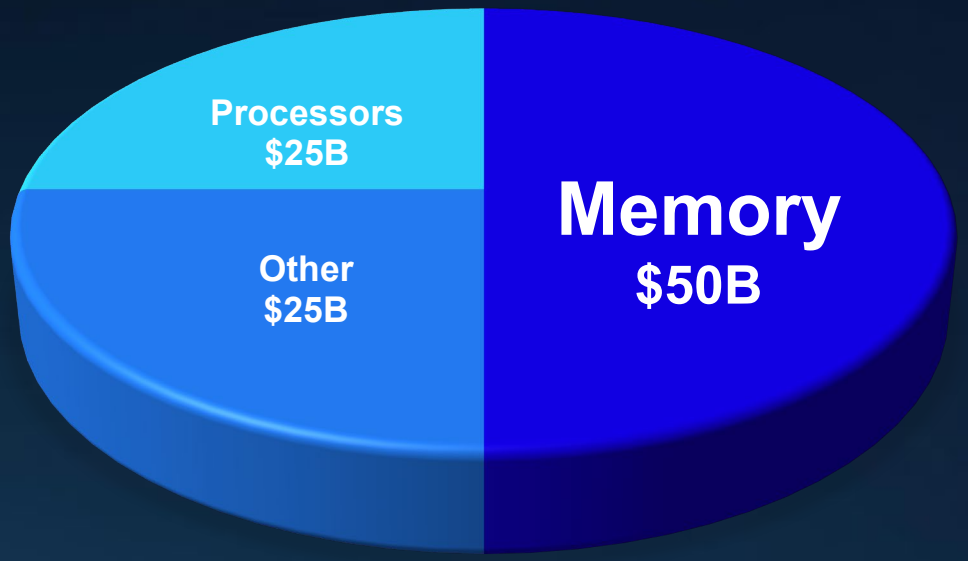
Flash Memory Summit

MEXT

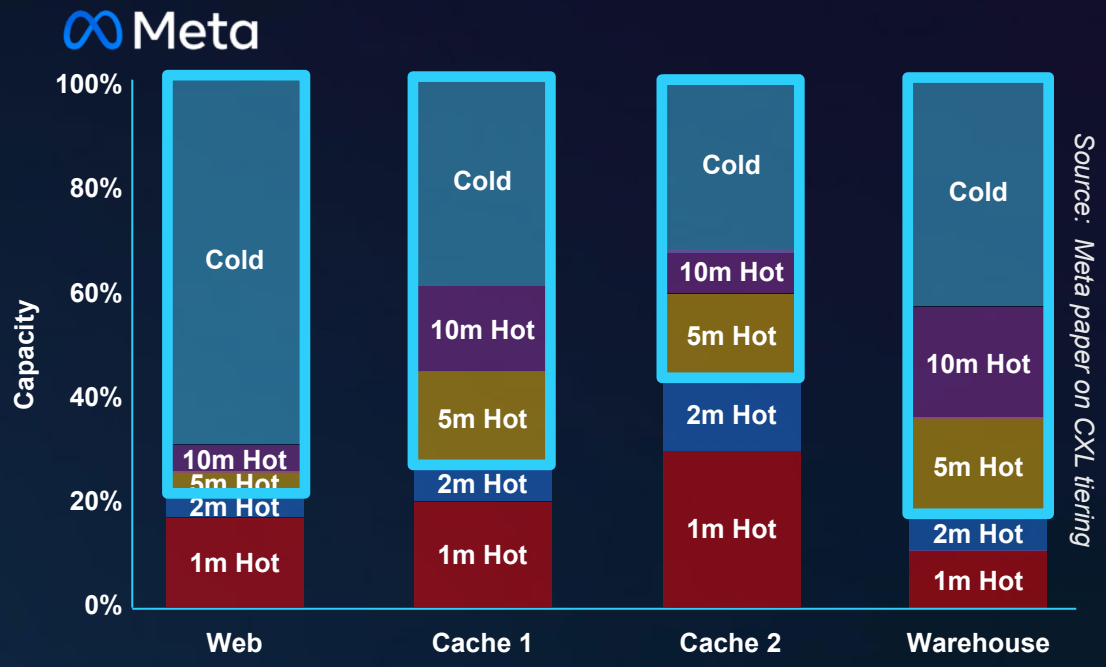
FMS

the **Future** *of* **Memory** *and* **Storage**

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

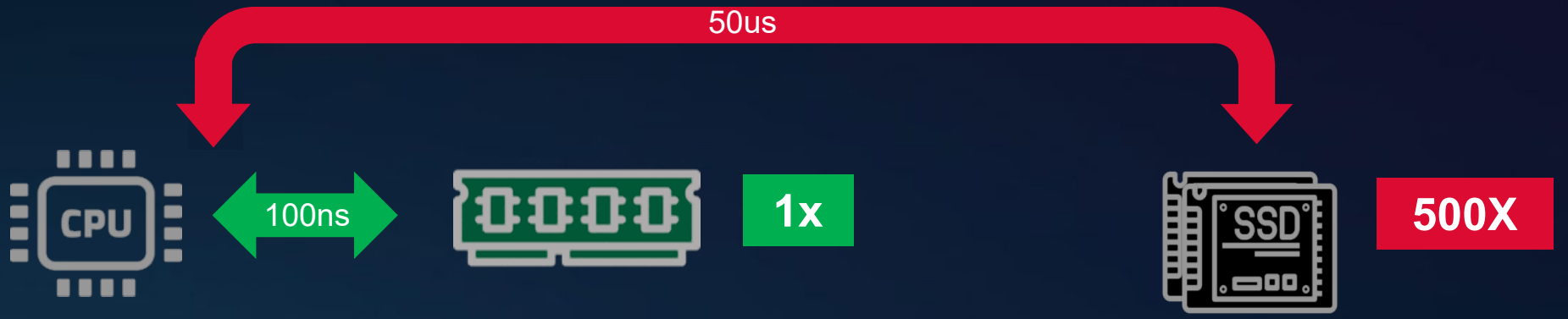
No Hardware or Software Changes

HW / architectural and SW changes drive up cost

Bring Flash into Memory Tier

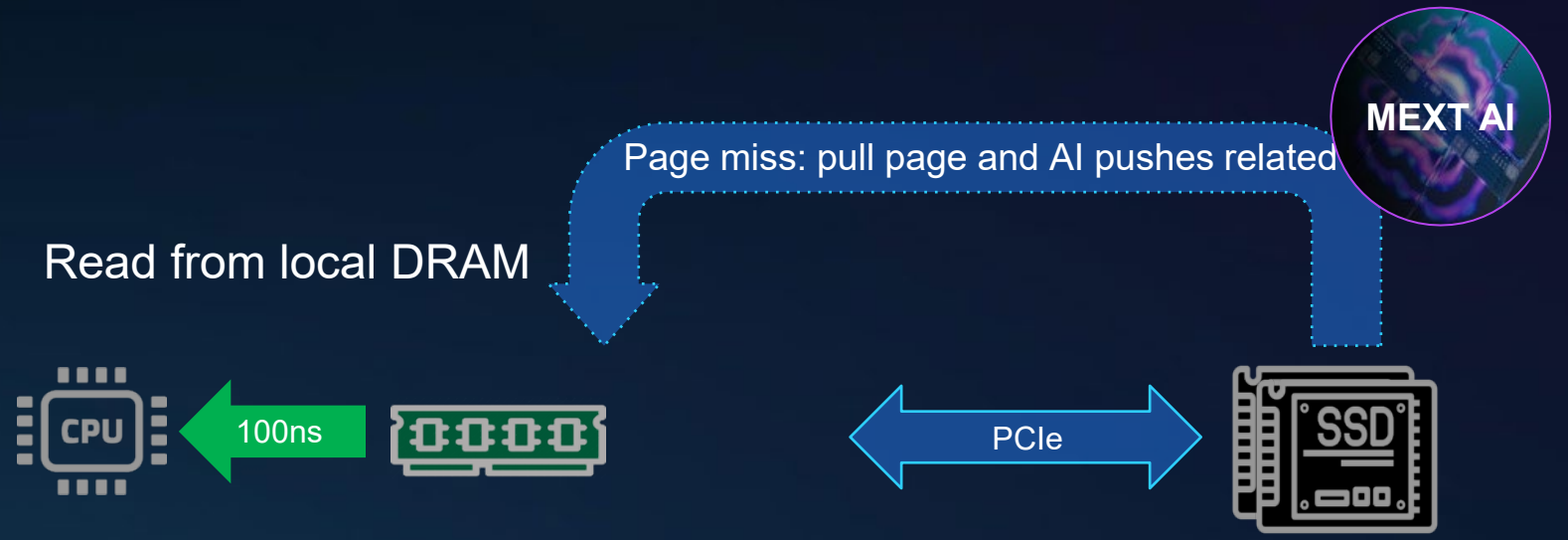
Flash is 95% lower cost-per-bit vs DRAM

“Great Idea, But It Won’t Work…”



AI-Powered Predictive Memory

AI engine predicts and transparently tiers memory

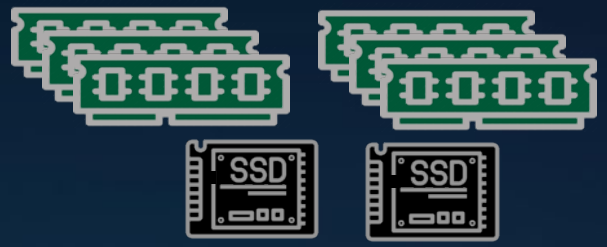


- AI Push Engine predicts what will be needed in advance
- Places memory pages optimally before they are requested
- Goal: all but initial miss happen as quickly as possible

AI-Powered Predictive Memory

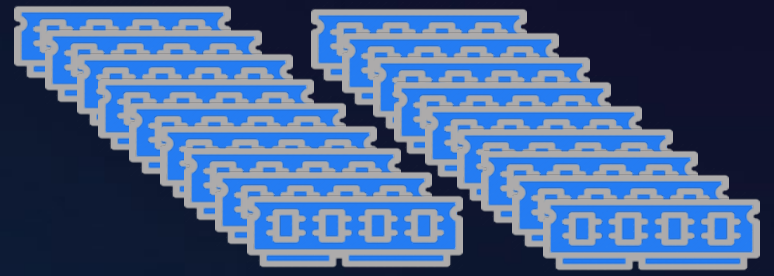
No HW / SW changes, use FLASH as memory

Without MEXT



MEXT

With MEXT



Plug-In

Lightweight, low-overhead Linux Driver

AI Engine

Ensemble of AI models enables transparent auto-tiering of memory pages

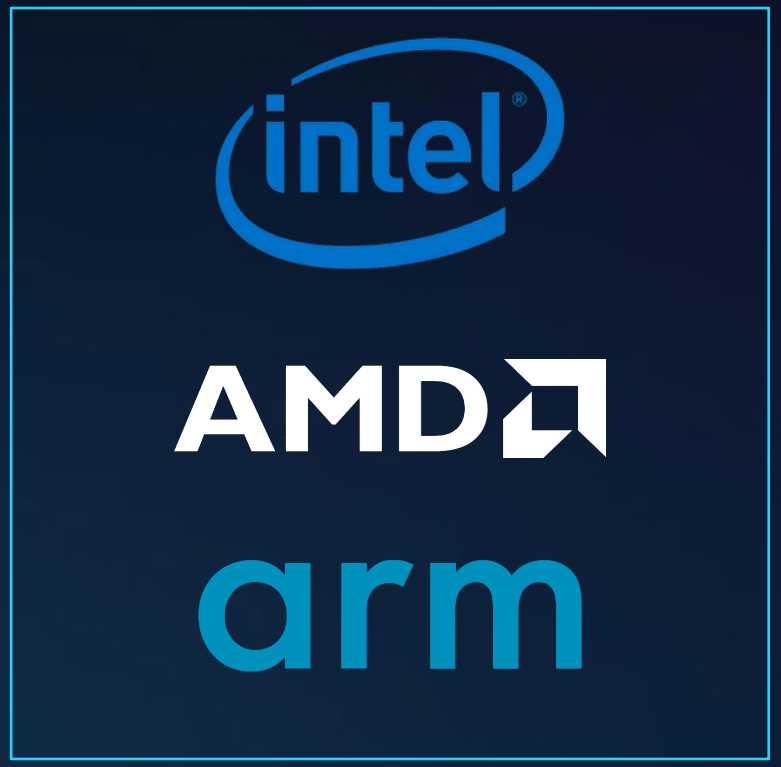
MEXTview™

Illustrates memory utilization and MEXT AI prediction accuracy

MEXT FMS

Software Deploys Anywhere in Minutes

On-Premises or Cloud Environments



Intel, AMD, or ARM Processors

VMs, Containers, or Bare Metal Configurations

MEXT FMS

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

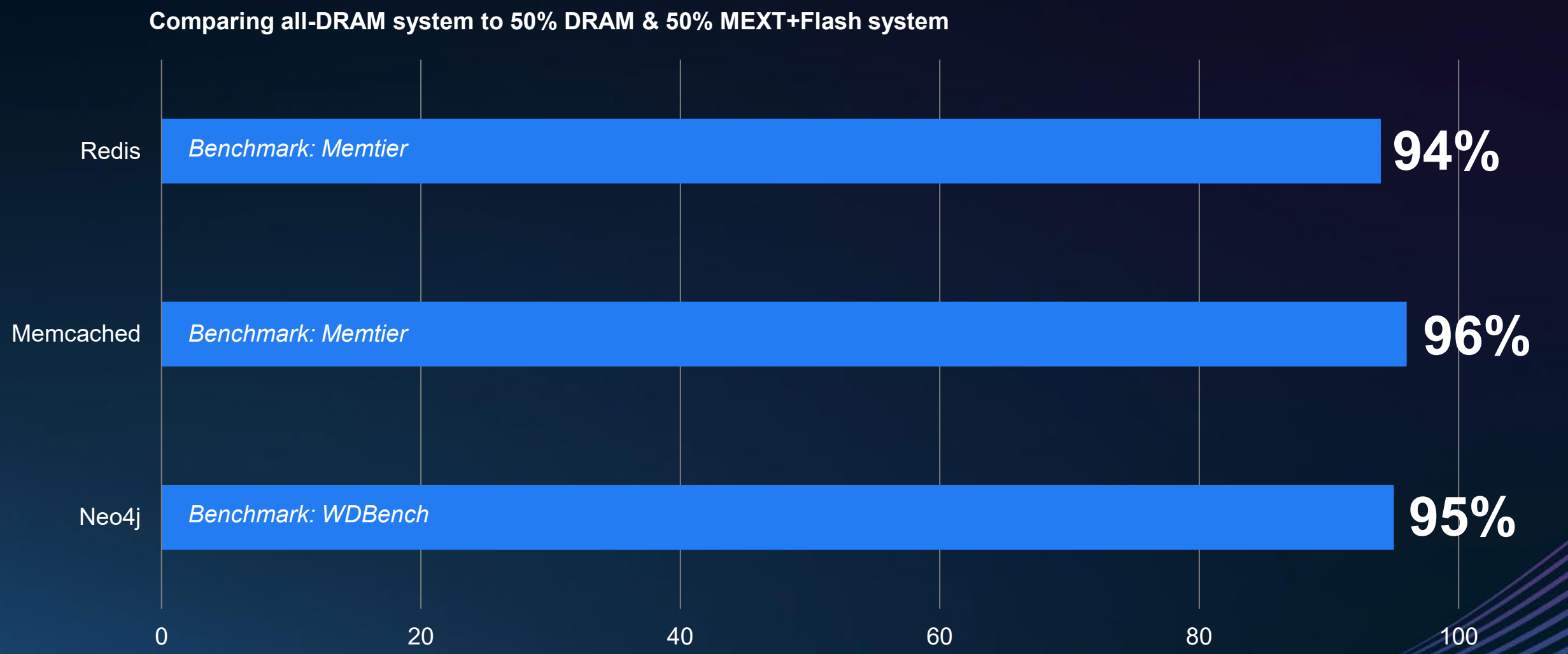
