

HBM 두께 완화 기술 및 SK하이닉스 HBM3E 양산 동향 분석 보고서

문서번호 CRSM-AI-2026-AUTO

작성일 2026-06-13

작성 Cressem AI 시스템 (자동 생성)

보안등급 사내 비밀 (Confidential)

버전 v1.0

목 차

HBM 두께 완화 기술 및 SK하이닉스 HBM3E 양산 동향 분석 보고서	3
개요/배경	3
HBM 고단 적층 및 두께 완화 기술 분석	4
결론/시사점	7

HBM 두께 완화 기술 및 SK하이닉스 HBM3E 양산 동향 분석 보고서

본 보고서는 고단 적층 HBM의 핵심 과제인 패키지 두께 제어 기술을 분석하고, SK하이닉스의 HBM3E 양산 현황 및 기술적 우위를 검토합니다. 특히 Advanced MR-MUF 공정을 중심으로 한 차세대 HBM의 기술적 진화 방향을 제시합니다.

개요/배경

1. AI 가속기 시장의 폭발적 성장과 HBM의 전략적 가치

최근 생성형 AI(Generative AI) 기술의 급격한 발전과 대규모 언어 모델(LLM)의 확산은 데이터센터와 클라우드 컴퓨팅 인프라의 패러다임을 근본적으로 변화시키고 있습니다. 이러한 변화의 중심에는 방대한 양의 데이터를 초고속으로 처리해야 하는 AI 가속기(AI Accelerator) 시장이 자리 잡고 있으며, 이를 뒷받침하기 위한 핵심 메모리 솔루션으로 고대역폭 메모리(HBM, High Bandwidth Memory)의 중요성이 그 어느 때보다 강조되고 있습니다.

기존의 DDR(Double Data Rate) 방식은 데이터 전송 통로인 인터페이스의 한계로 인해 AI 모델이 요구하는 초고속·대용량 데이터 병목 현상을 해결하기에 역부족입니다. 반면, HBM은 여러 개의 DRAM 칩을 수직으로 적층하고 TSV(Through Silicon Via, 실리콘 관통 전극) 기술을 통해 데이터 통로를 확보함으로써, 기존 메모리 대비 압도적인 대역폭(Bandwidth)과 낮은 전력 소모를 구현합니다. 이에 따라 NVIDIA의 H100, B200 및 B300 시리즈와 같은 차세대 AI 가속기 시장이 확대됨에 따라, HBM은 단순한 부가 메모리를 넘어 AI 시스템의 성능을 결정짓는 핵심적인 전략 자산으로 부상하였습니다 [출처: news.skhynix.co.kr].

2. 고단 적층 기술의 진화와 기술적 임계점(Technical Bottleneck)

AI 모델의 파라미터 수가 기하급수적으로 증가함에 따라, 단일 패키지 내에 더 많은 메모리 용량을 확보하기 위한 '고단 적층(High-stacking)' 경쟁이 가속화되고 있습니다. 현재 시장은 HBM3를 넘어 HBM3E(HBM3 Extended) 단계에 진입하였으며, 향후 HBM4로의 전환이 예견되어 있습니다. 적층 수가 8단(8-Layer)을 넘어 12단, 16단 이상으로 높아질수록 메모리 용량은 증가하지만, 이는 동시에 반도체 패키징 공정에서 전례 없는 기술적 난제를 야기합니다.

가장 먼저 직면하는 문제는 **전체 패키지 두께(Total Package Thickness)의 제한**입니다. 표준 규격에 따라 HBM 패키지의 전체 높이는 엄격하게 제한되어 있습니다. 적층 수가 늘어남에도 불구하고 두께를 일정 수준 이하로 유지해야 하므로, 개별 DRAM 칩의 두께를 극단적으로 얇게 만드는 기술이 필수적입니다. 칩이 얇아질수록 물리적 강도가 약해져 휨(Warping) 현상이 발생하기 쉬우며, 이는 공정 중 불량률을 높이는 직접적인 원인이 됩니다.

