Fundamental bounds on the fidelity of sensory cortical coding

https://doi.org/10.1038/s41586-020-2130-2

Received: 1 August 2017

Accepted: 21 January 2020

Published online: 18 March 2020

Check for updates

Oleg I. Rumyantsev[1,2,3]✉, Jérôme A. Lecoq[1,2,4,5], Oscar Hernandez[1,2], Yanping Zhang[1,2,6], Joan Savall[1,2,6], Radosław Chrapkiewicz[1,2], Jane Li[1,4], Hongkui Zeng[5], Surya Ganguli[1,3]✉ & Mark J. Schnitzer[1,2,3,4,6]✉

How the brain processes information accurately despite stochastic neural activity is a longstanding question[1]. For instance, perception is fundamentally limited by the information that the brain can extract from the noisy dynamics of sensory neurons. Seminal experiments[2,3] suggest that correlated noise in sensory cortical neural ensembles is what limits their coding accuracy[4–6], although how correlated noise affects neural codes remains debated[7–11]. Recent theoretical work proposes that how a neural ensemble's sensory tuning properties relate statistically to its correlated noise patterns is a greater determinant of coding accuracy than is absolute noise strength[12–14]. However, without simultaneous recordings from thousands of cortical neurons with shared sensory inputs, it is unknown whether correlated noise limits coding fidelity. Here we present a 16-beam, two-photon microscope to monitor activity across the mouse primary visual cortex, along with analyses to quantify the information conveyed by large neural ensembles. We found that, in the visual cortex, correlated noise constrained signalling for ensembles with 800–1,300 neurons. Several noise components of the ensemble dynamics grew proportionally to the ensemble size and the encoded visual signals, revealing the predicted information-limiting correlations[12–14]. Notably, visual signals were perpendicular to the largest noise mode, which therefore did not limit coding fidelity. The information-limiting noise modes were approximately ten times smaller and concordant with mouse visual acuity[15]. Therefore, cortical design principles appear to enhance coding accuracy by restricting around 90% of noise fluctuations to modes that do not limit signalling fidelity, whereas much weaker correlated noise modes inherently bound sensory discrimination.

The sensitivity and noise fluctuations of primary sensory neurons, such as photoreceptors or mechanoreceptors, limit the perception of weak stimuli[16–18], although disagreement persists about which downstream noise sources limit perceptual discriminations when sensory inputs exceed detection thresholds[4–14]. A groundbreaking experiment spurred this debate by identifying individual visual cortical neurons that signal visual attributes nearly as reliably as an animal's perceptual reports[2,3]. One proposed explanation is that similarly tuned cortical neurons might share positively correlated noise fluctuations that limit the perceptual improvements attainable by averaging signals from multiple cells with similar response properties[2,4] (Extended Data Fig. 1a–c).

Theoretical studies show that positively correlated noise limits the information that cells with similar sensory-evoked responses can encode[4,5,7], but this is not necessarily the case for ensembles of cells with diverse tuning properties[8–10] (Extended Data Fig. 1d–f). A recent framework based on a feedforward neural network asserts that, in the space of all possible neural ensemble dynamics, it is only noise in the dimensions of sensory representations that constrains coding fidelity[13,14] (Extended Data Fig. 1g–m). Previous experiments have examined noise in cell pairs, but this approach incurs substantial measurement errors[13,19,20] and the results were conflicting[4,6,21–23]. To our knowledge, no previous study has recorded neural ensemble noise patterns, related these to sensory signals, and tested the idea that only specific noise patterns confine the information encoded by large neural populations[13,14].

A multi-beam two-photon microscope

To make such measurements, we built a laser-scanning two-photon microscope with a 4-mm^2 field of view for imaging across the span of the mouse primary visual cortex (V1). The microscope has 16 photodetectors and 16 corresponding beams, which originate from one laser and are focused 500 μm apart in the specimen in a 4 × 4 array (Fig. 1). Four beams are active at any instant, and switching to a different four beams takes about 50 ns; this enables scanning of a larger area per unit time

[1]James H. Clark Center for Biomedical Engineering & Sciences, Stanford University, Stanford, CA, USA. [2]CNC Program, Stanford University, Stanford, CA, USA. [3]Department of Applied Physics, Stanford University, Stanford, CA, USA. [4]Department of Biology, Stanford University, Stanford, CA, USA. [5]Allen Institute for Brain Science, Seattle, WA, USA. [6]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. ✉e-mail: rumyantsev.oleg@gmail.com; sganguli@stanford.edu; mschnitz@stanford.edu

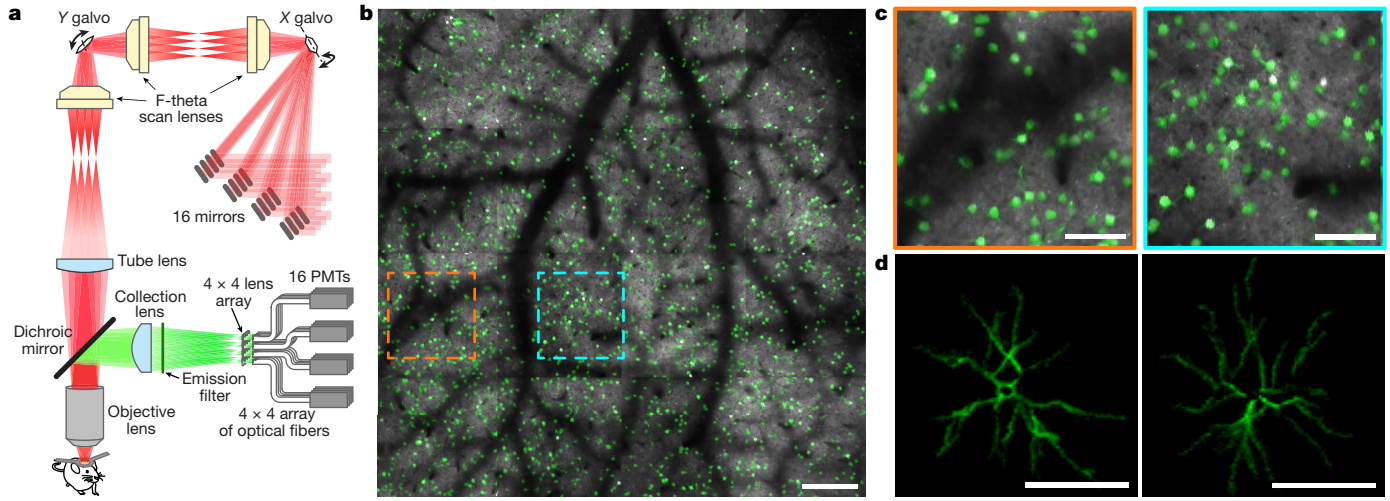


Fig. 1 | Two-photon Ca^{2+} imaging over a 4-mm² field of view. a, Schematic of the microscope. Sixteen laser beams converge on a pair of galvanometer mirrors (*X*- and *Y*-galvos). Sixteen photomultiplier tubes (PMTs) detect fluorescence. **b**, Two-photon image (greyscale, mean of 1,000 frames taken at 7.23 Hz) of GCaMP6f-expressing layer 2/3 pyramidal neurons in the visual cortex of an awake mouse. Overlaid are >2,000 neuronal somata (green) identified in the Ca^{2+} video. Boxed areas are magnified in **c. c**, Magnifications of the boxed areas in **b**. **d**, Example Ca^{2+}-activity sources (computationally identified), revealing dendrites. Images in **b**–**d** are representative of results from 10 mice. Scale bars: **b**, 250 µm; **c**, **d**, 100 µm.

than would be feasible with one beam and the same optics (Extended Data Figs. 2–4). Compared to 16 active beams, our approach yields fourfold greater fluorescence for any given time-averaged illumination power and delivers fourfold less heat to the brain for an equivalent rate of fluorescence emission (Supplementary Note). The active laser foci are ≥1 mm apart, so fluorescence scattering between the four active image tiles is <2%; scattering into inactive tiles can be corrected computationally using the 16 photocurrents (Extended Data Fig. 4). Our system images neocortical activity down to layer 5 with full-frame acquisition rates of 7.23–17.5 Hz (Supplementary Videos 1–3), whereas other two-photon microscopes with large fields of view attain similar imaging rates over smaller sub-fields[24–27] (Extended Data Fig. 2j, k).

Imaging studies across cortical area V1

We studied layer 2/3 pyramidal neurons, which project extensive connections from V1 to higher visual areas. In awake mice expressing the Ca^{2+}-indicator GCaMP6f in these neurons, we imaged around 1,000–2,000 cells concurrently as mice viewed, with one eye, a random sequence of moving gratings. Each grating was oriented at either +30° or −30° from vertical, lasted 2 s and spanned the central ~50 deg of the eye's visual field (Fig. 2a–c). There were 350 trials with each stimulus, but because locomotion modulates vision[28] we analysed only trials with locomotor speeds of less than 0.2 mm s^{-1} (217–332 trials per stimulus). From these recordings we extracted 8,029 neurons, mainly in V1 (1,031–2,191 cells in each of 5 mice; Extended Data Figs. 5, 6).

A total of 5,008 cells responded at least weakly to the stimuli, with activity rates and stimulus preferences consistent with those found in previous studies[28,29] (Extended Data Fig. 6a–d). These neurons likely had substantially overlapping inputs, because mouse V1 neurons respond to large portions of the visual field that are comparable in size to our stimuli[29]. Noise correlation coefficients in pairs of concurrently recorded cells were widely distributed, with positive mean values ($r = 0.06 \pm 0.01$; mean ± s.d.; 5 mice) as in most previous reports[6] (Fig. 2d–g, Extended Data Fig. 6e–i). Active cell pairs that on average responded similarly to the two stimuli had, on average, noise correlation coefficients about twice as large as those that responded dissimilarly (Fig. 2f, g).

To evaluate the significance of these correlations, we created trial-shuffled datasets in which the responses of each cell were permuted across different trials, thereby mimicking cells with statistically identical individual responses as in the real data but with uncorrelated noise fluctuations. Non-zero noise correlations in trial-shuffled datasets merely reflect the finite number of trials. Indeed, noise correlation coefficients were more narrowly distributed than in real data, although many deviated substantially from zero (Fig. 2d, g). This confirms the difficulty of measuring noise correlations given limited trials[13,19] and likely explains why previous studies of cell pairs yielded divergent results[4,6,21–23].

Evaluations of cortical coding fidelity

To study visual coding, we represented the dynamics using a population vector (one cell per dimension) and used the discriminability index, d', to assess the statistical confidence in distinguishing the stimuli on the basis of their evoked neural responses[30]. $(d')^2$ relates to the Fisher information that the cell ensembles convey about stimulus identity[8,13,30], which even for binary classifications (≤1 bit of Shannon entropy) can be infinite—that is, 100% confidence[31]. Theories of noise correlations and neural coding have largely examined pairwise discriminations, as error rates discriminating more than two stimuli are well approximated using d' values from all the pairwise comparisons[31].

To enable us to determine d' accurately despite having about 5- to 10-fold fewer trials than cells recorded per mouse, we created analyses to extract the primary, ensemble noise modes without measuring noise in cell pairs (Appendix). First, we performed a dimensional reduction by using partial least squares (PLS) analysis to identify and retain only five population vector dimensions in which the stimuli were highly distinguishable; retaining more than five dimensions only added noise and decreased the ability to distinguish the stimuli (Fig. 3a, b, Extended Data Figs. 5b, 7a–c). In this five-dimensional representation, the neural dynamics evoked by the two stimuli became distinguishable over the first ~0.5 s of stimulus presentation (Fig. 3b–d). Using an optimal linear decoder of the ensemble activity, d' values rose to a plateau within ~0.5 s of the stimulus onset; the optimal decoder then remained stable until stimulus offset (Extended Data Fig. 7d). In shuffled datasets the stimuli were even more distinguishable, as d' values attained greater values than in real datasets, indicating that correlated noise degrades stimulus representations in the real data.

We also evaluated decoders that ignore noise correlations. 'Diagonal decoders', which neglect off-diagonal elements of the noise covariance

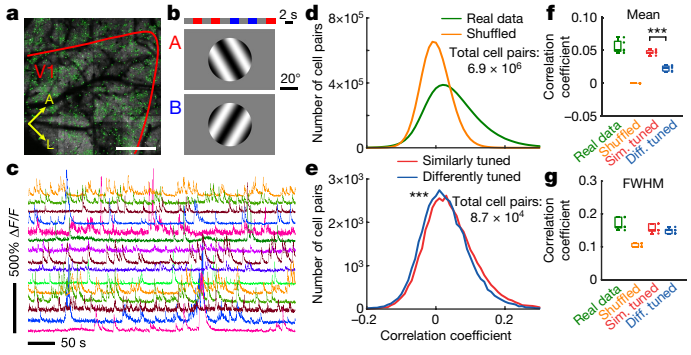


Fig. 2 | Noise correlations of cell pairs are difficult to estimate from hundreds of stimulus trials. a, Image of visual cortex, processed as in Fig. 1b. The coordinate system indicates anterior (A) and lateral (L) directions. The red line marks the area V1 boundary, found by retinotopic mapping. Scale bar, 500 µm. **b**, Top, in each trial, one of two randomly chosen stimuli (A or B) appeared for 2 s, followed by a uniform background for 2 s. Bottom, each stimulus was a drifting grating, oriented at either +30° or −30° from vertical. The analyses in **d**–**g** used 217–332 trials per stimulus in each of 5 mice. **c**, Example Ca^{2+} activity traces. *F*, fluorescence intensity. **d**, Histograms of noise correlation coefficients (Pearson's *r*) for concurrently imaged cell pairs (6,946,280 cell pairs; 5 mice), computed using the estimated spike count of each cell within [0.5 s, 2 s] of stimulus onset. *r* values are averages across both stimuli for real and trial-shuffled datasets. The latter histogram was Gaussian ($R^2 = 0.9982 \pm 0.0005$ (95% confidence interval)) with a variance around 50% of that of the real data, showing the difficulty of accurately determining pairwise noise correlations with hundreds of trials. Error bars estimated as counting errors are too small to see. **e**, Histograms of noise correlation coefficients differed significantly for cell pairs with similarly or differently tuned mean responses to the two stimuli, computed for the top 10% most active cells and by grouping cell pairs into those with positively or negatively correlated mean responses to the two stimuli. (***$P < 1.3 \times 10^{-6}$ for all 5 mice; two-tailed Kolmogorov–Smirnov test; 901 cells, 43,887 positively and 43,768 negatively correlated pairs). For exact *P* values for this and all subsequent figures, see Supplementary Information. **f**, **g**, Box plots of mean (**f**) and full width at half maximum (FWHM) (**g**) values of the colour-corresponding distributions in **d**, **e**. Circles indicate data points for 5 individual mice. Noise correlations in **f** were greater for cell pairs with similarly tuned responses (one-tailed Wilcoxon rank sum test, ***$P < 0.001$ for all 5 mice). Extended Data Fig. 6g–i shows results for all cell pairs. Boxes cover the middle 50% of values, horizontal lines denote medians, and whiskers span the full range of the data.

matrix[30], performed nearly as well as optimal linear decoders, although the decrement was statistically significant (Fig. 3d–h). Thus, although correlated neural noise degraded stimulus encoding, using the noise structure to improve decoding brought only modest benefit.

The stability of the optimal decoder across most of the stimulus duration suggested that, by integrating neural activity across the stimulus presentation, the brain might in principle average out noise in its sensory representations to improve discrimination. To test this, we examined the optimal linear decoder of the time-integrated neural responses over each trial, which indeed yielded greater d' values (Extended Data Fig. 7e). For comparison, we examined decoders of the cumulative set of neural responses that had occurred up to each moment in the stimulation trial (Fig. 3e–h). Cumulative decoders surpassed those using individual time-bins of neural activity, but not the simple decoder of time-integrated activity (Extended Data Fig. 7e). This suggests that there was little temporal structure in the sustained neural responses that might improve decoding beyond that attained using time-integrated activity, at least as reported by Ca^{2+} imaging.

We next examined how decoding varied with *n*, the number of cells analysed. In the absence of correlated noise, each additional cell used should linearly increase the Fisher information that is conveyed about the identity of the stimulus[5,12]. Trial-shuffled datasets confirmed this, as $(d')^2$ increased linearly with *n* (Fig. 3f, g). In real data, $(d')^2$ reached a plateau when *n* exceeded ~1,000 cells, for both instantaneous and cumulative decoders (Fig. 3f–i). This constitutes direct evidence of information saturation in large neural populations, without extrapolations from cell pairs.

Several control analyses bolstered these conclusions. First, we validated linear decoding as a way of assessing Fisher information. The noise covariance matrix was stimulus-independent, with similar matrix elements for both stimuli ($r = 0.81 \pm 0.16$; mean ± s.d.; 20 off-diagonal matrix elements for each of 5 mice). Thus, nonlinear decoders should have similar accuracy as the optimal linear decoder, which we confirmed by quantifying the additional information that an optimal quadratic decoder could extract from the data (Extended Data Fig. 7f–h). Second, we verified that there were a sufficient number of trials to estimate d' accurately. In every mouse the empirically determined values of d' approached a stable estimate with increasing numbers of trials and were stationary across the imaging session (Extended Data Fig. 7g, i, j). Third, we confirmed that alternative decoding methods using regularized regression yielded similar d' values and identical conclusions to those from PLS analysis (Extended Data Fig. 8a, b). Further, we used regularized regression to analyse publicly available neural activity datasets[32], which also showed that d' reached a plateau (Appendix). Fourth, we used simulations to verify that our decoders were robust to potential large sources of neural variability, such as common mode noise and gain modulation of visual responses (Extended Data Fig. 8c–h). Fifth, we mathematically derived the accuracy of d' determinations made via PLS analysis (Appendix). Altogether, numerous analyses and derivations upheld the information saturation that we found in ensembles of ~1,000 neurons or more.

The data also enabled us to test a framework for understanding cortical noise fluctuations based on a feedforward network[12,13]. In this framework, the encoded information, *I*, as a function of the ensemble size, *n*, obeys $I(n) = (I_0 n)/[1 + \varepsilon n]$, where the constant I_0 is the mean encoded information per cell in the shuffled data and the parameter ε characterizes the strength of information-limiting correlations[13]. Our data matched this prediction (Fig. 3f, g), verifying the existence of information-limiting correlations and establishing the effect size. The minimum set of cells needed to detect information saturation is approximately $2\varepsilon^{-1}$, which is around 800–1500 cells for the instantaneous decoders (Fig. 3h, i). This shows the importance of large recordings to adjudicate whether correlated noise limits coding accuracy, and likely explains why previous recordings of less than 350 cells did not observe information saturation[19,21].

Comparing neural coding to visual acuity

An additional benefit of recordings across V1 is to enable estimates of the attainable perceptual acuity given only the information encoded in the early visual cortex, which is important for fine discriminations of grating stimuli[33]. To approximate conditions more representative of the perceptual threshold, we examined another 5 mice that viewed the same grating stimuli as before but with ±6° orientations—closer to the discriminability limits.

As expected, these stimuli were harder to distinguish from their evoked neural activity stimuli (Extended Data Fig. 9). The asymptotic d' value (~2.5) for large *n* suggests that gratings presented at ±2.4° under otherwise identical viewing conditions would have the minimal, perceptibly distinct orientations ($d' \approx 1$). Behavioural studies of mouse visual spatial acuity under photopic illumination[15] yield similar predictions of ±2.3° (Methods). Direct measurements of mouse visual orientation sensitivity have been slender and used different stimuli from ours, but yielded similar values[34]. The fine agreement in these numbers is probably fortuitous, but the similar values estimated from cortical responses and behavioural studies[15,34] suggest that the information signalling limits of visual cortical coding likely have an important role in setting perceptual bounds.

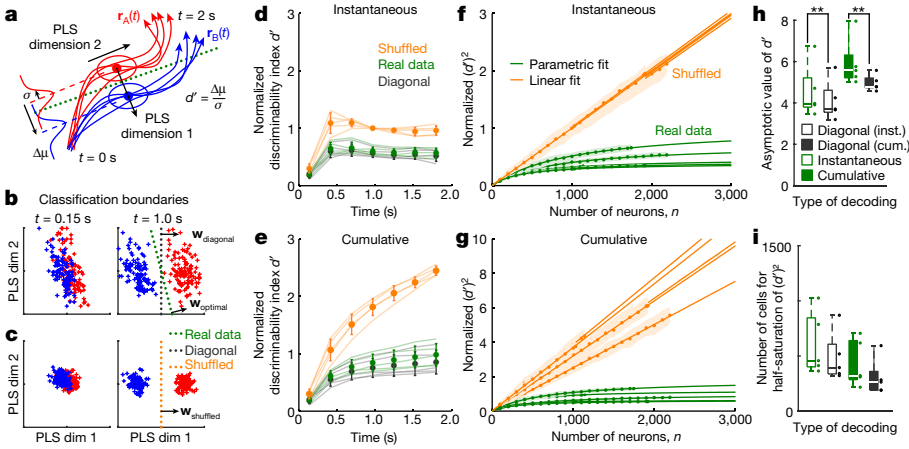


Fig. 3 | Correlated noise limits the information conveyed by cortical neural ensembles. a, Schematic of neural ensemble dynamics in a population vector representation of reduced dimensionality. Trajectories, $\mathbf{r}_A(t)$ and $\mathbf{r}_B(t)$, depict single-trial responses to different (red, blue) stimuli. At a fixed time after stimulus onset, the sets of responses to the two stimuli form two distributions of points (ellipses). At the bottom left are projections of these distributions onto a subspace, found by PLS analysis, in which responses to the two stimuli are most distinct. The green line indicates the optimal linear boundary for classifying stimuli in this subspace. The stimulus discriminability, d', equals the separation, Δμ, of the two distributions along the dimension orthogonal to this boundary, divided by the s.d., σ, of each distribution along this dimension. **b**, **c**, Neural ensemble responses, 0.15 s (left) and 1 s (right) after stimulus onset, in the two-dimensional space in which the sets of responses to the two stimuli are most distinct, for real (**b**) or trial-shuffled (**c**) datasets. The blue and red crosses denote individual trials (220 trials per stimulus); the green and orange lines mark the classification boundaries for real and trial-shuffled data, respectively; and the vertical black line in **b** is the classification boundary for diagonal discrimination, which ignores correlations in the responses of the cells. $\mathbf{w}_{optimal}$, $\mathbf{w}_{shuffled}$ and $\mathbf{w}_{diagonal}$ represent directions normal to the classification boundaries. $\mathbf{w}_{shuffled} = \mathbf{w}_{diagonal}$, as the corresponding classification boundaries are identical. **d**, Mean values of d' (coloured data points; $N = 5$ mice) plotted as a function of time after stimulus onset, for the classifiers in **b**, **c**. Error bars represent the standard deviation. Coloured lines show the d' values for individual mice, computed using the protocol of Extended Data Fig. 5b and averaged over 100 different randomly chosen subsets of 1,000 cells and randomly chosen assignments of trials to decoder training sets and test sets in each mouse. d' values are normalized by those obtained for trial-shuffled data (averaged across 0.83–1.11 s). **e**, Same as **d** but using cumulative decoding, which considers the full time-course of the activity of each cell up to time t. For each mouse, d' values in **e**–**h** have the same normalizations as in **d**. **f**, $(d')^2$ values during the interval 0.83–1.11 s from stimulation onset, plotted against the number of cells, n, used for analysis. Data points in **f**–**i** are averages over 100 different subsets of cells, and the shading in **f**, **g** indicates the standard deviation. For real data, $(d')^2$ values were well fit by the expression $(d')^2 = (d)^2_{shuffled}/(1+\varepsilon \times n)$, (green curves; $R^2 = 0.88 \pm 0.03$ (s.d); $\varepsilon = 0.0019 \pm 0.0007$; 5 mice), where ε is the fit parameter and $(d')^2_{shuffled}$ is the $(d')^2$ value for n cells in a linear regression to the shuffled data (orange lines). **g**, Same as **f**, but for $(d')^2$ values computed using cumulative decoding for the interval 0–1.11 s. **h**, **i**, Asymptotic d' values in the limit of many cells (**h**) and the number of cells at which $(d')^2$ attains half its asymptotic value (**i**) determined from curve fits as in **f**, **g** for instantaneous (open boxes) and cumulative (filled boxes) decoding. Optimal linear decoders (green) slightly but significantly outperformed diagonal decoders (black) (**$P < 10^{-11}$; one-tailed Wilcoxon rank sum test; $N = 100$ different assignments to decoder testing and training sets using all cells recorded in each mouse; dots are mean values from individual mice). Boxes cover the middle 50% of values, horizontal lines denote medians, and whiskers span the full range of the data. Analyses in **d**–**i** are based on 217–332 trials per stimulus in each of 5 mice and time bins of 0.275 s.

