

2026年の半導体・SPE市場の見通し

■半導体市場見通し

半導体市場はCY25にYoY+26%、7,917億ドルに達し、そのけん引役はAIに関連するGPU及びDRAM、NAND Flashであった。いずれも大きな要素は数量の増加以上に価格上昇が市場全体を押し上げた結果となった。この傾向はCY26も継続すると弊社では予想する。GPUに加えてCustom ASICの需要増、かつメモリに関しては弊社調査によればCY26中は少なくとも価格上昇は継続すると見る。それにより全体の半導体市場はCY26でYoY42%増となり1.1兆ドルの市場となると予想する。従来1兆ドルの台はCY30の時間軸が想定されていたが、AIの加速、それに伴うメモリの価格上昇により従来想定に対して早期に1兆ドルを突破する見込みである。

図1 半導体市場見通し



出所：WSTS実績を基にSMBC日興証券作成、予想

弊社ではAIへの需要の強さが継続することをベースシナリオとする。AIに関して、現時点では過去のITバブルとは異なり、少なくともまだ数年は需要の強さが継続すると想定する。2023年に半導体産業に出現したAI。NVIDIAのGPUについては生産の初期段階における生産不具合等があったものの、着実なロードマップを描いていると言えよう。

NVIDIAのGPUのみならず、それに伴うHBMの進化も著しい。2020年にリリースされたA100においてはHBM2EからGPUの世代交代と同時にHBM3、HBM3E(8hi/12hi)が量産適用されており、CY26にはHBM4の量産が開始されていると推察する。またGPU/Custom ASICとHBMを支えるCoWoSに代表されるパッケージの進化も同時に進展

SMBC日興証券株式会社
株式調査部 シニアアナリスト
はなや 武



している。

NVIDIAのA100からHopper世代、現行のBlackwell世代と、GPU自体の性能は飛躍的に向上してきた。またNVIDIAは半導体の進化のみならず、そのエコシステムの構築により、競合に対して高いハードルを設定している。GPUはGraphics Processing Unitであり、当初はあくまでも画像や動画をはじめとする映像を処理し、表示する役割を担うものであった。CPU(Central Processing Unit)はPCの頭脳であり、一つ一つの作業を高速で処理することを得意とするのに対して、GPUは複数の作業を同時並行で処理することを得意としている。映像を処理することは大量のデータを一括処理できるGPUの得意分野であった。その利点を活用し、大量データを同時に処理・分析することにより、AIという世界の主役となったのである。GPUを核として、エコシステムを支えるソフトウェア環境以外の進化も同時並行的に進められている。それが後述するHBMの進化及びパッケージの進化となる。

米国ハイパースケイラー(Microsoft、Amazon.com、Google、Meta)を中心にAIアクセラレータとしてのCustom ASICの開発競争が激しい。各社は主にBroadcom、Marvell、AlchipなどのFabless企業のサポートを受け、自社のCustom ASICの開発を行っている。Training AIにおいてはNVIDIAのGPUを用い、Inference AIにおいては自社内製のCustom ASICを適宜使用する考えがあると推察する。

汎用性が高く、並列処理に長けたGPUはTrainingにおいては現状の最適解であろう。しかし、Inferenceにおいてはアプリケーション、ソフトウェアとの互換性、効率性の観点から、アプリケーション開発を行っているハイパースケイラーの内製品の方が優れてくる部分も出て来よう。また、ハイパースケイラーはAIにおけるプラットフォームの座を狙う中で、ソフトウェアとハードウェアの内製化やブラックボックス化がプラットフォームにとっては定石の戦略であろう。

HBMの誕生は既に10年以上前に遡り、初代のHBMは2014年頃には開発が完了していた。一方、これほど広帯域で大容量、かつ高額なDRAMを必要とするデバイス市場は長らく誕生しなかった。それがAI市場の出現により一躍脚光を浴びた。量産適用されたのはNVIDIAのA100でHBM2Eが採用されたことからであろう。その後、HBM3、現行のHBM3Eと着実に進化し、DRAM市場のシェア変動の主因となった。

