人工知能学会研究会資料
SIG-AGI-015-08

Spiking Neural Network 技術の現状と課題に関する考察

A study on spiking neural network hardware technology

岡島　義憲　(フリー)

Yoshinori Okajima

Abstract: Introducing spiking signals to neural network operations is important and fundamental feature of the neuromorphic architectures, because of reducing energy-efficiency dramatically, and for emulating information operations of human cortex. Though, in terms of application accuracy, the Spiking Neural Network, SNN, has not been outperforming the Artificial Neural Networks, the reason of which is regarded as that the SNN is following ideas of existing Deep-Neural-Network topologies and their learning methodologies. The outline of this situation is reviewed and studied for seeing into next deployments of the SNN.

１．はじめに：SNN の経緯

Spiking Neural Network（SNN）は、電気インパルスによって生ずるシナプス後細胞の活動電位変動のモデル式をコンピューティング動作の演算に用いることを特徴とするニューロモーフィック技術の一つである.

ニューロモーフィックとの表現は、1990 年前後に、半導体集積回路の専門家であった Carver Mead[1]の造語であるが、当初は特定の技術を意味するのではなく、「生体の神経回路に学び(Bio-Inspired)、回路技術の新しいパラダイム開拓のヒントとする」という技術開発＆研究戦略であった.

1997 年に、機械学習の専門家であった W. Maass が、Spiking Signal を入力信号に採用した Neural Network の演算能力を多層の Artificial Neural Network と対比し、その計算能力は条件付きではあるが、Universal な能力を持ちうるとし、SNN を第三世代の Neural Network 技術と位置付けた[2].

ついで、2004 年に W. Maass は、脳学者であったスイス EPFL の H. Markram と、Von-Neumann 型コンピュータと対比させる形で、SNN をデータ・フローをリアルタイムに処理可能な Universal Machine と見做し得るとし、その数理を Liquid State Machine（LSM）との概念で定式化した [3] [4] [5].

基本となるのは、Spiking Signal モデルと Neuron モデルである. その概要を先ず本論２章に記載する.

2014 年以降、SNN 搭載用に設計されたニューラル・ネットワーク表現用の専用半導体チップの試作発表が相次いでいる.

まず、推論機能のみのチップが登場し、SNN の動作が非常に低消費電力であると認識された[11] [12]. 但し、Spiking Signal を適用するには、現状のコンピュータ用の「データ」を Spiking Signal に変換する必要があった. いずれも、生体の模倣ではない工学的な工夫である.

また、学習メカニズムとしては、STDP によるシナプス係数調整であり、ANN にて開花した教師有り学習を行うには、ANN にて調整した係数を SNN に移植するとの工夫を必要とした.

Winner-Takes-All (WTA)回路を実装することにより、教師無し学習は可能とされた[15] [16].

Network 表現については、神経細胞間の複雑な配線を既存の電子回路技術で大規模に実装することには大きな困難があるので、生体の模倣は行なえず、いずれの場合も工学的な既存技術の流用であった.

「発火」という生体内の量子化現象と電子回路内のパルスとの対応関係や、その間の Encode 方法やその意味論について多くの議論があり、そして、ASIC や FPGA による評価チップが発表され、それらを踏まえたベンチマーク議論が続いて来た. 現状においては、消費電力以外での ANN に対する SNN の明瞭な優位性は未だ見えていない[18] [19] [20].

WTA 回路とベンチマーク状況については、各々３章と４章で概観する.

SNN の優位性不足の根本には、大きく３点の見方が在る.

1) Network-Topology を同じとした条件での比較であるので、当然の結果とする[18].
2) 既に、ANNの認識精度は人間を上回っているの

* 連絡先 E-mail ： y.okajima@acoustic.zaq.jp

で、ニューロモーフィック戦略によるANNの性能凌駕には無理がある[20].

3) 生体の脳の機能を十分に実装できていない[14].

2005 年、H. Markram は、Bio-Inspired の次の知見を目指して、IBM 社との共同研究(Blue Brain Project)をスタートさせた. このプロジェクトは大脳新皮質の特徴的な構造体であるミニ・コラムやコラムの動作をシミュレーションすることであった[10].

大脳新皮質の局所回路に関する脳科学の調査は非常に多い[6] [7] [8] [9].

ミニ・コラムは生体が後天的な学習を始める前に形成されている構造であるため、今後、その知見の実装に向けての進展が予想される. この点については、５章にて考察する.

尚、本論は、回路アーキテクチャを考察の対象とするため、CMOS ベースのテクノロジーを用いた SNN 用のハードウエアについてのみ扱い、Memristor 等の新規製造技術とした製造コスト削減に関する技術は対象としない.

2. Neuron Model と Signal Model

2.1. Neuron Model

生体の中の神経は、前段の神経細胞群から様々な電気インパルスを受け続け、設定されたシナプス係数にて「電気インパルス」をフィルタリングし続け、それによって生ずる細胞膜電位差が閾値を超えた場合に発火する. このメカニズムを定式化したモデルを、一般に Integrate & Fire（I&F）と呼ぶ.

ANN の形式ニューロンにおいては、「発火」の概念は明瞭には存在せず、一般的な形式ニューロンの演算においては、「その神経細胞のシナプス係数群が成すベクトル(W_μ)と入力信号群が作る入力ベクトル(S_μ)が作る内積値 $\{\sum(W_\mu S_\mu)\}$ からバイアス値(B)の減じた値を活性化関数(φ)にて変換し、出力値(Q)を生成する」との定式化であった.

$$Q = \varphi\{\sum(W_\mu S_\mu) - B\} \qquad \cdots\cdots \text{式(A)}$$

この計算には、時間(t)の登場は必須ではない.

一方、SNN においては、「ニューロンには、$\sum(W_\mu S_\mu)$に相当する後シナプスが注入されるとし、神経細胞膜に蓄積される電荷の時間変化値$(C\Delta V(t))$を、神経細胞に注入される電流$(\sum(W_\mu S_\mu))$とリーク電流$((V(t) - V_{reset})/R_d)$の差によって説明できる」とする.

$$C\Delta V(t) = \left\{\sum(W_\mu S_\mu(t)) - (V(t) - V_{reset})/R_d\right\} \cdot \Delta t \qquad \cdots\cdots \text{式(B)}$$

ここで、抵抗値R_dは、細胞膜のリーク抵抗値である.

