

# Spectral receptive fields do not explain tuning for boundary curvature in V4

Timothy D. Oleskiw, Anitha Pasupathy, and Wyeth Bair

Department of Biological Structure and National Primate Research Center, University of Washington, Seattle, Washington

Submitted 31 March 2014; accepted in final form 22 July 2014

**Oleskiw TD, Pasupathy A, Bair W.** Spectral receptive fields do not explain tuning for boundary curvature in V4. *J Neurophysiol* 112: 2114–2122, 2014. First published July 23, 2014; doi:10.1152/jn.00250.2014.—The midlevel visual cortical area V4 in the primate is thought to be critical for the neural representation of visual shape. Several studies agree that V4 neurons respond to contour features, e.g., convexities and concavities along a shape boundary, that are more complex than the oriented segments encoded by neurons in the primary visual cortex. Here we compare two distinct approaches to modeling V4 shape selectivity: one based on a spectral receptive field (SRF) map in the orientation and spatial frequency domain and the other based on a map in an object-centered angular position and contour curvature space. We test the ability of these two characterizations to account for the responses of V4 neurons to a set of parametrically designed two-dimensional shapes recorded previously in the awake macaque. We report two lines of evidence suggesting that the SRF model does not capture the contour sensitivity of V4 neurons. First, the SRF model discards spatial phase information, which is inconsistent with the neuronal data. Second, the amount of variance explained by the SRF model was significantly less than that explained by the contour curvature model. Notably, cells best fit by the curvature model were poorly fit by the SRF model, the latter being appropriate for a subset of V4 neurons that appear to be orientation tuned. These limitations of the SRF model suggest that a full understanding of midlevel shape representation requires more complicated models that preserve phase information and perhaps deal with object segmentation.

shape processing; object recognition; ventral visual pathway; macaque monkey; computational model

VISUAL OBJECT PERCEPTION and recognition in primates is based on sensory information processing within the ventral visual pathway (Felleman and Van Essen 1991; Mishkin and Ungerleider 1982). Over the last half-century, studies of the primary visual cortex (V1) have identified local orientation and spatial frequency as the basis dimensions of form representation at the early stages in the ventral pathway (Campbell and Robson 1968; De Valois and De Valois 1990; Hubel and Wiesel 1959, 1965, 1968; Movshon et al. 1978; Schiller et al. 1976). At intermediate stages, in particular area V4, the representation has yet to be firmly established. Neurons in V4 have been shown to be selective for bars of different length, for radial or concentric gratings, for moderately complex shapes, and specifically for the curvature of segments of the bounding contour of shapes (Desimone and Schein 1987; Gallant et al. 1993; Hegdé and Van Essen 2007; Kobatake and Tanaka 1994; Nandy et al. 2013; Pasupathy and Connor 1999, 2001). No single model is widely accepted to account for these observations, but a common approach to explaining extrastriate responses in both the dorsal and ventral pathways is to model

them in terms of selectivity for simple combinations of the features that are represented at earlier levels. This amounts to using weighted combinations of V1-like channels to fit the observed data (Cadieu et al. 2007; David et al. 2006; Rust et al. 2006; Vintch 2013; Willmore et al. 2010). Here we examine whether an instance of this approach, known as the spectral receptive field (SRF) model (David et al. 2006), can account for complex curvature selectivity observed in V4 neurons.

The SRF model describes the tuning of V4 neurons in terms of a weighting function across orientation and spatial frequency bands in the power spectrum of the stimulus (David et al. 2006). This model has the elegant simplicity of combining V1-like signals in a manner that discards phase and thereby produces translation invariance, a key feature of V4 responses (Gallant et al. 1996; Pasupathy and Connor 1999, 2001; Rust and DiCarlo 2010). It has also been argued (David et al. 2006) that the SRF model can account for the ability of V4 neurons to respond to complex shapes in terms of contour features at a particular location within an object-centered reference frame (Pasupathy and Connor 2001). For example, some neurons may respond strongly to shapes with a sharp convexity to the upper right, while others may respond to shapes with a concavity to the left. These patterns of selectivity are well modeled by two-dimensional (2D) Gaussian tuning functions in a space defined by 1) the curvature of the boundary and 2) angular positions relative to object center (Pasupathy and Connor 2001). They are also well modeled by a hierarchical contour template model (Cadieu et al. 2007). Using the previously recorded data set on which both of these models were based, we examine whether the SRF model, the simplest of the three, can account for the contour selectivity observed in V4. We find that there are important features of the data that are not captured by the SRF model.

## MATERIALS AND METHODS

### *Experimental Procedures*

All animal procedures for this study, including implants, surgeries, and behavioral training, conformed to National Institutes of Health and US Department of Agriculture guidelines and were performed under an institutionally approved protocol. The data analyzed here are derived from a previous study (Pasupathy and Connor 2001) and consist of the responses of 109 single, well-isolated V4 neurons in two rhesus monkeys (*Macaca mulatta*) that were recorded while the animals fixated a 0.1° white spot on a computer monitor. After preliminary characterization of the receptive field (RF) location and preferred color of each cell, shape tuning was characterized with a set of 366 stimuli (Fig. 1). Each stimulus was presented in random order without replacement five times for most cells (91/109; 9 cells had 4 repetitions and 9 had 3 repetitions). Response rates were calculated by counting spike occurrences during the 500-ms stimulus presentation period. Spontaneous rates, calculated based on blank stimulus periods

Address for reprint requests and other correspondence: W. Bair, Dept. of Biological Structure and National Primate Research Center, Univ. of Washington, 1959 NE Pacific St., HSB G-520, UW Mailbox-357420, Seattle, WA 98195 (e-mail: wyeth0@uw.edu).

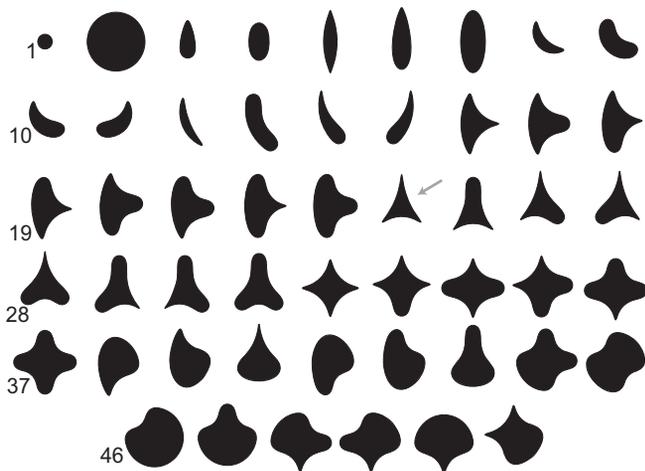


Fig. 1. Set of 51 shapes used to characterize V4 neurons. Each shape was presented at each of 8 orientations, or fewer for shapes with rotational symmetry. For example, the circles (top left) were shown in just 1 conformation because all rotations are identical. Shapes are numbered left to right, top to bottom, starting with 1 at top left. Gray arrow marks shape 24, referred to in RESULTS.

interspersed randomly during stimulus presentation, were subtracted from the average response rate for each stimulus.

