A BASIC MOS NEURAL-TYPE JUNCTION*
A PERSPECTIVE ON NEURAL-TYPE MICROSYSTEMS

N. El-Leithy#@, R.W. Newcomb#, and M. Zaghloul@

#Microsystems Laboratory
Electrical Engineering Department
University of Maryland

@Electrical Engineering and Computer Science Department
George Washington University

ABSTRACT

This paper presents a perspective on neural-type electronics as a novel device-circuit design approach inspired by neural structures. Such an approach attempts to provide a well-engineered, theory-tested substrate on which the emerging computational, adaptive, learning, massively-parallel architectures can be built. Further, the paper proposes an MOS neural-type junction as an example of the adopted design philosophy. The proposed circuit captures the temporal and spatial activity of neural elements through N input structures. Each of these structures comprises an initial adaptive weighting stage feeding into a charge-accumulating system. The net excitatory drive resulting from space and time effects stimulates a dynamic firing output stage. The circuit permits the implementation of information processing elements in which pulse frequency modulation is the coding mechanism. Some modifications to the design, implications and future prospects are outlined.

A. Perspective

MOTIVATION AND BACKGROUND:

While from the bounded level of our mind
Short views we take, nor see the lengths behind; (1)

The development of electronics is at an important crossroads. Considerable uncertainty surrounds what materials, fabrication, electronic processes, devices, circuit design techniques, systems, etc., are to be significant at the end of the century. As sub-micron geometries are approached, there are signs of crisis in a number of inter-related areas which confront the progression to very large integration densities and high speed systems. Major problems arise due to the fact that, at this quantum scale, no scaling theory exists which can simultaneously scale all the physical factors affecting device operation; hence, neither the effects of the electric fields nor the influence of the device environment can be treated satisfactorily (2,3,4).

Therefore, as the semiconductor technology approaches the complexity levels of biological systems, it becomes pertinent to ask what radical changes in microelectronics concepts are required if we are ever to match the power, efficiency and reliability of information processing mechanisms within the nervous system. Can we start departing from the design ethics of classical circuit and device theory in response to the emerging, highly demanding, intelligent systems that we aspire to construct?

Inspired by such questions, the neural-type microsystems (NTM) technology has evolved to provide novel device-circuit design concepts that capture the essence of neural signal processing in a simple, elegant form complying with state-of-the-art microelectronics trends (5). As such, NTM's offer potentialities for establishing a well-engineered and theory-tested substrate on which the various powerful computational, adaptive and learning models with their massively parallel, non-von Neumann architectures can be built.

*This research was supported by the US National Science Foundation under Grant ECS 85-06924.

The history of NTM's goes back to the early 1960's, when H. Crane introduced this design philosophy in his Ph.D. dissertation (6), leading to the conception of the neuristor (7). This concept was stimulated by the desire to construct new wiring schemes for highly miniaturized electronic systems, in order to overcome the problems of signal degradation arising from the passive characteristics and parasitic coupling of the interconnections. Though influenced by neural structures, it must be emphasized that Crane's ideas evolved as an approach more in the spirit of revolutionizing electronic design techniques than actually developing new methodologies of problem solving, learning and pattern recognition as was the main trend of investigations at that time (8,9,10).

However, at the time of its conception, practical electronic realizations of the neuristor were not available except for Nagumo's tunnel-diode transmission line (11), the semi-distributed line using unijunction transistors (12), the syneuristor of Waldron (13), the pnpn diode neuristor line (14), and the distributed superconductive line utilizing the Josephson effect (15). Such implementations lacked the integrability feature which led to a world-wide search for circuits that could be realized in VLSI with such coming forth in the 1970's from Poland in bipolar form (16) and from the US in MOS structures (17). Conceived applications of the resulting circuitry as general-purpose signal processors (18, 19), retinal-type sections for robotic vision (20), logic elements (21), and components of neural-type robotics (22,23), have been investigated.

Expanding upon earlier work, recent papers (28,36) as well as this paper attempt to refine the philosophy, establish the theory and possibly unfold a whole new field of electronic arts.

DESIGN THEME:

'Fired' at first sight with what the Muse imparts,
In fearless youth we tempt the heights of Arts, (1)

Departing from the prejudices and constraints of classical circuitry, neural-type electronics can be characterized by the following:

(1) Exploitation of the highly nonlinear device-circuit interactions which (through electronic inspiration) tend to capture the complex, dynamic behavior of neural elements. As an example, instead of controlling parasitic linkages among circuit elements, an attempt is made to allow the existing inter-device coupling to participate in achieving the desired operation. Ultimately, the design should lead to holistic integrated circuit layouts.

