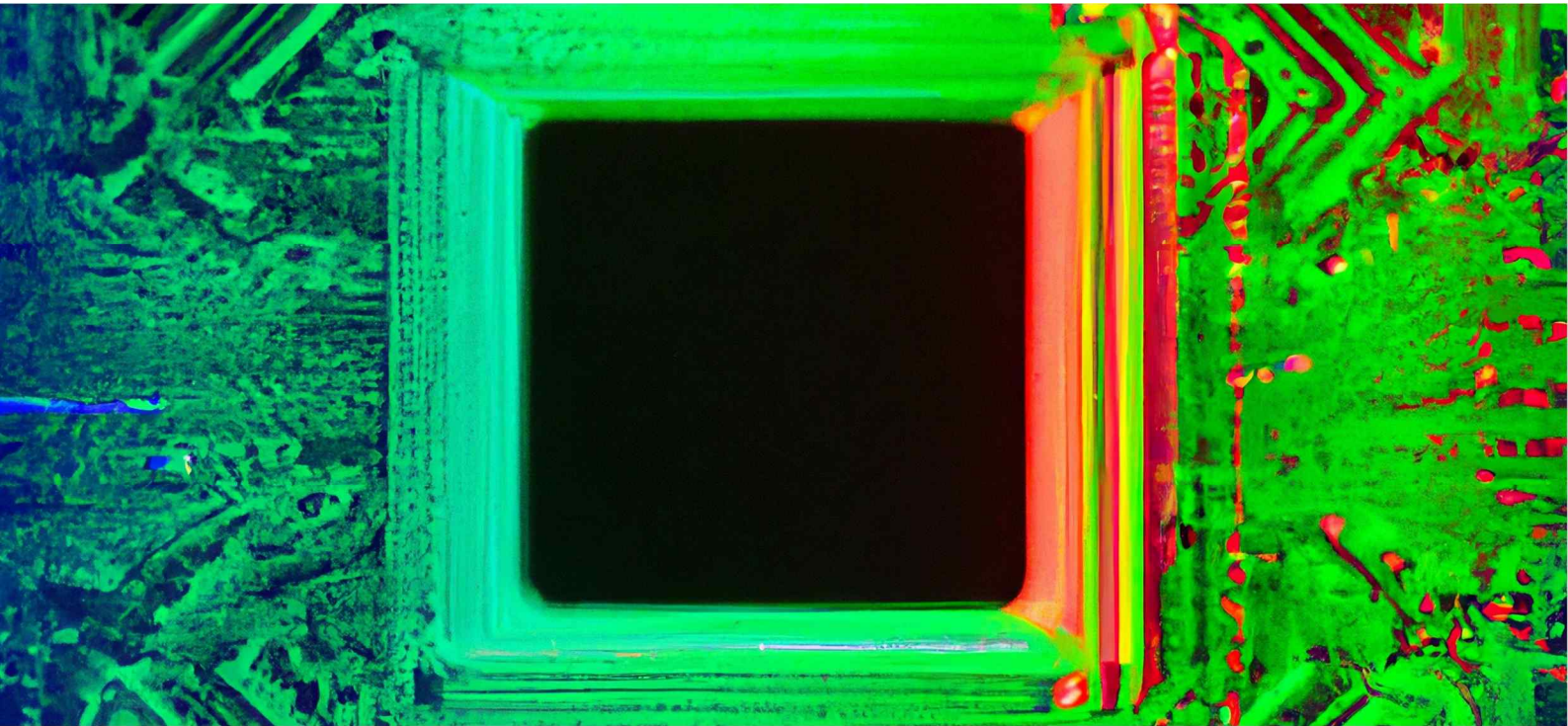




HBM 후공정 경쟁의 다음 격차는 장비가 아니라 냉각시간에서 만들어진다

[산업 분석] 반도체 후공정

조진현 쿨스 대표 jhcho@cools.co.kr

SUMMARY

HBM 경쟁의 무게중심이 전공정 미세화에서 적층·접합·열관리 등 첨단 후공정으로 이동했다. 2026년 3월 국내 장비기업이 2.5D 열압착 본더의 중국·대만 파운드리 공급 계획을 공개하면서 후공정 핵심장비의 접근 장벽은 사실상 해제됐고, 경쟁 기준은 장비를 확보했는지가 아니라 같은 장비에서 사이클타임과 열이력을 어떻게 관리하는지로 이동하고 있다. 국내 기업의 우위는 장비 접근성, 공정 노하우, 접합·냉각의 물리 원리 세 층으로 나뉜다. 앞의 두 층은 장비 공급 개방과 인력 이동으로 축소 경로에 들어섰고, 세 번째 층은 전 사업자가 공기제트 대류냉각을 쓰고 있어 아직 어느 곳도 차별화를 형성하지 못했다. 공기제트 냉각은 작동 유체의 물성 때문에 계면 열전달계수 $500\sim1,000\text{ W/m}^2\cdot\text{K}$ 에서 이미 상한이다. 고체 접촉 전도냉각은 동일 계면에서 20~50 배를 확보하며, 접합 1 회당 약 4 초인 냉각 구간을 0.5 초까지 줄이는 것이 물리 허용 최대치다. 쿨스의 서멀클러치는 1차 검증 장비에서 이 구간을 7.0 초에서 3.4 초로 51% 단축했다. 관건은 전환의 성립 여부가 아니라 시점이다.

1. 개요

고대역폭 메모리(High Bandwidth Memory, HBM) 경쟁의 무게중심이 전공정 미세화에서 적층·접합·열관리 등 첨단 후공정으로 이동하는 가운데, 2026년 상반기를 기점으로 후공정 핵심장비에 대한 접근성 격차가 빠르게 악화되는 국면에 진입했다. 장비를 확보했는지가 아니라 같은 장비에서 사이클타임과 열이력을 어떻게 관리하는지가 경쟁 기준으로 이동하고 있으며, 그 병목은 접합 후 냉각 구간에 놓여 있다.

2. 첨단 패키징 본더, 중국·대만 파운드리 공급 예정 공개

