A Basic Building Block Approach to CMOS Design of Analog Neuro/Fuzzy Systems

F. Vidal-Verdú[1], A. Rodríguez-Vázquez[2], B. Linares-Barranco[2] and E. Sánchez-Sinencio[3]

[1]Dept. de Arquitectura y Tecnología de Computadores y Electrónica.
Universidad de Málaga, Plaza El Ejido sn, 29013-Málaga, Spain.
[2]Dept. of Analog Design, Centro Nacional de Microelectrónica
Edificio CICA, Avda. Reina Mercedes sn, 41012-Sevilla, SPAIN
[3]Dept. of Electrical Engineering, Texas A&M University
College Station, TX 77843, U.S.A.

Abstract-This paper outlines a systematic approach to design fuzzy inference systems using analog integrated circuits in CMOS, standard VLSI technologies. Proposed circuit building blocks are arranged in a layered neuro/fuzzy architecture composed of five layers: fuzzification, T-norm, normalization, consequent, and output. Inference is performed by using Takagi and Sugeno's if-then rules, particularly where the rule's output contain only a constant term-- a singleton. A simple CMOS circuit with tunable bell-like transfer characteristics is used for fuzzification. Input to this circuit are voltage while output are current. Circuit blocks proposed for the remaining layers operate in current-mode domain. Innovative circuits are proposed for the T-norm and the normalization layers. The other two layers use current mirrors and KCL. All proposed circuits emphasize simplicity at the circuit level-- a prerequisite to increasing system level complexity and operation speed. A three-input, four-rule controller has been designed for demonstration purposes in a 1.6μm CMOS single-poly, double-metal technology. We include measurements from prototypes of the membership function block and detailed HSPICE simulations of the whole controller. These results operation speed in the range of 5MFlips with systematic errors below 1%.

I. INTRODUCTION

Software implementations of fuzzy inference systems typically operate below Kflip rate (flip stands for fuzzy logic inferences per second), not fast enough for many high-speed control problems, like those related to automotive engines [1]. During the last few years different authors have focused on the development of dedicated *hardware*, using IC technology [1], [2], [3], to overcome this drawback . In particular, *analog* circuits are worth considering for this application due to the intrinsically higher *speed* and lower *power* consumption than their digital counterparts [4]. Functional *efficiency* (measured as the device count for a given operation) of analog circuits is also much larger than for digital, due to the possibility of versatile exploitation of small analog devices (formed by a few transistors) for a wide variety of low-level linear and non-linear processing required for fuzzy inference. Finally, the intrinsically lower accuracy of analog circuits does not seem to be a major limitation for most fuzzy system applications, where *accuracy* requirements range from 6 to 9 bits-- affordable with

even the cheapest VLSI technologies [4],[5].

Previous proposals for analog fuzzy circuitry use bipolar transistors and/or linear resistors [2],[3] -- not readily implementable in the standard VLSI technology. Consequently, these ICs are expensive to produce, and not fully compatible to other conventional digital circuitry that may be needed to integrate together with the fuzzy circuitry, for complex control tasks. To overcome this drawback, all circuits proposed in this paper use MOS transistors as the only circuit primitive, and thus are fully compatible with the cheapest CMOS single- poly scaled technologies.

Since one of the major problems encountered in fuzzy systems is how to capture expertise, our approach focuses on algorithms that enable the incorporation of learning; in particular the *neuro-fuzzy* models proposed in [6]. We propose a number of CMOS building blocks supporting our design approach: membership function circuit, T-norm circuit, and normalization circuit. Detailed analysis of a complete bell- shaped membership function controller has been carried out to identify the main system parameters and error sources. Finally, we include measurements and HSPICE simulation results (for 1.6μm CMOS technology) to illustrate performance of the proposed building blocks and network architecture.

