February 2001

A retinomorphic chip with parallel pathways : encoding INCREASING, ON, DECREASING, and OFF visual signals

Kwabena A. Boahen
University of Pennsylvania, boahen@seas.upenn.edu

Postprint version. Published in *Analog Integrated Circuits and Signal Processing*, Volume 30, Issue 2, February 2002, pages 121-135.
Publisher URL: http://dx.doi.org/10.1023/A:1013751627357

A retinomorphic chip with parallel pathways : encoding INCREASING, ON, DECREASING, and OFF visual signals

Abstract
Retinomorphic chips may improve their spike-coding efficiency by emulating the primate retina's parallel pathways. To model the four predominant ganglion-cell types in the cat retina, I morphed outer and inner retina microcircuits into a silicon chip, Visio1. It has 104 x 96 photoreceptors, 4 x 52 x 48 ganglion-cells, a die size of 9.25 x 9.67 mm^2 in 1.2 μm 5V CMOS, and consumes 11.5 mW at 5 spikes/second/ganglion-cell. Visio1 includes novel subthreshold current-mode circuits that model horizontal-cell autofeedback, to decouple spatial filtering from local gain control, and model amacrine-cell loop-gain modulation, to adapt temporal filtering to motion. Different ganglion cells respond to motion in a quadrature 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 system, which implements Watson and Ahumada's model of human visual-motion sensing.

Keywords
neuromorphic systems, analog VLSI, mixed-mode design, CMOS imager, silicon retina, silicon neuron, spatiotemporal filtering, automatic gain control, contrast gain control, direction selectivity

Comments
Postprint version. Published in *Analog Integrated Circuits and Signal Processing*, Volume 30, Issue 2, February 2002, pages 121-135.
Publisher URL: http://dx.doi.org/10.1023/A:1013751627357

A Retinomorphic Chip with Parallel Pathways: Encoding INCREASING, ON, DECREASING, and OFF Visual Signals

Kwabena Boahen
Penn Bioengineering
3320 Smith Walk
Philadelpha PA 19104

February 12, 2001

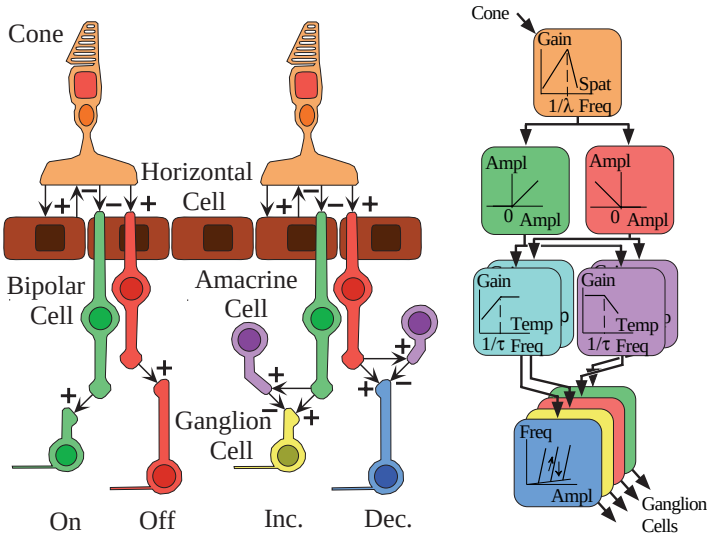
Abstract. Retinomorphic chips may improve their spike-coding efficiency by emulating the primate retina's parallel pathways. To model the four predominant ganglion-cell types in the cat retina, I morphed outer and inner retina microcircuits into a silicon chip, Visio1. It has 104×96 photoreceptors, $4 \times 52 \times 48$ 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/ganglion-cell. Visio1 includes novel subthreshold current-mode circuits that model horizontal-cell autofeedback, to decouple spatial filtering from local gain control, and model amacrine-cell loop-gain modulation, to adapt temporal filtering to motion. Different ganglion cells respond to motion in a quadrature 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 system, which implements Watson and Ahumada's model of human visual-motion sensing.

Keywords: Neuromorphic Systems, Analog VLSI, Mixed-Mode Design, CMOS Imager, Silicon Retina, Silicon Neuron, Spatiotemporal Filtering, Automatic Gain Control, Contrast Gain Control, Direction Selectivity

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 [28, 24, 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



(a) Retinal Structure: The visual signal is relayed from the photoreceptors (called cones) to the ganglion cells (whose axons form the optic nerve) by bipolar cells, with synaptic interactions occuring within two plexiform layers. Horizontal cells mediate lateral interactions in the first (outer) plexiform layer while amacrine cells mediate lateral interactions in the second (inner) plexiform layer; these interneurons are inhibitory. Four types of ganglion cells, which I call ON, OFF, INCREASING, and DECREASING, encode the retina's output in their spike trains. (b) Retinal Function: The visual signal is bandpass filtered spatially in the outer retina, then lowpass and highpass filtered temporally in the inner retina. These analog signals are first half-wave rectified, and then later encoded as spike trains. 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 firing rate is linearly proportional.

Figure 1. Retinal Circuits and Operations

longer distances and respond more transiently than midgets (also called β cells) [43, 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 [35]. 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 [35].

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 [27]. Complementary signaling overcomes three shortcomings of using a single channel to transmit both increases and decreases, measured relative to a baseline:[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.

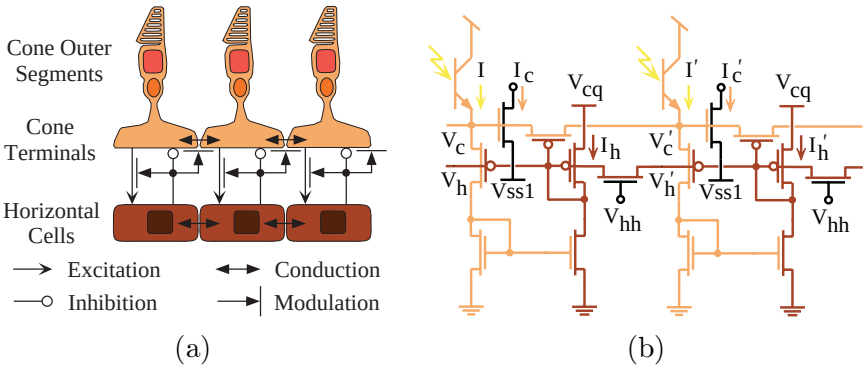
Except prior to the very first synapses, found in rod and cone terminals, complementary signaling is used throughout the retina to transmit information efficiently using vesicles or spikes.

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 potassium 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 [31].

Behaving like leaky integrators, these baseline-setting 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

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



(a) (b)

(a) Neural Microcircuitry: Cone terminals (CT) receive a photocurrent I that is proportional to incident light intensity from their outer segments (CO). To control CO to CT attenuation without changing the cone-network's space-constant, the horizontal cell (HC) modulates both CT gap junctions and membrane conductances. Cone-excitation modulation is realized by GABA auto-receptors in the HC membrane, which form a positive feedback loop. (b) CMOS Circuit: 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 CT and HC activities, which are represented by the currents I_{c} and I_{h}. This current is sunk from the CT node, V_{c}, to model HC shunting inhibition, and dumped on the HC node, V_{h}, to model CT excitation and HC autofeedback. V_{cq} sets the mean level of the cone current I_{c}.

Figure 2. Morphing the Outer Retina

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 Sections 2 and 3. These current-mode circuits operate in the subthreshold region [32], where small-signal conductances and transconductances are proportional to the current level [2], current-spreading diffusor networks are linear [12, 42, 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, 30] and at the network level. Previous attempts to make these two functions coexist in one

network produced undesirable side-effects [12, 3]. In particular, the circuit starts ringing if you attenuate redundant low-frequency temporal and spatial signals by increasing the negative-feedback loop's gain. And the receptive field expands alarmingly if you reduce photocurrent sensitivity by increasing the cone–to–cone conductance.

