Scientific Computing Technology
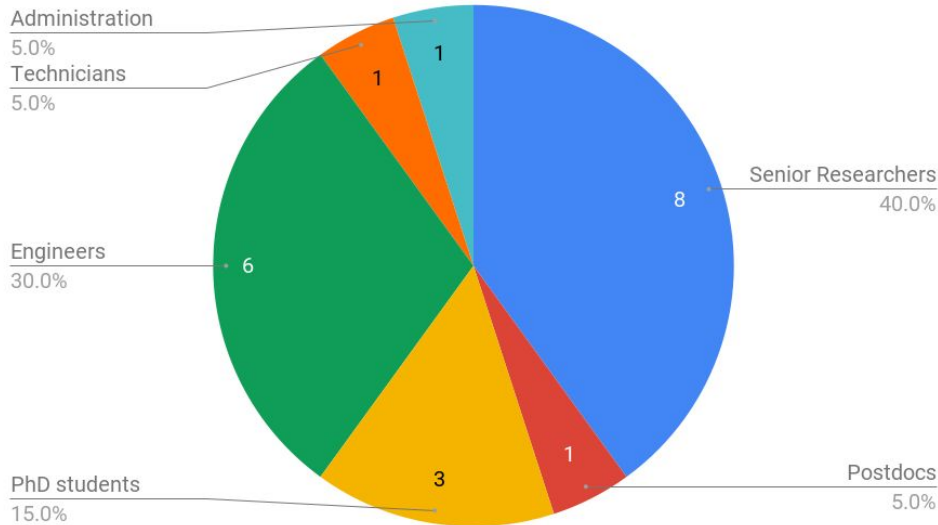
José M. Hernández

Our **mission**

Enable and accelerate science with computing technology

- **Infrastructure and support** for the particle physics, astroparticle and observational cosmology projects in the department
  - Transversal research line to the scientific research lines
- Focus on **data-intensive distributed computing** technologies
  - Massive storage systems, High-throughput workload management frameworks
  - Development and integration of Grid/Cloud/HPC Computing for HEP
- Highly parallel **processing and artificial intelligence** technologies
  - Highly **parallel** computing techniques (GPU accelerators)
  - **Machine learning** techniques (neural networks, data mining)
- We provide **hardware, software and human resources**
  - Specialized personal (development, integration, operations, support)
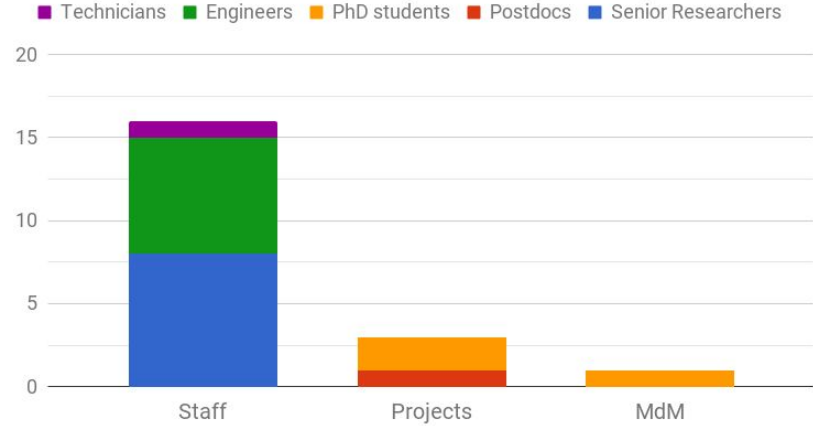 - Data centers in Madrid (CIEMAT) and Barcelona (PIC)

CIEMAT-FP Advisory Committee

People

- 20 people (3 women)
- Recent staff stabilization process
- Personnel is our most important asset

Personnel by category

Personnel funding sources

Technicians ■ Engineers ■ PhD students ■ Postdocs ■ Senior Researchers

Gender distribution by category

Women ■ Men



ry Co
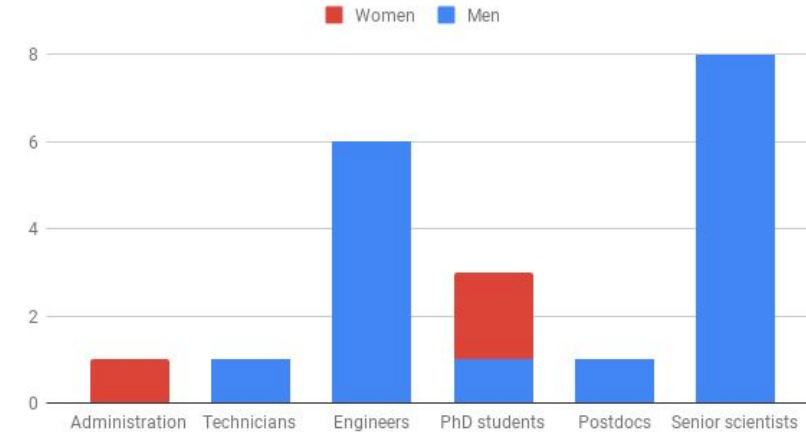
PIC scientific data center

Founded 2003, PIC is a scientific-technological centre specializing in data analysis and management technologies, operated through a collaboration between CIEMAT and IFAE.

Mission: Participate at highest level in LHC Computing (Tier-1) and leverage experience to support other excellent scientific activities.

Value: Experience from multidisciplinary collaborations to build scalable and robust common services for multiple experiments: leverage synergies, economies of scale, cross-fertilization.

- **Multidisciplinary team** to bridge the gap between the science and the data infrastructure - advanced computing services.

- **Service providers** for data preservation, analysis and sharing.

- Close collaboration with disciplinary researchers to **design & build architectures** for handling scientific data flows.

    - **Agile** interaction. Focus on prototype/feedback loop.

4
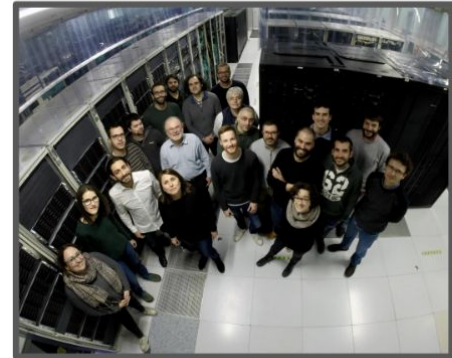
# PIC data center

200 Gbps connection to Research Networking

- Largest data mover in Spanish academic network: 100 PB in+out per year

Data processing services

- 12.000 CPUs, 18 GPUs (HTCondor cluster)

- 18 PB disk (dCache) + 69 PB tape (Enstore)

- Bigdata platform (Hadoop/HIVE/Spark)
   - Cosmology analysis web portal and data processing cluster

Two machine rooms, total ~150 kW$_{IT}$:

- ~120 kW in 150 m2 air-cooled room (high efficiency, renovated 2014-2016)

- ~30 kW in 25 m2 innovative liquid immersion cooling system (higher efficiency)
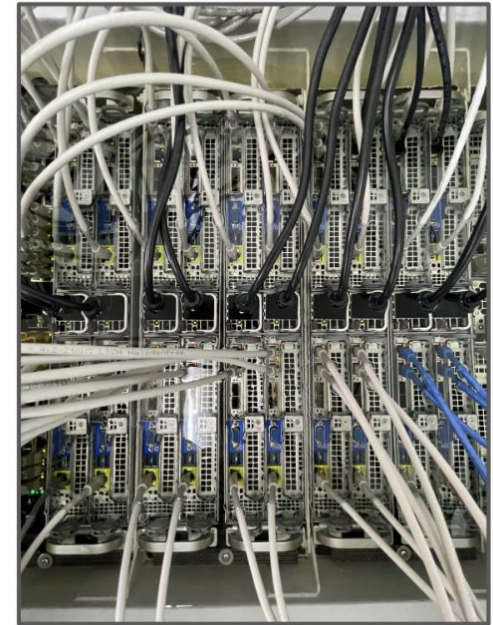
Spanish LHC Tier-1 at PIC

