2026. 08. 20

금융으로 만나는 새로운세상

IBKS 반도체 In-Depth

# 아직도 사이클 타령


IT/반도체 **김 윤 호**
02) 6915-5656
unokim88@ibks.com

RA **고 혁 진**
02) 6915-5675
ameise02@ibks.com


IBK투자증권에서 반도체 담당하는 김운호입니다.

2026년 7월부터 주가는 크게 하락했습니다. 발표된 2분기 실적, 3분기 이후 실적 가이던스 모두 어느때보다 더 좋을 수 없는 수준의 내용이지만 주가는 이를 반영하지 않고 있습니다. 기대보다는 걱정이 좀 더 크게 작용한 구간으로 보입니다. 펀더멘탈을 흔들만한 이슈는 보이지 않는데 오해를 불러 온 변수는 있었던 것 같습니다. 이 변수를 어떻게 이해하는 것이 올바른가에 대한 제 해석을 보고서에 담았습니다.

시장을 흔든 변수는 5가지로 압축했습니다.

1) AI 투자 성장률 둔화에 대한 우려입니다. 과거에 투자 성장률 둔화는 하락 사이클의 전조였습니다.
2) 효율적 AI 모델 출현. 기존 개발 방식과 투자의 축소로 해석될 여지가 있습니다.
3) 공급자가 만드는 순환 금융. 닷컴 버블을 연상시키는 상황입니다.
4) 투자 계획이 공격적이다. 수요보다 공급이 많을 수도 있다는 불안감이 생기는 변수입니다.
5) 중국발 공급 과잉. Wafer가 두 배씩 늘어나는 것은 분명히 위험 신호일 수 있습니다.

이에 대한 올바른 해석이 이번 보고서 내용입니다.

1) AI 투자 성장률 둔화는 투자 주체의 확산과 투자 대상의 변화를 알면 문제가 되지 않을 변수입니다. Agent AI에서는 CSP(클라우드 서비스업체) 이외 투자 주체(기업, 정부)가 늘어납니다. Hybrid 연산(Cloud+Edge)이 확산되며 Edge가 확대되기 때문입니다. 또한 비싼 GPU 대신 CPU와 메모리 투자 비중이 커집니다.

2) 효율적 AI 모델 출현에 대한 올바른 해석은 개발과 연산을 확인하면 됩니다. 모델 개발에는 적은 자원의 투입이 가능하지만 연산 규모가 커지면 Computing 파워에 대한 수요는 확대되는 것이 당연합니다. 이미 중국 Moonshot의 신규 가입자 중단을 통해서 확인한 바 있습니다.

3) 공급자가 만드는 순환 금융. 닷컴 투자 당시에는 장비업체들이 주도했지만, AI 투자에는 칩 메이커, 클라우드 서비스업체뿐만 아니라 금융기관도 참여하고 있습니다. AI 생태계 확장에 공감대가 형성된 결과로 보입니다. 미국과 중국의 대립 구도도 일조하고 있는 것으로 판단합니다.

4) 투자 계획이 공격적이다. 예상되는 수요를 고려하면 공격적일 수가 없어 보입니다. 메이커들은 수요 없는 투자를 하지 않습니다. 강한 수요를 LTA(장기공급계약)가 담보하고 있습니다. 아직 수요를 걱정할 단계는 아니라고 봅니다.

5) 중국발 공급 과잉은 오지 않을 것으로 봅니다. 생산능력으로만 보면 이미 공급과잉이어도 이상하지 않습니다. Wafer 생산량보다 무엇을 어떻게 생산하는지가 더 중요한 변수이기 때문입니다. CXMT 주력 제품은 모바일, PC이고, 생산 공정(선폭)은 5세대 늦습니다.

AI 수요는 아직 3부 능선에 있습니다. 메모리 수요가 폭발적으로 늘어날 것으로 보이는 Agent AI는 시작도 하지 않았습니다. 펀더멘탈을 걱정할 시기가 아니라고 생각합니다.

# CONTENTS

## 쏟아지는 걱정들... 알고 보면 기우

### 사이클로 보면 꼭지처럼 보이지만... 성장으로 보면... 과정이다.

고점 논란은 진행형

AI가 사이클 산업이라는 의견이 아직도 주식 시장에서는 논란거리다. 커지고 있는 메모리 시장, 그에 비해서 턱없이 부족한 공급 능력, 그리고 이러한 상황을 보증하는 LTA의 증가를 확인하면서도 꺼질 수 밖에 없는 시장이라고 단언하고 주가는 이를 반영하고 있다.

올바른 해석이 필요하다

주가 하락은 이러한 변수들을 반영한 것으로 보인다. 아직 성장 단계 초입이라 판단하고 있기 때문에 논란을 불러온 변수에 대해서 올바른 해석이 필요하다고 판단했고, 항목별로 올바른 해석을 정리했다.

투자 방향과 투자 대상이 바뀐 것이다

1. AI 투자 성장률 둔화 : AI 투자의 증가율 둔화는 기저효과가 높은 영향도 있지만 투자 주체의 변화, 투자 대상의 변화가 중요 변수이다. Agent AI 모델 확산에 따른 결과인데 연산이 클라우드뿐만 아니라 Edge, Device에서 진행되는 비중이 높아지고 있다. 또한 AI 연산에 CPU와 메모리가 더 중요한 역할을 하기 시작했다.

개발의 효율성과 연산은 별개 문제

2. 효율적 AI 모델 출현 : 효율적 AI 모델은 투자가 적고, 자원을 효율적으로 사용한다는 의미일 뿐이다. 연산 규모가 커지고, 동시 접속자 수가 많아지고, 처리 토큰량이 증가하면 Computing 파워에 대한 수요가 커질 수 밖에 없다.

누가 투자하는 것이 중요하지 않다. 수요가 있다는 것이 핵심

3. 공급자가 만드는 시장은 허수일까 : 누가 투자하는지가 아니라 왜 투자하는지가 변수이다. 공급자가 수요를 만들어서 없는 수요가 생기고 이것이 공급 과잉으로 이어질 것이라는 닷컴 버블의 양상은 AI 투자와 거리가 멀다. SPV를 활용한 Financing은 점차 확대될 것으로 보인다. 누가 투자하는 것이 중요한 것이 아니라 투자로 공급 과잉이 될 지가 변수이다. 그 가능성은 낮다고 판단한다.

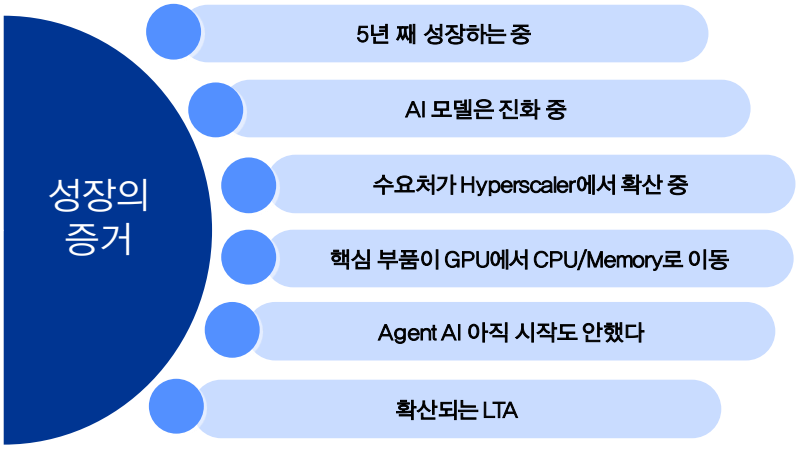
수요를 고려하지 않는 투자는 없다. 그래서 LTA를 한다

4. 투자 계획을 고려할 때 공급 과잉이 올 것이다 : 수요를 고려해서 투자한다. 누구도 공급 과잉을 바라지 않는다. 2분기 중에 발표한 메모리 업체들의 장기 투자 규모는 이례적으로 큰 규모이다. 수요가 빠르게 성장하는 것에 대응하는 것으로 해석한다. 수요 상황을 무시하고 무분별하게 투자한다면 공급 과잉을 피할 수 없겠지만 수요를 확인하지 않고 무작정 투자를 하지는 않는다. 이를 담보하기 위해서 LTA를 체결하는 것이다.

중국은 무시해도 된다

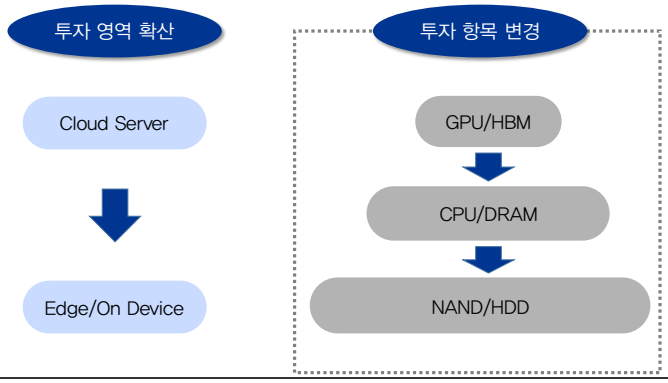
5. 중국발 공급 과잉 : CXMT가 상장 때 확보한 자금으로 공격적인 투자를 할 것이고 이에 따른 공급 과잉을 우려하고 있다. 공급 과잉 가능성은 낮다고 판단한다. 이미 Micron과 동일한 수준의 Wafer 생산 능력을 갖고 있지만 생산량은 1/4 수준이다. 공정 노드, 제품믹스에서 모두 열위에 있기 때문이다. Wafer 물량이 중요한 변수가 아니다.

그림 1. 성장 중인 증거(정황)



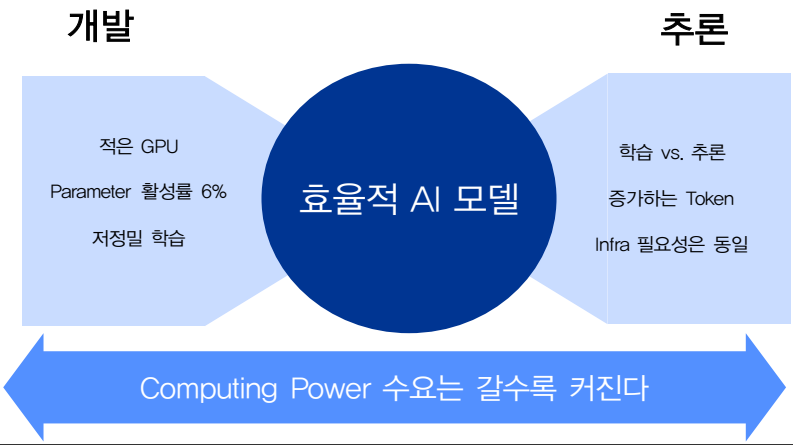

자료: IBK투자증권

그림 2. AI 투자 증가율 둔화 : 투자 주체와 투자 항목이 바뀌고 있다



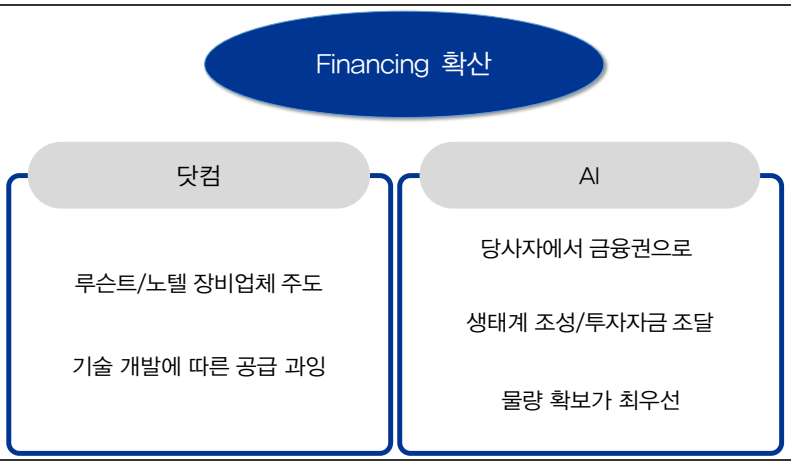

자료: IBK투자증권

그림 3. 효율적 모델 Computing 파워 확대는 필수적




자료: IBK투자증권

그림 4. 투자 주체가 아니라 수요가 핵심이다




자료: IBK투자증권

그림 5. 공급과잉은 요원


자료: IBK투자증권

그림 6. 중국발 공급 과잉 리스크는 제한적


자료: IBK투자증권

## 아직도 사이클 타령

아직도 이어지는 의문들

최근 여러 이슈들이 주가를 하락하게 하는 변수로 작용했다. 이는 아직 메모리 산업이 과거와 달리 성장 산업으로 변화된 것이 아니라, 여전히 사이클 산업에서 벗어나지 못했다는 의구심에서 비롯되었다 판단한다.

성장 국면이라는 근거는 향후 3년간 성장이 예상되고, 모델의 진화로 메모리에 대한 수요가 증가할 것으로 보기 때문

메모리가 과거와 달리 성장 산업의 성격으로 변화하고 있다는 것은 다음과 같은 변수에서 확인할 수 있다.

1. 시장 규모의 성장이 지난 사이클의 3년을 크게 벗어났다. 2022년을 저점으로 2030년까지 성장할 것으로 전망한다. 2031년부터 감소한다는 의미도 아니다. 5년 이상 성장한다면 사이클이 아니거나 예년보다 훨씬 길다는 의미이다. 이 상황에서 하락 사이클을 언급할 이유가 없다.

