

AI チップの計算能力競争とチップのサプライチェーンの分散化

TAcc+ スタートアップ分析チーム

ビッグデータと AI を結びつけ、アルゴリズムを効率的に実行し、指数関数的に増加する膨大なデータを処理するには、強力な計算能力が必要となる。このため、AI の演算処理向けに設計されたチップが登場し、急速に発展している。マーケットズアンドマーケットズの報告によれば、AI チップ市場の規模は 2023 年に 512 億ドルに達したが、2028 年には 1,318 億ドル、年平均成長率は 20.8% に達すると予測されている。

AI チップの計算能力競争

ChatGPT の爆発的な人気により、生成 AI の応用が急速に発展している。応用はテキスト、画像、音声、映像にまで及び、カスタマーサポート、アート、音楽、広告等々、さまざまな産業の生産性を向上させている。

AI アシスタントが持つ各方面での発展のポテンシャルに、多くの企業が、「キラーアプリケーション」としての爆発的な商機の可能性を感じている。マイクロソフトの Azure、Amazon の AWS、Alphabet の Google Cloud 主要 3 社をはじめとするクラウドサービス各社が投資を加速、拡大し、それに伴って AI サーバーの需要も急増している。

AI チップの計算能力競争は激化し、ほぼ半年ごとに計算能力を倍増させた新しいチップが登場している。現在、AI チップの専門市場では NVIDIA が圧倒的な優位を占め、市場シェアは 95% 以上に達している。

NVIDIA のジェンソン・フアン創設者兼 CEO は、2024 年 3 月 19 日に開催された NVIDIA GTC 2024 の席上、言語モデルをトレーニングする際に必要な FLOPS（1 秒間に実行できる浮動小数点数演算の回数）について紹介した。2017 年に Google が Transformer モデルを発表して以来、主要企業が発表するモデルにおける計算速度の成長要求はこれまで以上に急速に高まり、2023 年末にリリースされた GPT-MoE モデルでは、必要とされる FLOPS は毎秒 1.8 兆回にもなった。

フアン氏は具体例として、2021 年に NVIDIA の GPU を使ったスーパーコンピュータ Selene を挙げた。Selene は 4,480 基の A100 型 GPU を搭載し、処理速度は毎秒 3 EFLOPS（毎秒 100 京回計算する能力）、接続帯域幅 112TB/s である。これに対して、2023 年のスーパーコンピュータ Eos は 10,752 基の H100 型 GPU を搭載し、処理速度は毎秒 43 EFLOPS、接続帯域幅は 1,100TB/s に達している。わずか 2 年で、全体の処理速度が 10 倍以上に向上したのである。

NVIDIA は新世代の Blackwell アーキテクチャ GPU を発表した。革新的な生成 AI 応用の需要に対応することを目指し、この Blackwell チップが新たな

産業革命を推進するエンジンとなることを期待している。Blackwell は TSMC の 4 ナノメートルプロセスを採用し、2,080 億個のトランジスタを搭載している。NVIDIA は Blackwell アーキテクチャ GPU2 基と ARM アーキテクチャの Grace CPU1 基を組み合わせ、スーパーAI チップ GB200 を開発した。このチップは、AI トレーニングの速度が前世代の H100 GPU チップの 4 倍で、推論能力は 30 倍以上向上し、コストとエネルギー消費は 25 倍削減されている。

古くからのライバル企業である AMD も存在感を示している。NVIDIA が 2022 年 3 月に H100 GPU チップをリリースしたのに対し、AMD は 2023 年 6 月に AI の演算向けの GPU チップ MI300X および GPU と CPU を組み合わせた「アクセラレーテッド プロセッシング ユニット (APU)」の MI300A を発表した。MI300X は 1,530 億個のトランジスタ、192GB のメモリ容量、毎秒 5.3TB のメモリ帯域幅を備えている。AMD によれば、H100 に比べ、トランジスタ数は約 2 倍、メモリ容量は 2.4 倍、メモリ帯域幅は 1.6 倍になったという。