*Stimulus Design and Representation*

Stimulus design is described in detail by Pasupathy and Connor (2001). Briefly, stimuli were constructed by systematic combination of 4–8 contour segments each of which took 1 of 5 curvature values, resulting in 51 shapes (Fig. 1). To create radial variation, each shape was rotated by 8 increments of 45°, discarding duplications due to rotational symmetry. Shape stimuli were presented in the center of the RF of the cell under study and were sized such that all parts of the stimuli were within the estimated RF of the cell. Specifically, the outermost stimulus edges were at a distance of 3/4 of the RF radius, which was estimated based on the reported relationship between eccentricity and RF size (Gattass et al. 1988).

For modeling and fitting, each shape was generated as a discretized binary mask of 128 × 128 pixels and then convolved with a Gaussian filter of standard deviation 1 pixel (e.g., Fig. 2A). This image represents a 5° × 5° patch of the visual field to approximate the experi-

mentally used resolution (Pasupathy and Connor 2001). The cutoff frequency of this representation is 12.8 cyc/° (half of the 25.6 pixels/° resolution). Because the typical stimulus size was ~3° diameter in the electrophysiology study of Pasupathy and Connor (2001), we made the largest stimulus have a diameter of ~75 pixels within the 128-pixel field. Fourier transforms of stimulus images were computed with a 2D FFT algorithm. The magnitude of complex-valued Fourier components was subjected to the transformation  $t(x) = \log(x + 1)$  to attenuate the low-frequency power that is largely similar across all shapes (Fig. 2B). Because of the limited number of stimuli and trial repetitions, power spectra were downsampled to reduce the number of dimensions in the representation to facilitate model fitting (see below). Specifically, a spectral power sample (Fig. 2C) was created by summing over 7 × 7-pixel blocks within the spectrum, with the middle block centered on the DC bin, to achieve a 17 × 17 grid (the extra few pixels at the margins were ignored). This limited our frequency representation to 0–12 cyc/°, which exceeds the range used in a comparable study (David et al. 2006). Because of the even symmetry of the power spectrum, this resulted in a 17 × 9-pixel representation (as depicted in Fig. 2C), denoted  $P = \{P_s\}_{s \in S}$ , where the set of all shapes is denoted  $S$  ( $|S| = 366$ ). Overall, the aims of this representation were 1) to approximate the methods used during the original data recording, 2) to reduce the number of parameters to be fit (17 × 9, given the symmetry in the power spectra), and 3) to represent the vast majority of the frequency range that would be available to the visual cortex at the relevant eccentricities.

*Models*

*Spectral receptive field.* As proposed by David et al. (2006), an SRF model performs a linear combination of the spectral power of the stimulus in discrete bands to predict neural activity. Using the spectral power sample,  $P_s$ , of each shape and observed neuronal responses,  $r_s$ , the SRF model seeks a set of weights,  $\Phi^{SRF}$ , to minimize the residual error between model prediction  $P\Phi^{SRF}$  and  $r$ . Finding such a template can thus be cast as a linear least-squares optimization, i.e.,

$$\Phi^{SRF} = \underset{\Phi}{\operatorname{argmin}}(\|P\Phi - r\|) \tag{1}$$

where  $\|\cdot\|$  denotes the standard Euclidean norm. For procedural convenience, stimulus power spectra are encoded by a 153-element vector representing the coefficients of the 17 × 9 sampling of spectral power. As neural responses to 366 shape stimuli are considered,  $P$  is a 366 × 153 matrix. Vectors  $\Phi$  and  $r = \{r_s\}_{s \in S}$  are of 153 × 1 and 366 × 1 elements, respectively.

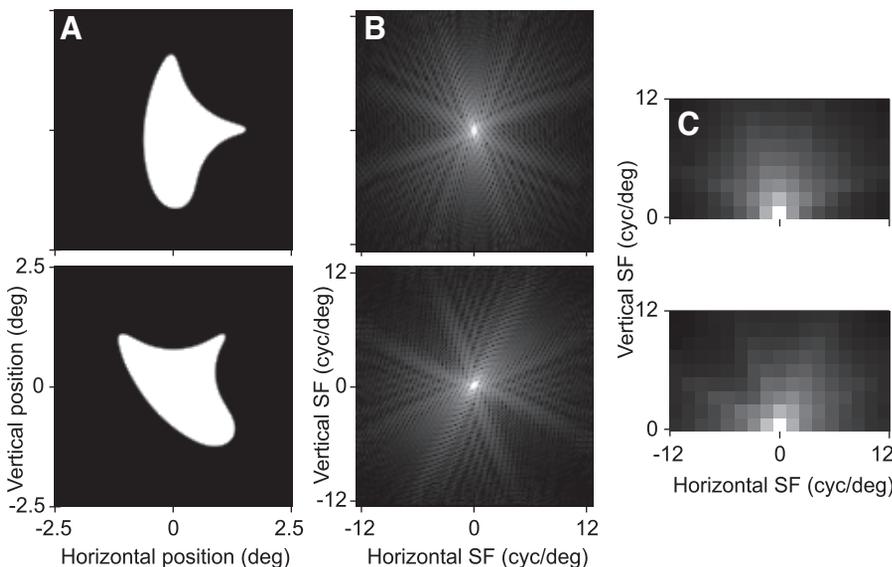


Fig. 2. Shape stimuli and their spectral power representation. A: 2 rotations of shape 18 (Fig. 1) are shown at 128 × 128-pixel resolution. B: log of the Fourier power spectra of A. C: downsampled spectral power representation of B used for fitting. SF, spatial frequency.

Because of the high ratio of model parameters to stimuli and correlations among stimuli, the matrix  $P$  is ill-conditioned, making standard least squares prone to overfitting. To correct for this, we used Tikhonov regularization (Press et al. 2007), i.e.,

$$\Phi^{\text{SRF}} = \underset{\Phi}{\operatorname{argmin}} (\|P\Phi - r\| + \lambda \|\Phi\|) \quad (2)$$

in place of Eq. 1, where  $\lambda$  denotes the regularization factor. We tested values of  $\lambda$  from 0.01 to 100, using 100 points that were evenly spaced on a log scale. The data were divided into 100 randomly chosen partitions of 75% training and 25% test data. Each partition was used to fit and test the model at each  $\lambda$  value. At each  $\lambda$ , we computed  $M_{\text{test}}^{\lambda}$  and  $M_{\text{train}}^{\lambda}$ , the average explained variance across all partitions in the testing and training data, respectively. For each cell, we defined  $\lambda'$  to be the value of  $\lambda$  that maximized  $M_{\text{test}}^{\lambda}$  and then defined the training and testing performance to be  $M_{\text{test}}^{\lambda'}$  and  $M_{\text{train}}^{\lambda'}$ , respectively. To verify that these methods were sufficient to reveal SRF maps like those reported previously (David et al. 2006), we simulated SRFs having a variety of sizes and shapes, tested them with the same shape set used in the electrophysiology study, and confirmed that we could recover the simulated fields.

