

解 説

ニューロモルフィック工学・脳型機械学習ハードウェアの行方

浅井 哲也

北海道大学 大学院情報科学研究科*

Whereabouts of Neuromorphic Engineering and Machine-learning Hardware

Tetsuya Asai

Graduate School of Information Science and Technology, Hokkaido University*

概要

精巧な神経ネットワークモデルを集積回路上に再構築してそこから逆に脳機能を解明しようとするサイエンス指向の「ニューロモルフィック工学」と、抽象度の高い脳モデルを専用集積回路や計算機・演算アクセラレータ上に実装して利活用を目指す応用指向の「脳型機械学習ハードウェア」。両者の変遷と最近の動向についてまとめ、今後それらの研究が何を目的にどこへ進むかについて考察する。

1. 背 景

ヒトと同じように対話ができるようなロボットを夢見る読者は多いだろう。著者もその一人であり、現在も同じ夢を見続けている。著者がこの分野に参入した1990年代初頭は、今で言う古典的なニューラルネットワーク研究（BP や連想記憶など）は少し下火になっていて、より抽象度を下げた本物の脳に近いモデル（例えば、オシレータ・スパイクベースのニューラルネットワークや、脳の階層構造を強く意識したニューラルネットワークなど）に研究トレンドが移行している最中であった。それら神経科学ベースの先端モデルの解析には様々な数学的・統計力学的テクニックが駆使され、我が師を含む理論計算を得意とする神経科学研究者はそれこそスーパースター的存在であったと記憶している。このような時代背景の中で、ニューロモルフィック工学^{1,2)}と呼ばれる新しい学問分野が誕生することになる。著者は、これまで神経科学ベースのニューロモルフィック工学研究、人工ニューラルネットワークの工学応用を目指した機械学習ハードウェア研究、および本業である集積回路システム研究の分野を渡り歩いてきた。それらの苦楽の経験から、近年のニューロモルフィック工学・脳型機械学習ハードウェア研究の動向を解説するとともに、可能性のあるハードウェアの未来・問題点を洗い出してそれら研究の行方を占ってみたい。

2. ニューロモルフィック工学の変遷

神経科学・モデル研究のトレンドの移行とほぼ時を同じくして、米 Caltech の Carver Mead¹⁾を中心に、神経科学ベースのモデルを集積回路上に忠実に再構築しようとする新しい研究・学問分野「ニューロモルフィック工学 (Neuromorphic Engineering)」が「鑄造」された^{1,2)}。この時期は、古典的なニューラルネットワークの応用限界が見えてきていたことと、それを実装するハードウェア技術が未成熟だったこと（かつ集積回路設計開発ができる機関が限られていたこと）が相俟って、応用視点の工学研究者および産業界はニューラルネットワークのハードウェア研究開発から撤退し始めた時期でもある。その時代の中で、ニューロモルフィック工学は、より脳に近いモデルを扱うことで古典的ニューラルネットワークの応用限界を打破できる可能性があったこと、およびこの頃から大学などの研究機関が設計した集積回路を (MOSIS⁴⁾と呼ばれる大学-企業間の仲介組織を通して) LSI ベンダーに格安で製造してもらえるようになり集積回路研究への参入が容易になった背景に後押しされて、将来の集積回路上の人工脳の実現を夢見た多くの若者達 (Carver Mead の子孫とも呼ばれる)・研究者がニューロモルフィック工学研究に参入した。

