ASTERICS @ CEA

Karl Kosack
Thierry Stolarczyk
Fabio Acéro
Jérémie Decock

CEA ASTROPHYSICS

200 people (25% of Irfu)

- Mixed staff: CEA, CNRS, Univ. Paris VII
- Head: Anne Decourchelle

Large involvement in Instruments

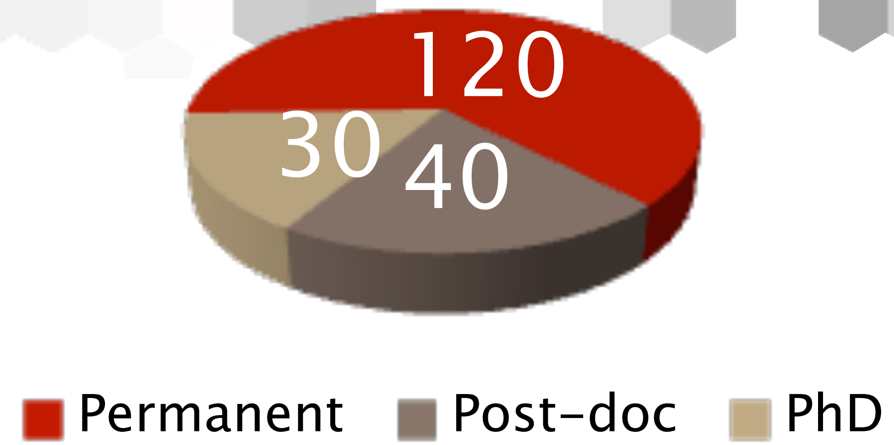
- CTA
- ARTEMIS
- Euclid
- JWST
- SVOM (largest group in France)
- E-ELT, Plato, Athena, Solar Orbiter
- and **many others currently running**… (HESS, Fermi, XMM-Newton, Planck, INTEGRAL, Kepler, …)

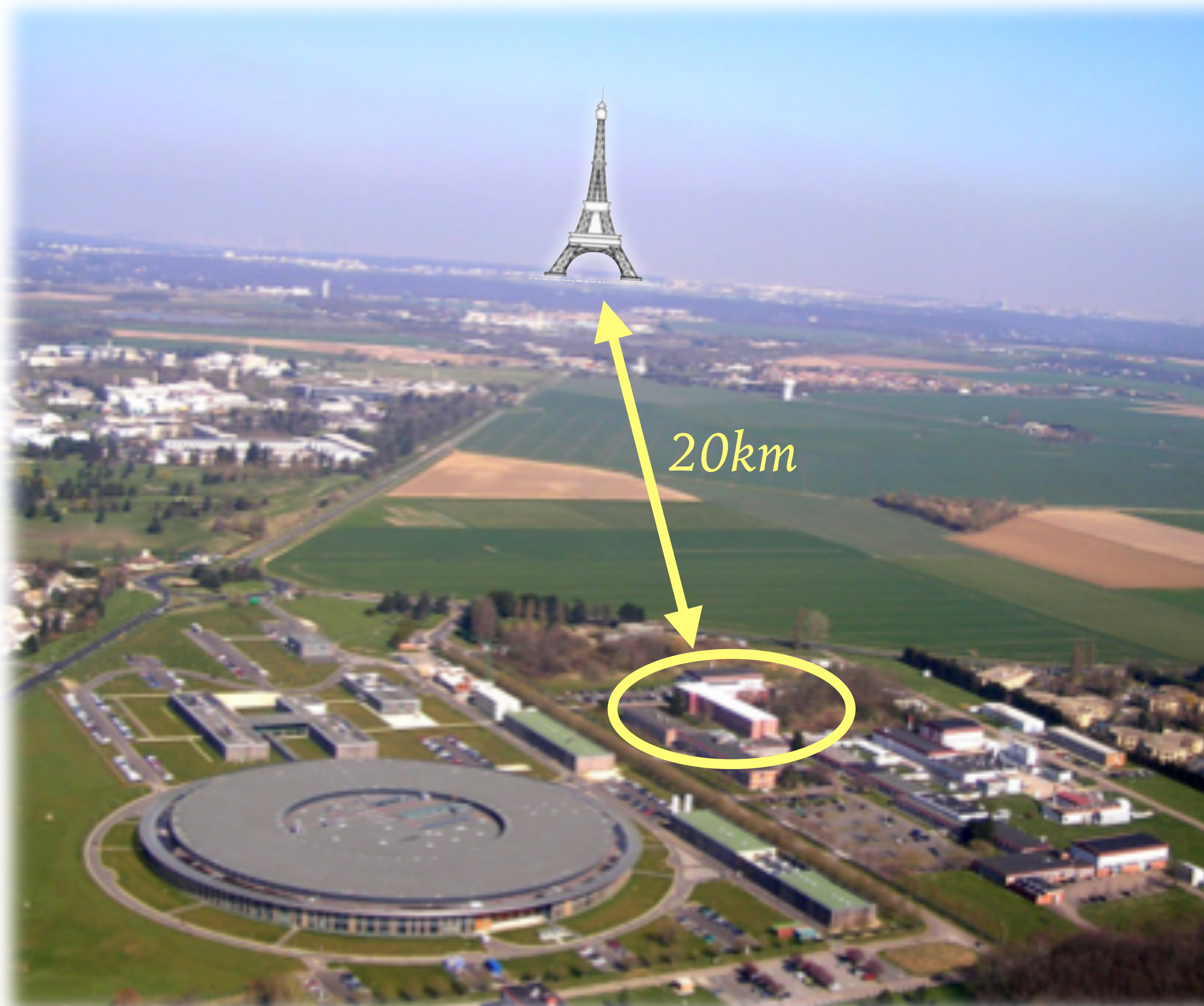
Detector R&D (IR, bolometers, X/γ)