**Angular position and curvature.** Pasupathy and Connor (2001) proposed an angular position and curvature (APC) model that performs a nonlinear computation over stimuli represented as a set of 4–8 points in the 2D space of angular position,  $\theta$ , and contour curvature,  $\kappa$ . Neural responses are predicted by evaluating a 2D Gaussian energy function (Von Mises in  $\theta$ ) at each of these points and taking the maximum. In particular,  $s_i = (\theta_i, \kappa_i)$  denotes the points defining a shape stimulus  $s$  for  $i = (1, \dots, I_s)$ , where  $I_s$  is the number of points. An APC model seeks the energy function parameters  $\Phi^{\text{APC}} = (\alpha, \mu_{\theta}, \sigma_{\theta}, \mu_{\kappa}, \sigma_{\kappa})$  that minimize the error with respect to the observed neural responses  $r_s$ . The APC model is fit through nonlinear optimization, i.e.,

$$\Phi^{\text{APC}} = \underset{(\alpha, \mu_{\theta}, \sigma_{\theta}, \mu_{\kappa}, \sigma_{\kappa})}{\operatorname{argmin}} \sum_{s \in S} \left[ \max_{i \in I_s} \left\{ \alpha e^{\frac{\cos(\theta_i - \mu_{\theta})}{\sigma_{\theta}} - \frac{(\kappa_i - \mu_{\kappa})^2}{\sigma_{\kappa}^2}} \right\} - r_s \right]^2. \quad (3)$$

Unlike SRF modeling, a global optima cannot be found deterministically. We estimated the optimal model parameters by performing gradient descent on the objective function. To avoid locally optimal solutions, descent was repeatedly conducted from random initializations ( $n = 100$ ) sampled from a uniform distribution over the angular position and curvature parameter space. Simulations reveal that global optima are consistently well approximated after only a few repeated descents.

Because responses of many V4 neurons depend on the curvature of three adjoining contour segments centered at a specific angular position (Pasupathy and Connor 2001), we also considered an APC model that includes three curvature dimensions and a single angular position dimension. We refer to this as the 4D APC model to distinguish it from the 2D APC model described above. The 4D APC model has nine parameters, which include the four additional parameters for the means and SDs of the Gaussian functions describing the two adjoining curvature dimensions. We used the same 75%/25% data partition scheme for fitting and testing our APC models as described above for the SRF model.

## RESULTS

The results are organized in two sections. We first examine whether there is direct evidence for the SRF model by testing a specific prediction that it makes about responses to stimuli subject to a 180° rotation. We then compare the ability of the curvature model and the SRF model to capture variance in the data and examine whether the two models are equally good at explaining tuning for boundary curvature.

### Response Similarity for 180° Stimulus Rotation

The SRF model predicts responses of V4 neurons on the basis of the spectral power coefficients of the visual stimuli; therefore, any SRF-like neuron would naturally yield equivalent responses, up to noise, to stimuli having identical power spectra. It turns out that any stimulus rotated by 180° has the same spectrum as the original stimulus. This follows intuitively because any visual stimulus can be described by its Fourier (sine and cosine) components and these components do not change their orientation, spatial frequency, or amplitude when rotated 180° in the spatial domain. Formally, denoting the Fourier transform  $F$  of a 2D shape image  $f$  as

$$F f(x, y) = \hat{f}(\xi, \zeta), \quad (4)$$

the spectral power of a 180° rotation of  $f$ , denoted  $f_R$ , is equal to the spectral power of  $f$ , i.e.,

$$\begin{aligned} |F f_R(x, y)|^2 &= |F f(-x, -y)|^2 \\ &= |\hat{f}(-\xi, -\zeta)|^2 \\ &= |\overline{\hat{f}(\xi, \zeta)}|^2 \\ &= \overline{\hat{f}(\xi, \zeta)} \cdot \hat{f}(\xi, \zeta) \\ &= |F f(x, y)|^2. \end{aligned} \quad (5)$$

The second step above follows from the time reversal property of the Fourier transform. The third step follows because the Fourier transform of a real-valued function is Hermitian (overbar denotes the complex conjugate), and the fourth and fifth steps simply apply the definition of the squared norm as the product of a complex value and its conjugate, e.g.,  $|y|^2 = y\bar{y}$ . This prediction of the SRF model, that neurons will respond the same to a shape and its 180° rotation, is counterintuitive in light of findings that many V4 neurons are tuned for the angular position of stimulus features around the boundary of a shape (Pasupathy and Connor 2001), the latter being a property that is grossly changed by 180° rotation. For example, if a neuron is tuned for a sharp convexity to the right, it would respond strongly to a shape such as that in Fig. 2A, *top*, but not to the 180° rotation of that shape (not shown).

To test this prediction of the SRF model, we identified all pairs of shapes in our stimulus set that were 180° rotations of each other. For example, the shape in Fig. 2A was presented at 8 rotations and thus contributed 4 such 180° rotation pairs. We assessed the amount of correlation,  $r_{180}$  (Pearson's  $r$  value) in these paired responses for each cell; data for three example cells are depicted in Fig. 3 (see legend for details). The first example cell (*b1601*; Fig. 3A) shows positive correlations for 180° rotations, *cell a8602* (Fig. 3B) shows no correlation, and *cell b2601* (Fig. 3C) shows anticorrelation. The first example would appear to be consistent with the idea that responses are similar for 180° rotations, whereas the third clearly contradicts this notion, suggesting that if a shape produces a larger than average response, its 180° rotation typically does not. However, the observed correlation must be interpreted relative to the amount of correlation between spectrally dissimilar stimuli, i.e., non-180° rotation pairs. To calculate this baseline correlation,  $r_{\text{baseline}}$ , we chose 4 of the 24 possible non-180° pairings at random for each shape (where 8 rotations were presented)

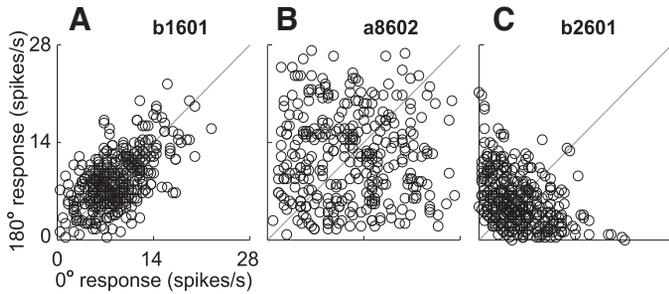


Fig. 3. Comparing responses to a shape and its 180° rotation. Mean responses to a shape and its 180° rotation are plotted against each other for 3 example neurons. Data points are derived from the 42 shapes for which 8 rotations were presented. Each 180° rotation pair contributes two points, (x,y) and (y,x), for a total of 336 points per neuron. The spectral receptive field (SRF) model predicts that  $y = x$ , up to noise, and thus a high positive correlation coefficient is expected. Nevertheless, our population contained neurons with positive correlation ( $r = 0.54$ ; A), no correlation ( $r = 0.00$ ; B) and negative correlation ( $r = -0.39$ ; C). *b1601*, *a8602*, and *b2601* indicate neuronal ID, where the first letter indicates the animal ID.

and calculated the bootstrap distribution of  $r$  values (Fisher  $z$ ) from repeated simulations ( $n = 100$ , which proved to be convergent). Figure 4A shows an example (*cell a6802*) in which the response correlation for 180° rotations is significantly positive ( $P < 0.05$ ) but not different from the correlation of non-180° pairings (Fig. 4B). It turns out that many cells show a positive baseline correlation because they respond better to some shapes than others regardless of orientation. This can arise simply from shapes that have similar attributes repeated along their boundaries (e.g., Fig. 1, *shape 24*) or from sensitivity to attributes that are not changed by rotation, such as surface area.

