ANALOG INTEGRATED CIRCUITS AND SIGNAL PROCESSING

An International Journal

Volume 9, No. 2, March 1996

Analog Integrated Circuits and Signal Processing, 9, 141–166 (1996)

Translinear Circuits in Subthreshold MOS

ANDREAS G. ANDREOU

Electrical and Computer Engineering, Johns Hopkins University, Baltimore MD 21218 USA

KWABENA A. BOAHEN

Computation and Neural Systems, California Institute of Technology, Pasadena CA 91125 USA

Abstract. In this paper we provide an overview of translinear circuit design using MOS transistors operating in subthreshold region. We contrast the bipolar and MOS subthreshold characteristics and extend the translinear principle to the subthreshold MOS ohmic region through a drain/source current decomposition. A fronthack-gate current decomposition is adopted; this facilitates the analysis of translinear loops, including multiple input floating gate MOS transistors. Circuit examples drawn from working systems designed and fabricated in standard digital CMOS oriented process are used as vehicles to illustrate key design considerations, systematic analysis procedures, and limitations imposed by the structure and physics of MOS transistors. Finally, we present the design of an analog VLSI "translinear system" with over 590,000 transistors in subthreshold CMOS. This performs phototransduction, amplification, edge enhancement and local gain control at the pixel level.

1. Introduction

The *Translinear* principle [1] exploits the exponential current-voltage non-linearity in semiconductor devices and offers a powerful circuit analysis and synthesis [2] framework. Originally formulated for bipolar transistors [1], this principle enables the design of analog circuits that perform complex computations in the current-domain including products, quotients, and power terms with fixed exponents [1], [2]. Translinear circuits perform these computations without using differential voltage signals and are amenable to device–level circuit design methodology.

Most of the work on translinear circuits to date, use bipolar transistors and the emphasis is on high precision and high speed. One fascinating aspect of translinear circuits is their insensitivity to isothermal temperature variations, though the currents in its constitutive elements (the transistors) are exponentially dependent on temperature. The effect of small local variations in fabrication parameters can also be shown to be temperature independent. An excellent up-to-date overview of translinear current–mode analog circuits using bipolar transistors can be found in [3].

The increased commercial interest in analog CMOS LSI and VLSI has renewed interest in the translinear principle for MOS circuit design. A generalized form of the translinear principle was recently proposed for MOS operating above threshold [4]; this extension however does not follow the original definition of a translinear circuit [1]. This extension is simply a design principle that exploits conservation of energy (KVL) around circuit loops which have specific topological properties. **A** novel class of translinear circuits that employs multiple input gates, with floating gate MOS transistors in subthreshold has been recently proposed and experimentally demonstrated [20].

Another exciting research area that emerged the last few years, is the synthesis of analog VLSI for sensory information processing systems [7], [8] employing MOS transistors operating in subthreshold region [5], [6], [7]. We have been exploring translinear circuits in subthreshold MOS for use in analog neuromorphic LSI and VLSI systems [11], [12], [9], [10]. In this biologically motivated computational paradigm, high processing throughput is attained through a trade-off between massive parallelism and lower speed in the circuits and therefore subthreshold CMOS operation is possible. Such architectures often necessitate the computation of linear and non-linear functions, and if a current–mode [11], [12] design methodology is adopted, the translinear principle offers an effective way for synthesizing circuits [13], [14] and systems [15], [16], [17], [18].

In this paper, we discuss *experimental* circuit designs based on the translinear properties of subthreshold MOS transistors in the saturation and ohmic regions. Our objective is to present a comprehensive overview on this subject, beginning with the basic devices and circuits, and following it through to the system level. The discussion of subthreshold MOS models, and their characteristics and limitations can be found in other excellent references (for example [5], [6]). However, a basic review of subthreshold MOS and bipolar operation, is provided since the large signal properties of the devices are ***key*** to the subject matter. Most of the circuit examples given, have been used in analog LSI and VLSI systems that have been fabricated and tested functional. The value of these circuits can only be fully appreciated in the context of the systems that employ them; references to the original journal articles are given.

The paper is divided into six sections. Section 2 contrasts the translinear properties of bipolar transistors with those of MOS transistors in subthreshold. Basic circuit techniques that employ MOS transistors in subthreshold saturation and ohmic regimes are introduced in section 3. In the same section, we discuss both translinear loops (TL) composed of generalized diodes and current sources, and translinear networks (TN) that include voltage sources as well. Section 4, focuses on an analog VLSI translinear system, a contrast-sensitive, silicon retina [17]. A discussion of MOS device limitations and deviations from the first order large signal models that ultimately affect circuit and system performance is presented in section 5. Section 6 concludes the paper.

2. Translinear Devices

We begin the discussion of translinear circuits in subthreshold MOS technology with the basic devices. A translinear element is a physical device whose transconductance and current through the device are linearly related, that is, the current is exponentially dependent to the controlling voltage. A two terminal p-n junction (diode), with its exponential I-V characteristics, is a translinear element and used often as an example in circuits [3]. Voltage gated, ion channels—conductances — are also translinear devices.

Three-terminal devices are termed "translinear" if the relationship between the current and the controlling voltage is exponential *and* the two terminals across which the controlling voltage is applied exhibit true diode-like behavior, i.e., increasing the voltage on one terminal is exactly equivalent to decreasing the voltage on the other terminal by the same amount. In this case, a loop of such devices consists of voltage drops across pairs of control terminals and we exploit the linear transconductance-current relationship. Bipolar transistors have both properties whereas MOSFETs do not.

The large-signal device model equations for both the bipolar transistor and MOSFET in subthreshold are discussed in Appendix A where the approximations made during their derivations are clearly stated and the symbols are defined. In the active-forward region of operation, the function of a bipolar transistor as a *transconductance amplifier* is captured by the following equation:

$$I_C = I_S\, e^{\frac{V_B - V_E}{V_t}} \tag{1}$$

where $V_t = (kT/q)$ and I_S is defined in Appendix A.

The magnitude of the transconductance from the base is identical to the magnitude of the transconductance from the emitter:

$$g_m \equiv \frac{\partial I_C}{\partial V_{BE}} = -\left.\frac{\partial I_C}{\partial V_E}\right|_{V_B=c} = \left.\frac{\partial I_C}{\partial V_B}\right|_{V_E=c} = \frac{I_C}{V_t} \tag{2}$$

We now contrast the operation of a bipolar transistor as a translinear element with that of an MOS transistor operating in subthreshold. Much like a bipolar transistor, the MOSFET in subthreshold has exponential voltage current characteristics (see Figure 1). There are however, two fundamental differences between MOSFET and bipolar devices that have implications in the design of translinear circuits.

1. Unlike a bipolar transistor, the current in a MOSFET is controlled by the surface potential, which is capacitively-coupled to the gate (front-gate) and bulk (back-gate) terminals.
2. The MOSFET, is symmetric with respect to the source and drain terminals while a bipolar is not.

In summary, the MOS transistor is a *four* terminal device with symmetric drain and source terminals, as result of lossless channel conduction, and an isolated control potential capacitively set by one or more control gates. As we will see in subsequent sections, the latter property of the MOS transistor is a mixed blessing when the design of translinear circuits is considered.

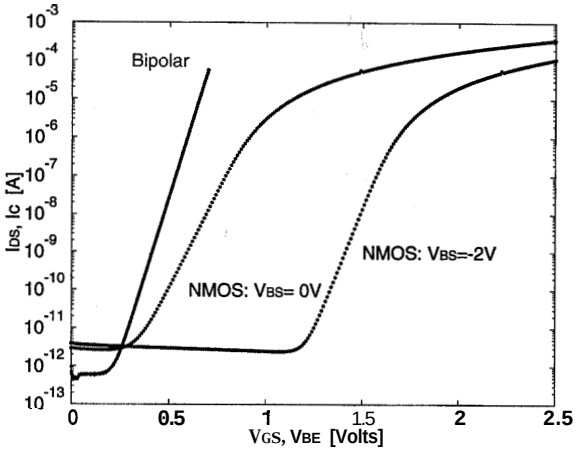

Fig. **1.** Measured current I_{DS} and I_C versus controlling voltage V_{GS} and V_{BE} respectively. The **MOS** transistor has dimensions of ($16 \times 16\mu m^2$) and is fabricated in a $1.2\mu m$ n-well CMOS process and is biased at a drain-source voltage, V_{DS}=1.5 Volts. The current is measured at two different substrate voltage bias conditions. The bipolar transistor is a vertical device with an emitter area of ($16 \times 16\mu m^2$) fabricated in a $2\mu m$ n-well **CMOS** process and biased with V_{CE}=1.5 Volts. T = 301.5 **K.**

It should be pointed out that the voltage difference that controls the current in a MOSFET to yield the translinear behaviour, is the potential difference between the channel surface potential @ and the potential at the source V_S and or drain V_D so that the current between the drain and source for an **NMOS** is given by:

$$I_{DS} \sim \left[\exp\frac{(\psi_s - V_S)}{V_t} - \exp\frac{(\psi_s - V_D)}{V_t}\right] \tag{3}$$

Since the MOS transistor has two "gates" the relationship between $@(V_B, V_G)$ and the bulk or gate terminal voltages V_B and V_G can be obtained using the simple capacitive divider model depicted in Figure 2. The introduction of the parameter $\kappa \equiv C^{\diamond}_{ox}/(C^{\diamond}_{ox} + C^{\diamond}_{dep})$ is convenient for modeling the effect of the two gates. Note that κ is a function of the surface potential ψ_S as $C^{\diamond}_{dep}$ is a function of the applied gate and substrate voltages.

In saturation i.e. when $V_{DS} \geq 4V_t$, and when the current controlling voltages are referenced to source, Equation 3 simplifies to:

$$\begin{aligned} I_{DS} &= SI_{no}\, e^{\left(\frac{C^{\diamond}_{dep}}{C^{\diamond}_{ox}+C^{\diamond}_{dep}}\right)\frac{V_{BS}}{V_t}}\, e^{\left(\frac{C^{\diamond}_{ox}}{C^{\diamond}_{ox}+C^{\diamond}_{dep}}\right)\frac{V_{GS}}{V_t}} \\ &= SI_{no}\, e^{\frac{(1-\kappa_n)V_{BS}}{V_t}}\, e^{\frac{\kappa_n V_{GS}}{V_t}} \end{aligned} \tag{4}$$

Equation *4* can be re-written as a function of dimensionless current quantities i_G and i_B. Each of these currents would correspond to the device current if the surface potential ψ_S could assume the voltage at the gate or bulk terminal. In essence these currents correspond to ideal diode junctions between the source and surface potential weighted by the appropriate capacitive divider ratio. Therefore, the equation for the drain current can be written as:

$$I_{DS} \equiv SI_{no}\, i_G^{\kappa_n}\, i_B^{(1-\kappa_n)} \tag{5}$$

In subsequent sections, we will see how the latter formulation facilitates the analysis of MOS translinear circuits and an extension of it will be used to analyze FGMOS translinear loops. Since the dimensioneless current quantities are related to the surface potential the will be called @-currents or ***psi-currents.***

The transconductance from the gate is given by:

$$g_m \equiv \left.\frac{\partial I_{DS}}{\partial V_G}\right|_{V_B,V_S=c} = \frac{\kappa I_{DS}}{V_t} \tag{6}$$

and from the local substrate (backgate) terminal:

$$g_{mb} \equiv \left.\frac{\partial I_{DS}}{\partial V_B}\right|_{V_G,V_S=c} = \frac{(1-\kappa) I_{DS}}{V_t} \tag{7}$$

The conductance g_s at the source is given by:

$$g_s \equiv \left.\frac{\partial I_{DS}}{\partial V_S}\right|_{V_G,V_B=c} = \frac{I_{DS}}{V_t} \tag{8}$$

The transconductances depicted in Equations 6, 7 and **8** are linear functions of the current –to a first order– and hence each MOS transistor in saturation has the equivalent of three different translinear elements. Note, that the source transconductance is equal to the gate transconductance by shorting the local substrate and the source of the transistor ($V_B = V_S$) in which case Equation 4 for the current becomes:

$$I_{DS} = SI_{no}\, e^{\frac{\kappa_n(V_G - V_S)}{V_t}} \tag{9}$$

In subsequent sections, we will see how shorting of the substrate to the bulk, partially circumvents the non-idealities in the translinear properties of MOS transistors and enables the design of near ideal loops.

The translinear properties of the bipolar and MOS transistors in subthreshold are evident in Figure 1. In

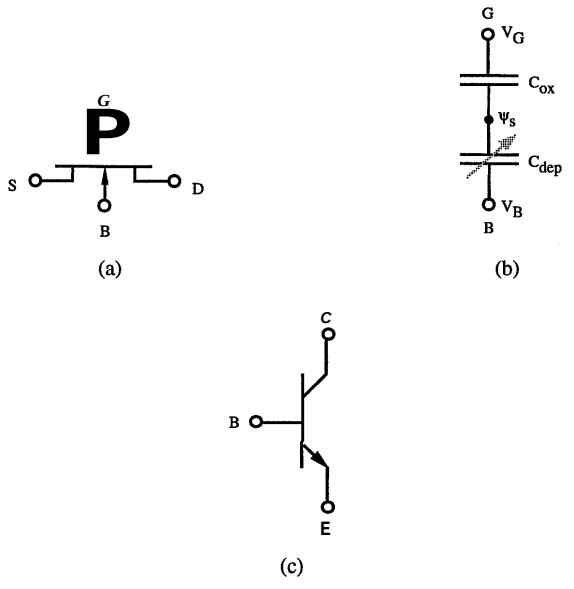

Fig. 2. (a) Symbol for an **NMOS** transistor. The current between the drain and source is controlled by the difference between the surface potential ψ_s and the potential at the source terminal. The surface potential ψ_s is set by the potential at the gate and bulk terminals through the capacitive divider between C_{ox} and C_{dep} shown in (b). An NPN bipolar transistor is shown in (c). The current between the collector and emitter is controlled by the voltage difference between the base and emitter nodes.

a logarithmic current scale, the transfer characteristics show linearity with respect to the controlling voltages. Plots of the normalized transconductance (g_m/I) are shown in Figure 3 and demonstrate how the bipolar device is an ideal translinear element while the MOS transistor in subthreshold only approximates it over a limited range.

