Opponent Inhibition: A Developmental Model of Layer 4 of the Neocortical Circuit

Andrew S. Kayser[3,4] and Kenneth D. Miller[1–5]

[1]Depts. of Physiology and [2]Otolaryngology
[3]Neuroscience Graduate Programs
[4]W.M. Keck Center for Integrative Neuroscience
[5]Sloan-Swartz Center for Theoretical Neurobiology at UCSF
University of California, San Francisco, CA 94143-0444

January 2, 2002

This is the final draft of a manuscript that appeared as
Neuron **33**, 131-142 (2002).

Address Correspondence To:
Kenneth Miller
Dept. of Physiology, UCSF
513 Parnassus
SF, CA 94143-0444
Phone: 415-476-8217
Fax: 415-476-4929
Email: ken@phy.ucsf.edu

Acknowledgements: This work was supported by R01-EY11001 from the NEI.

Summary

We model the development of the functional circuit of layer 4, the input-recipient layer, of cat primary visual cortex. The observed thalamocortical and intracortical circuitry codevelop under Hebb-like synaptic plasticity. Hebbian development yields opponent inhibition: inhibition evoked by stimuli anticorrelated with those that excite a cell. Strong opponent inhibition enables recognition of stimulus orientation in a manner invariant to stimulus contrast. These principles may apply to cortex more generally: Hebb-like plasticity can guide layer 4 of any piece of cortex to create opposition between anticorrelated stimulus pairs, and this enables recognition of specific stimulus patterns in a manner invariant to stimulus magnitude. Properties that are invariant across a cortical column are predicted to be those shared by opponent stimulus pairs; this contrasts with the common idea that a column represents cells with similar response properties.

One of the remarkable properties of the cerebral cortex is its division into functional columns of cells (Hubel and Wiesel, 1962; Mountcastle, 1957): cells sharing certain functional response properties are grouped vertically through the depth of cortex. In the primary visual cortex (V1) of the cat, preference for the orientation of a light/dark contrast edge and for stimulation through one or the other eye are two of the most notable response properties that show columnar organization. These properties vary continuously across the two tangential directions of cortex, giving rise to the well-known orientation and ocular dominance maps. This 2-dimensional tangential organization has been extensively investigated, and numerous modeling studies have examined its development (reviewed in Erwin et al., 1995; Miller, 1996). However, not all functional response properties show columnar organization, as we shall discuss. Few if any models have addressed the question of which functional response properties should show columnar organization.

Here we will examine a restricted version of this question, focusing on layer 4 (the input-recipient layer) of cat V1. We address the question of why some response properties, and not others, are invariant among the cells in a local region of layer 4. To be invariant across the depth of cortex, properties must be invariant in particular in layer 4; and conversely, properties that vary among nearby cells in layer 4 cannot show columnar invariance. Furthermore, given the strong vertical connections carrying information from layer 4 to the other layers (Callaway, 1998; Gilbert, 1983; Lübke et al., 2000), it seems likely that properties established in layer 4 will be shared by the other layers. Thus, we believe insight into the origins of columnar organization can be obtained by understanding layer 4. At the same time, we also address another question: how can we account for the development of the full, functional mature circuitry of layer 4? We will suggest that the answers to these two questions are closely intertwined.

Layer 4 of cat V1 consists largely of simple cells (Gilbert, 1983), cells whose receptive fields consist of separate ON (light-preferring) and OFF (dark-preferring) subregions (Fig. 1). The

Place Fig. 1 about here

alignment in visual space of these subregions confers upon the cell a preferred orientation. As suggested by Hubel and Wiesel (Hubel and Wiesel, 1962) and later confirmed (Ferster et al., 1996; Reid and Alonso, 1995; Tanaka, 1983), the orientation selectivity of simple cells originates in the arrangement of their inputs from the lateral geniculate nucleus (LGN): ON-subregions receive input from overlying ON-center LGN cells, and OFF-subregions from overlying OFF-center cells. We refer to this arrangement as "Hubel-Wiesel" connectivity.

In addition to a preferred orientation, simple cells have a preferred spatial phase (Fig. 1A). We distinguish between *relative* spatial phase, which refers to the position of the ON and OFF subregions with respect to the receptive field (RF) center, and *absolute* spatial phase, which refers to the position of the subregions with respect to the visual world. Neurons isolated on the same electrode (and presumably within the same cortical column) within cat V1 tend to have similar preferred orientations, but vary in both relative and absolute spatial phases (DeAngelis et al., 1999; Liu et al., 1992; Pollen and Ronner, 1981) (see Discussion). Importantly, then, some properties within a column, and within a local region of layer 4, are invariant (*e.g.* orientation) but others are not (*e.g.* relative and absolute spatial phases). Accounting for this organization constitutes one goal of our developmental model.

A second goal of our model is to reproduce the connectivity of the mature functional circuit of cat V1 layer 4. Reproducing the Hubel-Wiesel arrangement of geniculocortical connections is one component of this goal. The other component is to reproduce the intracortical connectivity of layer 4, about which less is known. Two key constraints are that the input to simple cells, both excitatory and inhibitory, comes primarily from other neurons with similar orientation tuning (Anderson et al., 2000a; Ferster, 1986), and that the excitation and inhibition are opponent (Anderson et al., 2000a; Ferster, 1988; Hirsch et al., 1998): a dark stimulus over an OFF subregion elicits synaptic excitation, whereas a light stimulus in the same position evokes synaptic inhibition. Figure 1B shows a cartoon of a model circuit that meets these requirements and that reproduces a wide range of experimental observa-

tions (Ferster and Miller, 2000; Kayser et al., 2001; Krukowski and Miller, 2001; Lauritzen et al., 2001; Troyer et al., 1998), including such functional response properties as the invariance of orientation tuning to stimulus contrast (Anderson et al., 2000b; Sclar and Freeman, 1982; Skottun et al., 1987; Troyer et al., 1998) and the temporal tuning properties of V1 cells (Krukowski and Miller, 2001). The model is also consistent with a number of recent experimental results suggesting that the orientation tuning of voltage responses in cat layer 4 simple cells is largely determined by the pattern of thalamic input they receive (Chung and Ferster, 1998; Ferster et al., 1996; Gillespie et al., 2001). Accounting for the simultaneous development of this model circuit and Hubel-Wiesel geniculocortical connectivity constitutes the second goal of our model.

To address the development of the layer 4 columnar circuit, we are guided by our previous results (Miller, 1994). We showed that a simple correlation structure in the LGN activity patterns, which could plausibly exist in the spontaneous (or dark) activity of LGN cells in the absence of vision, can give rise under Hebbian or "correlation-based" development to the Hubel-Wiesel pattern of LGN inputs to simple cells. This correlation structure is described in Fig. 2A. This scenario of Hebbian development guided by spontaneous activity is consistent with findings that the initial development of orientation selectivity in V1 does not depend on vision (Crair et al., 1998; Fregnac and Imbert, 1984), but can be prevented by alteration of input activities (Chapman and Gödecke, 2000). We argued (Miller, 1994) that the postulated correlation structure is a plausible one because it would naturally arise if spontaneous quantal noise in photoreceptors were the source of spontaneous LGN activity (as it is for spontaneous retinal activity in adult cats, Mastronarde (1989)), and we showed that the correlations expected under this scenario accurately predicted the preferred spatial frequencies of cortical cells. (Another commonly discussed source of correlated activity is the spontaneous "waves" of retinal activity that exist in very young animals, Wong (1999); these disappear just before the development of orientation selectivity and so are unlikely to play

Place Fig. 2 about here

a role in that development, see discussion in Miller (1994) and Erwin and Miller (1998).) Accordingly, we posit that this correlation structure exists in cat LGN when the layer 4 circuit is developing, which constitutes one key prediction of our model (see Discussion). Thus, the patterns of LGN activity in the present model are noise that has been filtered to have the postulated correlation structure (an example is in Fig. 2B). We now ask whether these LGN activity correlations are sufficient to account for the development of the layer 4 columnar unit of cat V1, *i.e.*, for the simultaneous development of the Hubel-Wiesel patterns of LGN inputs, the postulated intracortical connectivity, and local invariance of orientation but not spatial phase.

Results

The initial condition for the model is illustrated in Fig. 3A, left. We model a "column" of layer 4 cells as a group of 10 cells, 4 inhibitory and 6 excitatory. (We use 10 cells for computational speed and ease of presentation; simulations of up to 100 cells, and/or in which only 20% of cells are inhibitory, yield similar results.) Each cortical neuron initially receives unstructured geniculate input, with roughly equal strengths of ON and OFF inputs from each location. Initial intracortical connections are weak and all-to-all: *i.e.* every cell in each column is connected to every other cell in the column, excluding self-connections. These promiscuous initial connections are meant to simulate two ideas. First, initial connections are largely unstructured: very few cat layer 4 V1 neurons (approximately 15%) are orientation selective at 6-9 days of age (Albus and Wolf, 1984), modeled here by the unstructured projections from LGN; and layer 4→layer 4 connections are sparse at 5 days, but of adult complexity by 20 days (Callaway and Katz, 1992). Second, all possible connections among a local set of neurons are potentially available to the network, with selection of appropriate connections occurring by developmental rules described below. The cortex, of course, ex-

Place Fig. 3 about here

plores its possible connections through a process of sprouting-and-retraction, rather than by making all possible connections at once. We model this for simplicity by initially connecting all cells and allowing the network to prune connections over time, but models involving sprouting and retraction of connections will yield similar results if the developmental rules for selecting surviving synapses are similar (Miller, 1998). The two methods yield similar results because both involve an exploration of the space of possible connections, with retention or loss of connections determined by Hebbian rules.

