Event reconstruction for IACTs

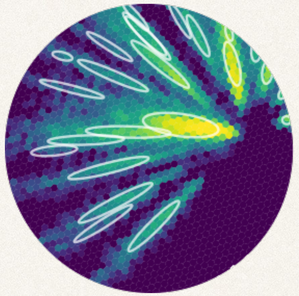
Atreyee Sinha, TIFR

Many slides taken from previous CTAO schools lectures by Maxmillian Linhoff

Data levels in CTAO (and others…)

- R0: Raw data (output of camera/ simtel simulations)
- R1: Corrected data
- DL0: First levels of calibration - pixel-wise time series for each telescope
- DL1: Image parametrization
- DL2: event reconstruction and classification
- DL3: event lists and IRFs - Lowest data level given to end users
- DL4: Science binned data
- DL5: Science products (spectra, lightcurve, significance…)
- DL6: catalogs

data volume reduction at each step



ctapipe

What we cover today

- How to go from waveform information to images
- Image parametrisation
- Image reconstruction (mono + stereo)
- Instrument Response Functions - briefly

Aim: “Invert” everything that happened to the gamma-rays in the atmosphere and the detectors.

What we want from reconstruction

- For each event, we want to know
  - Particle type
  - Primary energy
  - Arrival direction
  - Arrival time

How well we can do this reflects our detector performance.

Since we have no way of reaching ~10s of TeV in labs, we depend heavily on simulations

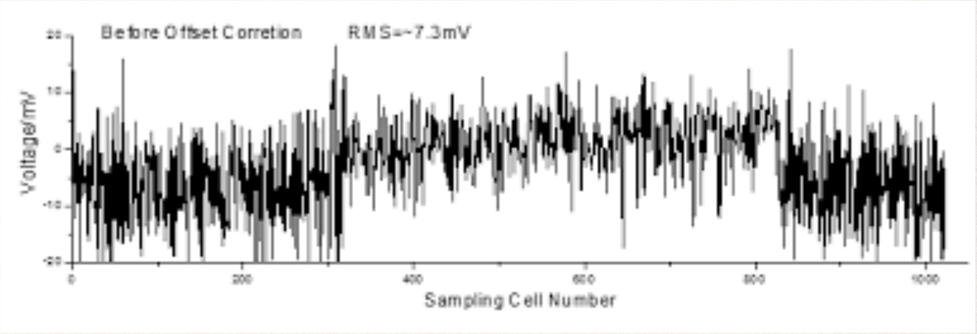
From raw data to images

What is raw data

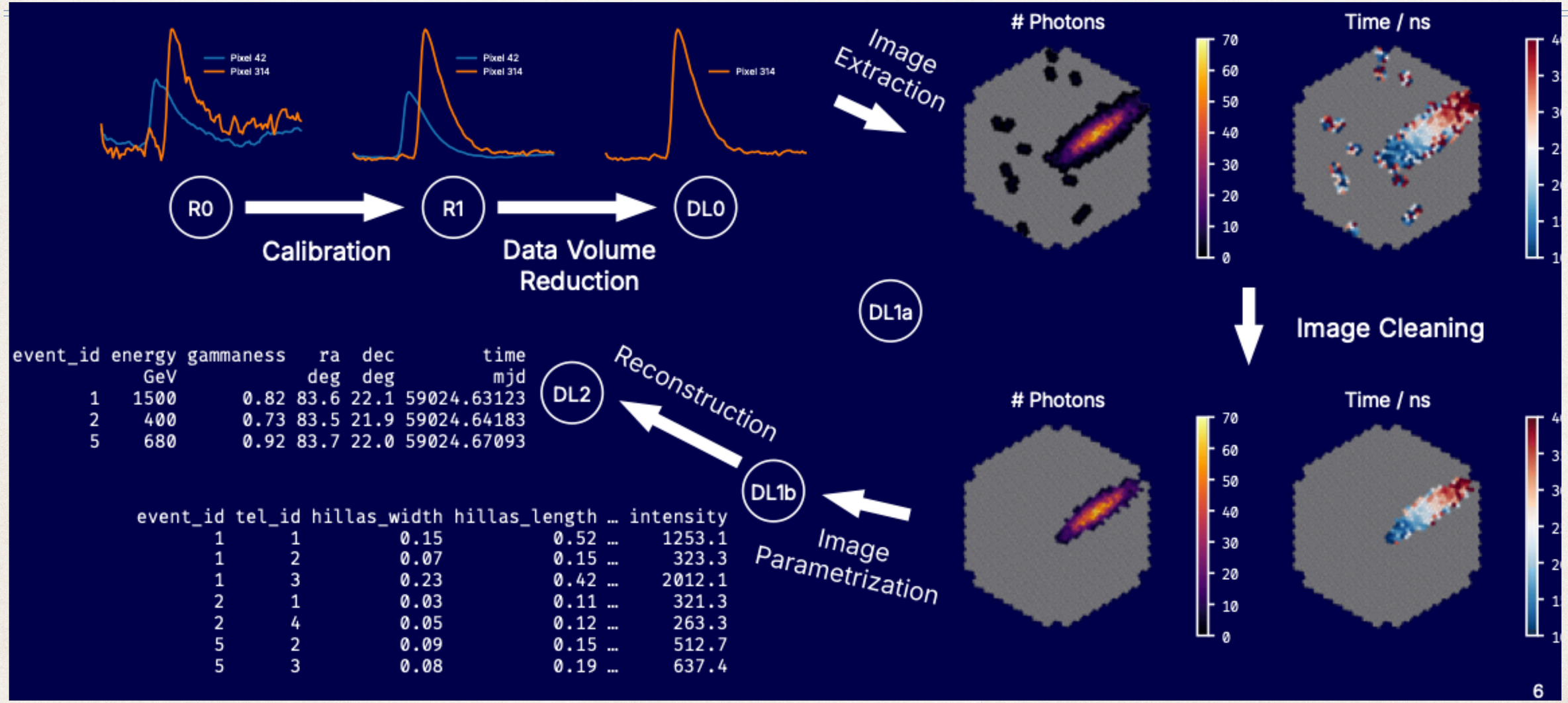
- PMT or SiPM signals are readout at high frequencies ~ 1 GHz
- Camera level triggers look for “interesting” data
- Telescope triggers are sent to the array trigger, which decides if data should be stored or not
- In case of a trigger, the cameras digitize the pixel data in a window around the trigger time.
- Seras apply two different amplification gains to the analog signal to increase dynamic range

Raw data calibration

- Basically correction for the readout systems
  - Eg: DRS4 chips baseline correction
- Removal of spikes, etc
- Dual gain channels
  - Increase dynamic range, use two channel per pixel with different gain (~20)
  - Use HG by default
  - If HG saturates, choose LG
  - Use overlap for cross calibration



