The Journal of Neuroscience, September 1992, 12(9): 3568–3581

Predicting Responses of Nonlinear Neurons in Monkey Striate Cortex to Complex Patterns

Sidney R. Lehky,[1] Terrence J. Sejnowski,[2] and Robert Desimone[1]

[1]Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, Maryland 20892 and [2]Computational Neuroscience Laboratory, The Salk Institute, La Jolla, California 92093

The overwhelming majority of neurons in primate visual cortex are nonlinear. For those cells, the techniques of linear system analysis, used with some success to model retinal ganglion cells and striate simple cells, are of limited applicability. As a start toward understanding the properties of nonlinear visual neurons, we have recorded responses of striate complex cells to hundreds of images, including both simple stimuli (bars and sinusoids) as well as complex stimuli (random textures and 3-D shaded surfaces). The latter set tended to give the strongest response. We created a neural network model for each neuron using an iterative optimization algorithm. The recorded responses to some stimulus patterns (the training set) were used to create the model, while responses to other patterns were reserved for testing the networks. The networks predicted recorded responses to training set patterns with a median correlation of 0.95. They were able to predict responses to test stimuli not in the training set with a correlation of 0.78 overall, and a correlation of 0.65 for complex stimuli considered alone. Thus, they were able to capture much of the input/output transfer function of the neurons, even for complex patterns. Examining connection strengths within each network, different parts of the network appeared to handle information at different spatial scales. To gain further insights, the network models were inverted to construct "optimal" stimuli for each cell, and their receptive fields were mapped with high-resolution spots. The receptive field properties of complex cells could not be reduced to any simpler mathematical formulation than the network models themselves.

As one ascends the visual pathways, from retina, through striate cortex, and on to parietal cortex or inferior temporal cortex, neuronal properties become increasingly complex. Cells at early stages of the pathways respond well to simple patterns such as spots, bars, or sinusoidal gratings. Responses of these early cells, which include retinal ganglion cells, lateral geniculate principal cells, and simple cells in V1 cortex, can be predicted reasonably well using the principle of linear superposition. That is, their response to a complex pattern is the sum of their response to simpler components. Already in V1, however, complex cells have nonlinear properties that cannot be predicted from their responses to small spots (Hubel and Wiesel, 1968). By the time one gets to inferior temporal cortex, cells often require complicated and highly specific spatial patterns (e.g., Gross et al., 1972; Schwartz et al., 1983; Desimone et al., 1984). Because these cells are so nonlinear, one cannot simply sum the responses to simple patterns such as spots and gratings to predict the response to complex images such as textures or faces. Therefore, finding an effective stimulus for such cells is a matter of extensive trial and error.

It would be desirable to have a method that can systematically capture the spatial properties of nonlinear visual cells. We have attempted to do this by using an optimization technique (Rumelhart et al., 1986) to create neural-like models of single units that can predict responses of those units to a wide range of complex stimuli. Such neural network models are useful for representing complicated, nonlinear input/output relationships. The general method was to measure responses of cells to a large and diverse stimulus set (400 patterns) and then create a network model for each cell that attempted to reproduce the spatial response properties of the cell. Once the response properties have been captured in an empirical model, it becomes possible to study the cell's input/output properties in a manner not possible when one is forced to work in "real time" during a recording session. Furthermore, the "hidden layer" in a network model might be helpful in characterizing the types of inputs a cell receives.

Although the ultimate area of interest is extrastriate cortex, at this time we shall focus on complex cells in V1. Complex cells are the first cells in the visual pathways that have strongly nonlinear spatial properties, and have been better studied than any cells in extrastriate cortex. Therefore, they provide an opportunity for evaluating the performance of the network modeling before applying it to neurons in areas of cortex that are less well explored.

Materials and Methods

Experimental methods

Animal preparation and recording procedure. Cells were recorded over a period of months from a single female cynomolgus monkey (*Macaca fascicularis*) weighing 3.3 kg. Most of the recording details have been described previously (Desimone and Gross, 1979). Briefly, 1 week prior to the first recording session, a post for holding the head and a recording chamber, both of stainless steel, were affixed to the skull with bone cement. This surgery was done using aseptic methods while the animal was under deep anesthesia induced by intravenous sodium pentobar-

Received Dec. 2, 1991; revised Apr. 8, 1992; accepted Apr. 14, 1992.

This work was funded in part by the McDonnell-Pew Program in Cognitive Neuroscience. Preliminary reports were presented at the 13th European Conference on Visual Perception in Paris, September 1990, and the 20th annual meeting of the Society for Neuroscience in St. Louis, October 1990. Portions of the modeling were completed while S.R.L. was a guest in the lab of John Maunsell, and portions of the text were written while visiting the Santa Fe Institute. We thank Mortimer Mishkin for his support during all aspects of the project.

Correspondence should be addressed to Dr. Sidney Lehky, Room 603, Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030.

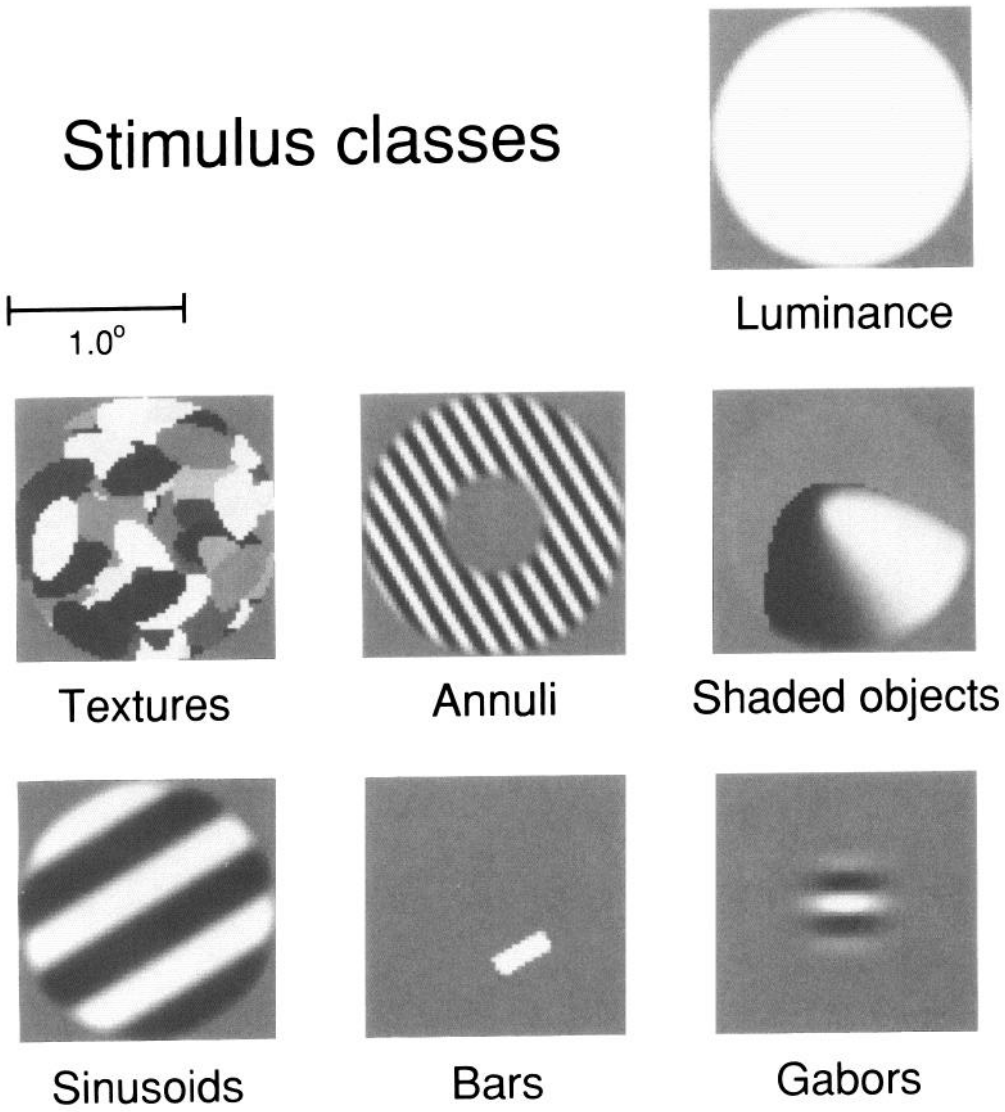

Figure 1. The 400 spatial patterns in our stimulus set were drawn from these seven classes. These are reproductions of the 64 × 64 pixel images actually shown to the monkey, and edges appear ragged due to aliasing. We classified the shaded objects and textures as complex stimuli and the others as simple stimuli, according to criteria discussed in the Results section.

bital. To gain access to the cortex, a 2 mm hole was drilled through the occipital skull within the boundaries of the recording well, leaving the dura intact. The hole was judged to be placed over dorsolateral striate cortex on the basis of skull landmarks, and the size and location of receptive fields. Between recording sessions, the recording chamber was filled with tetracycline ointment and covered by a stainless steel cap. Over the course of the experiment, several such holes were drilled as old holes filled with new bone growth. Typically, four electrode penetrations were made through each hole. The animal remained healthy and was not killed at the end of the series of recording sessions.

At the start of a session, the animal, initially sedated with ketamine, was anesthetized with 2.5% halothane in a 50:50 mixture of nitrous oxide and oxygen. It was then intubated with a endotrachial tube coated with Xylocaine jelly, and placed on a cushion and heating pad. The head was held in place by the post previously mentioned, so it was not necessary to use ear bars. At this point, the halothane anesthetic was discontinued and the animal was given a single bolus of 1 μg/kg sufentanil anesthetic (a synthetic opiate). The animal was afterward maintained on sufentanil at a rate of 1 μg/kg/hr. This dosage of sufentanil, together with the N_2O, produced a light grade of surgical anesthesia. The animal was paralyzed with pancuronium bromide included with the sufentanil infusion, and maintained by artificial respiration. No surgery or any potentially painful procedure was conducted following onset of paralysis. Temperature, end-tidal CO_2, and EKG were all monitored and maintained within normal limits. The pupils were dilated and accommodation blocked using a 1% solution of cyclopentolate.

The corneas were covered by gas-permeable contact lenses whose curvatures were selected, using a retinoscope, to focus the eyes on a computer display screen located 98 cm away. When inserting the lenses, care was taken that they were not free to slide about on excess mounting fluid. Only the eye contralateral to the recording electrode was used, the other eye being occluded. The fixation point for each eye was determined by projecting the image of an ophthalmoscope reticle centered on the fovea back through the ophthalmoscope onto the computer screen, using a corner cube prism.

Although the eyes were paralyzed, there were nevertheless small residual movements, which were monitored optically. Light from a 0.5 mW HeNe laser was attenuated by a neutral filter with an optical density of 4.0, reflected off a beam-steering mirror, passed through a 1 mm pinhole, and then reflected off an aluminum mirror (1.5 × 1.5 × 0.1 mm) glued to the edge of the contact lens and finally projected onto a sheet of graph paper mounted about 2 m from the monkey. Angular movement of the eye was calculated from the displacement of the light spot on the graph paper. Using this apparatus, three components of motion were apparent in the paralyzed eye: (1) a fast oscillation with an amplitude (peak-to-peak) of about 2 arcmin that appeared to be tied to heartbeat; (2) a larger, slower oscillation with an amplitude of about 5 arcmin that appeared to be tied to respiration; and (3) long-term drift of typically 15 arcmin/hr. All three motions were generally in the same direction. The long-term drift may have been caused by the accumulation of a slight hysteresis in the oscillations (so that the eye did not return to exactly the original position after each cycle). During the recording session, data collection was stopped every 15 min and the position of the stimulus on the screen was shifted by the amount required to compensate for the eye drift determined for that period.

Recording was performed using stainless steel microelectrodes pur-

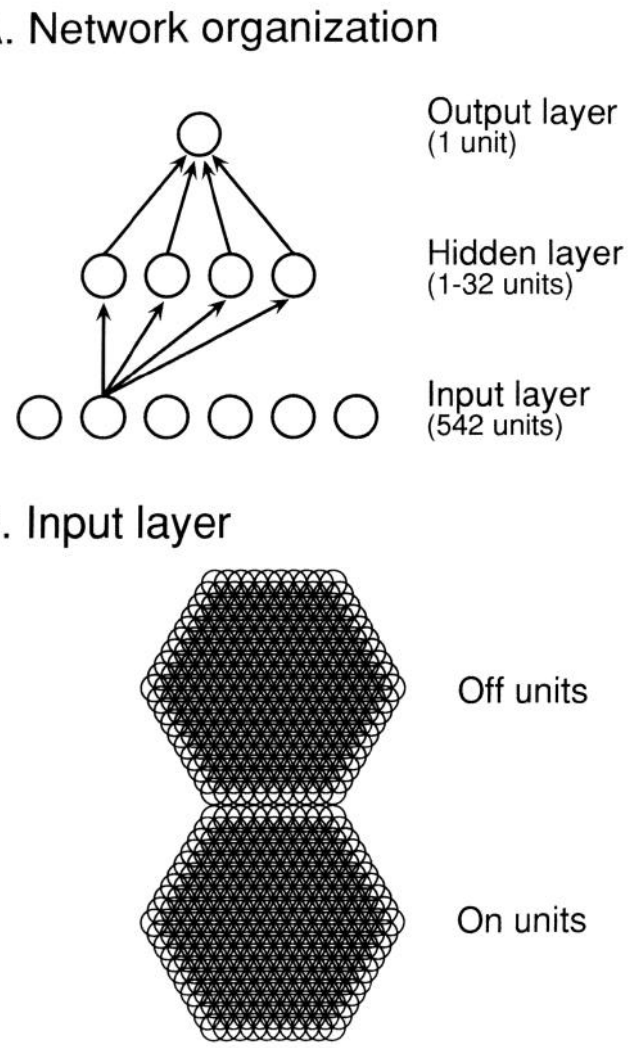

Figure 2. *A*, The networks had a three-layer, feedforward organization, without any lateral connections within a layer or feedback connections. Every unit in a layer connected to all units in the next layer. There were 542 units in the input layer, divided into on-center and off-center units, both with circular, antagonistic center/surround receptive fields. There was one output unit. The number of hidden units ranged from 1 to 32. *B*, The input units were organized into two hexagonal spatial arrays. In reality the two arrays were superimposed, and not separated as shown here. They subtended a visual angle of about 1° in diameter.

chased from Micro Probe Inc. (Clarksburg, MD), with the standard paralene and polyimide insulation supplemented with an additional varnish coat to increase stiffness near the tip. Impedance was typically 2.5 MΩ. Upon termination of recording, the animal was kept under observation until the paralytic wore off and it was breathing freely, at which time it was returned to its cage. The duration of a session was about 14 hr.

Stimuli. Stimuli were presented on a Zenith 1490 flat-screen color monitor. The gray-scale gamma function of the monitor, as well as red, blue, and green gamma functions, were measured with a Minolta CS-100 light meter. These data were used to calculate a library of linearized color lookup tables available to the display program, including both luminance and chromatic isoluminant scales. Mean luminance was set to 17.5 cd/m^2. The monitor displayed 640 × 480 pixels, which at 98 cm subtended a visual angle of 15.0° × 11.2°. The frame rate was 60 Hz.

An identical set of 400 spatial patterns were presented to each cell. They were all bounded within a circular region 1.5° in diameter, the rest of the screen being kept at mean luminance. The patterns could be divided into seven classes, as illustrated in Figure 1. These classes are described below.

1. *Luminance.* This class had three patterns, in which luminance was set to 0.1, 17.5, and 35.0 cd/m^2 within the circular boundary.

2. *Sinusoidal gratings.* Three spatial frequencies were used, 2.0, 4.0, and 8.0 c/degree. For each spatial frequency, there were six orientations, which were 0°, 30°, 60°, 90°, 120°, and 150°. Finally, for each spatial frequency and orientation, gratings were presented in two phases, which were 0° and 180°. Grating contrast was always 1.0. The pattern was linearly faded out over a 0.1° thick ring around the periphery of the stimulus in order to reduce the luminance discontinuity at the edge of the stimulus. Altogether, there were 36 patterns in this class.

3. *Gabor functions.* These had the same spatial frequencies and orientations as the sinusoidal gratings. However, in addition to the two phases used for the gratings, two additional ones were used here, 90° and 270°, for a total of four phases. The Gabor functions were localized within a Gaussian envelope with a space constant of 0.125°. There were 72 patterns in this class.

4. *Sinusoidal annuli.* These had the same parameters as the sinusoidal gratings, except that a 0.5° diameter "hole" was cut from the center. Both the inner and outer rims of the annuli were linearly faded over a range of 0.1° to reduce discontinuities. There were 36 patterns in this class.

5. *Bars.* The bars had dimensions 0.1° × 0.3°. They were presented at six orientations, ranging from 0° to 150° at 30° intervals. For each orientation, bars were presented at seven positions located at 0.15° intervals, displaced laterally in the direction orthogonal to the long axis of the bar. In addition, longer bars with dimensions 0.1° × 0.6° were presented in the six orientations, but only at a single, central position. All of these bars were presented in both black and white versions, having luminances of 0.1 and 35.0 cd/m^2, respectively, against a mean luminance background. This category had a total of 96 patterns.

6. *Shaded elliptic paraboloids.* These were 3-D synthetic surfaces having elliptical cross sections in the x–y plane and parabolic cross sections in the z–x and z–y planes. They were rotated in 3-D space and shaded according to the reflectance model described by Lehky and Sejnowski (1990). Each paraboloid was described by eight parameters. These were center coordinates (two parameters), center principal curvatures (two parameters), rotation (two parameters), and illumination direction (two parameters). All parameters were chosen with a uniform random distribution, except principal curvatures, which had a lognormal distribution with a mode of 6.0/degree. The stimulus disk was linearly faded to mean luminance along a 0.1° thick ring along the periphery of the stimulus, and the luminance distribution within the stimulus was balanced so that it averaged to mean luminance. There were 79 patterns in this class.

7. *Textures.* These were made by superimposing a large number of small, randomly located ellipses. The elliptical micropattern forming each texture came in three sizes, in which the minor axes of the ellipses had lengths of 0.12°, 0.24°, and 0.48°. The lengths of the major axes were twice those of the minor axes. Only one size of ellipse appeared in each texture. A texture made out of small ellipses contained 200 of them, each ellipse with random position, orientation, and luminance level. Textures containing medium or large ellipses were made the same way except that only 100 or 50 ellipses were superimposed. There were 26 textures at each spatial scale, for a total of 78 patterns in this class.

The stimuli were flashed with a rectangular temporal waveform having a duration of 200 msec. The interval between stimuli was 250 msec. Each of the 400 patterns was presented to the cell 30 times, for a total of 12,000 stimulus presentations. The 400 patterns were broken into 40 blocks of 10, and the blocks were presented in random order.

We acquired cells by manually moving a small, flashing bar about the screen with the computer mouse as the electrode was slowly advanced through the cortex. When promising, well-isolated single-unit activity was observed, the bar position giving the best response was determined from the loudness of the firing rate on the audio monitor. Those coordinates were passed to the display program, which centered the 1.5° stimulus field at that position and proceeded with the automatic display sequence. As described above, the stimulus position was manually adjusted every 15 min to compensate for eye drift. It took about 2.5 hr to collect data from a single cell. During a single session, one or two cells were studied.

Modeling methods

A separate network model was developed for each neuron studied. These network models were not intended to mimic the anatomical microcircuitry that underlies the neural responses, nor were they intended to

provide a realistic depiction of many physiological and biochemical processes known to occur in the nervous system. The goal of each network was to reproduce the functionality of the neuron (i.e., its input/output transfer function) without reproducing how this transfer function is actually implemented in the brain. Yet while these networks are abstract and simplified representations of the actual neural substrate, we believe that they may retain sufficient isomorphism with that substrate to be suggestive of how properties of individual visual neurons might arise in the brain. To create a network that reproduces the behavior of a neuron, we used an iterative optimization algorithm that takes as input a large set of input/output pairs collected as data, and then automatically adjusts the parameters (weights) of the network to create the proper nonlinear input/output function. The validity of the model was checked by testing it with inputs that were not part of the set used to create it, which is an essential test. While a number of such optimization algorithms are available, we chose to use back-propagation (Rumelhart et al., 1986). The mathematical details of the specific variant of the algorithm we used are given in Lehky and Sejnowski (1990).

