A detailed look at AMD’s new Epyc “Rome” 7nm server CPUs

reader comments

90 with 61 posters participating, including story author

Share this story

- It feels a little strange to just agree with a headline slide in a product demo, but we can’t find the lie here.
- A dramatic shot of an Epyc Rome processor mounted in a system, *sans* heatsink.
- This half-delidded graphic shows off Rome’s “chiplet” system-on-chip design.

When AMD debuted the 7nm Ryzen 3000 series desktop CPUs, they swept the field. For the first time in decades, AMD was able to meet or beat its rival, Intel, across the product line in all major CPU criteria—single-threaded performance, multi-threaded performance, power/heat efficiency, and price. Once third-party results confirmed AMD’s outstanding benchmarks and retail delivery was a success, the big remaining question was: could the company extend its 7nm success story to mobile and server CPUs?

Yesterday, AMD formally launched its new line of Epyc 7002 “Rome” series CPUs—and it seems to have answered the server half of that question pretty thoroughly. Having learned from the widespread FUD cast at its own internally generated benchmarks at the Ryzen 3000 launch, this time AMD made certain to seed some review sites with evaluation hardware well before the launch.

The short version of the story is, Epyc “Rome” is to the server what Ryzen 3000 was to the desktop—bringing

significantly improved IPC, more cores, and better thermal efficiency than either its current-generation Intel equivalents or its first-generation Epyc predecessors.

Performance

Rome offers far more CPU threads per socket than Intel's Xeon Scalable CPUs do. It also supports a higher DDR4 clockrate and offers 128 PCIe 4.0 lanes, each of which has twice the bandwidth of a PCIe 3.0 lane. This becomes increasingly important in large datacenter environments, which can frequently bottleneck on data ingest as much or more than on raw CPU firepower. Rome also significantly improved upon Epyc's original NUMA design, increasing efficiency and removing potential bottlenecks in multi-socket configuration.

While Rome still can't beat the highest-end Xeon parts for raw hardware clock rate or single-threaded performance, it comes far closer than the first Epyc generation did. This is largely due to a large array of architecture improvements, shown below in AMD's launch-day slides, which cumulatively add up to roughly 15% improvement in instructions executed per hardware clock cycle (IPC).

- The overall story with Rome's improved internal architecture comes down to more instructions executed with each CPU clock cycle.
- Rome offers both more DDR4 channels and higher DDR4 clock rates than its Xeon competitors.
- Rome improves on first-generation Epyc's prediction, fetch and decode with a new L2 branch prediction algorithm, more buffers, and improved associativity.
- Rome can schedule more integer executions, farther ahead, than its first-generation predecessor could.
- Vector and floating point execution scheduling is improved with Zen 2 due to wider data paths and decreased latency.

- Rome offers more cache throughput and larger structures than first-generation Epyc did.
- Epyc's NUMA design improved significantly from first-generation to Rome, increasing efficiency and removing potential bottlenecks in multiple-socket systems.

Ars did not receive hardware review units for this product launch. So, the following performance analysis relies on Rome benchmark data graciously provided by Michael Larabel, of well-known Linux-focused testing, reviews, and news site Phoronix. We'll largely be focusing on dual-socket builds using Rome's 64-core/128-thread Epyc 7742 and 32C/64T Epyc 7502, versus dual-socket builds of Intel's 28C/56T Xeon Platinum 8280, and 20C/40T Xeon Gold 6138.

- PyBench is a single-threaded benchmark, and the higher clock rate of the Xeon CPUs shows to good advantage here. (Data courtesy of Phoronix)
- Despite MKL-DNN being an Intel software package heavily optimized for Xeon CPUs, the Rome CPUs run neck and neck here. (Data courtesy of Phoronix)
- Intel's home-ground software optimization advantage for its MKL-DNN library shows heavily in this deconvolution batch test. (Data courtesy of Phoronix)

On single-threaded benchmarks such as PHPBench and PyBench, it's easy to see both AMD's promised 15% increase in IPC realized and the narrowed gap between their single-threaded performance and Intel's. Although Epyc Rome still loses out to Xeon Scalable here, the performance delta has shrunk from roughly 50% to 20%. Xeon Scalable also comes out on top in the MKL-DNN video encoding tests—which shouldn't be a surprise, since MKL-DNN is a software package written by Intel developers, utilizing their Math Kernel Library for Deep Neural Networks.

While it's easy to complain that Intel CPUs have an unfair advantage in MKL-DNN benchmarks, it is representative of the

kind of entrenched advantage Intel enjoys—and it's a real advantage. Someone with a heavily MKL-DNN focused workload is unlikely to care about what is or isn't fair.

Enlarge / Multithreading-friendly and vendor-neutral tests—such as x265 video encoding, or this OpenSSL library benchmark—heavily favored the massively multithreaded Rome CPUs. (Data courtesy of Phoronix)

On vendor-neutral and multithreading-friendly workloads such as x265 video and OpenSSL, the Rome CPUs significantly outperformed the Xeons across the board. Datacenters are notoriously conservative in design, and more resistant to vendor-shopping than small business or end users—but it's harder to ignore AMD's increasingly large multi-threaded performance wins, when Intel's single-threaded performance gap has been cut in half.

Listing image by AMD

https://arstechnica.com/?p=1548393