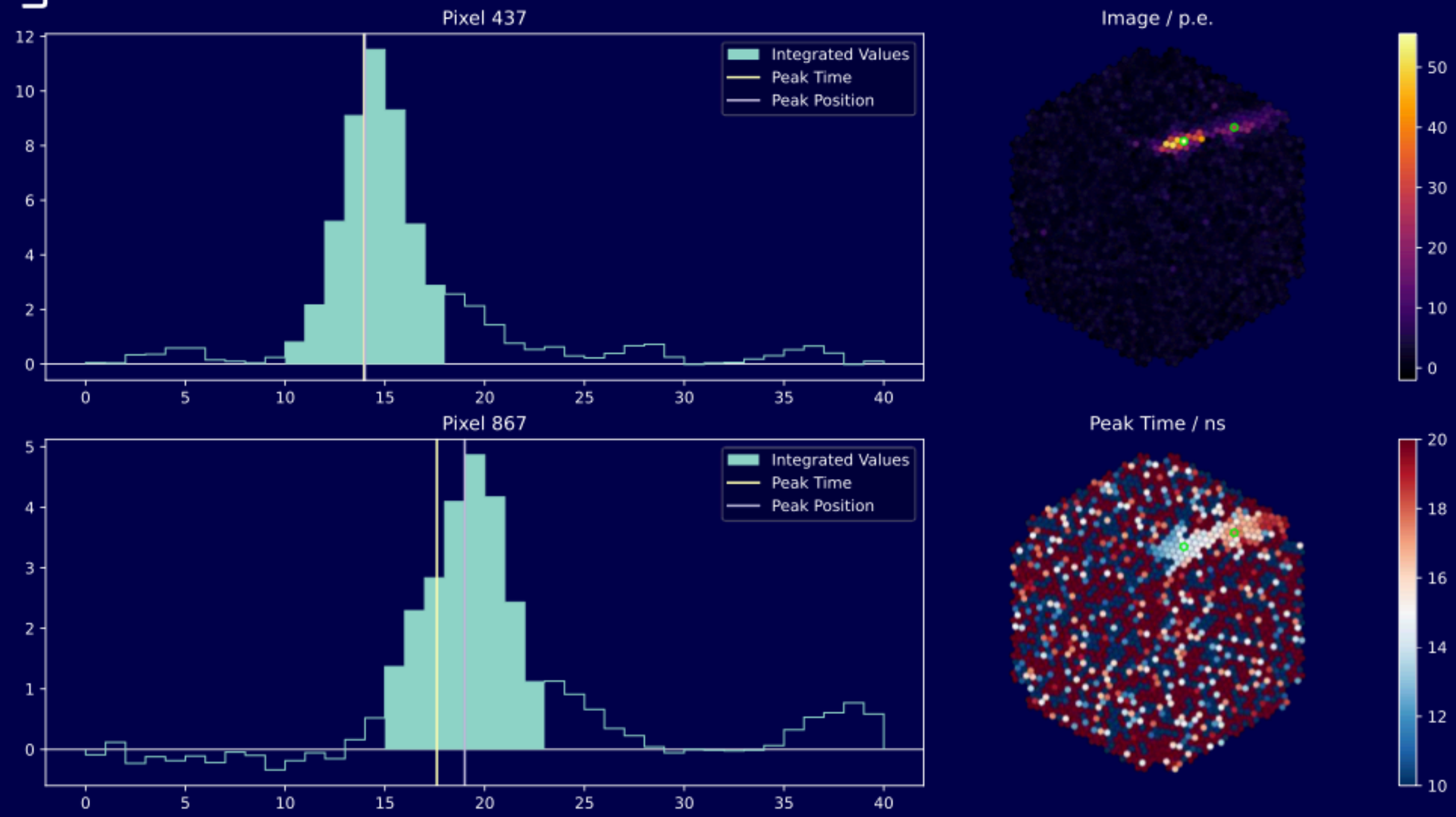
Data levels upto DL2



From DL0 to DL1

- DL0: full pulse for every pixel
- DL1: pulse arrival time and charge (in photo electrons) at each pixel
- Huge data volume reduction: ~ 100s of numbers to 2 / pixel
- Several algorithms exist: find peak, integrate around it…
- Special runs required for calibration
  - Flat field: pulse arrival time and charge intensity on each pixel for uniform illumination
  - Dark pedestal: Randomly triggered event without HV switched on, closed shutter
  - Single pe: Randomly triggered event with HV switched on, closed shutter
  - Sky pedestal: Randomly triggered event with HV switched on, open shutter

Image Extraction



DL1 for gamma and proton

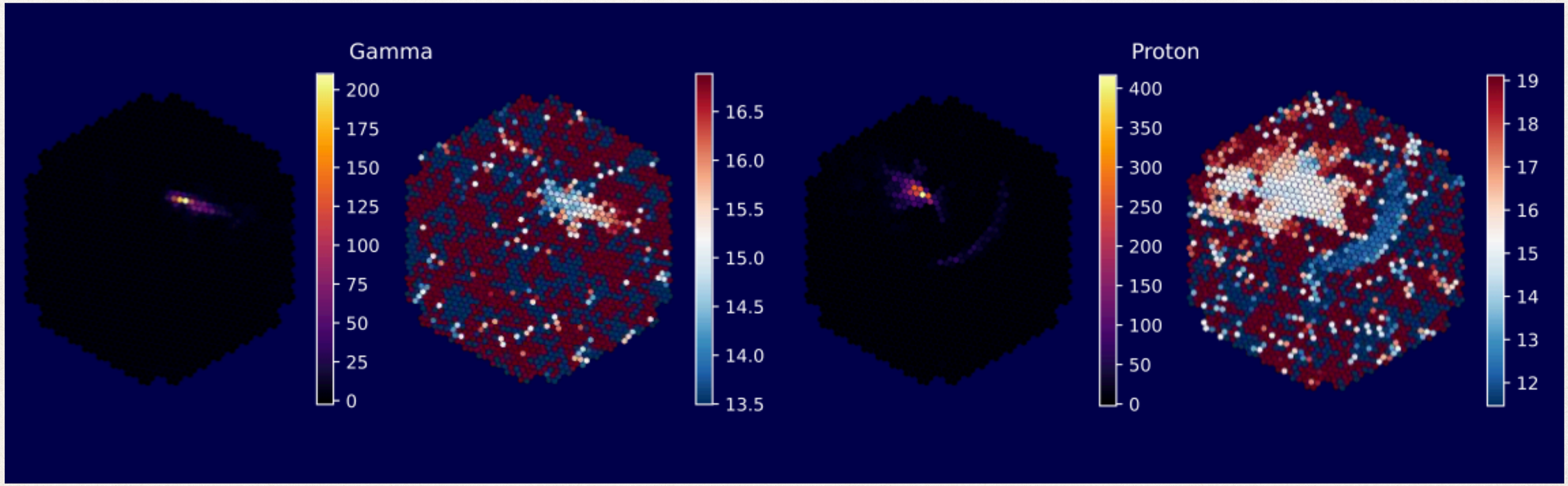


Image Cleaning

- Most pixels in any given event only contain noise
  ⇒ Identify pixels likely containing Cherenkov light
- Many different algorithms
- Can use intensity, peak time and noise level estimated from pedestal events
- Simple Example: tailcuts clean:
  1. Identify pixels above first threshold
  2. Remove pixels with less than 2 neighbors
  3. Add neighboring pixels above a second, lower threshold

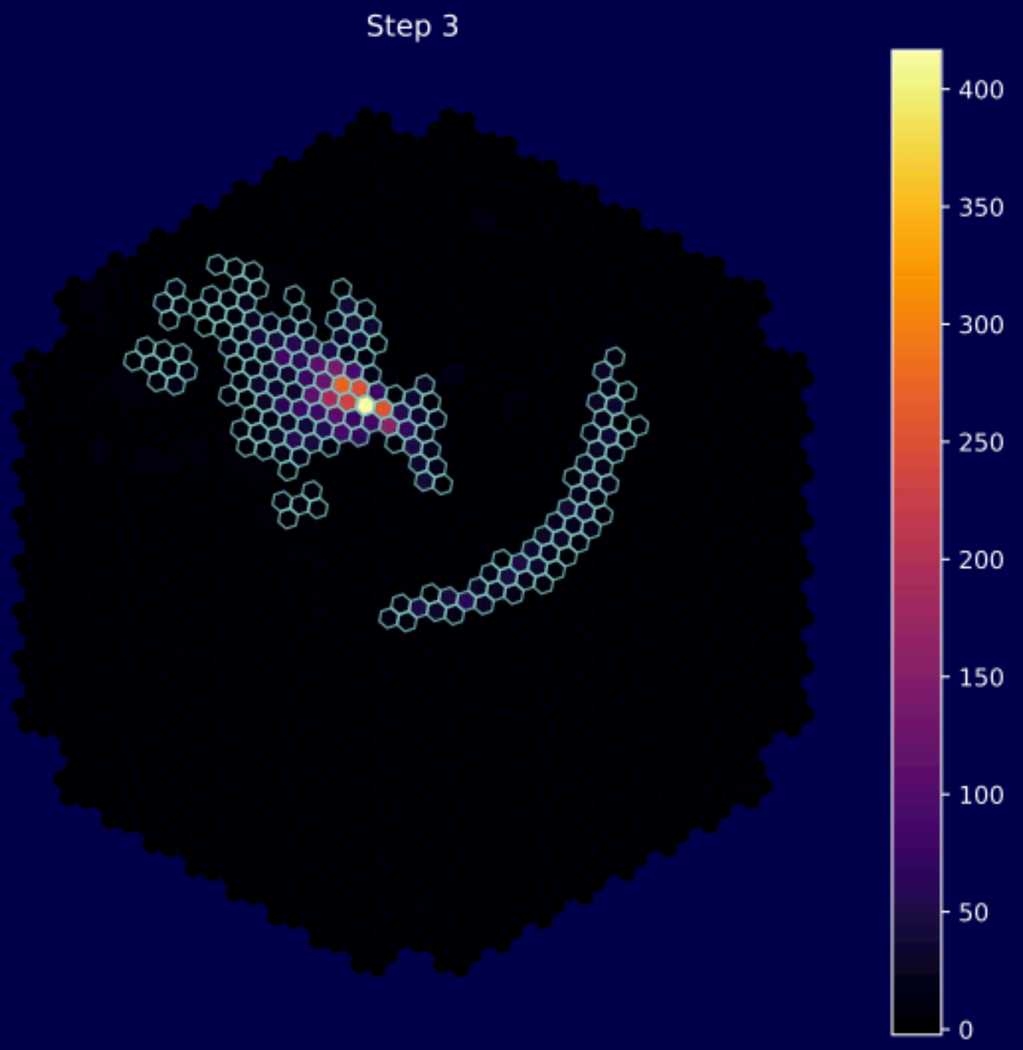
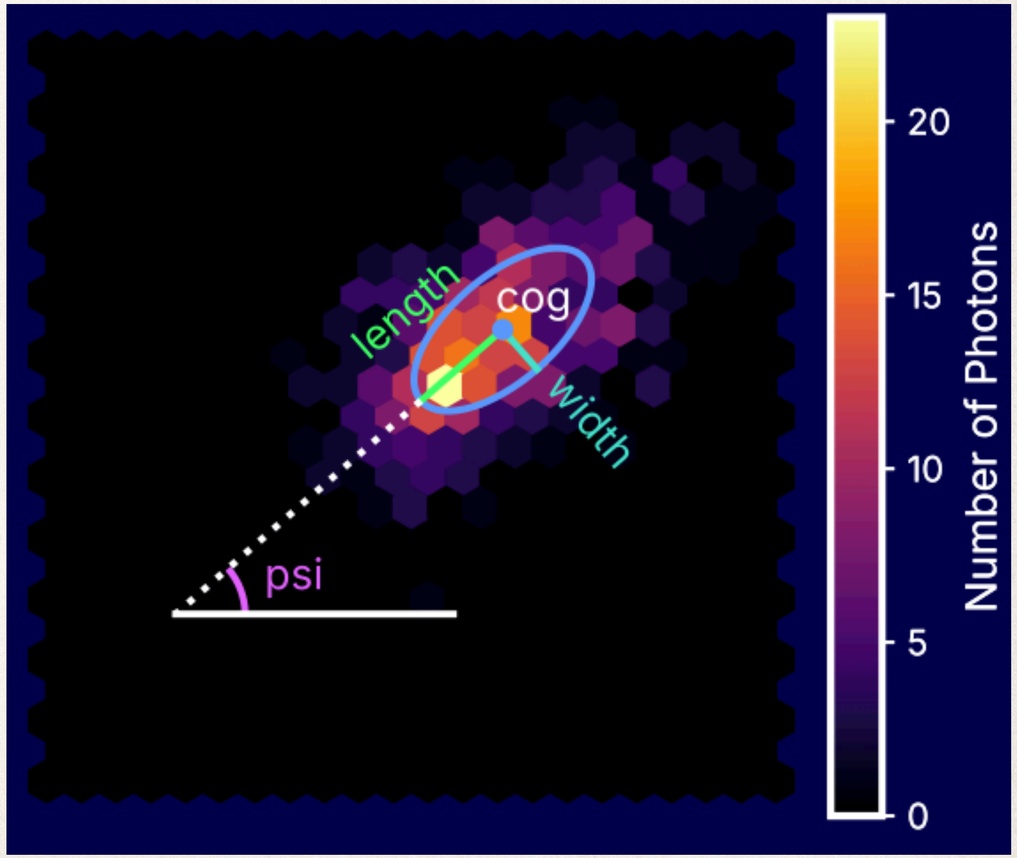


