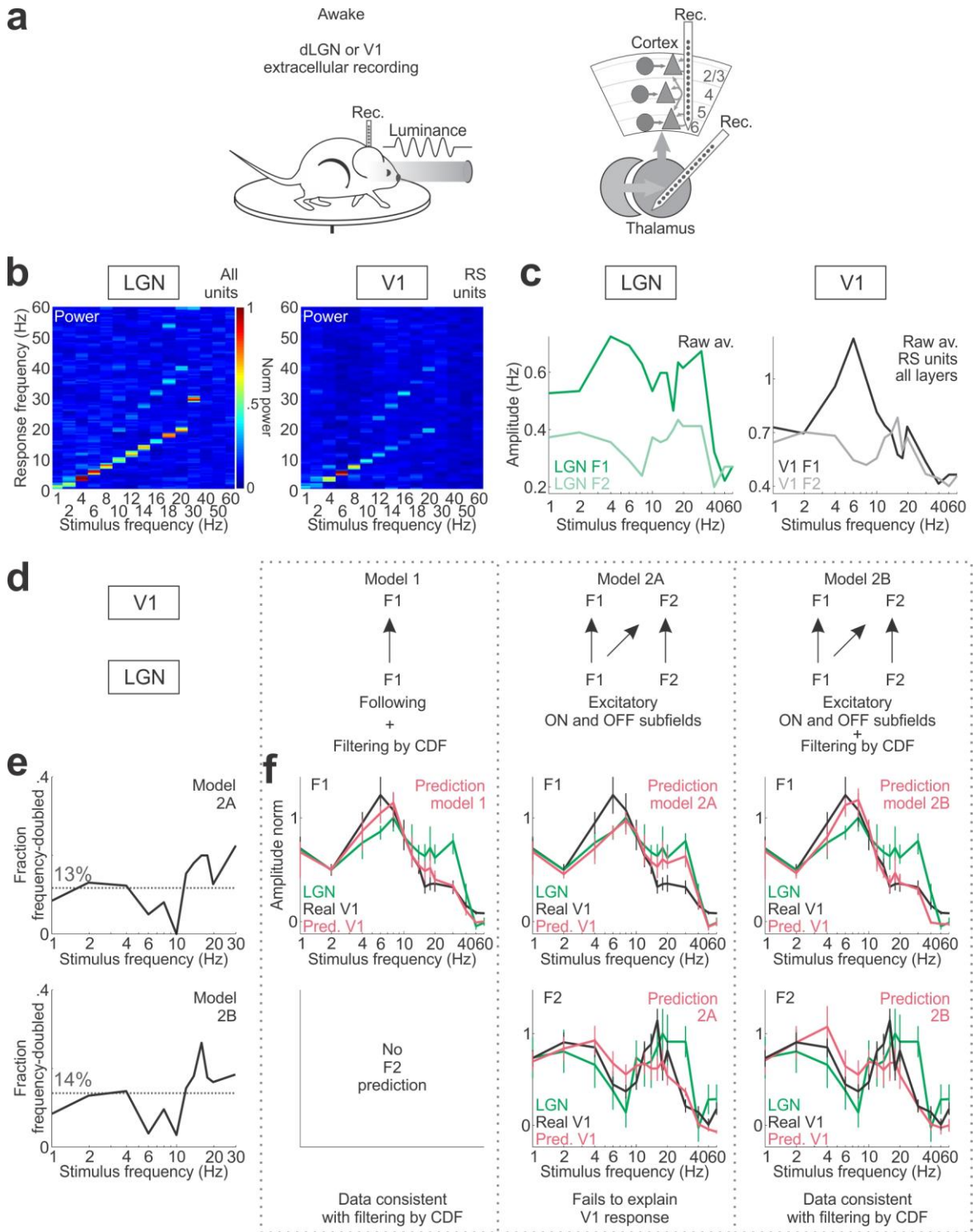
**Top:** Pink: CDF of RS units. Curve is zeroed at pre-stimulus baseline activity (arrow). Note both fast shut-off, fit by 12 ms decaying single exponential, and second small, slow recovery component. Dark blue, dotted blue, solid gray, dotted gray: single exponential decays with time constants as indicated; 1 and 100 ms time constants shown for comparison.