The initial steps in creating the model involve choosing a network architecture (i.e., the number of units and how they connect to each other) and choosing the properties of each unit. These will be described below. However, some method is still needed to select the connection weights that allow the network to do the task at hand. Weights typically number in the thousands, and a combinatorial explosion prevents one from trying out all possible combinations of weights to find the optimal configuration for the required input/output function. Neither would it be feasible to set the weights by hand using intuition. The back-propagation algorithm solves this problem by searching only a subset of the weight space, in a manner that continuously and systematically brings the network closer to the optimal configuration. The algorithm therefore is a purely formal technique for creating networks with specific input/output characteristics, and is not meant to mimic actual developmental or learning processes in the brain.

Network architecture. The network had a conventional three-layer organization, consisting of an input layer, a middle "hidden" layer, and an output layer (Fig. 2*A*). Each unit in a layer connected to every unit in the subsequent layer (i.e., this was a globally connected feedforward network). There were no feedback connections, nor were there any lateral connections between units in the same layer. All units had activities that could continuously range from 0.0 to 1.0. Excitatory and inhibitory inputs to a unit were added linearly and then passed through a sigmoid nonlinearity to produce the output for that unit.

The input layer had 542 units, divided into two 2-D hexagonal arrays, one with on-center units and the other with off-center units, which were spatially superimposed (Fig. 2*B*). The input arrays subtended a visual angle of close to 1°. Each unit had an antagonistic center-surround receptive field with circular symmetry. On-center units had excitatory centers and inhibitory surrounds, while off-center units had opposite polarity. The receptive fields were described by a difference of Gaussians:

$$R(x, y) = e^{-(x^2 + y^2)/\sigma^2} - 0.16e^{(x^2 + y^2)/2.5\sigma^2}, \quad (1)$$

where $\sigma = 0.05°$. (The equation for an off unit was the same except multiplied by -1.) These parameters were chosen so that the Fourier transform of the receptive field resembled typical spatial contrast sensitivity curves of neurons in the macaque monkey lateral geniculate nucleus, as measured by Derrington and Lennie (1984). The sensitivity of the receptive field was normalized such that an optimal spot of light coinciding with the field center and having unit intensity produced a response of 1.0 in the model neuron.

Spacing between receptive field centers in the input layer was, on average, 0.05°. The array spacing was not perfectly regular, but was randomly shifted by a random distance uniformly distributed over the range of ±0.0075°. This was done to reduce spatial aliasing of the input pattern by the sampling array (Yellott, 1982).

There was only one output unit. The activity of this unit in response to an input pattern was meant to replicate the response of the actual biological neuron to the same pattern. The number of units in the hidden layer was variable. We tried networks with anywhere from 1 to 32 hidden units.

Creating the network. The training set consisted of 360 spatial patterns, which were a random subset of the 400 patterns that had been presented to the monkey. The other 40 patterns were reserved to test the ability of the network model to generalize (i.e., respond correctly to stimuli not used in the creation of the model).

The target of the model was to reproduce the mean firing rate of the neuron in response to each stimulus pattern. No attempt was made to capture temporal aspects of recorded responses. Mean firing rate was calculated over the period starting 40 msec after stimulus onset (which was the typical latency of the neural response) and ending at stimulus offset, for a duration of 160 msec. Responses were averaged over the 30 repeats of each pattern. All firing rates were normalized so that the pattern (out of all 400) producing the largest response was set to 1.0.

In the initial state of the network, all synaptic weights were randomly set over the range of 0.0–1.0. From this starting point, the iterative optimization procedure went as follows. For each trial, responses of the input units to the stimulus image were computed by convolving their receptive fields with the image, which had been randomly chosen from the training set. (In reality, convolutions for all 400 stimuli were precomputed and stored for later repeated use.) These input responses were then propagated up through the hidden units to the output unit. At this point, the actual response of the model output unit was compared with the correct response (i.e., the response of the recorded neuron) for that pattern. The difference between the two was used in the back-propagation algorithm to adjust all synaptic weights slightly throughout the network in a manner so as to reduce this error. Over the course of many trials in which each of the 400 stimuli was repeatedly presented in random order, the synaptic weights were gradually optimized so that the output of the network approximated the recorded response for each image. We generally terminated the run after 100,000 training trials (250 repetitions for each stimulus).

As an additional detail in determining the error during each trial, the correct response consisted of the mean firing rate, plus a normally distributed random component whose standard deviation was equal to the standard error of the recorded neuron's responses. This means that for different trials involving the same input pattern, the target response was slightly different. The rationale for adding this random component was that the network ought not to be trained to a precision beyond that justified by the precision of the data.

Results

General results

Properties of the cells

We recorded from 25 cells. The locations of these cells were not histologically verified at the end of the experiment because, given the long period of time over which the experiment was conducted, it would not have been possible to recover electrode tracks. As indicated above, the recording sites were judged to be in dorsolateral striate cortex on the basis of skull landmarks as well as the topography, size, and properties of the receptive fields. We kept well away from the vertical meridian representation at the V1/V2 border. Based on the distance the electrode traveled after the onset of neural activity, we believe that the majority of cells were in the supragranular layers of the cortex. The receptive fields of all cells were located in the parafoveal representation of the visual field, with eccentricities of less than 5°. The median receptive field width was 0.4°, as indicated by bar responses measured full width at half height. All except two cells showed orientation tuning.

Color preferences were only informally examined during the initial cell acquisition phase of the recording protocol. About 80% of the cells appeared to respond strongly to both red and green bars, suggesting they were not narrowly color tuned. The rest responded preferentially to either red or green. In almost all cases, responses to white bars appeared about as good as responses to colored bars. For our purposes, therefore, there was in general no advantage to using colored stimuli. For the most part, we collected data using patterns characterized by gray scale luminance gradients, except in two cases in which we used a red luminance scale and a red–green isoluminant scale.

We classified 24 of the 25 cells as complex. This judgement was based on the fact that the spatial locations of responses to

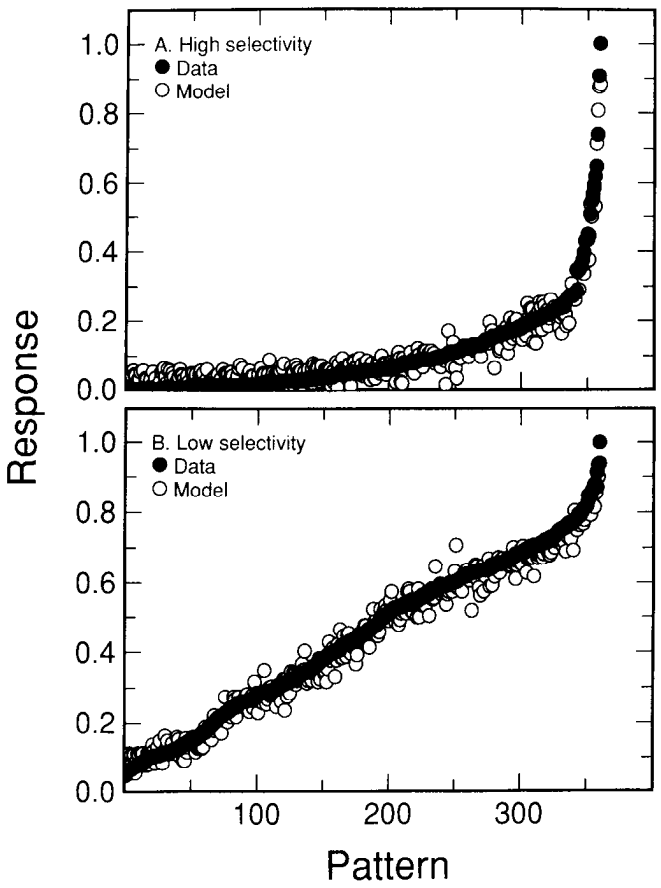

Figure 3. Responses of two neurons (*solid circles*) to all 360 stimulus patterns in the training set used to create network models for those neurons. Responses have been normalized to 1.0 and sorted by according to the actual response of the neuron. There is no order to the patterns along the abscissa other than this. Model predictions are given by *open circles*. The fluctuation of the *open circles* from the sequence of *solid circles* is an indication of how well the network was able to capture the responses of the neuron across the entire range of stimuli. *Top*, This neuron showed a high selectivity, responding well only to a few patterns. Maximum firing rate was normalized relative to 39 spikes/sec *Bottom*, This neuron showed a low selectivity, responding well to many patterns. Maximum firing rate was normalized relative to 83 spikes/sec.

black bars and white bars within the receptive fields were nearly the same, and that the response magnitudes to the two kinds of stimuli were also similar, being within a factor of two of each other (see, e.g., Fig. 5). The one remaining cell remained unclassified by this criterion because it did not respond strongly to any of the bar stimuli. A high incidence of complex cells and low incidence of simple cells in V1 of the macaque has been reported by Hubel and Wiesel (1968), who indicated that overall only 8% of neurons they recorded from were simple. Dow (1974) and Poggio (1972) have reported similar numbers.

Response to the best stimulus pattern had a median value of 59 spikes/sec and a range of 13–224 spikes/sec over the 25 cells we recorded from. The ratio of the best response to the worst for a given cell was usually in the range from 5:1 to 10:1. When the same pattern was presented at different times, the standard error of mean firing rates was typically 0.30 of the firing rate for patterns producing the smallest responses, and 0.05 for the best patterns, with a smooth gradation between. Stimuli that produced the best responses tended to be large and complex patterns, namely, the random textures and shaded 3-D elliptic paraboloids (see, e.g., Fig. 9). This preference for complex stimuli was a statistical bias, and not absolute.

