Analog Integrated Circuits and Signal Processing, ??, 1–14 (199?)

The Retinomorphic Approach: Pixel-Parallel Adaptive Amplification, Filtering, and Quantization

KWABENA A. BOAHEN

Physics of Computation Laboratory, MS 136-93, California Institute of Technology, Pasadena CA 91125

Received May 1, 1996. Revised ???? ?, 1996.
Editor: T. Lande

Abstract. I describe a vision system that uses neurobiological principles to perform all four major operations found in biological retinae: (1) continuous sensing for detection, (2) local automatic gain control for amplification, (3) spatiotemporal bandpass filtering for preprocessing, and (4) adaptive sampling for quantization. All four operations are performed at the pixel level. The system includes a random-access time-division multiplexed communication channel that reads out asynchronous pulse trains from a 64×64 pixel array in the imager chip, and transmits them to corresponding locations on a second chip that has a 64×64 array of integrators. Both chips are fully functional. I compare and contrast the design principles of the retina with the standard practice in imager design and analyze the circuits used to amplify, filter, and quantize the visual signal, with emphasis on the performance trade-offs inherent in the circuit topologies used.

Keywords: retinomorphic, neuromorphic, local gain control, spatiotemporal filtering, pixel parallel, adaptive neuron

1. Smart-Pixel Arrays

The migration of sophisticated signal processing down to the pixel level is driven by shrinking feature sizes in CMOS technology, allowing higher levels of integration to be achieved [1], [2]. New pixel-parallel architectures are required to take advantage of the increasing numbers of transistors available [3]. Inspired by the pioneering work of Mahowald and Mead [4], I describe in this paper a retinomorphic vision system that addresses this need by mimicking biological sensory systems.

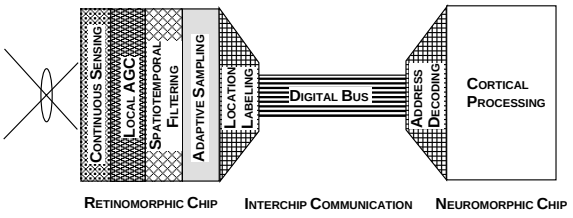
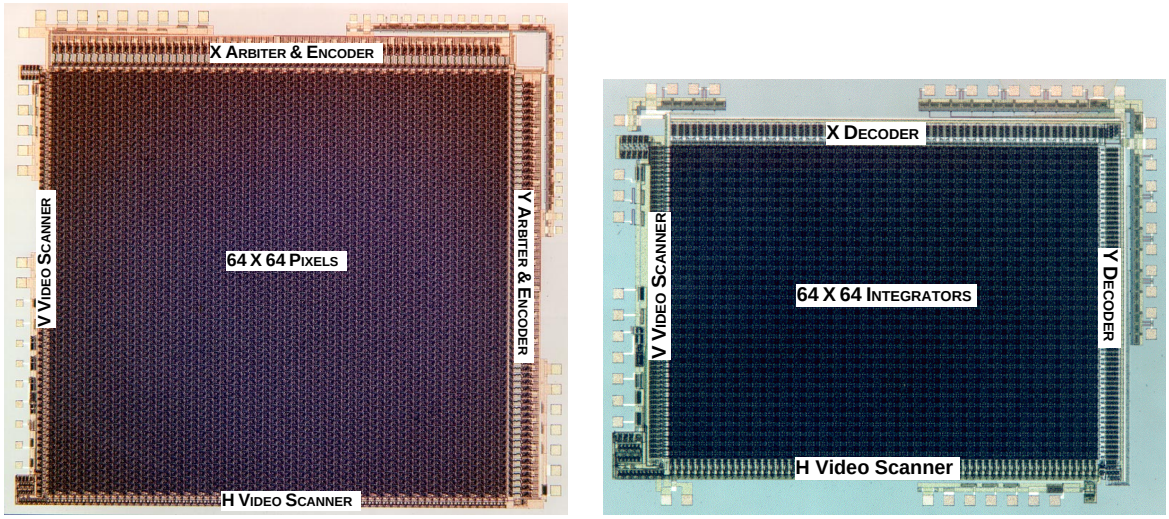


Fig. 1. System concept. The retinomorphic chip acquires, amplifies, filters, and quantizes the image. All these operations are performed at the pixel level. The interchip communication channel reads out asynchronous digital pulses from the pixels by transmitting the location of pulses as they occur. A second neuromorphic chip decodes these address events, and recreates the pulses.

In particular, my approach uses the system architecture and neurocircuitry of the nervous system as a blueprint for building integrated low-level vision systems—systems that are *retinomorphic* in a literal sense [5], [3]. Morphing of biological wetware into silicon-based hardware results in sensory systems that maximize information uptake from the environment, while minimizing redundancy in

Table 1. Specifications of the two-chip retinomorphic system. L is the minimum feature size, which was 2μm for this process; S/s is samples per second.

Specification	Imager	Postprocessor
Technology	2-μm 2-poly 2-metal p-well	
Number of pixels	64×64	
Pixel size (L^2)	53×49	31.5×23
Transistors/pixel	32	8
Die size (mm^2)	8.1×7.4	5.1×4.0
Supply	5 V	
Dissipation (0.2 MS/s)	230 mW (total)	
Throughput	2 MS/s	



(a) (b)

Fig. 2. Die micrographs. (a) Retinomorphic focal-plane processor. The core of this chip is a 64×64 array of pixels arranged on a hexagonal grid. Pixels generate pulses and communicate the occurence of these pulses by signalling on the column and row request lines. The arbiters ensure that pulses are read out of the array one by one, in an orderly manner, by selecting one pixel at a time with the column and row select lines. The encoders generate the addresses of the selected row and column; this pair of binary words uniquely identifies the location of the pulse. (b) Postprocessor. The core of this chip is a 64×64 array of diode-capacitor integrators. We can feed short current pulses to any integrator in the array by supplying its row and column addresses to the decoders. We use the scanners (shift registers) to read out analog currents from the array for display on a video monitor.

their output; that achieve high levels of integration, by performing several functions within the same structure; and that offer robust system-level performance, by distributing computation across several pixels.

I describe a retinomorphic system that consists of two chips: a focal-plane image processor that adaptively amplifies, filters, and quantizes the visual signal at the pixel-level, and a postprocessor that has a two-dimensional array of integrators. The system concept is shown in Figure 1. Both chips are fully functional; specifications and die micrographs are shown in Table 1 and in Figure 2, respectively.

I outline the general design principles of the retina, and contrast them with standard engineering practice in Section 2. Having defined the retinomorphic approach to imager design, I describe a design for a retinomorphic pixel in Section 3, and present test results from the complete two-chip neuromorphic system in Section 4. The communication channel used to transmit the pulse trains from chip to chip is described in detail a companion paper [6]; a brief description is already available [7]. My concluding remarks are presented in Section 5.

Table 2. Retinal design principles versus electronic-imager design principles.

