MODELING A MOTION PERCEPTION AND ESTIMATION SUITABLE FOR VLSI IMPLEMENTATION

Goce V. Shutinoski, Tomislav A. Dzhekov
Faculty of Electrical Engineering
P.O.Box 574 91000 Skopje
Republic of Macedonia

Keywords: motion perception, motion estimation, video signal processor, neuromorphic cells, VLSI

***Abstract:** Perception and evaluation of motion in real time are prerequisites for achieving vision chips capable to detect motion in the vision field, tracking objects and navigating through the 3-D space. They comprise fundamentals in early vision and are important step toward seeing as an intelligent act. Computation of the velocity vector field associated with object motion is one of the most exhaustive computational tasks in image processing and analysis that consumes over a half of the digital VSP chip available computational capability. So it is not surprising that a great deal of effort has been directed into providing an efficient solution to this problem. There are several application oriented classes of motion estimation algorithms such as pel-recursive, block-matching, frequency domain techniques and techniques based on neural networks implementation concept. In this paper neural correlation of motion detection mechanism similar to the visual system of invertebrates is adopted for VLSI implementation. More specifically, using VLSI implemented neuromorphic cells with the dynamics of the transmitter characterized by first order differential equation, the bipolar cells are formed. Combining these cells in suitable manner it is possible to provide simple mechanism to achieve direction sensitive elementary motion detectors. Both spatial and temporal adaptation responses of the VLSI model are investigated and SPICE simulation results are presented.*

1.Introduction

Vision is the most important sensorial interface between the surrounding environment and the central nervous system not only for vertebrates but for all except a few animals. Motion analysis, often referred to in the context of optical flow, is fundamental intermediate-level vision task [1]. Perception and evaluation of 3-D motion in real time are an integral part of vision and involve computation of the parameters of 3-D motion and prediction of future positions of moving objects. Visual motion detection is an early preatentative mechanism preceding many higher levels of information processing [1], [2]. In the invertebrate it can be used for target detection in camouflage or target tracking without elaborate perception of the important aspects such as form, while in the human visual system these basic tasks are strongly interrelated by a complex, 3-D form perception. Seeing is particularly important for a fast moving animals, the fly for example, and no artificial system has comparable optimal properties [1]. It is not surprising that many research efforts have been directed toward better understanding of motion perception in the visual system of invertebrates, as well as in vertebrates, so that accumulated knowledge can successfully be implemented in building proprietary vision chips.

While the techniques of classical motion analysis are still being improved interest in the field has been moved over the past decade to exploring some different approaches that may compensate the limitations of the traditional methods. It is well known that the 3-D motion interpretation and motion parameters derivation are strongly related to the estimation of optical flow. In [4] is presented neuromorphic silicon model that emulates vertebrate retina that can adapt locally to brightness changes, can detect edges and can compute motion. Excellent review of the state of the art retinal processing and neuromorphic vision chips can be found in [2].

Neuronal activity, in general, can be modeled in terms of the all-or-none response of a cell to incoming stimuli. The resulting firing patterns in networks of such neural cells arise from highly nonlinear processes related with neuron membrane activity, generation of spikes, their transmission and further higher level processing mechanisms. Certain important aspects of visual processing, which occur at retinal level, include spatial and temporal enhancement of stimuli. It has been shown that such spatio-temporal adaptation arises from inhibitory connection in the retina both between neighboring cells and via feedback loops to the cell under the consideration making a process of self-inhibition [3].

The simplified process of human vision at the retinal level is as follows. Light entering the eye through the pupil passes through the retina before landing on the rod and cone photoreceptors that convert it into neural signals that propagate forward to the several ganglion layers on the front of the retina, fig.1. The function of H cells layer has to be considered as averaging the incoming signal from their reception field (local averaging) which appear to be a form of lateral inhibition occurring in determining the bipolar cell response. Excitatory connections from these cells enable sensing the presence and temporal variations of stimuli, while spatially adjacent cell, by a process of comparison of past and present activities, stimulates corresponding ganglions to produce output that can be used by specialized cells to detect elementary motion. It is suggested that a pattern-motion cell combines motion information from multiple local sources and becomes activated when globally coherent motion in a prescribed direction is present in the cell's receptive field [5].

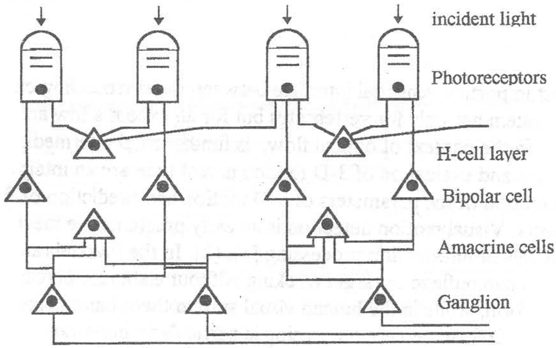


Fig.1 A retina structure

This paper presents a plausible elementary motion detector model similar to the fly visual system well described in [1]. First we start with short review of the dynamics of the simple cell and then by combining such cells in suitable manner, we present a gated dipole structure with possibility of both temporal and spatial adaptations. In the next section we explore CMOS capabilities for circuits implementation. More specifically, using VLSI implemented neuromorphic cells with the dynamics of the transmitter characterized by first order differential equation, the bipolar cells are formed. Combining these devices in more complex structures it is possible to provide simple mechanism to achieve direction sensitive elementary motion detectors [1], [3]. In the last section simulation results of both temporal and spatial behavior of the device are presented and appropriate conclusions are given.

2. A dynamic of cells

Let us consider the neuronal cell with the dynamics characterized by the following differential equation

$$\frac{dz(t)}{dt} = a(b - z(t)) - x(t)z(t) \qquad (1)$$

where **ab** is the transmitter production rate, **-az(t)** is the feedback inhibition and **b** denotes the maximum amount of a transmitter, which states that the input signal **x(t)** depletes the transmitter proportionally to its strength and to the available amount of transmitter **z(t).** Suppose that a signal transmitted to the postsynaptic cell is proportional not only to the input stimuli **x(t)** but also to the available amount of transmitter. Than the signal produced at the output can be written as

$$y(t) = x(t)z(t) \qquad (2)$$

Now, suppose that a step stimulus, $[x(t) = x,\ t>0]$, is applied and the transmitter is fully accumulated, $[z(0) = b]$. One can find that for t >0 the response of the cell is

$$y(t) = \frac{abx}{a+x}\left(1 + e^{-(a+x)t}\right) + bxe^{-(a+x)t} \qquad (3)$$

which means that after the overshoot [**bx**], depletion process produces a temporal adaptation and the output settles to the value

$$y(\infty) = \frac{abx}{a+x} \qquad (4)$$

The second problem is the question of spatial adaptation or the discrimination of spatial inputs from spatial background. In conjunction with this, concerning the motion estimation, is the aperture problem which arises as a consequence of uncertainty about both the position and orientation of a visual stimulus determination by a single cell, since the local direction of motion signaled by the activity of such cells is ambiguous [3], [5]. Thus, two degrees of freedom in motion measurements must be introduced which is equivalent to two channels processing mechanism. Consider two same channels with lateral inhibition connections as presented in [1] and depicted in fig.2. The interconnections at this second stage mathematically are described by following equations

$$\begin{aligned} \frac{dy_{on}}{dt} &= -Ay_{on} + (B - y_{on})[I + J_{on}]z_{on} - (D + y_{on})[Iz_{off}] \\ \frac{dy_{off}}{dt} &= -Ay_{off} + (B - y_{off})[I]z_{off} - (D + y_{off})[I + J_{on}]z_{on} \end{aligned} \qquad (5)$$

where, **A** is the rate of spontaneous decay of the cell activity, **B** and **-D** are upper and lower levels of saturation of the activity, $I+J_{on}$ is the net excitatory input or net inhibitory input depending on the cell under the consideration, **I** is the arousal signal feeding both channels, and notation **y** designates both cell and its activity. Applying signal $x = I+J_{on}$ to the left channel and $x = I$ to the other and taking into account slow time scale of the transmitter dynamics, one can finds the equilibrium states as

$$\begin{aligned} y_{on} &= \frac{B[I + J_{on}]z_{on} - DIz_{off}}{A + [I + J_{on}]z_{on} + Iz_{off}}, \quad z_{on} = \frac{ab}{a + I + J_{on}} \\ y_{off} &= \frac{BIz_{off} - D[I + J_{on}]z_{off}}{A + [I + J_{on}]z_{on} + Iz_{off}}, \quad z_{off} = \frac{ab}{a + I} \end{aligned} \qquad (6)$$

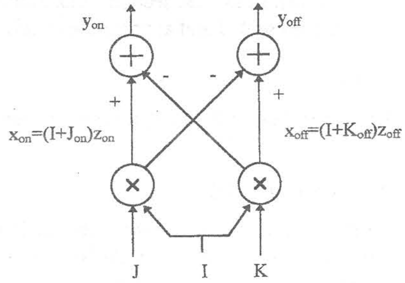


Fig.2 A gated dipole

Inspecting the above equations for $\mathbf{J_{on} = 0}$, $\mathbf{t<0}$, it is obviously that activities of both cells will equilibrate to the value

$$\mathbf{y_{on} = y_{off} = \frac{[B-D]Iz_{on}}{A+2Iz_{on}}, \; z_{on} = z_{off} = \frac{ab}{a+I}} \tag{7}$$

and if spatially homogeneous input is applied, than $\mathbf{J_{on}=K_{off}=I}$ and the overall response of the cells is

$$\mathbf{y_{on} = y_{off} = \frac{ab[B-D][I+J]}{A[a+I+J]+2ab[I+J]}} \tag{8}$$

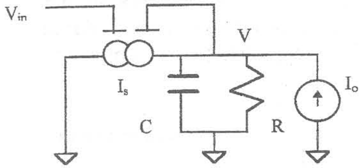
That is, because of lateral interactions of the cells spatial adaptation is generated.

3. On modeling a CMOS neural cell

Monolithic integration becomes a necessity for applications such as artificial neural networks where the large scale of components has to be implemented. According to the continuous-time properties of proposed system, modeled by the ordinary differential equations we decide to implement monolithic CMOS realization of elements in 2 μm n-well technology. This section presents most important design considerations for monolithic implementation of complex cell dynamics in CMOS technology.

The useful range of component values for this technology is determined by sizes and shapes of corresponding physical components. Absolute accuracy of integrated components is very poor ranging around 20% and depends on temperature variations, process of aging and component spreading. On the good side is tolerance of ratios between similar components (as low as 0.1%), encountering the influence of temperature, aging or even nonlinearities (cancellation of nonlinearities in a current mirror or differential pair). In n-well CMOS technology NMOS transistors are directly fabricated on the global p-type substrate whereas PMOS transistors are made on local n-type substrate (wells) that have been created on the global substrate with a local diffusion-implantation process. In standard CMOS technologies both passive quasi-linear resistors with area occupancy $\sim 10^3\ \mu m^2/K\Omega$ and active elements resistive components based on MOS ohmic region properties can be fabricated. Also thick oxide capacitors can be formed at the price of area occupation ($\sim 10^4\ \mu m^2/pF$), or by exploiting capacitive effects between gate and channel of an MOS transistor, ($\sim 10^3\ \mu m^2/pF$), [7].

Let us now consider the structure of fig.3. The capacitor potential V can be found by solving the first order differential equation of the form



$$\frac{dV(t)}{dt}=\frac{1}{R}\left[\frac{I_0}{RC}-V\right]-\frac{1}{C}I_s \qquad (9)$$

Fig.3 Circuit that emulates a cell dynamic

where $\mathbf{I_0}$ represents constant current source and $\mathbf{I_S}$ represents a voltage controlled current source. By a simple change of variables $\mathbf{a=1/R}$, $\mathbf{b=I_0/(RC)}$ and $\mathbf{I_S=Cx(t)z(t)}$, this is the same relation as (1). Thus, our goal is to design proprietary CMOS building blocks to implement the proposed cell dynamics.

4. A CMOS implementation

A transconductance current-mode approach is adopted for the implementation of the voltage controlled current source I_S. The actual realization comprises a high-output impedance transconductance element used as a controllable current source linear in the driving voltage V_{in} over a wide range and the double-stack transistor configuration of the differential pair [6],[8]. Fig.4 shows the schematic of the transconductance element. The device utilizes MOS transistor M1 biased in the linear triode region connected to a cascode transistor M2 via high-gain feedback circuit. Therefore, the triode transistor drain voltage is forced to a constant level largely independent of the input voltage. The supplied output current is proportional to the driving voltage V_{in} while invariant to the output voltage of the device, $[I_{tc}=G_m(V_{ctrl})Vin]$. The

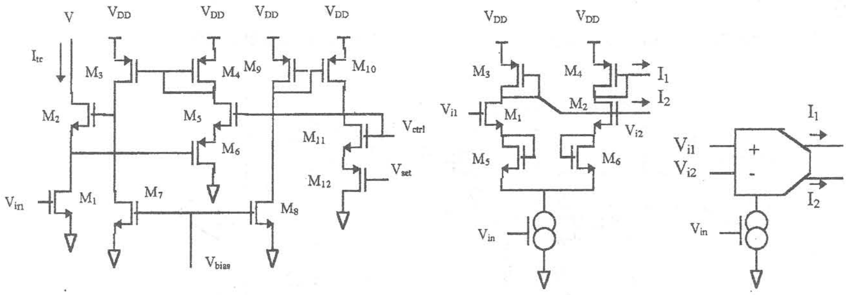


Fig.4 Transconductance element with high output impedance (symbol)

bias circuit that generates the control voltage for transconductance element is specified by an externally supplied voltage V_{ctrl}, allowing approximately linear control of the G_m value of the element, keeping the output voltage as low as possibly (~0.5V). The output current of the transconductance element feeds into a differential pair injecting the current $g_m(V_{i1}-V_{i2})I_{tc}$ differentialy into the diode-connected transistors M3, M4 output lines I_1, I_2. Fig.5 shows the dependence of the generated output current of the device ($I_{out}=I_1-I_2$) on the input voltage V_{i1} (excitatory input) and the control voltage V_{in} (averaging signal). Notice that V_{i2} acts as the selfinhibitory signal.

Fig.6 presents active components implemented resistor utilizing ohmic region of transistor M1. Providing that the applied resistor voltage V_R is grater than $\max\{V_{T1},V_{T2}\}$, while the other MOS transistors are properly biased in saturation region, the nonlinear term of the implemented element is canceled which produces satisfying linear resistor V/I dependence. Although, there have been many proposed improved mirror structures [7], [8], a simple four-

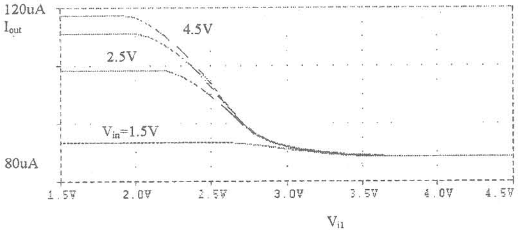


Fig.5

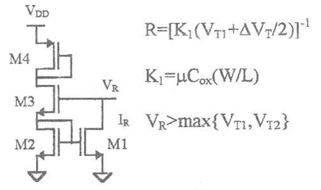


Fig.6 Active components resistor

transistor current mirrors are chosen since the device functionality and simulation validity is the most important concern in this application. Implemented current duplicating devices so as current summing structures have a wide range of linearity on the voltage at the summing node providing that this voltage is kept in the range of the values $(V_{DD}-V_T)/2$ and $(V_{DD}+3V_T)/2$.

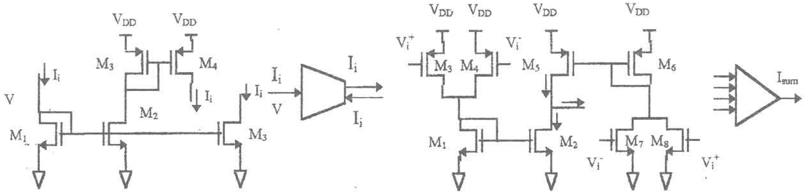


Fig.7 Current duplicator (symbol) and current summing element

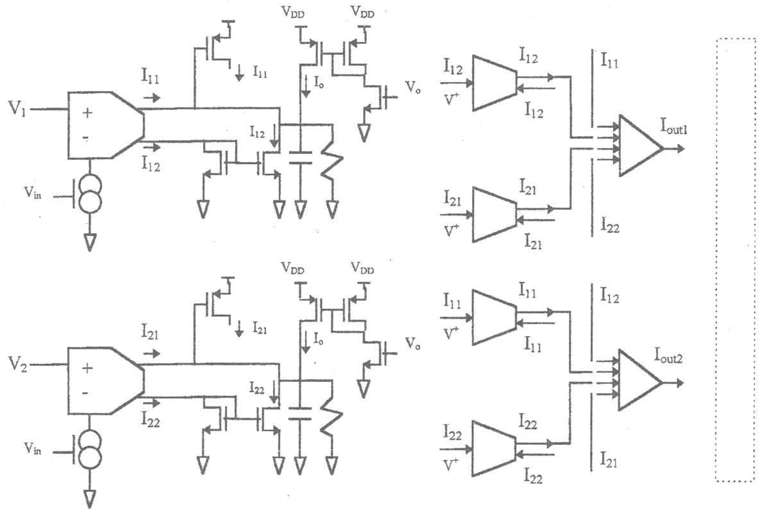


Fig.8 Simplified schematic of bipolar cell circuit

In the fig.8 is presented simplified schematic of a bipolar cell circuit. Notice that a similar second stage of cells is omitted.

5. Simulations and results

Extended structure of fig.8 was generated by a subcircuit model definition and simulations on Spice program were performed. We experimented with various input signal levels that represent sudden changes of environmental signal. Fig.9 shows temporal adaptation process of a simulated neuromorphic circuit over various levels of excitation in both directions. Notice that the temporal adaptation strongly depends on the accumulated transmitter and applied stimulus. The change of the stimulus strength produces the change of the overshoot value that is followed by a settling plateau. Fig.10 depicts the spatial adaptation process. Applying the same excitation input for the on-cell as above while holding the off-cell at a level of arousal signal resultants in overshoot response followed by a plateau for on-cell but the activity of the off-cell is suppressed. When the excitation to the on-cell becomes low, it responses undershoot and the off-cell will rebound. Upon a same stimulus to both of the cells they respond equally reaching approximately same activity level.

6. Conclusion

In this paper the analog current-mode design technique is used to realize the neuromorphic cell capable to respond on temporal and spatial variations of stimuli. SPICE simulation of 2μm CMOS components implemented circuits gives results that are in good accordance with the mathematical background. Concerning the actual motion direction and estimation of motion parameters, more realistic results should be expected using 2-D network of such cells.

References

[1] H.Ogmen, S.Gagne': "Neural network architectures for motion perception and elementary motion detection in the fly visual system", Neural Networks vol.3, pp.487-505, 1990;

[2] Spectrum, May 1996. pp.38-46 "Neuromorphic vision chips";

[3] J.G.Taylor:"A silicon model of vertebrate retinal processing", Neural networks vol.3, pp.171-178, 1990;

[4] J.Hutchison, C.Koch, J.Luo, C. Mead:"Computing motion using analog and binary resistive networks", IEEE Computer mag. vol.21 No.3, pp 52-63, 1988;

[5] J.Marshall:"Self-organizing neural networks for perception of visual motion", Neural networks, vol.3, pp.45-74, 1990;

[6] G.Cauwenberghs:"An analog VLSI recurrent neural network learning a continuous-time trajectory", IEEE Tran. on NN, vol. 7, No.2, pp.346-361, 1996.

[7] C.-Y.Wu, J.-F.Lan:"CMOS current-mode neural associative memory design with on-chip learning", IEEE Tran.on NN, vol.7,No.1, pp.167-181, 1996.

[8] A.Rodriguez-Vasquez, M. Delgado-Restituto:"CMOS design of chaotic oscillators using state variables:a monolithic Chua's circuit", IEE Tran. on CAS-II, vol.40, No.10, pp.596-613, 1993.

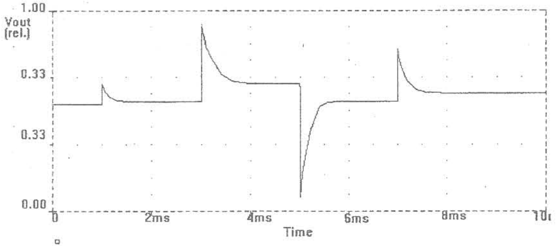
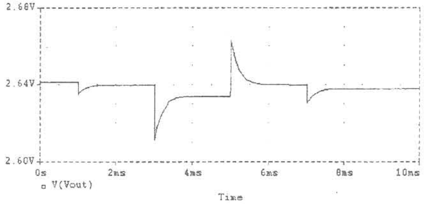


Fig.9 A temporal adaptation



Oposite response - rebounding

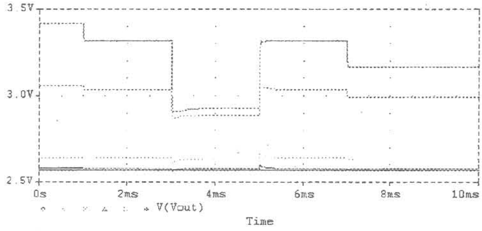


Fig10 Spatial adaptation (in conjunction with fig.9)