The network then develops under simple rules. Excitatory connections, both geniculocortical and intracortical, develop based on a Hebbian covariance rule; neurons that "fire together, wire together". Inhibitory connections learn based on a rule derived from the experimental work of Komatsu and colleagues (Komatsu, 1994, 1996; Komatsu and Iwakiri, 1993): when the presynaptic inhibitory neuron is active, the inhibitory synapse is strengthened if it is coactivated with other inhibitory inputs to the same postsynaptic cell, and weakened if the postsynaptic cell is simultaneously active (see Methods). As with most developmental models, some form of synaptic competition must be introduced to ensure that each cell maintains a reasonable level of received weights and of projected weights. One could design dynamical negative feedback loops to attain this, perhaps expressed in terms of biological elements such as neurotrophins, but if such feedback is strong the main effect would be to constrain the Hebbian dynamics to explore only a subspace of weights (the subspace in which all cells maintain their summed received and projected weights) (Miller and MacKay, 1994). Given our current ignorance of the actual biological dynamics that achieve competition (Turrigiano and Nelson, 2000), we feel it is simplest to directly constrain the Hebbian dynamics, thus modeling the fact of competition but not its mechanism. Thus, we demand that each cortical cell must maintain constant sums of projected weights and of each type (thalamocortical, intracortical excitatory, intracortical inhibitory) of received weight.

The final column resulting from this development (Fig. 3A, middle) demonstrates five

key points. First, each of the cells has developed a Hubel-Wiesel pattern of LGN inputs. Second, all of the cells develop the same preferred orientation. Third, the cortical connections are phase-appropriate: excitatory cells project to cells of the same absolute spatial phase, while inhibitory cells project to cells of the opposite absolute spatial phase. Fourth, two absolute spatial phases develop, because inhibitory cells become anticorrelated with the cells to which they project. Fifth, the circuit replicates a key function (Troyer et al., 1998) of the cat V1 layer 4 circuit in showing contrast-invariant orientation tuning (Fig. 3A, right). With different initial conditions, the detailed connections and preferred orientation change, but the results are always identical in these five respects.

A simple intuition lies behind these results. The development of Hubel-Wiesel RF structure in the geniculocortical afferents is driven by the correlation structure in the LGN, along with Hebbian rules by which "neurons that fire together wire together", as described above and in Miller (1994). The *alignment* of the orientations of the simple cell RFs of different cortical cells is driven by the intracortical connections. Hebbian dynamics lead to a final state in which excitatory connections link cells that have strongly correlated receptive fields, while inhibitory connections link cells that have strongly anticorrelated receptive fields. Consider two cortical cells, A and B. Suppose A and B share correlated LGN inputs. As a consequence they will tend to be coactive, so that under Hebbian rules an excitatory intracortical synapse between them would grow stronger, while an inhibitory synapse would grow weaker. Alternatively, if A and B have anticorrelated LGN inputs, they will tend not to be coactive, so that an excitatory intracortical synapse between them would grow weaker, while an inhibitory synapse would, if coactive with other inhibitory inputs, grow stronger. Thus, receiving correlated or anticorrelated LGN inputs promotes development of excitatory or inhibitory connections, respectively. Conversely, excitatory or inhibitory connections promote development of correlated or anticorrelated LGN inputs, respectively. Suppose A sends an excitatory connection to B. Then when cell A fires, cell B will tend to fire as well,

so that A and B will tend to reinforce the same input patterns by Hebbian rules and so develop correlated LGN inputs. If instead cell A sends an inhibitory connection to cell B, when cell A fires cell B will tend not to fire. Cell B will thus tend to respond to, and to reinforce under Hebbian rules, input patterns anticorrelated with those that excite A. Thus, there is a strong positive feedback mechanism leading to a state in which cells connected by strong excitatory connections receive correlated sets of LGN inputs, and cells connected by strong inhibitory connections receive anticorrelated LGN inputs.

Strongly correlated receptive fields are those of similar preferred orientation and similar absolute spatial phase, while strongly anticorrelated receptive fields are those of similar preferred orientation and roughly opposite spatial phase. Thus, maximizing correlation between cells connected by excitatory synapses, and maximizing anticorrelation between cells connected by inhibitory synapses, leads the network as a whole to align in orientation, but to vary in spatial phase. This result is apparent in Fig. 3. In Fig. 3A, note that only two relative spatial phases develop. A diversity of relative spatial phases and receptive field structures will develop if the initial LGN input projections show more retinotopic scatter between cortical cells (Fig. 3B), though again only two absolute phases develop.

Each of these results is quite robust to variation in the parameters of the model (see Methods), as expected from the simple and general intuition underlying the results. The results also robustly arise across different networks run with identical parameters but beginning from different random initial conditions. Figure 4 summarizes the connectivity across 128 such simulations, 64 without retinotopic scatter (Fig. 4A) and 64 with random retinotopic scatter (Fig. 4B). In both cases, excitatory connections arose predominantly between neurons of the same preferred orientation and phase, while inhibitory connections predominantly connected neurons of the same orientation but opposite phase; these tendencies were slightly weaker with retinotopic scatter than without. To characterize the degree to which cells became orientation selective, we defined an orientation selectivity index or OSI (see

Place Fig. 4 about here

Methods); for calibration, the OSI was $.51 \pm .01$ (mean $\pm$ stdev) for the cells in Fig. 3A and $.51 \pm .06$ for the cells in Fig. 3B, and cells with OSI$\geq$ 0.18 look well-tuned to the eye. Across the 64 simulations without scatter, the OSI was $.43 \pm .09$, while with scatter it was $.41 \pm .13$; thus cells robustly showed strong orientation selectivity. To characterize the tendency of preferred orientations to line up across the ten cells in a simulation, for each simulation we computed the standard deviation of the distribution of preferred orientations observed across the ten cells (see Methods). Across the 64 simulations, this standard deviation had a mean of 2.1 degrees, median of 0.51 degrees for the cases without retinotopic scatter, and a mean of 15.2 degrees, median of 9.2 degrees, for the cases with retinotopic scatter (for comparison, the standard deviation was 3.6 degrees for the cells of Fig. 3A and 9.5 degrees for Fig. 3B, and a uniform distribution of orientations would have standard deviation 52 degrees). Thus there was a very strong tendency for orientations to be aligned.

Extensions to Multiple Columns and Maps

We have experimented with simulations of multiple columns, with weak all-to-all coupling between neighboring columns. We have found that multiple columns will simultaneously develop as just described, and that adjacent columns will tend to develop similar orientation tuning. Furthermore, adjacent columns can differ in their absolute spatial phases, which could contribute to a greater diversity of absolute phases of cells recorded on the same electrode. While this coupling produces some continuity of preferred orientation, it does not produce a truly periodic map of orientations. Because excitatory connections tend to connect correlated RFs and inhibitory connections anticorrelated RFs, and because both correlated and anticorrelated RFs share the same preferred orientation (but have the same or opposite absolute spatial phase, respectively), it seems unlikely that any simple between-column wiring scheme would produce periodicity; periodicity of orientation requires that an uncorrelated RF, representing the orthogonal orientation, be preferred at some distance.

A similar problem appeared in our previous modeling of development of geniculocortical synapses (Miller, 1994), for similar reasons. In that work, maps that appeared periodic to the eye developed, but Fourier analysis showed that the maps often tended to be low-pass (having similar power at all frequencies below some cutoff) rather than band-pass or periodic (having a peak of power at some nonzero frequency). We expect the present model would develop maps similar to those in Miller (1994), if we could simulate at higher resolution (we have thus far been limited computationally to small networks – 12 x 12 columns – with only nearest-neighbor coupling between columns).

There seem to be at least two ways that genuinely periodic maps might be produced. First, periodicity of orientation tuning might be established by horizontal interactions among complex cells in other layers, which respond to stimuli of any phase; the results could then propagate back to layer 4. Horizontal interactions among complex cells that are excitatory at short range and inhibitory at longer range, or that otherwise yield spatially periodic activity patterns (Chiu and Weliky, 2001), can suffice to organize a periodic pattern of orientations (von der Malsburg, 1973). Second, plasticity rules that might decorrelate RFs of connected cells, by weakening connections between coactive cells, have recently been described in layer 4 (Egger et al., 1999). If such rules operated preferentially in the horizontal direction while the rules modeled here operated vertically, this could also lead to periodic organization. Alternatively, perhaps V1 has specialized structures built in that ensure periodicity. It is interesting that, when visual inputs are induced to innervate auditory thalamus, orientation maps develop in auditory cortex that show continuity but do not show obvious periodicity and appear to be low-pass rather than band-pass (Sharma et al., 2000); perhaps this is the result expected from generic Hebbian mechanisms without V1-specific specializations.

Discussion

These simulations for the first time address the development of a complete, functional local circuit from unstructured initial conditions, demonstrating how the key elements of the cat V1 layer 4 circuit can simultaneously arise. In so doing, they also address for the first time the origins of columnar invariance: the reasons why some properties are locally invariant in layer 4 and thus candidates for columnar invariance, while others vary locally in layer 4 and thus cannot show columnar invariance. The key elements of the cat V1 layer 4 circuit – Hubel-Wiesel patterns of LGN input, phase-specific intracortical connectivity that produces contrast-invariant orientation tuning, and local uniformity of preferred orientation with diversity of relative and absolute spatial phase – will all codevelop guided only by plausible LGN activity correlations and simple Hebb-type rules of synaptic modification. Hebbian development in the context of a mixture of excitatory and inhibitory neurons leads to the development of at least two anticorrelated receptive field structures: if inhibitory neuron A inhibits cell B, the two cells tend to develop anticorrelated receptive fields. Thus, properties that are shared by anticorrelated pairs of receptive fields – *e.g.*, preferred orientation – are candidates for columnar invariance, while properties that differ between anticorrelated pairs – *e.g.*, absolute spatial phase – will differ locally in layer 4. The model that arrives at these conclusions is very simple, but also very robust, depending primarily on these simple and general features of Hebbian development.