Origins of information-limiting noise

To identify why the information saturates, we analysed the neural noise structure by finding the principal eigenvectors of the neural noise covariance matrix and the mean amplitudes of visual signals encoded along each of these eigenvectors. This allowed us to decompose $(d')^2$ into a sum of signal-to-noise ratios, one for each eigenvector[13] (Methods). Although visual signal amplitudes increase linearly with ensemble size, n (Fig. 4a, b), certain noise eigenvalues might also increase with n, which could offset the greater signalling capacity of a larger ensemble and cause the information saturation.

We developed methods to determine the principal eigenvectors of the noise covariance matrix without needing accurate estimates of its matrix elements—a key distinction from previous analyses[13,19,20]. Contravening prevailing thinking, with our approach recordings of more cells enable accurate estimates of these eigenvectors and of d' using fewer trials (Extended Data Fig. 10). As n increased, mean ensemble responses to the two stimuli became increasingly distinct while staying aligned to the dimensions important for optimal decoding (Fig. 4b, c). In real but not in shuffled datasets the noise covariance matrix had 2–3 eigenvalues that also increased linearly with n (Fig. 4d, e). We examined how these particular noise eigenmodes related to the dimensions in which the neural ensembles represented visual signals.

In every mouse the visual signalling dimensions were nearly orthogonal to the largest noise mode, which therefore had almost no effect on coding fidelity even though it was around tenfold greater than any other noise mode (Fig. 4e–h; Extended Data Fig. 10). Instead, it was the third-largest noise mode that primarily aligned with the visual coding dimensions and thereby limited coding accuracy (Fig. 4f–h). These properties were sometimes seen, to a lesser extent, in the second-largest mode. The existence of noise eigenvectors that closely align to the dimensions used for visual representations and have eigenvalues that grow with n explains the information saturation for large n and why there was little performance decrement for decoders that did not account for correlated noise. Although these inferences rely on Ca^{2+} signals, not electrical recordings, this is unlikely to affect the conclusions, as variability in how spikes produce Ca^{2+} signals arises mainly from fluctuations in Ca^{2+} levels, photon emission and detection, which are statistically independent across cells and are not information-limiting.

A key question is how does information-limiting noise arise. Recent work examines this issue in a two-layer, feedforward network model with sensory inputs and intrinsic noise in both its input and its output layers[12]. As more cells are added to the output layer, the encoded information approaches a plateau, the value of which depends on the noise levels and synaptic weights[12] (Extended Data Fig. 1j–m). Our re-analysis of this model[12] revealed that the dimensionality of the space of receptive fields in the output layer equals the number of noise covariance matrix

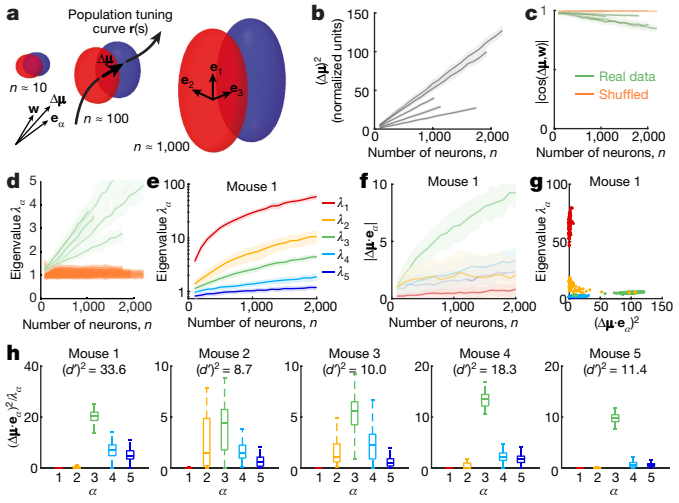


Fig. 4 | The largest noise mode is orthogonal to the dimensions encoding sensory information. a, Schematics of trial-to-trial variability in ensemble neural responses with increasing numbers of cells, n. Ellipsoids represent 1 s.d. fluctuations around mean ensemble responses, $\mathbf{r}(s)$, to two similar stimuli parameterized by a variable s (here the stimulus orientation). For large n, response variability along the tuning curve, $\mathbf{r}(s)$, increases proportionally to the separation, $\Delta\boldsymbol{\mu}$, between the two mean responses, leading to a saturation of d'. $\mathbf{w}$ is the normal to the optimal linear classification boundary between the two response sets. $\mathbf{e}_{1,2,3}$ are three eigenvectors, $\mathbf{e}_\alpha$, of the noise covariance matrix, averaged across both stimuli. The eigenvalues, λ_α, are the noise variances along each eigenvector. **b**, Mean values of $(\Delta\boldsymbol{\mu})^2$ plotted against n for 5 mice in units of the variance in the shuffled datasets, which have isotropic noise covariance matrices. Analyses in **b**–**h** used instantaneous decoding in the five-dimensional space found by PLS analysis and 100 different randomly chosen subsets of cells and assignments of trials to decoder training sets and test sets. Given these 100 sets of results, lines and shading in **b**–**f** denote mean ± s.d., **g** shows 100 individual results, and **h** has box plots. **c**, Cosine of the angle between $\Delta\boldsymbol{\mu}$ and $\mathbf{w}$ plotted against n, for 5 individual mice in real and trial-shuffled datasets. Because $\Delta\boldsymbol{\mu}$ is nearly collinear with $\mathbf{w}$, optimal linear decoding—which accounts for noise correlations—only modestly outperforms diagonal decoding, which does not (Fig. 3h). **d**, Eigenvalues, λ_α, for the eigenvectors best-aligned with $\Delta\boldsymbol{\mu}$ in 5 individual mice (green lines) increase linearly with n, revealing the origin of information-limiting correlations. For trial-shuffled data (25 orange lines, 5 eigenvalues for each of 5 mice), the noise variance along $\Delta\boldsymbol{\mu}$ is independent of n and is uniform for all eigenvectors of the noise covariance matrix. **e**, **f**, The geometric relationships between visual signals and noise indicate that the largest noise mode is not the one that is information-limiting. Each colour denotes a different eigenvector, $\boldsymbol{e}_\alpha$, of the noise covariance matrix in the reduced five-dimensional space, $\alpha\in\{1,2,3,4,5\}$. In each individual mouse (**e**) there were multiple eigenvalues, λ_α, of the noise covariance matrix that increased with n. Extended Data Fig. 10 shows results for all mice. Visual signals (**f**) also increased with n, as shown by decomposing $\Delta\boldsymbol{\mu}$ into components along the five eigenvectors, $\mathbf{e}_\alpha$. In each mouse the eigenvector with the largest eigenvalue, $\mathbf{e}_1$, was the least well aligned with the visual encoding direction, $\Delta\boldsymbol{\mu}$ (compare the red curves in **e**, **f**). **g**, A plot of noise values computed as in **e** against signal values computed as in **f**, using all recorded cells from mouse 1. The largest noise mode (red points) is an order of magnitude greater than the noise modes that limit neural ensemble signalling (green and yellow points), yet it is the least aligned with the signal direction. **h**, Signal-to-noise ratios for all five eigenvectors, computed using the values in **g**. $(d')^2$ equals the sum of these five signal-to-noise ratios. Boxes cover the middle 50% of values for the same 100 data subsets used in **e**–**g**, horizontal lines denote medians, and whiskers span 1.5 times the interquartile range. Analyses in **b**–**h** are based on 217–332 trials per stimulus in each of 5 mice.

eigenvectors for which the eigenvalues increase linearly with the number of output cells (Appendix). This shows that information-limiting correlations arise even in rudimentary networks, and reflect the co-propagation of signals and noise through the same synaptic connections.

Discussion

Our findings address longstanding questions about how the brain computes accurately despite neural noise[1], and help to resolve a 30-year-old puzzle by providing direct evidence that correlated noise limits cortical coding accuracy[2–4]. These results adjudicate against models in which noise correlations do not limit—or even improve—cortical ensemble coding[7,8]. Encoded visual signals in our recordings were orthogonal to the largest noise eigenmode, enhancing coding accuracy by restricting ~90% of noise fluctuations to dimensions that did not impede signalling. This strategy allows cortical codes to evade a majority of noise, although coding fidelity is ultimately bounded by the weaker correlated noise patterns that cannot be disambiguated from signal. (This strategy might not apply to sensory variables, such as full-field luminance, that animals rarely use for fine discriminations.) In support of these conclusions, mouse visual acuity measured using stimuli similar to ours[15,34] is around tenfold better than would be predicted from the total noise amplitude in the visual cortex, but fits with the amplitudes of the information-limiting noise modes.

Nevertheless, rigorous comparisons between the accuracies of sensory cortical coding and psychophysical discriminations will require concurrent evaluations in individual animals, using identical stimuli. Visual stimuli of greater size can increase d' values[32] by decreasing the mean level of shared inputs among responsive cells and thereby reducing ε, whereas stimuli of greater saliency should increase d' by increasing I_0. The recent history of sensory stimuli will also influence d' owing to sensory adaptation. Although specific values of d' will vary across stimulus types, information-limiting noise correlations and the saturation of information for large n arise generically from the propagation of signals and noise through common circuitry and place fundamental constraints on coding accuracy. Therefore, our experimental results likely reflect basic attributes of hierarchical networks and should generalize to diverse stimuli and sensory modalities.

The brain probably cannot learn its own correlated noise structure to decode sensory features optimally, as any particular sensory scene almost never repeats precisely. Nonetheless, decoders that ignore noise correlations can still be near optimal (Fig. 3d, e, h, Extended Data Fig. 9c), as predicted for large networks with information-limiting noise correlations[14]. Therefore, information-limiting cortical noise might help downstream circuits to readout diverse sensory features nearly optimally.

Future work should extend our experiments to different stimuli, sensory modalities and behavioural conditions. Together with our analyses tailored for large-scale recordings, microscopes that image multiple brain regions concurrently[24–26,35] will enable studies of noise correlations and information flow across successive cortical areas. Such measurements will help to address longstanding questions about the decoding strategies that the brain uses for perception, and the effect of attention on perceptual sensitivity and neural ensemble noise.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2130-2.

1. von Neumann, J. *The Computer and the Brain* 2nd edn (Yale Univ. Press, 1958).
2. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
3. Newsome, W. T., Britten, K. H. & Movshon, J. A. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).
4. Zohary, E., Shadlen, M. N. & Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140–143 (1994).

5. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
6. Cohen, M. R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–819 (2011).
7. Sompolinsky, H., Yoon, H., Kang, K. & Shamir, M. Population coding in neuronal systems with correlated noise. *Phys. Rev. E* **64**, 051904 (2001).
8. Abbott, L. F. & Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural Comput.* **11**, 91–101 (1999).
9. Shamir, M. & Sompolinsky, H. Implications of neuronal diversity on population coding. *Neural Comput.* **18**, 1951–1986 (2006).
10. Ecker, A. S., Berens, P., Tolias, A. S. & Bethge, M. The effect of noise correlations in populations of diversely tuned neurons. *J. Neurosci.* **31**, 14272–14283 (2011).
11. Oram, M. W., Földiák, P., Perrett, D. I. & Sengpiel, F. The 'Ideal Homunculus': decoding neural population signals. *Trends Neurosci.* **21**, 259–265 (1998).
12. Kanitscheider, I., Coen-Cagli, R. & Pouget, A. Origin of information-limiting noise correlations. *Proc. Natl Acad. Sci. USA* **112**, E6973–E6982 (2015).
13. Moreno-Bote, R. et al. Information-limiting correlations. *Nat. Neurosci.* **17**, 1410–1417 (2014).
14. Pitkow, X., Liu, S., Angelaki, D. E., DeAngelis, G. C. & Pouget, A. How can single sensory neurons predict behavior? *Neuron* **87**, 411–423 (2015).
15. Prusky, G. T., West, P. W. & Douglas, R. M. Behavioral assessment of visual acuity in mice and rats. *Vision Res.* **40**, 2201–2209 (2000).
16. Baylor, D. A., Lamb, T. D. & Yau, K. W. Responses of retinal rods to single photons. *J. Physiol. (Lond.)* **288**, 613–634 (1979).
17. Barlow, H. B. Retinal noise and absolute threshold. *J. Opt. Soc. Am.* **46**, 634–639 (1956).
18. Siebert, W. M. Some implications of the stochastic behavior of primary auditory neurons. *Kybernetik* **2**, 206–215 (1965).
19. Yatsenko, D. et al. Improved estimation and interpretation of correlations in neural circuits. *PLoS Comput. Biol.* **11**, e1004083 (2015).
20. Kanitscheider, I., Coen-Cagli, R., Kohn, A. & Pouget, A. Measuring Fisher information accurately in correlated neural populations. *PLoS Comput. Biol.* **11**, e1004218 (2015).
21. Ecker, A. S. et al. Decorrelated neuronal firing in cortical microcircuits. *Science* **327**, 584–587 (2010).
22. Reich, D. S., Mechler, F. & Victor, J. D. Independent and redundant information in nearby cortical neurons. *Science* **294**, 2566–2568 (2001).
23. Renart, A. et al. The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
24. Stirman, J. N., Smith, I. T., Kudenov, M. W. & Smith, S. L. Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. *Nat. Biotechnol.* **34**, 857–862 (2016).
25. Chen, J. L., Voigt, F. F., Javadzadeh, M., Krueppel, R. & Helmchen, F. Long-range population dynamics of anatomically defined neocortical networks. *eLife* **5**, e14679 (2016).
26. Sofroniew, N. J., Flickinger, D., King, J. & Svoboda, K. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *eLife* **5**, e14472 (2016).
27. Tsai, P. S. et al. Ultra-large field-of-view two-photon microscopy. *Opt. Express* **23**, 13833–13847 (2015).
28. Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
29. Bonin, V., Histed, M. H., Yurgenson, S. & Reid, R. C. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J. Neurosci.* **31**, 18506–18521 (2011).
30. Averbeck, B. B. & Lee, D. Effects of noise correlations on information encoding and decoding. *J. Neurophysiol.* **95**, 3633–3644 (2006).
31. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* 2nd edn, (John Wiley & Sons, 2006).
32. Stringer, C., Michaelos, M. & Pachitariu, M. High precision coding mouse visual cortex. Preprint at https://www.biorxiv.org/content/10.1101/679324v1 (2019).
33. Prusky, G. T. & Douglas, R. M. Characterization of mouse cortical spatial vision. *Vision Res.* **44**, 3411–3418 (2004).
34. Glickfeld, L. L., Histed, M. H. & Maunsell, J. H. Mouse primary visual cortex is used to detect both orientation and contrast changes. *J. Neurosci.* **33**, 19416–19422 (2013).
35. Lecoq, J. et al. Visualizing mammalian brain area interactions by dual-axis two-photon calcium imaging. *Nat. Neurosci.* **17**, 1825–1829 (2014).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Microscope design

We used a systems-engineering approach to design the two-photon microscope. To simulate its optical performance and assess component suitability, we used optical design software (ZEMAX) to simulate both ray and wave propagation through the optical pathway. To validate the multiplexing strategy (Extended Data Fig. 2b–d) and the computational un-mixing of crosstalk between image tiles (Extended Data Figs. 3c–e, 4a–c), we simulated fluorescence scattering in brain tissue using the non-sequential mode of ZEMAX. We created an optomechanical design of the microscope using CREO Parametric 3.0 CAD mechanical design software.

Laser source and control of illumination

We used an ultrashort-pulsed Ti:sapphire laser (MaiTai eHP DeepSee; Spectra Physics) with an 80 MHz repetition rate. We tuned the emission wavelength to 910 nm and used the laser's built-in pre-chirping module to attain pulses of 130 ± 20 fs duration (FWHM) at the sample plane. For general purpose routeing of the laser light to and within the microscope we used broadband dielectric mirrors (BB1-E03, Thorlabs). A computer-driven rotating half-wave ($\lambda/2$) plate (WP, AHWP05M-980; Thorlabs) controlled the laser beam polarization and hence the power transmitted through a polarizing beam splitter (PBS) (PBS102, Thorlabs) and into the microscope's illumination pathway (Extended Data Fig. 2d). To block all laser illumination to the microscope during the turnaround portion of the fast galvanometer mirror's scanning cycle, we used a custom laser chopper wheel (90:10 duty ratio), positioned after the PBS and synchronized in frequency and phase with the fast-axis galvometer cycle.

Multiplexing of the 16 illumination pathways

Owing to the powerful ultrafast lasers that are now commercially available, past users of two-photon microscopy have often had more than enough illumination power at their disposal but remained limited with regards to the imaging speeds and the fields of view that were attainable with a single beam and existing scanning hardware. We therefore developed a multi-beam, two-photon microscope that puts the (previously) excess laser power to good use, by using multiple beam paths that enable the coverage of larger fields of view at faster image-frame acquisition rates. The Supplementary Note, Extended Data Fig. 2j, k and Supplementary Fig. 1 quantitatively compare our imaging system to other recent approaches to large-scale two-photon microscopy.

To steer laser illumination into four different sets of four beam paths, we used three pairs of electro-optic modulators (EOM) (LM0202 3 × 3 mm 5W, LIV20 pulse amplifier; QIOptic) and PBS cubes (PBS102, Thorlabs) (Extended Data Fig. 2d). We drove each EOM with a high-voltage (310 V amplitude) square wave oscillation, with the period matched to that of the microscope's pixel clock. When imaging using the 4 × 4 set of beams, the square waves driving the second and third EOMs were both phase-shifted by ¼ period relative to the square wave driving the first EOM (Extended Data Fig. 2c). By toggling the beam exiting each EOM between the two linear orthogonal polarization states (the transition time between polarizations was around 50 ns), these three square-wave signals steered the beam from the laser successively into each of the four sets of four beam paths (that is, 16 total), with each set of four illuminated for ¼ of each pixel clock cycle (Extended Data Fig. 2b–d). Within each set, three beamsplitters (10RQ00UB.2 and 10RQ00UB.4, respectively, for S and P polarizations; Newport) divided the beam power equally between four different paths corresponding to four non-neighbouring image tiles in the 4 × 4 array (Extended Data Fig. 2b). Because the efficiency of two-photon fluorescence excitation increases as the square of the peak illumination intensity, this temporal multiplexing scheme enabled fourfold greater fluorescence excitation compared with an otherwise identical, 4 × 4 set of beams that were not multiplexed in time.

Illumination pathways

Each of the 16 beam pathways contained a pair of kinematically mounted mirrors, a 1:2 telescope implemented using a pair of lenses (AC254-500-B-ML, LA1464-B; Thorlabs), and a gimbal-mounted mirror (GMB1/M; Thorlabs). The 16 beam paths converged on a 6-mm-diameter, Ag-coated mirror mounted on a galvanometer scanner (6215HSM40B scanner, 671215HHJ-1HP driver; Cambridge Technologies). This galvanometer served as our slow-axis scanner.

To image the 16 beams striking the first scanning mirror onto an identical galvanometer scanning mirror serving as the fast-axis scanner, we used a pair of telecentric f-theta lenses designed to induce minimal group velocity dispersion with ultrashort-pulsed illumination (S4LFT0089/094; Sill Optics) in a 1:1 telescope configuration (Fig. 1a). A third, identical f-theta lens and a tube lens (f = 300 mm, G322-372-525, Linos) imaged all 16 beams striking the second scanning mirror onto the back aperture of the microscope objective. The objective focused the 16 beams to a square array of 4 × 4 foci, which together scanned a 2 mm × 2 mm specimen area at image frame acquisition rates up to around 8 Hz.

Alternatively, to enable image frame acquisition rates up to 20 Hz over a 2 mm × 2 mm specimen area, we used a resonant galvanometer scanner (6SC08KA040-02Y, Cambridge Technology, 8 kHz, 7 mm clear aperture) as the fast-axis scanner. The 8 kHz rate of resonant line-scanning allowed us to use a data acquisition scheme based on line multiplexing instead of pixel multiplexing. In this mode we used EOM3 to direct the laser illumination into one of its two optical output paths (Extended Data Fig. 1d, phase I and phase IV). During the resonant scanner turnaround times, we used EOM1 to redirect the laser illumination towards EOM2, the output pathway of which was blocked. During both the forward and backward motion of the resonant scanner a set of 4 laser beams scanned across a total of 8 image tiles—that is, 2 tiles per beam. By using a different set of 4 beams during the forward and backward scanning motions, we sampled one image line in all 16 image tiles during each cycle of the resonant scanner while using only 8 of the 16 beam paths. As with the pixel-multiplexing approach, only 4 beams were active at any instant in time.

