AMD EPYC™ 8004 Series Architecture Overview

Date	Version	Changes
Jul, 2023	0.1	Initial NDA partner release
Sep, 2023	1.0	Initial public release

Audience

This guide provides a high-level technical overview of 4th Gen AMD EPYC™ 8004 Series Processor internal IP.

Author

Santosh Shukla and Chris Karamatas (with support from the AMD FAE team and Anthony Hernandez)

Table of Contents

This page intentionally left blank.

Chapter 1

AMD EPYC™ 8004 Series Processors

AMD EPYC™ 8004 Series Processors are part of the fourth generation of AMD EPYC server-class processors. 4th Gen AMD EPYC processors feature AMD's latest "Zen 4" based compute cores with corresponding next-generation Infinity Fabric memory & I/O technology. AMD EPYC 8004 processors with new/smaller SP6 socket/packaging and lower TDP are ideal for telco and edge environments.

1.1   AMD EPYC 8004 Specifications

4th Gen AMD EPYC 8004 Series Processors offer a variety of configurations with varying numbers of cores, Thermal Design Points (TDPs), frequencies, cache sizes, etc. that complement AMD's existing server portfolio with further improvements to performance, power efficiency, and value. Table 1-1 lists some distinguishing features common to all AMD EPYC 8004 Series Processors.

AMD EPYC 8004 Series Processor Specifications	
Code name	"Siena"
Socket	SP6
Compute cores	Zen 4-based
Core process technology	5nm
Max number of cores (threads)	64 (128)
Max cores per Core Complex (CCX)	8
Max number of Core Complex Dies (CCDs)	4
Number of CCXs per CCD	2
Max L3 cache size (per CCX)	128 MB (16 MB)
Max # of memory channels	6 DDR5
Max memory speed	4800 MT/s DDR5
Max memory per socket	3 TB
Max # of high-speed I/O lanes	96
Max lanes Peripheral Component Interconnect	96 lanes PCIe® Gen 5
Max lanes Compute eXpress Links	48 lanes CXL 1.1+
Max Processor Frequency	3.15 GHz

Table 1-1: AMD EPYC 8004 Series Processor features

1.2   Operating Systems

AMD recommends using the latest available targeted OS version and updates. Please see AMD EPYC™ Processors Minimum Operating System (OS) Versions for detailed OS version information.

1.3   Processor Layout

4th Gen AMD EPYC processors incorporate compute cores, memory controllers, I/O controllers, RAS (Reliability, Availability, and Serviceability), and security features into an integrated System on a Chip (SoC). The AMD EPYC 8004 Series Processor retains the proven Multi-Chip Module (MCM) Chiplet architecture of prior successful AMD EPYC processors while making further improvements to the SoC components.

The SoC includes the Core Complex Dies (CCDs), which contain Core Complexes (CCXs), which contain the "Zen 4"-based cores. The CCDs surround the central high-speed I/O Die (and interconnect via the Infinity Fabric). The following sections describe each of these components.

1.4   "Zen 4" Core

AMD EPYC 8004 Series Processors are based on the new "Zen 4" compute core. The "Zen 4" core is manufactured using a 5nm process and is designed to provide an Instructions per Cycle (IPC) uplift and frequency improvements over prior generation "Zen" cores. Each core has a larger L2 cache and improved cache effectiveness over the prior generation. Each "Zen 4" core includes:

- Up to 32 KB of 8-way L1 I-cache and 32 KB of 8-way of L1 D-cache

- Up to a 1MB private unified (Instruction/Data) L2 cache. All caches use a 64B cache line size.

Each core supports Simultaneous Multithreading (SMT), which allows 2 separate hardware threads to run independently, sharing the corresponding core's L2 cache.

1.5   Core Complex (CCX)

Figure 1-1 shows a Core Complex (CCX) where up to eight "Zen 4"-based cores share a 16 MB L3 or Last Level Cache (LLC). Enabling Simultaneous Multithreading (SMT) allows a single CCX to support up to 16 concurrent hardware threads.

Figure 1-1: Top view of 8 compute cores sharing an L3 cache (8004 models models)

1.6 Core Complex Dies (CCDs)

Each Core Complex Die (CCD) in an AMD EPYC 8004 Series Processor contains two CCXs, as shown in Figure 1-3.

Zen4 Core	L2 Cache	Shared 16MB L3 Cache	L2 Cache	Zen4 Core
Zen4 Core	L2 Cache		L2 Cache	Zen4 Core
Zen4 Core	L2 Cache		L2 Cache	Zen4 Core
Zen4 Core	L2 Cache		L2 Cache	Zen4 Core
Zen4 Core	L2 Cache	Shared 16MB L3 Cache	L2 Cache	Zen4 Core
Zen4 Core	L2 Cache		L2 Cache	Zen4 Core
Zen4 Core	L2 Cache		L2 Cache	Zen4 Core
Zen4 Core	L2 Cache		L2 Cache	Zen4 Core

Figure 1-2: 2 CCXs in a single 4th Gen AMD EPYC 8004 CCD

1.7 I/O Die (Infinity Fabric™)

The CCDs connect to memory, I/O, and each other through an updated I/O Die (IOD). This central AMD Infinity Fabric™ provides the data path and control support to interconnect CCDs, memory, and I/O. Each CCD connects to the IOD via a dedicated high-speed Global Memory Interconnect (GMI) link. The IOD helps maintain cache coherency. AMD EPYC 8004 Series Processors support up to 4 xGMI (or G-links) with speeds up to 32Gbps. The IOD exposes DDR5 memory channels, PCIe® Gen5, CXL 1.1+, and Infinity Fabric links.

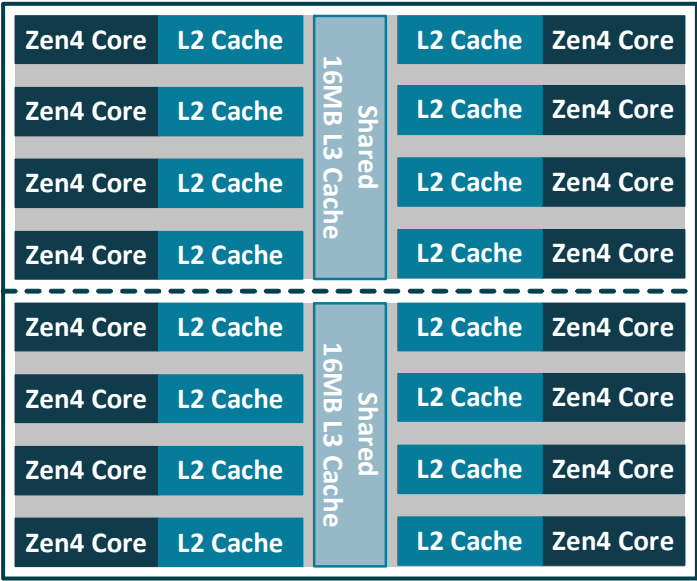All dies (chiplets) interconnect with each other via AMD Infinity Fabric technology. Figure 1-4 (which corresponds to Figure 1-2, above) shows the layout of a 64-core AMD EPYC 8004 processor. AMD EPYC 8004 processors have a maximum of 4 CCDs, with each CCD connecting to the IOD via its own GMI connection.
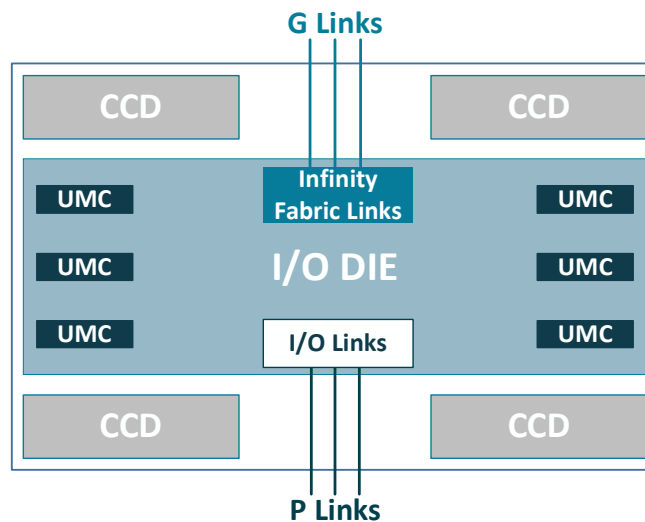

Figure 1-3: AMD EPYC 8004 processor internals interconnect via AMD Infinity Fabric (12 CCD processor shown)

The IOD provides six unified memory controllers that support DDR5 memory. The IOD presents links that the system OEM/designer can configure to support various IO interfaces.

1.8 Memory and I/O

Each UMC can support up to 2 DIMMs per channel (DPC) for a maximum of 12 DIMMs. OEM server configurations may allow either 1 DIMM per channel or 2 DIMMs per channel. AMD EPYC 8004 Series Processors can support up to 3TB of DDR5 memory. Memory interleaving on 2, 3, and 6 channels helps optimize for a variety of workloads and memory configurations.

Each processor has a set of 3 P-links and 3 G-links. AMD EPYC 8004 provides IO flexibility for OEMs to optimize their IO subsystems appropriately. There are 3 G-links and 3 P-links, for a total of 6 x16, or 96, high-speed I/O lanes. Additionally, there are also 8 PCIe Gen3 bonus lanes available.

The 3 G-links provide 48 lanes of PCI Gen5 support. The 3 P-links can be used for (48 lanes) CXL 1.1+ support or (48 lanes) PCI Gen5 support. Alternatively, you can use 32 of the lanes for SATA.

In summary, these links can support:

- Up to 6 x16 bit or 96 lanes of PCIe Gen 5 connectivity to peripherals.

- Up to 48 lanes (3 P-links) that can be dedicated to Compute Express Link (CXL) 1.1+ connectivity to extended memory.

- Up to 32 I/O lanes that can be configured as SATA disk controllers.

1.9 Visualizing AMD EPYC 8004 Series Processors (Family 19h Models A0h–AFh)

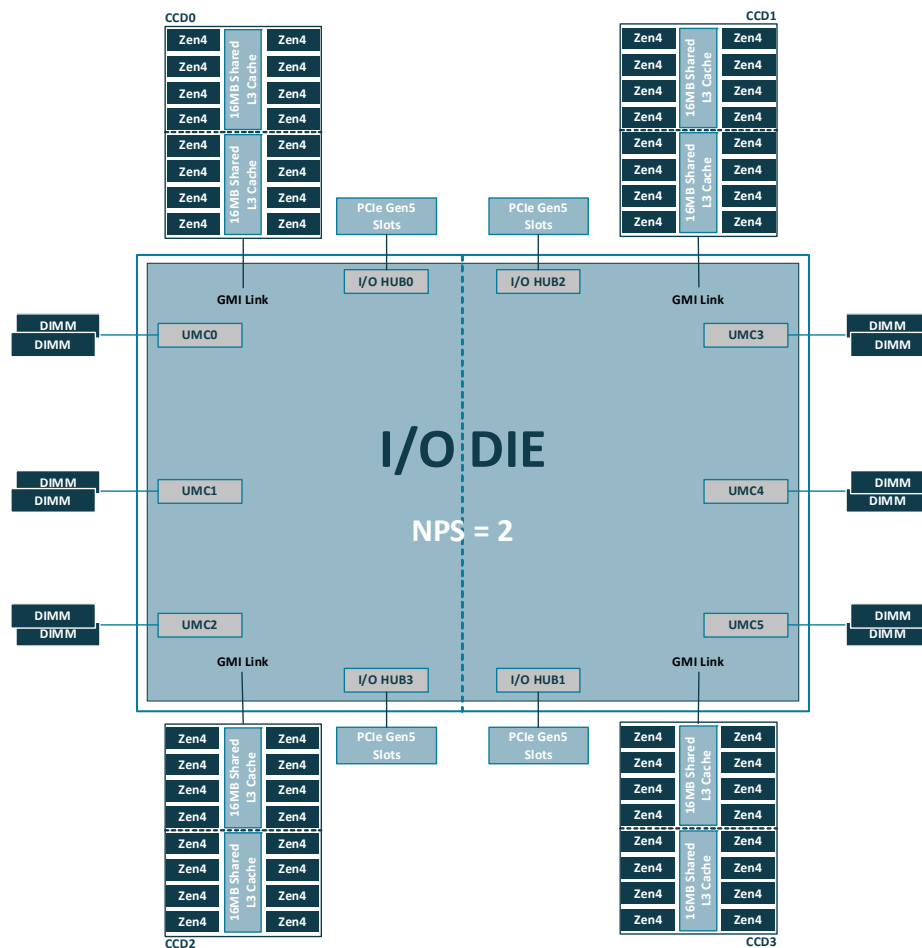AMD EPYC 8004 Series Processors have up to 4 CCDs that each contain two CCXs, as shown below.

Figure 1-4: The AMD EPYC 8004 System on Chip (SoC) consists of up to 4 CCDs and a central IOD

Note: For workloads requiring increased bandwidth, AMD provides Siena OPNs (e.g. AMD EPYC 8324) where each CCD connects to the I/O Die via two GMI3 interfaces.

1.10 NUMA Topology

AMD EPYC 8004 Series Processors use a Non-Uniform Memory Access (NUMA) architecture where different latencies may exist depending on the proximity of a processor core to memory and I/O controllers. Using resources within the same NUMA node provides uniform good performance, while using resources in differing nodes increases latencies.

