CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Nature Precedings

On the Coding of Negative Quantities in Cortical Circuits

Dana H. Ballard
Department of Computer Science
University of Texas at Austin
Austin, TX
dana@cs.utexas.edu

Janneke Jehee
Department of Psychology
Vanderbuilt University
Nashville, TN
janneke.jehee@vanderbuilt.edu

June 1, 2008

Müller's *Law of specific nerve energies* introduced the idea that nerves transmit information about specific sensory features. This concept has been refined by the notion of 'labeled lines,' specific cells that capture sub-features of a sensory or motor stimulus, such as Hubel and Weisel's opponent color cells. Such features can be visualized as representing a signed quantity that has positive and negative components that are encoded with separate nerve cells. We show that there are two important consequences when learning receptive fields using signed codings in circuits. The first is that in feedback circuits even simple operations need to be distributed across multiple distinct pathways. The second consequence is that such pathways are necessarily dynamic. Synaptic weights change during learning and must break and grow new circuit connections because the weights need to change sign during receptive field formation.

INTRODUCTION

Animals owe their abilities in large part to their nerve cells that conduct electric spikes over large distances and allow sensory input to effect distal motor responses in a timely manner,[1] and, over evolutionary timescales, the coding of these spikes has become increasingly specialized. For example in visual input, the distribution of light on the mammalian retina is coded by individual cells that are each responsible for the sensation in a small retinal location, which in turn codes the light coming from a small part of distal space. Müller described this general coding strategy as the *Law of specific nerve energies*.[20,23] However the coding strategy is even more refined: In both the retina and the lateral geniculate nucleus (LGN) part of the image code consists of what are termed ON and OFF cells.[38] The former are responsive to a small spot of light seated in a darker surround and the latter are responsive to a dark spot in a lighter surround. Hubel and Weisel termed this coding strategy *labeled lines*.[38] Individual axons of the cells code the the cells' response, thus a distal cell that is recipient to this input has only the 'label' of the axonal 'line' to determinine the meaning of the associated input.

The labeled line coding strategy is ubiquitous in mammalian cortex. The cortex is characterized by hierarchies, where more and more complex features of the stimulus are coded from simpler ones.[10,35] Classic examples are the simple cells in primate striate cortex that respond to spatially localized oriented photometric 'edges' and 'bars' in the visual stimulus. Hubel and Weisel first suggested that these responses could be constructed from the simpler and anatomically precedent ON and OFF cell responses by comprising the explicit collections of them that reflected the response. Subsequent experimental evidence suggests that this original suggestion is correct.[2,13,29] When the connections between the LGN and striate cortex simple cells are tested, the spatial disposition of the respective ON and OFF inputs to the simple cell conform to the simple cell's receptive field. Furthermore feedback connections also exhibit similar regular structure.[21] Sillito has shown that the feedback connections of cortical simple cells are inhibitory when the 'ON' field of a simple cell feedback connects to its corresponding input ON LGN cell, and excitatory when it connects to the corresponding OFF LGN cell.[36]

The labeled line strategy can be viewed from a general perspective and that is that it is a special way of coding negative numbers. In signaling the value of a feature, the cortex uses two neurons, one for the positive quantity and one for the negative quantity. To distinguish this characterization, we use the phrase *signed labeled lines* to specifically note that the quantities are part of a two-cell representation for signed numbers. This strategy is used throughout visual cortex. Simple edge cells, direction-sensitive cells,[25] opponent color cells,[16] disparity cells,[15] motion cells,[30] as well as many more types, all use the opponent encoding strategy.

The signed labeled line convention introduces difficulties even in simple computational algorithms, because the use of signed quantities can interact with the labeled line representation in unexpected ways. Most neural network models finesse these complexities since they combine pairs into a single model 'neuron' that has a signed output as well as synapses that can change sign. But the crucial question remains of how the model changes when these issues are addressed.

This paper introduces a methodology for dealing with signed labeled lines that explicitly acknowledges the need to represent positive and negative quantities. This allows us to demonstrate the signed labeled lines' consequences by modeling the feedback circuit between the striate cortex and the LGN. The model feedback circuit learns synaptic weights by training itself on natural image patches, appropriately filtered to reflect processing in the retina.[14] The learning algorithm that we use is based on matching pursuit[18] which has a simple geometric interpretation. This kind of algorithm was originally demonstrated the formation of simple cell receptive fields[24] and has been subsequently extended to cortical hierarchies.[27] Its importance is that it does not have to specify the connections in detail but instead relies on a general abstract principle that the synapses should be chosen to minimize the spikes that are need to code any particular input pattern. The algorithm's feedback circuit encoding has sometimes been seen as at odds with alternative models that create receptive fields by competition, but in fact these two models have been shown to be equivalent if the competition is managed at the level of the competing neurons' input rather than by lateral inhibition.[12] This connection makes the matching pursuit algorithm featured here very general.

We show that translating the learning algorithm to this more realistic context of separate signed inputs and synapses places additional demands on the neural circuitry, but also allows simpler interpretations of experimental results. Our principal results are twofold. First, although experimental observations in both the feedforward and feedback pathways have been separately characterized as 'push-pull,' we show that they both are direct consequences of an algorithm for receptive field formation. For feedforward ciruitry, we show that the push-pull structure reported by[13, 19] is needed to correctly match the input to a neuron's receptive field. For feedback circuitry, complexities reported in this system[21] can be explained by separate feedback pathways necessitated by the signed labeled line representation. This is a much simpler explanation of observed structure than has so far been offered. The second result is that, in the feedback circuit learning process, the synaptic weights regularly change sign. The consequences for neurobiology are that synaptic contacts must be made or retracted. While the fact of synaptic growth and retraction is now well established from experiments,[7, 8, 34] we demonstrate how it needs to happen in the context of an algorithm for receptive field formation. The model allows us to study the synapse changes quantitatively and monitor their progress throughout the receptive field formation process.

RESULTS

The model has a complete set of two-way connections between the LGN and V1. As described in the Methods section, the model cells are signed labeled lines in that LGN OFF cells respond only to positive local contrast and OFF cells respond only negative local contrast. The connections are initially set to random values but are learned during the course of being exposed to 10,000 - 20,000 image patches. Figure 1 shows the results of the receptive fields of the V1 cells that are learned.

To characterize these receptive fields, feedforward connection weights from ON-center type and OFF-center type LGN cells coding for the same spatial location are summed for each of the model's 128 V1 cells. These summed weights are shown in Fig. 1B. After training, the receptive fields show orientation tuning as found for simple cells in V1.