For the microscope objective lens, we used either an air objective lens (Leica, 5.0 × Planapo 0.5 NA; 19 mm working distance; anti-reflection (AR) coated for 400–1,000 nm light; transmission >90% at 520 nm, >75% at 910 nm) or a water-immersion lens optimized for large-scale two-photon imaging[26] (1.0 numerical aperture (NA) fluorescence collection, objective (Jenoptik; 2.5 mm working distance). The illumination beams underfilled the back aperture of the microscope objective lens, leading to an optical resolution of approximately 1.2 µm and 8 µm in the lateral and axial dimensions, respectively, as determined from the FWHM values of the microscope's optical point-spread function.

Fluorescence collection pathway

Fluorescence emanating from the sample returned through the objective lens, reflected from a dichroic mirror (FF735-Di02-58x82, Semrock) and passed through a collection lens (AC508-180-A, Thorlabs) and a fluorescence emission filter (FF02-525/40-25, Semrock).

The objective and the collection lens project a magnified image of the fluorescence foci in the sample. To optimize the efficiency of fluorescence detection, we designed a custom 4 × 4 lens array (4.5 mm pitch, plano-convex lenslets, custom injection-moulded in poly(methyl methacrylate) (AR-coated: reflectivity <0.5%, 450–650 nm) that efficiently coupled fluorescence emissions into a 4 × 4 array of 3-mm diameter (0.5 NA) plastic optical fibres (FF-CK-120, AR-coated, FibreFin) (Fig. 1a).

To capture the maximum amount of fluorescence near the edges of the large field of view, the outer lenslets in the array were slightly larger than the others, extending outward from the perimeter of the array. Because even the outer lenslets had a maximum numerical aperture (0.19 NA) much lower than that of the plastic fibres (0.5 NA), this lenslet design yielded a theoretical efficiency of >97% for coupling

fluorescence into the array of 16 optical fibres. The fibre array delivered the fluorescence to a set of 16 GaAsP photomultiplier tubes (PMT) (H10770PA-40, Hamamatsu). Each 400-mm-long fibre had a specified transmission efficiency of >98%, yielding an overall design efficiency of >95% for conveying fluorescence into the photomultiplier tubes.

Optomechanics

We custom-fabricated the majority of the structural components of the microscope at our laboratory's machine shop using high-strength 7075-aluminium alloy and computer numeric control machining. We used three-dimensional (3D) printing to create a cover for the microscope objective lens and a mount for the dichroic mirror. The optomechanical components were generally catalogue parts from standard vendors, mainly Thorlabs, Newport and Linos.

Data acquisition electronics

Owing to the unique multiplexing scheme of our microscope, data acquisition differs from that in a conventional two-photon microscope (Extended Data Fig. 3a). A major concern was to ensure that the signals from each of the four phases per pixel clock cycle were correctly assigned. This necessitated sampling the 16 PMTs sufficiently rapidly to ensure that the signals corresponding to different pixels and phases were not conflated. Hence, we chose a sampling rate of 50 MHz for each PMT. Because the duration of each of the four multiplexing phases was 400 ns, this sampling rate yielded 20 samples per pixel per multiplexing phase (Extended Data Fig. 3b).

To implement data sampling at this rate, we first converted the photocurrents from the 16 PMTs into voltage signals using a set of four transimpedance amplifiers, each with four input channels (SR445A, Stanford Research Systems). We then sampled the resulting voltage signals using a 16-channel, 50 MS/s analogue-to-digital converter (ADC; 14-bit-samples encoded in 2 bytes) module (NI 5751, National Instruments). The ADC connected to the NI FlexRIO field programmable gate array (FPGA) Module for PXI Express, which was controlled by a host computer (Win 64-bit, 2 Intel E5-2630 processors, 32 GB RAM, Lenovo) through a PCIe-PXIe link (NI PXIe-7962R, NI PXIe-1082 chassis, PXIe-PCIe8381 link, National Instruments) (Extended Data Fig. 3a). For each multiplexing phase, the FPGA module summed the digitally sampled values of the photocurrents into pixel intensities. All subsequent data manipulations involved only the pixel intensities, yielding a total data throughput rate of 60 MB s^{-1} or 105 MB s^{-1}, for image frame acquisition at 7.23 Hz or 17.5 Hz, respectively, as opposed to the 1.6 GB s^{-1} raw data stream. To eliminate any residual crosstalk between pixels resulting from the approximately 50-ns switching time of the EOMs, the software interface gave the user the flexibility to discard the first few samples of each pixel.

Instrument control

When imaging in pixel-multiplexing mode, we used ScanImage[36] software (version 3.8) to generate the analogue signals driving the galvanometer scanners and the digital line-clock and frame-clock signals (Extended Data Fig. 3a). Using the clock signals from ScanImage, the FPGA module generated signals to drive the EOMs. We created custom LabVIEW (National Instruments, version 2012 SP1, 32 bit) code to initiate the imaging sessions and control the data acquisition parameters. When imaging in line-multiplexing mode, we controlled the instrumentation fully using custom software written in LabVIEW. We synchronized laser line-scanning and data acquisition by using the clock of the resonant scanner as a master clock.

In both imaging modes, the FPGA module continually transmitted to the host computer the imaging data in packets of pixels, combined into image lines, via a high-speed direct memory access first-in first-out (DMA FIFO) data link. The host computer constructed image tiles from the image line data, accounting for the number of photodetection channels and temporal multiplexing phases. The computer then streamed the image data onto its hard drive (Extended Data Fig. 3a).

Mice

The Stanford Administrative Panel on Laboratory Animal Care (APLAC) approved all procedures involving animals, and we complied with all of the panel's ethical regulations. We analysed data acquired from 6 male and 4 female Ai93 triple transgenic *GCaMP6f-tTA-dCre* mice from the Allen Institute (*Rasgrf2-2A-dCre*/*CaMK2a-tTA*/Ai93), which expressed the Ca^{2+}-indicator GCaMP6f in layer 2/3 pyramidal cells[37]. Mice resided on a 12-h reverse light cycle in standard plastic disposable cages. Experiments occurred during the dark cycle. All animals in the experiment belonged to the same group, so blinding and random assignments were neither needed nor feasible.

For illustrative purposes only, we imaged a single *tetO-GCaMP6s*/*CaMK2a-tTA* mouse[38], which expressed the Ca^{2+}-indicator GCaMP6 s in a subset of neocortical pyramidal neurons (Supplementary Video 3).

Surgical procedures

At the start of surgery we gave adult mice (12–17 weeks old) buprenorphine (0.1 mg kg^{-1}) and carprofen (5 mg kg^{-1}) and anaesthetized them with 1–2% isoflurane in O_2. We implanted a glass window within a 5-mm-diameter craniotomy positioned over the right visual cortical area V1 and surrounding cortical tissue. The window was a round #1 cover glass (5 mm diameter, 0.15 ± 0.02 mm thickness, Warner Instruments) that we attached to a circular steel annulus (1 mm thick, 4.9 mm outer diameter, 4.4 mm inner diameter) using adhesive cured with ultraviolet light (NOA81, Norland Products). To fill the gap between skull and glass window we applied 1.5% agarose. We secured the window on the cranium with dental acrylic. We also implanted an aluminium metal bar atop the cranium, allowing the mice to be head-restrained during in vivo brain imaging. For two days after surgery, we gave the mice buprenorphine (0.1 mg kg^{-1}) and carprofen (5 mg kg^{-1}) to reduce post-surgical discomfort. Mice recovered for at least one month before any imaging experiments began.

Visual stimulation

Mice viewed visual stimuli on a gamma-corrected computer monitor (Lenovo LT2323p; 58.4 cm diagonal extent) that was 10 cm away from the left eye and spanned around 142° of this eye's accessible, angular field of view. We generated visual stimuli using the psychophysics toolbox libraries of the MATLAB (Mathworks; version 2017b) programming environment. Stimuli were sinusoidal drifting gratings (spatial frequency, 0.04 cycles per degree; stimulus angular diameter, 50 deg; drifting rate, 50 deg s^{-1}, centred on the left eye's visual field; stimulation duration, 2 s; amplitude modulation depth, 100%; screen background intensity, 50%; Fig. 2b). During each experiment, we presented the gratings at two different angles, ±30° or ±6° to the vertical, in a random sequence. Between successive stimuli, the monitor was uniformly illuminated at the background intensity for a 2-s inter-trial interval. To prevent light from the visual stimuli from entering the fluorescence collection pathway of the microscope, the stimuli used only the blue component of the RGB colour model, which was blocked by the fluorescence emission filter. We also placed a colour filter (Rosco, 382 Congo Blue) on the monitor screen. The mean luminance from the stimulus at the mouse eye was approximately 5×10^{10} photons mm^{-2} s^{-1}, which is more than two orders of magnitude higher than the transition threshold to photopic vision in mice[15].

Imaging sessions

To reduce the stress of head restraint, we head-fixed mice on a 100-mm-diameter Styrofoam ball that could rotate in two angular dimensions. We tracked the movement of the ball with an optical computer mouse. Because running or walking is known to alter visual processing in rodents[28], we ensured that all visual stimulation trials used for analysis were those when the mice were passively viewing the video monitor, without locomotion, by excluding all trials during which the mice had

an ambulatory speed of greater than 0.2 mm s^{-1}. We imaged the Ca^{2+} activity of neocortical layer 2/3 pyramidal neurons, 150–250 μm below the cortical surface. The pixel clock cycle duration was 1.6 μs, hence the pixel dwell time in each of the four multiplexing phases was 400 ns. Owing to the ~50-ns switching time of the EOMs, we discarded four samples at the start of each phase, removing any crosstalk between phases. Across the full duration of each imaging session, fluorescence intensities decreased by ~9% owing to photobleaching. The total laser illumination power was 280–320 mW, divided evenly amongst the 4 beams that were active at any instant in time. Hence, each of the 16 image tiles (each 500 μm × 500 μm in size) received a time-averaged power of 17.5–20 mW, for a time-averaged illumination intensity of 70–80 mW mm^{-2}. Previous Ca^{2+} imaging studies of layer 2/3 neocortical neurons with conventional two-photon microscopy[39–42] have used mean illumination intensities of 89–1,800 mW mm^{-1}.

For studies in which the visual stimulation comprised moving gratings oriented at ±30°, we used the air objective lens and the pixel-multiplexing approach to image acquisition. We acquired images with 1,024 × 1,024 pixels at a 7.23 Hz frame rate across the 2 mm × 2 mm field of view using the air objective lens. The total imaging duration per session was 2,800 s (about 20,000 two-photon image frames), resulting in 700 visual stimulation trials, 350 for each of the two visual stimuli.

For studies in which the moving grating stimuli were oriented at ±6° to vertical, we used the water-immersion objective lens and line-multiplexing to acquire images with 1,728 × 1,728 pixels at 17.5 Hz across the 2 mm × 2 mm field of view, which we averaged and downsampled on the FPGA module to 864 × 864 pixels (Extended Data Fig. 9a–c, e, Supplementary Videos 2, 3). The total imaging duration per session was around 1,500 s.

Image reconstruction

We wrote custom MATLAB (Mathworks; version 2017b) scripts to manipulate the experimental datasets directly from the computer hard drive, without loading all the data into the computer's random-access memory.

The first step of image reconstruction accounted for the differences in the gain values of the 16 PMTs. We determined the gain values by imaging a static fluorescence sample and then analysing the statistics of the photon shot-noise limited fluorescence detection. Specifically, we performed a linear regression between the mean signal from each PMT and its variance. In the shot-noise limited regime, the slope of this relationship equals the combined gain of the PMT, pre-amplifier and ADC. Knowledge of the pre-amplifier and ADC gain values enabled us to determine the PMT gain. Given these empirically determined PMT gain values, the first step of image reconstruction was normalization of the fluorescence signals from each PMT channel by its gain.

The second step in image reconstruction was un-mixing of the crosstalk between the different PMT channels (Extended Data Fig. 3). In principle, when using laser-scanning microscopes with multiple illumination beams, one can apply to the set of PMT signal traces an un-mixing matrix that represents the inverse of a pre-calibrated, empirically determined matrix of crosstalk coefficients between the different photodetection channels[43]. However, this approach assumes that the biological sample is uniform and hence that a single un-mixing matrix will apply equally well across the entire specimen. In practice, brain tissue is not optically uniform, and it is challenging to precisely determine the crosstalk matrix in image sub-regions with low fluorescence levels, such as in blood vessels. Furthermore, two-photon neural Ca^{2+} imaging routinely involves modest signal-to-noise ratios and consequently the application of the inverse crosstalk matrix introduces additional error, analogous to the errors introduced by deconvolution methods when applied to weak signals.

For these reasons, we used a more straightforward, conservative and computationally efficient method of image reconstruction. Because crosstalk was only present in our microscope near the boundaries between image tiles, for each of the four sub-frames per image we computationally reassigned the signals from the boundary regions between tiles to the nearest neighbour source tile from which the crosstalk signals originated according to Extended Data Fig. 3c. We empirically determined that boundary regions 50 pixels wide contained ~75% of the scattered fluorescence photons from each laser focus. Hence, computational re-assignment of the photons from these boundary regions enabled conservative estimates of cells' fluorescence signals, near continuous stitching of the images (Extended Data Fig. 3d, e), and high-fidelity extraction of neural activity (Extended Data Fig. 4).

Beyond each 50-pixel-wide boundary region, there were generally residual scattered fluorescence photons. Thus, for purposes of visual display only (Fig. 1b, c; Supplementary Video 1), we removed boundary artefacts left over after computational re-assignment (Extended Data Fig. 3c) by parameterizing the boundary with a smoothly decaying function:

$$\mathrm{sigmoid}(x) = \frac{1}{1 + e^{\left(\frac{x-d}{a}\right)}}$$

where x is the distance from the tile edge, $d = 70$ pixels is the width of the boundary region, and $a = 25$ pixels characterizes the smoothness of the boundary decay.

Image pre-processing

After image reconstruction, each dataset comprised 16 videos, each 256 pixels × 256 pixels × 21,000 frames for a typical experiment, corresponding to the 16 tiles of each image frame. To correct for lateral displacements of the brain during image acquisition, we applied a rigid image registration algorithm (Turboreg[44]; http://bigwww.epfl.ch/thevenaz/turboreg/) to each of the individual video tiles. We chose this approach because the application of a single, rigid image registration algorithm over the entire 2 mm × 2 mm field of view did not account for variations in tissue motion between the different image tiles. After image registration, for display purposes only we merged the 16 motion-corrected video tiles into images or videos of the entire field of view (Supplementary Videos 1–3). We performed all further analysis on individual tiles.

For display purposes only (Supplementary Video 2, 3), to minimize stitching artefacts during video playback we applied to each image frame a linear-blending stitching algorithm[45,46]. We then computationally corrected the movie for lateral displacements of the brain by using a piecewise rigid image registration algorithm[47]. To highlight the details for viewers using a typical computer monitor, we saved the processed video using a contrast (γ) value of 0.75.

Computational extraction of neural activity traces

To identify individual neurons in the Ca^{2+} imaging data, we separately analysed the 16 individual video tiles in each movie and applied an established algorithm for cell sorting based on the successive application of principal component and independent component analyses[35,48] (Mosaic software, version 0.99.17; Inscopix). We visually screened the resulting set of putative cells and removed any that were clearly not neurons (about 50% of candidate cells were removed). For the resulting set of cells, we created a corresponding set of truncated spatial filters that were localized to the cell bodies by setting to zero all pixels in the filter with values <5% of the peak amplitude of the filter. After thresholding, we removed any connected components containing less than 30 pixels. To obtain traces of neural Ca^{2+} activity, we applied the truncated spatial filters to the $(F(t) - F_0)/F_0$ movies (Extended Data Fig. 5), where $F(t)$ denotes the time-dependent fluorescence intensity of each pixel and F_0 is its mean intensity value, time-averaged over the entire movie.

For each cell, we used fast non-negative deconvolution to estimate the number of spikes fired in each time bin[49]. We then temporally down-sampled twofold the resulting traces by summing the estimated numbers

of spikes in pairs of adjacent time bins, yielding time bins of 0.276 ms. We performed all subsequent analysis on the down-sampled traces.

Moreover, previous work has shown that the activity of mouse visual cortical neurons differs substantially between behavioural states of passive viewing and viewing during active locomotion[28,35]. To ensure that all visual stimulation trials used for analysis were those when the mice were passively viewing the video monitor, we excluded from analysis all trials during which the mice were running or walking (at speeds greater than 0.2 mm s^{-1}). The resulting set of trials retained for data analysis in each mouse was 217–332 for each stimulus condition, except for the analysis of Extended Data Fig. 9a–c, e, which involved 122–167 trials per stimulus condition.

Trial-shuffled datasets

To create trial-shuffled datasets, we randomly permuted the activity traces of each cell across the full set of trials in which the same stimulus was presented, using a different random permutation for each individual cell. Thus, the trial-shuffled datasets preserved the statistical distributions of each cell's responses to the two stimuli, but any temporally correlated fluctuations in different cells' stimulus-evoked responses were scrambled. For analyses of trial-shuffled data, we averaged results over 100 different randomly chosen subsets of cells and/or stimulation trials, each of which was trial-shuffled with its own distinct permutations; exceptions to this statement are the analyses of Extended Data Figs. 8c–h, 10a, b, for which we averaged results over 30 such calculations instead of 100.

Noise correlations in the visual stimulus-evoked responses of pairs of cells

To compute correlation coefficients for the noise in the visual responses of a pair of neurons, we first integrated the estimated spike count of each cell between [0.5 s, 2 s] from the start of visual stimulation. After separating the trials for each of the two visual stimuli, we subtracted from each trace the mean stimulus-evoked response of the cell and then calculated the Pearson correlation coefficient, r, for the resulting set of responses from the two cells. We then averaged these noise correlation coefficients over the two stimulus conditions. Figure 2d, e and Extended Data Fig. 6e, g show statistical distributions of the resulting mean correlation coefficients across many cell pairs.

We compared the statistical distributions of mean correlation coefficients for two different sets of cell pairs, those with positive and those with negative covariance of their mean stimulus responses (that is, cell pairs with similar or dissimilar visual tuning) (Extended Data Fig. 6e, g). To visually highlight the differences between the two distributions (Fig. 2e), we also analysed only the most responsive cells, defined as those cells with the top 10% values of $\sqrt{\langle r_A \rangle^2 + \langle r_B \rangle^2}$, where r_A and r_B are the mean responses to the two stimuli.

Dimensionality reduction and computation of d' for neural responses to visual stimuli

To estimate how much information the neural activity conveyed about the stimulus identity, we used the metric d', which characterizes how readily the distributions of the neural responses to the two different sensory stimuli can be distinguished[50]. The quantity $(d')^2$ is the discrete analogue of Fisher information[30]. We evaluated three different approaches to computing d' values for the discrimination of the two different visual stimuli (Fig. 3).

In the first approach, which we termed 'instantaneous decoding' (Fig. 3d, f, Extended Data Figs. 7a, 9a), we chose for analysis a specific time bin relative to the onset of visual stimulation. To examine the time-dependence of d', we used the instantaneous decoding approach and varied the selected time bin from $t = 0$ s to $t = 2$ s relative to the start of the trial. The number of dimensions of the neural ensemble activity evoked in response to the visual stimulus was N_o, the number of recorded neurons ($N_o \approx 1{,}500$). Said differently, the set of estimated spike traces provided an N_o-dimensional population vector response to each stimulus presentation.

In the second approach, termed 'cumulative decoding' (Fig. 3e, g, Extended Data Figs. 7b, 9b), we concatenated the responses of each neuron over time, from the start of the trial up to a chosen time, t. In this case, the dimensionality of the population activity vector was $N_o \times N_t$, where N_t is the number of time bins spanning the interval [0 s, t].

In the third approach, termed 'integrated decoding' (Extended Data Fig. 7c), we examined the neural ensemble responses integrated over the interval from [0 s, 2 s] relative to stimulation onset. In the plots of d' against time as computed by instantaneous decoding, the interval [0.5 s, 2 s] is when the d' values have already reached an approximate plateau (Extended Data Fig. 7e). With integrated decoding, the dimensionality of the population vector response was N_o, the number of recorded neurons, as in the instantaneous decoding approach.

In each of the three decoding approaches, we arranged the traces of estimated spike counts into three-dimensional data structures (number of neurons × number of time bins × number of trials), for each of the two visual stimuli (Extended Data Fig. 5b).

A challenge was that calculation of d' in an N_o-dimensional population vector space would have involved estimation of a $N_o \times N_o$ noise covariance matrix with over a million matrix elements. Direct estimation of the covariance matrix would have been unreliable, because the typical number of cells per dataset, $N_o \approx 1{,}500$, was much larger than the typical number of trials $P \approx 600$. This issue was even more severe in the case of cumulative decoding, for which the population activity vector had $N_o \times N_t$ dimensions. However, we found mathematically that by reducing the dimensionality of the space used to represent the ensemble neural responses, one can reliably estimate eigenvalues for the largest eigenvectors of the noise covariance matrix, which govern how well the two visual stimuli can be discriminated based on the neural responses (Appendix).

Our approach to dimensional reduction relied on a PLS discriminant analysis[51]. The PLS analysis enabled us to find the dimensions of the population vector space that were most informative about which visual stimulus was shown. To determine how many dimensions were important for discriminating the two stimuli, we constructed an orthonormal projection operator, which projected the N_o-dimensional (or $N_o \times N_t$ dimensional) ensemble neural responses onto a truncated set of the N_R dimensions identified by the PLS analysis as being the most informative about the identity of the visual stimulus.

