



## Analog Integrated 2-D Optical Flow Sensor

ALAN A. STOCKER

Howard Hughes Medical Institute and Center for Neural Science, New York University, 4 Washington Place Rm 809, New York, NY 10003-1056, USA  
E-mail: alan.stocker@nyu.edu

Received June 21, 2004; Revised January 22, 2005; Accepted February 9, 2005

**Abstract.** I present a new focal-plane analog very-large-scale-integrated (aVLSI) sensor that estimates optical flow in two visual dimensions. Its computational architecture consists of a two-layer network of locally connected motion units that *collectively* estimate the optimal optical flow field. The applied gradient-based optical flow model assumes visual motion to be translational and smooth, and is formulated as a convex optimization problem. The model also guarantees that the estimation problem is well-posed regardless of the visual input by imposing a bias towards a preferred motion under ambiguous or noisy visual conditions. Model parameters can be globally adjusted, leading to a rich output behavior. Varying the smoothness strength, for example, can provide a continuous spectrum of motion estimates, ranging from normal to global optical flow. The non-linear network conductances improve the resulting optical flow estimate because they reduce spatial smoothing across large velocity differences and minimize the bias for reliable stimuli. Extended characterization and recorded optical flow fields from a  $30 \times 30$  array prototype sensor demonstrate the validity of the optical flow model and the robustness and functionality of the computational architecture and its implementation.

**Key Words:** recurrent network, neuromorphic, constraint satisfaction, regularization, parallel computation

### 1. Motivation

The ability to estimate motion using visual information is important for any natural and artificial agent behaving in a dynamical visual environment. Knowing the relative motions between different objects as well as between objects and the agent is crucial for a cognitive perception of the environment and thus a prerequisite for intelligent behavior. However, the demand for real-time processing and the limited resources available on freely behaving agents impose severe constraints that require *efficient* computational systems in terms of processing speed, energy consumption, and physical dimensions. These requirements favor parallel computational architectures. Implementations of such architectures become particularly appealing when image sensing and motion estimation circuitry can be combined within a single sensor. Furthermore, a sensor architecture that consists of a topographically uniform array of identical processing units (pixels) has the advantage that processing power scales with array size, thus keeping processing speed independent of spatial array resolution. Analog VLSI circuits are particularly suited for such architectures because they require significantly less power and silicon area than digital circuits for computational tasks of comparable complexity [1]. Furthermore, time-continuous analog processing matches the continuous

nature of visual motion information. Temporal aliasing artifacts do not occur while they can be a significant problem in clocked, sequential circuit implementations, in particular when limited to low frame-rates [2].

Visual information is – in general – locally ambiguous and noisy, and often does not allow a unique estimation of visual motion. A prominent ambiguity, commonly referred to as the “aperture problem” [3], is illustrated in Fig. 1a for an idealized motion scene under noise-free conditions. Observations through apertures showing zero-order (aperture A) or first-order spatiotemporal brightness patterns (apertures B and C) do not permit an unambiguous local estimate of the visual motion because an infinite number of possible flow vectors can explain the local percept. A model for visual motion is needed in order to resolve the ambiguities. *A priori* assumptions about the expected visual motion are part of such model. For example, a normal flow model assumes that from all possible optical flow vectors the shortest one is correct, which is the one that is perpendicular to the edge orientation in each aperture. Clearly, the computational task and the resulting quality of the optical flow estimate can vary substantially depending on the complexity of the chosen model.

Tracking distinct local brightness patterns as e.g. the corner of the triangle in aperture D, would allow the

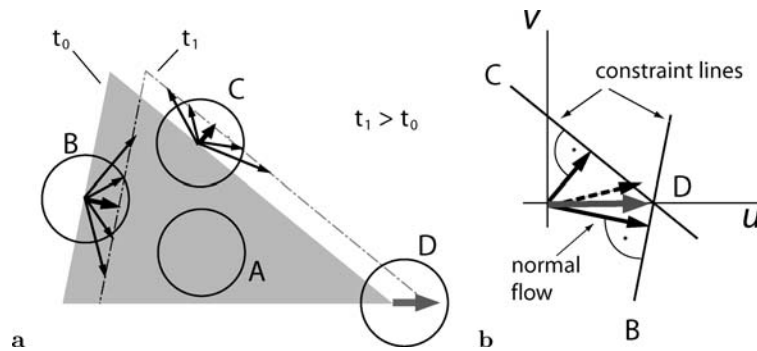


Fig. 1. The aperture problem. (a) Visual motion observed through the apertures A, B, and C is ambiguous with respect to the motion of the moving object. (b) Combining the local observations through aperture B and C permits to determine the object motion (intersection of constraints). Vector averaging of the normal flow field does not lead to the correct global motion (dashed arrow).

local estimation of visual motion without further assumptions, however requiring complex spatiotemporal filter stages to detect all possible patterns (feature detectors). Clearly, this is only possible at sparse locations. Alternatively, visual information can be spatially integrated over multiple apertures. Combining the local constraints from each aperture then ideally leads to a unique and well-defined estimate of object motion (see Fig. 1(b))<sup>1</sup>. It is important to realize that the vector average of normal flow usually does not coincide with this collective *intersection-of-constraints* (IOC) estimate. The IOC solution assumes that the motion signals to be integrated arise from a common visual motion source (object).

## 2. Review of Analog VLSI Visual Motion Sensors

Most of the known aVLSI motion sensors estimate visual motion only along a given single spatial orientation and/or they do not attempt to estimate visual speed explicitly. They can be classified in methods performing *explicit matching* in the time-domain e.g. [4–9], *gradient based* methods e.g. [10–12] and *implicitly matching* or *correlation based* methods that follow insect vision e.g. [13–18].

Only a few 2-D visual motion sensors have been reported that attempt to estimate visual speed. A first class of implementations estimates *normal flow*. Jiang and Wu [19] reported a correlation based approach. Motion is reported if the time-of-travel of the extracted edges between neighbor motion units matches a pre-set delay time. Since the tuning is very narrow, the sensor is fairly limited and cannot report arbitrary visual velocities without being continuously re-tuned. More practical sensors were reported by Kramer et al. [20], applying explicit matching in the time-domain. Two circuits are presented where the time-of-travel of a brightness edge is measured either by eliciting a monotonic decaying function with arrival of an edge and sampling

the functions value at the time the edge passes a neighboring pixel, or, by measuring the amount of overlap of two fixed-size pulses of neighboring pixels that each are triggered by the arrival of a brightness edge. A similar approach but different implementation was proposed by Etienne-Cummings et al. [21]. Where as in [20] temporal intensity changes are assumed to represent brightness edges, here, brightness edges are first extracted in the spatial domain before matching was performed in the time-domain. This has the advantage that also very slow speeds can be detected accurately. Mehta and Etienne-Cummings presented a gradient-based normal flow sensor [22]. In their implementation, the extraction of the spatiotemporal gradients and the normal flow estimation were performed in a line-sequential manner. The combination of an active pixel array with correlated double sampling and off-array computation permits high resolution and low noise estimation results, but it does not have the scaling property of a topographically homogeneous array processor.

None of the above approaches, however, performs *spatial integration* in order to incorporate more elaborate models of optical flow estimation. Tanner and Mead [23] described early a sensor with an array size of  $8 \times 8$  pixels that provides a global motion estimate. Measured motion data, however, was never explicitly shown and the sensor was reported to be fragile, even under well-controlled laboratory conditions. Nevertheless, it was the first hardware example of a collective computational approach to visual motion estimation. Subsequent attempts [24] failed to result in a more robust implementation. In a previous paper [25], we presented a first improvement that allowed the estimation of smooth optical flow besides a global motion estimate. The prototype implementation with a  $7 \times 7$  array was functional although it was rather limited by its small linear output range. A recent attempt by Lei and Chiueh [26] modified our previous approach [25] by incorporating segmentation properties but was not able

to demonstrate robust behavior under realistic illumination conditions.

The 2-D optical flow sensor described in this paper represents a significant improvement and further development of previous approaches. It provides a locally smooth optical flow field where the degree of smoothing is adjustable. Its increased linear range and a reasonable prototype array size encourages practical applications in robotics or surveillance. The sensor is a successful example of how to apply collective computation to aVLSI circuits in order to solve a perceptual task. We have previously presented other aVLSI implementations that extend the optical flow sensor presented here, and which permit the dynamical control of the local network connections in order to guide the process of motion integration [27–30]. To clarify the chronological order of development, however, note that all these implementations are based on the design of the optical flow sensor presented here, despite the fact that they have been published earlier. In fact, they all use the exact same schematics of the core motion pixel (later shown in Fig. 3).

### 3. Optical Flow Model

For analytical reasons, the input of the proposed model is defined as the spatiotemporal gradients  $E_x = \frac{\partial}{\partial x} E(x, y, t)$ ,  $E_y = \frac{\partial}{\partial y} E(x, y, t)$ , and  $E_t = \frac{\partial}{\partial t} E(x, y, t)$  of the image brightness distribution  $E(x, y, t)$ . Note, that image brightness depends on the particular characteristics of the imager, and is not necessarily proportional to the irradiance on chip. The extraction of these spatiotemporal gradients in the actual aVLSI implementation will be discussed later. The model's output is the optical flow field  $v(x, y, t) = (u(x, y, t), v(x, y, t))$ , which represents the instantaneous estimate of visual motion. To increase readability, space and time dependence will not be explicitly expressed in subsequent annotations.

The applied motion model assumes that the *brightness constancy constraint* [30] holds and that the optical flow varies smoothly in space. Following Horn and Schunck [31], these two constraints can be formulated as optimization problem for which the desired optical flow estimate is the optimal solution. It can be easily verified that using only these two constraints results in an ill-posed optimization problem for particular visual input patterns. Namely, when only zero- and first-order brightness patterns of equal orientation are present throughout the whole image (see Fig. 1). However, it is highly desirable for a physical system that the implemented function is well-posed under all possible input conditions. Otherwise, the system's behavior is unpredictable, driven by noise and non-idealities of the implementation. To avoid this, a

*bias constraint* is added, expressed as the convex cost function

$$B(v) = (u - u_{\text{ref}})^2 + (v - v_{\text{ref}})^2, \quad (1)$$

which measures the distance of the optical flow estimate to some predefined reference motion vector  $(u_{\text{ref}}, v_{\text{ref}})$ . The reference motion can be understood as the *a priori* expected motion in case the visual information content is unreliable or missing. There is evidence that the human visual system is applying a similar strategy [32]. In contrast to previous formulations [33, 34, 32], the reference motion is not necessarily assumed to be zero. For example, it could be adapted to the statistics of its visual environment; the reference motion could represent e.g. the statistical mean of the experienced visual motion. Such adaptation mechanisms, however, will not be further discussed in this article.

Combining the model of Horn and Schunck with the additional bias constraint (1) leads to an optical flow model that can be formulated as the following constraint optimization problem: Given the input  $E_{x_{ij}}$ ,  $E_{y_{ij}}$  and  $E_{t_{ij}}$  (on nodes  $i, j$  in a discrete, orthogonal  $n \times m$  image space) and a reference motion  $v_{\text{ref}} = (u_{\text{ref}}, v_{\text{ref}})$ , find the optical flow field  $v_{ij}$  such that the cost function

$$H(v_{ij}; \rho, \sigma) = \sum_{i=1}^n \sum_{j=1}^m [(E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}})^2 + \rho((\Delta u_{ij})^2 + (\Delta v_{ij})^2) + \sigma((u_{ij} - u_{\text{ref}})^2 + (v_{ij} - v_{\text{ref}})^2)] \quad (2)$$

is minimal.<sup>2</sup> The positive parameters  $\sigma > 0$  and  $\rho \geq 0$  determine the relative influence of each constraint. Boundary conditions are given by assuming the gradients of the flow field to be zero along the array boundary.

