Inputs for the European Strategy Update for Particle Physics

COMCHA Network

Introduction

The fields of particle and astroparticle physics have seen tremendous advances in computing and artificial intelligence (AI) over the past decade, driven by the ever-growing data demands of experiments like those conducted at the Large Hadron Collider (LHC). The scale and complexity of the data produced by modern particle detectors—often in the range of several petabytes per year—have pushed the limits of traditional computing infrastructure. This has led to a significant evolution in high-performance computing (HPC), data analysis techniques, and the adoption of AI-driven methods. For future colliders and telescopes, these advancements will play a crucial role not only in managing and processing data but also in optimizing experiment design, detector performance, and theoretical simulations.

The current situation in high-energy physics presents both opportunities and challenges. The LHC is expected to continue operating at least until 2040, with the High Luminosity LHC (HL-LHC) upgrades, which will provide greater collision rates, enabling more precise measurements. However, beyond that, there is a growing consensus that a new collider will be necessary to push beyond the current energy and precision frontiers. Cutting-edge detector technology and computational advances in machine learning, quantum computing, and data analysis are positioning the field for breakthroughs.

Several options for future colliders are being considered globally. Among them are the *Future Circular Collider (FCC)* at CERN, in its electron-positron and hadron-hadron versions, the *International Linear Collider (ILC),* a new electron-positron machine in Japan, the *Compact Linear Collider (CLIC)*, an electron-positron collider with more challenging technology, the Muon Colliders, exploiting a new concept in accelerator physics and the *Circular Electron Positron Collider (CEPC)*, a proposal for a Chinese electron-positron collider similar to the FCC–ee.

Whatever will be the decision about the next collider, the computing technologies will be crucial for the success of the experiment and the enhancement of the physics goals.

In addition to advancements in colliders, telescopes dedicated to particle physics—such as cosmic-ray observatories and neutrino detectors—are also pushing the boundaries of technology and computation. These telescopes, including large-scale installations like the IceCube Neutrino Observatory, Cherenkov Telescope Array (CTA), and Pierre Auger Observatory, focus on detecting high-energy cosmic particles, gamma rays, and neutrinos from distant astrophysical sources. The data volumes from these observatories rival those of collider experiments, demanding sophisticated computational infrastructures for real-time data processing, event filtering, and analysis.

State-of-the-art computing solutions in particle physics telescopes often mirror those used in collider experiments. Additionally, hardware accelerators, including GPUs and FPGAs, are increasingly used to handle the large datasets generated from arrays of telescopes, enabling efficient signal processing and real-time filtering of cosmic events. For instance, in experiments like IceCube, GPUs accelerate the reconstruction of neutrino interactions, while machine learning algorithms are used for event classification. The CTA project is investigating the suitability of FPGAs as deep learning accelerators for real-time triggering. These computational innovations are essential for detecting and analyzing the rare particle interactions that provide insight into high-energy astrophysical phenomena, further linking the goals of particle physics and astronomy in exploring the fundamental forces of the universe.

State of the Art in Computing and Artificial Intelligence in Particle Physics

Currently high-performance computing (HPC) is fundamental to the analysis of particle physics data. Moving forward, the need for even more computational power is clear as future experiments are expected to generate significantly more data than the LHC.

- Simultaneously, advancements in heterogeneous computing—exploiting not just CPUs, but also Graphics Processing Units (GPUs), Field-Programmable Gate Arrays (FPGAs), and other specialized accelerators—are becoming increasingly important. The use of GPUs are pioneer in the LHCb trigger system [1] and have become a key component of high-throughput data analysis pipelines in particle physics. They are being used extensively for parallelizable tasks like image recognition and deep learning. FPGAs are widely used in particle physics experiments due to their flexibility, speed, and ability to perform real-time data processing [2]. Their main role is in situations where low-latency, high-throughput, and deterministic processing are required, such as trigger, data acquisition systems and real time data processing.

- Artificial intelligence, particularly machine learning (ML), is playing an ever-larger role in particle physics, both in theoretical research and experimental analysis. AI techniques are now integral to tasks such as event generation, detector simulation, particle reconstruction and identification, event classification, and background noise reduction. The success of AI-driven methods in these areas is due to their ability to recognize patterns in large datasets that would otherwise be difficult to discern using traditional techniques. Examples can be found in Refs [3,4].

Beyond real-time data analysis, AI is also proving to be invaluable in detector optimization and design. Reinforcement learning, for example, has been explored to optimize detector configurations by simulating millions of potential layouts and evaluating their effectiveness in capturing the desired physics signatures [5]. This approach has the potential to revolutionize the way future colliders are designed, allowing for highly efficient, AI-tuned detectors. Fig. 1. shows some of the uses of machine learning in HEP, which include model building, theory generation, detector simulation and optimization, reconstruction tasks and data analysis.

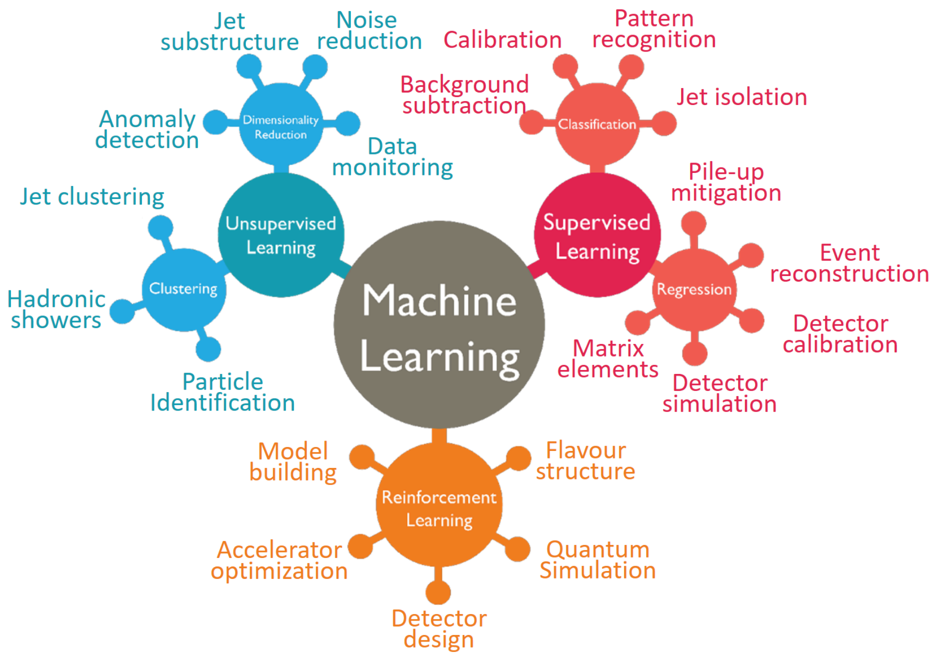

Figure 1. Some of the applications of machine learning in the HEP field.

- Quantum computing is another emerging technology that can make a substantial impact in the coming decades, especially for simulations of quantum field theories and the combinatorial optimization problems that arise in collider physics. Although quantum hardware is not yet ready for large-scale deployment in particle physics, ongoing research in quantum algorithms is promising, particularly for problems that classical computers struggle to solve efficiently [6].

- Neuromorphic systems, which are inspired by the architecture and function of the human brain, offer an emerging and potentially transformative approach to high-performance computing, particularly for data-heavy fields like particle physics. Neuromorphic computing mimics the brain's ability to process information efficiently through massively parallel architectures and low-power synaptic communication. As these systems continue to develop, they have the potential to play a key role in enabling both the scientific and environmental goals of next-generation collider experiments.

In summary, the prospects for computing and AI in future colliders are vast. From data acquisition and real-time analysis to detector design and theoretical simulations, advanced computing technologies will be indispensable. As these tools continue to evolve, they will push the boundaries of what is possible in particle physics, enabling scientists to probe deeper into the fundamental structure of matter and the universe.

Sustainable Computing

The increasing complexity and data demands in particle physics research—particularly for current and future colliders—require significant computational resources. This poses a challenge not only in terms of raw performance but also from an environmental perspective. The energy consumption associated with massive data processing infrastructures has led to growing concerns about the sustainability of high-performance computing (HPC) in scientific research. Hardware accelerators, such as GPUs and FPGAs, are promising options for improving the energy efficiency of computing in particle physics, contributing to more environmentally sustainable research practices.

The enormous amounts of data processed, stored and analyzed by high energy physics experiments require substantial computational power, typically provided by large-scale data centers and global computing grids. These infrastructures are energy-intensive, contributing to increased carbon footprints, particularly as future colliders are expected to produce even more data than the LHC.

Efforts to address this issue have focused on two main fronts: optimizing software algorithms to be more efficient and transitioning to hardware solutions that can deliver high performance at lower energy costs. GPUs and FPGAs are two such hardware accelerators that, when appropriately utilized, can significantly reduce the environmental impact of computing in particle physics.

GPUs for Energy-Efficient Data Processing

GPUs are well-known for their ability to handle highly parallelizable tasks. In particle physics, many data analysis tasks—such as event reconstruction, particle tracking, and simulation—are naturally suited to parallel processing. By using GPUs, researchers can achieve substantial speedups compared to traditional Central Processing Units (CPUs), often by factors of 10 to 100. More importantly, GPUs can process these tasks more efficiently, offering greater performance per watt of power consumed.

One key advantage of GPUs in this context is their ability to handle large-scale machine learning (ML) workloads, which are increasingly important in particle physics for tasks like particle identification, data denoising, and event classification. Training deep neural networks or running inference on experimental data can be highly energy-intensive when done on CPUs. However, GPUs are optimized for matrix operations and floating-point calculations, which are fundamental to ML tasks, thus providing both speed and energy savings.

By shifting computationally expensive tasks to GPUs, particle physics research can achieve a significant reduction in energy consumption. Some studies have shown that GPUs can deliver up to 5-10 times more energy-efficient performance compared to CPUs for similar workloads [7]. This becomes particularly relevant as future colliders. Efficient use of GPUs could thus help mitigate the environmental impact of data processing at these facilities.

FPGAs for Low-Power, High-Efficiency Computation

FPGAs represent another class of hardware accelerators that can play a crucial role in sustainable computing. Unlike GPUs, which are designed for general-purpose parallel computing, FPGAs are highly customizable and can be programmed to perform specific tasks with optimized energy efficiency. They excel in scenarios where deterministic, low-latency, and low-power performance is needed, making them ideal for real-time data processing in particle detectors.

In particle physics, FPGAs are already widely used for tasks like trigger systems, which filter out irrelevant collision events in real-time to reduce the amount of data that needs to be stored and analyzed. Future colliders will require even more advanced and energy-efficient trigger systems and real-time processing due to the increased data rates. FPGAs can be programmed to implement complex algorithms at the hardware level, allowing them to process large volumes of data with minimal power consumption.

Because FPGAs can be configured to perform only the necessary operations, without the overhead associated with general-purpose processors, they offer exceptional energy efficiency. For example, FPGAs have been shown to provide up to 50 times the performance-per-watt of traditional CPUs for specific workloads [8], such as filtering, sorting, and event reconstruction. This makes them an attractive option for sustainable computing in future collider experiments, where energy efficiency will be critical to managing the environmental footprint of data processing.

Combining GPUs and FPGAs for Maximum Sustainability

A hybrid approach that combines the strengths of GPUs and FPGAs can provide the best of both worlds—high-performance parallel computing from GPUs and highly efficient, task-specific computation from FPGAs. This approach is already being explored in some areas of particle physics, where both accelerators are used in tandem to optimize data analysis pipelines.

For example, FPGAs can be used at the front end of the data processing chain, handling real-time tasks like event triggering and filtering, which require low-latency, energy-efficient computation. GPUs can then take over for more compute-intensive tasks such as machine learning inference, simulation, and event reconstruction. By dividing the computational workload based on the strengths of each accelerator, researchers can achieve significant energy savings without sacrificing performance.

Sustainability Prospects for Future Colliders

As future collider projects begin to take shape, the role of hardware accelerators will become even more crucial for achieving environmentally sustainable computing. These colliders are expected to produce data on a scale that dwarfs even the LHC, placing unprecedented demands on computing infrastructure. GPUs and FPGAs, along with other emerging technologies like quantum computing and neuromorphic chips, will be key to ensuring that the next generation of experiments can be conducted in an energy-efficient manner.

Efforts to develop more energy-efficient software algorithms will also complement the use of hardware accelerators. For instance, researchers are already working on optimizing machine learning models to run more efficiently on GPUs, and low-power FPGA implementations of physics algorithms are being developed to further reduce energy consumption.

In summary, hardware accelerators like GPUs and FPGAs offer powerful tools for reducing the environmental impact of computing in particle physics. By leveraging their energy-efficient performance, researchers can address the growing demand for computational resources in future colliders while minimizing their carbon footprint. As the field continues to advance, the integration of these technologies will be essential for achieving both scientific and environmental goals.

Bibliography

[1] LHCb Collaboration, Future physics potential of LHCb; LHCb-PUB-2022-012; LHCb Upgrade GPU High Level Trigger Technical Design Report LHCB-TDR-021; LHCb potential to discover long-lived new physics particles with lifetimes above 100 ps, Eur.Phys.J.C 84 (2024) 6, 608
[2] S. P. Summers, Application of FPGAs to Triggering in High Energy Physics, https://cds.cern.ch/record/2647951
[3] ATLAS Collaboration, New techniques for jet calibration with the ATLAS detector, *Eur.Phys.J.C* 83 (2023) 8, 761, e-Print: 2303.17312 [hep-ex]
[4] F. Mokhtar, Progress towards an improved particle flow algorithm at CMS with machine learning, e-Print: 2303.17657 [physics.data-an]
[5] C. Fanelli, Design of detectors at the electron ion collider with artificial intelligence, *J.Phys.Conf.Ser.* 1525 (2020) 012098
[6] W. Y. Chan, Application of quantum computing techniques in particle tracking at LHC, *EPJ Web Conf.* 295 (2024) 12001
[7] D. Saikumar, Future of computing at the Large Hadron Collider, e-Print: 2210.13213 [hep-ph]
[8] Z. Que, Optimizing Graph Neural Networks for Jet Tagging in Particle Physics on FPGAs, FPL22 Proceedings 327-333, http://cds.cern.ch/record/2861082