A Preference for Phase-Based Disparity in a Neuromorphic Implementation of the Binocular Energy Model

Eric K. C. Tsang
eeeric@ee.ust.hk
Bertram E. Shi
eeber@ee.ust.hk
Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

The relative depth of objects causes small shifts in the left and right retinal positions of these objects, called binocular disparity. This letter describes an electronic implementation of a single binocularly tuned complex cell based on the binocular energy model, which has been proposed to model disparity-tuned complex cells in the mammalian primary visual cortex. Our system consists of two silicon retinas representing the left and right eyes, two silicon chips containing retinotopic arrays of spiking neurons with monocular Gabor-type spatial receptive fields, and logic circuits that combine the spike outputs to compute a disparity-selective complex cell response. The tuned disparity can be adjusted electronically by introducing either position or phase shifts between the monocular receptive field profiles. Mismatch between the monocular receptive field profiles caused by transistor mismatch can degrade the relative responses of neurons tuned to different disparities. In our system, the relative responses between neurons tuned by phase encoding are better matched than neurons tuned by position encoding. Our numerical sensitivity analysis indicates that the relative responses of phase-encoded neurons that are least sensitive to the receptive field parameters vary the most in our system. We conjecture that this robustness may be one reason for the existence of phase-encoded disparity-tuned neurons in biological neural systems.

1 Introduction

The accurate perception of the relative depth of objects enables both biological organisms and artificial autonomous systems to interact successfully with their environment. Binocular disparity, the positional shift between the image locations of an object in two eyes or cameras caused by the difference in their vantage points, is one important cue that can be used to infer depth.

Neurons in the mammalian visual cortex combine signals from the left and right eyes to generate responses selective for a particular disparity.
Barlow, Blackemore, & Pettigrew, 1967; Poggio, Motler, Squatrito, & Trotter, 1985). Ohzawa, DeAngelis, and Freeman (1990) proposed the binocular disparity energy model to explain the responses of binocular complex cells and found that the predictions of this model match measured data from the cat. Cumming and Parker (1997) showed that this model also matches data from the macaque. Although there are discrepancies between the model and the biological data (Ohzawa, DeAngelis, & Freeman, 1997; Chen, Wang, & Qian, 2001; Cumming, 2002), it remains a good first-order approximation.

In the model, a neuron can achieve a particular disparity tuning by a position or a phase shift between its monocular receptive field (RF) profiles of the left and right eyes. Based on an analysis of a population of binocular cells, Anzai, Ohzawa, and Freeman (1999a) suggest that the cat primarily encodes disparity via a phase shift, although position shifts may play a larger role at higher spatial frequencies. A preference for phase encoding is also consistent with human psychophysical data (Schor, Wood, & Ogawa, 1984; DeValois & DeValois, 1988). More recent data from the alert macaque suggest that a mixture of position and phase encodings is common (Prince, Cumming, & Parker 2002; Tsao, Conway, & Livingstone, 2003). Since computational studies indicate that populations of neurons tuned by either encoding can be used to recover disparity (Qian & Zhu, 1997), the natural question that arises is, What are the relative advantages of one encoding scheme over the other?

This letter suggests one possible advantage of neurons tuned by phase shifts: robustness in the presence of neuronal variability, which we discovered in the course of developing a neuromorphic implementation of disparity-tuned neurons. Although physiological measurements reveal a large diversity of responses from populations or neurons, most computational studies of the binocular energy model assume retinotopic arrays of neurons that are identical except for position or phase shifts. In our neuromorphic implementation, variability is unavoidable due to manufacturing tolerances. Thus, we face a similar problem as biological systems: computing with inexact computing elements. Solutions to this problem will lead to more robust engineering systems, as well as possible insights into biological computational strategies.

Section 2 reviews the binocular energy model and the encoding of disparity by position and phase shifts. Section 3 describes our implementation, which is constructed using a pair of silicon retinas that feed into a pair of silicon chips containing retinotopic arrays of neurons with monocular Gabor-type RF profiles. Section 4 presents measured results from the system. We find that the relative responses of neurons tuned by phase encoding are in better agreement with theoretical predictions than neurons tuned by position encoding. In our system, random transistor mismatch introduces mismatch between the nominally identical monocular neurons used to construct the binocular responses. We investigated the sensitivities of the position and phase models numerically to variations in different parameters of the monocular RF profiles. For most parameters, phase encoding
is more robust than position encoding. We also characterized the variations in the monocular RF profiles within one chip. Serendipitously, we find that the phase model is least sensitive to the parameters that vary the most in our system. Section 5 summarizes our results and future directions.