(2) Adaptability of most of the circuit components and interconnects in response to spatial and temporal information patterns. Toward this goal, one of the methods currently being investigated (by the authors) involves the use of floating-gate MOS structures. These allow dynamic nonvolatile changes of device and circuit characteristics.

(3) The basic theory relates to analog charge manipulations within the circuit: being established, maintained or destroyed continuously at various points in the structure. It is well known that stimulation of many forms of neural components occurs due to the passage of a quantity of charge through a membrane. Incidentally, analog circuit theorists (24), as well as device modelers (25,26) have come to realize the convenience and usefulness of describing the device and circuit activity in terms of charge as a means of unifying the various analytic techniques utilized. Such charge theory seems to evolve naturally as an appropriate description of the temporal and spatial signals traversing a neural-type network.

(4) Pulse frequency modulation is adopted since it is believed to be one of the best methods of information coding in the neural-type environment in terms of still unknown criteria, which could be component or structural simplicity, distortion attenuation, etc. The use of such a coding mechanism in neural-type electronics may reveal the contribution of the impulse generation process to the overall system dynamics.

Classes of NTM' s:

In NTM's three general classes of subsystems have been isolated functionally but not necessarily structurally: those dynamically generating the signals (neural-type cells NTC's); those actively transmitting the signals with no attenuation (neural-type lines NTL's); and those processing and combining or mixing the signals (neural-type junctions NTJ's). Most of the earlier realizations focused on obtaining electronic NTL's having the burst-like signal propagation properties of a nerve axon, namely: 1) threshold of stimulability, 2) uniform velocity

of propagation, 3) attenuationless transmission, 4) characteristic pulse generation, and 5) refractoriness. One such implementation, shown in Figure 1, consists of a cascade of three-MOS-transistor sections exhibiting zero power consumption during the resting state (no excitation) by virtue of the enhancement-mode MOS characteristics (27). In fact, an NTL is formed by discretely coupling NTC's (5), therefore, each isolated section of the cascade is, in effect, an MOS NTC shown separately in Figure 2. Such an NTC comprises an input MOS structure Q1, which establishes the cell threshold for incoming positive spikes and provides unilateral transmission and isolation, coupled to a dynamic firing output stage consisting of the Q2-Q3-RC structure. The Q2-R6C combination creates the functional nonlinearity necessary for pulse shaping, while the simple insertion of the feedback transistor Q3 allows for encoding the input signal amplitudes into output neural-type pulse repetition rates, whereby the cell produces a pulse train whose frequency is a function of how much above threshold the steady input stimulant is - quite reminiscent of firing zones in real neurons! Figure 3 illustrates a transient analysis run on MICROCAP for the MOS NTC (28). Obviously, the circuit is very nonlinear and its operation critically depends upon hysteresis generated through feedback and inherent parasitics of the transistors (28).

This brings the discussion to the NTJ which is the key information processor in a neural-type system handling both spatial and temporal patterns of activity.

B. BASIC MOS NEURAL-TYPE JUNCTION

Inspired by present day knowledge of the synaptic-dendritic tree structures, De Claris introduced the concept of neural-type junctions (NTJ's) in a mathematical form (29). From a circuit viewpoint, an NTJ handles large numbers of incoming pulse trains, processes them in an elaborate way whereby different inputs are given different weights, produces a net activity level through integrative action over time and space, then, depending on that level, stimulates a firing output stage which responds by generating a distinct pulse train. These features are incorporated in the circuit proposed by this paper and shown in Figure 4. The circuit consists of N input structures, each of which comprises an initial weighting stage feeding into a charge-accumulating system. All the N structures, if properly excited, can draw current through a single summing resistor or channel by means of a current-source coupler. The summing element then stimulates the dynamic output stage to generate the required response. The various properties of the MOS NTJ will be explained.

1) Input-Weighting:

A neural-type pulse train arrives at each input structure where it experiences initial processing (refer to Figure 5 which depicts a single input channel connected to the firing output stage). For each of the N input mechanisms, the threshold voltage V_T of Q1 (among other parameters which will be assumed constant) determines the transfer efficiency of the corresponding initial stage. Input weights are set via threshold voltage (V_T) adjustment, thus, the strength of signal transmission through Q1 can be controlled according to the desired operation.

