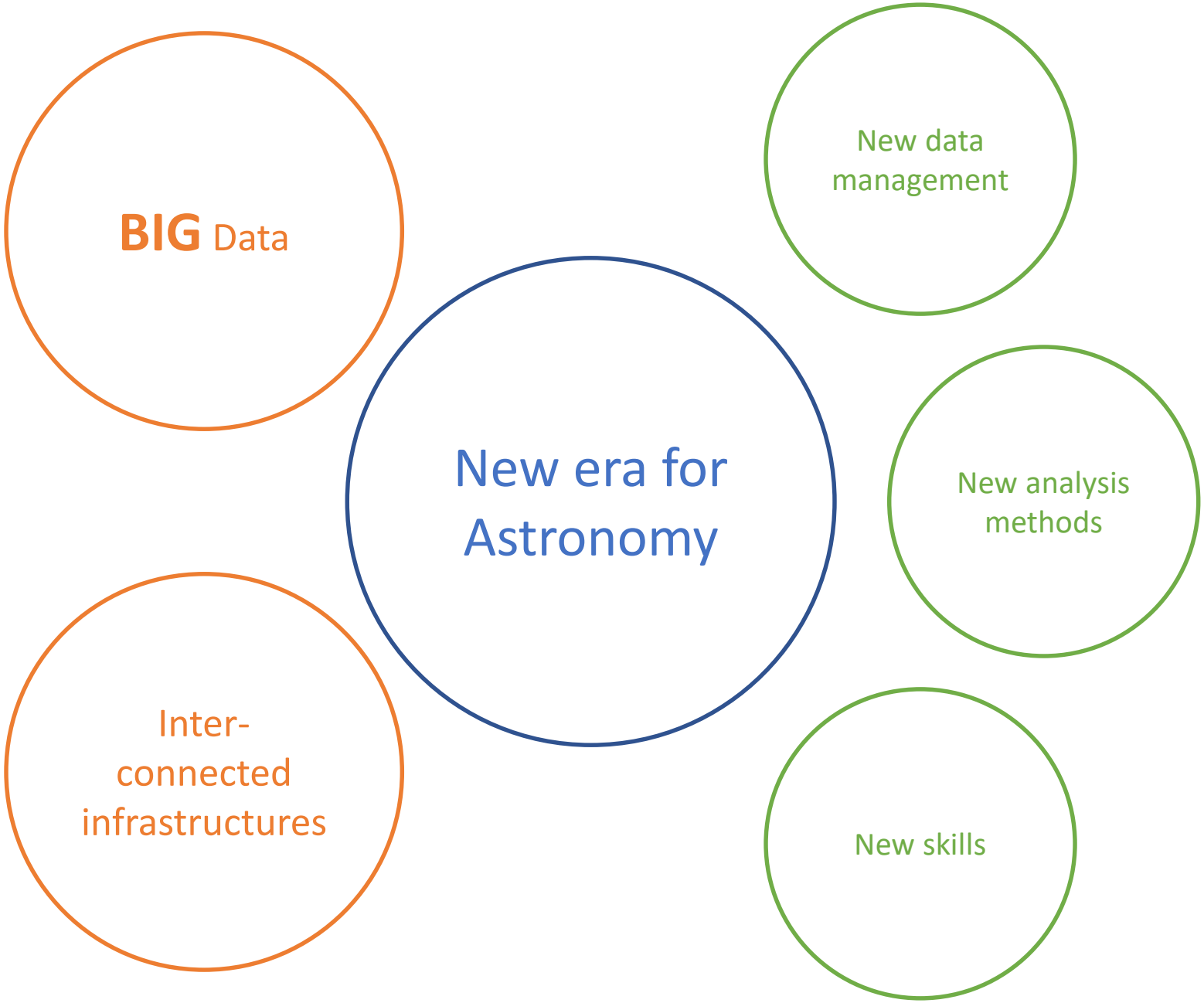
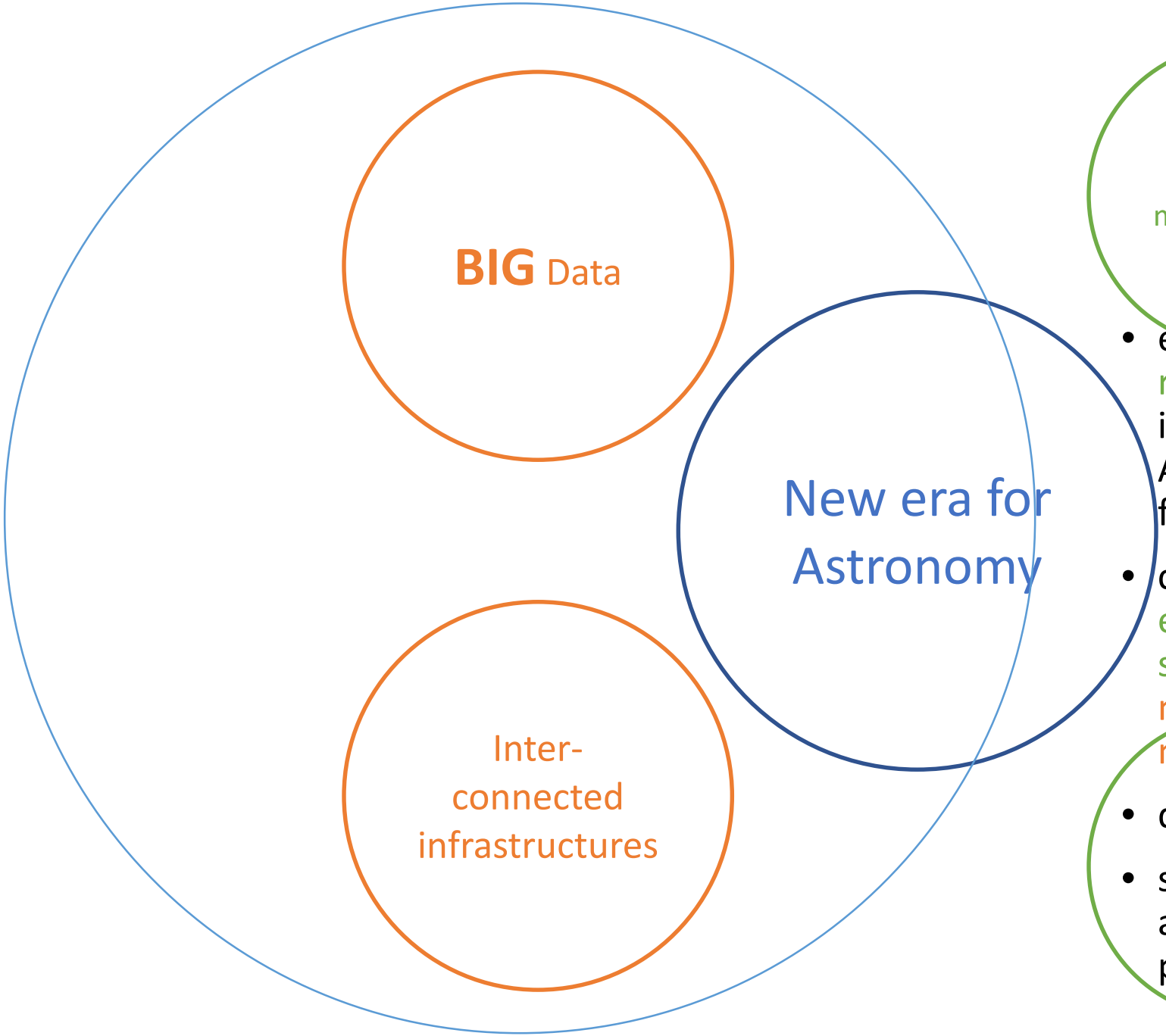
OBELICS
Observatory e-environments linked by common challenges
- towards a multi-messenger era

