Systems/Circuits

A Convolutional Subunit Model for Neuronal Responses in Macaque V1

Brett Vintch,[1] J. Anthony Movshon,[1] and Eero P. Simoncelli[1,2,3]

[1]Center for Neural Science, [2]Courant Institute of Mathematical Sciences, and [3]Howard Hughes Medical Institute, New York University, New York, New York 10003

The response properties of neurons in the early stages of the visual system can be described using the rectified responses of a set of self-similar, spatially shifted linear filters. In macaque primary visual cortex (V1), simple cell responses can be captured with a single filter, whereas complex cells combine a set of filters, creating position invariance. These filters cannot be estimated using standard methods, such as spike-triggered averaging. Subspace methods like spike-triggered covariance can recover multiple filters but require substantial amounts of data, and recover an orthogonal basis for the subspace in which the filters reside, rather than the filters themselves. Here, we assume a linear-nonlinear-linear-nonlinear (LN-LN) cascade model in which the first LN stage consists of shifted ("convolutional") copies of a single filter, followed by a common instantaneous nonlinearity. We refer to these initial LN elements as the "subunits" of the receptive field, and we allow two independent sets of subunits, each with its own filter and nonlinearity. The second linear stage computes a weighted sum of the subunit responses and passes the result through a final instantaneous nonlinearity. We develop a procedure to directly fit this model to electrophysiological data. When fit to data from macaque V1, the subunit model significantly outperforms three alternatives in terms of cross-validated accuracy and efficiency, and provides a robust, biologically plausible account of receptive field structure for all cell types encountered in V1.

Key words: **model; receptive field; subunits; V1**

Significance Statement

We present a new subunit model for neurons in primary visual cortex that significantly outperforms three alternative models in terms of cross-validated accuracy and efficiency, and provides a robust and biologically plausible account of the receptive field structure in these neurons across the full spectrum of response properties.

Introduction

The stimulus selectivity of neurons in primary visual cortex (V1) is often described using oriented linear filters. Simple cells combine signals from afferents arriving from the lateral geniculate nucleus (LGN) (Hubel and Wiesel, 1962; Alonso et al., 2001); the resulting behavior can be modeled with a linear filter and a rectifying nonlinearity in an "LN" cascade (Jones and Palmer, 1987; Heeger, 1992). Complex cell responses can be described by summing the responses of a set of simple cells with identical orientation tuning, which differ in spatial position and/or phase (Hubel and Wiesel, 1962; Movshon et al., 1978; Adelson and Bergen, 1985).

The simple-complex distinction may not be a true dichotomy (Mechler and Ringach, 2002), and "subspace" characterization methods (Stevenick and Bialek, 1988; Brenner et al., 2000; Paninski, 2004; Sharpee et al., 2004; Simoncelli et al., 2004) have shown that the responses of individual V1 cells may be fit by a set of linear filters numbering from one to more than a dozen (Touryan et al., 2002; Rust et al., 2005; Chen et al., 2007). These models usefully extend traditional accounts of simple and complex cells, but their biological interpretation is unclear; the extracted filters are constrained by the analysis to be orthogonal and therefore span the relevant subspace with diverse shapes and positions. Moreover, although these orthogonal filters can be used to constrain a model based on spatially shifted (convolutional) filters (Rust et al., 2005; Lochmann et al., 2013), the initial characterization of the subspace has many more degrees of freedom in its parameterization and thus requires large amounts of data for accurate estimation.

Received July 1, 2013; revised Sept. 8, 2015; accepted Sept. 23, 2015.

Author contributions: B.V., J.A.M., and E.P.S. designed research; B.V. performed research; B.V. analyzed data; B.V., J.A.M., and E.P.S. wrote the paper.

This work was supported by National Institutes of Health Research Grants EY04440 and EY22428 to J.A.M. and E.P.S., and a Howard Hughes Medical Institute Investigatorship to E.P.S. We thank Luke Hallum, Romesh Kumbhani, and Andrew Zaharia for help with experiments and for critical readings of earlier drafts of the manuscript; Yan Karklin for thoughtful conversations during the formative stages of the project; and Nicole Rust for sharing data.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Eero P. Simoncelli, Center for Neural Science and Howard Hughes Medical Institute, New York University, New York, NY 10003. E-mail: eero.simoncelli@nyu.edu.

DOI:10.1523/JNEUROSCI.2815-13.2015

 0270-6474/15/3514829-13$15.00/0

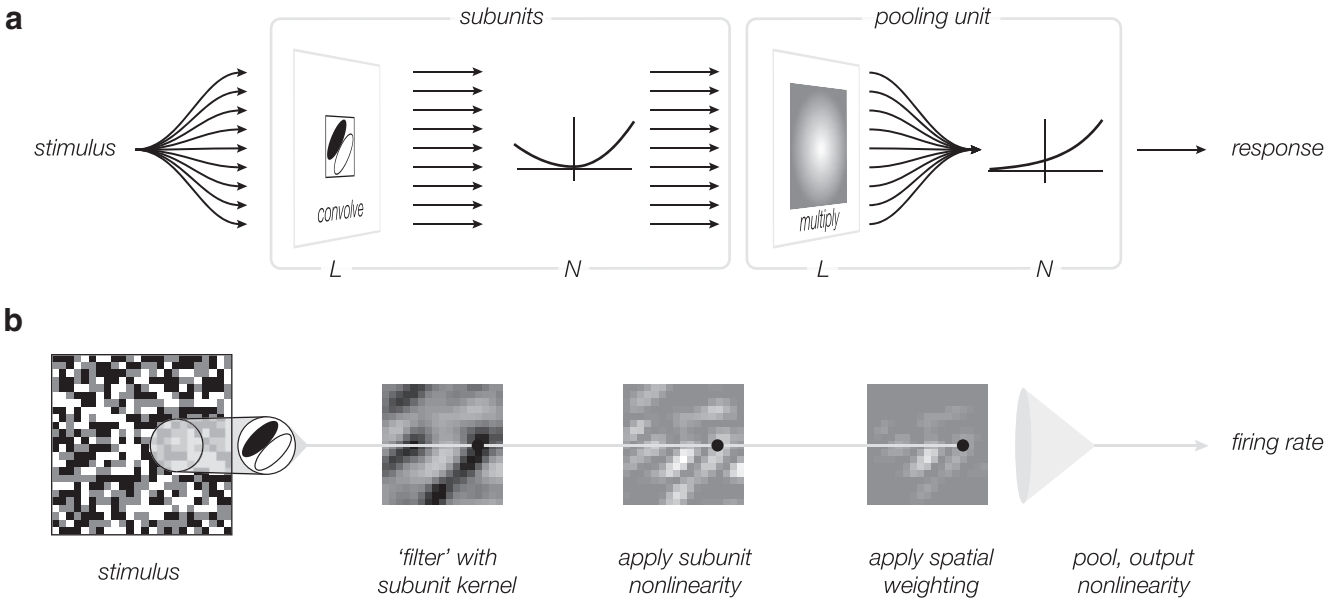

Figure 1. Subunit model for a single channel. ***a***, A signal flow diagram describes how stimulus information is converted to a firing rate. The stimulus is passed through a bank of spatially shifted, but otherwise identical, linear-nonlinear subunits. The activity of these subunits is combined with a weighted sum over space (and optionally time; not shown), and passed through an output nonlinearity to generate a firing rate. ***b***, Responses of intermediate model stages are depicted for an example input (for simplicity, a single frame is shown, rather than a temporal sequence). Each stage, except for the last, maps a spatially distributed array of inputs to a spatially distributed array of responses, depicted as pixel intensities in an image. The pooling stage sums over these responses and applies the final output nonlinearity.

We have developed a general model for V1 neurons (Fig. 1), along with a direct method for fitting it to spiking data. The model has two channels (one excitatory, one suppressive), each formed from a weighted sum of linear-nonlinear (LN) subunits, similar in concept to Hubel and Wiesel's original description of complex cell responses as resulting from a spatial combination of simple cells (Hubel and Wiesel, 1962), to Barlow and Levick's characterization of directionally selective retinal ganglion cells in the rabbit (Barlow and Levick, 1965), and to Victor and Shapley's subunit model for Y-type ganglion cells in cat retina (Victor and Shapley, 1979).

To make the fitting problem tractable, we assume that the subunit filters of each channel differ only in spatial position and that their nonlinearities are identical. The difference between the responses of the two channels is transformed with a rectifying nonlinearity to give the firing rate of the neuron. We have developed a method for directly and efficiently estimating the model, and have tested it on data from V1 neurons driven by spatiotemporal white noise. The results show that the fitted model outperforms previously published functional models (specifically, the LN, energy, and spike-triggered covariance [STC]-based models) for all cells, in addition to providing a more biologically reasonable account of the origins of cortical receptive fields. A brief account of some of this work has appeared previously (Vintch et al., 2012).

Materials and Methods

Electrophysiology

We recorded from 38 well-isolated single neurons in the primary visual area (V1) of adult macaque monkeys (*Macaca nemestrina* and *M. fascicularis*; 6 males), using methods that are described in detail previously (Cavanaugh et al., 2002). Typical experiments spanned 5–7 d during which animals were maintained in an anesthetized and paralyzed state through a continuous intravenous infusion of sufentanil citrate and vecuronium bromide. Vital signs (temperature, heart rate, end-tidal PCO_2 levels, blood pressure, EEG activity, and urine quantity, and specific gravity) were continuously monitored and maintained within physiological limits. Eyes were treated with topical gentamicin, dilated with topical atropine, and protected with gas-permeable hard contact lenses. Additional corrective lenses were chosen via direct ophthalmoscopy to make the retinae conjugate with the experimental display. All experimental procedures and animal care were performed in accordance to protocols approved by the New York University Animal Welfare Committee, and in compliance with the *National Institute of Health Guide for the Care and Use of Laboratory Animals*.

We recorded neuronal signals with quartz-platinum-tungsten microelectrodes (Thomas Recording) lowered through a craniotomy and durotomy centered between 10 and 16 mm lateral to the midline and between 3 and 6 mm behind the lunate sulcus. We recorded across all cortical depths. Receptive fields were centered in the inferior quadrant of the visual field, between 2 and 5 degrees from the center of gaze. The amplified signal from the electrode was bandpassed (300 Hz to 8 kHz) and routed to a time-amplitude discriminator, which detected and time-stamped spikes at a resolution of 0.1 ms.

