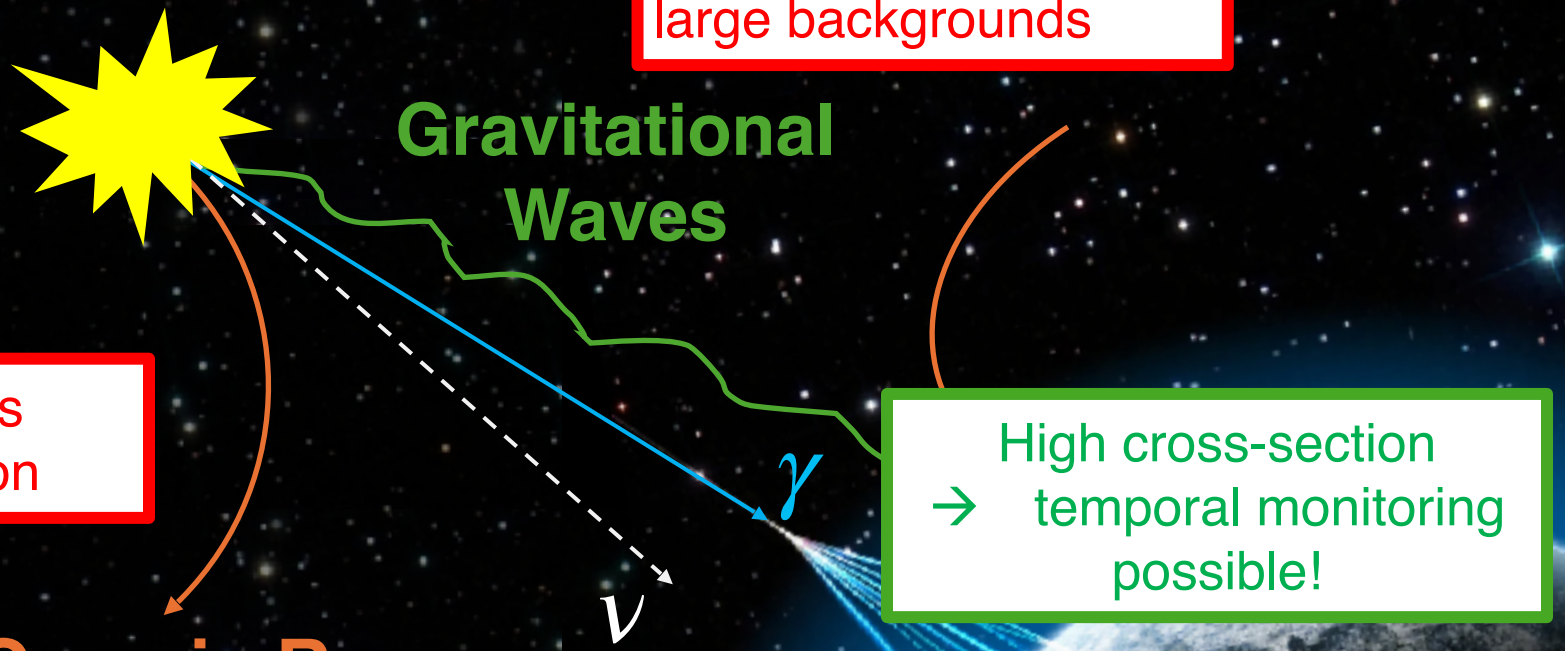


Transformer-Based Analysis for Next-Generation Compact Cherenkov Telescopes

Elli Jobst, Lukas Heinrich, David Paneque, Lea Heckmann,
Nicole Hartman, Jarred Green, Michele Peresano

Multi-Messenger Astronomy

Cosmic Accelerators create unique physical conditions, which are **not-reproducible on Earth**



High-energy particles produce γ-rays

- Hadronic Processes

$$p + p \rightarrow X + \begin{cases} \pi^+ & 1/3 \\ \pi^- & 1/3 \\ \pi^0 & 1/3 \end{cases}$$

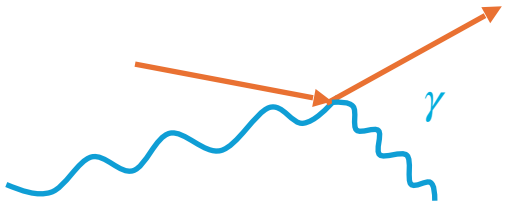
$$p + \gamma \rightarrow \Delta^+ \rightarrow \begin{cases} p + \pi^0 & 1/3 \\ n + \pi^+ & 2/3 \end{cases}$$

$$\pi^0 \rightarrow 2\gamma$$

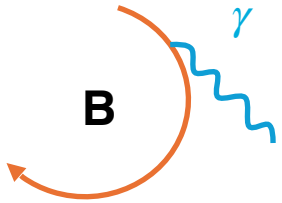
$$\pi^{\pm} \rightarrow \mu^{\pm} + \nu/\bar{\nu}$$

- Leptonic Processes

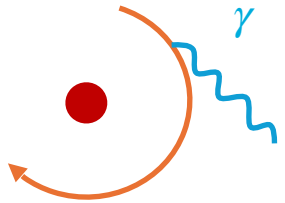
Inverse Compton Scattering



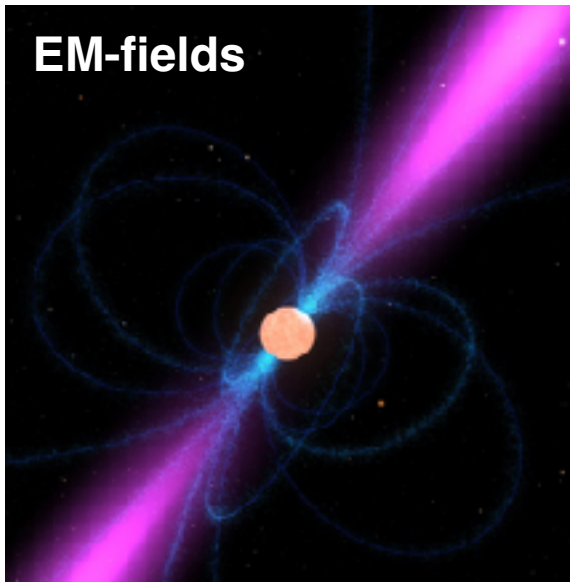
Synchrotron Radiation



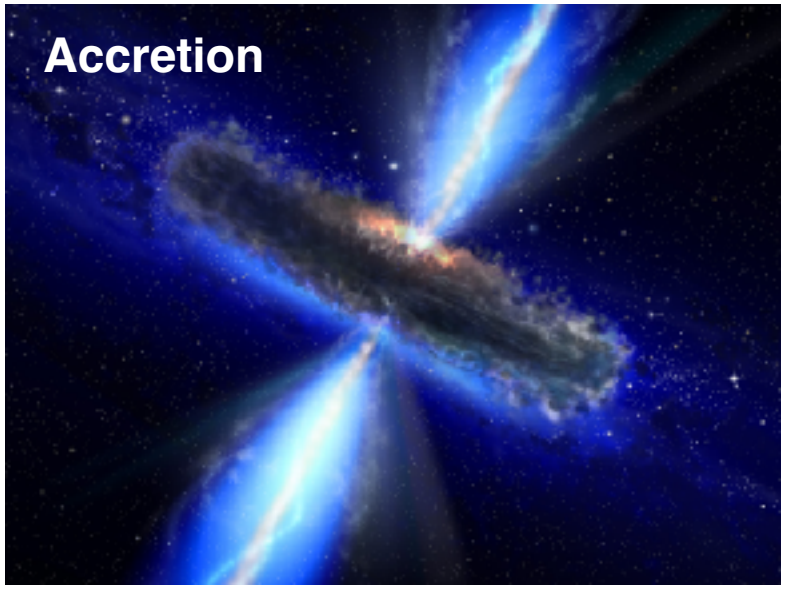
Bremsstrahlung



Studying the high-energy Universe with γ-rays







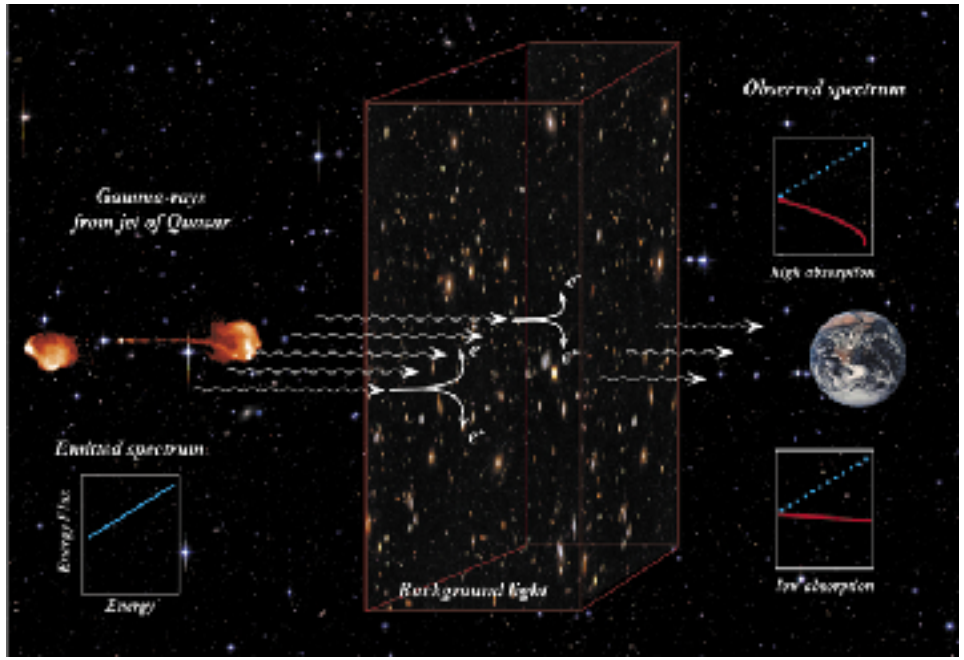
→ Use high-energy γ–rays to probe highly energetic objects

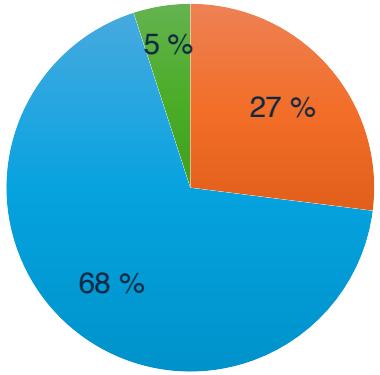
source direction

lower limit to energy

time evolution

Fundamental Physics with γ-ray Astronomy





- Dark Energy
- Dark Matter
- Ordinary Matter

- **Fundamental physics related to the propagation of the γ-rays**
  - Extragalactic Background Light (EBL)
  - Intergalactic Magnetic Fields (IGMF)
  - Lorentz Invariance Violation (LIV)
  - Search for Axion Like Particles (ALP)

- **Indirect searches of DM**

$$X + X \rightarrow \ ? \rightarrow \gamma\gamma$$
$$X + X \rightarrow SM \rightarrow \gamma$$

Ground-based γ-ray Astronomy

17th century
first optical telescopes

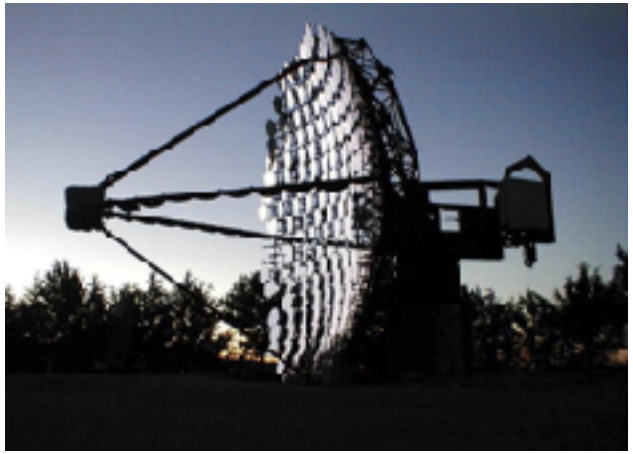
1930s first radio telescopes
Big push in the 1950s

1960s first X-ray telescopes
Big push in the 1970s and 1980s

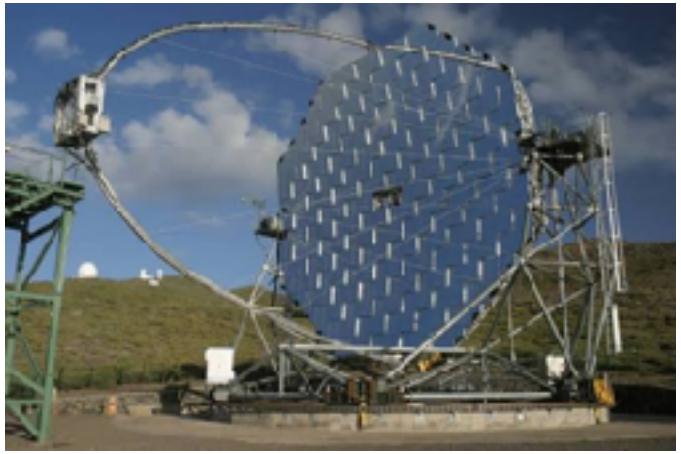
1970s
First space-based gamma-ray telescopes