- With CEA (LETI), ONERA, Industry…

Science:

- Cosmology, Galaxy evolution
- Star formation, Planets, ISM, plasmas
- **High energy phenomena** (25 people)





Irfu
Institut de recherche sur les lois fondamentales de l'Univers

cea SACLAY

CTA GROUP

12 People, working on CTA + other projects

- ➤ HESS, Fermi, Integral, SVOM, XMM-Newton, Antares
- ➤ *Science* : SNRs, X-Ray binaries, pulsars/PWNe, galactic diffuse flux, and GRBs

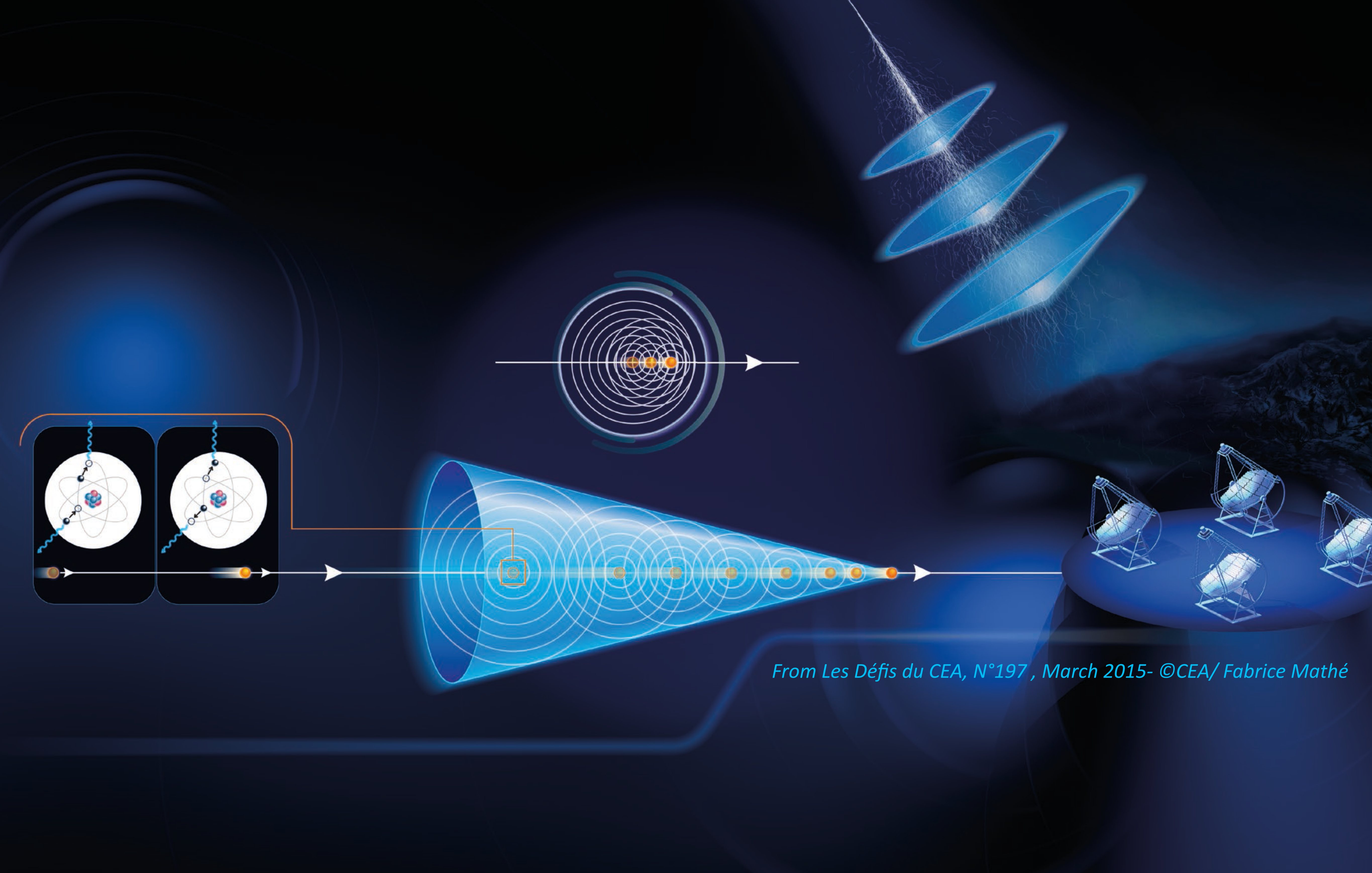
- 1 **ASTERICS** postdoc to join in April 2016
- 1 postdoc working as liaison between High-Energy Astro group and CosmoStat Group (J. Decock) for algorithm development

Main Contributions so far for CTA:

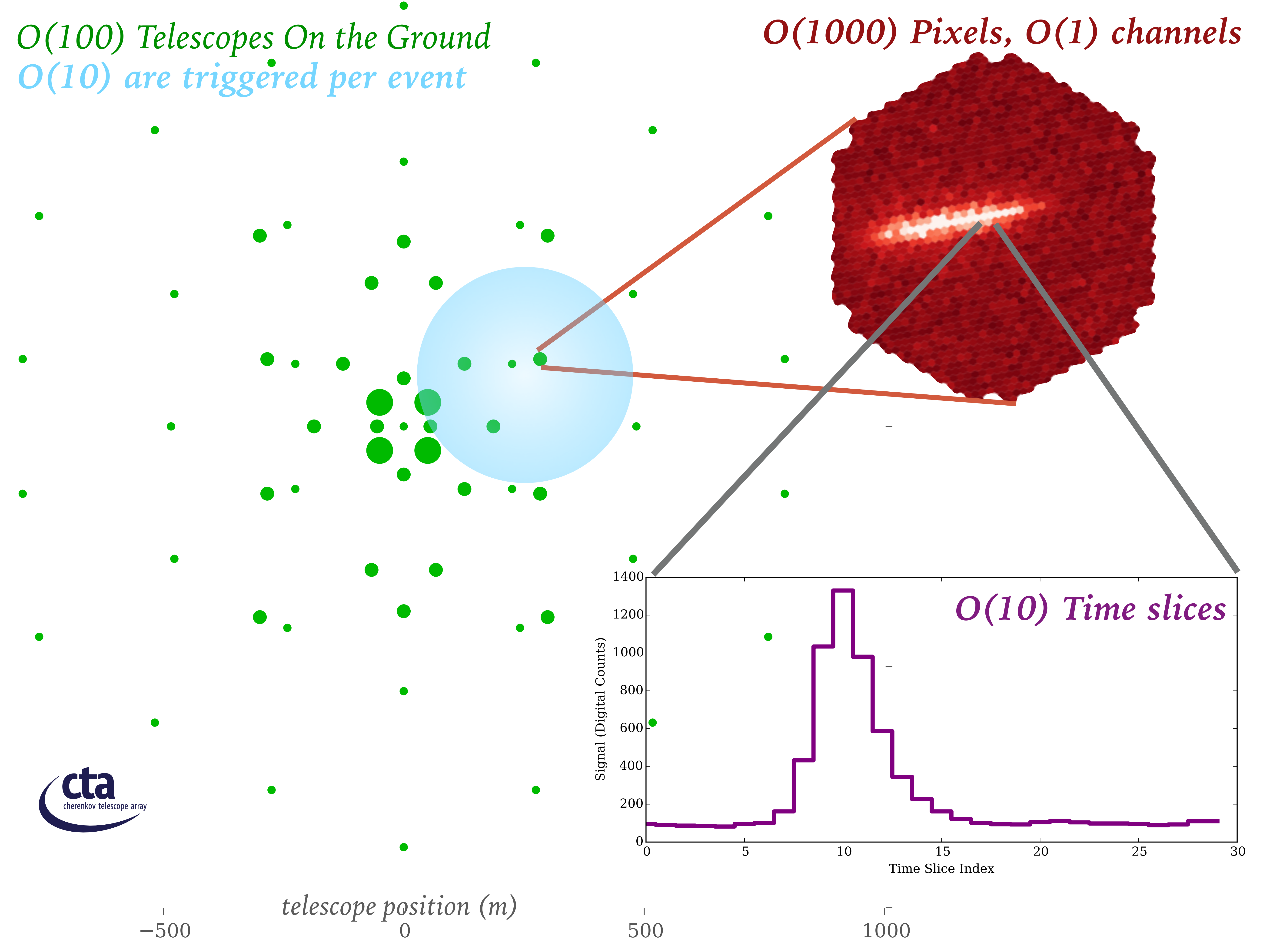
- CTA Data Pipelines coordination (KK)
- Site Infrastructure
- CTA Key Science

cta
cherenkov telescope array

credit: DESY/Milde Science Comm./Exozet



Using Cherenkov radiation from particles in extensive air showers to detect gamma and cosmic rays in the atmosphere



O(100) Telescopes On the Ground
O(10) are triggered per event
O(1000) Pixels, O(1) channels
O(10) Time slices
Signal (Digital Counts)
0
200
400
600
800
1000
1200
1400
0
5
10
15
20
25
30
Time Slice Index
cta
cherenkov telescope array
telescope position (m)
−500
0
500
1000

DATA VOLUME

Trigger rate is O(10,000) Hz
(really more like 30kHz)

Data volume is therefore
O(10 tels • 1000 pix • 10 times • 10000 Hz)

- = O(10) GB/s
- = 10x CERN ATLAS

Non-trivial data volume!

- need to reduce by a factor of >**20-100x** on-site (compression and suppression)
  - implies robust software,
  - streaming, possibly real-time
- even afterward estimate **4 PB/yr**
  - will want to re-process it all at least annually! (grows in time)
  - push I/O (and CPU) limits
  - parallelism is *strictly required*

Big* Data.

*** (well, at least for high-energy astrophysics! yes, SKA is much worse…)**

END PRODUCTS

Each "event" is reconstructed:

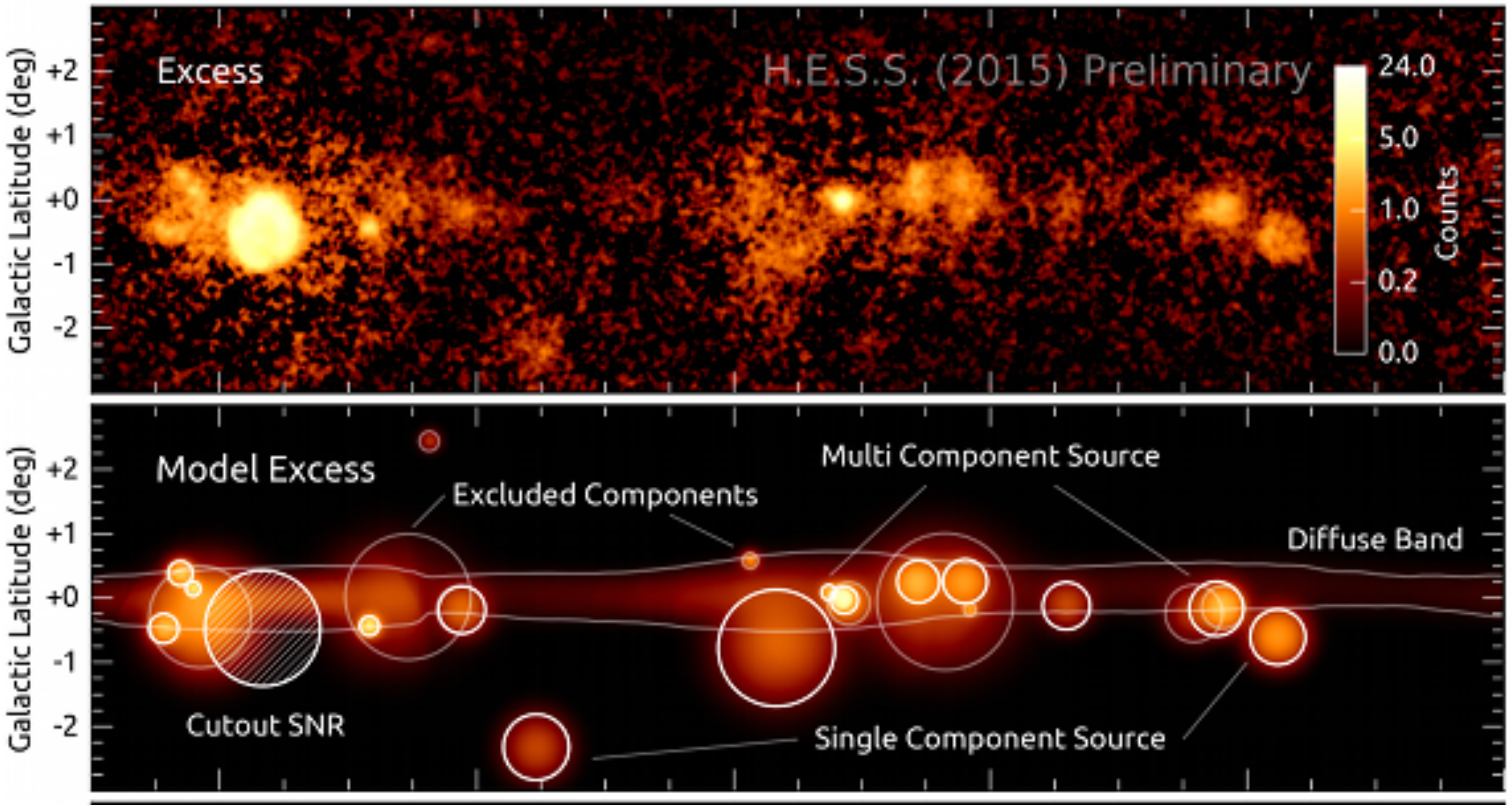
- Energy
- RA/Dec
- Type (gamma? Hadron? Electron? Muon?)

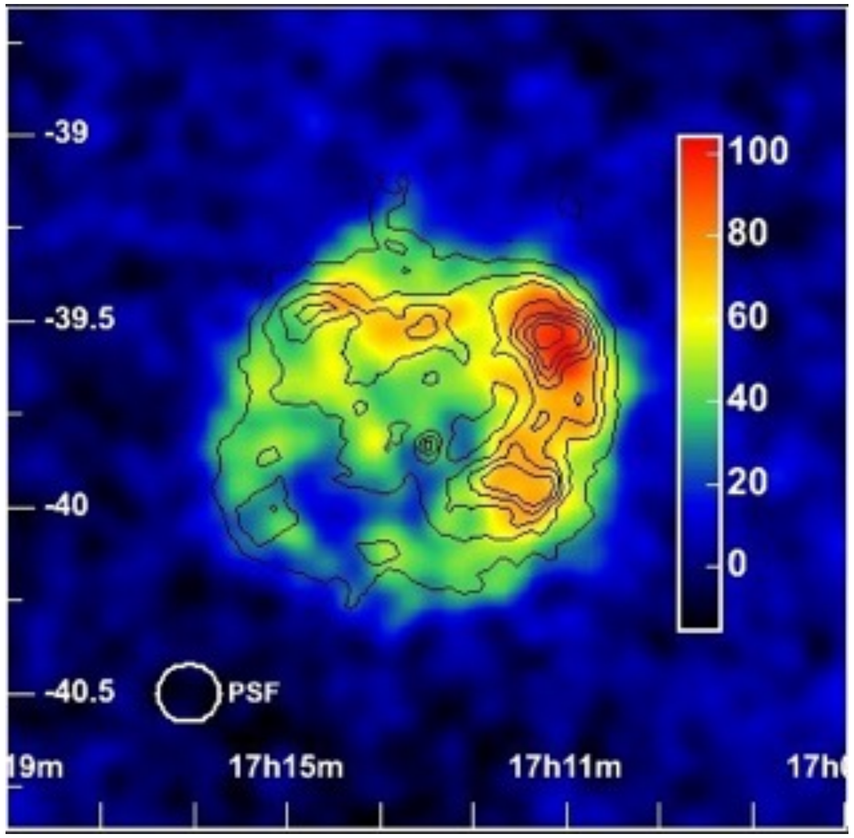
Final data products accumulate events and make *statistical* gamma-ray:

- images
- spectra
- lightcurves…

Challenges:

- CR background dominated
- instrument response varies with atmospheric conditions
- energy dependent PSF and FOV
- source confusion





TOOLS/ALGORITHMS 1

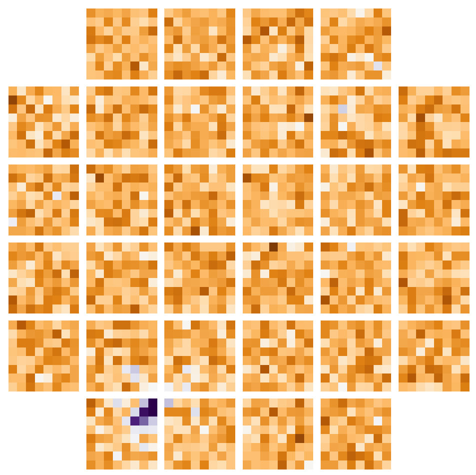
Generalization of "Images"

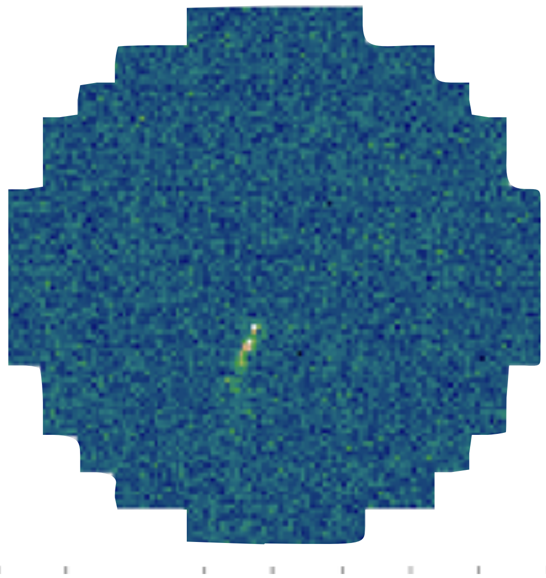
- non-square pixels, **triangular** or **cartesian basis**, **gaps** (6+ different camera types/geometries/technologies)
- data cubes in time, or images of *time parameters*

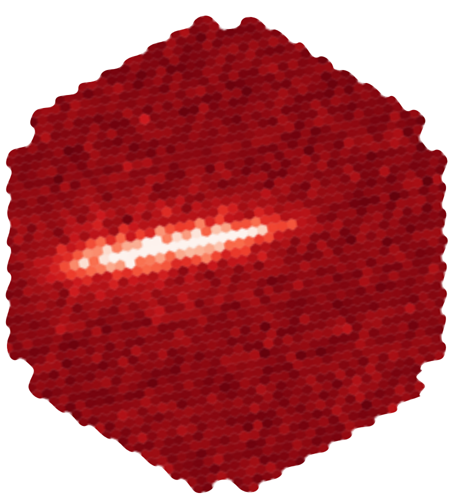
Image and Signal processing

- signal processing: peak finding, integration
- calibration (background, flatfield, time, optics…)
- image de-noising and in-painting (identification of signal region and missing information)
- image feature extraction (characterization of image)
- advanced techniques (wavelets, compressed sensing, and beyond)