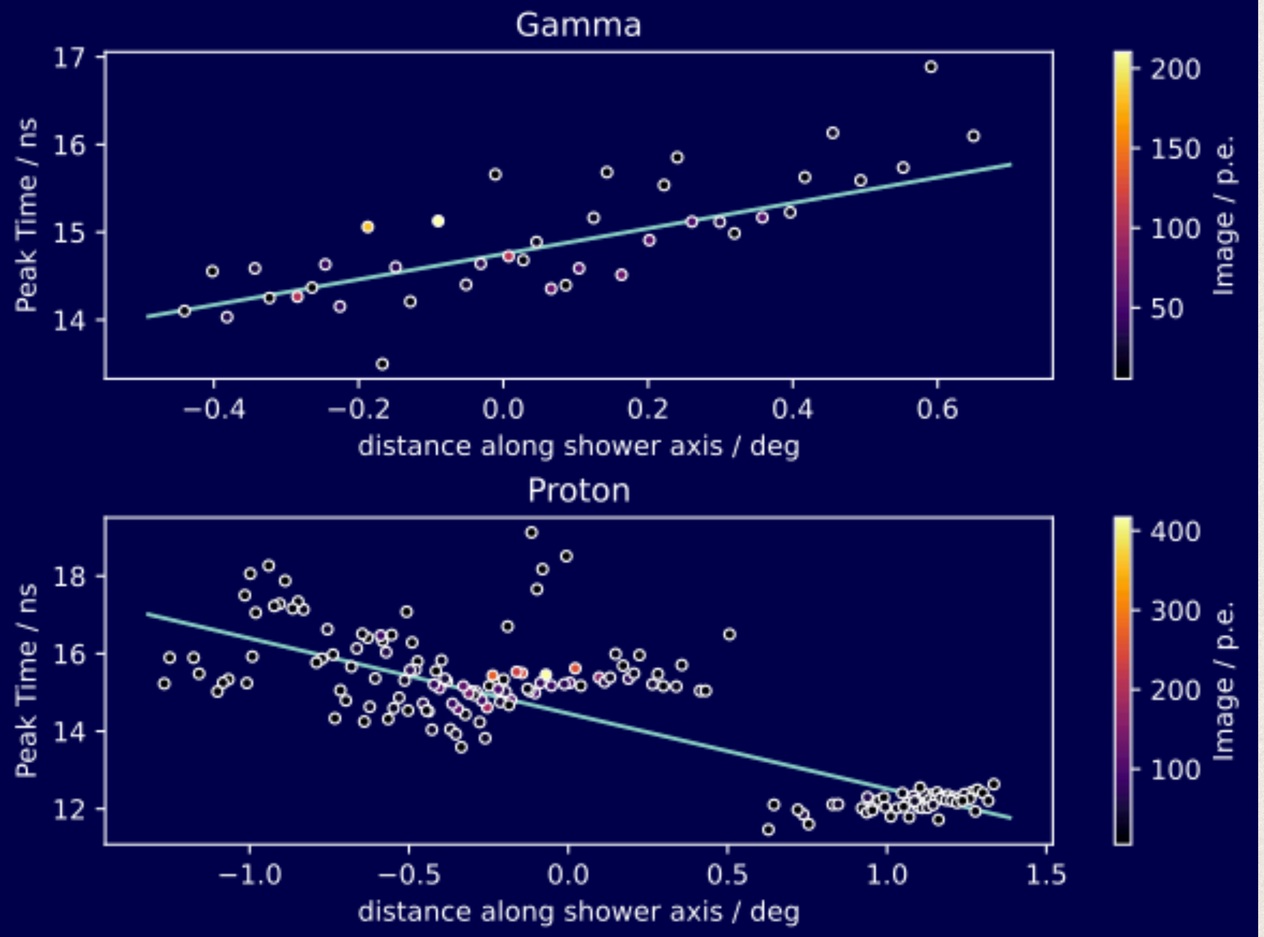
Image parametrisation

- Initial set of parameters proposed by Hillas in 1985
- Expanded over time
- Basically a PCA of a cleaned image
  - Weighted mean vector: center of gravity
  - Covariance matrix
  - Diagonalise covatriance matrix (eigenvalue problem)
    - Eigenvalues - length and width
    - Eigenvectors - orientation of major axis
  - Total number of p.e. : Intesnity
  - Higher level moments: skewness and kurtosis



