

About Salience Labs

Al requires new hardware. On-chip integrated optics can provide orders of magnitude performance improvement. Salience Labs has a worldclass solution backed by a decade of research from top research groups to address AI inferencing: a market expected to exceed \$50B by 2025. Salience is developing an ultra-high throughput, low latency Photonic Tensor Processing Unit. The work has its origins in labs at the University of Oxford and Münster University.

Salience Labs uses a proprietary amplitude-based approach to photonics, resulting in small, dense photonic chips clocking at 10's of GHz. This, combined with massively parallel performance (up to 64 beams parallel operations can be packed into a single beam of light), results in a highdensity processing unit.

Salience leverages multi-chip design, with the photonic processing mapping directly on top of Static Random Access Memory (SRAM). This novel 'on-memory compute' architecture is inherently faster and can be adapted to the applicationspecific requirements of different market verticals, making it ideal for realizing AI use-cases in communications, robotics, vision systems, healthcare and other data workloads. Our prototype photonic chip runs neural nets and was published in Nature. The team is currently designing a fully integrated commercial-scale prototype.

Why Now?

Forget Moore's law. At the present, the computational requirements for AI double every 3.5 months. This is vastly greater than anything present day semiconductor scaling can provide. Disruptive solutions are urgently required. Salience Labs brings one such solution – a photonic on-chip technology - to enable high-throughput AI inferencing.

Salience Labs' technology leverages the advancements that have been made in Silicon Photonics fabrication. The chip has been designed from first principles for volume manufacture, and Salience is currently fabricating with production-level foundries using standard CMOS processes.

Our Technology

Salience Labs is developing a massively parallel photonic tensor processing core, with a fully integrated photonic-electronic architecture that interfaces seamlessly with digital electronics. Our solution offers orders of magnitude improvement in throughput and latency of AI compute.

Our proprietary technology has at its core a photonic waveguide array that performs matrix vector multiplications that can clock at >10 GHz. The technology allows for performance to be scaled in multiple ways. Due to Salience's proprietary amplitude-based approach to photonics, the chip can be clocked at 10-100 GHz. In addition, this method allows for multiple vectors – up to 64 – to be encoded onto different colors of light using wavelength division multiplexing. This significantly increases the compute density. Our current on-chip demonstrator uses 4 vectors in parallel at speeds up to 14 GHz [Feldmann et al, Nature 589, 52–58, 2021].



In addition, Salience's multi-chip design stacks the photonic processing chip on top of SRAM in a highly distributed architecture. This ensures that the high-speed photonics chip is fully leveraged to enable a performance boost in AI compute.

Market Opportunity

The AI hardware market is predicted to reach over \$50B by 2025. In addition, AI hardware is becoming increasingly segmented and verticalized by use-case. Salience Labs is developing an application specific AI processor that leverages its high throughput and low latency for AI use-cases in communications, robotics, vision systems, healthcare and other data workloads. We are currently exploring how we can meet customer requirements across a number of use cases in these areas, and welcome any approaches on this front.

Progress to Date

- Founded in March 2021
- Raised \$11.5M Seed round of VC funding, with an additional \$1.3M in grant funding from the German government
- Own worldwide, exclusive licenses across 4 patent families covering the technology from University of Oxford and University of Münster
- Prototype chip and demonstration: fabricated a photonic chip that can run at up to 4 Teraops, demonstrating neural nets for handwritten digit recognition

Team

Salience Labs is a team of over 12 experienced engineers and is growing:

- worked at Oxford Sciences Innovation, a £600M deep-tech VC fund.
- from Münster University in Germany. Johannes designed and tested the first Salience Labs prototype.
- at Sondrel, Huawei and Imagination Technologies.

Consultants and Advisors

- founder and Director of Silicon Catalyst, former CEO of Sematech and PVMC.
- Corporation, \$5.7 billon)
- Fellow at Yale and got a doctorate at Oxford.

Ask

Seeking relationships with commercial partners who are exploring use cases requiring high throughput, low latency AI compute for dedicated workloads.





• Vaysh Kewada, CEO: Physics Masters from Imperial College London, former consultant at McKinsey & Co. Previously

• Johannes Feldmann, CTO: Postdoctoral researcher at Oxford University. Holds a PhD in Physics on photonic computing

• Crescenzo D'Alessandro, Chief Architect: Over 16 years industry experience as SoC architect and in hardware design

• Andrew Garrett, Software Architect: Over 20 years industry experience as a Software Architect at Sony, Nexsan and others.

• Daniel Armbrust: Chairman: over 35 years' experience in semiconductor R&D and operations executive positions. Co-

• Jalal Bagherli, Investor, Board Member: former CEO of Dialog Semiconductor (acq. 2021, Renesas Electronics

• Wolfram Pernice: Head of the Pernice Lab in Heidelberg, Germany, and a world leading expert in photonic integrated circuits, with >7000 citations in peer reviewed journals and 10's millions in grant funding. He was formerly a Postdoctoral

• Harish Bhaskaran: Leads the Advanced Nanoscale Engineering Lab at Oxford, holds the UKRI Manufacturing Fellowship. Expertise in Photonics and Nanomanufacturing, inventor of SRD® & PtSi technology. Formerly at IBM and Yale.