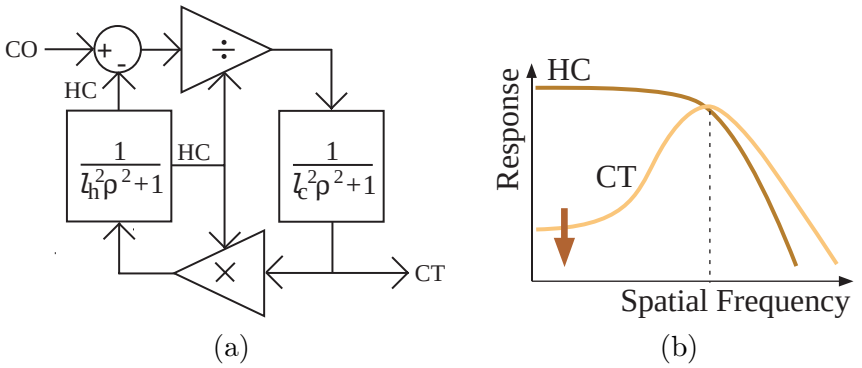
To overcome the high-gain negative-feedback outer retina circuit model's shortcomings, I searched the neurobiology literature for a retinal mechanism that could *decouple* spatiotemporal filtering, local gain control, and temporal stability. Horizontal-cell autofeedback, which was demonstrated by Kamermans and Werblin a few years ago in the tiger salamander [23], can achieve this. Horizontal cells (HC), 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 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 HCs, 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].

HC autofeedback can improve temporal stability by amplifying the cone signal, allowing us to decrease the strength of the cone–to–HC synapse. Thus, we can attenuate low-frequency signals while maintaining temporal stability. A lower cone–to–HC 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–HC synapse. More HC activity provides a larger boost to synaptic input from cones. Therefore, if HC activity is proportional to intensity, then the cone–to–HC synaptic strength becomes proportional to intensity as well, compensating for the decrease in cone sensitivity with increasing intensity when local automatic gain control is in effect. A neural microcircuit for the outer retina that includes HC autofeedback is shown in Figure 2a.

A system-level analysis of autofeedback, with the aid of the block diagram shown in Figure 3a, reveals that

$$
\begin{aligned}
i_{\mathrm{hc}} &= \left(\frac{A}{(\ell_{\mathrm{c}}^2 \rho^2 + 1)(\ell_{\mathrm{h}}^2 \rho^2 + 1) + A/B} \right) \frac{i_{\mathrm{co}}}{B} \\
i_{\mathrm{ct}} &= \left(\frac{\ell_{\mathrm{h}}^2 \rho^2 + 1}{(\ell_{\mathrm{c}}^2 \rho^2 + 1)(\ell_{\mathrm{h}}^2 \rho^2 + 1) + A/B} \right) \frac{i_{\mathrm{co}}}{B},
\end{aligned}
$$

where B is the attenuation from the cone outer segment (CO) to the cone terminal (CT) and A is the amplification from CT to the HC. Thus, HC lowpass-filters the CO signal while CT bandpass-filters it,



(a) System-Level Concept: Visual signals travel from the cone outer-segment (CO), to the cone terminal (CT), and on to the horizontal cell (HC), which closes the negative feedback loop. To realize local automatic gain control, HC modulates the attenuation from CO to CT. And to keep the loop-gain constant, HC also modulates the signal it receives from CT, giving rise to a positive-feedback loop. This compensation keeps spatial filtering invariant, while HC activity tracks the local average intensity $\langle I_{\rm co}\rangle$ and CT activity becomes proportional to contrast $I_{\rm co}/\langle I_{\rm co}\rangle$. (b) Frequency Responses: Both HC and CT act as lowpass filters, but HC has a longer space constant (i.e., $\ell_{\rm h} > \ell_{\rm c}$). Thus, HC inhibition eliminates low frequency signals, resulting in bandpass response in CT. This bandpass characteristic corresponds to an impulse response (i.e., receptive field) with a small central excitatory region and a large surrounding inhibitory region.

Figure 3. Outer Retina Spatial Filtering

as shown in Figure 3b. These filters have the same corner frequency, which is inversely proportional to the closed loop space-constant $\ell_A = (B/A)^{1/4}\sqrt{\ell_{\rm c}\ell_{\rm h}}$, where $\ell_{\rm c}$ and $\ell_{\rm h}$ are the space-constants of the cone and horizontal-cell networks. Therefore, spatial filtering is invariant if we make A proportional to B, which makes HC activity track the local average intensity. Hence, CT activity becomes proportional to contrast if we set $A = B = i_{\rm hc}$.

A novel current-mode CMOS circuit that uses HC autofeedback to decouple spatiotemporal filtering, local gain control, and temporal stability is shown in Figure 2b. I implement CT attenuation by modulating the conductances of both the cone's membrane and its gap junctions—through HC shunting inhibition—which leaves the cone-network's space constant unchanged. Currents represent signals instead of voltages, as this is a current-mode circuit. I use the subscripts c and h to denote cone and horizontal-cell; a pair of letters denotes coupling between cells or to ground, which is denoted by 0. And I use upper case symbols (e.g.,

Table I. Negative versus Positive Feedback

	Attenuation	Damping-Factor	Space-Constant	Sensitivity
Neg	$\frac{c_{\mathrm{c}0} I_{\mathrm{c}}}{c_{\mathrm{c}0} I_{\mathrm{c}} + c_{\mathrm{h}0} I_{\mathrm{h}}}$	$\frac{1}{\epsilon_{\mathrm{E}}} \frac{\sqrt{I_{\mathrm{c}} I_{\mathrm{h}}}}{Q_{\mathrm{I}} I_{\mathrm{c}} + I_{\mathrm{h}}/Q_{\mathrm{I}}}$	$L_{\mathrm{I}} I_{\mathrm{h}}^{1/4}$	$A_{\mathrm{I}} I_{\mathrm{h}}^{1/4}$
Pos	0	Q_{I}	L_{I}	$A_{\mathrm{I}} I_{\mathrm{c}}$

I_{c} is the mean cone activity, I_{h} is the mean horizontal cell activity, which is proportional to light intensity, and $\epsilon_{\mathrm{E}} \equiv U_{\mathrm{T}}/V_{\mathrm{E}}$ is the ratio between the thermal voltage and the Early voltage. Ideally, the attenuation, a measure of low-frequency signals' relative amplitude, should be zero; the damping-factor $Q_{\mathrm{I}} = \sqrt{c_{\mathrm{c}0}/c_{\mathrm{h}0}}$, a measure of temporal stability, should be less than one; the space constant $L_{\mathrm{I}} = (\alpha_{\mathrm{hh}} \alpha_{\mathrm{cc}})^{1/4}$, a measure of receptive field size, should be independent of I_{h}; and the sensitivity $A_{\mathrm{I}} = (L/\alpha_{\mathrm{cc}}) I_{\mathrm{h}}^{-1}$, a measure of photosignal amplification, should be inversely proportional to I_{h}.

I_{h}) to represent the mean (i.e., DC) signal value, while lower case ones (e.g., i_{h}) represent small instantaneous deviations from the mean.