**Bottom:** Magnitude of slow recovery component as a function of where the baseline of CDF is set. The higher the baseline (dotted gray line), the more pronounced the slow recovery component of CDF (from red to brown). In this study, we choose the baseline of the CDF (under anesthesia) to be the mean pre-stimulus baseline, because thalamic silencing does not affect spontaneous activity under anesthesia. However, in each experiment, the CDF relaxes to a baseline that depends on the strength of thalamic silencing; thus, the baseline of the CDF may be offset with respect to the pre-stimulus baseline. Curves are offset on ordinate for clarity. Inset: green and pink from Fig. 4c (right). Arrow shows CDF amplification of stimulus frequencies in theta band.

#### (b) CDF represented in the frequency domain.

**Top:** Pink: amplitude frequency spectrum of pink curve in (a). This pink filter is used to predict V1's response. Colors as in (a) but spectra are in frequency domain (all scaled).

**Bottom:** Frequency-domain forms of CDFs in bottom panel of (a) (colors are matched). Increasing the magnitude of the slow recovery component of CDF increases theta-band amplification (i.e., the 4-11 Hz component of the frequency-domain form of the CDF).



Supplementary Figure 8

Both fundamental (F1) and first harmonic (F2) components of V1 response are consistent with filtering by CDF.

This figure rules out an alternate model where responses at the fundamental (F1) frequency in dLGN are converted to first harmonics (F2) in V1 without filtering by the cortical decay function (CDF). We find that both F1 and F2 components in V1 are attenuated, consistent with filtering by the CDF. A transformation from F1 in dLGN to F2 in V1 should instead increase F2 in V1.

**(a)** Experimental set-up to measure frequency response in awake mice recording in either dLGN or V1.

**(b) Left:** Heatmap is median power of dLGN single-unit response to flicker stimulus in awake mice as a function of stimulus frequency and response frequency. F1 response is the diagonal where response frequency matches stimulus frequency. F2 is the diagonal where response frequency is twice the stimulus frequency.

**Right:** Same as left but for V1 regular-spiking (RS) units in awake.

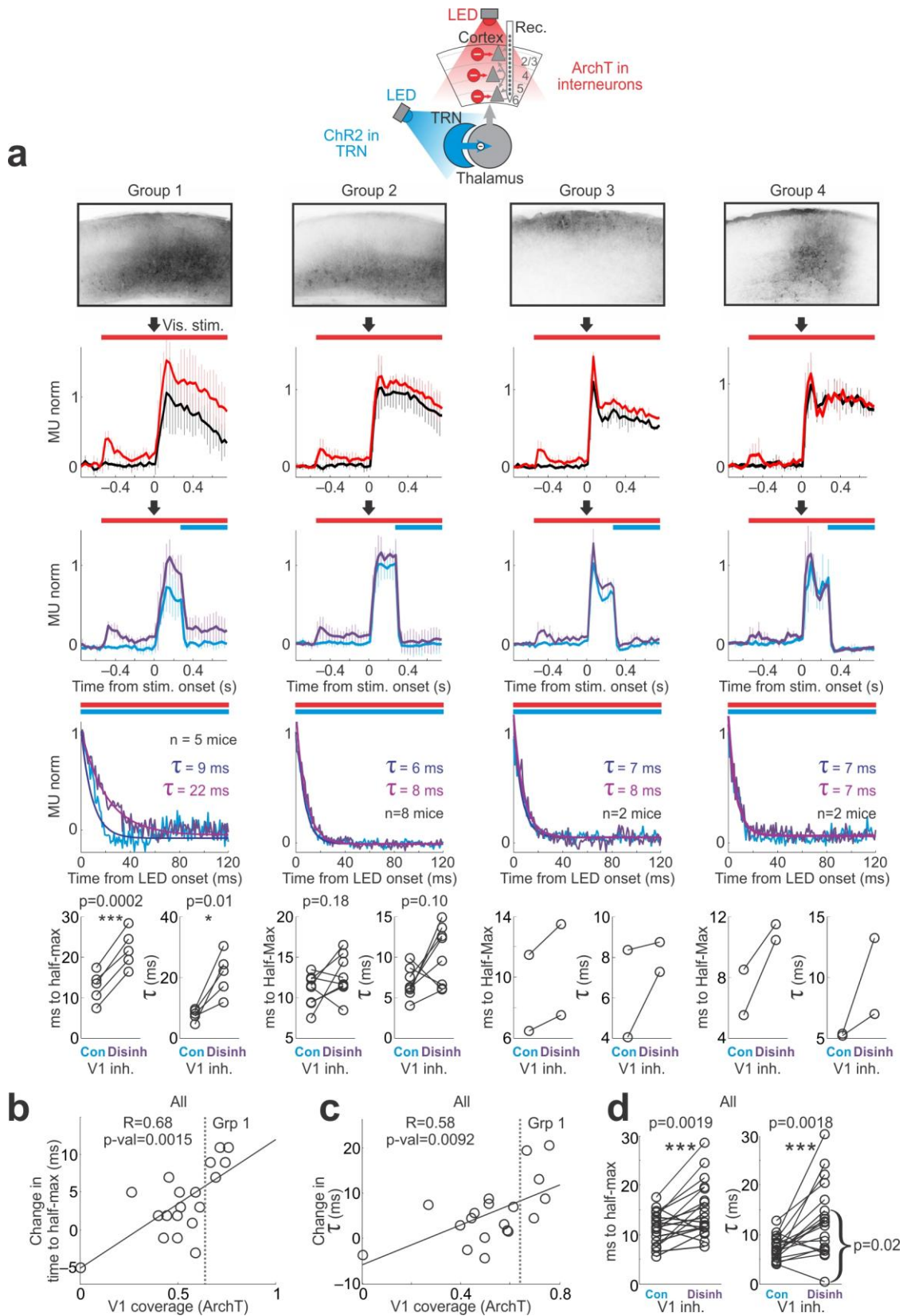
**(c)** Amplitude of F1 and F2 response components from heatmaps in (a). Includes all units, including units that are well-driven, as well as units that are not well-driven, by the stimulus, thus the low spike rate. **Left:** for dLGN relay units. **Right:** for V1 RS units. Consistent with data from cats<sup>31</sup>, V1's F1 response dips between 10 and 20 Hz, while the F2 response dips around 8 Hz.