W_μは、$(1/\text{抵抗})$の次元を持つパラメータ、$S_\mu(t)$は前シナプス細胞が伝える電圧パルスとなっている. $S_\mu(t)$は、k 回目の発火タイミングを${t_\mu}^k$とすると、ディラックのδ関数を使って、

$$S(t)_\mu = S_f \times \{\sum_k \delta(t - {t_\mu}^k)\}$$

と表せる. S_fは電圧の次元を持つ.

SNN の発火するタイミングは、膜電位 $V(t)$ が発火電位 V_{fire} を超えた時とされる. 発火後は、発火前の膜電位 である V_{reset} にリセットされる.

以上は、Integrate & Fire（I&F）の最も基本的なモデルであり、モデルの精度を高めた多数のヴァリエーションが存在するが、それらモデルの差異の影響は小さいとされている[21][22].

ここで注意すべきは、「発火」の導入によって、以下のコンセプトが自動的に導入されることとなった点である.

(1) 量子化

「発火」は、入力された刺激量に対し量子化操作を行ったこととなる. （抵抗値R_dによる減衰効果と、式（A）の値(B)を無視した場合には、式（A）と式（B）は同形式であり、SNN では発火により、式（A）の出力値(Q)を、V_{fire}を単位として除算し、整数の算出を行っていることが分かる.）

入力された信号の発火は、生体の神経回路内においても、また、電子回路にても、信号伝送時のノイズ耐性を大きくし、神経回路内を長距離伝送では大きな利点を持つ.

「発火」の導入と共に、SNN では、「発火頻度」、もしくは「発火確率」という随伴情報（時間と共に変化しうる変動パラメータ）を導入することがある[15]. （この変動パラメータは、ANN にて活性化関数を ReLU を用いた時の演算結果に相当しうる.）

ここで、注意すべきは、SNN にて扱う Spiking 信号と、上記の発火の随伴情報としての「発火頻度」と、実際の神経細胞にて起こる発火のバースト現象の関係である.

生体内の神経細胞の発火現象とは連続することがあり、連続するそのパルス個数は数えることが出来る（図１）.

電子回路にて、同様のパルスが入力されると仮定すると、その動作は、I&F がモデル化した時と同じとなり、発火タイミングを$(t_\mu{}^k)$として、

$$S(t)_\mu = S^f \times \{\sum_k \delta(t - t_\mu{}^k)\}$$

となる.

一方、現在の多くの SNN において、パルスは、通常、発火そのものではなく、「発火頻度」を表現することに使われる[21][22].

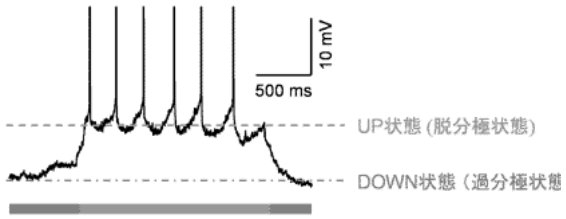


図１．神経の発火現象
（引用）池谷裕二氏の HP より掲載
http://gaya.jp/research/spontaneous_activity.h

(2)　低消費電力化

発火により、信号の電位レベルの High 側(値１)と Low 側(値 0)には、非対象な意味付けが可能となった. 値”1”は活性化状態、値”0”は非活性化状態の意味となる.

このような意味付けがハードウエア・レベルで行うことができると、Spiking Signal が伝わる先の回路は、容易に動作の停止や電源遮断を行えるようになり、消費電力を下げる回路テクニックを多用できることになる.　入力信号群が、よりスパースとなると、消費電力はより下がる.

この効果は、電子回路レベルで、そのような回路テクニックを使える場合に有効であり、SNN 動作を、既存の汎用コンピュータ・ハードウエアを用い、ソフトウエアによって表現する場合には、そのような低消費電力化は困難である.　SNN は、専用のハードウエア（ASIC）にての表現するメリットが大きい.

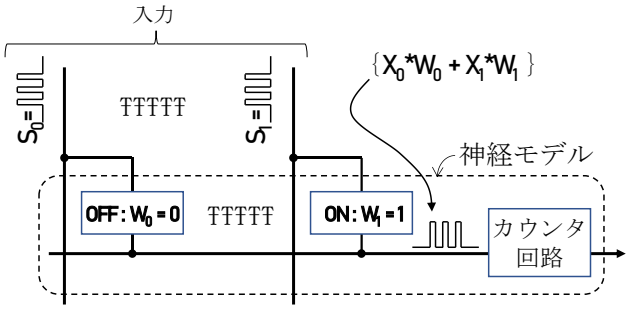


図２．積和回路

(3)　和算や積算が容易になる.

上記の「関連する数値情報」が、発火頻度ではなく、I&F に従った$t_\mu{}^k$に相当する発火動作を意味するとした場合には、図２に示したように非常にシンプルな電子回路にて、$\sum(W_\mu S_\mu)$ の積和演算を行うことが出来る.　その動作を、以下に示す.