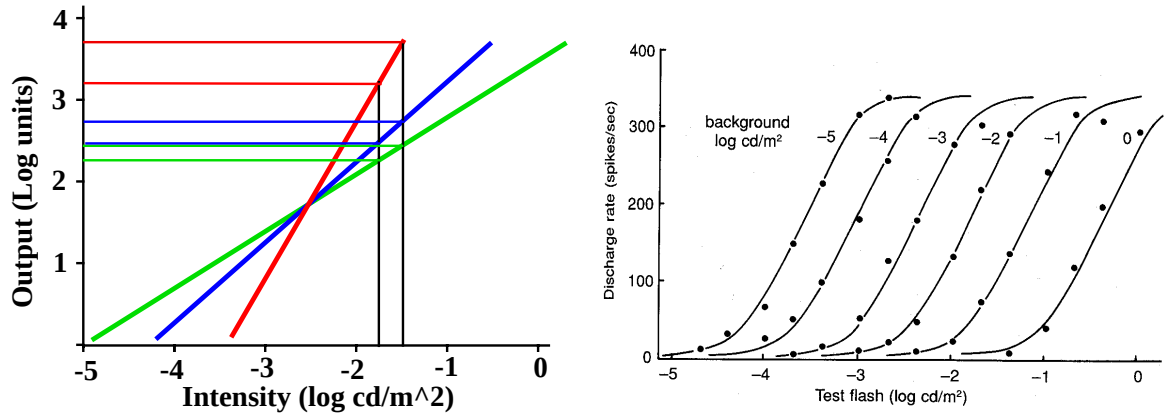
Operation	Standard	Retinal
Detection	Integrating	Continuous
Gain Control	Global	Local
Filtering	Allpass	Bandpass
Quantization	Fixed	Adaptive
Architecture	Serial	Parallel

2. Pixel-Parallel Processing

The functional and structural organization of the retina is radically different from that of standard human-engineered imagers. The principles of operation of the retina are outlined in Table 2; the principles of operation of standard imager technology also are listed for comparison.

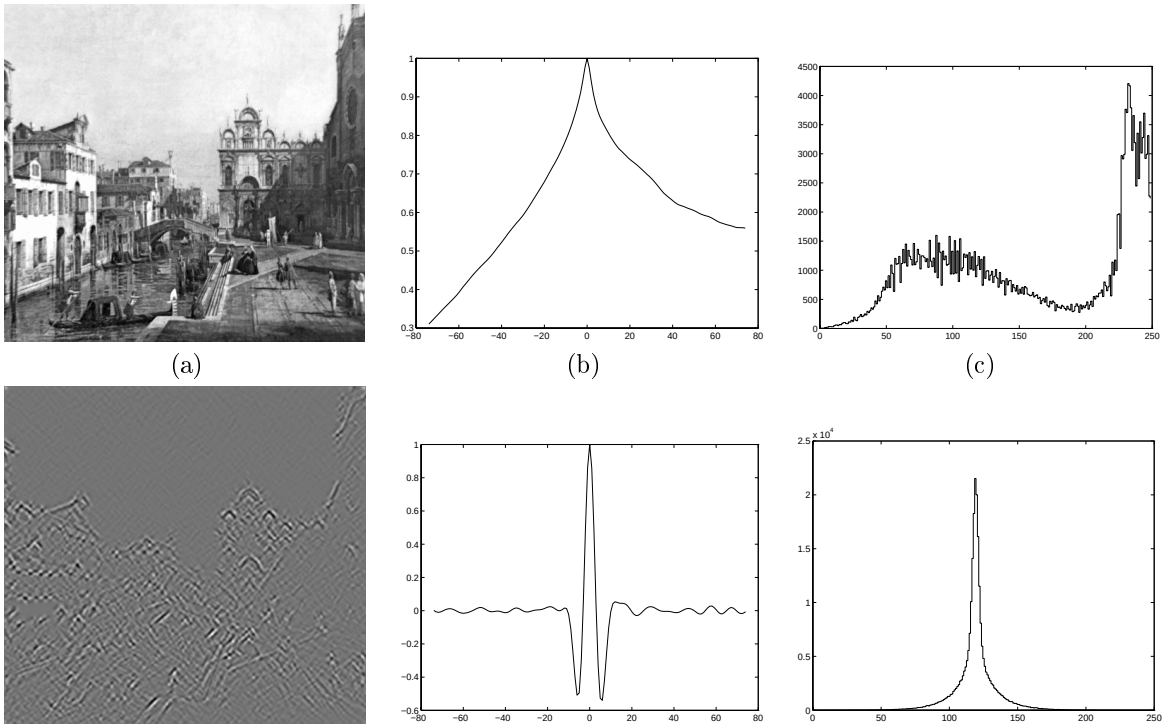
2.1. Sensing: Continuous Versus Integrating

Integrating detectors (e.g., charge-coupled devices (CCDs) [8] and photogates [9]) suffer from bloom-



(a) (b)

Fig. 3. Input–output transfer curves for light sensors. (a). As larger and larger input ranges are spanned, the slope decreases, and finer resolution is required to detect the same percentage change in the input signal. (b). Using transfer curves that can be centered at the local intensity level decouples dynamic range and resolution. Each curve spans only a 20-fold input range, since local variations in intensity are due primarily to changes in reflectivity: A black sheet has a reflectivity of 0.05, and a white sheet has a reflectivity of 0.95. These transfer curves were measured for the cat retina, and were reproduced from [19].



(a) (b) (c)

(d) (e) (f)

Fig. 4. Bandpass filtering. The top row shows the original 512 × 479 × 8-bit image (a), its autocorrelation (b), and its amplitude histogram (c). The bottom row shows the bandpass-filtered image (d), its autocorrelation (e), and its amplitude histogram (f). In the original image, pixels are highly correlated, and the correlation falls off slowly with distance. The distribution of amplitudes in the original image is broad and bimodal, due to the relatively bright overcast sky and the dark foreground objects. in contrast, the amplitude distribution for the filtered image is clustered around zero (119), and decays rapidly.

ing at high intensity levels and require a destructive readout (reset) operation. Continuous-

sensing detectors (e.g., photodiodes or phototransistors) do not bloom, and can therefore operate over a much larger dynamic range [10]. In addition, redundant readout or reset operations can be eliminated, with considerable power savings, because charge does not accumulate.

Continuous-sensing detectors have been shunned, however, because they suffer from gain and offset mismatches that give rise to salt-and-pepper noise in the image. However, Buhman et.al. have shown that the powerful learning capabilities of image-recognition systems can easily compensate for this fixed pattern noise [11].

The real benefit of using continuous sensors lies in their ability to perform analog preprocessing before quantizing the signal. A signal that takes on a discrete set of values at a discrete set of times (quantized in amplitude and time) carries less information than does a signal that takes on the full continuous spectrum of amplitudes and times. For instance, graded potentials in the nervous system can transmit information at the rate of 1650 bits per second—over four times the highest rate measured for spike trains [12].

The analog operations described in Sections 2.2 and 2.3 reshape the spectral distribution and the amplitude distribution of the analog signal, to transmit information efficiently through this bottleneck.

2.2. *Amplification: Local Versus Global Control*

Imagers that use global automatic gain control (AGC) can operate under only uniform lighting, because the 1000-fold variation of intensity in a scene with shadows exceeds their 8-bit dynamic range.[1] A charge-coupled device or photogate can achieve 12 bits (almost four decades) [9], and a photodiode or phototransistor can achieve 20 bits (six decades) [10], [13]—but the phototransistor's performance in the lowest two decades is plagued by slow temporal response. The dynamic range of the system's output, however, is limited by the cost of precision analog read-out electronics and A/D converters, and by video standards.

When AGC acts globally, the input dynamic range matches the output dynamic range, and the only way to extend the input range is to extend the output dynamic range. In practice, we must reduce the noise floor to improve resolution.

As shown in Figure 3, local AGC decouples dynamic range and resolution, extending the input dynamic range by mapping different parts of the input range to the limited output range, depending on the local intensity level. This solution is beneficial if the resolution required to discriminate various shades of gray (1 in 100 for the human visual system) is poorer than the resolution required to span the range of all possible input levels (at least 1 in 100,000 for the photopic range of human vision).

2.3. *Preprocessing: Bandpass Versus Allpass*

On average, natural images have a $1/f^2$ power spectrum for both spatial and temporal frequency [14], [15], whereas noise, due to quantum fluctuations, has a flat spectrum. Consequently, imagers that transmit the full range of frequencies present pass on useless information at high frequencies, where the signal-to-noise is poor, and pass on redundant information at low frequencies, where the signal-to-noise is good. Bandpass spatiotemporal filtering rejects the wideband noise, and attenuates the redundant low-frequency signals; this strategy is the optimal one for removing redundancy in the presence of white noise [16], [17], [18].

I illustrate in Figure 4b,d the redundancy reduction achieved, for a typical outdoor scene, by computing the correlation between pixel values.[2] The correlation is over 40% for pixels 60 pixels apart in the raw image. In the filtered image, pixels more than 10 pixels apart have less than 5% correlation. Comparison of the amplitude histograms before and after filtering (Figure 4c,f) demonstrates that bandpass filtering has two additional benefits.

First, bandpass filtering results in a sparse output representation. For our sample image, 24.4% of the pixels fall within $\pm 0.39\%$ of the full-scale range (i.e., $\pm$1LSB at 8-bit resolution); 77.5% of them fall within $\pm 5\%$ (i.e., ± 13 at -127 to $+127$ amplitude range). Hence, if we choose to ignore

amplitudes smaller than 5%, we need to transmit only 22.5% of the pixels. In practice, the degree of sparseness will depend on the cut-off frequency of the bandpass filter. Although rejection of high frequencies introduces some redundancy, this rejection is necessary to protect the signal from noise that is introduced by the signal source or by the circuit elements.

Second, bandpass filtering results in a unimodal amplitude distribution that falls off exponentially. For our sample image, the distribution is fit by a sum of two exponentials that change by a factor of $e = 2.72$ whenever the amplitude changes by 2.5 and 14.0, on a ± 128 scale; the rapidly decaying exponential starts out 4.5 times larger. Empirical observations confirm that this simple model holds for a wide range of images.

In contrast, the distribution of raw intensity values is difficult to predict, because gross variations occur from scene to scene, due to variations in illumination, image-formation geometry (surface and light-source orientation), and shadows [20]. These slowly changing components of the image are removed by local AGC and bandpass filtering. When the bandpass characteristics are fixed and the intensity is normalized, the parameters of the amplitude distribution vary much less, and the quantizer can exploit this invariance to distribute its codes more effectively.

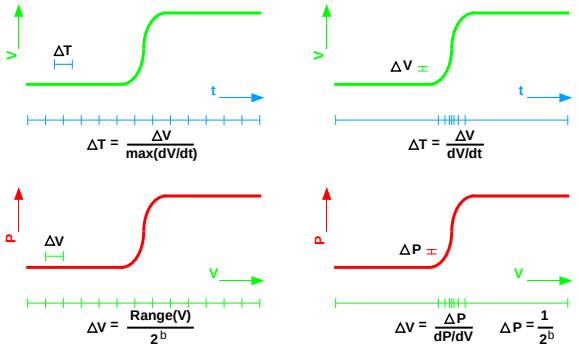


Fig. 5. Quantization in time and amplitude. Top row: Time intervals (ΔT) are set to match the maximum rate of change (left column). The signal is sampled repeatedly, even when $dV/dt \approx 0$—that is, when the change is insignificant (oversampling). Instead of fixing the time step, it is more efficient to fix the voltage step (ΔV), and to adapt the time intervals dynamically to achieve this change in voltage, as shown on the right. Bottom row: Amplitude intervals (ΔV) are uniformly distributed. The signal is sampled repeatedly, even though $dP/dV \approx 0$—that is, although the probability that the input amplitude falls in this interval is negligible. Instead of fixing the voltage step, it is more efficient to target a certain change in the cumulative probability ($\Delta P = 2^{-b}$, where b is the number of bits per sample), and to choose voltage intervals statistically to achieve this change in probability, as shown on the right.

2.4. *Quantization: Adaptive Versus Fixed*

The quantization intervals of traditional A/D converters are set to match the maximum rate of change and the smallest amplitude, as shown in Figure 5. This uniform quantization is optimum only when high frequencies dominate and all amplitudes are equally likely. As we have seen, neither case applies to natural scenes: the power spectrum decays with frequency, as in $1/f^2$, and the amplitude probability density decays exponentially. Therefore, uniform quantizers produce numerous redundant samples, because changes in the signal are relatively rare [15], and underutilize their large amplitude codes, because these signal amplitudes occur rarely in natural scenes [20].

Assuming that temporal changes are due primarily to motion, we can estimate the amount of redundancy from the spatial-frequency power spectrum and from the velocity distribution. The velocity distribution, measured for movies and amateur videos, is dominated by low velocities and falls off with a power law of 3.7 [15]. High velocities will be even more drastically attenuated in an active vision system that compensates for global motion, and that tracks objects [21]. After bandpass filtering, signals that change gradually over space are eliminated and rapid changes occur only rarely and over much more restricted areas.

Due to the absence of high speeds and of nonlocal intensity varaitions, the imager's output signals rarely change rapidly. Consequently, adapting the sampling rate to the rate of change of the signal greatly reduces the number of samples produced. Alternatively, this adaptation allows higher temporal bandwidths to be achieved for a given mean sampling rate.

Using the amplitude distribution of our bandpass-filter sample image, we can calculate the probability of failing to discriminate between a pair of samples drawn from that distribution: It is 0.0384 when the 2^8 quantization levels are uniformly distributed—an order of magnitude bigger

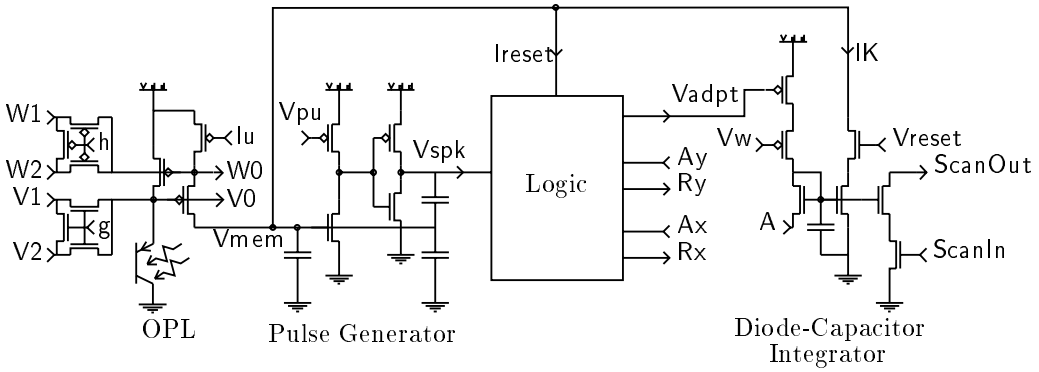


Fig. 6. Retinomorphic pixel. The outer-plexiform–layer (OPL) circuit performs spatiotemporal bandpass filtering and local automatic gain control (AGC) using two current-spreading networks. Nodes V0 and W0 are connected to their six nearest neighbors on a hexagonal grid by the delta-connected transistors. The OPL's output current is converted to pulse frequency by the pulse generator. The logic circuit communicates the occurrence of a pulse (Vspk) to the chip periphery using the row and column request and select lines (Ry/Ay and Rx/Ax), turns on Ireset to terminate the pulse, and takes Vadapt low, to feed a current pulse to the integrator; the logic circuit is described elsewhere [7], [6]. The integrator's output current (IK) is subtracted from the input to the pulse generator; the device in series with the integrator's output, whose gate is tied to a fixed bias Vreset, is used to isolate the integrator from the rapid volatges swings that occur at Vmem when spikes occur. The two series-connected transistors on the right are used to scan out the integrator's output for display on a video monitor.