두 번째 난제는 **열 관리(Thermal Management)**입니다. 칩을 수직으로 높게 쌓을수록 내부에서 발생하는 열이 외부로 방출되는 경로가 차단되어, 동작 온도(°C)가 급격히 상승하게 됩니다. 고온 환경은 데이터 전송 오류를 유발할 뿐만 아니라, 소자의 신뢰성을 저하시키고 전력 효율을 떨어뜨리는 결과를 초래합니다. 따라서 고성능 연산 시 발생하는 발열을 효과적으로 제어할 수 있는 방열 구조 설계와 소재 기술이 HBM 경쟁력의 핵심 지표가 되고 있습니다.

3. 패키징 기술의 패러다임 변화와 검사 솔루션의 필요성

이러한 기술적 한계를 극복하기 위해 글로벌 메모리 제조사들은 각기 다른 차세대 패키징 공정 기술을 도입하며 격돌하고 있습니다. SK하이닉스는 기존의 TC-NCF(Thermal Compression Non-Conductive Film) 방식의 한계를 넘어, 열 방출과 공정 효율을 극대화한 Advanced MR-MUF(Mass Reflow Molded Underfill) 기술을 통해 시장의 주도권을 확보하고 있습니다 [출처: files-scs.pstatic.net]. 반면 삼성전자는 TC-NCF 기술을 고도화하여 초고단 시장을 공략하는 등, 적층 방식에 따른 기술적 차별화가 뚜렷하게 나타나고 있습니다.

결과적으로 HBM 시장은 단순히 '얼마나 많이 쌓느냐'의 문제를 넘어, '어떻게 얇고, 차갑고, 안정적으로 쌓느냐'의 싸움으로 변모하였습니다. 이러한 공정의 복잡도 증가는 제조 공정 중 발생할 수 있는 미세 결함이나 전기적 특성 변화를 실시간으로 감지할 수 있는 고정밀 검사(Inspection) 및 테스트(Test) 기술의 중요성을 시사합니다. 특히 TSV 연결성, Die-to-Die 통신 안정성, 그리고 고온 동작 시의 열적 안정성을 검증하는 기술은 HBM 양산 수율을 결정짓는 최종 관문이 될 것입니다.

구분	주요 기술적 과제	영향 요소	비고
두께 제어	고단 적층 시 패키지 높이 제한	칩 박막화(Thinning), Warpage 제어	규격 준수 필수
열 관리	적층 구조 내 열 축적 현상	방열 소재(MUF 등), 열전도율 향상	신뢰성 및 전력 효율 직결
신뢰성 확보	TSV 및 범프(Bump) 연결 안정성	전기적 특성, 데이터 전송 오류율	검사 장비의 정밀도 요구

표 1. 개요/배경

HBM 고단 적층 및 두께 완화 기술 분석

1. 고단 적층(High-stacking)에 따른 패키지 두께 제한의 기술적 난제

AI 가속기 및 데이터센터용 프로세서의 성능이 기하급수적으로 향상됨에 따라, 메모리 역시 대역폭(Bandwidth) 확장을 위해 적층 수(Stacking number)를 높이는 방향으로 진화하고 있습니다. 현재 HBM3E 단계에서는 8단(8-Layer)을 넘어 12단(12-Layer) 및 16단 이상의 초고단 적층이 요구되고 있습니다. 그러나 여기서 발생하는 가장 치명적인 물리적 제약은 **전체 패키지 두께(Total Package Thickness)의 제한**입니다.

표준 규격에 따라 HBM 패키지의 전체 두께는 엄격하게 제한되어 있습니다. 적층 수가 증가하더라도 패키지 높이를 일정 수준 이하로 유지해야 하므로, 개별 DRAM Die의 두께를 얇게 만드는 **초박형화(Ultra-thinning)** 기술이 필수적입니다. Die가 얇아질수록 다음과 같은 세 가지 핵심적인 기술적 난제가 발생합니다.

첫째, **웨이퍼 및 Die의 물리적 취성(Brittleness) 증가**입니다. Die를 수십 μm 단위로 얇게 갈아내는(Grinding) 과정에서 웨이퍼가 휘어지는 **휨(Warpage)** 현상이 심화되며, 이는 적층 공정 중 Die의 파손(Cracking)이나 물리적 손상으로 이어져 수율(Yield)을 급격히 저하시킵니다.

둘째, **TSV(Through Silicon Via) 형성 및 정렬(Alignment)의 난이도 상승**입니다. 칩이 얇아질수록 수천 개의 TSV 홀을 관통하는 공정의 정밀도가 요구되며, 층간 정렬 오차(Misalignment)가 발생할 경우 전기적 연결성(Connectivity) 결함으로 직결됩니다.

셋째, **열 방출(Thermal Dissipation) 및 신뢰성 문제**입니다. 칩이 얇아지면 열 저항은 낮아질 수 있으나, 적층 구조가 복잡해짐에 따라 내부에서 발생하는 열이 외부로 배출되지 못하고 갇히는 열 정체 현상이 발생할 수 있습니다. 이는 고온 동작 시 소자의 특성 변화를 야기하고 패키지의 구조적 안정성을 해치는 원인이 됩니다.

2. 두께 완화 및 제어를 위한 핵심 기술적 접근법

이러한 물리적 한계를 극복하기 위해 반도체 제조사들은 크게 세 가지 방향의 기술적 접근을 시도하고 있습니다.

가. 초박형 Die 제조 기술 (Ultra-thinning Technology)

DRAM Die의 두께를 최소화하기 위해 고정밀 CMP(Chemical Mechanical Polishing) 및 Grinding 공정이 도입됩니다. 최근에는 웨이퍼의 평탄도를 극대화하여 얇아진 Die에서도 물리적 강도를 유지할 수 있는 기술이

핵심입니다. 특히 12단 이상의 적층에서는 개별 Die의 두께를 기존보다 더욱 줄여야 하므로, 박막 상태에서도 변형을 최소화하는 공정 제어가 필수적입니다.

나. 차세대 적층 소재 및 공정 도입 (Advanced Stacking Process)

기존의 방식으로는 얇아진 Die 사이의 간극(Gap)을 메우고 구조적 안정성을 확보하는 데 한계가 있습니다. 이에 따라 SK하이닉스가 주도하는 **Advanced MR-MUF(Mass Reflow Molded Underfill)** 기술과 삼성전자의 **TC-NCF(Thermal Compression Non-Conductive Film)** 기술이 두께 제어의 핵심 솔루션으로 대두되었습니다.

- **Advanced MR-MUF:** 액체 형태의 보호재(MUF)를 칩 사이에 주입하여 굳히는 방식으로, 칩 사이의 간극을 균일하게 채울 수 있어 적층 시 발생하는 물리적 압력을 줄여줍니다. 이는 휨(Warping) 현상을 최소화하고 전체 패키지 두께를 규격 내로 정밀하게 유지하는 데 매우 유리합니다.
- **TC-NCF:** 비전도성 필름(NCF)을 칩 사이에 삽입하고 열과 압력을 가해 접합하는 방식입니다. 필름의 두께를 조절함으로써 적층 높이를 제어하며, 최근에는 필름의 두께를 극도로 얇게 만들어 고단 적층 시의 두께 한계를 극복하려는 연구가 진행 중입니다.

다. 하이브리드 본딩 (Hybrid Bonding) 기술로의 진화

적층 수가 16단 이상으로 넘어가는 미래 시점에는 기존의 범프(Bump) 기반 접합 방식 자체가 두께의 걸림돌이 됩니다. 범프는 물리적인 높이를 점유하기 때문입니다. 이를 해결하기 위해 구리(Cu)와 구리를 직접 연결하는 **하이브리드 본딩(Hybrid Bonding)** 기술이 논의되고 있습니다. 이 기술은 중간 접합층(Interconnect layer)의 높이를 거의 '제로(Zero)'에 가깝게 만들 수 있어, 칩 간 간격을 극단적으로 줄이고 패키지 전체 두께를 혁신적으로 완화할 수 있는 차세대 핵심 기술로 평가받습니다.

