

# 半導体漫遊記

## 湯之上隆

(352)

されているのか？ 米マイクロソフト、米グーグル、米アマゾンなどクラウドメーカーの需要は満たされているのか？

ここで台湾の調査会社デジタイムリサーチのレポートによると、AIサーバーとはHBMを搭載していないAI半導体を2個以上搭載しているシステムであり、ハイエンドAIサーバーとはHBMを搭載したAI半導体（つまりNVIDIAのGPU等）を少なくとも4個搭載したシステムと定義している。

そしてChatGPレベルの生成AIは、ハイエンドAIサーバー上で動作する。さらにChatGPレベルの生成AIの開発と稼働には、3万台ものハイエンドAIサーバーが必要であるという。では、そのハイエンドAIサーバーはどのくらい出荷されているのだろうか？

23〜24年に、NVIDIAのGPUは1個4万円（約600万円）に高騰した。本格的な生成AIブームはこれから到来する。しかし依然としてGPUは大幅に不足している。

（微細加工研究所・所長）

2022年11月30日に米オープンAI社がChatGPTを公開して以降、生成AI（人工知能）が大ブームとなり、AI半導体としてNVIDIAのGPUが引っ張りだこになった。しかしGPUの生産において、TSMCの工程とDRAMを積層した広帯域メモリ（HBM）の2つがボトルネックとなり、GPU不足が続いていた。

TSMCの工程とは、別々に生産したGPU、CPU、HBMなどを12インチリコウエハから角形に切り出した基板に配置する工程のことである。その基板をシリコンインターポーザと呼ぶが、GPUが年々性能を向上させてきたために、GPUチップが巨大化し、HBMの個数も増えてきた。その結果、インターポーザも年々大型化することになり、1枚のウエハから取得できるインターポーザ

数が減少し、これがGPU不足の元凶となっていた。そのような中でTSMCがインターポーザのキャパシティを倍増し、SKハイニックスなどのDRAMメーカーがHBMを増産したことにより、GPUのリードタイム（生産にかかる

デジタイムリサーチのレポートによると、AIサーバーとはHBMを搭載していないAI半導体を2個以上搭載しているシステムであり、ハイエンドAIサーバーとはHBMを搭載したAI半導体（つまりNVIDIAのGPU等）を少なくとも4個搭載したシステムと定義している。

そしてChatGPレベルの生成AIは、ハイエンドAIサーバー上で動作する。さらにChatGPレベルの生成AIの開発と稼働には、3万台ものハイエンドAIサーバーが必要であるという。では、そのハイエンドAIサーバーはどのくらい出荷されているのだろうか？

# GPU祭り は序章

## DNAI 本格AIブームはこれから

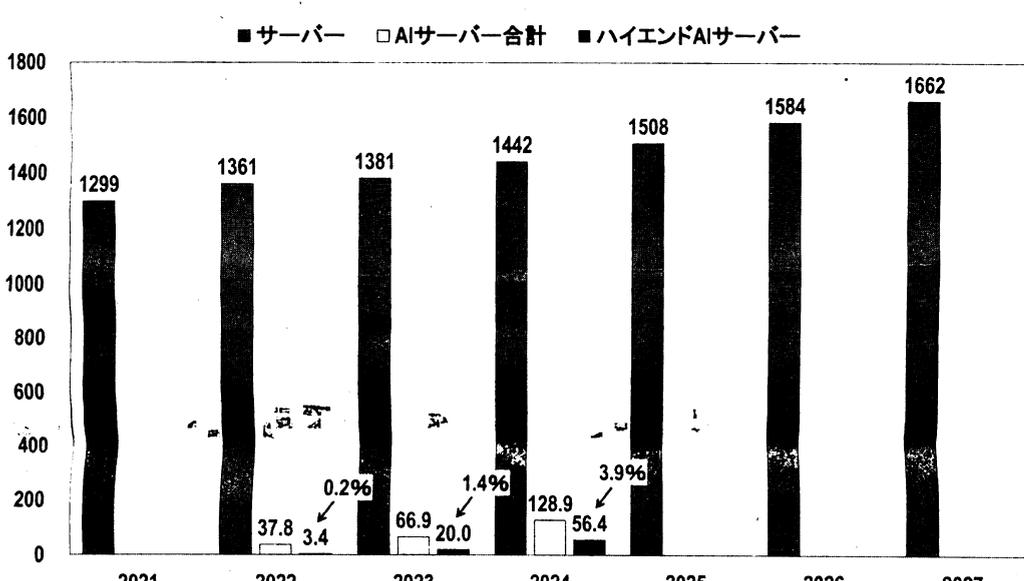
とも4個搭載したシステムと定義している。

そしてChatGPレベルの生成AIは、ハイエンドAIサーバー上で動作する。さらにChatGPレベルの生成AIの開発と稼働には、3万台ものハイエンドAIサーバーが必要であるという。では、そのハイエンドAIサーバーはどのくらい出荷されているのだろうか？

取得できるインターポーザ数が減少し、これがGPU不足の元凶となっていた。そのような中でTSMCがインターポーザのキャパシティを倍増し、SKハイニックスなどのDRAMメーカーがHBMを増産したことにより、GPUのリードタイム（生産にかかる

デジタイムリサーチのレポートによると、AIサーバーとはHBMを搭載していないAI半導体を2個以上搭載しているシステムであり、ハイエンドAIサーバーとはHBMを搭載したAI半導体（つまりNVIDIAのGPU等）を少なくとも4個搭載したシステムと定義している。

そしてChatGPレベルの生成AIは、ハイエンドAIサーバー上で動作する。さらにChatGPレベルの生成AIの開発と稼働には、3万台ものハイエンドAIサーバーが必要であるという。では、そのハイエンドAIサーバーはどのくらい出荷されているのだろうか？



サーバー、AIサーバー合計、ハイエンドAIサーバーの出荷台数  
 出所：MICおよびDIGITIMES Research: Servers Report Database, 2024, "Global annual server shipments, 2023-2024"を基に筆者作成