### 4. Optical Flow Network

The cost function (2) is convex for any given input. Thus, a linear dynamical system that performs *gradient descent* on the cost function is guaranteed to find the optimal solution. Gradient descent ( $\dot{u} \propto -\partial H/\partial u$ ,  $\dot{v} \propto -\partial H/\partial v$ ) results in a system of  $2n \times m$  linear differential equations

$$\begin{aligned} \dot{u}_{ij} &= -\frac{1}{C} [E_{x_{ij}}(E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}}) \\ &\quad - \rho(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{ij}) \\ &\quad + \sigma(u_{ij} - u_{\text{ref}})] \\ \dot{v}_{ij} &= -\frac{1}{C} [E_{y_{ij}}(E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}}) \\ &\quad - \rho(v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{ij}) \\ &\quad + \sigma(v_{ij} - v_{\text{ref}})] \end{aligned} \quad (3)$$

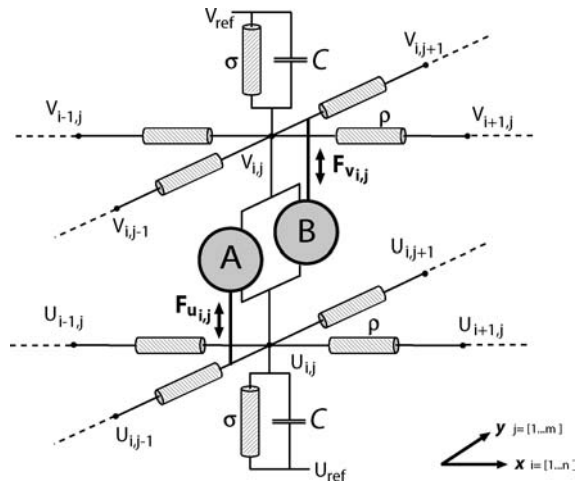


Fig. 2. A single unit of the optical flow network.

with  $C$  being a positive constant. These equations can be mapped to the voltage dynamics of an electronic network with two cross-coupled resistive layers. Figure 2 illustrates a single unit of such a network. Identifying the local estimate of optical flow with the voltages  $U_{ij}$ ,  $V_{ij}$  (with respect to some virtual ground  $V_0$ ), each term in straight brackets in (3) represents the sum of currents that charges up or down the capacitances  $C$  until equilibrium is reached. Each constraint of the cost function 2 has a physical counterpart:

- Smoothness is enforced by the two resistive layers with lateral conductances  $\rho$ .
- The bias constraint is implemented as the leak conductance  $\sigma$  to the reference motion represented by the potentials  $V_{\text{ref}}$  and  $U_{\text{ref}}$ .
- The brightness constancy constraint is enforced by “constraint-boxes”  $A$  and  $B$  that inject or sink the currents

$$\begin{aligned} F_{u_{i,j}} &\propto -E_{x_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}) \quad \text{and} \\ F_{v_{i,j}} &\propto -E_{y_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}), \end{aligned} \quad (4)$$

respectively [35]. These correction currents represent how much a potential optical flow estimate violates the brightness constancy constraint, and are computed in cross-coupled feedback loops.

Exploiting the natural dynamics of a physical system to solve an optimization problem is closely related to the principle of Hopfield networks [36]. The solution to the problem is represented by a stable attractor of the analog electronic network [37]. In contrast to Hopfield networks, the system has a single asymptotically stable attractor that, however, changes with changing visual input. We can assume that the state of the network closely follows the attractor if the network time-constant is negligible compared to the dynamics of the input. The appropriate control of the node capacitance

$C$  and the current levels in the implementation can ensure a close-to-optimal solution for reasonable input dynamics.

The characteristics of the model and therefore the computational behavior of the optical flow network are determined by the relative weights of the three constraints, which are defined by the lateral and vertical network conductances  $\rho$  and  $\sigma$ , respectively. According to the strength of these adjustable conductances, the network accounts for different models of visual motion estimation such as *normal flow*, *smooth optical flow* or *global flow*.

## 5. Circuit Architecture and Implementation

Figure 3 shows the complete circuit schematics of a single motion unit (pixel) of the optical flow sensor. All the transistors are operated in the sub-threshold regime, except those in the core of the multiplier circuits computing  $E_x u$  and  $E_y v$  (see Section 5.2). All variables are differentially encoded because the input (spatiotemporal brightness gradients) and the output (the components of the local optical flow vector) can take on positive and negative values. Referencing the voltages  $U_+$  and  $V_+$  to the fixed virtual ground  $V_0$  substantially reduces the implementation complexity by requiring only two single-line resistive layers. Note, that in this particular implementation, the reference motion for the bias constraint was chosen to be zero, thus  $U_{\text{ref}}, V_{\text{ref}} = V_0$ . However, an independent control of both potentials could be easily implemented.

### 5.1. Extraction of the Brightness Gradients

The first stage of processing consists of estimating the spatiotemporal brightness gradients. Each pixel includes a logarithmic, adaptive photoreceptor [38] with adjustable adaptation rate [39].

A hysteretic differentiator circuit [40] provides the currents  $E_{t-}$ ,  $E_{t+}$  that represent the rectified temporal derivative of the photoreceptor signal  $E_0$ . The adjustable source potentials  $\text{HD}_{\text{tweak}+}$ ,  $\text{HD}_{\text{tweak}-}$  allow to control the output current gain of the differentiator.

The spatial derivatives  $E_x$  and  $E_y$  are estimated as first-order approximations of the brightness gradients, thus the central differences between the photoreceptor outputs of nearest neighbors in each orthogonal direction. Because the photoreceptor encodes irradiance logarithmically, the spatial derivatives become independent of absolute illumination and basically represent the irradiance contrast between two neighboring image locations.<sup>3</sup> This property is very useful because it implies that the sensor ideally provides the same response for the same stimulus, independently of the lighting conditions.

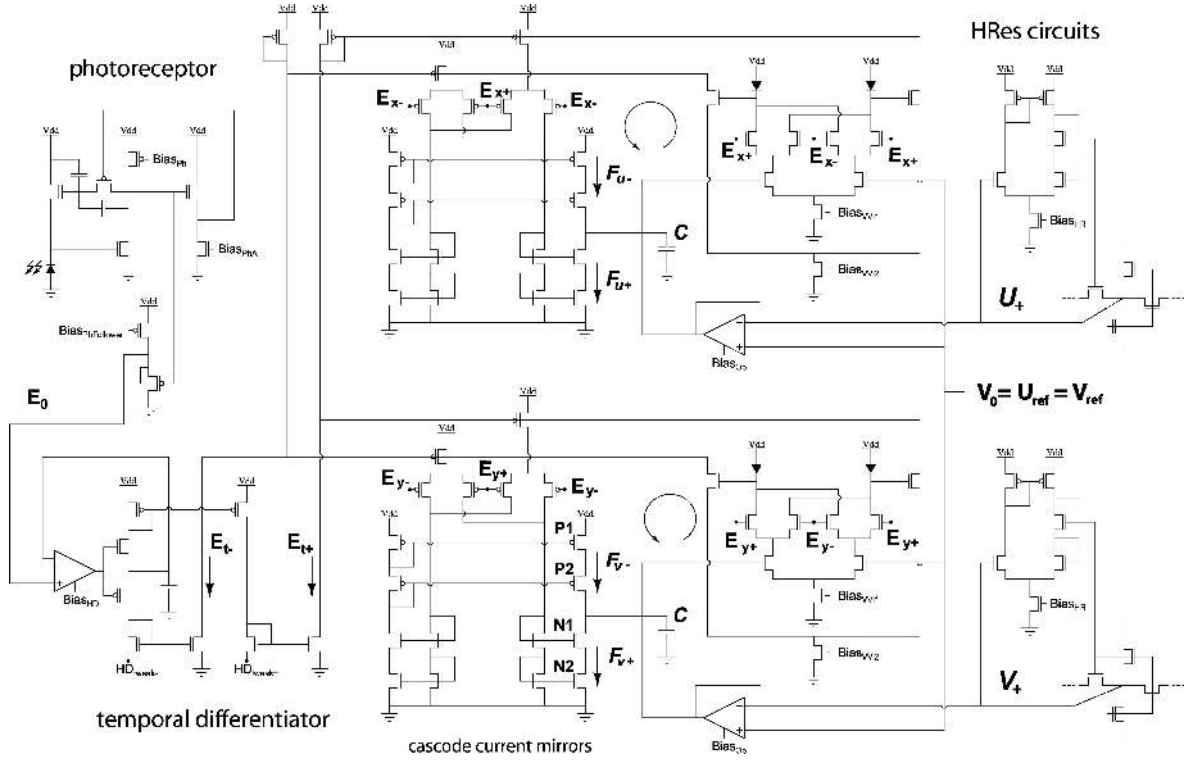
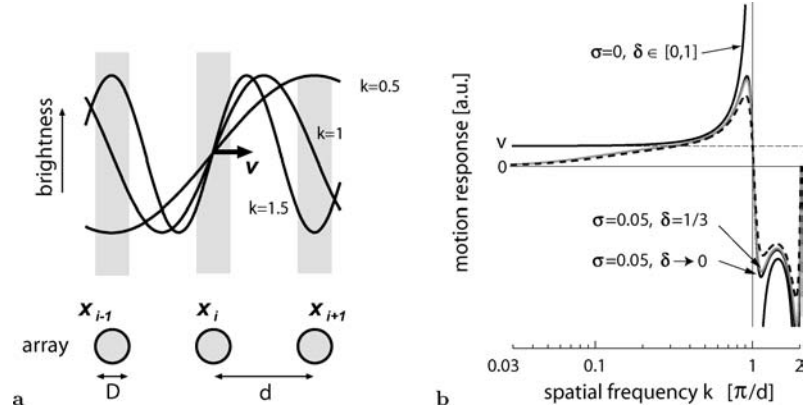


Fig. 3. Circuit schematics of a single motion unit.


 Fig. 4. Spatial sampling and its effect on the motion response. (a) Spatial sampling of sinusoidal brightness patterns of different spatial frequencies  $k$  (given in units of the Nyquist frequency  $[\pi/d]$ ) in one visual dimension. (b) The expected response of the optical flow sensor according to (8) as a function of the spatial frequency  $k$  of a sinewave stimulus moving with velocity  $v$  and the fill-factor  $\delta = D/d$ . The dashed curve is the time-averaged response ( $\sigma = 0.05$ ,  $\delta = 1/3$ ).

