Q.ANT

Q.ANT NPS

The Photonic AI Accelerator

Photonic Processor for energy-efficient High-Performance Computing and Real-Time AI Applications available in a 19″ Rack-mountable Server

Defining the Future of AI Processing

Join a transformative leap in technology – where light meets algorithms to redefine AI processing. Seize the exclusive opportunity to experience Q.ANT's first commercial Photonic AI Accelerator. Based on the innovative Light Empowered Native Arithmetic LENA architecture, this technology will go beyond traditional computing, promising to deliver up to 50 times the computing speed and 30 times the energy efficiency of conventional technologies – and thus reducing operational costs and environmental impact of data centers significantly.

The future is built on Q.ANT's Native Processing Server NPS, the first photonic 19" Rack-mountable server with a Photonic AI Accelerator as a PCIe card designed specifically for AI inference and advanced data processing. Plug & play ready to be integrated in datacenters and HPCs for immediate access to photonic computing. Upgradable with additional PCIe cards for even more processing power in the future. The gateway to a new era of computing with the power of light at ease.

Experience the platform where complex, non-linear mathematical models for AI, machine learning, physics simulations, time-series analysis, and graph problems can be executed with unparalleled performance, powered by the pure energy of light and based on the Q.ANT proprietary material Thin Film Lithium Niobate on Insulator. Don't just adapt to the next generation of computing – define it with Q.ANT.

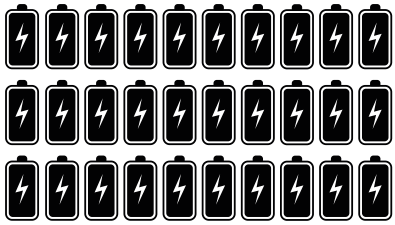
Leverage Q.ANT technology for harnessing the extraordinary potential of photonic AI acceleration. Test, innovate, and get hands-on with a technology that promises a sustainable and powerful future. Redefine the possibilities of AI processing – where cutting-edge efficiency will meet the brilliance of light.

30x
energy efficiency
50x
computing speed

The Potential of Photonic Computing

For AI computations, Q.ANT's **Native Processing Unit (NPU)** promises to **consume 30x less energy** than a Graphics Processing Unit (GPU).

GPU



NPU

High performance: Photonics can run at few tens of GHz bandwidth compared to few GHz in digital electronics, giving more operations performed per second.

Inherent parallelism: Using multiple wavelengths of light to run calculations on the same chip at the same time drastically increases compute density.

Better energy efficiency: Since only light, and no current, is flowing through the circuit, photonic chips have lower cooling requirements. Combining this with higher performance and compute density leads to energy savings.

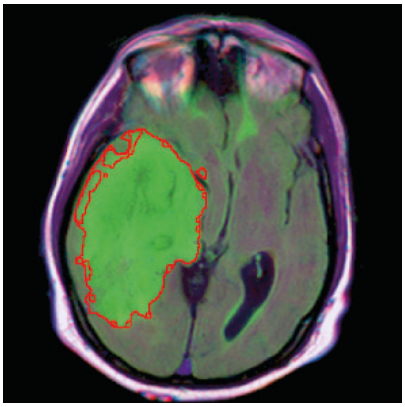
The Software Development Kit to use the NPS
The Q.ANT Toolkit

- The Q.ANT Native Processing Server features an interface known as Q.ANT Toolkit. This interface enables users to operate directly at the multiplication level or to leverage optimized neural network operations, such as fully connected layers or convolutional layers.
- The Toolkit offers a comprehensive collection of example applications that illustrate how AI applications can be enhanced.
- These examples can be used directly or as a foundation for creating own implementations.

Name	Description	Programming Language
Matrix Multiplication	Multiplication of a matrix and a vector	Python / C++
Image Classification	Classification of an image (e.g. based on ImageNet data set)	Python (Jupyter)
Semantic Segmentation	Segmentation of an image (e.g. based on brain MRI scan data set)	Python (Jupyter)
Attention based AI models	e.g. speech recognition	Python (Jupyter)

Powering real-world AI Applications
with Photonic Analog Computing

U-Net for **cancer detection in brain MRI scans** running on a **Native Processing Unit (NPU)**



Q.ANT's Native Processing Server, a photonic analog processor, is solving complex, real-world AI computations such as image recognition and segmentation tasks. From executing ResNet for object recognition in images to applying U-Net for identifying cancer regions in brain MRI scans: The Native Processing Server handles billions of operations with 99% consistency to conventional digital computation, demonstrating the viability of photonic analog computing.

Technical specifications for NPS

System / Subsystem	Feature
System node	X86 based 19” 4U commercially available rack system
Operating System	Linux Debian/Ubuntu with Long-Term Support
Network interface	Ethernet with up to 10 Gbit speed
Software interface	C/C++ and Python API
API to subsystem	Linux device driver
Photonic AI accelerator	• Full length PCIe card with 3 slot height • PCIe Gen3 x8 interface, shared memory & I/O windows • Upgradable with enhanced photonic integrated circuits • Upgradable with enhanced logic functions for performance improvement
Power consumption of photonic AI accelerator	45 W
Photonic integrated circuit (PIC)	Ultrafast photonic core based on z-cut Lithium Niobate on Insulator (LNoI)
Throughput of photonic AI accelerator	100 MOps
Cooling of photonic AI accelerator	Passive
Operating temperature range	15 to 35°C

All specifications may be subject to change without advanced notice.

Photonic Integrated Circuit PIC at the heart of NPS

At the heart of this innovation is Q.ANT’s proprietary Thin Film Lithium Niobate on Insulator PIC, which offers precise light control at the chip level.
Q.ANT controls the entire value chain from raw materials to fully functional systems to make these processors achieve superior mathematical and algorithmic performance.

www.qant.com/photonic-computing

Q.ANT GmbH | Handwerkstraße 29 | 70565 Stuttgart, Germany
+49 711 25245-0 | native-computing@qant.gmbh | www.qant.com