

Real-Time AI-Based Triggering for the SST-1M Telescope

Jakub Kvapil on behalf of the SST-1M Collaboration

Digital Twins for Nuclear and Particle Physics – NPTwins 2026
Trento, 11.-15. May 2026

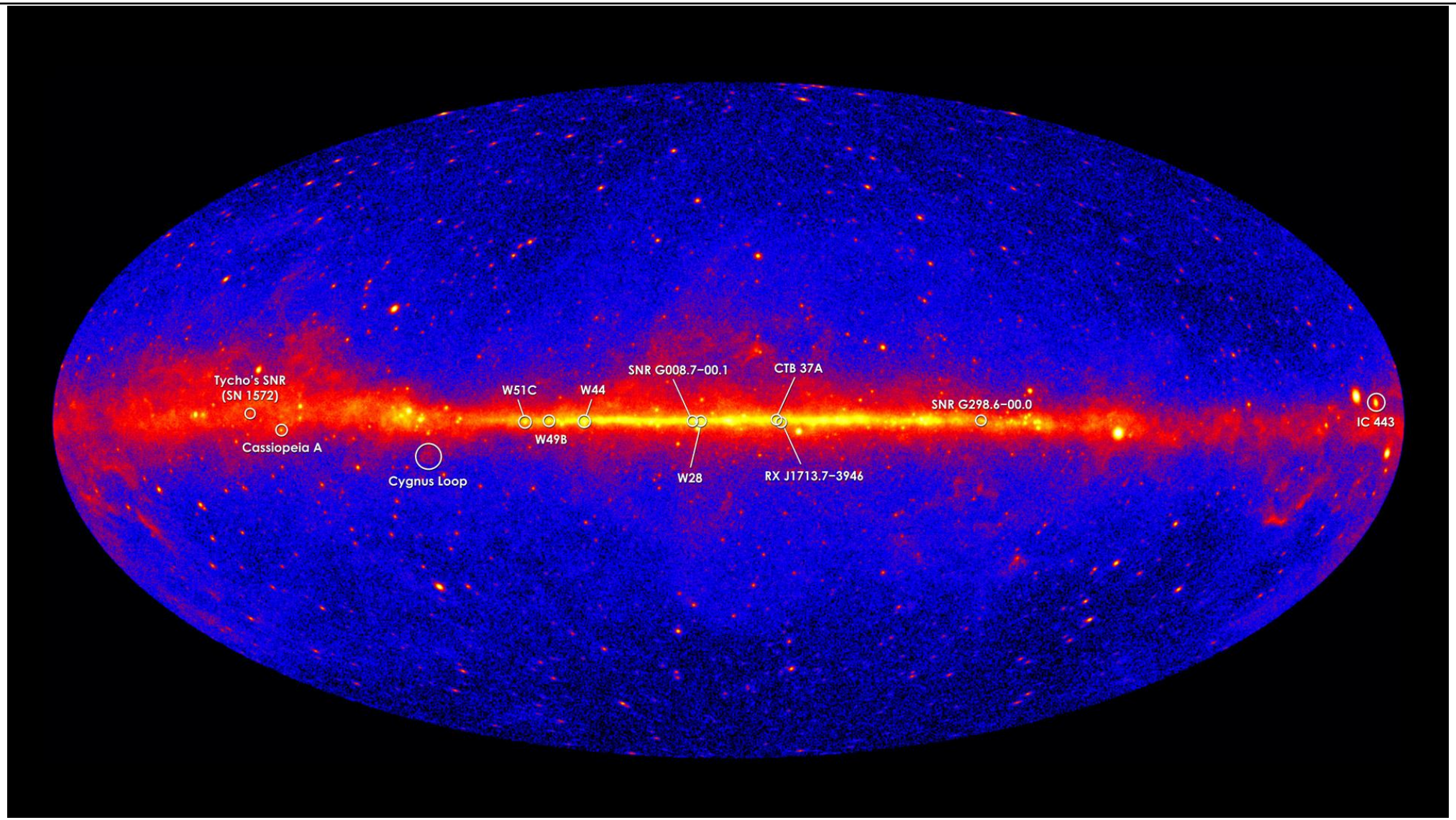
Hes·so GENÈVE
Haute Ecole Spécialisée de Suisse occidentale

SST-1M
Small Size Telescope
Single-Mirror

hepia
Haute école du paysage, d'ingénierie et d'architecture de Genève

Gamma-ray sources in the galaxy

NASA's Goddard Space Flight Center
Images courtesy of NASA/DOE/Fermi LAT Collaboration