Visual stimulation

We presented pixellated (XYT) noise stimuli on a gamma-corrected CRT monitor (Eizo T966; mean luminance, 33 cd/m^2), at a resolution of 1280×960 pixels, with a refresh rate of 120 Hz, positioned 114 cm from the animal's eyes. We generated stimuli pseudorandomly using Expo software (http://corevision.cns.nyu.edu) on an Apple Macintosh computer. The stimuli consisted of a pixel array (usually 16×16) filled with white, ternary noise that was continuously refreshed at 40 Hz. The width of the array was approximately double that of the receptive field (as measured with optimized drifting gratings) while still maintaining adequate pixel resolution to capture receptive field features. For a subset of cells, we measured responses to a repeated "frozen" sample of noise for cross-validation. Each sample of frozen noise had identical statistics to the main white-noise stimuli, lasted 25 s, and was repeated 20 times in succession.

Data for flickering bar (XT) noise stimuli were taken from Rust et al. (2005); these data were collected similarly, and details can be found in the original article. Stimuli were displayed using a Silicon graphics Octane-2 workstation at 100 Hz. Each frame consisted of a square region containing 16 adjacent parallel bars of the neuron's preferred orientation, ar-

ranged to cover its receptive field. On each frame, each bar was randomly assigned one of two luminance values (i.e., the bars were filled with white binary noise).

Model fitting

We consider four models for neuronal responses in primary visual cortex. All models were fit to the same data, which consisted of individual V1 cell spiking responses to the spatio-temporal pixel arrays.

LN model. The stimulus is filtered with a linear kernel whose spatial extent matches the stimulus (e.g., 16 bars, or 16 × 16 pixels) and whose temporal duration is typically 16 frames for XT stimuli and 8 frames for XYT stimuli. The filter output is then passed through an instantaneous nonlinearity that converts a simulated membrane potential to a spike rate. The model response (expected spike count in the t-th 25 ms time bin) is written as follows:

$$\hat{r}_{\mathrm{LN}}(t) = f(\mathbf{k}^T \mathbf{x}_t) \tag{1}$$

where $\mathbf{x}_t$ is a vector containing the spatiotemporal stimulus segment (intensity values over spatial position and time) preceding the t-th time bin, the superscript T indicates vector transposition, $\boldsymbol{k}$ is the vectorized filter kernel, and f is a scalar-valued nonlinear function. We estimate the filter kernel using the spike-triggered average (STA); the spike-count weighted average of preceding stimulus frames (Chichilnisky, 2001). We then use least-squares regression to estimate a piecewise-linear output nonlinearity over 9 equispaced nodes chosen to span the response distribution (this method is also used to determine the output nonlinearity for the other models, except for the Rust STC model which assumes a modified Naka-Rushton function).

Energy model. The energy model is the *de facto* standard description of complex cells in primary visual cortex (Adelson and Bergen, 1985). The responses of two filters with orthogonal phase preferences (odd-symmetric and even-symmetric) are squared and summed to generate phase-invariant selectivity. Similarly, the summed and squared responses of a second pair of filters are subtracted to capture suppression. Although this model is commonly used for complex cells, there is no standard method for fitting to white noise responses (for fitting methods using sparse bar stimuli, see Emerson et al., 1992). We have developed a novel optimization procedure to fit this model to spiking responses to white-noise stimuli. We search for the single even-symmetric filter whose squared response best matches the observed spike count in terms of mean square error. We then take the 2D Hilbert transform (Bracewell, 2000) yielding an odd-symmetric filter (orthogonal to the original) with identical power spectrum.

The model response is written as follows:

$$\hat{r}_{\mathrm{energy}}(t) = [(\mathbf{k}^T \mathbf{x}_t)^2 + (\mathbf{k}_H^T \mathbf{x}_t)^2] - [(\mathbf{s}^T \mathbf{x}_t)^2 + (\mathbf{s}_H^T \mathbf{x}_t)^2] \tag{2}$$

where $\boldsymbol{k}$ and $\boldsymbol{s}$ are the vectorized excitatory and inhibitory filter kernels, respectively, and the subscript H indicates a directional Hilbert transform. We use gradient descent to solve for the even-symmetric filters that minimize the squared response error.

To obtain the quadrature filters for $\boldsymbol{k}$ and $\boldsymbol{s}$, we first estimate the filter's predominant spatiotemporal orientation in the Fourier domain using orthogonal least-squares regression. This is obtained by seeking the eigenvector u with the largest eigenvalue of the matrix M^TM, where M is a 2- or 3-column matrix (for modeling XT or XYT stimuli, respectively) whose rows contain each of the Fourier frequencies weighted by the filters' Fourier amplitudes at those frequencies. The odd-symmetric filters are then computed by Hilbert-transforming the even-symmetric filters in the direction of this eigenvector.

Rust (STC-based) model. STC can be used in combination with spike-triggered averaging to obtain multiple filters spanning a response "subspace" (Steveninck and Bialek, 1988; Brenner et al., 2000; Simoncelli et al., 2004; Rust et al., 2005; Touryan et al., 2005; Fairhall et al., 2006; Pillow and Simoncelli, 2006; Schwartz et al., 2006; Chen et al., 2007). The general procedure uses an eigenvector decomposition of the covariance matrix of those stimuli that elicited spikes. We compute the stimulus covariance matrix as a spike-weighted sum of stimulus outer-products as follows:

$$\hat{C} = \frac{1}{t_{\max}} \sum_t r_t(\mathbf{x}_t \mathbf{x}_t^T) \tag{3}$$

where r_t is the spike count measured in the t-th 25 ms time bin, and $t_{\max}$ is the total number of time bins in the sum. The estimated model filters are the eigenvectors of this matrix whose eigenvalues differ significantly from those expected by chance (Rust et al., 2005). Specifically, larger eigenvalues are associated with stimulus attributes that excite the cell (when presented in either polarity); smaller eigenvalues are associated with stimulus attributes that suppress the cell. We find the relative weighting of each eigenvector by projecting the stimuli onto it, squaring the responses, and performing ordinary least-squares regression against the observed spike count. Whereas Rust et al. (2005) used nested significance testing to determine the relevant number of filters to include in the model, here we use cross-validated prediction error because it explicitly maximizes the measure of performance used throughout this paper: the r value of the measured spike count compared with model-predicted firing rate.

STA/STC analysis produces a set of excitatory and suppressive filters, but does not resolve how their responses should be combined to yield an estimated firing rate. Rust et al. (2005) proposed a phenomenological model built on the responses of the STA- and STC-derived filters. The STA response is half-wave-rectified, squared and weighted, and then added to the weighted responses of the fully squared STC filters responses to form an excitatory channel response, E_t. Similarly, the suppressive STC responses are squared and weighted, and combined to form a suppressive channel response, S_t. A Naka-Rushton nonlinear combination of these two non-negative channel responses is then used to compute a firing rate as follows:

$$\hat{r}_{\mathrm{STC}}(t) = \alpha + \frac{\beta E_t^{\rho} - \delta S_t^{\rho}}{\gamma E_t^{\rho} - \varepsilon S_t^{\rho} + 1} \tag{4}$$

where E_t and S_t are the excitatory and inhibitory channel responses for the t-th time bin. The 6 free parameters in this expression ($\alpha, \beta, \delta, \varepsilon, \gamma, \rho$) are fit to the data for each cell.

Subunit model. The subunit model linearly combines multiple channels (two channels, an excitatory and a suppressive channel, are used for analysis in this paper), which are themselves each a weighted sum of LN afferents, or subunits. Each channel contains an array of subunits, each consisting of the common linear filter (L) and instantaneous nonlinearity (N). The nonlinear subunit responses are pooled with a spatial weighting function. The output of the full model is generated from the linear combination of the channels as follows:

$$\hat{r}_{\mathrm{subunit}}(t) = f_r\left(\sum_c \sum_{m,n} w_c(m,n)\, f_{c,\theta}\left(\sum_{i,j,\tau} k_c(i,j,\tau) \cdot x(m-i, n-j, t-\tau)\right) + b\right) \tag{5}$$

where k_c is the filter kernel for the cth channel, $f_{c,\theta}$ is the subunit nonlinearity specified by parameter vector θ, w_c represents the spatial pooling weights, and f_r is a final rectifying nonlinearity, which transforms the summed channel responses (the "generator signal") into a firing rate. For the XT stimuli, the pooling is also performed over time, but for brevity, the equations in this section do not reflect this temporal weighting. The subunit nonlinearity, $f_{c,\theta}$, is parameterized as piecewise linear as follows:

$$f_{c,\theta}(s) = \sum_l \theta_{c,l} T_l(s) \tag{6}$$

where the T_l are a fixed set of 13 evenly spaced overlapping "tent" functions. This allows us to combine the subunit nonlinearities and pooling within the generator signal into a single summation as follows:

$$g(t) = \sum_c \sum_{m,n} w_c(m,n)\ \theta_{c,l} T_l\left(\sum_{i,j,\tau} k_c(i,j,\tau) \cdot x(m-i, n-j, t-\tau)\right) + b \tag{7}$$