- **Worldwide distributed computing** infrastructure for **processing, simulation and analysis** of LHC experimental data
  - Data at the exabyte scale
- Spain provides ~5% of WLCG resources (WLCG MoU)
  - Tier-1 center at PIC
- LHC computing technology, infrastructure and services transferred to other scientific projects
  - High throughput distributed computing
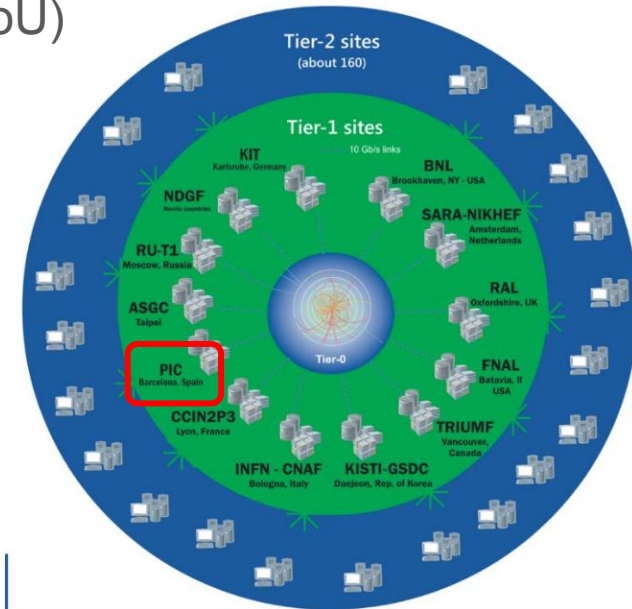
Other scientific projects at PIC

Gamma rays: CTA data center
- One of four off-site CTA data centers (DESY, INAF, CSCS and PIC)
- Leading off-site data management
  - Automated transfers ORM-PIC via dedicated 10Gbps network link
  - Automated replication to other sites using Rucio/FTS steered from PIC
  - Long-term preservation of the data

Gravitational waves: LIGO/VIRGO
- Joined through IFAE the LIGO/VIRGO Grid in Summer 2019
- Last 12 months contribution: 10% of the CPU (EU) and GPU (Global)

Einstein Telescope
- Active in the set up of computing activity

Cosmology: PAU Survey
- Main data center at PIC: data preservation and access for analysis
- Development of automated data handling system: nightly pipelines, science validation
- Data analysis: Photo-z, shapes, intrinsic alignment
- Machine Learning: Galaxy/star separation, noise removal

Cosmology: Euclid
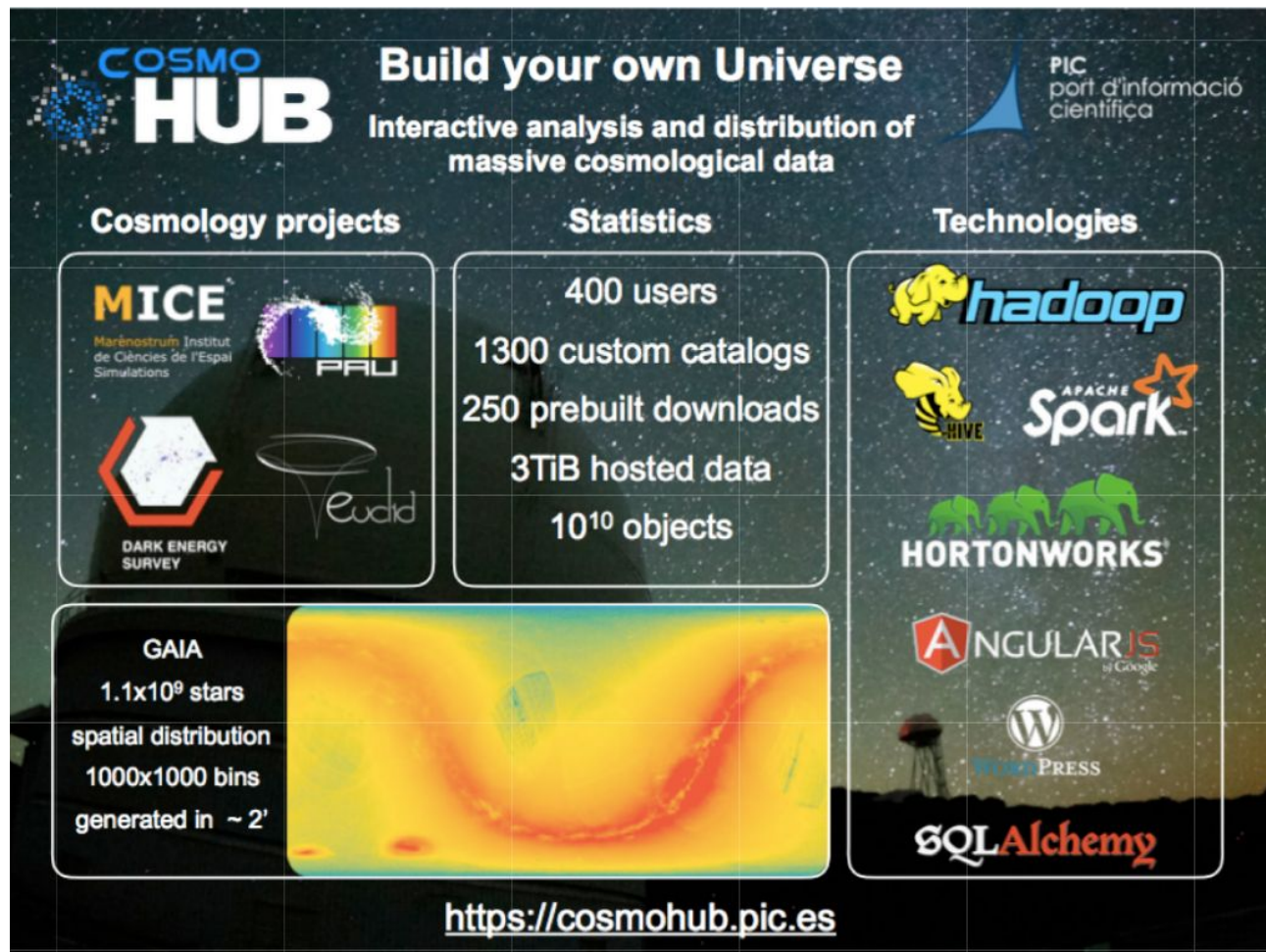- PIC is the Science Data Center Spain

Neutrino physics: DUNE
- Contributing computing resources to DUNE distributed computing system
  - 1.5 TB disk, 500 CPU cores

Services for astroparticle and cosmology projects

CosmoHub platform

- Web portal for interactive exploration of cosmology data
- Processing platform with very fast data query using industry big data techniques
- Real-time data analysis and visualization
- 40 TB catalogued data
- ~100 catalogs from different projects
- ~150 active users (>1500 total)
- ~13500 custom catalogs generated

Infrastructure: CIEMAT data center

- Located in Madrid, at CIEMAT main site
- Computing resources and user support
  - **4.5k CPU cores, 6 PB disk storage, 20 Gbps WAN** connectivity
- Tier-2 center for CMS
- Computing resources for local data analysis of the various physics projects in the department
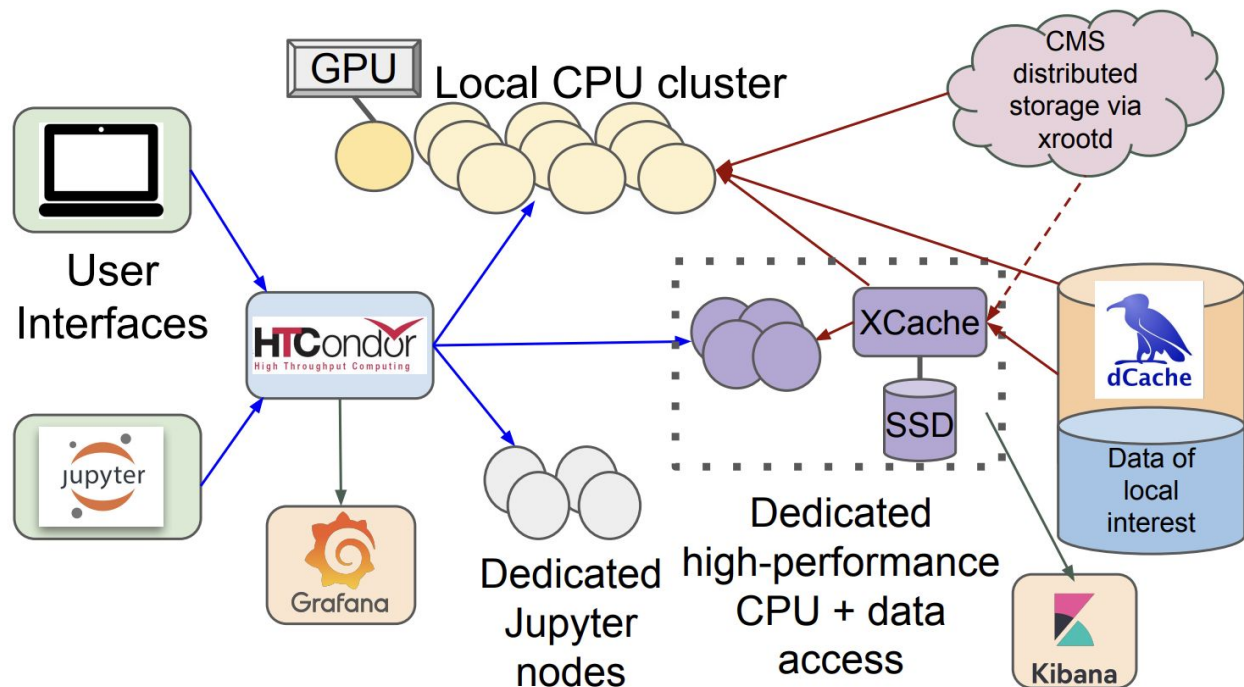- CMS analysis facility
- VIRGO online cluster