1.10.1 NUMA Settings

A user can adjust the system **NUMA Nodes Per Socket** (NPS) BIOS setting to optimize this NUMA topology for their specific operating environment and workload. For example, setting NPS=2 as shown in "Memory and I/O" on page 4 divides the AMD EPYC processor into halves, where each half has 2 CCDs, 3 UMCs, and 2 I/O Hubs. The closest processor-memory I/O distance is between the cores, memory, and I/O peripherals within the same quadrant. The furthest distance is between a core and memory controller or IO hub across different halves. The locality of cores, memory, and IO hub/devices in a NUMA-based system is an important factor when tuning for performance.

The NPS setting also controls the interleave pattern of the memory channels within the NUMA Node. Each memory channel within a given NUMA node is interleaved. The number of channels interleaved decreases as the NPS setting gets more granular. For example:

- A setting of NPS=2 configures the processor into two NUMA domains that groups half of the cores and half of the memory channels into one NUMA domain, and the remaining cores and memory channels into a second NUMA domain. Memory is interleaved across the three memory channels in each NUMA domain. PCIe devices will be local to one of the two NUMA nodes depending on the half that has the PCIe root complex for that device. AMD suggests NPS1 or NPS2 for most workloads.

- A setting of NPS=1 indicates a single NUMA node. This setting configures all memory channels on the processor into a single NUMA node. All processor cores, all attached memory, and all PCIe devices connected to the SoC are in that one NUMA node. Memory is interleaved across all memory channels on the processor into a single address space.

You may also be able to further improve the performance of certain environments by using the **LLC (L3 Cache) as NUMA** BIOS setting to associate workloads to compute cores that all share a single LLC. Enabling this setting equates each shared L3 or CCX to a separate NUMA node, as a unique L3 cache per CCD. A single AMD EPYC 8004 Series Processor with 4 CCDs can have up to 8 NUMA nodes when this setting is enabled.

Thus, a single EPYC 8004 Series Processor may support a variety of NUMA configurations ranging from one to four NUMA nodes, depending on how many UMCs are populated. NPS4 may be an option on some EPYC 8004 systems/processors; however, the asymmetry of UMCs means that NPS1 or NPS2 will usually provide better results.

If software needs to understand NUMA topology or core enumeration, it is imperative to use documented Operating System (OS) APIs, well-defined interfaces, and commands. Do not rely on past assumptions about settings such as APICID or CCX ordering.

Chapter 2

Processor Identification

Figure 2-1 shows the processor naming convention for AMD EPYC 8004 Series Processors and how to use this convention to identify particular processors models:
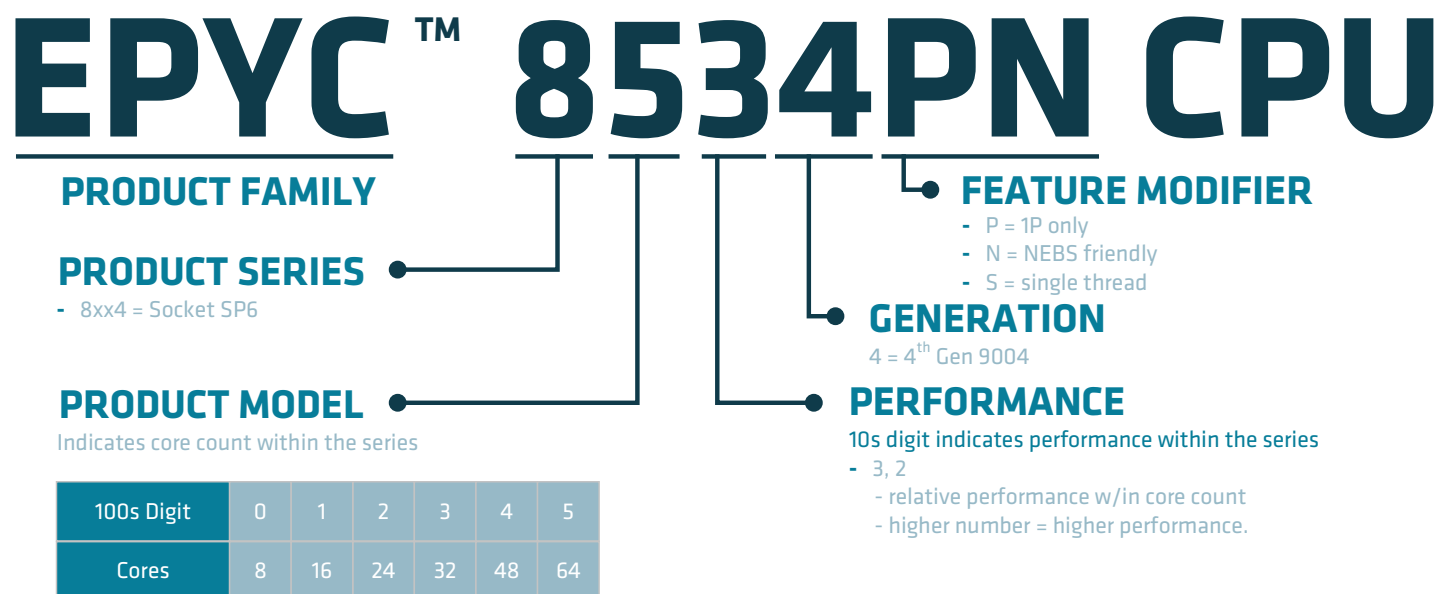

