Computing Challenges for Future Colliders

COMCHA
Computing challenges

Arantza Oyanguren, Luca Fiorini
IFIC – Valencia (Spain)

Contents

- Introduction
- Artificial Intelligence
- Triggers and hardware accelerators
- Simulations: physics and detectors
- Software and data analysis
- Sustainability
- Emerging technologies: quantum, neuromorphics, …
- COMCHA network: collaboration, training and skills
- Conclusions

Introduction

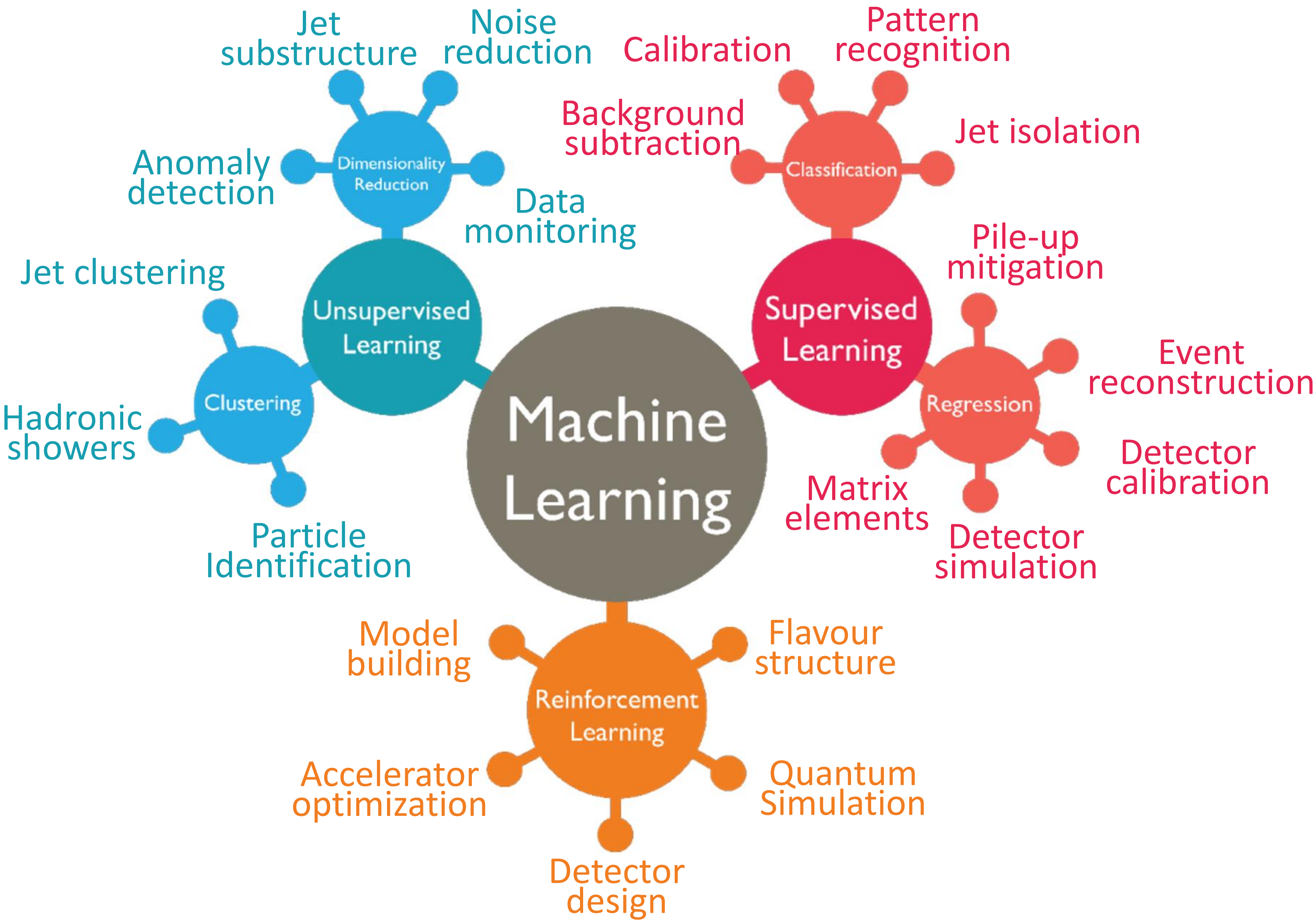
- We have built powerful accelerators and amazing detectors in high-energy physics, and plan to go even further.

- Software and computing is used during **every step** of the **data processing chain** → From accelerator and detector control, through trigger, to reconstruction, simulation and analysis.

The challenges…

- New challenge and unconventional accelerators (HL-LHC, lepton and hadron colliders)
- More sophisticated and performant detectors.
- More channels, additional information (including time!)
- More backgrounds: Increased pile up, beam-induced background
- Higher data rates: improved (and hw-less!) triggers.
- More physics precision needed for probing new physics (above TeV!).
- Need to explore unconventional signatures (what are we missing?).

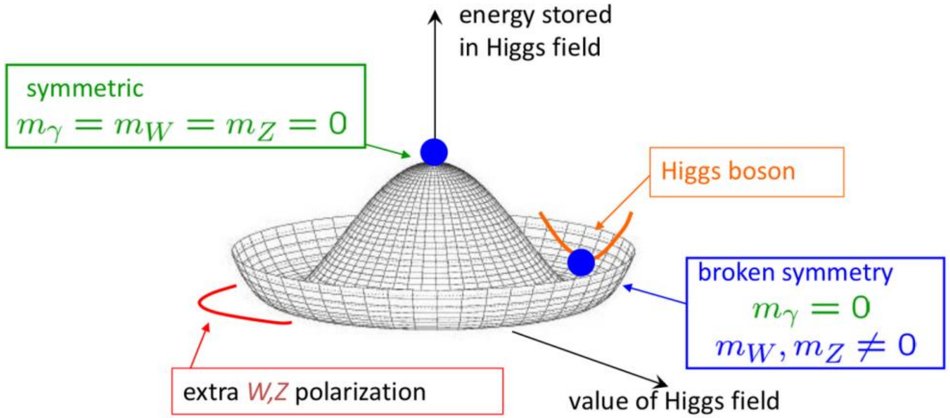
Artificial Intelligence

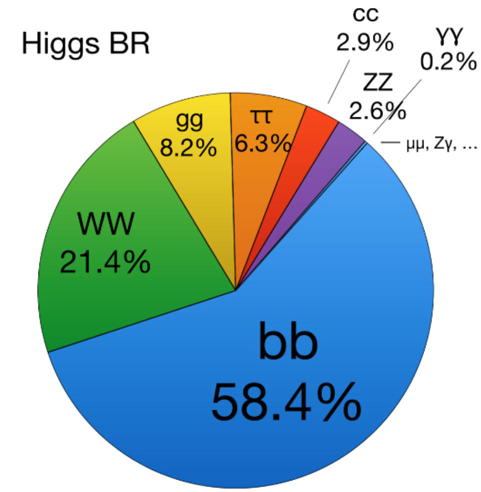


See https://iml-wg.github.io/HEPML-LivingReview/

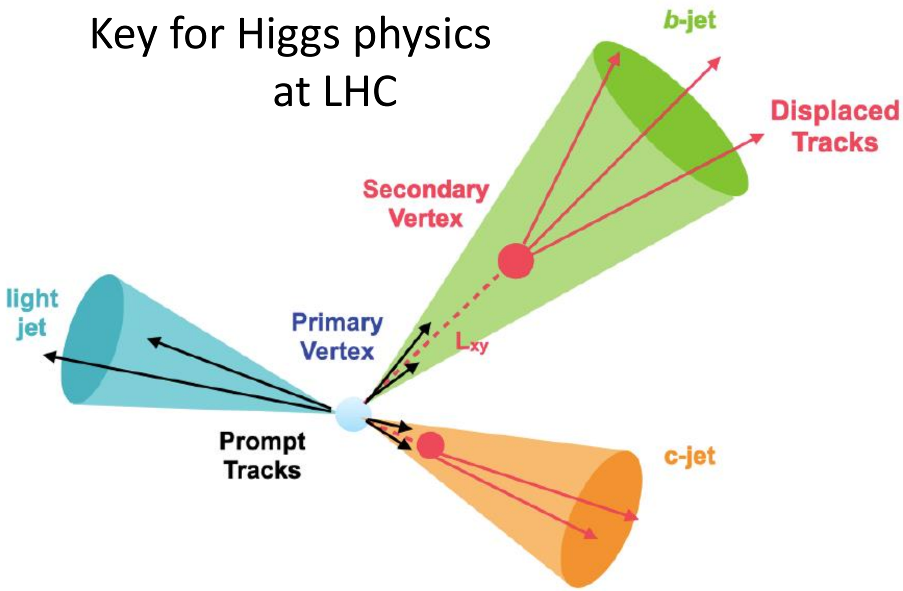
Artificial Intelligence

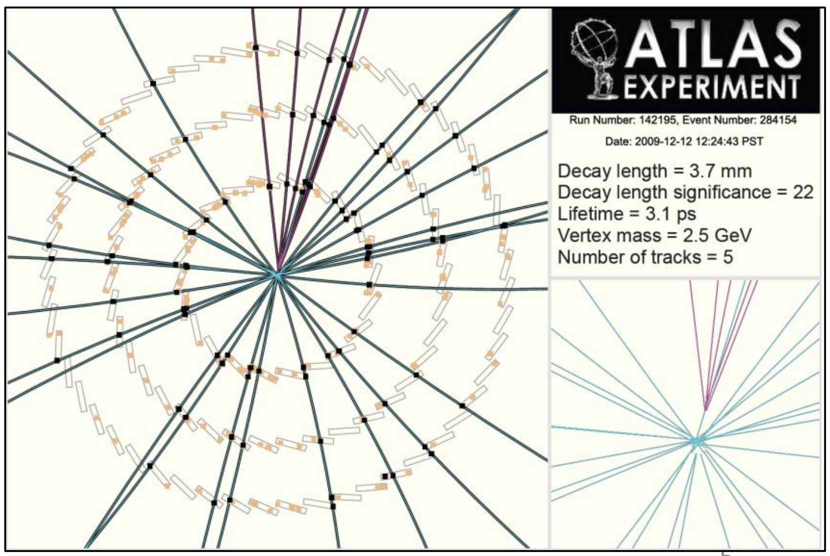
- **Jet Flavour Classification**: needed for Higgs physics





Key for Higgs physics at LHC



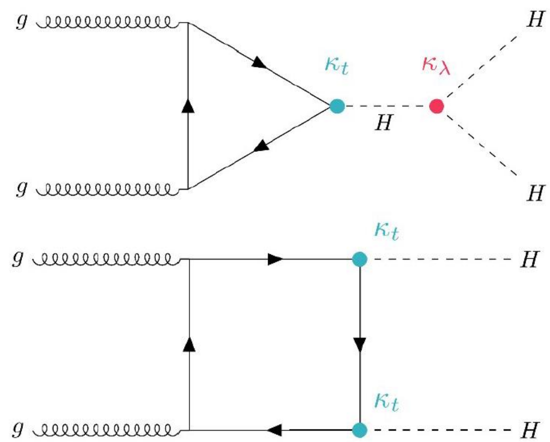


- ML development: Recurrent Neural Networks [JINST 15 (2020) 12, P12012]

Artificial Intelligence

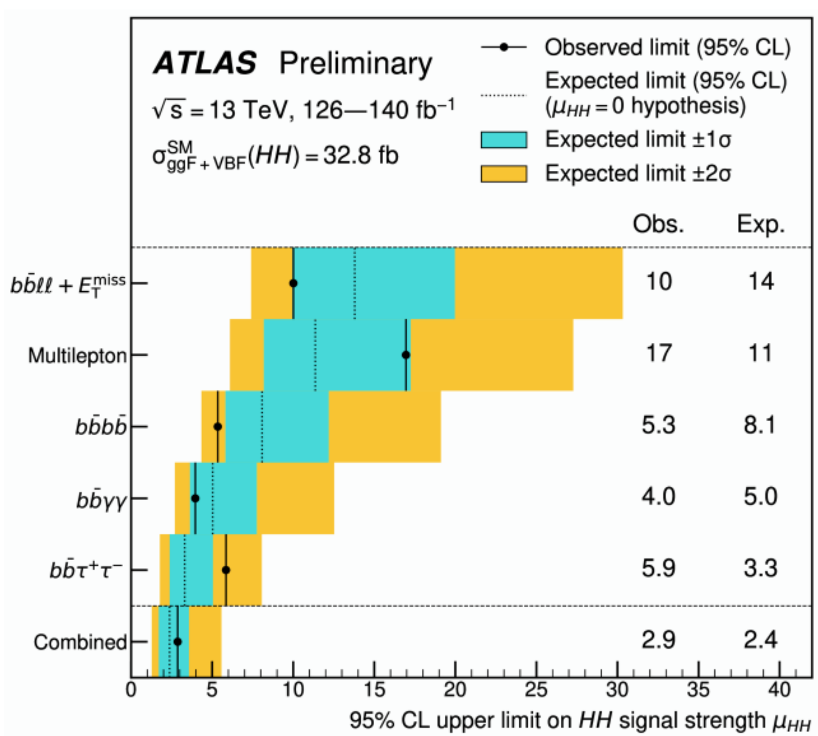
- Huge impact

→ ex: HH production

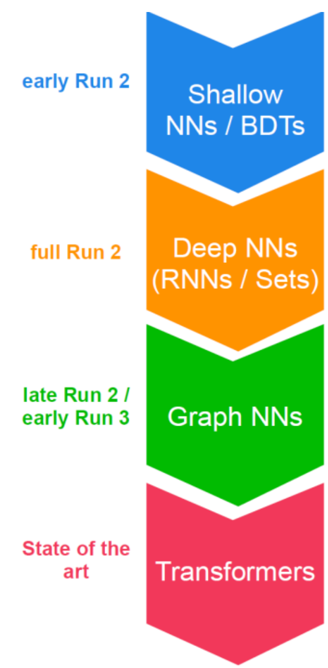


	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

$\sigma_{HH/SM}$ limit



→ Trends in flavour-tagging:



Trigger and hardware accelerators

Exploiting hardware accelerator technologies in event reconstruction and selection:

→ Use more than one kind of processor or cores to maximize performance and energy efficiency.
→ Exploit the high level of parallelism to handle particular tasks.

Graphic Processor Units (GPUs)

- Multicore processors, highly commercial
- High throughput (# processed events / time)
- Ideal for data –intensive parallelizable applications

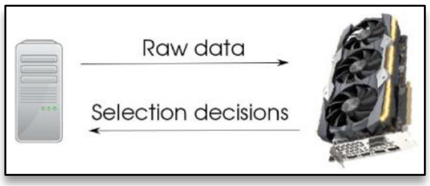
Field Programmable Gate Arrays (FPGAs)

- Programmable and flexible devices
- Low latency
- Low power consumption
- Ideal for compute- and data-intensive workloads

Trigger and hardware accelerators

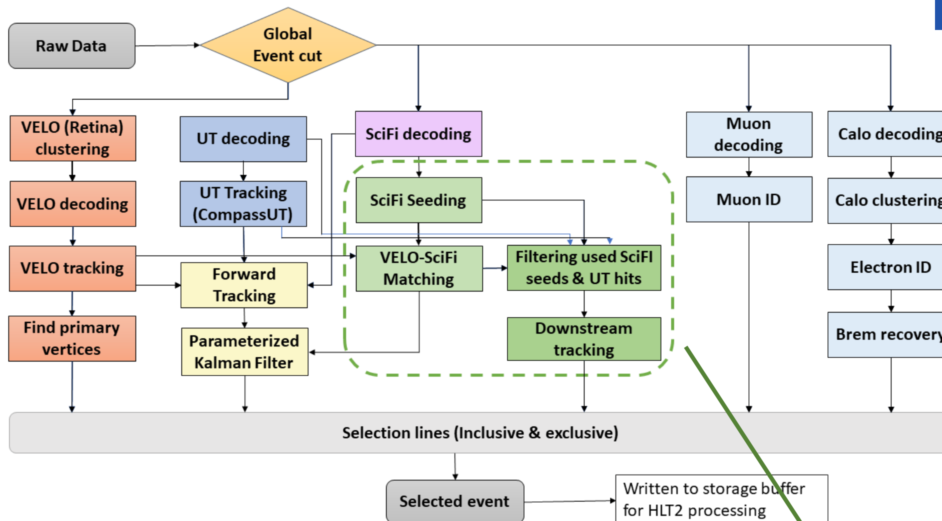
Allen: the LHCb high-level trigger 1 (HLT1) application on GPUs.

