Supporting Information
Sharpee et al. 10.1073/pnas.1217479110

Sharpee et al. 1073/pnas.
Sharpe

SI Text

Estimation of Linear–Nonlinear Models

Standard linear–nonlinear (LN) models with no position invariance were fitted using maximum likelihood with a flexible nonlinear function (1, 2) relating the neural firing to stimulus components along the maximally informative dimensions. The method begins by computing the spike-triggered average. After that the profile of the relevant stimulus feature is refined to increase predictive power of the model (either according to the mutual information between the stimuli and the neural responses or the likelihood of the model given the data). The optimization is based on a combination of simulated annealing and gradient descent following the procedure described in ref. 1. As a result, all stimulus correlations that are not associated with the neural response are removed, because they do not provide additional information about the neural response.

LN models with partial invariance can be estimated by a twostep maximum likelihood optimization (3) that iteratively updated both the relevant features and the locations that were most likely to have triggered a spike. Because of the large number of models that needed to be estimated for each neuron (12 ranges of position invariance at two different scales, each computation repeated for different jackknife estimates; see below), we have simplified the second step in the calculation by estimating the locations that were most likely to have triggered a spike just once, using the image frame that elicited the maximal number of spikes as the relevant feature. In cases in which more than one image elicited the maximal firing rate, these images were aligned to each other to maximize their dot product and averaged, producing an estimate of the "optimal image." This simplification of the method is similar to the "dejittering" technique that was previously developed to minimize the effects of time and spatial jitter on the estimates of neural feature selectivity (4). We found that this simplified approach produced models with statistically the same predictive power (Fig. S8) compared with models obtained by full two-step likelihood optimization.

Models with unlimited range of invariance were fitted using the same maximum likelihood approach as models without translation invariance (1, 2), but with respect to the amplitudes of the 2D Fourier transform applied to stimulus frames (5).

- 1. Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: Maximally informative dimensions. Neural Comput 16(2):223–250.
- 2. Kouh M, Sharpee TO (2009) Estimating linear-nonlinear models using Renyi divergences. Network 20(2):49–68.
- 3. Eickenberg M, Rowekamp RJ, Kouh M, Sharpee TO (2012) Characterizing responses of translation-invariant neurons to natural stimuli: Maximally informative invariant dimensions. Neural Comput 24(9):2384–2421.
- 4. Dimitrov AG, Sheiko MA, Baker J, Yen SC (2009) Spatial and temporal jitter distort estimated functional properties of visual sensory neurons. J Comput Neurosci 27(3): 309–319.

Four jackknife estimates (6) were generated for each type of model using different subsets of the unrepeated dataset. This choice of the number of jackknives was motivated by the need to balance the size of the "training" and "validation" datasets. Larger numbers of jackknives reduce the size of the validation set, leading to greater overfitting biases in model estimation. At the same time, reasonable estimates of variance in receptive field estimates can be obtained with four jackknives. Results are reported as averages across jackknife estimates (in the case of models with limited range of position invariance, individual jackknives for each neuron were realigned to maximize correlation with each other). The (signed) signal-to-noise was computed as a SD scaled by a factor of 3 according to the jackknife correction (6).

Computation of Information Explained by an LN Model

The amount of information accounted for by an LN model with two relevant features, \vec{v}_1 and \vec{v}_2 , was computed by projecting all stimuli onto the relevant stimulus features and forming the probability distributions of these projections x_1 and x_2 , respectively, both across all of the presented stimuli $P_{\vec{v}_1,\vec{v}_2}(x_1,x_2)$ and across stimuli that elicited a spike $P_{\vec{v}_1, \vec{v}_2}(x_1, x_2 | split e)$. For the latter, conditional distribution, the contribution of each stimulus is proportional to the number of spikes it elicited (there is no need to binarize the spike trains; one can work with spike counts obtained in each bin of the spike train). The Kullback–Leibler distance between these two probability distributions $I(\vec{v}_1, \vec{v}_2)$ =

$$
\int dx_1 dx_2 P_{\vec{v}_1, \vec{v}_2}(x_1, x_2 | spike) \log_2 \left[\frac{P_{\vec{v}_1, \vec{v}_2}(x_1, x_2 | spike)}{P_{\vec{v}_1, \vec{v}_2}(x_1, x_2)} \right]
$$
 yields the amount of

information per spike that is accounted for by the reduced model.

Both the overall information and the information accounted by different estimated models contain a positive bias, which decreases as more data are collected (7–9). To correct for this bias, we followed procedures described in ref. 7. Specifically, we computed information values based on different fractions of repeats (80–100%) and then used a linear extrapolation to find values predicted if an infinite number of repetitions could be obtained. The linear dependence is expected on theoretical grounds in the asymptotic regime of sufficiently large dataset sizes (7–9). This procedure was used to correct for bias in both the information per spike and the information accounted for by each type of model.

- 5. David SV, Hayden BY, Gallant JL (2006) Spectral receptive field properties explain shape selectivity in area V4. J Neurophysiol 96(6):3492–3505.
- 6. Efron B, Tibshirani RJ (1998) An Introduction to Bootstrap (Chapman and Hall, New York).
- 7. Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000) Synergy in a neural code. Neural Comput 12(7):1531–1552.
- 8. Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and information in neural spike trains. Phys Rev Lett 80(1):197–200.
- 9. Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. Neural Comput 7(2):399–407.