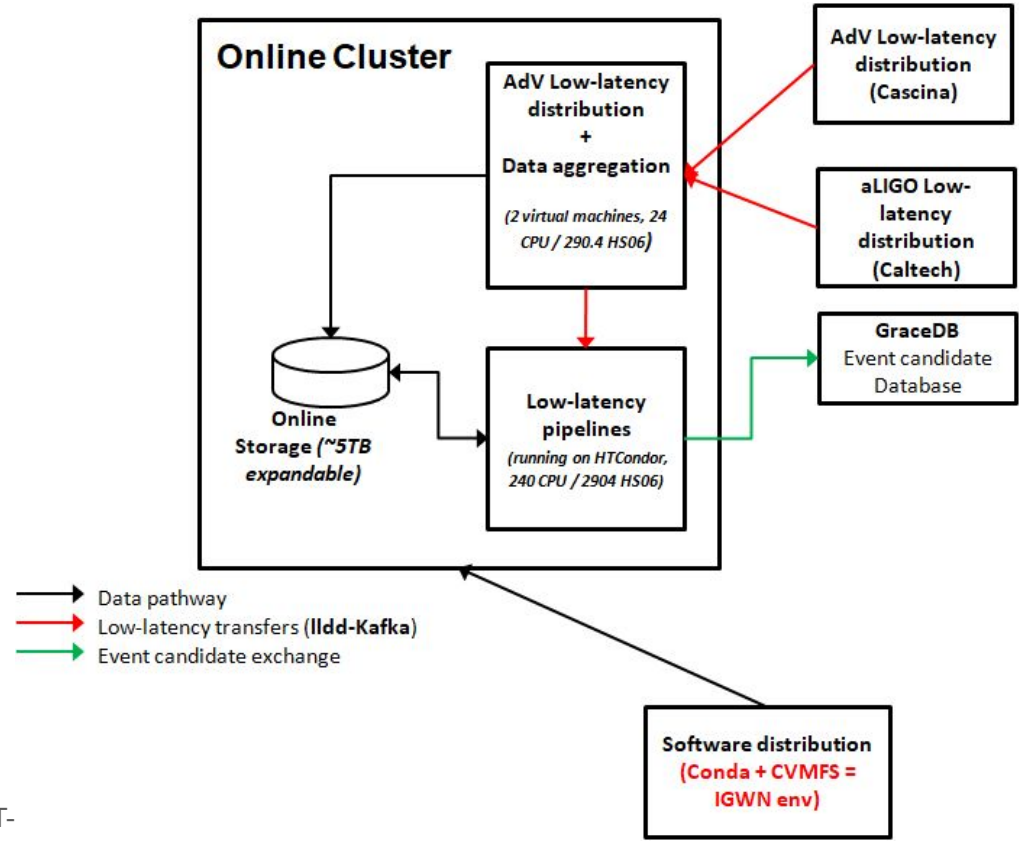# CMS **Analysis Facility**

- ○ High throughput storage (400 TB NVMe)
- ○ GPU servers (4x -Nvidia A100 GPUs)
- ○ Dedicated CPU batch system (~500 cores)
- ○ Infrastructure, services and support

Gravitational waves: LIGO/VIRGO low latency pipelines

- Deployed in 2023 cluster for running VIRGO low-latency pipelines
  - Extension of VIRGO online cluster
  - ~500 CPU cores

Funding

- **~5.5 M€** funding 2020-2023
- **Sources:** National R&D&I program, national infrastructure funds, direct assignments from ministry
- **Projects:** mostly LHC computing and PIC infrastructure (additional funding for PIC through IFAE)

Call	Project	Period	Funding [k€]
National Program R&D&I	LHC Computing Tier-1/Tier-2	2020-2023	850
National Program R&D&I	CMS Analysis Facity	2021-2024	308
National Program R&D&I	LHC Computing Tier-1/Tier-2	2023-2026	634
Ministry Direct assignment	LHC Computing upgrade	2023-2025	593
National scientific-technological equipment	PIC high performance storage system	2022-2023	923
ICTS equipment	PIC data preservation system	2023-2025	540
ICTS equipment	PIC data processing system	2025-2027	1000
Ministry Direct assignment	PIC CTA data center	2021-2023	400
National Program R&D&I	DUNE computing infrastructure	2021-2023	100
Ministry Direct assignment	AI infrastructure for quantum computing	2022-2025	200

12

Scientific production

In addition to providing computational support to the scientific projects of the department, the computing group produces as well scientific results, mostly innovation in data-intensive computing ([link](#)). Since 2020:

- 8 projects competitive funding (>3 M€)
- 3 ongoing PhD thesis
- 50 contributions to conferences and workshops
- 20 publications
- Management responsibilities in experiments
    - WLCG Grid deployment board chair, CMS computing resource board chair, CMS resource office coordination, CMS distributed computing coordination, etc

Prospects

- Upscaling PIC data center
- Support to particle, astroparticle and cosmology scientific projects
- Integration of supercomputing resources
- Deployment of general-purpose analysis facility
- Applied AI
- Quantum computing
- Technology transfer (within CIEMAT and outside)

CIEMAT-FP Advisory Committee

Upscaling PIC

- Member of Spanish ICTS (Scientific and Technological Singular Facility)
  - Data node of Spanish Supercomputing Network from september 2020
  - ICTS funding (0.6 M€) to extend long-term tape archive storage
  - Expected ICTS funding in 2024
    - Extend disk data storage (0.75 M€)
    - deploy an energy efficient data processing cluster (GPU cluster with immersion cooling, 0.25 M€)
- Funding from the Ministry for upscaling the LHC Tier-1 center throughout Run 3 towards the LHC high luminosity Phase (1.6 M€)

Particle physics: LHC

- Secured funding for next three years
  - 1 M€ from the national R&D program, complemented with 1.6 M€ with specific funding for LHC upgrades
  - Provide 5% Tier-1 for ATLAS, CMS, LHCb
  - ~20% annual increases 2024-2026

Utilization of supercomputing resources

- CIEMAT leads the agreement with the Barcelona Supercomputing Center (BSC) for the Spanish contribution to LHC computing
  - Successful (although lot of hard work) integration of MareNostrum 4 machine
    - PIC as entry point to BSC for the LHC distributed computing infrastructure
  - MN4 provides ~50% of the CPU contribution from Spain to LHC (~60M core-CPU hours per year)
- Integrate MN5
  - Recently inaugurated
  - 20x more powerful than MN4, pre-exascale EuroHPC supercomputer, total peak computational power of 314PFlops, ~800k CPU cores Intel Sapphire Rapids, ~4k Nvidia Hopper GPUs
  - Volunteered as early adopters

CFP general purpose Analysis Facility

- Build on CMS AF project and extend it to support other CFP scientific projects
  - Already informally used by other projects
- Integrate computing resources and expertise at PIC and CIEMAT sites
  - High throughput data-intensive analysis platforms
  - GPU resources for AI
  - Big Data platform (Hadoop/HIVE/Spark)
  - Multidisciplinary team to bridge the gap between the science and the data infrastructure
- Plan to apply for funding to the Spanish HEP R&D national program

