

転換点を迎える半導体メモリ —生成AI技術の成長を支えるその技術とバ リューチェーン—



三井物産戦略研究所
技術・イノベーション情報部インダストリーイノベーション室
小川 玲奈

Summary

- 半導体の機能を人間の作業にたとえると、ロジックは人間、メモリは机、ストレージは本棚に相当する。代表的なメモリであるDRAMの中で市場成長をけん引するのは、生成AI向けのHigh Bandwidth Memory (HBM) というメモリで、2029年のHBMの予想市場規模は440億ドルである。
- メモリはAI開発のボトルネックとみなされている。次世代HBMでは、製造の一部をロジックのファウンドリーが受託する見通しである。
- 次世代HBMから始まるバリューチェーンの変化により、要素技術の供給先の広がりや、新たなビジネスモデルの誕生が期待される。

1. 今知っておきたい、メモリの基礎知識

2030年に1兆ドルに達するといわれる半導体市場の中で、メモリは、ロジック（CPU、GPUなど）に次ぐ規模を有し、かつ高い成長が予想されている。代表的なメモリであるDRAM（Dynamic Random Access Memory：ディーラム）の市場規模は、2023年960億ドル、2029年には1,360億ドルになると予想されている¹。その成長をけん引するのは生成AI向けの High Bandwidth Memory（HBM）で、2023年54億ドルから2029年には440億ドル²と、6年間で8倍以上となる見通しである。本章では、メモリについて押さえておきたい基礎知識を紹介する。

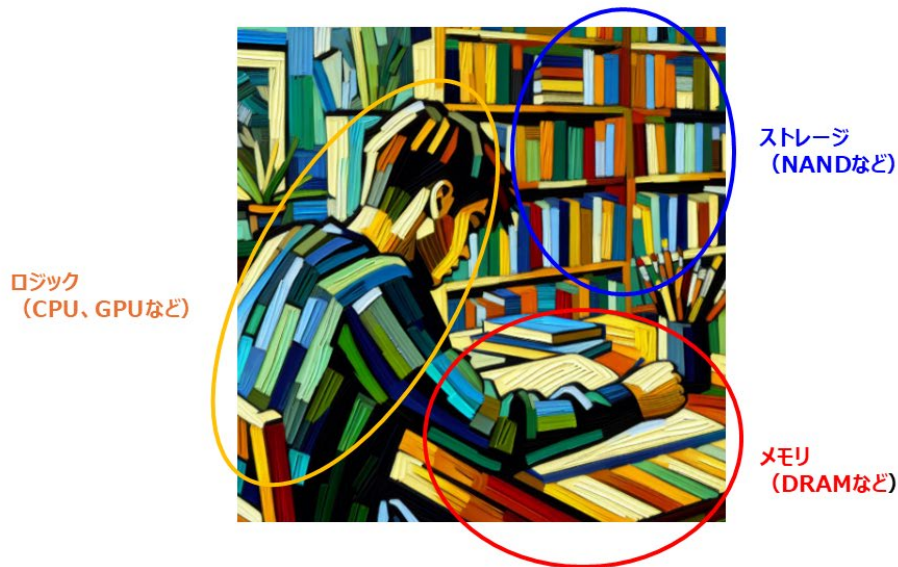
1-1. メモリとは

メモリとは、パソコンやデータセンターなどの中で演算を行うロジックにデータを提供するために、一時的にデータを保存する半導体記憶装置であり、RAM（Random Access Memory：ラム）ともいう。半導体の機能を人間の作業にたとえると、メモリは作業を行うためにデータを取り出して並べる机、ロジックは作業をする人間、ストレージはデータを長期間保存するための本棚に相当する（図表1）。

¹ 2023年の実績値は[世界半導体市場統計](#)（WSTS）、DRAMとHBMの市場予測は[Yole Intelligence](#)の予測値を引用。

² 出所：EE Times Japan, [世界メモリ市場、2025年に過去最高の2200億ドル超に 生成AIがけん引](#)

図表1：半導体デバイスの役割を人の作業にたとえたイメージ図

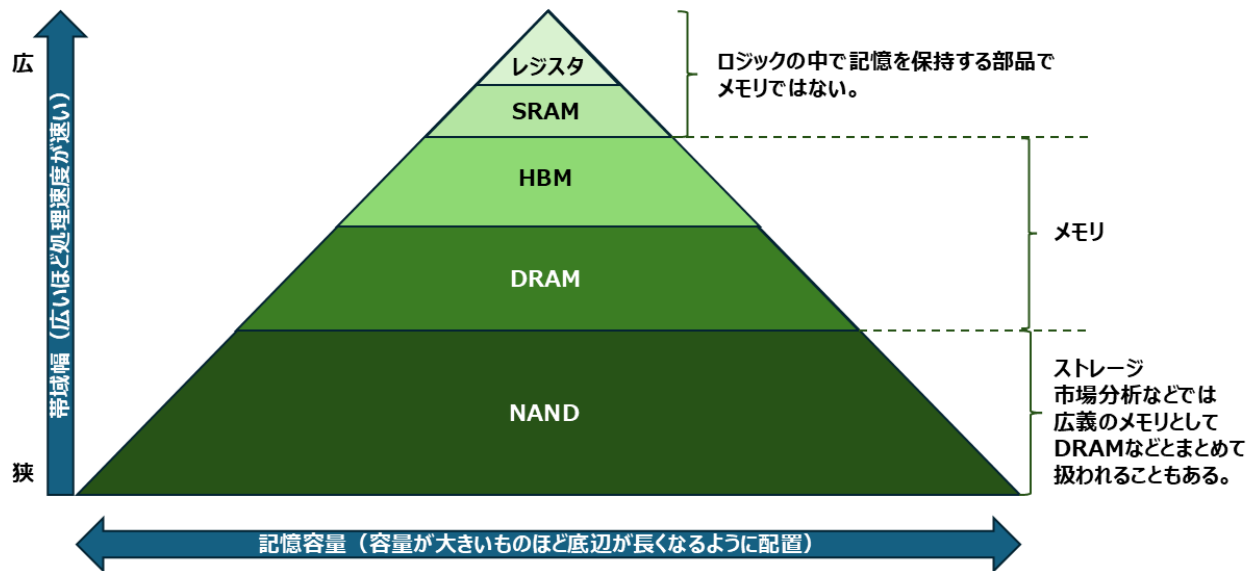


出所：Copilot にて三井物産戦略研究所作成

1-2. 各種記憶装置の種類と性能

記憶装置は、記憶容量と処理速度の速さによりさまざまな種類がある（図表2）。処理速度は帯域幅の広さで表現され、広いほど速い。処理速度と記憶容量にはトレードオフの関係があるため、用途によって最適な記憶容量と帯域幅を持つメモリが選択される。パソコン、データセンター、スマートフォンなどでは一般的にRAMとしてDRAMが用いられている。市場分析では、NAND（Not AND：ナンド）も広義のメモリとして扱われることがあるが、NANDは処理速度が遅いため、一般にストレージとして利用されている。また、記憶装置として、ロジックで演算を行うために一時的に記憶を保持する役割を持つレジスタやSRAM（Static Random Access Memory：エスラム）といった記憶装置も存在するが、これらはロジックの内部にあるため、メモリとは分けて考えられるのが一般的である。

