
Ampere Computing AmpereOne Aurora

STEVE MCDOWELL, CHIEF ANALYST
8/3/24

CONTEXT

Ampere [announced updates](#) to its AmpereOne processor roadmap, highlighting its upcoming product AmpereOne Aurora processor. The company also offered pricing updates and other details about its overall roadmap.

Key features of AmpereOne Aurora include:

- Up to 512 Ampere Cores, offering over three times the performance of current AmpereOne processors.
- A scalable AmpereOne Mesh for seamless compute connectivity.
- Integrated Ampere AI IP directly in the silicon hardware, along with high bandwidth memory.

AmpereOne Aurora is designed to scale across AI inference and training use cases, providing leading performance per rack for AI Compute and enabling deployment in any existing data center worldwide. This makes it adaptable for public clouds, enterprises, hyperscale data centers, and edge environments, addressing the global AI power crisis.

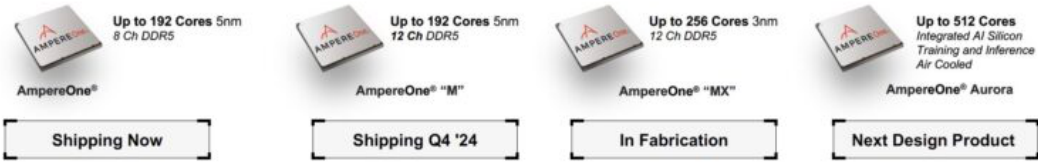
AMPERE'S UPDATED ROADMAP

Ampere Computing laid out a detailed and ambitious roadmap to enhance its processor lineup, with a focus on cloud-native AI computing. The roadmap outlines the progression of its current and future products, emphasizing performance improvements, AI integration, and competitive positioning.

The Ampere Roadmap: Powerful Roadmap with Rapid Innovation

Continued Commitment to Leadership Performance Per Rack for AI Compute in Air Cooled Environments

AmpereOne® Family



Ampere® Altra® Family



CURRENT PRODUCTS

AmpereOne (2023)

- **Core Configuration:** Currently shipping with up to 192 cores on a 5nm process.
- **Target Use:** Designed for high-performance computing tasks optimized for cloud environments.

UPCOMING PRODUCTS

AmpereOne M Series (Q4 2024)

- **Core Configuration:** Maintains up to 192 cores.
- **Platform:** Built for the 12-channel DDR5 platform showcased at Computex 2023.
- **Enhancements:** Improved memory channel capabilities, laying the groundwork for higher core counts and better performance.

AmpereOne MX (2025)

- **Core Configuration:** Increases to 256 cores.
- **Platform:** Continues utilizing the 12-channel DDR5 architecture.
- **Target Use:** Enhanced for handling more demanding workloads, providing greater compute power and memory bandwidth.

AmpereOne Aurora (2025)

- **Core Configuration:** 512 cores.
- **Performance:** Ampere promises an increase of three times the performance per rack compared to current AmpereOne processors, accounting for compute and power efficiency.
- **AI Integration:** Integrates a proprietary AI engine directly into the silicon, designed for AI training and inference tasks.
- **Memory:** Incorporates high bandwidth memory (HBM) on the SoC.
- **Cooling and Deployment:** Maintains an air-cooled design, ensuring deployability across various data centers without needing specialized cooling solutions.
- **Scalability:** Features a scalable mesh interconnect, allowing for seamless integration of multiple chiplets.

PRICING DETAILS

SKU/Model Details and Suggested Base Volume Pricing



Processor Model	Core Count	Frequency (GHz)	Usage Power*	SBV USD Price
AmpereOne® A192-32X	192	3.2	274	\$5,555
AmpereOne® A192-26X	192	2.6	211	\$5,237
AmpereOne® A160-28X	160	2.8	215	\$4,760
AmpereOne® A144-27X	144	2.7	200	\$4,284
AmpereOne® A144-26X	144	2.6	332	\$2,936
AmpereOne® A144-24X	144	2.4	185	\$3,967
AmpereOne® A128-34X	128	3.4	310	\$4,760
AmpereOne® A96-37X	96	3.7	293	\$4,761

- The 192-core AmpereOne part is priced significantly higher than the 128-core Ampere Altra Max, at over \$5000 compared to \$2500.
- Lower-core count and higher clock speed parts are priced over \$4700.
- The AmpereOne A144-26X, clocked at 2.6GHz, is priced at \$2936 but appears to consume a lot of power, indicating lower binned silicon running at a higher voltage.

NEW: AMPEREONE AURORA

Ampere’s AmpereOne Aurora is a new processor on the roadmap. It is designed to address the growing convergence of general-purpose and AI workloads in cloud environments, making it a critical component for modern data centers.

The Next Step in the AmpereOne Roadmap: AmpereOne® Aurora

Integrated

AI Acceleration Silicon
Directly in the SOC

Air Cooled

Deployable in All Existing
Data Centers

Efficient

Accelerate Our Industry’s
Climate Goals



- Up to 512 AmpereOne custom cores delivering over 3x the performance per rack of current AmpereOne processors
- AmpereOne Scalable Mesh that allows compute of all types to be seamlessly connected
- Ampere AI IP integrated directly in the SoC via Die-to-Die Interconnect with high bandwidth memory

KEY FEATURES

1. Core Configuration:

- **Up to 512 Ampere Cores:** The Aurora processor will feature a staggering 512-core configuration, delivering over three times the performance of current AmpereOne processors.

2. AI Integration:

- **Integrated AI Acceleration:** Aurora directly incorporates Ampere's proprietary AI engine into its silicon. This integration enhances AI inference and training capabilities, supporting retrieval-augmented generation (RAG) tasks and vector databases.

3. Memory and Interconnect:

- **High Bandwidth Memory:** To support the data-intensive nature of AI workloads, Aurora includes high bandwidth memory.
- **Scalable Mesh Interconnect:** The processor features a custom mesh interconnect that seamlessly connects CPU and AI cores.

4. Cooling and Deployability:

- **Air-Cooled Design:** Despite its high-performance capabilities, Aurora is designed to be air-cooled, enabling deployment in existing data centers without requiring specialized cooling solutions.

5. **Performance:**

- **Leading Performance per Rack:** Aurora is expected to provide top-tier performance per rack for AI compute, making it suitable for a wide range of AI inference and training use cases.

6. **Flexibility and Scalability:**

- **Scalability Across Use Cases:** Aurora can scale across various AI inference and training applications, offering powerful compute capabilities for different workloads.
- **Adaptable Memory Channels:** While the exact number of memory channels is not finalized, Aurora is expected to have at least 16, possibly up to 24 channels, ensuring balanced and efficient data processing.

COMPETITIVE LANDSCAPE

The AmpereOne Aurora will likely be available around the same time as AMD EPYC "Turin Dense" (192 cores/384 threads) and Intel Sierra Forest-AP (288 cores/288 threads).

Ampere's pricing strategy appears to be more competitive compared to AMD's and Intel's enterprise pricing, aligning more closely with their cloud pricing models.

ANALYSIS

Ampere's roadmap showcases a clear trajectory of innovation, focusing on increasing core counts, integrating advanced AI capabilities, and maintaining energy-efficient designs. The company's strategic positioning and competitive pricing aim to make Ampere a strong contender in the evolving landscape of high-performance computing and AI workloads.

The AmpereOne Aurora is a significant step forward for the company. The new processor combines high core counts, integrated AI acceleration, and scalable interconnects with an energy-efficient, air-cooled design. This processor is well-suited for modern data centers and enterprise applications, offering powerful performance and flexibility while maintaining sustainability.

Ampere's strategic focus on creating an open, interoperable, and accessible AI compute environment further enhances its appeal, positioning the Aurora processor as a key player in the evolving landscape of cloud and AI computing.

While initially targeting hyperscalers and cloud builders, Ampere is now positioning its processors as a viable option for enterprises and other organizations without the expertise or budget to design custom Arm chips. Ampere aims to be the go-to choice for those seeking the benefits of cloud-native Arm architecture without the need for in-house chip design.



© Copyright NAND Research.

NAND Research is a registered trademark of NAND Research LLC, All Rights Reserved.

This document may not be reproduced, distributed, or modified, in physical or electronic form, without the express written consent of NAND Research. Questions about licensing or use of this document should be directed to info@nand-research.com.

The information contained within this document was believed by NAND Research to be reliable and is provided for informational purposes only. The content may contain technical inaccuracies, omissions, or typographical errors. This document reflects the opinions of NAND Research, which is subject to change. NAND Research does not warranty or otherwise guarantee the accuracy of the information contained within.

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 organizations.

Contact NAND Research via email at info@nand-research.com or visit our website at nand-research.com.