Applied AI: activities at CIEMAT

- Applications based on artificial intelligence (AI) algorithms becoming ubiquitous
- These techniques are being exploited in several projects in the department and beyond
  - Background event rejection in the dark matter DEAP-3600 experiment
  - Event classification in CMS proton-proton collisions
  - Identification of gravitational waves in Ligo-Virgo detectors
  - Position reconstruction in the DarkSide-20K experiment
  - In oncology, investigation of the relationship between ovarian and bladder cancers, their phenotypes, and treatments
  - Prediction of the effects of the layout of an urban network of electric vehicle charging stations on traffic congestion, air quality, carbon footprint, and electric grid usage
  - Identification of anomalous thermoluminescence glow curves in dosimetry services
  - Assessment of the impact of the Low-Emission Zone "Madrid Central" on NO2 emissions
  - Emphasis on Explainable AI shedding light on the on the reasoning behind the decisions or predictions made by the AI
  - Course given at CIEMAT on Neural Networks Applied to Scientific-Technical Problems

Applied AI: new group at PIC

New group on AI with 3 people

- Interdisciplinary
- Close collaboration with different research groups

Day 6 - Image segmentation and object detection
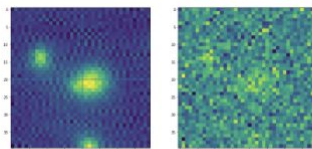
Martin Eriksen, Laura Cabayol

AI training
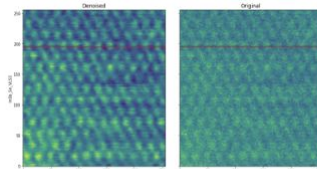
~10 weekly 2h sessions

Small groups (6-10). Focus on discussion.
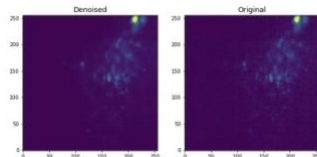
Participants from IFAE, and elsewhere

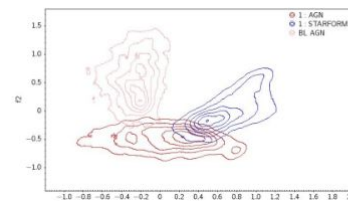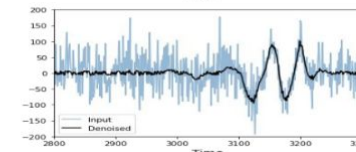*L. Reynolds and C. Ramírez (Cosmo), A. de Lorenzis and P. Casado (Neutrinos), M. Seglar (GW)*
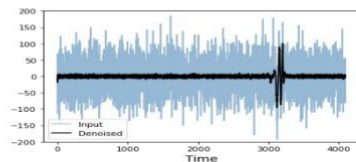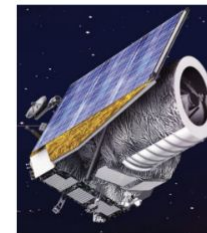
Galaxy surveys

Material science (InCAEM)

Bioimaging (ICFO)
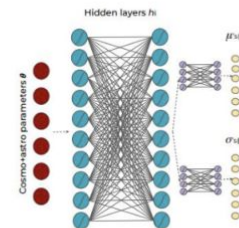
Galaxy evolution

Gravitational waves

Euclid weak lensing

Cosmological emulators

Quantum SPAIN
Quantum Machine Learning

6

Quantum Spain project

PIC participates in Quantum Spain project to deploy a quantum computer and in Spain. Part of the future user support team.

— Promoted by the Ministry of Economy through the Secretary of State for Digitization and Artificial Intelligence and financed with the Recovery Funds.
   22 M€, 01/01/22 – 31/12/25

— Goals: deployment of a quantum computer for spanish researchers and development of quantum algorithms.

— As a node of the RES, PIC gets funding for 1FTE to be trained as future support team.

Technology transfer

Within CIEMAT:

- Leading Scientific Computing R&D line in new CIEMAT strategic program
  - Federating computing resources of the three CIEMAT data centers (Madrid, Barcelona-PIC, Trujillo-CETA)
    - Deploying data transfer system
    - Unified portal for compute resource user access
- Several projects using AI technology

Outside CIEMAT:

- InCAEM project - In Situ Correlative Facility for Advanced Energy Materials
  - Data analysis system for a multidisciplinary platform in collaboration with ALBA synchrotron
- Genomic data processing in Parc Taulí Hospital

GOBIERNO DE ESPAÑA · MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES · Ciemat · Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas · cfp CIEMAT física de partículas

Summary

- Advanced computing technology is key for science exploitation in our field
- We have assembled over the past two decades a substantial computing infrastructure and a competent team
  - Bridging the gap between science and data infrastructure is essential for competitiveness and efficiency
- Positive perspectives for the next few years
  - Satisfactory funding and new projects