図表2：記憶装置の種類と性能



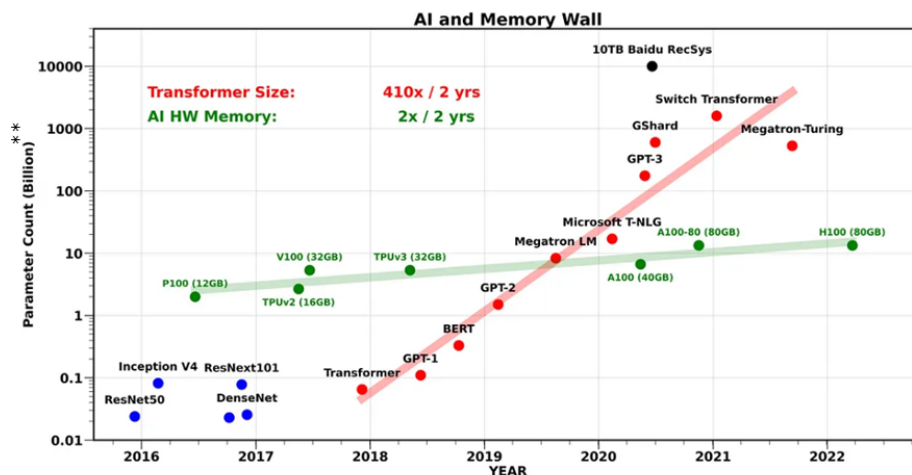
出所：各所資料から三井物産戦略研究所作成

1-3. 生成AI向けでメモリの成長をけん引するHBM

近年注目度が高まっている生成AI のデータ処理では、メモリの処理速度や、ロジックとメモリの間の通信速度が性能向上のボトルネックとなる「メモリの壁」が指摘されている³。その主要因は、メモリの帯域幅の改善が、生成AI の開発に必要なデータ処理速度に相当する「パラメータ」の増大と比較して桁違いに遅いことにある。2019年半ば、1つのAI半導体に搭載されているメモリの処理能力は、AIモデルで必要とされる処理能力に追い抜かれ、その後引き離され続けている（図表3）。

³ Medium, “AI and Memory Wall” (Amir Gholami著, <https://medium.com/riselab/ai-and-memory-wall-2cb4265cb0b8>)

図表3：メモリのデータ処理能力とトランスフォーマー*で必要とされるデータ量の比較



*トランスフォーマーとは、生成AI などにおいて、入力された情報のコンテキストを学習し、応答に変換するモデルのこと。

**縦軸の Parameter Count は、トランスフォーマー実現のために必要な情報量に相当する。これに、1つのAI半導体に搭載されているメモリで処理できるデータの量（緑）を、トランスフォーマーで必要とされるデータ量（赤）と重ねてプロットしている。

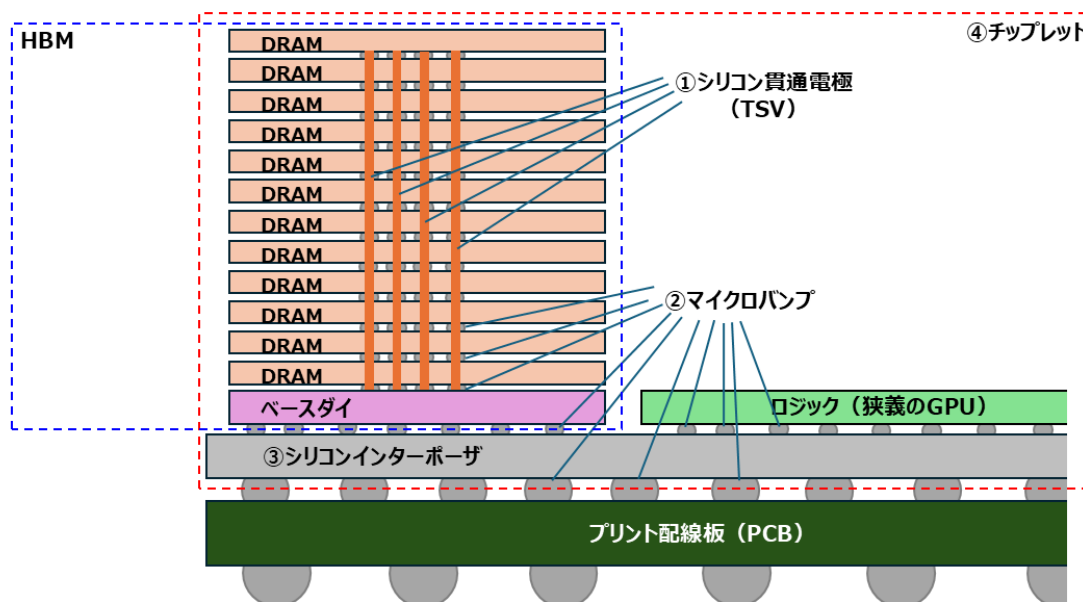
出所：Medium, “AI and Memory Wall”

(Amir Gholami著, <https://medium.com/riselab/ai-and-memory-wall-2cb4265cb0b8>)

HBMは、面積あたりの記憶容量を高めるために、DRAMを、制御や通信を担うベースダイという基板の上に積層したメモリである。（米）NVIDIAのGPU⁴ではさらに、ロジックとHBMを、シリコンインターポーザと呼ばれる基板を介して接続する（図表4）ことで、ロジックとメモリの間の通信時間を短縮させている。HBMの実現や進化には、製造プロセス、素材、設計といった幅広い分野にまたがる要素技術の進歩と製造現場への実装が欠かせない。具体的には、①シリコン貫通電極（TSV）、②マイクロバンプ、③シリコンインターポーザ、④チップレット、⑤電源電圧の品質安定化、⑥熱マネジメントといった技術の改良がHBMの性能に直結している（図表4、5）

