Pulse-Based 2-D Motion Sensors

Charles M. Higgins, Rainer A. Deutschmann, *Member, IEEE,* and Christof Koch, *Member, IEEE*

***Abstract*— We present two compact CMOS integrated circuits for computing the two-dimensional (2-D) local direction of motion of an image focused directly onto the chip. These circuits incorporate onboard photoreceptors and focal plane motion processing. With fully functional 14 × 13 and 12 × 13 implementations consuming less than 50 μW per pixel, we conclude that practical pixel resolutions of at least 64 × 64 are easily achievable. Measurements characterizing the elementary one-dimensional motion detectors are presented along with a discussion of 2-D performance and example 2-D motion vector fields. As an example application of the sensor, it is shown that the array as fabricated can directly compute the focus of expansion of a 2-D motion vector field.**

***Index Terms*— Analog VLSI, CMOS imager, focus of expansion, motion sensor, vision chip.**

I. Introduction

WEIGHING less than an ounce and consuming milliwatts of power, integrated CMOS vision sensors are ideal to serve as a replacement for a standard CCD camera and image processor for embedded applications. These integrated circuits incorporate photosensing and image processing in the focal plane for real-time application-specific processing. Such sensors are most advantageously applied on platforms like automobiles, robotics, and spacecraft, where power consumption and weight are primary concerns.

The analysis of visual motion by pixel-parallel integrated CMOS circuitry is particularly appropriate because it can be done in continuous time, eliminating the temporal aliasing problem found in sampled-time motion systems. While processing of visual motion is computationally intensive for a serial computer, pixel-parallel processing of motion allows real-time responses with relatively low computational power per processor.

The integrated motion sensors presented in this paper detect intensity edges at each pixel in the image and track them over time. Flow field vectors indicating motion are generated at the location of the moving edges. In order to obtain the densest focal-plane motion array possible, we have designed the sensors to output the local direction of motion (sign of velocity). Direction of motion is sufficient to locate the focus of expansion or contraction in an image, and to detect the axis and direction of rotation [1]. Direction of motion is also sufficient in many cases for motion segmentation. The average direction of motion over the entire image can produce an overall direction of translation, which can be used for image stabilization or as self-motion information. Since local direction of motion can be used for many of the same tasks as local velocity, the high pixel resolutions possible with these sensors may make them more practically useful than present lower-resolution velocity sensors. This work has been briefly described in [2].

Manuscript received February 19, 1998; revised February 23, 1999. This work was supported by the Center for Neuromorphic Systems Engineering as a part of the National Science Foundation's Engineering Research Center program, and by the Office of Naval Research. The work of R. A. Deutschmann was supported by the German National Merit Foundation. This paper was recommended by Associate Editor B. Linares-Barranco.

C. M. Higgins and C. Koch are with the Division of Biology, 139-74, California Institute of Technology, Pasadena, CA 91125 USA.

R. A. Deutschmann is with the Walter Schottky Institute, Technische Universität, Munich, Germany.

Publisher Item Identifier S 1057-7130(99)04878-8.

TABLE I
REVIEW OF INTEGRATED 2-D MOTION SENSORS

Authors	Reference	Year	Function
Tanner and Mead	[4]	1986	Global velocity
Delbrück	[5]	1993	Semi-local direction of motion
Etienne-Cummings et al.	[6]	1997	Local velocity
Deutschmann et al.	[8]	1998	Local direction of motion

II. Previous Work

Previous work in integrated motion sensors has primarily concentrated on one-dimensional (1-D) motion for simplicity. However, many of these 1-D motion sensors are not suitable for extension to the measurement of two-dimensional (2-D) motion due to circuitry size, interconnection complexity, or algorithmic limitations. Discrete-time (clocked) implementations of motion sensors, in general, consume much more power than continuous-time implementations. Thus, we restrict our review to continuous-time integrated sensors, which are capable of providing 2-D direction and/or velocity. The sensors in our survey are summarized in Table I. An extensive survey of motion sensors, including 1-D sensors, can be found in [3].

Any review of continuous-time motion sensors must begin with Tanner and Mead's 1986 gradient-based optical flow chip [4], which provides an estimate of the global 2-D velocity by satisfying a global image constancy constraint in an 8 × 8 photoreceptor array. The chip worked for high contrasts over a small velocity range, and suffered from fixed pattern noise due to low precision photoreceptors and derivative circuits. Due to the difficulty encountered in implementing the gradient method, many subsequent designs were based on correlation algorithms.

Delbrück's velocity-tuned retina [5] was built on a 26 × 26 hexagonal grid and incorporated delay lines in three primary directions. If the velocity of a moving edge matched that of the delay line, the output signals grew in strength; otherwise, they diminished. Because of the velocity-tuned design, multiple chips are required to measure motion over a large velocity range.

The velocity sensor work of Etienne-Cummings *et al.* [6] was based on analog photosensing and edge detection stages

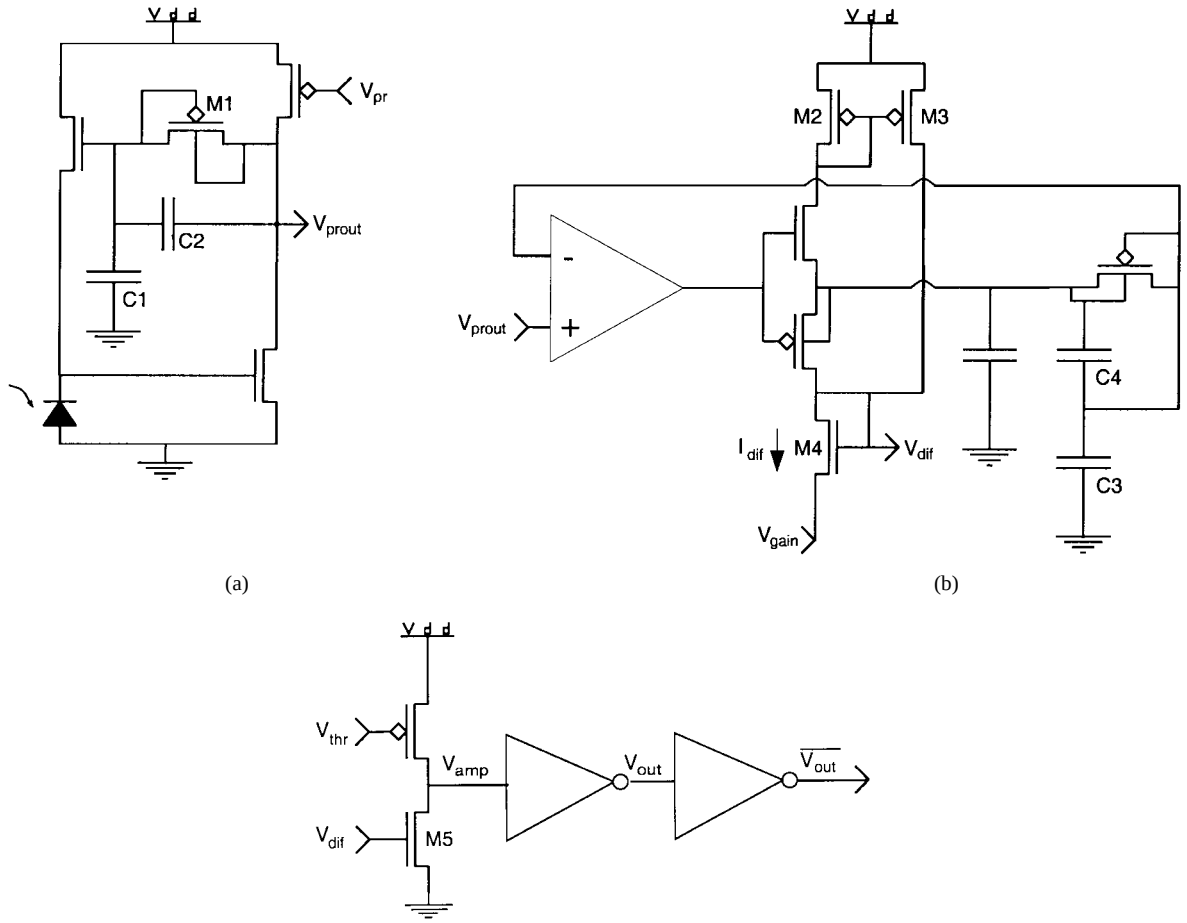


Fig. 1. TED circuitry. (a) Adaptive photoreceptor [9]—transistor $M1$ allows long-term adaptation to the light level and the ratio of $C1$ to $C2$ sets the transient gain. (b) Nonlinear differentiator [10]—transistors $M2$ and $M3$ have been added to the original circuit to allow response to both signs of temporal edges and the ratio of $C3$ to $C4$ sets the feedback gain. The operational amplifier is implemented with a standard 5-transistor transconductance amplifier. (c) Thresholding circuit—when $V_{\rm dif}$ exceeds the threshold set by $V_{\rm thr}$, $V_{\rm out}$ and $\overline{V}_{\rm out}$ will pulse.