V_T is a function of a number of parameters including gate material, gate insulation material, gate insulation thickness, channel doping, impurities at silicon-insulator interface, and voltage between source and substrate V_{SB}. Consequently, weights can be adjusted by controlling either the chief process parameters - namely, the doping concentration and the oxide thickness - or the chief circuit operational parameter - namely, V_{SB}.

A variable dc reverse bias voltage V_{SB} was applied experimentally to every Q1 in the circuit (not shown). A clever circuit design generating its own substrate bias as in (30) can also be used. Accordingly, the effective threshold becomes:

$$V_T = V_{TO} + \gamma[(V_{SB} + 2\phi_F)^{1/2} - (2\phi_F)^{1/2}] \qquad (30)$$

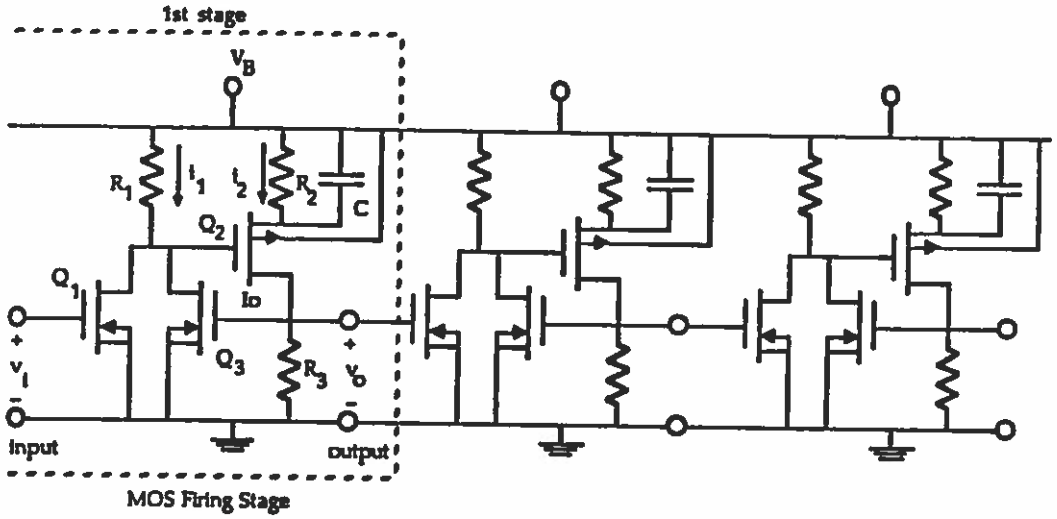


Figure 1 : An MOS Neuristor Line (27)

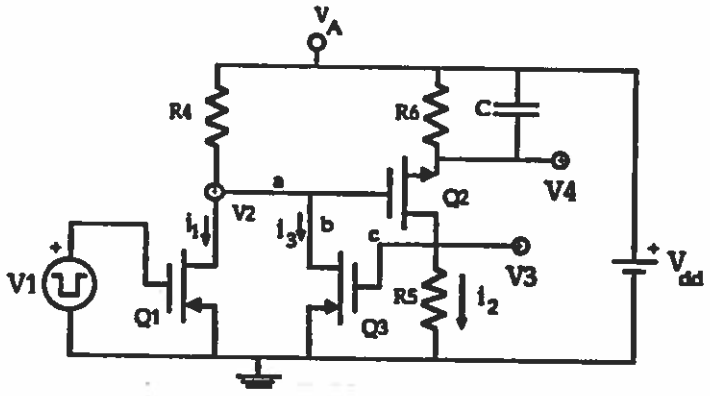


Figure 2 : A Dynamic Firing NTC (28)

