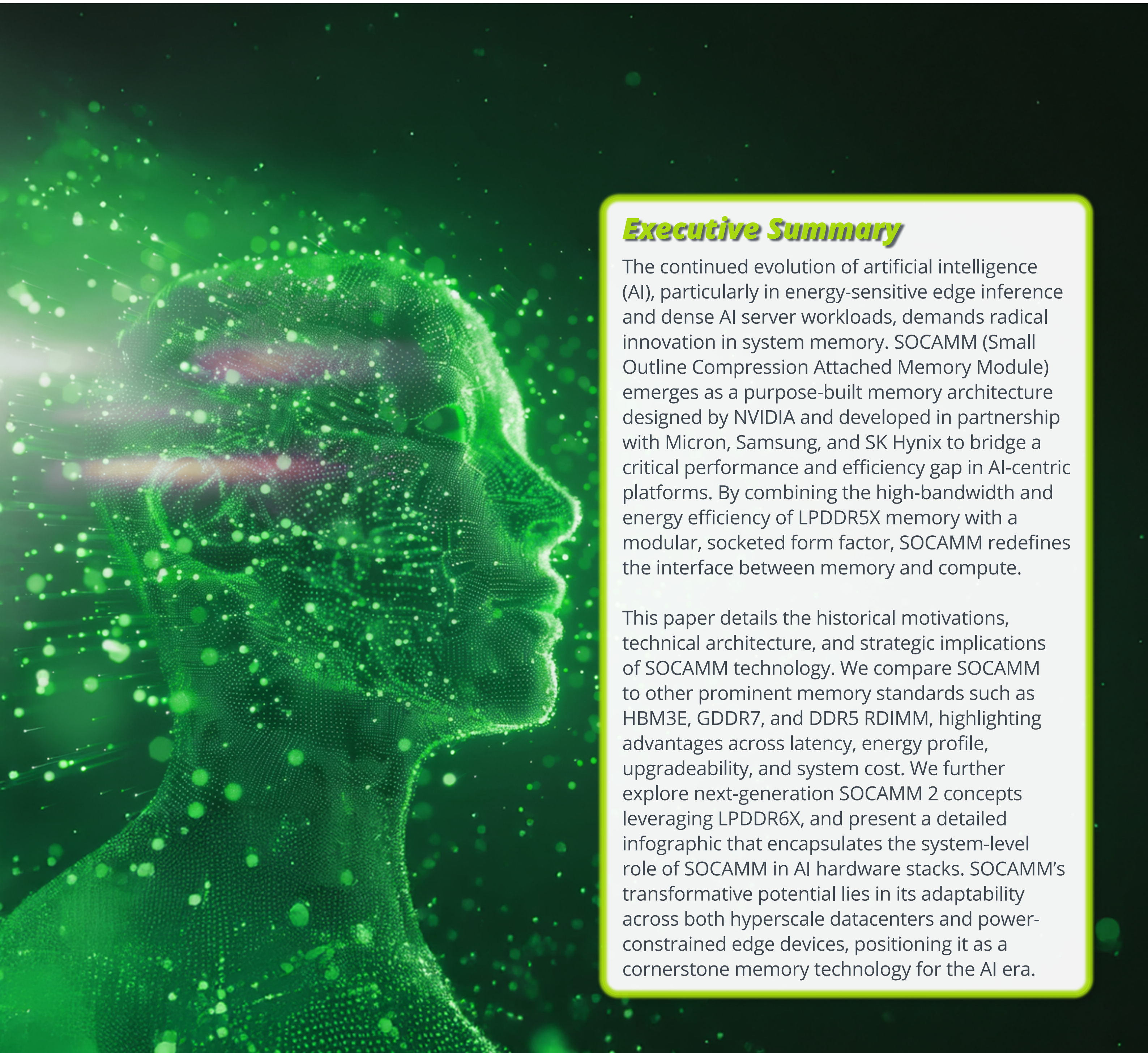
MEMORY
COMPETENCE
CENTER

.MEMPHIS
WE KNOW MEMORY

SOCAMM: THE NEW MEMORY KID ON THE AI BLOCK



Executive Summary

The continued evolution of artificial intelligence (AI), particularly in energy-sensitive edge inference and dense AI server workloads, demands radical innovation in system memory. SOCAMM (Small Outline Compression Attached Memory Module) emerges as a purpose-built memory architecture designed by NVIDIA and developed in partnership with Micron, Samsung, and SK Hynix to bridge a critical performance and efficiency gap in AI-centric platforms. By combining the high-bandwidth and energy efficiency of LPDDR5X memory with a modular, socketed form factor, SOCAMM redefines the interface between memory and compute.

This paper details the historical motivations, technical architecture, and strategic implications of SOCAMM technology. We compare SOCAMM to other prominent memory standards such as HBM3E, GDDR7, and DDR5 RDIMM, highlighting advantages across latency, energy profile, upgradeability, and system cost. We further explore next-generation SOCAMM 2 concepts leveraging LPDDR6X, and present a detailed infographic that encapsulates the system-level role of SOCAMM in AI hardware stacks. SOCAMM's transformative potential lies in its adaptability across both hyperscale datacenters and power-constrained edge devices, positioning it as a cornerstone memory technology for the AI era.

INTRODUCTION AND HISTORICAL CONSIDERATIONS

The acceleration of AI computing, particularly deep learning inference and model deployment at the edge, has placed unprecedented demands on memory subsystems. Historically, memory technologies such as DDR4 and DDR5 RDIMMs have supported general-purpose computing platforms with acceptable performance-per-watt metrics. However, as AI inference increasingly migrates from cloud-bound GPU clusters to resource-constrained edge devices and modular servers, a new memory paradigm is required – one that combines bandwidth, low latency, and efficiency within a small thermal envelope.

LPDDR memory has proven effective in mobile and automotive environments due to its low power characteristics. Yet its soldered integration restricts reuse, repair, and field upgrades, introducing lifecycle limitations and driving up total cost of ownership (TCO). SOCAMM was conceived to reconcile this contradiction: a compact, modular memory interface offering LPDDR-class power efficiency while enabling the flexibility and serviceability long associated with RDIMMs.

THE NECESSITY FOR SOCAMM DEVELOPMENT

The demand for SOCAMM is rooted in a confluence of architectural and commercial imperatives. First, edge inference applications – from autonomous machines and industrial AI controllers to smart retail systems and compact inference clusters – require memory solutions with extreme energy efficiency, physical modularity, and field replaceability. LPDDR5X excels in raw efficiency but has traditionally been locked into soldered-down configurations, reducing system adaptability.

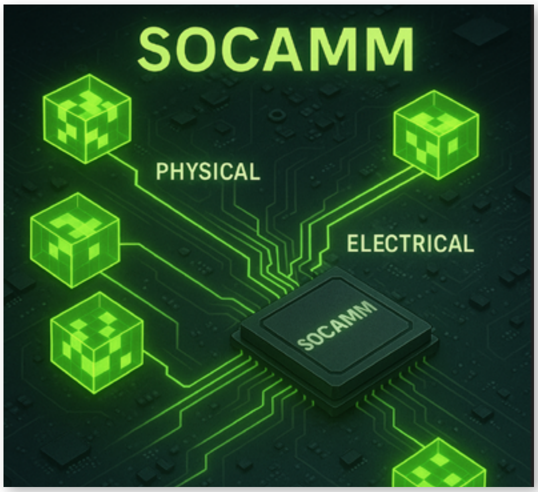
Second, the trend toward modular computing – epitomized by NVIDIA's Grace CPU and Grace-Hopper Superchips – necessitates high-bandwidth memory solutions that match the power and spatial profile of chiplet-based and AI-native architectures. SOCAMM introduces socketed LPDDR5X modules into these platforms, enabling AI OEMs and hyperscalers to dynamically scale capacity, swap failed memory units, or extend system lifespan without full board replacements.

Third, the AI market's need for right-sized, sustainable hardware has led to new procurement models emphasizing cost-per-inference, upgradability, and power-aware design. SOCAMM addresses all three metrics simultaneously.

SOCAMM ARCHITECTURE:
PHYSICAL, ELECTRICAL, AND SYSTEM INTEGRATION

Physical Design

SOCAMM modules measure approximately 14 × 90 mm, occupying roughly one-third the footprint of a standard RDIMM. The form factor is designed to align alongside modular compute packages such as NVIDIA's Grace CPU, enabling tight integration without obstructing airflow or thermally critical components. Modules are attached via screw-secured sockets, allowing easy replacement and enhanced mechanical stability.



Internal Stacking and Capacity

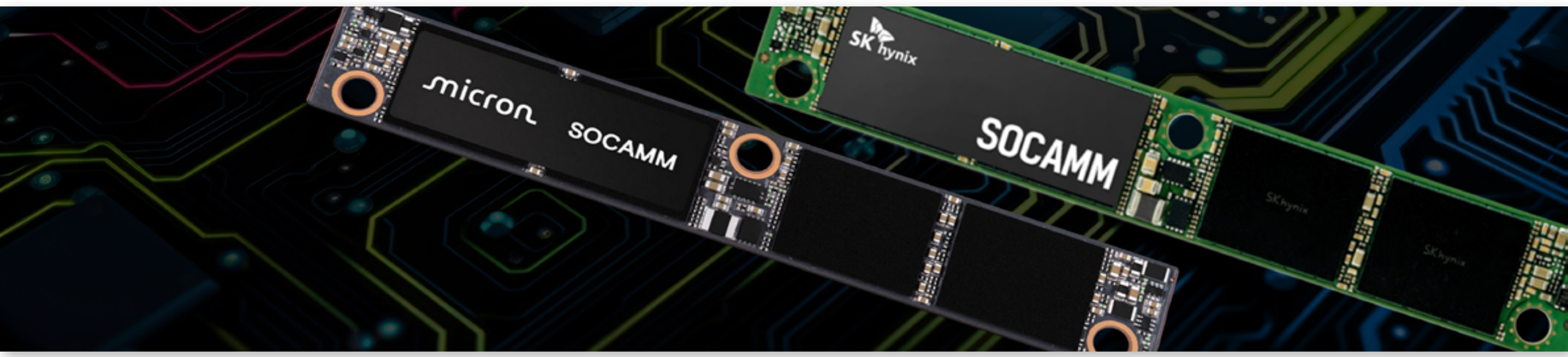
Each SOCAMM module can integrate up to 4 high-density LPDDR5X stacks, with each stack supporting up to 16 memory dies. This results in a total module capacity of 128 GB, comparable to mid-to-high capacity RDIMM configurations. These dies are organized using 64-bit channels to support the high-throughput requirements of AI inference.

I/O and Signaling

SOCAMM supports up to 694 contact pads, a substantial increase over the 260 pins typically found in DDR5 RDIMMs. This I/O density enables higher aggregate bandwidth per module. The memory operates at speeds of 6400 to 8533 MT/s, with next-generation SOCAMM 2 modules expected to exceed 10,000 MT/s with LPDDR6X signaling. These modules interface directly with CPUs and accelerators via a low-voltage LPDDR-compatible PHY and routing layer optimized for minimal signal reflection and jitter.

Power and Thermal Profile

LPDDR5X technology enables SOCAMM to operate at ~1.05V, delivering a power savings of 30–35% compared to DDR5 RDIMM solutions. Passive thermal designs suffice in many edge environments, while compact servers may employ directed airflow or vapor chamber solutions depending on system envelope.



