

半導体産業の 現状

半導体産業の情勢を形成する
動向と要因

目次

投資拡大により
1兆米ドル規模の産業へ

04

本格化する車載用半導体

14

専用シリコンのルネサンス

22

メモリは新たな境地へ

07

デカップリングへの適応：
レジリエンスを高めるための戦略

20

人工知能—規模から多様性へ

28



半導体が切り開く日本の産業の未来

本レポートが明らかにしているのは、半導体産業が既存の延長線上では捉えきれない構造変化の中にあるということです。長年にわたり「量産効率」と「スケール」が最大の競争軸であったこの産業は、今や急速に「用途特化型」「高難度少量対応型」「スピード重視の開発協業型」へと重心を移しています。これは、AIアクセラレーターやSDV、電動化によるパワー半導体の革新、データセンターの高密度化など、さまざまな最終市場の構造変化が同時多発的に進んでいることに起因しています。

これまで「例外」や「特殊対応」とされてきた専用多品種生産は、今や、先端ロジックや車載用途、通信、電力制御などのコア領域において標準的な要求となりつつあります。多く企業が、単一製品・大ロットの効率性ではなく、仕様変更を前提とした短サイクルの開発対応力、複雑性の高い設計への柔軟な製造プロセス、変種を吸収できる検証環境、そして何より、こうした構造変化を経済的に成立させる事業構造のしなやかさを求めています。そこではもはや、自社だけで完結できる領域は限られています。今後、企業が競争力を維持・強化するためには、組織の枠を超えて外部と連携し、構造的に変化に対応できるシステムを築くことが欠かせません。

このことは、実際の半導体業界の成功事例からも読み取れます。たとえば、AI処理向けに需要が急拡大している HBMは、その技術的複雑性故に、メモリメーカー単独では製品化が困難でした。HBMの市場は2028年に380億米ドルに達し、サーバーDRAM市場の半分以上、DRAM全体の約28%を占めるまでに成長すると見込まれています。この背景には、TSMCが複数のメモリメーカーと2年以上にわたりHBMの共同開発を進めている ように、ファウンドリー・パッケージ・設計の連携なしでは成立しない構造があります。こうした外部連携こそが、最新世代のHBM3E、HBM4などの市場投入を可能にしています。

また、データセンター向けのカスタムIC市場は2028年に240億米ドル規模に達すると予測されています。このカスタムICの多くは、GoogleのTPUやMetaのMTIAのように、BroadcomやMarvellなど外部のASIC設計パートナーとの連携によって開発されています。それらは、演算効率とコストを両立する“特定用途向け高性能チップ”として、標準GPUとの差別化を図り、少量生産を前提とした開発・設計・製造の共創モデルが標準となりつつあります。

さらに車載半導体領域でも、1台あたりの半導体搭載額は2019年の420米ドルから、2023年には800米ドル、2030年には1,350米ドルに達すると予測されています。この数値は、電動化や自動運転、ソフトウェアアップデートへの対応といった複雑化要因の累積を示しており、セグメントや車種によって異なるアーキテクチャへの柔軟な設計対応が前提になることを意味しています。SoCやMCU、メモリといった構成要素すべてにおいて、用途ごとの微細な差異に応じたバリエーション対応力が、車載半導体メーカーと半導体の装置・材料サプライヤーにも等しく求められています。

これらの変化に対して、個社単独のリソースで対応し続けることは、技術・事業・意思決定のいずれの面においても限界があります。特に、SDVやカスタムICのように設計と製造の境界が曖昧化するプロジェクトでは、装置や材料の仕様を含めた仕様共創体制が必要不可欠となっています。このような文脈においては、企業の間では“パートナー”としての役割が重要になっています。

PwCは、こうした外部連携と構造的再設計の両方において、日本企業と並走する立場をとっています。業界・国境・企業機能をまたいだ構造理解と、意思決定支援・実行設計・伴走型マネジメントの知見を融合することで、変革の構造的な立ち上げを支援することができる。変化が激しく、技術が高度化し、市場が不確実性を増すこの時代において、必要なのは正確な情報や優れた技術よりも、それらをつなぎ、活用し、柔軟に進化できる仕組みと考えます。

半導体産業は、かつてない広がりと深さで社会の中に浸透し始めています。そしてそれに応えるための企業構造は、もはや1社完結では成立しません。日本の製造現場には、世界に誇れる技術の蓄積と品質への徹底した姿勢があります。しかし、半導体産業の競争環境が激しく、複雑に変化する今、そうした個々の優れた取り組みを、拠点や部門、さらには企業の垣根を越えて統合し、スピード感をもって活用できる体制がなければ、グローバルの競争で生き残ることは難しいでしょう。分散していた強みを連携させ、再構築することが、これから競争力の源になります。そのためには、社内の見える課題に向き合うだけでなく、見えにくい摩擦や連携の不整合にも目を向け、構造的に解決を試みる姿勢が必要です。

変革は、構想するだけでは進みません。仕組みと、それを動かす関係性をどう構築するかが肝要です。私たちはその伴走者として、日本の半導体産業の新しい挑戦に寄り添いたいと考えています。

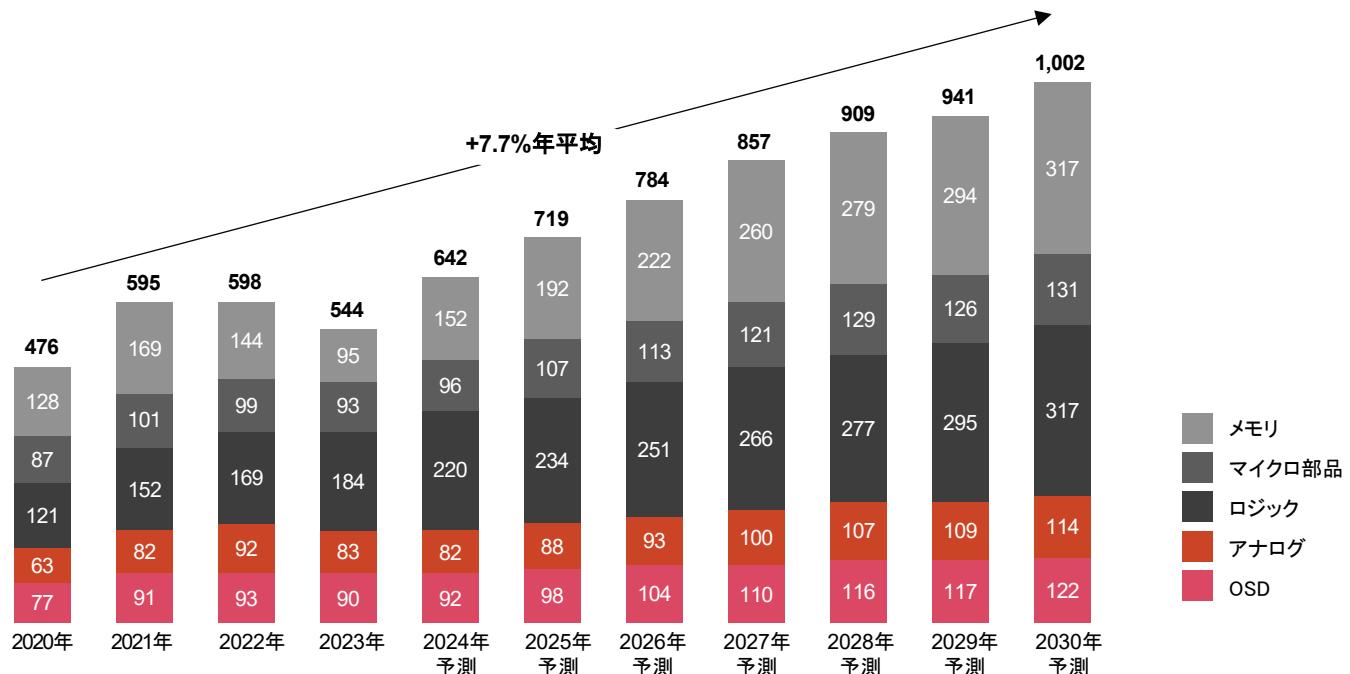
セクション1

投資拡大により1兆米ドル規模の産業へ

半導体は70年以上にわたり技術革新の原動力となり、世界中のあらゆる産業に抜本的な変革をもたらしてきました。パーソナルコンピューティングやスマートフォンに始まり、データセンターやクラウドコンピューティングに至るまで、経済全体にわたつて極めて重要な用途が半導体のイノベーションによって発展してきたのです。今後もこの半導体産業の軌道は、電動化、デジタル化、人工知能(AI)やモノのインターネット(IoT)技術の加速といったメガトレンドによって支えられていくことでしょう。

こうした動向によって半導体産業は長期にわたり持続的な成長を達成できるとされ、世界の年間収益は2024年で6,420億米ドル、10年後には1兆米ドルに達すると予測されています(図表1参照)。

図表1
部品別世界半導体市場、2020～2030年(単位:10億米ドル)



出所:Omdia 2024年第3四半期、OSD(オプトエレクトロニクス、センサー、ディスクリート)

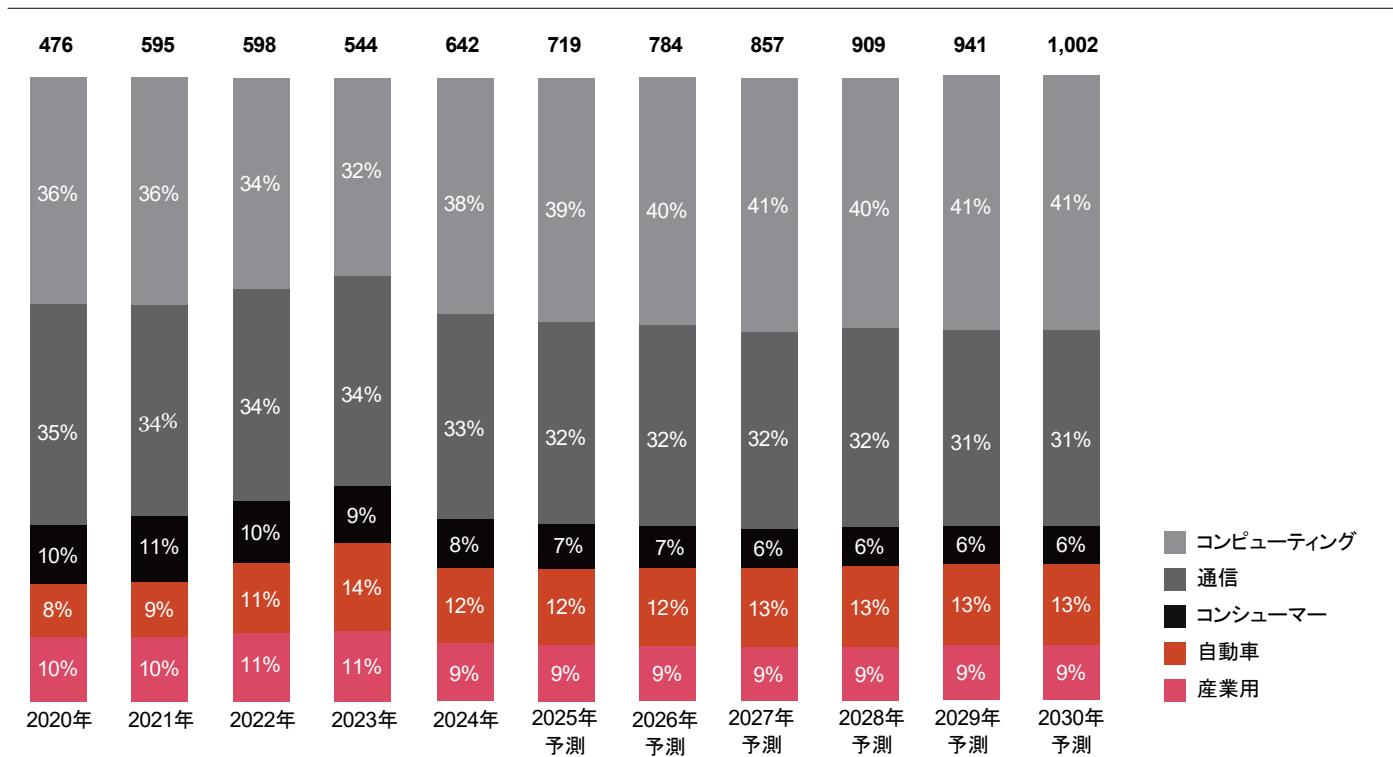
半導体産業は、規模に応じた効率的な製造、技術、世界的な流通による恩恵を受けて、長い間グローバルなサプライチェーンとして機能してきました。しかしこの5年間のコロナ禍と貿易摩擦の激化によって、自国生産への投資を拡大し、サプライチェーンの主権を確保する必要性が浮き彫りとなっています。

コロナ禍においては、企業による従業員のリモートワーク環境の整備、個人による家電製品への投資によって、半導体需要は過去最高水準まで急増しました。同時に、IoT技術の採用が拡大したこと、業界は業務を最新化しさらなる需要が生まれました。一部のセクターでは需要規模を過小評価していたため、サプライヤーは生産能力の確保に苦戦しました。そのため、2020年後半から2022年末にかけては世界的な半導体不足に見舞われました。一方、2023年は過剰供給と需要減少によって過剰在庫となり、半導体産業は不況に陥りました。現在では市場は安定しつつあり、半導体産業の収益は回復し2022年のピークを上回ると見込まれています。

半導体産業の主要な成長要因であるAI、IoT、自動車

半導体市場を構成する7種類の部品（メモリ、ロジック、マイクロ部品、アナログ、オプトエレクトロニクス、センサー、ディスクリート）のうち、収益の最大シェアを占めるのは引き続きメモリおよびロジック製品になるでしょう。その優位性は、コンピューティングやモバイル機器から産業用製品や自動車製品まで、幅広い用途に不可欠な貢献をしていることで説明できます。データの主導的役割が一層強まっていているこの世界では、より高速で効率的なメモリソリューションへの需要が高まり、業界全体でAI、IoT、クラウドコンピューティングの採用が拡大していることからその勢いはさらに増すことでしょう。さらに世界中でデータセンターが急増し続けていることから、大量のリアルタイムデータを処理するためのメモリおよびロジックの重要性が高まっています（図表2参照）。

図表2
用途シェア別世界半導体市場、2020～2030年（単位：10億米ドル）



出所：Omdia 2024年第3四半期

用途では、2024年以降、コンピューティング市場が通信を抜いて最大シェアとなり(前頁の図表2参照)、2030年までの年平均成長率(CAGR)は9%と予測されています。特にこの成長の大きな要因となっているのがAIです。その背景には、集中型計算タスクを管理するために業界から先端半導体によるソリューションを求める声が高まっていることがあります。機械学習、ニューラルネットワーク、データアナリティクスなどの技術は、半導体業界の未来を形作る上で革新的役割を果たすでしょう。ただし、こうした動向による影響の全貌はまだ明らかになっていません。

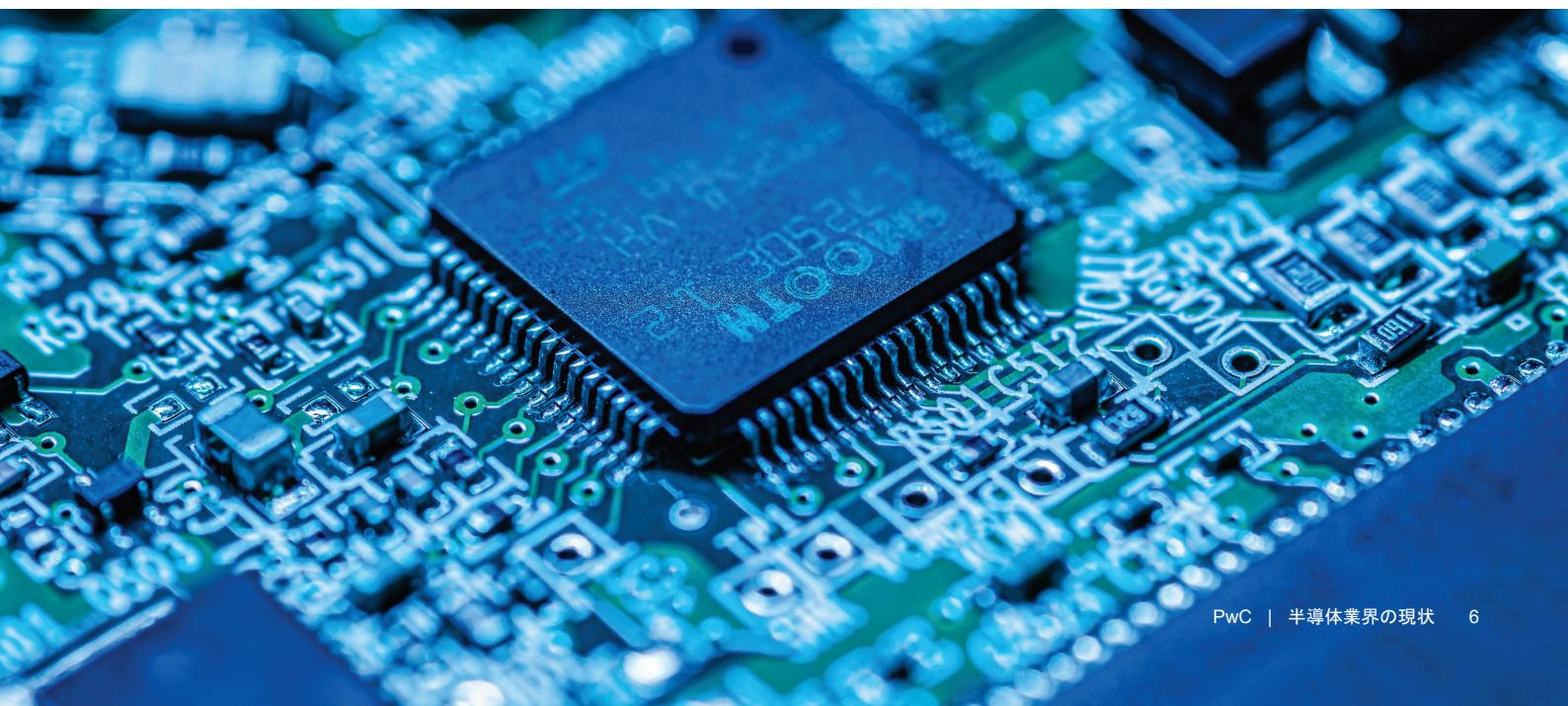
さらに、標準的な既製部品よりもカスタマイズされたソリューションを求める企業が増えてきていることから、カスタム集積回路(IC)への移行も大きな動向として見られます。この移行の理由は、データセンターからコンシューマーエレクトロニクスまでの用途において、それぞれに特化した性能、エネルギー効率、セキュリティ強化が必要とされているためです。その代表例としては、汎用プロセッサに比べ優れた統合性と性能を実現している、Appleが独自開発したMチップが挙げられます。

自動車セクターは2024年から2030年までのCAGRが10%と、最も急成長する半導体市場であり続けると予測されています。この成長予測を支える重要な要素は、米国および欧州市場の成長予測に一部短期的な調整があるとはいえ、自動車の電動化が継続していることです。PwC Autofactsによると、世界の小型乗用車カテゴリーにおけるバッテリー式電気自動車(BEV)の普及率は、2024年第2四半期に13.3%と推定され、2030年までに42.5%に上昇すると予想されています¹。BEVにはモーター駆動とバッテリーシステム管理にパワーエレクトロニクスICとパワーマネジメントICが必要です。BEVは高電圧システムのため、内燃エンジン車(ICE)に比べ搭載される半導体は2倍以上になります²。

ソフトウェア定義型自動車(SDV)への移行もまた、自動車セクターの形を変えつつあります。SDVは継続的に更新できる先端ソフトウェア機能に依存していることから、カスタマイズと柔軟性が向上する他、イノベーションサイクルも短縮されます。このようにハードウェアとソフトウェアを分離することが、より迅速な進歩へつながります。また、自動運転や快適性の向上といった動向により、高性能半導体への需要が高まっています。このように要件は変化し続けるため、半導体メーカーおよび部品メーカーは将来の車両構造を予測し、常に製品需要の先を行く必要があります。自動車1台あたりの車載用半導体は、2019年の420米ドルから2023年には800米ドルとほぼ倍増しています。2030年の予測は1,350ドルで、10年間で3倍になると見込まれています²。

1,350
米ドル

1台あたりの車載用半導体



セクション2

メモリは新たな境地へ

半導体市場においてメモリは極めて重要な役割を果たしており、さまざまな用途でデータストレージおよびデータ処理に欠かせない部品となっています。高密度、大容量、高速、および帯域幅拡張への需要が高まっており、メモリは長期にわたるイノベーションの軌道をたどっています。このイノベーションは、AI、IoT、クラウドコンピューティング、先端データアナリティクスなど、効率よく動作するために大量のメモリリソースを必要とするデータ集約型の用途が急増していることによって形成されます。過去20年間で主な半導体デバイスの中で最も急成長している分野として浮上してきたのが、メモリICであり、CAGRは8.6%です。デジタル経済の拡大に伴い、半導体の総収益に占めるメモリの割合は、2008年の18%から2024年までに推定25%に増加するとみられています²。このデータストレージおよびデータ処理の効率化への需要の高まりは、今後の半導体産業の成長において、メモリが重要な要素としての役割を果たすことを明確に示しています。



このデータストレージおよびデータ処理の効率化への需要の高まりは、今後の半導体産業の成長において、メモリが重要な要素としての役割を果たすことを明確に示しています

Glenn Burm
PwC、グローバル・セミコンダクター・セクター・リーダー

DRAMの進化: AI主導とデータ主導の成長を促進

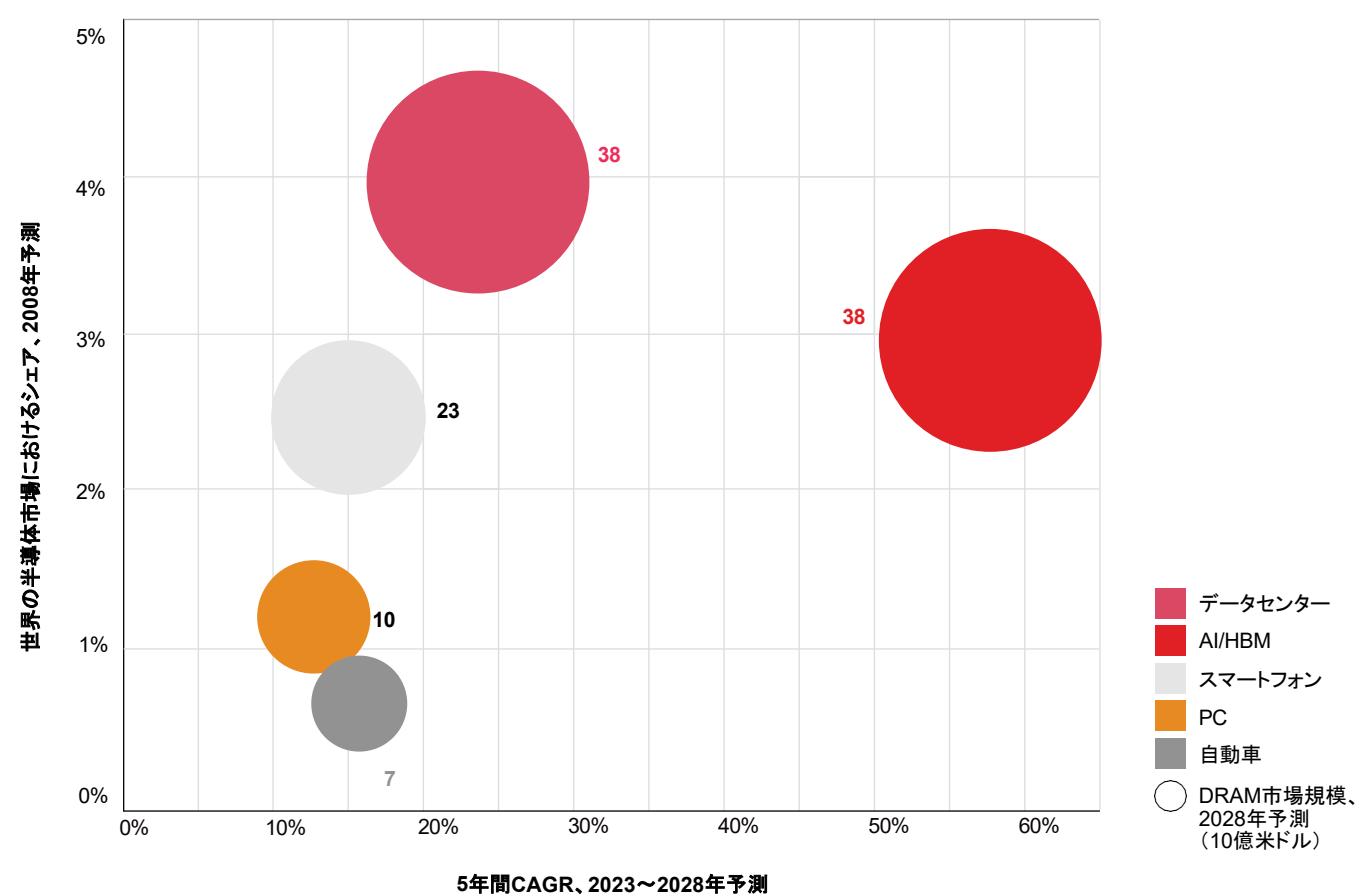
ダイナミック・ランダム・アクセス・メモリ(DRAM)はコンピューターなどの電子機器に使用されるメモリの一種であり、実際に使用中または処理中のデータが格納されます。長い間半導体業界の要であり、2024年には総収益の14%を占めたと推定されています。技術の進歩とデータ集約型の用途への需要の高まりを追い風に、DRAMは長年にわたり大きく進化してきました。当初はパーソナルコンピューティングやデータセンターで使用されていたDRAMの役割は、広帯域幅、高速、大容量のニーズが高まり続けているAI、機械学習、クラウドコンピューティングなどの新興分野にも対応すべく拡大しています。

広帯域幅メモリ(HBM)の市場価格は2028年までにデータセンター向けDRAMに匹敵すると予測されています。2028年までに世界の半導体市場の4.1%を占め、2023年から2028年の間のCAGRは57.5%と予測されています。これに対し、データセンター向けDRAMの市場シェアは4.2%とやや高いものの、CAGRは緩やかで22.3%にとどまるとの予測です。

一方、より成熟したセグメントであるスマートフォン向けDRAMの市場シェアは2.6%で、同期間のCAGRは15.3%とされています。これはAI主導のメモリ用途が爆発的に成長していることを明確に示しています(図表3参照)。

図表3

世界の半導体市場における用途別DRAM市場シェア(%)および5年間のCAGR(%)、2023～2028年



出所:Omdia 2023年第3四半期

広帯域幅メモリ(HBM): 次世代のメモリのイノベーションを加速

AIと高性能コンピューティングが発展し続ける中、従来のDRAMソリューションは限界に達しつつあります。こうした要求を満たせる重要なイノベーションとして登場したのがHBMです。並列コンピューティングやAIのワークロードに最適化されたHBMは、1,000以上の入出力チャネルを備えた非常に幅の広いインターフェイスを特長とし、従来のDRAMに比べ非常に高いデータ転送レートを実現しています。AIのトレーニングや推論タスクに不可欠なNVIDIAやAMDの画像処理装置(GPU)は、大規模なデータセットや複雑な計算をより効率的に管理するため、HBMに大きく依存しています³。

新プラットフォームの登場と性能の向上により、HBMはDRAM搭載量において年間50～100%増加しています。2022年に発表されたHBM3世代は、80GBのメモリを特長とするNVIDIAのH100に採用されました。2024年にはHBM3Eを搭載したB100が発売され、メモリ容量は最大192GBへ増加しました³。NVIDIAが最近発表したHBM4をベースとしたRubinプラットフォームはこの限界をさらに押し上げ、最大メモリ容量は764GBとなります²。

つまり4～5年間でメモリ容量は6～10倍と大幅に増加したことになります。HBM3Eが登場する前は、メモリ企業がベースダイを含むHBMの全ての部品を製造していました。しかし、HBM4の登場によってロジックチップとメモリチップの融合が始まれば、ファウンドリがベースダイの生産を開始するでしょう。各顧客の要求に応じてカスタム機能を組み込む必要があることから、ファウンドリとメモリ企業との連携は不可欠となるでしょう。例えば、TSMCはAI用途に使用されるHBMに関して、すでに2年以上複数のメモリチップメーカーと協業を行っています⁴。

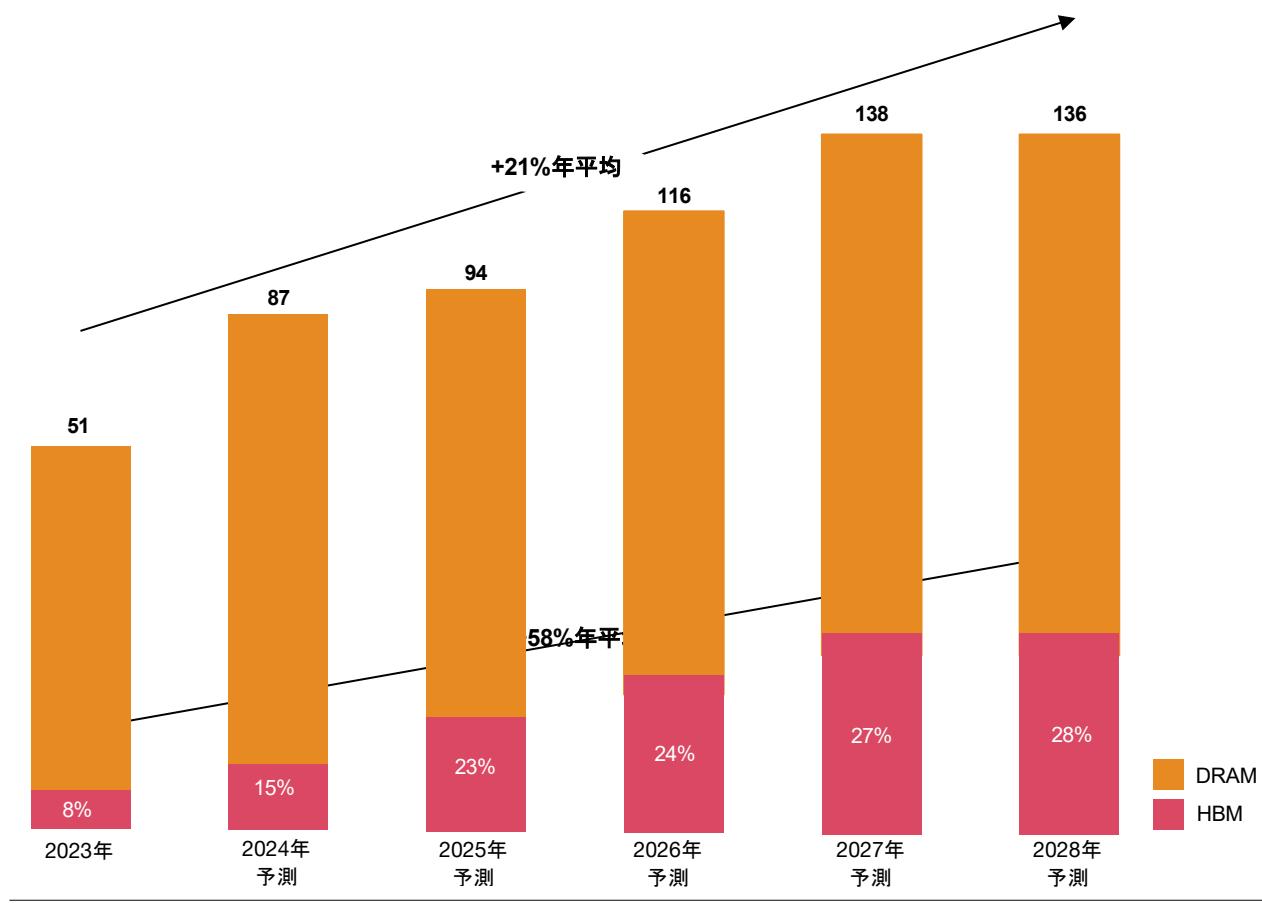
より広範な市場であるDRAMは標準化インターフェイスに対応した製品であり、市場はコストと規模によってけん引されています。一方HBMは、技術的要件や性能要求の進化が非常に早いため、クローズドループのエコシステムの中で運用されており、新規参入は困難です。HBMの仕様の進化はあまりに早く、設計は一層閉鎖的なまととなることから、ベンダーおよび顧客は新しい開発に迅速に適応することができます。その結果、製品の品質とシリコン貫通電極(TSV)歩留まりが性能と信頼性を確保する上で重要な要素となります。