**(d)** Simple models of dLGN-to-V1 connectivity affecting spatial receptive field structures. **Model 1:** F1 response in dLGN drives F1 response in V1. No consideration of F2. This is the model presented in Fig. 4.

**Model 2A:** F1 response in dLGN drives some F1 in V1 and some F2 in V1. F1-to-F2 transformation between dLGN and V1 is consistent with certain models<sup>2</sup> of the visual system. Also, F2 response in dLGN drives some F2 in V1. In Model 2A, no amplitude filtering by the CDF. **Model 2B:** Same as Model 2A, but Model 2B does include amplitude filtering by the CDF. See Methods for equations associated with models.

**(e)** Fraction of F1 response in dLGN that frequency-doubles in cortex to produce an F2 response in V1, derived from Model 2A (**top**) or 2B (**bottom**). See Methods for associated equations and derivation. Y axis is the fraction of response that is frequency-doubled as a function of stimulus frequency, according to best fit models. Above 30 Hz enters noise (so not shown). Dotted gray is average across stimulus frequencies.

**(f)** Different predictions for V1 F1 and F2 response components by Models 1, 2A and 2B. **Left:** Prediction by Model 1, as in Fig. 4. **Middle:** Best prediction for both F1 and F2 components of V1 response according to Model 2A. Prediction does not match actual V1 F1 response. **Right:** Best prediction for both F1 and F2 components of V1 response according to Model 2B. By including filtering by the CDF in this model, this model can better account for the average amplitude of both the F1 and F2 response components in V1 RS cells.



Supplementary Figure 9

## Slowing of V1 CDF correlates with widespread ArchT expression in inhibitory interneurons of visual cortex.

(a) All mice tested, subdivided according to ArchT expression profile in V1.

**Top Row:** Expression of ArchT-GFP in visual cortex. Overlay of one coronal section from each mouse in Group. Grp. 1: widespread ArchT (n = 5 mice; these are 5 mice in Fig. 5), Grp. 2: deep layers (n = 8 mice), Grp. 3: superficial layers (n = 2 mice), Grp. 4: radially restricted (n = 2 mice).

**Row 2:** Increase in gain of V1 multi-unit (MU) activity by ArchT photo-stimulation. Black: control, Red: ArchT photo-stimulation.