COMPARATIVE ANALYSIS: SOCAMM VS. HBM, GDDR7, DDR5 RDIMM

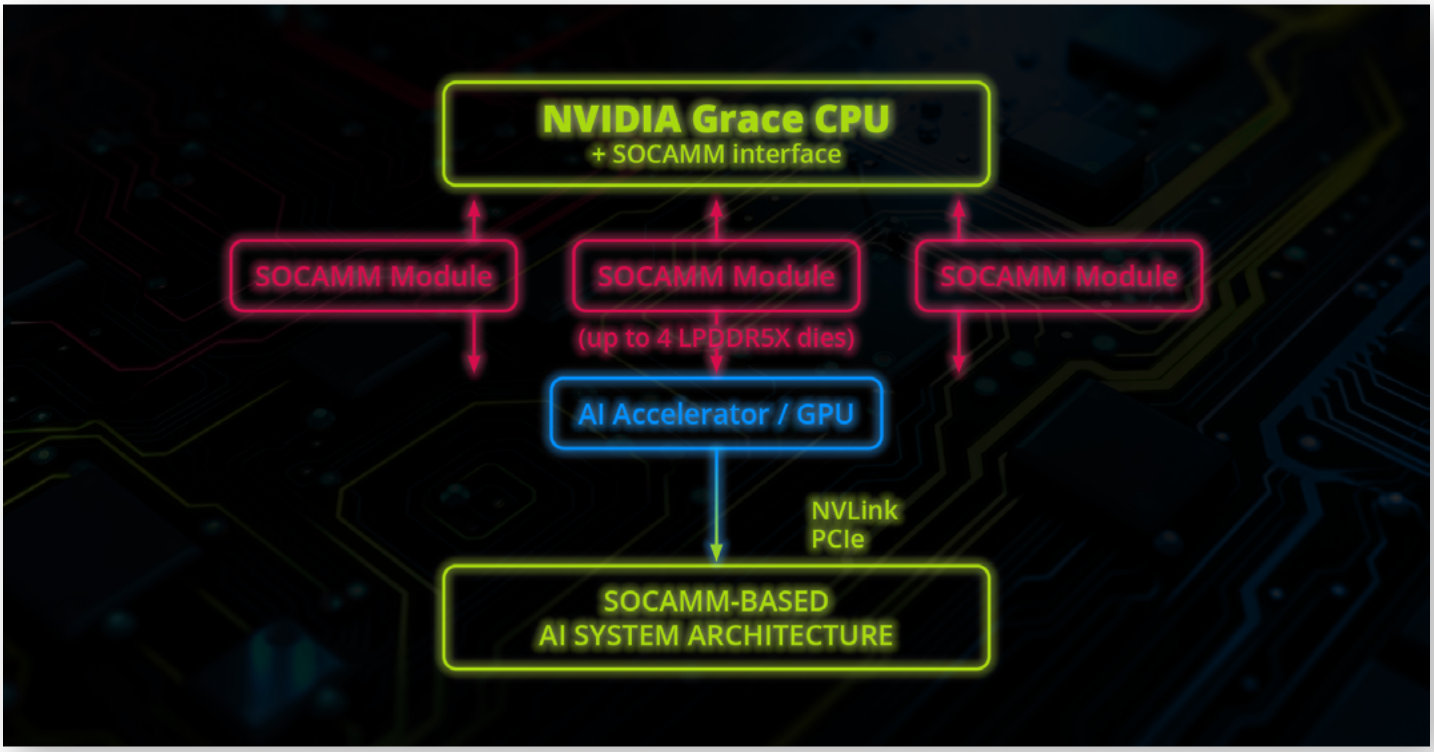
The following table provides a detailed comparison of SOCAMM against key memory technologies across technical and commercial metrics.

KPI	SOCAMM (LPDDR5X)	HBM3E/4 (Stacked)	GDDR7	DDR5 RDIMM
Form Factor	14×90 mm, socketed	Co-packaged/soldered	On-board BGA	DIMM (long)
Capacity per Module	Up to 128 GB	Up to ~36 GB/stack	Up to 32 GB (typ.)	Up to 256 GB+
Peak Bandwidth / Module	~100–150 GB/s	0.8–1.2+ TB/s	~150–200 GB/s	~50–80 GB/s
Data Rate (per pin)	LPDDR5X 6400–8533 MT/s	HBM3E/4 very high	28–32+ Gbps	4800–6400 MT/s
Power Efficiency	Excellent (LPDDR-class)	Good–Moderate	Good	Moderate–Poor
Latency (ns)	~30–45	~5–10	~10–20	~70–90
I/O Count	Up to ~694 pads	Very high TSV/I/O	High	288–560 pins (var.)
ECC / RAS	Platform-dependent	Yes (stack-level + link)	Limited/platform-dep.	Yes (server-grade)
Thermal Profile	Low–Moderate	High (aggressive cooling)	Moderate	Moderate–High
Modularity	High (socketed)	None	None	High (slot-based)
Upgrade Path	Direct swap/scale	Board/module respin	Board rework	DIMM swap
Cost / GB	Moderate	Very High	Moderate–High	Low–Moderate
Cost / GB/s	Favorable for inference	High	Moderate	Moderate
Supply Maturity (2025)	Early ramp	Mature	Initial ramp	Mature
Standardization	Emerging	JEDEC	JEDEC	JEDEC
Primary Use	AI inference, edge, AI PC	GPU training/HPC	Graphics/AI accelerators	General compute servers
Board Area Footprint	Small	Co-packaged	Moderate	Large
Sustainability Impact	High (upgradeable)	Low	Low	Moderate