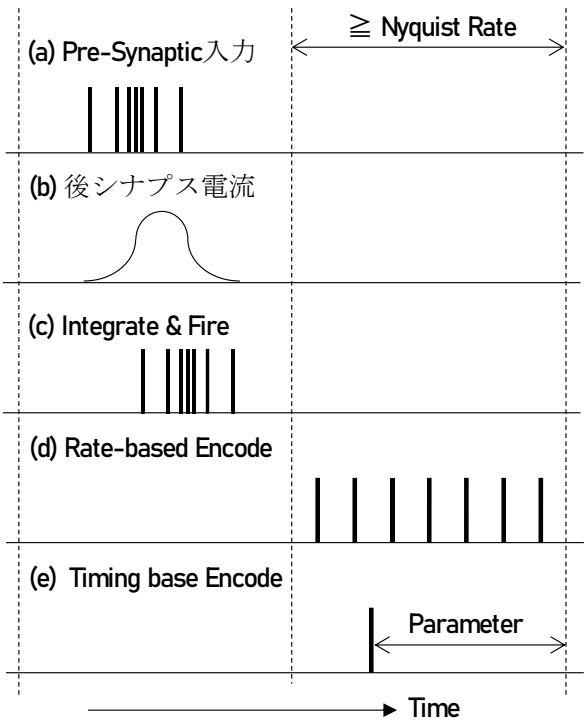


図３．Neural Encoding 手法

図中の S_0、S_1 は、入力信号(S_μ)の中の２本の信号である. また、各W_μは、記憶値にて制御され、$(1/抵抗)$の次元を持つスイッチである. ディジタル回路の場合は、抵抗値はゼロ又は無限大の２値であることが望ましいが、抵抗変化型のメモリを用いて、抵抗値をアナログ的に変化させるとの試みも多い.

入力信号（S_0、S_1）の電位レベルの High 側が値１、

Low 側が値 0 と意味付けが明瞭である場合、$\sum(W_\mu S_\mu)$の積和演算は、図２のように、複数の入力信号S_μを、抵抗をスイッチ(W_μ)で結び、そこに導通するパルス個数をカウントするだけの非常に簡易な回路で構成することができる.

この図の場合、各W_μに対して記憶値を１個しか記載していないが、各W_μに対してｍ個の記憶値を用意すると、各W_μに対して２$^{m-1}$の諧調をディジタルに設定することとなる.

シナプス接続を表現するスイッチ(W_μ)はメモリ素子でも表現可能である. そのような方式を In-Memory Computing、もしくは、Near-Data-Processing と呼ぶことがある.

2.2. Encoding 技術

神経細胞の複雑な結線をそのまま電子回路に置き換えることには電子回路の製造技術の面でも、また回路設計の面でも、現在は困難である. 従って、SNN の電子回路実装にては、通常、半導体技術にて培われたディジタル・データ通信の技術を流用する. 近年のディジタル・データ通信の技術は非常に高度であり、生体が進化によって獲得した手法よりも高度な技術ともいえるので、工学的には生体模倣する必要はないともいえる. 但し、通常のディジタル・データ通信の技術は、同期式の送受信回路を必要とするので、SNN が「Event-Driven の非同期 Spiking Signal」を基本とすることとは大きな相違がある.

そのような近年のディジタル・データ通信技術を用いる場合、ニューロン回路へ伝達される Spiking 信号は、プロトコルに定められたフォーマットに変換されてそのパケットに搭載されるデータとなる. 従って、データから入力する Spiking 信号への変換回路が必要である. この変換のことを Encode、もしくは、Transcode といい、様々な方式が提案されている[23][24] [25][26].

尚、ディジタル・データ通信に関しては、Encode（もしくは、Transcode）のための回路以外に、パケット・フォーマットに基づいて信号の伝達先アドレスを元に、途中のルーティングを決定し、伝送するルータ回路が必要である.

そのような前提で、発火信号が、次のニューロンの前シナプスに到達するまでの信号を図３に示した.

図３(a)には、シナプス前細胞の発火現象が記されている. これによって、注目する神経細胞に後シナプス電流(b)が注入され、Integrate & Fire 演算の結果、神経細胞の Model は発火現象(c)を演算出力する.

電子回路上は、発火により発生する Spike 信号によって、「発火個数」と「発火タイミング」という情報を伝えることができる. 通常は、「発火イベントのタイミング」と「発火頻度」をパルスによって伝送する.

SNN のコンセプトによっては、図３(c)の発火が注目する神経細胞の出力信号となり得るが、出力をネットワークの送る段階では、ニューロン・モデル演算の結果をディジタル数値として伝送する.

そのディジタル数値を受信したニューロンは、その数値を元に、図３(d) または、図３(e)のようにパルスを出力する.

図３(d)は、基準とする Timing-Window 内のパルス本数によって発火頻度を表現する Rate-Base Encoding と呼ばれる例である.

図３(e)は、基準 Timing との時間差に比例して発火頻度を表現する Timing-Base Encoding と呼ばれる[23][24].

ここで、Rate-Base Encoding の場合も、Timing-Base Encoding にても、共に、基準とする Time-Frame、もしくは、基準 Timing の想定が必要となってしまっていることに注目されたい. Time-Frame や基準 Timing は、通常、同期式回路の概念である.

この点は、Liquid State Machine の表現を目指した当初の SNN のコンセプトの後退である. その後退は、ディジタル・データ通信技術以外にも、「複雑化するニューロン・モデルに対応するために、ニューロン動作を CPU によって表現するのが現実的となった」との事情からも由来している.（CPU は、プログラムとデータを、クロック、又は、カウンタに従って一定量読み込んで演算を進めるので、基準 Timing に同期した動作となる.）

2.3. 非 CPU 方式（Fully-Parallel 方式）

多くの SNN 用のプロセッサは、ニューロンを表現する演算回路に設定するシナプス・パラメータ(W_μ)を複数セット用意して、1 個の CPU で複数ニューロンの電子回路表現としているが、今後、脳の中の多種の神経細胞の表現パラメータや役割/動作が明確となると、ニューロンを表現する演算回路をカスタム化し簡易化して、各々の搭載ニューロンに対応する演算回路をハードウエア上に実在させて、全ニューロンを平行動作させる方向が考えられる（Fully-Parallel 方式）. 前述の In-Memory Computing や

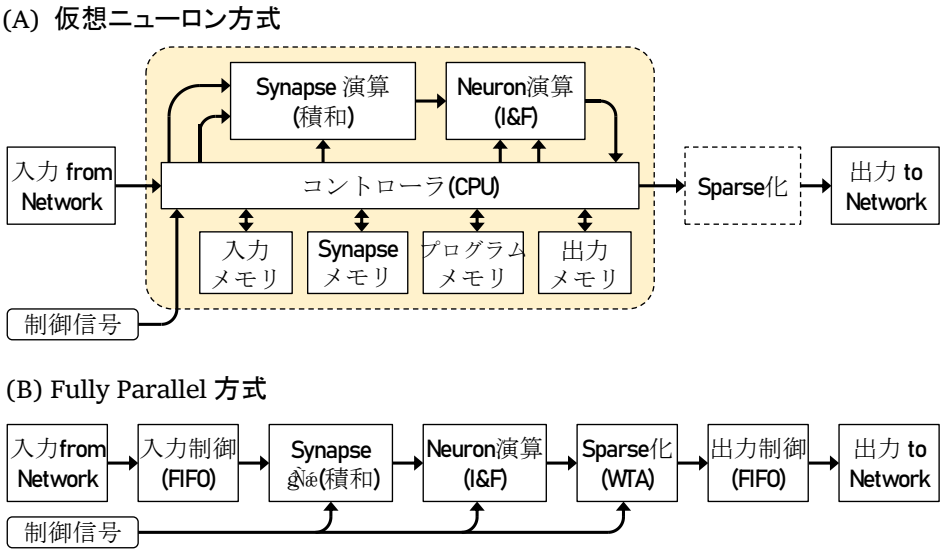


図４.
ニューロンを表現するコア回路の構成略図.

(A) CPU により、複数のニューロンを表現する場合.

(B) 表現するニューロン毎に回路リソースを持たせる場合.

Near-Data-Processing は、Fully-Parallel 方式の一種である.　但し、その場合も、ネットワーク表現に関しては、やはり、ディジタル・データ通信技術を用いることが現実的であるので、Encoder (Transcoder) を省くことはできない.

３. Winners-Takes-All と母集団

Spiking Neural Network は、2 章 2.1(2)で記したように、入力信号群がより Sparse 化すればより消費電力が下がる.

入力信号群を Sparse 化するための前段のニューロン集合のシナプス係数(W_μ)を調整する回路は、Winners-Takes-All (WTA)回路と呼ばれており、その具体的な実装として、ニューラル・ネットワークを使った Autoencoder が知られている[37][38][39][40].

WTA 回路は、教師無し学習における Classification 動作にて使われ得るが、必然的に、発火の強度を競い、Classification 動作に参加するニューロン母集団を定義しておく必要がある.

Fully Parallel 方式の時のその母集団を対象とした「Sparse 化のための Winner 検出機能」の位置付けを示すと、図５のようになる.

ここで、WTA 回路は、それ自体が別のニューラル・ネットワークとして、この母集団に付随する存在であることに注意されたい.　この付随するニューラル・ネットワークは工学的な工夫で追加された訳ではなく、教師無し学習を行うための必然性から追加されたネットワークである.

「発火確率」や「発火頻度」を定義する際にも、

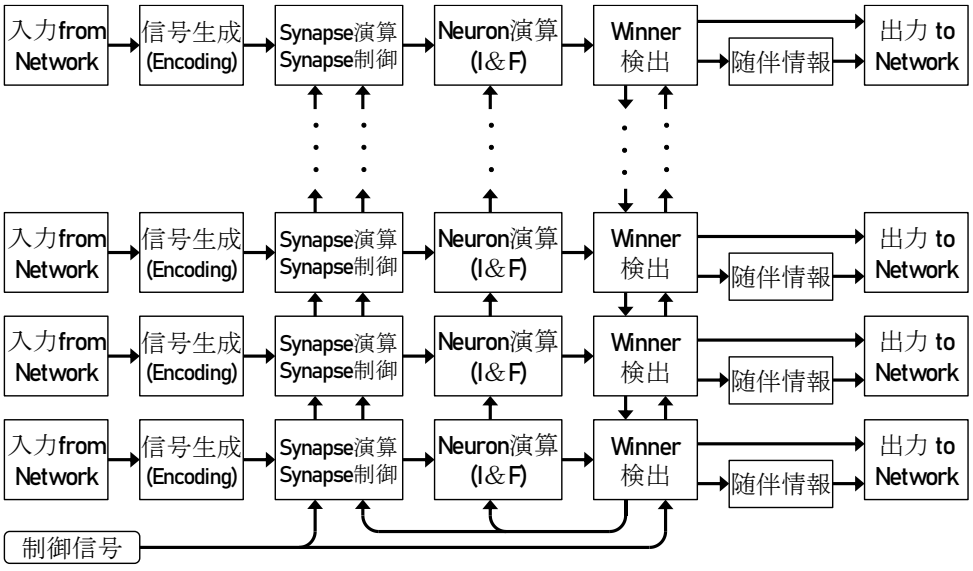


図５.
Winner-Takes-All 回路と、ニューロン演算回路、シナプス演算回路との関係.

Winner-Takes-All 回路は、ニューロン母集合の概念を必要とする.

WTA 回路がカバーするニューロン母集団と同様の母集団を定義する必要がある.

４．ANN とのベンチマーク

2014 年、①SpiNNaker[11]、②NeuroGrid [12]、③TrueNorth[13]、④BrainScales [14]等のニューロモーフィック・チップやそれを用いたシステムの発表がなされた. 更に、⑤Loihi[15]、⑥ODIN[16]と、多くの数の発表が続いている. （表１）

2015 年、米国 DOE は、「ニューロモーフィック・コンピューティングは、従来の CMOS 技術を dramatically に outperform するか？」とのテーマで開催した会議の結果を Report[17]にて以下のように結論付けた；

The development of a new brain-like computational system will not evolve in a single step; it is important to implement well-defined intermediate steps that give useful scientific and technological information.

ニューロモーフィックは、生体を模倣するための技術群であるが、回路技術、アルゴリズムに留まらず、関連するソフトウエア技術、人工シナプス素子に関する新規の製造技術や材料にも及ぶ技術群となってしまっている. それら全てが立ち上げることを必要とせずに部分的な採用を図り、現実的なロードマップにて、市場に受け入れられる姿を提示するできだとの合理的な内容であった. 但し、そのようなアドバイスが必要だったのは、「ANN に対する優位性を語るために、SNN は全てを結集するストーリを必要とする状況に陥っていた」ともいえる.

2018年に、N. Abderrahmanea[19]は、「既存のネットワークをSpiking方式に変換すると、搭載するメモリが減り、ハードウエア・コストが下がる. 」と評価していた.

但し、2019年に、K. Roy[18]は、以下の見解を述べている.

・Encoding技術によって、 従来のデータ・セットを使ってSNNを評価できるようになったが、それではSNNのメリットを引き出せない.

・SNNの実用的な価値については、長い間議論が続いているが、その為に、ニューロモーフィック・コンピューティングの開発は遅れており、ディープラーニングの急速な進歩によって、状況は悪化している.

・SNN では、強化学習の実装ができていない. 結果、ANN でトレーニングを行い得たシナプス係数を変換して SNN に適用しするという Conversion-based アプローチがとられている[27] [29]が、そのようにして、ANN と同等の精度を得ても、SNN の入力信号には Encoding 時間が追加となるので、Inference のレイテンシが伸び、エネルギー効率も下がる.

・SNN に、ANN の学習方法(BP 等)を実装し、ANN

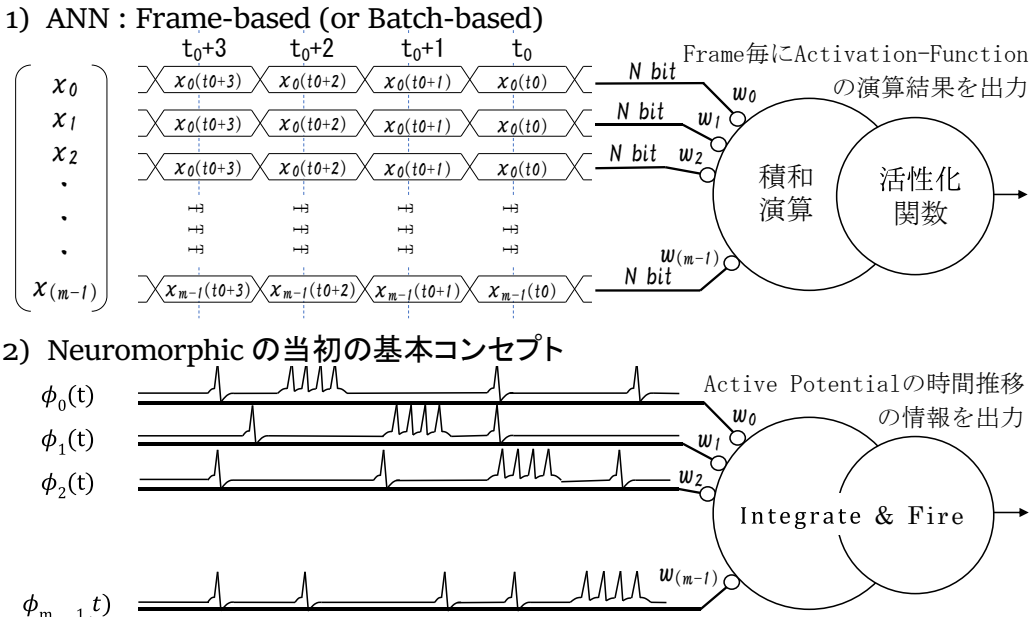


図６.
ANN ニューロンと、SNN のニューロンの基本動作の違い.

ANN はベクトルをバッチ処理し、SNN は各信号の時系列を処理する.

表１．SNN 方式を表現するための専用プロセッサ・チップ

Work	Technology	Application	Input data	Neuron model	I&F	学習	Network	SW言語
NeuroGrid_2014 Benjamin, et al.[11]	ASIC_180	Neuro Sciences	Spikes	Dimensionless model	Analog	(不明)	Programmable	NGPyton
TrueNorth_2014 P. A. Merolla, et al.[12]	ASIC_28	Classification	Frame-based	LIF	Digital	不可	Conv/FC/RNN	Corelets
SpiNNakerCMP_2014 S. B. Furber, et al.[13]	ASIC_130	Neuro Sciences	Spikes	LIF, IZH, HH	Digital	Program	Programmable	PyNN
BrainScales_2017 Karlheinz Meier [14]	ASIC_180 (Wafer Scale)	Neuro Sciences Classification	Frame-based	exp IF	Analog	STDP	Full Connection	PyNN
Loihi_2018 M. Davies, et al. [15]	ASIC_14	Classification	Spikes	CUBA LIF	Digital	STDP	Conv/FC/RNN	Loihi API
ODIN_2018 C. Frenkel, et al. [16]	ASIC_28	Classification	Spikes	Izhikevich	Analog	SDSP	Programmable	(不明)

用のワークロードでトレーニングし、ANN ベースの評価を行うというのでは、ハードウエアの進歩とネットワークサイズの拡張が著しい ANN に追いつくことはできない.

2020年に、L. Deng[20]は、以下のようにSNN-ANNのベンチマーク状況を表現している.

・SNNの認識精度はANNを上回ってはいない. 但し、それは、SNNのメリットを理解したベンチマーク・ワーク・ロードが無く、ANNベンチマーク用の静止画セットをSpiking用に変換してトレーニングし評価するという状況が続いていたためである. (「その状況では、Spatio-Temporalな信号を処理するSNNの性能をフェアーに評価し比較することはできない.」との意)

・既に、ANNの認識精度は人間を上回っているのに、Brain-Likeなメカニズムを導入して、ANNの性能を上回ろうとするには無理がある.

・ANNの成功は、成熟した学習モデル、多彩なベンチマークインフラ、プロセッサの性能向上が支えているが、SNNにはそれらの環境に劣る.

ハードウエア・コストと消費電力に関しては、多くの研究者が SNN の優位性を認めており、近年は、それらの利点を生かした Embedded-Use 向けの SNN の報告が増えてきている [34][35][36].

SNN 技術は、Signal 方式の信号技術をベースに新しいコンピューティング原理を目指して来たが、ネットワーク・トポロジーに新しい知見を追加するも

表２．SNN と ANN の比較 ： ※は、目標とする姿

	SNN	ANN
戦略	Event-Driven 非同期 LSM	Data Processing
入出力	Time Dependent Spiking Signal (Event-Driven, Asynchronous)	Frame Based Static Data (Vector)
ニューロンモデル	Integrate & Fire	McCulloch-Pitts の形式ニューロン (活性化関数)
シナプス・モデル (シナプス係数設定)	Dynamic Synapse (input Timing Dependent)	Static Synapse (Updated through Off-Lined Training)
State表現	Liquid State	Static State
演算器	(※) Integrated Memory-Operator	汎用ALU / 専用ALU
消費電力	1/10000 ～ 1/2	1 (Ref.)
予測精度	< 1	1 (Ref.)
Network トポロジー	(※) 自律学習により獲得	Designed by Human
課題	技術/製品のRoadmap化 Computerとしての明確化、 Software、それらの市場性	消費電力 / Computing Cost 学習の高速化 Network設計の自動化

のではなく、コンピューティング原理の発見という面では、一つの壁にぶつかったとの状況である.

SNN と ANN のネットワークが同じであるとすると、SNN の Spatio-Temporal な処理能力を生かすことはできなく、両者の性能に差が生じるとすると、ANN の活性化関数と、SNN の Integrate & Fire の違いが、静止画の分類や認識にどのような影響を与えるかの評価を行っているだけである.（表２）

５． SNN の次のステップに向けて

５．１　ミニコラムへの挑戦

大脳皮質には、ミニコラムと呼ばれる特徴的な繰り返し構造を持つ局所神経回路が存在することが知られており、長く注目されて来ている[6] [7][8][9][10].

窪田芳之は、2014 年、
「発火頻度が高く可塑性を持たない Fast Spiking (FS) バスケット細胞は、非錐体細胞の 40%を占め、錐体細胞の発火の制御、発火タイミングの制御を行うが、200-1000 個の錐体細胞を神経支配し、凡そ 500 個の前神経細胞によって神経支配されている. 等 」
と、蓄積した多くの知見をまとめた[6].

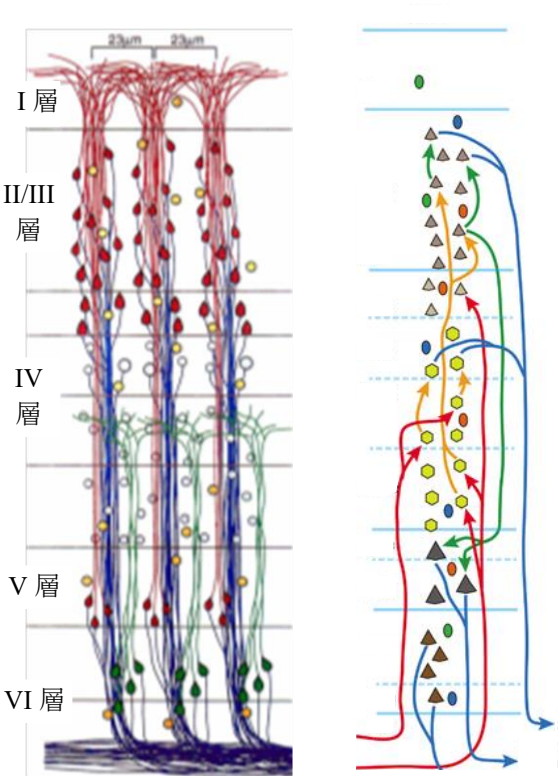
2015 年に、Markram.etal [23]は、14000 以上もの膨大な Wistar Rats のミニコラムの形態学的解析を行った結果を発表した. 形態と位置で分類すると 55 種、発火特性で 11 種に分け、それらのサブタイプとして 207 種ものニューロンを確認し、更にそれらのモデル・パラメータを抽出し、ミニコラムの形態学的再構成シミュレーションを行った[10].

これらから、ミニコラム（局所神経回路）は、内部に多種のニューロンを持ち、近隣のミニコラムからの制御を受け、錐体細胞経由で遠方に情報発信する存在であることは明らかである.

近年発表された SNN 用のプロセッサは、既に、その構造の実装に向けて、敢えて過剰ともいえる程のプログラム能力を持って開発されて来ている[15]. グリア細胞の実装の報告もある [44][45][46].

そこで、SNN を 1 個のニューロンをアトムとするネットワークではなく、１個のミニコラムをアトム回路とする形態を以下に考察する. ノード回路をより小さな電子回路で表現できれば、より多くのノードをチップに搭載できることとなり、結果ネットワーク規模を大きくできるとのメリットが期待できるからである.

図７. ミニコラムの構造



７図左： A Peters (1996) [7]より引用
７図右： 窪田芳之 (2014) [6]より引用

５．２　考察：ネットワークの観点から

ニューロモーフィックの戦略は、神経科学や数理科学、コンピュータ・サイエンスらの研究と、電子回路開発を同時進行させるという大胆な戦略であり、非常に難しい取り組みである. 電子回路開発は、産業上の価値観に左右されやすく、コスト/パフォーマンスを競う領域でもあるため、サイエンスによって解明される真実とギャップを生じやすく、結果として、ニューロモーフィックを卒業しての発明物でしかなくなってしまいがちである. 以下は、そのような打算の産物ではあるが、しかしながら、一つの現実的なアプローチではないだろうか？

1) ミニコラム回路モデルの作成方針

これは、ミニコラムの構造の知見と、その動作（入力に対する出力応答）を仕様としてまとめ、その動作を行う電子回路を考える、もしくは、その入力に対する出力のビヘイビアをアルゴリズムとして記述するとの作業である.（クリティカルな機能は、可能な限り回路化されている内部構造を持つ前提ではある.） これは、電子回路設計をボトムアップに進める場合の一般的な手法であり、手順である.

但し、入力に対して出力応答が組み合わせ論理と

して一意的に定まる通常の単純な電子回路と異なり、ミニコラムでは、動作が確率的となる．その理由は、

① 内部に、ダイナミックに特性を変化させる膨大なシナプス接合を持っており、その変動を正しく模倣することは困難であり、また、

② その動作振幅が小さく、熱雑音を加味しなくてはならない

からである．即ち、バラツキの設定や発生の機構を持ち、出力タイミング、及び、出力する量子量を、確率的に期待中心値の周辺に出力する動作をモデル化した回路であり、ビヘイビア・モデルである．

2）入出力信号のリストアップ

ミニコラムの構造は、第１層、第４層、第６層に、各々、大脳皮質内の他の領域、近隣のミニコラム、大脳辺縁系との間の多く神経線維を持つことが知られている．このことより、ミニコラムは、少なくとも３系統のネットワークと接続するノード回路であると想定しうる．ここでいう「系統」とは、ミニコラムがつながる第１層ネットワーク、第４層ネットワーク、第６層ネットワーク、等が、高速ディジタル通信技術で電子回路表現される前提で考える時の系統である．接続先のミニコラムがどこであるかは、通信パケットに設定する情報となる．

ミニコラムは、各ネットワークの系統に複数の接続数（つまり、ミニコラムの入出力端子数）を持つ可能性が高いが、外部端子とみなせるその数は、そのミニコラムが内部に持つ配線量やシナプス数に比べ数ケタ少ないため、リストアップ可能であることを期待し前提としている．

3）入出力信号の表現方法

ミニコラムから遠方へ発信される信号は、ミニコラム内の内部シナプス情報が確定的でないために、確率的に表現するしかないと想定される．ミニコラムは、近隣のミニコラムとの間で干渉的に動作するからである．その確率プロセスの表現方法と、確率情報としての入出力信号の表現方法を明確化する必要がある．

SNN アーキテクチャは、元々、情報を量子化して扱うという特徴を持っていた．Spiking 信号も、発火のタイミングと共に、発火頻度情報を伝える能力を持っていた．そこで、発火頻度の代わりに、発火確率を伝えるとすることが良い可能性がある．

4）内部回路の演算の表現方法

演算回路にとっては、整数の加算・積算は比較的シンプルに電子回路で表現できる．一方、小数点を扱う計算や、除算は大きな回路リソースを必要とする．従って、前記のミニコラム回路の入出力信号の表現方法においては、演算回路の簡略化の観点で検討が必要である．

CJS Schaefer, and S. Joshi は、2020 年に、整数ベースの演算とすることで、メモリ容量を 73.78%減らしても、精度の劣化を 1.04% に抑えることができたと報告[42]しており、伝達情報を全て整数とすることのメリットは非常に大きい．

６．むすび

ニューロモーフィック戦略の下で進化した Spiking Neural Network (SNN) は、音声やレーダー画像のように時間と共に変化する信号の表現とその信号処理に特徴があるため、Sensory-Motor 系への適用検討が先行していた[31][32] [33] .

しかし、SNN と ANN のベンチマークが示したように、ニューロモーフィック戦略下で人間を超えられるのかという問題も見えてきている．

そこで、SNN がミニコラムをアトム回路とする方向への展開を考察した．その場合、ミニコラムは複数系統の高速ディジタル・ネットワークに参加する素のレベルのネットワーク・ノードとなる．また、その場合、ネットワークは、ANN とは異なり非常にフラットとなる．フラットなネットワーク下では、「ネットワーク・ノード数（N）の二乗に比例して全体の価値が高まるとのメカトーフの法則が適用可能であるかもしれない」との期待がある．

SNN は、「入力信号をシナプス係数が表現するベクトルによってフィルタリングし、量子化された Spiking Signal を発する」とのミクロな動作の模倣が進んでいるが、脳の動作を模倣するとのレベルでのアーキテクチャに関する取り組みは未だ非常に少ない[32]．よりマクロな動作を表現しようとすると、ニューロン・タイプの余りの多様さと、コネクトームの複雑さが大きな壁である．

そこで、本論では、脳の全体を電子回路表現するとすると、この程度の粒度での表現が現実的ではないかとの候補としてミニコラムのビヘイビア表現を提案した．勿論、この内容には、ミニコラムから発信される素な情報が互いにどのように関連付けられ、我々の意識に上る記憶や経験、思考や知恵となるのかというメカニズムに欠けており、また、そもそも、このアプローチにて、ANN に勝る精度性能やレイテンシを得られるとの見通しはなく、コンピューティング原理としてのビジョンにも欠けている．内容は、あくまでもハードウエア検討戦略の考察である．

謝辞

本論執筆にあたっては、産業総合研究所の一杉博士と北海道大学院情報科学研究院の浅井教授に議論させていただき、様々に御示唆いただいたことに感謝する.

参考文献

[1] C. Mead; Neuromorphic electronic systems, in Proc. IEEE, vol. 78, no. 10, pp. 1629-1636. (1990)

[2] Wolfgang Maass; Networks of spiking neurons: The third generation of neural network models.(1997), in Neural Networks, 10(9), pp. 1659-1671.(1997)

[3] Wolfgang Maass, Thomas Natschläger, Henry Markram; Real-time computing without stable states: a new framework for neural computation based on perturbations, in Neural Computing, 14(11), pp. 2531-60. (2002)

[4] Maass W, Natschläger T, Markram H.; Fading memory and kernel properties of generic cortical microcircuit models, in J Physiol Paris. 2004 Jul-Nov;98(4-6):315-30. (2004)

[5] W Maass, H Markram; On the computational power of circuits of spiking neurons, in Journal of computer and system sciences (2004)

[6] 窪田芳之; 大脳皮質の神経細胞と局所神経回路, in 日本神経回路学会誌(2014)

[7] A Peters, C Sethares; Myelinated axons and the pyramidal cell modules in monkey primary visual cortex, in ournal of Comparative Neurology, Volume 365, Issue 2 (1996)

[8] Maruoka, et al.; Lattice system of functionally distinct cell types in the neocortex.in Science, NOV 3, (2017)

[9] 山川宏、荒川直哉、高橋亘一;全脳アーキテクチャに必要な新皮質マスターアルゴリズムの検討, in The 31st Annual Conference of the Japanese Society for Artificial Intelligence(2017)

[10] Henry Markram, et. al; Reconstruction and Simulation of Neocortical Microcircuitry, in Cell 2015, 163(2), pp. 456-92.(2015)

[11] S. B. Furber, et al.; The SpiNNaker project: A massively-parallel computer architecture for neural simulations, in Proc. IEEE, vol. 102, no. 5, (2014)

[12] Benjamin, et al.; Neurogrid: A mixed-analog-digital multichip system for largescale neural simulations, in Proceedings of the IEEE, 102(5), pp. 699-716. (2014)

[13] P. A. Merolla, et al.; A million spiking-neuron integrated circuit with a scalable communication network and interface, in Science, vol. 345, no. 6197, pp. 668-673. (2014)

[14] Karlheinz Meier; A mixed-signal universal neuromorphic computing system, in 2015 IEEE International Electron Devices Meeting (IEDM). (2015)

[15] M. Davies, et al.; Loihi: A neuromorphic manycore processor with on-chip learning., in IEEE Micro, 38(1), pp. 82-99. (2018)

[16] Charlotte Frenkel, et al.; A 0.086-mm2 12.7-pJ/SOP 64k-Synapse 256-Neuron Online-Learning Digital Spiking Neuromorphic Processor in 28nm CMOS, in IEEE Transactions on Biomedical Circuits and Systems journal. (2018)

[17] Schuller Ivan K. et al.l; Neuromorphic Computing - From Materials Research to Systems Architecture Roundtable. U.S., in Web. (2015)

[18] Kaushik Roy, et al.; Towards spike-based machine intelligence with neuromorphi computing, in Nature, Vol 575 (28), pp.608, (2019)

[19] Nassim Abderrahmanea, et al.; Design Space Exploration of Hardware Spiking Neurons for Embedded Artificial Intelligence, in Neural Networks Volume 121, pp. 366-386.(2020)

[20] Lei Deng, et al.; Rethinkig the perfomance comparison between SNNS and ANNS, in Neural Networks, Volume 121, pp. 294-307.(2020)

[21] Anthony Neville Burkitt; A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input, in Biological Cybernetics, 95(1), pp. 1-19. (2006)

[22] Wulfram Gerstner; Integrate-and-Fire Neurons and Networks, in The Handbook of Brain Theory and Neural Networks, Second edition, (M.A. Arbib, Ed.), Cambridge, MA: The MIT Press. (2002)

[23] Romain Brette; Philosophy of the spike: Rate-based vs. spike-based theories of the brain, in Frontiers in Systems Neuroscience, Vol 9, Article 151.(2015)

[24] Surya Narayanan, et al.; SpinalFlow: An Architecture and Dataflow Tailored for Spiking Neural Networks., in Computer Architecture (ISCA) 2020 ACM/IEEE 47th Annual International Symposium on, pp. 349-362. (2020)

[25] Srivatsa P, et al.; Spike Once: Improving

Energy-Efficient Neuromorphic Inference to ANN-Level Accuracy, in the 2nd Workshop on Accelerated Machine Learning (AccML). (2020)
[26] N. Abderrahmane, et al.; Neural coding and hardware architecture of spiking neural networks., in Euromicro conference on digital system design (DSD). (2019)
[27] Aboozar Taherkhania, et al.; A review of learning in biologically plausible spiking neural networks, in Neural Networks, Volume 122, Pages 253-272. (2020)
[28] Alan Diamond, et al.; Comparing Neuromorphic Solutions in Action: Implementing a Bio-Inspired Solution to a Benchmark Classification Task on Three Parallel-Computing Platforms, in Front. Neurosci. (2016)
[29] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy; Going deeper in spiking neural networks: VGG and residual architectures, in Frontiers in Neuroscience, vol. 13, p. 95. (2019)
[30] Vivienne Sze, et al.; Efficient Processing of Deep Neural Networks: A Tutorial and Survey, in Proceedings of the IEEE, Vol:105, No.:12, pp. 2295-2329. (2017)
[31] N. Abderrahmane, et al.; Information coding and hardware architecture of spiking neural networks, in 2019 22nd Euromicro Conference on Digital System Design (DSD). (2019)
[32] C. Eliasmith, et al.; A large-scale model of the functioning brain, in Science, vol. 338, no. 6111, pp. 1202-1205. (2012)
[33] S Yonekura, Y Kuniyoshi; Spike-induced ordering: Stochastic neural spikes provide immediate adaptability to the sensorimotor system, in Proc Natl Acad Sci USA, 117(22), pp. 12486-12496. (2020)
[34] Venkataramani, S.et al.; Efficient embedded learning for IoT devices, in 21st Asia and South Pacific Design Automation Conf, pp. 308-311. (IEEE) (2016)
[35] Charlotte Frenkel; Bottom-up and top-down neuromorphic processor design_Unveiling roads to embedded cognition, in PhD thesis of UCL-Université Catholique de Louvain. (2020)
[36] Bipin Rajendran, et al.; Low-Power Neruromorphic Hardware for Signal Processing Application., in the Special Issue on Learning Algorithms and Signal Processing for Brain-Inspired Computing, in the IEEE Signal Processing Magazine. (2019)
[37] Raphaela Kreiser, et al.; On-chip 教師無し学習 in Winner-Take-All networks of spiking neurons, in 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS). (2017)
[38] Wolfgang Maass; On the computational power of winner-take-all, in Neural computation. (2000)
[39] AlirezaMakhzani, Brendan Frey; Winner-take-all autoencoders, in Advances in Neural Information Processing Systems 28. (NIPS) (2015)
[40] T Fukai, S Tanaka; A simple neural network exhibiting selective activation of neuronal ensembles: from winner-take-all to winners-share-all, in Neural computation, MIT Press. (1997)
[41] G. Tang, et al.; Introducing Astrocytes on a Neuromorphic Processor: Synchronization, Local Plasticity and Edge of Chaos, in NICE '19: Proceedings of the 7th Annual Neuro-inspired Computational Elements WorkshopMarch 2019, No.:12 ,pp. 1-9. (2019)
[42] CJS Schaefer, and S. Joshi; Quantizing Spiking Neural Networks with Integers, in International Conference on Neuromorphic Systems, No. 11, Pages 1-8. (2020)