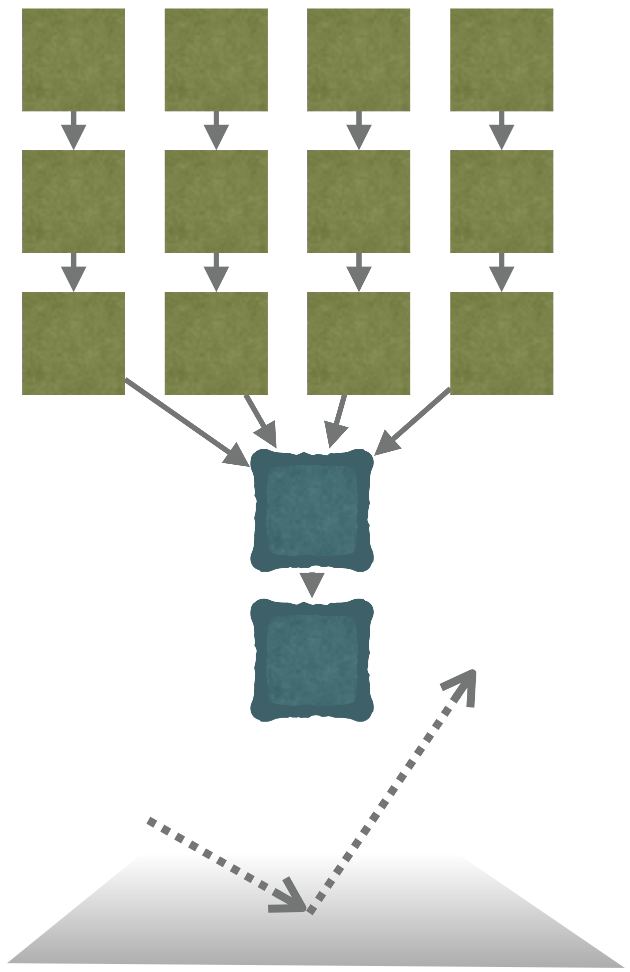
Current focus of our group





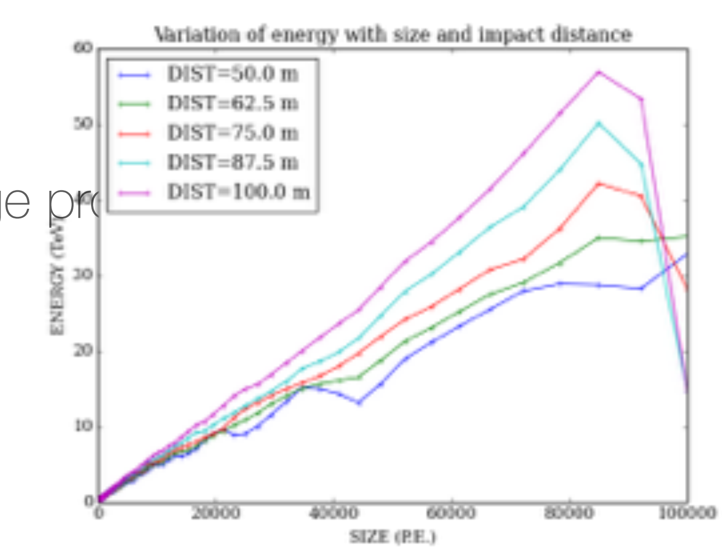


TOOLS/ALGORITHMS 2



Event Reconstruction:

- **data synchronization**
  (“join” operation on multiple telescope streams)
- **likelihood minimization** (with large dimensionality)
- ND **interpolation** (where N>3)
- 3D geometry and **linear algebra**
- **coordinate transformations**
  (in addition to standard astronomical ones)
  - ➤ detector plane for each telescope, including pointing corrections
  - ➤ nominal plane (common view of shower from all telescopes)
  - ➤ impact or ground plane (where the shower hits)
  - ➤ need for speed optimizations here (could contribute)
- **machine learning** (regression)
  - ➤ energy or shower determination from many input parameters
  - ➤ may explore even deep learning (convolutional neural networks, etc) for image pr



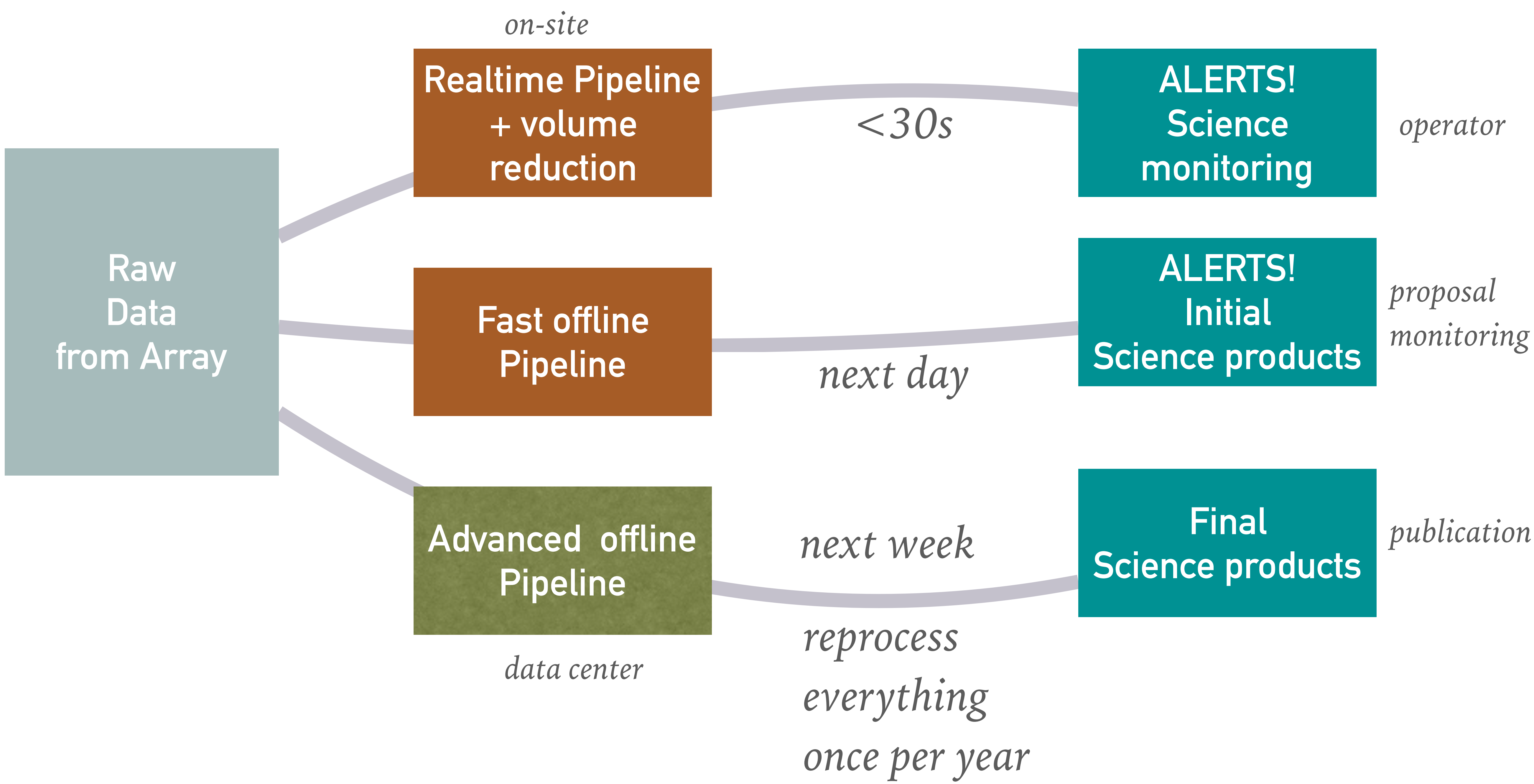
Event Classification: (gamma, proton, electron, muon?)

