

Emerging Computational Tools for Astrophysics Research

Fabio Vitello



Giornate INAF 2023

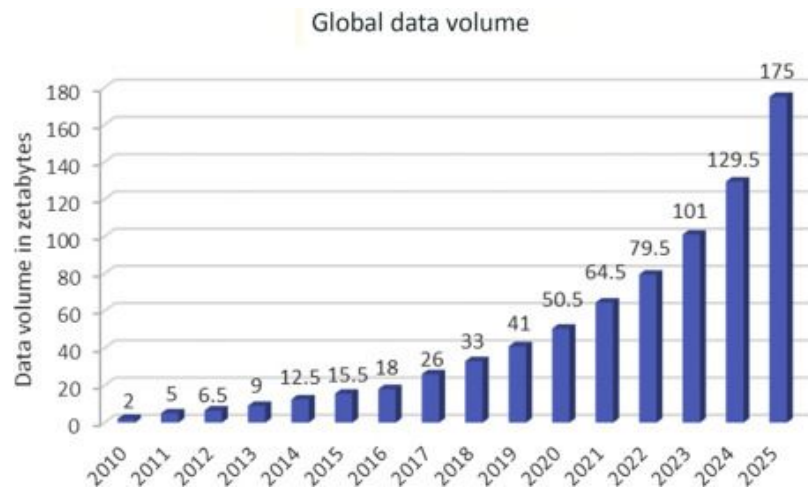
Transforming Astronomy Research

- Astrophysics research involves analyzing vast amounts of data from telescopes, instruments and numerical simulations.
- Emerging computational tools that can help on advancing discovery:
 - High Performance Computing (HPC),
 - Machine Learning
 - Quantum Computing,
 - Visualization Techniques.



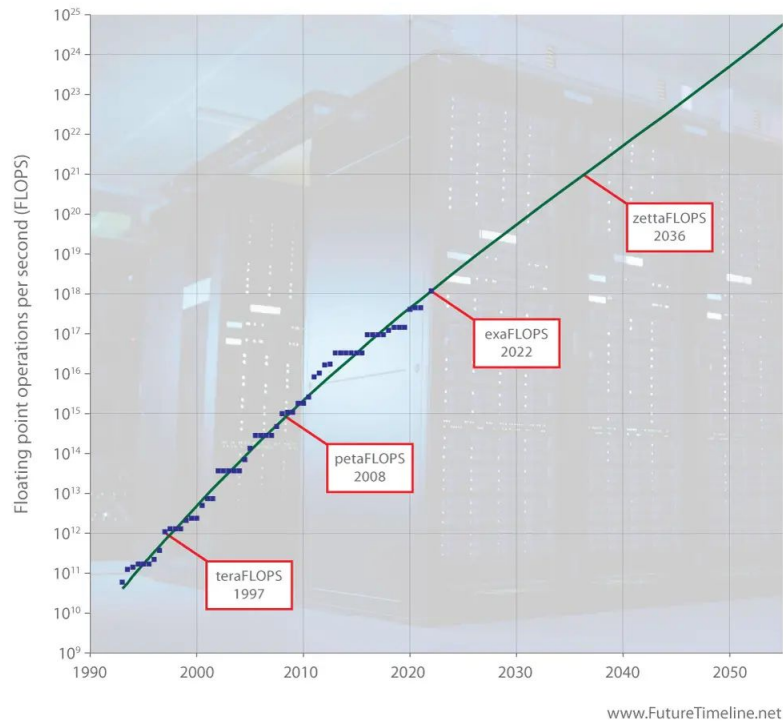
Advancing Discovery Through High Performance Computing

With the increasing complexity of astrophysical simulations and the need to analyze large datasets from next generation instruments, high performance computing has become an essential tool for astrophysicists.



Lei, Jiale, and Linghe Kong. "Fundamentals of big data in radio astronomy."
Big Data in Astronomy. Elsevier, 2020. 29-58.

Exascale systems



- Exascale refers to a level of performance capable of achieving at least one exaflop (10^{18} floating-point operations per second). This represents a significant increase in computing power compared to current systems, which are typically measured in petaflops (10^{15} FLOPS).
- Exascale systems are often designed to be highly energy-efficient, with advanced cooling and power management technologies.