2 The Binocular Energy Model

Ohzawa et al. (1990) proposed the binocular energy model to explain the response of binocular complex cells measured in the cat. The model is similar to the motion energy model previously proposed by Adelson and Bergen (1985) for human perception of visual motion. Anzai, Ohzawa, and Freeman further refined the model (1999a, 1999b, 1999c).

In the model, the response of a binocular complex cell is the linear combination of the outputs of four binocular simple cells, as shown in Figure 1. Anzai et al. (1999b) model the response of a binocular simple cell using a linear binocular filter followed by a half-squaring nonlinearity. The linear binocular filter sums the responses of two monocular filters modeled by Gabor functions.

Let $I(x)$ denote image intensity where $x \in \mathbb{R}^2$. The output of a Gabor monocular filter $m(c, \phi, I)$ is given by

$$m(c, \phi, I) = \int g(x, c, \phi)I(x)dx$$

$$g(x, c, \phi) = \kappa \exp \left( -\frac{1}{2}(x - c)^T C^{-1}(x - c) \right) \cos(\Omega^T (x - c) + \phi),$$

Figure 1: The binocular energy model of complex cell (adapted from Anzai et al., 1999a). The response of binocular complex cell is the linear combination of four binocular simple cell outputs. Each binocular simple cell is constructed from a linear binocular filter followed by half-squaring nonlinearity. The monocular filters combined by the binocular filter are modeled as Gabor functions.
where $\Omega \in \mathbb{R}^2$ and $C \in \mathbb{R}^{2 \times 2}$ control the spatial frequency tuning and bandwidth of the filter, and $\kappa$ controls the gain. These parameters are assumed to be the same in all of the simple cells that make up a complex cell. However, the center position $c \in \mathbb{R}^2$ and the phase $\phi \in \mathbb{R}$ vary between the two eyes and among the four simple cells. The linear binocular filter output is

$$b(c_R, c_L, \phi_R, \phi_L) = m(c_R, \phi_R, I_R) + m(c_L, \phi_L, I_L),$$

where the subscripts R and L differentiate the right and left eye center positions, phases, and image intensities. The simple cell output is given by

$$r_s = (|b(c_R, c_L, \phi_R, \phi_L)|^+)^2,$$

where $|b|^+ = \max\{b, 0\}$ is the positive half-wave rectifying nonlinearity.

While the response of simple cells depends heavily on the stimulus contrast and phase in addition to the disparity and position of the stimulus, the response of complex cells is largely independent of the phase and contrast. The binocular energy model posits that complex cells achieve phase invariance by linearly combining the outputs of four simple cells whose binocular filters differ in phase by $\pi/2$. If $r_c(c_R, c_L, \phi_R, \phi_L)$ is the complex cell output, then

$$r_c(c_R, c_L, \phi_R, \phi_L) = \sum_{k=0}^{3} r_s \left( c_R, c_L, \phi_R + \frac{k}{2}, \phi_L + \frac{k}{2}, \frac{\pi}{2} \right).$$

Since filters that differ in phase by $\pi$ are equal except for a change in sign,

$$r_s \left( c_R, c_L, \phi_R + 2\frac{\pi}{2}, \phi_L + 2\frac{\pi}{2} \right) = (|b(c_R, c_L, \phi_R, \phi_L)|^-)^2$$

$$r_s \left( c_R, c_L, \phi_R + 3\frac{\pi}{2}, \phi_L + 3\frac{\pi}{2} \right) = \left( |b \left( c_R, c_L, \phi_R + \frac{\pi}{2}, \phi_L + \frac{\pi}{2} \right)|^- \right)^2,$$

where $|b|^- = \max\{-b, 0\}$ is the negative half-wave rectifying nonlinearity. Thus, we can construct the four required simple cell outputs from positive and negative half-wave rectified outputs of two binocular filters.