[LHCB-TDR-021]

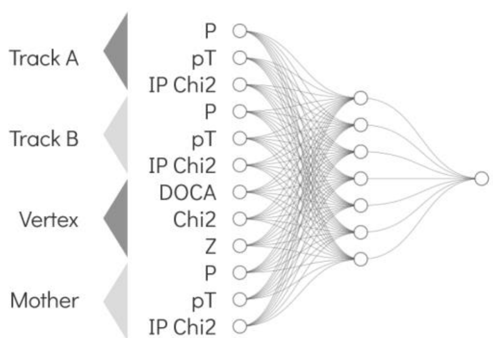


[*Allen project*]

[Com. Softw Big Sci 4, 7 (2020)]

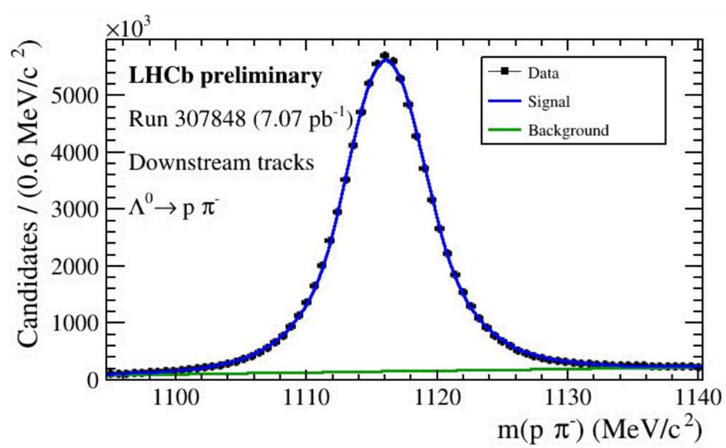


- New very challenging reco. algorithms (not possible in previous runs).
- A Neural Network working at 30 MHz!



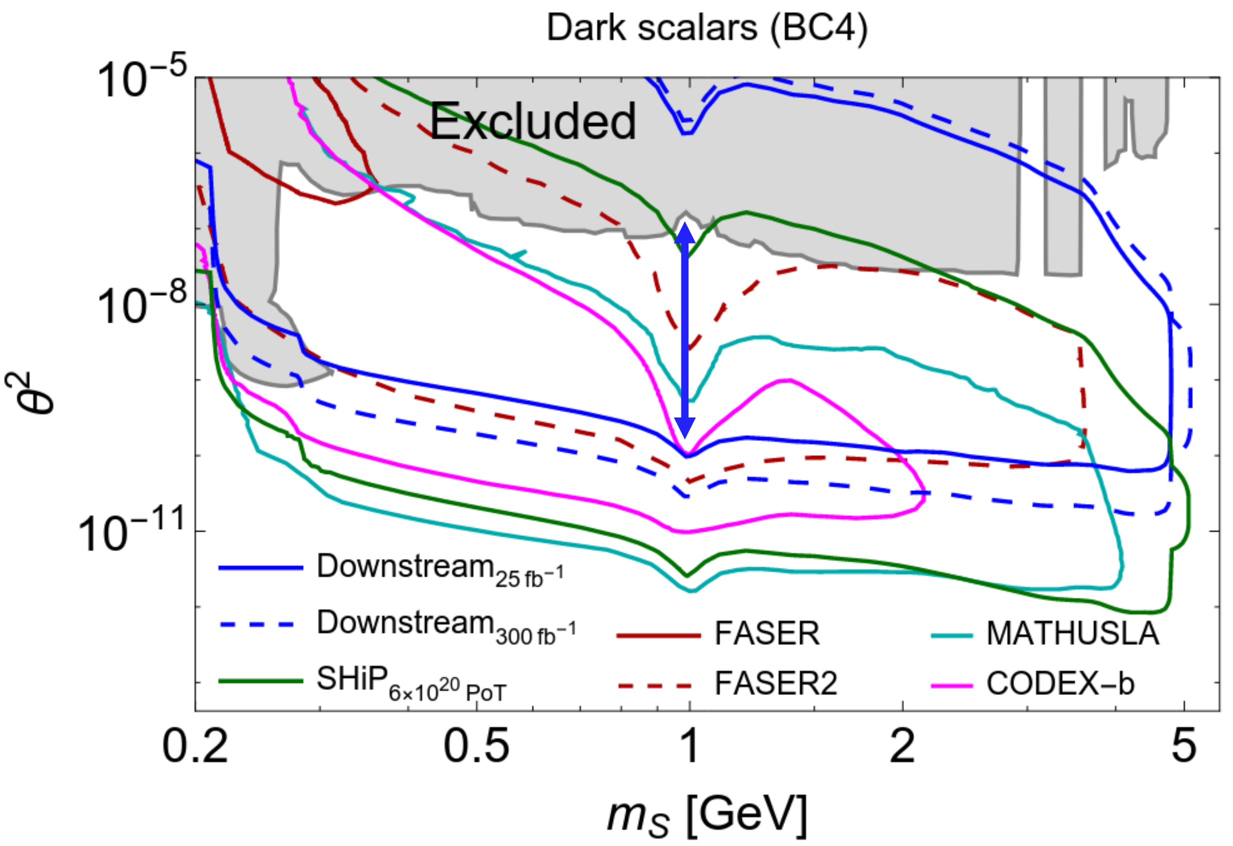
[Com. and Soft. for Big Science; 9,10, (2025)]

[LHCB-FIGURE-2024-035]



Trigger and hardware accelerators

- Extension of the LHCb capabilities to detect new long lived particles [Eur.Phys.J.C 84 (2024)6, 608]



Extension of two orders of magnitude in lifetime!

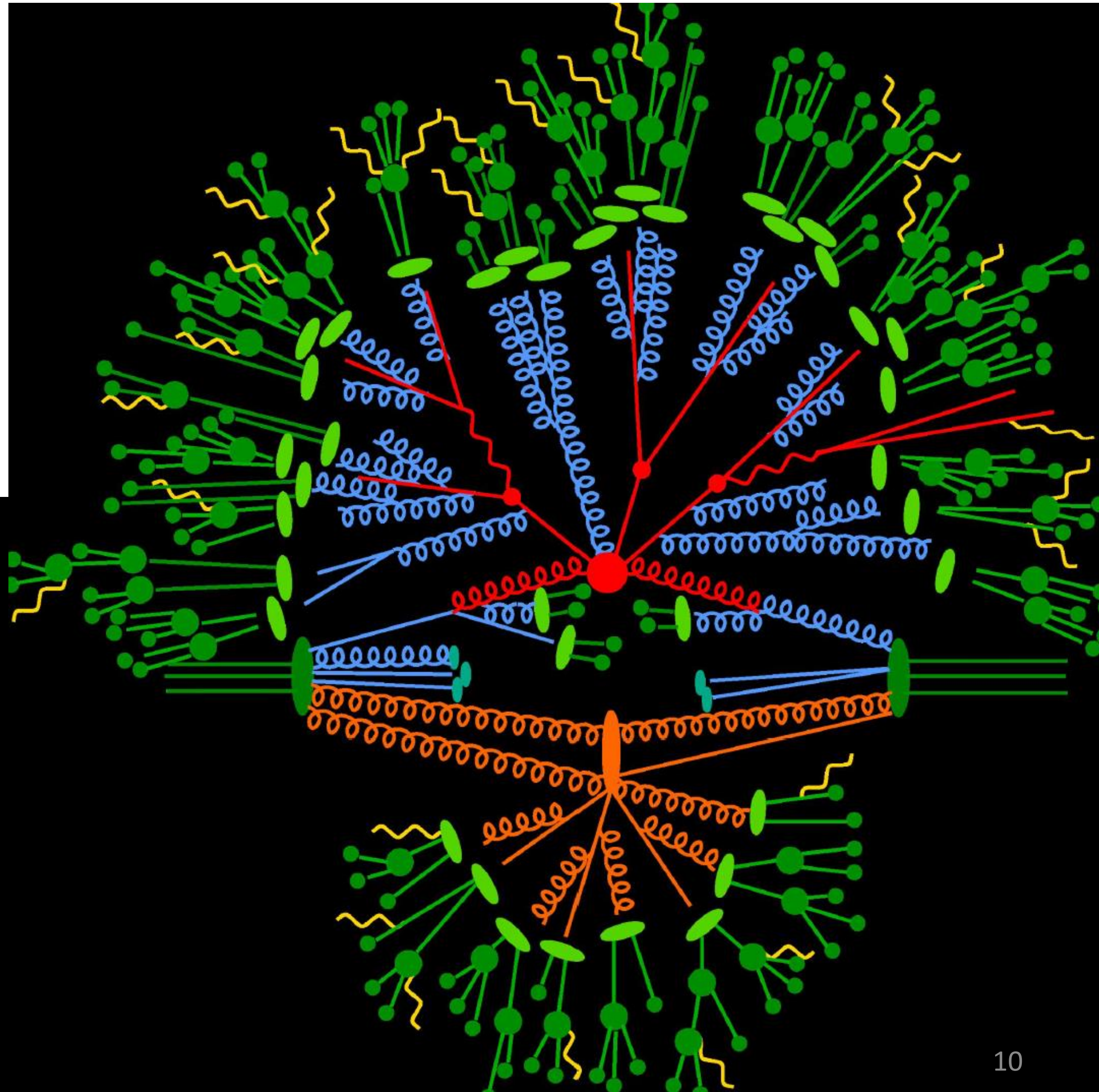
Similar enhancement for Heavy Neutral Leptons (HNL), Axion Like Particles (ALPs) and Dark photons.

Simulations: physics and detectors

https://montecarlonet.org/

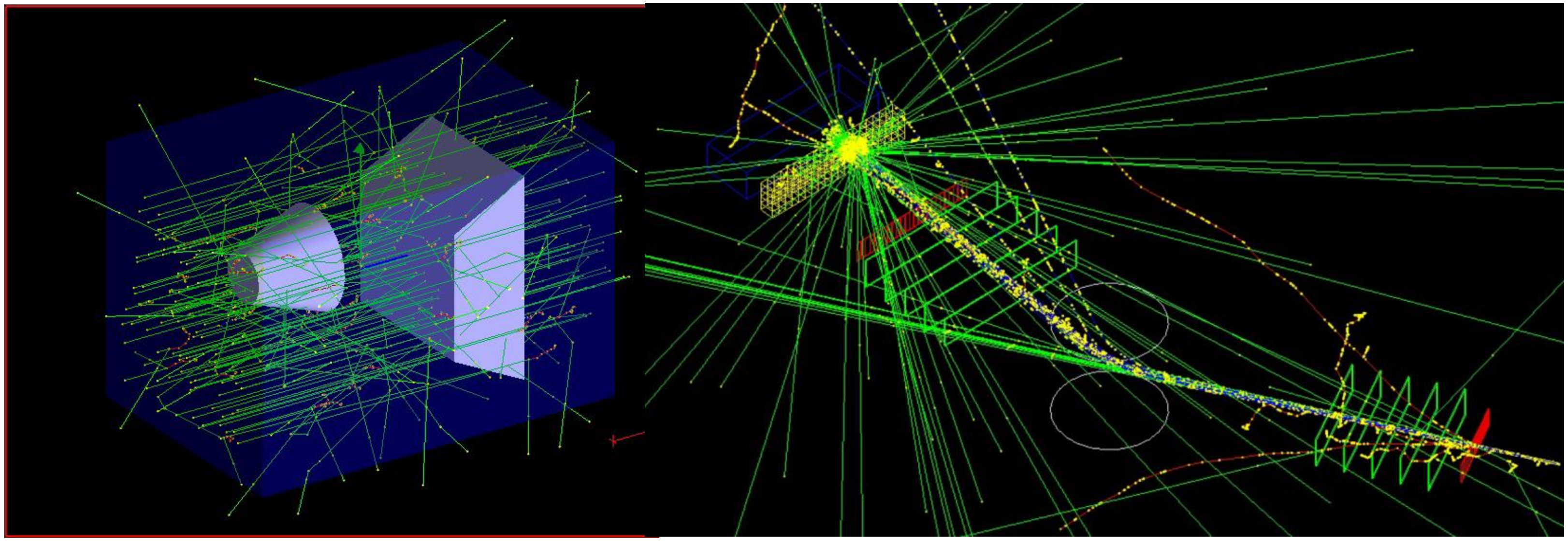
MCnet

Herwig
Pythia
MadGraph/aMC@NLO
Sherpa



hard scatter: matrix elements from first principles - incoming partons from parton-distribution functions(PDFs)
radiative corrections: resumming logarithms to all orders
multiple parton interactions: additional interactions between proton remnants
hadronisation: going colourless
hadron decays: from excited states to final-state particles
photon radiation: QED corrections

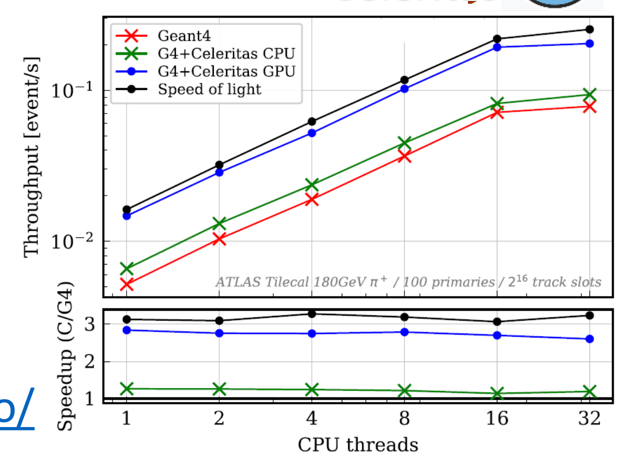
Simulations: physics and detectors



- Simulations take about 60% of the total CPU usage in a LHC experiment.
- Many of the physics results are limited by the statistics of the MC.

Trends:

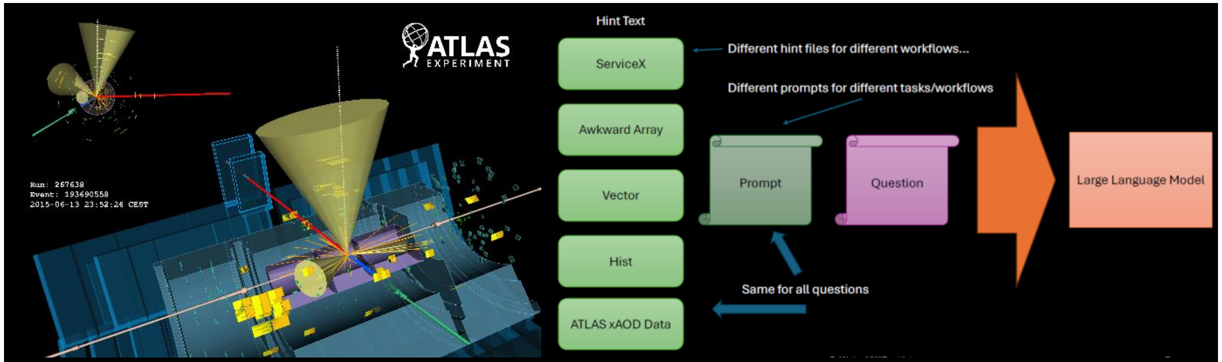
- ML and Generative models (GANs, VAEs, transformers).
- Fast simulations.
- Porting parts to accelerator platforms (GPUs, FPGAs)



https://celeritas-project.github.io/

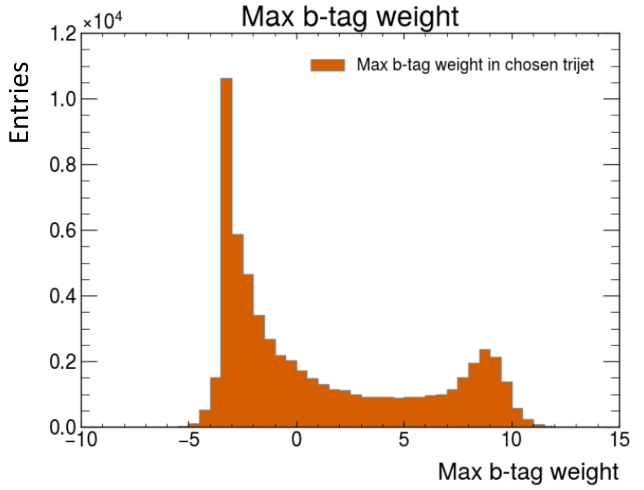
Software and data analysis

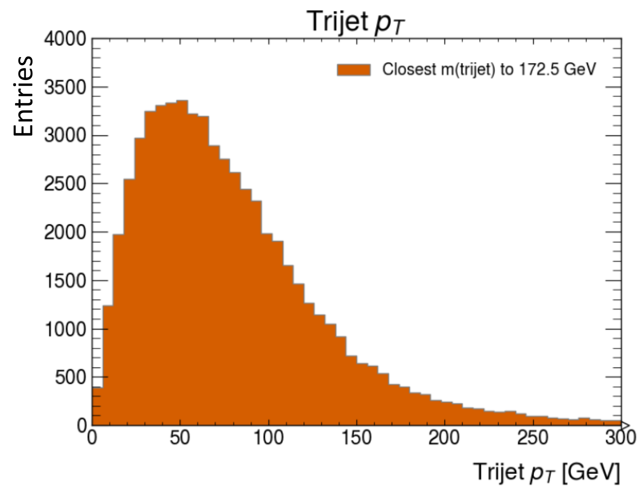
Potential of Large Language Models (LLMs):



Question 6: For events with at least three jets, plot the pT of the trijet four-momentum that has the invariant mass closest to 172.5 GeV in each event and plot the maximum b-tagging discriminant value among the jets in this trijet

1. **Loop** over all the events
2. **Require** the event to have three jets
3. **Fetch** the jets
4. **Form** all 3-pair combinations
5. **Pick** the 3-pair combination closest to mass 172.5 GeV (top mass)
6. **Fill** an appropriately defined histogram
7. **Fetch** the b-tagging discriminate for all three jets
8. **Fill** an appropriately defined histogram with the largest discriminate value.
9. **Plot** the histograms





G. Watts, ACAT 2025 https://github.com/gordonwatts/hep-data-llm

Software and data analysis

Making research data analysis open and reproducible:

reana — Home | Examples | Get Started | Documentation | News | Roadmap | Contact | Blog

Reproducible research data analysis platform

Flexible

Run many computational workflow engines.

Scalable

Support for remote compute clouds.

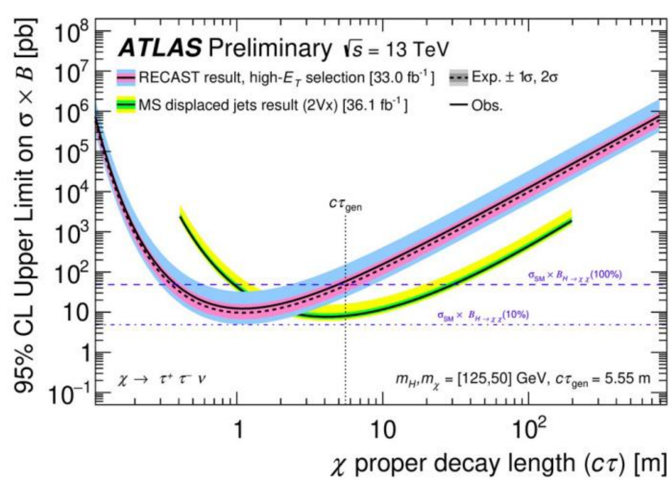
Reusable

Containerise once, reuse elsewhere. Cloud-native.

Free

Free Software. MIT licence. Made with ❤ at CERN.

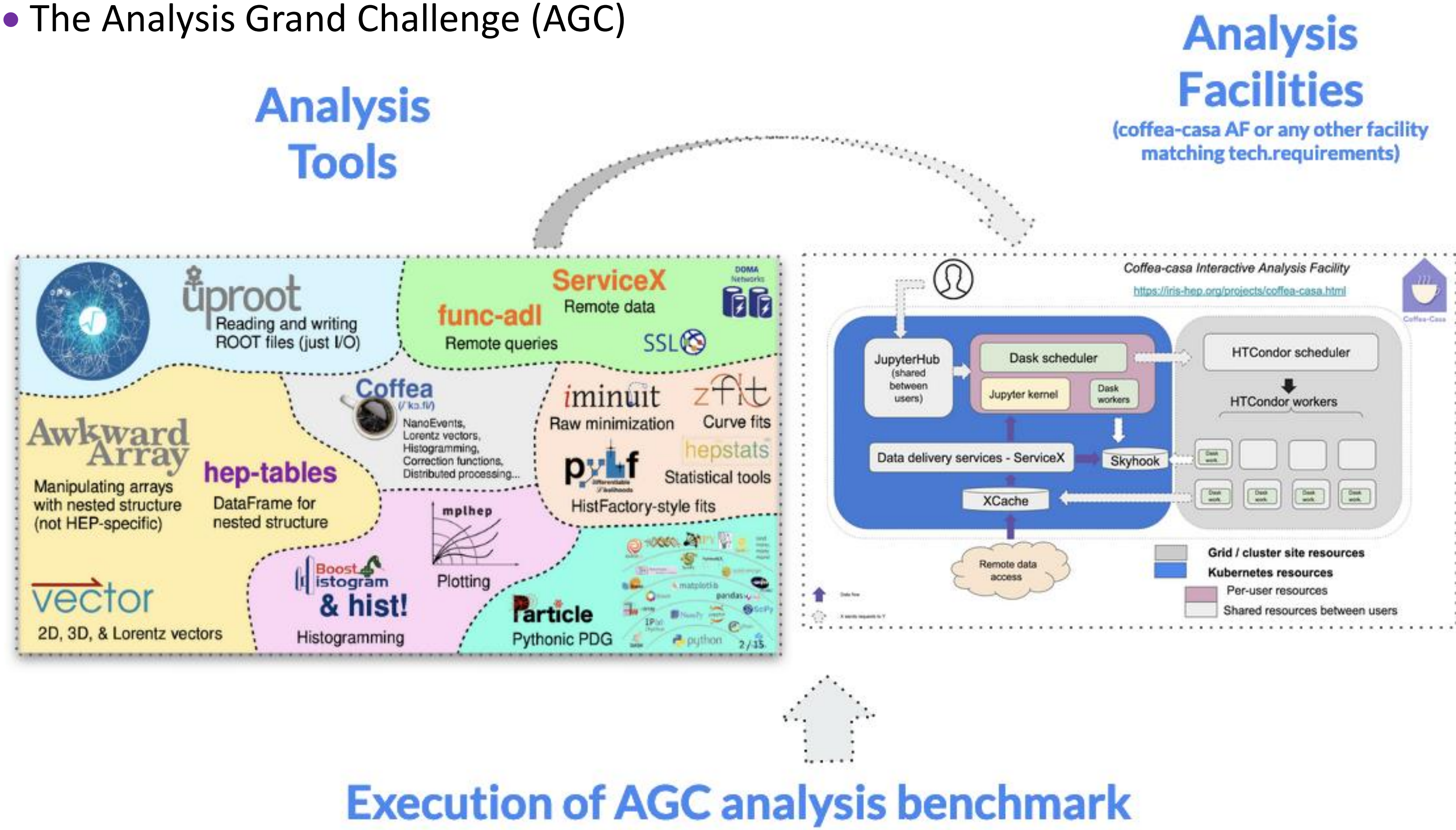
https://reanahub.io/



Tibor ˇSimko, CSC 2024

Software and data analysis

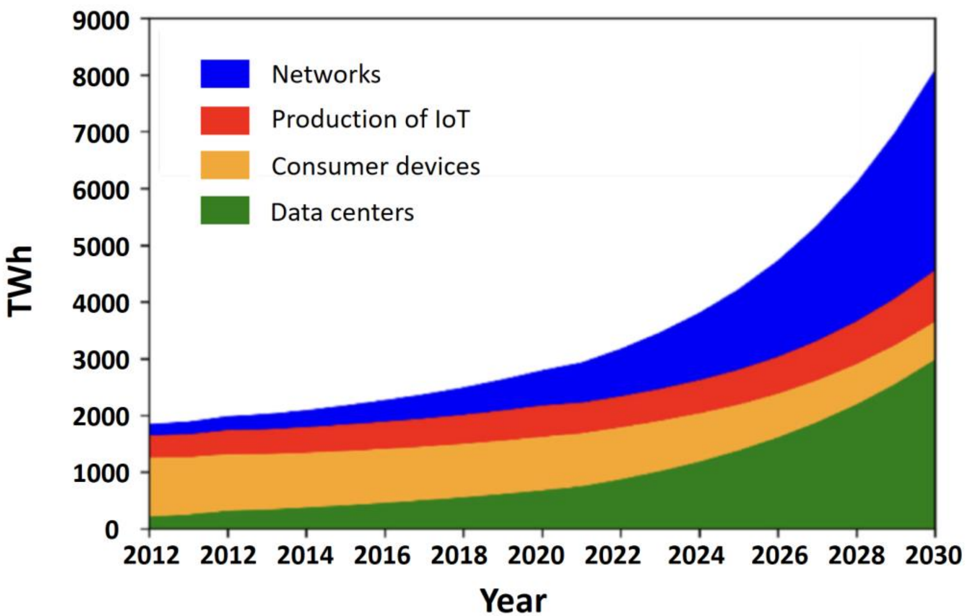
- The Analysis Grand Challenge (AGC)



https://iris-hep.org/projects/agc.html

Sustainability

- The energy consumption in data centers is rising significantly due to AI utilisation and other digital technologies.



★Energy consumption is an important figure of merit to consider!

