

# 에이전틱 AI(Agentic AI)와 피지컬 AI(Physical AI)의 부상이 DRAM 메모리 슈퍼사이클의 진폭과 지속성에 미치는 구조적 영향 분석

## 1. 서론: 반도체 메모리 시장의 구조적 변곡점

### 1.1 슈퍼사이클의 재정의: 순환적 호황에서 구조적 결핍으로

2025년을 기점으로 글로벌 반도체 메모리 산업은 과거 2017년에서 2018년 사이에 경험했던 '슈퍼사이클(Supercycle)'과는 질적으로 다른 국면에 진입하고 있다. 과거의 슈퍼사이클이 클라우드 서비스 제공업체(Hyperscalers)들의 데이터센터 증설이라는 단일 수요처의 양적 팽창에 기인했다면, 현재 도래하고 있는 사이클은 인공지능(AI) 기술의 패러다임 전환이 촉발하는 '질적 고도화'와 공급망의 '물리적 제약'이 맞물린 복합적인 현상이다. 특히, 단순히 텍스트와 이미지를 생성하는 수준을 넘어 자율적으로 의사결정을 수행하는 \*\*에이전틱 AI(Agentic AI)\*\*와, 가상공간의 지능을 물리적 세계로 확장하는 \*\*피지컬 AI(Physical AI)\*\*의 동시 부상은 메모리 반도체에 요구되는 성능의 임계치를 기존의 상식 이상으로 끌어올리고 있다.

본 보고서는 이러한 두 가지 차세대 AI 기술의 부상이 DRAM 산업의 수급 밸런스를 어떻게 붕괴시키며, 그 결과로 나타날 메모리 슈퍼사이클의 \*\*진폭(Amplitude)\*\*과 \*\*지속성(Duration)\*\*이 과거와 어떻게 차별화되는지를 심층 분석한다. 분석에 따르면, 에이전틱 AI에 의한 추론(Inference) 워크로드의 폭발과 피지컬 AI에 의한 엣지(Edge) 디바이스의 서버화(Serverization)는 메모리 수요를 다층화시키고 있으며, 이는 HBM(High Bandwidth Memory) 생산에 따른 웨이퍼 잠식 효과(Wafer Penalty)와 결합하여 2027년, 길게는 2028년까지 이어지는 역사상 가장 강력하고 긴 호황기를 예고하고 있다.<sup>1</sup>

## 1.2 기술적 동인: AI의 진화 단계와 메모리 요구사항

AI 기술은 현재 '반응형(Reactive)'에서 '능동형(Proactive)'으로, 그리고 '소프트웨어'에서 '하드웨어'로 그 외연을 급격히 확장하고 있다. 생성형 AI(Generative AI)가 사용자의 프롬프트에 수동적으로 응답하는 과정에서 막대한 데이터를 학습(Training)하는 데 집중했다면, 에이전틱 AI는 실시간으로 외부 환경과 상호작용하며 끊임없이 추론하고 기억을 유지해야 하는 과제를 안고 있다.<sup>3</sup> 이는 메모리 반도체의 역할을 단순한 데이터 저장소에서 시스템의 지능을 유지하는 '맥락의 저장소(Context Store)'로 격상시킨다.

동시에, 휴머노이드 로봇으로 대표되는 피지컬 AI는 데이터센터에 집중되었던 고성능 컴퓨팅 파워를 엣지단으로 분산시키며, LPDDR(Low Power DDR)과 같은 모바일 메모리의 성능 혁신과 용량 증설을 강제하고 있다. 테슬라의 옵티머스(Optimus)나 엔비디아의 프로젝트 GROOT와 같은 이니셔티브는 로봇 하나하나가 걸어 다니는 고성능 서버가 되어야 함을 의미하며, 이는 전례 없는 규모의 신규 메모리 시장(TAM, Total Addressable Market)을 창출하고 있다.<sup>5</sup>

---

## 2. 에이전틱 AI(Agentic AI): '상태 유지(Statefulness)'와 메모리 아키텍처의 혁명

### 2.1 생성형 AI의 한계와 에이전틱 AI의 부상

현재의 대규모 언어 모델(LLM) 기반 생성형 AI는 근본적으로 'Stateless(상태 비저장)' 아키텍처에 가깝다. 사용자가 질문을 입력하면, 모델은 학습된 가중치(Weights)를 바탕으로 확률적인 답변을 생성한 후 해당 상호작용을 종료한다. 비록 이전 대화 내용을 일부 참조하지만, 이는 제한적인 컨텍스트 윈도우(Context Window) 내에서 이루어지는 단기 기억에 불과하다.

반면, 에이전틱 AI는 명확한 목표(Goal)를 가지고, 이를 달성하기 위해 장기간에 걸쳐 일련의 작업을 자율적으로 수행한다. IBM과 주요 연구 기관들의 정의에 따르면, 에이전틱 AI는 ▲목표 설정 및 계획(Planning) ▲도구 사용(Tool Use) ▲반성 및 수정(Reflection) ▲기억 유지(Memory)라는 네 가지 핵심 기능을 수행한다.<sup>3</sup>

이러한 특성은 메모리 시스템에 완전히 새로운 부하를 가중시킨다. 에이전트는 복잡한 과업을 수행하기 위해 수십, 수백 단계의 중간 추론 과정을 거치는데, 이 과정에서 발생하는 모든

사고의 흐름(Chain of Thought)과 외부 도구(API, 브라우저 등)의 실행 결과를 실시간으로 저장하고 참조해야 한다. 즉, 에이전틱 AI는 'Stateful(상태 저장)' 시스템이며, 이는 메모리 용량(Capacity)에 대한 요구를 기하급수적으로 증가시키는 요인이다.<sup>8</sup>

2.2 KV-Cache 폭발과 '메모리 월(Memory Wall)'의 심화

에이전틱 AI의 핵심인 '긴 문맥(Long Context)' 처리는 기술적으로 \*\*KV-Cache(Key-Value Cache)\*\*의 크기 문제로 귀결된다. 트랜스포머(Transformer) 아키텍처에서 모델이 긴 문맥을 이해하기 위해서는 이전에 처리한 토큰들의 정보를 KV-Cache 형태로 메모리에 상주시켜야 한다.