Complex cells constructed according to the binocular energy model are selective for disparities in the direction orthogonal to their preferred orientation. For example, cells tuned to vertical orientations are selective for horizontal disparities. Their disparity tuning depends on the relative center positions and the relative phases of the two monocular filters that make up the binocular linear filter. Define the disparity, $d$, of a binocular stimulus to be the shift of the right eye stimulus with respect to the left eye stimulus, $I_R(x) \approx I_L(x - d)$. A binocular complex cell whose monocular filters are shifted by $\Delta c = c_R - c_L$ and $\Delta \phi = \phi_R - \phi_L$ will respond maximally when the input has a preferred disparity

$$D_{\text{pref}} \approx \Delta c - \frac{\Delta \phi}{\Omega}.$$
where $\Omega$ is the spatial frequency tuning in the horizontal direction. However, this is only an approximation, as the preferred disparity depends also on the frequency content of the input stimulus (Qian, 1994; Zhu & Qian, 1996). Disparity is encoded by a position shift if $\Delta c \neq 0$ and $\Delta \phi = 0$. Disparity is encoded by a phase shift if $\Delta c = 0$ and $\Delta \phi \neq 0$. The cell uses a hybrid encoding if both $\Delta c \neq 0$ and $\Delta \phi \neq 0$. Phase encoding and position encoding are equivalent for the zero-disparity-tuned cell. Assuming that the monocular RFs have the same centers, the complex cell output of Figure 1 is tuned to zero disparity.

3 Experimental Setup

3.1 System Overview. Figure 2 shows a block diagram of our binocular cell system, which uses a combination of analog and digital processing.

The visual front end consists of two silicon retina chips (Zaghoul & Boahen, 2004a,b) mounted behind 4 mm lenses that focus light onto them. Each retina contains a $60 \times 96$ pixel array of phototransistors that sample the incident light and convert it into a current. Processing circuits on the chip generate spike outputs that mimic the spiking responses of sustained-ON, sustained-OFF, transient-ON, and transient-OFF retinal ganglion cells. Our system uses only the sustained outputs, which respond continuously to a static image. ON and OFF cells respond to positive and negative contrasts. The ON and OFF cells are arranged in arrays of $30 \times 48$ neurons. Each chip has an asynchronous digital output bus that communicates the spikes in the array off chip using the address event representation (AER) protocol (Boahen, 2000).

The two Gabor-type chips perform the required monocular filtering operations (Choi, Shi, & Boahen, 2004). Each chip can process a $32 \times 64$ pixel input image, which is encoded as the difference in spike rate of ON and OFF input channels, which are supplied using the AER protocol. Our system uses only the upper left $30 \times 48$ pixels due to the lower resolution of the silicon retinas. Each chip filters its input image with both EVEN and ODD symmetric Gabor-type filters, corresponding to phases $0$ and $-\pi/2$. Four silicon neuron circuits at each pixel encode the positive (ON) and negative (OFF) components of the EVEN and ODD filter outputs as spike trains, which are communicated off chip using AER. We refer to these signals using the abbreviations $e^+$, $e^-$, $o^+$, and $o^-$.

The Gabor-type filters differ from Gabor filters in that the modulating function is not gaussian but still decays with distance from the origin. For a one-dimensional image, the RF profile can be approximated by

$$g(x, c, \phi) = \kappa \frac{\Delta \Omega}{2} \exp(-\Delta \Omega |x|) \cos(\Omega (x - c) + \phi), \quad (3.1)$$

where $x \in \mathbb{R}$ indexes pixel position. The parameters $\Omega \in \mathbb{R}$ and $\Delta \Omega \in \mathbb{R}$ control the spatial frequency tuning and bandwidth, $\kappa$ controls the gain,
Figure 2: Block diagram of the neuromorphic implementation. Connections via AER are denoted with oppositely oriented arrows. The outputs of the address filters are ON-OFF encoded voltage spike trains. The outputs of the binocular combination block are spike rate estimates obtained by binning. (A) The system configured for position encoding. (B) The system configured for phase encoding of positive disparity. Negative disparity can be encoded by swapping the routing between the ON and OFF channels from one retina in the binocular combination block.
\c c \in \mathbb{R} \text{ indicates the center position, and } \phi \in \mathbb{R} \text{ represents the phase of the profile. Since the distance between the points where Fourier transform of the RF profile drops to approximately half its peak value is } 2 \Delta \Omega, \text{ we refer to } \Delta \Omega \text{ as the half-bandwidth. The difference in the filter shape should not affect the resulting binocular complex cell responses significantly; Qian and Zhu (1997) have shown that the binocular complex cell responses in the energy model are insensitive to the exact shape of the modulating envelope. Because we expect primarily horizontal disparities, we tune the Gabor-type chips for vertical orientations.}

The two address filters and the binocular combination block combine the outputs of the two Gabor-type chips to implement two-phase quadrature binocular filters. Each address filter passes only those spikes from the four neurons at a desired retinal location and demultiplexes them as voltage pulses on four separate digital lines. The binocular combination block merges the spike trains from the two eyes into four lines that represent the outputs of two phase quadrature binocular filters differentially. By changing the retinal location selected from the left and right eyes in the AER address filters, we can change the position encoding of disparity. By altering the routing in the binocular combination block, we can change the phase encoding. We implement the address filters and the binocular combination block using Xilinx complex programmable logic devices (CPLDs).