In the reduced space with N_R dimensions, we calculated the $(d')^2$ value of the optimal linear discrimination strategy as:

$$(d'_{opt})^2 = \Delta\boldsymbol{\mu}^T \Sigma^{-1} \Delta\boldsymbol{\mu} = \Delta\boldsymbol{\mu}^T \mathbf{w}_{opt}$$

where $\Sigma = \frac{1}{2}(\Sigma_A + \Sigma_B)$ the noise covariance matrix averaged across two stimulation conditions, $\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_A - \boldsymbol{\mu}_B$ is the vector difference between the mean ensemble neural responses to the two stimuli and $\mathbf{w}_{opt} = \Sigma^{-1}\Delta\boldsymbol{\mu}$, which is normal to the optimal linear discrimination hyperplane in the response space[30].

To determine the optimal value of N_R for these computations of d', we split the data into three sets, each comprising a third of all trials. We used the first set to identify the PLS dimensions, the second 'training' set to find the optimal discrimination boundary defined by $\mathbf{w}_{opt}$, and the third 'test' set to estimate the discrimination performance d'. We then varied N_R and plotted the resulting d' values for both the training and test datasets (Extended Data Fig. 7a–c).

For all three decoding strategies, we chose $N_R = 5$ for all subsequent determinations of d', because the addition of further dimensions led to overfitting, as shown by the increase in discrimination performance using the training set and the decline in performance (that is, poorer generalization to previously unseen data) using the test set (Extended Data Fig. 7a–c).

After picking $N_R = 5$, for all further computations of d' we first chose a subset of neurons and divided the set of stimulation trials into two groups of equal size. We used the first group of trials to conduct the PLS analysis and the second group to determine d' and the eigenvalue spectrum of the noise covariance matrix (Extended Data Fig. 5b). To make plots of d' (Fig. 3d–g), we averaged d' values across 100 different randomly chosen subsets of cells, which we analysed independently for every time bin. For each subset of cells and every time bin, we randomly split the set of visual stimulation trials into two halves, one half for determination of the five-dimensional sub-space and decoder training, and the other half for decoder testing. In Fig. 3d, e, we kept constant the number of cells per subset. In Fig. 3f, g, we varied the number of cells per subset. For instantaneous and cumulative decoders in the experiment with visual gratings oriented at ± 30°, we used [0.83 s, 1.11 s] and [0 s, 1.11 s] time intervals, respectively (Fig. 3f–i). For the experiment with gratings oriented at ± 6°, the time intervals used for instantaneous and cumulative decoding were respectively [0.70 s, 0.94 s] and [0 s, 0.94 s] (Extended Data Fig. 9a–c).

To determine the asymptotic value of d' in the limit of many neurons, and the number of cells, $n_{1/2}$, at which $(d')^2$ attains half of its asymptotic value (Fig. 3h, i), we performed a two-parameter fit to the growth of d' with increasing numbers of neurons, n: $(d')^2 = (sn) / (1 + \varepsilon n)$. We determined the asymptotic value of d' as $(s/\varepsilon)^{1/2}$ and $n_{1/2}$ as ε^{-1}.

To verify that linear decoding is a near optimal decoding strategy, we confirmed that the noise covariance matrix Σ was stimulus-independent in the reduced, five-dimensional space used to calculate d' (Extended Data Fig. 7f). We found that the matrix elements of the noise covariance matrix were highly correlated across the two stimulus conditions (r: 0.81 ± 0.16, mean ± s.d., $N = 5$ mice). This indicates that other more complex, nonlinear decoding strategies are unlikely to substantially surpass the accuracy of the linear strategy, which we further confirmed via an analysis of quadratic decoding (Extended Data Fig. 7h).

We also verified that we had sufficient numbers of visual stimulation trials to estimate d' accurately (Extended Data Fig. 7g). For every mouse, d' approached an asymptote as the number of stimulation trials used for analysis was increased; this indicates that beyond a certain point the computed value of d' is insensitive to the number of trials. Moreover, we developed an analytic theory describing how the accuracy of our estimates of d' depends jointly on the numbers of neurons and experimental trials (Extended Data Fig. 10f–k, Appendix).

In addition to our analyses of real data, we also calculated $(d'_{\text{shuffled}})^2$ (Fig. 3b–g), the optimal linear discrimination performance using trial-shuffled datasets, which we created by shuffling the responses of each cell across stimulation trials of the same type. Owing to this shuffling procedure, the off-diagonal elements of Σ_A and Σ_B become near zero.

We further calculated the performance of a 'diagonal' discrimination strategy (Fig. 3b, d, e) that was blind to the noise correlations between neurons, using the actual (unshuffled) datasets[30]. For this sub-optimal strategy, $(d'_{\text{diagonal}})^2$ determines the separation of two response distributions obtained when the vector of decoding weights $\mathbf{w}$ is collinear with $\Delta\boldsymbol{\mu}$ (Fig. 3), which we calculated according to:

$$(d'_{\text{diagonal}})^2 = \frac{(\Delta\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\Delta\boldsymbol{\mu})^2}{\Delta\boldsymbol{\mu}^{\mathrm{T}}\Sigma_{\mathrm{d}}^{-1}\Sigma\Sigma_{\mathrm{d}}^{-1}\Delta\boldsymbol{\mu}}$$

where Σ_d is the diagonal covariance matrix.

Eigenvalues of the noise covariance matrix

To examine how the statistical structure of neural noise affects the ability to discriminate neural responses to the two different visual stimuli (Fig. 4, Extended Data Fig. 10a–e), we expressed $(d')^2$ in terms of the eigenvalues λ_α and eigenvectors $\boldsymbol{e}_\alpha$ of the noise covariance matrix Σ:

$$(d')^2 = \Delta\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\Delta\boldsymbol{\mu} = \sum_\alpha \left(\frac{|\Delta\boldsymbol{\mu} \cdot \boldsymbol{e}_\alpha|^2}{\lambda_\alpha} \right)$$

which can be viewed as a sum of signal-to-noise ratios, one for each eigenvector. Clearly, the eigenvectors well aligned with $\Delta\boldsymbol{\mu}$ are the most important for discriminating between the two distributions of neural responses. Noting that λ_α equals the noise variance along $\mathbf{e}_\alpha$, our data revealed noise modes that were well aligned with $\Delta\boldsymbol{\mu}$ and for which the variance increased linearly with the number of cells. The combination of these two attributes is what leads to the saturation of d' as the number of cells in the ensemble becomes large (Fig. 4). Notably, our analysis also uncovered noise modes with much larger variance that are not information-limiting, as they do not align well with $\Delta\boldsymbol{\mu}$.

Calculation of decoding weights

We calculated the vector of optimal linear decoding weights, $\mathbf{w}_{\text{opt}}$, in the reduced space identified by PLS analysis:

$$\mathbf{w}_{\text{opt}} = \Sigma^{-1}\Delta\boldsymbol{\mu}$$

For moving grating visual stimuli oriented at ±30°, $\mathbf{w}_{\text{opt}}$ was generally well aligned to $\Delta\boldsymbol{\mu}$, indicating that correlation-blind decoding performed near optimally (Figs. 3b, h, 4a, c). This was somewhat less the case with moving gratings oriented at ± 6° (Extended Data Fig. 9c). To assess the contributions of individual cells to the optimal decoder, we estimated the vector of decoding weights in the space of all neurons as:

$$\mathbf{w}_{\text{decoding}} = \frac{T^{\mathrm{T}}\mathbf{w}_{\text{opt}}}{||\mathbf{w}_{\text{opt}}||}$$

where T is a transformation matrix from the high-dimensional population vector space, in which the responses of each cell occupy an individual dimension, into the five-dimensional space identified by PLS analysis. Starting around 0.4 s after the onset of visual stimulation, $\mathbf{w}_{\text{decoding}}$ was largely time-invariant (Extended Data Fig. 7d).

L2-regularized regression

Because our method for computing d' via PLS analysis involved a dimensional reduction, we compared the d' values found with PLS analysis to those determined via a different method, L2-regularized regression[52], which does not depend on dimensional reduction (Extended Data Fig. 8a, b). This form of regression uses a regression vector, $\mathbf{b}$, that lies within the high-dimensional space of all ensemble neural activity patterns, but its length is limited by the use of an adjustable regularization parameter, k. For each subset of neurons considered, we randomly chose 90% of the visual stimulation trials for the determination of $\mathbf{b}$. We projected the neural responses from the remaining 10% of trials onto the dimension determined by $\mathbf{b}$. We then computed d' with the same formula as used with PLS analysis, except with $\mathbf{b}$ replacing $\mathbf{w}_{\text{opt}}$, the optimal linear discrimination hyperplane. Using this approach, we found the maximum value of d' across all values of k within the range $[1, 10^5]$. We averaged these maximal d' values across 100 different subsets of neurons and visual stimulation trials (Extended Data Fig. 8a).

Kullback–Leibler divergence

To assess the extent to which quadratic decoding might surpass the optimal linear decoder, we computed the Kullback–Leibler (KL) divergence[31] between the two distributions of ensemble neural responses to the two different visual stimuli (Extended Data Fig. 7h). The KL divergence is a generalization of d' to arbitrary distributions and, like d', provides an assessment of the statistical differences between two distributions. When the two distributions are Gaussians with equal covariance matrices, the KL divergence reduces to $(d')^2$, and linear decoding methods suffice to optimally discriminate between the two distributions[52]. By comparison, for two Gaussian distributions with different means and covariance matrices, $(d')^2$ is not equivalent to the KL divergence, and quadratic decoding methods are required to optimally discriminate between the two distributions[52].

To assess the potential benefits of quadratic decoding, we fit multivariate Gaussians to the two stimulus response distributions without assuming they had equal covariance matrices. We computed the KL divergence of the response distribution to stimulus A relative to the response distribution to stimulus B according to:

$$KL_{A||B} = \frac{1}{2}\left(\mathrm{tr}(\Sigma_A^{-1}\Sigma_B) + \Delta\boldsymbol{\mu}^T\Sigma_A^{-1}\Delta\boldsymbol{\mu} - N + \ln\left(\frac{\det(\Sigma_A)}{\det(\Sigma_B)}\right)\right)$$

where Σ_A, Σ_B are the noise covariance matrices for the two stimulation conditions, $\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_A - \boldsymbol{\mu}_B$ is the vector difference between the mean ensemble neural responses to the two stimuli, and N is the dimensionality of the response distribution (that is, the number of cells in the ensemble). The KL divergence saturated as N increased and was generally not much greater than $(d')^2$ (Extended Data Fig. 7h). This result was consistent with the finding that the noise covariance matrix was similar for the two different visual stimuli (Extended Data Fig. 7f) and supported the conclusion that quadratic decoding would achieve little performance gain beyond that of the optimal linear decoder.

Computational studies of the robustness of empirically determined d' values

To verify that our decoding methods were robust to the potential presence of effects such as common mode fluctuations and multiplicative gain modulation that could increase the trial-to-trial variability of neural responses, we compared the d' values obtained from PLS analysis versus L2-regularized regression using computationally simulated datasets of neural population responses (Extended Data Fig. 8c–h).

First, to examine the combined effects of information-limiting correlations and common mode fluctuations (Extended Data Fig. 8c–f), we studied a model of the neural ensemble responses in which the noise covariance matrix exhibited information-limiting noise correlations via a single eigenvector, $\boldsymbol{f}$, the eigenvalue of which grew linearly with the number of cells in the ensemble. In addition to this rank 1 component, we included a noise term that was uncorrelated between different cells, as well as a common mode fluctuation, yielding a noise covariance matrix with the form

$$\Sigma^* = \sigma^2 I + \varepsilon_{\mathrm{common}} J + \varepsilon \boldsymbol{f}^T\boldsymbol{f}$$

where $\sigma^2 = 1$ is the amplitude of uncorrelated noise, I is the identity matrix, J is a rank 1 matrix of all ones, and $\boldsymbol{f}$ is the information-limiting direction, a vector that we chose randomly in each individual simulation from a multi-dimensional Gaussian distribution with unity variance in each dimension. The amplitude of information-limiting correlations was $\varepsilon = 0.002$, approximately matching the level observed in the experimental data. In the model version without common mode fluctuations, we set $\varepsilon_{\mathrm{common}}$ to zero. In the version with common mode fluctuations, we set $\varepsilon_{\mathrm{common}} = 0.02$, ten times the value of ε. We chose the difference in the means of the two stimulus response distributions, $\Delta\boldsymbol{\mu}$, to be aligned with $\boldsymbol{f}$ (Fig. 3a) and to have a magnitude of 0.2, so that the asymptotic value of d' for large numbers of cells approximately matched that of the data. We compared the decoding results attained with and without the presence of common mode fluctuations in the neural responses.

Second, to study the possible effects of multiplicative gain modulation (Extended Data Fig. 8g, h), we compared two versions of a model in which the responses of the V1 neural population either were or were not subject to a multiplicative stochastic gain modulation but were otherwise statistically equivalent. We modelled the V1 cell population as a set of linear Gabor filters (see Appendix section 5). In the version with gain modulation, on each visual stimulation trial we multiplied the output of the Gabor filter by a randomly chosen factor, uniformly distributed between 50–150%, the value of which was the same for every cell but varied from trial to trial.

Estimates of perceptual acuity

We used the empirical determinations of d' based on visual cortical activity and the parameters of the moving grating visual stimuli to estimate the minimum perceptible orientation difference between the two stimuli. We compared the resulting values to those estimated from past behavioural measurements of visual acuity in mice[15,34], all of which agree well.

One behavioural study assessed how well three individual mice could discriminate the orientations of visual gratings[34]. The best trained of these three mice—that is, the mouse that performed the most sessions and had the smallest error bars in the threshold determination—had a behavioural threshold for orientation discrimination (4.6° ± 0.1°; $n = 7$ sessions) close to the value estimated from our neural data (4.8°). The second mouse had a 5.7° ± 0.6° threshold ($n = 4$ sessions), and the third mouse had a threshold of 6.9° ($n = 1$ session).

Another behavioural study examined visual acuity in 13 mice and determined the highest visual spatial frequencies the mice could discern[15]. To compare our results to this study, we used the fact that our grating stimuli had a low spatial frequency (0.04 cycles per degree) to approximate the perceptual challenge of estimating the grating orientation as being equivalent to that of estimating the orientation of the line of peak illumination intensity over the same viewing diameter. In the behavioural study of acuity[15], the mice used both eyes to view the stimulus, whereas in our studies mice viewed the stimulus with one eye, and we recorded neural activity from only one cerebral hemisphere. To account for these differences, we posited that neural noise fluctuations should be nearly independent across the visual streams from the two eyes, which would boost d' values by about a factor of √2 over those achievable with one eye. However, our determinations of d' from neural activity concern the discrimination of two distinct visual stimuli, which should also increase d' values by a factor of about √2 over those for a single stimulus viewed with one eye. Given these counterbalancing factors, we used the d' values to estimate the highest perceptible spatial frequency as $f \approx d'(\theta)/D \sin\theta$, where D is the diameter of the visual stimuli (50 deg; Fig. 2b) presented at orientations of $\pm\theta$. For the grating stimuli oriented at ±30° to vertical, $d' \approx 6$, yielding $f \approx 0.3$ cycles per degree. For the grating stimuli oriented at ±6°, which are more representative of the perceptual threshold, $d' \approx 2.5$ and thus $f \approx 0.48$ cycles per degree, comparable to the value of $f \approx 0.5$ cycles per degree attained from the behavioural studies at a unity d' value for the behavioural performance[15]. We converted values of f into the minimum perceptible orientation difference, $2\theta_{\mathrm{min}}$, between two grating stimuli oriented at $\pm\theta_{\mathrm{min}}$ by using $\theta_{\mathrm{min}} = \sin^{-1}(1/Df)$. This conversion yielded a prediction of $\theta_{\mathrm{min}} \approx 2.3°$ based on the behavioural studies of mouse visual acuity[15], as compared to $\theta_{\mathrm{min}} \approx 2.4°$ based on our neural data.

Computational simulations of activity in a two-layer neural network

To illustrate that cells whose receptive fields overlap exhibit shared noise correlations, we simulated a simple two-layer feed-forward network of linear neurons, with 14 input neurons and 3 output neurons (Extended Data Fig. 1j–m). The neurons in each layer were equally spaced along a linear axis. We defined the strengths of the connections, $\mathbf{w}_i$, between the input and output neurons such that the receptive field profiles of the different output neurons were spatially overlapping Gaussian functions of the linear separation between each output neuron and the input neurons (Extended Data Fig. 1j).

For the three example cells shown in Extended Data Fig. 1j, the unity-normalized overlap between their connection weight vectors was: $\mathbf{w}_1 \cdot \mathbf{w}_2 = 0.165$, $\mathbf{w}_1 \cdot \mathbf{w}_3 = 0.022$ and $\mathbf{w}_2 \cdot \mathbf{w}_2 = 0.038$. The activity of cells in the output layer, $\mathbf{r}$ was defined as: $r_i = [\mathbf{w}_i \cdot (\mathbf{x} + \mathbf{n})]$, where $\mathbf{x}$ is the mean activity of the input cells in response to a given stimulus, $\mathbf{n}$ is a noise term in which each element is Poisson-distributed with mean 0.1, and [] denotes rounding to the nearest integer. We simulated the activity

of this two-layer network across 10,000 time bins and calculated the noise correlation coefficients between three different pairs of output neurons.

Measurements of fluorescence scattering

To examine the extent of fluorescence scattering between active image tiles within one temporal phase of our multiplexed imaging scheme (Extended Data Fig. 4d–g), we measured the spatial distribution function, $P_S(x, y)$, governing the probability that a two-photon excited fluorescence photon will exit the cortical tissue surface at a point with lateral displacement coordinates (x, y) relative to the laser focus. To directly observe the distributions $P_S(x,y)$ of scattered fluorescence, we built a custom optical setup that used the Ti:sapphire laser beam to excite fluorescence in fixed cortical tissue slices from adult *GCaMP6f-tTA-dCre* mice and imaged the resulting distribution of fluorescence signals on a scientific grade CMOS camera (Orca Flash, Hamamatsu). Owing to the use of an imaging detector in this setup, the fluorescence detection pathway had to be optically corrected for field curvature and other image plane distortions, whereas the primary two-photon microscope (Fig. 1) had no such requirement. For this reason, our studies of scattering used an Olympus XLUMPLFLN objective lens (0.95 NA, 20×), which provided fluorescence images of ~1.2 mm in width. We positioned the laser focal spot on one side of the field of view, so as to image scattered fluorescence up to about 1.1 mm away from the focal spot (Extended Data Fig. 4e, f). We computed the mean $P_S(x, y)$ distribution, averaged over 100 different locations of the laser focus in each of 3 different brain slices, at tissue depths up to 600 μm beneath the surface of the slice. To determine the mean cross-sectional distribution of fluorescence as a function of the radial distance from the laser focus, $r = \sqrt{x^2 + y^2}$, we also averaged over all accessible polar angles. To compute the probability that a fluorescence photon excited in one active tile would scatter into an adjacent active tile, we integrated the circularly symmetric determinations of $P_S(x, y)$ over the portion of the image area yielding this form of crosstalk (Extended Data Fig. 4g).

Measurements of brain temperature during two-photon brain imaging

To perform temperature measurements in the brains of awake mice during two-photon imaging (Extended Data Fig. 2f), we surgically prepared *GCaMP6f-tTA-dCre* mice by performing a 5-mm-diameter craniotomy following the same procedures as described above. However, before placement of the cranial window, we inserted a flexible 200-μm-diameter thermocouple probe[53] (IT24P; Physitemp) into the brain, 100–200 μm beneath the dura, within ~0.75 mm of the centre of the field-of-view of the microscope. The thermocouple resided within a 5-mm-long plastic micropipette and extended ~2.5 mm beyond the tip of the micropipette.

Using ultraviolet-light curable glue (Loctite, 4305) and dental cement, we affixed the micropipette to the cranium at a shallow angle of 5° relative to the surface of the cranium. We then placed the glass cranial window onto the craniotomy and fixed the window in place with dental cement. The thermocouple probe was connected to a two-channel digital thermometer (CL3515R; Omega), which conveyed digitized temperature data (10 Hz sampling rate) to a computer via a USB port. We protected the wires of the thermocouple connecting to the digital thermometer using a 5-cm-long piece of flexible plastic tubing. We then commenced concurrent two-photon imaging (17.5 Hz image frame acquisition rate) and temperature recordings (Extended Data Fig. 2f).

Histology

To check whether in vivo two-photon imaging with the 16-beam instrument induced any brain tissue damage, we performed immunohistochemical analyses of post-mortem brain tissue sections (Extended Data Fig. 2g–i). We compared positive control tissue sections that we had deliberately damaged in vivo with high-power (2,680 mW mm^{-2}) laser illumination, negative control tissue sections that received no laser illumination, and experimental tissue sections that had undergone in vivo two-photon imaging at the highest intensity levels of laser illumination (80 mW mm^{-2}) used in this study for tracking neuronal Ca^{2+} dynamics.

We euthanized and intracardially perfused the mice in all three groups with phosphate buffered saline followed by a 4% solution of paraformaldehyde in phosphate buffered saline. To allow adequate time for expression of HSP70 following exposure to laser illumination[54], mice in the positive control and experimental groups were euthanized 21 h after the end of two-photon imaging. We sliced the fixed brain tissue using a vibratome (Leica VT1000 s) to obtain 100-μM-thick coronal sections. We immunostained the tissue sections with antibodies against glial fibrillary activation protein (1:2,500, rabbit anti-GFAP, Sigma HPA056030, Lot C115616) and heat shock protein 70 (1:400, mouse anti-HSP, Enzo ADI-SPA-810, Clone C92F3A-5, Lot 01031912) and then applied fluorophore-conjugated secondary antibodies (goat anti-rabbit-Alexa 594 (Invitrogen, A-11012, Lot 1933366) and goat anti-mouse-Alexa 488 (Invitrogen, A-11001, Lot 56881A)).

We also stained the sections with DAPI (Invitrogen, D1306), which labels cell nuclei by binding to DNA. After mounting the brain sections on glass slides, we visualized immunofluorescence using an epifluorescence macroscope (Leica, MZFL III) equipped with a plan 1.0× objective lens, a solid-state white light engine (Lumencor, Sola SM 5-LCR-VA), filter sets for imaging red and green fluorophores (Leica 10450756 and 10450212, respectively) and a CCD camera (QImaging, 01-QIClick-F-M-12). Brain sections from all three groups were imaged under identical optical conditions and with the same camera settings.