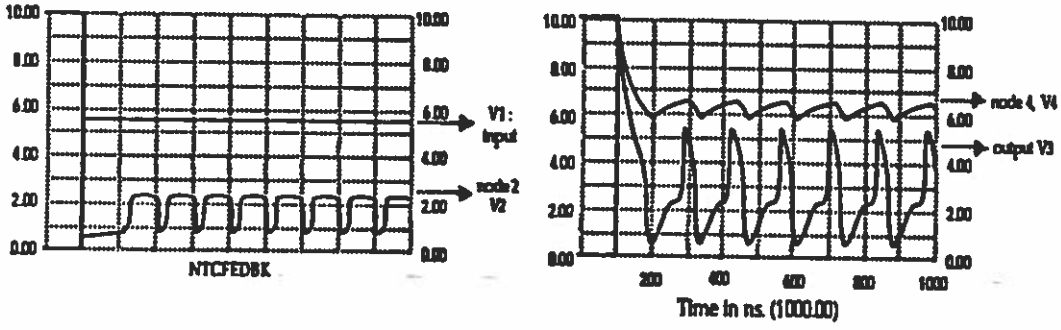


Figure 3 : A MICROCAP transient analysis for the NTC

THE KEY SIGNAL PROCESSOR CAPTURING BOTH SPATIAL AND TEMPORAL ACTIVITY

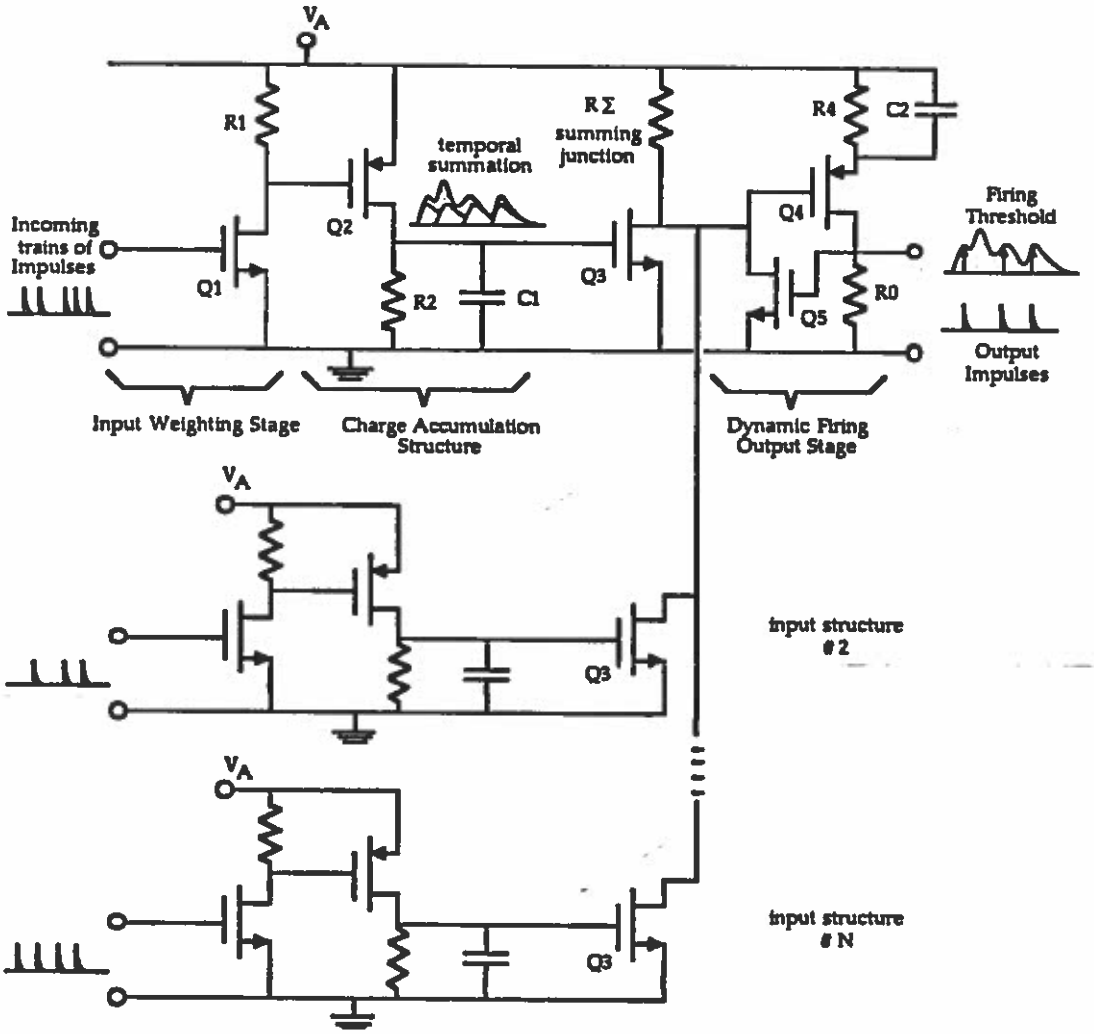


Figure 4 : The Basic MOS NTJ

where ϕ_F is a constant, V_{SB} is the substrate bias, V_{TO} is the threshold voltage for $V_{SB} = 0$, and

$$\gamma = \left(\frac{t_{ox}}{\epsilon_{ox}}\right)\sqrt{2q\epsilon_{si}N}$$

in which q is the electron charge, ϵ_{ox} is the dielectric constant of the gate insulator, ϵ_s is the dielectric constant of the Si substrate, and N is the concentration density of the substrate. Note that γ is a SPICE model parameter thus enhancing the design flexibility.

