Pulse Modulation VLSI Implementation of Clustering Algorithm Based on Stochastic Association Model

Takashi Morie*, Tomohiro Matsuura†‡, and Atsushi Iwata†
*Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology
2-4, Hibikino, Wakamatsu-ku, Kitakyushu 808–0196 JAPAN
E-mail: morie@brain.kyutech.ac.jp
Web: http://www.brain.kyutech.ac.jp/~morie
†Graduate School of Advanced Sciences of Matter, Hiroshima University
1-3-1, Kagamiyama, Higashi-Hiroshima 739–8526 Japan.
‡Now with Toshiba Co.Ltd., Kawasaki, Japan.

***Abstract*— The clustering algorithm employing "stochastic association", which we have already proposed, offers a simple and efficient soft-max adaptation rule. The adaptation process is the same as the on-line K-means clustering method except for adding random fluctuation in the distortion error evaluation process. This paper describes VLSI implementation of this new clustering algorithm based on a pulse modulation circuit architecture. Random fluctuation is added by a chaos generator circuit that we previously designed using a pulse-width/pulse-phase modulation approach. Experimental results using a fabricated chip have demonstrated that successful "stochastic association" clustering is achieved on a time scale of milliseconds.**

I. Introduction

Vector quantization (VQ) is a useful technique for data classification and compression. In VQ techniques, data vectors distributed in an M-dimensional space are encoded by using only a finite set of reference vectors $\boldsymbol{w} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_N)$ that minimize the average distortion error between the data and the reference vectors. A data vector $\boldsymbol{v}$ is represented by the best-matching or "winning" reference vector $\boldsymbol{w}_c$. Here, we treat on-line training, in which a stochastic sequence of incoming sample data points drives the adaptation procedure.

The straightforward approach to obtain such reference vectors is the well-known *on-line K-means clustering* algorithm, in which only the nearest reference vector to the sample vector is adjusted;

$$\Delta \boldsymbol{w}_i = \varepsilon \cdot \delta_{ic} \cdot (\boldsymbol{v}(t) - \boldsymbol{w}_i), \tag{1}$$

where, ε is the step size and δ_{ij} is the Kronecker delta. However, this simple clustering algorithm is often stuck in a local minimum. To avoid this difficulty, various "soft-max" adaptation rule that not only adjusts the "winning" reference vector but affects other reference vectors depending on their proximity to $\boldsymbol{v}$ [1-4].

We have already proposed a new efficient soft-max adaptation algorithm employing the *stochastic association (SA) model* and have demonstrated from simulation results that the algorithm is most efficient of the previously proposed algorithms [5]. We have also designed a CMOS VLSI circuit implementing the SA model for digital patterns, and demonstrated digital SA operation using a fabricated VLSI chip [6].

In this paper, we propose a new CMOS circuit using the pulse-width modulation approach for SA model and for clustering of analog patterns. The circuit described in this paper achieves on-chip training of reference vectors for clustering.

In Sec. II, the stochastic association model and the clustering algorithm using it are briefly reviewed. In Sec. III, a pulse-modulation circuit for the SA clustering is proposed, and the results of VLSI design are shown. Section IV describes experimental results using a fabricated chip, and the performance of the chip is demonstrated.

II. Stochastic Association Model and Clustering

In our stochastic association (SA) model, the association probability depends on the similarity between the input and the reference vectors. The SA algorithm extracts not only the reference vector most similar to the input but also other similar reference vectors with the probability depending on the similarity.

Figure 1 shows an architecture for clustering processing using the SA model. Stochastic fluctuation is added in the evaluation process of distortion error D_i between data vector $\boldsymbol{v}$ and reference vector $\boldsymbol{w}_i$.

The distortion error D_i can be the squared Euclidean distance $\|\boldsymbol{v} - \boldsymbol{w}_i\|^2$ or the Manhattan distance $\|\boldsymbol{v} - \boldsymbol{w}_i\|$. The evaluation result is represented by

$$R_i = D_i + \xi, \tag{2}$$

where ξ is random fluctuation. The winner-take-all circuit deterministically extracts the winning reference vector $\boldsymbol{w}_c$ by searching minimum R_i.

In the clustering algorithm using the SA model, which we call SA clustering, the winning reference vector is updated as expressed by eq. (1). Thus, as in the K-means algorithm, only one reference vector is adjusted for each adaptation step,

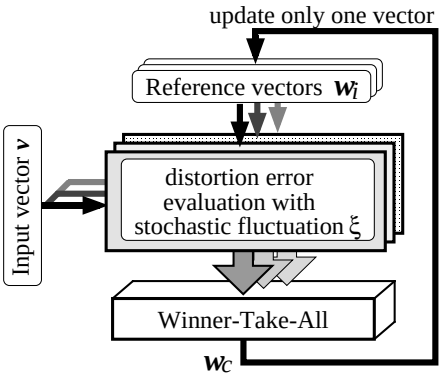


Fig. 1. Architecture for clustering processing using the SA model.

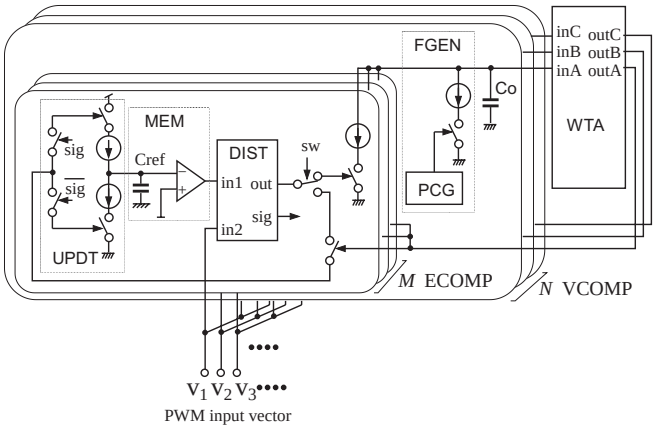


Fig. 2. Circuit for SA clustering.

but unlike the K-means algorithm, the adjusted vector is not always the most similar reference vector, and sometimes other similar vectors are adjusted.

III. Pulse-Modulation Circuit for SA Clustering

The proposed circuit for SA clustering is shown in Fig. 2. This circuit consists of a winner-take-all (WTA) circuit and N vector-comparators (VCOMPs), each of which includes M element-comparators (ECOMPs) and a random fluctuation generator (FGEN). The ECOMP has a distance calculation circuit (DIST), a reference vector element memory (MEM) and a memory updating circuit (UPDT). Each element of input vectors is represented by a pulse-width modulation (PWM) signal and the PWM signals are fed into the corresponding ECOMPs. The FGEN consists of a PWM chaos generator (PCG), which outputs PWM signals with chaotic pulse widths, and a switched current source driven by the PWM pulses.

The operation is as follows: In the jth ECOMP of the ith VCOMP, PWM signals representing the vector elements v_j and w_{ij} are fed into the DIST, and the DIST outputs a PWM signal representing the absolute difference $|v_j - w_{ij}|$ by using exclusive-OR logic operation and the sign bit of the difference. This PWM signal switches a current source, and a charge proportional to the pulse width of the PWM signal is extracted from capacitor C_0. All charges from all of the M ECOMPs are extracted from capacitor C_0, and the total charges extracted are proportional to the Manhattan distance between the input vector and the ith reference vector: $D_i = \sum_{j=1}^{M} |v_j - w_{ij}|$. In addition to D_i, a charge representing random fluctuation ξ_i is also extracted from capacitor C_0 from FGEN.

If the terminal voltage of capacitor C_0 is initialized at supply voltage V_{DD}, the final terminal voltage is given by

$$V_i = V_{DD} - \sum_{j=1}^{M} |v_j - w_{ij}| - \xi_i. \qquad (3)$$

From this equation, for a reference vector with larger similarity to the input vector has a probability with higher V_i. The WTA circuit searches the largest V_i, and the winning vector is determined. Thus, stochastic associative processing is achieved. Furthermore, by updating the winning reference vector using the UPDT circuit, SA clustering is achieved.

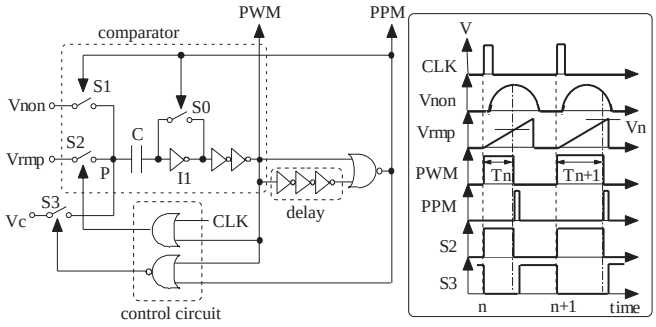


Fig. 3. PWM chaos generator circuit and operation sequence.

The PCG circuit can generate arbitrary synchronous chaotic pulse-width/pulse-phase modulation (PWM/PPM) signals by supplying an appropriate nonlinear voltage waveform $V_{non}(t)$ as shown in Fig. 3 [6, 7]. The circuit operation is based on the idea in which an arbitrary nonlinear input-output relationship is obtained by sampling the corresponding nonlinear voltage waveform using a PWM/PPM signal [8]. The circuit operation proceeds as follows. Here, time step n is defined by clock signal CLK, and V_n indicates a voltage at each time step. It is assumed that capacitor C holds the voltage difference between voltage V_n and the threshold voltage of the inverter I_1. At time step n, switch S_2 is turned on by CLK, and the linearly ramped voltage V_{rmp} is supplied at node P. When the voltage of node P reaches V_n, inverter I_1 inverts and PWM and PPM signals are generated. The PPM signal turns on S_0 and S_1, and capacitor C holds the voltage of V_{non} at that timing; this creates V_{n+1}. After S_1 turns off, S_3 turns on and node P is fixed at a constant voltage V_c. The same operation is repeated at the next time step. Thus, if the waveform V_{non} varies as $V_{non}(t) = f(t)$, then the pulse width of PWM signals or the pulse timing of PPM signals T_n is updated as follows: $T_{n+1} = f(T_n)$.

On the basis of Lazzaro's well-known current-mode WTA circuit [9], we designed a cascade-type voltage-output WTA