국내 장비기업 한미반도체는 2026년 3월 세미콘 차이나에서 2.5차원(2.5D) 열압착 본더 2개 기종을 공개하고, 해당 기종을 중국·대만 파운드리 기업에 공급할 예정이라고 발표했다. 2.5D TC 본더 40은 40밀리미터 × 40밀리미터 칩과 웨이퍼의 본딩을 지원하며, 120은 웨이퍼 및 기판 단위의 대형 인터포저를 처리한다.

주목할 지점은 대상 공정이다. 개방된 영역은 HBM 적층 공정에 국한되지 않고, 그래픽처리장치(GPU)·중앙처리장치(CPU)·HBM을 실리콘 인터포저 위에 단일 패키지로 통합하는 인공지능(AI) 패키징 본체까지 포함한다. 공급 대상도 메모리 기업이 아닌 파운드리 기업이다. 중국·대만의 AI 패키징 수탁 생산 역량이 함께 확장된다는 의미다.

앞서 2025년 6월에는 같은 기업이 중국 최대 D램 업체 창신메모리테크놀로지(CXMT)에 TC 본더를 납품한 사실이 보도된 바 있다. 다만 당시는 평가 단계의 개별 거래였다. 이번 사안은 공급 계획이 공개적으로 발표됐다는 점에서 성격이 다르다.

3. 규제는 전공정에, 시장은 후공정에

현행 대중국 통제는 전공정, 특히 노광 영역에 집중돼 있다. 극자외선(EUV) 영역은 접근이 차단돼 있고, 중국이 자체 개발한 액침 심자외선(DUV) 장비는 검증 단계다. 반면 후공정 접합 장비는 통제 목록의 핵심 축이 아니며 상업적 판단에 따라 공급된다.

그 결과 전공정 격차는 규제에 의해 유지되고, 후공정 격차는 시장에 의해 축소되는 비대칭이 성립한다. 국내 기업 우위의 상당 부분이 적층·접합·패키징 역량에 근거한다는 점을 고려하면 이 비대칭이 향하는 방향은 확정적이다.

장비 공급만으로 수율과 품질이 즉시 평준화되지는 않는다. 다만 동일하거나 유사한 장비 플랫폼을 쓰게 되면 온도·압력 제어, 다이 정렬 및 적층 자동화, 공정 데이터 축적, 장비 운영·유지보수 경험, 후공정 엔지니어 양성 속도의 격차는 시간 경과에 따라 축소된다. 모두 반복 시행으로 획득되는 성질을 갖는 항목이다.

4. 창신메모리, 자본시장 진입으로 조달 기반 확보

창신메모리는 2026 년 1 분기 글로벌 D 램 점유율 8%로 4 위에 올랐다. 전년 동기 3%에서 2 배 이상 확대된 수치이며, 매출은 700% 이상 증가했다. 7 월 27 일에는 상하이증권거래소 상장으로 579 억 2,000 만 위안, 약 12 조~13 조 원을 조달했다.

조달 자금은 우선 D 램 기술 고도화와 웨이퍼 생산능력 확대에 배정된다. 공개된 상장 자료에서 전용 HBM 사업은 별도 항목으로 명시되지 않았다. 다만 전공정 생산기반과 범용 D 램 현금흐름의 확대는 별도로 추진되는 HBM 및 첨단 후공정 투자를 뒷받침하는 자본적 기반을 형성한다. 이 영역에서는 개발 기간이 아니라 조달 기간이 추격 속도를 좌우한다.

상하이 신규 팹은 허페이 본사 대비 2~3 배 규모로 구축되고 있으며, 2026 년 하반기 장비 설치, 2027 년 본격 양산 일정 이 제시돼 있다. 업계에서는 전체 D 램 생산능력의 약 20%를 HBM3 생산에 배정할 가능성을 전망한다. 국내 메모리 3 사 및 미국 기업 출신 엔지니어 확보도 병행되고 있으며, 국내 인력만 약 200 명 규모로 보도됐다.

5. 우위의 3 층 구조: 두 층은 축소 경로, 한 층은 미형성

국내 기업이 첨단 패키징에서 확보한 우위는 성립 근거별로 세 층으로 분해된다. 제 1 층은 최신 세대 TCB 장비의 우선 확보와 장비기업과의 공동 개발 이력, 즉 장비 접근성이다. 제 2 층은 동일 장비에서 온도 프로파일, 가압 프로파일, 언더필 거동, 워피지 보상 조건을 축적한 데이터의 양, 곧 공정 노하우다. 제 3 층은 접합과 냉각의 물리 원리 자체다.

제 1 층은 공급 발표로 접근이 개방됐다. 제 2 층은 반복 생산과 인력 이동으로 이전되며, 200 명 규모의 인력 확보가 그 속도를 결정한다. 제 3 층에서는 현재 전 사업자가 예외 없이 열압착 후 공기제트 대류냉각을 채택하고 있어 어떤 사업자도 차별화를 형성하지 않았다.

여기에 축적의 역설이 겹친다. 공정 노하우는 그 노하우가 적용되는 물리 원리가 유지되는 한에서만 자산이다. 공기제트 냉각의 노즐 배치, 유량 분포, 유동 간섭, 온도 편차 보상 데이터는 고체 접촉 전도냉각 조건에서 재사용되지 않는다. 후자의 지배 변수는 계면 접촉 압력, 표면 조도, 접촉 열저항, 냉각체 열용량, 접촉·분리 사이클의 기계적 반복도로, 물리적으로 다른 변수 집합이다. 상실하는 자산의 절대량이 큰 쪽, 즉 축적이 많은 사업자의 전환 속도가 느려진다.

6. 공기제트 냉각은 이미 물리적 상한

냉각 성능은 최종적으로 계면 열전달계수(h)로 결정되고, 공기제트 충돌 냉각에서 h 의 상한은 작동 유체의 물성이 규정한다. 공기의 열전도율은 상온 0.026 W/m·K, 200°C 근방 0.037 W/m·K 이다. 열경계층 두께를 물리적으로 도달 가능한 최소치까지 줄여도 제트 충돌 영역의 h 는 500~1,000 W/m²·K 범위에 머문다.

이 상한은 노즐 설계, 분사 압력, 유량 증대로 개선되지 않는다. 유량을 높이면 소음, 진동, 파티클 비산, 인접 셀 간 유동 간섭이 증가해 접합 정밀도가 먼저 제약된다.

[표 1] 냉각 방식별 계면 열전달 성능 비교

구분	공기 자연대류	공기제트 충돌냉각	고체 접촉 전도냉각
작동 매체	공기	공기	고열전도 고체
매체 열전도율(W/m·K)	0.026	0.026~0.037	170~400
계면 h (W/m ² ·K)	5~25	500~1,000	20,000~50,000
상대 성능비	1	약 40	약 1,500
잔여 개선 여지	없음	없음	큼

* 고체 접촉 전도냉각의 h는 계면 접촉 압력, 표면 조도, 계면 매개층 조건에 의해 결정되며, 상기 범위는 계면 설계가 확보된 조건의 값이다. 자료: 쿨스

상한의 결과는 시간 구조로 나타난다. HBM 접합 1 사이클은 정렬, 승온, 가압 접합, 냉각, 헤드 상승으로 구성되는데, 승온과 가압은 히터 출력과 하중 제어로 단축돼 온 반면 냉각 구간은 단축되지 않아 현재 사이클에서 가장 구간을 점유한다. 장비 1 대당 시간당 처리량이 냉각 구간에 의해 결정되고, 적층 단수가 늘수록 먼저 접합된 하부 다이가 후속 사이클의 열에 반복 노출돼 누적 열노출량이 단수의 제공에 근접하는 형태로 증가한다. 생산성과 수율·신뢰성이 동일한 물리 변수에 묶여 서로 교환 관계에 놓이는 구조다.

7. 전도냉각의 물리 허용 최대치는 0.5 초 수준

냉각 구간 소요 시간은 열저항과 열용량의 곱, 즉 시상수에 비례한다. 계면 h가 500 W/m²·K에서 25,000 W/m²·K로 증가하면 계면 열저항은 1/50로 감소한다. 냉각체 내부 확산 저항과 방열부 배출 저항을 포함한 직렬 열저항 기준으로 산정해도 전체 시상수는 1/8 이하로 줄어든다. 접합 1 회당 약 4 초가 소요되는 냉각 구간을 0.5 초 수준까지 축소하는 것이 물리 허용 최대치이며, 이 값은 냉각 장치의 크기나 출력, 유량과 무관하게 결정된다.

실현 조건은 수치로 규정된다. 계면 접촉 열전달계수 20,000 W/m²·K 이상, 냉각체 열전도율 170 W/m·K 이상, 접촉·분리 사이클 시간 0.1 초 이하, 접촉 위치 반복도 ±10 마이크로미터 이내, 접촉 시 칩 위치 교란 1 마이크로미터 이내, 1 사이클 냉각체 온도상승 5K 이하다. 마지막 조건이 구조를 규정한다. 고체 접촉은 냉각에 유리한 동시에 승온에 불리하므로, 상시 접촉형 냉각판으로는 성립하지 않는다. 접촉과 분리를 사이클 내에서 능동 전환하는 구조가 전제된다.

반도체 접합·방열·계면 열제어 기술기업 쿨스가 개발한 서멀클러치는 이 조건에 맞춰 구성된 열제어 모듈이다. 승온·가압 구간에서는 냉각체를 가열부로부터 기계적으로 분리해 열손실 경로를 차단하고, 가압 접합이 완료된 시점에 냉각체를 정밀 접촉시켜 고체 전도경로로 열을 흡수한 뒤 방열부로 배출하고 분리한다.

“기존 냉각 장치와의 차이는 성능의 정도가 아니라 열전달 경로의 종류다. 공기를 사이에 둔 간접 냉각에서 고체가 직접 열을 전달하는 냉각으로 원리를 교체하는 것”이라고 조진현 쿨스 대표이사는 설명했다. 조 대표는 미국에서 열전달공학 박사학위를 취득한 뒤 국내 반도체 기업에서 28년간 열관리 및 기술개발을 담당했다.

1차 검증 장비에서는 TCB 냉각 구간 소요 시간이 7.0 초에서 3.4 초로 단축됐다. 단축률 51%로, 동일 접합 품질 기준 유지 조건에서 측정된 값이다. 계면 접촉 조건이 요구 수치에 도달할 경우 접합 1 회당 냉각 구간은 0.5 초 수준이 설계 도달점이며, 본딩 장비 1대당 생산성은 최대 3배, 다층 적층 시 하부 접합부 누적 열노출은 약 85% 감소한다는 것이 회사 측 설명이다.

“생산성과 누적 열노출이 동시에 개선되는 이유는 두 항목이 동일한 냉각 시상수에 의해 결정되기 때문”이라며 “생산성과 수율을 교환하지 않고 함께 개선한다. 교환 관계 자체가 해소된다”고 조 대표는 말했다.

8. 고단화·대면적화가 전환 필요성을 강화

전도냉각의 필요성은 경쟁 상황과 무관하게 제품 규격의 진행 방향만으로도 강화된다. 국제반도체표준협약기구(JEDEC)는 HBM 패키지 높이 기준을 기존 약 775 마이크로미터에서 900 마이크로미터 수준으로 확대하는 방안을 검토하고 있다. 높이 여유의 확대는 적층 단수 증가를 전제한 조치이며, 단수가 늘면 하부 접합부의 열이력은 규격 진행과 함께 악화된다.

차세대 HBM은 D램 다이 면적을 확대하는 방향으로 진행되고 있고, 이에 대응하는 대면적 본딩 장비도 이미 공개됐다. 다이 면적이 커지면 실리콘관통전극(TSV)과 입출력 인터페이스 수를 늘릴 수 있어 용량과 대역폭이 확장되는 반면, 온도 편차의 절대값이 증가하고 열팽창계수 불일치에 의한 워피지가 확대된다. 공기제트 냉각은 노즐 배치의 이산성 때문에 대면적에서 온도 균일도를 확보하기 어렵고, 고체 접촉 전도는 면 단위로 균일한 열전달 경로를 형성한다.

2.5차원 및 3차원을 포함한 첨단 패키징 시장은 2024년 460억 달러에서 2030년 794억 달러로 확대될 전망이다. 전환 판단의 대상 규모는 이 시장의 생산성 기준 전체다.

9. 결과를 가르는 변수는 성립 여부가 아니라 시점

전환 순서에 따른 결과는 세 갈래로 나뉜다. 국내 기업이 선행 전환하면 냉각 원리 영역에서 새로운 격차가 형성되고, 후 공정 장비 평준화가 완료돼도 동일 장비 내 시간 구조의 차이가 잔존해 처리량과 고단 적층 수율의 우위가 유지된다. 차세대 본딩 장비 규격에 대한 주도권도 함께 확보된다.

동시 전환의 경우 냉각 원리에서는 격차가 형성되지 않고, 경쟁은 다시 미세공정 수율과 생산능력 규모로 회귀한다. 후자는 국가 단위 자본 투입 규모의 문제다. 후발 기업이 선행 전환하면 공기제트 영역에서 축적한 우위의 상당 부분이 경쟁력으로 환산되지 않는다. 기존 기술 단계를 순차적으로 경유하지 않고 새 원리로 직접 이동하는 경로가 성립하기 때문이다.

“플랫폼 전환기에 결정적으로 작용한 변수는 축적의 크기가 아니라 전환 판단의 시점이었다”며 “피쳐폰 시대의 기술적 우위가 스마트폰 시대의 시장 지위로 이전되지 않은 것과 같은 구조”라고 조 대표는 지적했다.

실행 경로로는 단기와 중기가 제시된다. 단기적으로는 생산라인 전체를 변경하기 전에 단일 장비 또는 평가 장비에 모듈을 장착해 냉각 구간 소요 시간, 헤드 온도 균일도(면내 편차 $\pm 3K$ 이내), 접촉·분리 반복정밀도(± 10 마이크로미터 이내), 접촉 시 칩 위치 교란(1 마이크로미터 이내), 워피지, 하부 접합부 누적 열이력, 시간당 처리량, 기존 레시피 호환성을 확인할 수 있다. 중기적으로는 본딩 헤드, 히터, 냉각체, 방열부가 하나의 시간제어 시스템으로 작동하는 전용 장비 구조 개발이 과제로 남는다. 장비기업과 메모리 제조기업이 공동으로 사양을 확정하는 경로에서 규격 주도권이 결정된다.

쿨스는 반도체 접합·방열·계면 열제어 분야에서 1,500 건 이상의 특허 포트폴리오를 보유하고 있으며, 계면 접촉 조건, 클러치 구동 구조, 냉각체 소재 구성, 열배출 경로가 각각 권리화돼 있다. 사업 구조는 IP 라이선싱, 모듈·부품 공급, 장비 및 공정 솔루션의 3 개 축으로 구성된다. 기존 장비 개조형 모듈 공급 방식은 전환 결정의 최대 저항 요인인 기존 설비 매몰 비용 부담을 축소한다.

HBM 세대 격차와 미세공정 수율 격차는 유지되고 있다. 그러나 후공정 장비가 보급되면 장비를 확보하는 것보다 장비의 생산성을 얼마나 극대화하는가가 경쟁요소가 된다. 지금 필요한 판단은 전도냉각이 언젠가 필요한가가 아니라, 후공정 평준화가 본격화되기 전에 누가 먼저 이를 양산공정에 적용해 새로운 생산성 기준을 선점하는가에 있다.

자료 출처

- 주간한국·녹색경제신문, 세미콘 차이나 2026 2.5D TC 본더 40·120 공개 및 중국·대만 파운드리 공급 예정 발표, 2026.3.25.
- 카운터포인트리서치 메모리 트래커, 글로벌 D 램 매출 및 점유율, 2026.5. (2026 년 1 분기 글로벌 D 램 시장 970 억 달러)
- 창신메모리테크놀로지 상하이증권거래소 상장 공시, 2026.7.27.
- 아이뉴스 24, JEDEC 패키지 높이 규격 검토 및 율그룹 첨단 패키징 시장 전망, 2026.5.4.
- 시사저널 e, CXMT 국내 엔지니어 확보 규모, 2026.6.2.
- 파이낸셜뉴스, D 램·HBM 격차 및 추격 속도, 2026.7.29.
- 쿨스 내부 검증 데이터, TCB 냉각 구간 실측 결과, 2026.