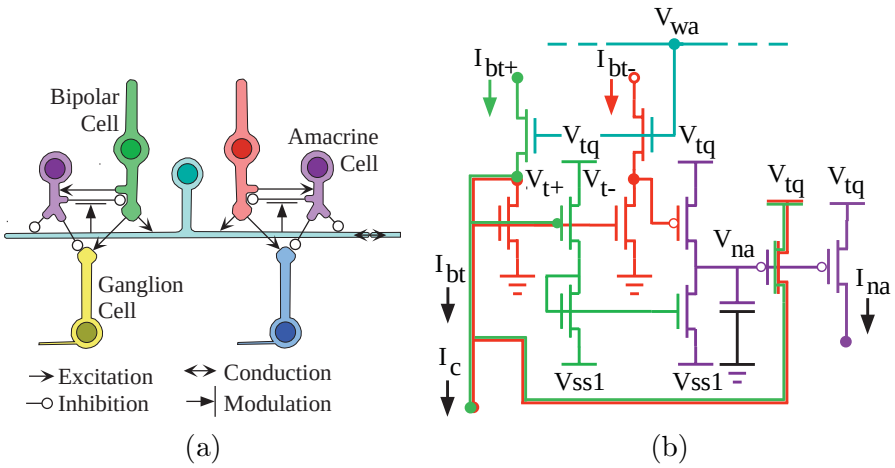
In the continuum limit, the outer retina circuit's small-signal (i.e., linear) behavior can be described in terms of the contrast signals $c \equiv i_{\mathrm{c}}/I_{\mathrm{c}}$ and $h \equiv i_{\mathrm{h}}/I_{\mathrm{h}}$ as follows:

$$\begin{aligned} \frac{i_0}{I_{\mathrm{h}}} - h + \alpha_{\mathrm{cc}} \nabla^2 c &= \frac{c_{\mathrm{c}0} U_T}{I_{\mathrm{h}}} \frac{dc}{dt} + c, \\ c + \alpha_{\mathrm{hh}} \nabla^2 h &= \frac{c_{\mathrm{h}0} U_T}{I_{\mathrm{h}}} \frac{dh}{dt}, \end{aligned}$$

where $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$. The absolute signal levels arise because the transistors' small-signal conductances and transconductances that model membrane conductances, gap junctions, and chemical synapses are linearly proportional to the DC current level in the subthreshold region. 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 [12]. $U_T \equiv kT/q$ is the thermal voltage. Notice that the behavior is entirely linear in the contrast signals c and h, apart from the input attenuation and the time-constants, which are controlled by I_{h}. However, the time-constants track each other, making it easy to maintain temporal stability.

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

$$\begin{aligned} \frac{i_0}{I_{\mathrm{h}}} - h + \alpha_{\mathrm{cc}} \frac{I_{\mathrm{c}}}{I_{\mathrm{h}}} \nabla^2 c &= \frac{c_{\mathrm{c}0} U_T}{I_{\mathrm{h}}} \frac{dc}{dt} + \epsilon_{\mathrm{E}} c \\ c + \alpha_{\mathrm{hh}} \frac{I_{\mathrm{h}}}{I_{\mathrm{c}}} \nabla^2 h &= \frac{c_{\mathrm{h}0} U_T}{I_{\mathrm{c}}} \frac{dh}{dt} + \epsilon_{\mathrm{E}} h, \end{aligned}$$



(a) Neural Microcircuit: The mechanism for modulating presynaptic inhibition at the bipolar terminal (BT) is not known—my model assigns this role to a wide-field amacrine cell. Full-wave rectification is realized by convergence of ON and OFF pathways. For moving stimuli, the gap-junction coupled WAs obtain signal energy by spatial integration. (b) Bipolar-Terminal Subcircuit: Unlike the neural circuit, complementary signaling is not used for BCs and NAs and modulation occurs before lowpass filtering. The NA signal, I_{na}, is subtracted from the unrectified cone signal, I_{c}, to obtain the BT signal I_{bt}. A four-transistor rectifier produces ON and OFF BT signals ($I_{\mathrm{bt}\pm}$). And a four-transistor subtractor takes the difference between $I_{\mathrm{bt}+}$ and $I_{\mathrm{bt}-}$ and supplies it to the NA node (V_{na})—V_{wa} modulates the current level in the subtractor.

Figure 4. Morphing the Inner Retina I

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 HC–to–cone synaptic gain, $A_{\mathrm{ch}} \equiv dc/dh$, is a thousand (i.e., $1/\epsilon_{\mathrm{E}}$) times smaller in the new circuit, while the cone–to–HC synaptic gain, $A_{\mathrm{hc}} \equiv dh/dc$, is infinite—for zero temporal and spatial frequencies (i.e., DC)—due to 100% positive feedback.[2] This positive feedback arrangement was previously used to realize a fast current-amplifier and buffer stage [40]. For nonzero frequencies, the cone–to–HC 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 I.

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

3. Inner Retina Model

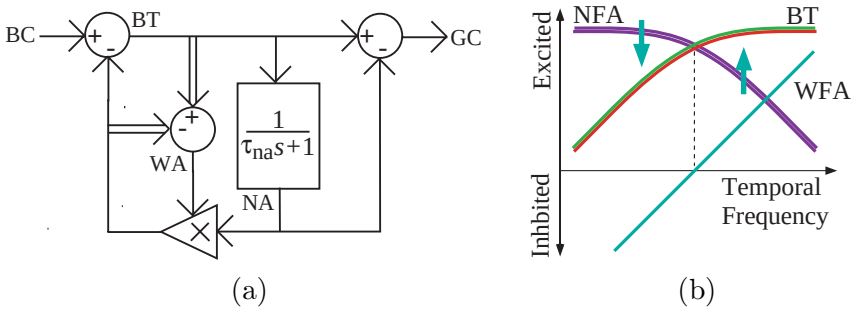
The inner retina performs lowpass and highpass temporal filtering and adapts its dynamics locally. Midget and parasol ganglion cells receive synaptic inputs from both bipolar and amacrine cells, but parasol cells receive more amacrine input (i.e., feedforward inhibition) [21, 13, 26], which accounts 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 [34, 35], 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 hypothesize that a change in closed-loop gain produces temporal adaptation, and that this gain is controlled by modulating the strength of feedback inhibition. Figure 4a illustrates this retinal mechanism for adapting temporal dynamics. A wide-field amacrine cell (WA) modulates the strength of feedback (i.e., presynaptic) inhibition at the bipolar terminals (BT). WA is excited by ON and OFF bipolars and inhibited by ON and OFF narrow-field amacrine cells (NA). All these synaptic interactions have been found in an anatomically identified amacrine cell type, called A19 [25]. 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.

My inner-retina circuit model, which was motivated by temporal adaptation, also accounts for contrast-gain control in the retina. The classic result is that ganglion cell responses compress in amplitude and time when they are driven with larger and larger step changes in intensity [36, 37]. Victor and Shapley accounted for both of these nonlinear effects by modeling the inner retina with a highpass temporal filter whose time-constant decreases with contrast [41]. At the functional level, the model I propose here is essentially identical to theirs. My contribution is to flesh out the neural mechanisms by which contrast is measured and the time-constant is adjusted. In the process, I assign specific roles to anatomically identified retinal cells and their synaptic connections.

A system-level analysis of loop-gain modulation, with the aid of the block diagram shown in Figure 5a, and the Laplace transform, reveals that

$$i_{\mathrm{na}} = \frac{\epsilon}{\tau_A s + 1} i_{\mathrm{bc}},\ i_{\mathrm{bt}} = \frac{\tau_A s + \epsilon}{\tau_A s + 1} i_{\mathrm{bc}},\ i_{\mathrm{gc}} = \frac{\tau_A s}{\tau_A s + 1} i_{\mathrm{bc}}; \qquad (1)$$



(a) System-Level Concept: Visual signals are relayed from cone terminal to ganglion cell (GC) by the bipolar cell (BC). They are also fed to a narrow-field amacrine cell (NA), which forms both negative-feedback and negative-feedforward loops. Feedback strength is modulated by a wide-field amacrine cell (WA) that computes the difference between full-wave rectified (double arrows) BT and NA signals. (b) Frequency Responses: NA inhibition eliminates low frequencies, resulting in a highpass response in BT that has the same corner frequency as the lowpass NA response. For frequencies above this corner frequency, BT excitation dominates and hence WA activity rises, and it boosts NA inhibition. For frequencies below the corner frequency, on the other hand, NA inhibition dominates and hence WA activity falls, and it throttles NA inhibition.