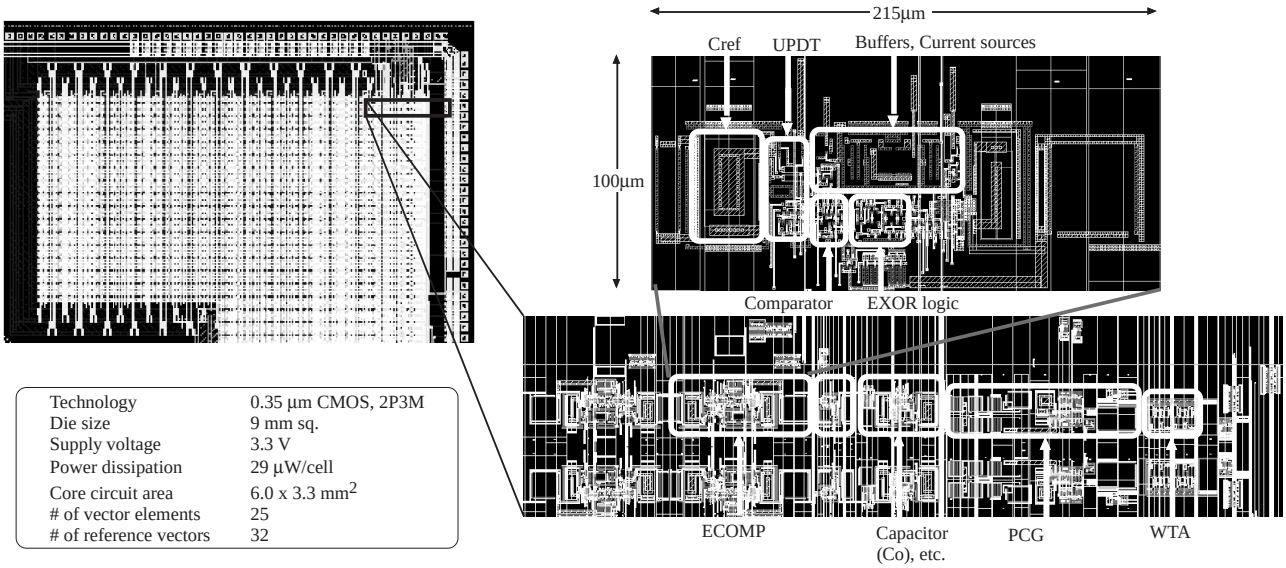


Fig. 4. Chip layout.

circuit. This circuit consists of N cells. Only the cell that receives the largest input outputs the $High$ voltage level, and all the other cells output the Low level.

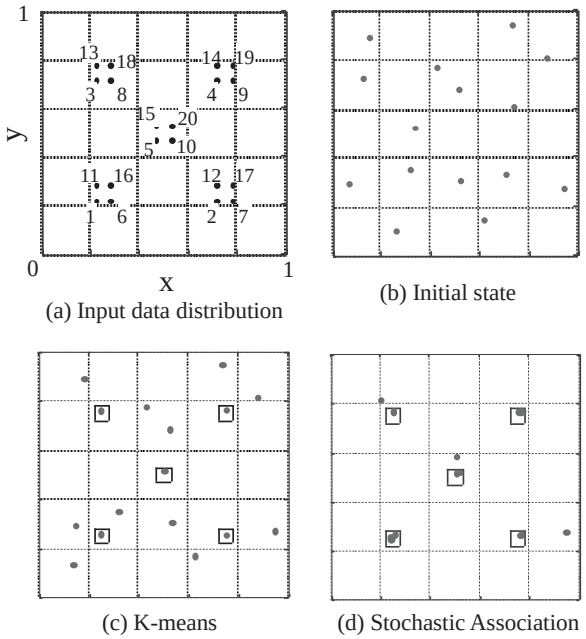
The core part of the chip layout and the unit cell layout are shown in Fig. 4. The fabrication technology used was a 0.35μm CMOS. The layout area of the core circuit is 6 $\times$ 3.3 mm^2 and the maximum numbers of vector elements and reference vectors processed in this chip are 25 and 32, respectively. The unit cell consists of distance evaluation and state update circuits (DIST and UPDT). The layout area of the unit cell is 100 $\times$ 215 μm^2, and the number of MOS transistors used is 134.

IV. Experiments using fabricated chip

A logic analysis system (Agilent 16702A) and arbitrary waveform generators (Tektronix AWG 2021) were used for measurements and experiments. Circuits based on the pulse modulation approach can be controlled by digital systems except the supply of analog voltage waveforms and analog bias voltages, which is an advantage of our pulse modulation approach.

Experiments of stochastic associative processing were performed using the fabricated chip. A parabolic voltage waveform was given as V_{non} to the PCG circuit, which generates logistic chaos. The experimental results demonstrated stochastic association characteristics, in which the association probability increases as the Manhattan distance decreases. It was also verified that less deterministic operation is achieved with a larger maximum pulse width of PWM chaos signals.

For the experiments of SA clustering, 20 data samples were distributed in five clusters in a 2-D space, as shown in Fig. 5(a). In the experiments, the data vectors were presented in the order shown in the figure. As an initial condition of the training, 15



(a) Input data distribution (b) Initial state

(c) K-means (d) Stochastic Association

Fig. 5. Experimental results of SA clustering.

reference vectors were scattered, as shown in Fig. 5(b). The number of updating epochs was set at 200.

A reference vector coordinate of 1 corresponds to a pulse width of 1100 ns, which is the maximum width. On the other hand, a coordinate of 0 corresponds to a pulse width of 100 ns, which is the minimum width. The offset of 100 ns is canceled by setting the same offset to input data vectors.