HBM市場は2028年までに急速に拡大し、CAGRはビット成長で64%、収益で58%となる見込みです。2028年までにHBMセグメントは380億米ドル規模となるでしょう。これはサーバー向けDRAM市場の約半分に相当し、1,360億米ドル規模のDRAM市場全体では27.6%を占めることとなります(次頁の図表4参照)。



图表4

HBMのビット需要(1ギガビット換算、単位:10億)、HBM市場およびDRAM市場、2023~2028年(単位:10億米ドル)



出所:Omdia 2024年第3四半期

収益性の向上がDRAM投資に新たな波を引き起す

DRAM供給のひっ迫によって価格と利益率が上昇していることから、メモリICベンダーは設備投資(CAPEX)を強化しています。DRAM産業の営業利益率(OPM)は2024年初頭に20%を超えており、同年末には30~40%に達すると予測されています。この収益性の急増により、2025年のCAPEXは2022年の330億米ドルをはるかに上回り、過去最高水準にまで達する可能性があります。とはいっても、その投資の大部分はHBMのバックエンド製造へと集中するため、2025年のDRAM生産能力は2022年と比べほぼ変わらないでしょう。

また、政府による支援策もこの状況を形成する要因となっています。2022年以降、米国、日本、韓国で補助金と税制上の優遇措置が導入され、その効果は2025年から明らかになるとされています。2025~2026年にはSamsungが平沢(ピョンテク)P4工場にて量産開始を予定している他⁵、2027年にはSK Hynixの龍仁(ヨンイン)工場、2026~2029年にはMicronのボイシ、広島、ニューヨークの新工場にてそれぞれ予定されています。

3D DRAMへの移行:2D技術の限界を打破

2017年に10nmノードが発表されて以降、DRAMのコスト改善に向けたペースは大きく鈍化しています。それ以前はコストの年間削減率は20~30%でした。ところが2017年から2023年にかけては年にわずか6%へと低下し、メーカーがさらにプロセスノードの小サイズ化を押し進めれば、このペースはさらに鈍化すると考えられています。こうした微細幅による2D DRAM技術のメリットが薄れつつあることは、10nm以下ではさらなるコスト削減はほんのわずかであろうことを示しています。

3D DRAMは2D DRAMには果たせなかつことを受け継いで、業界を長期的なコスト削減の軌道に乗せてくれるものと期待されています。第2世代の10nm以下の技術から始まる3D DRAMは、著しくコストを効率化できるでしょう。現在8層および16層の製品が開発中であり、2030年頃にはハイスタックの3D DRAMの量産が開始になると予測されています⁶。

NAND型フラッシュメモリの回復:市場の底からAI主導のスーパーサイクルへ

不揮発性記憶装置技術であるNAND型フラッシュメモリは、無電源状態でもデータを保持できるため、ソリッド・ステート・ドライブ(SSD)などの機器において重要な役割を果たしています。高密度、拡張性、およびビットあたりの低コストゆえ、コンシューマーエレクトロニクスやデータセンターなど、セクターやセグメントを問わず最新のストレージソリューションに欠かせないものとなっています。

過去10年間でNAND型フラッシュメモリのビット成長は大幅に伸び、2013年には371億個(1GB換算)であったのが2023年には7,449億個へと増加しました。この急増は主にスマートフォンやPC、サーバーの需要によるものであり、これらは高速かつ効率的なデータ処理を行うためにNANDベースのストレージに依存しています。3Gから5G/LTEへの移行や、より高速なPCプロセッサなどの重要な進歩が、さらなる成長に拍車をかけました。AIの採用が拡大するにつれ、特にトレーニングや推論など大規模なワークロード管理には膨大なストレージ容量が必要となるため、大容量SSDの需要が加速しています。2028年までに、NAND市場の収益はビット成長とスケーラブルで高性能なストレージへのニーズの高まりという後押しを受けて、1,150億米ドルに達すると予測されています²。



AIの採用が拡大するにつれ、特にトレーニングや推論など大規模なワークロード管理という観点から、大容量HBMの需要が加速しています“

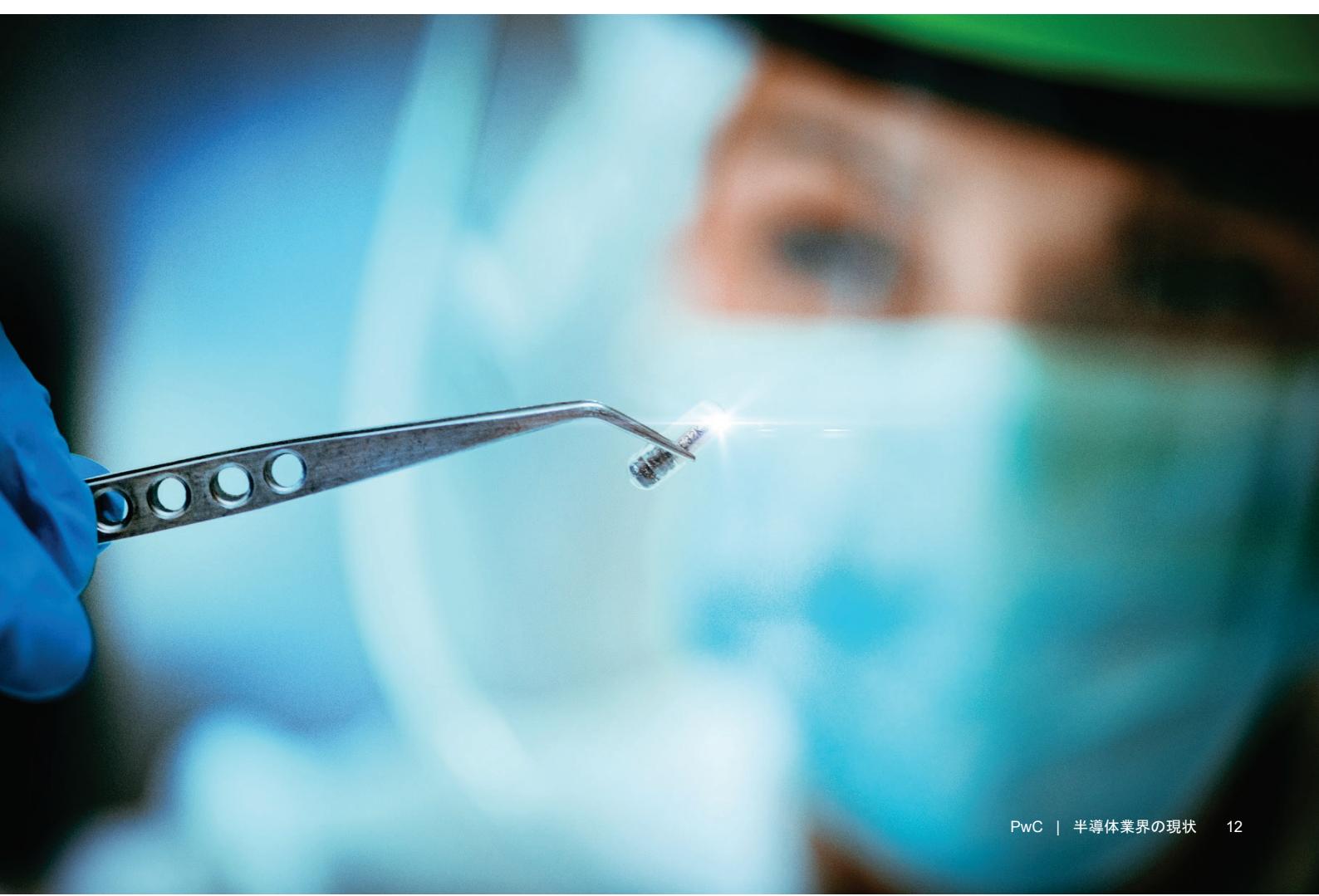
Yoo-Shin Chang
Strategy&韓国、パートナー

NAND市場は2023年第3四半期に底を打ち、その後回復に向かいました。ベンダーは過剰供給と価格下落に對処すべくウエハ生産を削減し、これによりビット成長は低下しましたが、これはNANDフラッシュメモリの生産数を意図的に減らすものでした。この供給縮小の戦略によって需要と価格のバランスが再び整い、2023年末までに平均販売価格(ASP)は安定しました。2024年は設備投資が13%減少したため、NANDフラッシュの供給量は引き続き抑制されており、生産拡大は制限されています²。

市場は2025年までに回復すると見られており、ベンダーはAI対応機器によって高まる需要予測に対応すべく、アイドル状態であった製造設備の稼働を再開しました。実際、2025年に予想される回復は、NANDの潜在的なスーパーサイクルの始まりを告げる可能性があり、ここから高成長が継続することが期待されています。

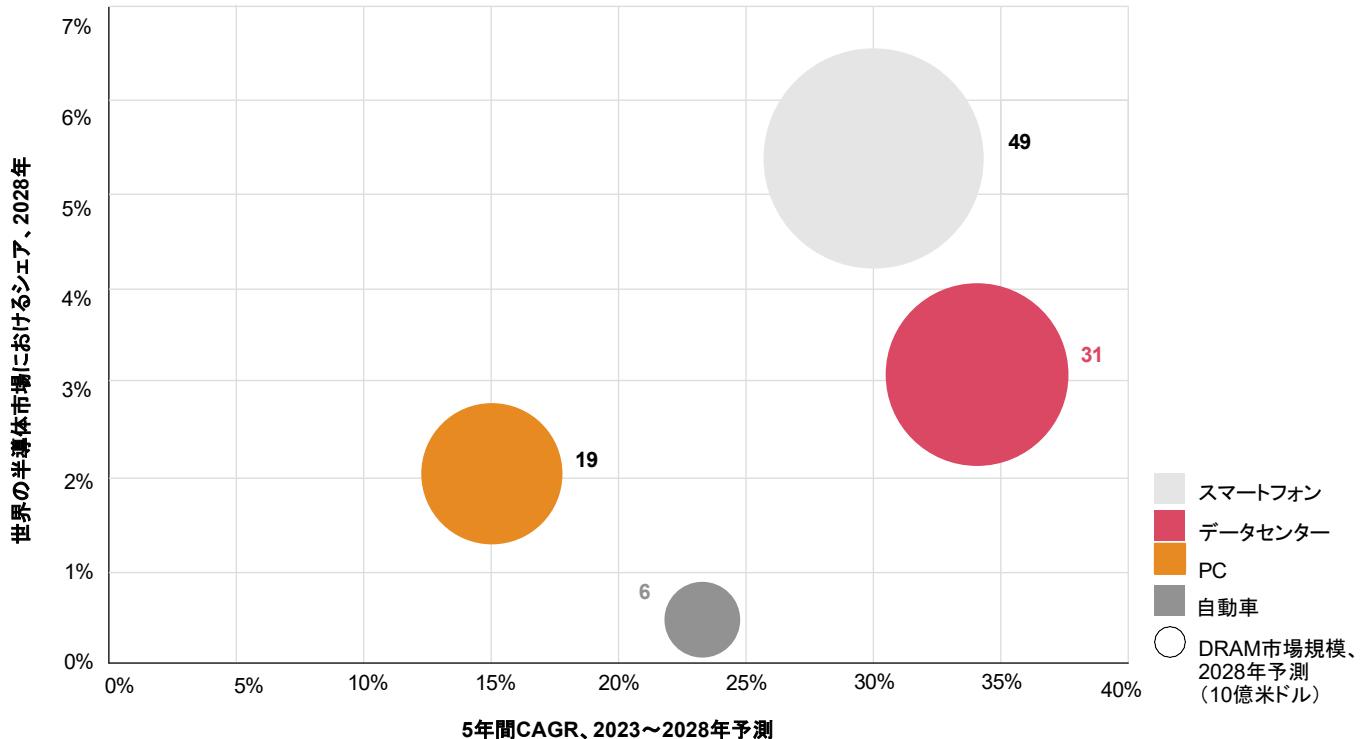
AIを搭載したスマートフォンにはさらに高度なストレージソリューションが必要となり、NAND需要を刺激し続けることでしょう。スマートフォンは引き続きNAND市場の最大セグメントであり、2028年までに世界の半導体市場の5.4%を占め、CAGRは30%と予測されています(次頁の図表5参照)。データセンターセグメントは急速に成長、2028年までに世界の半導体市場の3.5%に達し、大規模なAIワークロードとストレージインフラの拡大により、CAGRはさらに高い33.4%と見込まれています。PCは、AI用途によってより高度なストレージ機能へのニーズが生まれていることから、市場シェアは2.1%の予測です。車載メモリは、現在はNAND市場のごく一部を占めるに過ぎませんが、先端メモリソリューションへの依存が高まるとともに、2028年までのCAGRは23%と、力強い成長が見込まれています(次頁の図表5参照)。

NAND市場は、CAGR
23%と
力強い成長が
見込まれています。



図表 5

世界の半導体市場における用途別NAND型フラッシュメモリ市場シェア(%)および5年間のCAGR(%)、2023～2028年



出所:Omdia 2024年第3四半期

NANDストレージの拡張:QLC方式と1,000層技術への移行

NANDメモリセルはチップ内に層状に積層され、層が多いほどストレージ容量は大きくなります。2024年にはNANDベンダーの大半が200層を超えると予想されていますが、層数が増えるにつれて技術的な課題も発生します。1,000層のNAND開発が進む中、さらなる進歩を促進すべく新たな技術が導入され、今後10年間の半導体産業が形作られることでしょう²。

