

HBM3 고대역폭 메모리 적층 구조 및 검사 기술 분석 보고서

문서번호 CRSM-AI-2026-AUTO

작성일 2026-07-20

작성 Cressem AI 시스템 (자동 생성)

보안등급 사내 비밀 (Confidential)

버전 v1.0

목 차

HBM3 고대역폭 메모리 적층 구조 및 검사 기술 분석 보고서	3
개요 및 배경	3
HBM3 적층 구조 및 설계 원리	4
HBM3 주요 패키징 기술 비교	6
적층 공정 주요 결함(Defect) 분석	8
고정밀 검사 솔루션 요구사항	10
AI 기반 비전 검사 적용 방안	11
결론 및 향후 기술 전망	12

HBM3 고대역폭 메모리 적층 구조 및 검사 기술 분석 보고서

HBM3의 핵심인 3D 수직 적층 구조와 TSV(Through-Silicon Via) 기술을 분석하고, 적층 공정에서 발생하는 주요 결함 및 이를 검출하기 위한 고정밀 검사 솔루션을 제안합니다.

개요 및 배경

1. AI 가속기 시장의 폭발적 성장과 메모리 병목 현상

최근 생성형 AI(Generative AI) 및 대규모 언어 모델(Large Language Model, LLM)의 급격한 발전은 컴퓨팅 인프라의 패러다임을 근본적으로 변화시키고 있습니다. NVIDIA의 H100, B200과 같은 고성능 AI 가속기(AI Accelerator) 시장이 폭발적으로 성장함에 따라, 연산 장치(GPU/NPU)의 성능을 뒷받침할 수 있는 고성능 메모리 솔루션의 중요성이 그 어느 때보다 강조되고 있습니다 [출처: newsletter.semianalysis.com].

기존의 DDR(Double Data Rate) 기반 메모리 구조는 연산 장치와 메모리 사이의 물리적 거리가 멀고 데이터 전송 통로(Bus Width)가 제한적이어서, 연산 속도에 비해 데이터를 공급하는 속도가 따라가지 못하는 '메모리 벽(Memory Wall)' 현상을 심화시켰습니다. 이러한 데이터 병목 현상은 AI 모델의 추론(Inference) 및 학습(Training) 속도를 저하시키는 결정적인 요인으로 작용하며, 이를 해결하기 위해 초고대역폭(Ultra-High Bandwidth)을 제공하는 메모리 아키텍처가 필수적으로 요구되고 있습니다.

2. HBM3의 등장과 기술적 패러다임의 전환

이러한 시장의 요구에 부응하여 등장한 HBM3(High Bandwidth Memory 3)는 기존 DRAM의 한계를 극복하기 위한 혁신적인 설계 원리를 채택하고 있습니다. HBM은 단순히 용량을 늘리는 차원을 넘어, 여러 개의 DRAM 다이(Die)를 수직으로 쌓아 올리는 3D 스택(3D Stacked) 구조를 통해 데이터 전송 거리를 획기적으로 단축하고, 병렬 처리 능력을 극대화한 기술의 결정체입니다 [출처: hyeongcheol.tistory.com].

HBM3는 이전 세대인 HBM2E를 넘어선 고성능 규격으로, 1024-bit의 매우 넓은 인터페이스(Wide Interface)와 고속 데이터 전송 기술을 통해 테라바이트(TB/s) 급의 대역폭을 구현합니다 [출처: blogs.sw.siemens.com]. 이는 AI 워크로드에서 요구되는 방대한 양의 파라미터를 실시간으로 처리할 수 있는 최적의 균형(Bandwidth, Density, Energy Consumption)을 제공합니다 [출처: newsletter.semianalysis.com].

3. 패키징 기술의 고도화와 검사 중요성의 증대

HBM3의 성능은 단순히 설계 수치에만 의존하는 것이 아니라, 이를 물리적으로 구현하는 '첨단 패키징(Advanced Packaging)' 기술에 의해 결정됩니다. DRAM 다이를 수직으로 적층하고, 실리콘 관통 전극(Through-Silicon Via, TSV)을 통해 칩 간을 전기적으로 연결하며, 칩 사이의 간극을 메우는 언더필(Underfill) 공정 등은 극도로 높은 정밀도를 요구합니다.

특히 적층 단수가 높아짐에 따라 다음과 같은 기술적 도전 과제가 발생하고 있습니다:

- **열 관리(Thermal Management):** 고성능 연산 시 발생하는 막대한 열을 효과적으로 방출하기 위한 방열 구조 확보.
- **구조적 안정성(Structural Stability):** 적층 과정에서 발생하는 휨(Warping) 현상 및 물리적 변형 제어.
- **수율(Yield) 확보:** 수천 개의 미세한 TSV 연결점 중 단 하나의 결함만 발생해도 전체 스택을 폐기해야 하는 높은 불량 민감도.

따라서 HBM3 공정에서는 기존의 전공정(Front-end) 수준에 준하는 초정밀 검사 기술이 요구됩니다. 적층 전 개별 다이의 상태를 검증하는 Known Good Die(KGD) 확보부터, 적층 후의 전기적 특성 및 물리적 결함을 잡아내는

고정밀 광학/전기적 검사 솔루션은 HBM 제조 경쟁력을 결정짓는 핵심 요소가 되었습니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf].

본 보고서에서는 HBM3의 적층 구조와 설계 원리를 심층 분석하고, 이를 구현하기 위한 핵심 패키징 기술 및 공정 중 발생하는 주요 결함(Defect)을 고찰합니다. 나아가 이러한 난제를 해결하기 위한 고정밀 검사 솔루션과 AI 기반의 차세대 비전 검사 적용 방안을 제시함으로써, 향후 HBM4 및 하이브리드 본딩(Hybrid Bonding) 시대로의 기술 전환에 대한 대응 전략을 모색하고자 합니다.

HBM3 적층 구조 및 설계 원리

HBM3(High Bandwidth Memory 3)는 인공지능(AI) 및 고성능 컴퓨팅(HPC) 워크로드를 처리하기 위해 기존의 DDR(Double Data Rate) 메모리 구조를 근본적으로 탈피한 3D 적층(3D Stacking) 아키텍처를 채택하고 있습니다. 기존의 평면적(2D) 구조에서는 데이터 전송 경로가 길고 대역폭(Bandwidth) 확보에 한계가 있었으나, HBM3는 DRAM 다이(Die)를 수직으로 쌓아 올리고 이를 관통하는 미세 전극을 통해 데이터 전송 거리를 극단적으로 단축함으로써 초고속·대용량 데이터 처리를 실현합니다 [출처: hyeongcheol.tistory.com].

1. 3D Stacking 아키텍처와 DRAM Die의 수직 적층

HBM3의 핵심 설계 원리는 여러 개의 DRAM 다이를 수직으로 적층하여 단위 면적당 집적도를 극대화하는 것입니다. HBM3 규격에서는 기술의 성숙도에 따라 8단(8-High) 및 12단(12-High) 적층이 주를 이루며, 최근에는 16단 이상의 초고단 적층 기술도 논의되고 있습니다 [출처: 매뉴얼HBMAAdvancedInspectionSyst20260509001.pdf].

이 구조에서 각 DRAM 다이는 독립적인 메모리 셀을 보유하면서도, 적층을 통해 하나의 거대한 메모리 스택(Stack)을 형성합니다. 이러한 수직 적층 방식은 다음과 같은 설계적 이점을 제공합니다.

- **데이터 전송 경로의 최적화:** 칩 수평으로 배치할 때 발생하는 긴 배선(Interconnect) 길이를 수직 방향으로 전환함으로써 신호 지연(Latency)을 최소화합니다.
- **대역폭(Bandwidth)의 극대화:** 수천 개의 데이터 입출력 통로를 동시에 활용할 수 있는 넓은 인터페이스를 구축하여, HBM3E와 같은 최신 규격의 경우 1.2 TB/s 이상의 초고대역폭을 구현하는 기반이 됩니다 [출처: micron.com].
- **에너지 효율성 향상:** 데이터 이동 거리가 짧아짐에 따라 신호 전달에 필요한 전력 소모가 감소하며, 이는 단위 성능당 전력 효율(Performance per Watt) 개선으로 이어집니다 [출처: newsletter.semianalysis.com].

2. TSV(Through-Silicon Via)를 통한 수직 연결 기술

HBM3의 적층 구조를 완성하는 핵심 기술은 TSV(실리콘 관통 전극, Through-Silicon Via)입니다. 기존의 와이어 본딩(Wire Bonding) 방식이 금선(Gold Wire)을 이용해 칩의 가장자리를 연결하는 방식이었다면, TSV는 실리콘 웨이퍼 자체에 미세한 구멍을 뚫고 그 내부를 전도성 물질(주로 구리, Cu)로 채워 칩 내부를 관통하는 통로를 만드는 방식입니다 [출처: hyeongcheol.tistory.com].

TSV 기술은 HBM3의 고속 동작을 가능하게 하는 결정적인 요소로, 다음과 같은 메커니즘을 통해 작동합니다.

