A VLSI Design for Scalable High-Speed Digital Winner-Take-All Circuit

Myungchul Yoon

Abstract—A high speed VLSI digital Winner-Take-All (WTA) circuit called simultaneous digital WTA (SDWTA) circuit is presented in this paper. A minimized comparison-cell (w-cell) is developed to reduce the size and to achieve high-speed. The w-cell which is suitable for VLSI implementation consists of only four transistors. With a minimized comparison-cell structure SDWTA can compare thousands of data simultaneously. SDWTA is scalable with $O(m\log n)$ time-complexity for n of m-bit data. According to simulations, it takes 16.5 ns with 1.2V-0.13 μm process technology in finding a winner among 1024 of 16-bit data.

Index Terms—Winner-take-all circuit, digital WTA circuit, maximum selector circuit, scalable WTA architecture

I. INTRODUCTION

The Winner-Take-All (WTA) circuit is the circuit to identify the biggest one among multiple data. It is one of the most important building blocks in various areas such as neural networks, fuzzy systems and nonlinear filters. Many VLSI chips of these areas including neural network chips are implemented by analog circuits including analog WTA [1-3].

Many analog WTA circuits have been proposed [4-8], but digital WTA circuits have been relatively less researched than analog counterparts. Analog VLSI implementations of WTA circuit require less hardware than digital implementations. However, analog WTA circuits suffer from matching problems [9] and stability and convergence problems [10], especially for a large number of inputs. With the development of VLSI technologies, the number of inputs increases while the voltage level of V_{DD} as well as the input range decrease. Therefore, it is getting more difficult for analog WTA circuits to achieve a high-precision operation.

On the contrary, digital WTA circuits have many advantages compared to their analog counterparts. Unlike analog counterparts, digital WTA circuits are free from mismatch, stability, and convergence problems. Since the precision of digital circuits is determined only by the number of digits in digitization process, it is relatively easy to control the precision of WTA circuit. With the increase of integration level, the advantages of digital WTA circuits overweight the disadvantage so that switching analog WTA to digital WTA is necessary for more accurate operations for a large number of data. Digitization is the trends of times. Many studies have been carried out to digitize analog parts [11-14].

A general purpose high-speed digital WTA circuit which is scalable to use for a large number of inputs is presented in this paper. The proposed architecture, called Simultaneous Digital WTA (SDWTA) intends to minimize the hardware-overheads of digital WTA circuits and to achieve high speed operation. A small competition cell is devised for these purposes. To be suitable for VLSI implementation, the cell consists of only four transistors, so that SDWTA not only operates in high speed but also reduces hardware overheads.

The algorithm and circuits of SDWTA are described in the next section. The simulation results are presented in

```
//   Digital WTA algorithm for n of m-bit digital inputs
//
//   B_ij : 1 ≤ i ≤ n, 1≤ j ≤m    the j-th bit of i-th input
//   B_i1 : MSB,        B_im : LSB
//   W_ij: winner-state of the i-th input at the j-th bit
//   L_ij : loser-state of the i-th input at the j-th bit

1.        For all i,   L_{i0} = 0,   W_{i0} =1 ;
2.        From   j = 1   to   j = m    do {
3.             P̄_j = Σ_{i=1}^{n} W_{i,j-1} B_ij   ;
4.             For all i,   W_ij = W̄_{i,j-1} (B_ij + P_j) ;
5.                          L_ij = W̄_ij   ;
6.        }
7.     Return   W_im ;
```

Fig. 1. The algorithm for SDWTA.

Section III, and the concluding remarks are in Section IV.

II. SIMULTANEOUS DIGITAL WTA CIRCUIT

1. Algorithm for the SDWTA

The Simultaneous Digital WTA (SDWTA) circuit is based on the parallel bit-selection algorithm [15]. The algorithm used for SDWTA is given in Fig. 1. Let us consider n data, $B_1,...,B_n$ which are composed of m-bit binary numbers. Let us use the notation B_{ij} for the j-th bit of the i-th data. Let B_{i1} be the most significant bit (MSB), and B_{im} be the least significant bit (LSB) of B_i.

The algorithm consists of m-steps. In each step, the data are divided into two groups: a winner group and a loser group. At the j-th step, the j-th bits of all data are checked to distinguish winners from losers.

Let W_{ij} (L_{ij}) is the winner (loser) state of the data B_i at the j-th bit. $W_{ij} =1$ means that B_i is left in the winner group until the j-th bit is checked. $L_{ij} =1$ ($W_{ij} =0$) means B_i is in the loser group. At the beginning, all data are included in winner groups so that $W_{i0}=1$ ($L_{i0}=0$) for all of i.

Comparison of data proceeds from MSB to LSB data. At the j-th step, the j-th bits of all data is checked. If $W_{i,j-1}=1$ and $B_{ij}=1$, W_{ij} results in 1. If $W_{i,j-1}=1$ but $B_{ij}=0$, W_{ij} becomes 0 unless the passing signal (P_j) is generated. P_j is generated when $W_{i,j-1} \cdot B_{ij}=0$ for all i. When $P_j=1$, the states of $W_{i,j-1}$ are passed to W_{ij}.

After LSBs are checked, the data left in the winner group is the winner of the competition. If multiple data
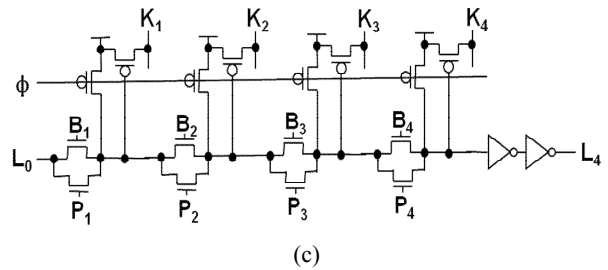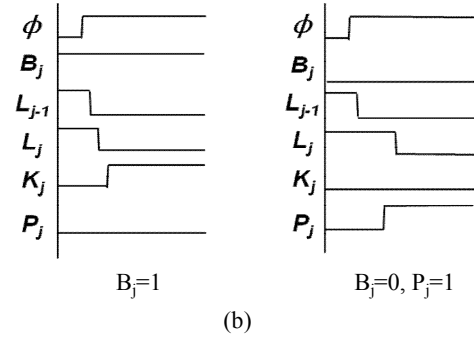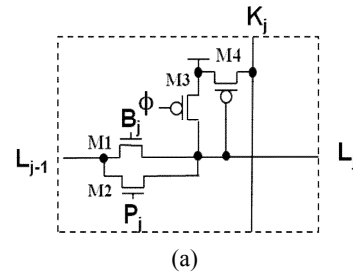

Fig. 2. The basic elements of SDWTA architecture (a) the circuit of a w-cell, (b) signal waveform for the w-cell, (c) the circuit of a 4-WTA cell block.

exist in the winner group, one of them is chosen according to a selecting rule given by its application.

2. Circuit Description

The SDWTA circuit consists of cell array and control circuits. Fig. 2(a) shows the proposed cell structure called w-cell. The w-cell consists of four MOS transistors so that the size is very small. The signals related to the cell operation play the roles such that

ϕ : global start/precharge   signal

L_j : loser state at the j-th bit. If $L_j =0$, it is in the winner group

B_j : the j-th bit of a datum B

P_j : signal to pass L_{j-1} state to L_j
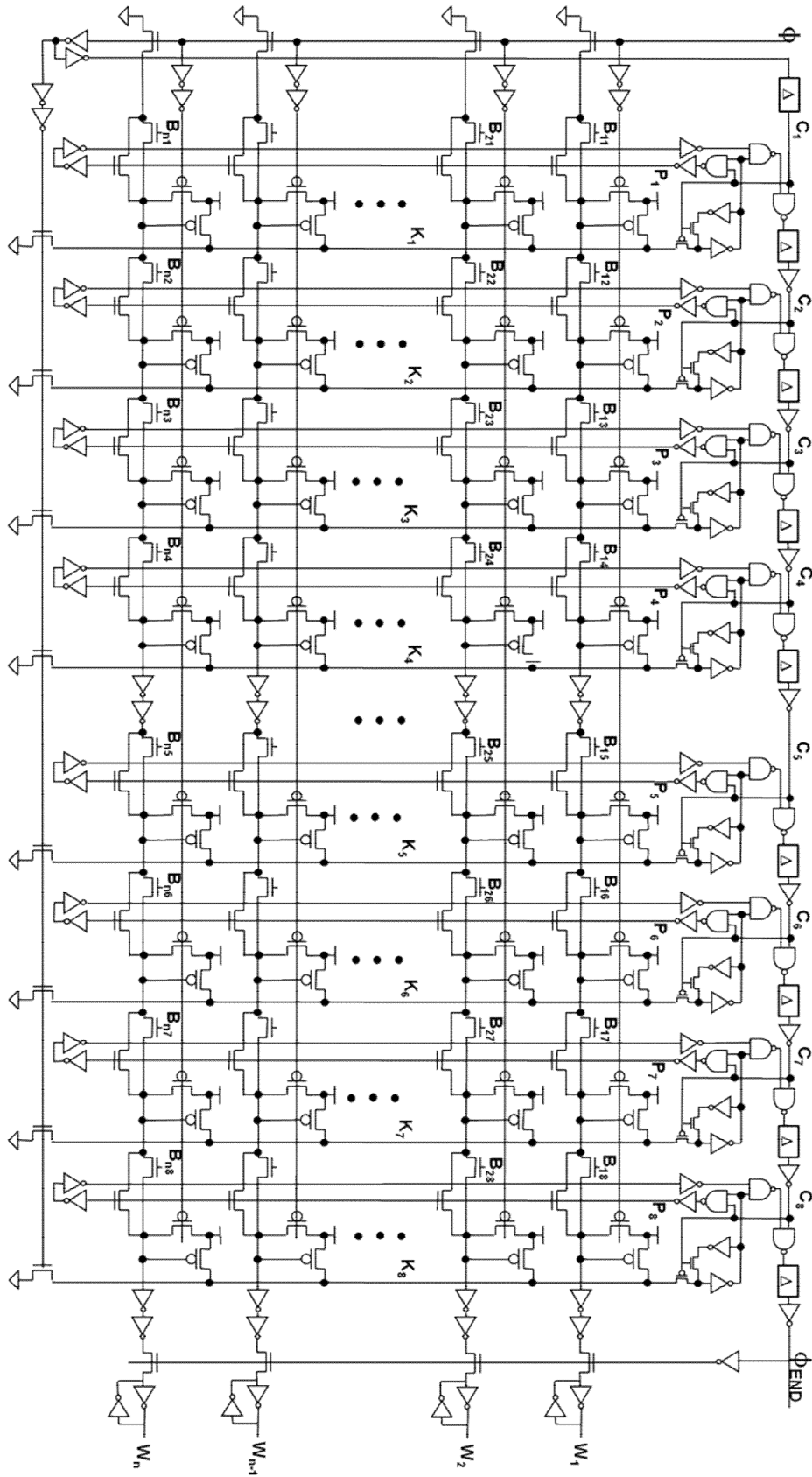
K_j : temporal signal used to determine P_j

Fig. 3. The SDWTA configuration for small-scale 8-bit data elements.

The two NMOS transistors M1 and M2 are used to perform the line 4 in the Fig. 1. Because NMOS transfers 0 better than 1, loser (L_j) states rather than winner (W_j) states are used as the input of the w-cell, i.e., 0 is passed to the next step only for winners. The M3 is used to precharge all L_j states to 1, before the circuit starts the WTA operation. Therefore, all data are set to losers before the circuit starts, and L_j's for the winners are become 0 during the WTA operation. M4 is used to pull up K_j-line which is attached to the j-th bit w-cells of all data. The K_j-line is reset to 0 initially, but pulled up to 1 when any one cell attached to the line becomes a winner (L_j=0). If K_j-line becomes 1, it kills the generation of P_j signal. If no cell becomes a winner, K_j-line remains at 0 and P_j signal is generated. This procedure corresponds to the line 3 in Fig. 1. When the j-th bits of all winner data are 0, the winner states do not changed. For this case, P_j signal is generated to pass the L_{j-1}-state to L_j. The signal waveforms for the cell operation are shown in Fig. 2(b).

Although the w-cell is the basic component of array, a W-block which is a chain of multiple w-cells is more useful as a building block of cell array. For an m-bit data, m of w-cells are connected serially. Because a long series of MOS transistors degrades the strength of signal, some voltage buffers are required in the middle of the chain.

A W-block as shown in Fig. 2(c) is composed of a series connected multiple w-cells and a voltage buffer. The number of w-cells in a W-block is designer's choice. In this paper the W-block with four w-cells is used as the basic W-block in building the cell array.

The overall SDWTA circuit diagram is presented in Fig. 3. For simplicity, it is drawn for the n of 8-bit data. It shows the arrangement of cell blocks and control circuits. The SDWTA is activated by ϕ signal. Before activation, i.e., when ϕ is 0, all L_{ij} are precharged to 1 and all P_j are reset to 0. If activated by ϕ, all precharge paths are closed and L_{i0} for all i become 0, so that the leftmost column starts competition. The circuits operate column by column and every column operation is identical. The control circuits located at the top in Fig. 3 control the progress of column operations from MSB to LSB.

