KIM STROHBEHN, RICHARD C. MEITZLER, ANDREAS G. ANDREOU, and ROBERT E. JENKINS

ANALOG IMAGE PROCESSING WITH SILICON RETINAS

Early vision algorithms map naturally on distributed physical systems configured to process signals in continuous time and to act on analog image values. Using analog very large scale integrated technology, we have constructed prototypes of several experimental systems that implement biologically inspired image processing functions, including edge enhancement, centroid computation, and image translation (motion computation). Recent experiments at the Applied Physics Laboratory indicate that these systems are already practical for use in feedback loops in selected control system applications.

INTRODUCTION

The visual systems of animals routinely accomplish such sophisticated image processing tasks as image acquisition, edge enhancement and detection, segmentation, motion estimation, and object recognition in real time and with minimum power dissipation. Further, these biological operations are superior to corresponding human-engineered functions in performance, robustness, and energetic efficiency. From careful studies of the visual systems of the cat, fly, frog, and turtle,[1] we now know enough about some of these neuromorphic early image processing strategies and architectures to use them as models for human-engineered systems. Early image processing in biological systems does not precisely restore information; rather, it transforms and decomposes the image into suitable representations (maps) that are useful for allowing the animal to interact with its environment (in a closed-loop configuration, where the environment is an integral part of a loop).

Most human-engineered image processing systems consist of an array detector, such as a charge-coupled device (CCD) camera, which senses or detects an image, and a digital processor, which reads and digitizes the image. The resulting array of numbers is then operated on by an image processing algorithm, implemented on either a special or general-purpose digital computer, to extract features of interest. Such systems produce sampled-data, discrete-time outputs and are programmable.

In contrast, biological systems perform image processing in continuous time on analog or quasi-analog signal values. Also, much of this computation occurs in a distributed fashion near the site of image detection, so that the animal can "read out" information based on reduced-bit-rate representations and thus avoid communication bottlenecks. We can demonstrate this strategy by examining edge enhancement in biological retinas, where computation is circumscribed by the inherent physics of the underlying substrate as well as limited energy resources and physical space.

When an animal views a scene, the retinal cells stimulate neurons to reduce the potential of their membranes and permit ions to diffuse through the intracellular fluid. The second- and higher-order spatial derivatives of the cell response to the diffusion at a particular location are proportional to the net flux of particles (current injection) introduced. If there is no net flux to this region, the distribution of ions is linearly graded. Biological retinas exploit this property to compute deviations from the average spatial intensity (i.e., to identify edges). They recognize contrast by sensing the local diffusivities of the medium, which are functions of the injected ion current. This "local gain control" mechanism makes the system nonlinear.

Artificial silicon retinas can be fabricated to mimic this behavior because diffusion is the dominant charge transport mechanism in the subthreshold metal-oxide-semiconductor (MOS) transistor. In subthreshold mode, the circuits are operated with the transistor gate voltages well below the point at which the mobile charge begins to limit the flow of current. In this operating regime, the transistors put out very small but highly controllable (via the terminal voltages) diffusion currents in the nanoampere range. Subthreshold operation enables ultra-low-power analog processing (and, therefore, ultra-low heat dissipation) with the high integration density of silicon transistors.

Our experiments at APL have indicated several applications for analog image processing systems that can acquire an image, perform local gain control and edge enhancement, and compute image translation. Our chips now put out a small number of signals—for example, x and y positions and velocity rather than some field quantity—because we have tried to implement as much functionality as possible on a single die.

PRINCIPLES OF ANALOG IMAGE PROCESSING

One of the central goals for this work is to use biological systems as models, an approach pioneered by Carver Mead's group at Caltech.[2,3] The efficiency and performance of many visual systems found in nature are still far beyond those possible with any current technology; therefore, we can learn—and borrow—much from natural systems.

Biological systems excel in terms of power consumption, size, and overall performance rather than the raw speed of the individual components. These attributes are also characteristic of neurons, the nerve cells that are the

basis for natural computation. Natural systems perform so well because their processing architecture is tailored to the computational units (neurons) as well as to the task at hand. Neuron-based systems are fully analog processors that process data continuously and in real time.

The fundamental computational unit of our silicon system (analogous to the neuron) is the MOS field-effect transistor (MOSFET). Figure 1 shows a diagram of two complementary MOSFETs, an n-type (charge carriers are electrons) and a p-type (charge carriers are holes). Their functions are identical except that increasing the voltage on the gate terminal increases the current through an n-type device and decreases the current through a p-type device. In a complementary MOS (CMOS) device, both types of MOSFETs are available on a single chip.

A MOSFET's operation is conceptually simple.[4–6] The gate terminal is insulated and thus conducts no current itself. Its function is to control the current flowing between the drain and source terminals. Also, the MOSFET is symmetric so that reversing the potentials on the drain and source switches the direction of the current but does not otherwise affect the device's behavior.

Although some of today's fastest microprocessors consist of MOSFETs, our systems use these devices quite differently from the way they are used in traditional digital computers. For example, most of the transistors in our image processors are operated in the subthreshold region, which is characterized by extremely low currents (100 nA or less). Such small currents mean that the power consumption and heat dissipation of our chips are orders of magnitude below those of a digital integrated circuit of similar computational capability. Thus, solar cells or batteries can easily power the system, and heat removal is unnecessary. As a result, our processors are much more compact than traditional digital systems, which enhances their portability and facilitates their inclusion in larger systems.

As another important benefit of subthreshold operation, the MOSFET has the highest possible gain in this region, that is, the current is an exponential function of the gate voltage. For an n-type transistor, the current from the drain to the source, I_{ds}, is given by

$$I_{ds} = \frac{W}{L} I_0 e^{\kappa V_{gate}/U_{th}} \left(e^{-V_{source}/U_{th}} - e^{-V_{drain}/U_{th}} \right), \quad (1)$$

where W and L are the transistor width and length, respectively, and are available as design parameters. The I_0 term (in amperes) takes into account carrier mobility. U_{th}, the thermal voltage, is approximately 0.026 V at room temperature and is calculated as kT/q, where k is Boltzmann's constant, T is absolute temperature, and q is the charge of an electron. Note that U_{th} is not the threshold voltage. V_{gate}, V_{source}, and V_{drain} are the voltages of the corresponding terminals shown in Fig. 1 relative to the substrate terminal. The dimensionless κ term accounts for capacitive coupling of the MOSFET gate to the substrate, which reduces the gate's effectiveness in controlling the surface potential under it; κ has typical values of 0.6 to 0.7 for our devices. The drain current equation for a p-type MOSFET is identical to Eq. 1 except that all voltages are reversed in sign.