Statistical tests

For comparison of the distributions of noise correlation coefficients in Fig. 2e and Extended Data Fig. 6g, we used two-tailed, two-sample Kolmogorov–Smirnov tests. In Figs. 2f, 3h and Extended Data Fig. 9c we used one-tailed Wilcoxon rank-sum tests. Supplementary Table 1 contains all *P* values associated with the figures and extended data figures.

Instrument availability

With support from the United States National Institute of Neurological Disorders and Stroke, we are currently converting the large-scale two-photon microscope (Fig. 1, Extended Data Fig. 2) into a research facility that is available to other laboratories and formally overseen by a steering committee. Researchers interested in this facility should please write to its principal investigator (M.J.S.) for more information.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Code availability

We used open source software routines for image registration[44] (http://bigwww.epfl.ch/thevenaz/turboreg/) and partial least squares analysis (https://www.mathworks.com/matlabcentral/fileexchange/18760-partial-least-squares-and-discriminant-analysis). Software code for extracting individual neurons and their Ca^{2+} activity traces from Ca^{2+} videos using principal component and then independent component analyses[35,48] is freely available (https://www.mathworks.com/matlabcentral/fileexchange/25405-emukamel-cellsort), although for convenience we used a commercial version of these routines (Mosaic software, version 0.99.17; Inscopix). We wrote all other analysis routines

in MATLAB (Mathworks; version 2017b). The primary software code used to support the findings of the study is available at Zenodo.org (https://zenodo.org/record/3593520#.XgWPu-hKg2w).

36. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
37. Madisen, L. et al. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).
38. Wekselblatt, J. B., Flister, E. D., Piscopo, D. M. & Niell, C. M. Large-scale imaging of cortical dynamics during sensory perception and behavior. *J. Neurophysiol*. **115**, 2852–2866 (2016).
39. Chettih, S. N. & Harvey, C. D. Single-neuron perturbations reveal feature-specific competition in V1. *Nature* **567**, 334–340 (2019).
40. Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
41. Chen, T. W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
42. Huber, D. et al. Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484**, 473–478 (2012).
43. Kim, K. H. et al. Multifocal multiphoton microscopy based on multianode photomultiplier tubes. *Opt. Express* **15**, 11658–11678 (2007).
44. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
45. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463–1465 (2009).
46. Brown, M. & Lowe, D. G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **74**, 59–73 (2007).
47. Pnevmatikakis, E. A. & Giovannucci, A. NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *J. Neurosci. Methods* **291**, 83–94 (2017).
48. Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* **63**, 747–760 (2009).
49. Vogelstein, J. T. et al. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol*. **104**, 3691–3704 (2010).
50. Bishop, C. M. *Pattern Recognition and Machine Learning* Vol. 1 (Springer, 2007).
51. Geladi, P. & Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **185**, 1–17 (1986).
52. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistcal Learning* (Springer, 2009).
53. Podgorski, K. & Ranganathan, G. Brain heating induced by near-infrared lasers during multiphoton microscopy. *J. Neurophysiol*. **116**, 1012–1023 (2016).
54. Graner, M. W., Cumming, R. I. & Bigner, D. D. The heat shock response and chaperones/heat shock proteins in brain tumors: surface expression, release, and possible immune consequences. *J. Neurosci.* **27**, 11214–11227 (2007).
55. Kalmbach, A. S. & Waters, J. Brain surface temperature under a craniotomy. *J. Neurophysiol*. **108**, 3138–3146 (2012).
56. Wang, H. et al. Brain temperature and its fundamental properties: a review for clinical neuroscientists. *Front. Neurosci*. **8**, 307 (2014).
57. Talan, M. Body temperature of C57BL/6J mice with age. *Exp. Gerontol*. **19**, 25–29 (1984).
58. Greenberg, D. S., Houweling, A. R. & Kerr, J. N. D. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci*. **11**, 749–751 (2008).
59. Karimipanah, Y., Ma, Z., Miller, J. K., Yuste, R. & Wessel, R. Neocortical activity is stimulus- and scale-invariant. *PLoS ONE* **12**, e0177396 (2017).

Acknowledgements We acknowledge a Stanford Graduate Fellowship (O.I.R.), research support from the Howard Hughes Medical Institute (M.J.S.), the Stanford CNC Program (M.J.S.), DARPA (M.J.S.), an NSF CAREER Award (S.G.), and the Burroughs-Wellcome (S.G.), McKnight (S.G.), James S. McDonnell (S.G.) and Simons (S.G.) foundations. NIH grants MH085500 and DA028298 to H.Z. funded development of the *GCaMP6f-tTA-dCre* and *Rasgrf2-2A-dCre* mice. NIH grant R24NS098519 (M.J.S.) supports our effort to make the 16-beam two-photon microscope an open resource available to other laboratories. We thank T. Moore, P. Jercog, J. C. Jung, D. Vucinic, B. F. Grewe, E. T. W. Ho, H. Kim, X. Pitkow and T. Zhang for discussions, D. Flickinger and K. Svoboda for providing design files for the water-immersion objective lens, C. Niell for providing *tetO-GCaMP6 s/CaMK2a-tTA* mice, and C. Irimia for animal husbandry.

Author contributions O.I.R., S.G. and M.J.S. designed experiments and analyses. O.I.R., J.A.L. and J.S. designed and built the microscope. O.I.R., J.A.L., O.H., Y.Z., R.C. and J.L. acquired and analysed data. S.G. developed theory and analysed data. H.Z. provided transgenic mice. O.I.R., S.G. and M.J.S. wrote the paper. All authors edited the paper. S.G. and M.J.S. supervised the research.

Competing interests M.J.S. is a scientific co-founder of Inscopix, which produces the Mosaic software used to identify individual neurons in the Ca^{2+} videos. J.A.L. is also an Inscopix stockholder.

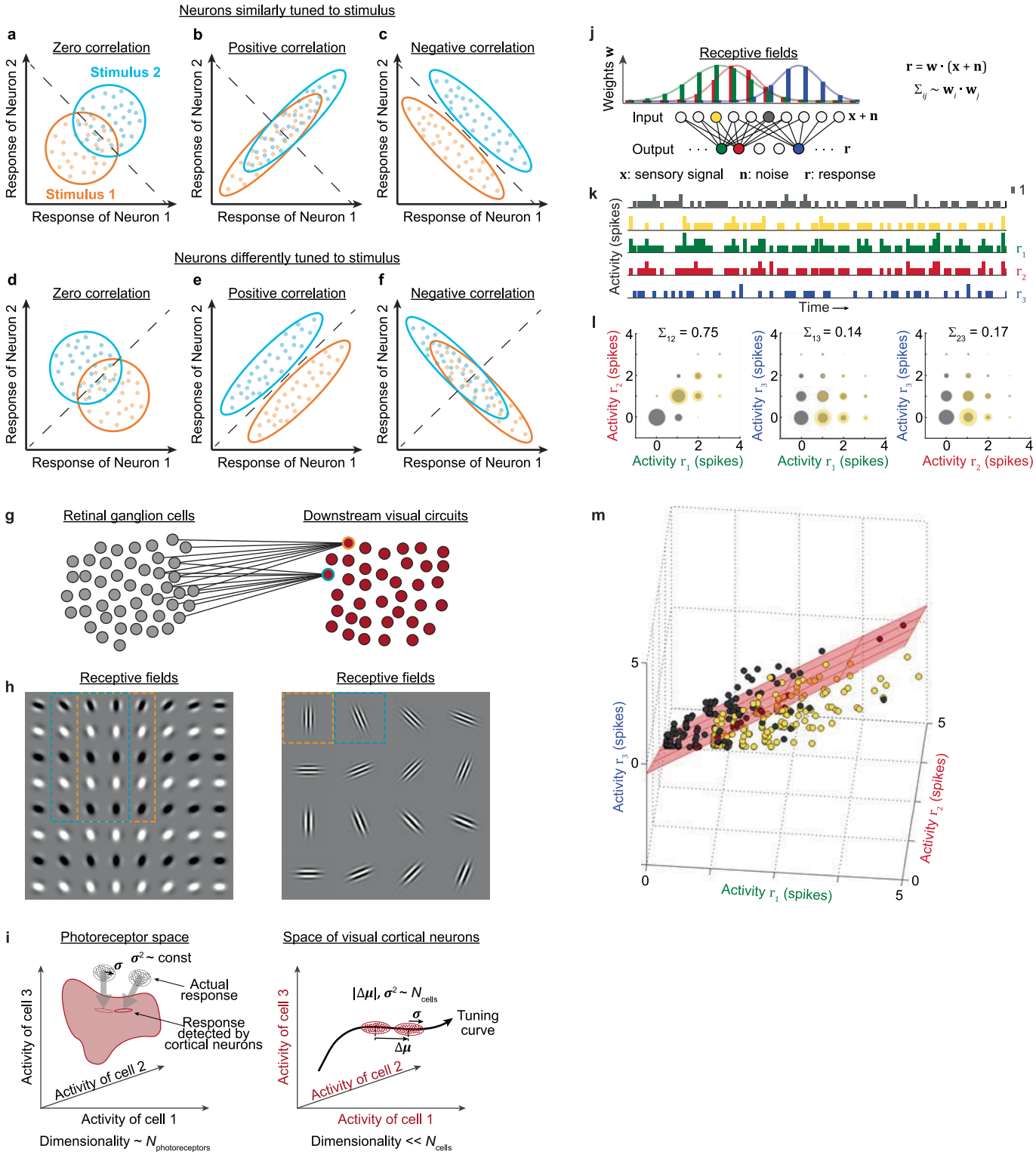
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2130-2.

Correspondence and requests for materials should be addressed to O.I.R., S.G. or M.J.S.

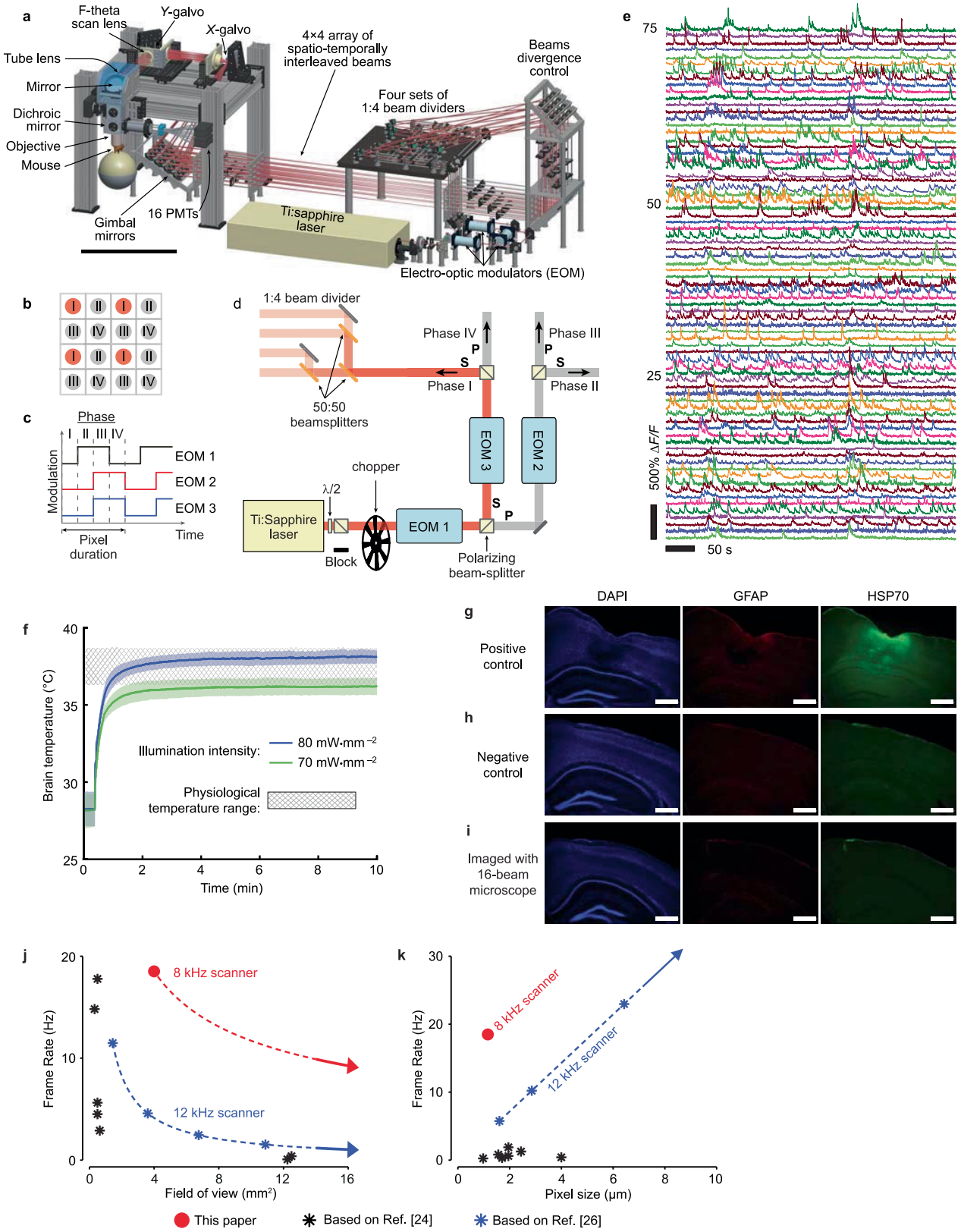
Peer review information *Nature* thanks Stefano Panzeri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.



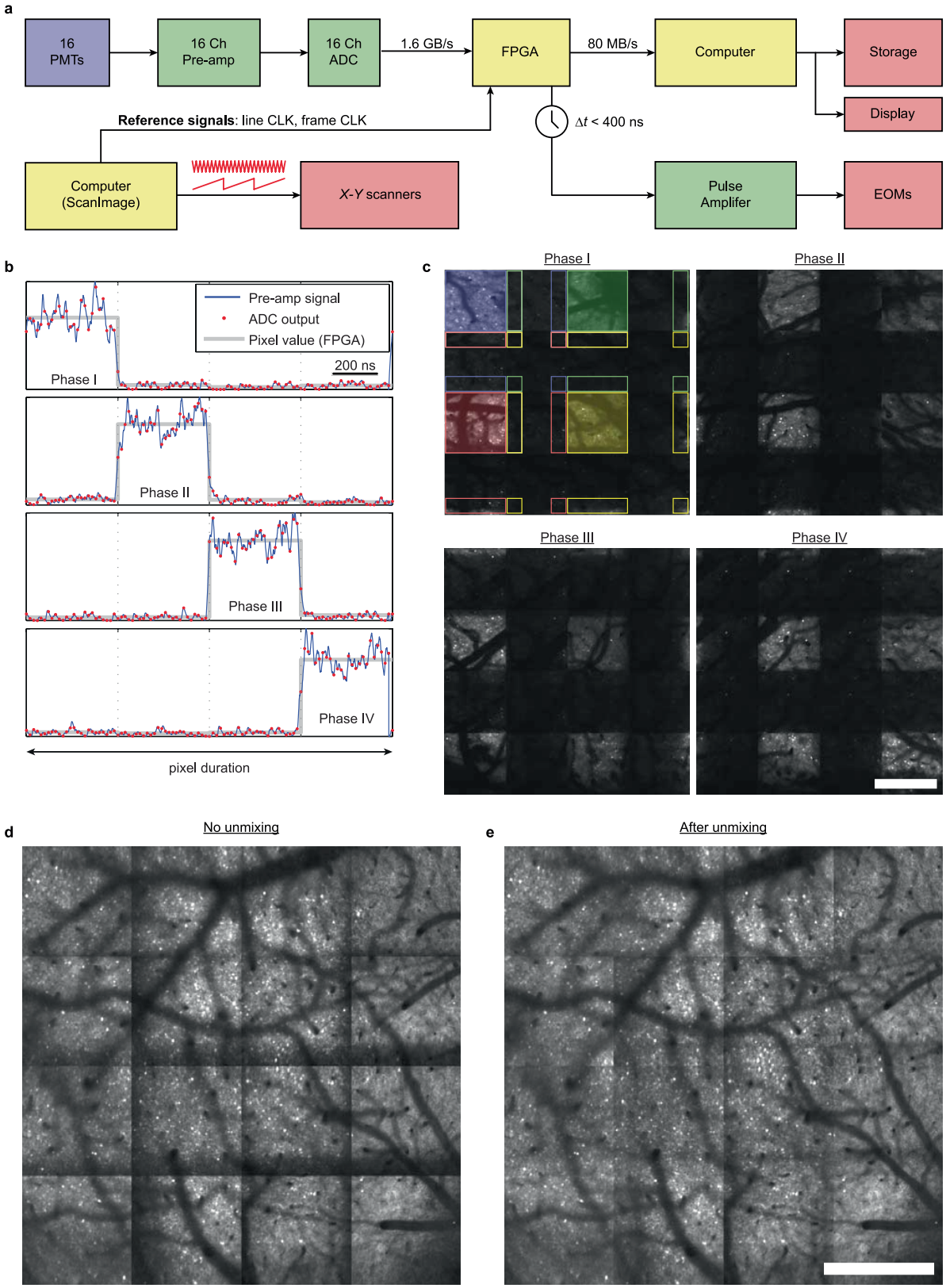
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | The discriminability of two sensory stimuli based on the activity patterns of two or more cells depends on the statistical relationship between the mean responses of the cells and their noise correlations, which in turn depends on visual neural circuitry. **a**–**f**, Schematics of the distributions of responses by two cells to two distinct stimuli in six different cases. Cyan dots indicate joint responses of the cell pair to stimulus 1; orange dots indicate responses to stimulus 2. Ellipses convey the shapes of the statistical distributions of the responses to each stimulus. Three types of noise correlation are depicted. In **a** and **d**, the two cells have statistically independent noise correlations. In **b** and **e**, the cells share positively correlated noise fluctuations. In **c** and **f**, the cells share negatively correlated noise fluctuations. In all six cases, dashed lines indicate optimal linear boundaries for stimulus discrimination. The information in **a**–**f** is based on similar plots published previously[5,11,30]. **a**–**c**, When both neurons have similar stimulus-response properties (for example, as schematized, when both cells have a smaller mean response to stimulus 1 than stimulus 2), positively correlated noise fluctuations (**b**) increase the overlap between the two response distributions and thereby impair stimulus discrimination, whereas negatively correlated noise fluctuations (**c**) improve stimulus discrimination as compared to the case with independent noise fluctuations (**a**). **d**–**f**, When both neurons have opposite stimulus tuning (for example, as schematized, when neuron 1 responds more vigorously to stimulus 1 and neuron 2 responds more vigorously to stimulus 2), positively correlated noise fluctuations (**e**) decrease the overlap between the two response distributions as compared to the case with independent noise fluctuations (**d**) and thereby improve stimulus discrimination, whereas negatively correlated noise fluctuations (**f**) impair stimulus discrimination by increasing the overlap of the two response distributions. **g**, Cells in visual cortical areas, denoted by red circles, integrate signals from earlier stages of the visual pathway, as schematized by the input connections to two example cortical neurons. Thus, as visual information propagates through neural circuitry, noise fluctuations become correlated between cells with similar receptive fields, leading to an upper bound on the amount of information that a neural ensemble can encode. **h**, Example receptive fields for cells in **g**. Cells in early stages of the visual processing pathway have relatively simple receptive fields. Integration of their activity patterns leads to more complex visual receptive fields in downstream visual areas. Dashed boxes enclose receptive fields (right) for the two example cells marked in **g**, as well as the receptive fields of cells providing visual inputs (left). **i**, A network's pattern of synaptic connectivity constrains the dimensionality of the activity in downstream visual circuits[12]. Left, in the early layers of the visual pathway, the dimensionality of ensemble activity is about the same order of magnitude as the number of photoreceptors. In downstream visual areas, due to the extraction of visual features, neural activity is constrained to a manifold of lower dimensionality (indicated by the red-shaded manifold in the space of all possible photoreceptor inputs). This manifold is determined by the set of receptive fields and hence the visual features that the downstream visual area detects. Grey ellipses (left) depict the distributions of photoreceptor responses to two distinct visual stimuli; after propagating through the visual circuitry these distributions are confined to the lower-dimensional manifold (red ellipses). Right, for a family of visual stimuli parameterized by a single variable, the mean neural ensemble responses lie along a corresponding tuning curve. Noise in the input circuitry propagates to downstream areas and leads to noise fluctuations in downstream neurons that are statistically correlated for cells with similar receptive fields. This, in turn, implies that the magnitude of noise fluctuations along the neural tuning curve becomes proportional to the number of cells in a neural ensemble and indistinguishable from the encoded visual signals, which also increase in proportion to the number of cells. This proportional growth of noise and signal ultimately limits the ability to discriminate two visual stimuli. Thus, for neural ensembles with more than a certain number of cells, the encoded information reaches an upper bound. **j**, We simulated a two-layer, linear feedforward neural network, to illustrate that information-limiting correlations are intrinsic to feed-forward neural networks with overlapping receptive fields[12]. Top, for three example output cells, the plot shows the synaptic weights of the inputs from cells in the first layer of the network. Bottom, diagram of connections between the two layers of the network. Symbols are defined as follows: **x** is the mean activity of cells in the first layer in response to a given stimulus; **n** is the noise in the activity of the input cells; **r** is the activity of the output cells. **k**, Digitized plots of spike counts for simulated activity in the network of **j**, for the two example input cells (yellow and black) and three example output cells (red, green, blue). The noise traces for the input cells came from independent Poisson random processes. External inputs to the network selectively drove either the yellow or the black cell, but owing to the presence of noise the two cells are occasionally active concurrently. **l**, Frequency plots of pairwise activity levels (rounded to the nearest integer) for pairs of output cells in the network of **j**. Yellow and black circles denote which of the two corresponding input cells received external input. The diameter of each circle denotes the number of time bins with a given pair of activity levels in the two cells. Σ values are noise correlation coefficients and are larger for pairs of output cells with greater overlap in their receptive fields. **m**, Plot of the distribution of activity responses in the output cell layer, for the three example cells coloured green, red and blue in **j**. Data points are coloured either yellow or black, to indicate whether the output activity is a response to stimulation of the yellow- or black-coloured cell in the input layer. The red plane denotes the optimal linear classification boundary between the two stimulation conditions.