2. AI 모델은 진화하고 있다. 메모리의 수요 규모가 AI 초기에 전망했던 것에 비해서 너무 커지고 있다. Generative AI에서 Agent AI로 진화하면서 Token 규모는 10,000배 커진다.

3. 더이상 Hyperscaler가 주도하지 않는다. Edge 수요가 커지고, 기업 수요, 소버린 AI를 위한 국가단위 수요가 확대되고 있다.

4. 투자가 GPU에 국한되지 않는다. CPU, DRAM/NAND 비중이 더 커지게 된다. Agent AI의 특징이다.

5. AI 모델의 정점이라고 전망하는 Agent AI는 아직 시작하지도 않았다.

6. 공급 증가에 한계가 드러났다. LTA가 확산되는 이유이다.

올바른 해석이 필요한 5개 이슈

이와는 별개로 성장이 꺾일 수 있다고 나오는 뉴스가 주가를 크게 흔들고 있다. 아래 5개가 대표적 변수라 판단하고, 이에 대한 올바른 해석이 무엇인지는 아래와 같다.

1. AI 투자 성장률 둔화 : 어디에 투자하는지를 살펴보자

2. 효율적 AI 모델 출현 : Computing 파워는 필연적으로 동반

3. 공급자가 만드는 시장은 허수일까 : 과거와 다른 투자 형태일 뿐

4. 투자 계획을 고려할 때 공급 과잉이 올 것이다 : 수요를 고려해서 투자한다

5. 중국발 공급 과잉 : 지나보면 허수

## AI 투자 성장률 둔화 : 투자 주체/투자 대상이 바뀌고 있다.

수요가 줄어들어서 사이클이 하락할 것이라는 시나리오

AI 투자 감소는 결국 올 것이고 이로 인해서 수급 상황은 개선될 것이다. 높은 DRAM 가격과 영업이익률도 하락할 것으로 보는 게 맞다. 하락 사이클을 언급할 때마다 등장하는 논리이다.

현재 업황은 수요/공급 모두 변수로 작용 중

사이클 산업에서 상승과 하락은 공급 부족과 공급 과잉을 의미한다. 지금 메모리 산업은 수요는 크고 공급이 제한적인 상황이라 수요, 공급 모두 변수로 작용하고 있다. 공급 변수는 차치하고 수요 변수를 중점으로 분석하기로 한다.

AI투자에서 Hyperscaler 비중은 낮아질 것

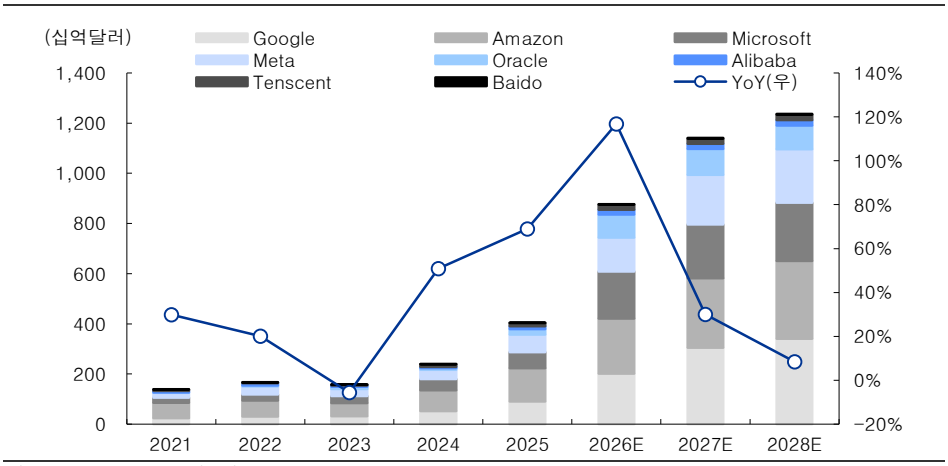
메모리 수요는 Hyperscaler가 주도하고 있다. AI 모델의 진화로 메모리 수요가 급증하고 있기 때문이다. 그럼에도 2027년 투자 증가율은 2026년 대비 낮을 것으로 전망한다. Hyperscaler들의 투자 증가율이 낮아지고 있기 때문이다.

Hyperscaler들의 투자 규모가 크게 증가하지 않을 것으로 전망하는 것은 Training, Inference AI 관련 인프라 투자 관련 수요가 성숙 국면에 진입했기 때문이다.

투자는 서버에서 Edge로 확대

이것이 투자 규모가 감소하는 상황으로 이어지는 것은 아닐 것으로 판단한다. 투자 주도 세력이 바뀌고 있기 때문이다.

그림 7. Hyperscaler 투자 규모 추이




자료: Bloomberg, IBK투자증권

## 1. Data Center에서 Edge로

Data Center 투자에서 Computing 비중 급증

Data Center 투자에서 항목별 비중 추이 변화는 Computing의 중요성을 잘 보여준다. AI 투자가 확산되기 전인 2022년, 2023년 Computing 비중은 35%에 못 미치는 수준이었다. 2024년에 46.4%, 2025년 51.6%, 2026년에는 59.1%로 전망한다.

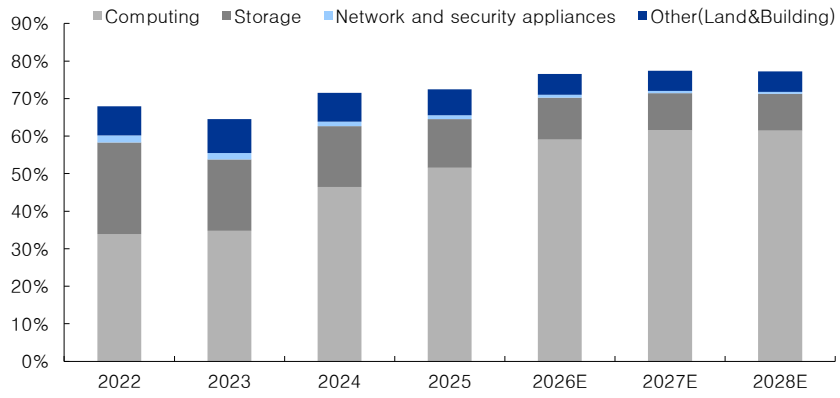
Computing 투자금액은 더 빠르게 증가하는 중

Computing 비중 증가는 둔화되고 있지만 투자 금액은 가파른 증가세를 보이고 있다. 2025년은 2024년 대비 68.1% 증가, 2026년은 2025년 대비 79% 증가, 2027년은 2026년 대비 27.2% 증가할 것으로 전망하고 있다. Hyperscaler들의 투자 증가율 둔화와 비슷한 맥락이다.

Computing 투자 증가율 둔화는 Agent AI 영향

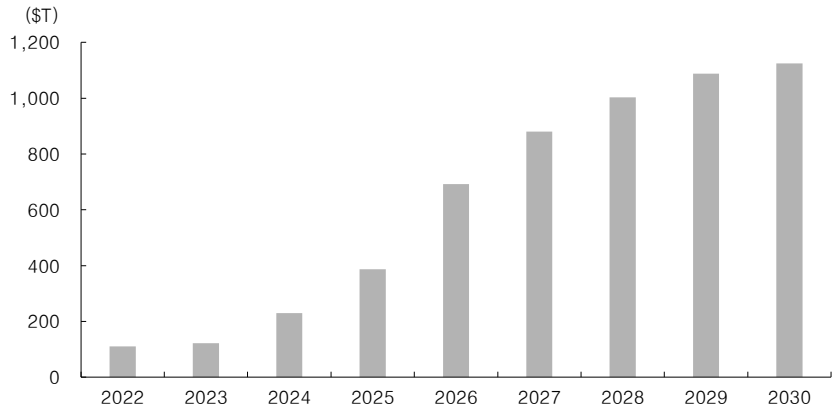
Computing에는 GPU뿐만 아니라 CPU, ASIC이 포함되고 메모리도 HBM, DRAM, NAND를 모두 포함하고 있다. Agent AI 시대로 진입하면서 CPU와 DRAM에 대한 의존도가 커지고, KV Cache 메모리에 대한 수요가 증가하면서 NAND 수요도 가파르게 증가하고 있어서 전체 투자 규모의 성장률 둔화가 메모리 수요 증가율이 둔화되는 것을 의미하지 않을 것으로 판단한다.

그림 8. Data Center 투자에서 항목별 비중




자료: OMDIA, IBK투자증권

그림 9. Computing 투자 규모 추이 및 전망




자료: OMDIA, IBK투자증권

## 2. GPU가 아닌 CPU

Agent AI는 분산 투자를 의미

AI 모델은 Generative AI, Reasoning AI, Agent AI로 진화하고 있다. Agent AI로 갈수록 Token 수는 엄청난 규모로 커지지만 연산 처리는 분산된다. 클라우드 중심이 아닌 Edge와 Hybrid 연산 방식이다.

Agent AI는 Hybrid agentic inference로 진화

Hybrid agentic inference는 민감한 파일과 기밀 데이터는 로컬 모델이 처리하고, 외부 리서치와 대형 모델 추론은 클라우드 모델이 처리하는 구조이다. Intel은 미래 Computing이 Data Center에 더 많은 compute, 로컬 머신에도 더 많은 compute가 함께 존재하는 Hybrid architecture로 진화한다고 강조했다.

이는 AI 투자를 확인할 때 서버 투자나 Hyperscaler들의 투자가 전부가 아니라는 게 달라진 점이다.

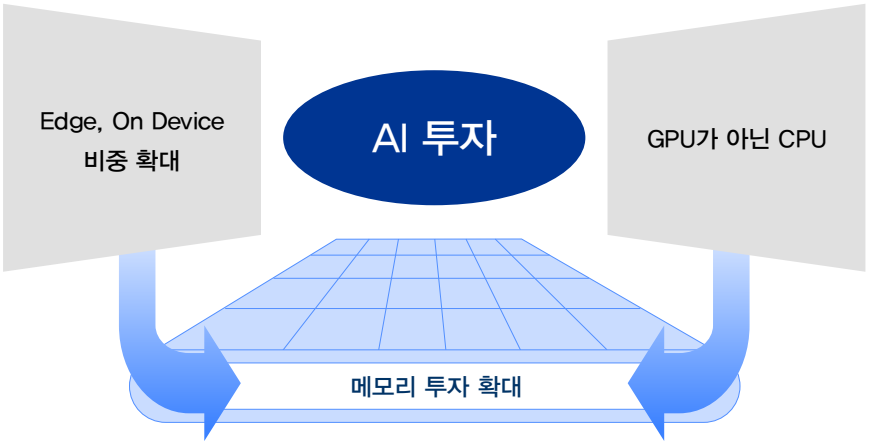
CPU가 중요한 자산으로 변신

다른 하나는 GPU보다 CPU 의존도가 커지고 있다. Agent AI 단계에서는 더 그렇다. 컨트롤 타워가 각각의 Agent 구동과 데이터 관리를 해야 하기 때문이다. 병렬 연산이 주요 역할인 GPU보다는 빠른 연산이 주요 역할인 CPU 수요가 커지는 이유이다.

투자 금액의 증가율은 둔화되더라도 금액은 지속적으로 성장한다. 투자 집행이 High End GPU가 아니라 CPU, DRAM으로 바뀐다면 DRAM에 대한 수요가 줄어들 것인지가 관전 포인트이다.

AI 투자 주체가 확산되고,
투자 대상이 바뀌고 있다.

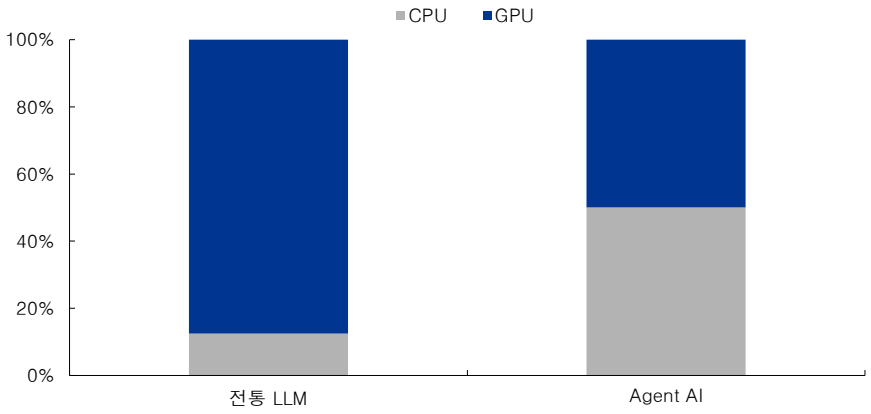
그림 10. AI 투자 방향 전환




자료: IBK투자증권

Agent AI는
CPU 중요성이 커진다.

그림 11. GPU/CPU workload 비율



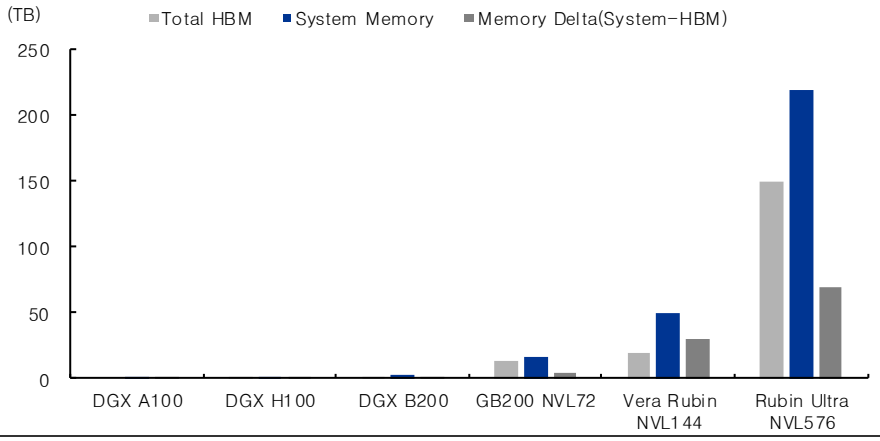

자료: TrendForce, IBK투자증권

## 3. 프로세서보다는 메모리

시스템 메모리
폭발적 증가세

메모리에 대한 투자 규모가 커지는 것은 NVIDIA의 서버 시스템에서도 확인할 수가 있다. GB200 NVL72 대비 Vera Rubin NVL144는 시스템 메모리가 3배, Rubin Ultra NVL 576은 10배 이상 소요될 것으로 전망한다.