The spike trains on the four output lines of the binocular combination block are binned over a 40 ms time window to estimate the spike rate then differenced to recover the two binocular filter outputs. The binocular filter outputs are then positive and negative half-squared to compute four simple cell outputs. The simple cell outputs are summed to compute the complex cell output. Spike binning is performed on the same Xilinx CPLD used to implement the binocular combination block. The final differencing, squaring, and summing operations are done by an 8051 microcontroller (MCU) running at 24 MHz. The final binocular cell responses are sent to a PC via the MCU serial port for analysis.

The remaining sections describe the design of the AER address filter and the binocular combination block in more detail and can be skipped without loss of continuity. The silicon retina and Gabor-type transceiver chip have been described elsewhere. The final steps computed by the microcontroller are not discussed further, as they are quite straightforward.

3.2 Address Filter. The AER output of each Gabor-type chip encodes the spike activity of all the neurons on that chip. The address filter extracts only those spikes corresponding to the four neurons at a selected retinal location.

The AER protocol communicates spikes from one chip, the transmitter, to another, the receiver. The transmitter signals the occurrence of a spike in the array by placing an address on the bus that uniquely identifies the neuron that spiked. Each neuron is assigned a unique X (column)
and Y (row) address pair. The time that the address appears on the bus encodes the spike time directly. As illustrated at the top of Figure 3, spike addresses are represented using a word serial format consisting of bursts of addresses, each of which encodes the locations of all simultaneously spiking neurons within one row. The first address within each burst corresponds to a chip address, identifying the chip that the spikes originate from. The next address identifies the row. Subsequent addresses identify the columns containing spiking neurons. Simultaneous spikes from different rows are handled by an arbiter that arranges them into sequential bursts. Three handshaking signals (ReqY, ~ReqX, and Ack) ensure that addresses are communicated correctly. The transmitter takes ReqY high and low to signal the start and the end of a burst. It takes ~ReqX low to signal the row and the subsequent column addresses. The receiver makes a transition on Ack to acknowledge every transition on either of the two request lines.

For the Gabor-type chip, each retinal position has four associated neurons, which are arranged in $2 \times 2$ blocks. ON and OFF neurons are in-
terleaved in alternating columns. EVEN and ODD neurons are located in alternating rows. Thus, the least significant bit (LSB) of the X address encodes ON-OFF, and the LSB of the Y address encodes EVEN-ODD.

As shown in Figure 3, the address filter extracts only those spikes corresponding to the four neurons at a desired retinal position and demultiplexes these spikes as voltage pulses on four separate wires. As addresses appear on the AER bus, the filter generates the Ack signal that is sent back to the Gabor-type chip. Two latches latch the row and column addresses of each spike, which are compared with the row and column addresses of the desired retinal location. In our addressing scheme, the retinal location is encoded on bits 1 to 6 of the address. Once the decoder detects a spike from the desired retinal location, it generates a voltage pulse, which is demultiplexed onto one of four output lines depending on the LSBs of the latched row and column addresses.

To avoid losing events, we implement the address filter on a Xilinx XC9500 series CPLD to minimize the time that the address filter requires to process each spike. We chose this series because of its speed and flexibility. The block delay in each macrocell is only 7 ns. The series supports in system programming enable rapid debugging during system design.

Because the AER protocol is asynchronous, we needed to pay particular attention to the timing in the signal path to ensure that addresses are latched correctly and to avoid glitches that could be interpreted as output spikes. First, we connected the ~ReqX signal from the Gabor-type chips that signals the validity of both row and column addresses to an ungated global clock pin (GCLK) to generate the signal that latches the row and column addresses. Because the GCLK pin is routed to every logic block directly, this ensures that the flip-flops latch the address bits simultaneously. Second, rather than using automatic place and route, we manually equalized the path delays in the address comparison circuits and in the 1 to 4 decoder. This was essential since the routing delays in the switching matrix can be on the same order as the block delays.