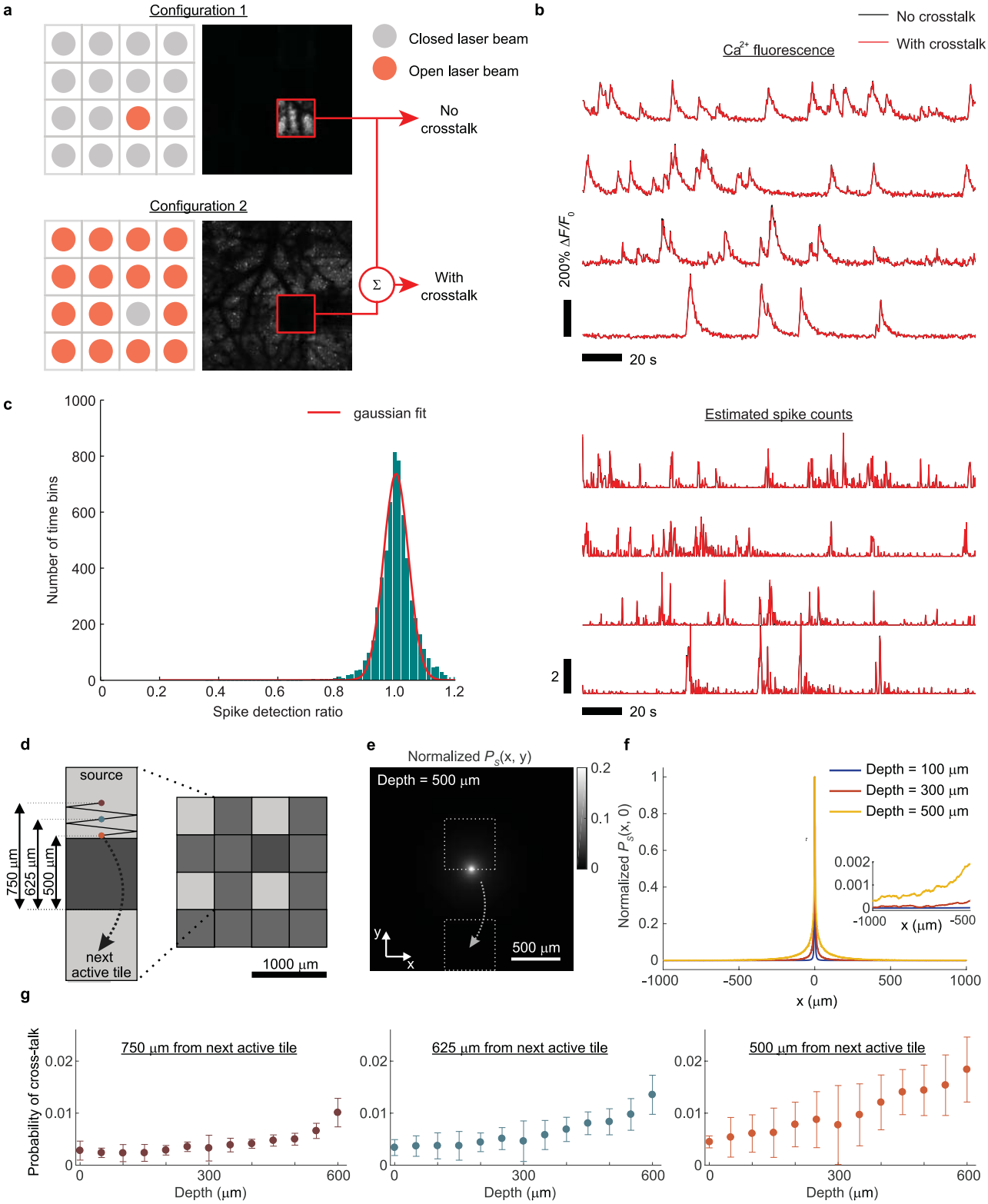
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Spatiotemporal multiplexing of the illumination beams permits imaging of large fields of view at fast frame rates without thermal damage to brain tissue. a, Computer-assisted design of the mechanical layout of the two-photon microscope. Scale bar, 0.5 m. **b**, In the pixel multiplexing mode of imaging, each of the 16 beams are assigned to one of four different temporal phases within each cycle of the pixel clock (Extended Data Fig. 3b). Alternatively, in the line-multiplexing mode of imaging, only 8 of the 16 beam paths are used (Methods). In neither imaging mode are neighbouring beams ever active concurrently (Extended Data Fig. 3c), minimizing fluorescence scattering between active image tiles and allowing scattering into inactive image tiles to be corrected computationally (Extended Data Figs. 3d, e, 4a–g). **c**, To switch between the different sets of active beams, square-wave electronic signals control a set of three electro-optic modulators (EOMs). **d**, A Ti:sapphire laser provides ultrashort-pulsed infrared illumination. A half-wave ($\lambda/2$) plate and a polarizing beam-splitter enable power control. Three pairs of EOMs and polarizing beam-splitters direct the light into one of four main optical paths, with only one path illuminated during each of the four multiplexing phases. In each of these four main paths, three 50:50 beam-splitters create four beams of equal intensity, yielding up to 16 total beams but with only four on at any instant. A chopper blocks all light during the turnaround portion of the galvanometer scanning cycle. **e**, Seventy-five example fluorescence traces of Ca^{2+} activity in layer 2/3 pyramidal cells of an awake mouse. **f**, Maintaining brain temperature within physiological ranges during in vivo two-photon imaging requires a proper balance between heat loss through the cranial window and heating induced by the laser illumination[53,55]. To directly verify that our cranial window preparation and imaging conditions properly balanced these two opposing effects, we measured brain temperature during two-photon imaging with the 16-beam microscope. For these studies we used an implanted thermocouple[53] and either the highest (blue trace) or lowest (green trace) time-averaged laser illumination intensity used for Ca^{2+} imaging elsewhere in this study (Methods). Consistent with previous work, before laser illumination commenced the brain temperature was about 9 °C below normal mouse body temperature[55], a state that is considered to be neuroprotective[56]. By about 100 s after the start of imaging, brain temperatures attained steady-state values within the physiological range of C57BL/6 mice[57] (grey shaded region; 36.3 °C–38.7 °C). Each trace is an average of three bouts of imaging for each of three separate mice. Coloured shading denotes the s.d. across the 9 individual measurements acquired at each illumination intensity. **g**–**i**, Fluorescence immunohistochemical analyses of tissue damage markers. To check whether in vivo imaging of brain tissue with the 16-beam instrument (4 mm^2 field of view) induced any tissue damage, we immunostained post-mortem brain tissue sections using antibodies to two different damage markers, glial fibrillary activation protein (GFAP) and heat shock protein 70 (HSP70), previously identified as indicators of laser-induced tissue damage[53]. We also stained the sections with DAPI, which labels cell nuclei. We compared positive control tissue sections (**g**) that we had deliberately damaged in vivo with high-power (2,680 mW mm^{-2}) laser illumination, negative control sections (**h**) that received no laser illumination, and experimental tissue sections (**i**) that had undergone in vivo two-photon imaging at the highest level of laser illumination (80 mW mm^{-2}) used in this study for tracking Ca^{2+} dynamics in neocortical layer 2/3 pyramidal neurons. Together, these analyses verified the functionality of the antibodies and revealed no signs of tissue damage from two-photon imaging. To image neurons in cortical layers deeper than layer 2/3, users have several options for doing so without delivering excess heat to the brain (Supplementary Video 3, Supplementary Note). Scale bars, 500 µm. Results shown are representative of those from 8 cerebral hemispheres of 4 different mice. **j**, **k**, Comparisons between recent large-scale two-photon microscopes[24,26]. The performance of a laser-scanning microscope closely relates to four main parameters: the scanner speed, image-frame acquisition rate, field of view, and pixel size (Supplementary Note). For microscopes that use a single laser beam to sweep in two dimensions across the field of view, these parameters obey the relationship $FOV = d \times v \times f^{-1}$, where FOV is the field-of-view area, d is the spacing between adjacent image lines (or equivalently the pixel width along the slow-axis of laser-scanning), v is the speed at which the beam is swept across the specimen by the fast-axis scanner, and f is the image-frame acquisition rate. By comparison, our approach using four active beams leads to an expression for the maximal field of view, $FOV = 4 \times d \times v \times f^{-1}$. These relationships enable performance comparisons with other recently published large-scale two-photon microscopes[24,26]. To illustrate, **j** shows a plot of the image-frame acquisition rate against the field-of-view area, given a line spacing of $d = 1.15$ µm. **k** shows how the image-frame acquisition rate depends on d for a 4 mm^2 field of view. Solid red circles denote the performance of our microscope in its line-multiplexing imaging mode using an 8-kHz resonant galvanometer (Methods). Black data points denote performance options of another large two-photon microscope, which uses pair of laser beams with temporally interleaved pulses[24], as calculated on the basis of its published capabilities. Blue data points and associated blue dashed lines show performance options for a third large-scale microscope[26], as calculated on the basis of its published capabilities.



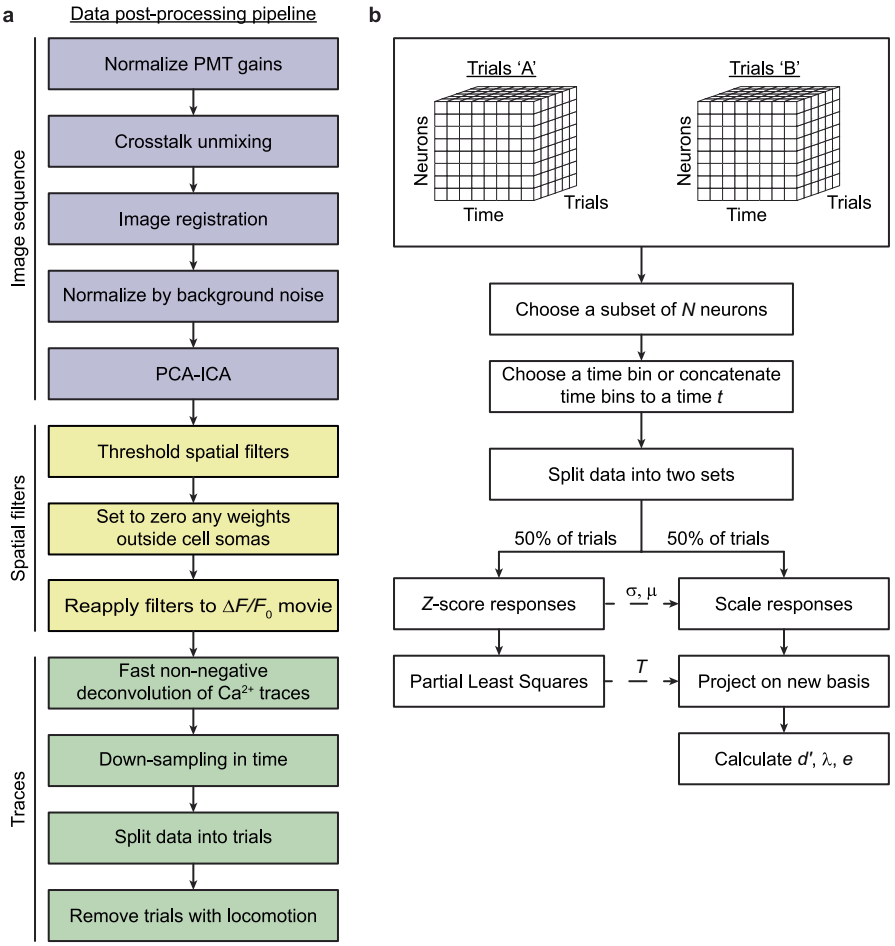
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Data acquisition and post-processing for two-photon imaging with 16 time-multiplexed excitation beams. a, Block diagram of the electronics for data acquisition and instrument control. PMT, photomultiplier tube; Pre-amp, pre-amplifier; ADC, analogue-to-digital converter; FPGA, field-programmable gate array; EOM, electro-optic modulator. **b**, Computer simulation of signal sampling in different stages of the pipeline in **a**. The ADC samples the analogue, pre-amplified and low-pass filtered signals (blue) from one of the PMTs at a rate of 5×10^7 samples per second. In each of the four temporal phases, the FPGA sums the digitized signals (red) from the ADC to yield the fluorescence intensity values of each image pixel (grey). **c**, Raw fluorescence images for each of the four excitation phases, acquired in an awake mouse expressing GCaMP6f in layer 2/3 cortical pyramidal cells and averaged over 100 frames (7.23 Hz acquisition rate). In each of the four phases, a distinct set of four PMTs detects most of the fluorescence emissions, creating four active image tiles within the 4 × 4 array. (Each of the four PMTs corresponds to one of the four laser beams that is active in that phase.) To illustrate, the four active tiles within the phase I image are shaded with a different colour (shaded large square regions). However, close to the boundaries of each active tile, some fluorescence photons are detected by the other 12 PMTs. During signal unmixing these photons are reassigned to corresponding pixels in the correct adjacent active image tile. For instance, within the phase I image photons detected in the areas outlined in colour (rectangles and small squares) are reassigned to the colour-corresponding active tiles. **d**, An image compiling the four sets of four active image tiles from the panels in **c**. **e**, During signal unmixing, we re-assign scattered fluorescence photons to their correct pixels of origin, using the method shown in **c**, by reassigning the boundary regions of 128 pixels width. The resulting image is displayed with the mean contrast equalized across tiles. Scale bars: **c**, **e**, 500 µm.

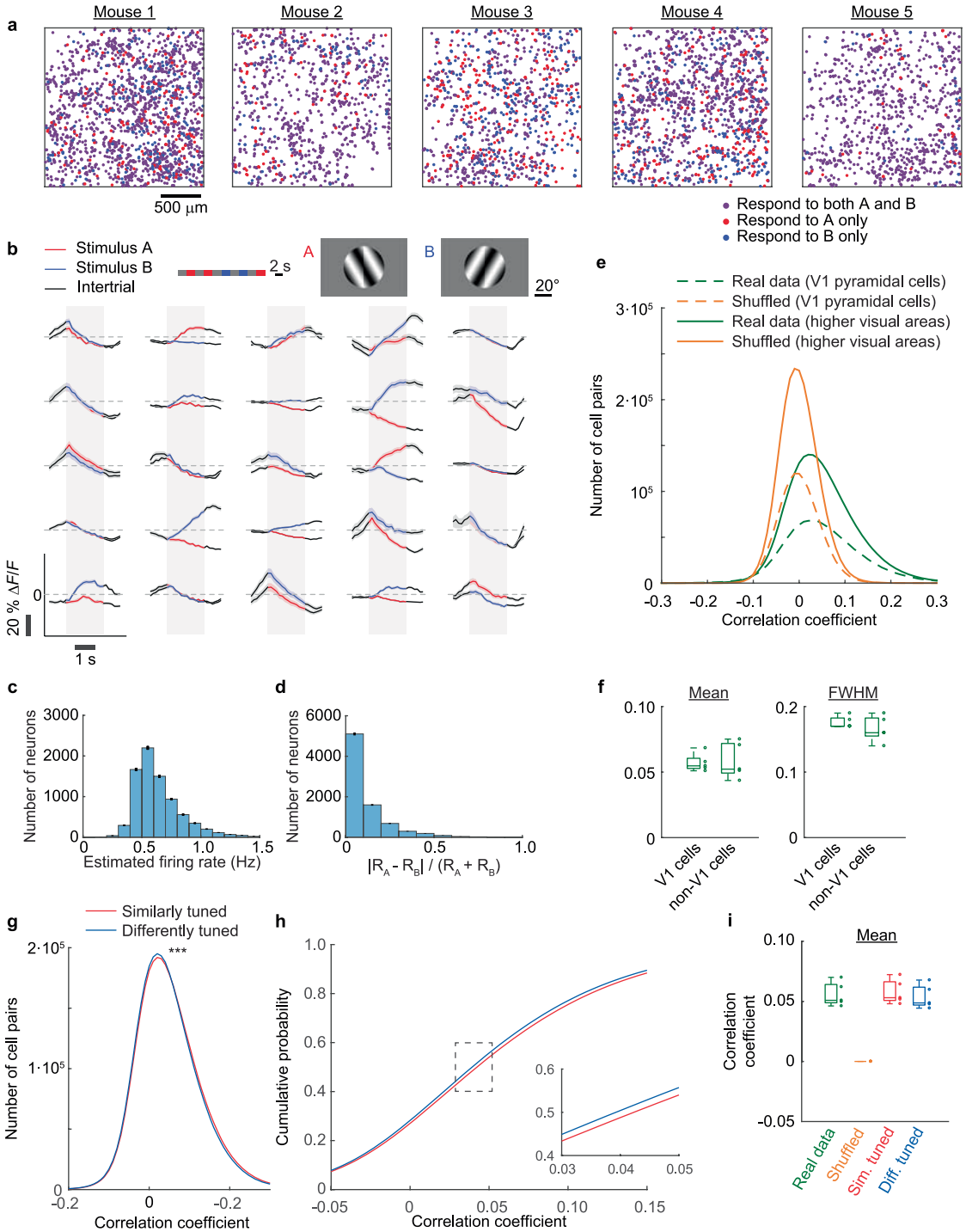


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Crosstalk un-mixing procedure for reconstructing the full field-of-view enables accurate estimation of neural activity traces. **a**, To quantify the extent of fluorescence scattering across image tiles, we acquired images in two distinct configurations that enabled us to distinguish fluorescence signals from any crosstalk due to fluorescence scattering across image tiles. Using an awake mouse expressing GCaMP6f in layer 2/3 cortical pyramidal cells, we first imaged with only one active laser beam and its corresponding PMT; the other 15 beams were blocked (configuration 1). In this configuration, there is no fluorescence scattering into the active image tile from the other 15 tiles, only the signals from the active tile. In configuration 2, we blocked the beam that had previously been active, unblocked the other 15 beams, operated the microscope with the normal multiplexing approach, and again sampled signals from all 16 PMTs. To estimate the extent of scattering into the tile with the blocked beam, we applied the computational un-mixing procedure to the raw image data. To estimate how much scattered fluorescence affects cell sorting, we first extracted individual cells and their Ca^{2+} activity traces from the first dataset, attained in configuration 1 without crosstalk. We then summed the images, frame by frame, from the two datasets, to create a mock dataset comprising unscattered plus scattered fluorescence signals, from which we again computationally extracted cells and their activity traces. This enabled a direct comparison between two datasets containing the exact same patterns of neural activity, with and without fluorescence scattering from other image tiles. **b**, Activity traces for four example cells, enabling comparisons of the Ca^{2+} activity traces (top), $\Delta F(t)/F_0$, and the resulting traces of the estimated spike counts (bottom), between the datasets with (red traces) and without (black traces) inter-tile scattering. The traces with and without inter-tile scattered fluorescence signals are nearly indistinguishable by eye. **c**, Histogram of the ratio of estimated spikes for the two datasets constructed in **a**, for all time bins (0.14 s per time bin) with an estimated spike count greater than 0.5. The mean ratio is 1.0 ± 0.06 (mean ± s.d.; $N = 31$ cells). Total number of time bins, 5,865. **d**–**g**, Studies of fluorescence scattering between the active image tiles in one temporal phase (Extended Data Fig. 2b) of the multiplexing scheme used for two-photon imaging. Throughout the paper, we corrected computationally for fluorescence scattering from active to inactive image tiles within each temporal phase of imaging (Extended Data Fig. 3c, Methods). This approach neglects the small amount of fluorescence scattering from active tiles to other active tiles, which in principle could also be computationally corrected using a more sophisticated method than the one we adopted. Hence, we examined experimentally the validity of our computational approach and the extent to which scattering between active tiles can be justifiably neglected. The amplitude of scattering between active tiles (**d**) varies with the location of each laser beam and its proximity to a tile boundary. We used fixed cortical tissue slices from adult *GCaMP6f-tTA-dCre* mice to measure the amplitude of such scattering effects when imaging at different depths within brain tissue. An image (**e**) of the spatial distribution of two-photon fluorescence excited 500 µm deep within a tissue slice shows that a majority of scattered fluorescence photons exits the brain tissue relatively near to the laser focus. By averaging over 100 different laser foci positions in each of 3 different brain slices, we determined the mean cross-sectional spatial profiles (**f**) of scattered fluorescence excited at different depths in tissue, as a function of the lateral displacement, x, from the laser focus. Profiles are shown normalized to unity at $x = 0$. The inset of **f** shows a magnified view of these cross-sectional profiles for $x \in [-1{,}000\ \mu m, -500\ \mu m]$, that is, up to 1 mm away from the laser focus. We used these empirically determined scattering profiles to compute the probability (mean ± s.d.; $N = 300$ laser focus positions) (**g**) that a fluorescence photon originating in one active image tile would scatter into an adjacent active tile. Even when the laser focus is on the boundary of an image tile, this probability remains less than 0.02 for all tissue depths ≤ 600 µm. For our studies of layer 2/3 cortical pyramidal cells in live mice, the probability of a fluorescence photon scattering between active tiles is less than 0.01. In conclusion, computational corrections for fluorescence scattering that account solely for scattering from active to inactive tiles—and neglect scattering between different active tiles—are empirically well justified.