- **machine learning** (classification)
  - ➤ decision trees (BDTs, Random Forests) or Support Vector Machines, etc.

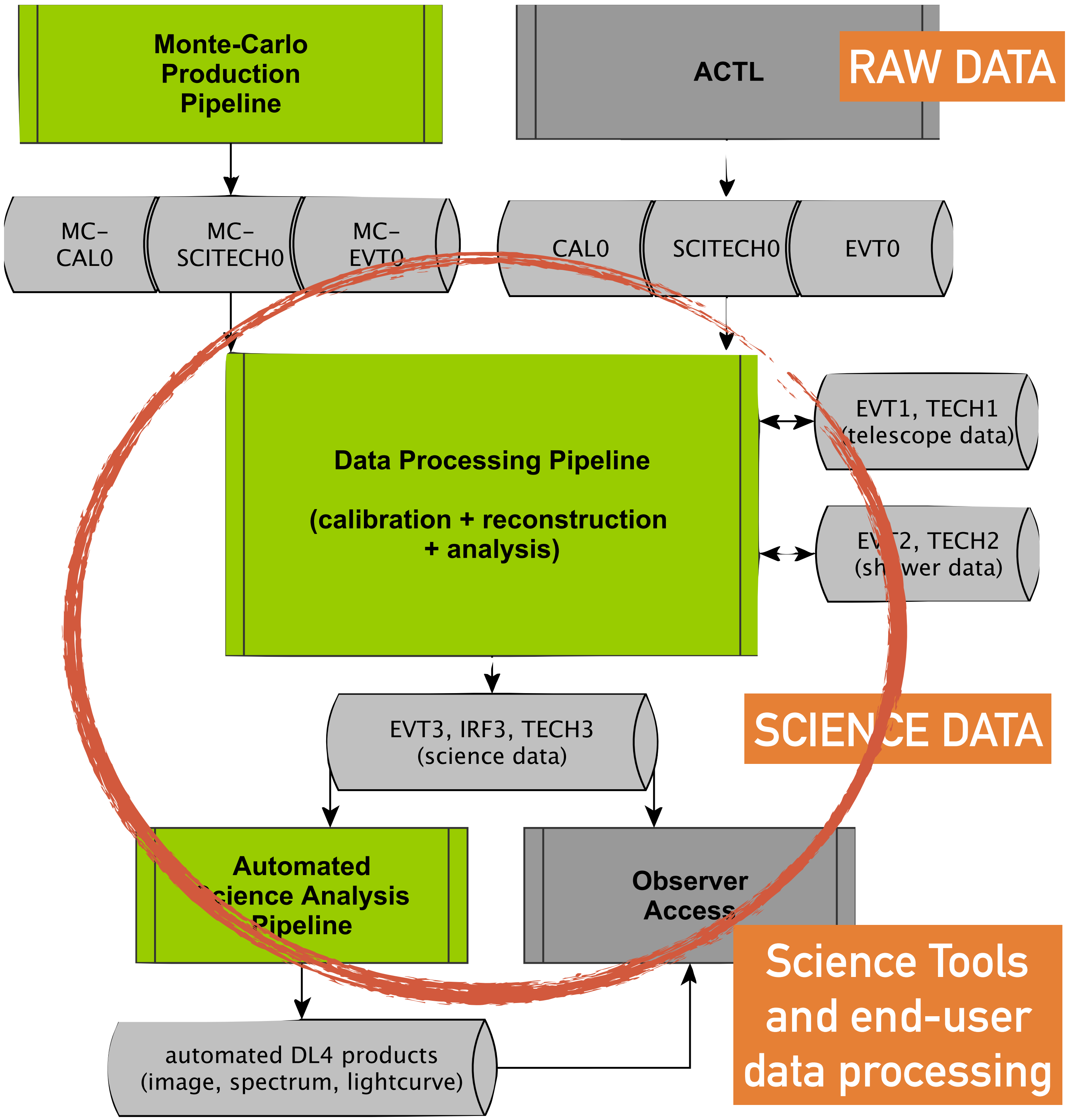
MAIN CHALLENGE

Efficient data processing!