While continuous-time temporal differentiation avoids any temporal sampling artifacts, discrete spatial sampling of the image brightness affects the visual motion estimate. This is illustrated in Fig. 4(a): Consider a simplified, one-dimensional optical flow network that is presented with a sine-wave grating stimulus moving with a fixed velocity  $v$ . The brightness distribution in the focal-plane is then given as  $E(x, t) = \sin(kx - \omega t)$  with  $x, t \in \mathbb{R}$ , where  $\omega = kv$  and  $k$  is the spatial frequency of the projected stimulus. The spatial gradient at location  $x_i$  is the central difference operator  $\Delta x_i = (x_{i+1} - x_{i-1})/2d$ , where  $d$  is the sampling dis-

tance (size of pixel). Assuming a non-zero sampling size  $D$  (size of photodiode), the discrete spatial gradient becomes

$$\begin{aligned}
 E_x(x_i, t) &= \frac{1}{2d} \left( \frac{1}{D} \int_{x_i+d-D/2}^{x_i+d+D/2} \sin(k(\xi - vt)) d\xi \right. \\
 &\quad \left. - \frac{1}{D} \int_{x_i-d-D/2}^{x_i-d+D/2} \sin(k(\xi - vt)) d\xi \right) \\
 &= \frac{2}{kDd} \sin(kd) \sin(kD/2) \cos(k(x_i - vt)).
 \end{aligned} \tag{5}$$

Temporal differentiation is continuous, thus the temporal gradient at location  $x_i$  simply is

$$\begin{aligned} E_t(x_i, t) &= \frac{\partial}{\partial t} \frac{1}{D} \int_{x_i-D/2}^{x_i+D/2} \sin(k(\xi - vt)) d\xi \\ &= -\frac{2v}{D} (\cos(k(x_i - vt)) \cdot \sin(kD/2)). \end{aligned} \quad (6)$$

According to (3), the response of a single unit in equilibrium with no lateral coupling ( $\rho = 0$ ) approximates normal flow and reduces to

$$V_{\text{out}}(x_i, t) = -\frac{E_t(x_i, t)E_x(x_i, t) + V_{\text{ref}}\sigma}{\sigma + E_x(x_i, t)^2} \quad (7)$$

Substitution of (6) and (5) in (7) and assuming  $V_{\text{ref}} = 0$  leads to

$$\begin{aligned} V_{\text{out}}(x_i, t) &= v \frac{kd}{\sin(kd)} \cdot \frac{\cos(kx_i - \omega t)^2}{\gamma + \cos(kx_i - \omega t)^2} \quad \text{with} \\ \gamma &= \sigma \frac{k^2 d^2 D^2}{4 \sin^2(kd) \sin^2(kD/2)}. \end{aligned} \quad (8)$$

Figure 4(b) shows the predicted peak output response (8) for different weights of the bias constraint  $\sigma$  and different fill-factors  $\delta = D/d$ , as a function of the spatial stimulus frequency  $k$ , given in units of the Nyquist frequency  $[\pi/d]$ . Note that the response strongly depends on  $k$ . For low spatial frequencies, the motion output well approximates the correct velocity. For very low frequencies, the local stimulus contrast diminishes and the non-zero  $\sigma$  biases the output response toward the reference motion  $V_{\text{ref}} = 0$ . As the stimulus frequency approaches the Nyquist frequency, the response increases to unrealistically high values and changes sign at  $k \in \mathbb{Z}^+[\pi/d]$ . Increasing the fill factor reduces the effect, although only marginally for reasonable values of  $\delta$  and small  $\sigma$ . Clearly, without the bias constraint ( $\sigma = 0$ ), Eq. (8) is not defined at spatial frequencies  $k \in \mathbb{Z}^+$ , and the motion response approaches  $\pm\infty$  in their vicinity.

## 5.2. Wide-Linear Range Multiplier

The design of the multiplier circuit computing  $E_x u$  and  $E_y v$  respectively, is crucial for a successful aVLSI implementation of the optical flow network. It has to operate in four quadrants, providing a wide linear output-range with respect to each multiplicand. Offsets should be minimal because they directly impose offsets in the optical flow estimate. Furthermore, the design needs to be compact in order to guarantee a small pixel size. The original Gilbert multiplier circuit [41] meets these requirements fairly well, with the exception of its small linear range when operated in sub-threshold. In above-threshold operation, however, its linear range is signif-

icantly increased due to a transconductance change of the individual differential pairs [27].

The proposed multiplier circuit (shown in detail in Fig. 5(a)) embeds a Gilbert multiplier in an outer differential pair. The idea is to operate the Gilbert multiplier above-threshold to increase the linear range but in addition, to re-scale the output currents such that the current level in the feedback loop remains sub-threshold. The scaling is approximately linear, thus

$$I_{\text{out}} = I_{\text{out}+} - I_{\text{out}-} \approx \frac{I_{b2}}{I_{b1}} I_{\text{out}}^{\text{core}}, \quad (9)$$

if  $\kappa n \approx 1$ , where  $1/n$  is the ideality factor<sup>4</sup> of the diodes  $D_{1,2}$  and  $\kappa$  is the slope factor of the nFETs in the embedding differential pair. Using the base-emitter junction of bipolar transistors to implement the diodes ensures an ideal ( $n = 1$ ) behavior [42]. Also, the voltage drop across the diodes is such that the gate voltage of the outer nFETs are typically within one volt below  $V_{dd}$ , meaning that their gate-bulk potentials are large. At such levels,  $\kappa$  asymptotically approaches unity because the capacitance of the depletion layer becomes negligibly small compared to the gate-oxide capacitance. Thus,  $\kappa n \approx 1$  and we can safely assume a linear scaling of the multiplier output currents with bias current  $I_{b2}$ .

Base-emitter junctions can be exploited either using native bipolar transistors in a genuine BiCMOS process or the vertical bipolar transistors in standard CMOS technology. Although a CMOS implementation of the diodes is desirable to avoid the more complex and expensive BiCMOS process, it is necessary to use a genuine BiCMOS process. The reason is that at current levels above  $10^{-6}$  A the vertical bipolar starts to deviate significantly from the desired exponential characteristics due to high-level injection caused by the relative light doping of the base (well) [42]. These are current levels that are significantly below the range where the multiplier core is preferably operated at. The exponential regime of the native bipolar, however, extends up to  $10^{-3}$  A. Figure 5(b) displays the measured base-emitter voltages  $V_{BE}$  as a function of the applied emitter current  $I_E$  for both, a vertical pnp-bipolar transistor in a typical p-substrate CMOS process and a native npn-bipolar transistor in a genuine BiCMOS process.

The output characteristics of the multiplier for sweeping either of the two input voltages are shown in Figs. 5(c) and (d). For mild above-threshold bias currents, the linear range reaches approximately  $\pm 0.5$  V and is slightly smaller for the upper differential input. Note, that the measurements were obtained from a test circuit with identical layout to the one within each pixel. Offsets are small. The circuit is compact and offers the possibility to control the linear range and the output-current level independently by the two bias voltages  $V_{b1}$  and  $V_{b2}$ . The disadvantages are the increased power

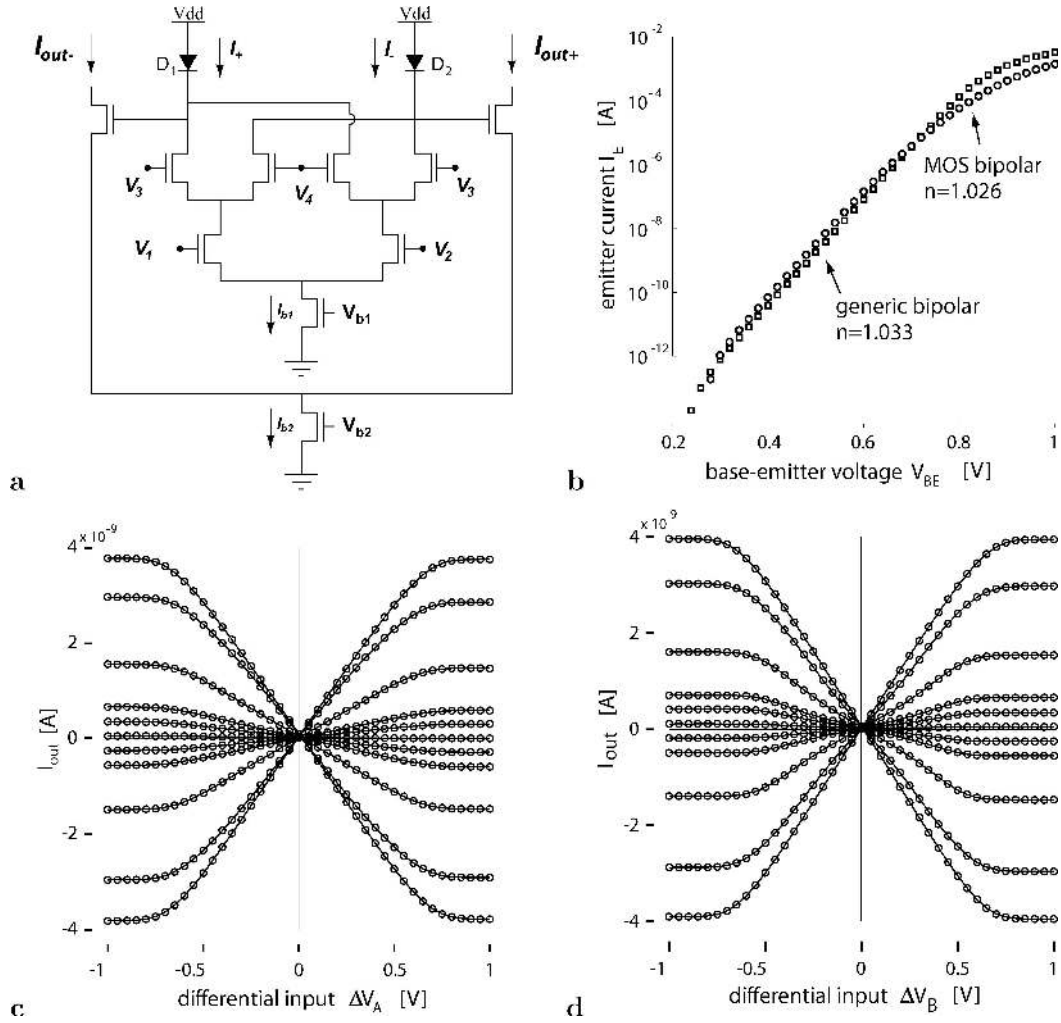


Fig. 5. Wide linear-range multiplier. (a) Circuit schematics. (b) The measured emitter currents for a native npn-bipolar transistor and the vertical pnp-bipolar of a pFET. (c) Measured output currents of the wide linear-range multiplier as a function of the applied input voltage at the lower differential pair  $\Delta V_A = V_2 - V_1$ . Each curve represents the output current for a fixed differential voltage  $\Delta V_B = V_4 - V_3 = [0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.75 \text{ V}]$ . (d) Same as (c) but now sweeping the upper differential pair input  $\Delta V_B$ .

consumption caused by the above-threshold operation and the need for BiCMOS process.

### 5.3. Output Conductance at the Feedback Node

Consider the current equilibrium at one of the capacitive nodes in Fig. 2. In steady state, all currents flowing onto this node represent the violations of the different constraints according to (3). To be completely true, this would require the current sources  $A$  and  $B$  to be ideal.