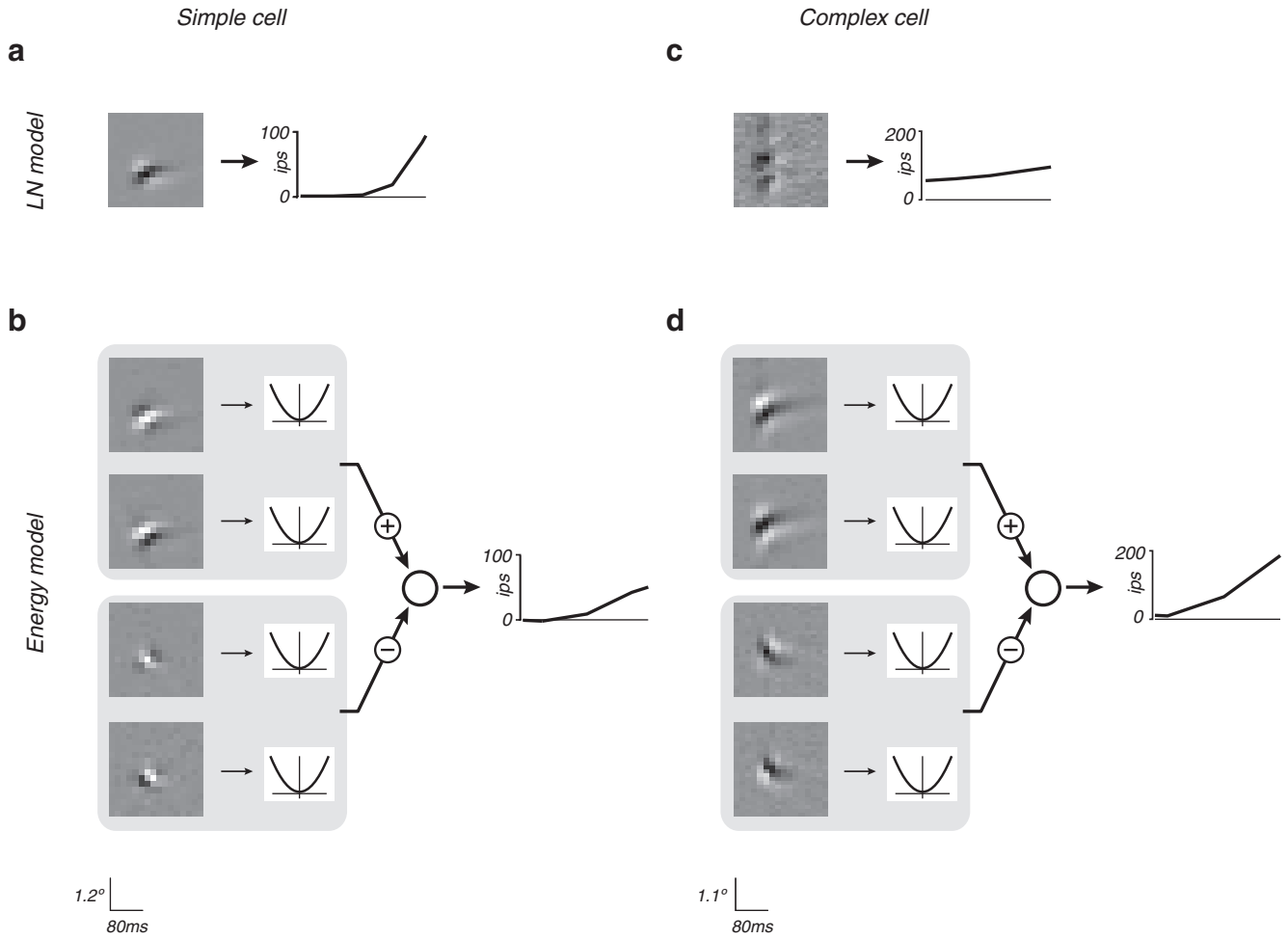

Figure 2. The LN and energy models fit to space-time (XT) data for an example simple cell and an example complex cell. ***a***, ***c***, LN model. XT filter is depicted as a grayscale image, with intensity indicating filter weight. Oblique filter orientation indicates a preference for moving stimuli (Adelson and Bergen 1985). In this figure and others, the input is in unspecified units, adjusted to span the range of linear responses that arise, given the combination of the experimental stimuli and fitted model parameters. The noisy appearance of the complex cell linear filter is indicative of a poor fit. ***b***, ***d***, Energy model, with two quadrature pairs of filters (one excitatory and one inhibitory). Display contrast for filters is normalized within each model (e.g., the suppressive filters are slightly weaker than the excitatory filters, and thus are plotted with proportionately less contrast). The energy model is invariant to phase, allowing it to capture this property of complex cells, but preventing it from describing simple cells ($r_{\text{simple}} = 0.08$, $r_{\text{complex}} = 0.41$).

To fit this model, we optimize over the parameters by minimizing mean square error of the generator signal $g(t)$ against the measured spike counts. Specifically, we solve the problem using coordinate descent, alternating between fitting the subunit nonlinearity and spatial pooling with alternate least squares (Young et al., 1976) and then optimizing over the subunit filters with gradient descent.

Because the generator is bilinear in $\theta_{c,l}$ and w_c for fixed k_c, these parameters are easy to optimize under most conditions (for an extensive treatment, see Ahrens et al., 2008). We further ensure that the solution path is well behaved through regularization: we use a cross-validated ridge prior for the spatial pooling to keep w_c compact, and a smoothing prior is used for $\theta_{c,l}$ to prevent $f_{c,\theta}$ from becoming jagged or discontinuous. This optimization problem is not convex, but like Ahrens et al. (2008), we find that the solution is stable for our data (for more details and validation against simulated data, see Vintch, 2013).

The bilinear form of the generator facilitates the calculation of the gradient for the subunit kernel, k_c. Specifically, between any two nodes of the piecewise nonlinearity $f_{c,\theta}$, the generator is a linear function of the kernel. Updating k_c causes the linear subunit outputs to drift between nodes; but because we regularize $f_{c,\theta}$ to be smooth, the optimization problem remains well behaved. After this iterative fitting procedure has converged, we complete the model fit by estimating a (piecewise linear) second stage output nonlinearity.

Subunit model initialization

The convergence time of the fitting procedure, particularly when a suppressive channel is included, benefits from a judicious choice of initial conditions. For the XT models, we use the fitted energy model to initialize the convolutional filters. Specifically, we invert the Fourier power of the excitatory and suppressive energy filters, assuming an odd filter with zero phase of the desired size.

The XYT models have many more parameters and must be fit to datasets with more samples. In general, we find that the energy model does not produce results of sufficient quality, especially in the suppressive channel, to be effective as an initialization tool. We therefore developed a novel technique that we call "convolutional STC" to generate initial filters (Vintch et al., 2012). We compute an STC matrix over a set of spatially overlapping stimulus blocks, each the size of the subunit model kernels. We use the eigenvectors with the largest and smallest eigenvalues to initialize each channel in the subunit model. This initialization procedure generally returns better cross-validated models than the method described for the XT models.

Model validation

We estimate each of the four models on "training" data and validate on the held-out set of "testing" data. This is essential because the spiking responses contain both a stimulus-driven component (the signal) and

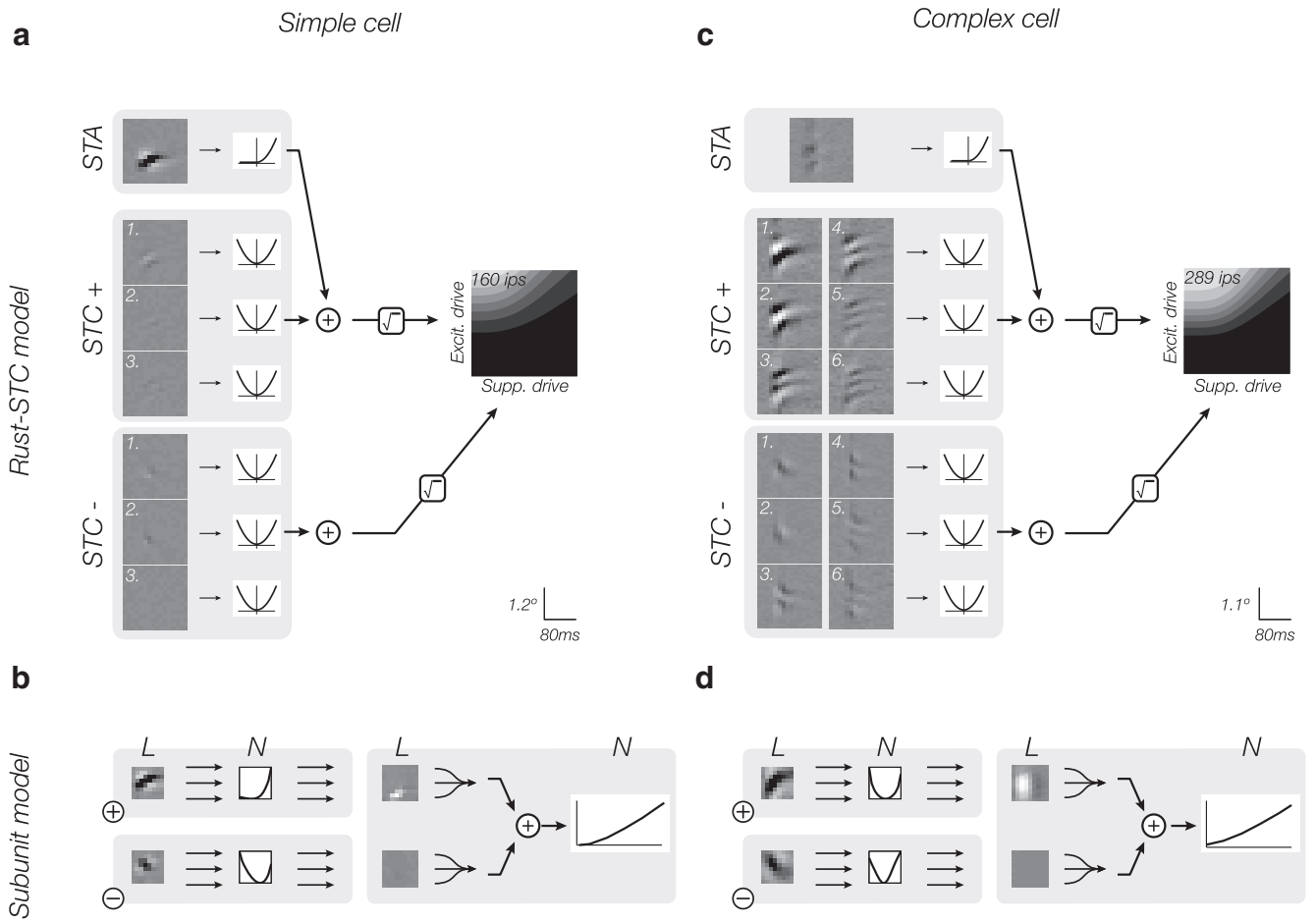

Figure 3. The Rust-STC and subunit models fit to XT data for the example simple and complex cells of Figure 2. ***a***, ***c***, Rust-STC model. Contrast of individual filters indicates their weighted contribution to the full model. The output nonlinearity is a joint function of the excitatory and suppressive drive (see Materials and Methods) and is depicted as a 2D image in which intensity is proportional to firing rate. Cross-validated model performance: $r_{simple} = 0.56$, $r_{complex} = 0.47$. ***b***, ***d***, Subunit model with two channels (excitatory and suppressive). The subunit filters are constrained to have unit norm, but the intensities in the image depicting the linear pooling weights indicate their relative contribution to the response. Cross-validated model performance: $r_{simple} = 0.55$, $r_{complex} = 0.42$.

random fluctuations (the noise). Models that are underconstrained (too many parameters for the amount of data) tend to fit the noise as well as the signal, showing high accuracy for the training data but worse performance on the testing data.

For XT data we performed "5-fold cross-validation": we use 4/5th of the data for training and test the fit on the remaining 1/5th of the data. We do this five times and average the fitting error over all five sets (the tranches of data correspond to randomly selected trials over time, rather than temporally contiguous sequences). Performance is measured as the correlation coefficient between the measured spike count and the predicted firing rate, and overfitting is assessed by comparing the training performance to the test performance. Although valuable for this dataset, this form of cross-validation is relatively cumbersome because the model must be estimated five times in succession.

