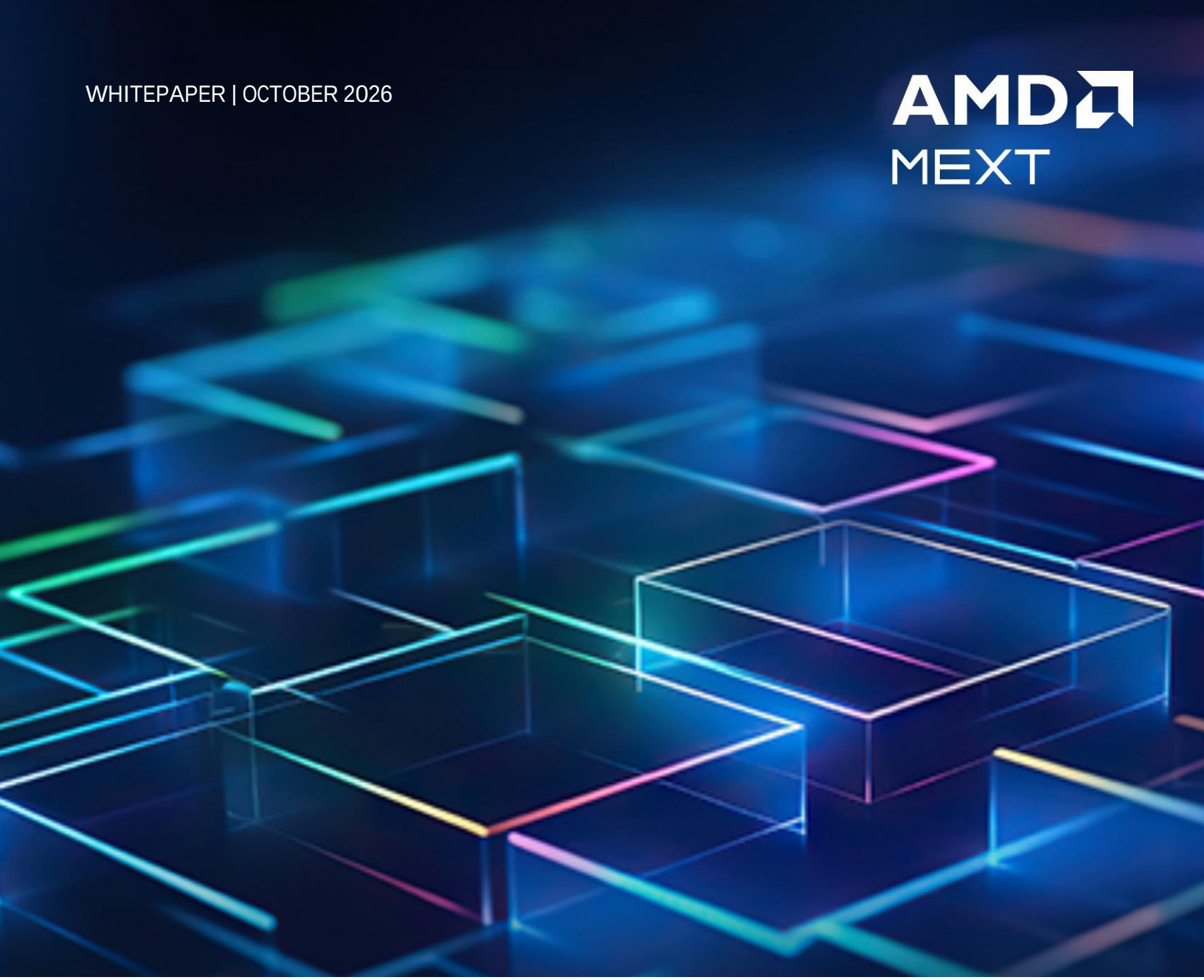
WHITEPAPER | OCTOBER 2026



DRIVING SUPERIOR PERFORMANCE PER-DRAM-DOLLAR WITH 50-75% LESS DRAM

An analysis of AMD MEXT software performance for Synopsys® PrimeTime®



EXECUTIVE SUMMARY

This whitepaper highlights how AMD MEXT Predictive Memory™ software improves the economics of memory-intensive electronic design automation (EDA) signoff workloads. Using AI-powered flash-backed memory expansion, MEXT can reduce the physical DRAM required for Synopsys® PrimeTime® while retaining a high percentage of all-DRAM performance. The result is substantially higher performance relative to the physical DRAM footprint, or higher performance-per-DRAM-dollar.