Different cells showed different degrees of selectivity; that is, some cells gave large responses to only very few patterns among all those presented to them, whereas others responded to a broader range of stimuli. This is shown in Figure 3. The top panel shows responses of a neuron with relatively high selectivity tested with 360 patterns (solid circles). These 360 patterns have been sorted according to their relative response magnitude and plotted in that order. One can see that responses are small for almost all stimuli but shoot up for a few patterns. The bottom panel shows responses of another neuron with low selectivity.

Figure 3 also shows network model predictions for neural responses to the 360 patterns (open circles). These 360 patterns formed the training set used to create the model for each neuron, so it is not surprising or particularly significant that the model predicted these data well (correlation of 0.95). More interesting would be the ability to predict novel stimuli not part of the training set, which will be discussed below in the section on modeling.

We were almost always able to hold cells for several hours. Over that period, it was not unusual to observe waxing or waning in response to all stimulus patterns. This sort of nonspecific shift in sensitivity was not significant for our modeling purposes, since we were interested only in relative responses to different stimuli. However, there was occasionally some drift in relative responses to different patterns during a session, which may have been caused by incomplete compensation for drift in eye position. Nevertheless, the small standard errors of the responses, given above, indicate that fluctuations in neuronal responses were not excessive.

Properties of the network models

A network model was created for each neuron we recorded. As there is not room to describe all networks, in this section we shall make a few general observations, and in the following sections describe the network model for two cells in greater detail.

For each cell, we created several networks that differed in the number of hidden units, ranging from 1 to 32. Note that since we always had just one output unit, our three-layer networks with one hidden unit were functionally equivalent to a two-layer network (i.e., a network with no hidden units). The properties of two-layer networks are qualitatively different from those of three-layer networks, since the input/output relationship in a two-layer net is linear (aside from the sigmoid transfer function of the output unit), whereas a three-layer network can represent strongly nonlinear input/output relationships.

Not surprisingly, the networks did very well in capturing the input/output relationship for the stimulus images in the training set (see, e.g., Fig. 3). We measured network performance by the correlation coefficient between the responses produced by the network to the input patterns and the responses measured from the recorded neuron. The median correlation (over the 25 cells) for networks with one hidden unit was 0.82, with 16 hidden units it was 0.95, and with 32 hidden units it was 0.98.

Neural network models with a sufficient number of hidden units can sometimes "memorize" each stimulus/response pair rather than extract regularities from the training set. The problem is analogous to fitting data with too many parameters. As a test of whether our networks had this problem, we created a new training set using synthetic data in which each stimulus

pattern was randomly paired with one of the responses recorded from the neuron. A network with 16 hidden units trained on this set of random data was able to reproduce it with a correlation of only 0.64, compared to a correlation of 0.95 for the actual data. This is an indication that the size of our data set was larger than the capacity of the networks to memorize individual items. The networks therefore had to deal with the data in a more general fashion.

A more direct measure of how well the networks extracted the input/output relationships of the recorded neurons is their ability to predict responses to stimuli that were not part of the training set used to create the model. A network that has memorized input/output pairs should not generalize. We tested prediction by measuring the network response to 40 new stimulus patterns that had not been part of the training set, but for which we had data. For networks with one hidden unit (again, essentially a two-layer network), the median correlation was 0.55 (range, 0.19–0.83). The ability of the model to generalize improved as more hidden units were added until there were 16 hidden units, at which point the median correlation was 0.78 (range, 0.40–0.94). Going to 32 units did not increase the ability of the models to generalize any further. For the reason stated above regarding memorization, we expect that prediction ability would eventually decline as the number of hidden units was increased even further. We also tested generalization for networks trained on the scrambled training set described in the preceding paragraph. As expected, there was zero correlation between the predicted and actual responses. Because networks with 16 hidden units appeared to work best, in discussions below we shall focus on networks of that size.

To provide a tougher test of the ability of the networks to generalize, we tried them on a difficult subset of the 40 patterns in the test set. All of the stimulus images we used could be divided into two groups, which we shall call "simple" and "complex." In the simple group fall the sinusoidal, Gabor pattern, annuli, bar, and mean luminance stimuli. These patterns were described by a small number of parameters, and on the basis of these parameters could be placed in an orderly sequence (according to orientation, spatial frequency, etc.) within each class. The complex group included the random textures and shaded surfaces. These patterns were defined by a large number of parameters, also selected randomly, and it was not possible to order them in any useful sequence. To predict responses to the simple patterns in the test set, it was only necessary for the networks to learn to interpolate across a smooth tuning function for a parameter. However, this cannot be said for the complex patterns. The ability of networks to generalize across complex, random patterns provides a particularly stringent measure of the degree to which they captured the response properties of neurons. Of the 40 test set patterns, 15 fell in the complex category. When the networks were tested for generalization to these 15 complex patterns, the median correlation between the predicted and the actual neuronal response was 0.65 (range, 0.08–0.84) with 16 hidden units in the network and 0.38 (range, −0.21–0.84) with one hidden unit, over the 25 cells we recorded from. Networks trained only on simple patterns were not able to predict responses to complex patterns.

The ability of the model to generalize depends on collecting massive amounts of data. This involves presenting a large number of patterns, and presenting each pattern many times to reduce the standard error of the responses. In the early phases of this study, we used only 100 patterns, presented 10 times each

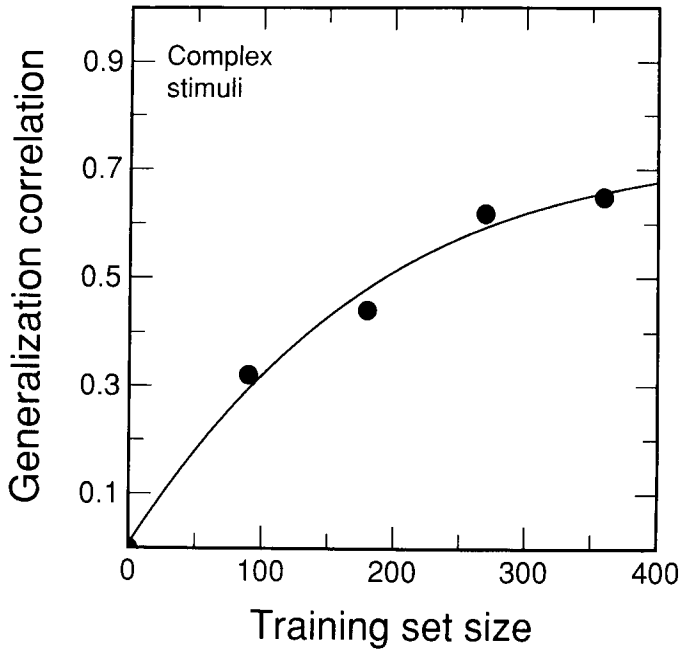

Figure 4. Ability of the networks to generalize, as a function of the number of patterns in the training set used to create them. By generalization, we mean the ability to predict responses to stimuli that were not part of the training set. Performance is measured as the median correlation between data and model for the 25 network models (one for each neuron we recorded from). The *curved line* is a fit of a third-order polynomial to the data.

to the neurons. Generalization to new patterns following training with this smaller data base was close to zero. As we reduced the standard error of the responses by increasing the number of repetitions of each stimulus to 30, and increased the diversity of the stimulus set by going from 100 to 400 patterns, generalization improved. Figure 4 shows the increase in the correlation coefficient between data and model for complex stimulus patterns as a function of training set size. It appears that performance of the models would have improved only slightly with an additional increase in the stimulus set. This suggests that a limiting factor in the model's ability to make predictions may be noise and fluctuations in the responses to each pattern used to create the model, rather than the number of patterns. Also, it is possible that the simple feedforward network architecture we used could be a limiting factor, and that model performance could be further improved by including lateral and feedback connections.

We tried to increase the amount of data in the training set for the networks artificially by including interpolated "data" in the set. For example, if responses were recorded to gratings with orientations of 0° and 30°, the response to a 15° grating could be interpolated and added to the training set, even though a 15° grating was never shown to the monkey. To test the usefulness of such a strategy, we trained networks with interpolated data added to the training set and measured generalization. Interpolation was only done for simple patterns and not the complex ones. By inserting synthetic data, we expanded the training set from 360 to 2145 patterns. Overall, this improved the ability of the network to generalize to new patterns slightly, raising the median correlation between the predicted and actual responses from 0.78 without interpolation to 0.85 with it. However, adding the interpolated data did not improve the ability of the

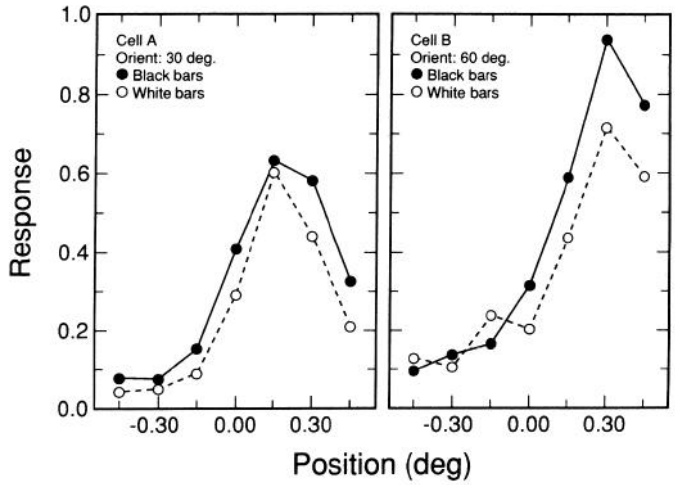

Figure 5. Spatial tuning curves of two example neurons to bar stimuli. The *curves* are shown for the bar orientation producing the strongest responses. Responses are normalized on a scale of 0.0–1.0 relative to the largest response produced by the 400 patterns in the stimulus set. The *abscissa* shows the distance of the bar from the center of the stimulus field along the direction orthogonal to the bar's orientation. *Positive numbers* indicate positions in the lower right quadrant, and *negative numbers* indicate positions in the upper left quadrant. Although for *Cell A* bar responses for 30° orientation are shown, responses at 0° were virtually identical, suggesting that the optimal orientation actually fell between 0° and 30°.

networks to predict responses to complex patterns in the test set. For our purposes, therefore, this did not turn out to be a useful technique.