We also collected data that allows us to perform a different kind of validation that is both more efficient and more illuminating. For a subset of cells that were presented with the XYT stimuli, we collected responses to 20 repeats of a 25 s stimulus that had not been presented before. We then tested the fitted model on each repeated trial and computed the average correlation coefficient over all 20 trials. That is, the model was fit to the longer, unrepeated stimulus presentation and tested on each of the shorter, repeated stimulus segments.

These cross-validation methods allow us to compare performance across models and provide a measure of overfitting, but they cannot tell us how well the models are performing on an absolute scale. Each recorded spike count includes many sources of noise (e.g., input noise, variability in spike generation, spike sorting errors), and this noise establishes a limit to the performance that any stimulus-driven model can achieve. We estimate this ceiling by computing an "oracle" response that seeks to minimize the effects of noise. Specifically, as a prediction of the response on a given trial, we take the average of responses on the other 19 trials. Many neurons are very unreliable, and their trial-to-trial variability is high. For these neurons, the oracle is expected to perform poorly, as is any stimulus-driven model. By comparing our models to the oracle maximum, we learn how much of the explainable (stimulus-driven) variance each model captures.

Assessing the parameters of the fitted subunit models

The structure of the subunit model is convenient for dissociating the spatiotemporal position of the receptive field from its tuning properties. We measure the spatial and temporal extent of the subunit filter and linear pooling for the XT models by fitting a 2D Gaussians with 6 parameters (1 for amplitude, 2 for spatial position, and 3 for the covariance matrix; Cholesky decomposition). The subunit filters are bandpass and contain both positive and negative components, so we first compute spatial power envelope before fitting a Gaussian. First, we find the 2D Hilbert transform in the direction of the subunit filter's preferred orientation. The power envelope is then the square-root of the sum of squares of the original subunit filter and its Hilbert transform. We compute the relative influence of the excitatory channel and the suppressive channel by comparing the SD of their independent responses over all stimuli frames.

For the cells tested with XYT stimuli, we also collected direction tuning curves for optimal drifting gratings. To determine the model predictions

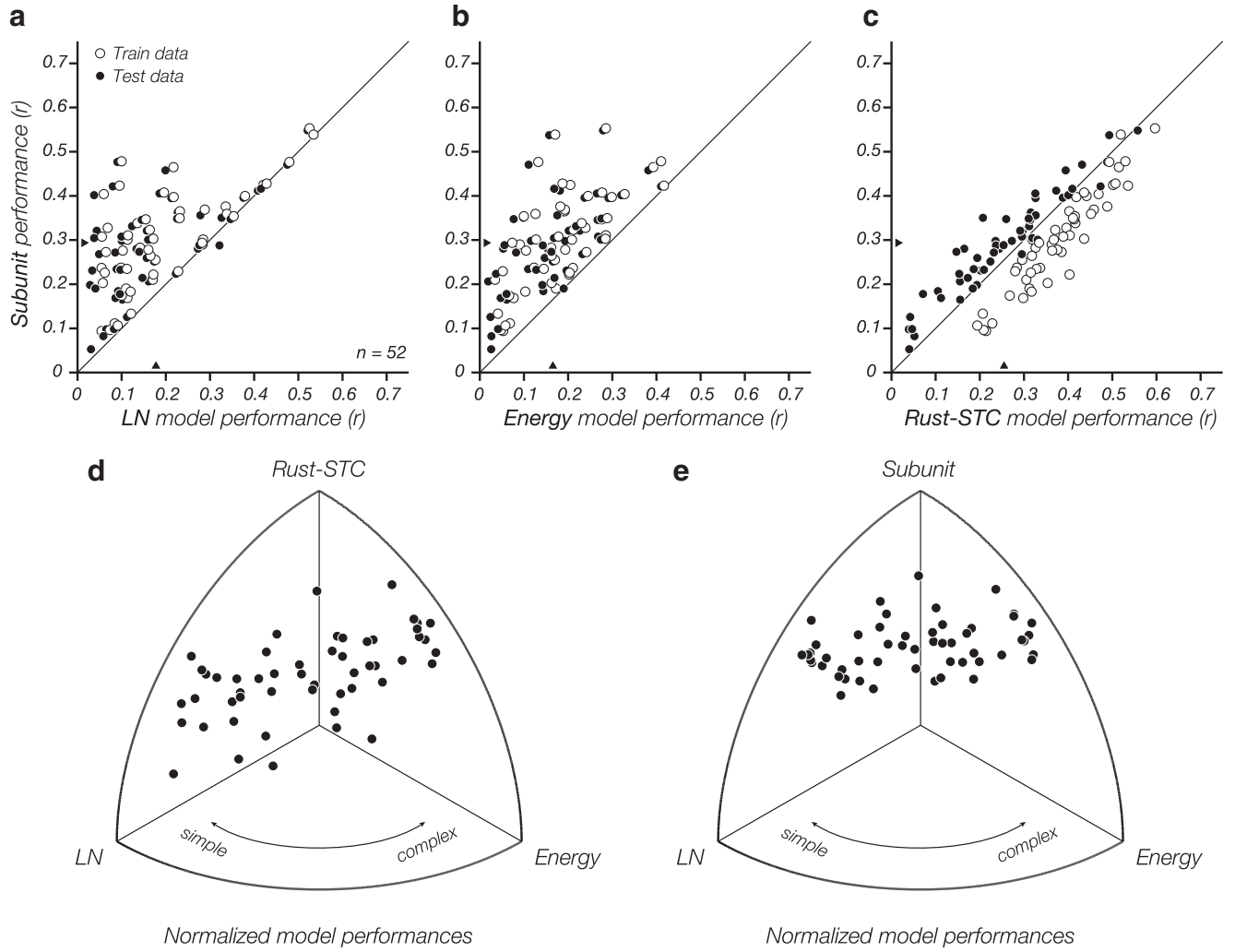

Figure 4. Comparison of models across all cells fit to space-time (XT) data. Model performance is measured as the correlation between the measured spike count and the model-predicted firing rate. Each point corresponds to a cell, with hollow symbols indicating performance on training data and solid symbols indicating performance on held out testing data (i.e., cross-validated). ***a***, ***b***, The subunit model outperforms both the LN and energy models. ***c***, The subunit model outperforms the Rust-STC model on cross-validated data (solid points), but not on training data (hollow points), a clear indication that the Rust-STC model is overfitting. ***d***, Relative performance of Rust-STC, energy, and LN models, plotted by projecting the 3D vector of *r* values for each cell onto the surface of the unit sphere. The relative performance of energy versus LN models gives an intrinsic indication of cell complexity. The Rust-STC model outperforms the energy model for most complex cells and has performance approximately comparable with the LN model for simple cells (the LN model is a special case of the Rust-STC model). ***e***, Relative performance of subunit, energy, and LN models. The subunit model consistently outperforms the other models, independent of cell type.

for direction tuning, we generated drifting sinusoidal stimuli near the model-optimal SF and TF and fed these sequences to each model along with the fitted parameters for each cell. Gratings have more effective contrast than the noise to which the models were fit. We therefore fit a new output nonlinearity for each model and each cell to account for the output scaling. For the subunit model, we also extended the dynamic range of the subunit nonlinearities, fitting each with a piecewise power-law function. Specifically, the nonlinearity for each channel is split at zero, and both the positive and negative side are fit with an independent nonlinearity of the form $\lfloor x - b \rfloor^p$, where b is an offset and p is a power bounded between 0 and 5.

Results

We analyzed the responses of neurons in macaque primary visual cortex to white-noise stimuli. One population of 52 neurons (data from Rust et al., 2005) was stimulated with an array of random black and white parallel bars at the neuron's preferred orientation, covering the receptive field. We label this class of stimuli "XT noise," referring to the two relevant dimensions of space (X) and time (T). A newly recorded second population of 38 neurons was studied with ternary pixel noise updated at 40 Hz. We refer to these stimuli as "XYT noise." For each cell, we fit the subunit model and three other previously published models: an LN model based on the STA, the "energy model" (Adelson and Bergen, 1985), and the "Rust-STC model" (Rust et al., 2005).

The LN cascade model is the simplest and most well-known descriptive model for response of sensory neurons. At each moment in time, the linear stage computes a weighted sum (or integral, if the stimulus space is continuous) over local stimulus values and recent time. The weights determine the stimulus selectivity of the model cell, and the linear operation can be interpreted as passive combination of incoming signals weighted by their associated synaptic efficacies. This response is then transformed with a rectifying nonlinear function, which can be interpreted as converting membrane voltage to firing rate. When applied to neurons from area V1, this framework is only practical for simple cells, which are known to have a monotonic contrast-polarity response functions. Complex cells, which respond well to stimuli of both polarities and exhibit position-invariant re-

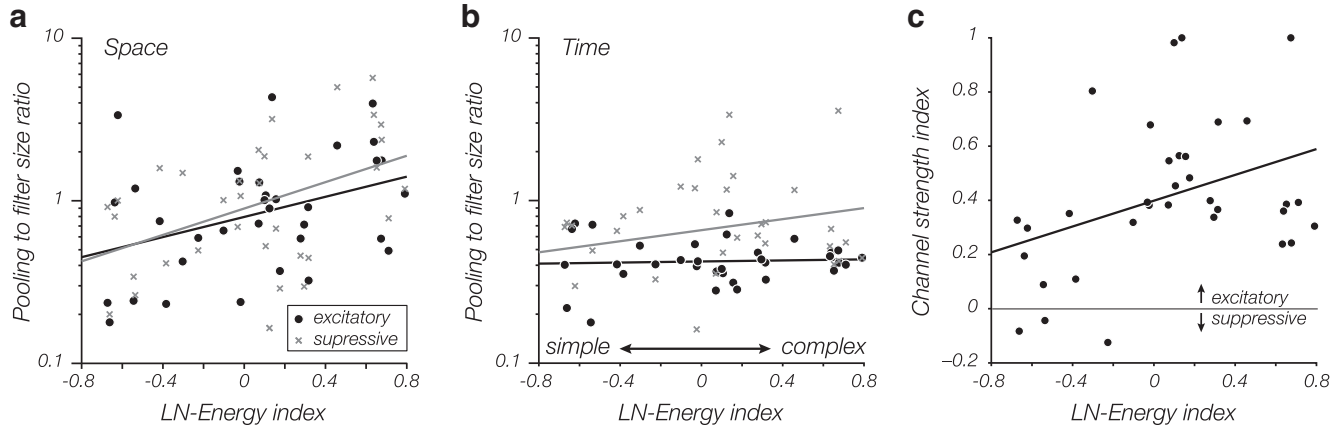

Figure 5. Dependence of model-estimated pooling extent and strength on cell complexity. The abscissa of each graph represents a complexity index (derived from the LN and energy models as $(E_{energy} - E_{LN})/(E_{energy} + E_{LN})$; see Materials and Methods). A value of −1 indicates a purely simple cell; and a value of 1 indicates a purely complex cell. ***a***, Relative spatial pooling size in the fitted subunit model increases with cell complexity. For each cell, we computed the ratio of SDs of 2D Gaussian envelopes fitted to the pooling weights and the subunit filter. As expected, complex cells pool over a larger relative region than simple cells. This same effect is seen for the suppressive channel (gray x's). ***b***, Relative temporal pooling size is not correlated with cell complexity, but pooling duration of the inhibitory channel is generally larger than that of the excitatory channel. ***c***, Relative strength of excitatory to inhibitory channels increases with cell complexity. Relative strength is computed as the ratio of SDs of the two channel responses (i.e., generator signals) over all stimuli.

sponses within their receptive fields, cannot be explained with a single filter, regardless of nonlinearity. Figure 2*a*, *c* shows model parameters for an example simple and complex cell, fitted to the XT data.