HBMにおいても今後進化を続けるであろう。2030年までの主な技術的な変化のポイントは(1)製造技術の変化、(2)ベースダイ(Logic Die)の進化と考える。

(1)製造方法については株式市場でも多くの議論が既になされてきた。基本的な製造方法の違いとしては、現行の主力モデルであるHBM3EにおいてSK Hynixはモールド装置を多用し、DRAM間のGap fillにモールド樹脂を流し込むAdvanced MR MUF(Mass Reflow Molded Under Fill)を採用している。一方、Samsung Elec及びMicronについては絶縁接着フィルムのNCFとTC Bonderを用いたTC-NCFを採用している。ただし、当初計画よりは遅れが生じている模様であるが、製造方法の変更が模索されている。Hybrid Bondingの採用である。現状 μ バンプによりDRAMダイの接続を行っているが、 μ バンプを排除し、Hybrid Bondingで直接結合する。

(2)ベースダイはHBMの最下層に置かれるHBMを制御するLogic半導体である。このベースダイについては従来、メモリメーカーがDRAMプロセスで内製していた自社画一のLogic半導体を搭載していた。一方、SK HynixはHBM4からTSMCへのベースダイの外部委託を開始する。それにより、従来メモリメーカーのレガシーNodeで作られていたベースダイの高性能化、省電力化が図られる。Micronも今後2027年頃にはTSMCへ外部委託を行う模様である。一方、Samsung Elec.は自社にLSI/Foundry事業を有するため、あくまでも自社完結の意向であり、陣営が分かれていく見通しである。

また、その次の世代のHBM4Eにおいては、NVIDIAやAMDなどGPU、AIアクセラレータメーカーである顧客に合わせてカスタムベースダイを導入する計画である。前述の通り、現状のNVIDIA一強の状況から、推論AIにおいてはハイパースケーラーも含めて多様化する方向にありそう。その方向性の中でHBM自体もカスタム化することで、顧客ごとの要求に対応しやすくなり、また電力効率等にも寄与することが期待される。

先端半導体は1971年に公表された世界で初めての商用マイクロプロセッサであるIntelの4004から過去50年強、Moore's lawに沿って微細化が中心的な技術進化であった。

それがオングストローム時代への突入を間近にし、技術的な限界が再び議論されている。またAIの登場によりMoore's lawによるトランジスタ数の増加だけではAI業界が求めるコンピューティングパワーを満たせない状態となってきた。

この状況下の技術進化としてMore than Mooreと呼ばれるパッケージ側でのコンピューティングパワーの進化論が活況を呈している。現状のAIサーバー向けにおいてはTSMCのCoWoS技術が標準となっているが、現状CoPoS(Chip-on-Panel-on-Substrate)、CoWoP(Chip-on-Wafer-on-Platform PCB)、EMIB-T、ガラスコアなど様々な研究開発が行われている。