当初のニューロモルフィック工学の成果は視覚、聴

¹⁾ 集積回路工学の分野では LSI 設計ルールの父とも呼ばれ、後の LSI の長期傾向指標となる「ムーアの法則³⁾」の名付け親としても知られる。

覚などのセンサ周りの機能モデルを集積回路化したものが主であったが⁵⁻⁸⁾、後に不揮発シナプス⁹⁻¹²⁾や非常に精巧なニューロンモデルの回路^{13, 14)}も集積回路に実装できるようになり、それらを組み合わせた小規模模脳の開発¹⁵⁾など、その後も躍進は続く。しかしある時期を境に、ニューロモルフィック工学は脳機能やその要素を精巧に模倣し応用する研究から、脳シミュレータ・エミュレータの研究開発へ舵を切り始める。そのきっかけは、2008年から始まった米 DARPA と IBM を中心とする「SyNAPSE プロジェクト¹⁶⁾」である。このプロジェクトは、猫の知能をスパコン上でシミュレートするというフェーズ 0、脳チップのプロトタイプ試作を行なうフェーズ 1、続くフェーズにてより大規模な脳チップを創り出すという計画であった (IBM が開発した脳チップ “TrueNorth”¹⁷⁾はこのフェーズに含まれる)。ここで重要なのは、「脳チップ」ができたとは言っても実際には「脳型シミュレータのチップ」を作ったにすぎないことである。もちろん、適切なパラメータ、結合重み、学習なしにはこれらのチップに知能は生まれない。要は、このシミュレータチップを使って一番難しい知能化の問題を解決する下地を作ろう—知能化できるかどうかはわからないけれど、とりあえずシミュレータチップを先に作ろう—という話なのである。問題の先送り感があったものの、米国の巨大国家プロジェクトとして動き出した以上、ニューロモルフィック研究者の目には、脳機能の仕組みはさておき、知能の入れ物 (シミュレータのチップ) を先に作れば良いのだ、それを活用すればニューロモルフィック工学研究は持続可能なのだ、と映ったことだろう。2012年には欧州でもニューロモルフィック工学研究を含む巨大プロジェクト「The Human Brain Project (HBP)¹⁸⁾」が立ち上がり、この入れ物中心の風潮がさらに加速されることになる。

さて、ニューロモルフィック工学が脳シミュレータのチップに傾いてしまった本当の理由は何であろうか？ それは、本質的にはサイエンス指向のニューロモルフィック研究者の実際の活躍の場が工学の舞台であり、さらに集積回路の設計には膨大な時間とコストがかかるため、集積回路を作る以上、それが何か目に見えるかたちで役立たないと (作ったことの恩恵が得られないと)、研究自体を継続することが難しくなってしまう環境・時代変化によると筆者は考える。つまり、サイエンス指向のニューロモルフィック研究者が工学の舞台上で既存技術との競争を強いられるような時代に突入したのである。そのような中で、多くのニューロモルフィック研究者が、集積回路を試作して論文を出版す

ること、それをもとに継続的に研究資金を得ることに時間を費やしてしまい、集積回路を作ることの本当の目的、つまり何のために脳機能を集積化してそれで何をするのか、それで既存の物・技術に勝てるのか、について客観的に考える時間がなくなってしまった、あるいは考えないようになってしまったように感じる。このような枠組みには当てはまらない、そもそも自分の研究は基礎分野なので競争や応用など意識していないしその必要もない、という立場のニューロモルフィック研究者も多いだろう。しかし、それら基礎を踏み重ねて構築する将来構想 (出口までの道筋) は持つべきであり、国民の血税を投入した競争的資金 (科研費などを含む) を使わせて頂く立場の研究であればこれはなおさらのことである。ニューロモルフィック工学分野で大型国家プロジェクト資金を獲得できる研究者がそう多くないことが、この問題設定・出口設定の難しさを表わしている。

SyNAPSE プロジェクトや HBP のシナリオに沿った場合、ニューロモルフィック工学分野の出口はシミュレータチップとその利活用となり、このシナリオに沿った研究をすれば中期的な持続可能性はあるかもしれない¹²⁾。シミュレータチップのアーキテクチャ開拓やそのプログラミング環境構築だけでも一大研究テーマとなり得る。しかし一歩引いて考えると、シミュレータチップの先—つまり、それをどう利活用するのかはどうしても見えにくいのである。ある日突然、“すごい”モデルの種が出てきて、それがシミュレータチップ上で人工脳として機能し、現在の技術を圧巻する、という可能性は個人的には捨てたくないが、そうなるための道筋が見えにくい中で、どのようにマイルストーンを設定し資金を獲得し社会的波及効果を高めるための企業参入も含めた出口指向のプロジェクトを立ち上げるかは、政治と同じくらい、いやそれ以上に難しく思われ、自ずとプロジェクト代表者には高い政治力も要求されている。

3. ニューラルネットの逆襲とハードウェア

そのような中、2010年頃から「ディープラーニング²¹⁾」による機械学習システムが工学の世界・産業界にも広く浸透し始める。所謂、古典と呼ばれたニューラルネットの逆襲である。それまで、ニューロモルフィック工学が脚光を浴びる中、古典ニューラルネットを用いた機械学習ハードウェア研究は完全に下火となって

¹²⁾ HBP の目標設定・成果を疑問視する声も少なくなく^{19, 20)}、計画の大幅見直しが行われる可能性があることにも注意されたい。

いた。2008年にシナプスとして利用可能な不揮発メモリデバイス（メモリスタ）が発見¹³され²³⁾、メモリスタのクロスバー構造により全結合のニューラルネットワークを自然なカタチで構築できる糸口が見えた。多くのニューロモルフィック研究者はまずメモリスタを利用したSTDPシナプスに関する研究¹¹⁾へ目を向けたが、メモリスタを扱うデバイス研究者やその価値に一早く目をつけた工学研究者らは、古典ニューラルネットワークを対象とした機械学習ハードウェアの研究開発に目を向けた²⁴⁻²⁶⁾。後者の理由は、古典ニューラルネットワークベースの機械学習のほうが仕組みやできることが分かりやすかったからだろう。その最中のニューラルネットワークの逆襲である。ディープラーニングの先端学習アルゴリズムははまだ流動的であり、専用ハードウェア化の時期ではない（後述のとおり、ソフトウェア+アクセラレータで処理すべきである）。しかし、重みの保持と積和演算だけでも先に専用ハードウェア化できれば、ソフトウェア学習により得られた結合重みをメモリスタに書き込むことでディープラーニングの恩恵がすぐに得られ、現在のフォンノイマンボトルネックを打破する高速・低電力な機械学習デバイスが実現できる見込みが高いと考えたのである。これは、現時点では持続可能性の高い良い選択のように思える。

ディープラーニングの浸透は、神経科学分野のみならず、コンピュータ科学やデータ科学などの情報科学分野の大規模な新規参入も誘引した。米国ではGoogle, Microsoft, Intel, NVIDIAなどのIT先端企業がディープラーニング先端研究開発に携わっていることは読者もご存知だろう²⁷⁾。欧州や国内でもベンチャーから大手まで多数のIT先端企業が参画している。ディープラーニングの素直な応用（画像・音声認識など）から、時系列予測、広い意味での学習型計算機の新価値創出、およびそれらのハードウェア開発など、研究開発分野は多岐に渡る。一方で、学術分野においてはディープラーニングのコモディティ化の傾向も見られ、舞台は工学分野の応用・実用化フェーズに移ったという見方もある²⁸⁾。

このような状況での脳型機械学習ハードウェアには、研究開発途上故の「柔らかさ」が要求される。まず前提として、ネットワークの構造や学習アルゴリズムが日々変化（進化）する状況においては、せっかくハードウェア

を作っても（アルゴリズム進化により）それがあつという間に風化してしまうようでは、ハードウェア設計に投資する意味はない。ならばソフトウェアで研究開発をすれば良いということになるが、ニューラルネットワーク計算は規模によっては負荷が重すぎて実用にはならないことが容易に予測され、そうなった場合はハードウェアの力を借りる必要がある。そのハードウェアの代表格がGPU（グラフィックスプロセッシングユニット）/GP-GPU（汎用GPU）である²⁹⁾。GPUは、多数（数十～数千個）の固定または浮動小数点演算ユニットとメモリ（主に、演算ユニット内ローカルメモリ、スレッド単位共有メモリ、およびグローバルメモリの三階層）のアレイからなる演算アクセラレータで、従来はホストCPUの代わりにグラフィックス処理を担うものであったが、近年はOpenCL³⁰⁾、CUDA³¹⁾の普及も相俟って様々な演算を担っている。CPU単体の処理の数十～数百倍の加速性能を示す例も珍しくないが、そのプログラミングには並列演算の思考が必要である。演算器へのタスク割り当てもさることながら、最も時間がかかるメモリ参照・書き込みについて、メモリの共有範囲を慎重に考慮してなるべくアクセス時間の短いメモリを利用するように工夫をしないと、思ったような性能を出すことはできない。これは、マルチコアプロセッサや並列分散クラスタのプログラミング・スケジューリングや、実社会における人員・リソースのマネジメントと同じで、特殊な才能・思考が必要である。このマネジメントに注力し過ぎると、本来すべきこと—ディープラーニングで何をするのか—が置き去りにされがちなので注意されたい。幸い、ディープラーニングに関する並列演算については、Intel, NVIDIA等から一般ユーザにも使い易いフレームワーク^{32, 33)}が提供されており、演算規模がそのフレームワークで扱える範囲であればリソースマネジメントにユーザが頭を悩ます必要はあまりない。

現在のディープラーニングの研究開発プラットフォームはGP-GPUが中心である。万能を誇るGP-GPUではあるが、演算器とメモリ階層、バスの構造に起因する「頭の固さ」は隠し切れない。ある演算器と演算器の間にバスが1本あれば数十倍速くなるのだが、それが無いので一旦共有メモリに転送せざるを得ない、あるいは何がなんでも行列演算のかたちにするために殆どの要素が“0”の疎行列を使って無駄な演算をしなければならないなど、とにかくエネルギーを無駄に使う強引に演算を加速するのである。その結果、巨大な冷却装置と数百～千Wクラスの電源が必要になる。これでは、ディープラーニングがスマートフォンに載る

¹³ 実際には、“発見”とされたデバイスは当時既に存在していた抵抗変化メモリ（resistive RAM: ReRAM）であり、これは本来のメモリスタの定義²²⁾—電荷 q と磁束 ϕ の関係：“ $\phi = f(q)$ ”—を直接的に表すものではなく、その微分特性を示すデバイスであった²³⁾。

時代は来なさそうだ。

GPUの固さを克服しえる、ソフトウェアのようにハードウェアの構成を容易に書き替えられるような柔軟なハードウェアが存在する。FPGA(再構成可能ゲートアレイ)と呼ばれる半導体チップである³⁴⁾。FPGA自体はハードウェアであり、内部の配線構造を物理的に変更することは勿論できないが、FPGA内部には論理演算回路の集団とそれらを接続する配線およびスイッチが二次元状にびっしりと配置されており¹⁴⁾、演算回路内の論理関数(参照テーブルと呼ばれるメモリにより表現される)とスイッチの接続を決めるメモリ(これら両方を併せて「構成情報メモリ」と呼ぶ)の値を書き替えることで、論理関数とそれらの接続のしかたを自由に変更できる。よって、FPGA上でハードウェアを設計するという事は、FPGA内部の構成情報メモリを所望の接続になるように書き替えることに相当する。現在は、C言語などの高級言語に似たハードウェア記述言語^{35,36)}を用いて回路の動作を記述し、それをコンパイル・配置配線(どの動作関数をどこの演算回路に割り当て、それらをどのように接続するかを最適化)すると、構成情報メモリの内容がほぼ全自動で生成されるような設計環境が整っており^{37,38)}、ユーザがその構成情報をFPGAに転送することでFPGAが動作するようになる。実行すべきアルゴリズムが更新されたとき、ハードウェアを作りなおすことなく、FPGAの構成情報のみを更新すればよいので、専用チップを設計・製作するよりもはるかに早く安くできるのである。

このFPGA技術は、機械学習分野においては、ビッグデータを対象とするデータマイニングシステム(MicrosoftのWeb検索エンジン「Bing」³⁹⁾など)や、トレンドの変化によってアルゴリズムを頻繁に更新する必要があるかつ演算速度が要求されるような自動証券取引システムなどに積極的に取り入れられている。ただし、デメリットもある。まず、コア回路の動作周波数は数十MHz~数百MHzと低い。これはハードウェアに柔軟性を持たせたことに起因する。一般的には、回路の構成次第ではソフトウェア処理の約十~百倍程度高速化することは可能であるが、そのような性能を出すためにはデジタル回路・システムの設計センス(ハードウェアを強く意識したハードウェア記述)が必要であり、ハードウェア設計の未経験者には少しとっ

つきにくい。また、ソフトウェア的(柔らかい)とは言え、FPGA自体は半導体チップであり、その扱いには電子回路の測定・評価のセンス、泥臭く時間のかかる作業が必要な場合もある。また、開発時間(設計、検証、デバッグ)はチップの設計・製作(1年スケール)よりも短いとは言え、同規模の演算を行うソフトウェア設計のざっと十~百倍程度の時間がかかる。また、外部インターフェースの設計やマクロの利用にも、ローレベルのハードウェア知識(チップI/Oからデバイスドライバ開発レベルまで)が必要である。

ニューラルネットの逆襲劇により、少くとも画像・音声認識の分野では実用化に向けたハードウェア研究開発が進むだろう。現時点ではクルマの自動運転に向けた車載システム⁴⁰⁾や、より自然なスマートフォン向け音声アシスタント⁴¹⁾などへの応用の明確な道筋が見えている。応用先がはっきりしてくると、それに向けたハードウェアの実現可能性も見えてくる(ロードマップを作成できるようになる)のだが、この流れの中では、ハードウェア研究者の側からディープラーニングの新しい価値を生みだすような研究は困難かもしれない。ハードウェアで何を演算するかは、ソフトウェア先行研究の成果に依存する。ソフトウェアベースの新価値・応用創出とそのハードウェア開拓の二兎を追うことはなかなか難しく、どちらかに注力したほうが効率的と思われるかもしれないが、それでもハードウェア研究者としては、なんとかハードウェアの視点ならではの価値・応用創出をしたいものである。その可能性について次章で述べる。

4. マイコンと脳型機械学習

上述のメモリスト、GPU、FPGAに続く第4のハードウェアプラットフォームとして、著者は近年マイコン(マイクロコントローラ)に再着目している。マイコンが身近な存在になって久しいが、拡張の容易さ、PCとの接続の容易さ、手頃な価格、豊富な入出力インターフェース、高級言語によるオブジェクト指向プログラミングが可能になったことを考えると、機械学習の最初の価値創出のためのトライ&エラーのプラットフォームとして魅力的に思える。そのお手軽さが最大のメリットであり、特に実世界を相手にするアプリ(センサやロボット・アクチュエータ系、インターネットのエッジ側アプリなど)において、ユーザが思いついたことをすぐに試すことができ、PCとの通信もUSBケーブル一本(あるいはWi-Fi)で済むためデバッグも容易である。このマイコンの利活用の機会について考えてみたい。

¹⁴⁾ 2015年6月に米Intelにより167億米ドルで買収された米Altera社の開発による先端FPGA(Stratix V)には、約36万個の論理演算回路、約700個の高精度演算ブロック、約2,600個のブロックRAM(20KB)がチップに搭載されている。

まず、現代の計算機ハードウェアの発展の歴史を振り返ってみると、IBMのような巨大企業と対等にベンチャー企業が善戦していた。設立当初の米 Apple Computer 社がその典型例である。良い意味でのハッカー集団が強い動機のもと、あるいはホビーの一環として始めたハードウェア・ソフトウェア設計のムーブメントが、最初は草の根的に、やがては爆発的に広がるようになる。1980年代初頭のことである。今のようなビジネスや生活に密着した実用アプリがない時代、何がそんなに面白いのか、それで何ができるのかは、世間にはなかなか理解されなかった時代である。彼らがコンピュータの虜になった理由は、単にゲームをしたかったからという人もいたかもしれないが、多くはプログラムできるハードウェア自体に無限の可能性を感じたからではないだろうか（少くとも、筆者はそうであった）。それで何ができるか具体的にはまだ見えないが、可能性だけは無限にあるという様相は、ニューロモルフィック工学のシミュレータチップのそれと酷似している。シミュレータチップで何ができるかはまだ見えないが、脳のような機能を生み出せる可能性がある、ということである。とは言え、当時のマイコンと同じように、シミュレータチップをユーザが自在に（かつお手軽に）使えるようになるまでにはまだ時間がかかりそうであるし、何よりも当時のような爆発的なユーザ・コミュニティの増加に繋るような社会状況にはないため、たとえ様相は酷似していても二匹目のドジョウ（シミュレータチップの普及と復興）は狙えそうにない。

一方、同じ状況に“脳型”のマイコンが置かれたとするとどうであろうか？ 近年の「Maker ムーブメント⁴²⁾」の活性化とクラウドファンディングの浸透に後押しされるかたちで、1980年代初頭と類似の脳型マイコンムーブメントを起こすことは可能だろうか？ 筆者はそれが可能であると考えている。僅かなメモリしか持たないマイコンに実装可能なニューラルネットは非常に貧弱である。例えば各層あたり数十個程度のニューロンからなる2層の全結合ネットワークを載せたとして、ニューロンの入出力はあるプロトコルに従ってシリアル入出力できるものとし、学習は教師なし（例えばオートエンコーダを実装）または教師あり（例えば最も単純なパーセプトロン型）と仮定しよう。マイコン1個あたりのネットワーク規模は貧弱ではあるが、これを複数カスケード接続することで、層あたりのニューロン数と層の深さをユーザが自由に設定できるような構成にする。なお、システム全体の入出力と教師あり学習の制御は別マイコンが担当する。これはニューラ

