

先取り★マーケットレビュー

今回のテーマ

「GTC2024」から見える

AI半導体の現状と将来

2024年3月27日（水）

楽天証券経済研究所

チーフアナリスト

今中 能夫

- エヌビディアの年次テクノロジー・カンファレンス「GTC2024」がアメリカ・サンノゼで3月18～21日に開催された。
- ジェイソン・ファンCEOの基調講演では、新アーキテクチャー「Blackwell」による「B200」「GB200」「GB200NVL72」の詳細が発表された。
- 1個の「**Blackwell GPU**」に2080億トランジスタが搭載されている（「H100」は800億トランジスタ）。これを2個連結して一つのパッケージに収めたものが「**B200**」、エヌビディアの自社製CPU「Grace」と「B200」（「Blackwell GPU」2個）を連結して1つのパッケージにしたものが「**GB200**」、36個のGrace CPUと72個のBlackwell GPU（即ち36個の「GB200」）を接続したものが「**GB200NVL72**」。

- MoE（Mixture of Experts：特定のタスクに特化した複数のexpertを入力に対して切り替えることで性能を上げる機械学習の手法）1.5テラのGPTを90日間トレーニング（機械学習）する場合、8000個の「H100」で15MW（メガワット）の電力が必要になる。
- これが「GB200NVL72」の場合は、2000個の「Blackwell GPU」（28個の「GB200NVL72」？）で4MWの電力で済む。大幅な電力消費の低減が実現できる。
- 今の主力「H100」と比較すると「B200」の推論性能は15倍、AIトレーニング性能は3倍。また、ファンCEOの基調講演の中で示されたグラフによれば、「H200」と「GB200」を比較すると推論性能は最大30倍高くなる。

- Blackwell GPUはTSMCのN4P（4ナノの改良版ライン）で生産される予定。「B200」の出荷開始は2024年後半だが、本格出荷は2025年に入ってからか。
- 「B200」は192GBのHBM3e（DRAMの最新規格「DDR5」をベースにした特殊メモリの最新型）を装備する。「GB200」は384GBのHBM3eを装備。「H100」の80GB（HBM2e）、「H200」の141GB（HBM3e）と比べて、GPUメモリの容量が拡大する。
- 「B200」「GB200」のメモリ帯域幅は8TB/s（毎秒8テラバイト）であり、「H100」の2TB/s、「H200」の4.8TB/sに比べて大幅に拡大。これが「B200」の高速化実現の要因の一つになっている。

●CNBCとファンCEOとのインタビューによれば、「Blackwell GPU」の価格は30,000~40,000ドル。CNBCのウェブ記事によれば、アナリストによれば「H100」の価格は25,000~40,000ドル。

●生成AIとAIは何に使うのか

✓計算処理

✓シミュレーション（気象、素材、タンパク質、地質、流体、空力など）

✓化学（新素材の研究開発）、薬品バイオの研究開発

✓設計

✓半導体生産

✓その他企業活動全般

エヌビディアのAI用GPUロードマップ（2023年10月）

2021年	2023年	2024年		2025年	(接続するCPUの) アーキテクチャー	特徴・強化点 (推論とトレーニング (学習))	
A100	→ H100		GH200NVL	GB200NVL	GX200NVL	Arm	学習と推論
			↑	↑	↑		
			GH200	GB200	GX200	Arm	推論
			↑	↑	↑		
		→ H200	→ B100	→ X100	x86	学習と推論	
		↘					
		L40S	B40	X40	x86	企業向け、推論	
	Quantum (エヌビディアの高速ネットワーク機器群)	400G	→ 800G	→ 1600G		InfiniBand AIインフラストラクチャー	
	Spectrum-X (エヌビディアのイーサネット・プラットフォーム)	400G	→ 800G	→ 1600G		企業向けイーサネットとハイパースケールAIインフラストラクチャー	

