October 13, 2026 05:51 AM GMT

Semiconductors | North America

Highlighting 7 Interesting Semiconductor Privates

The growth in AI has reshaped the semiconductor landscape, creating new markets and opportunities where private companies are increasingly relevant. This note highlights 7 companies innovating around themes like: AI factory efficiency, disaggregated inference, physical AI, and Security/sovereign AI.

PrivateCompany

Private Companies Highlighted: Axiado, Ayar Labs, d-Matrix, Eliyan, SambaNova, Sima.ai, and Tenstorrent. Ayar Labs, SiMa.ai, and d-Matrix will be attending the

Joseph Moore
Equity Analyst
Joseph.Moore@morganstanley.com +1 212 761-7516

Mason Wayne
Research Associate
Mason.Wayne@morganstanley.com +1 212 761-6012

Ella Tulchinsky
Research Associate
Ella.Tulchinsky@morganstanley.com +1 212 761-2222

Shane Brett
Equity Analyst
Shane.Brett@morganstanley.com +1 212 761-1022

Nicole Kozhukhov
Research Associate
Nicole.Kozhukhov@morganstanley.com +1 212 761-1636

Cate Folan
Research Associate
Cate.M.Folan@morganstanley.com +1 212 296-3520

We are seeing the rate of change in semiconductor technology accelerate rapidly due to the proliferation of AI. After a long period of the semiconductor TAM being dominated by consumer devices like PC's and smartphones the emergence of the datacenter and more recently AI end markets has opened up increasing opportunities for new and smaller companies to quickly gain global scale and relevance. We think incumbents are well positioned to address changing needs in the market and the shift from training to inference to agenetic workloads - that said as the size of the market grows there's room for more winners.

Below we highlight 4 themes we are watching in the private company landscape, and 7 companies with various exposures to those drivers.

Theme #1: Improving AI factory efficiency: Getting the most tokens, and therefore the most revenue, out of a limited power budget is a growing focus for neoclouds and hyperscalers. **Axiado** is gaining traction with its Trusted Control/Compute Unit (TCU), a single chip that handles server management and hardware security and uses AI to tune power and cooling, so more of each rack's power goes to compute. **Ayar Labs** is pioneering scale-up CPO solutions that cut power consumption and increase bandwidth for AI factories. **Eliyan** is addressing the chiplet interconnect and copper bottlenecks: its die-to-die IP enables high-bandwidth, low-power chiplet links on standard packaging, while NuLight optical interconnects and dual-port custom HBM aim to extend bandwidth between and into XPUs. In XPUs themselves, upstarts **d-Matrix, SambaNova** and **Tenstorrent** are pushing the frontier in terms of performance per watt through architectural innovations around memory and scale-up/scale-out topologies.

Theme #2: Disaggregated and fast Inference: An emerging trend which involves breaking up certain portions of the AI inference workload and running them on different processors depending on that workloads requirements. Disaggregation is a preferred way for the AI ecosystem to work around the increasing costs of memory, especially HBM. The most common form of disaggregation is to run the pre-fill (compute bound) portion of the workload on an Nvidia GPU, and the de-code

investors should be aware that the firm may have a conflict of

Research as only a single factor in making their investment decision.

For analyst certification and other important disclosures, refer to the Disclosure Section, located at the end of this report.

(memory bound) on an SRAM based accelerator like Cerebras or Groq. We have seen the TAM for disaggregation has accelerated sharply around the end of 2025 with the rise of agenetic coding and the largest players in the semis space are participating: NVIDIA acquired Groq around this time, AWS subsequently announced a disaggregated architecture pairing Trainiumalphawave for prefill with Cerebras for decode over EFA, and Cerebras has announced a similar approach with AMD.

In this note we outline the approaches to this market from: (1) **d-Matrix**, which recently joined Nvidia's NVLink Fusion ecosystem (link) and is targeting speculative decode as its entry point to disaggregation. In this model the GPU runs the large model while d-Matrix accelerates the smaller models that predict outputs in advance. **d-Matrix** next-generation product named Raptor stacks DRAM directly on top of compute logic, which the company says delivers SRAM-class bandwidth at roughly one-tenth of HBM's energy per bit with higher yields (2) **SambaNova**, whose RDUs (Reconfigurable Dataflow Units) and three-tier memory structure (on-chip SRAM, HBM and directly attached DDR) keep very large models closer to the chip and let customers switch between them quickly. The architecture is being used for on-prem fast inference in air-cooled racks, and in disaggregated inference environments alongside GPUs. (3) **Tenstorrent**, which is less focused on disaggregation, though it sees opportunities in that market, and is instead focused on building the best general-purpose system for AI performance. Its Tensix core pairs a matrix engine, a vector engine and local SRAM with small RISC-V cores that handle data movement, and hundreds of these tiles connect over an on-chip mesh network.

Theme #3: Embodied AI: Embodied AI is thought to be the next frontier in AI, spanning robotics, drones and smart infrastructure. While it is a much smaller market than the data center today, the TAM for manufacturing, autonomous systems and other applications that can be augmented by physical AI technologies creates a market opportunity of trillions of dollars over the longer term, in our view. **SiMa.ai** designs low-power inference silicon for physical AI. Its Modalix MLSoC combines a machine learning accelerator, a computer vision processor and Arm cores on a single chip, and runs physical AI models at under 10W. Those SoCs are enabled by its Palette Neat software, **SiMa.ai** is initially focused on robotics, drones and automotive. **Tenstorrent**'s RISC-V IP also addresses this market, as the company targets edge and automotive applications alongside the data center. It licenses its Ascalon CPU cores and Tensix AI IP so that customers can build custom silicon for robotics and autonomous systems.

Theme #4: Security, and sovereign AI: With heightened geopolitical tension around AI technologies and their growing importance in military technologies, we are seeing increased focus on two related areas: (1) security, including cybersecurity and IP protection for LLM builders, and (2) sovereign AI, where governments around the world want access to and control over their own AI infrastructure rather than relying on clouds in other regions. **Axiado**'s TCU (Trusted Control/Compute Unit) combines the baseboard management controller (BMC), hardware root of trust and TPM (Trusted Platform Module) functions with an on-chip AI engine in a single SoC. The product is vendor neutral, which makes it a possible building block for secure infrastructure across XPU suppliers. **Tenstorrent** is supplying RISC-V IP in a more

open model than vendors like Arm, allowing greater customization and control over the silicon with fewer licensing restrictions, giving customers more independence in designing their own chips.

Exhibit 1: We highlight 7 private companies targeting the AI TAM with a variety of hardware/software

Company	Market Opportunity / What They're Building	Notable Potential Competitors
Axiado	Trusted Control Unit (TCU) AI-driven hardware platform for servers & data centers — combines BMC, root-of-trust and TPM functions on one SoC.	ASPEED, Nuvoton, Lattice, Marvell, Microchip
Ayar Labs	Optical I/O chiplets (TeraPHY) + SuperNova light source Co-packaged optics replacing copper links to scale AI compute bandwidth and reduce power.	Broadcom, Marvell, Nvidia, Lumentum, Coherent, Celestial AI (Marvell), Lightmatter (private)
d-Matrix	Inference accelerators Digital in-memory compute purpose-built for low-latency, high-throughput generative AI inference in the data center.	Groq (Nvidia), Cerebras, SambaNova (private), Etched (private), AMD
Eliyan	Interconnect IP that moves more data into, and between, AI chips IP portfolio spans across optical, electrical, and memory to enable more efficent AI infrastructure	Synopsys, Cadence, Alphawave Semi (Qualcomm), Broadcom, Marvell, ams OSRAM
SambaNova	Inference accelerators + SambaNova Cloud Full-stack AI platform focused on fast, efficient inference for large models.	Nvidia, AMD, Groq, Cerebras, Etched (private), d-Matrix (private)
Sima.ai	Modalix MLSoC + Palette software Low-power platform for physical AI in robotics, automotive, drones and smart vision.	Nvidia, Qualcomm, Hailo (private), Ambarella
Tenstorrent	RISC-V CPUs + AI accelerators Open-source software stack, direct chip sales, and CPU/AI IP licensed to partners.	Nvidia, AMD, SiFive (private), Arm, Andes

For Further Reading:

Semiconductors: How can the AI ecosystem work around memory bottlenecks? (5 Oct 2026)

Cybersecurity: Agents on the Loose: Cyber Faces an Identity Crisis (20 Aug 2026)

US Public Policy & Global Thematics: AI Sovereignty: Policy, Market, and Economic Implications (19 Aug 2026)

Technology: Bringing More Light to Scale-Up Networks: An Updated Scale-Up Primer (13 Jul 2026)

Cerebras Systems: Unique capability in fast inference drives large economic opportunity; Initiate at OW (8 Jun 2026)

Humanoids: A $5 Trillion Global Market (29 Apr 2025)

The content addressing private companies is being provided for informational purposes only and does not constitute a solicitation or imply future research coverage if the company goes public. Content is based on unaudited information. No investment recommendation is provided as there is limited public information available for private companies. Investors should conduct their own due diligence and be aware that additional or different information may be provided by the company

Axiado - AI Chips for Hardware Security

Who is Axiado: Founded in 2017, Axiado designs and engineers AI-driven hardware security processors, secure management controllers, and platform-security silicon that protect servers, 5G, and data center infrastructure from vulnerabilities embedded in existing processors. The core of the product portfolio is the AX3080 TCU (Trusted Control/Compute Unit), which anchors platform security directly in hardware to secure the control and management of heterogeneous infrastructure.

Company Mission: Axiado's mission is to secure digital infrastructure by embedding a new breed of AI-driven hardware based platform security in servers, 5G, network and power management controller infrastructure. Axiado believes data security should be the default solution, not an option.

Market opportunity

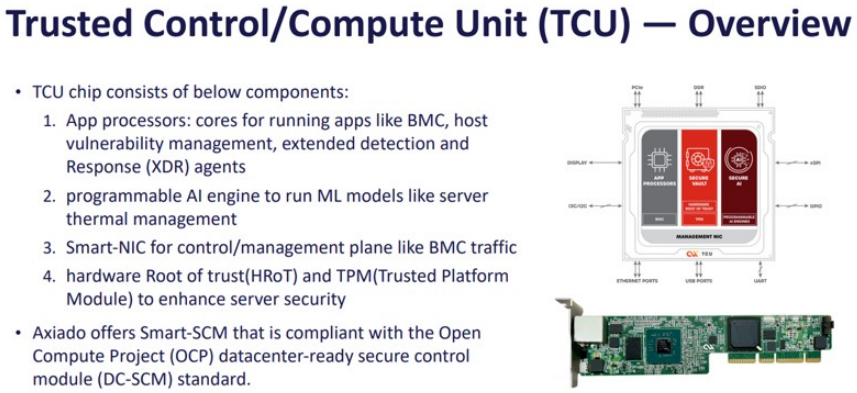
Axiado is focused on AI rack and platform security management. Cybersceurity is an increasing priority across enterprises globally, and with new threats from AI Axiado expects enterprise focus on this area to continue to increase. Axiado is building a new AI first security platform for the datacenter and see's an opportunity to replace fragmented TPM, BMC, LOM, firmware / software, and FPGA-based RoT approaches to a more purpose built and seamless product. The central idea is that rack and pod level infrastructure redefines what a management controller must see, secure, and coordinate - creating a shift from board / tray management toward centralized rack and pod management. It's also worth noting that adding accelerated compute capabilities can improve the overall performance of the system, with thermal management a compelling use case (see here).

Product overview: Axiado's TCU combines multiple board/rack management solutions into one product. with a number of form factors outlined below. The company is vendor neutral, and sees an opportunity across XPUs as well as CPU servers.

Exhibit 2: Axiado product overview



Exhibit 3: Axiado TCU overview



Source: Company Data [https://axiado.com/wp-content/uploads/2024/04/OCP24R-Increasing-Energy-Efficiency-using-DRL-BO.pdf], Morgan Stanley Research

Competitive differentiation

Axiado's differentiation is combining hardware, software, AI, and security in a vendor-neutral TCU platform. The company says it can attach across rack architectures and provide both root-of-trust and rack-level management.

Management has argued that rack vendors lack security expertise and that hyperscalers currently pay multiple vendors / people to handle functions Axiado can automate. The company also believes programmability is inappropriate for security, an advantage against FPGA-based solutions.

Notable Potential Competitors: Lattice Semiconductor/AMI, Aspeed, Nuvoton, Marvell, Microchip

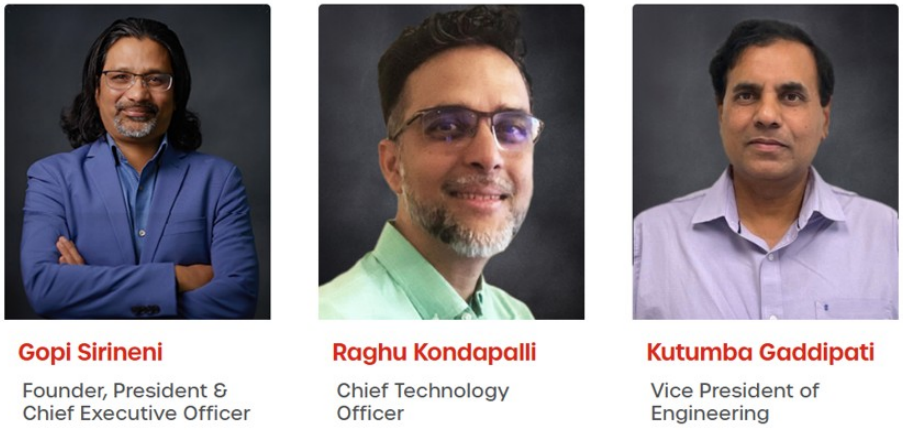
Recent developments

- **In November 2025 Bill Brennan, President and CEO of Credo Technology Group, joined Axiado's Board of Directors,** bringing over 25 years of leadership experience in the semiconductor and connectivity industries, including having guided Credo from a startup to a publicly traded company. His appointment reflects Axiado's push to build out its leadership bench with Tier-1 qualification and hyperscaler GTM experience.
- **In October 2025 at the OCP Global Summit, Axiado announced a collaboration with Inventec to integrate its TCU into Inventec's NVIDIA Blackwell GB200-based servers and Intel-based MGX platforms,** embedding hardware-anchored security and autonomous AI intelligence directly at the silicon level. The partnership pairs Axiado's SCM3002 card (built on the AX3000 TCU) for Intel/AMD platforms with the SCM3080-MT card (built on the AX3080 TCU) for NVIDIA platforms, unifying root-of-trust, BMC, TPM, secure networking, and AI inference into a single SoC. Embedded AI agents proactively detect anomalies, block attacks, and optimize system performance in real time. Capabilities the companies say software-only solutions cannot match, allowing threats to be stopped before the OS even boots.

Team

- **Gopi Sirineni , Founder, President & CEO,** founded Axiado in 2017 after serving as Vice President of Wi-Fi SON at Qualcomm (2012–2020) and as an Executive Board Advisor at Cognitive Systems Corp.
- **Raghu Kondapalli, Chief Technology Officer,** has served as Axiado's CTO since 2022, prior to which he was CTO of Intel's Data Center Group (2012–2020).
- **Kutumba Gaddipati,Vice President of Engineering,** joined Axiado in 2025, bringing prior experience as Director of SoC Engineering at SiMa.ai (2022–2025) and Director of SoC Architecture and Design at Akeana (2021–2022).

Exhibit 4: Axiado: Key Leadership



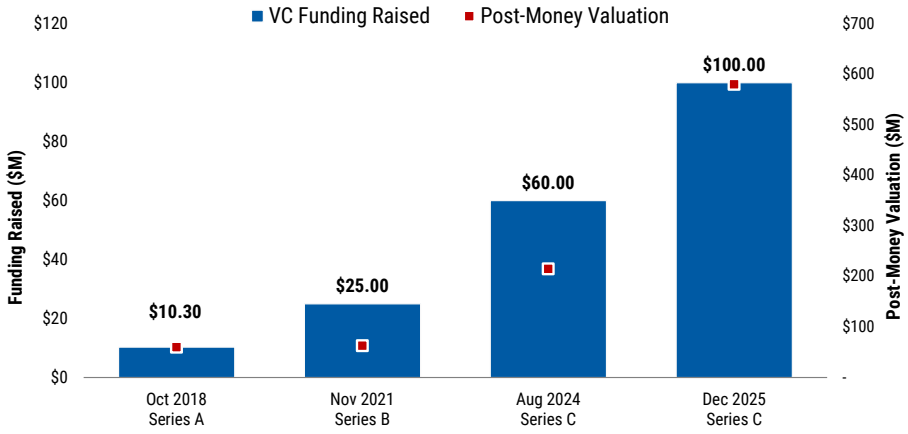
Source: Axiado

Investors and Funding

Axiado has raised a cumulative $195.3mn since 2018, the most recent round in December

2025 raised $100.0mn at a post money valuation of $0.58bn. Investors in the most recent round include Maverick Silicon (lead), Prosperity7 Ventures, Orbit Venture Partners, Crosslink Capital, and Nosterra Ventures.

Exhibit 5: Axiado Historical Funding Rounds (Sourced from PitchBook)



The valuation information provided here is included for illustrative purposes only. The information is based on publicly available information (funding rounds, company

Research.

Ayar Labs - Optical Interconnects for AI Data Centers

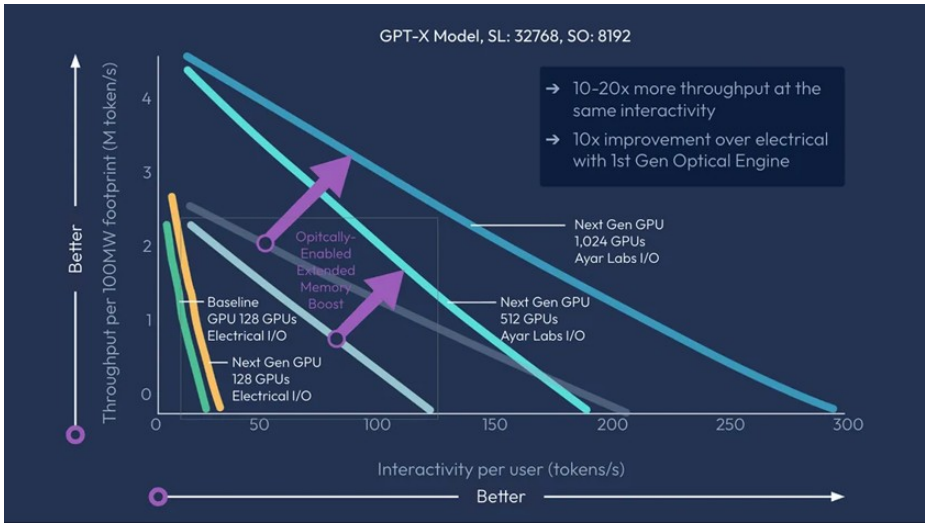
Who is Ayar Labs: Founded in 2015, designs and engineers optical interconnect chiplets and lasers that move data between chips using light instead of electrical signals, accelerating data movement for next-generation AI infrastructure. Ayar Labs Photonic Fabric delivers very high bandwidth at nanosecond latencies to break through the electrical I/O bottlenecks limiting AI data center compute scaling.

Company Mission: Ayar Labs' mission is to solve the foundational connectivity bottleneck in computing systems using a new generation of electronic/photonic integration and architectures.

Market opportunity

Ayar Labs is addressing the bandwidth and distance bottleneck in AI infrastructure through optical I/O and co-packaged optics. Management sees electrical I/O becoming the bottleneck, forcing more compute into racks and driving up power per rack. Removing the bandwidth-distance limitation could allow larger scale-up domains without the same constraints. The company sees optical engines, optical I/O, and CPO as increasingly central to AI infrastructure, with potential deployment in XPU-to-XPU, rack-to-rack, and future direct-attach optical systems. Compared to standard copper and pluggable optics Ayar sees 10x better bandwidth, 10x lower latency, and 3-5x the power efficiency. Ayar's solution is designed to be protocol agnostic, meaning the opportunity is applicable beyond any particular XPU or interface.

Exhibit 6: CPO value proposition for AI factories



Source: Ayar Labs

Product Overview: The company provides PIC + EIC solutions, external light sources, and optical I/O chiplets. The core product is the optical engine (TeraPHY), which integrates a

photonic integrated circuit (PIC) with an electronic integrated circuit (EIC) to convert electrical signals into light and back again directly at the chip package, replacing traditional copper I/O with in-package optical connectivity. This optical engine is powered by an external, remote light source (SuperNova), which keeps the laser off the hot compute die entirely — improving reliability and simplifying thermal management. Together, these components form Ayar Labs' full-stack optical interconnect platform that lets GPUs, CPUs, and switches communicate at high bandwidth and low latency across a chip, a rack, or multiple racks. Ayar's solution can deliver roughly 10–20x better performance-per-watt than copper interconnects while removing the reach limitations that constrain today's largest AI clusters.

Exhibit 7: Ayar Labs, Wiwynn Collaboration at OFC

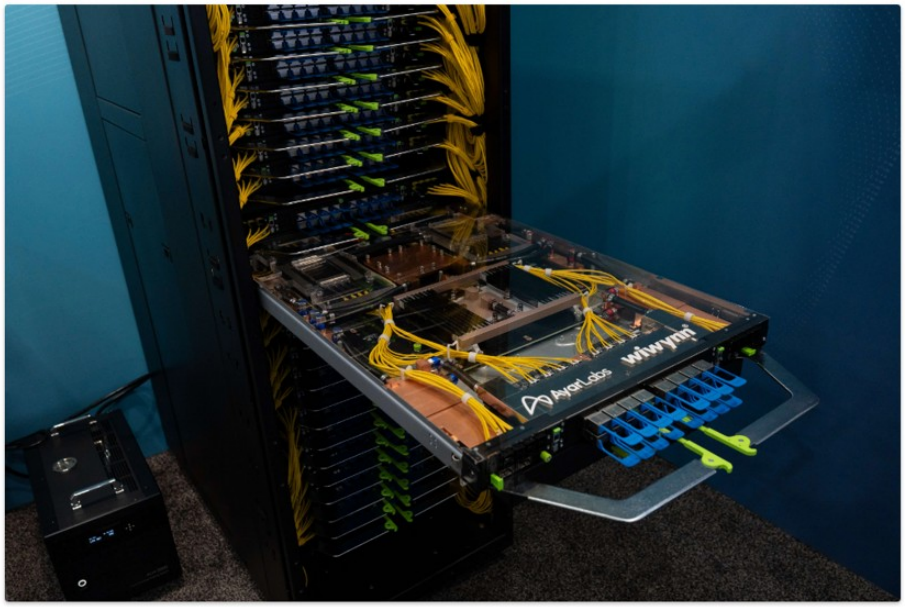


Figure 2. A L10 chassis designed and manufactured by Wiwynn that features Ayar Labs co-packaged optics and ELSFP light sources, as well as AI ASICs, advanced fiber management, and cold-plate liquid cooling.

