NEOCORTICAL LAYER 4 AS A PLURIPOTENT FUNCTION LINEARIZER

Oleg V. Favorov and Olcay Kursun

1Department of Biomedical Engineering
University of North Carolina School of Medicine
Chapel Hill, NC 27599-7575

and

2Department of Computer Engineering
Istanbul University
34320 Avcilar, Istanbul, Turkey

Running title: Layer 4 as a function linearizer

Corresponding author: Oleg V. Favorov, Ph.D.
Department of Biomedical Engineering
70 MacNider Building
CB#7575
University of North Carolina
Chapel Hill, NC 27599-7545
Email: favorov@bme.unc.edu
Telephone: (919) 966-9447
Fax: (919) 966-2963

Copyright © 2011 by the American Physiological Society.
ABSTRACT

A highly effective kernel-based strategy used in Machine Learning is to transform the input space into a new “feature” space where nonlinear problems become linear and more readily solvable with efficient linear techniques. We propose that a similar “problem-linearization” strategy is used by the neocortical input layer 4 to reduce the difficulty of learning nonlinear relations between the afferent inputs to a cortical column and its to-be-learned upper-layer outputs. The key to this strategy is the presence of broadly tuned feed-forward inhibition in layer 4: it turns local layer 4 domains into functional analogs of Radial Basis Function networks, which are known for their universal function approximation capabilities. Using a computational model of layer 4 with feed-forward inhibition and Hebbian afferent connections, self-organized on natural images to closely match structural and functional properties of layer 4 of the cat primary visual cortex, we show that such layer-4-like networks have a strong intrinsic tendency to perform input transforms that automatically linearize a broad repertoire of potential nonlinear functions over the afferent inputs. This capacity for pluripotent function linearization, which is highly robust to variations in network parameters, suggests that layer 4 might contribute importantly to sensory information processing as a pluripotent function linearizer: performing such a transform of afferent inputs to a cortical column that makes possible for neurons in the upper layers of the column to learn and perform their complex functions using primarily linear operations.

Keywords: feed-forward inhibition, orientation tuning, radial basis function, simple cell, V1
INTRODUCTION

A highly effective strategy used in the field of machine learning and pattern recognition for dealing with nonlinear problems is to transform the input space into a new higher-dimensional “feature” space, in which the problem becomes linear and thus more readily solvable with efficient linear techniques (Figure 1a). This kernel-based strategy (Schölkopf and Smola 2002) has given rise to such powerful methods as nonlinear Support Vector Machines, Kernel Principal Component Analysis, Kernel Independent Component Analysis, and others. In addition to the proven effectiveness of this “problem-linearization” strategy in machine pattern recognition, its relevance at the psychological level (Jäkel et al. 2009) strongly supports the proposals (Poggio 1990; Pouget and Sejnowski 1997; Maass et al. 2002; DiCarlo and Cox 2007) that such a strategy might also be used by the neocortex in the experience-driven development of its remarkable functional properties (Sur and Rubenstein 2005). Among specific applications, Poggio and Edelman (1990) used this strategy in their HyperBF model to recognize three-dimensional objects from any perspective based on a small sample of views. DiCarlo and Cox (2007) proposed that the same strategy, which they call an object-manifold untangling, is essential to the brain’s object recognition and categorization, demonstrating its effectiveness by accurately identifying objects from the firing rates of 100-300 neurons in the macaque inferior temporal cortex (Hung et al. 2005). Pouget and Sejnowski (1997) proposed that parietal cortical neurons reduce the complexity of sensorimotor transformations by generating such a representation of their sensory inputs from which diverse motor commands can be computed by simple linear summations.

FIGURE 1 NEAR HERE
This paper explores a potential implementation of the problem-linearization strategy in each individual cortical area. As a part of its development, each cortical area learns to perform certain nonlinear integrating functions over its afferent inputs from the lower-level cortical areas and/or the thalamus, which in combination with feedback inputs from the higher-level cortical areas determine the cells’ output. This function-learning task might – given its distinctly nonlinear nature – take advantage of a problem-linearization strategy to enable it to develop more perceptually or behaviorally advanced functions. This consideration leads to a specific hypothesis, which is the subject of this paper, that some form of the problem-linearization strategy is implemented in layer 4 of each cortical area.

Layer 4 (L4) is the principal initial recipient of the afferent input to a cortical area. It converts that input into a new form and sends it to the upper layers (layers 2 and 3, or L2/3) of the same cortical area for further processing. The product of that L2/3 processing is then sent to L4 of the next cortical area, where the two-stage information processing operation is repeated, but on a higher level, building on the advances made by the preceding cortical area (Rockland and Pandya 1979; Felleman and Van Essen 1991). The functional role played by L4 in this processing is unclear. Among the proposed hypotheses are: redundancy reduction (Barlow 1989); extraction of low-order directional derivatives of the input (Adelson and Bergen 1991); input-to-output information maximization (Linsker 1993; Okajima 2001); preservation of spatial relationships in the input (Li and Attick 1994); efficient sparse coding (Olshausen and Field 1996; Bell and Sejnowski 1997; Rehn and Sommer 2007); temporal coherence maximization (Hurri and Hyvärinen 2003); and tuning to particular input patterns as a step to generalization (Poggio and Bizzi 2004).
The hypothesis explored in this paper is that L4 reduces the difficulty of learning nonlinear relations between the afferent inputs to a cortical column and its to-be-learned upper-layer outputs by making those relations linear in the L4 output to the upper layers. In support of this hypothesis, we show that a prominent feature of L4, namely its feed-forward inhibition, makes local L4 domains into functional analogs of Radial Basis Function (RBF) networks, which are well known for their versatile function-linearization capabilities. We next formulate a mathematically explicit L4 model with such feed-forward inhibition, self-organize its plastic connections on natural images, and thus produce a model that closely mirrors L4 of the cat primary visual cortex, V1, in its structural and functional properties. We show that this L4 model does indeed have prominent function-linearization capabilities, suggesting the same for the real L4.

IMPLEMENTATION OF THE PROBLEM-LINEARIZATION STRATEGY IN L4

Pluripotent problem-linearization strategy

The basic strategy is depicted schematically in Figure 1b for a prototypical cortical column. The essential problem faced by the column is that the functions to be computed by its L2/3 cells over the afferent inputs to the column (we will refer to them as the “target” L2/3 functions) are much more elaborate than the afferent inputs – as indicated in Figure 1b, the target L2/3 functions are likely to have little, if any, linear correlation with the afferent inputs from which they are to be computed. To reduce the difficulty of having to learn complex nonlinear relations between the afferent inputs and the target L2/3 outputs, the afferent inputs can be first transformed in L4 in
such a nonlinear manner that will substantially raise linear correlations between the outputs of
the L4 cells and the target L2/3 functions, thereby making such functions imminently learnable
with linear techniques, including those based on biological Hebbian learning.

In their transformation of the afferent inputs, the set of L4 cells in the prototypical
cortical column of Figure 1b will have to “linearize” target functions for the large number of
cells comprising L2/3 of the column. Furthermore, since the L2/3 target functions are not
specified a priori, but are developed by L2/3 cells gradually in a process of experience-driven
self-organization and without providing any significant direct feedback to L4, the L4 cells will
have to linearize the potential L2/3 target functions “blindly.” Under such conditions an ideal L4
transform would be one in which any arbitrary function defined over the afferent inputs turns
linear in the “feature” space defined by the outputs of the L4 cells. However, while such an
“omnipotent” transform is theoretically possible (e.g., using radial basis function networks; Park
and Sandberg 1991; Kůrková 2003), it cannot be done in L4 due to the limited number of the
available neurons. Instead, L4 can pursue a strategy of maximizing the “pluripotency” of its
transform. That is, the L4 transform can be optimized to make linear as broad a repertoire of
functions over the afferent inputs as possible. The L2/3 cells will then learn their target
functions from this repertoire.

Deco and Obradovic (1995) offer a particular example of such a pluripotent approach to
function approximation, which we use as a conceptual prototype for developing the L4 model.
The DHL (Decorrelated Hebbian Learning) function approximator of Deco and Obradovic has
two stages: the first stage (analogous to L4) maps the input space with a set of radial basis
functions (RBFs) and the second stage (analogous to L2/3) approximates the target function with
a weighted sum of the RBF outputs. In the first stage, a combination of Hebbian afferent input
plasticity and anti-Hebbian plasticity of lateral interactions is used to distribute RBFs efficiently within the region of the input space from which the input patterns were drawn. The positions of the RBFs in the input space are influenced only by the presented input patterns, and not by the to-be-approximated function(s). Once RBFs are settled into a stable pattern, their outputs can be used in the second stage to approximate any desired function of the input patterns. The pluripotency of the RBF transform depends on the number of RBFs used to map the input space: the larger the number of RBFs, the more accurate the function approximation (Park and Sandberg 1991).

**L4 neuron as an RBF unit**

Here we show that L4 neurons can be considered as RBF units with regard to how they represent the identity (but not the strength) of sensory stimuli. To separate the effects on a neuron of the stimulus identity vs. its strength, we first consider stimuli of the same strength. An L4 neuron receives its afferent input from a set of thalamic or lower-level cortical neurons, which together define the high-dimensional afferent space of that L4 neuron. Any stimulus is represented as a vector in this afferent space by the stimulus-evoked activities of the afferent neurons (Figure 2a). All the stimuli of the same strength (which we express by the length, or the Euclidean norm $l_2$, of the stimulus vector) will lie on a hypersphere of $l_2$ radius in the afferent space.

**FIGURE 2 NEAR HERE**

To define an RBF on such a hypersphere, we have to specify the location of the RBF center on the hypersphere and our measure of proximity of stimuli to the RBF center. Since the RBF center and all the stimuli under consideration lie on the hypersphere, we can express the proximity of a stimulus to the RBF center simply by the cosine of the angle $\alpha$ between the two
vectors (Figure 2a). We also can control the radial extent of the RBF by thresholding \(\cos(\alpha)\), as shown in Figure 2b. Thus, our RBF takes the form:

\[
\phi(\vec{a}) = \frac{1}{1-\theta} \left[ \cos(\alpha - \theta) \right]^+
\]

(1)

where \(\alpha\) is the angle between the stimulus and RBF center vectors; \(\theta\) is the threshold; and \([ \cdot ]^+\) indicates that if the quantity in the brackets is negative, the value is to be taken as zero. This function is scaled to have maximum \(\phi(\vec{a}) = 1\).

We can express \(\cos(\alpha)\) in terms of activities of the afferent neurons and the coordinates of the RBF center:

\[
\phi(\vec{a}) = \frac{1}{1-\theta} \left[ \sum_{i=1}^{n} w_i a_i - \theta \right]^+
\]

(2)

where \(\vec{a} = [a_1 \ldots a_n]\) is the stimulus vector; \(a_i\) is the activity of the \(i^{th}\) input neuron; \(w_i\) is the \(i^{th}\) coordinate of the RBF center; and \(n\) is the total number of the afferent neurons. Since in the cortex the afferent input comes from excitatory cells, the activity of which cannot be negative, we specify that \(w_i \geq 0\) and \(a_i \geq 0\). We also specify here, for mathematical convenience, that the stimulus and RBF center vectors have a unit length.

To generalize to stimuli of any strength, we are guided by the basic observation that stronger stimuli tend to evoke stronger responses in L4 neurons. Reflecting this tendency, when we are dealing with a stimulus vector \(\vec{a}\) of length \(l_2\), we can compute the RBF response to the normalized vector \(\vec{a} / l_2\) and then scale the RBF response by \(l_2\):

\[
\Phi(\vec{a}) = l_2 \cdot \phi(\vec{a} / l_2) = \frac{l_2}{1-\theta} \left[ \sum_{i=1}^{n} w_i \frac{a_i}{l_2} - \theta \right]^+ = \frac{1}{1-\theta} \left[ \sum_{i=1}^{n} w_i a_i / l_2 - \theta \cdot l_2 \right]^+
\]

(3)
Finally, expressing $l_2$ in terms of activities of the afferent neurons, we obtain our radial basis function:

$$
\Phi(\vec{a}) = \frac{1}{1 - \theta} \left[ \sum_{i=1}^{n} w_i a_i - \theta \sqrt{\sum_{i=1}^{n} a_i^2} \right]^+ \tag{4}
$$

This RBF has a shape of a circular cone in a high-dimensional afferent space, with the cone apex at the origin of the coordinate system defined by the afferent inputs, the cone aperture determined by threshold $\theta$, and its axis determined by the RBF center vector $\vec{w}$. For an illustration, the shape of such a basis function in a two-dimensional afferent space is shown in Figure 2c.

Equation 4 has a straightforward biological interpretation, which is directly applicable to L4 neurons. The first term, $\sum_{i=1}^{n} w_i a_i$, can be interpreted as summation by an L4 neuron of synaptic inputs it receives from its afferent (thalamic or lower-level cortical) neurons. According to this interpretation, $w_i$ is the weight, or efficacy, of the synaptic connection from the $i^{th}$ afferent neuron. Connection weights $w_1 \ldots w_n$ determine the preferred direction $\vec{w}$ of the L4 neuron in the afferent space (Figure 2a).

The second term, $\theta \sqrt{\sum_{i=1}^{n} a_i^2}$, can be interpreted as a net feed-forward inhibition imposed on an L4 neuron by a local group of inhibitory cells that are directly driven by the afferent inputs to the local L4 neighborhood. This term is a sublinear sum of the activities of all the afferent neurons innervating the L4 neighborhood. Such sublinear summation can be accomplished by a combination of nonlinearities across multiple levels, including (1) individual afferent synapses onto the inhibitory L4 cells; (2) integration – including local shunting – of the afferent inputs in the dendritic arbors of the inhibitory cells; and (3) mutual inhibition among the
inhibitory cells that receive their inputs from partially overlapping subsets of the afferent neurons.

Such feed-forward inhibitory cells should have broadly tuned receptive field (RF) properties, essentially reproducing those of the afferent input. In fact, such feed-forward inhibition is a prominent feature of L4 functional architecture (Douglas et al. 1995; Miller et al. 2001; Porter et al. 2001; Alonso and Swadlow 2005). In particular, in the somatosensory barrel cortex feed-forward inhibition is mediated by fast-spiking inhibitory L4 neurons that do have multi-whisker RFs poorly tuned to the direction of whisker deflection, comparable to RFs of the barrels’ thalamocortical afferents (Kyriazi et al. 1996; Bruno and Simons 2002; Swadlow 2002, 2003). In the visual cortex some inhibitory cells in L4 have sharply tuned simple-cell RFs, but others have – as predicted by our model – orientation-insensitive RFs that do not have separate ON and OFF subregions, but respond to both dark and light stimuli (Hirsch et al. 2003; Lauritzen and Miller 2003). Such inhibitory neurons receive larger and faster excitatory inputs from the thalamic afferents than do excitatory L4 neurons and use AMPA receptors in their thalamic synapses that are different from those used by the excitatory L4 neurons (Sun et al. 2006; Cruikshank et al. 2007; Hull et al. 2009b). Instead of a linear I/V relationship observed in the excitatory cells, AMPA-mediated EPSCs in the inhibitory L4 cells show strong inward rectification tendency, which might contribute to their computing of the length of the afferent vector in equation 4. Another contribution to afferent vector length computing might come from voltage-dependent NMDA channels in the thalamic synapses on the inhibitory L4 cells, which stay closed at low levels of afferent activity (Hull et al. 2009b). At the neuron output level, stimuli co-applied to different parts of an inhibitory L4 cell’s RF produce substantial occlusion
(Brumberg et al. 1996), suggesting that such cells can produce inhibitory input to the excitatory L4 cells approximating the length of the afferent input vector.

Thus, to conclude, the biological evidence suggests that L4 neurons do perform something akin to equation 4 operation on their afferent inputs, which should endow them with RBF-like functional properties.

**Local L4 network model**

The focus of our L4 model is on a local group of L4 neurons that are innervated by the same set of afferent neurons. This focus limits the cortical extent of such groups (and thus the scope of the model) to 200-400 μm diameter macro- or even sub-macrocolumnar neighborhoods of L4 (Mountcastle 1978; Marino et al. 2005). Using the published estimates of L4 cell densities in different cortical areas (Beaulieu and Colonnier 1983; Feldmeyer et al. 1999; Keller and Carlson 1999; Budd 2000), a modeled group is estimated to comprise 600-2500 excitatory neurons.

We will model excitatory neurons in such an L4 group using equation 4. Inhibitory neurons in the group will be modeled implicitly by the feed-forward inhibition term of equation 4, assumed to be computed in the cortex by a local set of inhibitory cells from their own direct afferent inputs (Gibson et al. 1999; Hirsch et al. 2003; Lauritzen and Miller 2003; Sun et al. 2006). To optimize stimulus representation, a local group of L4 neurons, innervated by a particular set of afferent neurons, should have their preferred directions $\tilde{w}$ distributed evenly throughout the afferent space defined by the input neurons. This can be achieved through experience-driven self-organization of the afferent connections, which for this purpose can be made Hebbian; i.e., dependent in their strength on correlation of activities of the pre- and postsynaptic neurons (Miller et al. 1999). Thus in setting up our L4 model we postulate that the
strength of the connection between afferent neuron \( j \) and L4 neuron \( i \) reflects their correlation according to:

\[
 w_{ij} = \sqrt{\left[ \rho_{ij} \right]^2 / \sum_{k=1}^{n} \left[ \rho_{ik} \right]^2}
\]  

(5)

where \( \rho_{ij} \) and \( \rho_{ik} \) are the correlation coefficients between stimulus-evoked activities of target neuron \( i \) and afferent neurons \( j \) and \( k \), respectively. These correlation coefficients are computed only for those stimuli that evoke in the postsynaptic neuron \( i \) response \( F > 0 \). Connection weight \( w_{ij} \) is scaled in equation 5 so as to normalize the weights of all the afferent connections converging on an L4 neuron to a unit vector length. Experimental literature does not provide any detailed guidance on the precise relationship between the strength of a connection and the correlation of the pre- and postsynaptic cells. The particular form in equation 5 was chosen here because it produced slightly better function-linearization pluripotency in the model than did other alternatives (see Results).

To make different L4 neurons in the model choose different preferred directions \( \vec{\nu} \) in the afferent space, we link them via lateral connections. Depending on specific details of their implementation, lateral connections can endow networks of Hebbian neurons with a wide variety of functional capabilities (Fyfe 2005). In particular, lateral connections can diversify the afferent connections of the linked neurons if they are made anti-Hebbian (e.g., Foldiak 1990): i.e., the more positively correlated the stimulus-evoked activities of two L4 neurons are, the more negative (inhibitory) should be their connection; whereas the more negatively correlated the stimulus-evoked activities of two L4 neurons are, the more positive (excitatory) should be their connection. Incorporation of such lateral connections and their inputs into equation 4 gives us the formula for computing the output of L4 neuron \( i \):
\[
F_i(\tilde{a}) = \frac{1}{1 - \theta} \left[ \sum_{j=1}^{n} w_{ij} a_j - \theta \sqrt{\sum_{j=1}^{n} a_j^2 + \lambda \sum_{k,k \neq i} (-\rho_{ik}) \cdot F_k(\tilde{a})} \right]^+ 
\]

(6)

where \( \lambda \) is a lateral connection scaling constant, \( F_k \) is the output of L4 neuron \( k \), and \( \rho_{ik} \) is the correlation coefficient between outputs of L4 neurons \( i \) and \( k \). Experimental evidence in support of such anti-Hebbian lateral connectivity in L4 is reviewed in Discussion.

The model described by equations 5 and 6 faces two critical questions: (1) is it an accurate model of cortical L4; and (2) does it have significant function-linearization capabilities? The values of afferent and lateral connections \( w_{ij} \) and \( \rho_{ik} \) in equation 6 are determined by the choice of input patterns \( \tilde{a} \) used to calculate correlation coefficients in equations 5 and 6, which in turn reflect the statistical nature of the sensory environment from which those patterns are drawn. To answer the above two questions, we will develop the model connections on natural images, which should make the resultant model directly comparable to L4 of the primary visual cortex, V1.

**METHODS**

**Visual input patterns**

To generate realistic visual afferent inputs to L4 from the lateral geniculate nucleus (LGN), we modeled a layer of LGN-like neurons (Figure 3) based on the retinal/LGN model of Somers et al. (1995). This LGN layer comprises 91 neurons with retinotopically arranged ON-center receptive fields (RFs) and 91 neurons with retinotopically arranged OFF-center RFs, which together create a hexagonally shaped viewing window (Figure 3ab). The visual inputs to the LGN layer
originate from a set of five 500×335 pixel grayscale photographs containing natural texture-rich images of grass, bushes, trees, etc. (see Figure 3c for some examples). The photographs were not preprocessed in any way.

**FIGURE 3 NEAR HERE**

To generate a particular visual input pattern, the viewing window is centered on a particular location in one of the five photographs. The intensities of the pixels within the viewing window are then convolved with the RF profiles of the 182 LGN neurons. The RF profile is modeled as a difference of the “central” and the “surround” two-dimensional Gaussians, with a common space constant $\sigma$ for both dimensions:

$$R_{xy} = \left(\frac{1}{2\pi\sigma_{center}^2}\right)\exp\left(-\frac{D_{xy}^2}{2\sigma_{center}^2}\right) - \left(\frac{1}{2\pi\sigma_{surr}^2}\right)\exp\left(-\frac{D_{xy}^2}{2\sigma_{surr}^2}\right)$$  \hspace{1cm} (7)

where $\sigma_{center} = 0.8833$ and $\sigma_{surr} = 3\sigma_{center}$ (this yields the center width of 4 pixels). $D_{xy}$ is the distance between a pixel at the (x, y) location in the image and the ($x_0$, $y_0$) image location of the RF center. If $D_{xy} > 3\sigma_{surr}$, $R_{xy} = 0$ (i.e., the RF diameter is restricted to 15.8 pixels).

Thus, the activity of an ON-center LGN neuron with the RF center at the ($x_0$, $y_0$) location in the image is computed as:

$$a_{ON} = \left[0.1 + \sum_{y} \sum_{x} R_{xy} \cdot I_{xy}\right]^+$$  \hspace{1cm} (8)

where $I_{xy}$ is the grayscale intensity of the pixel at (x, y) location in the image ($0 \leq I_{xy} \leq 1$). The activity of an OFF-center LGN neuron is computed as:

$$a_{OFF} = \left[0.1 - \sum_{y} \sum_{x} R_{xy} \cdot I_{xy}\right]^+$$  \hspace{1cm} (9)
RF centers of the 91 ON-center LGN neurons are arranged in a hexagonal pattern, spaced one pixel apart (Figure 3b). RF centers of the 91 OFF-center LGN neurons coincide with the RF centers of the ON-center LGN neurons.

**L4 output**

L4 was modeled as a group of 182 neurons (unless otherwise noted) of the type described above by equations 5 and 6. This choice of having the same numbers of cells in the L4 and LGN layers will allow us to directly compare their function-linearization pluripotencies and thus estimate the pluripotency gain due to the L4 transform. The temporal behavior of each neuron is described by the following differential equation as a leaky integrator:

$$\tau \frac{d}{dt} F_i = -F_i + \frac{1}{1-\theta} \left[ \sum_{j=1}^{182} w_{ij} a_j - \theta \sqrt{\sum_{j=1}^{182} a_j^2} + \lambda \sum_{k=1, k \neq i}^{182} u_{ik} \cdot F_k \right]^+$$

(10)

where $\tau$ is a time constant. This differential equation was solved numerically using Euler updates with a stepsize $\Delta t = 1$ ms, after first ascertaining that the solutions are stable by comparing them with solutions obtained with smaller stepsizes. Explicitly, the Euler update for an equation $\tau (d/dt)x = -x + g(x)$ is $x(t + \Delta t) = (1 - \Delta t / \tau) \cdot x(t) + (\Delta t / \tau) \cdot g(x)$. The time constant $\tau$ was set to 4 ms. The response of the L4 network to a given afferent input pattern was computed in 20 time steps.

**Development of L4 connections**

L4 network connections were initialized by setting all lateral L4-L4 connections to zero and by setting the LGN connections of each L4 neuron to be proportional to the activities of the LGN cells, which were evoked by placing the LGN viewing window in a random location in one of
the database images. LGN connections of different L4 neurons were initialized with different randomly chosen views.

Responses of the L4 cells to visual input patterns were next used to modify the initial connections. The afferent connections in the model are Hebbian. Since the strength of a Hebbian connection basically reflects temporal correlation in activities of the pre- and postsynaptic cells, rather than committing to any particular explicit formulation of the Hebbian rule, we chose to measure directly the correlations of activities of cells in the network in response to a representative set of visual input patterns, and then change the connection strengths accordingly. Because a change in connection strengths in the network will usually result in a change in correlation of cells’ activities, a number of updates of the network’s connections are necessary in order to drive the network’s connectivity into a steady state, in which connection strengths will match the correlations of stimulus-evoked activities of the pre- and postsynaptic cells.

Thus, the connections were modified in 100-1000 update steps. At each step, the L4 network was stimulated with 1000 visual input patterns, which were generated by placing the LGN viewing window in random locations in any of the five database images. Activities of the 182 LGN cells and outputs of the 182 L4 cells in response to these 1000 visual patterns were used to compute correlation coefficients between all pairs of LGN-L4 and L4-L4 neurons, which were used to update the afferent and lateral connections.

At update step $s$, the weight of the afferent connection from LGN cell $k$ to L4 cell $i$ was updated, based on correlation $\rho_{ik}(s)$ of their outputs during step $s$ and according to equation 5, as:

$$w_{ik}(s) = \frac{\sqrt{[\tilde{w}_{ik}(s)]^2}}{\sum_{j=1}^{182} [\tilde{w}_{ij}(s)]^2}$$

where

$$\tilde{w}_{ik}(s) = (1 - \mu_{aff}) \cdot \tilde{w}_{ik}(s-1) + \mu_{aff} \cdot \rho_{ik}(s)$$
Correlation coefficient $\rho_{ik}$ was computed between the activity of LGN cell $k$ and the output of L4 cell $i$ averaged over 20 time steps of responding to each input pattern. The weight of the lateral connection between L4 cells $i$ and $k$ was updated, based on correlation $\rho_{ik}(s)$ of their time-averaged outputs during step $s$, as:

$$u_{ik}(s) = (1 - \mu_{lat}) \cdot u_{ik}(s - 1) - \mu_{lat} \cdot \rho_{ik}(s)$$ (13)

Correlation coefficients were computed only for input patterns that evoked time-averaged $F_i > 0.05$ in the postsynaptic cell. $\mu_{aff}$ and $\mu_{lat}$ are adjustment rate constants controlling how fast the connection weights will converge to stable values. In all experiments we used $\mu_{aff} = 0.01$ and $\mu_{lat} = 0.1$, which produced the fastest convergence to stable values.

**Pluripotency test**

An L4 network with developed afferent and lateral connections was evaluated for its function-linearization potential. We define the pluripotency of an L4 transform as the capacity to represent linearly any function of the afferent input patterns in the domain of potential target functions. Central to our definition of pluripotency is the correlation coefficient $\rho_T$ between a given target function $T(\bar{a})$ and its approximation by a weighted linear sum of activities of the L4 neurons:

$$A(\bar{a}) = \sum_j \rho_j \cdot F_{LA_j}(\bar{a})$$ (14)

In the absence of experimental evidence of any more efficient function learning than Hebbian in the cortex, we assume that connections of L4 cells with L2/3 cells are Hebbian, and the best that can be done is to set the weight $\rho_j$ equal to Pearson’s correlation coefficient between the target function $T(\bar{a})$ and the output of the $j^{th}$ L4 cell. Thus we define pluripotency $P$ of an L4
transform as the expected value of $\rho_T$, averaged across the entire domain of potential target functions that might be developed by the L2/3 neurons of the same cortical column.

However, we do not know the nature of the potential L2/3 target functions and therefore cannot in practice measure pluripotency $P$ directly. What we can do in practice is to estimate $P$ on arbitrarily defined target functions, which might or might not be among the potential L2/3 target functions. For example, we can define target functions by randomly selecting a set of $n$ test input patterns $\{\tilde{a}_1...\tilde{a}_n\}$ as points in the input space and assigning an arbitrary function value $T(\tilde{a})$ to each of these patterns, or points. Such a pluripotency-estimating approach is, in fact, used in Machine Learning to measure the capacity of a given learning algorithm to learn functions on a given input domain (Bartlett and Mendelson 2002; Shawe-Taylor and Cristianini 2004; Ambroladze et al. 2007; Zhu et al. 2009). One such commonly used measure of function-learning capacity is called the Rademacher complexity, and this is the measure we adopted for our pluripotency estimation task. A formal definition of the Rademacher complexity is provided in the Appendix, together with derivation of specific computational procedures (equations 21, 22, and 23) for applying it to our model L4 network as a measure of its pluripotency, $R^2_\alpha(A)$.

As is explained in the Appendix, $R^2_\alpha(A)$ can be thought of as an estimate of the fraction of variance in the target variable that is explained by its L4-based approximating function $A$. $R^2_\alpha(A)$ provides an estimate of the pluripotency of the L4 transform’s linearization of a broad range of potential L2/3 target functions with symmetric probability distributions and defined on sets of points in the afferent input space.
RESULTS

L4 network pluripotency

LGN and lateral connections in the model were optimized by developing them under various settings of the model parameters $\theta$ and $\lambda$ and finding those settings under which the model L4 achieves its highest performance on the Rademacher complexity test of function linearization pluripotency.

Figure 4 compares pluripotency – expressed by the squared Rademacher complexity $\mathcal{R}_n^2(A)$ – of the L4 network with pluripotencies of the LGN layer and the viewing window pixels. Since $\mathcal{R}_n^2(A)$ is measured on functions that are defined on sets of $n$ points in the input space, its value will depend on the number of points used. In Figure 4, $\mathcal{R}_n^2(A)$ was measured on sets of different sizes, from $n = 25$ to $n = 400$. First, we plot $\mathcal{R}_n^2(A)$ measured on the intensities of the pixels in the viewing window, i.e., the pixels that provide the afferent input to the LGN neurons in the model. As the plot shows, the pixel layer is essentially incapable of function linearization: its $\mathcal{R}_n^2(A)$ is close to zero for all but very small sets of test patterns. (By its definition, $0 \leq \mathcal{R}_n^2(A) \leq 1$.) Next, we also plot in Figure 4 $\mathcal{R}_n^2(A)$ measured on the outputs of 182 LGN neurons and observe that $\mathcal{R}_n^2(A)$ of the LGN layer very closely matches that of the pixel layer. That is, the LGN transform of the raw image has absolutely no function-linearization capability of its own. Finally, we plot $\mathcal{R}_n^2(A)$ measured on the outputs of 182 L4 neurons. In contrast to the pixel and LGN layers, L4 has much higher $\mathcal{R}_n^2(A)$ values, thus indicating that L4 has well developed pluripotent function-linearization capabilities.
As can be expected, $R_n^2(A)$ of L4 gradually declines when the number $n$ of the test input patterns is increased, since the limited number of neurons in L4 restricts its function-approximation capacities. The contrast in $R_n^2(A)$ between L4 and LGN is most revealing at $75 < n < 150$ and in the rest of the paper we will measure the Rademacher complexity at $n = 100$.

L4 pluripotency depends greatly on experience-driven self-organization of LGN-L4 connections. If the values of these connections are assigned randomly, the L4 network loses all its pluripotency: instead of having $R_{100}^2(A) = 0.69 \pm 0.004$ with self-organized connections, L4 with random LGN connections has $R_{100}^2(A) = 0.034 \pm 0.0004$. The reason for this effect is that LGN-L4 connections are Hebbian and, under influence of competitive lateral interactions among L4 cells, each L4 cell develops a different set of LGN connections $\mathbf{w}$, which gives that cell a particular preferred direction in the afferent space. Natural images occupy only a small fraction of the afferent space. Once LGN-L4 connections are trained on such images (Figure 3c), the L4 cells will chose preferred directions within the subspace of the natural images. L4 cells will thus be able to map this subspace most densely, rather than spreading over the entire – but overwhelmingly “uninhabited” – afferent space. This is important for L4 pluripotency because the L2/3 target functions $T(\mathbf{a})$ are defined exclusively just over the subspace of natural images.

To demonstrate that L4 pluripotency is confined to the subspace of natural images, we can test it on input patterns drawn randomly from the entire afferent space of the network. In that case, L4 pluripotency drops from $R_{100}^2(A) = 0.69$ to $R_{100}^2(A) = 0.00004 \pm 0.00002$.

The model's ability to linearize nonlinear functions also depends on feed-forward inhibition and lateral interactions among L4 neurons. As Figure 5a shows, without the feed-
forward inhibition (i.e., $\theta = 0$), the model L4 network is incapable of function linearization ($R^2_{100}(A) = 0.10 \pm 0.034$). L4 with optimal feed-forward inhibition but without lateral connections (i.e., $\lambda = 0$) also fails to develop pluripotency ($R^2_{100}(A) = 0.15 \pm 0.005$), for an obvious reason that without lateral connections L4 neurons cannot adequately diversify their afferent connections and responses to input patterns. **It takes both optimal feed-forward inhibition and lateral interactions for the L4 model to achieve high pluripotency** of $R^2_{100}(A) = 0.69 \pm 0.004$.

**FIGURE 5 NEAR HERE**

How effective is our L4 model in developing its pluripotency compared to the DHL function approximator of Deco and Obradovic (Deco and Obradovic 1995)? As described above, the DHL function approximator also utilizes a pluripotent approach to function linearization in its first stage, in which RBFs are distributed efficiently in the input space via a combination of Hebbian and anti-Hebbian learning. We trained the first stage of the DHL approximator, comprising 182 RBF units, on the activity patterns evoked in the 182-cell LGN layer by the same natural images used to develop the L4 model. We were able to achieve the top DHL performance on our pluripotency test with RBF parameter $\sigma = 0.02$. As plotted in Figure 5a, this performance ($R^2_{100}(A) = 0.58 \pm 0.017$) was lower than the top performance of the equal-size L4 model ($R^2_{100}(A) = 0.69$; the difference is statistically significant at $p < 0.001$), thus suggesting that L4 utilizes an efficient approach to function linearization.

The pluripotency performance of the L4 model can be raised further by increasing the number of cells in the L4 network. For example, when we increased the number of L4 cells from 182 to 400 and trained this expanded model with the same parameter settings ($\theta = 0.675$ and $\lambda = \ldots$)
3), the pluripotency performance increased from $R^2_{100}(A) = 0.69$ to $R^2_{100}(A) = 0.84 \pm 0.004$ (Figure 5a).

Can comparable levels of pluripotency performance be achieved with recurrent, rather than feed-forward, inhibition? To answer this question, we replaced feed-forward inhibition in equation 6 with a similar term that expressed the vector length of the outputs of the 182 L4 cells:

$$F_i(\vec{a}) = \frac{1}{1 - \theta} \left[ \sum_{j=1}^{182} w_{ij} a_j - \eta \sqrt{\sum_{k=1}^{182} F_k^2(\vec{a}) + \lambda \sum_{k,k \neq i} (-\rho_{ik} \cdot F_k(\vec{a}))} \right]^+$$  \hspace{1cm} (15)

The highest pluripotency performance obtained in the L4 model with such recurrent inhibition ($R^2_{100}(A) = 0.34 \pm 0.013$ at $\eta = 0.75$) was much lower than the performance of the feed-forward inhibition model (Figure 5a).

The dependence of the model pluripotency on the strength of the feed-forward inhibition is plotted in Figure 5b. To generate this plot, the afferent and lateral connections were developed multiple times under identical conditions with the exception of the feed-forward inhibition parameter $\theta$, which was varied systematically from its minimal possible value of zero to the highest value of 0.9. As this plot shows, the L4 network develops its maximal or near-maximal pluripotency at a wide range of $\theta$ values. Only at low settings of $\theta$ (at which feed-forward inhibition is mostly ineffective) and at very high settings of $\theta$ (at which feed-forward inhibition is so strong that it prevents L4 from responding to at least some of its input patterns) does the model pluripotency plunge from its high levels. Thus Figure 5b suggests that high pluripotency is a readily emergent property of L4-like networks that have robust feed-forward inhibition.

Turning to lateral connections, panels c and d of Figure 5 show that in addition to contributing to pluripotency by diversifying afferent connections of L4 cells, lateral connections also increase pluripotency by dynamically shaping the response of the L4 cell population to each
stimulus. That is, Figure 5c shows in the L4 network developed under the optimal values of parameters $\theta$ and $\lambda$, but tested for pluripotency under varying $\lambda$ values, that the weaker the lateral stimulus-evoked interactions in L4, the lower the L4 pluripotency. By itself this plot does not rule out the possibility that even in the absence of lateral stimulus-evoked interactions among the L4 neurons we might still be able to generate high levels of pluripotency, by simply increasing the value of $\theta$ and thereby reducing the sizes of the RFs in the afferent space. To test this possibility, we used the model with optimally developed thalamic and lateral connections and measured its pluripotency while systematically varying its $\theta$ parameter. The results, plotted in Figure 5d, show that when lateral stimulus-evoked interactions are made ineffective ($\lambda = 0$) in the network with optimally developed LGN connections, no amount of feed-forward inhibition will generate by itself the high levels of pluripotency.

Interestingly, in Figure 5d the L4 pluripotency shows greater dependence on parameter $\theta$ than in Figure 5b. In Figure 5b, at each plotted $\theta$ value the L4 was tested for pluripotency only after its connections were first allowed to self-organize into a stable pattern at that $\theta$. In contrast, in Figure 5d the network was allowed to self-organize only once under one optimal $\theta$ value ($\theta = 0.675$), and then it was tested under different $\theta$ values. Significantly, by changing the value of $\theta$ after the development of the network’s connections we no longer have these connections completely self-organized (i.e., if allowed, they would readjust to reflect the changed $\theta$ value). Thus the difference in the two plots indicates that a completed self-organization of the network’s connections is necessary for achieving maximal pluripotency.

Lastly, we explore three deviations from our L4 model design to address three questions. First question: How important for the L4 pluripotency is the specific form of feed-forward inhibition and LGN synaptic plasticity used in the model? The RBF theoretical origins of our
model specify that feed-forward inhibition should be computed as the length of the afferent input vector (equation 4), and LGN connections of an L4 cell should have a unit vector length (equation 5). We trained an alternative L4 model, in which feed-forward inhibition was computed more simply as a sum of activities of all 182 LGN cells and the strength of the connection between LGN cell $j$ and L4 a cell $i$ was computed also more simply as:

$$w_{ij} = \left[ \rho_{ij} \right]^+ / \sum_{k=1}^{182} \left[ \rho_{ik} \right]^+$$

(16)

The best pluripotency performance we could obtain from this simplified model was $R_{100}^2(A) = 0.58$, which is 16% lower than the best pluripotency performance we could obtain in the original model. While significant, such a relatively small reduction in the pluripotency performance suggests that biological L4 neurons should be able to achieve maximal or near-maximal pluripotency using sublinear summation approximations of the LGN activity and connection vector lengths.

Second question: In designing our model L4 neuron, we scaled its RBF by the length $l_2$ of the stimulus vector $\hat{a}$ (see equation 3). Does this scaling impact the model pluripotency? Without this scaling, the output of the model L4 cell would be a function of simply the angle $\alpha$ between the stimulus vector $\hat{a}$ and the cell’s preferred direction $\hat{w}$:

$$\Phi(\bar{a}) = \frac{1}{1 - \theta} \left[ \cos \alpha - \theta \right]^+ = \frac{1}{1 - \theta} \left[ \hat{w} \cdot \left( \frac{\hat{a}}{\sqrt{l_2}} \right) - \theta \right]^+$$

(17) ($\hat{w}$ has a unit vector length). The best pluripotency performance we were able to obtain from the model in which L4 cells were based on such an RBF was $R_{100}^2(A) = 0.70$, which is not significantly different from the performance of our original model ($R_{100}^2(A) = 0.69$), in which L4 cells are based on the $l_2$-scaled RBF of equation 3.
Finally third question: In our model we use anti-Hebbian lateral connections to diversify afferent connections of L4 cells. Might we be able to diversify these connections by other means? Winner-take-all networks, for example, avoid the use of plastic lateral connections. We explored a simple, biologically plausible alternative to anti-Hebbian lateral connections by replacing them with uniform-strength inhibitory connections. That is, we replaced the anti-Hebbian term $+ \lambda \sum_{k \neq i} (-\rho_{ik}) \cdot F_k(\tilde{a})$ in equation 6 with a fixed lateral inhibition term $- \lambda \sum_{k \neq i} F_k(\tilde{a})$:

$$F_i(\tilde{a}) = \frac{1}{1 - \theta} \left[ \sum_{j=1}^{n} w_{ij} a_j - \theta \sqrt{\sum_{j=1}^{n} a_j^2} - \lambda \sum_{k \neq i} F_k(\tilde{a}) \right]^+$$ (18)

While the pluripotency performance of the L4 model with such fixed lateral inhibition was found to be inferior to the performance of our original model ($R_{100}^2(A) = 0.63$ instead of 0.69), the difference was small enough (9%) to make this alternative model potentially interesting.

Having established that our L4 model does exhibit significant pluripotent function-linearization capabilities, we turn to the second critical question: is it an accurate model of cortical L4? To answer this question we compare the emergent structural and functional properties of the model L4 with those of the real L4 of cat V1. The comparison is done for a particular representative maximally-pluripotent 182-cell L4 network model whose afferent and lateral connections were developed in 1000 batch steps with $\theta = 0.675$ and $\lambda = 3$.

LGN-L4 connections and receptive fields

Figure 6 shows LGN connectional patterns and RFs of six exemplary model L4 neurons. RFs were generated using the reversed correlation method of Jones and Palmer (1987) by measuring
the response of an L4 cell to 1×3pixel rectangular dark and bright stimuli placed at all pixel positions in the viewing window. To display a two-dimensional RF, each pixel in the viewing window was color-coded by the difference between the cell’s responses to the bright and dark stimuli centered on that pixel. The four neurons on the left in Figure 6 are representative of the majority of L4 neurons in the model. Their RFs have one or more prominently elongated ON or OFF subfields. These ON and OFF subfields of an RF arise from rows of ON-center and OFF-center LGN cells. Such a pattern of parallel ON and OFF strips makes the model RFs closely resemble RFs of the simple-cell class of neurons in cat’s L4 (Hubel and Wiesel 1962; Martinez et al. 2005). The correspondence between LGN connectional patterns of the model L4 cells and their RFs is quite straightforward for the first four cells shown in Figure 6 and consistent with the rules of connectivity between LGN cells and cat V1 simple cells as defined by Alonso et al. (2001). Still, as Figure 6 shows, even for these cells RFs tend to underestimate the full extent and richness of LGN connectional patterns.

**FIGURE 6 NEAR HERE**

RFs of the two rightmost cells in Figure 6 (#88 and #115) are much less informative about their underlying LGN patterns. These cells represent a small minority of L4 cells in the model whose LGN patterns do not fit the classic simple-cell pattern of parallel ON and OFF strips. The RF of cell #88 looks like a typical simple-cell RF and offers no indication that its LGN pattern has two OFF regions almost at a right angle to each other. Cell #115 responds very little to RF-mapping stimuli, giving no indication of its branching LGN pattern.

Figure 7 shows LGN connectional patterns of all 182 L4 cells in the model. Each cell has a unique LGN pattern, and a large majority (75%) of the cells have LGN patterns of the type represented by the four leftmost cells in Figure 6. The patterns (and correspondingly RFs) vary
in the orientation of their strip-like subfields, covering the full range of possible orientations. The number of subfields in an LGN connectional pattern varies from 2 to 4. The average number of subfields in an LGN pattern is 2.70, which is close to the average of 2.45-2.65 for the number of subfields in the RFs of cat’s simple cells (DeAngelis et al. 1993; Troyer et al. 1998). The majority of simple cells have 2 or 3 subfields, and in both the model and in V1 only ~10% of cells have more than 3 subfields (Jones and Palmer 1987). The average length-to width ratio of the dominant subfields in the model RFs is ~3.8, which is also not far from the average aspect ratio of 4.3-4.5 reported for cat’s simple RFs (Gardner et al. 1999; Troyer et al. 1998).

FIGURE 7 NEAR HERE

About 25% of L4 cells in the model acquired LGN connectional patterns distinct from the simple-cell type illustrated by four cells in Figure 6. Two examples are illustrated in Figure 6 by cells #88 and #115, and another two are marked by boxes in Figure 7. These LGN patterns seem to reflect either termination of visual lines or their junctions. Most numerous (15-20%) are end-inhibition LGN patterns, in which one or more subfields do not extend across the entire length of the LGN field (e.g., see the top boxed-in pattern in Figure 7). These patterns resemble RFs of hypercomplex, or end-stopping, cells (Hubel and Wiesel 1962; Dreher 1972), which also make up ~25% of cells in V1 (Tolhurst and Thompson 1981). The rest (<10%) of the LGN patterns might be generally described as branching. Such LGN patterns have not been described in the literature. However, it should be pointed out that our knowledge of LGN connectional patterns onto V1 cells is mostly indirect, inferred from the RFs mapped in V1 (e.g., Hubel and Wiesel 1962) and supported by recordings of only limited numbers of synaptically connected LGN and V1 neurons (e.g., Alonso et al. 2001). As the examples of cells #88 and #115 in Figure 6 suggest, V1 neurons with branching LGN patterns – which would be relatively infrequently
encountered during experiments – can easily be passed over unrecognized because they would
either have unremarkable simple-cell RFs (e.g., cell #88) or appear to be simply unresponsive to
visual stimulation (e.g., cell #115).

To reveal the importance of different components of the model for the development of its
LGN connectional patterns, we trained the model under several alternative conditions. Figure 8a
shows LGN connectional patterns acquired by 12 exemplary L4 cells when the model was
trained on random-dot images. These LGN patterns are clearly different from simple-cell
patterns in Figure 7. In each, the LGN field is broken down into a mosaic of amorphously
shaped ON-center and OFF-center subregions. This outcome demonstrates that the reason why
our L4 model acquires its simple-cell LGN patterns when trained on natural images is because of
the prominence of local lines and edges in such images.

In Figure 8b, the model was trained on natural images, but without feed-forward
inhibition ($\theta = 0$) and lateral connections ($\lambda = 0$). In this case L4 cells did acquire simple-cell
LGN patterns, showing that Hebbian plasticity of LGN-L4 connections alone is sufficient to
produce simple-cell RFs (see also Lee et al. 2000). However, almost all L4 cells acquired LGN
patterns containing just two subfields: the average number of subfields is 2.10, which is much
smaller than the average in the real L4.

FIGURE 8 NEAR HERE

In Figure 8c, the model was trained on the same sequence of natural images as in Figure
8b, also without lateral connections ($\lambda = 0$), but with feed-forward inhibition ($\theta = 0.675$). LGN
patterns of most of the L4 cells changed little from Figure 8b to Figure 8c, but some did acquire
an extra subfield. As a result, the average number of subfields increased to 2.28. Training the
model without feed-forward inhibition ($\theta = 0$), but with lateral connections ($\lambda = 3$) also increased
the number of LGN subfields (average = 2.40), as can be seen in Figure 8d. In addition, lateral interactions led to emergence of branching and end-inhibition LGN patterns in 10% of L4 cells.

It is when they are both present, in the complete model, that feed-forward inhibition and lateral connections together generate the richest set of LGN patterns (Figure 8e).

Finally, to show the effect of the viewing window size, we trained the model on a larger LGN layer, expanded from 182 to 338 cells (Figure 8f). As would be expected, this allowed L4 cells to increase the number of subfields in their LGN patterns to an average = 3.19.

Orientation tuning

As can be expected from their striped LGN connectional patterns and RFs, the model L4 cells exhibit high sensitivity to orientation of elongated stimuli. Figure 9 plots the responses of five representative L4 cells to stimulation with sinusoidal gratings. Three of these cells (#32, 6 and 35) have typical LGN connectional patterns made up of 2, 3 or 4 strip-like ON-center and OFF-center subfields (see panel a in Figure 9). Two other cells (#88 and 115) belong to the small (<10%) minority of the model L4 cells with unconventional branching LGN patterns.

Panel b of Figure 9 shows the responses of the five cells to the standard gratings test used to classify V1 neurons (Skottun et al. 1991). As this panel shows, all five cells prominently modulate their response to the grating of the optimal spatial frequency and orientation depending on its spatial phase. In contrast, the same grating presented at the orientation orthogonal to the optimal one fails to evoke responses at any spatial phase. The combination of having discrete ON-center and OFF-center subregions in the LGN connectional patterns (and RFs) and a high
degree of modulation of responses to grating stimuli identify all the five cells in Figure 9 as clearly belonging to the class of simple cells (Hubel and Wiesel 1962; Skottun et al. 1991). Panel c shows orientation tuning plots of the five cells, which were obtained at three different stimulus intensities (or image contrast): 0.33, 0.67 and 1.0, where 1.0 is the maximal contrast present in the training natural images. Orientation tuning of the majority-representing top three cells is contrast invariant, as it is in the real V1 simple cells (Sclar and Freeman 1982). The other two cells in Figure 9 exhibit some small tightening of their orientation tuning with increased grating contrast. For example, the half-width at half-height (HWHH) of cell #88 declines from 36° to 28° to 24° when the stimulus strength is raised from 0.33 to 0.67 to 1.0. Such dependence of orientation tuning on contrast is atypical of the model L4 cells and reflects the fact that grating stimuli are not well suited for the cells with unconventional LGN connectional patterns, generating only very weak afferent drive and responses (note the greatly reduced ordinate scales for cells #88 and #115).

Panel d shows the effect of the spatial frequency of the grating stimuli on cells’ orientation tuning. The optimal spatial frequency varies among the model L4 cells between 0.06 and 0.16 cycles/pixel, with the average near 0.12 cycles/pixel. To convert these numbers into cycles/degree, we note that the diameter of the central RF region of the LGN cells in the model is equal to 4 pixels, whereas it is approximately equal to 30' in the central retina (Somers et al. 1995). Thus equating one pixel to ~7.5' of visual angle, the average optimal spatial frequency of the model L4 cells becomes ~1.0 cycles/degree. This average is close to the average optimal spatial frequency of 0.86 cycles/degree of V1 cells at 5° eccentricity (Movshon et al. 1978).

In panel d, orientation-tuning curves are plotted for gratings of below-optimal spatial frequency and above-optimal frequency. For each of the five cells the two curves are
superimposed in the same plot. These plots reveal that orientation tuning of all five cells is
tighter for gratings of higher spatial frequency. Similar relationship between orientation tuning
and spatial frequency obtains for simple cells in cat’s V1 (Vidyasagar and Siguenza 1985).

Figure 10a summarizes orientation tuning of the 136 (75%) L4 cells in the model that
exhibit the classical Hubel/Wiesel-type of LGN connectional patterns and simple-cell RFs. The
average orientation tuning is plotted for grating stimuli of two intensities: at the maximal contrast
present in the training natural images, and at 1/3 of the maximal contrast. The two curves show
that orientation tuning of these cells is contrast invariant, as it is in the real V1 simple cells (Sclar
and Freeman 1982). The average half-width at half-height (HWHH) measure of orientation
tuning of these cells is 15°, which matches the average of 16° reported for simple cells in cat’s
V1 (e.g., Rose and Blakemore 1974; Gardner et al. 1999).

FIGURE 10 NEAR HERE

Local L4 connections

How important are feed-forward inhibition and lateral connections for L4 orientation tuning?
Figure 10 plots average orientation tuning of L4 cells with simple-cell LGN patterns in the
model developed either without feed-forward inhibition (Figure 10b) or without lateral
connections (Figure 10c). In each case, L4 neurons have much broader orientation tuning than in
the full model. Without feed-forward inhibition, L4 cells have above-zero response to stimuli of
all orientations and average HWHH = 28°. Without lateral connections, L4 cells have average
HWHH = 30°. Thus, the presence of both feed-forward inhibition and lateral interactions is
required for the emergence of biologically accurate orientation tuning in our L4 model.
Figure 10d reproduces cortical inactivation experiments of Ferster et al. (1996) and Chung and Ferster (1998), in which cortical spiking activity was silenced through cooling or electrical shock. Without intracortical interactions, the intracellularly recorded membrane potential of cat’s L4 cells mostly reflects their LGN input, which was found to be already as tuned to orientation as the cell’s output (measured intracellularly). Figure 10d plots orientation-tuning curves averaged over the 136 L4 cells in the model that have classical simple-cell RFs. Plotted superimposed are orientation-tuning curves of (1) the cells’ LGN input in response to gratings of optimal frequency, and (2) the cells’ output prior to its rectification. Comparable to findings of Ferster and colleagues, the curves match closely, revealing only small tightening of orientation tuning of cells’ output by intracortical processing of the already well-tuned LGN inputs. The tightening is most pronounced in the vicinity of the optimal stimulus orientation, which is consistent with the experimental evidence of iso-orientation inhibition (Ferster 1986).

Lateral interactions are mediated in the model by anti-Hebbian connections among L4 cells. The strength and physiological sign of these connections is determined by the correlation of activities of the pre- and postsynaptic cells in response to the training set of natural images. Figure 11 plots the distribution of these correlation coefficients for all possible pairs among 182 L4 cells in the model. The plot shows first that L4 cells are highly decorrelated in their stimulus-evoked behaviors. Second, negative correlations are much more numerous than the positive correlations. And third, since the strength of anti-Hebbian connections in the model is set equal to the negative of the correlation coefficient (see equation 6), the plot in Figure 11 can also be viewed as horizontally flipped plot of lateral connection strengths among L4 cells. Thus, excitatory lateral connections are much more numerous (73%) than the inhibitory connections (27%). These numbers are consistent with the observed abundance of excitatory
interconnections among neighboring spiny cells in L4 (Anderson et al. 1994a; Stratford et al. 1996; Tarczy-Hornoch et al. 1999). Most of the excitatory lateral connections in Figure 11 have near-zero strengths and would be hard to detect experimentally. Indeed, Thomson et al. (2002) detected monosynaptic EPSPs in 17% of simultaneously recorded pairs of excitatory neurons in L4 of cat V1 (for a comparison, the top 17% of the most negatively correlated neuron pairs are indicated in Figure 11 by an arrow).
DISCUSSION

Relationship to RBF networks

We built an effective pluripotent function linearizer with common-use neural network components (Fyfe 2005). They include: (1) a cell’s output is computed as a weighted sum of its afferent inputs; (2) afferent connection weights are developed using a Hebbian rule; and (3) diversity of afferent connectional patterns among the cells in the network is achieved via anti-Hebbian lateral connections. However, by adding a stimulus strength-scaled threshold \( \theta \) (equation 3), we converted a conventional neural network into an RBF-like network (Lowe 2003), because such a threshold makes the network cells behave similar to RBF units.

Importantly, RBF networks are recognized as highly capable universal function approximators (Park and Sandberg 1991; Kůrková 2003). Pouget and Sejnowski (Pouget and Sejnowski 1997) and Poggio and colleagues (Poggio 1990; Vetter et al. 1995; Poggio and Bizzi 2004) have already argued – more abstractly than in this paper – that building perceptual and motor functions by linear summation of RBF-like units can be an effective cortical strategy. Their suggested possible means by which cortical networks might implement RBFs, however, were more complex than the \( \theta \) thresholding used in this paper (Poggio 1990; Poggio and Hurlbert 1993; Poggio and Bizzi 2004).

The network we developed is in some important regards more complex than a basic RBF network. In particular, in a high-dimensional afferent space defined by the afferent inputs, the region of the afferent space in which a standard RBF unit has an above-zero output (this region can be thought of as the unit’s RF in the afferent space) has the shape of a hypersphere. In our network with a stimulus strength-scaled threshold \( \theta \), a cell’s RF in the afferent space has the shape of a circular cone, rather than a hypersphere. A network with such “conical” basis
functions (CBF) is equivalent to an RBF network that has vector rescaling pre- and post-
processing stages. That is, in response to any given input pattern, that pattern is first normalized
by its vector length, next the RBF network computes the output to this unit-length input pattern,
and finally the RBF output is scaled by the original input vector length to produce the final
output. Thus, in a CBF network the direction of an input vector in the input space specifies the
direction of the output vector in the output space, whereas the length of the input vector does not
affect the direction of the output vector, only its length.

Another difference between RBF networks and our network concerns lateral
interconnections among the cells in the network. Whereas in a basic RBF network these
connections (or their functional equivalents) are used during the network development to
optimize the distribution of RBF centers in the input space, in our network lateral connections
also participate in stimulus-evoked dynamics, interactively shaping the response of the network
to each presented input pattern. As we show in Figure 5cd, such recurrent dynamics is
imperative for achieving maximal pluripotency in the network.

Resemblance to L4 of cat’s V1

The building components of our L4 model reflect the following properties of the real L4: (1)
strong feed-forward inhibition (Miller et al. 2001), which performs stimulus strength-scaled
thresholding in the model; (2) presence of inhibitory cells with strong direct thalamic inputs
(Cruikshank et al. 2007) and unoriented RFs (Hirsh et al. 2003), which implement feed-forward
inhibition; and (3) high density of excitatory interconnections among the cells in the L4 network
(Anderson et al., 1994a; Tarczy-Hornoch et al. 1999; Thomson et al. 2002).
When it is developed on natural images and LGN-like input patterns, our network acquires additional structural and functional properties that closely match the properties of L4 of the cat primary visual cortex: (1) self-organization of LGN connections to L4 cells into narrow parallel ON-center and OFF-center strips, producing simple-cell RFs (Hubel and Wiesel 1962; Alonso et al. 2001); (2) comparable numbers of RF subfields and aspect ratios (Jones and Palmer 1987; DeAngelis et al. 1993; Gardner et al. 1999); (3) emergence of end-inhibition RFs/hypercomplex cells (Hubel and Wiesel 1962; Dreher 1972; Tolhurst and Thompson 1981); (4) prominent phase modulation of cells’ responses to grating stimuli of optimal orientation (Skottun et al. 1991); (5) narrow orientation tuning of comparable half-width at half-height (Rose and Blakemore 1974); (6) contrast invariance of orientation tuning (Sclar and Freeman 1982); (7) comparable average optimal spatial frequency of grating stimuli (Movshon et al. 1978); (8) narrower orientation tuning for grating stimuli of higher spatial frequencies (Vidyasagar and Siguenza 1985); (9) narrow orientation tuning of LGN inputs to L4 cells, close to orientation tuning of their outputs (Ferster et al. 1996; Chung and Ferster 1998); and (10) presence of iso-orientation inhibition (Ferster 1986).

As was shown in Figures 8 and 10, the presence of both feed-forward inhibition and anti-Hebbian lateral connections is required in order for L4 cells in the model to develop the biologically accurate diversity of multi-subfield RFs and acquire orientation tuning matching in sharpness that of real L4 neurons.

While we find the RF and stimulus-response properties of the model and real L4 cells to be in close qualitative and quantitative agreement, we also note two exceptions. The first one is that the RF subfield aspect ratios are somewhat smaller in the model (the average ratio in the model is 3.8, whereas it is 4.3-4.5 in the real V1). We tentatively attribute this difference to the...
absence of L6-to-L4 connections in the model, which in the real cortex can substantially elongate
the L4 RFs via excitatory synapses on spiny cells of L4 (Anderson et al. 1994b; Tarczy-Hornoch
et al. 1999; Binzegger et al. 2004). The second exception to the otherwise close match between
the model and V1 concerns the range of preferred spatial frequencies. The highest preferred
spatial frequency that we find in the model is ~1.3 cycles/degree, whereas Movshon et al.
(Movshon et al. 1978) report 3 cycles/degree in V1. This difference might, however, reflect
insufficient realism in our modeling of the LGN layer and its inputs to L4 (see discussion in
Lauritzen and Miller 2003), rather than problems with the model of L4 itself.

One of the major features of our model is the presence of extensive interactions among
L4 cells. Indeed, high density of interconnections among neighboring excitatory cells is a
prominent feature of L4 architecture (Anderson et al. 1994a; Stratford et al. 1996; Tarczy-
Hornoch et al. 1999; Feldmeyer et al. 1999; Petersen and Sakmann 2000; Thomson et al. 2002;
Binzegger et al. 2004). These excitatory connections are represented in equation 6 in the term
\[ \lambda \cdot \sum (-\rho_{ik}) \cdot F_k \] by those among the presynaptic cells \( k \) that have a negative correlation \( \rho_{ik} \) with
the postsynaptic cell \( i \). More than 70% of the cells in a local group have negative correlations,
although many of those are close to zero (see Figure 11). In the model, excitatory
interconnections are a part of the mechanism that drives L4 cells to diversify their afferent
connections and efficiently map their common afferent space. Importantly, such a role requires
lateral excitatory connections in L4 to be modifiable by experience (so that their strength can
reflect the correlation \( \rho_{ik} \) in the behaviors of the pre- and postsynaptic cells), but in an unusual
way: i.e., rather than being conventionally Hebbian, these L4 connections have to be anti-
Hebbian.
Experimental support for the model’s proposal that interconnections among excitatory L4 cells should be anti-Hebbian comes from Egger et al. (1999; also see Sáez and Friedlander 2009 for more recent support). In paired recordings from synaptically connected spiny stellate neurons in L4 of the barrel field they found that correlated pre- and postsynaptic action potentials, synchronized to occur within 25 ms of each other, reliably induce in those cells long-term synaptic depression (LTD) rather than long-term potentiation (LTP). This LTD following correlated activity is specific for L4 excitatory neuron connections: in the same slice preparation, the same correlated activity induced LTP in pairs of pyramidal neurons that were studied in layers 2/3. The same conditions also produce LTP in layer 5 (Markram et al. 1997). Egger et al. (1999) did not identify the conditions that would produce potentiation of the L4 connections, but at least a recovery from depression – if not potentiation – must take place in order for those connections to exist.

The model’s anti-Hebbian mechanism for diversification of afferent connections also posits that those excitatory L4 cells that are positively correlated in their stimulus-evoked activities should inhibit each other. This lateral inhibition is represented in equation 6 in the term \( \lambda \cdot \sum (-\rho_{ik}) \cdot F_k \) by those among the presynaptic cells \( k \) that have a positive correlation \( \rho_{ik} \) with the postsynaptic cell \( i \). In the cortex, such anti-Hebbian lateral inhibition can be mediated disynaptically by either basket of neurogliaform cells of L4 (Tarczy-Hornoch et al. 1998) or even monosynaptically, as reported by Lee and Sherman (2009), via L4-specific group II metabotropic glutamate receptors.

Because of a large disparity in the numbers of excitatory and inhibitory neurons in L4 (i.e., six or more excitatory neurons per one inhibitory neuron), the same inhibitory neuron will have to mediate disynaptic links of more than one excitatory cell while somehow keeping their
links functionally separate. A possible solution to this “cross-talk” problem is suggested by a 
study of Ren et al. (2007), who recorded pairs of neighboring (<75 μm) excitatory (i.e., 
pyramidal) neurons in layer 2/3 of the mouse visual cortex and showed that in ~30% of such 
pairs a single action potential in one pyramidal cell evoked a large inhibitory postsynaptic 
current (IPSC) in the other cell. Ren et al. showed that such inter-pyramidal IPSCs do not 
involves generation of somatic action potentials in the inhibitory neurons, but instead are 
generated by direct axo-axonic activation of the presynaptic terminals of inhibitory (possibly 
basket) neurons, which in turn synapse on the target pyramidal cell (for contradictory findings 
see, however, Hull et al. 2009a). Similar axo-axonically mediated inhibition, which bypasses 
dendrites and somata of inhibitory neurons, might operate among pairs of L4 excitatory cells.

Another possible neural-circuit implementation of anti-Hebbian inhibition between 
excitatory cells derives from a reinterpretation of its mathematical expression in equation 6. 
That is, the lateral (anti-Hebbian) input term in equation 6 can be expanded into two terms:

$$\lambda \sum_k (1 - \rho_{ik}) \cdot F_k = \lambda \sum_k \left(1 - \rho_{ik}\right) \cdot F_k - \lambda \sum_k F_k$$  (19)

Note that the connection weight $$u_{ik} = (1 - \rho_{ik})$$ cannot be negative (i.e., inhibitory). This means 
that lateral input in equation 6 can be equivalently implemented by a combination of (1) inputs 
from exclusively excitatory anti-Hebbian connections of the type described by Egger et al. 
(1999) and (2) a fixed-connection input from one or just a few inhibitory cells that summate the 
output activities of all the excitatory cells in the local L4 network. In fact, a single inhibitory cell 
can carry out both feed-forward $$\theta \left(\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$$ and feedback $$\lambda \sum_k F_k$$ inhibition (Ahmed et al. 
1997).
Comparison with existing L4 models

The key feature of our model is the presence of feed-forward inhibition, the magnitude of which is proportional to the overall strength of the stimulus activating the local L4 network (the length of the afferent input vector), but is insensitive – invariant – to any spatial details of the stimulus patterns. In the model this untuned feed-forward inhibition performs the stimulus strength-scaled thresholding function, which endows the network with RBF-like properties and pluripotent function-approximation capabilities (see above).

A large and growing number of experimental studies in the somatosensory barrel and visual cortices (e.g., Kyriazi et al. 1996; Bruno and Simons 2002; Swadlow 2002; Hirsch et al. 2003; Sun et al. 2006; Cruikshank et al. 2007; Hull et al. 2009b; but see Cardin et al. 2007) have established the prominent presence of untuned feed-forward inhibition in L4 and its control over the responses of excitatory L4 cells to sensory stimuli. These experimental studies have motivated some successful applications of feed-forward inhibition to L4 modeling. For example, Kyriazi and colleagues (Kyriazi and Simons 1993; Kyriazi et al. 1996; see Miller et al. 2001 for a larger-picture review) used untuned feed-forward inhibition in a one-barrel model with randomly connected excitatory and inhibitory cells, which were activated by spike trains previously recorded from thalamic neurons, to successfully explain the behaviors of barrel neurons in response to various patterns of whisker stimulation.

Lücke (2009) used untuned feed-forward inhibition – computed as a sum of stimulus-evoked activities of all the afferent sources – to self-organize LGN-L4 connections in his model of a V1 cortical column. When trained on natural images, transmitted to L4 via an LGN-like layer similar in design to ours, the L4 cells in Lücke’s model acquire realistic simple-cell RFs. Although Lücke used a soft winner-take-all mechanism to diversify RFs among the L4 cells,
rather than the anti-Hebbian mechanism used in our model, the two models develop similar
assortments of RF shapes, sizes, orientations, etc., comparable to the visual cortical L4.
However, L4 cells in Lücke’s model are linear (their output is not rectified, as in our model) and
do not have any lateral interactions, which should make them incapable of pluripotent function
linearization and contrast-invariant orientation tuning.

Lauritzen and Miller (2003) used untuned feed-forward inhibition to extend an older
model of Troyer et al. (1998). Untuned feed-forward inhibition plays the central role in their
model in making orientation tuning of the model cells invariant to stimulus contrast. One
limitation of this model is that its LGN-L4 connections are hard-wired to produce RF properties
matching those of simple cells, rather than developing them via stimulus-driven self-
organization. On the other hand, one component of Lauritzen and Miller’s model that our model
lacks is an anti-phase inhibition; i.e., inhibition between L4 cells in whose RFs ON subregions of
one neuron overlap OFF subregions of the other neuron and vice versa. In their model such
inhibition sharpens the spatial frequency tuning of simple cells, reduces low temporal frequency
responses and increases network stability. The presence of such anti-phase inhibition has been
demonstrated experimentally in cat V1 by Ferster (1988), Hirsch et al. (1998), Anderson et al.
(2000), and Monier et al. (2003).

L4 cells with anti-phase RFs are negatively correlated in their activities (when one is
active, the other is silent and vice versa) and as a result in our model they are linked by
excitatory – rather than inhibitory – anti-Hebbian lateral connections (an example can be seen in
Figure 11). The absence of anti-phase inhibition in our current model will have to be addressed
in future. In fact, we can expect anti-phase inhibition to enhance L4 function-linearization
pluripotency in response to moving stimuli. To explain, because of a major contribution of
relatively slow-acting NMDA receptor-mediated synaptic transmission to stimulus-evoked activities of L4 neurons (Crair and Malenka 1995; Gil and Amitai 1996; Fleidervish et al. 1998; Feldmeyer et al. 1999; Hull et al. 2009b), optimally oriented moving grating stimuli, for example, will depolarize any given simple L4 cell not only at the times when the grating matches the ON and OFF subregions of the cell’s RF, but also during the in-between periods when the grating is traveling over the antagonistic RF subregions (see, for example, Anderson et al. 2000). If not for the anti-phase inhibition, which goes up during these anti-phase periods, simple cells would fire more or less continuously during the grating motion, thereby losing some spatial phase information about the stimulus that might be of potential value to the upper-layer neurons.

What our study contributes to the field of feed-forward inhibition models of L4 is an appreciation of their potential for pluripotent function linearization. Specifically, we show that feed-forward inhibition can make L4 neurons behave as RBF units (equations 1-4), thus enabling them to approximate nonlinear functions. We also show that high pluripotency is a readily emergent property of L4-like networks with robust feed-forward inhibition, quite tolerant of a wide range of feed-forward inhibition strengths (Figure 5b). In contrast, the other well-studied class of visual cortical L4 models (Teich and Qian 2006), in which feed-forward inhibition is replaced with recurrent inhibition, is much less pluripotent according to our simulations (Figure 5a). We find dynamical shaping of the responses of L4 neurons to impinging stimuli, achieved via lateral interactions among the neighboring L4 neurons, also to be necessary for maximal pluripotency. Finally, we find that visually-driven self-organization of our model network makes it not only highly pluripotent, but also comparable to L4 of cat V1 in its RF and orientation-tuning properties.
Ours is not the only way to build a pluripotent function linearizer that would be consistent with L4 structural and functional properties. A principally different approach was developed by Maass et al. (2002, 2004) based on the Liquid State Machine (LSM) class of neural networks. LSM models of cortical microcircuits, encompassing ~100 μm diameter cortical neighborhoods, are made up of spiking Hodgkin-Huxley point neurons with conductance-based synapses that have a biologically realistic short-term plasticity. Connections between neurons are assigned randomly, but with the probabilities matching those reported for specific cortical layers, such as L2/3, L4, or L5 (Haeusler and Maass 2007; Haeusler et al. 2009). In response to external inputs, such randomly-connected circuits of spiking neurons exhibit near-chaotic nonlinear dynamics, generating spiking patterns across the network that happen to represent linearly a diverse range of functions over the input patterns. The best linearization performance is achieved at the edge of chaos (Legenstein and Maass 2007).

Similar to our proposal, Maass and colleagues also view L4 as a pluripotent function linearizer. Their LSM-based approach to function linearization is appealing in its simplicity and biological plausibility, but it remains to be established how well its pluripotency compares to that achieved by our RBF-based approach. Also, because all the connections in the LSM networks are random, they lack realistic RFs. However, as Maass et al. (2004) point out, connections in LSM networks can be allowed to self-organize and cells then might acquire realistic RF properties.

Model limitations and future developments

Our aim in this paper was to introduce the idea that L4 might play the role of a pluripotent function linearizer in cortical information processing. We limited the model developed in this
paper to a local L4 neighborhood and spatial – rather than spatiotemporal – afferent input patterns. In the future, the model should be expanded to spatiotemporal input patterns and to larger L4 territories comprising a continuum of local L4 neighborhoods of the type modeled here. An extension of the function-linearizing capabilities of our model network to the time domain should give the network new parallels to cortical L4 (if L4 does perform the task of a pluripotent function linearizer) with regard to time integrating and differentiating processes, stimulus-response dynamics, stimulus motion representation, temporal frequency tuning, and other time-related L4 properties (including anti-phase inhibition).

Also the design of our pluripotent function-linearizing network should be expanded from dealing with only a relatively small set of afferents innervating just a local L4 neighborhood to dealing with a large spatially distributed complement of afferents innervating L4 of an entire cortical area. This expansion of the model will require an integration of the local function-linearization neural circuitry developed in this paper with pericolumnar circuitry, which in sensory cortical areas is responsible for experience-driven formation of somatotopic or retinotopic maps, maps of preferred stimulus orientation, spatial frequency, direction of motion, ocular dominance, etc.

In conclusion, the fact that an efficient pluripotent function linearizer, designed on a few generic neurally-guided principles, exhibits emergent structural and functional properties that closely resemble those of cortical L4 strongly supports our initial kernel-inspired conjecture. It suggests that L4 has effective function-linearization capabilities and that its major function is to perform a transform of its afferent input enabling the upper layers to learn and compute complex functions using operations that are to a large degree linear. This possibility should be tested directly by simultaneously recording responses of local groups of 50 or more L4 neurons in the
visual cortex to an ecologically representative variety of natural images and measuring their
group performance on a pluripotency test of the type used in this paper. With such experimental
demonstration of pluripotent function-linearization capabilities of L4, our model would offer a
highly explanatory conceptual framework for understanding L4 structural and functional
properties and its contribution to cortical information processing.

**APPENDIX**

**Estimating L4 pluripotency using the Rademacher complexity**

The *Rademacher complexity* is increasingly used in machine learning to analyze statistical
properties of kernels and to measure the capacity of the hypotheses space of a learning algorithm
on a given input domain (Bartlett and Mendelson 2002; Shawe-Taylor and Cristianini 2004;
Ambroladze et al. 2007; Zhu et al. 2009). The *empirical Rademacher complexity* of a class of
functions $F$ on a sample $S = (z_1 \ldots z_n)$ drawn from the domain space $Z$ is defined as:

$$
\hat{R}_S(F) = \mathbb{E}_\sigma \left[ \frac{2}{n} \sup_{f \in F} \left| \sum_{i=1}^{n} \sigma_i f(z_i) \right| \right]
$$

(20)

where $\sigma_1 \ldots \sigma_n$ are independent (Rademacher) random variables with equal probability $p = 0.5$ of
having values of 1 or -1. The *Rademacher complexity* of the function class $F$ for sample size $n$
is:

$$
R_n(F) = \mathbb{E}_Z \left[ \hat{R}_S(F) \right]
$$

(21)

In measuring Rademacher complexity of the $F_{L4}$-based linear approximating function $A$
(equation 14), $z_i$ will correspond to an afferent input pattern $\tilde{a}_i$, while instead of approximating
L2/3 target functions $T(\tilde{a})$, the L4 transform will be evaluated for its capacity to approximate
Rademacher variables $\sigma_i$ with weighted linear sums of activities of the L4 neurons (equation 14).
Since the weights $\rho_j$ in equation 14 are uniquely determined by correlation of $\sigma$ and $F_{L4_j}$, *supremum* vanishes in equation 20:

$$\hat{R}_S(A) = \mathbb{E}_\sigma \left[ \frac{2}{n} \sum_{i=1}^{n} \sigma_i \cdot A(\tilde{a}_i) \right]$$  \hspace{1cm} (22)

Another important consideration is that we will want to use Rademacher complexity to compare function-linearization capabilities of different networks (e.g., L4 vs. LGN, or L4 networks with different strengths of feed-forward inhibition or lateral interactions). To be able to do that, we will have to scale $A(\tilde{a})$ to zero-mean and $\frac{1}{2}$ standard deviation (any fixed standard deviation would work, but we chose $\frac{1}{2}$ to cancel the 2 in the numerator of equation 22):

$$A(\tilde{a}) = 0.5 \left( \sum_j \rho_j \cdot F_{L4_j}(\tilde{a}) - m \right) / s$$ \hspace{1cm} (23)

where $m$ and $s$ are the mean and standard deviation, respectively, of $A(\tilde{a})$ in equation 14, measured on the $n$ input patterns $\tilde{a}_1...\tilde{a}_n$ in equation 22.

An interesting consequence of using the scaled $A(\tilde{a})$ functions in equation 22 is that it makes $\hat{R}_S(A)$ a close estimate of the correlation coefficient between Rademacher variables $\tilde{\sigma}$ and $A(\tilde{a})$. This means, in accordance with our definition of pluripotency $P$, that Rademacher complexity for sample size $n$, $R_n(A)$, in effect estimates the pluripotency of the L4 transform with regard to its capacity to linearize sets of $n$ Rademacher variables.

The usefulness of Rademacher complexity lies in the fact that any other symmetric random variable besides the Rademacher variable will have the same complexity for a given class of functions $F$ up to a constant (Ledoux and Talagrand 1991; Ambroladze et al. 2007). In other words, Rademacher complexity provides an estimate of the pluripotency of the L4...
transform linearization of a broad range of potential L2/3 target functions with symmetric
probability distributions and defined on sets of points in the afferent input space.

In this paper we estimated the pluripotency of the model L4 network by computing
Rademacher complexity for samples of various sizes. A sample \( S \) of \( n \) test input patterns
\( \{\tilde{a}_1, \ldots, \tilde{a}_n\} \) was obtained by placing the LGN viewing window in randomly chosen locations in
the five database images and using only those \( n \) input patterns whose LGN vector length was
greater than 0.3 (to avoid low-contrast, nondescript views). The empirical Rademacher
complexity \( \hat{R}_n(A) \) was computed using equation 22, in which \( A(\tilde{a}_i) \) were computed using
equation 23 and expectation \( E_\sigma \) was obtained from 30 random sets of Rademacher variables \( \tilde{\sigma} \).
Rademacher complexity \( R_n(A) \) was computed using equation 21, based on 30 randomly
selected test input pattern samples \( S \).

Since in our application Rademacher complexity in effect estimates the expected
correlation between the target variable \( \tilde{\sigma} \) and its linear L4-based approximation \( A(\tilde{a}) \), we chose
to use squared Rademacher complexity \( R_n^2(A) \), because then it can be interpreted as the
coefficient of determination, estimating the fraction of variance in \( \tilde{\sigma} \) explained by \( A(\tilde{a}) \).

Acknowledgments

The authors thank Dan Ryder, Douglas G. Kelly, Barry L. Whitsel and Mark Tommerdahl for
helpful discussions.

Grants

Supported, in part, by Army Research Office grant W911NF-08-1-0308 and by Istanbul
University grant YADOP-5323.
REFERENCES


Budd JML. Inhibitory basket cell synaptic input to layer IV simple cells in cat striate visual cortex (area 17): a quantitative analysis of connectivity. *Vis Neurosci* 17: 331-343, 2000.


Gil Z, Amitai Y. Adult thalamocortical transmission involves both NMDA and non-NMDA receptors. *J Neurophysiol* 76: 2547-2554, 1996.


Figure 1. Problem linearization strategy of transforming the input space into a “feature” space. (a) The strategy is illustrated on a classification problem. The curved decision boundary separating two classes of data samples (little open and black squares) in the input space is made linear – and thus easier to learn – by mapping the data samples into a nonlinear transform of the input space. (b) Hypothesized function linearization strategy in a prototypical cortical column. Input from afferent sources (thalamic or lower-level cortical neurons) is distributed across the L4 cells in the column, which in turn distribute their outputs across the L2/3 cells in the same column. Although the target functions to be learned by L2/3 cells are likely to have a complex nonlinear form in the afferent space ($\rho$, the correlation coefficient between an afferent source and a target function, is near zero) and therefore difficult to learn, the L4 cells transform the afferent input space into a new space, in which the target functions become linear and therefore easier to learn.

Figure 2. Afferent input transformation by the model neuron. (a) In the two-dimensional afferent input space defined by the stimulus-evoked activities of two afferent sources, $a_1$ and $a_2$, any stimulus pattern is represented by an activity vector. All stimulus patterns of a unit vector length (i.e., their $l_2 = 1$) lie on a quarter-circle (rather than on a full circle, since afferent inputs cannot be negative). Suppose a particular point on this quarter-circle (at $w_1$ and $w_2$ coordinates) is chosen as an RBF center. Then the distance between this RBF center and any given unit-length stimulus-evoked input pattern is determined solely by the angle $\alpha$ formed by their two vectors. (b) The $\phi(\alpha)$ measure of the distance between the RBF center vector and any given
unit-length stimulus vector, plotted as a function of the angle $\alpha$ between them. $\varphi(\alpha)$, defined by

equation 1, is a thresholded and scaled cosine of angle $\alpha$, plotted here for threshold $\theta = 0.7$. The
angle $\alpha$ cannot be greater than 90° due to the afferent space constraints. (c) The basis function
$\Phi(\vec{a})$ of equation 4, plotted in the two-dimensional afferent space defined by the afferent inputs
$a_1$ and $a_2$. Parameters used: $\theta = 0.95$, $w_1 = 0.83$, $w_2 = 0.55$.

Figure 3. LGN layer model. (a) RF profiles of ON-center and OFF-center model LGN cells. (b) The map
of the RF centers (little white boxes) of the 91 ON-center cells of the LGN layer. Note that RF centers are arranged in a hexagonal pattern. RF centers are shown superimposed on a 13×13 pixel field, in which each pixel is shown as a black-edged gray square. RF centers of the OFF-center LGN cells match the RF centers of the ON-center cells. (c) Four exemplary 300×250 pixel subregions of the larger natural images that were used to activate the LGN layer and train the model connections.

Figure 4. Performance of the model network on the pluripotency test. Rademacher complexities $R_n^2(A)$ of the pixel, LGN, and L4 layers of the model are plotted as a function of $n$ (the number of test input patterns). Note the contrast between very poor performances of the pixel and LGN layers vs. high performance of L4.

Figure 5. Importance of the feed-forward inhibition and lateral connections to the pluripotency of the model L4. (a) Performance of the model network on the pluripotency test (expressed by the Rademacher complexity measure $R_{100}^2(A)$) under seven different conditions:
LGN – $R_{100}^2(A)$ of the LGN output; L4 – $R_{100}^2(A)$ of the outputs of 182 L4 cells under the optimal choice of parameters $\theta$ and $\lambda$; L4 w/o FFI – feed-forward inhibition is ineffective ($\theta = 0$) during the model training and testing; L4 w/o LAT – lateral connections are ineffective ($\lambda = 0$) during the model training and testing; DHL – $R_{100}^2(A)$ of the DHL function approximator on the same test functions; L4 (400) – $R_{100}^2(A)$ of 400-cell L4; L4 w REC – $R_{100}^2(A)$ of 182-cell L4 with recurrent, but not feed-forward, inhibition. Each bar is an average of $R_{100}^2(A)$ obtained in 10 models, each developed from a different randomly chosen initial state and tested on a different set of test functions. Error bars are standard deviations. (b) Pluripotency performance of the model self-organized and tested under the same optimal value of $\lambda$ but different values of $\theta$. (c) Pluripotency performance of the model self-organized under the optimal values of $\theta$ and $\lambda$, but tested for pluripotency using varying values of the lateral connection scaling constant $\lambda$. (d) Pluripotency performance of the model self-organized under the optimal values of $\theta$ and $\lambda$, but tested for pluripotency using varying values of the feed-forward inhibition threshold $\theta$. The top curve shows pluripotency in the network with the optimal $\lambda$; the lower curve shows pluripotency in the same network with ineffective lateral connections; i.e., $\lambda = 0$. Note that in the absence of lateral interactions, no setting of $\theta$ can produce the same high levels of pluripotency as reached under optimal $\lambda = 3$; i.e., feed-forward inhibition cannot fully substitute for stimulus-evoked dynamics with regard to pluripotency.

**Figure 6. LGN connections and RFs of six exemplary model L4 cells.** For each L4 cell, the LGN layer is shown as a hexagonally shaped mosaic of 91 small hexagons, each of which represents a pair of co-localized ON- and OFF-center LGN cells. The relative positions of the
cell-representing hexagons in the LGN field reflect the relative positions of their RF centers in
the visual field (which are arranged in a hexagonal retinotopic pattern, as shown in Figure 3b).
The color of each LGN cell-representing hexagon reflects the strength of its connection to the
target L4 cell and whether the connection comes from the ON-center (red) or OFF-center (blue)
LGN cell. LGN connectional patterns of the shown L4 cells subdivide the LGN field mostly into
long but narrow alternating strips of ON-center and OFF-center connections. RFs of the four
leftmost cells appear as contrast-enhanced versions of the LGN connectional patterns. These
RFs exhibit from 1 to 4 highly elongated ON and OFF subfields. The two rightmost cells are
examples of model L4 cells that developed unusual, branching LGN patterns. Significantly, the
exceptional nature of these LGN patterns would not be guessed from their RFs, which appear
either conventional (cell #88) or just poorly defined (cell #115).

Figure 7. LGN connectional patterns of all 182 L4 cells of the model network.

Figure 8. LGN connectional patterns of 12 exemplary L4 cells developed under a variety of
conditions. (a) The L4 model with full-strength feed-forward inhibition and lateral connections
was trained on random-dot images rather than on natural images. Note that L4 cells failed to
acquire oriented LGN patterns. (b-d) The L4 model was trained on natural images while its
feed-forward inhibition (d), lateral connections (c), or both (b) were made ineffective. (e) The
full L4 model was trained on the same sequence of natural images as in b-d. (f) The diameter of
the LGN layer was increased from 11 cells to 15 cells and the full L4 model was trained on the
same sequence of natural images as above. Note a gradual increase in complexity of LGN
patterns from b to f.
Figure 9. Orientation tuning of five model L4 cells. (a) LGN connectional patterns of the five cells. (b) Responses of the cells to grating images (shown above each plot). A cell’s response $F$ to a grating image is plotted as a function of the grating’s spatial phase relative to the LGN viewing window. Left – responses to optimally oriented gratings of the optimal spatial frequency. Right – responses to orthogonally oriented gratings. (c) Orientation tuning in response to gratings of the optimal spatial frequency. Plotted superimposed are tuning curves in response to three grating contrasts: 0.33 (dotted), 0.67 (gray), and 1.0 (black). $\Delta F$ – difference between the maximal and minimal responses of a cell to gratings of the same orientation, but varied across the full 0-360º range of spatial phases. In all the plots, the minimal response was always zero, making $\Delta F$ effectively the maximal response to the gratings of a given orientation. (d) Orientation-tuning curves in response to gratings of two spatial frequencies: dotted curve – spatial frequency lower than optimal; solid curve – spatial frequency higher than optimal.

Figure 10. Average orientation tuning of the model L4 cells. (a) Average orientation tuning of all 136 L4 cells with simple-cell RFs (gray curve–maximally contrasted grating stimuli; black curve–grating stimuli at 1/3 of the maximal contrast). Note that stimulus contrast does not change the tuning width. (b) Average orientation tuning of all the simple cells (n=163) in the model in which feed-forward inhibition was turned off ($\theta = 0$) during training and testing. (c) Average orientation tuning of all the simple cells (n=174) in the model in which lateral connections were turned off ($\lambda = 0$) during training and testing. (d) Average orientation tuning of the LGN input to simple cells in the full L4 model ($\Delta_{\text{LGN}}$–gray curve). Also plotted is average orientation tuning of cells’ unrectified output ($\Delta V$–black curve).
Figure 11. Correlation of stimulus-evoked responses among the L4 cells in the model network. For each of the 16471 possible pairs of L4 cells in the model, the correlation coefficient was computed between their responses ($F$) to 1000 image patches randomly taken from our training set of five natural images. Plotted is the distribution of the correlation coefficient values obtained from all cell pairs. Also shown are LGN connectional patterns of 4 pairs of L4 cells with the most positive and most negative correlations. Note that although the LGN patterns of the negatively correlated cell pairs are almost complete negatives of each other, their responses are only very weakly anti-correlated.
a)

FEATUER SPACE

\[ \Phi_2(x_1, x_2) \]

input space

\[ x_1 \]

\[ x_2 \]

b)

cortical column

L2/3

\[ \Sigma \]

L4

\[ \Phi \]

Afferent sources

\[ 1 \]

\[ \ldots \]

\[ n \]

correlation with target L2/3 functions:

\[ \rho = 1 \]

\[ |\rho| > 0 \]

\[ \rho \approx 0 \]
a. $\theta = 0.675 \quad \lambda = 3 \quad$ Random-dot images

b. $\theta = 0 \quad \lambda = 0 \quad$ Natural images
c. $\theta = 0.675 \quad \lambda = 0$
d. $\theta = 0 \quad \lambda = 3$
e. $\theta = 0.675 \quad \lambda = 3$
f. $\theta = 0.675 \quad \lambda = 3 \quad$ Larger LGN layer