

社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
CAS2001-125, DSP2001-208, CS2001-148(2002-03)

アナログ行列演算器を用いた音声認識回路

菅原 崇行 宮永 喜一 吉田 則信

北海道大学大学院工学研究科
〒060-8628 北海道札幌市北区北 13 条西 8 丁目

E-mail: {taka, miyanaga, yoshida}@ice.eng.hokudai.ac.jp

あらまし デジタル信号処理を含む多くのアルゴリズムで行列演算は用いられる。デジタル LSI で行列演算回路を設計した場合、加算と乗算の繰り返し処理によりハードウェアリソースを大量に消費する。本稿で設計したアナログ行列演算器は、大部分がコンデンサで構成される非常に単純な回路であり、同機能のデジタル回路に比べ面積、速度、消費電力の面で有効性が見込まれる。さらに本稿では、アナログ行列演算器を用いた音声認識システムの設計について検討する。

キーワード アナログ回路, 行列演算, 音声認識

An Analog C-matrix Circuit used for Speech Recognition System

Takayuki SUGAWARA, Yoshikazu MIYANAGA, and Norinobu YSHIDA

Graduate School of Engineering, Hokkaido University
Kita 13 Nishi 8, Kita-ku, Sapporo-shi, 060-8628 Japan

E-mail: {taka, miyanaga, yoshida}@ice.eng.hokudai.ac.jp

Abstract Calculation of matrices is applied to much algorithm including digital signal processing. When the matrix calculation is designed with digital circuits in LSI, a hardware resource is consumed in large quantities by reiteration processing of addition and multiplication. The proposed circuits are quite simple and designed with capacitors. By using this circuit, a simple matrix calculation can be realized with analog circuits. In case of a new LSI design with analog circuits, the small number of circuit elements, the high speed calculation and the low power consumption can be realized. Furthermore, the speech recognition system is designed using this circuit.

Key words Analog Circuit, Calculation of Matrices, Speech Recognition

1. まえがき

行列演算は基本的な演算の一つであり、多くのデジタル信号処理に利用されている。デジタル回路での行列演算器を考えたとき、加算器と乗算器によって構成することができるが、繰り返し計算を行う必要があるため、行列の要素が多くなるにつれ演算時間が増大する。演算時間の短縮を考えた場合並列処理・パイプライン処理を行うことが考えられるが、面積の増大という問題があり、演算時間と回路面積のトレードオフについて考慮する必要がある [1][4]。

一方で、アナログ回路の特性を利用した演算器を用いて集積回路を設計した場合、単純な回路構成で設計できるため素子数の減少、高速化、低消費電力化が可能となる。本稿で利用する行列演算器は、大部分がコンデンサで構成される、極めて単純な回路である。すべてのコンデンサは格子状に配列されているため、この回路をCマトリクスと呼ぶ。

従来の音声認識システムはソフトウェアやDSPチップにより実現することが前提条件であり [2]、システムのハードウェア化、さらには、ハードウェア化に向けたシステムの方式・アーキテクチャの最適設計などはあまり検討されていなかった。その中で我々は、音声認識システムのハードウェア化を目指し、エージェント型特徴クラスタリング・ラベリングの構築 [6]、Cマトリクスを用いたハードウェアの検討を行ってきた [7]。本稿では音声分析もCマトリクスを用いて行い、一括してアナログ回路によって処理される、音声認識システムの設計について検討する。

2. Cマトリクス

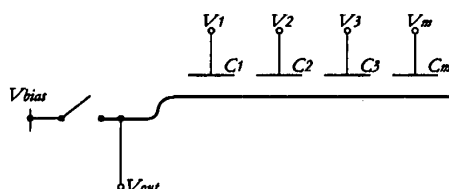


図 1: コンデンサ列

図1のようなコンデンサ列について考える。スイッチが閉じている間、負ノードには V_{bias} が印加さ

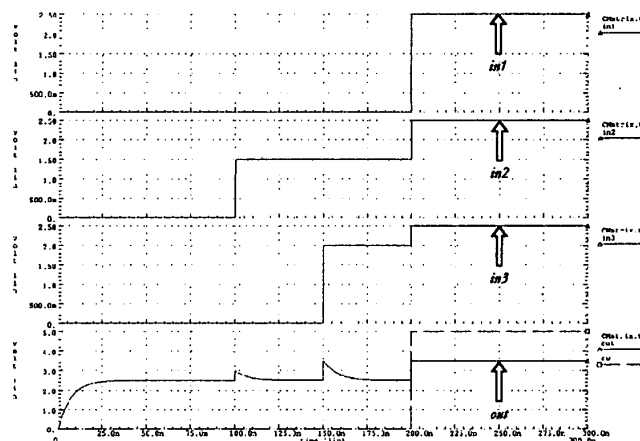


図 2: コンデンサ列のSPICEシミュレーション結果

れている。正ノードに $V_1, V_2, \dots, V_m$ が入力されたとき、各ゲート間に保存される電位差は $V_1 - V_{bias}, V_2 - V_{bias}, \dots, V_m - V_{bias}$ である。次にスイッチが開いたとき、負ノードは V_{bias} から切り離され、各正ノードの入力が $V'_1, V'_2, \dots, V'_m$ に変化したならば:

$$V_{out} = \frac{\sum_{l=1}^m C_l (V'_l - V_l)}{\sum_{l=1}^m C_l} + V_{bias} \quad (1)$$

図2に簡単なSPICEシミュレーション結果を示す。3つのコンデンサからなり各容量は1pF, 2pF, 3pFである。入力電圧はそれぞれ0V, 1.5V, 2Vから2.5Vの変化する。またバイアス電圧は2.5Vである。出力電圧は期待通り3.5Vとなる。

Cマトリクスはコンデンサ列を数行並べたものである。

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (2)$$

このような k 行 n 列の行列演算は、図3で表すCマトリクスで行える。実際に設計を行う場合フリッジ効果を考慮する必要があるため、チップのデザインルールで求まるコンデンサの最小方形などを単位コンデンサとしてメッシュ状に配置する。各コンデンサの容量は単位コンデンサ間の配線を変える事で決定できる。そのため a_{kn} は正数でなければならない。

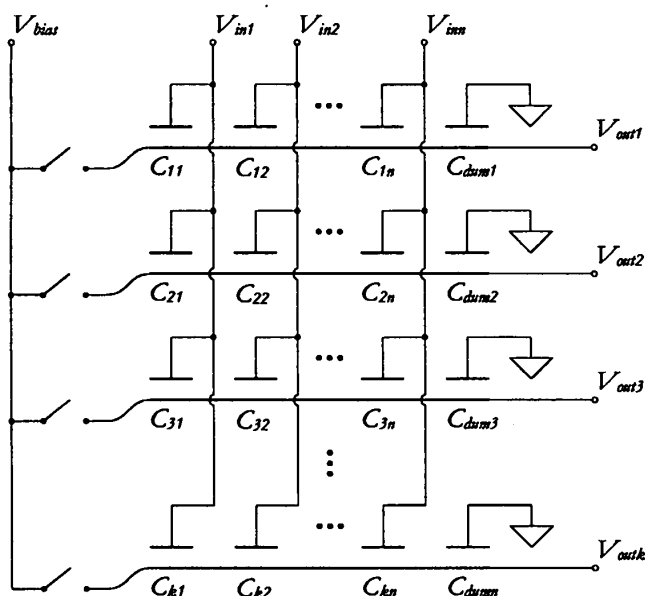


図 3: C マトリクス

容量の最小値を C_0 として:

$$C_{kn} = a_{kn} C_0 \quad (3)$$

入力電圧の変化量 $V_{in,n} = x_n$ とすると各出力は:

$$V_{out,k} = \frac{C_0}{C_{all}} \sum_{j=1}^n a_{kj} x_j + V_{bias} \quad (4)$$

よって, V_{bias} を基準として $V_{out,k}$ の比は (2) 中の y_k の比と同じになる。各行の総容量が同じ値 C_{all} になるように, ダミーコンデンサの容量 $C_{dum,n}$ を設定する。また, 出力電圧が負にならないようにバイアス電圧 V_{bias} をかける。

3. 音声認識システム

我々の音声認識システムのハードウェア化について, すでに図 4 のようなアーキテクチャを考えている。この中で, 音声の特徴を波形から求める部分, 音声の特徴をクラスタリングする部分, さらに定常的な音声のラベリングを行う部分をアナログ LSI 化することを目的として, 後者に関しては文献 [7] において提案されている。

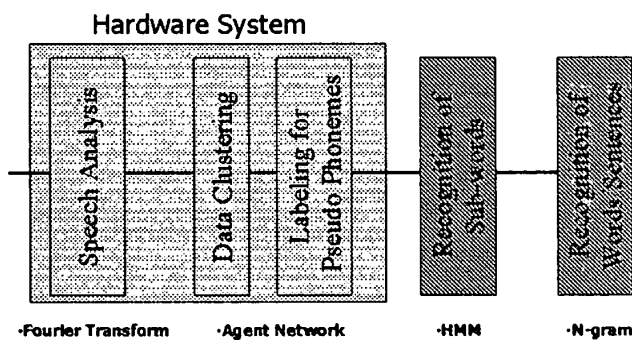


図 4: 音声処理・認識システムの概要

3.1. 音声の特徴推定

音声の特徴推定には, 従来より線形予測理論に基づく方法が多数提案されているが, 本システムでは行列演算で表現できる離散フーリエ変換 (DFT) を利用する。音声信号を $s(n)$ 窓関数を $w(n)$ とすると出力である対数パワースペクトル $S_r(k)$ は:

$$S(k) = \sum_{n=1}^N s(n)w(n)e^{-j\frac{2\pi k}{N}n} \quad (5)$$

$$S_r(k) = \log |S(k)| \quad (6)$$

$$1 \leq k \leq \frac{N}{2}$$

3.2. クラスタリング・ラベリングシステム

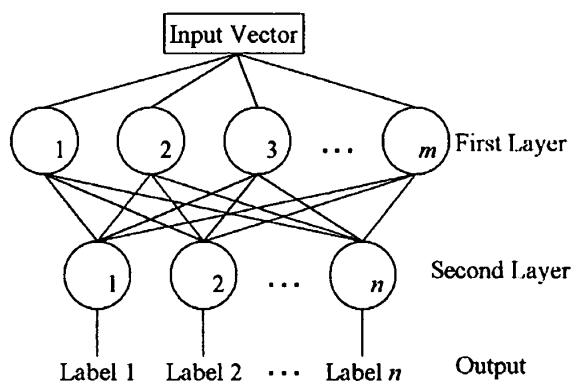


図 5: クラスタリング・ラベリングシステム

本システムで用いるクラスタリング・ラベリングシステムは 2 つの層で構成されている (図 5)。図中

で第1層は入力データに従って、自己組織化アルゴリズムに基づいた特徴を出力するクラスタリング層、第2層は第1層の出力が入力される層で、教師付きアルゴリズムに基づいた重みをかけて足しあわせるラベリング層である。ネットワークの学習は、同一の動作をするソフトウェアシステムを構築し、文献[5]の手法により決定する。認識時には各ノード内部情報は更新されない。さらに各ネットワーク内では一切通信が行われない。

認識時に用いる計算式を示す。クラスタリング層には m 個のクラスタノードがあり、各々のノードはパターンベクトル $\mathbf{x}_i$ ($i = 1, 2, \dots, m$) をもつ。それぞれのノードは p 次元の入力ベクトル $\mathbf{y} = (y_1, y_2, \dots, y_p)$ とパターンベクトル $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ とのユークリッド距離 D_i ($i = 1, 2, \dots, m$) に基づいた類似度 S_i ($i = 1, 2, \dots, m$) を次のように計算する:

$$D_i = \sqrt{\sum_{j=1}^p (y_j - x_{ij})^2} \quad (7)$$

$$S_i = \begin{cases} 1 - (D_i/D_s)^2 & D_i < D_s \\ 0 & D_i \geq D_s \end{cases} \quad (8)$$

D_s は非線形問題に対応させるため設けた閾値である。ラベリング層は n 個のノードをもち、クラスタリング層の出力 S_i に m 次元の重みベクトル $\mathbf{w}_t = (w_{t1}, w_{t2}, \dots, w_{tm})$ ($t = 1, 2, \dots, n$) をかけて足し合わせる。システムの入力 $\mathbf{z} = (z_1, z_2, \dots, z_n)$ はその符号である:

$$R_t = \sum_{i=1}^m w_{ti} S_i \quad (9)$$

$$z_t = \begin{cases} 0 & R_t < 0 \\ 1 & R_t \geq 0 \end{cases} \quad (10)$$

4. アーキテクチャ

図6に提案するシステムの構成を示す。入力される音声信号は連続であるので、リングカウンタとMOSスイッチを用いてサンプリングを行う[8]。サンプリングされた信号は、DFTのポイント数に従って離散フーリエ変換部に並列に入力し、対数パワースペクトルを求める。さらにクラスタリング・ラベリング部に並列に入力し音声認識を行う。

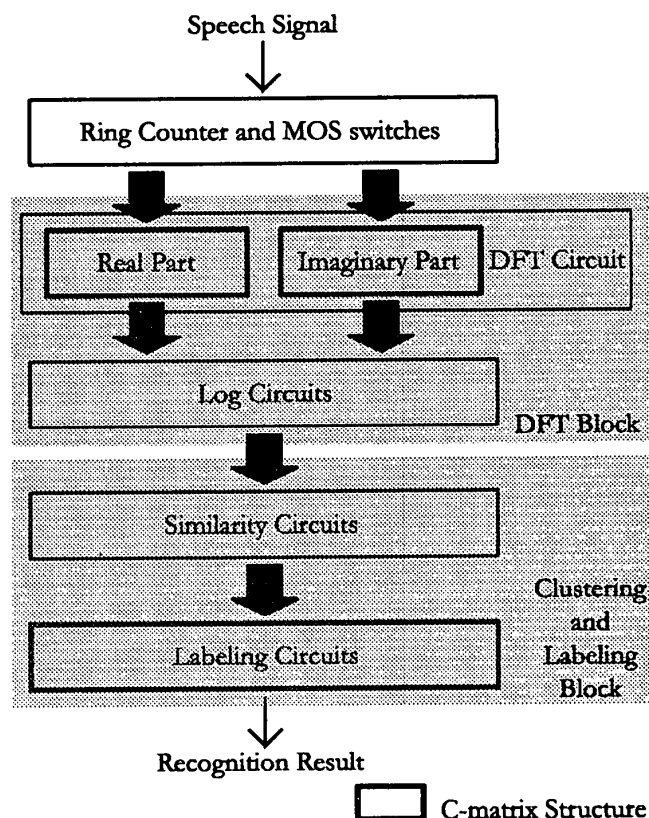


図6: 音声処理・認識システムの構成

4.1. 離散フーリエ変換部

DFTを行うのに実部と虚部に分けて計算するので、 k 行 n 列の C マトリクスを2つ用意する。ポイント数が N である場合、 $k = \frac{N}{2} + 1, n = N + 1$ である。 C マトリクスを用いて行列演算を行う場合、各重みは単位コンデンサを連結した数によって決まるので正数でなければならない。式(5)を変形して、 $a_{kn} = w(n) \cos \frac{2\pi kn}{N} n, b_{kn} = -w(n) \sin \frac{2\pi kn}{N} n$ が DFT における重みとなるが、これらの比をもとめる精度内で整数に丸め込む。仮に 256 階調で演算を行うとすれば:

$$c_{max} = \max\{a_{kn}, b_{kn}\} \quad (11)$$

$$c_{min} = \min\{a_{kn}, b_{kn}\} \quad (12)$$

実部を計算する C マトリクスについて考える。

$$c_{kn} = \text{round} \left\{ \frac{255a_{kn}}{c_{max} - c_{min}} \right\} \quad (13)$$

とする。Cマトリクスの1行を比較コンデンサ列として用い、その容量は:

$$C_{cmp,n} = |c_{min,n}| \quad (14)$$

$$c_{min,n} = \min\{c_{1n}, c_{2n}, \dots, c_{kn}\} \quad (15)$$

その他の容量は:

$$C_{kn} = c_{kn} + C_{cmp,n} \quad (16)$$

式(11)-(16)により、すべての容量が正数であることを保証する。各入力電圧の変化量が $V_{in,n}$ であった場合、比較コンデンサ列の出力 V_{cmp} とその他の出力 $V_{out,k}$ は:

$$V_{cmp} = \frac{\sum_{n=1}^N C_{cmp,n} V_{in,n}}{C_{all}} + V_{bias} \quad (17)$$

$$V_{out,k} = \frac{\sum_{n=1}^N c_{kn} V_{in,n} + \sum_{n=1}^N C_{cmp,n} V_{in,n}}{C_{all}} + V_{bias} \quad (18)$$

$$V_{out,k} - V_{cmp} = \frac{\sum_{n=1}^N c_{kn} V_{in,n}}{C_{all}} \quad (19)$$

虚部についても同様である。

$V_{out,k}$ と V_{cmp} は差動電圧として絶対値回路[3]に入力される。絶対値回路は2対の差動トランスコンダクタンス増幅器と2対のカレントミラーからなり、入力電圧差が V である場合出力電流 I_{out} は:

$$I_{out} = I_0 \tanh \left\{ \frac{\kappa |V|}{2} \right\} \quad (20)$$

この式が描く曲線に対数曲線の近似として利用する。実部と虚部からくる絶対値回路の出力電流を、電流和としてダイオード接続されたnMOSに入力し、出力電圧を得る。

4.2. クラスタリング・ラベリング部

クラスタリング層はニューロンMOSFETとMOSFETの基本特性を用いて類似度を出力する[7]。式(7)での x_i の成分は、正数に丸め込む。ラベリング層はDFTを行うのと同様に比較コンデンサ列を持つCマトリクスを用いが、 $V_{out,n}$ と V_{cmp} を比較器に入力する。式(9)での w_t は式(11)-(16)と同様に変形してコンデンサの容量とする。比較器の出力は:

$$\begin{cases} 0 & V_{out,k} < V_{cmp} \\ 1 & V_{out,k} \geq V_{cmp} \end{cases} \quad (21)$$

なので、式(9),(10)と同様な演算である。

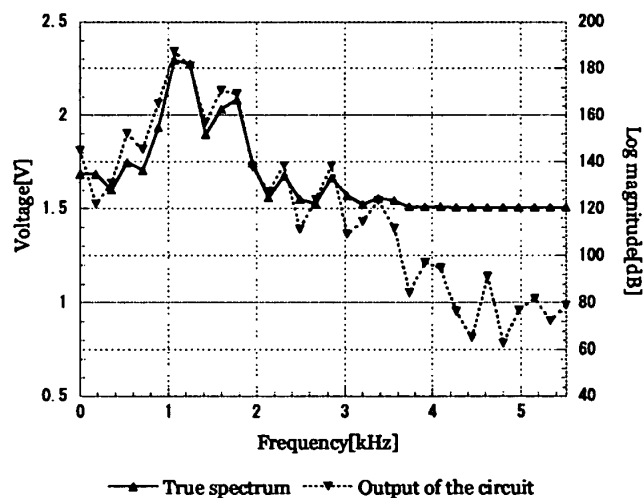


図 7: シミュレーション結果

5. シミュレーション

ポイント数を 64、演算精度を 256 階調とし離散フーリエ変換部を設計し SPICE シミュレーションを行った。 $V_{dd} = 5V$, $V_{bias} = 2.5V$ とした。入力は、11.25kHz で標本化された合成音声信号 64 ポイントの振幅を 2.5V が基準となるように 0 から 5V の間に丸めたものである。第 1 ホルマント周波数=1.0kHz, 第 2 ホルマント周波数=1.7kHz, 第 3 ホルマント周波数=3.0kHz である。窓関数にはハミング窓を使用した。SPICE シミュレーションによる出力と実際の対数パワースペクトルの比較を図 7 に示す。

高いエネルギーを持つ周波数では、アナログフーリエ変換器の出力はコンピュータシミュレーションによる出力とほぼ同等の推移をみせ、第 1, 第 2 ホルマンとは十分に強調されている。しかし周波数のエネルギーが低いところでは差があまりなく、第 3 ホルマントはわずかに隆起するにとどまっている。

低エネルギーの部分で出力の精度が極端に悪くなる原因は二つ考えられる。第一に C マトリクスにおいて、総コンデンサ容量 C_{all} が増加するに従って出力電圧が減少する。そのため各出力電圧の差が小さくなる。第二に、絶対値回路に入力される電圧差が小さい場合、対数特性が著しく悪化するためである。

6. まとめ

行列演算を行う C マトリクスについて構成, 及び動作原理を説明した。また利用例とし DFT 回路, ラベリング回路について説明をし, DFT 回路については設計を行いシミュレーションを示した。デジタル回路において, 行列演算は繰り返し加算, 乗算を行うためリソースの消費量が増加するが, C マトリクスを用いれば高速な並列演算を行うことができる。実時間での応答が求められている音声認識等の信号処理システムでの利用に有効であると考えている。

7. 謝辞

本研究の一部は半導体理工学センター (STARC) の助成により実施されている。

8. REFERENCES

- [1] C.Mead and L.Conway, "Introduction to VLSI System", Addison-Wesley Publishd Co., 1980
- [2] L.Rabiner and B-H Juang, "Fundamentals of Speech Recognition", AT&T, 1993
- [3] C.Mead(臼井支朗, 米津宏雄訳), "アナログ VLSI と神経システム", 株式会社トッパン, 1993.
- [4] T.Nishitani and K.K.Parhi, "LSI Signal Processing VIII", IEEE Signal Processing Society, 1995
- [5] 宮永喜一, 枅内香次, "自己組織化と教師によるネットワークの高速・高精度学習について", 信学論 (A), vol.J78-A, no.11, pp.1475-1484, Nov. 1995
- [6] Yosahikazu Miyanaga, Shinya Gozen, Noriyuki Ohtsuki, "A Robust Speech Analysis in Speech Recognition", IEEE World Computer Congress (WCC) 2000, vol.1, pp.706-709, August, 2000
- [7] 樺沢正之, 宮永喜一, 吉田則信, "C マトリクスを用いたクラスタリングシステムの設計", 電子情報通信学会技術研究報告, Vol.100 No.573, pp.57-64, 2001 年 1 月
- [8] 菅原崇行, 樺沢正之, 宮永喜一, 吉田則信, "短時間フーリエ変換器のアナログ回路構成について", 電子情報通信学会技術研究報告, Vol.100 No.573, pp.65-70, 2001 年 1 月