

■10 群 (集積回路) - 9 編 (新概念集積回路)

6 章 ニューラルネットワーク集積回路

(執筆者：森江 隆) [2018 年 12 月 受領]

■概要■

ニューラルネットワークの集積回路化に関しては、1990 年代にホップフィールド(Hopfield)型ニューラルネットワークモデルのアナログ LSI 化に始まって、その発展形であるボルツマンマシンや、バックプロパゲーション (誤差逆伝搬) 学習を行う多層パーセプトロンモデルのアナログ方式及びデジタル方式での集積回路化が活発に行われた。デジタル方式では PC に組み込むアクセラレータボードも市販されているが、その後の汎用コンピュータの急速な高性能化に伴い、ニューラルネットワーク専用チップとしての存在意義が薄れてしまった。1990 年代後半から 2000 年代前半はニューラルネットワーク研究にとって「冬の時代」になり、ハードウェア研究も活性度が低下した。

本章 6-1 及び 6-2 節で述べるニューラルネットワーク LSI は、発火頻度や発火平均に基づきニューロン出力をアナログ値 (またはデジタル値) で表現し、ニューロンの機能として単純な閾値素子 (マッカーピッツ型) またはシグモイド型入出力関数を用いたアナログ型モデルを、単純な積和演算と非線形変換の組合せにより実現する方式である。この方式は主に 1990 年代までに研究された方式であるが、2010 年代になって盛んになった深層学習ベースの人工知能 (AI) を実現する深層ネットワークを実装するハードウェア方式もこの方式を採用することが多い。

一方で、神経細胞内の複雑なダイナミクスを実現するモデルの集積回路化の研究は長く続けられてきたが、2000 年代に入って、このようなダイナミクスの結果発生するスパイクパルスを示的に扱って、その発火タイミングにより、時間軸上にアナログ情報を表現する「スパイクニューロンモデル」が注目され、それを実現する専用 LSI の研究が盛んになった。このモデルは、従来型のアナログ・ニューロンモデルと比べて生体模倣の忠実度が高く、より高度な機能の実現が期待できる。また、このモデルは非同期動作であるため、処理速度においても従来の同期型のネットワークに比べて高速処理も期待できる。2010 年代後半に入ってこのような方式は「ニューロモルフィック」アプローチと呼ばれる。ただし、「ニューロモルフィック」という考え方自体は神経素子模倣という意味で、1980 年代から提案されている。

アナログ情報をスパイクタイミングで表現する手法は、ハードウェア化の観点からも非常に有用である。ニューロンの出力は発火、非発火の 2 値であるため、既存のデジタルシステムとの相性が良く、従来のアナログ型ニューラルハードウェアの制御性に関する欠点を改善する方式としても期待できる。

学習法に関しては、1990 年代に盛んに実装が試みられたヘップ型学習及びその拡張版 (バックプロパゲーション学習を含む) に対して、2000 年代のスパイクニューロンモデルの台頭に併せて、生物学的なニューロンが行っているスパイクタイミング依存型シナプス可塑性 (Spike-Timing Dependent synaptic Plasticity : STDP) を実装する試みが盛んになった。STDP は非同期スパイク間の位相差に応じて荷重値を変化させる学習則であり、学習機能を有するスパイクニューロンモデルに適した学習法である。以上のモデルを実現する回路を 6-3 節で紹

介する。ただし、2010年代から盛んに研究開発されて実用化された深層学習ではバックプロパゲーション学習法が用いられる。

ニューラルネットワーク LSI としては、ほかに網膜機能を模擬するいわゆる「シリコン網膜」の研究が 1980 年代以降続けられてきた。これを 6-4 節で紹介する。

実際の神経細胞及び神経回路ではノイズを有効に利用していることが知られている。これに関するモデル及び回路を 6-5 節で紹介する。

【本章の構成】

本章では、アナログ方式ニューラルネットワーク集積回路 (6-1 節)、デジタル方式ニューラルネットワーク集積回路 (6-2 節)、スパイクングニューラルネットワーク集積回路及び STDP 回路 (6-3 節)、シリコン網膜と視覚処理用ニューラルネットワーク集積回路 (6-4 節)、雑音を利用するニューラルネットワーク集積回路 (6-5 節) について述べる。

次節以降の解説は 2009 年時点の内容である。

■10 群-9 編-6 章

6-1 アナログ方式ニューラルネットワーク集積回路

(執筆: 森江 隆) [2009年9月 受領]

6-1-1 アナログニューラルネットワーク LSI の目的

アナログ回路でニューラルネットワークを構成する目的は大きく分けて 2 つある。一つは、デジタル方式に比べて回路構成要素が少なく済むので、アナログ回路によって高速化・高集積化・低消費電力化を目指そうという立場であり、もう一つはアナログ回路のダイナミクスを積極的に利用して、ダイナミクスを有するニューラルネットワークを高速に実行しようとする立場である。

前者はデジタル回路でもできることをアナログ回路で実行しようとするので、集積度、演算精度などの点からデジタル方式との競合にさらされる。微細化の進展の恩恵が直接的に受けられるデジタル方式に比べて、アナログ方式は高速化・高集積化の点で不利であるが、アナログ量を直接扱えること、及びサブスレッショルド動作による超低消費電力化が可能であることから、生体埋め込みデバイスなどの応用ではアナログ方式に有利な面があると考えられる。また、LSI プロセス技術の微細化が限界に近付きつつあると言われている現在、今後の量子効果デバイスなどによる新機能デバイス技術の恩恵を受けるという点で、アナログ方式の潜在的な能力に期待が寄せられている。

一方、後者の立場は抵抗ネットワークによる連想メモリやシリコン網膜などの専用ニューラルネットを構成するもので、アナログでしかできない回路構成法である。また、カオスなどの本質的にアナログ動作で実現される機能を利用する場合もアナログニューラルネットワーク LSI が適している。

2000 年代に入って、6-3 節で述べるパルス方式の集積回路が盛んに研究されはじめたので、時間及び電圧・電流ドメインで連続量を扱う純粋なアナログ方式のニューラルネットワーク LSI の研究はやや下火になった感があるが、アナログメモリ素子などを用いたアナログ方式のニューラルネットワーク LSI の研究は引き続き続けられている。

6-1-2 アナログニューラルネットワーク LSI の特徴

アナログニューラルネットワーク LSI にはデジタル方式と比べて次のような特徴がある。

- 通常、完全並列アーキテクチャをとるので、ネットワーク動作として高速化が期待でき、フィードバックを有するネットワークに適する。
- 構成素子が少ないので、チップ内に多数のニューロンを集積できる。
- ネットワーク回路のダイナミクスを利用して、高速に最適化問題を解くことができる。
- 新機能デバイスを組み込んで、より高性能化・高集積化が図れる。

一方で、以下のような課題がある。

- (1) 実用的なアナログ記憶素子が未だ開発されていない。
- (2) 演算精度を確保しにくい。

デジタル方式では 1 つのニューロン回路を時分割で多重化して等価的に多数のニューロンを構成できる、いわゆる「仮想ニューロン方式」を採用することができるが、アナログ方式

では時分割化が難しいので、完全並列アーキテクチャ、すなわち「実ニューロン方式」を採用するのが普通である。したがって、高速ではあるが、チップまたはボード上に設けられたニューロンユニット数以上のネットワークを構成できない。ただし、アナログ量を少数のデバイスで表現するので、デジタルに比べて1チップ上に集積できるニューロン数は2桁以上大きいという特長がある。

アナログ LSI では、製造プロセスでの不均一性や温度変動などによるデバイス特性のばらつき、変動が直接回路特性に効いてくるので、外部からネットワークのパラメータをセットしても計算通りの動作をすることは期待できない。したがって、学習機能はこのチップ内のパラメータのばらつきを補償するためにも用いられる。

6-1-3 シナプス荷重の記憶方法

ニューラルネットワーク LSI で重要な要素であるシナプス荷重記憶デバイスについて考えてみよう。理想的なデバイスは以下のような条件を備えたものである。

- (1) アナログ情報を長期間保持可能（不揮発性が望ましい）。
- (2) 書き込み・消去特性が線形（学習機能を内蔵する場合）。
- (3) 微小刻みで記憶更新可（学習機能を内蔵する場合、分解能で10ビット以上）。
- (4) 小占有面積。
- (5) 記憶量の複製が容易。
- (6) 高速読み書き。

提案されている具体的な結合荷重実現方法は以下の通りである。残念ながら、上の条件をすべて満たす方法はまだ現れていない。

- (1) 固定の抵抗値を利用： 最も簡単なシナプス荷重の実現法である。一定の記憶を埋め込んだホップフィールドネットの連想メモリなどが構成できる。初期視覚処理チップはこの構成だが、この場合は記憶デバイスという目的で使っていない。
- (2) デジタルメモリ利用： デジタルメモリと D/A 変換器の組合せはアナログメモリというよりもデジタル技術の流用であるが、条件の(1)から(3)及び(5)を満たす最も現実的な解である。この方法では高精度化が容易であるけれども、それに伴ってチップ上での占有面積が大きくなるという問題がある。
- (3) キャパシタの蓄積電荷の利用： (2)、(4)及び(6)の条件を満たす。回路が単純なため高集積化が容易で、かつ高速なアクセス及び記憶更新が可能なので、多くのアナログ NN LSI で採用されている¹⁾。しかし、記憶保持時間が極めて短いという問題がある。A/D、D/A 変換器を組み合わせるリフレッシュを行う方法²⁾は有効だが、回路が複雑になるという難点がある。キャパシタを用いる場合のもう一つの問題は、高分解能な書き込み制御が難しいという点にある。単純なチャージポンプ回路では5ビット程度の分解能しか得られない³⁾。高分解能を実現するために PWM 信号を用いた例も提案されている⁴⁾。
- (4) デジタルメモリとキャパシタの併用： ニューラルネットとしての処理系ではキャパシタ型アナログメモリを用い、荷重値はデジタル RAM に蓄えて必要に応じてリフレッシュするという手法である⁵⁾⁶⁾。この場合、D/A 変換器などの大面積を必要とする回路を各シナプスに配置する必要はないので、すべての荷重をデジタルメモリで直接実現する方法よりも集積化が容易である。しかし、キャパシタ型の精度の問題は残る。

(5) EEPROM セル利用： EEPROM,フラッシュメモリで使われている不揮発性メモリ素子を利用する。MNOS⁷⁾と浮遊ゲート素子があるが、後者がこれまで頻繁に検討されており⁸⁾、荷重可変型のチップとしても商品化された⁹⁾。この方法は(1)及び(4)の条件を満たしている。長時間(4ビット精度で10年以上)アナログ値を保持することができるうえ、素子の占有面積もかなり小さくできるので、アナログニューラルネットワークチップの記憶デバイスとして有望である。通常構造のデバイスでは、書き替え特性に極めて非線形性が強いが、線形性を高める手法やデバイス構造も提案されている^{10),11)}。書き込み時間は数 μ 程度である。

6-1-4 LSI 開発例

(1) ホップフィールドネットワークチップ

ホップフィールドネットワーク^{12),13)}は1980年代初頭から盛んに研究されてきたが、最初にLSI化が試みられたモデルでもある。アナログ回路での基本的な構成は図1・1に示すように、抵抗アレイより成るシナプス荷重とオペアンプより成るニューロン素子との組合せである。このネットワークのダイナミクスは

$$C_i \frac{du_i}{dt} = \frac{u_i}{R_i} + \sum_{j \neq i} \frac{y_j}{r_{ij}}, \quad y_i = f(u_i) \quad (1 \cdot 1)$$

で表される。

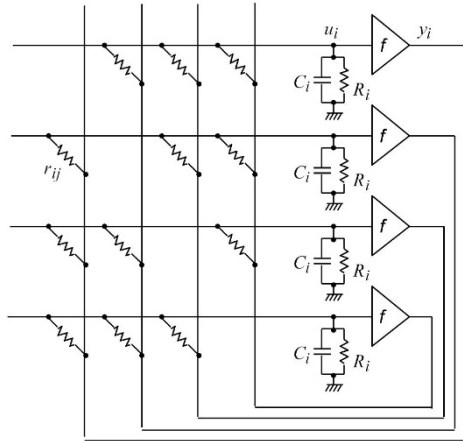


図1・1 ホップフィールドネットワーク回路

1980年代にベル研, CALTEC, ジェット推進研究所(JPL)を中心にチップ化が試みられた¹⁴⁾¹⁹⁾。ホップフィールドネットワークを利用した連想メモリについては、理論的には興味深いものがあり、今日に至るまで様々な検討がなされているが、実用的観点からは応用が直ちに開けないこともあってほとんど利用されていない。そのため、ハードウェア構成に関しては80年代後期の開発例以降大きな進展はみられない。ただし、新デバイスの機能確認としての試作や、単なる連想メモリとしてではなく、脳型連想モデルとして使われる可能性もあり、基本的なモデ

ルであるだけに注目すべきモデルであることは変わらない。

連想メモリと並んでホップフィールドネットワークの重要な応用として、巡回セールスマン (TSP) 問題などの組合せ最適化問題をネットワークに埋め込んで近似解を解く手法が 1980 年代に提案されており、理論的研究がなされている。これがアナログ専用ハードウェアで実時間で実行できれば、様々な有用な応用が期待できる。

(2) ボルツマンマシンチップ

ボルツマンマシンはホップフィールドネットを学習機械として拡張したモデルであり、確率的動作をする。ホップフィールドモデルと同様、ネットワーク動作にダイナミクスを有するために、通常の逐次型コンピュータでの実行は効率が悪く、アナログ LSI による動作が有効である。学習則は基本的にはヘップ則であるから、回路化が比較的簡単にできるという利点もある。そのため、最初に検討された学習機能内蔵型アナログニューラルネットワークチップはボルツマンマシンであった。最初に開発されたボルツマンマシンチップのニューロン部の回路構成を図に示す²⁰⁾。

ボルツマンマシンのハードウェア化の課題は、ニューラルネットワークの確率動作の実現方法である。原則として各ニューロンが互いに相関のない確率動作をしなければならない。アナログ LSI ではニューロンの閾値にランダムノイズを加えるという方法が考えられる。まず、熱雑音などを増幅してノイズ源としたボルツマンマシンチップが検討された²¹⁾、次にリニアフィードバック・シフトレジスタ (LFSR) を用いた疑似ランダムビット列を利用したものが発表された^{22),23)}。

(a) キャパシタ記憶を利用した疑似ボルツマンマシン

ノイズ源の問題にあまり深入りせず、ボルツマンマシンの高集積化・高速化を追求したチップが報告されている^{1),24)}。一連のチップの最終的なものは 400 ニューロン、40 k シナプスを搭載し、2 TCPS、80 GCUPS という動作速度を達成している。このような桁違いの性能を実現できた理由は以下の通りである。

- 乗算を行わない。

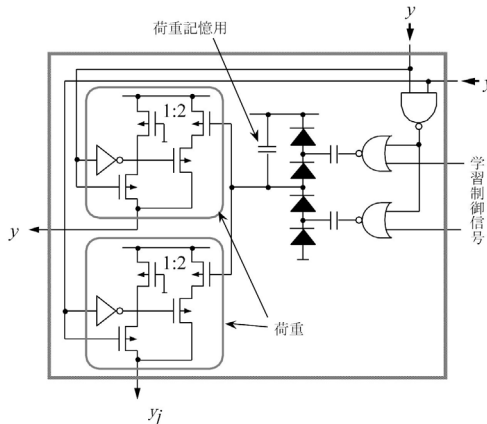


図 1・2 単純化ボルツマンマシンチップのシナプス回路

- 結合荷重は揮発性で低精度のキャパシタ蓄積電荷を利用する。
- 確率動作の非相関性をあきらめる。

まず、ボルツマンマシンではニューロンの状態が 0 か 1 のデジタル値であるから、ニューロンは単なるスイッチでよい。学習も単なるヘップ学習であるから、乗算は不要である。したがって、チップ面積増大・消費電力増大の根源であった乗算器が不要になる。

次に、結合荷重記憶デバイスにはキャパシタを用いている。シナプス部の回路構成の概要を図 1・2 に示す。

荷重変更方法はパルスによる単純なチャージポンピングであり、精度は 5 ビット程度である。

(b) 決定論的ボルツマンマシンチップ

確率動作をするボルツマンマシンを平均場近似理論によりニューロン動作をアナログ化したものが、平均場近似モデル (Mean-Field Theory Model) または決定論的ボルツマンマシン (Deterministic Boltzmann Machine : DBM) と呼ばれるモデルである。ニューロンの状態をアナログ値をとるよう拡張し、ニューロンの入出力関数をシグモイド関数とすればよいので、アナログ回路での実現に適している。決定論的・同期的動作をするので原理的にはネットワークとして高速な処理が期待できる。しかし、回路化に際しては、アナログ乗算器が必要になる。DBM に関しては幾つかのチップ試作例が報告されている^{23),25)27)}。

(3) BP 学習内蔵チップ

階層型 BP ネットワークはダイナミクスを持たないうえ、演算精度が要求されるため、ボルツマンマシンとは対照的にデジタル LSI での実現の方が容易である。そのため、BP ネットについてはアナログ LSI 化の試みはあまり多くない²⁸⁾³¹⁾。

デジタルメモリを用いた BP ネットチップセットの報告例がある³⁰⁾。BP 学習とヘップ学習がチップ上で実行できる。フィードフォワード計算はアナログだが、学習はデジタル回路で行う。

(4) 対比ヘップ学習チップ

BP ネットと決定論的ボルツマンマシン (DBM) はネットワーク構造も学習法も全く異なったモデルであるが、実は対比ヘップ学習 (Contrastive Hebbian Learning) を行うシナプスユニットを用意すれば、ほとんど同一の回路構成で両者のモデルを実行できるチップを作ることができる²⁷⁾。回路アーキテクチャを図 1・3 に示す。対比 BP 学習を行うシナプスユニットを中央にアレイ状に並べる。このユニットは 2 入力 2 出力の回路で、内部にシナプス荷重を保持・修正するシナプス荷重制御 (WPU) 回路を組み込んでいる。一方から入った信号は荷重で重み付けされ、反対方向に出力される。2 つの入出力は全く対称である。荷重は WPU 回路内部で 2 つの入力の積に比例した量だけ変化する。ユニットチップ上で左側に縦に並んでいるのがニューロンユニットであり、ニューロンの非線形入出力関係を実現する回路である。最下列横に並んでいるのが、サブニューロンユニットである。これは BP ネットを実現するときは逆伝播誤差を生成するために使われ、DBM を構成するときはニューロンの非線形入出力関係を実現する回路となる。その切替えは端子 MO で行われる。

6-1-5 まとめ

今後も様々なニューラルネットワークモデルのハードウェア化の試みが望まれるが、アナログ回路の利点を活かしたアプローチが重要である。また、高集積化・低消費電力化へ向

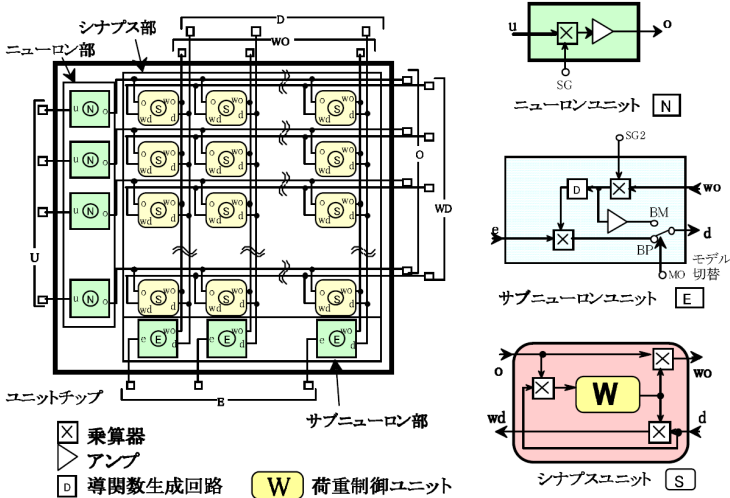


図 1・3 対比ヘップ学習チップのアーキテクチャ

けたデバイス・回路レベルでの研究課題も多く残されている。ただ、直ちに実用的な特定モデルのチップを作るというよりも、将来新たに有効なモデルが見出されたときに直ちにハードウェア化ができるように、アナログメモリをはじめとする要素デバイスや回路技術を検討していくことが重要である。

■参考文献

- 1) Y. Arima, K. Mashiko, K. Okada, T. Yamada, A. Maeda, H. Notani, H. Kondoh, and S. Kayano : "A 336-Neuron 28k-Synapse Self-Learning Neural Network Chip with Branch-Neuron-Unit Architecture," IEEE J. Solid-State Circuits, vol.26, pp.1637-1644, 1991.
- 2) B. Linares-Barranco, E.Sánchez-Sinencio, A. Rodríguez-Vázquez, and J.L. Huertas : "Modular Analog Continuous-Time VLSI Neural Networks with On Chip Hebbian Learning and Analog Storage," in IEEE Proc. of Int. Symp. Circuits and Systems (ISCAS), pp.1533-1536, 1992.
- 3) 村崎充弘, 他 : "336 ニューロン, 28 k シナプス搭載学習機能付きニューラルネットチップと Branch-Neuron-Unit アーキテクチャ," 信学技報, ICD91-108, 1991.
- 4) T. Morie, O. Fujita, and Y. Amemiya : "Analog VLSI Implementation of Adaptive Algorithms by an Extended Hebbian Synapse Circuit," IEICE Trans. Electron., vol.E75-C, no.3, pp.303-311, 1992.
- 5) T. Morishita, Y. Tamura, T. Otsuki, and G. Kano : "A BiCMOS Analog Neural Network with Dynamically Updated Weights," IEICE Trans. Electron., vol.E75-C, pp.297-302, 1992.
- 6) S. Eberhardt, T. Duong, and A. Thakoor : "Design of Parallel Hardware Neural Network Systems from Custom Analog VLSI 'Building Block' Chips," in Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp.11-183-190, 1989.
- 7) J.P. Sage, R.S. Withers, and K.E. Thompson : "MNOS/CCD Circuits for Neural Network Implementations," in IEEE Proc. of Int. Symp. Circuits and Systems (ISCAS), pp.1207-1209, 1989.
- 8) A. Kramer, V. Hu, C. K. Sin, B. Gupta, R. Chu, and P. K. Ko : "EEPROM Device as a Reconfigurable Analog Element for Neural Networks," in IEEE Int. Electron Devices Meeting, pp.259-262, 1989.
- 9) M. Holler, S. Tam, H. Castro, and R. Benson : "An Electrically Trainable Artificial Neural Network (ETANN) with

- 10240 “Floating Gate” Synapses,” in Proc. Int. Joint Conf. on Neural Networks (IJCNN), pp.II-191-196, June 1989.
- 10) O. Fujita, Y. Amemiya, and A. Iwata : “Characteristics of Floating Gate Device as Analogue Memory for Neural Networks,” *Electron. Lett.*, vol.27, pp.924-926, 1991.
 - 11) T. Morie, O. Fujita, and K. Uchimura : “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, 1997.
 - 12) J.J. Hopfield : “Neural Networks and Physical Systems with Emergent Collective Computational Abilities,” *Proc. Natl. Acad. Sci. USA*, vol.79, pp.2554-2558, 1982.
 - 13) J.J. Hopfield : “Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons,” *Proc. Natl. Acad. Sci. USA*, vol. 81, pp. 3088-3092, 1984.
 - 14) L.D. Jackel, R.E. Howard, H.P. Graf, B. Straughn, and J.S. Denker : “Artificial Neural Networks for Computing,” *J. Vac. Sci. Technol. B*, vol.4, pp.61-63, 1986.
 - 15) H.P. Graf, L.D. Jackel, R.E. Howard, B. Straughn, J.S. Denker, W. Hubbard, D.M. Tennant, and D. Schwartz : “VLSI Implementation of a Neural Network Memory with Several Hundreds of Neurons,” in J. S. Denker, editor, *AIP Conf. Proc. 151 Neural Networks for Computing*, pp.182-187, 1986.
 - 16) W. Hubbard, D. Schwartz, J. Denker, H.P. Graf, R. Howard, L. Jackel, B. Straughn, and D. Tennant : “Electronic Neural Networks,” in J. S. Denker, editor, *AIP Conf. Proc. 151 Neural Networks for Computing*, pp.227-234, 1986.
 - 17) M.A. Sivilotti, M.R. Emerling, and C.A. Mead : “VLSI Architecture for Implementation of Neural Networks,” in J. S. Denker, editor, *AIP Conf. Proc. 151 Neural Networks for Computing*, pp.408-413, 1986.
 - 18) A.P. Thakoor, A. Moopenn, J. Lambe, and S.K. Khanna : “Electronic Hardware Implementations of Neural Networks,” *Appl. Opt.*, vol.26, pp.5085-5092, 1987.
 - 19) R.E. Howard, D.B. Schwartz, J.S. Denker, R.W. Epworth, H.P. Graf, W.E. Hubbard, L.D. Jackel, B.L. Straughn, and D.M. Tennant : “An Associative Memory Based on an Electronic Neural Network Architecture,” *IEEE Trans. Electron Devices*, vol.ED-34, no.7, pp.1553-1555, 1987.
 - 20) J. Alspector, B. Gupta, and R.B. Allen : “Performance of a Stochastic Learning Microchip,” in D.S. Touretzky, editor, *Advances in Neural Information Processing Systems 1*, pp.748-760, Morgan Kaufmann, 1989.
 - 21) J. Alspector, R.B. Allen, V. Hu, and S. Satyanarayana : “Stochastic Learning Networks and their Electronic Implementation,” in D. Z. Anderson, editor, *Neural Information Processing Systems*, pp.9-21, 1987.
 - 22) J. Alspector, J.W. Gannet, S. Harber, M.B. Parker, and R. Chu : “A VLSI-Efficient Technique for Generating Multiple Uncorrelated Noise Sources and Its Application to Stochastic Neural Networks,” *IEEE Trans. Circuits & Syst.*, vol.38, pp.109-123, 1991.
 - 23) J. Alspector, A. Jayakumar, and S. Luna : “Experimental Evaluation of Learning in a Neural Microsystem,” in J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems 4*, pp.871-878. Morgan Kaufmann, 1992.
 - 24) Y. Arima, M. Murasaki, T. Yamada, A. Maeda, and H. Shinohara : “A Refreshable Analog VLSI Neural Network Chip with 400 Neurons and 40K Synapses,” *IEEE J. Solid-State Circuits*, vol.27, pp.1854-1861, 1992.
 - 25) C.R. Schneider and H.C. Card : “Analog CMOS Deterministic Boltzmann Circuits,” *IEEE J. Solid-State Circuits*, vol.28, pp.907-914, 1993.
 - 26) H. Won, S. Sato, K. Nakajima, and Y. Sawada : “A Pulsating Neural Network,” in *Ext. Abs. of Int. Conf. on Solid State Devices and Materials (SSDM)*, pp.379-381, 1994.
 - 27) T. Morie : “Analog VLSI Implementation of Self-learning Neural Networks,” in G. Cauwenberghs and M.A. Bayoumi, editors, *Learning on Silicon*, chapter 10, pp.213-242, Kluwer Academic, Norwell, MA, 1999.
 - 28) B. Furman and A.A. Abidi : “An Analog CMOS Backward Error-Propagation LSI,” in *Proc. 22nd Asilomar Conf. on Signals, Systems and Computers*, pp.645-648, 1990.
 - 29) F.M.A. Salam and M. Choi : “An All-MOS Analog Feedforward Neural Circuit with Learning,” in *IEEE Proc. of Int. Symp. Circuits and Systems (ISCAS)*, pp.2508-2511, 1990.
 - 30) T. Shima, T. Kimura, Y. Kamatani, T. Itakura, Y. Fujita, and T. Iida : “Neuro Chips with On-chip Back-Propagation and/or Hebbian Learning,” *IEEE J. Solid-State Circuits*, vol.27, no.12, pp.1868-1876, 1992.

- 31) T. Morie and Y. Amemiya : “An All-analog Expandable Neural Network LSI with On-chip Backpropagation Learning,” IEEE J. Solid-State Circuits, vol.29, no.9, pp.1086-1093, 1994.

■10 群-9 編-6 章

6-2 デジタル方式ニューラルネットワーク集積回路

(執筆者：肥川宏臣) [2009年9月 受領]

6-2-1 デジタル方式の概要

(1) プラットフォーム

ANN のハードウェア実装には、Digital Signal Processing (DSP), Application Specific Integrated Circuit (ASIC) や Field Programmable Gate Array (FPGA) を用いた実装方法がある。DSP を用いた場合、ANN の処理はソフトウェアによる逐次処理となるので、ANN に内在する並列性を利用することができない。ASIC による実装は高速で大容量のネットワークの実装が可能であるが、量産向けのプラットフォームである。FPGA は再構成可能なデバイスで、近年その大容量化、低コスト化が進み、多くの研究者がハードウェア ANN のプラットフォームとして利用している。

(2) 実装における問題点

ハードウェア実装のポイントは、その性能を維持しつつ、如何に大規模な ANN を実装するかという点にある。例えば、1 層当たり n 個のニューロンを持つネットワークで完全結合を行った場合、その複雑さは $O(n^2)$ となる。そのため、ANN のデジタルハードウェアに関する研究は、(1)ニューロンなどの構成要素のハードウェア化、(2)ネットワークレベルの実装、に分類することができる。(1)については、主に演算の単純化によるハードウェアの縮小を目的としている。演算の単純化により学習性能が悪化してしまうが、単純化と性能劣化のバランスが問題となる。(2)については、ANN 全体の効率的なハードウェア化を目的とするが、大規模ネットワークの実装ではシリアル処理を導入することになるので、回路規模(実装できるプロセス数)と並列性の改善が課題となる。

6-2-2 構成要素の実装

ANN ハードウェア実装には、多数のニューロンとニューロン間を接続するシナプス結合、可変荷重回路が主な構成要素となる。

まず問題となるのが、シナプス結合で使用される乗算回路である。ANN では、多数の乗算が使われるうえに、乗算回路はサイズが比較的大きいため、乗算回路はネットワーク全体の大きさに大きく影響する。次にニューロンで問題となるのが、非線形活性化関数の実現である。最も簡単な実装方法は関数値を Look Up Table (LUT) に格納することであるが、ニューロン数が多くなると LUT として使用するメモリの容量が大きくなってしまふ。また、ANN の一種である自己組織化マップ (Self-Organizing Map : SOM) の実装では、入力ベクトルと各ニューロンが持つ重みベクトルとのベクトル間距離の計算や勝者ニューロンを決めるための Winner-Takes-All 回路、近傍関数の実装が問題となる。

(1) 数値表現による回路の小型化

通常の数値演算を ANN の演算に使った場合、必要とされる数値演算精度と演算速度、ハードウェアコストとのバランスを考慮して演算方式を決める必要がある。例えば、浮動小数点演算と固定小数点演算で同じ学習能力を持つ ANN を FPGA に実装した場合、固定小数点演算を

用いた方が浮動小数点演算による回路に比べほぼ半分の大きさになる．回路の大きさだけでなく固定小数点演算の方が演算処理が早く，ANN 自体の学習速度も速くなる²⁾．

更に回路規模を小さくするために大幅な制限を課した ANN も提案されている．まず，活性化関数を 2 値化³⁾ や 3 値化⁴⁾ することが，回路縮小に有効である．前者ではニューロンの出力値は +1, -1, 後者の場合 +1, 0, -1 しかとらないので，関数値を格納するメモリ (LUT) が不要になるとともに，ニューロン出力とシナプス荷重との乗算回路を加減算回路に置き換えることができる．また，重み係数を 2 のべき乗に制限することでシナプス乗算器を省略したのも提案されている⁵⁾．この場合，乗算はシフトレジスタで実現できる．

(2) 自己組織化マップの構成要素

自己組織化マップの実装では，従来用いられてきたユークリッド距離をマンハッタン距離に置き換えることで乗算や平方根の計算を省くことができるため大幅な回路の簡単化が可能となる⁶⁾．また，近傍関数についても従来のガウシアン関数を矩形関数に置き換えたり係数を 2 のべき乗の値をとる単調減少関数とすることで，学習機能を損なわず回路規模の縮小を行うことができる^{7),8)}．

(3) パルスニューラルネットワーク

有効なハードウェア実装の一つに，パルスの振幅や，幅，密度，周波数で信号を符号化するパルス列信号を用いたアーキテクチャがある⁹⁾．このアーキテクチャでは，信号をランダムパルスの発生確率で表し，確率的演算 (Stochastic Computing) を用いる^{10),11)}．シナプス荷重やニューロンの出力値は一定期間におけるランダムパルスの発生数，平均パルス数により表される．AND や OR といった簡単な論理演算で乗算と加算の近似ができるため，大幅な回路規模の縮小が可能である．また，加算の場合，加算される 2 つの信号に含まれるパルス数が少ないと出力は入力における個々のパルス計数の単純合計に等しいが，パルスが頻繁 (より高い信号レベル) になると，パルスが互いにマスクする機会が多くなり，出力パルス数は単純合計より少なくなり，徐々に最大値に飽和することになる．この性質 (Statistical Saturation) から，確率加算は非線形加算となるため，ニューロンの非線形活性化関数が容易に実現される¹²⁾．

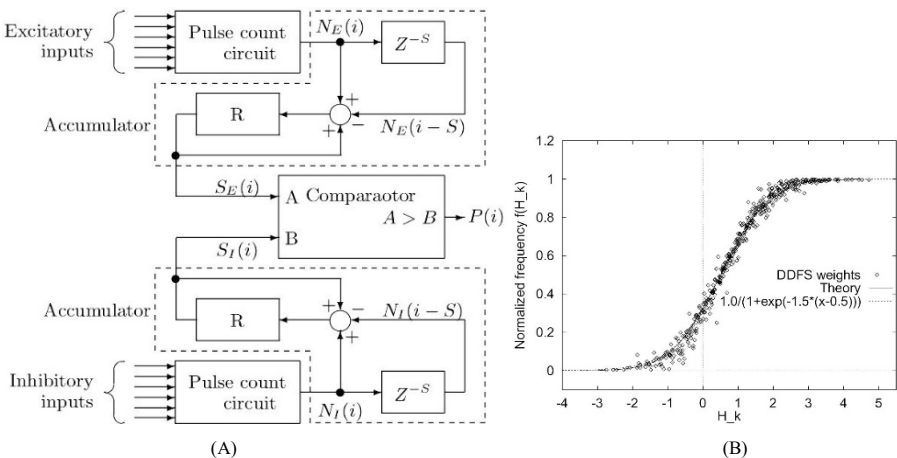


図 2・1 多数決ニューロン¹³⁾. (A)構成, (B)活性化関数

しかし、得られる活性化関数の変形が困難な点や数値演算に比べると演算精度が良くないといった欠点がある。これらの問題を解決するために、加算に多数決回路を用いたニューロンとシナプス乗算に Direct Digital Frequency Synthesizer (DDFS) を用いた ANN が提案され^{13),14)}、FPGA へ実装されている。従来の確率演算を用いたものに比べ、安定した動作のニューロンとなるとともに、簡単に活性化関数の形を調整できる。図 2・1 に多数決ニューロンの構成とその活性化関数を示す。

(4) 学習機能

ANN の重要な性質として、学習機能が挙げられる。多くのハードウェア ANN は Back Propagation (BP) 学習をハードウェアとして実装しているが、数値表現の工夫やパルス信号を用いることで回路の単純化が行われている。しかし、単純化により学習成功率が低下してしまう問題があり、その解決方法が研究課題となっている。一例を挙げると、学習性能の改善に教師信号へのランダムノイズ付加が有効であることが知られているが、活性化関数の微分関数にランダムノイズを加えることでも学習性能を改善することができる⁴⁾。

その他の学習アルゴリズムとして、同時摂動 (Simultaneous Perturbation) を使った学習方法¹⁵⁾ や主成分分析 (Principal Component Analysis : PCA) 学習のハードウェア化が行われている¹⁶⁾。

6-2-3 ネットワークレベルの実装

ANN 全体をハードウェアとして実装するためには、(1)ニューロン、シナプス乗算を行う専用回路を組み合わせた完全並列処理システムとする、(2)様々な演算回路で構成されるプロセッサを多数用いたアレイプロセッサに ANN の処理をマッピングする方法がある。

(1) 完全並列ネットワーク

並列性の高いハードウェア ANN は、開発したニューロン、シナプス回路の動作確認用のものが多いため、比較的小規模ネットワークであることが多い。しかし、7 ビットの重みを持つ 100 万個のシナプスにより結合された 1000 個のニューロンが並列に動作するシステムも開発されている¹⁷⁾。このシステムでは、信号レベルはパルス密度変調された信号で表され、パルス信号の特徴を生かして演算の単純化を行っている。そして、このシステムに Winner-takes-all ネットワークを実装し動作実験を行い、 360×10^9 Connection Per Second (CPS) での動作が可能であることが示されている。

(2) アレイプロセッサ実装

アレイプロセッサは、ANN だけでなくデジタル信号処理アルゴリズムなど大量のデータを扱う並列性の高いアルゴリズムの実装にも用いられる手法で、複数のプロセッサ (Processing Element : PE と呼ばれる) による並列処理を行う。PE には ANN 処理に特化したものを使用するが、より汎用性の高いプロセッサを用いたものもある。すべてのプロセッサが同じ処理を行う場合、Single Instruction Multiple Data (SIMD) アレイと呼ばれる。SOM をプロセッサアレイに実装したもの¹⁸⁾ やフィードフォワード型 ANN をレイヤごとに分けて処理するシステム¹⁹⁾ など多くのアレイシステムが提案されている。これらのシステムでは、ニューロン数が PE の数より多いネットワークについても各 PE に複数のニューロンの処理を順次担当させることで実行可能だが、当然、ANN 全体の完全並列処理はできなくなるので実行速度が低下してしまう。

(3) シストリックアレイ実装

アレイプロセッサでは PE 間の通信が問題となる。PE 数に比例して並列性が高くなり大量のデータが処理できるようになるが、同時に大量のデータを PE 間で送受信しなければならないことになる。しかし、PE 間通信に十分な通信帯域幅を確保するのも PE 数に比例して困難になる。シストリックアレイは、隣接するセル間でデータ転送することで効率的に大量のデータ処理を行う並列処理システムで、この PE 間通信の問題を解決している。シストリックアレイは、ANN プラットフォームに適している^{20),21)}。

6-2-4 応用と展望

ハードウェア化されたニューラルネットワークは大量のデータをリアルタイムに処理できる。そのため、画像処理やロボット制御、各種認識システムへの応用が進められている。VLSI 技術の発展に伴い、より実践的なシステムの開発が期待される。

■参考文献

- 1) A. Omondi and J. Rajapakse : “FPGA Implementations of Neural Networks,” Springer, 2006.
- 2) A.W. Savich, M. Moussa, S. Areibi : “The Impact of Arithmetic Representation on Implementing MLP-BP on FPGAs: A Study,” IEEE Trans. on Neural Networks, vol.18, no.1, pp.240-252, 2007.
- 3) J.H. Kim and S.K. Park : “The geometrical learning of binary neural networks,” IEEE Trans. on Neural Networks, vol.6, no.1, pp.237-247, 1995.
- 4) 肥川宏臣 : “ハードウェア化に適した学習機能付き 3 値多層ニューラルネットワーク,” 電子情報通信学会論文誌, vol.J81-D-II, no.12, pp.2811-2818, 1998.
- 5) M. Marchesi, G. Orlandi, F. Piazza, and A. Uncini : “Fast neural networks without multipliers,” IEEE Trans. on Neural Networks, vol.4, no.1, Jan. 1993.
- 6) M.S. Melton, T. Phan, D.S. Reeves, D.E. V.den Bout : “The TinMANN VLSI Chip,” IEEE Trans. on Neural Networks, vol.3, no.3, pp.375-384, 1992.
- 7) M. Pormann, U. Witkowski and U. Ruckert : “Implementation of Self-Organizing Feature Maps in Reconfigurable Hardware,” FPGA implementations of neural networks, Springer, 2006.
- 8) H. Hikawa : “FPGA implementation of self organizing map with digital phase locked loops,” Neural Networks, vol.18, no.5-6, pp.514-522, 2005.
- 9) W. Maass and C.M. Bishop, Eds. : “Pulsed neural networks,” Cambridge, MA, USA: MIT Press, 1999.
- 10) D.E. Van Den Bout et al. : “A Digital Architecture Employing Stochasticism for the Simulation of Hopfield Neural Nets,” IEEE Trans. on Circuits and Systems, vol.36, no.5, pp.732-738, May 1989.
- 11) S. Sato, K. Nemoto, S. Akimoto, M. Kinjo, and K. Nakajima : “Implementation of a New neurochip Using Stochastic Logic,” IEEE Trans. on Neural Networks, vol.14, no.5, pp.1122-1127, 2003.
- 12) Leonardo M. Reyneri : “A performance analysis of pulse stream neural and fuzzy computing systems,” IEEE Trans. on Circuits and Systems, vol.42, no.10, pp.642-660, Oct. 1995.
- 13) 肥川宏臣 : “周波数変調パルスと多数決ニューロンによる学習機能付き多層ニューラルネットワーク,” 電子情報通信学会論文誌, vol.J82-A, no.7, pp.1005-1015, 1999.
- 14) H. Hikawa : “A New Digital pulse-Mode Neuron with adjustable Activation Function,” IEEE Trans. on Neural Networks, vol.14, no.1, pp.236-242, 2003.
- 15) Y. Maeda, T. Tada : “FPGA Implementation of a Pulse Density Neural Network With Learning Ability Using Simultaneous Perturbation,” IEEE Trans. on Neural Networks, vol.14, no.3, pp.688-695, 2003.
- 16) Y. Hirai and K. Nishizawa : “Hardware implementation of a PCA learning by an asynchronous PDM digital circuit,” Proc. of IEEE IJCNN 2000, Jul. 2000.
- 17) Y. Hirai : “A 1,000-neuron system with one million 7-bit physical interconnections,” Proc. of the 1997 conference on Advances in neural information processing systems 10, pp.705-711, 1998.
- 18) D.C. Hendry, A. Duncan, and N Lightowler : “IP Core Implementation of a Self-Organizing Neural Network,”

IEEE Trans. on Neural Networks, vol.14, no.5, pp.1085-1096, 2003.

- 19) S. Himavathi, et al. : “Feedforward Neural Network Implementation in FPGA Using Layer Multiplexing for Effective Resource Utilization,” IEEE Trans. on Neural networks, vol.18, no.3, pp.880-888, 2007.
- 20) S.Y. Kung : “Digital Neural Networks,” Printice Hall, 1993.
- 21) C. Lehmann, M. Viredaz, and F. Blayo : “A generic systolic array building block for neural networks with onchip learning,” IEEE Trans. Neural network, vol.4, no.3, pp.400-407, 1993.

■10 群-9 編-6 章

6-3 スパイキングニューラルネットワーク回路の概要

(執筆著：森江 隆) [2009年9月 受領]

6-3-1 スパイキングニューラルネットワーク回路の概要

広義のスパイキングニューロンモデルは、スパイクパルスで情報を表現するニューロンモデルを指す¹⁾²⁾。その最も簡単なものは積分発火型 (Integrate-and-Fire) である。複数のニューロンから非同期的に出力されたスパイク信号を、それぞれシナプスを介してあるニューロンが受け取ると、ニューロンの内部にシナプス荷重に比例した大きさのシナプス後電位 (PSP : Post-Synaptic Potential) が生成される。これらを時空間的に積分した総和が閾値に達した時点でスパイクパルスを出力するモデルである。シナプス結合には興奮性と抑制性が存在し、結合荷重の符号で表現される。ニューロンがスパイクを出力することを神経科学の用語を用いて「発火 (Firing)」と呼ぶ。ニューロンは発火後、一定期間、入力にตอบสนองしない「不応期」に入る。生成されたスパイクは伝播遅延時間を経て、他のニューロンへ伝達される。

ほかに、スパイクレスポンスモデルは特定形状の PSP を時空間積分するものであり、共鳴発火型 (Resonate-and-Fire) モデルは周期的に状態が変化し、所定閾値位相で発火するモデルである。いずれも、生体ニューロンの精密なモデルである Hodgkin-Huxley モデルを単純化し、取り扱いを容易にしたものである。

ハードウェア実装の観点からは、通常のアナログ・ニューロンモデルが出力をアナログ値で表現する方式であるのに対し、PWM (Pulse-Width Modulation), PDM (Pulse-Density Modulation), PPM (Pulse-Phase Modulation) などのパルス変調方式により情報を表現するモデルに対応し、デジタル方式との相性が良く、集積回路化に適した方式と言える。

この方式では、チップ間伝送に関してスパイクパルスで効率良く実現する Address-Event Representation (AER) と呼ばれる方式が提案されており³⁾、しばしば用いられる。これは、ネットワークから出力されるスパイクパルスを、同時発火しないように時間領域を微小な間隔に区切り、各区切りの中にスパイクが存在する場合、そのニューロンの番号 (アドレス) を伝送する表現形態である。

また、スパイクタイミングを利用してシナプス荷重を変更する学習則として「スパイクタイミング依存シナプス可塑性 (STDP : Spike-Timing Dependent Synaptic Plasticity)」がよく知られており、それを実現する回路も多く発表されている⁴⁾。

ここでは、積分発火型ニューロン回路と、STDP 学習則を実現する回路を紹介する⁵⁾。

6-3-2 積分発火型スパイキングニューロン回路

積分発火型のニューロンモデルを実現する CMOS 回路の一例を図 3・1 に示す。また、回路のタイミングチャートを図 3・2 に示す。他のニューロンからのスパイク i_i が入力されると PSP の制御信号 (psp_cont) が生成される。 psp_cont が high の期間、トランスコンダクタンスアンプ (OTA) が駆動され、ニューロン部のキャパシタ C を充電、または放電することにより PSP を生成する。したがって、PSP の時空間加算はこのキャパシタ上で行われる。OTA の出力電流は $V_{\omega_{ij}} - V_{ref}$ に比例し、 $V_{\omega_{ij}} - V_{ref} > 0$ の場合は興奮性、 $V_{\omega_{ij}} - V_{ref} < 0$ の場合は抑制性となる (モ

デルとの対応を考えるとシナプスの荷重値は $V_{\omega_{ij}} - V_{ref}$ を指すが、今後は便宜上 $V_{\omega_{ij}}$ を荷重値と呼ぶこととする)。

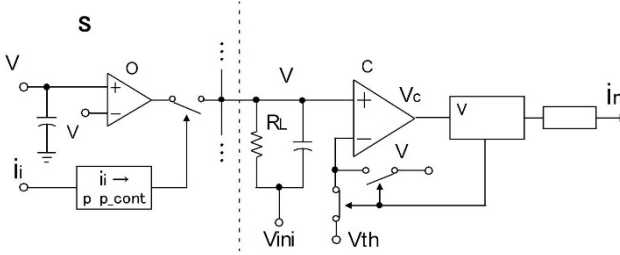


図 3・1 積分発火型スパイクニューロン回路

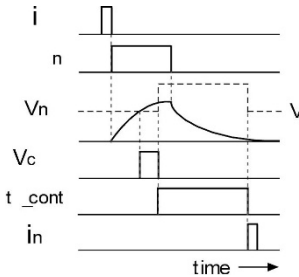


図 3・2 積分発火型スパイクニューロン回路のタイミングチャート

ニューロン部において、内部電位 V_n はキャパシタ C と並列に接続されたリーク抵抗 R_L によって静止電位 V_{ini} に落ち着く。 V_n はコンパレータ (CMP) によって閾値 V_{th} と比較され、 $V_n > V_{th}$ となると、 V_{CMP} が high となり、後段の回路によってスパイクパルスが生成される。同時に閾値制御信号 (th_con) によって一定期間閾値が上昇し、不応期が形成される。本回路において不応期と伝播遅延時間は等しく、不応期が終わると他のニューロンにスパイクが伝播される。

積分発火型のニューロン回路を対称結合を有するホップフィールド型のネットワークに適用し、連想メモリを構成した例も報告されている。

6-3-3 STDP 回路

STDP 機能を実現する回路の基本原理解を 図 3・3 に示す。この回路は 2 個の入力スパイクで動作

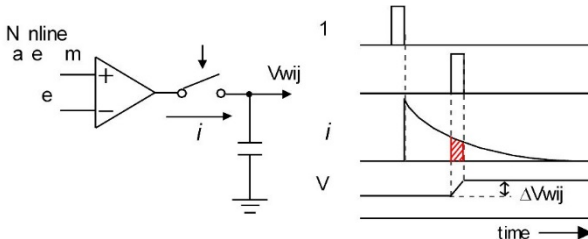


図 3・3 STDP 回路の基本原理解

し、最初のスパイクを非線形波形を生成するトリガ、もう一方のスパイクをサンプリングパルスとして扱う。回路自身が非線形波形を生成するため、非同期的動作が可能となる。

(1) 対称型 STDP 回路

図 3・4 に対称型の STDP 回路を示す。本回路はスパイク検出部、荷重更新部から構成される。回路のタイミングチャートを図 3・6 (a) に示す。スパイク検出部において、T-FF (Toggle Flip-Flop) は *pre* と *post* のスパイクによって二度状態が反転する。状態値の反転は後段の遅延反転回路 (D & I) と NOR ゲートによって検出される。結果として、早く到達したスパイクは荷重更新部の *in1* へ、遅く到達したものは遅延時間 t_d を経て *in2* へ入力される。荷重更新部は *pre* と *post* の時間間隔のみに依存し、荷重値 $V_{w_{ij}}$ を更新する。一方で、*pre* と *post* が同時に入力された場合は T-FF の状態値は一度しか変化しない。この場合は、回路内のリセット回路により、T-FF の状態は再び反転する。

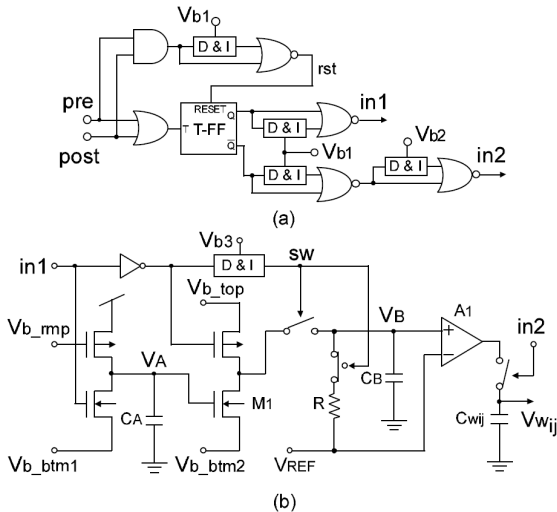


図 3・4 対称型 STDP 回路。(a) スパイク検出部、(b) 荷重更新部

入力スパイク $in1$ が荷重更新部に与えられると、図 3・4 (b) のキャパシタ C_A の端子にランプ波形 V_A が生成される。また、同時に遅延反転回路によって制御信号 SW が high となる。ランプ波形は後段のトランジスタ M_1 によって非線形波形に変形され、トランスコンダクタンスアンプ A_1 の入力端子へ与えられる。キャパシタ C_B の端子電圧 V_B は SW が low となると、抵抗 R によって基準電位 V_{REF} へ落ち着く。非線形波形が生成されている間に入力スパイク $in2$ が与えられると、 A_1 がキャパシタ $C_{w_{ij}}$ を充放電する。もし、 V_B が V_{REF} に落ち着いた後で $in2$ が入力された場合、荷重値は更新されない。STDP 関数の形状は回路に供給されるアナログのバイアス値によって決定される。 V_{b3} 、 V_{b_rmp} は STDP 関数の時間窓、 V_{b_top} 、 V_{b_btm2} 及びサンプリングパルスである $in2$ のパルス幅は、荷重値の更新量を決定するパラメータである。

(2) 非対称型 STDP 回路

図 3・5 に非対称型の STPD 回路を示す。本回路において、トランスコンダクタンスアンプ A_1

は $V_1 - V_{REF}$ に比例して $V_{w_{ij}}$ を増加させるのに対し、 A_2 は $V_2 - V_{REF}$ に比例して $V_{w_{ij}}$ を減少させる。言い換えれば、前者はシナプス増強、後者はシナプス減弱を実現する。

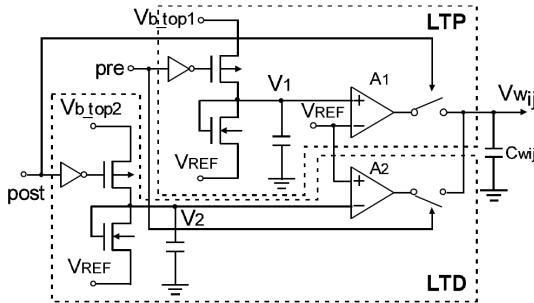


図 3・4 非対称型 STDP 回路

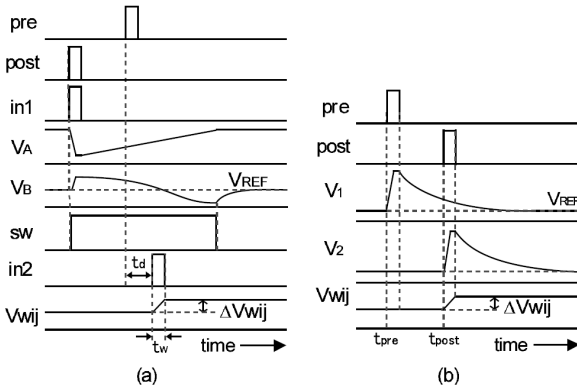


図 3・6 タイミングチャート。(a) 対称型 STDP 回路、(b)非対称型 STDP 回路

図 3・6 (b)に非対称型 STDP 回路のタイミングチャートを示す。 $t_{post} - t_{pre} > 0$ の場合には、入力スパイク pre によって非線形波形 V_1 が生成されると同時に A_2 が駆動される。しかし、 $V_2 - V_{REF} = 0$ であるため、 A_2 は $V_{w_{ij}}$ を更新できない。この場合、 $V_{w_{ij}}$ は A_1 のみによって更新される。また、 $t_{post} - t_{pre} = 0$ の場合には A_1 と A_2 が同時に駆動されるため、 $C_{w_{ij}}$ に流れる電流は相殺される。対称型の回路と同様に、STDP 関数の形状は V_{b_top1} 、 V_{b_top2} などのバイアス値によって決定される。

6-3-4 スパイクングニューラルネットワーク集積回路

試作チップの測定結果を図 3・7 及び図 3・8 に示す。図 3・7 に対称型 STDP 特性と非対称 STDP 特性が示されている。また、対称型 STDP 回路での PSP 振幅の変化を図 3・8 示す。スパイク時間差が 0 s のときはシナプス荷重が増加するため PSP は徐々に増加するが、スパイク時間差 90 s のときにはシナプス荷重が減少するため PSP は徐々に減少していくことから、STDP 学習がなされていることが分かる。

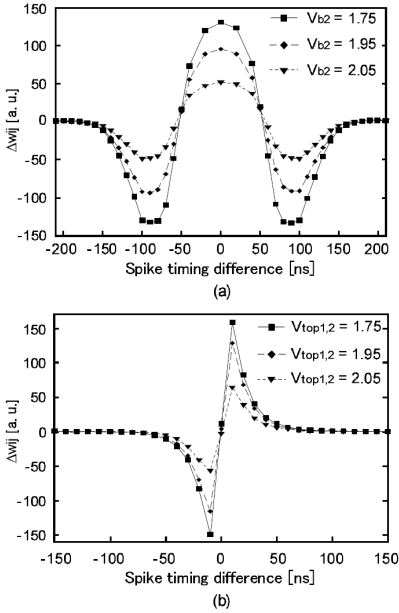


図 3・7 DP 回路測定結果. (a) 対称型 STDP 特性, (b) 非対称 STDP 特性

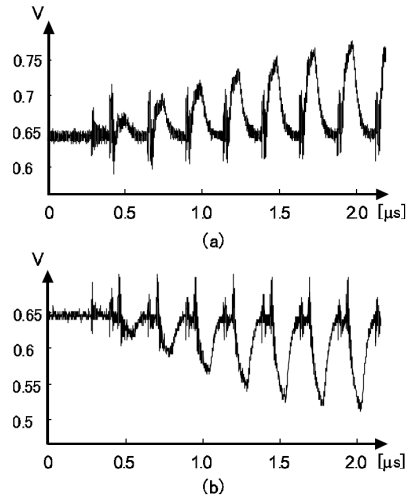


図 3・8 STDP 回路測定結果. 対称型 STDP 回路での PSP 振幅の変化. (a) スパイク時間差 0 ns, (b) スパイク時間差 90 ns

上記のスパイクングニューラルネットワーク集積回路の設計例として試作 LSI チップの諸元とレイアウト例をそれぞれ表 3・1 と図 3・9 に示す。

このように、STDP 機能をアナログ集積回路で実現し、荷重を変化させることができる。これを利用して連想記憶などの脳型処理モデルを実現することができる。ただし、既存の CMOS 集積回路技術ではキャパシタに電荷を蓄えるという揮発性記憶機能しか実現できないので、フラッシュメモリや抵抗変化型メモリ（メモリスタとも称される）などの不揮発性メモリ素子技術を採用する必要がある⁹⁾。

表 3・1 STDP 機能搭載スパイクングニューラルネットワーク LSI 諸元

製造技術	TSMC 0.25 μm 1-Poly 5-Metal CMOS
ニューロン数	25
対称型 STDP シナプス数	600
非対称型 STDP シナプス数	300
学習周期	200 ns
消費電力	280 μW /ニューロン回路 250 μW /対称型 STDP シナプス回路 100 μW /非対称型 STDP シナプス回路

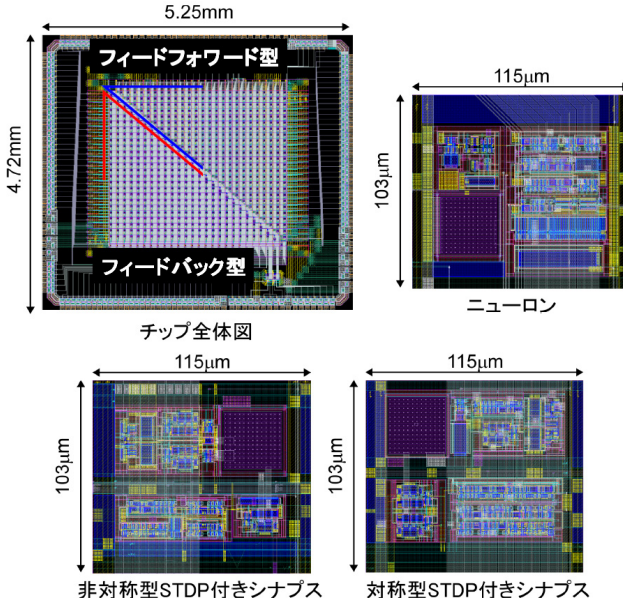


図 3・9 STDP 機能搭載スパイクングニューラルネットワーク LSI のレイアウト結果

■参考文献

- 1) W. Maass : “Networks of Spiking Neurons: The Third Generation of Neural Network Models,” Neural Networks, vol.10, no.9, pp.1659-1671, 1997.
- 2) W. Maass and C. M. Bishop, editors : “Pulsed Neural Networks,” MIT Press, Cambridge, MA, 1999.
- 3) K. A. Boahen : “Point-to-Point Connectivity Between Neuromorphic Chips Using Address Events,” IEEE Trans. Circuits & Syst. II, vol.47, pp.416-434, 2005.
- 4) G. Indiveri, E. Chicca, and R. Douglas : “A VLSI Array of Low-power Spiking Neurons and Bistable Synapses with Spike-timing Dependent Plasticity,” IEEE Trans. Neural Networks, vol.17, no.1, pp.211-221, 2006.
- 5) H. Tanaka, T. Morie, and K. Aihara : “A CMOS Spiking Neural Network Circuit with Symmetric/Asymmetric STDP Function,” IEICE Trans. Fundamentals, vol.E92-A, no.7, pp.1690-1698, 2009.
- 6) G. S. Snider : “Spike-Timing-Dependent Learning in Memristive Nanodevices,” in IEEE Int. Symp. Nanoscale Architectures (NANOARCH), pp.85-92, Jun. 2008.

■10 群-9 編-6 章

6-4 シリコン網膜と視覚処理用ニューラルネットワーク集積回路

(執筆著者：亀田成司) [2009年7月 受領]

生物は様々な感覚系を駆使することで自らを取り巻く環境を知覚し行動している。なかでも視覚系は時空間的に膨大な量の情報を扱うことから外界知覚において中心的な役割を果たしている。生体視覚系に見られる情報処理機構や機能を神経回路網（ニューラルネットワーク）として表現し集積回路化する視覚処理用ニューラルネットワーク集積回路の研究分野において、特に生体網膜を模倣設計した集積回路はシリコン網膜と呼ばれている^{1),2)}。

6-4-1 シリコン網膜

脊椎動物の視覚系で最も初段に位置するのが網膜である。網膜では側方向に接続を持つ細胞層が階層的に結合することで神経回路網が形成される（図 4・1）。視細胞層で受容された視覚情報は、外網状層及び内網状層に形成された神経回路網で時空間的に処理され、神経節細胞層を経て大脳皮質の視覚野に出力される³⁾。

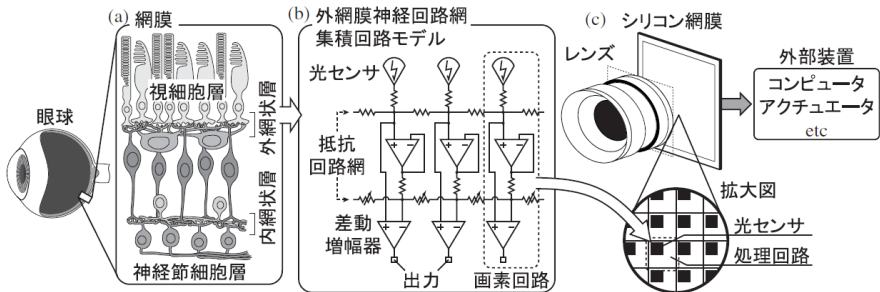


図 4・1 シリコン網膜と生体網膜の対比。(a)生体網膜の断面模式図、(b)外網膜神経回路網の集積回路モデル、(c)シリコン網膜の構成

このような、網膜の超並列回路構造及び機能を集積回路により模倣設計した視覚デバイスがシリコン網膜である（図 4・1(c)）。シリコン網膜は各画素に光センサと処理回路を持つことで超並列回路を形成し、光センサで受容した画像情報を実時間で処理する。パソコンや DSP (Digital Signal Processor) などを利用した従来の画像処理システムに比べて、高速処理、小型、低消費電力といった特徴を持つ。1988年に Mead と Mahowald により開発された世界初のシリコン網膜は、脊椎動物の外網膜の構造を模倣した側抑制型受容野を有し、入力画像の空間コントラスト強調（空間微分）機能を持つ⁴⁾。その後、光センサに順応機構を組み込み高感度化を試みたチップ⁵⁾や、より詳細な生理学的知見に基づいたラブラシアン-ガウシアン ($\nabla^2 G$) 型受容野を有するチップが開発されている^{6),7)}。また、脊椎動物の網膜や昆虫の視覚系における動き検出の知見に基づいたチップも開発されている^{8)~10)}。更に、網膜における機能分化に着目して、持続性応答経路（空間微分）と一過性応答経路（時間微分）の複数の視覚経路を持ったチップも開発されている^{11),12)}。多くのシリコン網膜の光センサには、神経回路網の非同期的な動作を表現するために、連続時間型（光電効果により生成した電流変化出力を利用）が用いられている

が応答感度や精度は一般的に低い。そこで、サンプル型光センサ（光電効果による電荷を一定時間蓄積（露光）し出力）を用いたシリコン網膜が開発されている¹²⁾。サンプル型は CMOS イメージセンサに使用されている方式で、露光時間の制約のため応答は離散的になるが、連続時間型に比べ高い応答感度を持ち、精度補償の機構を組み込みやすい。

6-4-2 高次視覚処理機能の実現

網膜から大脳皮質の視覚野への過程における高次視覚処理機能を模倣した視覚処理用ニューラルネットワーク集積回路の研究も行われている。生体視覚系では、網膜で時空間に処理された視覚情報は、大脳皮質の一次視覚野へ送られる。一次視覚野では、方位、運動方向、色などの特徴情報抽出及び情報統合が行われる。そして、高次の視覚領野に至る過程において機能分化が進み、各々異なる視覚領野において立体視、運動視、色覚、形態視といった高次視覚処理が分散的に行われている³⁾（図 4-2 (a)）。

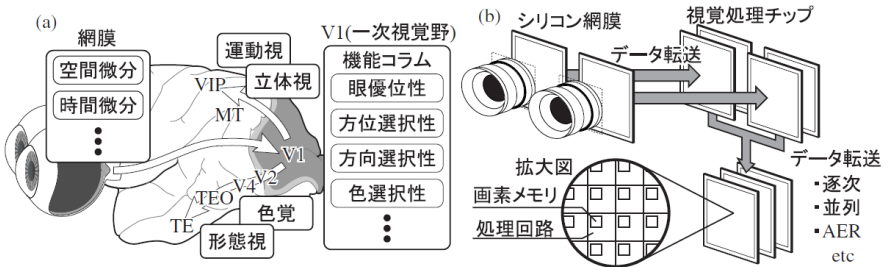


図 4-2 生体視覚系を模倣するシリコン視覚野. (a)生体視覚系の情報経路と機能分化, (b)シリコン視覚野の構成例

このような、視覚野に見られる低次から高次への階層構造や機能分化に伴う分散処理を模倣するために、複数の集積回路チップで構成した視覚システムをマルチチップ視覚システムやシリコン視覚野などと呼ぶ（図 4-2 (b)）。階層構造を複数のチップで表現し、処理を分散的に行うことで、視覚野に見られるような高次視覚機能が高解像度に行うことができる。また、チップの接続経路を組み合わせることで視覚処理機能の変更も可能である。ただし、チップ間で視覚情報を転送する必要があるため、転送時間及び精度がシステム性能に直接影響する。

このチップ間情報転送に光センサの特性が密接に関係している。光センサがサンプル型の場合の情報転送は、露光時間に同期した逐次転送により比較的容易に実行できる。一次視覚野の方位選択性を模倣した逐次転送方式のマルチチップ視覚システムが開発されている¹³⁾¹⁴⁾。この転送経路を並列にすれば、それだけ高速な転送が可能になる¹⁴⁾。一方、多くのシリコン網膜で採用されている連続時間型光センサの場合、離散的な逐次転送方式では時間連続な情報は転送できない。そこで考えられたのが AER (Address Event Representation) 通信方式である¹⁵⁾。AER 方式は、送信側チップは視覚情報を非同期のパルス頻度信号に変換し、各画素のパルス（イベント）を位置（アドレス）情報に符号化して多重送信を行う通信手法である。受信側チップは、積分発火 (Integrate-and-Fire) 型ニューロンのように、各画素でパルス頻度信号を積分することで視覚情報を再構成する。現在、AER 方式を利用したシリコン視覚野が積極的に開発され、方

位選択, 立体視, オプティカルフロー, 注意選択などの高次視覚処理に適用されている^{16)~18)}. また, 連続時間型光センサの応答精度の問題を改善した時間微分チップが開発され, 実用的な応用に向けた研究も行われている¹⁹⁾.

これらの視覚システムを構成する視覚処理チップは視覚系の特定の構造を模倣しているので汎用的な処理は基本的にできないが, 再構成可能なチップも開発されている. 隣接画素との接続構造を持つニューラルネットワークモデルである CNN (Cellular Neural Network) に基づくチップ及びチップシステムは各画素に汎用的なニューロン回路, 各種メモリ, 制御ユニットを内蔵することで初期視覚処理やテンプレート処理などの様々な視覚処理を実行できる^{20),21)}. また, 各画素に光センサやデジタル演算装置を内蔵したデジタルビジョンチップは生体模倣ではないが各種視覚処理を実行可能でロボットハンドの高速視覚制御などに応用されている²²⁾.

6-4-3 応用と展望

視覚処理用ニューラルネットワーク集積回路は処理の高速性, 小型, 低消費電力などの特徴から, 工学的にはロボットビジョンや認識処理装置の前処理デバイスとして応用されている^{19),22),23)}. 脳研究においても, 生体視覚系の構造を積極的に模倣したシリコン網膜やシリコン視覚野は視覚機能を解明するためのエミュレータとしての役割が期待されている¹⁸⁾. また, 近年ではインプラント技術としての人工眼の研究が行われている^{24),25)}. 現状の人工眼は光強度程度の情報しか扱えないが, シリコン網膜の技術を適用することで生体視覚系に適応した人工眼が実現できるであろう. 現在, 集積回路技術はトランジスタの微細化や高速化, 3次元集積技術による大容量化や高密度化など目覚ましい発展を続けている. この3次元集積技術を利用して網膜の3次元ネットワーク構造を模倣集積する研究も進められている²⁶⁾.

我々が生活する視覚環境に適応して進化してきた生物の視覚処理機構を工学的に模倣する試みは意義深いが, 生物とは性質の違う集積回路による実現は容易ではない. しかし, その違いを理解し, 集積回路の利点を上手く活用できれば, 特定の側面において生体視覚系に近づき超越する視覚システムを構築することは十分可能だろう. そして, 発展を続ける集積回路技術とナノテクノロジーやバイオテクノロジーなどの研究領域との幅広い融合により, 生体視覚系を模倣した集積回路システムの更なる進化が期待される.

■参考文献

- 1) C. Mead : "Analog VLSI and Neural Systems," Addison-Wesley, Reading, MA, 1989.
- 2) A. Moini, ed. : "Vision Chips, Kluwer Academic Publishers," MA, 1999.
- 3) 三島擠一(総編集) : "眼の事典," 朝倉出版, 2003.
- 4) C. Mead and M. Mahowald : "A Silicon Model of Early Visual Processing," Neural Networks, vol.1, no.1, pp.91-97, 1988.
- 5) M. Mahowald : "Silicon retina adaptive photodetectors," Proc. SPIE, Visual Information Processing: From Neurons to Chips, vol.1473, pp.52-58, 1991.
- 6) K.A. Boahen and A.G. Andreou : "A contrast sensitive silicon retina with reciprocal synapses, NIPS, vol.4, pp.764-772, 1992.
- 7) T. Yagi, T. Matsumoto, and H. Kobayashi : "Parallel analog image processings: solving regularization problems with architecture inspired by the vertebrate retinal circuit," Neural Network Systems Techniques and Applications, ed. C.T. Leondes, pp.201-285, Academic Press, CA, 1998.

- 8) T. Delbruck : "Silicon retina with Correlation-Based, Velocity-Tuned Pixels," IEEE Trans. Neural Networks, vol.4, no.3, pp.529-541, 1993.
- 9) A. Moini, A. Bouzerdoum, K. Eshraghian, A. Yakovlev, X.T. Nguyen, A. Blanksby, R. Beare, D. Abbott, and R.E. Bogner : "An Insect Vision-Based Motion Detection Chip," IEEE J. Solid-State Circuits, vol.32, no.2, pp.279-284, 1997.
- 10) J. Kramer, R. Sarpeshkar and C. Koch : "Pulse-based analog VLSI velocity sensors," IEEE trans. Circuits and Systems-II, vol.44, no.2, pp.86-101, 1997.
- 11) K.A. Boahen : "A Retinomorph Chip with Parallel Pathways: Encoding ON, OFF, INCREASING, and DECREASING Visual Signals," Analog Integrated Circuits and Signal Processing, vol.30, no.2, pp.121-135, 2002.
- 12) S. Kameda and T. Yagi : "An analog VLSI emulating sustained and transient response channels of the vertebrate retina," IEEE trans. Neural Networks, vol.14, no.5, pp.1405-1412, 2003.
- 13) K. Shimonomura and T. Yagi : "A multi-chip aVLSI system emulating orientation selectivity of primary visual cortical cell." IEEE Trans. Neural Networks vol.16, no.4, pp.972-979, 2005.
- 14) S. Kameda and T. Yagi : "An analog silicon retina with multichip configuration," IEEE trans. Neural Networks, vol.17, no.1, pp.197-210, 2006.
- 15) K.A. Boahen : "Point-to-Point Connectivity Between Neuromorphic Chips using Address-Events," IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing, vol.47, no.5, pp.416-434, 2000.
- 16) C.M. Higgins and C. Koch : "A Modular Multi-Chip Neuromorphic Architecture for Real-Time Visual Motion Processing," Analog Integrated Circuits and Signal Processing, vol.24, pp.195-211, 2000.
- 17) G. Indiveri, R. Murer and J. Kramer : "Active Vision Using an Analog VLSI Model of Selective Attention," IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing, vol.48, no.5, pp.492-500, 2001.
- 18) R. Silver, K. Boahen, S. Grillner, N. Kopell, and K.L. Olsen : "Neurotech for Neuroscience: Unifying Concepts, Organizing Principles, and Emerging Tools," J. Neuroscience, vol.27, no.44, pp.11807-11819, 2007.
- 19) P. Lichtsteiner, C. Posch, and T. Delbruck : "A 128×128 120 dB 15 μ s Latency Asynchronous Temporal Contrast Vision Sensor," IEEE J. Solid-State Circuits, vol.43, no.2, pp.566-576, 2008.
- 20) L.O. Chua and T. Roska : "Cellular neural networks and visual computing," Cambridge University Press, UK, 2002.
- 21) R.C. Galan, F. Jimenez-Garrido, R. Dominguez-Castro, S. Espejo, T. Roska, C. Rekeczky, I. Petras, and A. Rodriguez-Vazquez : "A Bio-Inspired Two-Layer Mixed-Signal Flexible Programmable Chip for Early Vision," IEEE trans. Neural Networks, vol.14, no.5, pp.1313-1336, 2003.
- 22) 鏡 慎吾, 小室 孝, 渡辺義浩, 石川正俊 : "ビジョンチップを用いた実時間視覚処理システム VCS-IV," 信学論 (D-I), vol.J88-D-I, no.2, pp.134-142, 2005.
- 23) M. Yagi and T. Shibata : "An Image Representation Algorithm Compatible With Neural-Associative-Processor-Based Hardware Recognition Systems," IEEE Trans. Neural Networks, vol.14, no.5, pp.1144-1161, 2003.
- 24) W. Liu and M.S. Humayun : "Retinal Prosthesis," 2004 ISSCC, pp.218-219, 2004.
- 25) L. Theogarajan, J. Wyatt, J. Rizzo, B. Drohan, M. Markova, S. Kelly, G. Swider, M. Raj, D. Shire, M. Gingerich, J. Loewenstein, and B. Yomtov : "Minimally Invasive Retinal Prosthesis," 2006 ISSCC, pp.24-25, 2006.
- 26) M. Koyanagi, Y. Nakagawa, K.-W. Lee, T. Nakamura, Y. Yamada, K. Inamura, K. Park, and H. Kurino : "Neuromorphic Vision Chip Fabricated Using Three-Dimensional Integration Technology," 2001 ISSCC, pp.270-271, 2001.

■10 群-9 編-6 章

6-5 雑音を利用するニューラルネットワーク集積回路

(執筆著者：浅井哲也) [2009年9月 受領]

工学、とりわけ集積回路工学においては、半導体素子のばらつきや外界からの雑音などの「ゆらぎ」の要素を極力排除するように回路を設計することが常識とされている。一方、自然が長い年月をかけて創り出した生物の神経系は、雑音を排除するのではなく活用する方向に進化してきた。本節では、生物の神経系の「ゆらぎを利用する」という姿勢に学んで、そのしくみを集積回路の設計に役立てようとする試みを2つ紹介する。具体的には、雑音を利用してパルス密度変換、及び位相同期を行うニューラルネットワーク集積回路¹⁾²⁾について簡潔に紹介する。

6-5-1 雑音を利用してパルス密度変換を行うニューラルネットワーク集積回路

ニューラルネットワークの重要な構成要素の一つとして、積分発火型の神経細胞モデルが挙げられる。この神経細胞モデルは、与えられたアナログ入力を時間積分し、その積分値がある閾値を超えたときに出力パルスを生成し、その直後に積分値をリセットするものである。この動作を繰り返すと、アナログ量を出力パルスの間隔に変換できるので、この神経細胞モデル単体の機能は「パルス密度変換 (1 ビット A/D 変換)」であると言ってよい。

このモデルを用いて、図 5・1 に示すようなニューラルネットワークを構成する³⁾。図中の大きな白丸 (○) は積分発火型の神経細胞を表す。これらの神経細胞は共通のアナログ入力、及びネットワーク外部から独立した雑音を受ける。また、アナログ入力の受け口となる各神経細胞のシナプス (図中の小さな白丸:○) は、それぞればらつきを持っている ($\alpha_i \neq \alpha_j, i \neq j$)。これらの神経細胞の出力パルスは、図中の Σ ユニットのより加算される。加算されたパルス列は、ネットワークの出力となると同時に、抑制性のシナプス (図中の小さな黒丸:●) を介して個々の神経細胞にフィードバックされる (負のフィードバック、 $k < 0$)。

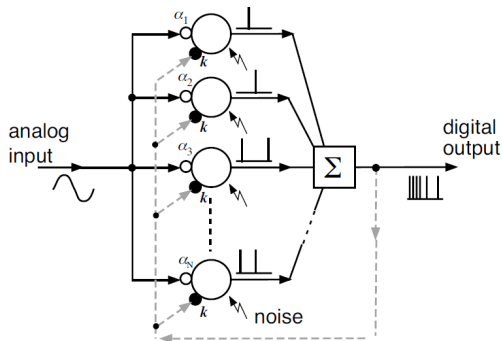


図 5・1 雑音を利用してパルス密度変換を行うニューラルネットワークの例

ニューラルネットワークを集積回路上に構築すると、ゆらぎの要素 (主に、回路素子のばらつきと外部雑音) が必ず現れる。このゆらぎを極限まで小さくすることが現在の集積回路設計における基本思想であるが、ここではあえてゆらぎを残しておく。素子ばらつきの大きい、か

雑音に鋭敏なサブスレッショルド CMOS 回路を用いてニューラルネットワーク集積回路を構成すれば、ゆらぎの要素は自然に取り込まれる (図 5・2)。適度な雑音強度と負帰還の強さを設定しておくことで、神経細胞間で弱い競合現象が起こり、外部雑音によってある神経細胞がパルスを生じできなかったときにのみ、別の神経細胞がパルスを生じ、ネットワーク全体として神経細胞が互いを補いあいながら動作するようになる。その結果、ネットワーク出力のパルス間隔が均等になり、ネットワークをパルス密度変換器として見た場合の SNR が向上する (図 5・3)¹⁾。この競合現象は、シナプス結合重みのばらつきによって発現する。したがって、素子ばらつきや外界からの雑音が避けられないような環境では、シナプス結合重みのゆらぎを排除するのではなく、ゆらぎを少しだけ残した方がよいのである。

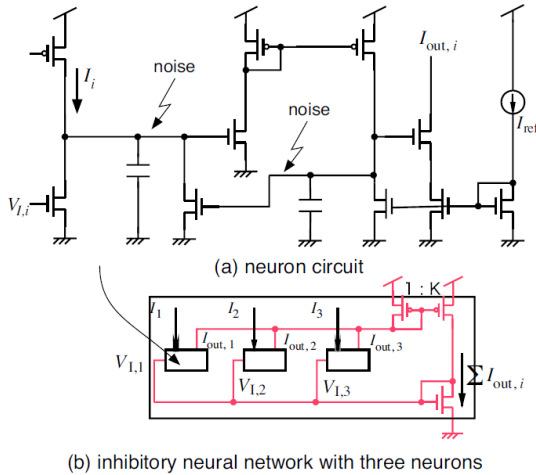


図 5・2 サブスレッショルド領域で動作する MOSFET によるニューラルネットワーク回路

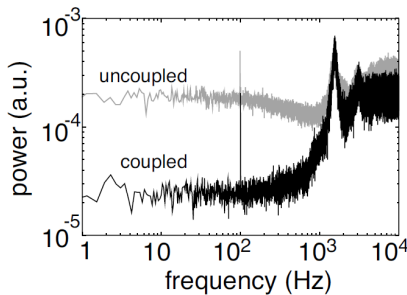


図 5・3 パルス密度変調におけるパワースペクトル (抑制なし: uncoupled, あり: coupled)

6-5-2 雑音を利用して位相同期を行うニューラルネットワーク集積回路

神経細胞の集団に雑音を加えることでそれらが同期発火するという現象は古くから知られている。特に、「相互結合を持たない“独立した”神経細胞群が雑音により位相同期する⁴⁾」と

いう性質は、工学的な応用価値が高いと思われる。例えば、個々の神経細胞を集積回路上に置かれた独立クロック源として考えれば、雑音を加えることでそれらを同期させられるかもしれない。この性質をクロックスキューなどの問題解決の糸口にできないだろうか。

雑音による位相同期を集積回路上で確認するために、Wilson-Cowan オシレータを利用した CMOS 神経細胞回路が設計された (図 5・4)²⁾。この回路は、シュミットトリガ-インバータとして動作する正帰還 OTA、及びバッファから成るリングオシレータである。外部雑音 (V_{noise}) により位相が変調されるような回路 (図中左の C, r_0 と buffer) が付加されている。図 5・5 はこの回路の出力波形の一例であり、外部雑音により位相がゆらぐ様子が分かる。

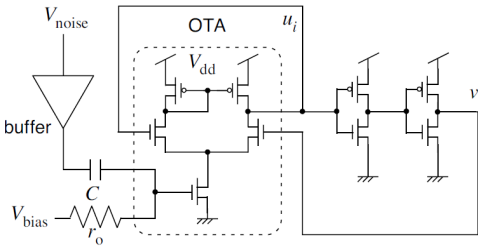


図 5・4 Wilson-Cowan 型オシレータ回路

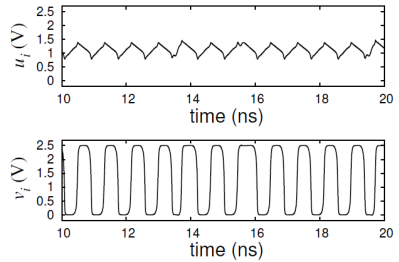


図 5・5 オシレータの出力波形の例

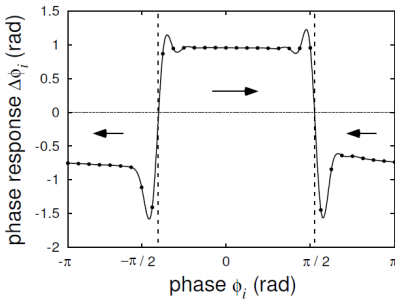


図 5・6 位相応答曲線

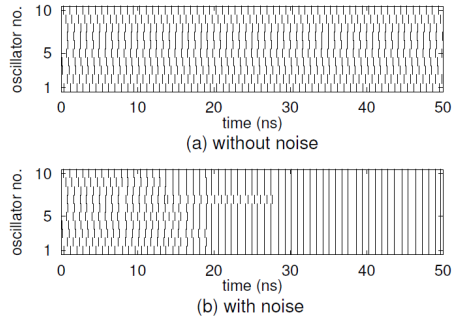


図 5・7 10 個のオシレータの位相同期の例

図 5・6 は、この回路の位相応答曲線 (Phase Response Curve : PRC) の一例である。PRC は、オシレータの位相が ϕ_i の状態で外部入力を与えたとき、それによる位相の変化量 ($\Delta\phi_i$) を示すものである²⁾。この例では、パルスを与える直前の位相が $-\pi/2 < \phi_i < \pi/2$ のとき、 $\Delta\phi_i$ は正となっている。つまり、この位相でパルスを与えると、オシレータの位相が進む。それ以外の位相では $\Delta\phi_i$ は負となり、パルスを与えると位相は遅れる。ここで注目すべきことは、「ランダムなタイミングでオシレータにパルスを加えると、位相が $\phi_i \approx \pi/2$ の周りに収束する」という性質である。パルスが与えられるたびに、図 5・6 中の矢印の方向に位相がシフトするため、オシレータの位相が $\pi/2$ 近傍に集まる。したがって、複数のオシレータを用意し、ランダムなタイミングでそれらのオシレータにパルスを加えれば、上記の理由により、すべてのオシレータの位相が $\pi/2$ 近傍に集まる。図 5・7 にその様子を示す。ランダムパルスを加えない場合は、10

個のオシレータの位相はばらばらである (図 5・7 (a)) (各位相が 0 となった時刻に縦線を記してある)。一方, ランダムパルスを加えると, 次第にオシレータの位相が次第にそろっていく。

各オシレータの素子ばらつき (固有振動数のばらつき) に対する同期構造はさほど不安定でないため²⁾, この回路の集団を集積回路上に実装すれば, それらは外部から共通の雑音を加えることで, 強制同期させられる。つまり, ここでも雑音が有効利用されている。

■参考文献

- 1) A. Utagawa, T. Asai, T. Hirose, and Y. Amemiya : “An inhibitory neural-network circuit exhibiting noise shaping with subthreshold MOS neuron circuits,” IEICE Trans. Fundamentals, vol.E90-A, no.10, pp.2108-2115, 2007.
- 2) A. Utagawa, T. Asai, T. Hirose, and Y. Amemiya : “Noise-induced synchronization among sub-RF CMOS analog oscillators for skew-free clock distribution,” IEICE Trans. Fundamentals, vol.E91-A, no.9, pp. 2475-2481, 2008.
- 3) D.J. Mar, C.C. Chow, W. Gerstner, R.W. Adams, and J.J. Collins : “Noise shaping in populations of coupled model neurons,” Neurobiology, vol.96, no.18, pp.10450-10455, 1999.
- 4) H. Nakao, K. Arai, and K. Nagai : “Synchrony of limit-cycle oscillators induced by random external impulses,” Phys. Rev. E, vol.72, no.2, 026220, 2005.