삼성전자의 기술 분석에 따르면, LLaMA-3.1 70B 모델을 기준으로 16K 길이의 컨텍스트 윈도우를 처리할 때, 요청(Request) 수가 증가함에 따라 KV-Cache가 차지하는 메모리 용량은 선형적이 아니라 폭발적으로 증가한다.<sup>10</sup> 예를 들어, 다수의 에이전트가 동시에 복잡한 추론을 수행할 경우, 단일 GPU에 탑재된 80GB 또는 144GB의 HBM(High Bandwidth Memory) 용량은 순식간에 포화 상태에 이르게 된다.

표 1: 생성형 AI와 에이전틱 AI의 메모리 워크로드 특성 비교

| 특성     | 생성형 AI<br>(Generative AI)     | 에이전틱 AI<br>(Agentic AI)           | 메모리 영향<br>(Memory Impact)                         |
|--------|-------------------------------|-----------------------------------|---|
| 운영 모드  | 반응형 (Reactive)                | 능동형/자율형<br>(Proactive/Autonomous) | 지속적인<br>백그라운드 연산 및<br>메모리 점유                      |
| 트래픽 패턴 | 버스트(Burst)<br>트래픽             | 지속적(Persistent)<br>트래픽, 잦은 호출     | 메모리 대역폭 상시<br>점유율 상승                              |
| 데이터 상태 | Stateless (세션<br>종료 시 휘발)     | Stateful (장기 기억<br>및 상태 유지 필수)    | 저장<br>공간(Capacity) 및<br>지속성(Persistence)<br>요구 급증 |
| 핵심 병목  | 연산<br>능력(Compute) 및<br>학습 대역폭 | 메모리<br>용량(Capacity) 및<br>추론 대역폭   | HBM 부족 심화 및<br>대용량 DDR/CXL<br>수요 촉발               |

|       |                               |                                   |  |
|-------|-------------------------------|-----------------------------------|--|
| 주요 연산 | 행렬 곱셈 (Matrix Multiplication) | 검색(Retrieval),<br>벡터 연산, 논리<br>추론 | 벡터 DB 및<br>RAG(검색 증강<br>생성)용 메모리<br>필요 |
|-------|-------------------------------|-----------------------------------|--|

자료: IBM <sup>3</sup>, Aalpha <sup>11</sup>, F5 <sup>12</sup>, 삼성전자 <sup>10</sup> 데이터를 바탕으로 재구성.

이러한 KV-Cache의 폭발은 기존의 GPU 중심 아키텍처가 가진 한계를 적나라하게 드러낸다. UC 버클리의 RISELab 연구에 따르면, 트랜스포머 모델의 파라미터 수는 2년마다 410배 증가하는 반면, GPU 메모리 용량은 동기간 2배 증가에 그치고 있다.<sup>13</sup> 에이전틱 AI 시대에 이르러, 이 격차는 단순한 성능 저하가 아닌 시스템의 '동작 불능'을 야기할 수 있는 치명적인 병목이 되고 있다.

## 2.3 이종 메모리 시스템(Heterogeneous Memory System)과 CXL의 부상

HBM의 용량 한계와 높은 비용 문제를 해결하기 위해, 업계는 \*\*이종 메모리 아키텍처(Heterogeneous Memory Architecture)\*\*로의 전환을 서두르고 있다. 이는 계층화된(Tiered) 메모리 시스템을 구축하여, 데이터의 접근 빈도와 중요도에 따라 서로 다른 종류의 메모리에 데이터를 배치하는 전략이다.

### 2.3.1 Lake Tahoe 프로젝트와 CXL 메모리 풀링

삼성전자, Montage Technology, H3 Platform이 협력하여 OCP(Open Compute Project)에서 공개한 'Lake Tahoe' 프로젝트는 에이전틱 AI를 위한 차세대 메모리 아키텍처의 청사진을 제시한다.<sup>14</sup> 이 프로젝트는 AMD EPYC Venice 플랫폼을 기반으로 CXL(Compute Express Link) 기술을 적극 활용한다.

- **CXL 스위치 기반 풀링:** Lake Tahoe 아키텍처의 핵심은 CXL 스위치를 통해 여러 호스트(CPU/GPU)가 거대한 메모리 풀(Memory Pool)을 공유하는 것이다. H3 Platform이 시연한 CXL 메모리 어플라이언스는 2TB 용량의 CXL 모듈 22개를 장착하여, 단일 새시에서 수십 테라바이트(TB) 급의 메모리를 제공한다.<sup>15</sup>
- **동적 할당(Dynamic Allocation):** 에이전틱 AI는 작업의 난이도에 따라 필요한 메모리 양이 급격하게 변동한다. 삼성전자가 개발한 'DC MFM(Data Center Memory Fabric Manager)' 소프트웨어는 이러한 변동성에 대응하여, 애플리케이션의 요구에 따라 CXL 메모리를 실시간으로 할당하거나 회수한다. 이는 물리적인 서버 증설 없이도 논리적인 메모리 용량을 무한대에 가깝게 확장하는 효과를 제공한다.<sup>15</sup>

- **성능 효율성:** CXL 스위칭 기술을 적용한 메모리 풀링 시스템은 기존의 인피니밴드(Infiniband) 네트워크 대비 데이터 처리량(Throughput)은 4.6배 향상시키고, 꼬리 지연시간(Tail Latency)은 82% 감소시키는 것으로 확인되었다.<sup>15</sup> 이는 실시간성이 중요한 에이전틱 AI의 추론 성능을 보장하는 데 결정적인 역할을 한다.