Varying V_{SB} to adjust V_T according to the above equation allows us to adaptively control the strength of the signal flow complying with a certain problem definition. Moreover, nonvolatile threshold changes in response to past history of operation or Hebbian learning rules are possible through replacing each Q1 by a floating-gate transistor (FAMOS). FAMOS's and double-gate MOS devices are quite developed and popular in EEPROM integrated circuits (31, 32). This approach is being thoroughly investigated by the authors.

2. Temporal Summation:

The shown input structures of figure 4 are excitatory in nature although inhibitory effects can be achieved through the use of complementary p-channel transistors. For figure 5, if an incoming pulse is larger than the threshold voltage V_T of Q1, it turns Q1 on, hence, drawing current through R1 which eventually turns Q2 on when the voltage of node 1 falls sufficiently below VA. Thus, the voltage-controlled current source behavior of Q2 allows an amount of current flow dependent on the extent to which Q1 is drawing charge to its channel. This current flow begins to build charge on the capacitor C1. The rate of such charge build-up is determined by the R2.C1 time constant as well as the transistor parasitics. If a second pulse arrives fast enough, before C1 is allowed to fully discharge by the trailing edge of the initial exciting pulse, the same mechanism of current flow occurs. This results in dumping more charge on C1. Through such a cumulative effect, an input pulse train is decoded into charge build-up over time within the circuit. If the gate voltage of Q3 (node 3) established by charge accumulation on C1 exceeds the device threshold voltage V_T, the current-source coupler Q3 turns on. Consequently, a chain of activities in the dynamic output stage identical to that witnessed in the NTC of figure 2, will be initiated. A MICROCAP transient analysis run illustrating these mechanisms is shown in Figure 6. In this figure, we monitor node 6: the input, node 3: voltage build-up on C1, node 4: the summing point, and node 7: the output. Without the time integration property of the input structure, an incoming pulse train of relatively low amplitude may not tirgger an output. However, with such time effects, an output signal will always be generated after some delay, provided that the input frequency is high enough. The time delay after which an output is generated, is a function of the circuit parameters, device parasitics and intensity of stimulation.

3) Spatial Summation:

All the input systems comprising the weighting stages R1-Q1, the charge-accumulators Q2-R2-C1, and the current-source couplers Q3 are tied to node 4 [gate voltage of Q4]. Hence, the current flow in the summing or junction resistor $R\Sigma$ is the sum of the drain currents of all the Q3's.

$$I_{R\Sigma} = \sum_{i=1}^{N} I_{DSQ3i}$$

where N is the number of input transmission channels and I_{DSQ3i} is the transistor channel current of each Q_{3i}.

In this way, the gate-to-source voltage of Q4 can be influenced by the activity of all the input channels. Because of the excessive current drainage through the summing junction $R\Sigma$, an improved circuit results by replacing it with a diode-connected n-channel transistor having $V_T = 0$, as shown in Figure 5.