than the the minimum confusion rate of $1/256 = 0.0039$, which occurs when the quantization levels

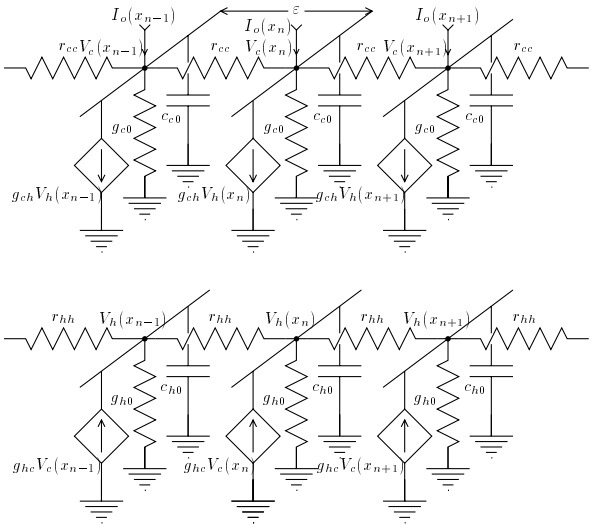


Fig. 7. Linear circuit model of the outer plexiform layer (OPL) circuit. Two resistive networks model the intercone and the inter–horizontal-cell electrical synapses (gap junctions), and transconductances model the reciprocal chemical synapses between cones and horizontal cells. ϵ is the pixel size; it relates the modeled quantities, which are in current per unit area, sheet resitance, conductance per unit area, and capacitance per unit area, to quantities in the real circuit. The model is analyzed in the continuum limit, where $\epsilon \to 0$.

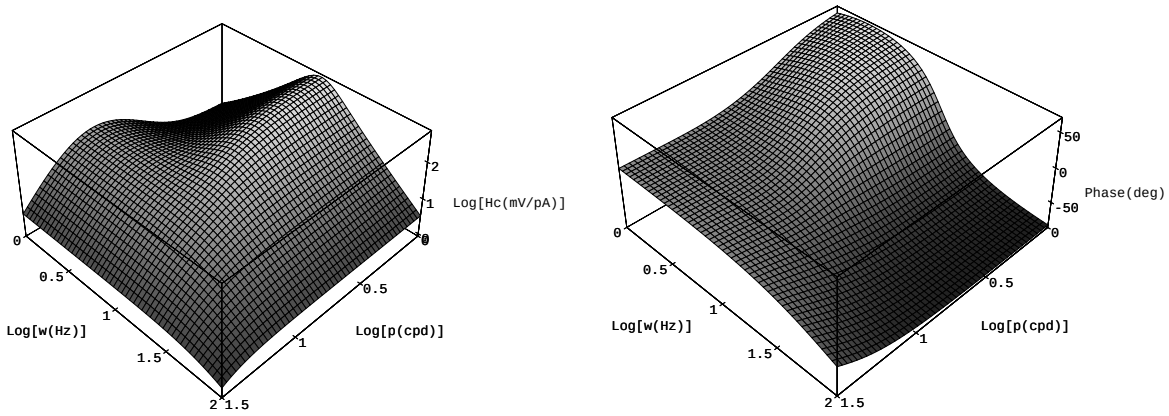
are chosen to make it equally likely that we will draw a sample from each interval. In fact, the confusion rate of 0.0384 can be achieved with just $\log_2(1/0.0384) = 4.7$bits per sample if the quantization levels are optimally distributed.

A quantizer that assigns its codes to probable amplitudes, rather than to improbable ones, maximizes the probability of discriminating between any two amplitude levels drawn from the input distribution; thus, information is maximized when all codes are equiprobable [22].

2.5. *Architecture: Parallel Versus Serial*

In addition to differing in the aforementioned design principles, biological and human-made vision systems use radically different architectures. The retina performs the four operations listed in Table 2 in a pixel-parallel fashion, whereas most synthetic imagers perform only detection in the pixel; the few that also amplify and quantize the signal perform these operations pixel-serially, and set the gain, sampling rate, and step size to be the same for all pixels [9], [23], [24].

In sharp contrast to human-engineered imagers, the retina adapts its gain, sampling rate, and step size locally, to maximize information uptake; it



(a) (b)

Fig. 8. Spatiotemporal sensitivity of linear OPL model. These three-dimensional surface plots show (a) magnitude and (b) phase versus spatial frequency (p) and temporal frequency (w).

also whitens the signal in space and time, to minimize the redundancy in its output samples. I present a pixel design in Section 3 that performs continuous sensing, bandpass spatiotemporal filtering, local AGC, and adaptive quantization.

3. A Retinomorphic Pixel

I designed the pixel circuit shown in Figure 6 using the retinomorphic approach; it senses, amplifies, filters, and quantizes the visual signal. In general terms, this retinomorphic pixel operates as follows.

The transducer is a vertical bipolar transistor; its emitter current is proportional to the incident light intensity [10]. Two current-spreading networks [5], [25], [26], [27] diffuse the photocurrent signals over time and space; the first layer (node V0) excites the second layer (node W0), which reciprocates by inhibiting the first layer. The result is a spatiotemporally bandpass-filtered image [28], [29], [30]. The second layer computes a measure of the local light intensity, and feeds back this information to the input layer, where the intensity information is used to control light sensitivity. The result is local AGC [5].

A pulse generator converts current from the excitatory layer into pulse frequency. The diode-capacitor integrator computes a current that is proportional to the short-term average of the pulse frequency; this current is subtracted from the pulse generator's input. The difference becomes larger as the input changes more rapidly, so pulses are fired more frequently. Hence, the more rapidly the input changes, the more rapidly the pulse generator fires.

Adding a fixed charge quantum to the integrating capacitor produces a multiplicative change in current—due to the exponential current–voltage relation in subthreshold. Hence, the larger the current level, the larger the step size. The result is adaptive quantization. I also use the diode-capacitor integrator in the postprocessor to integrate the pulses, and to reconstruct the current level that was encoded into pulse frequency.

I discuss the behavior of these circuits in detail in Sections 3.1 through 3.3. The performance of the solutions adopted are analyzed, with emphasis on the tradeoffs inherent in the circuit topologies chosen.

3.1. Preprocessing: Spatiotemporal Bandpass

Using the small-signal model of the coupled–current-spreading networks shown in Figure 7, I found that

$$I_o + \nabla^2 V_c / r_{cc} = g_{c0} V_c + c_{c0} \dot{V}_c + g_{ch} V_h, \quad (1)$$

$$g_{hc} V_c + \nabla^2 V_h / r_{hh} = g_{h0} V_h + c_{h0} \dot{V}_h, \quad (2)$$

in the continuum limit. Here, V_c is the voltage in the excitatory network, which models retinal cones; V_h is the voltage in the inhibitory network, which models retinal horizontal cells (HCs); and I_o is the photocurrent [30]. These functions are now continuous functions of space, (x, y),

and of time, t; $\nabla^2 f$ is the Laplacian of f (i.e., $\partial^2 f/\partial x^2 + \partial^2 f/\partial y^2$), and $\dot{f}$ is the temporal derivative of f (i.e., $\partial f/\partial t$). Models similar to this one were proposed and analyzed in [28], [29], [31].