3.3 Binocular Combination Block. The binocular combination block combines the eight monocular spike trains (four from each eye) to compute two-phase quadrature binocular filter outputs. For a zero-disparity-tuned cell, whose required routing is shown in Figure 2A, we set the AER address filters so that they extract spikes from monocular neurons corresponding to the same retinal location in the left and right eyes ($\Delta c = 0$). To compute the output of the first binocular filter B1, the binocular combination block sums the outputs of the left and right eye EVEN filters by merging the spikes from the e− channels onto a positive output line, B1+, and the spikes from the e channels onto a negative output line, B1−. Although the difference between the spike rates on B1+ and B1− encodes the B1 filter output, the B1+ and B1− spike rates do not represent the ON and OFF components of B1 since they may be positive simultaneously. To compute the output of the second
filter, B2, the binocular combination block merges the spikes from the ODD channels similarly.

We can electronically reconfigure the system to construct the binocular filter outputs required for neurons tuned to nonzero disparities. For position encoding, we change the relative addresses selected by the AER address filters to set \( \Delta c \neq 0 \), but leave the binocular combination block unchanged. For phase encoding, we leave the AER address filters unchanged and alter the routing in the binocular combination block. Because the RF profiles of the Gabor-type chips have two phase values, altering the routing can yield four distinct binocular filters with monocular filter phase shifts of \( \Delta \phi = -\pi/2, 0, \pi/2, \) and \( \pi \). Neurons with phase shifts of \(-\pi/2\), whose required routing is shown in Figure 2B, are tuned to positive disparities. Neurons with phase shifts of \( \pi/2 \), which can be obtained by swapping the routing of the ON and OFF signals from one eye in Figure 2B, are tuned to negative disparities. Neurons with phase shifts of \( \pi \), which can be obtained by swapping the routing of the ON and OFF signals from one eye in Figure 2A, are tuned inhibitory complex cells whose outputs are minimized by zero-disparity stimuli.

We implement the binocular combination block using a Xilinx XC9500 series CPLD. Inputs control the monocular phase shift of the resulting binocular filter by modifying the routing. For simplicity, we merge spikes using inclusive OR gates without arbitration. Although simultaneous spikes on the left and right channels could be merged into a single spike, the probability that this will happen is negligible, since the width of the voltage pulse representing each spike (~32 ns) is much smaller than the interspike intervals, which are on the order of milliseconds.

4 Results

This section describes the measured responses from the binocular complex cell implementation described in the previous section, the results of a sensitivity analysis of the responses to mismatch in the parameters of the monocular RFs, and a characterization of the monocular RF mismatch in our system.

4.1 Binocular Receptive Field Maps-Technique. Figure 4 shows our experimental setup for measuring the binocular RFs of the complex cells. Both Gabor-type chips were tuned for vertical orientations with similar RF parameters, as listed in Table 1. The RF parameters were estimated by fitting the monocular Gabor-type chip responses to equation 3.1.

We presented each retina with a 2.4 cm wide vertical dark bar on a bright background using two LCD monitors placed 55 cm from the retinas. This arrangement meant that the bar spanned four adjacent retinal positions, corresponding to about one-third of the period of the sinusoidal modulation in the RF profile and two-thirds of the width of the modulating envelope,
A Preference for Phase-Based Disparity

Figure 4: The experimental setup for measuring the binocular RF maps. The monitors were moved closer to the retinas to enable more detail of the hardware system to be shown. The insert board converts the AER address format used by the silicon retina to that used by the Gabor-type chips by inserting a chip address into each burst.

Table 1: Tunings for Two Gabor-Type Chips on the Experimental Setup.

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>2152</td>
<td>2265</td>
</tr>
<tr>
<td>$\Delta \Omega$</td>
<td>0.3059</td>
<td>0.2494</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>0.5226</td>
<td>0.5748</td>
</tr>
</tbody>
</table>

which is defined as $2(\Delta \Omega)^{-1}$. Using two monitors gives us precise control over the input disparity. We stimulated $31 \times 31$ pairs of left and right stimulus positions. The visual angle between two adjacent stimulus positions was 0.935 degree.

We characterized the responses of five binocular complex cells: one cell tuned to zero disparity and two pairs of cells tuned to positive and negative disparities by phase or position encoding. We obtained the zero-disparity-tuned cell and the phase-encoded cells by combining the spikes from pixels with (row, column) address of (15, 24) in the left and right chips. For the position-encoded cells, we combined the spikes from pixel (15, 24) in the right chip with the spikes from pixels (15, 20) and (15, 28) in the left chip.