SOCAMM 2 AND LPDDR6X ROADMAP

SOCAMM 2 will leverage LPDDR6X and introduce PMICs, bandwidth scaling beyond 160 GB/s, and improved thermal and electrical integrity. Expected to launch in 2026, SOCAMM 2 will support adaptive performance for AI PCs, edge clusters, and microservers.

SYSTEM ARCHITECTURE WITH SOCAMM



Typical System Block Diagram

Application Domains

- Compact AI workstations (e.g., NVIDIA “Project Digits”)
- Grace Blackwell server platforms
- Edge-AI inferencing modules in medical, automotive, and manufacturing
- Data center cold-storage memory expansion (hyperscalers)

STRATEGIC AND COMMERCIAL OUTLOOK

SOCAMM 2 will leverage LPDDR6X and introduce PMICs, bandwidth scaling beyond 160 GB/s, and improved thermal and electrical integrity.

Expected to launch in 2026, SOCAMM 2 will support adaptive performance for AI PCs, edge clusters, and microservers.

CONCLUSION

SOCAMM represents a pivotal memory architecture combining modularity, power efficiency, and scalability for modern AI workloads. Its evolution into SOCAMM 2 with LPDDR6X marks a long-term trend toward flexible, sustainable, and high-performance memory platforms in AI computing.

REFERENCES:

1. TrendForce, „SOCAMM Ignites a New Memory War After HBM: Samsung and SK Hynix Join the Race," TrendForce, Jul. 2025. Click here

2. Micron Technology, „Micron Innovates the Data Center and Edge with Industry-First LPDDR5X SOCAMM Module," Micron Investor Relations, Jun. 2025. Click here

3. T. Warren, „Micron and SK Hynix Unveil LPDDR5X SOCAMM: Up to 128GB for AI Servers," Tom's Hardware, Jul. 2025. Click here

4. NVIDIA Corporation, „NVIDIA Grace Hopper Superchip and Blackwell Architecture Overview," NVIDIA Technical Brief, Internal Report, 2024.

5. S. Garg et al., „High-Density LPDDR and Emerging Memory Interfaces for Scalable Edge-AI Inference," arXiv preprint, arXiv:2108.07879, Aug. 2021. Click here

6. N. Florous, „NVIDIA Picks Micron for Massive SOCAMM Rollout," LinkedIn Post, Jul. 2025. Click here

7. JEDEC Solid State Technology Association, „LPDDR6 and Emerging Modular DRAM Interface Specification (Draft)," JEDEC Standards Committee, Jun. 2025. Click here

8. SemiAnalysis, „Memory Bandwidth vs. Cost: SOCAMM vs. HBM vs. GDDR," SemiAnalysis Research, Mar. 2025. Click here

9. Hardware Times, „LPDDR vs DDR5 vs GDDR7: Technical Breakdown for AI Applications," Hardware Times, May 2025. Click here

10. AnandTech, „Micron LPDDR5X SOCAMM Engineering Sample Preview," AnandTech, Jul. 2025. Click here

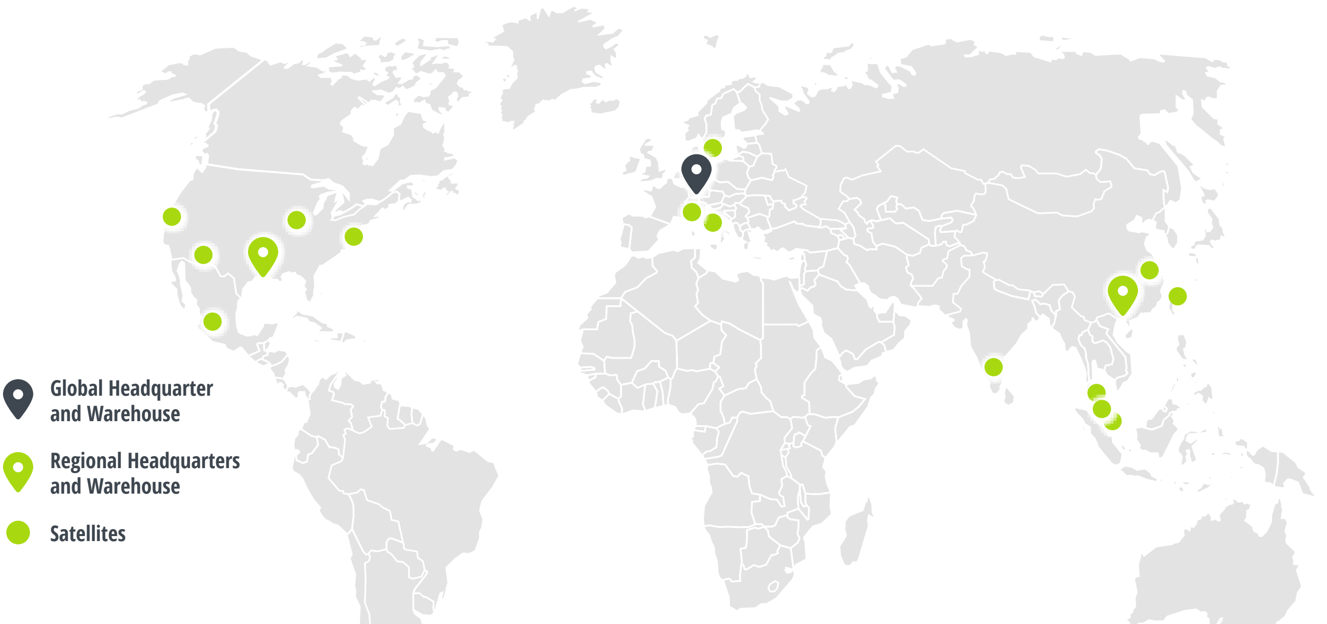
11. Patrick Kennedy, "Micron SOCAMM Memory Powers Next-Gen NVIDIA Grace GB300 Servers," Serve The Home, March 24, 2025. Click here

ABOUT US

MEMPHIS Electronic has been in the memory business for over 30 years. Due to our focus on memory only, we developed into a Memory Competence Center with an unmatched line card of over 18 different memory manufacturers (Samsung, Nanya, SK Hynix, Winbond, Huawei, SkyHigh, Ramxeed, Intelligent Memory, Apacer, Longsys, ESMT, Biwin and many more). We combine this with comprehensive supply chain solutions.

From legacy to latest components and modules, from standard to specialty memories – if it's a memory, we can help. Memory experts in 17 locations worldwide provide regional support and manufacturer recommendations, to ensure customers find the most suitable technology solution for every project.

MEMPHIS LOCATIONS



EMEA
Frankfurt am Main, DE
Copenhagen, DNK
Geneva, CH
Milan, IT

AMERICA
Houston, TX
Boston, MA
Chicago, IL
Phoenix, AZ
San Jose, CA
Guadalajara, MEX

APAC
HongKong, HKG
Shanghai, CN
Taipei, TW
Hsinchu, TW
Singapore, SGP
Penang, MY
Kuala Lumpur, MY
Bengaluru, IN

CONTACT US

For further information, data sheets or samples, please reach out.
We are only an email away: sales@memphis.de

MEMPHIS Electronic GmbH
www.memphis.de
info@memphis.de