Assuming infinite spatial extent and homogeneous initial conditions, I take Fourier transforms in space and time, and solve the equations to obtain

$$\tilde{H}_c = \frac{1}{g_{ch}} \frac{\ell_h^2 \rho^2 + i\tau_h \omega + \epsilon_h}{(\ell_c^2 \rho^2 + i\tau_c \omega + \epsilon_c)(\ell_h^2 \rho^2 + i\tau_h \omega + \epsilon_h) + 1},$$

where $\tilde{H}_c(\rho, \omega) \equiv \tilde{V}_c / \tilde{I}_o$. $\tilde{f}(\rho_x, \rho_y, \omega)$ denotes the Fourier transform of $f(x, y, t)$, $\rho = \sqrt{(\rho_x^2 + \rho_y^2)}$ is spatial frequency, and ω is temporal frequency (both in radians) [30]. Here, $\tau_c = c_{c0}/g_{ch}$ and $\tau_h = c_{h0}/g_{hc}$ are the time constants associated with the HC-to-cone coupling and with the cone-to-HC coupling, respectively; $\ell_c = (r_{cc} g_{ch})^{-1/2}$ and $\ell_h = (r_{hh} g_{hc})^{-1/2}$ are the space constants of the decoupled networks, with transconductances replaced by conductances to ground; and $\epsilon_c = g_{c0}/g_{ch}$ and $\epsilon_h = g_{h0}/g_{hc}$ are the ratios of leakage conductance to the transconductance. The reciprocals of ϵ_c and ϵ_h are the open-loop voltage gains from the HC to the cone, and vice versa.

The spatiotemporal frequency response of the excitatory cone network obtained from this analysis is plotted in Figure 8. The set of parameters values used was as follows: $\ell_c = 0.05°$, $\ell_h = 0.2°$, $\tau_c = 30$ms, $\tau_h = 200$ms, $\epsilon_c = 0.3$, $\epsilon_h = 0.1$, $g_{ch} = 0.2$pA/mV. Observe that the temporal-frequency response is bandpass at low spatial frequencies (flicker sensitivity), and that the spatial-frequency response is bandpass at low temporal frequencies (grating sensitivity).

Fig. 9. CCD Camera (top row) versus retinomorphic imager (bottom row) under variable illumination. The CCD camera has global automatic gain control (AGC), whereas the retinomorphic imager has local AGC and performs bandpass filtering.

However, the overall response is not linearly separable; that is, it is not simply the composite of a bandpass spatial filter and a bandpass temporal filter. The spatial tuning becomes lowpass at high temporal frequencies, and the temporal tuning becomes lowpass at high spatial frequencies [30]. The same behavior is observed in physiological data measured from cats [32] and psychophysical data measured from humans [33].

There are tradeoffs among small low-frequency response, large dynamic range, and high sensitivity. The circuit requires a high-gain cone-to-HC synapse (i.e., small ϵ_h) to attenuate the cone's response to low spatial and temporal frequencies, since $\tilde{H}_c(0,0) = \epsilon_h / g_{ch}$. However, increasing the gain of the cone-to-HC synapse decreases the dynamic range of the cone, (i.e., $V_c < \epsilon_h V_{\text{lin}}$, where V_{lin} is the linear range.) It also makes the circuit ring since, $Q = (\epsilon_c \sqrt{\frac{\tau_h}{\tau_c}} + \epsilon_h \sqrt{\frac{\tau_c}{\tau_h}})^{-1}$.

Smith and Sterling realized this constraint on the loop gain, and proposed using feedforward inhibition to second-order cells (bipolar cells) to attenuate the low frequency response [31]. Alternatively, to maintain temporal stability, we can decrease the gain of the HC-to-cone feedback synapse ($1/\epsilon_c$), or reduce the HC's time constant (τ_h). Unfortunately, both changes reduce the peak sensitivity of the cone $\tilde{V}(0, \hat{\omega}) = Q\sqrt{(\tau_h/\tau_c)}$. The circuit implementation shown in Figure 6 has high gain from the excitatory cone node (V0) to the inhibitory HC node (W0), giving it small DC response and high sensitivity, but poor temporal stability.

3.2. *Amplification: Local Automatic Gain Control*

I achieve local AGC by making the intercone conductance ($1/r_{cc}$) proportional to the local average of the photocurrent. This adaptation is realized in the circuit simply by the fact that (Vdd-V0) is equal to the sum of the gate–source voltages of two devices. The currents passed by these devices represent the activity in the inhibitory network, I_h, which is equal to the local average of the in-

tensity, and represent the activity of the excitatory network, I_c, which is equal to the Laplacian of the smoothed intensity profile (see Equation 2). Hence, by the extended translinear principle [25], the current that spreads in the excitatory network is proportional to the product, $I_c I_h$, of these currents. Since I_h scales with the intensity, the internode coupling in the excitatory cone network will scale accordingly [5].

It remains to show that the response of the excitatory cone network is proportional to the intercone resistance [34]. Closed-form solutions for the impulse response may be obtained in the one-dimensional space:

$$V_c(x) = r_{cc} I_o \frac{L}{2\sqrt{2}} e^{-|x|/L} \sin(|x|/L - \pi/4),$$

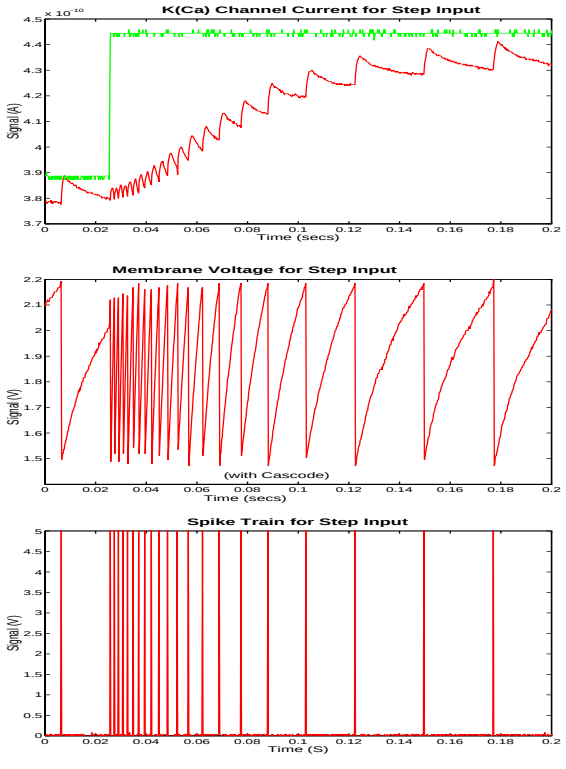


Fig. 10. Adaptive neuron's step response. Top: The neuron's input current and the integrator's output current. Middle: Input voltage ramping up between the reset (1.5V) and threshold levels (2.2V). Bottom: The spike train. The difference between the input current and the integrator's output current ramps up the input voltage as the surplus current charges the input capacitance.

where $L = \sqrt{\ell_c \ell_h} = (r_{cc} g_{ch} r_{hh} g_{hc})^{-1/4}$ is the effective space constant of the dual-layer network. These solutions are valid for the case $g_{c0} = g_{h0} = 0$, which is a fairly good approximation of the actual circuit. Linear system theory thus predicts that the gain of the cone is equal to the product of the space constant and the intercone coupling resistance.

This analysis also reveals that we compromise receptive-field size constancy by using the intercone coupling to implement local AGC, because the space constant depends on the intercone resistance: $L = (r_{cc} g_{ch} r_{hh} g_{hc})^{-1/4}$. Thus, as we increase r_{cc} to increase the gain, the receptive field contracts.