Source: Ayar Labs [https://ayarlabs.com/blog/building-a-production-ready-optically-connected-rack-for-ai-scale-up/]

Competitive differentiation

Ayar labs brings optical engines directly onto advanced packages, integrating photonics with the XPU itself. This tight physical and integration creates a set of core competencies spanning photonics design, advanced packaging, and high-volume manufacturing that is difficult for a component-level optics supplier to replicate without also owning the packaging and system-level integration work. Ayar believes that the first generation of scale-up CPO will be dominated by TSMC's COUPE, and that mixing and matching EICs and PICs across different foundries or vendors will not be straightforward in the near term. Ayar Labs in one of just a few companies that have been public in building on TSMC's COUPE platform for several years.

On the technology roadmap itself, the company is skeptical that electro-absorption modulator (EAM)-based approaches can be productized in high volume within five years, and has instead bet on micro-ring-based optical systems paired with remote laser

architectures. Because the laser sits away from the compute die, this design sidesteps thermal constraints that would otherwise limit how memory and compute can be positioned relative to the optical engine, giving the company more architectural flexibility in how it lays out future systems, and reducing the risk that laser reliability issues affect the compute silicon.

Ayar labs is positioned as an alternative to switch/networking vendors who bundle optics into a broader portfolio, being able to bundle is an advantage on paper, but management believes that customers building next-generation GPU/XPU fabrics will prefer to avoid being locked into any single vendor's ecosystem. Management does expect large GPU/AI-accelerator vendors to build some internal CPO capability over time given how strategically important optical I/O is becoming.

NPO approaches should prove complementary rather than directly competitive, and sees them as more likely to absorb overflow demand during periods of tight co-packaged-optics supply than to displace the tightly integrated approach in mainstream scale-up deployments.

Notable Potential Competitors: Broadcom, Marvell, Nvidia, Lumentum, Coherent, Celestial AI (Marvell), Lightmatter (private)

Team

- **Mark Wade, Co-Founder and Chief Executive Officer**, leads Ayar Labs' strategy and growth as the company commercializes optical I/O chiplets for AI infrastructure; he previously served as Ayar Labs' CTO and SVP of Engineering, and co-invented the company's core photonics technology at MIT and UC Berkeley between 2010 and 2015, holding a PhD in Electrical Engineering from the University of Colorado Boulder.
- **Lisa Cummins Dulchinos, Chief Financial Officer,** oversees financial strategy and has led multiple successful funding rounds for the company, building the financial infrastructure to support Ayar Labs' scale-up as a VC backed hardware company.
- **Vladimir Stojanovic, Co-Founder and Chief Technology Officer**, led the team that designed the world's first processor to communicate using light prior to founding Ayar Labs, and previously was a professor of EECS at UC Berkeley and a key developer of Rambus's high-speed link technology, holding a PhD from Stanford University.

Exhibit 8: Ayar Labs: Ley Leadership



Mark Wade

Co-Founder and
Chief Executive Officer



Lisa Cummins Dulchinos

Chief Financial Officer



Vladimir Stojanovic

Co-Founder and
Chief Technology Officer

Source: Ayar Labs

Recent developments

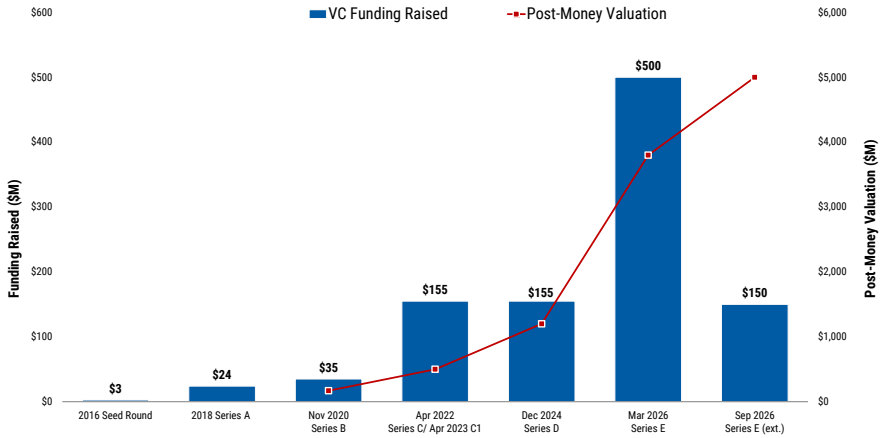
- **In June 2026 Ayar Labs joined the NVIDIA NVLink Fusion ecosystem**, making its co-packaged optics products optically and electrically compatible with NVIDIA's optical and SerDes technologies — enabling hyperscalers to deploy heterogeneous compute, including custom silicon, within NVIDIA rack-scale platforms while preserving NVLink-based architecture investments. The announcement built on Ayar Labs' close of a $500M Series E round (which included NVIDIA as a participant).
- **In March 2026 Ayar Labs and Wiwynn announced a strategic partnership to build optically connected, rack-scale AI systems, unveiling a joint reference design at OFC 2026** featuring a 100% liquid-cooled, HVDC-enabled architecture that scales to 1,024 accelerators per rack using Ayar's TeraPHY optical engines and SuperNova remote light sources — a design the companies plan to further showcase at Computex Taipei.
- **In September 2025 Ayar Labs and Alchip unveiled co-packaged optics for AI datacenters**. The companies showed working silicon, debuting a multi-chip package (MCP) built with Alchip that integrates eight optical engines, using TSMC's COUPE platform.
- **In October 2025 Ayar Labs formed a strategic partnership with GUC** to integrate co-packaged optics (CPO) into GUC's advanced ASIC design services, targeting hyperscaler AI infrastructure.
- I**n September 2025 Ayar Labs and Alchip unveiled co-packaged optics for AI datacenters**, enabling ultra-high bandwidth, low latency, and efficient multi-rack AI cluster scale-up.
- **In July 2025 Ayar Labs expanded its U.S. and Taiwan operations** and appointed Vivek Khanzodé as VP of Engineering to support high-volume CPO growth.
- **In March 2025 Ayar Labs unveiled the Universal Chiplet Interconnect Express (UCIe) optical interconnect chiplet,** aimed at maximizing AI infrastructure

performance and efficiency.

Fund-raising

Ayar Labs has raised a cumulative ~$1bn since 2018, the most recent rounds in 2026 raised $650.0mn at a post money valuation of between $4-5bn+ . Investors in the most recent round include Neuberger Berman (lead), ARK Invest, Artisan Partners, Greycroft, Insight Partners, Qatar Investment Authority, Sequoia Global Equities, 1789 Capital, AMD, Alchip Technologies, MediaTek, and NVIDIA.

Exhibit 9: Ayar Labs: Historical Equity Funding



The valuation information provided here is included for illustrative purposes only. The information is based on publicly available information (funding rounds, company

Research.

d-Matrix - In-Memory Compute for Fast AI Inference

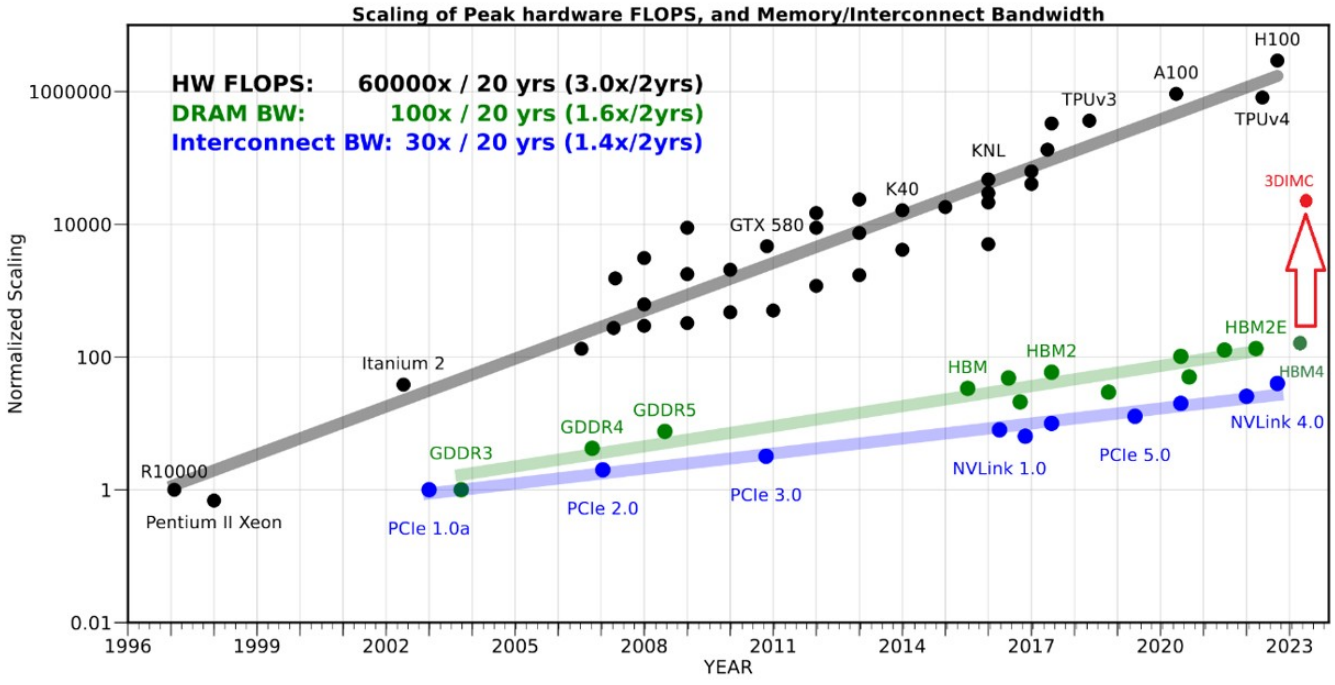
Who is d-Matrix: Founded in 2019, d-Matrix builds an XPU platform for generative AI inference, using digital in-memory computing (DIMC) techniques and a chiplet based architecture to reduce the memory compute bottleneck.

Company Mission: d-Matrix's mission is to dramatically reduce the cost, power consumption, and latency of deploying large language models and other transformer-based workloads at scale, making generative AI commercially viable for enterprises everywhere.

Market opportunity

d-Matrix is targeting low-latency inference and the emerging premium token market, particularly as workloads become more interactive, agenetic, and memory-bound. Management views disaggregated inference as the path forward and highlighted three forms: prefill/decode disaggregation, expert disaggregation, and speculative decoding. The company is targeting speculative decode to start .

Exhibit 10: The growing demand for inference should increase the value of solutions to the memory wall



Memory bandwidth improvement has lagged behind compute performance gains over the past 30 years. d-Matrix 3DIMC technology promises to help close the gap.