The progress of column is controlled by the column control signal (C_j). In the j-th column operation, the winner cells pull-up K_j-line. The state of K_j-line is latched by C_j signal so that the C_j must be activated after
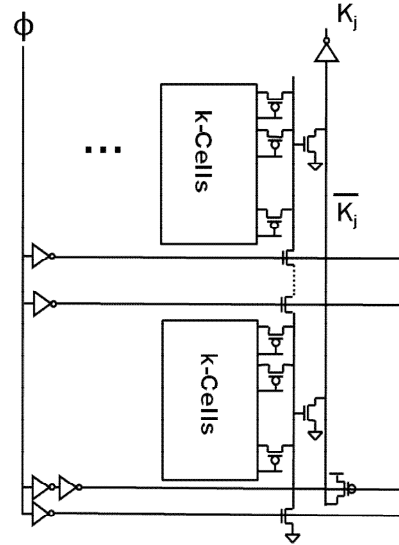

Fig. 4. The enhanced column structure for large-scale SDWTA circuit.

the state of K_j-line is clearly developed. For this purpose, the Δ-delay circuit is used to postpone the C_j-activation. ϕ_{END} signal plays a role to latch the result of competition as well as to inform the end of competition.

3. Enhancement for Scalability

If t_c is the worst case delay for the evaluation of one column, the speed of WTA circuit (T_d) for m-bit data is roughly becomes

$$T_d = mt_c$$

The most critical signal affecting t_c is K_j. Because all cells in a column are attached to a K_j-line, the load is very large. The worst case delay happens when the line is pulled-up by only one w-cell. For the high speed operations, the circuit in Fig. 3 is suitable for small-scale WTA that compares up to tens of data elements. For more than hundreds of data, the delay for pulling-up the K_j-line rapidly increases so that the speed of SDWTA is greatly degraded. For large–scale WTA operation, the column structure needs to be modified to reduce the delay. Fig. 4 shows the enhanced column structure for mid-scale and large-scale SDWTA circuits. To reduce the delay in pulling-up the K_j signal, the large load is distributed to two levels. The cells in a column are grouped as k-cell blocks so that the long $\overline{K_j}$ line is

Table 1. Comparison to some existing WTA implementations

Parameter	SDWTA	WTA after					
		[17]	[13]	[18]	[3]	[19]	[20]
design type	digital	digital	hybrid	hybrid	analog	analog	analog
input	voltage	voltage	voltage	voltage	voltage	current	current
technology	0.13 μm	0.6 μm	0.35 μm	0.6 μm	0.09 μm	0.35 μm	0.35 μm
supply voltage (V)	1.2	4.0	3.3	3~5	1.8	3.3	3.3
# of inputs	1024	128	16	8	2	1024	2
# of bits	16	6	5	4	-	-	-
resolution	1/64k	1/128	1/64	1/16	1/3600 (0.5 mV)	1/1024 (1.22 nA)	1/33 (1.8 nA)
worst case delay	16.5 ns	13 ns	135 ns	60 ns	30 ns	1 μs	34 ns
average delay/bit	1.03 ns	2.67 ns	27 ns	15 ns	-	-	-

pulled-down by a large NMOS when any one cell in a group becomes a winner. By this means, the size of pull-up PMOS (M4) in w-cell can be kept small while the speed could be increased by adjusting the size of NMOS pull-down transistor. As we can see in the next section, SDWTA with the enhanced column structure can extend the high speed operation to more than thousands of data.

Similarly to Fig. 4, the enhanced column structure could be extended to hierarchical pull-up and pull-down structures. For more than ten thousands data, the capacitive load of K_j-line could be distributed to three or four levels. If the load is well distributed to each level, the delay for developing K_j-line is proportional to log n. Since the delay of K_j-signal is the critical factor in the delay of a column operation (t_c), the total delay of SDWTA circuit (T_d) has $O(m \log n)$ complexity.

III. EXPERIMENTS

The speed and scalability of the SDWTA circuit are estimated by SPICE simulation. The simulation is performed by HSPICE with IBM's "1.2V-0.13 μm 8RF-LM" model parameters [16]. The minimum size transistors (W=0.13 μm) are used for w-cell except the pull-up PMOS (M4) of which the size is W=0.26 μm.