followed by digital motion circuitry. The disappearance of an edge at one pixel and its subsequent reappearance at another pixel was used to compute time of travel. Each pixel in a 5×5 array computed a local velocity vector over two orders of magnitude of velocity and down to a theoretical minimum contrast of 20%. The motion circuitry in this design was laid out in three localized blocks: a high fill-factor photoreceptor array, an edge-detector array, and a motion-processing array. Due to the wiring required between the three stages of processing, this design cannot be easily scaled to large array sizes. The largest fabricated array of these sensors to date is 9×9 pixels [7].

Recently, Deutschmann *et al.* [8] have presented a 15×15 motion sensor array that computes motion by a real-time multiplication of temporal and spatial derivatives. Due to the multiplicative algorithm used, the motion output is not only dependent on stimulus orientation and velocity, but also on stimulus spatial frequency and contrast. Because spatial frequency and contrast may vary over the image plane, the velocity cannot be unambiguously determined from the sensor output. For this reason, this output can be interpreted as giving information about direction of motion only. Unlike the sensors described in this paper, the output of this gradient-based sensor has no persistence time; that is, the motion output has nonzero magnitude only as long as the motion is actually present.

III. Temporal Edge Detector

The motion circuits we present here rely on the output of a temporal edge detector (TED) located in each pixel, which creates a sharp voltage pulse when the light intensity at that pixel changes abruptly. The pulses created by the TED are interpreted as moving spatial edges, and are used by the motion circuits with the assumption that every TED will respond with a single pulse as an edge crosses that pixel. Any other feature-detecting circuit that conforms to this assumption could also be used as a front end for the motion computation.

The TED circuit we have used consists of three stages: an adaptive photoreceptor circuit, which transduces local light intensity at a pixel into a voltage, a nonlinear differentiator circuit, which produces an analog voltage pulse for an abrupt change in photoreceptor voltage, and a thresholding circuit, which converts the differentiator output into a larger voltage pulse. The photoreceptor has been fully characterized in [9]. The subsequent stages are discussed below. All three TED stages are shown in Fig. 1; example oscilloscope traces for a passing edge are shown in Fig. 2.

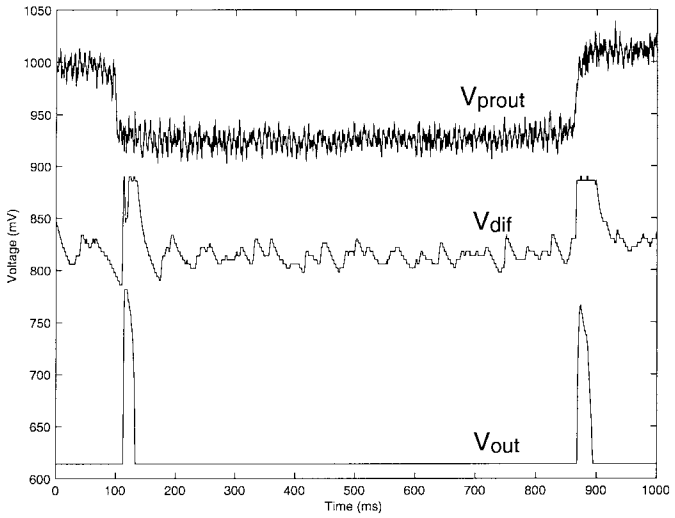


Fig. 2. Oscilloscope traces from the TED. The top trace is the photoreceptor output $V_{\rm prout}$ (shifted down by 100 mV for display), shown as a black bar on a white background passes the pixel. 120-Hz noise from the AC lighting is clearly visible. The middle trace is the corresponding nonlinear differentiator output $V_{\rm dif}$. Peaks for both signs of edges can be seen, along with subthreshold variations due to the lighting flicker. The small peak immediately following the response to the falling edge grows in size as stimulus speed increases, and eventually results in two above-threshold pulses for the falling edge. The bottom trace is the output $V_{\rm out}$ of the thresholding circuit (scaled by 10 and shifted up by 600 mV for display).

The nonlinear differentiator circuit of [10] has been modified to allow response to both signs of derivative. This makes the TED responsive to both dark-to-bright and bright-to-dark edges, which greatly increases the density of the flow fields produced. An additional current mirror [$M2$ and $M3$ in Fig. 1(b)] adds the current generated by rising photoreceptor transitions to the original current path. While this has the desired effect of allowing response to both signs of edges, it also results in a secondary response to a falling edge as the adaptive element recovers, seen weakly in the middle trace in Fig. 2. Due to the different current pathways activated by rising and falling edges, the differentiator output is not identical for the two edge types.

The thresholding circuit [Fig. 1(c)] is used to threshold and amplify the analog voltage pulse $V_{\rm dif}$ from the differentiator. This is necessary because of the small magnitude of the change in $V_{\rm dif}$, and because $V_{\rm dif}$ does not rest at ground, and so cannot be used directly as input to the digital motion circuitry without leakage problems. The circuit consists of a high-gain-inverting two-transistor amplifier followed by two CMOS inverters. Neglecting the Early effect, the bias $V_{\rm thr}$ simply sets a current threshold which transistor $M5$ must overcome to lower the output voltage. A small increase in $V_{\rm dif}$, causing $M5$ to cross the threshold, will cause a large downward swing in the amplifier voltage $V_{\rm amp}$, which is converted to a larger pulse by the first inverter. The second inverter makes available the opposite polarity of pulse; both polarities are used in the motion circuitry.

Kramer *et al.* [10] showed that the differentiator output current $I_{\rm dif}$ is proportional to stimulus *temporal contrast* C_t

$$C_t = V \times C_s$$

where V is speed and C_s is spatial contrast. Since transistors $M4$ and $M5$ make up a current mirror and $V_{\rm thr}$ sets a current threshold for pulse initiation, only edges above a certain temporal contrast will be detected by the TED. Due to the increased differentiator response at higher temporal contrast, the pulses from the thresholding circuit will be wider at higher speeds.

When visually stimulated by edges with temporal contrast near the threshold, the TED will fire probabilistically. The finite transition region between zero and unity firing probability is caused by noise and hysteresis in this highly nonlinear analog circuit. At this time, there exists no adequate analytical model of probabilistic TED performance, taking these factors into account.

IV. Motion Computation

In this section, two circuits are described for determining the direction of motion of an intensity edge. Using only nearest-neighbor connectivity, these circuits use TED pulses to produce voltages encoding the direction of motion, and then convert these voltages into a bidirectional current for output. Since both 2-D motion circuits presented are equivalent to two orthogonally-placed 1-D sensors, for simplicity we will describe only 1-D motion circuitry.

The output of each 1-D motion sensor represents the order in which the associated photoreceptors were crossed by a moving spatial edge. Like all local motion sensors, these sensors suffer from the aperture problem, and thus can only respond to the component of optical flow normal to the local gradients of intensity. The final result of this 1-D computation is the sign of the projection of the normal flow vector onto the sensor orientation. Two such sensors placed orthogonally effectively compute the vector sign of the local normal flow in the coordinate system of the chip.

The motion algorithms used in these circuits fall into the broad class of correlation-based motion algorithms, but due to the strong nonlinearities involved in the TED thresholding and digital motion computation, cannot be shown to be equivalent to biologically plausible algorithms such as that proposed by Reichardt [11].