The procedure of clustering is as follows:

1) Initialize the voltages of capacitor C_0 for summation,
2) Perform dummy operation of chaos generator (10 cycles)

in order to guarantee the random correlation among the chaos generators in all units,
3) Initialize reference vectors (Fig. 5(b)),
4) Perform stochastic association,
5) Observe the WTA output,
6) Update the winning reference vector,
7) Repeat (4)-(6) for given update epochs,
8) Output the coordinates of reference vectors.

On-line K-means clustering can be performed by stopping chaos generation. In this case, because no random fluctuation is added to the distance evaluation, the nearest neighbor reference vector is always updated. Figure 5(c) shows an example of on-line K-means clustering results. As expected, only one reference vector represented each data cluster, and the others were almost never updated.

In this experimental setup, an annealing process could not be performed because of performance limitation of the digital control system. Instead of a normal annealing process, in the initial phase, the chaos circuit was always operated, after the middle phase, chaos generation was performed every other updating; i.e, SA and K-means clustering processes were repeated alternatively. This procedure emulates an annealing process.

Figure 5(d) shows an example of SA clustering results. The initial distribution of reference vectors was the same as that for the on-line K-means clustering. All reference vectors except 3 vectors moved to the data cluster regions. This is because reference vectors far from the cluster regions were sometimes updated due to random fluctuation by chaotic signals. Compared with the on-line K-means algorithm, the SA algorithm gives much better clustering results near the optimum. This result demonstrates the excellent performance of the SA algorithm. The total time of this SA clustering was less than 2 ms. Thus, very high-speed real-time clustering can be achieved by using this VLSI chip.

V. Conclusions

We designed a CMOS VLSI chip that performs stochastic associative processing. The key process is adding random fluctuation for stochastic operation, and it is achieved by a chaos generator circuit designed using a pulse modulation approach.

The fabricated VLSI chip successfully performed stochastic associative processing and on-chip clustering training using it. Experimental results using the fabricated chip demonstrated that our stochastic association clustering has much superior performance to the on-line K-means clustering algorithm.

Acknowledgments

The authors would like to thank Prof. Masataka Hirose for his support and encouragement. This work was supported in part by Grants-in-aid for the Core Research for Evolutional Science and Technology (CREST) from Japan Science and Technology Corporation (JST).

References

[1] K. Rose, E. Gurewitz, and G. C. Fox, "Statistical mechanics and phase transitions in clustering," *Physical Review Letters*, vol. 65, no. 8, pp. 945–948, 1990.

[2] T. Kohonen, *Self-Organization and Associative Memory*. Berlin: Springer-Verlag, 1984.

[3] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten, ""Neural-gas" network for vector quantization and its application to time-series prediction," *IEEE Trans. Neural Networks*, vol. 4, pp. 558–569, 1993.

[4] S. Rovetta and R. Zunino, "Efficient training of neural gas vector quantizers with analog circuit implementation," *IEEE Trans. Circuits & Syst. II*, vol. 46, pp. 688–698, 1999.

[5] T. Morie, T. Matsuura, M. Nagata, and A. Iwata, "An efficient clustering algorithm using stochastic association model and its implementation using nanostructures," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. Cambridge, MA: MIT Press, 2002, pp. 1115–1122.

[6] T. Yamanaka, T. Morie, M. Nagata, and A. Iwata, "A CMOS stochastic associative processor using PWM chaotic signals," *IEICE Trans. Electron.*, vol. E84-C, no. 12, pp. 1723–1729, 2001.

[7] K. Murakoshi, T. Morie, M. Nagata, and A. Iwata, "An arbitrary chaos generator core circuit using PWM/PPM signals," in *Asia and South Pacific Design Automation Conf.*, Tokyo, Jan. 2000, pp. 23–24.

[8] T. Morie, S. Sakabayashi, H. Ando, M. Nagata, and A. Iwata, "Pulse modulation circuit techniques for nonlinear dynamical systems," in *Proc. Int. Symp. on Nonlinear Theory and its Applications (NOLTA'98)*, Crans-Montana, Sept. 1998, pp. 447–450.

[9] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of O(n) complexity," in *Advances in Neural Information Processing Systems 1*, D. S. Touretzky, Ed. Morgan Kaufmann, 1989, pp. 703–711.