AMD internal testing on a 1P AMD EPYC™ 9755 platform showed that a configuration with 256GB DRAM plus 256GB of MEXT Memory™ delivered 96.7% of the performance of a 512GB DRAM-only baseline[1]. Reducing DRAM further to 192GB retained 92.5% of baseline performance[2], while a 128GB DRAM configuration retained 83.7%[3]. These results translate to approximately 1.93x[1], 2.47x[2], and 3.35x[3] performance-per-DRAM-dollar, respectively.

50%-75%	83.7%-96.7%	1.93x-3.35x
less physical DRAM	of all-DRAM performance	performance-per-DRAM-dollar

INTRODUCTION

Signoff is one of the most computationally demanding steps in EDA workflows, requiring precise modeling of timing paths across billions of transistors and interconnects. These analyses involve massive, highly interconnected data structures that strain system memory, often pushing DRAM capacity and bandwidth to their limits. As design complexity grows with each process node, the sheer volume of timing data leads to frequent paging, reduced throughput, and underutilized compute resources, creating a significant bottleneck in design turnaround time and overall engineering efficiency. Optimizing memory usage without compromising accuracy has therefore become a critical challenge, directly impacting schedule predictability, tool utilization, and total compute cost.

HARDWARE CONFIGURATION

- **CPU:** AMD EPYC™ 9755 (Zen 5), 128 cores / 128 threads, single-socket configuration with SMT disabled
- **Memory:** 1.5 TB DDR5 installed as 12 × 128 GB DIMMs, with all 12 memory channels populated at 6400 MT/s (but capped at 512GB for the purposes of this testing using the kernel boot argument mem=512G)
- **Storage:** 3.84 TB Samsung PCIe Gen4 NVMe for OS, /tmp, and /var; 1.60 TB Kioxia NVMe dedicated to the MEXT memory pool

SOFTWARE STACK

- **Operating system:** Red Hat Enterprise Linux (RHEL) 8.10 (Ootpa)
- **Kernel:** Linux 6.17.1, locally built with CONFIG_FPROBE=y (6.17.1-fprobe.el8)
- **AMD MEXT:** MEXT Predictive Memory™ software version 1.2
- **EDA application:** Synopsys PrimeTime 2022.03-SP2, configured for up to 128 application threads
- **Compiler & system tuning:** GCC 8.5.0 (RHEL 8.10 system default) with the epyc-eda tuned profile active for all runs

AMD MEXT PREDICTIVE MEMORY SOFTWARE

How It Works

The MEXT software-only solution makes flash storage appear as DRAM-like memory to the OS.

Here's how it works:

- MEXT continuously monitors which memory pages in DRAM are actively being utilized (hot) and which have gone cold
- MEXT offloads the cold memory from DRAM to flash
- MEXT AI engine mitigates the effects of flash latency to keep the system performant
  - AI engine continually predicts which offloaded pages might be requested by the application soon (in other words, which pages are likely to soon go from cold to hot)
  - AI engine transparently moves those pages back into DRAM before the requests are even made
  - As a result, the application stays performant because from its perspective, the relevant memory pages are always already resident in DRAM

THE MARKET: DRAM CHALLENGES

Server memory (DRAM) comes with some major challenges.

First, it is now the most expensive data center component, with cost-per-bit increasing dramatically and showing no signs of slowing.[4]

Second, predicting how much is needed for a particular job is very difficult, and getting it wrong can result in out-of-memory (OOM) errors and job failures; as a result, most teams over-provision to avoid the "memory cliff".

Third, for larger memory systems, the granularity of available memory sizes is quite coarse; even if the exact requirement is known, it's impossible to fine-tune the purchased memory to the requirement plus a small buffer (e.g., 10%).

Lastly, even if the precisely correct memory size is somehow implemented, utilization can often be quite poor.

In short, DRAM is extremely expensive, often over-provisioned, hard to fine-tune, and usually poorly utilized.

MEXT Value

MEXT addresses the major DRAM challenges outlined above.

First, by leveraging low-cost flash as memory, MEXT enables much more cost-effective computing.

Second, MEXT prevents OOM crashes by enabling applications to go out to flash-as-memory if DRAM ever runs out.

Third, MEXT enables more precise right sizing of memory sizes to true requirements.

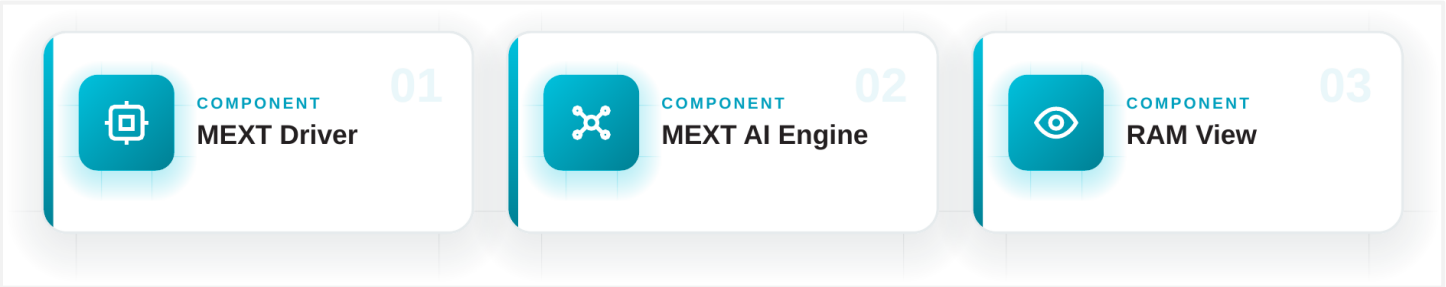
Lastly, MEXT solves the utilization issue by intelligently tiering across DRAM and flash to keep DRAM mostly all hot.

Seamless Implementation

MEXT is a software-only solution that works with any configuration: on-premises or in any cloud, across virtualized / bare-metal / containerized environments, and requiring no changes to the OS or applications. Installation takes less than 5 minutes.

AMD MEXT SOLUTION COMPONENTS

The solution consists of 3 primary components: the MEXT Driver, the MEXT AI Engine, and the RAM View™ Observability Platform.



Driver

The MEXT Driver is a dynamically loadable kernel module (which does not alter the standard Linux kernel) that sends process and memory page telemetry data to the AI engine.

MEXT AI Engine

The AI engine is a user-space process that feeds predictions of which memory pages should be pushed from flash to DRAM – making predictions / inferences in under a fraction of a second. It runs entirely on the local Linux operating system (on a single CPU core) and does not require a GPU.

It was inspired by modern AI techniques based on neural networks. Instead of using these techniques to predict words or natural language patterns, it predicts sequences of future memory page accesses. It consists of a family of models that work together, combining extremely lightweight heuristic predictors with more powerful neural-network models. For any given workload, it automatically adjusts to use the model or group of models that performs best. Continuous observation of which predicted pages were actually used by the application also enables the engine to acquire real-time feedback regarding model accuracy, supporting ongoing adaptation and self-optimization.

RAM View Observability Platform

RAM View is an observability / visualization tool that illustrates how much memory the application is consuming including real-time insights into page hotness / warmness / coldness. It illustrates the proportion of memory pages that were accessed <15s ago, <1min ago, <2min ago, <3min ago, <5min ago, and >5min ago.

METHODOLOGY

Testing used Synopsys PrimeTime 2022.03-SP2 on a single-socket AMD EPYC 9755 system. The 512GB DRAM-only baseline and each MEXT configuration ran the same Large GFX Top-Level static timing analysis workload. Elapsed time was measured in seconds, with lower time indicating higher performance. Each scenario was run three times and the averages were compared. The MEXT scenarios reduced the DRAM presented to the operating system to 256GB, 192GB, or 128GB, then added 256GB, 315GB, or 378GB of MEXT Memory, respectively, to maintain approximately 512GB of effective memory. Note: MEXT reserves 2 of 128 cores, so MEXT runs ran 128 threads on 126 CPUs while baseline had 128 on 128.

RESULTS

Test Configuration	Average Time (Seconds)
512GB DRAM-only baseline	14,040s (3-run average)
256GB DRAM + 256GB MEXT memory	14,516s (3-run average)
192GB DRAM + 315GB MEXT memory	15,182s (3-run average)
128GB DRAM + 378GB MEXT memory	16,783s (3-run average)

Performance

The 256GB DRAM + 256GB MEXT configuration achieved 96.7% of baseline performance[1]. The 192GB DRAM + 315GB MEXT configuration achieved 92.5%[2], and the 128GB DRAM + 378GB MEXT configuration achieved 83.7%[3]. These comparisons use average elapsed time across three runs per scenario; lower elapsed time is better.