Figure 5. Inner Retina Temporal Filtering

for NA, BT, and the parasol ganglion cell (GC), respectively, where,

$$\tau_A \equiv \epsilon\tau_{\rm na},\ \epsilon \equiv 1/(1 + I_{\rm wa}/I_{\rm wa0}). \tag{2}$$

$\tau_{\rm na}$ is the time-constant of NA and $I_{\rm wa}/I_{\rm wa0}$ is the modulation of its response by WA, whose response is $I_{\rm wa0}$ for zero frequency (i.e., static input). Thus, NA lowpass-filters the BC signal while BT highpass-filters it—these filters have the same corner frequency, which is inversely proportional to the closed loop time-constant τ_A. Through the dependence of τ_A on the modulation level, the corner frequency becomes proportional to WA activity, as shown in Figure 5b.

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, $\epsilon i_{\rm bt}$, increases as the loop gain is reduced to lower the corner frequency—and thus the BT response asymptotically approaches an allpass filter. However, NA's residual grows in the same way, and cancels out BT's residual in the parasol 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 wide-field amacrine cell centers the corner frequencies of the highpass and lowpass filters on the input spectrum. With BT–to–WA and NA–to–WA synapses of equal strength, WA's activity is given by

$$\alpha_{\mathrm{ww}}\nabla^2 i_{\mathrm{wa}} + \updownarrow i_{\mathrm{bt}} \updownarrow - \frac{I_{\mathrm{wa}}}{I_{\mathrm{wa0}}} \updownarrow i_{\mathrm{na}} \updownarrow = 0 \tag{3}$$

where α_{ww} is the coupling strength between WAs—normalized by the synaptic strength—and $\updownarrow i \updownarrow$ is the full-wave–rectified version of i; I have neglected WA's leakage conductance. Notice that WA modulates the inhibition it receives from NA, forming an autofeedback loop analogous to that found in HCs. For low spatial frequencies (i.e., on average), the Laplacian is close to zero, and we have

$$\frac{I_{\mathrm{wa}}}{I_{\mathrm{wa0}}} \approx \frac{\langle \updownarrow i_{\mathrm{bt}} \updownarrow \rangle}{\langle \updownarrow i_{\mathrm{na}} \updownarrow \rangle} = \frac{|i_{\mathrm{bt}}|}{|i_{\mathrm{na}}|} = \sqrt{\tau_{\mathrm{na}}^2 \omega^2 + 1},$$

where $\langle i \rangle$ is the local spatial average of i. Notice that $I_{\mathrm{wa}} = I_{\mathrm{wa0}}$ when $\omega = 0$, hence the loop-gain is unity at DC and $i_{\mathrm{bt}} = i_{\mathrm{na}} = i_{\mathrm{bc}}/2$. Substituting this expression into Equation 2 yields:

$$\tau_A = \frac{\tau_{\mathrm{na}}}{1 + \sqrt{\tau_{\mathrm{na}}^2 \omega^2 + 1}} \approx \begin{cases} \tau_{\mathrm{na}}/2 & \omega \ll 1/\tau_{\mathrm{na}} \\ 1/\omega & \omega \gg 1/\tau_{\mathrm{na}} \end{cases} .$$

Hence, the corner frequency $1/\tau_A$ tracks the input frequency ω if it is higher than NA's intrinsic cut-off frequency $1/\tau_{\mathrm{na}}$.

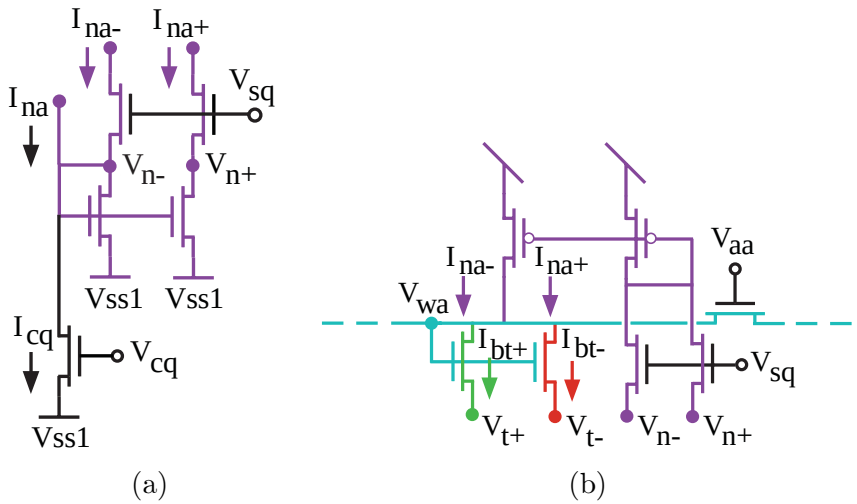
Such temporal adaptation could account for the variation of ganglion cells' temporal dynamics with eccentricity. Neurobiologists have observed that both midgets and parasols become more transient with increasing eccentricity, and the midget's sustained component decreases [43, 14]. Given that the input spectrum shifts to higher temporal frequencies with increasing eccentricity [18], this variation in temporal characteristics with eccentricity can be produced by a temporal adaptation mechanism like the one proposed here.

Temporal adaptation in my inner-retina circuit model also results in amplitude compression, or so-called contrast-gain control. This compression is evident if we transform the s-domain equations for i_{bt} and i_{na} (Equation 1) into the time-domain:

$$\frac{\tau_{\mathrm{na}}}{1+w}\frac{di_{\mathrm{bt}}}{dt} + i_{\mathrm{bt}} = \frac{\tau_{\mathrm{na}}}{1+w}\frac{di_{\mathrm{bc}}}{dt} + \frac{i_{\mathrm{bc}}}{1+w}, \tag{4}$$

$$\frac{\tau_{\mathrm{na}}}{1+w}\frac{di_{\mathrm{na}}}{dt} + i_{\mathrm{na}} = \frac{i_{\mathrm{bc}}}{1+w} \tag{5}$$

where $w \equiv I_{\mathrm{wa}}/I_{\mathrm{wa0}}$. For a uniform intensity over a large area, with $i_{\mathrm{bc}} = I_0$, $i_{\mathrm{na}} = i_{\mathrm{bt}} = I_0/2$ and $I_{\mathrm{wa}} = I_{\mathrm{wa0}}$. Immediately after the intensity is stepped to I_1, we have $i_{\mathrm{na}} = I_0/2$ and $i_{\mathrm{bt}} = I_1 - I_0/2$. Substituting



(a) Narrow-Field Amacrine Subcircuit: A reference current $I_{\rm cq}$, which matches the cone current's mean level, is subtracted from a copy of $I_{\rm na}$ produced by the bipolar-terminal (BT) circuit (shown in Figure 4b), and the difference is rectified to obtain ON and OFF narrow-field amacrine (NA) signals ($I_{\rm na\pm}$). (b) Wide-Field Amacrine Subcircuit: Copies of the ON and OFF BT ($I_{\rm bt\pm}$) and NA ($I_{\rm na\pm}$) signals, obtained from the BT and NA circuits by connecting a second transistor in parallel, are sunk from or mirrored onto the WA node ($V_{\rm wa}$) to excite or inhibit it, respectively. Neighboring WA nodes are connected together by transistors to model their gap junctions.

