Financiado por la Unión Europea NextGenerationEU
Gobierno de España – Ministerio de Ciencia, Innovación y Universidades
Plan de Recuperación, Transformación y Resiliencia
Generalitat Valenciana – Conselleria de Educación, Universidades y Empleo
GVANEXT

ASFAE
ASTROFÍSICA Y FÍSICA DE ALTAS ENERGÍAS

Computación Avanzada para el procesado intensivo de Big Data en ATLAS

*Operación del experimento ATLAS durante el Run 3 del LHC y explotación de sus datos para el estudio del bosón de Higgs y el quark top.*

*Aplicación de técnicas de machine learning al análisis de datos del LHC: integración de GPU y FGPA en la Grid; aplicación al trigger de estos procesadores, entre otras aplicaciones.*

Retos tecnológicos para el descubrimiento con el detector LHCb mejorado del CERN

# Computing and AI for collider experiments

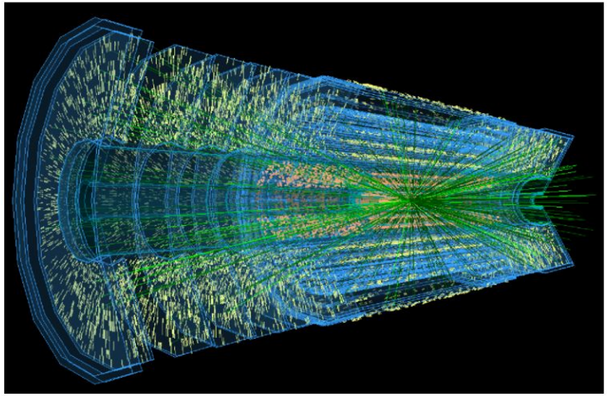
Miguel Villaplana Pérez

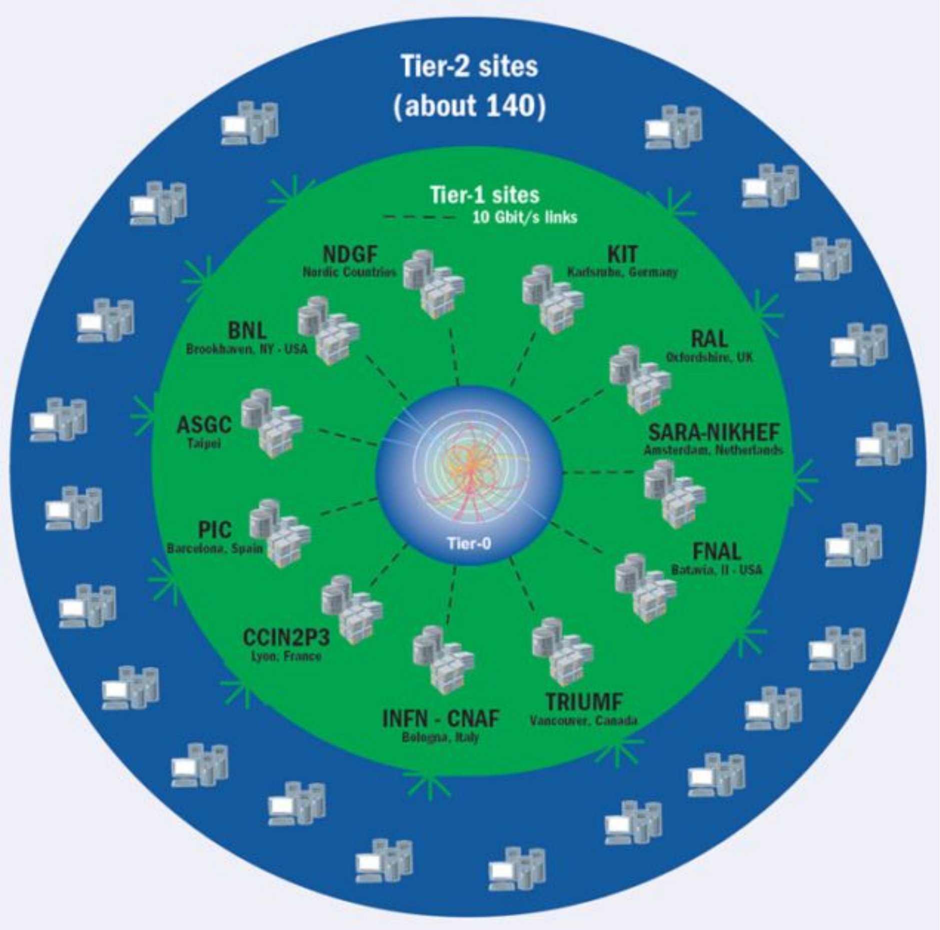
Instituto de Física Corpuscular

June 2024, Zaragoza

CSIC Consejo Superior de Investigaciones Científicas
Vniversitat de València
IFIC Institut de Física Corpuscular
UC Universidad de Cantabria
IFCA Instituto de Física de Cantabria

# LHC computing model






The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.
Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Plan de Recuperación, Transformación y Resiliencia
GENERALITAT VALENCIANA
Conselleria de Educación, Universidades y Empleo
GVANEXT

# Spanish computing sites for the LHC (WLCG-ES sites)

- Tier 1 (ATLAS, CMS, LHCb):
  - PIC-Barcelona
- Federated Tier2s
  - ⇨ 60% IFIC-Valencia
  - 25% IFAE-Barcelona
  - 15% UAM-Madrid
  - 75% Ciemat-Madrid
  - ⇨ 25% IFCA-Santander
  - 50% USC-Santiago
  - 50% UB-Barcelona*

*UB Tier2: Now it is working as a Tier3 where MC simulation is running

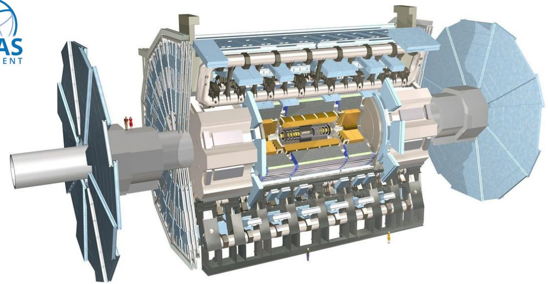

- LHC sites pledges in the last 5 years:

https://wlcg-cric.cern.ch/core/federation/list/

- **Integrated in the WLCG project (World Wide LHC Computing GRID) and strictly following the experiments computing models.**
- **We represented the 4-5% of the total Tier-2s and 5% of the total Tier-1s resources, with the budget reduction now the 3% for Tier2s and 4% for Tier1!!!"**


The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.
Financiado por la Unión Europea NextGenerationEU
Gobierno de España — Ministerio de Ciencia, Innovación y Universidades
Plan de Recuperación, Transformación y Resiliencia
Generalitat Valenciana — Conselleria de Educación, Universidades y Empleo
GVANEXT — Fondos Next Generation en la Comunitat Valenciana

# ATLAS Tier-2 @ IFIC



- Computing infrastructure for ATLAS Tier2 is dedicated to storage and data processing for the ATLAS experiment at LHC at CERN
- IFIC dedicates 61 computing servers with **4384 cores**, 15 storage servers with **4 PBytes** of disks, and several machines for file transfer control, information systems and management
- International collaboration with high standards on availability, storage, processing, monitoring and updates

- At the top of availability and reliability ranks: IFIC Tier2 is a Nucleus
  - Nucleus are Tier2s with a big amount of storage and very good network connection, passing job production on to smaller Tier2s (Satellites)
- Now we have RedIRIS-Nova at 100 Gbps -> IFIC’s WAN connectivity increased to 100 Gbps
- IFIC Tier 2 by the numbers:
  - We have processed more than 300 Bill. events in these last 5 years
  - Steady state of more than 5.000 running job slots since 2019, typically using 2GB per job slot
  - Mainly running with either 8 or 1 cores (“multi-core” or “single-core”) per job, depending on type of job

- **IFIC has 60% of the Spanish ATLAS Tier2 resources**
  - Pledges (C-RRB 2022/04):
  - **Current status:**
    - CPU (HS06): 44768 (28% come from machines > 5 years old)
    - Disk (TB): 3402 (60% of of storage resources at IFIC older than 6 years!)

