



리서치센터 리포트
바로가기

Tech (OVERWEIGHT)

HBM 디스펙 논란, 용량과 대역폭의 분리

- HBM의 단수는 12단에서 8, 4단으로 축소. 다만 사양 축소는 용량이며 대역폭 요구는 오히려 강화. HBM 사양 변화는 메모리 계층화 과정에서 접근할 필요
- HBM의 부가가치는 코어다이에서 베이스다리로 이동하고 있으며 베이스다이의 사양은 계속 확장 중
- HBM 사양 변화는 단기적으로 메모리 기업에 중립, 장기적으로 베이스다이 등 로직 영역으로의 확장 기회. 삼성전자에 상대적으로 유리한 구조 변화

WHAT'S THE STORY?

끝나지 않는 HBM 사양 축소 논란: 7월부터 시작된 HBM 사양 축소(despec) 논란이 9월까지 이어지고 있다. 엔비디아는 차세대 제품에서 기존 12단 중심의 HBM 구성을 8단까지 낮추는 방안을 검토하고 있으며, 최근에는 AI 고객사로부터 4단 적층 구조의 주문이 시장에서 거론되고 있다. 공개적으로 확인된 Rubin Ultra의 검토안에는 HBM4와 HBM4E의 8단 제품이 포함되고 있다.

우리는 HBM의 적층 사양 축소가 정해진 방향성이라고 믿는다. 그러나 이 변화를 단순한 HBM 수요 감소로 해석할 것이 아니라 AI 반도체에서 메모리의 역할이 계층별로 분화되는 첫 단계라고 판단한다.

AI의 발전은 계산의 계층화를 요구: 최근 Frontier AI가 보여주는 방향은 단순히 모델의 크기를 키우는 것이 아니다. 긴 시간 동안 여러 단계를 거쳐 문제를 해결하고, 필요한 연산만 선택적으로 수행하며, 컨텍스트를 반복적으로 불러와 사용하는 방향으로 발전하고 있다. 양자화, KV cache 압축과 재사용, 선택적 활성화(sparse computation)와 같은 기술도 같은 방향이다. 이러한 변화에서는 매 순간 거대한 모델 전체를 GPU 위에 올려놓는 것보다 필요한 데이터를 얼마나 빠르게 호출하고 이동시키느냐가 중요해진다.

AI 계산의 계층화가 메모리의 계층화를 촉발: 데이터가 어디에 상주하고 어디에서 계산되는지에 따라 메모리의 위치가 달라진다. GPU와 가장 가까운 HBM에는 지연속도와 대역폭에 민감한 데이터가 남고, 상대적으로 덜 민감한 데이터는 서버디램과 SOCAMM으로 이동한다. 더 장기간 유지하거나 반복적으로 재사용하는 KV cache는 CMX와 같은 별도 계층까지 확장된다. 모든 데이터를 HBM에 상주시키는 것은 HBM의 높은 원가와 제한된 공급을 감안할 때 가능하지도 않고 경제적으로 필요하지도 않다. AI 시스템은 이미 SRAM, HBM, 서버디램/SOCAMM, context storage를 하나의 연속된 계층으로 활용하기 시작했다. NVIDIA가 CMX를 도입하고 NVHBM으로 메모리 인터페이스 구조 자체를 바꾸려는 것도 같은 방향이다.

(다음 페이지에 계속)