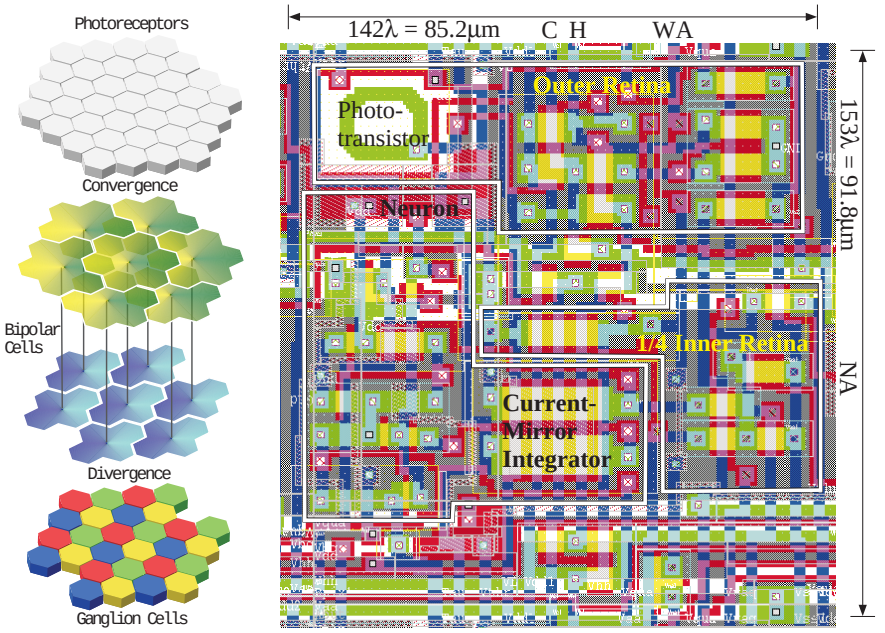
Figure 6. Morphing the Inner Retina II

these values in Equation 3 yields $I_{\rm wa}/I_{\rm wa0} = (I_1 - I_0/2)/(I_0/2)$, which equals $2c+1$, where $c \equiv (I_1 - I_0)/I_0$ is the contrast. I have assumed that the outer-retina's temporal filtering is infinitely fast and its intensity adaptation is infinitely slow, for simplicity. Replacing $i_{\rm bc}$ with $(1+c)I_0$, its derivative with $cI_0\delta(t)$, and w with $2c+1$, in the equation for $i_{\rm bt}$ above yields

$$\frac{1}{2}\frac{\tau_{\rm na}}{1+c}\frac{di_{\rm bt}}{dt} + i_{\rm bt} = \frac{I_0}{2}\left(1 + \frac{\tau_{\rm na}}{1+c}c\delta(t)\right).$$

It is evident from this result that the time-constant shortens and the driving force compresses as contrast increases. In my model, this effect occurs at the bipolar terminal, which passes it on to both sustained (midget- or X-like) and transient (parasol- or Y-like) ganglion cells. Victor and Sharpley did observe contrast-gain control in both X and Y ganglion cells in their experiments with cats [37].

A current-mode CMOS circuit I designed to implement inner retina temporal adaptation is shown in Figure 4b. It does not include feedfor-



(a) (b)

(a) Imager Array Architecture: 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) Pixel Layout: 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., $I_{\mathrm{na}+}$ in Figure 6a). 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.

Figure 7. Visio1 Chip Design

ward NA inhibition onto transient 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. To compare the energies of the highpass- and lowpass-filtered signals, it full-wave rectifies them, takes

the difference, and integrates it over space, using additional circuitry shown in Figure 6.

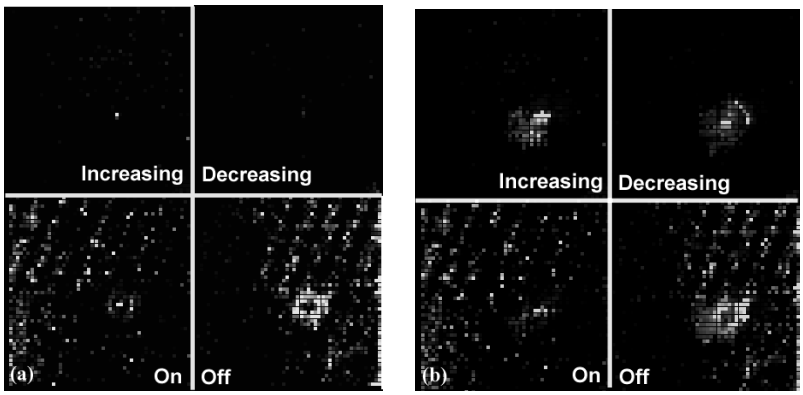
In contrast, a similar time-constant adaptation scheme proposed by Liu uses separate highpass and lowpass filters, integrates the difference between their *peak* responses over time, and adjusts their time-constants directly by changing the amplifiers' bias currents [29]. This scheme—implemented with a voltage-mode circuit—requires the amplifiers' transconductances to be matched and adapts over several cycles. Since the edge moves on after a single response cycle, it is imperative that adaptation occur instantaneously and this information propagate quickly to neighboring cells. Both objectives are achieved by my retinomorphic circuit.

My inner-retina circuit model computes signal energy rapidly by using spatial integration, since, for a moving stimulus, all phases of the response are available at different locations at the same instant. And it uses the same averaging network to distribute the results quickly. This collective computation also makes the retinomorphic approach more robust, compared to peak-detection. Unfortunately, its CMOS implementation in Visio1 has a low-frequency temporal instability—together with other shortcomings—that was discovered during testing.

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; its die size is $9.25 \times 9.67\text{mm}^2$. Visio1's architecture and pixel layout are described in Figure 7. Each pixel includes an adaptive silicon neuron that generates spikes [9], which are read out by an asynchronous address-event transmitter interface [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.

In addition to horizontal-cell autofeedback in the outer retina and amacrine-cell loop-gain modulation in the inner retina, Visio1 models cone–to–bipolar convergence, which makes the receptive-field center more Gaussian-like [38]. 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



(a) (b)

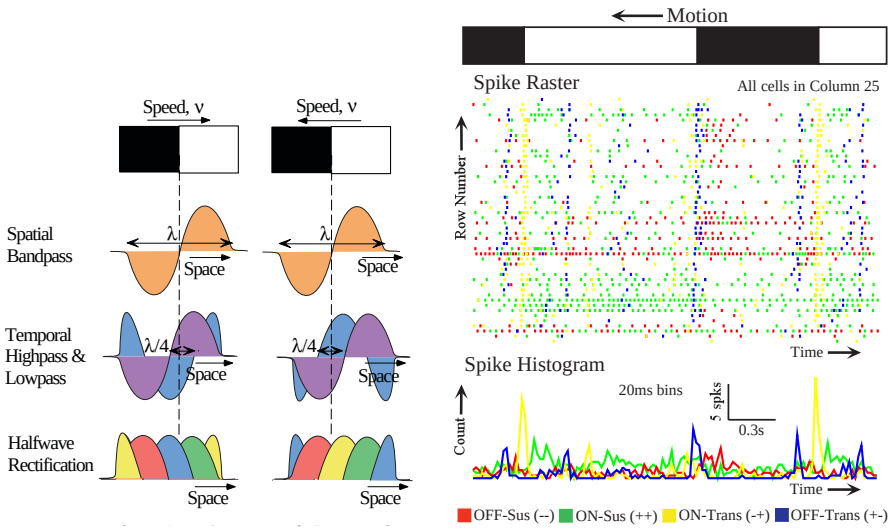
(a) Light spot stationary (located where the single active INCREASING cell is): ON cells pick up the increased signal at the spot's location while OFF cells pick up the decreased signal in surrounding region, due to lateral inhibition. Fixed-pattern noise, due to transistor mismatch, is also evident. (b) Light spot moving up and to the right: INCREASING cells pick up the increases at the excitatory center's leading edge and at the inhibitory surround's trailing edge, while DECREASING cells pick up the decreases at the inhibitory surround's leading edge and excitatory center's trailing edge. The mean spike rate was 5 spikes/neuron/sec.

Figure 8. Response of Retinomorphic Chip to Light Spot

by convergence at the ganglion-cell level, which I am yet to include—it would also reproduce the parasol cells' larger receptive fields.

Visio1 does not include synaptic interactions at the ganglion-cell level; these will be added in the next version. Unlike in Figure 1a, its ON and OFF ganglion cells are driven by narrow-field amacrine (NA) signals ($I_{\mathrm{na}\pm}$ in Figure 6a), while its INCREASING and DECREASING ganglion cells are driven by bipolar-terminal (BT) signals ($I_{\mathrm{bt}\pm}$ in Figure 4b). Driving all four ganglion cell types with BT signals and adding feedforward NA inhibition onto the transient cells, like the retina does, will make their responses more transient. Thus, these ganglion-cell level synaptic interactions result in a more efficient encoding, as they remove redundant low-frequency signals.

Visio1's light-spot responses are shown in Figure 8; these measurements demonstrate the effects of horizontal-cell (HC) and amacrine-cell inhibition. HCs inhibit neighboring spatial locations, giving rise to the inhibitory surround which is evident in the OFF response. While NAs inhibit focally, but with a delay, terminating the response in the INCREASING and DECREASING cells. However, the spot escapes this temporal lateral inhibition when it moves, giving rise to INCREAS-



(a) Expected Responses: Bandpass spatial filtering passes a particular wavelength λ and rejects the others. Highpass and lowpass temporal filtering translates the response in space by $\lambda/4$. Half-wave rectification produces a sequence of four responses that peak $\lambda/4$ apart, coinciding with the INCREASING (yellow), ON (green), DECREASING (blue), and OFF (red) parts of the cycle. Reversing the direction reverses the temporal contrast, interchanging the yellow and blue responses. (b) Observed Responses: 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, which combines all the spike trains from this single-trial multiple-neuron recording, confirms that the four ganglion cell types provide a quadrature representation.

Figure 9. Spike Rasters and Histograms for Moving Edges

ING and DECREASING responses at the leading and trailing edges of the excitatory center and the inhibitory surround. 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 [33].

A raster plot of spike trains recorded from neurons in a single column is shown in Figure 9b. 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 below 10Hz—several time less than in the real retina. Nevertheless, the sequence in which the four types fire is as predicted in Figure 9a. 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.

I expected to use the variable current-gain between the rectifier and the subtractor in my inner retina circuit (see Figure 4b) to modulate the loop-gain, but that proved futile. The currents in the subtractor are $A_{\mathrm{wa}} I_{\mathrm{bt}\pm}$, with $A_{\mathrm{wa}} \equiv e^{(V_{\mathrm{pn}}+V_{\mathrm{tq}}-V_{\mathrm{wa}})/U_T}$, where V_{pn} accounts for unequal pFET and nFET currents when $V_{\mathrm{tq}} = V_{\mathrm{wa}}$, due to differences in body-effect and zero-bias currents. Keeping in mind that the drain conductance g_{d} of the subtractor's output transistors is proportional to their current levels, we find that the voltage change v_{na} produced at node V_{na} for a given current difference $i_{\mathrm{bt}} \equiv I_{\mathrm{bt}+} - I_{\mathrm{bt}-}$ is:

$$v_{\mathrm{na}} = \frac{A_{\mathrm{wa}} i_{\mathrm{bt}}}{g_{\mathrm{d}}} = \frac{A_{\mathrm{wa}} i_{\mathrm{bt}}}{A_{\mathrm{wa}}(I_{\mathrm{bt}+} + I_{\mathrm{bt}-})/V_{\mathrm{E}}} = \frac{i_{\mathrm{bt}}}{I_{\mathrm{bt}+} + I_{\mathrm{bt}-}} V_{\mathrm{E}}.$$

It is *independent* of the current-gain A_{wa}.

Surprisingly, varying V_{wa} does not directly influence the relationship between the narrow-field amacrine and bipolar terminal signals in my circuit. Infact,

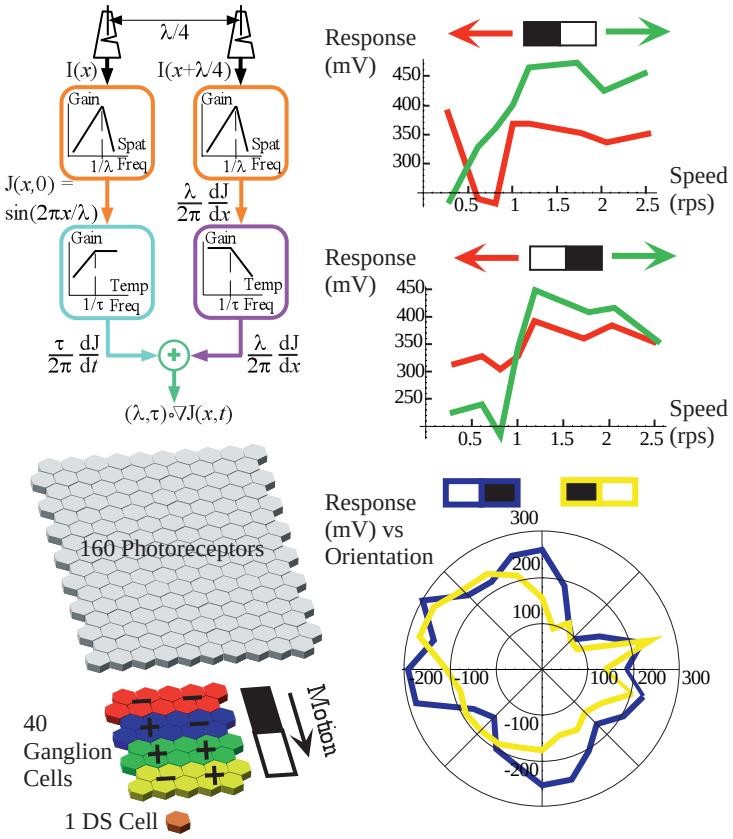
$$i_{\mathrm{na}} = g_{\mathrm{m}} v_{\mathrm{na}} = \frac{\kappa V_{\mathrm{E}}}{U_T} \frac{I_{\mathrm{na}}}{I_{\mathrm{bt}+} + I_{\mathrm{bt}-}} i_{\mathrm{bt}},$$

where $g_{\mathrm{m}} = \kappa \mathbf{I}_{\mathrm{na}}/U_{\mathrm{T}}$ is the feedback device's transconductance. Hence, to modulate the loop gain, we must change either I_{na} or $I_{\mathrm{bt}\pm} \equiv I_{\mathrm{bt}+} + I_{\mathrm{bt}-}$. I_{na}, the narrow-field amcrine current's mean level, is set by I_{cq}, the mean cone current level, which is constant (set by V_{cq} in Figure 2).

However, we can increase the mean bipolar terminal signal levels, $I_{\mathrm{bt}\pm}$, by increasing V_{wa}, and thereby reduce the loop-gain. This dependence arises because the rectifier's behavior is described by the following circuit equations [5]:

$$\begin{aligned} I_{\mathrm{bt}+} I_{\mathrm{bt}-} &= I_{\mathrm{bq}}^2; \\ I_{\mathrm{bt}+} - I_{\mathrm{bt}-} &= I_{\mathrm{c}} - I_{\mathrm{na}}, \end{aligned}$$

where $I_{\mathrm{bq}} = I_0 e^{\kappa V_{\mathrm{wa}}/2U_T}$. $I_{\mathrm{bt}\pm}$ will have to be several hundred times larger than I_{na} to achieve a loop-gain of unity, due to the large ratio between the subthreshold MOS transistor's transconductance and its drain conductance (i.e., $\kappa V_E/U_T$). But, of course, the mean currents themselves also contribute to wide-field amacrine excitation! Hence,



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' saturation above $1/\tau$ (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 places away in our case (lower-left). These spatial quadrature phase-shifts match the ganglion cells' temporal quadrature 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 (upper right). At the best speed, its direction-tuning is broad and centered around $-60°$ (lower right).

Figure 10. Direction-Selective Neurons

the ratio $I_{\rm na}/(I_{\rm bt+} + I_{\rm bt-})$ must remain close to unity, which means that the gain stays large.

The wide-field amacrine network cannot reach equilibrium at DC if the loop-gain is high because the bipolar-terminal signal is attenuated by the loop gain. According to the theory, $i_{\rm bt} = i_{\rm na} = i_{\rm bc}/(1+w)$, at DC, where w is the modulation. But the wide-field amacrine is excited by $i_{\rm bt}$ and inhibited by $wi_{\rm na}$. Hence, at DC, excitation and inhibition are balanced when $w = 1$. When w is large, WA is inhibited whenever the cone current deviates significantly from the mean, since $|i_{\rm na}| \approx |I_{\rm c} - I_{\rm cq}|$ and $|i_{\rm bt+}| \approx 0$. And hence $V_{\rm wa}$ increases, reducing current levels in the subtractor, which eventually drop below the leakage currents.