### 2.3.2 벡터 데이터베이스(Vector DB)와 검색 증강(RAG) 가속

에이전틱 AI는 의사결정을 위해 방대한 지식 베이스를 검색해야 한다. 이를 위해 텍스트 데이터를 벡터(Vector) 형태로 변환하여 저장하는 벡터 데이터베이스가 필수적이다. 벡터 DB는 고속의 유사도 검색(Similarity Search)을 지원해야 하므로, 디스크가 아닌 메모리상(In-Memory)에 상주하는 것이 유리하다.

CXL 기반의 대용량 메모리는 HBM보다 저렴한 비용으로 수 테라바이트 규모의 벡터 DB를 메모리에 올릴 수 있게 해 주며, 이는 에이전틱 AI의 '장기 기억(Long-term Memory)'을 구현하는 하드웨어적 토대가 된다.<sup>16</sup> 삼성전자와 마이크론 등 주요 메모리 제조사들은 이러한 수요를 겨냥해 CMM-D(CXL Memory Module - DRAM) 제품의 양산을 서두르고 있으며, 이는 2025년 하반기부터 본격적인 시장을 형성할 것으로 전망된다.<sup>18</sup>

---

## 3. 피지컬 AI(Physical AI): 엣지(Edge)로 확장되는 고성능 메모리 수요

### 3.1 엠바디드 AI(Embodied AI): 로봇, 물리 세계의 서버가 되다

피지컬 AI는 인공지능이 디지털 공간을 넘어 물리적 하드웨어(Body)를 통해 현실 세계와 상호작용하는 '엠바디드 AI(Embodied AI)'를 의미한다. 자율주행차에서 시작된 피지컬 AI의 흐름은 이제 휴머노이드 로봇으로 진화하고 있으며, 이는 메모리 반도체 수요의 지형도를 데이터센터 중심에서 엣지(Edge) 영역으로 확장시키고 있다.

로봇은 클라우드에 전적으로 의존할 수 없다. 보행 중 균형을 잡거나, 날아오는 물체를 피하는 것과 같은 실시간 제어 작업은 밀리초(ms) 단위의 초저지연성을 요구하기 때문이다. 따라서 로봇 내부에는 고성능 GPU와 대용량 메모리가 탑재된 강력한 온디바이스(On-device) AI 컴퓨터가 내장되어야 한다. 젠슨 황 엔비디아 CEO는 이를 "AI의 다음 물결"이라고 명명하며, 로봇 산업이 가져올 파괴적 혁신을 강조했다.<sup>20</sup>

## 3.2 피지컬 AI 하드웨어의 스펙 혁명: Tesla & Nvidia

주요 테크 기업들이 공개한 최신 로봇용 하드웨어 스펙을 분석해보면, 피지컬 AI가 요구하는 메모리 사양의 급격한 상향 평준화를 확인할 수 있다.

### 3.2.1 테슬라 AI5 (HW5.0): 자율주행을 넘어 로봇 두뇌로

테슬라가 개발 중인 차세대 FSD(Full Self-Driving) 컴퓨터인 'AI5'는 자율주행차뿐만 아니라 휴머노이드 로봇 '옵티머스'의 두뇌로도 사용될 예정이다. 공개된 정보에 따르면, AI5는 삼성전자와 TSMC의 선단 공정(3/4nm 급)을 통해 생산되며, 기존 HW4.0 대비 압도적인 성능 향상을 목표로 한다.

- **성능 향상:** AI5는 HW4.0 대비 연산 능력은 약 8배, 전력 효율은 3배 향상될 것으로 추정된다.<sup>21</sup>
- **메모리 스펙 폭증:** 가장 주목할 점은 메모리 사양이다. AI5는 HW4.0 대비 메모리 용량은 **9배**, 대역폭은 **5배** 증가할 것으로 알려졌다.<sup>21</sup> 이는 로봇이 시각, 촉각, 고유수용성 감각 등 멀티모달(Multi-modal) 데이터를 실시간으로 융합(Sensor Fusion)하고 처리하기 위해 막대한 메모리 버퍼가 필요하기 때문이다.
- **생산 및 도입 일정:** 2026년 초 프로토타입 공개 및 2027년 양산을 목표로 하고 있으며, 이는 메모리 슈퍼사이클의 중후반부 수요를 견인할 핵심 요인이다.<sup>23</sup>

### 3.2.2 엔비디아 Jetson Thor: 로봇 전용 컴퓨팅 플랫폼

엔비디아가 발표한 휴머노이드 로봇용 플랫폼 'Jetson Thor'는 최신 블랙웰(Blackwell) 아키텍처를 기반으로 설계되었다.