**Row 3:** V1 MU activity with thalamic silencing. Blue: control, Purple: ArchT photo-stimulation. Arrow is onset of moving grating. Blue horizontal bar (blue LED) is photo-activation of TRN (thalamic silencing). Red horizontal bar (amber LED) is ArchT photo-stimulation. Error bars are s.d. across PSTH.

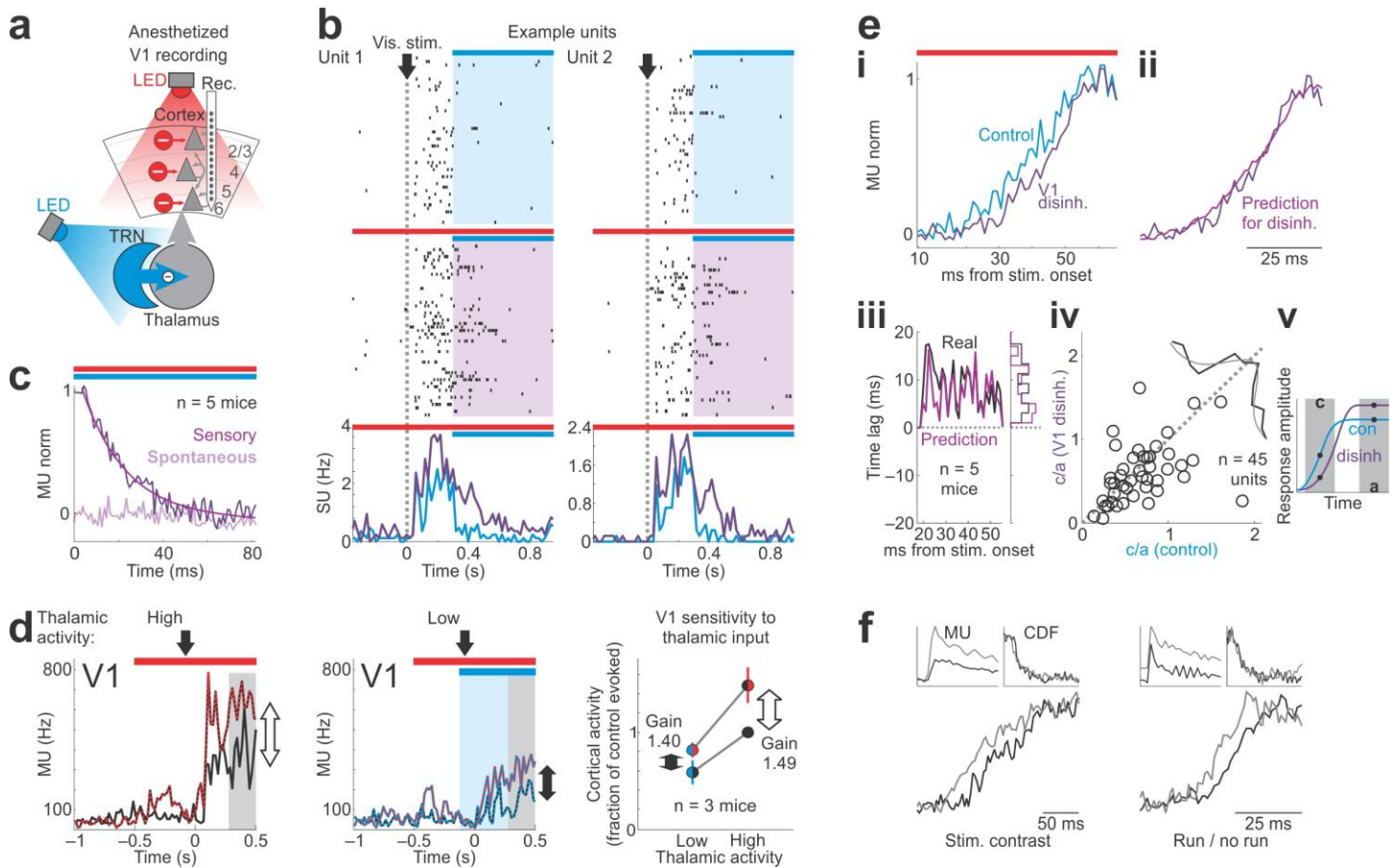
**Row 4:** Time course of V1 MU (baseline-subtracted) after silencing thalamus. Blue: control, Purple: ArchT photo-stimulation.