Late 1980s
First significant detection by ground-based Cherenkov telescopes

Whipple, 1989 Crab Detection



MAGIC, since 2004

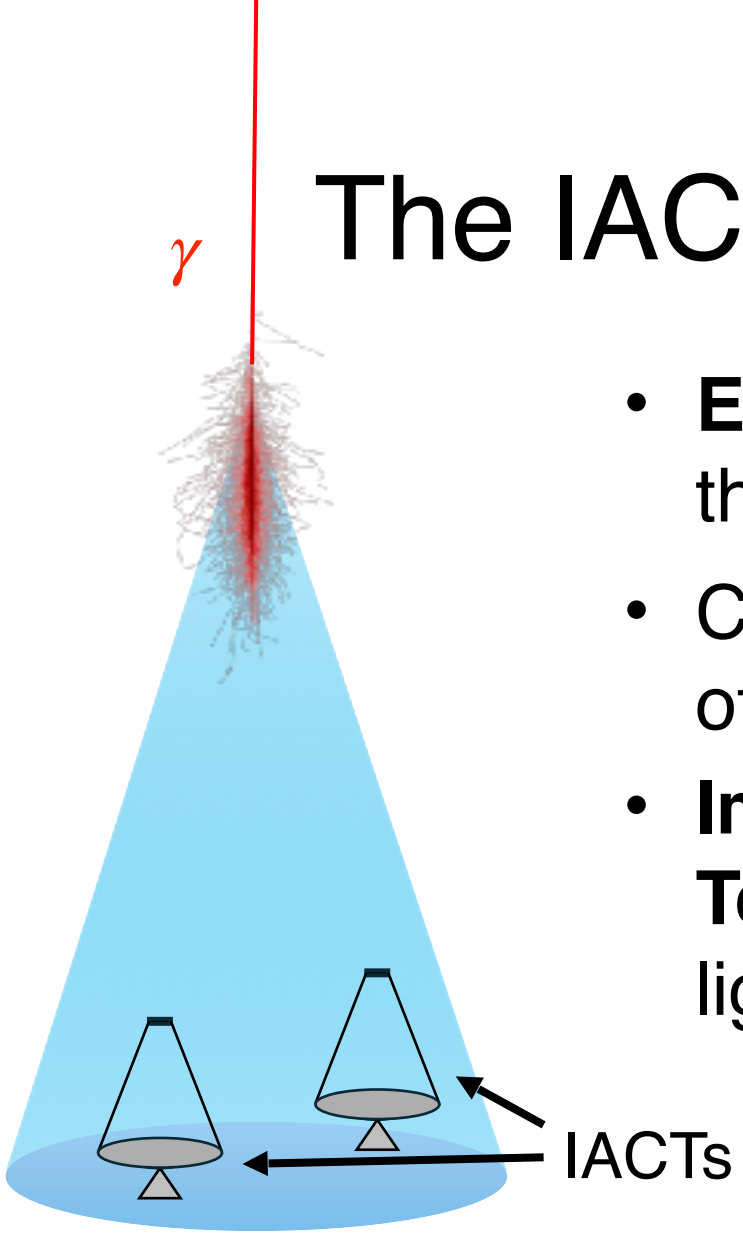


LST-1, since 2018



Novel discipline → lots of room for improvement & potential discovery

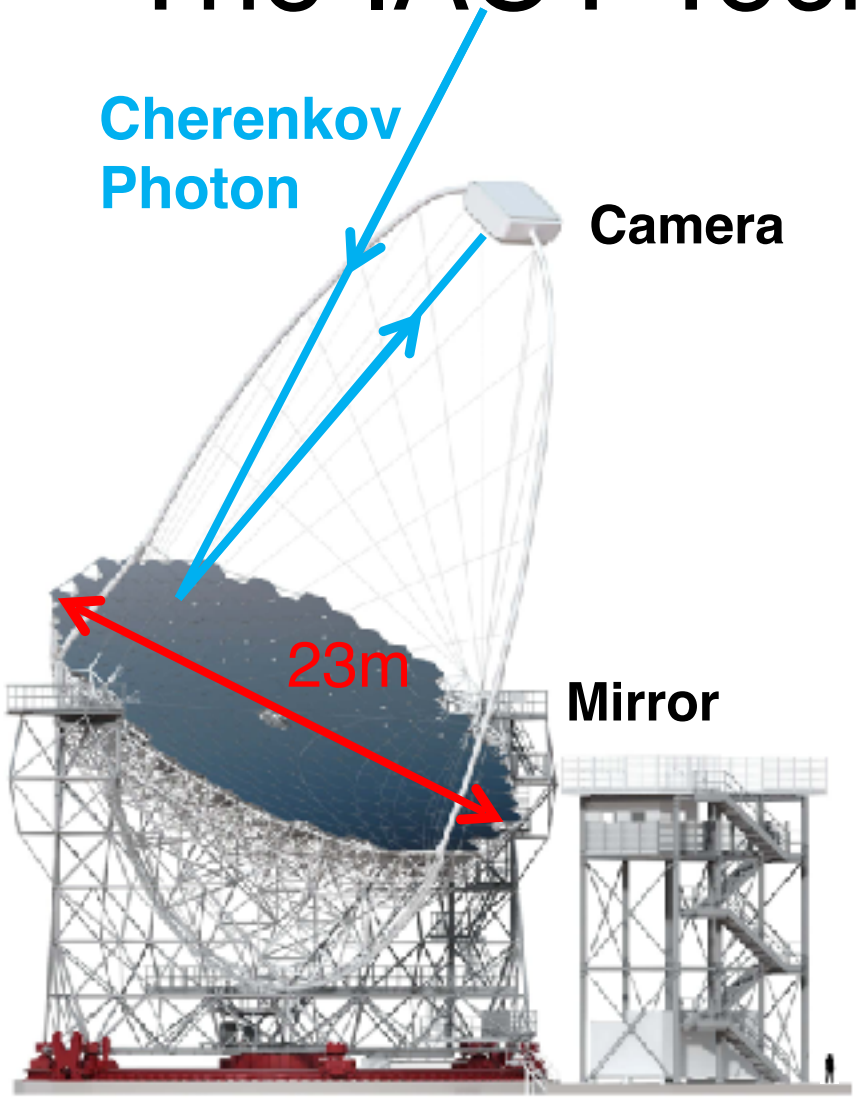
The IACT Technique



- **Extensive Air Showers** are produced when the γ-rays interact in the atmosphere
- Charged particles faster than the local speed of light emit **Cherenkov Radiation**
- **Imaging Atmospheric Cherenkov Telescopes (IACTs)** observe the Cherenkov light to infer the original particle‘s properties

→ **Use atmosphere as a calorimeter**

The IACT Technique



From the CTAO Website, 2026



CTAO LST subarray, inaugurated two weeks ago!

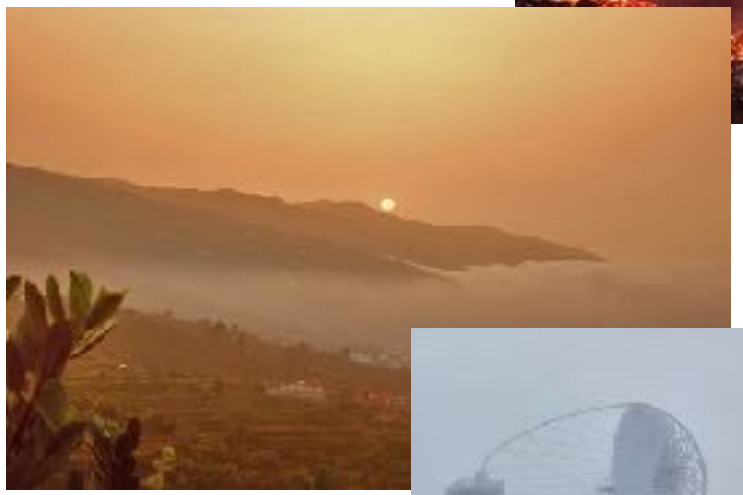
Large mirror dishes

Multiple telescopes observing together

Challenges of current IACTs

- High construction and operation cost
- Degradation of hardware due to harsh weather
- Higher maintenance required
- Inaccurate description of the telescope in simulations







The Small Virtual Telescope (SVT)

Single, compact telescope **($d_{mirror} \approx$ 5m)**

Inside a dome

- safe from the elements
- easy to maintain

Cheap and easy to construct and operate

- low construction cost

$$\text{cost} \propto d_{mirror}^{2.6}$$

- no on-site shifters required

Issues:

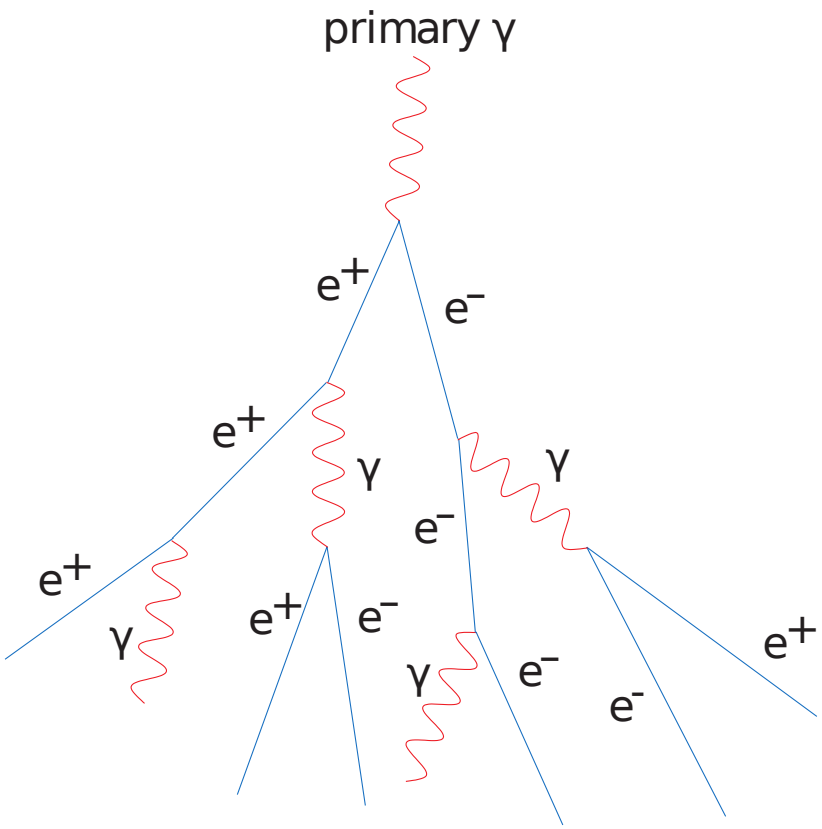
- data analysis is difficult at low energies
- worse telescope performance with a single small telescope

„Can we develop a **different data analysis** with new **deep learning techniques** to improve the telescope performance enough to make such a **compact telescope** feasible?“

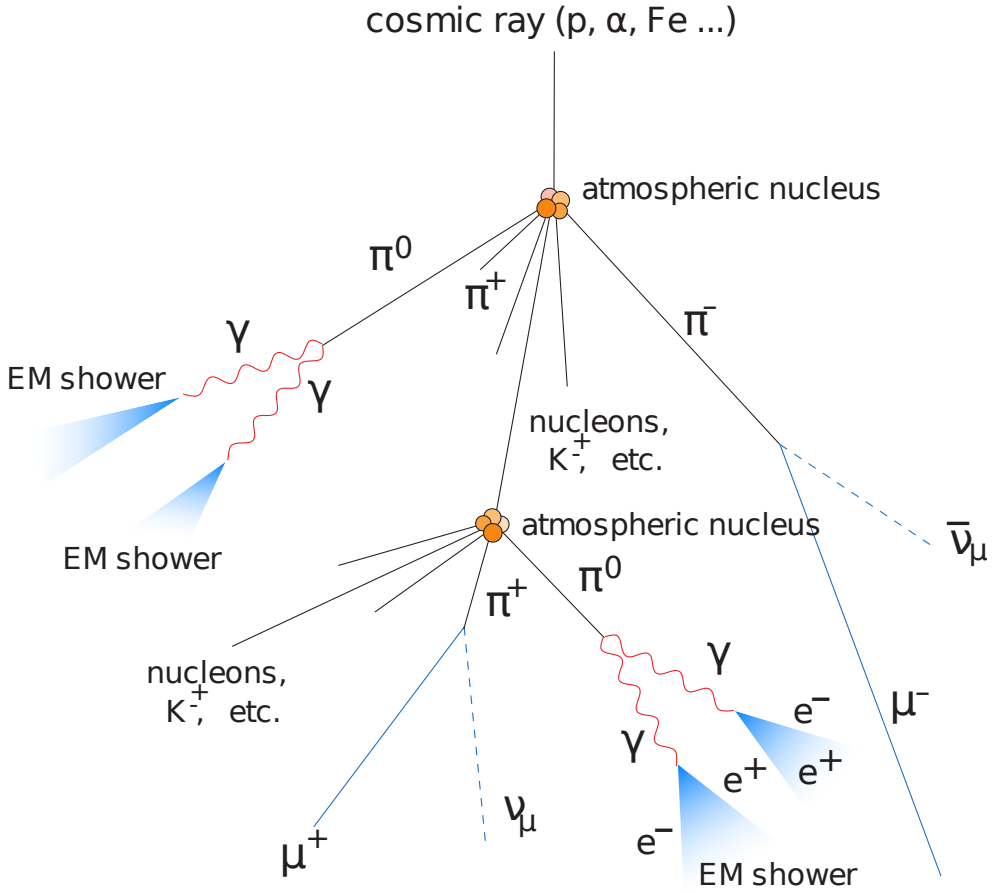
The Dataset

Extensive Air Showers

Electromagnetic (EM) Shower



Hadronic Shower

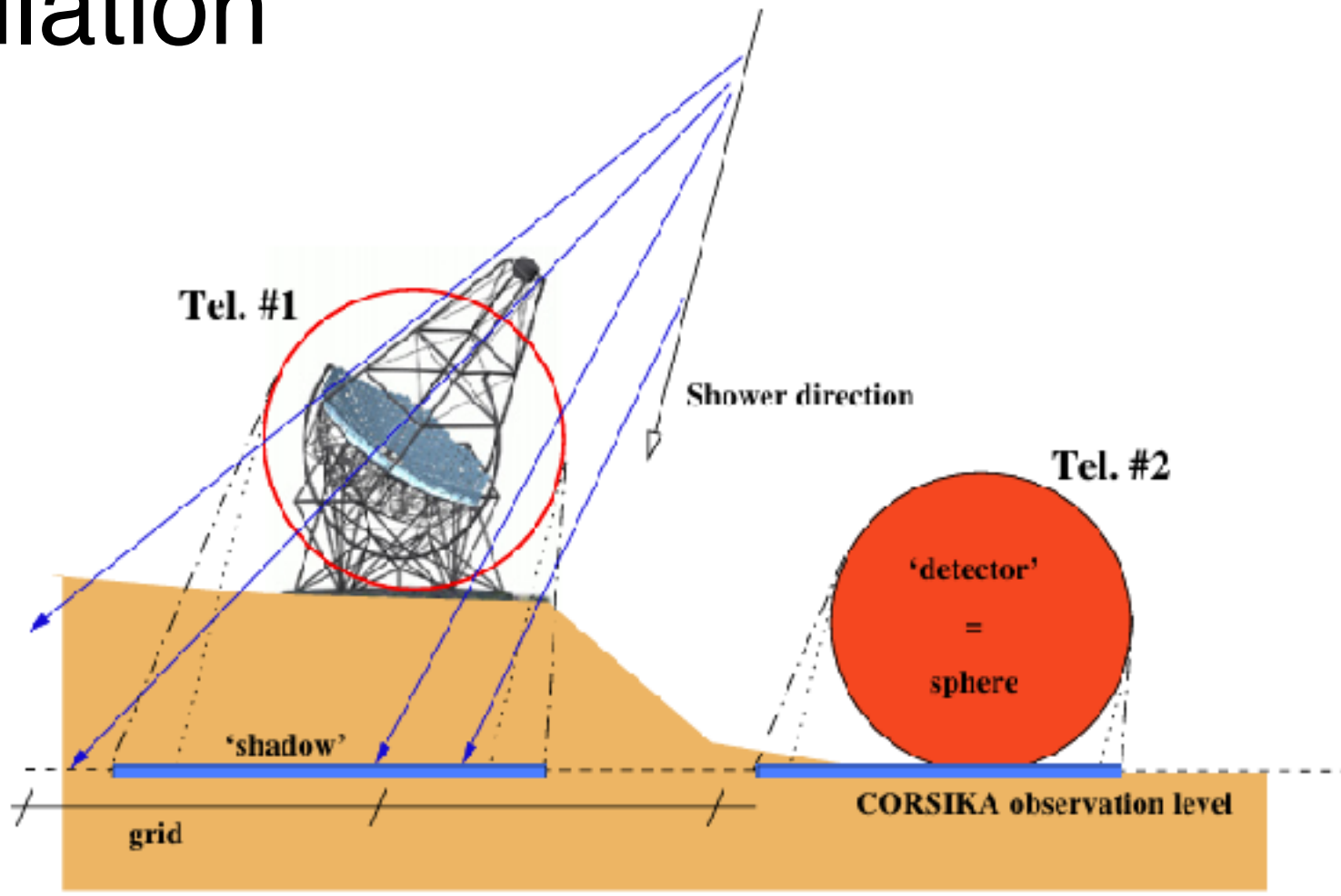


From A. Hahn, PhD Thesis (2025)

Air Shower Simulation

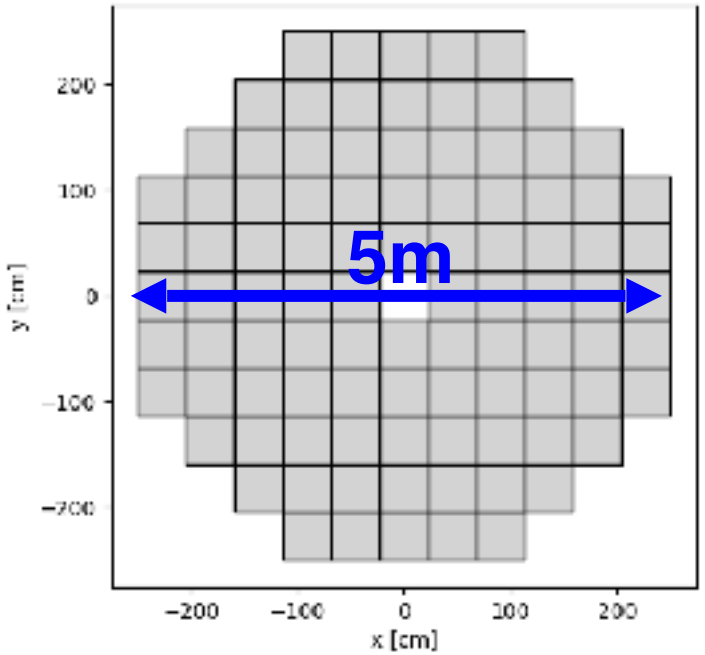
CORSIKA

- γ–rays and protons
- flat (in log-space) energy spectrum for training data
- realistic point-source spectrum (Crab Nebula) for testing data

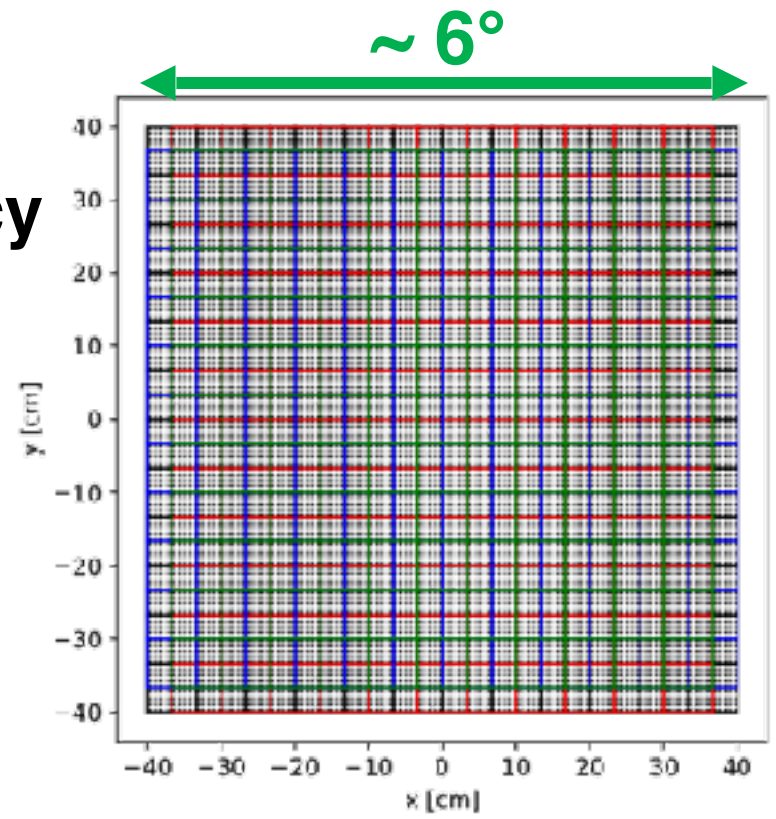


From K. Bernlöhr (2008)

Telescope Specifications



- 5m **Dish Diameter**
- 7.5m focal length
- 100% Reflectivity
- **100% Quantum Efficiency**



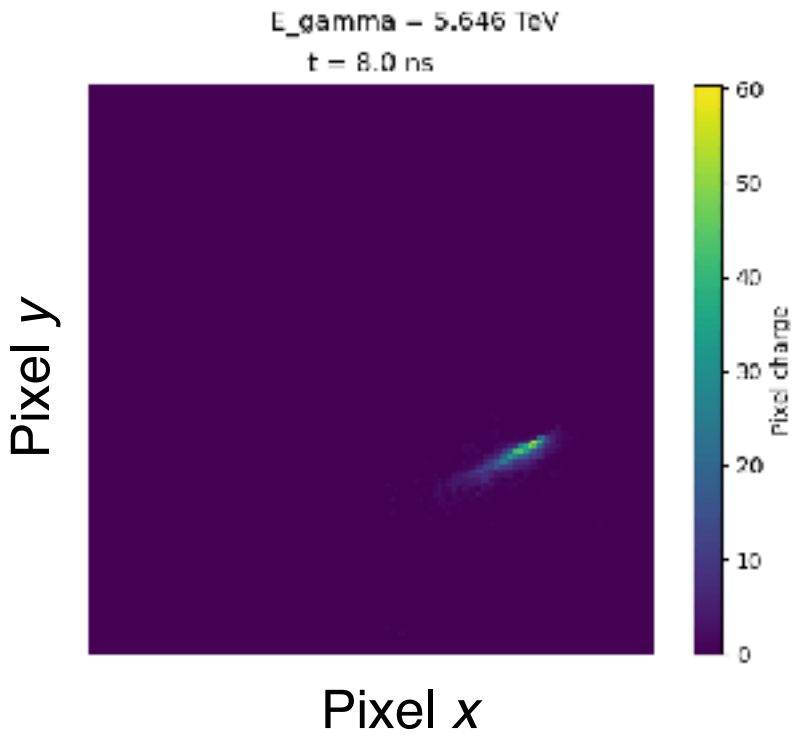
- large **Field of View** of ~ 6° and high-resolution camera with 120x120 pixels
- realistic Night Sky Background light

→ **Goal: keep everything ideal to explore what is possible with deep learning**

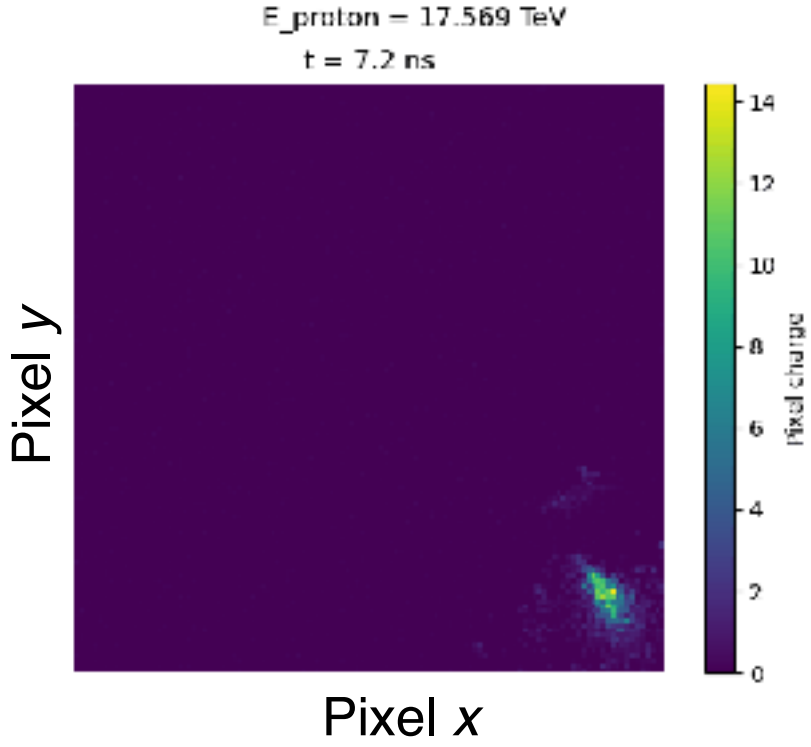
Raw data

We obtain a 128-frame movie for each event

γ–ray



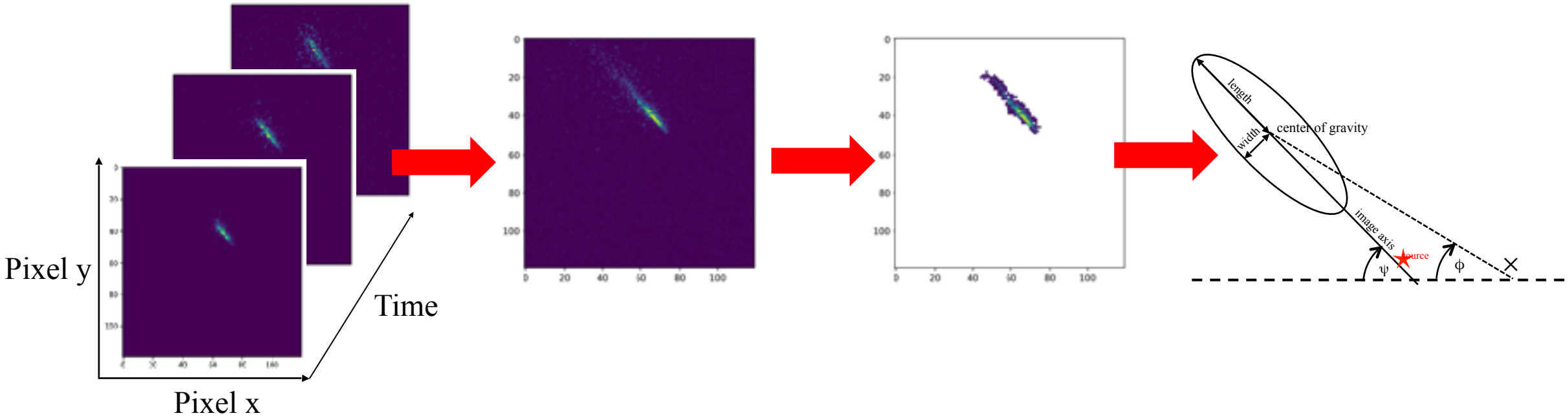
proton



The Standard Analysis

The Standard Event Analysis

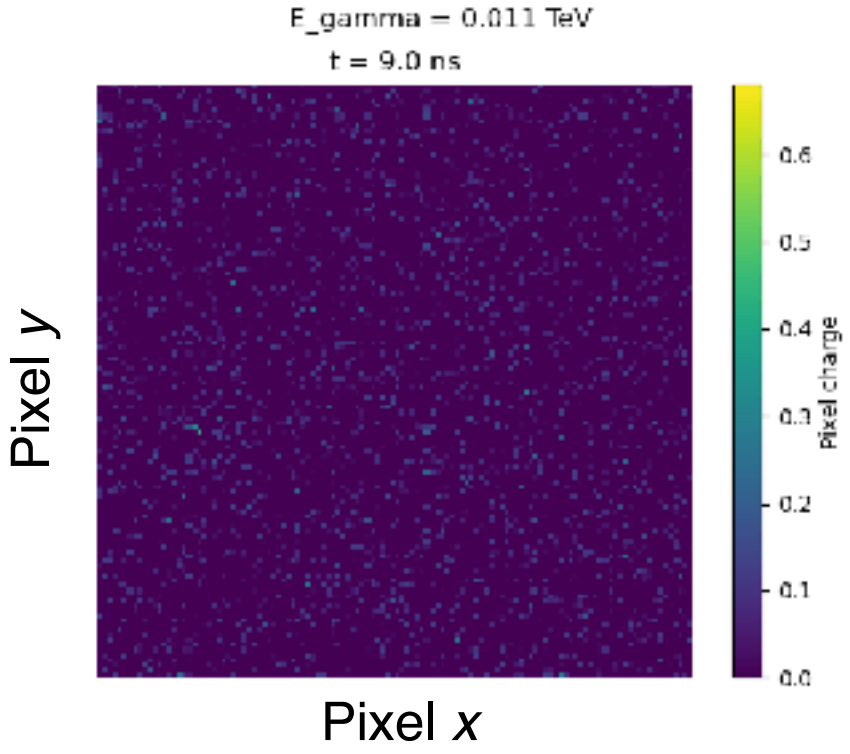
Raw Data	Calibrated Image	Cleaned Image	Image Parameters
(128x120x120) = $O(10^6)$			$O(10)$



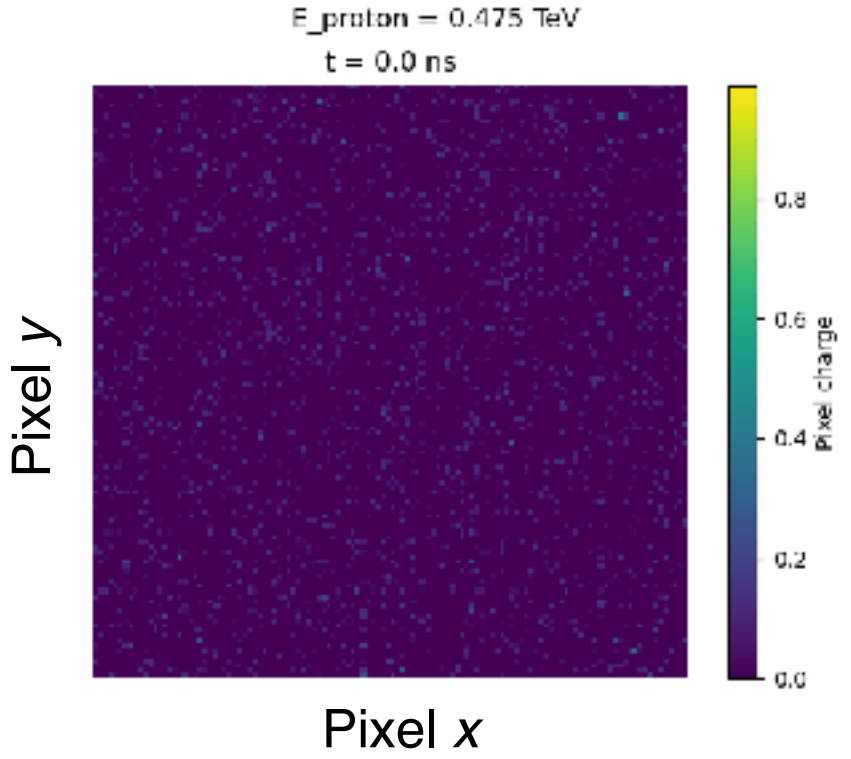
→Huge loss of information contained in the original movie!

At low energies…

γ–ray



proton



Telescope Performance

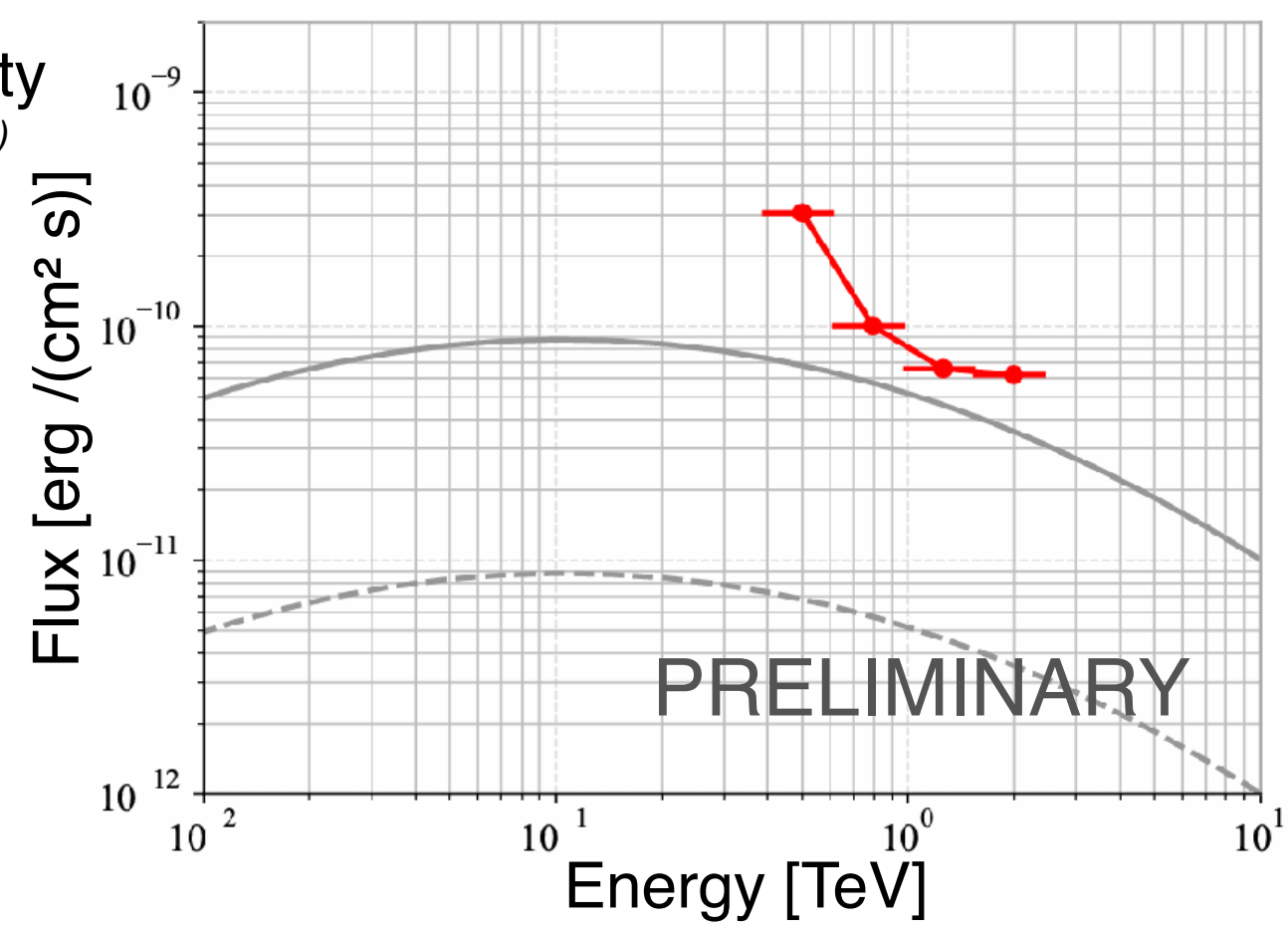
Telescope Performance - Sensitivity

How many photons would we need to observe from this source in order to get a 5σ detection?

1h differential sensitivity

after Formula (17) from Li & Ma (1983)

Standard Analysis for the SVT



100% Crab

10% Crab

Deep Learning Analysis

What type of architecture?

Convolutional Neural Networks

- suited for image processing
- rigid architecture

Graph Neural Networks

- great flexibility
- difficult to scale up

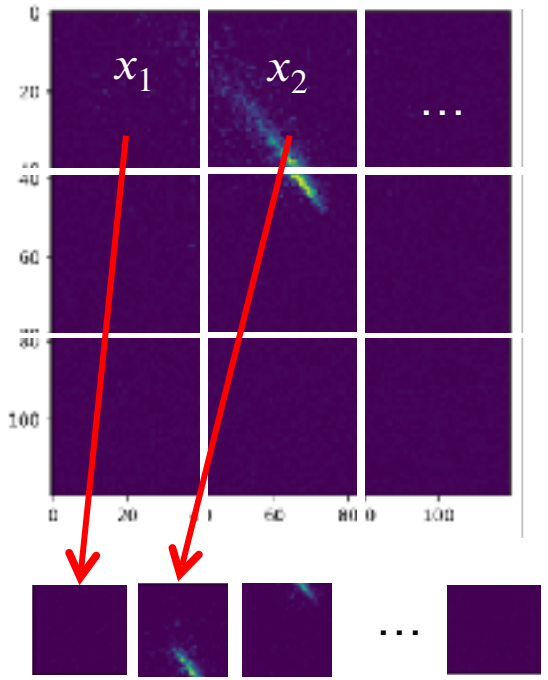
Recurrent Neural Networks

- suited for sequence processing
- difficult to scale up
- information loss for long sequences

→ **Transformers combine these features without the disadvantages**

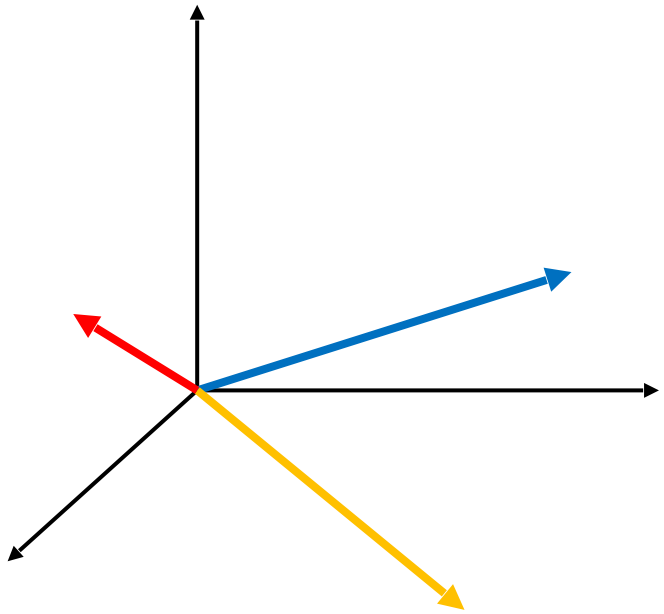
Vaswani, Ashish et al. (2017). "Attention Is All You Need"

Vision Transformer



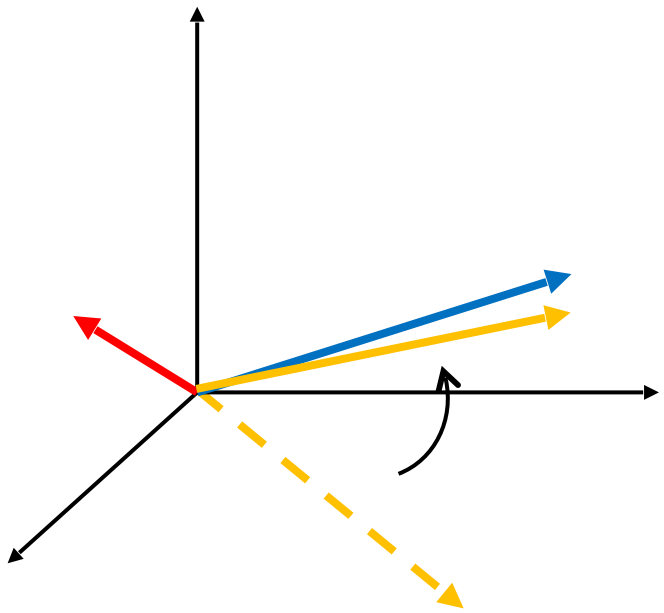
- Sequence of image patches
- Converted into a vector in the ***embedding*** space
- Positional encoding is added

Vision Transformer



- each patch is a vector in the embedding space
- each dimension is a concept

Vision Transformer

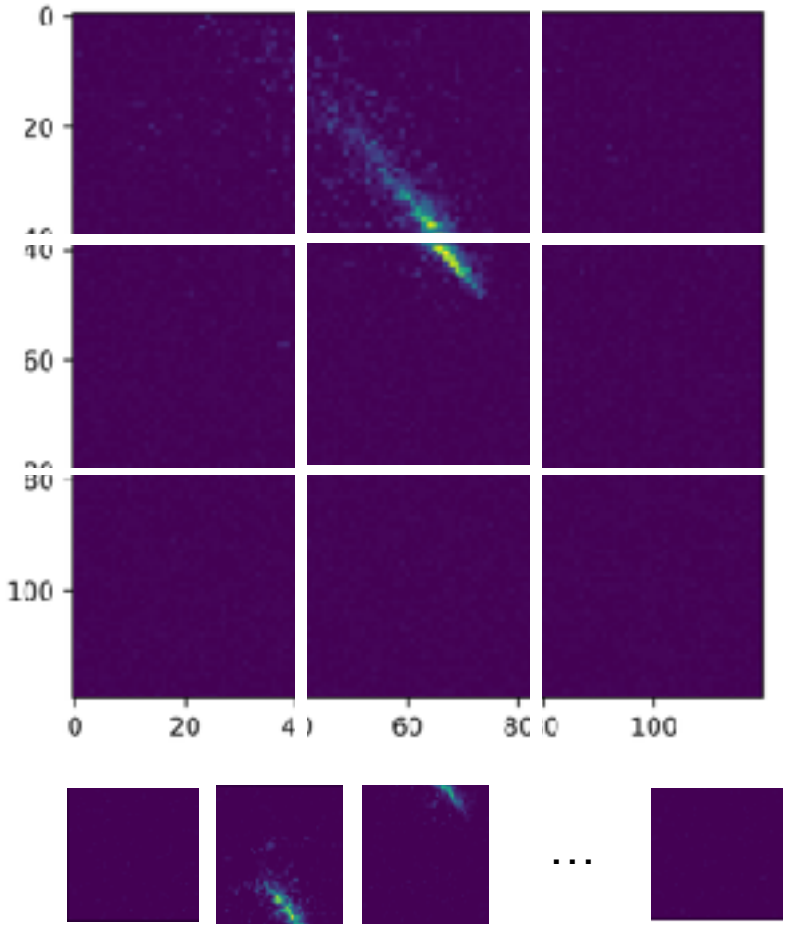


- each patch is a vector in the embedding space
- each dimension is a concept

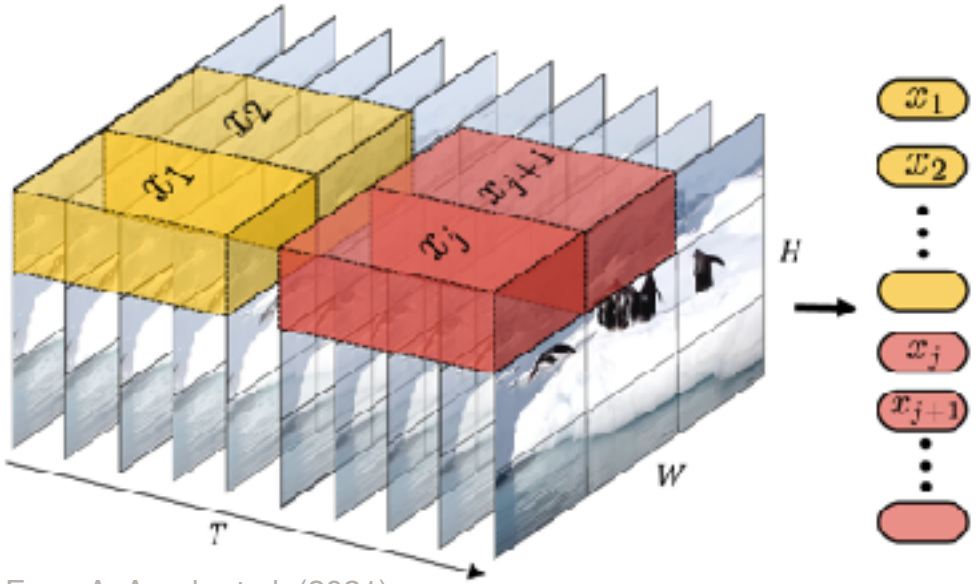
- **Self-Attention Mechanism**: vectors are updated based on context

→ **Transformer can learn which parts of the input are important**

Video Vision Transformer



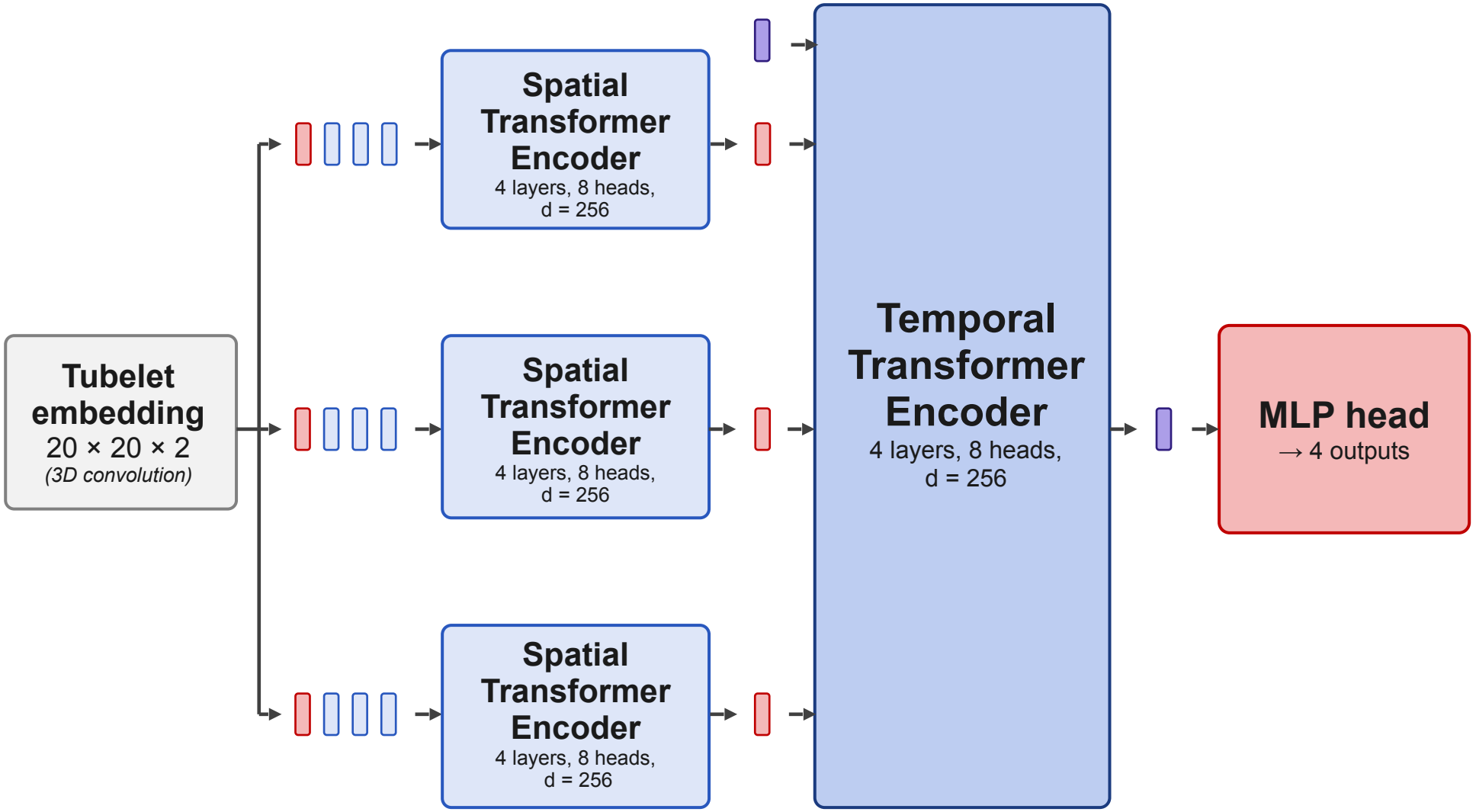
EXTEND INTO 3D



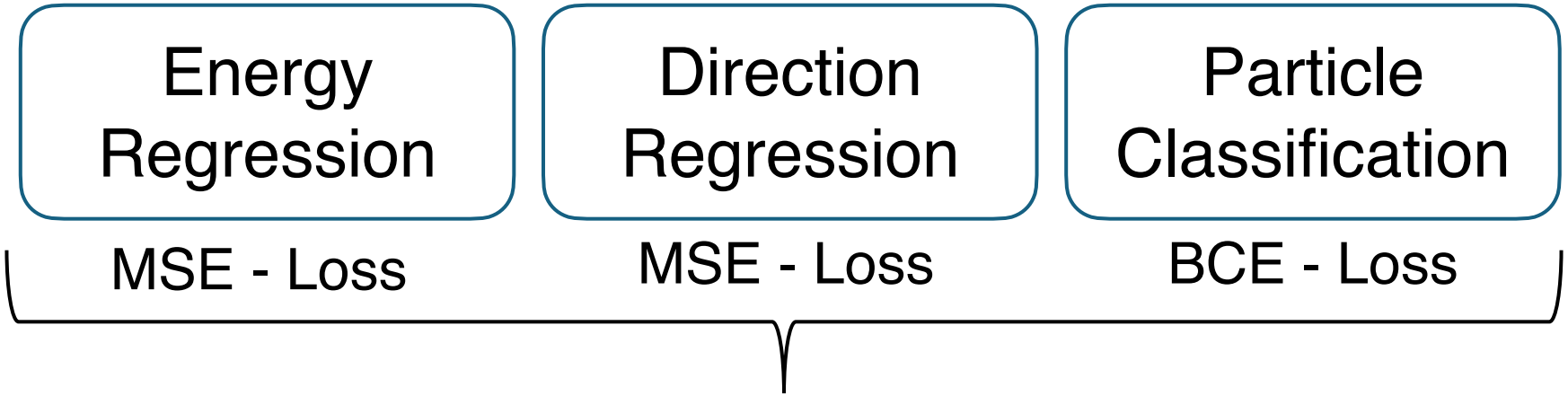
From A. Arnab et al. (2021)

Tubelet embedding
20 × 20 × 2
(3D convolution)

ViViT – Factorized Encoder Model



Composite Model



→ Now: predict all labels at the same time!

- utilizes correlations in the data
- needs a combined regression-classification loss

Composite Model – Loss Function

- Regression loss for a sample i:

$$\mathscr{L}_{r,i} = \frac{\left(\hat{E}_i - E_i\right)^2 + \left(\hat{x}_i - x_i\right)^2 + \left(\hat{y}_i - y_i\right)^2}{3}$$

- Classification loss for a batch B with N samples:

$$\mathscr{L}_C(\vec{C}, \hat{\vec{C}}) = -\frac{1}{N}\sum_{i=1}^{N}\left[C_i \log\left(\hat{C}_i\right) + \left(1 - C_i\right) \cdot \log\left(1 - \hat{C}_i\right)\right]$$

All losses are divided by their own value to get unit gradients!

Composite Model – Loss Function

Total loss for *B*:

$$\mathscr{L}_B = \frac{\vec{\mathscr{L}}_r \cdot \vec{m} + \mathscr{L}_C}{2}$$

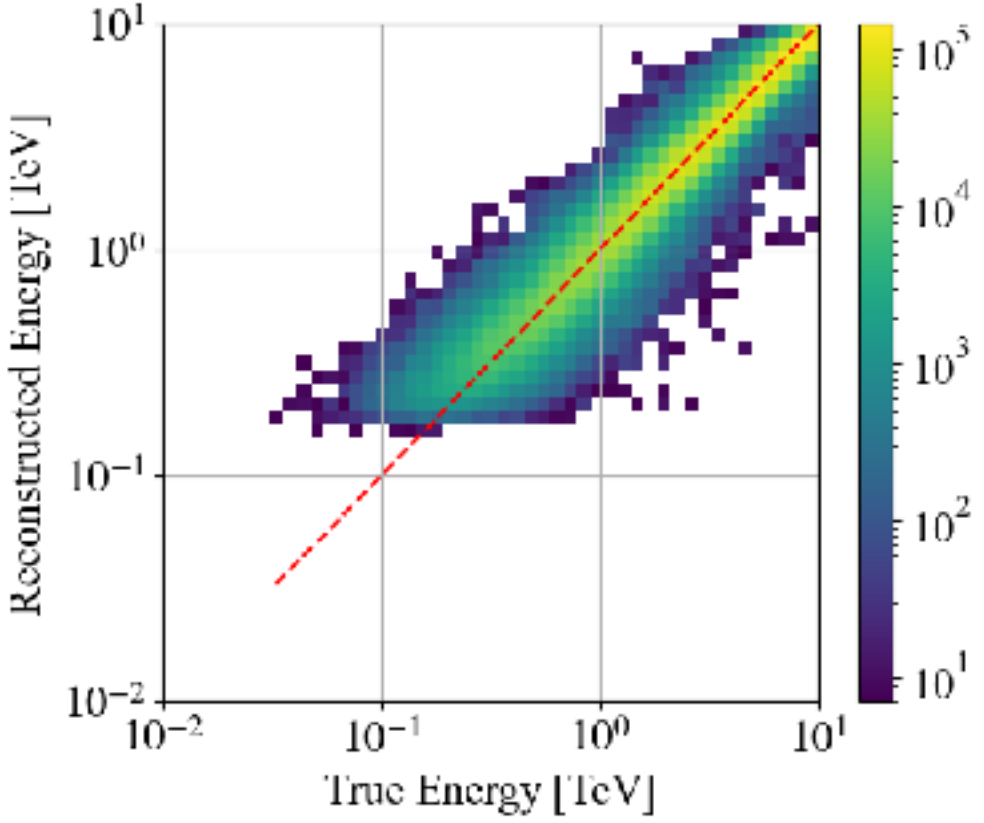
with $m_i = \left(\sum_{j \in B} C_j \right)^{-1}$ for γ–rays , else $m_i = 0$

Energy and direction regression is only trained on γ–ray events, classification on both γ–ray and protons!

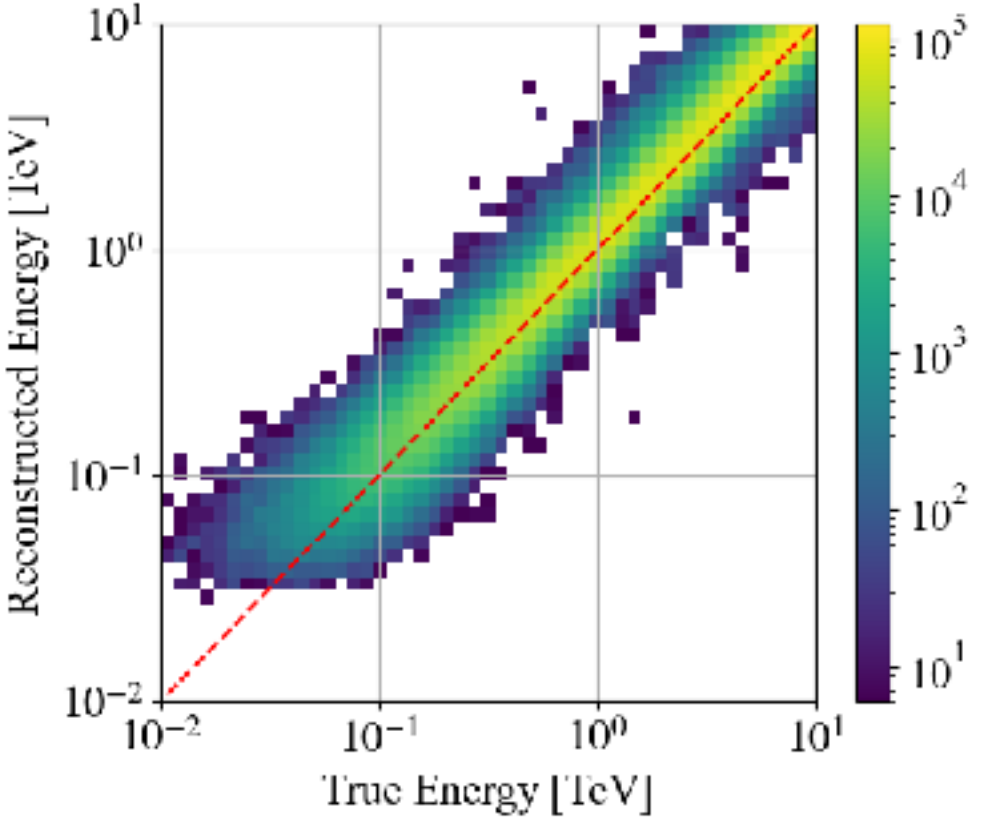
Results

Energy Reconstruction

SVT Standard Analysis

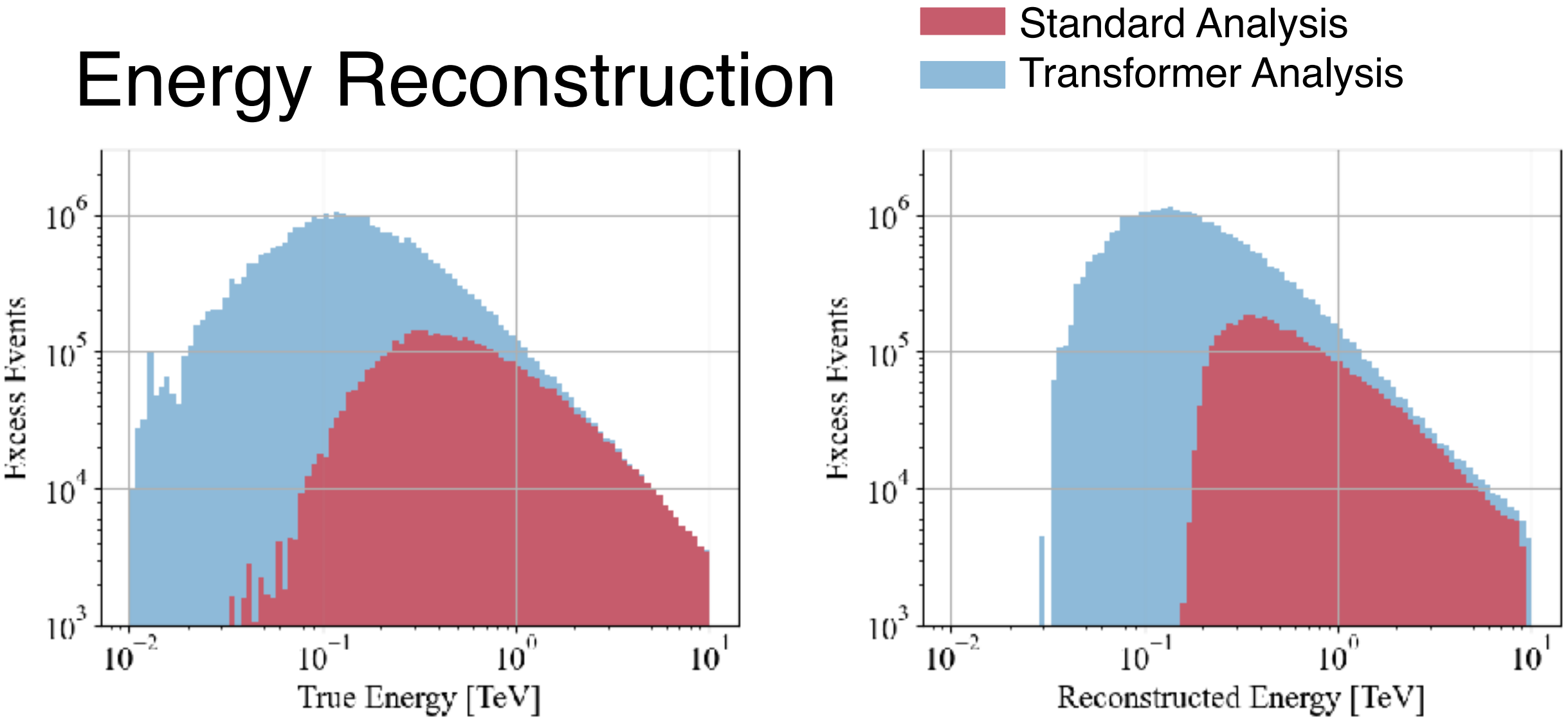


SVT Transformer Analysis



→ **We are able to reconstruct events with smaller photon content!**

Energy Reconstruction



→ **We can reconstruct events with energies about 3 times lower!**

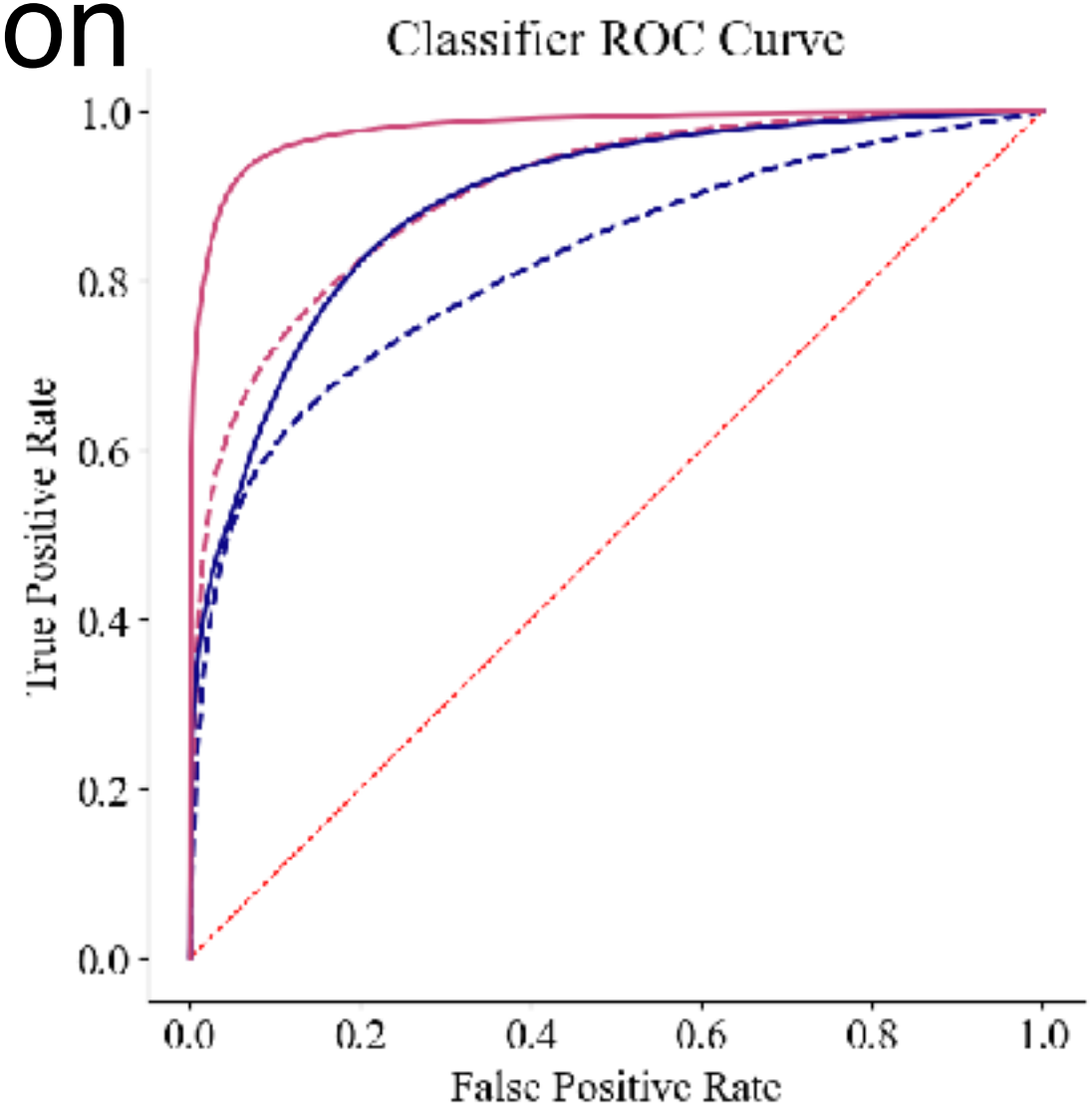
Particle Separation

320GeV – 1TeV

100GeV – 320GeV

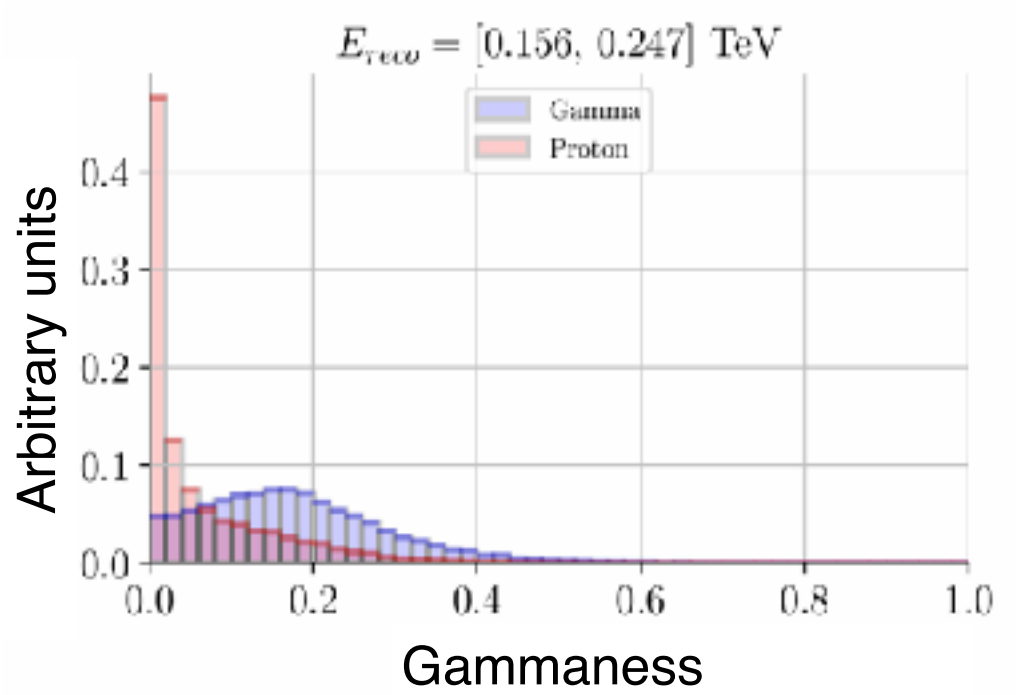
Standard Analysis

Transformer Analysis

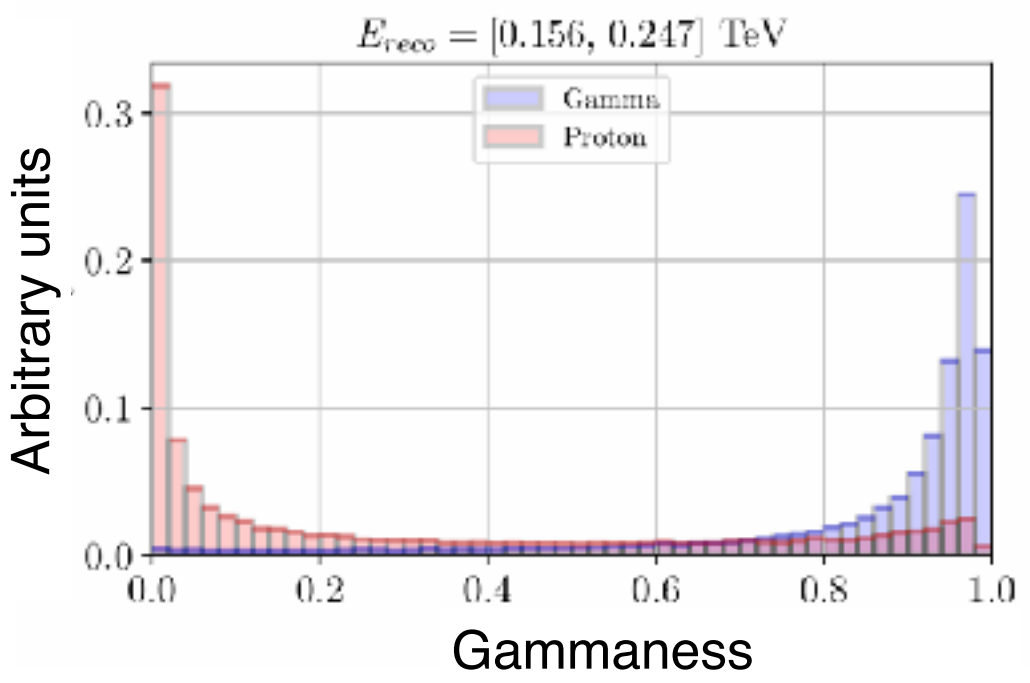


Particle Separation

SVT Standard Analysis



SVT Transformer Analysis



→ **The transformer analysis is able to separate the distributions much better, even at low energies!**

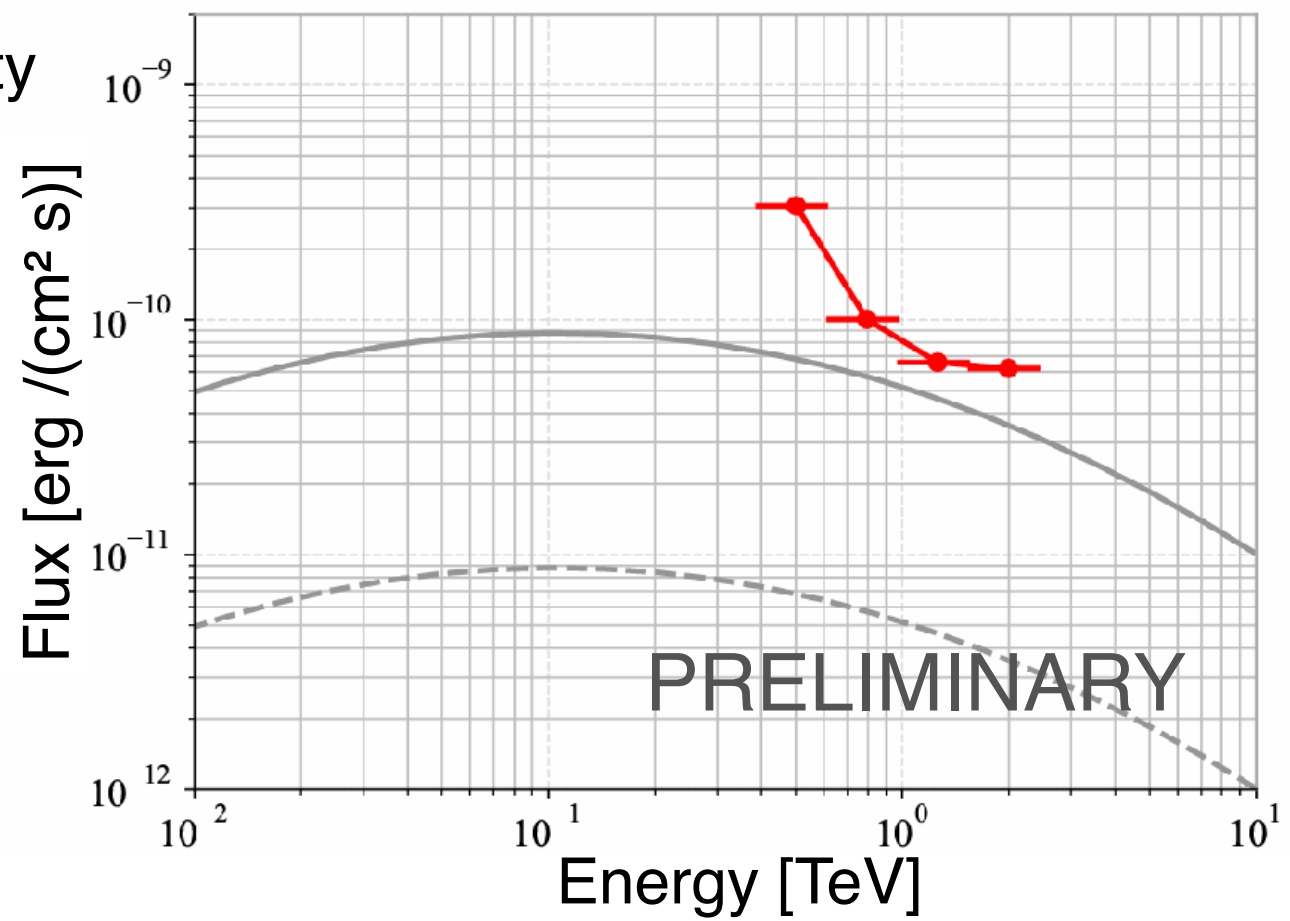
Telescope Performance - Sensitivity

How many photons would we need to observe from this source in order to get a 5σ detection?

1h differential sensitivity

after Formula (17) from Li & Ma (1983)

Standard Analysis for the SVT



100% Crab

10% Crab

Telescope Performance - Sensitivity

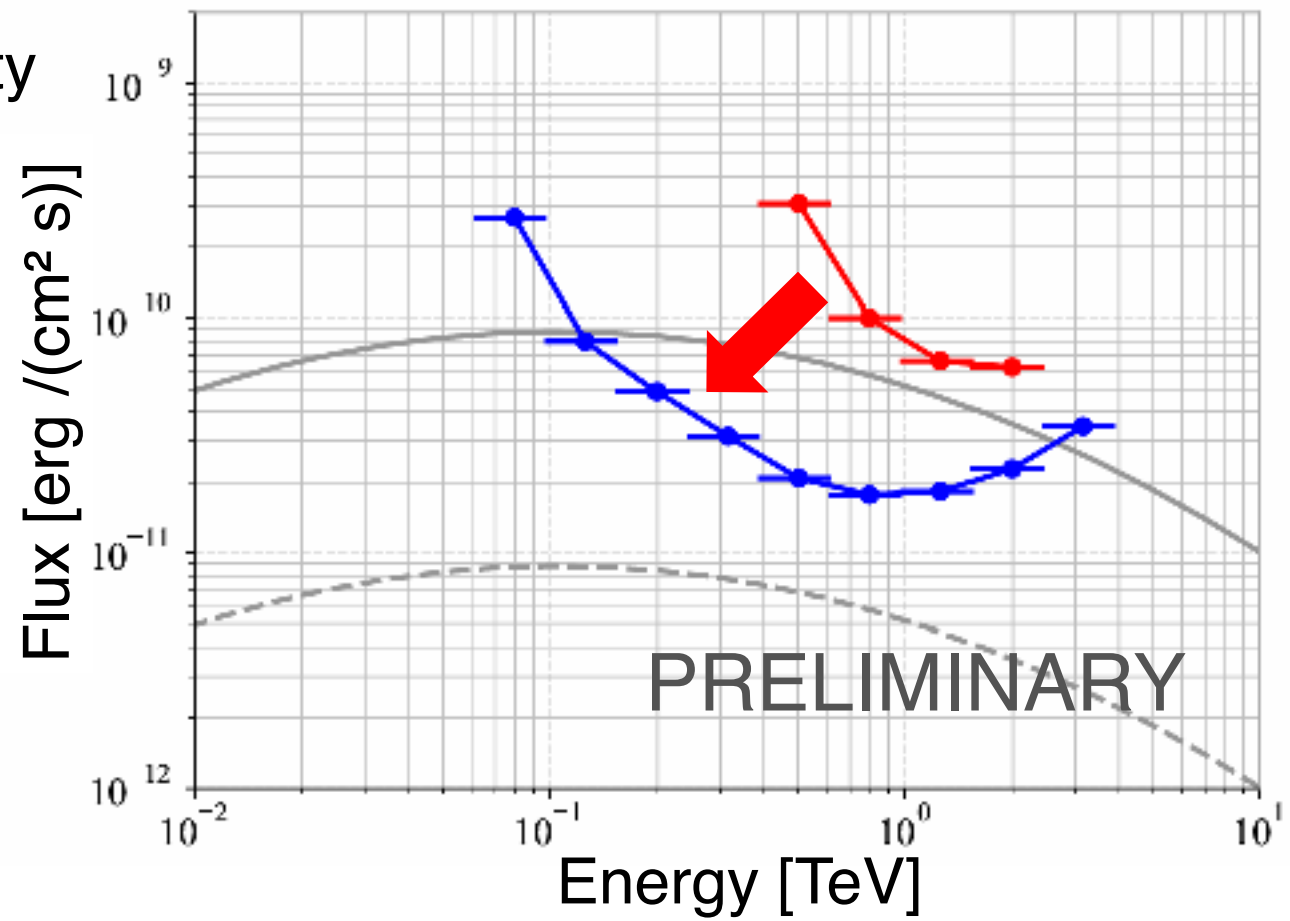
How many photons would we need to observe from this source in order to get a 5σ detection?

1h differential sensitivity

after Formula (17) from Li & Ma (1983)

Standard Analysis for the SVT

Transformer Analysis for the SVT

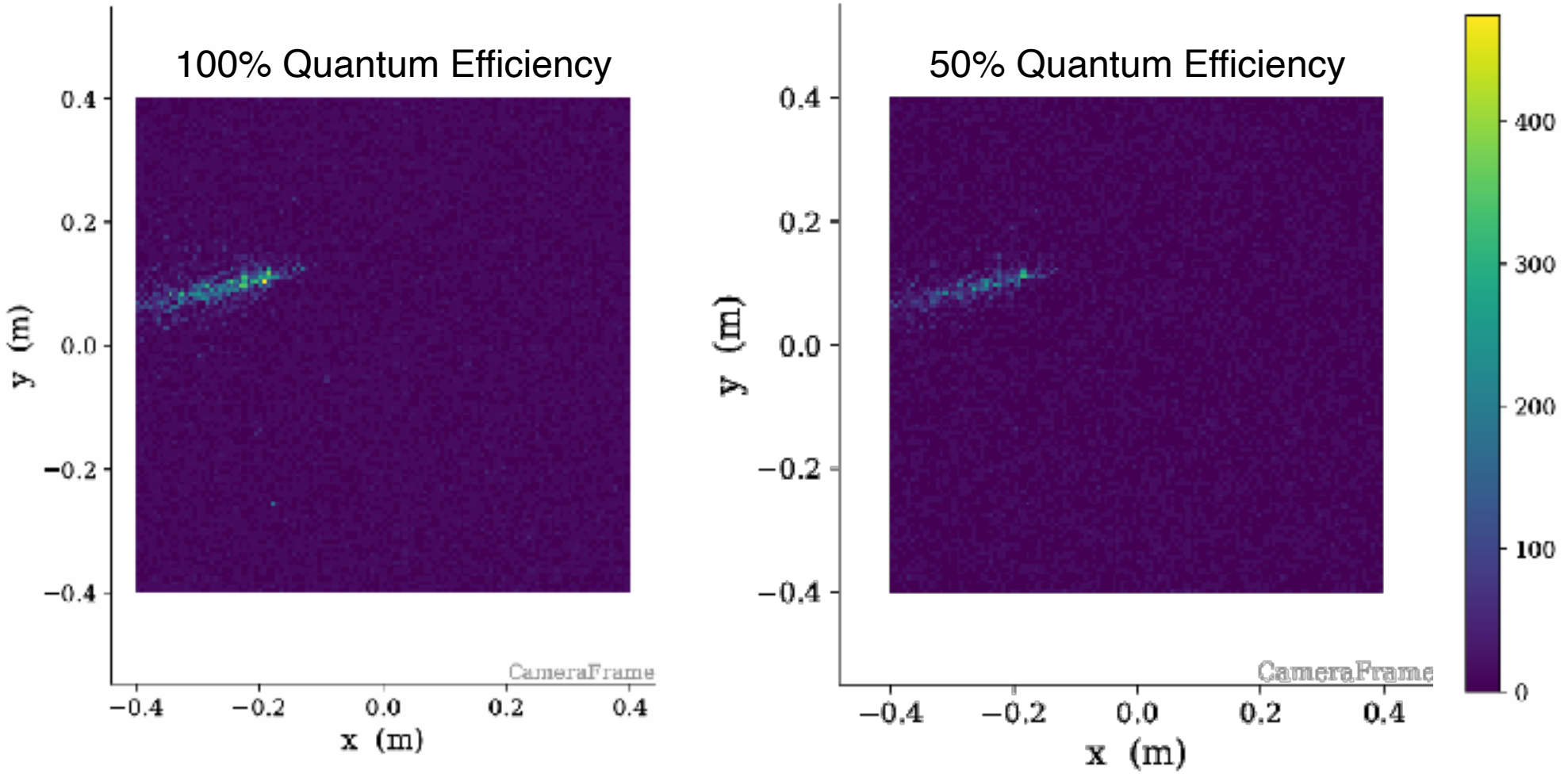


100% Crab

10% Crab

Design Optimization

Example: Quantum Efficiency

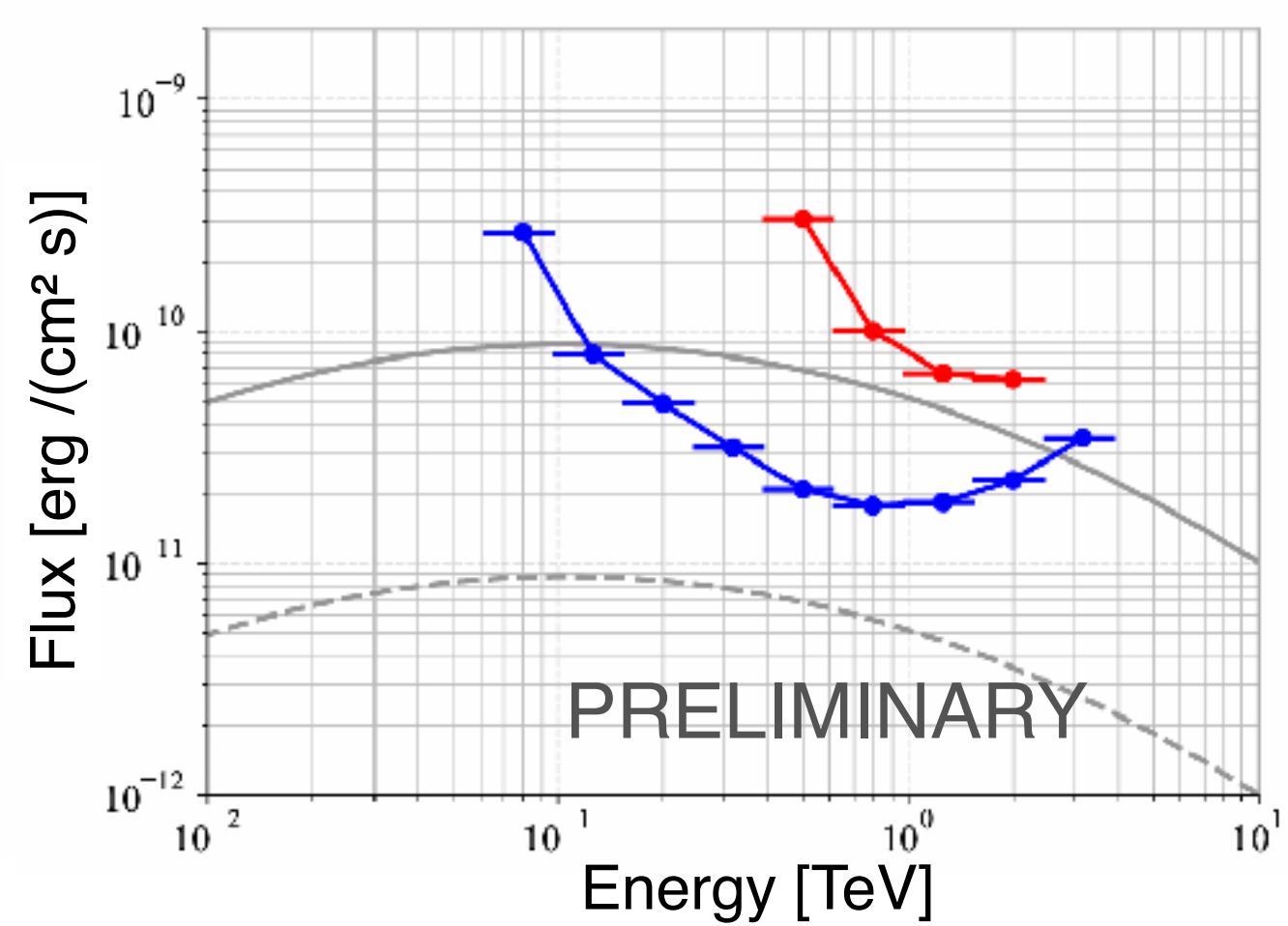


Example: Quantum Efficiency

100% Quantum Efficiency

Standard Analysis

Transformer Analysis



100% Crab

10% Crab

Example: Quantum Efficiency

100% Quantum Efficiency

Standard Analysis

Transformer Analysis

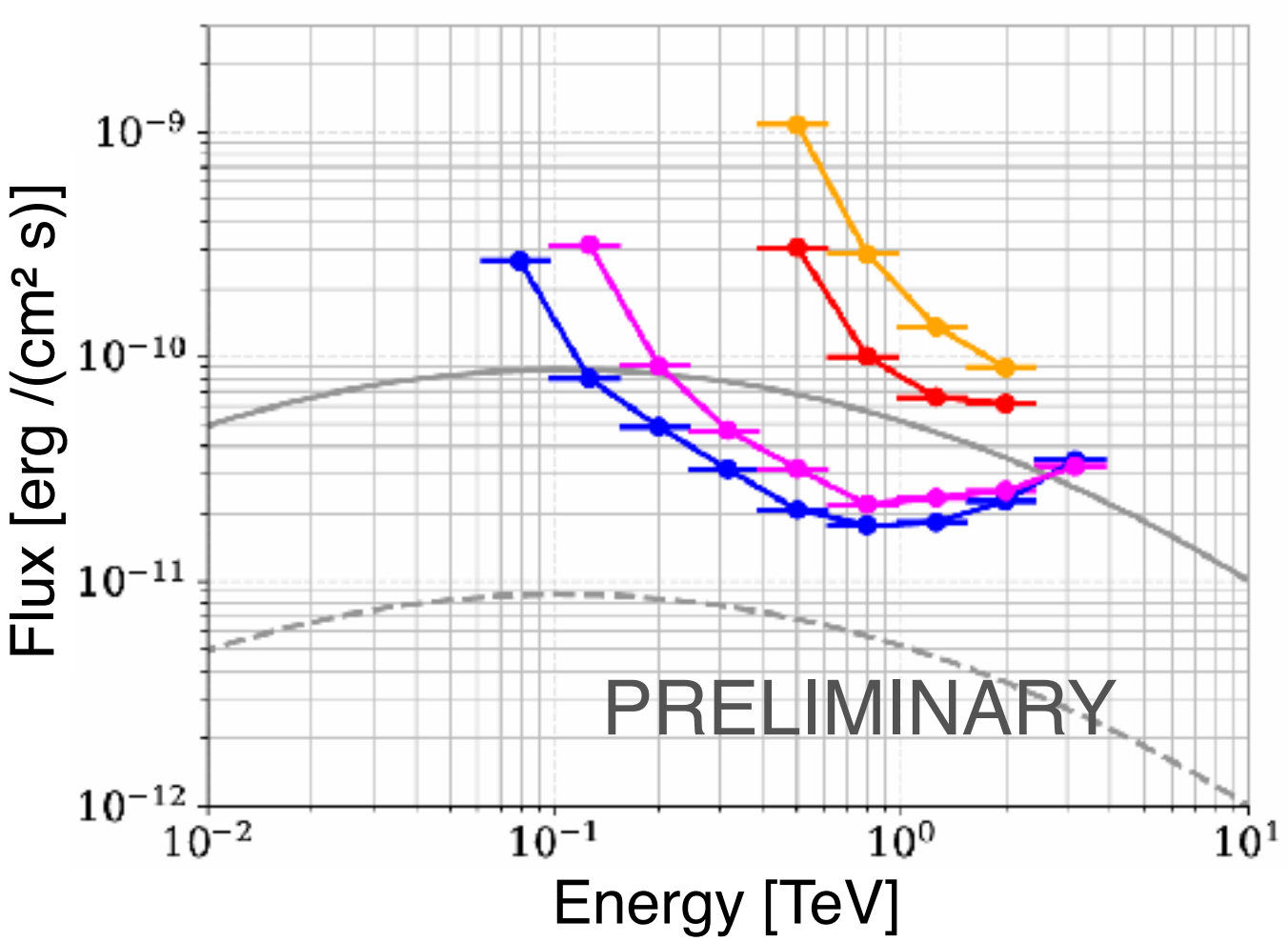
50% Quantum Efficiency

Standard Analysis

Transformer Analysis

100% Crab

10% Crab



Example: Quantum Efficiency

100% Quantum Efficiency

Standard Analysis

Transformer Analysis

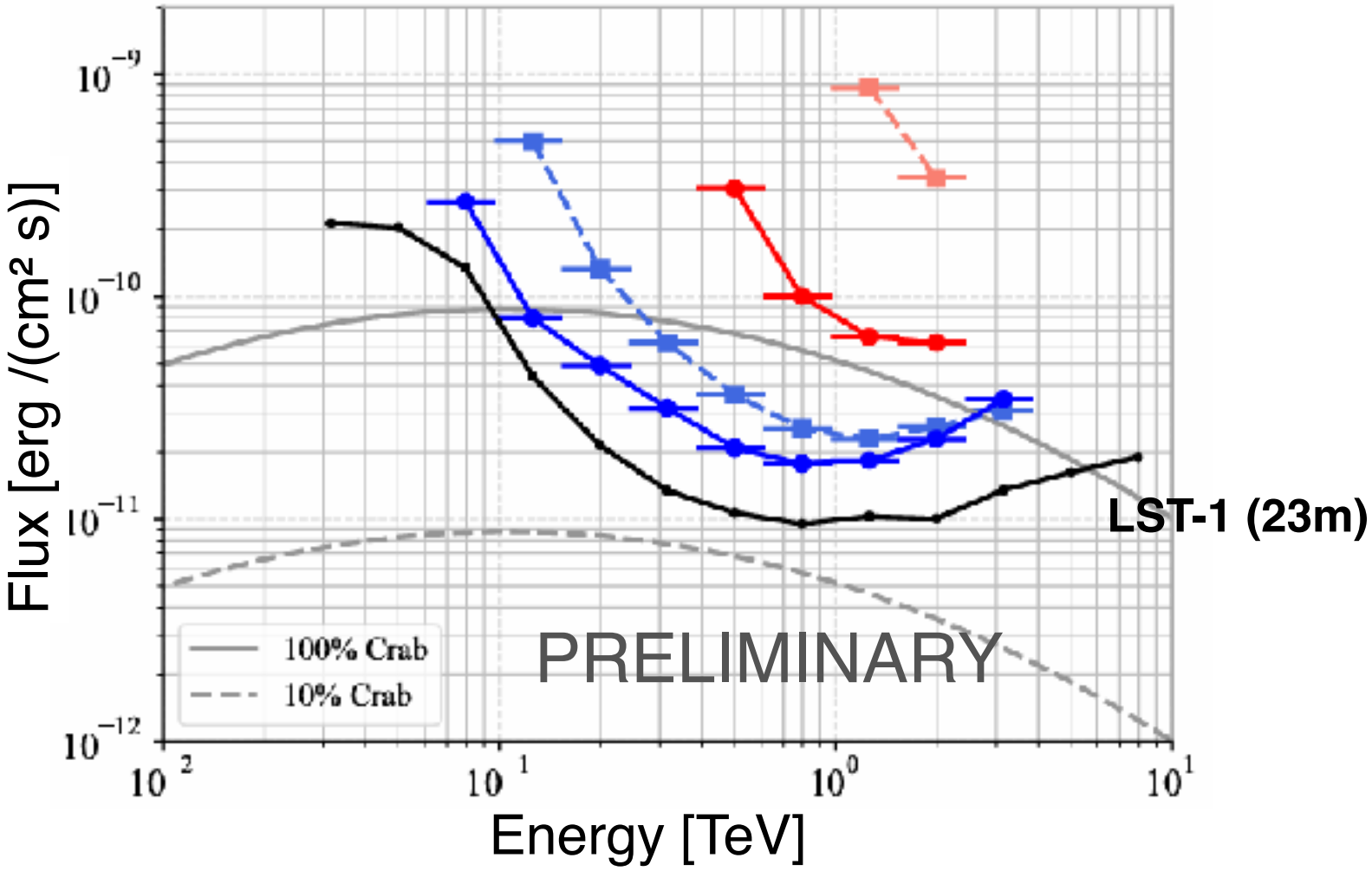
LST-1 Quantum Efficiency

Standard Analysis

Transformer Analysis

100% Crab

10% Crab



LST-1 (23m)

Conclusions

→ **one order of magnitude improvement in telescope sensitivity and large decrease of the energy threshold!**

Increase the sensitivity of current instruments

Construct compact yet sensitive instruments in the future

https://arxiv.org/abs/2608.31148

Outlook

In the mid 2030s:

- many new instruments coming online >10^7alerts/night
- the only sensitive instruments for VHE γ-rays are the CTAO telescopes

→possibility to build a **longitudinally spread array of compact telescopes** with full sky coverage

→ young field of research, potential for discovery by **extending observation into the time domain**

Stay tuned!