⁴ Graphic Processing Unitの略。画像処理を高速で行うロジックだったが、生成AIを含む機械学習の演算の高速化も可能であることからAI半導体として広く用いられるようになった。半導体製品としては、HBMとロジック（狭義のGPU）がシリコンインターポーザで接続され、プリント配線板上に載った状態で販売されているため、図表4全体をGPUと呼ぶことが多い。

図表4：HBMとNVIDIAのGPUの概要



⑤電源電圧の品質安定化および⑥熱マネジメントはHBM、GPUの全体に関係する要素技術のため、図中の記載なし。

出所：各所資料から三井物産戦略研究所作成

図表5：HBMに必要とされる要素技術の概要

要素技術名	技術の概要	技術的課題・開発動向
①シリコン貫通電極（TSV）	シリコン製のチップに貫通孔を開け、導電体を充填することで形成される電極。これにより、上下に積層したチップをワイヤーよりも短い距離で接続できる。	微細化により、貫通孔と電極の形成（レジスト、エッチング、めっき、平坦化など）の難易度が上がっている。
②マイクロバンブ	半導体デバイスと回路基板の電極同士を電気的に接続するために形成するはんだの突起。	バンブによって生じる接合距離が集積密度や電力損失低下の律速となってしまうため、バンブを使わずに接合するハイブリッドボンディングの適用が検討されている。
③シリコンインターポーザ	シリコンの基板に配線を形成したもの。この上に複数のチップが実装される。	接続部分のみにインターポーザを使うシリコンブリッジや、反り問題解消のためのガラスインターポーザへの置き換え検討などがある。
④チップレット*	大きく、複雑な集積回路を形成するために、他のチップ（チップレット）と情報のやりとりを行うことを前提に設計された半導体集積回路ブロックのこと。	AI半導体向けでは大規模な集積回路を開発するためのチップ間接続の標準化が進み、設計ツールも成熟してきている。
⑤電源電圧の品質安定化	HBMによる高速大容量のデータ送信では大電流が必要。その安定動作のためには、大電流でも安定した電圧を供給できる電源が必要とされる。	シリコンよりも高耐圧、低損失な窒化ガリウム（GaN）などが電源ユニット用のパワー半導体として検討されている。
⑥熱マネジメント	HBMが作動できる温度を超えることのないよう、デバイス構造、電源回路、監視システムによるデバイス内外の熱の制御を行うこと。	HBMそのものが大電流で動作するため発熱が大きく、積層されているため、単層の半導体よりも放熱が難しい。

*チップレットについては Biz Tech フォーカス 2025 チップレット

https://www.mitsui.com/mgssi/ja/report/detail/_icsFiles/afieldfile/2025/02/07/2501btf_ogawa_ishiguro.pdfにて詳細を論じている。

出所：各所資料から三井物産戦略研究所作成

2. DRAMおよびHBMの構造とバリューチェーンの変化

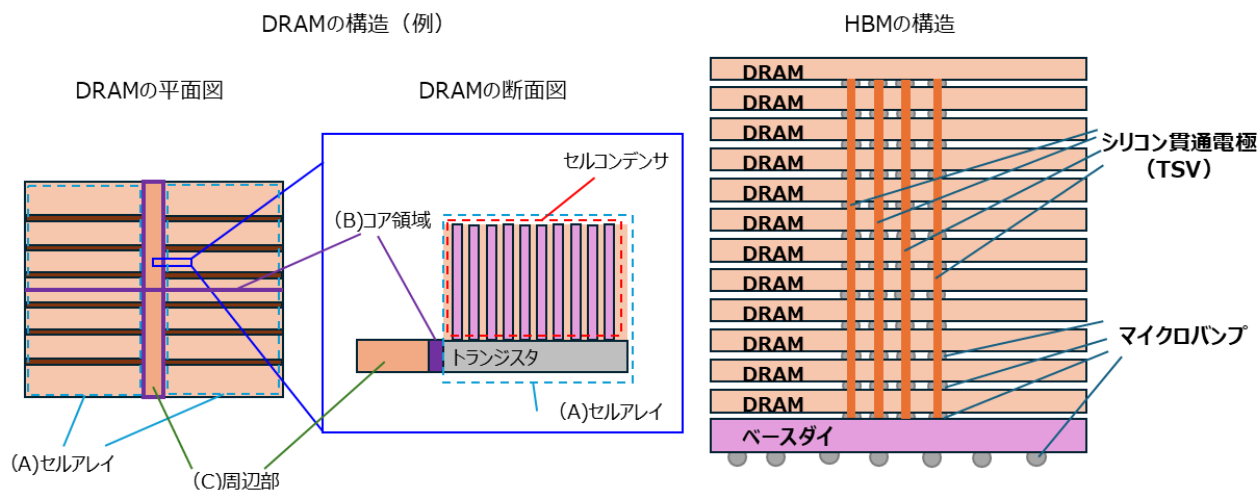
本章ではDRAMおよびHBMの構造と、HBMの開発ロードマップに見るバリューチェーンの変化について説明する。

2-1. DRAMおよびHBMの構造

DRAMのチップは、データを記憶する（A）セルアレイ、データへのアクセスを司る（B）コア領域、チップ外への通信を行うための（C）周辺部からなる（図表6左）。セルアレイは、セルコンデンサ（capacitor）とトランジスタからなり、チップ面積の大半を占める。DRAMの微細化は、セルコンデンサとトランジスタの微細化である点が、トランジスタのみの微細化に注力できるロジックとは大きく異なる。

HBMは、ベースダイの上にDRAMを積層し、TSVとマイクロバンプで接続した構造（図表6右）とすることで、DRAMの処理速度を損ねずに、チップ面積あたりの容量を大きく向上させている。これにより、HBMは生成AI向けのメモリに求められる大容量と高速処理の両立を実現できるメモリとして採用された。

図表6：DRAMおよびHBMの構造



出所：各所資料から三井物産戦略研究所作成

2-2. HBMの開発ロードマップに見る、バリューチェーンの変化

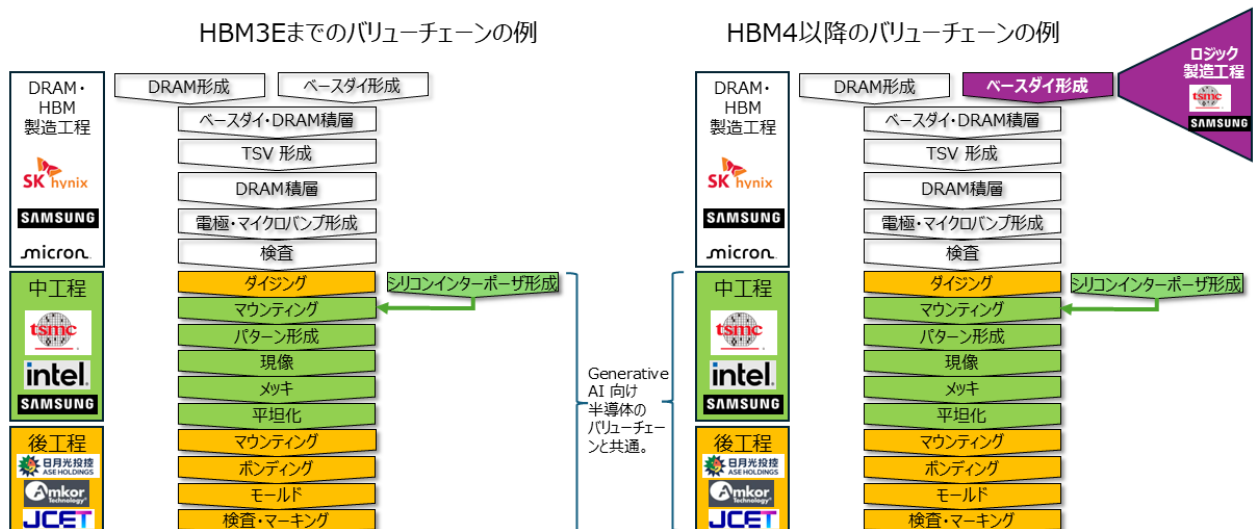
HBMは業界としてロードマップを作り、それに従って開発を進めている（図表7）。2016年の第1世代のリリースから2025年6月の時点で3つの世代を経ており、2025年6月時点で最新の世代はHBM3Eである。HBMを開発・販売しているのは（韓）SK Hynix、（韓）Samsung Electronics、（米）Micron Technology の3社で、いずれも大手DRAMメーカーである。2025年末から2026年にかけてHBM4のリリースが予定されている。これまでのHBMは、DRAMのプロセスのみで製造されてきたが、HBM4からは、ベースダイをロジックのプロセスで製造する（図表8）。このプロセス変更により、帯域幅がHBM3Eの2倍となる。

図表7：HBMの開発ロードマップ

期間（西暦）	2016-2018	2018-2020	2020-2022	2022-2024	2024-2026	2026-2028	2028-2030
製品	HBM	HBM2	HBM2E	HBM3	HBM3E	HBM4	HBM4E
帯域幅	256GB/s	307GB/s	461GB/s	819GB/s	1.0TB/s	2.0TB/s	2.5TB/s
ダイ1枚あたりのメモリ容量	8GB		16GB		24GB		32GB
積層数	4 or 8			8 or 12		12 or 16	
HBMの総容量	4 or 8 GB		8 or 16 GB	16 or 24 GB	24 or 36 GB	36 or 48 GB	48 or 64 GB
ベースダイの製造プロセス	DRAMプロセス					ロジックプロセス	

出所：ISSCC 2025 Forum 3.4: Memory Solutions for AI Era: High-Bandwidth Memories and Dense-Data Stage
から三井物産戦略研究所作成

図表8：HBMのバリューチェーンの変化



出所：各所資料から三井物産戦略研究所作成

HBM4のベースダイの製造は主に、最先端ロジック製造の最大手ファウンドリーである（台）TSMCが担う見通しである。ロジックの製造設備を持たないSK Hynix とMicron Technologyだけでなく、自社内にファウンドリーを持つ Samsung Electronics も、TSMCとの連携を模索しているとの観測がある。

3. さらに高速・低消費電力なメモリの開発動向

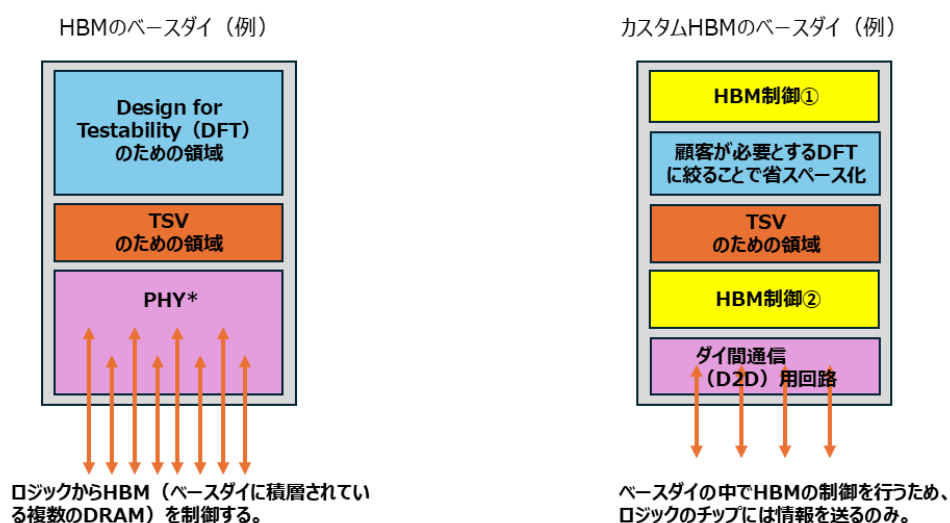
本章ではHBM4リリースの先に社会実装が予想される技術として、カスタムHBMとコンピュータインメモリについて解説する。

3-1. カスタムHBM

HBM4の製造をファウンドリーが受託することは、メモリ製造の水平分業の始まりを意味する。TSMCなどのファウンドリーは、同じ製品を大量に製造するメモリメーカーと異なり、ウエハごと、あるいは1枚のウエハの中に異なる設計のチップを製造するサービスを提供している。このサービスをベースダイに応用することで、HBMのカスタム化が可能になると考えられている。

従来のHBMでは、どの顧客に対しても同じチップで対応できるよう、積層されたDRAMが正常に動作していることを確認するための回路（DFT）がベースダイの多くを占めており、ベースダイの面積の制約から、HBMの制御はロジックで行っていた。これに対し、カスタムHBMでは、顧客にとって必要最低限なDFTに絞ることで、HBMを制御するための回路をベースダイに収める余地が生まれる。これにより顧客は、HBM制御回路をロジックで設計する必要がなくなる。さらに、従来はロジックからHBMを制御するために送られていたデータの通信も不要になり、HBMとロジックダイ間の通信は、演算に必要とするデータを送信するだけで済む（図表9）。以上より、カスタムHBMは、ロジックの設計負荷低減に加えて、送信にかかる電力の削減というメリットをもたらす。

