

R.W. Newcomb, J.D. Lohn, "Analog VLSI for Neural Networks," in Handbook of Brain Theory and Neural Networks, M. Arbib Ed., Bradford Books, MIT Press, 1995.

Analog VLSI for Neural Networks

Robert W. Newcomb and Jason D. Lohn

Handbook of Brain Theory and Neural Networks
M. Arbib Ed., Bradford Books, MIT Press, 1995.

1. Introduction

One of the most promising strategies for implementing neural networks is through the use of electronic analog VLSI (very large scale integration) circuits. An analog circuit is one that processes a continuum of real-valued signals of continuous time, in contrast to a digital circuit which processes integer-valued (most often binary) signals in discretized time. VLSI refers to an integrated circuit design and manufacturing technology whereby hundreds of thousands to millions of active components (most often transistors) are placed on a chip on the order of 100 mm^2 in area and 0.5 mm thick. Because Artificial Neural Networks (ANNs) attempt to behave similarly to the brain with its millions of neurons, VLSI is the most appropriate presently available technology for their hardware implementations. Furthermore, both VLSI circuits and biological neurons are of the same class, that is, fundamentally analog.

Although during the 1980s digital ANNs held most interest, the first neural-like circuits were analog ones constructed by Dr. Otto Schmitt in the late 1930s using vacuum tube analog computer circuits (Schmitt, 1937). These were extended to rather cumbersome transistor circuits after the Second World War to obtain artificial neurons where much of the emphasis was placed upon the initiation and propagation of action potentials. The circuits developed to accomplish these tasks rely upon nonlinearities for implementing amplitude saturation, pulse repetition saturation, threshold effects, and dynamics for effecting time domain changes upon the action potentials (Reiss, 1962). But, because of the large size of the circuits used for just one neuron, very little was done to make full ANN systems until the advent of integrated circuits (ICs). In a number of research centers around the world during the mid 1960's considerable interest began to develop in the design of analog IC neurons and systems built from them. Of importance to the signal processing capabilities in this development has been the recent emphasis upon the synaptic combining of signals via weight matrix summations as opposed to the axon propagation of action potentials. Present analog ANNs consist of synaptic weights, implemented by amplifier gains, summation of the weighted signals, implemented by the use of Kirchhoff's laws (most conveniently the current law, denoted KCL), activation functions realized by amplifier nonlinearities, and in many cases dynamics, via capacitors, for smoothly transitioning from an initial state to a desired equilibrium.

The work horses of analog VLSI ANNs are the Differential Voltage Controlled Current Source (DVCCS) and capacitors. The DVCCS is used for making synaptic weights and activation functions and capacitors are used for dynamics. A DVCCS takes a voltage difference as input, and gives an output current as a function

of that difference. DVCCS gains can realize the weights when operating on small signals in a linear fashion and can also realize saturation nonlinearities when operating on large signals. In both cases the DVCCS output currents can be conveniently summed by KCL. Using capacitors in conjunction with DVCCSs it is known that any linear circuit can be realized (Bialko, 1971) so that any desired filtering of ANN signals is available, as may be needed in special applications such as ANN retinas (Mead, 1989). Along with the DVCCS and the capacitor it is also convenient to have resistors for conversion of currents to voltage and voltage divisions, as well as devices for creating and scaling currents (called current sources and current mirrors, respectively). Except for passive resistors and capacitors (both of which are generally avoided in VLSI due to large area or non-ideal characteristics) all of these devices can basically be constructed from VLSI transistors, which are discussed in Section 2.

First we present a complete ANN analog circuit to give an overview of the circuits discussed in later sections. Figure 1 shows an analog circuit suitable for the VLSI realization of the Hopfield continuous time equations which, for the n th neuron of a set of N , are (for simplicity of notation we omit subscripts n on device parameters but not on the output currents and state)

$$C \frac{dx_n}{dt} + Gx_n = \sum_{i=1}^N g_{mi} \cdot R_i I_{ni} + I_{\text{bias}} \quad n = 1, \dots, N \quad (1a)$$

$$i_n = g(x_n) \quad (1b)$$

In Eq. (1) $g(\cdot)$ is any of the activation functions available (see Eqs. (2-4)) with i_n being its current output; x_n is the n th neuron's state variable; I_{ni} is the current output of the i th neuron which is fed to the input of the n th neuron; $g_{mi}R_i$ are the synaptic weights and I_{bias} is the bias input. With reference to Fig. 1 and Eqs. (1) it will be shown below that VLSI circuits can be constructed to make this Hopfield class of analog neural networks, as well as any other analog ANN (e.g. ART II, pulsed Hebbian, biological mimics). The nonlinear function $g(\cdot)$ in Eq. (1b) can be realized via a DVCCS (see Fig. 5) exhibiting square law, exponential, or sigmoidal processing. In this simple model of a neuron, these nonlinearities can be thought to correspond to the activation processes in the cell body. The weighted inputs from the synapses to the cell body can be thought to correspond to varying amounts of currents linearly summing, via KCL, at the input node to the left of the activation function DVCCS in Fig. 1. On the right side of Eq. (1a) the weights $g_{mi}R_i$ are the current gains of DVCCSs operating as linear amplifiers with resistor inputs, the resistors being used to convert the neuron output currents to voltages (and constructed using one of Figs. 2(b), 4(a), 6(a) depending upon their Ohmic value). The bias input I_{bias} , also on the right of Eq. (1a), is constructed as a

constant current source made of a transistor (Fig. 4(a)). Using a resistor-capacitor branch connected to this same input node of the activation function amplifier, we obtain the dynamics of the analog circuit, shown on the left side of Eq. (1a) incorporating the derivative. Because each neuron output current i_n (which can be positive or negative) needs to be sent to each of the other neurons, it needs to be repeated N times, this being accomplished by the bidirectional current mirror (see Fig. 4(d)) in multi-output form on the right of Fig. 1. This simply reproduces N copies of the neuron output current i_n irrespective of what load is presented to it. For adjustments, as may be needed for adaptive ANNs, the g_{mi} can be made voltage variable by variation of the gain of the associated DVCCS (via the tail current introduced in Fig. 5 below). Using VLSI layouts, as shown in Fig. 2, for constructing the components of Fig. 1 with the following transistor circuits, analog VLSI neural systems can be fabricated which realize this and other analog ANNs.

