

「確率的連想」によるベクトル量子化器の学習と パルス変調方式によるLSI回路実現

松浦知宏 森江隆 永田真 岩田穆

広島大学 工学部

〒739-8527 東広島市鏡山1-4-1

TEL:0824-24-7686 FAX:0824-22-7195

E-mail: {matsuura, morie, nagata, iwa}@dsl.hiroshima-u.ac.jp

この論文では、「確率的連想」アルゴリズムを用いた、ベクトル量子化器の効率的な学習方法を提案する。確率的連想による学習は soft-max 法の一つである。この手法を用いると、最も優れた学習性能の一つとして知られているニューラルガス法と同等の学習性能が、少ない計算量で得られる。さらに、パルス変調方式を用いた、確率的連想を実行するLSI回路の構成法と、この回路を用いたベクトル量子化器の学習システムを提案する。

確率的連想, ベクトル量子化, soft-max 法, パルス変調方式

A Vector Quantization Training Using Stochastic Association Algorithm and its LSI Circuit Using Pulse-Width Modulation

T. Matsuura, T. Morie, M. Nagata, and A. Iwata

Faculty of Engineering, Hiroshima University

1-4-1 Kagamiyama, Higashi-Hiroshima, 739-8527

TEL : +81-824-24-7686 FAX : +81-824-22-7195

E-mail: {matsuura, morie, nagata, iwa}@dsl.hiroshima-u.ac.jp

This paper proposes a efficient training method for vector quantizers using a "stochastic association" algorithm. It is a new soft-max adaptation rule. The new algorithm can achieve efficient adaptation with low computational power as high as the "neural gas" algorithm, which is reported as one of the most efficient clustering methods. This paper also proposes an LSI circuit using pulse-width modulation for stochastic processing and proposes a training system for vector quantizers using its circuit.

stochastic association, vector quantization, soft-max method, pulse-width modulation

1 はじめに

情報圧縮技術等への応用が可能なベクトル量子化は、データ空間を、有限の数の参照ベクトル（コードブックベクトル）で代表させる処理である。ここでどのようにして、最も歪みの小さい、最適な参照ベクトルを選ぶかが問題となる。そこでデータベクトルに対して、歪みの小さくなる方向に参照ベクトルを少しずつ更新させる学習法が提案されている。

データベクトルを v ，参照ベクトルのセットを $w = (w_1, \dots, w_N)$ とする。ここで $v, w \in \mathbb{R}^D$ である。また、 v を量子化する参照ベクトルを w_c とする。このとき、 v と w_c との平均歪み E は、次式で与えられる。

$$E = \int \|v - w_c\|^2 p(v) d^D v \quad (1)$$

ここで $p(v)$ はデータベクトルの確率分布である。参照ベクトルの学習は、式 (1) を最小にする最適化問題である。

参照ベクトルの学習について、様々なアルゴリズムが提案されているが、ここでは毎回一つのデータベクトルに対して学習を進めるオンライン学習を扱う。最も基本的な方法が K-means 法である。K-means 法は、あるベクトル $v(t)$ が与えられると、それに一番近い参照ベクトル w_c だけを更新させる。このとき、時刻 t での参照ベクトルの更新量 Δw_i は次式で与えられる。

$$\Delta w_i = \epsilon \cdot \delta_{ic} \cdot (v(t) - w_i) \quad (2)$$

ここで、 ϵ は毎回の更新量の絶対値を決めるパラメータ（ステップサイズ）であり、 δ_{ic} はクロネッカーのデルタである。しかし、K-means 法はローカルミニマムに陥りやすい。これを回避するために、soft-max 法と呼ばれる効率のよい学習法がいくつか提案されている。これは v に最も近い参照ベクトルのみでなく、全ての参照ベクトルに対して v との類似度に依存した更新をする方法である。

以下によく知られている soft-max 法の例を述べる。

1. 最大エントロピー法 [1]

