Architecture of the ATON AI processor using 3D signal propagation

N. Kälin[1, 2], R. Kraneis[3], M. Müller[3], G. Scholz[1], A. Waag[3], and C. Werner[1]

[1] Ostfalia University of Applied Sciences, Wolfsburg, Germany
[2] Eastern Switzerland University of Applied Sciences (OST), Rapperswil, Switzerland
[3] Institute of Semiconductor Technology, Nitride Technology Center, TU Braunschweig, Braunschweig, Germany

A fundamental difference between biological and current artificial information processing is that biological nervous systems are developed in three spatial dimensions, whereas conventional computer chips process information only in two dimensions. The essentially planar arrangement of today's standard silicon chips results from the photolithography-based manufacturing process, in which semiconductor structures are created by exposing a flat monocrystal, the wafer, to light.

The two-dimensionality of current computing chips has led to the fact that the performance of modern artificial information processing is generally achieved through high clock frequencies combined with digital and therefore highly accurate value representation. In contrast, natural nervous systems process information more slowly, but with a significantly higher degree of parallelism [1]. Although a high degree of parallelism is generally possible also on conventional chips, power consumption here typically increases linearly with the number of operations n, i. e. power consumption is in $\mathcal{O}(n)$.

In this contribution, the authors present the ATON processor architecture. Here, the actual computing operations are mainly realized in the optical domain. Previous work has already shown that with this approach, electrical power consumption is in $\mathcal{O}_{avg}(\sqrt{n})$ [2].

Although the semiconductor chips associated with the ATON architecture are also structured two-dimensionally, signal propagation occurs orthogonally to the chip surface using the third spatial dimension. Signal emission is realized via highly efficient gallium nitride microLEDs, which generate a spatially high-resolution optical pattern that is then processed in the remaining parts of the processor. Each microLED can be viewed as an independent information source (neuron). With current technology, it is already possible to manufacture highly efficient microLEDs with an edge length of one micrometer (and even less). Consequently, more than 10^8 signal emitters can be integrated on a single square centimeter of chip area. This enables massively parallel information processing, which comes much closer to the requirements of neuro-inspired information processing than today's conventional digital computers.

In addition to the spatial distribution of the signal emitters, it is also possible to modulate microLEDs in the gigahertz regime. Therefore, the amount of information that can be processed with an ATON processor per unit of time is immense. It significantly shifts the boundaries known from current digital technology.

Another essential feature of the ATON architecture is that signals in the optical domain can cross each other without mutual interference. This eliminates many of the problems associated with crossing interconnects known from purely electrical signal processing. In addition, problems in the electrical domain with parasitic loads leading to systematic errors and limiting signal dynamics are minimized.

Current research questions focus particularly on concepts and technologies required to realize the actual computing operations within the ATON processor; for this, the light signals emitted by the microLEDs must be modulated, superimposed, and finally measured in a multi-stage process. Due to the high technical complexity and long turnaround times of the to be manufactured components, a digital twin is created, which allows the processor to be tested and evaluated before all parts of the processor are physically available.

In this conference contribution, the authors present the current state of their research as well as possible applications of their technology.

[1] E Müller *et al* 2023 *Network Neuroscience* **7** (2): 844–863. doi: 10.1162/netn_a_00308

[2] M Müller *et al* 2025 *Neuromorph. Comput. Eng.* **5** 024005. doi: 10.1088/2634-4386/adcbcc