Source: d-Matrix [https://www.d-matrix.ai/scaling-ai-inference-with-3dimc/]

Product overview: d-Matrix builds XPUs around a memory-centric architecture designed

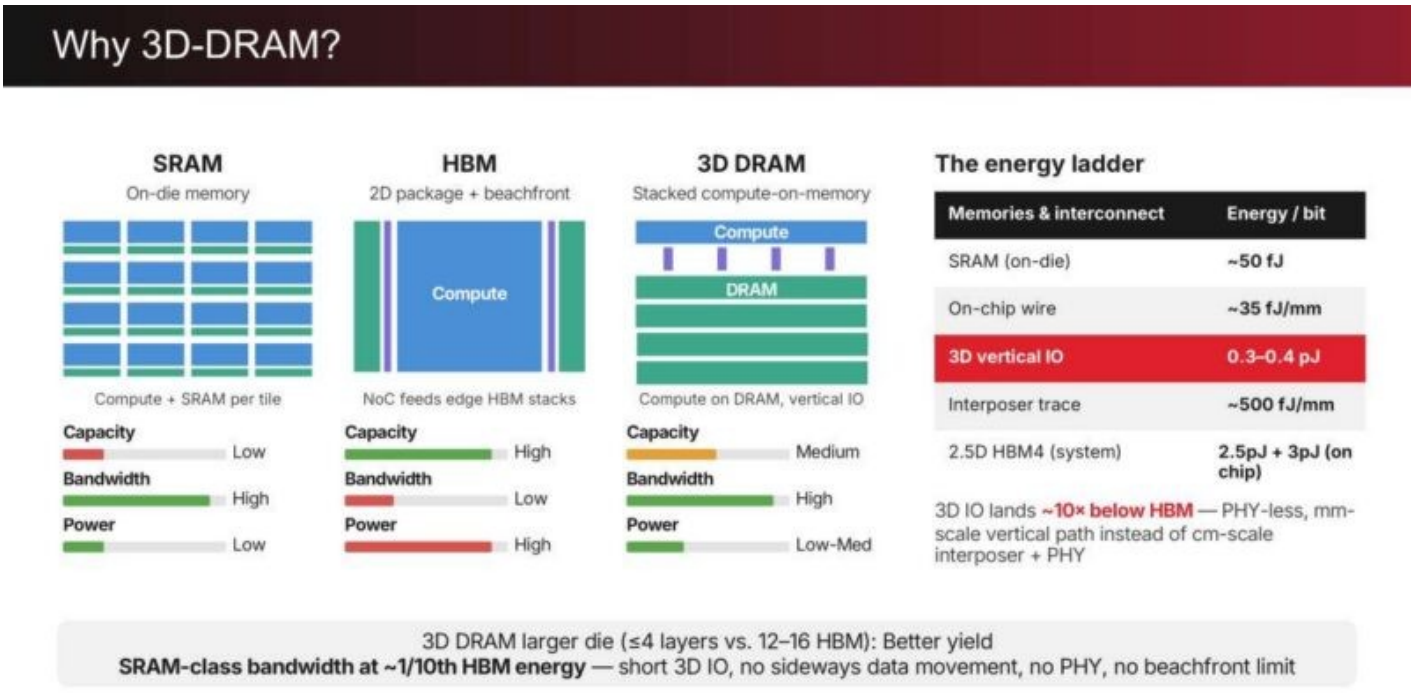
to improve token throughput and latency by reducing memory bottlenecks. This is enabled by Digital In-Memory Compute (DIMC), d-Matrix's proprietary technology that embeds multiply-accumulate logic directly into the SRAM bit cell itself, rather than moving weights back and forth between separate memory and compute units as in a conventional architecture. Because computation happens where the data already lives, DIMC delivers higher memory bandwidth than HBM, while retaining full digital precision (unlike earlier analog in-memory compute attempts).

The current gen product,"Corsair", is an SRAM-based product in production now, the product is a chiplet-based PCIe Gen5 accelerator card combining eight 6nm chiplets, 2GB of on-chip SRAM, and up to 256GB of off-chip LPDDR5X capacity memory, purpose-built for generative AI inference.

The next-gen product, Raptor, adds DRAM stacked directly on top of TSMC 4nm logic dies and is capable of running 2.8T-parameter Kimi 3 on a single rack. The compute is d-Matrix's proprietary Digital In-Memory Compute (DIMC); Raptor will use RISC-V IP from Andes for the control core, which also handles vector compute and activation-function offload. d-Matrix 3D-DRAM can achieve better yield than HBM, with SRAM class bandwidth at 1/10th the power consumption.

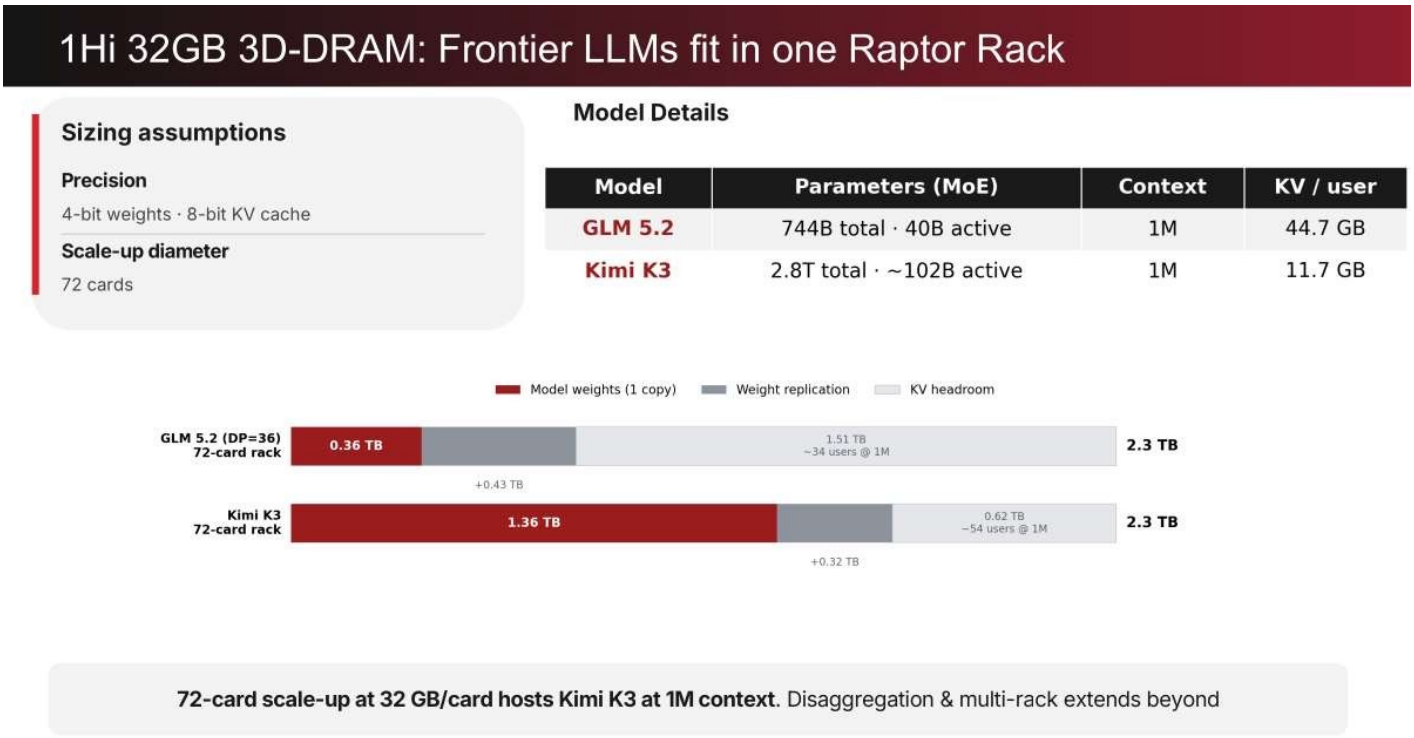
For scale-up, d-Matrix has developed its own die-to-die and card-to-card interconnects, DMX Link connects chiplets within a package at roughly 1 TB/s, while DMX Bridge cards link multiple Corsair cards together into a single all-to-all cluster (up to 16 chiplets). For scale-out d-Matrix invested JetStream, an ethernet-based NIC that connects multiple Corsair-equipped servers into a larger inference cluster. In early 2026, d-Matrix acquired the data center business of GigaIO, adding FabreX-based composable PCIe fabric technology to extend rack-scale scale-up further — building on the companies' prior joint "SuperNODE" reference system, which paired Corsair with GigaIO's fabric to demonstrate GPU-free inference at rack scale.

Exhibit 11: 3D DRAM can offer SRAM bandwidth at 1/10th HBM energy



Source:d-Matrix [https://www.servethehome.com/d-matrix-raptor-3d-dram-accelerator-for-generative-inference-at-hot-chips-2026/]

Exhibit 12: d-Matrix "raptor" can run frontier LLMs on a single rack



Source: d-Matrix [https://www.servethehome.com/d-matrix-raptor-3d-dram-accelerator-for-generative-inference-at-hot-chips-2026/]

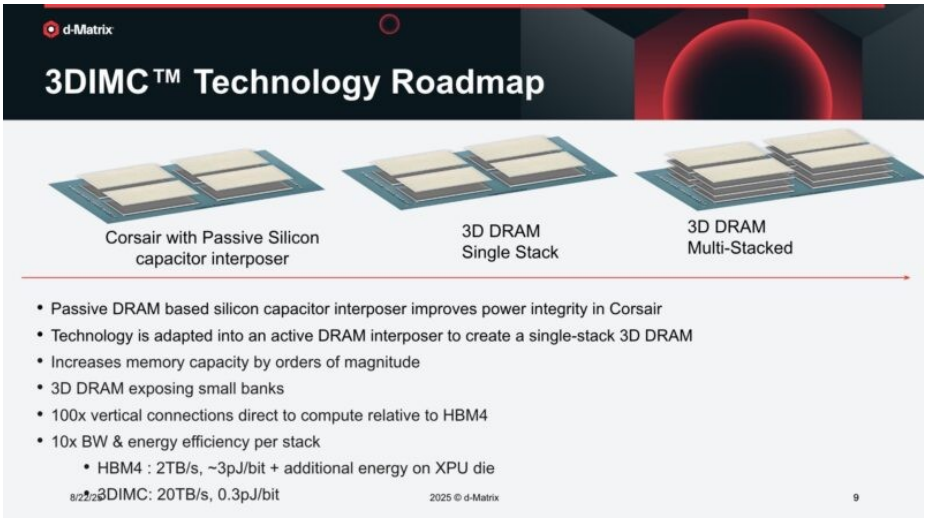
Competitive differentiation

d-Matrix's core differentiation is the memory architecture, with 3D DRAM an industry first technology per the company, positioning the company between HBM enabled GPUs and SRAM based approaches - offering customers higher speeds than HBM and greater capacity than SRAM. This allows d-Matrix to enable disaggregated inference with fewer processors, which becomes more important as memory demands from new models continue to scale. The company has also built its own ethernet-based scale-out interconnect as an alternative to the NvLink networking standard, aiming to give customers flexibility across vendors.

On the software side, the company is investing heavily in its own kernel and SDK strategy to shorten the time it takes to port new models onto its hardware. Kernels are written in C, a widely-understood language that lets frontier AI labs get models running on the hardware quickly, with AI coding tools now capable of accelerating that process further. Management believes this SDK-first approach differentiates the company from other specialized inference hardware players.