HBM 고단 적층 구조 단면도 (Typical 12-Layer Stack)

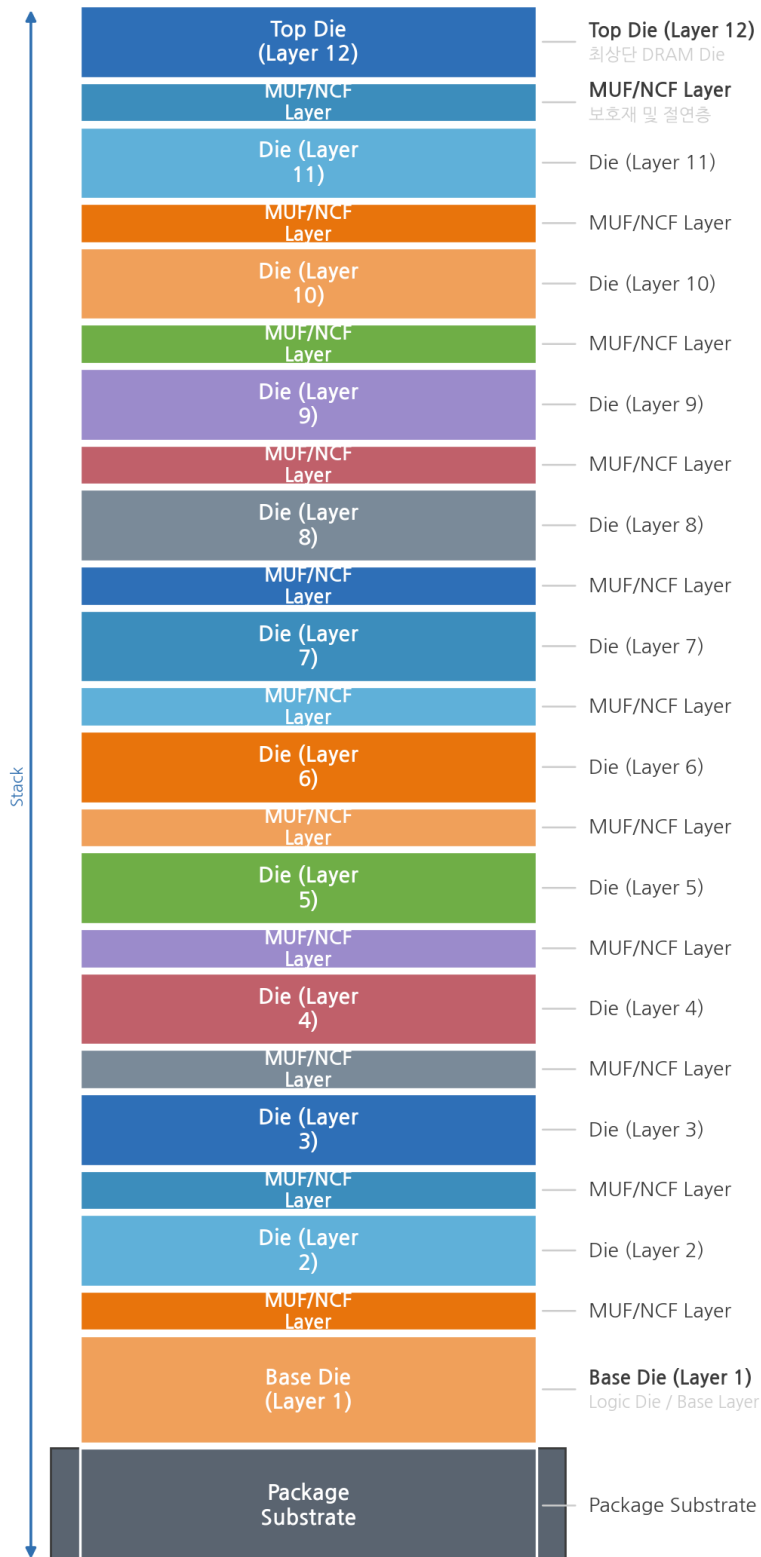


그림 1. HBM 고단 적층 구조 단면도 (Typical 12-Layer Stack)