III. NEURO-FUZZY SYSTEM ARCHITECTURE

Circuit building blocks are arranged in the *layered* architecture of Fig.1 [6], where it is implicitly assumed that inference is performed using Takagi and Sugeno's *singleton* algorithm [7]. Referring to Fig.1 we can identify the catalog of analog functions needed to implement neuro/fuzzy controllers:

Layer 1: It contains a node per each fuzzy label of the input variables. Layer input is $x^T=[x_1, x_2 \ldots , x_M]$, and layer output is the matrix of matching degrees,

$$[s_{ji}] = [\mu_{ji}(x_j)] \qquad 1 \leq j \leq M, 1 \leq i \leq N \qquad (1)$$

where $\mu_{ij}(x)$ is the membership function associated to the i-*th* linguistic label of the j-*th* input variable. Hence, each node in this layer realizes a nonlinear transforma-
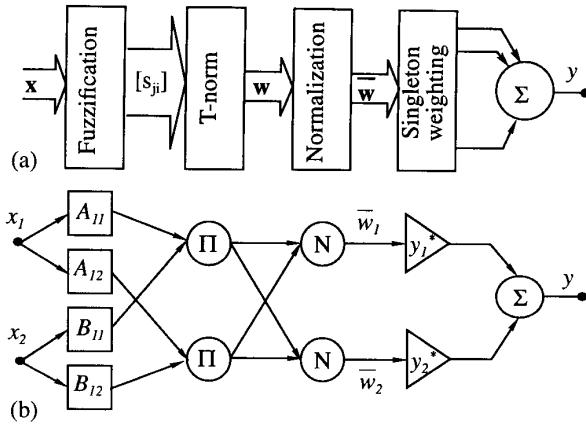
Fig.1: (a) Singleton neuro-fuzzy architecture; (b) Exemplary architecture for two input and two rules.

tion.

Layer 2: It maps this matrix of matching degrees into the vector of firing rule activities, $w^T=[w_1, w_2..., w_N]$. Each vector component is calculated by a corresponding processing node as follows,

$$w_i = min\ (s_{1i}, s_{2i}, ..., s_{Mi}) \qquad (2)$$

Layer 3: Each node in this layer calculates the averaged firing activity of its corresponding rule, starting from the vector of firing activities, as follows,

$$\bar{w}_i = w_i / (\sum_{i=1,N} w_i) \qquad 1 \le i \le N \qquad (3)$$

Layer 4: This layer multiplies each component of $\bar{w}$ by its corresponding singleton, thus obtaining the vector of graded rule's consequents,

$$z_i = \bar{w}_i y_i^* \qquad 1 \le i \le N \qquad (4)$$

Layer 5: This contains a single node which aggregates the consequent outcome of the individual rules, to obtain the inferred output.

$$y(x) = \sum_{i=1,N} z_i \qquad (5)$$

Nodes in layers 1 and 4 are *adaptive*, while the remaining have a fixed function. By setting the parameters that control the shapes and locations (inside the universe of discourse intervals) of the membership functions and the singletons, Fig.1 can learn a prescribed input-output mapping [6].

IV. CMOS PREMISE BUILDING BLOCKS

A. Membership Function Circuits (Layer 1)

CMOS PWL membership function circuits have been proposed in [8]. Herein we will consider *bell-like* functions with continuous derivative -- a feature which may render advan-

tages for learning purposes. The exact bell shape [6],

$$\mu(x) = \cfrac{1}{1 + (\cfrac{x-E}{\Delta})^{2B}} \qquad (6)$$

where Δ, E, and B are the adaptation parameters, uses involved analog circuitry [9]. However, membership functions with bell-like shape and continuous derivative are realized in a simple manner by the transconductance mode circuit of Fig.2(a), consisting of two interconnected source-coupled MOS differential pairs. This structure is similar to that used in high-speed folding ADCs [10] and exploits the operation of the MOS differential amplifier as a current switch with soft transition region -- depicted in Fig.2(b).

Fig.2(a) obtains a bell-like membership function through a linear, KCL combination of two soft-limiter characteristics: one with positive slope and the other with negative slope as Fig.2(c) illustrates. A square-law model of the MOS transistor [4], obtains the following expressions for the transition regions in Fig.2(c),

$$\mu(x) = \begin{cases} \sqrt{\beta I_u} x_1 \sqrt{1 - \cfrac{\beta x_1^2}{I_u}} & -F < x_1 < F \\ \\ (-\sqrt{\beta I_u})\, x_2 \sqrt{1 - \cfrac{\beta x_2^2}{I_u}} & -F < x_2 < F \end{cases} \qquad (7)$$

where $x_1 = x-E_1$, $x_2 =x-E_2$, $\beta=kW/L$, $F=(I_u/2\beta)^{1/2}$, k is a technological parameter (whose value for NMOS transistors in a typical technology is about 50 A/v ^{2}) and W and L are the transistor width and length. The unitary current I_u in (1) is a normalization value which corresponds to the largest matching degree value $(m_{ji}=1)$.