Network examples

Since the network models for all 25 cells were qualitatively similar, we show two representative cells as examples. The networks described below had 16 hidden units, the number that gave the best generalization. In order to give some idea of the properties of these two cells as characterized by conventional means, Figure 5 shows spatial tuning curves to black bars and white bars having the optimal orientation. The graphs indicate that both cells respond well to white bars and black bars at the same location, which is a characteristic of complex cells. Both cells responded best to bars located in the lower right quadrant of the stimulus field, as shown in Figure 6.

It should be kept in mind that the models were constructed on the basis of recorded responses to flashed stimuli. It seems likely that if the temporal conditions of stimuli presentation had been different (e.g., if the stimuli were drifted across the screen rather than flashed), various aspects of the models could have been quantitatively different.

Connection weights

Figures 7 and 8 show the connection weights for the network models of the two cells. Each of the 16 hourglass-shaped objects in the figures shows the connection weights from the input units to one hidden unit.

Examining the weights for the different hidden units, two types of organization are apparent. In some, the excitatory and inhibitory weights are organized as large, elongated blobs (e.g., the hidden unit shown top row, fourth column in Fig. 7). In others, the pattern of excitatory and inhibitory weights appears to be at a finer spatial scale (e.g., second row, fourth column in Fig. 7). Such a division of hidden units into high spatial frequency/low spatial frequency classes was an almost universal occurrence for the network models of various cells.

The "high spatial frequency" hidden units have a very complicated organization of weights, which look rather unbiological. It is possible that what we are seeing there is a patchwork of many small regions, each with a simple organization, all jumbled together. If that is the case, then the apparent complexity of the high-frequency units may be an artifact of the connectivity we chose for the network. In our globally connected network, each hidden unit receives input from all input units across the entire model "retina," 1° across. However, one would expect that hidden units responding well to high spatial frequencies would receive inputs from a more localized area of the visual field than

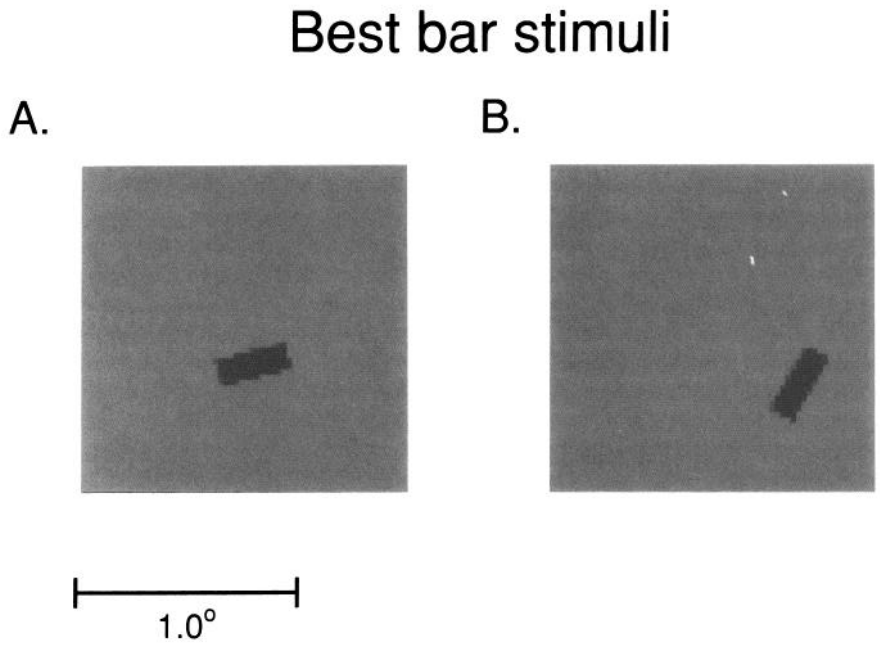

Figure 6. Bars producing the best responses for the two example cells, based on data of the type shown in Figure 5. Since responses of *Cell A* were almost identical for 0° and 30°, we show the optimal orientation for that cell as being halfway between, at 15°.

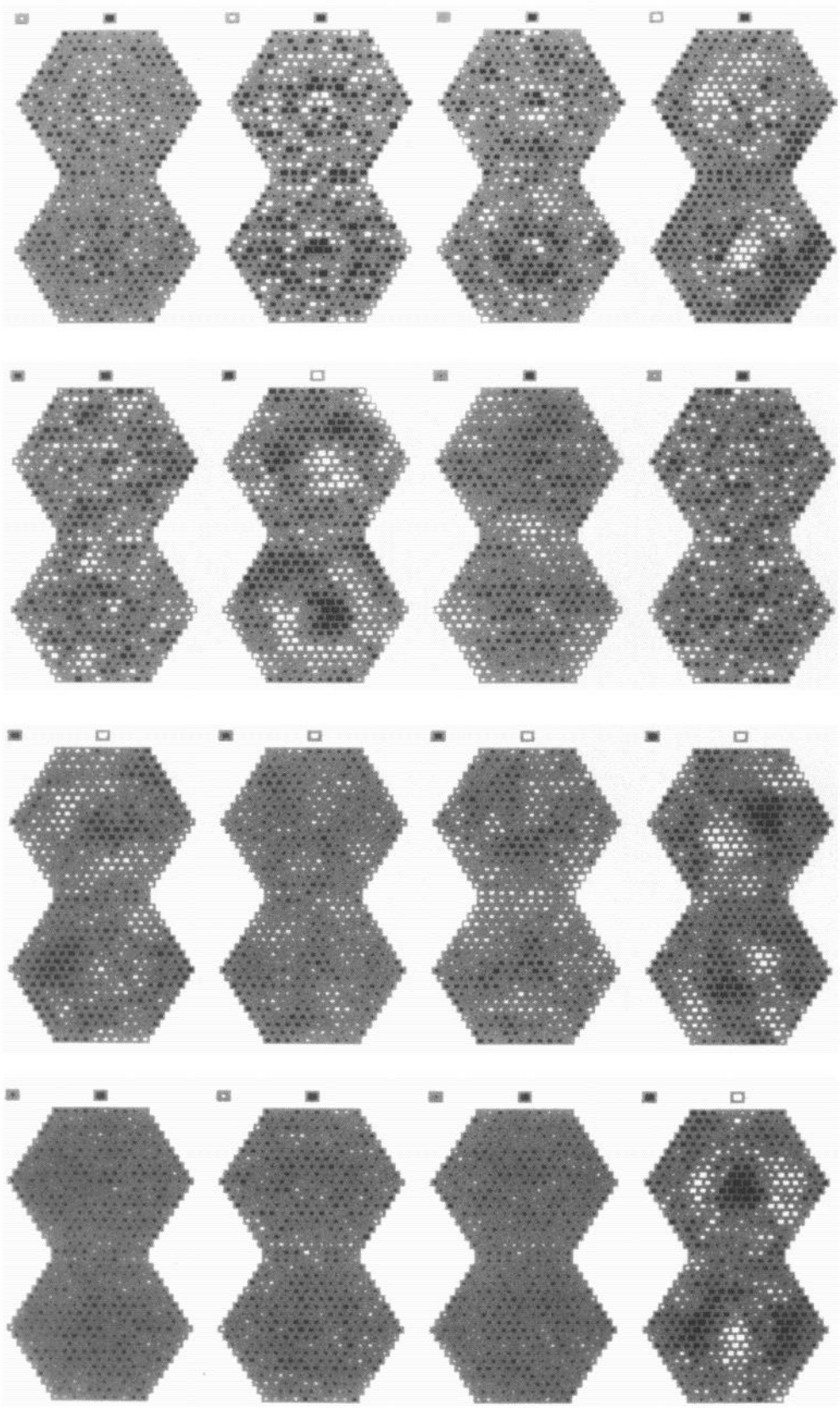

Figure 7. Diagram of the weights underlying a network model with 16 hidden units for one neuron (Cell A in Fig. 5). *Solid squares* indicate inhibitory weights, and *open squares* show excitatory ones. Each of the hourglass icons shows the weights for one hidden unit. The *lower hexagon* shows weights from the 247 on-center input units to that hidden unit, and the *upper hexagon* shows weights from the 247 off-center units. The *single square at top center* of each icon shows the weight from that hidden unit to the single output unit. The *single square at the top left* of each icon shows the bias on the hidden unit, essentially equivalent to setting a threshold for that unit.

