HAETAE

Heterogeneously Integrated Multi-Material Photonic Chiplets for Neuromorphic Photonic Transfer Learning AI Engines

Grant Agreement	101194393
Programme	HORIZON-JU-Chips-2024-3-RIA
Duration	01/10/2024 – 31/09/2027 (36 Months)
Budget	Overall Cost: 2,899,956.00 € EU Contribution: 1,499,956.00 €
Coordinator	Aristotle University of Thessaloniki, GR
Contact	**Prof. Nikolaos Pleros,** email: npleros@csd.auth.gr **Dr. Miltiadis Moralis-Pegios,** email: mmoralis@csd.auth.gr **Dr. Apostolos Tsakyridis,** email: atsakyrid@csd.auth.gr
Website	http://www.haetae.eu

Consortium

ARISTOTLE UNIVERSITY OF THESSALONIKI	Aristotle University of Thessaloniki	Greece
imec	Interuniversitair Micro-Elektronica Centrum	Belgium
AKHETONICS	Akhetonics	Germany
DGIST	Daegu Gyeongbuk Institute of Science & Technology	Republic of Korea
KAIST 1971	Korea Advanced Institute of Science and Technology	Republic of Korea

The challenge

Artificial Intelligence (AI) is booming and so does also its energy footprint. The exponentially increasing scale of Deep Learning (DL) models comes at the cost of high computational power and energy requirements, with the daily power consumed by highly acclaimed Large Language Models (LLMs) like ChatGPT-3 being already at 0.5 GWh. With current projections forecasting that the computational power requirements will double every 5-6 months , the way-to-the-rescue was initially being sought through innovative electronic AI hardware architectures, including neuromorphic and analog-in-Memory Computing (AiMC). Designing, however, an AI hardware roadmap around electronics implies inherent energy constraints due to the speed and power limits of the electronic interconnects inside the circuits. *Meeting the compute power requirements of next-generation AI applications without yielding an energy boom necessitates probably nothing less than a "tectonic shift" in the underlying computing hardware*. This deadlock has driven the emergence of neuromorphic photonics that were theoretically predicted to allow for orders of magnitude improvements in energy and size efficiencies compared to electronic AI platforms with the predictions relying on solid scientific principles and assumptions. Delving, however, deeper in the performance metrics of recent neuromorphic photonic demonstrations reveals a significant discrepancy between projected and achieved efficiencies: instead of the 10s of fJ/MAC expectations, computations with light can still not be offered at energy levels lower than a few pJ/MAC.

Mission statement / Objectives

This is where HAETAE steps in, aiming to overcome these shortcomings and release a versatile and energy-efficient photonic AI processor that can turn the promise of neuromorphic photonics into a tangible reality. To realize its ambitious goals, HAETAE will follow a holistic hardware/software co-design approach targeting the following objectives:

- **Develop a multi-material photonic integration platform that synergizes the best-in-class computational technologies**. HAETAE aims to rely on 3D Si/Si$_3$N$_4$ exploiting the low-loss interconnect and passive circuit substrate of Si$_3$N$_4$ and synergize it with i) SiGe/Ge for high-speed modulation, weighting functions and photodetection, ii) Si-based non-volatile Micro Electro Mechanical systems (MEMS) for zero-energy non-volatile weighting, iii) InP actives for on-chip amplification and non-linearities, providing analogue all-optical reconfigurable Non-Linear Activation Functions (NLAF) without requiring any digital interfaces between cascaded neural layers!

➢ **Adopt the best-in-class linear optical circuit architecture for versatile, loss- and fidelity-optimized photonic Matrix Vector Multiplication (MVM) engines**: HAETAE adopts a newly designed and loss-optimized linear optical Xbar circuit architecture based on a 3D Si/Si_3N_4 layer, following the principles of the conventional Xbar, for its photonic MVM chiplets in order to take advantage of its lowest insertion-loss performance and highest circuit-size scalability perspectives, its error-tolerant design that safeguards 100% accuracy in targeted computations, and its WDM-compatible credentials for boosting computational performance through wavelength parallelism.

➢ **Align photonic Neural Network (NN) technology with Transfer Learning DL models and architectures**. HAETAE aims to repaint the optical computing landscape by aligning its photonic NN technology along the powerful transfer learning (TL) paradigm that seems to perfectly fit the idiosyncrasy of photonic NN hardware. TL extends along the use of pre-trained backbone NNs that are followed by a shallow retrainable layer for task-adaptation purposes, suggesting that its synergy with photonics can i) maximize energy gains through the use of energy-efficient statically-weighted photonic MVM chiplets as its pre-trained photonic backbone NNs, ii) adapt to different applications through the use of a limited number of photonic TMM engines as its shallow photonic retrainable NN layer. The transition into a Photonic TL era will align with the development of an optics-informed NN training model library, designed to optimally streamline DL models for photonic hardware.

➢ **Translate its photonic-based innovations into tangible market opportunities by validating its technology platform along a set of communication and computing application areas**: i) in the Data Centre (DC) communication sector, demonstrating a photonic TL-enabled compute engine as the AI-enhanced DSP module in inter- and intra-DC optical interconnections applications, ii) in the large language models sector, efficiently processing physical languages showcasing the capabilities of its re-trainable photonic MVM chiplet to allow for real-time language translation, sentiment analysis and content filtering tasks, iii) in the inter-DC security sector, achieving line-rate cybersecurity threat detection.

Target technology breakthroughs

HAETAE's cross-cutting workplan, extending from material technology to photonic accelerator architectures and related applications, aims to enable impressive beyond-the-state-of-the-art advancements in multiple fronts. Specifically, HAETAE envisions to develop:

➢ **MEMS phase and amplitude modulators** projected to (i) achieve nearly zero power consumption by migrating from volatile to non-volatile designs, transforming photonic MEMS to electronic memristor equivalents (ii) offer more than 50x higher dynamic range compared to alternative photonic amplitude modulation components, reaching 50 dB.

➢ **All-optical NLA modules** projected to leapfrog current SotA by offering (i) a 5x higher operating frequency reaching 50 GHz, (ii) 80x in area efficiency reaching 1100 GSymbols/s/mm^2 (iii) 6x in energy efficiency reaching 2.5 pJ/NOP.

➢ **A converged PTL engine** projected to surpass current SotA NN hardware accelerators by: (i) 20x in operating frequency reaching 50 Gbaud, (ii) 15x in energy efficiency reaching 18 fJ/MAC, (iii) 20% higher footprint efficiency reaching 2450 GMAC/s/mm2.

➢ **An all-optical DSP module replacement** is expected to offer 50% lower power consumption than currently available electronic counterparts, reaching ~2W, when targeting 800G OSFP modules.

Consortium

HAETAE's consortium is strategically assembled to include the entire technology development chain, comprising a high-quality blend of industrial and academic partners which will work complementary towards satisfying all possible technological requirements and exploitation paths. *HAETAE's team has been built for success by bringing together, across the Republic of Korea and the European Union* :

- Three leading universities (AUTH, KAIST, DGIST) with pioneering outcomes in the area of neuromorphic photonics, combining expertise along all critical integration platforms employed in HAETAE – KAIST and DGIST on MEMS-based and AUTH on silicon-based PNNs,
- A renowned R&D centre (IMEC) with a track-record on developing cutting-edge Silicon Photonic technology, and finally
- An established SME (AKHETONICS) with strong R&D in the fields of all-optical computing.

The synergy between all partners ensures that the development work always remains along the right direction: system design, modelling, testing and evaluation on the basis of valid market application scenarios are indispensable activities to enable HAETAE to maximize its exploitation potential, providing a credible path for breeding innovation into tangible outcomes.