Frontier



Hewlett Packard Enterprise Frontier, or OLCF-5, is the world's first and fastest exascale supercomputer, hosted at the Oak Ridge Leadership Computing Facility (OLCF) in Tennessee, United State.

It is based on the Cray EX and is the is the world's fastest supercomputer. Frontier achieved an Rmax of 1.102 exaFLOPS, using AMD CPUs and GPUs.

Measured at 62.68 gigaflops/watt, In November 2022 Frontier topped the Green500 list for most efficient supercomputer.

Leonardo

LEONARDO, the pre-exascale Tier-0 EuroHPC supercomputer is classified in 4^o position in the Top500 List with 250 petaflops. It is hosted by Cineca in the Bologna Technopole.

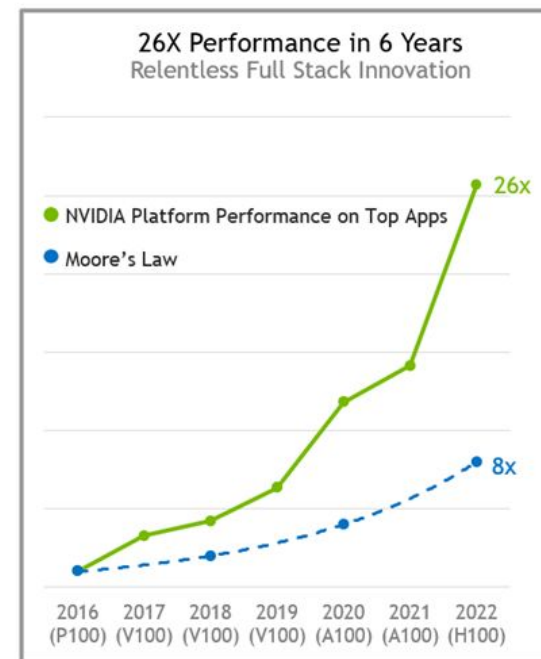
The system consists of an Atos BullSequana XH2000 computer, with close to 14,000 Nvidia Ampere GPUs and 200Gb/s Nvidia Mellanox HDR InfiniBand connectivity.

It is constructed with two main partitions: Booster Module and Data-centric Module.



GPU porting to enable exascale

- GPUs provide significantly more compute cores and higher memory bandwidth than CPUs.
- GPUs scales more naturally to larger core counts than CPUs, allowing continued performance improvements as more powerful GPUs and systems become available.
- While GPUs use more power than CPUs for a given compute workload, **power per unit of performance continues improving** over time. This enables potential exascale capabilities within reasonable power budgets, increasing sustainability of advanced systems and **reducing their environmental impact.**



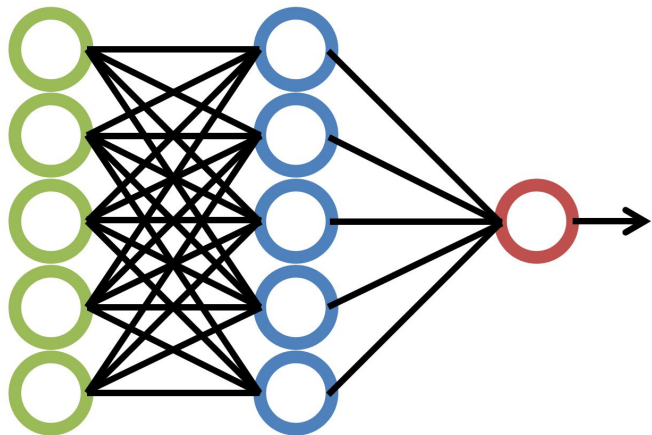
New processor architectures: RISC-V

- RISC-V is an open source instruction set architecture standard. It provides flexibility to customize processors with optimized instruction sets, cache/memory subsystems, extensions, etc.
- Architecture extensibility allows to optimize processors for specific astrophysical applications, such as simulating star formation models or reconstructing images from telescope data.
- Some RISC-V designs aim for high energy efficiency and low power. Benefits for astrophysics include minimizing wasted power during I/O- or memory-intensive phases of computation. Lower environmental footprint and operating costs.



ML - Unlocking Discovery in Massive Data

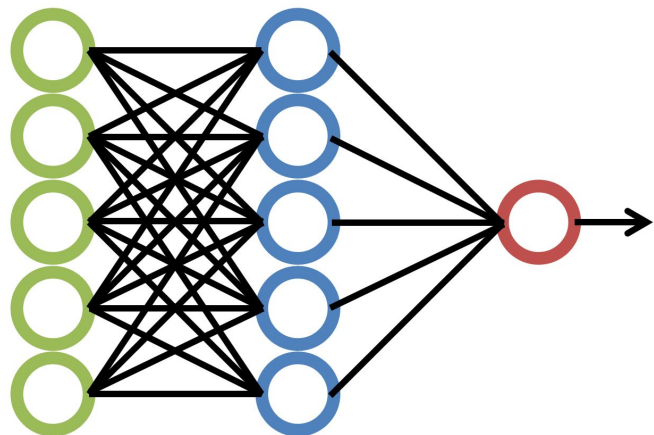
ML frameworks can be used in astrophysics research for a variety of applications, including:



- **Image Analysis:** ML frameworks can be used to analyze images and extract information about the objects and structures within them. This can include tasks such as **object detection**, **segmentation**, and **classification**.
- **Data Mining:** ML frameworks can be used to mine large datasets and identify patterns and trends within the data. This can include tasks such as **clustering**, **classification**, and **anomaly detection**.

ML - Unlocking Discovery in Massive Data

ML frameworks can be used in astrophysics research for a variety of applications, including:



- **Simulation and Modeling:** ML frameworks can be used to build and train models of astronomical phenomena. This can help to make **more accurate predictions** about future observations.
- **Time-Series Analysis:** ML frameworks can be used to analyze time series data from astronomical observations. This can include tasks such as **signal processing, feature extraction, and classification**.

ML - Unlocking Discovery in Massive Data

ML frameworks that can be used in astrophysics research:

- **TensorFlow:** TensorFlow is an open-source ML framework developed by Google. It provides a wide range of tools and libraries for building and training ML models, and can be used for tasks such as image classification, object detection, and time-series analysis.
- **PyTorch:** open-source ML framework that provides an easy-to-use interface for building and training deep learning models, and supports a wide range of applications, including image analysis, natural language processing, and graph analysis.
- **Scikit-learn:** Scikit-learn is a popular ML library that provides a variety of tools for data analysis and model training. It can be used for a wide range of tasks, including classification, regression, clustering, and dimensionality reduction.

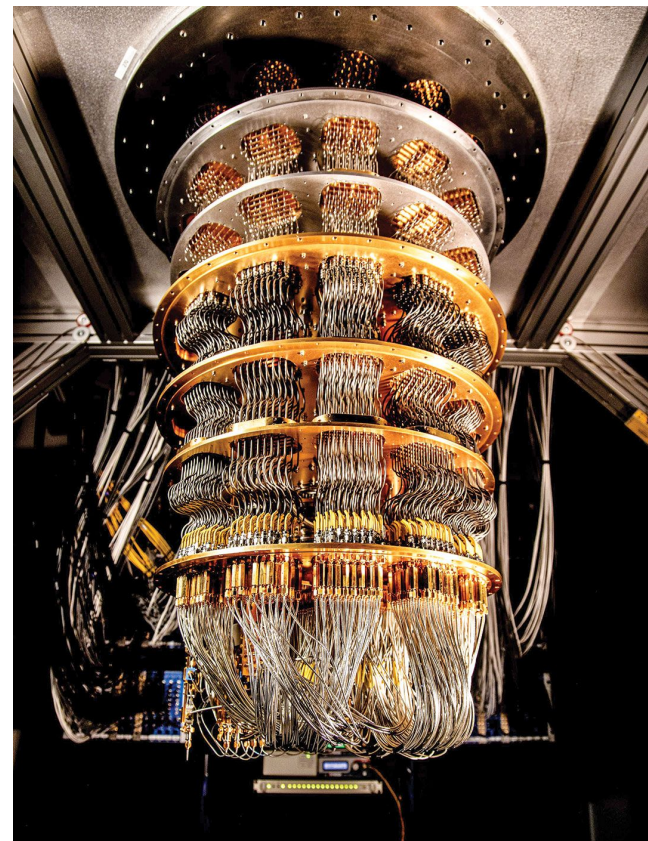
ML - Unlocking Discovery in Massive Data

ML frameworks that can be used in astrophysics research:

- **Keras:** Keras is a high-level ML library that provides a user-friendly interface for building and training deep learning models. It is built on top of TensorFlow and can be used for a variety of applications, including image processing, natural language processing, and time-series analysis.
- **Theano:** Theano is a mathematical library that can be used for building and training ML models. It provides a variety of tools and libraries for numerical computation, and can be used for tasks such as neural network training, optimization, and symbolic differentiation.

Quantum computing

Quantum computers exploit quantum **superposition** and **entanglement** to solve complex problems intractable for classical computing. They can factor huge numbers, optimize difficult functions, simulate quantum systems with far more precision.



Quantum computing

Promising application of quantum computing in astrophysics includes:

- **Simulations:** Quantum computers can simulate quantum mechanical systems with exponential speedup over classical computers, enabling to simulate the behavior of large-scale systems such as galaxy formation and black hole accretion.
- **Data Processing:** Quantum computing could enable faster and more efficient processing of large astronomical datasets. Quantum algorithms can be used to **search large databases**, perform **clustering** and **classification** tasks, and optimize the analysis of astronomical data.
- **Machine Learning:** Quantum machine learning (QML) can be used to develop new algorithms for analyzing and understanding data. QML can be used for **clustering**, **classification**, and **anomaly detection**.

Visualization Tools for Astrophysics

Data visualization is a critical tool for astrophysics research, as it allows researchers to **discovers** hidden patterns, **insights** and relationships within data and **communicate** complex datasets and simulations.

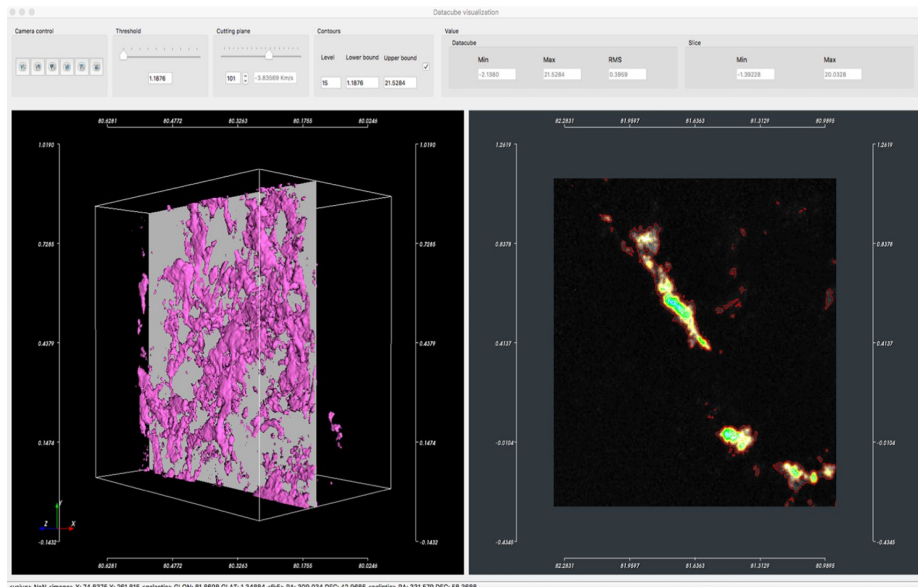
Common visualization techniques:

- **2D and 3D plots:** can reveal patterns in the data.
- **Volume rendering:** allows researchers to visualize astronomical data in 3D space, which can help to reveal hidden structures and relationships within the data.
- **Interactive visualizations:** allow researchers to explore and analyze complex datasets in a more intuitive and interactive way.

Visualization Tools for Astrophysics

Visual Analytics: Combines visualization techniques with data analysis and machine learning algorithms to enable users to gain insights from complex and large datasets.

Valuable tools to discover patterns, trends in multivariate data.

