Peak Insights

Peak Insights represent the thought leadership of Portfolio Manager Chris Smith and the Antero Peak Group. Through Peak Insights, the team seeks to offer unique access to innovative leaders and a view into important trends that will impact markets and businesses.

HBM Opportunity Within Memory

After two years of cyclical declines, the global compute and memory market is at an inflection point, creating a unique opportunity for our thematic investing framework. We believe we are at one of these points of structural change within the DRAM (dynamic random-access memory) market driven by the emergence of High Bandwidth Memory (HBM) and its applicability to AI-related compute. AI and HBM in particular are going to alter the memory landscape for many years, potentially driving higher growth and profitability for the companies involved , creating a secular structural growth opportunity within the cyclical memory market.

Memory Background

DRAM is the primary memory in modern computers and graphic cards. DRAM can be thought of as the short-term memory that allows a processor to quickly access and process information. As compute speed grows thanks to faster logic chips, input/output speeds from memory have become a critical constraint for overall performance. This dynamic has led to the creation of new DRAM microarchitectures. HBM, a 3D version of stacked DRAM, is very fast and energy efficient and will increasingly be paired alongside AI graphics processing unit (GPU) chips.

DRAM has historically been a hyper-cyclical market marked by booms and busts where the latter periods often resulted in negative gross margins and large operating losses for industry participants. Over the last several decades, a large number of US and Japanese players entered this fiercely competitive market, but failed to survive as returns were driven down by new market entrants. Today, the industry is consolidated to three large players: Korea-based Samsung Electronics and SK Hynix and US-based Micron Technology. Greater consolidation has led to better supply discipline during economic downturns and made the industry more attractive from a long-term investing standpoint.

DRAM, as an industry, has seen annual binary digit (bit) growth (a measure of the quantity of information produced) in the mid-teens with total through-cycle revenue growth in the single digits. Price, while cyclical, has tended to decline over time. Each successive generation of product and process node advancement leads to faster data transfer rates, higher memory densities and improved energy efficiency. Technological advancements in DRAM production have been highly deflationary as each successive generation lowered the cost per bit.

This marks a shift in this deflationary paradigm as the cost to produce HBM is substantially higher due to its larger die size, 3D stacking and additional expensive components.

Investment Risks: Investments will rise and fall with market fluctuations and investor capital is at risk. Investors investing in strategies denominated in non-local currency should be aware of the risk of currency exchange fluctuations that may cause a loss of principal. These risks, among others, are further described near the back of this document, which should be read in conjunction with this material.

Christopher Smith
Portfolio Manager

21 Years Investment Experience

AI and HBM's Role

While the growth of AI applications and AI infrastructure is unlikely to be linear, it is still in very early stages with a long runway for growth. Lisa Su, the CEO of Advanced Micro Devices, made headlines in December 2023 when she forecasted an addressable market for datacenter AI chips of $400B in 2027 from $45B in 2023, suggesting a 70% compound annual growth rate (CAGR). While this number is eye-popping, the mathematical rationale and basis for this estimate is not unreasonable. Estimates can justify 10 million AI servers in place by 2027 using what we believe are reasonable assumptions. Exhibit 1 does a great job of walking through, piece by piece, how one might arrive at this figure.

Exhibit 1: Usage and Complexity: Reasonable Assumptions Can Justify 10 Million AI Servers By 2027

	2023 - 2027 INCREASE	RATIONALE
Model complexity	~100X	10X every 2 years. Deceleration vs today (10X every year).
Chip efficiency	~120X	6X every 18MM (every new generation). In line with today.
Chip time per unit of usage	**0.8X**	**Model sized divided by chip efficiency**
Users	~5X	200MM to 1B. Internet users grew from 200MM to 800MM from 1999 to 2003.
Usage per user	~5X	From 1min of full inference server usage per user per day in 2023 to 5min in 2027, as use cases are still limited. Imagine, for instance, GenAI getting into gaming or office productivity.
Usage	**25X**	**Users x usage per user**

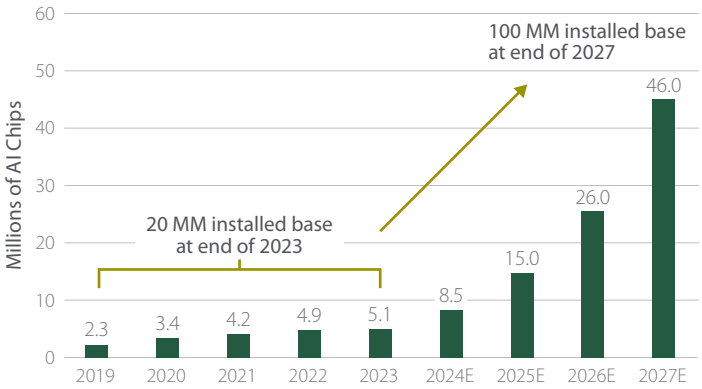
	INCREASE IN THE SERVER INSTALLED BASE			# OF SERVERS 2023		2027
Training	√(Usage Increase)	x	Chip time / unit of usage	1.7MM	4X	6.8MM
	√25 x 0.8 = 4					
	Assumes # of users of the same training is growing as much as the volume of training					
Inference	Usage increase	x	Chip time / unit of usage	0.18MM	20X	3.6MM
	25 x 0.8 = 20					
	Total AI servers			**1.9MM**	**5X**	**10MM**

A rapid but reasonable adoption and development of AI services can easily require 10MM AI servers in 2027

Source: New Street Research.

Assuming 8-10 chips per server, 10MM AI servers implies approximately 80MM to 100MM AI chips. In order to get to an installed base of 10MM AI servers in 2027, AI chip deployments would have to grow rapidly.

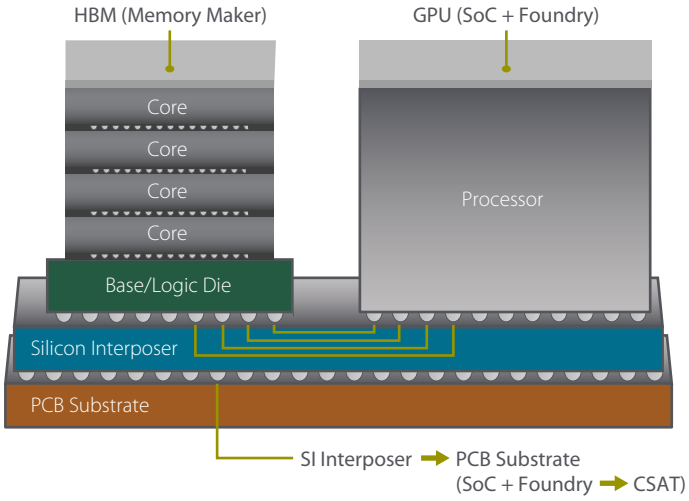
Exhibit 2: Global Datacenter AI Chip Installed Base



Source: New Street Research, published 6 Jan 2024.

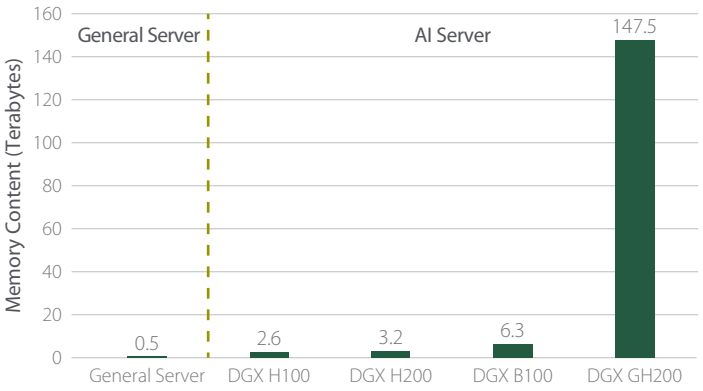
Given the inordinately large amount of data that needs to be processed by large clusters of AI accelerators in both training and inferencing, memory bandwidth has become a key constraint and enabler of performance. HBM's 3D vertical stacking of DRAM dies enable much wider memory interfaces, allowing for the "high bandwidth" in "high bandwidth memory." In addition, HBM requires less power than alternative technologies. All of this makes HBM particularly well suited and necessary for the training and inferencing associated with AI Large Language Models.

Exhibit 3: How HBM Attaches To A GPU



Source: Citi Research.

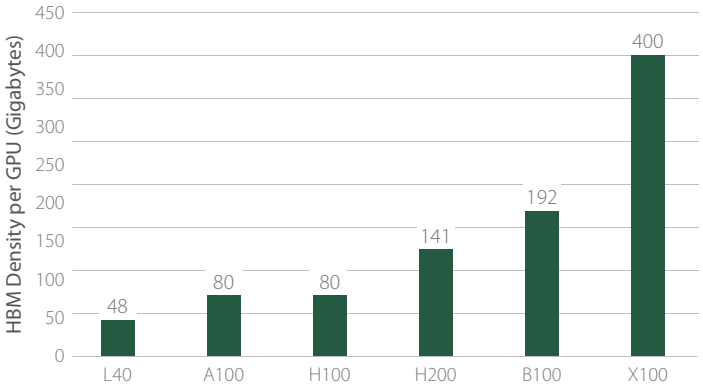
Exhibit 4: DRAM Content Per System: General Server Versus AI Servers



Source: Citi Research/NVIDIA, published 1 Jan 2024.

Beyond the rapid growth of AI accelerators themselves, memory content of AI accelerators is 5X-294X larger than that of a traditional server (Exhibit 4). This is key as memory consumption per accelerator is set to grow enormously to yield better accelerator performance. As an example, NVIDIA's X100 GPU is expected to be released sometime in 2025 and will have 5X the HBM content as its current H100 GPU. The H100 is currently the most coveted flagship GPU and has been hailed as the world's most advanced chip, but it is priced at a massive $30,000 USD.

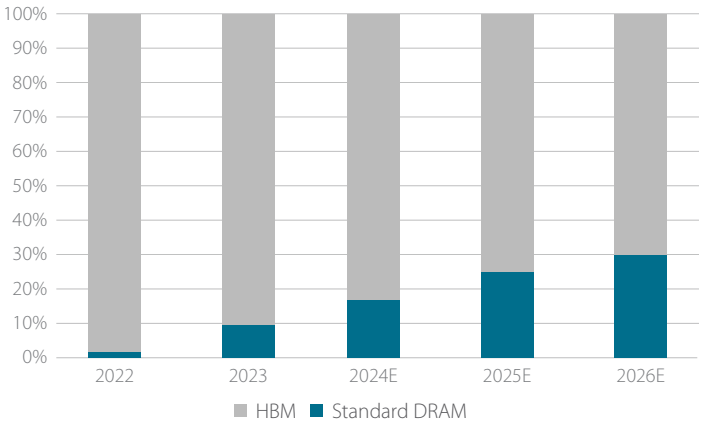
Exhibit 5: HBM Density Per NVIDIA AI GPU



Source: Citi Research/NVIDIA.

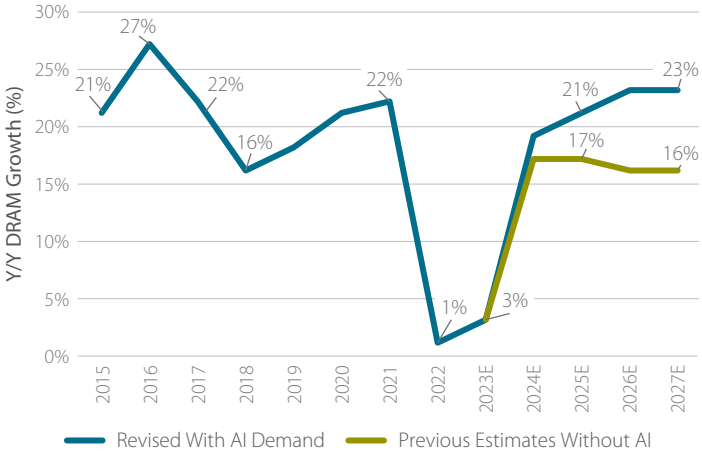
All of these factors could drive a 7X increase in HBM between 2023 and 2025 and an increase in HBM as a percent of the total DRAM market from the single digits to ~25% (Exhibit 6), driving an acceleration in the overall DRAM bit growth. Exhibit 7 shows the projected growth inclusive of AI demand versus previous numbers.

Exhibit 6: HBM Share Of DRAM Market



Source: Arete Research/Antero Peak Group. As of 31 Dec 2023. Estimates are based on the team's analysis and are subject to material revision.

Exhibit 7: Global DRAM Demand Growth—Revised For New AI Demand



Source: Citi Research, published 1 Jan 2024.

Memory has historically been a commodity that was *standardized* to fit into Intel CPUs, but in the future, it will be increasingly optimized and *customized* to meet the requirements of different AI chips. The increased product diversification and complexity of memory products driven by AI will lead to memory looking more like the logic foundry market—an attractive one to us at the Antero Peak Group. Customization, in our view, is synonymous with more differentiation, higher price and higher margin. This should benefit the HBM companies who have historically been faced with commodity pricing curves for their legacy memory products.

Where Are We In The Cycle?

We are at the beginning of an upturn in the memory cycle after a prolonged and severe downturn which began in 2022. Demand for smartphones, PCs and general-purpose servers—the key end-markets for DRAM—has shifted to positive year-over-year growth in Q4 2023 after two years of declines. The COVID-driven demand surge in 2020 and 2021 led to overconsumption of goods, overproduction of components (including DRAM) and the build-up of excess inventory. The DRAM industry reduced production in 2023 by about 30% from highs to combat weak demand and excess inventory. The industry is still working down that excess inventory, but we believe there is now line of sight to normalization within the next couple of quarters. The supply cuts that we have seen combined with the bottoming out of end-markets and ongoing inventory normalization has led to pricing beginning to rise sequentially. In the last few cycles, trough to peak pricing increased between 77%-225% (Exhibit 8). In the current cycle, pricing has only increased 16% so far, suggesting we have quite a bit more room to go.

Exhibit 8: Trough to Peak DRAM Spot Price Change By Cycle

	2013 - 2014	2016 - 2018	2020 - 2021	2023 - CURRENT
Trough	$2.35	$1.54	$2.15	$1.19
Peak	$4.19	$5.00	$3.81	$1.38
Trough to Peak Increase	78%	225%	77%	16%

Source: Antero Peak Group. As of 31 Jan 2024.

Furthermore, unlike prior cycles, HBM presents an additional catalyst to the DRAM market today. While we have already discussed how HBM has attractive growth characteristics fueled by AI, HBM also acts as a functional supply cut for the DRAM market. Because the die size is 2X as big as double data rate (DDR) and yields are lower, for every bit of capacity allocated to HBM, the overall DRAM market loses 2 bits. Thus, if 7% of bit capacity were to move to HBM by 2025 (a reasonable estimate), overall DRAM capacity inclusive of HBM would decline by 14%, accelerating price growth. We believe existing supply discipline, the effective capacity reduction from the shift to HBM and a demand recovery is likely to drive a meaningful amount of additional pricing growth from here.

Summary

DRAM sits at an attractive inflection point—the beginning of an AI-driven memory upcycle with cyclical and secular tailwinds. We expect memory products to become semi-customized in the future due to the emergence of AI. This should result in increased product diversification and complexity. As products become more customized to meet specific needs, the ability of semiconductor companies to differentiate their products and charge a higher price will become apparent. Customers will be less able to switch between competing memory providers, leading to higher margins, stickier relationships and higher returns on capital over time. We see acceleration in the growth of leading-edge DRAM and are focusing our research on the companies in the space with exposure to HBM memory, including Samsung, SK Hynix, Micron and ASML—the only company in the world that makes the equipment necessary for the chip companies to make leading edge memory. In short, our strong belief in the coming importance of AI implies strong growth in HBM and an upcoming paradigm shift in the DRAM market.

For more information: Visit www.artisanpartners.com

Investment Risks: Non-diversified portfolios may invest larger portions of assets in securities of a smaller number of issuers and performance of a single issuer may have a greater impact to the portfolio's returns. Use of derivatives may create investment leverage and increase the likelihood of volatility and risk of loss in excess of the amount invested. High portfolio turnover may adversely affect returns due to increased transaction costs and creation of additional tax consequences. Securities of small- and medium-sized companies tend to have a shorter history of operations, be more volatile and less liquid and may have underperformed securities of large companies during some periods. International investments involve special risks, including currency fluctuation, lower liquidity, different accounting methods and economic and political systems, and higher transaction costs. These risks typically are greater in emerging and less developed markets, including frontier markets. Investors investing in strategies denominated in non-local currency should be aware of the risk of currency exchange fluctuations that may cause a loss of principal. These risks, among others, are further described in Artisan Partners Form ADV, which is available upon request. This is a marketing communication.

This summary represents the views of the portfolio managers as of 31 Jan 2024. Those views may change, and Artisan disclaims any obligation to advise investors of such changes. For the purpose of determining the portfolio's holdings, exposures are delta-adjusted at the issuer level and may include multiple securities of the same issuer. The holdings mentioned above comprised the following percentages of a representative account within the Antero Peak Strategy Composite's total net assets as of 31 Dec 2024: NVIDIA Corp 10.1%. Securities named in the commentary, but not listed here are not held in the portfolio as of the date of this report. Portfolio holdings are subject to change without notice and are not intended as recommendations of individual securities.

Theme categorizations are at the sole discretion of the team. Themes and constituents are as of the date indicated and subject to change.

CAGR represents compound annual growth rate, the year-over-year growth rate over a specified period of time. It is calculated by taking the nth root of the total percentage growth rate, where n is the number of years in the period being considered. **Binary Digit (Bit)** is the smallest unit of data that a computer processes. **Die** is a thin piece of silicon that contains an integrated circuit which has been cut out of the wafer.

This material is provided for informational purposes without regard to your particular investment needs and shall not be construed as investment or tax advice on which you may rely for your investment decisions. Investors should consult their financial and tax adviser before making investments in order to determine the appropriateness of any investment product discussed herein.

Artisan Partners Limited Partnership (APLP) is an investment adviser registered with the U.S. Securities and Exchange Commission (SEC). Artisan Partners UK LLP (APUK) is authorized and regulated by the Financial Conduct Authority and is a registered investment adviser with the SEC. APEL Financial Distribution Services Limited (AP Europe) is regulated by the Central Bank of Ireland. APLP, APUK and AP Europe are collectively, with their parent company and affiliates, referred to as Artisan Partners herein. Artisan Partners is not registered, authorized or eligible for an exemption from registration in all jurisdictions. Therefore, services described herein may not be available in certain jurisdictions. This material does not constitute an offer or solicitation where such actions are not authorized or lawful, and in some cases may only be provided at the initiative of the prospect. Further limitations on the availability of products or services described herein may be imposed.

This material is only intended for investors which meet qualifications as institutional investors as defined in the applicable jurisdiction where this material is received, which includes only Professional Clients or Eligible Counterparties as defined by the Markets in Financial Instruments Directive (MiFID) where this material is issued by APUK or AP Europe. This material is not for use by retail investors and may not be reproduced or distributed without Artisan Partners' permission.

In the United Kingdom, issued by Artisan Partners UK LLP, 25 St. James's St., Floor 10, London SW1A 1HA, registered in England and Wales (LLP No. OC351201). Registered office: Phoenix House, Floor 4, Station Hill, Reading Berkshire RG1 1NB. In Ireland, issued by Artisan Partners Europe, Fitzwilliam Hall, Fitzwilliam Pl, Ste. 202, Dublin 2, D02 T292. Registered office: 70 Sir John Rogerson's Quay, Dublin 2, D02 R296 (Company No. 637966).

Australia: This material is directed at wholesale clients only and is not intended for, or to be relied upon by, private individuals or retail investors. Artisan Partners Australia Pty Ltd is a representative of APLP (ARBN 153 777 292) and APUK (ARBN 603 522 649). APLP and APUK are respectively regulated under US and UK laws which differ from Australian laws and are exempt from the requirement to hold an Australian financial services license under the Australian Corporations Act 2001 in respect to financial services provided in Australia.

Canada: This material is distributed in Canada by APLP and/or Artisan Partners Distributors LLC, which conduct activities in Canada under exemptions from the dealer, portfolio manager and investment fund manager registration requirements of applicable Canadian securities laws. This material does not constitute an offer of services in circumstances where such exemptions are not available. APLP advisory services are available only to investors that qualify as "permitted clients" under applicable Canadian securities laws.

© 2025 Artisan Partners. All rights reserved.

For Institutional Investors Only—Not for Onward Distribution

ARTISAN PARTNERS

1/15/2025 – A25143L-vXUS