Notable Potential Competitors: Broadcom, Cerebras, SambaNova (private), Nvidia (groq), AMD, Etched (private)

Exhibit 13: d-Matrix 3DIMC technology roadmap

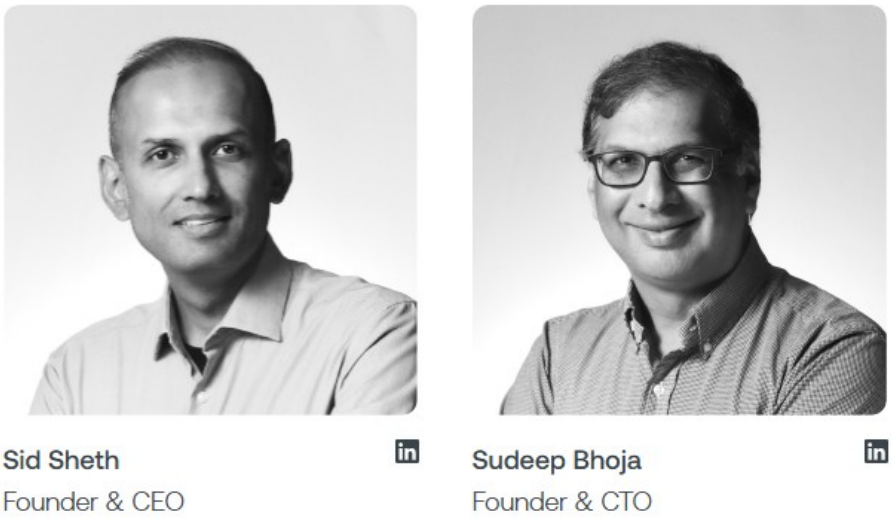


Team

- **Sid Sheth, Co-Founder, President and Chief Executive Officer,** co-founded d-Matrix in 2019 to address the growing memory bottleneck in AI inference, betting early on the shift toward transformer-based generative AI workloads; he previously served as SVP/GM of the Networking Business Unit at Inphi Corp. (2011–2019) and as Director of Marketing for Network Connectivity at Broadcom (2007–2011).
- **Sudeep Bhoja, Co-Founder and Chief Technology Officer,** leads d-Matrix's chip architecture and technical strategy, having co-founded the company alongside Sheth after years of working together on networking silicon; he previously served as CTO of Networking Interconnect at Inphi Corp. (2012–2019) and as Technical

Director at Broadcom (2006–2012).

Exhibit 14: d-Matrix: Key leadership



Recent developments

- **In September 2026, d-Matrix announced a multi-year collaboration with NVIDIA to connect its next-generation Raptor XPUs to NVIDIA's AI infrastructure through NVLink Fusion**, building on NVIDIA's MGX rack architecture alongside Vera CPUs, BlueField-4 DPUs, ConnectX-9 SuperNICs and Spectrum-X networking. The racks are designed to run next to NVIDIA GPU systems such as Vera Rubin NVL72 for disaggregated inference, and d-Matrix is working with Astera Labs on connectivity. Integrated MGX systems are expected initially in the fourth quarter of 2027.
- **In August 2026 d-Matrix acquired Wallaroo.ai**, adding AI inference deployment and orchestration software to its platform
- **In August 2026 d-Matrix presented its Raptor 3D-DRAM accelerator** at Hot Chips 2026, showcasing a stacked memory architecture delivering roughly 100 TB/s of bandwidth at a fraction of HBM4's power consumption.
- **In July 2026 d-Matrix and Parasail announced a deployment combining the Corsair platform with NVIDIA Hopper and Blackwell GPU architectures** within Parasail's cloud data center.
- **In April 2026 d-Matrix acquired the data center business of GigaIO,** including the SuperNODE platform and FabreX PCIe-based memory fabric, extending its platform beyond silicon into full system-level rack-scale infrastructure and adding a new engineering hub in Carlsbad, California.
- **In March 2026 d-Matrix and Alchip announced a collabora** companies describe as the world's first 3D DRAM solution fo at delivering up to 10x faster inference than HBM4-based sol upcoming Raptor accelerator.
- **In November 2025 d-Matrix partnered with Andes Techno** Andes' RISC-V CPU IP into Raptor, d-Matrix's next-generation

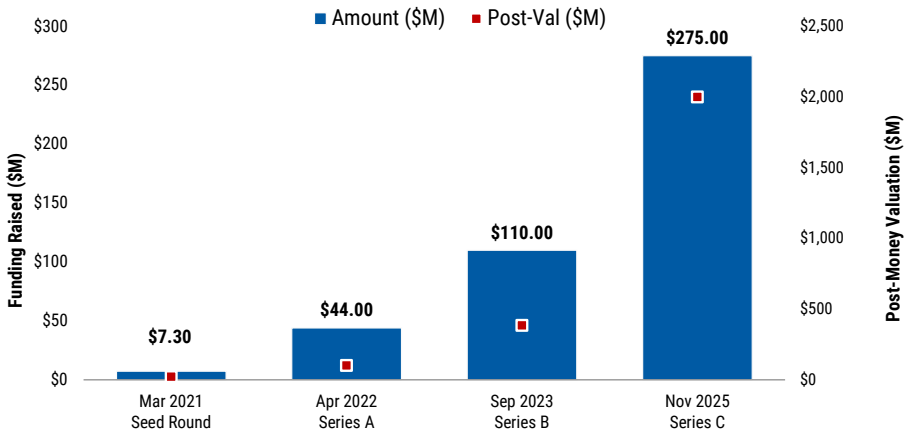


- **In October 2025 d-Matrix announced SquadRack**, a rack-scale reference architecture developed with Broadcom and Arista, offering 3x cost-performance and 10x token speed for AI inference at the Open Compute Summit.

Fund-raising

d-Matrix has raised a cumulative $446.3mn since 2021, the most recent round in November 2025 raised $275.0mn at a post money valuation of $2.00bn. Investors in the most recent round include Bullhound Capital, Triatomic Capital, and Temasek (co-leads), with Qatar Investment Authority, EDBI, M12, Nautilus Venture Partners, Industry Ventures, and Mirae Asset.

Exhibit 15: d-Matrix Historical Funding Rounds



The valuation information provided here is included for illustrative purposes only. The information is based on publicly available information (funding rounds, company

Research.

Eliyan - Breaking down the memory wall with chiplet interconnects

Who is Eliyan? Founded in 2021, Eliyan designs and engineers interconnect technology that moves more data into, and between, AI chips. The portfolio spans three areas: NuLink electrical die-to-die (D2D), chip-to-chip (C2C) and chip-to-module (C2M) PHY IP; NuLight 2D µVCSEL optical interconnects; and a dual-port custom HBM (cHBM) architecture. Simultaneous bi-directional (SBD) signaling raises bandwidth per electrical connection, NuLight targets the "Copper Wall" in AI scale-up fabrics, and dual-port cHBM increases bandwidth available to memory and I/O.

Company Mission: Eliyan's mission is to addresses the data movement limits that constrain AI systems.

Market opportunity

Eliyan is targeting the memory and I/O bottlenecks in AI infrastructure, including die-to-die, chip-to-chip, rack-to-rack, and HBM connectivity. The company believes electrical links are reliable and low-cost, but that as AI scale-up fabrics grow, copper faces simultaneous limits in bandwidth density, power and reach, which it refers to as the Copper Wall. It sees a large opportunity in reducing hop count, lowering AI pod power and improving bandwidth density, with opportunities across ASIC IP, chiplets, memory expansion, scale-up EIC chipsets and optical engines.

Product overview:

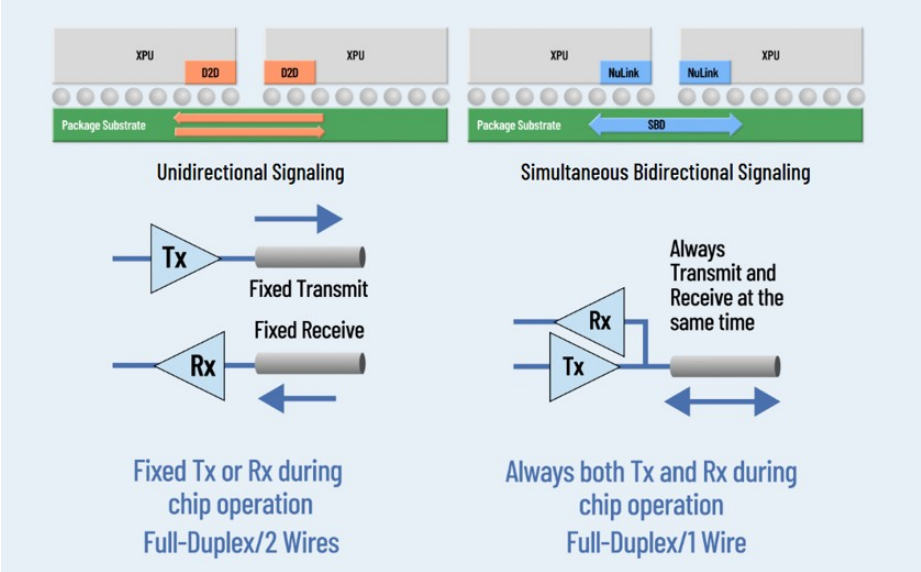
NuLight (optical interconnect): NuLight uses two-dimensional arrays of micro vertical-cavity surface-emitting lasers (2D µVCSELs) to move data optically. Eliyan states that NuLight delivers 8x the bandwidth density, 5x the power efficiency (one-fifth the energy per bit) and 50x the reach of copper links, which would allow more bandwidth between accelerators over larger physical systems without a commensurate increase in power. NuLight's slow-n-wide approach sends data over four 56Gbps NRZ optical lanes instead of one 224Gbps PAM4 lane.

Dual-port cHBM: Eliyan's dual-port custom HBM architecture allows memory and I/O chiplets to use all four edges of the XPU, so memory and scale-up I/O can share the package perimeter rather than competing for scarce die-edge bandwidth.

NuLink (electrical interconnect): Eliyan designs and licenses chiplet interconnect PHY IP that lets multi-die AI chips achieve advanced packaging performance without requiring an expensive silicon interposer. The NuLink family covers D2D (and, via NuLink-X, C2C and C2M) links built around patented SBD signaling, which transmits and receives over the same connection at the same time, which can double the aggregate bandwidth per connection. Because NuLink achieves this on standard organic substrates rather than

costly, size-limited silicon interposers, it extends the reach between chiplets by up to 10x, allowing more HBM or compute chiplets on a single package. NuLink also underpins Eliyan's Universal Memory Interface (UMI), which lets a single ASIC design connect interchangeably to DDR, GDDR or HBM memory chiplets, while NuGear chiplets adapt standard HBM PHY interfaces to NuLink. The technology is protocol-agnostic, supporting both UCIe and BoW. The latest product, NuLink-XD, is a 3nm 224G PAM4 SerDes optimized for chip-to-chip and chip-to-module signaling in markets including NPO and CPO.

Exhibit 16: Eliyan Bidirectional Signaling



Competitive differentiation

In optical Eliyan's NuLight's 2D µVCSEL arrays use less constrained GaAs supply (compared to InP lasers), enable high bandwidth density, with low optical-engine power, and high reliability. In comparison with other optical approaches NuLight offers higher bandwidth density, lower optical-engine power and better reliability than linear-drive optics (LDO) and DWDM silicon photonics (OCI), with the reliability advantage stemming from its patented optical-array repair technology.

Eliyan's differentiates in electrical interconnects by delivering silicon-interposer-class bandwidth on standard, low-cost substrates, removing the need for more expensive advanced packaging process. This is enabled by a patented simultaneous bidirectional signaling architecture that effectively doubles bandwidth per connection at lower power than conventional PHY approaches. A technical approach that took years of dedicated architectural work to develop and is protected by a large patent portfolio.

Because the company's interconnect works on widely available substrates rather than a narrow set of advanced packaging suppliers, it also offers customers a more diversified, lower-cost supply chain, and can improve manufacturing yield and time-to-market versus interposer-based approaches.

Notable Potential Competitors: Synopsys, Cadence, Alphawave Semi (Qualcomm),

Broadcom, Marvell, ams OSRAM

Team

- **Ramin Farjadrad, Co-Founder and Chief Executive Officer**, founded Eliyan in 2021 after recognizing that multi-die packages needed a fundamentally different PHY architecture; he previously served as CTO and VP of Networking/Automotive at Marvell (2019–2021) and as Co-Founder and CTO of Aquantia Corp. (2005–2019).
- **Patrick Soheili, Co-Founder and Chief Strategy and Business Officer**, leads Eliyan's go-to-market and partnership strategy; he has served as Managing Director at Ridgeview Venture since 2008 while also VP of Product Management and Corporate Development at eSilicon (2008–2020).
- **Syrus Ziai, Co-Founder and Chief Operating Officer**, oversees Eliyan's operations and execution; he previously served as VP of Engineering at Nuvia (2019–2021) and SVP of Engineering at PsiQuantum (2016–2019).

Exhibit 17: Eilyan: key leadership



Source: Eilyan

Recent developments

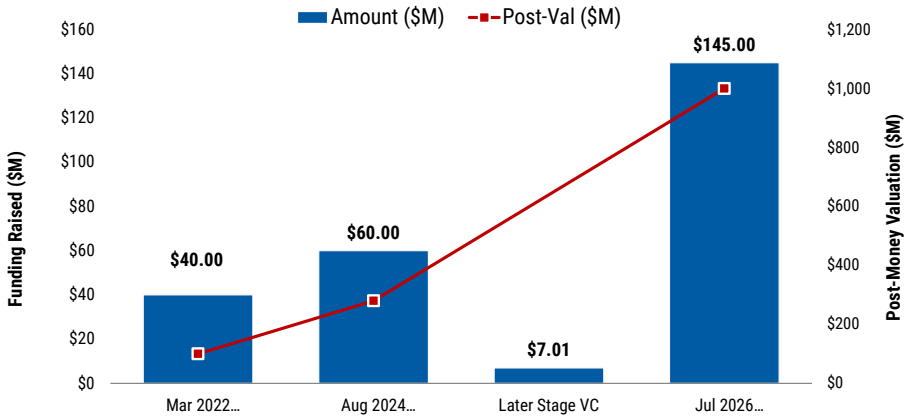
- **In July 2026, Eliyan introduced NuLink-XD, a power-reach configurable 224G PAM4 SerDes** built on TSMC's 3nm process, purpose-built for chip-to-chip and chip-to-module connections feeding into next-generation optical interconnects (including near-package optics and co-packaged optics) — delivering up to 40% lower power than prior designs to support AI scale-up fabrics spanning multiple racks.
- **In November 2025, Eliyan joined the Arm Total Design ecosystem**, supporting the Foundation Chiplet System Architecture (FCSA) that Arm contributed to the Open Compute Project, while also leading a memory-interconnect contribution extending OCP's BoW 2.0 chiplet interconnect specification to support higher-bandwidth memory applications.

Fund-raising

Eliyan has raised approximately $395 million to date and was most recently valued at $1 billion with the close of its $145 million Series C in July 2026 (led by Seligman Ventures). Strategic investors include AMD, Arm, Coherent, Meta, Cisco Investments, Lumentum,

Intel Capital, Samsung Catalyst Fund, Micron Ventures, and SK hynix.

Exhibit 18: Eliyan:



The valuation information provided here is included for illustrative purposes only. The information is based on publicly available information (funding rounds, company

Research.

SambaNova - Memory Centric Computing for Fast Inference

Who is SambaNova? Founded in 2017, SambaNova designs and engineers XPUs, and XPU systems and software platform for enterprise AI, enabling organizations to develop and deploy AI models cloud and on-prem. The core of the product portfolio is the Reconfigurable Dataflow Unit (RDU), which uses a software-defined, dataflow-driven architecture to optimize performance and efficiency for both AI training and inference.

Company Mission: SambaNova's mission is to improve AI hardware by letting software dictate hardware configuration, delivering AI systems that make the most advanced capabilities accessible without requiring unlimited resources.

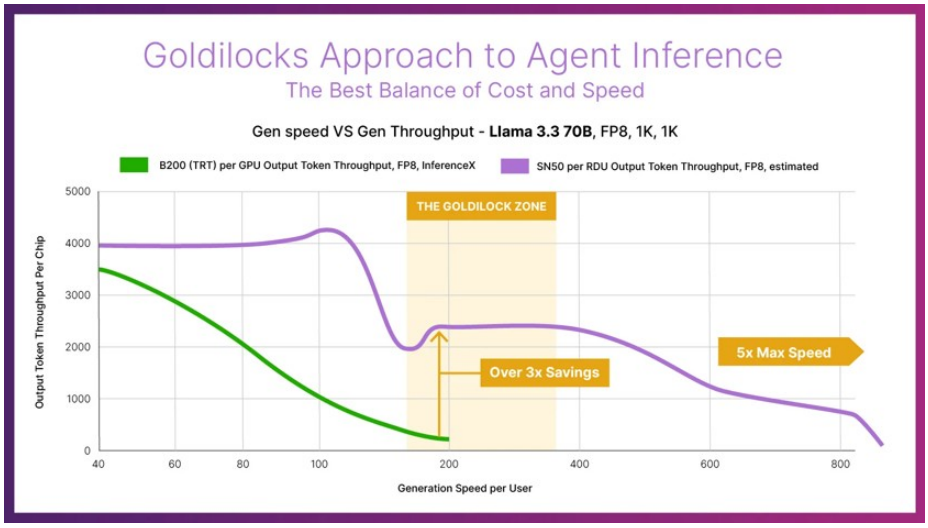
Market opportunity

SambaNova is focused on low-cost, low-latency inference for large models. . As models surpass 1T-parameters and premium token markets are emerging around speed, quality, and always-on agentic workloads. The company sees a need for new hardware solutions to best serve those workloads.

Product overview: SambaNova designs and engineers XPUs built around the Reconfigurable Dataflow Unit (RDU), a processor architecture that's different from the kernel-by-kernel execution model used by GPUs. Rather than moving data between memory and compute units for each operation, the RDU is configured to stream data through a pipeline of compute and memory connected by an on-chip network. Fusing hundreds of operations into a continuous dataflow graph rather than separate kernel calls.

The core of the product portfolio is the dataflow architecture combined with a three-tier memory system (on-chip SRAM, HBM, and directly-attached DDR DRAM), which lets the RDU store large AI models across tiers and switch between them quickly, an advantage in agentic workloads that call on multiple models across a task. The company's software stack (SambaFlow/SambaNova Suite) helps map standard PyTorch and TensorFlow models onto this hardware without requiring hand-written kernels. Commercially, this is packaged into SambaRack (a 16-RDU, air-cooled rack-scale system), SambaCloud (a managed inference platform), and SambaManaged (a turnkey inference deployment for third-party data centers).

Exhibit 19: SambaNova sees significant cost per token advantages vs GPU only solutions



Competitive differentiation

SambaNova's core differentiation is its memory architecture, its RDUs use SRAM, HBM, and DDR memory, reducing dependency on the CPU and PCIe bus. Management contrasted this with GPU architectures and CUDA's kernel-by-kernel temporal mapping, while SambaNova spatially maps workloads.

SambaNova's dataflow architecture fuses many operations into a single streaming pipeline and keeping intermediate results on-chip to minimize memory bottlenecks.

The company also points to power efficiency as a differentiator, positioning its systems as viable in standard air-cooled data centers rather than requiring the liquid cooled infrastructure.

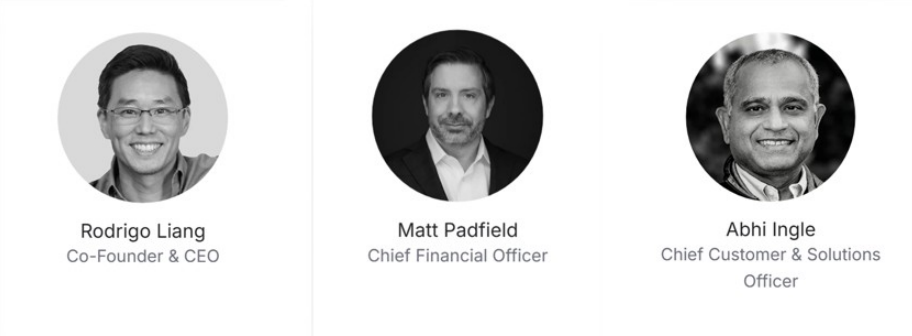
Notable Potential Competitors: NVIDIA, AMD, Nvidia (Groq) , Cerebras, Etched (private), d-matrix (private)

Team

- **Rodrigo Liang, Co-Founder and Chief Executive Officer**, co-founded SambaNova in 2017 after leading a nearly 1,000-person chip design team at Oracle, where he was responsible for SPARC processor and ASIC development; he began his career at Hewlett-Packard and holds BS and MS degrees in Electrical Engineering from Stanford University.
- **Abhi Ingle, Chief Product and Strategy Officer**, leads global product strategy, innovation, and corporate development at SambaNova; he previously built multiple billion-dollar businesses at AT&T and served as Chief Strategy and Corporate Development Officer at Qualtrics through its pre-IPO scale-up, and holds an MBA from Harvard Business School.
- **Matt Padfield, Chief Financial Officer,** joined SambaNova in June 2026 to build the company's financial infrastructure for its next phase of growth; he previously

served as Group Vice President of Corporate FP&A at Coherent and as VP of FP&A and Corporate Development at Synaptics, where he also served as acting CFO and led over $1 billion in M&A activity.

Exhibit 20: Sambanova; Key Leadership



Source: Sambanova

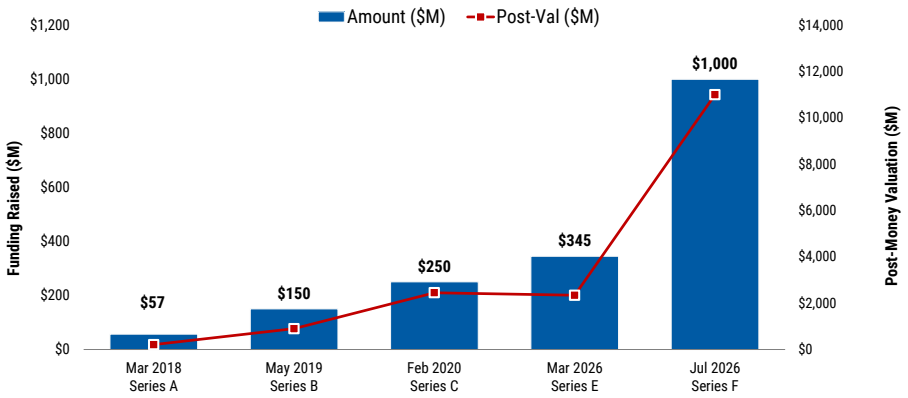
Recent developments

- **In June 2026, Vista Equity Partners and Cambium Capital launched Vector Core Compute (VC2),** the first commercially available enterprise inference cloud for disaggregated inference, backed by a $3.5 billion compute commitment to SambaNova; the platform combines Intel Xeon CPUs, SambaNova SN40 RDUs, and Nvidia Blackwell GPUs, with Together.ai as its first commercial customer and early access extended across Vista's 90+ portfolio companies.
- **In June 2026, at Computex, SambaNova, Intel, and Foxconn demonstrated the first live disaggregated inference system for AI agents** — pairing Nvidia GPUs for prefill with SambaNova's SN40 RDU for decode — running at Vector Core Compute's new data center with Together.ai as the first commercial customer, and announced plans for production rack-scale systems combining Intel Xeon processors with the upcoming SN50 RDU.
- **In April 2026, SambaNova and Intel outlined plans for an integrated hardware system** combining Intel Xeon 6 processors with SambaNova's SN50 RDU for large-scale agentic AI workloads, alongside a distribution agreement with TEPCO Systems to deploy AI infrastructure in Japan; Intel also moved to increase its ownership stake in SambaNova to approximately 9%.
- **In February 2026, SambaNova introduced its fifth-generation SN50 chip** — claimed to offer higher performance than Nvidia's B200-based systems at a lower total cost of ownership — alongside a $350 million Series E round and a multi-year strategic collaboration with Intel, following the collapse of earlier acquisition talks between the two companies.