DeepSpeed

Animation Rendering

MOONRAY

MEXT FMS

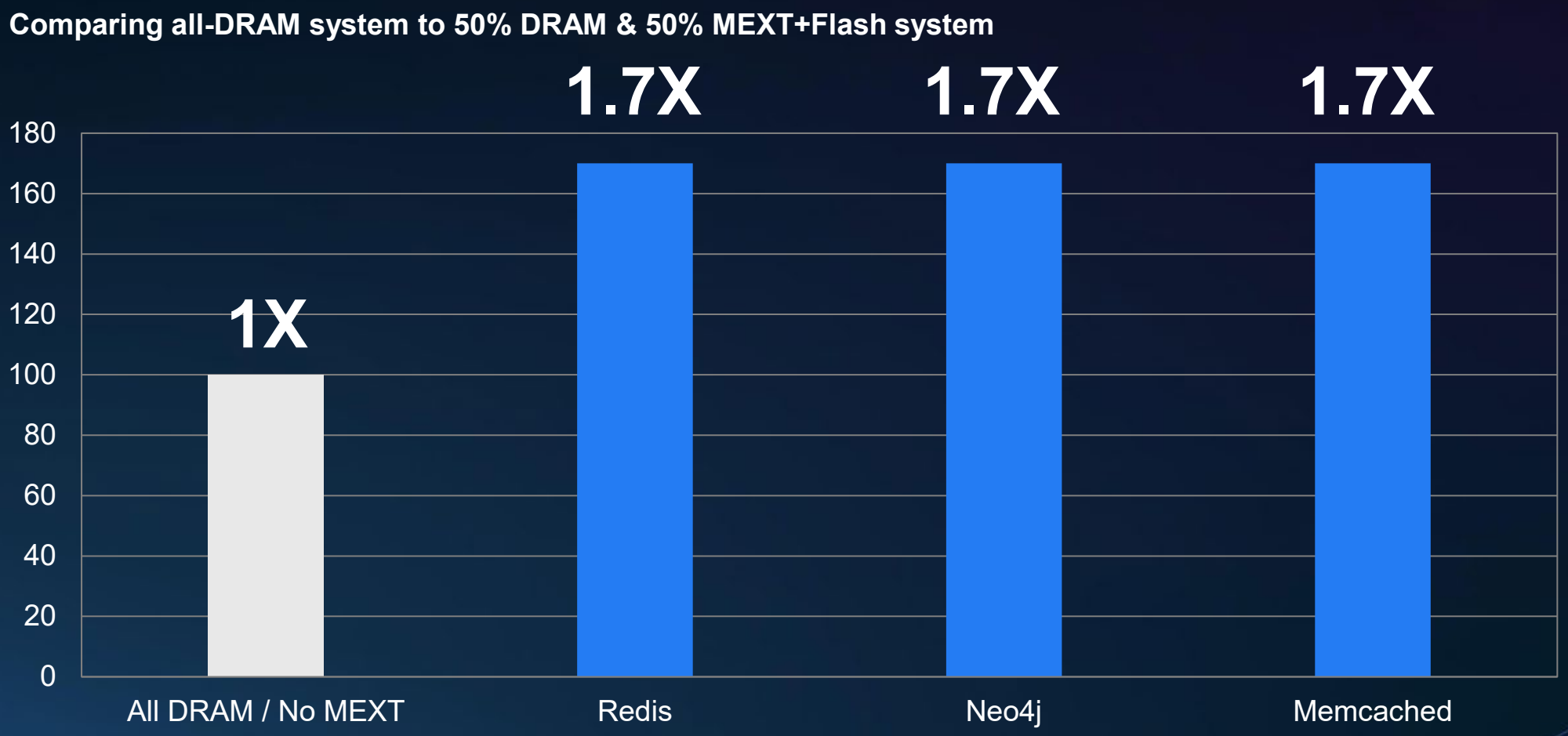
Performance Summary



MEXT FMS

Testing Performed in AWS and on AMD System

Driving Superior Performance / $



MEXT FMS

Better Performance / $ is a Winning Formula

AI-Powered Predictive Memory

Radically lowering the cost of computing