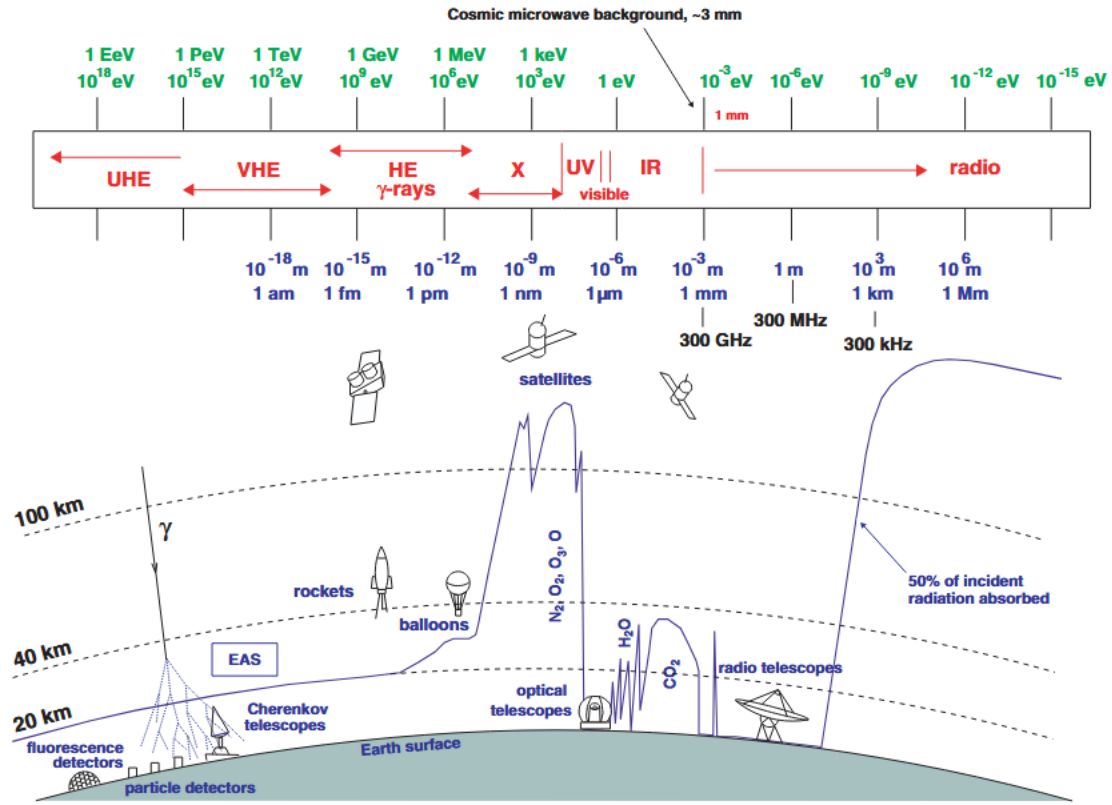
Imaging Atmospheric Cherenkov Telescopes



CTAO: https://www.eso.org/public/images/eso1841i/

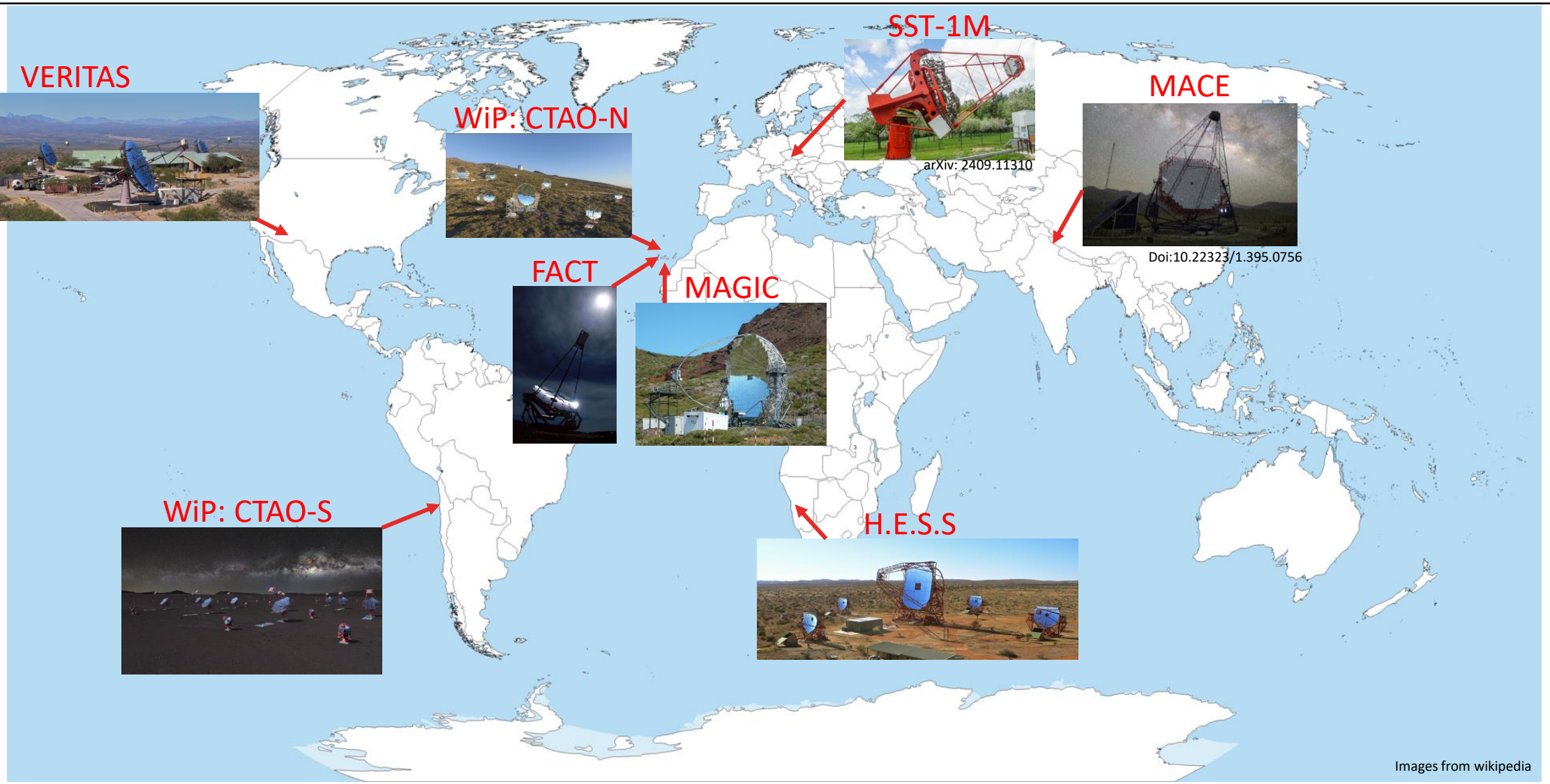
- Primary ray: p, He, ions, **γ**
- Secondary: π, K, μ, e