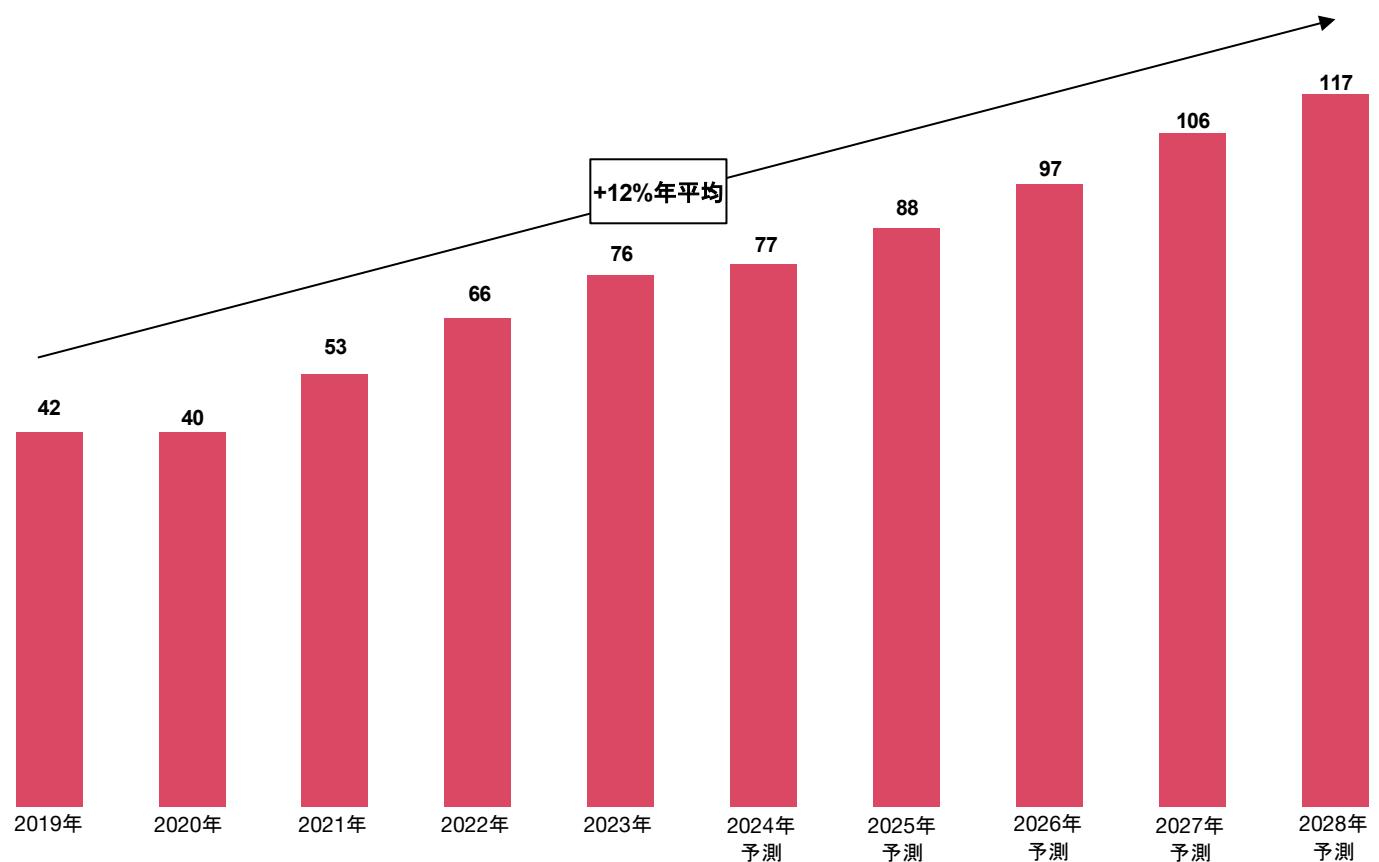
特にAIワークロードによる大容量ストレージへの需要の高まりに対応しようと、NANDベンダーは1つのメモリセルに3ビットを格納するトリプルレベルセル(TLC)テクノロジーから、1つのセルに4ビットを格納し、より低コストで高密度なストレージを実現できるクアッドレベルセル(QLC)テクノロジーへと移行しつつあります。2029年までに、QLCはNAND市場の50%以上を占めると予想されています²。データセンター向けQLCベースのエンタープライズSSDではSK Hynix/Solidigmが先行しており⁷、また、Micronなどの主要ベンダーは、メモリセル間の干渉を低減しクライアントSSDの拡張性を向上させるチャージ・トラップ・フラッシュ(CTF)技術を採用しています。TLCと比べると、QLCは性能と耐久性の点においてデメリットがあります。ただしこれらの課題は、特定のワークロードに合わせたオーバープロビジョニングやハイブリッドソリューションといった手法によって対処されつつあります。

セクション3

本格化する車載用半導体

車載用半導体市場は、EVの普及とSDVの台頭により大きく変貌しつつあります。2023年には世界の車両生産台数がコロナ禍以降初めてコロナ禍前の水準を上回ったことは、自動車市場が回復し、正常な状態に戻ったことを示しています。今後5年間のCAGRは8.9%で、2028年末までに1,170億米ドルの収益に達すると予測されています(図表6参照)。

図表6
車載用半導体市場、2019～2028年(単位:10億米ドル)



出所:Omdia 2024年第3四半期

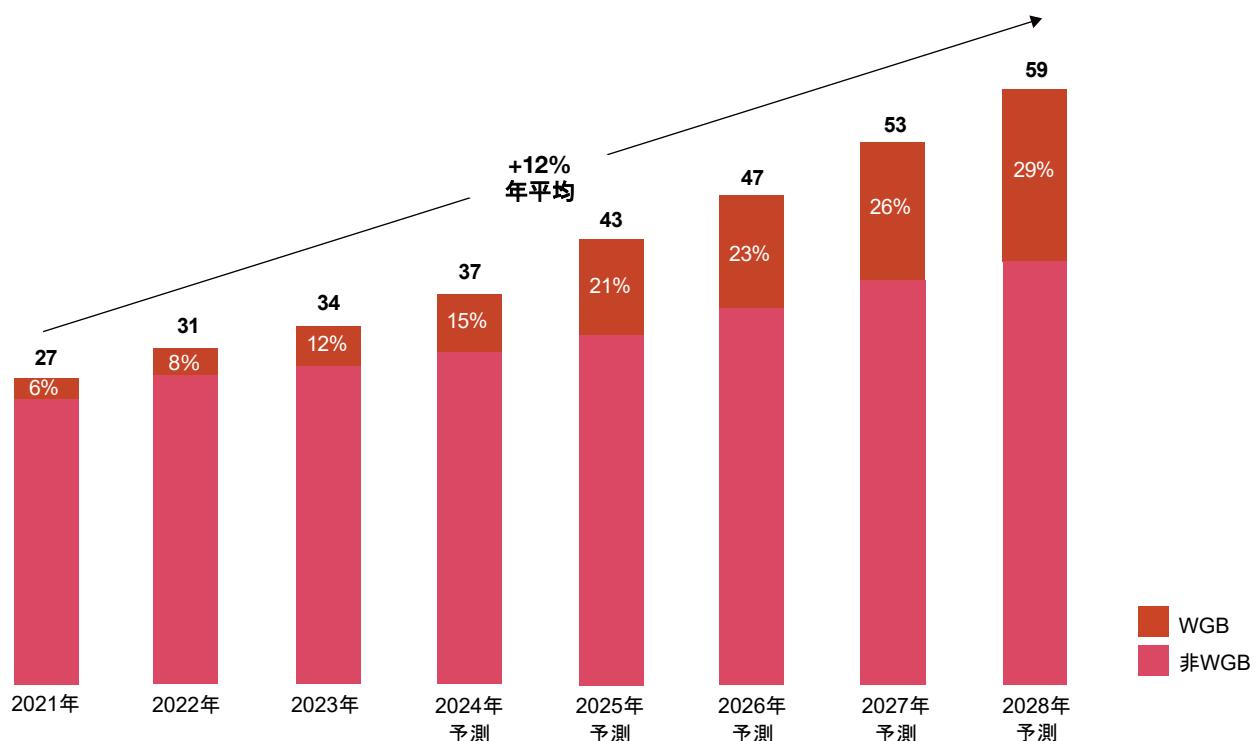
電動化が後押し:EV革命を牽引するパワー半導体

車両の電動化は、パワー半導体に大きな市場機会をもたらしました。パワー半導体はさまざまな車両用途に不可欠であり、特にメインインバータ、DC/DCコンバーター、車載充電器(OBC)、バッテリー管理システム(BMS)などのEVシステムで重要な役割を果たしています。テクノロジー調査・アドバイザリーグループのOmdiaによると、車両1台あたりのパワー半導体の価値は、EVの方が内燃エンジン車(ICE)よりも6倍高いとのことです。その結果、2023年の自動車市場における売上高は前年比30.2%増の210億米ドルに達し、今後も二桁成長が続く見通しです。

EVの普及とバッテリー技術の進歩がこの分野のイノベーションを押し進めています。炭化ケイ素(SiC)や窒化ガリウム(GaN)などのワイドバンドギャップ(WBG)半導体や、酸化ガリウム(Ga₂O₃)などの新興材料が、自動車業界に変革をもたらしています。WBGデバイスは従来のシリコン(Si)に比べてエネルギー効率に優れ、電力密度が高く、高温でも機能するため、インバータ、充電器、DC/DCコンバーターなどの重要システムに最適です。例えばSiCとGaNであれば、スイッチング速度をより早くし、より優れた熱管理が可能となるため、車両の航続距離と性能の向上に欠かせない、部品の小型化と軽量化を実現できます。

Infineon、STMicroelectronics、Onsemi、ROHM、Nexperiaなど大手のパワー半導体業者はWBG技術に一層力を注いでいます。そのため、パワーエレクトロニクス分野でのシェアは大幅に伸びています。2023年のWBGの車載パワー市場シェアは12%でしたが、Omdiaの予測によれば、年間収益は2028年までに168億米ドルに成長し、29%のシェア拡大とされています(図表7参照)。

図表7
パワーディスクリートおよびモジュール市場、2021～2028年(単位:10億米ドル)



出所:Omdia 2024年第3四半期

シリコンとWBG:競争か共存か?

コンパクトシティEVやハイブリッド車のような低電力でコスト重視の用途にはシリコンが今も主流ですが、高性能なEVにはSiCのようなワイドバンドギャップ材料が不可欠となってきています。SiCが特に適しているのは、電力密度とエネルギー効率が非常に重要な、航続距離の長い高性能モデルのメインインバータです。ただし、SiCは結晶成長過程でエネルギー消費が高く複雑であるため、依然としてコストの高さが課題となっています。しかも硬くてもろいゆえ、加工は難しく、歩留まりもよくありません。

一方で、大手企業が新しい生産能力に多額の投資を行っているため、SiCの価格は今後数年で下落するとみられています。近年、既存企業に対抗すべくSiC市場に参入したサプライヤーが中国だけで50社以上にのぼることも、この動向を後押ししています。競争の激化は、SiC製造に必要な多額の設備投資と相まって、この統合を助長し、産業が発展するにつれて競争環境の様相を変えていく可能性が高いでしょう。

新たな生産能力が入ってきたことで、供給過剰と急激な価格下落の可能性という懸念が生じている一方、200mm SiCウエハへの移行には依然として大きな課題があります。プロセスが複雑で設備の利用も限られているため、リードタイムが長く、能力増強に遅れが生じています。こうした遅れは当面は供給過剰のリスクを軽減しますが、将来の市場ダイナミクスは、特に欧州と米国における今後数年間のEV需要の伸びによって大きく異なるでしょう。

これに対し、GaNはより大きなシリコンウエハと互換性があり、既存の製造インフラを利用すれば、コストを削減できる可能性が多いにあります。このため、GaNは特に急速充電器や電力コンバーターにとって魅力的です。部品サイズと重量を減らせるため、車載充電システムにも最適です。現在、電圧範囲の拡大と信頼性の向上を目指して研究が進んでおり、技術の進歩とともにGaNは高電圧用途でSiCと対抗することになるでしょう。Infineonによる世界初の300mm GaNウエハ技術の開発など、最近のブレークスルーは製造コストの削減に向けた大きな一歩であり、GaNを今後の強力な競争相手として位置づけています。GaN市場も、2023年のInfineonによるGaN Systemsの買収や2024年後半のRenesasによるTransphormの買収予定など、M&A活動が盛んであることから、統合が進むとみられています。

自動車産業が進化し続ける中で、パワーエレクトロニクスにおいては、さまざまな車両タイプの特定用途のニーズに基づいて技術が共存すると思われます。すなわち、どの半導体技術を選ぶかは、性能、電力、コストに関する具体的な要件に大きく依存するということです。これらの技術が進化するにつれて、さまざまな用途で相互に補完し合いながら、EVの設計と効率の可能性の限界を押し広げていくことになるでしょう。



パワー半導体市場の中で変りゆく情勢

パワー半導体のバリューチェーンはいくつかの重要な段階で構成され、それぞれの段階が最終製品の性能とコスト構造に関与しています。バリューチェーンは基板の作成と結晶成長から始まりますが、ここではSiやSiCなどの原料を高品質のウエハに加工し、半導体デバイスの基盤を作ります。次に、ウエハ処理中に、これらのウエハ上に半導体構造を形成し、基本的な電気特性を与えます。その後ウエハはひとつひとつのチップへと切断(ダイシング)され、ディスクリート部品として使用される他、パワーモジュールに組み立てられます。最終段階では、最適な熱管理、耐久性、性能を確保するため、デバイスを保護構造で囲むパッケージングが行われます。複数のチップを組み合わせてより高レベルの電力を処理するパワーモジュールには、パッケージングが特に重要となります。

EVパワーエレクトロニクスのメーカーでは、それぞれ特定の電力要件と効率要件に合わせて設計されるメインインバータ用のディスクリートソリューションとパワーモジュールソリューションがどちらも使用されます。SiCをベースとしたパワーモジュールとディスクリートソリューションでは、基板作製段階と結晶成長段階が特に重要であり、パワーパッケージの総付加価値の35~45%に寄与します(次頁の図表8参照)。これは、高電圧車載用途においてSiCが優れた効率と性能を実現するのに欠かせない、高品質なSiC基板を製造することがいかに複雑であるかを示しています。SiC技術の進歩と生産の拡大とともに、これらの初期段階に起因する価値は減少すると見られています。製造プロセスの向上と規模の経済性によってコストが削減されれば、SiCをベースとしたソリューションの競争力が高まり、より広範囲なEVへの採用が進むことでしょう。