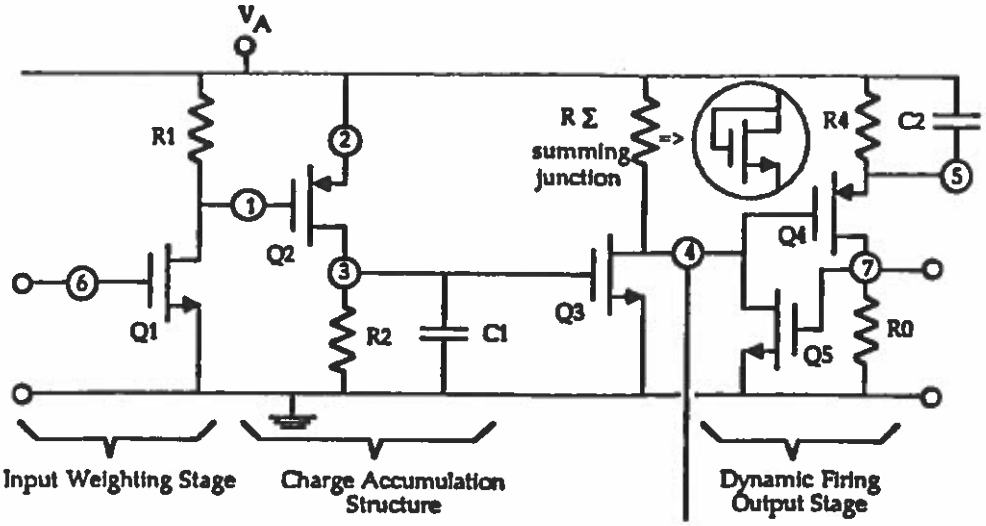


figure 5 : A single input structure coupled to a firing output stage.

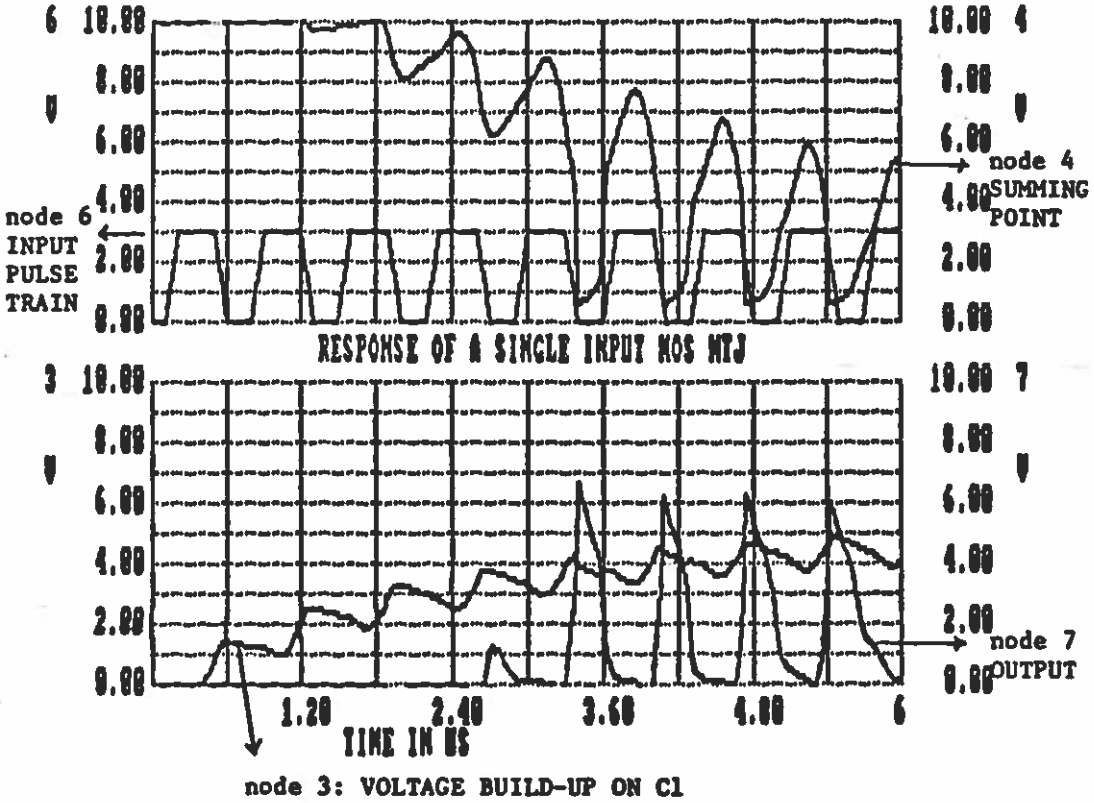


Figure 6 : Monitoring nodes 6, 4, 3, and 7

4. Output Encoding of the NET INPUT DRIVE:

The net drive to Q4 is, in effect, the result of both time and space integration - time effects are coupled through Q3 while spatial effects are combined in the junction resistor RΣ (or channel QΣ). The dynamic firing output stage [Q4-R4-C2-R0-Q5], through its Q5 feedback mechanism and hysteretic properties, generates a neural-type pulse train (bottom trace of figure 6) whose frequency is a function of the drive intensity. The ability of the output stage (which is, actually, an NTC) to encode information as pulse repetition rates is illustrated in Figure 3.

