

Nvidia の成功に学ぶ(その 6)

2021年4月21日 | [岡島義憲の集積回路の明日に向けて](#)
(情報統合技術研究合同会社)

今回のブログでは、米 Nvidia 社の直近の市場認識と事業戦略に関する筆者の注目点をまとめる。同社は、2月24日(米国時間)に発表した2021年度期の決算概要にて、売上高が前年度比53%増(166億7,500万ドル)、営業利益が同59%増(45億3,200万ドル)、純利益が同55%増(45億3,200万ドル)と、大幅な増収増益を報告した([参考資料 1](#))。過去5年間に、売上額を急増(5年で3.55倍)させた最大の要因は、データセンター向けGPUと、ゲームや仮想通貨のデータマイニングの汎用GPUであるという([参考資料 2](#)、[3](#))。ウェーハファウンドリ企業からのウェーハ供給が滞っている状況でも、同社業績は驚異的に伸びている。

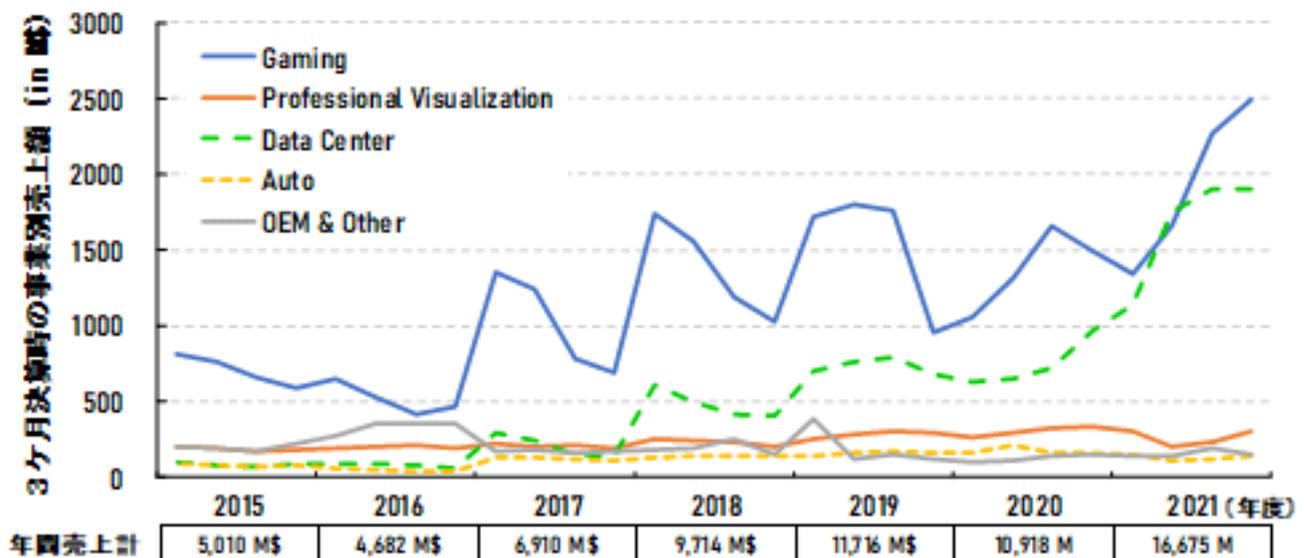


図1 Nvidia社の事業分野毎の3ヵ月決算の推移(同社の年次決算は、1月末)
2020年後半より、データセンター向け売り上げが急増している。

出典 Nvidia社が米国SECに提出した[各年のFORM-10K資料](#)より、筆者が作成した。

ホモジナスな並列化によるアクセラレーション戦略

同社の技術戦略は、この20年間一貫して「GPU(並列データ処理アクセラレータ回路)によるCPUの処理の高速化を汎用化する」とのことであった。General-Purpose GPU (GPGPU)のためのソフトウェア環境整備を進め、GPUのアプリケーションを画像処理以外の領域にも広めた。「全てのCPUにGPUを付ける」を目指すかのような勢いである。同社のArm社買収の動きも、そのように考えると分かり易い。CPU設計と協調すると、パワーマネジメントの裁量の余地が更に広がる。

但し、アクセラレータ回路コアに実装すべき要素回路は、処理するデータや適用するアルゴリズムによって違いうる。従って、同社は、PC ゲーム用(汎用 GPU)、Professional Graphics 処理、データセンタ用、等のマーケット・セグメントに適合させた GPU 商品をラインナップしてきている。