- Measure Cherenkov radiation produced by secondary showers of primary cosmic ray



Satalecka, Konstancja. (2010). Multimessenger studies of point-sources using the IceCube neutrino telescope and the MAGIC gamma-ray telescope.

Imaging Atmospheric Cherenkov Telescopes in the world



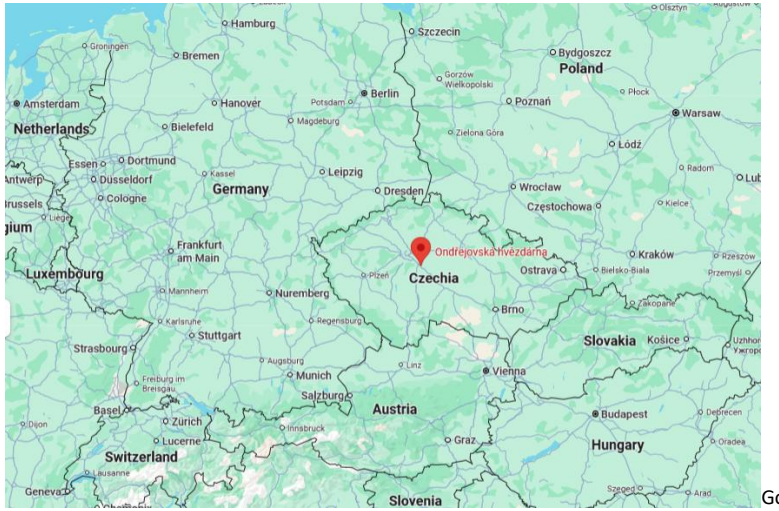
SST-1M Collaboration

SST-1M
Single-Mirror
Small Size Telescope

https://sst-1m.space/

HEPIA RA&D

- Collaboration of 16 institutes, primarily from Switzerland (4), Czech Republic (5), and Poland (5)
- Developed as a design of Small-Sized Telescope for the Cherenkov Telescope Array Observatory (CTAO), alternative design was selected
- Currently operates as an independent collaboration
- Two prototypes constructed and currently operating at the Ondřejov Observatory of Czech Academy of Sciences
- Ongoing discussions to relocate telescopes to either to Pierre Auger Observatory (Malargüe, Argentina) or India Astronomical Observatory (Hanle, India)



Google maps

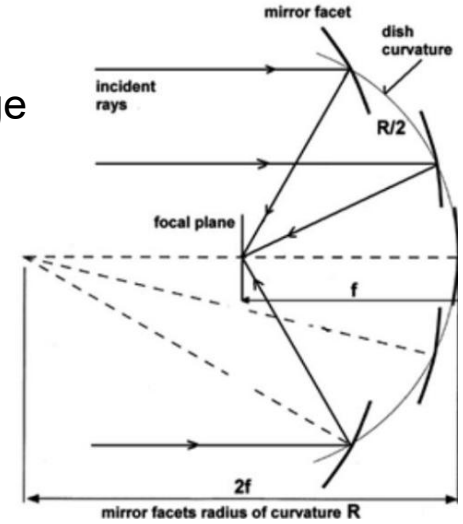


Telescopes are 155 meters apart

DOI: 10.1088/1475-7516/2025/02/047

SST-1M Telescopes

- Optimized for γ-ray showers in the 1–300 TeV range
- Mechanical design
  - Lightweight (~8.6 t)
  - Cost-effective
  - Modular (transportable in a shipping container)
- Optical design (Davies-Cotton)
  - 18 hexagonal mirror facets
  - High reflectivity in 300–550 nm range



Patrik Čechvala, Triggering Discoveries in HEP III

Telescope 1

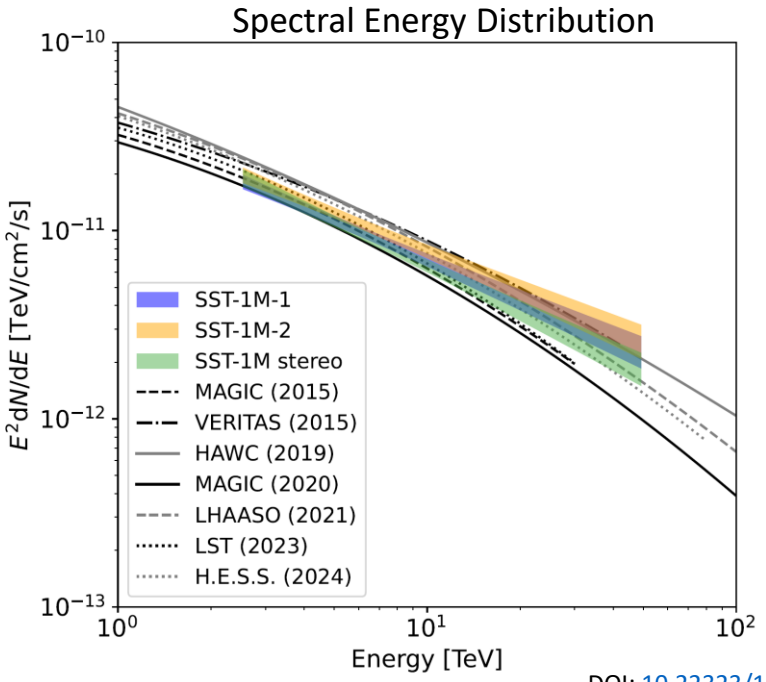


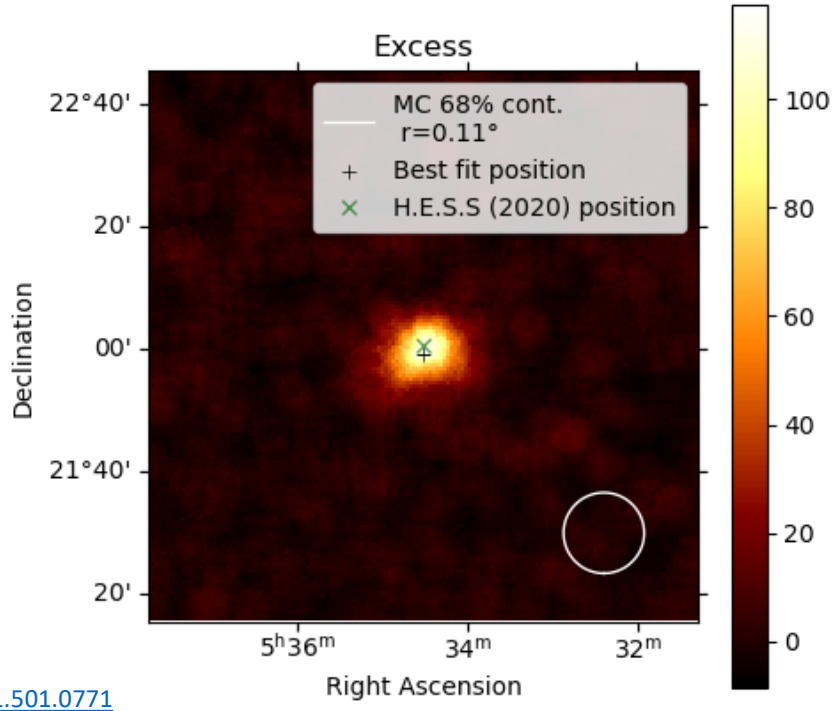
Telescope 2



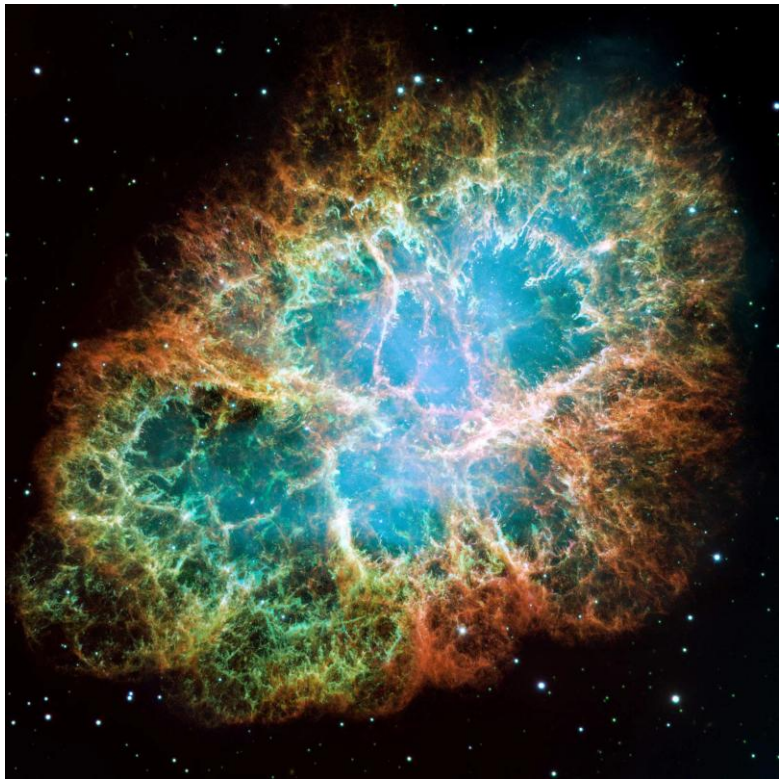
SST-1M Observation – Crab Nebula

- Crab Nebula has a well characterized spectrum
  - Widely used as a benchmark for calibration and analysis