Timing parameters

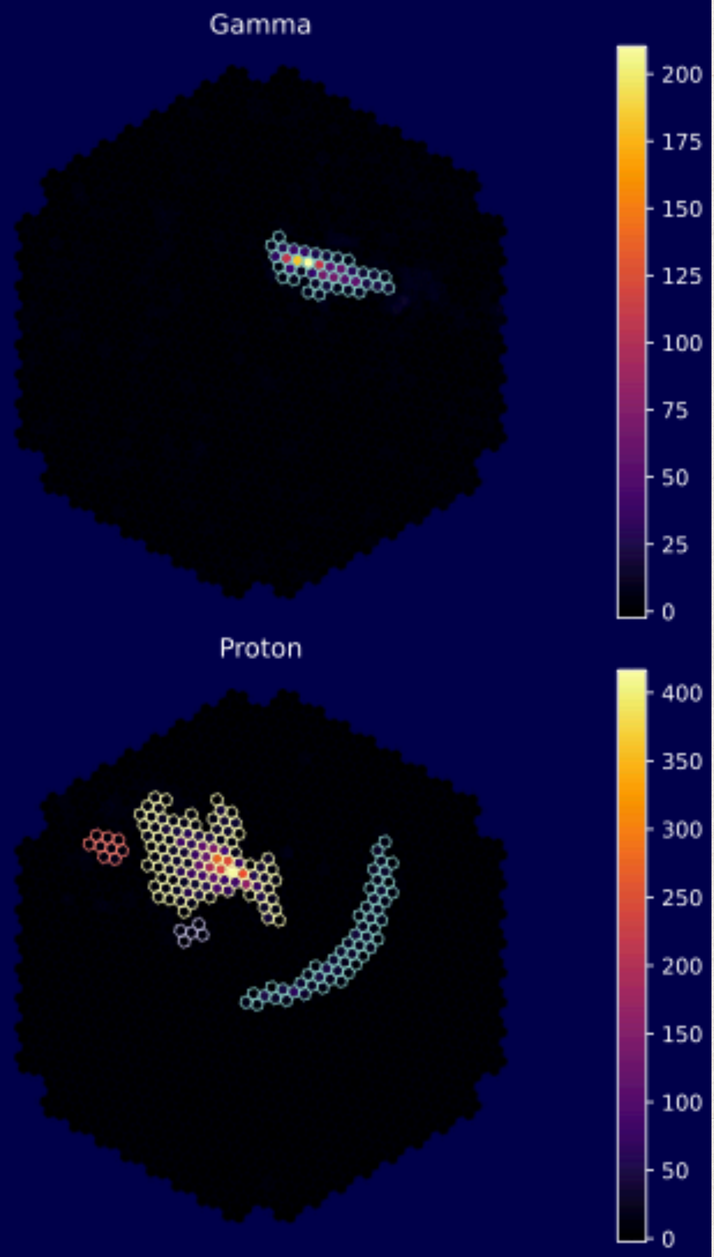
- After computing the Hillas Parameters, we can perform a fit of the peak times along the shower axis
- Slope is highly correlated with impact distance
- Sign of the slope is very useful for monoscopic direction reconstruction
- Good agreement with linear fit only for electromagnetic cascades, hadrons much more erratic



Morphology & Concentration

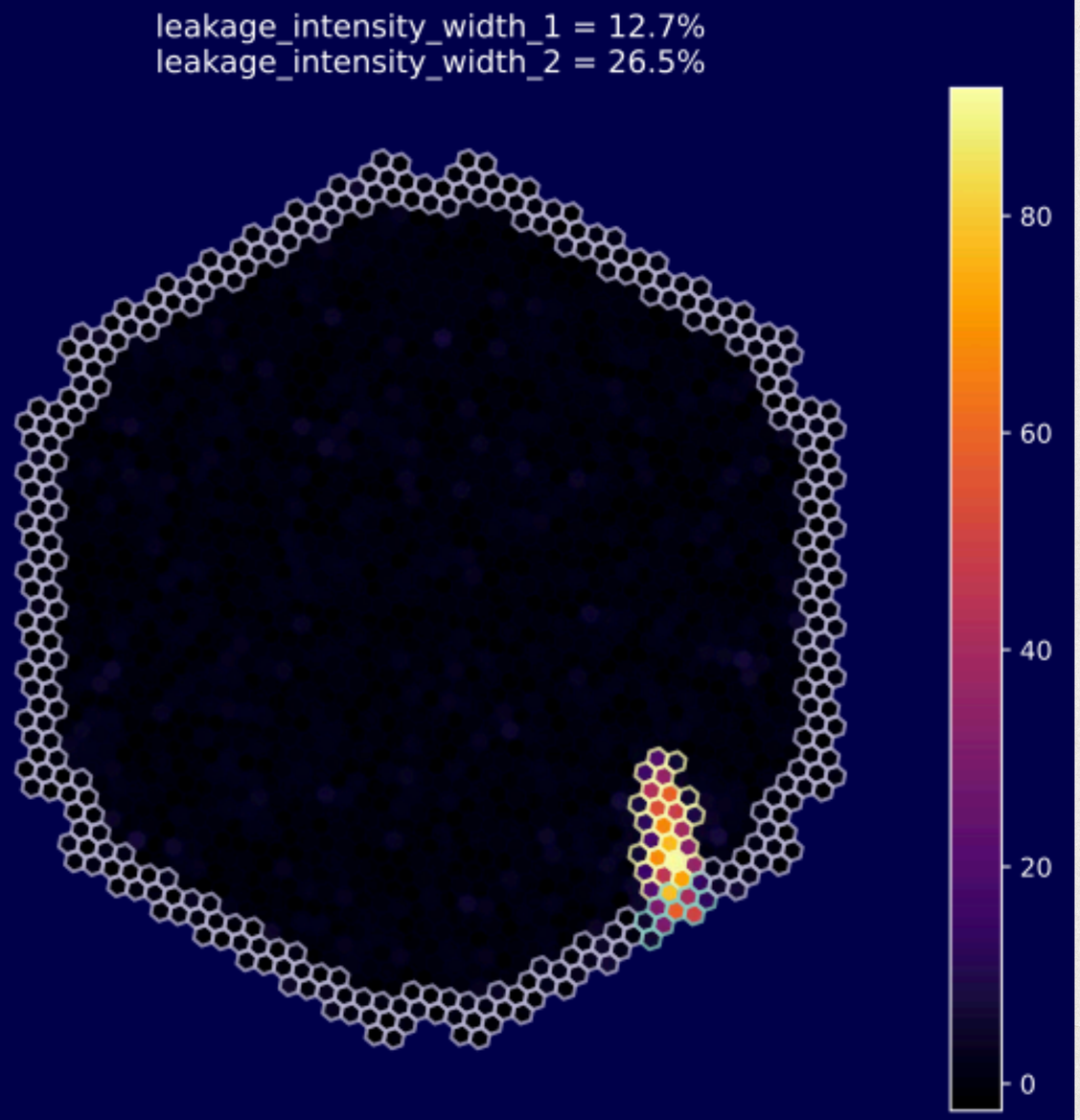
- For gamma-ray showers, we expect only a single blob of light
- Hadronic cascades have higher probability of sub-showers and muon rings
- Compute the number of “islands”, groups of connected pixels after cleaning
- Algorithm: breadth-first search of neighboring pixels
- Number of pixels after cleaning, number of all, small, and large islands
- Concentration features: ratios of intensity
    - inside of the Hillas ellipse
    - the brightest pixel
    - the three pixels closest to the cog