Thomas Vuillaume,
The new era of multi-messenger astrophysics,
Groningen, 28 March 2019


BIG Data
New data management
New era for Astronomy
New analysis methods
Inter-connected infrastructures
New skills



New data management

Obelics
Observatory E-environments Linked by common ChallengeS

- enabling interoperability and software re-use for the data generation, integration and analysis of the ASTERICS ESFRI and pathfinder facilities

New analysis methods

- creating an open innovation environment for establishing open standards and software libraries for multi-wavelength and multi-messenger data
- developing common solutions

New skills

- studying advanced analysis algorithms and software frameworks for data processing

The OBELICS galaxy

Building a community: workshops

- Three workshops bringing together representatives of the ESFRI and major industrial partners
- Creating a community to face the common data challenges
  - Invited talks
  - Tutorials and live demos
  - Panel discussions
- Creation of a machine learning community in astronomy and astroparticle physics

EU-TO
APPEC
egi
HELIX NEBULA THE SCIENCE CLOUD
E4 COMPUTER ENGINEERING
INDIGO - DataCloud
HELIX NEBULA THE SCIENCE CLOUD
AARC
CERN
OROBIX TOOLS WITH A MIND
COMMON WORKFLOW LANGUAGE
NVIDIA
Atos
ORACLE
IBM
EOSCpilot The European Open Science Cloud for Research Pilot Project
IVOA

Training the community: schools

- Three summer schools “Advanced software programming for Astrophysics and Astroparticle physics”
- Python oriented
- Training young and seniors to develop good, reusable and efficient software together to tackle the data analysis challenges of the community

Training the community: schools

- > 60 students every year
  - From all over the world
  - From all communities

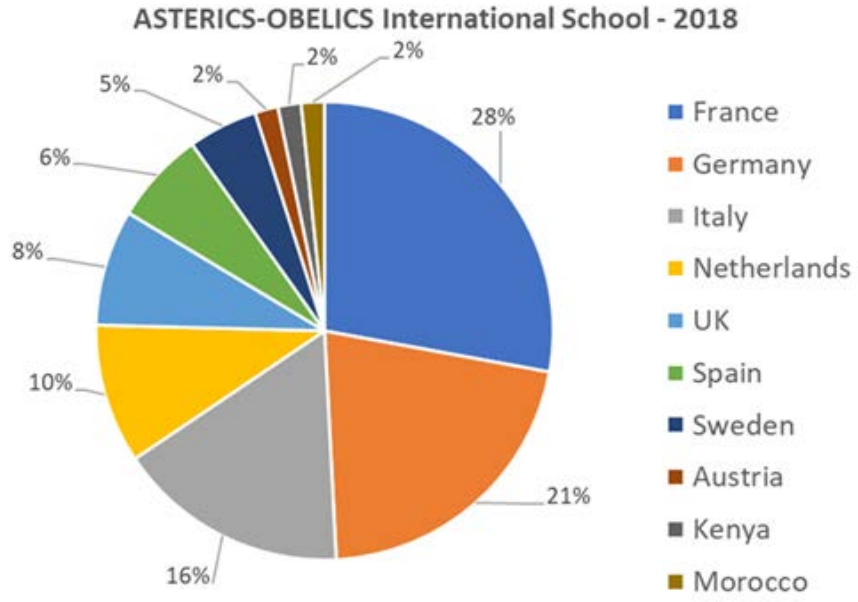
- Lectures
- Hands-on

- Tutors experts in all wavelengths and messengers

- Great feedback from the students

- **To be continued…**

  - **Third one in two weeks**
  - **And more with ESCAPE2020**



Covered topics:

- Good coding style
- Git
- Python data science libraries (numpy, pandas, scipy, astropy…)
- Data visualisation
- Machine Learning
- Julia

Addressing the challenges

-

OBELICS highlights,

a non-exhaustive list

CORELib: A COsmic Ray Event LIBrary

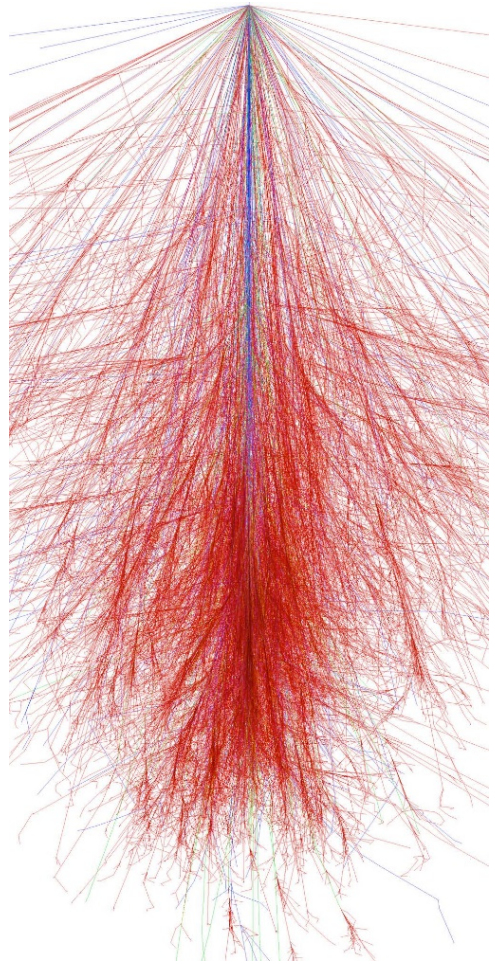
- Library of cosmic events generated by CORSIKA
- Can be used “as-is”
- Save a lot of computing resources

Pilot production
(available via SFTP)

Energy range (GeV)	Number of events
200-1000	10^7
10^3-10^4	10^7
10^4-10^5	10^6
10^5-10^6	10^5
10^6-10^7	10^4
10^7-10^8	10^3
10^8-10^9	10^2

Full production
(accessible via GRID)

Energy range (GeV)	Number of events
200-1000	15 x10^5
10^3-10^4	15 x10^5
10^4-10^5	15 x10^5
10^5-10^6	15 x10^5
10^6-10^7	15 x10^5
10^7-10^8	15 x10^5
10^8-10^9	15 x10^5



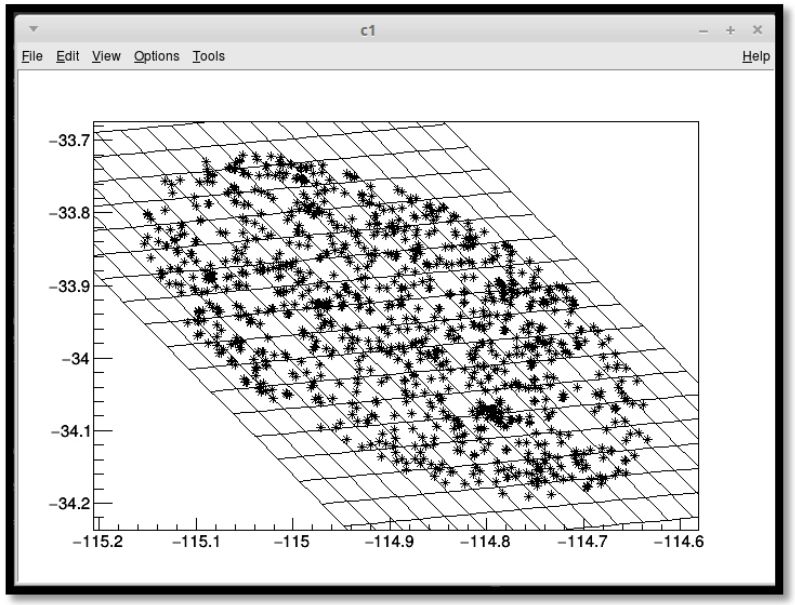
Contact: Cristiano Bozza

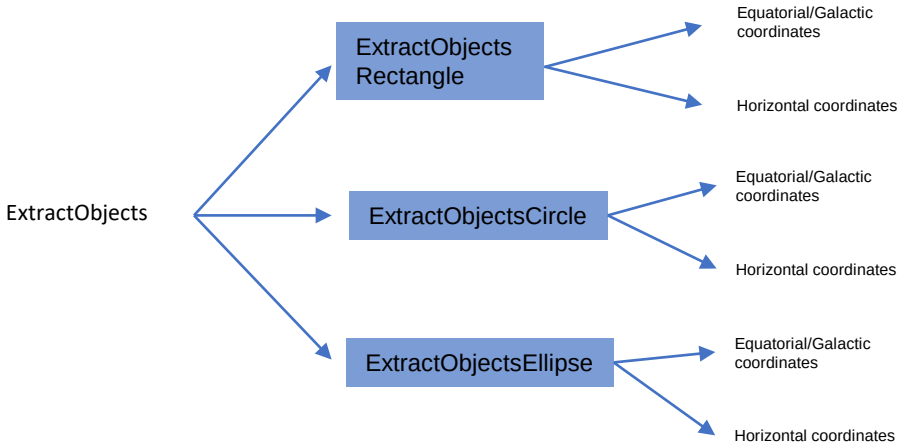
KM3NeT

INFN
Istituto Nazionale di Fisica Nucleare

ROAst: ROot extensions for ASTronomy

- extend the ROOT capabilities adding packages and tools for astrophysical research:
  - access to astronomical catalogues
  - coordinate conversion tools
  - high-precision Moon and Sun position models relative to the Earth
  - graphical tools to produce commonly used plots (general and partial skymaps)