- Observation September 2023 – March 2024





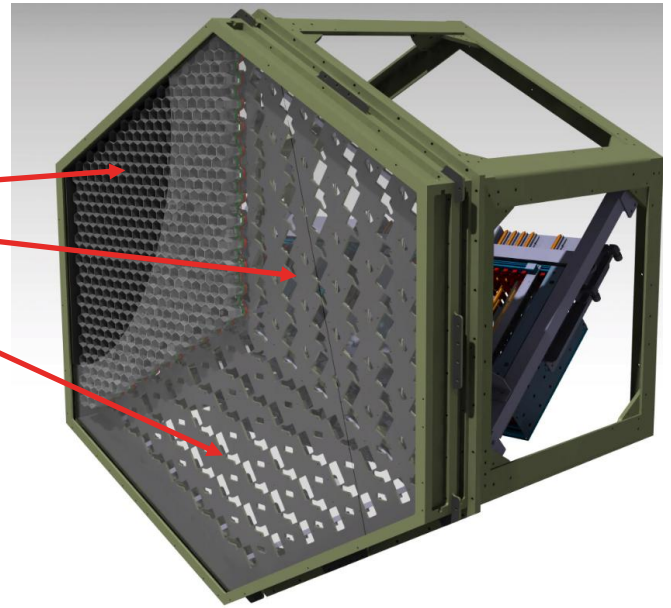
DOI: 10.22323/1.501.0771

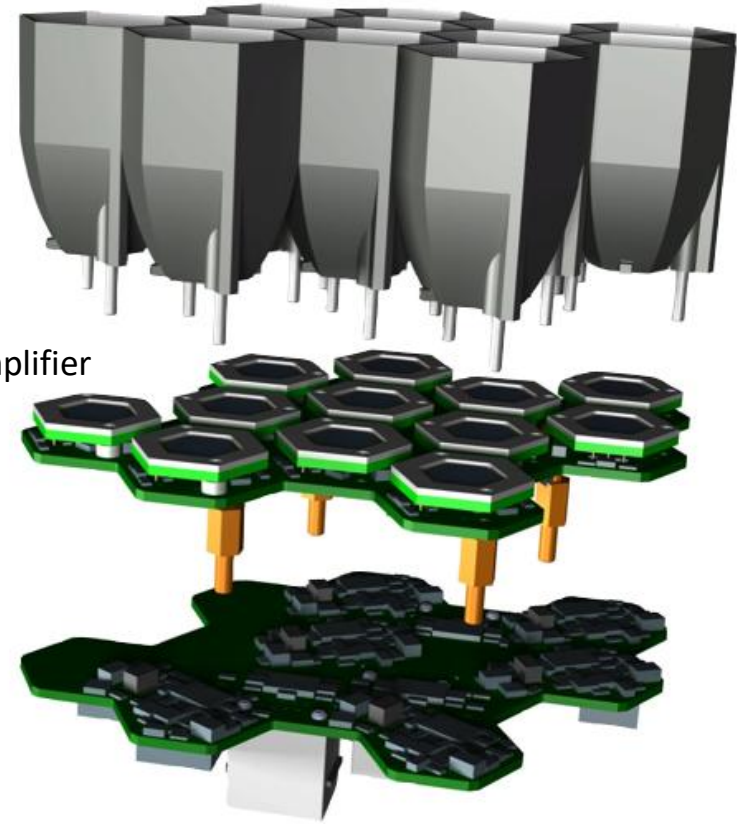


NASA/ESA Hubble Space Telescope. (n.d.). *Messier 1 (Crab Nebula)*. Retrieved from https://science.nasa.gov/mission/hubble/science/explore-the-night-sky/hubble-messier-catalog/messier-1/