By making $E_1=E-\Delta$ and $E_2=E+\Delta$, (7) provides a first-order approximation to the bell-shape of (6) with the slope at the *crossover* [6] points (parametrized by B in (6)) given by:
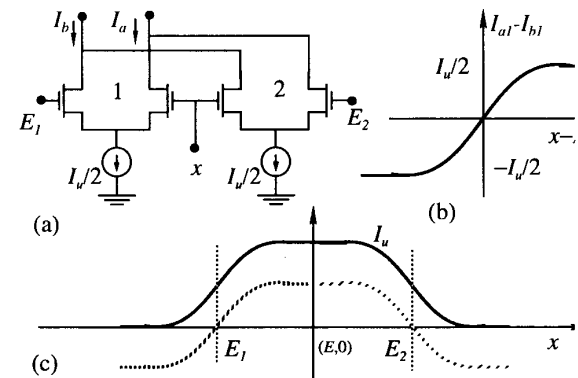
$$g_m = \sqrt{\beta I_u} \qquad (8)$$

Fig.2: Bell-like CMOS membership function: (a) Circuit structure; (b) D ferential pair transfer characteristics; (c) Membership function shape.

Thus, the membership function shape and position ·can be tuned by proper setting of the reference voltages E_1 and E_2 and the transistor widths and lengths.

B. T-Norm Nodes (Layer 2)

The calculation of the minimum among the matching degrees in fuzzy inference rules is functionally equivalent to obtaining the complement of the maximum among the complements of these matching degrees,

$$w = mim(s_1, s_2, \ldots, s_M) = \overline{max(\overline{s_1}, \overline{s_2}, \ldots, \overline{s_M})} \qquad (9)$$

where the upper bar denotes complement, calculated from the original variable as follows,

$$\bar{z} \equiv 1 - z \qquad (10)$$

Note that the complement operator is easily realized in current-mode, by using KCL, with 1 in (10) being the unitary current, I_u. Let us now consider the implementation of the maximum operator. The classical approach used in analog computation for voltage-mode circuits is based on the following steady-state equation,

$$-i_o + \sum_{k=1,M} A u_{-1}\left(i_{ik} - i_o\right) = 0 \qquad (11)$$

where i_o is the output, i_{ik} are inputs, and $u_{-1}(\bullet)$ denotes *rectification operator*. Fig.3(a) illustrates this concept while Fig.3(b) shows a conceptual CMOS current-mode schematic for it. This circuit is similar to that presented in [11] for the *winner-take-all* operation. However, contrary to the winner-take-all, the circuit of Fig.3(b) is designed not only to select the maximum among a set of input currents, but also to propagate that maximum current to the output node. Fig.3(c) illustrates circuit operation. The maximum current determines the value of the common gate voltage, V_g. The only input transistor that operates in saturation region is that which is driven by maximum
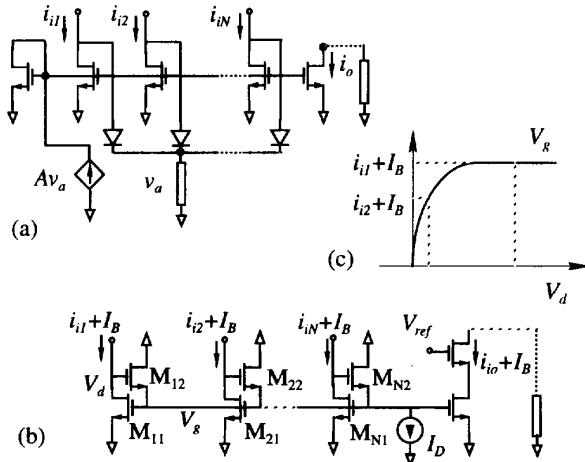
input current. All the rest operate in ohmic region.

Fig.3 requires careful analog design to reduce errors due to channel length modulation effects, which appear if transistor output nodes are not equipotential. In particular, our design approach yields 0.3% error for 15μA current.

C. Rule Antecedent.

Fig.2(a) obtains a bell-like membership function shape and its complementary shape, as Fig.4(a) illustrates. Thus, it can be directly connected to Fig.3(b) to calculate rule antecedents. Fig.4(b) shows the corresponding conceptual schematics. Actual circuit implementations use cascode transistors and proper biasing for increased accuracy, up to 99%. In particular, we follow the strategy to adjust transistors sizes to match voltages in case an average current flows through the transistors in the differential pairs.