The population results of this analysis for the data set of 109 cells are shown in Fig. 5A, where  $r_{180}$  is plotted against  $r_{\text{baseline}}$ . The significance level is set at  $2\sigma$  of baseline correlation. Note that most neurons ( $n = 68$ ) lie near the line of equality, e.g., *a6802* (from Fig. 4; *point 4* in Fig. 5A). Interestingly, some cells, e.g., *b1601* (from Fig. 3A; *point 1* in Fig. 5A), fall significantly above equality, indicating possible selectivity for

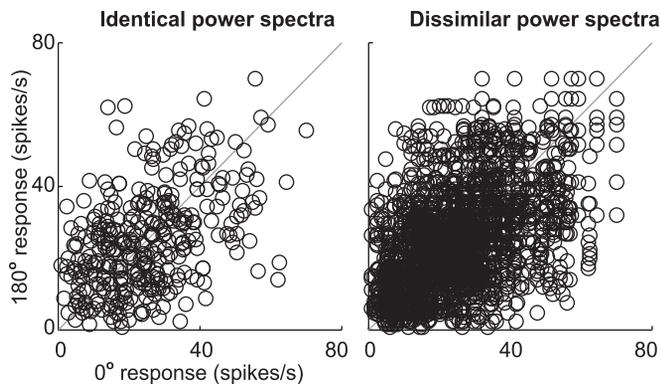


Fig. 4. Computing baseline correlation between non-180° shape rotations. A: mean responses of *neuron a6802* to pairings of stimuli that are spectrally identical, i.e., 180° rotation, (same analysis and plot format as in Fig. 3). B: similar to A, but responses are plotted for all pairings for a given shape that are not 180° rotations. There are 24 such pairings for each shape with 8 rotations, compared with only 4 pairings for 180° rotations. Correlation coefficients are similar in A and B,  $r = 0.46$  and  $r = 0.47$ , respectively, suggesting that for this neuron average responses are higher for some shapes than for others, regardless of rotation.

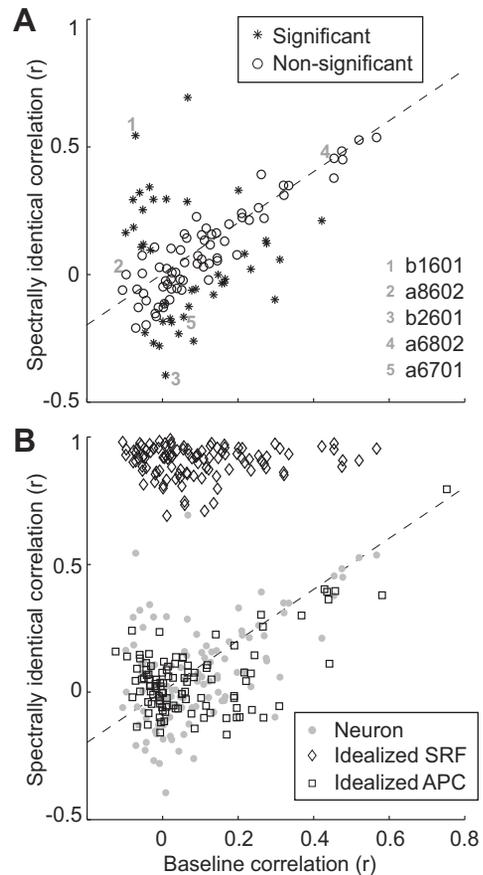


Fig. 5. Comparing response correlation for 180° rotations to baseline correlation. A: for each of the 109 neurons, the correlation coefficient for responses to spectrally identical shapes (180° rotations) is plotted against the baseline correlation value (the correlation between responses to non-180° pairings, see RESULTS). Points fall on both sides of the line of equality (dashed line). Points plotted as asterisks indicate neurons for which the  $y$ -axis deviates from the  $x$ -axis by  $>2$  SD, based on a bootstrap estimate of the baseline correlation distribution. Points numbered 1–3 correspond to the 3 example neurons in Fig. 3, *point 4* corresponds to the example in Fig. 4, and *point 5* corresponds to an example neuron shown below in RESULTS. B: the data from A are replotted (gray filled circles) and are compared to predictions for an idealized SRF model and an idealized angular position and curvature (APC) model for each of the 109 neurons. As expected, points for the idealized SRF prediction have much higher correlation for spectrally identical stimuli, because the SRF model predicts identical mean responses for such stimuli.

features that are preserved across 180° rotations and are potentially consistent with an SRF model.

We compared the scatter of data in Fig. 5A to that expected from an idealized SRF model that includes realistic (Poisson) noise. We did this by setting an underlying mean firing rate (target rate) for each shape and then deriving from it a measured rate by sampling a spike count from the target rate five times with Poisson statistics (variance equal to mean). To embody the SRF model, we set the target rates equal for pairs of shapes that were 180° rotations, choosing randomly between the two experimentally observed rates. From these measured rates, we computed  $r_{180}$  as described above. We repeated this process 100 times and determined the average correlation (using Fisher  $z$ ). In Fig. 5B, the results of this statistical simulation are plotted together with the actual data and against the same  $r_{\text{baseline}}$  values. The results indicate that hypothetical SRF units show much higher values of  $r_{180}$  than were observed

in our data. This suggests that, while a few cells (e.g., *neuron b1601*) show consistency with the SRF model, the vast majority of neurons from our population do not.

We performed a similar simulation using the response rates predicted by the APC model (see MATERIALS AND METHODS) for comparison to the neuronal data and to the responses of the idealized SRF model. Each cell was fit to the APC model (see *Model Fitting and Performance*), and the resulting predicted mean responses were used as the target rates. Observed rates were computed from the average of five Poisson samples. The result (Fig. 5B) shows that the APC model predicts a much lower  $r_{180}$  value than the SRF model, and that the predicted values are approximately consistent with the range of values found for the neurons.

In summary, the SRF model makes a distinct prediction about  $180^\circ$  rotations that the APC model does not, and with respect to this prediction the SRF model is far less consistent with our data than the APC model.

### Model Fitting and Performance

Although the SRF model fails to predict the differences in neuronal responses to shapes and their  $180^\circ$  rotations, previous reports show that both the SRF and the APC models account for only part of the variance of V4 responses (Pearson's  $r$  values of 0.32 for the SRF model of David et al. 2006 and 0.57 for the 4D APC model of Pasupathy and Connor 2001). We thus wanted to establish 1) what fraction of the variance is captured by the SRF model across the entire set of shapes, and how this compares to that previously reported for the SRF and APC models, and 2) whether the cells that are well fit by the SRF model, in terms of amount of explained variance, are also the ones that are well fit by the APC model.