3. Translinear Circuits

In this section we discuss translinear circuits that employ translinear elements, both MOS operating in subthreshold and bipolar transistors. We follow the convention proposed by Barrie Gilbert in [3] and make a distinction between a Translinear Loop (TL) and a Translinear Network (TN).

3.1. TranslinearLoops

In "strictly" TLs the translinear principle [1] can be stated as follows:

> In a closed loop containing an equal number of oppositely connected translinear elements, the product of the current densities in the elements connected in the Clockwise (CW) direction is equal to the corresponding product for elements connected in the Counter Clockwise (CCW) direction.

As an example let us consider the circuit of Figure 4 consisting of four ***ideal*** diodes in the loop X–Z–Y–W–**X.** Following the translinear principle, we can write:

$$\frac{\prod_{CCW} J}{\prod_{CW} J} = 1 \quad \text{or} \quad J_2 J_4 = J_1 J_3 \qquad (10)$$

Note that the translinear principle is derived by beginning with Kirchoff's voltage law or the principle of conservation of energy, so that:

$$\sum_{i=1}^{i=N/2} V_{D(2i-1)} - \sum_{i=1}^{i=N/2} V_{D(2i)} = 0 \qquad (11)$$

Equation 10 follows from Equation 11 if the voltages are summed around loops of translinear devices.

In a circuit graph composed of two terminal elements such as ideal diodes (see Figure 4), there is a direct relationship between the voltage difference among each pair of nodes transversed by the translinear device, and the current in the arc that joins the nodes. This is a consequence of having the voltage nodes that control the current be the same as the current–output nodes of the device. In practical systems, the ideal diodes in Figure 4, would correspond to base-emitter junctions of bipolar transistors with shorted collector-base terminals.

Analogous behaviaur can be obtained using translinear three terminal devices such as bipolar transistors, MOS transistors in subthreshold, or any other device that yields diode-like characteristics. However, in three terminal devices, the diode-control nodes in the circuit need not correspond to the current path. In bipolar transistors the diode control nodes are available and thus they can be used explicitely in constraint equations such as Equation 11. This is not true for MOS transistors! **As** we have seen already, one of the diode control nodes, (namely the node corresponding to the surface potential ψ_s) is not ***directly*** accessible. The situation becomes even more complex in MOS transistors with a floating gate (FGMOS) (see Figure 5) coupled to one or more controlling gates (see [19], [20] and references therein). At first sight, the floating gate, appears to make the situation worse, but actually it opens the possibility for a new class of translinear circuits

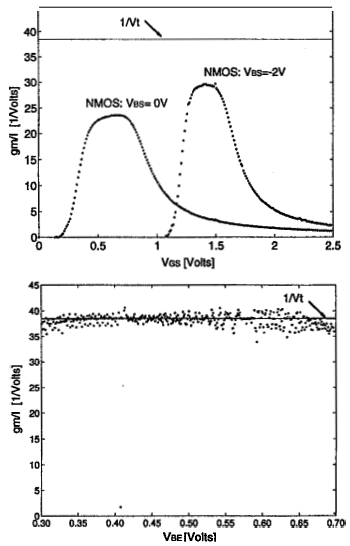

Fig. **3.** Normalized transconductance curves. The transconductance is computed through numerical differentiation of the data in Figure 1, and subsequent smoothing. (Top) For the MOS transistor; (Bottom) for the bipolar transistor. The dramatic decrease of the transconductance in the MOS transistor at low gate-source voltages is attributed to the leakage current.

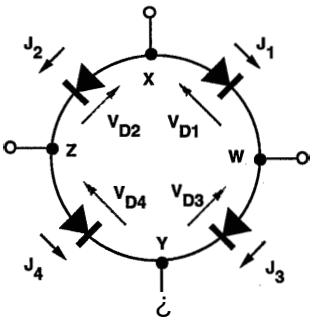

Fig. **4.** A Translinear loop using ideal p-n junctions (diodes).

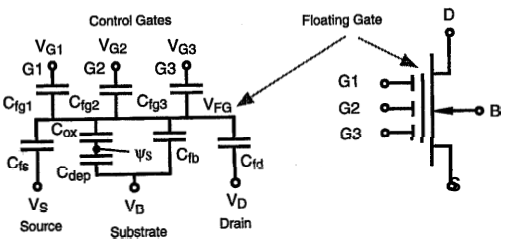

Fig. **5.** Capacitive model and symbol for a floating gate MOS (FG-MOS) transistor. The device depicted in this figure has three control gates G_1, G_2 and $G3$.

proposed and experimentally demonstrated recently by Minch et.al. [20].

Essentially the physical structure of FGMOS transistors offer an extra degree of freedom which can be exploited systematically through another set of constraint equations of the form:

$$V_{FGi} = \frac{Q_{FGi}}{C_{Ti}} + \sum_{j=1}^{j=N} \Lambda_{ij} V_{Gj} \qquad (12)$$

where V_{FGi}, Q_{FGi} are the floating gate voltage and charge on the (ith) transistor and V_{Gj} is the voltage of the (jth) control gate. The total capacitance seen in the floating gate is C_{Ti} and Λ_{ij} is a design parameter that depends on the ratio of the control gate to floating gate capacitance, i.e. $\Lambda_{ij} \equiv C_{fgj}/C_{Ti}$. The details of a systematic analysis procedure for FGMOS translinear circuits can be found in [20].

3.1.1. *Analysis of Translinear Circuits with MOS Transistors in Saturation* The current mirror is a trivial example of a translinear circuit; it has a single loop with two translinear elements, one CCW and the other CW.

Two current mirrors implemented with complementary devices and connected back-to-back yield the circuit shown in Figure 6. This loop includes four three-terminal devices and corresponds to the ideal diode example of Figure 4. The circuit can be readily recognized as a BiCMOS implementation of an AB stage in a digital oriented CMOS process where only one type (NPN) of bipolar transistors is available [23]. A composite structure made of an MOS in subthreshold and an NPN bipolar yields a pseudo–PNP device with good driving capabilities. Translinear loops using both PNP and NPN bipolar transistors were first studied by Fabre [21].

Applying the translinear principle to the loop X–Z–Y–W–X of Figure 6 yields the following constraint

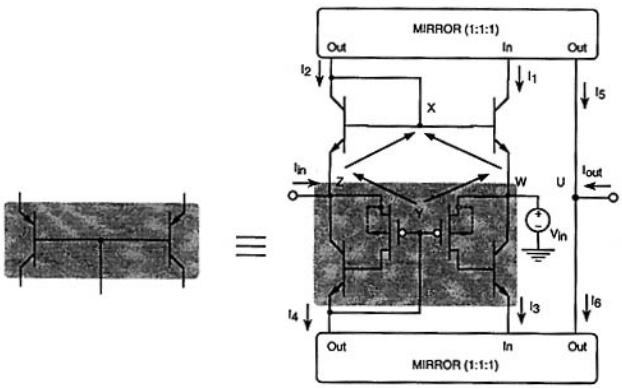

Fig. 6. The translinear loop of Figure 4, implemented using composite bipolar and subthreshold MOS transistors. The loop is employed in a current conveyor configuration where the bidirectional output current I_{out} equals to the bidirectional current I_{in}.

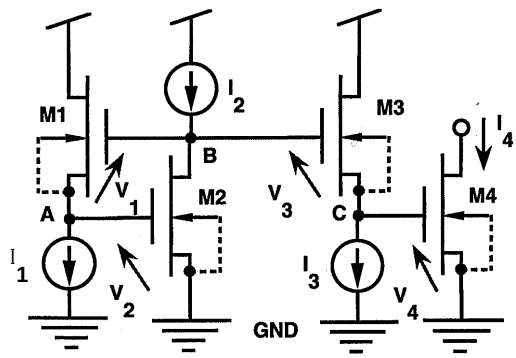

Fig. 7. A translinear circuit that performs one-quadrant normalized multiplication. I_1, I_2 and I_3 are the inputs and I_4 is the output.

equation for the currents in the circuit:

$$I_2 I_4 = I_1 I_3 \tag{13}$$

This classical four junction loop can be combined with two current mirrors to implement a current conveyor [22] where $I_{out} = I_{in}$ and $V_Z = V_{in}$.

Our second example is the MOS transistor one-quadrant multiply-divide circuit shown in Figure 7. A large number of these CMOS multipliers have been employed in the implementation of a correlation-based motion-sensitive silicon retina [24].

Applying the translinear principle to the loop GND–A–B–C–GND, we find a total of four equivalent diode junctions and obtain

$$I_1 I_2 = I_3 I_4 \quad \text{or} \quad I_4 = \frac{I_1 I_2}{I_3} \tag{14}$$

The above relationship can also be derived by summing the voltages around the loop (conservation of energy)

$$V_1 + V_2 - V_3 - V_4 = 0$$

Replacing the gate-source voltages for M_1, M_2, M_3, M_4 with their respective drain-source currents using Eq. (9) (assuming all devices are in saturation, have $V_{SB} = 0$, have negligible drain conductance, have identical κ, have identical I_0 and geometry S), we obtain:

$$\frac{kT}{\kappa q}\left(\ln\left(\frac{I_1}{S\,I_0}\right) + \ln\left(\frac{I_2}{S\,I_0}\right) - \ln\left(\frac{I_3}{S\,I_0}\right) - \ln\left(\frac{I_4}{S\,I_0}\right)\right) = 0$$

or

$$\ln\left(\frac{I_1 I_2}{S\,I_0}\right) = \ln\left(\frac{I_4 I_3}{S\,I_0}\right)$$

from which Eq. (14) readily follows. Note that the assumption of identical κ holds true to a first order because $V_{SB} = 0$ and the gate of all transistors are within a few hundred millivolts from each other.

Yet another way of viewing the function of this circuit is that of a *log-antizog* block. Transistors M_1 and M_2 do the *log-hg*, M_4 does the *antilog-ing* and M_3 is a level shifter.

Another single quadrant multiplier is shown in Figure 8. This circuit was proposed and its function experimentally demonstrated in [29]. The operation of the circuit can be understood by noting that a single transistor (M_4) can perform a single quadrant multiplication because the voltages on the gate and bulk control the current in a multiplicative fashion (see Equation 4). Since in subthreshold the transistors saturate at only a few V_t of drain source voltage, the bulk terminal of the device can be connected to the drain without turning on the bulk-source junction.

An expression for the output current I_4 can be obtained by applying the translinear principle around the four loops (Vdd-A-Vdd), (Vdd-B-Vdd),(Vdd-C-Vdd),(Vdd-D-Vdd) to obtain the following equations

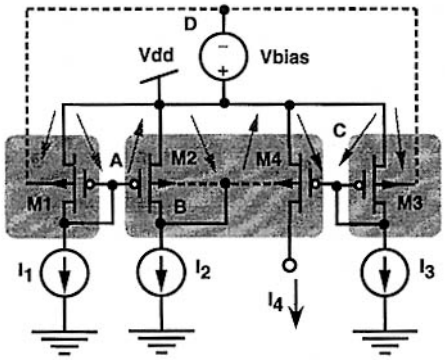

Fig. 8. **A** four transistor translinear circuit that performs a one-quadrant normalized multiplication and exploits the back-gate in an **MOS** transistor. Device pairs $M1$, $M3$ and $M2$, $M4$ share local substrate terminals (in this case n-wells). I_1, I_2 and I_3 are the inputs and I_4 is the output.

for the ***psi-currents*** introduced in Equation 5:

$$\begin{aligned} i_{G1} &= i_{G2} \\ i_{B2} &= i_{B4} \\ i_{G4} &= i_{G3} \\ i_{B3} &= i_{B1} \end{aligned} \tag{15}$$

The actual currents in the four MOSFETs M_1,M_2,M_3,M_4, can be written as a function of the ***psi-currents:***

$$\begin{aligned} I_1 = I_{DS1} &= SI_{no}\, i_{G1}^{\kappa_1}\, i_{B1}^{(1-\kappa_1)} \\ I_2 = I_{DS2} &= SI_{no}\, i_{G2}^{\kappa_2}\, i_{B2}^{(1-\kappa_2)} \\ I_3 = I_{DS3} &= SI_{no}\, i_{G3}^{\kappa_3}\, i_{B3}^{(1-\kappa_3)} \\ I_4 = I_{DS4} &= SI_{no}\, i_{G4}^{\kappa_4}\, i_{B4}^{(1-\kappa_4)} \end{aligned} \tag{16}$$

where the devices have been assumed to have the same S and I_{no}. If now the assumption is made that $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4$, Equations 15 and 16 yield the. following expression for the output current I_4 in terms of the input currents I_1, I_2 and I_3:

$$I_1 I_4 = I_2 I_3 \qquad \text{or} \qquad I_4 = \frac{I_2 I_3}{I_1} \tag{17}$$

In the original implementation [29] it was suggested that to improve accuracy, the voltage on the local substrate (n-well) of devices ***M3*** and M_1 should be set at a value close to that of node ***B***, the local substrate of of devices M_2 and M_4. This is indeed necessary, as the bulk voltage determines κ of all transistors, which

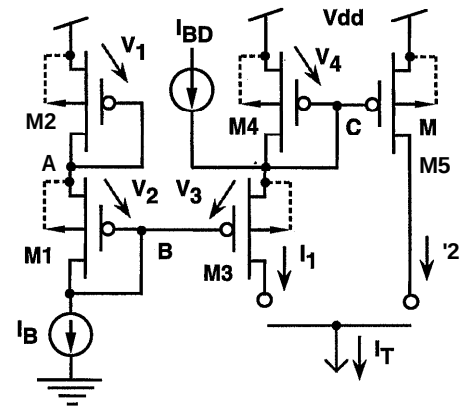

Fig. 9. Circuit that converts a bidirectional current on a single wire into two unidirectional currents on separate wires. This is a current-mode absolute value circuit. The sign of the bidirectional input current is assumed to be positive when it adds positive charge to node ***C***. The bidirectional current I_{BD} is the input, the unidirectional currents I_1 and I_2 are the outputs, and I_B sets the operating point of the circuit.

was assumed to be the same for all transistors. Another implicit assumption here is that the gate voltage is approximately the same for all transistors.

Our next example addresses the problem of converting a bidirectional current to two unidirectional currents which is the equivalent to a current-mode half-wave rectification. A translinear circuit that computes this nonlinear function is shown in Fig. 9. The bidirectional current I_{BD} is steered through transistor M_3 when $I_{BD} > 0$ and through transistors $M_{4,5}$ when $I_{BD} < 0$. Concentrating on transistors $M_{1,2,3,4}$, we identify a loop (VDD-A-B-C-VDD) and apply the translinear principle to yield:

$$\begin{aligned} I_B I_B &= I_1 (I_1 - I_{BD}) \\ I_2 &= I_1 - I_{BD} \end{aligned}$$

M_5 form a simple current mirror.

These equations may be solved for I_1 and I_2 in terms of I_B and I_{BD}:

$$\begin{aligned} I_1 &= \frac{1}{2}\left(I_{BD} + \sqrt{4{I_B}^2 + {I_{BD}}^2}\right) \\ I_2 &= \frac{1}{2}\left(-I_{BD} + \sqrt{4{I_B}^2 + {I_{BD}}^2}\right) \end{aligned} \tag{18}$$

Which shows that $I_1 \simeq |I_{BD}|$ and $I_2 \simeq 0$ when $I_{BD} \gg I_B$ and vice versa. The absolute value is ob-

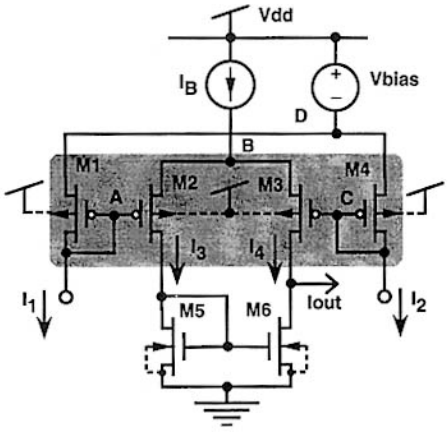

Fig. *10.* **A** translinear circuit that computesthe normalized difference of two current signals. I_1 and I_2 are the inputs and the bidirectional current I_{out} is the output normalized to $I_1 + I_2$.

tained by connecting the two output wires together in which case:

$$I_T \equiv I_1 + I_2 \simeq |I_{BD}| \quad \text{if } I_{BD} \gg I_B$$

This circuit has been employed in a CMOS integration of an autoadaptive linear recursive network for the separation of sources [14].

The next translinear circuit performs a current-ratio computation. This functional block, is part of the readout amplifier in an analog VLSI system that integrates monolithically a one dimensional array of photodiodes and selective polarization film to form a *polarization contrast* retina [25].

The simple translinear circuit in Figure 10 is excellent for rescaling differential current signals and thus computing the contrast. I_1 and I_2 represent currents fi-om two selected photodiodes. The heart of the computation circuit will be recognized as a Gilbert gain-cell [3] implemented in subthreshold MOS.

The analysis of this circuit is typical for translinear circuits that involve differential current signals. Application of the translinear principle around the loop A-B-C-D-A yields:

$$I_1 I_4 = I_3 I_2 \quad \Rightarrow \quad \frac{I_3}{I_1} = \frac{I_4}{I_2}$$

$$\frac{I_3 - I_4}{I_1 - I_2} = \frac{I_3 + I_4}{I_1 + I_2}$$

and

$$AI^* = \frac{I_B}{I_{in}} AI$$

similarly for $\mathbf{A}I^* \equiv I_3 - I_4$, $I_B \equiv I_3 + I_4$. The differential output current $\mathbf{A}I^*$ is a scaled version of the differential input current $\boldsymbol{AI}$. The voltage between node $\boldsymbol{B}$ and Vdd should be such that the current source I_B stays in saturation.

The mirror composed of transistors M5 and M6 converts the unidirectional differential signal $\boldsymbol{AI}$ to the bidirectional signal I_{out} so that:

$$I_{out} \equiv -I_B \frac{\boldsymbol{AI}}{I_{in}}$$

3.1.2. Analysis of Circuits with MOS *Transistors in the Ohmic Regime* In this subsection, we extend the translinear principle to subthreshold MOS transistors operating in the ohmic region. In Appendix A, (see Fig. 25), we show how the source-drain current of a MOS transistor can be decomposed into a source component I_{Q_s} and a drain component I_{Q_d}, and that these components superimpose linearly to yield the actual current $I_{SD} \equiv I_{Q_s} - I_{Q_d}$.

In the ohmic region, these components are comparable. Decomposition and linear superposition may be used to exploit the intrinsic translinearity of the gate-source and gate-drain "junctions." This is the basis for extending the translinear principle to the ohmic region. On the otherhand, in the saturation region, we can exploit the translinearity of the gate-source "junction" directly because the drain component is essentially zero and decomposition is of no consequence.

The translinear circuits based on subthreshold ohmic operation are only possible because of the symmetry between drain and source operation of an MOS transistor. One could argue that decomposition is also possible with bipolar devices. However, while the difference of two exponentials is the *exact* form for MOS devices, it is only an *approximation* for bipolars, due to the fact that the forward and reverse current gains of the device never reach unity [46], [47] (see Equations 45 in Appendix A). This distinction is fundamental and important difference between MOS and bipolar transistors arising from lossless transport in a MOS channel versus lossy transport in bipolars due to recombination in the base. It is possible to use CMOS compatible lateral bipolar transistors as symmetric devices [26] but at the expense of a large base current that increases the power dissipation in the system.

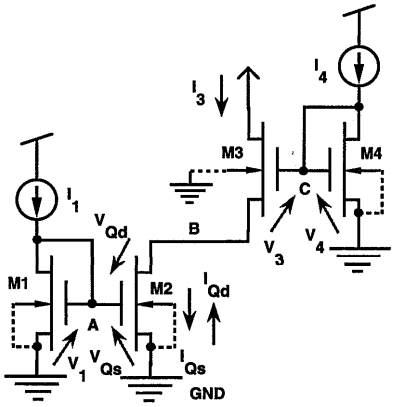

Fig. II. **A** translinear circuit that employs subthreshold **MOS** transistors in saturation and ohmic regime and computes the product of two input currents I_1 and I_2 normalized to $I_1 + I_2$.

To demonstrate the application of the translinear principle to circuits that include MOS transistors in the ohmic regime, consider the one-quadrant current-correlator circuit in Fig. 11. Transistor M_2 operates in the ohmic region. Proper circuit operation requires that the output voltage is high enough to keep *M3* in saturation. This circuit was first introduced by Delbrück [27] and later incorporated in a larger circuit that implements the non-linear Hebbian learning rule in an auto-adaptive network [28], [30] and in a micropower auto-correlation system [3 13.

An expression relating the output current, I_3, to the input currents, I_1 and I_4, can be derived by treating the source-gate and the drain-gate "junctions" of the ohmic device as separate translinear elements and applying the translinear principle. For the two loops formed by nodes GND-A-GND and GND-A-B-C-GND in Fig. 11, we obtain

$$\begin{aligned} I_1 &= I_{Q_s} \\ I_3 I_{Q_s} &= I_4 I_{Q_d} \end{aligned} \tag{19}$$

source-drain current of the MOS transistor can be decomposed into a source component I_{Q_s} and a drain component I_{Q_d}—controlled by their respective "junction" voltages V_{Q_s} and V_{Q_d}. These opposing components superimpose linearly to give the actual current passed by M_2, i.e.,

$$I_{DS2} = I_{Q_s} - I_{Q_d}$$

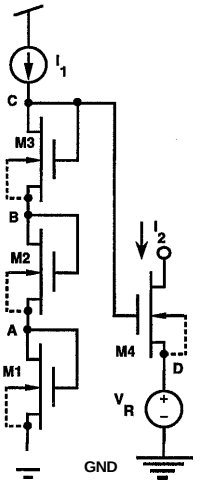

Fig. 12. **A** current-mode circuit -translinear network- that implements a normalized cubic non-linearity. I_1 is the input, I_2 **is** the output and the voltage source V_R normalizes the result.

$$I_3 = I_{Q_s} - I_{Q_d} \tag{20}$$

and (20), the output current is given by:

$$I_3 = \frac{I_1 I_4}{I_1 + I_4} \tag{21}$$

3.2. *Translinear Networks*

In the previous section we have discussed "strictly" translinear loops **(TL)**. Translinear networks [3] differ from translinear loops in that they contain independent voltage sources and the following equation can be employed in their analysis:

$$\frac{\prod_{CCW} J}{\prod_{CW} J} = \mathcal{G}\, e^{E/V_t}$$

or

$$J_2 J_4 \ldots J_N / J_1 J_3 \ldots J_{N-1} = \mathcal{G}\, e^{E/V_t} \tag{22}$$

where E is the independant voltage source and $\mathcal{G}$ is a constant coefficient that lumps device design and fabrication parameters. The above extension to the translinear principle was proposed by Hart [33].

We begin the discussion of TNs using a simple circuit that has the topology of a current mirror and incorporates a voltage source in the loop (see Figure 12). If the

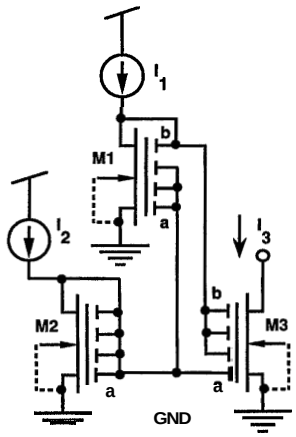

Fig. 13. **A** translinear circuit using FGMOS transistors to compute the ratio of a cubic to square functions. I_1 and I_2 are the inputs and *13* is the output.

input current is I_1, the output current is I_2, and V_R is a constant voltage source, application of the translinear principle around the loop (GND-A-B-C-D-GND) yields:

$$I_2 = \frac{{I_1}^3}{\mathcal{G}\, e^{V_R/V_t}} \tag{23}$$

The voltage source is necessary for circuit operation and it normalizes appropriate the output current. This circuit has been employed in a small system that implements the Herault-Jutten independent component analyzer [14].

An FGMOS-based circuit that has the same functionality as the circuit in Figure 12, is shown in Figure 13. We begin the analysis of the circuit by noting that the current in the channel of an FGMOS, is controlled by multiple gates that can be thought as extensions to the front gate of the transistor. **As** such, Equation 5 can be re-written for an N-input **NMOS** transistor as:

$$I_{DS} \equiv S\, S_Q\, I_{no}\, i_B^{(1-\kappa_n)} \prod_{j=1}^{j=N} i_{Gj}^{\kappa_n/N} \tag{24}$$

where it has been assumed that all N gates of the i-th transistor have the same strength, i.e. have the same coupling capacitance to the floating gate. The charge Q_{FG} on the floating gate is incorporated through a geometry related multiplicative constant S_Q so that when the charge Q_{FG} is zero, $S_Q = 1$.

An expression for the output current I_3 can be obtained by applying the translinear principle around three loops that include the floating gates and source of transistor $M3$ together with the floating gate nodes and sources of the other transistors. When the three loops are traversed, the following equations for *psi-currents* are obtained:

$$\begin{aligned} i_{3a} &= i_{2a} \\ i_{3a} &= i_{1a} \\ i_{3b} &= i_{1b} \end{aligned} \tag{25}$$

We have adopted a notation where for example, i_{3a} denotes the *psi-current* in device 3 controlled by the voltage on its gate a. Using Equation 24, the current at the source of M_1, M_2, M_3, can be expressed as functions of *psi-currents* so that:

$$\begin{aligned} I_1 &= I_{DS1} = S\, S_{Q1}\, I_{no}\, i_{1a}^{3\kappa_1/4}\, i_{1b}^{\kappa_1/4} \\ I_2 &= I_{DS2} = S\, S_{Q2}\, I_{no}\, i_{2a}^{\kappa_2} \\ I_3 &= I_{DS3} = S\, S_{Q3}\, I_{no}\, i_{3a}^{\kappa_3/4}\, i_{3b}^{3\kappa_3/4} \end{aligned} \tag{26}$$

where the devices are assumed to have the same S and I_{no}; the back-gate contribution to the current in each device is eliminated as all transistors have the source shorted to the substrate. Now, by making the assumption that $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4$, and no charge on the floating gates, Equations 25 and 26 yield the following expression for the output current I_3 in terms of the input currents I_1 and I_2:

$$I_3 = \frac{I_1^3}{I_2^2} \tag{27}$$

The assumption of equal κ is reasonable so long the voltage on the floating gate is such that all devices stay in subthreshold. An alternative way of obtaining a functional description of the circuit can be found in the paper by Minch et. al [20].

A TN that incorporates bipolar transistors, an MOS in subthreshold, and an independent voltage source V_{XY} is shown in Figure 14. We will use Equation 22 to derive the relationship between I_1, I_2, I_{ZW} and V_{XY}. The MOS transistor is assumed to be ideal with $\kappa = 1$. A discussion of networks with non-ideal devices will be done in the next section.

A ratio relationship between I_{ZW} and $I_1 - I_2$ can be derived by employing Equation 22 applied to the two loops (X–Z–Y–X) and (X–Y–W–X) to yield the following equations:

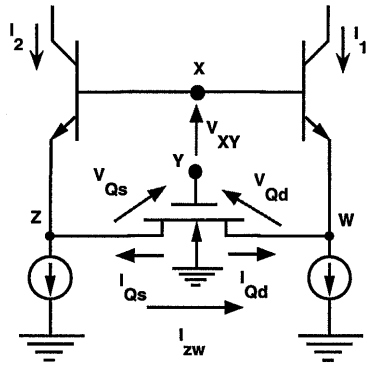

Fig. 14. A BiCMOS translinear network that exploits the MOS subthreshold ohmic characteristics.

$$I_{Q_s} = I_2\, \mathcal{G}\, e^{-V_{XY}/V_t}$$
$$I_{Q_d} = I_1\, \mathcal{G}\, e^{-V_{XY}/V_t} \tag{28}$$

Using the adopted conventions for current decomposition in **NMOS** transistors (see Figure 25), I_{Q_s} and I_{Q_d}

$$I_{ZW} = I_{Q_d} - I_{Q_s}$$

together with Equation 28 we obtain the following expression that relates the currents in the circuit.

$$I_{ZW} = (I_1 - I_2)\, \mathcal{G}\, e^{-V_{XY}/V_t} \tag{29}$$

It is immediately apparent that the current ratio $I_{ZW}/(I_1 - I_2)$ can be controlled both by a fixed parameter ($\mathcal{G}$) that is designed prior to the fabricaition of the circuit and a variable quantity (e^{-V_{XY}/V_t}) that can be programmed (post-fabrication) during circuit/system operation. This property will be utilized in the design of **linear** MOS transistor-only spatial averaging networks.

3.2.1. Translinear Spatial Averaging Networks Often, models of neural computation necessitate the realization of spatial averaging networks [7]. To demonstrate the analogies between linear and translinear networks as well as their subtle and important differences, we begin with networks that employ linear conductances, voltages and currents and contrast them with translinear current-mode [16] networks.

A voltage–mode circuit model for a loaded network is shown in Figure 15(left) for which:

$$I_{PQ} = (G_1/G_2)(I_Q - I_P)$$

This is a lumped parameter model where G_1 and G2 correspond to resistances per unit length. The voltages on nodes ***P*** and Q referenced to ground, represent the state of the network and can be read out using a differential amplifier with the negative input grounded.

The equivalent circuit using idealized non-linear conductances is shown in Figure 15(right). The difference in currents through the diodes D_1 and D_2 are linearly related to the current through the *diffusor* MOS transistor.' This relationship can be derived from Equation 55 describing subthreshold conduction, and the ideal diode characteristics where $I_D = I_S \exp[V_D/V_t]$. An expression can be derived for the current I_{PQ} in terms of the currents I_P and I_Q, the reference voltage V, and the bias voltage V_C, when diodes are replaced by transistors:

$$I_{PQ} = \left(\frac{SI_{n0}}{I_S}\right) \exp\left[\frac{\kappa_n V_C - V_r}{V_t}\right](I_Q - I_P) \tag{30}$$

The current I_{n0} and S is the zero intercept current and geometry factor respectively for the diffusor transistor M_h. I_S is the reverse saturation current for the diode that is assumed to be ideal. The currents in these circuits are identical if

$$\frac{G_1}{G2} = \left(\frac{SI_{n0}}{I_S}\right) \exp\left[\frac{\kappa_n V_C - V_r}{V_t}\right]$$

Increasing V_C or reducing V, has the same effect as increasing G_1 or reducing G_2. The state of this network is represented by the charge at the nodes ***P*** and Q. Since the anode of a diode is the reference level (zero negative charge), the currents I_P and I_Q represent the result. Unfortunately, the anode of a diode or a diode connected transistor is not a good current source.

When diodes are not explicitly available in the process, diode connected PMOS or NMOS transistors can be used as shown in Figure 16. When the loads are PMOS, the current current I_{PQ} is given in terms of voltages normalized to (kT/q):

$$I_{PQ} = \left(\frac{S_h I_{n0}}{S_v I_{p0}}\right) \exp(\kappa_n v_C - \kappa_p v_r)({I_Q}^{1/\kappa_p} - {I_P}^{1/\kappa_p}) \tag{31}$$

When NMOS transistors are used as loads, there is the additional benefit, that of exploiting the current conveying properties of a single transistor [16], to obtain the current outputs I_P and I_Q, on nodes that are low

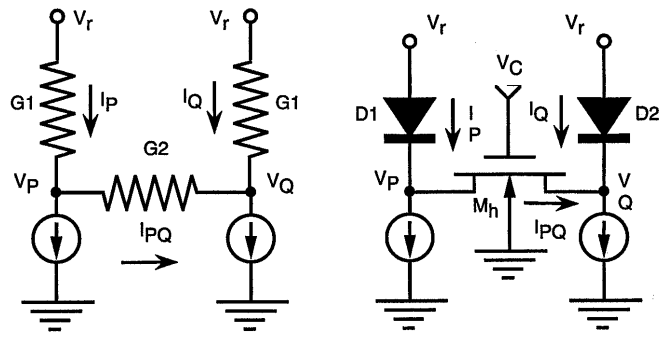

Fig. 15. Building blocks for linear loaded networks. Using segments that employ ideal (left) linear and (right) non-linear elements.

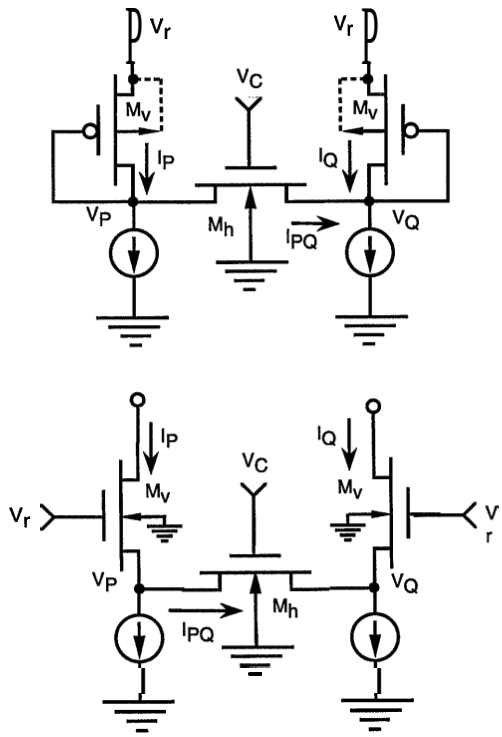

Fig. 16. Current-mode building blocks for linear loaded networks using (top) **PMOS** transistor implementation, (bottom) **NMOS** single transistor current-conveyor implementation.

conductance (the drain terminal are now excellent outputs for the currents). Using Equations 8.45 in [16], the current I_{PQ} is given as:

$$I_{PQ} = \left(\frac{S_h}{S_v}\right) \exp(\kappa_n v_C - \kappa_n v_r)(I_Q - I_P) \qquad (32)$$

where S_h and S_v are geometry parameters for transistors M_h and M_v, respectively.

The one dimensional MOS transistor-only network corresponding to the Helmholtz equation shown in Figure 17 can model the averaging that occurs at the horizontal cells layer of the outer retina. This is equation is the basis of the well known silicon retina architecture proposed by Mahowald and Mead [34], [7].

Summing the currents at node j we get:

$$I_j^* = I_{ij} - I_{jk} + I_j \qquad (33)$$

Using the results from the previous section for the currents I_{ij} and I_{jk} given by Equation 32 substituted in Equation 33 yield:

$$I_j^* = I_j + \left(\frac{S_h}{S_v}\right) \exp(\kappa_n v_C - \kappa_n v_r)(2I_j - I_i - I_k) \qquad (34)$$

Normalizing internode distances to unity the above equation can be written on the continuum as:

$$I^*(x) = I(x) + \lambda \frac{d^2 I(x)}{dx^2}$$

This equation yields the solution to the following optimization problem: Find the smooth function $I(x)$ that best fits the data $I^*(x)$ with the minimum energy in its first derivative. Input is the currents $I^*(x)$ and output the currents $I(x)$.

The parameter $\lambda \equiv \left(\frac{S_h}{S_v}\right) \exp(\kappa_n v_C - \kappa_n v_r)$ is the cost associated with the derivadve energy—relative to the squared-error of the fit.

The diffusive network in Fig. 17 was recently described in terms of "pseudo-conductances" [35]. We

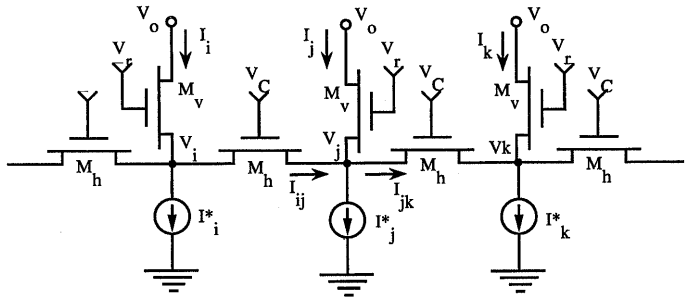

Fig. 17. A one-dimensional **MOS** translinear network to perform local aggregation –spatial averaging–. The back-gate terminals of all devices are connected to the substrate.

have used the chargelcurrent-based formulation first proposed in [15] to explain its behaviour. This current-mode approach relies an intuitive understanding of the device physics and yielded the insight which enabled us to extend the translinear principle to subthreshold MOS transistors in the ohmic region as well as the decomposition of the current into dimensionless components corresponding to ideal junction. We now have a comprehensive ***current-mode*** approach for analyzing subthreshold MOS circuits. The essence of this approach is the representation of variables and parameters by charge, current, and diffusivity. Voltages and conductances are not used explicitly.

Bult and Geelen proposed an identical network for linear current division above-threshold and used it in a digitally-controlled attenuator [36]; they also analyzed its subthreshold behavior. However, they stipulate that all gate voltages must be identical and control the division by manipulating the geometrical factor W/L of the devices. We have shown here, and previously in [15], that this constraint may be relaxed in subthreshold without disrupting linear operation. This is a real bonus because it allows us to modify the divider ratio or space constant of the network after the chip is fabricated by varying $(V, - V_r)$. Tartagni et al. have demonstrated a current-mode centroid network [37] using subthreshold MOS devices whose operation is described by the current division principle.

3.3. *A general result for MOS translinear loops*

Three of the circuits discussed in the previous subsection, namely the translinear multiplier of Figure **8,** the MOS implementation of the Gilbert gain stage in Figure 10, and the current correlator (Fig. 11) have been experimentally shown to exhibit near "exact" translinear behaviour even though they are build from MOS transistors and they do not have their source connected to the local substrate.

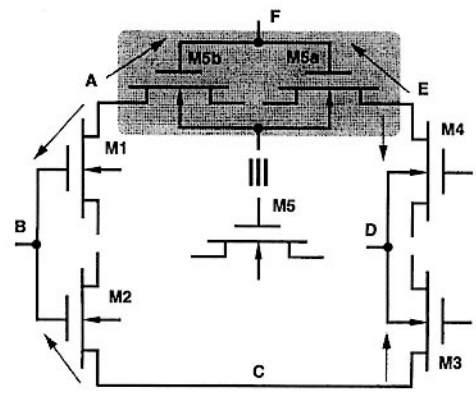

Fig. 18. Translinear loop composed of five **MOS** transistors in subthreshold. All devices are in saturation except device M_5 which is in the ohmic regime and therefore can be decomposed as two devices in saturation, back to back sharing same gate and substrate.

A recent result by Eric Vittoz [32] can be employed to partially explain this rather surprising behaviour. He considers translinear loops constructed from MOS transistors in subthreshold saturation with common substrate connection (similar to the one shown in Figure 18). If the pairing of transistors in the CW and CCW direction, is such that they have their gates connected to gates and sources connected to sources and they are alternated (much like even-numbered and odd-numbered devices in Figure 18) Vittoz shows that the translinear loop does not suffer from the MOS transistor non-ideal translinear behaviour. He notes also that loops containing transistors in the ohmic regime can also be included in this formulation as they can be decomposed as two parallel connected saturated de-

vices sharing common gate and common substrate (see Figure *25.)*

However, to account for the near "exact" operation of the multiplier in Figure 8, Vittoz's argument must be extended to include loops that go through the back gate of the MOS transistors as illustrated in Figure 18. The global substrate restriction can thus be removed and replaced by a local substrate connection, and the result still holds true. In a standard CMOS process, this will of course be possible only for one type of devices.

Now, we will re-examine the operation of the circuits in Figure 10, Figure 8, and Figure 1l.

Consider the largest loop (A–B–C–D–A) in Figure 10. Devices M_1 and M_2 have common gate and common bulk and so do devices $M3$ and M_4. When adjacent devices are paired in different ways we observe that $M3$ and M_2 share the same source and bulk which is the case also for M_1 and M_4.

In the the largest loop (A–B–C–D–A) of translinear multiplier circuit of Figure 8 in we can verify that transistors M_1 and M_2 as well as $M3$ and M_4 share common gate and source. The alternative pairing, finds M_4 and M_2 sharing same bulk and source which is also the case for $M3$ and M_1.

When devices in the loop are operating in the ohmic regime, such as M_2 in the circuit of Figure 11, we can verify that the loop (GND–B–C–GND) incorporates two adjacent sets of devices $M3$ and M_2 share same bulk and source/drain while $M3$ and M_4 share bulk and gate; the bulk in this circuit is the same for all devices.

3.4. Translinear circuit dynamics

The dynamics of translinear circuits and systems have not been discussed in this paper. However, it was pointed out in [45], that in networks with non-linear conductances without complementary non-linear reactances, the state equations that describe the dynamics of the system are non-linear. Given an architecture and a particular network, a method was outlined to test for stability [45].

4. A Translinear System: A Contrast Sensitive Silicon Retina

Image acquisition under naturally occuring, uncontrolled lighting conditions is required by autonomous robotis, in prosthetic devices for the blind, and autonomous motor-vehicle navigation. Today this task is accomplished in two separate steps. First the light intensity is recorded through a standard imager such as a CCD camera. The intensity field is subsequently processed outside the camera to discard any absolute luminance information and form a representation where only relative illumination, i.e. *contrast,* is retained. Additional processing such as edge extraction and or low bit–rate encoding may follow. However, even though the precision necessary for these tasks rarely exceeds 8 bits, the signal itself has a very large dynamic range, many orders of magnitude, which makes the problem difficult. This issue becomes acute when the illumination varies within a single frame something not uncommon in natural scenes (see Figure 19). The detrimental effects of non–uniform illumination in the performance of a face recognition system have been investigated experimentally by Buhman, Lades and Eeckman [38].

We will now present one solution to the problem of robust image acquisition and preprocessing under variable illumination conditions: a *neuromorphic* analog VLSI silicon retina. This is a contrast-sensitive edge-enhancing imager that includes a rudimentary, yet effective, *local gain control* mechanism at the pixel level. The architecture is inspired by the processing performed in the outer plexiform layer of the vertebrate retina [15], [17]. The resulting image captured with such a system is shown in Figure 20.

The biologically motivated solution is attractive from a computational perspectiye because *contrast,* an invariant representation of the visual world, has been obtained with a front-end that is robust, small, and extremely low power (a few mW). There is also another benefit; the output representation has limited range and therefore subsequent processing/communication stages are not burdened with handling and processing signals of wide dynamic range. A performance comparison between the contrast sensitive silicon retina front end [15] and a conventional camera, in a face recognition experiment is reported in [38].

4.1. Biological Organization

The analog silicon system in the core of the array is modeled after neurocircuitry in the distal part of the vertebrate retina—called the outer-plexiform layer. Figure 21 illustrates interactions between cells in this layer [39]. The well-known center/surround receptive field emerges from this simple structure, consisting of just two types of neurons. Unlike the ganglion cells

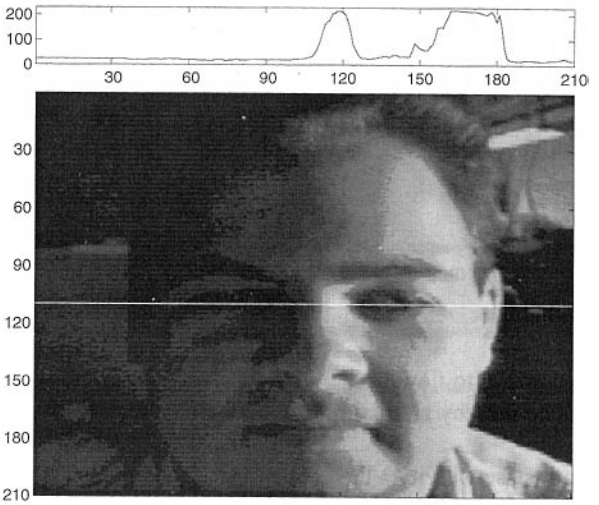

Fig. 19. (Bottom) "Mark" as captured by a conventional camera. (Top) Intensity profile at image line **110** (white line). The light source is positioned to the right side of the image and it introduces a large gradient in illumination within a single frame. This is clearly shown in the intensity histogram. The dynamic range of the scene exceeds the dynamic range of the camera. Aperture control on the camera provides a rudimentary global gain control mechanism. Information in this image is lost at this very first step because there is no gain control (adaptation) at the pixel level.

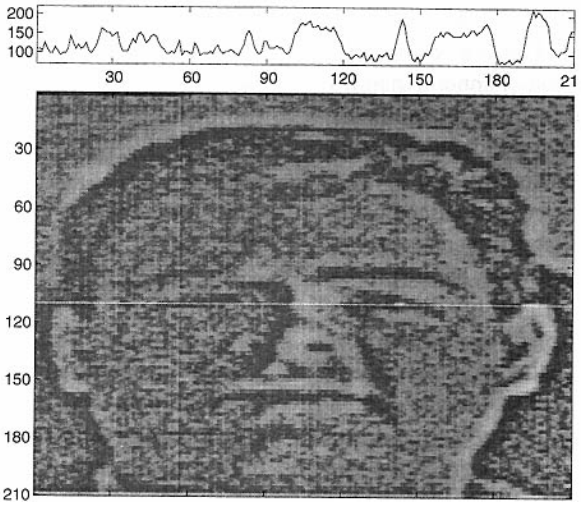

Fig. 20. (Bottom) "Mark" as captured by the translinear silicon retina. (Top) Histogram for the output of the system. The light source is again positioned to the right side of the image and it introduces a large gradient in illumination within a single frame. The image captured by the silicon retina discards absolute illumination and preserves only local contrast information through local gain control at the pixel level. Unlike the image in Figure 19, the presence of a large illumination gradient does not degrade image acquisition here.

in the inner retina and the majority of neurons in the nervous system, the neurons that we model here have graded responses (they do not spike); thus this system is well-suited to analog VLSI.

The photoreceptors are activated by light; they produce activity in the horizontal cells through excitatory chemical synapses. The horizontal cells, in turn, suppress the activity of the receptors through inhibitory chemical synapses.. The receptors and horizontal cells are electrically coupled to their neighbors by electri-

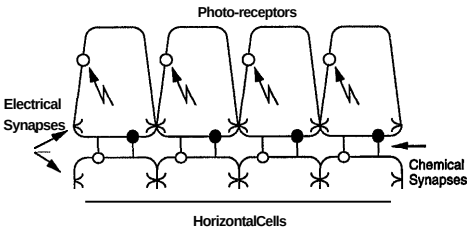

Fig. 21. One-dimensional model **of** neurons and synapses in the outer-plexiform layer. Based on the red-cone system in the turtle retina.

cal synapses. These allow ionic currents to flow from one cell to another, and are characterized by a certain conductance per unit area.

In the biological system, contrast sensitivity –the normalized output that is proportional to a local measure of contrast– is obtained by shunting inhibitiori. The horizontal cells compute the local average intensity and modulate a conductance in the cone membrane proportionately. Since the current supplied by the cone outer-segment is divided by this conductance to produce the membrane voltage, the cone's response, ,will be proportional to the ratio between its photoinput and the local average, i. e. to contrast. This is a very simplified abstraction of the complex ion-channel dynamics involved. The advantage of performing this complex operation at the focal plane is that the dynamic range is extended (local automatic gain control).

The biological system, is mapped onto silicon using circuits of minimal complexity that exploit native properties of subthreshold MOS transistors. High computational throughput at low levels of energy dissipation is achieved by employing *low precision* analog processing in a massively parallel analog architecture that exploits the translinear properties of subthreshold MOS in saturation and ohmic regime.

4.2. Silicon System Architecture

The core of the silicon retina is an array of pixels with a six-neighbour connectivity (see Figure 22). The wiring is included in the layout of the cell (see Figure 24) so that they may be tiled in a hexagonal tesselation to form the focal plane processor. This is a mesh processor architecture where two layers of processors, C and ***H***, communicate both intra and inter layer through local paths. This parallel processing scheme features locality of reference and thus minimizes communication costs.

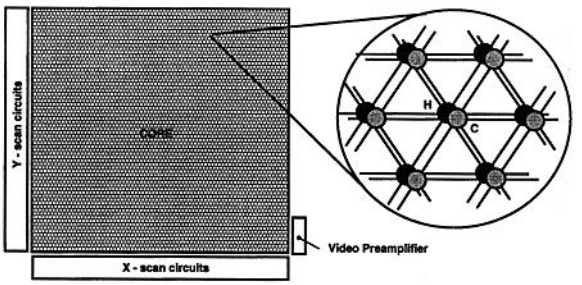

Fig. *22.* Floorplan and system organization. It comprises of two functional components, the core, and the support circuitry. Focal plane processing is performed in the core area.

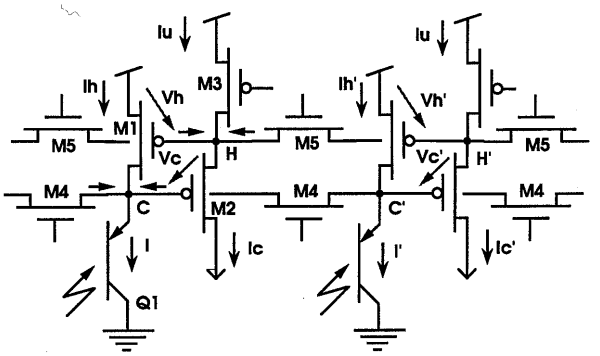

Fig. 23. One-dimensional implementation of outer-plexiform retinal processing. There are two diffusive networks implemented by transistors $M4$ and M_5, which model electrical synapses. These are coupled together by controlled current-sources (devices M_1 and M_2) that model chemical synapses. Nodes ***H*** in the upper layer correspond to horizontal cells while those in the lower layer ***(C)*** correspond to cones. The bipolar phototransistor Q_1 models the outer segment of the cone and $M3$ models a leak in the horizontal cell membrane. Note that the actual system has a six neighbor connectivity.

Support circuitry in the periphery extracts the data from the core and interfaces with the display. The chip incorporates a video pre-amplifier and some digital logic for scanning the processed images out of the array. This circuitry is discussed in detail in the paper by Mead and Delbrück [41]. Standard **NTSC** video is produced off-chip using an **FPGA** controller and a video amplifier.

The basic analog MOS circuitry for a one dimensional pixel with two neighbor connectivity is shown in Figure **23.** We begin with the non-linear aspects of system operation, its *contrast sensitivity.* The non-linear operation that leads to a local gain-control mechanism in the silicon system is acheived through a mechanism that is qualitatively similar to the biological counterpart, but quantitatively different (see discus-

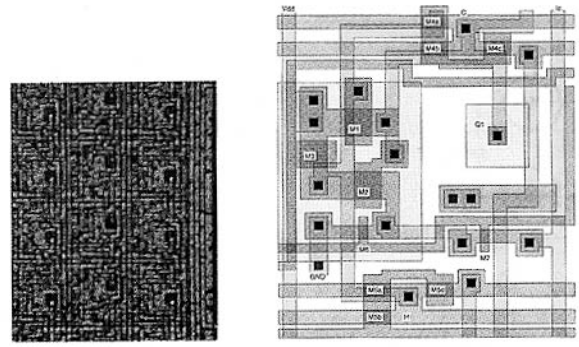

Fig. 24. (Left) Photomicrograph of the chip. The surface is covered by second metal except where there are openings for the phototransistors (the dark square areas). Note the hexagonal connectivity of the pixels. (Right) Layout df the basic cell.

sion in [15]). Refering to Figure 23, the output current $I_c(x_m, y,)$ at each pixel, can be given (approximately) in terms of the input photocurrent $I(x_m, y,)$ and a local average of this photocurrent in a pixel neighborhood (M,N). This region may extend beyond the nearest neighbor. The fixed current I_u supplied by transistor *M3* normalizes the result and Ψ is a parameter.

$$I_c(x_m, y_n) = I_u \frac{I(x_m, y_n)}{\left(I(x_m, y_n) + \Psi \sum_{\mathrm{M,N}} I(x_i, y_j)\right)} \quad (35)$$

At any particular intensity level, the outer-plexiform behaves like a linear system that realizes a powerful second-order regularization algorithm [40] for edge detection. This can be seen by performing an analysis of the circuit about a fixed operating point. To simplify the equations we first assume that $\hat{g} = \langle I_h \rangle g$, where $\langle I_h \rangle$ is the local average. Now we treat the diffusors (devices M_4) between nodes C and C' as if they had a fixed diffusitivity $\hat{g}$. The diffusitivity of the devices M_5 between nodes H and H' in the horizontal network is denoted by h. Then the simplified equations describing the full two-dimensional circuit on a square grid are:

$$I_h(x_m, y_n) = I(x_m, y_n) + \hat{g} \sum_{\substack{i=m\pm1 \\ j=n\pm1}} \{I_c(x_i, y_j) - I_c(x_m, y_n)\}$$

$$I_c(x_m, y_n) = I_u + h \sum_{\substack{i=m\pm1 \\ j=n\pm1}} \{I_h(x_m, y_n) - I_h(x_i, y_j)\}$$

Using the second-difference approximation for the laplacian, we obtain the continuous versions of these equations

$$I_h(x, y) = I(x, y) + \hat{g} \nabla^2 I_c(x, y) \quad (36)$$

$$I_c(x, y) = I_u - h \nabla^2 I_h(x, y) \quad (37)$$

with the internode distance normalized to unity. Solving for $I_h(\mathrm{x}, \mathrm{y})$, we find

$$\hat{g} h \nabla^2 \nabla^2 I_h(x, y) + I_h(x, y) = I(x_i, y_j) \quad (38)$$

This is the ***biharmonic*** equation used in computer vision to find an optimally smooth interpolating function $I_h(\mathrm{x}, \mathrm{y})$ for the noisy, spatially sampled data $I(xi, y_j)$; it yields the function with minimum energy in its second derivative [40]. The coefficient $\mathrm{h} = \hat{g}h$ is called the regularizing parameter; it determines the trade–off between smoothing and fitting the data.

4.3. Layout Considerations

The two-layer architecture for the silicon retina can be accomodated in a cell area of 80h x 94h using a single poly two metal technology. In the implementation reported in [15] and here, a double poly, double metal technology is used and the cell area is 66h x 73h. First metal and polysilicon wires are used for interconnects; second metal is used to cover the entire array, shielding the substrate fi-om undesirable photogenerated carriers. Transistors are implemented using both polysilicon layers.

The system has been fabricated with 230 x 210 pixels on a 9.5 x 9.3 mm die in a 1.2μm n–well double metal, double poly, digital oriented CMOS technology. The chip incorporates 590,000 transistors in the 48,000 pixels and support circuitry, with the core operating in subthreshold/transition region consuming less than 100mW.

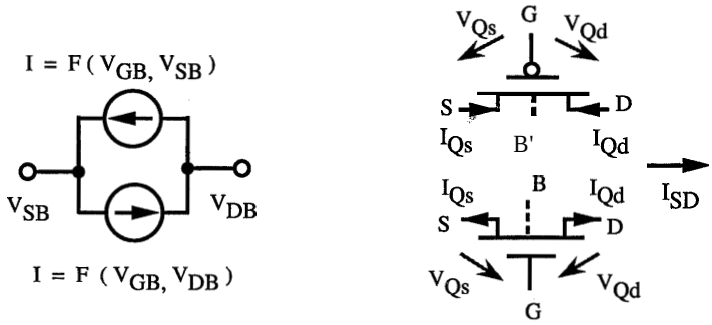

Fig. 25. Large signal model for an NMOS transistor (left). Adopted conventions for current decomposition in PMOS and NMOS devices (right).

4.4. Discussion

The silicon retina, presented in this paper is essentially an *analog floating-point* processor. As a first step, the system computes the range (the voltages on the H nodes correspond to the value of the exponent in floating-point data representation). This is the operating point of the system and is a function of the spatial coordinates and this is how **local** automatic gain control is achieved. At an operating point, sophisticated spatial filtering is performed to smooth the sampled data and enhance the edges. Having separated the problem of precision and dynamic range, the signal processing within the range can be done with low precision analog hardware.

The benefit of a robust architecture on the ultimate system performance is evident in the design of the silicon retina. The regularization properties inherent in the architecture mitigate the intrinsic random variations in the device characteristics, leading to robust performance. This methodology allows analog VLSI implementations using the poorly matched, small geometry, nano-power devices available in garden-variety digital VLSI CMOS technology. Thus we see how a neuromorphic architecture can account for the properties of the computational substrate, and yield robust operation in the presence of noise (structural variability) in the individual devices, The translinear propeties of the MOS transistor in subthreshold ohmic and saturation are key to an area efficient implementation which is commensurate with integration at the focal plane.

5. Subthreshold MOS Device Limitations

As we have seen earlier, linear relationships between conductance/transconductances and current, and the equivalence of controlling the current from the collector and emitter terminals, are the basis of bipolar translinear circuit design. This was also shown to be applicable to MOS transistors in subthreshold as well. The absence of a base current makes it in a sense an "ideal" element.

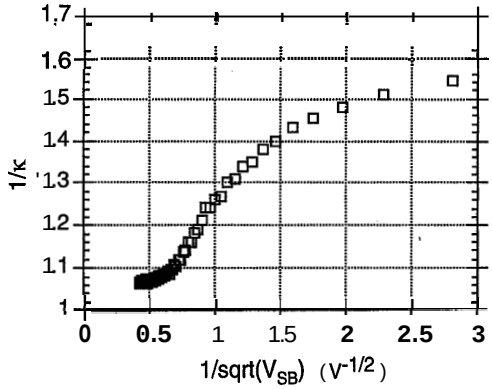

Fig. 26. Experimentally determined variation of parameter κ as a function of the substrate voltage V_{SB}.

However, one must be aware of certain characteristics of the MOS transistor that have a detrimental effect on circuit behaviour. Some of them are already depicted in the large signal model equations. Device characteristics that are not modeled in the simplified transistor model are discussed with the help of experimental data from measurements on transistors operating in subthreshold.

5.1. Transconductance Limitations

Unlike bipolar transistors, the gate and the source of an MOS transistor are not equally effective in controlling the current. Increasing their voltages by the same amount, does not change the current by the same

amount because the transconductances are different (compare Eqs. (6) and (8)). Whereas changing the source voltage changes the the barrier by an equal amount, only a certain fraction (κ) of changes in the gate voltage affects the surface potential and hence the barrier height. This behaviour is depicted in Figure 1 and manifests itself as a slope for the current characteristics that has lower value compared to that of a bipolar transistor. An even lower value $(1-\kappa)$ for the slope is seen from the back gate terminal.

The experimental data from Figure 1 suggest that κ can be pushed closer to unity by biasing the device with a large surface potential (i.e. large V_{SB} and V_{GB}). The dependence of κ on the substrate voltage is measured and plotted in Figure 26. As the substrate is reversed bias, the depletion capacitance $C^{\diamond}_{dep}$, in Equation 58 becomes smaller and hence the gate has larger influence on the channel conductance. Under these conditions, the subthreshold MOS device becomes closer to an ideal translinear element whose transfer characteristics come closer to those of bipolars.

Another way to circumvent the potential divider problem is to move the local substrate voltage together with the source voltage—which is exactly equivalent to increasing V_{GB} by the same amount and does not change V_{SB}. In practice, this may be achieved simply by shorting the source to the local substrate. Clearly, this is only possible with devices in separate wells. For this reason, it is preferable to bias the device at large values of surface potential as described in the previous paragraph, since that works for both types of transistors and is more area efficient (there is no need for separate wells and a triple well process).

Another departure from the simplified model equations presented earlier is evident when we consider the dependence of the drain current on the voltage at the source of the transistors. According to the model (equation 55) this dependence should follow the exponential law and the parameter in the exponent is the inverse of the thermal voltage. An experiment to verify this was conducted and the results are shown in Figure 27 where the slope of the curve has the value of only 35.4 V^{-1}. This suggests that the source conductance cannot be adequately described through Equation 8 and that another parameter $\eta \approx 0.9$ must be introduced in front of the source and drain voltages in the original device Equation 55. Both NMOS and PMOS transistors in different fabrication processes show similar behaviour that is correlated to the zero bias leakage current of the junctions.

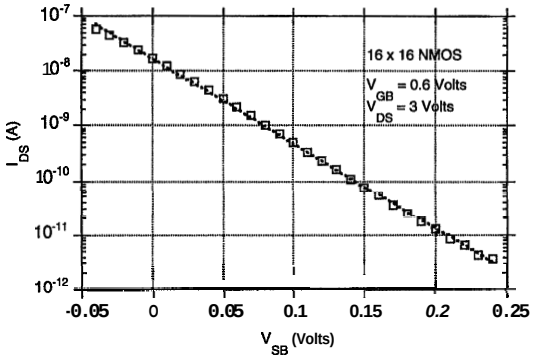

Fig. **27.** Measured drain current I_{DS} as a function of the source voltage V_{SB} for a $(16 \times 16\mu m)$ NMOS transistor fabricated in a $2pm$ n-well CMOS process. The solid line is an exponential function fit to the data. The experiments were performed at a temperature such that the thermal voltage $V_t = 0.0259$ Volts and hence $1/V_t = 38.5V^{-1}$.

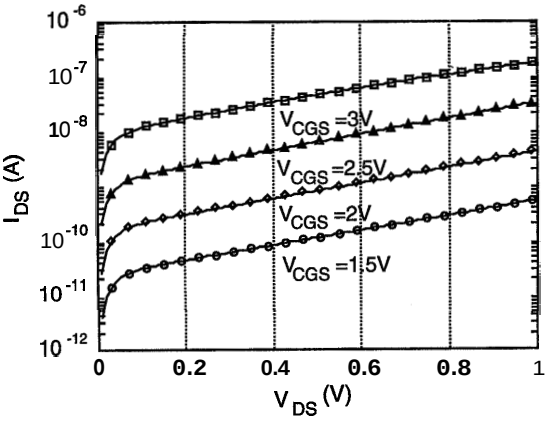

Fig. 28. Drain saturation characteristics of an NMOS FGMOS transistor. The dimensions of the device are $W = 8pm$ and $L = 4\mu m$ and the first-poly second-poly coupling area is $2pm \times 2\mu m$. Note that the current is plotted in a logarithmic scale.

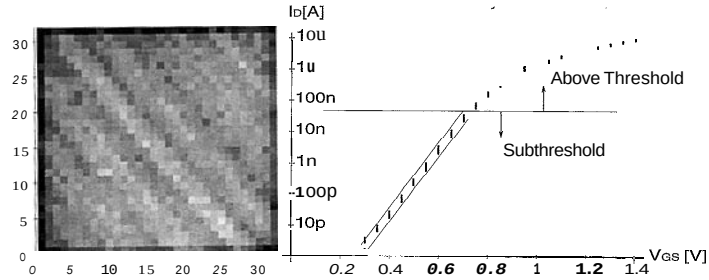

Fig. 29. (Left) Density plots of currents in a ***32 x 32*** array of ***4pm x 4pm*** **NMOS** transistors. Each transistor is represented by a square pixel. Current level is coded by the shade of gray, where the minimum and maximum values are represented by black and white, respectively. The current at a nominal current level of 100nA is obtained by setting V_{GS} to be the same for all transistors in the array. The devices are biased in saturation. (Right) Measured drain current I_D versus gate-source voltage V_{GS} for ***32*** small geometry transistors ($4 \times 4\mu m$) fabricated in a ***2pm*** n-well **CMOS** process; drain-source voltage of V_{DS}=1.5 Volts. The fuzziness in the current, (mismatch between devices), is constant in subthreshold (on a $\log(I)$ scale) and decreases as the device enters the transition and above threshold regime.

5.2. Output Conductance Limitations

The non-zero output conductance, or what is often called the Early effect also degrades the performance of the circuits. All well known circuit techniques that reduce the output conductance, such as cascoding and regulated cascoding are beneficial to translinear circuits. A second effect contributes to the drain conductance of FGMOS transistors. As pointed out in [19], the parasitic coupling of the drain and source to the floating gate through capacitances C_{fgd} and C_{fgs} (see Figure 5) yields an exponential dependance of the output current on the drain voltage. This is evident in the experimental data shown in Figure 28 for NMOS devices; PMOS transistors exhibit similar behaviour. The experimental data in Figure 28 are fit to an output conductance model and the overlap capacitances C_{fgd} and C_{fgs} estimated as $0.15 fF/\mu m$ [19]; this number could be as large as $0.5 fF/\mu m$.

5.3. Device Matching Limitations

Another important area of concern in designing translinear computational circuits, is the poor matching characteristics of MOS transistors in subthreshold (much worse than bipolars). This is more acute in analog VLSI systems applications, where small geometry transistors must be used, typically $4\mu m \times 4\mu m$ or $6\mu m \times 6\mu m$, (in a one micron process) to achieve high densities. Low power design condiserations suggest that it is preferable to operate the devices in the region where the transconductance per unit current is maximum [16], i.e., in the subthreshold and transition regions. Small device geometries and high transconductance per unit current makes the drain current strongly dependent on spatial variations of process-dependent parameters, particularly I_0. Characterization of the fabrication process and the matching properties of the basic devices is thus of paramount importance because it provides the necessary information for designing low power systems. The experimental data in Figure 29 show that there are three different effects that can be responsible for the poor matching characteristics of MOS transistors in subthreshold; these were discussed in [42]. After discounting the two deterministic effects, we are left with the random variations.

Random mismatch in the subthreshold region can be characterized in terms of the simple model parameters I_o and κ. The parameter κ is very stable, with a normalized standard deviation of $\sigma(\kappa)/\langle\kappa\rangle \approx 0.3\%$, where (.) denotes mean value. The small variations in κ suggest that doping and gate-oxide thickness are extremely uniform. The fuzziness in the data points in Figure 29(Right) is therefore not due to changes in slope. This implies that characteristics are displaced from one transistor to the next, implicating the flat-band voltage which depends on the contact potential, and this fixed interface charge density $Q_0^\diamond$ (implanted ions and trapped electrons). The variations in the current can of course be related to this fixed-charge distribution through the transconductance and the gate-oxide capacitance.

$$\frac{\sigma(I)}{\langle I\rangle} = \kappa \frac{\sigma(Q_0^\diamond)/C_{ox}^\diamond}{kT/q} \tag{39}$$

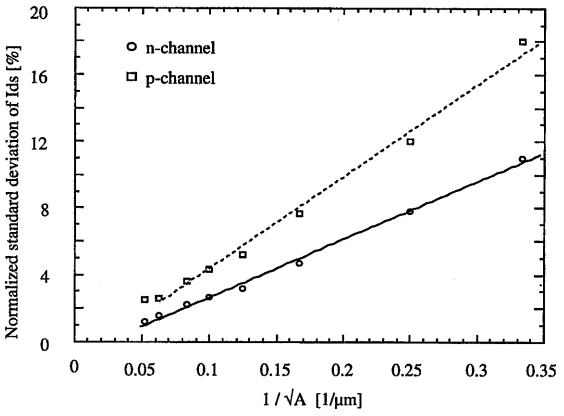

Fig. 30. Dependence of normalized standard deviation of I_{DS} on transistor size; the lines are best fits to the data. All devices have square geometries with area $\mathbf{A} = L^2$. Notice that the normalized standard deviation of I_{DS} saturates at large transistor geometries.

Figure 30 shows the dependence of the normalized standard deviation of the drain current, $\sigma(I)/(I)$, on transistor size. Each data point represents measurements from approximately 1000 transistors. The normalized standard deviation of the current is inversely proportional to the square root of the device area, **A,** and is given by:

$$\frac{\sigma(I_D)}{\langle I_D \rangle} = \sigma_0 \frac{1}{\sqrt{W \times L}} = \sigma_0 \frac{1}{A} \qquad (40)$$

where σ_0 is the mismatch per unit length for a given device type and process.

5.4. Noise Limitations

Shot noise in the MOS transistor operating in subthreshold has a one-sided power spectrum given in [49]:

$$S_{i,shot}(\omega) = 4qI_{DS} \qquad (41)$$

For a device in saturation the noise is exactly half.

For sub-threshold currents between 1 nA and 100nA, flicker noise for mid-to-low frequencies must be included. **A** model for flicker noise is given by [49].

$$S_{i,flick}(\omega) = \frac{M g_m^2}{C_{ox}^{\diamond} WL} \frac{2\pi}{LC} \qquad (42)$$

M a process-dependent constant with a typical value of 4.0 x $10^{26} C^2/m^2$.

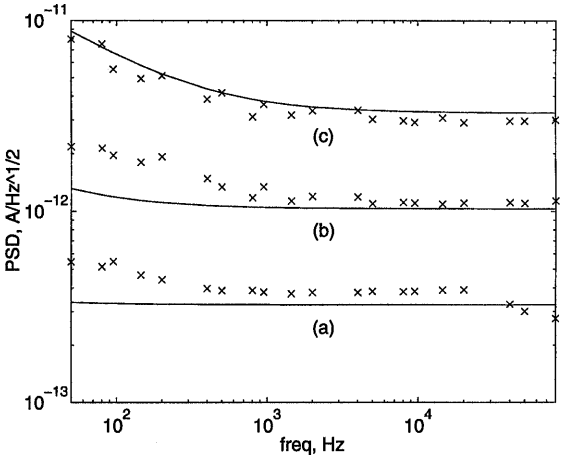

Fig. 31. Power spectral density for a PMOS transistor in saturation, where $W = 1148\,\mu$m and $L = 4\,\mu$m. The model is given **by** solid lines, the data are marked **by** x's. The three curves correspond to nominal current values of (a) 1nA, **(b)** 10nA, and (c) 100nA for an equivalent square device. Some amount of excess noise is evident at low current levels.

Assuming shot and flicker noise are independent, a complete noise model for a transistor operating in the subthreshold region is

$$S_{i,noise}(\omega) = S_{i,shot}(\omega) + S_{i,flick}(\omega) \qquad (43)$$

Figure 31 shows the noise power spectral density for a PMOS transistor. One free process-dependent parameter M is used to model the flicker noise. Note that, at low enough current levels, flicker noise cannot be detected within the audio frequency range. This property is seen for curve (a) of Figure 31 in which there is little evidence of flicker noise for frequencies above 50 Hz.

5.5. Bandwidth Limitations

The maximum useful frequency of operation possible with an MOS transistor, is determined by its transition frequency f_T defined [49] as $(g_m/2\pi C)$ where C is the total input capacitance i.e. the capacitance per unit area times the area of the device:

$$C = W\,L\,\frac{C_{ox}^{\diamond}\,C_{dep}^{\diamond}}{C_{ox}^{\diamond} + C_{dep}^{\diamond}} \approx W\,L\,C_{ox}^{\diamond}$$

so that for subthreshold operation:

$$f_T \equiv \frac{\kappa I_{DS}}{(kT/q)} \frac{1}{2\pi C}$$

The maximum value of drain current I_{DSmax} with the MOS transistor still in subthreshold region is given by [5]:

$$I_{DSmax} \equiv \frac{2}{\kappa} \frac{W}{L} \mu \, C_{ox}^{\diamond} \, (kT/q)^2$$

From the above equations a maximum transition frequency in subthreshold f_{Tmax} can be approximated to:

$$f_{Tmax} < \frac{\mu \, (kT/q)}{\pi L^2} \tag{44}$$

where μ is the effective carrier mobility and L is the device channel length. The transition frequency of a device is essentially the bandwidth (as determined by the internal gain and parasitic capacitances of the transistor). For six to ten micron length devices (typical in analog VLSI today), functional systems in the hundreds of kHz range are possible while for submicron devices, the limit extends to the MHz range.

6. Conclusions

In this paper we have provided a comprehensive overview of the application of the translinear principle to MOS circuits operating in subthreshold. Our research was aimed at exploring different ideas on neuromorphic analog network computations and their VLSI implementations. The results of our investigation are encouraging; analog circuits designed with components of limited precision, when assembled in large networks following a design methodology, based on translinear circuit techniques, can successfully perform linear and non-linear computation with energetic efficiency unmatched by any other digital counterparts. Our 590,000 transistor analog VLSI, contrast sensitive, silicon retina is another step towards the direction envisioned by Barry Gilbert: [44] ". . . convergence of IC technology capabilities and neural network requirements makes wafer–scale integration of meganetworks a very real possibility . . .".

Acknowledgements

The research was partially supported by NSF grant ECS-9313934; Paul Werbos is the program monitor, and by a contract from the Army Night Vision Laboratory at Fort Belvoir. The final version of this document was prepared while one of the authors (AGA) was on a sabbatical leave at Caltech. We thank Carver Mead for his continuing support and encouragement. Chip fabrication was provided by MOSIS.

Appendix A: Device Models

A.1 Bipolar Transistor Model

The Ebers-Moll model [46], [47] for an npn bipolar transistor is:

$$I_E = -I_F + \alpha_R I_R \tag{45}$$
$$I_C = \alpha_F I_F - I_R$$

$$I_F = I_{ES}(e^{\frac{qV_{BE}}{kT}} - 1) \tag{46}$$
$$I_R = I_{CS}(e^{\frac{qV_{BC}}{kT}} - 1)$$

$$\alpha_F I_{ES} = \alpha_R I_{CS} \tag{47}$$

where I_C and I_E are the collector and emitter currents respectively and

V_{BE} is the base to emitter voltage,
V_{BC} is the base to collector voltage,
I_{ES} is the saturation current of emitter junction with zero collector current,
I_{CS} is the saturation current of collector junction with zero emitter current,
α_F common-base current gain.
α_R common-base current gain in inverted mode, i.e. with the collector functioning as an emitter and the emitter functioning as a collector.

By convention, the currents for bipolars are positive when flowing into its terminals.

Combining Eqs. (45), (46), and (47), the collector current can be expressed as:

$$I_C = \alpha_F I_{ES} \left((e^{\frac{qV_{BE}}{kT}} - 1) - \frac{1}{\alpha_R}(e^{\frac{qV_{BC}}{kT}} - 1)) \right) \tag{48}$$

For an ideal device with common-base current gain, α_F, and common-base current gain in inverted mode, α_R, very close to unity, the above equation becomes:

$$I_C = I_{ES}(e^{\frac{qV_{BE}}{kT}} - e^{\frac{qV_{BC}}{kT}}) \tag{49}$$

However, regular bipolar transistors do not have both α_F and α_R near unity.

When the collector to base voltage equals zero or the collector is reverse biased with respect to the base, the above equation simplifies to the familiar:

$$I_C = A_E J_{ES}(e^{\frac{qV_{BE}}{kT}} - 1)$$
$$\approx A_E J_{ES} e^{\frac{qV_{BE}}{kT}} = I_S e^{\frac{V_{BE}}{V_t}} \quad (50)$$

where A_E is a design parameter, the area of the emitter junction. J_{ES} and I_S are the saturation current density and current for the emitter respectively. In this case, $I_R \ll I_F$ and the equations above give $I_C = -\alpha_F I_E$. Using the relation $I_E + I_C + I_B = 0$ (KCL) we get the familiar result

$$I_C = \frac{\alpha_F}{1-\alpha_F} I_B \equiv \beta_F I_B \quad (51)$$

where β_F is the common-emitter current gain.

A.2 MOS Transistor Model

A charge-based formulation [7], [16] that preserves the symmetry between the source/drain terminals of an MOS transistor is presented. Which terminal of the device actually serves as the source or the drain is determined by the circuit, the bias conditions—and even the input signals. This symmetric view of an MOS transistor enabled us to extend the translinear principle to operation in the subthreshold ohmic regime [16].

The MOS device has a very simple current-charge relationship because diffusion and drift are both proportional to the concentration gradient. As shown in Appendix A of [7] and in [16], this yields a quadratic expression for the current that consists of two independent opposing components I_{Q_s} and I_{Q_d}—in the absence of velocity saturation and channel-length modulation effects. These components are related to charge densities at the source $Q_s^\diamond$ and $Q_d^\diamond$ at the drain of the device.

The device drain–source current can thus be written as:

$$I \equiv I_{Q_s} - I_{Q_d}$$
$$= \mu \frac{W}{L}\left[\left(\frac{1}{2}\frac{Q_s^{\diamond 2}}{C_{ox}^\diamond + C_{dep}^\diamond} + \frac{kT}{q}Q_s^\diamond\right) - \left(\frac{1}{2}\frac{Q_d^{\diamond 2}}{C_{ox}^\diamond + C_{dep}^\diamond} + \frac{kT}{q}Q_d^\diamond\right)\right] \quad (52)$$

W is the width, L is the length of the channel and (μ) is the effective channel mobility. The capacitances $C_{ox}^\diamond$ and $C_{dep}^\diamond$ are the gate oxide and depletion area capacitances of the channel. A key property of the MOS device that makes this possible is lossless channel conduction. Unlike a bipolar transistor, the controlling charge on the gate is isolated from the charge in transport by the almost infinite gate-oxide resistance. Therefore, there is no recombination between the current-carrying charge in the channel and the current-modulating charge on the gate.

The familiar ohmic/saturation dichotomy introduced in voltage-mode design can be reformulated in terms of the opposing drain and source driven current components. In saturation, $|I_{Q_d}| \ll |I_{Q_s}|$ and $I \approx I_{Q_s}$ and therefore the current is independent of the drain voltage. In ohmic, $I_{Q_d} \sim I_{Q_s}$ and $I = I_{Q_s} - I_{Q_d}$ and therefore the current depends on the drain voltage as well as the source and gate voltages. The functional dependence of the current components on the terminal voltage is fixed and remains the same throughout the ohmic and saturation regions.

The charge densities at the source and drain terminals can be related to the terminal voltages. In general the charge-voltage relationship is much more complicated than the current-charge one because both the mobile charge and the depletion charge are involved in the electrostatics. The device current in Equation 52 can thus be written as a function F, of the terminal voltages with a general functional form for the current-voltage relationship valid for all the regions of operation given by:

$$I_{SD} \propto \mathcal{F}(V_{GB}, V_{SB}) - \mathcal{F}(V_{GB}, V_{DB}) \quad (53)$$

This functional form was first introduced by [48] for above threshold operation and is also discussed in [49]. For an n-type device, $\mathcal{F}$ is a nonpositive, monotonically decreasing function of V_{GB} and a monotonically increasing function of V_{SB}.

In subthreshold region, the following factorization of F : is also possible [6], [7].

$$I \propto \mathcal{G}(V_{GB})\,[\mathcal{H}(V_{SB}) - \mathcal{H}(V_{DB})] \quad (54)$$

where $\mathcal{G}$ and 31 are exponential functions. This shows that the source-driven and drain-driven components are controlled independently by V_{SB} and V_{DB}. However, V_{GB}, acting through the surface potential, also controls both components in a symmetric and multiplicative fashion. In this mode of operation the MOS transis-

tor has been called a *diffusor* [15] in analogy with the variable conductance electrical junctions in biological systems.

An expression for the current in an NMOS transistor operating in subthreshold can thus be written [6], [7] as:

$$\begin{aligned} I_D &\equiv I_{DS} \\ &= I_{n0} S \exp(\kappa_n V_{GB}/V_t) \\ &\quad \times [\exp(-V_{SB}/V_t) - \exp(-V_{DB}/V_t)] \end{aligned} \tag{55}$$

and for a PMOS

$$\begin{aligned} I_D &\equiv I_{SD} \\ &= I_{p0} S \exp(-\kappa_p V_{GB}/V_t) \\ &\quad \times [\exp(V_{SB}/V_t) - \exp(V_{DB}/V_t)] \end{aligned} \tag{56}$$

The terminal voltages V_{GB}, V_{SB}, V_{DB} are referenced to the substrate. The constant I_0 depends on mobility (μ) and other silicon physical properties. S is a geometry factor, the width W to length L ratio the device.

For devices that are biased with $V_{DS} \geq 4\ V_t$, (saturation) the drain current is reduced to:

$$\text{IDS} = S I_{no} \exp(1-\kappa_n) V_{BS}/V_t) \exp(\kappa_n V_{GS}/V_t) \tag{57}$$

This shows explicitly the dependence on V_{BS} and the role of the bulk as a ***back-gate*** that underlies this. This equation, having only the dependence on V_{GS}, and V_{BS}, is used for circuit designs where devices operate in saturation as transconductance amplifier. However, channel-length modulation (Early effect) —which we have ignored completely—becomes significant in saturation. So the device equations must be augmented with terms that model this effect to accurately predict the output conductance.

The parameter κ is defined as

$$\kappa = \frac{C_{ox}^{\diamond}}{C_{ox}^{\diamond} + C_{dep}^{\diamond}} \tag{58}$$

The physical significance of κ is apparent if the observation is made that that the oxide and depletion capacitances form a capacitive divider between the gate and bulk terminals that determines the surface potential [7]. Lighter doping reduces $C_{dep}^{\diamond}$, and pushes the divider ratio closer to unity. A larger surface potential also reduces $C_{dep}^{\diamond}$. The parameter κ takes values between 0.6 and 0.9.

Notes

1. The diffusor is a term adopted in [15] to describe the exploitation of diffusion transport in MOS transistors to spread signals in a manner analogous to gap junctions between neural cells.

References

1. B. Gilbert, "Translinear circuits: A proposed classification." ***Electronics Letters,*** 11(1), pp. 14–16, 1975; errata, 11(6), p. 136.
2. E. Seevinck, "Analysis and synthesis of translinear integrated circuits." ***Studies in Electrical and Electronic Engineering,*** vol. 31. Elsevier: Amsterdam, 1988.
3. B. Gilbert, "Current-mode circuits from a translinear viewpoint: a tutorial." In Chapter 2 of C. Toumazou, F. J. Lidgey and D. G. Haigh, eds., ***Analogue IC Design: The Current-Mode Approach,*** IEEE Circuits and Systems Series, vol. 2. Peter Peregrinus Ltd: London, 1990.
4. E. Seevinck and R. J. Wiegerink, "Generalized translinear circuit principle." ***IEEE Journ. of Solid-state Circuits,*** SC-26(8), pp. 1198–1102, August 1991.
5. E. A. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operation." ***IEEE Journ. of Solid-state Circuits,*** SC-12(3), pp. 224–231, June 1977.
6. E. A. Vittoz, "Micropower techniques." In Y. P. Tsividis and P. Antognetti, eds., ***VLSI Circuits for Telecommunications.*** Prentice Hall: 1985.
7. C. A. Mead, ***Analog VLSI and Neural Systems.*** Addison-Wesley: Reading, MA, 1989.
8. C. A. Mead, "Neuromorphic electronic systems." ***Proceedings IEEE,*** 78(10), pp. 1629–1636, Oct. 1990.
9. A. G. Andreou, "Low power analog VLSI systems for sensory information processing," in B. Sheu, Edgar Sanchez-Sinencio and M. Ismail, eds., ***Microsystems Technologies for Multimedia Applications: An Introduction.*** IEEE Press: Los Alamitos, CA, 1995.
10. K. A. Boahen, "Retinomorphic vision systems." ***Proceedings of MicroNeuro-96.*** IEEE Computer Society Press: Los Alamitos, CA, 1996.
11. A. G. Andreou and K. A. Boahen, "Synthetic neural circuits using current domain-signal representations." ***Neural Computation,*** 1, pp. 489–501, 1989.
12. A. G. Andreou, "Synthetic neural systems using current-mode circuits." ***Proceedings of the IEEE 1990 International Symposium on Circuits and Systems,*** New Orleans, pp. 2428–2432, May 1990.
13. A. G. Andreou, K. A. Boahen, P. O. Pouliquen, A. Pavasović, R. E. Jenkins, and K. Strohbehn, "Current-mode subthreshold MOS circuits for analog VLSI neural systems." ***IEEE Trans. on Neural Networks,*** 2(2), pp. 205–213, March 1991.
14. M. Cohen and A. G. Andreou, " Current-mode subthreshold MOS implementation of the Herault–Jutten autoadaptive network." ***IEEE Journ. of Solid-State Circuits,*** 27(5), pp. 714–727, May 1992.

15. K. A. Boahen and A. G. Andreou, "A contrast sensitive silicon retina with reciprocal synapses." In J. **E.** Moody, **S.** J. Hanson, and R. **P.** Lippmann, eds., ***Advances in Neural Information Processing Systems 4***. Morgan Kaufmann Publishers: San Mateo, CA, 1992.
16. A. G. Andreou and K. A. Boahen, "Neural information processing (II)." In Chapter 8 of ***Analog VLSI Signal and Information Processing,*** M. Ismail and T. Fiez, eds. McGraw-Hill: 1994.
17. A. G. Andreou and K. A. Boahen, "A 48,000 pixel, 590,000 transistor silicon retina in current-mode subthreshold CMOS." ***Proceedings of the 37th Midwest Symposium on Circuits and Systems,*** pp. 97–102, Lafayette, Louisiana, August 1994.
18. A. G. Andreou and K. A. Boahen, "A 590,000 transistor, 48,000 pixel contrast sensitive, edge enhancing CMOS imager-silicon retina." ***Proceedings of the 16th Conference on Advanced Research in VLSI,*** pp. 225–240, Chapel Hill, North Carolina, March 1995.
19. K. Yang and A. G. Andreou, "A multiple input differential amplifier based on charge sharing on a floating gate MOSFET." ***Journal of Analog Integrated Circuits and Signal Processing,*** 6, pp. 197–208, November 1994.
20. B. **A.** Minch, C. Diorio, P. Hasler, and C. A. Mead, "Translinear circuits using subthreshold floating-gate MOS transistors." ***Journal of Analog Integrated Circuits and Signal Processing,*** this issue.
21. A. Fabre, "Dual translinear voltage/current convertor." ***Electronics Letters,*** 19(24), pp. 1030–1031, 1983.
22. A. Fabre, "Wideband translinear current convertor." ***Electronics Letters,*** 20, pp. 241–242, 1984.
23. P. M. Furth and A. G. Andreou, "A high-drive low-power BiCMOS buffer using compound PMOS/NPN transistors." ***Proceedings of the 36th Midwest Symposium on Circuits and Systems,*** pp. 1369–1372, Detroit, Michigan, August 1993.
24. A. G. Andreou, K. Strohbehn, and R. Jenkins, "A silicon retina for motion computation." ***Proceedings of the IEEE 1991 International Symposium on Circuits and Systems,*** Singapore, pp. 1373–1376, June 1991.
25. Z. Kalayjian, A. G Andreou, L. Wolff, and J. Williams, "Integrated 1-D polarization imagers." ***Proceedings of the 29th Annual Conference on Information Sciences and Systems,*** Baltimore, MD, March 1995.
26. **X.** Arreguit, **E.** A. Vittoz, F. A. van Schaik, and A. Mortara, "Analog implementation of low-level vision systems." ***Proceedings of European Conference on Circuit Theory and Design,*** pp. 275–280. Elsevier Publishing: September 1993.
27. T. Delbrück, "'Bump' circuits for computing similarity and dissimilarity of analog voltages." ***Proc. Int. Joint Con. on Neural Networks,*** Seattle WA, pp. 475–479, 1991.
28. M. Cohen and A. G. Andreou, "MOS circuit for nonlinear Hebbian learning." ***Electronics Letters,*** 28(9), pp. 809–810, April 1992.
29. M. van der Gevel and J. C. Kuenen, "Square-root **X** circuit based on a novel, back-gate-using multiplier." ***Electronics Letters***, 30(3), pp. 183–184, February 1994.
30. M. **H.** Cohen and A. G. Andreou, "Analog CMOS integration and experimentation with an autoadaptive independent component analyzer." ***ZEEE Transactions on Circuits and Systems: Part II: Analog and Digital Signal Processing,*** 42(2), pp. 65–77, February 1995.
31. T. **S.** Lande, J. A. Nesheim, and Y. Berg, "Auto correlation in micropower analog CMOS." ***Journal of Analog Integrated Circuits and Signal Processing,*** 7, pp. 61–68, January 1995.
32. E. A. Vittoz, "Analog VLSI implementation of neural networks." In ***Handbook of Neural Computation,*** Institute of Physics Publishing and Oxford University Press, USA.
33. B. L. Hart, "Translinear circuit principle: a reformulation." ***Electronics Letters,*** 15(24), pp. 801–803, Nov. 22, 1979.
34. Mead, C. A. and Mahowald, M. A., "A silicon model of early visual processing." ***Neural Networks,*** 1, pp. 91–97, 1988.
35. **E.** Vittoz and X. Arreguit, "Linear networks based on transistors." ***Electronics Letters,*** 29, pp. 297–299, February 4, 1993.
36. K. Bult and G. J. G. M. Geelen, "An inherently linear and compact MOST-only current division technique." ***IEEE J. Solid-State Circuit,*** SC-27(12), pp. 1730–1735, December 1992.
37. M. Tartagni and P. Perona, "Computing centroids in current-mode technique," ***Electronics Letters,*** 21, October 1993.
38. J. M. Buhmann, M. Lades, and R. Eeckman, "Illumination-invariant face recognition with a contrast sensitive silicon retina." In J. D. Cowan, G. Tesauro, and J. Alspector, eds., ***Advances in Neural Information Processing Systems 6***. Morgan Kaufmann Publishers: San Mateo, CA, 1994.
39. J. E. Dowling, ***The Retina: An Approachable Part of the Brain.*** The Belknap Press of Harvard University: Cambridge, MA, 1987.
40. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory." ***Nature,*** 317, pp. 314–319, 1985.
41. C. A. Mead and T. Delbrück, "Scanners for visualizing activity in analog VLSI circuitry." ***Analog Integrated Circuits and Signal Processing,*** 1(2), Oct. 1991.
42. A. Pavasović, A. G. Andreou, and C. R. Westgate, "Characterization of subthreshold MOS mismatch in transistors for VLSI systems." ***Journal of Analog Integrated Circuits and Signal Processing,*** 6, pp. 75–84, June 1994.
43. **B.** Gilbert, "A monolithic 16-channel analog array normalizer." ***IEEE Journ. of Solid-state Circuits,*** SC-19(6), 1984.
44. B. Gilbert, "Nanopower nonlinear circuits based on the translinear principle." In ***Workshop on: Hardware implementation of neural nets and synapses,*** San Diego, CA, January 1988.
45. F. J. Pineda and A. G. Andreou, ***An Analog Neural Network Inspired by Fractal Block Coding,*** Advances in Neural Information Processing Systems, Vol. 7. Morgan Kaufmann Publishers: San Mateo, CA, 1995.
46. J. J. Ebers and J. L. Moll, "Large-signal behavior of junction transistors." ***Proceedings IRE,*** 42, pp. 1761–1772, December 1954.
47. R. S. Muller and T. I. Kamins, ***Device Electronics for Integrated Circuits.*** John Wiley and Sons: New York, 1977.
48. J. E. Meyer, "MOS models and circuit simulation." ***RCA Review,*** 32, pp. 42–63, March 1971.
49. Y. Tsividis, ***Operation and Modeling of the MOS Transistor.*** McGraw-Hill: New York, 1987.

Andreas G. Andreou received the M.S.E. and Ph.D. in electrical engineering and computer science from Johns Hopkins University, Baltimore, in 1983 and 1986, respectively. From 1987 to 1989 he was a postdoctoral fellow and associate research scientist at Johns Hopkins where he became Assistant Professor in 1989 and Associate Professor in 1993. His research interests are in the areas of device physics, integrated circuits and neural computation.

Dr. Andreou is a member of Tau Beta Pi and a member of I.E.E.E.

Kwabena A. Boahen is a Ph.D. student at Caltech in the Computation and Neural Systems program after completing a B.S./M.S.E. degree in electrical and computer engineering at Johns Hopkins University, Baltimore, MD. His research at Caltech involves analog VLSI models of biological computation, with an emphasis on retinal computation and chip-to-chip communication.

Mr. Boahen is a member of Tau Beta Pi and a student member of the I.E.E.E.