An important parameter for this circuit is the input range, which limits the universe of discourse for input variables. The input range limits can be calculated as follows,

$$x \geq V_T + \sqrt{\frac{I_D L_{M_{i2}}}{k W_{M_{i2}}}} + \sqrt{\frac{(I_Q + I_B) L_{M_{i1}}}{k W_{M_{i1}}}}$$

$$x \leq V_A - V_T - \sqrt{\frac{I_Q}{2 k W_{Md_{ijk}}}} \qquad (12)$$

We can optimize the range with large transistor sizes and small current, but the most determinant factor is the voltage V_A. This voltage can be enlarged by using the cascode mirror of Fig 4(c), where square-law model calculations obtain:
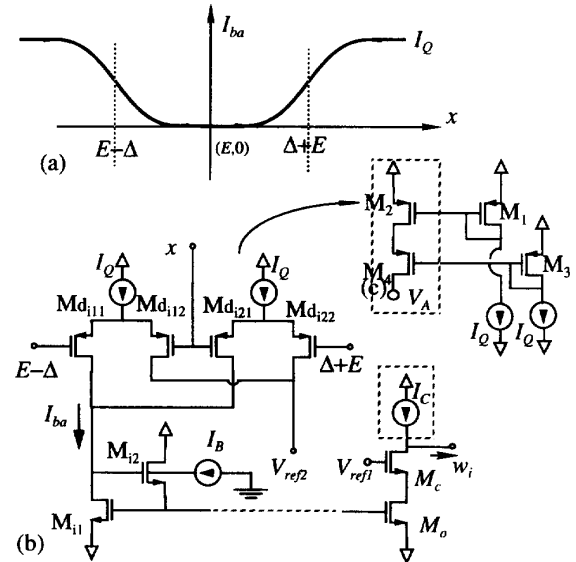

Fig.3: CMOS current-mode maximum/propagate circuit: (a) Concept; (b) Basic schematic; (c) Illustrating operation principle.

Fig.4: Rule antecedent circuit: (a) Membership function circuit output; Basic schematic; (c) I_Q current source implementation.

$$\delta = \sqrt{\frac{I_Q}{k}} \left(\sqrt{\frac{L_{M_3}}{W_{M_3}}} - \sqrt{\frac{L_{M_4}}{W_{M_4}}} \right) \qquad (13)$$

where $\delta = V_{dd} - V_A$; V_A can be enlarged by proper sizing of M_3 and M_4 transistors.

Thus the university of discourse can be made to cover 45% of the total excursion between supply voltage. Further extension achieves up to 100% of the excursion by using both p- and n-channel differential pairs in the membership function circuit.

V. CMOS CONSEQUENT BUILDING BLOCKS

A. Normalization Circuitry (Layer 3)

Using analog dividers to evaluate (3) is impractical -- analog dividers are costly and inaccurate. A convenient alternative uses feedback to maintain constant a sum of vector components [2], [3], [12]. Unfortunately, transient response of this normalization scheme is rather poor -- a negative consequence of feedback. In particular, it obtains times around 1μs (90%) when used in the CMOS 1.6μm 3-input 4-rule controller of Section VI.

On the other hand, the normalization circuit operation can be summarized as follows

$$w_i = F(w, \overline{w}) \qquad (14)$$

and

$$|\overline{w}| = \sum_{i=1,4} \overline{w}_i = A \qquad (15)$$

where $F(\bullet)$ is an increasing monotonic function of $\overline{w}_i$ and A is a real constant. Fig.5 (illustrated for a case with 4 inputs) presents a circuit which realizes this function without feedback, and hence yields much better transient response than previous proposals. The proposed circuit consists of two source coupled NMOS arrays; the one at the bottom implements a non-linear I/V conversion, and produces a voltage input for each output transistor of the top array, whose drain current is finally replicated by a PMOS current mirror. Square- law calculations on this circuit give
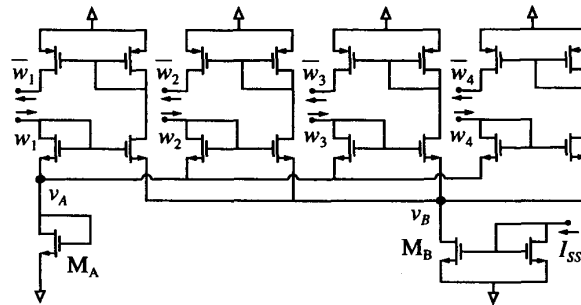