The direct tests of this cat V1 model are to more directly assess the predicted phase-specificity of intracortical connectivity – *i.e.*, cells linked by excitatory connections should have similar absolute spatial phase, while those linked by inhibitory connections should have roughly opposite absolute spatial phase – and to measure the correlations of LGN activity during development of orientation selectivity to determine whether they have the predicted form. The two existing cross-correlation studies of pairs of simple cells recorded on the same

electrode seem consistent with our prediction of phase-specific connectivity, although numbers were small in both cases. Thus, DeAngelis et al. (1999) reported that cross-correlations indicative of monosynaptic excitatory connections were restricted to neuron pairs with similar space-time receptive fields, which in particular meant similar preferred orientations and absolute phase. Liu et al. (1992) reported that 3 of 3 cell pairs with antiphase receptive fields (as assessed by temporally antiphase responses to drifting sinusoids) had cross-correlations indicative of mutual inhibition. This is the result expected from our model: inhibitory cells monosynaptically inhibit cells of opposite phase, while excitatory cells drive inhibitory cells of the same phase and thus disynaptically inhibit cells of opposite phase. Liu et al. (1992) also reported that 11/13 cell pairs that showed 80-100^o temporal phase differences in their responses had cross-correlations showing no sign of interaction, again consistent with the phase-specific connectivity predicted here. While the correlations existing in LGN during development of orientation selectivity have not yet been measured, a recent experiment tested the role of the predicted LGN correlations in driving orientation selectivity in ferrets by blocking all ON-center input activity while leaving OFF-center activity intact. This disruption indeed prevented development of orientation selectivity (Chapman and Gödecke, 2000), as the model predicts.

We have only considered inputs from a single eye, but given two eyes, maximally anticorrelated receptive fields would be expected to be those representing the same eye and same orientation but having opposite absolute spatial phase. (Both correlations, for well-matched receptive fields, and anticorrelations, for roughly opposite receptive fields, are expected to be weaker between the eyes than within the eyes. For example, in the presence of vision, corresponding points in the two eyes do not always see the same stimulus, because objects are located at different disparities; averaging over scenes would reduce both correlations and anticorrelations between the eyes relative to those within the eyes. Before vision and somewhat before the development of orientation selectivity, correlations exist between the two eyes' ac-

tivities in the LGN that are weaker than within-eye correlations (Weliky and Katz, 1999). Correlations at the time that orientation selectivity develops have not yet been measured.) Thus, the same principles would be expected to predict that ocular dominance is locally invariant, consistent with the fact that ocular dominance shows columnar organization.

Diversity of Receptive Field Properties

The model with retinotopic scatter can produce a diversity of relative spatial phases of receptive fields, in agreement with DeAngelis et al. (1999), who found that relative phase varied roughly randomly between two cells recorded on the same electrode. However, a probable weakness of the model is that it produces only two absolute spatial phases. The experimental data on this point suggests that multiple absolute phases are found at single recording sites, though the data are not conclusive. DeAngelis et al. (1999) found that correlation coefficients between the two spatial receptive fields of cell pairs ranged fairly uniformly between -1 and 1, which is the result expected if absolute phases show a random distribution. However, this measure is affected by differences in the spatial positions of the two receptive fields as well as by differences in their absolute phases. For example, Fig. 2 of DeAngelis et al. (1999) shows two cells that appear to have roughly identical absolute phase, but due to spatial position differences the correlation coefficient of their spatial receptive fields is only 0.23. Thus it is difficult to draw a firm conclusion as to the distribution of absolute phases from the similarity data. Earlier studies examined the *temporal* phase of response of cells to sinusoidal gratings, reporting that pairs of cells recorded on the same electrode tended to respond either in antiphase (180^o out of phase) or in quadrature (90^o out of phase) (Liu et al., 1992; Pollen and Ronner, 1981). Such measurements reflect a combination of absolute spatial phase and the temporal phase of response. For example, two cells could respond with 90^o temporal phase difference if they had the same absolute spatial phase, but one received non-lagged LGN inputs and one received lagged LGN inputs (Saul

and Humphrey, 1992).

Both the similarity data, showing that spatial correlation coefficients can approach -1, and the finding of cells in temporal antiphase seem consistent with the idea of the present model that each of two opposite absolute phases can be found among cells recorded on the same electrode. However, both the similarity data, showing many cell pairs with weak correlation coefficients, and the finding of neuron pairs in quadrature suggest that nearby cells may have uncorrelated as well as correlated or anticorrelated receptive fields. As discussed previously, the experiments suggest, consistent with the model, that only the strongly correlated or strongly anticorrelated pairs show signs of connections (excitatory or inhibitory, respectively); the uncorrelated pairs appear to be unconnected. This suggests that our "columns" may be but one strongly connected set among many such sets that coexist in the same local spatial region, so that a more complete model would have multiple spatially overlapping "columns", with stronger connectivity within a column and weaker connectivity between columns. Indeed, as noted previously, in simulations of development of multiple 10-cell "columns" with weak connections between adjacent columns, adjacent columns can develop similar preferred orientation but differ in their absolute spatial phases. A more complete model would also have a greater diversity of spatial and temporal receptive field types, which would provide more dimensions along which receptive fields may differ and become decorrelated. In the model, the preferred spatial frequency of cells is determined by the correlation structure of the LGN inputs, and there is no diversity in this structure – correlations between inputs are simply determined by the distance between two inputs. In reality, there are multiple subtypes of LGN cells and diversity of LGN receptive field shapes and sizes, which are likely to lead to diversity of correlations and in turn to diversity of cortical preferred spatial frequencies. In addition, we have considered only two temporal types of LGN inputs: an ON and an OFF cell representing the same location respond at opposite temporal phases to a sinusoidally-modulated stimulus. In reality there is a bimodally-distributed

continuum of ON-center temporal response types and similarly for OFF-center (Wolfe and Palmer, 1998), where the modes represent lagged and non-lagged cells.

The key question for a model is why, as more complexity is considered, preferred orientation should remain locally invariant as more variability arises in other receptive field properties. We suggest that this may arise for the same reason we have already seen in the more simplified context: even as other response properties vary, cells will tend to maximize correlations (for excitatory connections) or anticorrelations (for inhibitory connections) by maximizing subregion overlap or antioverlap, and this in turn requires cells to maintain a roughly common preferred orientation. Alonso et al. (2001) studied connections from LGN to cortical simple cells, and showed that spatial phase was the strongest determinant of a connection – i.e. an LGN cell's ON-center would overlap an ON-subregion of a simple cell to which it connected – but that other properties, such as a match of center size to subregion width or a match of temporal response properties, more weakly constrained connectivity. This suggests that variations in these other properties have a weaker impact on the correlations that guide development. If the same principles apply to intracortical connections, then preferred spatial frequency (the cortical analogue of center size) and temporal response properties might vary between, or even within, weakly connected columns, but subregions would still become arranged to maximize overlap or anti-overlap and therefore align their orientations. (Preferred spatial frequency also shows some columnar organization (Issa et al., 2000), but this organization is much weaker than that of orientation and there is approximately a ± 0.5 octave variation in preferred spatial frequency among cells at the same recording site, A. Emondi, A. Kurgansky, S. Rebrik and K.D. Miller, unpublished results.) Such variation of more weakly constrained response properties might be sufficient to account for the experimental data. A central issue for future theoretical work is to explore whether and under what conditions such a scenario can indeed generate greater receptive field diversity while maintaining roughly constant preferred orientations.

Comparison to Previous Work

The present work knits together elements of our previous work. We have previously shown that a correlation structure like that used here will lead to development of a Hubel-Wiesel pattern of LGN inputs to simple cells, in the context of fixed intracortical connectivity (Miller, 1994), but we did not address the simultaneous development of intracortical connections. We have also previously shown that the model intracortical circuit cartooned in Fig. 1 can account for many aspects of the functional responses of cat V1 layer 4 (Ferster and Miller, 2000; Kayser et al., 2001; Krukowski and Miller, 2001; Lauritzen et al., 2001; Troyer et al., 1998), but we did not address the development of this circuit. Here we are for the first time addressing the development of a fully functional circuit, including the simultaneous co-development of geniculocortical and intracortical connectivity, as well as proposing, based on this, more general principles that may underly the development of properties that show columnar invariance.

Some previous models have examined codevelopment of horizontal intracortical connections and LGN connections in a 2-dimensional model cortex (Bartsch and van Hemmen, 2001; Sirosh and Miikkulainen, 1997). These models examined development of horizontal patterns of intracortical connections (across the 2-dimensional map), but did not address the development of the local intracortical functional circuit. Olson and Grossberg (1998), in work contemporary with the present work (*e.g.*, see Kayser and Miller, 1998), showed that, if intracortical circuitry were fixed so that cortical cells came in mutually inhibitory "dipole" pairs, and LGN inputs then developed given this fixed circuitry, then the mutually inhibitory pairs would develop antiphase receptive fields. However, they did not study how such intracortical and LGN connectivity could co-develop.

Generalizations for Layer 4 of the Cortical Circuit

The present results suggest direct generalizations from the functional properties of cat V1 to more general properties of the cortical layer 4 circuit, which in turn can be tested experimentally. The only thing specific to cat V1 in our model is the nature of the inputs received – the inputs come in two varieties, ON-center and OFF-center, and have a correlation structure similar to that shown in Fig. 2A – yet this is sufficient to account for at least a functioning skeleton of the layer 4 cat V1 circuit. This suggests the generalization that layer 4 more generally develops through simple Hebb-like rules, instructed simply by the correlations in the activities of its inputs. Given a co-developing mix of excitatory and inhibitory cells, this leads naturally to opponent inhibition: cells become excited by one coactive input pattern and are inhibited by the input pattern most anticorrelated with this pattern.

Opponent inhibition has been a common theme in studies of the visual system (*e.g.*, Anderson et al., 2000a; Ferster, 1988; Hirsch et al., 1998; Palmer and Davis, 1981), where the fundamental opponency between light and dark stimuli has a neural correlate in ON- and OFF-center inputs, but to our knowledge it has not previously been considered that opponency might be a more general cortical property. What might constitute opponent stimuli in other systems? In the auditory system, spectro-temporal receptive fields that resemble the spatial receptive fields of simple cells are observed (*e.g.*, deCharms et al., 1998), raising the possibility that opponency might be observed between cells with antiphase spectro-temporal receptive fields (that is, pairs in which one cell is excited at the frequencies and delays that inhibit the other). More generally, opponency requires that the two opponent stimuli activate largely nonoverlapping spacetime patterns of inputs, which patterns nonetheless converge in the same local region of cortex. Application to other systems depends on considering thalamic activity patterns in this light, as well as on directly studying "who inhibits whom" in cortex.

What is the function of opponent inhibition in layer 4? Provided that this feedforward inhibition dominates feedforward excitation (as suggested by the prolonged inhibition following nonspecific stimulation of LGN, Ferster and Jagadeesh (1992), or cortex, Chung and Ferster (1998)), it enables layer 4 cells to respond specifically to the input pattern that excites them, in a manner invariant to stimulus magnitude (Troyer et al., 1998) (Fig. 5); this is the generalization of contrast-invariant orientation tuning in V1. The argument, which is spelled out in the legend of Fig. 5, is essentially very simple: suppose a cell is excited by an input pattern **A**, and receives opponent inhibition driven by the anticorrelated input pattern $\bar{\mathbf{A}}$. The set of inputs active in pattern **A** will be essentially non-overlapping with the set active in pattern $\bar{\mathbf{A}}$. If an input pattern **B** is uncorrelated with **A** but shares some inputs with **A**, it will also tend to share some inputs with $\bar{\mathbf{A}}$; because the inhibition is dominant, this will yield a net inhibition. Thus a pattern must be very close in structure to **A** to yield net excitation. The degree of "closeness" required increases with the strength of inhibition but is largely invariant to changes in stimulus magnitude. Thus a cell can respond to its preferred stimulus **A** even when presented at very low magnitude, but will not respond to some partially overlapping stimulus **B** even when presented at high magnitude.

Place Fig. 5 about here

Layer 4 cells project strongly to layers 2/3 (Callaway, 1998; Gilbert, 1983; Lübke et al., 2000), where in V1 complex cells are found (Gilbert, 1983) – cells that respond to stimuli of a given orientation, independent of spatial phase. This suggests a further generalization, that layer 2/3 cells may be sensitive to the stimulus attributes that anticorrelated opponent pairs have in common, and insensitive to the attributes that differentiate such pairs.

Conclusion

In sum, we predict that experiments that ask "who inhibits whom?" in layer 4 of other cortical areas (*e.g.* primary somatosensory or auditory cortex) will find the answer to be neurons that represent patterns that are anticorrelated in the input structure, as expected

from general considerations of Hebbian learning; that the features that distinguish these opponent pairs will not be invariant within a column, while features that they have in common will be candidates for columnar invariance; that this opponent inhibition will endow layer 4 cells with magnitude-invariant pattern recognition; and that layer 2/3 cells may be sensitive to the features that opponent pairs have in common and insensitive to the features that distinguish them. Further study will certainly complicate this simple picture. For example, the concepts of "correlated" and "anticorrelated" activity patterns will need further refinement: on what time scales must neurons be coactive (or not) to be correlated or anticorrelated with respect to the Hebbian plasticity rules operating in neocortex (Feldman, 2000; Markram et al., 1997; Song et al., 2000)? Nonetheless, we believe that these ideas provide a novel and testable framework for studying the organization of cerebral cortex that is solidly rooted both in the many person-years of experimental study of cat V1 and in simple consequences of Hebbian development.

Methods

Cortical cells are arranged in a column of 6 excitatory cells and 4 inhibitory cells. (Columns with 16 excitatory cells and 4 inhibitory neurons, or 80 and 20, have also been used, with similar results.) Two 16×16 arrays of LGN cells, representing ON and OFF LGN neurons respectively, provide input. All cortical cells receive connections from both ON and OFF LGN neurons over a circle of diameter 13 centered on the cell's retinotopic position; the strength of connection is scaled by an arbor function identical to that used previously (Miller, 1994), with maximum value set to 1. All cortical cells are centered on the same retinotopic point unless retinotopic scatter is introduced (*e.g.*, Fig. 3B). To model scatter (Albus, 1975; Hubel and Wiesel, 1974), RF position is randomly chosen from a uniform distribution within an area of radius 3 (half the arbor function radius). Intracortical weights are initially all-to-

all; each cortical cell, whether excitatory or inhibitory, receives projections from all other cells in the column (no self-connections are made). All weights are initially assigned randomly from a uniform distribution ranging from 0.4 to 0.6, scaled by the arbor function in the case of geniculocortical weights. Each class of weight (*i.e.* geniculocortical and the 4 types of intracortical postsynaptic weights – see below) is then multiplicatively renormalized such that the sum of all weights of each type received by a cell matches a preset value for each cell, as described below.

We construct LGN activity patterns with appropriate correlations as follows (Goodhill, 1993). An LGN neuron with two-dimensional position α and of type ON (N) or OFF (F) is randomly assigned an initial activity $r_0^{N/F}(\alpha)$ of -0.5 or 0.5. These activities are then correlated with strength $h = 0.2$ (where $h = 0.5$ is perfect correlation, and $h = 0.0$ is no correlation) between ON and OFF layers: $r_{cor}^{N/F}(\alpha) = (1-h)r_0^{N/F}(\alpha) + hr_0^{F/N}(\alpha)$. Values of $h \geq 0.1$ gave essentially identical results. The values $r_{cor}^{N/F}$ are then convolved with a difference-of-Gaussians function $C(\alpha) = e^{-\alpha^2/\sigma^2} - (1/9)e^{-\alpha^2/(3\sigma)^2}$ for $\sigma = 1.54$ grid units; C is applied directly to the ON LGN, but multiplied first by -1 for the OFF LGN. This yields a set of spatially correlated activities $r_{sc}^{N/F}$. The function C embodies our hypothesis about the second-order structure of LGN activity patterns: ON (OFF) cells tend to fire with other ON (OFF) cells at short distances, but to fire with OFF (ON) cells at larger distances (Fig. 2A,B). Wrap-around boundary conditions are applied at LGN grid borders. Finally, all activities are rectified at zero to give rise to LGN spike rates $s_\alpha^{N/F} = \left[r_{sc}^{N/F}(\alpha)\right]^+$ (here, $[x]^+ = x,\ x > 0;\ = 0$, otherwise). Note that the correlation structure that results is not precisely like C^{ORI} of Fig. 2A, because it has an additional positive-going phase at further retinotopic separations, but this correlation structure also meets the conditions we have defined (Miller, 1994) for development of simple cell receptive fields.

We give each cortical cell an index (from 1 to 10) and let $\mathcal{E}, \mathcal{I}$ represent the indices of the excitatory or inhibitory cells respectively. Let w_{xy} represent the weight from cortical cell y

to cortical cell x, and $w_{x\alpha}^{N/F}$ represent the weights from ON/OFF LGN cells α to cortical cell x. Given the LGN activities $s_{\alpha}^{N/F}$, the activity v_x of cortical cell x is computed as follows:

$$\frac{dv_x}{dt} = -v_x + \sum_{y\in\mathcal{E}} w_{xy} f^E(v_y) - \sum_{y\in\mathcal{I}} w_{xy} f^I(v_y) + \sum_{\alpha;J\in\{N,F\}} w_{x\alpha}^J s_{\alpha}^J \tag{1}$$

The functions f^E and f^I are linear approximations to a sigmoid of the form $f(v) = 0$, $v \leq 0$; $= \gamma v$, $0 < \gamma v < f_{\max}$; $= f_{\max}$, $\gamma v \geq f_{\max}$. For f^E, $\gamma = f_{\max} = 1$; for f^I, $\gamma = 1.5$ and $f_{\max} = 2$, reflecting the higher gain and peak firing rate of inhibitory neurons. For a given LGN input pattern, we iterate cortical activity until a steady state is reached, and then apply plasticity rules to this combination of LGN and cortical activity; a new LGN input pattern is then generated and the process repeated.

Excitatory connections, both geniculo- and intra-cortical, develop based on a Hebbian covariance rule, so that cells that "fire together, wire together". For a given intracortical excitatory weight w_{xy}, for instance, if $\bar{v}$ is the recent time-average of the activity v, then

$$\delta w_{xy} \propto \begin{cases} v_x > \bar{v}_x \text{ or } v_y > \bar{v}_y & : \quad (v_x - \bar{v}_x)(v_y - \bar{v}_y) \\ \text{otherwise} & : \quad 0 \end{cases} \tag{2}$$

Inhibitory connections learn based on a rule derived from the experimental work of Komatsu (Komatsu, 1994, 1996; Komatsu and Iwakiri, 1993). In rat visual cortical slice he found:

1. In the presence of blockers of excitatory neurotransmission (DNQX and APV), a stable increase in the initial slope of the IPSP (LTP) occurs in response to brief tetani (Komatsu and Iwakiri, 1993);
2. This LTP is not dependent on post-synaptic voltage (Komatsu, 1994);
3. LTP induction is prevented by the application of GABA_B, but not GABA_A, antagonists (Komatsu, 1996);

4. If APV is removed from the bath solution, tetanic stimulation or application of NMDA to the post-synaptic neuron induces a decrease in the initial slope of the IPSP (Komatsu and Iwakiri, 1993) (LTD);

We translated these findings into the following computationally convenient rules for inhibitory plasticity:

1. LTP occurs when the postsynaptic cell is both receiving more inhibition than its recent time average, irrespective of the membrane voltage, and the presynaptic input is greater than its recent time average.
2. LTD occurs when the postsynaptic cell is depolarized, and the presynaptic input is greater than its recent time average.
3. Under other conditions, no change occurs.

This rule takes the following form for an inhibitory weight w_{xy}:

$$\delta w_{xy} \propto -[v_x - \bar{v}_x]^+[v_y - \bar{v}_y]^+ + [i_x - \bar{i}_x]^+[v_y - \bar{v}_y]^+ \tag{3}$$

where $i_x = \sum_{z \in \mathcal{I}} w_{xz} v_z$ is the summed inhibitory input received by the postsynaptic cell.

In practice, the above weight changes are accumulated over some number of input patterns (usually 40) before the values of the weights are updated. Additionally, before the constraints described below are applied, for the first 200 batches (sets of 40 patterns) of each simulation we separately normalize (1) all the geniculocortical and (2) all the intracortical weight changes such that their respective rms values match preset learning rates (generally an rms value of 0.001 for each). This normalization controls the speed with which the simulations develop. After the first 200 batches, the last normalization factor used is applied for the remainder of the simulation. We ran a total of 15000 batches per simulation.

To implement competition, we follow weight updates with subtractive normalization of various weight sums (Miller and MacKay, 1994). We conserve the sum of each weight

type received by each cell: geniculocortical weights, intracortical excitatory weights, and intracortical inhibitory weight. The conserved weight sums were: geniculocortical: 1.0; $E \rightarrow E$: 0.125; $E \rightarrow I$: 0.5; $I \rightarrow E$: 2.25; $I \rightarrow I$: 0.25. These numbers were chosen by running a simulation with all weight sums equal to 1.0, then multiplicatively scaling weight sums in the final state to determine those sets of sums that gave contrast-invariant orientation tuning, and choosing one of the latter. The resulting numbers reflect the overall dominance of inhibition, as required to achieve such tuning (Troyer et al., 1998) and as suggested by experiments (Chung and Ferster, 1998; Ferster and Jagadeesh, 1992). Inhibitory synaptic strengths were multiplied by a factor which ramped linearly from 0.2 to 1.0 over the first 6000 batches and remained 1.0 thereafter. This increase in inhibitory strength, which mimics the experimentally observed increase of inhibitory strength over development (Ben-Ari et al., 1997; Luhmann and Prince, 1991), was included because of concern that the dominant inhibition could initially prevent excitatory cells from responding to any pattern and thus from learning. We also conserved the total weight projected by each cortical cell. This was set for each cell to the value found after initial normalization of received weights to the above values. Over 64 columns, the conserved projected weights were: excitatory, 0.44 ± 0.02 (mean $\pm$ stdev); inhibitory, 3.56 ± 0.14. The normalization was applied either pre- or post-synaptically at alternating time points, for practical reasons: simultaneously satisfying both pre- and postsynaptic constraints is computationally difficult, but with a small learning rate is well approximated by alternation.

Weights are bounded below by 0, and above by a maximum value. This maximum is 0.018 times the value of the arbor function at a given retinotopic position for geniculocortical weights, and 0.5 times the received weight sum for each intracortical weight type. Weights in these simulations ultimately tend to go to the maximal or minimal allowed values (Miller and MacKay, 1994) so, given the constraint on total synaptic weight received, the choice of the maximum value determines the number of nonzero synapses ultimately received, *i.e.* the

degree of pruning of the initial all-to-all connectivity. The chosen values lead to pruning of somewhat more than 1/2 of geniculocortical connections (effects of varying this parameter were examined in Miller (1994, Fig. 10)), and to pruning of intracortical connections so that each cell tends to receive intracortical input from two excitatory and two inhibitory cells. Results are not sensitive to these values; the main requirement is that more than 1/2 of both geniculocortical and intracortical connections be pruned, so that geniculocortical inputs can develop clean ON/OFF segregation and intracortical connections can become restricted to those that are phase appropriate. (These arguments are true of the 10-cell simulations; we have not systematically studied other cases.) To simultaneously constrain individual weights and weight sums, we follow the algorithm of Erwin and Miller (1998).

The network is quite robust to variation in parameters. The standard deviation of the range of orientations preferred by cells within a column is unaffected by varying the sizes of any one of the weight sums (geniculocortical, $E \rightarrow E$, $E \rightarrow I$, $I \rightarrow E$, $I \rightarrow I$) between 0.25 and 4, while leaving the other four fixed; as mentioned above, the weight sums chosen were selected primarily to ensure contrast-invariant orientation tuning in the mature circuit. Varying "averaging" parameters such as the number of stimulus presentations over which we averaged before updating the weights, or the number of activity iterations in cortex for each stimulus presentation, had no effect on cortical alignment of orientations provided the number of iterations was 5 or greater. Even differences between geniculocortical and intracortical learning rates of up to 5 orders of magnitude had only minimal effects on orientation alignment, provided the network was shown enough stimuli to allow development at very low learning rates. The ramping-up of inhibitory strength could proceed over anywhere from 3000 to all 15,000 batches without altering results. A thorough analysis of parameter dependence of results in the 10-cell model can be found in Kayser (1999).

To determine the orientation selectivity index (OSI) and preferred orientation or spatial frequency for a cell, we take the Fourier transform of the LGN receptive field, defined as the

difference between ON and OFF weights at each point in the receptive field. The preferred orientation or preferred spatial frequency is taken to be the orientation or spatial frequency corresponding to the maximum of the Fourier transform. To compute the OSI, we form an orientation tuning curve $R_j = R(\theta_j)$ as the maximum of the amplitude of the Fourier transform in each of 18 10-degree orientation bins, where the j^{th} bin is centered on θ_j. We then take the Fourier transform of this tuning curve, $\tilde{R}_n = \sum_j e^{2\pi i n j/18} R_j$. The OSI is then defined to be $\sqrt{2}|\tilde{R}_1|/\sqrt{\sum_i |\tilde{R}_i|^2}$. The factor of $\sqrt{2}$ is included so that the OSI is 1 if all the power is in the first harmonic (note, because R_j is real, $|\tilde{R}_1| = |\tilde{R}_{-1}|$ where $\tilde{R}_{-1} = \tilde{R}_{17}$). We showed in Miller (1994) that, without the factor of $\sqrt{2}$, cells appeared strongly orientation selective to the eye for OSI$\geq$ .13, which corresponds to OSI$\geq$ .18 with the present normalization. To determine the spread of preferred orientations across a column, we computed the standard deviation of the distribution of preferred orientations as follows. Preferred orientations were in the range 0 to 180 degrees. We considered a line ranging from 0^o to 360^o, and placed each preferred orientation θ twice on this line, once at θ and once at $180^o + \theta$. We then found the shortest line segment that contained all of the column's preferred orientations once. We then computed the ordinary mean and standard deviation about this mean of the orientations on this segment.

To determine orientation tuning curves, we display fixed, oriented gratings of 36 orientations and 8 spatial phases, with spatial frequency equal to the mean preferred spatial frequency across cells in multiple simulations with identical parameters. The steady-state response of each cortical neuron to each stimulus is determined as described above. Responses are averaged over spatial phases to yield the neuron's orientation tuning curve. To determine the LGN response to a grating of a given contrast, we equate the mean activity of the LGN inputs during the developmental simulations (about 0.275 in our arbitrary units) with a rate of 15 Hz, corresponding to experimentally observed background rates (Levine and Troy, 1986). We then determine the firing rates in response to a sinusoidal grating of a

given contrast from the data of Sclar (Sclar, 1987) for LGN responses to 2 Hz gratings. We assume LGN firing rates are sinusoidally modulated across space about the background rate of 15 Hz, except that negative rates are set to zero, with ON-center and OFF-center LGN cells showing opposite phase of the modulation; the strength of this sinusoidal modulation ($0.3932, 0.5977, 0.9366$ and 1.3262 for 10, 20, 40, and 80% contrast, respectively) is set so that the first harmonic of the resulting rectified sinusoid equalled Sclar's value.

References

Albus, K. (1975). A quantitative study of the projection area of the central and the paracentral visual field in area 17 of the cat. I. The precision of the topography. *Exp. Brain Res.*, 24, 159–179.

Albus, K. and Wolf, W. (1984). Early post-natal development of neuronal function in the kitten's visual cortex: A laminar analysis. *J. Physiol.*, 348, 153–185.

Alonso, J. M., Usrey, W. M., and Reid, R. C. (2001). Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *J. Neurosci.*, 21, 4002–4015.

Anderson, J. S., Carandini, M., and Ferster, D. (2000a). Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J. Neurophysiol.*, 84, 909–926.

Anderson, J. S., Lampl, I., Gillespie, D., and Ferster, D. (2000b). The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science*, 290, 1968–1972.

Bartsch, A. P. and van Hemmen, J. L. (2001). Combined Hebbian development of geniculocortical and lateral connectivity in a model of primary visual cortex. *Biol. Cybern.*, 84, 41–55.

Ben-Ari, Y., Khazipov, R., Leinekugel, X., Caillard, O., and Gaiarsa, J. L. (1997). $GABA_A$, NMDA, and AMPA receptors: A developmentally regulated 'ménage à trois'. *Trends Neurosci.*, 20, 523–529.

Callaway, E. (1998). Local circuits in primary visual cortex of the macaque monkey. *Ann. Rev. Neurosci.*, 21, 47–74.

Callaway, E. M. and Katz, L. C. (1992). Development of axonal arbors of layer 4 spiny neurons in cat striate cortex. *J. Neurosci.*, 12, 570–582.

Chapman, B. and Gödecke, I. (2000). Cortical cell orientation selectivity fails to develop in the absence of on-center retinal ganglion cell activity. *J. Neurosci.*, 20, 1922–1930.

Chiu, C. and Weliky, M. (2001). Spontaneous activity in developing ferret visual cortex in vivo. *J. Neurosci.*, 21, (in press).

Chung, S. and Ferster, D. (1998). Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron*, 20, 1177–89.

Crair, M. C., Gillespie, D. C., and Stryker, M. P. (1998). The role of visual experience in the development of columns in cat visual cortex. *Science*, 279, 566–570.

DeAngelis, G. C., Ghose, G. M., Ohzawa, I., and Freeman, R. D. (1999). Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. *J. Neurosci.*, 19, 4046–4064.

deCharms, R. C., Blake, D. T., and Merzenich, M. M. (1998). Optimizing sound features for cortical neurons. *Science*, 280, 1439–1443.

Egger, V., Feldmeyer, D., and Sakmann, B. (1999). Coincidence detection and changes of synaptic efficacy in spiny stellate neurons in rat barrel cortex. *Nature Neurosci.*, 2, 1098–1105.

Erwin, E. and Miller, K. D. (1998). Correlation-based development of ocularly-matched orientation maps and ocular dominance maps: Determination of required input activity structures. *J. Neurosci.*, 18, 9870–9895.

Erwin, E., Obermayer, K., and Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: A critical comparison. *Neural Comput.*, 7, 425–468.

Feldman, D. (2000). Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron*, 27, 45–56.

Ferster, D. (1986). Orientation selectivity of synaptic potentials in neurons of cat primary visual cortex. *J. Neurosci.*, 6, 1284–1301.

Ferster, D. (1988). Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *J. Neurosci.*, 8, 1172–1180.

Ferster, D., Chung, S., and Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, 380, 249–252.

Ferster, D. and Jagadeesh, B. (1992). EPSP-IPSP interactions in cat visual cortex studied with in vivo whole-cell patch recording. *J. Neurosci.*, 12(4), 1262–1274.

Ferster, D. and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Ann. Rev. Neurosci.*, 23, 441–471.

Fregnac, Y. and Imbert, M. (1984). Development of neuronal selectivity in the primary visual cortex of the cat. *Physiol. Rev.*, 64, 325–434.

Gilbert, C. D. (1983). Microcircuitry of the visual cortex. *Ann. Rev. Neurosci.*, 6, 217–247.

Gillespie, D. C., Lampl, I., Anderson, J. S., and Ferster, D. (2001). Dynamics of the orientation-tuned membrane potential response in cat primary visual cortex. *Nature Neurosci.*, 4, 1014–1019.

Goodhill, G. J. (1993). Topography and ocular dominance: A model exploring positive correlations. *Biol. Cybern.*, 69, 109–118.

Hirsch, J. A., Alonso, J.-M., Reid, R. C., and Martinez, L. (1998). Synaptic integration in striate cortical simple cells. *J. Neurosci.*, 18, 9517–9528.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160, 106–154.

Hubel, D. H. and Wiesel, T. N. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter and magnification factor. *J. Comp. Neurol.*, 158, 295–306.

Issa, N. P., Trepel, C., and Stryker, M. P. (2000). Spatial frequency maps in cat visual cortex. *J. Neurosci.*, 20, 8504–8514.

Kayser, A. S. (1999). *Studies of the development and mature function of a candidate layer 4 cortical circuit*. PhD thesis, University of California, San Francisco.

Kayser, A. S. and Miller, K. D. (1998). The development of the cat layer 4 orientation circuit: Origins of columnar organization and contrast invariance. *Soc. Neurosci. Abstr.*, 24, 261.

Kayser, A. S., Priebe, N. J., and Miller, K. D. (2001). Contrast-dependent nonlinearities arise locally in a model of contrast-invariant orientation tuning. *J. Neurophysiol.*, 85, 2130–2149.

Komatsu, Y. (1994). Age-dependent long-term potentiation of inhibitory synaptic transmission in rat visual cortex. *J. Neurosci.*, 14, 6488–6499.

Komatsu, Y. (1996). $GABA_B$ receptors, monoamine receptors, and postsynaptic inositol trisphosphate-induced Ca^{2+} release are involved in the induction of long-term potentiation at visual cortical inhibitory synapses. *J. Neurosci.*, 16, 6342–6352.

Komatsu, Y. and Iwakiri, M. (1993). Long-term modification of inhibitory synaptic transmission in developing visual cortex. *NeuroReport*, 4, 907–910.

Krukowski, A. E. and Miller, K. D. (2001). Thalamocortical NMDA conductances and intracortical inhibition can explain cortical temporal tuning. *Nature Neuroscience*, 4, 424–430.

Lauritzen, T. Z., Krukowski, A. E., and Miller, K. D. (2001). Local correlation-based circuitry can account for responses to multi-grating stimuli in a model of cat V1. *J. Neurophysiol.*, 86, 1803–1815.

Levine, M. W. and Troy, J. B. (1986). The variability of the maintained discharge of cat dorsal lateral geniculate cells. *J. Physiol.*, 375, 339–359.

Liu, Z., Gaska, J. P., Jacobson, L. D., and Pollen, D. A. (1992). Interneuronal interaction between members of quadrature phase and anti-phase pairs in the cat's visual cortex. *Vision Res.*, 32, 1193–1198.

Lübke, J., Egger, V., Sakmann, B., and Feldmeyer, D. (2000). Columnar organization of dendrites and axons of single and synaptically coupled excitatory spiny neurons in layer 4 of the rat barrel cortex. *J. Neurosci.*, 20, 5300–5311.

Luhmann, H. and Prince, D. (1991). Postnatal maturation of the GABAergic system in rat neocortex. *J. Neurophysiol.*, 65, 247–263.

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275, 213–215.

Mastronarde, D. N. (1989). Correlated firing of retinal ganglion cells. *Trends Neurosci.*, 12, 75–80.

Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *J. Neurosci.*, 14, 409–441.

Miller, K. D. (1996). Receptive fields and maps in the visual cortex: Models of ocular dominance and orientation columns. In Domany, E., van Hemmen, J. L., and Schulten, K., editors, *Models of Neural Networks III*, pages 55–78. Springer-Verlag, New York. Available as ftp://ftp.keck.ucsf.edu/pub/ken/miller95.ps.

Miller, K. D. (1998). Equivalence of a sprouting–and–retraction model and correlation–based plasticity models of neural development. *Neural Comput.*, 10, 529–547.

Miller, K. D. and MacKay, D. J. C. (1994). The role of constraints in Hebbian learning. *Neural Comput.*, 6, 100–126.

Mountcastle, V. B. (1957). Modality and topographic organization of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.*, 20, 408–434.

Olson, S. and Grossberg, S. (1998). A neural network model for the development of simple and complex cell receptive fields within cortical maps of orientation and ocular dominance. *Neural Networks*, 11, 189–208.

Palmer, L. A. and Davis, T. L. (1981). Receptive-field structure in cat striate cortex. *J. Neurophysiol.*, 46, 260–276.

Pollen, D. A. and Ronner, S. F. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science*, 212, 1409–1411.

Reid, R. C. and Alonso, J. M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378, 281–284.

Saul, A. B. and Humphrey, A. L. (1992). Evidence of input from lagged cells in the lateral geniculate nucleus to simple cells in cortical area 17 of the cat. *J. Neurophysiol.*, 68, 1190–1208.

Sclar, G. (1987). Expression of "retinal" contrast gain control by neurons of the cat's lateral geniculate nucleus. *Exp. Brain Res.*, 66, 589–596.

Sclar, G. and Freeman, R. D. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Exp. Brain Res.*, 46, 457–461.

Sharma, J., Angelucci, A., and Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature*, 404, 841–847.

Sirosh, J. and Miikkulainen, R. (1997). Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Comput.*, 9, 577–594.

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., and Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behavior. *J. Neurophysiol.*, 57, 773–786.

Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3, 919–926.

Tanaka, K. (1983). Cross-correlation analysis of geniculostriate neuronal relationships in cats. *J. Neurophysiol.*, 49, 1303–1318.

Troyer, T. W., Krukowski, A., Priebe, N. J., and Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: Feedforward tuning and correlation-based intracortical connectivity. *J. Neurosci.*, 18, 5908–5927.

Turrigiano, G. G. and Nelson, S. B. (2000). Hebb and homeostasis in neuronal plasticity. *Curr. Opin. Neurobiol.*, 10, 358–364.

von der Malsburg, C. (1973). Self-organization of orientation selective cells in the striate cortex. *Kybernetik*, 14, 85–100.

Weliky, M. and Katz, L. C. (1999). Correlational structure of spontaneous neuronal activity in the developing lateral geniculate nucleus *in vivo*. *Science*, 285, 599–604.

Wolfe, J. and Palmer, L. A. (1998). Temporal diversity in the lateral geniculate nucleus of cat. *Vis. Neurosci.*, 15, 653–675.

Wong, R. O. (1999). Retinal waves and visual system development. *Annu. Rev. Neurosci.*, 22, 29–47.

Figure Legends

Figure 1

Receptive Field and Circuit Structure. **A**. Cartoon of receptive fields (RFs) of simple cells preferring vertically-oriented stimuli. ON subregions are indicated by white, OFF subregions by black. Position of each RF in visual space is indicated by position within the figure – ie. within each RF pair, the RFs are slightly offset. The RFs within the left pair share the same relative spatial phase (relative to the RF center), but have differing absolute spatial phases (relative to a fixed point in the visual world), while the RFs within the right pair have differing relative phases, but share the same absolute phase. **B**. Cartoon of the postulated mature circuit. Here, all four indicated RFs are meant to be centered on the same spatial position, so that identical receptive fields represent cells with the same absolute spatial phase, while opposite receptive fields represent cells with opposite absolute spatial phase. Excitatory weights tend to link neurons of the same absolute spatial phase, while inhibitory weights tend to connect neurons of the opposite absolute spatial phase. The circuit model cartooned here, excepting inhibitory-to-inhibitory connections, was previously shown to account for many experimental observations, including the contrast-invariance of orientation tuning (Kayser et al., 2001; Krukowski and Miller, 2001; Lauritzen et al., 2001; Troyer et al., 1998).

Figure 2

The Role of Correlated LGN Input. **A**. Development of geniculocortical (GC) connections. It was previously demonstrated (Miller, 1994) that a simple correlation structure is sufficient to generate simple cell receptive fields (upper right). The required correlation structure is summarized by the shape of the function C^{ORI} (lower left), which represents

the *difference* between one LGN neuron's correlation with another of the same center type and another of the opposite center type, each at the given retinotopic separation. The illustrated shape of C^{ORI} means that activity patterns tend, stochastically, to resemble that shown at upper left: when ON neurons are firing, they tend to be most coactive with other ON neurons at small retinotopic distances, but with OFF neurons at larger distances. If LGN activities show such a correlation structure, then simple cells arise naturally through a Hebbian learning rule. **B**. Sample LGN input patterns at a random time step, instantiating a correlation structure like C^{ORI}. The activities of individual LGN cells (in arbitrary units, see Methods) are indicated by the scale bar.

Figure 3

Development and Mature Function of the Cortical Circuit. **A**. Geniculocortical weights for each of the ten cells in the column, at two different times, are shown by the circular afferent receptive fields. Grayscale denotes the *difference* between ON and OFF weights from a given retinotopic position to the given cell, normalized separately in each panel to cover the range between maximum and minimum ON/OFF difference across all cells in the panel. ON-dominated and OFF-dominated positions are white and black, respectively, while positions with equal ON and OFF weights are gray. Intracortical weights to and from a representative cell (#1) are represented by circles and squares, respectively, next to geniculocortical weights. Diameter/length of circles/squares corresponds to size of weight (normalized to maximum for that weight type, see Methods). Top six cells are excitatory ("E"), bottom four are inhibitory ("I"). **Left**: Initial condition. ON and OFF weights are roughly equal and of moderate size. Intracortical connections are all-to-all (excepting self-connections) with weights of a given type of roughly equal size between all cell pairs. **Middle**: Final condition. The 10 cells in the column have now all developed simple-cell geniculocortical structure with strong ON/OFF

segregation (gray regions correspond to very small or zero weights). The intracortical weights are very largely, though not completely, phase-specific: excitatory (inhibitory) projections are made to cells of similar (opposite) absolute spatial phase. We number cells in left-to-right and then top-to-bottom order, *e.g.* cell 2 is at upper right and cell 9 at lower left. **Right**: Top: the full 10×10 matrix of intracortical weights in the final condition is shown, normalized to the maximum allowed for the given type of weight, on a scale from 0 (white) to 1 (maximum allowed weight, black). Projecting cell is shown on horizontal axis, receiving cell on vertical axis, with numbering of cells as in middle panel. Connections show almost perfect phase specificity. Bottom: Excitatory neurons within the developed column display contrast-invariant orientation tuning. Response of cell 1 to gratings shown at 10%, 20%, 40%, and 80% contrast is indicated by increasing gray-level. Numbers at top right in each case indicate mean and standard deviation of best Gaussian fit to each response distribution, showing invariance of tuning widths with contrast. **B**. As in **A**, except that the position of each receptive field is shifted from the site's average location by a random amount (representing retinotopic scatter of the RF centers, Hubel and Wiesel (1974), Albus (1975)). The circles in the left and middle plots indicate the RF position of cell #6. In the final condition, excitatory (inhibitory) connections are again made between cells of similar (opposite) absolute spatial phase.

Figure 4

Robustness of the model results. Total intracortical weight connecting two cells vs. their difference in preferred orientation (ΔOri) and in absolute phase ($\Delta phase$), shown across all connected pairs of cells in 64 simulations without retinotopic scatter in A, and in 64 simulations with random retinotopic scatter in B. Left shows all excitatory weights and right shows all inhibitory weights. Bins are colored white when there were no weights in the given

bin. A cell's preferred orientation and phase were taken to be the orientation and phase corresponding to the peak of the Fourier transform of the cell's LGN receptive field (defined as the difference between ON-center and OFF-center strength from each location); therefore when orientations are significantly different, phase differences become less meaningful.

Figure 5

Why dominant opponent inhibition yields magnitude-invariant form recognition. A. The problem of magnitude-invariant form recognition. A simple cell RF is stimulated either by an optimally-oriented bar at low contrast (left; low contrast denoted by thin dashed lines), or an orthogonally oriented bar at high contrast (right; high contrast denoted by thick dashed lines). The optimal bar weakly excites the ON-center LGN inputs to the cell; the orthogonal bar strongly drives a subset of these ON-center inputs, while pushing a subset of OFF-center inputs down to zero firing rates. Because this suppression is limited at zero firing rate, the orthogonal bar evokes a net excitatory input. The two stimuli can thus be arranged to give the same pulse of LGN input to the cell. Yet the cell should respond to its optimal stimulus even at low contrast, and fail to respond to the orthogonal stimulus even at high contrast. We generalize this as follows: if a cell is wired to respond to input pattern **A** (the inputs evoked by the vertical bar), then some uncorrelated pattern **B** (the inputs evoked by the horizontal bar) will randomly share some inputs with **A**. Because firing rates are rectified at zero and background firing rates are low, **B** at high magnitude will evoke net excitation, strongly driving the inputs it shares with **A** while only weakly lowering the drive of inputs that it inappropriately stimulates. **B. How dominant opponent inhibition solves the problem of magnitude-invariant form recognition.** Let $\bar{\mathbf{A}}$ be the input pattern most anticorrelated with **A**; in the example of (A), this would be the inputs evoked by a dark bar over the ON subregion. Because input firing rates are non-negative, $\bar{\mathbf{A}}$ is

roughly orthogonal to **A**: inputs active in **A** are near 0 firing rate in $\bar{\mathbf{A}}$, and vice versa. We show them as orthogonal. We show two orthogonal dimensions, the **A** and $\bar{\mathbf{A}}$ directions, in the space of possible input patterns. Unit-length vectors in the **A** and $\bar{\mathbf{A}}$ directions are shown in the inset in the upper left quadrant. Dominant opponent inhibition means that a cell selective for **A** receives strong inhibition from cells driven by $\bar{\mathbf{A}}$, with weight $\mathbf{w} > 1$; the vector $-\mathbf{w}\bar{\mathbf{A}}$ is shown along the negative **barA** axis. Assume the cell responds if $\mathbf{A} - \mathbf{w}\bar{\mathbf{A}} > 0$; the vector $\mathbf{A} - \mathbf{w}\bar{\mathbf{A}}$ is shown in the lower right quadrant. The dashed line in the upper right quadrant is the line $\mathbf{A} - \mathbf{w}\bar{\mathbf{A}} = 0$ (this line is perpendicular to $\mathbf{A} - \mathbf{w}\bar{\mathbf{A}}$, and runs through the origin); input vectors that lie below this line have $\mathbf{A} - \mathbf{w}\bar{\mathbf{A}} > 0$ and thus evoke a response. Non-negative firing rates mean that input patterns can only be located in the upper-right quadrant. Thus, the set of possible input patterns that evoke a response is shown by the shaded region. The shape of this region shows that, for all magnitudes of stimuli, the cell only responds to input patterns that are close in shape (*i.e.* in vector direction) to the optimal pattern **A**. Stated more simply, because an uncorrelated input pattern **B** randomly shares some inputs with both **A** and $\bar{\mathbf{A}}$, it evokes net inhibition and the cell will not respond to it at any magnitude.

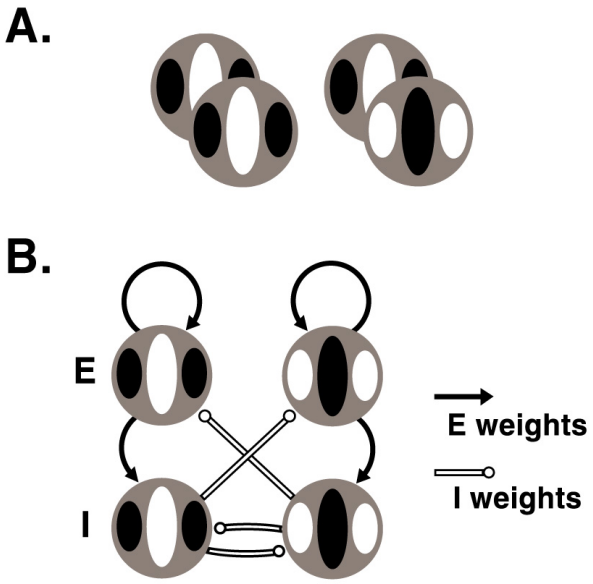

Figure 1:

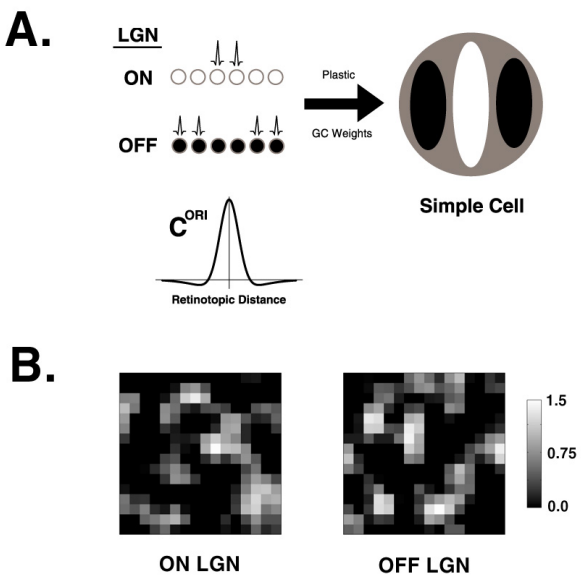
A.
LGN
ON
OFF
Plastic
GC Weights
Simple Cell
C^ORI
Retinotopic Distance
B.
1.5
0.75
0.0
ON LGN
OFF LGN

Figure 2:

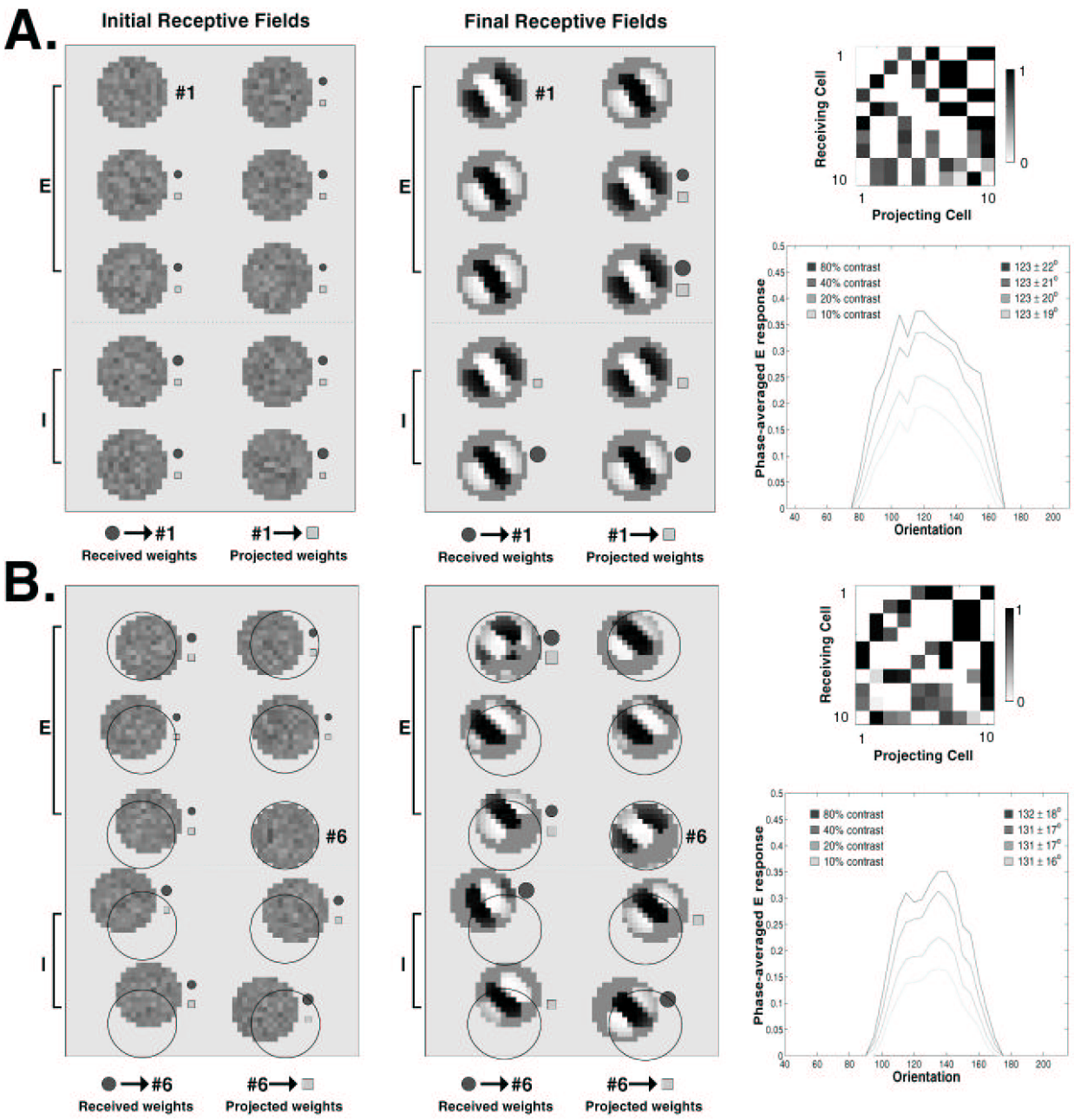
A.
Initial Receptive Fields
Final Receptive Fields
#1
E
I
Received weights
Projected weights
Receiving Cell
Projecting Cell
Phase-averaged E response
Orientation
80% contrast
40% contrast
20% contrast
10% contrast
123 ± 22°
123 ± 21°
123 ± 20°
123 ± 19°
B.
#6
132 ± 18°
131 ± 17°
131 ± 17°
131 ± 16°

Figure 3:

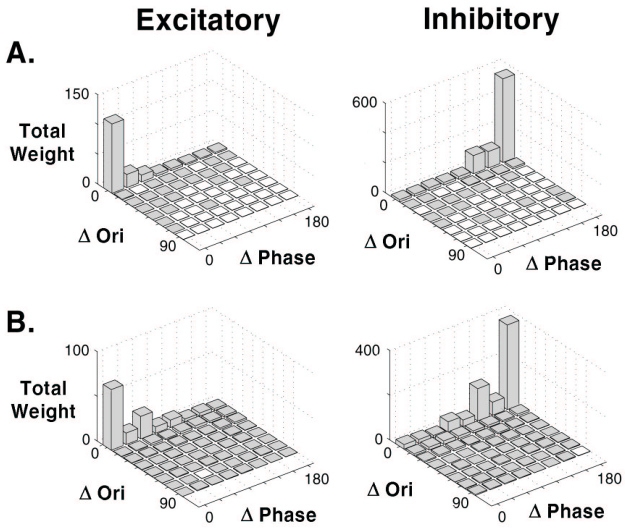
Excitatory
Inhibitory
A.
Total Weight
150
0
0
Δ Ori
90
0
180
Δ Phase
600
0
0
Δ Ori
90
0
180
Δ Phase
B.
Total Weight
100
0
0
Δ Ori
90
0
180
Δ Phase
400
0
0
Δ Ori
90
0
180
Δ Phase

Figure 4:

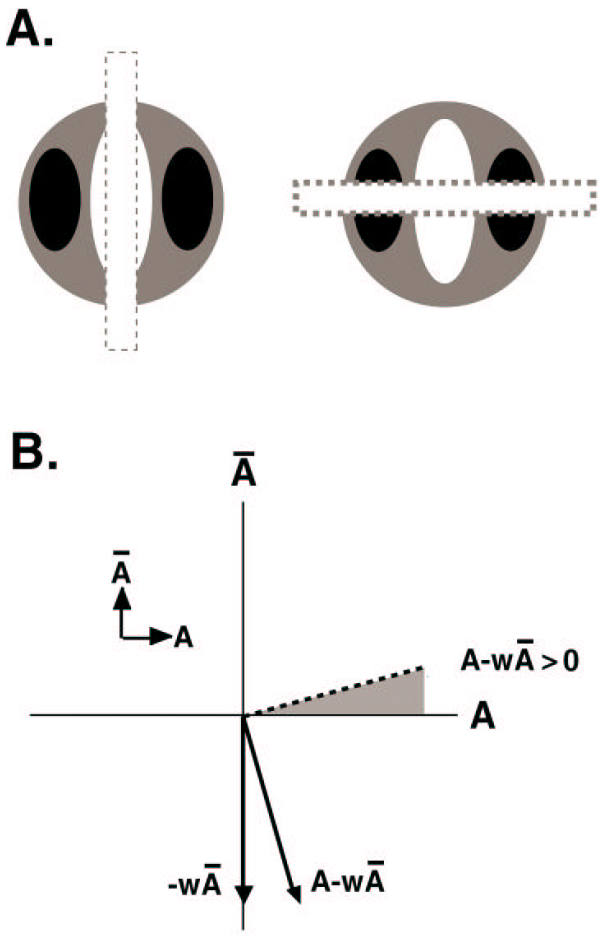

Figure 5: