CEAS Space Journal
https://doi.org/10.1007/s12567-025-00612-z

ORIGINAL PAPER

Concept of an in-orbit AI-system based on optical computing

Felix Kübler[1] · Mingwei Yang[1,2] · Lennart Mannteuffel[1] · Okan Akyüz[1] · Janik Wolters[2] · Enrico Stoll[1]

Received: 10 January 2025 / Revised: 11 March 2025 / Accepted: 31 March 2025

Abstract
Machine Learning (ML), or Artificial Intelligence (AI) in general, is among today's fastest-growing methods to handle complex or computationally intensive tasks. ML is commonly implemented with Artificial Neural Networks (ANNs) on conventional computer systems that can limit their full potential. Even with access to specialized hardware such as graphics cards or Tensor Processing Units (TPUs), the demand for more computing power constantly increases. Although these hardware requirements can be met for terrestrial applications, an extraterrestrial or in-orbit application is considerably more challenging. Additional requirements for energy budget, thermal control, and radiation resistance can usually not be met, especially for small spacecraft. The benefits of an AI system for fast onboard data processing would, however, be remarkable. An optical approach to this problem can potentially be the solution. Optical computers promise to be much more energy efficient and better suitable for the mentioned space requirements. An implementation of an optical computing device on a spacecraft has not been done and can be considered as a technological leap. This work, along with the project Optical Computing for Machine Learning in Orbit (OMLO) of the Technical University of Berlin (TU-Berlin), aims to specify and conceptualize such a system.

Keywords Artificial intelligence · Machine learning · Optical computing · Hybrid computing system · Satellite payload · Payload concept

1 Introduction

Since the first approach to an Artificial Intelligence (AI) system in 1956 by Newell and Simon [1], the technology has been continuously improved. Modern Machine Learning (ML) algorithms, such as deep or reinforcement learning, combined with an increase in computing power, have allowed enormous advances, especially in recent years. Today, ML is one of the fastest-growing methods to handle complex or computationally intensive tasks. A study commissioned by the Committee on Industry, Research and Energy (ITRE) of the European Parliament [2] emphasizes the versatility of AI and the great potential it presents for use cases ranging from automotive to healthcare. AI-aided systems are generally on the rise and are expected to become even more prominent in the future. The benefits of AI also apply to the space sector, specifically to in-orbit, deep space, or planetary exploration applications [3]. The uplink and downlink of a spacecraft are the limiting factors for data exchange with a ground station. The transmission of data over a long distance also has considerable latency. Using an AI approach in the ground segment might be beneficial for some use cases, but it will

✉ Felix Kübler
felix.kuebler@tu-berlin.de

Mingwei Yang
mingwei.yang@dlr.de

Lennart Mannteuffel
l.mannteuffel@tu-berlin.de

Okan Akyüz
okan.akyuez@tu-berlin.de

Janik Wolters
janik.wolters@dlr.de

Enrico Stoll
e.stoll@tu-berlin.de

[1] Raumfahrttechnik, Technische Universität Berlin, Marchstraße 12-14, 10587 Berlin, Germany

[2] Institute of Optical Sensor Systems, Deutsches Zentrum für Luft- und Raumfahrt, Rutherfordstraße 2, 12489 Berlin, Germany

Published online: 02 May 2025

also require a vast amount of data that needs to be acquired via the transmission bottleneck. An AI system located on the spacecraft itself can solve this problem. Data can be evaluated or preprocessed in real time, allowing for an increase in autonomy. The Remote Agent Experiment (RAX) on Deep Space 1 (DS1) in 1999 [4] and the Autonomous Sciencecraft Experiment (ASE) on Earth Observing One (EO-1) in 2000 [5] are some of the early examples in these fields. The Mixed-Initiative Activity Plan Generator (MAPGEN) system used for operation planning of the Mars Exploration Rover (MER) in 2003 [6] was one of the ground-based approaches, which was later replaced by the on board Autonomous Exploration for Gathering Increased Science (AEGIS) system. The AEGIS system was first tested on MER-B (Opportunity) in 2009 [7] and has been in routine use on the Curiosity rover since 2016 [8]. On May 18th, 2022 it was also deployed on the Perseverance rover with additional features [9]. As stated by Francis et al. [8], the detection of geological scenes can take up to hundreds of seconds due to the limited computing resources on the Perseverance rover. Computer vision, such as object detection and recognition, is commonly implemented with Artificial Neural Networks (ANNs) or specifically Convolutional Neural Networks (CNNs). Neural Networks (NNs) are mathematically based on Vector-Matrix Multiplication (VMM) and require dedicated hardware to ensure good performance. This hardware can range from conventional graphics cards or Graphics Processing Units (GPUs) to specialized Tensor Processing Units (TPUs) as well as large amounts of Random-Access Memory (RAM). For on-board data processing, these options are only available to a limited extent. The harsh space environment adds strict requirements for power consumption, thermal management, and radiation hardness that are generally not met.

Optical computing systems can offer a solution to these restrictions. VMM requires Multiply-Accumulate (MAC) operations that can be passively realized by controlled attenuation (multiplication) and subsequent superposition (addition) of optical signals [10]. Such linear-optics implementations have been demonstrated based on different free-space configurations [11] and show a promising development in the field of integrated photonic circuits [12]. Most free-space implementations are based on the Stanford multiplier design [13], supporting computational operations including VMM, convolution, correlation, and discrete Fourier transform. In the context of ANNs and CNNs enables VMM or convolution in a single immediate calculation. Moreover, the feasibility of multi-layered NNs has been demonstrated by introducing an atomic non-linearity to separate the otherwise linear-optics layers [14]. Consequently, optical processors have the potential to achieve very high computing power with comparatively low energy consumption. Various wavelengths can even be combined for efficient multiplexing and high-speed optical data transmission. In addition, they exhibit natural resistance against ionizing radiation.

As noted by Caulfield and Dolev [15], an optical computer will inevitably depend on electronic components to some extent. However, there is a desire for a complementary approach, aiming to replace conventional electronic calculation bottlenecks with dedicated optical hardware. These hybrid optical computing systems hold the potential to establish a new standard for ML, particularly in space environments. Despite the recognized utility of similar systems in ML, the conjunction with AI approaches in the space segment remains mostly unexplored.

The purpose of this paper is to explore the potential of (hybrid) optical computing systems for ML with a specific focus on space applications. The primary objective is to conceptualize and design a system capable of performing Optical Vector–Matrix Multiplication (OVMM) within the constraints of a satellite payload. To illustrate the practicality of the proposed system, a use case for ML in space will be identified and evaluated, serving as a demonstrative scenario for training and assessing the NN.

2 Application

The chair of Space Technology (RFT) at the Technical University of Berlin (TU-Berlin) has a large expertise in conceptualizing, developing, and deploying satellite systems. In the last 33 years, 30 satellites have been successfully launched and operated within the scope of 19 different missions. Each mission was specifically dedicated to demonstrating new payloads or novel satellite bus concepts. In addition, another 4 satellites will be launched in the following years.

Among all RFT satellite buses, the TUBiX10 [16], TUBiX20 [17], and TUBiX5 bus are more recent examples. The number in each acronym indicates the total mass of the bus, for example, TU-Berlin innovative neXt generation satellite bus (10 kg) (TUBiX10). These bus systems have been used for multiple missions including Spectrum AnaLysis SATellite (SALSAT) [18], TU Berlin Infrared Nanosatellite (TUBIN) [19], and Nanosatellites in Formation Flight (NanoFF) [20]. They have demonstrated their reliability and can be considered flight-proven systems. While all current RFT satellites have been deployed into Low Earth Orbit (LEO), inter-terrestrial missions have yet to be implemented.

With an increase in system and mission complexity, the adoption of AI capabilities is becoming increasingly more significant. They could, for example, improve the autonomy of spacecraft, which is crucial to managing the transmission distances to Earth and the delays they entail. AI can provide

the desired autonomy for various applications. In particular, object detection and classification offer significant advantages for automated decision-making [8], as well as for pre-evaluation and filtering of data before the downlink.

As part of the project Highly integrated data processing unit for AI-applications for small satellites (AITHER) [21] at the RFT, a processing unit for space-specific AI applications is currently under development. Solely constructed from Commercial Off-The-Shelf (COTS) components, the goal is to find a trade-off between component cost, processing power, and power consumption. Similar to other dedicated hardware, such as GPUs and TPUs, conventional computing hardware could become the limiting factor.

Addressing this challenge in the project Optical Computing for Machine Learning in Orbit (OMLO), we propose an alternative solution replacing traditional computing hardware with optical alternatives. Specifically, CNNs are planned to be utilized in the imagined system. Within the scope of the project, we further desire to integrate the optical components into a common satellite bus. Hence, intermediate hardware to interface both technologies must also be developed. The envisioned system encompasses various components, including mechanisms for data collection, storage, preprocessing, and the conversion of conventional binary data into an optical signal. The intricate aspects of controlling and configuring the optical components within this system are also integral considerations.

As an in-orbit use case, we will focus on object classification based on image data. This could, for example, be used for cloud detection in an Earth observation context. Images containing clouds will be filtered before the downlink, leaving only unobstructed views of the Earth. Thus, the utilization of the downlink bottleneck can be improved. To initially demonstrate the Proof-of-Concept (POC), the Modified National Institute of Standards and Technology (MNIST) dataset [22] is used. It is well-established and ideal for comparison with other works. This dataset contains handwritten digits and is not directly related to an in-orbit application. However, the general approach to AI training and object classification can be transferred to other use cases of object classification.

The project can be divided into three major work packages: implementation of the optical computing setup, satellite integration, and data processing via software. In the following sections, concepts for each of these packages will be outlined.

3 Optical Implementation

The Stanford multiplier [13] outlines an optical computing system, capable of various computational tasks, including VMM, convolution, and others. Typically implemented using Spatial Light Modulator (SLM) in a free-space setup, it executes the convolution process of a CNN in a single step, in contrast to the sequential scanning approach of conventional systems. This relates to a complexity of $O(N)$ versus $O(N \times K)$ respectively, where N is the size of a single-channel image and K is the number of kernel entries.

The output of this optical convolution can then be used as input for the fully connected layers of the NN. Yang et al. [14] demonstrated this process by successfully classifying handwritten digits from the MNIST database. In their experiment, the convolution result was digitalized and fed into a NN on conventional hardware for further processing. However, by integrating an atomic non-linearity immediately after the convolution step, it would be possible to use a fully optical NN instead. This behavior is equivalent to the non-linear transitions between the layers of a common NNs. It is used to separate the otherwise linear operations.

For an in-orbit application, Yang et al.'s proposed free-space setup will in part be miniaturized to create an Optical Convolution Neural Network (OCNN) payload demonstrator. To reduce the overall payload size, the SLM designated for the input layer is planned to be replaced by four silicon-based and Printed Circuit Board (PCB) mounted ring modulators developed by The Leibniz Institute for High Performance Microelectronics (IHP). Further details are discussed in the following Sect. 3.2. At this time, these ring modulators are still undergoing testing and are not available to us. They are, therefore, emulated by an SLM for the time being. It is important to note that each ring modulator can only encode a single value, limiting the whole input to four values at a time rather than the N-pixel array of an SLM. This strict input limitation renders the convolution process of the proposed optical system impractical. In Optical Convolution Networks (OCNs), the acceleration potential scales with the size of the convolution kernel. Employing the convolution theorem

$$f(x) * h(x) = \mathcal{F}^{-1}\{\mathcal{F}[f(x)] \cdot \mathcal{F}[h(x)]\} \quad \text{given two arbitrary functions } f(x) \text{ and } h(x) \tag{1}$$

and the Fourier transform properties of lenses, the kernel can be chosen as big as the feature map to which it is applied. Compared to spatial kernels in digital CNNs that mostly rely on 3×3 kernels, Fourier kernels in OCNNs can be much larger and commonly reach sizes of 100×100 [23].

Due to the hardware constraints introduced by the ring modulators, we are thus limited to 2×2 kernel matrices. During a single optical convolution step only $2 \times 2 = 4$ compared to $100 \times 100 = 10{,}000$ calculations are executed. Therefore, the performance is expected to lag by approximately 3 orders of magnitude. Consequently, the Fourier transform properties do not offer significant benefits and the OVMM approach will be pursued instead. The convolution process can be implemented conventionally

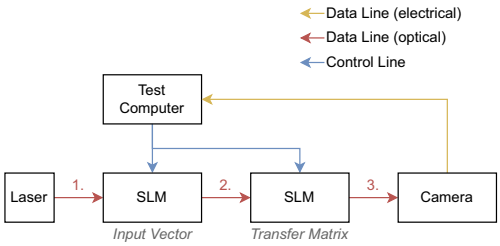


Fig. 1 Experimental free-space configuration with two Spatial Light Modulators (SLMs)

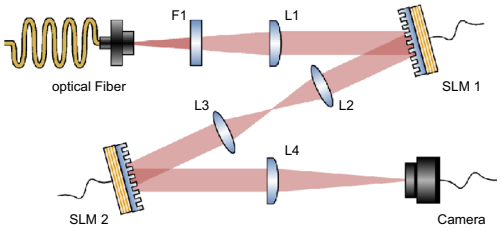


Fig. 2 Experimental free-space setup based on the Vector-Matrix Multiplier proposed by Spall et al. [11]. Components: Absorption filter ($F1$), collimation lens ($L1$), convex lens ($L2$ & $L3$), cylindrical lens ($L4$), Spatial Light Modulator (SLM)

using the sequential sliding-window approach, while each filter multiplication is performed optically. Moreover, the OVMM approach extends beyond convolution and can also be applied to other calculations, such as the fully connected layer in the desired CNN.

In total, two different free-space configurations are planned. The first SLM implementation is used to demonstrate the general feasibility of hybrid (conventional/optical) calculations for ML. The second and integrated approach shall further show its applicability as a spacecraft payload.

3.1 SLM implementation

The overall concept of the first optical free-space experiment is depicted in Fig. 1. It consists of a laser, 2 SLMs, a Complementary Metal-Oxide-Semiconductor (CMOS) camera, and a test computer. For the signal lines, we differentiate between optical data (red), electrical data (yellow), and electrical control/configuration (blue) lines.

First, the laser provides a non-modulated light beam as the signal source at stage 1. The beam is reflected off the first SLM while the input vector is encoded onto the beam. The now modulated signal at stage 2 is reflected off the second SLM that encodes the transfer matrix. The optical superposition of both modulation processes results in a signal equal to the multiplication of the input vector and the transfer matrix at stage 3. This signal is captured by the camera and fed back to the test computer. The SLMs are controlled by the test computer to configure both the input vector and the transfer matrix of the VMM.

Figure 1 neglects components such as lenses crucial for generating the input vector and summation during the Vector-Matrix Multiplication (VMM). A more detailed schematic of the optical components is illustrated in Fig. 2.

The laser beam is guided into the system via an optical fiber and its intensity is reduced using an absorption filter $F1$. The diverging light is then collimated using a collimation lens $L1$. The distance from the optical fiber to the lens $L1$ is important for beam expansion, the higher the distance, the larger the diameter of the beam. The width of the beam determines the usable area on the SLMs in subsequent steps. The beam is reflected by the first SLM which imprints the desired input vector. The individual values of the input vector are modulated as light intensity in a specific region of the beam. The plane of the first SLM is imaged by two successive biconvex lenses $L2$ and $L3$ onto the next SLM. At the second SLM, the beam is reflected again, and the beam intensity is modulated based on a transfer matrix. The information encoded in the light beam already corresponds to a partial result matrix, composed of the multiplied components of the input vector and the transfer matrix. Using a cylindrical lens $L4$, the beam is focused into a line and captured by the camera. This focus superimposes the rows of the result matrix, thus corresponding to the summation of the rows. Finally, the line captured by the camera is the result of the Vector-Matrix Multiplication.

Figure 3 illustrates the Optical Vector–Matrix Multiplication concept for a total of 4 input values. For simplicity, this graphic does not display all the optical components that are not required for the mathematical operation. The areas in the beam for modulating the input vector $\vec{x}$ and the transfer matrix A must be of the same size. The input vector is treated as a row vector during the optical computation. To fully cover the transfer matrix with elements from the input vector, duplicate rows are added to the input vector. This results in the illustrated 4×4 matrix encoded by the first SLM. After the light propagates through the second SLM, the spatial profile of the beam is equivalent to the element-wise products of $\vec{x}$ and A. The cylindrical lens focuses the beam into a line while performing a Fourier transform in that direction. The thin line is equal to the summation of the previously individual products. Finally, the camera captures the result vector $\vec{y}$ of the finished OVMM.

3.2 Integrated implementation

In principle, the implementation procedure is similar to the previous setup described in Sect. 3.1. However, to potentially integrate optical computation into a satellite payload, the input SLM is replaced by better suited hardware. The

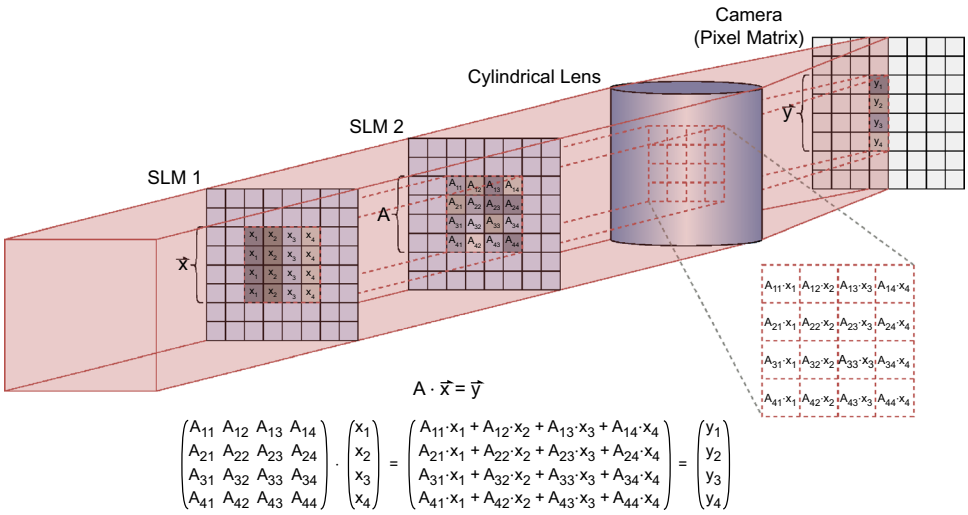
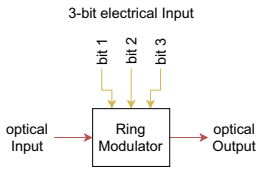
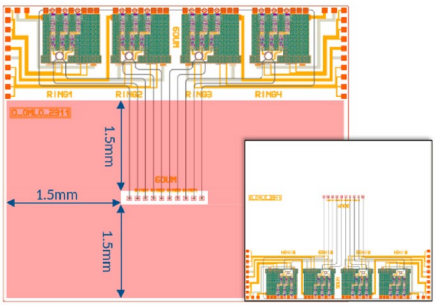


Fig. 3 Concept of the Optical Vector–Matrix Multiplication with two SLMs, a cylindrical lens and a camera. Exemplary for a vector limited to 4 entries. Based on the concept proposed by Spall et al. [11]



(a) Concept of a single channel ring modulator.



(b) PCB layout of ring modulator group consisting of four individual ring modulators by IHP.

Fig. 4 Schematic and PCB layout of the ring modulators/ Optical Digital-to-Analog Converters (ODACs) designed and produced by The Leibniz Institute for High Performance Microelectronics (IHP)

SLM responsible for the modulation of the transfer matrix continues to be used until a better alternative is found. The input vector will be modulated by silicon-based ring modulators produced by IHP. The following two graphics in Fig. 4 show the conceptual design of these ring modulators.

As depicted in Fig. 4a, a ring modulator consists of an electrical input, an optical input, and an optical output. The optical signal passes through the ring modulator, modulating its intensity based on the electronic input. The input resembles a 3-bit parallel interface, where the modulated light intensity is proportional to the configured 3-bit input value. The 3-bit interface was selected to implement Pulse-Amplitude Modulation 2 (PAM2) or PAM4 encoding, and possibly even PAM8. The modulators are grouped in sets of four and manufactured on silicon dies displayed in Fig. 4b. This group size is the limiting factor for the size of the input vector and the transfer matrix. In principle, larger chips with more ring modulator channels could be manufactured. However, the production constraints are cost, measurability, and assembly limitations. The silicon die area available to us only accommodated a variant featuring four modulators. The hardware incorporates digital-to-analog as well as electronic-to-optical signal conversion, giving it the name Optical Digital-to-Analog Converter (ODAC). The combination of these two properties within a single component is a groundbreaking innovation that significantly contributes to a more compact design. The circuit is implemented in a fully analog fashion and is currently expected to exceed modulation speeds of 10 GHz, making it ideal for high-speed optical computing. To provide a high-speed data stream to the 3-bit parallel interface of each ODAC, we require dedicated hardware. It needs to handle and buffer incoming data and perform data processing for signal generation of the parallel control lines. A specifically developed payload or Electronic Ground Support Equipment (EGSE) board described in Sect. 4 will be used for that purpose.

Figure 5 outlines the overall concept of the partially integrated free-space setup. A laser provides a non-modulated light beam as the signal source at stage 1. The ODACs located on the EGSE hardware encode the input vector onto the beam. Similar to the previous configuration, the now modulated signal at stage 2 is reflected by the SLM that encodes the transfer matrix, thereby performing the multiplication. The result at stage 3 is subsequently captured by a camera. The input and result data

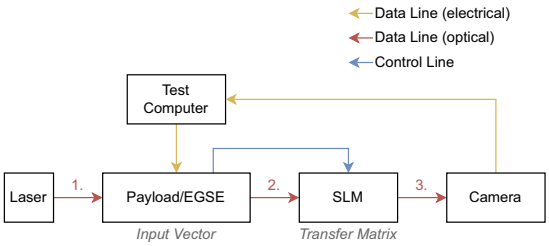


Fig. 5 Experimental, partially integrated configuration with ring modulators/ODACs and payload/EGSE

are still managed by a computer. In contrast, the ODACs and the SLM are controlled by the EGSE directly.

The experimental free-space setup for the proposed concept is detailed in Fig. 6. It consists of a laser array that provides four separate unmodulated light beams. Using optical fibers, they are fed into the ODACs where the input vector is encoded. Each vector component is encoded by a separate ODAC. The modulated light intensity is equivalent to its 3-bit parallel input value supplied by the EGSE. After the input vector has been encoded, the optical signals exit the ODAC via optical fibers. A fiber array terminates the fibers and subjects the beams to the free-space environment. The beams are then focused into a single beam by an aspheric lens $L1$ and reflected by a SLM. During the reflection, the spatial profile of the beam is modulated based on a transfer matrix. From this point on, the procedure is equivalent to the pure SLM implementation in Sect. 3.1. The beam is focused into a line by a cylindrical lens $L2$ and the result is captured by a camera.

To increase data throughput in the future, the camera will be replaced by high-speed photodiodes.

4 Onboard integration

Intending to create an ANN for an application in orbit, a concept for onboard integration is to be designed. The EGSE was already referenced in the context of the free-space setup in the previous Sect. 3.2. A detailed design and the connection to a suitable satellite bus have yet to be addressed further.

4.1 Payload concept

Several satellite missions have already been developed and launched by the Department of Space Technology (RFT) at the Technical University of Berlin (TU-Berlin). Among other satellite buses, the TU-Berlin innovative neXt generation satellite bus (10 kg) (TUBiX10) platform is particularly suitable for an OMLO payload.

It was originally designed for the S-Band communication network with distributed nano satellites (S-NET) mission [24], and was later adapted to host other payloads for the Spectrum AnaLysis SATellite (SALSAT) [18], and On-Orbit Verification Cube Satellite (OOV-Cube) [25] missions. A rendering of the TUBiX10 bus used for the S-NET mission is displayed in Fig. 7. It has a cubic design with dimensions of $240 \times 240 \times 240\ \text{mm}^3$ and a maximum mass of 10 kg. It can therefore be categorized as a nanosatellite. Due to its successful usage in multiple missions, the TUBiX10 bus can be deemed a trusted and flight-proven system.

The spin-off company Rapid Cubes GmbH (RC) from the RFT has licensed the TUBiX10 satellite bus and offers both its development and the development of dedicated payloads. We therefore aim to use RC for the production of our payload/EGSE hardware. This will

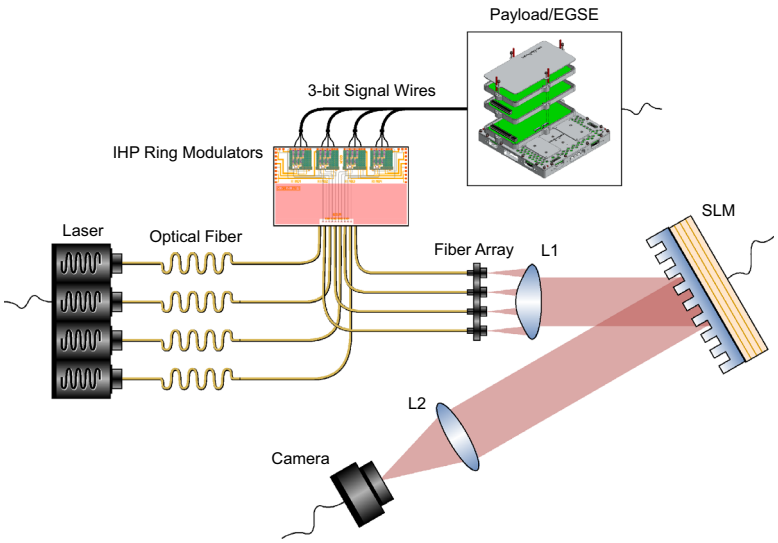
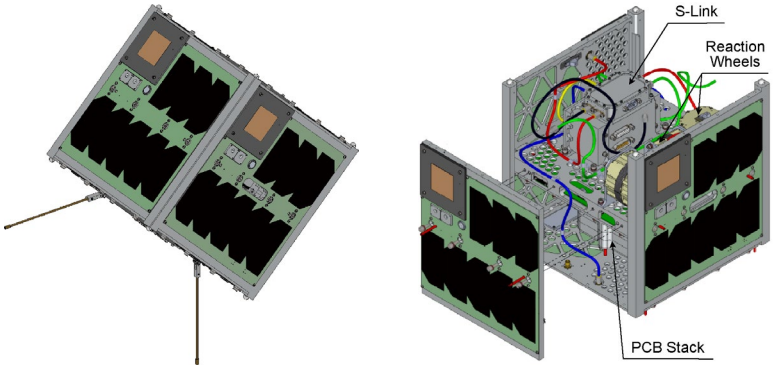


Fig. 6 Partially integrated free-space setup with ring modulators/ODAC from IHP derived from the concept (Fig. 2) in Sect. 3.1. Components: Laser Array, Ring Modulators, Fiber Array, Aspheric Lens ($L1$), SLM, Cylindrical Lens ($L2$), Camera



(a) Computer render of S-NET satellite. (b) S-NET satellite interior explosion.

Fig. 7 TUBiX10 satellite bus used by the S-NET mission

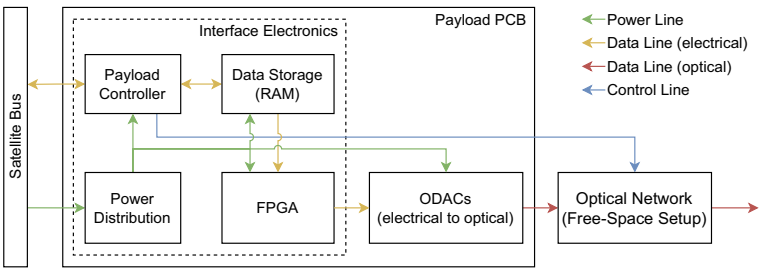


Fig. 8 Concept for the integration of the ODACs into a TUBiX10 payload

ensure compatibility with the TUBiX10 satellite bus for a future mission.

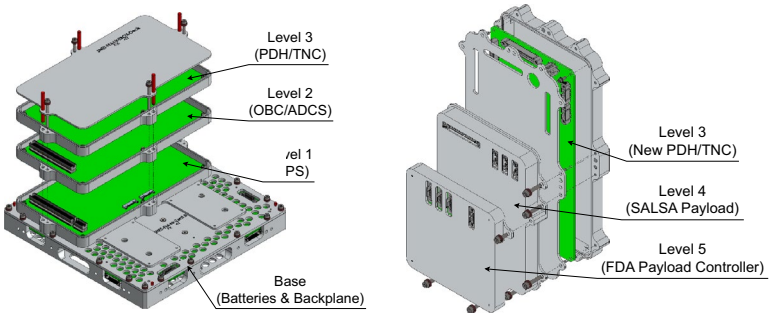
An overview of the conceptual payload is shown in Fig. 8. Inside the payload, we differentiate mainly between interface electronics and peripherals. The payload will be connected to the satellite bus via data and power lines.

4.2 Electronic Interface

The interface electronics of the payload PCB consist of a payload controller, data storage, Field Programmable Gate Array (FPGA) and Power Distribution Unit (PDU) as outlined in Fig. 8. The controller is responsible for exchanging data with the satellite bus. Data intended for the OVMM is processed and temporarily stored in external data storage, such as Random-Access Memory (RAM). A FPGA retrieves the data from the data storage and encodes it as a 3-bit parallel interface signal for the downstream ODAC. It is important to note that both data storage and FPGA have to operate at high speeds to allow the operation of OVMM components in the GHz range. In addition to managing satellite bus data, the payload controller manages the OVMM components. Through control lines, it configures the free-space devices by, for example, changing the weights of the transfer matrix SLM. Another part of the interface electronics is the power supply. It provides all necessary voltages for both the components of the interface electronics and those of the peripherals.

4.3 Peripherals

For the peripherals, a distinction between the Optical Digital-to-Analog Converter (ODAC) and the free-space components of the OVMM is made. Illustrated in Fig. 8, the ODACs are planned to be located on the payload PCB close to its dedicated interface electronics. They are produced as System-on-Modules (SoMs) and are therefore fully integrable into our proposed payload stack. The FPGA provides a 3-bit parallel interface signal to each ODAC. The ODACs modulate an optical signal based on an electrical input from the interface FPGA. There will be a total of four ODACs on a SoM, resulting in 12-bit parallel interface input signals and four optical output signals. The modulated optical signals are directed to the subsequent OVMM for further processing. The processed data or the result vector of the OVMM is captured by photodetectors and manually evaluated with laboratory measuring instruments. A laboratory computer will also store the result data for data analysis. Components regarding the OVMM exist exclusively as a free-space setup. Currently, they cannot be integrated into the payload board due to their size and complexity. The feedback of the output signal is planned to be captured by laboratory equipment only and will not yet be implemented as part of the payload.



(a) S-NET PCB Stack (b) SALSAT PCB Stack

Fig. 9 TUBiX10 payload stacks for S-NET and SALSAT missions

4.4 Integration into satellite-bus

The payload board is designed to be TUBiX10-compatible, potentially for use as a payload in future satellite missions. Figure 9 shows the PCB stack of the TUBiX10 satellite bus of the S-NET and SALSAT mission. The levels 1 and 2 of the stack are reserved for the Electrical Power System (EPS), the On-Board Computer (OBC) and the Attitude Determination and Control System (ADCS). Level 3 houses the Payload Data Handler (PDH) for S-NET or the SBand Terminal Node Controller (TNC) for SALSAT. An additional payload can be mounted on this stack as level 4. This was, however, contingent upon modifications to level 3, which was originally not intended for the passage of power and signal lines to further levels.

Figure 9b illustrates the PCB modification required for the SALSAT mission, which was based on a flight spare of the S-NET mission. The new level 3 adapted the old S-NET PDH to a new SALSAT SBand TNC that allows additional payloads. The Spectrum Analysis of LEO Satellite Allocations (SALSA) payload [26] has been added as level 4 and the controller electronics for a Fluid Dynamic Actuator (FDA) payload as level 5 of the stack. Levels 1 and 2 remained unchanged. For OMLO, the desired payload is intended to be integrated in the same manner. Since both the payload and the satellite bus are developed by the same subcontractor Rapid Cubes GmbH, we assume that the level connectors are compatible without any additional modifications. The connection to the structure, power supply, and data linkages are thus ensured.

Thermal integration of the payload board will be achieved through a dedicated layer in the PCB. Copper pads at the PCB's fastening points ensure contact of this layer with the satellite's mass.

Another aspect that is usually considered during integration is radiation resistance. Conventional components such as controller, FPGA, and the ODAC Integrated Circuits (ICs) will be tested under the influence of radiation once they are fully integrated. The optical components are, however, not susceptible to radiation and special shielding or similar measures can be omitted for these components. In contrast, the electronic components within the optical devices remain at risk from radiation and may introduce errors. This will also be addressed as part of the evaluation process.

5 Data processing

The intended hybrid computing system will initially do most of its calculations on conventional hardware. Some specific calculations regarding convolution or NN-related Vector–Matrix Multiplication will be outsourced to the optical hardware. The number of optical calculations is planned to be increased over time. All AI related tasks are planned to be pre-trained on conventional hardware and transferred to the optical components for their execution. AI training on optical hardware can also be considered for future use cases, but will not be practiced in the scope of this paper. Due to the hardware limitation of only 4 optical input channels, input vector sizes must be reduced to 4×1. The transfer matrix implemented by an SLM, as described in Sect. 3.1, is indirectly limited to a size of $n \times 4$ by the input vector. The conversion and preparation of data for these limitations play a significant role in the convolution and dense layer of the proposed AI system.

5.1 Convolution layer

Figure 10 illustrates the data processing steps of the convolution layer. A $n \times n$ input image is split into 2×2 overlapping chunks based on the sliding-window approach. The chunks are then flattened into 4×1 vectors and optically multiplied with a pre-trained 4×4 convolutional matrix. Each row of this matrix represents a separate filter or kernel. The resulting output vectors of this operation contain the convolved values for each filter applied to a specific chunk. As the final step of this layer, the filtered vectors are forwarded to the next, fully connected layer. Restructuring the output vectors into a human-readable matrix is unnecessary given that the fully connected layer will be trained on the raw filtered output.

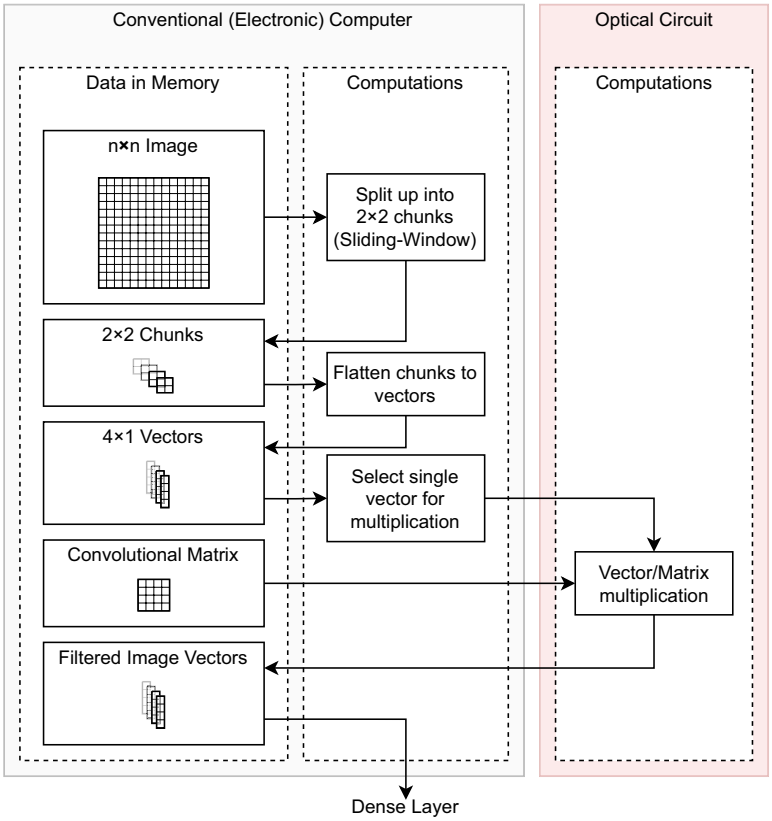


Fig. 10 Process diagram for the convolution of an image

The convolutional matrix containing the filter can be reused for all image chunks, partially eliminating the need for a reconfiguration of the optical components. If, however, more than 4 filters shall be used for the convolution, the rather slow SLM does need to be reconfigured. This might result in a theoretical bottleneck depending on its update rate. Another limitation of the input vector is the 3-bit value range per vector element.

The limitation is also introduced by the hardware of the IHP ODAC chip. This means that the entirety of our input can only have $4 \times 3 = 12$ bit of information. For a single convolutional layer, this limitation does not pose a challenge. However, for an additional convolutional layer, $2 \times 2 \times 4 = 16$ inputs are required. If the previous 4×4 convolutional matrix shall be reused, the number of inputs must be reduced. By quantizing the filtered outputs from the previous layer and using three instead of four filters, the information limit can be attained.

For our Proof-of-Concept (POC), only a single convolutional layer will be used. This number will, however, increase over the duration of our experiments.

5.2 Dense layer

The dense layer represents the fully connected network of the proposed CNN. Similar to the convolution layer, the dense matrix has to be split into smaller, computable chunks. The process for computing the Vector–Matrix Multiplication related to the dense layer is outlined in Fig. 11.

First, the dense matrix is divided into a block matrix based on the divide-and-conquer algorithm for Matrix-Matrix Multiplication by Volker Strassen [27]. All blocks contain a 4×4 matrix. Each of these submatrices is then multiplied with a Filtered-Image-Vector (FIV) generating an intermediate result vector for each submatrix/FIV combination. This VMM is again performed on the optical hardware. The intermediate vectors are combined to construct 4×4 matrices that can be multiplied with a 1-vector (all elements are equal to 1). This step is equivalent to the summation of the individual vectors. The result vectors are fed back into the pool of intermediate vectors to repeat the summation until only a single vector remains for each row. The resulting vectors can then be appended to yield the raw output of the VMM. By applying a MinMax (MM) function to the raw output, a probability vector is calculated from which the final prediction can be derived.

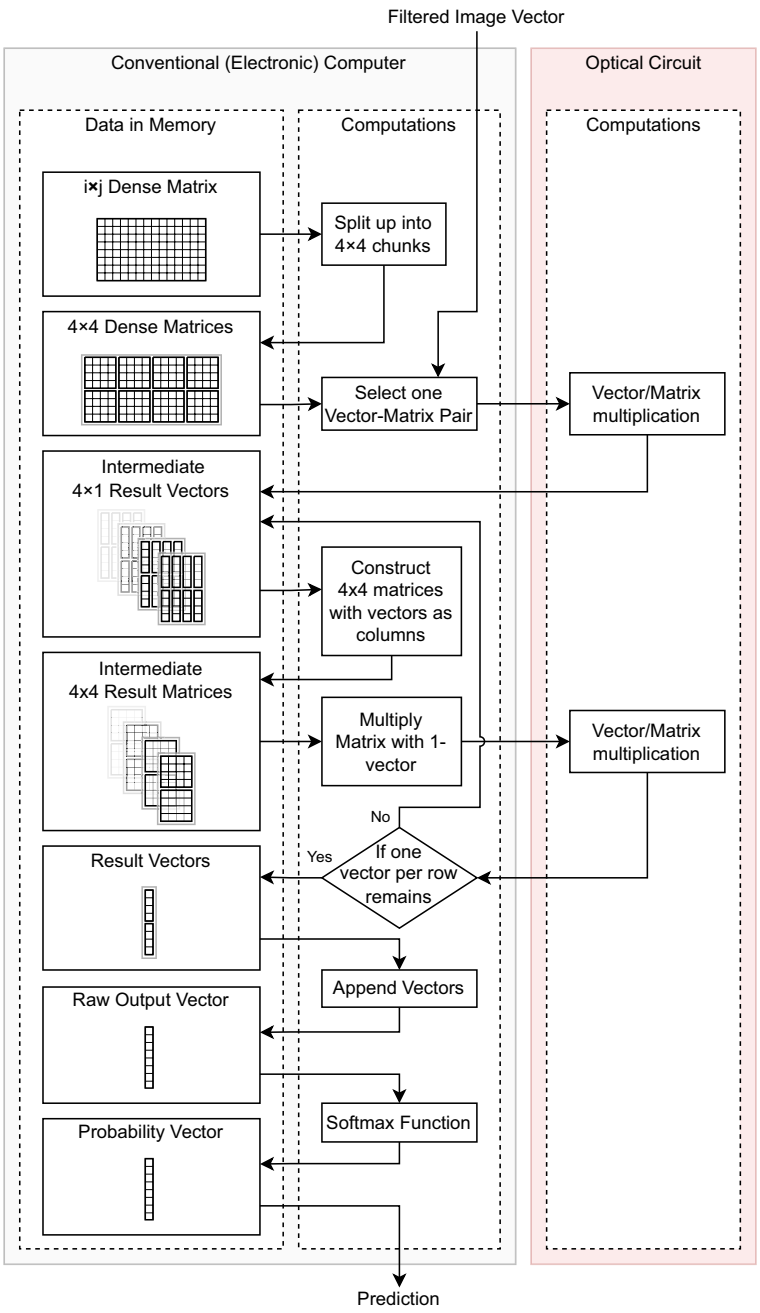


Fig. 11 Process diagram for the dense layer of the Convolutional Neural Network (CNN)

6 Preliminary results

During the conceptual phase, we were unable to conduct extensive experiments on the proposed system. Preliminary experimental setups and simulations can nonetheless provide valuable insight into the concept's feasibility.

Using a SLMs configuration similar to the setup described in Sect. 3.1, a 4×4 OVMM was performed. Each matrix entry was discretized to 4 bits and the vector values were represented as single-bit values for simplicity. The first results show an accuracy of 84.5 %, demonstrating the plausibility of the OVMM, while leaving room for further improvements.

The data processing approach proposed in Sect. 5 was simulated using the software TensorFlow, with all optical calculations carried out internally on conventional hardware. Training was performed using the MNIST dataset over 5 epochs. With a 3-bit discretized vector and matrix, we achieved an accuracy of 91.58 %.

These results highlight the promising potential of our concept. However, the combination of both technologies still presents challenges that are not addressed by these results.

7 Conclusion

The proposed satellite payload concept for an optical computing system offers a new approach to an AI system in a space environment. The hybrid design will bridge the gap between conventional and optical components. Although the final goal is a fully optical computing system, it has to be integrable into a common satellite bus for now. With ongoing development, the number of optical components can be increased until an entirely optical system is realized.

An optical system can offer high computing performance in relation to its power consumption. Feldmann et al. were able to demonstrate Tera-Multiply-Accumulate per Second (TMAC/s) speeds with their photonic tensor core using only about 20 mW optical power output [28]. Modern spacecraft CPU such as the Quad-Core LEON4FT processer developed by the European Space Agency (ESA) can theoretically provide up to 1 GFLOPS at 5 W [29]. Assuming two operations per MAC the speed difference between these two systems is evident. The 20 mW demonstrated by Feldmann does not consider the power consumption of the electronic equipment. The total power consumption of the photonic tensor core is therefore much higher. Miniaturization and efficiency optimization are still important fields of research.

Yet, it can be emphasized that the potential of optical computing scales with the size of the matrix. The performance increases quadratically, while energy consumption scales linearly, making it more efficient for large matrices. Given a constant matrix, multiple vectors can also be transmitted simultaneously at different frequencies. In this case, performance increases in proportion to the number of wavelengths used. As a result, the system's potential improves significantly for larger matrices. Thus, there will be an unavoidable pivot point at which the number of parallel computations surpasses that of conventional systems in relation to their power consumption.

In our system, the computing speed is theoretically limited by the update speed of the SLMs, which takes approximately 0.1 ms. The vector encoding SLM will be replaced by the ring modulators that operate in the GHz range, leaving the matrix encoding SLM as a bottleneck. Since SLMs are also not suitable for integration into a satellite bus, the utilization of a Mach-Zehnder Interferometer (MZI) instead of the second stage SLM should be considered. In addition, this change in hardware might also improve the total computation speed. This leaves the camera recording the OVMM result as the speed-limiting component. In later implementations, it will be replaced by high-speed photodiodes to further improve the calculation speed. With a transition to faster multiplication and result stages, the speed of the payload's electrical components will become the main speed constraint of the hybrid system.

An all-optical setup could be implemented but would have to be specifically designed for a single dedicated use case. Our approach uses a single Optical Vector-Matrix Multiplication stage that will be reused for all calculations. It is therefore highly reconfigurable and universal in its application. On the downside, it creates a type of delay or clock cycle introduced by the electronic hardware for data distribution and storage.

At this early conceptual stage, a flight-ready implementation is not immediately feasible. The goal is merely a Proof-of-Concept. The preliminary experimental results in Sect. 6 can nonetheless serve as an indication of the concept's plausibility. Following the implementation of the proposed system, extensive testing and evaluation will be conducted to identify any necessary changes to the concept and its hardware. This phase will involve optimization and redesign based on efficiency, reliability, and compatibility with other system components. Once these necessary modifications are finalized, we will begin seeking suitable cooperation partners, both from industry and academic institutions, to further advance the development process. Subsequent to the optimization and redesign, the miniaturization and integration of the optical circuit is planned. This will involve refining the design to reduce the size without compromising performance and functionality.

Additionally, a feedback mechanism for the optical output signal will be incorporated into the system design. As a result, a fully self-contained system will be realized, with the potential to evolve into a flight-ready payload in the future.

The power efficiency of the system will also improve with further development. In our current free-space setup, energy consumption is neglected for the time being.

Based on these developments, we aim to establish a common usage of optical computing systems within the space segment. The potential is enormous and must be explored and utilized.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The only data recoreded within the scope of this paper was disclosed in the "Preliminary results" section (6). There is no additional data.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Newell, A., Simon, H.A., Shaw, C.: The logic theory machine: a complex information processing system. IRE Trans. Inf. Theory **2**(3), 61–79 (1956). https://doi.org/10.1109/TIT.1956.1056797
2. European Parliament's Committee on Industry, Research and Energy: Opportunities of Artificial Intelligence (2022). https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652713/IPOL_STU(2020)652713_EN.pdf
3. Jonsson, A., Morris, R.A., Pedersen, L.: Autonomy in space: current capabilities and future challenge. AI Mag. (2007). https://doi.org/10.1609/aimag.v28i4.2066
4. Bernard, D.E., Gamble, E.B., Rouquette, N.F., Smith, B., Tung, Y.-W., Muscettola, N., Dorias, G.A., Kanefsky, B., Kurien, J., Millar, W., et al.: Remote agent experiment ds1 technology validation report. Ames Research Center and JPL, pp. 8–9 (2000)
5. Chien, S., Sherwood, R., Tran, D., Cichy, B., Rabideau, G., Castaño, R., Davis, A., Mandl, D., Frye, S., Trout, B., Shulman, S., Boyer, D.: Using autonomy flight software to improve science return on earth observing one. J. Aerosp. Comput. Inf. Commun. (2005). https://doi.org/10.2514/1.12923
6. Ai-Chang, M., Bresina, J., Charest, L., Chase, A., Hsu, J.C.-J., Jonsson, A., Kanefsky, B., Morris, P., Rajan, K., Yglesias, J., Chafin, B.G., Dias, W.C., Maldague, P.F.: MAPGEN: mixed-initiative planning and scheduling for the mars exploration rover mission. IEEE Intell. Syst **19**(1), 8–12 (2004). https://doi.org/10.1109/MIS.2004.1265878
7. Estlin, T.A., Bornstein, B.J., Gaines, D.M., Anderson, R.C., Thompson, D.R., Burl, M., Castaño, R.C., Judd, M.: Aegis automated science targeting for the MER opportunity rover. ACM Trans. Intell. Syst. Technol. (TIST) (2012). https://doi.org/10.1145/2168752.2168764
8. Francis, R., Estlin, T., Doran, G., Johnstone, S., Gaines, D., Verma, V., Burl, M., Frydenvang, J., Montaño, S., Wiens, R.C., Schaffer, S., Gasnault, O., DeFlores, L., Blaney, D., Bornstein, B.: Aegis autonomous targeting for ChemCam on mars science laboratory: deployment and results of initial science team use. Sci. Robot. **2**(7), 4582 (2017). https://doi.org/10.1126/scirobotics.aan4582
9. NASA: SuperCam gains new artificial intelligence capabilities with AEGIS update (2023). https://mars.nasa.gov/mars2020/mission/status/446/supercam-gains-new-artificial-intelligence-capabilities-with-aegis-upgrade/
10. McMahon, P.L.: The physics of optical computing. Nat. Rev. Phys. **5**, 717–734 (2023). https://doi.org/10.48550/arXiv.2308.00088
11. Spall, J., Guo, X., Barrett, T.D., Lvovsky, A.I.: Fully reconfigurable coherent optical vector-matrix multiplication. Opt. Lett. **45**(20), 5752 (2020). https://doi.org/10.1364/OL.401675
12. Shen, Y., Harris, N.C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., Sun, X., Zhao, S., Larochelle, H., Englund, D., Soljačić, M.: Deep learning with coherent nanophotonic circuits. Nat. Photonics **11**, 441–446 (2017). https://doi.org/10.1038/nphoton.2017.93
13. Goodman, J.W., Dias, A.R., Woody, L.M.: Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms. Opt. Lett. **2**(1), 1–3 (1978). https://doi.org/10.1364/OL.2.000001
14. Yang, M., Robertson, E., Esguerra, L., Busch, K., Wolters, J.: Optical convolutional neural network with atomic nonlinearity. Opt. Express **31**(10), 16451–16459 (2023). https://doi.org/10.1364/OE.490070
15. Caulfield, H.J., Dolev, S.: Why future supercomputing requires optics. Nat. Photonics (2010). https://doi.org/10.1038/nphoton.2010.94
16. Barschke, M., Yoon, Z., Brieß, K.: TUBiX—the TU Berlin innovative next generation nanosatellite platform. In: International Astronautical Congress (Vol. 64), Beijing, China (2013). https://www.researchgate.net/publication/262006877_TUBiX_-_The_TU_Berlin_Innovative_Next_Generation_Nanosatellite_Platform
17. Barschke, M., Baumann, F., Ballheimer, W., Gordon, K., Nitzschke, C., Brieß, K.: TUBiX20—the novel nanosatellite bus of TU Berlin. In: IAA Symposium on Small Satellites for Earth Observation (Vol. 9), Berlin, Germany (2013). https://www.researchgate.net/publication/262006892_TUBiX20_-_The_Novel_Nanosatellite_Bus_of_TU_Berlin
18. Großhans, J., Huu, Q.V., Wüstenberg, P., Pust, M., Lange, S., Balke, A., Stoll, E.: SALSAT: first year in orbit—preliminary assessment of the nanosatellite bus, subsystems, and the payloads. In: 72nd International Astronautical Congress (IAC), Dubai, United Arab Emirates (2021). https://www.researchgate.net/publication/369453144_SALSAT_First_year_in_orbit_-_Preliminary_assessment_of_the_nanosatellite_bus_subsystems_and_the_payloads
19. Baumann, F., Ballheimer, W., Gordon, K., Nitzschke, C., Brieß, K.: TUBIN—a novel nanosatellite bus with uncooled infrared imager payload. In: German Aerospace Congress (Vol. 61), Berlin, Germany (2012). https://www.researchgate.net/publication/262103495_TUBIN_-_A_Novel_Nanosatellite_Bus_with_Uncooled_Infrared_Imager_Payload
20. Roychowdhury, D., Kapitola, S., Boyer, F., Legenza, A., Diez, J., Palmera, B., Kübler, F., Freymutha, J., Grau, S., Stoll, E.: Lessons learned from the initial operations phase in the nanoff cubesat formation flight mission. In: International Astronautical Congress 2024, Milan, Italy (2024). https://www.researchgate.net/publication/385343550_Lessons_Learned_from_the_Initial_Operations_Phase_in_the_NanoFF_CubeSat_Formation_Flight_Mission
21. TU-Berlin, Chair of Space Technology: AITHER—Hochintegrierte OnBoard-Datenverarbeitungseinheit für KI-Anwendungen auf Kleinsatelliten (2022). https://www.tu.berlin/raumfahrttechnik/forschung/aktuelle-projekte/aither
22. LeCun, Y., Cortes, C., Burges, C.J.C.: THE MNIST Database of handwritten digits (1998). https://yann.lecun.com/exdb/mnist/
23. Gene, J., Park, S., Shin, H.C., Sohn, J.M.: Hybrid optical convolutional neural network with convolution kernels trained in the spatial domain. Neurocomputing **573**, 127251 (2024). https://doi.org/10.1016/j.neucom.2024.127251
24. Yoon, Z., Frese, W., Briess, K.: Design and implementation of a narrow-band intersatellite network with limited onboard resources for IoT. Sensors **19**, 4212 (2019). https://doi.org/10.3390/s19194212
25. ESA: OOV-Cube—on-orbit verification cube satellite (2024). https://www.esa.int/Enabling_Support/Space_Transportation/Ariane/Ariane_6_flies_OOV-Cube_Internet_of_wild_Things

26. Grosshans, J., Buscher, M., Vu, H.Q., Balke, A., Maaß, A., Lohse, A.: SALSA—a novel spectrum analyzer board for LEO satellite allocations based on SDR technology. In: 2018 AIAA SPACE and Astronautics Forum and Exposition, Orlando, USA (2018). https://doi.org/10.2514/6.2018-5283
27. Volker, S.: Gaussian elimination is not optimal. Numerische Mathematik **13**, 354–356 (1969). https://doi.org/10.1007/BF02165411
28. Feldmann, J., Youngblood, N., Karpov, M., Gehring, H., Li, X., Gallo, M., Fu, X., Lukashchuk, A., Raja, A., Liu, J., Wright, D., Sebastian, A., Kippenberg, T., Pernice, W., Bhaskaran, H.: Parallel convolution processing using an integrated photonic tensor core. Optica **7**, 1812–1819 (2020). https://doi.org/10.48550/arXiv.2002.00281
29. Jalle, J., Hjorth, M., Andersson, J., Fossati, R.W.L.: DSP benchmark results of the GR740 rad-hard quad-core leon4ft. In: Analogue and Mixed-Signal Integrated Circuits for Space Applications, Gothenburg, Sweden (2016). https://indico.esa.int/event/102/contributions/62