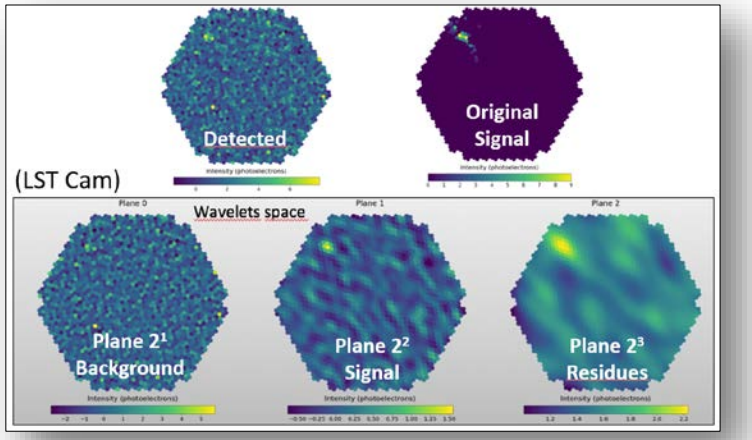


Contact: Cristiano Bozza, INFN

PyWI - Python Wavelet Imaging

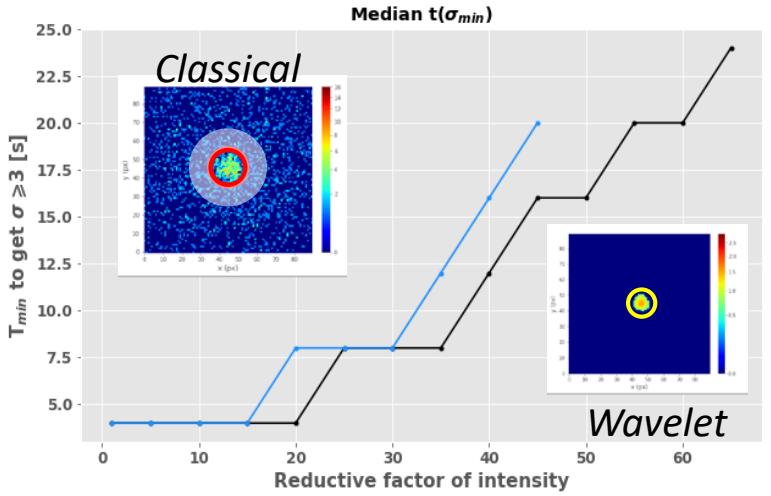
- Adapt tools from CEA Cosmostat (www.cosmostat.org) used in Euclid and SKA to CTA
  - ISAP library (2D, 3D wavelet transform and filtering)
  - Bundle in a Python library

Camera images



- Night sky background filtering
- Sensitivity gain

Sky images



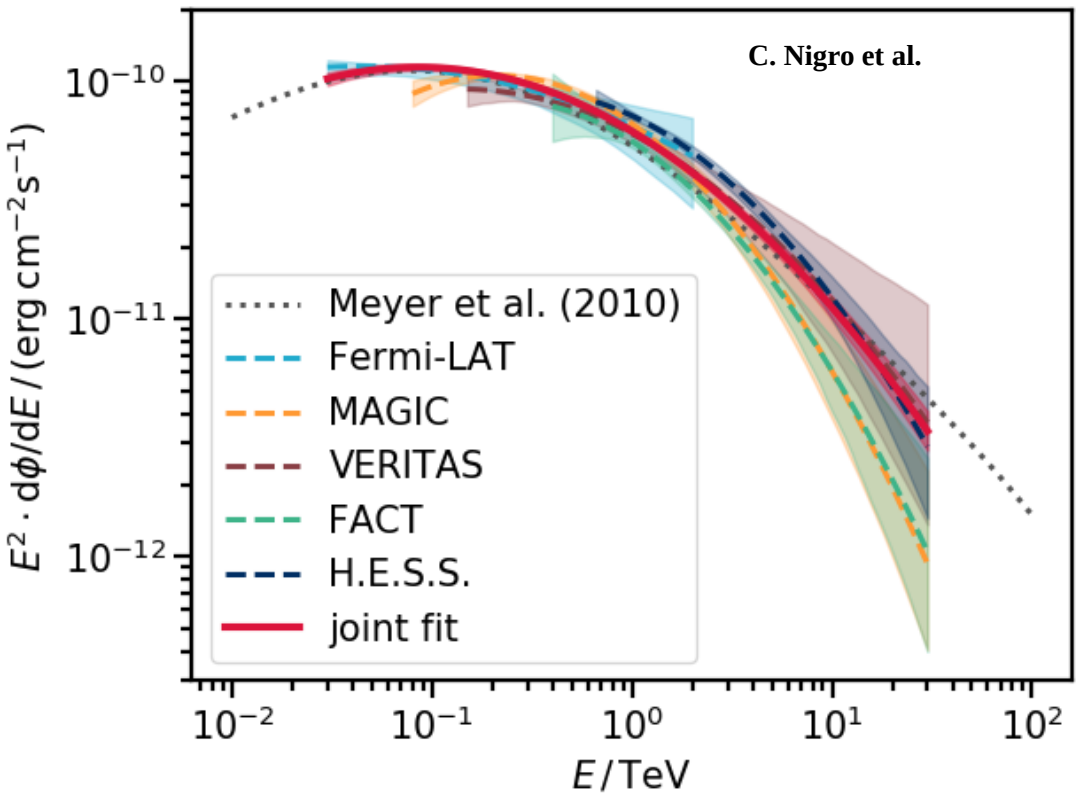
- +50% on detection threshold
- Gain on detection time