그림 12. AI 서버 메모리 규모




자료: OMDIA, IBK투자증권

그림 13. Compute 집약적에서 메모리 집약적 AI로

| | AI 가속 시대(2018~21) | HBM 확장 시대(2022~25) | 메모리 용량 시대(2026+) |
|---|---|---|---|
| Driver | • 딥러닝 적용 확대<br>• GPU 가속 주류화<br>• 컴퓨트가 성능을 정의 | • 거대 모델 학습 붐<br>• GenAI 추론 부상<br>• 대역폭 병목 | • Agentic 워크플로우 등장<br>• Multimodal 추론 확대<br>• Context가 병목 |
| Result | • 컴퓨트 중심 인프라<br>• CPU-메모리 중심 시스템<br>• GPU는 가속기 역할 | • GPU 중심 아키텍처<br>• HBM 우선순위 상승<br>• Rack-level HBM 확장 | • Rack-level 메모리 풀<br>• 시스템 메모리 확장<br>• 메모리 계층 아키텍처 |

Model

| | DGX A100 | DGX H100 | DGX B200 | GB200 NVL72 | Vera Rubin NVL144 | Rubin Ultra NVL576 |
|---|---|---|---|---|---|---|
| Year | 2020~21 | 2023 | 2024~25 | 2025 | 2026(예상) | 2027+(예상) |
| Total HBM | ~320GB HBM2E | ~640GB HBM3 | ~1.4TB HBM3E | ~13TB HBM3E | ~20TB HBM4 | ~150TB HBM4E |
| System DRAM | ~1TB DDR4 | ~2TB DDR5 | 2~4TB DDR5 | ~17TB LPDDR5X | 50TB+ SOCAMM2 | ~220TB SOCAMMx |

자료: OMDIA, IBK투자증권

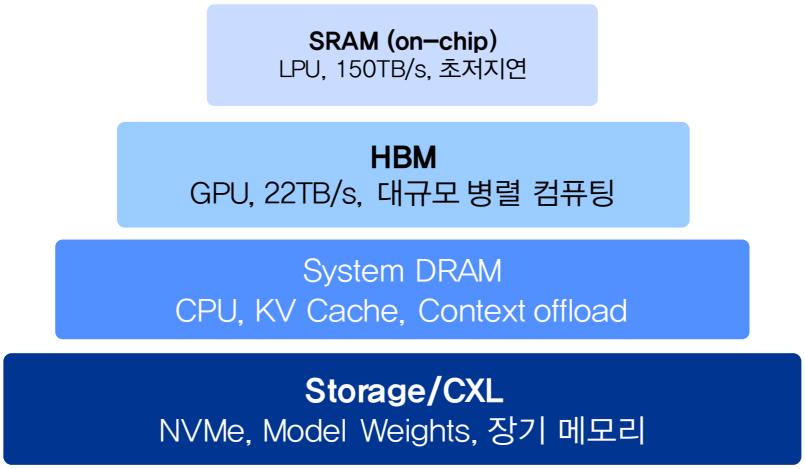
**Full Stacking이 진행되고 있다**

메모리에 대한 수요가 증가하는 것은 AI 서버를 구성하는 시스템의 변화에서도 찾을 수 있다. GPU, HBM 중심의 수요에서 LPU, DPU가 신규로 진입하면서 SRAM, NAND에 대한 수요가 커지게 되고, CPU Rack이 추가되면서 DRAM에 대한 수요도 지속될 것으로 전망한다.

**Rack Solution 확산**

NVIDIA가 Vera Rubin NVL72에서 제시했고, AMD는 Helios로 대응하고 있고, Intel도 CPU 전용에서 GPU와 RDU를 포함한 Rack 시스템을 공개했다.

**그림 14. 풀 스택 메모리 아키텍처**




자료: OMDIA, IBK투자증권

**표 1. 풀 스택 메모리 시사점**

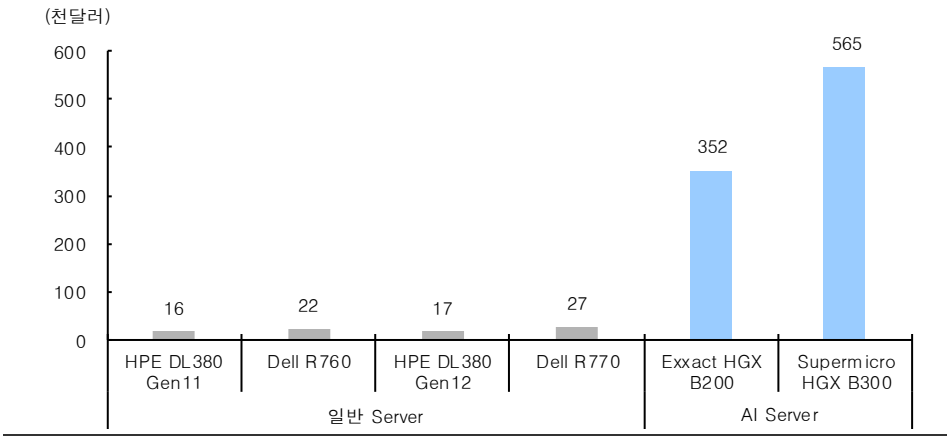
| 구분 | 내용 |
|---|---|
| LPU ≠ HBM 대체 | Ruibn GPU는 Prefill을 담당하고 Groq LPU는 Decode 담당. 경쟁이 아니라 분업 구조. Nvidia Dynamo는 이기종 하드웨어 전반에 워크로드를 동적으로 라우팅 |
| Agentic AI는 풀 스택 필요 | Multi-agent real-time 소통은 초당 1,500개 이상의 토큰을 필요. GPU 단독으로는 해당 지연시간 요구를 충족할 수 없어 SRAM이 메모리 계층에서 필수적 |
| HBM 수요는 계속 성장할 것 | Rubin Ultra는 HBM4E 1TB 탑재해 출시 예정. 학습 및 Prefill 워크로드가 계속 확장되며, SRAM이 단순히 처리할 수 없는 영역에서 HBM 수요는 지속 증가 |

자료: OMDIA, IBK투자증권

동일한 투자 금액으로 CPU 서버를 30배 구매 가능

High End GPU 장착 서버 1대로 일반 CPU 서버 30대를 구매할 수 있다. 동일한 투자 금액으로 CPU/DRAM 서버를 훨씬 더 많이 투자할 수 있다. 투자 금액의 증가폭이 크지 않더라도 메모리 투자는 지속적으로 성장한다고 보는 것이 맞다.

**그림 15. 일반 Server vs AI Server 가격**




자료: 각 사, IBK투자증권
주: 각 사 공식 홈페이지 가격 기준

**표 2. 일반 Server vs AI Server 비교**

| 구분 | 모델 | Compute | 메모리 | 가격(달러) |
|---|---|---|---|---|
| 일반 Server | HPE DL380 Gen11 | 16 Core | 128GB DDR5 | 15,625 |
| | Dell R760 | 16 Core | 64GB DDR5 | 22,125 |
| | HPE DL380 Gen12 | 16 Core | 128GB DDR5 | 16,675 |
| | Dell R770 | 16 Core | 128GB DDR5 | 26,691 |
| AI Server | Exxact HGX B200 | **8× B200** | **1.44TB HBM3e** | 351,899 |
| | Supermicro HGX B300 | **8× B300** | **2.30TB HBM3e** | 565,384 |

자료: 각 사, IBK투자증권
주: 각 사 공식 홈페이지 가격 기준

**표 3. 일반 Server vs AI Server BOM 비교** (단위: 달러)

| 구분 | Sapphire Rapids Server | NVIDIA DGX H100 |
|---|---|---|
| CPU | 1,850 | 5,200 |
| 8 GPU + 4 NV Switch Baseboard | 0 | 195,000 |
| Memory | 3,930 | 7,860 |
| Storage | 1,536 | 3,456 |
| SmartNIC | 654 | 10,908 |
| Chassis(Case, backplanes, cabling) | 395 | 563 |
| Motherboard | 300 | 360 |
| Cooling | 275 | 463 |
| Power Supply | 300 | 1,200 |
| Assembly and Test | 495 | 1,485 |
| Markup | 689 | 42,000 |
| Total | 10,424 | 268,495 |

자료: SemiAnalysis, IBK투자증권

# 효율적 AI 모델 출현 : Computing 파워는 필연적으로 동반

## 1. 효율적 AI 모델의 탄생

DeepSeek V3 개발비는 557.6만 달러

DeepSeek와 Moonshot은 AI 투자의 효율성을 화두로 던진 대표적인 LLM 개발회사이다. 처음 시장에 충격을 준 모델은 DeepSeek의 R1, V3 모델이다. V3 모델의 개발 비용을 공개한 내용은 아래와 같다.

1) 사전학습 : 266만 H800 GPU 시간
2) 후속학습 : 10만 GPU 시간
3) 사전학습 데이터 : 14.8조 Token
4) 최종 학습비 : 557만 6,000달러. H800 임대로 시간당 2달러로 가정

실제 투입비만 산정. 인건비, 이전 개발비는 제외

이는 최종 성공한 모델에 직접 투자한 금액만을 의미하기 때문에 개발비, 인건비, 데이터 수집에 소요된 비용, 이전 모델 개발비 등이 모두 제외된 것이어서 다른 대형 LLM 모델 개발비와 동일선상에서 비교할 수는 없다.

효율적 모델인 이유. MoE, MLA, FP8 등 효율적 자원 사용

그럼에도 효율적인 모델이라고 평가받는 것은 아래와 같은 변수 때문이다.

1) MoE(Mixture-of-Experts) 구조 : 모델의 파라미터 대비 파라미터 활성 비율은 5.5%에 불과
2) MLA(Multi-head Latent Attention) 사용 : 정보를 낮은 차원의 잠재 표현으로 압축해서 저장. KV Cache 메모리 축소
3) FP8 저정밀 학습 : 16 Bit 대비 GPU 연산 처리량, 메모리 사용량 감소. 학습에 사용한 GPU는 2,048개
4) GPU간 통신을 계산과 동시에 처리
5) Multi-Token Prediction : 일반 LLM의 예측 Token은 1개
6) 인력 투입 축소 : 사람이 데이터를 입력하는 것이 아니라 강화학습을 이용

모델 구조가 효율적이지만 한 질문당 총 추론 비용은 커질 수 있다. 추론 과정에서 많은 Token을 생성시키 때문이다.

## 2. 추가적인 하드웨어 확보가 동반되어야 하는 구조

개발은 효율적이지만 컴퓨팅 파워는 필수

DeepSeek가 효율적이라는 말은 하드웨어 투자가 적게 필요하다는 의미가 아니라 동일 성능을 더 적은 연산비로 구현했다는 것을 의미한다. Computing 하드웨어에 대한 필요성을 검증하는 변수는

1. 활성 파라미터가 적어도 전체 모델은 저장해야 함

활성 파라미터가 적지만 필수적으로 저장할 공간은 필요

DeepSeek-V3/R1은 총 6,710억 파라미터 가운데 토큰당 약 370억 파라미터만 활성화하는 MoE 모델이다. 계산량은 줄지만, 어떤 Expert가 호출될지 모르기 때문에 전체 모델 가중치는 GPU·CPU 메모리나 고속 스토리지에 준비돼 있어야 한다. 여기에 KV Cache, 활성값, 통신 버퍼, 런타임 여유 메모리가 추가된다. 따라서 최고 성능으로 서비스하려면 여전히 다수의 고성능 GPU와 대용량 HBM이 필요하다.

**표 4. 정밀도에 따른 필요 저장 용량**

| 정밀도 | 이론상 저장 용량 |
|---|---|
| BF16·FP16, 파라미터당 2바이트 | 약 1.34TB |
| FP8, 파라미터당 1바이트 | 약 671GB |
| 4비트 양자화 | 약 336GB |

자료: OMDIA, IBK투자증권

2. 학습 자체도 2,048개 H800 GPU를 사용

GPU 2,000개 이상 학습시간 280만 시간

DeepSeek-V3는 소규모 서버에서 학습된 모델이 아니다. 공식 기술보고서에 따르면 2,048개의 NVIDIA H800 GPU를 사용했고, 전체 학습에 약 278만 8,000 GPU 시간이 소요되었다.

이를 단순 환산하면 2,048개 GPU를 약 57일 연속 가동한 규모의 GPU가 필요하다는 의미이다. GPU뿐만 아니라 고속 네트워크, 전력, 냉각, 스토리지도 필요하고, 학습 데이터 14.8조 Token의 저장·전처리 인프라도 필요하다. 즉, 기존 미국 최첨단 모델보다 효율적이었다는 것이지 하드웨어가 거의 필요 없었다는 의미는 아니다.

3. 실제 추론 서비스는 학습보다 더 큰 누적 투자가 될 수 있음

추론량이 증가할수록 Token량은 더 많아진다

모델 학습은 일회성에 가깝지만 추론은 사용자가 질문할 때마다 계속 발생한다. 특히 R1 같은 추론 모델은 일반 ChatBot보다 긴 추론 토큰을 생성하기 때문에 한 요청당 연산량이 커질 수 있다.

DeepSeek가 공개한 대규모 추론 시스템 구조에서는 입력 처리(prefill)에 최소 4개 노드·32개 H800, 출력 생성(decode)에는 배포 단위당 18개 노드·144개 H800와 같은 대규모 Expert 병렬화 구성이 사용되었다. 이는 높은 동시접속 처리량을 위한 운영 구조이지, 모델이 1~2개의 GPU로 충분하다는 의미가 아니다.

4. 많은 데이터는 GPU 외 인프라도 요구

인프라 투자는 별개

대규모 모델 개발에는 다음 투자가 동반된다.

**표 5. 모델 개발 시 투자 인프라 및 역할**

| 인프라 | 역할 |
|---|---|
| GPU·AI 가속기 | 학습 및 추론 연산 |
| HBM | 모델 가중치·KV 캐시 저장 |
| 고속 인터커넥트 | MoE 전문가 간 데이터 교환 |
| 대용량 SSD | 데이터셋·체크포인트 저장 |
| 분산 파일시스템 | 다수 GPU에 데이터 동시 공급 |
| 전력·냉각 | 대규모 클러스터 운영 |

자료: OMDIA, IBK투자증권

DeepSeek도 고성능 분산 파일시스템인 3FS, MoE 통신 라이브러리 DeepEP, FP8 행렬연산 라이브러리 DeepGEMM을 별도로 개발했다. 이는 모델 알고리즘뿐 아니라 Data Center 전체를 최적화했다는 의미이다.

5. 효율 개선이 오히려 총 하드웨어 수요를 확대한다.

싸면 더 많이 쓰게 된다

서비스 단가가 낮아지면 사용량이 크게 늘어나다. 제번스 역설이다.

1) 모델 효율이 개선되면 2) Token당 비용이 하락하고, 3) AI 서비스 이용을 증가시킨다. 결국 AI 서비스 이용이 증가하며 4) 추론 Token 총량이 증가하게 되고 이는 5) 전체 GPU·전력·스토리지 수요 증가로 이어지는 것이다.

DeepSeek 같은 효율적인 모델은 GPU 한 개당 처리량을 높이지만, AI를 더 저렴하고 광범위하게 보급해 산업 전체의 Data Center 투자 수요를 오히려 확대할 것으로 판단한다.

## 투자 주체가 아니라 수요가 핵심이다.

26년 들어 투자 공조 확대

AI 모델 업체와 Chip Maker, CSP(Cloud Service Provider)와 투자 계약은 2023년 3건, 2024년 4건, 2025년 4건이 있었고, 2026년에는 8월초까지 17건이 발표되었다. 투자 규모도 점차 커지고 있다.

AI 모델이 감당하기 힘든 수준의 투자 규모

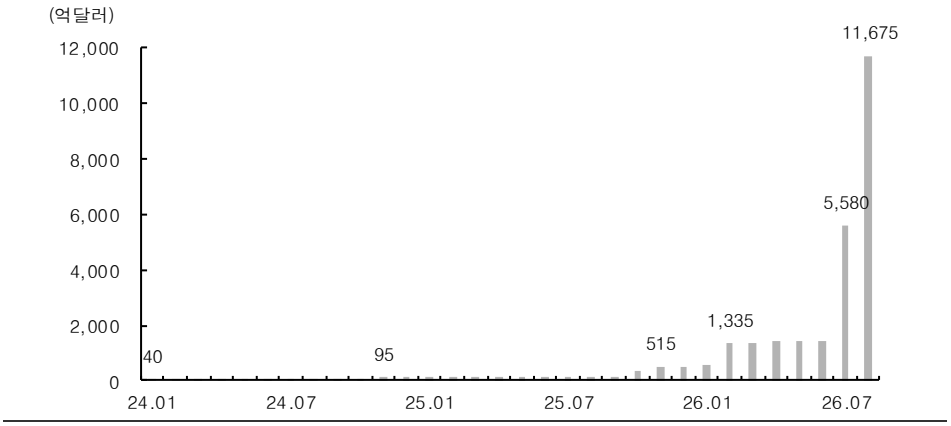
AI 모델업체들의 투자 규모가 커지면서 Financing의 필요성이 커졌고, AI 투자의 최대 수혜 업체인 NVIDIA, AMD가 투자에 적극적이다. CSP인 Microsoft, Amazon, Google, Meta도 같은 입장이다.

금융 그룹의 다양한 Vehicle 가세

AI 투자 구도가 바뀌고 있다. AI 모델 회사들이 직접 투자하던 방식에서 하드웨어 공급업체, CSP, Private Credit, IB, 가상 자산 채굴업체까지 확장되고 있다. 가상 자산 채굴업체가 진입한 배경은 전력 수급이 원인이다. 여유가 있는 전력을 공급하고 지분을 확보하는 방식이다.

최근 NVIDIA는 5,000억달러 규모의 AI Financing을 진행하고 있다. 이는 하나의 프로젝트가 아니라 여러 금융 회사를 중심으로 다양한 금융수단(Vehicle)을 통해서 자금을 확보할 계획이다.

**그림 16. AI 밸류체인 순환금융 규모 추이(누계)**




자료: 언론보도, IBK투자증권
주: 시리즈 투자, 회사채 등 자체 자금조달, 신디케이트 대출, 규모 미공개 건은 제외, 논의 중인 건은 보도일 기준 포함

**닷컴 때와 다른 점은 금융 그룹으로 확산된 것**

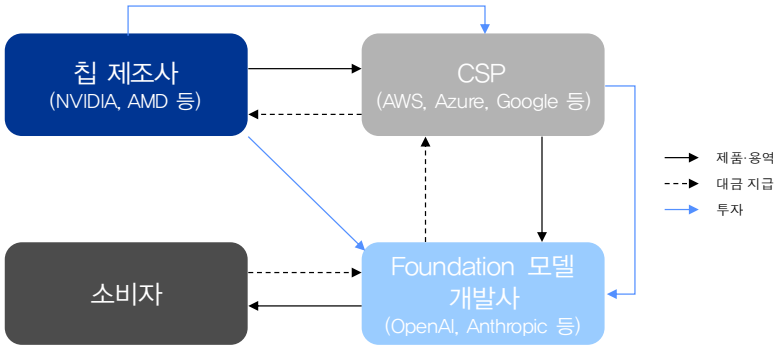
과거 닷컴 버블 당시 루슨트, 노텔이 ISP, Telecom 회사에 장비 투자를 지원한 것과 유사한 면도 있지만 NVIDIA가 독자적으로 지원하는 것이 아니라 금융 인프라를 이용한다는 점에서 차이가 있다.

이러한 움직임이 과거 닷컴 버블과 다른 요인은 수요가 일시적이지 않고, 투자를 줄일 수 있는 효율적인 방법보다 Computing 수요가 훨씬 더 빠르게 성장한다는 점이다.

**시장 규모가 가장 큰 변수. 투자 주체는 변수가 아님**

투자에 대한 우려의 핵심은 과잉투자인지 아닌지에 대한 검증이다. 투자 주체가 누구인지는 중요한 변수가 아니다. 과거에도 누가 투자했느냐가 아니라 공급 과잉으로 이어진 공격적인 투자 규모가 문제였다.

**그림 17. AI 밸류체인 내 순환금융 구조**




*칩 제조사·CSP가 고객사에 공급한 자본이 자사 제품·서비스 구매를 거쳐 매출로 환류*

자료: IBK투자증권

## 1. NVIDIA : 생태계 구축 1인자

생태계 확장이 목표

NVIDIA 투자의 핵심은 리소스가 부족한 AI 모델 업체들에게 최대한 자원을 지원해주면서 AI 생태계를 확장하는 것이다. 이는 결과적으로 NVIDIA가 GPU를 판매할 시장이 커지는 효과도 동반된다.

직접투자에서 Financing으로

투자 방식은 현물 투자 등 직접 투자에서 금융그룹 컨소시엄을 통한 자금조달로 확대되고 있다.

**표 6. NVIDIA AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| Inflection AI 투자 | – | 신규 라운드 투자 |
| xAI 투자 | – | 시리즈 C 투자 참여. NVIDIA Spectrum Ethernet Networking Platform 활용해 Data Center 규모 두 배 확장하고, 20만 개 Hopper GPU 탑재 발표 |
| Lambda 투자 | – | 시리즈 D 투자 참여. 인프라 및 소프트웨어 확장하고 GPU 추가 배포 발표 |
| Open AI 투자 | 최대 1,000억달러 | 차세대 모델 학습 및 실행을 위해 Vera Rubin 등 10GW 규모 NVIDIA 시스템 도입<br>→ LOI 체결 후 2026년 2월 300억 달러 지분투자로 규모 축소 |
| Antrhopic 투자 | 최대 150억달러 | Microsoft 전략적 파트너십 일환으로 각 최대 100억, 50억달러 투자 |
| CoreWeave 투자 | 20억달러 | 2030년까지 5GW 이상 AI Factory 구축 가속화 지원 |
| Open AI 투자 | 800억달러 | Pre Value 7,300억달러로 1,100억달러(Amazon 500억달러, NVIDIA 300억달러, Softbank 300억달러) 신규 투자 |
| Nebius 투자 | 20억달러 | 2030년말까지 5GW 이상 용량 구축을 위한 가속 Computing 플랫폼 도입 지원 |
| IREN 투자 | 최대 21억달러 | IREN 주식 21억달러 규모 주식매입권 확보하고, DSX 기반 최대 5GW 규모 AI 인프라 구축 지원 |
| Sharon AI 투자 | – | 최대 4만 개 GB300 기반 72MW Data Center 구축 협력. NVIDIA가 수익 공유 및 Credit 지원 모델 제공 |
| Naver 투자 | 최대 10억달러 | 2028년까지 DSX AI Factory 설비를 55MW에서 200MW로 확장하고 NVIDIA AI 인프라 구축을 1GW까지 확장 발표 |
| Hut 8 투자 | 최대 500억달러 | Hut 8이 건설 중인 1GW 규모 텍사스 Data Center 임대 계약 체결. NVIDIA는 네오클라우드에 재임대 계획 |
| SB Energy 투자 | 1,050억달러 | 최대 8GW 규모 Ohio Data Center 구축 관련 Open AI 임대료 등 1,050억달러 지급보증 |
| SB Energy 투자 | 15억달러 | Ohio Data Center 구축을 위한 투자 |
| Open AI 투자 | 3,500억달러 | Open AI의 칩 구매 자금 조달 논의 |
| Lancium 투자 | 최대 30억달러 | 20억달러를 투자해 전력 인프라 업체 Lancium의 지분 20% 우선 확보 후 추가 투자 계획 |
| Open AI 등 투자 | 5,000억달러 | Apollo, Blackstone 등과 협력해 개별 vehicle로 나눠 5,000억달러 규모 자금 조성 논의 |

자료: 언론보도, IBK투자증권
주: 현재 거래 체결 전으로 논의 중인 사항은 음영처리

## 2. AMD : NVIDIA 뒤를 따라서

NVIDIA와 유사한 행보

AMD는 NVIDIA만큼 투자 규모가 크지는 않지만 xAI, Open AI, Meta, Anthropic에 Instinct GPU 공급과 관련된 투자를 꾸준히 지속하고 있다. Agent AI 모델에서는 CPU의 역할이 더욱 중요해지면서 AMD도 Rack solution인 Helios를 적극 확대할 계획이다.

Chip Maker로서 시장 확대 필요성은 NVIDIA와 유사한 상황이다. NVIDIA에 비해서 AI 생태계 내 입지가 좁아서 투자 규모는 낮은 수준이다. 향후 Agent AI가 보편화되는 시기에는 CPU 투자 비중이 높아지면서 생태계 조성을 위한 투자 규모도 확대될 것으로 예상한다.

**표 7. AMD AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| xAI 투자 | - | 시리즈 C 투자 참여. NVIDIA Spectrum Ethernet Networking Platform 활용해 Data Center 규모 두 배 확장하고, 20만 개 Hopper GPU 탑재 발표 |
| Liquid 투자 | 2.5억달러 | 시리즈 A 투자 참여. AMD Instinct GPU에서 모델 학습 · 배포 발표 |
| Open AI 투자 | AMD 보통주 최대 1.6억 주 | 차세대 AI 인프라 구축을 위해 AMD Instinct GPU 6GW 계약 체결 |
| Meta 투자 | AMD 보통주 최대 1.6억 주 | 차세대 AI 인프라 구축을 위해 AMD Instinct GPU 6GW 계약 체결 |
| TensorWave 투자 | 3.5억달러 | AMD Instinct MI355X GPU 클러스터 기반 AI 인프라 확장 |
| Anthropic 투자 | 최대 50억달러 | 27년 상반기부터 AMD MI450 GPU 2GW 규모 구매 계획 |

자료: 언론보도, IBK투자증권

## 3. Amazon : 클라우드 1위 업체

CSP이자 칩 메이커

가장 큰 CSP 업체인 Amazon은 최대 AI 모델업체인 Open AI, Anthropic에 투자 협력을 강화하고 있다. AI 모델업체들이 CSP의 최대 고객이기 때문이다. 또한 자체 개발 AI Chip Trainium3의 판매도 목적이다.

**표 8. Amazon AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| Anthropic 투자 | 40억달러 | 클라우드 제공 및 AWS Trainium 및 Inferentia 칩 형태 Computing 인프라 제공 |
| Anthropic 투자 | 40억달러 | AWS를 주요 클라우드 제공 업체로 유지하고 AWS Trainium 및 Inferentia을 통한 모델 학습 |
| Open AI 투자 | 800억달러 | Pre Value 7,300억달러로 1,100억달러(Amazon 500억달러, NVIDIA 300억달러, Softbank 300억달러) 신규 투자 |
| 채권 발행 | 약 538억달러 | 2026년 약 2,000억달러 CapEx 계획 속 AI·AWS 인프라 투자 재원 확보 |
| Anthropic 투자 | 50억달러, 추가 최대 200억달러 | Trainium 3 포함 최대 5GW 용량 확보 예정 및 향후 10년간 AWS 기술에 1,000억달러 이상 투자 발표 |
| 채권 발행 | 최소 250억달러 | AI 투자를 위한 자금 조달 |

자료: 언론보도, IBK투자증권

## 4. Google : 직접 투자 용도

자체 투자 자금 확보

자금 조달에 가장 적극적인 업체 중 하나다. AI 연구 개발 영역에서 압도적 선두 업체였지만 Generative AI 상용화에서 Open AI에 뒤쳐지고, Agent AI도 Anthropic이 선점한 상황이기 때문에 이를 따라 잡기 위한 투자 규모가 가파르게 증가하고 있다. 이를 위한 자금 조달 규모도 CSP 업체 중에서는 가장 큰 규모이다.

**표 9. Google AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| Anthropic 투자 | - | 시리즈 C 투자 참여 |
| 채권 발행 | 약 315억달러 | AI 관련 지출 증가를 위한 100년 만기 채권 발행 |
| 증자 | 최대 847.5억달러 | AI 인프라 강화를 위해 기존 800억달러 조달 계획에서 조달 금액 상향 조정 |
| 채권 발행 | 250억달러 | AI 인프라 투자 지원을 위한 채권 발행 |
| Anthropic 투자 | - | Nexus Data Centers의 Hubbard Campus에 대한 Anthropic 임차료 등 지급보증 |

자료: 언론보도, IBK투자증권
주: 현재 거래 체결 전으로 논의 중인 사항은 음영처리

## 5. Microsoft : Open AI 초기 투자자

Amazon과 유사

Open AI에 초기 투자한 이유로 Generative AI 시장에 가장 손쉽게 진입한 CSP이다. 하지만 Microsoft 자체 모델인 Copilot 성능이 떨어져서 추가적인 투자가 동반되고 있다. 또한 CSP 사업 확대를 위한 AI 인프라 투자가 동반되는 중이다.

**표 10. Microsoft AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| Inflection AI 투자 | - | 신규 라운드 투자 |
| G42 투자 | 15억달러 | Azure 플랫폼에서 AI 애플리케이션 및 서비스 운영하고 Microsoft와 협력해 AI 솔루션 제공 |
| Antrhopic 투자 | 최대 150억달러 | NVIDIA, Microsoft 전략적 파트너십 일환으로 각 최대 100억, 50억달러 투자 |
| Mistral 투자 | 수십억달러 | Mistral의 유럽 기반 GPU 활용해 클라우드 및 AI 서비스 제공 |

자료: 언론보도, IBK투자증권

## 6. Broadcom : XPV 컨소시엄 주도

**표 11. Broadcom AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| Open AI 투자 | - | Open AI가 가속기와 시스템을 설계하고, Broadcom이 10GW 규모 맞춤형 AI 가속기 및 시스템 개발·배포 |
| Anthropic 투자 | - | TPU 확보를 목적으로 Apollo, Blackstone 등이 SPV에 350억달러 Debt 제공.<br>Broadcom은 선순위 채권 300억달러 잔존가치 보증 |

자료: 언론보도, IBK투자증권

## 7. 투자 자금 확보가 목적 : 메타, Oracle, Intel

**표 12. Meta AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| JV 설립 | 270억달러 | 블루아울 캐피탈과 Hyperion Data Center 개발을 위한 JV 구성 |
| JV 설립 | 140억달러 | 블랙록과 Hyperion Data Center 개발을 위한 JV 구성 |
| 채권 발행 | 300억달러 | AI 인프라 확대를 위한 회사채 발행 |
| 채권 발행 | 250억달러 | AI 인프라 투자 확대 일환 |

자료: 언론보도, IBK투자증권

**표 13. Oracle AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| 채권 발행 | 180억달러 | AI Data Center 인프라 확대를 위한 회사채(만기 2030~2065년) 발행 |
| 신디케이트 대출 | 380억달러 | Vantage Data Centers가 개발하는 Stargate 프로젝트 건설 자금조달을 위한 4년 만기 신디케이트 대출 주선 |
| 채권 발행 | 450억달러 | 주요 고객사 계약 수요 충족을 위한 추가 용량 구축을 위해 250억달러 회사채, 200억달러 주식 배분 계약 체결 |
| 전환우선주 발행 | 50억달러 | Series D 전환우선주 발행. AI 인프라 투자 재원 추가 확보 목적 |

자료: 언론보도, IBK투자증권

**표 14. Intel AI 관련 투자 및 자금조달 현황**

| 구분 | 규모 | 내용 |
|---|---|---|
| 채권 발행 | 65억달러 | 아일랜드 반도체 생산공장 지분 매입 자금 조달 |
| 증자 | 200억달러 | AI 인프라 및 파운드리 설비 투자를 위해 200억 달러 규모 유상증자 시행 |

자료: 언론보도, IBK투자증권

## 8. Google 사례는 대규모 Financing 표준이 될 것

AI XPV 구축

2026년 6월 Broadcom은 Apollo, Blackstone과 Anthropic·Open AI 등 frontier AI labs에 2028년까지 20GW 이상의 Computing 자원을 지원하는 AI XPV(XPU+ Networking을 장기 자본과 결합해 AI Labs의 Compute 구축을 지원) 플랫폼을 구축했다.

Compute가 첫 사례. Google TPU 350억 달러. 부채로 인식. 금융권은 이자수익

AI XPV의 첫 번째 거래는 Google TPU(Tensor Processing Unit) 확보를 위한 350억달러 자금 조달이다. TPU는 Google과 Broadcom이 2016년부터 공동 개발한 AI 칩이다. 이번 350억달러는 TPU 구매를 담당하는 SPV(Special Purpose Vehicle, 특수목적기구) 'Compute'에 Debt 형태로 제공된다. Compute는 350억달러를 통해 1GW 규모의 TPU 약 100만 개를 확보했다. TPU는 Fluidstack 기반 Data Center에 2026년 중반 탑재될 예정이다.

Broadcom은 잔존가치 보증

이번 거래에서 Broadcom은 잔존가치 보증을 제공해 SPV의 자금조달을 용이하게 하는 역할을 한다. Anthropic이 TPU 리스비를 지급하지 못하거나 담보인 TPU로 채권 회수가 불가능해질 경우 Broadcom이 채권자의 선순위 트랜치 약 300억달러 손실을 보증한다.

Data Center는 TeraWulf, Hut8이 담당

칩을 수용할 Data Center 구축은 TeraWulf, Hut 8 등 가상화폐 채굴 업체가 맡는다. 채굴 업체들은 전력이 싼 지역에 대규모 부지를 기확보했기 때문에 전력 병목 해소에 유리하다. 2025년 8월 Google은 TeraWulf가 짓는 Lake Mariner 지역 360MW 규모의 Data Center에 리스 지급보증을 약정했다. 임차인은 Fluidstack HPC이며, Fluidstack HPC가 Data Center 리스료를 지급하지 못할 경우 Google이 보증을 하게 된다.

최종 사용자는 Anthropic

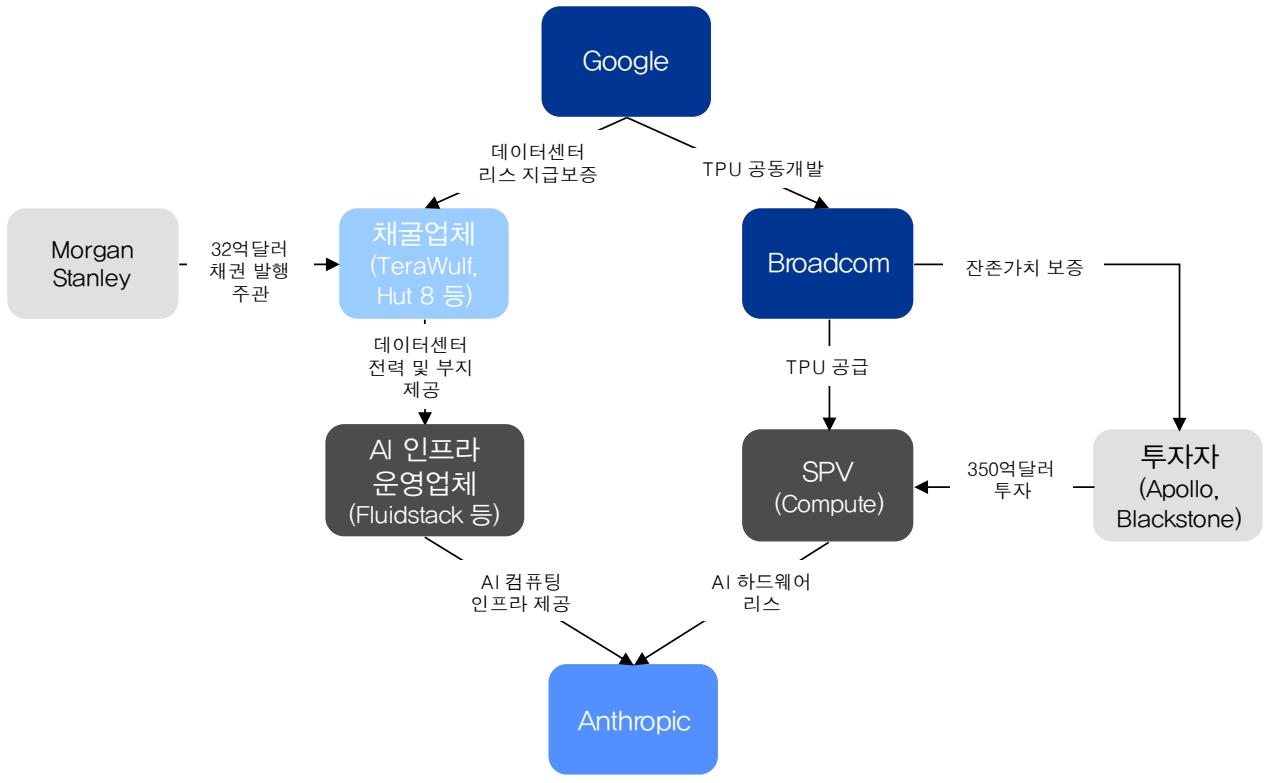
Data Center는 Anthropic向 TPU가 탑재될 예정이다. TeraWulf는 Google의 지급보증을 통한 신용보강으로 Morgan Stanley 주관 하에 32억달러 담보부 채권을 발행했다. 지급보증 제공 대가로 Google은 TeraWulf의 신주인수권을 확보했다. 이러한 방식으로 Google이 지급보증을 제공한 TPU Data Center는 2.4GW에 달하는 것으로 파악된다.

표 15. Google/Broadcom 인프라 금융 참여자 간 이해관계

| 관계 | 참여자 간 이해관계 |
|---|---|
| Broadcom↔Anthropic | Broadcom : Anthropic의 대규모 수요를 기반으로 Google TPU 및 네트워킹 부품의 장기 수요 확보<br>Anthropic : Google TPU 기반 Compute 확보 |
| 금융회사↔SPV | Apollo·Blackstone : SPV에 부채**/사모신용을 제공해 이자수익 확보**<br>SPV : TPU 구매자금 조달. Broadcom의 잔존가치 보증 제공으로 리스크 축소하며 자금조달 용이 |
| Anthropic↔SPV | Anthropic : 350억달러 상당의 칩을 직접 선매입할 필요 없이 리스 방식으로 확보<br>SPV : 부채로 TPU를 확보해 Anthropic에 임대하여 리스 수익으로 부채 상환 |
| Google↔Anthropic | Google : Anthropic의 대규모 TPU 수요를 확보해 **TPU 생태계 확대 및 수익 창출**<br>Anthropic : 대규모 TPU Compute를 확보해 모델 학습·추론에 활용 |
| Google↔채굴업체 | 채굴업체 : 기존 전력·부지 인프라를 AI DC로 전환해 **장기 임대수익이라는 신규 BM** 확보<br>Google : Fluidstack의 의무를 보증해 프로젝트 자금조달 가능성을 높이고, 신주인수권을 확보.<br>직접 **Capex를 전부 부담하지 않고 TPU가 들어갈 외부 Data Center CAPA를 확보** |
| Morgan Stanley↔채굴업체 | Morgan Stanley : TeraWulf의 채권 발행을 주관해 **인수·주선 수수료** 확보<br>TeraWulf : 대규모 DC 건설 Capex 조달. Google의 지급보증이 Fluidstack 신용을 보강해 **자금조달이 용이** |

자료: IBK투자증권

그림 18. Google/Broadcom 인프라 금융 투자 구조도




자료: IBK투자증권

**표 16. AI 밸류체인 내 순환금융 현황 1**

| 시기 | 투자사 | 피투자사 | 규모 | 내용 |
|---|---|---|---|---|
| 2023.05 | Google | Anthropic | – | Google, 시리즈 C 투자 참여 |
| 2023.06 | Microsoft /NVIDIA | Inflection AI | – | Microsoft, NVIDIA, 신규 투자 라운드 투자 |
| 2023.09 | Amazon | Anthropic | 40억달러 | 클라우드 제공 및 AWS Trainium 및 Inferentia 칩 형태 컴퓨팅 인프라 제공 |
| 2024.04 | Microsoft | G42 | 15억달러 | Azure 플랫폼에서 AI 애플리케이션 및 서비스 운영하고 Microsoft와 협력해 AI 솔루션 제공 |
| 2024.11 | Amazon | Anthropic | 40억달러 | AWS를 주요 클라우드 제공 업체로 유지하고 AWS Trainium 및 Inferentia을 통한 모델 학습 |
| 2024.12 | NVIDIA /AMD | xAI | – | NVIDIA, AMD 시리즈 C 투자 참여. NVIDIA Spectrum Ethernet Networking Platform 활용해 Data Center 규모 두 배 확장하고, 20만 개 Hopper GPU 탑재 발표 |
| 2024.12 | AMD | Liquid | 2.5억달러 | AMD, 시리즈 A 투자 참여. AMD Instinct GPU에서 모델 학습 · 배포 발표 |
| 2025.02 | NVIDIA | Lambda | – | NVIDIA, 시리즈 D 투자 참여. 인프라 및 소프트웨어 확장하고 GPU 추가 배포 발표 |
| 2025.08 | Intel | 채권 발행 | 65억달러 | 아일랜드 반도체 생산공장 지분 매입 자금 조달 |
| 2025.09 | NVIDIA | Open AI | 최대 1,000억달러 | Open AI, 차세대 모델 학습 및 실행을 위해 Vera Rubin 등 10GW 규모 NVIDIA 시스템 도입 → LOI 체결 후 2026년 2월 300억 달러 지분투자로 규모 축소 |
| 2025.09 | Oracle | 신디케이트 대출 | 380억달러 | Oracle, Vantage Data Centers가 개발하는 Stargate 프로젝트 건설 자금조달을 위한 4년 만기 신디케이트 대출 주선 |
| 2025.10 | AMD | Open AI | AMD 보통주 최대 1.6억 주 | Open AI, 차세대 AI 인프라 구축을 위해 AMD Instinct GPU 6GW 계약 체결 |
| 2025.10 | Meta /Blue Owl | JV 설립 | 270억달러 | Hyperion Data Center 개발을 위한 JV 구성 |
| 2025.10 | Broadcom | Open AI | – | Open AI가 가속기와 시스템을 설계하고, Broadcom이 10GW 규모 맞춤형 AI 가속기 및 시스템 개발·배포 |
| 2025.10 | Meta | 채권 발행 | 300억달러 | AI 인프라 확대를 위한 회사채 발행 |
| 2025.11 | NVIDIA /Microsoft | Antrhopic | 최대 150억달러 | NVIDIA, Microsoft 전략적 파트너십 일환으로 각 최대 100억, 50억달러 투자 |
| 2026.01 | NVIDIA | CoreWeave | 20억달러 | CoreWeave, 2030년까지 5GW 이상 AII Factory 구축 가속화 지원 |
| 2026.02 | Oracle | 채권 발행 | 450억달러 | 주요 고객사 계약 수요 충족을 위한 추가 용량 구축을 위해 250억달러 회사채, 200억달러 주식 배분 계약 체결 |
| 2026.02 | JPMorgan 등 | 전환우선주 발행 | 50억달러 | Series D 전환우선주 발행. AI 인프라 투자 재원 추가 확보 목적 |
| 2026.02 | Google | 채권 발행 | 약 315억달러 | AI 관련 지출 증가를 위한 100년 만기 채권 발행 |
| 2026.02 | NVIDIA /Amazon | Open AI | 800억달러 | Pre Value 7,300억달러로 1,100억달러(Amazon 500억달러, NVIDIA 300억달러, Softbank 300억달러) 신규 투자 |
| 2026.02 | AMD | Meta | AMD 보통주 최대 1.6억 주 | Meta, 차세대 AI 인프라 구축을 위해 AMD Instinct GPU 6GW 계약 체결 |
| 2026.03 | Amazon | 채권 발행 | 약 538억달러 | 2026년 내 목표 중인 인공지능 인프라 구축 관련 2,000억달러 투자 일환 |
| 2026.03 | NVIDIA | Nebius | 20억달러 | Nebius, 2030년말까지 5GW 이상 용량 구축을 위한 가속 컴퓨팅 플랫폼 도입 지원 |

자료: 언론보도, IBK투자증권

**표 17. AI 밸류체인 내 순환금융 현황 2**

| 시기 | 투자사 | 피투자사 | 규모 | 내용 |
|---|---|---|---|---|
| 2026.04 | Amazon | Anthropic | 50억달러, 추가 최대 200억달러 | Anthropic, Trainium 3 포함 최대 5GW 용량 확보 예정 및 향후 10년간 AWS 기술에 1,000억달러 이상 투자 발표 |
| 2026.04 | Meta | 채권 발행 | 250억달러 | AI 인프라 투자 확대 일환 |
| 2026.05 | NVIDIA | IREN | 최대 21억달러 | NVIDIA, IREN 주식 21억달러 규모 주식매입권 확보하고, DSX 기반 최대 5GW 규모 AI 인프라 구축 지원 |
| 2026.06 | Google | 증자 | 847.5억달러 | AI 인프라 강화를 위해 기존 800억달러 조달 계획에서 조달 금액 상향 조정 |
| 2026.06 | AMD | TensorWave | 3.5억달러 | AMD Instinct MI355X GPU 클러스터 기반 AI 인프라 확장 |
| 2026.06 | NVIDIA | Sharon AI | – | 최대 4만 개 GB300 기반 72MW Data Center 구축 협력. NVIDIA가 수익 공유 및 Credit 지원 모델 제공 |
| 2026.06 | Micron | Anthropic | – | Micron의 메모리 및 스토리지 공급 및 공동 설계 추진 |
| 2026.06 | Broadcom | Anthropic | – | TPU 확보를 목적으로 Apollo, Blackstone 등이 SPV에 투자하는 350억달러 Debt 제공. Broadcom이 선순위 채권 300억달러 잔존가치 보증 |
| 2026.07 | AMD | Anthropic | 최대 50억달러 | Anthropic, 27년 상반기부터 AMD MI450 GPU 2GW 규모 구매 계획 |
| 2026.07 | Amazon | – | 최소 250억달러 | AI 투자를 위한 자금 조달 |
| 2026.07 | NVIDIA | Naver | 최대 10억달러 | Naver, 2028년까지 DSX AI Factory 설비를 55MW에서 200MW로 확장하고 NVIDIA AI 인프라 구축을 1GW까지 확장 발표 |
| 2026.07 | NVIDIA | Hut 8 | 최대 500억달러 | NVIDIA, Hut 8이 건설 중인 1GW 규모 텍사스 Data Center 임대 계약 체결. NVIDIA는 네오클라우드에 재임대 계획 |
| 2026.07 | Microsoft | Mistral | 수십억달러 | Mistral의 유럽 기반 GPU 활용해 클라우드 및 AI 서비스 제공 |
| 2026.07 | Meta /BlackRock | – | 140억달러 | Hyperion Data Center 개발을 위한 JV 구성 |
| 2026.08 | Google | – | 250억달러 | AI 인프라 투자 지원을 위한 채권 발행 |
| 2026.08 | Intel | 유상증자 | 200억달러 | AI 인프라 및 파운드리 설비 투자를 위해 200억 달러 규모 유상증자 시행 |
| 2026.08 | NVIDIA | SB Energy | 1,050억달러 | 최대 8GW 규모 Ohio Data Center 구축 관련 Open AI 임대료 등 1,050억달러 지급보증 |
| 2026.08 | NVIDIA | SB Energy | 15억달러 | Ohio Data Center 구축을 위한 투자 |
| 논의 중 | Google | Anthropic | – | Nexus Data Centers의 Hubbard Campus에 대한 Anthropic 임차료 등 지급보증 |
| 논의 중 | NVIDIA | Open AI | 3,500억달러 | Open AI의 칩 구매 자금 조달 논의 |
| 논의 중 | NVIDIA | Lancium | 최대 30억달러 | 20억달러를 투자해 전력 인프라 업체 Lancium의 지분 20% 우선 확보 후 추가 투자 계획 |
| 논의 중 | NVIDIA | Open AI 등 투자 | 5,000억달러 | Apollo, Blackstone 등과 협력해 개별 vehicle로 나눠 5,000억달러 규모 자금 조성 논의 |

자료: 언론보도, IBK투자증권

## 수요 없는 공급 확대는 없다

메이커별 생산 계획은 이전 계획보다는 앞당겨지는 추세

메모리 업체들의 공급 계획이 빠르게 수정되고 있다. 국내 업체들은 다소 진행이 느릴 것으로 예정되었던 용인 반도체 클러스터 완공 시기가 각각 예정 대비 빨리 진행되고, 기존 증설 계획도 완공 시점을 앞당기고 있다.

2028년까지 메모리 메이커들의 생산능력은

1) 삼성전자 P4가 완전 가동되어 100K/월에 이를 것으로 예상하고, P5 라인 40K/월 추가되고,

2) SK하이닉스는 용인 Fab1 가동으로 2028년말까지 100K/월까지 확대되고,

3) Micron은 ID1을 최대 80K/월 추가할 것으로 예상한다.

투자는 수요를 기반으로 진행

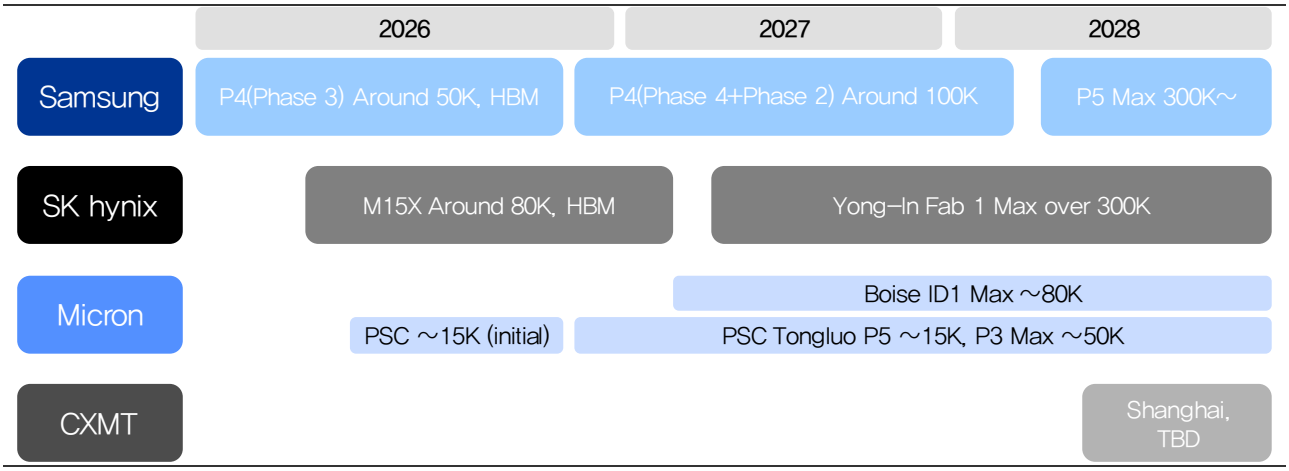
메이커들의 생산량 확대에서 가장 중요한 변수는 수요이다. 수요가 확보되지 않은 상황에서 집행되는 투자는 없다. 그리고 반도체 투자에서 주목해야 할 부분은 클린룸 완공 및 장비 반입, 양산까지 1년 이상의 차이가 있다는 것이다.

클린룸을 확보한 이후에도 15개월이 지나야 완전 가동

부지 확보 이후 건물을 짓고, 클린룸을 완공하기까지가 기초이고, 장비 반입에 소요되는 시간이 3개월 내외, 장비 테스트 및 조정하는데 2개월이 소요된다. 이후 Wafer를 투입하고 최종 제품을 확보하는데 걸리는 기간 4개월이 추가된다. 완성된 제품으로 고객사 품질 테스트를 통과할 때까지 걸리는 시간을 포함하면 수개월이 더 추가되기도 한다.

Full Capacity까지 걸리는 기간은 신공정일 경우 15개월까지 소요될 수 있다. 물량 확대가 단기간 내에 어려운 이유이다.