- want to improve sensitivity *and* data processing speed!



OBELICS CONTRIBUTIONS



Clearly some strong connections between OBELICS and CTA Data Processing needs

CEA group

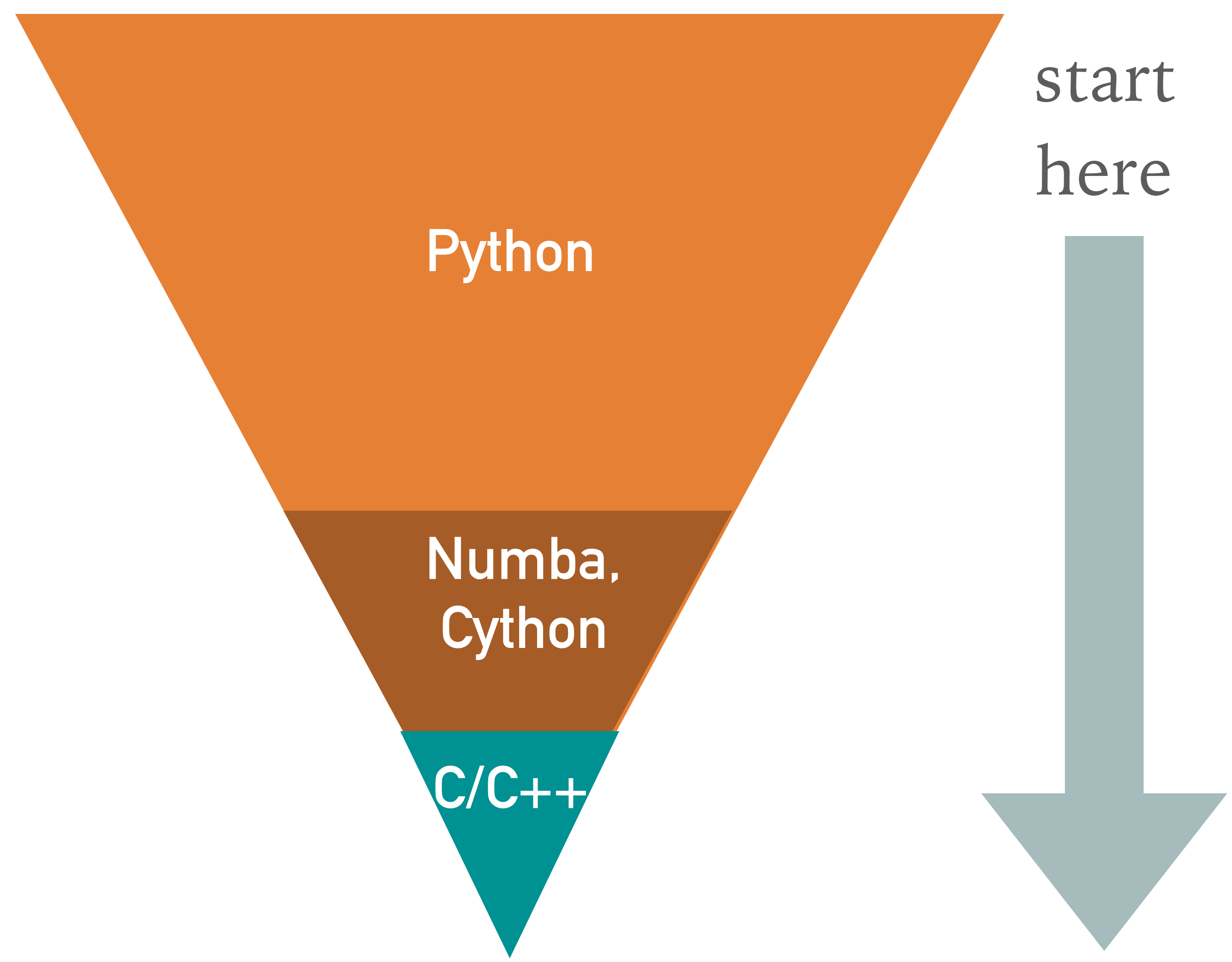
- Declared on task 3.4
- strongly involved in the data processing pipeline for CTA
- contribute to first software library (tools from CTA that could be used by others)
  - signal processing?
  - sparsity?
  - machine learning?
  - Higher level science tools?

BUILDING A FRAMEWORK

Bottom-Up approach

Most current frameworks did it this way (if they use python at all)

Top-Down approach



Our approach: start early with python and high-level API

common “core” package

ctapipe will be **glue** between various components. Provides common APIs and user interfaces packaging, etc.

- Develop new code quickly
- incorporate existing code from prototypes in a common place
- Unify documentation