C. CONCLUSION - A LOOK INTO THE FUTURE

But more advanced, behold with strange surprise
New distant scenes of endless science rise! (1)

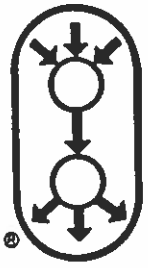
This paper presented an MOS neural-type junction module that handles temporal and spatial activities in a highly nonlinear, but elegant, manner. It is an indication of how electronic arts can capture complex neural signal manipulations. The processing power of the proposed structure, though not analytically investigated, is potentially immense. Thorough analysis and mathematical formulation of the circuit behavior are ongoing. The aim is to establish a design methodology by means of which specific functions can be represented to solve associated problems. Implementation of adaptive properties by tailoring the analog charge storage capacity of double-gate MOS structures is also under study. It is apparent that the concept of analog charge manipulation seems to be very descriptive of the circuit operation. In order to further emphasize the charge handling theory, all the circuit resistors can be replaced by MOS transistors as was indicated in the case of RΣ - a process which will certainly enhance the chip integrability feature of the design. Charge control models of MOS device operation are currently being developed by the authors.

However, it is already apparent that the highly successful silicon MOS technology will reach the end of its evolutionary travel at about the 0.1μm scale. The search for new device materials and novel electronic properties is likely to accelerate in the next stages of microelectronics. A major development concerns the III-V based materials which can now be configured into a variety of heterostructures with a wide range of exploitable physical phenomena (33,35). The current move to GaAs and related compounds is stimulated not only by their low power, high speed advantages but, also, by their ability to form quantum well architectures, in which materials can be fabricated with "atomic" structures specially configured for their electronic properties (34). Advanced fabrication and patterning techniques may offer the opportunity to tailor the inter-device interactions and externally control the device parasitics in order to build functional adaptive systems that are quite immune to the problems of metallic interconnections and system partitioning. These architectures would be characterized by nonclassical signal processing (through charge transfer mechansims) inducing certain device and system properties highly reminiscent of neural structures. The authors believe that the merging of the evolving neural network concepts with the emerging three-dimensional superlattices (34,35) may offer revolutionary, challenging and exciting prospects for the future of electronics.

References

1. "From an Essay on Criticism," by Alexander Pope.
2. Hoeneisen, B., and Mead, C.A., "Fundamental limitations in microelectronics-I.MOS Technology,"Solid State Electronics, Vol. 15, pp. 819-829, 1972.
3. Barker, J.R., in Physics of Non-linear Transport in Semiconductors, D.K. Ferry, J.R. Barker,C. Jacoboni, editors, New York: Plenum Press, 1980.
4. Solid State Sciences Committee, National Research Council, Microstructure, Science, Engineering and Technology, National Academy of Sciences, Washington, D.C., 1979.
5. El-Leithy, N. and Newcomb, R.W., "Overview of neural-type electronics," Conference Proceedings 28th Midwest Symposium on Circuits and Systems, Louisville, KY, August 19-20, 1985.
6. Crane, H.D., "Neuristor Studies," Stanford Electronics Lab., Stanford University, CA, Technical Report #1506-2, July 11, 1960.
7. Crane, H.D., "Neuristor: a novel device and system concept," Proc. IRE, Vol. 50, No. 10, pp.2048-2060, 1962.