ルネットであるので、このハードウェアをプログラムするということは、これら脳型マイコンの集団をどのように教育するかということであるが、このこと自体に新しさを感じる Maker 人口は少なくないだろう。教育次第では色々なモノになれる可能性と、プログラミング（教育）のしきいの低さ、豊富な入出力デバイスが入手可能であることを考慮すると、脳型マイコンの Maker コミュニティへの浸透可能性は高いと思われる。カスケード段数を多くすることで多機能化が見込める事例が出てくれば、複数マイコンの機能を FPGA 上に実装した柔らかい脳型 FPGA の製品開発に着手する良い動機となり、クラウドファンディングへの展開も視野に入る。その浸透により、実世界を相手にする様々な取り組みが強い動機とともに自発的になされれば、数十万人規模（潜在的にはその2桁以上）と言われる Maker コミュニティの中から面白いアプリ出てくる可能性、あるいはキラーアプリが見つかる可能性が高まる。こんなところから機械学習ハードウェアの新価値や新しい応用を拾えるのかという意見もあるかもしれないが、このような場を創ることもまた、脳型ハードウェア研究を活性化するために必要であると考えられる。何より、実物を扱う研究はワクワクするし、研究モチベーション維持のためのポジティブなエネルギーを与えてくれるのである。興味のある読者は、本業とは別に手を出されてみては如何だろうか。

5. 脳型デバイス研究について

半導体や分子などのデバイス研究は、情報処理応用をターゲットとした場合、自ずと計算機としての応用を考えることになる。現在の計算機の構成要素はスイッチ（電圧により ON/OFF を制御できるスイッチ）素子とメモリ素子、およびそれらの配線であり、現在それらはシリコン CMOS デバイス・プロセスにより具体化されている。新規デバイスにとって CMOS デバイスは超えねばならないモノであり、当然、大多数のデバイス研究がスイッチ素子やメモリ素子の実現を目指すことになる。この方向性は正しい。その中で、スイッチやメモリのみならず、計算機の構造そのものに目を向けるデバイス研究が増えてきている。その対象となっているのが脳である。

さて、情報処理の世界では、ハードウェア・ソフトウェアともに機能はボトムアップ的に拡張していくものであり、役に立つものを作ろうとすると普通はこうなる。一方、デバイス学の根底にある材料科学の世界では、機能はもともと材料が持っているか、あるいは突然あらわれる。よって、その材料で何ができるか、何

に应用できるか最初はわからない。何らかの機能が生まれた後で应用を探してゆくというトップダウン的な考え方が、特に新材料を扱うデバイス研究に浸透している。よって、機能創発に対するアプローチが情報の世界とデバイスの世界では相反するのである。情報処理回路とデバイスの分野は外から見ると同じに見えるかもしれないが、実際には分野間のギャップが大きい。この点については、回路設計からの歩み寄りの努力、特に若い世代がCMOS回路だけでなくポストCMOSデバイスの理解を深める努力が必要であると考え。スイッチやメモリ以外で具体的に「これが必要」というものを明確に示せれば議論が進むのだが、どのようなものが使えるのかを回路設計側が正しく把握しなければそもそも議論にならないのである。

デバイス研究の進展により、ニューロモルフィックな神経素子、シナプス素子およびそれらの基本的な配線技術が揃ったとしよう。そこまでは順調に進むと想像するが、その先については、先に述べたニューロモルフィック工学が直面しているのと同じ問題に直面するので注意が必要である。なお、それら要素演算に限っては、既存のデバイス・回路や計算機に対する利点は現時点ではまだなさそうである。一方、ディープラーニングの構成要素を新規デバイスで創るという場合は、出口は見えるものの、競合相手であるCMOSを凌駕する性能を出さなければ立場が苦しくはなるが、実用化フェーズ以前の基礎デバイス研究については、ディープラーニングや機械学習を出口目標として設定しても今のところ違和感はない。

最後に、新探求デバイス¹⁵をターゲットとした脳型計算機の階層構造について考えてみる。従前の情報処理システムは、電気電子素子を構成要素とし、その小～中規模な組み合わせにより要素回路層が構成され、さらに要素回路の組み合わせによるシステム層、システムの組み合わせによるアプリケーション層、という階層を持つ。新探求デバイスをターゲットとした脳型計算機も同じような階層を持つものと仮定すると、その基本構成要素は、スパイクを生成・伝搬させる神経膜に相当する素子（ニューロン・軸索・樹状突起デバイス）と、ニューロン同士を適切な強度で結ぶ可塑性接合素子（シナプス）、の二つであり、これらを階層の底辺（基盤層）とする。その上に、基本構成要素の小～中規模な組み合わせによるセンシング・初期情報処理や様々な要素神経情報処理（特徴抽出など）を行う中間層が

存在し、さらにその前処理階層の組み合わせた高度な脳型情報処理を行う上位層が存在するだろう。だとすると、基盤層と中間層はニューロモルフィック新探求デバイスが担い、上位層は（当面は）FPGAやCMOSが担う、という構図が現実的ではないだろうか。基盤層の要素と中間層における情報処理は恐らく今後不変である。また、実物の神経素子のゆらぎを考慮すれば、中間層ではCMOSほどの演算精度は不要と予測できる。よって、基盤層と中間層には「精度はさほど高くないがCMOSよりも高密度」というような新探求デバイスの活躍の機会がある。一方、出口となる上位層は、中間層と出口の両方の変化に適応可能であることが求められるため、FPGAのような「柔らかい」ハードウェアやCMOSを当面必要とするだろう。このように、新探求デバイスとFPGA/CMOSの棲み分けをしておくことは重要である。筆者の経験上、新しいデバイスを使って新しい情報処理に挑戦しようとすると、得てしてその試み自体に傾倒して出口設定や他デバイスとの接点をあいまいに考えがちになるからである。脳型新探求デバイスの研究に携わる方々には、そのデバイスで今できることのみならず、上述の上位層（出口）を意識しながらそれを支える基盤層・中間層を構築することにも是非目を向けて頂ければ幸いである。

6. ま と め

以上、少々駆け足ではあったが、ニューロモルフィック工学・脳型機械学習ハードウェアの変遷、およびそれらの可能性のある未来について論じてみた。

ニューロモルフィック工学については、出口設定が重要な課題となっており、シミュレータチップをシミュレータ以外の用途でどう活用するかを本気で考えなければならない時期に来ている。脳に相当する驚くような機能をシミュレータチップ上に実現する夢は個人的には持ち続けたいが、それまでどのようにして研究を持続させるか、よりポジティブには、どのようにすれば脳機能をシミュレータチップ上に発現させるかを戦略的に考える必要があり、個々の研究よりは、その戦略立てのための研究をすることのほうが優先されるべきだろう。

ディープラーニングのハードウェア研究は今後も躍進を続ける。ハードウェアアクセラレータのみならず、柔らかいハードウェアやマイコンを駆使した新価値創出のための開発インフラ・コミュニティ作りも重要であることを述べた。

最後に、脳型デバイス研究について述べた。CMOS

¹⁵ 新探求デバイス (Emerging Research Device) とは、CMOSの微細化に伴い直面する課題を打破するための新機能デバイスおよび新原理デバイス⁴³⁾のことを言う。

以外の素子を用いてそれを凌駕する演算システムを構築することは簡単ではないが、その主要因は、現代のブール代数に基づく計算方式そのものかもしれない。もしそうなら、集積デバイス学、回路設計学、システム設計学、計算科学などの分野を横断的に学び研究できる環境を、多くの若手研究者・学生に提供しない限り、脳型デバイス・アーキテクチャに関する大きなブレークスルーは期待できない（それぞれ個別の学問の専門性・知識を高めるだけでは駄目である）。これら問題解決のために、筆者が参画している国際半導体ロードマップ委員会 (ITRS⁴³⁾) 等における今後の議論とリーダーシップの発揮に注力したい。

謝辞

本解説に関わる研究の一部は、科学研究費補助金(20111004, 24360145, 25110015, 代表：浅井 哲也)の助成を受けたものである。

参 考 文 献

- 1) https://en.wikipedia.org/wiki/Neuromorphic_engineering/
- 2) Indiveri, G. (2004): *Neuromorphic Engineering*, Springer-Verlag.
- 3) Moore, G.E. (1965 and 1998): Cramming more components onto integrated circuits, *Proc. IEEE*, Vol.86, No.1, pp.114-117.
- 4) <https://www.mosis.com/>
- 5) Mead, C. (1989): *Analog VLSI and Neural System*, Addison-Wesley.
- 6) Mead, C., Ismail, M. (1989): *Analog VLSI Implementation of Neural Systems*, Kluwer Academic Publishers.
- 7) Mahowald, M. (1989): *An Analog VLSI System for Stereoscopic Vision*, Kluwer Academic Publishers.
- 8) Moini, A. (2000): *Vision Chips*, Kluwer Academic Publishers.
- 9) Diorio, C., Hasler, P., Minch, B., Mead, C. (1998): Floating-gate MOS synapse transistors. in *Neuromorphic Systems Engineering*, Kluwer Academic Publishers.
- 10) Snider G.S. (2007): Self-organized computation with unreliable, memristive nanodevices, *Nanotechnology*, Vol.18, No.36, 365202.
- 11) Linares-Barranco, B., Serrano-Gotarredona T. (2009): Memristance can explain spike-time-dependent-plasticity in neural synapses, *Nature Precedings*.
- 12) Jo, S.H., Chang, T., Ebong, I., Bhadviya, B.B., Mazumder, P., Lu, W. (2010): Nanoscale memristor device as synapse in neuromorphic systems, *Nano Letters*, Vol.10, No.4, pp.1297-1301.
- 13) Kohno, T., Aihara, K. (2008): A design method for analog and digital silicon neurons —mathematical-model-based method—, *AIP Conference Proc.*, Vol.1028, pp.113-128.
- 14) Indiveri, G., *et al.* (2011): Neuromorphic silicon neuron circuits, *Front. Neurosci.*, Vol.5, No.73, pp.1-23.
- 15) Neftci, E., Chicca, E., Indiveri, G., Douglas, R. (2011): A systematic method for configuring VLSI networks of spiking neurons, *Neural Computation*, Vol.23, pp.2457-2497.
- 16) <https://en.wikipedia.org/wiki/SyNAPSE>
- 17) Merolla, P.A., *et al.* (2014): A million spiking-neuron integrated circuit with a scalable communication network and interface, *Science*, Vol.345, No.6197, pp.668-673.
- 18) <https://www.humanbrainproject.eu/>
- 19) Theil, S. (2015): Why the Human Brain Project went wrong—and how to fix it, *Scientific American*, Vol.313, No.4.
- 20) <http://gendai.ismedia.jp/articles/45728>
- 21) 岡谷貴之 (2015) : 深層学習. 東京, 講談社.
- 22) Chua, L.O. (1971) Memristor—the missing circuit element, *IEEE Trans. Circuit Theory*, Vol.18, pp.507-519.
- 23) Strukov, D.B., Snider, G.S., Stewart, D.R., Williams, R.S. (2008): The missing memristor found, *Nature*, Vol.453, No.1, pp.80-83.
- 24) Cantley, K.D., Subramaniam, A., Stiegler, H.J., Chapman, R.A., Vogel, E.M. (2012): Neural learning circuits utilizing nano-crystalline silicon transistors and memristors, *IEEE Trans. Neural Networks*, Vol.23, No.4, pp.565-573.
- 25) Park, S., *et al.* (2013): Neuromorphic speech systems using advanced ReRAM-based synapse, *Proc. in IEDM*, pp.25.6.1-4.
- 26) Prezioso, M., Merrih-Bayat, F., Hoskins, B.D., Adam, G.C., Likharev, K.K., Strukov, D.B. (2015): Training and operation of an integrated neuromorphic network based on metal-oxide memristors, *Nature*, Vol.521, pp.61-64.
- 27) <http://www.theplatform.net/2015/08/25/a-glimpse-into-the-future-of-deep-learning-hardware/>
- 28) <http://business.nikkeibp.co.jp/article/report/20150217/277612/>
- 29) https://ja.wikipedia.org/wiki/Graphics_Processing_Unit
- 30) <https://ja.wikipedia.org/wiki/OpenCL>
- 31) <https://ja.wikipedia.org/wiki/CUDA>

- 32) Chetlur, S., Woolley, C., Vandermersch, P., Cohen, J., Tran, J., Catanzaro, B., Shelhamer, E. (2014): cuDNN: efficient primitives for deep learning, arXiv, arXiv:1410.0759.
- 33) <https://01.org/intel-deep-learning-framework>
- 34) <https://ja.wikipedia.org/wiki/FPGA>
- 35) Thomas, D., Moorby, P. (2002): The Verilog hardware description language, Springer-Verlag.
- 36) Lipsett, R., Schaefer, C.F., Ussery, C. (1989): VHDL: Hardware description and design, Springer-Verlag.
- 37) https://en.wikipedia.org/wiki/Altera_Quartus
- 38) https://en.wikipedia.org/wiki/Xilinx_Vivado
- 39) <http://www.theplatform.net/2015/08/27/microsoft-extends-fpga-reach-from-bing-to-deep-learning/>
- 40) http://car.watch.impress.co.jp/docs/topics/20150826_717652.html
- 41) <http://www.bloomberg.com/news/articles/2015-10-02/apple-acquires-artificial-intelligence-startup-to-boost-siri>
- 42) <http://makezine.jp/aboutmake>
- 43) <http://www.itrs2.net/>