**Bottom:** Quantification of shut-off time course (stats. are two-sided paired t-tests; data distributions were assumed to be normal and variances were assumed to be equal across groups, but this was not formally tested – see figure for all data points).

(b) Coverage (fraction of pixels expressing ArchT within visual cortex) vs. change in time to half-max of V1 shut-off as a result of ArchT photo-stimulation. Grp. 1 includes the 5 mice with greatest ArchT coverage of visual cortex. R is correlation coefficient; p is p-value of correlation. Note strong correlation between expression of ArchT and slowing of CDF.

(c) Same as (b), but metric is time constant of single exponential fit to V1 shut-off. R is correlation coefficient; p is p-value of correlation. Note strong correlation.

(d) Quantification across all mice tested. Significant slowing of V1 shut-off with ArchT photo-stimulation across all mice (p = 0.00018) and even across mice when excluding the 5 with the greatest ArchT coverage (i.e., excluding 5 with the strongest effect on CDF, p = 0.02). Two-sided paired t-tests comparing interleaved conditions. Data distributions were assumed to be normal and variances were assumed to be equal across groups, but this was not formally tested.



Supplementary Figure 10

### Reducing cortical inhibition to slow the CDF slows onset of V1 response.

(a) Schematic of dual optogenetic manipulation, as in Fig. 5.

(b) Two example single units at moment of thalamic silencing with (purple) or without (blue) V1 disinhibition.

(c) No further effect of silencing thalamus during cortical disinhibition on spontaneous activity in V1. Dark purple curve is CDF during cortical disinhibition from Fig. 5d. Light purple: Same time window as dark purple, i.e., activity in V1 at moment of silencing thalamus and during cortical disinhibition, but here without visual stimulus (spontaneous activity).

(d) **Suppressing cortical inhibitory interneurons does not differentially increase cortical sensitivity to low, relative to high, levels of thalamic activity.**

Fig. 5 suggests that suppressing the activity of inhibitory interneurons in cortex leads to prolonged dynamics (CDF) within cortical recurrent circuits. To show that the site of the prolonged CDF is cortical recurrent circuits, we need to rule out a change in cortical sensitivity to thalamic input as an alternate explanation for the prolonged CDF. If thalamic silencing were incomplete and if suppressing cortical inhibitory interneurons were to increase the gain of V1's response differentially to low levels of thalamic input, we would expect a larger response in V1 when levels of thalamic input are low, that is, after onset of the blue LED to suppress activity in thalamus. This change in the transfer function between thalamus and cortex could explain the prolonged CDF observed during cortical disinhibition **only if the gain of the V1 response is higher at low levels of thalamic input than it is at high levels of thalamic input**. If the gain of the V1 response is increased by cortical disinhibition equally for high and low levels of thalamic input, then we expect the activity in V1 to be scaled up