- **통합 메모리:** Jetson Thor는 128GB의 LPDDR5X 통합 메모리를 탑재하며, 대역폭은 273GB/s에 달한다.<sup>24</sup> 이는 일반적인 고성능 게이밍 PC의 메모리 용량(32~64GB)을 훨씬 상회하는 수준이다.
- **FP8 정밀도 지원:** 생성형 AI 모델의 효율적인 구동을 위해 8비트 부동소수점(FP8) 연산을 지원하며, 이를 통해 1000억 개 이상의 파라미터를 가진 거대 모델을 로봇 내부에서 실행할 수 있다.<sup>26</sup>

표 2: 주요 피지컬 AI 플랫폼의 메모리 사양 비교

| 플랫폼                      | 제조사    | 타겟 애플리케이션         | 주요 프로세서 아키텍처                     | 메모리 유형                | 메모리 용량                      | 메모리 대역폭      | 특징                          |
|--------------------------|--------|-------------------|----------------------------------|-----------------------|-----------------------------|--------------|-----------------------------|
| <b>Tesla AI5 (HW5.0)</b> | Tesla  | 자율주행차, Optimus 로봇 | 자체 설계 NPU (Samsung/TSMC 파운드리)    | GDDR7 또는 LPDDR 6 (추정) | HW4 대비 9배 증가 (약 100GB + 예상) | HW4 대비 5배 증가 | 에너지 효율 극대화, 비디오 메모리 집약적     |
| <b>Jetson Thor</b>       | Nvidia | 휴머노이드 로봇 (GROOT)  | Blackwell GPU + ARM Neoverse CPU | LPDDR 5X              | 128 GB                      | 273 GB/s     | 통합 메모리 구조, 생성형 AI 최적화       |
| <b>Orin AGX (기존)</b>     | Nvidia | 자율주행, 산업용 로봇      | Ampere GPU                       | LPDDR 5               | 64 GB                       | 204 GB/s     | 현재 주력 제품, Thor 대비 성능 1/8 수준 |

자료: Tesla AI5 유출 정보 <sup>21</sup>, Nvidia Jetson Thor 공식 스펙 <sup>24</sup> 종합.

### 3.3 LPDDR6: 엣지 AI를 위한 차세대 표준

피지컬 AI의 고성능화는 모바일 DRAM 표준의 진화를 가속화하고 있다. 배터리로 구동되는 로봇은 전력 효율(Power Efficiency)이 무엇보다 중요하기 때문에, PC나 서버용 DDR 메모리보다는 LPDDR(Low Power DDR) 메모리를 선호한다.

JEDEC이 최근 발표한 **LPDDR6(JESD209-6)** 표준은 이러한 요구사항을 완벽하게 충족시킨다.

- **대역폭 혁신:** LPDDR6는 핀당 속도가 최대 **14.4Gbps**에 달하며, 이는 기존 LPDDR5X의 **8.5Gbps** 대비 약 **70%** 향상된 수치이다. **64비트 버스** 기준으로 최대 **38.4GB/s**의 대역폭을 제공한다.<sup>27</sup>
- **아키텍처 변화:** 기존의 **16비트 채널 4개** 구조에서 **24비트 채널 4개** 구조로 변경되어 채널 효율성을 높였으며, **AI 연산**에 최적화된 새로운 명령어 세트를 지원한다.<sup>28</sup>
- **온디바이스 AI 활성화:** LPDDR6의 도입은 로봇뿐만 아니라 **AI PC, AI 스마트폰** 등 모든 엣지 디바이스에서 에이전트 기능을 수행할 수 있는 기반을 마련해 준다. 시놉시스(Synopsys) 등 주요 IP 업체들은 이미 TSMC의 N2P 공정에서 LPDDR6 IP의 실리콘 검증을 마쳤으며, 2026년부터 본격적인 상용화가 시작될 예정이다.<sup>29</sup>

---

## 4. DRAM 메모리 슈퍼사이클의 구조적 특성 분석

에이전트 AI와 피지컬 AI가 견인하는 수요의 변화는 공급 측면의 제약 요인과 맞물려, 과거와는 전혀 다른 양상의 메모리 슈퍼사이클을 만들어내고 있다. 이는 단순한 경기 순환이 아닌, 산업 구조의 재편에 가깝다.

### 4.1 공급 측면: HBM의 '웨이퍼 잠식(Wafer Penalty)' 효과

이번 슈퍼사이클의 가장 강력한 공급 측 요인은 **HBM(High Bandwidth Memory)** 생산이 일반 DRAM 공급 능력을 물리적으로 잠식한다는 점이다. 이를 업계에서는 '웨이퍼 페널티(Wafer Penalty)'라고 부른다.

1. **다이 크기(Die Size) 증가:** HBM 코어 다이(Core Die)는 동일 용량의 DDR5 다이 대비 베이스 다이(Base Die)와 TSV 영역 등으로 인해 칩 크기가 약 **35~45%** 더 크다.<sup>30</sup> 이는 웨이퍼 한 장당 생산할 수 있는 칩의 개수(Net Die)가 그만큼 줄어든다는 것을 의미한다.
2. **공정 난이도와 수율 손실:** HBM 제조에는 미세 구멍을 뚫는 TSV 공정과 칩을 적층하는 본딩(Bonding) 공정이 추가된다. 이 과정에서 발생하는 수율 손실을 감안하면, HBM 1개를 생산하는 데 필요한 웨이퍼 투입량은 동일 용량의 일반 DRAM 대비 약 **3배**에 달한다는 것이 트렌드포스와 업계의 정설이다.<sup>32</sup>
3. **라인 전환(Conversion)의 가속화:** 삼성전자와 SK하이닉스는 폭증하는 HBM 수요를 맞추기 위해, 수익성이 낮은 레거시 DDR4/DDR5 라인을 HBM용 선단 공정(1b, 1c nm) 라인으로 전환하고 있다. SK하이닉스의 경우 2027년까지 전체 DRAM 생산량의 약 **40%**가 HBM이 될 것으로 전망된다.<sup>33</sup>

이러한 현상은 결과적으로 **\*\*풍선 효과\*\***를 야기한다. HBM 생산을 늘릴수록 일반 서버용



DDR5, PC/모바일용 LPDDR 메모리의 공급은 줄어들 수밖에 없다. 특히 2025년부터 에이전틱 AI용 추론 서버 수요가 폭발하면서 DDR5 고용량 모듈 수요가 급증할 때, 시장은 심각한 공급 부족(Shortage)에 직면하게 될 것이다.