EPYC™ 8534PN CPU

PRODUCT FAMILY

PRODUCT SERIES
- 8xx4 = Socket SP6

PRODUCT MODEL
Indicates core count within the series

100s Digit	0	1	2	3	4	5
Cores	8	16	24	32	48	64

FEATURE MODIFIER
- P = 1P only
- N = NEBS friendly
- S = single thread

GENERATION
4 = 4th Gen 9004

PERFORMANCE
10s digit indicates performance within the series
- 3, 2
  - relative performance w/in core count
  - higher number = higher performance.

Figure 2-1: AMD EPYC SoC naming convention

2.1   CPUID Instruction

Software uses the `CPUID` instruction (`Fn0000_0001_EAX`) to identify the processor and will return the following values:

- **Family:** 19h identifies the "Zen 4" architecture

- **Model:** Varies with product. For example, EPYC Model 10h corresponds to an "A" part "Zen 4" CPU.

  - **8xx4:** Models A0h–AFh

- **Stepping:** May be used to further identify minor design changes

For example, `CPUID` values for Family, Model, and Stepping (decimal) of 25, 17, 1 correspond to a "B1" part "Zen 4" CPU.

2.2   New Software-Visible Features

AMD EPYC 8004 Series Processors introduce several new features that enhance performance, ISA updates, provide additional security features, and improve system reliability and availability. Some of the new features include:

- 5-level Paging

- AVX-512 instructions on a 256-bit datapath, including BFLOAT16 and VNNI support.

- Fast Short Rep STOSB and Rep CMPSB

Not all operating systems or hypervisors support all features. Please refer to your OS or hypervisor documentation for specific releases to identify support for these features.

Please also see the latest version of the *AMD64 Architecture Programmer's Manuals* or *Processor Programming Reference (PPR) for AMD Family 19h*.

2.2.1   AVX-512

AVX-512 is a set of individual instructions supporting 512-bit register-width data (i.e., single instruction, multiple data [SIMD]) operations. AMD EPYC 8004 Series Processors implement AVX 512 by "double-pumping" 256-bit-wide registers. AMD's AVX-512 design uses the same 256-bit data path that exists throughout the Zen4 core and enables the two parts to execute on sequential clock cycles. This means that running AVX-512 instructions on AMD EPYC 8004 Series will cause neither drops on effective frequencies nor increased power consumption. On the contrary, many workloads run more energy-efficiently on AVX-512 than on AVX-256P.

Other AVX-512 support includes:

- Vectorized Neural Network Instruction (VNNI) instructions that are used in deep learning models and accelerate neural network inferences by providing hardware support for convolution operations.

- Brain Floating Point 16-bit (BFLOAT16) numeric format. This format is used in Machine Learning applications that require high performance but must also conserve memory and bandwidth. BFLOAT16 support doubles the number of SIMD operands over 32-bit single precision FP, allowing twice the amount of data to be processed using the same memory bandwidth. BFLOAT16 values mantissa dynamic range at the expense of one radix point.

Chapter 3

Resources

3.1 Resources

Please see the following resources for additional information about AMD EPYC 8004 Series processors:

- AMD EPYC™ 8004 Series Server Processors

- AMD64 Architecture Programmer's Manual

- AMD EPYC™ Tech Docs and White Papers

- *BIOS & Workload Tuning Guide for AMD EPYC™ 8004 Series Processors* (available from AMD EPYC Tuning Guides)

- Memory Population Guidelines for AMD Family 19h Models A0h–AFh Socket SP6 Processors - Login required; please review the latest version if multiple versions are present.

- Socket SP5/SP6 Platform NUMA Topology for AMD Family 19h Models 10h–1Fh and Models A0h–AFh - Login required; please review the latest version if multiple versions are present.

This page intentionally left blank.