INTEL LAUNCHES "EMERALD RAPIDS" 5ᵀᴴ GENERATION XEON

STEVE MCDOWELL, CHIEF ANALYST
DECEMBER 15 , 2023

CONTEXT

Intel unveiled its new Emerald Rapids processors, part of its 5th-Gen Xeon Scalable lineup at its recent "AI Everywhere" event. The new processors arrive with multiple features designed to enhance performance across workloads, including AI and HPC.

This is the second Xeon family upgrade in less than a year, providing more compute power and faster memory while maintaining the same power envelope as the previous generation.

The Intel Emerald Rapids processors bring several performance improvements over their predecessors, particularly in computing power, energy efficiency, and memory handling.

Here are some critical aspects of these improvements:

1. **Enhanced Core Performance:** Emerald Rapids processors offer a significant boost in performance, primarily due to their advanced core designs and higher core counts. With up to 64 cores in the top configurations, these processors can handle more simultaneous tasks, making them ideal for high-performance computing environments.

2. **Refined 10nm Process Technology:** A refined 10nm manufacturing process allows for more efficient and powerful chiplets. This technological advancement not only increases the computational power but also improves energy efficiency.

3. **Increased Cache Size:** A larger L3 cache in the Emerald Rapids processors enables faster data access and processing, reducing latency and improving overall system performance.

4. **Improved Memory Support:** These processors support the CXL 1.1 memory protocol, which enhances memory bandwidth and efficiency. This leads to better handling of memory-intensive tasks and applications.

5. **Higher Memory Bandwidth:** Emerald Rapids CPUs support faster DDR5 memory, offering a higher memory bandwidth than previous generations. This increase in memory speed further boosts the overall system performance.

6. **Enhanced Interconnect Speeds:** The UltraPath Interconnect (UPI) links in Emerald Rapids have been improved, offering faster data transfer rates. This is particularly beneficial in multi-processor configurations where rapid communication between CPUs is crucial.

7. **Performance Per Watt Improvements:** Emerald Rapids delivers a more favorable performance-to-power consumption ratio, making these processors more energy-efficient while still providing robust performance.

8. **Benchmarks and Comparisons:** Emerald Rapids shows a performance improvement of around 1.21X over the Sapphire Rapids generation in various benchmarks. This increase is due to architectural enhancements, increased core counts, and enhanced memory and interconnect speeds.

These performance improvements position Emerald Rapids as a strong contender in the server processor market, especially for data center applications, high-performance computing, and other scenarios where high computational power and efficiency are critical.

FEATURES

Intel's Emerald Rapids processors, part of the 5th-Gen Xeon Scalable lineup, come with various features and specifications designed to bolster their performance in server and data center environments.

- **Core and Thread Configuration**

  - **High Core Count:** The flagship model, like the Xeon Platinum 8592+, offers up to 64 cores and 128 threads, catering to high-performance and multi-threaded workloads.

- **Memory and Cache**

  - **Enhanced Cache Memory:** Emerald Rapids significantly increases the L3 cache compared to its predecessors, with the top models featuring up to 320MB of L3 cache.

  - **DDR5 Memory Support:** These processors support faster DDR5 memory, with speeds up to DDR5-5600 for high-performance models, improving overall memory bandwidth.

- **Process and Microarchitecture**

- **Intel 7 Process:** Emerald Rapids leverages the 'Intel 7' process, a refined version of Intel's node, enhancing performance and energy efficiency.

- **Raptor Cove Microarchitecture:** A slight enhancement over the previous architecture, contributing to improved processing efficiency.

- **Die Architecture**

  - **Simplified Die Design:** Moving from a quad-chiplet to a two-die design, Emerald Rapids offers a more consistent latency profile while reducing complexity.

  - **Transistor Count:** The new die design features 61 billion transistors for dual-XCC models.

- **Platform and Compatibility**

  - **Eagle Stream Platform:** Maintains backward compatibility with the existing Sapphire Rapids platform, simplifying integration and upgrade paths for users.

  - **Socket Compatibility:** Uses the LGA4677 socket, aligning with previous-generation infrastructure.

- **Performance and Power Efficiency**

  - **Thermal Design Power (TDP):** Ranges from 125W to 350W for standard models, with liquid-cooling-optimized chips peaking at 385W.

  - **Performance Enhancements:** Claims of up to 42% improvement in AI inference, 21% in general compute workloads, and 36% higher performance-per-watt compared to previous generations.

- **Acceleration Technologies**

  - **Built-in Accelerators:** Emerald Rapids introduces purpose-built accelerators for tasks like compression, encryption, and data analytics, improving performance in these areas.

- **Additional Features**

  - **PCIe 5.0 Support:** Includes support for 80 lanes of PCIe 5.0, providing substantial bandwidth for high-speed peripherals and storage devices.

  - **Memory Expansion and Support:** CXL Type 3 memory support enables new memory tiering and interleaving options, expanding capacity and bandwidth.

- **Advanced Vector Extensions (AVX) and AMX Licensing:** Provides performance boosts under heavy vectorized workloads, enhancing processing capabilities for specific applications.

EMBEDDED ACCELERATION

Intel's Emerald Rapids processors incorporate several acceleration technologies to boost performance for specific workloads. These technologies are a key aspect of Intel's strategy to enhance the capabilities of their server CPUs, making them more versatile and efficient in handling specialized tasks. Here's a breakdown of these acceleration technologies:

- **Purpose-Built Accelerators:**

  o Emerald Rapids features dedicated regions on the chip that act as accelerators for specific tasks.

  o These accelerators are designed to significantly enhance performance in areas traditionally relying on discrete accelerators.

- **Configurable Accelerator 'Devices':**

  o Each Emerald Rapids chip can be configured with a variable number of accelerator 'devices'.

  o The '+' models have at least one accelerator of each type enabled by default.

  o Customers can upgrade to more units of each type of accelerator for additional fees, offering flexibility based on specific workload requirements.

- **Types of Accelerators:**

  - **Intel® Advanced Matrix Extensions (Intel® AMX):** Enhances deep learning training and inference performance, ideal for natural language processing, recommendation systems, and image recognition.

  - **Intel® QuickAssist Technology (Intel® QAT):** Offloads encryption, decryption, and compression tasks to free up processor cores, enabling systems to serve more clients or use less power. In 4th Gen Xeon processors, Intel QAT provides high performance in compressing and encrypting single data flows.

- **Intel® Data Streaming Accelerator (Intel® DSA):** Optimizes streaming data movement and transformation, crucial for storage, networking, and data-intensive workloads. This helps speed up data movement across CPU, memory, caches, and connected devices.

- **Intel® In-Memory Analytics Accelerator (Intel® IAA):** Increases the efficiency of database and analytics workloads, enhancing query throughput and reducing memory footprint for in-memory database and big data analytics. It's suitable for in-memory databases and open-source data stores.

The acceleration technologies in Intel's Emerald Rapids CPUs represent a significant step towards enhancing the processor's capabilities beyond general compute tasks.

The ability to configure and customize the accelerator mix allows users to tailor the processor to their specific application needs, ensuring optimal performance for diverse workloads.

These processors feature AI acceleration in each core, tackling demanding AI workloads efficiently. They offer up to 42% higher inference performance and sub-100 millisecond latency for large language models under 20 billion parameters, even before incorporating discrete accelerators.

Compared to their predecessors, the 5th Gen Xeon processors show a 21% average performance improvement in general computing and a 36% increase in performance per watt across various workloads. Users upgrading from older generations, following a typical five-year refresh cycle, could see up to a 77% reduction in TCO.

These accelerators provide specialized performance boosts by targeting specific areas like compression, encryption, data movement, and analytics, making Emerald Rapids a versatile choice for a wide range of data center applications.