To check the scalability, the delays of the SDWTA circuit for various numbers of 16-bit data are simulated for three structures; S1, S2, and S3. S1 is for the small scale WTA which uses the circuit as in Fig. 3. S2 and S3 are for the mid-scale and the large-scale WTA which adopt the enhanced column structure in Fig. 4. S2 uses eight w-cell groups (k=8) while S3 uses sixteen w-cell groups (k=16).
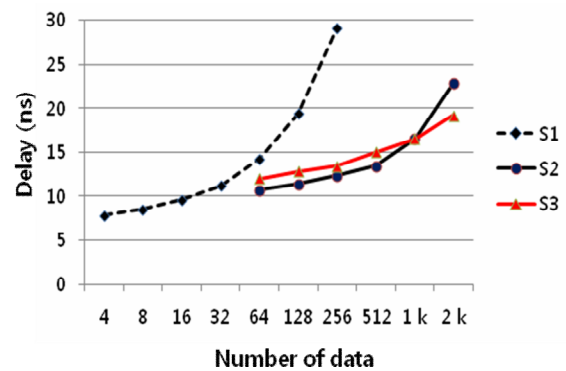

Fig. 5. Simulation results for the speed of SDWTA.

Fig. 5 shows the simulation results. It shows that S1 is suitable up to WTA of 32 data. The delay of S1 for 4 of 16-bit data is 7.9ns and it increases to 11.2ns for 32 data. For more than 32 data, the delay increases rapidly along with the number of data. For more than 50 data, the simulation results show that adopting the enhanced column structure is necessary to increase the speed of the WTA circuit. S2 is faster than S3 up to 1024 data, but S3 becomes faster than S2 for the larger number of data. The optimum number k for grouping depends on the number of data and the size of pull-down NMOS. Fig. 5 shows that the SDWTA architecture with the enhanced column structure can find the winner among 2k of 16-bit data in 20 ns with 0.13 μm technology. It implies that the SDWTA is a scalable architecture which can operate at high speed from tens to thousands of inputs.

In Table 1, the speed of the SDWTA is compared to those of several existing WTA implementations. The SDWTA is about 100 times faster than analog WTA implementations in [3, 19, 20]. Most of analog WTAs focus on the precision of the WTA circuits, because input

ranges decrease while the number of inputs increases. The precision of digital circuit is determined by bit-length so that the design of a digital WTA can devote more efforts to speed-enhancement. Some WTA implementations [13, 18] adopt hybrid architecture which use DAC to convert digital inputs to analog inputs for WTA circuit. The SDWTA does not need such a conversion, but also operates about 30 times faster than those implementations. The digital WTA in [17] uses the two-dimensional bit-propagating (2DBP) scheme. From Fig. 5, the delay of S2 with 128 inputs is 11.36ns which is equivalent to 4.26 ns for 6-bit inputs. Compared to the delay of 2DBP in Table 1, the SDWTA is about 3 times faster than the 2DBP scheme. The results of comparisons show that SDWTA is an architecture competitive in speed and superior in scalability to other implementations.

IV. CONCLUSIONS

A new architecture for the digital WTA circuit called SDWTA is presented in this paper. SDWTA is a high-speed scalable architecture that compares thousands of data at the same time. The basic cell structure and block structures are devised for the SDWTA architecture.

The speed of SDWTA is estimated by SPICE simulation for 0.13 μm process technology. According to the simulation, the SDWTA takes 7.9 ns in finding the winner among 4 of 16-bit data, and 16.5 ns among 1024 data. It is proven by simulations that the SDWTA architecture is scalable to hundreds of thousands data.

REFERENCES

[1] M. A. G. Andreou, K. A. Boahen, A. Pavasovic, P. O. Pouliquen, R. E. Jenkins, and K. Strohbehn, "Current-mode subthreshold MOS circuits for analog VLSI neural systems", *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp.205-213 1991.

[2] R. Kalim and D. M. Wilson, "Semi-parallel rank-order filtering in analog VLSI", *Proc. IEEE ISCAS'99*, vol. 2, pp.232 -235 1999.

[3] M. Rahman, K. L. Baishnab, and F. A. Talukdar, "A high speed and high resolution VLSI Winner-take-all circuit for neural networks and fuzzy systems" *IEEE ISSCC2009,* pp. 1-4, 2009.

[4] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, D. S. Touretzky, *Winner-Take- All Networks of O(n) Complexity*, vol. 1, pp.703 -711 1989 :Morgan Kaufmann.

[5] J. A. Startzyk and X. Fang, "CMOS current-mode winner-take-all circuit with both excitatory and inhibitory feedback", *Electron. Lett.*, vol. 29, no. 10, pp.908-910 1993.

[6] S. P. DeWeerth and T. G. Morris, "CMOS current-mode winner-take-all circuit with distributed hysteresis", *Electron. Lett.*, vol. 31, no. 13, pp.1051-1053 1995.

[7] G. Indiveri, "A current-mode hysteretic winner-take-all network, with excitatory and inhibitory coupling", *Analog Integr. Circuits Signal Process.*, vol. 28, pp.279 -291 2001.

[8] D. Moro-Frias, M. T. Sanz-Pascual, and C. A. de La Cruz-Blas, "A novel current-mode Winner-Take-All topology," *Circuit Theory and Design (ECCTD), 2011 20th European Conference on* , vol., no., pp.134,137, 29-31 Aug. 2011.

[9] N. Kumar, P.O.Pouliquen, and A. G. Andreou, "Device mismatch limitations on the performance of a Hamming distance classifier" in *Proc. 1993 IEEE Int. Workshop on Defect and Fault Tolerance in VLSI Systems*, pp. 327-334 1993.

[10] P. V. Tymoshchuk, and M. P. Tymoshchuk, "Stability and convergence analysis of model state variable trajectories of analogue KWTA neural circuit," *Direct and Inverse Problems of Electro-magnetic and Acoustic Wave Theory (DIPED), 2011 XVth International Seminar/Workshop on* , vol., no., pp.26,35, 26-29 Sept. 2011.

[11] K. Uchimura, Fish, A. Milrud, V. Yadid-Pecht, O., "A high- speed digital neural network chip with low-power chain-reaction architecture," *IEEE J. Solid State Circuits,,* vol. 27, no.12, pp. 1862-1867, 1992.

[12] A. Schmid, Y. Leblebici, and D. Mlynek, "Mixed analogue-digital artificial-neural-network architecture with on-chip learning", *IEE Proc. Circuits Devices syst.,* vol. 146, no. 6, pp.345-349 1999.

[13] M. A. Abedin, Y. Tanaka, A. Ahmadi, T. Koide, and H. J. Mattausch, "Fully Parallel Associative Memory Architecture with Mixed Digital-Analog Match Circuit for Nearest Euclidean Distance Search" *IEEE APCCAS2006,* pp. 1309-1312, 2006.

[14] J. Kim, K. Hwang, and W. Sung, "X1000 real-time phoneme recognition VLSI using fee-forward deep neural networks," *IEEE ICASS7,* pp. 7510-7514, 2014.

[15] A. Kapralski, "The maximum and minimum selector SELRAM and its application for developing fast sorting machines," *Computers, IEEE Transactions on*, vol. 38, no. 11, pp. 1572-1577, Nov 1989.

[16] The MOSIS Service, http://www.mosis.com/ [Online]

[17] M. Ogawa, K. Ito, and T. Shibata, "A general-purpose vector-quantization processor employing two-dimensional bit-propagating winner-take-all" *IEEE Sym. VLSI Circuits Digest of Tech. Papers,* vol. 35, no.11, pp. 244-247, 2002.

[18] C. K. Kwon, and K. Lee, "Highly parallel and energy-efficient exhaustive minimum distance search engine using hybrid digital/analog circuit techiques," *VLSI Systems, IEEE Transactions on*, vol. 9, no. 5*,* pp. 726-729, 2001.

[19] J. Kim, K. Hwang, and W. Sung, "32x32 winner-take-all matrix with single winner selection," *Electronics Letters*, vol. 46, no. 5*,* pp. 333-335, 2010.

[20] A. Fish, A. Milrud, V. Yadid-Pecht, O., "High-speed and high-precision current winner-take- all circuit," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 52, no. 3, pp. 131-135, March 2005.

Myungchul Yoon received the BS and MS degrees in electronics engineering from Seoul National University, Korea, in 1986 and in 1988 respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of Texas at Austin in 1998. From 1988 to 2002, he was with Hynix Inc. Icheon, Korea as technical research staff at Semiconductor R&D Lab. and Mobile Communication R&D Lab. From 2005 to 2006, he was with DGIST, Korea as technical staff at the Information Technology R&D Division. Since 2006, he has been with the Department of Electronics Engineering, Dankook University, Cheonan, Korea, where he is a professor. His research interests are in low-power VLSI design, embedded systems, mobile communication, and wireless personal area networks.