The images shown in Figure 9 demonstrate the effects of bandpass filtering and local AGC. These data are from the retinormorphic chip described in [5]; images of the same scenes acquired with a CCD camera are included for comparison [11]. [3] Bandpass filtering removes gradual changes in intensity and enhances edges and curved surfaces. It also reduces the variance of the amplitude distribution by setting uniform areas to the mean level (gray). Local AGC extends the dynamic range by boosting the gain in the dark parts of the scene. Thus, the retinomorphic chip picks up information in the shadows, whereas the output of the CCD camera is zero throughout that region.

Unfortunately, the retinomorphic chip's output is noisier in the darker parts of the image, due to the space constant decreasing with increasing gain. When the space constant decreases, wideband salt-and-pepper noise is no longer attenuated, because the cutoff frequency shifts upward. The dominant noise source is the poor matching among the small ($4L \times 3.5L$; where L, the minimum feature size, is 2μm) transistors used—it is not shot noise in the photon flux. Nevertheless, when it replaced the CCD as the front-end of a face-recognition system, the 90×90–pixel OPL chip improved the recognition rate from 72.5% to 96.3%, with 5% false positives, under variable illumination [11].

3.3. *Quantization: Adaptive Neuron Circuit*

I built an adaptive neuron circuit by taking a pulse generator and placing a diode-capacitor integra-

tor around it in a negative-feedback configuration (See Figure 6). The pulse-generation circuit has a high-gain amplifier (two digital inverters) with positive feedback (capacitive divider) [35]. The high-gain amplifier serves as a thresholding device, and the positive-feedback network provides hysteresis. In addition, there is a reset current (Ireset) produced by the logic circuit that terminates the spike.

The response of the adaptive neuron circuit to a 14% change in its input current is shown in Figure 10; these data demonstrate the adaptive sampling rate, the adaptive step size, and the integration of pulse trains by the diode-capacitor integrator. Other designs for adaptive neurons are described in [36], [37].

The diode-capacitor integrator is based on the well-known current-mirror circuit. A large capacitor at the input of the mirror integrates charge, and the diode-connected transistor leaks charge away. This circuit's temporal behavior is described by the following nonlinear differential equation in the current domain:

$$\frac{\mathrm{Q_T}}{I_{\mathrm{out}}(t)}\frac{dI_{\mathrm{out}}}{dt} = I_{\mathrm{in}}(t) - \frac{1}{\mathrm{A}}I_{\mathrm{out}}(t),$$

where $\mathrm{U_T} \equiv \mathrm{kT/q}$ is the thermal voltage, $\mathrm{A} = \exp(V_{\mathrm{A}}/\mathrm{U_T})$ is the current gain of the mirror, and $\mathrm{Q_T} \equiv \mathrm{CU_T}/\kappa$ is the charge required change the current by a factor of $e = 2.72$ [10], [38]. This circuit has a time-constant $\tau = \mathrm{Q_T}/I_{\mathrm{out}}(t)$, that is proportional to the current level due to exponential current–voltage relation in the subthreshold region.

The output produced by a periodic sequence of current pulses is

$$I_{\mathrm{out}}(t_0 + nT) = \frac{1}{1/\hat{I}_T + (1/I_{\mathrm{out}}(t_0) - 1/\hat{I}_T)(1+\alpha)^{-n}}, \quad (3)$$

immediately after the $(n+1)$th pulse; it decays as

$$I_{\mathrm{out}}(t) = \frac{I_{\mathrm{out}}(t_0 + nT)}{\frac{I_{\mathrm{out}}(t_0+nT)}{\mathrm{AQ_T}}(t - (t_0 + nT)) + 1},$$

during the interspike interval, $t_0 + nT < t < t_0 + (n+1)T$, where $\alpha \equiv (\exp(q_\alpha/\mathrm{Q_T}) - 1)$ is the percentage by which the output current is incremented by each spike, and $\hat{I}_T \equiv \alpha \mathrm{AQ_T}/T$ is the steady state [38].

We can use the current-mirror gain A to control the decay rate, because the time scale is set by $\tau = AQ_T/I_{\mathrm{out}}$. The fixed quantity of charge q_α supplied by each current pulse multiplies the current by $\exp(q_\alpha/\mathrm{Q_T})$, since it takes $\mathrm{Q_T}$ to change the current by a factor of $e = 2.72$. Hence, the incremental change in the output current caused by a spike is not fixed: It is proportional to the output current level at the time that the spike occurs. The peak output current levels attained immediately after each spike converge to $\hat{I}_T \equiv \alpha \mathrm{AQ_T}/T$ when $(1+\alpha)^{-n} \ll 1$. Therefore, the equilibrium output current level is proportional to the pulse frequency.

The complete adaptive neuron circuit is described by two coupled differential equations:

$$\mathrm{C_{mem}}\frac{dV_{\mathrm{mem}}}{dt} = I_{\mathrm{in}} - I_{\mathrm{K}} - \mathrm{Q_{th}}\delta(V_{\mathrm{mem}} - \mathrm{V_{th}}), \quad (4)$$

$$\frac{\mathrm{Q_T}}{I_{\mathrm{K}}}\frac{dI_{\mathrm{K}}}{dt} = \mathrm{q_\alpha}\delta(V_{\mathrm{mem}} - \mathrm{V_{th}}) - \frac{1}{\mathrm{A}}I_{\mathrm{K}}, \quad (5)$$

where I_{in} is the current supplied to node Vmem by the OPL circuit, and $\mathrm{C_{mem}}$ is the total capacitance connected to that node; I_{K} is the current subtracted from node Vmem by the integrator; $\mathrm{C_{Ca}}$ is the integrator's capacitance; $\mathrm{Q_T} = \mathrm{C_{Ca}U_T}/\kappa$ is the charge required to change the integrator's output current by a factor of $e = 2.72$; and $\mathrm{Q_{th}}$ is the repolarization charge (i.e., the charge that we must supply to Vmem to bring the latter from the reset level to the threshold level ($\mathrm{V_{th}}$)).

The parasitic coupling capacitance between Vmem and the integrator's input node is not included in these equations. This capacitance can have a large influence on the circuit's behavior [38]. In this particular design, however, the cascode device between the integrator's output and the pulse generator's input (tied to Vreset) eliminates virtually all coupling.

For a constant input current, these equations may be integrated to obtain

$$\mathrm{Q_{th}} = I_{\mathrm{in}}\Delta_n - \mathrm{AQ_T}\ln\left(\frac{I_{\mathrm{K}n}}{\mathrm{AQ_T}}\Delta_n + 1\right),$$

where $\Delta_n \equiv t_{n+1} - t_n$ is the interspike interval.

When adaptation is complete, the interspike intervals become equal, and we have $I_{\mathrm{K}n} = \alpha \mathrm{A} \mathrm{Q}_{\mathrm{T}} / \Delta_n$ (from Equation 3). Hence,

$$\Delta_n = (\mathrm{Q}_{\mathrm{th}} + \mathrm{A}\mathrm{q}_\alpha)/I_{\mathrm{in}} = \gamma \mathrm{Q}_{\mathrm{th}}/I_{\mathrm{in}}$$

(remember that $\mathrm{q}_\alpha = \mathrm{Q}_{\mathrm{T}} \ln(1+\alpha)$). This result is understood as follows. During the interspike interval, Δ_n, the input current must supply the charge Q_{th} to the capacitors tied to Vmem, and must supply the charge $\mathrm{A}\mathrm{q}_\alpha$ removed by the integrator, where q_α is the quantity of charge added to the integration capacitor by each spike. Notice that firing-rate adaptation reduces the firing rate by a factor of $\gamma \equiv 1 + \mathrm{A}\mathrm{q}_\alpha/\mathrm{Q}_{\mathrm{th}}$.

It is preferable to have $I_{\mathrm{K}}(t) < I_{\mathrm{in}}(t)$ for all t, because V_{mem} stays close to the threshold, making the latency shorter, and less variable, and keeping the integrator's output device in saturation. The circuit operates in this regime if $\gamma < 2/\alpha$ [38]. A tradeoff is imposed by my desire to operate in this regime: If we want a large adaptation-attenuation factor γ, we must use a small charge quantum q_α—and must turn up A to compensate—making the number of spikes required to adapt large.

4. Overall System Performance

The output of the postprocessor—after image acquisition, analog preprocessing, quantization, interchip communication, and integration of charge packets in the receiver's diode-capacitor integrators—is shown in Figure 11. The sparseness of the output representation is evident.

When the windmill moves, neurons at locations where the intensity is increasing (white region invading black) become active; hence, the leading edges of the white vanes are more prominent. These neurons fire more rapidly as the speed increases because they are driven by the temporal derivative. The time constant of the receiver's diode-capacitor integrator is intentionally set shorter than that of the sender, so temporal integration occurs at only high spike rates. This mismatch attenuates low frequency information, and results in an overall highpass frequency response that eliminates the fixed-pattern noise and enhances the imager's response to motion. The mean spike rate was 30Hz per pixel, and the two-chip system dissipated 190 mW at this spike rate.

5. Discussion

I have described the design and performance tradeoffs of a retinormorphic imager. This VLSI chip embodies four principles of retinal operation.

First, the imager adapts its gain locally to extend its input dynamic range without decreasing its sensitivity. The gain is set to be inversely proportional to the local intensity, discounting gradual changes in intensity and producing an output that is proportional to contrast [5]. This adaptation is effective because lighting intensity varies by six decades from high noon to twilight, whereas contrast varies by at most a factor of 20 [20].

Second, the imager bandpass filters the spatiotemporal signal to attenuate low-frequency spatial and temporal signals, and to reject wideband noise. The increase in gain with frequency, for frequencies below the peak, matches the $1/f^2$ decrease in power with frequency for natural image spectra, resulting in a flat output power spectrum. This filtering improves information coding efficiency by reducing correlations between neighboring samples in space and time. It also reduces the variance of the output, and makes the distribution of activity sparse.

Third, the imager adapts its sampling rate locally to minimize redundant sampling of low-frequency temporal signals. In the face of limited communication resources and energy, this sampling-rate adaptation has the additional benefit of freeing up capacity, which is dynamically reallocated to active pixels, allowing higher peak sampling rates and shorter latencies to be achieved [39].

And fourth, the imager adapts its step size locally to trade resolution at high contrast levels, which rarely occur, for resolution at low contrast levels, which are much more common. The proportional step size in the adaptive neuron, which results in a logarithmic transfer function, matches an exponentially decaying amplitude probability density, making all quantization intervals equiprobable. Hence, it maximizes the expected number of signals that can be discriminated, given their probability of occurrence.

For independent samples, information is linearly proportional to bandwidth, and is logarithmically proportional to the signal-to-noise ratio

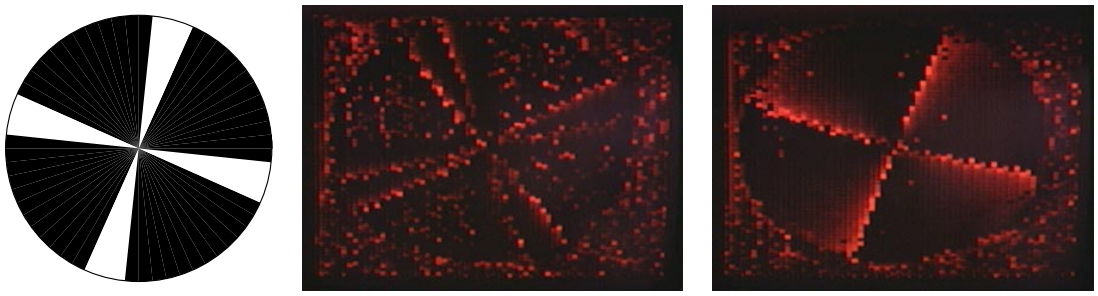


Fig. 11. Video frames from postprocessor chip showing real-time temporal integration of pulses. The stimulus is a windmill pattern (left) that rotates counterclockwise slowly (middle) and quickly (right).

(SNR) [22]. We increase bandwidth by making the receptors smaller and faster, so that they can sample more frequently in space and time. As an unavoidable consequence, they integrate over a smaller volume of space–time and therefore SNR degrades. There is therefore a reciprocal relationship between bandwidth and noise power (variance) [40]. Since their goal is to maximize information, biological sensory systems aggressively tradeoff SNR for bandwidth, operating at SNRs close to unity [12], [40].

With this optimization principle in mind, I have proposed compact circuit designs that realize local AGC, bandpass filtering, and adaptive quantization at the pixel level. The overriding design constraints are to whiten the signal, which makes samples independent; to minimize the pixel size and capacitance, which makes sampling more dense and more rapid; and to minimize power consumption, which makes it possible to acheive very-large scale integration. Hence, all circuits use minimal-area devices and operate in subthreshold, where the transconductance per unit current is maximum. I realized extremely compact implementations by modeling these circuits closely after their biological counterparts [5], [38].

I analyzed three limitations in these simple circuit designs.

First, attenuating low frequencies by using a high-gain receptor-to-HC synapse (ratio of $g_{hc}/g_{h0} \equiv 1/\epsilon_h$) results in temporal instability. To break this tradeoff, we must regulate the gain dynamically.

Second, controlling the gain by changing the receptor-to-receptor coupling strength compromises the receptive field size. To decouple these parameters, we must change one of the synaptic strengths (transconductances, g_{ch} or g_{hc}) proportionally.

And third, attenuating the firing rate by using an integrator with a long time constant results in extremely slow adaptation, because we must use a small charge quantum to avoid sending the integrator's output above the input level. The circuit would adapt more rapidly, and fire fewer spikes in the process, if it maintains a uniformly high firing rate until the integrator catches up with the input, and then switches abruptly to a low firing rate.

6. Conclusions

Taking inspiration from biology, I have described an approach to building machine vision systems that perform sophisticated signal processing at the pixel level. These retinomorphic systems are adaptive to their inputs, and thereby maximize their information-gathering capacity and minimize redundant information in their output data stream.

These optimization principles are radically different from those that drive the design of conventional video cameras. Video cameras are designed to reproduce any arbitrary image to within a certain worst-case error tolerance, whereas biologics exploit the statistical properties of natural spatiotemporal signals, giving up worst-case performance to get better average-case performance.

Optimizing average-case performance maximizes the discrimination ability of biologics. Consequently, biomorphic systems promise superior solutions for human-made systems that perform perceptive tasks, such as face recognition and object tracking, energy efficiently.

Notes

1. I am assuming a linear encoding—a practice that is the standard. This assumption limits the dynamic range to 2^b for a b-bit encoding.
2. I performed bandpass filtering by convolving the input image with the laplacian of a Gaussian with $\sigma = 2.5$ pixels. I calculated the autocorrelation of the images by subtracting out the mean, shifting a copy of the image up or right by 1 to 75 pixels, multiplying corresponding pixels, and summing; I normalized the results to yield a maximum of unity. Rightward shifts are plotted on the positive axis (0 to 75), and upward shifts are plotted on the negative axis (0 to -75).
3. CCD Camera Specifications: COHU Solid State RS170 Camera (142320), auto iris, gamma factor enabled, 512×480 pixels, 8-bit gray-level outputs. Lens Specifications: COSMICAR TV lens, ES 50mm, 1:1.8. This comparison, and the face recognition studies, were done in collaboration with Frank Eeckman, Joachim Buhman, and Martin Lades of the Lawrence Livermore National Labs, Livermore CA.

