AI Hardware Investment Research

Investment Thesis

We are in the early innings of AI inference demand. Even if frontier model training stopped today, current models are capable enough to fuel years of enterprise implementation and increased usage. Coding assistants already generate constant inference, and voice agents and autonomous workflows will only expand this further. The probability that inference growth slows meaningfully appears low.

This cycle is worth playing. We may be 12 months late, but the opportunity is multi-year. One approach is to invest at the hardware layer.

Why hardware? AI inference is built directly on a defined set of physical components. If we identify the right enablers of inference, we can participate in this cycle regardless of which model or inference implementation wins. This differs from the cloud era, where x86 hardware was commoditised because the bottleneck was software implementation, not compute.

The economic opportunity is substantial. AI inference attacks large budgets. Cloud apps addressed on-prem infrastructure with TCO advantages; AI addresses headcount budgets with even stronger relative TCO. A 10x ROI on a $120k FTE implies $12k ACV per role replaced. As scope expands, we could see far higher ACVs with strong willingness to pay.

Why current architectures create bottlenecks. LLMs process tokens sequentially, which fundamentally rebalances computing requirements. Components that were once minor are now critical. Memory bandwidth, interconnect speed, and packaging density face excessive demand relative to the old cloud paradigm.

Why look beyond Nvidia? First, lesser-known players may offer greater returns. Second, customers will continue diversifying away from Nvidia to improve purchasing power. There are opportunities in companies that succeed regardless of who provides the compute nodes.

Core Assumptions

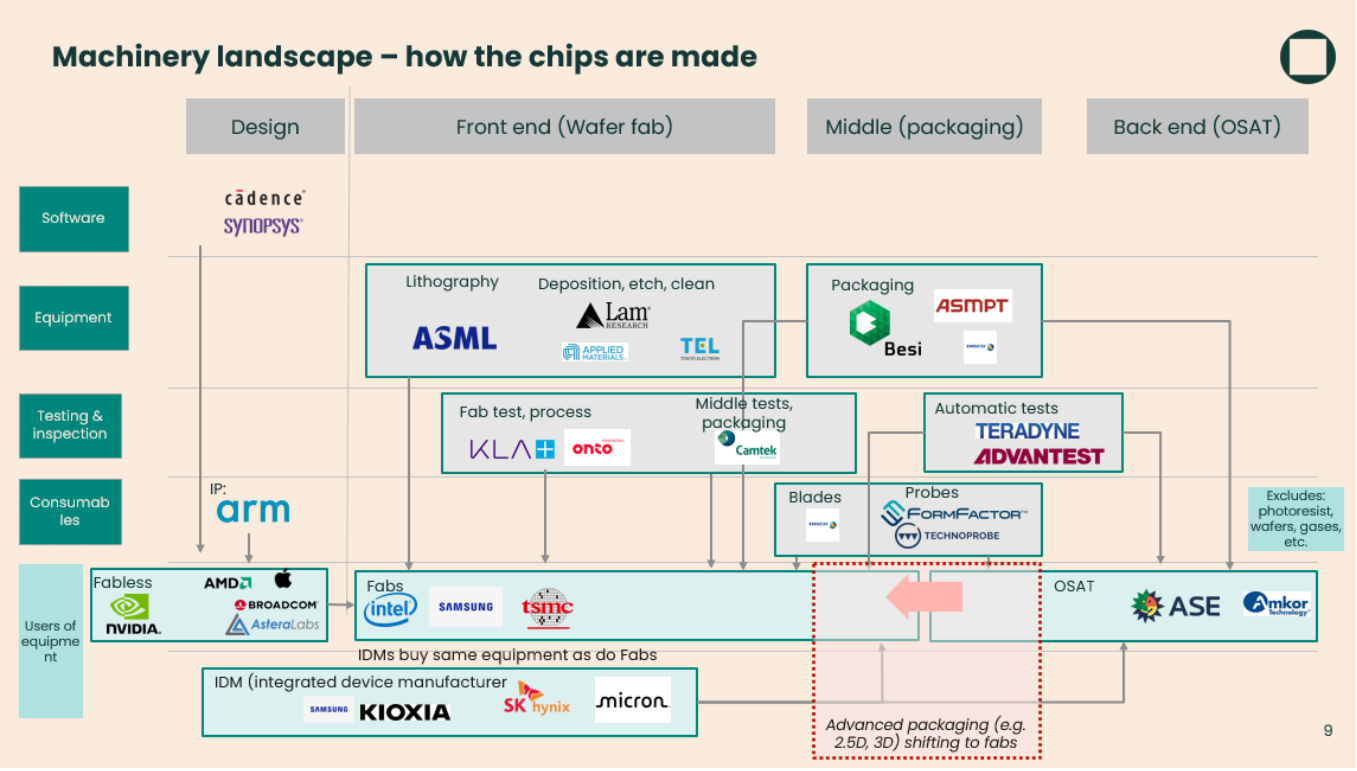
1. We're in early innings for inference demand
2. Current infrastructure cannot satisfy demand
3. Willingness to pay for inference will continue to increase
4. Context and memory become more important (persistent state vs. zero-shot RAG per call)

How Inference is Evolving

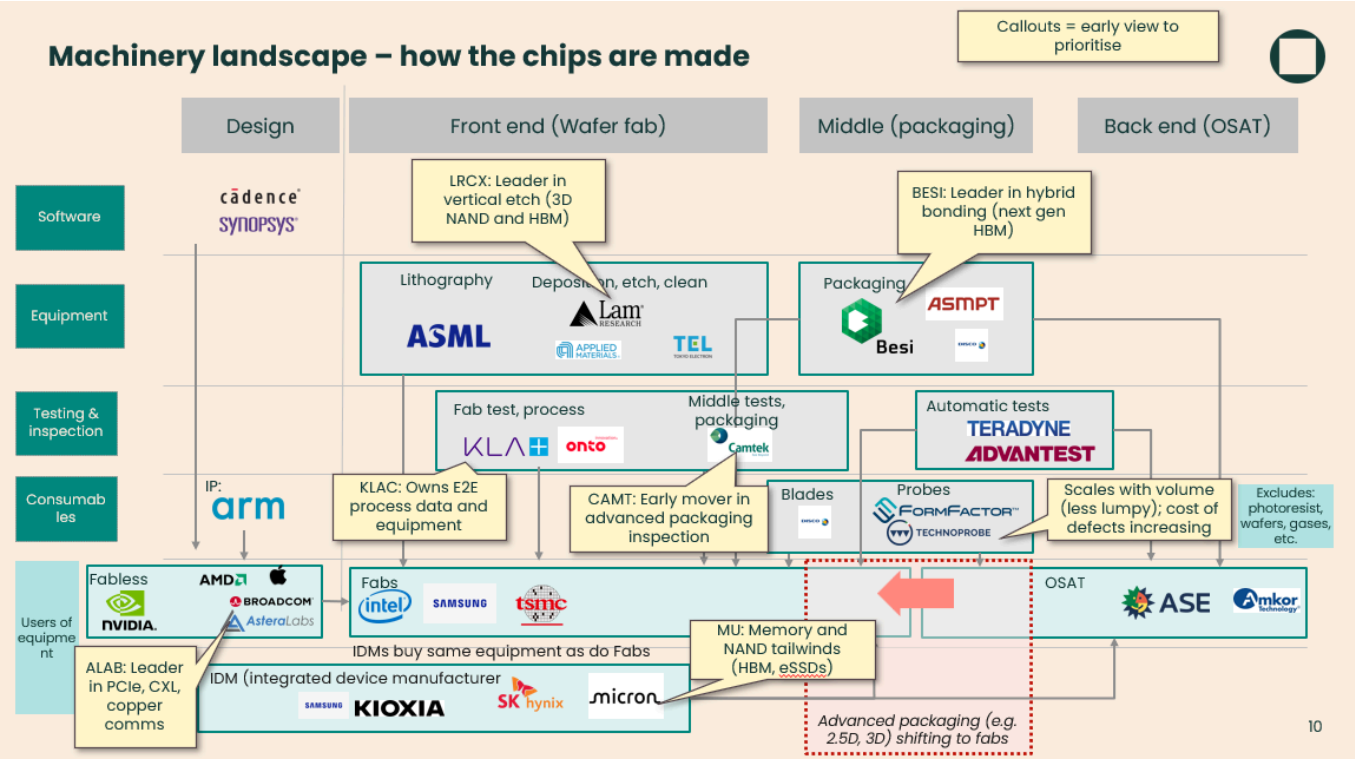
AI inference is becoming more continuous and higher skill. Context windows and persistent state become more important as use cases mature.