Fig.5: Normalization function: CMOS circuit for open-loop normalization.

$$\overline{w}_i = \beta_{top} \left[\frac{w_i}{\beta_{bot}} + (\frac{1}{4}) \sum_{j=1,4} \left(\sqrt{\frac{w_j}{\beta_{bot}}} - \sqrt{\frac{\overline{w}_j}{\beta_{top}}} \right) \right]^2 \qquad (16)$$

This expression fulfills (13), while (14) is forced by KCL.

Main error sources in Fig.5 are channel length modulation and common mode rejection. Proper design yields 0.8% error using our design approach.

B. Output Circuitry (Layers 4 and 5)

These are realized in single manner in current domain (addition is then realized by KCL) using current mirrors, and tuned in either digital or analog manner using state-of-the-art analog current-mode techniques [8], [9]. In particular, singleton weighting is easily obtained by means of current mirrors with different sized input and output transistors, where the ratio of these sizes gives the singleton value.

Analog programmability can be incorporated using techniques similar to that in [8]. Fig.6 illustrates the incorporation of digital programmability. In this figure, we compose the desired transistor size by combining transistors of different sizes, using NMOS transistors as voltage controlled switches to achieve external control of this combination by digital signals. The global output is

$$y = \sum_{i=1,4} (3s_{i2} + 2s_{i1} + s_{i0}) \overline{w}_i \qquad (17)$$

where $(3s_{i2}+2s_{i1}+s_{i0})$ is the singleton value for rule i, and s_{ij} take logical values 1 or 0 associated to V_{dd} and V_{ss} voltage, respectively.

VI. PRACTICAL RESULTS

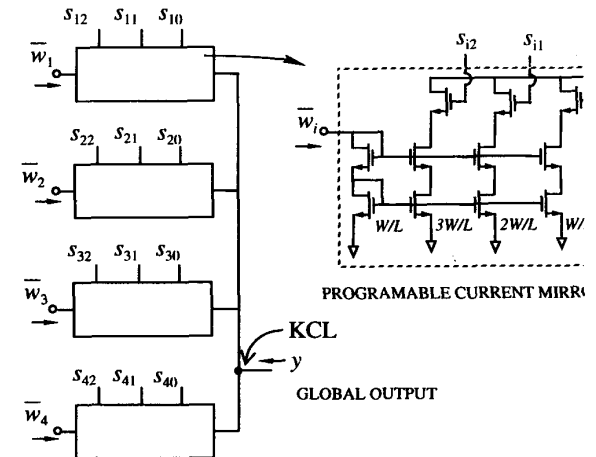We have designed a 3-input four-rule controller in a CMOS n-well single-poly double-metal 1.6μm technology.

Fig.6: Rule output weighting and aggregation to obtain for global outp

Fig.7 shows measurements from prototypes of the membership function circuit, which demonstrate tunability of the shape (Fig.7(a)) and the position (Fig.7(b)). Remaining results given in this section are HSPICE simulation results (for the netlist extracted from the physical layout) using level 6 transistor models.

The controller works with $V_{ss}=-2.5v$ and $V_{dd}=2.5v$, and bias currents of $I_Q(=I_s/2)=15\mu A$, $I_B=10\mu A$ (see Fig.4(b)), $I_D=0.5\mu A$ (see Fig.3(b)) and $I_C=35\mu A$ (see Fig.4(b)). The bias current of the normalization circuit, I_{ss}, may be variable (a correcting factor is then added to output) or constant when the number of rules changes, depending on dynamic requirements. The maximum value used was $62.5\mu A$ for a sixteen rule controller. The minimum transistor size is $W=10\mu m/L=5\mu m$, while maximum size is $W=200\mu m/L=5\mu m$ (in the differential pairs).

A. Rule Circuit

Input range or universe of discourse is determined by (12). We adjusted transistor sizes to obtain a CMR range around 2V. A simulation test was realized to measure it, as well as its errors, where membership function shape is the same for all input and is moved along universe of discourse. Fig.8 shows the simulation results. Significant errors in output indicate universe of discourse limits. We can measure an input range larger than 2V and deviations along universe of discourse of membership function values below 1%.

A second test was realized to determine the circuit dynamic response. We forced the output circuit to go from minimum to maximum value by exciting with a step signal for proper membership function location. Under these conditions, rising time (99% settling) was around 250ns and falling time (also 99%) was around 100ns.

B. Complete Controller

A first test was realized to prove the validity of the design approach. For a controller of one input and four rules, different singleton values were given to each rule. Fig.9 shows the controller output (top), as well as normalization circuit output (bottom).

Finally, a test was realized assuming all singletons equal to 1, so that global output is equal to I_{ss} (see Fig.5). We can then measure deviations in this theoretical output as input change, as well as transient responses for step input signals. This test was realized for controllers with different numbers of input and rules. Nominal errors below 1% were measured, and Figs. 10 and 11 summarize transient response results. T_1 denotes the transitory associated to a rising rule output, while T_2 denotes the transitory associated to a falling rule output.

VI. CONCLUSIONS

A set of innovative building blocks for analog fuzzy controllers design has been presented. The simplicity of these blocks, as well as the compact design they achieve (less than 3 mm^2 for a three input-four rules controller), permit operation speeds of up to 5MFlips with low power consumption. Besides this, the great modularity of this approach enables increased controller complexity, by adding rules and/or input, with no extra design effort. Programmability is easy to add to this approach, as shown in Section IV for singletons. A similar method can be used to program membership function slopes, by spliting transistors of differential pairs, while their locations are tunable by means of gate voltages in transistors of membership function circuits. Finally, the neural architecture of the controller, as well as the continuously differentiable functions that the membership function circuit gives, will permit introducing further learning capabilities.

REFERENCES

[1] K. Namakura et al.: "Fuzzy Inference and Fuzzy Inference Processor". IEEE Micro, pp. 37-48, Oct. 1993.
[2] T. Miki et al.: "Silicon Implementation for a Novel High-Speed Fuzzy Inference Engine: Mflips Analog Fuzzy Processor". Journal of Intelligent and Fuzzy Systems, Vol. I, pp. 27-42, 1993.
[3] T. Yamakawa: "A Fuzzy Inference Engine in Nonlinear Analog Mode and Its Application to a Fuzzy Logic Control". IEEE Trans. on Neural Networks, Vol. 4, pp. 496-522, May 1993.
[4] E. Vittoz: "The Design of High-Performance Analog Circuits on Digital CMOS Chips". IEEE Journal of Solid-State Circuits, Vol. 20, pp. 657-665, Jun. 1985.
[5] M.J.M. Pelgrom et al.: "Matching Properties of MOS Transistors". IEEE Journal of Solid-State Circuits, Vol. 39, pp. 1433-1440, June 1990.
[6] J.S.R. Jang and C.T. Sun: "ANFIS: Adaptive-Network-Based Fuzzy Inference System". IEEE Trans. on Systems, Man and Cybernetics, Vol. 23, pp. 665-685, May 1992.
[7] T. Takagi and Sugeno: "Derivation of Fuzzy Control Rules from Human Operator's Control Action". Proc. of the IFAC Symp. on Fuzzy Information, Knowledge Representation and Decision Analysis, pp. 55-60, July 1989.
[8] A. Rodríguez-Vázquez and M. Delgado-Restituto: "Generation of Chaotic Signals using Current-Mode Techniques". Journal of Intelligent and Fuzzy Systems.Vol.2, January 1994 (to appear).
[9] C. Toumazou et al. (editors): "Analog IC Design: The Current Mode Approach". Peter Peregrinus 1990.
[10] J. Van Valburg and R.J. Van de Plassche: "An 8-b 650Mhz Folding ADC". IEEE Journal of Solid-State Circuits, Vol.39, Dec. 1992.
[11] J. Lazzaro, R. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all Networks of O(n) Complexity", Advances in Neural Information Processing Systems, Vol. 1, D. S. Touretzky, Ed. Los Altos, CA: Morgan Kaufmann, 1989.
[12] M. Sasaki et al.: "Current-Mode Analog Fuzzy Hardware with Voltage Input Interface and Normalization Locked Loop". IEICE Trans. Fundamentals, Vol. E75-A, pp. 650-654, June 1992
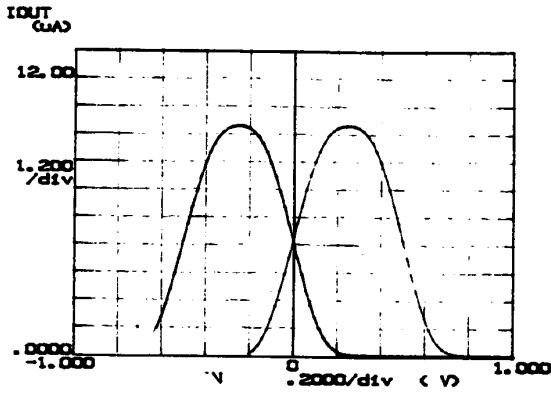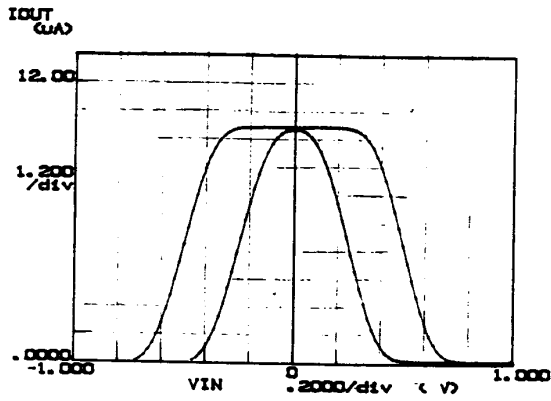
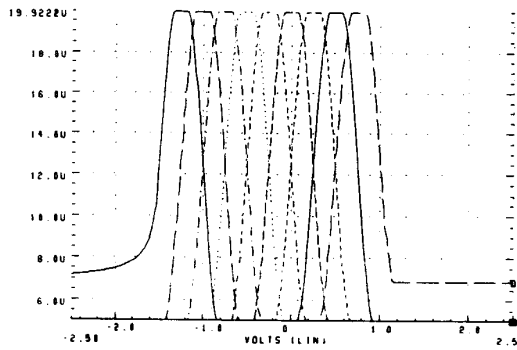Fig.7: Illustrating membership function tunability for a 1.6µm CMOS prototype: (a) Shape; (b) Position.
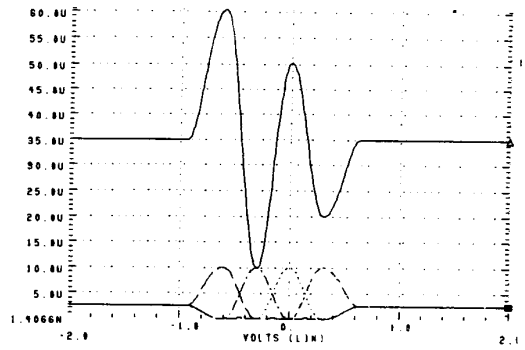

Fig.8: Rule circuit output

Fig.9: Global DC controller output (top) and normalized rule circuit outp (bottom)
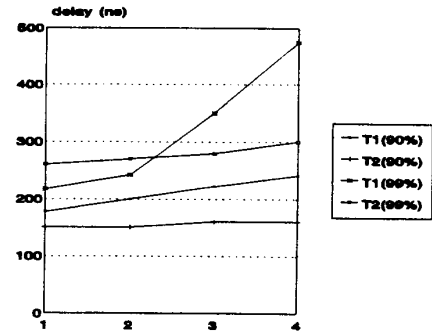

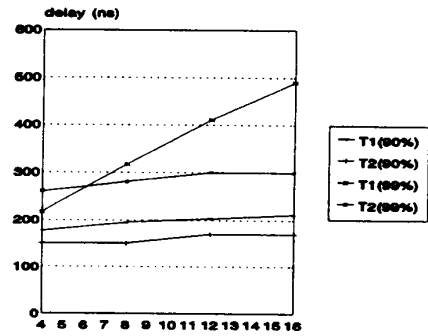Fig.10: Controller transient response vs number of connected inputs (fc rules) (I_{SS}=22.5µA)

Fig.11: Controller transient response vs. connected rules (one input). I_{SS} (4rules)=22.5µA; I_{SS} (8rules)=35µA; I_{SS} (12rules)=47.5µA; I_{SS} (16rules)=62.5µA

123