図表9：HBMとカスタムHBMのベースダイ比較



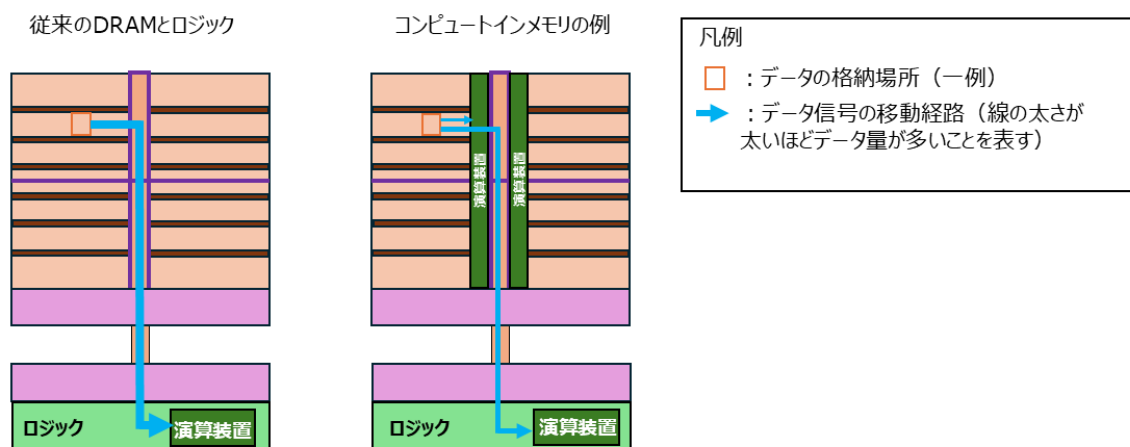
*PHY（ファイ）：情報を伝送するための回路。物理層ともいう。

出所：各所資料から三井物産戦略研究所作成

3-2. コンピュートインメモリ (Compute-in-Memory)

さらなる生成AI向けメモリの低消費電力化が可能な技術として、コンピュートインメモリの注目度も高まっている。半導体の消費電力は、信号の移動距離に比例する。従来のDRAMは、演算を行うためのデータをロジックに伝送するため、信号の移動距離が長く消費電力が大きい。これに対し、コンピュートインメモリでは、DRAMの中にロジックの機能の一部（演算装置）を入れ込むことで、DRAMからロジックへの伝送を減らし、消費電力の低減が可能となる（図表10）。DRAMベースのコンピュートインメモリ技術は、SK Hynix や Samsung Electronicsが「Processing-in-Memory (PIM)」という名称で開発・提案している。一方、TSMCなどDRAMメーカー以外の半導体メーカーでも、DRAMよりも消費電力の低い磁気抵抗メモリ（MRAM）などの記憶装置に演算装置を入れる検討が行われている。

図表10：従来のDRAMとコンピュートインメモリの違い



出所：各所資料から三井物産戦略研究所作成

4. まとめ～新たなビジネスの可能性

本稿では、メモリの基本知識からひもとき、生成AIの性能向上のボトルネックたるメモリに技術革新が求められていること、現在生成AI向けで採用されているHBMでは、2025年から2026年にかけてリリースされるHBM4から、ベースダイの製造にロジックのプロセスが用いられること、これによりHBMのバリューチェーンが変わることを述べた。

バリューチェーンの変化によって生まれる新たな商流に加え、新しいアイデアを持つメーカーと、異な

る技術・産業分野の人や技術をつなぐ機能の重要性も高まりそうだ。

当レポートに掲載されているあらゆる内容は無断転載・複製を禁じます。当レポートは信頼できると思われる情報ソースから入手した情報・データに基づき作成していますが、当社はその正確性、完全性、信頼性等を保証するものではありません。当レポートは執筆者の見解に基づき作成されたものであり、当社および三井物産グループの統一的な見解を示すものではありません。また、当レポートのご利用により、直接的あるいは間接的な不利益・損害が発生したとしても、当社および三井物産グループは一切責任を負いません。レポートに掲載された内容は予告なしに変更することがあります。