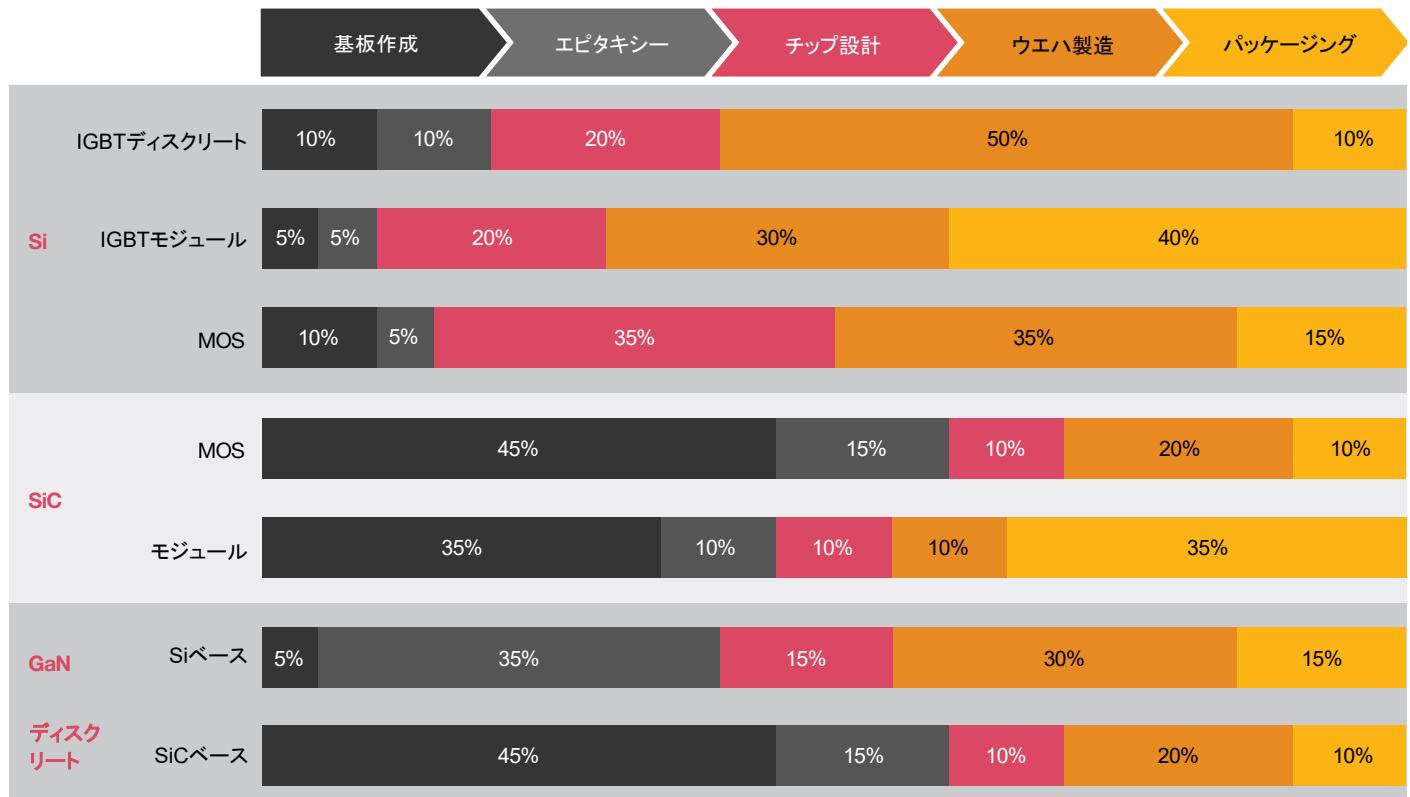
パワーモジュールソリューションでは、パッケージング段階(モジュール設計と組み立ても含まれます)が総付加価値の35~40%を占めます(次頁の図表8参照)。このステップは、モジュールの高熱ストレス耐性および機械的ストレス耐性を確保し、信頼性と性能の両方を向上させるために不可欠です。半導体メーカーはパッケージングの重要性が高まっていることを認識し、独自のパワーパッケージの開発に力を入れています。このようにして半導体メーカーはチップ設計と製造にとどまらず、バリューチェーンのコントロールを広げることができます。この垂直統合によって、製品の差別化、性能の最適化、競争力強化を実現できます。バリューチェーンにいる企業がこれらの戦略を追求するにつれて、競争情勢は進化し続けるでしょう。

“ SDVへの流れは、継続的に更新、改善が可能なソフトウェアを用いて車両機能を実装することができます普及していることを反映しています。ハードウェアとソフトウェアの大部分が分離されているため、この開発によりさらにカスタマイズを増やし、顧客のための柔軟性を高め、イノベーションサイクルを短縮することができます”

Tanjeff Schadt
Strategy&ディジタル、パートナー

図表8

バリューチェーンの各ステップが寄与する典型的な価値、パワー半導体タイプ別



出所: 2023年初頭の調査に基づくPwC分析

ハードウェアからソフトウェアへ: SDVと半導体需要

SDVへの流れは、継続的に更新、改善が可能なソフトウェアを用いて車両機能を実装することがますます普及していることを反映しています。ハードウェアとソフトウェアの大部分が分離されているため、この開発によりさらにカスタマイズを増やし、顧客のための柔軟性を高め、イノベーションサイクルを短縮することができます。半導体によって大きく定義される現代の電気／電子(E／E)アーキテクチャは、SDVを実現するための基盤を形成します。自動車産業がゾーンアーキテクチャやセントラル・コンピュート・アーキテクチャへ移行するとともに、高性能プロセッサの需要も高まるでしょう。この移行により、現在分散ECUシステムを管理している従来のコンピューティング機器およびパワーマネジメント機器の必要性が減ることで、車両エレクトロニクスの効率化が進むでしょう。

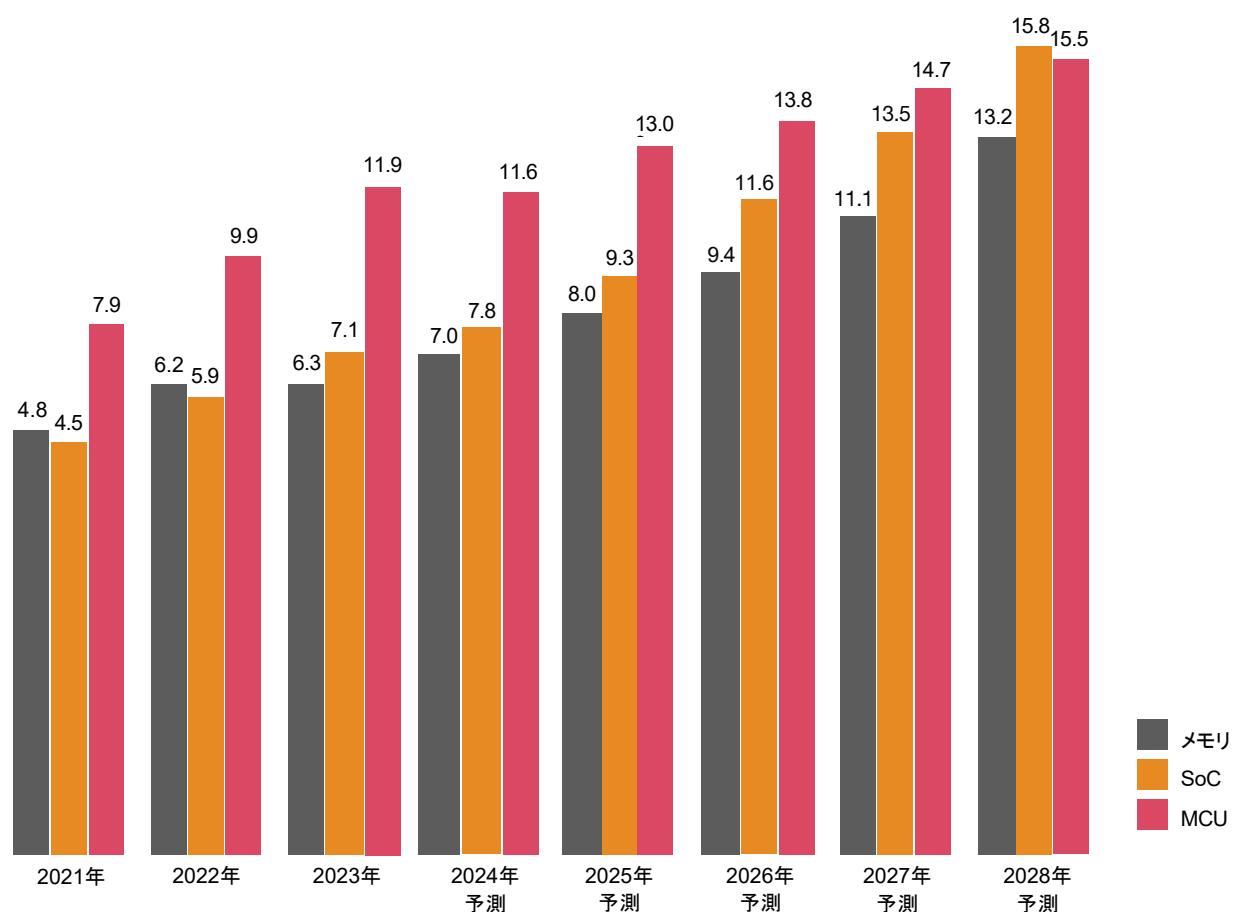
SDVの核となるのは、一般にシステム・オン・チップ(SoC)と呼ばれる先端プロセッサです。これらのSoCは、中央処理装置(CPU)、メモリ、周辺機器などのさまざまな機能をシングル・チップ・アーキテクチャに統合します。このような統合は、リアルタイムデータ処理、先進運転支援システム(ADAS)の制御、セキュリティモジュール、インフォテインメントシステムなど、SDVに必要な複雑なソフトウェアタスクを管理するために不可欠です。さらに、GPUとニューラル処理装置(NPU)へのニーズが高まっていることは、これらが先端機能の中心となる機械学習アルゴリズムを強化する役割があることを反映しています。車載用SoC市場は2028年に160億米ドルに達すると見込まれており、その間のCAGRは17%との予測です(次頁の図表9参照)。

SoCは中央処理に不可欠ですが、マイクロコントローラー(MCU)は特定の制御タスクと周辺インターフェイスの処理に依然として非常に重要です。MCUはリアルタイム操作に特化しており、エンジン制御、組み込みセンサー、バッテリー管理システムなど、正確なタイミング、電力効率、信頼性が特に必要とされるシステムに導入されることが多いです。

最新のMCUは、イーサネット、Wi-Fi、Bluetooth、車両と外部ネットワーク間のリアルタイムデータ交換に依存するV2X(車両とあらゆるものと接続する)通信など、先進的な接続オプションをサポートしています。

メモリソリューションはSDVで生成される大量データの保存とアクセスに不可欠です。これらの部品は複雑な車両ソフトウェアシステムの高速データアクセスおよびストレージ要件をサポートします。大容量の高速メモリによって大量のセンサーデータやソフトウェアアプリケーションが格納され、リアルタイムの処理や意思決定が可能となります。車両用半導体市場全体におけるメモリチップのシェアは、2023年は8%でしたが、2028年には11%に増加し、関連収益は130億米ドルに達すると予測されています(図表9参照)。

図表9
車両用SoC、MCU、メモリの部品タイプ別予測、(単位:10億米ドル)



出所:2023年初頭の調査に基づくPwC分析

セクション4

デカップリングへの適応: レジリエンスを高めるための戦略

世界の政治情勢が変化する中、半導体産業は地政学的緊張の中心となっていました。輸出管理規制と現地調達要求は従来のグローバルなサプライチェーンを困難にしています。テック企業の間では、こうした混乱に対するレジリエンスを構築する重要性が高まっています。この不安定な環境においては、安定したサプライチェーンの確保、リスク軽減、競争力維持が今や成功のための重要な要素となっています。

米国と中国の技術的・経済的分断(デカップリング)

近年、米国と中国は技術面、経済面において、徐々に事実上のデカップリングへと向かっています。このような動向は、サプライチェーンのさまざまな段階で両国に大きく依存している半導体産業に大きな影響を及ぼします。その結果、デジタル技術とコネクティビティ技術に重点を置いた米国中心の技術圏と中国中心の技術圏がどちらも出現しています。グローバルサウスでは、中国のテック企業がデジタルインフラの大規模プロジェクトを数多く推進しています。

米国による関連法制措置には、輸出管理改革法(2018年)、外国企業説明責任法(2021年)、安全機器法(2021年)、米国CHIPS法(2022年)などがあります。現在は米国国防権限法(NDAA)第5949条の施行規則案について議論されています。これは米国政府機関が、特定の中国製半導体を搭載した電子機器やデバイスを取得することを禁止するものです。可決されれば、これらの禁止事項は2027年12月23日に施行されます。過去には、当初政府部門に課された規制が数年のうちにより広範な民間市場に拡大することがしばしばあったことから、これらの半導体規制も最終的にはより広範な産業に影響を及ぼす可能性があるでしょう。

その一方で中国本土でも同様の規制を構築しており、特に「中国製造2025(Made in China 2025)」(2015年以降)、ITアプリケーションイノベーション(ITAI)プログラム(2016年以降)、輸出管理法(2020年)、データセキュリティ法(2021年)、反外国制裁法(2021年)、外国国家免責法(2023年)などが挙げられます。中国政府は中国の自動車メーカーに対し、2025年までに半導体現地調達率を25%まで引き上げることをすでに要請しています。また、対外依存度を下げるべく半導体能力の向上も急速に進めています。

世界の半導体製造において特定の地域が重要な役割を果たしていることを考えれば、地政学的緊張の高まりによって主要なサプライチェーン経路やパートナーシップに重大なリスクをもたらすことは明らかです。緊張が高まれば、経済的圧力や貿易摩擦も高まり、企業活動やサプライチェーンの安定に悪影響が及ぶおそれがあります。その影響がどの程度になるかは、そうした展開の激しさやスピードによることは確かですが、それでも、半導体産業における生産と流通の一貫性維持には広範囲にわたり影響するでしょう。

中国政府は中国の
自動車メーカーに対し、
2025年までに半導体
現地調達率を

25%

に引き上げることをすでに
要請しています。

地政学レジリエンスを高める戦略

半導体に係る地政学レジリエンスは、グローバル市場にサービスを提供するテック企業にとって重要な成功要因になりつつあります。地政学リスクを認識し、効果的な戦略を採用することで、企業はイノベーションと競争力を維持しつつ、レジリエンスを高めることができます。世界情勢が変化し続ける中、これらの課題を先取りすることが長期的な成功の鍵となります。

マルチソース戦略

サプライチェーンの混乱を最小限に抑えるため、大手企業はマルチファブ戦略およびマルチソーシング戦略を採用しています。製造拠点とサプライヤーを多様化することで、地政学的事象が事業に影響を及ぼすリスクを軽減することができます。

リスクベースのアプローチ

企業はリスクベースのアプローチによって、半導体部品のライフサイクル全体にわたるリスクを特定、評価、軽減することができます。地政学的混乱に対するリスク基準が強化されたこのアプローチにより、半導体、モジュール、サブコンポーネント、または製品レベルで潜在的な問題に対して事前に対処することができます。



製品ローカリゼーションのアーキタイプ

規制が進むにつれ、政府は特定の市場に合わせたローカルまたはハイブリッドのソリューションを義務付ける可能性があります。こうした変化に対応するために、テック企業はコンプライアンスを強化し、技術スタックのレジリエンスを強化する戦略を採用することができます。プロアクティブ(事前対策的)なアプローチとしては、地域によって製品を区別する他、現地要件を満たすために各半導体のドロップイン(置き換えが簡単な)代替品を使用するなどがあります。あるいは、グローバル製品を維持し、必要時のみ調整を行うというリアクティブな(受け身の)スタンスをとってもよいでしょう。

人材戦略

ローカリゼーションを成功させるには、相当の人員リソースが必要となります。企業が地域の能力構築に取り組むことにより、スキルの高い専門家の需要が増加します。こうした取り組みを支えるためには、強力な人材パイプラインが必要です。2023年のStrategy&による調査「人材ギャップを解消(Bridging the talent gap)」⁸によると、欧州連合(EU)が設定したグローバル市場シェア20%という地域目標⁹を達成するには、欧州の半導体環境だけでも2030年までに約35万人の専門家が必要とのことです。

