Press-Release

Q.ANT Launches First Commercial Photonic Processor - for Energy-Efficient High-Performance Computing and Real-Time AI Applications

Available Now: Q.ANT's first Native Processing Unit is poised to provide at least 30x energy efficiency improvements and substantial performance boost bringing Data Center sustainability within reach.

Stuttgart – November 19, 2024 – Q.ANT, the leading startup for photonic computing, today announced the launch of its first commercial product – a photonics-based Native Processing Unit (NPU) built on the company's compute architecture LENA - Light Empowered Native Arithmetics. The product is fully compatible with today's existing computing ecosystem as it comes on the industry-standard PCI-Express. The Q.ANT NPU executes complex, non-linear mathematics natively using light instead of electrons, promising to deliver at least 30 times greater energy efficiency and significant computational speed improvements over traditional CMOS technology. Designed for compute-intensive applications such as AI Inference, machine learning, and physics simulation, the Q.ANT NPU has been proven to solve real-world challenges, including number recognition for deep neural network inference (see the recent press release regarding Cloud Access to NPU).

"With our photonic chip technology now available on the standard PCIe interface, we're bringing the incredible power of photonics directly into real-world applications. For us, this is not just a processor—it's a statement of intent: Sustainability and performance can go hand in hand," said Dr. Michael Förtsch, CEO of Q.ANT. "For the first time, developers can create AI applications and explore the capabilities of photonic computing, particularly for complex, nonlinear calculations. For example, experts calculated that one GPT-4 query today uses 10 times more electricity than a regular internet search request. Our photonic computing chips offer the potential to reduce the energy consumption for that query by a factor of 30."

Q.ANT's breakthrough relies on its proprietary LENA platform, which includes Thin-Film Lithium Niobate (TFLN) on Insulator chips. Q.ANT has been developing this photonic material since its foundation in 2018. This platform enables precise light control at the chip level. By controlling the entire value chain from wafer to finished processors and leveraging its deep understanding of light, Q.ANT achieves mathematical and algorithmic density surpassing conventional CMOS technology. For instance, a Fourier transform that requires millions of transistors in traditional computing can be accomplished with a single optical element.

"Q.ANT's novel approach to photonic processing is a groundbreaking step towards addressing the escalating energy demands of the AI era," said Eric Mounier, PhD. Chief Analyst, Photonics & Sensing at Yole Group. "This breakthrough is made possible by using optimal materials for optical circuits, which Q.ANT has engineered over the last few years. This new processor generation finally gives access to superior mathematical operations, which have been too energy-demanding on traditional GPUs. The first impact is expected in AI inference and training performance, paving the way for high-efficiency, sustainable AI computing." [1]

The Q.ANT NPU is poised to reduce the calculation requirements for machine learning applications in computer vision or the training and inference of large language models (LLMs.):
- Test runs of the Q.ANT NPU demo system in the cloud with MNIST datasets showed that Q.ANT's native computing approach achieves accuracy comparable to linear networks with lower power consumption.

- Simulations of Kolmogorov-Arnold-Networks (KAN) showed that 43% fewer parameters are required and the number of operations can be reduced by 46%, establishing it as a more efficient choice for AI inference.
- Further tests and simulations on image recognition show that the Q.ANT NPU can train significantly faster and achieve accurate recognition with only 0.1 million parameters and 0.2 million operations. A conventional approach struggles to achieve acceptable results even with 5.1 million parameters and 10 million operations.

Furthermore, it enables faster solutions for partial differential equations in physics simulations, simplifies time series analysis, and improves efficiency in solving graph problems. Unlike standard CMOS technology, the Q.ANT NPU processes data via light, allowing for more power-efficient mathematical operations. While a conventional CMOS multiplier requires 1,200 transistors to perform a simple 8-bit multiplication, the Q.ANT NPU achieves this with a single optical element.

Availability and Packaging

The Q.ANT NPU can be ordered now for delivery in February 2025. The Q.ANT NPU is available as a turnkey Native Processing Server (NPS) that is fully compatible with a traditional server environment and can be integrated into any HPC or data center. By providing early-adopter access to its photonics technology, Q.ANT aims to accelerate breakthroughs in computing and unlock new frontiers in AI and scientific research in a field that will open up significant performance increases and huge energy-saving potential in only a few years' time.

Integrating with existing AI Software Stacks

The Q.ANT Native Computing Unit's intuitive interface, known as the Q.ANT Toolkit, seamlessly integrates with existing AI Software Stacks and allows developers to operate at various levels, from multiplication to optimized neural network operations. It also provides a comprehensive collection of example applications. For pricing and further details or to place a pre-order, contact native-computing@qant.gmbh.

Press-Release
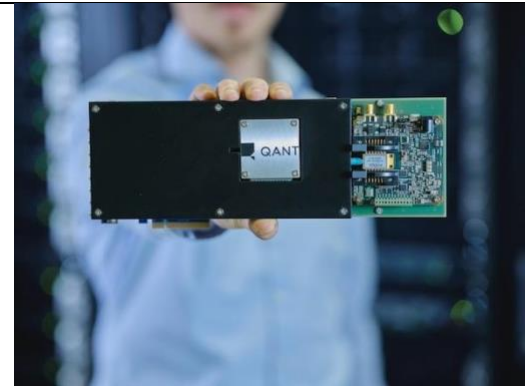
Images

Caption: Photonic Processor on PCI-Express: the Q.ANT NPU is designed for compute-intensive applications such as AI Inference, machine learning, and physics simulation. (image: Q.ANT GmbH)

Caption: The Q.ANT NPU is available as a turnkey Native Processing Server (NPS), fully compatible and ready to plug into any HPC or data center. (image: Q.ANT GmbH)

Caption: Seamlessly integrating into HPC Environments, the Q.ANT NPU promises to deliver greater energy efficiency and significant computational speed improvements over traditional CMOS technology. (image: Q.ANT GmbH)

Press-Release

About Q.ANT

Q.ANT is a pioneer in light-based data processing. Q.ANT's Native Sensing and Native Computing technologies are based on Q.ANT's Para.Digm framework for generating, processing and detecting light, overcoming the limitations of existing technologies and opening up new areas of application in various sectors such as High-Performance Computing (HPC), artificial intelligence, medical technology, aerospace, mechanical engineering and the process industry. Q.ANT emerged as an independent start-up from TRUMPF's research laboratories in 2018 and is headquartered in Stuttgart, Germany.

1. *Source: Optical Computing report, Yole Intelligence, 2024*

Media Contacts

Edith Laga
Fellow PR & Public Affairs, Q.ANT
edith.laga@qant.gmbh
+49 157 830 407 51
Q.ANT GmbH, Handwerkstr. 29, D-70565 Stuttgart, Germany

Toni Sottak
Wired Island International
toni@wiredislandpr.com
+1 843 530 4442