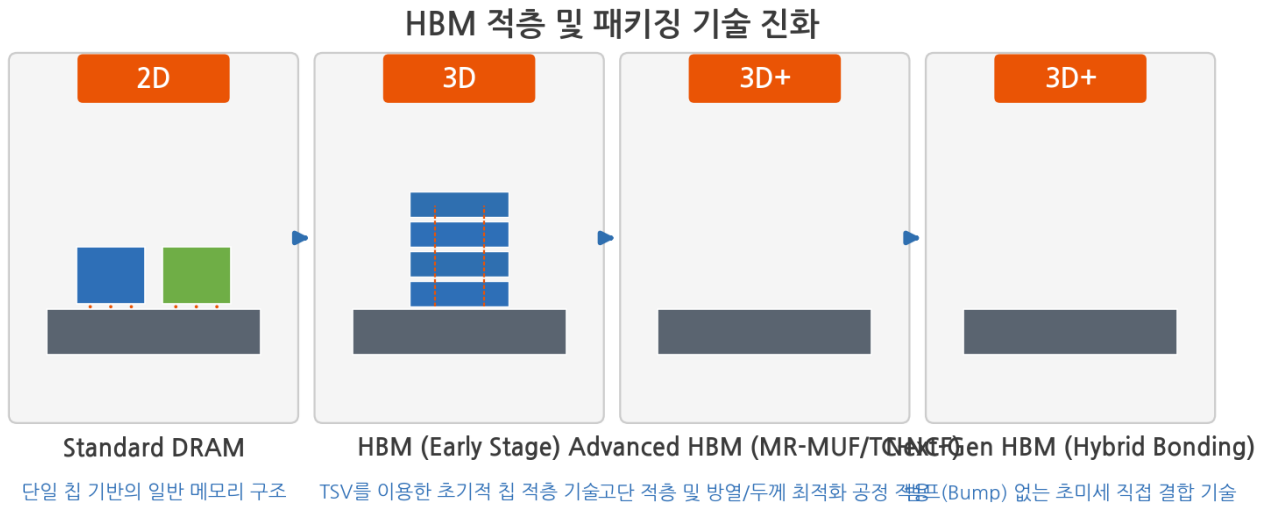


그림 2. HBM 적층 및 패키징 기술 진화

HBM 기술 진화 로드맵 (HBM3E to HBM4+)

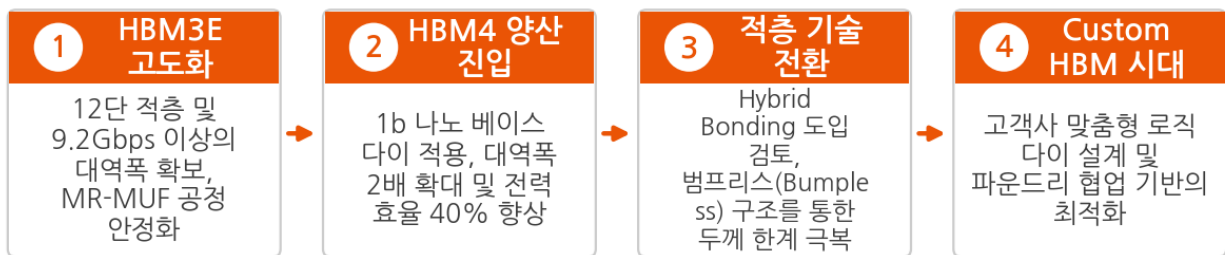


그림 3. HBM 기술 진화 로드맵 (HBM3E to HBM4+)

기술적 과제와 검사 기술의 중요성

HBM4와 그 이후의 차세대 메모리는 공정 난이도가 기하급수적으로 상승함에 따라, 미세한 결함이 전체 패키지의 불량으로 이어지는 리스크가 커집니다. 특히 하이브리드 본딩과 같은 신규 공정이 도입될 경우, 기존의 전기적 특성 검사뿐만 아니라 나노미터(nm) 단위의 정밀한 광학 검사(Optical Inspection)와 물리적 결함을 잡아낼 수 있는 고도화된 머신 비전 솔루션이 필수적입니다. 따라서 미래 HBM 시장의 주도권은 단순히 메모리 셀을 잘 만드는 것을 넘어, 초고단 적층 공정에서 발생하는 물리적/전기적 변수를 완벽하게 제어하고 검증할 수 있는 기술력에 달려 있습니다.

결론/시사점

본 보고서에서 분석한 바와 같이, AI 가속기 시장의 폭발적인 성장과 함께 고대역폭 메모리(HBM)는 단순한 메모리 반도체를 넘어 AI 시스템의 성능을 결정짓는 핵심 인프라로 자리 잡았습니다. 특히 적층 수가 증가함에 따라 발생하는 물리적 한계인 **전체 패키지 두께 제한(Thickness Constraint)**과 **발열 제어(Thermal Management)** 문제는 차세대 HBM 시장의 주도권을 결정짓는 가장 치명적인 기술적 변곡점이 될 것으로 전망됩니다.

기술적 측면에서 요약하자면, SK하이닉스는 **Advanced MR-MUF** 공정 기술을 통해 고단 적층 시 발생하는 휨(Warping) 현상을 억제하고, 우수한 열전도율을 확보함으로써 12단 이상의 고단 제품에서도 안정적인 수율과 성능을 증명하고 있습니다 [출처: 보고서SK하이닉스HBM3E양산동향및기술경쟁력분석20260509001.pdf]. 이는 경쟁사가 채택하고 있는 TC-NCF 방식과 차별화되는 지점으로, 향후 HBM4로 이어지는 초고단 적층 경쟁에서 SK하이닉스가 'First Mover'로서의 지위를 공고히 할 수 있는 핵심 동력이 될 것입니다.

결론적으로 HBM 시장의 승패는 단순히 '얼마나 높게 쌓느냐'가 아닌, '**제한된 두께 내에서 얼마나 효율적으로 열을 방출하며 전기적 신뢰성을 유지하느냐**'에 달려 있습니다. 이에 따라 다음과 같은 시사점을 도출할 수 있습니다.

1. **공정 기술의 고도화 및 표준화**: HBM4 및 그 이후 세대로 갈수록 하이브리드 본딩(Hybrid Bonding) 등 새로운 접합 기술의 도입이 가속화될 것이며, 이는 패키징 공정의 난이도를 기하급수적으로 높일 것입니다.

2. **검사 솔루션(Inspection Solution)의 중요성 증대**: 적층 구조가 복잡해지고 TSV(Through Silicon Via) 및 I/O 연결 밀도가 높아짐에 따라, 미세 결함을 잡아내기 위한 고정밀 광학 검사 및 전기적 특성 검사(Electrical Test)의 중요성이 극대화될 것입니다. 특히 Die-to-Die 통신 신뢰성과 열적 안정성을 실시간으로 검증할 수 있는 고도화된 검사 장비는 양산 수율을 결정짓는 필수 요소입니다.

3. **통합적 품질 관리 체계 구축**: 고객사(NVIDIA 등)의 요구 규격이 더욱 엄격해짐에 따라, 설계 단계부터 패키징, 최종 검사에 이르기까지 데이터 기반의 통합적인 품질 관리 체계가 요구됩니다.

결국, 차세대 HBM 시장은 독보적인 적층 기술력을 보유한 메모리 제조사와, 그 공정의 미세한 오류를 완벽히 포착해낼 수 있는 고성능 검사 솔루션 기업 간의 기술 협력이 시장의 성패를 가르는 핵심 축이 될 것입니다.