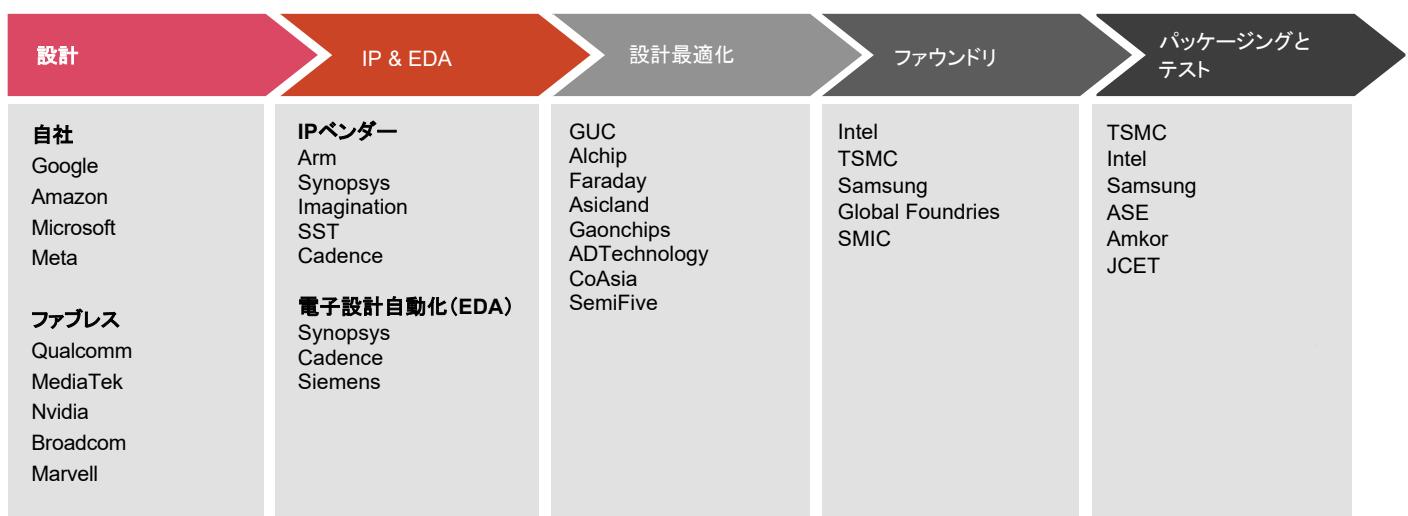
セクション5

専用シリコンのルネサンス

特定用途向けの専用ICの市場は、性能、効率、およびセキュリティへの需要の高まりによって、今後10年間で復活すると見られています。この需要の急増により、設計サービス、ファウンドリ、電気設計自動化(EDA)ツールなどのバリューチェーンが拡大しています。GUC、Alchip、ADTechnologyなどの企業が専門的な設計サービスを提供している一方で、オープン・コンピュート・プロジェクトのチップセット市場などの取り組みでは、事前設計されたプロセッサ部品へのアクセスがデモクラタイズされています。現在では、Fraunhofer Institute for Integrated Circuits(IIS)などの小規模なファウンドリサービスが、数百万ではなく数万規模のカスタムICの生産に対応しているため、ハイペースケール対応能力を持たない企業でもこの技術を利用できるようになっています。知的財産権の可用性の拡大、ソフトウェアツールの開発、そして設計コストの低減により、カスタムIC市場はさまざまな業界で成長し続けると見られています(図表10参照)。

しかし、特に設計コストの高い先端半導体に関しては拡張性が今も重要な課題として残っています。エレクトロニクス産業のコンサルティング企業であるIBSの調査によれば、10nmチップの設計コストは1億7,000万米ドル以上、5nmチップは5億米ドル以上です。データセンター産業の情勢は進化し続けており、データセンター・コンピューティングの四大ユーザーでデータセンターのIT設備投資の30~40%を占めるAmazon、Meta、Microsoft、GoogleにカスタムIC開発に適した環境を提供しています²。これらの企業は規模と性能のニーズがあるアプリケーションを運用しており、カスタムIC開発を戦略的優先事項としています。

図表 10
カスタムIC開発のバリューチェーン(主要企業)



出所: 一般に公開されている情報

映像処理用カスタムIC

カスタムICの開発にあたり最初にターゲットとなったのは、非常に大規模な動画配信需要のあった映像処理でした。Googleの試算では、世界のインターネットトラフィックの60%が動画配信に起因するとされています⁸。過去5年間、同社はビデオ・コーディング・ユニット(VCU)と呼ばれるカスタム・ビデオ・エンコーディングICを開発してきました。これらのICにより、GoogleはYouTubeに必要なサーバー数を大幅に減らすことに成功しました。VP9ビデオ・コーディング・フォーマットを使用して、Skylakeベースのサーバーのラック数台を20個のVCUを備えたサーバー1台に置き換えたとGoogleは報告しています。カスタムICの設計と製造への初期投資があったにもかかわらず、このVCUプロジェクトによって、3年間でYouTubeのコストが20倍から33倍削減されたとGoogleでは試算しています⁸。



MetaとTencentも独自のビデオプロセッサを開発し、性能が飛躍的に向上したことが報告されています。NETINTなど他の動画配信企業はASICを導入し、サーバー1台あたり、1ワットあたり、1米ドルあたりの性能を最大限引き出しています(次頁の図表11参照)。

ネットワークおよびセキュリティ用途用カスタムIC

他にもカスタムIC開発にあたり主要な分野となっているのが、計算指向型で大規模なワークロードを処理するネットワークやセキュリティ用途です。AmazonはOmdiaとの対話の中で、同社のインフラの約20%がネットワーク処理とセキュリティ処理の専用となっていることを示しました。このような状況から、データ処理装置(DPU)上のイーサネットコントローラと統合されたカスタムIC、通称Nitroカードが開発されました。自社開発したDPUは、仮想マシンデータプレーン処理(カプセル化やルーティングなど)、暗号化、およびその他のネットワーク機能を処理しています。AmazonはNitroカードの機能を拡張し、ストレージ制御機能、セキュリティ監視機能、システム制御機能、アナリティクス機能を追加しました。これらのタスクをカスタムICにオフロードすることで、AmazonはCPUコアを解放し、今度はそれをIaaS(Infrastructure as a Service)として企業に販売しました。このプロジェクトの成功は、ハードウェアベースの信頼の基点とNitroハイパーバイザーのためのカスタム・セキュリティ・チップ開発へつながりました。



AI処理用カスタムIC

AIは最も性能集約型で商業的に重要な用途の1つであり、これによりあらゆる大手クラウド・サービス・プロバイダーはカスタムICの開発を進めることができます。これらのカスタムチップはAI用途の高速化や効率性の向上によって、プロバイダーに競争力を与えるよう設計されています。Googleは、テンソル処理装置(TPU)を使用してAI用のカスタムICを開発した最初のクラウドプロバイダーであり、当初はAI推論専用として導入されました。Googleは2024年までに、主にAI推論タスク用として100万個以上のTPUを導入すると予想されています²。



躍進:中国のシリコン競争におけるAIカスタムチップ

GPUの対中輸出規制が続く中、中国はAI用のカスタムIC開発が不可欠となっていました。中国のクラウドプロバイダーは競争力を維持するため、AI推論とトレーニングの両方に最適化されたカスタムICを構築しています。

2023年、TencentはカスタムASIC、Zixiao v1を推論タスク用のNVIDIA A10 GPUの代替品として位置づけ、導入を拡大し始めました。TencentはAIトレーニングにもZixiao v2Proを導入しており、その性能はNVIDIA L40 GPUに匹敵すると主張していますが、これは間違いなく米国の制裁に反発したことです。

Huaweiも2019年のAscend 910を皮切りに、AIワークロード向けのカスタムシリコンを開発しています。しかし、米国の制裁によってサプライチェーンが混乱していることから、SMICとの協業で新しいバリエントを開発することとなりました。2023年8月、HuaweiとiFLYTEKは、SMICのN+2 7nmプロセスで製造されたとされるAscend 910Bを搭載したStarDesk AI Workstationを発表しました。以来、Huaweiは数百ユニットから2万ユニットまでのクラスターサイズで10万個以上のAscend 910Bを導入してきました²。

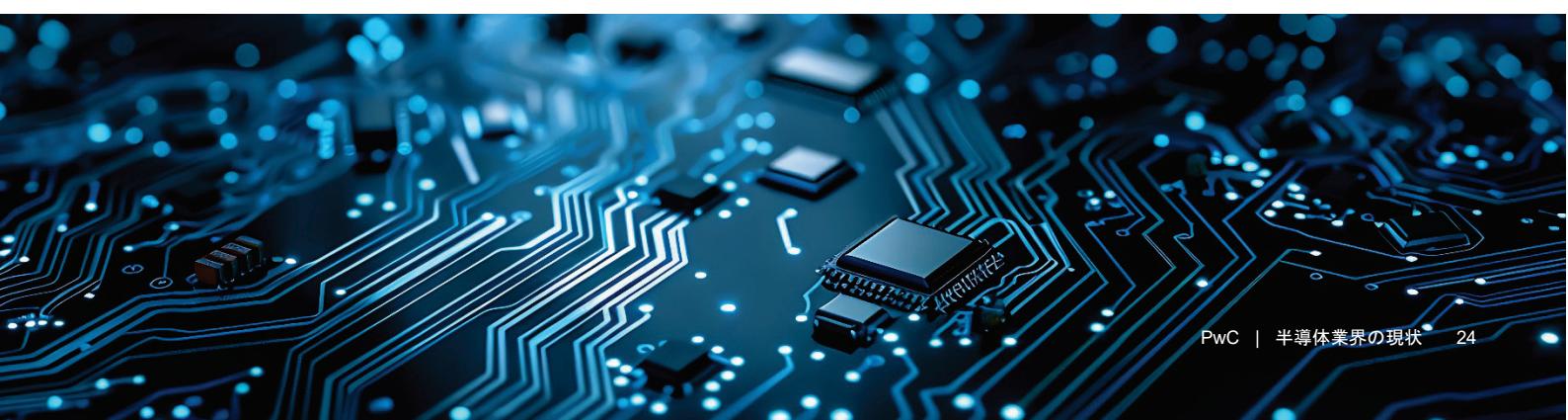
図表11
専用ICを有する代表的な企業

企業	チップタイプ	チップ名	タスク
映像処理	Google	映像処理装置	映像処理およびエンコード
	Meta	映像処理装置	映像処理およびエンコード
	Tencent	映像処理装置	映像処理およびエンコード
	Netint	映像処理装置	映像処理およびエンコード
ネットワーク／セキュリティ	AWS	データ処理装置	データブレーン処理 暗号化などのネットワーク機能
	Cisco	ネットワーキング	ネットワーキング、スイッチング、およびセキュリティ
AI処理	Google	AIアクセラレータ	AI処理(トレーニングと推論)
	AWS	AIアクセラレータ	AI処理
	Microsoft	AIアクセラレータ	AI処理
	Meta	AIアクセラレータ	AI処理
	Tencent	AIアクセラレータ	画像および音声認識
	Huawei	AIアクセラレータ	AIワークロード
	Baidu	AIアクセラレータ	AIコンピューティング
	Alibaba	AIアクセラレータ	AI推論

出所: 一般に公開されている情報

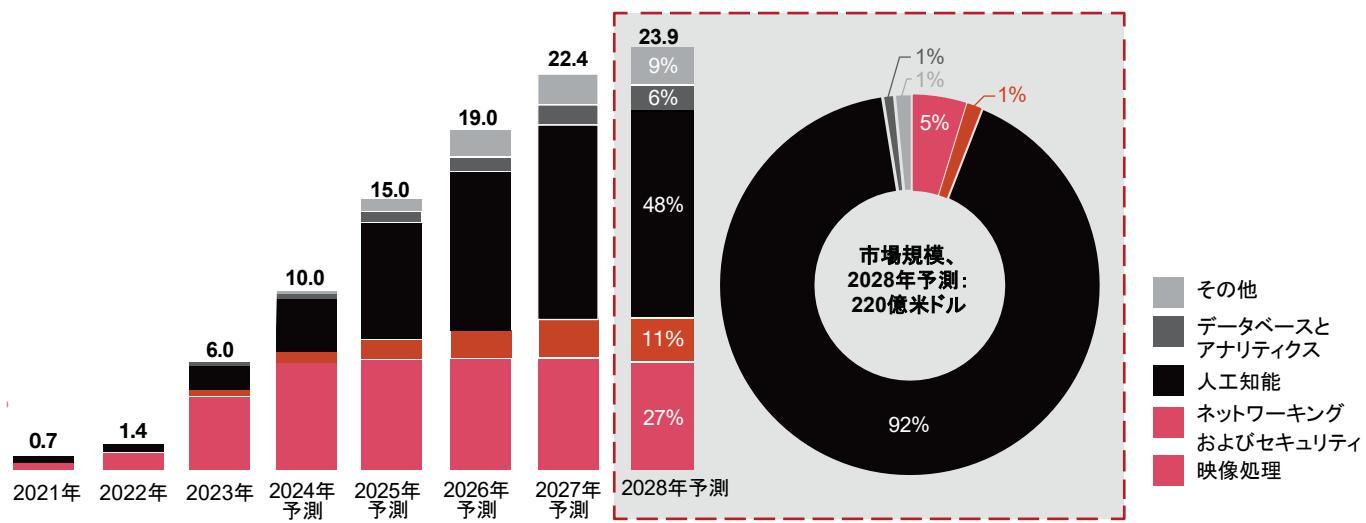
次の波: セキュリティ、Webサービス、データベース、アナリティクス

カスタムIC開発の次の波では、セキュリティ、Webサービス、データベース、アナリティクスにおけるワークロードがターゲットになるとされています。これらの用途の規模が拡大するにつれて、カスタムICは計算効率を最適化し、性能を最大化するための鍵となるでしょう。例えば、データベース処理用に設計されたカスタムチップを使用することで、クエリの応答時間を短縮し、1台のサーバーで対応できるユーザー数を増やすことができます。この分野の初期実験はすでに進行中であり、Microsoftなどの企業ではフィールド・プログラマブル・ゲート・アレイが導入されています。同様に、急速にデジタル化する世界経済においては、サイバーセキュリティを強化する必要がありますが、セキュリティ用途用のカスタムICの開発が促進されるでしょう。ソブリンクラウドの動きにより、セキュリティ要件はすでに確立されており、これらのチップの設計を形作ることになるでしょう。



図表12

2021～2028年のデータセンター業界におけるカスタムIC(単位:100万個)および2028年の市場規模(単位:10億米ドル)



出所:Omdia 2024年第3四半期

コンピューティング市場情勢への影響

Omdiaの予測では、こうした動向によって、2028年までに年間約2,500万個のカスタムユニット需要が生まれるとのことです。これらのチップの処理能力を利用できるよう設計されたサーバーには、システムあたり平均10個以上のプロセッサが搭載されるでしょう。その結果、これらの専用プロセッサによって勢いづいたサーバーの年間導入台数は、2028年までに200万台を超えるとされています。2023年と2024年はAIへの投資がようやく大きな成果を上げ始める時であり、この期間は映像処理チップが最も広く導入されるでしょう。また映像処理チップは、AI向けに最適化されたチップよりもはるかに安価です。AI向けのチップは高価なHBMメモリと多数のコンピューティングコアが搭載されています(図表12参照)。

金額ベースでは、データセンター市場におけるカスタムICの年間需要は2028年までに約240億米ドルに達すると予測されており(図表12参照)、これはGPU分野でNVIDIAと競合できなかった半導体設計者や製造業者に大きな可能性をもたらすでしょう。Broadcom、Marvell、Intelなどの企業は、データセンターフィールドにおけるカスタムICの需要の高まりに乗じて戦略を調整しています。

特定用途向けの専用ICを最初に自社開発したのは大手データセンター事業者でしたが、この動向は自動車や医療などの産業にも広がりつつあります。自動車産業では、多くの企業がNVIDIA、Intel、Huaweiなどのベンダーと提携し、市販のプロセッサを最適化しています。例えばTeslaは、自動運転におけるニューラルネットワーク演算用に、冗長性や安全機能を組み込んだ独自のプロセッサを設計しています¹⁰。デンソーは車載用途向けのカスタムプロセッサを開発し、世界中の自動車メーカーに供給しています。BYDもバッテリー管理システム(BMS)、パワートレイン制御、リアルタイムのセンサーデータ処理用のMCUを中心に、半導体製造を社内に統合しました。

他の企業も市販のプロセッサの最適化から独自設計の開発へと移行するにつれて、車載用カスタムIC市場の価値は飛躍的に増大するでしょう。2028年には、SoCの収益158億米ドルのうち、かなりの割合を特定用途向けの専用ICが占めると予想しています。

データセンター市場における
カスタムICの年間需要は、
金額ベースで

約24億
米ドル
に達すると予測されています。

決定方法:カスタマイズか自社能力強化か

カスタムIC開発にこだわるかどうかは、各企業の戦略的ニーズ、運用規模、および技術に対してどのレベルで管理が必要かによって決まります。そこで企業は専門知識とリソースを含む社内能力を評価して、自社開発の複雑さが管理可能であるか、あるいは外部パートナーと協業する方がより適切な道であるかを判断しなければなりません。カスタマイズによるメリットを得つつリスクを軽減したい企業は、ファブレス半導体企業や統合デバイスメーカー(IDM)と提携することでカスタマイズされたソリューションが得られます。正確な技術仕様と性能仕様を満たせるよう、この提携には完全にカスタマイズされたサービスから共同設計までさまざまあります。

社内能力が高い企業であれば、カスタムICの自社開発には大きなメリットがあります。自社開発であれば、消費電力の削減や性能向上など、特定の目的に合わせた最適化が可能となり、具体的なユースケースに合わせてより深くカスタマイズできるようになります。設計プロセスを所有することで、IPセキュリティの管理が強化され、独自技術の保護が容易になります。さらに自社開発はイノベーションを促進し、知的財産のライセンス供与やニッチ製品の開発など、新たなビジネスチャンスも生み出します。

カスタマイズ性が高まることで、ハードウェアとソフトウェアを密接に統合させ、製品を特定のユースケースにより適応させることができます。このようなハードウェアとソフトウェアの相乗効果は、従来のハードウェアメーカーがソフトウェア主導の企業になりつつあるという抜本的な変化をもたらしています。このような移行は、製品提供を強化するだけでなく、新しい収益モデルを創出し、市場での競争優位性を構築します。

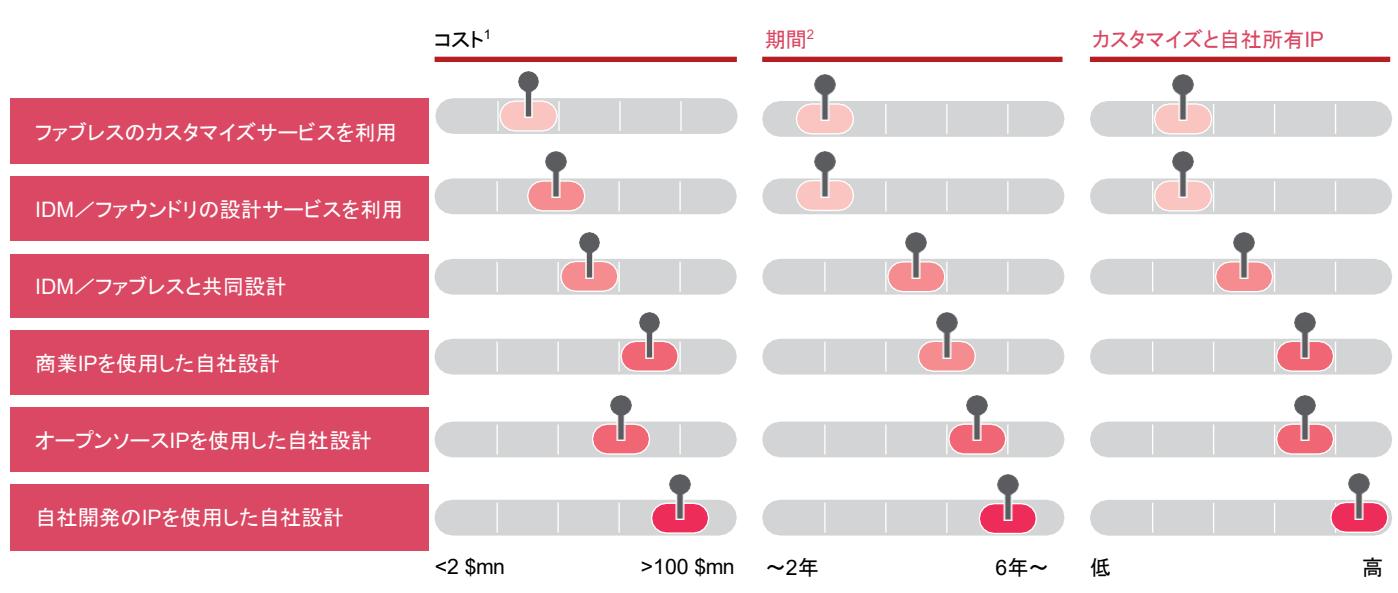


このようなハードウェアとソフトウェアの相乗効果は、従来のハードウェアメーカーがソフトウェア主導の企業になりつつあるという抜本的な変化をもたらしています”

Tom Archer

PwC米国 グローバル・テクノロジー・コンサルティング／アライアンス・リーダー

図表 13
カスタムIC開発への投資方法の評価

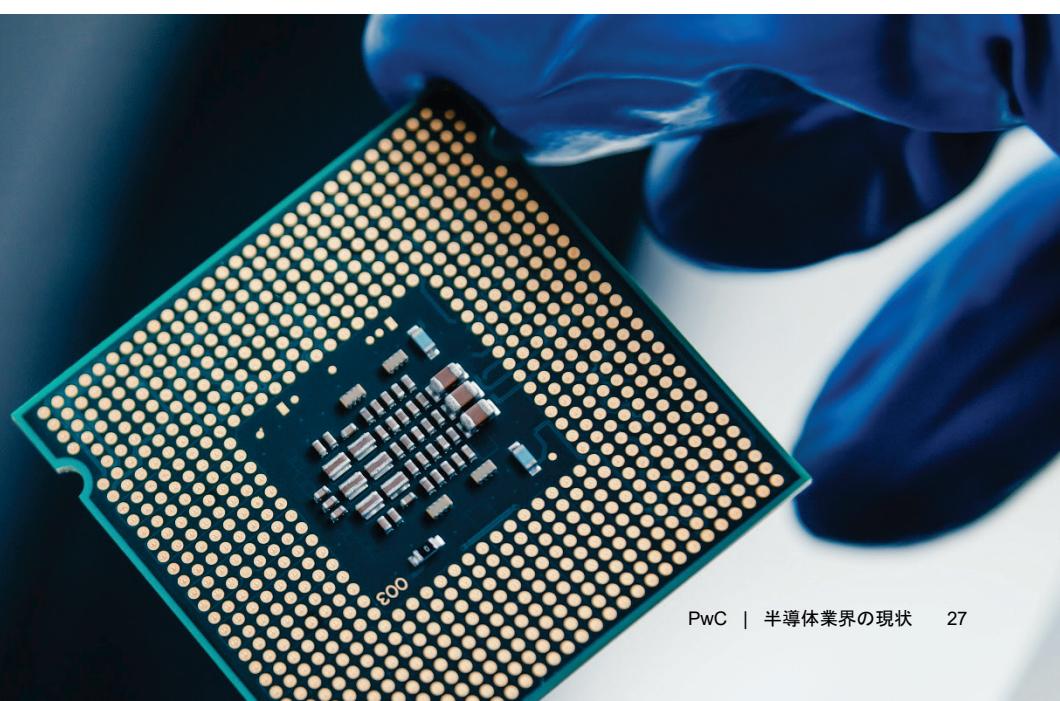


1 半導体技術(ノードサイズ)、製品タイプ、サプライヤーに応じた詳細なコスト

2 必要な能力を構築するフェーズを含みます

出所: PwCによる分析

最終的にどの方法を採用するかは、企業の目標、用途の複雑さ、カスタマイズと市場投入までの時間とのバランスによって決まります。商用IPまたはオープンソースIPを使用すれば、必要な柔軟性は維持しつつ開発期間を短縮できます。中小企業であれば、定評のある企業との提携がカスタムIC市場に参入する効率的な方法となります。一方、大企業はIPを完全に所有することで、独自のユースケースに合わせたソリューションを創り出し、最適化できるというメリットが得られる可能性があります(図表13参照)。



セクション6

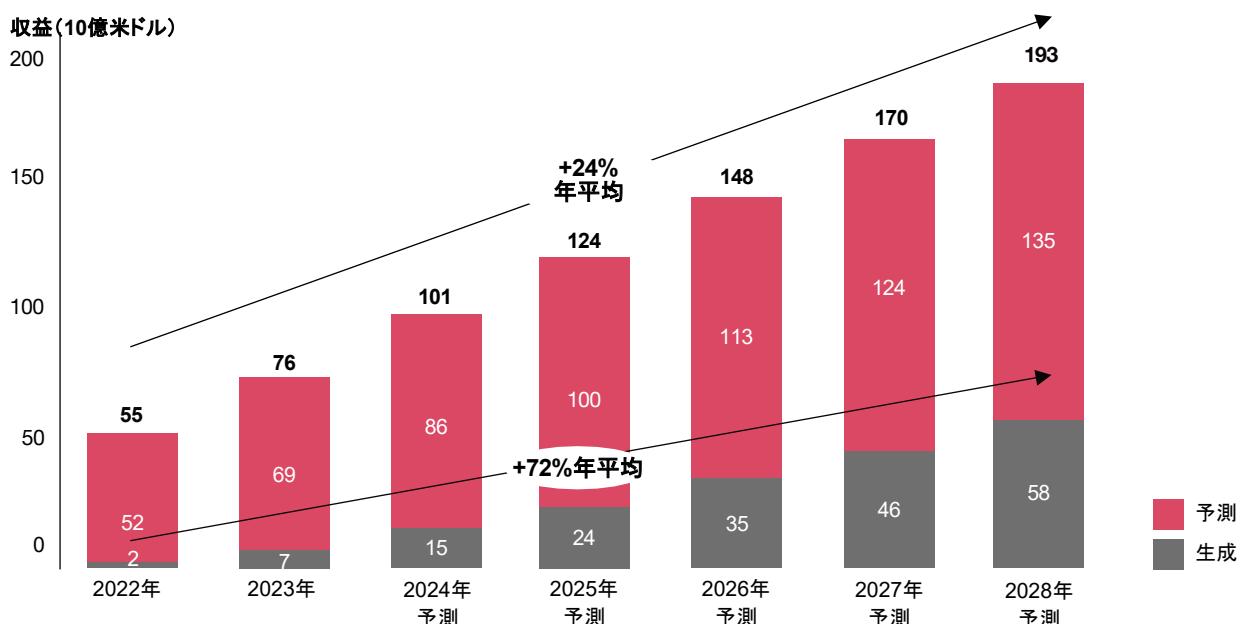
人工知能——規模から多様性へ

人工知能(AI)は急速に普及し、世界中の産業や経済に大きな影響を与えています。特に半導体産業に対しては、2つの大きな機会を提供しています。まず1つに、AI支援を活用したチップ設計と製造プロセスの最適化により、効率性の向上、エラーの減少、市場投入までの時間短縮が期待できます。そしてもう1つ、AI用途、特に予測AIおよび生成AIの爆発的な成長に伴い、先端半導体部品の需要が急増しており、今後の主要な成長要因となっています。

自動品質検査やサプライチェーン最適化などの予測AIの用途は大幅に成長し、2028年までに年間収益が1,350億米ドルに達すると予測されています(図表14参照)。これらのシステムは既存のデータを利用して結果を予測し、ビジネスプロセスを強化します。

対照的に、生成AIは学習したデータパターンに基づいて新たなコンテンツを作成するよう設計されています。既存のデータセットから学習して、テキスト、画像、音楽などを生成するアプリケーションを強化します。このタイプのAIシステムは、初期の採用からマス市場への普及へと進んでおり、2023年以降は54%のCAGRで成長し、2028年には580億米ドルの収益を計上すると予測されています(図表14参照)。

図表 14
予測AIおよび生成AIソフトウェア市場、2022～2028年(単位:10億米ドル)



出所:Omdia 2024年第3四半期

2019年から2021年の間にピークを迎えたAIの規模の競争は、2021年にGoogleのSwitch-Cが1兆6,000億個のパラメータで記録を打ち立てて以降減速しています¹¹。GPT-4や同様のモデルはこの記録を超えていませんが、典型的なモデルのサイズは増えています。2023年2月にMetaのLLaMaモデルが発表されてからは、12より小さなAIモデル（パラメータ数が50～700億個）が、特にオープンソースコミュニティで急増しています。この「ミッシングミドル」カテゴリーは、研究者や個人がハイエンドの個人用ハードウェアで探索できるほど小さいため、現在、イノベーションの温床となっています。

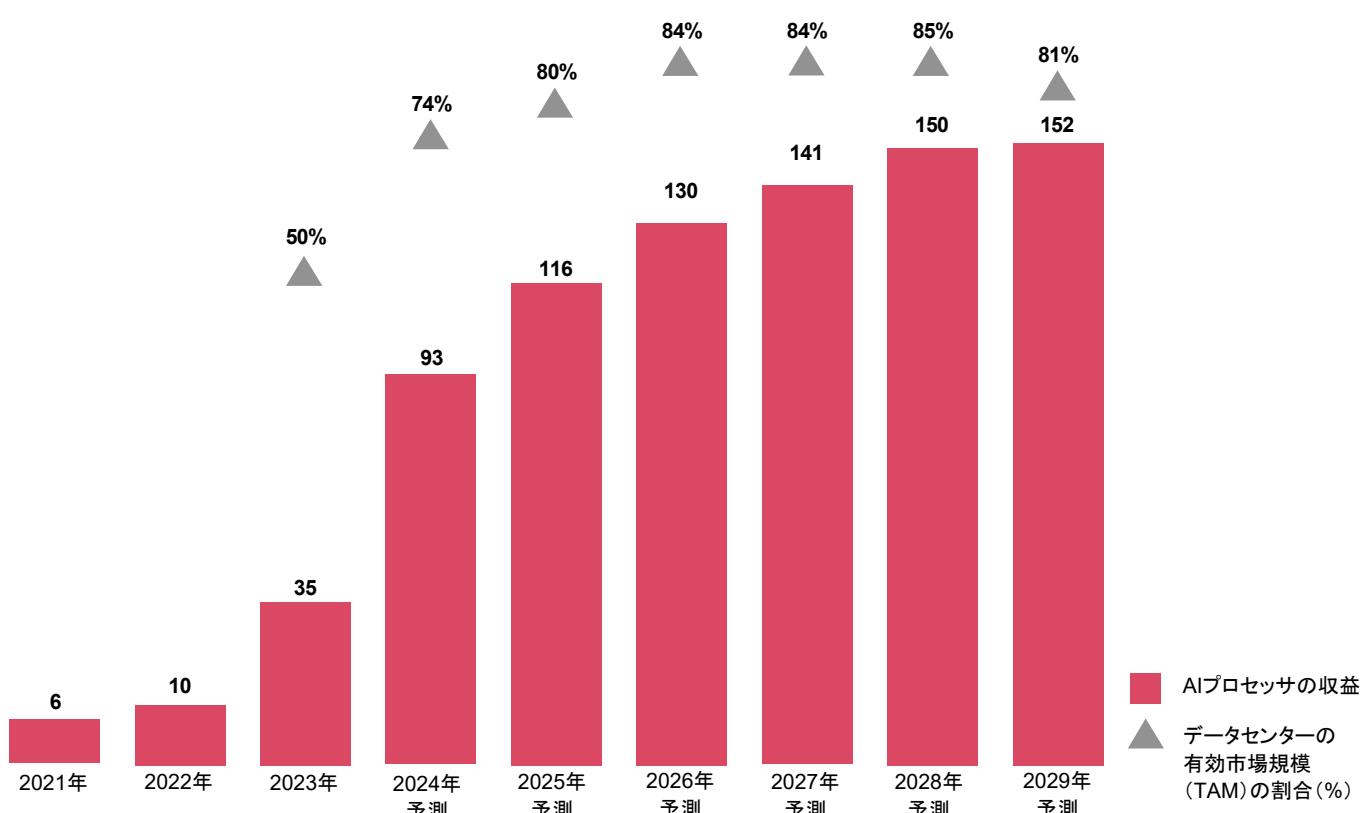
AIアクセラレータでデータセンターを変革

AIの採用、特に予測AIから生成AIへの移行は、特にAIアクセラレータやHBMといった半導体需要の急増へとつながっています。YOLO-v5など500M未満のウェイトクラスの従来の機械学習モデルと小規模なニューラルネットワークは、Phi、Gemma、Mistral-7Bなどの5～70Bのウェイトクラスの大規模なモデルへと置き換えられています。AIコンピューティングの需要が急増する中、NVIDIAのデータセンター事業が生み出す2024年度の収益は、その全てがコンピューティングやAIと厳密に関連しているわけではないとはいえ、800億米ドル以上にのぼると予測されています。その際立った2023年第2四半期の決算発表前は総収益がこの半分以下であったことを考えれば、この予測は極めて衝撃的です。

