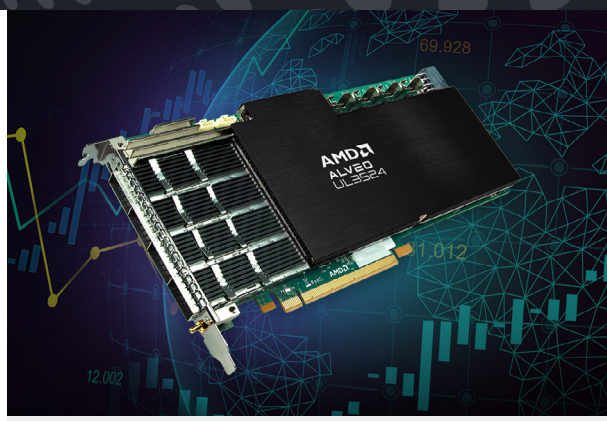
ALVEO™ UL3524 ACCELERATOR CARD

FPGA Accelerator for Ultra-Low Latency Trading



OVERVIEW

Today's leading trading firms, market makers, hedge funds, brokerages, and exchanges are continuously looking for new ways to improve tick-to-trade performance for competitive advantage in financial markets. The Alveo UL3524 FPGA accelerator card combines ultra-low latency networking with adaptable hardware to accelerate trading strategies at nanosecond speed.

The Alveo UL3524 card is powered by a purpose-built FPGA for electronic trading, based on the production-proven 16nm UltraScale+ architecture. The device features a breakthrough transceiver architecture that achieves less than 3ns transceiver latency† and is 7X faster than previous generation FPGA technology††, delivering high-performance trade execution.

Equipped with 4 network ports at 10/25Gb/s data rates, the card comes in a Full-Height, ¾ Length (FH¾L) PCIe® form factor, deployable in 1U, 2U, and 4U servers. The Alveo UL3524 card is now shipping and in production deployment.

HIGHLIGHTS

Purpose-Built for Ultra-Low Latency (ULL) Performance

- Custom FPGA device and new transceiver architecture for fast trade execution
- Less than 3ns transceiver latency† and 7X performance vs. previous generation††

Hardware Flexibility and AI-Enabled Trading Strategies

- FPGA fabric to accelerate diverse strategies and evolving algorithms
- Open-source PyTorch development flow available for low latency AI

For Diverse FinTech Applications

- For ULL algorithmic trading, pre-trade risk analysis, and data delivery services
- Enabling an ecosystem of custom FinTech solutions and ULL infrastructure

Less than 3ns transceiver latency†

KEY APPLICATIONS

TARGET USERS

- Proprietary Traders
- Hedge Funds
- Market Makers
- Brokerages
- Market Data Vendors
- Exchanges

USE CASES

- Ultra-Low Latency Trading
- Pre-Trade Risk Analysis
- Market Data Delivery & Distribution

HFT Solutions, LLC, Ex. 2017
Citadel Securities LLC v. HFT Solutions, LLC, IPR2026-00212
Page 1 of 5

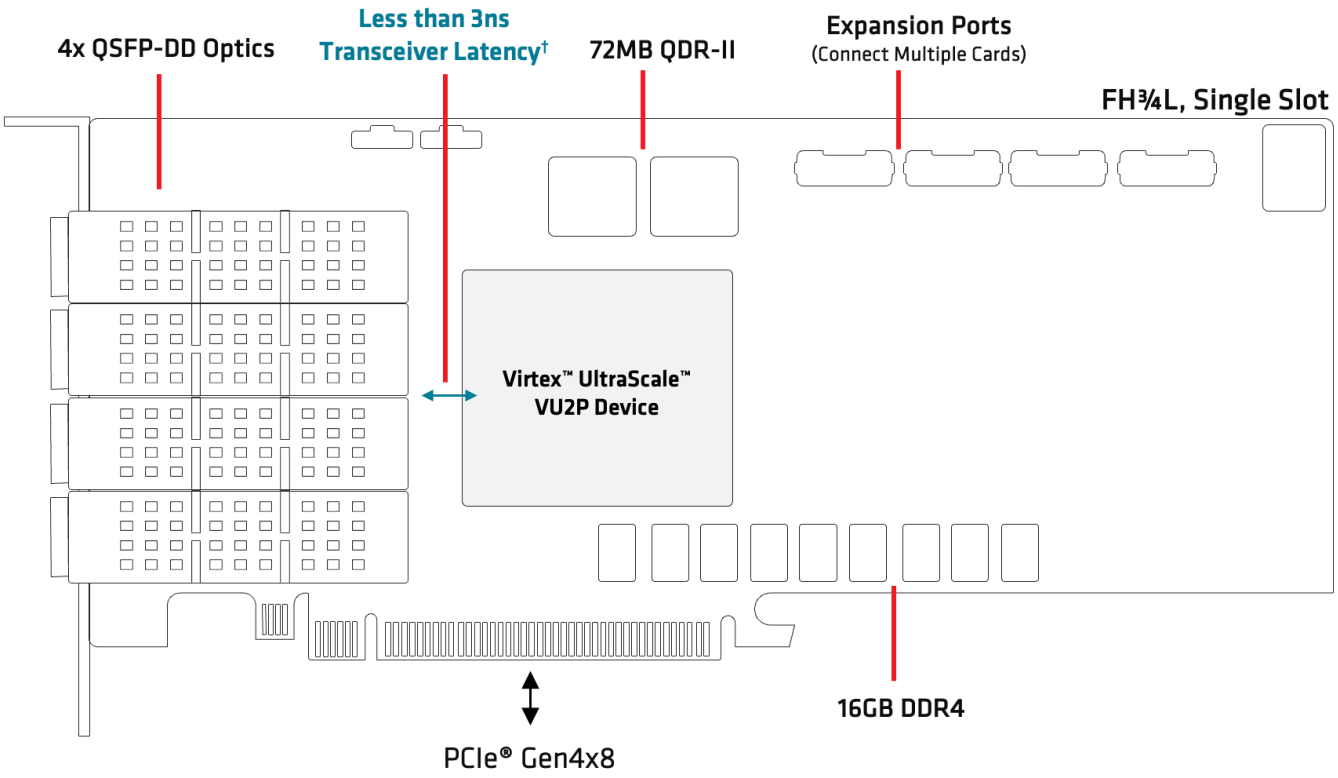
PURPOSE-BUILT FOR ULTRA-LOW LATENCY PERFORMANCE

The accelerator card features a purpose-built Virtex™ UltraScale+™ FPGA device with a new transceiver architecture. Running at 644MHz clock speeds and 16-bit operation, the platform is architected for ultra-low latency, simplifying timing closure and accelerating time to deployment of high-performance trading systems.

Developers can quickly measure performance of the platform in their own environment using a standard reference design from AMD, along with a documented benchmarking methodology and simulation testbench. Request access to the Alveo™ UL3524 lounge to get started.

Alveo™UL3524 Accelerator Card

Powered by Purpose-Built Ultra-Low Latency Transceiver Architecture

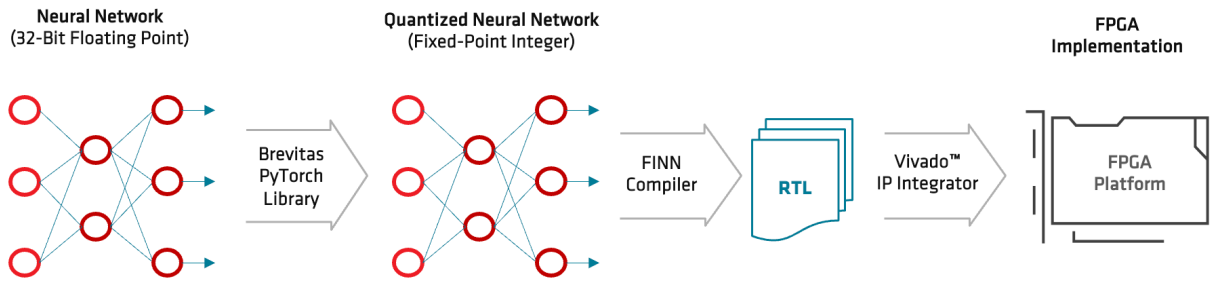


HARDWARE FLEXIBILITY AND AI-ENABLED TRADING STRATEGIES

Featuring 64 ultra-low latency transceivers, 787K LUTs of FPGA fabric, and 1,680 DSP slices of compute, the Alveo™ UL3254 accelerator card is built to accelerate custom trading algorithms in hardware, where traders can tailor their design to evolving strategies and market conditions. Supported by traditional FPGA flows using the Vivado™ Design Suite, the product comes with a variety of reference designs and performance benchmarks so FPGA designers can quickly explore key metrics and develop their own trading strategies to spec, backed by global support from AMD domain experts.

With the increasing adoption of AI in the algorithmic trading market, AMD is making available the open-sourced and communiy-supported FINN development framework. Using PyTorch and neural network quantization techniques, FINN enables developers to reduce the size of their AI models while still retaining accuracy, compile to hardware IP, and integrate their network model into the trading algorithm's datapath for low latency performance. As open-source initiatives, these solutions give developers flexibility and accessibility to the latest advancements as the project evolves.

Open-source Frameworks for Low Latency AI Inference



FOR DIVERSE FINTECH APPLICATIONS

With adaptable hardware, the Alveo UL3524 card is ideal for algorithmic trading, accelerated pre-trade risk checking, real-time market delivery, and more.

For **ultra-low latency algorithmic trading**, financial firms can implement custom strategies in hardware, prioritizing tick-to-trade execution speed or complex algorithm acceleration. Developers can integrate their UDP/TCP stack, feed handler, order book, and trading algorithms.

For **pre-trade risk management**, financial firms can develop custom logic to help ensure SEC 15c3-5 regulatory compliance at ultra-low latency while evaluating market trends, risk of the trading algorithm, and trade execution performance.

For real-time **market data delivery** and distribution, the Alveo UL3524 card can be deployed by exchanges and brokerages to deliver price data, historical data, market depth, news, and other parameters including volatility and options data—all at ultra-low latency performance.

Algorithmic Trading

Pre-Trade Risk Management

Market Data Delivery

SPECIFICATION

CARD FEATURES AND SPECIFICATIONS	
FPGA Resources	• 787K Look-up Tables (LUTs) • 1,722K registers • 1,680 DSP Slices • 256Mb embedded memory (76Mb Block RAM, 180Mb UltraRAM)
Transceivers	• 8 GTYP transceivers (32.75Gb/s) • 64 GTF ultra-low latency transceivers (25.78Gb/s)
On-Board Memory	• 16GB DDR4, 64b +8b ECC at 2666 MT/s • 72MB, 2x288 Mb QDR II+, 550Mhz
Interfaces	• 4x QSFP-DD (32 x 10/25G ports)
Expansion Ports	• Four ARF6 supports additional 32x10/25G ports to connect multiple cards • Two Pico-Clasp connectors for sideband
Clock Synchronization	• 1 PPS In
Configuration and Debug	• 2 Gbit QSPI, JTAG over Micro USB
PCIe® Interface	• PCIe Gen4 x8 (x16 physical connector)
Form Factor	• Full-height, ¾ Length (FH¾L) • Single slot
Power and Thermal	• 180W Electrical • 125W TDP • Passive cooling
Product SKU	• A-UL3524-P16G-PQ-G

In Production and Deployment-Ready

TAKE THE NEXT STEP

- For pricing and availability, contact your local sales representative or complete the Alveo™ UL3524 Inquiry Form.
- To request software licensing and technical documentation, visit the Alveo UL3524 Lounge Request Form.

†Testing conducted by AMD Performance Labs as of 8/16/23 on the Alveo UL3524 accelerator card, using Vivado™ Design Suite 2023.1 and running on Vivado Lab (Hardware Manager) 2023.1. Based on the GTF Latency Benchmark Design configured to enable GTF transceivers in internal near-end loopback mode. GTF TX and RX clocks operate at same frequency of ~644MHz with a 180 degrees phase shift. GTF Latency Benchmark Design measures latency in hardware by latching value of a single free running counter. Latency is measured as the difference between when TX data is latched at the GTF transceiver and when TX data is latched at the GTF receiver prior to routing back into the FPGA fabric. Latency measurement does not include protocol overhead, protocol framing, programmable logic (PL) latency, TX PL interface setup time, RX PL interface clock-to-out, package flight time, and other sources of latency. Benchmark test was run 1,000 times with 250 frames per test. Cited measurement result is based on GTF transceiver "RAW Mode", where PCS (physical medium attachment) of the transceiver passes data 'as-is' to FPGA fabric. Latency measurement is consistent across all test runs for this configuration. System manufacturers may vary configurations, yielding different results. ALV-10

††Based on simulation comparison between Virtex™ UltraScale+™ GTY transceivers and ultra-low latency GTF transceivers.

DISCLAIMERS

The information contained herein is for informational purposes only and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD's products are as set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale.

COPYRIGHT NOTICE

© Copyright 2023 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, Alveo, Artix, EPYC, Kintex, Kria, Radeon, Ryzen, Spartan, Versal, Vitis, Virtex, Vivado, Zynq, and other designated brands included herein are trademarks of Advanced Micro Devices, Inc. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies. AMBA, AMBA Designer, ARM, ARM1176JZ-S, CoreSight, Cortex, and PrimeCell are trademarks of ARM in the EU and other countries. PCIe, and PCI Express are trademarks of PCI-SIG and used under license. PID2336007