The "energy model" (Adelson and Bergen, 1985) is a standard model for complex cell responses. The model computes the sum of squares of responses of two space-time oriented linear filters, a quadrature pair that have the same frequency response magnitude but different symmetry (one is even-symmetric, the other odd-symmetric). This LN-L construction is a quantitative instantiation of the qualitative description given by Hubel and Wiesel (1962) and can be thought of as combining the responses of four simple cells with identical retinotopic location and orientation and frequency selectivity, differing only in the precise arrangement of excitatory and suppressive regions. The response of this model to sinusoidal gratings is phase insensitive while retaining selectivity for spatial orientation and direction. We developed a method to fit this model to spiking data recorded during exposure to white-noise stimuli (see Materials and Methods). Figure 2*b*, *d* shows model parameters for the example simple and complex cells.

Neurons in V1 have been traditionally categorized as simple or complex (Hubel and Wiesel, 1962), but their properties may lie on a continuum (Mechler and Ringach, 2002; Rust et al., 2005). The model developed by Rust et al. (2005), which we will denote as Rust-STC because of its reliance on spike-triggered covariance for filter estimation, combines aspects of the LN and energy models into a single framework, in which a set of LN responses are additively combined to form excitatory and suppressive signals, which are then combined using a spiking nonlinearity that includes both subtractive and divisive interactions (Rust et al., 2005). The linear filters are obtained from STA and covariance (STC) analyses. Unlike the LN or energy models, the combination of linear and energy-like elements allows the Rust-STC model to accommodate the range of simple and complex cells, as can be seen for the example simple cell and complex cell (Fig. 3*a*,*c*).

A drawback of the Rust-STC model is that, unlike the LN and energy models, the parameters are more difficult to interpret as biological elements, primarily because the STC analysis forces the filters to be orthogonal. As a consequence, pairs of filters beyond the first pair are delocalized in both space-time and spatiotemporal frequency (Fig. 3*a*,*c*), properties that are generally not observed in real cells (Rust et al., 2005; Vintch et al., 2012). Furthermore, the Rust-STC model has many more degrees of freedom than the LN or energy models. As a result, obtaining an accurate fit to the responses of an individual neuron requires a large amount of data (Rust et al., 2005).

We have developed a subunit model that retains many of the advantages of the Rust-STC model, but with an architecture that has a more natural biological interpretation and requires far less data to achieve a satisfactory fit (Vintch et al., 2012). The model sums over two channels, one excitatory and one suppressive; each channel is constructed from a bank of LN "subunits" with identical receptive field structure, which differ in their spatial and temporal position (in other words, the set of linear filters perform a convolution). For each channel, the set of linear filter responses are passed through a common output nonlinearity before they are linearly weighted and summed to generate the channel response. Because of this structure that enforces local computations, the model operates on the same stimulus space as the Rust-STC model with many fewer parameters. The model framework, along with examples of the computation performed at each stage, is depicted in Figure 1. We fit the linear filter and nonlinearity, as well as the spatial and temporal pooling functions (weights), for both an excitatory and suppressive channels for each cell (see Materials and Methods). Fitted model parameters for the example simple and complex cell are shown in Figure 3*b*, *d*. The model fitted to the simple cell makes primary use of one channel, with an asymmetric (approximately half-wave-rectifying) nonlinearity, and a pooling region that is highly concentrated. For the complex cell, on the other hand, the model uses a symmetric (approximately quadratic) nonlinearity, and a much broader pooling region.

When averaged across all cells presented with the XT stimuli, the subunit model produced the most accurate cross-validated fits of the tested models ($< r_{LN} > = 0.18$, $< r_{energy} > = 0.17$, $< r_{STC} > = 0.26$, $< r_{subunit} > = 0.29$; Fig. 4*a*–*c*). The LN, energy, and subunit models exhibit cross-validated (testing) performance that differs only slightly from their corresponding training

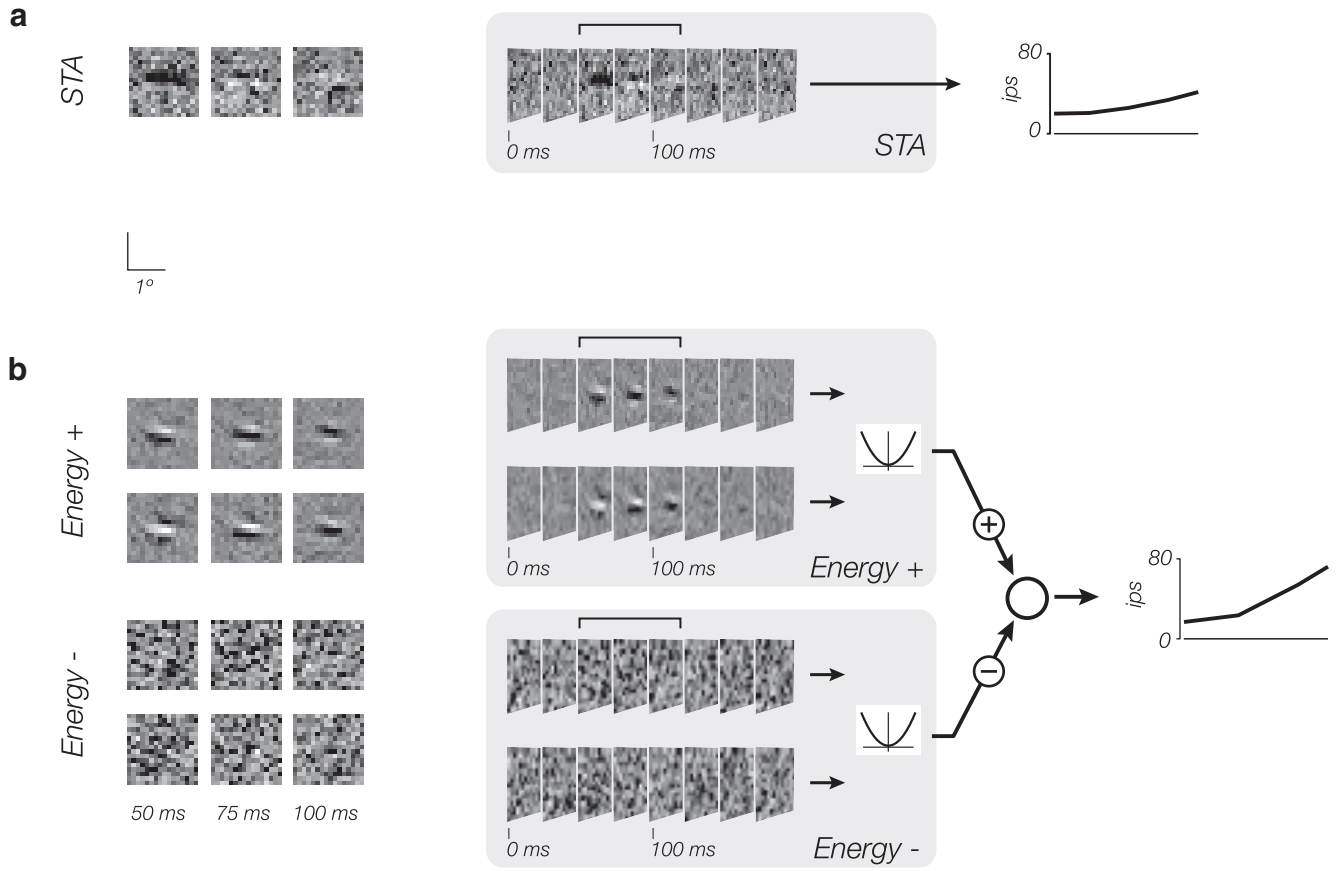

Figure 6. The LN and energy models fitted to space-space-time (XYT) data from an example complex cell. ***a***, The LN model shows selectivity for spatial patterns over time. The 3D filter kernel is depicted on the right as a stack of spatial filters, each associated with one (25 ms) temporal frame. Three relevant time slices (50, 75, and 100 ms) are replotted on the left for clarity. ***b***, The energy model, with two quadrature pairs of filters (one excitatory, one inhibitory).

performance (Fig. 4*a–c*), indicating that these data are sufficient to constrain them. The inferior performance of the LN and energy models arises because each has a form that is too restrictive to account for the behaviors found across all cells. In particular, The LN model only captures behavior of simple cells, and the energy model captures complex cells (Fig. 2). On the other hand, both the Rust-STC and subunit models are able to capture the behaviors of both cell types (Fig. 4*d*,*e*).

The performance improvement of the subunit model over the Rust-STC model arises primarily because of overfitting in the latter. On training data, the Rust-STC model outperforms the subunit model, but it underperforms on a held-out test set (Fig. 4*c*). Test performance as a percentage of training performance was 96% and 67% for the subunit and Rust-STC, respectively, which are significantly different from one another (two-tailed *t* test; $p \ll 0.05$). This overfitting behavior of the Rust-STC model is a result of the substantially larger parameter space. Although both models compute responses as a function of a 16×16 pixel stimulus space, the Rust-STC model uses a set of linear filters that are the full size of this space, whereas the subunit model uses two 8×8 filters that are applied convolutionally. In addition to the larger filter size, the Rust-STC model generally uses more than two filters (the number is adapted for each cell), which are estimated from a 256-dimensional covariance matrix. To examine whether the filter size alone accounted for the overfitting, we fit each cell with a Rust-STC model restricted to the central 8×8 pixel region of the stimulus. As expected, the average cross-validated performance of this model was significantly worse ($< r_{\text{reducedSTC}} > = 0.23$, which is 89% of the unrestricted performance). However, despite the fourfold reduction in the dimensionality of the parameter space, the restricted model still overfit the data, exhibiting test performances that were 74% that of the training performance on average (which was not significantly different from the unrestricted Rust-STC model; two-tailed *t* test).