The model does retain the all-important feature of separate ON and OFF cells and, as a consequence, important structure emerges. The feedforward connections to simple cells respect the simple cells' receptive field[2] and the feedback connections from a simple cell target the appropriate LGN cells.[21] Both of these properties are observed as a result of the learning process in our model. The lower right portion of Figure 1 shows the detailed connectivity between 16 LGN cells and one simple cell after training. Here the color blue codes the synaptic strengths between OFF cells and red is used to code for synaptic strengths between the ON cells and the simple cell. What the figure shows is that for a representative learned receptive field, all the LGN cells that connect to it from an 8×8 array of OFF cells and an 8×8 array of ON cells connect to the appropriate part of the V1 cell's receptive field with the appropriate synaptic strength.

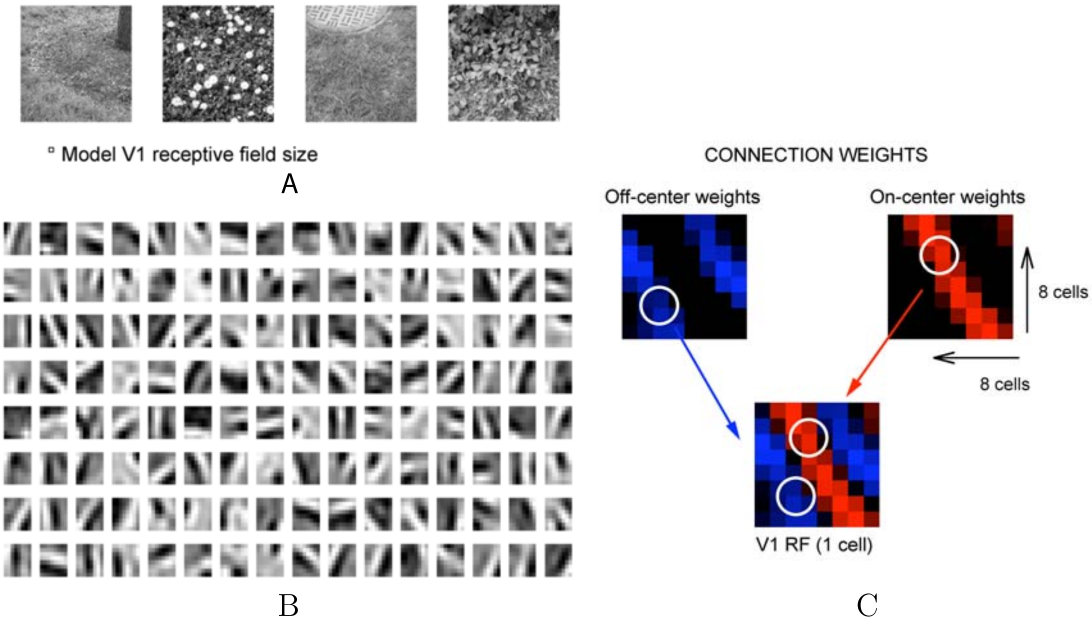

Figure 1: Learning receptive fields with signed labeled lines. A) Subset of natural images used for training. The small square immediate below denotes model V1 receptive field size. B) V1 receptive fields after training where ON and OFF responses are combined to produce a gray scale image. Black depicts off-regions in the model V1 receptive field, white depicts on-regions. C) A detail from the feed forward connections in the model making the the connections of different sign explcit. Blue denotes OFF center connections and red denotes ON center connections.

One important difference between the model and the cortical circuitry is that the model agglomerates what are known to be many intermediate connections. Thus the LGN input to V1 terminates in layer IV and the feedback connections to the LGN originate in layer V and VI. However this important distinction is glossed over in the model which just has its LGN cells reciprocally connected to V1 cells. Furthermore our model uses cells that can have both excitatory and inhibitory synapses, even though this is not possible biologically. The understanding is that to produce inhibition, there must be an intermediate stage where the excitatory connection excites an inhibitory cell and vice versa. Rather than complicate the circuit diagrams, we allow model cells to have both kinds of connections.

Signed Labeled Lines and Projections In Fig. 1 all the feed forward connections from the LGN to the model V1 cell are trying to make that cell produce a spike, that is they are all excitatory connections. What about inhibitory connections? Reid and Alonso[29] showed that ON and OFF cells that did not connect to the appropriate parts of the V1 receptive field did not make excitatory connections, but there remains the possibility that they may make, by some route, inhibitory connections. Our model suggests that indeed this should be the case, and why by using the notation for signed labeled lines developed in the Methods section.

A basic step in the model is to compute the projection

$$\sum_{i=1}^{N} x_i w_i.$$

In terms of our new notation this can be rewritten as

$$\sum_{i=1}^{N} (x_i^+ w_i^+ + x_i^+ w_i^- + x_i^- w_i^+ + x_i^- w_i^-)$$

but since all the inputs are treated identically, lets just concentrate on one such input and drop its subscript, so that the focus is on

$$x^+ w^+ + x^+ w^- + x^- w^+ + x^- w^-.$$

where in this case w^+ is an excitatory synapse and w^- is an inhibitory synapse. Taken at face value, this implies that there are four possible synapses that could be constructed to represent all the different possibilities for a term in the original dot product as shown in Fig 2. However when the receptive field is formed, only one of $\{x^+ w^+, x^+ w^-\}$ can be non zero, and the desired term in the dot product is positive. Lets assume that the positive term that maximizes the projection is $x^+ w^+$. Then for the projections to be calculated correctly, there needs to be a subtraction for the incorrect input x^-. Thus the synaptic connection $x^- w^-$needs to be included where $|w^-| = |w^+|$. If it is not included, then inputs that should be discounted will not be, and as a consequence, those inputs will be recorded as better matches to a neuron's receptive field than is in fact the case. These pathways are denoted with red arrows in Figure 2C. As a side note, this inhibition can be handled in at least two ways. Either 1) the four connections can be present at a single cell or 2) two cells can be used one collecting x^+w^+ and the other collecting x^-w^- followed by each cell laterally inhibiting the other. Lateral connections between simple cells are known to exist but the specificity implied by the need to represent the dot product correctly has not been established. Nonetheless a prediction of the signed labeled line model is that this specificity has to appear in some form of which the two possibilities just discussed are the prime candidates and there is evidence for both.[13,19,37]

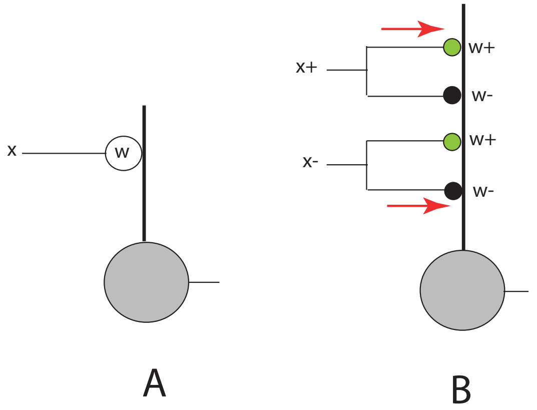

Figure 2: The feed forward pathway connections. The dot product computation illustrates the difference between conventional neural models and signed labeled lines. In the diagram thick lines denote dendrites and thin lines denote axons. *Green* circles = excitatory synapses. *Black* circles = inhibitory syanpses. A) If synapses and inputs could change sign then they could be handled simply with a single contact. B) In the actual case there are four possibilities, each of which needs a separate synapse. Only one of $w+$ and $w-$ can be non-zero at any one time and the same holds for x^+ and $x-$. Complementary pairs are required be non-zero to faithfully represent a dot product as shown by red arrows for the case of x^+w^+ and x^-w^-.

Signed Labeled Lines in Feedback Circuits The algorithm elaborated upon in the Methods section represents input by rapidly and sequentially selecting a handful of neurons to represent it. The algorithm is conceptually simple: one of the neurons that best matches the input is selected first, then that neuron's contribution is subtracted from the input via a feedback signal with the result that the remainder is in the form of new input and the process is repeated. However handling negative feedback in the signed labeled line system is far from straightforward and must be handled on a case by case basis. As will be demonstrated, net result is that the different cases need to be realized in separate circuitry. To illustrate the signed labeled line solution, consider the central calculation of the matching pursuit circuit described graphically in Figure 8. In the feed forward pathway the projection of the input onto the largest vector must be calculated. The result is given by $\mathbf{x} \cdot \mathbf{w_1}$ in standard notation and we have termed this quantity β. The feedback is given by the difference between the input vector $\mathbf{x}$ and its projection β into the closest vector described its synapses. Where $\mathbf{w_1}$ is the closest such vector, this difference is given by:

$$\mathbf{x} - \beta \mathbf{w_1}$$

Note that the need to deal with subtraction is a central requisite of this algorithm but not of course specialized to it. Any algorithm that required subtraction will have this issue.

Since all the components of the vector are treated identically, for simplicity of notation, again we will focus on just one vector component. Thus in the subsequent calculations all the variables are scalars. The difference between the input and vector projection for a single component can be indicated by $x - \beta w$ in standard notation. In signed labeled line notation we have

$$\begin{pmatrix} x^+ \\ x^- \end{pmatrix} - \beta \begin{pmatrix} w^+ \\ w^- \end{pmatrix}$$

where of course only one of x^+ and x^- can be non zero at any one time. Similarly only one of w^+ and w^- can be simultaneously zero. We illustrate the circuitry for x^+ nonzero. The other case is handled symmetrically. For each case we indicate the resultant circuit pathway with colored arrows as shown in Figure 3.

Case 1: $x^+ > \beta w^+$

$$\begin{pmatrix} x^+ \\ 0 \end{pmatrix} - \beta \begin{pmatrix} w^+ \\ 0 \end{pmatrix} = \begin{pmatrix} x^+ - \beta w^+ \\ 0 \end{pmatrix}$$

This is a simple case. The feedback pathway is inhibitory and has value w^+.

Case 2: $x^+ < \beta w^+$

$$\begin{pmatrix} x^+ \\ 0 \end{pmatrix} - \beta \begin{pmatrix} w^+ \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \beta w^+ - x^+ \end{pmatrix}$$

This case is a little tricky but important. The result uses $-x^+$. To realize this, x^+ has to be fed into the negative side, i.e., the opponent neuron, with an inhibitory connection, and the feedback to that neuron has to be positive or excitatory. As shown by,[14] this component of the circuit can be a form of rebound that introduces a temporal transient when the inputs are suddenly disturbed.

Case 3: $w^- > 0$

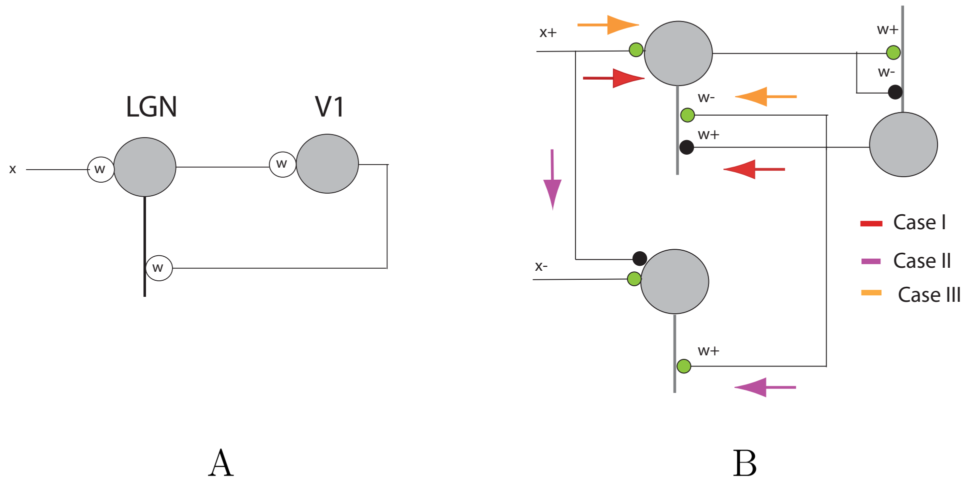

Figure 3: The feedback pathway connections. A) An inhibitory feedback circuit feedback circuit is simple to describe with signed synapses. However special care must be taken when using signed labeled lines. Analyzing the feedback to a single cell requires treating x^+ separately from x^-, but here only the three cases for x^+ are analyzed (and depicted in red) as the cases for x^- are symmetric. B) Case I: In the simplest to understand case $x^+ > \beta w^+$, the feed forward circuit computes the projection β and the feedback component is inhibitory. Case II: When $x^+ < \beta w^+$, things are more complicated as the feedback must *excite* the complementary LGN cell. Case III: When $w^- > 0$ the feedback is excitatory also but to the x^+ cell.

$$\begin{pmatrix} x^+ \\ 0 \end{pmatrix} - \beta \begin{pmatrix} 0 \\ w^- \end{pmatrix} = \begin{pmatrix} x^+ + \beta w^- \\ 0 \end{pmatrix}$$

This is another simple case. The feedback pathway is excitatory and has value w^+. By considering x^-, the need for three more pathways can be demonstrated for a total of six overall.

With six parallel feedback pathways a concern is whether they would interfere. A case by case analysis conforms that the circuit will function as desired. Lets examine the $x+$ three cases. Case I does not interfere with Case II because when the values are appropriate for Case I, the circuitry on the complementary side is held off. Similarly when Case II is appropriate, the circuitry for Case I is held off by virtue of the relative values. As for Case III, when the synapse $w^- > 0$, its complement w^+ is 0 so none of the circuitry is activated. These relationships might be more complicated if the circuitry had to operate in parallel with multiple, simultaneous feedback pathways. However a fundamental property of the algorithm is that only one coding (V1) neuron is analyzed per iteration. Owing to this property, the cases hold for each of the model LGN neurons.

The analysis has revealed six separate cases but one can wonder whether they are all used by the algorithm. In other words, is image data such that some of the cases do not occur? The simulation conforms that all six cases are used. Figure 4 shows this result. Each time a V1 cell is selected, it must send feedback to each of the 8 × 8 × 2 LGN cells that it is connected to. For each of those cells, only one of the six cases will come up. For this reason we can create a color coded image with the rule that, for each V1 coding neuron, the last time it was selected, for each of its feedback targeted LGN neurons, we can color code the route that the feedback took. To unpack this explanation a bit more, imagine that each of the positions in Figure 4 represents all six possible pathways to the LGN at that location. The color denotes, for a particular feedback moment in time, which of the six pathways was actually used. The colors in part A of Figure 4 reveal that typically all six cases are present. Furthermore they are used extensively. Figure 4B shows a histogram of the routes over a large sample of cells.

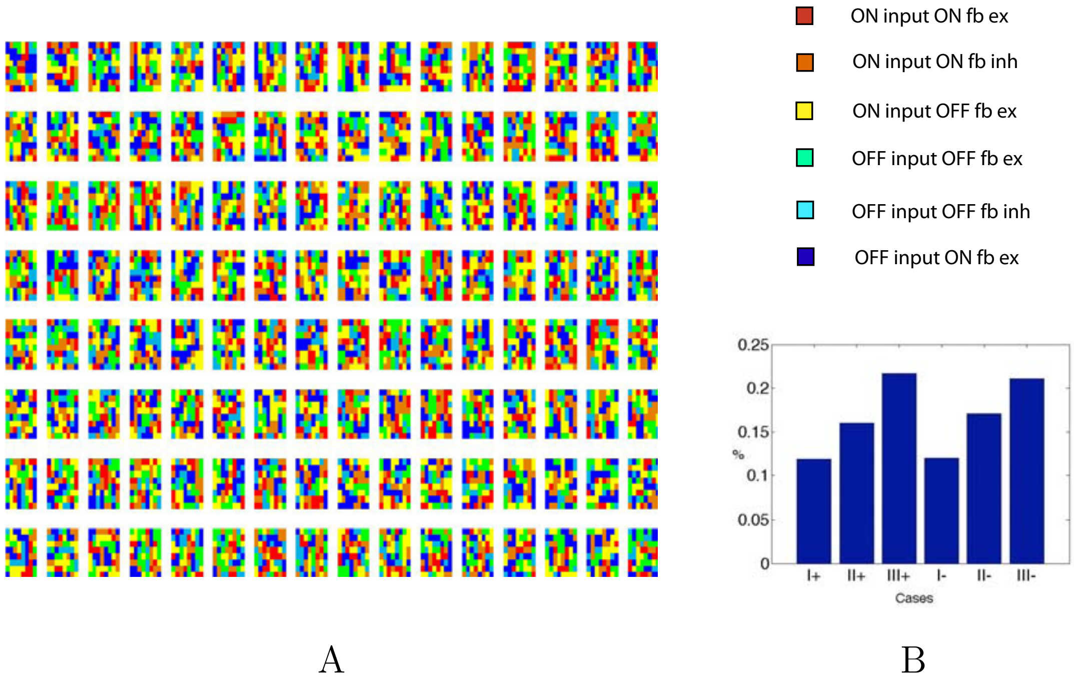

Figure 4: Tracking instantaneous feedback routes. Colors denote the six different routes that feedback can travel. A) For each model neuron, the last time it was chosen its feedback pathway for each of its synapses is labeled with a color denoting the route to each of the LGN neurons selected. The colors show that all six cases are realized. B) A histogram of the frequency of usage of the different cases.

Dynamic Synapses Given that we can explicitly represent signs of synapses, one very important consideration is; Do the connections in the circuitry need to change from one polarity to another? Recent research on the formation of synapses is showing that their formation and maintenance is very dynamic.[7,8,34] The simulations show that this is indeed the case; connections can be required to change from inhibitory to excitatory and vice versa throughout the learning process . In the model, all possible connections are present initially and just their strength is modulated by the algorithm. Perhaps quite naturally, in the course of learning their final values, the synaptic weights change sign fairly often.

Figure 5 shows this by testing the polarity of the weights every 3,000 image samples. As is evident, a large fraction of the synapses change their values. During the first 3,000 iterations about 4,000 of the total of 8,192 feedforward synapses change their values. If they are not needed they drift towards zero, but if they are needed an excitatory contact may have to be replaced by an inhibitory contact or vice versa. The figure shows the change from excitatory to inhibitory as black and the opposite change as white. Most of the changes are in the early stages, but the synapses can change even near the end of the learning process. The model is noncommittal as to how synapse changes are accomplished. The synapses need to change throughout the learning process, but the number decreases to less than .05% per learning rule update (An update refers to the selection of a neuron in the matching pursuit process - See Methods). However at the beginning the rate of sign changes may seem low at 5%, but remember that this is for each neuron that is selected, so in fact the cortical connection process needs to be very dynamic. What perhaps might have been expected, but nonetheless is very interesting to observe, is that the progress of receptive field formation is highly correlated (r=0.97) with the number of polarity changes, as shown in Figure 5F. This hints that the rate of polarity change could be a highly informative developmental measure.

Figure 6 summarizes the average rate of polarity change per iteration for all the synapses for a single V1 cell. That is each time any neuron is updated the number of sign changes in its synapses are recorded. Thus the figure reflects the average behavior of all the neurons in the model. However the model allows us to be much more specific about these changes. To demonstrate this capability, we track the behavior of an individual neuron's synapse as is done for model neuron #54 (out of 128) in Figure 6 which shows the course of each of its 64 synapses. The x-axis records the updates, that is each time that particular neuron was selected for modification (In the course of the learning algorithm there were intervening periods where other neurons were chosen). The simulation data for model neuron #54 shows that for the first 100 updates, 19 of 64 synapses changed from one polarity to another. For example synapse location (4,3) started out as excitatory (+1), switched to inhibitory (-1) around update 50 and then switched back to excitatory and finished as an inhibitory connection. By comparison, synapse (4,4) was always inhibitory and synapse (1,6) was always excitatory.

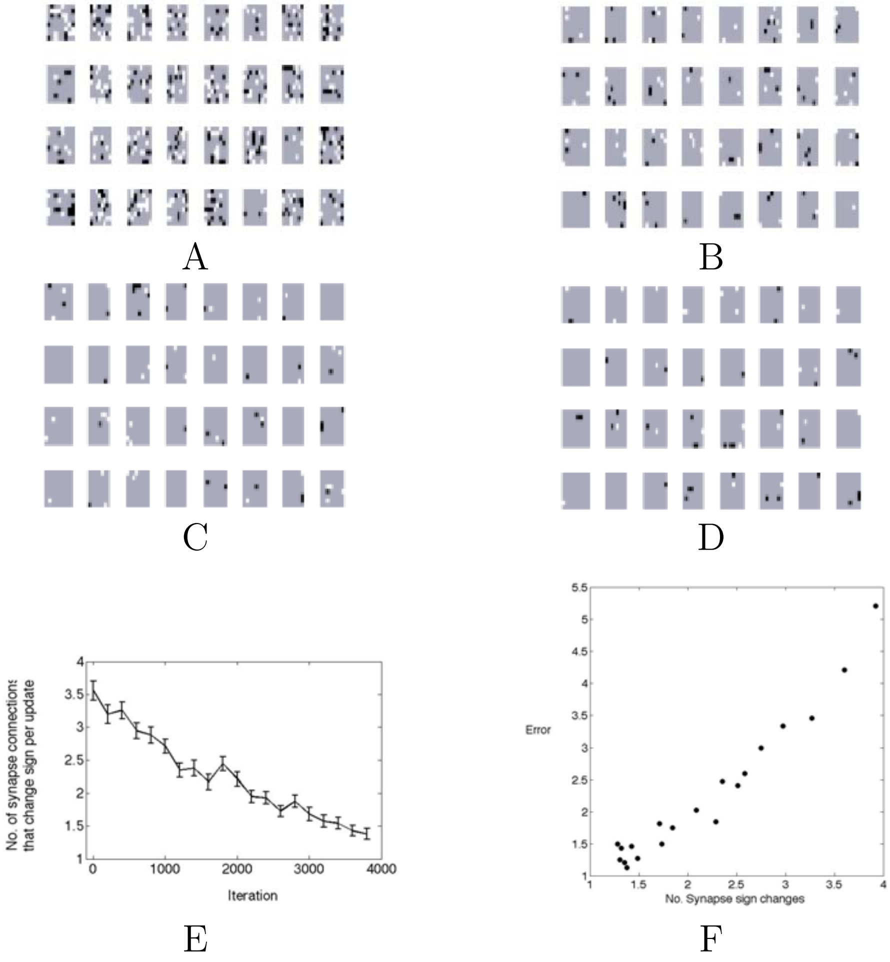

Figure 5: The changes in polarity of receptive fields during learning. A change from positive to negative is denoted by black and a change from negative to positive is denoted by white. A) After the first 3000 image patches B) After the second 3000 image patches C) After the third 3000 image patches D) After the forth 3000 image patches. E) The points plotted show the average number of synapses that had to change from a base of 256. Fifty samples are used in computing the standard error bars. Thus initially, on every learning update, about 5% of the synapses need to change signs. At the end of learning this number is down to less than .5%. F) The change in synapse polarity is tightly correlated with the residual error in fitting receptive fields (r = 0.97), suggesting that the changes in polarity can be used to track the progress of receptive field formation.

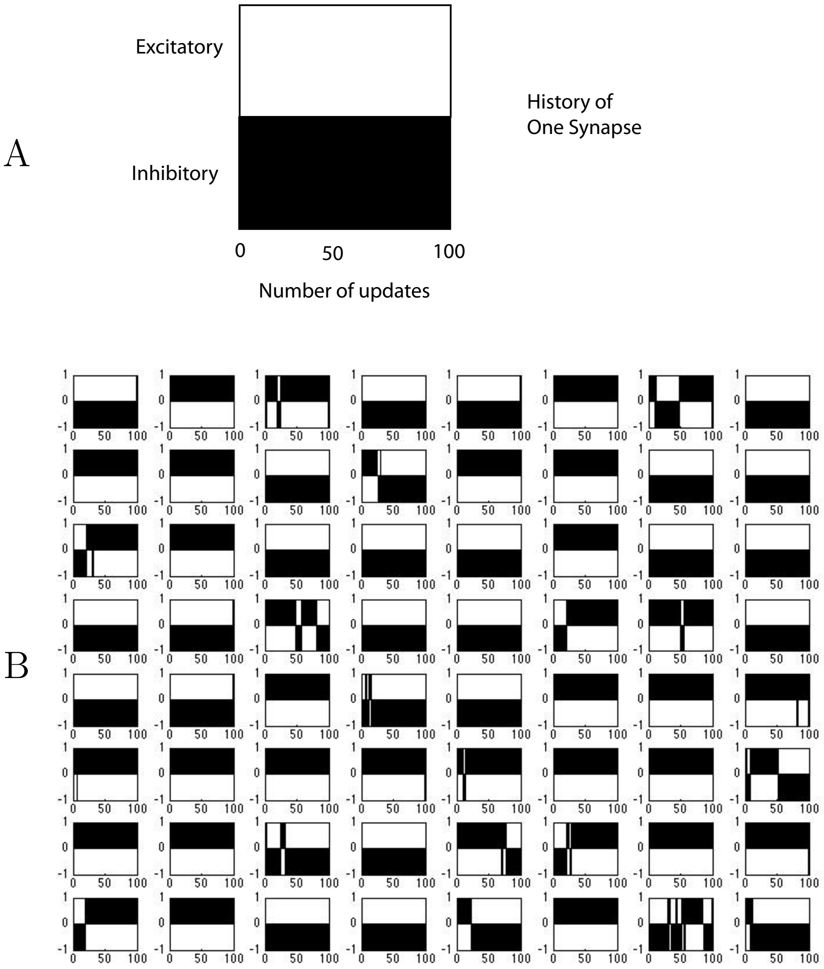

Figure 6: Tracking the polarity of connections to a single cell. A) The black portion of the legend denotes the polarity of a synapse (in this case inhibitory) as a function of the number of updates. B) The state of all 128 possible synapses for a particular model simple cell as a function of the times that it was chosen to represent an input image stimulus. The figure represents two possible synapses at each location. Most locations can be represented by a single synapse that does not change sign, but at 19 positions the synapses need to change sign during the computation, some several times such as synapse position $(8, 7)$.

DISCUSSION

Most neural network simulations ignore the detailed constraints of real neurons, blithely assuming that synapses can change signs and that huge precision is available in the intracellular signaling. The assumption is that the mathematics is important and the implementational issues are just unimportant details. Our simulations support this by showing that when these details are taken into account, the results that have been obtained with the more abstract models do indeed extend to the more detailed setting. However the labeled line model provides an important new vista into neural coding of dynamical circuits in at least several aspects.

Multiple, separate feedback pathways Surprisingly, from the standpoint of the model its feedback travels along different pathways depending on whether the it is negative or positive. This is at least testable and may have already been tested. Sillito et al[21] has observed that cortical feedback to the LGN is phase reversed, meaning that if the cortical simple cell connects back to an LGN cell of the opposite polarity as measured with respect to the cortical cell's receptive field, then that connection is excitatory. They recognize that this is a push-pull circuit, and speculate on its function as "gain control and linearity in the transfer of input to the cortex," but from our perspective a potential function is much simpler. The phase- reversed connections occur as Case II of our signed labeled lines feedback, and thus are a direct consequence of an algorithm which is trying to represent stimuli in an economical way and compute synapse strengths via negative feedback.

Excitatory and inhibitory synapses are exchanged in receptive field formation Learning in the labeled line model can require that an excitatory synapse be replaced by an inhibitory one and vice versa. This means that these synapses must be coupled somehow, so that the state of one can be available in some form to its complement. The simulations herein do not address the mechanism for accomplishing this but it needs to be done. This observation is not as evident from signed representations. Furthermore the number of synapses that have to change sign form iteration to iteration is significant, being about 3 %. and, if there were a way of measuring synaptic dynamics *en mass*, this could be tested.

One issue that is not simple to explain is that the synapses can be set with so few updates. After about 400 updates per model neuron the synapses have converged to their final values. Given that an update in our simulation might only take 20 $\sim$ 100 ms, it is hard to explain why the biological process seems to take much longer. One way this could arise is if there were overhead in setting up the synapses in the first place; our model does not represent this difficulty. Another slowdown factor might be that the amount a synapse can change per update is much less than assumed by the model. In any case the model provides the beginning of a processes of simulating alternate hypotheses.

Signed labeled lines and squaring Some abstract models of motion detection require a squaring function to overcome the fact that while the signed signal may be uncorrelated, its absolute value is usefully correlated.[33] However, just how does neurobiology come up with such a function? In the labeled line representation, this is much less of a problem than in signed representations as the signal is easily rectified by treating the 'negative' part of the signal as positive.

Summary New techniques that allow the elucidation of the details of cortical circuitry are showing that the cortical matrix of cells is very complex,[17,22,39] so to decipher it, its is likely that all useful constraints will need to be brought to bear. We show here that the interaction of a standard algorithm with the basic cortical coding of signed information can explain experimental observations of push-pull circuitry in both feed forward and feedback pathways. as to the One important thing to keep in mind is that although the model is much more detailed than the majority of neural models that used signed representations for synapse and neuronal outputs, it is still very abstract in that it ignores many of the still more detailed aspects of cortical architecture.[28,32] This architecture is obviously used for many functions in the course of implementing complex behaviors and those functions must be represented in additional circuitry to that assumed by our model. Furthermore it is well known that the feedback loop from striate cortex to LGN is complicated by many intermediate connections. For example the input to striate cortex terminates in layer IV whereas the output to the LGN originates from layers V and VI. In our model this complexity is summarized in single model neurons that receive both input and provide output. Along these lines there is another area in the simulation would need to be refined, and that is the fact that in the cortex the number of excitatory synapses outnumbers the number of inhibitory synapses. One estimate[32] is that the ratio of excitatory synapses to inhibitory synapses is on the order of 84:16. Since the ratio in the model is very close to 1:1, this means that there must be a pooling of inhibition where by multiple network inhibitory connections are handled by registering them as excitatory on an intermediate cell that then has a single inhibitory connection of the net value on the original destination cell.

METHODS

The model consists of two layers shown by Figure 7. The first layer, which would correspond to the lateral geniculate nucleus, consists of on-center type and off-center type units. Similar to geniculate cells, on-center type units code for brighter stimulus regions and off-center type units code for darker regions. We assume that either the on-center unit or its off-center counterpart coding for the same spatial location is active at any given time step in the model. The model's next higher level, which corresponds to an orientation column in primary visual cortex, receives input from model LGN through feedforward connections. In each feedforward-feedback cycle of the model, the feedforward receptive field that best matches the input, or equivalently the most likely prediction, is selected with high probability. The selection is made on the basis of the projections of the image, seen as a vector onto the synaptic weights, also see as vectors. A specific projection between an input $\mathbf{I} = (x_1, \ldots, x_n)$ and a neuron with synapses $\mathbf{w} = (w_1, \ldots, w_n)$ can be expressed as

$$\sum_{i=1}^{N} x_i w_i$$

or more compactly as the dot product $\mathbf{I} \cdot \mathbf{w}$. This expression has some recent experimental evidence.[3,4] Variations impose some non-linearity on the result.[26,31]

Once a neuron with weights $\mathbf{w}$ is chosen on the basis of its projection, the learning rule moves it a little closer to the input vector i.e.

$$\Delta\mathbf{w} = \alpha(\mathbf{I} - (\mathbf{I} \cdot \mathbf{w})\mathbf{w})$$

The repeated application of the learning rule produces the receptive fields shown in Fig. 1.

The selected neuron spikes and feeds its prediction back to model LGN. Weights of feedback connections follow the structure of feedforward connections, as has been found experimentally.[21,36] LGN neurons then compute the error between the higher-level prediction and the actual input, and the process is repeated in the next feedforward-feedback cycle. Thus, lower-level error detectors correct higher-level predictions, while higher-level responses update lower-level error signals in each feedforward-feedback pass of the model.

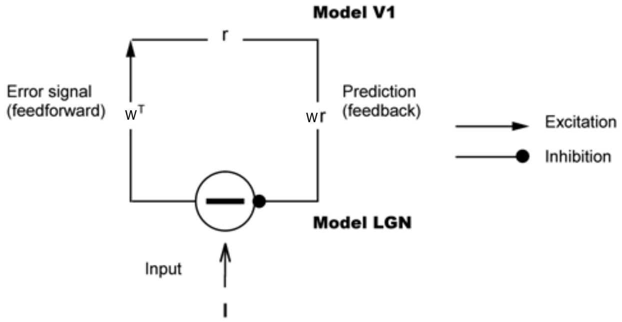

Figure 7: Hierarchical model for predictive coding. A) Higher-level units attempt to predict the responses of units in the next lower level via feedback connections, while lower-levels signal the difference between the prediction and the actual input. Feedforward connections encode the synaptic weights represented by a matrix W^T. Higher-level units maintain the current estimate of the input signal $\mathbf{r}$ and convey the top-down prediction $W\mathbf{r}$ to the lower level via feedback connections. Difference detectors compute the difference $\mathbf{I} - W\mathbf{r}$ between current activity $\mathbf{I}$ and the top-down prediction $W\mathbf{r}$.

Each cell in the model encodes scalar information using one spike only, where the time from spike arrival to a reference signal is the information carrier. We do not implement the reference signal explicitly but argue that the model could easily be amplified to take this into account. Model activity is updated every 20 milliseconds. As the model does not incorporate neural structures earlier than the LGN, we add 30 milliseconds to the input to account for the delays before the LGN.[6] Connection weights of the model are adapted to the input by minimizing the description length of the joint distribution of inputs and neural responses.[14] This not only improves the sparseness of the neural code, but also tends to optimally capture input statistics. Thus, for any given input, the model converges to a set of connection weights that are optimal for predicting that input. The model is trained on image patches extracted from natural scenes, the motivation being that receptive field properties might be largely determined by the statistics of their natural input.[5, 9, 11, 27]

The feed forward connections from 64 cells in the LGN to a single cell in V1 are depicted in Figure 1. The learning algorithm connects a complete set of synapses to the V1 cell initially, that is 128 synapses altogether, half from the ON calls and half from the OFF cells. However after learning only the appropriate set of LGN cells have large weights as shown in the Figure. This replicates the experimental finding of Alonso and Reid.[29] They used antidromic simulation in paired recordings to confirm this connection arrangement. The experimental finding is very significant since it confirms the original suggestion by Hubel and Weisel that the connections could be formed in this manner. What our work shows is that a Hebbian learning rule based on sparse coding principles is able to produce this arrangement.

Signed Labeled Line Encoding Notation The response of an cell can be characterized mathematically in terms of a function of the inputs multiplied by synaptic 'weights,' numbers representing the strength of a synapse. Thus if the input to such a cell is represented by a vector $\mathbf{x}$ and the synapses as a vector $\mathbf{w}$, the the response β can be given by

$$\beta = f(\mathbf{x} \cdot \mathbf{w}) \tag{1}$$

where f is a function that captures any nonlinearities in the response and $\mathbf{w} \cdot \mathbf{x}$ is the projection of $\mathbf{x}$ onto $\mathbf{w}$ or equivalently, the dot product between $\mathbf{x}$ and $\mathbf{w}$ given by

$$\mathbf{w} \cdot \mathbf{x} = \sum_{i=1}^{N} x_i w_i$$

While the above expression models neuronal responses, and has experimental support for at least excitatory synapses,[3, 4] it is cast at a level of abstraction that avoids the crucial issue associated with labeled lines and that is the representation of positive and negative coefficients. Let us illustrate these issues

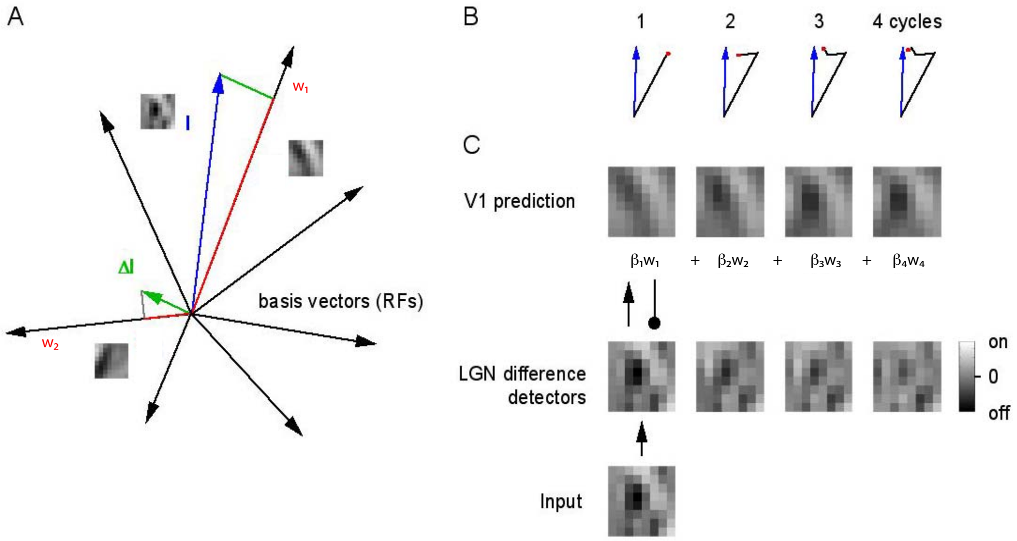

Figure 8: A geometrical explanation of the matching pursuit learning model. A) An 8×8 filtered image patch can be represented as a blue vector with 128 coordinate values (in our notation). The neuron whose receptive field is most like the patch, in this case $\mathbf{w_1}$, is chosen to represent the patch. Since there are also 128 synapse strengths, or weights, these can be represented as a vector also. The difference between them is termed the residual (green) and is sent back to the LGN as feedback and the process repeats. A very small number of repetitions produces an accurate representation. B) Four steps in the vector approximation. C) The evolution of the approximation in pictorial terms. The green vector is also the basis for the learning algorithm. After each vector is chosen, it is moved closer to the input by adding the residual into its synaptic weight vector. The weight vectors are normalized to unity, reflecting a constraint that limits the total strength of the synapses.

with the example of a single term in the expression in Equation 1 above. Suppose a neuron is receiving input at a single synapse that can be expressed mathematically as the product wx. Figute 2A shows how this multiplication would be implemented with signed quantities. Both the axonal input x and the synaptic strength w can be signed quantities, so a single 'synapse' suffices to represent the calculation. However in the more detailed model that respects the representation of positive and negative quantities by separate cells the calculation cannot be done so easily. Besides the separate inputs, a further complication (from the standpoint of mathematical operations) is that synapses cannot change sign. An inhibitory synapse cannot become excitatory and vice versa. Let us illustrate this complication in detail. Now x can be either positive or negative, denoted with $\{x^+, x^-\}$ as can w, denoted with $\{w^+, w^-\}$. Thus to compute the product, four connections are required, representing all the combinations of positive and negative signs. Figure 2B shows these possibilities. Note that the figure is still a level of abstraction above the biological implementation of this relationship since a given set of synapses from any one neuron can only be excitatory or inhibitory. Thus at least one additional cell is required to change the inhibition to a an excitation of an inhibitory cell. Note also that the ± notation is an *algebraic* device for keeping track of opponent quantities. For example w^+ denotes a the strength of a synapse. Whether or not it turns out to be excitatory or inhibitory depends on circuit and algorithm details.

When using signed labeled lines, the realization of elementary operations is not so straightforward and requires some care. To see this it helps to develop a notation for signed labeled line vectors. In standard vector notation, an example of a vector with two components is: $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. A simple example showing the subtraction of two vectors is shown as follows,

$$\begin{pmatrix} 1 \\ -2 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ -3 \end{pmatrix}$$

This is standard vector mathematics, but now lets introduce a convention that allows us to keep track of the fact that positive and negative components are represented by different cells. To express the subtraction in terms of labeled line notation, lets use a separate component for each of the positive and negative side, as illustrated in the dot product example. Thus

$$\mathbf{x} = \begin{pmatrix} x_1^+ \\ x_1^- \\ x_2^+ \\ x_2^- \end{pmatrix}$$

and the above example becomes

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 3 \end{pmatrix}$$

Note that each vector component must be either positive or negative so that in the labeled line notation, one of the two corresponding pairs is always zero. Furthermore note that in subtracting two vectors the result can be arbitrary in the sense that the resultant component that is non-zero depends on the signs and magnitudes of the vectors.

ACKNOWLEDGEMENTS

This research was supported by NIH grant R01-02983

References

[1] J. M. Allman. *Evolving Brains*. W. H. Freeman, 1998.

[2] J.-M. Alonso, W. M. Usrey, and R. C. Reid. Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *Journal of Neuroscience*, 2001.

[3] R. Araya, K. B. Eisenthal, and R. Yuste. Dendritic spines linearize the summation of excitatory potentials. *PNAS*, 103:18799–18804, 2006.

[4] R. Araya, J. Jiang, K. B. Eisenthal, and R. Yuste. The spine neck filters membrane potentials. *PNAS*, 103:17961–17966, 2006.

[5] J. J. Atick. Could information theory provide an ecological theory of sensory processing? *Network*, 1992.

[6] W. Bair, J. R. Cavanaugh, M. A. Smith, and J. A. Movshon. The timing of response onset and offset in macaque visual neurons. *Journal of Neuroscience*, 2002.

[7] Jennifer Bourne and Kristen M. Harris. Dothinspineslearntobemushroomspinesthatremember? *Current Opinion in Neurobiology*, 17:381–386, 2007.

[8] Jennifer N. Bourne and Kristen M. Harris. Balancing structure and function at hippocampal dendritic spines. *Annual Review of Neuroscience*, 31:47–67, 2008.

[9] Y. Dan, J. J. Atick, and R. C. Reid. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *Journal of Neuroscience*, 1996.

[10] Daniel J. Felleman and David C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47, 1991.

[11] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 1987.

[12] George F Harpur, Richard W Prager George F Harpur, and Richard W Prager. Development of low entropy coding in a recurrent network. *Network: Computation in Neural Systems*, 7:277–284, 1996.

[13] Judith A. Hirsch. Synaptic physiology and receptive field structure in the early visual pathway of the cat. *Cerebral Cortex*, 13:63–69, 2003.

[14] J Jehee, C Rothkopf, J Beck, and D Ballard. Learning receptive fields using predictive feedback. *Journal of Physiology-Paris*, 100:125–132, 2006.

[15] S. LeVay and T. Voight. Ocular dominance and disparity coding in cat visual cortex. *Visual Neuroscience*, 1:395–414, 1988.

[16] M. S. Livingstone and D. H. Hubel. Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, 4:309–356, 1984.

[17] L. Luo, E. M. Callaway, and K. Svoboda. Genetic dissection of neural circuits. *Neuron*, 57:634–660, 2008.

[18] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41:3397–3415, 1993.

[19] Luis M. Martinez, Qingbo Wang, R. Clay Reid, Cinthi Pillai, Jose-Manuel Alonso, Friedrich T. Sommer, and Jusith A. Hirsch. Receptive field structure varies with layer in the primary visual cortex. *Nature Neuroscience*, 8:372–379, 2005.

[20] J. Müller. *Zur vergleichenden Physiologie des Gesichtssinnes des Menschen und der Tiere*. Leipzig: C. Knobloch, 1826.

[21] Penelope Murphy, Simon C. Duckett, and Adam M. Sillito. Feedback connections to the lateral geniculate nucleus and cortical response properties. *Science*, 286:1552, 1999.

[22] J. J. Nassi, D. C. Lyon, and E.M. Callaway. The parvocellular lgn provides a robust disynaptic input to the visual motion area mt. *Neuron*, 50:319–327, 2006.

[23] Ulf Norrsell, Stanley Finger, and Clara Lajonchere. Cutaneous sensory spots and the "law of specific nerve energies": history and development of ideas. *Brain Research Bulletin*, 48:457–553, 1999.

[24] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996.

[25] G. A. Orban, H. Kennedy, and J. Bullier. Velocity sensitivity and direction selectivity of neurons in areas v1 and v2 of the monkey: influence of eccentricity. *Journal of Neuroscience*, 56:462–480, 1986.

[26] J. W. Pillow and P. Simoncelli E. Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. *Journal of Vision*, 6:414–428, 2006.

[27] Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience*, 2:79–87, 1999.

[28] I. Reichova and M. Sherman. Somatosensory corticothalamic projections: distinguishing drivers from modulators. *Journal of Neurophysiology*, 92:2185–2197, 2004.

[29] R. C. Reid and J.-M. Alonso. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 1995.

[30] H. R. Rodman and T. D. Albright. Coding of visual stimulus velocity in area mt of the macaque. *Vision Research*, 27:2035–2048, 1987.

[31] Tatyana O. Sharpee, Hiroki Sugihara, Andrei V. Kurgansky, Sergei P. Rebrik, Michael P. Stryker, and Kenneth D. Miller. Adaptive filtering enhances information transmission in visual cortex. *Nature*, 439:936–942, 2006.

[32] M.S Sherman and R. W. Guillery. *Synaptic Organization of the Brain.* Oxford University Press, 2003.

[33] Eero P. Simoncelli and David J. Heeger. A model of neuronal responses in visual area mt. *Vision Research*, 38:743–761, 1996.

[34] John Smythies. *The Dynamic Neuron.* MIT Press, 2002.

[35] L. G. Ungerleider and M. Miskin. Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, editors, *Analysis of Visual Behavior*, pages 549–586. MIT Press, 1982.

[36] W. Wang, H. E. Jones, I. M. Andolina, T. E. Salt, and A. M. Sillito. Functional alignment of feedback effects from visual cortex to thalamus. *Nature Neuroscience*, 2006.

[37] Xin Wang, Yichun Wei, Vishal Vaingankar, Qingbo Wang, Kilian Koepsell, Friedrich T. Sommer, and Judith A. Hirsch. Feedforward excitation and inhibition evoke dual modes of firing in the cat's visual thalamus during naturalistic viewing. *Neuron*, 55:465–478, 2007.

[38] T. N. Wiesel and D. H. Hubel. Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *Journal of neurophysiology*, 29:1115–1156, 1966.

[39] Y. Yoshimura, J. L. Dantzker, and E.M. Callaway. Excitatory cortical neurons form fine-scale functional networks. *Nature*, 433:868–873, 2005.