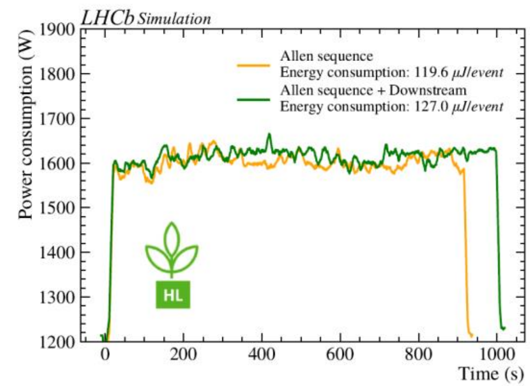
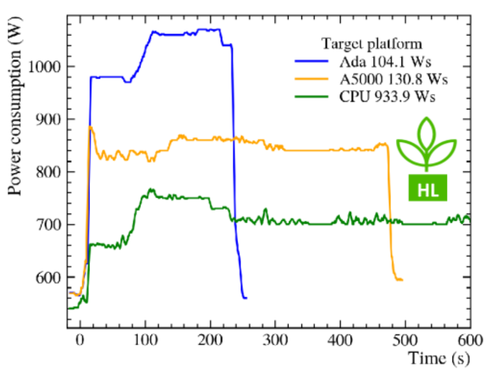
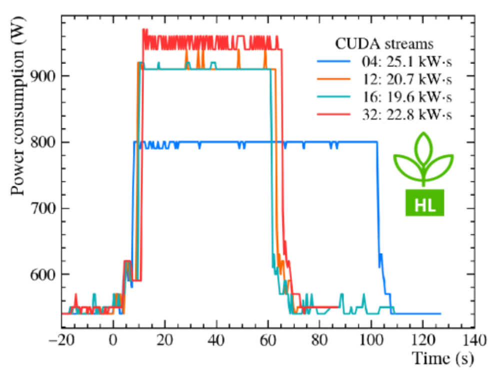


Fig. 17: Power consumption with **Allen** software running over 3.2M $B_s \to \phi\phi$ events without (blue) and with (orange) **Downstream** algorithm. The power consumption is measured using metered rack PDU AP8858EU3 with an average readout frequency of 2 Hz. The moving average filter with window of 20 points is applied. The measurements are obtained using the NVIDIA RTX 6000 Ada Generation GPU card.

[Com. and Soft. for Big Science; 9,10, (2025)]

- Dependence with hw and sw utilization!





★ New hardware platforms (FPGAs, optimized hybrid systems can help with that!)

COMCHA network

COMCHA
Computing challenges

RED2022-134769-T

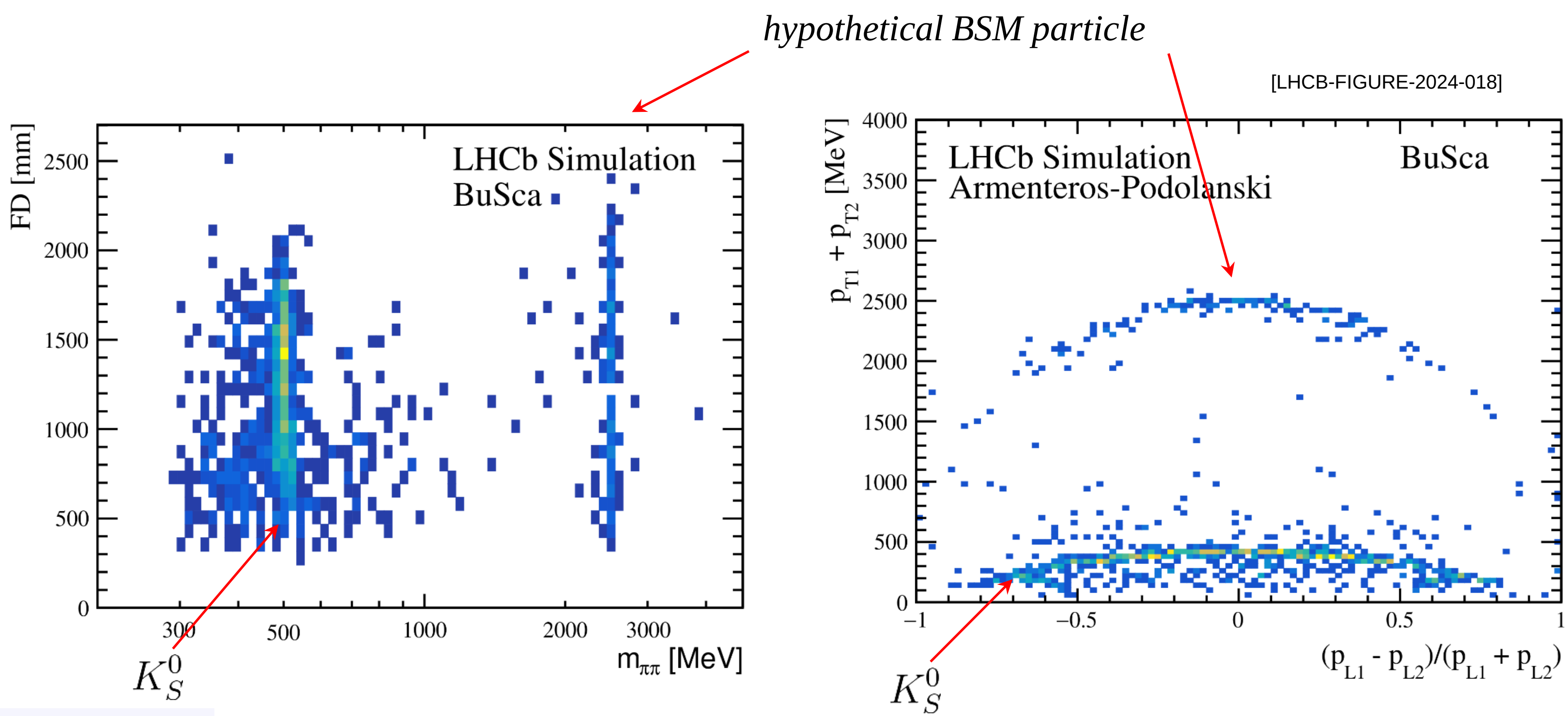
National Network on Computing Challenges

COMCHA network

Project:

BuSca: a *Buffer Scanner* at HLT1 to detect LLPs beyond the SM

Real time analysis prior any trigger decision!
→ Streamlined candidate reconstruction and selection with a neural network at 30 MHz!!

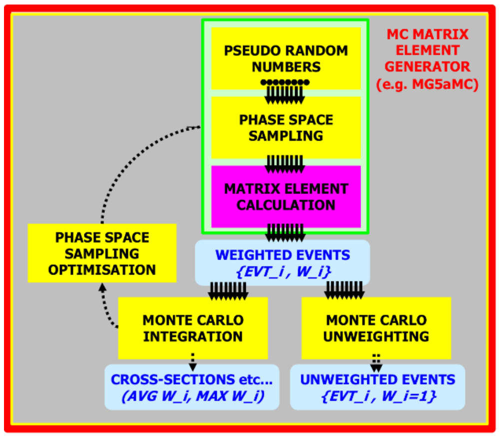


A. Oyanguren

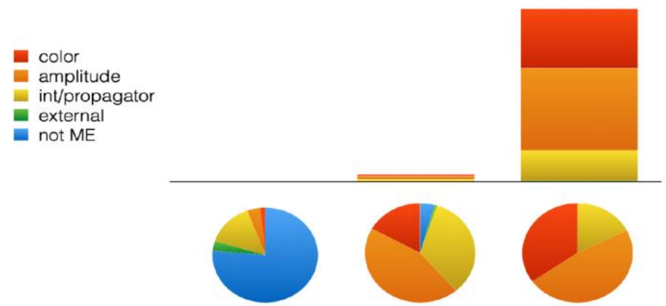
COMCHA network

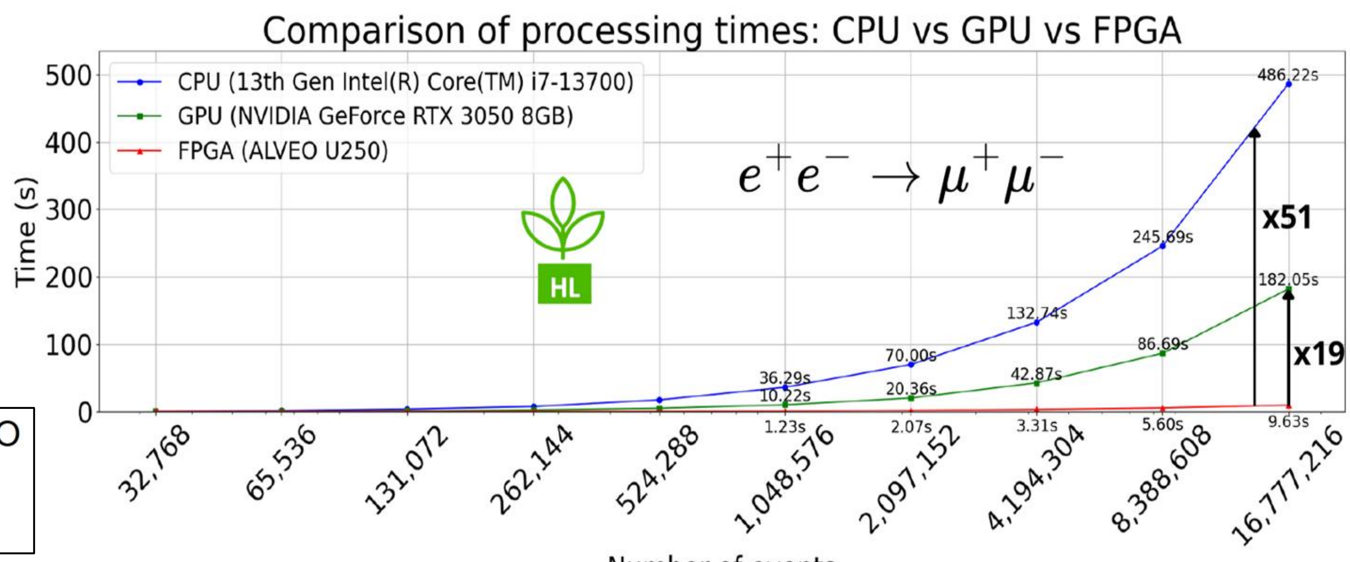
Project: HEP generators in hybrid architectures

Porting *MadGraph5*, a generator for SM and BSM phenomenology, to FPGAs



	$gg \to t\bar{t}$	$gg \to t\bar{t}gg$	$gg \to t\bar{t}ggg$
madevent	13G	470G	11T
matrix1	3.1G (23%)	450G (96%)	11T (>99%)



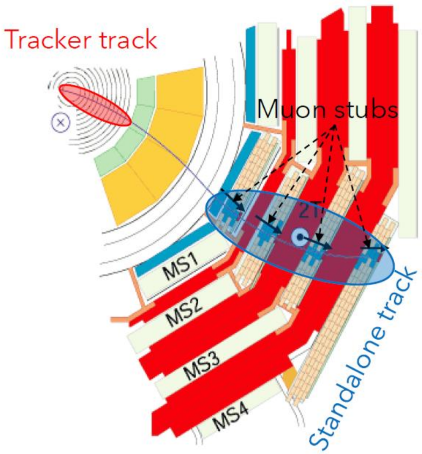


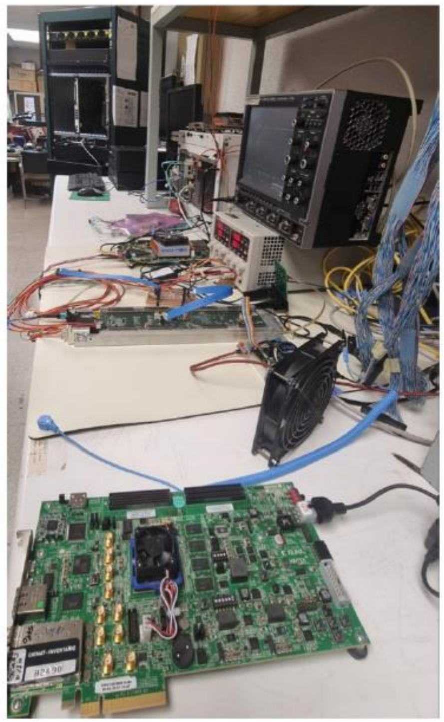
L. Fiorini,
S. Folgueras,
A. Oyanguren,
A. Valero

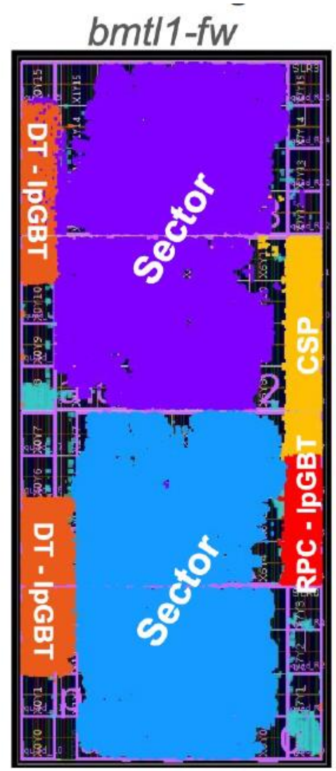
MG5_aMC@NLO

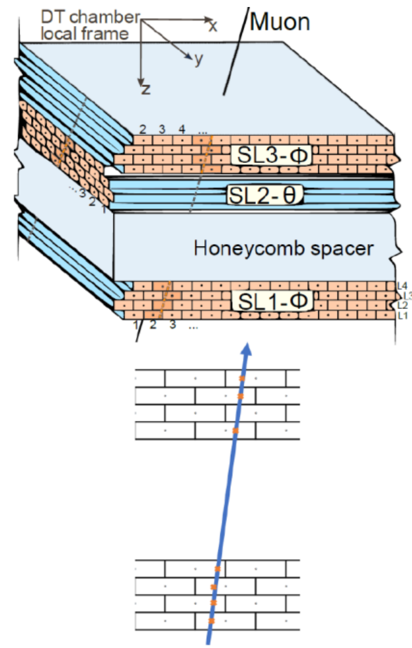
COMCHA network

Project: *Real-time muon tracking algorithm on FPGAs for the Upgrade CMS*









- Highly-optimized FPGA algorithms: High clock speeds (~500 MHz) with minimal resource utilization.
- AI utilization for FPGAs.
- Enable reconstruction for exotic signatures beyond capabilities of the current system: hadronic showers, slow-particles.

C. Fernández, S. Folgueras

COMCHA network

UNDER DEVELOPMENT

Project: Signal Reconstruction Algorithms in FPGAs with NN

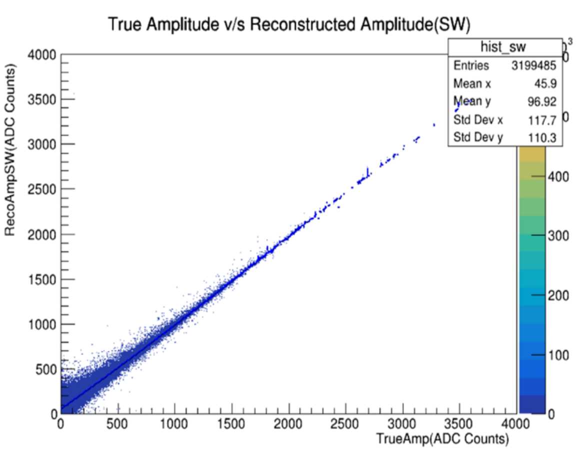
PyTorch, Keras → hls4ml → AMD Vivado, GHDL, OSVVM, V → cocotb, ROOT Data Analysis Framework

SW Dev & Train
- Developing and training of ML models in SW

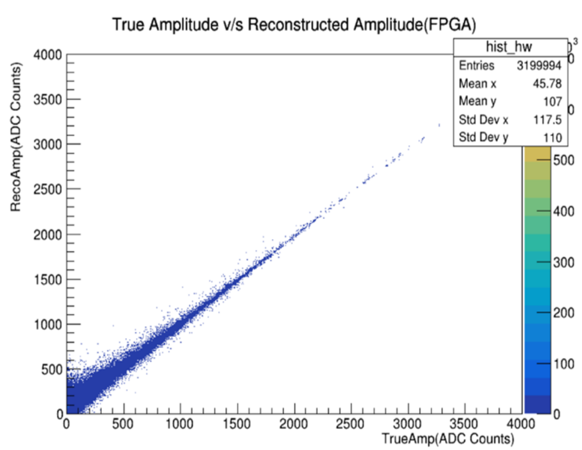
FW Prototyping
- Prototype models in FW
- Analysis of resources, bit widths and latencies

FW Fine-tuning
- Fine-tuning in RTL (VHDL) of the selected model
- Pre-post-processing
- Logic wrapping the ML model
- Formal verification

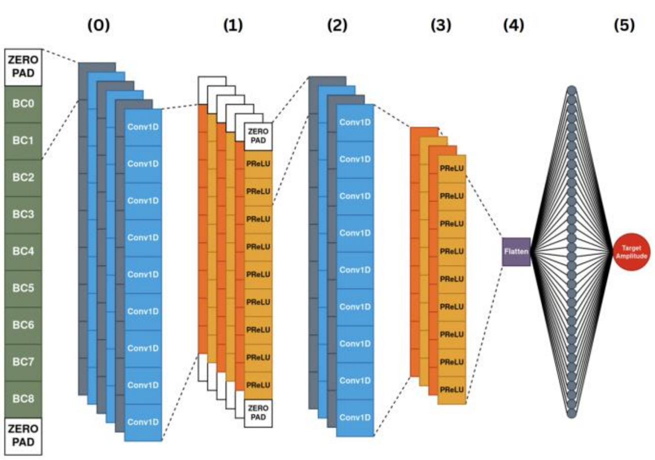
Physics Analysis
- FW/SW Cosimulation
- Physics data analysis
- Cosimulation made in containerised environment (Docker)



SW output: True vs Reconstructed Amplitude



FPGA output: True vs Reconstructed Amplitude



L. Fiorini, A. Valero

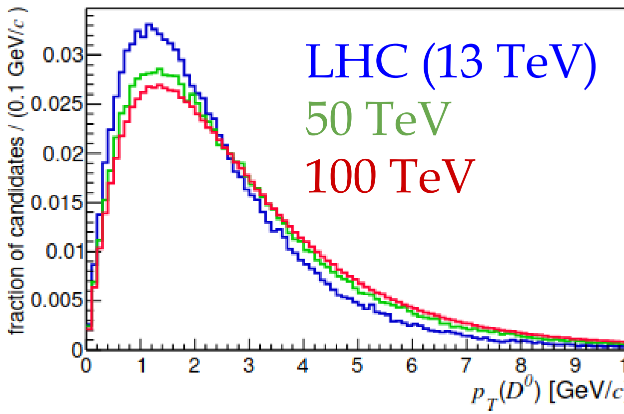
COMCHA network

UNDER DEVELOPMENT

UNIVERSIDADE DA CORUÑA

<u>Project:</u> *Improving Flavor Physics*

Improving trigger, reconstruction and flavor tagging at LHCb.