出所：「NVIDIA Investor Presentation October 2023」26ページより楽天証券作成

エヌビディアのAI用GPUロードマップ（「GTC2024」基調講演を踏まえた楽天証券による修正後）

2021年	2023年	2024年	2024~2025年	(接続するCPUの) アーキテクチャー	特徴・強化点 (推論とトレーニング (学習))	
		GH200NVL	GB200NVL	Arm	学習と推論	
		↑	↑			
		GH200	GB200	Arm	推論	
		↑	↑			
A100	→	H100	→	B200	x86	学習と推論
		↘		↓		
		L40S		(B100)	x86	企業向け、推論
	Quantum (エヌビディアの高速ネットワーク機器群)	400G	→	800G		InfiniBand AIインフラストラクチャー
	Spectrum-X (エヌビディアのイーサネット・プラットフォーム)	400G	→	800G		企業向けイーサネットとハイパースケールAIインフラストラクチャー

出所：「NVIDIA Investor Presentation October 2023」26ページに楽天証券加筆

注：GTC2024のエヌビディアCEOによる基調講演では、「X100」に対するコメントがなかったため、2025年の「X100」系統の部分を省いた。

H100/H200性能比較

大規模言語モデルを使った推論速度（H100に対して）	
Llama2 70B	1.9倍
GPT-3 175B	1.6倍
Llama2 13B	1.4倍
ハイパフォーマンスコンピューティングの計算速度	
Dual x86 CPUと比較した時のH200の計算速度	110倍
A100を1とした時のパフォーマンス	
H100	1.7倍
H200	2.0倍
エネルギー効率（H100との比較）	
エネルギー消費	50%
TCO（Total Cost of Ownership、IT関連設備の総保有コスト）	50%

出所：エヌビディア資料より楽天証券作成

H100/B200性能比較（H100に対するB200の性能比較）

リアルタイム大規模言語モデル推論	15倍
AIトレーニング性能	3倍
高速データ処理	2倍

出所：エヌビディア資料より楽天証券作成

エヌビディアの業績

	2023年1月期	2024年1月期	2025年1月期 楽天証券予想 (前回)	2025年1月期 楽天証券予想 (今回)	2026年1月期 楽天証券予想 (前回)	2026年1月期 楽天証券予想 (今回)
売上高	26,974	60,922	110,000	110,000	170,000	170,000
前年比	0.2%	125.9%	80.6%	80.6%	54.5%	54.5%
営業利益	4,224	32,972	70,000	70,000	118,000	118,000
営業利益率	15.7%	54.1%	63.6%	63.6%	69.4%	69.4%
前年比	-57.9%	680.6%	112.3%	112.3%	68.6%	68.6%
当期純利益	4,368	29,760	58,800	58,800	98,800	98,800
前年比	-55.2%	581.3%	97.6%	97.6%	68.0%	68.0%
EPS	1.75	11.95	23.61	23.61	39.68	39.68
配当	0.16	0.16	0.16	0.16	0.16	0.16
PER	527.6	77.4	39.2	39.2	23.3	23.3

株価 925.61 ドル (2024年3月26日)

時価総額 2,282,554 百万ドル (2024年3月26日)

発行済株数 2,490 百万株 (完全希薄化後、Diluted)

発行済株数 2,466 百万株 (完全希薄化前、Basic)

単位：百万ドル、%、倍

出所：会社資料より楽天証券作成。

注1：当期純利益は親会社株主に帰属する当期純利益。

注2：EPSは完全希薄化後 (Diluted) 発行済株数で計算。ただし、時価総額は完全希薄化前 (Basic) で計算。

- スーパー・マイクロ・コンピューターは3月19日付けで、公募増資200万株と、30日以内に30万株を追加売却する追加オプションを発表した。
- 公募価格は875ドル、手取り概算は17.5億ドル（追加オプションを除く）。
- 資金使途は、仕入れ資金の増強、生産能力の拡大、研究開発投資の増加など。
- また、会社側は3月18日付けで、「B100」「B200」「GB200」搭載の新型AIサーバーとそれらに向けた冷却装置を発売すると発表。

