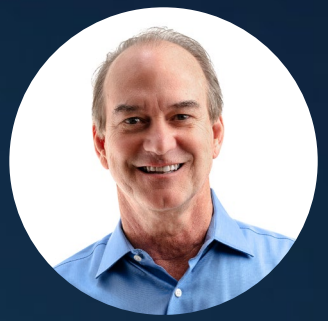
MEXT

AI-Powered Predictive Memory:
DRAM Performance at Flash Cost



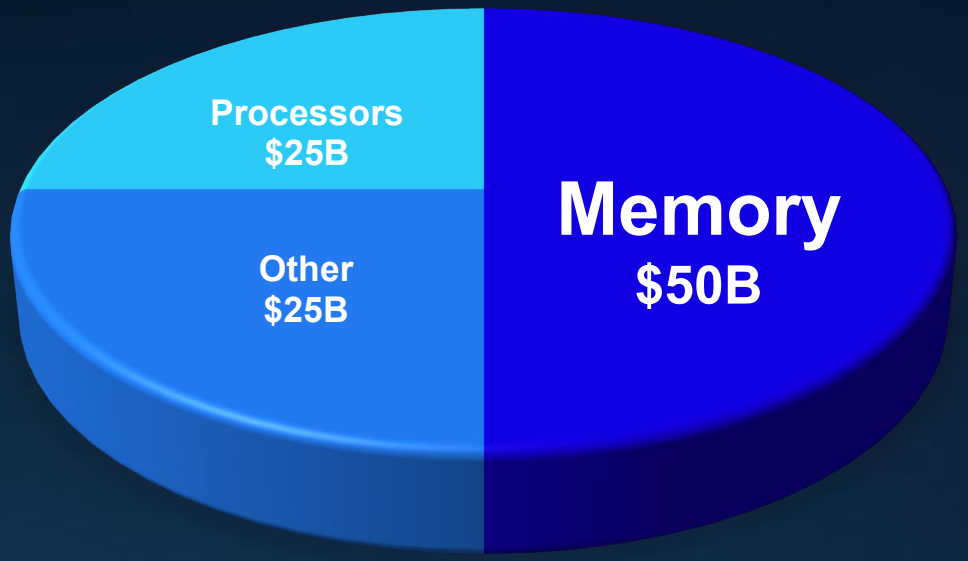
Gary Smerdon
Founder and CEO, MEXT
August 6th, 2025

MEXT FMS

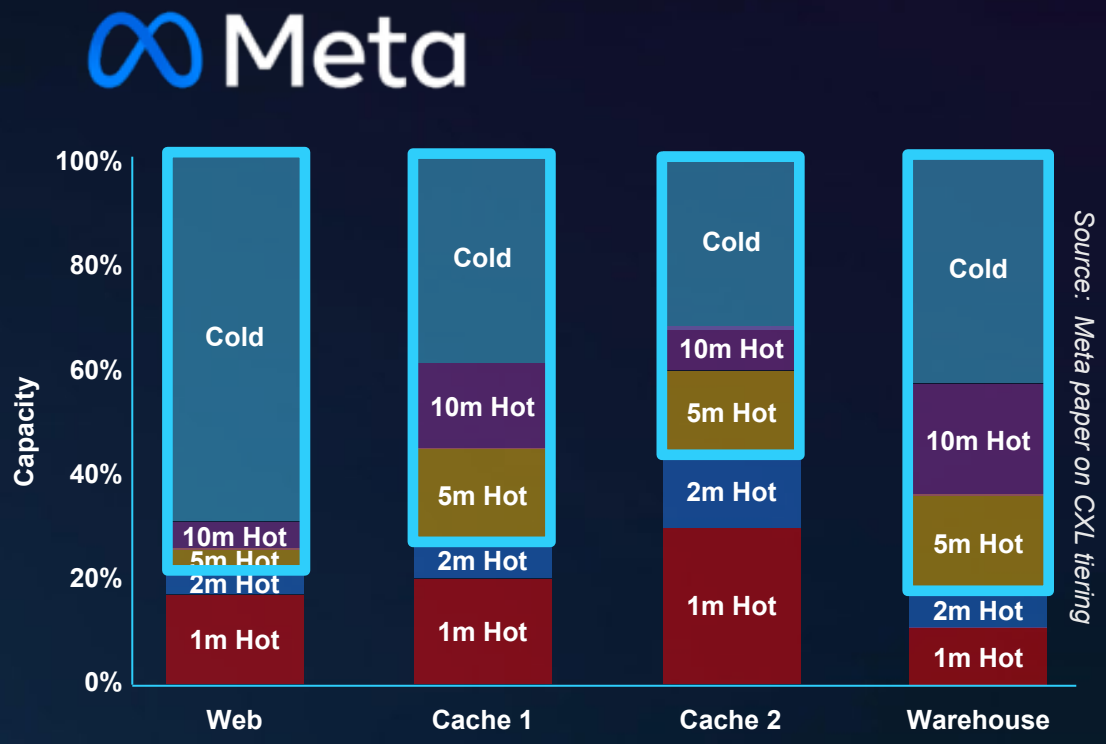
Market Context & Intro to MEXT

- Modern applications (AI, large-scale databases) are putting tremendous pressure on data center computing resources
- Examples: Redis, Neo4j, Spark, Moonray, Oracle, Memcached, DeepSpeed…
- Workload scale keeps increasing, data volume and velocity keeps increasing
- Requires massive amounts of memory, storage, and processing power
- As a result, the cost of computing is skyrocketing
- MEXT founded in 2023
- Our mission: radically lower the cost of computing for large-memory workloads
- Tackling one of the largest cost components: server memory (DRAM)
- Core product: AI-Powered Predictive Memory (NOT persistent memory)

DRAM is a Huge Cost, But Poorly Utilized



DRAM = **50%** of server cost



Memory is often **< 50%** utilized

MEXT FMS

3 Keys to Drive Down Memory Cost

Increase DRAM Utilization

Eliminate overprovisioning and stranded DRAM

No Hardware or Software Changes

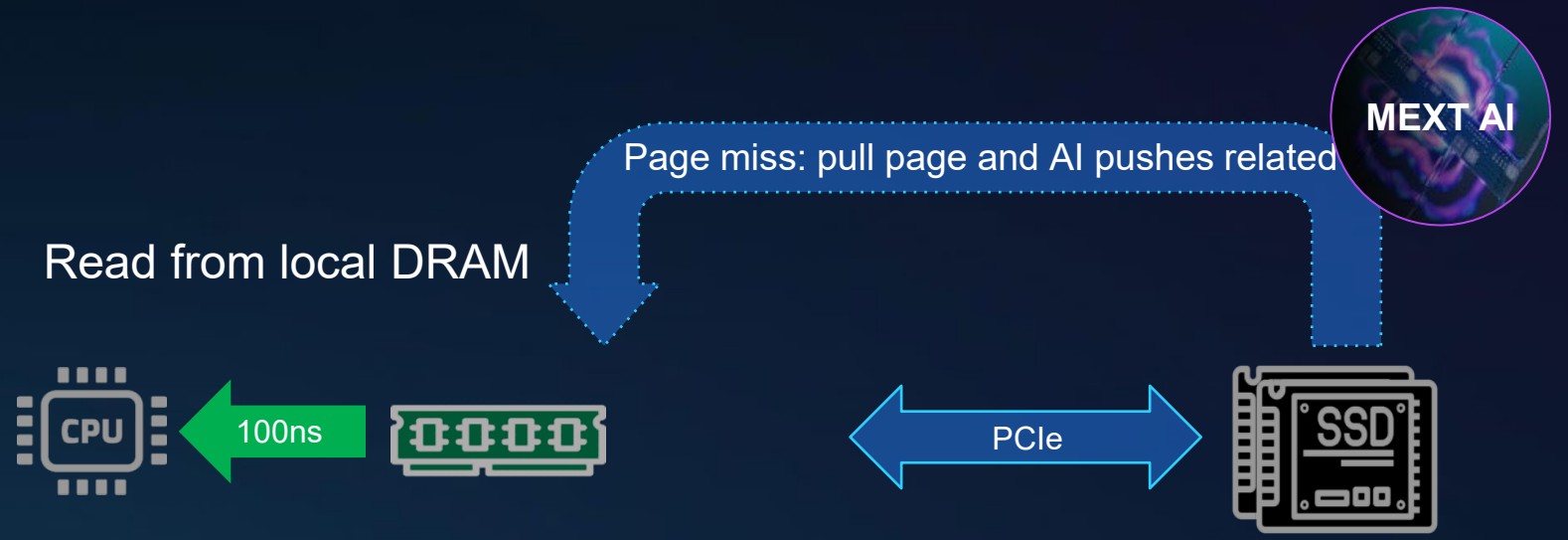
HW / architectural and SW changes drive up cost

Bring Flash into Memory Tier

Flash is 95% lower cost-per-bit vs DRAM

AI-Powered Predictive Memory

AI engine predicts and transparently tiers memory



- AI Push Engine predicts what will be needed in advance
- Places memory pages optimally before they are requested
- Goal: all but initial miss happen as quickly as possible

Workload Examples

Relational In-Memory Database

ORACLE

Analytics

APACHE Spark™

In-Memory Key Value

redis

Caching Key Value

MEMCACHED

Graph Database

neo4j

AI-Native Vector Database

weaviate

AI Large Language Model

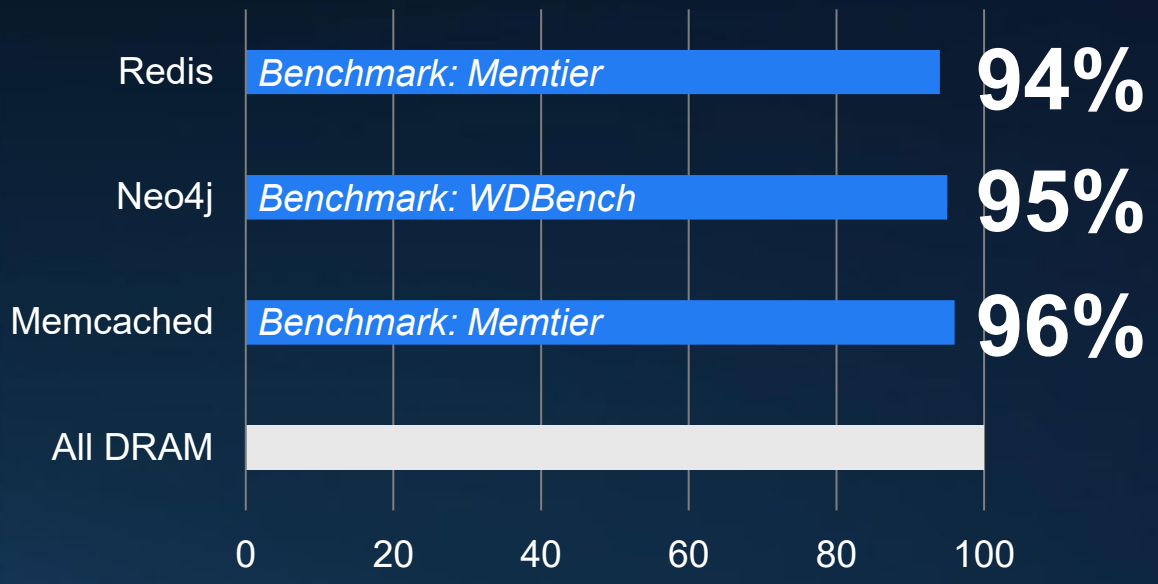
DeepSpeed

Animation Rendering

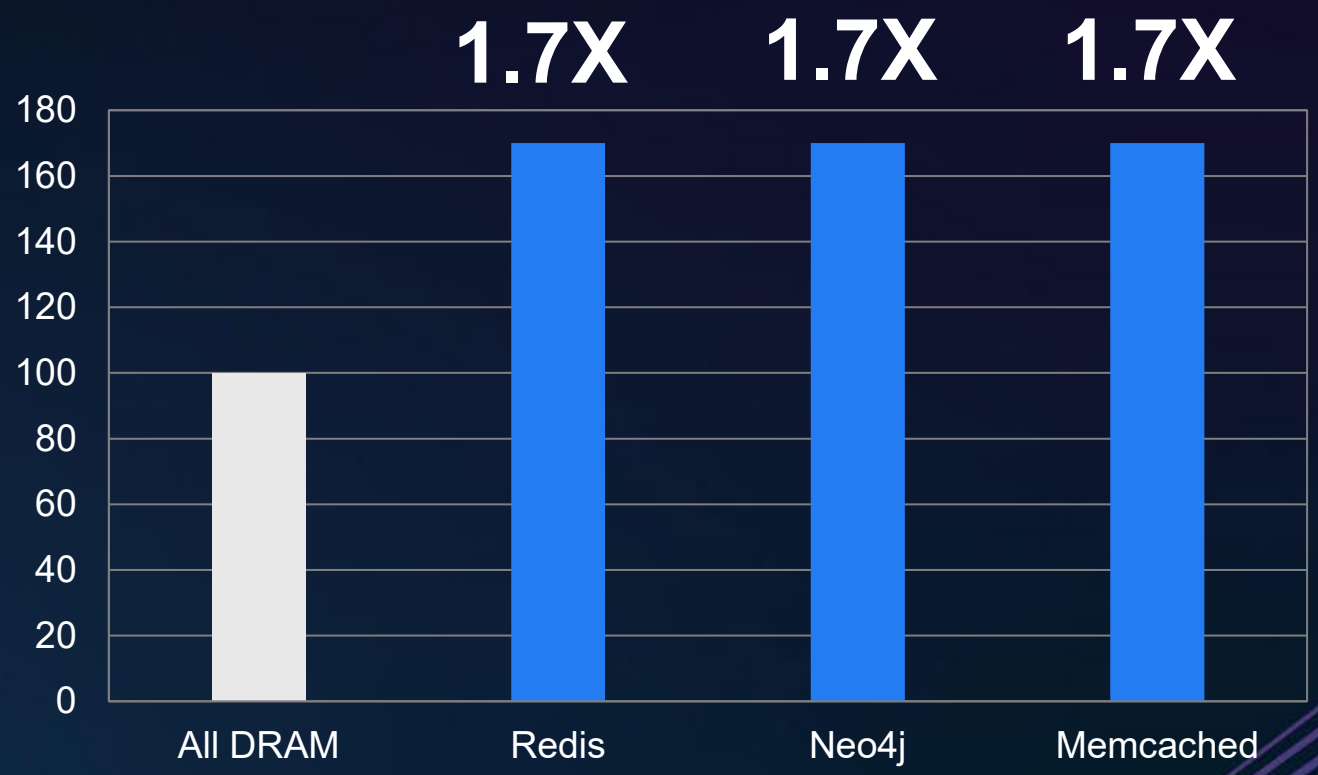
MOONRAY

MEXT FMS

Performance



Performance / $



MEXT FMS

Comparing all-DRAM system vs. 50% DRAM & 50% MEXT+Flash system

Customers

Customer	Workload	Outcome
Leading Global Animation	Animation / Rendering	2X Memory – More Productivity
Major Financial Institution	Graph Analysis	Faster Scanning of TBs Data
Large SaaS Company	Database	Lower Cloud $ / More Memory
Leading EDA	EDA Tools	Larger Systems - No OOM

Partners

AMD

aws

Google Cloud

KIOXIA

ORACLE CLOUD

MEXT

FMS

Takeaways

IT teams can leverage MEXT to either:

- **Reduce DRAM by 50%** or more for radically lower computing costs and power expenditure
- Keep DRAM the same but leverage MEXT + Flash to **double the effective memory capacity** in the system, leading to superior application performance for memory-bound workloads
- Get an **“insurance policy”** to never run out of memory for workloads that generally use their memory well but occasionally experience spikes