平均歪みを系の自由エネルギーと置き換えて統計力学を導入した方法。全ての参照ベクトル w_i は、入力ベクトル v との距離に依存して更新される。 β が小さいときは、類似度が小さいものも更新されるが、 β が大きくなると、類似度の大きいものしか更新されなくなる。

$$\Delta w_i = \epsilon \cdot \frac{e^{-\beta(v-w_i)^2}}{\sum_{j=1}^N e^{-\beta(v-w_j)^2}} \cdot (v(t) - w_i) \quad (3)$$

2. コホネンの自己組織化マップ [2]

自己組織化マップによる学習では、ベクトル空間と位相空間の両方の概念が導入される。ベクトル空間

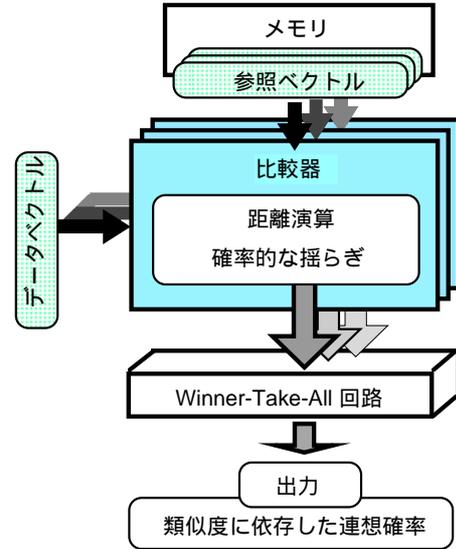


図 1: 確率的連想処理アーキテクチャ

は実際のベクトルの位置を表し、位相空間はそれぞれのベクトルを格子状に割り当てたときの関係を表す。ベクトル空間上で最も近い参照ベクトルが選ばれたとき、位相空間上の近傍参照ベクトルもその距離に依存した関数 $h_\sigma(i, c)$ によって更新する。

$$\Delta w_i = \epsilon \cdot h_\sigma(i, c) \cdot (v(t) - w_i) \quad (4)$$

3. ニューラルガス法 [3]

全ての参照ベクトルを、データベクトルとの距離の小さい順にソートして得られた順番 $k_i(v, w)$ の関数 $h_\lambda(k_i)$ に従って更新する方法。この手法は、最大エントロピー法やコホネンの自己組織化マップと比較して、少ない学習回数で、小さな平均歪みが得られることが報告されている。

$$\Delta w_i = \epsilon \cdot h_\lambda(k_i(v - w_i)) \cdot (v(t) - w_i) \quad (5)$$

しかしすべての参照ベクトルを、ソートしてその順位に応じて更新するので、参照ベクトルの数が増大すると計算量が大きくなる。これに対して、上位のソートだけで十分な学習が実現できる、効率的なニューラルガス法 [4] の報告もある。

この論文では、我々が既に提案している“確率的連想”アルゴリズムを用いた、ベクトル量子化器の効率的な学習法を提案する。確率的連想による学習は soft-max 法の一である。この手法を用いると、最も優れた学習性能の一つとして知られているニューラルガス法と同等の学習結果が、少ない計算量で得られる。

さらにパルス変調方式を用いた、確率的連想処理を実行する LSI 回路の構成法と、この回路を用いたベクトル量子化器の学習システムを提案する。

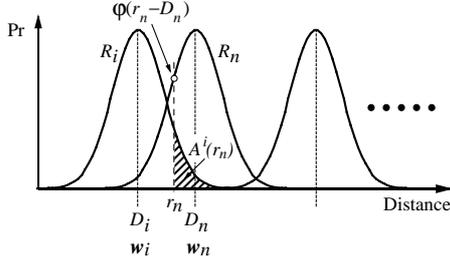


図 2: 参照ベクトル間の位置関係と確率分布

2 確率的連想アルゴリズム

一般的な連想処理では、記憶されている参照ベクトルからデータベクトルに最も近いベクトルを一つだけ出力する。これは K-means 法における、入力ベクトルに最も近い参照ベクトルを選びだす処理と一致する。この場合、同じ入力を与えられたときは、その処理を何回繰り返しても、同じ出力結果しか得られない。