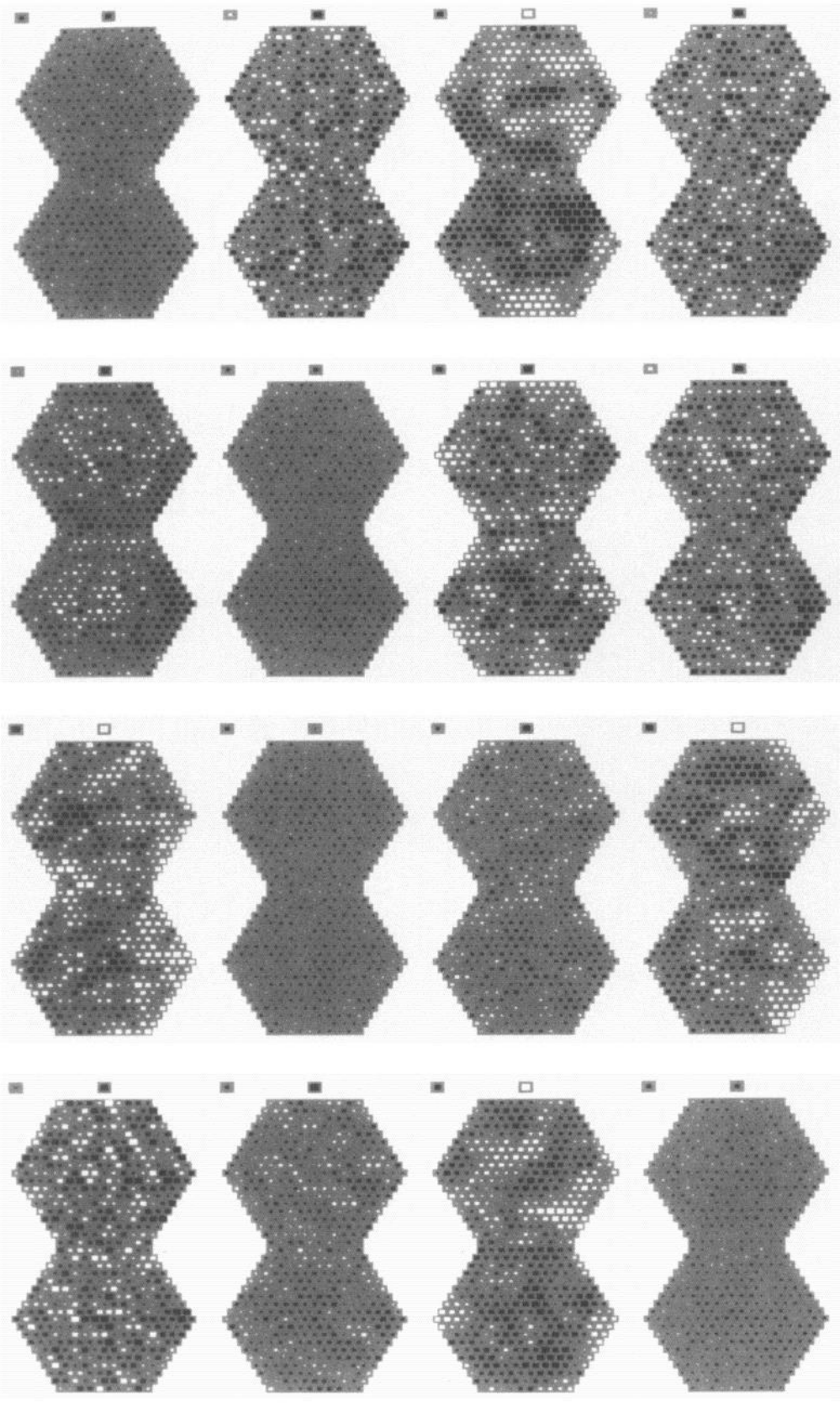

Figure 8. Diagram of the weights underlying the network model for a second neuron (Cell B in Fig. 5). See Figure 7 for display conventions.

would low-frequency units. Since there is nothing in the network to enforce such a local connectivity for hidden units, perhaps what forms in some cases is a mosaic of local domains.

We examined this possibility by creating networks that were partially connected. In addition to hidden units that were connected to the entire input layer, we created hidden units connected to 0.5° and 0.25° patches of the inputs in a manner that tiled the entire field with a high degree of overlap. Although the locally connected hidden units did form a simple organization of weights, this did not prevent some of the globally connected hidden units from continuing to form a complicated mosaic pattern. Furthermore, there was no improvement in the ability of the network to predict the data, so this approach was abandoned. However, even though this approach did not seem very effective, it may be that a cleaner separation of the mosaic patterns could have been obtained by a more extensive search through different network architectures.

The on- and off-center inputs to many low spatial frequency hidden units were complementary. This can be seen in Figures 7 and 8, where the top and bottom hexagons of some of the hourglass figures look like inverted copies of each other. This has the effect that if a light spot excites such a hidden unit at a given location, a dark spot will inhibit it. In effect these hidden units acted as linear subunits of the nonlinear neuron we modeled. Such complementarity of weights is not forced by the optimization algorithm, and indeed, it was not a characteristic of all hidden units, but it is interesting that they were so common. Neither is such an organization of weights an idiosyncratic product of this data set, for the same complementarity was seen in a previous model (Lehky and Sejnowski, 1988), which was entirely synthetic in that it involved no recorded data at all. It is not known why the "high spatial frequency" hidden units never showed this complementary organization.

It is notable that the size of the weights in the networks remains substantial across the entire stimulus field of the network, about 1° across. Such was the case in all the networks we created. This would indicate that, according to our modeling, there was a fairly broad region influencing these cells' responses. White noise analysis of complex cells in cat striate cortex (Szulborski and Palmer, 1990) has also found a large response region, extending 5° or 6°, although measured at slightly greater eccentricities than we did.

Generalization

Generalization for these two networks is shown in Figure 9. This compares neural responses with model predictions for the 40 patterns in the test set, which were not part of the corpus used to create the model. The actual response of the neuron for each of the 40 patterns is indicated on the horizontal axis of Figure 9, and the model prediction is indicated on the vertical axis. If the data and model predictions were identical, all points would fall on the 45° line. The ability of the network to generalize to new stimuli is an important test of the model.

Lesions

We looked at the effects of "lesioning" away hidden units upon the networks' ability to generalize. Lesions were accomplished by setting all weights associated with a given hidden unit to zero. Each of the 16 hidden units within each network was removed one at a time, so that the network always had 15 functioning hidden units. Removing any single hidden unit usually had negligible effect on network performance. Of the 400

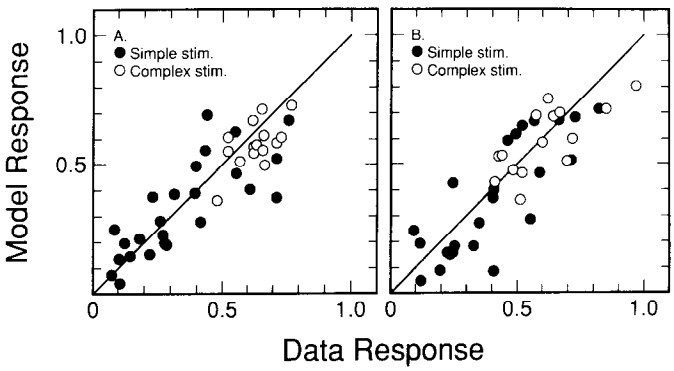

Figure 9. Ability of the network model to generalize, for the two example neurons. The *abscissas* show the neurons' normalized firing rate for 40 patterns that were not part of the training set, and the *ordinate* shows the model predictions for those patterns. Ideally, all dots would fall on the 45° line. Aside from the modeling, this figure also shows that neural responses were systematically higher for complex stimuli (shaded objects and textures) than for simple stimuli. *A*, Model predictions based on network shown in Figure 7. *B*, Model predictions based on network shown in Figure 8.

units lesioned in the 25 networks, in slightly more than half the cases there was a change in correlation coefficient of 0.01 or less.

Removal of a few "critical" hidden units did have a significant effect, however. Defining "large" as any change whose absolute value ≥ 0.10, 30 out of 400 (7.5%) of lesioned hidden units caused large changes in networks' ability to generalize. Of these, removal of only three led to improvements in network performance, and those units were within networks that had among the worst performances to begin with. The change in correlation upon removal of a single unit ranged between -0.28 and 0.15.

The critical hidden units tended to be associated more with the "low spatial frequency" class of receptive field organization, described previously. Only 1 out of the 30 critical hidden units was clearly a "high spatial frequency" unit. Although we don't have a quantitative criterion for classifying high or low spatial frequency units, at least a quarter of the hidden units are clearly "high spatial frequency." This means that this class is underrepresented among the critical hidden units.

Obviously there are many more opportunities for exploring the effects of "lesioning." For example, methods exist for systematically identifying individual connections within the network that contribute little to performance (Le Cun et al., 1990). Upon removing those connections and retraining the resulting smaller network (having fewer degrees of freedom), generalization commonly improves over the original model.

Spot responses

Once a network model has been created that captures the response properties of a cell, it becomes possible to do simulated experiments on it. One such simulated experiment we performed was to map the response of the network to a small stimulus "spot" applied to its input field. The spot could either be white or black, against a gray background. The size of the spot was 1.4 arcmin across; thus, we were stimulating the model of the neuron at a higher resolution than would have been practical for the actual neuron. A motivation for this spot mapping was to gain some sense of how the network behaved as a whole, for examination of all the weights in Figures 7 and 8 provide