  to the total intensity

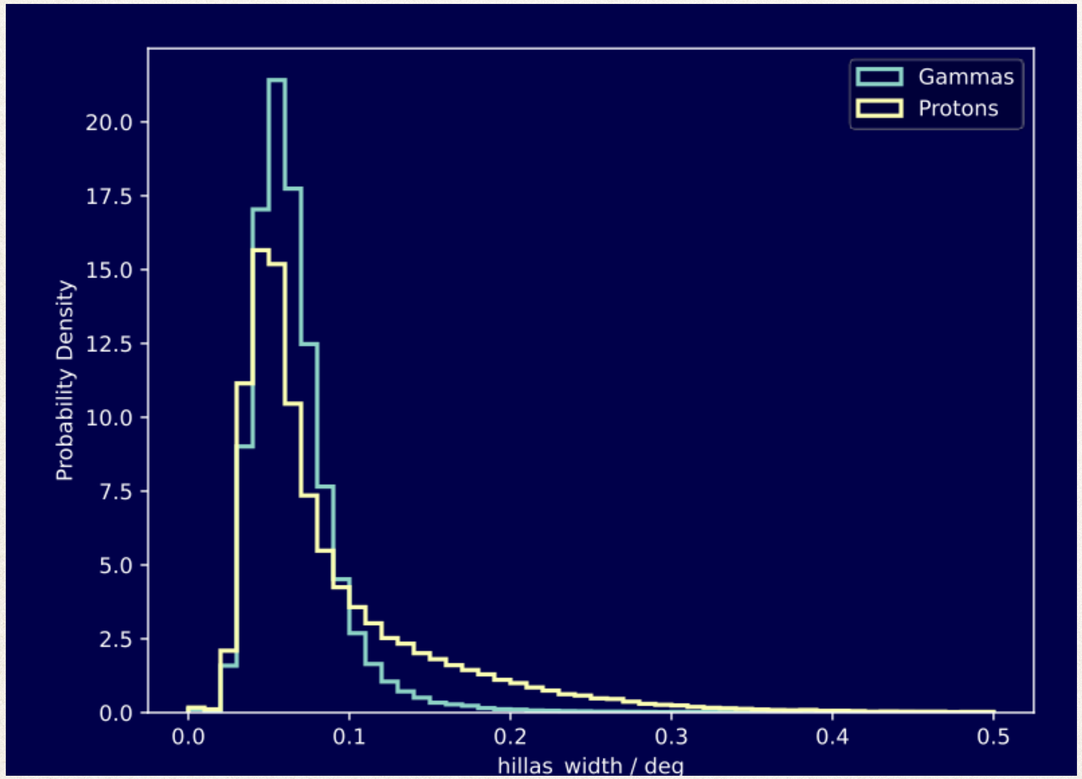


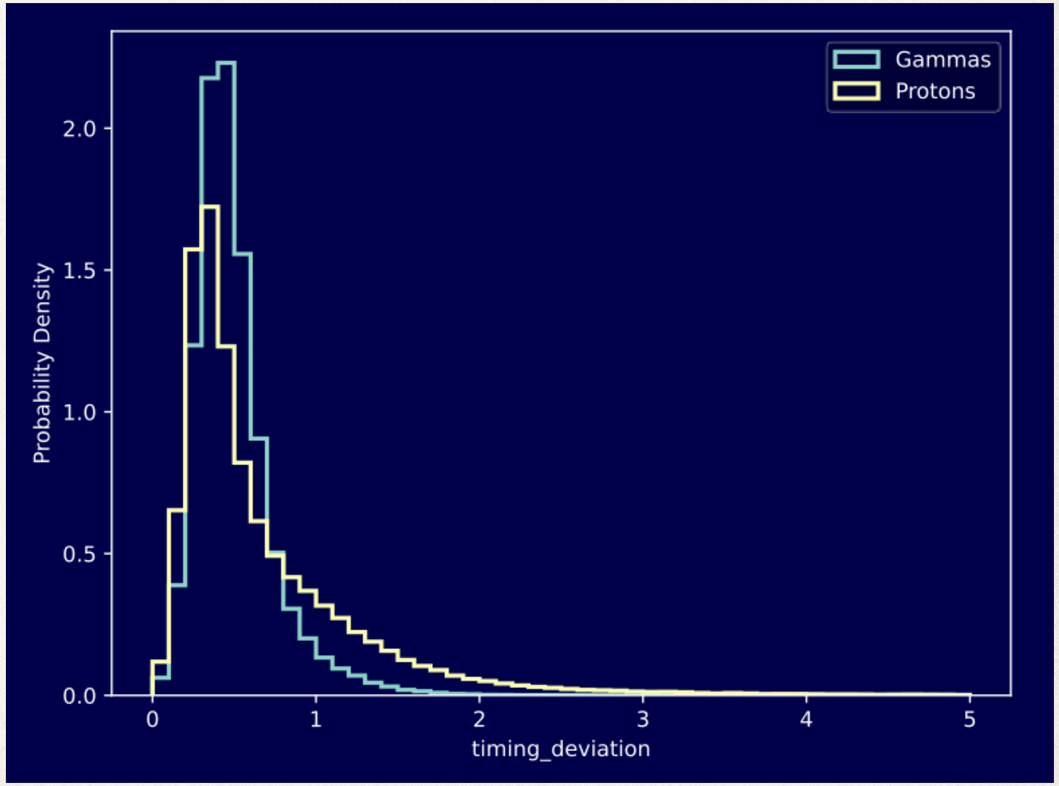
Leakage

- Ratios of photoelectrons or pixels after cleaning in the outer edge of the camera
- To describe how well an event is *contained* in the telescope's FoV
- Mainly important for energy estimation and discarding non-contained events



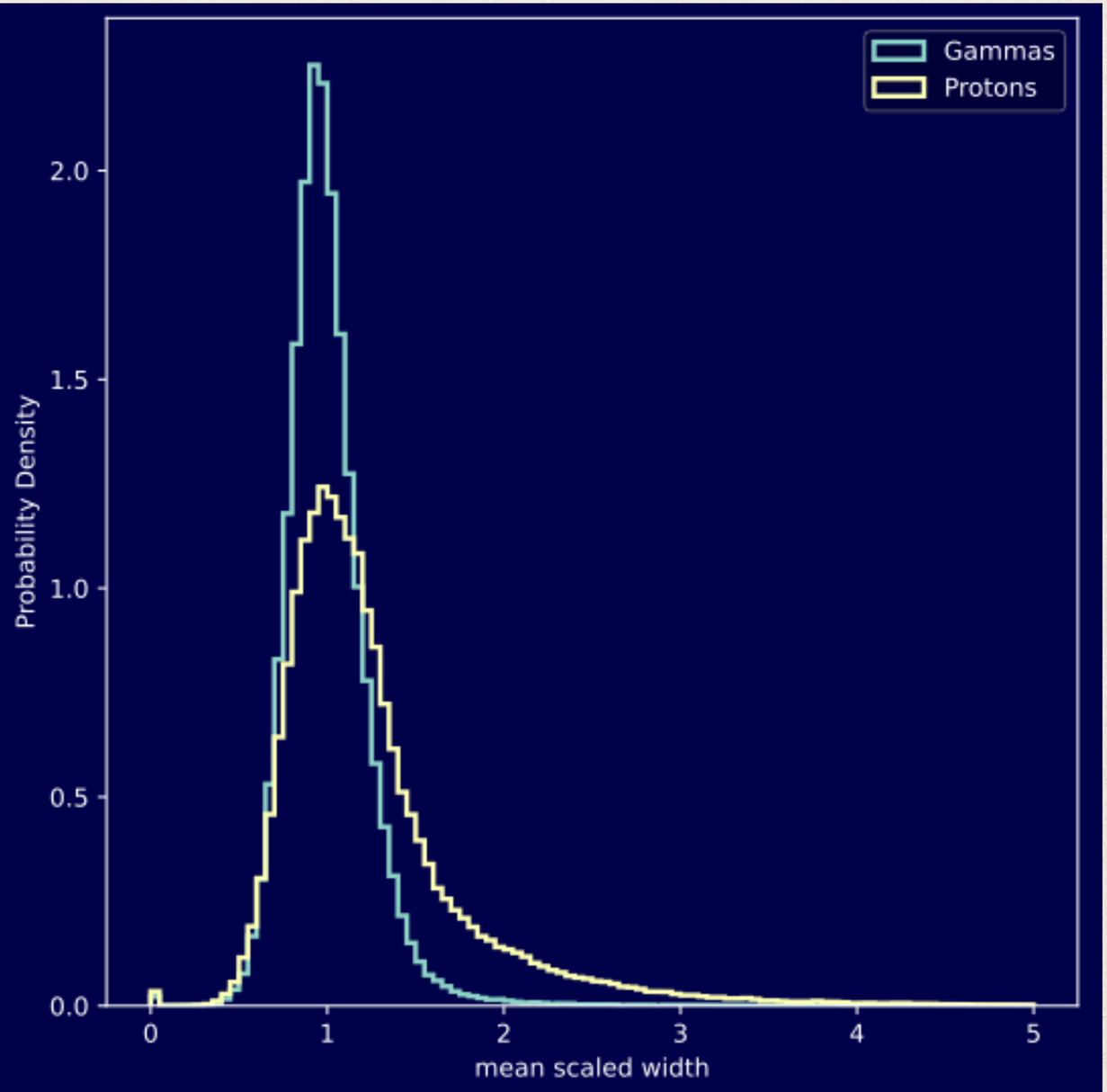
Gamma vs hadron eg



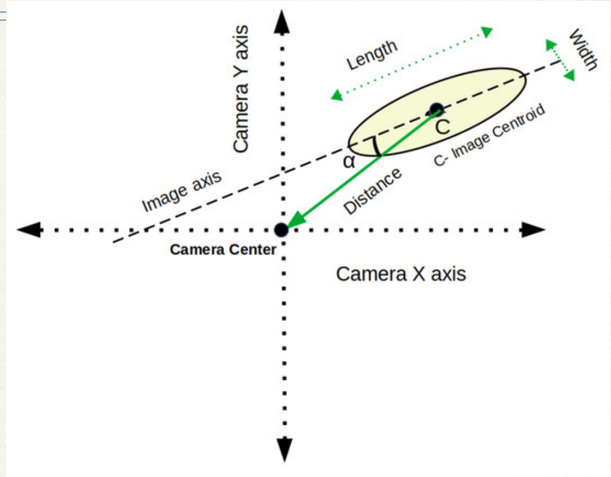


Mean Scaled Parameters

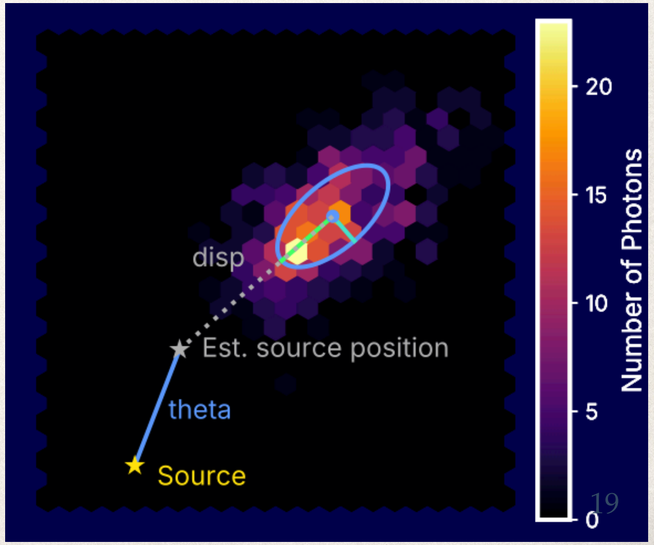
- This is a higher-level analysis step “from the old days, before machine learning”
- Still helpful to gain insights into parametrization
- Idea: scale image parameter by the mean of this image parameter for gamma rays (depending on image intensity, impact distance, zenith angle)
- Can also be averaged over telescopes
  ⇒ one number per subarray event
- Values around 1 are “gamma like”
- Values aways from 1 are “hadron like”
- Well suited for manual cut analysis or machine learning models that cannot deal with many features
- Less useful in e. g. tree-based learners