SST-1M Camera

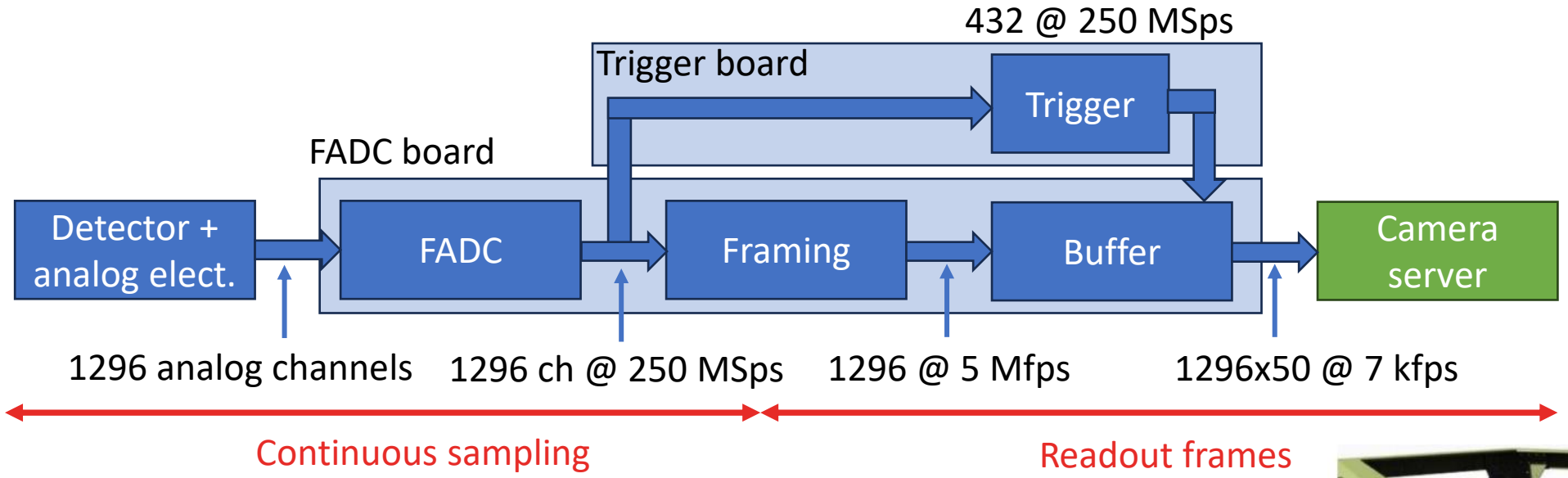
- Fully digital camera: trigger and readout (FADC 250MSps@12bits, Xilinx Virtex 7)
- Silicon photomultiplier (SiPM)-based camera design
  - Optimized for fast Cherenkov light flashes (ns scale)
  - Custom hexagonal Hamamatsu Photonics SiPMs
    - 1296-pixel array, pixel apothem 12.15 mm
- Readout architecture
  - Sampling frequency of 250 MHz
  - 3 readout segments
  - Each segment includes:
    - 9 FADC boards
    - 1 trigger board
  - One trigger board operates as the





