PHOTONICS21

Tematys
Exploration of photonics markets

Photonics for High Performance Computing (HPC)

Published by:
Photonics21 – European Technology Platform

Project coordinator
Photonics21 Secretariat
c/o VDI Technologiezentrum GmbH
VDI Platz 1, 40468 Düsseldorf – Germany
Mail: secretariat@photonics21.org
Website: www.photonics21.org
X: twitter.com/Photonics21
Linked-in: linkedin.com/company/Photonics21

Katharina Flaig-Rüttgers, VDI Technologiezentrum GmbH
Sylvie Rijkers-Defrasne, VDI Technologiezentrum GmbH

TEMATYS – 6, cité de Trévise
75009 Paris – France
www.tematys.com

Benoît d’Humières
Thierry Robin

Cover graphics:
Photonics21 / Ocean

Design and layout:
Steven Randall, Ocean
Studio 2.6 , The Leathermarket,
11–13 Weston Street, London SE1 1 ER
www.ocean-design.com

Editorial support:
Sam Young, Matter PR
Clarendon House, 52 Cornmarket Street
Oxford, UK, OX1 3HJ
www.matterpr.com

Brussels / Düsseldorf / Paris 2024

Photonics21 has received financial support from the European Commission in the scope of the Coordination and Support Actions “Boosting Europe’s Sovereignty in Technology by driving Photonics from Research to Market – Photonics21” – Grant Number 101016520 and „Photonics from Research to Market – Empowering Europe’s strategic autonomy, supporting the Green Deal and securing resilience – Grant Number 101134961.

Photonics for
High Performance Computing (HPC)

Contents

1. HPC: Operation and architecture

Introduction

High-performance computing (HPC) refers to computing systems with extremely high computational power that are able to solve immensely complex and demanding problems.[1]

HPC involves the use of supercomputers and parallel processing techniques to solve complex computational problems. The term is most commonly associated with computing used for scientific research or computational science but also refers to engineering applications with high computing needs (such as computational fluid dynamics and the building and testing of virtual prototypes).

HPC is key to Europe's future prosperity, digital transformation and resilience. With €7 billion in funding, the European Commission is strengthening investments in supercomputing. The Commission aims to build up supercomputing and data processing capacities by buying world-class supercomputers and supporting an ambitious HPC research and innovation agenda.

Figure 1: MareNostrum5 supercomputer
Source: https://eurohpc-ju.europa.eu/

In December 2023, MareNostrum 5, the latest world-class European supercomputer, was inaugurated in Barcelona, Spain. Currently ranked fifth in the world, it is hosted at the Barcelona Supercomputing Centre and will be accessible to a wide range of European scientific and industry users starting in March 2024. MareNostrum 5 has a peak performance of 314 Petaflops. It is also one of the most energy-efficient supercomputers in Europe.

Performances

The capability of high-performance computers is commonly measured in floating-point operations per second (FLOPS) instead of million instructions per second (MIPS). Since 2017, HPC has performed over 10^{17} FLOPS (a

[1] https://digital-strategy.ec.europa.eu/en/policies/high-performance-computing

hundred quadrillion FLOPS, 100 petaFLOPS or 100 PFLOPS). For comparison, a desktop computer performs hundreds of gigaFLOPS (10^{11}) and tens of teraFLOPS (10^{13})[2].

A list of the most powerful high-performance computers can be found on the TOP500 list[3], which ranks the world's 500 fastest high-performance computers, as measured by the High-Performance LINPACK (HPL) benchmark. Not all computers are listed, either because they are ineligible (e.g. they cannot run the HPL benchmark) or because their owners have not submitted an HPL score. The TOP500 list is updated twice a year, in June and in November. In November 2023, the total combined performance of all 500 high-performance computers exceeded the Exaflop barrier with seven exaflops (Eflops), up from 5.2 exaflop/s (Eflops) 6 months ago. Frontier remains the No. 1 system in the TOP500. This HPE Cray EX system is the first system to have a performance exceeding one Exaflop/s. It is installed at the Oak Ridge National Laboratory (ORNL) in the US, where it is operated by the Department of Energy (DOE).

Architectures

HPC systems are characterised by their ability to perform high-level computations at high speeds, involving a large number of processors working in parallel. There are mainly two types of architecture for HPC:

- The most commonly used is **Cluster Computin**g: A cluster is a group of standalone computers connected through a local area network (LAN), working together as a single, integrated computing resource on the same set of tasks.
- The other is **Massively Parallel Processing (MPP)**: In this configuration, nodes are arranged to execute commands or calculations in parallel. Much like processing by graphics processing units (GPU), the idea is to scale data processing and computation volume across a system. It is suitable for large-scale simulations and complex computations requiring high parallelism. The MPP architecture accounts for **12% of the top 500 but around 50% of total performance** (in FLOPS).

These are architectures where all the computers (nodes) are located in the same place. This is not to be confused with a distributed architecture that requires a vast amount of computing resources scattered geographically. The distributed architecture used in this case is generally referred to as grid computing, and each node (computer) may be located remotely and operate independently. It is, therefore, a network of computers that are used when they are otherwise inactive.

A number of special-purpose systems have been designed and dedicated to a single problem. This allows the use of integrated circuits such as specially programmed Field-Programmable Gate Array (FPGA) chips or custom Application-Specific Integrated Circuits (ASICs), allowing better price/performance ratios by sacrificing generality. Examples include Belle, Deep

[2] GPU Database: https://www.techpowerup.com/gpu-specs/
[3] https://www.top500.org

Blue, and Hydra for playing chess, Gravity Pipe for astrophysics, and MDGRAPE-3 for protein structure prediction and molecular dynamics. Broadly, there are three main components of an HPC system:

- **Compute:** As the name suggests, the compute component focuses on the processing of data, the execution of software or algorithms, and solving problems. A cluster of computers (and all the processors, dedicated circuits, and local memory entailed therein) performing computations would fall under the "compute" umbrella.
- **Network:** Successful HPC architecture must have fast and reliable networking, whether for ingesting external data, moving data between computing resources, or transferring data to or from storage resources. In HPC systems, the speed and flexibility of the interconnect become very important; this is where photonics comes in.
- **Storage:** Cloud storage, with high volume and high access and retrieval speeds, is integral to the success of an HPC system. While traditional external storage is the slowest component of a computer system, storage in an HPC system will function relatively fast to meet the needs of HPC workloads.

Note on Energy Efficiency: There are a lot of efforts to make HPC more energy efficient, reducing the environmental impact. A leading solution is photonics, which uses light to improve data transfer and significantly enhance energy efficiency.

2. HPC applications and market evaluation

High-performance computing (HPC) is crucial for specialised applications such as scientific research, simulation, machine learning, and, most importantly, for Artificial Intelligence (AI). HPC is essential for complex research, including climate and molecular modelling, as well as astronomical simulations. It underpins AI innovations by enabling rapid processing of massive training datasets, which is essential for advanced machine learning algorithms.

Genomic sequencing relies on high-performance computing to perform a huge number of calculations per second, which is critical for mapping the genetic makeup of organisms.

HPC is also essential in feature films and animation for rendering life-like computer-generated imagery (CGI) and realistic special effects. In finance, HPC drives analytical engines and AI, supporting everything from investment analysis to insurance risk assessment and fraud detection.

The applications of high-performance computing vary and span various fields. HPC is often necessary for handling large datasets, performing complex calculations, or both. The table below presents a comprehensive list of important applications.

Sector	Examples
Scientific Research and Simulation	Climate modelling; Molecular modelling in various materials and biological systems; Astronomical simulations;
Government, Security and Defense	Code breaking; Encryption for secure communication; Weapon simulation;
Healthcare and Biotechnology	Genomic sequencing to understand genetic diseases and develop personalised medicine; Drug discovery; Processing of medical images;
Energy Exploration and Production	Analysis of seismic data; Analysis of the location of oil and gas reserves;
Finance and Insurance	Risk management; High-frequency trading; Risk modeling;
Artificial Intelligence and Big Data	Processing large datasets; Training deep learning models;
Automotive and Aviation Industries	Crash simulations; Aerodynamic modeling; Virtual testing of vehicles for safety improvements;
Entertainment, Gaming and Media	Computer-generated imagery (CGI); Creating immersive environments for AR/VR;
Weather Forecasting and Environmental Analysis	Predicting paths and impacts of severe weather events; Studying pollution patterns, deforestation effects and biodiversity changes.

The market shares in terms of applications are given in the figure below.

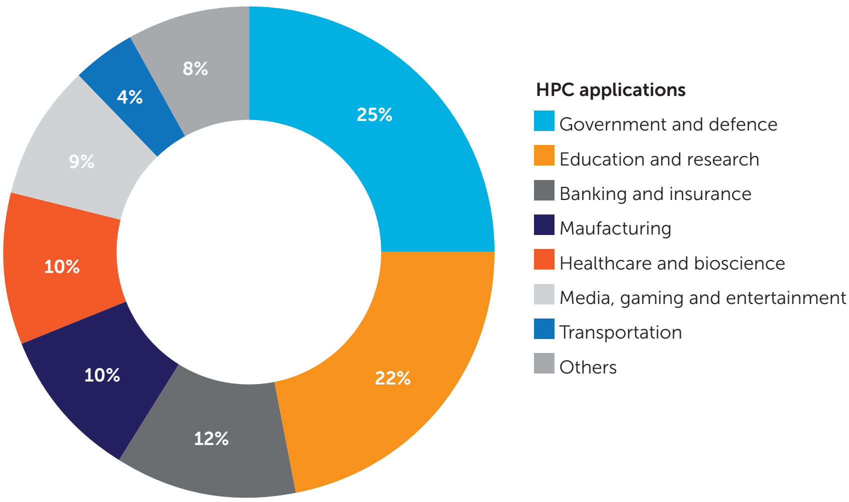

Figure 2: Shares of applications of high-performance computing (HPC) Source: Tematys/Photonics21, 2023.

HPC systems are crucial for addressing complex and computationally intensive tasks, facilitating progress in science, technology, and other areas with significant societal impact. As computational demands increase, the importance of HPC is expected to grow even further across these diverse fields.

3. Overview of photonics for HPC

Figure 3 below gives an overview of current and future applications of photonics technologies in High-Performance Computing.

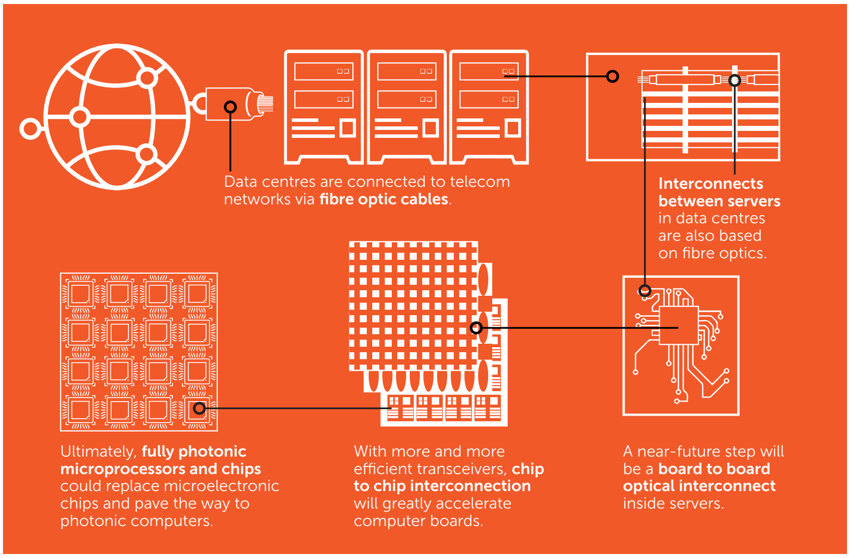

Figure 3: Overview of current and future applications of photonics technologies in High-Performance Computing Source: Tematys/Photonics21, 2023

Connection to telecom networkss

High-performance computing systems are connected to networks to enable remote use of computing resources. They benefit greatly from fibre optic connections, which enhance performance through faster data transfer, reduced latency, increased reliability, scalability, and energy efficiency. These benefits are critical in the high-demand context of HPC.

Figure 4: Rack-to-rack optical interconnects used in HPC Source: iStock

HPC systems typically comprise multiple compute nodes interconnected within a data centre. Since 2008, with rare exceptions, photonics has been utilised to create high-speed and energy-efficient interconnects between these nodes. Fibre optical cables can transmit large amounts of data at high speeds over longer distances with minimal latency and reduced power consumption compared to electrical interconnects. This enables faster communication and data transfer between compute nodes, facilitating efficient parallel computing in HPC clusters.

Memory and storage acceleration

HPC workloads frequently require intensive data access and storage operations. Photonics can improve memory and storage performance in HPC systems by using standard interconnect protocols like Compute Express Link (CXL) over optics. This can significantly accelerate communication between processors and memory/storage subsystems, reducing latency and improving overall system throughput. CXL over optics is particularly beneficial for large-scale simulation, big data analytics, and machine learning applications where rapid data access and processing is critical. An example of this emerging technology is Photowave from Lightelligence, which connects devices over a low-latency optical CXL fabric.

Co-Package optics

Future pluggable optics may have limitations in supporting increasing capacity due to required electrical and optical densities, thermal management, and energy efficiency. The Co-Package Optics (CPO) technology platform aims to overcome these challenges by bringing optics closer to the chipset. The main goal of future designs is to use light to bring data to the point where it is centrally processed. CPO is an innovative concept that brings the optics and the switch ASIC very close together.

Board-to-board and chip-to-chip optical interconnects

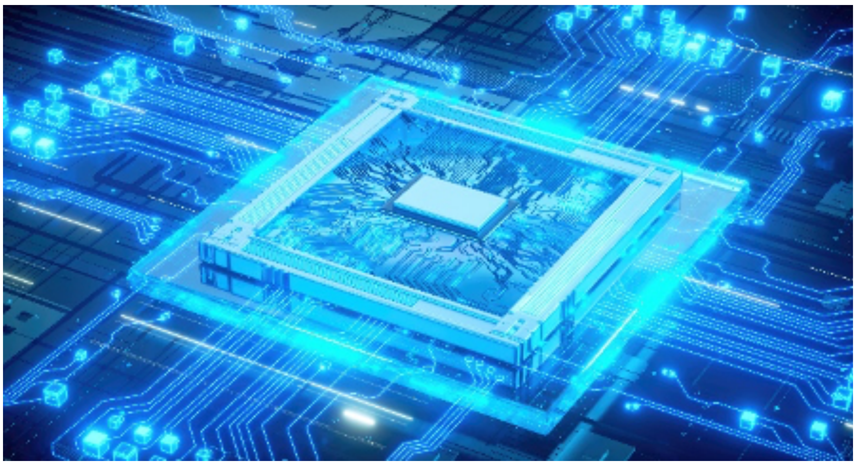

Figure 5: Concept of chip-to-chip optical interconnects
Source: iStock

Large-scale simulation, big data analytics and Artificial Intelligence models are rapidly growing, outpacing the capabilities of traditional architectures (copper-based electrical interconnects) for chip-to-chip or board-to-board communication, creating a bottleneck in scaling HPC. Optical interconnects are emerging as a crucial technology to address this issue. They aim to replace or supplement copper wires with optical channels, offering solutions to the limitations of electronic interconnects. These optical channels are more energy-efficient, have lower cooling needs, and can handle greater bandwidth demands, making them vital for high-performance computing, especially in AI, big data, and cloud computing.

The new disaggregated architecture in HPC separates compute, memory, and storage components. It uses advanced optical I/O technology to connect CPUs, GPUs, and other processors with memory and storage, enabling faster transmission speeds.[4]

Optical computing

Supercomputers' performance improves tenfold every four years. However, traditional electronic-based computing faces limitations due to Moore's law and escalating energy costs. To address future demands, there is a shift towards optical computing, replacing electrons with photons. Significant strides have been made in silicon photonics, enhancing data exchange and enabling optical operations essential for HPC. These advancements make it easier to integrate multiple silicon structures on a single chip for various optical operations, including the implementation of logic gates.

Photonics also hold huge promise in AI, particularly for hardware-level neural networks, which outperform software implementations. Photonic integrated circuits are being developed to demonstrate neuromorphic operations. These circuits leverage non-linear behaviours that can have similar characteristics to the transfer function of a neuron or gate in some operating regimes.

However, challenges remain due to the complexity and cost of fabricating photonic components and integrating them with standard electronics. Despite these challenges, optical computing has the potential to revolutionise computation by providing faster speeds, greater efficiency, reduced data loss, and improved environmental sustainability.

[4] https://www.yolegroup.com/product/report/co-packaged-optics-for-datacenter-2023/

4. HPC – Position of Europe

In terms of HPC system vendors

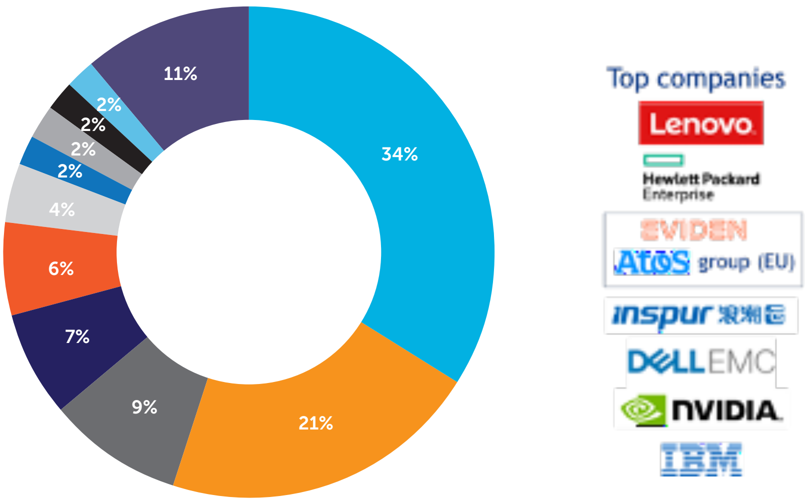

Figure 6: Vendors of HPC (High-Performance Computing) systems: share and top companies
** including with Nvidia/Mellanox*
Source of data: https://www.top500.org/statistics/list/

Vendors system share

- Lenovo
- HPE
- EVIDEN
- Inspur
- DELL EMC
- Nvidia
- NEC
- Fujitsu
- Sugon
- IBM
- Others

EVIDEN, a part of the French ATOS group, is the only European vendor of HPC systems, holding approximately 10% of the system share in the TOP500 list. Intel remains the primary provider of processors for TOP500 systems, accounting for 68% of the systems on the November 2023 list, while AMD processors are used in 28% of the systems.

In terms of HPC infrastructure

The European High-Performance Computing Joint Undertaking (EuroHPC JU) is a legal and funding entity created in 2018 to enable the EU and EuroHPC participating countries to coordinate their efforts and pool their resources with the objective of making Europe a world leader in supercomputing. In July 2021, the Council adopted the EuroHPC JU Regulation, bringing an investment of €7 billion for the period 2021 – 2027. The EuroHPC Joint Undertaking's existing supercomputers are Discoverer in Bulgaria, MeluXina in Luxembourg, Vega in Slovenia, Karolina in Czechia, LUMI in Finland, LEONARDO in Italy, Deucalion in Portugal and MareNostrum in Spain.

EuroHPC JU is pooling European resources to develop top-of-the-range supercomputers for processing big data. One of the European supercomputers, LUMI, located at CSC in Finland, is ranked fifth on the TOP500 list. LEONARDO is ranked sixth, and MareNostrum is eighth. EuroHPC JU's capabilities will be extended to exascale performance and beyond with two supercomputers: JUPITER in Germany and JULES VERNE in France.

In terms of photonics for HPC

Europe currently lags behind other global leaders in the development of photonics for HPC. There is a huge funding gap in private investments for start-ups[5] in new photonic interconnects and optical computing: this will induce slower progress in this field, especially compared to the US. This shortfall would have to be offset by investment from the Member States or the EU.

[5] See the part on investments in the Photonics market and industry report, Photonics21, 2024.

5. Summary and conclusions

HPC systems are complex and require careful planning and optimisation in terms of architecture, operation, and technology choice. Their continued evolution is pivotal in advancing research and development across many scientific and industrial fields.

HPC is indispensable in all EU strategic economic sectors, including automotive, manufacturing, and space, as well as in the new data-driven EU digital economy. The European supercomputing infrastructure is a strategic resource for the future of EU industry, SMEs, and job creation. Even more so, combining HPC with Artificial Intelligence and Big Data will provide unprecedented opportunities for transforming EU businesses, public services, and society.

Photonics plays a crucial role in HPC due to its potential to overcome several limitations of traditional electronic systems. Photonics provides faster data transmission and lower latency, which are essential for the high data processing demands of HPC. It also improves energy efficiency, addressing the power consumption and heat dissipation challenges in HPC environments.

Optical interconnects are becoming a critical solution for the next generation of HPC systems, particularly driven by the increasing demands of AI. The rapid increase in the size of training datasets is becoming a significant obstacle to scaling AI models. This growth could potentially impede AI advancements. Photonics, particularly optical I/O, which can efficiently handle vast data volumes, offers a viable solution to this challenge. The adoption of optical interconnects is increasingly viewed as essential in HPC systems, particularly for improving data movement in AI applications.

In fact, Artificial Intelligence (AI) is playing a crucial role in various scientific fields, particularly in photonics, creating a mutually beneficial relationship. In photonics, AI is revolutionising traditional methods. For instance, the University of Texas uses AI for efficient 'inverse design' in nanophotonic structures, surpassing conventional trial-and-error approaches. The demand for computing power in AI is increasing, as highlighted by OpenAI's CEO Sam Altman. This challenge is being addressed by integrated photonics solutions such as Lightmatter's photonic AI accelerators, which offer efficient processing capabilities. The integration of AI and photonics is creating a positive feedback loop, where advancements in one field propel the other, leading to significant technological progress in both areas.[6]

As data and computational needs continue to grow exponentially, photonics is increasingly recognised as a key enabler for the next generation of HPC and AI, securing its place as the backbone of these systems.

[6] https://spie.org/news/photonics-focus/janfeb-2024/exploring-the-ai-photonics-virtuous-circle

PHOTONICS21

Tematys

Exploration of photonics markets