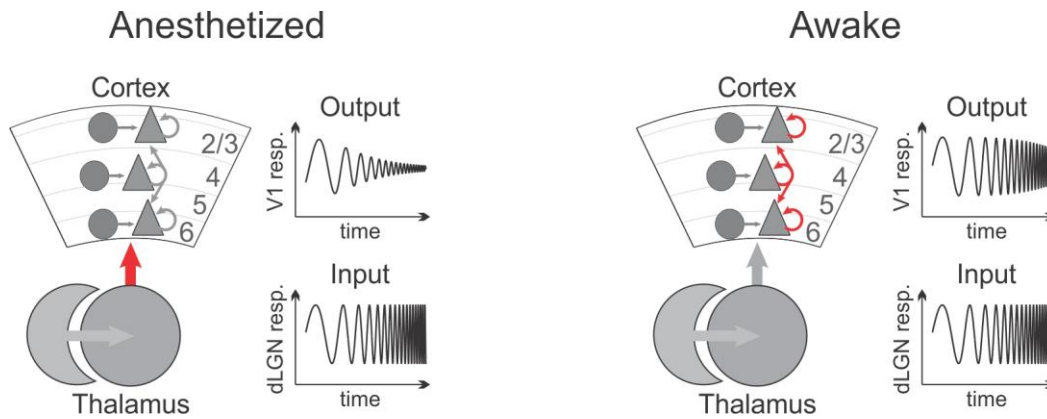


by a constant factor. Scaling the CDF by a constant factor does not change its time course. Thus, a simple change in gain of V1's response to thalamic input cannot explain the observed change in V1 dynamics. Here we show a control experiment indicating that cortical disinhibition does not make V1 more sensitive to low, relative to high, levels of thalamic input. Thus, the transfer function between thalamus and cortex as a function of the level of activity is only scaled by a constant factor. Therefore the source of prolonged cortical dynamics (the prolonged CDF) is likely mechanisms of cortex independent of thalamic input.

**First Two Columns:** Multi-unit (MU) PSTH responses recorded in V1 for different interleaved LED conditions. Both the blue LED and the amber LED turn on before the visual stimulus in this experiment. **Black:** Control (no V1 disinhibition). **Red-Black:** Amber LED to suppress inhibitory interneurons in cortex (V1 disinh.). **Blue-Black:** Low thalamic activity is achieved by low-intensity blue LED illumination of TRN over first 500 ms of visually evoked response. **Red-Blue:** Both amber and blue LEDs (to partially suppress thalamic activity during V1 disinhibition). **Left:** No suppression of visually evoked response in thalamus; thus, response to high-contrast visual stimulus (moving grating onset) is high thalamic activity. Two-headed white arrow indicates increase in visually evoked activity at steady-state (gray shaded area) upon V1 disinhibition. **Middle:** Lowering thalamic activity through partial suppression of visually evoked response in thalamus by blue LED. Two-headed black arrow indicates increase in visually evoked activity at steady-state (gray shaded area) upon V1 disinhibition. **Right:** V1 disinhibition does not produce a greater gain increase at low, relative to high, levels of thalamic activity. Amplitude of cortical activity during low or high thalamic activity. Circle colors same as colors in plots to left. Gain, computed during steady-state (gray shaded area) is (black arrow) Red-Black over Black or (white arrow) Red-Blue over Blue-Black. Circles show mean response amplitude at steady-state (gray shaded areas), normalized to the response in control (black circle), across 3 mice.

**(e) Slowing the CDF slows V1's onset response. (i)** Effect of control (blue) or slowed (purple) CDF on time course of response in V1 (MU) to visual stimulus onset. Interleaved LED conditions. Same mice as in Fig. 5d. Note lagged onset response following V1 disinhibition. **(ii)** Purple curve is from left. Lighter purple: Prediction of lagged V1 response to stimulus onset given the measured change in the CDF during V1 disinhibition (Methods). **(iii)** Quantification of average lag ( $n = 5$  mice) over the initial 50 ms of the onset response (black) and of predicted lag (light purple). Positive values indicate delayed onset response. Note good match between experimental (black) and predicted lag (light purple). Left axis y-axis shows histogram of experimental (black) and predicted (light purple) lags. **(iv)** Slowing of single-unit (SU) onsets upon slowing the CDF, computed in a manner analogous to Fig. 5f scatter plot (the plot shows the ratio of SU firing rate during initial 50 ms of the onset response, called time window c, divided by firing rate during the subsequent 50 ms, which was 50-100 ms after response onset, called time window a). **(v)** Cartoon of onset response showing how higher values of this ratio (c over a) for control (blue) indicate faster rise in control conditions vs. V1 disinh. (purple). P-value (two-sided paired t-test; data distribution was assumed to be normal and variances were assumed to be equal across groups, but this was not formally tested – see figure for all data points) for SU slowing is  $p = 0.04$ .

**(f)** Increasing gain of V1 response by varying vis. stim. contrast (**Left**, gray = high contrast, black = low contrast) or behavioral state (**Right**, gray = running, black = non-running) leads to faster, not slower, rise of onset response.



Supplementary Figure 11

**Rate-limiting afferent or recurrent dynamic control of responses in visual cortex.**

Sites of rate-limiting processes are colored in red. Under anesthesia (left), depression at thalamic afferent synapses is rate-limiting and strongly low-pass filters input from dLGN. In awake mice (right), both low- and high-frequency input from dLGN pass equally well through thalamic afferent synapses. In the awake state, the dynamics of intra-cortical recurrent circuits become rate-limiting.