HBM 디스펙에는 공급 부족의 영향도 존재: 이 변화가 반도체 공급 부족 때문인지, 서버 BOM 최적화 때문인지, 순수한 알고리즘 발전 때문인지를 명확히 구분하기는 어렵다. 아마 세 가지가 동시에 작용했을 것이다. 다만 공급 제약이 분명히 영향을 미쳤다는 정황은 있다. 어떤 메모리 업체도 디스펙 이후 27년 HBM 수요를 하향하지 않았다는 점이다. 엔비디아가 HBM 12단 대신 8단을 쓰더라도 2027년에 구매할 총 HBM 용량은 비슷하다. 가격도 근거 중 하나이다. 27년 HBM 가격은 26년 대비 평균 50% 인상이 예상된다. 가격이 오르는 국면에서 사양을 낮추면서도 물량을 줄이지 않는다는 것은 HBM 용량 쪼개기가 있었다는 증거다. 애초에 대역폭은 HBM의 가장 큰 장점이므로, HBM을 충분히 구할 수 있다면 쓰지 않을 이유가 없다. 결국 AI 모델 기술의 변화와 만성적 HBM 쇼티지가 복합적으로 작용한 변화로 보는 것이 타당하다. 단수 하락이 HBM형 디램 웨이퍼를 범용으로 돌려 공급량 증가 효과가 있을 것이라는 반론이 가능하다. 4단 제품의 목적 중 하나는 동일한 HBM 예산으로 가속기 대수를 늘리는 데 있고, HBM의 쇼티지는 27년에도 풀리지 않을 것이다.

HBM에서 대역폭과 용량의 역할 분리: HBM은 애초부터 용량보다 대역폭을 위해 만들어진 디램이다. 이름 그대로 High Bandwidth Memory다. 다만 23년 이후 LLM의 파라미터와 컨텍스트 규모가 급격히 확대되면서 HBM에는 대역폭뿐 아니라 용량 확대도 동시에 요구되었을 뿐이다. 하지만 메모리 계층이 세분화되기 시작한다면 두 역할을 반드시 하나의 메모리가 담당할 필요는 없게 된다. HBM은 GPU에 가장 가까운 계층에서 낮은 지연속도와 높은 대역폭을 담당하고, 대규모 용량 담당은 서버디램, SOCAMM, CMX 및 SSD로 분산할 수 있다. AI의 성능 개선 과정에서 총 메모리 용량의 중요성이 사라지는 것이 아니라, 필요한 용량을 서로 다른 메모리 계층이 나누어 담당하는 것이다.

HBM 디스펙의 본질은 대역폭의 디스펙이 아니라 용량의 디스펙에 가깝다. 오히려 핀 스피드와 스택 대역폭에 대한 요구는 계속 높아지고 있다. 실제로 SK하이닉스는 HBM4E에서 핀당 최대 16Gbps를 구현했다고 발표했고, 삼성전자 역시 HBM4E에서 14Gbps, 향후 16Gbps까지의 확장을 제시하고 있다. 엔비디아의 HBM4E의 큐브당 용량 요구량은 줄었지만 대역폭 요구량은 늘었다.

HBM의 가치 중심은 코어 다이에서 베이스 다이로 이동: 이 변화는 HBM 내부에서도 역할 분화를 일으킨다. 과거 HBM 경쟁에서 가장 눈에 띄는 지표는 코어다이의 공정과 적층 수였다. 8단에서 12단, 16단으로 올라가는 것이 제품의 용량과 ASP를 결정했다. 그러나 용량 중요도가 둔화되고 대역폭이 더 중요해지면 차별화의 중심은 베이스 다이로 이동한다.

삼성전자가 Hot Chips 2026에서 제시한 Custom HBM의 방향은 이를 잘 보여준다. 기존 HBM에서는 메모리 컨트롤러가 XPU(AI반도체)에 존재하고 PHY를 통해 HBM과 통신하지만, 삼성전자는 메모리 컨트롤러를 HBM의 베이스다이로 옮기고 기존 PHY를 보다 작은 D2D 인터페이스로 대체하는 구조를 제안했다. 이를 통해 XPU 면적도 아끼고 통신 효율도 개선할 수 있다. 엔비디아의 NVHBM도 같은 방향이다. 역시 커스텀 메모리 컨트롤러를 HBM 베이스 다이로 이동시킨다.

베이스다이의 고부가화, 코어다이의 범용화: 코어다이가 완전한 범용 기술이라고 말할 수는 없다. 특히 방열 구조는 코어다이의 업체별 핵심 차별화 요소이다. 다만 과거에는 코어다이 적층 경쟁력이 결정적이었지만 이제 베이스다이의 로직과 인터페이스로 이동하고 있을 뿐이다. 하이닉스와 마이크론은 27년 HBM4E부터 TSMC의 5nm 파운드리를 사용할 것으로 보인다. 분명한 스펙 업그레이드이다. 물론 엔비디아 등 고객사가 IP와 구조를 제공할 가능성이 높지만, 기본적으로 베이스다이의 디자인과 개발, 샘플은 메모리 업체들이 담당하고 있고, SK하이닉스는 TSMC를 넘어 인텔 파운드리를 이용할 생각도 하고 있다.

예전에는 HBM 가격을 Gb당 가격으로 비교했지만 앞으로는 단순히 Gb당 원가로 비교하기 힘들 것이다. 적층 수가 줄어 디램 용량이 감소하더라도 고성능 logic base die의 가치가 상승한다면 HBM의 Gb당 가격은 오히려 높아질 수 있다.

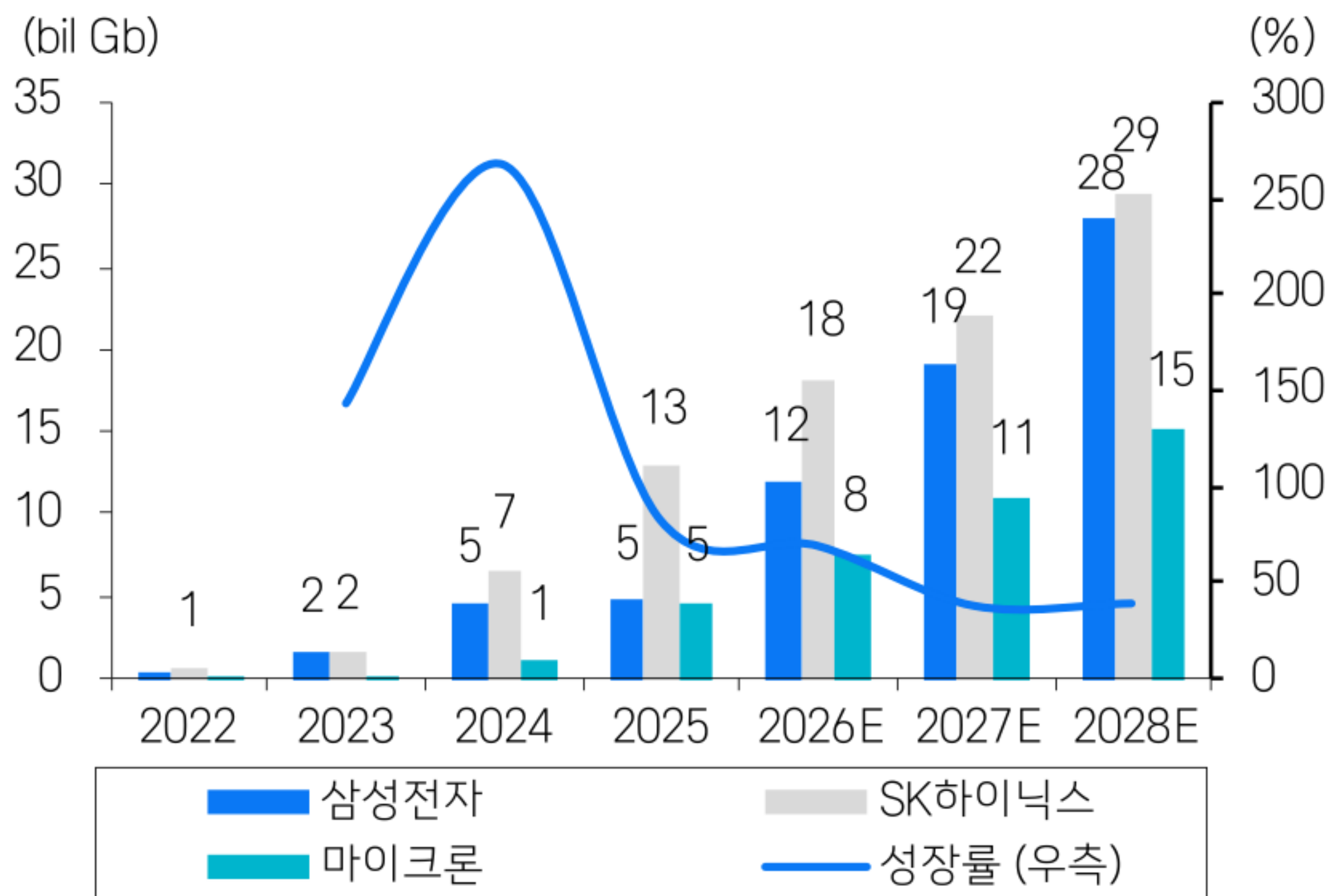
메모리 계층화의 방향성: HBM 4/8단 디스팩과 베이스다이 고도화는 별개의 사건이 아니다. HBM의 대역폭/용량 기능 분화를 드러낸 사건이고, 그 분화는 코어 다이와 베이스 다이의 역할 분리로 이어진다. 더 확장하자면 HBM 디스팩을 다른 메모리 사양과 구분해서 생각할 수도 없다. 고객사들은 HBM+DIMM+SSD의 배분 문제로 AI서버의 메모리 구성을 하게 될 것이며 이 세 종류의 메모리를 모두 만드는 메모리 3사의 TAM은 변하지 않는다. HBM의 용량이 줄어드는 만큼 다른 메모리 계층의 수요가 증가한다면 삼성전자, SK하이닉스, Micron처럼 디램과 낸드를 모두 생산하는 종합 메모리 업체의 전체 TAM에 미치는 영향은 HBM 단품 수요 감소보다 훨씬 작을 수 있다.

리스크 - 데이터 이동의 주도권과 표준화: 메모리 계층화가 메모리 업체에 무조건적인 기회만을 의미하지는 않는다. 메모리가 여러 계층으로 나뉘수록 가장 중요한 부가가치는 계층 간 데이터를 누가 이동시키고 통제하느냐에서 발생한다. 데이터 이동의 컨트롤은 모델 압축, 풀링, 티어링, 간단한 프로세싱, 스트리밍 오프로드까지 모두 관여해야 한다. 현재까지 이 영역의 주도권은 메모리 업체보다는 NVIDIA, 브로드컴, 마벨과 같은 컨트롤러 업체에 더 가깝다. Astera Labs와 Montage Technology 등도 메모리의 연결과 데이터 이동 컨트롤 영역으로 사업을 확장하고 있다. HBM 베이스다이가 앞으로 데이터 이동 측면에서 더 많은 역할을 담당하겠지만 전체 시스템의 데이터 컨트롤 중심이 곧바로 메모리 업체로 이동한다고 보기는 어렵다.

또 하나의 변수는 인터페이스 표준화다. 커스텀 HBM이 확대되더라도 D2D 인터페이스에 대한 산업 표준을 만든다면 메모리 업체 차별화가 제한될 수 있다. 엔비디아의 NVHBM은 베이스다이 부가가치를 엔비디아가 가져갈 가능성도 함께 열어 준다. 반대로 고객마다 다른 인터페이스를 요구한다면 HBM은 고객별 특화된 제품으로 발전할 가능성이 높아진다. 아직 결론을 내리기에는 이르다.

메모리, HBM 디스팩을 걱정할 필요는 없다: HBM 디스팩은 메모리의 중요도 하락을 의미하지 않을 뿐더러 HBM 캐파의 하향이나 공급 과잉 전환을 의미하지도 않는다. HBM은 여전히 AI메모리 계층화의 중심에 있다. 삼성전자의 발언대로 HBM에서의 경쟁 우위가 AI 시대의 메모리 기술 우위로 이어지기 때문에 메모리 업체들은 여전히 HBM캐파를 후순위로 미루지도 않을 것이다. 특히 단수 증가가 제한적이고 베이스다이가 고도화된다는 것은 고단화에 불리한 NCF방식의 수율 개선이 가능하고, 베이스다이 설계기술과 파운드리의 이익 레버리지가 가능한 삼성전자에 유리한 기술 변화이다. 우리는 이 변화를 단기적으로 메모리에 중립, 장기적으로 메모리의 영역 확장 기회, 특히 삼성전자에 유리한 구도라고 판단한다.

HBM 출하량 전망



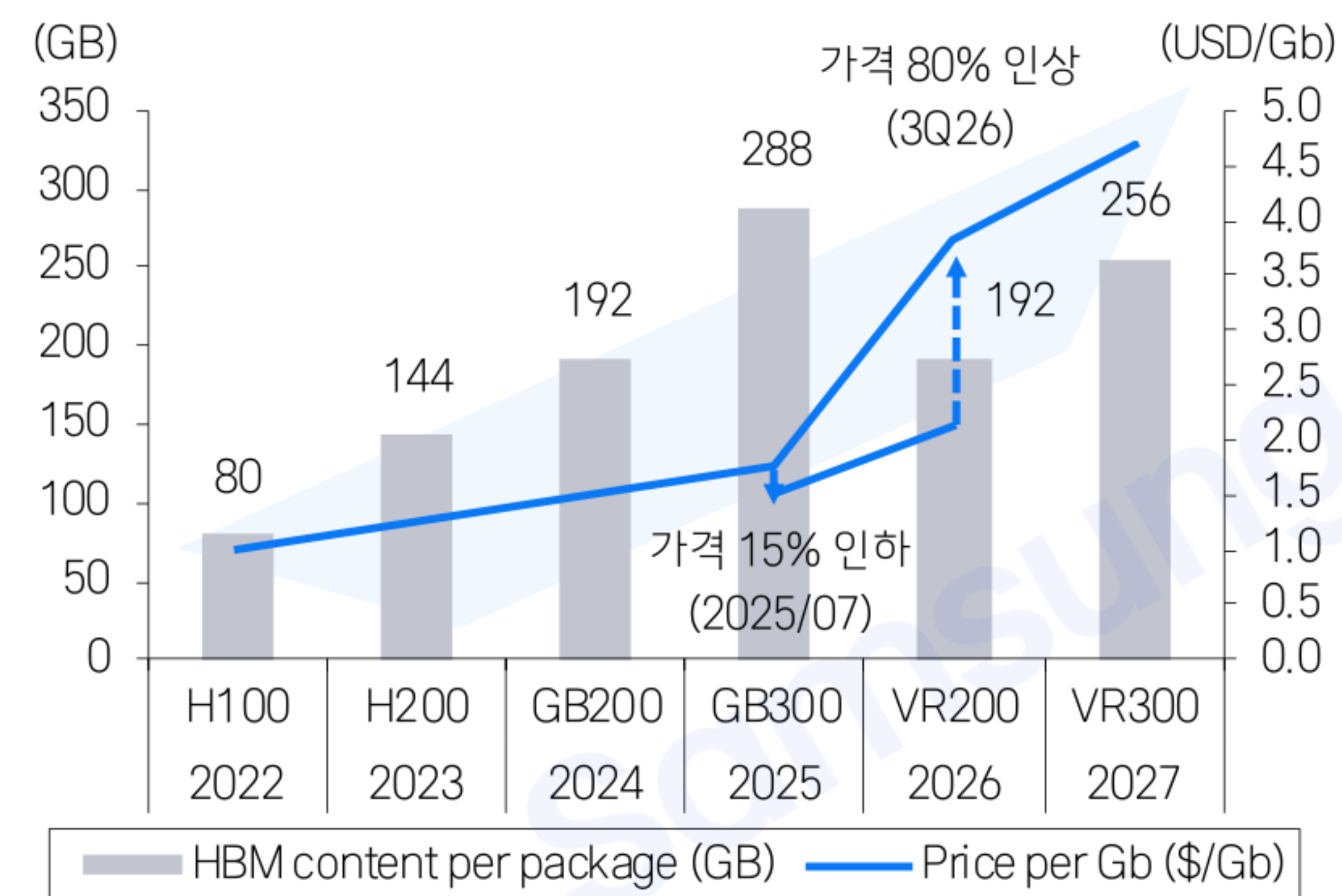
자료: 삼성증권 추정

NVL72 BOM cost 변화 추정

(\$k)	GB300	VR200 Old	VR200 New	Old VR Chg(%p)	New VR Chg(%p)
Compute	16.0%	14.1%	17.0%	-1.9%	1.0%
Memory	30.1%	40.5%	35.1%	10.4%	5.0%
Network	15.0%	14.3%	15.1%	-0.6%	0.2%
Power & Other	38.9%	31.1%	32.8%	-7.9%	-6.1%
Compute	274	482	551		
Memory	514	1,388	1,138		
Network	256	491	491		
Power & Other	665	1,064	1,064		
Total	1,709	3,426	3,245		

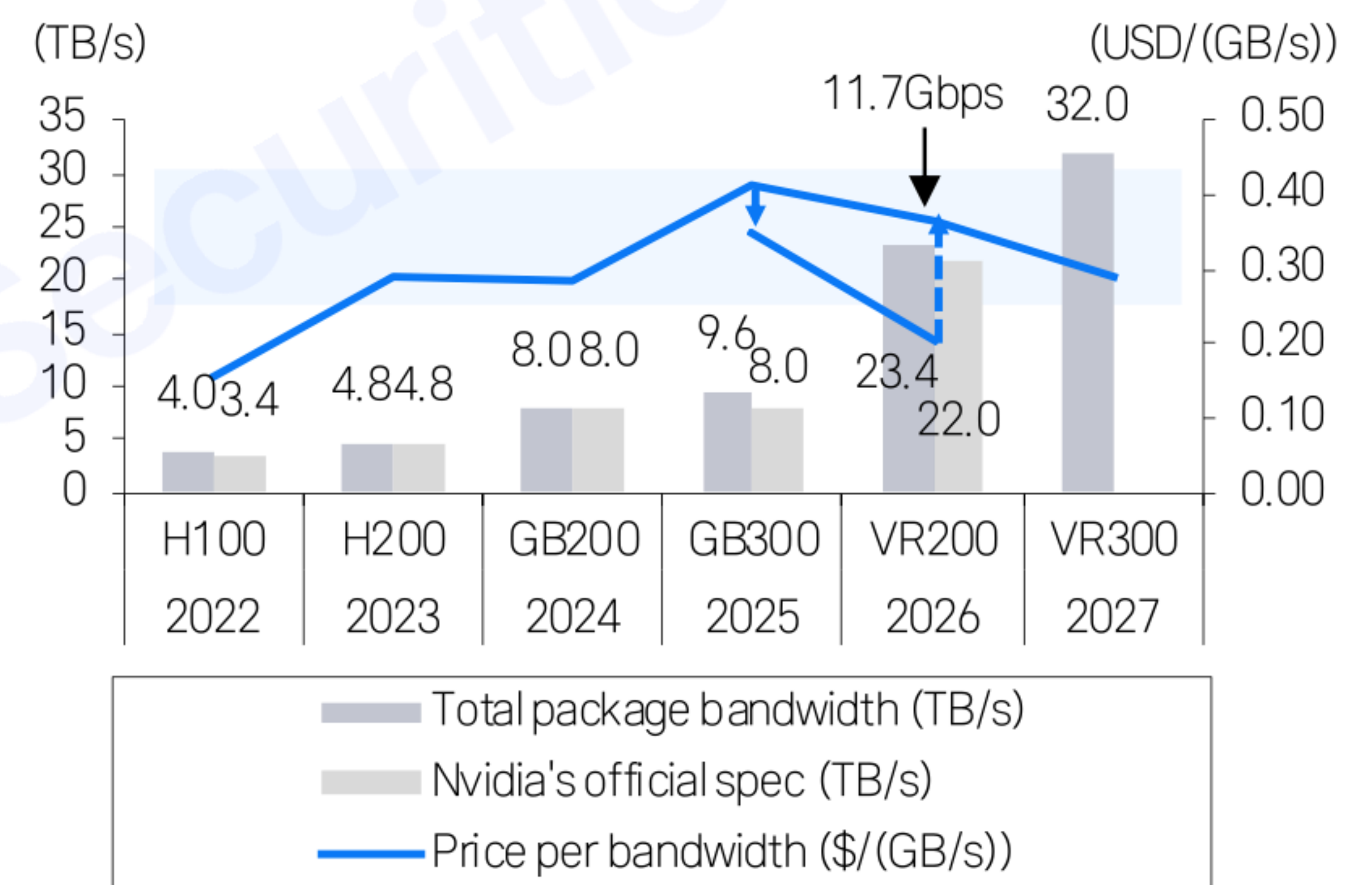
자료: 삼성증권 추정

HBM 용량당 가격은 세대를 거듭하며 지속 증가



자료: 삼성증권 추정

반면 HBM 대역폭당 가격은 비슷하게 유지

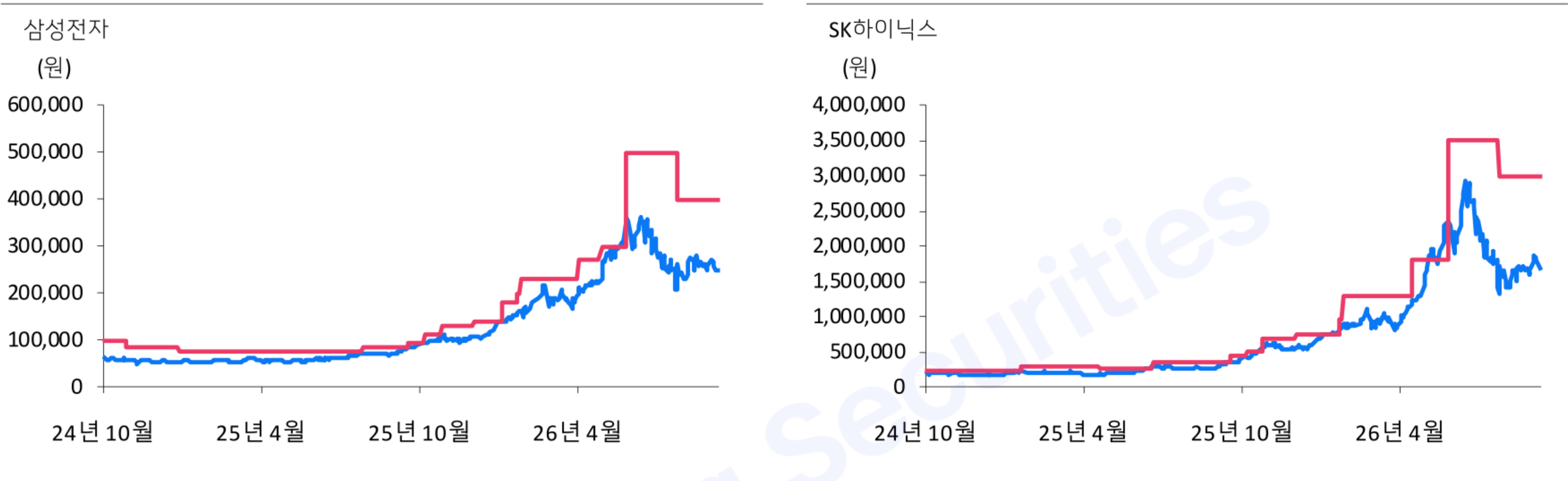


자료: 삼성증권 추정

Compliance notice

- 당사는 2026년 9월 16일 현재 삼성전자와(과) 계열사 관계에 있습니다.
- 본 조사분석자료의 애널리스트는 2026년 9월 16일 현재 위 조사분석자료에 언급된 종목의 지분을 보유하고 있지 않습니다.
- 당사는 2026년 9월 16일 현재 위 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
- 본 조사분석자료에는 외부의 부당한 압력이나 간섭 없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.
- 본 조사분석자료는 당사의 저작물로서 모든 저작권은 당사에게 있습니다.
- 본 조사분석자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다.
- 본 조사분석자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.
- 본 조사분석자료는 기관투자가 등 제3자에게 사전 제공된 사실이 없습니다.

2년간 목표주가 변경 추이



최근 2년간 투자의견 및 목표주가 변경 (수정주가 기준)

삼성전자												
일 자	2024/9/12	11/1	2025/1/2	8/1	9/23	10/13	10/31	12/8	2026/1/9	1/27	1/30	4/7
투자의견	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY
TP (원)	100000	83000	74000	85000	93000	110000	130000	140000	180000	200000	230000	270000
과리율 (평균)	-39.49	-33.92	-21.93	-15.05	-6.95	-10.64	-21.96	-16.54	-18.69	-19.23	-19.75	-20.25
과리율 (최대or최소)	-35.30	-29.28	-1.89	-1.76	1.51	-5.36	-14.54	0.71	-15.39	-18.80	-5.22	-16.30
일 자	5/4	6/2	7/31									
투자의견	BUY	BUY	BUY									
TP (원)	300000	500000	400000									
과리율 (평균)	-3.39	-40.44										
과리율 (최대or최소)	16.33	-27.50										
SK하이닉스												
일 자	2024/9/12	2025/1/24	4/24	6/25	9/23	10/13	10/30	12/8	2026/1/27	1/30	4/21	6/2
투자의견	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY	BUY
TP (원)	240000	280000	250000	340000	430000	500000	700000	750000	950000	1300000	1800000	3500000
과리율 (평균)	-23.78	-30.26	-14.53	-17.52	-14.81	-3.33	-19.05	-11.48	-10.42	-26.68	-3.42	-36.95
과리율 (최대or최소)	-6.04	-21.96	11.40	3.82	-0.47	11.60	-11.43	2.27	-9.37	-10.31	31.28	-16.60
일 자	7/30											
투자의견	BUY											
TP (원)	3000000											
과리율 (평균)												
과리율 (최대or최소)												

투자기간 및 투자등급: 삼성증권은 기업 및 산업에 대한 투자등급을 아래와 같이 구분합니다.

* 2023년 7월 27일부로 기업 투자 등급 기준 변경

기업		산업	
BUY (매수)	향후 12개월간 예상 절대수익률 15% 이상 그리고 업종 내 상대매력도가 평균 대비 높은 수준	OVERWEIGHT(비중확대)	향후 12개월간 업종지수상승률이 시장수익률 대비 5% 이상 상승 예상
HOLD (중립)	향후 12개월간 예상 절대수익률 -15%~ 15% 내외	NEUTRAL(중립)	향후 12개월간 업종지수상승률이 시장수익률과 유사한 수준 (±5%) 예상
SELL (매도)	향후 12개월간 예상 절대수익률 -15% 이하	UNDERWEIGHT(비중축소)	향후 12개월간 업종지수상승률이 시장수익률 대비 5% 이상 하락 예상

최근 1년간 조사분석자료의 투자등급 비율 2026.06.30 기준
매수(85.5%)·중립(14.5%)·매도(0%)

Samsung Securities

삼성증권

삼성증권주식회사

서울특별시 서초구 서초대로74길 11(삼성전자빌딩)
Tel: 02 2020 8000 / www.samsungpop.com

삼성증권 Family Center: 1588 2323

고객 불편사항 접수: 080 911 0900



Member of
**Dow Jones
Sustainability Indices**
Powered by the S&P Global CSA