A. Inhibit, Trigger, and Inhibit (ITI)

The basis of the ITI sensor is a voltage signal that is pulled high (triggered) by one TED and pulled low (inhibited) by a neighboring TED. A passing edge must cross both TED's to be detected. If the triggering TED is crossed last, the voltage signal will remain high, indicating an edge's passage in the sensor's preferred direction. If the inhibiting TED is crossed last, the signal will go briefly high and be left low after the edge's passage. The ITI sensor was inspired by the FTI sensor of Kramer [12].

The 1-D ITI sensor is diagramed in Fig. 3. The local TED pulse created by an edge crossing any pixel triggers the direction voltage for both directions. The TED pulse from the neighboring pixel, which is next crossed by the edge, inhibits the direction voltage in the incorrect direction, leaving only the correct direction of motion activated. Since the output of the sensor is calculated as the difference of the two direction

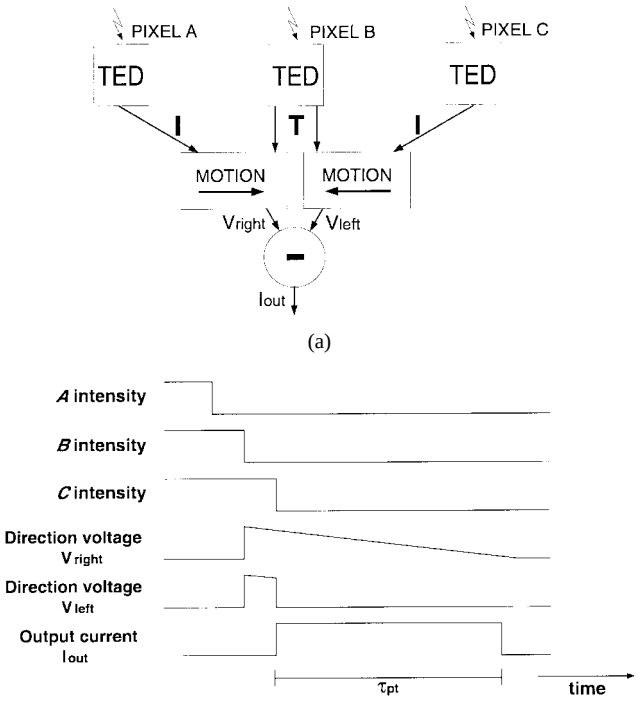


Fig. 3. ITI sensor. (a) Block diagram. (b) Simulated traces. An edge crossing the sensor from left to right pulls high (triggers) direction voltages for both directions $V_{\rm right}$ and $V_{\rm left}$ in pixel B. The same edge subsequently crossing pixel C pulls low (inhibits) the incorrect direction voltage $V_{\rm left}$. The resulting positive output current $I_{\rm out}$ indicates rightward motion. Pixels B and A interact similarly to detect leftward motion, resulting in a negative output current. Note that no speed information is encoded in the magnitude of $I_{\rm out}$.

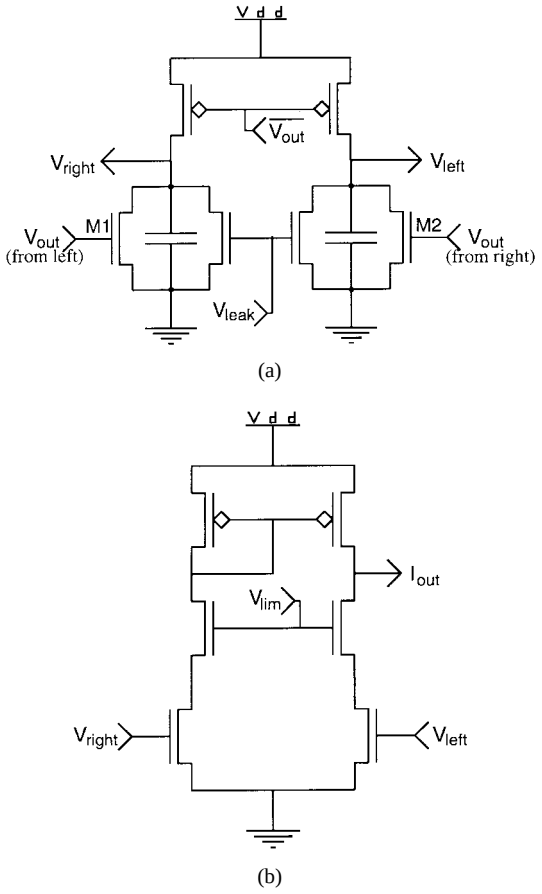


Fig. 4. ITI circuits. (a) Motion circuit. The local TED signal $\overline{V}_{\rm out}$ is used to pull high the direction voltages $V_{\rm right}$ and $V_{\rm left}$ when an edge passes. The signals $V_{\rm out}$ come from neighboring pixels, and are used to pull the direction voltages low. The bias $V_{\rm leak}$ sets the persistence time $\tau_{\rm pt}$ of the motion signal. (b) Current output circuit. The state of the direction voltages determines the sign of the bidirectional output current $I_{\rm out}$, whose magnitude is set by $V_{\rm lim}$.

voltages, no output is produced until one of the two direction voltages is inhibited.

Note that only two wires are required for communication between a pair of neighboring pixels; they carry inhibiting TED pulses in opposite directions. Since these wires connect to both left and right neighbors, a total of four wires (two entering and two leaving) connect each 1-D pixel to its neighbors.

The ITI motion circuit schematic is shown in Fig. 4(a). The signal $\overline{V}_{\rm out}$ comes from the local TED and is used to pull up (trigger) both direction voltages $V_{\rm left}$ and $V_{\rm right}$ to $V_{\rm dd}$. The voltage signals $V_{\rm out}$ connected to the gates of $M1$ and $M2$ come from neighboring pixels to the left and right, respectively, and are used to pull down (inhibit) the direction voltages. Through the transistors with constant gate bias $V_{\rm leak}$, the capacitors are linearly discharged, thus determining the persistence time $\tau_{\rm pt}$ of the motion vectors. The current output circuit in Fig. 4(b) computes the difference between the two direction voltages. The sign of the current output $I_{\rm out}$ is determined by which of the two direction voltages $V_{\rm left}$ and $V_{\rm right}$ is high. If both voltages are high, only a small mismatch current output is produced. The magnitude of the current output is set by the bias $V_{\rm lim}$. Since $V_{\rm lim}$ is small compared to $V_{\rm dd}$, $I_{\rm out}$ maintains a constant magnitude for almost all of its duration.

B. Facilitate, Trigger, and Compare (FTC)

The idea behind the FTC sensor is to open up a time window upon the occurrence of an edge at one pixel and then to determine at which neighboring pixel the edge next appears. A passing edge must cross two pixels in order for its direction to be determined.

A 1-D FTC sensor is diagramed in Fig. 5. Building blocks are a facilitation circuit to provide a voltage signal for a fixed time after edge detection, a direction-of-motion circuit for each direction, and a comparator to produce the output current.

When an edge crosses a pixel, its facilitation signal is activated. If the edge then crosses the next pixel within the facilitation time, the appropriate direction voltage will be triggered and then decays linearly over time. The incorrect direction voltage will not be triggered because it has not been facilitated before. A bidirectional output current is generated depending on which direction voltage is higher. If both direction voltages are low, no output current is generated.

Note that, again, only two wires are required for communication between a pair of neighboring pixels; the facilitation signal and trigger pulse are sent leftward. Since these signals are also received from the right, a total of four wires (two entering and two leaving) connect each 1-D pixel to its neighbors.

The FTC circuit schematic is shown in Fig. 6. When the local TED voltage $V_{\rm out}$ pulses, transistor $M1$ is switched

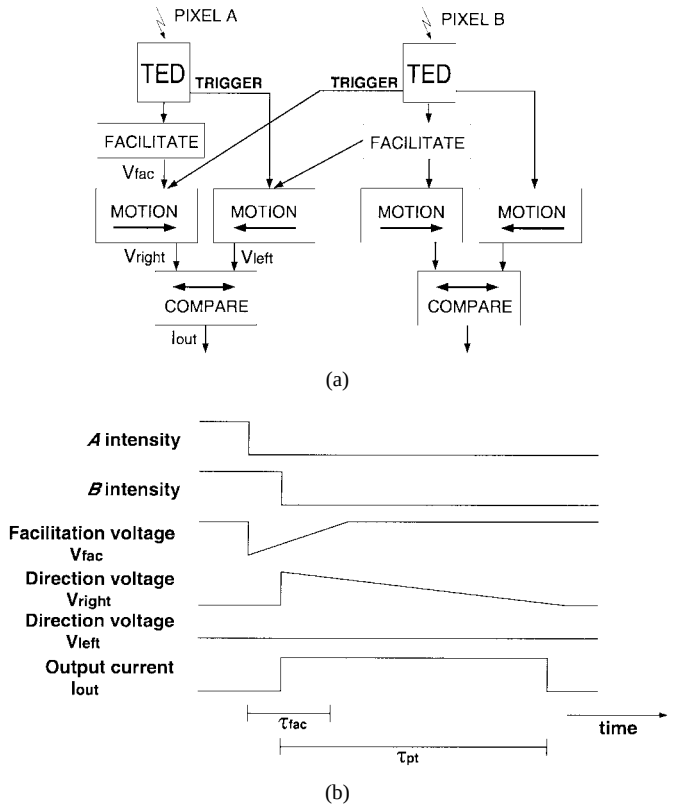


Fig. 5. FTC sensor. (a) Block diagram. (b) Simulated traces. An edge moving left to right first crosses pixel A, opening a facilitation time window ($V_{\rm fac}$ goes low). When the edge reaches pixel B, the rightward direction voltage $V_{\rm right}$ is triggered. The incorrect direction voltage $V_{\rm left}$ is not activated because it is triggered before facilitation arrives from pixel B. Voltages $V_{\rm left}$ and $V_{\rm right}$ are compared and a positive current $I_{\rm out}$ is generated, indicating rightward motion. A leftward-moving edge would result in a negative output current. Note that no speed information is encoded in the magnitude of $I_{\rm out}$.

on briefly, thus discharging the facilitation capacitor $C1$ to ground. Subsequently $C1$ leaks up to $V_{\rm dd}$ within the time $\tau_{\rm fac}$ (usually about 50 ms), depending on the bias setting $V_{f\_{\rm leak}}$ on transistor $M2$. At the same time the opposite polarity TED pulse $\overline{V}_{\rm out}$ is used to switch on transistor $M3$. If at this moment the neighboring pixel to the right is facilitated, the capacitor indicating leftward motion $C2$ is charged up to $V_{\rm dd}$. The voltage $V_{\rm left}$ on capacitor $C2$ decays to ground within $\tau_{\rm pt}$, depending on the bias voltage $V_{\rm leak}$ on the leakage transistor $M4$. Rightward motion is detected correspondingly with the facilitation voltage $V_{\rm fac}$ and the TED pulse $\overline{V}_{\rm out}$ from the rightward neighboring pixel, thus generating a voltage $V_{\rm right}$. Leftward and rightward motion indicator voltages $V_{\rm left}$ and $V_{\rm right}$ are compared in the current output circuit, a standard transconductance amplifier. A constant positive current $I_{\rm out}$ is generated if $V_{\rm right}$ is greater than $V_{\rm left}$, indicating rightward motion. For leftward motion $I_{\rm out}$ is negative. The magnitude of $I_{\rm out}$ is determined by the bias transistor $M5$. The time $I_{\rm out}$ stays on is given by the persistence time $\tau_{\rm pt}$. If both $V_{\rm left}$ and $V_{\rm right}$ are at ground (no motion), the output current $I_{\rm out}$ is zero.

V. Motion Sensor Characterization

Both circuits were fabricated in a 1.2-μm CMOS process on a MOSIS tiny chip (2.2 mm $\times$ 2.2 mm). The layout of the ITI chip is shown in Fig. 7, and has the same general floorplan as the FTC chip. Details of each circuit implementation are shown in Table II. The photodiode light collection area, total number of transistors, and the sum of all capacitors used to implement each pixel are shown. The minimum power dissipation is consumed by the chip with no stimulus. The maximum power dissipation is consumed with a full-field stimulus while all vectors persist. The higher minimum power dissipation of the ITI pixel is accounted for by leakage current in minimum-length digital transistors.

Both chips incorporate serial pixel scanners for real-time readout of both the motion vector outputs and the raw photoreceptor responses. With a standard PC, we are able to scan out and display the motion vector field and the corresponding image with a frame rate above 50 Hz, limited by the graphics display speed of the PC. Both chips are also capable of computing a real-time average of any subarray of the output vector field due to the programmability of the serial scanners. This fact will be exploited in Section VI to allow direct computation of the focus of expansion.

A. Elementary Motion Sensor

Since each 2-D sensor is composed of two orthogonally placed noninteracting 1-D sensors, we can completely characterize 2-D sensor performance by analyzing the contrast, speed, and orientation response of the elementary 1-D direction-of-motion sensors.

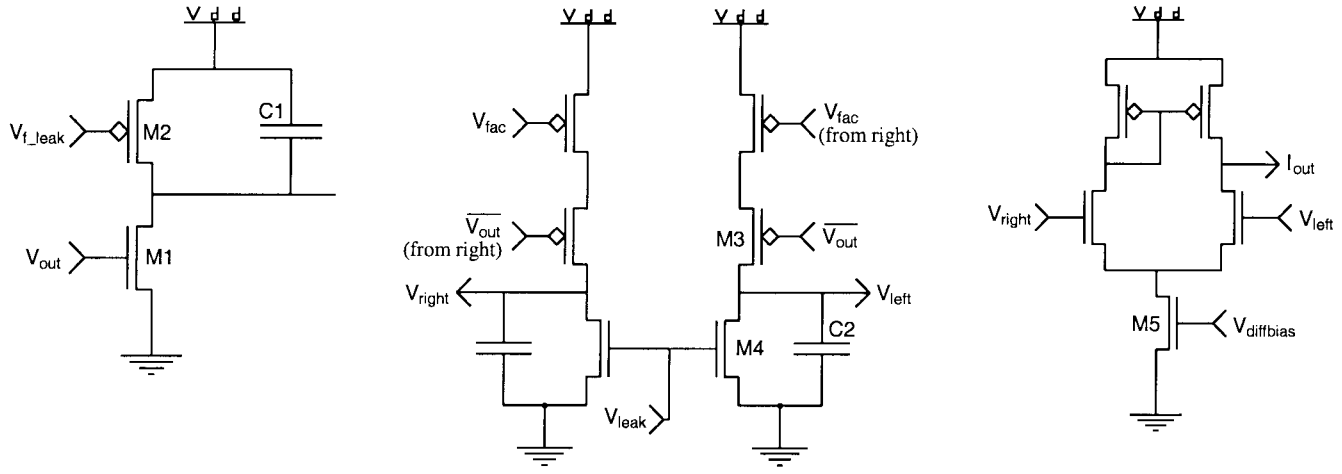
The motion computation algorithms used here are independent of stimulus spatial frequency until the point of spatial aliasing, so a spatial frequency response curve is not shown. Aliasing will occur at higher spatial frequencies than 0.5 cycles/pixel for both sensors.

Recall that the output of each 1-D sensor is a current which ideally assumes one of three values: positive, zero, or negative. The data shown were collected in AC lighting by using an 8-mm focal-length lens to focus a stimulus onto the chip from a distance of 23.5 cm. Although the sensors will respond to both dark-to-bright and bright-to-dark edges, the data presented below were obtained for bright-to-dark edges only.

1) Speed and Contrast Sensitivity: Here, we report the performance of a single 1-D sensor from each chip in response to a full-field square-wave grating stimulus as speed and contrast are varied and the angle is held constant.

In Fig. 8, we plot the probability of correct responses to the stimulus for both sensors over 50 trials. From the data it can be seen that, for contrasts above 20%, both sensors compute the correct direction of motion over approximately two orders of magnitude in speed.

The velocity and contrast range over which pulses are reliably produced by the TED places bounds on the speed response of the motion sensor. We see as expected that the low-speed threshold is inversely proportional to contrast. At a very high speed, the TED's on both chips intermittently produce *two* pulses for each edge, due to the small secondary peak discussed in Fig. 2, leading to incorrect responses. Double-pulsing begins at approximately the same speed for all contrasts. Due to the highly nonlinear nature of the TED



(a) (b) (c)

Fig. 6. FTC circuit. (a) Facilitation circuit. The facilitation pulse $V_{\rm fac}$ is pulled low by the local TED pulse $V_{\rm out}$ and leaks back to $V_{\rm dd}$ with a time constant set by $V_{f\_{\rm leak}}$. (b) Motion circuit. Direction voltages $V_{\rm left}$ and $V_{\rm right}$ are pulled high only when facilitation voltage and TED output are both active. (c) Current output circuit. The sign of the current $I_{\rm out}$ is determined by which of the two direction voltages $V_{\rm left}$ and $V_{\rm right}$ is higher.

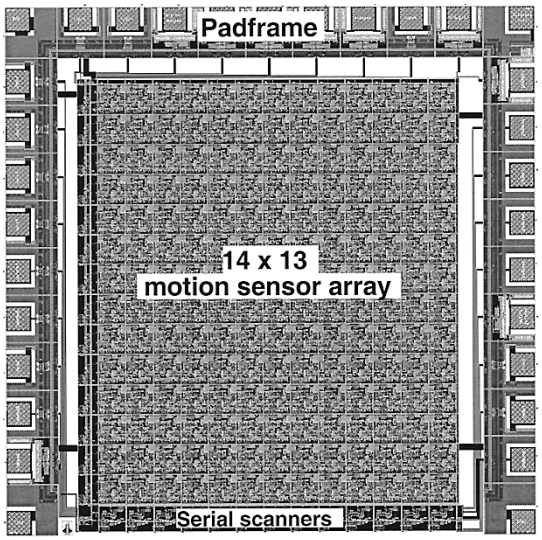


Fig. 7. ITI layout. The serial scanners are on the left and bottom; the rest of the chip consists of a 14 × 13 array of 182 identical pixels. This design was fabricated on a MOSIS tiny chip (2.2 × 2.2 mm) in a 1.2-μm CMOS process.

TABLE II
COMPARISON OF ITI AND FTC MOTION PIXELS

	ITI	FTC
Pixel Size	110 × 120 μm^2	128 × 119 μm^2
Array Size	14 × 13	12 × 13
Light Collection Area	15 × 15 μm^2	15 × 15 μm^2
Number Transistors	49	47
Total Pixel Capacitance	1.2 pF	3.2 pF
Minimum Power	8 μW	0.6 μW
Maximum Power	47 μW	29 μW

circuitry, no adequate analytical model of the double-pulsing phenomenon exists.

While Fig. 8 shows the maximum possible dynamic range of speed with the current TED implementation, it is possible to adjust this range to a significantly lower or higher speed for a given application.

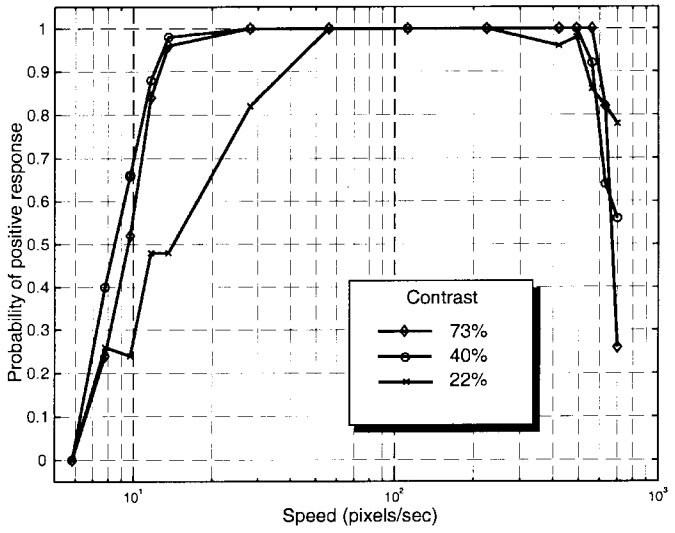
2) Orientation Response: Here we characterize a single 1-D sensor as the angle of a full-field square-wave grating stimulus is varied while leaving the contrast and speed constant.

For a medium-speed stimulus, Fig. 9 shows that the ITI sensor consistently responds with a positive direction of motion for 150° of orientation centered on the positive sensor orientation, a negative direction for 150° centered on the negative sensor orientation, and transitions smoothly in the 30° near orthogonal to both sensors. This orientation response shows a good representation of the sign of stimulus velocity. The FTC sensor's response is qualitatively similar.

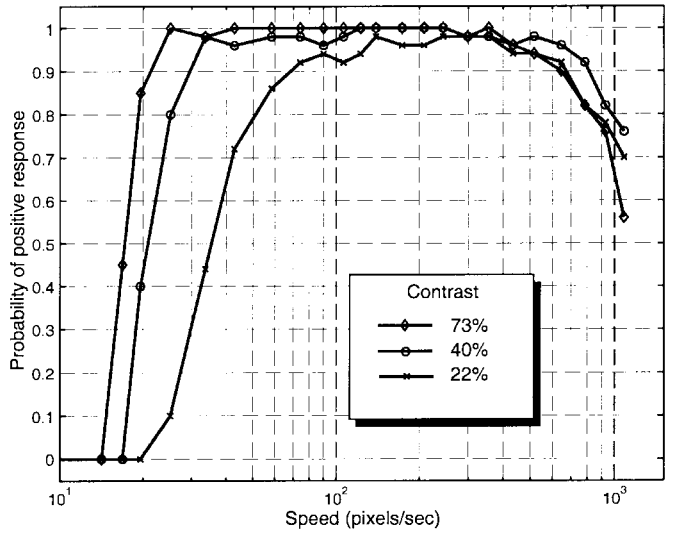
As the stimulus angle changes from the sensor's preferred direction toward orthogonal, the time between stimulation of sensor photoreceptors decreases exactly as if the speed of the stimulus were increasing. However, stimulus temporal contrast is not changing, and thus the TED pulses remain the same. For this reason, the orientation response cannot be predicted from the speed response.

As the stimulus angle increases, a point is reached at which the TED pulses from neighboring pixels begin to overlap. For overlapping pulses, the FTC sensor will trigger both directions of motion, resulting in a zero output. Thus, the sign transition bandwidth depends on the TED pulse width. Because the TED pulse width increases with stimulus speed, so will the sign transition bandwidth for the FTC sensor.

The ITI sensor will continue with the correct output until the TED pulses from neighboring pixels are fully overlapping. At this point, the ITI sensor will also trigger both direction voltages and produce a zero output. However, due to variations in the response delays of the three TED's involved in the ITI sensor, a further increase of the stimulus angle may be required to produce the opposite sign output. The amount of this further increase will depend on stimulus speed, since the TED response delays are fixed. For this reason, the sign

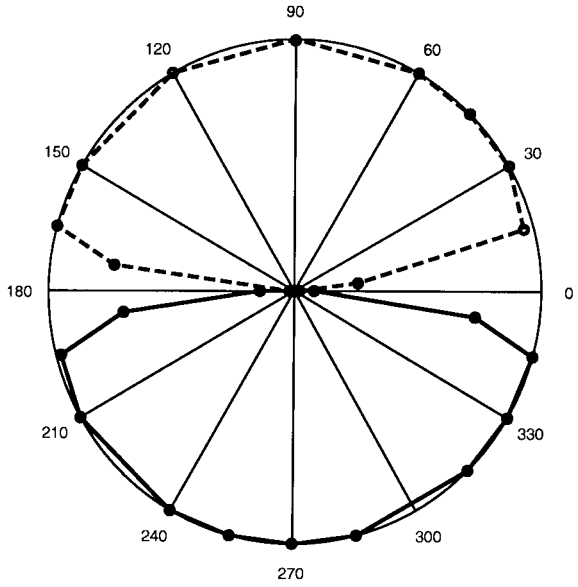


(a)

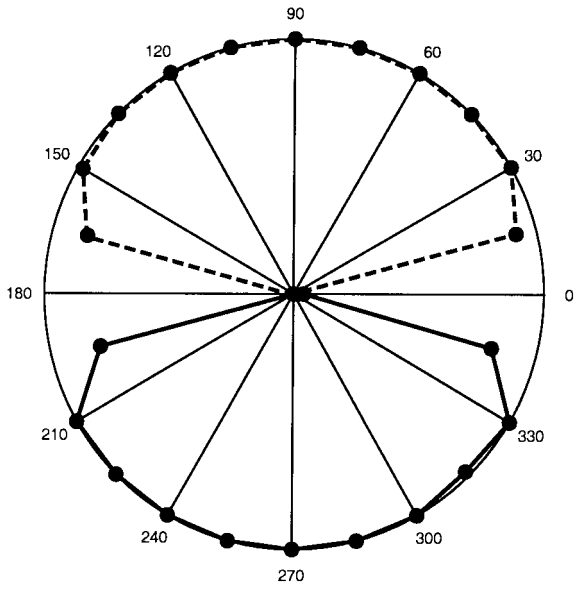


(b)

Fig. 8. Single-pixel speed sensitivity for different contrasts at constant stimulus angle (optimal orientation for a positive response). (a) ITI sensor. (b) FTC sensor. The probabilities have been calculated from 50 stimulus presentations. The cutoff at low speeds is due to low temporal contrast. Note the reliable response at high contrast over two orders of magnitude in speed. At very high speeds, the error is due to double-pulsing in the TED's.



(a)



(b)

Fig. 9. Single-pixel orientation response. (a) ITI sensor. (b) FTC sensor. The probability of positive (dashed line) and negative (solid line) responses is plotted as the stimulus direction is varied. The probabilities have been calculated from 50 stimulus presentations. The radial angle indicates stimulus direction in degrees. A 90° stimulus is parallel to the positive-facing sensor, and a 270° stimulus to the negative-facing sensor. A radius value on the circle indicates a response probability of 1.0. Stimulus spatial contrast was 74%, and speed was 140 pixels/s.

transition bandwidth of the ITI sensor will also increase with stimulus speed.

B. 2-D Array Performance

Above, we have characterized a single 1-D elementary motion sensor over a variety of stimulus contrasts, speeds, and orientations. Here we describe the performance of an array of 2-D sensors made up of these elementary 1-D sensors.

As in any 2-D sensor made up of two independent 1-D sensors, the X and Y components of a direction vector may appear and disappear at slightly different times. The time between appearance of the X and Y vector components depends on the angle of the stimulus and is always less than one pixel transit time: between 1.4 and 140 ms for the stimuli presented in this paper. Since both vector components decay independently, it is possible for one component to disappear before the other. This is dependent upon the matching of the decay (V_{leak}) transistors. For typical matching seen in fabricated designs, the time between the disappearance of the two components is less than 5% of the total persistence time of the vector.

Since, as shown in Section V-A-2, each 1-D sensor tends to respond with a zero value only in a small range of stimulus angles near orthogonal, each 2-D sensor will effectively prefer diagonal vectors (i.e., both X and Y components nonzero).

Although eight directions of motion can be represented by a single 2-D sensor, the four diagonal directions will be seen for a much wider range of stimulus angles. This is the desired mode of operation, since the sensors were designed to report the sign of 2-D velocity.

The elementary 1-D sensor measurements were taken in the absence of serial scanner noise, which is incurred because of a high-speed clock line that must be run to each pixel. This clock switching noise, often seen in such mixed-signal designs, couples slightly into the photoreceptor and requires a decrease in the response speed of the nonlinear differentiator. When scanners are running continuously, the minimum temporal contrast that can be detected is unchanged. However, the high-speed response threshold is decreased by approximately 100 pixels/s due to earlier onset of double-pulsing. The orientation response is also unchanged.

Due to spatial variations in fabrication across the die, each nominally identical motion sensor has a somewhat different response. On the ITI chip, 96% of all pixels have a low-speed temporal contrast threshold (defined as the point of activation probability 0.5) within 5 pixels/s of the most common value. High-speed response thresholds are more variable, with approximately 70% of pixels within 100 pixels/s of the most common value. Orientation centers (the point of equal positive and negative activation) vary by 15°. Orientation bandwidths (the 0.5 probability bandwidth of the transition) vary by 30°. The variance of all parameters increases as the contrast decreases. All of the above variations are due to differences in TED responses, and are highly fabrication-process dependent.

In order to convey a sense of the real-time performance of the 2-D motion sensor arrays, we present image and flow-field samples from the 12×13 FTC chip. In Fig. 10, data is shown as a hand is moved across the chip's visual field. The outline of the hand is clearly visible in the image from the on-chip CMOS photoreceptor array, and the flow field vectors allow determination of the hand's direction of motion. The length of the trail of vectors following the hand is proportional to the persistence-time setting of the chip. In Fig. 11, the FTC chip's response to an approaching circle is shown. In this case, we expect to see an expanding flow field with all vectors pointing away from the center of the circle. Because the stimulus moves slightly down as it approaches the sensor, fewer vectors in the upward direction are produced. The focus of expansion can be clearly located, and the angular resolution of the direction-of-motion sensor is demonstrated. Note the preference for diagonal vectors. The ITI chip gives rise to qualitatively similar responses.

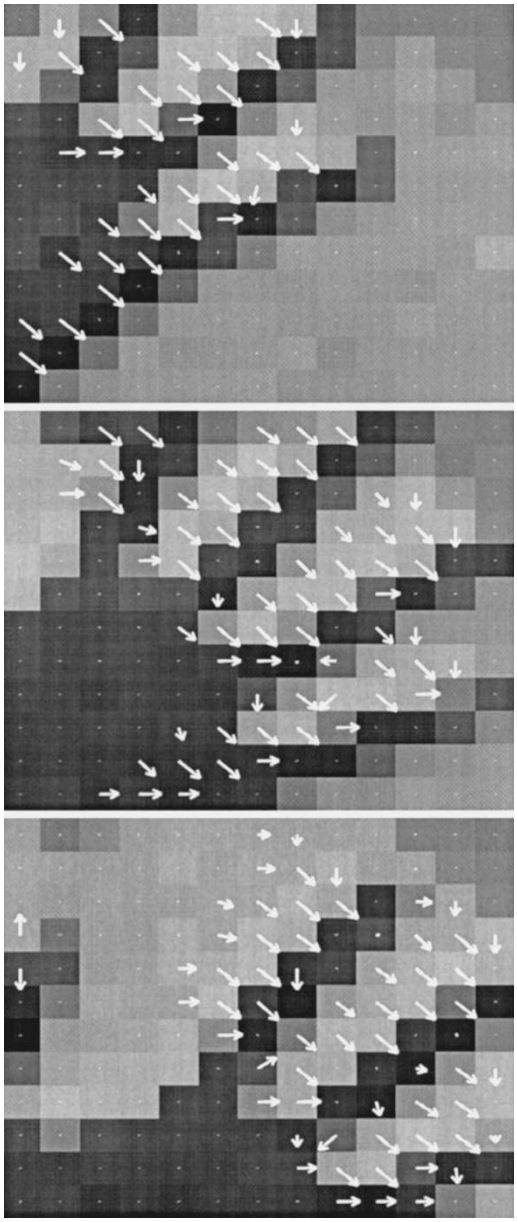


Fig. 10. Sample flow fields from the FTC chip for a moving hand: gray value image (raw photoreceptor outputs) and motion vector field were serially scanned from the sensor chip and displayed without any averaging. Pixel output currents were converted to voltages with a current sense amplifier. The frame rate was 50 Hz.

VI. Application to Focus of Expansion

In rotating, expanding, or contracting flow fields, the singular point is defined as the point at which the motion field vanishes [13]. Optical flow vectors grow smaller in magnitude as the singular point is approached and change sign as it is passed. The vector sum of the sign of optical flow can be used to locate the singular point [14] (see Fig. 12 for a more precise definition). When a stimulus producing such a flow field is centered over the chip (i.e., the singular point is coincident with the center pixel of the array), the vector sum of the sign of optical flow is zero because all vector components have an opposite-sign counterpart in another portion of the field. As such a field moves away from the chip center, the number of vectors of each sign becomes imbalanced, resulting in an increase in each component of the vector sum proportional to the number of pixels off-center.

Due to the fact that it is possible to program the serial scanners to compute a sum of all pixel outputs, the sensor arrays presented in this paper can compute the singular point location directly. Because velocity is not taken into account, uniform velocity over the field is not required for proper function as long as the stimuli are within the velocity limits of the sensor. For proper operation of the algorithm, the flow

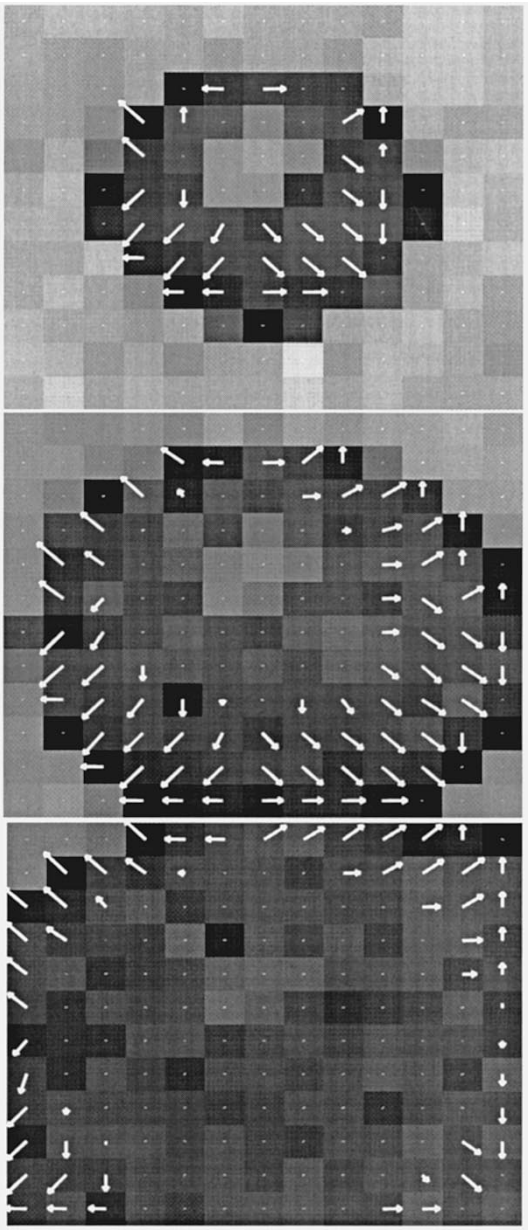


Fig. 11. Sample flow fields from the FTC chip for an approaching circle: gray value image (raw photoreceptor outputs) and motion vector field were serially scanned from the sensor chip and displayed without any averaging. Pixel output currents were converted to voltages with a current sense amplifier. The frame rate was 50 Hz.

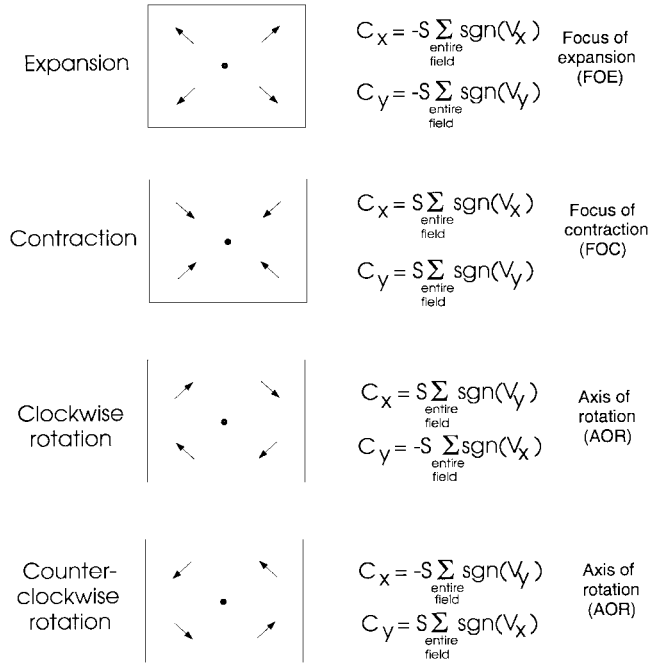


Fig. 12. Flow fields with singular points: for each of the four types of flow field, an example is given of the flow pattern, and the equation which can be used to calculate the singular point is shown. (C_x, C_y) is the position of the singular point *relative to the center of the chip*, (V_x, V_y) is a velocity vector in the array, and S is a constant of proportionality.

field must cover the entire field of view of the chip. This is a reasonable assumption for self-motion, since the entire visual scene will move relative to the sensor. Due to the persistence time of the elementary motion sensor, it is only necessary to have motion stimuli cross each portion of the visual field periodically.

In Fig. 13, a measurement is shown of the ITI chip output while an expanding spiral stimulus is presented. This stimulus creates an expansive flow field, but without the velocity variation seen in such a field generated by self-motion. The normalized X component of the chip output, averaged over one period of the stimulus, is plotted against X position of the FOE relative to the chip center. This output is independent of the Y position of the FOE. It varies linearly with FOE position at approximately 0.3-μA/pixel until the FOE moves off of the chip. The maximum output current is approximately 1.7-μA. With this output and the corresponding Y output, it is possible to specify the location of the FOE. The chip has also been tested with contracting and rotating stimuli and works equally well. However, it is important to note that the chip output in this mode cannot be directly used to detect the *type* of flow field being presented.

Because the sum of the entire array is being measured, the current magnitude bias V_{lim} is greatly reduced in this mode of operation. In addition, the serial scanners are static while this output is being produced. These factors together result in a significant reduction in power consumption. During this mode of operation, the entire ITI chip consumes approximately 2 mW (10 μW/pixel).

VII. Discussion

In this paper, we have presented two compact integrated CMOS hardware implementations for the real-time computation of 2-D visual direction of motion. The primary advances of the new sensors are in density, scalability, and power consumption. We have characterized the performance of the individual sensors, and shown a simple application of the sensor array to the location of the focus of expansion. Each individual pixel computes the direction of motion over two orders of magnitude in speed and down to 20% contrast while consuming less than 50 μW.

We have demonstrated fully functional prototype sensors fabricated in a 1.2-μm standard CMOS process on a 2.2 mm $\times$ 2.2 mm silicon die. The fully pixel-parallel sensor design

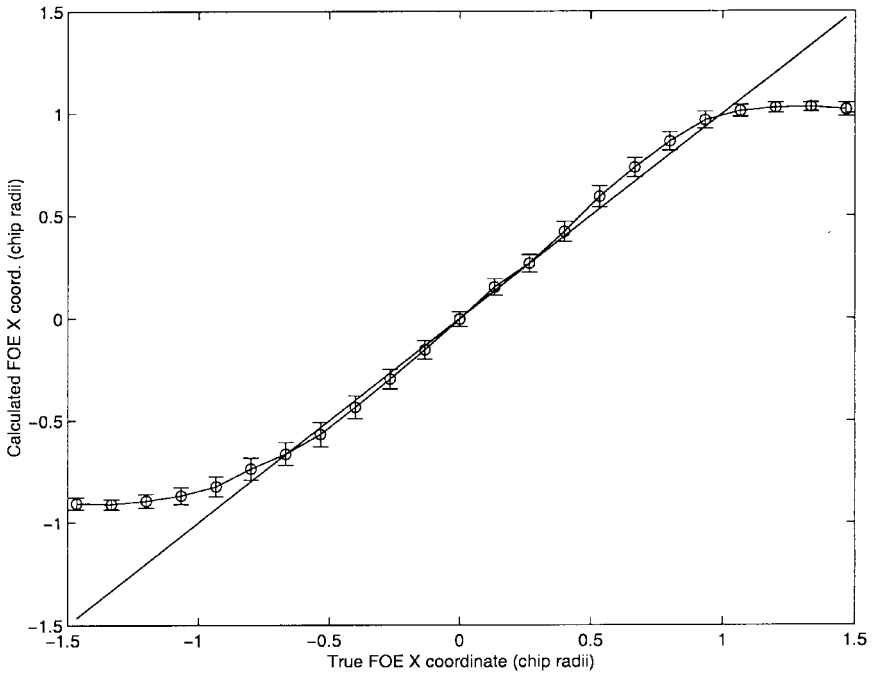


Fig. 13. FOE measurement: as an expanding spiral stimulus is presented to the ITI chip, the sum of all vector outputs is measured. Because the output current fluctuates slightly as the stimulus passes, an average of the chip's output is taken over one period of the stimulus while the position of the FOE remains constant. Plotted is the mean and standard deviation of this one-period average over 100 stimulus presentations. FOE positions are shown relative to the center of the array. The output current has been normalized to units of chip radii; the maximum output current is approximately 1.7-μA. The X output current is linear in the X position of the FOE, and saturates near the chip boundary when the FOE is completely off of the chip. This output is independent of the Y position of the stimulus.

scales directly to larger arrays. On a 9.4 mm $\times$ 9.7 mm die in the same process, straightforward replication of the sensor pixels will yield an array larger than 64×64 pixels consuming less than 200 mW. We are currently converting the sensors to an 0.5-μm CMOS process, which will allow 2-D arrays of more than 16 000 pixels.

The sensors presented in this paper grew out of the 1-D velocity sensor research of Kramer *et al.* [10]. While compact relative to many previous velocity sensors, the effort in these 1-D sensors was primarily concentrated on maximizing the velocity and contrast range. As such, these sensors were still too large to be used in practical 2-D focal plane arrays. In addition, many of the 1-D velocity sensors developed respond inappropriately to orthogonal stimuli, require large capacitors, or use long-range connectivity, making redesign necessary. In order to extend this work to 2-D motion sensors, we retained the basic pulse-based concept but redesigned the motion circuits in order to solve these problems. Using the same fabrication technology, the 2-D ITI direction-of-motion pixel requires 20% less layout area than the *1-D* FS velocity sensor [10] while consuming approximately the same power. Compared to the velocity sensor of Etienne–Cummings *et al.* [6], the ITI sensor requires 24% less layout area and consumes eight times less power per pixel.

Offboard spatial averaging can be used to improve the regularity of the flow field produced the sensor array at the cost of decreased spatial resolution. Because each elementary 1-D sensor will change its output from positive to negative at a slightly different stimulus angle, the angle of a spatially averaged vector output will change more smoothly as the stimulus angle changes than that of a single pixel. Temporal averaging of flow fields from this sensor array yields no significant improvement in performance on *transient* stimuli. Temporal averaging can be used to greatly increase the probability of detection for low temporal contrast *periodic* stimuli.

The two motion circuits presented in this paper are different in implementation but quite similar in performance. Based on our evaluation, there is no reason to prefer one sensor over the other. The only significant performance differences stem from the different implementations of the temporal edge detectors. Since both motion circuits are digital, their intrinsic speed limits lie in leakage currents for very low velocities and rise times for very high velocities. Neither limit has been approached by the temporal edge detectors demonstrated here. The primary avenue for advance in these sensors lies in improvements to the front-end edge detection circuits. It is worth noting that in a 2.0-μm process, the same TED circuitry used in this paper has been measured to perform well over nearly *four* orders of magnitude in stimulus velocity and down to 11% contrast. Due to the tighter controls placed on modern high-resolution processes, it is very likely that the TED design will also perform better in a 0.5-μm process than we have demonstrated here.

The CMOS sensors described here output a motion vector field, which can be scanned serially into a computer for further processing. This significantly reduces the image processing burden on the computer, but relies on serial processing for data reduction. We are, therefore, working toward systems in which information about the 2-D motion vector field can be directly output from the chip. A simple form of this idea is the

FOE calculation shown in Section VI. Many such quantities can be computed without the need for serial scanners and their associated switching noise, thus making maximum use of the sensor's capabilities. Using this technology, efficient and compact smart sensors that produce high-level motion information without the need for serial processing are feasible.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their help in clarifying this paper.

REFERENCES

[1] C. Fermüller and Y. Aloimonos, "Direct perception of three-dimensional motion from patterns of visual motion," *Science,* vol. 270, pp. 1973–1976, Dec. 22, 1995.

[2] R. Deutschmann, C. Higgins, and C. Koch, "Real-time analog VLSI sensors for 2-D direction of motion," in *Proc. Int. Conf. Artificial Neural Networks.* New York: Springer-Verlag, 1997, pp. 1163–1168.

[3] R. Sarpeshkar, J. Kramer, G. Indiveri, and C. Koch, "Analog VLSI architectures for motion processing: From fundamental limits to system applications," *Proc. IEEE,* vol. 84, pp. 969–987, 1996.

[4] J. Tanner and C. Mead, "An integrated analog optical motion sensor," in *VLSI Signal Processing II.* New York: IEEE Press, 1986, pp. 59–76.

[5] T. Delbrück, "Silicon retina with correlation-based, velocity-tuned pixels," *IEEE Trans. Neural Networks,* vol. 4, pp. 529–541, May 1993.

[6] R. Etienne-Cummings, J. Van der Spiegel, and P. Mueller, "A focal plane visual motion measurement sensor," *IEEE Trans. Circuits Syst. I* vol. 44, pp. 55–66, 1997.

[7] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M. Zhang, "A foveated visual tracking chip," in *Proc. Int. Solid-State Circuits Conf.,* 1997, San Francisco, CA.

[8] R. A. Deutschmann and C. Koch, "Compact real-time 2-D gradient based analog VLSI motion sensor," presented at Int. Conf. Advanced Focal Plane Arrays and Electronic Cameras, Zurich/Switzerland, 1998.

[9] T. Delbrück and C. Mead, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits," Dept. Computation Neural Syst., California Inst. Technol., Tech. Rep. 30, 1993.

[10] J. Kramer, R. Sarpeshkar, and C. Koch, "Pulse-based analog VLSI velocity sensors," *IEEE Trans. Circuits Syst. II,* vol. 44, pp. 86–101, 1997.

[11] J. P. H. van Santen and G. Sperling, "Elaborated Reichardt detectors," *J. Opt. Soc. Amer. A,* vol. 2, pp. 300–320, 1985.

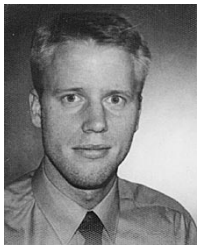
[12] J. Kramer, "Compact integrated motion sensor with three-pixel interaction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, pp. 455–460, 1996.

[13] A. Verri, M. Straforini, and V. Torre, "Computational aspects of motion perception in natural and artificial vision systems," *Philos. Trans. R. Soc. London B,* vol. 337, pp. 429–443, 1992.

[14] C. M. Higgins and C. Koch, "An integrated vision sensor for the computation of optical flow singular points," in *Advances in Neural Information Processing Systems.* Cambridge, MA: MIT Press, 1998, vol. 11.

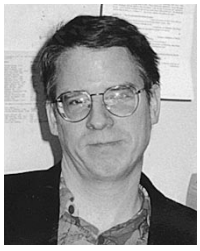
Charles M. Higgins received the Ph.D. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, CA, in 1993.

He was with the Radar Systems Group, MIT Lincoln Laboratory, Cambridge, MA, until 1996, when he returned to Caltech as a postdoctoral Research Fellow in the Division of Biology. His current research is in the area of analog VLSI vision chips and his research interests include analog computation, asynchronous digital systems, hardware implementations of biological neural systems, and autonomous robotics.



Rainer A. Deutschmann (M'98) received the Diploma degree in physics from the University of Technology, Munich, Germany, in 1997. He is currently working toward the Ph.D. degree at the Walter Schottky Institute for Semiconductor Physics, Garching, Germany.

He was a visiting Scientist at the California Institute of Technology, Pasadenta, CA, in 1996–1997. His research interests include quantum semiconductor structures, mixed-signal VLSI design, and neurophysics.



Christof Koch (M'89) was born in 1956 in Kansas City, KS. He received the Ph.D. degree in biophysics in 1982 from the University of Tuebingen, Germany, where he studied physics and philosophy.

After four years at MIT, Cambridge, MA, he joined California Institute of Technology, Pasadena, CA, where he is currently a Professor of computation and neural systems. He has written and edited five books, including a textbook on biophysics, and more than 200 technical articles on analog VLSI, vision and visual algorithms, computational neuroscience, and the neuronal correlates of consciousness (co-authored with Francis Crick).