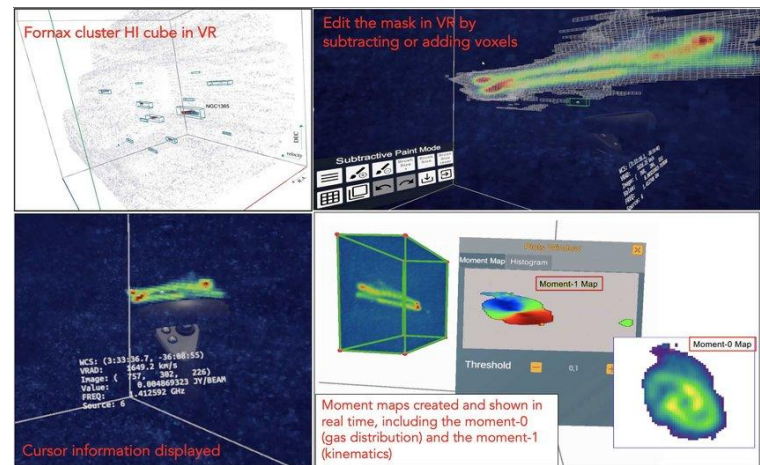


Vitello, F., et al. "Vialactea visual analytics tool for star formation studies of the galactic plane." *Publications of the Astronomical Society of the Pacific* 130.990 (2018): 084503.

Visualization Tools for Astrophysics

Virtual Reality (VR) can be a powerful tool for visualizing and interacting with large and complex datasets. By using VR headsets and hand controllers, users can navigate through the data, zoom in and out, and interact with 3D objects and visualizations in real-time.

VR can help users to develop a better spatial understanding of astronomical data by allowing them to see the data in 3D space.



Marchetti, Lucia, et al. "iDaVIE-v: immersive Data Visualisation Interactive Explorer for volumetric rendering." *arXiv preprint arXiv:2012.11553* (2020).

Visualization Tools for Astrophysics

Visualizing petabyte-scale data can be a significant challenge.

However, here are a few strategies that can help to visualize big data at this scale:

- **Data Sampling:** Sample the data to reduce its size, while still preserving important features and patterns. This can be done by selecting a subset of the data points or by aggregating the data into coarser-grained units. This can be helpful for exploratory analysis or for generating quick, rough visualizations of the data.
- **Data Reduction:** use data reduction techniques to transform the data into a lower-dimensional space, while preserving important features and patterns, that can be more easily visualized. This can be done using techniques such as principal component analysis (PCA) or t-SNE. This can be helpful for creating visualizations that highlight the structure and relationships within the data.

Visualization Tools for Astrophysics

- **Parallel visualization** is a technique that involves distributing the computational load of visualizing large and complex datasets across multiple processors or computing nodes. This can enable users to interactively explore and analyze datasets that would be too large or complex to visualize on a single machine.
- **Remote visualization** is a technique that allows users to visualize and interact with large and complex datasets from a remote location, typically over a network connection. This technique can be useful for researchers who need to analyze data that is stored on remote servers or high-performance computing clusters.

Challenges & Opportunities

- Complex exascale hardware architectures with millions of cores, deep memory hierarchies and high-speed networks requires **software optimized for scalable performance**.
- Larger problem sizes introduce **need for innovative algorithms**.
- Higher complexity demands strategies for **resilience, manageability and sustainability**.
- One of the main challenges in using machine learning (ML) models is the issue of "**black box**" models, where the internal workings of the model are not readily interpretable or explainable.



Challenges & Opportunities

- Though emerging, quantum computing has reached "quantum supremacy" and continues scaling toward usefulness. Challenges remain around building **stable quantum systems**, developing intuitive software tools, and finding "**killer apps**" driving mainstream adoption.
- Porting large, complex astrophysics codes written originally for CPUs to run efficiently on GPUs introduces numerous **software engineering challenges** around memory management, data transfers, work distribution, kernel fusion, pointer chasing, and more.
- Partnerships between astrophysicists, software engineers, hardware architects, and programmers create opportunities of **co-design** to tailor architectures, specifications and software stacks.



fabio.vitello@inaf.it