Hewlett Packard Enterprise | AMD

Business white paper

Check if the document is available in the language of your choice.

REALIZE A MORE PRODUCTIVE EDA ENVIRONMENT

Improving the economics of semiconductor design with HPE Apollo 2000 Gen10 Plus Systems and AMD EPYC™ processors

TABLE OF CONTENTS

INTRODUCTION

Few industries are more competitive than modern electronics manufacturing and chip design. Consumers expect devices to be faster, cheaper, and more reliable with each generation. Whether large or small, electronics manufacturers rely on electronic design automation (EDA) to enable these improvements.

High-performance computers are used in all phases of the EDA cycle from system-level design to logic, analog design, simulation, and layout. For even mid-sized projects, verifying proper device functionality is one of the largest challenges faced by chip designers. As engineers make changes to a design, they need to run extensive computer simulations to verify functionality. By most estimates, regression testing and verification account for roughly 80% of simulation workloads in modern electronic design environments.[1] Given the enormous cost of committing a design to silicon, projects must be error-free before tape-out. The performance and capacity of the EDA simulation environment directly affect product quality, time-to-market, downstream support costs, and IT costs—all impacting the bottom line.

EDA firms compete based on the effectiveness of their design environments. In this white paper, we explain how high-performance HPE Apollo 2000 Gen10 Plus Systems powered by 2nd generation AMD EPYC processors can provide electronics manufacturers with a decisive advantage. HPE Apollo 2000 Gen10 Plus Systems can help customers increase simulation capacity, improve throughput and productivity, and reduce TCO in EDA server farms.

CUSTOMER CHALLENGES

Device simulation becomes more difficult as designs become larger. As the number of registers and memory in a device increases (n), the number of states to be modeled increases exponentially (2^n). System-on-a-chip (SoC) designs are frequently in the range of hundreds of millions, or even billions of gates, making verification more challenging with each product generation as designs become more complex.

In addition to size and complexity, reliability and security are also important considerations. Products such as sensors for autonomous vehicles, embedded control systems, and medical devices need to work flawlessly. This demands higher levels of verification coverage and increased simulation to ensure quality and reliability.

Figure 1 illustrates the challenge faced by EDA design centers. Bringing innovative new products to the market and improving reliability requires more simulation capacity. However, firms simultaneously face pressure to shorten design cycles to meet time-to-market objectives with limited budgets for hardware and software.

[1] Based on internal estimates from HPE's internal VLSI design environment, 2019.

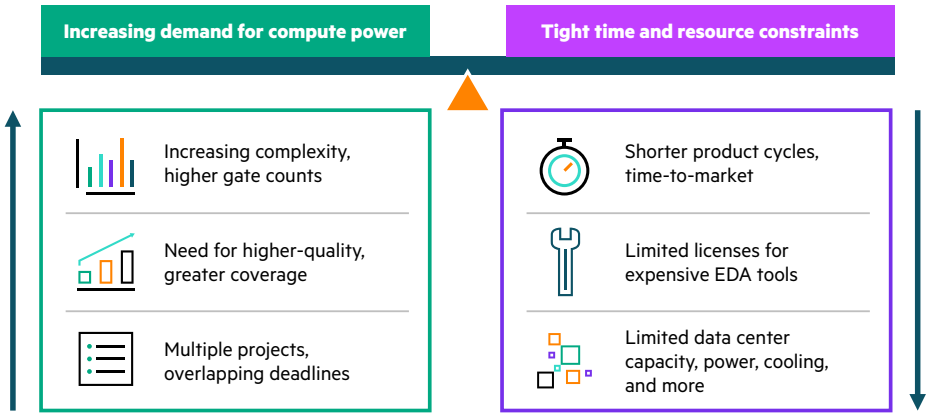


FIGURE 1. EDA firms need more simulation capacity but face tight resource constraints

EDA software tools need to simulate multiple aspects of device functionality over different periods. Chip designers typically run tools from leading EDA vendors, including Cadence®, Synopsys®, and Mentor Graphics®. Workloads are diverse, with some simulations running for minutes or hours, while others run for days, or even weeks on large server farms.

Table 1 describes some typical verification workloads and their characteristics.