その後、NVIDIA は 2024 年 3 月に Blackwell GPU と GB200 を発表した。その性能は AMD 製品を大きく上回っていた。また、NVIDIA は新しいチップのリリースサイクルを、それまでの 2 年に 1 回から 1 年に 1 回へと短縮することを宣言した。

フアン氏は 2024 年の COMPUTEX TAIPEI の講演で、2025 年に Blackwell Ultra、2026 年には次世代 GPU プラットフォーム Rubin、2027 年にはそのアップグレード版である Rubin Ultra を発表する予定であることを明らかにした。この発表から、新しい GPU プラットフォームを 2 年ごとにリリースし、その翌年にアップグレード版を提供するという進化パターンが示されていることが分かる。一方、AMD のリサ・スーCEO も、2024 年第 2 四半期の財務報告において、AMD が毎年新しい AI チップをリリースする計画であると公表した。2024 年第 4 四半期に MI325X、2025 年に MI350、2026 年に MI400 をリリースする予定だという。

両社は AI チップのハードウェア性能だけでなく、全体の計算性能とコスト効率においても競争を繰り広げている。業界の専門家であるダグ・オラフリン氏は、現時点で NVIDIA が AI チップ市場を支配できている理由は、顧客に最も包括的なソリューションを間断なく提供しているからだと指摘している。そのソリューションには、最高のチップ、最適なネットワーク機器、そして最も優れたソフトウェア CUDA が含まれる。これに対して、AMD も自社のソフトウェア ROCm をオープンソース化し、CUDA を ROCm に変換するためのツールを顧客に提供し始めた。リサ・スー氏はインタビューの中で、生成 AI に必要な高い計算能力は、高性能のチップ、大容量のメモリとソフトウェアによって支えられる必要があると述べた。

GPU チップの二大メーカーが AI チップ市場で争う中、CPU チップ市場のトップ企業である Intel も黙ってはいなかった。Intel は 2023 年 12 月に AI チップ Gaudi3 を発表した。

Intel は 2019 年に Habana Labs を買収した後、直ちに Gaudi チップの開発に

着手した。パット・ゲルシンガーCEO は発表会の席上、AI PC が AI サーバーの需要急増に続く次の主役となると述べ、特に Windows ノートパソコンとパーソナルコンピュータ向けに設計された Core Ultra チップ、および第 5 世代 Xeon サーバーチップを強調して紹介した。

これら 2 つのチップには、グラフィック処理能力を向上させ、AI プログラムの実行を加速させる AI 専用コンポーネント「ニューラルプロセッシングユニット (NPU)」が内蔵されている。Core Ultra は、各コンピュータがオフライン時に、単体で小規模な AI の演算によるタスク処理能力を強化する。Xeon は GPU と組み合わせて使用されることが多く、生成 AI システムのトレーニングやデプロイに利用される。

端末デバイスの AI 計算能力を向上させるために、Intel は 2016 年にスタートアップ企業である Movidius を買収し、同社の製品である「ビジュアルプロセッシングユニット (VPU)」を手に入れた。この VPU は、低消費電力で AI の演算を実行できる。同社は買収後まもなく USB メモリに似た形状のデバイスを発売した。これは、USB 接続で各種端末デバイスに外付けすることで、デバイスの AI の演算能力を向上させることができる。

その後、Intel は Meteor Lake シリーズのプロセッサに VPU 設計を導入し、AI の演算性能を向上させるとともに、CPU の負荷と消費電力を軽減させた。以上から、Intel はデータセンターへの AI クラウドサービスの導入と、オフライン端末デバイスへの AI 機能導入の計算能力の需要に、同時に対応していることがうかがえる。

Alphabet、IBM、Meta、マイクロソフト、Amazon など他のテクノロジー大手も、AI チップ設計の分野に積極的に参入している。その主な目的は、AI チップを手に入れる新たなルートを開拓し、自社のニーズに合った、よりコスト効率の高いソリューションを開発することである。