**Supplementary Table 1: Network time constant (NTC) fit to CDF across conditions.**

Condition		Mean NTC (ms) and s.e.	Mean ms to half-max	# mice
Anesthetized, across layers	L2/3	9.8 ± 1.7	4.8 ± 1.1	26
	L4	9.0 ± 2.2	5.4 ± 0.9	26
	L5a	8.9 ± 1.3	6.8 ± 0.9	26
	L5b	15.7 ± 2.5	9.4 ± 0.9	26
	L6	7.6 ± 1.5	6.8 ± 0.9	26
		1D bal. ANOVA p = 0.01	1D bal. ANOVA p = 0.01	
		L5a faster than L6	L2/3 and L4 faster than L5b	
Anesthetized, LED onset delay	200 ms	11.8 ± 2.6		6
	300 ms	9.0 ± 0.6		26
	600 ms	13.8 ± 2.1		3
	900 ms	8.1 ± 4.5		2
		1D unbal. ANOVA p=0.14		
Anesthetized, stimulus contrast	0.1	10.4 ± 3.5		3
	0.4	11.3 ± 3.1		4
	0.7	10.0 ± 1.8		3
	1	9.0 ± 0.6		26
		1D unbal. ANOVA p=0.63		
Anesthetized, stimulus type	Moving grating	9.0 ± 0.6		26
	Static, all	11.0 ± 0.6		8
			two-tailed unpaired t-test p=0.03	
	Static, full-field flash	10.1		1
	Static, reversing checkerboard	12		1
	Static, grating	11.0 ± 2.7		5
Anesthetized, fractional suppression	Static, 100 ms full-field flash	7.5		1
	30-50%	10.5 ± 1.0		4
	50-70%	14.0 ± 8.8		2
	70-90%	12.0 ± 2.1		5
	> 90%	17.4 ± 3.1		5
		1D unbal. ANOVA p=0.65		
Brain state, anesthetized vs. awake	Anesthetized, moving grating	9.0 ± 0.6		26
	Awake, moving grating	10.5 ± 0.7		9
		two-tailed unpaired t-test p=0.15		
Awake, across layers	L2/3	6.1 ± 0.5	4.1 ± 1.2	11
	L4	7.1 ± 2.3	3.2 ± 0.6	11
	L5a	9.7 ± 2.5	5.6 ± 1.8	11
	L5b	14.1 ± 1.9	14.0 ± 2.6	11
	L6	14.3 ± 2.1	6.1 ± 0.8	11
			1D bal. ANOVA p=0.03	1D bal. ANOVA p=9.2e-5
		5b slower than 2/3, 4	L5b slower than all others	
Awake, stimulus contrast	0.1	15.0 ± 3.4		5
	0.4	14.4 ± 2.5		7
	0.7	11.6 ± 2.9		5
	1	10.5 ± 0.7		9
		1D unbal. ANOVA p=0.24		
Awake, stimulus size	5 degrees	14.0 ± 4.3		4
	7 degrees	14.6 ± 2.9		3
	15 degrees	13.7 ± 1.6		6
	Full-field	10.5 ± 0.7		9
		1D unbal. ANOVA p=0.21		
Awake, stimulus type	Moving grating	10.5 ± 0.7		9
	Static, all	13.1 ± 1.0		11
			two-tailed unpaired t-test p=0.04	
	Static, full-field flash	13.0 ± 1.6		9
	Static, reversing checkerboard	13.5 ± 1.9		5
	Static, grating	12.5 ± 2.7		3
Awake, behavioral state	Static, 100 ms full-field flash	14.0		1
	Running	9.4 ± 3.1		4
	Non-running	10.3 ± 3.0		4
		two-tailed paired t-test p=0.16		
Awake, evoked vs. spontaneous	Moving grating	10.5 ± 0.7		9
	Spontaneous baseline	13.7 ± 1.5		7
		two-tailed unpaired t-test p=0.08		
Awake, V1 vs. V2	V1	10.5 ± 0.7		9
	V2	15.8 ± 2.4		3
		two-tailed unpaired t-test p=0.003		

**Supplementary Table 1:** Network time constant (NTC) fit to CDF across conditions. CDF distributions were assumed to be normal and variances were assumed to be equal across groups (see Supplementary Fig. 4d for CDF distribution and discussion of normality).