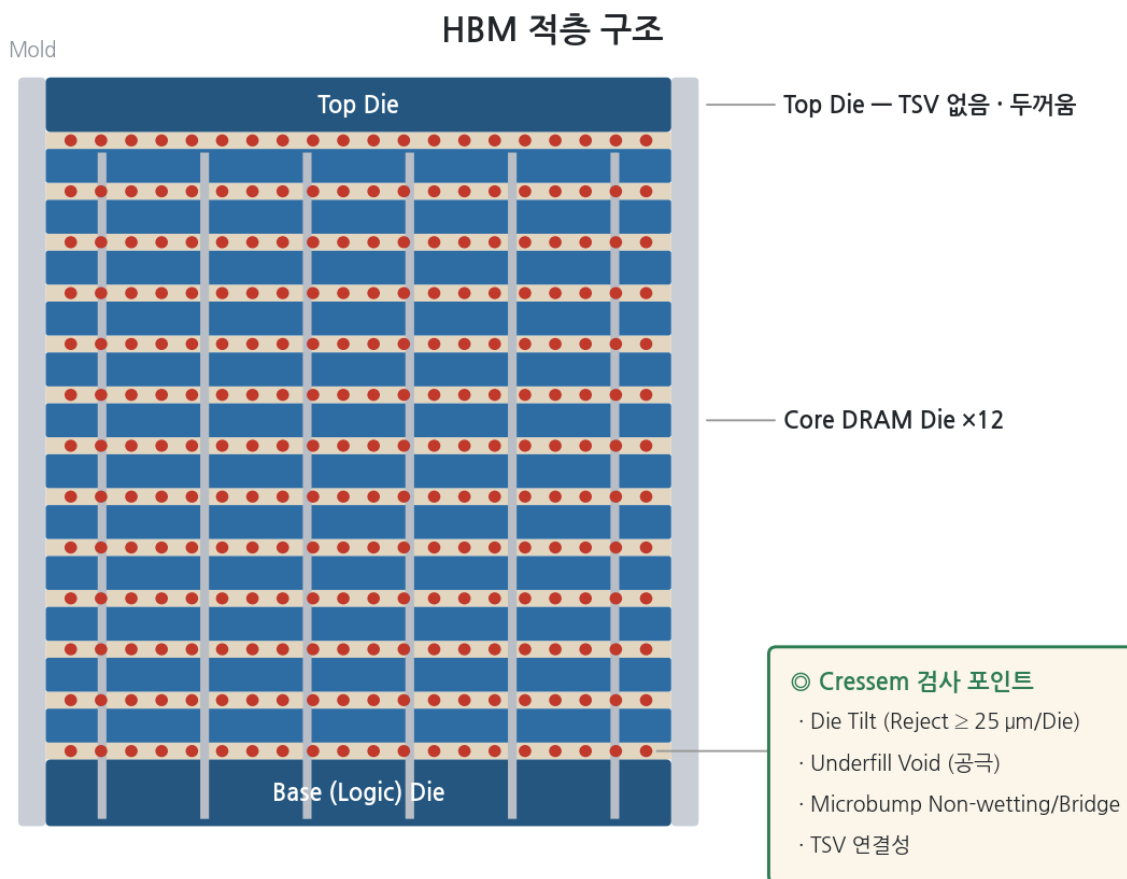
- **고밀도 I/O 구현:** 칩의 가장자리가 아닌 면(Face) 전체를 활용하여 수천 개의 TSV를 배치할 수 있습니다. 이를 통해 HBM3는 1024-bit 이상의 매우 넓은 데이터 버스(Bus) 폭을 확보할 수 있습니다 [출처: blogs.sw.siemens.com].
- **전기적 특성 최적화:** TSV는 신호 경로의 인덕턴스(Inductance)와 커패시턴스(Capacitance)를 낮추어 고주파수 대역에서의 신호 무결성(Signal Integrity, SI)을 유지하는 데 유리합니다.

- **열 관리(Thermal Management):** 수직으로 연결된 TSV는 데이터 경로 역할뿐만 아니라, 적층된 칩 사이에서 열을 수직으로 전달하는 열 통로(Thermal Path) 역할도 수행하여 고성능 동작 시 발생하는 발열을 제어하는 데 기여합니다.

3. Base Die와 Core Die의 역할 분담

HBM3 스택은 최하단의 **Base Die(또는 Logic Die)**와 그 위에 적층되는 여러 개의 **Core Die(DRAM Die)**로 구성됩니다.

- **Base Die (Logic Die):** 스택의 기반이 되는 이 다이는 메모리 컨트롤러와 HBM 스택 사이의 인터페이스 역할을 수행합니다. 단순한 물리적 지지체를 넘어, 데이터의 경로를 제어하고 신호를 정렬하며, 외부 인터페이스 규격에 맞게 데이터를 변환하는 로직 기능을 포함합니다. 최근 HBM4로의 전환 과정에서는 이 Base Die가 파운드리 공정의 로직 공정과 결합된 'Customized Logic Die' 형태로 진화하며, 메모리와 로직 간의 경계가 더욱 모호해지는 추세입니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf].
- **Core Die (DRAM Die):** 실제 데이터가 저장되는 메모리 셀 영역입니다. 적층 단수가 높아질수록 각 Core Die 사이의 간극을 어떻게 제어하고, TSV와 Micro-bump를 얼마나 정밀하게 정렬(Alignment)하느냐가 전체 수율을 결정짓는 핵심 요소가 됩니다.



Cressem · 결정론 렌더 · 읽기용 개략도

그림 1. HBM 적층 구조

4. 설계 시 고려되는 주요 물리적 변수

HBM3의 고단 적층 설계에서는 단순한 전기적 연결을 넘어, 물리적 안정성을 확보하기 위한 고도의 설계 최적화가 요구됩니다.

첫째, **Warpage(휨) 제어**입니다. 적층 단수가 12단, 16단으로 증가함에 따라 열팽창 계수(CTE) 차이로 인해 웨이퍼나 패키지가 휘어지는 현상이 심화됩니다. 이는 TSV의 정렬 오차를 유발하거나 본딩 품질을 저하시키는 원인이 됩니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf].

둘째, **Micro-bump의 정밀도**입니다. TSV를 통해 연결된 각 Die 사이에는 전기적 연결을 완성하기 위한 미세한 범프(Micro-bump)가 배치됩니다. 이 범프의 크기와 피치(Pitch)가 극도로 미세해짐에 따라, 접합부의 Void(공극)나 Bridge(단락) 현상을 방지하기 위한 정밀한 공정 설계가 필수적입니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf].

셋째, **열 방출 효율**입니다. 적층 구조 특성상 중심부에 위치한 Die는 외부로의 열 방출이 어려워 열 축적 문제가 발생할 수 있습니다. 이를 해결하기 위해 SK하이닉스의 MR-MUF(Mass Reflow-Molded Underfill)와 같이 액체 형태의 보호재를 채워 열전도율을 높이는 방식이나, 차세대 Hybrid Bonding(Cu-to-Cu) 기술을 통해 범프 없이 직접 구리를 접합하여 열 저항을 최소화하는 설계가 도입되고 있습니다 [출처: 보고서SK하이닉스HBM3E양산동향및기술경쟁력분석20260509001.pdf].