一方、我々は単電子デバイスを用いて、連想処理に確率的な動作を積極的に導入し、入力ベクトルと参照ベクトルとの類似度に依存した確率で出力を得るシステムを提案している [5, 6]。我々はこれを「確率的連想処理システム」と呼んでいる。このシステムでは、データベクトルに最も類似した参照ベクトルが選ばれるとは限らず、類似度に依存した確率で他の類似した参照ベクトルも選ばれる。

確率的連想処理システムを実現するアーキテクチャを図 1 に示す。参照ベクトルを保存するメモリ、データベクトルと参照ベクトルとの類似度を評価する比較器と Winner-Take-All (WTA) 回路から構成される。この比較器では、一つのデータベクトル v と N 個の参照ベクトル w_i との距離 D_i (例えばユークリッド距離 $\|v - w_i\|^2$) を演算する。確率的連想処理では、距離 D_i の演算結果に積極的に確率的な揺らぎ ξ を導入する。その演算結果は次式で与えられる。

$$R_i = D_i + \xi \quad (6)$$

ここで ξ は、確率分布 $\varphi(\xi)$ に従う乱数である。 R_i が r_i という値を取る確率は次式で与えられる。

$$\Pr\{R_i = r_i\} = \varphi(r_i - D_i) \quad (7)$$

WTA 回路は、全ての比較器の演算結果 R_i から、最小の R_i (勝者の参照ベクトル) を決定論的に選び出す。そして勝者の参照ベクトルのみを、K-means 法と同様に式 (2) に従って更新する。

ここで、参照ベクトル w_n が勝者になる確率を求めてみよう。今 R_n が r_n という値をとって w_n が勝者になる確率は、図 2 に示すように、 $R_i > r_n (\forall i \neq n)$ となる確

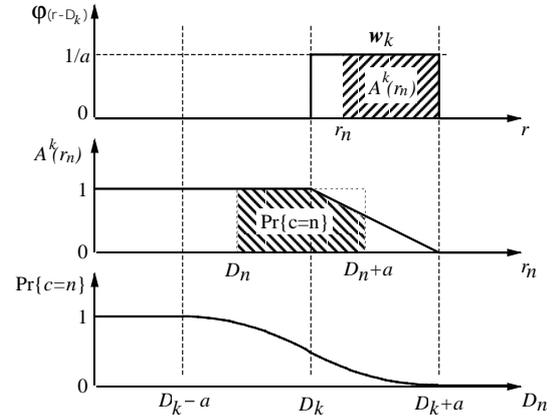


図 3: 一様分布を仮定した場合の連想確率

率と $\varphi(r_n - D_n)$ との積で与えられる。よって、次式が得られる。

$$\Pr\{c = n\} = \int_{-\infty}^{\infty} dr_n \varphi(r_n - D_n) \prod_{i \neq n} A^i(r_n) \quad (8)$$

$$A^i(r_n) \equiv \int_{r_n}^{\infty} \varphi(r - D_i) dr. \quad (9)$$

ここで、次式のような一様分布を仮定して、具体的な \Pr を求めてみよう。

$$\varphi(x) = \begin{cases} 1/a & \text{for } 0 < x < a, \\ 0 & \text{for } x \leq 0, a \leq x, \end{cases} \quad (10)$$

まず (8) 式より、次式が得られる。

$$\Pr\{c = n\} = \int_{D_n}^{D_n+a} \frac{dr_n}{a} \prod_{i \neq n} A^i(r_n) \quad (11)$$

w_n の分布が他の参照ベクトルの分布と重複しない場合、 $\Pr\{c = n\}$ は 0、または 1 の値を取る。

$$\Pr\{c = n\} = \begin{cases} 1 & \text{if } \forall i; D_i - D_n > a, \\ 0 & \text{if } \exists i; D_n - D_i > a. \end{cases} \quad (12)$$