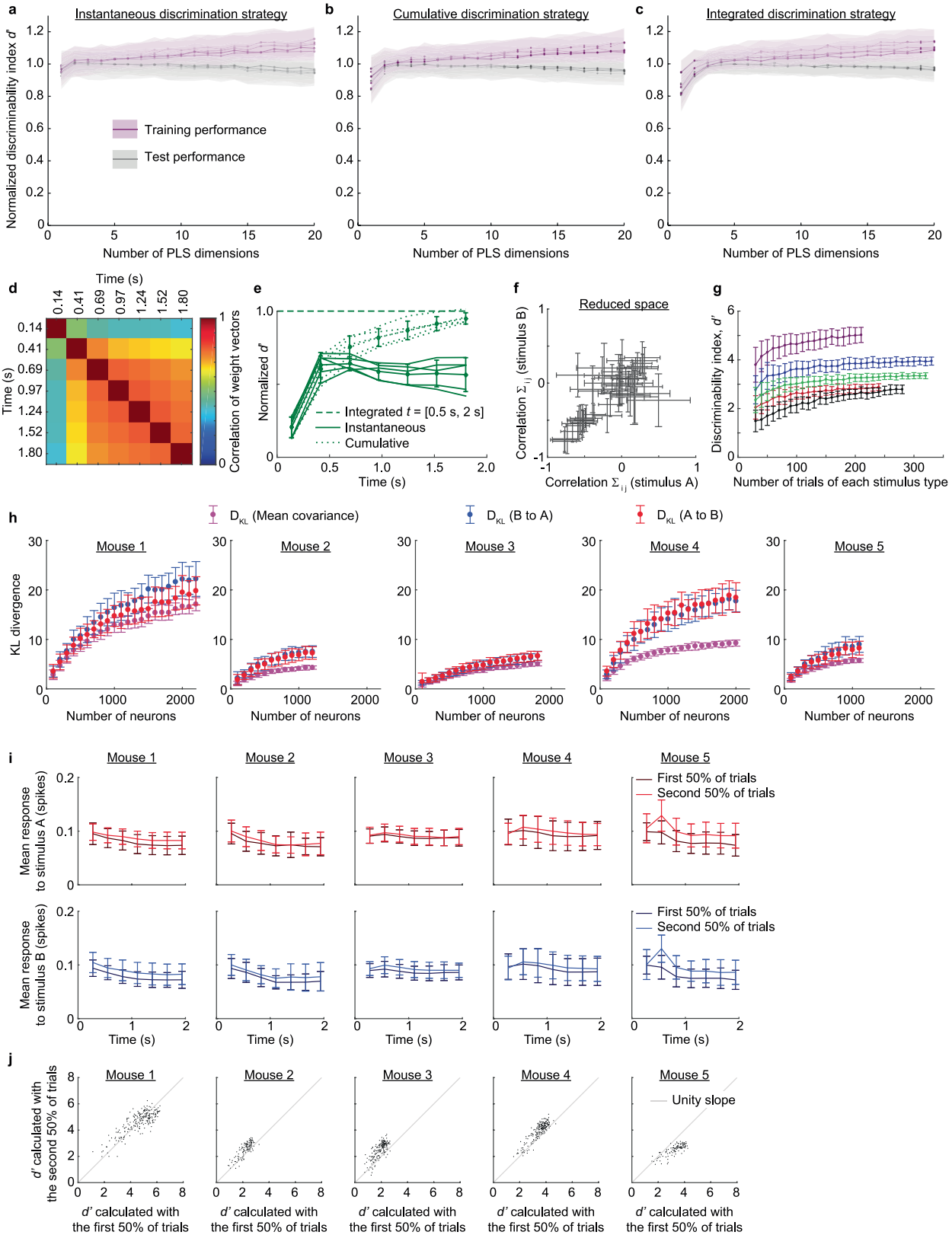


Extended Data Fig. 5 | Pipeline of offline data processing and procedures for reducing the dimensionality of the neural ensemble activity data and calculating the decoding accuracy. a, Pipeline of the offline procedures we applied to the acquired fluorescence signals to attain traces of neural activity. Steps coloured purple involve algorithms that use raw or processed image data. Steps coloured yellow involve algorithms that use cells' spatial filters as their input arguments. Steps coloured green involve algorithms that use cells' activity traces as their inputs. Purple steps, starting from the raw photocurrents from each of the 16 PMTs (sampled at 50 MHz and assigned to individual image pixels corresponding to a 400-ns laser dwell time), we normalized the photocurrent signals by the gain of each individual PMT, to equalize the image intensity scale across the entire image. We then un-mixed scattered fluorescence, as shown in Extended Data Fig. 3, and applied an image registration routine (TurboReg[44]) to the videos from the individual image tiles. To highlight Ca^{2+} transients against baseline fluctuations, we used the fact that the two-photon fluorescence increases of GCaMP6 during Ca^{2+} transients are many times the s.d. of background noise. Thus, we converted the fluorescence trace of each pixel, $F(t)$, into a trace of z-scores, $\Delta F(t)/\sigma$. Here $\Delta F(t) = F(t) - F_0$ denotes the deviation of the pixel from its mean value, F_0, and σ denotes the background noise of the pixel, which we estimated by taking the minimum of all standard deviation values calculated within a sliding 10-s window[35]. After transforming the movie data into this $\Delta F(t)/\sigma$ form, we identified neural cell bodies and processes using an established cell-sorting algorithm that sequentially applies principal and independent component analyses (PCA and ICA) to extract the spatial filters and time traces of individual cells[48]. Yellow steps, for all spatial filters corresponding to individual cell bodies, we thresholded the filters at 5% of each filter's maximum intensity and set to zero any filter components with non-zero weights outside the soma. To attain neural activity traces, we then reapplied the set of resulting filters to the $\Delta F(t)/F_0$ movies. Green steps, to estimate the most likely number of spikes fired by each cell in each time bin, we applied a fast non-negative deconvolution algorithm to the $\Delta F/F_0$ trace of the cell[49]. For each neuron, we down-sampled (2×) the activity traces to time bins of 0.275 s by averaging the values within adjacent time bins. To make comparisons across similar behavioural states, we removed all trials during which the mouse was moving. **b**, Neural responses for each visual stimulus (A and B) are represented as matrices of size $N_{neurons} \times N_{trials} \times N_{time\,bins}$. To calculate the accuracy of stimulus discrimination, we first randomly chose a subset of neurons from the dataset. For decoding using the 'instantaneous' strategy (Fig. 3, Extended Data Figs. 7–10), we then chose a specific time bin, whereas for the 'cumulative' decoding strategy we treated all the different time bins up to a specific time, t, as independent dimensions of the population activity vector. We then split the trials in half, into a training set and a test set, each with equal numbers of trials with the A and B stimuli. We took the neural activity traces in the training set and normalized them by the s.d. of the cell's activity about its mean, to create to a set of z-score traces. We then performed PLS analysis to identify a low-dimensional basis that well captured the separation between the neural responses to the two sensory stimuli. Using the activity data in the test set, we applied the same normalization and dimensional reduction procedures and values as for the training set. We used the resulting distributions of responses to calculate d' values and the eigenvectors of the noise covariance matrix. For each mouse we repeated this entire procedure for 100 different randomly chosen subsets of neurons.



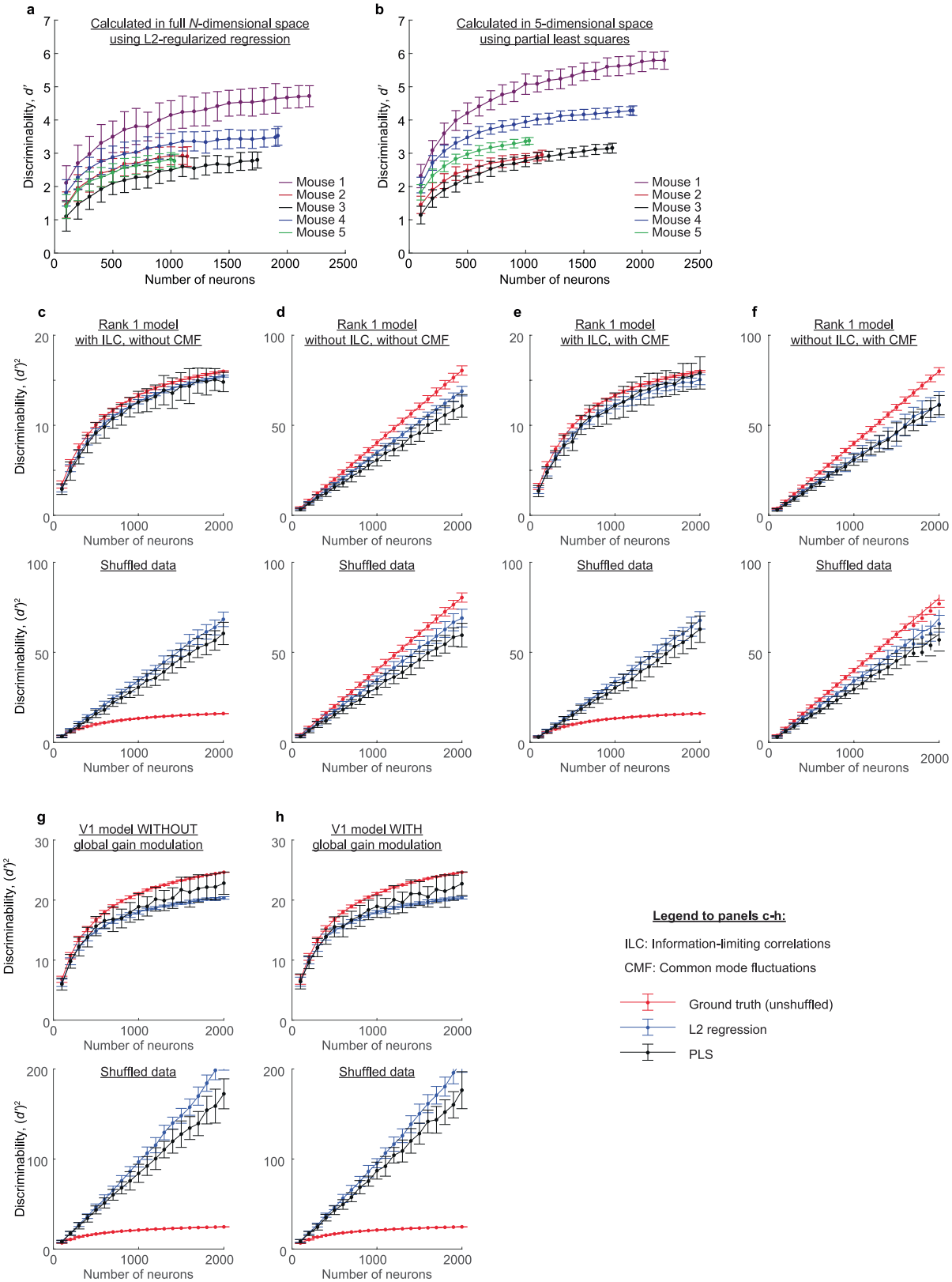
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Distributions of pairwise noise correlation coefficients do not differ significantly between pyramidal neurons in area V1 and higher-order visual areas. a, Anatomical maps of visual cortical neurons that responded to each of the two stimuli. For these maps (but for no other analyses in the paper), we denoted a cell as responsive to one of the stimuli if, in at least one time bin during the 2-s-stimulation period (0.275 s per bin), the difference between the cell's mean response and its mean activity trace during the inter-trial intervals was more than twice the sum of the s.e.m. values for these two traces. Cells that responded to stimulus A only are shown red, those that responded only to stimulus B only are shown blue, those that responded to both stimuli are shown purple. **b**, Mean Ca^{2+} responses ($\Delta F/F$) of 25 example neurons to the two different moving grating stimuli, oriented at ± 30°. Ca^{2+} activity traces are shown coloured during the stimulation period (marked with light grey shading) and black otherwise. Coloured shading about each trace denotes the s.e.m. over 217 trials of each type. The inset shows a schematic of the two stimuli, which appeared for 2 s per trial and were presented in random order. **c**, **d**, Histograms of the estimated mean spiking rates of individuals neurons during visual stimulation (**c**) and the absolute values of the differential responses of the individual neurons to the two visual stimuli, $|R_A - R_B| / (R_A + R_B)$ (**d**), where R_A and R_B denote the mean responses of a cell to stimuli A and B, respectively. The distributions of cells' activity rates and preferences for one stimulus over the other were consistent with previous studies of rodent visual cortical neurons[28,29,38,58,59]. Data shown are for $N = 8{,}029$ individual cells from $N = 5$ mice. Error bars are s.d. as estimated on the basis of counting errors. **e**, Histogram of noise correlation coefficients, r, between pairs of layer 2/3 pyramidal neurons, computed as in Fig. 2d, for V1 cell pairs (dashed lines) and cells pairs in higher-order visual areas (solid lines). The histograms show mean values across the two different visual stimuli for both the real neural activity traces, and for trial-shuffled data in which each cell's responses to each stimulus presentation were randomly permuted across the set of all presentations of the same stimulus. r values were computed on the basis of cells' responses integrated over $t = [0.5\text{ s}, 2\text{ s}]$ from the start of each trial. Histogram bin, 0.01. ($N = 1{,}331{,}109$ V1 cell pairs from 5 mice; $N = 2{,}428{,}437$ cell pairs from higher-order visual areas in 5 mice). **f**, Box-and-whisker plots of the mean and FWHM values of the distributions in **e** (real data only). Both statistical metrics are similar for the two classes of visual cortical neurons. Open circles denote individual data points for $N = 5$ mice. **g**, **h**, Histograms (**g**) and cumulative probability distributions (**h**) of noise correlation coefficients for all cell pairs (based on all recorded V1 and higher-order visual cortical neurons) with similar or differently tuned mean evoked responses to the two visual stimuli. Unlike Fig. 2e, which shows these distributions for only the most active cells (the highest decile), here the distributions include all cell pairs with either positively (red curves) or negatively (blue curves) correlated mean responses to the two stimuli. Within these two groups of cell pairs, we computed the noise correlation coefficient, r, for each cell pair. Owing to the extremely large number of cell pairs, the two distributions of r values differed significantly (***$P < 10^{-13}$ for all 5 individual mice; two-tailed Kolmogorov–Smirnov test; 3,482,186 positively correlated cell pairs in total; 3,464,094 negatively correlated pairs), even though the effect size was tiny and the two distributions were nearly identical. This result shows the difficulty of detecting information-limiting correlations by measuring pairwise noise correlations, because the variance in the individual r values is much greater than the difference between the mean values of the two distributions. **i**, Box-and-whisker plots of the mean values of the correlation coefficients in **g**, **h**. Open circles mark individual data points for $N = 5$ mice. **b**–**i** are based on 217–332 trials per stimulus condition in each of 5 mice. In **f**, **i**, boxes cover the middle 50% of values, horizontal lines denote medians, and whiskers span the full range of the data.



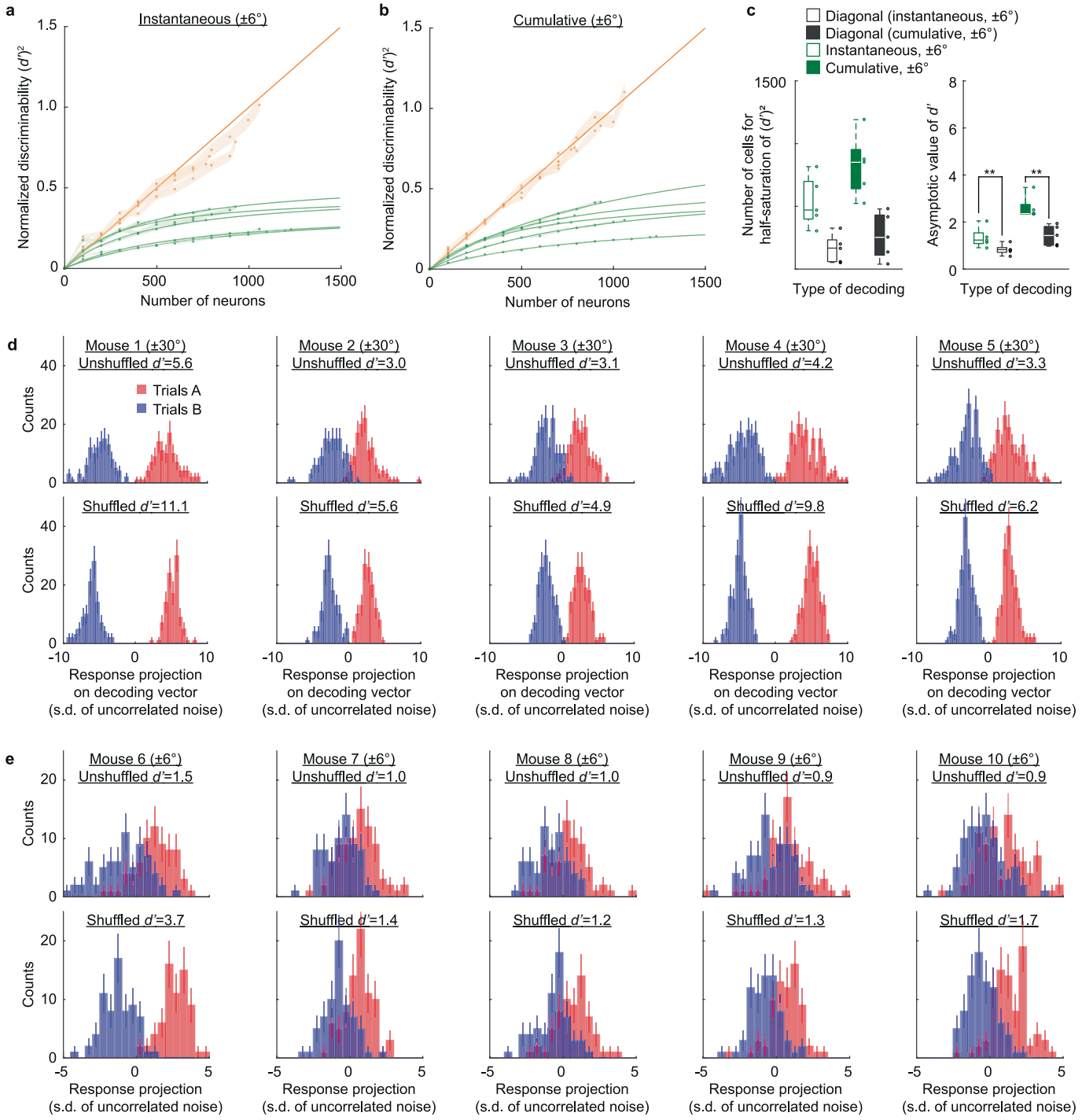
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Temporal integration of neural activity improves decoding performance, but quadratic and linear decoding yield identical biological conclusions. a–**c**, To identify how many PLS dimensions were needed to determine d' accurately, we divided data from each of 5 mice into three equally sized portions. We performed PLS analysis using trials in the first third. Onto the PLS dimensions thereby identified, we projected the neural ensemble activity in the second third of the data (training data). We retained only the first N_R dimensions of this projection and computed d' in the reduced space (magenta data points) by identifying a hyperplane for optimal stimulus discrimination. Finally, we applied this discrimination strategy to the remaining third of the data (test data) and again calculated d' (grey points). Plots show mean values of d' as a function of N_R for the interval [0.83 s, 1.11 s] from stimulus onset ($N = 5$ mice; error bars denote s.d. across 100 different subsets of 1,000 neurons per mouse). We normalized d' values to that found for $N_R = 5$ on the test dataset. For $N_R > 5$, discrimination performance declines owing to overfitting for all discrimination strategies: instantaneous (**a**), cumulative (**b**) and integrated (**c**). Hence, throughout the rest of the study we used $N_R = 5$ for all calculations of d'. **d**, Pearson correlation coefficients between the optimal linear decoding weights attained using instantaneous decoding at different time bins after the onset of grating stimuli (±30° orientations). These weights were highly correlated for different time bins, especially across the interval [0.5 s, 2 s], during which d' reaches a plateau. Further, optimal decoders for each time bin yielded nearly equivalent decoding performance when applied to data from other time bins. For instance, the optimal decoder for the fourth time bin ($t = 0.97$ s), when applied to any other of the last five time bins, yielded a performance within less than 2% of that of the optimal instantaneous decoder in all mice. When applied to the first and second time bins, the decoder from the fourth time bin yielded decoding performances that were, respectively, 83 ± 11% and 90 ± 3% (mean ± s.d.; $N = 5$ mice; 217–232 trials per stimulus) of that of the optimal decoders. **e**, Plots of d' versus time after stimulus onset, for instantaneous and cumulative decoding strategies (Fig. 3). For each mouse that viewed gratings oriented at ±30°, we chose 100 random subsets of 1,000 cells and normalized d' values by those obtained using a time-integrated decoding strategy, which involved optimal linear discrimination over one interval, [0.28 s, 1.94 s], covering most of the visual stimulation period. Green traces, mean d' values for individual mice using a time bin of 275 ms. Error bars, s.d. across 5 mice. **f**, In the five-dimensional space used after truncating ensemble neural responses to the five leading PLS dimensions, the distributions of noise in the responses to the two stimuli were highly similar. Specifically, non-diagonal elements, Σ_{ij}, of the noise covariance matrices for the two stimulus conditions were highly correlated (r: 0.81 ± 0.16; mean ± s.d.; $N = 5$ mice), as computed for the interval [0.83 s, 1.11 s] after stimulus onset. This similarity argues that a linear discrimination strategy to classify the two sets of ensemble neural responses is near optimal, as confirmed in **h**. Values of Σ_{ij} are plotted as mean ± s.d., computed across 100 different randomly chosen subsets of 1,000 neurons per mouse. **g**, Using optimal linear decoding, d' values saturated as the number of trials analysed increased. Colours denote individual mice. Data points were calculated for the interval [0.83 s, 1.11 s] after stimulus onset. Error bars, s.d. across 100 different randomly chosen subsets of 1,000 cells per mouse and stimulation trials. **h**, To check whether our results depended on our use of linear decoding, we tested whether quadratic decoding might yield different conclusions. We examined the KL divergence[31], a generalization of $(d')^2$ that makes no assumption about the statistical distributions under consideration. We computed the KL divergence, which equals $(d')^2$ for linear decoders, by using Gaussian approximations to the distributions of ensemble neural responses to the two different stimuli, and we plotted the results as a function of the number of cells, n, in the ensemble. First, to recapitulate our determinations of $(d')^2$ (magenta data points), we computed the KL divergence under the assumption the two different response distributions had distinct means but identical noise covariance matrices, which we estimated as the mean noise covariance matrix averaged over the two different stimulus conditions. This is equivalent to computing $(d')^2$. Next, we relaxed the assumption that the two noise covariance matrices were equal and computed the KL divergence between the distributions of neural responses to stimulus B relative to those to stimulus A (blue points), and vice versa (red points) (Methods). For all mice, KL divergence values saturated with increasing n and, except in one mouse, were not much larger than $(d')^2$ values. Thus, quadratic decoders (which are optimal for discriminating two Gaussian distributions with different means and covariances) will yield the same basic conclusions as linear decoders (which are optimal for discriminating two Gaussian distributions with the same covariance matrix). Data points and error bars denote mean ± s.d. values computed in each mouse across 50 different randomly chosen subsets of cells and assignments of visual stimulation trials to decoder training and testing (Extended Data Fig. 5b). **i**, Mean neural responses, averaged across all cells, to stimuli A (top) and B (bottom) for the first and second halves of the experimental trials in each mouse. Error bars, s.d. across the set of trials. **j**, d' values computed for each mouse using instantaneous decoders trained on the first half of the trials and tested on the second half (x axis), plotted with d' values for an instantaneous decoder trained on the second half of the trials and tested on the first half (y axis). **a**–**j** are based on 217–332 trials per stimulus condition in each of 5 mice.