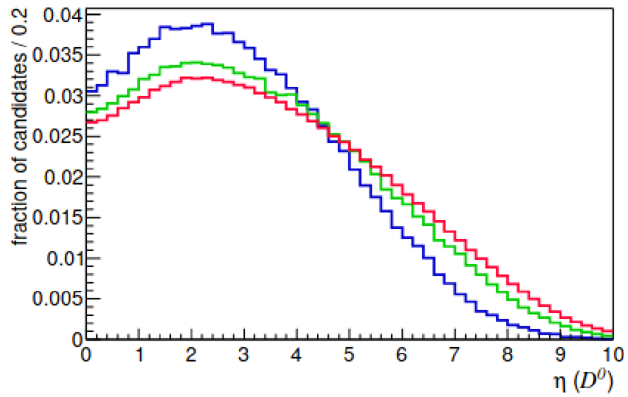
Key for a future FCC-hh!

- Very good geometrical acceptances (https://arxiv.org/abs/2012.02692), but big challenges:

- Higher multiplicities above higher pT's (https://link.springer.com/article/10.1140/epjc/s10052-019-6904-3)

  → More background, more challenging to trigger: Use handles like particle detachment, higher computing power.

  → Flavor tagging also more challenging → full event interpretation like recent LHCb's inclusive flavor tagging (https://arxiv.org/abs/2508.20180) would be more important than ever.



D. Martinez Santos, V. Chobanova

COMCHA network

Collaboration, training and skill development in Spain

Barcelona (URL)

Valencia (IFIC)

U. Oviedo

COMCHA Computing challenges

To subscribe:

comcha@pegaso.ific.uv

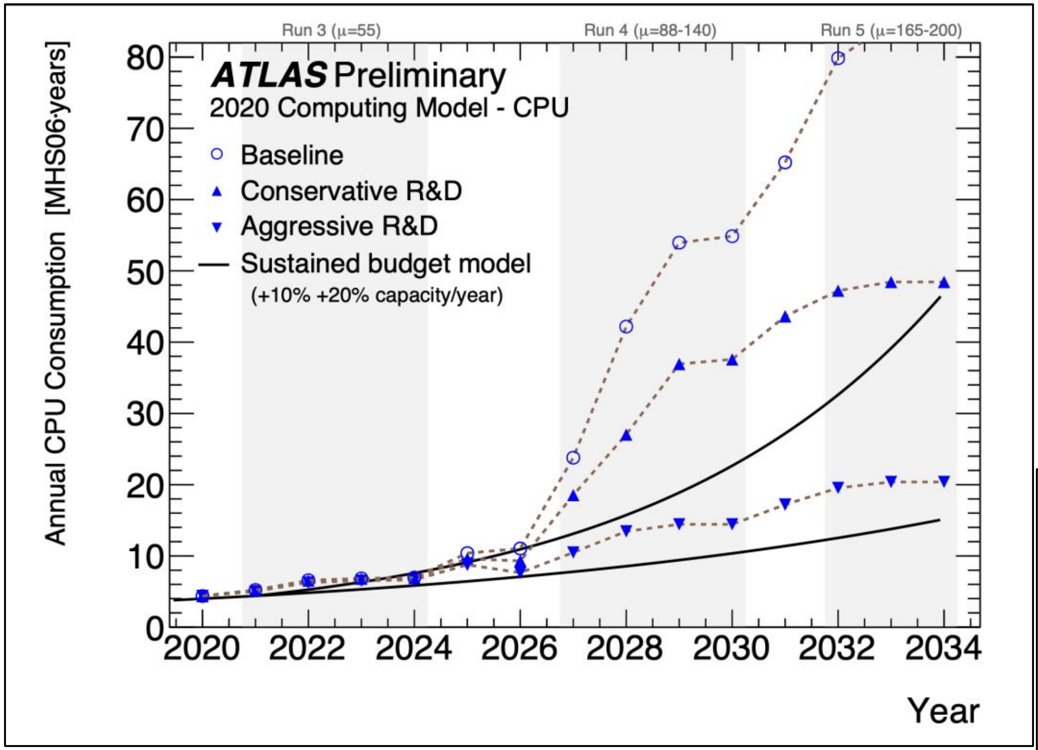
- Training on advanced computing.
- Grants for students (HEP-IRIS, GSoC, AIHUB)
- Promote collaborative projects (Funding!: Transición Digital, AI, EU ALIES)
- International visibility (HSF, HEP-IRIS) (conferences)

Contact:
Arantza.Oyanguren@ific.uv.es
Luca.Fiorini@ific.uv.es

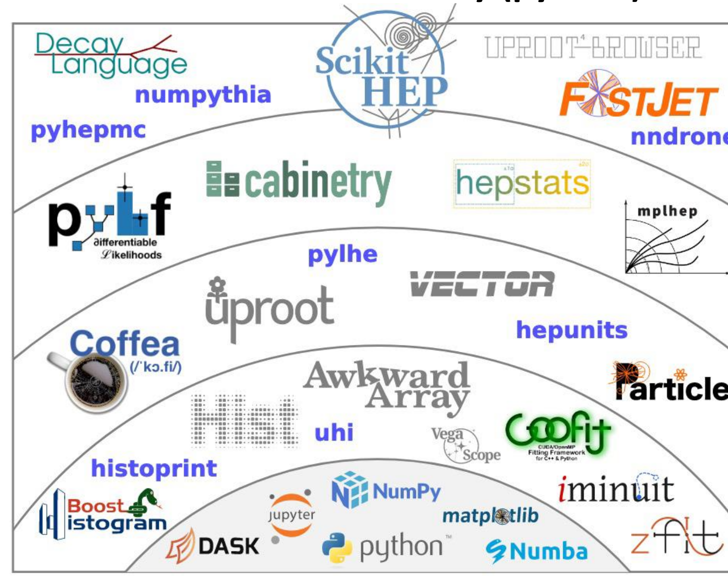
Conclusions

- Prospects for the next decade:

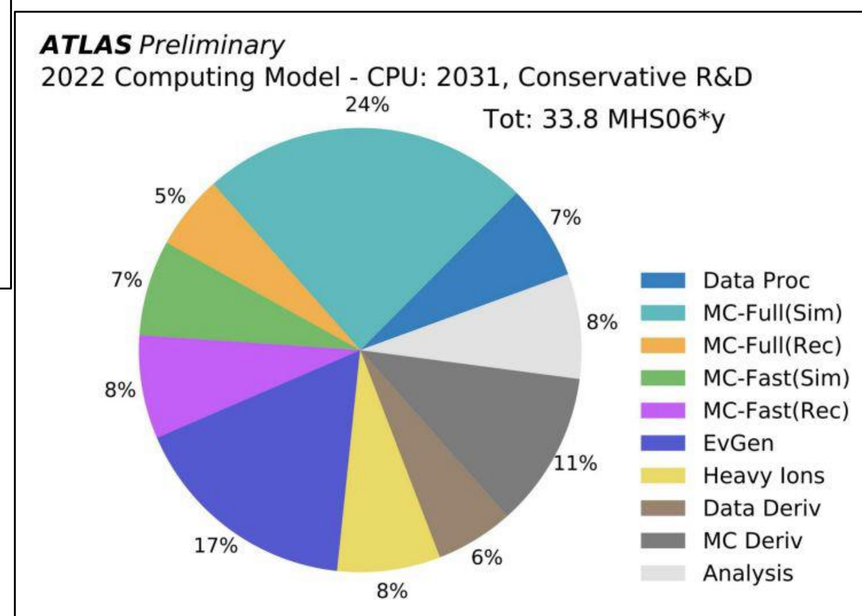
★ **Huge data volumes**



★ **Software diversity (python)**



★ **Computing tasks**



★ **Hardware tendencies**

- ARM at least 30% more energy efficient on HEP workloads
- Hybrid systems including GPUs and FPGAs

Conclusions

“We are in the era of computing, and its transformative power will be decisive for the success of future colliders.”

ChatGPT 5

COMPCHA
Computing challenges