We performed an empirical evaluation of both SRF and APC models by fitting to, and predicting, recorded neural responses to our stimuli. We partitioned our data into training and testing sets for cross-validation, and we measured model performance in terms of explained variance ( $r^2$ ) for both sets. Bootstrap validation estimates (Fig. 6A) show that although the SRF model outperforms both APC models across training data sets, it underperforms both the 2D and 4D versions of the APC model on the test data sets. This is a hallmark of overfitting: the SRF model has  $\sim 30$  times the number of parameters ( $9 \times 17$

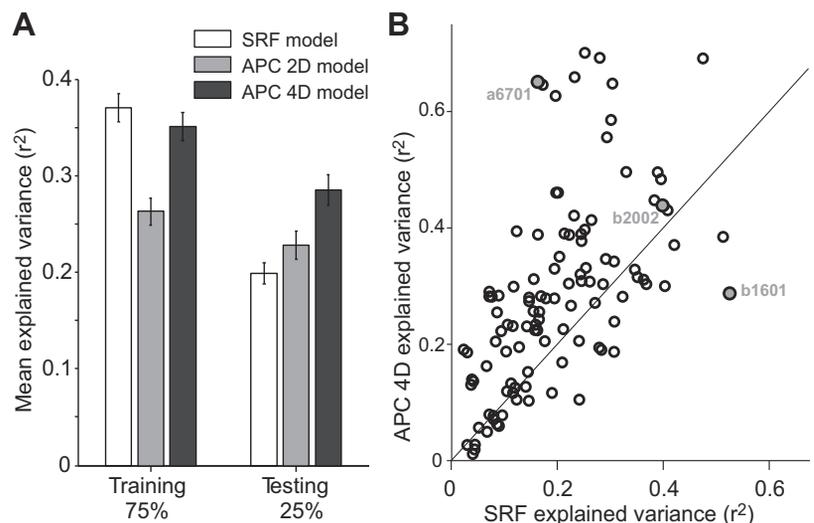
spectral weights) compared with the 2D APC model (5 parameters) and 17 times that of the 4D APC model (9 parameters). When comparing only the testing validation performance across all neurons (Fig. 6B), the responses of the majority of neurons (77 of 109) are better predicted by the 4D APC model than the SRF model, with a significantly higher average explained variance (mean 0.09, SD 0.13, paired  $t$ -test,  $P < 0.0001$ ).

Although the performance of the SRF model was relatively weak, this does not appear to reflect particular limitations of our stimulus set, because the performance was favorable to, and in fact better than, that reported previously (David et al. 2006): our mean  $r$  value was 0.43 ( $n = 109$ ), compared with their mean of 0.32.

Another important feature of the scatter in Fig. 6B is the paucity of points near the upper right corner. This implies that the neurons best explained by the APC model are not also those that are best explained by the SRF model. For example, *neuron b1601* (Fig. 6B, bottom right) was among the most SRF-like cells: its responses were best fit by the SRF model ( $r^2 = 0.54$ ) and were also among the most consistent with the prediction regarding  $180^\circ$  rotations examined above (Fig. 5A), but its responses were not well explained by the APC model ( $r^2 = 0.2$ ). On the other hand, points do fall near the extreme lower left in Fig. 6B, representing neurons that are poorly fit by both models. This is expected under the simple assumption that some neurons do not respond well to the stimulus set, or have very noisy responses. Discarding the neurons that were poorly fit by either model ( $r^2 < 0.15$ ), there was no significant correlation between the explained variance of the APC and SRF models ( $r = 0.17$ ,  $P = 0.17$ ,  $n = 65$ ). This suggests that these distinct models do not capture the same features of the response.

To understand the tuning properties of neurons that were well fit by the SRF model and compare them to those that were well fit by the APC model, it is useful to examine the raw responses and fit parameters for several example neurons. The responses of the SRF-like neuron, *b1601*, for each of the 366 shapes are plotted in Fig. 7A, where red indicates the strongest responses and blue the weakest. This neuron tended to respond most strongly to shapes that were oriented horizontally, and the strongest responses were often offset in the diagram by 4 rows,

Fig. 6. Fit performance of SRF and APC models. *A*: mean explained variance is plotted for the SRF model and the 2D and 4D versions of the APC model for training data (*left*) and testing data (*right*). The training and testing partitions were 75% and 25% of the data, respectively. The SRF model explained more variance in the training data than the APC models, but both APC models outperformed the SRF model on the testing data on average. Error bars show SE. *B*: explained variance values for the 4D APC model are plotted against those for the SRF model for all 109 neurons. Three examples are indicated: *b1601*, which was better fit by the SRF model; *a6701*, which was better fit by the APC model; and *b2002*, which was about equally well fit by both models.