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

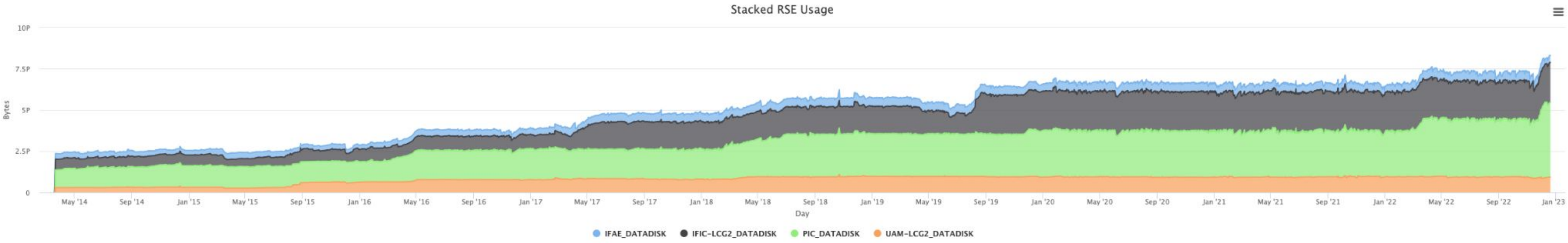
Gobierno de España · Ministerio de Ciencia, Innovación y Universidades

Plan de Recuperación, Transformación y Resiliencia

Generalitat Valenciana · Conselleria de Educación, Universidades y Empleo

GVA NEXT

# Storage: overview




- ~13 PB of ATLAS data stored in the Spanish Tier 1 and Tier 2 centers
  - > 4 PB stored at IFIC!

- However, 60% of storage resources at IFIC older than 6 years!
  - New servers currently being installed at IFIC (800TB)
  - Broken disks: 61 in 2021; 64 in 2022
  - Older disks are a data loss risk
  - Drop of support of older servers in new versions of kernel

➯ **A computing engineer hired with ASFAE funds to support the infrastructure**

➯ **Currently in the process of buying new storage and CPU resources with ASFAE funds (~160k€)**

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

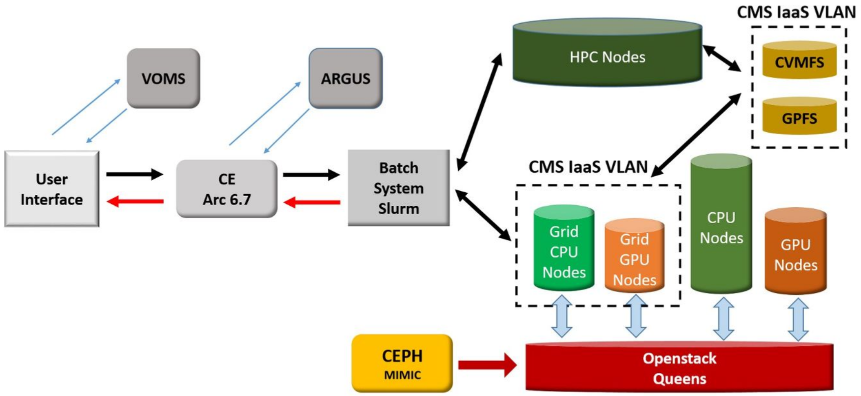
Plan de Recuperación, Transformación y Resiliencia

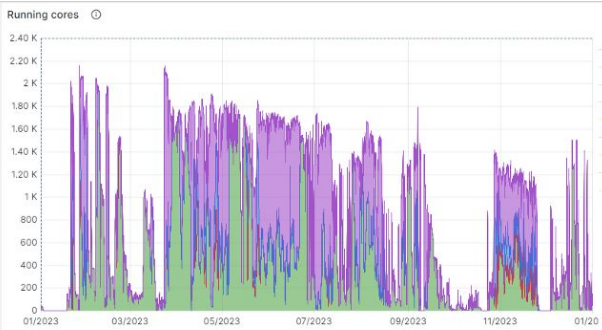
GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo

GVANEXT Fondos Next Generation en la Comunitat Valenciana

# CMS Tier-2 @ IFCA

- The IFCA Tier2 is implemented on the Opensource Suite of Cloud OpenStack.
    - Integrated with the rest of the IFCA computing infrastructure.
- **IFCA provides a IaaS (Infrastructure as a Service) to the Tier2 project of CMS.**
    - Allows to easily benefit from already deployed services.
- Different resources can be used and shared through the BatchSystem:
    - Grid Worker nodes (IaaS).
    - **GPU** nodes can also be served by the cloud system (IaaS).
    - Opportunistic running on the **HPC Altamira** node.
- Worker Nodes are cloud machines building singularity containers to run CMS jobs.
    - CMS software loaded through cvmfs cache.
    - Output is stored in GPFS distributed file system.
    - Containers deleted after execution.
- CE takes care of the User Subject, Group or Role, and mapping to a defined queue at arc.conf file.




| Name | Mean | Last * | Max | Min | Total |
|---|---|---|---|---|---|
| Analysis | 462 | 44.8 | 1.98 K | 0 | 4.05 Mil |
| Cleanup | 0.0281 | 0 | 16.6 | 0 | 246 |
| Harvesting | 0.000251 | 0 | 1.60 | 0 | 2.20 |
| LogCollect | 0.209 | 0 | 59.6 | 0 | 1.83 K |
| Merge | 0.315 | 0 | 52.6 | 0 | 2.76 K |
| Processing | 40.3 | 0 | 1.86 K | 0 | 353 K |
| Production | 373 | 0 | 2.10 K | 0 | 3.27 Mil |
| TOTAL | 876 | 44.8 | 2.16 K | 0 | 7.67 Mil |

In the process of hiring a computer scientist.

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

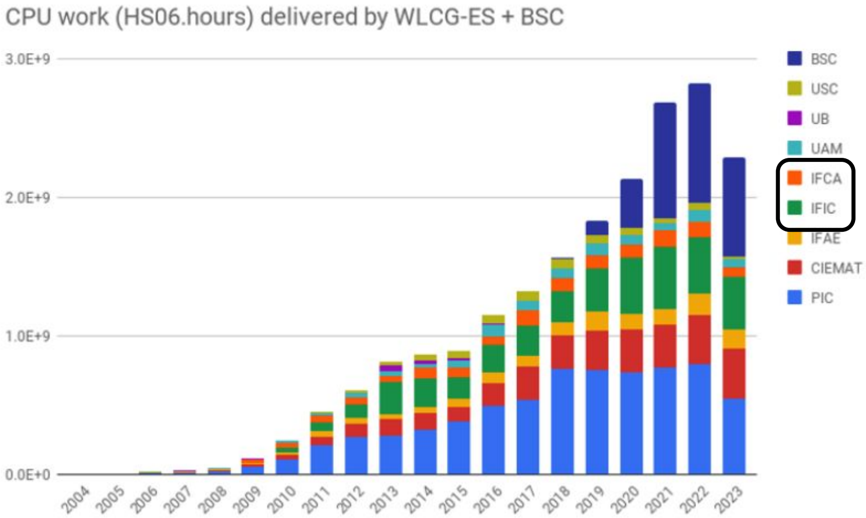
GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

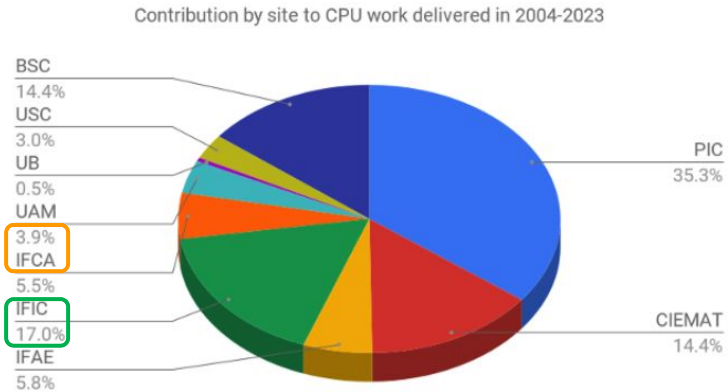
Plan de Recuperación, Transformación y Resiliencia

GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo

GVANEXT Fondos Next Generation en la Comunitat Valenciana

# CPU delivered by Spain to WLCG




~2000 Million hours delivered during 2004-2023

(20x10$^{9}$ HS06.hours; average CPU core power ~10 HS06)

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

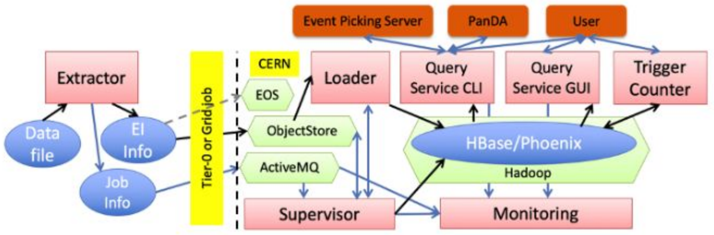
Plan de Recuperación, Transformación y Resiliencia

GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo

GVANEXT Fondos Next Generation en la Comunitat Valenciana

# IFIC Event Index

- A system designed to be a complete catalogue of ATLAS events, real and simulated data.
- Partitioned architecture, following data flow:
  - **Data Production:** extract event metadata from files produced at Tier-0 or on the Grid
  - **Data Collection:** transfer EventIndex information from jobs to the central servers at CERN
  - **Data Storage**: provide permanent storage for EventIndex data and fast access for the most common queries.



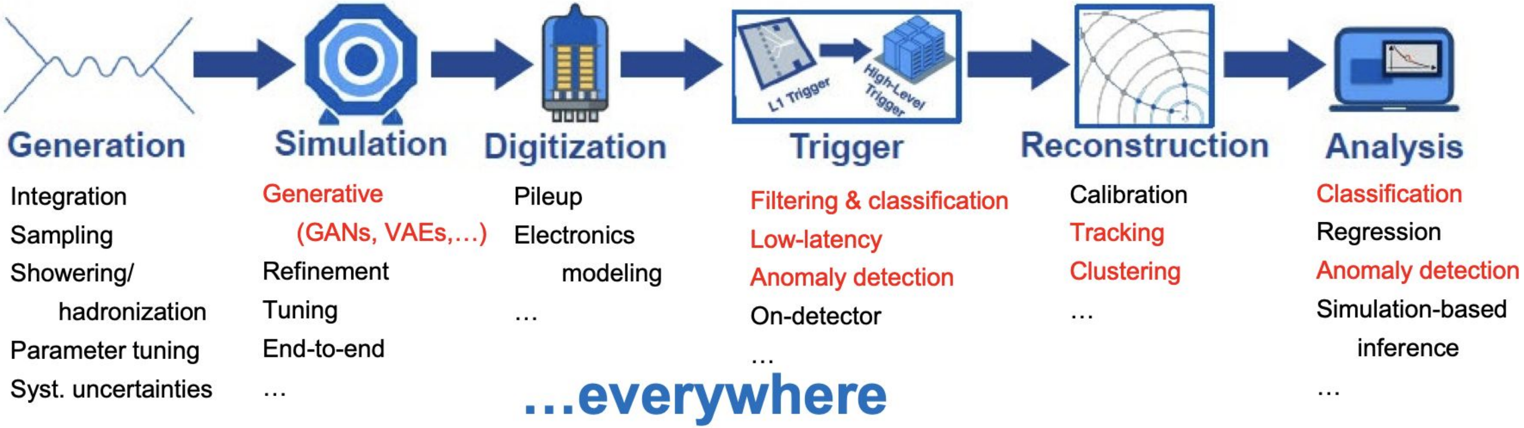

| Data Production | Data Collection | Storage & Queries |
| --- | --- | --- |
| • Extract event metadata by means of Tier 0 and Grid jobs.<br>• Store production independently of ATLAS DDM.<br>• Coordinated by **IFIC**. | • Collect, validate and ingest production.<br>• Designed, developed and maintained by **IFIC**. | • Permanent storage.<br>  ○ Hbase/Phoenix.<br>• Client APIs and tools.<br>• Storage and Trigger Counter designed, developed and maintained by **IFIC**. |

- Use Cases:
  - Event Picking.
  - Production consistency checks (Duplicate event and overlap detection).
  - Trigger checks and event skimming
- The new system is in operation since last Spring 2022 and performing excellently.

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.
Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Plan de Recuperación, Transformación y Resiliencia
GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo
GVANEXT Fondos Next Generation en la Comunitat Valenciana

# AI in collider experiments

- Main objectives of today's Particle Physics program are:
  - Probing the SM with increasing precision ⇒ search for anomalies as evidence for BSM physics
  - Searching directly for BSM physics
- Require the processing, identification, storage and analysis of rare and/or complex signals hidden in immense amount of data (background)
  - Eg. at the LHC, ~99.999% of the data has no interest
- ML used since the 80's & 90's for (offline) event and particle identification, energy estimation, flavor tagging
- Since then, hardware and software technology progress lead to extensive HEP R&D adaptations and applications…




The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA · MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Plan de Recuperación, Transformación y Resiliencia

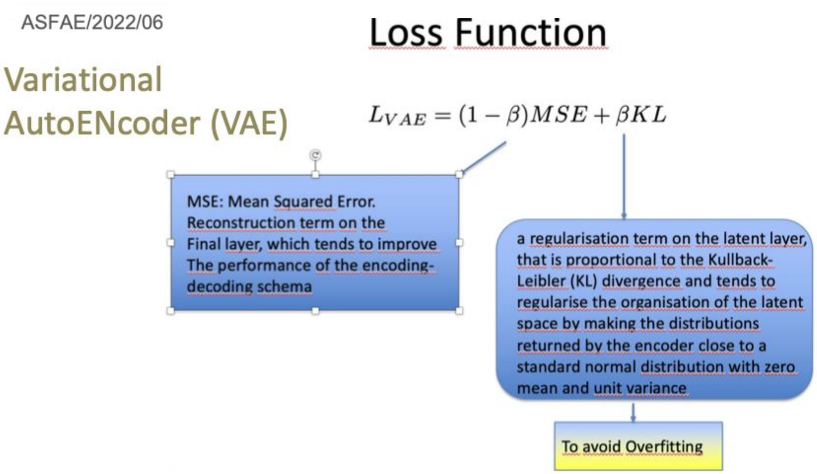
GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo

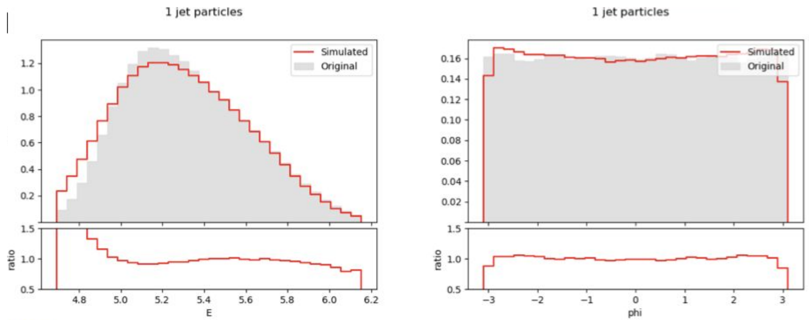
GVANEXT Fondos Next Generation en la Comunitat Valenciana

# Generative models for simulation

- Classical simulation of proton-proton collisions implies: PS generation, Hadronization/fragmentation, Pass the particles through the detector
- Time consuming and expensive
- Alternative are high fidelity fast generative models, eg. GANs, VAEs
  - Able to sample high dimensional feature distributions by learning from existing samples

Variational AutoEncoder (VAE) for pp->ttbar with 6 jet in final state (IFIC)




E, $\phi$ of first jet using $\beta$-VAE with $\beta$=0.001

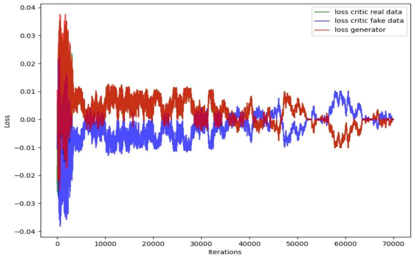
CHEP2023

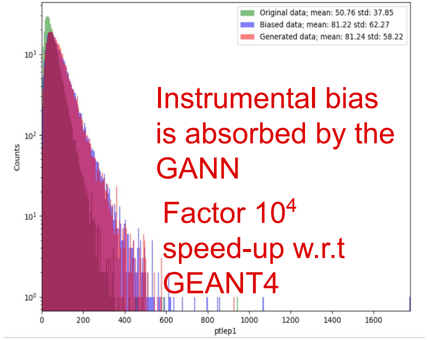
The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Plan de Recuperación, Transformación y Resiliencia
GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo
GVANEXT Fondos Next Generation en la Comunitat Valenciana

# Simulation and ML-oriented detector design at IFCA

- Realistic CMS MC simulation using Wasserstein Generative Adversarial Neural Networks (WGANN)



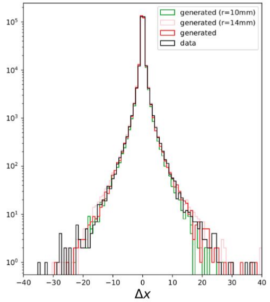



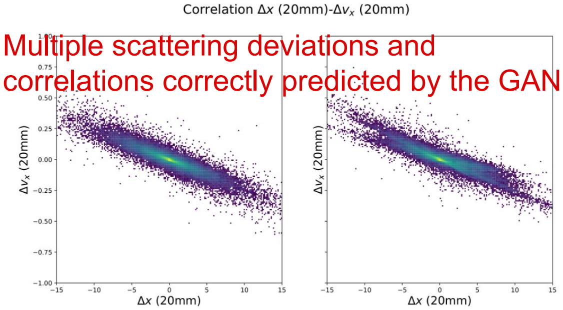

Instrumental bias is absorbed by the GANN

Factor $10^4$ speed-up w.r.t GEANT4

http://hdl.handle.net/10261/279966

- Ultra-fast muon propagation simulation using GANNs



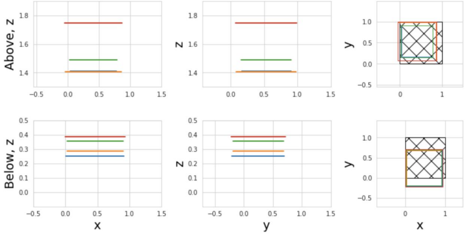




Multiple scattering deviations and correlations correctly predicted by the GANN

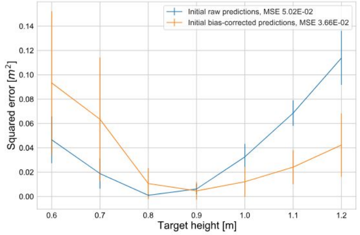
https://doi.org/10.31526/jais.2024.487 Factor $10^4$ speed-up w.r.t GEANT4

- Machine-Learning Oriented Design of Experiments using differential programming (MODE Collaboration)
- Optimization of a Muon detector setup for industrial application of muon tomography

Z position of active planes




Target height optimization




Differential Programmig workflow

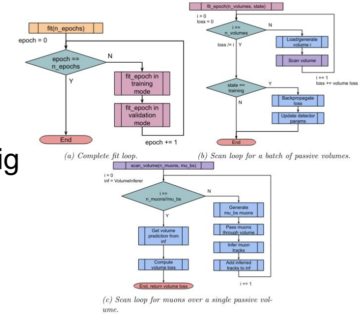
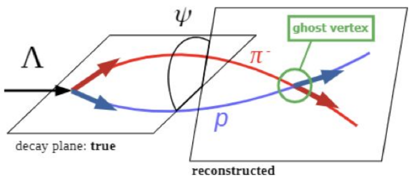


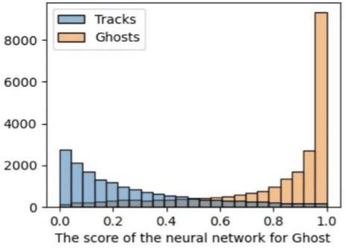
Figure 6: Breakdown of the fitting procedure of detectors in TomOpt

https://arxiv.org/abs/2309.14027 (Accepted in Machine Learning: Science and Technology)

# NN and BDTs in the LHCb Trigger

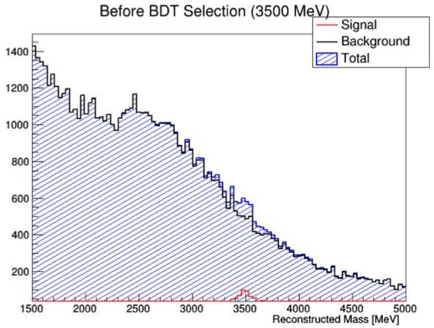
- Ghost killer: Removal of fake tracks originating from spurious hits in the detector
  - NN with single hidden layer (14 nodes) & 8 features, trained on MC

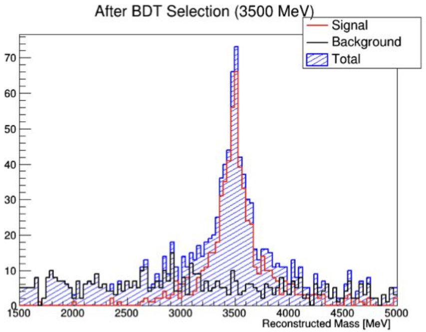


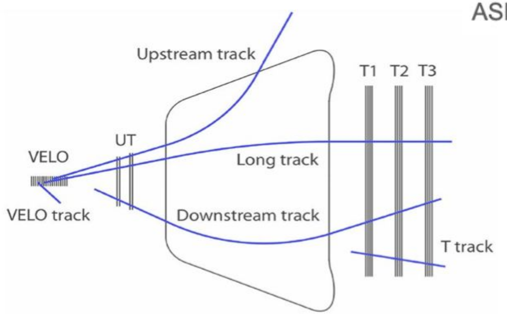

- Offline/HLT2 combinatorial background suppression

Example: search for dark Higgs in $B^+ \to K^+ H(\to \mu^+ \mu^-)$ decays

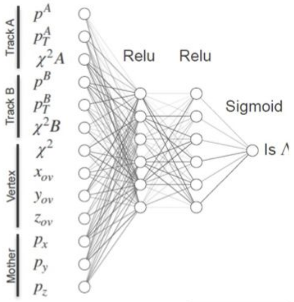


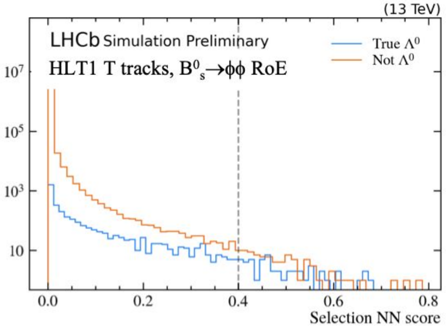




- T tracks:




  - HLT1: NN for track pair selection



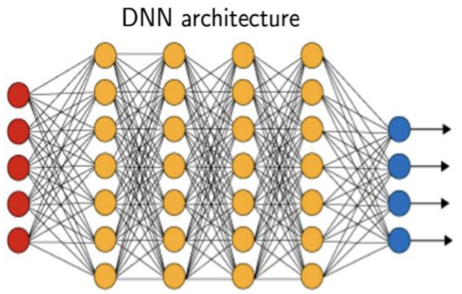

CUDA implementation in place, optimization ongoing

  - HTL2: Binary CatBoost BDT for track filtering

- BDT for offline ghost vertex reduction

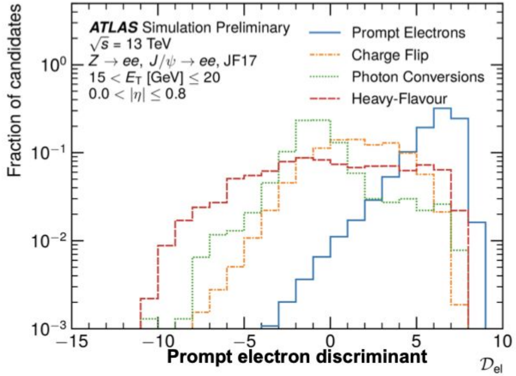
# DNN for electron ID in ATLAS

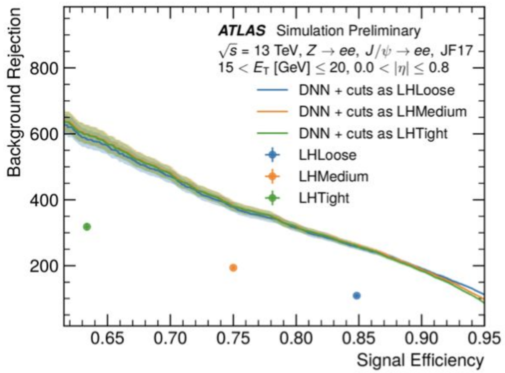
- Architecture and configuration detailed in ATL-PHYS-PUB-2022-022
- Neural Networks are powerful in signal to background discrimination → Replacing previous electron identification based on a Likelihood-based approach (LH), while using similar high-level input variables as LH




- **DNN architecture**
  - 5 hidden layers with 256 nodes each, activation function: leaky ReLU, batch normalization
- **Output layer**
  - Six outputs (signal + 5 background classes) with softmax activation for multiclass classification
- Outstanding discrimination, flexible discriminants built out of the DNN scores
- Discriminating performance comparing signal efficiency ($\varepsilon$) and background rejection ($1/\varepsilon$)
- At ~75% of signal efficiency, DNN outperforms LLH by a factor >2 background rejection

| Class | Description |
|---|---|
| Prompt Electrons (El) | Prompt isolated electrons coming from Z, W and J/psi |
| Charge Flip (CF) | Prompt with incorrectly reconstructed charge |
| Photon Conversion (PC) | Electrons coming from prompt photons |
| Heavy-Flavour (HF) | Electrons coming from a b- or c- hadron decay |
| Light-Flavour e/gamma | Electron coming from a u-, d- or s- hadron |
| Light-Flavour Hadrons | Undecayed hadrons |




The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Plan de Recuperación, Transformación y Resiliencia

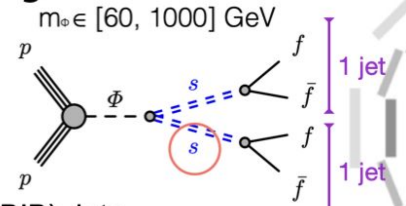
GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo

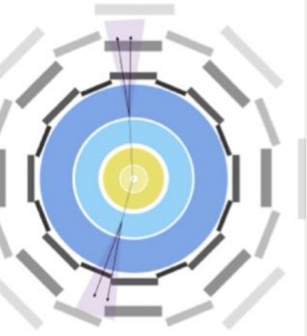
GVANEXT Fondos Next Generation en la Comunitat Valenciana

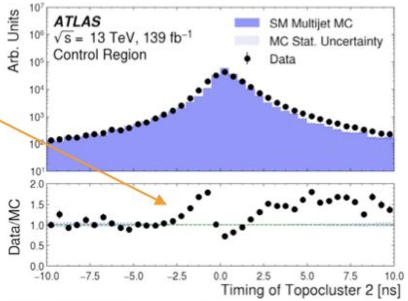
JHEP06(2022)005

# LLPs decaying into displaced hadronic jets

- Hidden Sector with a heavy boson decaying to long-lived scalars
- Signature: **2 displaced jets** in the ATLAS calorimeter
- Adversarial Neural Network separates signal from two types of bkgs
  - training on a set of signal MC, SM background MC and beam-induced background (BIB) data
  - mismodelling in input variables (eg. jet timing) had a big impact in the data/MC behaviour of the NN scores
  - Adversarial NN running in a control region avoids the NN to use data/simulation differences

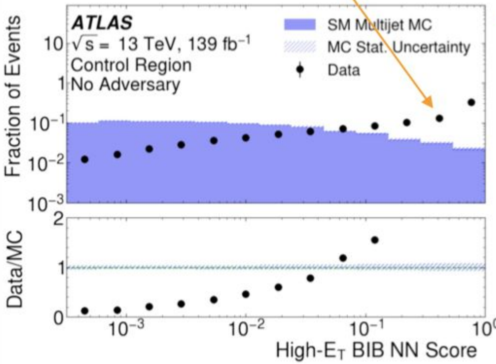


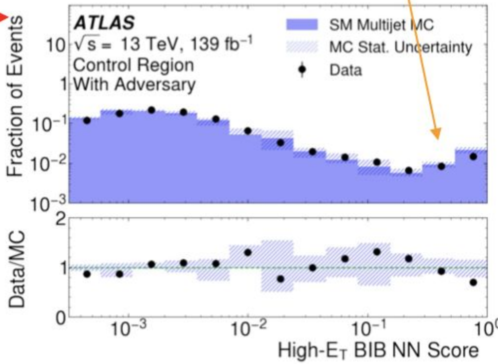



**mismodelling in jet timing**

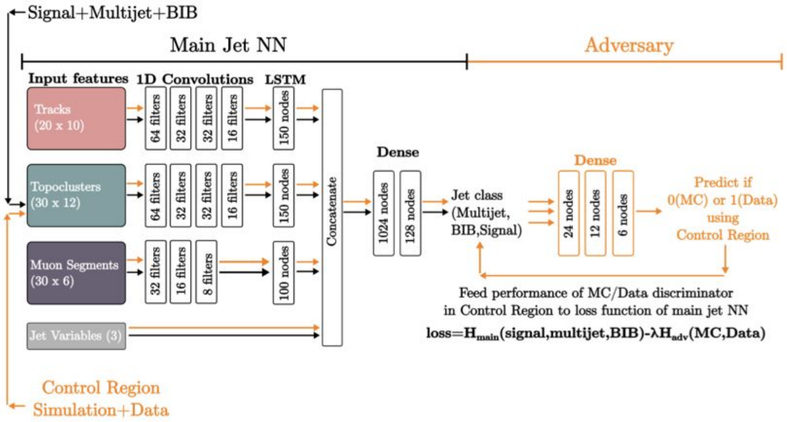
**large impact in the BIB NN score**




**adding Adversarial NN**

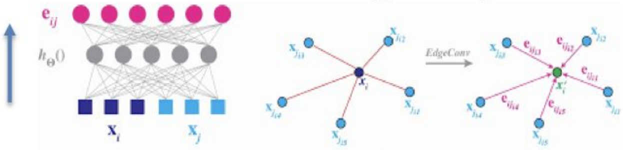



**impact in the BIB NN score becomes tiny**



# Anomaly detection

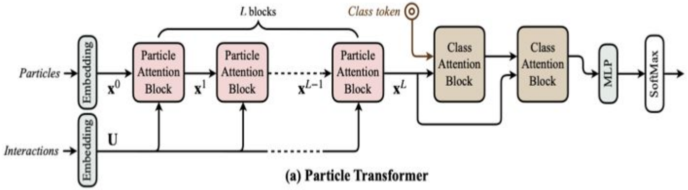
- Much progress in recent years developing powerful architectures for supervised DL tasks, such as
  - **ParticleNet (PNet)** Graphs that enhance the correlations among closest neighbours (pairwise features)




  - **Particle Transformer (ParT)** The self-attention mechanism allows to focus on the relevant correlations among the different objects of the event

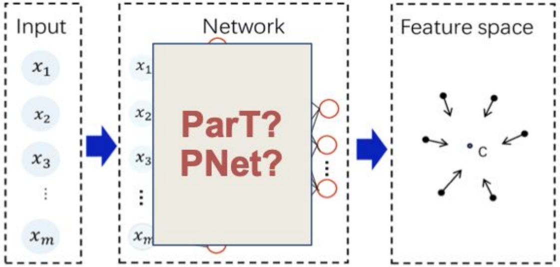
Four-momentum

Higher-level features $m_{ij}^2$, $\Delta R$



(a) Particle Transformer

- Support Vector Data Description technique (SVDD) can be used to adapt any classifier into an anomaly detector
  - Add an output layer (output space)
  - Take Loss function as distance to a center in the output space
  - Background events get closer to the center




- Are the best-performing classifiers also the best anomaly/outliers detectors?
  - Trained only with "background" events (unsupervised)
- ***DarkMachines collab.*** already explored anomaly detection in the HEP context
  - PNet and ParT not explored yet for exotic searches
- Use pp collisions at 13 TeV
  - Detector simulation performed with Delphes 3 using a simplified ATLAS detector card
  - **Input variables**: four-momentum and type of objects, and the missing transverse momentum of the event

**Two other techniques are also being explored to adapt any classifier, known as *DROCC* and *Smearing***

**Work still in progress but results look promising**

# Summary

### Computing Infrastructure for ATLAS and CMS

- Excellent performance of IFIC and IFCA' Tier2
- ASFAE funds invested in
  - Further development of the computing infrastructure for next years' pledges (HL-LHC)
  - Replacing decaying infrastructure and future points of failure in storage
  - Computing engineer for support of Valencian infrastructure

### Artificial Intelligence in collider experiments

- ML/DL is a fundamental tool everywhere in today's (and future) HEP experiments
- Discussed a wide range of applications at CMS, ATLAS and LHCb experiments within the framework of the ASFAE projects
- Of critical importance for fully exploiting the physics potential of LHC during Run 3 and beyond

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Plan de Recuperación, Transformación y Resiliencia

GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo

GVANEXT Fondos Next Generation en la Comunitat Valenciana

# Thanks

IFIC
INSTITUT DE FÍSICA CORPUSCULAR

# LHC HL-LHC Plan and estimation

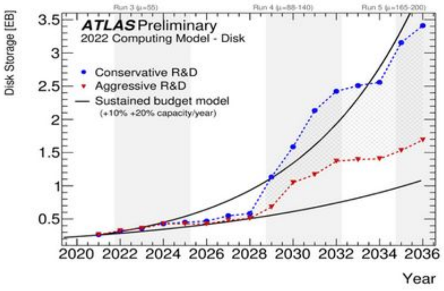


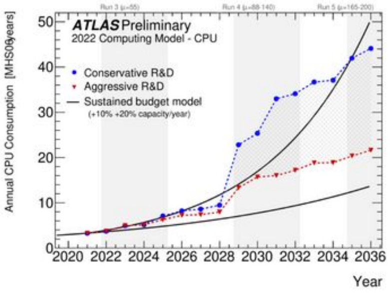

The original schedule for HL-LHC had pileup = 200 already in 2027, a large jump.

Revised schedule from January 2022:

- First full year of HL-LHC running at PU=140 is 2030.
- PU=200 reached only in Run 5

⇨ Nevertheless, x10 more events, bigger and more complex events (x5) -> **Unprecedented challenge!**




- **ATLAS Flat Budget projections** from 2022 show that future needs
  - Optimization (both speed and flexibility) of the experiment (e.g.reconstruction, simulation) and non-experiment (e.g. generation) software
  - Optimization of the available hardware infrastructure usage
  - Storage is the most expensive resource to deploy and operate

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Plan de Recuperación, Transformación y Resiliencia
GENERALITAT VALENCIANA
Conselleria de Educación, Universidades y Empleo
GVANEXT
Fondos Next Generation en la Comunitat Valenciana

# WLCG-ES: a success story

- Two decades contributing to LHC distributed computing infrastructure (Worldwide LHC Computing Grid, WLCG) and R&D at the highest level
  - ~5% of WLCG resources (20k CPU-cores, 15 PB disk, 20 PB tape), ~1500M CPU hours delivered since 2004
  - Providing 1 of the 13 Tier-1 sites worldwide (PIC)
  - Federated Tier-2 sites for ATLAS (IFIC, IFAE, UAM), CMS (CIEMAT, IFCA), LHCb (USC, UB)
  - Among the most reliable sites in WLCG
- A large effort from HEP community and institutions
  - ~26 M€ funding (direct costs) from HEP national program since 2001
  - Funding from institutions of the same order
    - Funding personnel, electricity, infrastructure
- Large community of experts in distributed high throughput computing
  - Contributions to LHC computing, development, integration, operations, management
  - Leverage expertise and infrastructure to support other projects in HEP/astro/cosmo (CTA, MAGIC, DUNE, DarkSide, PAU, Euclid, Virgo, etc)
  - **We have generated a big strategic asset for our community!**

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Plan de Recuperación, Transformación y Resiliencia

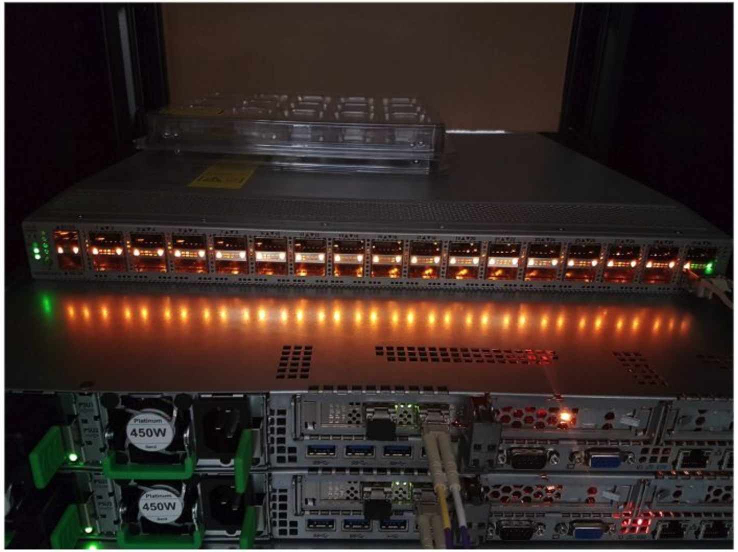
GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo

GVANEXT

# IFIC Networking



- University of Valencia and IFIC has a 100 Gbps connection.
- The Spanish Academic Network provider (RedIRIS) has upgraded the connection at IFIC institute.





The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Plan de Recuperación, Transformación y Resiliencia

GENERALITAT VALENCIANA
Conselleria de Educación, Universidades y Empleo

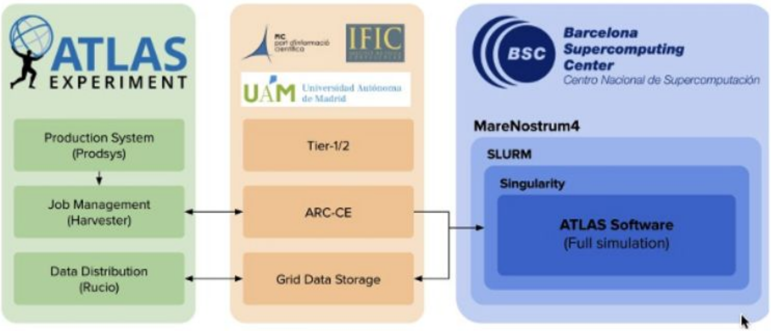
GVANEXT
Fondos Next Generation en la Comunitat Valenciana

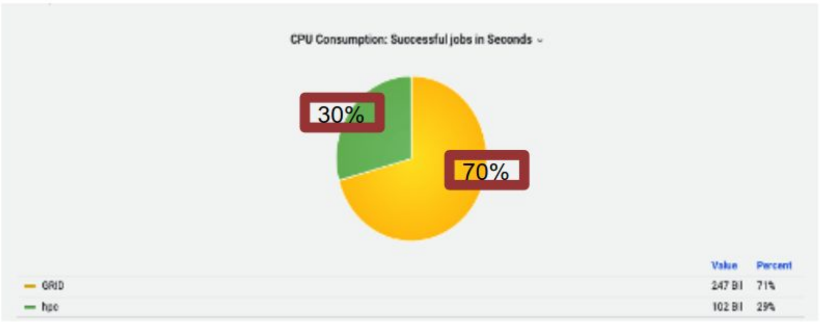
# HPC resources



❖ Use of the Mare Nostrum4 (HPC) by ATLAS Tier-1 and Tier-2s:

- ➢ Using ARC-CE at PIC, IFIC and UAM to interconnect Mare Nostrum and ATLAS production system.
- ➢ Only simulation workflow validated - singularity containers, pre-placed at MareNostrum's GPFS.
- ➢ Mare Nostrum accepts only SSH protocol for job submission and data transfer.



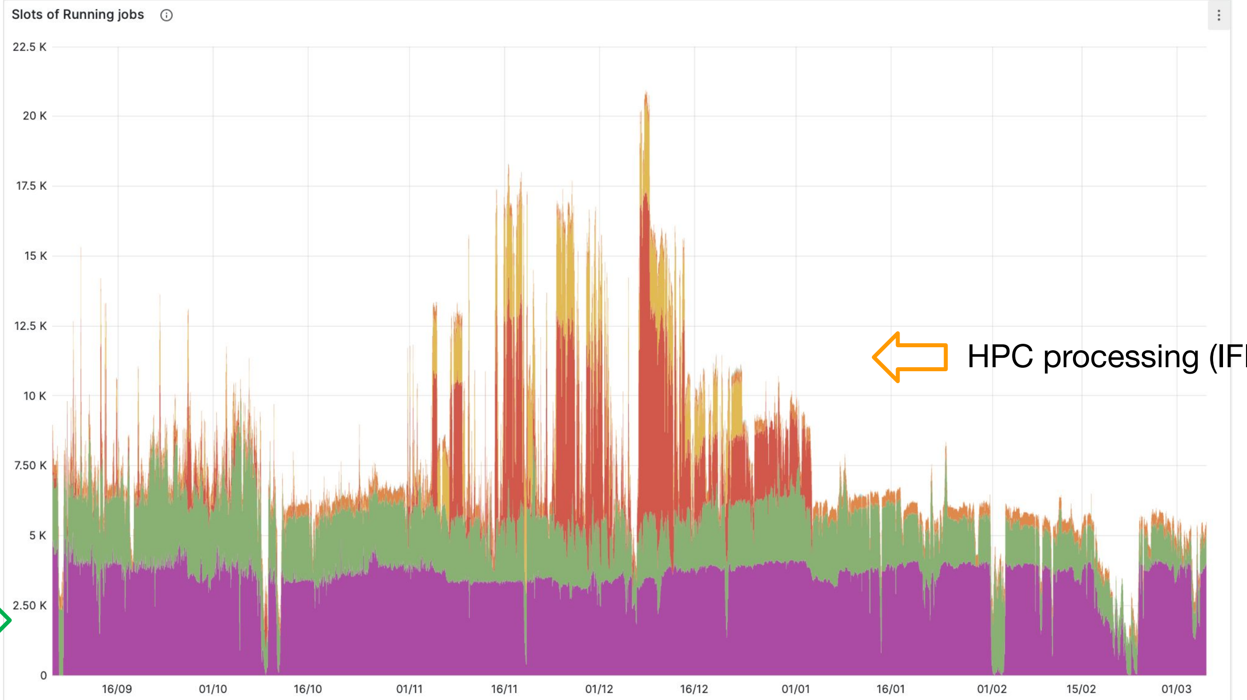




- ➢ Proportion of HS06 (s) provided by GRID resources (yellow) and the MareNostrum 4 HPC (green) in total contribution to the ATLAS computing by the Spanish cloud.
- ➢ 30 million hours approved at Mare Nostrum4 every year by ATLAS through Spanish gateways, which corresponds to 50% of the simulation jobs assigned to Spain.

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.
Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Plan de Recuperación, Transformación y Resiliencia
GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo
GVANEXT

# IFIC Tier 2 efficiency running ATLAS jobs

- Running jobs by site the last 6 months




| | min | max | avg |
|---|---|---|---|
| IFIC | 0 | 5.20 K | 3.48 K |
| ifae | 0 | 7.34 K | 2.09 K |
| IFIC_MareNostrum4 | 0 | 12.0 K | 1.14 K |
| UAM_MareNostrum4 | 0 | 4.02 K | 485 |
| UAM | 0 | 561 | 419 |
| UAM-LCG2_BOINC | 0 | 70.8 | 2.20 |
| IFIC_Lusitania2 | 0 | 16 | 0.803 |

HPC processing (IFIC-MareNostrum)

IFIC processing

# Analysis Facilities



- Guiding principle: Help physicist minimize time-to-insight, enabling iterative exploration of the data:
  - In the future, when processing 10x lumi/evts, avoid physicists 10x waiting time!!
  - Boost productivity and competitiveness of our physics communities
- Typically consist of:
  - Local access to the reduced data samples (e.g. PHYSLITE) with low latency from compute
  - Dedicated storage resources (of order of several 100s of TB).
  - CPU resources used interactively and/or via a batch system (mostly HTCondor).
    - Future User Interface (UI) to be designed in a flexible way, with user-friendly interfaces that do not discourage users.
    - Future implementation of **Jupyter Notebook** instances that will be spawned via a dedicated portal.
  - SW delivery mostly via CVMFS with increasing presence of containers.
  - Expert data/code manager: critical liaison role (not a final user, nor an infrastructure expert, however facilitating technology/access).
  - GPU resources available, but often not dedicated.
    - IFIC: **ARTEMISA** infrastructure (https://artemisa.ific.uv.es)
  - Network:
    - LAN of multiple 25Gbps to support intense data throughput.
    - WAN of 100Gbps connectivity with the WLCG dedicated network for data lake access.

The ASFAE's research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Plan de Recuperación, Transformación y Resiliencia

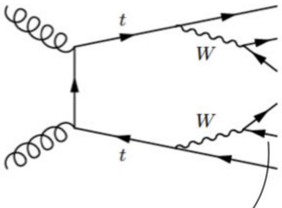
GENERALITAT VALENCIANA Conselleria de Educación, Universidades y Empleo

GVANEXT Fondos Next Generation en la Comunitat Valenciana

# Generative Models for simulation



- Classical simulation of proton-proton collisions implies:
  - PS generation
  - Hadronization/fragmentation
  - Pass the particles through the detector
- Time consuming and expensive
  - Last above step is particularly expensive (eg. dense materials)
  - Billions of events
- Alternative are high fidelity fast generative models, eg. GANs, VAEs & NFs
  - Able to sample high dimensional feature distributions by learning from existing samples, eg. classical simulation
  - Generate SM background and BSM physics scenario and process the data in a easy-format (sequence of 4 – vectors)
  - Metrics to asses performance & syst. errors to be defined
- Use case:
  - pp→ttbar with 6 jets in the final state




Datasets used in this work have been taken from a uptodate repository; the ones generated by DarkMachines community. LHCsimulationProject, Feb 2020, doi:10.5281/zenodo.3685861. Available here

## Loss Function

Variational AutoENcoder (VAE)

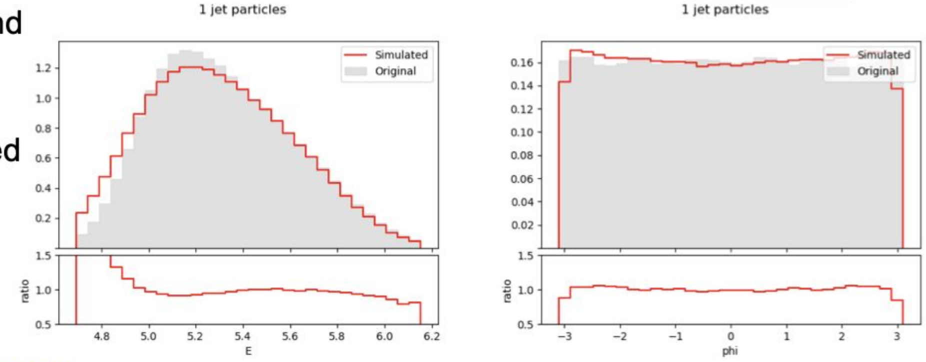
$$L_{VAE} = (1-\beta)MSE + \beta KL$$

MSE: Mean Squared Error. Reconstruction term on the Final layer, which tends to improve The performance of the encoding-decoding schema

a regularisation term on the latent layer, that is proportional to the Kullback-Leibler (KL) divergence and tends to regularise the organisation of the latent space by making the distributions returned by the encoder close to a standard normal distribution with zero mean and unit variance

To avoid Overfitting

Results with β-VAE

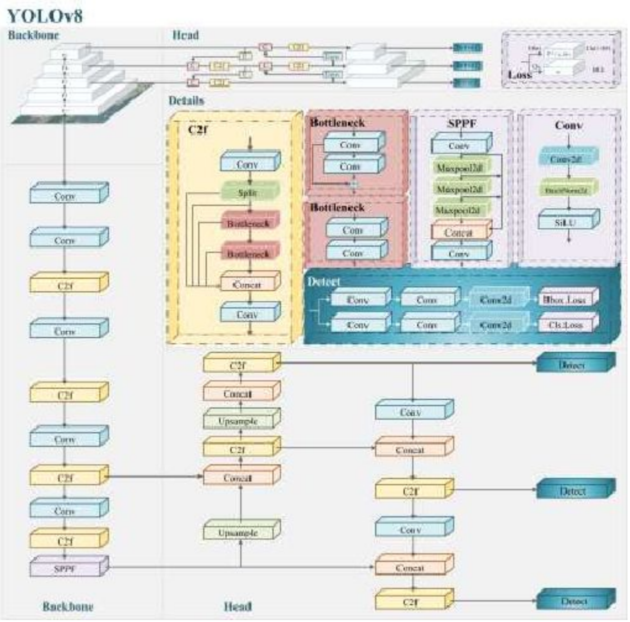



CHEP2023

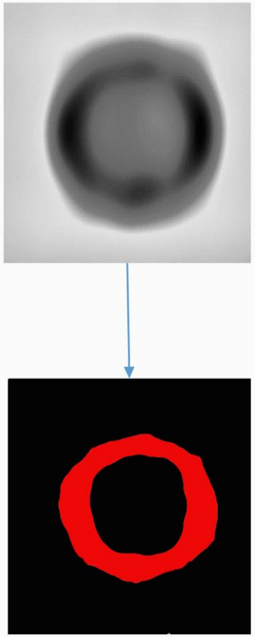
E, ϕ of first jet using β-VAE with β=0.001

# Artificial Inteligence in High Energy Physics at IFCA

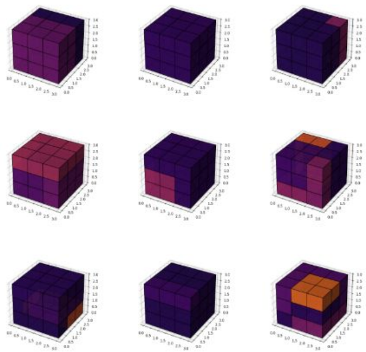
- Object segmentation using YOLOv8 and U-Nets in the context of muon tomography with a LGAD demonstrator




Furnace wall wear

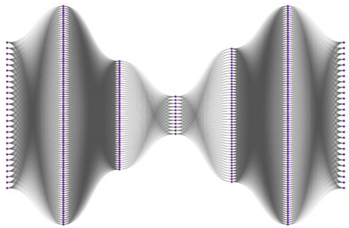


- Anomaly detection using Variational Auto-Encoders in the context of muon tomography applied to cargo inspection



Regularity patterns (examples)

VAE design after optimization

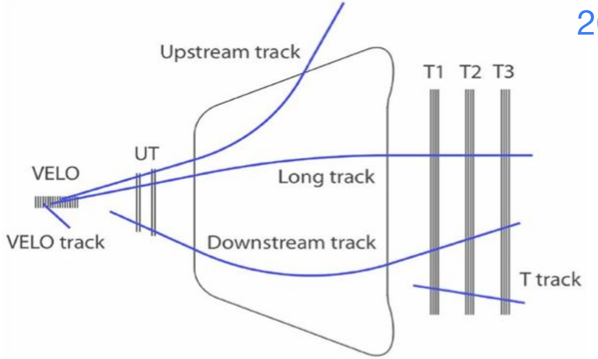


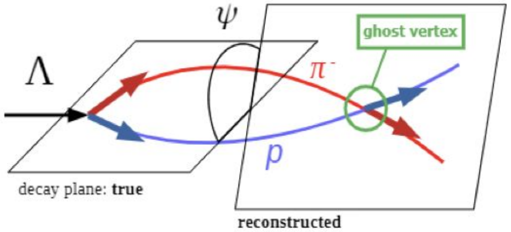
- System trained with geometrical patterns with density regularity (boxes, pallets, etc)
- Triggering alarm for abnormal cases

# T tracks

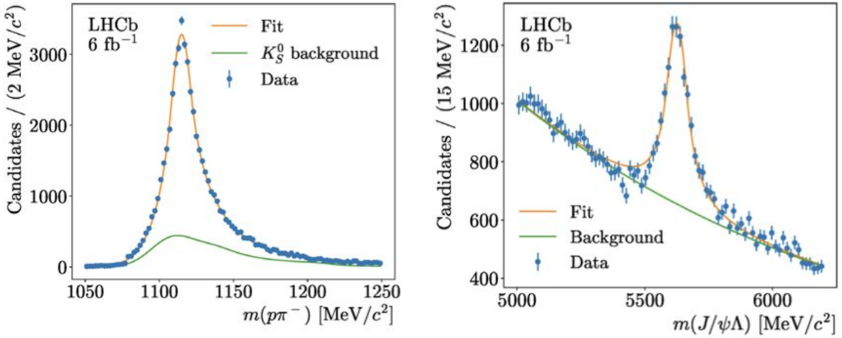


- Unmatched SciFi segments to Long nor Downstream tracks historically unused
- Challenging
  - Short lever arm & weak B field
  - Large extrapolation through strong and inhomogeneous B field
  - Poor momentum resolution ~ 20%
  - Large combinatorics, ~1500 2-track combinations/event @ 10 MHz
  - “Ghost vertices” due to closing-track topologies
- Feasible selection and offline reconstruction for physics
- Clear benefits for physics with strangeness (~40% of the decays) and BSM LLPs with $\tau$ > 100 ps
- Could be triggered?



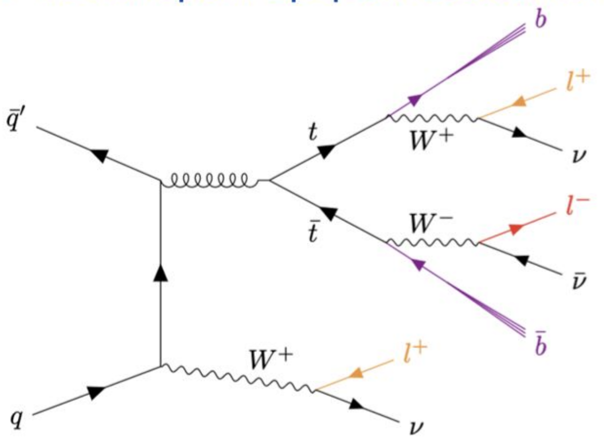

$\Lambda_b \rightarrow \Lambda\, J/\psi$ (Run 1 & 2) arXiv:2211.10920 [hep-ex]

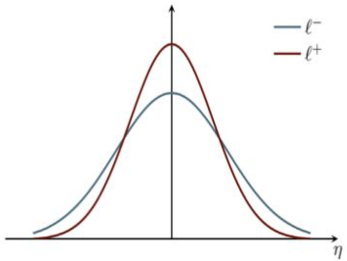


# ttbarW leptonic charge asymmetry



- ML for lepton-top quark association




JHEP **07** (2023) 033

Phys. Lett. B **736** (2014) 252

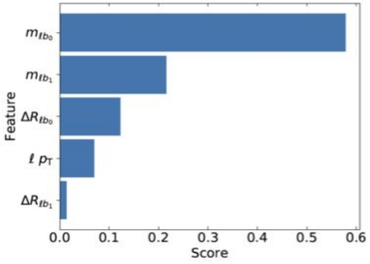
$$A_c^\ell = -13.2 \pm 0.1 \text{ (theory) } \%$$

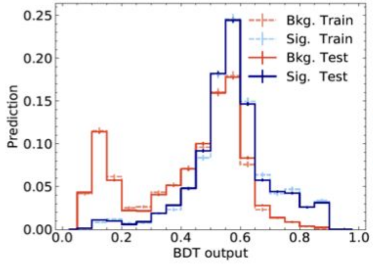
inclusive NLO+PS

$$A_c^\ell = \frac{N(\Delta|\eta_\ell| > 0) - N(\Delta|\eta_\ell| < 0)}{N(\Delta|\eta_\ell| > 0) + N(\Delta|\eta_\ell| < 0)} \quad \text{with} \quad \Delta|\eta_\ell| = |\eta_{\ell^+}| - |\eta_{\ell^-}|$$

- Charge asymmetry between the leptons coming from top and antitop quarks: enhanced in ttbarW events compared to ttbar
- Experimental challenge in ttbarW 3l final state:
  - Identify the charged leptons coming from top and antitop quarks
  - The correct even lepton is selected using GBDT

Odd lepton: always from (anti)top quark
Even lepton: need to select the correct one




- For each event, trained even leptons (object level MVA, per lepton)
- The accuracy of the BDT for selecting the correct lepton is 71%