■NAND Flashでも需給のタイト化定着、価格上昇が継続すると予想

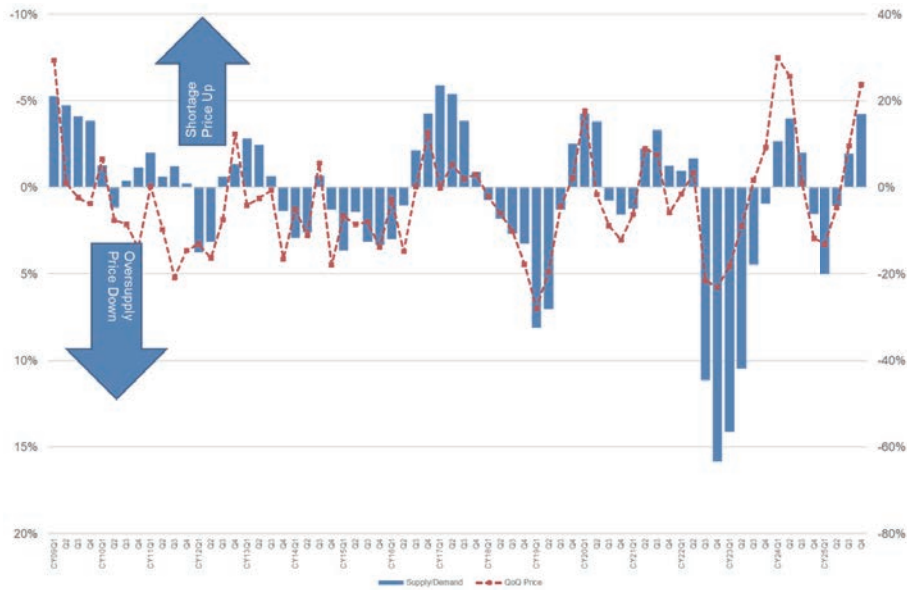
現状のNANDサイクルの認識を示す。CY25年初はNAND各社の早期の稼働低減、供給減により、サイクルが通常に比べて短期的かつ早期に進んだという認識の一方で、足元ではサイクルの進み具合は鈍化している。過去に生じていた設備投資の増加がみられず、供給増の見通しが持ちにくい。それゆえに、価格上昇、NANDメーカーの業績拡大、好調期が継続、長期化すると考える。

過去のNAND Flash市場の需給バランスの状況と価格推移を比較すると、基本的には価格動向と需給動向は同様のトレンドを描いてきている。また市場の拡大、好調局面、言い換えれば価格が上昇する局面は4~6四半期で終わることが常であった。

その主因は価格上昇、需給ひっ迫局面で、大手メーカーが能力増強投資、投資競争を行ってきたことと考える。他社に先んじる形で設備投資を意思決定し、実施できれば市況の良さを一層享受でき、またシェアアップを図れるという考えが背景にありそう。その点では過去においてはSamsung Elec.がその傾向が強かったという認識である。

一方で現在は設備投資に動く可能性は低いと考える。Samsung Elec.の状況はHBMにおいて後塵を拝し、巻き返して全力を挙げ、それが実を結びつつあるようにみえる。またクリーンルームスペースの制約もあり、NAND Flashへの投資が加速する蓋然性は現時点では非常に低い。また弊社の推察では、Samsung Elec.のNAND事業は2025年7~9月期にようやく損益分岐点を越えた程度であり、DRAMの収益性を考えればNAND Flashへの投資の意思決定は難しいであろう。wafer inputベースでは能力を減少させながらマイグレーション投資を行い、供給増をする見通しである。しかし、大規模な投資にはピオンテック5の建設完了を待

図2 NAND Flash 需給と価格



出所：Gartner 及び WSTS よりSMBC日興証券作成

出典：Gartner, Forecast: NAND Flash Market Statistics, Supply and Demand, Worldwide, 2023-2029, 4Q25 Update, Joseph Unsworth アナリスト他、2025年12月17日

つ必要があると考える。

スマホ /PC 向けについては NAND Flash の価格の上昇、またそれ以外にも DRAM、AP の価格上昇が見込まれる中で今後 BOM コストの上昇が大きくなると想定される。その場合、想定される NAND 市場のシナリオは主に3つであると考え。① NAND への価格圧力の強まり、② NAND 以外への価格圧力の強まりで NAND への価格は市場動向に左右される、③ NAND の台当たり搭載容量の増加ペースが減速する。①についてはスマホ /PC 向けの価格上昇のモメンタムはその他のアプリケーションに比べて緩慢になる可能性はあるが、生産能力が大きく増えない中でまたその他のアプリの需要が強く、値上げ余力も大きい中で、スマホ /PC においても価格が大幅に下落するとは考えにくい。そうすると②と③の組み合わせの可能性が高まっている。

AI 向けの NAND 市場 GPU を待たせるな

AI ストレージにとって重要なことは非常に高単価な GPU のアイドル状態を極力短くすることであろう。Training においては、大規模なデータを収容するために中断することなく拡張できるストレージが必要である。また、特に高いランダム読み取りパフォーマンスが必要になる。

一方で、Inference においては RAG (Retrieval-Augmented Generation) の活用がポイントになりそうである。RAG とは生成 AI の活用、Inference 段階において特定のコミュニティ (各企業など) が生成 AI を導入する際に、そのコミュニティにより一層適した内容にするための技術である。企

業を例に取れば各企業、部署、部門ごとに AI をより有効に活用するために、最新データやその企業、部署で使われている専門用語などのデータを取り込むことでより負荷を低減し、ハルシネーションを防ぐなど精度と信頼性の向上に資する。従来の ChatGPT に代表される汎用的な LLM (大規模言語モデル) により専門性が高く、特定タスク処理のためのデータベースを組み合わせたものになる。

その場合、AI サーバーは多様化し、サーバー台数としての伸びが期待できる。その際には eSSD の需要の拡大をけん引する可能性が高いと考える。また、単に台数の成長だけでなく、GPU 等 AI アクセラレータの稼働率を最大限に発揮させるためレイテンシーの低減は必須事項であり、消費電力の問題は切っても切り離せない。その点で SSD の優位性を出しやすく、また各社の SSD 開発状況を確認してもここ数年の読み出し、書き込み性能の向上は加速していると考えられ、今後の主戦場である eSSD 市場への取り組みが活発化していると言えよう。

SPE 市場見通し

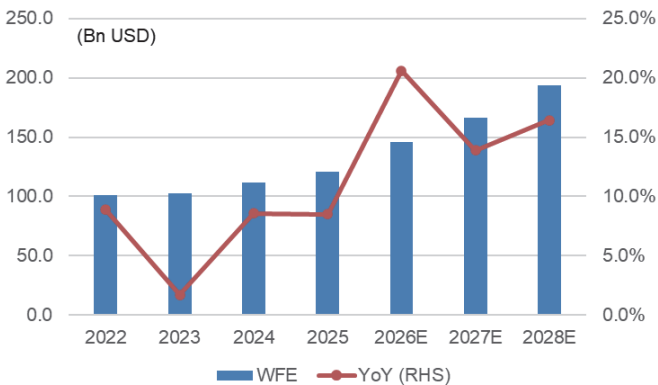
前工程装置はこれまで AI に近い存在ではあるものの、直接的な恩恵が後工程及びテスト工程に比べて希薄と認識されてきた。つまりは AI 関連需要が拡大してきた初期段階においては業界では GPU/HBM/CoWoS の数量の増加に対して既存の前工程で十分に賄えてしまった一方で、後工程の変化及びテスト工程の拡大により特定のプロセスのみに集中的に装置需要が創出された。一方現在は AI による物量の増加、市場のひっ迫、数の不足が生じてきている。

AI の需要の強さから前工程投資が加速する局面であると考えている。

WFE を考える際の主要な構成要素である Logic/Foundry、DRAM、NAND 向けはそれぞれ拡大を続けると予想する。タイミングには差はあるものの、全般的な投資の拡大が今後3年継続するものと考えられる。

前述の通り、半導体市場が早期に1兆ドルを突破する中、WFE に関してはCY26は Capital Intensity が下落する見込みである。CY23においては Capital Intensity は20% に到達、WFE についても拡大を続けているが、前述の通り、ここ数年の AI 市場の勃興による半導体市場の拡大を支えているのは、数量増というよりは価格上昇であるため、これまで以上の物量を必要とする状況にはCY25前半までは至らなかった。AI によって打ち消されていたが、民生品向け及び汎用的な半導体に関してはコロナ後の大幅な在庫調整が長引いていたと考えられる。ただし、ここ半年において閾値を超えてきた認識で、必要数量の増加、需給のひっ迫が生じ、それによりメモリ価格は早期に反応し、急上昇した。一方で WFE はそう簡単には増やすことができない。装置に関しては基本的には POR 取得時点で価格が固定されるため、メモリのような価格上昇には追従できない。

図3 WFE 見通し



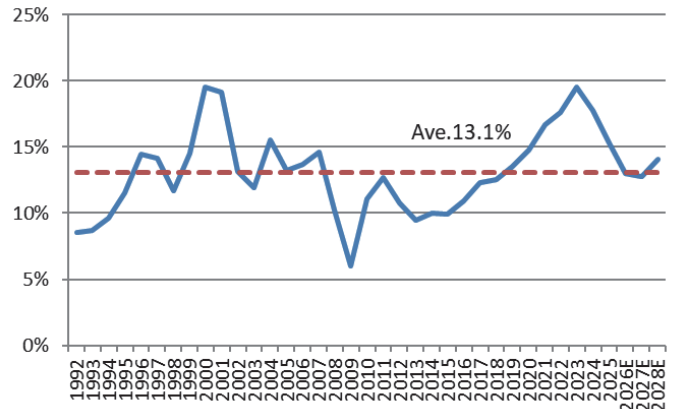
出所：Gartner を基に SMBC 日興証券作成、予想

出典：Gartner, Forecast: Semiconductor Capital Spending, Wafer Fab Equipment and Capacity, Worldwide, 4Q25 Update

ただし、半導体メーカーの設備投資の意欲が高まることで、今後3年 WFE は拡大を続けると想定する。Capital Intensity は CY25 で 15%、CY26 には 13% に低下、CY28 に 14% 程度まで若干の回復を見込む。

まず DRAM については、HBM では Samsung Elec. の巻き返しで投資の前倒しを試みている。また、これまでの HBM のみの投資から Conventional DRAM のタイト化に伴

図4 Capital Intensity



出所：WSTS、Gartner を基に SMBC 日興証券作成

う価格の急上昇、Conventional DRAM の収益性は HBM を既に超えている状況などもあり、設備投資見通しは強さが今後3年間は継続すると想定する。新棟の貢献として WFE が CY27 にどの程度入り切るのははまだ流動的な部分はあるが、CY27/28 にかけて強さが継続すると予想する。

一方で、NAND Flash 向け設備投資については依然として段階的にしか生じないと想定する。CY27 後半もしくは CY28 には DRAM メーカーにもクリーンルームに余裕が生じるため、NAND への投資が開始される可能性がある。それに伴い、CY28 には NAND 投資が加速すると予想する。

Logic/Foundry の投資に関して最も注目されていた TSMC の CY26 の設備投資見通しは、N3 のひっ迫により想定以上に大きく上振れとなった。定性的には今後3年間の投資の強さが言及されており、TSMC の売上成長は AI に支えられている限り高水準の設備投資が継続しよう。Samsung Elec. の米国テ일러工場については、Tesla 及び Qualcomm の受注により拡大を急いでいるようである。また、Intel に関しては 14A の開発量産成功について未だ流動的な部分があり、順調に進めばアップサイドが生じよう。

特記事項：ガートナーに帰属するすべての記述は、ガートナーの顧客向けに発行された配信購読サービスの一部として発行されたデータ、リサーチ・オピニオン、または見解に関する SMBC 日興証券による解釈であり、ガートナーによるレビューは行われておりません。ガートナーの発行物は、その発行時点における見解であり、本誌発行時点のものではありません。ガートナーの発行物で述べられた意見は、事実を表現したのではなく、事前の予告なしに変更されることがあります。

(事務局注：2026年3月31日現在の内容です。)