COMPETITIVE ENVIRONMENT

The competitive environment for Intel's Emerald Rapids processors is intense and multifaceted, primarily due to the rapid advancements and growing demands in the server and data center markets. Here's an overview of the key aspects of this competitive landscape:

- **AMD's EPYC Genoa:**
  - AMD's EPYC processors, especially the Genoa lineup, are a significant competitor to Emerald Rapids. Genoa offers up to 96 cores, surpassing Emerald Rapids' maximum of 64 cores. This core count advantage makes AMD's offering particularly compelling for high-density, parallel computing workloads.

- AMD has been steadily eroding Intel's market share in the server processor domain, challenging Intel's traditional dominance.

- **Performance and Efficiency:**

  - While Emerald Rapids boasts improvements in performance-per-watt and AI inference capabilities, it faces the challenge of matching or exceeding AMD's performance in general compute workloads and power efficiency.

- **Market Segmentation and Demand:**

  - Intel's strategy with Emerald Rapids involves targeting various market segments, including cloud computing, networking, and storage. However, the highest-tier chips often showcase the most advanced features and constitute a smaller portion of the market than mid-range offerings.

  - The demand in these segments is dynamic, with a growing emphasis on energy efficiency, processing power, and specialized capabilities like AI acceleration.

- **Built-in Acceleration Technologies:**

  - Intel's focus on built-in accelerators for specific tasks, such as encryption and data analytics, positions Emerald Rapids uniquely against competitors. These accelerators could be a key differentiator in workloads that benefit from such specialized processing.

- **Future Developments:**

  - AMD's roadmap includes the upcoming Zen 5-powered Turin processors, while Intel is preparing its Granite Rapids processors for launch. These future developments indicate an ongoing and evolving competitive scenario.

  - The competition is not just about core counts and raw performance but also involves energy efficiency, total cost of ownership, and workload-specific optimizations.

Emerald Rapids enters a highly competitive server processor market, contending with formidable offerings from AMD, as well as emerging ARM-based solutions. Its success will depend on how well it balances performance, efficiency, and specialized capabilities to meet the diverse needs of the data center and server markets.

ANALYSIS

Emerald Rapids, marking Intel's continued evolution in its Xeon Scalable Processor family, brings significant improvements in core performance and energy efficiency, critical factors in modern data centers.

With up to 64 cores per processor, Emerald Rapids offers an impressive parallel processing capability, essential for handling the demanding workloads prevalent in enterprise computing today. Adopting a refined 10nm process technology not only underscores Intel's commitment to advancing its manufacturing prowess but also contributes to notable improvements in computing power and power efficiency.

A standout feature of the Emerald Rapids processors is its use of specialized accelerators for tasks such as compression, encryption, and data analytics. The ability to configure processors with varying numbers of accelerator 'devices' offers flexibility based on workload requirements.

However, the launch of Emerald Rapids has its challenges. The processor enters a market experiencing a recession, except in the AI segment, where competitors like Nvidia hold a strong position. Moreover, Intel faces stiff competition from AMD's Epyc series and various Arm-based server CPUs, which have made significant inroads in the server market.

The performance improvements in Emerald Rapids, while substantial, must be viewed in the context of this competitive environment. Intel's struggles in the GPU segment, especially against Nvidia's dominance and AMD's credible offerings, further complicate its market position.

While Emerald Rapids is a testament to Intel's engineering capabilities and its ability to innovate in core performance and efficiency, its success will ultimately depend on how well it can navigate the competitive pressures and market dynamics. The server processor market is more contested than ever, and Intel's Emerald Rapids will be a key product to watch in determining the company's standing in this fiercely competitive arena.

Emerald Rapids demonstrates Intel's commitment to delivering high-performance solutions for AI, data center, and server workloads. These advancements align with Intel's strategy to address the growing demand for AI processing while maintaining a focus on performance, power efficiency, and security. Intel is delivering precisely what's required in today's hypercompetitive compute market.

NAND Research is a technology-focused industry analyst firm providing research, customer content, market and competitive intelligence, and custom deliverables to technology vendors, investors, and end-customer IT