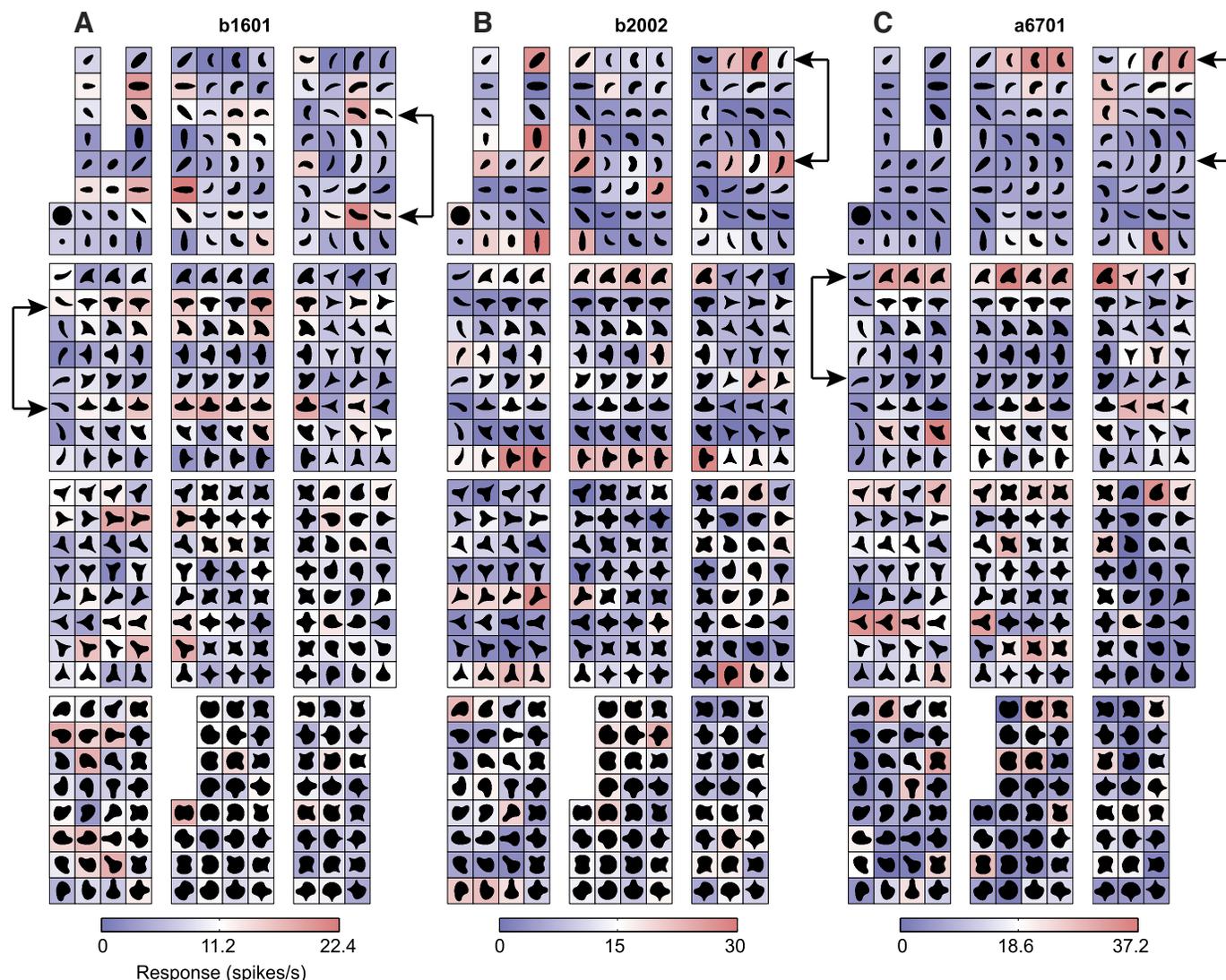


Fig. 7. Shape tuning maps for 3 example neurons. *A*: mean firing rate of neuron *b1601* to each shape (drawn in black) is indicated by the color surrounding the shape. Dark blue and dark red indicate the lowest and highest responses, respectively (see scale bar at bottom). This neuron responded best to shapes that had a horizontal alignment, and 180° rotations of the same shape often gave roughly similar responses (black arrow pairs). All rotations (up to 8) of each shape are arranged contiguously within a single column in 1 block. *B*: responses for example neuron *b2002*, which tended to prefer shapes with a vertical or right-leaning alignment. Sometimes responses to 180° rotations were similar (e.g., black arrow pair). *C*: responses for example neuron *a6701*, which was well fit by the APC model and poorly fit by the SRF model (Fig. 6*B*). Shapes associated with the strongest responses did not elicit strong responses when rotated by 180° (compare top and bottom arrows in each arrow pair).

which corresponds to 180° of stimulus rotation. We will see below (Fig. 8*A*) that the SRF map for this neuron reflects this apparent preference for horizontal orientation. A second example neuron (Fig. 7*B*) that was moderately well fit by both models (*b2002* in Fig. 6*B*) responded strongly to stimuli that were oriented vertically or tilting somewhat toward the right. Here some but not all of the stimuli evoking the strongest responses were separated by 180°, consistent with the moderate fit of the SRF model. A contrasting example (Fig. 7*C*) shows a neuron that did not display a clear preference for overall orientation. In particular, the strongest responses are not separated by 180° rotations, consistent with the poor fit of the SRF model (*a6701* in Fig. 6*B*). All of the shapes that evoke strong responses from this cell include a concavity to the right side of the shape. This type of tuning is well captured by the APC model, as indicated by the relatively high explained variance value (*a6701* in Fig. 6*B*).

The SRF maps for the example neurons just described are shown in Fig. 8. As described in MATERIALS AND METHODS, we fit SRF maps over a broad range of regularization values,  $\lambda$ , computing training and test performance at each value to assess and minimize the influences of overfitting. For neuron *b1601* (Fig. 8*A*, top), the training performance declined with increasing  $\lambda$  while the testing performance increased to a maximum and subsequently fell to an asymptote. This behavior is expected, and it held for all neurons (Fig. 8*D* shows population average). For each neuron, SRF maps are shown (below the performance plots) for low, optimal (highest test performance), and high regularization values. Each map shows spectral weights as a function of horizontal and vertical spatial frequency. In this representation, frequency increases with distance from the origin, and power at a particular orientation lies along a line radiating from the origin. At low  $\lambda$  (Fig. 8, *A–C*, bottom), the maps have a salt-and-pepper appearance that fits