→ Higher sensitivity and faster detection = better alert response and generation

Contact: Thierry Stolarczyk, CEA

A common data format for VHE astronomy

- Current IACT data and software are mostly p
- CTA will operate as an **open observatory wit**
- Development of a common data format:
  - event lists (photon-like events) + instrument re
  - in FITS
  - Discussed openly at https://gamma-astro-data
- CTA science tool prototypes ***Gammapy*** *and*
- Conversion of current IACTs data to a comm
  - CTA science tools validation
  - Legacy data that allow reproducible and multi-i
- First multi-instrument spectral analysis of the Crab Nebula:
  - Common data format + *Gammapy*

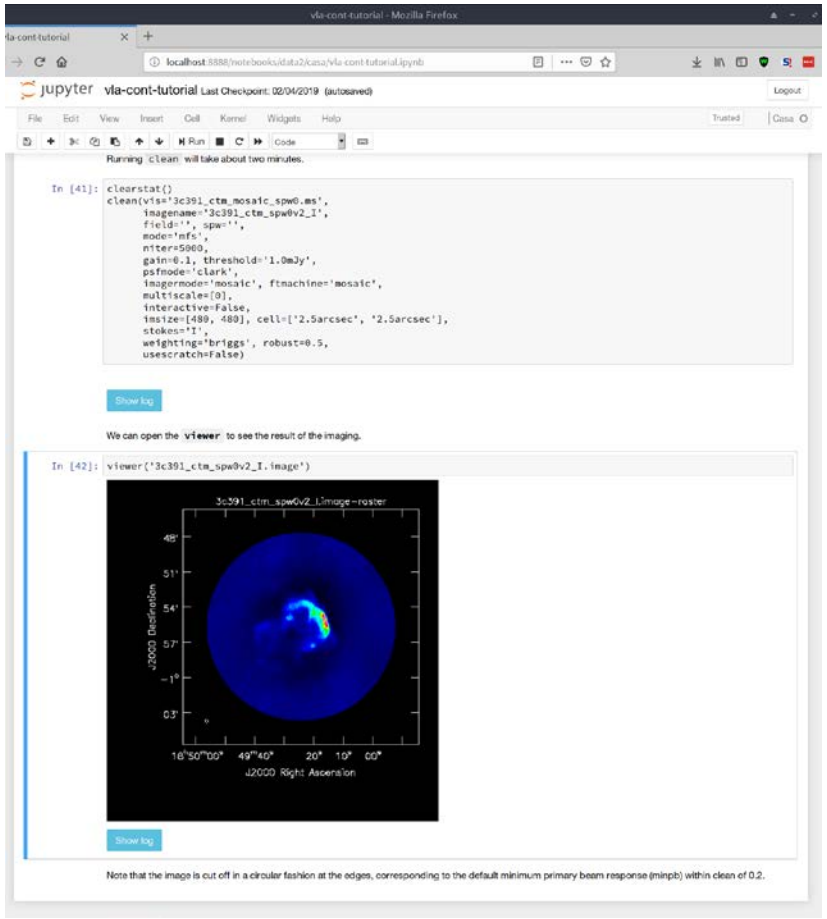


See Léa Jouvin's talk

Contact: Léa Jouvin, IFAE

CASA in Jupyter notebooks

- CASA: Leading data reduction package for Radio Astronomy (ALMA, VLA)
- JIVE recently added VLBI capabilities
- Jupyter notebooks allow easy remote execution where data is
- CASA data format (MSv3) also to be used for SKA.



VLA tutorial running as Jupyter notebook

Contact: Aard Keimpema, JIVE

CASA in Jupyter notebooks

With Recipe Integration:

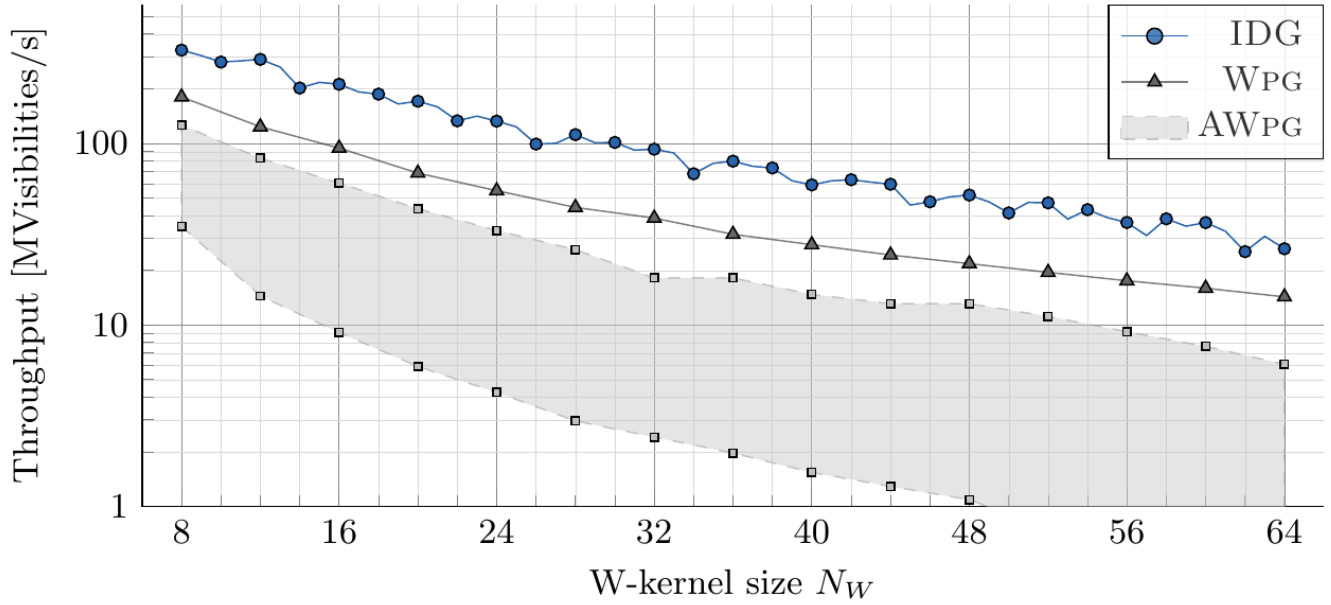
- Recipe: Minimal-recomputation for exploratory data analysis (Nikolic, Small and Kettenis) (also part of OBELICS)
- Allows skipping steps that have already been calculated when executing notebooks
- Available as docker and singularity containers
- https://github.com/aardk/jupyter-casa

Contact: Aard Keimpema, JIVE

Image Domain Gridding

The Low Frequency Array (LOFAR) requires very computationally intensive data analysis

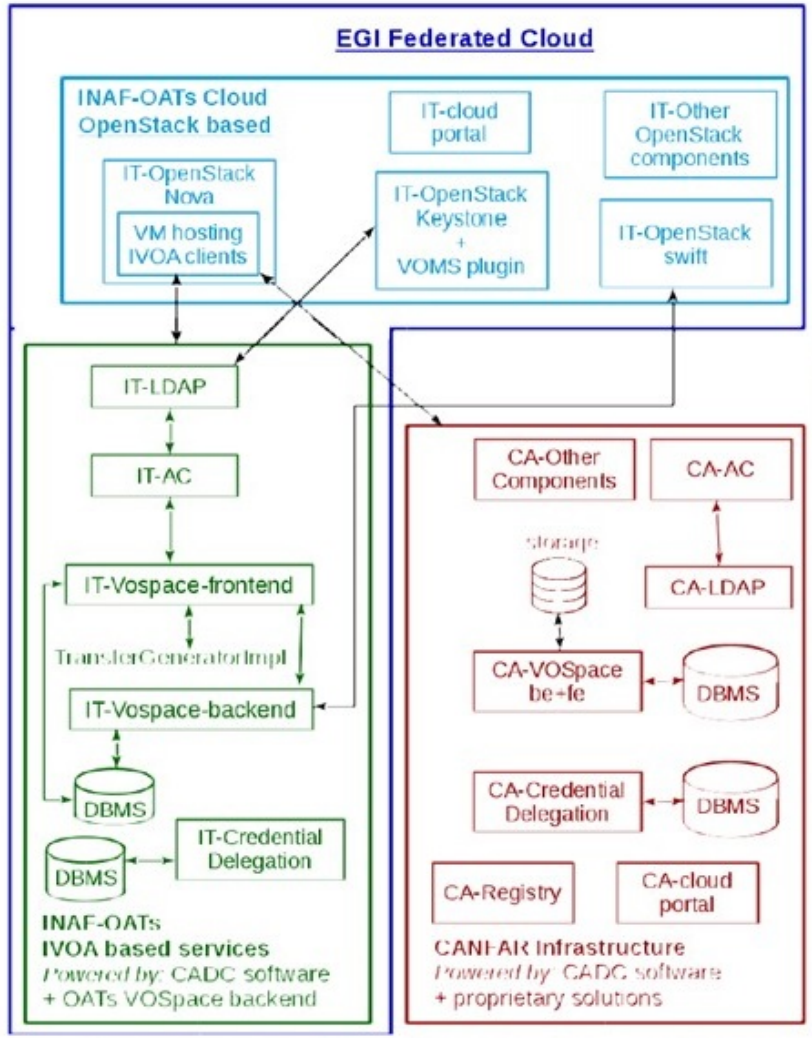
- Image Domain Gridding designed for massively parallel hardware
- Optimised for GPUs computations
- Integrated into WSClean (imager)
- Will be used in the preFactor pipeline currently commissioned by the Radio Observatory



Contact: Sebastiaan van der Tol & Bram Veenboer, ASTRON

Authentication & Authorisation, Workflows

- Evaluation of available workflow systems
- Evaluation of available A&A systems
- Design and implementation of several dedicated workflows embedding A&A functions:
  - Data processing workflow for CTA
  - Science gateway as web-based tool for high-energy astrophysics
  - Data processing workflows at INAF-IA2 for GIANO and HARPS-N pipelines and studies for their integration within the CWL (Common Workflow Language) framework
- Deployment of tools for federation of geographically distributed clouds
- Coordination with EGI and IVOA for standardization



Contact: S.Bertocco, A.Bignamini (INAF-OATs); A.Costa (INAF-OACt)

Fluorescence radiation in Corsika

- Extensive Air Showers (EAS) generate Cherenkov and Fluorescence radiation. Fluorescence is normally neglected in Cherenkov observatories
  - Wanted to check this assumption and investigate its possible uses
  - Implemented fluorescence in CORSIKA, the most widely used EAS code
  - Studied Fluorescence contamination in Cherenkov-signals
  - Checked first constrains for using Cherenkov telescopes in “fluorescence mode”
  - Code potentially useful for a large number of situations

- *Detailed simulation of the effect of fluorescence radiation in Cherenkov Telescopes observations*, D. Morcuende, J. Rosado, J.L. Contreras, F.

Data and Software preservation through Containerisation

Docker Images for Software Development

- Fully reproducible environments
- Isolated workspaces
- Reusable base images
- Integrated into GitLab CI Pipelines in KM3NeT

Singularity Images in Production Pipelines

- Seamless integration into the (file) system of the host computer
- Completely independent base environment
- Perfect fit for Grid computing with heterogeneous systems
- "Software installation" is basically just copying a prebuilt Singularity image
- Already used in ANTARES production pipelines and currently tested in KM3NeT

Data Provenance

- File history is part of the data preservation
- Provides historical records of the data and its origins, like
  - container IDs (e.g. Singularity Registry)
  - specific software and library versions
  - and additional parameters (e.g. command line arguments or configuration files)
- Data and Software stored together in a single image file
- Reproducibility of results through containerised analysis chain

Contact: Tamas Gal, Kay Graf, FAU

KM3NeT

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

Data format studies for CTA

- Benchmarks of data formats for CTA low-level data
  - Historical data format
  - New custom data formats for CTA data
  - HDF5

- Benchmarks of data compression for CTA low-level data

- Development of a new compression algorithm for digitalized signals from physics experiments
  - Test on CTA MC data = speed-up of 19 compared to LZMA (keeping the best compression ratio)
  - *Polynomial data compression for large-scale physics experiments, Aubert, P. et al 2018*

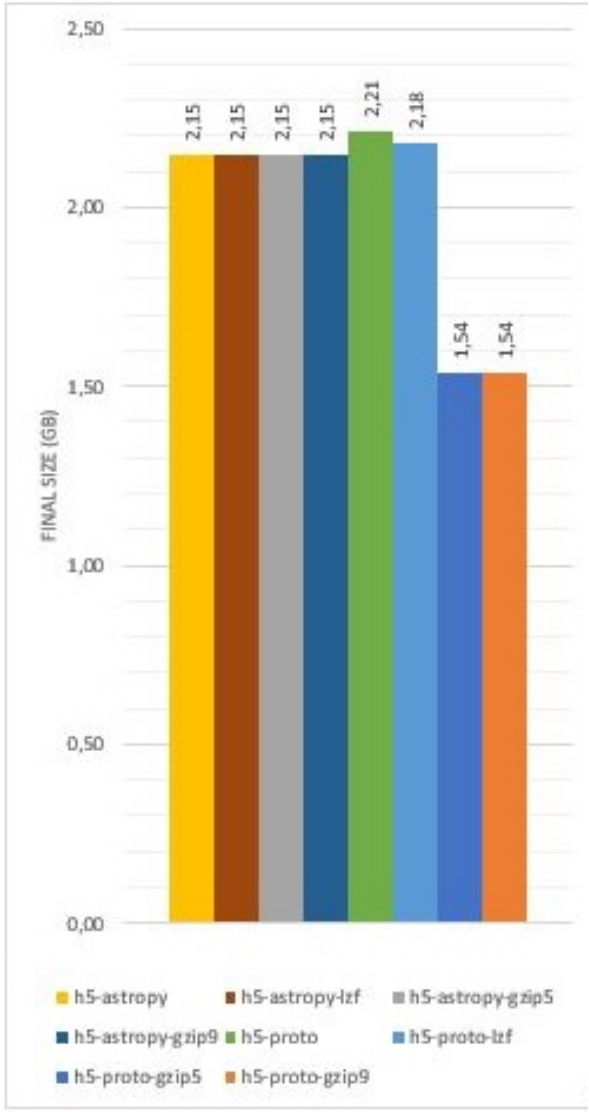


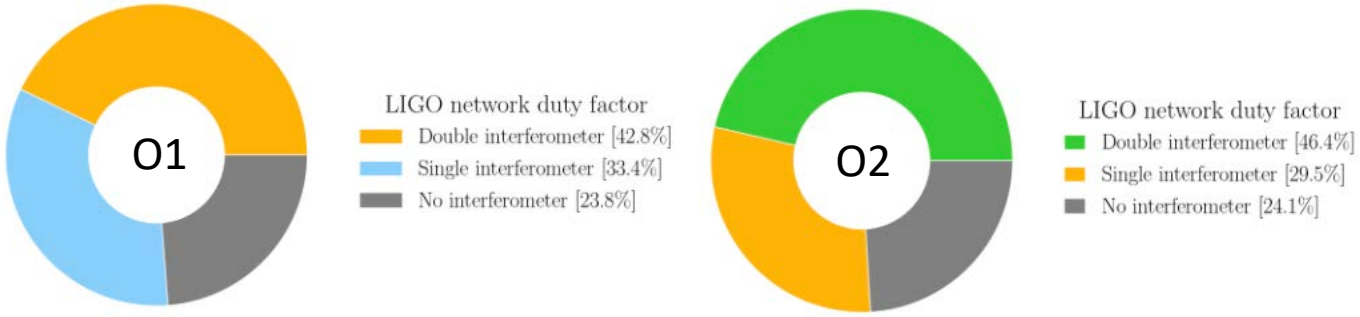
Figure 1: Final size.

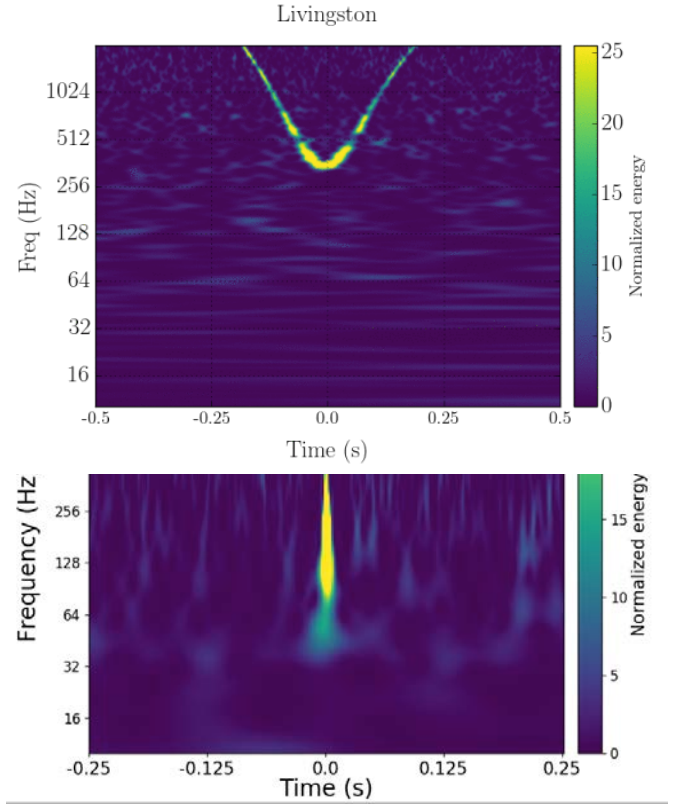
Contact: J.L. Contreras, UCM & P. Aubert, LAPP

Detecting Virgo's glitches with deep learning

- Study, identify and reduce the transient noise present in the gravitational wave detectors using deep learning techniques
  - Raw time-series as input instead of spectrograms
  - Both strain data and auxiliary channels

- Goal: analyze single-detector data
  - Detect more events



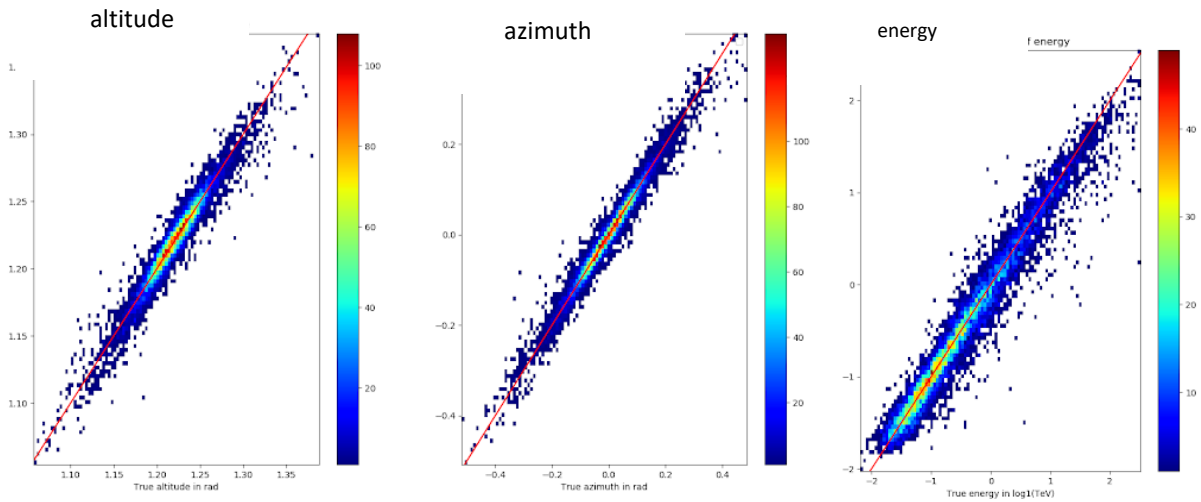


Contact: Agata Trovato, APC

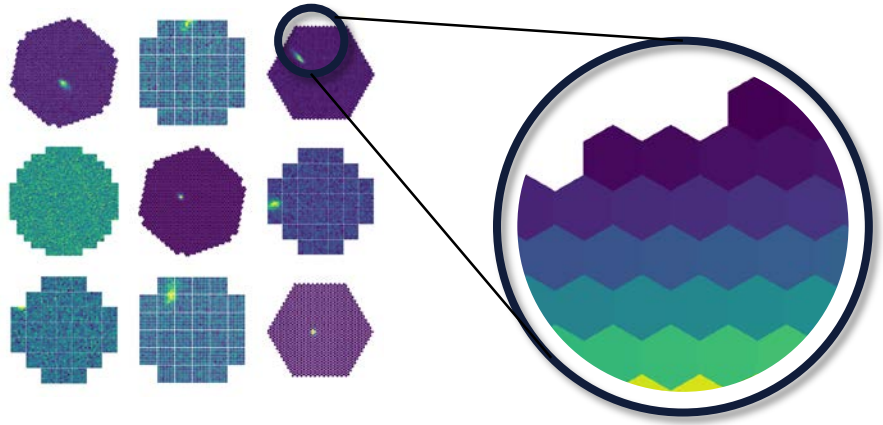
GammaLearn: Deep Learning for CTA

- Goal: Using state-of-the-art image analysis
  - Improve sensitivity and resolution
  - Fast reconstruction = fast alert generation

- GammaLearn framework
  to ease R&D and production
  https://lapp-gitlab.in2p3.fr/GammaLearn



- The hexagonal problem



- Indexed Convolution package to deal with any pixel layout
- *Indexed Operations for Non-rectangular Lattices Applied to Convolutional Neural Networks, Jacquemont, M. et al 2019*
- https://github.com/IndexedConv
- Great interest for other physics detectors

Contact: Thomas Vuillaume, LAPP

And many more developments…