サーバー売上高ランキング（2023年7-9月期）

順位		サーバー売上高	市場シェア	前期比	前年比
1	ODMダイレクト	6,361.0	24.9%	-19%	-4%
2	デル・テクノロジーズ	3,497.0	13.7%	1%	-30%
3	IEITシステムズ（Inspur）	2,943.7	11.5%	50%	30%
4	HPエンタープライズ	2,316.8	9.1%	-10%	-31%
5	スーパーマイクロ	1,882.2	7.4%	-3%	21%
6	ファーウェイ（xFusionを含む）	1,528.8	6.0%	22%	-10%
7	レノボ	1,331.5	5.2%	-2%	-39%
8	H3C	1,160.4	4.5%	46%	26%
9	IBM	911.0	3.6%	-15%	0%
10	Cisco	729.9	2.9%	5%	-14%
	その他	2,886.2	11.3%	3%	29%
	合計	25,548.5	100.0%	-1%	-7%

単位：100万ドル

出所：OMDIA2023年12月6日付けプレスリリースより楽天証券作成

注：ODMダイレクトは大手クラウドサービス会社等の内製サーバーを受託生産するサーバーメーカー、電子機器組立業者等の合計。

サーバー各社のAIサーバー売上高

	2023年1-3月期	2023年4-6月期	2023年7-9月期	2023年10-12月期
スーパー・マイクロ・コンピューター	3.7億ドル	約10億ドル	11億ドル以上	20億ドル以上
	2023年2-4月期	2023年5-7月期	2023年8-10月期	2023年11月-2024年1月期
デル・テクノロジーズ			5億ドル以上	8億ドル

単位：億ドル

出所：スーパーマイクロは、会社資料、発言をもとにした楽天証券推定。デルは決算電話会議での会社側の発言による。

スーパー・マイクロ・コンピューターの業績

	2022年6月期	2023年6月期	2024年6月期 会社予想 (前回)	2024年6月期 会社予想 (今回)	2024年6月期 楽天証券予想 (前回)	2024年6月期 楽天証券予想 (今回)	2025年6月期 楽天証券予想 (前回)	2025年6月期 楽天証券予想 (今回)
売上高	5,196	7,123	10,500	14,500	14,700	14,700	26,000	26,000
前年比	46.1%	37.1%	47.4%	103.6%	106.4%	106.4%	76.9%	76.9%
営業利益	335	761			1,370	1,370	2,500	2,500
営業利益率	6.4%	10.7%			9.3%	9.3%	9.6%	9.6%
前年比	170.2%	127.2%			80.0%	80.0%	82.5%	82.5%
当期純利益	285	640			1,170	1,170	2,130	2,130
前年比	154.5%	124.6%			82.8%	82.8%	82.1%	82.1%
EPS	4.87	10.93			19.98	19.98	36.38	36.38
配当	0.00	0.00			0.00	0.00	0.00	0.00
PER	210.6	93.8			51.3	51.3	28.2	28.2

株価 (NASDAQ) 1,025.06 米ドル (2024年3月26日)

時価総額 60,017 百万ドル (2024年3月26日)

発行済株数 58.550 百万株 (公募200万株後の発行済み株式数。追加オプション30万株を除く)

単位：百万ドル、ドル、%、倍

出所：会社資料より楽天証券作成。

注1：当期純利益は親会社株主に帰属する当期純利益。

注2：EPSは完全希薄化後発行済み株式数で計算。ただし、時価総額は完全希薄化前発行済み株式数で計算。

注3：会社予想は予想レンジの平均値。

注4：表中の会社予想（前回）は2024年6月期1Q決算時の会社側ガイダンス、会社予想（今回）は2024年6月期2Q決算時の会社側ガイダンス。

●HBMとDRAMの重要性を再認識したい

- ✓HBM（HBM3e）の増産とDRAM（DDR5）のウェハの増産が必要になっている。2025年は過去最大規模のDRAM投資か。
- ✓NAND投資もそれに続く可能性がある。
- ✓マイクロン・テクノロジー

マイクロン・テクノロジーの業績

	2023年8月期 1Q	2Q	3Q	4Q	2024年8月期 1Q	2Q	3Q会社予想
売上高	4,085	3,693	3,752	4,010	4,726	5,824	6,600
前年比	-46.9%	-52.6%	-56.6%	-39.6%	15.7%	57.7%	75.9%
営業利益	-209	-2,303	-1,761	-1,472	-1,128	191	573
営業利益率	-5.1%	-62.4%	-46.9%	-36.7%	-23.9%	3.3%	8.7%
前年比	赤転	赤転	赤転	赤転	赤字	黒転	黒転
当期純利益	-195	-2,312	-1,896	-1,430	-1,234	793	189
前年比	赤転	赤転	赤転	赤転	赤字	黒転	黒転

単位：100万ドル、%

出所：会社資料より楽天証券作成。

注1：当期純利益は親会社株主に帰属する当期純利益。

注2：会社予想は予想レンジの平均値。

マイクロン・テクノロジーの業績

	2021年8月期	2022年8月期	2023年8月期	2024年8月期 楽天証券予想 (前回)	2024年8月期 楽天証券予想 (今回、暫定値)	2025年8月期 楽天証券予想 (前回)	2025年8月期 楽天証券予想 (今回、暫定値)
売上高	27,705	30,758	15,540	22,800	25,000	31,000	40,000
前年比	29.3%	11.0%	-49.5%	46.7%	60.9%	36.0%	60.0%
営業利益	6,283	9,702	-5,745	-400	800	6,000	8,600
営業利益率	22.7%	31.5%	-37.0%	-1.8%	3.2%	19.4%	21.5%
前年比	109.2%	54.4%	赤転	赤字	黒転	黒転	975.0%
当期純利益	5,861	8,687	-5,833	-700	400	5,400	7,300
前年比	118.1%	48.2%	赤転	赤字	黒転	黒転	1725.0%
EPS	5.26	7.80	-5.24	-0.63	0.36	4.85	6.55
配当	0.00	0.43	0.46	0.46	0.46	0.46	0.46
PER	22.6	15.2	-22.7	-189.1	330.9	24.5	18.1

株価 (NASDAQ) 118.80 ドル (2024年3月26日)

時価総額 131,155 百万ドル (2024年3月26日)

発行済株数 1,114 百万株(希薄化後、Diluted)

発行済株数 1,104 百万株(希薄化前、Basic)

単位：100万ドル、%

出所：会社資料より楽天証券作成。

注1：当期純利益は親会社株主に帰属する当期純利益。

注2：EPSは完全希薄化後（Diluted）発行済株数で計算。ただし、時価総額は完全希薄化前（Basic）で計算。

マイクロン・テクノロジー：テクノロジー別売上高

売上高	2022年8月期 1Q	2Q	3Q	4Q	2023年8月期 1Q	2Q	3Q	4Q	2024年8月期 1Q	2Q
DRAM	5,587	5,719	6,271	4,809	2,829	2,722	2,672	2,755	3,427	4,158
NAND	1,878	1,957	2,288	1,688	1,103	885	1,013	1,205	1,230	1,567
その他	222	110	83	146	153	86	67	50	69	99
合計	7,687	7,786	8,642	6,643	4,085	3,693	3,752	4,010	4,726	5,824

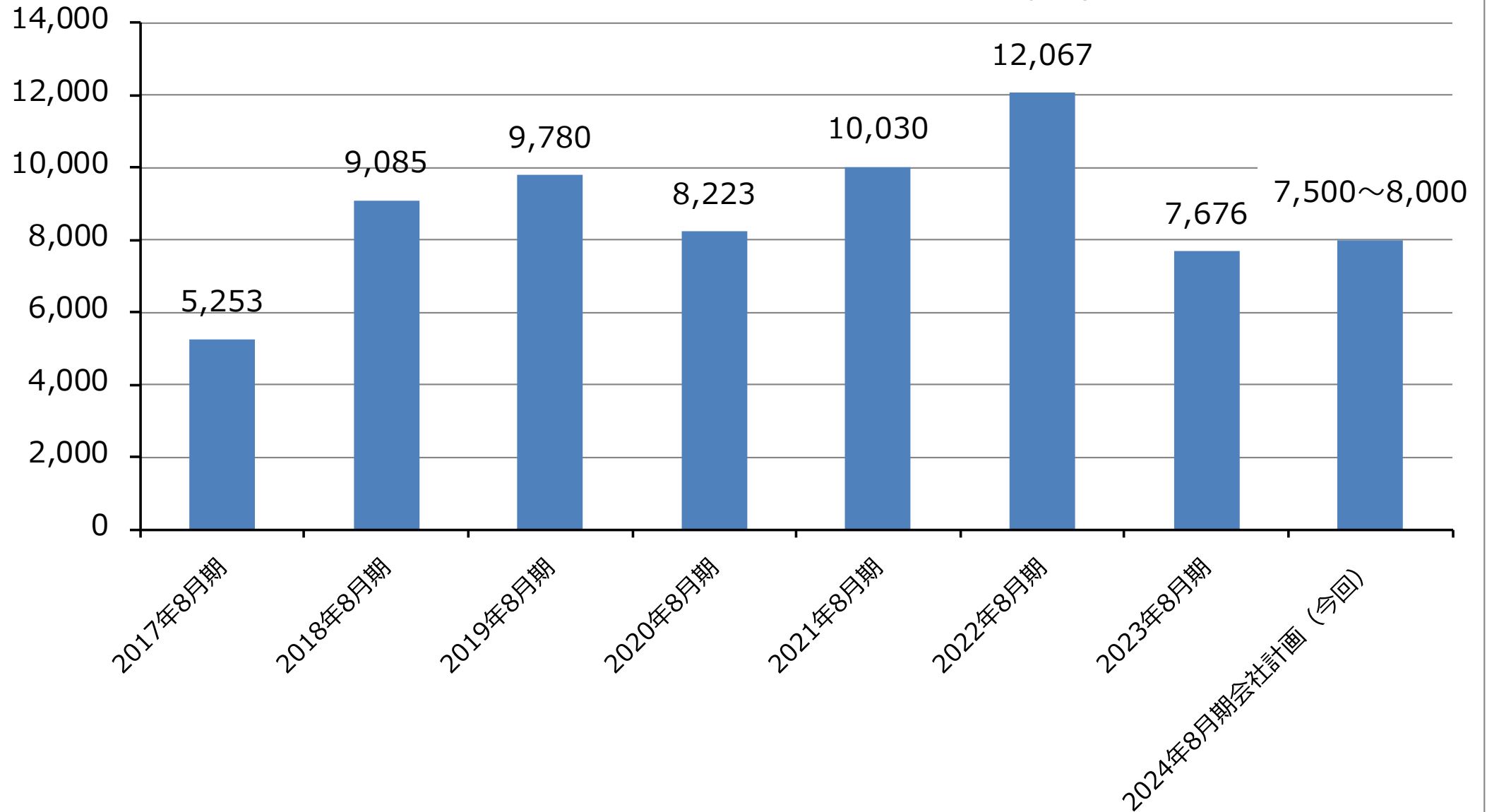
前年比	2022年8月期 1Q	2Q	3Q	4Q	2023年8月期 1Q	2Q	3Q	4Q	2024年8月期 1Q	2Q
DRAM	37.7%	28.7%	15.1%	-21.0%	-49.4%	-52.4%	-57.4%	-42.7%	21.1%	52.8%
NAND	19.3%	18.6%	26.3%	-14.4%	-41.3%	-54.8%	-55.7%	-28.6%	11.5%	77.1%
その他	55.2%	-22.5%	-48.8%	-31.1%	-31.1%	-21.8%	-19.3%	-65.8%	-54.9%	15.1%
合計	33.2%	24.9%	16.4%	-19.7%	-46.9%	-52.6%	-56.6%	-39.6%	15.7%	57.7%

単位：100万ドル、%

出所：会社資料より楽天証券作成

マイクロン・テクノロジーの設備投資：年度ベース

(単位：100万ドル、出所：会社資料より楽天証券作成)

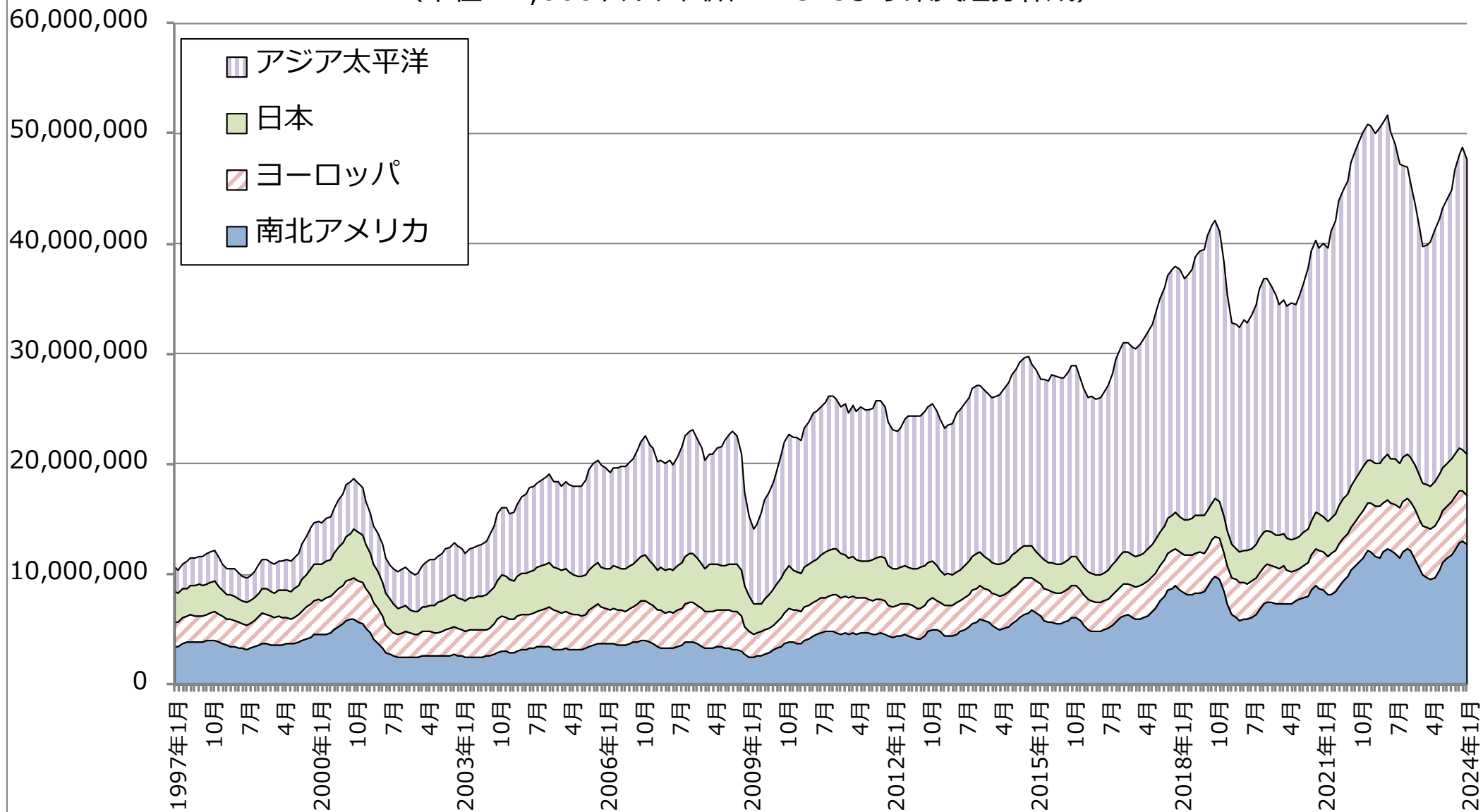


●半導体製造装置の注目点

- ✓4ナノ、DRAMの前工程投資、HBM増産投資が増加すると予想される：
東京エレクトロン、アプライド・マテリアルズ、ASMLホールディング
- ✓ディスコの「中工程」が重要になる：ウェハを薄くするグラインダを前工程に近いクリーン度の高い工程で使う。グラインダの価格が2倍になる。
- ✓検査が重要：レーザーテック（フォトマスク欠陥検査を精緻に行う必要がある）、アドバンテスト（GPUテスト、メモリ・テストの需要増加に期待）
- ✓AI用GPUもHBMも、ダイサイズの大きさが生産の困難と半導体製造装置の需要増加に結びついている。

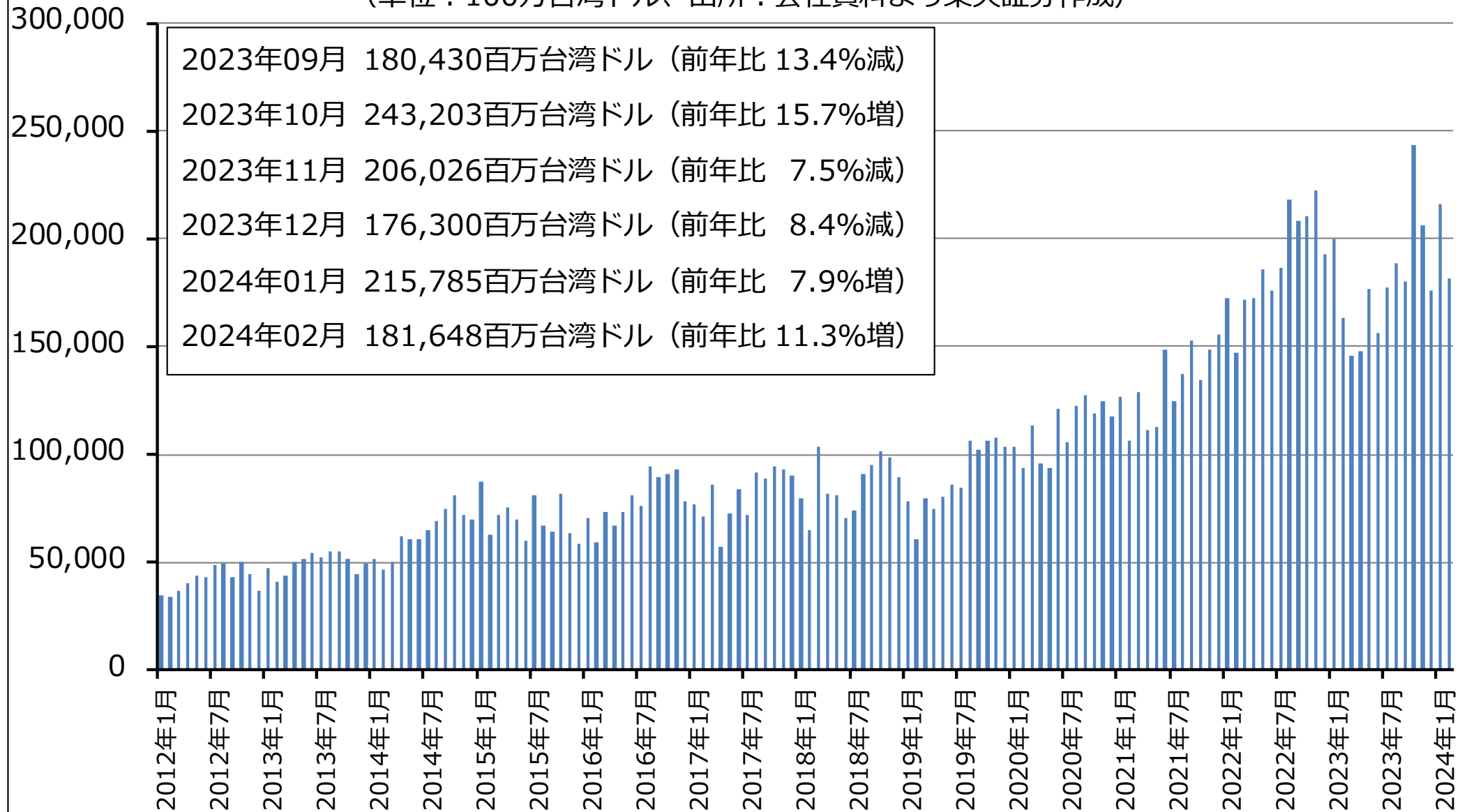
世界半導体出荷金額（3カ月移動平均）

（単位：1,000ドル、出所：WSTSより楽天証券作成）



TSMCの月次売上高

(単位：100万台湾ドル、出所：会社資料より楽天証券作成)



日本製半導体製造装置販売高（3カ月移動平均）

（出所：日本半導体製造装置協会、単位：100万円、%）