The subunit model also has the advantage that its structure dissociates receptive field position and spatial extent from stimulus tuning properties. A simple cell and a complex cell with identical subunit filters would exhibit identical frequency and orientation tuning properties but would differ in the size of their receptive fields and their invariance to stimulus phase or position. We measured the spatial and temporal extent of the subunit filter and pooling map for each cell by fitting a 2D Gaussian envelope to each (see Materials and Methods). As expected, pooling for most complex cells covers a larger spatial region than that of simple cells (Fig. 5*a*). This is true in both the excitatory channel and the suppressive channel ($r_{\text{excitatory}} = 0.42$, $p < 0.05$; $r_{\text{suppressive}} = 0.39$, $p < 0.05$). The extent of pooling over time was not correlated with cell type (Fig. 5*b*), but over all cells the suppressive channel tended to sum over time more broadly than the excitatory channel ($p < 0.05$, *t* test).

The smaller pooling envelope for simple cells should not be interpreted to mean that simple cells have less suppression. For each cell, we calculate a "channel strength" index that is 0 if excitation and suppression are balanced, and near 1 if excitation is much stronger than suppression (see Materials and Methods).

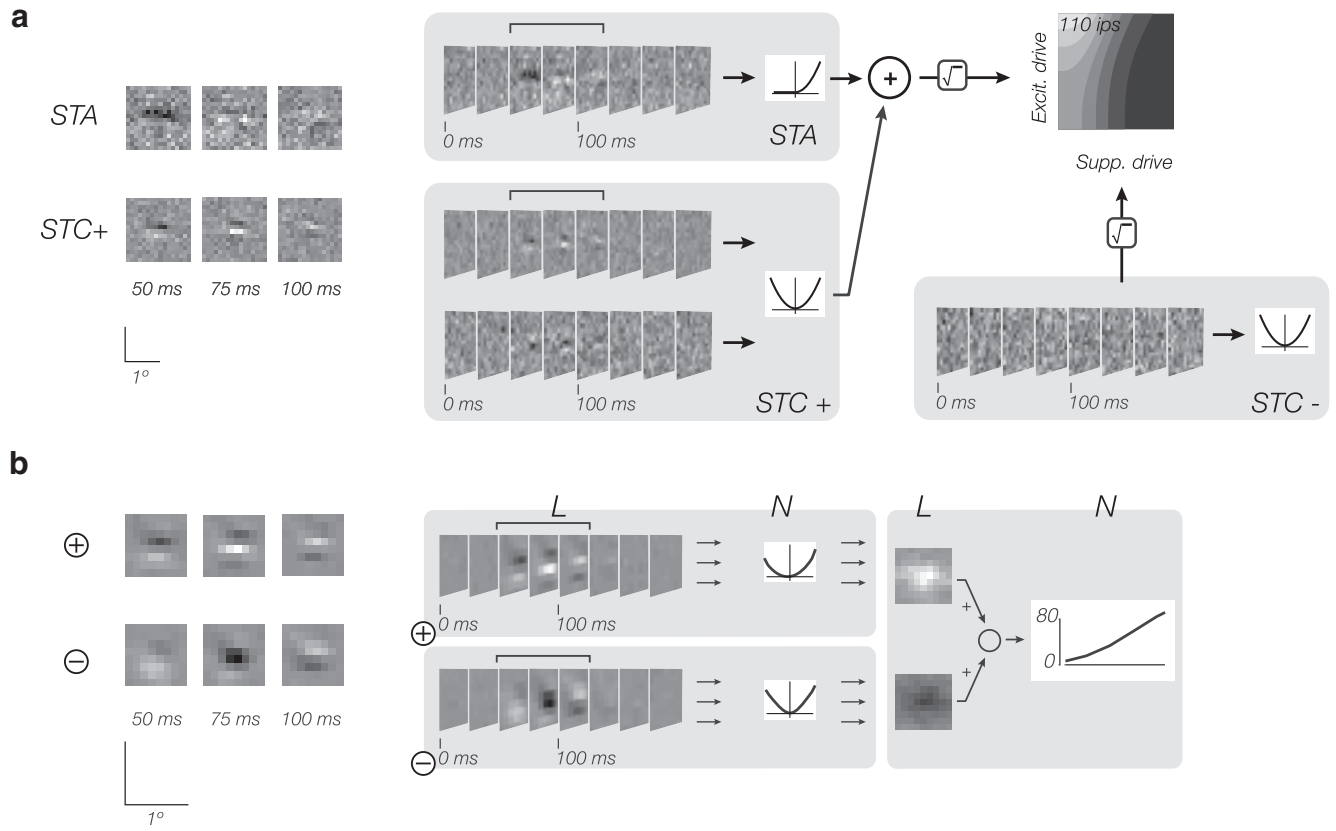

Figure 7. The Rust-STC and subunit models fitted to XYT data from the example complex cell of Figure 6. Conventions follow Figure 6. ***a***, Rust-STC model. The high dimensionality of the XYT filters means that the parameters of this model are difficult to estimate and the resulting filters are very noisy. ***b***, Subunit model. The convolutional nature of the subunit model allows it to cleanly capture the structure of the complex cell receptive field. There is a clear selectivity for upward motion over time and suppression by opposite-direction motion at a lower spatial frequency.

Although our extracellular recordings do not allow us to distinguish between excitatory and inhibitory inputs, the data show a tendency for simple cells to be more balanced and for complex cells to have a stronger excitatory drive (Fig. 5*c*; $r = 0.38$; $p < 0.05$).

The subunit model is easily extended to handle stimuli with two spatial dimensions in addition to time. We presented 38 cells with ternary pixel noise at 40 Hz on a 16 × 16 grid. We fit these data with a subunit model whose filter kernels covered an 8 × 8 pixel region, and 8 frames. Because of the relatively slow frame rate, we used a pooling stage that combined subunit responses only over space.

For a subset of neurons, we also measured responses to 20 repeats of an identical ("frozen") noise stimulus. Cross-validated performance for these repeat experiments was computed by fitting to subsets of 19 of the repeated trials, and measuring correlation (r value) between the measured spike count of the 20th trial and the model-predicted firing rate.

The four model fits for an example complex direction-selective cell are shown in Figures 6 and 7 This cell responds best to horizontal features that drift upward over time, a preference that can be observed in the filter kernels of all four models. The Rust-STC model filters for this cell are much noisier than those for any of the other models, a result of the high dimensionality of the parameter space. For this example cell, the subunit model had the best cross-validated performance ($r_{LN} = 0.16$, $r_{energy} = 0.28$, $r_{STC} = 0.43$, $r_{subunit} = 0.54$).

As with the XT data, the subunit model outperformed the other models in terms of average cross-validated correlation across all XYT cells ($< r_{subunit} > = 0.27$, $< r_{STC} > = 0.16$, $< r_{energy} > = 0.12$, $< r_{LN} > = 0.12$; Fig. 8). For these stimuli, the difference in model performance was more substantial than for the XT stimuli: the subunit model performed ~70% better than the Rust-STC model on average. The Rust-STC model also showed a larger gap between training and testing performance than for the other models (Fig. 8*d*, open and closed circles), again indicating a high degree of overfitting. Most of the increase in performance for the subunit model comes from the model structure and filters; replacing all of the subunit nonlinearities with best-fit second-order polynomials (thus matching the shape of the filter nonlinearities in the Rust-STC model) does not significantly affect overall subunit model performance ($p = 0.98$, two-tailed t test). The specific choice of nonlinearity may affect model interpretation (Kaardal et al., 2013) but does not materially affect model performance.

We can also compare the performance of the subunit model to the upper bound derived from the frozen noise measurements, which we refer to as the "oracle" prediction. The oracle used the mean response over 19 repeated presentations to predict the firing rate for the 20th presentation (Fig. 8*a*; we do this for all 20 trials and take the average). On average, the subunit model performed 76% as well as the oracle model, whereas the Rust-STC model only performed 49% as well. Thus, the subunit model is able to capture a significant percentage of the explainable (stimulus-driven) variance in each cell's binned spike counts, and can do so much more efficiently (i.e., with less training data) than the Rust-STC model.

We also examined the ability of each model to predict direction tuning curves for drifting gratings. We simulated responses

to sine-wave gratings at the optimal spatial and temporal frequency for each cell, and used each model, with parameters fitted to the white-noise training data, to predict the grating responses averaged across phase (i.e., the DC response). Gratings have higher effective contrast than the white noise used to estimate the model parameters, so we estimated a new output nonlinearity that mapped the model drive (sum of excitatory and inhibitory channels) to the observed spike count. We then compared these model-predicted direction tuning curves to the actual measured tuning curves (Fig. 9*a* shows this for an example cell).

The subunit model was consistently more accurate than the other three in predicting the direction tuning curves (Fig. 9*b*). The superior performance of the subunit model comes both from its estimates of direction preference (Fig. 9*c*) and of tuning width (Fig. 9*d*). All models predict tuning curves that are broader than those directly measured, although the subunit model was the most accurate. The cells for which the model provided accurate tuning curve estimates are also the cells for which the subunit model predicts novel white-noise stimuli well ($r = 0.49$, $p = 0.025$; Pearson r value, data not shown). These cells were also the cells with the highest average firing rate ($r = 0.57$, $p = 0.007$; Pearson r value) and the highest number of total recorded spikes ($r = 0.59$, $p = 0.005$; Pearson r value).