2. Transistors and VLSI Layouts

The key circuit component in analog VLSI is the transistor, a three-or four-terminal device that can behave as a switch in digital circuits and an amplifier in analog ones. Transistors may be fabricated using a variety of technologies: BJT (bipolar junction transistor), MOS (metal oxide semiconductor), CMOS (complementary MOS), and others. For neural network implementations, MOS and BJT devices have been used the most; when both occur together the process is called BiCMOS and is the most prevalent present day analog VLSI technology.

Figure 2 shows the circuit symbols for those transistors of most interest to ANN VLSI along with a top view of an IC layout of each. The fabrication details can be found in (Geiger, 1990), however for our purposes, it is enough to know only a few aspects about their operation.

In the MOS transistor the drain current, I_D , which flows from the outside into the drain, D , and then through the device to the source, S , is controlled by the voltage at the gate, G , with respect to the source, V_{GS} , when the latter is “above” threshold voltage, V_{th} . As the threshold voltage can be used as a fine control on the ANN weights we note that it is dependent upon the voltage of the bulk, B , to source voltage, V_{BS} , where the bulk material is that of the substrate into which the transistor is embedded. The two types of MOS transistors, NMOS and PMOS, are distinguished by their internal conduction mechanisms with the currents and voltages of the latter being ideally the negative of the former in the complementary case desired for CMOS fabrications. Since the channel can be formed by enhancing or depleting charge,

we have enhancement and depletion mode transistors of each of the NMOS and PMOS types, where the distinction is that the threshold voltages of depletion mode devices are generally of opposite sign to those of enhancement mode devices. Depletion mode transistors are not as common in analog VLSI due to the extra fabrication steps needed but can be used to obtain more flexible designs. MOS transistors can be operated such that between the drain and source a resistor is seen whose value depends upon the gate to source voltage, giving a voltage variable resistor useful for adaptation. More commonly the MOS transistor is operated in its saturation mode where instead of a resistor between drain and source a current source is seen, this current source depending on the gate to source voltage in a square law fashion conveniently allowing for quadratic weights. By operating an MOS transistor at very low gate to source voltages, called sub-threshold, exponential behavior is obtained; sub-threshold operation is convenient for low power designs but is not too robust. In all of these cases the drain current is proportional to the width to length ratio, $\frac{W}{L}$, of the channel which acts as a design parameter that is very easily set in a VLSI layout. For bipolar transistors exponential nonlinearities are obtained and the VLSI design parameter is the emitter area to which the collector current is proportional.

Besides the active transistors, passive capacitors are used to obtain dynamics needed for realization of the ANNs that are described by differential equations, such as ART-II, the Hopfield continuous time neural nets and derivatives needed for back-propagation. Several types of capacitors are available in VLSI. The primary one is realized by an oxide between two conductive layers (presently polysilicon but possibly metal on top and doped silicon on the bottom), as shown in Fig. 3(a). These capacitors are linear time-invariant capacitors and satisfy the standard law of $i = Cdv/dt$ with $C = C_{OX}WL$ for which C_{OX} is the capacitance per unit area of the (gate) oxide used. The area, WL , of the polysilicon plate serves as a design constant. Unfortunately, to obtain capacitance of useful values requires considerable area, and, hence, capacitors often take up a good portion of analog VLSI neural networks.

At times one also needs linear resistors, for transformation of currents to voltages or for biasing, in which case the most common means of VLSI implementation is via strips of polysilicon, often in snake form to optimize layout (see Fig. 3(b)). The conductance G is given through the sheet resistance R_s (in Ohms/square, a material constant) by $G = W/(LR_s)$, in which L is the length (distance between contact pads) and W is the width of the polysilicon. These resistors also take up considerable area and, thus, are avoided, but for small values of resistance are sometimes invaluable (for values of 10 to 100 Ohms). Means

of transistor construction of larger valued resistors are given in the next section.

3. Primary Circuits

Two of the key components in an ANN are the weights and the nonlinear activation functions. A weight can be realized by a DVCCS operating in its linear region while a sigmoidal nonlinear activation function can be realized by operating a DVCCS over its full nonlinear range. We consider, as background, current sources, current mirrors, and resistors constructed as diode connected transistors. These are all used for biasing the transistors, that is, setting the modes of operation of transistors, while the current mirrors and sources are used for various adjustments, as in adapting weights.

Figure 4(a) shows how a current source can be constructed from a voltage source of voltage V and an MOS transistor operating in its saturation region. We note that 1) the current I can be adjusted by varying the above-threshold voltage V , 2) current sources of one polarity are changed into ones of opposite polarity by reversing the attachment points or by interchanging NMOS and PMOS, and 3) one needs to maintain the saturation mode of operation which is achieved by application of sufficient bias voltage across the current source nodes. If the transistors of Fig. 4(a) are operated in their Ohmic regions, with small V_{DS} , then the same circuits give voltage variable resistors of conductance $G(V)$, which is useful for making small area resistors (10-1000 Ohms) as well as adaptive adjustments. On the left of Fig. 4(b) is shown a diode connected transistor which acts as a nonlinear resistor. If the transistor is turned on and is an enhancement mode device, then it is always in the saturation region and we have quadratic behavior, that is, $I = (K \frac{W}{L})(V - V_{th})^2 1(V - V_{th})$, where $1(\cdot)$ is the unit step function and K is a material constant. If a depletion mode transistor is used in Fig. 4(b) then the mode of operation is Ohmic (rather than saturation) and when turned on the law changes to a linear one, becoming $I = (K \frac{W}{L})(V - V_{th})V$ for $V_{th} < V$ with $V_{th} < 0$. Since this single transistor essentially passes current in only one direction, whereas currents which can assume any polarity are often needed, placement of a PMOS transistor back to back with an NMOS one leads to the bidirectional resistor of the right side of Fig. 4(b).