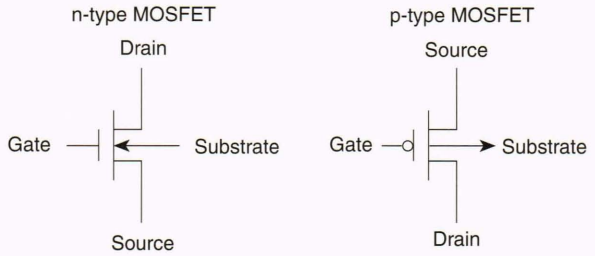


Figure 1. Symbols for n-type and p-type MOSFETs. Each transistor consists of a semiconductor substrate and source and drain regions that have been implanted with impurities that either donate electrons (n-type) or accept electrons (p-type). A metal electrode or gate spans the region from the source to the drain. A voltage applied to the gate lowers the resistance to charge flow, and charge diffuses through the channel under the gate from the high charge density at the source to the low charge density at the drain. The current developed is an exponential function of the applied voltage for subthreshold operation.

The subthreshold current is dominated by the diffusion of charge carriers from a high-concentration region (the source) to a low-concentration region (the drain), which is exponentially related to voltage.[5,6] It is this exponential behavior that enables our designs to perform important functions with only a fraction of the power and chip surface area required by digital circuits. Our use of innate device characteristics is an important similarity between our processors and natural ones.

Neural computation is based in part on the diffusion of ions across permeable membranes and Boltzmann statistics, which describes the exponential decrease in the density of particles in thermal equilibrium with a potential gradient. However, the properties of ion channels and digital integrated circuits are fundamentally different. In particular, ion channels are physical systems that do not have to obey solid-state Fermi statistics, that is, the ions do not have to occupy a certain energy state at thermal equilibrium, as do electrons in semiconductors or metals. Thus, even though the characteristics of gated channel conductance in transistors depend exponentially on charge diffusion, the exponential dependence of the natural system is steeper (the equivalent of a κ greater than 1) because of a collective phenomenon called correlated charge transport. The importance of this fundamental difference and its effect on the behavior of biological computational systems are not presently clear. However, from our experience with the synthesis of complex silicon systems, correlated transport in the active elements could yield reliable operation in noisy environments as well as good noise margins at very low power (100 mV). Operating at reduced power also reduces the energy required per computation.

Another important similarity between our analog image processors and nature's arrangement is that both use local processors to perform many computations in parallel. This scheme allows a much faster overall rate of computation even though the individual processors do not operate at high speed. Most digital imaging systems suffer from computational bottlenecks as many sensors attempt to access a central processor at the same time.

With no controls needed on processor access, parallel computation also simplifies issues related to the timing and coordination of individual circuits.

At times, some very "unnatural" features appear in our designs as an unavoidable consequence of differences between neuron-based and CMOS systems. We cannot stack our devices in multiple layers like biological systems, which means that the devices cannot be interconnected as fully as in nature. Also, the differences between MOSFETs and neurons can significantly affect the implementation of biological algorithms in silicon. In addition, because our designs are more specialized than biological systems, we can often obtain a simpler solution by deviating from purely natural types of processing. Nevertheless, neuromorphic-like computation gives our systems unique advantages in terms of size and power for their processing abilities.

A SECOND-GENERATION SILICON RETINA

Several years ago, Boahen and Andreou[7] implemented a model of early image processing in the vertebrate retina in CMOS circuitry, hereafter referred to as the silicon retina. We have extended this work to show that subthreshold analog circuitry can mimic the functionality of the first two retinal cell layers (the outer plexiform). To demonstrate this concept, we constructed a circuit with a minimum of transistors that continuously computes a powerful image processing transformation.

In vertebrates, the retina performs vital image processing in addition to collecting light. Most important among these functions, as we mentioned before, are local gain control and enhancement of image edges. Thus, the outer plexiform output is primarily contrast-sensitive rather than intensity-sensitive. To understand how this edge enhancement works, consider the response of the retina to a small point of light that stimulates a single photoreceptor cell (for example, a cone cell). The stimulated cell is excited above its baseline activity level and inhibits its neighboring cells, i.e., it attempts to decrease their activity. In other words, the retinal cells engage in a local competition in which active cells attempt to turn off their neighbors. The receptive field of the photoreceptor cell is described as on-center, off-surround because the stimulated cell is turned on (is at the center), and a stimulus in the surrounding area of the cell tends to turn it off. In effect, the on-center, off-surround receptive fields act as a high-pass spatial filter to enhance edges.

This type of processing is frequently represented by a linear systems model that convolves the image with the well-known difference-of-Gaussians (DOG) filter. The resulting output is the difference between two smoothed images, one smoothed with a Gaussian kernel (weighting function) with a wide response that computes a local average, and the other smoothed with a narrower kernel that estimates local activity. The difference produces a high-pass spatial response, with the additional benefit of local smoothing to provide noise immunity. In biological retinas, resistive, leaky connections between cells produce similar local smoothing. The retina also compresses the range of intensity of the incident light (a nonlinear operation), limiting the dynamic range of signals presented for further operations and thus facilitating processing by subsequent stages. Beyond the outer plexiform layer, the bipolar and ganglion cells provide even more processing prior to transmission of information along the optic nerve. These later processing stages have not yet been implemented in our circuitry, but are discussed in detail (as is the entire retina) by Dowling.[1]

Figures 2a and b show, respectively, a schematic model of the retina's outer plexiform and the corresponding circuit for a single silicon retina pixel. The cones, the retina's color receptors, are interconnected by gap junctions so that each cone's activity also spreads to its neighbors. The gap junctions are effectively leaks in the cell membrane that provide some resistance to ion flow, and thereby constitute a resistive network. The horizontal cells below the cones are also excited by cone activity and spread their activity through another resistive network of gap junctions. However, connections from the horizontal cells back to the cones are inhibitory, producing negative feedback. This latter mechanism produces the on-center, off-surround receptive field just discussed. Since the conductance of the gap-junction network determines its spatial response, arbitrarily large kernels can be implemented with only nearest-neighbor connections (i.e., cells in a neighborhood do not have to be directly connected to respond to a stimulus). These two cell layers and the inhibitory/excitatory connections between them constitute a simple yet powerful structure for computing a DOG-like convolution.