The Equipment Landscape

Chips move through four stages: Design → Front End (Wafer Fab) → Middle (Packaging) → Back End (OSAT). Our focus is primarily on Middle and Back End—advanced packaging and inspection/test—where complexity is increasing fastest.



Priority Companies



Priority Research

Memory

The memory wall is the primary bottleneck. KV caches for inference require HBM bandwidth that current supply cannot meet.

Company	Thesis Alignment	Founder-Led
Micron (MU)	US-based memory IDM; one of three HBM suppliers (with SK Hynix, Samsung); sold out through 2026; direct beneficiary of KV cache demand	No (but CEO co-founded SanDisk)

Fabless Components

As bandwidth requirements increase, signal integrity over copper becomes harder. Leading edge retimers and switches become no-regrets purchases.

Company	Thesis Alignment	Founder-Led
Astera Labs (ALAB)	Leads high-frequency retimers and PCIe switches; wins regardless of which GPUs ship; COSMOS software adds diagnostics moat	**Yes**

Equipment

Advanced packaging (HBM stacking, chiplets) requires new manufacturing processes. Equipment providers benefit from tech transitions.

Company	Thesis Alignment	Founder-Led
BE Semiconductor (BESI)	Near-monopoly in hybrid bonding; if HBM4 height limits require hybrid bonding, BESI is the bottleneck	**Yes**
Lam Research (LRCX)	Etch/deposition leader; benefits from node transitions and new fab builds for AI capacity	No

Inspection & Test

Advanced packaging has lower yields. Every defect in a $40k GPU matters. Inspection scales with complexity and intolerance for defects.

Company	Thesis Alignment	Founder-Led
Camtek (CAMT)	Back-end inspection for advanced packaging; benefits from CoWoS and HBM expansion	**Yes** (Israeli)
KLA (KLAC)	Process control leader; "MRI for chips"; scales with yield management needs at advanced nodes	No
FormFactor (FORM)	Probe cards are consumables for chip testing; recurring revenue tied to test volume	No
Technoprobe (TPRO)	Probe card competitor; benefits from same dynamics as FormFactor	**Family-led**

Storage

If context windows and model weights grow, fast storage for spillover and checkpointing becomes more critical.

Company	Thesis Alignment	Founder-Led
Pure Storage (PSTG)	Flash storage for AI data pipelines; benefits from data gravity as AI generates more data	Founder is Chief Visionary Officer

---

Product type comparison

Vendor	Ticker	Component	Est. Price	Purchasing Dynamic	Spend Type
Astera Labs	ALAB	Retimer Chip	$40-$75	1-to-Many: scales linearly with GPU shipments	Volume
FormFactor	FORM	Probe Card	$75k-$150k	Consumable: wears out from testing volume	Consumable
Technoprobe	TPRO	Probe Card	$50k-$100k	Consumable: same as FormFactor	Consumable
Pure Storage	PSTG	FlashBlade	$155k+	Data Gravity: bought as data buckets fill	Capex
Camtek	CAMT	Inspection System	$800k-$1M	Back-End Capacity: packaging line expansion	Capex

Vendor	Ticker	Component	Est. Price	Purchasing Dynamic	Spend Type
BE Semiconductor	BESI	Hybrid Bonder	$2M+	Tech Transition: enables HBM4/chiplets	Capex
Lam Research	LRCX	Etch/Deposition	$3M-$5M	Node Transition: new fabs and smaller nodes	Capex
KLA	KLAC	Plasma Inspector	$5M-$10M+	Yield Management: defect intolerance	Capex
Micron	MU	HBM Stack	$300-$400	Memory Demand: KV cache requirements	Volume

---

Founder-Led Reference

Public - Founder as CEO:

- Astera Labs (ALAB) - Jitendra Mohan
- BE Semiconductor (BESI) - Richard Blickman (since 1995)
- Camtek (CAMT) - Rafi Amit (co-founder, prev. CTO)

Public - Founder Involved:

- Pure Storage (PSTG) - John Colgrove (Chief Visionary Officer)
- Technoprobe (TPRO) - Family-led (founder's nephew is CEO)

Other Notable Founder-Led in AI Hardware:

- NVIDIA (NVDA) - Jensen Huang
- Super Micro (SMCI) - Charles Liang
- Monolithic Power (MPWR) - Michael Hsing
- Dell (DELL) - Michael Dell
- Broadcom (AVGO) - Henry Samueli (Chairman/CTO)
- Arista (ANET) - Andy Bechtolsheim (Chairman/Chief Architect)

Other Companies in the AI Hardware Universe

Memory

Company	Ticker	Notes
SK Hynix	000660.KS	HBM leader; ahead of Micron on HBM3E
Samsung	005930.KS	Investing heavily for HBM4; catching up
Kioxia	Private	NAND flash; former Toshiba Memory
Solidigm	Private	NAND (SK Hynix / Intel JV)
Rambus	RMBS	Memory interface IP and security

Compute / Logic

Company	Ticker	Notes
NVIDIA	NVDA	Dominant GPU provider; **Founder-led**
AMD	AMD	GPU alternative; MI300 competing for AI
Intel	INTC	Gaudi accelerators; foundry ambitions
Broadcom	AVGO	Custom ASICs for hyperscalers; networking
Marvell	MRVL	Custom silicon; data infrastructure
Qualcomm	QCOM	Edge AI; mobile inference
Cerebras	Private	Wafer-scale AI chips; **Founder-led**
Groq	Private	LPU inference chips; **Founder-led**
Google (TPU)	GOOGL	In-house AI accelerators

Foundry

Company	Ticker	Notes
TSMC	TSM	Advanced packaging (CoWoS); node leadership

Equipment

Company	Ticker	Notes
ASML	ASML	EUV lithography monopoly
Applied Materials	AMAT	Deposition, etch, CMP
Advantest	6857.T	Test equipment; duopoly with Teradyne
Teradyne	TER	Test equipment for semiconductors

EDA / Design Software

Company	Ticker	Notes
Synopsys	SNPS	Chip design software; **Founder involved**
Cadence	CDNS	Chip design and verification
Siemens EDA	SIEGY	Mentor Graphics acquisition

Interconnect / Networking

Company	Ticker	Notes
Arista Networks	ANET	Data centre switches; **Founder involved**
Cisco	CSCO	Networking infrastructure
Ciena	CIEN	Optical networking
Amphenol	APH	Connectors and cables

Company	Ticker	Notes
TE Connectivity	TEL	Connectors for high-speed data
Coherent	COHR	Optical components; transceivers
Lumentum	LITE	Optical components
Innolight	300502.SZ	Optical transceivers (China)
Fabrinet	FN	Optical manufacturing; **Founder involved**
Credo Technology	CRDO	High-speed connectivity IP
Alphawave Semi	AWE.L	Connectivity IP; **Founder-led**

Components / Power

Company	Ticker	Notes
Monolithic Power	MPWR	Power management ICs; **Founder-led**
Vicor	VICR	Power modules for AI servers; **Founder-led**
Infineon	IFNNY	Power semiconductors
Microchip Technology	MCHP	MCUs, analog, FPGA
Lattice Semiconductor	LSCC	Low-power FPGAs
Silicon Motion	SIMO	SSD controllers; **Founder-led**
Phison Electronics	8299.TW	SSD controllers; **Founder-led**
Montage Technology	688008.SS	Memory interface chips; **Founder-led**

Storage

Company	Ticker	Notes
NetApp	NTAP	Enterprise storage
VAST Data	Private	AI-native storage; **Founder-led**
WEKA	Private	High-performance file system; **Founder-led**
DDN	Private	HPC and AI storage; **Founder-led**

Systems / Servers

Company	Ticker	Notes
Dell Technologies	DELL	AI servers; **Founder-led**
HPE	HPE	AI infrastructure
Super Micro	SMCI	GPU server specialist; **Founder-led**

Power / Cooling Infrastructure

Company	Ticker	Notes
Vertiv	VRT	Data centre power and cooling
Eaton	ETN	Power management
Schneider Electric	SBGSY	Data centre infrastructure
CoolIT Systems	Private	Liquid cooling for data centres

Template

See Templates/Company Template for the standard analysis framework.