

Artificial Neural Network LSI の設計法

森江 隆

九州工業大学大学院生命体工学研究科*¹

積和演算と非線形関数変換により処理される Artificial Neural Network LSI の設計法について述べる。回路実現手法として、デジタル方式、アナログ方式、両者を混載したミックスト・シグナル方式、パルスの時間軸情報を用いるパルス変調方式について、網羅的な開発例を示すのではなく、特徴的な開発事例を示しながら回路・アーキテクチャ設計の課題などを解説する。

1. はじめに

$$y_i = f(x_i) \quad (2)$$

生体の神経細胞の動作を忠実にまねて脳・神経系の機能を実現しようとする、いわゆる“neuromorphic”なアプローチや、スパイクタイミングを利用した情報処理、実数演算が本質的な役割を果たすカオスシステムなど、興味深い系のハードウェア化は他稿に譲ることにし、ここでは、神経細胞の動作を抽象化し、積和演算と、シグモイド関数に代表される非線形関数処理の組み合わせで情報処理を行う、いわゆる Artificial Neural Network (ANN) の集積回路 (LSI) の設計法について議論する。

1980年代後半から、バックプロパゲーション学習を行う多層パーセプトロンやホップフィールドネットワークをはじめとする様々な ANN の LSI 実現の試みが盛んになされてきた。回路の方式は、10年ほど前までは純アナログ方式、純デジタル方式、およびパルス密度変調方式が多かったが、近年、アナログとデジタル双方の利点を共有しようとする、パルス幅変調方式やミックスト・シグナル方式などが提案されている。その結果、指数関数的に性能向上が進む汎用プロセッサに対して優位性を主張できるような専用 LSI も試作されるようになってきた。

このように、ANN LSI 開発の歴史は長いが、本稿では紙面の制限から、引用する文献は概ね 1995 年以降に限る。それ以前のニューラルネットワーク LSI の概要については、他を参照されたい¹⁾。

本稿で取り上げる ANN の演算は、基本的に積和演算 (MAC: Multiply-and-accumulate) と非線形関数演算である。これは、よく知られているように、

$$x_i = \sum_j w_{ij} y_j, \quad (1)$$

と表される。ここで、 x_i 、 y_i 、および w_{ij} は、それぞれニューロンの内部状態、出力値、およびシナプス荷重である。また、 $f(\cdot)$ は出力関数 (または活性化関数) であり、よく用いられるのが、ロジスティック関数などのシグモイド関数だが、非単調関数を用いるモデルもある。

ハードウェア化されている他の情報処理での演算と比較して、このような ANN 演算の最大の特徴は、積和演算の項数が極めて大きいことにある。例えば、画像認識などではこれが 1,000 のオーダーになる。通常の画像処理では、演算されるベクトルの数は画素数に等しくなるので膨大だが、結合は隣接部などに限られており、積和演算の項数はせいぜい数 10 程度であるのと対照的である。そこで、ANN LSI ではこの膨大な積和演算をいかに効率的に行うかが設計のポイントになる。

演算は、(1) メモリ部からニューロン値およびシナプス荷重値を読み出し、(2) 積和演算を行い、(3) 結果を非線形処理して、(4) 演算結果を再びメモリに格納する、という手順をとる。そこで、メモリ部と演算部との結合形態と、デジタルハードウェアで最も演算量の多い乗算をどのように実現するかが課題となる。

以下、第 2 節で各種回路構成の特徴を整理し、第 3, 4 節で上記の課題をどのように解決するかを、公表されている開発例を基に解説する。なお、以下の記述で筆者らの研究内容がやや強調されている点は了承されたい。

2. 各種回路方式の特徴

2.1 アナログ方式

アナログ方式は状態値を電圧、電流、または電荷により連続量で表現する。クロック動作による離散時間系も可能だが、連続時間系も構築できることが大きな特徴である。明らかに、実際の神経細胞の動作はアナログ的であるから、それを模倣するにはアナログ方式が適しており、実際、これまで多くの ANN LSI はアナログ方式で設計されている²⁾。加算・乗算・非線形変換などを少数の素子で実現できるため 1 チップ上に多くの処理ユニットを搭載でき、超並列演算が可能である。これにより、デジタル方式に比べ、高速化が可能になるという利点がある。

しかし、一般にアナログ回路は、ノイズや温度変化による特性ドリフトなど回路や素子の非理想的な性質に影響を受けやすい。演算精度は 5-8 ビット程度に限定される^{*2}。また、制御性が低く、システムに汎用性を持たせることが難しい。制御はやはりデジタル的に行うのがよいが、既存のデジタルシステムと結合するには A/D, D/A 変換が必要で、そのインタフェース部のオーバーヘッドが無視できない、という欠点もある。