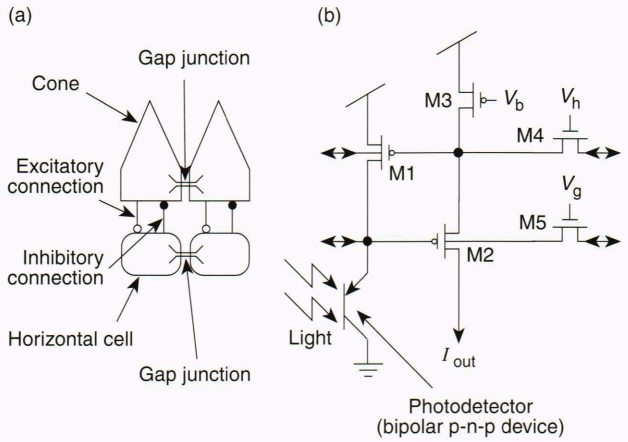


Figure 2. Turtle retina and silicon retina structures. (a) In the outer plexiform of the biological retina, activity in the cones spreads to neighboring cones through a gap-junction resistive network and also to a horizontal layer of cells below with their own resistive network. Connections from the horizontal cells back to the cones provide negative feedback. Gap-junction networks perform local spatial averaging of the illumination level within the visual scene. The resulting receptive field is called on-center, off-surround. (Reprinted from Andreou et al.[8] by permission. ©1991 IEEE.) (b) Simplified, one-dimensional unit pixel of a silicon implementation of the turtle retina consists of a photodetector and five MOSFETs. A bias voltage V_b is input to the M3 transistor, which produces a base-level current in the system. M4 and M5, respectively, are smoothing networks for M1 and M2 and are biased by voltages V_h and V_g. Light striking the photodetector produces an excitatory current that reduces the gate voltages on M2 and M1; M1, in turn, produces negative feedback by increasing the voltage on the gate of M2. The circuit output is the drain current of M2, which represents the processed input image.

Our retina circuit attempts to mimic this economy and functionality with only seven MOSFETs and a photodetector. These elements constitute a pixel. The photodetector is a vertical bipolar transistor with a floating base that consumes approximately 22% of the area within a pixel. The two gap-junction networks, formed by the transistors M4 and M5, respectively, are independently biased from external voltage sources for easy control of the diffusion of charge. With this design, we can adjust the sizes of both the on-center and off-surround receptive fields and control the amount of localization or smoothing performed by the retina.

To more easily explain the operation of this circuit, we have depicted it in Fig. 2b as a simplified unit cell with five MOSFETs, which provide one-dimensional connectivity. In actual practice, transistors M4 and M5 are replicated to provide two-dimensional connectivity, but the operation does not differ conceptually from the one-dimensional case.

The p-type transistors M1 and M2 form the negative feedback loop analogous to the inhibitory connections between the horizontal cell and the cone or photoreceptor in the natural retina. Incident photons detected at the reverse-biased base-collector junction of the photoreceptor (bipolar transistor) produce an emitter current that discharges the gate of the source follower M2. The discharge is smoothed by the M5 diffusion network. The resulting decrease in voltage can be interpreted as an excitatory response. The gate voltage of the inverting amplifier M1, which is smoothed by the M4 diffusion network, decreases along with the gate voltage of M2. The decrease in M1's gate voltage increases the drain current of M1, which then charges up the M2 gate and provides the negative feedback. Just as retinal cells compete, M2 competes with the M4 network for the bias current provided by current-source transistor M3 to produce the output drain current. If an adjacent output is providing a large amount of current, M4 takes available output current from M2, thereby producing the on-center, off-surround response. In addition, the outputs are all normalized to the level set by the bias voltage V_b. As a result, the output level is relatively independent of the overall illumination level, which is another highly desirable property for our silicon retina.

The silicon retina carries out "current-mode" computations[8] in which current, rather than voltage, is the variable of interest. This design philosophy is directly drawn from biological systems, which use ion flow (i.e., currents) for their computations. Also like the natural retina, the silicon retina is sensitive to image contrast rather than absolute intensity. By providing for local gain control, the circuit allows resolution of detail across images that contain large nonuniformities in average intensity. The amount of gain localization is also controlled by the transistor bias levels.

The pixel shown in Fig. 2b can be connected in a hexagonal, two-dimensional array to maximize its number of nearest neighbors. Using the second-difference approximation to the Laplacian operator ∇^2, we can obtain an approximate continuous equation for the retina:

$$\lambda \nabla^2 \nabla^2 I_{M1}(x,y) + I_{M1}(x,y) = I_{photo}(x,y)\,, \qquad (2)$$

where λ is determined by the product of the spatial constants of the upper and lower diffusive networks, and I_{M1} is the drain current through M1. The spatial constants are set by V_h and V_g, the biases on the gates of M4 and M5, respectively. Equation 2 can be recognized as the biharmonic equation.[8] The solution I_{M1} of this equation is an estimate of the image for the input I_{photo}. Since this estimation problem is ill-posed in general, a biharmonic term (double Laplacian) is commonly added as a smoothness constraint. By adjusting the λ term, we can select the relative extent of smoothing versus sensitivity to edges. The output current I_{out} is a normalized, high-pass-filtered version of the image estimate solution:

$$I_{out} = I_{M3} - \nabla^2 I_{M1}\,. \qquad (3)$$

Several versions of the retina have been fabricated in CMOS processes with a 2-μm feature size. The same design was scaled down and fabricated in a 1.2-μm process on a 1-cm^2 die; this chip holds 590,000 transistors in a 48,000-pixel array and is fully functional. To our knowledge, this fully analog processor is the largest ever fabricated and certainly performs more computations per silicon area and per unit of energy than any other silicon system, commercial or experimental. If we serially scan out the values at each pixel[9] and display the result on a standard television monitor, we can easily see the edge-enhancing properties of this design.

Figure 3b shows the response of a 50 $\times$ 50 pixel silicon retina to a photograph of Saturn and its moons (Fig. 3a). This figure serves to illustrate several properties of silicon retinas. First, the retinal response is very "flat" compared with a conventional image because background illumination variations are replaced by a constant background level. In fact, the input image was processed under a severe illumination gradient (a desk lamp was cocked at an angle to the photograph), but the output of the retina is insensitive to illumination level as long as it receives enough photons to detect sufficient contrast. To test this invariance, we varied the room lighting while observing the retinal output on the monitor. As expected, the output remained constant over many orders of illumination magnitude.

The primary edges detected in this image are along the edges of Saturn, the foreground moon, and the rings. Note the response to an edge: the retina pixels are brighter than average along the bright side of the edge and darker than average along the dark side of the edge, and the interiors of both light and dark objects are replaced by the background level. These properties are just what we would expect. In present versions of the retina, the gain of the input phototransistors is rather nonuniform, which causes a fixed-pattern input noise that competes with actual image edges and limits the edge sensitivity. We are therefore investigating several ways of reducing fixed-pattern noise. In this case, we chose a smoothing level that reduced the fixed-pattern edges somewhat but still permitted detection of the interesting edges in the input image consistent with the 50 $\times$ 50 pixel spatial resolution limits.

Even though the static retina output is not yet particularly impressive compared with that of a video camera,

the present retina is mature enough for practical application in certain position control systems.

VIDEO-CENTROID COMPUTATIONS

The centroid of an image is a very useful quantity for image-tracking applications. Since finding the centroid is an averaging process, the solution is robust to noise as well as insensitive to minor variations in the apparent image due to changes in illumination level (but not gradient). In addition, centroid computation is consistent with our retina paradigm, i.e., it uses many simple, local computing elements in parallel to calculate a global quantity. Thus, this particular problem is a good candidate for an analog very large scale integrated (VLSI) implementation.