HBM3 주요 패키징 기술 비교

HBM3(High Bandwidth Memory 3)의 성능을 결정짓는 핵심 요소는 단순히 DRAM 다이(Die)를 수직으로 쌓는 것을 넘어, 적층된 칩 사이의 간극을 어떻게 메우고(Underfill), 열을 어떻게 효과적으로 배출하며(Thermal Management), 물리적 안정성을 어떻게 확보하느냐에 달려 있습니다. 현재 HBM 시장의 양대 패키징 기술인 TC-NCF(Thermal Compression Non-Conductive Film) 방식과 MR-MUF(Mass Reflow Molded Underfill) 방식은 각각의 구조적 특성에 따라 열 방출 효율, 수율(Yield), 그리고 공정 난이도 측면에서 뚜렷한 차이를 보입니다.

1. TC-NCF (Thermal Compression Non-Conductive Film) 기술 분석

TC-NCF 방식은 칩과 칩 사이에 비도전성 필름(Non-Conductive Film)을 삽입한 후, 열과 압력을 가하여 접합하는 방식입니다. 이는 기존의 범프(Bump) 접합 공정을 고도화한 형태로, 삼성전자가 12단 이상의 초고단 적층 시장을 공략하기 위해 지속적으로 고도화해 온 핵심 기술입니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf].

- **구조적 메커니즘:** 각 DRAM 다이 사이에 얇은 NCF 필름을 배치합니다. 이후 본딩 헤드가 상단 다이에 압력을 가하면서 동시에 열을 전달하여, 필름을 녹임과 동시에 범프 간의 전기적 연결을 완성합니다.
- **장점 및 특성:** NCF 필름이 칩 사이의 간격을 물리적으로 지지해주기 때문에, 적층 단수가 높아짐에 따라 발생할 수 있는 칩의 휨(Warping) 현상을 제어하는 데 유리합니다. 또한, 필름의 두께를 미세하게 조절함으로써 전체 패키지의 높이를 규격 내로 유지하기 용이합니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf].
- **한계점:** 필름이 칩 사이를 채우고 있기 때문에, 액체 형태의 소재를 사용하는 방식에 비해 열전도율(Thermal Conductivity)이 상대적으로 낮을 수 있습니다. 이는 고성능 연산 시 발생하는 발열을 외부로 배출하는 데 있어 병목 현상으로 작용할 수 있으며, 고온 동작 시의 열적 안정성 확보를 위한 정밀한 공정 제어가 필수적입니다.

2. MR-MUF (Mass Reflow Molded Underfill) 기술 분석

MR-MUF 방식은 SK하이닉스가 선도적으로 도입하여 HBM 시장의 기술적 변곡점을 만든 공정입니다. 이 방식은 칩들을 먼저 적층한 후, 칩 사이의 빈 공간에 액체 상태의 보호재(Underfill)를 대량으로 주입하여 굳히는 방식입니다 [출처: 보고서SK하이닉스HBM3E양산동향및기술경쟁력분석20260509001.pdf].

• **구조적 메커니즘:** 먼저 칩들을 수직으로 정렬한 후, 리플로우(Reflow) 공정을 통해 범프를 일괄 접합합니다. 이후 칩 사이의 미세한 틈새로 액체 형태의 MUF 소재를 침투시켜 전체 구조를 견고하게 고정합니다.

• **장점 및 특성:**

• **우수한 방열 성능(Thermal Management):** 액체 상태로 주입되는 MUF 소재는 기존 NCF 필름 대비 열전도율이 우수합니다. 이는 HBM3/3E와 같이 데이터 전송 속도가 9.2Gbps를 상회하는 고속 동작 환경에서 발생하는 막대한 열을 효과적으로 분산시켜, 칩의 열적 안정성을 극대화합니다 [출처: 보고서SK하이닉스HBM3E양산동향및기술경쟁력분석20260509001.pdf].

• **공정 효율 및 수율 향상:** 칩 적층 시 가해지는 물리적 압력을 최소화할 수 있어, 고단 적층 시 발생하는 휨(Warping) 현상을 줄이고 생산 수율을 높이는 데 기여합니다 [출처: 보고서SK하이닉스HBM3E양산동향및기술경쟁력분석20260509001.pdf].

• **한계점:** 액체 소재가 미세한 TSV(Through-Silicon Via) 주변이나 좁은 간극 사이로 균일하게 침투해야 하므로, 내부 공극(Void) 발생 여부를 확인하기 위한 고도의 검사 기술이 요구됩니다.

