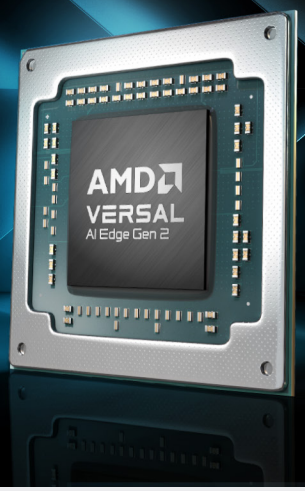
END-TO-END ACCELERATION FOR EMBEDDED AI

AMD
together we advance_



OVERVIEW

AMD Versal™ AI Edge Series Gen 2 adaptive SoCs deliver end-to-end acceleration for AI-driven embedded systems—all in a single device built on a foundation of enhanced safety and security. Combining world-class programmable logic with a new high-performance processing system of integrated Arm® CPUs and next-generation AI Engines, these devices enable all three phases of compute in embedded AI applications: preprocessing, AI inference, and postprocessing.

Embedded systems interact with the world in real time and face tight operating constraints. To respond to real-world conditions in milliseconds and meet size and power targets, all phases of processing need hardware acceleration, but until now, no single hardware architecture is well optimized for all three phases. Multi-chip solutions carry significant overhead, so a single-chip heterogeneous processing solution that includes programmable logic for flexible, real-time preprocessing is required to maximize performance.

Designed for a broad range of embedded markets, including those with high-security, high-reliability, long-lifecycle, and safety-critical applications, Versal AI Edge Series Gen 2 adaptive SoCs are designed to meet ASIL D / SIL 3 operating requirements and are compliant with numerous other safety and security standards.

HIGHLIGHTS

PREPROCESSING, AI INFERENCE, AND POSTPROCESSING – IN ONE DEVICE

- World-class programmable logic for flexible, real-time preprocessing
- Next-generation AI Engines for efficient AI inference –up to 3X TOPS/watt[1]
- Up to 10X scalar compute[2] with high-performance CPUs for postprocessing
- Supported by new DDR5/LPDDR5X memory controllers

FOR HIGH-SECURITY, SAFETY-CRITICAL APPLICATIONS

- Up to 100k DMIPs of compute at ASIL D / SIL 3 (random) operation[3]
- ASIL D / SIL 3 operation from processing system to NoC to DDR memory
- New application security unit and DDR inline crypto for run-time security
- Secure boot and device configuration through the platform management controller

HARDENED IMAGE AND VIDEO PROCESSING TO SAVE SPACE AND POWER

- New image signal processor tiles – over 1 Gpix/s throughput per tile
- Enhanced video codec unit supports HEVC & AVC encode & decode[4]
- Video processing pipeline for horizontal mirroring, tone mapping, and more
- Integrated 4-core Arm Mali™-G78AE GPU for real-time display/HMI

KEY APPLICATIONS

AUTOMOTIVE

ADAS

Autonomous Driving

INDUSTRIAL AND SMART CITY

Autonomous Mobile Robots

Industrial PCs

Edge AI Box

AEROSPACE AND DEFENSE

Avionics, UAS, UAM

Mission Computing

Detection and Tracking

HEALTHCARE

Ultrasound

Endoscopy

3D Imaging

FEATURES

FEATURE	HIGHLIGHTS
Processing System (PS) of Integrated CPUs	• Up to 8x Arm Cortex-A78AE application processors – up to 200k DMIPs • Up to 10x Arm Cortex-R52 real-time processors • Support for USB 3.2, DisplayPort™ 1.4, 10G Ethernet, PCIe® Gen5, and more
AI Engines	• New AIE-ML v2 tile architecture – up to 3X TOPS/watt vs AIE-ML[1] • Expanded data type options – new FP8, FP16, MX6, MX9 support • MX6 & MX9: Shared-exponent data types designed to boost throughput and perf/watt
Programmable Logic (PL)	• Low-latency, deterministic, parallel processing • Fully customizable to enable differentiated, proprietary algorithms • Field-upgradeable: Adaptable to changing conditions and evolving workloads
Functional Safety	• ASIL D / SIL 3 (random) operation from PS through NoC to DDR memory • Up to 100k DMIPs of compute at ASIL D / SIL 3 (random) operating levels[3] • Entire device designed to ASIL D / SC3 for systematic faults
Security	• New application security unit provides run-time HSM security • Platform management controller manages secure boot and device-level services • DDR memory controllers support inline encryption (AES-XTS or AES-GCM)
Image Signal Processors (ISPs)	• Support up to five camera streams with >1 Gpix/s total throughput per ISP tile • Up to 3 ISP tiles per device – over 3 Gpix/s ISP throughput in largest device • Save power and programmable logic resources vs. soft ISP implementation
Video Codec Unit (VCU)	• Hardened VCU tile contains both encoder & decoder instance • Encoder & decoder support HEVC & AVC up to 4K60, 4:4:4, 12-bit[4] • Multistream support through time division multiplexing
Video Processing Pipeline (VPP)	• Hardens color space conversion, tone mapping, horizontal mirroring, and more • Two symmetrical video pipelines per tile – use independently or blend and combine • Offered in the Versal AI Edge Series Gen 2 2VE3358 device
Integrated GPU	• 4-core Arm Mali-G78AE GPU with up to 268 GFLOPs of compute (FP32 MACs)[5] • Four shader cores in 2 slices – configurable as 1 or 2 independent partitions • Support for: OpenGL® ES 3.2, OpenGL SC 2.0, Vulkan® 1.2, Vulkan SC, OpenCL™ 3.0
DDR5/LPDDR5X Memory Controllers	• Support for DDR5 @ 6400 Mb/s and LPDDR5X @ 8533 Mb/s • Up to 170 GB/s memory bandwidth in the largest devices[6] • Flexible pin planning – swap hard controller pins to support other interfaces
Programmable I/O	• New high-performance X5IO support DDR5/LPDDR5X, LVDS, and other standards • New MIPI C-PHY support (4.5 GS/s) to complement 4.5 Gb/s D-PHY support • HDIO and MIO support lower speeds and logic levels up to 3.3V
Network on Chip (NoC)	• High-bandwidth software-programmable network on chip • Data movement alternative to PL-based routing • Assured quality of service (QoS) to prioritize critical traffic
32G High-Speed Serial Transceivers	• Production-proven 32G GTYP transceivers • Up to 20 PL-facing transceivers per device • 4 additional PS-facing transceivers per device for PS-based 10 GbE, PCIe Gen5
100G Multirate Ethernet	• Channelized for 1x100 GbE, 2x50 GbE, 1x40 GbE, 4x25 GbE, or 4x10 GbE • Integrated FECs for robust error correction (KR FEC, KR4 FEC, KP4 FEC) • FEC bypass mode for custom use
PCIe Gen 5	• PL-based support for PCIe Gen5x4, Gen4x8, and other configurations • Hardened PCIe controller IP blocks integrated into programmable logic • Up to 4 PL-based controllers per device; additional PCIe Gen5 controllers in PS

NEXT STEPS

For more information on AMD Versal AI Edge Series Gen 2, visit **www.amd.com/versal-ai-edge-gen2**

ENDNOTES

1. Based on AMD internal performance and power projections for the AIE-ML v2 compute tile architecture in the Versal AI Edge Series Gen 2 using the MX6 data type, compared to performance specifications and AMD Power Design Manager power results for the AIE-ML compute tile architecture featured in the first-generation Versal AI Edge Series using the INT8 data type. Assumptions: 2 row, 8 column sub-arrays. Operating conditions: 1 GHz F_{MAX}, 0.7V AIE operating voltage, 100°C junction temperature, typical process, 60% vector load, % activations = 0 < 10%. Actual performance will vary when final products are released in market. Performance projections as of March 2024. (VER-023)
2. Based on AMD internal pre-silicon performance estimates for combined total DMIPs of the Versal AI Edge Series Gen 2 and Versal Prime Series Gen 2 processing system when configured with 8 Arm Cortex-A78AE applications cores @2.2 GHz and 10 Arm Cortex-R52 real-time cores @1.05 GHz, compared to the published combined total DMIPs of the processing system in the first-generation Versal AI Edge Series and Versal Prime Series. Versal AI Edge Series Gen 2 and Prime Series Gen 2 operating conditions: Highest available speed grade, 0.88V PS operating voltage, split-mode operation, maximum supported operating frequency. First-generation Versal AI Edge Series and Versal Prime Series operating conditions: Highest available speed grade, 0.88V PS operating voltage, maximum supported operating frequency. Actual DMIPs performance will vary when final products are released in market. (VER-027)
3. Based on pre-silicon performance estimates, the application processing unit (APU) in the processing system in the Versal AI Edge Series Gen 2 and Versal Prime Series Gen 2 is projected to offer up to 100k DMIPs of compute while meeting ASIL D / SIL 3 safety standards when configured with 8 Arm Cortex-A78AE application cores @ 2.2 Ghz, with all cores operating in lockstep mode. (VER-028)
4. Video codec acceleration (including at least the HEVC (H.265), H.264, VP9, and AV1 codecs) is subject to and not operable without inclusion/installation of compatible media players. (GD-176)
5. Based on Arm published product specifications for the Versal AI Edge Series Gen 2 and the Versal Prime Series Gen 2 devices, respectively configured with a 4-core Arm Mali-G78AE GPU, maximum operating frequency 1050 MHz, 64 FP32 per ops/clock/core, and 4 texels per ops/clock/core. Actual Versal AI Edge Series Gen 2 and Versal Prime Series Gen 2 product performance will vary when final products are released in market. (VER-030)
6. Based on AMD engineering pre-silicon performance estimates for the Versal AI Edge Series Gen 2 2VE3858 device with 5x 32B memory controllers and expected maximum LPDDR5X memory data rate of 8.533 GB/s, compared to an in-production first-generation Versal AI Edge Series VE2802 device with 3x 64b memory controllers operating at the published maximum LPDDR4X memory bandwidth of 102.4 GB/s. Actual memory bandwidth calculations for the Versal AI Edge Series Gen 2 devices are subject to change when final products are released in market. (VER-031)

DISCLAIMERS

COPYRIGHT NOTICE