The earliest work on centroid-computing chips of this type was reported by Deweerth.[10] The core element for a simple centroid computation is simply an array of transconductance amplifiers, each of which is biased by a photodetector (Fig. 4a). (A transconductance amplifier generates an output current that is a function of the difference of two input voltages.) The gate of one MOSFET of the input differential pair is connected to the output line (as a follower), while the gate of the other is connected to a resistive divider. Applying known voltages V^+ and V^- on either end of the divider produces an output that varies linearly along its length and thus can be used to encode position. The left MOSFET can be considered as carrying the output current of the pair, while the right one carries the position current.

To perform the actual centroid computation, we configure the transconductance amplifiers as followers with a common position output node (Fig. 4b). The network attempts to satisfy the following equation:

$$0 = \sum_i I_{\text{photo}}(i) \tanh\left[\frac{V(i) - V_{\text{out}}}{2U_{\text{th}}}\right], \tag{4}$$

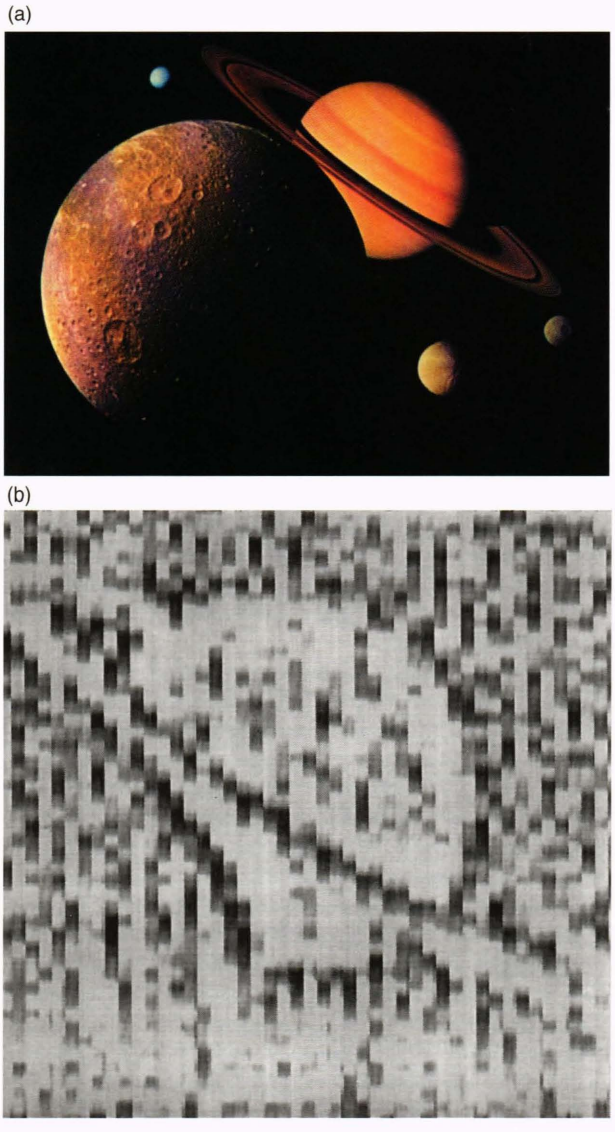


Figure 3. (a) Photographic montage of Saturn and its moons (courtesy of the Planetary Society). (b) Digitized video output of a 50 × 50 pixel retina chip for the Saturn image in (a).

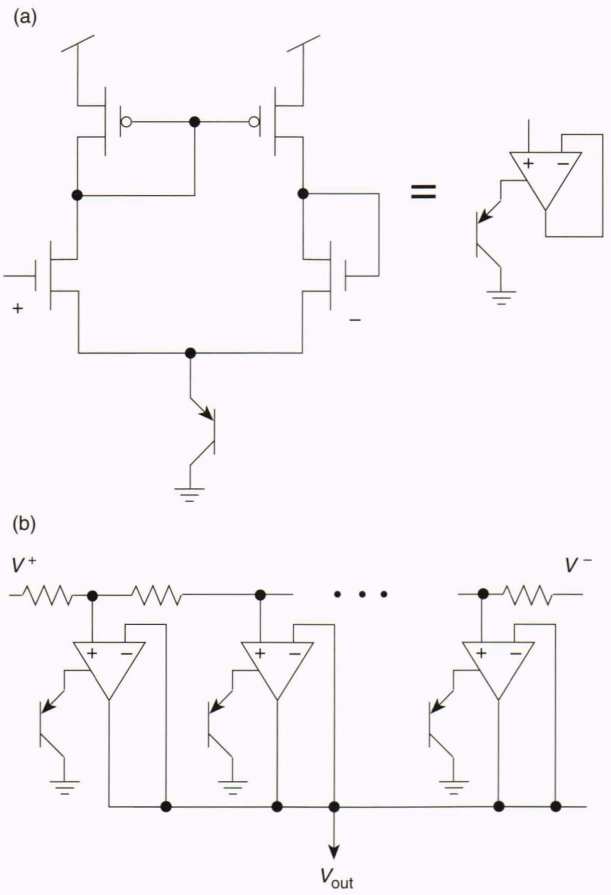


Figure 4. (a) Schematic of a simple centroid-computing circuit, consisting of a transconductance amplifier biased with a phototransistor. Application of known voltages on either end of the resistive divider produces a linear voltage that can encode position. (b) Schematic of the centroid computation circuitry used in the video-centroid chip. The output is the image centroid for small signals or the weighted median for large signals, both of which are excellent position estimates for images with a well-defined centroid.

where $V(i)$ is the position voltage at the noninverting (+) input of the ith amplifier, $I_{photo}(i)$ is the corresponding ith photocurrent, and U_{th} is the thermal voltage divided by κ. The slope of the tanh function is the transconductance of the amplifier. For small signals, tanh can be replaced by the transconductance, which is linear with the bias current (i.e., photocurrent) in the subthreshold region. If we apply Kirchoff's current law (conservation of charge) at the output node, we can readily see that the solution to Eq. 4 is the centroid of the photocurrent distribution. The network stabilizes at a point where the output V_{out} is equal to the solution of Eq. 4, which is the image centroid for small signals or the weighted median for large signals. Both the centroid and the median yield excellent position estimates for images with a well-defined centroid as shown in Fig. 5, which is a plot of the y-channel position output versus a laser spot position on one of our experimental chips (the x-channel results are similar).