Monoscopic image reconstruction

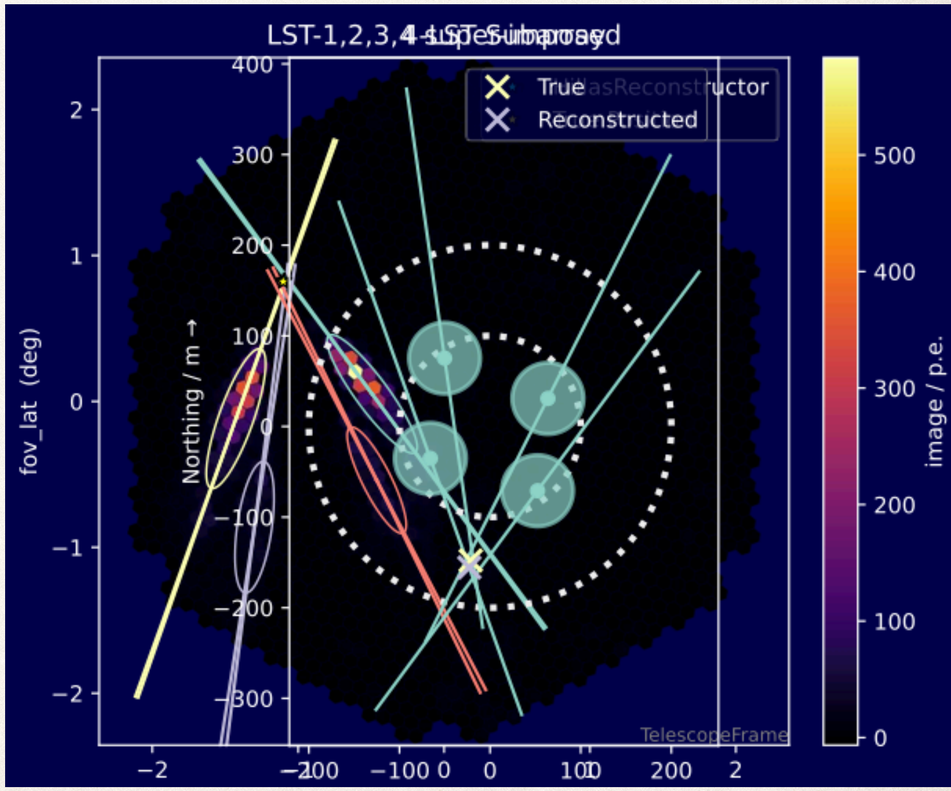


- Many methods
  - “alpha” cuts: source position knows
  - “disp” method: source anywhere in FoV



Stereo Image reconstruction

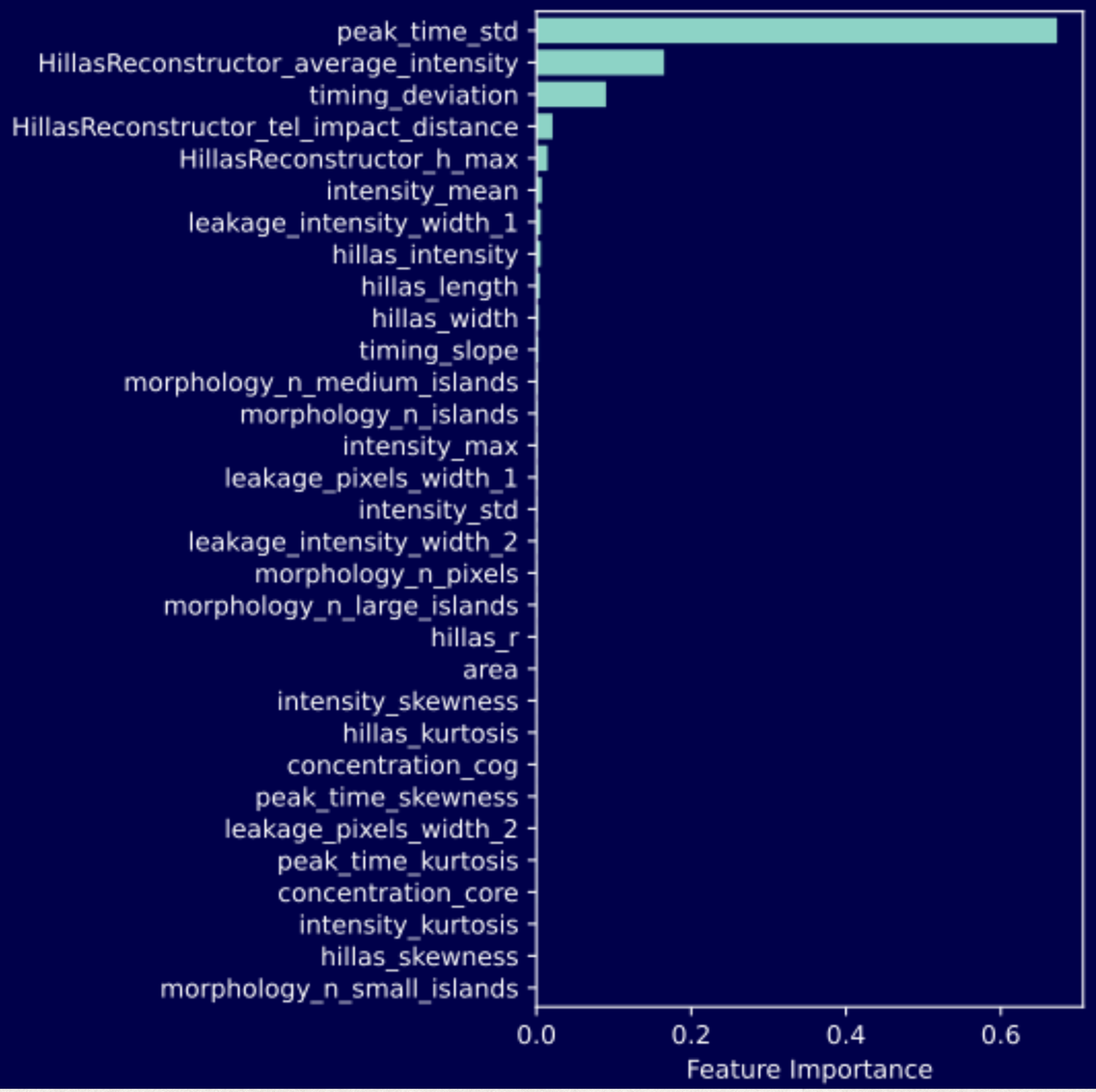
- Geometrical reconstruction of 3D shower axis
  - Direction in horizontal coordinates: altitude, azimuth
  - Impact point on ground: x,y
- Least squares intersections of planes spanned by each telecsope
- Requires at least 2 telescopes, more the merrier

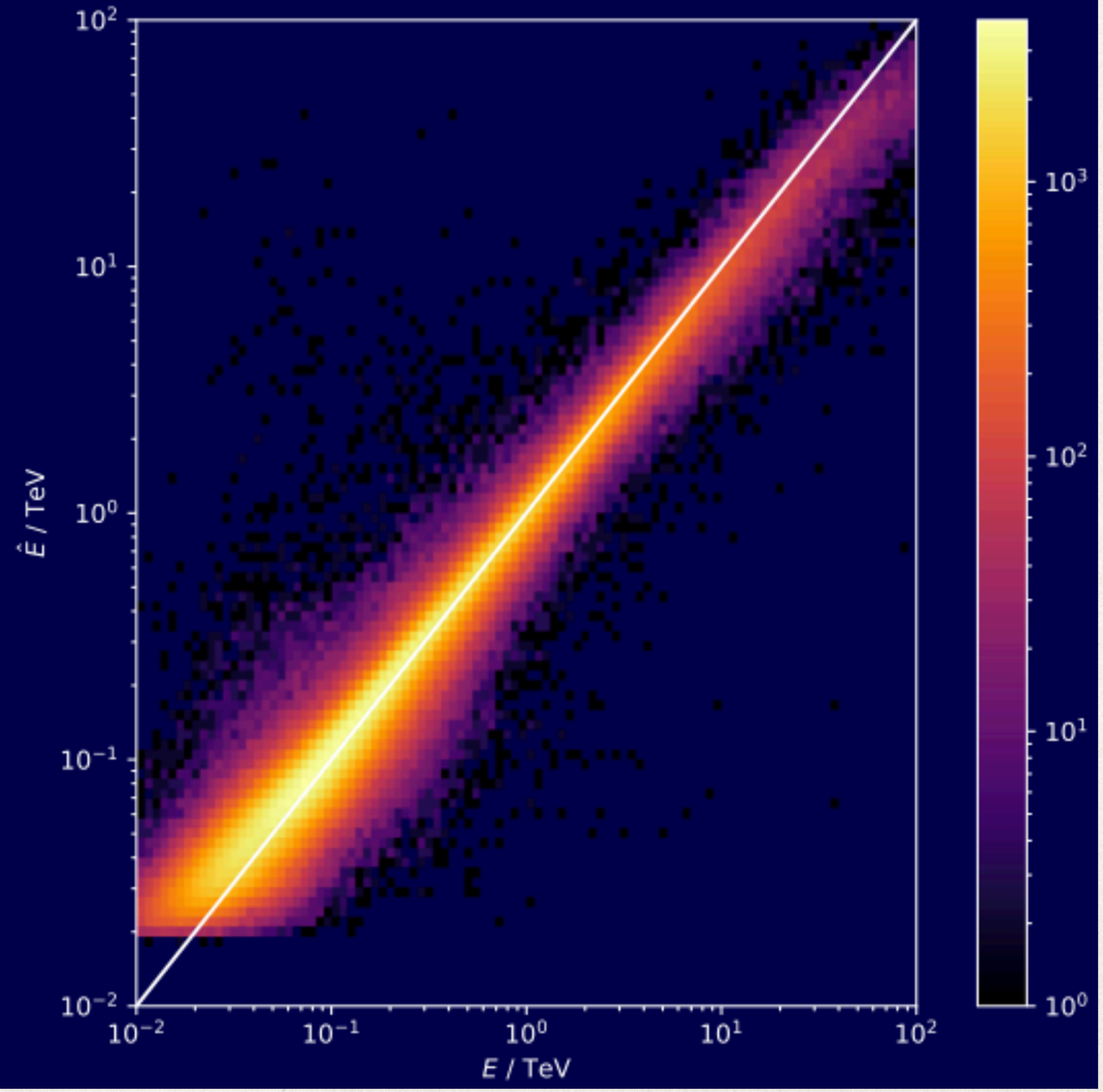


Energy reconstruction

- There is no way to “calibrate” IACT energy reconstruction purely by measurements
- Tree based ensemble methods have proven robust and performance
  - Random forest
  - Boosted decision trees

Energy Reconstruction





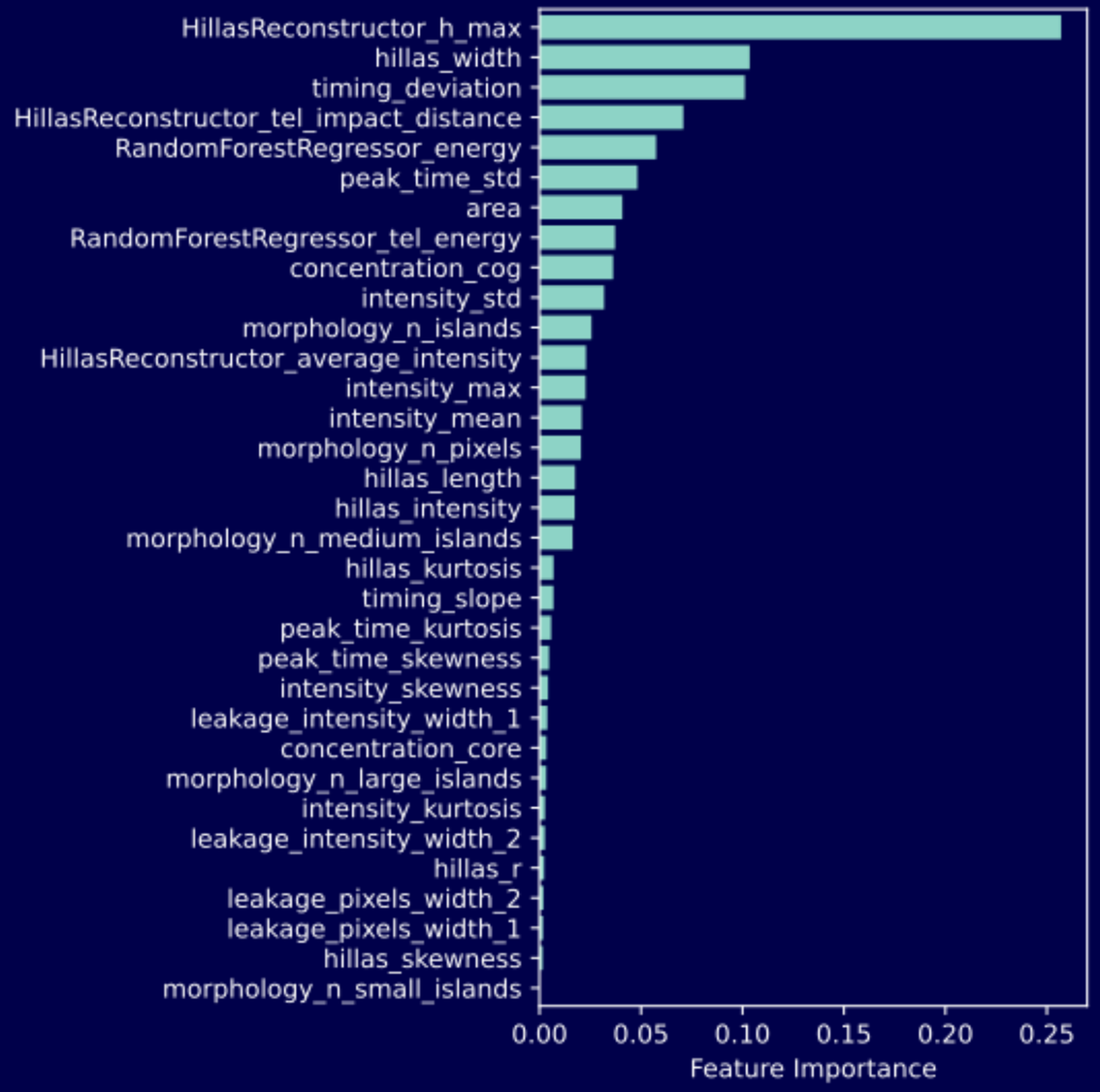
Particle Type Classification

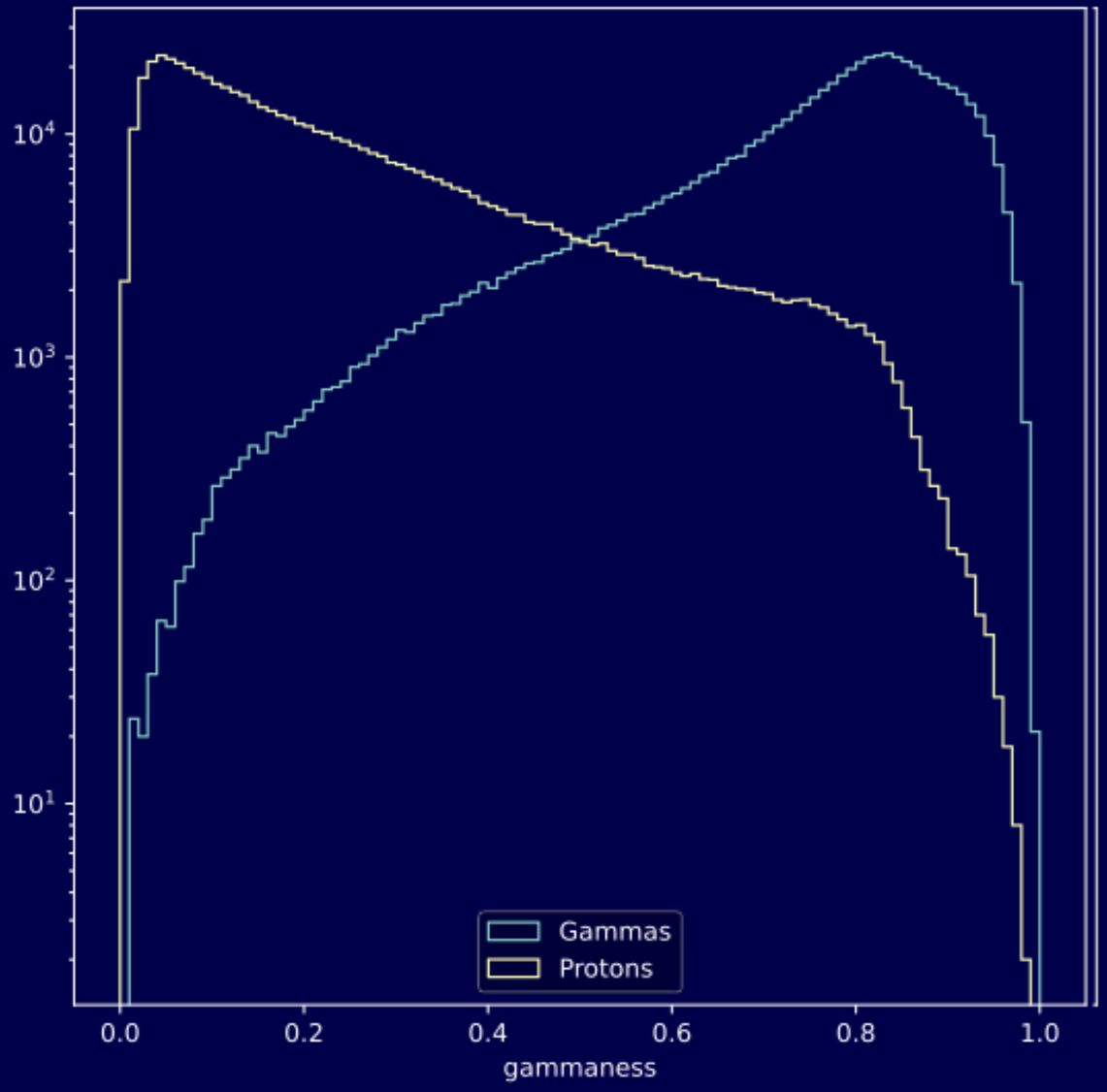
- As a gamma-ray observatory, for almost all science cases we can simplify to binary classification:

  Signal, gamma rays ⇔ background, everything else

- Training needs labelled datasets for signal and background
- Signal is almost always taken from simulated gamma-ray events
- Background can either be simulations as well or taken from observations of “dark spots”
- Hadronic air showers:
  - are much more expensive to simulate
  - interaction models have known issues in describing air showers (muon puzzle)
- Using real observations for background training comes with its own caveats:
  - Model might learn to differentiate simulations from observations
  - Needs available observations matching conditions of observations to be analyzed
- MAGIC usually uses observed background, LST-1 uses simulated protons

Particle Type Classification



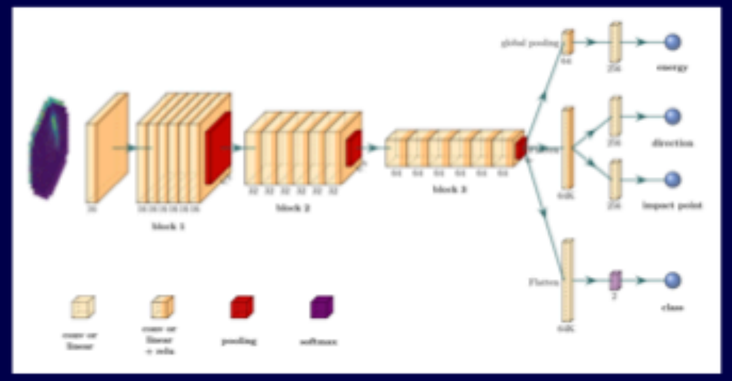


Advanced reconstruction methods

Modern Machine Learning Approaches

- Several groups are working on using Deep Learning using Convolutional Neural Networks for IACT event reconstruction:
  - https://gammalearn.pages.in2p3.fr/pages/ (based on pytorch)
  - https://github.com/ctlearn-project/ctlearn (based on tensorflow)

- Predicting reconstructed properties directly from DL1 images or even DL0 waveforms
- Shown to outperform Image Parameters + Random Forests under controlled conditions (simulations, good quality data)
- Seem to struggle more with data/monte carlo mismatches, changing conditions, in general less robust



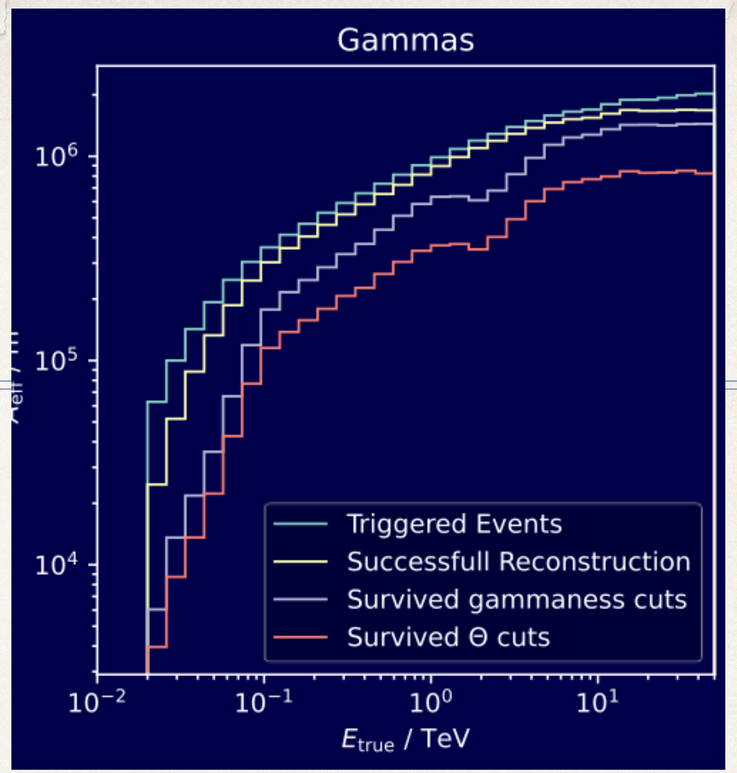
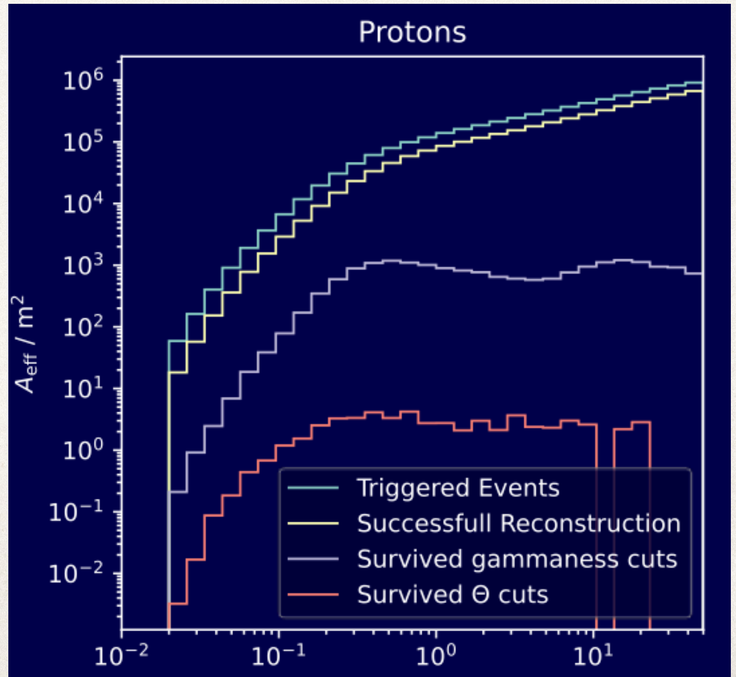
PyTorch

TensorFlow

Likelihood-based Methods: ImPACT + FreePACT

- Based on simulations, it is possible to have a "template" function that predicts the image of a telescope given the true parameters of the air shower (energy, direction, impact, $X_{\max}$) and the telescope properties (pointing, position on the ground).
- These templates can be used to perform a likelihood fit on the actual event data
- Initial method: ImPACT with look-up tables of the mean image built from specialized simulations interpolated at runtime
- FreePACT: a neural network trained to predict the likelihood of pixel values from the true gamma ray properties and the position and direction of the pixel replaces the template library
    - Better description of the likelihood (not just mean)
    - Can be trained on standard simulations

Schwefer, G., Parsons, R., & Hinton, J. (2024). A hybrid approach to event reconstruction for atmospheric Cherenkov Telescopes combining machine learning and likelihood fitting. *Astroparticle Physics*, *163*, 103008. doi:10.1016/j.astropartphys.2024.103008

Beyond event reconstruction

- We now have a list of air shower events with predicted properties.
  - Which parameters to use for my telescope system?
  - Apply gamma hadron cuts
    - Alternatively, use event types
  - Construct the instrument response functions
  - Apply the cuts to real data
  - And now the fun begins - you start to do science!