TABLE 1. Different types of EDA verification workloads

Category	Verification type	Description
Digital abstractions	Gate level simulations (GLS)	Models may consist of billions of gates. Simulation runtimes can range from hours to weeks, depending on the model and simulator.
	Register transfer level (RTL)	Models typically consist of millions of lines of C-like code executing 10–100K simulated cycles per second (cps). Runtimes range from seconds to multiple days.
	Transaction level model (TLM)	Models consist of up to 1 million lines of C++-like code running 10K–1M simulated cps. Runtimes range from seconds to hours.
Analog abstractions	Transistor level (SPICE)	Models consist of analog primitives: resistors, capacitors, transistors, and others.
	Verilog-AMS/ VHDL-AMS (CAMS)	Models consist of behavioral code and operate on voltage and current values in an analog simulator to solve a network.
	System-level verification	Models consist of C-like code executed on a digital simulator.

Many EDA tools are single-threaded. To optimize throughput and server utilization, customers tend to run many simulations on multicore servers. To achieve high throughput, customers need:

- High clock frequencies
- Large amounts of physical memory
- Large amounts of L3 cache per core
- Low latency and high bandwidth to cache and memory

Semiconductor manufacturers need to deliver ever more complex designs, get to market faster, and continuously improve product quality—all with limited resources.

EDA SOFTWARE LICENSING

A specific challenge faced by electronic manufacturers is the high cost of software tools. Software license costs for EDA environments are typically much higher than hardware costs. Because of this cost disparity, IT administrators tend to be much more concerned about using software resources efficiently than maximizing server utilization. The cost of engineering talent is also an important consideration and organizations need to maximize their productivity. Figure 2 provides a simplified view of a typical design environment.

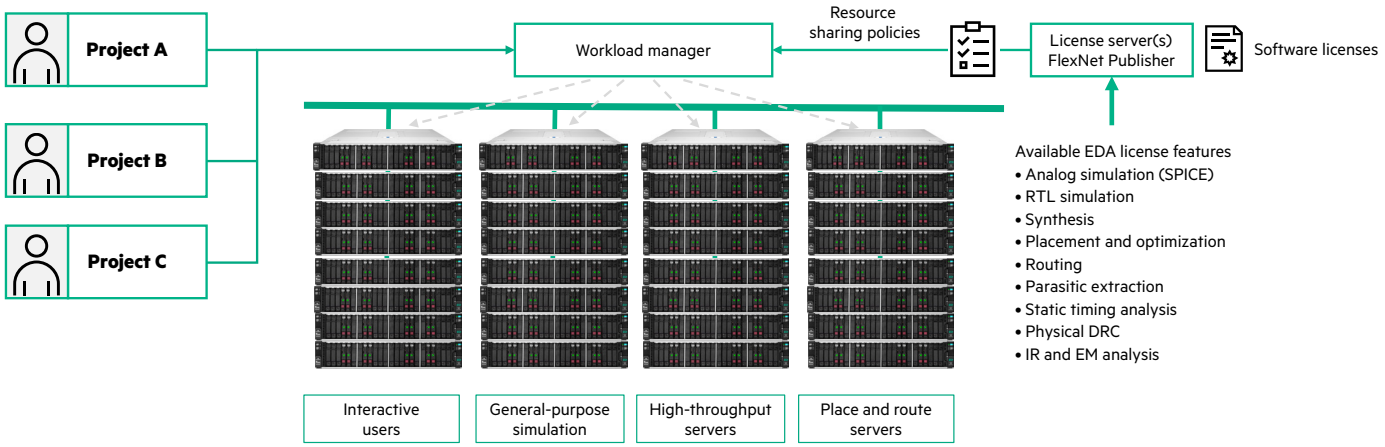


FIGURE 2. A typical EDA environment

Project teams typically work on multiple and sometimes overlapping designs, and need access to EDA software tools and servers to run them on. Flexera® FlexNet Publisher (formerly FLEXlm) is used in EDA environments to meter and manage software licenses. As each tool runs, it contacts a license server and checks out a license. Tools return license features when executions complete. In some cases, a simulation may consume multiple license features.

A single license for a verification tool can cost multiple thousands of dollars per year.[2] Regression tests can involve millions of discrete simulations and completing these quickly requires a large number of licenses. For high-demand tools, a design environment may have hundreds of license features. Overall license costs can easily exceed $1 million annually for a single tool.

Because licenses are expensive, design firms have a strong incentive to keep these licenses fully utilized. Workload management software plays a critical role, coordinating with license servers, and scheduling various batch and interactive jobs. The scheduler seeks to ensure that project deadlines are met and that resources are shared according to policy, optimizing both licenses and infrastructure resources.

Not only is it important to minimize the idle time for licenses, but it is also essential to use the licenses efficiently by running simulations as quickly as possible. A key metric for EDA firms is the number of simulations run per day per license. High-value tools need to execute on server nodes that deliver the highest possible throughput to maximize cost-efficiency.

[2] Price estimate provided by HPE VLSI design environment manager.

AMD EPYC 7002 series processors deliver exceptional performance and scalability for EDA workloads[6]

- World's first 7 nm x86 server CPU
- Highest available core count to maximize parallelism
- World's first PCIe Gen4 capable x86 server CPU
- Eight memory channels per socket
- World's first x86 server processor with DDR4 3200 memory support
- Leadership L3 cache per core

THE AMD ADVANTAGE

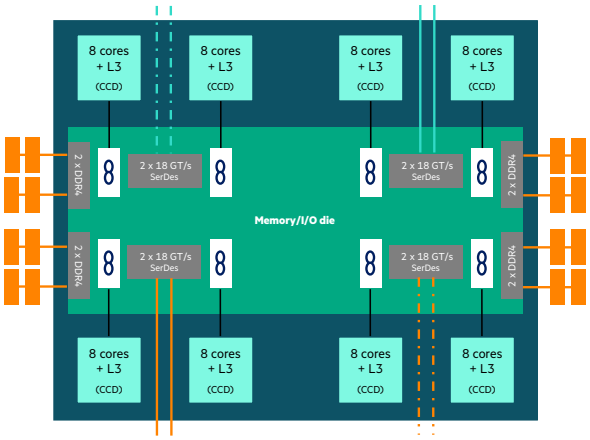
Built on 7 nm technology, the AMD EPYC 7002 series processors bring together high core counts, large memory capacity, extreme memory bandwidth, large cache sizes, and massive I/O with the right ratios to enable exceptional HPC workload performance. For EDA users, this can translate into higher-quality designs, reduced regression runtimes, and better license utilization.

While AMD EPYC processors are the choice of next-generation exascale supercomputers,[3] they are also highly affordable, often delivering superior performance to alternative processors while easily fitting within the budgets of design environments of all sizes.

An ideal architecture for memory-intensive EDA workloads

The unique architecture shown in Figure 3 is the key to the EPYC processor's throughput advantage. The 9-die system-on-a-chip (SoC) features 8 core complex die (CCD); provides up to 8 cores and 32 MB of cache per CCD. This design places large amounts of L3 cache close to compute cores delivering optimal throughput for clock and cache-sensitive RTL and verification workloads. The advanced 7 nm process enables clock frequencies to scale to up to 3.90 GHz, helping minimize license checkout time and enabling users to get more productivity from expensive license features.

While other processors share relatively small amounts of L3 cache across multiple cores, AMD EPYC 7002 series processors offer up to 256 MB of L3 cache. It provides a direct path between each core and associated L3 cache to speed throughput and help reduce latency.[4] This combination of more L3 cache per core, direct channels to cache, more memory channels, and faster memory combines to deliver exceptional throughput.



- Industry leading 7 nm technology*
- Up to 64 cores/128 threads
- Up to 3.90 GHz boost clock speed**
- Up to 256 MB L3 cache
- 8 DDR4 memory channels per socket
- 3200 MT/s DDR
- 128 PCIe Gen4 lanes per socket
- Configurable NUMA operation
- Embedded AMD Secure Processor

* AMD EPYC is the world's first high-performance 7 nm x86 CPU powered by "Zen 2" Processor Core. amd.com/en/press-releases/2018-11-06-amd-takes-high-performance-datacenter-computing-to-the-next-horizon

** Max. boost for AMD EPYC processors is the maximum frequency achievable by any single core on the processor under normal operating conditions for server systems.

FIGURE 3. AMD EPYC 7002 series high-level processor design

For EDA workloads, high-frequency AMD EPYC 7Fx2 processors shown in Table 2 will be of interest. These parts deliver leadership per-core performance while offering large amounts of L3 cache per core.

TABLE 2. AMD EPYC 7Fx2 series high-frequency processors

EPYC 7002 model	Cores/threads	Base speed	Boost speed[5]	L3 cache	Power (Watts)	L3 cache per core
7F72	24/48	3.20 GHz	Up to 3.70 GHz	192 MB	240	8 MB
7F52	16/32	3.50 GHz	Up to 3.90 GHz	256 MB	240	16 MB
7F32	8/16	3.70 GHz	Up to 3.90 GHz	128 MB	180	16 MB

While the performance for EDA applications depends on the tool and design simulated, industry-standard benchmarks illustrate the advantage of AMD EPYC 7002 series processors.

[3] AMD EPYC-based systems have been chosen as the basis of exascale supercomputers. Design wins include Frontier, a collaboration between the U.S. Department of Energy, ORNL, and HPE expected to be delivered in 2021. AMD EPYC processors will also power El Capitan, a collaboration between U.S. DOE, LLNL, and HPE expected in early 2023.

[4] CCX is a term used in AMD CPUs and stands for CPU complex or core complex. It refers to a group of four CPU cores and their CPU caches (L1, L2, and L3). The number of cores per CCX varies by processor. In the case of the 7F32 and 7F52 processors, each CCX contains a single core, providing an independent path to L3 cache for each core to avoid contention and maximize cache per core. In the case of other parts, CCXs contain two or more cores developer.amd.com/wp-content/resources/56827-1-0.pdf.

[5] Max. boost for AMD EPYC processors is the maximum frequency achievable by any single-core on the processor under normal operating conditions for server systems.

[6] amd.com/en/processors/epyc-7002-series

Figure 4 shows relative SPECrate® 2017_fp_base scores per core, comparing EPYC 7Fx2 high-frequency parts to alternative processors with the same core counts on dual-processor systems. EPYC processors' superior performance is a result of high-frequency processors, fast DDR4 memory supporting 3200 MT/s, 33% more memory channels per processor (8 versus 6), and large amounts of L3 cache per core.

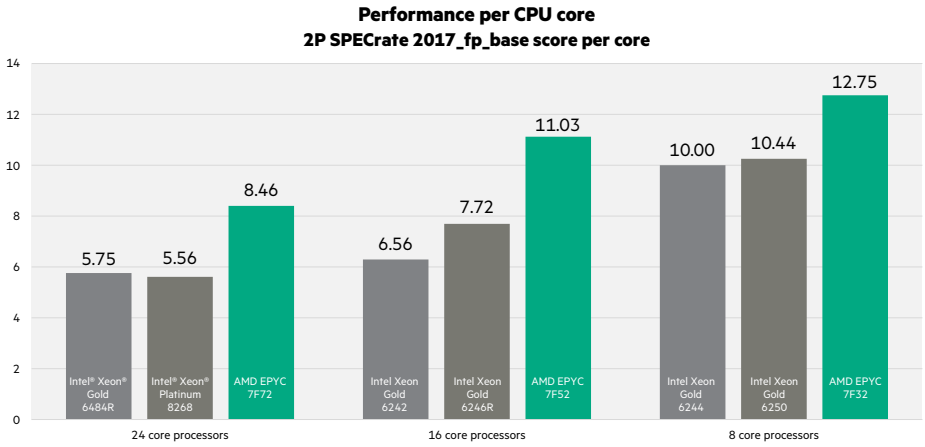
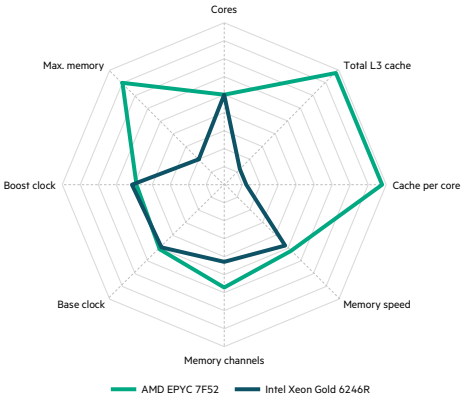


FIGURE 4. EPYC 7Fx2 series high-frequency parts versus best-in-class competitors[7]

Many EDA workloads are particularly sensitive to memory and L3 cache performance. Table 3 illustrates the unique advantages of the EPYC processor, providing better performance across multiple points of comparison.

TABLE 3. AMD EPYC provides advantages across multiple dimensions.

	AMD EPYC 7F52[8]	Intel Xeon Gold 6246R[9]
Number of cores	16	16
Total L3 cache	256 MB	35.75 MB
L3 cache/core	16 MB	2.23 MB
Memory speed	3200 MT/s	2933 MT/s
Memory channels	8	6
Base clock (GHz)	3.50 GHz	3.40 GHz
Boost clock (GHz)[10]	3.90 GHz	4.10 GHz
Max. memory	4 TB	1 TB



[7] SPEC and SPECrate are trademarks of the Standard Performance Evaluation Corporation. All rights reserved. All stated results are as of July 13, 2020, see spec.org for more information. ROM-725 configurations as follows:
Intel Xeon Gold 6248R (48C) scoring 276 SPECrate 2017_fp_base (276/48 = 5.75 score/core) spec.org/cpu2017/results/res2020q2/cpu2017-20200608-22806.html
Intel Xeon Platinum 8268 (48C) scoring 267 SPECrate 2017_fp_base (267/48 = 5.56 score/core) spec.org/cpu2017/results/res2019q2/cpu2017-20190524-14457.html
AMD EPYC 7F72 (48C) scoring 406 SPECrate 2017_fp_base (406/48 = 8.46 score/core) spec.org/cpu2017/results/res2020q2/cpu2017-20200316-21224.html
Intel Xeon Gold 6242 (32C) scoring 210 SPECrate 2017_fp_base (210/32 = 6.56 score/core) spec.org/cpu2017/results/res2019q2/cpu2017-20190430-13303.html
Intel Xeon Gold 6246R (32C) scoring 247 SPECrate 2017_fp_base (247/32 = 7.72 score/core) spec.org/cpu2017/results/res2020q1/cpu2017-20200303-21209.html
AMD EPYC 7F52 (32C) scoring 353 SPECrate 2017_fp_base (353/32 = 11.03 score/core) spec.org/cpu2017/results/res2020q2/cpu2017-20200316-21248.html
Intel Xeon Gold 6244 (16C) scoring 160 SPECrate 2017_fp_base (160/16 = 10.0 score/core) spec.org/cpu2017/results/res2019q2/cpu2017-20190528-14869.html
Intel Xeon Gold 6250 (16C) scoring 167 SPECrate 2017_fp_base (167/16 = 10.44 score per core) spec.org/cpu2017/results/res2020q2/cpu2017-20200608-22682.html
AMD EPYC 7F32 (16C) scoring 204 SPECrate 2017_fp_base (204/16 = 12.75 score per core) spec.org/cpu2017/results/res2020q2/cpu2017-20200316-21244.html

[8] amd.com/en/products/cpu/amd-epyc-7f52

[9] ark.intel.com/content/www/us/en/ark/products/199353/intel-xeon-gold-6246r-processor-35-75m-cache-3-40-ghz.html

[10] Max. boost for AMD EPYC processors is the maximum frequency achievable by any single-core on the processor under normal operating conditions for server systems.

HPE Apollo 2000 Gen10 Plus System with 4x HPE ProLiant XL225n Plus Servers powered by AMD EPYC achieved 10 world records on SPECpower_ssj® 2008, making it the most energy-efficient multinode system server in the world.[16]

HPE APOLLO 2000 GEN10 PLUS SYSTEM

HPE Apollo 2000 Gen10 Plus System is a dense, multiserver platform delivering tremendous performance, throughput, and workload flexibility in a small data center space footprint. Based on industry-leading 2nd generation AMD EPYC 7002 processors, HPE Apollo 2000 Gen10 Plus Systems deliver twice the density of traditional rackmount servers. Each chassis supports up to four dual-processor HPE ProLiant XL225n Gen10 Plus hot-plug servers, each with 2 TB of high performance 3200 MT/s DDR4 memory in just two rack units (2U).[11]

For EDA environments, HPE Apollo 2000 Gen10 Plus Systems provide the ideal blend of features. They offer exceptional simulation performance, expanded power capacity with 3000W power supplies, N+N redundant power, and increased thermal capacity and airflow to reliably support long-running, high-throughput EDA simulations.

HPE Apollo 2000 Gen10 Plus System

- Up to 4 x AMD EPYC servers per 2U chassis
- 2 AMD EPYC 7002 series processors per server
- Up to 64 cores and 128 threads per CPU
- 2 TB memory per server (8 TB in 2U)
- Eight memory channels for superior throughput
- 3200 MT/s DDR4 memory
- Up to 16 MB of L3 cache per core
- Comprehensive management tools (APM/RCM)
- Security anchored in HPE iLO 5 and Silicon Root of Trust
- 3000W power, N+N redundancy
- Enhanced thermal efficiency for HPC workloads

HPE Apollo 2000 Gen10 Plus System front view with multiple storage options

HPE Apollo 2000 Gen10 Plus System real view with up to 4 hot-pluggable dual-processor servers per chassis for maximum density and flexibility

FIGURE 5. HPE Apollo 2000 Gen10 Plus features

With support for the full family of AMD EPYC 7002 series processors, EDA IT administrators can configure systems to precisely meet workload demands. Customers can choose high-frequency EPYC 7F52 processors with fewer cores per processor to help optimize per-core performance or select high-throughput parts such as the EPYC 7742 processor with 64 cores.

Fast I/O is also critical for EDA server farms to help ensure that file and network I/O do not emerge as bottlenecks. HPE Apollo 2000 Gen10 Plus Systems offer PCIe Gen4, providing twice the throughput of the previous generation.[12] HPE offers a variety of high-performance PCIe options, including 200 Gbps Mellanox® and HPE HDR InfiniBand adapters,[13] multiport 100GbE adapters, and high-performance NVMe SSDs. Multiple storage options are available inside the chassis, ranging from 0 to 24 SFF SAS/SATA hard drives.

Comprehensive server security and management

For security-conscious design environments, HPE Apollo 2000 Gen10 Plus Systems provide runtime firmware validation that authenticates critical firmware at start up. Only HPE offers industry-standard servers with firmware anchored into silicon with HPE iLO 5[14] and Silicon Root of Trust. Tied into the Silicon Root of Trust is the AMD Secure Processor, a dedicated security processor embedded in the AMD EPYC SoC.

Customers can also take advantage of optional HPE Apollo Platform Manager (APM), a rack-level power and system management solution for HPE Apollo servers providing an enhanced graphical interface for ease of system management.[15] An optional HPE Apollo 2000 Rack Consolidation Module kit allows iLO aggregation at the chassis level that can be daisy-chained to connect to a top of rack (TOR) management switch.

HPE Performance Cluster Manager (HPCM) is a complete integrated cluster management solution for HPE Apollo systems. HPCM provides system setup, hardware monitoring, management (aggregating system metrics along with remote management from HPE iLO) and cluster health management, image management, software updates, and power management.

[11] Internal benchmarking has shown that for some EDA tools, setting the memory speed to 2933 MT/sec can offer better performance for some tools, since 2933 MT/sec is a multiple of the EPYC Infinity Fabric clock speed of 1467 MHz, March 2020.

[12] PCIe 4.0 delivers 16.0 GT/s, twice the transfer speed of PCIe 3.0 en.wikipedia.org/wiki/PCI_Express

[13] HPE HDR InfiniBand adapters are based on standard Mellanox ConnectX-6 technology.

[14] HPE iLO is a remote server management processor embedded in the system boards on HPE ProLiant servers providing lights-out operation.

[15] h20195.www2.hpe.com/v2/GetDocument.aspx?docname=c04111481

[16] psnow.ext.hpe.com/doc/a50001386enw

HPE Apollo 2000 Gen10 Plus Systems deliver sustained high-performance across multiple cores.