To create a centroid-computing chip, we modified the architecture of Fig. 4 to establish a two-dimensional phototransistor array with current summing lines along each row and column.[11] These lines were then used as the inputs to the centroid computation circuitry. The summing operation allows each pixel to contribute to both the x and y centroids, in contrast to Deweerth's design in which alternating x and y pixels were used. In addition, summation is an averaging process, which helps to eliminate errors in the centroid caused by unusually conductive MOSFETs. Mismatch in transistor currents for identical drain, gate, and source voltages is unavoidable and must be accounted for in the design.[12] We solved the mismatch problem by aggregation (i.e., averaging) in the same way as nature, which has to deal with nonuniformities in neuron properties that are far greater than those in MOSFETs. Aggregating the response of many parallel elements is a common technique in biological computation to improve the robustness of the output signal.

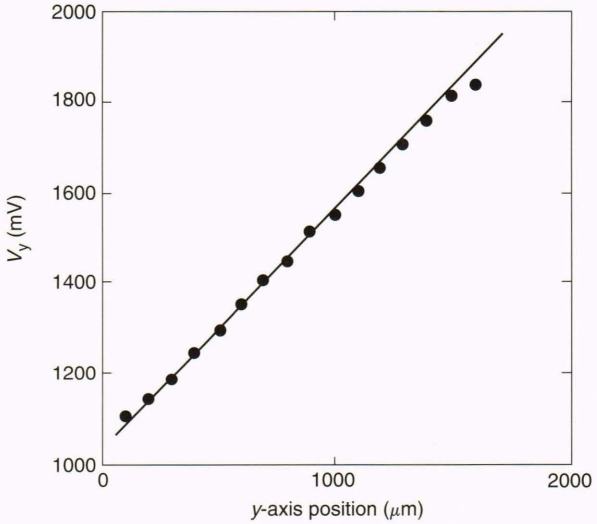


Figure 5. The y-channel position sensitivity (output voltage) of an experimental video-centroid chip is a linear function of the position of a laser spot, indicating that the network computations provide excellent position estimates for images with a well-defined centroid. Data shown are for a 489-pW laser. The x-channel results are similar.

In addition to summing lines, our centroid chips add scanning circuitry to read out the photocurrent at each pixel. Because the chip output is both a video image and the centroid of that image, we designated our design the video-centroid (VC) chip. The video circuitry allows us to view the same image as the chip and thus to ensure that the system is tracking the intended target. Note that incorporation of a small amount of circuit complexity helps reduce complexity in other parts of the system, for example, by eliminating the need for additional beamsplitters and methods of aligning the viewing and tracking optical components.

The first-generation VC chip is currently being used to generate image-position error signals for an image-motion stabilization system at the National Solar Observatory in Sunspot, New Mexico. It has successfully replaced a half-height rack of digital processing equipment formerly used for tracking. Although power consumption is not an issue here, the VC chip's power dissipation is many orders of magnitude below that of its digital predecessor. Finally, the chip's frequency response is very fast compared with that of the photodiode sensor used in the digital system because the sensor array does not have to be read out prior to position computation. Since the VC chip's scanning circuitry is independent of the centroid computation, it can achieve bandwidths up to 40 kHz. For closed-loop systems, the mechanical positioning elements are the only limitations. Figure 5 shows a plot of the VC chip position sensitivity for static inputs. The bandwidth is dependent on illumination, as might be expected from Fig. 4.

Our next-generation VC chip incorporated several improvements, most notably, replacement of the single phototransistor with the silicon retina described previously.[13] We named this design the retina-VC chip, or RVC chip. Because the retina is contrast-sensitive, the VC chip architecture could accommodate lower contrast images as well as scenes with illumination gradients. Scanner circuitry was also improved to enhance the speed and quality of the video readout. With these modifications, we were able to increase the size of the pixel array while still maintaining a standard video output. The RVC chip included an on-chip video driver, thereby reducing off-chip components to a single field-programmable gate array chip containing the video timing circuitry, a transistor, and fewer than 10 discrete resistors and potentiometers for biasing the chip. The entire centroid computation system fit on a printed circuit board approximately 4 in^2, which is considerably more compact than an equivalent digital system. Figure 6 shows a block diagram of the RVC chip system.

We tested the RVC chip's sensitivity by using it to view star fields through an 11-in. Celestron telescope set at *f*/10. Stars of magnitude +3 were the dimmest objects which could be viewed reliably. (Fainter stars were obscured by nonuniformities in the retina resulting from the transistor mismatch discussed earlier.) Since the RVC chip does not use an integrating detector, which builds up an image over time, it is much less sensitive than a CCD detector and is appropriate only for bandwidths greater than 1 kHz.[13] However, converting the photodetector to an

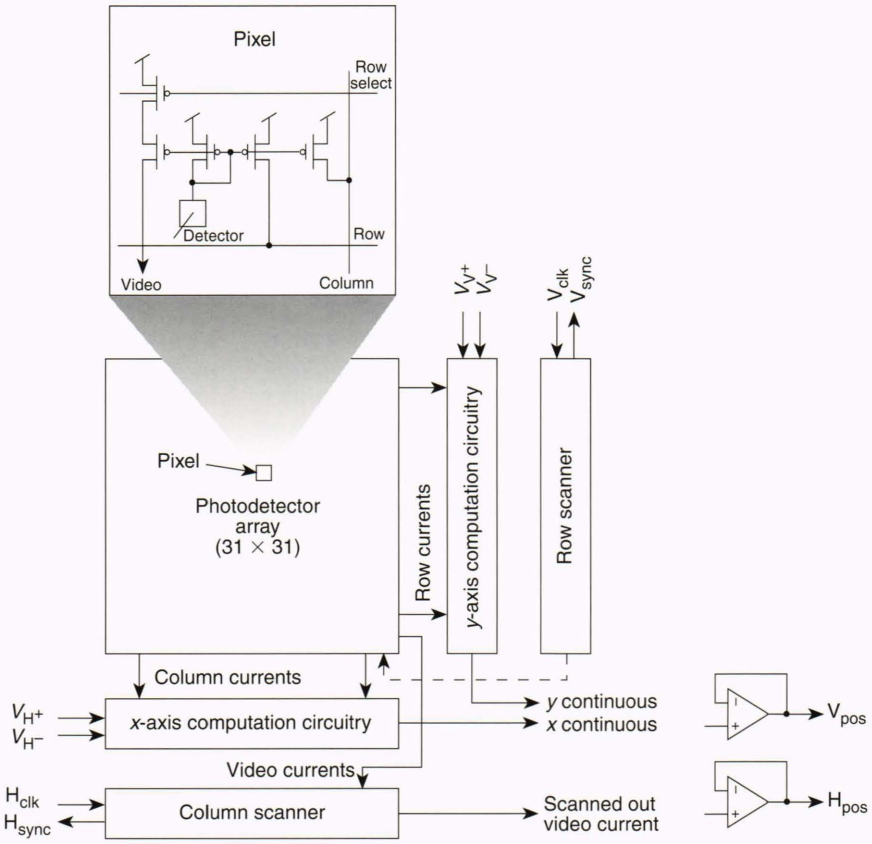


Figure 6. Retina video-centroid chip architectural block diagram. The detector array consists of our silicon retina, which provides continuous-time x and y signals and a scanned video signal. Current summing lines on each row and column enable each pixel to contribute to both the x and y centroids. Scanning circuitry reads out the photocurrent at each pixel. Global x and y computations, such as image centroid, are performed at the periphery. V_{H+} and V_{H-} are the voltages in the positive and negative horizontal directions; V_{V+} and V_{V-} are the voltages in the positive and negative vertical directions; H_{clk} and V_{clk} are the horizontal and vertical clocks; H_{sync} and V_{sync} are the horizontal and vertical synchronization pulses; H_{pos} and V_{pos} are the horizontal and vertical position outputs.

integrating detector is straightforward. Its sensitivity and fixed-pattern noise would then be similar to those of a CCD detector, but no readout would be required for the position estimate information. RVC chip attributes such as small size, low power dissipation, and simplicity extend the scope of potential applications even further.

The centroid computation on this RVC chip was largely ineffective because the silicon retina enhanced the high-frequency portions of the image and removed the low-frequency information. As a result, the image became insensitive to the sort of aggregating centroid computation that we are performing (see Ref. 13, p. 13, for a full derivation). We corrected the problem in the next design by using a current-mode absolute value circuit on the retina output prior to the centroid computation. The absolute value circuit effectively restores the DC value removed by the retina, allowing the centroid to be computed.

MOTION COMPUTATION WITH CORRELATING DETECTORS

We have also fabricated designs that can estimate both image position and velocity on-chip. Instead of a centroid calculation, these systems use a correlation algorithm that is inspired by biological image processing systems. The system outputs a continuous voltage proportional to either image displacement or velocity, which can be used as the feedback signal in a closed-loop tracking system without further processing.

Subthreshold analog VLSI techniques have been applied to the motion computation problem in a number of instances. Some earlier work[14] also attempted to directly implement a local gradient algorithm of the type proposed by Horn and Schunck,[15] but explicit computation of spatial and temporal derivatives, combined with the global constraint solution required by the algorithm, resulted in complex circuitry and a large pixel size. Other implementations also lack the simplicity of a biologically inspired solution (see, for example, Horiuchi et al.[16] and Etienne-Cummings, Van der Spiegel, and Mueller[17]). A notable exception is the work by Benson and Delbrück,[18] who used nearest-neighbor inhibition to produce a direction-sensitive and speed-proportional output. Unfortunately, their chip output was a pulse-width modulated signal with a limited range of velocity sensitivity, which is less easily integrated into a larger system.

Our approach is based on correlation, which is the method used in the visual systems of various insects.[19] The correlation algorithm, which is equivalent to Horn and Schunck's gradient approach,[20] is well-suited to our preference for current-mode circuits and does not require the explicit computation of derivatives. A basic unit of the Reichardt detector is shown in Fig. 7. The detector consists of two branches, each of which computes the product of that branch's pixel delayed by a time ϵ and the instantaneous value of the neighboring pixel. The right- and left-branch products are maximum for leftward and rightward motion, respectively. The detector output is the difference of these two products and, in the continuous limit for very close pixel spacings, is given by[21]

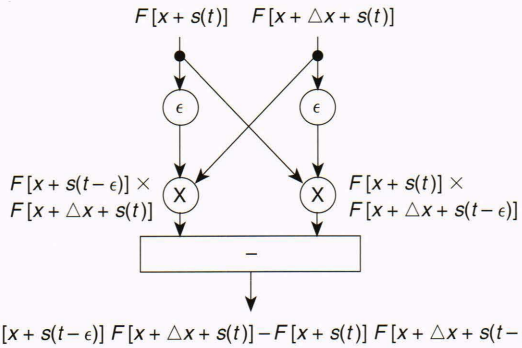


Figure 7. The Reichardt detector estimates both image position and velocity using a biologically inspired correlation algorithm. Each detector branch computes the product of that branch's pixel delayed by a time ϵ and the instantaneous value of the neighboring pixel. The detector output is the difference of the right- and left-branch products, which are maximum for leftward and rightward motion, respectively. $F(x)$ defines the input image. $F[x+s(t)]$ is the image brightness distribution at position x and time t for an image displacement of $s(t)$ over time.

$$-\epsilon \frac{ds}{dt}\left[\left(\frac{\partial F}{\partial x}\right)^2 - F\frac{\partial^2 F}{\partial x^2}\right]dx\,, \qquad (5)$$

where the input image is described by $F(x)$ and the image displacement over time is $s(t)$. The output is then proportional to the image velocity, ds/dt. Although the output also depends on the image itself, through the partial derivative terms, the image velocity factor weights the output for spatial frequency and contrast, thus emphasizing the more information-rich portions of the image.

To improve the robustness of the system, we again exploited the benefits of collective computation and implemented a one-dimensional array of 47 Reichardt detectors in silicon.[22] Figure 8 is a block diagram of our architecture showing two linked detectors. To make the design more space-efficient, each pixel was actually part of two detector elements, forming the right branch of one and the left branch of another. A silicon retina was used to enhance the image. In addition, all computations were current-mode except for the delay circuit, which was the most challenging element from a circuit design standpoint. The individual left- and right-branch outputs were summed on global current lines, in much the same way as with the VC chips, and were then fed into an off-chip instrumentation amplifier to generate the velocity signal.

The one-dimensional test chip was completely functional and correctly computed the image velocity; however, transistor mismatch caused a non-zero output, even for a stable image. This offset proved to be troublesome when we attempted to include the Reichardt chip in a closed-loop image stabilization system because it also varied with image movement and thus could not be separated from the true velocity signal. When the output was subsequently integrated to obtain image position information (the required control variable), the offset errors were also integrated and rapidly swamped the true signal.

To compensate for the offset error problems, we designed a second-generation chip that replaced the delay

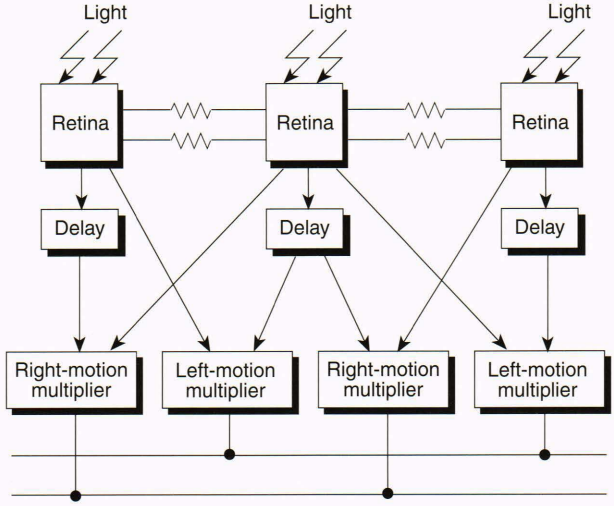


Figure 8. Two linked detectors of the 47-detector one-dimensional Reichardt detector array, which correlates image position and motion. A silicon retina enhances the image. All computations are current-mode except for the delay circuit. The individual left- and right-branch outputs are summed on current lines, similar to the video-centroid chip, and output off-chip for velocity calculations. To eliminate drift, the delay circuit was replaced with a sample-and-hold circuit, which makes the chip's output proportional to image displacement rather than velocity. (Reprinted from Meitzler et al.[23] by permission. © 1993 IEEE.)

circuit with a sample-and-hold circuit.[23] The sample-and-hold alters the Reichardt detector response equation to

$$s\left[\left(\frac{\partial F}{\partial x}\right)^2 - F\frac{\partial^2 F}{\partial x^2}\right]dx \qquad (6)$$

so that the chip's output is now proportional to the image displacement s, rather than velocity. Because image displacement cannot be computed while the image is being sampled, we maximized the duty cycle of the chip by using a sample-and-hold circuit with as long a hold time as possible. This approach also minimized the time spent resampling the image. The sample-and-hold circuit used is attributable to Vittoz et al.[24] and has a decay rate 10 to 20 times lower than that of a simple capacitor and switch arrangement. With the low-leakage sample-and-hold, the image needs to be resampled at most once per minute.

A one-dimensional sample-and-hold design was fabricated and again found to be functional. Testing is planned for a more practical, two-dimensional implementation in which two one-dimensional arrays are tiled perpendicular to each other. By arranging these arrays so that their photosensitive axes are in the center of the chip, we can sample along the horizontal and vertical centerlines of the image to compute displacement. Because sample-and-hold chips contain only nearest-neighbor connections, they are best suited for closed-loop systems where the displacement signal is used for stabilization, i.e., the image does not move more than 1 pixel. This limitation is necessary because excursions greater than 1 pixel will result in the correlation of unrelated portions of the scene and an unpredictable output.

Applications for the correlating motion chips are similar to those for the centroid chips, i.e., image tracking and stabilization systems where high bandwidth is important and space and power are at a premium. The sample-and-hold chip will be used with a system to stabilize solar granulation images, which will be carried aloft in a balloon. This project places strict limitations on weight and power dissipation yet demands high performance—in other words, an ideal candidate for our analog image processing chips. Future space missions that require high-bandwidth image stabilization loops are also appropriate applications for chips with integrating detectors instead of phototransistors.

FUTURE WORK

The development of subthreshold analog circuitry and low-power techniques for image processing is still largely in its infancy. Results to date demonstrate that we can design tailored systems with exceptional performance for their small size and energy requirements.

Further advances in this field hinge on a number of unresolved and developing issues. Our most immediate problem is to minimize the input fixed-pattern noise. The most promising approach is to combine a CCD-like integrating detector[25] with our present retinal circuitry. The sensitivity of this type of active retinal pixel should be comparable to that of CCD cameras, and fixed-pattern noise should be less than 0.1%. Such a sensor should find wide application in tracking systems.

On a more fundamental level, our algorithms will continue to evolve as the fields of neuroscience and psychology expand human knowledge of the visual system. Our challenge will be to map these discoveries into novel architectures in silicon. As our systems necessarily expand in complexity, we will have to consider multichip implementation, which raises issues of communication bandwidth and robust signaling across noisy wires. Research is under way at APL and elsewhere to develop practical solutions to these very difficult problems.

Another area we are investigating is improvements in design methodology and practices. Current transistor models for simulating subthreshold circuits are frequently of limited use for verifying a circuit's operation prior to fabrication.[26] We are also evaluating the larger variety of devices made available by improvements in fabrication technology. Understanding how to use devices such as floating-gate MOSFETs[27] will allow us to expand our repertoire of tools and design approaches.

A number of other applications besides image processing are being targeted for this technology. Subthreshold analog image processors have been proposed for use in laser tracking for optical communications, guidance systems, polarization-based imaging systems, and pattern-recognition neural networks. As fabrication technology and design skills advance, we expect these compact, low-power systems to play an expanding role.

REFERENCES

[1] Dowling, J. E., *The Retina: An Approachable Part of the Brain*, Balknap Press (Harvard University Press), Cambridge, MA (1987).

[2] Mead, C. A., *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, MA (1989).

[3] Mead, C. A., "Neuromorphic Electronic Systems," *Proc. IEEE* **78**(10), 1629–1636 (1990).

[4] Sedra, A. S., and Smith, K. C., *Microelectronic Circuits*, Third Ed., Saunders College Publishing, Fort Worth, TX (1991).

[5] Tsividis, Y. P., *Operation and Modeling of the MOS Transistor*, McGraw-Hill, New York (1987).

[6] Andreou, A. G., and Boahen, K. A., "Neural Information Processing II," in *Analog VLSI: Signal and Information Processing*, M. Ismail and T. Fiez (eds.), McGraw-Hill, New York, pp. 358–413 (1994).

[7] Boahen, K. A., and Andreou, A. G., "A Contrast Sensitive Silicon Retina with Reciprocal Synapses," in *Advances in Neural Information Processing Systems*, J. E. Moody, S. J. Hanson, and R. P. Lippmann (eds.), Vol. 4, Morgan Kaufmann, San Mateo, CA, pp. 764–772 (1991).

[8] Andreou, A. G., Boahen, K. A., Pouliquen, P. O., Pavasović, A., Jenkins, R. E., and Strohbehn, K., "Current-Mode Subthreshold MOS Circuits for Analog VLSI Neural Systems," *IEEE Trans. Neural Networks* **2**(2), 205–213 (1991).

[9] Mead, C. A., and Delbrück, T., "Scanners for Visualizing Activity of Analog VLSI Circuitry," *Analog Integr. Circuits Signal Process.* **1**(2), 93–106 (1991).

[10] Deweerth, S. P., "Analog VLSI Circuits for Stimulus Localization and Centroid Computation, *Int. J. Comput. Vision* **8**(3), 191–202 (1992).

[11] Strohbehn, K., *Analog VLSI Video Computation Organization*, JHU/APL S1A-035-92 (1992).

[12] Pavasović, A., *Subthreshold Region MOSFET Mismatch Analysis and Modeling for Analog VLSI Systems*, Ph.D. Thesis, The Johns Hopkins University, Baltimore, MD (1991).

[13] Strohbehn, K., *RVC Silicon Retina*, JHU/APL S1A-168-92 (1992).

[14] Tanner, J. E., and Mead, C. A., "An Integrated Analog Optical Motion Sensor," in *VLSI Signal Processing, II*, S. Y. Kung, R. E. Owen, and J. G. Nash (eds.), IEEE Press, New York (1986).

[15] Horn, B. K. P., and Schunck, B. G., "Determining Optical Flow," *Artif. Intell.* **17**, 185–203 (1981).

[16] Horiuchi, T., Lazzaro, J., Moore, A., and Koch, C., "A Delay-Line Based Motion Detection Chip," in *Advances in Neural Information Processing Systems*, Vol. 3, Morgan Kaufmann, San Mateo, CA, pp. 406–412 (1990).

[17] Etienne-Cummings, R. R., Van der Spiegel, J., and Mueller, P., "Real-Time 2-d Analog Motion Detector VLSI Circuit," in *Int. Joint Conf. Neural Networks*, IEEE and INNS, IEEE, Piscataway, NJ, pp. IV426–IV431 (1992).

[18] Benson, R. G., and Delbrück, T., "Direction Selective Silicon Retina That Uses Null Inhibition," in *Advances in Neural Information Processing Systems*, J. E. Moody, S. J. Hanson, and R. P. Lippmann (eds.), Vol. 4, Morgan Kaufmann, San Mateo, CA (1991).

[19] Reichardt, W., "Autocorrelation, a Principle for the Evaluation of Sensory Information by the Central Nervous System," in *Sensory Communication*, Symp. Principles of Sensory Commun., W. A. Rosenbluth (ed.), MIT Press, Cambridge, MA, pp. 303–317 (1961).

[20] Andreou, A. G., Strohbehn, K., and Jenkins, R. E., *A Proposed Scheme for the Analog VLSI Implementation of the Hassenstein-Reichardt Motion Detector*, JHU/ECE-88/07, The Johns Hopkins University, Baltimore, MD (Nov 1988).

[21] Reichardt, W., and Guo, A., "Elementary Pattern Discrimination (Behavioural Experiments with the Fly *musca domestica*)," *Biol. Cybern.* **53**(5), 285–306 (1986).

[22] Andreou, A. G., Strohbehn, K., and Jenkins, R. E., "Silicon Retina for Motion Computation," in *1991 IEEE Int. Symp. Circuits and Systems*, IEEE, Singapore, pp. 1373–1376 (1991).

[23] Meitzler, R. C., Andreou, A. G., Strohbehn, K., and Jenkins, R. E., "A Sampled-Data Motion Chip," in *36th Midwest Symp. Circuits and Systems*, IEEE, Detroit, MI, pp. 288–291 (1993).

[24] Vittoz, E. A., Oguey, H., Maher, M. A., Nys, O., Dijkstra, E., and Chevroulet, M., "Analog Storage of Adjustable Synaptic Weights," in *VLSI Design of Neural Networks*, U. Ramacher and U. Ruckert (eds.), pp. 47–63, Kluwer Academic Publishers, Boston, MA (1991).

[25] Fossum, E. R., "Active Pixel Sensors: Are CCDs Dinosaurs?" in *SPIE Proc. Charge-Coupled Devices and Solid State Optical Sensors III*, Vol. 1900, M. M. Blouke (ed.), SPIE, San Jose, CA, pp. 2–14 (Feb 1993).

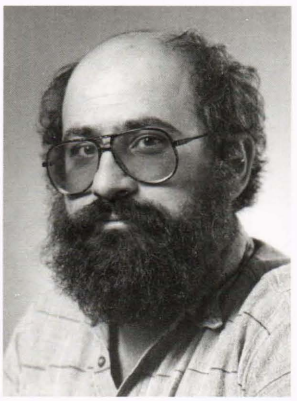
[26] Meitzler, R. C., and Andreou, A. G., "On the Simulation of Analog VLSI Systems Operating in the Subthreshold and Transition Regions," in *Proc. Tenth Biennial UGIM Symp.*, IEEE, Research Triangle Park, NC, pp. 145–150 (1993).

[27] Yang, K., and Andreou, A. G., "Subthreshold Analysis of Floating-Gate MOSFET's," in *Proc. Tenth Biennial UGIM Symp.*, IEEE, Research Triangle Park, NC, pp. 141–144 (1993).

ACKNOWLEDGMENTS: This work was supported by the APL Independent Research and Development Program. The chips were fabricated by the Metal-Oxide-Semiconductor Implementation Service (MOSIS) at the University of Southern California. Bernie Kluga provided invaluable assistance during RVC chip testing with the APL Astronomy Club's C-11 telescope, and fabricated the image-motion camera installed at Sunspot, NM. Xiao Li Sun performed electrical testing on the VC chips. David Rust provided interesting application opportunities and enthusiastically supported this developing technology. We thank the Analog VLSI Group at The Johns Hopkins University Homewood Campus, and especially Philippe Pouliquen and Kwabena Boahen for endless stimulating and useful discussions and ideas. Bedo Pain at the Jet Propulsion Laboratory introduced us to the idea of the CCD-like active pixel.

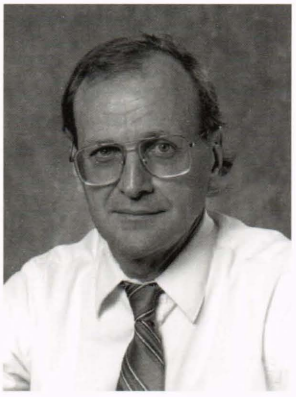
THE AUTHORS

KIM STROHBEHN received B.S., M.S., and Ph.D. degrees in electrical engineering from Iowa State University in 1975, 1976, and 1979, respectively. Dr. Strohbehn is a Senior Engineer and member of the Computer Science and Technology Group in the Space Department. He also teaches the Analog VLSI course with Dr. Andreou in The Johns Hopkins University G. W. C. Whiting School of Engineering Graduate Program. His interests include the application of analog VLSI to astronomical instrumentation.



ANDREAS G. ANDREOU received M.S.E. and Ph.D. degrees in electrical engineering and computer science from The Johns Hopkins University in 1983 and 1986, respectively. From 1987 to 1989, he was a postdoctoral fellow and associate research scientist at JHU. He is currently an Associate Professor of electrical and computer engineering at JHU and a member of the APL Senior Staff in the Computer Science and Technology Group. His research interests are device physics, analog VLSI, and neural computation.

RICHARD C. MEITZLER received a B.S. degree in electrical engineering from the Pennsylvania State University in 1988 and an M.S. in electrical engineering from The Johns Hopkins University in 1990. He is currently a Ph.D. candidate in the Electrical and Computer Engineering Department at The Johns Hopkins University, investigating analog CMOS circuits for vision applications and MOS device modeling.



ROBERT E. JENKINS received a B.S. degree in engineering and an M.S. degree in physics from the University of Maryland in 1965. He is a lecturer in The Johns Hopkins University School of Engineering and a Principal Staff Engineer at APL. He is also a member of the Program Committee for the Part-Time Master's Degree Program in Electrical Engineering at JHU. He joined APL in 1961 and is currently Group Supervisor of the Space Department's Computer Science and Technology Group and program manager for the Space Department's IR&D program. He has over 20 years of experience in computing systems.