However, in the aVLSI implementation a small but present output conductance remains at these feedback nodes. This conductance can be understood as imposing a second bias constraint on the estimation problem, causing some extra current that shifts the current equilibrium. It biases the capacitive node to a reference voltage that is intrinsic and depends on various parameters like the strength of the feedback current or

the Early voltages of the transistors. In general, this reference voltage is not identical with  $V_{\text{ref}}$  and cannot be properly controlled. Thus, the superposition of the two bias currents can have an asymmetric effect on the final motion estimate. Since the total correction currents are typically weak, the effect is significant.

The aim, therefore, is to reduce the output conductance at the feedback nodes as much as possible. The applied cascode current mirror circuits lead to a substantial decrease in output conductance compared with simple current mirrors used in previous related implementations [23–25].

Neglecting any junction leakages in drain and source, we find the total output conductance to be

$$g_o = g_{oN} + g_{oP} = \frac{F_{x+} \cdot U_T}{V_{E,N1} V_{E,N2}} + \frac{F_{x-} \cdot U_T}{V_{E,P1} V_{E,P2}}, \quad (10)$$

where  $F_{x+} - F_{x-}$  is the total feedback current of either one of the two feedback loops,  $U_T$  the thermal voltage and  $V_{E,x}$  the Early voltages of the different transistors in the cascode current mirror (see transistor labels in Fig. 3).

#### 5.4. Effect of Non-Linearities

A linear translation of the network architecture (Fig. 2) into silicon is hardly possible. In the following, some important non-linearities of the implementation and their effect on the optical flow estimate are discussed.

**5.4.1. Saturation in the Feedback Loop** The first important non-linearity is caused by the saturation of the multiplier circuit in the feedback loop. As shown in Fig. 5, the output current saturates for larger input voltages. What does this mean in terms of the expected motion output of a single motion unit?

A single motion unit ideally satisfies the brightness constancy constraint

$$E_x u + E_t = 0 \quad (11)$$

for a given non-zero spatial brightness gradient  $E_x$ . Now, assume that the multiplication  $E_x u$  is replaced with a non-linear function  $f(u)|_{E_x}$  with  $E_x$  constant, where  $f$  describes the saturating output characteristics of the proposed multiplier circuit. For the sake of simplicity let us assume that the output of the multiplier core ( $I_+ - I_-$ ; see Fig. 5(a)) shows the same saturation characteristics as if it were operated in sub-threshold, which is only qualitatively correct because the characteristics change above-threshold [27]. Then, the multiplier output (9) can be rewritten as  $f(u)|_{E_x} = c_{E_x} I_{b2}/I_{b1} \tanh(u)$ , where the constant  $c_{E_x}$  is proportional to a given  $E_x$ . Since  $f(u)$  is one to one, solving (11) for the motion response  $u$  leads to

$$u = f^{-1}(-E_t) = -\operatorname{atanh}\left(\frac{I_{b1}}{I_{b2}} E_t c_{E_x}^{-1}\right). \quad (12)$$

Figure 6 illustrates the expected motion response where each curve is for a different value of the bias current  $I_{b2}$ . The motion output increasingly over-estimates the stimulus velocity for increasing speed values. The more the output of the multiplier circuit saturates and is below the true multiplication result, the more the feedback loop compensates by increasing  $u$  in order to make the equilibrium (11) hold. Increasing  $I_{b2}$  decreases the slope of the motion response. Thus, the bias current of the outer differential pair acts as gain control that permits to match the limited linear output range to the expected maximal motion range.

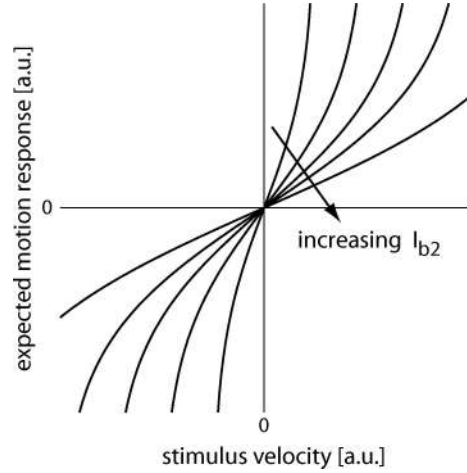


Fig. 6. Expected motion response due to saturation of the multiplier circuit. The figure illustrates the expected speed tuning curves according to (12). The different curves correspond to different values of the bias current  $I_{b2}$ .

**5.4.2. Non-Linear Bias Conductance** The quadratic measure of the bias constraint in the cost function (2) results in an ohmic bias current that is proportional to the constant conductance  $\sigma$ . Sufficiently low and compact ohmic conductances, however, are difficult to implement in CMOS technology. The bias conductance of the optical flow unit is implemented using the transconductance  $g_m$  of a simple amplifier circuit (see Fig. 3). Its characteristics are ideally given as

$$I(\Delta V) = I_b \tan h\left(\frac{\kappa}{2U_T} \Delta V\right), \quad (13)$$

with  $I_b$  being the bias current controlled by the voltage Bias<sub>OP</sub>,  $U_T$  the thermal voltage and  $\kappa$  the common slope factor. Figure 7(a) illustrates the output of a saturating resistor according to (13). The transconductance  $g_m = \frac{I_b \kappa}{2U_T}$  characterizes the small ohmic regime ( $<100$  mV) and the bias current  $I_b$  determines the saturation current level of the circuit. Changing the resistor from an ohmic to a saturating behavior changes the bias constraint (1) from a quadratic toward an absolute-value function, precisely given as the electrical co-content (Fig. 7(b)). From a computational point of view, this implies that the optical flow estimate is no longer penalized proportionally to the amplitude of its components. Rather, beyond the small ohmic regime, a constant bias current is applied. Therefore, the bias constraint rather has a *subtractive* than a *divisive* effect on the optical flow estimate. As discussed earlier, it seems reasonable to strengthen the bias constraint locally according to the degree of ambiguity. Only, the degree of ambiguity is not reflected by the amplitude of the motion estimate. Thus, it seems most sensible to penalized the motion estimate independently of its amplitude, which actually



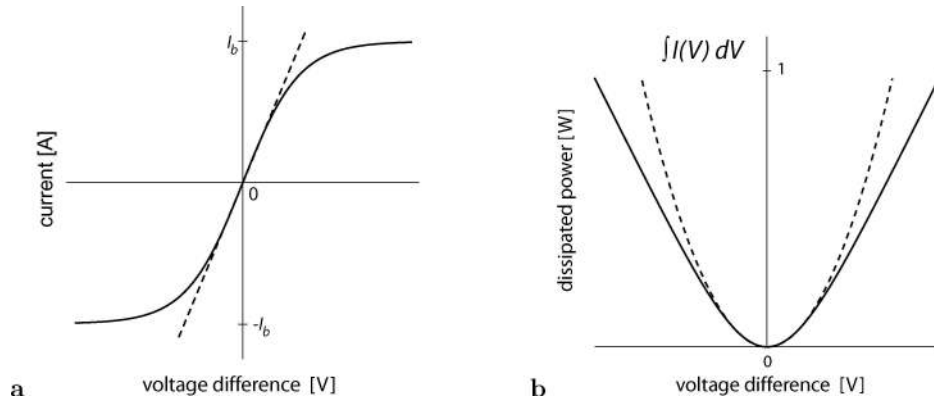


Fig. 7. Ohmic versus saturating resistor: (a) The characteristic curves for an ohmic (dashed line) and a saturating resistor. (b) The electrical co-contents (integrals).

means to apply a constant bias current. Just recently, we have shown that the human visual system uses a bias that penalizes high speeds even less [43]. As presented later, the non-linear bias conductance improves the output behavior of the optical flow sensor for low contrast and low spatial frequency stimuli compared to a quadratic formulation of the bias constraint. Global asymptotic stability of the optical flow network is still guaranteed because the cost function remains convex.

**5.4.3. Non-Linear Smoothing of the Optical Flow** The smoothness constraint is enforced by HRes circuits [40], which have characteristics equivalent to (13). The saturation current level is controlled by the bias voltage  $Bias_{HR}$ . Similarly to the bias constraint, the smoothness constraint turns from a quadratic (for small voltage differences) to an absolute-value cost function. Saturating resistors privilege smoothing across small voltage differences, thus between image locations that show only little differences in the components of their optical flow vectors. If the difference becomes too big, the smoothing conductance is continuously reduced in proportion to the local gradient of the optical flow components. Consequently, smoothing is reduced at boundaries of different motion sources where the optical flow is likely to differ substantially, which partially preserves motion discontinuities. Because the non-linear smoothing is independent in each component of the optical flow estimate, smoothing is not rotationally invariant but rather slightly increased for visual motion along the diagonal of the intrinsic coordinate frame of the array. In contrast to resistive fuse circuits [44], non-linear smoothing conductances preserve the convexity of the cost function (2), thus ensure global asymptotic stability.

## 6. Sensor Measurements

The optical flow sensor has been fully characterized using two array implementations. Both share the same technology and an identical pixel design, and they only

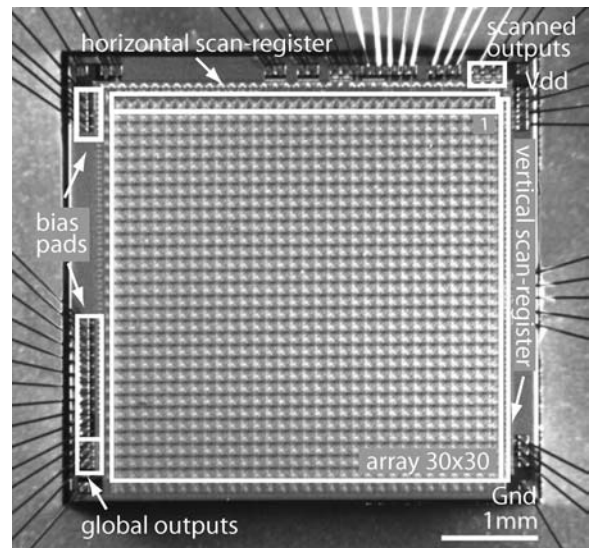


Fig. 8. Micro-photograph of the optical flow sensor. Scanning circuitry allows to read-out the optical flow field as well as the photoreceptor signals. The global optical flow estimate is directly accessible (scanning circuitry de-activated) at separate output pads when the smoothness conductance is set high ( $Bias_{HR} > 0.8$  V).

differ in their array sizes. Figure 8 shows a micro-photograph of the larger array. Speed-tuning, contrast and spatial frequency dependence are based on measurements from the smaller array while the flow fields are recorded from the larger array. Specifications are summarized in Table 1.

Stimuli were visual and directly projected onto the sensor via an optical lens system. They were either generated electronically and displayed on a computer screen, or were physically moving objects (e.g. for high temporal frequencies). Using a dark-room test environment including an optical bench permitted the precise control over all relevant parameters. On-chip irradiance during all experiments was within one order of magnitude with a minimal value of  $9 \text{ mW/m}^2$  for low contrast computer screen displays. At such low values, the rise time of the photoreceptor circuit is in the order of a few

Table 1. Sensor specifications.

<i>Chip:</i>	
Technology	AMS 0.8 $\mu\text{m}$ BiCMOS, 2M, 2P
Threshold voltage 20/20	0.8 (−0.84) V
Supply voltage	5 V
Die size(s)	$2.55 \times 2.55 \text{ mm}^2/4.72 \times 4.5 \text{ mm}^2$
Array size(s) (motion units)	$10 \times 10/30 \times 30$
<i>Pixel:</i>	
Size	$124 \times 124 \mu\text{m}^2$
Active elements	119
Power dissipation	52 $\mu\text{W}$
Outputs (scanned)	optical flow, photoreceptor

milliseconds [45]. This is sufficiently fast with respect to the applied stimulus speeds in those particular measurements. Unless indicated differently, all circuits were biased in the sub-threshold regime. The shown optical flow signals are always referenced to the virtual ground potential  $V_0$ , which was typically set around 2 V.

### 6.1. Response Characteristics

Moving sinewave and squarewave gratings were applied to characterize the motion response for varying speeds, contrasts and spatial frequencies. Orientation tuning and the ability to find the intersection-of-constraints solution were also tested. The presented data in this section constitute the global motion signal, thus the unique, collectively computed solution provided by all units of the array. Unless indicated differently, each data point represents the mean value of 20 measurements, each being the average output over one cycle of a grating. Standard deviations are only shown where they exceed the size of the data points.

6.1.1. Speed Tuning Figure 9(a) shows the speed-tuning of the optical flow sensor for sinewave and squarewave gratings of equal contrast and fundamental

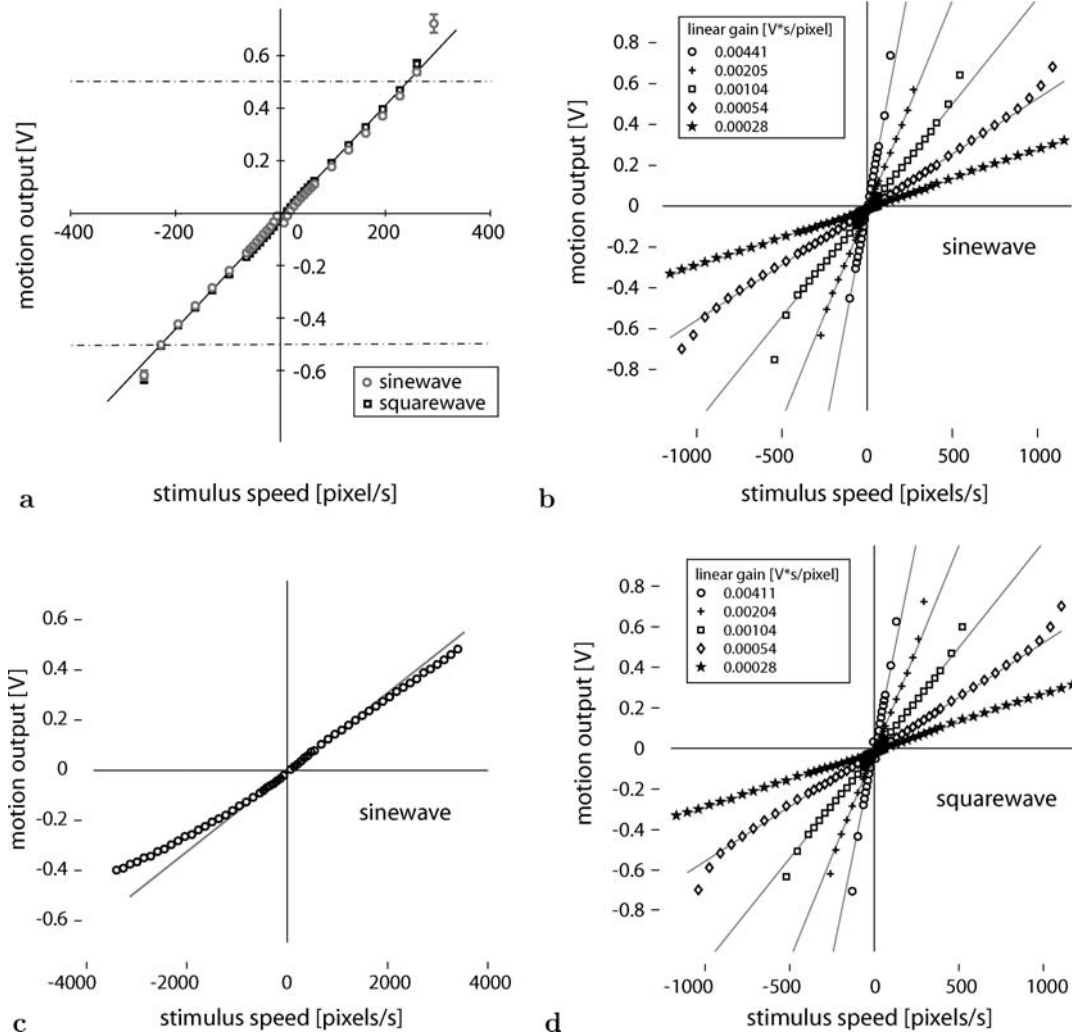


Fig. 9. Speed tuning of the optical flow sensor. (a) Speed tuning of the optical flow sensor for sinewave and squarewave gratings (spatial frequency 0.08 cycles/pixel, contrast 0.8). (b) Speed tuning for the sinewave grating for different values of the multiplier bias voltage ( $\text{Bias}_{V_{I2}} = [0.5, 0.53, 0.56, 0.59, 0.62 \text{ V}]$ ). Values in inlets indicate the slopes of the linear fits. (c) Output in the low gain limit ( $\text{Bias}_{V_{I2}} = 0.67 \text{ V}$ ). (d) Same as (b) but squarewave gratings.

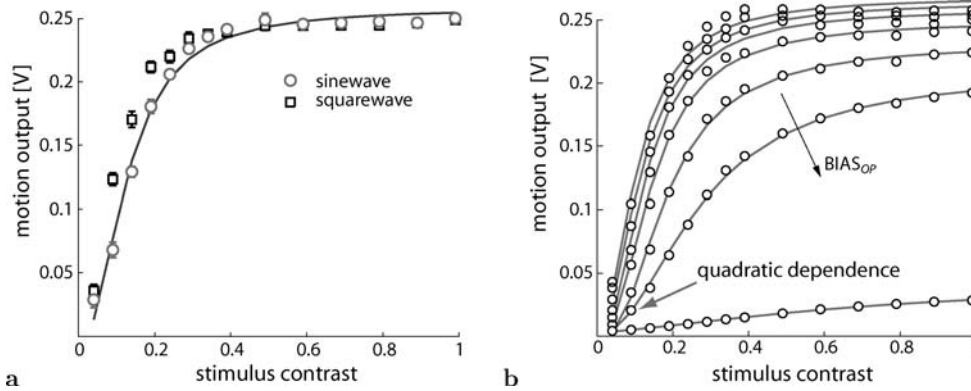


Fig. 10. Contrast dependence of the motion output. (a) The motion output was measured as a function of stimulus contrast for sinewave and squarewave gratings. (b) Increasing the bias voltage  $Bias_{OP} = [0.15, 0.23, 0.26, 0.29, 0.32, 0.35, 0.45 \text{ V}]$  leads to an increased contrast dependence of the motion output, here measured for sinewave grating stimuli.

spatial frequency. Negative speed values indicate motion in the opposite direction. The tuning is almost identical for both type of gratings. It clearly exhibits the qualitatively predicted behavior due to the saturation of the multipliers in the feedback loop (compare with Fig. 6). The tuning is linear in a range of approximately  $\pm 0.5 \text{ V}$  as indicated by the dash-dotted lines. Beyond the linear range, the output quickly increases/decreases and finally hits the voltage rails on either side. Due to the above-threshold operation of the wide linear-range multiplier core, the tuning reveals a more pronounced linear-range and a sharper transition to the super-linear behavior than predicted by the simplified qualitative analysis (12).

Figures 9(b) and (d) illustrate how the bias current of the wide linear-range multiplier affects the sensitivity of the optical flow sensor. Increasing the voltage  $Bias_{V12}$  leads to a smaller response gain and thus maps the linear output-range to a larger maximal speed range. The solid lines represent the linear fits to the individual measurements. The slope of the fits approximately scales by a factor of one half for a fixed increase in bias voltage  $Bias_{V12}$  of 30 mV (see measured gain values in inlets), and thus is inversely proportional to the bias current. Comparing the measured responses for sinewave (Fig. 9(b)) and squarewave gratings (Fig. 9(d)) does not reveal any significant differences. Setting  $Bias_{V12}$  high (0.67 volts) permits to measure speeds up to 5000 pixels/sec (Fig. 9(c)). However, the response becomes slightly asymmetric because the output conductances increase with increasing current densities in the feedback loop, and interfere with the explicit bias conductances of the circuit (10).

**6.1.2. Contrast Dependence** The output of the optical flow sensor depends continuously on stimulus contrast, measured as  $(E_1 - E_2)/(E_1 + E_2)$  where  $E_1$  represents

the brightest and  $E_2$  the darkest intensity level of the gratings. Figure 10(a) shows the contrast dependence of sinewave and squarewave gratings with identical spatial frequency (0.08 cycles/pixel) and constant speed (30 pixels/s).

Below a contrast value of about 0.3, the output signal continuously decreases towards the reference motion  $V_{ref}$  with decreasing contrast. The squarewave stimulus shows a slightly sustained resistance against this drop-off. A least-squares fit according to 7 was applied to the response to the sinewave grating. Although it well describes the overall response, the measured output curves rise faster and exhibit a more extended plateau towards lower contrast. This improved output behavior originates from the non-linear implementation of the bias conductance. The effective conductance decreases with increasing motion output signal while 7 assumes a constant  $\sigma$  that leads to a stronger bias at high contrast. Figure 10(b) shows the influence of the bias strength on the contrast tuning using the same sinewave grating stimuli as before. Increased values of the voltage  $Bias_{OP}$  lead to a contrast dependent output over the complete contrast range. For high values of  $Bias_{OP}$  the bias conductance dominates the denominator in 7 and the motion estimate becomes quadratically dependent on contrast. It reduces to the multiplication of the spatial and temporal brightness gradients, hence  $v \approx E_x E_t$ . In fact, such simple spatiotemporal gradient multiplication does not need the costly feedback architecture and can be implemented in a compact feed-forward way [10, 11]. However, the expected quadratic dependence can be observed only for very low contrasts because the pFET Gilbert multipliers computing the product of the spatial and temporal gradient quickly saturate for higher contrast levels.

**6.1.3. Spatial Frequency Tuning** The third stimulus parameter tested was spatial frequency. Figure 11(a)

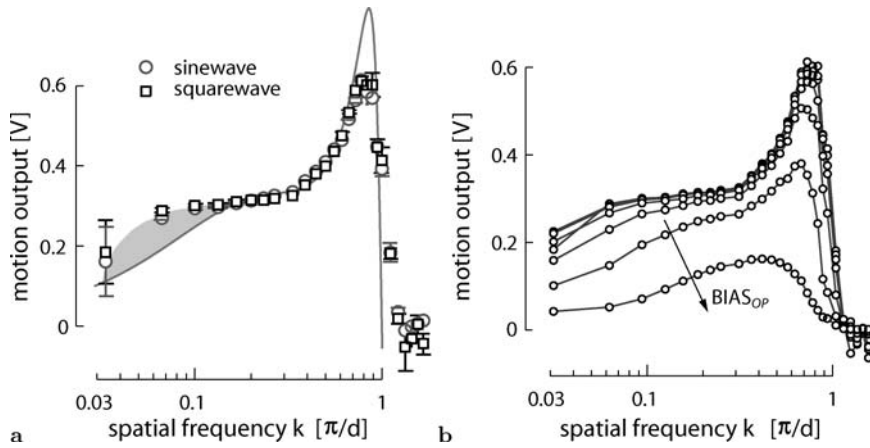


Fig. 11. *Spatial frequency tuning* (a) The spatial frequency response was measured for sinewave and squarewave gratings moving with constant speed. Spatial frequency  $k$  is given in units of the Nyquist frequency  $[\pi/d]$ . For frequencies  $k < 1$ , the response closely follows the expected response (8). The non-linear bias conductance extends the range for accurate estimation at low spatial frequencies (gray area). (b) Increasing the bias voltage  $Bias_{OP} = [0.1, 0.15, 0.2, 0.25, 0.30, 0.35, 0.40 \text{ V}]$  reduces the overall response amplitude and shifts the peak towards  $k = 0.5$ .

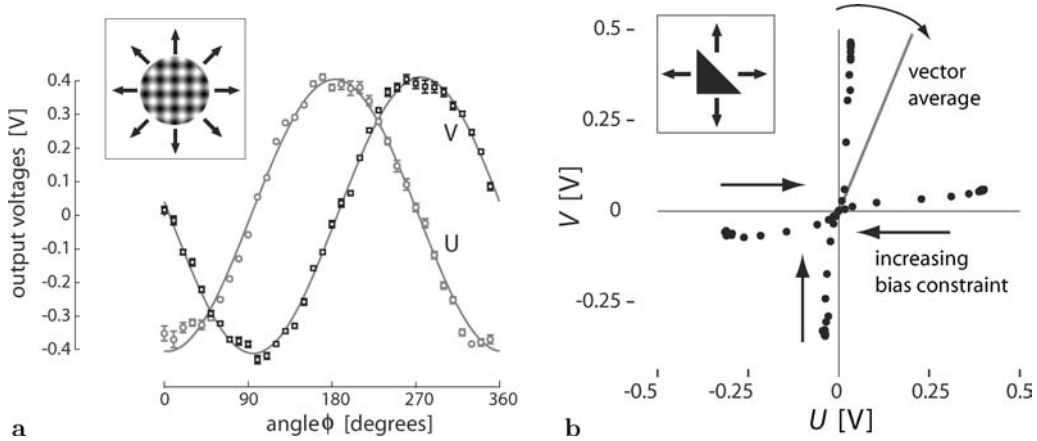


Fig. 12. *Solving the aperture problem*. (a) The optical flow sensor exhibits the expected cosine and sinewave tuning curves for the components  $U$  and  $V$ . The sinusoidal plaid stimulus (spatial frequency 0.8 cycles/pixel, contrast 0.8) was moving at a velocity of 30 pixels/s. (b) The output is also shown for a triangular object moving only in orthogonal directions. The optical flow sensor approximately solves the aperture problem. Data points along each trajectory correspond to particular bias voltages  $Bias_{OP} = [0, 0.1, 0.15, 0.2, 0.25, 0.3, 0.33, 0.36, 0.39, 0.42, 0.45, 0.48, 0.51, 0.75 \text{ V}]$ .

shows the motion output to sinewave and squarewave gratings (0.8 contrast, 30 pixels/s) as a function of spatial frequency. The waveform difference has only a minimal effect on the tuning curve. A least-squares fit according to (7) was performed to the response to the sinewave grating for frequencies  $k < 0.75$ . For spatial frequencies  $k < 0.06$ , the spatial brightness gradients on the chip surface become very small and thus the bias constraint dominates. Again, the non-ohmic implementation significantly improves the response. The range for correct motion estimation is extended down to frequencies as low as  $k = 0.06$ . The shaded area indicates the improvement compared to the fit, which assumes an ohmic conductance. As spatial frequencies exceed the Nyquist frequency ( $k > 1$ ), the responses rapidly drop below zero and remain small negative. Because the test-setup did not allow to project high spatial

frequency stimuli with sufficiently high contrast, the response beyond the Nyquist frequency could not be properly measured.

As shown in Figure 11(b), the strength of the bias voltage  $Bias_{OP}$  affects the amplitude of the motion output as well as the spatial frequency for which the motion response is maximal. Recall that increasing  $Bias_{OP}$  decreases the motion output the more, the smaller the local spatial gradient is. Because  $E_x$  is largest at  $k = i + 0.5$  where  $i \in \mathbb{Z}^+$  the spatial frequency for which the motion output is maximal shifts towards this value with increasing values of  $Bias_{OP}$  (see Fig. 4(a)).

**6.1.4. Spatial Integration** A key property of the optical flow sensor is its ability to accurately estimate two-dimensional visual motion. Figure 12(a) shows the orientation tuning of the optical flow sensor when

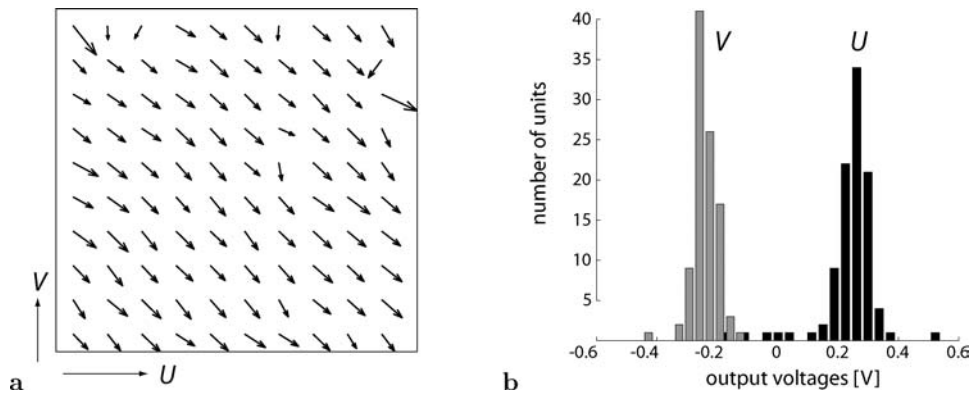


Fig. 13. Flow field offsets. (a) The measured time-averaged optical flow for uniform visual motion ( $\rho = 0$ ). (b) Histograms for both components of the optical flow vectors with mean values of  $\bar{U} = 0.25$  V and  $\bar{V} = -0.24$  V and standard deviations of 0.09 V and 0.04V, respectively.

presented with an orthogonal sinusoidal plaid stimulus. The output fits well the expected sine and cosine functions.

Also, the optical flow sensor approximately *solves the aperture problem* for a single motion source (single object moving on a non-structured background). To demonstrate this, the sensor was biased to compute global motion. A high contrast stimulus was presented, consisting of a dark triangle on a light background moving in either of the four orthogonal directions. Its particular object shape requires to compute the intersection-of-constraints solution in order to correctly estimate object motion. Figure 12(b) shows the applied stimulus and the global motion estimates of the sensor for a constant positive or negative object motion in the two orthogonal directions. Each data point represents the estimate for a particular value of the bias conductance. For  $\text{Bias}_{OP} < 0.25$  V, the optical flow sensor almost perfectly reports the true object motion in either of the four tested directions although with a small amplitude asymmetry. The remaining directional deviation of less than 10 degrees is the result of the output conductance at the feedback nodes. Note, that the vector average of normal flow would lead to a much larger deviation in direction (22.5 degrees) for this particular stimulus. As  $\text{Bias}_{OP}$  increases, the reported speed decreases rapidly while the direction of the global motion estimate shifts more and more toward the vector average.

**6.1.5. Array Offsets** To conclude the characterization of the optical flow sensor, the fabrication induced output variations between the individual units of the optical flow network were investigated. A diagonally oriented moving sinewave grating was applied such that both components ( $u$ ,  $v$ ) of the normal flow vector were equally large in amplitude. Each unit of the sensor was recorded in isolation with the smoothness constraint being completely disabled ( $\text{Bias}_{HR} = 0$  V). The stimulus

speed and the bias settings were chosen such that the output values were clearly within the linear range of the optical flow units.

Figure 13(a) shows the resulting time-averaged (20 stimulus cycles) optical flow field. It reveals some offsets between the optical flow vectors of the different units in the array. Units that exhibit the most variations seem to be located preferably at the array boundaries where device mismatch is usually pronounced due to asymmetries in the layouts. Because motion is computed component-wise, deviations due to mismatch are causing errors in speed as well as orientation of the optical flow estimate. Figure 13(b) shows the histogram of the output voltages of all motion units. Very similar results were found for other chips of the same fabrication batch. The outputs consistently approximate normal distributions as long as the motion signals are within the linear range of the circuit, which is in agreement with randomly induced mismatch due to the fabrication process. Offset values were measured for completely isolated motion units. Already weak coupling among the units noticeably increases the homogeneity of the individual responses.

## 6.2. Flow Field Estimation

On-chip scanning circuitry allows to simultaneously read-out the optical flow estimate (voltages  $U_+$ ,  $V_+$ ) as well as the photoreceptor voltage at each pixel. Scanning was performed at 67 frames/s (100kHz clock frequency) but can be as high as 1000 frames/s before the cut-off frequency of the follower based scanning circuitry starts to significantly impair the signals. In both examples  $\text{Bias}_{OP}$  was 0.35 V.

Figure 14 demonstrates how different bias values of the HRes circuits can control smoothness of the optical flow estimate. The same “triangle” stimulus as in Fig. 12(b) was presented, moving from right to left with constant speed. Optical flow was recorded for three

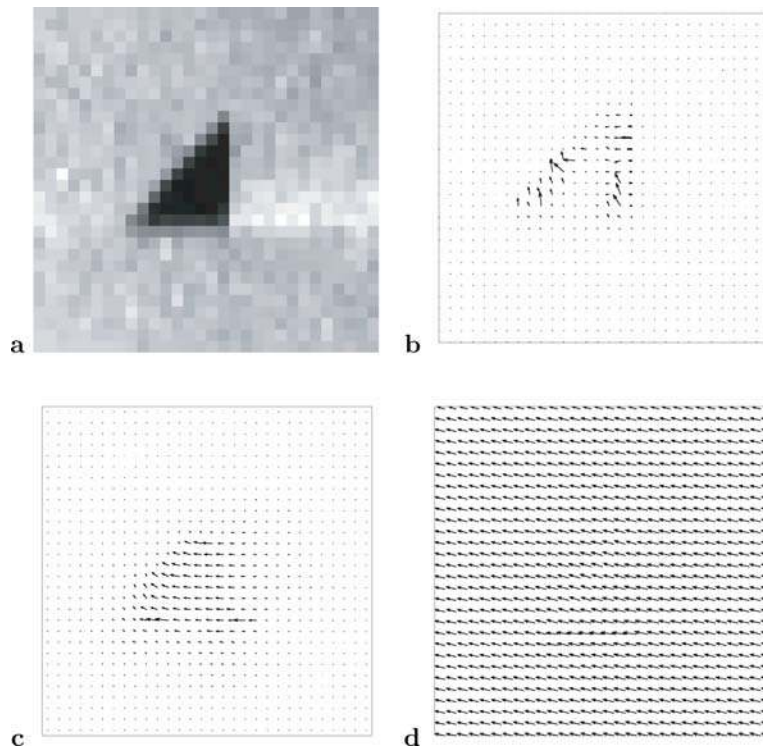


Fig. 14. Varying the smoothness of the optical flow estimate. (a) The photoreceptor signal represented as gray-value image. (b-d) Sampled optical flow estimates for increasing weights of the smoothness constraint ( $\text{Bias}_{HR} = [0.25, 0.38 \text{ and } 0.8 \text{ V}]$ ).

different smoothness strengths and snap-shots representing the momentary optical flow estimate are shown when the object approximately passed the center of the visual field. Figure 14(a) represents the photoreceptor output voltages while (b–d) display the optical flow fields defined by the voltage distributions  $U_{ij}$  and  $V_{ij}$  (relative to  $V_0$ ) for increasing values of  $\text{Bias}_{HR}$ . For low values, spatial interaction amongst motion units was limited and the result is a noisy normal flow estimate. As the voltage is increased, the optical flow estimate becomes smoother and finally represents an almost global estimate that well approximates the correct two-dimensional object motion.

The last example illustrates how the sensor behaves for “natural” stimuli under real-world conditions. Figure 15 shows a sampled sequence of the optical flow sensor’s output while observing two tape rolls passing each other on an office table from opposite directions. Again, each frame (a–i) is a snap-shot in chronological order (every fourth frame is shown) that represents the instantaneous outputs of the sensor. The photoreceptor output is encoded as gray-scale image and overlaid with the estimated optical flow field.

The optical flow estimate matches well the qualitative expectations according to the previous analysis. Rotation of the rolls could not be perceived because of the missing spatial brightness structures on their surfaces. The flow fields are mildly smooth leading to

perceived motion also e.g. in the center-hole of each roll. The chosen value for the smoothness conductance ( $\text{Bias}_{HR} = 0.41 \text{ V}$ ) permits the sensor to estimate the perceptually correct object motion in the middle of the tape rolls, whereas at the outer borders of the rolls, the smoothness kernel is not large enough to cover areas with spatial brightness gradients sufficiently distinct orientation. The aperture problem holds and the estimated optical flow field tends toward a normal flow estimate.

The “tape rolls” sequence (and to some degree also the “triangle” sequence) demonstrates another interesting behavior of the sensor. The optical flow field seems partially sustained and trailing the trajectory of the rolls (see e.g. the optical flow field on the left-hand side in Fig. 15(c)), indicating a long decay time-constant of the optical flow signal. The time-constant is determined by the capacitance and the total conductance at the capacitive nodes of each motion unit. The total conductance, however, is not constant and is given mainly by the transconductance of the feedback loop and the bias conductance<sup>5</sup>. It follows from the network dynamics (3) that the transconductance in each feedback loop is approximately proportional to the square of the spatial brightness gradient. Because the background and the table in the tape rolls sequence are unstructured the transconductance is dropping low once the high contrast boundary of the tape rolls has passed. Thus, the

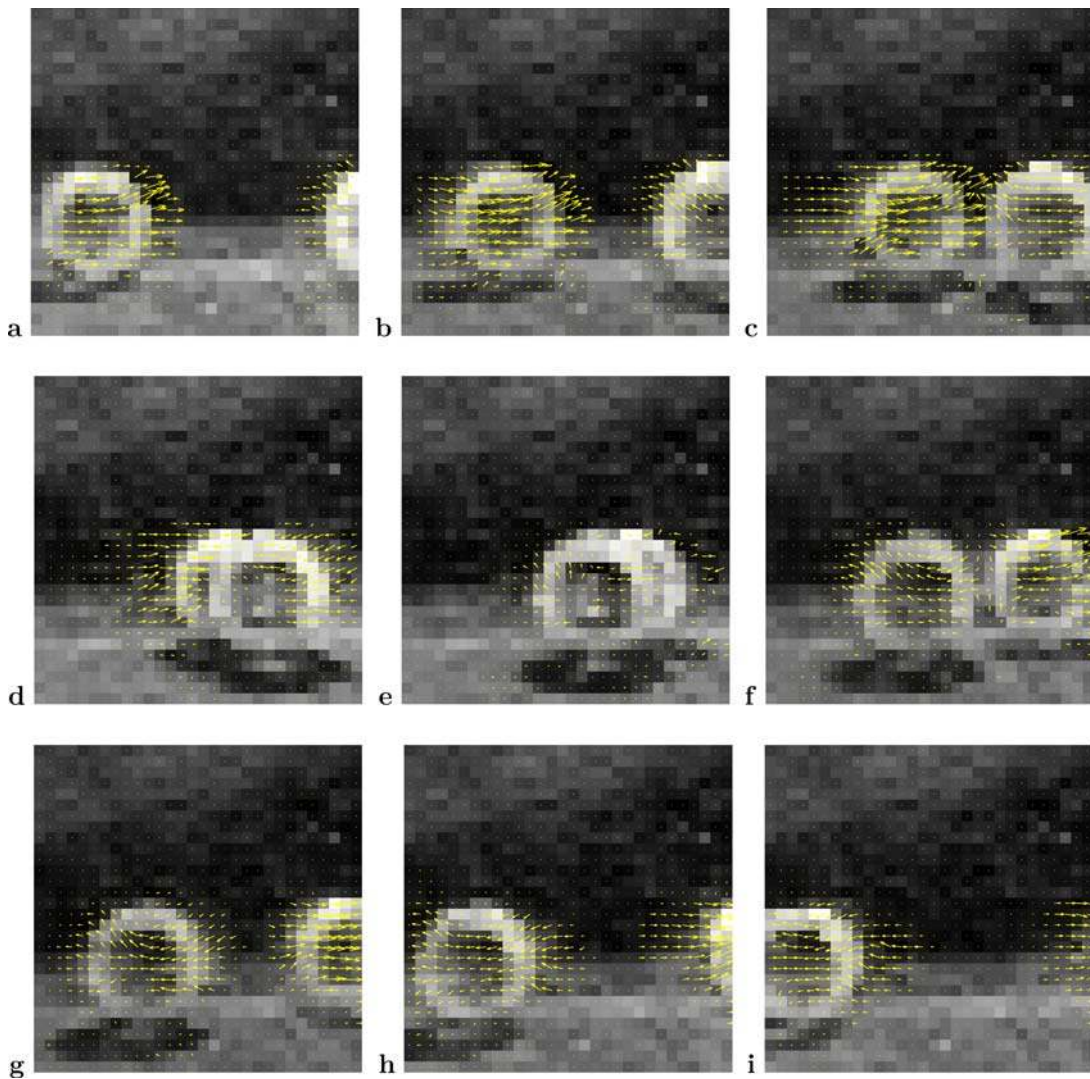


Fig. 15. Output sequence of the optical flow sensor for a real-world scene. The scene—two tape rolls rolling over a flat surface from opposite directions—was observed by the optical flow sensor. The individual frames show the output of the sensor sampled at 67 Hz (every fourth frame is shown). The frames are in chronological order (a-i) where the gray-scale images represent the photoreceptor signals. The instantaneous estimate of the optical flow field is superimposed.

time-constant becomes long because it is now solely determined by the bias conductance  $\sigma$ —which is typically small. This leads to the slow decay of the estimated optical flow fields. Consequently, a fast decay should result in situations where strong spatial brightness gradients are present right after the moving object has passed. This can be observed in Figure 15(f) where after occlusion, the high-contrast back edge of the second roll immediately induced new motion information.<sup>6</sup> In subsequent frames (Figs. 15(g)–(i)) the trailing flow field is building up again.

## 7. Discussion

The presented optical flow sensor is a successful example of a collective and distributed visual processing system implemented in aVLSI. The sensor's nearest-

neighbor connected network architecture applies an error correction strategy. Each optical flow unit in the network acts in a feedback loop, comparing its own local visual information with the result of a collective estimate of the optical flow field and subsequently corrects the estimate to match the local observation. The optical flow sensor is one of the very few aVLSI implementations that follows such strategy. It proves that there is no conceptual reason why such an approach cannot be successfully implemented in aVLSI circuits known for their inherent susceptibility to device mismatch and noise. Mathematically, the sensor solves a *constraint satisfaction* (optimization) problem. It finds the optical flow estimate that optimally matches the visual information with respect to the implicit optical flow model defined by different constraints on the visual information and the desired estimate. Such optimization

problems can be computationally very expensive to implement on sequentially computing, digital hardware.

The optical flow sensor embeds a motion model that exceeds the complexity of models applied in other aVLSI motion sensors. Spatial integration of visual information provides a robust and perceptually more correct optical flow estimate. Also, the bias constraint guarantees that the optimization problem is well-posed and that the sensors response degrades gracefully at very low stimulus contrasts. Interestingly, many illusions in visual motion perception of humans can be qualitatively well explained by a bias constraint [32, 43]. Such similarities should not be over-emphasized; the optical flow sensor is at most a functional model of human visual motion processing. Nevertheless, they indicate that the model seems a valuable choice for real-world visual conditions. The non-linear aVLSI implementation of the network conductances has a positive effect on the optical flow model. The optical flow estimate becomes less biased for high contrast stimuli without reducing robustness (see Fig. 11(a)) compared with a linear bias conductance (thus a quadratic constraint formulation). And the non-linear smoothness conductance reduces smoothing across motion discontinuities.

The logarithmic photoreceptor circuit substantially participates in the robust behavior of the sensor. Because of the logarithmic compression, the spatial gradients of the image brightness basically encode the local contrast in the image. In principal, this property makes the behavior of the optical flow sensor *independent of the absolute illumination* in the observed visual scene.

Another feature of the model is its adaptive dynamical behavior. The non-zero time-constant of the optical flow sensor leads to the integration of visual motion information over time. Since the time-constant depends on the amplitude of the spatial brightness gradients and thus on the confidence in the visual information present, the actual temporal integration window becomes large for weak and short for reliable visual input (see Fig. 15). At any time the model assumes the optical flow to change smoothly over time with an integration window adapted to the confidence and thus the signal/noise ratio of the visual input.

A few examples already demonstrate the potential of the sensor for smart visual interfaces [46] or robotics applications [47], in particular when tuned to provide a global optical flow estimate. The low dimensional signal of a global optical flow estimate eliminates the need for high transmission bandwidth and complex interfaces to further processing stages, allowing the construction of simple yet powerful systems. Other applications for surveillance and navigation tasks are imaginable.

The network architecture can be further evolved by including mechanisms that locally control the con-

ductances  $\sigma_{ij}(t)$  and  $\rho_{ij}(t)$ . This can permit a richer, more object related computational behavior of the system [27]. We have already shown approaches where the local smoothness conductances are dynamically adapted to limit spatial integration to the extents of individual motion sources [28]. Also the local control of the bias conductance  $\sigma_{ij}(t)$  permits interesting attentionally guided visual motion processing [29]. Motion units not belonging to the attentional focus can be suppressed by increasing  $\sigma_{ij}$ , thus literally shunting the units. We know that such attentional inhibition is involved in the processing of visual motion in primate's visual motion system [48], for example. Although in detail not biological plausible, such extended aVLSI systems of the optical flow sensor would give us the possibility to study in a more systematic way the bottom-up and top-down interplay in a real-time perceptual system.

## Acknowledgments

This article presents work that was exclusively conducted at the Institute of Neuroinformatics INI<sup>7</sup> over the last few years. I want to thank in particular Rodney Douglas, the late Jörg Kramer, and all other members of the hardware group of INI for help and support. Thanks to Eero Simoncelli, Patrik Hoyer and Giacomo Indiveri for helpful comments on the manuscript. Chip fabrication was provided through the multi-project-wafer service kindly offered by EUROPRACTICE. Funding was largely provided by ETH Zürich (research grant no. 0-23819-01). The author is currently supported by the Howard Hughes Medical Institute.

## Notes

1. This is for the ideal and noise-free case. In reality, the local constraints will impose an optimization problem of finding the solution that best fits the local constraints.
2. The discrete gradients are defined as e.g.  $(\Delta u_{ij})^2 = (\frac{u_{ij} - u_{i+1,j}}{h})^2 + (\frac{u_{i,j} - u_{i,j+1}}{h})^2$ , where  $h$  is the grid constant.
3. The difference in log-space turns into a ratio in linear space.
4. Shockley equation:  $I_{\text{diode}} = I_0(\exp(\frac{1}{n} V_d \frac{q}{kT}) - 1)$ .
5. Neglecting the smoothness conductance.
6. Note, that a static background provides motion information (zero motion) as long as it has spatial structure.
7. University and ETH Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

## References

1. C. Mead, "Neuromorphic electronic systems." *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, 1990.
2. P. Nesi, F. Innocenti, and P. Pezzati, "Retimac: Real-time motion analysis chip." *IEEE Trans. on Circuits and Systems - 2: Analog and Digital Signal Processing*, vol. 45, no. 3, pp. 361–375, 1998.
3. E.C. Hildreth, *The Measurement of Visual Motion*. MIT Press, 1983.



4. T. Horiuchi, J.P. 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.), 1991, vol. 3, pp. 406–412.
5. R. Etienne-Cummings, S. Fernando, N. Takahashi, V. Shotonov, J. Van der Spiegel, and P. Müller, "A new temporal domain optical flow measurement technique for focal plane VLSI implementation." in *Computer Architectures for Machine Perception*, December 1993, pp. 241–250.
6. R. Sarpeshkar, W. Bair, and C. Koch, "Visual motion computation in analog VLSI using pulses." in *Advances in Neural Information Processing Systems*, vol. 5, pp. 781–788, 1993.
7. J. Kramer, "Compact integrated motion sensor with three-pixel interaction." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 4, pp. 455–460, 1996.
8. C. Higgins and C. Koch, "Analog CMOS velocity sensors." *Procs. of Electronic Imaging SPIE*, vol. 3019, February 1997.
9. C. Higgins, R. Deutschmann, and C. Koch, "Pulse-based 2D motion sensors." *IEEE Trans. on Circuits and Systems 2: Analog and Digital Signal Processing*, vol. 46, no. 6, pp. 677–687, 1999.
10. T. Horiuchi, B. Bishofberger, and C. Koch, "An analog VLSI saccadic eye movement system." in *Advances in Neural Information Processing Systems 6*, J. Cowan, G. Tesauero, and J. Alspector (Eds.), Morgan Kaufmann, 1994, pp. 582–589.
11. R. Deutschmann and C. Koch, "An analog VLSI velocity sensor using the gradient method." in *Procs. IEEE Intl. Symposium on Circuits and Systems*, IEEE, 1998, pp. 649–652.
12. R. Deutschmann and C. Koch, "Compact real-time 2-D gradient-based analog VLSI motion sensor." in *Proc. SPIE Int. Conference on Advanced Focal Plane Arrays and Electronic Cameras*, 1998, vol. 3410, pp. 98–108.
13. A. Andreou and K. Strohhahn, "Analog VLSI implementation of the Hassenstein-Reichardt-Poggio models for vision computation," in *Procs. of the 1990 Intl. Conference on Systems, Man and Cybernetics*, 1990.
14. R. Benson and T. Delbrück, "Direction-selective silicon retina that uses null inhibition." in *Neural Information Processing Systems 4*, D. Touretzky, (Ed.), MIT Press, 1992, pp. 756–763.
15. T. Delbrück, "Silicon retina with correlation-based velocity-tuned pixels." *IEEE Trans. on Neural Networks*, vol. 4, no. 3, pp. 529–541, 1993.
16. R. Harrison and C. Koch, "An analog VLSI model of the fly elementary motion detector." in *Advances in Neural Information Processing Systems 10*, M. Kearns and S. Solla (Eds.), MIT Press, 1998, pp. 880–886.
17. M. Ohtani, T. Asai, H. Yonezu, and N. Ohshima, "Analog velocity sensing circuits based on bio-inspired correlation neural networks," in *Microelectronics for Neural, Fuzzy and Bio-Inspired Systems*, Granada, Spain, 1999, pp. 366–373.
18. S.-C. Liu, "A neuromorphic aVLSI model of global motion processing in the fly." *IEEE Trans. on Circuits and Systems 2*, vol. 47, no. 12, pp. 1458–1467, 2000.
19. H.-C. Jiang and C.-Y. Wu, "A 2-D velocity- and direction-selective sensor with BJT-based silicon retina and temporal zero-crossing detector." *IEEE Journal of Solid-State Circuits*, vol. 34, no. 2, pp. 241–247, 1999.
20. J. Kramer, R. Sarpeshkar, and C. Koch, "Pulse-based analog VLSI velocity sensors." *IEEE Trans. on Circuits and Systems 2*, vol. 44, no. 2, pp. 86–101, 1997.
21. R. Etienne-Cummings, J. Van der Spiegel, and P. Mueller, "A focal plane visual motion measurement sensor." *Trans. on Circuits and Systems I*, vol. 44, no. 1, pp. 55–66, 1997.
22. S. Mehta and R. Etienne-Cummings, "Normal optical flow measurement on a CMOS APS imager." in *Int. Symposium on Circuits and Systems*, Vancouver: IEEE, May 2004, vol. 4, pp. 848–851.
23. J. Tanner and C. Mead, "An integrated analog optical motion sensor." in *VLSI Signal Processing*, 2, S.-Y. Kung, R. Owen, and G. Nash, (Eds) IEEE Press, 1986, p. 59 ff.
24. A. Moore and C. Koch, "A multiplication based analog motion detection chip." in *SPIE Visual Information Processing: From Neurons to Chips*, B. Mathur and C. Koch (Eds.), SPIE, 1991, vol. 1473, pp. 66–75.
25. A. Stocker and R. Douglas, "Computation of smooth optical flow in a feedback connected analog network." in *Advances in Neural Information Processing Systems 11*, M. Kearns, S. Solla, and D. Cohn (Eds.), MIT Press: Cambridge, MA, 1999, pp. 706–712.
26. M.-H. Lei and T.-D. Chiueh, "An analog motion field detection chip for image segmentation." *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 5, pp. 299–308, 2002.
27. A. Stocker, "Constraint optimization networks for visual motion perception - analysis and synthesis." Ph.D. dissertation, Dept. of Physics, Swiss Federal Institute of Technology, Zürich, Switzerland, March 2002.
28. A. Stocker, "Analog VLSI focal-plane array with dynamic connections for the estimation of piece-wise smooth optical flow." *IEEE Transactions on Circuits and Systems-1: Regular Papers*, vol. 51, no. 5, May 2004, pp. 963–973.
29. A. Stocker and R. Douglas, "Analog integrated 2-D optical flow sensor with programmable pixels," in *IEEE Int. Symposium on Circuits and Systems*, Vancouver: IEEE, May 2004, vol. 3, pp. 9–12.
30. J. Limb and J. Murphy, "Estimating the velocity of moving images in television signals." *Computer Graphics Image Processing*, vol. 4, pp. 311–327, 1975.
31. B. Horn and B. Schunck, "Determining optical flow." *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.
32. Y. Weiss, E. Simoncelli, and E. Adelson, "Motion illusions as optimal percept." *Nature Neuroscience*, vol. 5, no. 6, pp. 598–604, 2002.
33. A. Yuille and N. Grzywacz, "A mathematical analysis of the motion coherence theory." *Intl. Journal of Computer Vision*, vol. 3, pp. 155–175, 1989.
34. E. Simoncelli, E. Adelson, and D. Heeger, "Probability distributions of optical flow." in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 1991, pp. 310–313.
35. J. Harris, C. Koch, E. Staats, and J. Luo, "Analog hardware for detecting discontinuities in early vision." *Intl. Journal of Computer Vision*, vol. 4, pp. 211–223, 1990.
36. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons." *Procs. National Academic Sciences U.S.A.*, vol. 81, pp. 3088–3092, May 1984.
37. J. Hopfield and D. Tank, "Neural computation of decisions in optimization problems." *Biological Cybernetics*, no. 52, pp. 141–152, 1985.
38. T. Delbrück and C. Mead, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits." Caltech Computation and Neural Systems Program, Tech. Rep. 30, 1994.
39. S.-C. Liu, "Silicon retina with adaptive filtering properties." in *Advances in Neural Information Processing Systems 10*. MIT Press, November 1998, pp. 712–718.
40. C. Mead, *Analog VLSI and Neural Systems*. Addison-Wesley: Reading, MA, 1989.
41. B. Gilbert, "A precise four-quadrant multiplier with subnanosecond response." *IEEE Journal of Solid-State Circuits*, vol. 3, no. 4, pp. 363–373, December 1968.
42. P. Gray and R. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd ed. Wiley and Sons: New York, 1993.

43. A. Stocker and E. Simoncelli, "Constraining a Bayesian model of human visual speed perception," in *Advances in Neural Information Processing and Systems NIPS*, vol. 17. MIT Press: Vancouver, 2005, to appear.
44. J. Harris, C. Koch, and J. Luo, "A two-dimensional analog VLSI circuit for detecting discontinuities in early vision." *Science*, vol. 248, pp. 1209–1211, June 1990.
45. T. Delbrück, "Investigations of analog VLSI visual transduction and motion processing." Ph.D. dissertation, Department of Computational and Neural Systems, California Institute of Technology, Pasadena, CA, 1993.
46. J. Heinzle and A. Stocker, "Classifying patterns of visual motion—a neuromorphic approach." in *Advances in Neural Information Processing Systems 15*, S. T. S. Becker and K. Obermayer (Eds.), MIT Press: Cambridge, MA, 2003, pp. 1123–1130.
47. V. Becanovic, G. Indiveri, H.-U. Kobialka, Plöger, and A. Stocker, *Mechatronics and Machine Vision 2002: Current Practice*, ser. Robotics. Research Studies Press, 2002, ch. Silicon Retina Sensing guided by Omni-directional Vision, pp. 13–21.
48. S. Treue and J. Maunsell, "Attentional modulation of visual motion processing in cortical areas MT and MST." *Nature*, vol. 382, pp. 539–541, August 1996.



**Alan A. Stocker** received the MS degree in Material Sciences and Biomedical Engineering and the

PhD degree in Physics from the Swiss Federal Institute of Technology in Zürich, Switzerland, in 1995 and 2001, respectively. In 1995, he was with the Institute for Biomedical Engineering at the Helsinki University of Technology, Finland, where he performed his master's thesis. In the end of 1995, he joined the Institute of Neuroinformatics in Zürich where he graduated with a dissertation about the analysis and the design of analog network architectures for visual motion perception. Since 2003, he is a postdoctoral fellow at the Center for Neural Sciences of New York University. His research interests are in computational neuroscience with a focus on models for primate's visual motion system. Furthermore he is interested in the analysis and development of neurally inspired, computationally efficient hardware models of visual perception and their applications in robotics.