特にデータセンターでのAIシリコンの需要が急速に成長したことで、CPUとコプロセッサ（GPUまたはドメイン固有のASICなど）とのこれまでの関係は変化してきました。CPUは、GPUのビジネス・ロジック・コプロセッサになりつつあり、またはさまざまな専用アクセラレータのオーケストレータとして機能しています（図表15参照）。

図表15

クラウドおよびデータセンター向けAIプロセッサ市場、2021～2029年（単位：10億米ドル）



80%
以上

(金額ベース)のデータ
センターのプロセッサが
AIアクセラレータになるか
AI機能搭載のものとなるで
しょう。

私たちの予測では、2028年または2029年までには金額ベースでデータセンター向けプロセッサの80%以上がAIアクセラレータになるかAI機能搭載のものとなり、約1,500億米ドルの有効市場規模に貢献するとしています。この成長を促す主要因は特定のユースケースではなく、Transformerモデルアーキテクチャの採用と関係があります。もともとは機械翻訳用に設計されましたが、以来、Transformerモデルはほぼ全ての用途で最も広く使われているAIアーキテクチャとなりました。しかしその設計には、推論スループットはメモリ帯域幅に依存しているものの、メモリ要件がコンテキストウインドウのサイズに応じて二次的にスケーリングされるという重大な制限があります。現行のAI手法では、トレーニングプロセスのある時点で全対全通信が必要であり、最も遅いマシンまたはネットワークリンクの速度まで動作が遅くなってしまうことから、帯域幅、レイテンシー、そして特に予測可能なレイテンシーは非常に重要です。

直感的な印象とは異なり、十分なメモリと入出力帯域幅を実現する方が、計算能力(浮動小数点演算/秒、FLOPS)を提供するよりも大量の電力を消費します。メモリはFLOPSの約10倍、入出力ではさらに10倍の電力コストがかかります。メモリ容量と高速アクセスはどちらも不可欠であるため、単にDRAMを追加するだけでは不十分であり、広帯域幅メモリ(HBM)をアクセラレータに直接統合する必要があります。こうしたニーズからNVIDIAのB200やAMDのMI300などのフラッグシップGPUが生まれましたが、B200の場合は1キロワットを超える大きな電力を消費します²。特に半導体レベルで大幅な効率性向上を達成することは、AI用途の運用コスト削減のために不可欠です。

GPUの優位性を揺るがすカスタムアクセラレータ

データセンターにおける電力需要の増加は、GPUの資本コストと相まって、カスタムシリコンプロジェクトやAIアクセラレータのスタートアップ企業の波に拍車をかけています。これは1991年、当時ソニーの最高技術責任者(CTO)であった牧本次生氏が述べたように、エレクトロニクス産業におけるカスタマイズ(技術フロンティアでの運用時)フェーズと標準化(需要安定時)フェーズのサイクル現象を反映しています。AIは、この「牧本ウエーブ」の新たなフェーズの火付け役となり、IBM、Tesla、Huawei、Appleといったハイパースケール・クラウド・プロバイダーからは独自のカスタムアクセラレータASICが発表されました。名前が明かされていないMarvellの「Customer C」も、2026年までにAIカスタムASICを増強すると見られています¹³。

これまでのところ、カスタムASIC、特にGoogleのCloud TPUは、GPU市場を独占しているNVIDIAに唯一大きく食い込んでいます。GoogleのTPUとMetaのトレーニングおよび推論アクセラレータ(MTIA)のASICアウトソーシングパートナーであるBroadcomは、2023年のAI関連収益が3倍に増え、NVIDIAのデータセンター事業よりも速いペースで成長しています¹¹。

“

特に半導体レベルで大幅な効率性向上を達成することは、AI用途の運用コスト削減のために欠かせないでしょう”

内村公彦

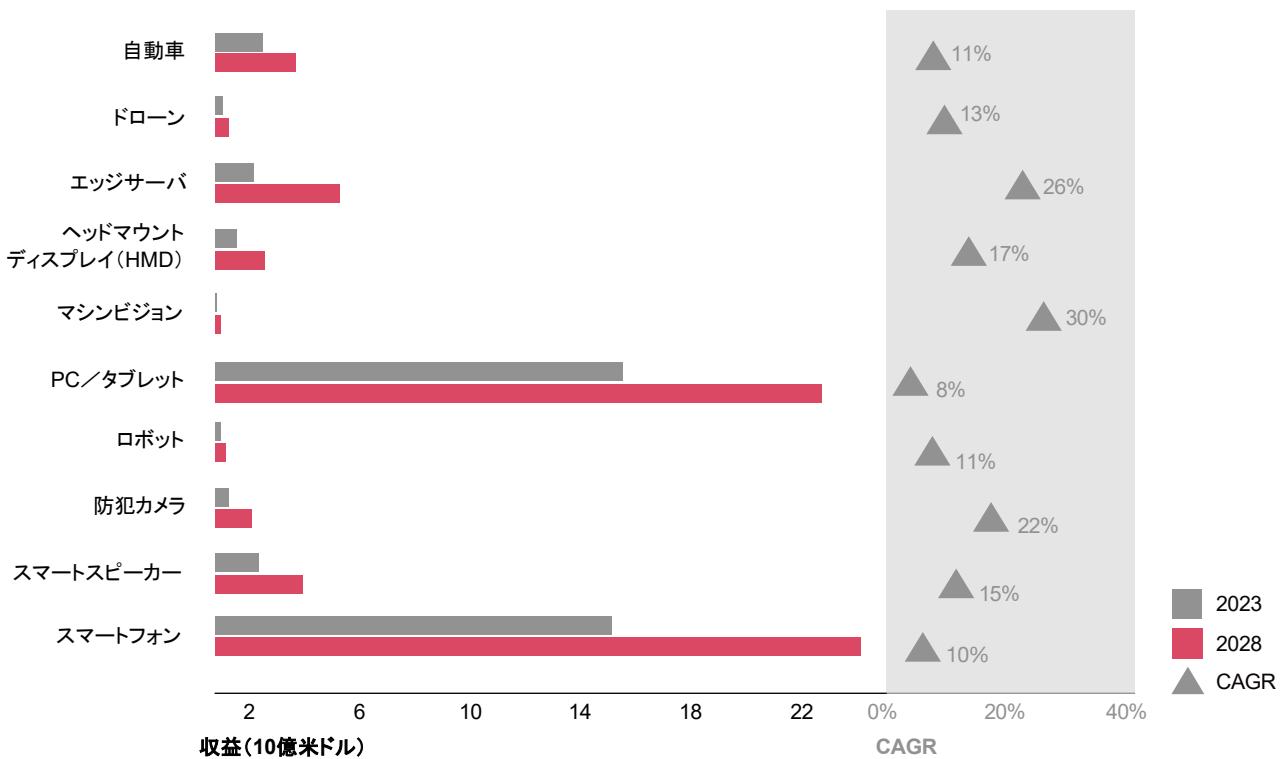
PwCJapanグループ、パートナー

AIチップのスタートアップ企業は、製品ソフトウェア開発キット(SDK)、ツール、サービス、垂直型ソリューションの構築に必要なソフトウェア開発者を引き付けることができないことが大きな原因となって、なかなかインパクトを与えることができずにいます。そうしたことから、ファウンドリとASICパートナーには、顧客が参加しやすくなるよう統合サービスを提供するという重要な機会が生まれています。これらのサービスにはパッケージングとリソグラフィの他、隣接するIPも含めなければならず、重要な点はソフトウェア領域にまで及ぶべきだということです。堅牢な開発者ツールチェーンを提供することは、導入の課題を克服する上で重要な要素です。

エッジにもAIを：アクセラレータシリコンの進化

AI用途の成長は、AIアクセラレータシリコンのエッジコンピューティングやクライアントコンピューティングへの普及によっても支えられています。スマートフォンはアーリーアダプターであり、2017年にはAppleとQualcommによって早くも専用のAIアクセラレータコアがSoCに組み込まれていました。2023年までにはスマートフォンの約66%に何らかの形でAIアクセラレーションが搭載され、今や120米ドル以下のデバイスにも搭載されるほど普及しています。Qualcommなどの企業は、モバイル機器で70～100億のパラメータモデルを実行できるようになりました¹⁴。しかし、PCは後れをとっています。IntelのMeteor Lake CPUが登場するまでは、PCのAIアクセラレーションは消費電力の大きいGPUベースのゲーム機か、はじめからAI機能を内蔵しているApple Silicon Macに限られていました。その後、Intel、AMD、QualcommがAIアクセラレーションを搭載したPC用CPUを発表しました。これらのプロセッサは、スマートフォンに見られるSoC設計にますます似てきています(図表16参照)。

図表16
デバイスタイプ別エッジデータセンター向けAIプロセッサ市場、2023～2028年(単位:10億米ドル)



出所:Omdia 2024年第3四半期

次にやるべきことは？：障壁の克服とTransformer後のAIの再考

今後数年間において重要な問いは、カスタムシリコンを採用できる企業が増えるかどうか、そして投資の最低基準がどうなるかです。約12カ月でテープアウトするまでの開発費(NRE)として5,000万米ドルという数字が一般的に言われていますが、真の課題は、チップのサポートに必要なソフトウェアツールの開発と保守にあるかもしれません。これがAIチップのスタートアップ企業にとって大きな障壁となっていることは確かです。

それでもう一つ差し迫った問題が、「Transformerモデルの後には何が来るか」ということです。Striped HyenaやMambaなど現在いくつかの代替案が検討されており、これはTransformerのメモリを大量に消費する自己注意機構を、かつてのリカレント・ニューラル・ネットワークから借用した概念であるステートマシンに置き換えることを目指しています。さらに、3値(1.5ビット)の数値表現に移行することで、Transformerの使用を拡張することができます。もしブレークスルーによってAIが再びコンピュートバウンドになったとしても、主要GPUへの広範な世界的投資はそれほど影響を与えないかもしれません、その大規模な並列処理能力は依然として価値を提供することでしょう。競争力を維持できるよう、業界関係者は進行中のAI研究に深く関与し続け、将来の技術シフトに備える必要があります。



文末脚注

1. PwC 2024: Electric Vehicle Sales Review Q2-2024(英語)
2. Omdia analysis and research Q3 2024(英語)
3. AnandTech March 2024:NVIDIA Blackwell Architecture and B200/B100 Accelerators Announced: GoingBigger With Smaller Data(英語)
4. Taipeitimes September 2024:Samsung, TSMC collaborating in HBM solutions(英語)
5. Trendforce August 2024(英語)
6. TrendForce June 2024:SK Hynix's 5-layer 3D DRAM Yield Reportedly Hits 56.1%(英語)
7. SK Hynix press release May 2019; SOLIDIGM press release July 2023(英語)
8. PwC 2023:Bridging the talent gap(英語)
9. 欧州委員会(2022年)。欧州半導体法、2023年8月16日検索
10. Tesla FSDチップ:企業ウェブサイトより
11. William Fedus, Barret Zoph, Noam Shazeer:Switch Transformers:Scaling to Trillion Parameter Models with Simple and Efficient Sparsity(英語)
12. Center for Security and Emerging Technology 2023:Meta's Language Models Leak(英語)
13. TrendForce April 2024:Marvell's AI Business Reportedly Accelerates, Potentially Benefiting TSMC(英語)
14. Forbes:Qualcomm On-Device AI Powers Future Products From Phones To PCs March 2024(英語)

筆者について

Glenn Burnは、PwC半導体部門およびStrategy&韓国のグローバルリーダーです。技術、メディア、通信、プラットフォーム分野の業界アドバイザーであり、Strategy&/PwCとSamsung Groupとのグローバル関係を統括しています。クライアントの成長戦略、投資戦略、価値創造、デジタル戦略をサポートしています。

PwC Semiconductor Center of Excellence (CoE) は、PwCグローバルネットワーク内の半導体専門家によるグローバル専門チームであり、韓国、ドイツ、米国、日本などの主要地域を担当しています。PwC Semiconductor CoEは、半導体エコシステム全体にわたりクライアントが直面する課題に対処する革新的なソリューションの提供に向け取り組んでいます。

Omdia Semiconductor Researchには、貴重なご支援と素晴らしい洞察力で本レポートに貢献いただきましたことに深く感謝申し上げます。

お問い合わせ先

PwC Japanグループ
<https://www.pwc.com/jp/ja/contact.html>



www.pwc.com/jp

PwC Japanグループは、日本におけるPwCグローバルネットワークのメンバーファームおよびそれらの関連会社(PwC Japan有限責任監査法人、PwCコンサルティング合同会社、PwCアドバイザリー合同会社、PwC税理士法人、PwC弁護士法人を含む)の総称です。各法人は独立した別法人として事業を行っています。複雑化・多様化する企業の経営課題に対し、PwC Japanグループでは、監査およびブローダーアシュアランスサービス、コンサルティング、ディールアドバイザリー、税務、そして法務における卓越した専門性を結集し、それらを有機的に協働させる体制を整えています。また、公認会計士、税理士、弁護士、その他専門スタッフ約12,700人を擁するプロフェッショナル・サービス・ネットワークとして、クライアントニーズにより的確に対応したサービスの提供に努めています。PwCは、社会における信頼を構築し、重要な課題を解決することをPurpose(存在意義)としています。私たちは、世界149カ国に及ぶグローバルネットワークに370,000人以上のスタッフを擁し、高品質な監査、税務、アドバイザリーサービスを提供しています。詳細は www.pwc.com をご覧ください。

本報告書は、PwCメンバーファームが2024年11月に発行した『State of the semiconductor industry』を翻訳したものです。翻訳には正確を期しておりますが、英語版と解釈の相違がある場合は、英語版に依拠してください。

オリジナル(英語版)はこちらからダウンロードできます。<https://www.pwc.com/gx/en/industries/technology/state-of-the-semicon-industry.html>
日本語版発刊年月: 2025年5月 管理番号: I202412-09

© 2025 PwC. All rights reserved. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details. This content is for general information purposes only, and should not be used as a substitute for consultation with professional advisors.