4.2 Binocular Receptive Field Maps—Results. Figure 5 shows the binocular RF maps measured as described in the previous section. The map
for the zero-disparity-tuned neuron (Figure 5E) shows the expected peak in the center of the map. Maps for neurons tuned to positive (Figures 5C and 5F) or negative (Figures 5A and 5D) disparities show this peak shifted above or below the main diagonal. For position encoding (Figures 5D and

Figure 5: Binocular RF maps shown as images where the gray level of pixel encodes the response of the complex cell to a pair of spatial bar stimuli applied to the left and right Gabor-type chips. Darker gray levels indicate higher firing rates. (A, D) The RF maps of neurons tuned to negative disparities by phase and position encoding. (B) The regions of the maps corresponding to different disparities. (C, F) The RF maps of neurons tuned to positive disparities by phase and position encoding. (E) The map of the neuron tuned to zero disparity, which is the same for both position and phase encoding.
5F), the peaks are qualitatively similar to the peak of the zero-disparity neurons, only shifted. The responses of the neurons tuned by phase encoding (Figures 5A and 5C) are qualitatively different from the response of the zero-disparity-tuned neuron, exhibiting suppression in their response to the opposite disparity.

4.3 Disparity Tuning Curves. Figure 6 shows the disparity tuning curves for complex cells tuned to negative, zero, and positive disparities by phase and position encoding, which we computed by integrating the responses in Figure 5 along lines of constant disparity. We also computed
the ideal disparity tuning curves by assuming that the monocular responses to inputs at a position $x$ had the form

$$m(c, \phi, x) = \frac{\Delta \Omega}{2} \exp\left(-\Delta \Omega |x|\right) \cos(\Omega (x - c) + \phi),$$

where the parameters $\kappa$, $\Omega$, and $\Delta \Omega$ were obtained by least-squares fit of the expression to the measured monocular responses from one eye, and $\phi$ was either 0 or $\pi/2$.

Although we do observe the expected shifts in locations of the peak responses for different disparity tunings, the relative sizes of the peaks for the phase-encoded neurons appear to match the theoretical predictions of the disparity energy model better than those of the position-encoded neurons. The model predicts that the disparity tuning curves for the position-encoded neurons should be identical in size and shape, just shifted to the left or the right. However, the measured responses show three peaks with varying heights.

These deviations between expected and measured responses are due primarily to mismatch between the monocular RF profiles of the Gabor-type chips. Because subsequent processing is implemented in digital hardware, there is no variability introduced by these stages, except due to quantization, which is negligible here. The disparity energy model generally assumes that the left and right monocular RFs have the same gain and spatial frequency tunings. However, random transistor mismatch within the Gabor-type chips causes the gain and spatial frequency tuning to vary from neuron to neuron.

The presence of this mismatch and its observed effect on the relative responses of neurons tuned to different disparities raises a question: Which of the two encoding schemes is more robust in the presence of these variations? As described below, we addressed this question in two parts. First, how does the variation in the parameters controlling the monocular RF profiles affect the relative outputs of complex cells? Second, how much do the RF parameters vary from neuron to neuron on the chip?

### 4.4 Sensitivity Analysis.

We model the mismatch by assuming the responses of the monocular filters have the same parameterized form, but that the parameters vary from pixel to pixel. For simplicity, we assume one-dimensional input images consisting of a single spatial impulse and model the response of the four neurons at the origin of retinal coordinates to an impulse at pixel $x$ by

\[
\begin{align*}
    h_{e+}(x) &= \kappa_{e+}[h_e(x) - \xi_c]^+ \\
    h_{e-}(x) &= \kappa_{e-}[h_e(x) - \xi_c]^-
\end{align*}
\]

\[
\begin{align*}
    h_{o+}(x) &= \kappa_{o+}[h_o(x) - \xi_o]^+ \\
    h_{o-}(x) &= \kappa_{o-}[h_o(x) - \xi_o]^-
\end{align*}
\]
where
\[ h_l(x) = \frac{\Delta \Omega}{2} e^{-\Delta \Omega |x|} \cos(\Delta \Omega x + \phi_l) \]
and
\[ h_o(x) = \frac{\Delta \Omega}{2} e^{-\Delta \Omega |x|} \sin(\Delta \Omega x + \phi_o). \]

We examined the effect of mismatch in five RF parameters: the spatial frequency \( \Omega \), the half-bandwidth \( \Delta \Omega \), the offsets \( (\xi_l, \xi_o) \), the phase offsets \( (\phi_l, \phi_o) \), and the gains \( (\kappa_{l+}, \kappa_{l-}, \kappa_{o+}, \kappa_{o-}) \). Mismatch in the analog transistors implementing the Gabor-type filters introduces mismatch in the spatial frequency, half-bandwidth, phases, and offsets. Mismatch in the transistors of the input integrators and output spiking neurons introduces gain mismatch. We show in the next section that this mathematical model captures approximately 90% of the RF variability on the Gabor-type chips.