Fund-raising

SambaNova has raised a cumulative $2.48bn since 2018, the most recent round in July 2026 raised $1,000mn at a post money valuation of $11bn. Strategic investors include Intel, Google Ventures, Vista Equity Partners, and SoftBank

Exhibit 21: SambaNova: Funding History



The valuation information provided here is included for illustrative purposes only. The information is based on publicly available information (funding rounds, company

Research.

SiMa.ai - Purpose-built for Physical AI

Who is SiMa.ai? Founded in 2018, SiMa.ai is an AI software company that builds it's own chips. The company designs and engineers Machine Learning System-on-Chip (MLSoC) SOCs and companion agenetic AI software that bring computer vision, transformer, and generative AI workloads to physical AI devices across industrial, automotive, robotics, and aerospace applications. The core of the product portfolio is the Modalix MLSoC paired with the Palette Software Suite, which together deliver a software-centric "ONE Platform for Physical AI" supporting any model, or modality at industry-leading performance per watt.

Company Mission: SiMa.ai's mission is to help customers deploy high-performance AI/ML physical AI applications with a high bar for security and privacy with partners, radically simplifying physical AI deployment through one unified platform.

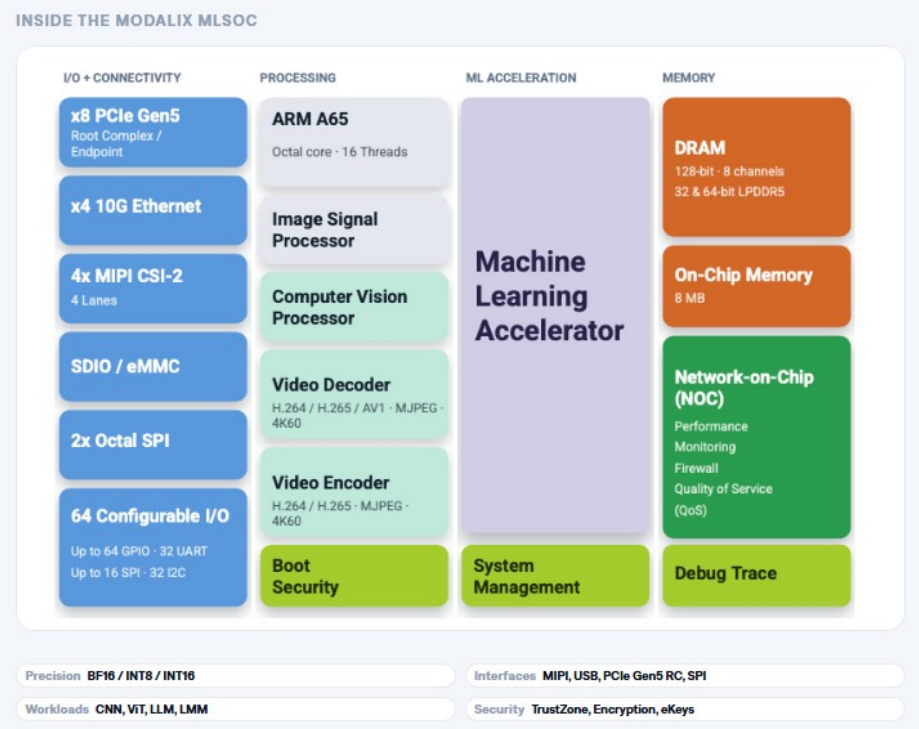
Market opportunity

SiMa.ai is targeting physical AI markets including drones, robotics, automotive, manufacturing, AI-enabled appliances, and healthcare. The company's view is that reasoning models will enable a transition from rules-based computer vision to LLM-based physical AI. Management has argued that the market needs AI-first silicon and software rather than legacy embedded architectures.

Automotive use cases include in-car AI assistants that avoid cloud dependency, run under 10W, support easy LLM upgrades, and deliver low latency. In drones, SiMa.ai positions against incumbents by offering a new class of intelligent drones with real-time Physical AI — integrating vision, sensor fusion, and multimodal AI into a unified architecture for full end-to-end drone applications.

Product overview: SiMa.ai designs and engineers SOCs called Machine Learning Systems-on-Chip (MLSoC) purpose-built for physical AI. SiMa.ai's MLSoC combines a dedicated accelerator, a computer vision processor, and an application compute unit (Arm cores) on a single chip. SiMa.ai expects to release its next generation MLSoC in the first half of 2028.

Exhibit 22: SiMa.ai MLSoC

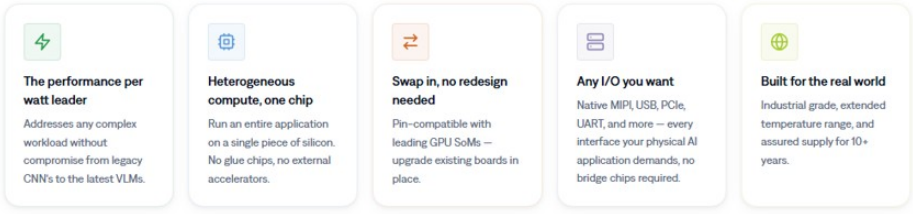


Competitive differentiation

SiMa.ai's key differentiation is low-power reasoning on-device for physical AI applications. SiMa.ai expects to be able to run a 1,000 dense TOPS engine (2x better than the competition) at 80W and has argued it can match Nvidia Orin SoM pricing while consuming meaningfully less power. The company built pin-to-pin replacement SoMs and believes Nvidia is moving toward higher-end Thor platforms while much of the market needs lower-power solutions to enable physical AI applications.

Software scalability is also a major differentiator versus legacy embedded vendors and FPGA models. SiMa.ai recently launched its agentic software called Palette Neat for physical AI that allows developers to describe sensors, cameras, radar, and systems in English, drastically reducing development time.

Exhibit 23: SiMa's key differentiation



Notable Potential Competitors: Nvidia, Qualcomm, Hailo (private), Ambarella

Team

- **Krishna Rangasayee, Founder and CEO**, founded SiMa.ai in 2018 after serving as COO of Groq and spending 18 years at Xilinx, where he rose to SVP and GM of the overall business and grew it to $2.5 billion in revenue at 70% gross margin; he holds an MSEE from Mississippi State University and completed the executive management program at Stanford's Graduate School of Business.
- **Gopal Hegde, Senior Vice President of Engineering and Operations**, joined SiMa.ai in 2020 to lead engineering and operations after more than 30 years in the industry, including as VP/GM of the Server Processor Business Unit at Marvell/ Cavium, COO at Calxeda, and senior engineering leadership roles at Cisco and Intel; he holds an MSECE from the University of Massachusetts Amherst.
- **Harry Kroeger, President of Automotive Business**, joined SiMa.ai in 2023 to lead its expansion into automotive after three decades in the industry, most recently as a member of the board of management at Bosch overseeing a $25 billion automotive electronics business, and 21 years at Mercedes-Benz; he has also served on the boards of Tesla and Rivian.

Exhibit 24: SiMa.ai: Key Leadership



Recent developments

- **In October 2026, SiMa.ai partnered with Wobot.ai to make Wobot's reasoning-based video analytics, available on the Modalix MLSoC as a PCIe card or system-on-module.** The solution processes video on-site at the edge rather than in the cloud, which SiMa.ai says helps address data-privacy requirements in regions such as Europe, Japan and Korea. Wobot.ai's software is deployed across QSR, retail, hospitality and manufacturing sectors.
- **In August 2026, SiMa.ai and ARK Electronics launched the ARK-SiMa.ai Modalix MLSoC Bundles.** The partnership creates production ready, drone-oriented hardware combining SiMa.ai's Modalix System-on-Module with ARK's made-in-the-USA flight controllers and sensors. The systems are designed to be form-factor compatible with the industry's leading GPU-based edge platform, giving drone OEMs a lower power, lower cost drop-in alternative.
- **In June 2026, SiMa.ai released Palette Neat, the first agentic development environment for Physical AI** applications as part of its Palette Software Suite. With Palette Neat, customers can build new or port existing applications seamlessly within days/hours instead of months.
- **In April 2026, SiMa.ai partnered with ARC Aerospace to power its Ruby FC and**

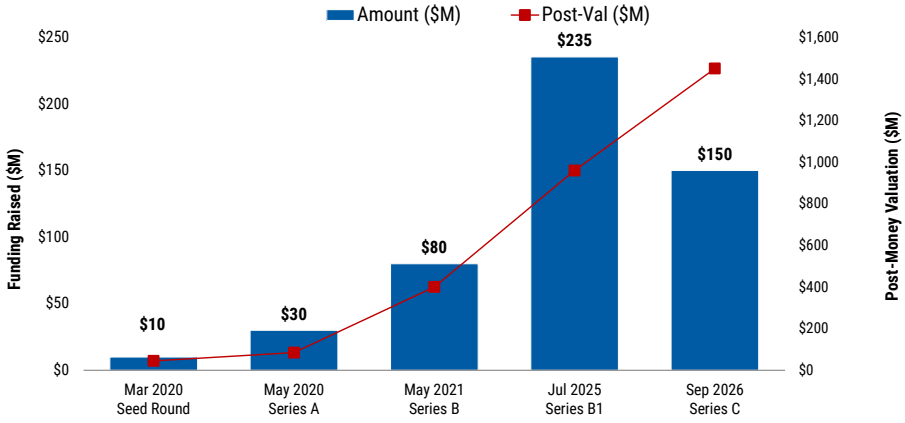
Ruby CEP flight computer series with the Modalix MLSoC family, supporting domestic U.S. production of next-generation flight computers for autonomous systems and counter-drone missiles, including navigation in GPS-denied environments and real-time target acquisition and tracking.

- **In March 2025, SiMa.ai accelerated the release of its second-generation MLSoC Modalix family**, built on TSMC's 6nm process, unifying computer vision, transformers, and generative AI into a single software-centric platform for the first time.

Fund-raising

SiMa.ai has raised about $500mn in total since its founding in 2018, most recently a $150mn Series C announced on September 28th 2026 at a $1.45bn valuation (up from $960mn after the prior round in mid-2025, per PitchBook). The round was co-led by Fidelity Management & Research Company and Amplify, with participation from Dell Technologies Capital, Maverick Capital, Point72, StepStone Group and others. Strategic investors include Micron Technology, which made an undisclosed investment in April 2026, and Dell Technologies Capital.

Exhibit 25: SiMa.ai: Equity funding history



The valuation information provided here is included for illustrative purposes only. The information is based on publicly available information (funding rounds, company

Research.

Tenstorrent - RISC-V Processors for Open AI Compute

Who is Tenstorrent? Founded in 2016, Tenstorrent designs and engineers AI processors, high-performance RISC-V CPUs, configurable chiplets, and open-source software for AI and HPC workloads, spanning everything from desktop development boxes to scalable data center compute. The core of the product portfolio is the Tensix core architecture, which combines an array math unit with RISC-V "baby" processors and a network-on-chip to deliver general-purpose, highly networked AI compute rather than narrowly specialized silicon.

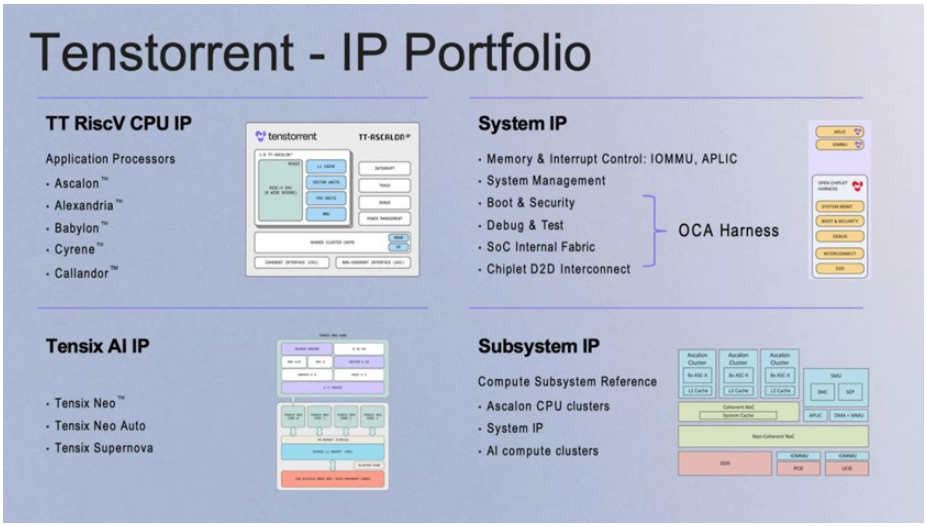
Company Mission: Tenstorrent's mission is general-purpose AI computing, the company positions itself as "anti-specialized," betting that as AI models evolve, hardware built for flexibility across chip types and workloads will outlast narrowly specialized accelerators.

Market opportunity

Tenstorrent is focused on building open AI compute, RISC-V CPU IP, chiplets, and AI systems for edge, automotive, and datacenter. The company sees a large opportunity in providing alternatives to proprietary AI infrastructure. Management has argued that customers increasingly want architectural flexibility and lower-cost licensing options.

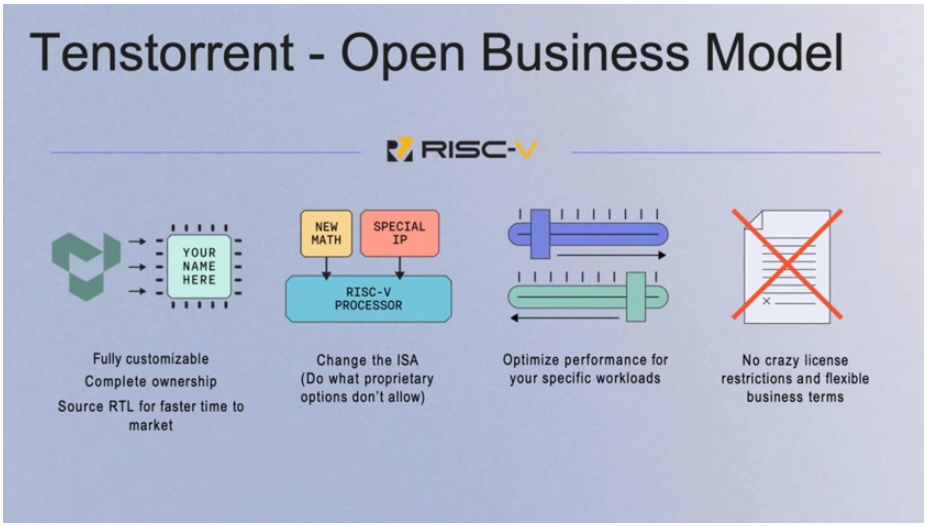
Product overview: Tenstorrent designs and engineers AI accelerators, high-performance RISC-V CPUs, and configurable chiplets. The core of the product portfolio is the Tensix core — a compute tile combining a matrix engine, a vector engine, local SRAM, and five small embedded RISC-V cores. Hundreds of these tiles are connected via an on-chip mesh network, letting any core access any other core's local memory directly rather than routing everything through a shared cache or DRAM — a design intended to reduce the memory-latency bottleneck. The current-generation Blackhole chip pairs 120 Tensix cores with 16 larger, Linux-capable RISC-V cores that manage the chip independently of a host CPU, and ships with an on-chip Ethernet mesh for scale-out rather than a proprietary interconnect. The accompanying software stack, TT-Metalium, is fully open-source under an MIT license, in contrast to Nvidia's proprietary CUDA. Beyond finished hardware, Tenstorrent also licenses its Tensix and Ascalon (RISC-V CPU) IP directly to customers building their own custom silicon, giving the company multiple ways to monetize the same underlying architecture across finished accelerators, chiplets, and licensable IP.

Exhibit 26: Tenstorrent IP Portfolio



Source: Tenstorrent's 10-year journey with RISC-V | embedded world 2026 [https://www.youtube.com/watch?v=A3Z3JIZcMvo]

Exhibit 27: Tenstorrent Open Business Model



Source: Tenstorrent's 10-year journey with RISC-V | embedded world 2026 [https://www.youtube.com/watch?v=A3Z3JIZcMvo]

Competitive differentiation

Tenstorrent's differentiation is its open platform: RISC-V CPU cores, AI compute, chiplets, servers, and open-source software. Management argued that it licenses the whole stack, enabling customers to build what they want. This is positioned against Arm's more restrictive architecture licensing model.

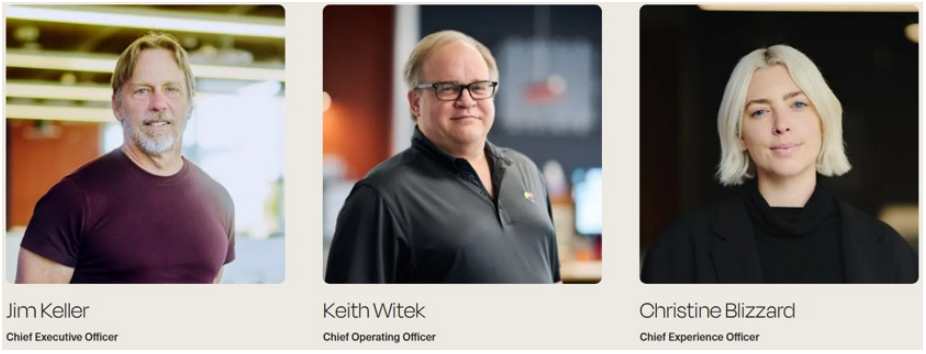
The company has emphasized that chiplets allow customers to move faster and that Tenstorrent can provide CPU, AI, and I/O building blocks. Larger incumbents such as AMD, Intel, and Xilinx have similar technology, but Tenstorrent is making it standard and broadly available.

Notable Potential Competitors: Nvidia, AMD, SiFive (private), Arm, Andes

Leadership Team

- **Jim Keller, Chief Executive Officer**, joined Tenstorrent as CTO and President in 2020 before becoming CEO in 2023; before Tenstorrent, he was Senior Vice President of Silicon Engineering at Intel (2018–2020) and previously led chip architecture at AMD (Zen), Apple (A4/A5), and Tesla's Autopilot hardware team, making him one of the most recognized chip architects in the industry.
- **Keith Witek, Chief Operating Officer**, joined Tenstorrent in January 2023 to scale global operations, finance, and IP licensing; he previously led Strategic Alliances for Google's consumer electronics group, served as SVP of Corporate Development and General Counsel at SiFive, held technology-enablement and legal roles at Tesla, and spent 13 years at AMD, culminating as Corporate VP of Strategy and Corporate Development.
- **Christine Blizzard, Chief Experience Officer**, leads Tenstorrent's brand, marketing, and communications; she previously founded y3k, a brand marketing studio, and worked as a Senior Creative Producer for Google's Events and Experiences Team.

Exhibit 28: Tenstorrent: Key Leadership



Source: Tenstorrent

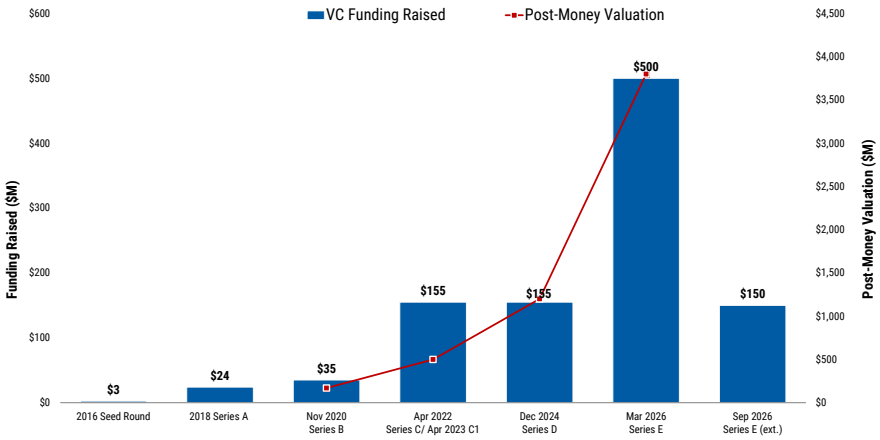
Recent company developments

- **In September 2026, Tenstorrent and ai& launched JapanFold**, a sovereign drug-discovery platform running on Tenstorrent Galaxy superclusters, giving Japanese researchers and pharmaceutical companies access to leading open-source structural biology models (including Boltz-2, OpenFold3, and RFdiffusion 3) with all computation kept entirely within Japan and near-linear performance scaling on Tenstorrent hardware.
- **In July 2026, Tenstorrent partnered with Stealthium** to deliver runtime observability for AI infrastructure running on its hardware.In July 2025, Tenstorrent acquired Blue Cheetah Analog Design, after previously licensing its die-to-die interconnect IP, to strengthen its in-house chiplet interconnect capabilities.
- **In June 2026, Tenstorrent launched TT-Ascalon S, a new RISC-V CPU core**, and expanded its presence across Japan.
- **In May 2026, Tenstorrent launched TT-QuietBox 2**, a liquid-cooled desktop AI workstation built on four Blackhole chips with a fully open-source software stack, able to run models up to 120 billion parameters locally starting at $9,999.
- I**n April 2026, Tenstorrent's Blackhole chip reached general availability**, built on the Tensix core architecture with on-chip Ethernet-based scale-out in place of a proprietary interconnect.

Fund-raising

Tenstorrent has raised approximately $2.4 billion in cumulative funding since 2016. Strategic and institutional investors across its financings include Samsung Securities, Samsung Catalyst Fund, Hyundai Motor Group, LG Technology Ventures, Bezos Expeditions, Fidelity, Baillie Gifford, and others.

Exhibit 29: Tenstorrent: Equity Funding history



The valuation information provided here is included for illustrative purposes only. The information is based on publicly available information (funding rounds, company

Research.