References

1. B Hoeneisen and C Mead. Fundamental limitations in microelectronics-i: Mos technology. *IEEE J. Solid-State Circ.*, 15:819–829, 1972.
2. C Mead. Scaling of mos technology to submicrometer feature sizes. *J. VLSI Signal Processing*, 8:9–25, 1994.
3. K Boahen. Retinomorphic vision systems. In *Microneuro'96: Fifth Int. Conf. Neural Networks and Fuzzy Systems, Laussanne Switzerland*, Los Alamitos, CA, Feb 1996. EPFL/CSEM/IEEE, IEEE Comp. Soc. Press.
4. M Mahowald and C Mead. The silicon retina. *Scientific American*, 264(5):76–82, 1991.
5. K Boahen and A Andreou. A contrast-sensitive retina with reciprocal synapses. In J E Moody, editor, *Advances in Neural Information Processing*, volume 4, San Mateo CA, 1991. Morgan Kaufman.
6. K Boahen. Communication neuronal ensembles between neuromorphic chips. *In preparation.*
7. K Boahen. A retinomorphic vision system. *IEEE Micro Magazine*, Oct 1996.
8. K Fujikawa, I Hirota, H Mori, T Matsuda, M Sato, Y Takamura, S Kitayama, and J Suzuki. A 1/3 inch 630k-pixel it-ccd image sensor with multi-function capability. In John H. Wuorinen, editor, *Digest of Technical Papers*, volume 38 of *IEEE International Solid-State Circuits Conference*, pages 218–219, San Francisco, CA, 1995.
9. A Dickinson, B Ackland, E El-Sayed, D Inglis, and E R Fossum. Standard cmos active pixel image sensors for multimedia applications. In William Dally, editor, *Proceedings of the 16th Conference on Advanced Research in VLSI*, pages 214–224, Chapel Hill, North Carolina, 1995. IEEE Press, Los Alamitos CA.
10. C Mead. A sensitive electronic photoreceptor. In H. Fuchs, editor, *1985 Chapel Hill Conference on VLSI*, pages 463–471, Rockville MD, 1985. Computer Science Press, Inc.
11. J Buhman, M Lades, and Eeckman F. Illumination-invariant face recognition with a contrast sensitive silicon retina. In J D Cowan, G Tesauro, and J Alspector, editors, *Advances in Neural Information Processing*, volume 6, San Mateo CA, 1994. Morgan Kaufman.
12. R R de Ruyter van Steveninck and S B Laughlin. The rate of information transfer at graded-potential synapses. *Nature*, 379:642–645, Feb 1996.
13. T Delbruck and C Mead. Photoreceptor circuit with wide dynamic range. In *Proceedings of the International Circuits and Systems Meeting*, IEEE Circuits and Systems Society, London, England, 1994.
14. D J Field. Relations between statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am.*, 4:2379–2394, 1987.
15. D Dong and J Atick. Statistics of natural time-varying scenes. *Network: Computation in Neural Systems*, 6(3):345–358, 1995.
16. J Atick and N Redlich. What does the retina know about natural scenes. *Neural Computation*, 4(2):196–210, 1992.
17. D Dong and J Atick. Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, 6(2):159–178, 1995.
18. J H van Hateren. A theory of maximizing sensory information. *Biol. Cybern.*, 68:23–29, 1992.
19. B Sakman and 0 D Creutzfeldt. Scotopic and mesopic light adaptation in the cat's retina. *Pflúgers Archiv fúr die gesamte physiologie*, 313:168–185, 1969.
20. W A Richards. A lightness scale for image intensity. *Appl. Opt.*, 21:2569–2582, 1982.
21. M 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.
22. C E Shannon and W Weaver. *The Mathematical Theory of Communication.* Univ. Illinois Press, Urbana IL, 1949.
23. C Jansson, I Per, C Svensson, and R Forchheimer. An addressable 256×256 photodiode image sensor array with 8-bit digital output. *Analog Integr. Circ. & Sig. Proc.*, 4:37–49, 1993.
24. B Fowler, A E Gamal, and D Yang. A cmos area image sensor with pixel-level a/d conversion. In John H. Wuorinen, editor, *Digest of Technical Papers*, volume 37 of *IEEE International Solid-State Circuits Conference*, pages 226–227, San Francisco, California, 1994.
25. A Andreou and K Boahen. Translinear circuits in subthreshold mos. *J. Analog Integrated Circ. Sig. Proc.*, 9:141–166, 1996.
26. E Vittoz and X Arreguit. Linear networks based on transistors. *Electronics Letters*, 29:297–299, 1993.
27. K Bult and G J Geelen. An inherently linear and compact most-only current division technique. *IEEE J. Solid-State Circ.*, 27(12):1730–1735, 1992.

28. P C Chen and A W Freeman. A model for spatiotemporal frequency responses in the x cell pathway of cat's retina. *Vision Res.*, 29:271–291, 1989.
29. S Ohshima, T Yagi, and Y Funashi. Computational studies on the interaction between red cone and h1 horizontal cell. *Vision Res.*, 35(1):149–160, 1994.
30. K Boahen. Spatiotemporal sensitivity of the retina: A physical model. Technical Report CNS-TR-91-06, California Institute of Technology, Pasadena CA, 1991.
31. R G Smith. Simulation of an anatomically defined local circuit: The cone-horizontal cell network in cat retina. *Visual Neurosci.*, 12(3):545–561, May-Jun 1995.
32. C Enroth-Cugell, J G Robson, D E Schweitzer-Tong, and A B Watson. Spatiotemporal interactions in cat retinal ganglion cells showing linear spatial summation. *J. Physiol.*, 341:279–307, 1983.
33. D H Kelly. Motion and vision ii: Stabilized spatiotemporal threshold surface. *J. Opt. Soc. Am.*, 69(10):1340–1349, 1979.
34. K Boahen. Toward a second generation silicon retina. Technical Report CNS-TR-90-06, California Institute of Technology, Pasadena CA, 1990.
35. C Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, Reading, MA, 1989.
36. M Mahowald and D Douglas. A silicon neuron. *Nature*, 354(6345):515–518, 1991.
37. J Lazzaro. Temporal adaptation in a silicon auditory nerve. In D Tourestzky, editor, *Advances in Neural Information Processing*, volume 4, San Mateo CA, 1992. Morgan Kaufmann.
38. K Boahen. The adaptive neuron and the diode-capacitor integrator. *In preparation.*
39. K Boahen. Retinomorphic vision systems ii: Communication channel design. In *IEEE Int. Symp. Circ. & Sys.*, volume Supplement, Piscataway, NJ, May 1996. IEEE Circ. & Sys. Soc., IEEE Press.
40. W R Softky. Fine analog coding minimizes information transmission. *Neural Networks*, 9(1):15–24, 1996.

Kwabena A. Boahen is a doctoral student at the California Institute of Technology, Pasadena, CA, in the Computation and Neural Systems program. He earned a BS/MSE degree in Electrical Engineering from the Johns Hopkins University, Baltimore, MD. His current research interests include mixed-mode multichip VLSI models of biological sensory systems, and asynchronous digital interfaces for interchip connectivity.