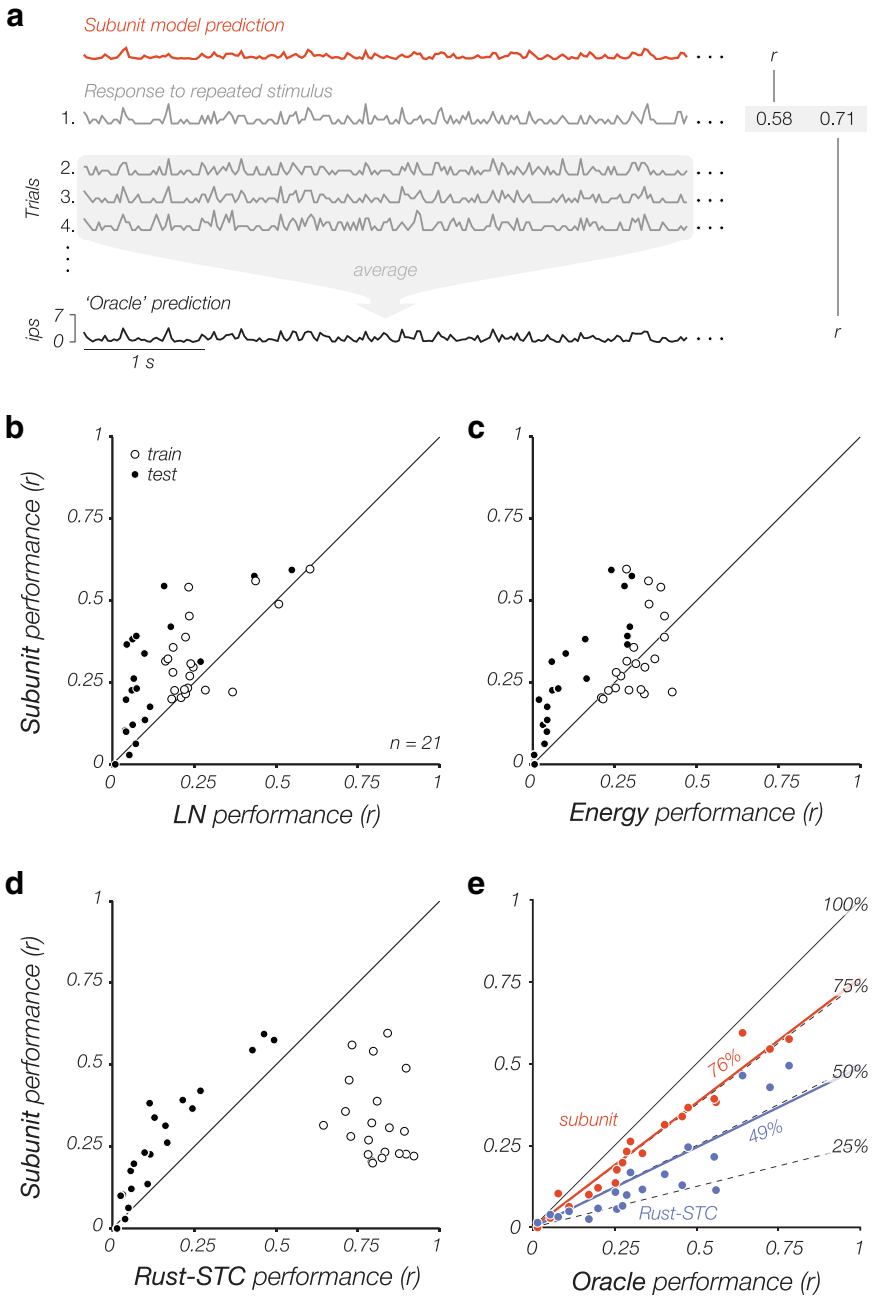

Figure 8. Comparison of model performances across a subset of XYT cells ($n = 21$), for which we obtained responses to a 1000-frame stimulus repeated 20 times. ***a***, Model performance is computed as the average correlation (r) between the observed spike count for each of the 20 trials and the model-predicted firing rate. We also computed performance for an "oracle," which predicts responses on a given trial using a rate estimated by averaging the other 19 trials. This provides an approximate upper-bound on performance of any stimulus-driven model. ***b–d***, The subunit model outperforms all other models on cross-validated testing data (solid), but not on training data (hollow). This indicates significant overfitting for the other models, especially the Rust-STC model. ***e***, Comparison of subunit and Rust-STC models to the oracle. On average, the subunit model (orange points) captures 76% of the variance explained by the oracle, and the Rust-STC model (purple points) captures 49%.

Discussion

We have developed a generalized subunit model for neurons in primary visual cortex, along with a method to directly and efficiently estimate the parameters of this model from measured spike counts. The model includes as special cases the LN model that is widely used for simple cells, as well as the energy model that is widely used for complex cells. It in turn is a special case of a general subspace model. When fit to a relatively short segment of responses to white-noise stimuli, the new method significantly outperforms both of these classic models in terms of the accuracy of predicted responses on a held out test set, over a large set of neurons from macaque primary visual cortex. We also compared the performance of our model to that of Rust et al. (2005), which provides good fits for both simple and complex cells. The subunit model performed significantly better on cross-validated data than the Rust-STC model, primarily because it is more compactly parameterized and thus less susceptible to overfitting. As the number of samples grows, the Rust-STC model could perform as well or better than the subunit model because of the flexibility provided by its larger number of parameters.

Subunit models date to the work of Hubel and Wiesel (1962), who proposed that the absence of distinct excitatory and suppressive subregions in most complex cell receptive fields could be explained by combining the responses of a set of spatially distributed simple cells having similar stimulus preferences; this model was further developed and quantified by Movshon et al. (1978). In the retina of the cat, Hochstein and Shapley (1976) showed that spatial frequency responses of Y cells arose from spatial filter significantly smaller than the receptive field size, and proposed a

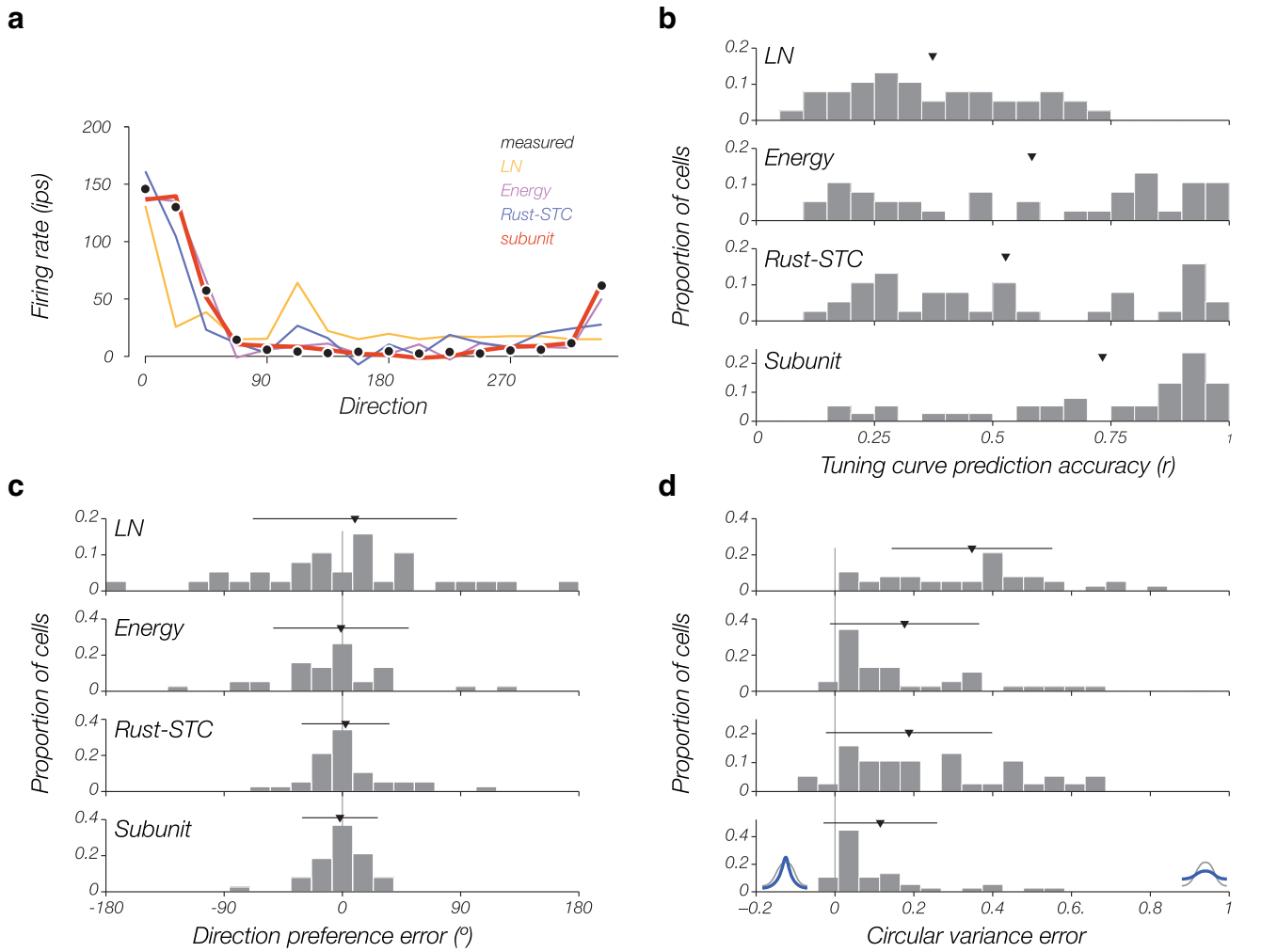

Figure 9. Predicting responses to drifting gratings. ***a***, Black dots indicate the measured tuning curve for drifting gratings over 16 directions. Colored lines indicate the model-predicted tuning curves for all four model types (model predictions are rescaled for illustration purposes; see Materials and Methods). ***b***, For all cells, we quantify model tuning curve accuracy as the correlation between the actual tuning curve and the model-predicted tuning curve. Here, a histogram is plotted over all cells ($n = 38$). Triangles represent population means. Error bars (horizontal lines) indicate SD. ***c***, Error in direction preference is smallest for the subunit model, but all models are unbiased in their error. ***d***, Error in circular variance measures the width of the tuning curves (flat curves have a CV of 1). All models tend to predict flatter tuning curves than those measured from drifting gratings, but the subunit model is better than the other models on average. Black and blue curves at the bottom illustrate actual and model predicted tuning curves, respectively.

model for these responses based on a sum of nonlinear subunits. Adelson and Bergen (1985) introduced the spatiotemporal "energy model" as a simplification of the Hubel and Wiesel (1962) description that used only two filters, of differing phase, whose responses were squared and summed. Fukushima (1980) suggested that the spatial pooling of rectified filter responses is a canonical operation that is repeated in all visual cortical areas to impart translation-invariance. Consistent with our own motivations, convolutional processing was introduced to artificial neural networks as a means of greatly reducing the parameterization, and thus enabling efficient training (LeCun et al., 1989). This notion was generalized to include pooling over other dimensions by Perrett and Oram (1993), and is widely used in modern object recognition systems (Poggio and Edelman, 1990; LeCun and Bengio, 1995; Wallis and Rolls, 1997; Riesenhuber and Poggio, 1999; Jarrett et al., 2009). It has also been proposed as a means of representing statistical quantities (local variances or covariances), which can support the representation and recognition of visual texture (Freeman and Simoncelli, 2011; Freeman et al., 2013).