それらの中でも、現在最も強化しているセグメントは、「データセンター用」だろう。昨年 5 月に発表した「A100(TSMC 社の 7nm プロセス採用)」は、チップ面積 826mm²、搭載トランジスタ数 540 億以上、HBM (High Bandwidth Memory) の容量 40GB、2 チャンネルの HBM からの合計転送バンド幅は 1.56TB/s という驚くべきチップである。スパコンから AI をカバーし、ディープラーニングにて主流とみられる半精度(16 ビット)計算の時に、1 チップで 1 ペタ Flops 以上のスペックを持つ([参考資料 4](#))。

同社の GPGPU は、「ホモジニアスな並列化による CPU 動作のアクセラレーション」である(図 2)。同じマイクロアーキテクチャを持つ「プログラマブルな処理回路群」を並列化している。

一方、スマートフォンや車載 ECU のチップメーカーは、通常、ヘテロジニアスな並列化を進めている。消費電力の削減を重視する(データフォーマットやアルゴリズムごとに最適化した、様々なアクセラレーションコアを搭載して最適な回路で処理した方が、消費電力がより小さくなる)からだが、データセンターやスパコンといえども、消費電力は問題なはずである。

この路線の違いは、今後、どのように進むのか? 端末系はヘテロジニアス、データセンター系はホモジニアスと分かれるのか? それとも、消費電力や性能以外の、他の要因が市場での優劣を左右するのか? その点について、筆者の考えを以下にまとめてみた。

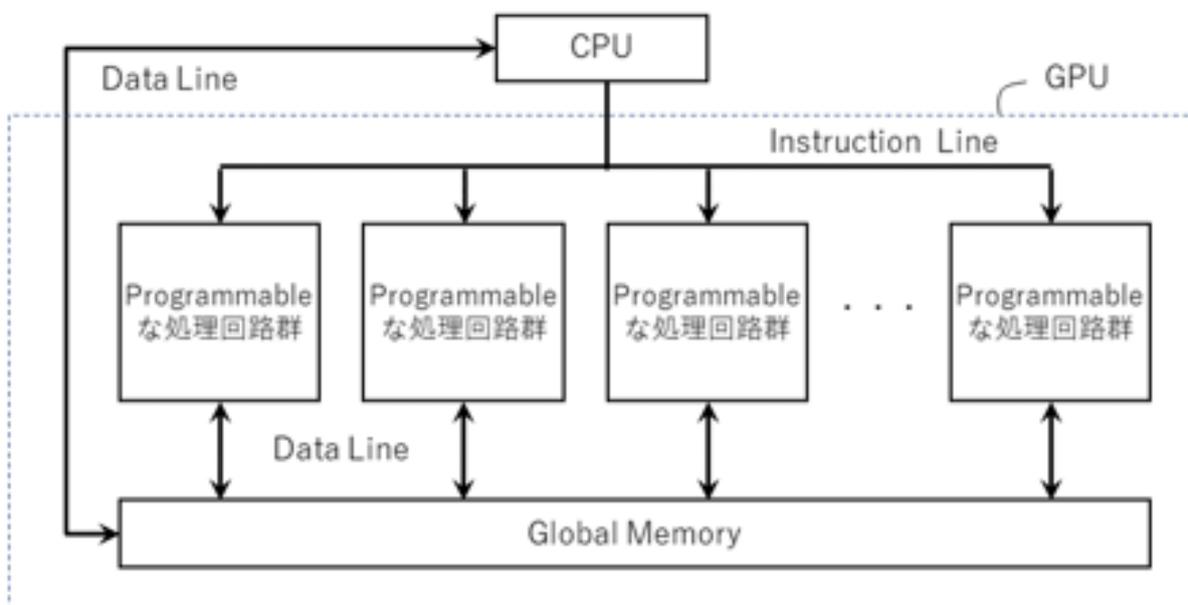


図 2 並列化の一般形 Nvidia 社は「ホモジニアスな並列化によるアクセラレーション」を戦略とするが、スマートフォンや車載 ECU チップは、通常、ヘテロジニアスな並列化によってアクセラレーションする。データフォーマットやアルゴリズムごとに最適化した、様々なアクセラレーションコアを搭載することで、最適な回路で処理すれば消費電力が小さくなるからである

出典: [Nvidia の成功に学ぶ\(その 5\)](#) (2021/03/19)

業績好循環の背景の観点から：ソフトウェア開発者の支持取り付けの効果

Nvidia 社は、自らの並列処理回路アーキテクチャを流布すべく、グラフィック・ボードメーカーやアプリケーション・ソフトウェアの開発者達を強かにサポートし、現在の成功に至っている。成功要因は、「開発者達の支持を取り付けた事」だとの面があっただろう。

同社は、プラットフォーム戦略として、以下に関する技術を重視しているという。

- ・GPU 回路
- ・相互接続技術 (NV-Link、NV-Switch)
- ・ソフトウェア (GPGPU、各種ライブラリ、等)
- ・重要アプリケーション (3D グラフックスや AI) のアルゴリズムへの知見
- ・システム・サポート、および
- ・サービス

それらを総合する存在である「プラットフォーム」が、独自の価値を提供し、その価値が多くの開発者から支持されているのだという。

上記にリストアップされた技術はいずれも大きな技術であるが、同社の「アーキテクチャへのこだわり」は、内部の専門家を継続的に育成するという点で、同社の社内でも、また、顧客側の開発者でとつても「費用対効果が高い教育戦略」であったのではないだろうか？「アーキテクチャ」の変化が著しいと、過去の教育投資が無駄になるからである。その結果、社内／社外の両面で、同社のアーキテクチャへの支持が強まったのではないかと思われる。

近年、ディープラーニング技術をベースとした AI 技術には、GAFA (Google、Amazon、Facebook、Apple) と呼ばれるサービスよりの企業や自動車の電装メーカーからの参入が相次いでいる。それら企業は、図 3 の上層に位置する企業である。

一方、Nvidia 社は、集積回路のコア回路から戦略をスタートさせており、図 3 の下層部分 (テクノロジー開発) から、より上層にサポートを拡大させて来た。

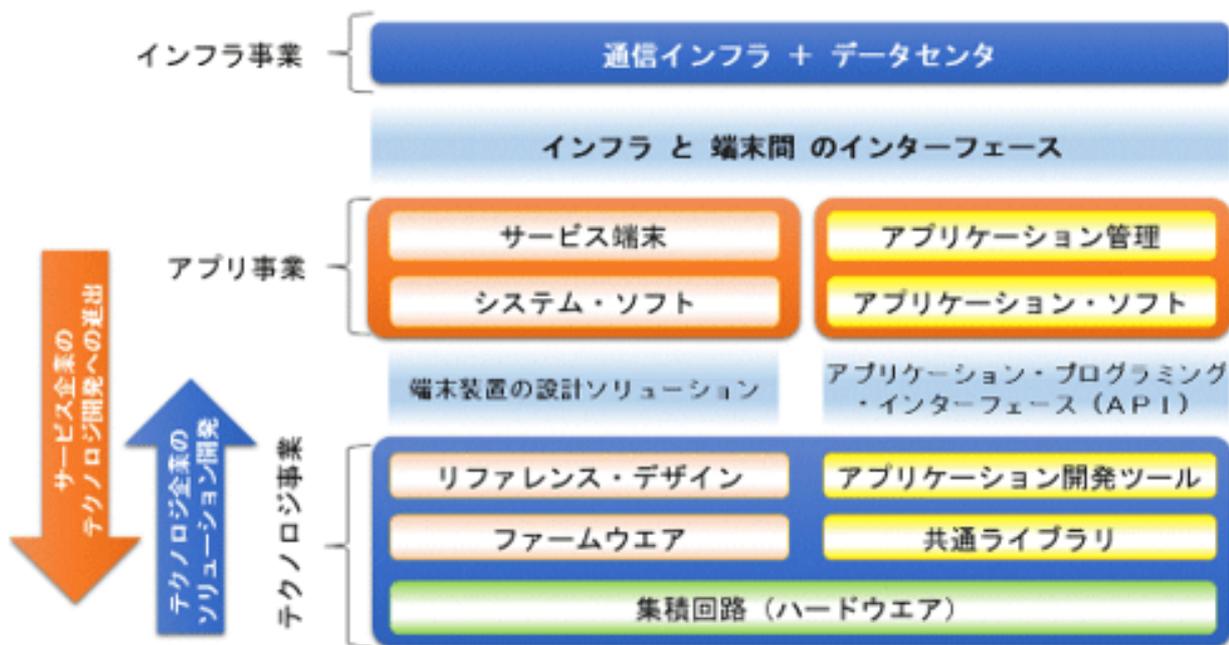


図3 集積回路設計へのR&D予算の投入競争 寡占化したテクノロジー企業も、サービス企業も、次世代半導体開発に対するR&Dを強化している 出典:筆者作成

今後、AI分野では、GAFA、電装メーカー、GPUメーカーを含む様々な企業が入り乱れた大競争が展開すると思われるが、

- ・ スケーラブルな回路であること
- ・ 世界中の開発者達からの支持を得てきた(幅広くツール提供を行い、開発文化として広める)

ことを2大理由に、データセンターには Nvidia 社のチップが浸透しつつある。

AIでは、データセンターで行う学習(トレーニング)のデータ処理と、端末で行う推論のデータ処理を同じニューラルネットワークで行うことが望ましいが、データセンターを制覇する技術は、端末チップに搭載すべきAIアクセラレータ回路を限定する可能性があり、Nvidia社のコア技術は今後端末側にも及ぶ可能性があるのではないと思われる。