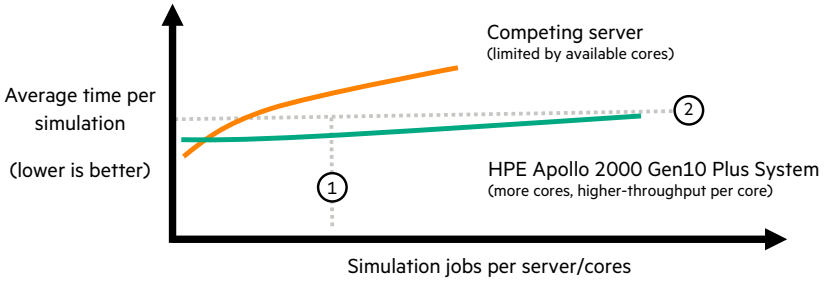
EDA users can reduce regression runtime, help maximize license utilization, and reduce TCO by delivering more simulation capacity with a smaller data center footprint.

PERFORMANCE WHERE IT MATTERS

Figure 6 illustrates the advantage that HPE Apollo System 2000 Gen10 Plus Systems provide for EDA environments. To reduce simulation times and maximize license utilization, semiconductor companies usually look for processors that deliver the maximum per-core performance.

Typically, as additional simulations are run across more cores on the same processor, performance degrades (as shown in the orange line in Figure 6). This degradation occurs for a variety of reasons, including the sharing of L3 cache, limited memory bandwidth, and NUMA effects as cores access memory on remote processors.

Because software licenses are a precious resource, EDA users cannot afford poor per-core simulation performance. They often choose to operate toward the left side of the curve shown in Figure 6, running fewer simulations per server to keep simulation performance high. The problem with this approach is that it results in underutilized servers and a higher data center footprint, increasing TCO.



Two key advantages of AMD EPYC processors:

1. Achieve higher simulation throughput for a fixed number of cores to reduce regression runtimes
2. Run more simulations per server while optimizing license utilization for reduced TCO

Note: This conceptual drawing reflects performance measurements from internal HPE digital simulation benchmarks comparing the AMD EPYC 7F72 running on a HPE ProLiant XL225n Gen10 Plus Server vs. a competing HPE ProLiant Server.

FIGURE 6. HPE Apollo 2000 Gen10 Plus Systems offer high throughput and scalability for EDA workloads

While results vary depending on the EDA tool, internal testing at HPE has shown that HPE Apollo 2000 Gen10 Plus Systems can deliver per-core performance that is equal to or better than other x86 servers.[17] This is true especially when additional simulations are run on the same server (the green curve shown in Figure 6). EDA firms can achieve higher simulation throughput for a fixed number of cores, reduced regression runtime, and improved designer productivity. With improved throughput, firms can also choose to run more simulations per server while still using licenses efficiently, significantly reducing TCO.

With HPE Apollo 2000 Gen10 Plus Servers, EDA users can:

- Reduce regression runtimes to help maximize productivity
- Enable high verification throughput to improve design quality
- Increase EDA software licenses utilization to help minimize cost
- Significantly reduce data center footprint by running more simulations per server

[17] Internal HPE digital simulation benchmarks have shown that AMD EPYC processors deliver superior per-core performance when multiple simulations are run on the same processor. The comparison was made between an AMD EPYC 7F52 processor (3.9 GHz Max. Boost Clock) vs. an Intel Gold 6246R processor (4.1 GHz Max. Turbo Boost Frequency), June 2020.

PURPOSE-BUILT FOR EDA WORKLOADS

Whether large or small, silicon design firms are dealing with multiple challenges, including increasing design complexity, time-to-market pressures, and the high cost of engineering talent and software tools. Electronic devices increasingly require more thorough verification as new applications demand higher levels of reliability and safety.

HPE Apollo 2000 Gen10 Plus Systems powered by AMD EPYC 7002 series processors provide an important new tool and added flexibility for organizations needing to improve the productivity and efficiency of their chip design environments.

By deploying HPE Apollo Gen10 Plus Systems, customers can:

- Accelerate the design process to meet time-to-market pressures
- Improve product quality and meet more stringent reliability requirements with the capacity to run more simulation and verification workloads within available timeframes
- Increase value from limited IT budgets by deploying cost-effective, higher throughput systems that deliver improved server farm utilization, more efficient software license utilization, and better engineering productivity

LEARN MORE AT

hpe.com/servers/apollo2000

To learn more about AMD EPYC 7002 series processors, visit amd.com/en/processors/epyc-7002-series.

Make the right purchase decision. Contact our presales specialists.

Chat **Email** **Call**

Get updates

Hewlett Packard Enterprise

a50002333ENW, August 2020