DOI 10.1140/epjc/s10052-017-4609-z
DOI 10.1088/1475-7516/2025/02/047

SST-1M Readout chain

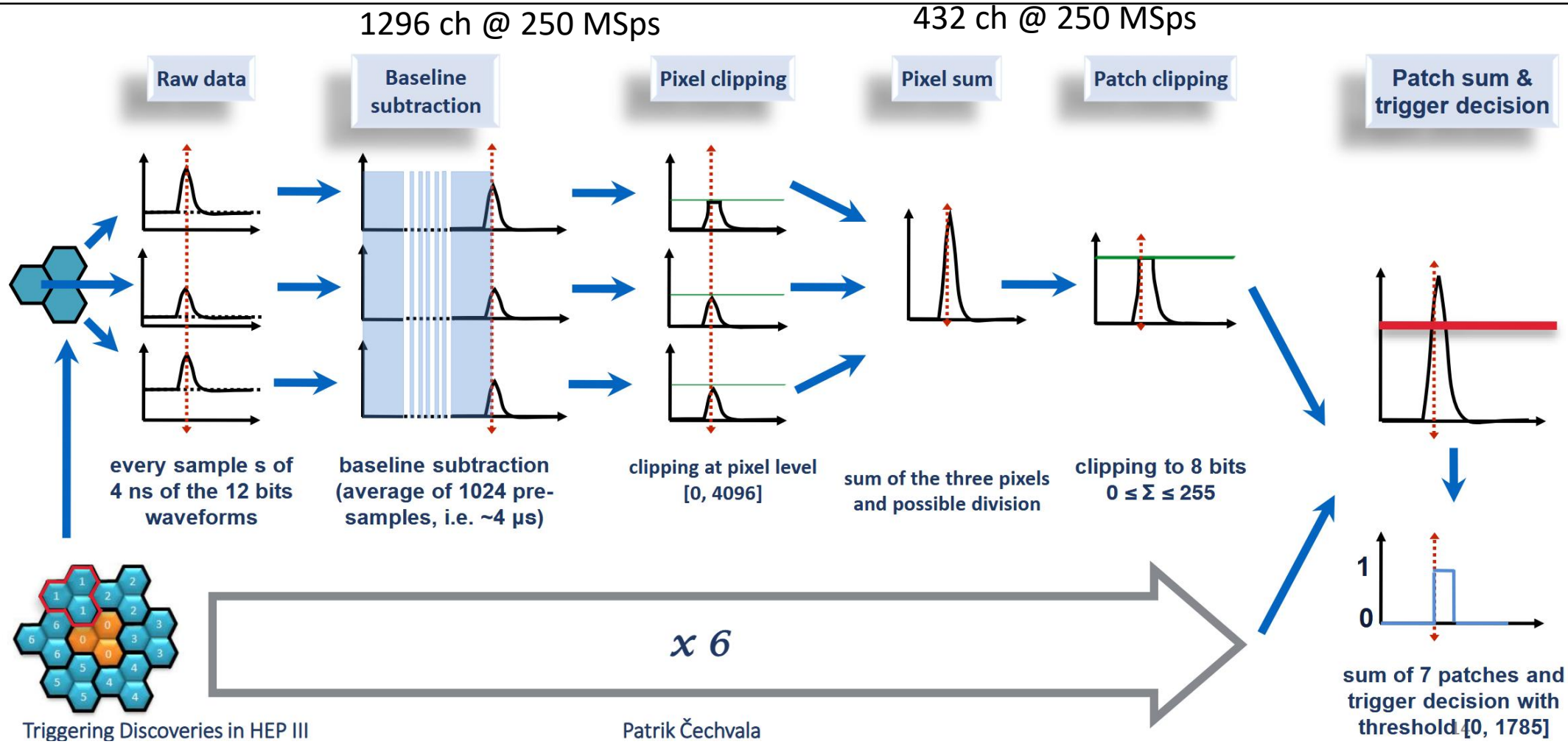


- 12-bit ADC, 250 MHz sampling, 50 samples per frame
- 10G Ethernet data readout interface
  - Maximum readout rate ~7 kHz frames
  - Required background rejection factor for trigger ~714



DOI 10.1088/1475-7516/2025/02/047

SST-1M Trigger chain



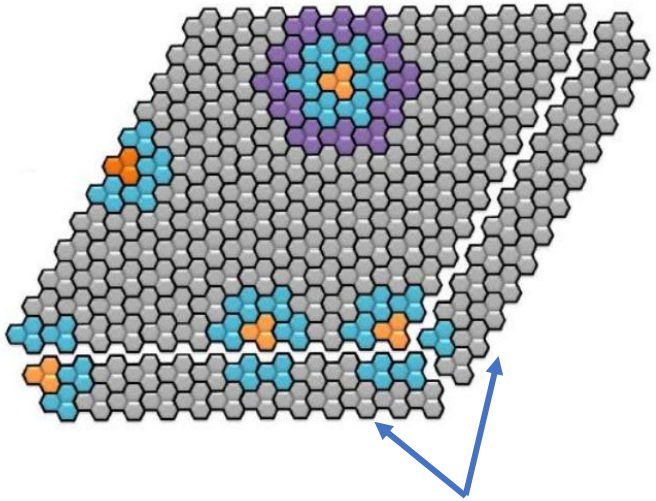
Patrik Čechvala, Triggering Discoveries in HEP III

- Default trigger scheme is called PATCH7 – triggered if group of 7 triplets sum (21 pixels) crosses threshold

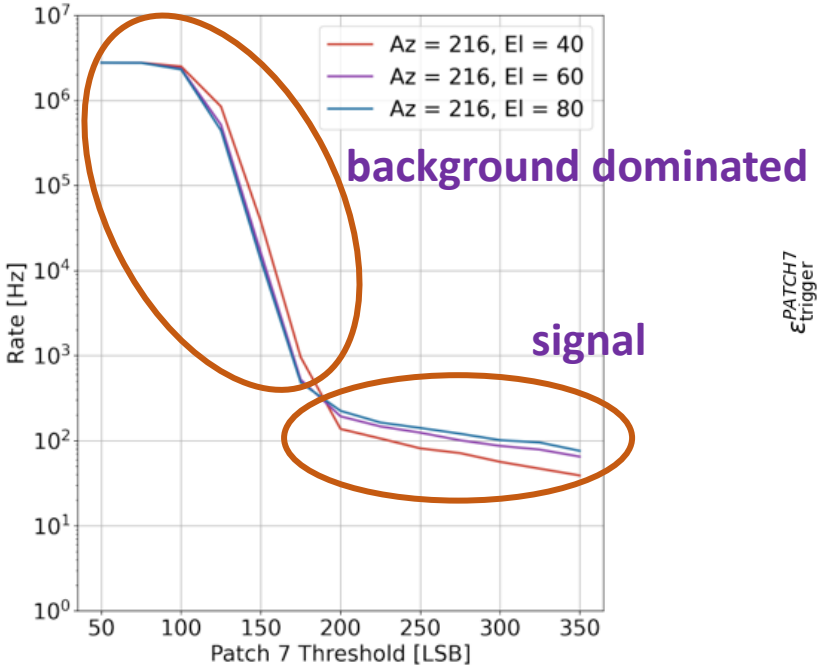
SST-1M: PATCH7 trigger

- Default trigger is PATCH7 – sum of 21 pixels
  - For the study, threshold is set to corresponds to 7 kHz background rate which results in integrated signal efficiency of about 25%
- Can performance be improved?

