Analog VLSI Architectures for Motion Processing: From Fundamental Limits to System Applications

RAHUL SARPESHKAR, JÖRG KRAMER, GIACOMO INDIVERI, AND CHRISTOF KOCH

Invited Paper

This paper discusses some of the fundamental issues in the design of highly parallel, dense, low-power motion sensors in analog VLSI. Since photoreceptor circuits are an integral part of all visual motion sensors, we discuss how the sizing of photosensitive areas can affect the performance of such systems. We review the classic gradient and correlation algorithms and give a survey of analog motion-sensing architectures inspired by them. We calculate how the measurable speed range scales with signal-to-noise ratio (SNR) for a classic Reichardt sensor with a fixed time constant. We show how this speed range may be improved using a nonlinear filter with an adaptive time constant, constructed out of a diode and a capacitor, and present data from a velocity sensor based on such a filter. Finally, we describe how arrays of such velocity sensors can be employed to compute the heading direction of a moving subject and to estimate the time-to-contact between the sensor and a moving object.

I. INTRODUCTION

Various applications in automotive navigation, robotics, and remote sensing require sensors for processing visual motion that are small, consume little power, and work in real time. Considering the type of environments humans are typically exposed to, we shall use the term "real time" in its common anthropocentric meaning, i.e., for time delays not exceeding a few tens of milliseconds. Since motion-sensing algorithms have a large computational overhead, most real-time machine-vision applications require special-purpose parallel hardware for computing motion across the entire image. Parallel hardware implementations are particularly attractive if image acquisition and motion computation can be integrated on a single silicon chip. Such *smart-vision* chips could be used in stand-alone applications, such as tracking cars, or as front-ends in conventional machine vision systems [1], [2].

Because image irradiance is a continuous function of time, asynchronous circuit implementations are preferable to clocked implementations. The latter introduce temporal aliasing artifacts that can significantly compromise time-sensitive computations, such as those associated with optical flow.

Analog processing is more economic in terms of silicon area and power than digital processing of comparable complexity, and thus makes higher pixel densities possible. Its main drawback is its lack of precision, but high-precision motion processing is usually not possible, because of noisy input data and fundamental computational problems associated with the estimation of the velocity field from the optical flow (at best, estimating optical flow is numerically ill-conditioned, and at worst, ill-posed [3], [4]).[1]

In this paper, we move in bottom-up fashion from low-level issues having to do with photoreceptors to higher-level topics having to do with properties of the entire motion flow field. Most of the issues we discuss are general and pertain to all analog VLSI motion processing systems, though some reflect our own biases and the particular details of circuits that we have built.

The first stage of all visual motion-processing systems is phototransduction. Consequently, in Section II we start by discussing an issue that has been largely ignored but which we believe to be important—how the size of the light-collection area of a photoreceptor affects the photoreceptor's, and consequently the motion sensor's, signal-to-noise ratio (SNR). In Section III, we review the correlation and gradient motion-processing algorithms, since they have inspired the architectures of many analog VLSI chips.

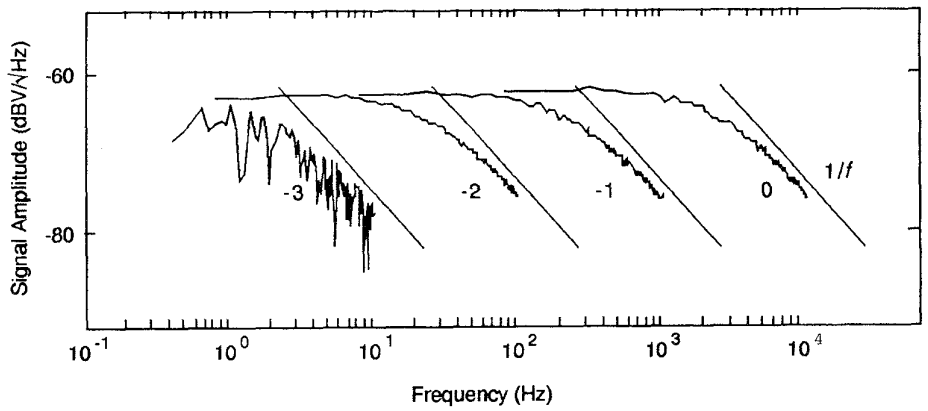
Manuscript received September 1, 1995; revised February 15, 1996. This work was supported by Daimler–Benz, the Office of Naval Research, the Center for Neuromorphic Systems Engineering (part of the NSF Engineering Research Center Program), and by the Office of Strategic Technology of the California Trade and Commerce Agency.

The authors are with the Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA 91125 USA.

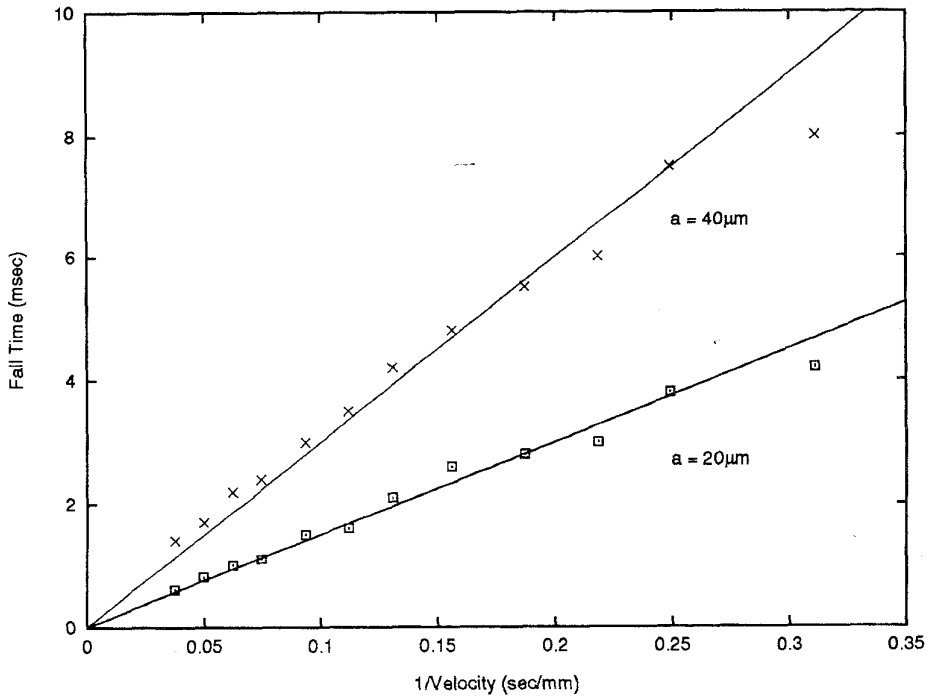
C. Koch is also with the Division of Biology, California Institute of Technology, Pasadena, CA 91125 USA.

Publisher Item Identifier S 0018-9219(96)04996-1.

[1] Given the topic of this paper, we make no distinction here between the optical flow field induced by the time-varying image irradiance and the underlying 2-D velocity field, a purely geometrical concept [3], [5].



(a)



(b)

Fig. 1. *Frequency response and voltage decay time of a single CMOS photoreceptor.* (a) The irradiance was attenuated in steps of factors of ten by interposing neutral density filters between a LED and the photoreceptor. The number given near each curve is the logarithm of the attenuation factor. The slope of each response curve asymptotes to a $1/f$ dependence, characteristic of a first-order low-pass filter. The dBV units refer to signal amplitude with respect to a 1 V signal in decibels. (b) Fall time of the photoreceptor's output voltage signal as a function of the velocity of a moving edge for photoreceptors with different sizes. The fall time is inversely proportional to the velocity of the edge and proportional to the size of the photoreceptor's light-collecting area. This confirms our notion that the fall time is ultimately limited by the time-of-travel of the edge image across the light-collecting area if the photoreceptor's bandwidth is large enough to follow the dynamics of the stimulus. The optical stimulus was a sharp high-contrast edge on a rotating drum with diameter $d = 11.4$ cm, whose surface was situated $l = 29$ cm away from the lens imaging it onto the chip. The focal length of the lens was $f = 113$ mm. The velocity of the edge image on the chip was computed as $(f/l)(\pi d/T)$ where T is the rotation period of the drum.

In Section IV, we present a survey of various analog VLSI chips that have been reported in the literature. In Section V we analyze what determines the minimum and maximum speeds that a sensor can handle. We calculate how these limits depend on the SNR for a classic correlation (Reichardt) sensor. In Section VI we show how a nonlinear filter with an adaptive time constant as implemented with a simple diode-and-capacitor circuit can be used to design a motion sensor that operates over a large dynamic range. We present theory and data that illustrate the large dynamic

range that is achievable. In Section VII we describe an implementation of such a motion sensor and present data that shows the working of the sensor over a wide range of velocities, contrasts and light levels. In Section VIII, we discuss how such a motion sensor can be used in single-chip systems for extracting high-level information from the motion-flow field, such as the focus-of-expansion or the time-to-contact with a moving object. We conclude the paper with a short summary in Section IX.

In this paper, we focus primarily on one-dimensional (1-D) motion-sensing problems. The issues we discuss are all pertinent to two-dimensional (2-D) motion-sensing problems as well; for 2-D problems, however, motion integration [6] (e.g., solving the *aperture* problem [7], [8]) would need to be implemented on chip as well. Part of this work has been previously reported in [9], [10], and [38].

II. Photoreceptor Sizing and System Performance

Photoreceptors are an integral part of all visual motion-sensing circuitry. The transduction of light to electric current is usually accomplished with a photodiode held at a fixed or nearly fixed reverse bias. The light collection area of the photodiode is an important parameter that affects the performance of the motion sensor in many ways. We shall start by assuming that each pixel has a square light-collection area of $a \times a$. We illustrate how various system parameters scale with a. We shall only show the forms of the dependence on a without constant factors.

Our theory and measurements are based on the adaptive photoreceptor characterized extensively in [11], but our discussion is of a general nature and applies to many common photoreceptor designs.

A. Spatial and Temporal Filtering

For a photodiode, the photocurrent I is proportional to Ea^2, where E is the irradiance and a^2 is the light-collection area. The total capacitance on the photodiode sensing node is given by the sum of the junction capacitance—proportional to the area a^2 of the photodiode's pn junction—and a contribution from parasitics due to the sensing and amplifying circuitry that is part of every photoreceptor. Thus the total capacitance C is proportional to (a^2+p), where p is the area-independent contribution from the parasitics. To a good approximation, the photoreceptor can be characterized as a first-order temporal filter. Its bandwidth B (defined in the usual fashion as the frequency at which the response is reduced by 3 dB) is inversely proportional to the total capacitance and increases linearly with the photocurrent.[2] Thus

$$B \sim \frac{a^2}{a^2+p}E. \tag{1}$$

The linear dependence of the bandwidth on the irradiance E can be seen from Fig. 1(a), showing the response of a CMOS photoreceptor to small-signal, white noise for different irradiance levels. The data were provided by Delbrück [11]. For a given irradiance E, the bandwidth rises with increasing a but eventually asymptotes to a constant value independent of a.

A sharp moving edge imaged onto a photoreceptor is spatially filtered due to the finite size of the photoreceptor's light-collecting area. Consequently, the rise and fall times of the photoreceptor's output waveform in response to the motion of a sharp irradiance gradient increase with a. In fact, if the photoreceptor does not impose any bandwidth limitations on the rise and fall times, they are limited by the time-of-travel of the edge across the width a. Thus if we define t_{rf} to be the 10–90% rise or fall time, then

$$t_{rf} \sim \frac{a}{v} \tag{2}$$

where v is the velocity of the moving step edge and the constant of proportionality is $0.9 - 0.1 = 0.8$. Fig. 1(b) shows this $1/v$ dependence for the fall times of two photoreceptors with $a = 40$ μm and $a = 20$ μm, respectively. The fits to the measured data have constants of proportionality of 0.75, and the ratio of the two slopes is two. These numbers agree well with our rather simple theoretical considerations.

Typical visual patterns include different spatial frequencies. Thus it is of interest to find out how the photoreceptor responds to a given spatial frequency input. A sinusoidal input with spatial frequency f_x and contrast c that is moving with velocity v will be spatially and temporally filtered by the photoreceptor. The effect of the square-box spatial filter is given by a sinc function ($\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$). The low-pass temporal filter is first-order and characterized by the bandwidth B. Photoreceptors are often designed such that their transient output voltage is proportional to the contrast $c = \Delta E/E$ and independent of the absolute irradiance E. The photoreceptor output voltage V_{ph} is then given by

$$V_{\mathrm{ph}} \sim c\left(\frac{\sin(\pi a f_x)}{\pi f_x}\right)\left(\frac{1}{\sqrt{\left(1+\left(\frac{f_x v}{B}\right)^2\right)}}\right). \tag{3}$$

B. Noise

The thermal noise at the photoreceptor output is caused by the amplification of shot-noise diffusion currents in the photodiode and transistors in the photoreceptor. The theory and measurements of thermal noise in subthreshold MOS devices, resistors, and photoreceptors may be found in [12]. The mean-square thermal voltage noise at the output of the photoreceptor is proportional to kT/C and thus given by

$$V_{\mathrm{th}}^2 \sim \frac{kT}{a^2+p}. \tag{4}$$

In practice however, for ac-driven light sources that are typical for indoor environments, optical line noise is the dominant source of noise in the photoreceptor. In the United States, the line frequency of 60 Hz results in 120 Hz

[2] The bandwidth is proportional to g/C, where $g = I/V_k$ is the conductance of the photodiode sensing node and V_k is a multiple of the thermal voltage kT/q. Since V_k is almost always constant, the bandwidth is proportional to I/C.

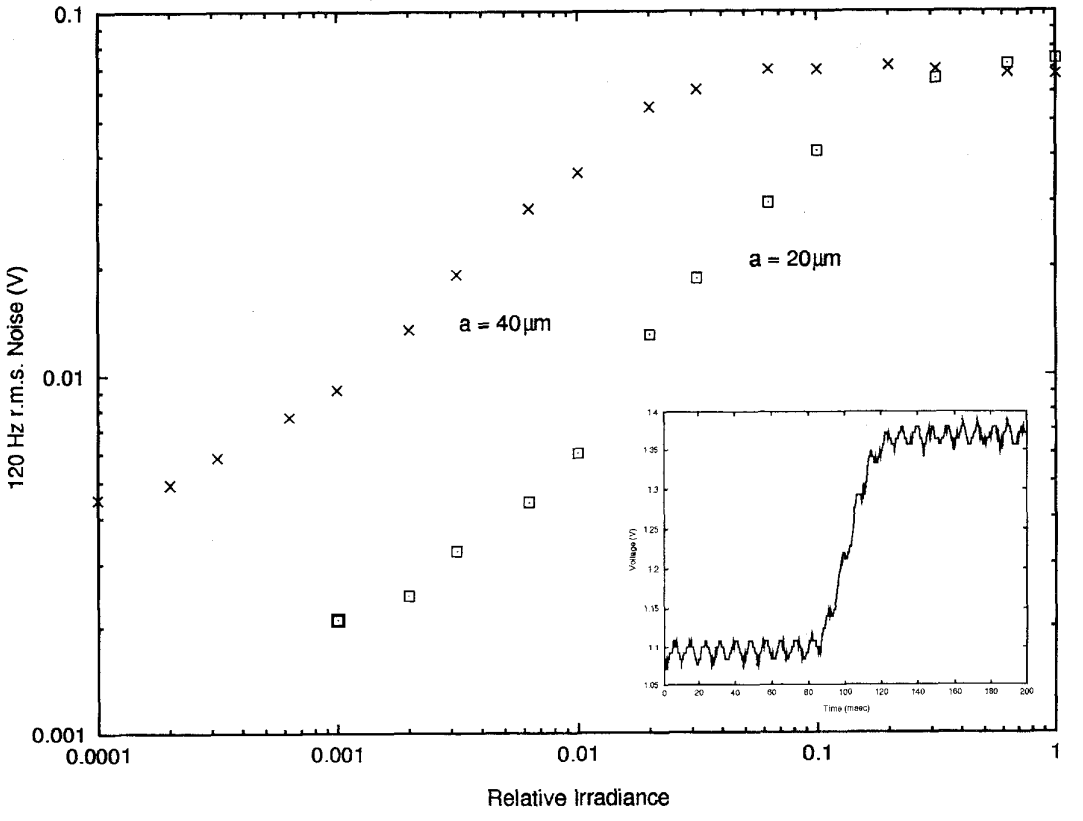


Fig. 2. *Photoreceptor noise as a function of image irradiance.* The output noise of the photoreceptors of Fig. 1(b) due to 120 Hz line noise of fluorescent ambient lighting shows that at high irradiance levels this type of noise is independent of the photosensitive area. The noise was calculated from measurements of the power of the first seven harmonics of 120 Hz (1–7f). Neutral-density filters were interposed between the light source and the photoreceptors to attenuate the irradiance level in a controlled manner. The noise at high irradiance levels is identical for the two photoreceptors with light-collecting areas of 40 μm $\times$ 40 μm and 20 μm $\times$ 20 μm, respectively, since the signal gain of the photoreceptor is invariant with its light-collection area. However, because of its lower bandwidth (due to the parasitic capacitance on the photodiode sensing node), the smaller photoreceptor starts to filter out the noise at higher light levels. The inset shows a typical scope trace of a photreceptor's output voltage in response to a high contrast edge.

line noise.[3] Fig. 2 shows 120 Hz noise measurements for the two photoreceptors with $a = 40$ μm and $a = 20$ μm as a function of irradiance for fluorescent lighting. The inset shows a scope trace of the output voltage of a photoreceptor in response to an edge signal in the presence of 120 Hz incandescent optical line noise. The relative modulation of the photodiode current $\Delta I/I$ is equal to the irradiance modulation if the photoreceptor is not bandwidth-limited. Since the voltage gain of the photoreceptor is only dependent on $\Delta I/I$ and not on the absolute current level I, both photoreceptors have the same amount of gain with respect to optical line noise at high light levels, where there are no filtering effects. However, at low light levels the slow response of the photoreceptor filters out the 120 Hz signal. The data shown in Fig. 2 confirms the prediction of (1) that the 20 μm photoreceptor has a smaller bandwidth than the 40 μm photoreceptor, and therefore starts filtering out the 120 Hz signal at higher light levels.

[3] In other countries, the line noise is at 100 Hz due to a line frequency of 50 Hz.

C. Signal-to-Noise (S/N) Considerations

We first discuss S/N considerations for the case in which thermal noise is dominant in the photoreceptor (dc-lighting or sunlight conditions). Subsequently, we treat the case in which optical line noise dominates (ac-lighting conditions).

1) Thermal Noise: If thermal noise is the dominant form of noise in the photoreceptor, we see from (4) that we can reduce it by increasing a. From (1) we observe that increasing a improves the bandwidth B and reduces the effect of temporal filtering on the signal. From noise and temporal-filtering considerations, therefore, it is advantageous to have a very large photoreceptor width a. However, (3) implies that increasing a increases spatial blurring, thereby reducing spatial resolution and the output signal of the photoreceptor. From spatial filtering considerations, therefore, it is disadvantageous to increase a. In order to understand which of these effects dominates, we need to examine how they scale with a.

According to (4), the thermal noise power is reduced as $1/(a^2 + p)$ by increasing a. Equation (3) shows that the

rising bandwidth with increasing a [see (1)] improves the SNR by increasing the signal. The sinc function has its first zero at $f_x = 1/a$ and we may approximate it by a low-pass spatial filter with a pole at $1/a$. Hence, the lowest spatial frequencies are unattenuated, while spatial frequencies significantly above $1/a$ are attenuated in power like $1/a^2$. Thus, the ratio of signal power to noise power scales like (a^2+p) for low spatial frequency inputs, and like $(a^2+p)/a^2$ for high spatial frequency inputs, if we assume constant bandwidth. Given that the bandwidth improves with increasing a, the SNR rises even more strongly with increasing a. Therefore, we conclude that if thermal noise is the dominant form of noise in the photoreceptor, the SNR at its output is improved by having a large a. Intuitively, this makes sense, since the way to eliminate intrinsic noise is to average the input over as large a photon-collecting area as possible.

However, it is not sufficient to view the output of the photoreceptor in isolation: with increasing a, the thermal noise from the photoreceptor becomes reduced until the thermal noise behavior starts to be dominated by subsequent motion-processing stages and the input-referred thermal noise becomes independent of a. Assuming that this happens for large values of a, we note that the temporal bandwidth is also independent of a. The overall S/N ratio of the system then starts to degrade with increasing a, because of the $1/a^2$ signal attenuation due to spatial filtering.

2) Optical Line Noise: If optical line noise is the dominant form of noise in the photoreceptor, increasing the size of the photoreceptor does not reduce the noise. In fact, it may be shown that the bandwidth improvement with increasing a increases the 120 Hz noise more than the input signals with lower temporal frequencies. Furthermore, the input signals suffer from spatial filtering and the global optical line noise does not. Therefore, we conclude that if optical line noise is the dominant form of noise in the photoreceptor, the SNR at its output is degraded by having a large a. Note that this conclusion is exactly the opposite of the one we reached for thermal noise. In indoor environments the optical line noise typically dominates the thermal noise, as can be seen from the inset of Fig. 2. It is therefore advantageous to work with a small a in these cases.

D. General Rectangular Designs

Our discussion centered on the common case of photoreceptors with light-collection areas of $a \times a$. Typically, such square photoreceptors are chosen for 2-D grids. For 1-D arrays of motion sensors with associated photoreceptor circuitry, the SNR of the photoreceptors is optimized by light-collection areas with dimensions $b \times a$, where $b \gg a$ and a is along the motion-computing direction. This strategy allows us to reduce thermal noise (by making b as large as possible) without causing increased spatial filtering of the signal (by keeping a as small as possible). For large enough areas ($ab \gg p$), the bandwidth is almost independent of area because the junction capacitance is much greater than any parasitic capacitance [see (1)]. In the presence of optical line noise, the SNR is independent of b for given a and $ab \gg p$, due to the fact that neither the signal nor the noise is affected by b.

Before we conclude our discussion of photoreceptor sizing, it is worth mentioning that for 2-D images, the details of the analysis differ because of concerns like edge orientation, but the general tradeoffs with respect to a remain valid.

III. Motion Algorithms

Algorithms for estimating optical flow can be divided into two main groups [6], [13]. In *intensity*-based methods, the image irradiance, or some linearly filtered version of it, is directly used to estimate the optical flow throughout the image. The two most popular types of intensity-based motion algorithms are *gradient* methods and *correlation* or *spatio-temporal energy* methods. Motion algorithms of the second group are known in computer vision as *token*-based methods and are associated in the psychophysical literature with *long-range* motion. They first extract particular features in the image, such as edges, corners, or higher-level features by using nonlinear operators, and then estimate velocity in a subsequent stage at sparse locations throughout the image by tracking these features across time. All motion-sensing algorithms suffer from the fact that estimating optical flow is at the very least a numerically ill-conditioned and frequently an ill-posed problem [3], [4]. One key difficulty with methods working on discrete space or time intervals (or both) is known as the *correspondence* problem: which feature at a given time or location corresponds to what feature at a different time or location [14], [15].

As will be shown in Section IV, various analog VLSI velocity sensors have been designed using intensity-based motion algorithms. Accordingly, we will briefly summarize the main features of correlation and gradient algorithms and present their most popular versions. A more detailed review may be found in [13], and a comparison of the pros and cons of these algorithms may be found in [7].

A. Correlation or Reichardt Algorithm

The Reichardt algorithm was first proposed to explain the optomotor response to moving patterns in beetles and flies [16]. In more recent years, similar algorithms—termed second-order or spatio-temporal energy methods—have been proposed to explain motion perception in humans and other primates [17]–[19]. Common to all these methods is that the image irradiance $E(x,t)$ is passed through a linear spatio-temporal filter and multiplied with a delayed version of the filtered irradiance from a neighboring receptor. These algorithms calculate a quadratic functional that depends on the velocity and the contrast of the optical input.

The output of the original Reichardt model [16] is obtained by first multiplying the spatially filtered brightness signal from one photoreceptor with a delayed or low-pass-filtered version of the signal from an adjacent photoreceptor, as shown in Fig. 3. This quadratic signal is large when

Table 1 Implemented Analog VLSI Motion Sensors

Reference	Year	Input	Algorithm	Dimension	Array Size	Interaction
[25]	1986	log (irradiance)	gradient	2	8 × 8 rectangular	global
[28]	1991	center surrounded	correlation	1		global
[29]	1991	temporal ON edges	binary correlation	1	27	global
[34]	1992	temporal ON edges	time-of-travel	1	47 × 47 rectangular	local
[30]	1993	temporal edges	correlation	3 × 1	26 × 26 hexagonal	global
[31]	1993	spatial edges	binary correlation	1		local
[35]	1993	spatial edges	time-of-travel	2 × 1	5 × 5 rectangular	local
[10]	1995	temporal ON edges	time-of-travel	1		local
[36]	1995	temporal ON edges	time-of-travel	1	8	local

Most sensors operated on tempral or spatial edges, rather than directly on image irradiance. One sensor [25] estimated the global average of a 2-D velocity field. All other circuits estimated 1-D motion along one or more noninteracting directions. Earlier implementations averaged the motion signals globally across the sensor array to compute uniform motion, while some of the newer ones provided local motion information at each pixel site to estimate the entire optical flow field.

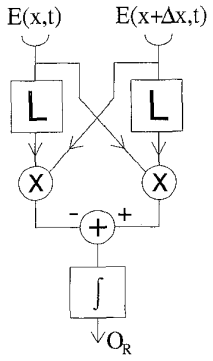


Fig. 3. *Architecture of a classic Reichardt motion sensor.* The inputs $E(x,t)$ and $E(x+\Delta x,t)$ represent the image irradiances at the locations x and $x+\Delta x$, respectively. The blocks labeled L represent the first-order low-pass filters with time constant τ. The irradiance at one input is multiplied with the low-pass-filtered irradiance of the neighboring input. The difference of the results for both directions (opponency stage) is integrated over time to yield the direction-dependent output signal O_R.

the low-pass filter delay compensates for the time-of-travel due to motion; consequently, for a given stimulus, it is maximum for a particular speed in one direction of motion. In an opponency stage, the outputs of two multiplication stages, corresponding to opposite directions of motion, are subtracted and the result is then integrated in time to yield the output of the sensor. Since this operation is akin to an autocorrelation, this sensor is also known as a *correlation* sensor.

Note that this sensor is sensitive to the direction of motion, as coded in the sign of the output, but *not* to velocity as such. Rather, it may be viewed as a linearly-separable spatial-frequency and temporal-frequency filter. A sinusoidal stimulus pattern, $E(x,t)$, may be represented as $E(x,t) = E_0(1 + c\sin(kx - \omega t))$, with c denoting the image contrast, k the angular spatial frequency, $\omega = kv$ the angular temporal frequency, and v the velocity of the pattern. If the spacing between neighboring photoreceptors is Δx and the time constant of the low-pass filter is τ, it is straightforward to show that the output of a Reichardt sensor is given by

$$O_R \sim c^2 \sin(k\Delta x)\left(\frac{\omega\tau}{1+(\omega\tau)^2}\right). \tag{5}$$

Thus the sensor has its maximum output at the temporal frequency $1/\tau$ and at the angular spatial frequency $\pi/2\Delta x$, which corresponds to half the Nyquist spatial-sampling rate. The temporal frequency $\omega = kv$ is signed to be positive for velocities in one direction and negative for velocities in the other.

B. Gradient Algorithm

Gradient methods or, more generally, differential methods, exploit the relationship between the velocity and ratios of first-order or higher-order temporal and spatial derivatives of the image irradiance distribution and are commonly used in machine vision [8], [20]–[23]. These methods yield a direct estimate of the optical flow field, but the implementation of spatial derivatives with offset-prone analog circuits is problematic.

Gradient algorithms are based on *a priori* assumptions about the image brightness (see Section VIII). The simplest gradient algorithm estimates velocity from the constraint $dE(x,t)/dt = 0$, i.e., that the image brightness does not change over time. In the 1-D case, the velocity then amounts to

$$v = -\frac{\partial E/\partial t}{\partial E/\partial x}. \tag{6}$$

More sophisticated versions [23], [24] require the computation of higher-order spatial and temporal derivatives; such computations, however, are not suited to implementation with analog VLSI.

IV. Review of Analog VLSI Motion Sensors

In the following, we review VLSI implementations of continuous-time, analog motion sensors that implement intensity-based and token-based algorithms. We restrict ourselves to sensors that incorporate the photoconversion stage and the motion-processing circuitry on a single chip. The reviewed sensors are summarized in Table 1.

Among the first circuits operating on a gradient algorithm was an "optical mouse" chip for estimating uniform image velocity in two dimensions [25].[5] The circuit consisted of an 8 × 8 array of velocity-sensing elements. Their outputs were averaged to correct for offset effects and the aperture

[5]The first optical mouse chip built implemented a digital algorithm [26].

problem. The circuit effectively measured a single velocity value within a limited range for high contrasts, whereas with decreasing contrast the reported motion tended to zero for a given image velocity. The precision of the local velocity measurements was low due to large circuit offsets and even with the implemented signal aggregation across the entire array the performance was poor.

Due to the intrinsic difficulty of accurately computing local spatial and temporal derivatives using low-precision hardware [27], subsequent work concentrated on implementing correlation algorithms, since they show superior numerical stability. A 1-D spatial-correlation sensor was built [28] where a spatially filtered version of the input signal was fed into a circuit that multiplied the delayed signal of one pixel with the undelayed signal of the adjacent pixel on either side. The two signals were averaged along the array and used to interpolate the correlation peak in the subpixel range. For a high-contrast bar stimulus the output signal was approximately linearly dependent on speed. Temporal aliasing was observed when the image was displaced by more than one pixel within a single time delay interval. The spatial response of the input filter was illumination-dependent. Because the shape of the correlation peak for the implemented algorithm depended on the stimulus, the circuit could only measure the velocity of a stimulus to which it had been calibrated.

Since all purely correlation-based methods yield output signals that strongly depend on contrast and illumination, as well as on velocity, general interest shifted toward hybrid methods that perform correlation-type motion computations on image tokens extracted by low-level feature detectors. The circuits described in the remainder of this section used edges as image tokens. They either identified rapid temporal-irradiance transients as temporal edges, or sharp spatial-irradiance gradients as spatial edges.

One such sensor was inspired by the auditory system of the barn owl and integrated in a 27 element 1-D array [29]. At each pixel site, a binary voltage pulse of fixed width was triggered in response to a temporal dark-bright or ON edge. Pulses from adjacent pixels were propagated through two parallel delay lines from opposite directions. The velocity of the edge stimulus was extracted from the position of their meeting point (*coincidence detection*). For given bias settings, the circuit measured velocity over a range of one order of magnitude across a decade of dc illuminance down to low contrasts. It was not operable under ac incandescent lighting, however, because the optical line noise caused spurious edge signals to be triggered. This problem could only be solved by using additional filtering circuitry. Other drawbacks of the system were the limited detectable velocity range for a given delay setting and the large area consumption of the delay lines.

A 2-D scheme, implemented as a hexagonal array of 26×26 pixels [30], compared the velocity of temporal edges traveling across the image plane with the preset velocity of pulses propagating through delay lines along three directions. If the apparent velocity of an edge along a delay line matched the propagation speed of a pulse in the delay line, this pulse was periodically reinforced and its amplitude increased. If the velocities did not match, the pulses in the delay line gradually decayed. The pulse amplitude at each pixel was measured with a nonlinear circuit. The output signal in response to a traveling edge increased along the delay lines, while the velocity-tuning curve sharpened up. The circuit responded down to low contrasts. The output signal of a given element depended on its position and the apparent velocity, contrast, and sharpness of the moving edge. A population of such arrays, tuned to different velocities, or a single array with adaptive tuning, would be necessary to measure velocity over an extended range. This scheme is therefore not suitable for the monolithic implementation of dense velocity-sensing arrays.

A binary correlation scheme was implemented in a 1-D array [31]. It used a spatial-edge detector, that convolved the image with a difference-of-exponentials kernel, implemented with two resistive grids with different resistances [32]. An edge of sufficiently high contrast triggered a voltage pulse of fixed amplitude and width. This pulse was then timed against the pulse from an adjacent pixel that was delayed by a fixed time interval. The output signal was a current pulse with a width equal to the overlap time of the two pulses. The system worked robustly for high-contrast edges under ac or dc lighting. However, it was not sensitive enough to respond to low-contrast edges and its response to nonoptimum velocities was ambiguous.

While such correlation-based algorithms are interesting models of how motion computations may be carried out in neurobiological circuits, most of them are unsuitable for implementation in integrated circuits that compute extended optical flow fields, because they are tuned to a narrow velocity range and tend to be expensive in silicon area. Algorithms that measure the time-of-travel of an image token between two fixed locations (inversely related to the velocity) are more attractive for implementation in compact arrays.

A circuit inspired by a model for 1-D direction selectivity in the rabbit retina [33] was implemented as a 2-D array of 47×41 pixels [34]. Inhibitory connections between neighboring pixels in one direction suppressed the response for motion in that direction, called the null direction. In the other direction, the preferred direction, voltage pulses were triggered by temporal ON edges and terminated by inhibition from the neighboring pixel. For low speeds and contrasts, however, the pulse amplitudes and widths were strongly dependent on speed and contrast and the pulses decayed before inhibition set in. For a given stimulus, the output pulse width therefore increased with speed at low speeds and decreased again with higher speeds, where it was limited by inhibition. In the null direction, the output was efficiently inhibited, unless the travel time of the edge was shorter than the onset delay of the inhibition, in which case spurious outputs were observed. The circuit responded down to low contrasts, but contrast-independent velocity measurements were only possible for high contrasts in a very limited velocity range.

Another time-of-travel chip was implemented in a 5×5 array [35]. It used a spatial-edge detector as an input stage and measured the time it took an edge to cross a pixel by tracking the maximum of the edge-detector output signal [35]. For sharp black-white edge stimuli, the chip showed an approximately inverse relationship of pulse width and speed over three orders of magnitude of speed and six orders of magnitude of ambient lighting. The contrast threshold for edge detection was given as 5%, but the performance of the motion sensor at intermediate and low contrasts was not reported. Velocity was computed along two perpendicular axes in the focal plane with two pairs of motion cells per pixel. No interaction between the two directions was implemented, i.e., no attempt was made to estimate 2-D motion. Due to the large amount of circuitry needed by this algorithm, it does not appear to be unsuitable for implementation in dense arrays.

Two time-of-travel algorithms were implemented with more compact circuits, that unambiguously encoded 1-D velocity over considerable velocity, contrast, and illumination ranges. Both used a temporal-edge detector as input stage, that responded to ON edges [9], [10]. In one circuit [10], [31], an edge signal generated a voltage pulse of fixed amplitude and width at each pixel location. The pulses from two adjacent locations were fed into two motion circuits, one for each direction. For motion in the preferred direction, such a motion circuit output a pulse whose width corresponded to the time interval during which the input pulses overlapped. For motion in the null direction no output pulse was generated. The other circuit [36] was based on a scheme where three adjacent edge detectors contributed to the measurement of a velocity value. In the preferred direction of motion, an edge signal from the first detector facilitated the response of the second detector, whose signal then triggered a binary output voltage pulse. An inhibition signal from the third detector terminated the pulse. In the null direction, inhibition set in first, thereby suppressing the triggering of an output pulse, before being released by the facilitation signal. The outputs of both circuits were contrast-invariant down to approximately 15% contrast and illumination-invariant over more than two orders of magnitude under ac incandescent room illumination.

V. Limits of Operation of Motion Sensors

When comparing the merits of different motion sensors, it is crucial to understand the ultimate limits on their performance. These limits are important in determining the sensor's dynamic range of operation.

For a given pixel spacing Δx, the maximum and minimum speeds that a motion sensor can transduce correctly are determined by the smallest and largest time intervals, $\Delta t_{\min}$ and $\Delta t_{\max}$, that it can handle respectively, according to

$$v_{\max} = \frac{\Delta x}{\Delta t_{\min}} \tag{7}$$

and

$$v_{\min} = \frac{\Delta x}{\Delta t_{\max}}. \tag{8}$$

The values of $\Delta t_{\min}$ and $\Delta t_{\max}$ may be set by two types of limits, here referred to as *parametric* and *noise* limits. We shall discuss the parametric limits briefly and the noise limits in more detail. This choice is made not because parametric limits are less important in practice, but because they are more straightforward to understand and less fundamental than noise limits.

A. Parametric Limits

The maximum or minimum value of any one circuit parameter in a motion sensor sets limits on $\Delta t_{\min}$ or $\Delta t_{\max}$. For example, in digital systems $\Delta t_{\min}$ is the clock period and $\Delta t_{\max}$ may be the number of periods that can be counted by the largest counter in the system.[6] In many pulse-based analog systems, $\Delta t_{\min}$ is set by the smallest width, rise or fall times of particular pulses in the computation. Similarly, $\Delta t_{\max}$ is set by the largest width of particular pulses or by the total number of stages in a pulse-delay line. Parametric limits may be altered by changing the values of the parameters that determine them. In some cases, the technology imposes limits that the user cannot go beyond; for instance, it is hard to achieve rise and fall times shorter than a few nanoseconds in current CMOS technology.

B. Noise Limits

Several motion sensors, such as the Reichardt sensor described earlier, do not have any inherent parametric limits aside from the ones imposed by the implementation technology. If there was no noise in the system, $\Delta t_{\min}$ would be zero and $\Delta t_{\max}$ would be ∞ and the sensors could operate over an infinite velocity range. So how does noise limit their operation? We will carry out an intuitive analysis to answer this question for the Reichardt sensor.

Imagine that an infinitely narrow bar of contrast c moves a distance Δx from one location to another and we want to determine its velocity v. We choose an impulsive event, because we want the input to have as fine a resolution as possible so as to be limited by the resolution of the sensor rather than by any limit in the input. Analog motion sensors, like the Reichardt sensor, do not have clocks to measure time; rather, they have filters with state variables. Changes in the values of these state variables between input events encode the passage of time.

In the classic Reichardt sensor (Fig. 3), the arrival of the bar triggers a first-order, low-pass impulse response of the type $ce^{-t/\tau}$, in a filter with time constant τ. After a time equal to the time-of-travel $\Delta x/v$, the same event arrives at a neighboring location and triggers a short pulse that is

[6] In practice, in most digital systems, the time interval is fixed by the clock period and Δx is measured to estimate velocity. In this case, $\Delta x_{\max}$ and $\Delta x_{\min}$ limits should be used instead of $\Delta t_{\min}$ and $\Delta t_{\max}$ limits.

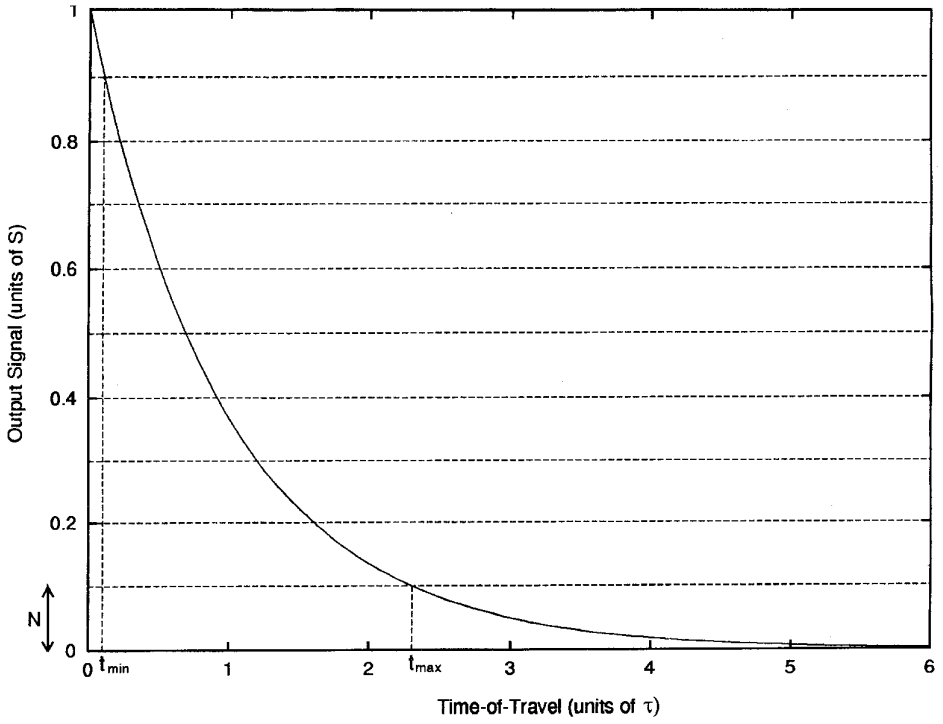


Fig. 4. *Dynamic range of a Reichardt motion sensor.* The dynamic range of any motion sensor is limited by the minimum and maximum time delays $\Delta t_{\min}$ and $\Delta t_{\max}$ it can process. We can define these time delays to correspond to output signal values one noise floor below the maximum and one noise floor above the minimum signal, respectively. The graph shows these limits for the impulse response $S \times \exp(-\Delta x/v\tau)$ of a classical Reichardt sensor as a function of the normalized time of travel of a bar between the two input receptors. They are a function of the sensor's SNR (aasumed to be 10 in this example).

multiplied with the filter output to produce the output value

$$V_R = S \times \exp\left(-\frac{\Delta x}{v\tau}\right). \tag{9}$$

The value S is a measure of the signal amplitude at the output and is proportional to the square of the contrast. The output of the sensor encodes velocity for a given stimulus because it is a function of the time-of-travel. For the following, we will assume that the output of the system is noisy, characterized by its root-mean-square (rms) noise N. In other words, output values that are within N of each other are not resolvable. In effect, the noise imposes a resolution grid with spacing N on the output signal, as illustrated in Fig. 4.

We define the minimum resolvable time interval $\Delta t_{\min}$ and the maximum resolvable time interval $\Delta t_{\max}$ to correspond to the outputs being within N of the value S for $v = \infty$, and within N of the value zero for $v = 0$, respectively. Though the system will respond to smaller times-of-travel, it is not possible to resolve the range of velocities between $\Delta x/\Delta t_{\min}$ and ∞, since they all map near S. Likewise, all velocities between zero and $\Delta x/\Delta t_{\max}$ cannot be resolved for the same reason. In order to compute these minimum and maximum times-of-travel, we need to invert

$$Se^{-\frac{\Delta t_{\min}}{\tau}} = S - N \tag{10}$$

and

$$Se^{-\frac{\Delta t_{\max}}{\tau}} = N. \tag{11}$$

If we define S/N to be the SNR,[7] it follows from (10) and (11) that the dynamic range of the Reichardt motion sensor is

$$D_R = \frac{\Delta t_{\max}}{\Delta t_{\min}} = \frac{\ln(S/N)}{\ln\left(\frac{S/N}{S/N-1}\right)}. \tag{12}$$

Using the Taylor series expansion of the natural logarithm, we observe that for large S/N values $D_R \approx (S/N)\ln(S/N)$.

The maximum level of noise tolerated by this sensor is given by $S/N = 2$, since in this case $\Delta t_{\min} = \Delta t_{\max} = \tau \ln 2$ and $D_R = 1$. Expressions (10)–(12) are valid only if $S/N \geq 2$. We will now show how D_R can be improved by introducing nonlinear adaptation.

VI. Nonlinear Adaptive Filtering

The previous analysis suggests the characteristic properties of a motion sensor with wide dynamic range: in order to reduce $\Delta t_{\min}$, we need a sharply decaying slope, i.e., a small time constant, at the beginning of the filter's

[7] In most of the literature, the SNR is defined as S^2/N^2 instead of S/N. We depart from convention because our choice avoids having too many square roots in mathematical expressions.

impulse response. In order to increase $\Delta t_{\max}$, the decay must slow down with time, i.e., the time constant must increase. Both of these requirements are met by an adaptive nonlinear filter whose time constant is small when its output voltage is large and whose time constant is large when its output voltage is small. In fact, there is a very simple way to build such a filter in electronics. All it takes is a parallel diode-and-capacitor circuit fed by an input current. This circuit is the nonlinear adaptive analog of a first-order filter built with a parallel RC circuit. Fig. 5(a) illustrates the difference between a simple exponential decay from a parallel RC circuit and the decay from an exponential-element-and-capacitor circuit. In general, any element with an exponential I-V characteristic in parallel with a capacitor exhibits similar behavior.

The dynamics of the latter circuit may be computed in closed form by solving a simple nonlinear differential equation. Here we state only the results. If the exponential element has an I-V characteristic given by

$$I = I_K e^{\frac{V}{V_K}} \tag{13}$$

where V_K and I_K are the voltage and current parameters that characterize the exponential element, respectively, then after the input current I_0 is switched off, the current in the exponential element $I_{\text{out}}(t)$ is given by

$$I_{\text{out}}(t) = \frac{I_0}{1 + \frac{I_0 t}{CV_K}} \tag{14}$$

where C is the capacitance. For

$$t \gg CV_K/I_0 \tag{15}$$

$I_{\text{out}}(t) = CV_K/t$ and is independent of the initial input current I_0. For large I_0, this happens very quickly. The independence of the dynamics on the initial condition is a very useful property for a motion sensor, since we would like measured velocities to be the same irrespective of the contrast of the input signal, especially for strong input signals.

The voltage on the exponential element is often more convenient to use and report than the current. It is given by

$$V_{\text{out}}(t) = V_K \ln\left(\frac{I_{\text{out}}(t)}{I_K}\right). \tag{16}$$

As discussed above, for large times the current has a $1/t$ dependence and consequently the voltage has a logarithmic dependence on time.

Fig. 5(b) shows measured voltage traces of a diode-and-capacitor element and theoretical plots according to (14) and (16) for various initial currents I_0. The convergence of the waveforms irrespective of initial conditions and the logarithmic behavior at large times may be easily seen.

If the rms noise voltage on the capacitor due to all noise sources is N, then by an analysis very similar to the one in Section 5.2, we can show that

$$\Delta t_{\min} = \frac{CV_K}{I_K}\left(e^{-\frac{(S-N)}{V_K}} - e^{-\frac{S}{V_K}}\right) \tag{17}$$

and

$$\Delta t_{\max} = \frac{CV_K}{I_K}\left(e^{-\frac{N}{V_K}} - e^{-\frac{S}{V_K}}\right). \tag{18}$$

Here S is the initial voltage computed from the initial current I_0 by

$$S = V_K \ln\left(\frac{I_0}{I_K}\right). \tag{19}$$

S is thus a measure of the signal output amplitude. The dynamic range $D_R = \Delta t_{\max}/\Delta t_{\min}$ is then given by

$$D_R = \frac{e^{\frac{(S-N)}{V_K}} - 1}{e^{\frac{N}{V_K}} - 1}. \tag{20}$$

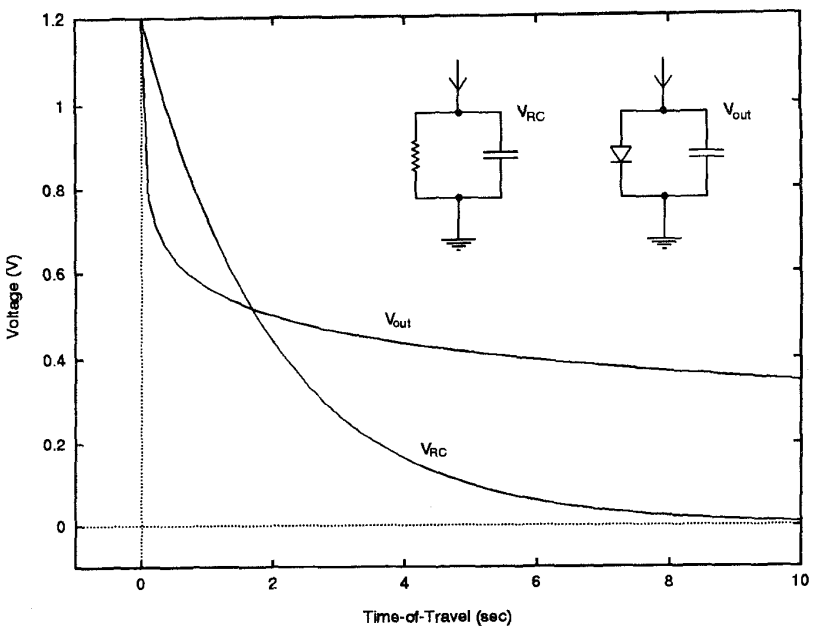
Note that for this filter, the dynamic range is determined by the SNR *and* by the value of N with respect to V_K. If N is small and S is large compared with V_K (i.e., S/N is large), then $D_R \approx (V_K/N) \times \exp((S/N - 1)N/V_K)$. If N is large compared with V_K and S/N is large, then $D_R \approx \exp((S/N - 2)N/V_K)$. In both cases, the exponential scaling with S/N is superior to the scaling of the Reichardt sensor of $(S/N)\ln(S/N)$. As before, when the noise level reaches half of the signal strength, $D_R = 1$ and the sensor becomes inoperable.

VII. The FS Velocity Sensor

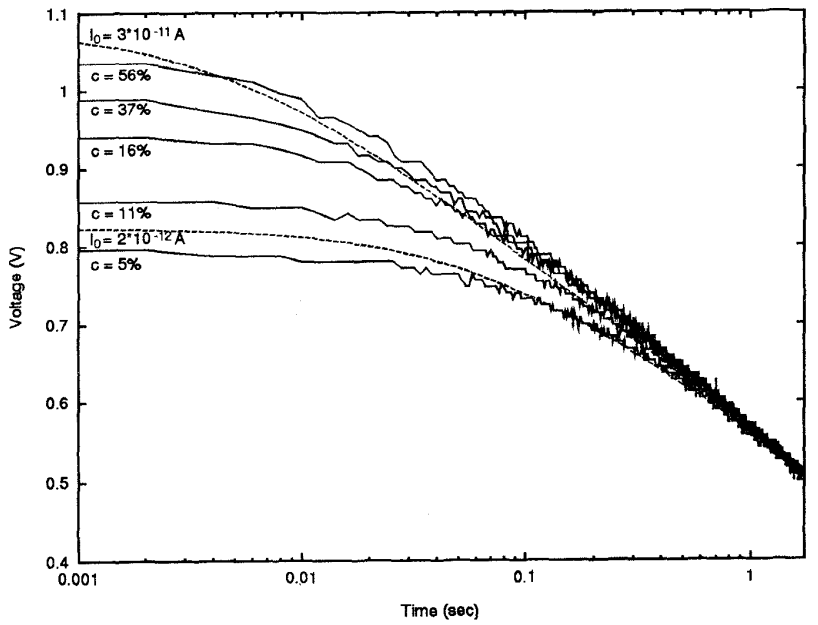
The facilitate-and-sample (FS) velocity sensor [10] uses the diode-capacitor dynamics of the nonlinear filter described in the previous section to obtain a wide dynamic range. Unlike the Reichardt algorithm, the FS algorithm is token-based. In a first stage, the circuit extracts edges from the image brightness distribution. These can be either temporal or spatial edges, depending on the implementation. The sensor then estimates the time it takes for an edge to travel between two adjacent photoreceptor locations. This time is not computed by correlating edge responses obtained from the two locations; rather, the nonlinear-filter response caused by an edge at one location is sampled by a narrow pulse caused by the arrival of the edge at the adjacent location. This sampling scheme results in the sensor output being independent of contrast for large times-of-travel [see Fig. 5(b)], whereas a correlation scheme would result in a dependence on contrast for all times-of-travel, as in the Reichardt sensor.

The FS sensor is schematically shown in Fig. 6(a). It consists of two temporal-edge detectors (E), two pulse-shaping circuits (P), two motion circuits (M), and a direction-selection circuit (D). Since the FS sensor has been previously described in detail [9], [10], we give here only a brief description of the different elements and some performance data.

Each temporal-edge detector (E) comprises an adaptive photoreceptor circuit converting irradiance transients into voltage transients [11] and a circuit that converts the positive voltage transients, corresponding to ON edges, into current impulses. Each current impulse generates two voltage pulses, which we call the fast pulse and the slow



(a)



(b)

Fig. 5. *Properties of a nonlinear diode-capacitor filter.* (a) Comparison of the calculated impulse responses of a parallel RC circuit as implemented by the low-pass filter of the Reichardt sensor (see Fig. 4) and a nonlinear adaptive filter built with an exponential element and a capacitor in parallel. The response of the nonlinear filter has a much steeper slope initially and thus a higher sensitivity than that of the RC filter; subsequently, its time constant increases so that its decay becomes eventually much slower than that of the RC filter. The use of such a nonlinear filter in a motion sensor results in a much larger dynamic range than the use of a RC filter, since the sharp slope near the origin reduces $\Delta t_{\min}$, while the slow decay after long delays increases $\Delta t_{\max}$ for a given noise level. For the RC filter a time constant $\tau = 2$ s and a maximum signal $S = 1.2$ V was assumed. The nonlinear filter curve is derived from (14) and (16) with $I_K = 3.5 \times 10^{-16}$ A, $V_K = 95$ mV, $I_0 = 9^{-11}$ A, and $C = 1.4 \times 10^{-12}$ F. (b) Measured dynamics of such a nonlinear filter for different initial current conditions. The circuit was activated by current pulses from a temporal-edge detector, such as the one used in the FS motion sensor in response to moving edges of different contrasts c. The measured data (solid lines) is compared with theoretical curves (dashed lines) derived from (14) and (16). Irrespective of the initial condition, the responses eventually converge to the same logarithmic time response. The larger the intial current I_0 the more rapid is the convergence. If I_0 encodes the contrast in the image, as for the measured data, the behavior implies that at sufficiently high contrasts and for sufficiently low velocities, the motion sensor's output is invariant with contrast.

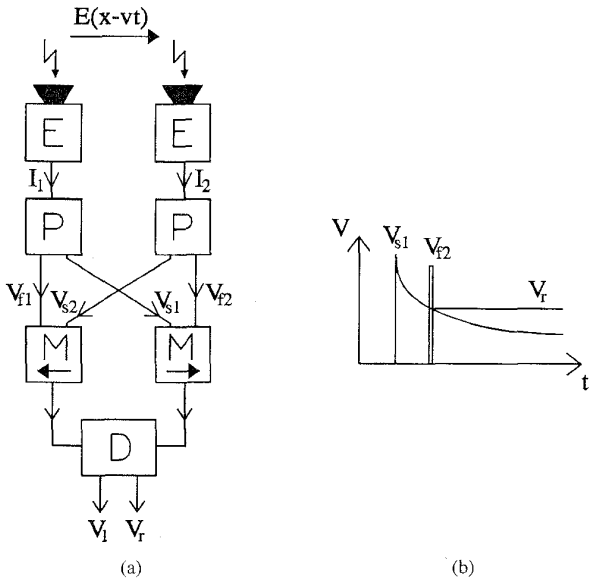


Fig. 6. *The FS velocity sensor.* (a) *Block diagram.* Temporal-edge detectors (E) generate current impulses in response to fast image irradiance (i.e., brightness) transients. Pulse-shaping circuits (P) convert the current impulses into voltage pulses. Voltage pulses from adjacent pixels are fed into two motion circuits (M) computing velocity for opposite directions (V_l and V_r) along one dimension. A direction-selection circuit (D) suppresses the response in the null direction to prevent temporal aliasing. (b) *Voltage signals.* The analog output voltage of the motion circuit for rightward motion (V_r) equals the voltage of the slowly decaying facilitation pulse (V_{s1}) at the time of arrival of the narrow and fast sampling pulse (V_{f2}). For leftward motion, the fast sampling pulse precedes the slow facilitation pulse and the output voltage is low. The analog output voltage thus encodes velocity for rightward motion only.

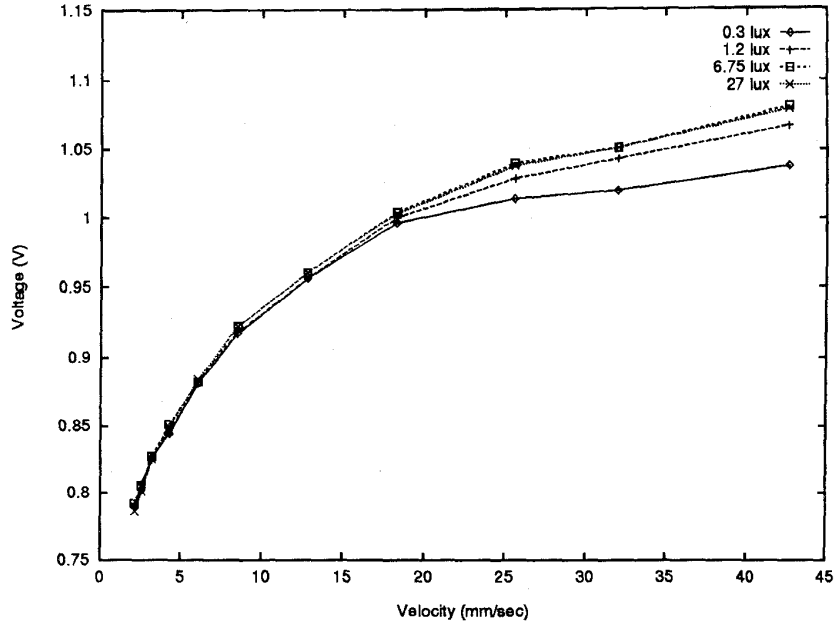
pulse respectively, in the pulse-shaping circuit (P) that is coupled to E. The fast pulse is a narrow spike that is a sharpened version of the input current impulse. The slow pulse is the output of a parallel diode-capacitor circuit in response to the current impulse, as described by (14) and (16). Each motion circuit (M) uses the fast pulse generated by an edge at one location to sample and report the voltage of the slow pulse output at the other location. If for a given edge the onset of the slow pulse precedes the fast pulse, the edge is said to move in the motion circuit's preferred direction, otherwise it is said to move in its null direction. Note that the sampled voltage in the null direction contains no information on the speed of the edge triggering the sampling signal, since it is related to the time of arrival of the previous edge. The motion circuit on the right receives the slow pulse V_{s1} from the left edge detector and the fast pulse V_{f2} from the right edge detector. Thus its preferred direction is from left to right [Fig. 6(b)]. Conversely, the motion circuit on the left receives the slow pulse V_{s2} from the right edge detector and the fast pulse V_{f1} from the left edge detector. Thus its preferred direction is from right to left. If the optical stimulus is not spatially aliased, the motion circuit for which the edge moves in the preferred direction reports the higher voltage than the one for which it moves in the null direction. The direction-selection circuit (D) makes use of this fact to set the output voltage for the null direction to zero in order to prevent temporal aliasing.

As we have seen in Section VI, the sampled value of the slow pulse in the preferred direction is a monotonic function of the time-of-travel and, if condition (15) is met, independent of the height of the input current impulse, i.e., of edge contrast. Also, since the photoreceptor's transient output voltage is quite invariant with illuminance, so is the motion sensor's output. This means that the output voltage of the motion circuit for which the edge moves in the preferred direction then also encodes speed. Fig. 7 illustrates the performance of the FS sensor for different illuminances and contrasts.

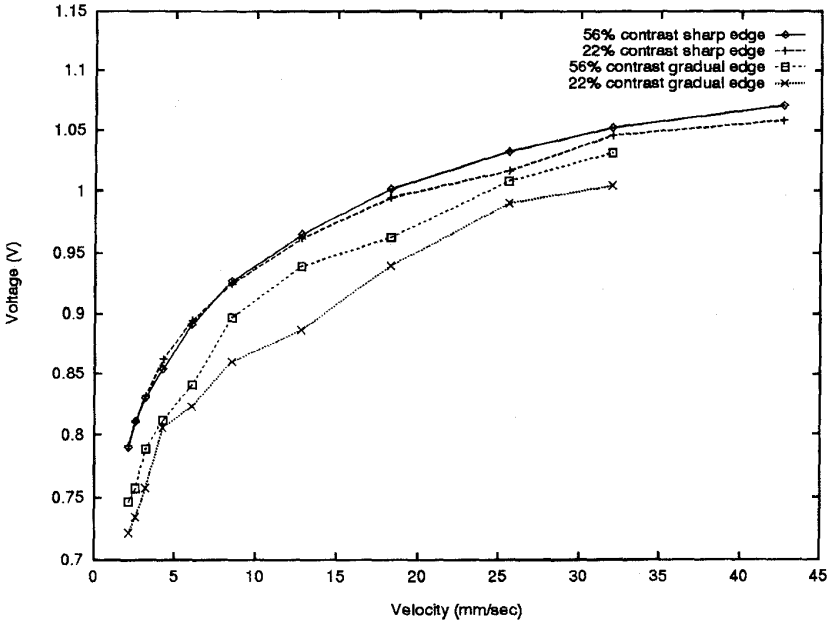
In Fig. 8 velocity measurements over a more extended range are plotted. The response to a 56% contrast edge is compared with the response to electronically generated current pulses fed into the pulse-shaping circuits to simulate an ideal moving edge. The limits of operation of the FS sensor are determined by a combination of parametric and noise considerations that have to do with the details of the photoreceptor, edge-detection, and motion circuits. For stimulus-independent speed-encoding, $\Delta t_{\min}$ is limited by condition (15), whereas direction-selectivity is lost when the width of the sampling pulse (here 5.3 μs) exceeds the time-of-travel of an edge between two adjacent photoreceptors and a high signal is sampled in the null direction. We were unable to measure the $\Delta t_{\max}$ limit for the optical input stimulus, because it is so large that it exceeds the capabilities of our experimental apparatus. The data for the electronic input shows that the motion-computing stage has low intrinsic noise and a dynamic range of at least seven orders of magnitude, and that $\Delta t_{\max}$ is ultimately limited by leakage currents, rather than by noise. The optical input stage, however, reduces the dynamic range of the sensor, because of the significantly higher noise from the photoreceptor and edge-detection circuitry and because of the larger width of the sampling pulse, causing spurious responses in the null direction at high speeds. Yet we still obtain at least three orders of magnitude of dynamic range for high-contrast optical edges. As we have seen (Fig. 7), more severe dynamic range limitations, especially at high velocities, arise from the nonuniformity of the input signals, as derived from edges with different sharpnesses and contrasts, under a variety of lighting conditions. If progress were made on a good adaptive algorithm for finding edges or other features, the dynamic range of analog VLSI motion sensors could be greatly improved, because the effective "noise" at the input would then be reduced.

A. Scaling Considerations

An elementary cell of a 1-D array of FS sensors consists of an edge detector, a pulse-shaping circuit, two motion circuits, and a direction-selection circuit. It comprises 41 transistors and eight capacitors, totaling 7.5 pF capacitance. FS sensors have been fabricated using a 2 μm n-well CMOS process provided by the MOSIS prototyping service, where the total area consumption of an elementary cell is about 0.05 mm^2. The smallest chips ("tiny" chips) available



(a)



(b)

Fig. 7. *Response of the FS velocity sensor for different illuminations and contrasts.* The output voltage for motion in the preferred direction is shown as a function of image velocity. The stimuli were printed on paper wrapped around a rotating drum. The pixel spacing of 300 μm corresponds to an optical angle of 1.28°. The image velocity was computed in the same way as described in Fig. 1(b). Each data point represents the average of five successive measurements under ac incandescent illumination. (a) Response to a 56% contrast sharp edge for different illumination levels. (b) Response to sharp and gradual edges of different contrasts. Contrast is defined here as the ratio of the difference to the sum of the image irradiances on either side of the edge.

through the MOSIS service with a usable area of about 2.5 mm^2 (leaving room for read-out circuitry and pads) can accommodate 50 pixels. The largest MOSIS chip would be able to contain 1250 pixels on 62.5 mm^2.

For an arrangement of elementary motion cells operating in two dimensions, the edge detectors and the pulse-shaping circuits may be shared among all directions, but each cell would need four motion circuits and two direction-

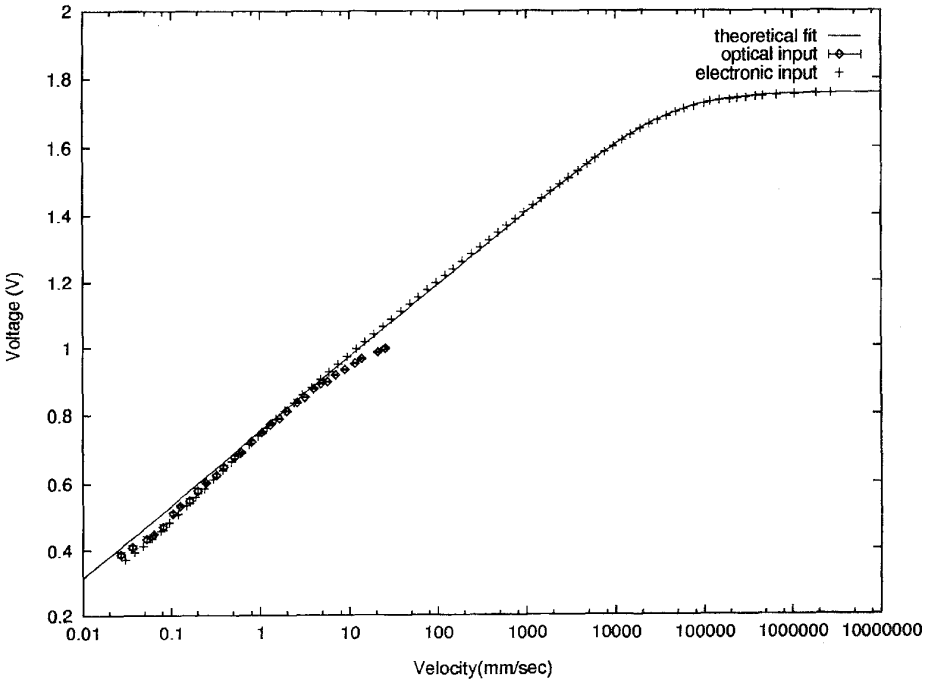


Fig. 8. *Response of the FS velocity sensor for optical and electronic stimuli.* The output voltage for motion in the preferred direction, shown as a function of image velocity for a 56% sharp edge, is compared to that measured for stimulation of the motion-computing circuitry with 50 μs long current impulses to simulate the outputs of adjacent edge detectors to idealized edges. A theoretical fit to the latter data, based on (14) and (16), is also shown. The rms error of the data points for optical input, as indicated by error bars, averages to about 3 mV, whereas for electronic stimulation each data point has an rms error of less than 1 mV. The FS motion sensor shows excellent performance over more than seven orders of magnitude of velocity for such standardized electronic input signals due to its adaptive time constant and low intrinsic noise. For optical stimulation, our measured data is limited to three orders of magnitude for the following reasons: 1) at high speeds, the finite width of the sampling pulse causes spurious responses in the null direction, 2) at high speeds, our rotating test drum runs unevenly and wobbles, and 3) we are unable to run the test drum at very low speeds.

selection circuits. This would increase the cell size by about 35%.

If the circuits were fabricated in a state-of-the-art 0.7 μm process, the cells could probably be reduced to about a quarter of their current size, so that a 128×128 array would fit onto 16 mm $\times$ 16 mm chip area.

VIII. System Applications

As discussed in the previous sections, we designed analog VLSI motion sensors whose outputs are nearly independent of absolute light level and stimulus contrast, if these are sufficiently high, and that are able to achieve satisfactory performance over a large dynamic range in velocity. Because they integrate both the photosensing as well as the motion-computation stage on a single chip, they represent a small, power-lean, and low-cost alternative to conventional computer-vision systems.

Two significant problems need to be tackled next. First, so far we have made no attempt to integrate the optical flow over space. In particular, we have not designed any circuitry to solve the aperture problem at the level of our motion sensors. Second, such sensors lack a high degree of precision in their output values because of temporal noise (see Section II), spatial fixed-pattern noise in motion sensing arrays due to offsets and because the performance of the FS sensor degrades at low contrasts and irradiance levels. The rms voltage noise of an FS velocity-sensing element averages to about 3 mV over the velocity range shown in Fig. 8, that spans three orders of magnitude and maps onto a voltage range of 0.6 V. This corresponds to 7 or 8 b of resolution for such a stimulus.

To fully exploit our motion sensors at a system level, we need to develop parallel image-processing architectures for applications that rely mainly on *qualitative* properties of the optical flow, rather than on the precise value of each flow vector. Given the noise inherent in cortical neurons, it might be argued that optical flow computations carried out in the central nervous system can also only exploit such properties. Furthermore, analog VLSI devices are most efficient at solving well-defined custom problems, unlike their digital counterparts that are programmable and therefore more versatile. Thus we need to choose application domains where analog VLSI solutions are attractive compared to standard, digital machine vision systems. As an example, we will focus on an application domain that is becoming important in both research and industrial communities—vehicle navigation. Specifically, we will show how parallel architectures based on analog

VLSI motion sensors can be used to estimate integrative features of optical flow fields extracted from typical vehicle navigation scenes [3]. The two features that we shall discuss are *heading direction* and *time-to-contact*. Among other features that can be estimated are motion discontinuities, to help segregate objects (e.g., cars) from the background [37]–[39].

In order to analyze the computational properties of the optical flow and to determine the most suitable architectures for analog VLSI implementations, we performed software simulations on sequences of images obtained from a commercially available camera a 64×64 pixel silicon retina placed on a moving van. The camera uses adaptive CMOS photoreceptors similar in overall design to the ones incorporated into our motion sensors. Fig. 9 shows an example of such an image and the associated optical flow field.

In our simulations, the optical flow fields were generated by implementing a gradient algorithm based on the *image-brightness-constancy* equation (see Section III-B for the 1-D formulation). If $E(x, y, t)$ represents the image irradiance at time t and position (x, y), the brightness-constancy equation expresses the constraint that

$$\frac{d}{dt}E(x, y, t) = 0 \tag{21}$$

in subregions of the image. By explicitly computing the temporal and spatial derivatives, (21) can be rewritten as

$$\nabla E \cdot \mathbf{v} + \frac{\partial}{\partial t}E = 0 \tag{22}$$

and the velocity vector $\mathbf{v}$ can be computed as

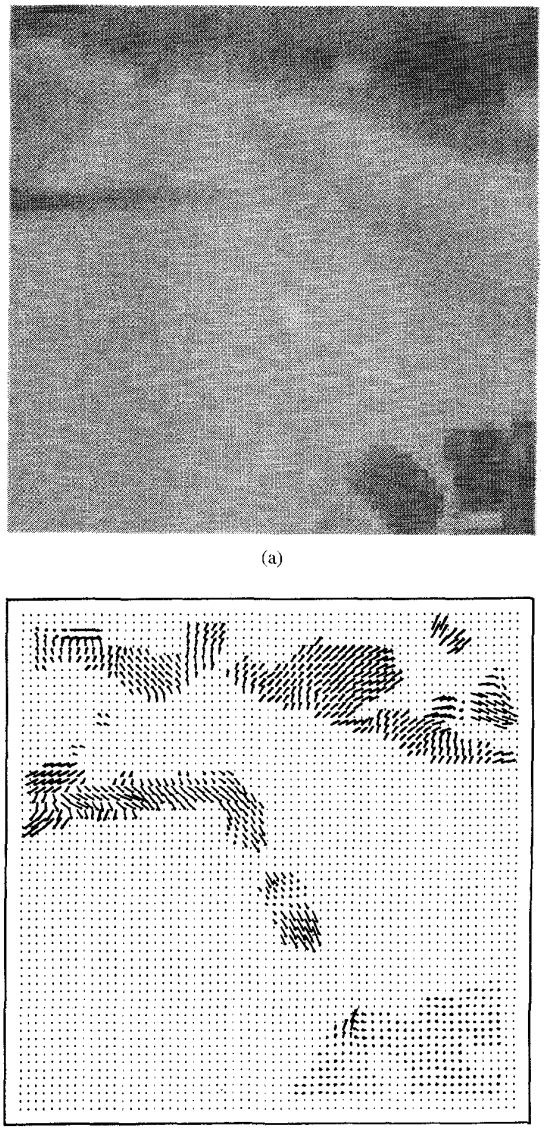
$$\mathbf{v} = -(\nabla E^t \cdot \nabla E)^{-1} \nabla E^t \frac{\partial}{\partial t}E \tag{23}$$

wherever the determinant of the square matrix $(\nabla E^t \cdot \nabla E)$ exceeds a certain threshold [7], [8], [24]. At locations where it remains below the threshold, the optical flow is not estimated. Thus the data is sparse and only reliable for areas in the image with high contrast. Having defined the types of images to be analyzed and the types of processing to be performed, we can use *a priori* information to selectively integrate such sparse data to estimate the targeted features of the optical flow field.

A. Heading Direction

We are primarily interested in measuring the heading direction of a vehicle along the horizontal axis, since this is the axis along which the vehicle may be controlled. Hence, we will consider arrays of 1-D velocity sensors to reduce the computational complexity of the problem. A further simplification can be made if we restrict the task to purely translational ego motion. Because cars and other vehicles will soon contain integrated sensors for determining the rate of rotation around the three body axes, we assume here that we can use this extra-retinal information to remove the components of the optical flow due to rotation.

In the case of pure translational motion, the heading direction coincides with the *focus-of-expansion* (FOE) and is defined as the spatial location on the image plane from which all the velocity vectors diverge. If we only examine the horizontal component of the optical flow, the FOE corresponds to the point at which the vectors change direction. If we assign positive values to those pointing in one direction and negative values to vectors pointing in the opposite direction, the FOE location will correspond to the point closest to the *zero crossing* in the array. Unfortunately, even when there is no rotational motion involved, real image sequences may give rise to optical flow fields that contain more than one zero-crossing. Nevertheless, by using *a priori* information on the spatial



(a)

(b)

Fig. 9. *Optical flow associated with a highway scene.* (a) Image of a straight stretch of highway obtained with a 64×64 silicon retina mounted on a moving van (Courtesy of B. Mathur of Rockwell International). The photoreceptors used in this camera are similar to the ones used in our velocity sensors. (b) Optical flow field calculated from the image in (a) and the succeeding image in the sequence using a standard machine-vision algorithm discussed in (23).

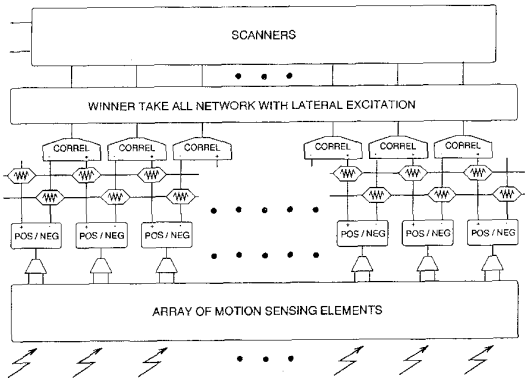


Fig. 10. *Block diagram of the 1-D analog VLSI FOE architecture.* The outputs of an array of FS sensors corresponding to opposite directions of motion are fed into differencing transconductance amplifiers. The positive and negative half currents of the amplifiers are spatially smoothed with separate diffusor networks. The outputs of the diffusor networks provide the input to a correlation network that detects the adjacency of large currents corresponding to opposite directions of a motion. The correlation network signals the presence and strength of zero-crossing in the optical flow field. A winner-takes-all network with positive feedback and lateral spreading selects the FOE that corresponds to the steepest zero-crossing, and that is near the previously selected FOE.

and temporal characteristics of the optical flow image (e.g., by making the assumption that the direction of heading of a vehicle changes smoothly), it is possible to detect the correct position of the FOE.

An analog VLSI architecture designed to extract the FOE from a 1-D image is shown in Fig. 10 [38], [40]. The two directional outputs of an array of FS motion sensors are differenced with transconductance amplifiers [41]. The positive and negative half currents from the amplifiers are separately smoothed with diffusor networks [42] to reduce noise and offsets from the sensor data. An abrupt change in the direction of the optical flow vectors is reported by bump-like correlator circuits [43]. These circuits detect the adjacency of large opponent currents (positive half current versus negative half current) belonging to neighboring pixels, i.e., they perform zero-crossing detection, and yield information on the slope of the zero-crossing as well. A winner-take-all network with distributed excitation [44] chooses the position of the steepest zero-crossing as the location of the FOE. The winner-take-all network has built-in lateral spreading and positive feedback such that the position of the FOE that is chosen is always near that of the previous choice; consequently, the position of the FOE shifts smoothly in time as one would expect from a real-world situation. The scanners report the spatial response of the winner-take-all network outputs off chip, so that the position of the FOE may be shown on a display.

Fig. 11 shows an example of the FOE computation for the data of Fig. 9 obtained from software simulations that model the architecture shown in Fig. 10.

B. Time-to-Contact

The time-to-contact τ represents the duration before a car collides with an obstacle or another moving object, assuming that the relative velocity remains constant. It

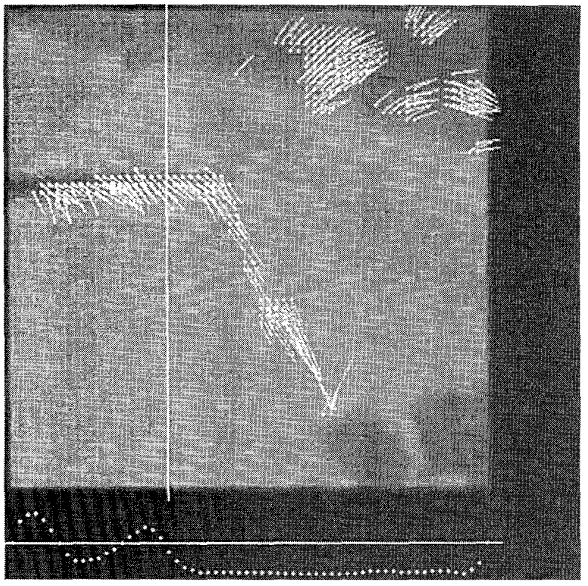


Fig. 11. *FOE of the highway scene.* Determination of the FOE of the highway sequence from the optical flow data of Fig. 9 using a simulation of the VLSI architecture of Fig. 10. The horizontal components of the extracted optical flow vectors, summed along the image colums over each frame, are displayed at the bottom of the figure. The horizontal coordinate of the FOE is computed as the abscissa of the zero-crossing with maximum steepness and closest to the abcissa of the previously selected FOE. The vertical line indicates the computed position of the FOE.

therefore represents a very useful parameter for correcting speed and direction of a vehicle, or for initiating airbag deployment. It has been shown [45] that the time-to-contact can be directly computed from the optical flow without knowledge of the FOE for a general motion of the car with respect to the object.

Imagine that a car with a mounted camera is moving with a speed v toward a wall at a distance Z_0. Under these conditions

$$\tau = \frac{Z_0}{v}. \tag{24}$$

The velocity field $\mathbf{V}$ in the camera's image then expands uniformly away from the FOE and the velocity vectors increase in magnitude proportionally with the distance from the FOE. Using the 2-D version of Gauss's divergence theorem, it has been shown [46] that τ can be estimated in a very robust manner for an arbitrary translatory motion toward a plane at distance Z_0 by measuring the normal velocity component around some closed contour. If we consider as closed contour a circle C of radius r_0 and assume a linear velocity field $\mathbf{V}$ so that $\nabla \cdot \mathbf{V}$ is constant everywhere, independently of the exact location of the circle, then

$$\int_C \mathbf{V} \cdot \mathbf{n} ds = 2\pi r_0^2 \frac{v}{Z_0} = \frac{2\pi r_0^2}{\tau} \tag{25}$$

where $\mathbf{n}$ denotes the unit normal vector along the contour. This means that a simple integration of the output of a number of 1-D motion sensors arranged along a circular contour that measure the radial velocity component di-

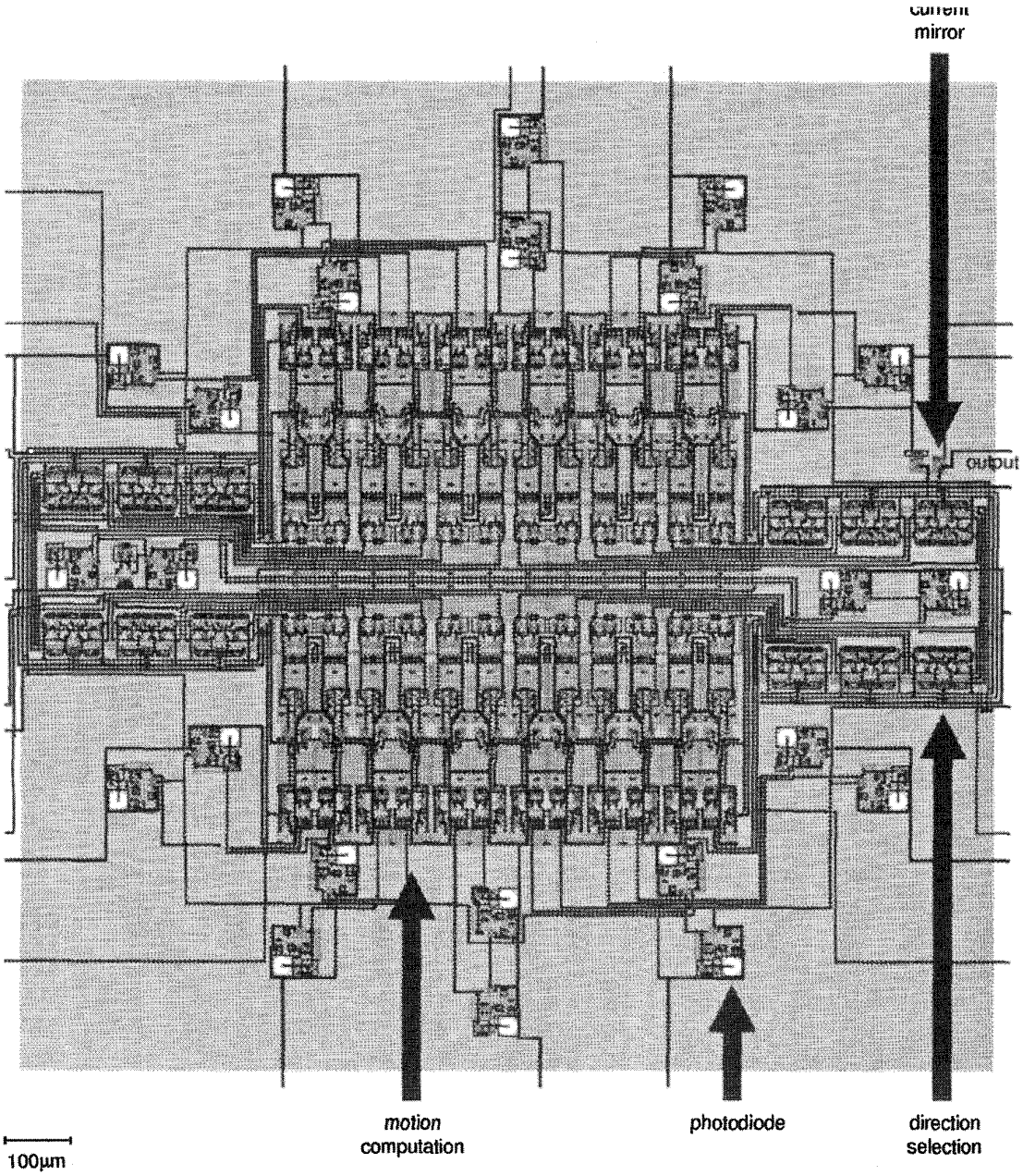


Fig. 12. *The time-to-contact chip.* Layout of an analog VLSI chip for the determination of time-to-contact. The 12 pairs of photoreceptors arranged on a circle are coupled to 12 FS motion sensors that estimate the radial components of the optical flow field. The pulse-shaping circuits and motion circuits are located in the central part of the chip and the direction-selection circuits are located on the left and right sides. Their output currents are summed for each direction (outward and inward) and subtracted (using a current minor) to yield the final output that is inversely proportional to the time-to-contact. The size of the layout is 1.6 mm × 1.6 mm as implemented with a 2 μm CMOS process.

rectly yields the time-to-contact, *without* computing any spatial or temporal derivatives. Furthermore, because we are integrating over the output of a number of elementary motion sensors, each individual measurement can be affected by a certain amount of noise without substantially deteriorating the overall result. In the more general case of an observer approaching a slanted surface (leading to a quadratic velocity field), as long as the observer is heading directly for this slanted surface, one can exploit the symmetry of the velocity field, arriving at the same result as above. Similar integration schemes can be exploited for designing circuits sensitive to rotations [46]. This type of computation therefore lends itself remarkably well to VLSI implementation.

We designed a parallel architecture in which 12 FS velocity sensors, arranged on a circle, estimate the radial

component of the velocity field (Fig. 12). For such an arrangement, the time-to-contact can be approximated by

$$\tau = \frac{N \cdot r}{\sum_{k=1}^{N} v_k} \tag{26}$$

where N denotes the number of velocity-measuring elements on the circle, r the radius of the circle, and v_k the radial velocity components at the locations of the elements. The output voltages of the FS sensors are converted into currents that linearly encode velocity. The currents are summed together, so that the total output current is inversely proportional to the time-to-contact. The algorithm is robust in that errors due to device mismatch, noise, and erroneous velocity measurements are decreased through averaging. Using a stimulus of high-contrast concentric rings centered on the focus-of-expansion or just simulating an approaching motion with a rotating spiral we obtained reasonably accurate estimates of the time-to-contact [40]. As yet, we have not accounted for the aperture problem, i.e., we are not able to extract the radial velocity component of a randomly oriented edge, but due to the integrative properties of the sensor we expect to be able to find reasonable estimates for the time-to-contact in more general scenes and to reliably determine the direction of motion (i.e., expansion or contraction) in them.

IX. Conclusion

We have shown in this article how analog VLSI velocity sensors with on-chip photoreceptors can be built and how they can be used in massively parallel architectures to evaluate integrative features of the optical flow. Given the robust, very compact, and low-power nature of such smart vision sensors, they represent a true alternative—for certain applications—to more conventional digital image architectures [1].

More specifically, we have emphasized the following points: photoreceptor sizing is important because of the tradeoffs between spatial filtering and the reduction of noise. The limits of operation of motion sensors are determined by the maximum and minimum values of parameters in the system and/or by noise. One can carry out quantitative calculations of the dynamic range of motion sensors as a function of the SNR of the system. These calculations show why it is advantageous to use nonlinear adaptive filters such as those built with a diode and a capacitor to obtain a wide dynamic range, as evidenced by the data from our FS sensor. The determination of features like focus-of-expansion and time-to-contact that can be estimated from integrative properties of the optical flow are promising targets for parallel analog VLSI implementations. This is because such features are particularly important for mobile systems, where processing time, size, and power consumption are often important issues that favor analog VLSI implementations over traditional digital ones.

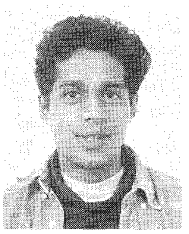
Acknowledgment

The frequency response data shown in Fig. 1(a) were provided by T. Delbrück. The image sequences used in Section VIII were recorded with a camera developed by Rockwell International Corporation and were provided by B. Mathur. Fabrication of the integrated circuits was provided by MOSIS.

References

[1] C. Koch and H. Li, Eds., *Vision Chips: Implementing Vision Algorithms with Analog VLSI Circuits.* Los Alamitos, CA: IEEE Comput. Soc., 1994.
[2] E. R. Fossum, "Architectures for focal plane image processing," *Opt. Eng.*, vol. 28, pp. 865–871, Aug. 1989.
[3] A. Verri and T. Poggio, "Motion field and optical flow: Qualitative properties," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 11, pp. 490–498, 1989.
[4] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, Sept. 1985.
[5] B. K. P. Horn, *Robot Vision.* Cambridge, MA: MIT Press, 1986.
[6] T. Poggio, W. Yang, and V. Torre, "Optical flow: Computational properties and networks, biological and analog," in *The Computing Neuron*, R. Durbin, C. Miall, and G. Mitchison, Eds. Wokingham, U.K.: Addison-Wesley, pp. 355–370, 1989.
[7] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comp. Vision*, vol. 12, pp. 43–77, 1994.
[8] A. Verri, M. Straforini, and V. Torre, "Computational aspects of motion perception in natural and artificial vision systems," *Phil. Trans. R. Soc. Lond. B*, vol. 337, pp. 429–443, 1992.
[9] J. Kramer, R. Sarpeshkar, and C. Koch, "An analog VLSI velocity sensor," in *Proc. 1995 IEEE Int. Symp. on Circ. and Syst.*, Seattle, WA, May 1995, pp. 413–416.
[10] ——, "Pulse-based analog VLSI velocity sensors," to be published in *IEEE Trans. Circ. and Syst. II.*
[11] T. Delbrück and C. A. Mead, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits," CNS Memo No. 30, Calif. Inst. Technol., June 1995.
[12] R. Sarpeshkar, T. Delbrück, and C. A. Mead, "White noise in MOS transistors and resistors," *IEEE Circ. and Devices*, vol. 9, pp. 23–29, 1993.
[13] E. C. Hildreth and C. Koch, "The analysis of visual motion: From computational theory to neuronal mechanisms," *Ann. Rev. Neurosci.*, vol. 10, pp. 477–533, 1987.
[14] S. Ullman, *The Interpretation of Visual Motion.* Cambridge, MA: MIT Press, 1979.
[15] ——, "Analysis of visual motion by biological and computer systems," *IEEE Comput.*, vol. 14, pp. 57–69, 1981.
[16] B. Hassenstein and W. Reichardt, "Systemtheoretische analyze der zeit-, reihenfolgen- und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers chlorophanus," *Z. Naturforsch*, vol. 11b, pp. 513–524, 1956.
[17] E. H. Adelson and J. R. Bergen, "Spatio-temporal energy models for the perception of motion," *J. Opt. Soc. Amer. A*, vol. 2, pp. 284–299, 1985.
[18] A. B. Watson and A. J. Ahumada, "Model of human visual-motion sensing," *J. Opt. Soc. Amer. A*, vol. 2, pp. 322–341, 1985.
[19] J. P. H. van Santen and G. Sperling, "A temporal covariance model of motion perception," *J. Opt. Soc. Amer. A*, vol. 1, pp. 451–473, 1984.
[20] H. H. Nagel, "Analysis techniques for image sequences," in *Proc. 4th Int. Joint Conf. Patt. Recog.*, Kyoto, Japan, Nov. 1978.
[21] C. Fennema and W. Thompson, "Velocity determination in scenes containing several moving objects," *Comp. Graph. Image Process*, vol. 9, pp. 301–315, 1979.
[22] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185–204, 1981.
[23] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," *Biol. Cybern.*, vol. 60, pp. 79–87, 1988.
[24] A. Verri, F. Girosi, and V. Torre, "Differential techniques for optical flow," *J. Opt. Soc. Amer. A*, vol. 7, pp. 912–922, 1990.
[25] J. Tanner and C. Mead, "An integrated analog optical motion sensor," in *VLSI Signal Processing, II*, S. Y. Kung, Ed. New York: IEEE Press, 1986, pp. 59–76.

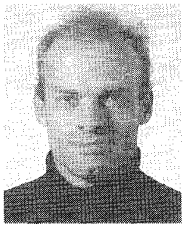
[26] R. F. Lyon and M. P. Haeberli, "Designing and testing the optical mouse," *VLSI Design*, vol. 3, pp. 20–30, 1982.
[27] T. Horiuchi *et al.*, "Computing motion using analog VLSI chips: An experimental comparison among different approaches," *Int. J. Comp. Vision*, vol. 8, pp. 203–216, 1992.
[28] A. G. Andreou, K. Strohbehn, and R. E. Jenkins, "Silicon retina for motion computation," in *Proc. 1991 IEEE Int. Symp. Circ. and Syst.*, Singapore, June 1991, pp. 1373–1376.
[29] T. Horiuchi, J. Lazzaro, A. Moore, and C. Koch, "A delay line based motion detection chip," in *Advances in Neural Information Processing Systems 3*, R. Lippman, J. Moody, and D. Touretzky, Eds. San Mateo, CA: Morgan Kaufman, 1991, pp. 406–412.
[30] T. Delbrück, "Silicon retina with correlation-based, velocity-tuned pixels," *IEEE Trans. Neur. Networks*, vol. 4, pp. 529–541, 1993.
[31] R. Sarpeshkar, W. Bair, and C. Koch, "Visual motion computation in analog VLSI using pulses," *Advances in Neural Information Processing Systems 5*, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufman, 1993, pp. 781–788.
[32] W. Bair and C. Koch, "An analog VLSI chip for finding edges from zero-crossings," *Advances in Neural Information Processing Systems 3*, R. Lippman, J. Moody, and D. Touretzky, Eds. San Mateo, CA: Morgan Kaufman, 1991, pp. 399–405.
[33] H. B. Barlow and W. R. Levick, "The mechanism of directionally selective units in the rabbit's retina," *J. Physiol.*, vol. 178, pp. 447–504, 1965.
[34] R. G. Benson and T. Delbrück, "Direction selective silicon retina that uses null inhibition," *Advances in Neural Information Processing Systems 4*, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufman, pp. 756–763, 1992.
[35] R. Etienne–Cummings *et al.*, "A new temporal domain optical flow measurement technique for focal plane VLSI implementation," in *Proc. Comp. Arch. Machine Perception 1993*, M. Bayoumi, L. Davis, and K. Valavanis, Eds., pp. 241–250, 1993.
[36] J. Kramer, "Compact integrated motion sensor with three-pixel interaction," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 18, pp. 455–460, 1996.
[37] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive networks," *IEEE Computer*, vol. 21, pp. 52–64, 1988.
[38] G. Indiveri, J. Kramer, and C. Koch, "System implementations of analog VLSI velocity sensors," in *Proc. 5th Int. Conf. Microelectronics for Neural Networks and Fuzzy Systems (MicroNeuro'96)*, Lausanne, Switzerland, Feb. 1996.
[39] J. Kramer, R. Sarpeshkar, and C. Koch, "Analog VLSI motion discontinuity detectors for image segmentation," submitted for publication to *Proc. 1996 IEEE Int. Symp. on Circ. and Syst.*, Atlanta, GA, May 1996.
[40] G. Indiveri, J. Kramer, and C. Koch, "Parallel analog VLSI architectures for computation of heading direction and time-to-contact," in *Advances in Neural Information Processing Systems 8*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996.
[41] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
[42] K. A. Boahen and A. G. Andreou, "A contrast sensitive silicon retina with reciprocal synapses," in *Advances in Neural Information Processing Systems 4*, J. E. Moody, S. J. Hanson, and R. P. Lippmann, Eds. San Mateo, CA: Morgan Kaufman, 1992, pp. 764–772.
[43] T. Delbrück, "'Bump circuits' for computing similarity and dissimilarity of analog voltages," in *Proc. Int. Joint Conf. Neur. Networks*, Seattle, WA, July 1991, pp. 475–479.
[44] T. G. Morris, D. M. Wilson, and S. P. DeWeerth, "Analog VLSI circuits for manufacturing inspection," in *Proc. 1995 Conf. Advanced Res. in VLSI*, Chapel Hill, NC, Mar. 1995.
[45] M. Tistarelli and G. Sandini, "On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 15, pp. 401–410, Apr. 1993.
[46] T. Poggio, A. Verri, and V. Torre, "Green theorems and qualitative properties of the optical flow," A.I. Memo No. 1289, MIT, 1991.
[47] J. Luo, C. Koch, and B. Mathur, "Figure-ground segregation using an analog VLSI chip," *IEEE Micro*, vol. 12, pp. 46–57, 1992.



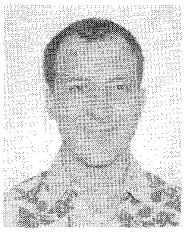
Rahul Sarpeshkar was born in Bangalore, India, in 1968. He received the B.S. degrees in electrical engineering and physics from the Massachusetts Institute of Technology, Cambridge, MA, in 1990. Since then he has been a Ph.D. student at the Carver Mead Analog VLSI Lab, California Institute of Technology, Pasadena, CA.

He has been working on noise devices, circuits and systems, adaptive silicon cochleas, and in visual motion detection. His research interests are in analog circuit design for sensory systems and in the fundamental limits of computing due to physics. He has authored 11 publications and holds one patent.

Mr. Sarpeshkar received an award from MIT for his B.S. thesis, which covered sense amplifiers in DRAM's. He is a member of Tau Beta Pi and Sigma Pi Sigma.

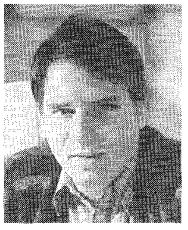


Jörg Kramer received a degree in physics from the Swiss Federal Institute of Technology (ETHZ), Lausanne, in 1987, and the M.S. degree in applied optics from the Imperial College, London, U.K., in 1988. In 1993 he received the Ph.D. degree in physics from ETHZ for his work on an optoelectronics project. Since 1993 he has been a postdoctoral fellow at the California Institute of Technology, where he works on analog VLSI vision systems.



Giacomo Indiveri was born in Sassari, Italy, in 1967. He received the Laurea degree in electrical engineering from the University of Genova, Italy, in 1992.

From 1992 to 1995 he was a Research Fellow with the National Research Program on Bioelectronic Technologies of the University of Genova's Department of Biolelectronic and Electronic Engineering. Since 1994 he has also been a Research Fellow in the Biology Department at the California Institute of Technology, where he works on the design of analog VLSI subthreshold neuromorphic architectures for low-level visual tasks and motion detection. His research interests are in the areas of neural computation, analog VLSI, and biological signal processing.



Christof Koch received the Ph.D. degree from the University of Tübingen, Germany. He was a postdoctoral fellow at MIT's Artificial Intelligence Laboratory.

In 1986 he joined the California Institute of Technology. His research focuses on understanding the biophysical mechanisms underlying motion and visual attention in cortical networks in the mammalian visual system. His laboratoy builds neuromorphic, analog, smart vision chips to solve a variety of aplied vision problems. He has published three books, over 170 technical papers, and holds numerous patents in the area of vision chips.