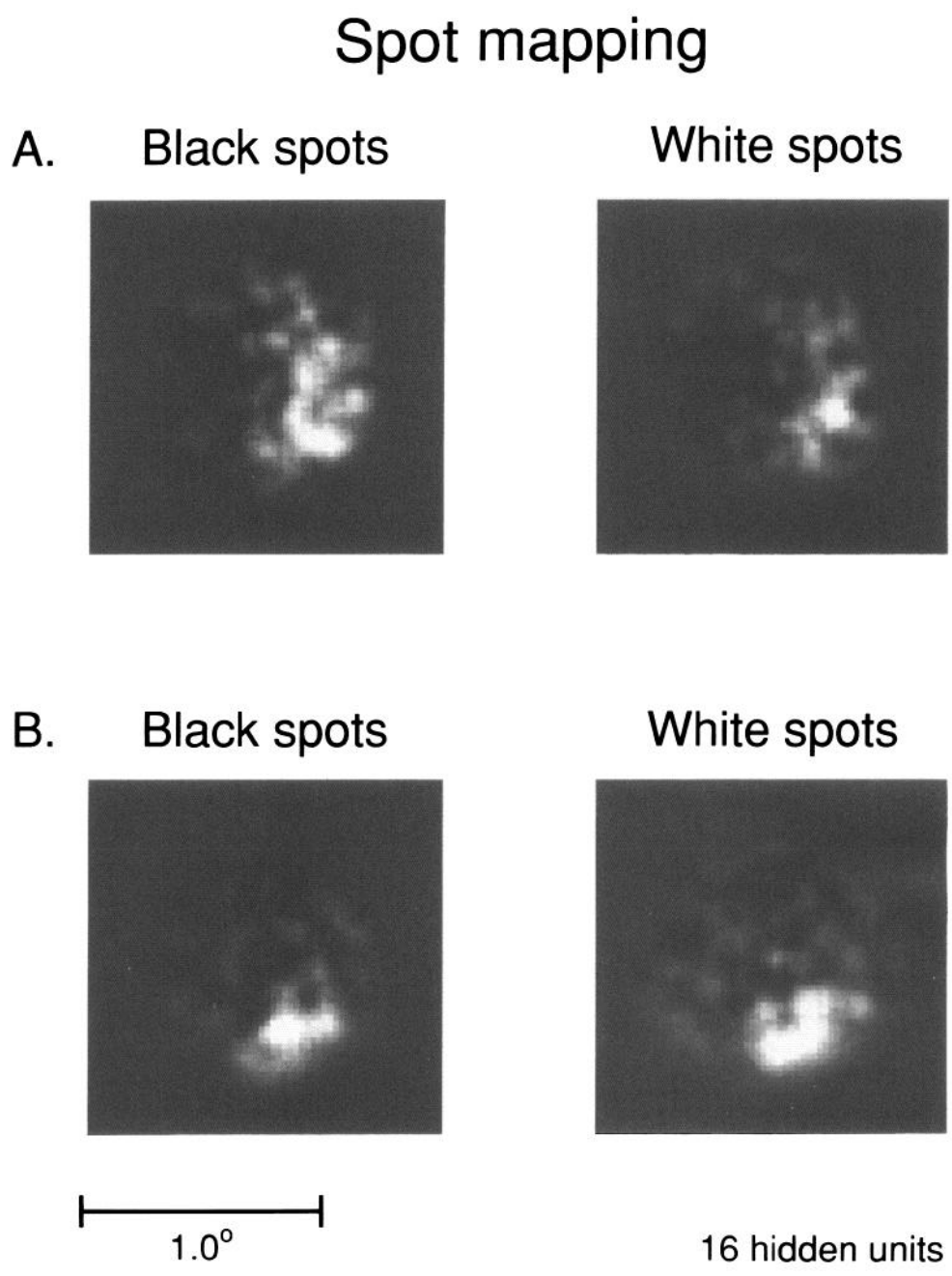

Figure 10. Results of simulated experiments in which we mapped responses of network models for the two example neurons to small spots of light 1.4 arcmin wide presented at locations across the model stimulus field. The *gray level* at the edges of the squares (where the spot was off the model "retina") indicates the level of spontaneous activity in the network model, and lighter or darker regions indicate excitation or inhibition caused by the spot stimulus. Responses have been normalized to a scale of 0.0–1.0. The substantial spatial overlap of responses to white and black spots shows that the models have captured a defining feature of complex cells in striate cortex. *A*, Spot responses for the network shown in Figure 7. *B*, Spot responses for the network shown in Figure 8.

little intuition about how the network would respond to a specific stimulus at a specific location.

The results of the spot mapping are shown in Figure 10 for the two exemplar cells. In all cases, responses to the spots have been normalized to 1.0, because the absolute responses to such tiny stimuli were very small. The regions of high spot sensitivity can be seen to form rather amorphous blobs without a large amount of structure, which is typical of the spot mappings for network models of other cells. For some cells, spot responses were more localized than seen here, and for other cells less. At the coarse level, it can be seen that regions responding to black spots and white spots are in the same general area, and often overlap substantially, showing that the model has captured one of the defining features of complex cells. On the other hand, there are finer modulations in the responses that appear to be complementary between the two (i.e., the peak in one corresponds to a trough in the other). However, this fine structure is model dependent, being influenced by receptive field diameters of units in the model "retina."

The peaks in the spot mapping of the model matched the positions of the best bar stimuli in the data only roughly—they all fall in the lower right quadrant of the stimulus field (see Fig. 6*A,B*, which corresponds to Fig. 10*A,B*, respectively). The difference in position between the two is about 0.15°. This may be overstating the difference a bit, because the best bar position is only approximately known (i.e., we have data only for the limited number of bar positions and orientations that were in our stimulus set). Also, it is possible that the center of the optimal bar did not correspond to the center of the spot response because it may have been a corner, end, or edge of the bar that

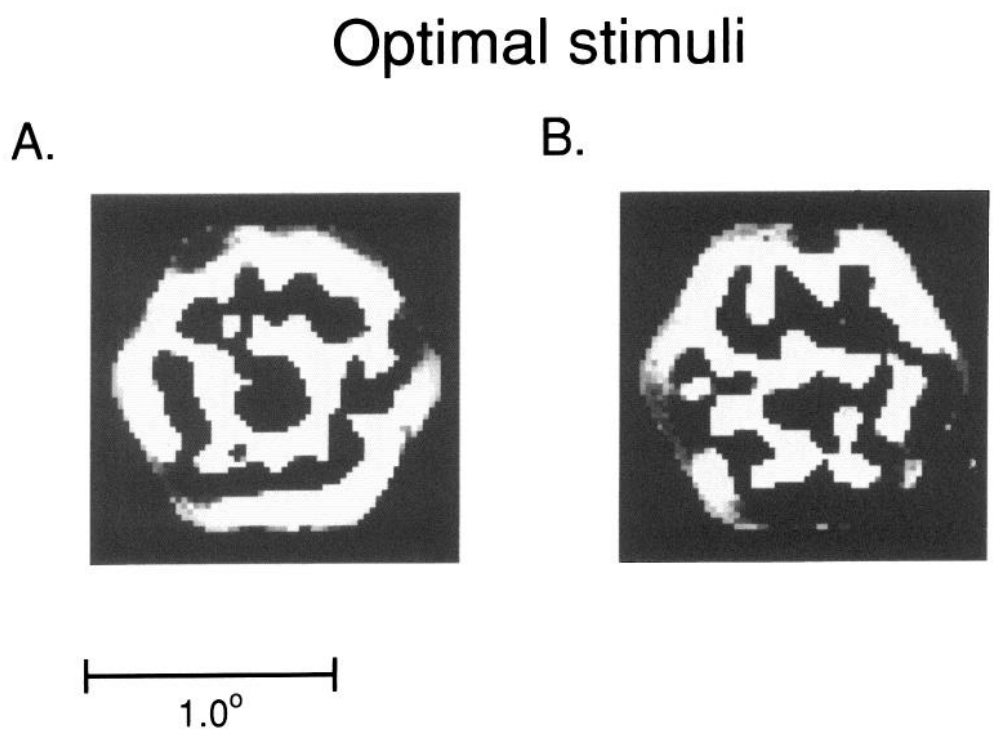

Figure 11. Optimal stimuli found by inverting network models for the two example neurons. *Black borders* indicate the edge of the model "retina." The configuration of the optimal stimuli appears unrelated the spot responses in Figure 8, an indication of strong nonlinearities in the responses of the neurons. *A,* Optimal stimulus for the network shown in Figure 7. This image produced a response 1.85 times that of the best stimulus pattern in the data set. *B,* Optimal stimulus for the network shown in Figure 8, producing a response 1.46 times the best stimulus pattern in the data set.

was actually the effective stimulus. In any case, the rough correspondence between bar response data and calculated spot responses provides additional reassurance that the model is behaving in a sensible manner.

Optimal stimulus

The optimal stimulus for the network model of a neuron can be found by inverting the network (see Linden and Kinderman, 1990, for the inversion technique). In essence, this was done by using the same optimization algorithm used to create the network, except that instead of changing the weights of the trained network, we held those constant and changed the values of the pixels in the input image so as to maximize the output of the network. Starting with the pixels set either randomly or at a uniform gray level, a pattern gradually emerged as the algorithm adjusted the pixel values to produce the greatest output. Responses to these optimal stimuli were typically about 50% larger than the best pattern in our stimulus set.

Optimal stimuli for network models of the two example neurons shown in Figure 11 are typical of the patterns we calculated. They form highly irregular striped patterns that fill the entire 1° visual field subtended by the model. The substantial difference between the optimal stimulus and the relatively unstructured spot responses in Figure 10 is indicative of the high degree of nonlinearity in the cell (in a linear system both should look similar). These images were highly reproducible when created multiple times for the same network model (starting with different random pixel configurations). They were also reproducible for different network models of the same neuron having 8, 16, or 32 hidden units, correlations between patterns being greater than 0.90.

An aspect of the optimal images that did appear to be model dependent was the width of the irregular stripes. The width of these stripes was approximately the same as the center diameter of the circular center/surround units in the input layer of the model (which were set at 0.10°). We created a network in which the input units had twice this diameter, and the resulting optimal image for this network had thicker stripes. The dependence of the optimal image on the spatial scale of the input units is not surprising. The upper network layers never "see" the raw image, but only the image after it undergoes a bandpass spatial filtering in the input layer. This initial filtering will affect the spatial frequency content of the network's optimal stimulus, but nevertheless the actual spatial organization of the stimulus will be largely determined by the weights in the network (which in turn are determined by the data).

Whether the optimal stimulus calculated for the network model of a neuron is in reality an extremely good stimulus for the actual neuron is unverifiable until we have computers fast enough to do the calculations while the recording session is still in progress. In the meantime, it is best to look at the properties of these optimal stimuli in more general and qualitative terms. The most robust aspects of the calculated optimal patterns are as follows. First, they always extended over the entire 1° visual field of the network, much broader than the spot responses calculated for the model (Fig. 10), or the bar spatial tuning curves in the data (Fig. 5). We were never able to localize a small region (by masking in various ways) that seemed to be doing most of the work. Second, the predominant structure within the images always appeared to be irregular stripes, as opposed to disconnected blobs, for example. Finally, there was no indication the calculated optimal stimuli could be described mathematically by some simple function (e.g., a Gabor function) corrupted by noise in various ways. The patterns seemed intrinsically irregular.

Discussion

We have created neural network models of individual complex cells in monkey primary visual cortex. These models were able to predict, with moderately high correlation coefficients, the recorded responses to complicated spatial patterns not part of the data set used to create the models. This indicates that the

models have captured a significant portion of the response properties of the cells. While these network representations of the neurons we recorded from are obviously very simplified relative to the actual state of affairs in the brain, we believe they still show a predictive power that surpasses more conventional models of complex cells (Glezer et al., 1980; Spitzer and Hochstein, 1985b). In fact, these latter models have not been demonstrated to predict responses to novel stimuli at all.

The networks were created by training them on our recorded data, and by presenting them with the same input stimuli presented to the monkey. Others have also reported neural networks being trained on actual data, specifically Krauzlis and Lisberger (1990) for cerebellar cells, and Hertz et al. (1991) for cells in inferior temporal cortex. Although in those two models the network did well on the training set, there was no report of the ability of the network to generalize, which we feel is an essential test of how well the model captures the properties of the cell.

