

AMD MEXT

PREDICTIVE MEMORY SOFTWARE

MARKET CONTEXT

Memory is now an industry-wide problem: rising costs, constricted capacity, limited availability. To make matters worse, across many IT environments, significant shares of provisioned DRAM typically remain inactive for meaningful periods of time.

AMD MEXT was built to address 3 imperatives: 1) fix the DRAM under-utilization issue, 2) do so with limited hardware or software changes (be a drop-in solution), and 3) bring flash into the memory domain.

The core innovation – called AMD MEXT Predictive Memory – leverages AI to transparently tier memory pages across DRAM and NVMe flash, effectively enabling flash to serve as an extension of DRAM. This helps customers expand usable memory capacity through the simple addition of MEXT software, or reduce the physical DRAM required for a workload to run (filling in the gap with MEXT-plus-flash functioning as memory).

THE CUSTOMER OUTCOME

MORE CAPACITY

Expand usable capacity for memory-hungry workloads.

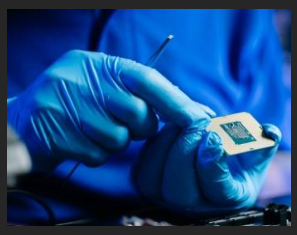
BETTER ECONOMICS

Improve performance-per-DRAM-dollar.

SEAMLESS INTEGRATION

Deploy in less than 5 min with no application or OS changes.

TARGET WORKLOADS



EDA / Chip Design



Agentic AI



Database & Analytics



Media & Entertainment

SOLUTION COMPONENTS

MEXT DRIVER

Low-overhead Linux kernel module that installs in less than 5 min

MEXT AI ENGINE

Ensemble of models including lightweight heuristic predictors up to transformers-based models

RAM VIEW OBSERVABILITY TOOL

Tool that comes with MEXT that illustrates real-time memory utilization / temperatures

Public

HOW IT WORKS

1 – COLD PAGE OFFLOAD	A lightweight Linux driver is installed and starts gathering memory-access telemetry. Cold memory pages are offloaded to NVMe flash.
2 – AI-POWERED PREDICTION	An ensemble of AI models continually predicts which pages that have previously been offloaded to NVMe flash will be needed next by the application.
3 – PUSH BACK TO DRAM	Pages are predictively and transparently moved to DRAM before they are needed, keeping the working set in DRAM and maintaining performance.

PERFORMANCE RESULTS: EDA Static Timing Analysis

For this EDA static timing analysis example, running with a 50% DRAM reduction (768 GB) and 50% MEXT memory (768 GB) provides **~1.74 better performance-per-DRAM-dollar** compared with the baseline system (1.5 TB DRAM and no MEXT memory on the same AMD EPYC™ 9755 CPU).[1]

ALL-DRAM BASELINE	MEXT-ENABLED
1.5TB Memory Capacity	**1.5TB** Effective Memory Capacity
100% DRAM 1.5TB All-DRAM	**50% DRAM, 50% MEXT** 768GB DRAM, 768GB MEXT Memory
1x Performance-Per-DRAM-Dollar 100% Performance / 100% DRAM Cost	**1.74x Performance-Per-DRAM-Dollar** ~87% Performance / ~50% DRAM Cost

RIGHT-SIZE DRAM

Use a smaller physical DRAM footprint while maintaining comparable workload performance.

EXTEND EXISTING SYSTEMS

Add usable memory capacity through software and dedicated NVMe flash.

UNLOCK MORE WORK

Help memory-bound environments improve on throughput and performance-per-dollar.

GETTING STARTED

1) Identify your memory-constrained workload.
2) Reach out to your AMD sales representative to start a no-cost proof-of-concept.
3) Evaluate how MEXT can reduce your DRAM footprint or extend usable memory capacity.

LEARN MORE
amd.com/mext

1: 9xx5-298a: AMD Internal testing as of 10/1/2025 on Synopsys PrimeTime, an EDA signoff workload, on an AMD EPYC 9755 CPU with 1.5TB DRAM only memory as the baseline vs. the same AMD EPYC 9755 CPU using 768GB DRAM and 768GB MEXT memory pooled from a 3.84TB Samsung NVMe drive.

Testing measured performance, measured in seconds (lower is better) in matched NPS1/NPS2 pairs.

System configurations:
AMD EPYC 9755
- AMD Volcano CRB reference board. BIOS RVOT1004E.
- 1P, 128 cores per socket, 24 memory channels
- NVMe for MEXT memory: Samsung MZQL23T8HCLS-00A07, 3.84 TB, PCIe Gen4x4, TLC
- Base/Boost frequencies of 2.7 GHz/4.1 GHz, respectively
- L1/L2/L3 cache of 12 MiB/ 256 MiB/ 1 GiB, respectively
- NPS1/NPS2
- 1.5TB Memory using 12x 128 GB DDR5 DIMMs with 6400 MT/s speeds
- OS version: Red Hat Enterprise Linux 8.10
- Kernel: Linux Kernel 6.17.1
- GCC version: 8.5.0 (RHEL 8.10 system default)
- SMT: off
- All other settings default

Performance per DRAM dollar reflects normalized performance divided by the relative DRAM capacity used and does not account for NVMe, MEXT licensing, other system costs, or total cost of ownership.