Low-frequency oscillations arise when the subtractor's current are overwhelmed by leakage currents. Assuming leakage from $V_{\rm dd}$ to $V_{\rm na}$ dominates, it charges the capacitor, reducing $I_{\rm na}$. If $I_{\rm na} = I_{\rm c} > I_{\rm cq}$ initially, inhibition decreases while excitation increases. Consequently, $V_{\rm wa}$ eventually repolarizes, restoring current levels in the subtractor, which discharges $V_{\rm wa}$ and initiates another cycle. Else, if $I_{\rm na} = I_{\rm c} < I_{\rm cq}$, both excitation and inhibition increase, but inhibition remains dominant. Hence, $V_{\rm wa}$ never repolarizes. Assuming leakage to ground dominates, on the other hand, $V_{\rm wa}$ still either oscillates or is permanently inhibited—it is just that the inequalities for $I_{\rm c}$ are reversed.

For nonstatic images, we would expect the instability to disappear—it did not show up in simulations. This was indeed the case. To determine how the loop-gain and the time-constant change, let us start with the KCL equation for node $V_{\rm na}$:

$$C\frac{dv_{\rm na}}{dt} + g_{\rm d}v_{\rm na} = A_{\rm wa}i_{\rm bt}.$$

Substituting $v_{\rm na} = i_{\rm na}/g_{\rm m}$ and $g_{\rm m} = \kappa I_{\rm na}/U_{\rm T}$ gives,

$$\frac{CU_T}{\kappa I_{\rm na}}\frac{dI_{\rm na}}{dt} + \frac{A_{\rm wa}(I_{\rm bt+} + I_{\rm bt-})}{I_{\rm na}}\frac{U_T}{\kappa V_{\rm E}}i_{\rm na} = A_{\rm wa}i_{\rm bt}.$$

As the rectifier ensures that $i_{\rm bt} = i_{\rm c} - i_{\rm na}$, we obtain the following differential equations for $i_{\rm na}$ and $i_{\rm bt}$:

$$\begin{aligned}\frac{\tau_{\rm na}}{A_{\rm wa}}\frac{di_{\rm na}}{dt} + (1+\epsilon)i_{\rm na} &= i_{\rm c},\\ \frac{\tau_{\rm na}}{A_{\rm wa}}\frac{di_{\rm bt}}{dt} + i_{\rm bt} &= \frac{\tau_{\rm na}}{A_{\rm wa}}\frac{di_{\rm c}}{dt} + \epsilon i_{\rm c},\end{aligned}$$

where $\tau_{\rm na} \equiv CU_T/\kappa I_{\rm na}$ and $\epsilon \equiv (U_T/\kappa V_{\rm E})((I_{\rm bt+} + I_{\rm bt-})/I_{\rm na}) \ll 1$. The differential equation for $i_{\rm bt}$ was obtained by substituting $i_{\rm na} = i_{\rm c} - i_{\rm bt}$ into the $i_{\rm na}$ equation.

These circuit equations differ from the model equations (Equations 4 and 5)—the circuit's cone current i_{c}, which is the analog of the model's bipolar cell current i_{bc}, is not modulated. Only the time-constants are modulated—due to the changing drain-conductance—much as in Liu's approach [29]. Setting $A_{\mathrm{wa}} = \omega\tau_{\mathrm{na}}$ equalizes the driving forces for these two equations when the temporal frequency is ω, making the amplitudes of i_{bt} and i_{na} the same.

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. An example is shown in Figure 9—swapping red and green or yellow and blue produces the other three sequences. Using Visio1 as a frontend, direction-selective (DS) cells can be built simply by wiring up these four distinct receptive fields, using virtual connections [10], as shown in Figure 10. What exactly do such DS cells compute? And what is the effect of temporal adaptation?

To compute direction of motion, we can project the gradient in spacetime $\nabla I(x,t) \equiv (\partial I/\partial x, \partial I/\partial t)$ onto the vector (λ, τ), where the x-axis is perpendicular to the edge, which moves in the positive x direction with velocity $\nu \equiv (\partial I/\partial t)/(\partial I/\partial x)$. That is,

$$(\lambda,\tau)\cdot\nabla I(x,t) = \lambda\frac{\partial I}{\partial x} + \tau\frac{\partial I}{\partial t} = (\lambda+\tau\nu)\frac{\partial I}{\partial x}.$$

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

We can maximize the projection by adapting either the length constant λ or the time-constant τ to make (λ, τ) point along the gradient. To do this, we must equalize the spatial and temporal components of the projection:

$$\left|\lambda\frac{\partial I}{\partial x}\right| = \left|\tau\frac{\partial I}{\partial t}\right| \Leftrightarrow \frac{\lambda}{\tau} = \left|\frac{\partial I/\partial t}{\partial I/\partial x}\right| \equiv \nu.$$

In so doing, we obtain a response of $2\lambda(\partial I/\partial x)$ in the prefered direction—independent of speed—and the response in the null-direction disappears. The time-constant adaptation version of this strategy is accomplished by my inner retina temporal adaptation mechanism, which acts to equalize highpass and lowpass responses (see Figure 4a).

Translating the bandpass-filtered image $J(x,t)$ by $\lambda/4$ in the negative x direction yields the spatial derivative—scaled by $\lambda/2\pi$. To obtain this approximate result, assume that the spatial filter passes the frequency $2\pi/\lambda$—and rejects all other frequencies. In that case, the derivative is

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

Highpass temporal filtering yields an image $K_{\mathrm{hp}}(x,t)$ that is the temporal derivative of a lowpass filtered version $K_{\mathrm{lp}}(x,t)$ of the original image—scaled by $\tau/2\pi$. To obtain this approximate result, assume that the highpass and lowpass filters have the same corner frequency $2\pi/\tau$. In that case, making use of the Laplace transform, the derivative is

$$\mathcal{L}\left\{\frac{\partial K_{\mathrm{lp}}}{\partial t}\right\} = sH_{\mathrm{lp}}(s)J(x,s) = \frac{2\pi}{\tau}H_{\mathrm{hp}}(s)J(x,s) = \frac{2\pi}{\tau}\mathcal{L}\{K_{\mathrm{hp}}(x,t)\}$$

since the filters' transfer functions are related by $H_{\mathrm{hp}}(s) = \tau s H_{\mathrm{lp}}(s)/2\pi$, if the frequency is in Hz. Hence, we obtain an exact temporal derivative if we replace the original image with the lowpass filtered one.

Therefore, to obtain the projection of the spacetime gradient $\nabla I(x,t)$ onto (λ,τ), pass the translated bandpass-spatial-filtered image through a lowpass temporal filter and the untranslated image through a highpass temporal filter and sum their outputs together (see Figure 10). And maximize the projection by equalizing the output amplitudes of these two filters through temporal adaptation. This adaptation makes the retina exquisitely sensitive to motion by matching the inner retina's delay to the time-of-flight across outer-retina receptive fields. I implemented this algorithm by wiring up Visio1's ganglion cells to silicon neurons on a second neuromorphic chip [9] as shown in Figure 10. I measured clear preferences for contrast polarity and motion direction for speeds spanning one decade, even though temporal adaptation was not working correctly.

This retinomorphic motion algorithm is a practical version of Watson and Ahumada's Hilbert-transform-based model of human visual-motion sensing [44]. 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 [1] closely-

related spatiotemporal energy model is quite similar [20, 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 [22, 39]. 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.

Acknowledgements

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 Whitaker 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. Adelson, E. H. and J. R. Bergen: 1985, 'Spatiotemporal Energy Models of the Perception of Motion'. *J Opt Soc Am* **2**(2), 284–299.