Figure 4(c) shows current mirrors which allow current in one section of an ANN to determine that in another, perhaps for adjusting weights. The gains are easily set by layout design, being ratios of widths to lengths when the transistors are maintained in saturation. These mirrors use the diode connection of Fig. 4(b) to set the gate-source voltages for the input and output transistors to be equal, of value $V_{GS} =$

$V_{\text{th}} + \sqrt{(L/KW)I_{\text{in}}}$. The current mirrors of Fig. 4(c) allow current to flow in only one direction. However, by placing a P-mirror on top of an N-mirror as shown in Fig. 4(d) we can get a bidirectional current mirror which is convenient for realizing weights. By replacing all of the MOS devices by BJTs in the sources, in the mirrors, or in the diodes of Fig. 4, similar BJT devices also can be constructed. Further, by placing several output transistors on one input transistor, multiple output current mirrors are easily constructed and of considerable use for distributing current in current mode VLSI ANNs (as in Fig. 1).

Figure 5(a) shows the basic configuration of a DVCCS. In Figure 5 the tail current, I_{T} , is steered between I_1 and I_2 by the differential pair consisting of identical transistors T_1 and T_2 (of NMOS or NPN types), with the steering controlled by the voltage difference of the input voltages, $V_{\text{d}} = V_1 - V_2$. The difference of the transistor currents, $I_{\text{d}} = I_1 - I_2$, is designed to be a function of V_{d} and I_{T} , independent of any devices connected, such as loads or the current mirror. To obtain the current output as this difference, the current mirror is used along with KCL at the output node so that $I_{\text{out}} = -I_{\text{d}}$. The function of I_{out} versus V_{d} realized depends upon the NMOS or NPN transistors and their modes of operation used to form the current difference. In all cases the gates/bases are the leads to the left (in T_1) and right (in T_2), the drains/collectors are at the top and the sources/emitters are at the bottom. In practice there is some loading by whatever is attached in which case three current mirrors are used for isolation as shown in Fig. 5(b) where, as an example, NMOS transistors have been inserted for T_1 and T_2 . In Fig. 5(b) the output sign is changed, from that of Fig. 5(a), and a gain k is introduced through the upper mirrors so that $I_{\text{out}} = +kI_{\text{d}}$.

For the possible nonlinearities of I_{d} versus V_{d} there are several design alternatives. Considering first NMOS transistors operated in saturation, a sigmoidal ANN activation function can be made by the use of sufficiently large bias voltages and small enough input difference voltage V_{d} , when $V_{\text{d}}^2 < \sqrt{I_{\text{T}}/(K\frac{W}{L})}$. Upon noting that $I_{\text{T}} = I_1 + I_2$, the difference current is

$$I_{\text{d}} = (K\frac{W}{L})\sqrt{(2I_{\text{T}}/(K\frac{W}{L})) - V_{\text{d}}^2}V_{\text{d}} \quad V_{\text{d}}^2 < I_{\text{T}}/(K\frac{W}{L}) \quad (2)$$

When used for $V_{\text{d}}^2 \ll \sqrt{2I_{\text{T}}/(K\frac{W}{L})}$ the difference current I_{d} in Eq. (2) is linearized to

$$I_{\text{d}} = g_{\text{m}}V_{\text{d}} \quad g_{\text{m}} = \sqrt{2I_{\text{T}}(K\frac{W}{L})} \quad (3)$$

Typical orders of magnitude are $O(I_{\text{T}}) = 10^{-3}$, $O(K) = 10^{-4}$, $10^{-2} < O(\frac{W}{L}) < 10^2$, giving $10^{-5} < O(g_{\text{m}}) < 10^{-2}$ over a limited range of input V_{d} . However, if a wider range of linear relationship is desired, then it is best to operate the NMOS transistors in the Ohmic region, which is achieved by clamping their drain-source voltages to be equal and small using npn-NMOS pairs.

If in Fig. 5(a) T_1 and T_2 are NPN transistors, then, upon again noting that $I_T = I_1 + I_2$

$$I_d = I_T \tanh(V_d/(2V_T)) \quad (4)$$

where $V_T \approx 0.025$ is the thermal voltage at room temperature. This characteristic is quite nicely sigmoidal, as well as mathematically convenient (being infinitely differentiable), leading to the BJT DVCCS having considerable importance for VLSI construction of ANN activation functions, especially for back propagation circuits. An almost identical result occurs if NMOS transistors are used in the subthreshold mode.

In some instances it is necessary to convert the output of a DVCCS into a voltage (giving a Differential Voltage Controlled Voltage Source, DVCVS), as when voltage output for an activation function is desired. Such can be accomplished by directing DVCCS output current into a resistor, perhaps made by other DVCCSs or an MOS transistor as per Fig. 4(b). However one of the best ways to do this is to attach the gates of a CMOS pair, of the type shown in Fig. 5(c), onto the output of the DVCCS. Since the CMOS pair allows no current at its input, the DVCCS can of course no longer act as a current source but its output voltage is determined by other factors (specifically, the channel length modulation effect through the Early voltage). Other voltage amplifiers are available in the literature (Geiger, 1990) but obtaining gain comparable to those of the ever useful operational amplifiers (op-amps) (with voltage gain $\approx 10^6$), takes considerable chip area and design expertise. Thus op-amps are not generally reasonable in VLSI for ANNs.

Since the DVCCS and the capacitor are sufficient to generate all linear circuits, we can construct many of the components of ANNs using them. For example, Fig. 6(a) shows construction of a resistor with resistance on the order of 100 Ohms, complementing the sizes available from snake resistors. Although the connection looks like positive feedback it is stable due to our sign convention of currents entering the device terminal (reversing the + and - terminals very conveniently yields a negative resistor). This DVCCS resistor has another advantage of being variable since g_m can be controlled by the tail current which in turn can be controlled by the gate voltage of an MOS transistor realizing the current source. In Fig. 6(b), a degree one low-pass filter, and in Fig. 6(c) a degree two low pass filter (formed from two of degree one with feedback), are shown in which any stable, or even unstable, degree one or two low pass filters can be obtained by proper choice of the parameters (Kardontchik, 1992). Any higher degree filter is easily realized using state-variable design theory. However, ANNs also require nonlinearities and, as we have seen above, several nonlinearities are available, such as square-law and sigmoidal tanh ones. To build other nonlinearities it is convenient to have multipliers, which can also be constructed from the DVCCS. Thus we note that if the tail current

I_T is made to vary with another voltage, then Eqs. (2) and (4) show that the output current varies in a multiplicative manner with that voltage. For the NMOS differential pair operated in the saturation region, this is especially convenient if the tail current is made by the current source of Fig. 4(a). The reason is that the square in the current source current, Fig. 4(a), cancels the square-root in the output difference current I_d for small V_d^2 , effectively letting multiplication take place. Although the input difference voltage, V_d , can take any sign, the tail current is limited to be positive which means this type of multiplier is a two quadrant multiplier. A much better multiplier is based upon the four-quadrant Gilbert multiplier (Geiger, 1990, p. 737) which uses two DVCCSs with the transistors of their tail current sources forming another differential pair. Assuming linear operation in the saturation region of the transistors and inputs V_x and V_y bounded by $V_x^2 \ll 2I_1(K\frac{W}{L}), \ll 2I_2(K\frac{W}{L}), V_y^2 \ll 2I_T(K\frac{W}{L})$, with K, I_1 and I_2 as in Eq. (2) the Gilbert multiplier gives $I_{out} \approx V_x V_y$. This Gilbert multiplier has successfully been used to multiply voltage determined weights with neuron output voltages (Linares, 1993, p. 446). By solving the Gilbert multiplier equation for V_x one can in principle obtain division, that is $V_x \approx I_{out}/V_y$, where now, however, I_{out} must become an input, something which is rather difficult to accomplish, though possible if one were to use high voltage gain operational amplifiers. Thus dividers are even less recommended than multipliers in VLSI.

4. Applications

Analog VLSI implementations of neural-like processing have been applied to domains such as neural modeling, visual processing, and associative memories. In this section, we briefly highlight some of these applications and invite the interested reader to learn more by way of the following references: (Mead, 1989; Zornetzer, 1990; Sánchez, 1992, 1993). Other applications include constrained optimization (Tank, 1986), Fourier transform computations (Culhane in El-Leithy, 1989), oscillators (Linares in El-Leithy, 1989), Hebbian learning (Meador, 1991), A/D conversion (Yuh in Sánchez, 1993), data compression (Fang in Sánchez, 1992), pattern recognition (Salam, 1991), and fuzzy controllers (Yamakawa in Sánchez, 1993).

4.1 Silicon Neurons

Analog VLSI circuits have been applied to the task of neural modeling whereby a given neurophysiological phenomenon is mimicked. Neural modeling by way of hardware implementations serves to test the suitability of a given physical medium, at speeds unobtainable by computer simulation. It also provides a testbed useful

for experimentation and testing (e.g. drug effects) on a given neurophysiological system.

One approach in which nerve cell characteristics were modeled in hardware is the silicon neuron of (Mahowald 1991). In the circuits comprising the silicon neuron, the neuron's ability to self-generate electrochemical impulses are emulated. For example, in one circuit, a differential pair and transistor are used to model the activation of voltage dependent membrane conductances and the flow of potassium ions, respectively. The same circuit also employs a follower-integrator to allow for time dependence adjustments of the ion currents. A second circuit, similar to the first, was designed to model the flow of sodium ions. In addition to an activation signal, an inactivation signal was needed, and was implemented with a differential pair. In both cases, the sigmoidal nature of the subcircuits closely matches the observed membrane conductances. When the above circuits are integrated into a single VLSI chip, the authors estimate the area of a single neuron to be less than 0.1 mm^2 . Also, power dissipation is small – 60 mW per neuron. Other approaches have been taken by (Moon in Sánchez, 1992).

Circuit realizations for a set of low-level electrochemical processes occurring within synapses have also been constructed. Using dynamics derived from actual neurophysiological data, second messenger chemical “pools” (ion concentrations) (Hartline in El-Leithy, 1989) were simulated (Tsay, 1993) using VLSI analog multipliers and DVCCs. Measurements taken on these chips are in line with neurophysiological data.

4.2 Silicon Retina

One of the best matches to date between analog VLSI circuitry and a biologically-based application is the silicon retina of (Mead 1989). This chip implements the first stages of invertebrate retinal processing and produces signals similar to those found in real retinas.

These CMOS circuits model photoreceptor cells (for computing light intensity and transducing it to an electrical signal), horizontal cells (to average photoreceptor outputs both spatially and temporally), and bipolar cells (to compute the difference between the photoreceptor and horizontal cell outputs). The ANN photoreceptors are fabricated as vertical bipolar transistors, which are parasitic to the CMOS fabrication process. A special multiplexor circuit is used to scan data out of the photo transistor array. The published results show good agreement with natural retinas including edge responses and Mach bands.

Another silicon retina implementation which includes tuned pixels is discussed in (Delbrück in Sánchez, 1993).

4.3 Associative Memories

An associative memory is a memory system whereby the location of the target memory cell need not be specified in order to retrieve its contents, rather recall is done by associating with properties of the data. ANN associative memories (such as the Hopfield Net), store patterns in weights such that when a 'noisy' pattern is presented, the complete pattern is produced from the memory. An analog VLSI implementation of a heteroassociative memory is briefly described below.

Boahen and his co-workers describe the implementation of a three-layer, 46 neuron heteroassociative memory (Boahen in El-Leithy, 1989). Using current mode circuits operating in subthreshold conduction, the chip contains a regular array of cells, each cell containing two synapses and a one bit weight memory cell. Inverters are used for thresholding neurons, current sources are used in the bias circuit, and a multiplier circuit is used in the synapse.

A class of adaptable associative memories is realized using DVCCSs incorporating Gilbert multipliers for transconductance weights in (Linares in Sánchez, 1993).

5. Discussion and Future Outlook

Analog VLSI offers the ANN world the distinct advantages of speed and real-time processing when compared to digital technology while suffering from relatively large size requirements and lack of standard cells. It also offers the ability to make continuous and speedy adjustments for adaptive neural networks and those needing efficient calculations of derivatives, as in back-propagation ANNs. Although the absolute error for analog components is typically larger than 5% the relative precision can usually be controlled to be under 0.1% when implemented in VLSI. Roughly this is the equivalent of eight bit digital resolution at hundreds to thousands of mega-Hertz. When working with the primary circuits discussed here, such as the DVCCS and the current mirrors, voltage and current differences matter most, so that it is the relative tolerance that is critical. In any event, because ANNs are by conception fault tolerant, precision is not usually of concern, even though the absolute precision of good digital circuits is not obtained.

VLSI neurons can have their dimensions comparable to those of biological neurons with considerably faster signal processing. However real neurons take full advantage of their three dimensional nature whereas most present day VLSI structures are essentially planar being embedded on the surface of a wafer. And even though connection wires can be routed under other wires in multiple metal VLSI constructs, and there

do exist some prototype 3-D processes, the technology is still quite limited in terms of three dimensional patterning and interconnecting. Consequently one of the major areas for the future is how to fabricate efficiently three-dimensional patterns. As we have seen above, there are somewhat standard components available for analog VLSI construction of ANNs with their characteristics largely at the disposal of the designer. This very necessity of obtaining variations in characteristics is what makes the analog world less constrained than the digital world, but by the same token it negates the use of standardized cells so advantageously used in digital hardware. Nevertheless, ANNs are amenable to a mixture of analog and digital realizations when it comes to the field of pulse-coded ANNs – the action potentials can be standardized and then realized by digital pulses while the synaptic effects can be most conveniently realized by analog devices since real-valued weights are involved.

A large number of other devices not mentioned above but of interest to specialized areas of ANNs are presently available for VLSI. Among such devices are charge coupled devices (CCDs), possibly for axon-like propagation or enzyme effect mimicking, floating-gate devices for long term storage of weights, and JFETs for less delicate fabrications. It should be noted that the MOS devices take minimal area but they are very subject to damage by static charge that can puncture the very thin gate oxide. For the future there are the very small resonant tunneling devices which use a different substrate than the silicon of present VLSI and molecular devices which probably show the greatest long range potential due to their minimal size and general signal handling.

6. Acknowledgments

The authors wish to thank Professor Arbib for the opportunity to present these ideas as well as Mr. C.-P. Chew, in clarifying various points especially concerning subthreshold models, and Mr. J.-D. Yuh for assistance on VLSI structures.

7. References

Bialko, M., and Newcomb, R. W., 1971, Generation of All Finite Linear Circuits Using the Integrated DVCCS, IEEE Transactions on Circuit Theory, Vol. CT-18, No. 6, November, pp. 733 - 736.

El-Leithy, N., and Newcomb, R. W., Eds., 1989, Special Issue on

Neural Networks, IEEE Transactions on Circuits and Systems, Vol. 36, No. 5, May.

* Geiger, R. L., Allen, P. E., Strader, N. R., 1990, VLSI Design Techniques for Analog and Digital Circuits, New York: McGraw-Hill. (* for general VLSI electronic circuits)

Graf, H. P., Jackel, L. D., 1989, Analog Electronic Neural Network Circuits, IEEE Circuits and Devices Magazine, Vol. 5, No. 4, July, pp. 44-55.

Kardontchik, J. E., 1992, Introduction to the Design of Transconductor-Capacitor Filters, Boston:Kluwer Academic Publishers.

Lee, B. W., Sheu, B. J., 1991, Hardware Annealing in Analog VLSI Neurocomputing, Norwell, Mass.: Kluwer Academic Publishers.

Linares, B., Sánchez, E., Rodriguez, A., Huertas, J., 1993, A CMOS Analog Adaptive BAM with On-Chip Learning and Weight Refreshing, IEEE Transactions on Neural Networks, Vol. 4, No. 3, May, pp. 445 - 455.

Mahowald, M., and Douglas, R., 1991 A Silicon Neuron, Nature, Vol. 354, December 19/26, pp. 515 - 518.

* Mead, C. A., 1989, Analog VLSI and Neural Systems, Reading, Mass.: Addison-Wesley. (* for VLSI layout, transistor physics and neural circuit applications)

Meador, J., Watola, D., and Nintunze, N., 1991, VLSI Implementation of a Pulse Hebbian Learning Law, Proceedings of the 1991 IEEE International Symposium on Circuits and Systems, Singapore, June, pp. 1287 - 1290.

Mueller, P., van der Spiegel, J., Blackman, D., et. al, 1989, Design and Fabrication of VLSI Components for a General Purpose Analog Neural Computer, in Analog VLSI Implementation of Neural Systems (Mead, C. A., and Ismail, M., Eds.), Boston: Kluwer.

Reiss, R. F., Ed., 1964, Neural Theory and Modeling, Stanford, CA.,:Stanford University Press.

Schmitt, O. H., 1937, An Electrical Theory of Nerve Impulse Propagation and Mechanical Solution of the

Equations of Nerve Impulse Propagation, The American Journal of Physiology, Vol. 119, No. 2, June 1, pp. 399 - 400.

Salam, F. and Wang, Y., 1991, A Real-Time Experiment Using a 50-Neuron CMOS Analog Silicon Chip with On-Chip Digital Learning, IEEE Transactions on Neural Networks, Vol 2, No. 4, July, pp. 461 - 464.

Sánchez, E., and Newcomb, R. W., Eds., 1992 & 1993, Special Issues on Neural Network Hardware, IEEE Transactions on Neural Networks, Vols. 3 & 4, No. 3, May.

Tank, D. W., and Hopfield, J. J., 1986, Simple "Neural" Optimization Networks: An A/D Converter, Signal Decision Circuit, and a Linear Programming Circuit, IEEE Transactions on Circuits and Systems, Vol. CAS-33, No. 5, May, pp. 533 - 541.

Taylor, G. W., 1978, Subthreshold Conduction in MOSFET's, IEEE Transactions on Electron Devices, Vol. ED-3, No. 3, March, pp. 337 - 350.

Tsay, S.-W., 1993, Implementation of Hartline Pools and Neural-Type Cells by VLSI Circuits, Advances in Control Networks and Large-Scale Parallel Distributed Processing Models, Vol. 2, (M. Fraser, Ed.), Norwood, NJ: Ablex.

* Zornetzer, S. F., Davis, J. L., and Lau, C., Eds., 1990, An Introduction to Neural and Electronic Networks, San Diego: Academic Press. (* for broad overview including VLSI neural circuits)

Figures

Figure 1. n th Neuron for a VLSI Hopfield Type ANN of N Neurons.

Figure 2. VLSI Transistors and Layouts: a) NMOS, b) PMOS, c) NPN BJT.

Figure 3. Passive Components and Layouts: a) MOS Capacitor, b) Snake Resistor.

Figure 4. MOS Current Sources, Resistors, and Current Mirrors: a) Voltage Variable Current Sources (Saturation Mode) $I = (K\frac{W}{L})(V - |V_{th}|)^2$ for $0 \leq V - V_{th} \leq V_{DS}$ or Voltage Variable Resistor (Ohmic mode) $I = G(V)V_{DS}$, $G(V) = (2K\frac{W}{L})(V - |V_{th}|)$ for $0 \leq V_{DS} \ll V - V_{th}$, b) Unidirectional, $I = (K\frac{W}{L})(V - V_{th})^2 \cdot 1(V - V_{th})$, and Bidirectional Resistors, c) NMOS and PMOS Unidirectional Current Mirrors, $I_{out} = ((L_2/W_2)/(L_1/W_1))I_{in}$, d) Bidirectional Current Source.

Figure 5. DVCCS: a) Basic Configuration and Circuit Symbol, T=M (MOS) or Q (BJT), b) Improved Saturation Mode DVCCS, c) CMOS Inverter (to Attach to a DVCCS Output to Form a DVCVS).

Figure 6. Circuits Using the DVCCS: a) Resistor ($R = 1/g_m$),

b) Degree One Low-Pass Filter, $(V_{\text{out}}/V_{\text{in}}) = (g_m/C)/(s + (g_m/C))$; s=complex frequency

c) Degree Two Low-Pass Filter, $(V_{\text{out}}/V_{\text{in}}) = [(g_{m1}/C_1)(g_{m2}/C_2)]/[s^2 + ((g_{m1}/C_1) + (g_{m2}/C_2))s + (g_{m2}/C_2)(g_{m1} - g_{m3})/C_1]$.

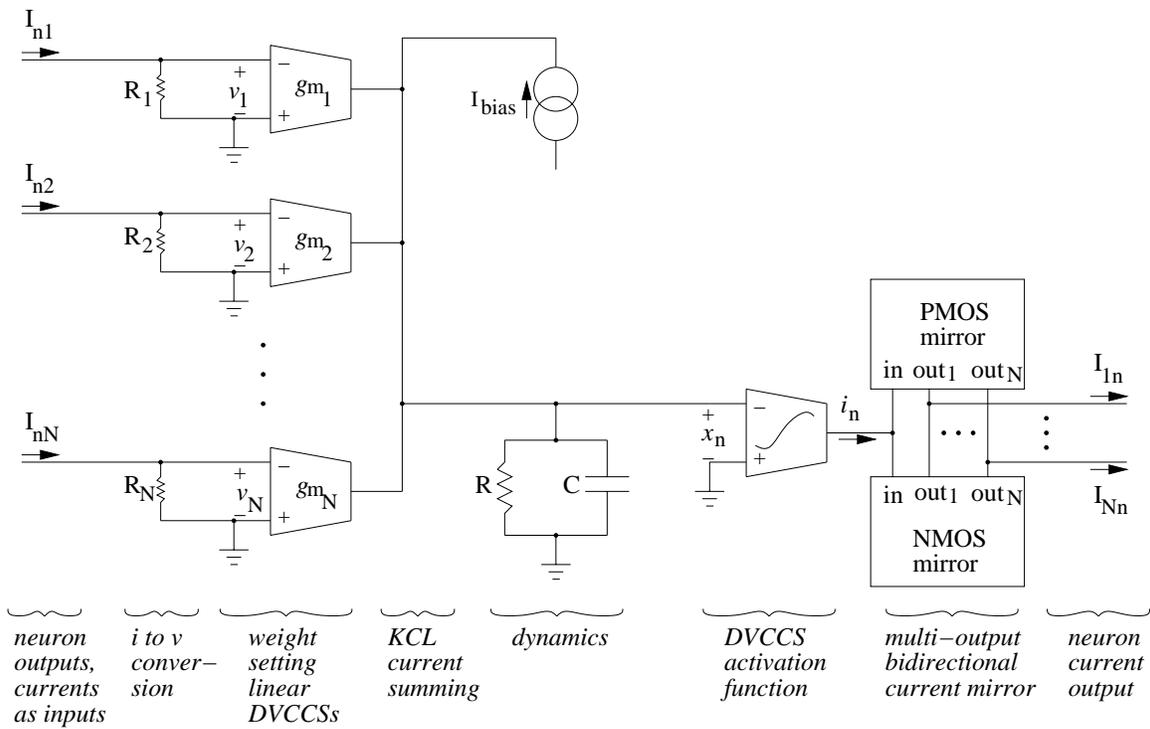
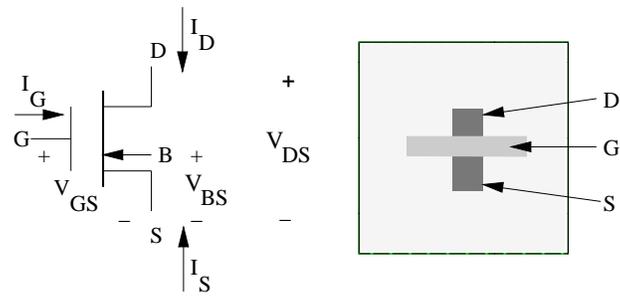
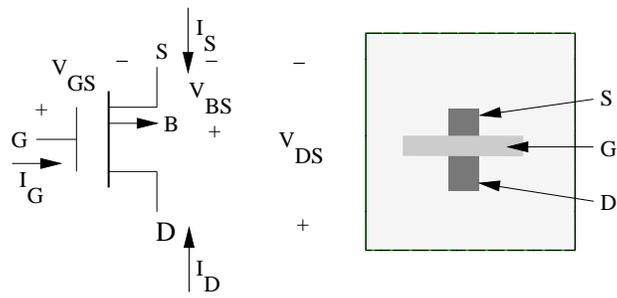


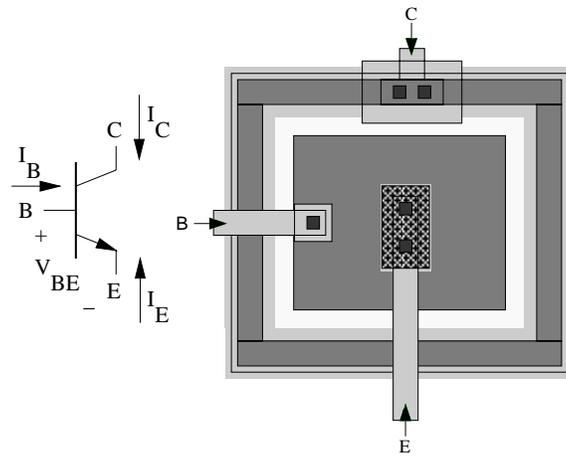
Figure 1:



(a)

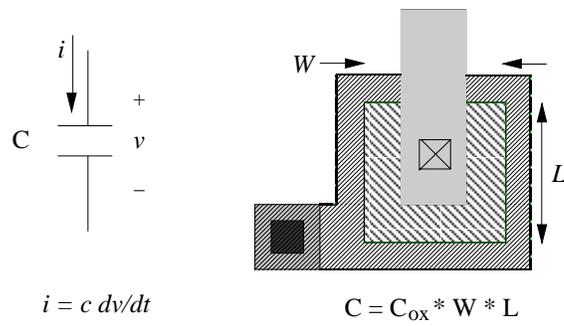


(b)

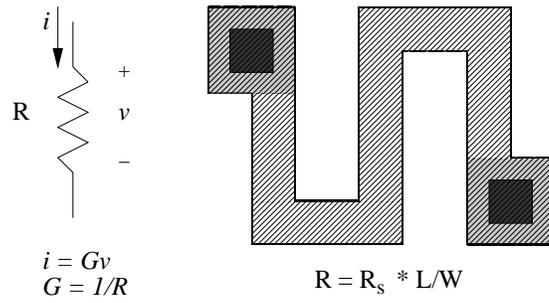


(c)

Figure 2:



(a)



(b)

Figure 3:

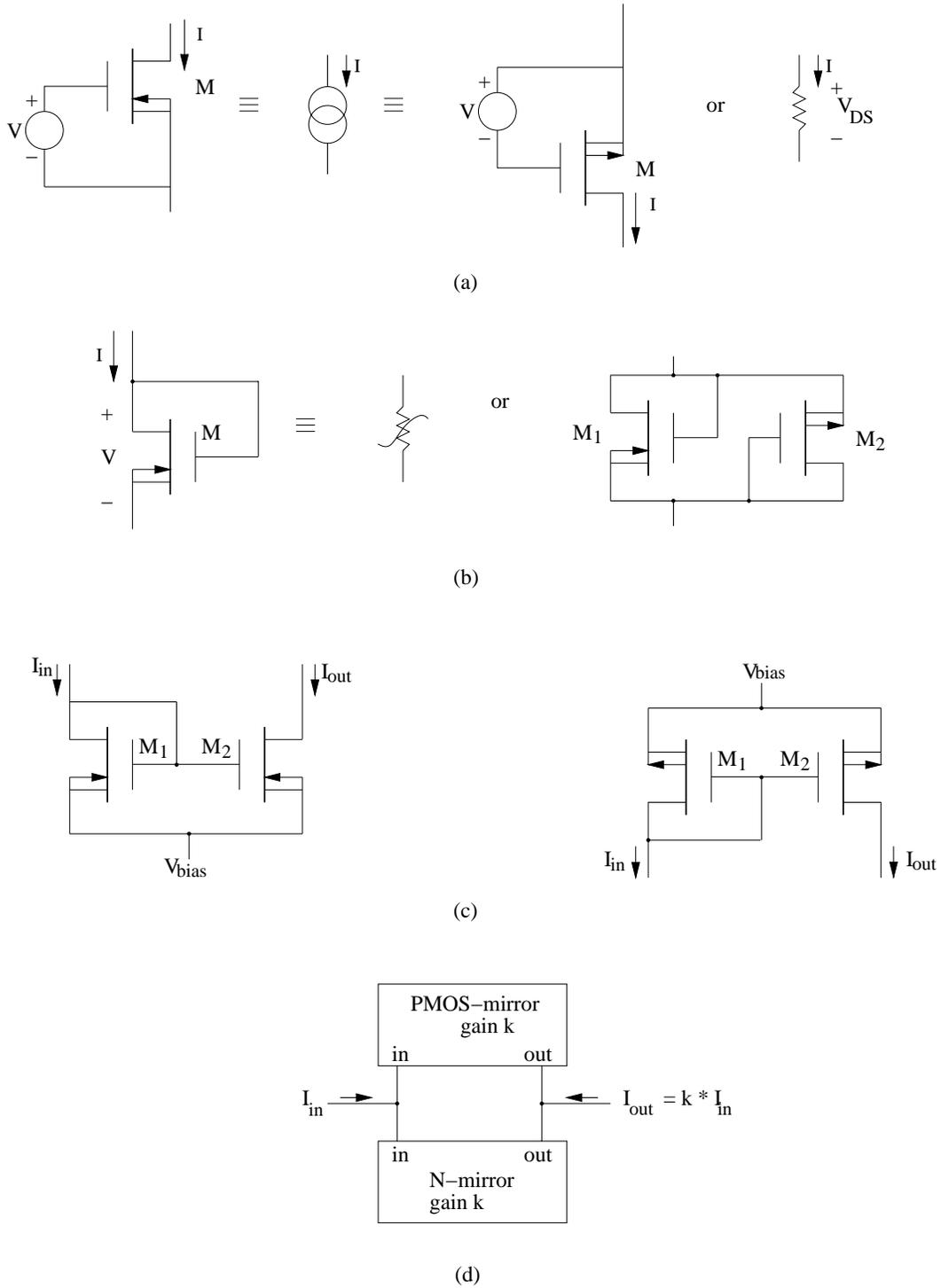


Figure 4:

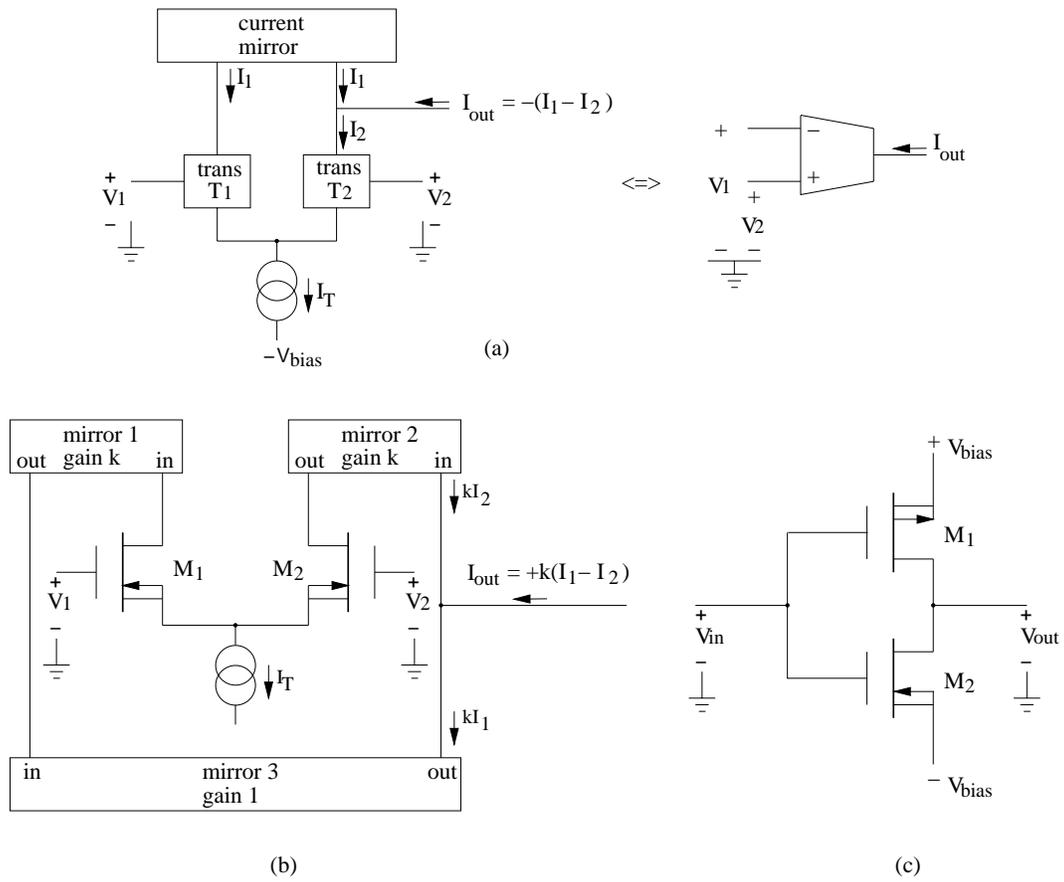
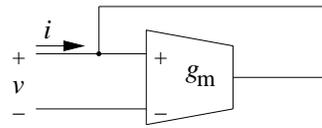
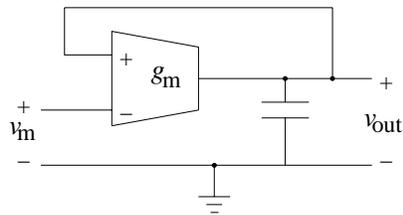


Figure 5:

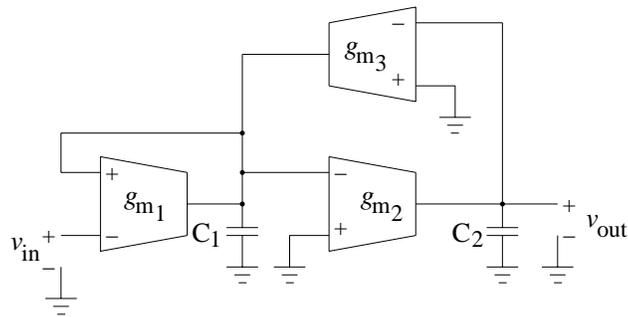


$$i = g_m v$$

(a)



(b)



(c)

Figure 6: