

Q.ANT bringt Generative KI auf photonische Hardware

ISC High Performance 2026: Generative KI und Recurrent Networks laufen auf Q.ANTs Photonik-Prozessor der zweiten Generation

Stuttgart, 23. Juni 2026 – Q.ANT, Pionier im Bereich kommerzielles photonisches Computing, hat heute erstmals hochkomplexe, produktionsrelevante KI-Workloads auf eigener photonischer Hardware demonstriert.

Auf der ISC High Performance 2026 in Hamburg zeigt Q.ANT erfolgreich ein Diffusionsmodell und ein Recurrent Neural Network auf seinem photonischen Prozessor Native Processing Unit NPU der zweiten Generation. Damit belegt Q.ANT, dass die eigene photonische Architektur die volle Bandbreite moderner KI-Modelle unterstützt: von generativer Bildsynthese bis hin zu sequenzieller Zeitreihenprognose.

Die Demonstrationssysteme auf der ISC zeigen darüber hinaus die Fortschritte im derzeit entstehenden Ökosystem. Bereits Anfang dieses Jahres realisierte Daisytuner, ein Compiler-Entwickler, ein Objekterkennungsmodell direkt aus PyTorch auf photonischen Prozessoren von Q.ANT. Damit wurde erstmals ein KI-Modell aus einem Standard-ML-Framework erfolgreich für photonische Hardware kompiliert.

Diese Durchbrüche zeigen, dass der photonische Prozessor NPU über grundlegende Algorithmen hinausgewachsen ist und inzwischen ein Niveau erreicht, das echte kommerzielle Anwendungen adressiert. Auf Q.ANT-Hardware zielen diese High-Performance-Computing-Workloads darauf ab, bei äquivalenten Matrixoperationen bis zu 30-mal energieeffizienter als klassische Prozessoren zu arbeiten – auf Ebene des photonischen Chips.

„Die photonische Architektur von Q.ANT verändert die Energiegleichung für das KI-Ökosystem“, sagt Dr. Michael Förtsch, Gründer und CEO von Q.ANT. „Wenn man mit Licht statt mit Transistoren rechnet, reduziert man den Energieverbrauch an der Quelle. Und es ist hinlänglich bekannt, dass die Energie der Engpass für die Zukunft von KI ist; diesen muss die Branche überwinden. Unsere jüngsten Demonstrationen generativer KI zeigen, dass photonische Hardware die mathematische Last anspruchsvollster moderner KI-Workloads tragen kann.“

Generative KI auf photonischer Hardware

Um die Eignung für generative KI zu verdeutlichen, führt Q.ANT-Hardware ein Diffusionsmodell zur Image-to-Image-Synthese aus. Diese Klasse generativer Workloads ist durch iterative, parallelisierte Matrixoperationen geprägt. Die Anwendung unterstreicht die Eignung photonischer Systeme in einer der rechenintensivsten Architekturen der KI und markiert das erste Mal, dass ein Diffusionsmodell dieser Komplexität auf photonischer Hardware ausgeführt wurde.

Diffusionsmodelle erzeugen Bilder über wiederholte Forward Passes eines Deep Neural Networks mit Matrizenberechnungen. Indem Q.ANT die Rechenschritte mit Licht statt mit Transistoren ausführt, geht die photonische Hardware über grundlegende Demonstratoren hinaus und verarbeitet die lineare Algebra, die im Kern moderner KI-Anwendungen liegt.

„Diffusionsmodelle sind weit verbreitete und rechenintensive Ansätze in der modernen, generativen KI. Sie basieren auf sich wiederholenden, sehr umfangreichen Rechenoperationen, um schrittweise eine kohärente Ausgabe zu erzeugen“, sagt Professor Dr. Björn Ommer, Leiter der Computer Vision & Learning Group an der Ludwig-Maximilians-Universität München (LMU), an dessen Lehrstuhl die Diffusionsmodelle entwickelt wurden. „Wenn photonische Hardware solche Arbeitslasten effizient und zuverlässig ausführen könnte, ist das ein spannender Hinweis darauf, dass alternative Rechenarchitekturen in der Zukunft der generativen KI eine wichtige Rolle spielen.“

Zeitreihenprognose mit xLSTM

Als Beleg für die Erweiterbarkeit auf strukturell andere Architekturen präsentiert Q.ANT zudem das Zeitreihen-Prognosemodell TiRex, entwickelt vom österreichischen Frontier-AI-Lab NXAI, das die Extended Long Short-Term Memory (xLSTM)-Architektur für Enterprise-Anwendungen kommerzialisiert.

„Wir arbeiten bei TiRex von Beginn an daran, das Verhältnis von Leistung und Energieverbrauch neu auszubalancieren. Es auf photonischer Hardware von Q.ANT laufen zu sehen, ist beeindruckend und eröffnet ein neues Kapitel im Computing“, sagt Lukas Fischer, Head of Applied Research bei NXAI. „xLSTM-Architektur auf photonischen Systemen könnte neu definieren, was energieeffiziente KI überhaupt bedeutet.“

Im Unterschied zu transformerbasierten Modellen ist xLSTM ein Recurrent Neural Network, das Muster in sequenziellen Daten erkennt und zukünftige Werte über lange Zeithorizonte vorhersagen kann. Das kommerzielle TiRex-Modell von NXAI mit produktionsoptimierten Gewichten eignet sich unter anderem für Finanzmarktanalysen, Supply-Chain-Optimierungen, Wettervorhersagen und Verkehrsflusssimulationen.

Mit den Demonstrationen von xLSTM und dem Diffusionsmodell zeigt Q.ANT, dass die eigene Hardware in den anspruchsvollsten Klassen neuester KI besteht und arbeiten kann. Ihre erfolgreiche Ausführung belegt die Breite der adressierbaren KI-Anwendungsfälle.

Aufbau eines photonischen Computing-Ökosystems

Die ISC-Demonstrationssysteme sind der jüngste Schritt in einer Reihe von Third-Party-Integrationen, kommerziellen Partnerschaften und Installationen von Q.ANT.

Im Mai kündigte Q.ANT die erste kommerzielle Partnerschaft für seine Hardware mit dem deutschen Cloud-Service-Provider IONOS an. Im April stellte Q.ANT-Partner Daisytuner die Entwicklung eines Compilers vor, der Standard-KI-Toolchains für eine Live-Objekterkennung nutzt. Führende Institutionen wie das Leibniz-Rechenzentrum München und das Jülich Supercomputing Centre, zwei der wegweisenden europäischen HPC-Zentren, betreiben Q.ANT-Hardware im laufenden Produktivbetrieb.

Das sich um Q.ANT formierende Ökosystem spiegelt das grundlegende Vertrauen wider, dass photonisches Computing zur Infrastruktur der nächsten Compute-Generation wird.

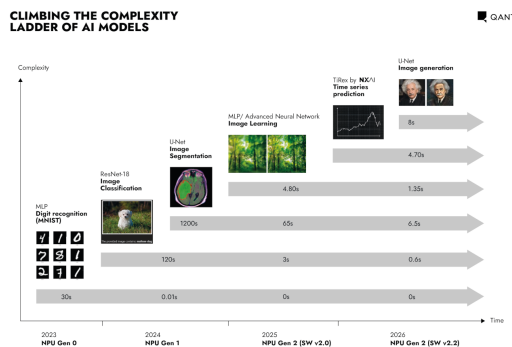
Über Q.ANT

Q.ANT ist ein in Stuttgart ansässiges Photonik-Unternehmen, das photonische KI-Beschleuniger für KI und Hochleistungsrechner entwickelt und eine skalierbare Alternative zu transistorbasierten Systemen bietet. Seine Native Processing Units (NPUs) verwenden photonische integrierte Schaltkreise auf Lithium-Niobat Basis, um mathematische Operationen direkt optisch auf dem Chip durchzuführen, was eine energieeffiziente Co-Prozessierung für komplexe Rechenaufgaben ermöglicht. Q.ANT betreibt in Zusammenarbeit mit IMS CHIPS eine eigene TFLN-Chip-Pilotlinie.

Kontakt für die Medien:

Johannes Manger und Veit Mathauer
Sympra GmbH (GPRA) Agentur für Public Relations
qant@sympra.de | +49 711 947670

Jörg Kochendörfer
Q.ANT GmbH, PR
joerg.kochendoerfer@qant.gmbh | +49 160 5619730



Climbing the complexity ladder of AI models



Q.ANT Working on the Diffusion Model



Q.ANT Native Processing Unit NPU-2

Diese und weitere Bilder stehen zum Download [hier](#).