## 4.2 수요 측면: 삼중고(Triple-Whammy)의 도래

과거의 사이클이 모바일(2010년대 초)이나 클라우드 서버(2017-2018) 중 하나의 킬러 앱에 의해 주도되었다면, 이번 사이클은 세 가지 거대한 수요의 파도가 시차를 두고, 그러나 중첩되어(Overlapping) 몰려오는 형국이다.

1. 제1파: AI 학습(Training) 수요 (2024~): 엔비디아 H100/Blackwell 등 가속기 기반의 HBM 수요. 여전히 공급이 수요를 따라가지 못하는 상태이다.
2. 제2파: 에이전틱 AI 추론(Inference) 수요 (2025 하반기~): LLM이 에이전트화되면서 추론 서버의 대형화가 진행된다. 128GB 이상의 고용량 DDR5 RDIMM과 CXL 메모리 모듈 수요가 폭발한다. 최태원 SK그룹 회장 역시 추론 시장의 확대가 가져올 메모리 수요 급증을 경고한 바 있다.<sup>35</sup>
3. 제3파: 피지컬/엣지 AI 수요 (2026~2027): 테슬라 옵티머스의 양산과 AI PC/폰의 교체 주기가 맞물리며 LPDDR6 등 고성능 엣지 메모리 수요가 가세한다.

## 4.3 사이클의 진폭(Amplitude)과 지속성(Duration) 전망

### 4.3.1 진폭: 역사상 가장 높은 매출과 이익률

이번 슈퍼사이클의 진폭, 즉 시장 규모와 이익률은 과거 2018년의 고점을 훌쩍 뛰어넘을 것으로 분석된다.

- 가격(Price)의 구조적 상승: HBM은 일반 DRAM 대비 5~7배 높은 가격 프리미엄을 형성하고 있다. 전체 DRAM 매출에서 HBM이 차지하는 비중이 2024년 10% 미만에서 2026년 30% 이상으로 증가함에 따라, 블렌디드 ASP(평균판매단가)는 급격히 상승할 것이다.<sup>34</sup>
- 매출 전망: 글로벌 DRAM 시장 매출은 2025년 약 1,210억 달러를 기록하며 전 고점에 근접하고, 2026~2027년에는 2,000억 달러를 돌파하며 사상 최대치를 경신할 것으로 예측된다.<sup>36</sup> 이는 AI가 기업의 '선택'이 아닌 '생존 필수 요소'가 되면서 가격 저항이 낮아졌기 때문이다. 모건스탠리는 이를 두고 "단순한 사이클이 아닌 패러다임 시프트"라고

평가하며, 삼성전자와 SK하이닉스의 이익이 시장 컨센서스를 30~50% 상회할 수 있다고 전망했다.<sup>38</sup>

#### 4.3.2 지속성: 2028년까지 이어지는 장기 호황

일반적인 메모리 사이클은 1.5~2년의 상승 후 공급 과잉으로 꺾이는 패턴을 보였다. 그러나 이번 사이클은 2027년, 길게는 2028년까지 이어지는 '장기 슈퍼사이클(Longer-duration Supercycle)'이 될 가능성이 높다.

- 재고 소진과 공급 지연: 트렌드포스에 따르면 2025년 말 기준 DRAM 공급사들의 재고는 3.3주 수준으로 역사적 최저점(2018년 수준)에 도달할 것으로 보인다.<sup>1</sup> 반면, 주요 3사(삼성, SK, 마이크론)의 신규 팹(Fab) 증설 효과는 2027년 이후에나 본격화된다. 클린룸 건설과 장비 반입에 소요되는 물리적 시간이 절대적으로 부족하기 때문이다.<sup>39</sup>
- 수요의 바통 터치: 2025년까지는 학습용 수요가, 2026년부터는 에이전틱 AI의 추론 수요가, 2027년부터는 휴머노이드 로봇 수요가 차례로 시장을 견인하며 수요의 공백을 허락하지 않을 것이다. 특히 골드만삭스는 로봇 시장이 2035년까지 380억 달러 규모로 성장할 것으로 보며, 이는 장기적인 메모리 수요의 하방 경직성을 제공한다.<sup>40</sup>

표 3: 주요 투자은행 및 조사기관의 슈퍼사이클 전망 종합

| 기관                    | 예상 사이클 지속 기간 | 전망 요약 및 근거         | 주요 코멘트   |
|-----------------------|--------------|--------------------|--|
| <b>Morgan Stanley</b> | 2025 ~ 2027+ | 에이전틱 AI로 인한 구조적 변화 | "과거 사이클과 비교 불가. 2027년 정점까지 강력한 상승세 지속. 이익률(Earnings)이 밸류에이션을 압도할 것" <sup>2</sup> |
| <b>Goldman Sachs</b>  | ~ 2027       | 데이터센터 전력 및 메모리 수요  | 2026년 말 데이터센터 가동률 95% 상회 예상. 2027년까지 매우タイト한 수급 전망<br><small>41</small>           |

|                     |               |                        |   |
|---------------------|---------------|------------------------|---|
| <b>TrendForce</b>   | ~ 2027 (Peak) | HBM 웨이퍼 잠식<br>및 재고 소진  | 2025년 말 공급사<br>재고 3.3주(역대<br>최저). 일반 DRAM<br>가격도 HBM 효과로<br>동반 상승 예상 <sup>1</sup> |
| <b>SK Hynix</b>     | ~ 2030        | 커스텀 HBM 및 AI<br>메모리 성장 | 연평균 30% 이상의<br>AI 메모리 시장<br>성장. 2030년까지<br>장기적인 성장세<br>유지 <sup>42</sup>           |
| <b>TechInsights</b> | ~ 2028        | AI가 주도하는<br>구조적 성장     | 2026년 이후에도 AI<br>수요가 메모리<br>시장의 병목이 되며<br>가격 상승 견인 <sup>44</sup>                  |

## 5. 경쟁 구도 및 기업별 대응 전략

슈퍼사이클의 혜택은 준비된 기업에게 집중될 것이다. 주요 메모리 3사는 각기 다른 전략으로 에이전틱/피지컬 AI 시장을 공략하고 있다.

### 5.1 삼성전자: CXL 생태계 주도와 턴키(Turnkey) 전략

삼성전자는 HBM 시장에서의 초기 대응 지연을 만회하기 위해 '규모의 경제'와 '차세대 표준 선점'에 집중하고 있다.

- **CXL 올인:** 삼성은 에이전틱 AI 시대의 핵심이 될 CXL 메모리(CMM-D) 시장을 가장 공격적으로 개척하고 있다. 레드햇(Red Hat) 등과 협력하여 CXL 메모리 동작 검증을 완료했으며, 2025년 말부터 양산에 돌입한다.<sup>18</sup> 이는 HBM의 용량 한계를 보완하는 솔루션으로, 삼성의 주력인 일반 DRAM 생산 능력을 고부가가치화할 수 있는 전략이다.
- **HBM4 및 턴키 솔루션:** 2026년 양산을 목표로 HBM4 개발을 가속화하고 있으며, 메모리-파운드리-패키징을 일괄 제공하는 턴키 서비스를 통해 고객사의 공급망 복잡성을 줄여주는 전략을 취한다. 이는 테슬라나 구글 같은 빅테크가 자체 칩(Custom Silicon)을 개발할 때 매력적인 제안이 될 수 있다.<sup>45</sup>

## 5.2 SK하이닉스: HBM 기술 리더십과 커스텀 메모리

SK하이닉스는 현재 HBM 시장의 절대 강자로서의 지위를 유지하면서, 단순 공급사를 넘어 AI 인프라 파트너로 진화하고 있다.

- **CAPEX 집중:** 청주 M15X 팹과 용인 클러스터에 투자를 집중하여 2026년부터 생산 능력을 대폭 확충한다. 이는 HBM 생산으로 인한 일반 DRAM 공급 부족을 해소하고 시장 점유율을 방어하기 위함이다.<sup>33</sup>
- **커스텀 HBM (Custom HBM):** 고객사의 AI 가속기 아키텍처에 최적화된 맞춤형 HBM을 개발하여 경쟁사와의 차별화를 꾀하고 있다. 이는 로직 다이(Logic Die)에 고객사의 IP를 통합하는 형태로, 메모리 반도체의 부가가치를 극대화하는 전략이다.<sup>46</sup>

## 5.3 마이크론: 엠티 AI와 공급망 다변화

마이크론은 HBM3E 시장에 성공적으로 안착하며 3강 체제를 굳혔다. 특히 미국 기업이라는 지정학적 이점을 활용하여 서구권 데이터센터 고객들을 공략하고 있다.

- **LPCAMM2 등 폼팩터 혁신:** 마이크론은 AI PC와 온디바이스 AI 시장을 겨냥해 성능과 전력 효율을 개선한 LPCAMM2 모듈을 업계 최초로 상용화했다. 이는 피지컬 AI 기기의 공간 제약을 해결해 줄 수 있는 핵심 솔루션이다.

---

# 6. 결론: AI 패브릭(Fabric) 시대의 메모리

## 6.1 요약 및 시사점

에이전틱 AI와 피지컬 AI의 부상은 메모리 산업에 있어 단순한 수요처의 추가가 아닌, 존재 이유의 재정의의 의미를 지닌다. AI가 '생각(Thinking)'을 넘어 '행동(Acting)'하고 '움직이는(Moving)' 단계로 진화함에 따라, 메모리는 시스템의 성능과 지능을 결정짓는 가장 중요한 자원이 되었다.

본 보고서의 분석 결과, 다가오는 메모리 슈퍼사이클은 다음과 같은 특징을 가질 것으로

전망된다.

1. **진폭의 확대:** HBM의 가격 프리미엄과 일반 DRAM의 공급 부족이 겹치며, 매출과 이익 규모는 2018년의 2배 수준에 이를 것이다.
2. **지속성의 장기화:** 에이전틱 AI(2025~2026)와 피지컬 AI(2027~)가 릴레이 하듯 수요를 견인하며, 사이클의 정점은 2027년 이후로 늦춰질 것이다. 구조적 결핍은 2028년까지도 해소되기 어려울 수 있다.

## 6.2 향후 과제

메모리 기업들은 HBM 생산 능력 확충이라는 당면 과제와 함께, CXL 및 LPDDR6와 같은 차세대 기술 표준을 선점해야 하는 이중의 과제를 안고 있다. 투자자들은 단순한 비트 성장률(Bit Growth)보다는, 이중 메모리 아키텍처와 엣지 AI 플랫폼에서 어떤 기업이 기술적 해자(Moat)를 구축하고 있는지를 주시해야 할 것이다. AI는 이제 툴(Tool)에서 세상을 움직이는 패브릭(Fabric)이 되었고, 그 패브릭의 씨실과 날실은 바로 반도체 메모리이다.

### 참고 자료

1. [News] Record-Low Memory Stock Could Signal Supercycle Ahead, Likely Peaking by 2027, 11월 29, 2025에 액세스, <https://www.trendforce.com/news/2025/09/30/news-record-low-memory-stock-could-signal-supercycle-ahead-likely-peaking-by-2027/>
2. Memory Semiconductor Supercycle Gains Momentum as AI Demand Surges, 11월 29, 2025에 액세스, <https://www.chosun.com/english/industry-en/2025/09/29/GWQDNWLN6BGYNPCMYEYJ7Y3GFI/>
3. Agentic AI vs. Generative AI - IBM, 11월 29, 2025에 액세스, <https://www.ibm.com/think/topics/agentic-ai-vs-generative-ai>
4. What Are AI Agents? | IBM, 11월 29, 2025에 액세스, <https://www.ibm.com/think/topics/ai-agents>
5. How NVIDIA Isaac GR00T N1 Is Redefining Humanoid Robotics? - Analytics Vidhya, 11월 29, 2025에 액세스, <https://www.analyticsvidhya.com/blog/2025/03/nvidia-isaac-gr00t-n1/>
6. NVIDIA Announces Isaac GR00T N1 — the World's First Open Humanoid Robot Foundation Model — and Simulation Frameworks to Speed Robot Development, 11월 29, 2025에 액세스, <https://nvidianews.nvidia.com/news/nvidia-isaac-gr00t-n1-open-humanoid-robot-foundation-model-simulation-frameworks>
7. What is agentic AI? | Micron Technology Inc., 11월 29, 2025에 액세스, <https://www.micron.com/about/micron-glossary/agentic-ai>
8. The Memory for Agentic AI (Stateful) | by Vishal Anand | Oct, 2025 | Medium, 11월 29, 2025에 액세스,

- <https://er-vishalanand.medium.com/the-memory-for-agentic-ai-stateful-89bb3e7a2639>
9. Managing memory for AI agents, 11월 29, 2025에 액세스,  
<https://tsaiprabhanj.medium.com/managing-memory-for-ai-agents-56db285d892a>
  10. with Agentic AI and Memory Centric Computing, 11월 29, 2025에 액세스,  
[https://files.futurememorystorage.com/proceedings/2025/20250807\\_DRAM-304-1\\_SO.pdf](https://files.futurememorystorage.com/proceedings/2025/20250807_DRAM-304-1_SO.pdf)
  11. Agentic AI vs. Generative AI: What's the Difference - 2025 - Aalpha Information Systems, 11월 29, 2025에 액세스,  
<https://www.aalpha.net/blog/agentic-ai-vs-generative-ai-differences/>
  12. Agentic AI Memory Systems Are a Bellwether for Network Traffic Growth - F5, 11월 29, 2025에 액세스,  
<https://www.f5.com/company/blog/agentic-ai-memory-systems-are-a-bellwether-for-network-traffic-growth>
  13. AI and Memory Wall. (This blogpost has been written in... | by Amir Gholami | riselab | Medium, 11월 29, 2025에 액세스,  
<https://medium.com/riselab/ai-and-memory-wall-2cb4265cb0b8>
  14. Samsung Electronics Highlights Open Collaboration for the AI Era at ..., 11월 29, 2025에 액세스,  
<https://semiconductor.samsung.com/news-events/tech-blog/samsung-electronics-highlights-open-collaboration-for-the-ai-era-at-ocp-global-summit-2025/>
  15. Heterogeneous Memory System Architecture for Next gen Agentic AI Server Platform Montage, H - YouTube, 11월 29, 2025에 액세스,  
<https://www.youtube.com/watch?v=9D8TAFpJLpI>
  16. The Cambridge Report on Database Research - arXiv, 11월 29, 2025에 액세스,  
<https://arxiv.org/html/2504.11259v1>
  17. 2025 Program at a Glance - FMS: the Future of Memory and Storage, 11월 29, 2025에 액세스,  
<https://futurememorystorage.com/program/program-at-a-glance?date=2025-08-06>
  18. Samsung Electronics Presents Vision for AI Memory and Storage at FMS 2025, 11월 29, 2025에 액세스,  
<https://semiconductor.samsung.com/news-events/tech-blog/samsung-electronics-presents-vision-for-ai-memory-and-storage-at-fms-2025/>
  19. Scaling The Storage Pyramid With Micron's Datacenter Memory And Flash, 11월 29, 2025에 액세스,  
<https://www.nextplatform.com/2025/03/25/scaling-the-storage-pyramid-with-microns-datacenter-memory-and-flash/>
  20. NVIDIA Powers Humanoid Robot Industry With Cloud-to-Robot Computing Platforms for Physical AI, 11월 29, 2025에 액세스,  
<https://nvidianews.nvidia.com/news/nvidia-powers-humanoid-robot-industry-with-cloud-to-robot-computing-platforms-for-physical-ai>
  21. Tesla teases new AI5 chip that will revolutionize self-driving, 11월 29, 2025에 액세스,