例えば Alphabet は、2023 年 12 月に最先端の AI モデル「Gemini」を発表した。このモデルは、動画、音声、テキストなど、各形式の情報を処理することができる。この新技術は、同社の AI アシスタント製品「Gemini」の 2024 年版に搭載されている。

Alphabet はまた、自社開発の新世代のカスタム AI チップ「TPU v5p」も発表した。このチップは、前世代の AI チップと比較してトレーニング速度が約 2 倍に向上し、新版 AI モデルの演算ニーズに対応できる。

IBM は、2023 年 10 月に人間の脳の働きを模倣した AI 推論チップである「NorthPole」を発表した。このチップの最大の特徴は、すべてのメモリがチップ上に統合されていることで、追加のメモリ接続が不要であるため、データ転送速度がデータ処理速度に追いつかないことで生じるボトルネック（フォン・ノイマン・ボトルネック）を回避できる点である。このチップは、大量のデータをリアルタイムで処理する必要があるエッジコンピューティングのアプリケーションシナリオに適している。例えば、自動運転車のナビゲーション装置、ロボット操作、交通状況のモニタリング、衛星モニタリング、

情報セキュリティ検出などである。

大手テクノロジー企業だけでなく、多くのスタートアップ企業も AI チップの開発に取り組んでいる。以下にいくつかの例を挙げる。

●SambaNova Systems

SambaNova Systems は、従来の CPU および GPU アーキテクチャの計算能力向上のコスト効率が徐々に低下していることに着目した。ハイエンド級の GPU 価格が上昇している一方、供給が間に合っていないという状況下で、データセンター事業者に計算能力を高める新たなソリューションを提供した。

同社は、データ分析および AI のハードウェアとソフトウェア、コンピュータープロセッサの開発に焦点を当て、さらに AI アルゴリズムを活用してカスタムチップの設計能力向上に力を注ぎ、チップの計算能力、消費電力、サイズ制御などの面で GPU を上回る性能を追求している。また、すでに Alphabet、ソフトバンク、Intel から投資を受け、同社の製品は 2024 年 1 月に大手半導体メーカーのアナログ・デバイセズに採用されている。

●Cerebras Systems

Cerebras Systems は、NVIDIA、AMD、Intel などがウェハを小さなチップに分割する手法を取っているのとは異なり、ウェハ全体を使って巨大なチップを製造している。このアプローチにより、AI の演算に最適化された新しいコンピュータシステムを構築している。現在主流の設計では、チップを再接続するために多大な電力とコストが必要だが、同社の設計では、接続とネットワークのコストおよび消費電力を削減できる。

同社は、OpenAI のサム・アルトマン CEO や、Benchmark などのベンチャーキャピタルから投資を受け、2023 年 7 月 20 日にはアラブ首長国連邦のテクノロジー分野のホールディング企業と提携し、AI スーパーコンピュータを構築することを発表し、また 2024 年 3 月 13 日には新型の WSE-3 ウェハスケール AI チップもリリースした。WSE-3 は H100 チップに匹敵する性能を持ち、その性能は H100 の 62 基分に相当する。同社は 2024 年下半期の上場を目指し、IPO（新規公開株式）の申請を行ったということだ。

●Graphcore

Graphcore は、小規模な AI モデルの演算に適した、AI トレーニングに特化したチップ製品「IPU（Intelligence Processing Unit）」の開発に取り組んでいる。これは、同社の創業者が、GPU は CPU よりも AI の演算に適してはいるが、AI 専用設計されたものではないと考えたからだ。だが、同社はマイクロソフトやセコイア・キャピタルなどから投資を受けたものの、新しい設計が既存のサプライチェーンに参入することは難しい。また、ChatGPT の大規模言語モデル（LLM）では高い演算能力が要求される。

さらに、GPU は IPU よりも LLM に適しているうえ、米中の半導体戦争が同社の中国市場での事業展開を妨げたため、同社の製品は長い間市場で支持を得られない状況が続いていた。2024 年には外部による買収の機会を模索

し、同年7月にソフトバンクグループに買収された。

●Groq

Groq は、AI 推論に特化したチップの設計に注力している。このチップは、カスタムコンパイラ（コンパイラとは、あるプログラム言語で書かれた原始コードを別のプログラム言語に変換できるコンピュータプログラム）を使用し、さまざまな言語モデルをサポートできる。

同社のチップは低消費電力で、AI、機械学習、高性能計算の分野における複雑なワークロードを加速でき、競合製品と比較してコスト効率が高い。同社は近年、AI ソリューション企業の Maxeler Technologies や Definitive Intelligence を相次いで買収し、2024 年には新部門の「Groq Systems」を立ち上げて、顧客や開発者コミュニティを積極的に拡大している。

●Mythic

Mythic は、DAC（デジタルアナログコンバータ）、ADC（アナログデジタルコンバータ）、およびメモリ内計算をサポートする組み込み型フラッシュメモリにより、アルゴリズムの加速を実現した。成熟した製造工程により、AI 推論向けに特化したアクセラレータチップを開発することで、デジタルアーキテクチャを採用している競合製品よりも、優れたコストパフォーマンスと低消費電力のメリットを提供することを目指している。

近年、NVIDIA が AI チップ市場で圧倒的な地位を占めていることにより、ベンチャーキャピタルは AI チップ関連のスタートアップへの投資リスクに対する意識を高めている。Mythic のデイブ・リック CEO によれば、資金調達は一時困難に直面したものの、幸運にも 2024 年 8 月にシスコ、サムスン、ブラックロックが主導するシリーズ D ラウンドで、6 億 4 千万ドルの投資獲得に成功したという。

地政学が引き起こすチップのサプライチェーンの分散化

台湾は、1960 年代に軽工業の輸出を推進し、1970 年代には基幹産業と重工業の発展、1980 年代には経済自由化と技術集約型産業の成長、1990 年代にはアジア太平洋オペレーションセンターの発展と情報産業の成長を促進してきた。1991 年にソビエト連邦が解体し冷戦が終結すると、世界の政治経済情勢は、政治体制によるグループ間の対立から開放と和解へと転じた。これにより国際貿易量が急増し、世界各地の産業が連携・統合を進め、徐々にグローバルな産業チェーンと分業体制が形成された。各国の企業はそれぞれの強みを活かして、グローバルロジスティクスを通じて資源を統合し、製品の製造・販売を行ってきた。

台湾はこのような主観的・客観的条件が整った結果、世界のハイテク産業のサプライチェーンにおける重要なクラスターとして発展してきた。

調査会社 TrendForce の調査結果によれば、2024 年第 1 四半期時点で、世界のウェア OEM 企業トップ 10 の収益状況は下表のとおりである。台湾企業は

トップ 10 に 4 社がランクインし、市場シェアは合計で 69.6%に達している。
なかでも TSMC はリーディングカンパニーとして市場シェア 62.3%を占め、
台湾が世界の半導体産業の専門分業体制におけるチップ製造部門で、非常に
重要な地位を占めていることを示している。

表 1 世界のウェハ OEM 企業トップ 10 の売上高と市場シェア

ラン キン グ	会社名	売上高（単位：百万ドル）			市場シェア	
		2024Q2	2024Q1	四半期 成長率	2024Q2	2024Q1
1	TSMC	20,819	18,847	10.5%	62.3%	61.7%
2	サムスン	3,833	3,357	14.2%	11.5%	11.0%
3	SMIC	1,901	1,750	8.6%	5.7%	5.7%
4	UMC	1,756	1,737	1.1%	5.3%	5.7%
5	グローバ ルファウ ンドリー ズ	1,632	1,549	5.4%	4.9%	5.1%
6	華虹グル ープ	708	673	5.1%	2.1%	2.2%
7	タワーセ ミコンダ クター	351	327	7.3%	1.1%	1.1%
8	VIS	342	306	11.6%	1.0%	1.0%
9	PSMC	320	316	1.2%	1.0%	1.0%
10	Nexchip	300	310	-3.2%	0.9%	1.0%
トップ 10 合計		31,962	29,172	9.60%	96%	96%

出典：TrendForce

2018 年以降、米中貿易摩擦が米中貿易戦争へと発展し、世界の政治経済情勢は、開放と和解から政治体制によるグループ間の対立と衝突へと転じた。2020 年初頭には新型コロナウイルス感染症が世界中で猛威を振るい、感染拡大と各国の入国制限により、世界中のサプライチェーンは大混乱に陥った。

さらに、2022 年 2 月 21 日にはロシアとウクライナの間で長年繰り返されてきた小規模な衝突が全面戦争に発展し、世界のエネルギーと食糧供給に深刻な影響を及ぼした。2023 年にはパナマ運河が 70 年ぶりに深刻な干ばつに見舞われ、世界の輸送能力が大打撃を受けた。同年 10 月 7 日にはイスラエルとハマスの戦争が勃発し、紅海の航路が遮断されたため、貨物船は喜望峰またはマゼラン海峡を通るルートに変更することを余儀なくされ、輸送コストが大幅に増加した。

東アジア地域においても、南シナ海、台湾海峡、東シナ海、日本海周辺の政治情勢が緊張を高めている。貿易戦争、感染症、戦争、自然災害が原因でサプライチェーンが中断するリスクが増大するなか、多くの国はそれら非常事態に対処するため、自国のサプライチェーンを強化する政策を立てている。中でも米中の競争によって焦点となっているハイテク産業は最も重要な分野である。

2023 年 10 月 14 日、TSMC の創業者であるモリス・チャン氏は、同社の運動会に出席した際、2019 年の運動会で自身が述べた予測を振り返った。2019 年、チャン氏は米中のテクノロジー戦争が拡大するにつれ、TSMC は地政学的な要因により、米中両国が争って提携を求めてくる企業となるだろうと述べていた。そして 4 年後、彼はその予測が現実となっただけでなく、これまでの半導体産業のグローバル化は徐々に消失し、自由貿易が後退するだろう、そして各先進国は自国の安全保障を優先する方向へと転じ、TSMC は各国が争って「自国の土地」に工場建設を誘致する企業となるだろうと断言した。

またチャン氏は、競争という観点から見ると、競合他社も自身の地政学的な優位性を活かし、TSMC に追いつき追い越そうと積極的に動くことが予測され、今後数年間、TSMC はさらに厳しい競争に直面するだろうと指摘し、現在蓄積された優位性が、今後 20～30 年間維持される保証はないため、TSMC の経営チームは万全の態勢を整え、慎重に対処し、これからもトップを維持してほしいと語った。

表 2 海外投資による TSMC の OEM ウェハ工場計画

地点	稼働開始 予定	工場建設計画	投資状況
日本	2024 年	28/22/16/12 ナノメートルプロセス	総投資額は約 3 兆円。TSMC、Sony、トヨタなどの企業が参画して設立した JASM が建設。日本政府は 1.2 兆円を助成。
日本	2027 年	40/7/6 ナノメートルプロセス	
米国	2025 年	4 ナノメートルプロセス	総投資額は 650 億ドル超。米国政府

米国	2028 年	2 ナノメートルプロセス	は 66 億ドルを助成し、さらに 50 億ドルの融資および 25% の投資税額控除を提供する可能性がある。
米国	2029 年	2 ナノメートルまたはそれ以上の先進プロセス	
ドイツ	2027 年	28/22/16/12 ナノメートルプロセス	総投資額は 100 億ユーロ超。TSMC とボッシュ、インフィニオン、NXP などの企業が共同出資により設立した ESMC が建設。ドイツ政府は 50 億ユーロを助成し、地域のインフラ整備の費用も負担する。