8. McCullouch, W.S., and Pitts, W. Bulletin Math. Biophysics, Vol. 5, pp. 115-133, 1963.
9. Minsky, M. and Papert, S., "Perceptrons: An Introduction to Computational Geometry," MIT Press, Cambridge, MA, 1969.
10. Rosenblatt, F., "Principles of Perceptrons," Spartan, Washington, D.C., 1962.
11. Nagumo, J., Armimoto, S. and Yoshizawa, S., "An active pulse transmission line simulating a nerve axon," Proc. IRE, Vol. 50, No. 10, pp. 2062-2070, 1962.
12. Ambroziak, A., "Semidistributed neuristor line using unijunction transistors," Solid State Electronics, Vol. 7, pp. 259-265, Pergamon Press, 1964.
13. Waldron, M.B., "Syneuristor - A device and concept," Technical Report #6560-22, Information Systems Laboratory, Stanford University, Jan. 1971.
14. Cote, A.J., Jr., "Simulating nerve networks with four layer diodes," Tunnel diode and Semiconductor circuits; J. Carroll, editor, McGraw-Hill, 1963.
15. Reible, S.A., and Scott, A.C., "Pulse propagation on a superconductive neuristor," J. of Appl. Physics, Vol. 46, No. 11, pp. 4935-45, November 1975.
16. Wilamowski, B.W., Czarnul, Z., and Bialko, M., "Novel Inductorless Neuristor Line," Electron Lett., Vol. 11, No. 15, pp. 355-356, 1976.
17. Newcomb, R.W., "MOS Neuristor Lines," in Constructive Approaches to Mathematical Models, C.V. Coffman and G.J. Fix, editors, Acad. Press, 1979.
18. Kohli, C.K., and Newcomb, R.W., "Voltage controlled oscillations in the MOS neural line," Proceedings of the 20th Midwest Symposium on Circuits and Systems, Part 1, August, 1977.
19. Czarnul, A., Kiruthi, G., and Newcomb, R.W., "MOS neural pulse modulator," Electroncis Letters, December, 1979.
20. Ajmera, R.C., Newcomb, R.W., Chitale, S.A., and Nilsson, A., "VLSI Brightness module for robot retinas," Proc. of the 26th Midwest Symposium on Circuits and Systems, pp. 70-73, Puebla, August, 1983.
21. Gutierrez, J., Mulet, J., and Warzansky, W., "Boolean multiple-valued logic based on neuristor properties," Proc. of the 26th Midwest Symposium on Circuits and Systems, pp. 68-69, Puebla, August, 1983.
22. Chande, A., DeClaris, N., and Newcomb, R.W., "Neural-type robotics - An Overview," Proc. of the 26th Midwest Symposium on Circuits and Systems, pp. 68-69, Puebla, August, 1983.
23. Niznik, C. and Newcomb, R.W., "Computer networking capacity in robotic neural systems," Computer Communications, Vol. 7, No. 2, pp. 85-91, April, 1984.
24. Allen, P.E. and Sanchez, E., Sinencio, "Switched Capacitor Circuits," NY: Van Nostrand Reinhold, 1983.
25. Sparkes, J.J., "A study of the charge control parameters of transistors," Proc. of the IRE, Vol.48, No. 10, pp. 1696-1705, October, 1960.
26. Hodges, D.A. and Jackson, H.G. (eds.), "Analysis and Design of Digital Integrated Circuits, McGraw-Hill Book Co., pp. 206-228, 1983.
27. Kohli, C.K., and Newcomb, R.W., "An integrable MOS neuristor line," Proc. of the IEEE, pp. 1630-1632, November, 1976.
28. Newcomb, R.W. and El-Leithy, N., "Semistate description of an MOS neural-type cell," Proc. of the MTNS Meeting, Phoenix, AZ, June 15-19, 1987.
29. DeClaris, N., "Neural-type junctions: A new circuit concept," Proc. of the IEEE MidwesternSymposium on Circuits and Systems, pp. 268-271, 1976.
30. Ong, D.G., "Modern MOS Technology: Processes, Devices and Design," NY: McGraw-Hill, pp. 219-220, 1984.
31. Kahng, D., and Sze, S.M., "A floating gate and its application to memory devices," The Bell System Technical Journal, Vol. XLVI, No. 6, August, 1967.
32. Wang, S.T., "On the I-V characteristics of Floating-gate MOS transistors," IEEE Transactions on Electron Devices, Vol. ED-26, No. 9, 1979.
33. Barbe, D.F., ed., "Very Large Scale Integration," New York: Springer-Verlag, 1980.
34. Capasso, F., "Band-gap engineering: From physics and materials to new semiconductor devices," Scien, Vol. 235, No. 4785, pp. 172-176, January, 1987.
35. Ploog, K., and Dohler, G.H., "Compositional and doping superlattices in III-V semiconductors," Advances in Physics, Vol. 32, No. 3, pp. 285-359, 1983.
36. El-Leithy, N., and Newcomb, R.W., "A Semistate Model for Neural-type Junction Circuits," Proc. of the MTNS Meeting, Phoenix, AZ, June 15-19, 1987.



IEEE FIRST INTERNATIONAL CONFERENCE ON NEURAL NETWORKS

Sheraton Harbor Island East
San Diego, California
June 21-24, 1987

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

EDITORS:
MAUREEN CAUDILL
CHARLES BUTLER

IEEE CATALOG NUMBER 87TH0191-7

SAN DIEGO SECTION
SYSTEMS, MAN, AND CYBERNETICS SOCIETY
CONTROL SYSTEMS SOCIETY
ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY