An Efficient Clustering Algorithm Using Stochastic Association and Its Implementation Using 3D-Nanodot-Array Structures

Takashi Morie*, Tomohiro Matsuura†, Makoto Nagata‡, 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
† Graduate School of Advanced Sciences of Matter, Hiroshima University
‡ Faculty of Engineering, Kobe University

Abstract – This paper describes a clustering algorithm for vector quantizers using a "stochastic association model". It offers a simple, powerful 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. For hardware implementation of this process, we propose a nanostructure with 3D-nanodot arrays, whose operation is described by a single-electron circuit. It positively uses fluctuation in quantum mechanical tunneling processes.

1 Introduction

Vector quantization (VQ) techniques are used in a wide range of applications, including speech and image processing and data compression. VQ techniques encode a data manifold $V \subseteq \Re^D$ by using only a finite set of reference vectors $w = (w_1, \cdots, w_N)$. A data vector $v \in V$ is represented by the best-matching or "winning" reference vector w_c.

Various clustering algorithms to obtain the best reference vectors which minimize the average distortion error have been reported. Here, we treat on-line training, in which the data point distribution is not given a priori, but instead a stochastic sequence of incoming sample data points drives the adaptation procedure.

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

$$\Delta w_i = \varepsilon \cdot \delta_{ic} \cdot (v(t) - w_i), \qquad (1)$$

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, a common approach is to introduce a "soft-max" adaptation rule that not only adjusts the "winning" reference vector but affects other reference vectors depending on their proximity to v. The well-known "soft-max" clustering algorithms are the *maximum-entropy* (ME) algorithm [1], the *Kohonen's self-organization map* (SOM) [2], and the *neural-gas* (NG) algorithm [3].

In this paper, we propose a new efficient soft-max adaptation algorithm [4]. It employs the *stochastic association model* that we have proposed related to single-electron circuits [5-7]. It is demonstrated from simulation results that our clustering algorithm is as powerful as the other algorithms. We also propose a nanostructure based on a single-electron circuit for implementing the stochastic association model [8-11].

2 Stochastic association algorithm

A usual associative memory is defined as a system that *deterministically* extracts the vector most similar to the input vector from the stored reference vectors. This just corresponds to the process choosing the winning reference vector for a certain data vector in all conventional clustering algorithms.

In our stochastic association (SA) model, the association probability depends on the similarity between the input and the reference vectors [5]. 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.

In the SA algorithm, stochastic fluctuation is added in the evaluation process of distortion error D_i between data vector v and reference vector w_i. The

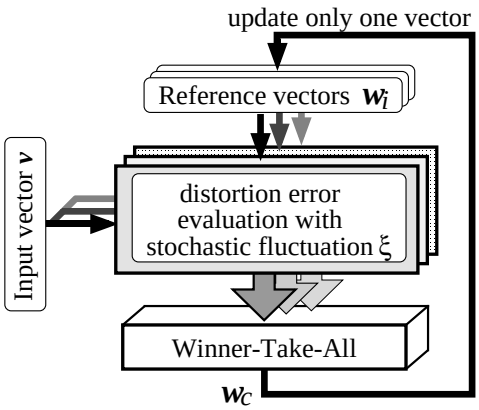

Figure 1: Architecture for clustering processing using the SA model.

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

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

where ξ is a random variable with probability distribution function $\varphi(\xi)$. Therefore, the evaluation result R_i is also considered as a random variable. The probability that R_i has value r_i is represented by

$$\Pr\{R_i = r_i\} = \varphi(r_i - D_i). \tag{3}$$

The winning reference vector w_c is determined by

$$c = \arg \min_i \{R_i\}. \tag{4}$$

If the winning reference vector is updated as expressed by eq. (1), the SA model can provide a new soft-max adaptation rule. Figure 1 shows an architecture for clustering processing using the SA model. The distortion error between the input vector and each stored reference vector is evaluated in parallel with stochastic fluctuation. The winner-take-all circuit deterministically extracts the winner, and only the winning reference vector is updated as in the K-means algorithm. However, unlike the K-means algorithm, the adjusted vector is not always the most similar reference vector, and sometimes other similar vectors are adjusted. The total adjusting tendency in the SA algorithm seems similar to the NG or ME algorithm because the probability of reference vector selection is determined by the neighborhood ranking and the distances between each reference vector and a given data vector.

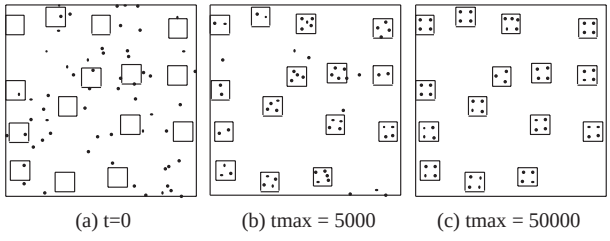

Figure 2: Test problem and clustering results by SA and ME algorithms. Data samples uniformly distribute in square regions, and points represent reference vectors. Both algorithms use the same initial state. The number of total updating steps is denoted by t_{max}.

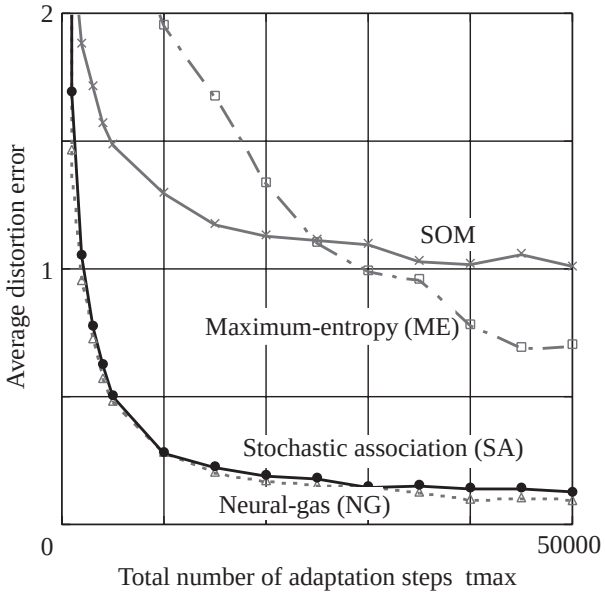

Figure 3: Clustering performance of SA algorithm compared with other clustering approaches.

3 Simulation results

In order to test the performance of the SA algorithm and to compare it with the other soft-max approaches, we performed several simulations.

One of the simulation results are shown in Figs. 2 and 3. For the detail of the simulation, see [4].

The data clusters are of square shape within a two-dimensional input space as shown in Fig. 2.

Figure 2 shows an example of clustering by the SA algorithm compared with that by the ME algorithm. The result of the SA algorithm demonstrates nearly perfect clustering for $t_{max} = 50000$. In contrast, the clustering result by the ME algorithm is not so good although the parameters used were optimized.

The simulation results of clustering performance are shown in Fig. 3. The performance of the SA algorithm is nearly equal to that of the NG algorithm, which is the most efficient clustering method

in this test problem. The number of adaptation steps to reach the steady state and the distortion error at the steady state in the SA algorithm are nearly the same as those in the NG algorithm. Here, all the clustering algorithms including the SA algorithm use an annealing procedure to escape local minima; the parameters that determine the updating step size and the region of proximity were gradually reduced during adaptation.

Consequently, comparing with the other soft-max algorithms, the SA algorithm has nearly the best clustering performance. Moreover, because only one reference vector is adjusted per adaptation step, the computational power required by the SA algorithm is much less than that required by the other soft-max algorithms. If the number of reference vectors is N, the total updating processes of all reference vectors in the SA algorithm are $1/N$ times as many as those in the other algorithms. Thus, the SA algorithm is the most efficient clustering method.

4 Nanostructure

The key for implementing the SA model is adding random fluctuation as expressed by eq. (2). We have already proposed single-electron circuits and nanostructures evaluating Hamming distance for the SA model [5, 6, 8-10]. We have also proposed its CMOS VLSI implementation. [7]

Figure 4(a) and (b) show a nanostructure and the corresponding single-electron circuit, respectively, which are the most sophisticated version of our circuits and structures [11, 4]. The nanostructure consists of plural (M) dot structures arranged on a MOS transistor gate electrode. Each dot structure consists of 1-D dot arrays A_h ($D_1, \cdots, D_n, D_c, D_n, \cdots, D_1$) and A_v (D_{v1}, D_{v2}, D_{v3}), where n means the number of dots at a side of A_h. (From Monte Carlo single-electron circuit simulations, n should be more than 3). The dot diameter assumed is around 1 nm. The capacitance C_o corresponds to the gate capacitance of an ultrasmall MOS transistor. It is assumed that an electron e_M can be introduced in array A_h. Electron e_M, which is initially located at D_c, can move along array A_h through tunneling junctions C_j, but it cannot move to A_v through the normal capacitor C_2. Digital (High/Low) voltages V_{di} and V_{ri} ($i = 1, 2, \cdots, M$) are applied at both edges of A_h, which correspond to elements of data and reference vectors, respectively. Each dot structure simply works as an exclusive-NOR logic gate (bit comparator) with random fluctuation as explained below.

If the two digital data bits (H or L) are matched, electron e_M stabilizes at center dot D_c, otherwise e_M moves to an off-center position. After stabilizing e_M, by changing voltages V_{di}, V_{ri} and back-gate voltage V_{bg}, vertical dot array A_v detects whether e_M stays at D_c or not; only if e_M stays at D_c, A_v is polarized and an electron is induced at the gate electrode of C_o. The total number of induced electrons (N_e) is proportional to the number of dot structures with matched bits; thus the Hamming distance can be measured by counting the induced electrons using the ultrasmall MOS transistor. (If one of the input digital data is applied through an inverter, the number of unmatched bits can be calculated).

The detail of operation stabilizing e_M is as follows: Because of the charging energy of e_M itself, the total energy as a function of the position of e_M in array A_h has two peaks at the midpoints of each side of the array, and has minimal values at D_c and both of D_1 as shown in Fig. 4(c). The energy barrier height for e_M at D_c is assumed larger than the thermal energy at room temperature.

In *L-L state*, the energy at D_1 rises up, thus e_M is most strongly stabilized at D_c. On the other hand, in *H-L(L-H)* or *H-H state*, the energy barrier is lower than that of *L-L state*, thus e_M can more easily overcome the barrier by using thermal noise. Figure 4(d) shows the relation between operation temperature and time (t_M) required until e_M moves to D_1, which was obtained by Monte Carlo single-electron circuit simulation. The moving process assisted by thermal noise is purely stochastic, thus t_M scatters in a wide range. However, because the energy barrier height in *H-L(L-H) states* is lower than that in *H-H state* as shown in Fig. 4(c), there exists a certain time span t_0 within which e_M in *H-L(L-H) states* moves to D_1 while e_M in *H-H state* stays at D_c. At room temperature (300K), t_0 is several microseconds in this case although t_0 depends on the tunneling resistance.

If the detection process starts after t_0, nearly perfect exclusive-NOR (bit comparison) operation is achieved. On the other hand, if the detection timing is shifted from t_0, arbitrary amount of fluctuation can be added in the bit comparison results, as shown in Fig. 5 [11]. Thus, we utilize quantum mechanical tunneling processes assisted by thermal noise in this structure, which is similar to a phenomenon known

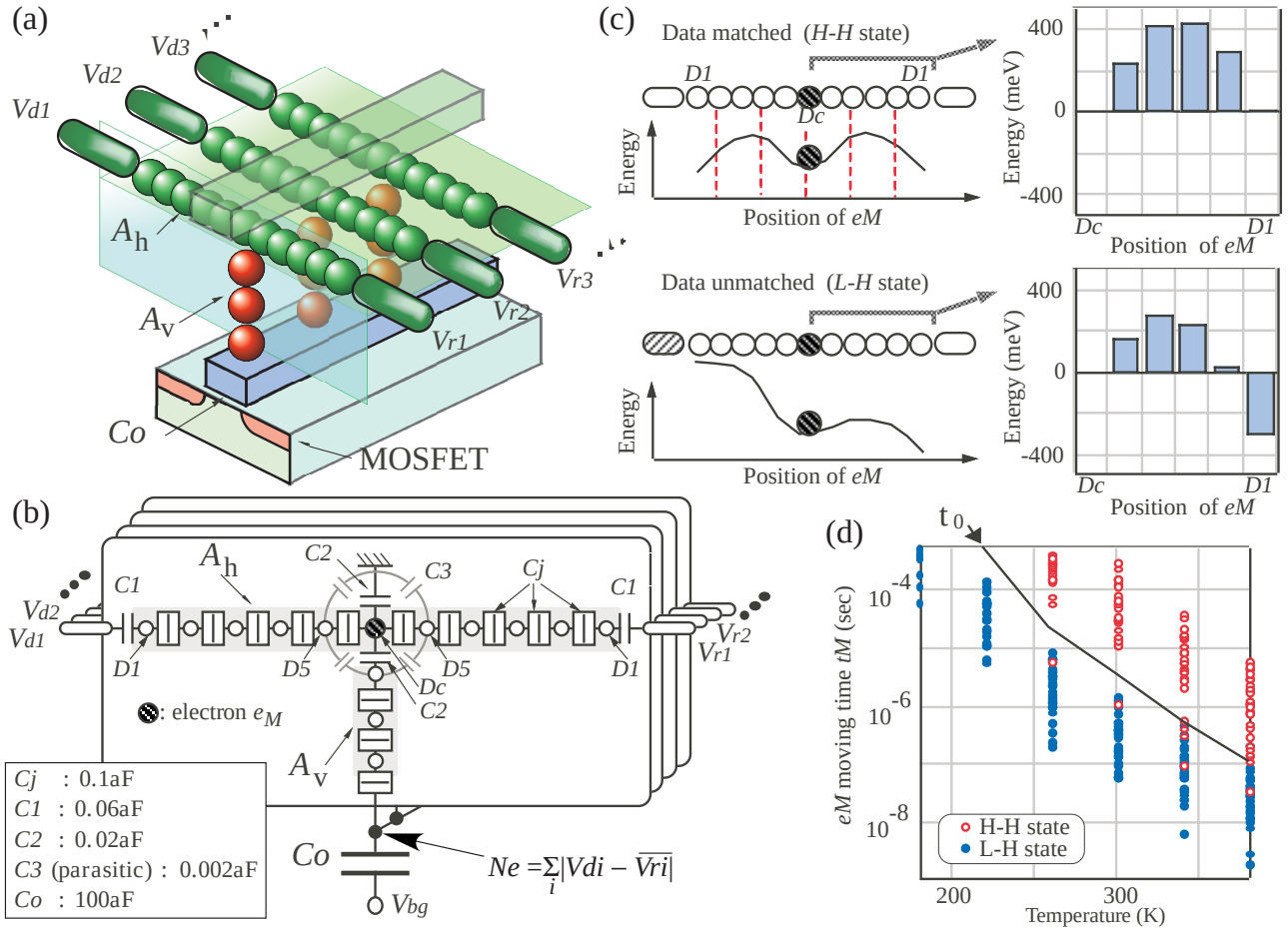

Figure 4: Nanostructure evaluating Hamming distance. (a) Schematic of nanostructure, where dot arrays are extremely enlarged compared with a MOSFET to emphasize the dot structures. (b) Single-electron circuit. (c) Potential profile in dot array A_h. (d) e_M moving time for bit comparator operation.

as *stochastic resonance*.

Although digital data are treated in the above explanation, analog data can be treated in the same circuit by using pulse-width modulation (PWM) signals, which have a digital amplitude and an analog pulse width. Therefore, instead of the Hamming distance, the Manhattan distance can be evaluated by using this nanostructure. Because random fluctuation is naturally added in our nanostructure, it can implement the calculation expressed by eq. (2).

The proposed nanostructure has not yet been fabricated using the present VLSI technology, but the basic technology related to nanocrystalline floating-dot MOSFET devices, which are closely related to our structure, is now being developed [12-14]. Furthermore, well-controlled self-assembly processes using molecular manipulation technology, especially using DNA [15], would be utilized to fabricate our nanostructure.

5 Conclusions

The stochastic association algorithm offers a simple and powerful soft-max adaptation rule for clustering. Although our new method is the same as the simple on-line K-means clustering except for adding random fluctuation in the distortion error evaluation process, it has the most efficient adaptation performance; the computational effort is much smaller compared with the conventional algorithms.

By employing the nanostructure proposed in this paper, very high performance clustering hardware could be constructed.

Acknowledgments

The authors wish 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) from 1997 to 2001.

References

[1] K. Rose, E. Gurewitz, and G. C. Fox, *Physical Review Letters*, **65**, 945–948 (1990)

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

[3] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten, *IEEE Trans. Neural Networks*, **4**, 558–569 (1993)

[4] T. Morie, T. Matsuura, M. Nagata, and A. Iwata, in T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, 1115-1122, Cambridge, MA (2002). MIT Press . This paper can be obtained from NIPS ONLINE (http://www-2.cs.cmu.edu/Groups/NIPS/NIPS2001/papers/#part5).

[5] M. Saen, T. Morie, M. Nagata, and A. Iwata, *IEICE Trans. Electron.*, **E81-C**, 30–35 (1998)

[6] T. Yamanaka, T. Morie, M. Nagata, and A. Iwata, *Nanotechnology*, **11**, 154–160 (2000)

[7] T. Yamanaka, T. Morie, M. Nagata, and A. Iwata, *IEICE Trans. Electron.*, **E84-C**, 1723–1729 (2001)

[8] T. Morie, T. Matsuura, S. Miyata, T. Yamanaka, M. Nagata, and A. Iwata, *Superlattices & Microstructures*, **27**, 613–616 (2000)

[9] T. Matsuura, T. Morie, M. Nagata, and A. Iwata, in *Ext. Abs. of Int. Conf. on Solid State Devices and Materials (SSDM)*, 306-307, Sendai, Japan (2000)

[10] T. Morie, T. Matsuura, M. Nagata, and A. Iwata, in *Extended Abstracts, 4th International Workshop on Quantum Functional Devices (QFD2000)*, 210-213, Kanazawa, Japan (2000)

[11] T. Morie, T. Matsuura, M. Nagata, and A. Iwata, *J. Nanosci. Nanotech.*, **2**, 343–349 (2002)

[12] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbé, and K. Chan, *Appl. Phys. Lett.*, **68**, 1377–1379 (1996)

[13] A. Kohno, H. Murakami, M. Ikeda, H. Nishiyama, S. Miyazaki, and M. Hirose, in *Ext. Abs. of Int. Conf. on Solid State Devices and Materials (SSDM)*, 124-125, Sendai, Japan (2000)

[14] R. Ohba, N. Sugiyama, J. Koga, K. Uchida, and A. Toriumi, in *Ext. Abs. of Int. Conf. on Solid State Devices and Materials (SSDM)*, 122-123, Sendai, Japan (2000)

[15] R. A. Kiehl, in *Extended Abstracts, 4th International Workshop on Quantum Functional Devices (QFD2000)*, 49-51, Kanazawa, Japan (2000)

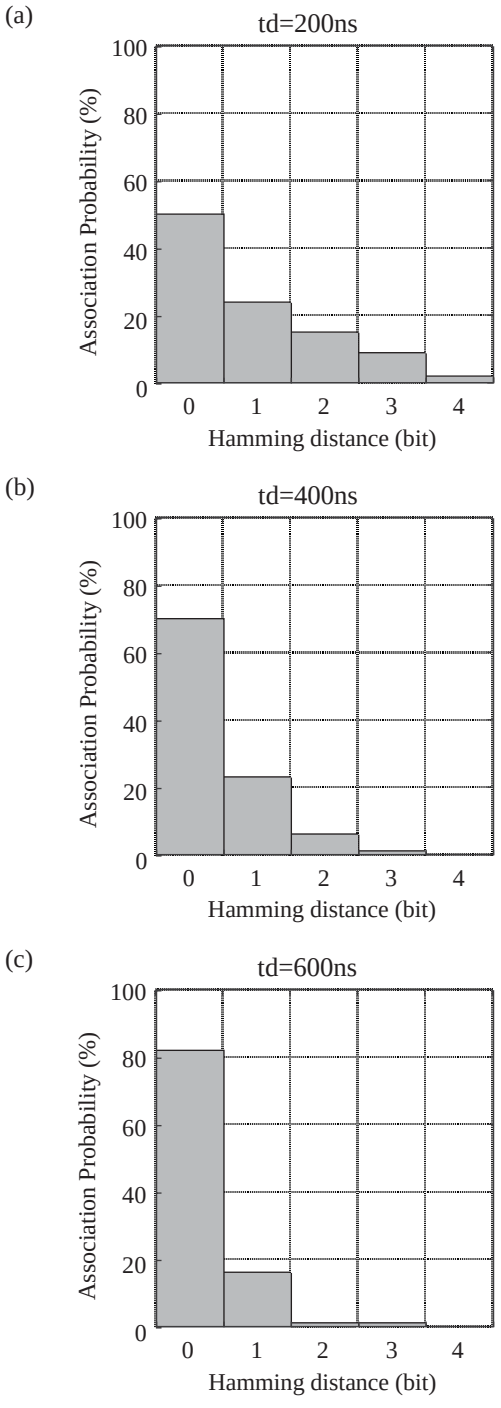

Figure 5: Association probability distribution as a function of Hamming distance for various detection timing t_d. The distribution is broadened by fluctuation due to thermal noise when t_d is different from t_0.