

メモリズムプロセッサの技術解説書

目次

- [1. イントロダクション](#)
- [2. ノイマン型コンピュータの課題](#)
- [3. 前処理や複雑なアルゴリズム](#)
- [4. 今後の情報処理社会とメモリズムプロセッサ](#)
- [5. DBPプロセッサ](#)
- [6. XOPプロセッサ](#)
- [7. SOPプロセッサ](#)
- [8. 情報処理の分業化と理想のコンピュータ](#)
- [9. 特許実績](#)
- [10. これまでの理論検証に関して](#)
- [11. 半導体業界の動向](#)
- [12. 用語解説](#)

1. イントロダクション

メモリズムプロセッサは今後の本格的ビッグデータ社会、IoT社会、AI社会を見据えて開発されたプロセッサである。

特に情報処理が「超高速」である事と「超省電力」である事は、「情報処理上のあらゆる課題を解決」するといっても過言ではないほど重要な意味を持っている。

然しながら、「メモリズムプロセッサを正しく理解するのは難しい、馴染みにくい」と感じる方や、誤解をしている方も大勢いる。

本説明書は、以上のような背景からメモリズムプロセッサの要点を易しく解説するものである。

2. ノイマン型コンピュータの課題

現在我々が日常使っているコンピュータはノイマン型コンピュータである。

ノイマン型コンピュータは算術演算を機械化する事を目的として誕生した。

このコンピュータは汎用性があるので様々な分野で利用され、今日の情報化社会を築き上げた事は紛れもない事実である。

然しながら苦手な処理(効率が悪く時間のかかる仕事)も少なくない。

苦手な処理の代表的な処理は、「膨大な情報(データ)の中から意図するデータの検出を伴う情報処理」である。

「データの検出を伴う情報処理」の一例をあげれば、「検索・照合・認識・認証・クラス分け・並べ替え」などである。

ノイマン型コンピュータは、算術演算などの情報処理に都合の良い構成であるが、演算機能とメ

メモリが物理的に分離されている構成(アーキテクチャ)であるのでバスボトルネック問題が生じ、「データの検出を伴う情報処理」はバスボトルネックの影響を真面に受ける事になる。

3. 前処理や複雑なアルゴリズム

従って「検索・照合・認識・認証・クラス分け・並べ替え」などの処理を高速に行うためには、事前にインデックスを作成する事や複雑なアルゴリズムを用いて情報処理をする事で対処する以外ない。

然しながらこれらのインデックスやアルゴリズムは、前処理や更新処理が必要であるので、リアルタイム性が犠牲になるばかりでなくシステム全体のパフォーマンスや電力性能を劣化させる結果となり、専門性も高くなるので恒久的な解決策とは言えない。

先に説明したバスボトルネックの問題は、現在のコンピュータの誕生間もなくの頃から問題視されてきたが、これまで有効な解決策もなく放置されてきた。

いわば急場凌ぎのようなコンピュータの利用方法となっている事に留意する必要がある。

4. 今後の情報処理社会とメモリズムプロセッサ

「検索・照合・認識・認証・クラス分け・並べ替え」などの「データの検出を伴う情報処理」は、ビッグデータ、IoT、AIなど今後の情報処理社会に深く関わるので、性能や機能の向上が不可欠である。然しながらムーアの法則で知られる半導体の微細化技術が限界に近づき、従来型プロセッサ(古典的プロセッサ)であるCPUやGPUの継続的な性能向上が期待出来ず、新しい考え方の情報処理プロセッサが必要になる。

「データの検出を伴う情報処理」に適した構成(アーキテクチャ)にするためには、メモリと演算機能を一体にする以外方法はない。

メモリズムプロセッサは、以上のような問題を解決するために発明されたメモリと演算機能が一体となったプロセッサである。

CPUやGPUは、算術演算と論理演算が可能なALU(Arithmetic and Logic Unit)や、浮動小数点演算が可能なFPU(Floating Point Unit)の2種類の演算機能で構成されている。

これらの演算機能はトランジスタ数が多く、高コストで消費電力も大きいので、1つのチップにCPUの場合数十個、GPUの場合数千個の実装が限界となる。

このように、高コストで消費電力の大きな演算機能に「検索・照合・認識・認証・クラス分け・並べ替え」などの単純処理の繰り返しは、CPUやGPUに勿体ない処理をさせている事になり留意する必要がある。

一方メモリズムプロセッサの演算機能は「検索・照合・認識・認証・クラス分け・並べ替え」に特化し、トランジスタ数が少なく、省スペース・低コスト・省電力であるので、1つのチップで百万個を超える超並列処理も可能となり、「検索・照合・認識・認証・クラス分け・並べ替え」に向けた構成となっている。

メモリズムプロセッサは、今後の情報処理の進化に不可欠な情報処理である「検索・照合・認識・認証・クラス分け・並べ替え」などの処理を効率的に実現出来るよう構成されたコプロセッサやアクセラレータと考えればよい。

従来形プロセッサであるCPUやGPUは、「OS処理、算術処理、制御処理、通信処理」が得意であるのでこれらの処理を行い、「検索・照合・認識・認証・クラス分け・並べ替え」などの苦手な情報処理を、コプロセッサやアクセラレータであるメモリズムプロセッサに肩代わりさせる事が可能になる。

メモリズムプロセッサは、対象とするデータ(情報)の種類の違いにより以下の3つのタイプのプロセッサが用意されている。

5.DBP(DataBase Processor)プロセッサ

DBPは汎用データの検索や照合を高速化する事を目的としたプロセッサである。

従来のCPUとメモリによるデータの検索や照合は、CPUによるメモリのスキャンが必要になる。

例えばメモリに記憶されたデータの中から、「オール0」のデータを検出する事を考えてみよう。CPUにより「オール0」のデータを検出するためのスキャンは、処理の大半(例えば99%)が意図するデータではない、言い換えれば99%が無駄な仕事になっている。

この問題を解決するためには、インデックスを事前に用意するなどの前処理が必要になり、真の意味のリアルタイム性が失われる事になる。

さらにこのインデックスは、基データが変化する度に更新をする必要がある。

この更新処理は大きな負荷となるので、更新処理を高速化するには高性能な(電力消費の大きな)CPUを使用する必要がある。

従って検索システムは大型化し、消費電力が大きく、専門性も高くなるので、高価で開発期間も長期間になる。

DBPは以上のような情報処理上の問題を根源的に解決するプロセッサである。

DBPはメモリ上のデータの行列配列を反転する考え方で、データの列幅を大幅に拡大する事と超省スペース(超低コスト)の演算器をメモリチップ上に搭載させる構成のプロセッサである。

この構成とする事により、インデックスが不要になり、電力が削減され、超高速にデータの検索や照合が可能になる。

またDBPを必要個数分並列接続して利用する事で任意のデータサイズにスケールアップ可能である。

DBP自身が検索・照合を自己完結型で処理するので、DBPの数が増えても(データ容量が増えても)一定の時間で処理が可能になる。

従ってDBPを使用した検索システムは小型軽量、省電力化されて専門性も低くなるので、低コスト、短時間でシステムを提供する事が可能になる。

6.XOP(Cross Operating Processor)プロセッサ

XOPは2組のデータ群を効率的に比較する事を目的としたプロセッサである。

CPUによるn個のデータとm個のデータ群の比較演算は、 $(n \times m) / 2$ 回の演算が必要になり、nならびにmが大きくなるとCPUの負担は膨大になる。

XOPはこのようなCPUによる比較演算の問題を解消するためのプロセッサである。

チップ内には、XデータならびにYデータを記憶する2組のメモリ部と、2組のメモリからのデータ線の交点にデータを比較演算する演算器が大量に組み込まれている構成である。

一例として、XYがそれぞれ1Kのメモリであれば、1M(100万)個の演算器が組み込まれる事になる。

そして、XならびにYのデータが一致もしくは類似(マッチ)している場合に、そのアドレスを外部に出力する構成である。

XYのメモリの容量を増やす事により、1K×1Kの演算のみならず、1億×1億などのデータ比較を可能にし、1M個の演算器が並列に比較演算を行うので、CPUによる逐次比較を根源的に進化させる事が可能になる。

XOPの2組のデータのうち1組を既知のデータとする事により、ヒストグラムの作成、ソーティングの実行、相関の検証、クラス分けなど、CPUの負担の大きな処理を肩代わりする事が可能になる。

7. SOP(Set Operating Processor)プロセッサ

SOPはパターン認識技術の標準化や汎用化を目指したプロセッサである。

パターン認識技術は、知識処理の原典とも呼ぶべき技術であると広く理解されているものの、CPUやGPUはパターン認識が苦手である。

その原因の一つとして、パターン認識技術の基本的な手段はパターンマッチングであるが、CPUやGPUによるパターンマッチングは効率が悪く利用する事が出来ないからである。

パターンには文字列や時系列データなど1次元配列からN次元配列までの配列がある。

これまでこれらの配列データに対するパターンマッチングの利用について、1次元配列である文字列は正規化表現で知られるパターンマッチングが広く利用されている。

正規化表現は文字列のみでなくDNAの塩基配列の解析などにも利用されている。

然しながら、2次元配列である画像のパターンマッチングは古くから利用されてきたが、現在はテンプレートパターンが得やすい物体の位置決めや外観検査などの分野に利用されるのみとなっている。

その原因は、画像データのパターンマッチングを行うためのテンプレートパターンの作成が複雑で最適なパターンが求めにくい事、そして致命的な原因は、CPUやGPUによるパターンマッチング処理は時間がかかり過ぎる事である。

従って、顔の検出にはHaarアルゴリズム、人体検出にはHOGアルゴリズム、汎用検出にはSIFTなどの複雑なアルゴリズムを駆使するか、深層学習のような機械学習に頼らざるを得ない状況となっている。

然しながら「検出や認識の対象物」そして「検出や認識の目的」に適したアルゴリズムや機械学

習を選択して利用する必要があり専門性が高く、開発期間、開発コストに課題がある。

SOPは以上のような問題を解決するために、これまで困難とされていたパターンマッチングをハードウェアで実現し、パターン認識技術の根源的な進化と標準化を目的としたプロセッサである。

SOPは1次元からN次元のパターンマッチングが可能であるが、ニーズの多い2次元配列の画像のパターンマッチングについて説明をする。

SOPは従来の集合演算(ブール演算)を「位置を加えた集合演算」に拡張して集合演算を実現するものである。

位置を含めた集合演算とする事により、画像上のエッジ、コーナ、領域、パターンなどの検出=広義のパターンマッチングが可能になる。

つまり画像上の様々な物体そのものや特徴を、複雑なアルゴリズムを利用する事や学習をさせる事なく超高速に検出する事が可能になる。

SOPをASICで実現した場合、試作レベルのASICでもノートパソコンのCPU(TDP15W程度)に比較して1万倍以上高速にパターンマッチングが実現出来る事が確認されている。

従って1回の集合演算(広義のパターンマッチング)は数 μ 秒程度となり、しかも集合演算時の電力は1W程度、非演算時の電力はほぼ0Wとなる。

今後ASICの製品化を目指す事により100万倍もの電力性能が期待される。

超高速で広義のパターンマッチングが可能であるので、画像認識分野で従来から問題とされているテンプレートパターンの最適化問題や画像の回転やスケールリングに伴う問題をクリアする事が可能になる。

また超高速でしかも超省電力であるので、従来の機械学習や深層学習では実現困難な新しい認識手法を提供する事が可能になる。

1次元のSOPは正規化表現による文字列データやDNAの塩基配列のパターンマッチングによる高速な検出(認識)や解析、3次元のSOPは立体空間、例えば分子構造や日常空間のパターンマッチングによる高速な検出(認識)や解析に利用する事が可能である。

8. 情報処理の分業化と理想のコンピュータ

アムダールの法則が示唆する通り、情報処理システムの一部が並列化によって高速化されても情報処理システムの性能向上は余り期待出来ない。

システムの多くの部分が並列化して高速化される事により、情報処理システムの性能は大幅に向上する事になる。

以上のような背景から、メモリズムプロセッサは3つのタイプのプロセッサが用意され、3つのプロセッサを組み合わせして利用する事により多くの効果が発揮出来る事になる。

従来のコンピューティングがクラウドなど分散処理であったのに対し、メモリスムプロセッサを利用するコンピューティングは、専門のプロセッサが得意な処理を実施する分業型コンピューティングに進化する事を意味する。

従来型プロセッサとメモリスムプロセッサがコンビネーションしたコンピューティングにより、これまでの情報処理に比較して様々なメリットを提供する事が可能になる。

ASIC化したメモリスムの場合の性能やメリットの一例を示す。

- ・「データ検出を伴う処理」の処理性能が数千倍から数万倍以上の向上
- ・「データ検出を伴う処理」の電力が数十分の1などの省エネ効果
- ・「データ検出を伴う処理」の性能向上による精度や速度などのAI性能の向上
- ・「データ検出を伴う処理」の省エネ効果により軽薄短小性の向上
- ・インデックスを作成する事や複雑なアルゴリズムが不要になるので、システムや製品の開発期間や開発コストを圧縮する事が可能
- ・半導体チップ化により、システムや製品のコストが削減し経済性が向上

以上の様な情報処理上の様々なメリットにより情報処理(ソフトウェア)の標準化が図られ、ソフトウェアが人に優しく(易しく)なり情報処理の進化が加速する。

これが筆者が考えるノイマン型コンピュータの理想の姿である。(用語解説も参照の事)

9. 特許実績

DBP、XOP、SOPとも国内外の特許を取得済み。

10. これまでの理論検証に関して

DBP、XOP、SOPの特許理論はFPGAにより検証済みである。

画像認識用2次元SOP(2D-SOP)はASICのRTLやGDSなどのIPが完成済みである。

11. 半導体業界の動向

半導体の微細化技術が限界に近づきCPUやGPUの性能向上が期待出来なくなった今、世界中で量子コンピュータなどの新しいコンピュータの研究が加速している。量子コンピュータの場合、実現出来る処理が組み合わせ問題の解決などに限定されている事や商品化するまでの時間が問題である。

このような背景から、メモリと演算機能を一体化したインメモリコンピューティングやニアメモリコンピューティング技術の研究が活発化している。

残念ながら日本でのこれらの研究は皆無の状況である。

現在のコンピュータの苦手な処理の代表格である「データ検出を伴う情報処理」を肩代わりする事を目的とし体系付けされた3つの種類のプロセッサによるインメモリコンピューティングやニアメモリコンピューティング技術は、全世界で弊社以外ない。

また、メモリズムプロセッサは現在の半導体生産ラインで直ぐに生産出来るので、半導体メーカーにとって設備投資の負担が無い事も大きな魅力であり、メモリズムプロセッサは安価に市場に供給する事が可能になると考えている。

12. 用語解説

メモリズムプロセッサに関連する説明であるので、必ずしも用語の意味を正確に伝えられていない事にご注意。

ノイマン型コンピュータ

現在我々が日常的に使用しているコンピュータである。

このコンピュータが発明される切っ掛けは、大砲の弾道などの計算を機械化し精度を上げるなど軍事的なニーズによるものである。

イギリスの数学者アラン・チューリングの示した計算手法(チューリングマシン)を基に、現在のコンピュータの原型を作った米国の数学者フォン・ノイマンに敬意を表してノイマン型コンピュータと呼ばれている。

非ノイマン型コンピュータ

量子コンピュータなどノイマン型コンピュータ以外のコンピュータの総称。

CPU(Central Processing Unit)、**中央処理装置**とも呼ばれる。

情報処理の中核をなすユニット(デバイス)であり、算術演算のみならずあらゆる情報処理に利用されている。

後述するノイマンバスボトルネックがあるので苦手な処理も少なくない。

現在のCPUは、マルチコア化による情報処理の高速化とキャッシュメモリによる情報処理の効率化が鍵になっている。

今後微細化技術の限界により、CPUの性能は上がりにくくなってくるものと考えられる。

メモリ(Memory)

情報(データ)を記憶する半導体デバイス。記憶するデータは、CPUに与える命令データと、データベースのデータや画像や音声などの利用データの2種類がある。

現在汎用的に利用されるメモリはDRAM、SRAM、FLASHメモリの3種類であるが、より高速なメモリと不揮発性のメモリのニーズが多く、様々なタイプのメモリの研究が進められている。

メモリズムプロセッサはDRAM、SRAM、FLASHメモリで実現可能であるが、半導体メモリであれば新しいタイプのメモリに組み込む事も可能である。

ノイマンバスボトルネック

CPUとメモリが物理的に分離されている事に起因して、CPUとメモリ間のデータの授受が制約を受け、CPU本来の性能を発揮出来なくなる事象の総称。メモリズムプロセッサは、この問題が

最も影響を及ぼす「データの検出」を解決する事を目的としたプロセッサである。

GPU(Graphics Processing Unit)

元々は画像処理を高速に実現する事を目的に開発されたが、最近はAIなどの処理に欠かせない算術演算を高速に実現するためのデバイスとして広く利用されている。

1チップに数千個の演算器を搭載する事で算術演算が高速化されるが、大電力消費が課題であり省電力化の研究がなされている。

データ検出のためのインデックス

特定のデータを高速で検出するためには、BTreeや転置インデックスなどのインデックスを利用する必要がある。

これらのインデックスを用いる事により検索性能は大幅に向上するが、基のデータの変化に追従して更新処理を行う必要がある。更新処理は大きな負担になる事が多いので、システムとしてのパフォーマンスを劣化させる事になる。

