Retinomorphic Chips that see Quadruple Images

Kwabena Boahen
Penn Bioengineering
3320 Smith Walk
Philadelphia PA 19104-6392 USA
kwabena@neuroengineering.upenn.edu

Abstract

Retinomorphic chips may improve their spike-coding efficiency by emulating the primate retina's parallel pathways. To this end, I recreated retinal microcircuits in a chip, Visio1*, that models the four predominant ganglion-cell types. It has* 104×96 *photoreceptors,* $4 \times 52 \times 48$ *ganglion-cells, a die size of* $9.25 \times 9.67\text{mm}^2$ *in* 1.2μm 5V CMOS*, and consumes* 11.5mW *at* 5 *spikes/second/neuron.* Visio1 *includes novel subthreshold current-mode circuits that use horizontal-cell autofeedback to decouple spatiotemporal bandpass filtering from local gain control and use amacrine-cell loop-gain modulation to adapt highpass and lowpass temporal filtering. Different ganglion cells respond to motion in a stereotyped sequence, making it possible to detect edges of one contrast or the other moving in one direction or the other. I present results from a multichip 2-D motion architecture, which implements Watson and Ahumada's model of human visual-motion sensing.*

1. Parallel Pathways in the Retina

The presence of visual pathways specialized for spatial and temporal resolution—called parvocellular and magnocellular pathways in primates—has been confirmed both physiologically and anatomically [27, 23, 17]. Neurons in these pathways pool signals over either space or time to average out quantum fluctuations, maintaining the same noise level as they trade poor resolution in one domain for good resolution in the other [15]. As the light intensity drops, pooling occurs over larger distances or longer times to maintain the signal-to-noise ratio [6].

There is a continuum of spatial and temporal resolutions within each pathway, however, due to the variation of spatiotemporal characteristics of midget and parasol retinal ganglion cells with eccentricity [14]. They range from small and sustained in the fovea, where fine details of an object stabilized by tracking are resolved, to large and transient in the periphery, where sudden motion in the surroundings is captured. At a given eccentricity, parasols (also called α cells) cover two to three times longer distances and respond more transiently than midgets (also called β cells) [38, 14].

In terms of actual numbers and sampling densities, midgets and parasols make up 90% of the total ganglion cell population and occur in a ratio of about 9:1 [34]. Nine times fewer parasols are required to tile the retina because their dendritic fields are three times larger. The remaining 10% of the cells form a heterogeneous group and project mainly to the midbrain [34].

Activity in each pathway is encoded by a pair of complementary channels, served by ON- and OFF-midgets or by ON- and OFF-parasols. The ON channel signals increases in amplitude by increasing vesicle-release or spike-discharge rates; the OFF channel signals decreases in amplitude in a similar fashion [26]. Complementary signaling overcomes three shortcomings of using a single channel to transmit both increases and decreases:[1]

- Elevated spike-discharge rates and vesicle-release rates must be maintained in the quiescent state.
- Decreases are transmitted with lower fidelity, because quantum fluctuations (i.e., shot noise) decrease only as the square root of the mean rate.
- Decreases are transmitted with lower speed, because quanta are infrequent and membrane repolarization and transmitter removal are passive.

Except prior to the very first synapses, found in rod and cone terminals, complementary signaling is used throughout the retina to transmit information efficiently using vesicles or spikes.

[1] The rod pathway does fine with a single channel because its baseline activity is virtually zero.

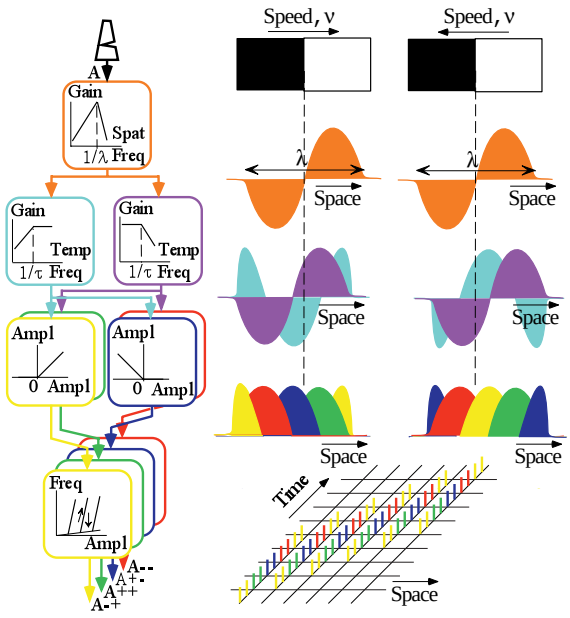


Figure 1. Retinal Operations

The visual signal is bandpass filtered spatially, then lowpass and highpass filtered temporally (first panel, numbering clockwise). These analog signals are half-wave rectified, then quantized. Spatial filters adapt to light intensity—their gain is inversely proportional. Temporal filters adapt to stimulus speed—their time-constant is inversely proportional (i.e., $\tau = \lambda/\nu$, where λ is the wavelength selected by the spatial filter). And spiking neurons adapt to their input's rate of change—their firng rate is linearly proportional. Four channels provide a quadrature representation: increasing (-+ or yellow), positive (++ or green), decreasing (+- or blue), and negative (-- or red), as demonstrated by a moving edge (second panel). Reversing the direction reverses the temporal contrast, interchanging the yellow and blue responses (third panel). Responses occur in the same order in time, at a given location, as they do in space, at a given time (fourth panel).

In addition to reducing quiescent firing rates by using complementary channels, I postulate that further savings in spikes and vesicles are achieved by adapting the baseline—both at the ganglion-cell level and at the network-level. At the ganglion-cell level, the baseline is set by calcium-dependent pottasium channels, which shunt the input current [7, 9]. At the network-level, it is set by presynaptic inhibition from amacrine cells, which terminate vesicle release at ON and OFF bipolar-cell terminals [30]. Behaving like leaky integrators, these amacrine cells follow the signal's temporal average. To ensure signal fluctuations are not masked by a dynamic baseline, a long time-scale average must be computed when change is slow. And, to ensure the signal crosses the baseline when its derivative changes sign, a short time-scale average must be computed when change is rapid. Thus, spikes and vesicles are conserved, as frequent quanta produced by fast signals are discharged for short durations while infrequent quanta produced by slow signals are discharged for long durations.

Retinomorphic chips, which perform adaptive pixel-parallel quantization [7, 9], may improve their spike-coding efficiency by emulating parallel pathways in the retina. To this end, I have recreated retinal microcircuits serving the magnocellular and parvocellular pathways in a chip, Visio1, that models the four predominant ganglion cell types. Visio1 performs the operations shown in Figure 1 at the pixel-level. Anatomically identified neural microcircuits that perform these operations and their CMOS neuromorphs are described in the Section 2 and 3. These current-mode circuits operate in the subthreshold region [31], where small-signal conductances and transconductances are proportional to the current level [2], current-spreading diffusor networks are linear [12, 37, 4], and the generalized translinear principle holds [5]. Visio1's design and performance are described in Section 4, and its application in a multichip, real-time, 2-D motion-sensing system is described in Section 5. Section 6 concludes the paper.

2. Outer Retina Model

The outer retina performs spatiotemporal bandpass filtering and adapts its gain locally—both at the receptor level [16, 29] and at the network level. My previous attempts to make these two functions coexist in one network produced undesirable side-effects [12, 3]. In particular, while attempting to attenuate redundant low-frequency temporal and spatial signals, I found that the high loop-gain required in a negative-feedback circuit resulted in temporal instability. And, while attempting to extend dynamic range, I found that controlling sensitivity by changing the cone–to–cone conductance caused the receptive field to expand alarmingly.

The shortcomings of my high-gain negative-feedback outer retina circuit model inspired me to search for a retinal mechanism that could *decouple* spatiotemporal filtering, local gain control, and temporal stability.

Horizontal-cell autofeedback, which was demonstrated by Kamermans and Werblin a few years ago in the tiger salamander [22], is a promising neural mechanism. Horizontal cells, which are known to use the inhibitory neurotransmitter GABA, also express GABA-gated Cl-channels. These channels have a reversal potential of -20mV and therefore depolarize the

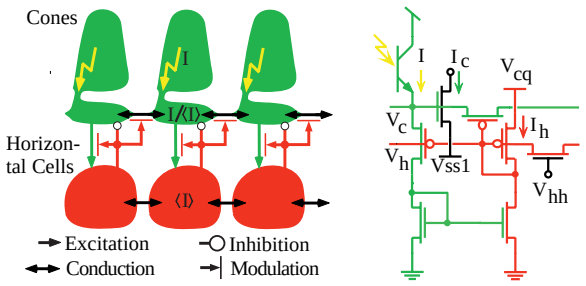


Figure 2. Morphing the Outer Retina

(a) Cones receive a photocurrent that is proportional to incident light intensity, I, from their outer segments. Horizontal cells, whose activity is proportional to the local spatiotemporal average light intensity, $\langle I \rangle$, modulate gap junctions between cones and modulate their membrane conductances as well by making inhibitory synapses. They also modulate the excitatory synapses they receive from cones, giving rise to a positive-feedback loop. As a result, the cone's activity becomes proportional to spatiotemporal contrast (i.e., $I/\langle I \rangle$). (b) A pMOS transistor, with its source tied to V_c and its gate tied to V_h, produces a current proportional to the product of the cone and horizontal cell activities, which are represented by the currents I_c and I_h. This current is shunted from the cone node, V_c, to model horizontal-cell inhibition, and dumped on the horizontal-cell node, V_h, to model cone excitation and horizontal-cell autofeedback.

	Attenuation	Q − Factor
Neg	$\frac{c_{c0} I_c}{c_{c0} I_c + c_{h0} I_h}$	$\frac{1}{\epsilon_E}\left(\sqrt{\frac{c_{c0} I_c}{c_{h0} I_h}} + \sqrt{\frac{c_{h0} I_h}{c_{c0} I_c}}\right)^{-1}$
Pos	0	$\sqrt{\frac{c_{c0}}{c_{h0}}}$
	Sensitivity	**Space − Constant**
Neg	$\alpha_{hh}^{1/4}(\alpha_{cc} I_h)^{-3/4}$	$(\alpha_{hh}\alpha_{cc} I_h)^{1/4}$
Pos	$\alpha_{hh}^{1/4}\alpha_{cc}^{-3/4}(I_c/I_h)$	$(\alpha_{hh}\alpha_{cc})^{1/4}$

Table 1. Negative versus Positive Feedback

I_c is the mean cone activity and I_h is the mean horizontal cell activity, which is proportional to light intensity. Ideally, the attenuation, a measure of low-frequency signals' relative amplitude, should be zero; the Q-factor, a measure of temporal stability, should be less than one; the sensitivity, a measure of photosignal amplification, should be inversely proportional to I_h; and the space constant, a measure of receptive field size, should be independent of I_h.

cell when they are opened, forming a positive-feedback loop. Kamermans and Werblin showed that this autofeedback loop accounted for the extremely slow dynamics of horizontal cells, increasing their time constant from 65ms to 500ms. My analysis of the tradeoffs involved in outer-retina design has yielded two new hypotheses about the role of autofeedback [8].

Horizontal-cell autofeedback can improve temporal stability by amplifying the cone signal, allowing us to decrease the strength of the cone–to–horizontal-cell synapse. Thus, we can attenuate low-frequency signals while maintaining temporal stability. A lower cone–to–horizontal-cell synaptic transconductance extends the cone's dynamic range as well.

Autofeedback can also make receptive-field size independent of sensitivity by modulating the effective strength of the cone–to–horizontal-cell synapse. More activity in the horizontal cells provides a larger boost to synaptic input from cones. Therefore, if horizontal cell activity is proportional to intensity, then the cone–to–horizontal-cell transconductance becomes proportional to intensity as well, compensating for the increase in the cone–to–cone conductance with intensity.

A novel current-mode CMOS circuit that uses horizontal-cell autofeedback to decouple spatiotemporal filtering, local gain control, and temporal stability—and the hypothetical neural microcircuit it is based on—is shown in Figure 2. Its small-signal (i.e., linear) behavior is described by the equations:

$$\begin{aligned} i_0 + \frac{\alpha_{cc} I_h}{I_c}\nabla^2 i_c &= i_h + \frac{c_{c0} U_T}{I_c}\frac{di_c}{dt} + \frac{I_h}{I_c} i_c \\ \frac{I_h}{I_c} i_c + \alpha_{hh}\nabla^2 i_h &= \frac{c_{h0} U_T}{I_h}\frac{di_h}{dt}, \end{aligned}$$

in the continuum limit. Currents represent signals instead of voltages, as this is a current-mode circuit. The subscripts c and h denote cone and horizontal-cell; a pair of letters denotes coupling between cells or to ground, which is denoted by 0. Upper case symbols represent the mean signal value, while lower case ones represent small instantaneous deviations from the mean. Thus, it is evident how the absolute signal levels modulate the transistors' small-signal conductances and transconductances, which model membrane conductances, gap junctions, and chemical synapses. α_{cc} and α_{hh} are determined by the relative size of the transistors coupling the cone nodes together and the bias voltage applied to the transistors coupling the horizontal-cell nodes together. $U_T \equiv kT/q$ is the thermal voltage.

In contrast, my earlier design [12, 3], which relied on high-gain negative feedback, is described by the equations:

$$i_0 + \alpha_{cc}\nabla^2 i_c = i_h + \frac{c_{c0} U_T}{I_c}\frac{di_c}{dt} + \epsilon_E \frac{I_h}{I_c} i_c$$

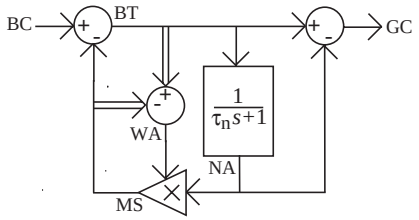


Figure 3. Adapting Temporal Dynamics

The narrow-field amacrine cell (NA) feeds inhibition back onto the bipolar-cell terminal (BT) and forward onto the ganglion cell (GC). Feedback inhibition is modulated (MS) by a wide-field amacrine cell (WA). WA is excited by BTs and inhibited by NAs in both ON and OFF pathways (double lines). It pools these signals over a large area and equalizes their average amplitudes by adjusting the loop gain. If the BC signal changes too rapidly, NA inhibition falls and BT excitation increases, causing WA to turn up the gain and boost NA inhibition. Conversely, if it changes too slowly, NA inhibition increases and BT excitation decreases, causing WA to turn down the gain and cut NA inhibition.

$$i_c + \alpha_{hh}\nabla^2 i_h \quad = \quad \frac{c_{h0}U_T}{I_h}\frac{di_h}{dt} + \epsilon_{\mathrm{E}}\frac{I_c}{I_h}i_h,$$

where $\epsilon_{\mathrm{E}} \equiv U_T/V_{\mathrm{E}}$ is the ratio between the thermal voltage (25mV at room temperature) and the Early voltage (typically about 25V). In comparison, the horizontal-cell–to–cone synaptic gain, $A_{ch} \equiv di_c/di_h$, is a thousand (i.e., $1/\epsilon_{\mathrm{E}}$) times smaller in the new circuit, while the cone–to–horizontal-cell synaptic gain, $A_{hc} \equiv di_h/di_c$, is infinite—for zero temporal and spatial frequencies (i.e., DC)—due to 100% positive feedback.[2] For nonzero frequencies, the cone–to–horizontal-cell synaptic gain is proportional to I_{h} (and hence to light intensity) in the new circuit, due to the modulatory effect of horizontal-cell autofeedback. The intensity-dependencies of the characteristics of these two circuits are compared in Table 1.

3. Inner Retina Model

The inner retina performs lowpass and highpass temporal filtering and adapts its dynamics locally. Midget and parasol ganglion cells receive synaptic inputs from both bipolar and amacrine cells, but parasol cells receive more amacrine input (i.e., feedforward inhibition) [20, 13, 25], which accounts for their more transient response. Parasols also have larger dendritic fields and are driven by bipolar cells with larger dendritic and axonal fields than those that drive the midgets [33, 34], which accounts for their larger receptive fields. Presumably, both midget and parasol bipolar cells receive presynaptic amacrine input (i.e., feedback inhibition) at their terminals.

I postulate that the dynamics of the amacrine cell that feeds inhibition forward to ganglion cells and back to bipolar cells is adapted by a second amacrine cell, which modulates the feedback-loop gain. Figure 3 illustrates this novel mechanism for adapting temporal dynamics. The wide-field amacrine cell (WFA) is excited by ON and OFF bipolars and inhibited by ON and OFF narrow-field amacrine cells (NFA). All these synaptic interactions have been found in an anatomically identified amacrine cell type, called A19 [24]. These cells have thick dendrites, a large axodendritic field, and are coupled together by gap junctions. Hence, they can integrate and distribute signals rapidly over a large area.

The NFA produces a lowpass-filtered version of the bipolar-cell signal while the bipolar terminal (BT) produces a highpass-filtered version—these filters have the same corner frequency, $1/\tau_A$. A frequency-domain small-signal analysis, using Laplace transforms, yields the following responses:

$$i_{\mathrm{n}} = \frac{\epsilon}{\tau_A s + 1} i_{\mathrm{b}},\ i_{\mathrm{t}} = \frac{\tau_A s + \epsilon}{\tau_A s + 1} i_{\mathrm{b}},\ i_{\mathrm{g}} = \frac{\tau_A s}{\tau_A s + 1} i_{\mathrm{b}};$$

for the NFA, the BT, and the parasol ganglion cell, respectively. Where,

$$\tau_A \equiv \epsilon\tau_{\mathrm{n}},\ \epsilon \equiv 1/(I_{\mathrm{w}}/I_{\mathrm{w0}} + 1),$$

τ_{n} is the time-constant of the NFA cell and $I_{\mathrm{w}}/I_{\mathrm{w0}}$ is the modulation of its response by the WFA cell, whose response is I_{w0} for zero frequency (i.e., static input). Hence, the corner frequency is proportional to the WFA cell's response, through the dependence of the closed-loop time-constant on the modulation level.

Feedforward inhibition produces a purely transient response in parasol cells, whereas midgets, which do not receive feedforward inhibition, have a sustained component. This residual activity, $\epsilon(I_{\mathrm{w}})i_{\mathrm{b}}$, increases as the loop gain is reduced to lower the corner frequency—and thus the BT response asymptotically approaches an allpass filter. However, the NFA cell's residual grows in the same way, and cancels out the BT's residual in the ganglion cell, irrespective of the gain setting. Thus, a purely highpass response is achieved—this is impossible with a finite-gain negative feedback loop.

The WFA cell centers the corner frequencies of the highpass and lowpass filters on the input spectrum. Setting gains of the BT–to–WFA and NFA–to–WFA synapses equal and normalizing them to unity, and neglecting the leakage conductance, the WFA's activity

[2] These synaptic gains are obtained by setting the temporal and spatial derivatives equal to zero and differentiating the first and second equation, respectively.

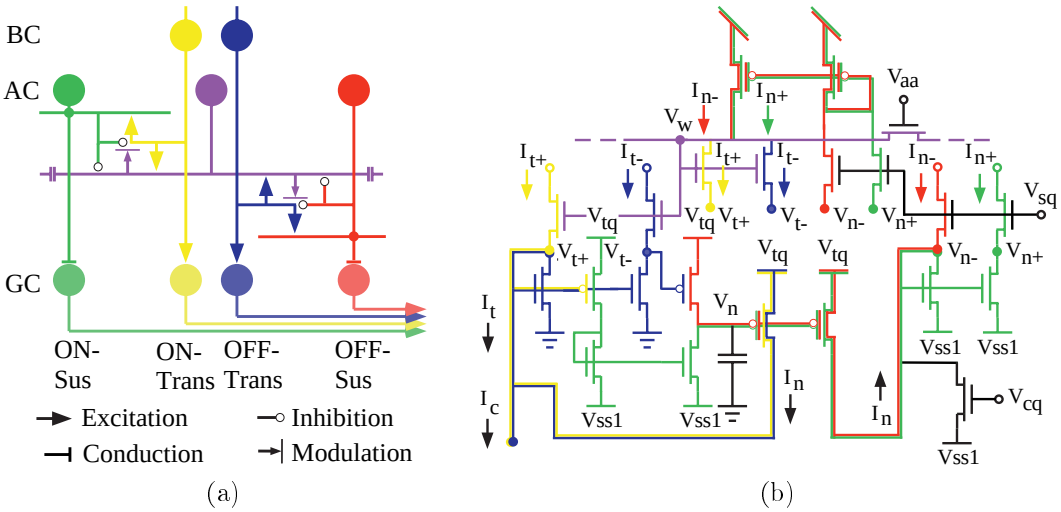


Figure 4. Morphing the Inner Retina

(a) Bipolar cells (BC) excite both narrow- and wide-field amacrine cells (AC) and excite ganglion cells (GC) as well. Narrow-field amacrines (NFA) inhibit BCs and wide-field amacrines (WFA), and drive GCs as well. WFAs modulate the NFA–to-BC synapses. ON and OFF pathways are served by two sets of circuits, but they drive a single WFA-cell syncytium. (b) Unlike the neural circuit, complementary signaling is not used for BCs and NFAs and modulation occurs before lowpass filtering. The NFA signal, I_{n}, is subtracted from the unrectified cone signal, I_{c}, to obtain the BT signal I_{t}. Four-transistor rectifiers produce ON and OFF NFA ($I_{\mathrm{n}\pm}$) and BT ($I_{\mathrm{b}\pm}$) signals. Copies made by a transistor connected in parallel are sunk from or mirrored onto the WFA node (V_{w}) to excite or inhibit it, respectively. A four-transistor subtractor takes the difference between $I_{\mathrm{t}+}$ and $I_{\mathrm{t}-}$ and supplies it to the NFA node (V_{n})—V_{w} modulates the current level in the subtractor.

is given by

$$\alpha_{\mathrm{ww}}\nabla^2 i_{\mathrm{w}} + \updownarrow i_{\mathrm{t}} \updownarrow - \frac{I_{\mathrm{w}}}{I_{\mathrm{w0}}} \updownarrow i_{\mathrm{n}} \updownarrow = 0$$

where α_{ww} is the coupling strength between WFAs and $\updownarrow i \updownarrow$ is the full-wave–rectified version of i. For low spatial frequencies (i.e., on average), the Laplacian is close to zero, and we have

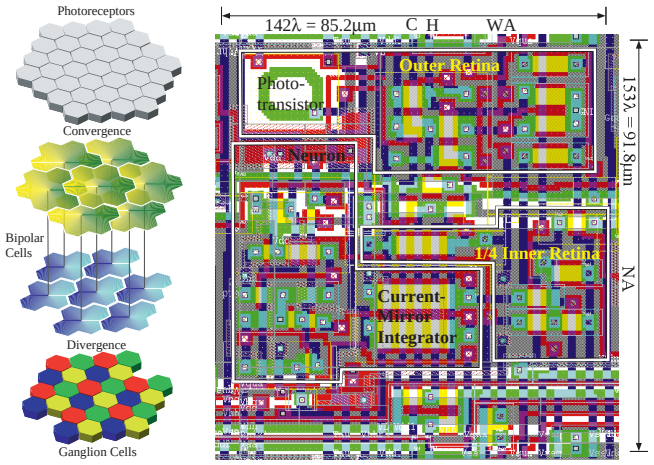
$$\frac{I_{\mathrm{w}}}{I_{\mathrm{w0}}} \approx \frac{\langle \updownarrow i_{\mathrm{t}} \updownarrow \rangle}{\langle \updownarrow i_{\mathrm{n}} \updownarrow \rangle} = \frac{|i_{\mathrm{t}}|}{|i_{\mathrm{n}}|} = \sqrt{\tau_{\mathrm{n}}^2\omega^2 + 1} \approx \tau_{\mathrm{n}}\omega,$$

assuming $\omega \gg 1/\tau_{\mathrm{n}}$; $\langle i \rangle$ is the local spatial average of i. Consequently, $\tau_A(\omega) \approx 1/\omega$—the corner frequency matches the input frequency. For the low-frequency extreme, $\omega \ll 1/\tau_{\mathrm{n}}$, we have, $I_{\mathrm{w}} \approx I_{\mathrm{w0}}$ and $\tau_A = \tau_{\mathrm{n}}/2$.

Temporal adaptation accounts for the variation of ganglion cells' temporal dynamics with eccentricity. Both midgets and parasols become more transient as they adapt to higher frequencies, and the midget's sustained component decreases. Hence, given that the input spectrum shifts to higher temporal frequencies with increasing eccentricity [18], the observed variation in temporal characteristics with eccentricity follows.

A novel current-mode CMOS circuit that implements retinomorphic temporal adaptation—and the hypothetical neural microcircuit it is based on—is shown in Figure 4. It does not include feedforward NFA inhibition onto transient GCs and feedforward BC excitation onto sustained GCs. Nevertheless, like the retina, it produces a highpass response by placing a lowpass filter in a negative feedback loop. Thus, corner frequencies of the highpass and lowpass responses are automatically matched. It compares their energy by full-wave rectifying, taking the difference, and integrating it over space using the WFA cell. Spatial integration is rapid since—for a moving stimulus—all phases of the response are available at different locations at the same instant. It is also robust due to the collective nature of the computation.

In contrast, a similar time-constant adaptation scheme proposed by Liu uses separate highpass and lowpass filters, integrates the difference between their *peak* responses over time, and adjusts their time-constants directly by changing the amplifiers' bias currents [28]. This scheme—implemented with a voltage-mode circuit—requires the amplifiers' transconductances to be matched and adapts on a time-scale several times the period of a response cycle. Since the edge moves on after a single response cycle, it is imperative that adaptation occur instantaneously and this information propagates quickly to neighboring cells. Both objectives are achieved by my retinomorphic circuit.



(a) (b)

Figure 5. Chip Architecture and Pixel Layout

(a) Bipolar cells, which connect 7 photoreceptors to 4 ganglion cells (one of each type), subsample the image by a factor of 4. A receptor either supplies its two identical output currents to one bipolar cell or to two neighboring cells. Starting with an $N \times N$ receptor array, we end up with $4 \times N/2 \times N/2$ ganglion-cell arrays. (b) The pixel has 39 MOS transistors: 8 in the outer retina, 6 in the inner retina, and 16 in the neuron, plus 9 transistors in three delta-structures (with a common T-shaped gate) that connect receptor (C), horizontal-cell (H), and wide-field-amacrine cell (WA) nodes to their six nearest neighbors. Unlike the schematic in Figure 2b, the phototransistor's current is mirrored first. The inner-retina circuit is spread across four pixels, with 6 transistors in two of them and 5 in the other two, for a total of 22. The subcircuit shown here produces the on-sustained output (i.e., I_{n+} in Figure 4b). In two of the pixels, a poly1-poly2 capacitor—attached to the narrow-field amacrine node (NA)—replaces the well-transistor above the current-mirror integrator. Pixels are tiled hexagonally by flipping those in every other column vertically.

4. Chip Design and Testing

I designed and fabricated a $4 \times 52 \times 48$ ganglion-cell chip, called Visio1, in a 1.2μm ($\lambda = 0.6\mu$m) double poly, double metal, n-well CMOS process (AMI's fab, available through MOSIS); it is $9.25 \times 9.67\text{mm}^2$. Visio1's architecture and pixel layout are described in Figure 5. It includes an asynchronous address-event transmitter interface, which reads out spikes generated inside the pixels [11]. Ganglion-cell type is determined by decoding the LSB's of row and column addresses. Visio1 was used in a real-time motion processing application, as described in Section 5.

Visio1 models cone–to–bipolar convergence, which makes the receptive-field center more Gaussian-like [35]. According to my simulations, the steeper frequency roll-off that results produces 60dB attenuation at 3.75 times the peak spatial frequency—compared to 32 times the peak without convergence. High spatial frequencies must be eliminated to preserve the signal–to–noise ratio after highpass temporal filtering, as these components produce proportionately high temporal frequencies when the stimulus moves. Further signal–to–noise enhancement may be realized by convergence at the ganglion-cell level, which I am yet to include—it would also reproduce the parasol cells' larger receptive fields.

A raster plot of spike trains recorded from neurons in a single column is shown in Figure 6. Unfortunately, crosstalk tended to make all the neurons fire when a certain activity level was exceeded, so I had to keep their mean firing rates extremely low—an order of magnitude less than in the real retina. Nevertheless, the sequence in which the four types fire is as predicted in Figure 1. There is considerable variability—especially among the sustained cells—due to transistor mismatch, which was systematic, producing striations oriented at about 30 degrees from vertical, like those described in [32]. Transient cells are more synchronized than sustained ones, indicating that their inputs are larger—temporal adaptation does not appear to be working as expected. Also, the second half of the response sequence is delayed—indicating either a refractory period (unlikely) or blurring in the outer retina, which occurs when its temporal bandwidth is exceeded.

Testing revealed a low-frequency oscillation in the inner retina circuit, which arose because the WFA cell does not modulate the loop-gain as intended. A MOS transistor's drain conductance and transconductance are both linearly proportional to its channel current in the subthreshold region, and hence their ratio is independent of the current level [2]. Therefore, the voltage gain and the normalized-current gain are fixed. That is, $i_n/I_n = A_E i_t/I_t$, where $A_E \equiv V_E/U_T$ is about one thousand. As the average current levels I_n and I_t do not change, modulation cannot change the loop gain—but it does change the time-constant directly by changing the drain conductance, as in Liu's approach [28].

Oscillations occur because balanced WFA excitation and inhibition is unstable in the presence of high-gain NFA negative feedback. Balance is achieved when the sustained and transient paths split the cone signal equally (e.g., $I_{t+} = I_{n+} = (I_c - I_{cq})/2$, if $I_c > I_{cq}$)—which requires a BT–to–NFA synaptic gain of unity. However, high synaptic-gain zeroes mismatch between the NFA and the cone at DC, creating an imbalance between excitatory and inhibitory inputs to the WFA cell (e.g., $I_{n+} \approx I_c - I_{cq}, I_{t+} \approx 0$, if $I_c > I_{cq}$). Conse-

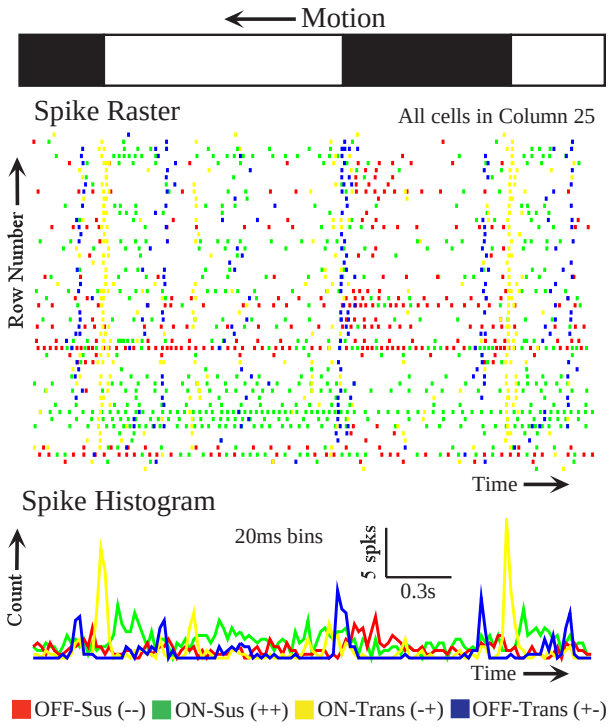


Figure 6. Spike Rasters and Histograms

Spike trains produced in response to vertical bars moving at $12°/s$ (equals 32 pixels/s) were recorded from 4×45 neurons in the same column. During this 2.65s recording, each neuron fired 13 spikes, on average. The histogram combines all the spike trains from this single-trial multiple-neuron recording.

quently, the WFA's input is given by

$$\langle \updownarrow i_t \updownarrow \rangle - \langle \updownarrow i_n \updownarrow \rangle \propto |I_c - I_n| - |I_n - I_{cq}|$$

for $I_w = I_{w0}$. It is inhibited whenever the DC cone current deviates significantly from the baseline, and hence V_w increases, reducing current levels in the subtractor to zero—and leakage currents take over.

At DC, inhibition overwhelms excitation in the WFA cell, producing either permanent hyperpolarization or slow, leakage-current–driven oscillation. Assuming leakage from V_{dd} to V_n dominates, it charges the capacitor, reducing I_n. If $I_n = I_c > I_{cq}$ initially, inhibition decreases while excitation increases. Consequently, V_w eventually repolarizes, restoring current levels in the subtractor, which discharges V_w and initiates another cycle. Else, if $I_n = I_c < I_{cq}$, both excitation and inhibition increase, but inhibition remains dominant. Hence, V_w never repolarizes. Assuming leakage to ground dominates, on the other hand, V_w still either oscillates or is permanently inhibited—it is just that the inequalities for I_c are reversed. For nonstatic images, we would expect the instability to disappear—it did not show up in simulations. This was indeed the case.

5. Detecting Motion Direction

The four ganglion-cell types respond to a moving edge in stereotyped sequences, which make it possible to distinguish edges of one contrast or the other moving in one direction or the other. Hence, using Visio1 as a frontend, direction-selective (DS) cells can be built simply by wiring up these four distinct receptive fields—no delays are required [10]. An example is shown in Figure 7—swapping red and green or yellow and blue produces the other three receptive fields, as explained in Figure 1. What exactly do such DS cells compute? And what is the effect of temporal adaptation?

Translating the bandpass-filtered image $\lambda/4$ in the negative x direction yields the spatial derivative—scaled by $\lambda/2\pi$. An approximate result obtained by assuming that the filter passes the spatial frequency $2\pi/\lambda$—and rejects all other frequencies. In that case, the derivative is

$$\frac{\partial J}{\partial x} = \frac{\partial}{\partial x} \sin\left(\frac{2\pi x}{\lambda} - \nu t\right) = \frac{2\pi}{\lambda} \sin\left(2\pi \frac{x - \lambda/4}{\lambda} - \nu t\right)$$

where the x-axis is perpendicular to the edge, which moves in the positive x direction with velocity ν.

Passing the translated image through a lowpass temporal filter and the untranslated image through a highpass temporal filter yields the temporal derivative—scaled by $\tau/2\pi$. On condition that these filters have the same corner frequency $2\pi/\tau$, since multiplying the lowpass' transfer function by $(\tau/2\pi)s$ yields the highpass'. And hence, the result is exact if we replace the original image with the lowpass filtered one.

Summing these spatial and temporal derivatives yields the projection of the gradient in spacetime onto the vector (λ, τ)—or $(-\lambda, \tau)$, when the lowpass and highpass temporal filters are swapped. That is,

$$\frac{1}{2\pi}(\lambda, \tau) \cdot \nabla J(x, t) = \frac{\lambda}{2\pi}\frac{\partial J}{\partial x} + \frac{\tau}{2\pi}\frac{\partial J}{\partial t} = \frac{1}{2\pi}(\lambda + \tau\nu)\frac{\partial J}{\partial x}$$

as shown in Figure 7. Hence, when λ and ν have the same sign (the preferred direction), the projection's amplitude is large. And when λ and ν have the opposite signs (the null direction), the projection's amplitude is small.

And temporal adaptation, which equalizes the output amplitudes of the lowpass and highpass temporal filters, makes (λ, τ) point along the gradient. That is,

$$\left|\frac{\lambda}{2\pi}\frac{\partial J}{\partial x}\right| = \left|\frac{\tau}{2\pi}\frac{\partial J}{\partial t}\right| \Leftrightarrow \frac{\lambda}{\tau} = \left|\frac{\partial J/\partial t}{\partial J/\partial x}\right| \equiv \nu$$

Hence, adaptation maximizes the spacetime-gradient's projection onto the sustained and transient channels by matching inner retina time-constants to the edge's time-of-flight across outer-retina receptive fields. In so doing, it also maximizes the amplitude of prefered-direction cells—making them independent of speed—and zeroes the amplitude of null-direction cells.

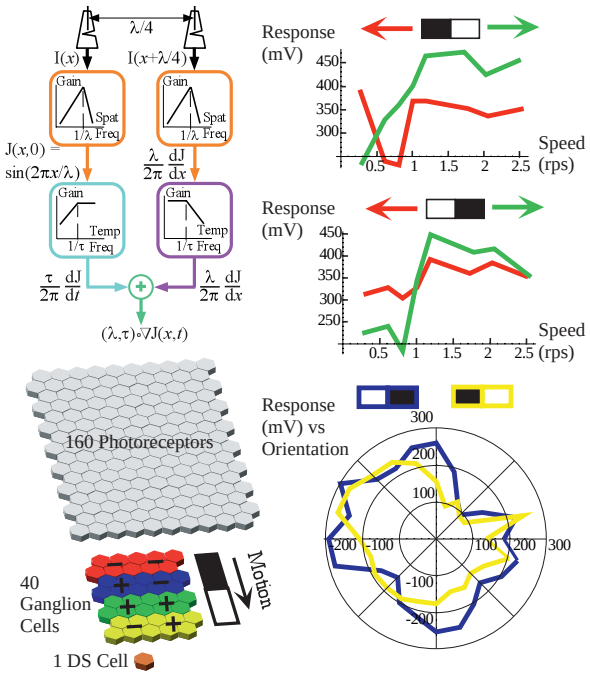


Figure 7. Direction-Selective Neurons

Retinal operations—bandpass spatial filtering and highpass temporal filtering—compute the gradient in spacetime, assuming the bandpass passes only the wavelength λ and the lowpass temporal filter compensates for the highpass' finite bandwidth, $1/\tau$ (upper-left panel). The amplitude of the gradient's projection onto the vector (λ, τ) is obtained by summing signals from four types of ganglion cells $\lambda/4$ apart—two cells in our case (lower-left). These 90°-multiple spatial phase-shifts match the ganglion cells' 90°-multiple temporal phase-shifts when a black–white edge moves in the direction shown, and hence all four responses peak simultaneously. Measurements reveal a clear preference for contrast polarity and motion direction in a target cell with this receptive field, for speeds from 4 to 40°/s—equivalent to 10.7 to 107 pixels/s (right panel). At the best speed, its direction-tuning is broad and centered around −60°.

This retinomorphic motion algorithm is a practical version of Watson and Ahumada's Hilbert-transform-based model of human visual-motion sensing [39]. In their model, the spatially and temporally bandpass filtered image is Hilbert-transformed spatially and temporally, and summed with itself. A Hilbert transform phase-shifts each frequency component by 90°—its amplitude remains unchanged. I approximate a spatial Hilbert transform by translating by $\lambda/4$—a phase difference of exactly 90° for the peak frequency. And I approximate a temporal Hilbert transform by equalizing the highpass and lowpass output amplitudes—which have a phase difference of exactly 90°. Etienne-Cummings et al's implementation of Adelson and Bergen's closely-related spatiotemporal energy model is quite similar [1, 19]—but it lacks the adaptive temporal dynamics provided by the Hilbert transform.

6. Conclusions

I reverse-engineered outer and inner retina microcircuits and morphed them into CMOS circuits to implement parallel visual pathways on a silicon chip. These micropower current-mode retinomorphic circuits are fairly compact, allowing several levels of processing to be performed at the focal plane. By going a step further than previous retinomorphic chips [7, 9] and including inner-retina processing, I modeled the four predominant ganglion-cell types in the primate retina.

In addition to improving spike-coding efficiency, these specialized visual channels provide more robust primitives for computing optical flow than differentiation and division, which most gradient-based algorithms call for [21, 36]. Adaptive inner-retina temporal dynamics extend the dynamic range for motion—just like adaptive outer-retina amplification extends the dynamic range for intensity. Unfortunately, my present design failed to modulate the loop-gain; I am redesigning it to rectify this. Nevertheless, and despite extremely limited mean firing rates of 5Hz, I demonstrated direction-selectivity over one decade of speed.

7. Acknowledgments

I was a doctoral student at Caltech, in Carver Mead's lab, when I begun this project, where it was funded by ONR; DARPA; and NSF's ERC Program. It is currently funded by start-up funds from University of Pennsylvania's Engineering and Medical Schools (through the Institute of Medicine and Engineering) and the Lindback Foundation. I thank Masahide Nomura, Eduardo Ros Vidal, and Rufin VanRullen for their help in programming the projective-field processor board to implement DS receptive fields at the 1998 Telluride Neuromorphic Engineering Workshop.

References

[1] E H Adelson and J R Bergen. Spatiotemporal energy models of the perception of motion. *J Opt Soc Am*, 2(2):284–299, 1985.

[2] A G Andreou and K A Boahen. Synthetic neural circuits using current-domain signal representations. *Neural Computation*, 1:489–501, 1989.

[3] A G Andreou and K A Boahen. A 48,000 pixel, 590,000 transistor silicon retina in current-mode subthreshold cmos. In *Proc. 37th Midwest Symposium on Circ. and Sys.*, pages 97–102, Lafayette, Louisiana, 1994.

[4] A G Andreou and K A Boahen. Neural information processing (ii). In M Ismail and T Fiez, editors, *Analog VLSI Signal and Information Processing*, chapter 8. McGraw-Hill, 1994.

[5] A G Andreou and K A Boahen. Translinear circuits in subthreshold mos. *J. Analog Integrated Circ. Sig. Proc.*, 9:141–166, 1996.

[6] J Atick and N Redlich. What does the retina know about natural scene. *Neural Computation*, 4(2):196–210, 1992.

[7] K A Boahen. Retinomorphic vision systems. Microneuro'96: Fifth Int. Conf. Neural Networks and Fuzzy Systems, Los Alamitos CA, 1996. IEEE Comp. Soc. Press.

[8] K A Boahen. *Retinomorphic Vision Systems: Reverse Engineering the Vertebrate Retina.* PhD thesis, California Institute of Technology, Pasadena CA, 1996.

[9] K A Boahen. The retinomorphic approach: Pixel-parallel adaptive amplification, filtering, and quantization. *Analog Integr. Circ. and Sig. Proc.*, 13:53–68, 1997.

[10] K A Boahen. Massive connectivity between neuromorphic chips using address-events. *In preparation*, 1999.

[11] K A Boahen. A throughput-on-demand address-event transmitter for neuromorphic chips. In *Conference on Advanced Research in VLSI*, Los Alamitos CA, 1999. IEEE Computer Soc Press.

[12] K A Boahen and A Andreou. A contrast-sensitive retina with reciprocal synapses. In J E Moody, editor, *Advances in neural information processing 4*, volume 4, San Mateo CA, 1991. Morgan Kaufman.

[13] E Cohen and P Sterling. Microcircuitry related to the receptive field center of the on-beta ganglion cell. *J Neurophysiol*, 65(2):352–9, 1991.

[14] L J Croner and E Kaplan. Receptive fields of p and m ganglion cells across the primate retina. *Vision Res*, 35(1):7–24, 1995.

[15] L J Croner, K Purpura, and et al. Response variability in retinal ganglion cells of primates. *Proc Natl Acad Sci U S A*, 90(17):8128–30, 1984.

[16] T Delbruck and C A Mead. Photoreceptor circuit with wide dynamic range. In *Proceedings of the International Circuits and Systems Meeting*, 1994.

[17] A M Derrington and P Lennie. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J Physiol (Lond)*, 357:219–40, 1984.

[18] M P Eckert and G Buchsbaum. Efficient coding of natural time-varying images in the early visual system. *Phil. Trans. Royal Soc. Lond. Biol*, 339(1290):385–395, 1993.

[19] R Etienne-Cummings, J van der Spiegel, P Mueller, and et al. Vlsi implementation of cortical visual motion detection using an analog neural computer. In D Touretzky, editor, *Advances in Neural Information Processing Systems 7*, volume 7, San Mateo CA, 1997. Morgan Kaufman.

[20] M A Freed and P Sterling. The on-alpha ganglion cell of the cat retina and its presynaptic cell types. *J Neurosci*, 8(7):2303–20, 1988.

[21] B K P Horn and B G Schunck. Determining optical flow. *Artif. Intell.*, 17:185–204, 1981.

[22] M Kamermans and F Werblin. Gaba-mediated positive autofeedback loop controls horizontal cell kinetics in tiger salamander retina. *J Neurosci*, 12(7):2451–63, 1992.

[23] E Kaplan and R M Shapley. X and y cells in the lateral geniculate nucleus of macaque monkeys. *J Physiol (Lond)*, 330:125–43, 1982.

[24] H. Kolb and R. Nelson. Functional neurocircuitry of amacrine cells in the cat retina. In A. Gallego and P. Gouras, editors, *Neurocircuitry of the Retina: A Cajal Memorial*, pages 215–232. Elsevier, New York, 1985.

[25] H Kolb and R Nelson. Off-alpha and off-beta ganglion cells in cat retina: Ii. neural circuitry as revealed by electron microscopy of hrp stains. *J Comp Neurol*, 329(1):85–110, 1993.

[26] S W Kuffler. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.*, 16:37–68, 1953.

[27] A G Leventhal and R W Rodieck. Retinal ganglion cell classes in the old world monkey: morphology and central projections. *Science*, 213(4512):1139–42, 1981.

[28] S Liu. Silicon model of motion adaptation in the fly. In *3rd UCSD-Caltech Symposium on Neural Computation.* ftp://ftp.ini.unizh.ch/pub/shih/symp.pdf, 1996.

[29] S Liu. Silicon retina with adaptive filtering properties. In *Advances in Neural Information Processing Systems*, San Mateo CA, 1998. Morgan-Kaufman.

[30] G Maguire, P Lukasiewicz, and et al. Amacrine cell interactions underlying the response to change in the tiger salamander retina. *J Neurosci*, 9(2):726–35, 1989.

[31] C A Mead. *Analog VLSI and Neural Systems.* Addison Wesley, Reading MA, 1989.

[32] A Pavasović, A G Andreou, and C R Westgate. Characterization of subthreshold mos mismatch in transistors for vlsi systems. *J Analog Integrated Circ & Sig Proc*, 6:75–84, 1994.

[33] J. L. Polyak. *The Retina.* Chicago Press, Chicago IL, 1941.

[34] R W Rodieck. The primate retina. *Comp. Primate Biol.*, 4:203–278, 1988.

[35] R G Smith. Simulation of an anatomically define local circuit — the cone-horizontal cell network in cat retina. *Visual Neurosci.*, 12(3):545–561, May-Jun 1995.

[36] J Tanner and C Mead. Optical motion sensor. In *Analog VLSI and Neural Systems*, chapter 14, pages 229–255. Addison-Wesley, Reading MA, 1989.

[37] E Vittoz and X Arreguit. Linear networks based on transistors. *Electronics Letters*, 29:297–299, 1993.

[38] M Watanabe and R W Rodieck. Parasol and midget ganglion cells of the primate retina. *J Comp Neurol*, 289(3):434–54, 1989.

[39] A B Watson and A J Ahumada, Jr. Model of human visual-motion sensing. *J Opt Soc Am*, 2(2):322–342, 1985.