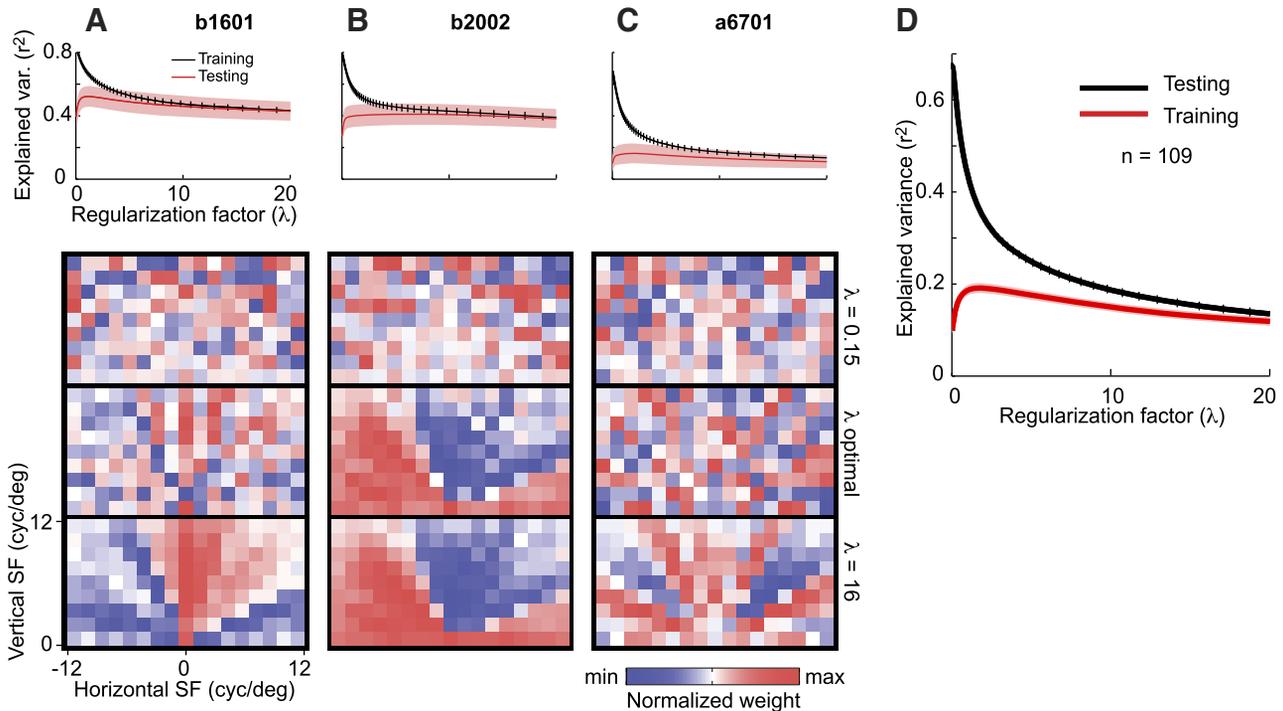


Fig. 8. SRF maps depend on regularization. SRF maps (image panels) are shown for the 3 example cells of Fig. 7 (A–C) and for 3 levels of the regularization parameter (low to high, top to bottom, respectively). Red indicates positive weights, and blue indicates negative weights (see scale bar near *bottom* of C). Top row shows maps for low  $\lambda$  (0.15). These maps produce the best performance on the training data (black line, top panels, described below) but substantially worse performance on the test data (red line, top panels) because of overfitting. Second row shows maps at the optimal  $\lambda$  (best test performance) for each neuron ( $\lambda = 1.52, 5.09,$  and  $0.38$  for A–C, respectively). Third row shows maps for high  $\lambda$  (16). Each map shows an example of the SRF given a random selection of training/testing partition (75/25%). Top panels plot the average performance (across 100 random training/test partitions) on the training and test data as a function of  $\lambda$ . Shaded area shows SD. D: average performance on the training and test data as a function of  $\lambda$  across all 109 neurons. Shaded area shows SD.

the training data well, but they strongly underperform on the testing data and thus are not likely to reflect a true receptive field. At high  $\lambda$  (here = 16, but maps were similar over a broad range), the training and test performance become nearly equal, suggesting that the features remaining in the maps are those that best generalize beyond the training set. Indeed, the  $\lambda = 16$  map for *neuron b1601* (Fig. 8A, *bottom*) has a red streak along the vertical axis, indicating a preference for horizontal orientation, which is apparent in Fig. 7A. The high- $\lambda$  map for *neuron b2002* (Fig. 8B, *bottom*) has a red streak along the horizontal axis that expands upward in the left quadrant, indicating a preference for vertical to right-leaning orientation, as observed in Fig. 7B. In contrast, the SRF map for *neuron a6701* (Fig. 8C, *bottom*) has red streaks at multiple orientations, and, most notably, the performance (Fig. 8C, *top*) is substantially lower at all  $\lambda$  compared with the first two examples.

In summary, the correspondence between the coherent structure within the SRF maps (Fig. 8) and the raw shape responses (Fig. 7) suggests that our SRF fits provide a useful characterization for some neurons, but that these neurons also appear to be ones that display sensitivity to the overall orientation of a shape.

## DISCUSSION

We examined whether the selectivity of V4 neurons for boundary curvature can be simply explained in terms of tuning for the spatial frequency power spectrum as quantified by the SRF model. We found that the responses of curvature-tuned V4 neurons are inconsistent with the SRF model on several

counts. First, the SRF model predicts identical responses to 180°-rotated stimuli, but most V4 neurons, especially those that are curvature tuned, do not exhibit this property. Second, compared with the curvature-based model, the SRF model captured significantly less of the variance in V4 responses for a set of parametrically designed 2D complex shapes. Finally, the V4 neurons that were particularly well fit by the SRF model were also those that could be roughly described as showing simple orientation tuning, and were not among the best fit by the curvature model.

A previous attempt to show that the SRF model could unify V4 neuronal selectivity from studies using disparate stimulus sets (David et al. 2006) was motivated by several attractive features of the model. The SRF model describes V4 tuning in terms of sensitivity to particular frequency bands within the power spectrum of the visual input. Because the frequency bands can be labeled in terms of orientation and spatial frequency, the SRF model can be viewed as a simple extension of the representation present in V1, where neurons are tuned to stimulus orientation (Hubel and Wiesel 1968) and spatial frequency (Albrecht et al. 1980; Campbell et al. 1969; De Valois and De Valois 1990; Movshon et al. 1978). This has the advantage that the circuit implementation of a V4 neuron in terms of the SRF model would be a relatively straightforward combination of V1 outputs. Another key feature of the SRF model is the second-order nonlinearity inherent to the power spectrum that discards phase information and can thereby produce phase- and position-invariant responses, approximating similar characteristics of V4 neurons (Gallant et al. 1996;

Pasupathy and Connor 1999, 2001; Rust and DiCarlo 2010). However, the simplification of discarding phase information before integrating across frequency bands ignores a key feature of V4 curvature selectivity. Specifically, a V4 neuron may respond preferentially to a sharp convexity pointing upward relative to the object center but not to that same feature pointing downward; the SRF model cannot reproduce this important aspect of curvature tuning because phase-insensitive Fourier power models predict identical responses for pairs of stimuli that are 180° rotations of each other. We directly examined the responses of V4 neurons to such pairs of stimuli and found that this prediction did not hold, in contradiction to the SRF model. We conclude that a defining characteristic of the SRF model—that phase information is dropped before combining spatial frequency components across the image—is inconsistent with curvature selectivity in V4.

Because all current models of V4 have limitations, it is important to consider how the SRF model compares to alternatives in its ability to explain the variance of neuronal responses to the same stimulus set. We fit SRF maps to V4 responses to a set of simple shapes that parametrically explored a space of contour curvature and angular position. Our SRF maps were roughly consistent with those reported previously (David et al. 2006; see their Figs. 1–3). Our maps often showed tuning for multiple orientations, similar to theirs, and our maps explained a larger fraction of the response variance than their maps did. One difference was that their spatial resolution was 12 cyc/RF, whereas ours was about three times higher (12 cyc/°, with typical RF sizes  $\sim 3^\circ$ ). Nevertheless, the SRF model captured less response variance on average than our APC model, which had far fewer parameters. Two observations are particularly worth noting. First, none of the cells best fit by the curvature model (20 cells for which  $r^2 > 0.4$ , 4D APC model) was better fit by the SRF model. This suggests that the SRF model does not capture the key features of curvature selectivity that are represented in the curvature model. Second, a closer examination of the cells best fit by the SRF model reveals that they would be well described as orientation selective, consistent with examples of David et al. (their Figs. 1b and 3a). Thus the SRF model does not provide a sufficient framework to understand curvature tuning in V4; nevertheless, it may serve an important role in describing cells in V4 whose tuning is largely in the orientation dimension. Future work will be required to understand how these different types of tuning operate together in V4.

Although the contour curvature model provides a good fit to the responses of many V4 neurons, it has the limitation of being a descriptive model and does not point to any obvious implementation in terms of biologically plausible circuitry. One model to derive curvature selectivity in V4 from inputs coming from V1 and V2 would involve first coarsely defining an object, i.e., segmenting it, and then assessing the orientation progression along its boundary. The latter step is captured by the model of Cadieu and colleagues (discussed below). The former step, segmentation, is more challenging but could be achieved by a set of “grouping cells” like those proposed by Craft et al. (2007) as a mechanism for creating border ownership signals in V2. Grouping cells group together concentric contour segments, and a set of such cells captures the coarse shape of an object. This is equivalent to finding the set of largest disks that would just fit within a bounding contour, a

method proposed for computing the medial axis of a shape (Blum 1967). Grouping cells are hypothesized in V4 and could send lateral connections to curvature-sensitive neurons. Inputs from the set of grouping cells would specify the centroid of the stimulus in a graded fashion. Further experiments are needed to explore this possibility, but preliminary results from our laboratory suggest that the earliest responses in V4 encode the overall size of the stimulus, which supports this hypothesis.