データ検出のための複雑なアルゴリズム

画像の認識を行うためのアルゴリズムとして、顔の認識はHaarアルゴリズムが利用されている。汎用の画像認識は局所的特徴量に注目したHOGアルゴリズムやSIFTアルゴリズムが利用されている。いずれのアルゴリズムも複雑で専門性が高い処理となっている。

機械学習

通常の情報処理は、プログラムによりCPUに命令を与える事で実現されるが、複雑性が高くプログラムを作成出来ない情報処理、例えば認識処理機能においては、その機能に沢山のデータを学習させる事により認識処理機能を実現させようとするものである。

認識する対象物やその用途において様々な手法が存在し、利用可能な状態に最適化するまでには試行錯誤を繰り返す必要があり、専門性が高い。

深層学習(Deep learning)

ニューラルネット方式の機械学習の一つである。

深層学習の最大の特徴は特徴量の自動獲得であり、この特徴により従来型の機械学習より認識精度を向上出来る事から広く利用されている。

深層学習の課題として、学習のためのコンピューティングの負担が大きく、GPUなど演算パワーの大きなプロセッサを利用する必要がある。

また深層学習の手法も様々であり、標準化出来るレベルに至っていない。

メモリプロセッサ

ノイマン型コンピュータの苦手な処理(=CPUやGPUが苦手な処理)の代表的な処理である「情報(データ)の検出を伴う処理」をCPUやGPUから肩代わりするためのプロセッサ。

ビッグデータ、IoT、AIなど今後の情報化社会に必要な不可欠なデバイスとして期待されている。

インメモリコンピューティング

ノイマンバスボトルネックを解消するために、メモリチップの中に演算機能を埋め込んだタイプの半導体チップによるコンピューティングの総称である。

ニアメモリコンピューティング

ノイマンバスボトルネックを解消するために、メモリの近くに演算機能を埋め込んだタイプの半導体チップによるコンピューティングの総称である。

半導体微細化技術の限界が間近になりCPUやGPUの性能が上りにくくなった現在、これらの半導体チップの研究が活発化している。

メモリズムプロセッサは、これらのコンピューティングに類するものである。

補ノイマン型コンピュータ

CPUとメモリを使ったノイマン型コンピュータ以外のコンピュータを、一律に「非ノイマン型コンピュータ」と呼ぶ傾向があるが、メモリズムプロセッサはノイマン型コンピュータの苦手な処理を肩代わりするコンピューティングであるので、「補ノイマン型コンピュータ」と呼ぶのが相応しい。

ノイマン型コンピュータの理想

コンピュータの利用方法について考えてみる。

コンピュータは一般の人々が日常的に利用する用途と、気象庁が利用する天気予報、地震の際の建築物の強度計算や創薬のためのシミュレーション、更には暗号処理や組み合わせ問題など専門家が利用する特殊な用途に大別される。

専門家が求めるコンピュータは算術演算などの更なる進化が求められているが需要は限られている、必要性が高いものはスーパーコンピュータやハイパフォーマンスコンピュータ(HPC)、場合によっては量子コンピュータを利用すればよい。

一方メモリズムプロセッサが目指す「データの検出処理」はメールの検索、パソコン内の古いデータの検索、更にはWEB検索、そして画像や音声などの認証など、世界中の人々が日に何回も利用する処理、利用したい処理である。

以上のような背景から筆者が主張するノイマン型コンピュータの理想は専門家が求めるコンピュータではなく、一般ユーザから専門家まで市場性の高い機能に焦点を当てた理想のコンピュータである。

これまで「情報(データ)の検出を伴う処理」に関する問題解決は未着手のままとなっていた。その要因は、従来の情報処理技術の常識ではインデックスや複雑なアルゴリズムは水や空気のように在って当たり前な存在であるとともに、CPUやGPUに勿体ない情報処理をさせている事に気が付かなかつたからと思われる。

仮にこの事に気が付いても、その解決方法を生み出すまでに至らなかったものと推測する。

この事実を受け入れノイマン型コンピュータの宿命とも言うべきバスボトルネックの問題を解決する事がノイマン型コンピュータを理想の姿にする事になる。