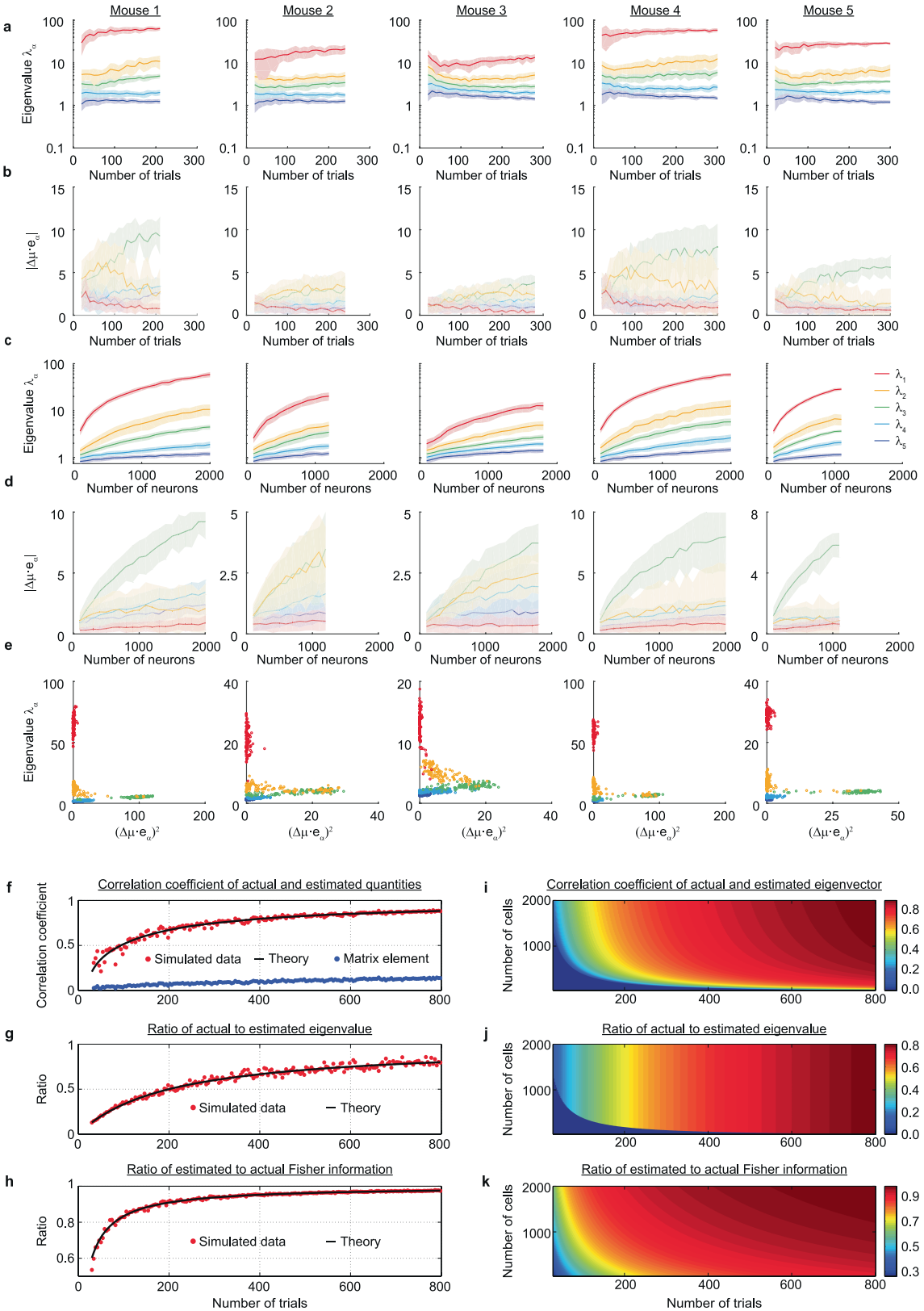
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | PLS-based decoding methods are robust to multiplicative gain modulation and common mode fluctuations in the neural ensemble dynamics and yield identical conclusions to regularized regression. a, **b**, To test whether PLS analysis and dimensionality reduction might lead to underestimates of d', we compared d' values determined using an L2-regularized regression (L2RR) performed in the full space of neural responses (**a**) to those found by PLS analysis (**b**). The two methods yielded similar estimates of d', which both saturated with increasing numbers of neurons. Plots show d' values (mean ± s.d.) for neural responses within [0.83 s, 1.11 s] after stimulus onset, computed across 100 different randomly chosen subsets of neurons and visual stimulation trials (Extended Data Fig. 5b). For PLS analyses, we used half of the trials in each subset for decoder training and the other half for testing. For L2RR we used 90% of the trials in each subset to determine the regression vector and the other 10% to determine d'. We varied the regularization parameter, k, within $[1, 10^5]$ and used the maximum d' value so obtained, as determined independently for each mouse, subset of neurons, and subset of trials (217–332 trials per stimulus condition in each of 5 mice). **c**–**h**, The conclusions of our study depend on comparisons of decoding performance between real and trial-shuffled datasets. Thus, we checked whether our PLS-based decoding methods would robustly detect information-limiting correlations in models in which such correlations were present but weak; avoid reporting information-limiting correlations in models lacking such correlations; and be robust to the potential presence of other strong sources of neural trial-to-trial variability—such as common mode fluctuations and multiplicative gain modulation—even when they make an order-of-magnitude greater contribution to neural variability than the information-limiting noise fluctuations. We studied these issues using two different computational models (Methods). For both models we plotted empirically determined $(d')^2$ values as a function of the number of neurons in the ensemble. We compared determinations of $(d')^2$ using PLS-based decoding and those made using L2RR to the actual ground truth values of $(d')^2$ in each model. In each panel, the top and bottom plots show results for unshuffled and trial-shuffled datasets, respectively. Data points and error bars denote mean ± s.d. values across 30 different simulations. To examine the combined effects of information-limiting noise correlations and common mode fluctuations (**c**–**f**) we studied a model of neural ensemble responses in which the noise covariance matrix exhibited information-limiting noise correlations via a single eigenvector **f**, the eigenvalue of which grew linearly with the number of cells in the ensemble. In addition to this rank 1 component, we included a noise term that was uncorrelated between different cells, as well as a common mode fluctuation, yielding a noise covariance matrix with the form $\Sigma^* = \sigma^2 I + \varepsilon_{common} J + \varepsilon \mathbf{f}^T \mathbf{f}$, where $\sigma^2 = 1$ is the amplitude of uncorrelated noise, I is the identity matrix, J is a rank 1 matrix of all ones, reflecting a common mode fluctuation, and **f** is the information-limiting direction, a vector that we chose randomly in each individual simulation from a multi-dimensional Gaussian distribution with unity variance in each dimension. The amplitude of information-limiting correlations was $\varepsilon = 0.002$, approximately matching the level observed in the experimental data. We chose the difference in the means of the two stimulus response distributions, $\Delta\boldsymbol{\mu}$, to be aligned with **f** (Fig. 3a) and to have a magnitude of 0.2 so that the asymptotic value of d' for large numbers of cells approximately matched that of the data. We compared decoding results attained with and without the presence of the common mode fluctuations in the neural responses. In the version of the model without common mode fluctuations, we set ε_{common} to zero. In this case (**c**) both PLS- and L2RR-based decoders correctly detected the saturation of information in the real data but not in trial-shuffled datasets. (See Extended Data Fig. 10h, k for theoretical results showing how the accuracy of d' estimates from PLS analysis depends on the numbers of neurons and experimental trials in this particular model.) To verify that our methods would not incorrectly report an information saturation when it was in fact absent, we next set $\varepsilon = 0$ and confirmed that in the absence of information-limiting noise correlations (**d**), neither decoder detected a saturation of information in the real or shuffled data. In the version of the model with common mode fluctuations, we set $\varepsilon_{common} = 0.02$, ten times the value of $\varepsilon = 0.002$. In this case (**e**), both PLS- and L2RR-based decoders correctly detected the information saturation in the real but not in the shuffled data. To verify that common mode fluctuations alone cannot induce an illusory saturation of information (**f**), we set $\varepsilon = 0$ while maintaining $\varepsilon_{common} = 0.02$ and confirmed that neither PLS- nor L2RR-based decoders reported an illusory information saturation. Overall, these results indicate that our methods accurately detect the presence of weak information-limiting correlations buried within common mode noise that can be an order of magnitude larger, without falsely detecting information-limiting correlations when they are absent. To study the possible effects of multiplicative gain modulation (**g**, **h**), we compared two versions of a model in which the responses of the V1 neural population either were or were not subject to a multiplicative stochastic gain modulation but were otherwise statistically equivalent. We modelled the V1 cell population as a set of Gabor filters (see Appendix section 5). In the model version with gain modulation, on each visual stimulation trial we multiplied the output of each Gabor filter by a randomly chosen factor, uniformly distributed between 50%–150%, the value of which was the same for all cells but varied from trial to trial. In the model version without gain modulation (**g**) both PLS- and L2RR-based decoders detected the information saturation in the real but not in the trial-shuffled datasets. When we added global gain modulation to the model (**h**) both decoders correctly found the information saturation in the real but not in the shuffled datasets.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Moving grating visual stimuli oriented at ±6° are harder to distinguish on the basis of their evoked neural ensemble responses than gratings oriented at ±30°, but also reveal the saturation of information signalling in large neural populations. **a**, $(d')^2$ values determined using an 'instantaneous' decoder for the interval [0.70 s, 0.94 s] from visual stimulation onset, plotted as a function of the number of cells, n, in the ensemble in mice presented moving gratings oriented at ±6°. Data points represent mean values determined across 100 different subsets of cells, and the shading represents s.e.m. As in Fig. 3f, g, we fit the $(d')^2$ values as a function of n using a one-parameter fit, $(d')^2 = (d')^2_{shuffled}/(1 + \varepsilon \times n)$, where $(d')^2_{shuffled}(n)$ is the empirically determined value of $(d')^2$ for the same number of cells in the shuffled data, and ε is the fit parameter. For each mouse, for both real and trial-shuffled data we normalized $(d')^2$ values by the value of $(d'_{shuffled})^2$ for $n = 1{,}000$ neurons. Goodness of fit: $R^2 = 0.41 \pm 0.17$ (s.d). $N = 5$ mice. $\varepsilon = 0.0021 \pm 0.0008$ (s.d.), 122–167 trials per stimulus condition for each mouse. **b**, Same as **a**, but using the 'cumulative' decoding strategy over the [0 s, 0.94 s] time interval. **c**, Box-and-whisker plots of the asymptotic values of d' in the limit of many neurons (right) and the number of cells at which $(d')^2$ attains half its asymptotic value (left) as determined from parametric fits to the data of **a** and **b** for the instantaneous (open boxes) and cumulative (filled boxes) decoding strategies. Optimal linear decoders (green data) slightly but significantly outperformed diagonal decoders (black data) (**$P < 0.0001$; one-tailed Wilcoxon rank sum test; $N = 100$ different randomly chosen assignments of trials to decoder training and test sets in each mouse; 122–167 trials per stimulus condition for each mouse; open circles denote mean values from $N = 5$ individual mice). **d**, **e**, Histograms for the real (unshuffled) and shuffled datasets of the ensemble neural responses to each of the two visual stimuli, projected onto the direction of the optimal decoding vector determined by PLS analysis, as computed in each mouse viewing moving gratings oriented either at ±30° (**d**) or ±6° (**e**), using all imaged neurons and the instantaneous decoding approach. Error bars denote counting errors. Values on the x axes are plotted for each mouse in units of the s.d. of its neural ensemble responses along the decoding vector for the shuffled data. For each mouse, the histograms have approximately equal shapes for the two visual stimuli, are unimodal and approximately symmetric about their mean values, bolstering the use of linear decoding and d'. This analysis involved 217–232 trials per stimulus condition per mouse in **d** and 122–167 trials per stimulus condition per mouse in **e**.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Hundreds of experimental trials sufficed to estimate the statistical structure of signals and noise in visual cortical coding. **a**, **b**, PLS analysis represents ensemble neural responses in a low-dimensional subspace that helps for understanding visual discrimination (Fig. 4). On the basis of Extended Data Fig. 7a–c, computations here used the five most informative PLS dimensions. Each column shows results from an individual mouse that viewed gratings oriented at ±30° (217–332 trials per stimulus). Each colour denotes a different eigenvector, $\mathbf{e}_\alpha$, of the noise covariance matrix in the five-dimensional subspace. α denotes the dimension index, {1,2,3,4,5}. As illustrated in Fig. 4e, each mouse had multiple eigenvalues, λ_α, of the noise covariance matrix that increased with the number of cells, n, used for analysis. As shown in Fig. 4f, visual signals—defined as the mean separation, $\Delta\boldsymbol{\mu}$, between the two response distributions—also increased with n. **a**, **b** show eigenvalues λ_α (**a**) and signal components $|\Delta\boldsymbol{\mu}\cdot\mathbf{e}_\alpha|$ (**b**) plotted against the number of trials analysed. Both signal and noise estimates plateau, indicating that there were sufficient trials to accurately estimate signal and noise structure in the reduced five-dimensional space. Throughout **a**–**d**, lines and shading denote mean ± s.d. across 100 different randomly chosen subsets of cells and assignments of trials to decoder training and testing, except in **a**, **b** we used all cells from each mouse and 30 different assignments of trials. **c**, **d**, The statistical relationships between visual signals and noise show the largest noise mode is not information-limiting. Each mouse had multiple eigenvalues, λ_α, of the noise covariance matrix (**c**) that increased with n, the number of cells. Visual signals (**d**) also increased with n, as shown by decomposing $\Delta\boldsymbol{\mu}$ into components along the five eigenvectors, $\boldsymbol{e}_\alpha$. In every mouse the eigenvector with the largest eigenvalue, $\boldsymbol{e}_1$, was the least well aligned with the signals, $\Delta\boldsymbol{\mu}$ (compare red curves in **c**, **d**). **e**, Plots of noise values, computed as in **c**, versus signal values, computed as in **d**, based on all recorded neurons from each mouse and the same 100 subsets of data used in **c**, **d**. The largest noise mode (red points) was generally an order of magnitude greater than noise modes that limited neural ensemble signalling (green and yellow points). **f**–**k**, In **a**–**e** and throughout much of the paper, we analysed populations of up to 2,191 neurons using 217–332 trials with each stimulus, which sufficed to accurately determine the Fisher information, $(d')^2$, and principal eigenvectors of the noise covariance matrix (Fig. 4). By comparison, there were insufficient trials to accurately determine noise covariance matrix elements—that is, noise correlations between cell pairs (Fig. 2d). To explain this, we derived the accuracy with which d' and principal noise covariance eigenvectors and eigenvalues can be estimated through PLS analysis of recordings of n neurons across P trials, using the computational model of Extended Data Fig. 8c (Appendix section 6 has derivations of results in **f**–**k**). The central idea, illustrated in **f**, is that one can estimate accurately the principal noise covariance eigenvector, because it has a large eigenvalue, λ, that grows linearly with n ($\lambda \cong cn$, where c is a constant). The theory predicts that the correlation coefficient, $\mathcal{C}$, between estimated and actual eigenvectors is given by $\mathcal{C}^2 = \frac{cP-1/(cn)}{cP+1}$, for $c^2Pn > 1$. Otherwise, $\mathcal{C} = 0$. **f** shows predictions for $\mathcal{C}$ (black curve) versus the number of trials, P, for $n = 2{,}000$ and $c = 0.005$. We chose this c value to fall within the lower range of growth rates for experimentally determined eigenvalues, **c**. The predicted $\mathcal{C}$ values match those describing the accuracy (red points) with which we could estimate the principal noise covariance eigenvector in the computational model. However, correlation coefficients (blue points) between estimated and actual individual elements of the noise covariance matrix were unsatisfactory, even with 800 trials. **i** shows predicted values of $\mathcal{C}$ as a joint function of n and P. Iso-contours of $\mathcal{C}$ are hyperbolic, revealing a tradeoff such that recording more cells enables accurate estimation of noise eigenvectors using fewer trials. We also derived how accurately one can estimate eigenvalues of the noise covariance matrix, as quantified using the ratio, $\mathfrak{R}_\lambda = \lambda/\hat{\lambda}$ where $\lambda = cn$ is the actual eigenvalue in the model and $\hat{\lambda}$ is the estimate based on P trials. The theory predicts $\mathfrak{R}_\lambda = \frac{cP}{cP+1}$ when $c^2Pn > 1$; otherwise we set $\mathfrak{R}_\lambda = 0$, because we cannot accurately estimate the corresponding eigenvector when $c^2Pn < 1$. **g** plots predictions of $\mathfrak{R}_\lambda$ (black curve) versus P (for $n = 2{,}000$ cells and $c = 0.005$), which match the accuracy with which we estimated the model eigenvalues from simulated data (red dots). **j** shows $\mathfrak{R}_\lambda$ predictions as a joint function of n and P. We also studied how well one can estimate the Fisher information, $(d')^2$, via PLS analysis of data with fewer trials than recorded neurons. We examined the ratio, $\mathfrak{R}$, of the d' estimate to its actual value using the model and simulated data of Extended Data Fig. 8c and found $\mathfrak{R}^2 = \frac{1+n\varepsilon}{1/\mathcal{C}^2_{\mathrm{PLS}}+n\varepsilon}$, where $\mathcal{C}^2_{\mathrm{PLS}} = \frac{\Delta s^2 P + 4(\varepsilon+1/n)}{\Delta s^2 P + 4(\varepsilon+1)}$ is the predicted correlation coefficient between the PLS regression vector and the optimal one. Here Δs^2 and ε determine the Fisher information in the model of Extended Data Fig. 8c via $(d'_{\mathrm{opt}})^2 = \frac{n\Delta s^2}{1+n\varepsilon}$. As in Extended Data Fig. 8c, we used $\varepsilon = 0.002$ to match the growth rate of $(d')^2$ in experimental data with increasing n, and $\Delta s^2 = 0.04$ to approximate the magnitude, $\frac{\Delta s^2}{\varepsilon}$, of $(d')^2$ in the data for large n. $\mathcal{C}^2_{\mathrm{PLS}}$ increases monotonically with P and n, confirming that PLS regression improves as n and P increase. As $\mathcal{C}^2_{PLS}$ nears 1, so does $\mathfrak{R}^2$, indicating that PLS analysis can accurately estimate $(d')^2$. **h** shows predictions for $\mathfrak{R}^2$ versus P for $n = 2{,}000$ cells (black curve). The theory matches the accuracy with which we estimated $(d')^2$ via PLS analyses of the simulated model data (red dots). **k** shows predicted $\mathfrak{R}^2$ values versus n and P. Iso-contours of $\mathfrak{R}^2$ are hyperbolic, indicating recordings of more neurons permit accurate estimates of $(d')^2$ based on fewer trials.

nature research

Corresponding author(s): M.J. Schnitzer, O. Rumyantsev, S. Ganguli

Last updated by author(s): Dec 27, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For data acquisition, we used custom routines written in LabView software (National Instruments, version 2012 SP1, 32 bit). For instrument control in the pixel-multiplexing acquisition modes, we also used open-source ScanImage software (version 3.8).
Data analysis	We used open source software routines for image registration (http://bigwww.epfl.ch/thevenaz/turboreg/), cell sorting, and partial least squares analysis (https://www.mathworks.com/matlabcentral/fileexchange/18760-partial-least-squares-and discriminant-analysis). Software code for extracting individual neurons and their calcium activity traces from calcium videos by using principal component and then independent component analyses is freely available (https://www.mathworks.com/matlabcentral/ fileexchange/25405-emukamel-cellsort), although for convenience we used a commercial version of these routines (Mosaic software, version 0.99.17; Inscopix Inc.). We wrote all other analysis software in Matlab (2017b). The primary software code used to support the findings of the study is available at Zenodo.org (https://zenodo.org/record/3593520#.XgWPu-hKg2w).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We designed the study such that each of the main results would come from 5 biological replicates, i.e., 5 different mice, for each of 2 different experimental conditions. The experimental results from all 10 mice were similar, affording confidence in the findings.
Data exclusions	To minimize false positives during cell sorting, we adopted a conservative approach during manual classification of candidate cells, such that we accepted for analysis only those candidates whose spatial forms and temporal dynamics were plainly those of neurons. Due to the known modulatory effects of locomotion on mouse visual processing, we used for analysis only those experimental trials during which the mice had no locomotor activity.
Replication	We reproduced the main results of our study across 10 different mice under 2 different experimental conditions (5 mice in each group).
Randomization	We split the datasets randomly into training and test subsets (usually 50% each; see Methods for details). We determined measured quantities by averaging across multiple realizations of such a split. For analysis performed on subsets of neurons, we chose subsets randomly and averaged results across multiple subsets. See Methods and Extended Data Fig. 5b for details.
Blinding	All animals in the experiment belonged to the same experimental group, so blinding was neither needed nor feasible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	We immunostained tissue sections with antibodies against glial fibrillary activation protein (1:2500 dilution, rabbit anti-GFAP, Sigma HPA056030, Lot C115616) and heat shock protein 70 (1:400 dilution, mouse anti-HSP, Enzo ADI-SPA-810, Clone C92F3A-5, Lot 01031912) and then applied fluorophore-conjugated secondary antibodies (goat anti-rabbit-Alexa 594 [Invitrogen, A-11012, Lot 1933366] and goat anti-mouse-Alexa 488 [Invitrogen, A-11001, Lot 56881A]).
Validation	We performed positive control experiments to validate the abilities of these antibodies to detect laser-induced tissue damage (Extended Data Fig. 2g).

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	We analyzed data acquired from 6 male and 4 female Ai93 triple transgenic GCaMP6f-tTA-dCre mice from the Allen Institute (Rasgrf2-2A-dCre/CaMK2a-tTA/Ai93), which expressed the calcium indicator GCaMP6f in layer 2/3 pyramidal cells. Mice were 12-17 weeks of age when we implanted the cranial window in preparation for brain imaging. For illustrative purposes only, we imaged a single tetO-GCaMP6s/CaMK2a-tTA mouse42, which expressed the calcium indicator GCaMP6s in a subset of neocortical pyramidal neurons (Supplementary Video 3).

Wild animals	No wild animals were used.
Field-collected samples	There were no field-collected samples.
Ethics oversight	The Stanford University APLAC approved all procedures involving animals.

Note that full information on the approval of the study protocol must also be provided in the manuscript.