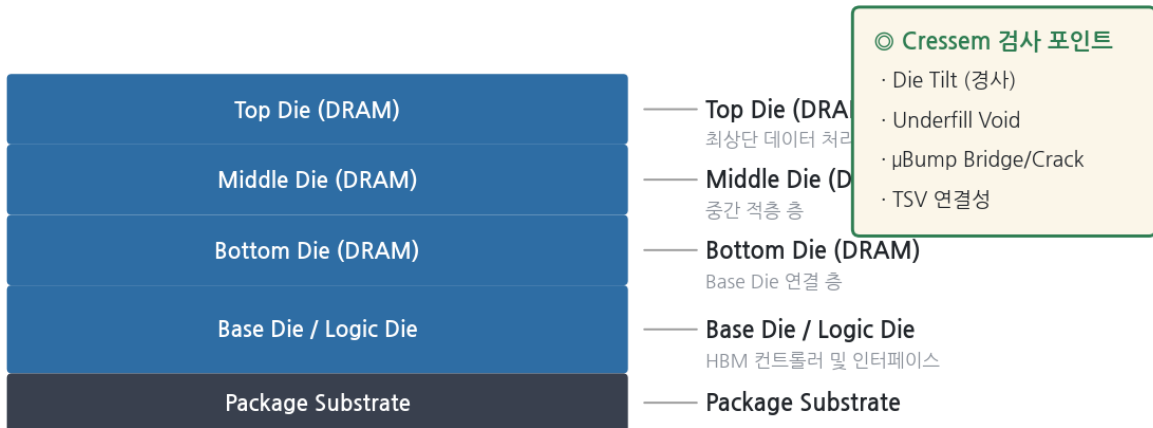
3. TC-NCF vs MR-MUF 기술 비교 요약

두 방식은 각각 '물리적 제어(NCF)'와 '열적/공정 효율(MUF)'이라는 서로 다른 강점에 집중하고 있습니다. 아래 표는 HBM3 공정 관점에서의 주요 성능 지표를 비교한 것입니다.

비교 항목	TC-NCF (Thermal Compression NCF)	MR-MUF (Mass Reflow Molded Underfill)
주요 적용 기업	삼성전자 등	SK하이닉스 등
방식 핵심	고체 필름 압착 및 열 압착	액체 보호재 대량 주입 및 리플로우
Warping(휨) 제어	우수 (필름의 물리적 지지력 활용)	양호 (압력 최소화 공정)
Thermal(방열) 성능	보통 (필름 소재의 열전도 한계)	매우 우수 (액체 Underfill의 높은 열전도)
공정 난이도	고단 적층 시 두께 제어 핵심	액체 침투 균일성 및 Void 제어 핵심
수율(Yield) 특성	정밀한 압력/온도 제어에 의존	공정 단순화 및 열 안정성을 통한 확보

표 1. TC-NCF vs MR-MUF 기술 비교 요약

HBM3 MR-MUF 적층 구조 단면도



Cressem · 결정론 렌더 · 읽기용 개략도

그림 2. HBM3 MR-MUF 적층 구조 단면도

4. 검사 기술적 시사점

패키징 방식의 차이는 검사 장비(Inspection Equipment)의 요구 사양을 결정짓는 결정적인 요인이 됩니다.

TC-NCF 방식에서는 필름의 두께 불균일성이나 압착 과정에서 발생하는 칩의 미세한 위치 오차(Misalignment)를 잡아내기 위한 고해상도 광학 검사가 중요합니다. 반면, **MR-MUF** 방식에서는 액체 소재가 칩 사이를 채울 때 발생하는 미세한 기포인 **Void(공극)**나, TSV 연결부의 전기적 무결성을 검증하는 것이 핵심 과제입니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf]. 특히 HBM4로 진화하며 적층 단수가 16단 이상으로 늘어날 경우, 이러한 공정별 결함 특성을 실시간으로 판별할 수 있는 AI 기반 비전 솔루션의 중요성은 더욱 증대될 것으로 전망됩니다.

적층 공정 주요 결함(Defect) 분석

HBM(High Bandwidth Memory)의 성능이 고단 적층(High-stacking) 기술을 통해 극대화됨에 따라, 적층 공정 중 발생하는 물리적·전기적 결함(Defect)은 수율(Yield)을 결정짓는 가장 치명적인 요소가 되었습니다. HBM3 및 차세대 HBM4로 진화할수록 적층되는 DRAM 다이(Die)의 수가 증가하고 전체 패키지의 두께는 제한되므로, 공정 중 발생하는 미세한 변형이나 불량은 전체 스택(Stack)의 폐기로 이어지는 막대한 경제적 손실을 야기합니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf]. 본 섹션에서는 적층 공정에서 발생하는 3대 핵심 결함인 Warpage(휨), Misalignment(정렬 오차), Void(공극)를 중심으로 그 발생 원인과 기술적 영향력을 심층 분석합니다.

1. Warpage (휨 현상)

Warpage는 적층 공정 중 발생하는 웨이퍼 또는 패키지 전체의 물리적 변형을 의미합니다. 이는 HBM의 고단 적층화가 진행됨에 따라 가장 심각하게 대두되는 이슈 중 하나입니다.

- **발생 원인 및 메커니즘:** HBM은 서로 다른 열팽창 계수(Coefficient of Thermal Expansion, CTE)를 가진 실리콘(Si) 다이, 언더필(Underfill) 소재, 그리고 기판(Substrate)이 층층이 쌓인 구조입니다. 공정 중 발생하는 열 사이클(Thermal Cycle) 과정에서 각 소재 간의 열팽창 차이로 인해 내부 응력(Internal Stress)이 축적되며, 이 응력이 해소되는 과정에서 구조물이 휘어지는 현상이 발생합니다 [출처: ...]

보고서SK하이닉스HBM3E양산동향및기술경쟁력분석20260509001.pdf]. 특히, TC-NCF(Thermal Compression Non-Conductive Film) 방식처럼 물리적 압력을 가하는 공정이나, 적층 단수가 12단 이상으로 높아질수록 누적되는 응력의 크기는 기하급수적으로 증가합니다.

- **공정 및 품질에 미치는 영향:**

- **Die-to-Die 접합 불량:** 힘이 발생하면 다이 간의 수평도가 깨지며, 이는 후속 본딩(Bonding) 공정에서 균일한 압력 전달을 방해합니다.

- **TSV 연결성 저하:** 물리적 변형은 실리콘 관통 전극(Through-Silicon Via, TSV)과 마이크로 범프(Micro-bump) 사이의 접촉 불량을 유도하여 전기적 단선(Open) 또는 단락(Short)을 유발합니다 [출처: HBM3 PPA Performance Evaluation by TSV Model with Micro-Bump and Hybrid Bonding | IEEE].

- **검사 난이도 상승:** Warpage가 심할 경우, 광학 검사 시 초점 심도(Depth of Field)를 벗어나게 되어 정확한 2D/3D 이미지 획득이 어려워지며, 이는 검사 데이터의 신뢰성 상실로 이어집니다 [출처: 매뉴얼HBMAdvancedInspectionSyst20260509001.pdf].

2. Misalignment (정렬 오차)

Misalignment는 적층되는 각 Die 또는 Bump의 중심축이 설계된 위치에서 벗어나는 결함을 의미합니다. HBM의 초고속 데이터 전송을 가능하게 하는 TSV 및 마이크로 범프 구조에서는 극도로 미세한 정렬 정밀도가 요구됩니다.

- **발생 원인 및 메커니즘:** 적층 공정 중 사용하는 본더(Bonder) 장비의 스테이지(Stage) 정밀도, Pick-and-place 과정에서의 물리적 충격, 그리고 열 변형에 의한 위치 이동 등이 주요 원인입니다. 특히 차세대 공정인 하이브리드 본딩(Hybrid Bonding)의 경우, 범프(Bump) 없이 구리(Cu)와 구리를 직접 맞닿게 하는 Cu-to-Cu 접합 방식을 사용하므로, 기존 범프 방식보다 훨씬 높은 수준의 정렬 정밀도가 요구됩니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf].

- **공정 및 품질에 미치는 영향:**

- **신호 무결성(Signal Integrity, SI) 저하:** 정렬 오차로 인해 접합 면적이 감소하거나 접촉 저항(Contact Resistance)이 불균일해지면, 고속 신호 전달 시 임피던스 불일치가 발생하여 데이터 전송 오류가 증가합니다.

- **전력 무결성(Power Integrity, PI) 문제:** 전력 공급을 위한 VDD/VSS 라인의 정렬이 어긋날 경우, 전류 밀도가 특정 지점에 집중되는 현상이 발생하여 소자의 수명 단축 및 열적 불안정성을 초래합니다.

- **기하학적 불량:** 미세 정렬 오차는 인접한 회로 패턴과의 간섭을 유발하여 전기적 단락(Short)을 일으키는 직접적인 원인이 됩니다.

3. Void (공극)

Void는 다이 사이의 접합부나 언더필(Underfill) 내부, 혹은 TSV 내부 등에 형성되는 미세한 빈 공간(Air Gap)을 의미합니다.

- **발생 원인 및 메커니즘:**

- **언더필 공정 이슈:** 액체 형태의 보호재를 주입하는 과정에서 기포(Bubble)가 포집되거나, 소재의 점도(Viscosity) 및 유동성 문제로 인해 특정 영역이 채워지지 않을 때 발생합니다. SK하이닉스의 MR-MUF(Mass Reflow-Molded Underfill) 기술은 액체 소재를 동시에 채워 이 문제를 완화하려 하지만, 여전히 미세 공정에서는 제어가 까다로운 영역입니다 [출처: HBM3-0315-1 - SK Hynix].

- **본딩 공정 이슈:** 하이브리드 본딩 시 Cu-to-Cu 접합면 사이의 이물질(Particle)이나 미세한 단차로 인해 완벽한 접합이 이루어지지 않을 때 공극이 형성됩니다 [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf].

- **공정 및 품질에 미치는 영향:**

- **열 방출 효율(Thermal Dissipation) 저하:** HBM은 고성능 연산 시 막대한 열이 발생합니다. Void는 열전도율이 매우 낮은 공기층을 형성하므로, 열 경로(Thermal Path)를 차단하여 국부적인 Hot-spot을 생성하고 이는 결국 칩의 열적 파괴로 이어집니다 [출처: HBM (High Bandwidth Memory)의 구조와 설계 원리].
- **구조적 취약성:** Void는 외부 충격이나 열 팽창 시 응력이 집중되는 지점이 되어, 패키지의 균열(Crack)이나 박리(Delamination)를 유발하는 시발점이 됩니다.

HBM 적층 결함 발생 및 영향 프로세스

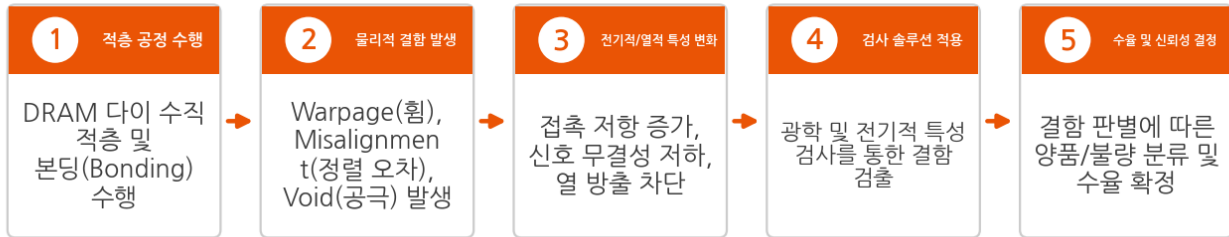


그림 3. HBM 적층 결함 발생 및 영향 프로세스

[요약: 주요 결함 특성 비교]

결함 항목	주요 발생 원인	핵심 영향 (Impact)	검사 핵심 기술
Warp(휨)	소재 간 CTE 차이, 열 응력 누적	접합 불량, TSV 단선, 검사 초점 이탈	3D AOI, 변형을 측정
Misalignment (정렬 오차)	장비 정밀도 한계, 열적 위치 이동	신호 무결성(SI) 저하, 단락(Short) 발생	Sub-micron급 정렬 검사
Void (공극)	언더필 기포, 접합면 이물질(Particle)	열 방출 저하(Hot-spot), 구조적 균열	고해상도 X-ray, 광학 비전

표 2. 요약: 주요 결함 특성 비교

고정밀 검사 솔루션 요구사항

HBM3(High Bandwidth Memory 3)의 적층 단수가 증가하고 데이터 전송 속도가 Gbps 단위로 급격히 상승함에 따라, 기존의 검사 방식으로는 미세 결함을 검출하거나 신호의 무결성을 보장하는 데 한계가 발생하고 있습니다. 특히 8단(8-Layer) 및 12단(12-Layer) 이상의 고단 적층 구조에서는 미세한 물리적 변형이나 전기적 특성 저하가 전체 패키지의 치명적인 불량으로 이어질 수 있습니다. 따라서 차세대 HBM 검사 솔루션은 Sub-micron(1 μ m 미만)급의 초고해상도 광학 검사 능력과 고속 신호 전달을 검증할 수 있는 전기적 특성 검사 기준을 동시에 충족해야 합니다.

1. 초고해상도 광학 검사(Optical Inspection) 요구사항

HBM3의 적층 공정, 특히 TSV(Through-Silicon Via) 및 마이크로 범프(Micro-bump) 기반의 접합부 검사를 위해서는 극도로 높은 수준의 광학적 정밀도가 요구됩니다.

- **Sub-micron급 AOI(Automated Optical Inspection) 성능:** HBM3는 수천 개의 TSV를 통해 칩 간 연결이 이루어집니다. 이 연결 부위의 미세한 이물질(Particle), 브릿지(Bridge, 단락), 또는 정렬 오차(Misalignment)를

탐지하기 위해서는 Sub-micron 단위의 해상도를 가진 2D/3D AOI 시스템이 필수적입니다. 특히 차세대 공정인 하이브리드 본딩(Hybrid Bonding)으로의 전환을 고려할 때, 구리(Cu) 접합면의 미세 결함을 잡아낼 수 있는 광학 엔진의 성능이 핵심적인 지표가 됩니다. [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf]

- **고정밀 정렬 및 위치 제어:** 다단 적층 시 발생하는 미세한 위치 오차를 측정하고, Die-to-Die 간의 정렬 상태를 실시간으로 모니터링할 수 있는 나노급 위치 정밀도 스테이지와 오차 보정 알고리즘이 요구됩니다. [출처: PDF Hbm 반도체 초고집적 하이브리드본딩 스택장비 기술 동향]

- **실시간 양불 판별을 위한 고속 비전:** 생산 라인의 Tact Time(생산 주기)을 극대화하기 위해서는 고해상도 카메라와 정밀 조명 제어 기술이 결합된 고속 광학 엔진이 탑재되어야 합니다. 이는 검사 데이터의 폭증 속에서도 실시간으로 양품과 불량(Void, Bridge, Misalignment)을 판별할 수 있는 기반이 됩니다. [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf]

2. 고성능 전기적 특성 검사(Electrical Test) 요구사항

HBM3는 데이터 전송 속도가 9.2Gbps 이상으로 매우 높으며, 1024-bit의 광대역 인터페이스를 사용하므로 신호의 품질을 검증하는 전기적 검사 기준이 매우 까다롭습니다.

- **Signal Integrity(SI) 및 전력 무결성(PI) 검증:** 고속 신호 전달 과정에서 발생하는 신호 왜곡, 노이즈, 전압 강하 등을 측정하여 신호 무결성(Signal Integrity)을 보장해야 합니다. 특히 HBM3e 및 HBM4로 진화함에 따라 데이터 전송 속도가 1.5TB/s(Target) 수준까지 높아질 것으로 예상되므로, 고속 신호 환경에서의 안정적인 데이터 전송 능력을 검증하는 것이 필수적입니다. [출처: HBM3e and HBM4: IC design guide for next-generation high bandwidth memory]

- **High-Speed Probe Card 및 테스터 기술:** 늘어난 채널 수와 고속 데이터 전송 속도를 지원하기 위해서는 고밀도/고성능 프로브 카드(Probe Card)와 테스터 기술이 뒷받침되어야 합니다. 이는 적층된 Die 간의 Die-to-Die 통신 및 전기적 연결성을 완벽히 검증하기 위함입니다. [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf]

- **Known Good Die(KGD) 확보를 위한 Pre-stacking Test:** HBM은 적층 후 불량이 발견될 경우 전체 스택을 폐기해야 하므로 경제적 손실이 매우 큼니다. 따라서 적층 공정 전, 개별 Die의 전기적 특성을 완벽히 검증하여 불량 Die가 적층되는 것을 원천 차단하는 KGD 확보 기술이 검사의 핵심 요구사항입니다. [출처: 보고서_삼성전자HBM4전환에따른검사기술동향및대응전략20260509001.pdf]

3. 검사 솔루션 핵심 기술 비교 요약

구분	주요 요구사항	핵심 기술 요소	검사 목적
광학 검사 (Optical)	Sub-micron 해상도, 3D 형상 측정	AOI, 초고해상도 카메라, 정밀 조명, AI 비전	미세 결함(Void, Bridge, Particle) 및 정렬(Alignment) 오차 검출

표 3. 검사 솔루션 핵심 기술 비교 요약

AI 기반 비전 검사 적용 방안

HBM(High Bandwidth Memory)의 적층 단수가 12단, 16단 이상으로 급격히 증가함에 따라, 기존의 Rule-based(규칙 기반) 검사 방식은 한계점에 도달했습니다. HBM 공정에서 발생하는 결함은 미세화된 패턴과 복잡한 3D 구조로 인해 매우 정밀하며, 검사 과정에서 발생하는 데이터의 양 또한 폭증하고 있습니다. 이러한

환경에서 수율(Yield)을 최적화하고 생산성을 극대화하기 위해서는 Deep Learning(딥러닝) 기반의 AI 비전 검사 솔루션 도입이 필수적입니다.

1. Deep Learning 기반 결함 분류(Defect Classification) 전략

전통적인 광학 검사 방식은 사전에 정의된 알고리즘(예: 크기, 모양, 명암 차이)에 따라 불량을 판별합니다. 그러나 HBM의 TSV(Through-Silicon Via) 주변이나 Micro-bump 접합부에서 발생하는 결함은 형태가 매우 불규칙하며, 정상적인 공정 노이즈와 구분이 모호한 경우가 많습니다.

AI 기반 결함 분류는 다음과 같은 메커니즘을 통해 검사 정밀도를 혁신합니다.

- **특징 추출(Feature Extraction)의 고도화:** CNN(Convolutional Neural Networks) 및 최신 Transformer 기반 아키텍처를 활용하여, 육안이나 단순 알고리즘으로 식별하기 어려운 미세한 결함의 패턴을 학습합니다. 이를 통해 Void(공극), Bridge(단락), Misalignment(정렬 오차), Particle(이물) 등의 결함을 높은 정확도로 분류합니다. [출처: 분석_크레셈CRESSEM20260509001.pdf]

- **비정상 탐지(Anomaly Detection):** 양품(Golden Sample) 데이터만을 학습시켜, 양품의 범주에서 벗어나는 모든 패턴을 결함으로 간주하는 방식을 적용할 수 있습니다. 이는 미처 정의되지 않은 새로운 유형의 결함(Unknown Defect)이 발생했을 때 매우 효과적입니다.

- **데이터 증강(Data Augmentation) 및 합성:** 고가의 HBM 웨이퍼나 패키지 샘플을 무분별하게 사용할 수 없으므로, GAN(Generative Adversarial Networks)과 같은 생성 모델을 활용하여 가상의 결함 이미지를 생성함으로써 학습 데이터셋의 불균형 문제를 해결합니다. [출처: pdf_인하대학교컴퓨터공학과배승환20260509001.pdf]

2. 과검(Over-kill) 방지 및 수율 최적화(Yield Optimization)

반도체 검사 공정에서 가장 큰 비용 손실 중 하나는 **과검(Over-kill)** 문제입니다. 과검이란 실제로는 양품임에도 불구하고, 검사 장비가 미세한 노이즈나 공정상의 허용

결론 및 향후 기술 전망

HBM(High Bandwidth Memory) 기술은 AI 가속기 시장의 폭발적인 수요에 대응하기 위해 단순한 메모리 성능 향상을 넘어, 패키징 기술의 혁신을 통해 그 가치를 증명하고 있습니다. HBM3 단계에서 확립된 수직 적층 구조와 고대역폭 데이터 전송 기술은 향후 차세대 규격인 HBM4로 진입함에 따라 더욱 급격한 기술적 패러다임의 변화를 맞이할 것으로 전망됩니다.

HBM4로의 진화는 적층 단수의 증가(16단 이상)와 더불어, 기존의 범프(Bump) 기반 접합 방식을 탈피한 **하이브리드 본딩(Hybrid Bonding, Cu-to-Cu)** 기술의 도입을 예고하고 있습니다. 하이브리드 본딩은 범프 없이 구리(Cu) 패드와 패드를 직접 맞붙이는 방식으로, 적층 높이를 획기적으로 낮추면서도 데이터 전송 효율을 극대화할 수 있는 핵심 기술입니다. 하지만 이러한 변화는 검사 공정의 난이도를 기하급수적으로 높이는 도전 과제를 안겨줍니다.

향후 HBM 기술 로드맵과 그에 따른 검사 기술의 대응 방향은 다음과 같습니다.

구분	HBM3 / HBM3E (현재)	HBM4 (차세대)	검사 기술 대응 방향
주요 본딩 방식	TC-NCF / MR-MUF	Hybrid Bonding (Cu-to-Cu)	범프 없는 접합면의 미세 정렬 및 Void 검사
적층 단수	8단 ~ 12단	16단 이상	고단 적층에 따른 Warpage(휨) 및 두께 제어

인터페이스	DRAM-centric	Customized Logic Die 결합	파운드리 공정 결합에 따른 Interface 검사
데이터 속도	9.2 Gbps+	1.5 TB/s+ (Target)	고속 신호 무결성(SI/PI) 및 전기적 특성 검증

표 4. 결론 및 향후 기술 전망

HBM4 시대의 검사 솔루션은 단순히 불량을 찾아내는 수준을 넘어, 다음과 같은 세 가지 핵심 역량을 갖추어야 합니다.

1. 초고정밀 광학 검사(Ultra-high Precision Optical Inspection): 하이브리드 본딩 도입 시 Cu-to-Cu 접합면의 미세한 이