- <https://www.teslarati.com/elon-musk-reveals-new-information-tesla-ai5-chip-for-self-driving/>
22. Everything We Know About HW5 / AI5: Tesla's Next-Gen FSD Computer, 11월 29, 2025에 액세스,  
<https://drive-parts.com.ua/en/vse-shcho-my-znaiemo-pro-hw5-ai5-kompiuter-fs-d-nastupnoho-pokolinnia-vid-tesla/>
  23. Tesla Optimus Humanoid Robot: Key Takeaways From the Q3 Earnings Call | Nasdaq, 11월 29, 2025에 액세스,  
<https://www.nasdaq.com/articles/tesla-optimus-humanoid-robot-key-takeaways-q3-earnings-call>
  24. Jetson Thor | Advanced AI for Physical Robotics - NVIDIA, 11월 29, 2025에 액세스,  
<https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-thor/>
  25. NVIDIA Jetson AGX Thor Developer Kit Hands-on Blackwell for Robotics - Page 2 of 3, 11월 29, 2025에 액세스,  
<https://www.servethehome.com/nvidia-jetson-agx-thor-developer-kit-blackwell-for-robotics/2/>
  26. Unlock Faster, Smarter Edge Models with 7x Gen AI Performance on NVIDIA Jetson AGX Thor, 11월 29, 2025에 액세스,  
<https://developer.nvidia.com/blog/unlock-faster-smarter-edge-models-with-7x-gen-ai-performance-on-nvidia-jetson-agx-thor/>
  27. What is JESD209-6 and why is it important for edge AI? - EE World Online, 11월 29, 2025에 액세스,  
<https://www.eeworldonline.com/what-is-jesd209-6-and-why-is-it-important-for-edge-ai/>
  28. LPDDR 6, a significant improvement-Electronics Headlines-EEWORLD, 11월 29, 2025에 액세스, <https://en.eeworld.com.cn/mp/lcbank/a403113.aspx>
  29. Synopsys teases 'silicon bring-up' of next-gen LPDDR6 IP fabbed on TSMC's new N2P process node - TweakTown, 11월 29, 2025에 액세스,  
<https://www.tweaktown.com/news/108337/synopsys-teases-silicon-bring-up-of-next-gen-lpddr6-ip-fabbed-on-tsmcs-new-n2p-process-node/index.html>
  30. AI-Related HBM Demand Squeezing Out DDR5 Capacity And Tightening Wafer Supply - MediaTek In The Eye Of The Storm - Wccftech, 11월 29, 2025에 액세스,  
<https://wccftech.com/ai-related-hbm-demand-squeezing-out-ddr5-capacity-and-tightening-wafer-supply-mediatek-in-the-eye-of-the-storm/>
  31. HBM to be 20% of DRAM market in 2024 ... - eeNews Europe, 11월 29, 2025에 액세스,  
<https://www.eenewseurope.com/en/hbm-to-be-20-of-dram-market-in-2024/>
  32. RAM: WTF?, 11월 29, 2025에 액세스, <https://gamersnexus.net/news/ram-wtf>
  33. [News] SK hynix Reportedly to Double DRAM Capacity in 2H26 to Match Samsung, Pulls Back on NAND - TrendForce, 11월 29, 2025에 액세스,  
<https://www.trendforce.com/news/2025/10/02/news-sk-hynix-reportedly-to-double-dram-capacity-in-2h26-to-match-samsung-pulls-back-on-nand/>
  34. The HBM Wars: SK Hynix's Dominance, Samsung's Roadmap, and the Looming Threat of Cyclicity - Patsnap Eureka, 11월 29, 2025에 액세스,



- <https://eureka.patsnap.com/insight/the-hbm-wars-sk-hynixs-dominance-samsungs-roadmap-and-the-looming-threat-of-cyclicality>
35. SK Group pivots from scale to efficiency to meet soaring AI market demand, 11월 29, 2025에 액세스,  
<https://www.koreatimes.co.kr/business/tech-science/20251103/sk-group-pivots-from-scale-to-efficiency-to-meet-soaring-ai-market-demand>
  36. Dynamic Random-Access Memory Market Size, Share & Industry Forecast 2035, 11월 29, 2025에 액세스,  
<https://www.researchnester.com/reports/dynamic-random-access-memory-dram-market/3460>
  37. Dynamic Random Access Memory (DRAM) Market Size, Share & Trends Report 2030, 11월 29, 2025에 액세스,  
<https://www.mordorintelligence.com/industry-reports/dynamic-random-access-memory-market>
  38. Morgan Stanley: “The AI Memory Super Cycle Has Arrived — and It's Unlike Anything We've Seen Before.” - Syz Blog, 11월 29, 2025에 액세스,  
<https://blog.syzgroup.com/syz-the-moment/morgan-stanley-the-ai-memory-super-cycle-has-arrived-and-its-unlike-anything-weve-seen-before>
  39. Memory industry to maintain cautious capex in 2026 - Evertiq, 11월 29, 2025에 액세스,  
<https://evertiq.com/news/2025-11-13-memory-industry-to-maintain-cautious-capex-in-2026>
  40. The global market for humanoid robots could reach \$38 billion by 2035 | Goldman Sachs, 11월 29, 2025에 액세스,  
<https://www.goldmansachs.com/insights/articles/the-global-market-for-robots-could-reach-38-billion-by-2035>
  41. AI to drive 165% increase in data center power demand by 2030 | Goldman Sachs, 11월 29, 2025에 액세스,  
<https://www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030>
  42. SK Hynix Forecasts 30% Annual Growth in AI Memory Market Through 2030 | by ODSC, 11월 29, 2025에 액세스,  
<https://odsc.medium.com/sk-hynix-forecasts-30-annual-growth-in-ai-memory-market-through-2030-1c9f19dad250>
  43. Exclusive-SK Hynix expects AI memory market to grow 30% a year to 2030, 11월 29, 2025에 액세스,  
<https://www.business-reporter.com/industrial-innovations/exclusive-sk-hynix-expects-ai-memory-market-to-grow-30-a-year-to-2030>
  44. 2026 Memory Outlook Report - TechInsights, 11월 29, 2025에 액세스,  
<https://www.techinsights.com/outlook-reports-2026/memory-outlook-report>
  45. Samsung Accelerates 1c DRAM Device Integration, Aims to Mass Produce HBM4 by 2026, 11월 29, 2025에 액세스,  
<https://www.semicone.com/article-326.html>
  46. Feature: How the move to custom HBM boosts supply visibility, 11월 29, 2025에 액세스,  
<https://www.mobileworldlive.com/ai-cloud/feature-how-the-move-to-custom-hb>



[m-boosts-supply-visibility/](#)

47. Exploring the AI Ecosystem: The Future and SK hynix's Vision, 11월 29, 2025에  
엑세스,

<https://news.skhynix.com/exploring-the-ai-ecosystem-how-sk-hynix-is-driving-ai-s-future/>