By developing networks using the actual experimental inputs and outputs, the approach here differs from that of Zipser and Andersen (1988) and of Anastasio and Robinson (1989). In both of those studies, networks were created to simulate a particular task using synthetic inputs and outputs. Following training, the hidden layers were examined for units that qualitatively resembled those recorded in cortical areas likely to be involved in the task. This alternative approach is useful when one has in advance a good idea of the key parameters underlying the task. Those models involved the use of eye position information, or the control of eye position, in tasks for which simple linear models give reasonable first-order approximations. This makes it easier to incorporate into the model input and output representations that reflect prior knowledge about the problem. We do not have these advantages in constructing models of the processing of spatial patterns or, ultimately, object recognition, topics that at present have very weak conceptual foundations.

In this first attempt, we modeled only mean firing rate without trying to include any temporal structure of the responses. It would be possible to take time into account by splitting the data into a number of time bins and assigning a separate output unit in the network for each bin, rather than just having a single output unit. Some reports (Richmond et al., 1987, 1989) would suggest that including temporal aspects of the neural response would improve network performance. On the other hand, breaking up the response into smaller time bins would reduce the signal-to-noise ratio within each bin, since less data are being pooled. This raises the possibility that training the networks to reproduce the fine temporal structure would actually reduce their ability to generalize. How temporal information affects network performance must ultimately be decided by further modeling. A likely possibility, supported by the data of Kruger and Becker (1991), is that there is an optimal temporal integration time, perhaps reflecting some integrative time constant in visual cortex. Networks trained at the temporal resolution given by that bin size would have the maximum ability to predict responses.

The receptive fields of complex cells are often believed to be created by the nonlinear addition of linear subunits (e.g., see models of Glezer et al., 1980; Spitzer and Hochstein, 1985b). Interestingly, the fact that the connection weights for on- and off-center inputs were largely complementary for some of the hidden units in our models suggests that these model units had near-linear properties. However, some hidden units were obviously not linear and even the largely linear ones sometimes had complicated patterns of connection weights. Thus, this modeling suggests that the cells providing inputs to complex cells may often have a greater complexity and variety than previously believed. Ultimately it may be possible for network models to make specific predictions about the cells providing input to complex cells, which could then be tested experimentally.

As was outlined in the introductory remarks, we decided to develop neural network models of single cells in order to construct a more comprehensive description of their responses than has been available in the past. Previous studies have, in the main, attempted to examine complex cells using a restricted set of simple patterns, such as bars or sinusoidal patterns, which are useful stimuli for linear systems analysis (e.g., Schiller et al., 1976; Movshon et al., 1978; Dean and Tolhurst, 1983; Spitzer and Hochstein, 1985a; Pollen et al., 1988). While these studies have been helpful in revealing the general features of these cells, they do not appear to have characterized them to the extent that they could predict their responses to arbitrary stimuli. Optican and Richmond (1986) and Richmond et al. (1989) report that they have used the responses of complex cells to 1-D Walsh patterns to predict responses to stimuli that are the sum of two patterns. However, it is not clear if the model can predict responses to 2-D patterns or other complex patterns.

On the other hand, if one tries to probe nonlinear properties more fully by presenting cells with a richer set of stimuli, such as textures and 3-D surfaces, the problem remains how to integrate all this information into a useful characterization. In extrastriate cortex, where cells have been studied with very complex patterns (e.g., Desimone et al., 1984), one is sometimes left with descriptions that a cell responded well to this pattern, or that pattern, without any ability to predict responses to any other patterns.

The network models described here appear to offer promise as a solution to the limitations described above. The power of this technique comes from its ability to integrate large volumes of data acquired using a wide diversity of arbitrary, complex, stimuli into a single description of a cell. Obviously, the more information collected, the better the characterization will be, and a limiting factor for developing this type of network model appears to be the technical difficulties of recording sufficient data.

Besides this network approach, another technique that has been reported to produce a very general characterization of neuronal properties, including nonlinear properties, is white noise analysis (Marmarelis and Marmarelis, 1978). White noise analysis has been applied by Szulborski and Palmer (1990) to complex cells in striate cortex. They derived a series of second-order kernels having elongated center-surround organizations. However, these models have not yet been tested by using them to generate responses to any novel stimuli, and thus it is still not clear whether second-order kernels are sufficient to model complex cell properties fully. Given the paucity of simple cells in primate striate cortex, we chose to use circularly symmetric units as the primitives in our input layer, but in principle the primitives could have been other types of units, such as oriented ones. However, even if we had constructed a model in which the input layer consisted of oriented units in a variety of positions, orientations, phases, and so on, it still would have been necessary to use an adaptive algorithm to set the weights by which they all converged to form a complex cell. We believe it

would be virtually impossible to hand-tune the weights from all these subunits to create a practical model that quantitatively predicts an extensive data set. The point here is that these neural network techniques can be a flexible tool for fleshing out the details of various assumptions one may wish to incorporate in a model. It might be possible, for example, to use these algorithms to create network models of individual cells that incorporated details about cortical microcircuitry (Lund, 1988), which again would be extremely difficult if all connections had to be set by hand.

The question naturally arises after all this modeling: do we now understand the function of these cells? The answer is no. It seems unreasonably optimistic to expect that by taking data from a small number of cortical units, and subjecting those data to any sort of mathematical transform, no matter how complicated and nonlinear, the role of those units within the neural economy will somehow pop out. In particular, there is no reason to believe that characterization of a cell's receptive field, in itself, reveals the cells's function, a point made in a previous study (Lehky and Sejnowski, 1988). The neural network models presented here can be thought of as very elaborate characterizations of receptive fields.

Rather than trying to infer function solely from low-level single-cell data, broader psychological and computational considerations must be included as well. It is our hope that the synthesis of such top-down constraints with bottom-up modeling of individual neurons will lead to new hypotheses for perceptual mechanisms that can be experimentally tested.

References

Anastasio TJ, Robinson DA (1989) Parallel distributed processing in the vestibulo-ocular system. Neural Computation 1:230–241.

Dean AF, Tolhurst DJ (1983) On the distinctness of simple and complex cells in the visual cortex of the cat. J Physiol (Lond) 344:305–325.

Derrington AM, Lennie P (1984) Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleus of macaque. J Physiol (Lond) 357:219–240.

Desimone R, Gross CG (1979) Visual areas in the temporal cortex of the macaque. Brain Res 178:363–380.

Desimone R, Albright TD, Gross CG, Bruces C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051–2062.

Dow BM (1974) Functional classes of cells and their laminar distribution in monkey visual cortex. J Neurophysiol 37:927–946.

Glezer VD, Tsherbach TA, Gauselman VE, Bondarko, VM (1980) Linear and non-linear properties of simple and complex receptive fields in area 17 of the cat visual cortex. Biol Cybernet 37:195–208.

Gross CG, Rochamir CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex in macaque. J. Neurophysiol 35: 96–111.

Hertz JA, Richmond BJ, Optican LM (1991) Neural decoding. In: Pigments to perception: advances in understanding visual processes, NATO ASI series A, vol 203 (Lee B, Valsberg A, eds), pp 437–446. New York: Plenum.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol (Lond) 195:215–243.

Krazulis RJ, Lisberger SG (1990) Visual motion signals for pursuit eye movements identified on floccular Purkinje cells in monkey. Soc Neurosci Abstr 20:900.

Kruger J, Becker JD (1991) Recognizing the visual stimulus from neuronal discharges. Trends Neurosci 14:282–286.

Le Cun Y, Denker JS, Solla SA (1990) Optimal brain damage. In: Advances in neural information processing systems II (Touretzky DS, ed), pp 598–605. San Mateo: Morgan Kaufmann.

Lehky SR, Sejnowski TJ (1988) Network model of shape-from- shading: neural function arises from both receptive and projective fields. Nature 333:452–454.

Lehky SR, Sejnowski TJ (1990) Neural network model of visual cortex for determining surface curvature from images of shaded surfaces. Proc R Soc Lond [Biol] 240:251–278.

Linden A, Kinderman J (1990) Inversion of multilayer nets. In: Proceedings of the International Joint Conference on Neural Networks, Washington, DC (Caudill M, ed), pp 425–430. Hillsdale, NJ: Erlbaum.

Lund JS (1988) Anatomical organization of macaque monkey striate cortex. Annu Rev Neurosci 11:253–288.

Marmarelis PZ, Marmarelis VZ (1978) Analysis of physiological systems: the white noise approach. New York: Plenum.

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

Optican LM, Richmond B (1986) Temporal encoding of pictures by striate neuronal spike trains. II. Predicting complex cell responses. Soc Neurosci Abstr 12:431.

Poggio GF (1972) Spatial properties in striate cortex of unanesthetized macaque monkey. Invest Ophthalmol 11:368–377.

Pollen DA, Gaska JP, Jacobson LD (1988) Responses of simple and complex cells to compound sine-wave gratings. Vision Res 28:25–39.

Richmond BJ, Optican LM, Gawne TJ (1989) Neurons use multiple messages encoded in temporally modulated spike trains to represent pictures. In: Seeing contour and colour (Kulikowski JJ, Dickinson CM, Murray IJ, eds), pp 705–714. Oxford: Pergamon.

Richmond RJ, Optican LM, Podell M, Spitzer H (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. 1. Response characteristics. J Neurophysiol 57:132–146.

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Parallel distributed processing, Vol 1 (Rumelhart DE, McClelland JL, eds), pp 318–362. Cambridge, MA: MIT Press.

Schiller PH, Finlay BL, Volman SF (1976) Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. J Neurophysiol 39:1288–1319.

Spitzer H, Hochstein S (1985a) Simple and complex cell response dependencies on stimulation parameters. J Neurophysiol 53:1244–1265.

Spitzer H, Hochstein S (1985b) A complex-cell receptive field model. J Neurophysiol 53:1266–1286.

Schwartz EL, Desimone R, Albright TD, Gross CG (1983) Shape recognition and inferior temporal neurons. Proc Natl Acad Sci Biol 80:5776–5778.

Szulborski RG, Palmer LA (1990) The two dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells. Vision Res 30:249–254.

Yellott JI (1982) Spectral-analysis of spatial sampling by photoreceptors: topological disorder prevents aliasing. Vision Res 22:1205–1210.

Zipser D, Andersen RA (1988) A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature 331:679–684.