では、そのコア技術のマイクロアーキテクチャの次のステップは、どうなるのか？筆者は、Nvidia社の最近の動向を、一つの仮説を元に今後注視したいと思っている。

ネットワーク技術の強化

Nvidia社は、現在、ネットワーク技術の強化にも注力しているかのように見える。約1年前(2020年4月16日)、同社はイスラエルのネットワーク装置企業 Mellanox Technologies社を買収し(参考資料5)、同社の技術を元にした BlueField DPU (Data Processing Unit)という呼び名のネットワークの制御と監視のためのチップを早速市場投入した(参考資料6)。DPUは、ネットワークプロセッサをコアとし、ネットワークパケットの処理と高度な QoS(Quality of Service)を行うインターフェースチップのように筆者には見えている。

GPU と Arm 社の技術、DPU を合わせると、プロセッサベースのデータ処理ハードウェアとしてはフルセットが集まることになる。同社は、CPU と GPU の統合の次に、その合体回路へのネットワークインターフェース回路の付与を考えているのではないだろうか？

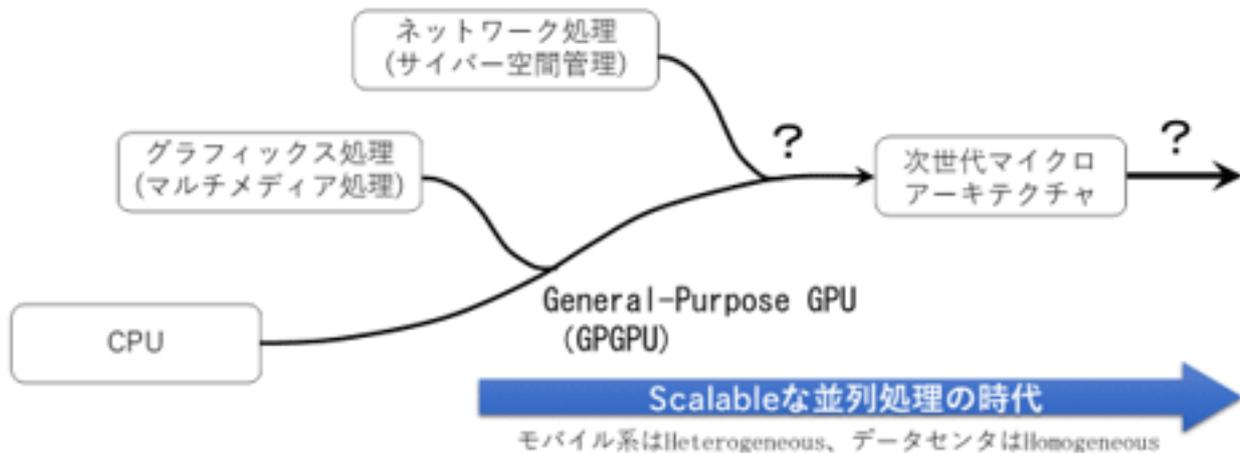


図4 マイクロアーキテクチャの進化(筆者の考え) CPU は、GPU と合体時に強力なスケーラビリティを獲得した。今後、ネットワーク回路とも合体を目指すのではないだろうか？更にその後、何を指すのかを考えるべき段階にある。出典:筆者作成

IoT の時代に、端末(PC、スマホ、各種の業務用装置、自動車等)の集積回路は、データセンターの集積回路群と協調動作する方向にある。アプリケーションを機能させるためには、データセンターとの接続は必須となるからである(参考資料7)。

ネットワーク機能の統合は、どのような統合回路を生むのだろうか？単なる2チップの合体なのか、それとも、アーキテクチャのレベルでの新しいアプローチが現れるのか？その動向は、非常に気になる「次のステップ」である。

そして、更にその先もあるだろう。集積技術の大変革の動きは、現在、大競争の最中にあり、次の次のストーリーも、その大競争の中で進化するのだろうと筆者は思う。

参考資料

1. [Nvidia 社の 2021 年 2 月 24 日のプレスリリース](#)
2. [NVIDIA、売り上げ最高も もろ刃の仮想通貨ブーム](#)、日本経済新聞 (2021/02/25)
3. [Nvidia DGX A100](#), Nvidia 社のホームページ
4. [AI データセンターの性能が 20 倍に、NVIDIA が Ampere 世代の GPU「A100」を発表](#)、MONOist (2020/05/15)
5. [Wikipedia の Mellanox Technologies の項](#)。Intel や Xilinx、Microsoft も Mellanox Technologies 買収に動いていたと言われている。
6. [Nvidia BlueField データプロセッシングユニット](#) Nvidia 社のホームページ
7. [NVIDIA が大賭けするデータ処理装置 \(DPU\) とは](#)、Axion