그림 19. 2026~2028 주요 메모리 업체 증설 전망




자료: OMDIA, IBK투자증권

표 18. 글로벌 DRAM fab CAPA 전망 (단위: K/month)

| | 1Q25 | 2Q25 | 3Q25 | 4Q25 | 1Q26 | 2Q26 | 3Q26E | 4Q26E | 1Q27E |
|---|---|---|---|---|---|---|---|---|---|
| **Samsung** | **610** | **630** | **650** | **675** | **675** | **680** | **675** | **695** | **685** |
| 11L | – | – | – | – | – | – | – | – | – |
| 12L | – | – | – | – | – | – | – | – | – |
| 13L | – | – | – | – | – | – | – | – | – |
| 15L | 150 | 155 | 160 | 160 | 160 | 160 | 155 | 155 | 150 |
| 16L | 70 | 70 | 75 | 75 | 75 | 70 | 70 | 70 | 65 |
| 17L | 95 | 95 | 95 | 95 | 95 | 95 | 95 | 95 | 95 |
| P1(Pyeongtaek) | 90 | 95 | 100 | 100 | 90 | 90 | 90 | 90 | 90 |
| P2(Pyeongtaek) | 105 | 105 | 105 | 105 | 105 | 105 | 100 | 100 | 95 |
| P3(Pyeongtaek) | 100 | 110 | 110 | 110 | 110 | 110 | 105 | 105 | 105 |
| P4(Pyeongtaek) | – | – | 5 | 30 | 40 | 50 | 60 | 80 | 85 |
| **SK hynix** | **465** | **500** | **525** | **530** | **530** | **545** | **565** | **580** | **593** |
| M10 | – | – | – | – | – | – | – | – | – |
| HC2 | 190 | 190 | 190 | 190 | 190 | 190 | 190 | 190 | 190 |
| M14 | 165 | 160 | 155 | 150 | 150 | 160 | 160 | 160 | 160 |
| M16 | 110 | 150 | 180 | 190 | 190 | 190 | 190 | 190 | 190 |
| M15X | | | | | | 5 | 25 | 40 | 50 |
| **Micron** | **300** | **300** | **300** | **300** | **300** | **300** | **305** | **315** | **315** |
| Manassas | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| Inotera | 88 | 88 | 88 | 88 | 88 | 88 | 88 | 88 | 88 |
| MMJ (Elpida) | 88 | 88 | 88 | 88 | 88 | 88 | 88 | 88 | 88 |
| MMT (Rexchip) | 97 | 97 | 97 | 97 | 97 | 97 | 97 | 97 | 97 |
| **PSC** | | | | | | | **5** | **15** | **15** |
| **Nanya** | **55** | **55** | **60** | **60** | **60** | **60** | **63** | **63** | **63** |
| **PSC** | **30** | **33** | **35** | **35** | **35** | **35** | **35** | **35** | **35** |
| **Winbond** | **25** | **25** | **27** | **28** | **28** | **29** | **35** | **35** | **35** |
| **CXMT** | **200** | **240** | **270** | **290** | **295** | **300** | **300** | **300** | **300** |
| **Total** | **1,685** | **1,783** | **1,867** | **1,918** | **1,923** | **1,949** | **1,983** | **2,038** | **2,041** |

자료: OMDIA, IBK투자증권

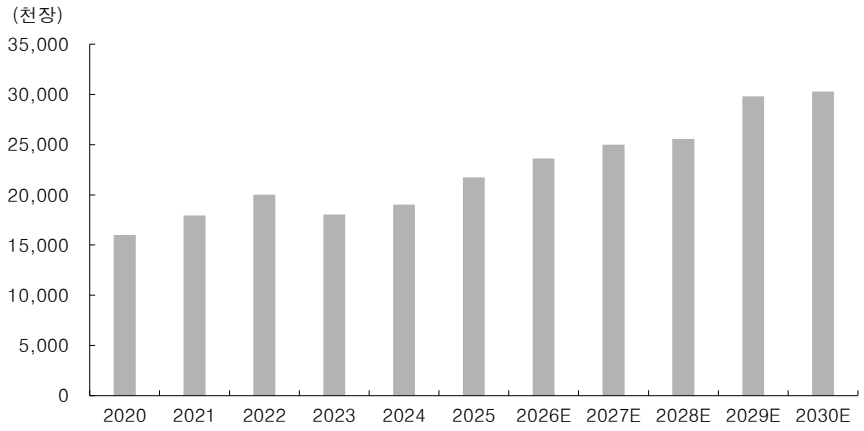
27년 114K/월
28년 49K/월 증가

OMDIA에 따르면 부지런히 라인을 증설해도 2027년 전 세계 DRAM Wafer는 114K/월 증가하고, 2028년에는 49K/월 증가할 것으로 전망된다. 시장 수요에 충분히 대응하기 어려울 전망이다.

29년 353K/월 증가로 전망

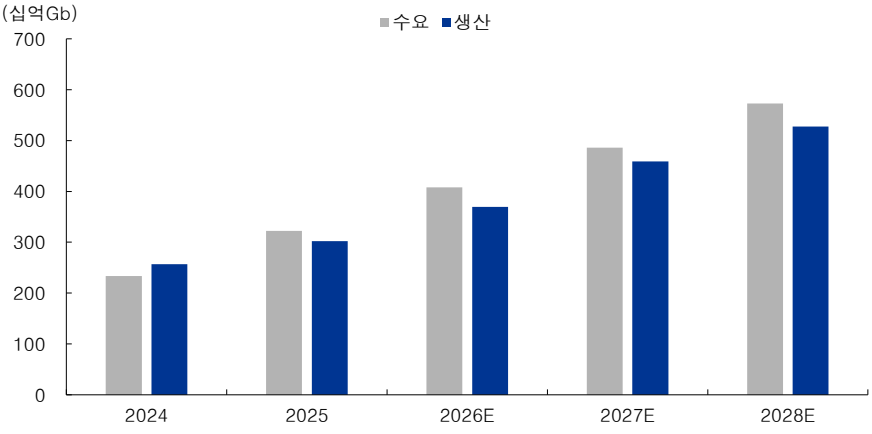
2029년에는 353K/월 증가할 것으로 예상되지만 유동적일 전망이다. 이 물량은 2028년 Wafer 생산량의 16.6% 수준이다. 이 물량도 공급 과잉을 불러올 수준은 아닌 것으로 판단한다.

**그림 20. 주요 메모리 업체의 Wafer 생산 추이**




자료: OMDIA, IBK투자증권

**그림 21. DRAM 수요 공급 추이**




자료: OMDIA, IBK투자증권

## 중국의 물량이 위험하다?

CXMT 위협은 없다

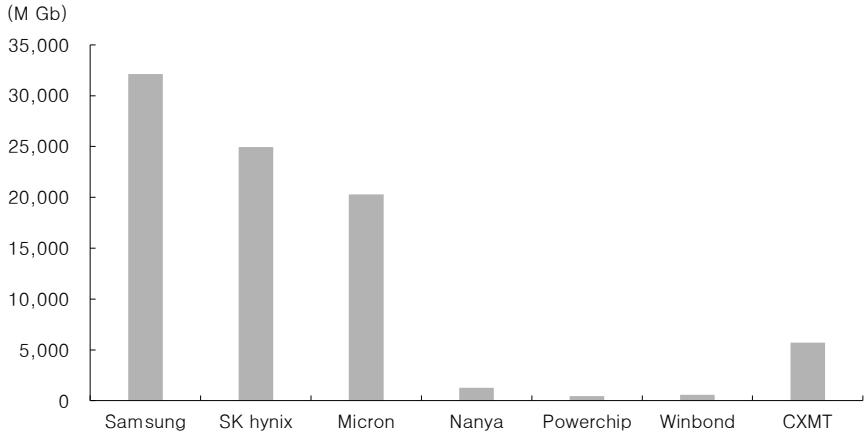
최근 주가하락을 야기한 변수 중의 하나는 CXMT가 상장을 통해서 확보한 자금으로 Wafer를 증설하면서 공급 부족 상황이 해소될 수 있다는 가정이다. 결론부터 말하자면 그 가능성은 아주 낮은 것으로 판단한다. 이유는 Wafer CAPA와 생산은 정비례하지 않기 때문이다. 이는 현재 CXMT의 여러 지표를 통해서 확인할 수 있다.

26년 1분기 출하 CXMT 점유율 6.7%

26년 1분기 기준 CXMT의 출하량 기준 점유율은 6.7%이다. 삼성전자가 37.6%, SK하이닉스 29.2%, Micron 23.8%이다. Micron과 4배 가까이 차이가 난다. 메이저 3사 중에서 Micron과 비교하는 것은 현재 Wafer 기준 생산 능력이 유사하기 때문이다.

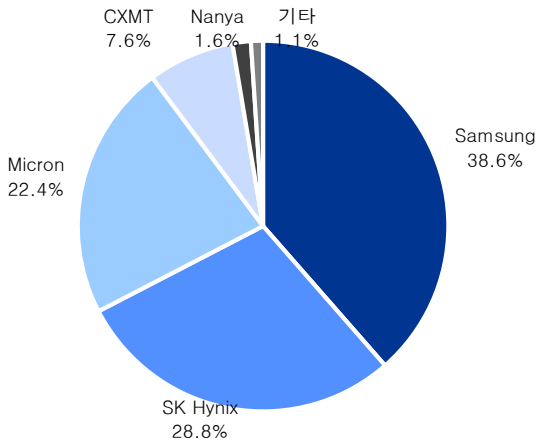
26년 1분기 매출액 기준 시장 점유율도 출하량과 크게 차이가 나지 않는다. CXMT 시장 점유율은 7.6%이다.

그림 22. 26년 1분기 주요 업체 DRAM 생산량




자료: OMDIA, IBK투자증권

그림 23. 26년 1분기 DRAM 매출액 점유율




자료: OMDIA, IBK투자증권

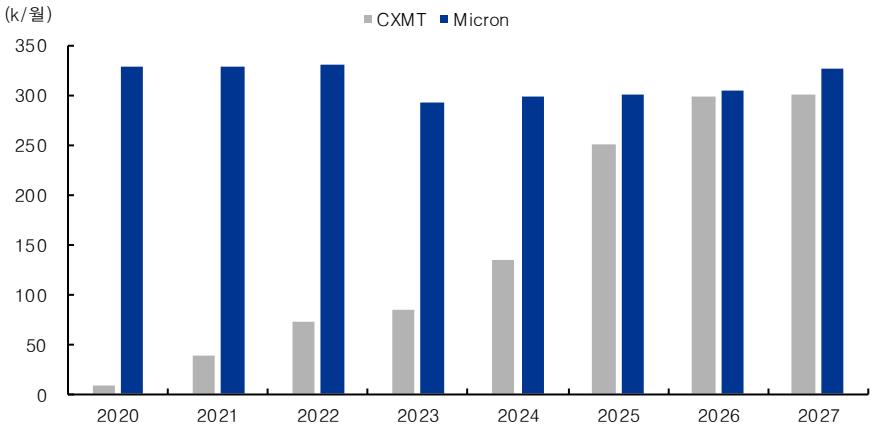
## 1. CXMT Wafer 생산 능력은 이미 Micron 수준

CXMT Wafer 생산능력은 Micron 수준

26년 2분기 기준 CXMT의 Wafer 생산능력은 300K/월이다. 이는 Micron과 동일한 수준이다. 2024년 규모의 2배 수준으로 증설 속도는 무척 빠르게 진행되었다. 다만 매출액, 출하량의 시장 점유율은 Wafer 생산 능력 대비 현저히 낮은 수준이다.

이에 대한 원인은 낮은 공정 기술, 낮은 Gb/Wafer, 모바일/PC 중심의 제품믹스, 높은 DDR4 비중 등이다.

그림 24. Micron/CXMT Wafer 생산능력 비교



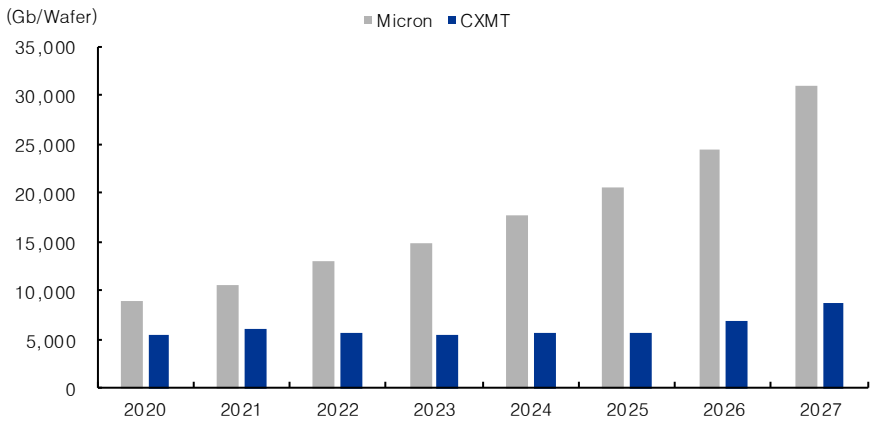

자료: OMDIA, IBK투자증권

## 2. Wafer 1장당 생산 Bit은 Micron의 1/4 수준

CXMT Gb/W는 Micron의 1/4

기술의 완성도 및 효율성은 Wafer 1장으로 얼마나 많은 Bit을 생산하느냐가 중요하다. 2025년 기준 Micron은 Wafer 1장에서 20,612Gb을 생산하지만, CXMT는 5,746Gb을 생산한다. 4배 차이이다. Wafer 증가가 출하량 증가를 의미하지 않는 뚜렷한 지표이다.

그림 25. Micron과 CXMT Gb/Wafer 추이




자료: OMDIA, IBK투자증권

## 3. 공정 기술 간극이 너무 크다

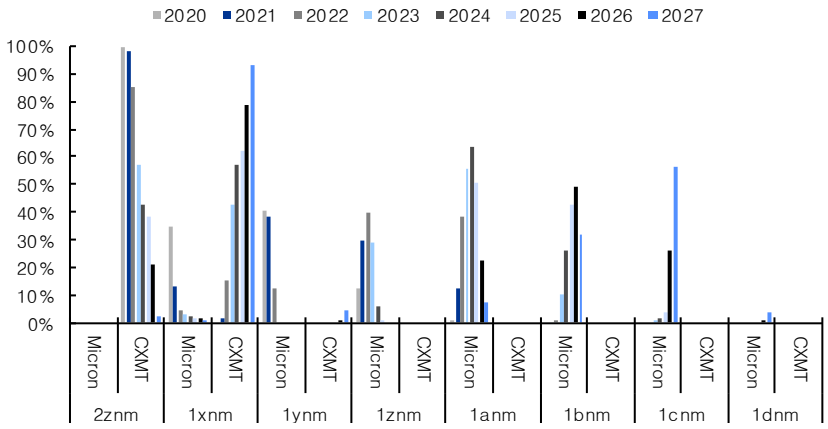
메이저 주력 공정은 1b nm에서 1c nm로 전환 중

DRAM 메이커들은 선단공정의 세대 진화를 통해서 전력, 속도 등의 성능을 개선한다. 지금 선도업체들의 주력 공정은 1b nm이다. 2027년 말에는 1c nm가 주력 공정으로 전환될 것으로 전망한다. 업체별로 비중 차이가 있지만 1c nm가 주력이 될 것으로 보는데는 이견이 없다.

CXMT 주력 공정은 5세대 뒤진 1x nm

CXMT의 주력 공정은 1x nm이다. 2027년말까지도 유지될 전망이다. 세대는 1x nm, 1y nm, 1z nm, 1a nm, 1b nm, 1c nm 순으로 진화한다. 세대로 따지면 5세대 차이가 벌어져 있다. 전력, 칩 사이즈, 스피드 모든 부분에서 경쟁사 대비 열위에 있다.

그림 26. Micron CXMT 선단 공정 비교



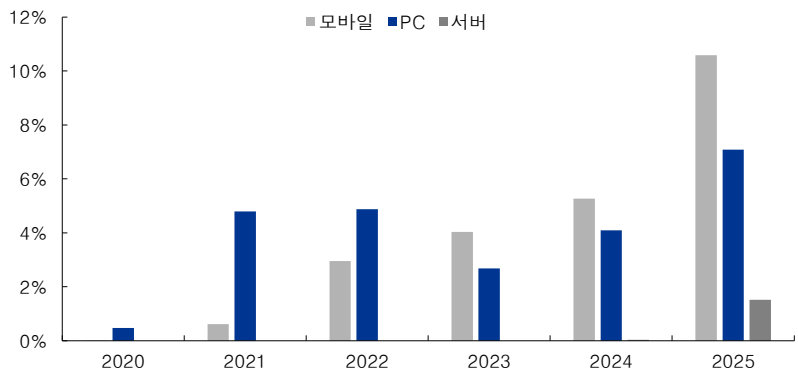

자료: OMDIA, IBK투자증권

## 4. 서버 시장 점유율은 2%

CXMT 서버 시장 점유율 2% 서버는 시장의 50.8%

DRAM을 사용하는 주요 어플리케이션은 서버, 모바일, PC이다. 전체 시장 중 서버 규모는 50.8% 수준이다. CXMT는 낮은 성능 제품 비중이 높아서 서버 시장 점유율은 2%에 불과하다. 모바일은 중국 내수 시장 공략으로 11%, PC는 7%이다. 서버에서 요구하는 성능이 가장 높기 때문이다.

그림 27. CXMT 어플리케이션별 시장 점유율




자료: OMDIA, IBK투자증권

## 5. 주력 제품은 모바일, DDR5는 9.4%

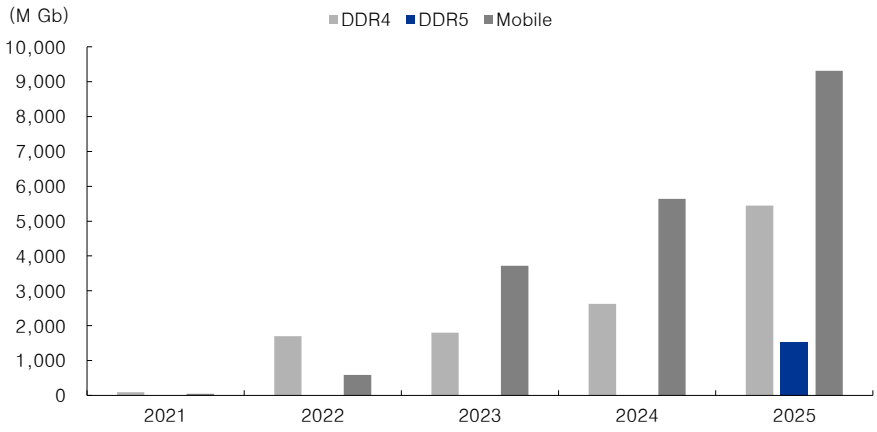
시장 평균 대비 편중된 제품별 비중

2025년말 출하 기준 CXMT의 주력 제품은 모바일이다. 비중은 57.2%이고 DDR4가 33.4%, DDR5는 9.4%이다. DDR5 매출은 2025년에 처음 발생했다.

Micron은 DDR4 15.8%, DDR5 42.1%, 모바일 32.5%, 그래픽 9.2%이다.

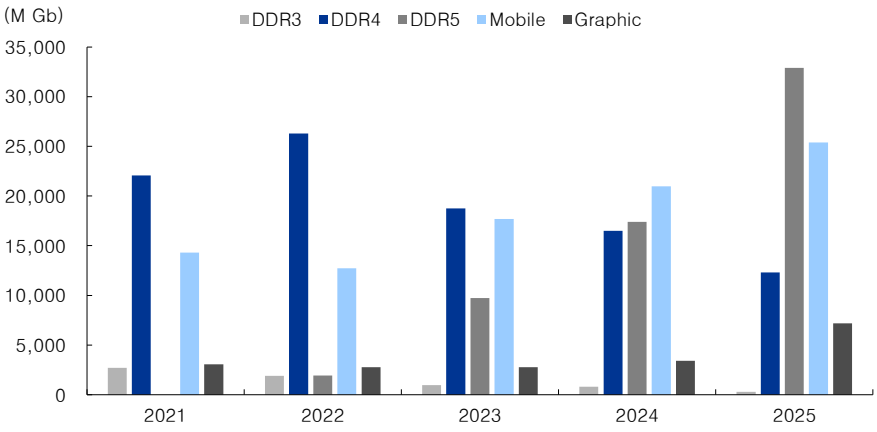
DRAM 시장은 DDR4가 14.9%, DDR5 39.3%, 모바일이 34.4%, 그래픽이 10.1%이다.

그림 28. CXMT 제품별 출하 추이




자료: OMDIA, IBK투자증권

그림 29. Micron 제품별 출하 추이




자료: OMDIA, IBK투자증권

## 6. Wafer가 늘어나도 서버 시장 진입은 어려울 것

증설 목표는 현재 대비 2배 이상

CXMT는 IPO 당시 확보한 자금 12.5조원 중 1.6조원을 Wafer 증설에 투자하겠다고 발표한 것과 달리 최근 베이징 인근 지역에 증설 계획 뉴스가 공표되고 있다. IPO로 확보한 1.6조원 포함 공모 당시 초과조달금 약 6조원 중 일부와 지방정부 금융 등 추가 조달을 통해 진행하는 것으로 파악된다. Wafer 증설이 출하량 증가, 제품믹스 개선, 선단 공정 발전을 의미하는 것은 아니다.

DRAM 시장을 주도하는 어플리케이션은 AI 서버 시장이다. 서버에 공급되는 DRAM은 가장 높은 성능이 요구된다. 공정 기술, 품질 안정성, 안정적 공급 능력이 모두 동반되어야 시장에 진입할 수 있다.

뚜렷한 한계

CXMT의 한계는

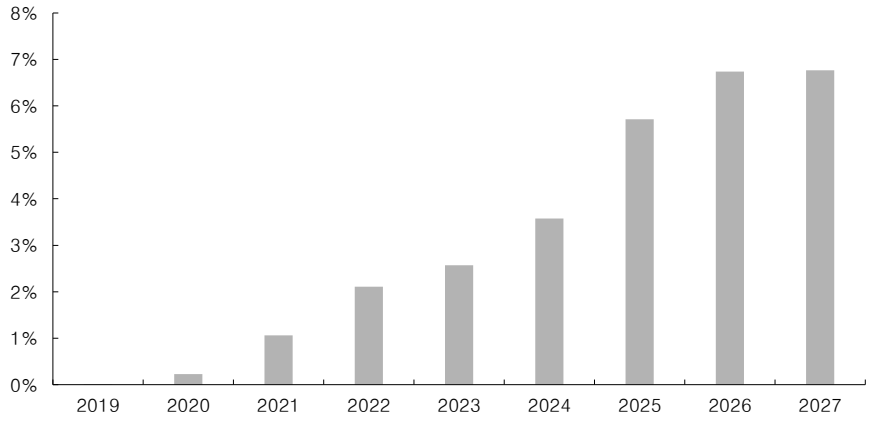
1. 선단 공정이 5세대 뒤처져 있고,
2. Gb/Wafer의 성장 속도가 느리고,
3. DDR5 전환 속도가 느린 것이다.

CAPA보다 R&D가 더 급하다

확보된 자금의 상당 부분을 R&D 개발에 사용하겠다는 것은 선도업체 대비 기술력의 차이가 크다는 의미이다. 지금까지 선도업체들이 발전해 온 과정을 고려하면 후발업체가 동일한 수준으로 기술 격차를 줄이기는 쉽지 않을 것으로 판단한다.

이를 고려해서 OMDIA는 2027년까지 CXMT 시장 점유율이 7%를 넘기기 어려울 것으로 전망하고 있다.

**그림 30. CXMT 생산량 기준 시장 점유율 추이 및 전망**




자료: OMDIA, IBK투자증권

Compliance Notice

동 자료에 게재된 내용들은 외부의 압력이나 부당한 간섭없이 본인의 의견을 정확하게 반영하여 작성되었음을 확인합니다.
동 자료는 기관투자가 또는 제3자에게 사전 제공한 사실이 없습니다.
동 자료는 조사분석자료 작성에 참여한 외부인(계열회사 및 그 임직원등)이 없습니다.
조사분석 담당자 및 배우자는 해당종목과 재산적 이해관계가 없습니다.
동 자료에 언급된 종목의 지분율 1%이상 보유하고 있지 않습니다.
당사는 상기 명시한 사항 외 고지해야 하는 특별한 이해관계가 없습니다.

## Note

## Note

# IBKS Research Center

| 성명 | 직급 | 담당업종 | 전화 | 이메일 |
|---|---|---|---|---|
| 용대인 | 전무(부문장) | 총괄 | 6915-5400 | daeinyong@ibks.com |
| 이승훈 | 상무대우(본부장) | AI/인터넷/게임 | 6915-5680 | dozed@ibks.com |
| **투자분석부** | | | | |
| 번준호 | 연구위원 | Strategy | 6915-5670 | ymaezono@ibks.com |
| 정용택 | 수석 Economist | Economy | 6915-5701 | ytjeong0815@ibks.com |
| 정형주 | 연구위원 | 채권/크레딧 | 6915-5654 | hj.jeong@ibks.com |
| 김예슬 | 연구원 | 글로벌 전략 | 6915-5526 | chchu93@ibks.com |
| **기간산업분석부** | | | | |
| 이동욱 | 연구위원 | 에너지/소재 | 6915-5671 | treestump@ibks.com |
| 남성현 | 연구위원 | 유통·식자재/지주 | 6915-5672 | rockrole@ibks.com |
| 이현욱 | 연구원 | 자동차/2차전지 | 6915-5659 | hwle1125@ibks.com |
| 장성호 | 연구원 | 조선/기계/운송 | 6915-5661 | cpszoo9080@ibks.com |
| **혁신기업분석부** | | | | |
| 김운호 | 연구위원 | IT/반도체 | 6915-5656 | unokim88@ibks.com |
| 김태현 | 연구위원 | 유틸리티/통신/방산 | 6915-5658 | kith0923@ibks.com |
| 조경진 | 연구위원 | 화장품 | 6915-5464 | ckjins@ibks.com |
| 조정현 | 연구원 | 건설/부동산 | 6915-5660 | controlh@ibks.com |
| **코스닥리서치센터** | | | | |
| 이건재 | 연구위원 | 코스닥 리서치 | 6915-5676 | geonjaelee83@ibks.com |
| 조은애 | 연구위원 | 코스닥 헬스케어 | 6915-5689 | goodkid@ibks.com |
| 정이수 | 연구위원 | 코스닥 제약/바이오 | 6915-5677 | ysjeong306@ibks.com |
| 강민구 | 연구원 | 코스닥 반도체/장비 | 6915-5473 | kmg@ibks.com |
| 김혜빈 | 연구원 | 코스닥 로봇 | 6915-5669 | hyebhinkim@ibks.com |
| 유창근 | 연구원 | 코스닥 스몰캡 | 6915-5655 | changsic12@ibks.com |

신뢰를 바탕으로 함께 성장 하는 IBK투자증권

IBK기업은행 금융그룹
IBK투자증권

서울특별시 영등포구 여의도동 국제금융로 6길 11
대표번호 02-6915-5000
고객지원부 1588-0030, 1544-0050

| | | | |
|---|---|---|---|
| IBKS Family Office | 02) 536-4070 | IBK WM센터 대구 | 053) 752-3535 |
| 영업부 | 02) 6915-2626 | IBK WM센터 광주 | 062) 382-6611 |
| 강남센터 | 02) 2051-5858 | IBK WM센터 일산 | 031) 904-3450 |
| 강남역 금융센터 | 02) 532-0210 | IBK WM센터 판교 | 031) 724-2630 |
| 분당센터 | 031) 705-3600 | IBK WM센터 평촌 | 031) 476-1020 |
| IBK WM센터 강남센트럴 | 02) 556-4999 | IBK WM센터 천안 | 041) 569-8130 |
| IBK WM센터 목동 | 02) 2062-3002 | IBK WM센터 부산 | 051) 741-8810 |
| IBK WM센터 도곡 | 02) 2057-9300 | IBK WM센터 창원 | 055) 282-1650 |
| IBK WM센터 한남동 | 02) 796-8500 | IBK WM센터 울산 | 052) 271-3050 |
| IBK WM센터 중계동 | 02) 948-0270 | IBK WM센터 시화공단 | 031) 498-7900 |
| IBK WM센터 반포자이 | 02) 3481-6900 | IBK WM센터 남동산단 | 032) 822-6200 |
| IBK WM센터 동부이촌동 | 02) 798-1030 | | |