Performance-Per-DRAM-Dollar

The tested MEXT configurations reduced physical DRAM capacity by 50%, 62.5%, and 75% relative to the 512GB DRAM-only baseline. MEXT Memory supplied the additional effective capacity through a dedicated KIOXIA 1.60TB NVMe device. Normalizing performance to the remaining DRAM footprint yields approximately 1.93x performance-per-DRAM-dollar at 256GB DRAM, 2.47x at 192GB DRAM, and 3.35x at 128GB DRAM.[1,2,3]

CONCLUSION

For memory-intensive signoff workloads, AMD MEXT software provides a practical way to reduce physical DRAM while preserving a high percentage of application performance. In this study, PrimeTime retained 83.7% to 96.7% of the 512GB all-DRAM baseline[1,2,3] while using 50% to 75% less physical DRAM. The result is a substantial improvement in performance relative to the DRAM footprint, giving semiconductor teams more flexibility to right-size memory resources while maintaining predictable time-to-results.

DISCLAIMER

©2026 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, EPYC, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Synopsys and PrimeTime are trademarks or registered trademarks of Synopsys, Inc. KIOXIA is a trademark of KIOXIA Corporation. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies. Supported features may vary by operating system. Please confirm with the system manufacturer for specific features. No technology or product can be completely secure.

ENDNOTES

1. 9xx5-298: AMD internal testing from 9/18/2026 to 9/22/2026 using Synopsys PrimeTime 2022.03-SP2, an EDA signoff workload, on a 1P AMD EPYC 9755 CPU. The 512GB DRAM-only baseline averaged 14,040 seconds versus 14,516 seconds with 256GB DRAM and a 256GB MEXT memory pool. Performance was calculated as 14,040 / 14,516 = 0.967, or 96.7% of baseline performance. Lower elapsed time is better. Results are averages of three runs per scenario.
   Performance-per-DRAM-dollar was calculated using the tested memory configurations with pricing available on https://www.trendforce.com/price/dram/dram_spot as of October 12th, 2026. Results may vary based on workload, memory capacity and utilization, software versions, operating-system settings, storage configuration, system configuration, and other factors.
   Performance per DRAM dollar claims are not meant to equate to performance per system dollar claims or overall TCO claims. They simply represent tthe performance achieved (normalized to all-DRAM performance) divided by the physical DRAM capacity as a fraction of the baseline.
2. 9xx5-299: AMD internal testing from 9/18/2026 to 9/22/2026 using Synopsys PrimeTime 2022.03-SP2, an EDA signoff workload, on a 1P AMD EPYC 9755 CPU. The 512GB DRAM-only baseline averaged 14,040 seconds versus 15,182 seconds with 192GB DRAM and a 315GB MEXT memory pool. Performance was calculated as 14,040 / 15,182 = 0.925, or 92.5% of baseline performance. Lower elapsed time is better. Results are averages of three runs per scenario.
   Performance-per-DRAM-dollar was calculated using the tested memory configurations with pricing available on https://www.trendforce.com/price/dram/dram_spot as of October 12th, 2026. Results may vary based on workload, memory capacity and utilization, software versions, operating-system settings, storage configuration, system configuration, and other factors.
   Performance per DRAM per dollar claims are not meant to equate to performance per system dollar claims or overall TCO claims. They simply represent the performance achieved (normalized to all-DRAM performance) divided by the physical DRAM capacity as a fraction of the baseline.
3. 9xx5-300: AMD internal testing from 9/18/2026 to 9/22/2026 using Synopsys PrimeTime 2022.03-SP2, an EDA signoff workload, on a 1P AMD EPYC 9755 CPU. The 512GB DRAM-only baseline averaged 14,040 seconds versus 16,783 seconds with 128GB DRAM and a 378GB MEXT memory pool. Performance was calculated as 14,040 / 16,783 = 0.837, or 83.7% of baseline performance. Lower elapsed time is better. Results are averages of three runs per scenario.
   Performance-per-DRAM-dollar was calculated using the tested memory configurations with pricing available on https://www.trendforce.com/price/dram/dram_spot as of October 12th, 2026. Results may vary based on workload, memory capacity and utilization, software versions, operating-system settings, storage configuration, system configuration, and other factors.
   Performance per DRAM dollar claims are not meant to equate to performance per system dollar claims or overall TCO claims. They simply represent the performance achieved (normalized to all-DRAM performance) divided by the physical DRAM capacity as a fraction of the baseline.
4. CounterPoint Research: https://counterpointresearch.com/en/reports/memory-price-tracker-feb-2026