To examine the robustness of the relative responses, we need to establish metrics on the quality of the relative responses. We concentrate on the relative responses of populations of neurons tuned to different disparities but similar frontoparallel positions since we can compare these to estimate disparity. Perhaps the simplest method classifies a stimulus into three categories depending on which neuron has the largest response, as shown in Figure 7. In this case, we are interested in how monocular RF mismatch affects the locations of decision boundaries \( L \) and \( R \). Although this is only one possible method for estimating disparity, other methods should display a similar dependence on the parameter mismatch.

We examined how much these boundaries shifted in response to monocular RF parameter mismatch by perturbing the RF parameters in the left and right neurons by additive gaussian offsets with zero mean and varying percentage deviations. We used the mean of the left and right RF parameters listed in Table 1 as the nominal parameters. Deviations in the offsets, which are nominally zero, were expressed as a percentage of the peak value of the EVEN filter impulse response. Similarly, the deviations in the phase offsets

![Figure 7: Decision boundaries obtained by comparing relative responses of neurons tuned to negative, zero, and positive disparities. \( L \) and \( R \) denote boundaries between negative/zero and zero/positive disparity stimuli. The parameters \( D = R - L \) and \( C = (L + R)/2 \) describe the distance between and the centroid of the decision boundaries.](http://www.mitpressjournals.org/doi/pdfplus/10.1162/089976604774201604)
were expressed as a percentage of the maximum phase shift before phase wraparound, \( \pi \). We found that in general, variations in \( L \) and \( R \) were coupled, but that the variations in distance \( D = R - L \) between the boundaries and their centroid \( C = (L + R)/2 \) were approximately independent.

Deviations in the decision boundary parameters \( D \) and \( C \) are expressed as a percentage of the nominal value of \( D \). Thus, a 100% shift in \( C \) indicates that the centroid of the left and right boundaries has shifted so much that the original and the perturbed regions corresponding to fixation share no overlap if \( D \) remains constant.

The variations in the decision boundary parameters increase linearly for small RF parameter deviations. We define the sensitivity of each decision boundary parameter in response to variations in each RF parameter as the slope of this line, which we estimate by linear least-squares fitting. If the sensitivity is less than one, we say that the decision boundary parameter is robust to variations in the RF parameter.

Figure 8 compares the sensitivities of the position and phase models. Except for one combination (the sensitivity of \( D \) to variations in \( \Omega_1 \)), the phase model is more robust than the position model. In addition, the sensitivity of neurons tuned by phase encoding is always less than one, indicating that it is robust to the variation in the monocular RF profiles. The distance \( D \) is more sensitive to variations than the centroid \( C \).

### 4.5 Intrachip Variability of RF Parameters

The previous section showed that in general, the phase model is more robust than the position model. However, since there are RF parameters for which phase encoding is more sensitive than position encoding, it is important to establish how much each RF parameter varies within one chip. Although transistor parameters will vary from chip to chip, the effect of this interchip variation can be minimized by adjusting the external bias voltages appropriately. However, intrachip variability makes it impossible to match the responses of all pairs of neurons on two different chips and is therefore the limiting factor.

We characterized the variability by measuring the variability of RF parameters obtained by least-squares fits to the RF profiles measured at seven neurons located in row 16 and columns 26, 28, 30, 32, 34, 36, and 38 of one chip. We chose a restricted range of columns and rows to avoid the edge effects due to the chip boundaries. The least-squares fits of the RF profiles of these neurons are illustrated in Figure 9.

We measured the quality of the fits by the percentage accuracy

\[
A_p\% = \left( 1 - \frac{\sum_j (h_e[j,k] - \hat{h}_e[k])^2 + (h_o[k] - \hat{h}_o[k])^2}{\sum_j (\hat{h}_e[k]^2 + \hat{h}_o[k]^2)} \right) \times 100\%,
\]

where

\[
h_e[k] = \sum_j (h_{e+}[j,k] - h_{e-}[j,k]) \quad \text{and} \quad h_o[k] = \sum_j (h_{o+}[j,k] - h_{o-}[j,k])
\]
Figure 8: The sensitivity of position and phase encoding. (A) Sensitivity of the distance $D$ between the decision boundaries to mismatch in different RF parameters. (B) Sensitivity of the centroid $C$.

are the measured one-dimensional EVEN and ODD RF profiles of the neurons, their fitted RF profiles are $\hat{h}_e[k]$ and $\hat{h}_o[k]$, and the indices $j$ and $k$ correspond to the row and column position. The fits shown in Figure 9 closely matched the measured profiles. The percentage accuracy for about half of the neurons was over 90% (the average value was 88.37%). It turns out that the phase variability is quite small. Fixing the phase offsets $\phi_e$ and $\phi_o$ to be zero only degraded the quality of the fits by 1.2%.

The percentage standard deviations are plotted in Figure 10. The filter gains vary the most, followed by the half-bandwidth. The spatial frequency, phase, and offset vary the least. This figure also shows that there is a complementary relationship between the sensitivity of the distance between the decision boundaries, $D$, for phase-encoded neurons and the RF parameter variability. Although the gains vary the most, the decision boundaries are not very sensitive to their variations. On the other hand, the decision bound-
Figure 9: The least-squares fit of measured Gabor-type RF profiles. (A–G) The fits for the EVEN and ODD profiles corresponding to the seven neurons located in row 16 and columns 26, 28, 30, 32, 34, 36, and 38 of a chip. The discrete data points, +, represent the measured responses. The solid lines represent the fits.
Figure 10: Comparison between the sensitivity of the distance $D$ between decision boundaries using phase-encoded neurons to variations in different RF parameters and the measured variation of the parameters in the system.

aries are relatively more sensitive to variations in the spatial frequency and offset, but these do not vary much on the chip.

5 Conclusion

We implemented a neuromorphic multichip model of disparity-tuned complex cells in the mammalian primary visual cortex based on the disparity energy model. We evaluated two possible mechanisms for disparity tuning, phase, and position offsets between the monocular RF profiles and demonstrated that our system prefers phase encoding, because the relative responses of neurons tuned to different disparities are less sensitive to mismatch in the monocular RF profiles. We found that the phase model is least sensitive to the RF parameters that vary the most in our system.

Biological systems would also benefit from this type of robustness, as they must also extract information from the responses of a population of neurons. Intuitively, the better robustness of our phase-encoded neurons arises because binocular disparity-selective neurons tuned to different disparities are constructed from the same sets of monocular outputs but in different combinations. Thus, different binocular neurons are subject to the same variations. On the other hand, our position-encoded neurons are constructed from different sets of monocular outputs. The biological analog of the large variability in the gain of RF profiles we measured may be the large variation in the firing rate of real neurons. Thus, our result implies that phase encoding is one mechanism that enables real cells to deal with firing rate variability. Our system architecture is similar to a recent cortical
model of stereopsis, which proposes that monocular neural outputs from layer 4 of V1 are combined binocularly in layer 3B (Grossberg & Howe, 2003). Interestingly, the population response of phase-encoded neurons is also more robust in another sense: the peak of the population response is an accurate estimate of the disparity regardless of the detailed luminance profile of the stimulus, while this is not true for the population responses of position-encoded neurons (Chen & Qian, 2004).

This implementation is an initial step toward the development of a multichip neuromorphic system capable of extracting depth information about the visual environment using silicon neurons with physiologically based functionality. Currently, the system can only compute the output of a single cell tuned to a particular disparity and retinal position. However, the estimation of disparities in a visual scene requires a population of neurons tuned to different disparities and retinal positions. Although we could accomplish this by exploiting the electronic reconfigurability of the system using time multiplexing, this approach is inefficient and time-consuming. Our current work seeks to implement this population in a fully parallel manner. By combining the AER outputs from the left and right Gabor-type chips onto a chip containing a two-dimensional array of neurons with the appropriate address remapping, we can implement an array of neurons tuned to the same disparity but varying retinal locations. Additional chips could represent neurons tuned to other disparities. Computing the neuronal outputs in parallel will enable us to investigate the roles of additional processing steps such as pooling (Fleet, Wagner, & Heeger, 1996; Qian & Zhu, 1997) and normalization (Albrecht & Geisler 1991; Heeger, 1992).

Acknowledgments

This work was supported in part by the Hong Kong Research Grants Council under grants HKUST6218/01E and HKUST6200/03E. It was inspired by an initial project at the 2002 Telluride Neuromorphic Workshop performed with Y. Miyawaki. We thank K. A. Boahen for supplying the silicon retina and AER receiver chips used in this work and for helpful discussions and T. Choi for his assistance in building the system.

References


Received August 18, 2003; accepted February 2, 2004.