Methods for fitting V1 models directly to physiological recordings originate with the work of Jones and Palmer, who used reverse correlation (De Boer and Kuyper, 1968; Marmarelis and Marmarelis, 1978) to obtain spatial or spatiotemporal receptive fields for simple cells (Jones and Palmer, 1987). Emerson et al. (1992) introduced a sparse noise methodology for fitting the energy model to complex cell responses. Lau et al. (2002) fit complex cell responses using an artificial neural network model that included multiple unconstrained filters, revealing families of oriented receptive fields of differing phase similar to the energy model. Parallel work by Touryan et al. (2002) used the spike-triggered covariance (Stevenick and Bialek, 1988) to estimate orthogonal pairs of oriented filters underlying complex cell responses, and later work by Rust et al. (2005) and extended this analysis to reveal larger sets of orthogonal filters which explain the responses of most complex cells. This generality comes at a cost: STC methods assume Gaussian stimulus ensembles (Paninski, 2004), and recording durations that strain existing experimental capabilities (although we note that Park and Pillow, 2011, have recently introduced regularization methods for STC that

partially alleviate these). Information-theoretic methods have been introduced to find orthogonal filters that capture the subspace of stimuli responsible for neural responses (Paninski, 2004; Sharpee et al., 2004). Compared with STC, these methods are less restrictive in terms of the stimuli (e.g., they have been used with natural movies), but the estimation of information generally makes them even more data intensive.

The method introduced here greatly reduces the data requirements of the subspace methods, by exploiting the simplified parameterization that results from banks of spatially shifted (convolutional) filters. Although one can proceed by first finding a response subspace (using one of the methods described in the previous paragraph, such as STC) and then solving for a filter whose shifted copies span that same subspace (Rust et al., 2005; Lochmann et al., 2013), such a two-step procedure does not generally realize the gains in accuracy that should accompany the reduction in dimensionality because the errors introduced in the first step are substantial, and the second step does not take them into account. Our method is also related to several analysis techniques in recent literature that incorporate the concept of subunits, but without building an explicit receptive field model. Local spectral reverse correlation computes a spike-triggered local Fourier spectrum over localized, stimulus windows (Nishimoto et al., 2006). The method computes both localized frequency and orientation tuning profiles (which can be linked to the responses of suitably frequency-tuned filters), as well as a position map. But as with the subspace methods, local spectral reverse correlation requires substantial amounts of data to achieve clean results. In addition, the spectral characterization of local regions combines excitatory and suppressive influences, complicating biological interpretation. Similar benefits and drawbacks may be attributed to the method of Sasaki and Ohzawa (2007), which estimates local second-order "subunit" kernels that respond in a contrast-invariant manner to windowed stimulus regions.

Perhaps the method most similar to our own is that of Eickenberg et al. (2012), who used an information-theoretic objective function to fit a model constructed from a logical OR-like combination of shifted filter responses. As in our case, the authors found that this model was substantially more data-efficient than a general subspace model. But the OR-like combination of shifted filter responses differs substantially from the LN-LN construction of our model, and the authors found that it was most suitable for characterizing responses of complex cells.

The model and methods presented in this paper offer a number of possibilities for extension and generalization. In particular, our current fitting procedure optimizes squared error between model response and observed spike counts. Incorporating a spike generation stage, even a simple one such as a Poisson process, would more accurately capture the precision associated with different spike count observations. For the purposes of this work, we assumed that neurons were held in a steady state of gain and adaptation by the continuous white-noise stimulation. It was therefore unnecessary to build into the model the known nonlinear mechanisms of gain control and adaptation (Sclar et al., 1990; Carandini et al., 1997). For a more comprehensive account of V1 responses, these would need to be included. In this and other ways, the general structure of the model reflects computations found in earlier stages of visual computation, such as the retina (Hochstein and Shapley, 1976; Victor and Shapley, 1979), and perhaps later stages, such as V2 or V5 (MT). We are currently exploring application of like models to suitable data (Vintch et al., 2012; Freeman et al., 2013).

In conclusion, we take pleasure in observing that the subunit model can readily be interpreted in biological terms. The field of LN subunits can be viewed as a population of afferents from upstream layers or visual areas, and the channel response arises from a linear combination of these afferents, weighted by their associated synaptic efficacies. The fitted subunit filters are tightly localized in space and time, like receptive fields in visual cortex (Hubel and Wiesel, 1962). Previous efforts to use high-dimensional models to capture the diverse properties of cortical receptive receptive fields, such as the Rust-STC model (Rust et al., 2005) or related work based on maximally informative stimulus dimensions (Sharpee et al., 2004), are opaque in connecting the internal components of the models to biological mechanisms. The subunit model presented here provides an accurate account of the data in a more readily interpretable framework.

References

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2:284–299. CrossRef Medline

Ahrens MB, Paninski L, Sahani M (2008) Inferring input nonlinearities in neural encoding models. Network 19:35–67. CrossRef Medline

Alonso JM, Usrey WM, Reid RC (2001) Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. J Neurosci 21:4002–4015. Medline

Barlow HB, Levick WR (1965) The mechanism of directionally selective units in rabbit's retina. J Physiol 178:477–504. CrossRef Medline

Bracewell RN (2000) The Fourier transform and its applications. New York: McGraw-Hill.

Brenner N, Bialek W, de Ruyter van Steveninck R (2000) Adaptive rescaling maximizes information transmission. Neuron 26:695–702. CrossRef Medline

Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci 17:8621–8644. Medline

Cavanaugh JR, Bair W, Movshon JA (2002) Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. J Neurophysiol 88:2530–2546. CrossRef Medline

Chen X, Han F, Poo MM, Dan Y (2007) Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1). Proc Natl Acad Sci U S A 104:19120–19125. CrossRef Medline

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network 12:199–213. CrossRef Medline

De Boer E, Kuyper P (1968) Triggered correlation. IEEE Trans Biomed Eng 15:169–179. CrossRef

Eickenberg M, Rowekamp RJ, Kouh M, Sharpee TO (2012) Characterizing responses of translation-invariant neurons to natural stimuli: maximally informative invariant dimensions. Neural Comput 24:2384–2421. CrossRef Medline

Emerson RC, Bergen JR, Adelson EH (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. Vision Res 32:203–218. CrossRef Medline

Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, Berry MJ 2nd (2006) Selectivity for multiple stimulus features in retinal ganglion cells. J Neurophysiol 96:2724–2738. CrossRef Medline

Freeman J, Simoncelli EP (2011) Metamers of the ventral stream. Nat Neurosci 14:1195–1201. CrossRef Medline

Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. Nat Neurosci 16:974–981. CrossRef Medline

Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36:193–202. CrossRef Medline

Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Vis Neurosci 9:181–197. CrossRef Medline

Hochstein S, Shapley RM (1976) Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. J Physiol 262:265–284. CrossRef Medline

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154. CrossRef Medline

Jarrett K, Kavukcuoglu K, Ranzato MA, LeCun Y (2009) What is the best

multi-stage architecture for object recognition? In: IEEE International Conference on Comput Vis (ICCV), pp 2146–2153. Kyoto, Japan. CrossRef

Jones JP, Palmer LA (1987) The two-dimensional spatial structure of simple receptive fields in cat striate cortex. J Neurophysiol 58:1187–1211. Medline

Kaardal J, Fitzgerald JD, Berry MJ 2nd, Sharpee TO (2013) Identifying functional bases for multidimensional neural computations. Neural Comput 25:1870–1890. CrossRef Medline

Lau B, Stanley GB, Dan Y (2002) Computational subunits of visual cortical neurons revealed by artificial neural networks. Proc Natl Acad Sci U S A 99:8974–8979. CrossRef Medline

LeCun Y, Bengio Y (1995) Convolutional networks for images, speech, and time series. In: The handbook of brain theory and neural networks (Arbib MA, ed.), pp 276–278. Cambridge, MA: MIT.

LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. Neural Computation 1:541–551. CrossRef

Lochmann T, Blanche T, Butts DA (2013) Construction of direction selectivity through local energy computations in primary visual cortex. PLoS One 8:e58666. CrossRef Medline

Marmarelis PZ, Marmarelis VZ (1978) Analysis of physiological systems: the white noise approach. Boston: Springer.

Mechler F, Ringach DL (2002) On the classification of simple and complex cells. Vision Res 42:1017–1033. CrossRef Medline

Movshon JA, Thompson ID, Tolhurst DJ (1978) Receptive field organization of complex cells in the cat's striate cortex. J Physiol 283:79–99. CrossRef Medline

Nishimoto S, Ishida T, Ohzawa I (2006) Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation. J Neurosci 26:3269–3280. CrossRef Medline

Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. Network 15:243–262. CrossRef Medline

Park M, Pillow JW (2011) Receptive field inference with localized priors. PLoS Comput Biol 7:e1002219. CrossRef Medline

Perrett DI, Oram MW (1993) Neurophysiology of shape processing. Image Vision Comput 11:317–333. CrossRef

Pillow JW, Simoncelli EP (2006) Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. J Vis 6:9. CrossRef Medline

Poggio T, Edelman S (1990) A network that learns to recognize 3D objects. Nature 343:263–266. CrossRef Medline

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025. CrossRef Medline

Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal elements of macaque V1 receptive fields. Neuron 46:945–956. CrossRef Medline

Sasaki KS, Ohzawa I (2007) Internal spatial organization of receptive fields of complex cells in the early visual cortex. J Neurophysiol 98:1194–1212. CrossRef Medline

Schwartz O, Pillow JW, Rust NC, Simoncelli EP (2006) Spike-triggered neural characterization. J Vis 6:13. CrossRef Medline

Sclar G, Maunsell JH, Lennie P (1990) Coding of image contrast in central visual pathways of the macaque monkey. Vision Res 30:1–10. CrossRef Medline

Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput 16:223–250. CrossRef Medline

Simoncelli EP, Paninski L, Pillow J, Schwartz O (2004) Characterization of neural responses with stochastic stimuli. Cogn Neurosci 3:327–338.

Steveninck RDRV, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc R Soc B Biol Sci 234:379–414. CrossRef

Touryan J, Lau B, Dan Y (2002) Isolation of relevant visual features from random stimuli for cortical complex cells. J Neurosci 22:10811–10818. Medline

Touryan J, Felsen G, Dan Y (2005) Spatial structure of complex cell receptive fields measured with natural images. Neuron 45:781–791. CrossRef Medline

Victor JD, Shapley RM (1979) The nonlinear pathway of Y ganglion cells in the cat retina. J Gen Physiol 74:671–689. CrossRef Medline

Vintch B (2013) Structured hierarchical models for neurons in the early visual system. PhD Thesis, Center for Neural Science, New York University.

Vintch B, Zaharia AD, Movshon JA, Simoncelli EP (2012) Efficient and direct estimation of a neural subunit model for sensory coding. Adv Neural Inf Process Syst 25:3113–3121. Medline

Wallis G, Rolls ET (1997) Invariant face and object recognition in the visual system. Prog Neurobiol 51:167–194. CrossRef Medline

Young F, De Leeuw J, Takane Y (1976) Regression with qualitative and quantitative variables: an alternating least-squaresmethod with optimal scaling features. Psychometrika 43:279–281.