2. Andreou, A. G. and K. A. Boahen: 1989, 'Synthetic Neural Circuits using Current-Domain Signal Representations'. *Neural Computation* **1**, 489–501.
3. Andreou, A. G. and K. A. Boahen: 1994a, 'A 48,000 pixel, 590,000 transistor silicon retina in current-mode subthreshold CMOS'. In: *Proc. 37th Midwest Symposium on Circ. and Sys.* Lafayette, Louisiana, pp. 97–102.
4. Andreou, A. G. and K. A. Boahen: 1994b, 'Neural Information Processing (II)'. In: M. Ismail and T. Fiez (eds.): *Analog VLSI Signal and Information Processing.* McGraw-Hill, Chapt. 8.
5. Andreou, A. G. and K. A. Boahen: 1996, 'Translinear Circuits in Subthreshold MOS'. *J. Analog Integrated Circ. Sig. Proc.* **9**, 141–166.
6. Atick, J. and N. Redlich: 1992, 'What does the retina know about natural scene'. *Neural Computation* **4**(2), 196–210.
7. Boahen, K. A.: 1996a, 'Retinomorphic Vision Systems'. In: *Microneuro'96: Fifth Int. Conf. Neural Networks and Fuzzy Systems.* Los Alamitos CA.
8. Boahen, K. A.: 1996b, 'Retinomorphic Vision Systems: Reverse Engineering the Vertebrate Retina'. Ph.D. thesis, California Institute of Technology, Pasadena CA.
9. Boahen, K. A.: 1997, 'The Retinomorphic Approach: Pixel-Parallel Adaptive Amplification, Filtering, and Quantization'. *Analog Integr. Circ. and Sig. Proc.* **13**, 53–68.
10. Boahen, K. A.: 1999a, 'Massive Connectivity Between Neuromorphic Chips Using Address-Events'. *In preparation.*
11. Boahen, K. A.: 1999b, 'A Throughput-On-Demand Address-Event Transmitter for Neuromorphic Chips'. In: *Conference on Advanced Research in VLSI.* Los Alamitos CA.
12. Boahen, K. A. and A. Andreou: 1991, 'A contrast-sensitive retina with reciprocal synapses'. In: J. E. Moody (ed.): *Advances in neural information processing 4*, Vol. 4. San Mateo CA.
13. Cohen, E. and P. Sterling: 1991, 'Microcircuitry related to the receptive field center of the on-beta ganglion cell'. *J Neurophysiol* **65**(2), 352–9.
14. Croner, L. J. and E. Kaplan: 1995, 'Receptive fields of P and M ganglion cells across the primate retina.'. *Vision Res* **35**(1), 7–24.
15. Croner, L. J., K. Purpura, and et al: 1984, 'Response variability in retinal ganglion cells of primates'. *Proc Natl Acad Sci U S A* **90**(17), 8128–30.
16. Delbruck, T. and C. A. Mead: 1994, 'Photoreceptor Circuit with Wide Dynamic Range'. In: *Proceedings of the International Circuits and Systems Meeting.*
17. Derrington, A. M. and P. Lennie: 1984, 'Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque'. *J Physiol (Lond)* **357**, 219–40.
18. Eckert, M. P. and G. Buchsbaum: 1993, 'Efficient coding of natural time-varying images in the early visual system'. *Phil. Trans. Royal Soc. Lond. Biol* **339**(1290), 385–395.
19. Etienne-Cummings, R., J. Van der Spiegel, and P. Mueller: 1999, 'Hardware implementation of a visual-motion pixel using oriented spatiotemporal neural filters'. *IEEE Transactions on Circuits & Systems II-Analog & Digital Signal Processing* **46**(9), 1121–36.
20. Etienne-Cummings, R., J. van der Spiegel, P. Mueller, and et al: 1997, 'VLSI Implementation of Cortical Visual Motion Detection using an Analog Neural Computer'. In: D. Touretzky (ed.): *Advances in Neural Information Processing Systems 7*, Vol. 7. San Mateo CA.

21. Freed, M. A. and P. Sterling: 1988, 'The ON-alpha ganglion cell of the cat retina and its presynaptic cell types'. *J Neurosci* **8**(7), 2303–20.
22. Horn, B. K. P. and B. G. Schunck: 1981, 'Determining optical flow'. *Artif. Intell.* **17**, 185–204.
23. Kamermans, M. and F. Werblin: 1992, 'GABA-mediated positive autofeedback loop controls horizontal cell kinetics in tiger salamander retina'. *J Neurosci* **12**(7), 2451–63.
24. Kaplan, E. and R. M. Shapley: 1982, 'X and Y cells in the lateral geniculate nucleus of macaque monkeys'. *J Physiol (Lond)* **330**, 125–43.
25. Kolb, H. and R. Nelson: 1985, 'Functional Neurocircuitry of Amacrine cells in the Cat Retina'. In: A. Gallego and P. Gouras (eds.): *Neurocircuitry of the Retina: A Cajal Memorial*. Elsevier, New York, pp. 215–232.
26. Kolb, H. and R. Nelson: 1993, '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.
27. Kuffler, S. W.: 1953, 'Discharge Patterns and Functional Organization of Mammalian Retina'. *J. Neurophysiol.* **16**, 37–68.
28. Leventhal, A. G. and R. W. Rodieck: 1981, 'Retinal ganglion cell classes in the Old World monkey: morphology and central projections'. *Science* **213**(4512), 1139–42.
29. Liu, S.: 1996, 'Silicon Model of Motion Adaptation in the Fly'. In: *3rd UCSD-Caltech Symposium on Neural Computation*.
30. Liu, S.: 1998, 'Silicon Retina with Adaptive Filtering Properties'. In: *Advances in Neural Information Processing Systems*. San Mateo CA.
31. Maguire, G., P. Lukasiewicz, and et al.: 1989, 'Amacrine cell interactions underlying the response to change in the tiger salamander retina.'. *J Neurosci* **9**(2), 726–35.
32. Mead, C. A.: 1989, *Analog VLSI and Neural Systems*. Reading MA: Addison Wesley.
33. Pavasović, A., A. G. Andreou, and C. R. Westgate: 1994, 'Characterization of Subthreshold MOS Mismatch in Transistors for VLSI Systems'. *J Analog Integrated Circ & Sig Proc* **6**, 75–84.
34. Polyak, J. L.: 1941, *The Retina*. Chicago IL: Chicago Press.
35. Rodieck, R. W.: 1988, 'The Primate Retina'. *Comp. Primate Biol.* **4**, 203–278.
36. Shapley, R. and J. D. Victor: 1979, 'Nonlinear Spatial Summation and the Contrast Gain Control of Cat Retinal Ganglion Cells'. *J. Physiol.* **290**, 141–161.
37. Shapley, R. and J. D. Victor: 1981, 'How Contrast Gain Control Modifies the Frequency Responses of Cat Retinal Ganglion Cells'. *J. Physiol.* **318**, 161–179.
38. Smith, R. G.: 1995, 'Simulation of an anatomically define local circuit — The cone-horizontal cell network in cat retina'. *Visual Neurosci.* **12**(3), 545–561.
39. Tanner, J. and C. Mead: 1989, 'Optical Motion Sensor'. In: *Analog VLSI and Neural Systems*. Reading MA: Addison-Wesley, Chapt. 14, pp. 229–255.
40. Temes, G. C. and W. H. Ki: 1987, 'Fast CMOS Current Amplifier and Buffer Stage'. *Electronics Letters* **23**(13).
41. Victor, J. D.: 1987, 'The Dynamics of the Cat Retinal X Cell Centre'. *J. Physiol.* **386**, 219–246.
42. Vittoz, E. and X. Arreguit: 1993, 'Linear networks based on transistors'. *Electronics Letters* **29**, 297–299.
43. Watanabe, M. and R. W. Rodieck: 1989, 'Parasol and midget ganglion cells of the primate retina.'. *J Comp Neurol* **289**(3), 434–54.

44. Watson, A. B. and A. J. Ahumada, Jr: 1985, 'Model of Human Visual-Motion Sensing'. *J Opt Soc Am* **2**(2), 322–342.