もし仮に w_k の分布が w_n の分布とだけ重複する場合、つまり $|D_n - D_k| < a$ でかつ、他のベクトル $w_i (i \neq k, n)$ が $D_i - D_n > a$ を満たすとき、

$$A^k(r) = \begin{cases} 1 & \text{for } r < D_k, \\ (D_k + a - r)/a & \text{for } D_k < r < D_k + a, \\ 0 & \text{for } D_k + a < r, \end{cases} \quad (13)$$

より、次式を得る。

$$\Pr\{c = n\} = \begin{cases} 1 & \text{for } X < -1, \\ 1 - (X + 1)^2/2 & \text{for } -1 \leq X < 0, \\ (X - 1)^2/2 & \text{for } 0 \leq X < 1, \\ 0 & \text{for } 1 \leq X, \end{cases} \quad (14)$$

ここで $X = (D_n - D_k)/a$ である。以上の関係を図 3 に示す。

このように、 w_n が勝者となる確率は各参照ベクトル w_i の位置関係に依存しているため、確率的連想はニューラルガス法に類似している。しかし同様に、各参照ベクトル w_i 間の距離にも依存しているため、確率的連想は最大エントロピー法にも類似しているといえる。

確率的連想処理は K-means 法と同様に一回の学習で更新される参照ベクトルは一つだけだが、最も類似した参照ベクトルが更新されるとは限らず、他の類似した参照ベクトルもある確率で更新される。

ハードウェア化を考えると、確率的連想による学習は他に比べて実現性が容易である。最大エントロピー法は更新量が距離の関数であり、また全ての参照ベクトルを更新するためハードウェア化が複雑になる。ニューラルガス法はソートの計算に時間がかかり、同じく全ての参照ベクトルを更新する必要がある。それに対し、確率的連想では一回の学習で更新される参照ベクトルは一つだけであるため、学習に要する演算は少ない。ただし、確率的な揺らぎを加える特別な回路が必要となる。

3 確率的連想によるベクトル量子化器の学習のシミュレーション

3.1 シミュレーションの評価方法

確率的連想処理を用いた学習の性能評価のために、数値シミュレーションを行った。評価は文献 [3] で行われている各学習法の比較にしたがって行った。具体的な条件は以下の通りである。

- データベクトル

対象にするベクトルの次元は $D = 2$ 、すなわち平面上に分布する 2 次元のベクトルとする。データベクトルは 15 個の重ならない正方形の領域クラスタに一樣に分布する。参照ベクトルの数は、 $N = 4 \times (\text{クラスタ数}) = 4 \times 15$ とする。参照ベクトルの初期値はランダムに与えられる。

- 評価値 α

上記のようなデータベクトル分布が与えられたときの参照ベクトルの最適解は、各クラスタに 4 個ずつ均等な位置をとる配置である。この時の平均歪みを E_0 とする。また、学習回数の上限値を $t = t_{max}$ と定め、そのときの平均歪みを $E(t_{max})$ とする。この E_0 と $E(t_{max})$ を用いて、評価値 α を次のように定める。

$$\alpha = \frac{E(t_{max}) - E_0}{E_0} \quad (15)$$

α が 0 に近いほど、最適解に近いといえる。

- 比較対象となる各学習法の更新量
コホネンの自己組織化マップでは

$$h_\sigma(i, j) = \exp(-\|w_i - w_j\|^2 / 2\sigma(t)^2) \quad (16)$$

ニューラルガス法では

$$h_\lambda(k) = \exp(-k/\lambda(t)) \quad (17)$$

をそれぞれ用いる。パラメータ $\sigma(t)$ $\lambda(t)$ および最大エントロピー法の $\beta(t)$ は以下に示す学習回数 t に依存する式に従って変化させる。

$$x(t) = x_i(x_f/x_i)^{t/t_{max}} \quad (\text{for } \sigma, \lambda, \beta) \quad (18)$$