純アナログ方式の場合、アナログ量をいかに記憶するか、すなわちアナログメモリの実現が、従来より最大の問題であり、現在に至っても解決されていない。通常は、一時的な記憶に限定して、キャパシタに電荷として記憶させる DRAM 方式が、汎用 CMOS プロセスで実現できるのでよく用いられる。しかし、長期間の記憶が必要な場合はデジタルメモリを併用することになる。この場合、上記と同様に D/A, A/D 変換が必要になることから、そのオーバーヘッドが問題になる。他の記憶手段としては、従来より不揮発性メモリとして使われてきた浮遊ゲート型デバイス、実用化が進む強誘電体メモリ (FeRAM)、近年開発が活発化している磁性体メモリ (MRAM) などがある。浮遊ゲートデバイスを用いてアナログメモリを構成する試みは従来より数多くなされており³⁾⁻⁷⁾、学習機能内蔵チップも試作されているが⁸⁾、より実用を目指した高速・低消費電力チップを実現するためには、浮遊ゲートデバイスよりも、FeRAM⁹⁾ や MRAM での開発が期待される場所である。

ところで、一般的な情報処理を見てみると、その

^{*2}従来より、ANN LSI の演算精度は、デジタル回路で実装したときのビット精度で換算して議論されることが多い。

ハードウェア化の歴史はアナログをデジタルで置き換える努力の繰り返しであった。現在の多くの情報処理ハードウェアはアナログ部を可能なかぎり少なくし、できるだけ多くの処理をデジタル部で行う設計方針をとっている。一方、生体の脳・神経系では、ノイズだらけの環境下で演算精度の低い多数のアナログ素子が協調して、信頼性の高い、頑健な知能情報処理を行っている。この方式を真似ればアナログ方式でも人の脳に迫る機能を持ったハードウェアができるはずである、と考えるのは妥当と思われる。問題は、生体と全く異なったハードウェアシステムであるシリコン LSI でどこまで生体を真似るか、という点である。その意味では、アナログ方式 ANN LSI は基盤技術としての有用性は認められるものの、やや中途半端な実現法であるように思われ、今後の neuromorphic なアプローチに期待がかかるのである。

2.2 デジタル方式

デジタル方式では、各演算素子は 0/1 の 2 値しか取らず、それらを必要精度分だけまとめて 2 進数で情報を表現するため、状態値は離散的である。また、通常はクロック同期動作なので、時間軸上でも離散的になる。状態が 0/1 の組み合わせなので、ノイズに影響されにくい。論理演算と算術演算を単純なスイッチ素子の組み合わせで実現するので、微細素子を大量にチップ上に配置し、大規模な演算を高速に実行することができる。こうして、CMOS デバイスの微細化によって、着実にデジタル LSI の高集積化と高速化が進んできた。しかし、近年、1 チップに数 1,000 万個のトランジスタが集積され、それが 1GHz を越えるスピードで動作するようになると、チップ動作上、消費電力が大きな問題となってきた。応用面でも、モバイル機器の普及やユビキタス社会の到来、環境への配慮などから低消費電力化が最大の関心事になっている。

デジタル方式はアナログ方式と対照的な特質を持つ。制御がしやすく、既存のデジタルシステムとの相性も良い。しかし、後述のように回路規模が大きくなるため、実際の処理に必要なとされるニューロン数よりも少数のプロセッサエレメント (PE) しかチップ上に搭載できず、時分割処理 (仮想ニューロン方式) を採用することになる。

近年のデジタル方式 ANN LSI 開発では、FPGA (Field Programmable Gate Array) を用いた例が多い^{10), 11)}。試作期間が不要で、設計から実動作まで

の期間が極めて短いことから、従来からプロトタイプングによく用いられてきたが、高集積なFPGAが比較的安価に入手できるようになったため、評価に耐えるANN LSI開発が可能になったと考えられる。

ここで、ニューラルネットワーク以外の情報処理用LSIについても触れておこう。近年、超高速の汎用プロセッサだけでなく、「メディアプロセッサ」と呼ばれる画像処理や画像圧縮などに特化したLSIが開発されている。このようなプロセッサはチップ内での並列処理化が盛んになされているため、デジタル方式ANN LSIとの差が小さくなっているように感じられる。

最近の高性能メディアプロセッサは、MPEG-2/4、JPEG2000など演算量の多い画像処理をワンチップで可能としており、処理性能は50GOPSを越えている^{*3}。例えば、集積回路のオリンピックといわれるISSCC (International Solid-State Circuit Conference) で今年発表された、車載用の画像認識用途を狙ったメディアプロセッサの一つは、100MHz動作で128個のPEを搭載し、各PEが1クロック当たり4命令を実行できることから、51.2GOPSの処理性能を備えている¹²⁾。このようなメディアプロセッサが、極めて長い語長(≥128ビット:VLIW)を処理するPEをせいぜい100個程度搭載するのに対し、ANN LSIは比較的短い語長(≤16ビット)のPEを多数(>1,000)搭載するという違いは、処理モデルの違いを反映していると思われる。

2.3 パルス密度変調方式

パルス密度変調(PDM)方式は、スパイクパルスの密度でアナログ量を表現するもので、生体の神経系での情報表現(の一つ)として知られ、以前からこの方式でのLSI化が盛んになされてきた。近年もこの方式で、1,000ニューロン、100万シナプスという大規模なシステムが構築されている¹³⁾。デジタル回路要素を用いながら、擬似的に連続時間ダイナミクスが実現できるところが特長だが、多数のパルスでアナログ量を表現するという事は、回路素子の

^{*3}プロセッサのアーキテクチャや処理方式によって様々な処理性能指標があるが、メディアプロセッサなどではGOPS (Giga Operations Per Second) がよく使われる。しかし、operationの意味を明確にしないと正確な比較ができない。ANN LSIではシナプスの計算が毎秒何回できるか、すなわちCPS(Connections Per Second) がよく使われる。1シナプス演算は1回の積和演算(乗算と加算)なので、1GCPS=2GOPS と言えそうだが、メディアプロセッサでANN演算をしたときにこれが成り立つかどうかは明らかでない(おそらく、多くのメディアプロセッサでは成り立たないであろう)。

オン・オフが頻繁に起こることを意味するので、低消費電力化が課題と思われる。

2.4 パルス幅・位相変調方式

パルス幅変調(PWM)またはパルス位相変調(PPM)方式は、電圧または電流に0/1の2値をもち、時間軸方向にパルス幅またはパルス位相(タイミング)によりアナログ量を表現するパルス信号を用いる回路技術である¹⁴⁾。この方式では、デジタル回路(インバータやゲート回路など)とアナログ回路(キャパシタやソースフォロワ回路など)を組み合わせる回路を構成する。したがって、デジタル回路要素については、Si CMOS技術の微細化に沿ってスケールアップできるので、純アナログ方式に比べるとチップ上の占有面積を小さくできる可能性がある。低電圧動作も可能で、高い制御性を有し、既存のデジタルシステムとのマッチングもよい。ノイズやクロストークなどにも強いという利点もある。一方、値を時間軸方向に展開するので、動作速度が遅いという欠点がある。PWM方式ANN LSIの開発例はこれまでにいくつか報告されている¹⁵⁾⁻¹⁹⁾。

また、PWM/PPM信号とアナログ電圧との変換過程で任意の非線形変換を実現することが可能である²⁰⁾。これを用いれば、離散時間・連続状態の非線形ダイナミカルシステムを高い制御性で実現することができる。この原理を利用して非線形関数を生成し、任意カオス発生器、抵抗ヒューズネットワーク、非線形振動子ネットワークなどをLSI化している²¹⁾⁻²⁴⁾。

2.5 ミックスト・シグナル方式

積和演算部など並列性が高く演算量が多い部分は、回路規模が小さくなる利点を生かしてアナログ方式やパルス変調方式を用い、チップ外部とのインタフェースやメモリ部分はデジタル方式で実現する方式であり、現実的な解として近年開発例が多い²⁵⁾⁻²⁸⁾。

ここで、ミックスト-アナログ/デジタル(Mixed A/D)方式とミックスト-PWM/デジタル(Mixed P/D)方式を比較してみよう。PWM信号はアナログとデジタルの中間の性質を有しているため、A/D変換の中間段階ともいえる。そのため、Mixed P/D方式では、A/D変換が簡単になる。高速クロック動作のデジタルカウンタでパルス幅をクロック数に変換することで、容易にPWM信号をデジタル値に変換でき、アナログ電圧からデジタルに変換する通常のA/D変換器に比べると面積的にはかなり小さくで

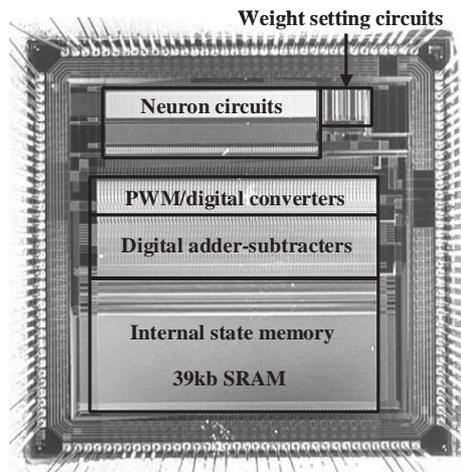


図 1: ミックスド-PWM/デジタル方式による Convolutional Neural Network LSI の設計例

きる。もっとも、パルス-デジタル変換において分解能を上げるために高速のカウンタを用いると消費電力が無視できなくなる。ここでも低消費電力化は大きな課題である。著者らが Mixed P/D 方式により設計した ANN LSI チップの例を図 1 に示す¹⁹⁾。図からわかるように、PWM-デジタル変換部の面積は全体の中でかなり小さい。一方、単純な比較は無理があるが、Mixed A/D 方式では A/D 変換部がチップ上の半分の面積を占める例も報告されている。

