Arm Neoverse V1 Platform: A revolution in high performance computing

arm

AT A GLANCE

To meet the rising demands of contemporary applications, semiconductor designers, OEMs and service providers are increasingly turning to Arm Neoverse solutions to provide the foundation for their next-generation infrastructure.

Arm Neoverse delivers leading performance and scalability to computing infrastructure while dramatically reducing power consumption and total cost of ownership.

WHY NEOVERSE V1 PLATFORM?

- Leading performance per core for HPC, HPC-in-the-Cloud and AI/ML-assisted workloads.
- +50% IPC uplift over Neoverse N1.
- SVE and Bfloat16 vector performance uplifts with the added software benefit of write-once, use-forever.
- Flexible CMN-700 mesh interconnect fabric for attaching high-bandwidth DDR5/HBM3 memory systems or custom accelerators.
- Versatile platform with full design control for SoC architects to differentiate products.

Leading Per-Core Performance with Scalable Vector Extension

Achieving exascale performance requires powerful CPUs, high-bandwidth memory, fast I/O systems, and often customized on-die or off-chip accelerators. These capabilities must be supported within extremely tight thermal, power, and cost constraints to deliver additional value to the innovations in the cloud.

Arm Neoverse is a diversified portfolio of processor and system IP designed to meet these challenges. CPUs based on Neoverse processor IP can be optimized for server, HPC and AI/ML-assisted workloads, delivering cutting-edge performance to meet the specific needs of the end-user. Neoverse also provides implementation flexibility for designing processors capable of delivering cloud, HPC and AI/ML services with best-in-class power efficiency and total cost of ownership (TCO).

As the first V-series processor, Neoverse V1 delivers a 50 percent integer performance uplift over Neoverse N1. It also includes the first Arm implementation of Scalable Vector Extension (SVE) allowing developers to write vector-length agnostic code. Neoverse V1 doubles the vector execution capabilities to 512 bits/core, delivering nearly 2X speedup in applications deploying vectors. Moreover, benchmarking the auto-vectorizing capabilities of SVE using a specialized test suite for vectorizing compilers shows that Neoverse V1 with SVE can deliver up to 4X speedup over Neoverse N1 deploying NEON vector instructions. For HPC centers, the result is unmatched performance, improved scalability, and reduced power consumption with a much-improved TCO compared to traditional deployments. Neoverse V1 is seeing excellent traction in HPC markets with various international initiatives launching efforts to develop their own SoCs for HPC deployments. This includes:
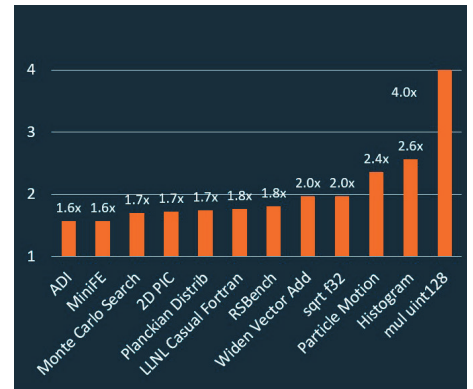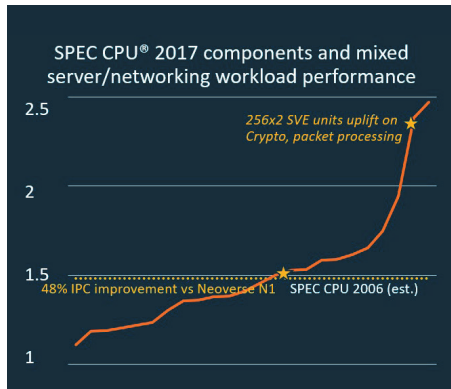
- SiPearl–a high-performance, energy efficient processor being developed through the European Processor Initiative, and based on Neoverse V1

- Neoverse V1 is also being designed in K-AB21, a supercomputer underway in the Republic of Korea that aims for a 2.5x performance uplift over contemporaries while lowering power by 60%

Neoverse V1 Speedup over N1 - A Generation Leap

Neoverse V1 shows excellent performance speedup over Neoverse N1 on a multitude of server and networking workloads. However, even more impressive is vector and AI/ML workload performance where the improvement is 2x to 4x over Neoverse N1.

Neoverse V1 IPC performance uplift over Neoverse N1

SVE vector workload performance: V1 speedup over N1

SPEC CPU® 2017 components and mixed server/networking workload performance

256x2 SVE units uplift on Crypto, packet processing

48% IPC improvement vs Neoverse N1   SPEC CPU 2006 (est.)

ADI 1.6x, MiniFE 1.6x, Monte Carlo Search 1.7x, 2D PIC 1.7x, Planckian Distrib 1.7x, LLNL Casual Fortran 1.8x, RSBench 1.8x, Widen Vector Add 2.0x, sqrt f32 2.0x, Particle Motion 2.4x, Histogram 2.6x, mul uint128 4.0x

Source: Arm

An Architecture for the Most Demanding Workloads

Release: 2020
Manufacturing node: 7nm/5nm
CCIX: 1.1
PCI3 gen 5
DDR5
HBM2e
SVE 2 x256b
Bfloat16
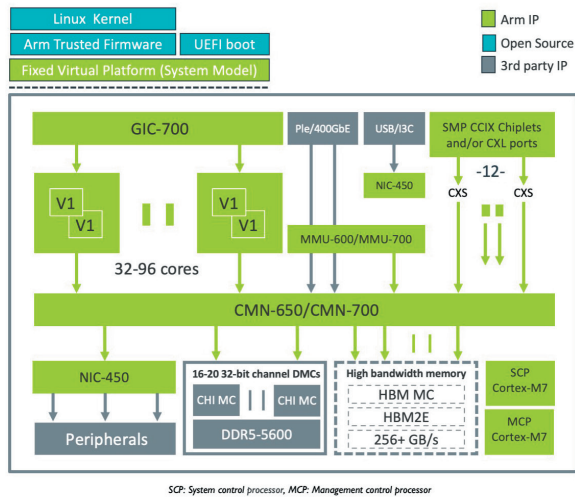
SCP: System control processor, MCP: Management control processor

Reference Design using Neoverse V1

CMN-700

CMN (Coherent Mesh Network) 700 is the high-speed interconnect connecting all CPU, caches, memory, IO and accelerator elements. To support the extreme bandwidth requirements of HBM memory, CMN-700 has been architected to support 3 TB/s, which is 2.5X the cross-sectional mesh bandwidth of previous generation mesh CMN-600. CMN-700 supports both the CCIX and CXL protocols, for chiplet, multichip and multisocket support. CMN-700 is architected to support PCIe Gen5 transfer rates along with a low-latency path to DDR5 and HBM3/2 memories.
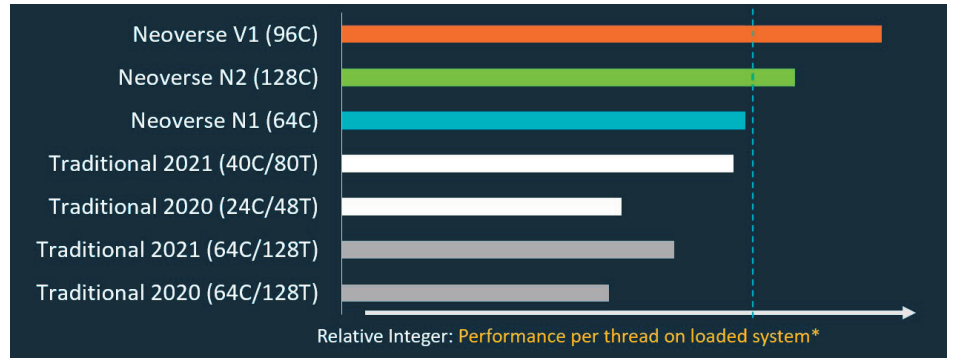
Bfloat16

Neoverse V1 has native support for Bfloat16, which along with the wider SVE-enabled vector data path, helps accelerate the ML training and inference without having to translate data into other formats.

Scalable Vector Scalable Vector Extension

Vector processing is one of the most powerful tools in computing, but traditional processors require developers to code for vectors of specific width. SVE allows developers to deploy and run code with vector length ranging from 128 bits to 2048 bits on Neoverse V1 processors without recompiling, a huge boost in both flexibility and code longevity for software developers.

Single Thread per Core = Superior Performance and Economics

Neoverse V1, like other Arm processors, is built upon the fundamental concept that performance can be maximized within efficient power budgets using densely packed, single threaded, highly efficient CPU cores. Since all execution units are fully available to the software being executed on a single-threaded CPU core, the performance in a fully-loaded system is much higher and predictable compared to solutions that deploy multiple threads in a single core. Moreover single-threaded systems have much less vulnerability to security threats such as side-channel attacks, making them a preferred choice for cloud providers. Neoverse V1 processors deliver the highest absolute single-thread and performance in a fully loaded system compared to all the processors in the Neoverse family.

Neoverse V1 (96C)
Neoverse N2 (128C)
Neoverse N1 (64C)
Traditional 2021 (40C/80T)
Traditional 2020 (24C/48T)
Traditional 2021 (64C/128T)
Traditional 2020 (64C/128T)

Relative Integer: Performance per thread on loaded system*

Diversity in Design

While Arm specifies the architecture for its processor IP, our semiconductor partners determine the ultimate design and performance characteristics of Neoverse processors. Neoverse V1 processors therefore range in core count, speed, cache size, I/O and other features. This approach accelerates innovation while giving developers and end-users greater choice.

For more information, please visit www.arm.com/products/silicon-ip-cpu/neoverse/neoverse-v1.

* Traditional current data is measured by Arm. Traditional next data is projected by Arm.
  Arm Neoverse performance data is estimated by Arm pre-silicon, on Arm Neoverse V1 and N2 Reference Designs:
  Neoverse V1: 96 cores, 2.6GHz, 10xDDR5-4800
  Neoverse N2: 128 cores, 3.0GHz, 8xDDR5-4800
  Neoverse N1: 64 cores, 2.5GHz, 8xDDR4-3200