Alternatives to the APC and SRF models considered here include a set of biologically inspired hierarchical models (Cadieu et al. 2007; Rodríguez-Sánchez and Tsotsos 2012; Serre et al. 2005). The model of Cadieu et al. has been shown to account for the curvature tuning of V4 neurons using the same data set examined here—Fig. 10A of Cadieu et al. (2007) shows that their model performed similarly to the 4D APC model in terms of explained variance. The Cadieu model, however, does not operate in an object-centered system and does not explicitly represent curvature. Curvature is built up as a combination of oriented segments, and translation invariance is achieved in small steps of positional invariance implemented by using the max-function. The model of Rodríguez-Sánchez and Tsotsos (2012) explicitly represents curvature tuning at intermediate stages in the visual hierarchy and implicitly uses an object-centered coordinate system. These models may provide a useful foundation for testing the nature of an object-centered representation and for developing a more complete model that encompasses novel recent findings related to object segmentation in V4 that have yet to be modeled (Bushnell et al. 2011).

In conclusion, it is essential to seek out the simplest models, and the SRF model is therefore an important point of comparison. However, responses of V4 neurons appear to reflect the solutions to some of the most difficult problems in visual object recognition, those of translation invariance and object segmentation, so it may be unsurprising if simple combinations of V1 outputs do not account for V4 responses. To advance our understanding of V4, it will be important to 1) develop a mechanistic implementation that explains curvature responses, 2) extend such models to handle complex scenes, and 3) conduct experiments to further characterize those V4 neurons that are not well explained by either the APC or SRF models.

## GRANTS

This work was funded by National Institutes of Health (NIH) Grant R01 EY-018839 (A. Pasupathy), National Science Foundation CRCNS Grant IIS-1309725 (W. Bair and A. Pasupathy), and NIH Office of Research Infrastructure Programs Grant RR-00166 (A. Pasupathy). T. D. Oleskiw was funded by NIH (Computational Neuroscience Training Grant 5R90 DA-033461-03) and by the Natural Sciences and Engineering Research Council of Canada (NSERC, PGS-D).

## DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

## AUTHOR CONTRIBUTIONS

Author contributions: T.D.O., A.P., and W.B. conception and design of research; T.D.O., A.P., and W.B. analyzed data; T.D.O., A.P., and W.B. interpreted results of experiments; T.D.O., A.P., and W.B. prepared figures; T.D.O., A.P., and W.B. drafted manuscript; T.D.O., A.P., and W.B. edited and revised manuscript; T.D.O., A.P., and W.B. approved final version of manuscript.

## REFERENCES

- Albrecht DG, De Valois RL, Thorell LG.** Visual cortical neurons: are bars or gratings the optimal stimuli? *Science* 207: 88–90, 1980.
- Blum HA.** Transformation for extracting new descriptors of shape. In: *Models for the Perception of Speech and Visual Form*, edited by Wathen-Dunn W. Cambridge, MA: MIT Press, 1967, p. 362–380.
- Burkhalter A, Van Essen DC.** Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *J Neurosci* 6: 2327–2351, 1986.
- Bushnell BN, Harding PJ, Kosai Y, Pasupathy A.** Partial occlusion modulates contour-based shape encoding in primate area V4. *J Neurosci* 31: 4012–4024, 2011.
- Cadiou C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T.** A model of V4 shape selectivity and invariance. *J Neurophysiol* 98: 1733–1750, 2007.
- Campbell FW, Cooper GF, Enroth-Cugell C.** The spatial selectivity of the visual cells of the cat. *J Physiol* 203: 223–235, 1969.
- Campbell FW, Robson JG.** Application of Fourier analysis to the visibility of gratings. *J Physiol* 197: 551–566, 1968.
- Craft E, Schütze H, Niebur E, von der Heydt R.** A neural model of figure-ground organization. *J Neurophysiol* 97: 4310–4326, 2007.
- David SV, Hayden BY, Gallant JL.** Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* 96: 3492–3505, 2006.
- Desimone R, Schein SJ.** Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J Neurophysiol* 57: 835–868, 1987.
- De Valois RL, De Valois KK.** *Spatial Vision*. Oxford, UK: Oxford Univ. Press, 1990.
- Felleman DJ, Van Essen DC.** Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1: 1–47, 1991.
- Gallant JL, Braun J, Van Essen DC.** Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259: 100–103, 1993.
- Gallant JL, Connor CE, Rakshit S, Lewis JW, Van Essen DC.** Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76: 2718–2739, 1996.
- Gattass R, Sousa AP, Gross CG.** Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8: 1831–1845, 1988.
- Hegd  J, Van Essen DC.** A comparative study of shape representation in macaque visual areas V2 and V4. *Cereb Cortex* 17: 1100–1116, 2007.
- Hubel DH, Wiesel TN.** Receptive fields of single neurones in the cat’s striate cortex. *J Physiol* 148: 574–591, 1959.
- Hubel DH, Wiesel TN.** Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol* 28: 229–289, 1965.
- Hubel DH, Wiesel TN.** Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195: 215–243, 1968.
- Kobatake E, Tanaka K.** Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867, 1994.
- Mishkin M, Ungerleider LG.** Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav Brain Res* 6: 57–77, 1982.
- Movshon JA, Thompson ID, Tolhurst DJ.** Spatial summation in the receptive fields of simple cells in the cat’s striate cortex. *J Physiol* 283: 53–77, 1978.
- Nandy AS, Sharpee TO, Reynolds JH, Mitchell JF.** The fine structure of shape tuning in area V4. *Neuron* 78: 1102–1115, 2013.
- Pasupathy A, Connor CE.** Responses to contour features in macaque area V4. *J Neurophysiol* 82: 2490–2502, 1999.
- Pasupathy A, Connor CE.** Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86: 2505–2519, 2001.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP.** *Numerical Recipes: The Art of Scientific Computing (3rd ed.)*. New York: Cambridge Univ. Press, 2007.
- Rodr guez-S nchez AJ, Tsotsos JK.** The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. *PLoS One* 22: 22, 2012.
- Rust NC, DiCarlo JJ.** Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30: 12987–12995, 2010.
- Rust NC, Mante V, Simoncelli EP, Movshon JA.** How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9: 1421–1431, 2006.
- Schiller PH, Finlay BL, Volman SF.** Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J Neurophysiol* 39: 1288–1319, 1976.
- Serre T, Kouh M, Cadiou C, Knoblich U, Kreiman G, Poggio T.** *A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex*. CBCL Paper 259/AI Memo 2005-036. Cambridge, MA: MIT, 2005.
- Vintch B.** *Structured Hierarchical Models for Neurons in the Early Visual System (PhD thesis)*. New York: Center for Neural Science, New York Univ., 2013.
- Willmore BD, Prenger RJ, Gallant JL.** Neural representation of natural images in visual area V2. *J Neurosci* 30: 2102–2114, 2010.