文献 [3] で最適値とされている各パラメータの値について以下に示す。

最大エントロピー法

$$\beta_i = 1, \beta_f = 10000$$

自己組織化マップ

$$\sigma_i = 2, \sigma_f = 0.01$$

ニューラルガス法

$$\lambda_i = 10, \lambda_f = 0.01$$

上記のようなパラメータのとき、いずれの学習においても、学習の始めのうちは、 w_c 以外の参照ベクトルの更新量も大きい。逆に $t = t_{max}$ に近づくと w_c のみ更新するようになり、K-means 法と等価な学習になる。

- ステップサイズ ϵ

各学習法で共通のパラメータであるステップサイズ ϵ も以下のように学習回数 t に依存して変化させる。

$$\epsilon(t) = \epsilon_i(\epsilon_f/\epsilon_i)^{t/t_{max}} \quad (19)$$

ここで、 $\epsilon_i = 0.5$ 、 $\epsilon_f = 0.005$ とした。このとき、更新量は学習の始めのうちは大きく、 $t = t_{max}$ に近づくとつれて次第に減少する。

確率的連想処理による学習でも、パラメータは学習回数 t に依存させる。ステップサイズは式 (19) を用いる。もう一つのパラメータは確率的な揺らぎの度合いである。今回行ったシミュレーションでは、ユークリッド距離 $\|v - w_i\|^2$ に標準偏差が γ のガウス分布に従う揺らぎを与えた。

$$\varphi(\xi) = N(0, \gamma^2) \equiv \frac{1}{\sqrt{2\pi}\gamma} \exp(-\frac{\xi^2}{2\gamma^2}) \quad (20)$$

γ も式 (18) に従って変化させた。パラメータは、 $\gamma_i = 0.2$ 、 $\gamma_f = 0.0001$ とした。

□ 入力ベクトルの分布 (一様分布) • 参照ベクトル

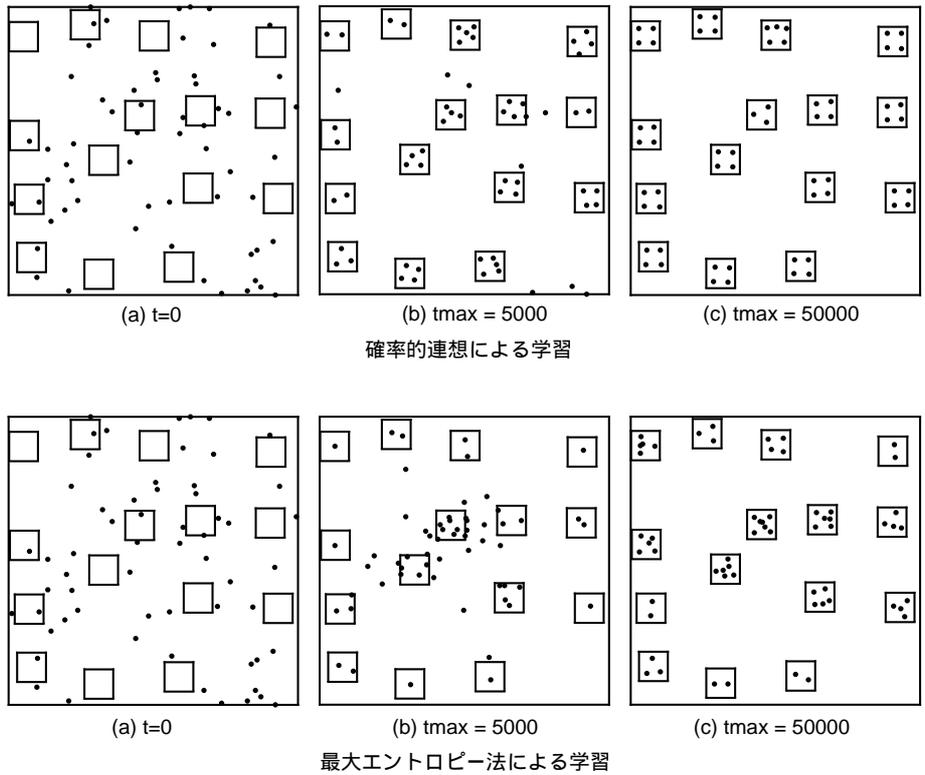


図 4: 確率的連想学習と最大エントロピー法による学習結果の一例

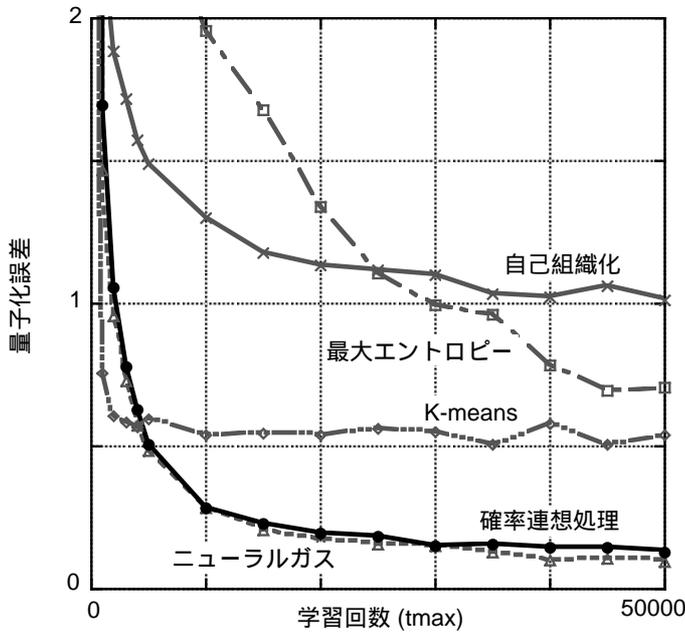


図 5: 確率的連想学習と他の学習法との比較

が 4 個ずつ均等に配置するときが最適な位置である。両方とも学習回数を増やすにつれて、最適な位置に近づいているが、確率的連想による学習の方が最適解により近づいていることがわかる。

確率的連想による学習と他の学習法との学習性能の比較結果を図 5 に示す。ここで初期状態による評価値の偏りを防ぐため、文献 [3] と同様に、参照ベクトルと入力ベクトルの初期分布を 50 回ランダムに変えてシミュレーションを繰り返し、評価値 α はその平均値で求めた。図 5 からわかるように、確率的連想による学習は最も効率の良い学習法であるニューラルガス法と同程度の学習性能を有している。確率的連想による学習で安定化する学習回数と、そのときの平均歪みはニューラルガス法とほぼ同じである。

しかし、ニューラルガス法と違い、確率的連想処理は類似度に応じたソーティングを行う必要がない。さらに一回の学習において更新される参照ベクトルは一つだけである。このことから、確率的連想による計算コストはニューラルガス法よりも少なくすむ。結論として、確率的連想による学習の方が効率的な学習法であるといえる。

3.2 シミュレーション結果

図 4 に確率的連想と最大エントロピー法によるシミュレーション結果の一例を示す。各領域内に参照ベクトル

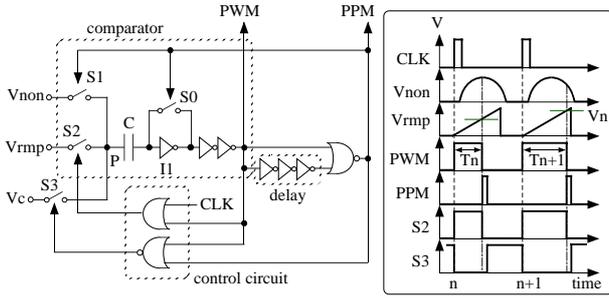


図 6: PWM カオス発生回路と動作タイミング図

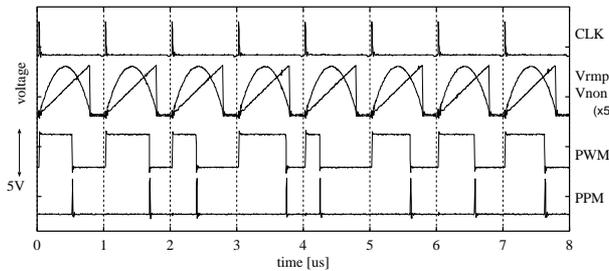


図 7: ロジスティック写像を用いた PWM カオス信号の発生

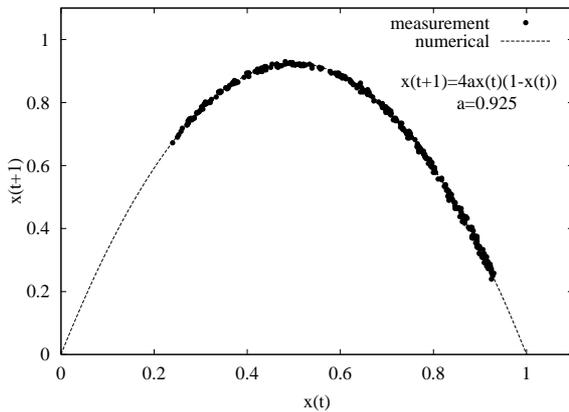


図 8: ロジスティック写像のリターンマップ

4 パルス変調方式による LSI 回路構成

第 2 節で述べたとおり，確率的な揺らぎの実現法を別にすれば，確率的連想処理は比較的ハードウェア化が容易である。ここでは我々が既に提案しているパルス幅変調 (PWM) 回路を用いて，ベクトル量子化器の学習に適用可能な，確率的連想処理のための LSI 回路構成を提案する。

4.1 PWM カオス発生回路

確率的な揺らぎをハードウェア的に実現するために多数の乱数発生回路が必要になる。デジタル回路でも，乱数発生回路は構成できるが，回路面積や，相関関係のない乱数を発生させることを考えると，実現は容易ではない。ここでは我々が開発した PWM カオス発生回路 [7] を用いる構成を考えた。PWM カオス発生回路の回路図を図 6 に示す。この回路は，任意の時間波形を有する電圧源 V_{non} (回路外部から与える) をパルス幅変調/パルス位相変調 (PWM/PPM) 信号でキャパシタ C にサンプリングし，次のタイムステップでそのサンプリング電圧をもとに新たな PWM/PPM 信号を生成していくという回路である。 V_{non} に非線形波形を与えれば，それに応じたダイナミクスが実現でき，カオスを生成できる。図 7 に示すように，クロック CLK に同期して，パルス幅がカオス的に変化する PWM 信号が得られる。ロジスティック写像を形成した場合のリターンマップを図 8 に示す。この場合，ほぼ一様分布をもつ疑似乱数が生成できる。この回路は面積 100μ 角以下で，1MHz 以上のレートで相関のないカオス的なパルス幅を持つ PWM 信号を出力できるので，チップ上に多数の乱数生成源を配置することができる。

4.2 確率的連想処理回路

図 9 に回路構成を示す。この回路はパターン比較器，WTA 回路とメモリから構成される。パターン比較器は，EXOR ゲート，PWM カオス発生回路とスイッチト電流源で構成される。データは PWM 信号で表現される。まず入力データと記憶データの各要素 v_j, w_{ij} を EXOR ゲートに通すことで， $|v_j - w_{ij}|$ の幅を有する PWM 信号が得られる。類似度が大きいほどこの値は小さくなる。この信号をスイッチト電流源により電流値に変換する。さらに PWM カオス発生回路からの疑似乱数としての PWM 信号 ξ_i が加わることで，確率的な動作が実現できる。キャパシタに保持される電荷 V_i は次式で与えられる。

$$V_i = V_{DD} - \sum_{j=1}^D |v_j - w_{ij}| - \xi_i \quad (21)$$

このように，この回路ではマンハッタン距離にほぼ一様分布の揺らぎが加えられるが，数値シミュレーションでは，この計算法でも前節のシミュレーション結果と同様な結果を得ている。

この式からわかるように，類似度が大きいほど V_i の値が大きくなる確率が高い。そして WTA 回路で最大の V_i を検出することで，確率的連想処理機能を実現する。さらに，WTA の出力から勝者の参照ベクトル w_c を決定し，式 (2) に従って w_i の更新を行う。これを繰り返し実

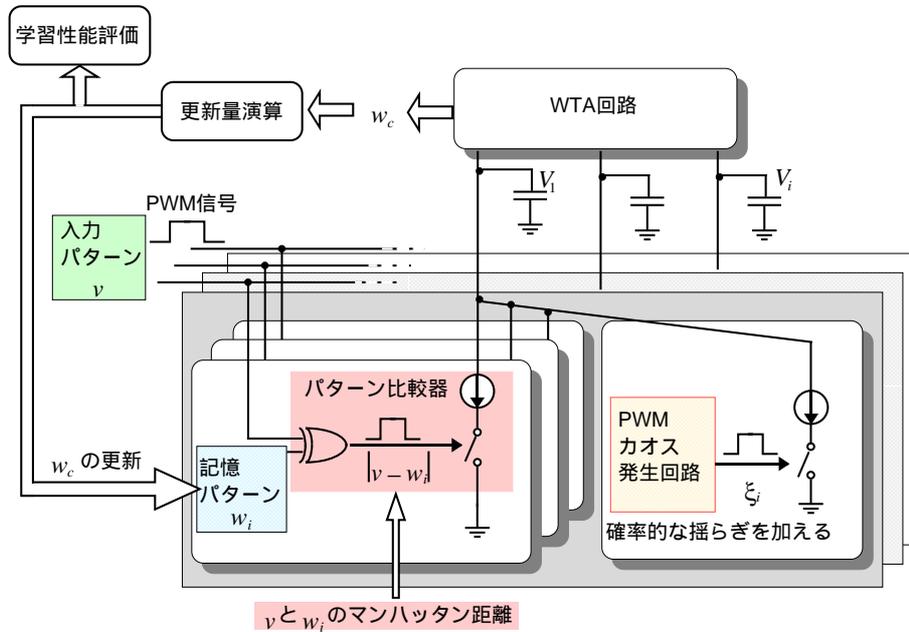


図 9: パルス変調方式による確率的連想処理回路

行することで、ベクトル量子化器の学習が実現できる。

5 まとめ

確率的連想処理によるベクトル量子化器の学習法を提案した。この学習法は類似度演算に確率的な揺らぎが加わる点を除いては最も簡単な K-means 法と同等であるが、ニューラルガス法と同等の学習性能を得ることができる。さらに、この学習法はソーティングを行う必要がなく、一回の学習において更新される参照ベクトルは一つだけであるので、計算コストはニューラルガス法よりもはるかに少なく、ハードウェア化も容易である。このことから、確率的連想は非常に効率的な学習法であるといえる。

また、パルス変調方式による確率的連想の LSI 回路構成を提案した。今後、この回路を設計し、ベクトル量子化器の学習システムを構築する予定である。

参考文献

- [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*, Springer-Verlag, Berlin, 1984.
- [3] T. M. Martinez, 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.*, vol. 46, pp. 688–698, 1999.
- [5] M. Saen, T. Morie, M. Nagata, and A. Iwata, "A Stochastic Associative Memory Using Single-Electron Tunneling Devices," *IEICE Trans. Electron.*, vol. E81-C, no. 1, pp. 30–35, 1998.
- [6] T. Yamanaka, T. Morie, M. Nagata, and A. Iwata, "A Stochastic Associative Memory Using Single-Electron Devices and Its Application to Digit Pattern Association," in *Ext. Abs. of Int. Conf. on Solid State Devices and Materials*, pp. 190–191, Hiroshima, Sept. 1998.
- [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.*, pp. 23–24, Tokyo, Jan. 2000.