3. アーキテクチャ：演算ユニットとメモリの構成

積和演算 (PE) 部と、ニューロン値やシナプス荷重値を記憶するメモリ部の構成はチップアーキテクチャを決定するキーポイントの一つである。

これには、図 2 に示すように、外部メモリを用いる方式 (a) と、チップ上にメモリを内蔵する方式がある。内蔵メモリ方式はまた、PE 部とメモリ部を分ける (集中メモリ) 方式 (b) と、各 PE 内にメモリを内蔵する (分散メモリ^{*4}) 方式 (c) が考えられる。

外部メモリ方式 (図 2(a)) では、外部メモリを増やしさえすれば、任意の規模のネットワークを実行できる。しかし、チップ間の接続ピン数の制限から接続バスが処理のボトルネックになる。そのため、メモリへのアクセス回数の低減と処理の高速化のために、ニューロ演算の特徴を利用してデータ転送量を削減する手法が提案されている。これはシグモイド関数変換時の単調性を利用して、積和演算の結果

^{*4}このアーキテクチャはメモリベースプロセッサなどとも呼ばれる。

が単調に進むように演算の仕方を工夫し、飽和領域に達した段階で計算を止めるという手法である²⁹⁾。さらにデータの差分情報のみを転送することで演算量とデータ転送量の削減を図っている。この手法はまたバックプロパゲーション学習にも適用されている³⁰⁾。

内蔵メモリの場合は、プロセッサとメモリ間に広いバス幅を設定することができ、両者のデータ転送のボトルネックを解消することができる。これは一般にシステム LSI の利点の一つでもある。一例を示すと、ゲーム機プレイステーション 2 に搭載された画像処理プロセッサ (Graphics Synthesizer) では、16 個のプロセッサ (画素処理回路) と 4MB の DRAM が 2,560 ビットのバスで接続されている。

一般にメモリは、一つの大きなセルアレイ構成にした方が周辺制御回路を共有できるため集積度が向上するが、メモリセルには行・列方向でしかアクセスできないという欠点が生じる。また、アクセスが高速になると、バスを駆動するための消費電力が無視できなくなる。一方、分散メモリ方式ではチップ面積の使用効率は低下するものの、各メモリセルに個別にアクセスできるため、超並列動作に適している。なお、ANN LSI およびメディアプロセッサともに、各 PE にキャッシュ的なローカルメモリ (数 kB) を搭載する方式はよく用いられる^{12),31)}。

4. 基本回路構成法

ANN 演算において、積和演算と非線形変換演算を比較すると、処理量は前者が圧倒的に多い。1,000 入力ニューロンでは、1,000 個の積和演算に対し、1 個の非線形変換演算でよい。そこで、積和演算は並列化が欠かせないが、非線形演算は逐次処理でも十分な場合が多い。

さて、積和演算をいかに効率的に実行するのだが、過去にはモデル (精度) を限定して実装した発表も多い。すなわち、シナプス荷重を 3 値 (-1, 0, +1) にしたり、ニューロン出力値を 2 値にする (すなわち、活性化関数をステップ関数にする) ことにより、乗算を省くことができる。これは劇的に回路規模を削減できる。一般のモデルには適用できないが、実用に使えればよいと割り切ることは重要である。

最近では、荷重を 3 値にして乗算を省きながら、オフセットをアナログ的に調整することで処理性能の劣化を防ぐ方法も提案されている³²⁾。また、カーネル関数をガウシアンから 2 値の均一関数にするこ

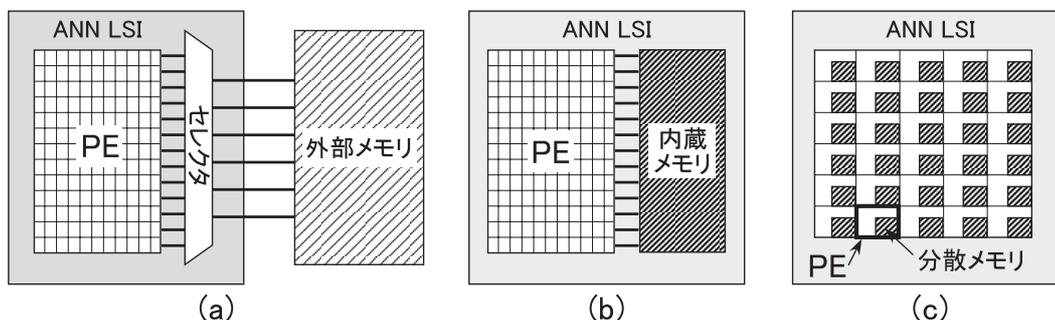


図 2: 処理エレメント (PE) とメモリとの構成法

とで乗算を省き、しきい値処理だけですむ LSI の提案もされている³³⁾。

4.1 積和演算回路

加算については、アナログ的には電流加算または電荷加算、PWM 方式では電荷加算が主に利用される。いずれも、1本のラインに電流または電荷を集めるだけなので、回路は極めて簡単でかつ容易に並列処理ができる。デジタル方式では加算器を用いるが、基本的に2入力しかないので、必要な個数の加算器を用意して並列処理を行うか、時分割で共有することになる。

図3に、デジタル方式、アナログ方式、PWM方式のそれぞれの積和演算回路の典型的な構成例を示す。

デジタル方式の場合(図3(a))、人手でやる筆算と同様に、桁数(2進数なのでビット数)の2乗の回路ユニットが必要で、各回路ユニットで各項の積の部分積を求めていく。デジタル回路ブロックでは乗算器が最も素子数、チップ面積、消費電力を消費する。乗算結果は2つずつ加算器で加算される。アナログ方式やPWM方式のように、1本のラインで加算を行うことができないので、通常は乗算器と加算器のセットで累算器を構成する。

アナログ方式では、電圧または電流方向にアナログ量を表現する。よく知られているように、MOSトランジスタのゲート電圧(V_{gs})-ドレイン電流(I_d)の関係は、飽和領域では、

$$I_d = \beta(V_{gs} - V_{th})^2 \quad (3)$$

と表されるので、両者は比例関係にない。ここで、 β は比例定数(ゲート長、ゲート幅などに依存する)、 V_{th} はしきい値である。そこで、少なくとも一方の入力に比例した結果を得られるように図3(b)に示すような差動対回路を用いる。飽和領域で動作させるために、おおざっぱに言ってゲート長をサブミクロ

ンサイズにできないため、スケーリング(デザインルールに対応してトランジスタサイズを縮小すること)は難しい。乗算結果は電流値で与えられ、1本のラインで加算し、抵抗により電圧に変換され、次のニューロン回路への入力となる。加算が容易に行えることはアナログ方式の最大の特長であるが、定常的な電流は消費電力の増大を招くので、演算時のみ電流を流すなどの制御が必要になる。

PWM方式では、図3(c)に示すように、PWM信号で定電流源をスイッチすることにより、パルス幅と電流値の積の形で電荷を発生させる。積和演算結果はキャパシタの電荷量または端子電圧として得られる。端子電圧は、線形に変化する参照ランプ電圧とコンパレータで比較することにより、PWM信号に変換されて出力される。

デジタル方式とPWM方式での性能比較を表1に示す。ただし、デジタル方式ではCADの自動回路生成に頼っているので、最適化されているわけではない。また、PWM方式では必要とされる精度によっては、電流源の電流量を絞ってさらに低消費電力化も可能である。しかし、通常は乗算による消費電力よりも、キャパシタ電圧をPWM信号に変換するコンパレータなど他の箇所の消費電力が大きいので、LSI全体の消費電力はアーキテクチャに依存する。とはいえ、オーダとして積和演算について比較すると、PWM方式はデジタル方式よりも100倍低速だが、チップ上に占める面積は1/100であり、消費電力は1/10,000である。したがって、100倍の並列化によって低速性を補えば、消費電力を1/100倍にできることになる。なお、アナログ方式の性能も概ねPWM方式と同程度と考えられる。

一般のANNモデルでは正負の加算を実行できないけれども、それには2通りの方式が考えられる。正・負の量をそれぞれ独立に加算して、最後

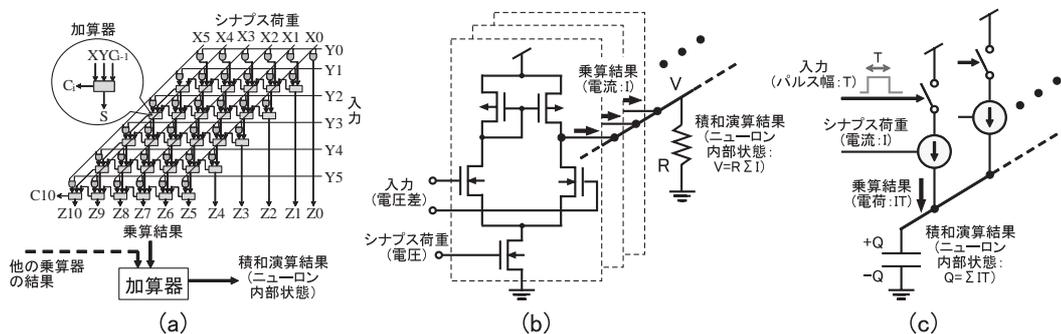


図 3: 各種方式での積和演算回路の例。(a) デジタル方式 (最も簡単なアレイ型並列乗算器を用いた例), (b) アナログ方式 (二象限乗算器 (差動対) を用いた例), (c) PWM 方式 (スイッチト電流源を用いた例)

表 1: デジタル方式と PWM 方式での乗算器の性能比較 (6 ビット精度)。VDEC・ローム 0.35 μm ルール, 電源電圧 3.3V を仮定。デジタルは動作周波数 100MHz で論理合成・自動配置配線を行った結果。PWM 方式は電流値 1 $\mu\text{A}(\text{max})$, PWM 幅 640ns(max), 動作周波数 1MHz で見積もった。

回路方式	デジタル	PWM	デジタル/PWM 比
演算時間 (μs)	0.01	1	1/100
消費エネルギー (pJ/OP)	370	2.1	~170
消費電力 (μW)	37,000	2.1	~17,000
レイアウト面積 (μm^2)	24,000	200	120
トランジスタ数	3,488	3	~1,700

に両者の差を計算する方式¹⁷⁾, および単一のキャパシタに正負のスイッチト電流源を接続し, 電荷を充・放電することにより実現する方式である。後者の方が加算のためのキャパシタが 1 個でよいので省面積になるが, 加減算それぞれの電圧範囲は 1/2 になるので, 精度は低下する。

最後に注目すべき技術として, 最近, ミックスド・シグナル方式での積和演算において, ビットシリアル演算を用いる手法が提案されていることを指摘しておく^{25), 26)}。ANN 演算では要求される演算精度が必ずしも高くないので, 図 3(a) に示したような並列乗算器ではなく, ビットシリアル演算回路を用い, かつアナログ回路も併用することで, 多少の速度低下を犠牲にしても, 回路規模が小さくなることで並列性を高め, 結果として高い処理性能を引き出すことができる。

4.2 非線形関数変換回路

デジタル方式では, メモリのアドレスを入力にしデータを出力とする Look-up table (LUT) がよく用いられる。これは任意の関数を実現でき, 処理時間がメモリアクセスのみなので, 高速である。しかし,

N ビット精度の場合アドレスが 2^N 個必要になることから, 精度を上げるに従い素子数が膨大になる。一方, 関数を数学的に計算する方法もある。これは少ない素子数で実現することも可能だが, 計算を行うために演算時間がかかるという欠点がある。

アナログの場合, 素子の特性を生かした回路を構成し, 非線形変換を実現する。回路はコンパクトだが, 実現できる非線形変換関数が制限される。例えば, シグモイド型関数やその導関数は差動回路で比較的簡単に生成できるが, 関数形状を任意に制御することは難しい。また, 任意の非単調関数を生成することは困難である。

パルス変調方式の場合, PWM 信号を用いた任意非線形変換の原理を図 4 に示す²⁰⁾。これは, 任意の非線形電圧波形を PWM または PPM 信号でサンプル・ホールドすることで実現するパルス-電圧変換である。

PWM 信号から PWM 信号への非線形変換は次のようにして実現する。時間領域で任意の電圧波形 $F(t)$ を有する参照電圧 V_{non} がスイッチ SW を介してキャパシタ C に供給される。スイッチ SW は入力

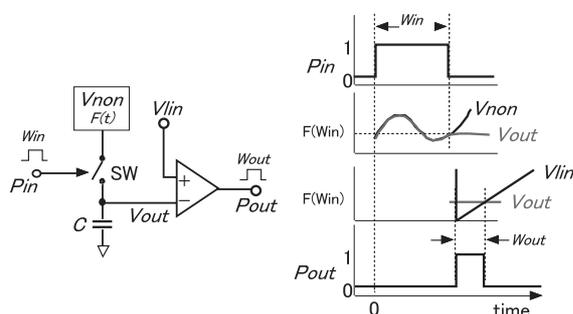


図 4: PWM 信号を用いた任意非線形変換の原理

パルス P_{in} により制御される．ここで， P_{in} の立ち上がりで $V_{non} = F(0)$ とする．スイッチ SW はパルス幅の期間 W_{in} で導通する．入力パルス P_{in} の立ち下がりスイッチが開くと，キャパシタの端子電圧 V_{out} は $F(W_{in})$ に等しくなり，これを保持する．電圧 V_{out} は線形に変化する参照ランプ電圧 V_{lin} とコンパレータで比較され，電圧に比例したパルス幅 W_{out} を有する P_{out} が生成される．この結果， $W_{out} = F(W_{in})$ が成り立ち，非線形参照電圧波形と同形の非線形関数で入出力 PWM 信号のパルス幅が変換される．この回路では，PWM 信号ではなく，むしろその立ち下がり位置する PPM 信号を用いる方が無駄な電流を抑えることができ，低消費電力化には有効である．

この方法が有用なのは，任意のアナログ入出力関数を作り出すのが難しいのに対し，時間の関数としての電圧波形 $F(t)$ を作り出すのが容易なことによる．例えば，電圧波形をアナログ発振回路で作ってもよいし，デジタル回路でルックアップテーブル法や関数生成法によりデジタル波形を作り出し，それを D/A 変換器によりアナログ波形にしてもよい．アナログ波形生成器は多数の PWM 非線形変換回路で共通に使用できるので，システム全体に対する波形生成のオーバーヘッドはわずかなものにできる．

5. おわりに

以上，簡単に ANN LSI 設計手法について概観し，最近の動向を解説した．私見だが，純アナログ方式はパルス変調技術と融合し，neuromorphic なアプローチを取り入れ，アナログでなければできない処理を実現する方向に向かうべきと考える．また，デジタル方式もアナログやパルス変調方式と融合し，ニューラルネットワーク的な意味でより効率的な，つまり既存のデジタル方式のように厳密な数値計算を追求せず，ニューロモデルに基づいて必要な精度の計算

のみを行うようなフレキシブルなハードウェアに向かっていくことが必要と思われる．最終的には，デジタル，アナログ，パルス変調のいずれの方式も取り込んだアーキテクチャと回路設計技術が発展していくことが望まれる．

実用的な観点からいえば，これまでいくつかの ANN LSI が商用化されたが，成功したとは言い難い状況にあった．一般の知的情報処理ではニューラルネットワーク以外の処理も必要で，それらを含めると汎用プロセッサを採用する方が効率だったためと考えられる．また，専用チップはコストの観点から，応用がかなりはつきりしていて，相当規模の出荷量が期待できるものでないとビジネス的に難しかった．

しかし，今日のようなシステム LSI の時代では状況が変わってくると思われる．ANN 処理回路は知能処理 IP^{*5} としての価値は大いにある．特に，昨今の知能ロボット技術の隆盛に伴って，高性能・低消費電力のロボットコントローラ・認識用 LSI，つまり，まさにロボットの「脳」としての機能を果たす LSI の要求が高まっている．そのような LSI を実現するためには，LSI 設計手法の研究と共に，ハードウェア化に適した知能処理モデルの研究が何より重要である．モデル・アルゴリズム，アーキテクチャ，ハードウェア実現それぞれの研究分野の連携が今後より一層望まれる．

参考文献

- [1] 岩田穆, 雨宮好仁 (1995): ニューラルネットワーク LSI, 電子情報通信学会
- [2] Cauwenberghs, G. and Bayoumi, M. A., editors (1999): Learning on silicon: Adaptive VLSI neural systems, Kluwer Academic
- [3] Nakajima, K., Sato, S., Kitaura, T., Murota, J. and Sawada, Y. (1995): Hardware implementation of new analog memory for neural networks, IEICE Trans. Electron., Vol. E78-C, No. 1, pp. 101-105
- [4] Kosaka, H., Shibata, T., Ishii, H. and Ohmi, T. (1995): An excellent weight-updating-

^{*5}IP: Intellectual Property のこと．直訳は知的資産だが，システム LSI の世界では，再利用可能な回路ブロックの設計資産のことを指す．プロセッサ，メモリ，専用処理回路などの IP を売買して，それらを組み合わせて LSI チップ上にシステムを作り上げる手法が確立されつつある．

- linearity EEPROM synapse memory cell for self-learning neuron-MOS neural networks, IEEE Trans. Electron Devices, Vol. 42, pp. 135–143
- [5] Tran, H. V., Blyth, T., Sowards, D., Engh, L., Nataraj, B. S., Dunne, T., Wang, H., Sarin, V., Lam, T., Nazarian, H. and Hu, G. (1996) : A 2.5V 256-level non-volatile analog storage device using EEPROM technology, IEEE Int. Solid-State Circuits Conf. Dig., pp. 270–271
- [6] Kim, K. and Lee, K. (1998) : An 8b resolution $360\mu s$ write time nonvolatile analog memory based on differentially balanced constant-tunneling-current scheme (DBCS), IEEE Int. Solid-State Circuits Conf. Dig., pp. 336–337
- [7] Kinoshita, S., Morie, T., Nagata, M. and Iwata, A. (2001) : A PWM analog memory programming circuit for floating-gate MOSFETs with $75\mu s$ programming time and 11b updating resolution, IEEE J. Solid-State Circuits, Vol. 36, No. 8, pp. 1286–1290
- [8] Morie, T., Fujita, O. and Uchimura, K. (1997) : Self-learning analog neural network LSI with high-resolution non-volatile analog memory and a partially-serial weight-update architecture, IEICE Trans. Electron., Vol. E80-C, No. 7, pp. 990–995
- [9] Kobayashi, D., Shibata, T., Fujimori, Y., Nakamura, T. and Takasu, H. (2002) : A ferroelectric analog associative memory technology employing hetero-gate floating-gate-MOS structure, Symposium on VLSI Technology Digest of Technical Papers, pp. 116–117
- [10] Mihu, I. Z., Brad, R. and Breazu, M. (2001) : Specifications and FPGA implementation of a systolic Hopfield-type associative memory, Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp. 228–233
- [11] Lim, A. B., Rajapakse, J. C. and Omondi, A. R. (2001) : Comparative study of implementing ICNNs on FPGAs, Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp. 177–183
- [12] Kyo, S., Koga, T., Okazaki, S., Uchida, R., Yoshimoto, S. and Kuroda, I. (2003) : A 51.2GOPS scalable video recognition processor for intelligent cruise control based on a linear array of 128 4-way VLIW processing elements, IEEE Int. Solid-State Circuits Conf. Dig., pp. 48–49
- [13] Hirai, Y. and Yasunaga, M. (1996) : A PDM digital neural network system with 1,000 neurons fully interconnected via 1,000,000 6-bit synapses, Proc. Int. Conf. on Neural Information Processing (ICONIP), pp. 1251–1256
- [14] Iwata, A. and Nagata, M. (1996) : A concept of analog-digital merged circuit architecture for future VLSI's, IEICE Trans. Fundamentals., Vol. E79-A, No. 2, pp. 145–157
- [15] El-Masry, E. I., Yang, H. K. and Yakout, M. A. (1997) : Implementations of artificial neural networks using current-mode pulse width modulation technique, IEEE Trans. Neural Networks, Vol. 8, No. 3, pp. 532–548
- [16] Bor, J. C. and Wu, C. Y. (1998) : Realization of the CMOS pulsewidth-modulation (PWM) neural network with on-chip learning, IEEE Trans. Circuits & Syst. II, Vol. 45, No. 1, pp. 96–107
- [17] Morie, T., Funakoshi, J., Nagata, M. and Iwata, A. (1999) : An analog-digital merged neural circuit using pulse width modulation technique, IEICE Trans. Fundamentals., Vol. E82-A, No. 2, pp. 356–363
- [18] Chen, L. and Shi, B. (2000) : CMOS PWM VLSI implementation of neural network, Proc. Int. Joint Conf. on Neural Networks (IJCNN), Vol. III, pp. 485–488
- [19] 是角圭祐, 森江隆, 野村修, 真継優和, 岩田穆 (2002) : アナログ・デジタル融合アーキテクチャによる convolutional network LSI の設計, 日本神経回路学会 第 12 回全国大会 (JNNS2002 とつとり) 論文集 CD, pp. 17–20
- [20] Morie, T., Sakabayashi, S., Ando, H., Nagata, M. and Iwata, A. (1998) : Pulse modulation circuit

- techniques for nonlinear dynamical systems, Proc. Int. Symp. on Nonlinear Theory and its Applications (NOLTA'98), pp. 447–450
- [21] Morie, T., Sakabayashi, S., Nagata, M. and Iwata, A. (2000) : CMOS circuits generating arbitrary chaos by using pulse width modulation techniques, IEEE Trans. Circuits & Syst. I, Vol. 47, No. 11, pp. 1652–1657
- [22] Murakoshi, K., Morie, T., Nagata, M. and Iwata, A. (2000) : An arbitrary chaos generator core circuit using PWM/PPM signals, Asia and South Pacific Design Automation Conf., pp. 23–24
- [23] Morie, T., Miyake, M., Nagata, M. and Iwata, A. (2001) : A 1-D CMOS PWM cellular neural network circuit and resistive-fuse network operation, Ext. Abs. of Int. Conf. on Solid State Devices and Materials (SSDM), pp. 90–91
- [24] Ando, H., Morie, T., Nagata, M. and Iwata, A. (2002) : An image region extraction LSI based on a merged/mixed-signal nonlinear oscillator network circuit, European Solid-State Circuits Conference (ESSCIRC), pp. 703–706
- [25] Szabó, T., Antoni, L., Horváth, G. and Fehér, B. (2000) : A full-parallel digital implementation for pre-trained NNs, Proc. Int. Joint Conf. on Neural Networks (IJCNN), Vol. II, pp. 49–54
- [26] Genov, R. and Cauwenberghs, G. (2001) : Massively parallel inner-product array processor, Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp. 183–188
- [27] Waheed, K. and Salam, F. M. (2001) : A mixed mode self-programming neural system-on-chip for real-time applications, Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp. 195–200
- [28] Schemmel, J., Schürmann, F., Hohmann, S. and Meier, K. (2002) : An integrated mixed-mode neural network architecture for megasynapse ANNs, Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp. 2704–2709
- [29] Aihara, K., Fujita, O. and Uchimura, K. (1995) : A sparse memory-access neural network engine with 96 parallel data-driven processing units, IEEE Int. Solid-State Circuits Conf. Dig., pp. 72–73
- [30] Saito, O., Aihara, K., Fujita, O. and Uchimura, K. (1998) : A 1M synapse self-learning digital neural network chip, IEEE Int. Solid-State Circuits Conf. Dig., pp. 94–95
- [31] Kondo, Y., Komori, S. and Kyuma, K. (1999) : A 1.2GFLOPS neural network processor for large-scale neural network accelerator systems, Cauwenberghs, G. and Bayoumi, M. A., editors, Learning on Silicon, pp. 243–259. Kluwer Academic
- [32] Khan, A. H. (2002) : Multiplier-free feedforward networks, Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp. 2698–2703
- [33] Aibe, N., Yasunaga, M., Yoshihara, I. and Kim, J. H. (2002) : A probabilistic neural network hardware system using a learning-parameter parallel architecture, Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp. 2270–2275