Fig. S1. Schematic representation of LN models. In addition to the model shown in Fig. 1A, we also used and compared performance of models with the following structures. (A) The LN model based on one relevant stimulus feature and partial translation invariance. The range of translation invariance is determined by center positions of intermediate units whose responses are pooled according to logical OR or MAX operations. (B) The LN model based on one relevant stimulus feature and unlimited translation invariance. The relevant feature in this model describes the optimal weighting of Fourier amplitudes to maximally modulate the neural response. This corresponds to the relative weighting of orientation values for eliciting the neural responses, without taking into account the relative spatial positions of the corresponding edges within an image. (C) Same as B but for two relevant stimulus features. This model allows for potentially nonlinear interactions between two different linear combinations of Fourier amplitudes.

Fig. S2. (A) Absence of clustering in curvature preferences. Each circle corresponds to a different neuron. The circles are positioned at the estimated receptive field center position measured in degrees relative to the fixation point (marked by ×). The circle radius is given by 10 ρ ^{-0.4}, where ρ is the curvature measured relative to the frame size. This function was chosen for presentation purposes, so that values for different neurons can be visible on the same scale. (B) There is no dependence of eccentricity on the degree of invariance when the range of invariance is measured in units of receptive field size ($P = 0.0929$).

Fig. S3. Control analysis showing that the trade-off between curvature tuning and position invariance is not due to the estimation procedure. We analyzed a set of model cells that had the same preferred curvature regardless of the range of invariance. These model spike trains were subjected to identical analysis to how V4 data were analyzed, from receptive field estimation to receptive field fitting with curved Gabor models. The derived curvature values show no dependence on the range of position invariance ($P = 0.58$, linear correlation). The curvature index 1 is parameter ρ (Materials and Methods) normalized such that the frame size equals 1.

Fig. S4. There were no significant differences in the mean firing rate (A) for neurons best described by either a position-specific or position-invariant model $(P = 0.59)$. These models also yielded similar percentages of information explained (B) and variance (C), with corresponding P values of 0.91 and 0.34 (Mann–Whitney test).

Fig. S5. Using information per spike as another measure of sparseness in neural responses. Both the dependence on threshold (A) and threshold (B) mirror those obtained with sparseness quantified as normalized variance in the firing rate (Fig. 2). (A) The information per spike was lower for neurons exhibiting position invariance compared with neurons with partial position invariance (from 5% to 15%, P = 0.0037, Mann–Whitney test). The dependence of information per spike on invariance range was better described by an inverse quadratic function than a linear function: I_{spike} ~ (x+c)^{−2}, where x is the invariance range and c is a constant (solid line shows the best fit, $P = 0.039$, correlation between x and I_{spike}). (B) There was no correlation between threshold and information per spike ($P = 0.65$).

Fig. S6. The distribution of curvature in natural scenes. Curvature is calculated as the angular deviation of a triplet of adjacent oriented elements from the axes (dotted lines) that define the spatial layout of the triplet. Natural images were filtered with a Laplacian of Gaussian filter to extract an edge map. The spatial scale of the filter is defined by the SD (σ) of the Gaussian function. In this case $\sigma = 4$. The edge map was then analyzed using a hexagonal mosaic of circular apertures shown in the inset. Orientation energy underlying each circular aperture was extracted from the two-dimensional Fourier transform of the corresponding image segment. This was passed through a binary threshold to obtain a final map of significant orientations in the image patch. Patches were analyzed at three edge-orientation sampling scales defined by the center-to-center distance d between adjacent apertures (Inset). The frequency of occurrence is plotted as a function of curvature normalized to lie within the [0 1] interval. Values near 0 indicate low curvature, midrange values near 0.5 indicate high curvature, and values near 1 indicate texture-like configurations where the oriented elements deviate maximally from the backbone axes.

Fig. S7. Distribution of curvature preferences across the V4 population. The distribution of curvature values derived from MIID1 and MIID2 by fitting the curved Gabor model. These values represent the marginal distribution from data in Fig. 1D.

Fig. S8. (A) Example estimation of the relevant stimulus feature ("template") of a model cell with partial translation invariance. The position invariance of this model neuron was modeled by comparing the template to stimuli shifted by one pixel in horizontal and/or vertical dimensions (a total of nine shifted position). The model template (Upper) is compared with the estimated templates (Lower). The dejittering estimation algorithm (Materials and Methods) produces templates that can be off-centered (Lower Left) but yield good correlation coefficients (0.78) after realignment with the model template (Lower Right). (B) Across the population of V4 neurons, the dejittering algorithm yields models with the same predictive power as the two-step maximum likelihood optimization (3) ($P = 0.69$, linear correlation). We use the dejittering algorithm because it is computationally less intensive than the two-step maximum likelihood optimization and because of the large number of models that needed to be estimated for each neuron (144 models for each neuron across different ranges of invariance and different jackknives).