出典：中央通訊社

世界の主要先進国および大手テクノロジー関連企業は、現在チップのサプライチェーンの現地化に力を入れ、第 2、第 3 の供給ルートを構築して、前述したリスクによるサプライチェーン寸断の危機を回避しようとしている。

チップのサプライチェーンは、従来のグローバル化から地域化へと分散しつつある。表 2 で示した TSMC の海外工場建設地点からも、東アジア、北米、西ヨーロッパが、分散化の第一波における 3 大チップ生産エリアであることがわかる。さらに、中国やインドなども積極的にサプライチェーンの構築を進めており、将来的には最低でも 5～10 の地域にサプライチェーンが分散されると予測される。

チップについては、サプライチェーンの上流・下流の企業も、地域や国境を越えた工場建設に協力している。そのような企業には、例えば 2023 年以降に AI サーバーの需要急増によって株価が急騰したサーバー ODM 大手の QCI、Wistron、Supermicro がある。

QCI は、HP などのサーバーブランドから、中国製造におけるリスクの分散を求められたため、「チャイナ・プラスワン戦略」に基づく分散計画に対応し、タイに工場を建設した。その結果、2024 年に HP から大規模な注文を受け、今後も生産能力を拡大する予定である。タイ工場は、QCI グループ内で台湾に次いで 2 番目に大規模なサーバーの一体化生産拠点になる見込みである。

Wistron の会長は、2024 年初頭に行われた旧暦の忘年会の席上、同社が現在メキシコとベトナムに工場を建設中であり、さらにインドと米国カリフォルニア州にも工場建設計画があることを表明した。

Supermicro は、NVIDIA との密接なパートナーシップを通じて業績が大幅に向上し、株価は 2023 年初頭から 2024 年初頭にかけて 6 倍以上に急騰した。同社のチャールズ・リャン総裁兼 CEO はインタビューで、アジアには大規模な通信およびデータ関連の需要があることを指摘し、同社は引き続き台湾の生産能力を拡大し、発展の重心を米国から台湾へ移す計画であると述べた。

出典：

- AI 晶片：前沿技術與創新未來。張臣雄。2021。
- AI 晶片為何如此重要 (上)。黃則普。2021。
- AI 黎明時刻 蘇姿丰領 AMD 正面對決。工商時報。2023。
- CPU vs. GPU：發揮兩者的最大效益。Intel。
- IBM 發表模擬人類大腦運作，號稱效能超越 4 奈米 GPU 的 AI 推論晶片 NorthPole。iThome。2023。
- Nvidia (NVDA) -全球最大圖像處理器公司。太股達人。2022。
- TrendForce：第二季供應鏈急單挹注晶圓代工利用率，全球前十大晶圓代工產值季增 9.6%、VIS 排名上升兩位。TrendForce。2024。
- 台積電評估德國設廠，外界預料首選薩克森矽谷。財經新報。2021。
- 直擊 | AMD 推 AI 新晶片，劍指輝達一哥地位！誰說沒客戶？微軟、Meta 都來了。數位時代。2023。
- 英特爾也強調 AI！Meteor Lake 系列處理器導入 VPU 設計。科技新報。2023。
- 英特爾發表新款 AI 晶片與 Nvidia 和 AMD 競爭。經濟日報。2023。
- 黃仁勳 2024 Nvidia GTC 完整演講，揭露在 AI 宇宙的最大野心。商周。2024。
- 黃仁勳旋風席捲全台「Jensanity」！獨家解析！黃仁勳 COMPUTEX 演講在說什麼？想搭上 AI 趨勢就該看這部影片！曲建仲。2024。
- 業界不樂見一家獨大，兩虎相爭煙硝味濃；AMD 挑戰輝達，供應鏈下半年開新局。科技新報。2024。
- 輝達市值狂飆！AI 晶片戰，AMD 能打破 Nvidia 壟斷嗎？天下雜誌。2024。