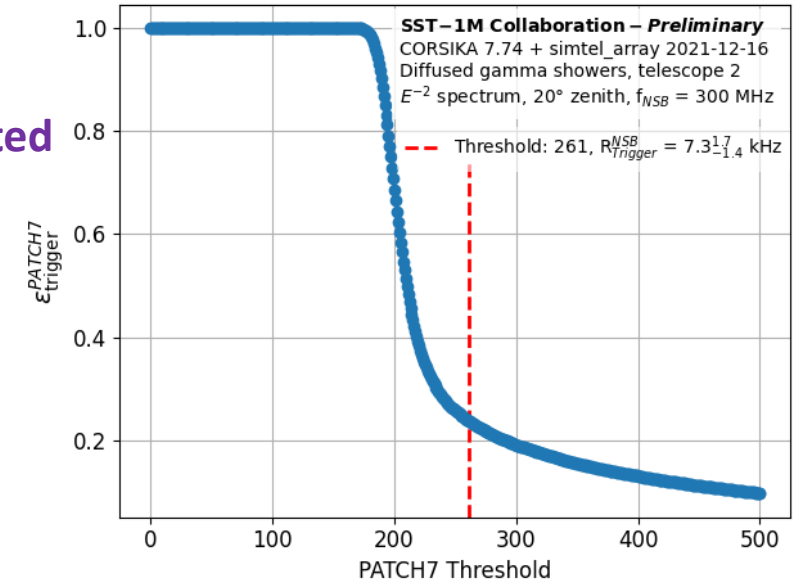
1/3 of the camera pixels

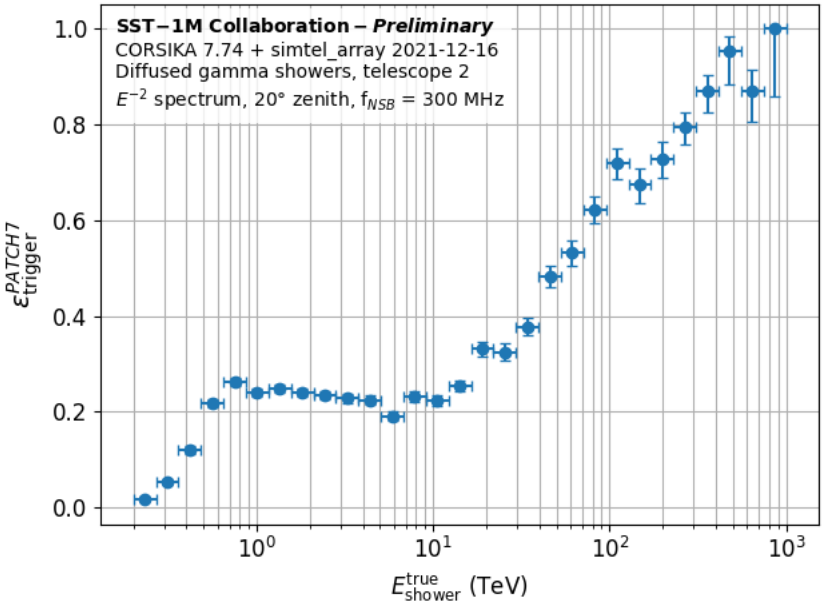


Overlap between sectors required to ensure lossless trigger



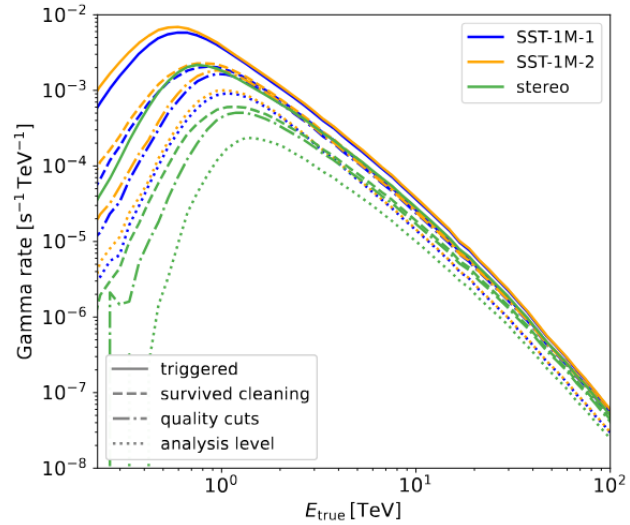
Alispach et al., in press

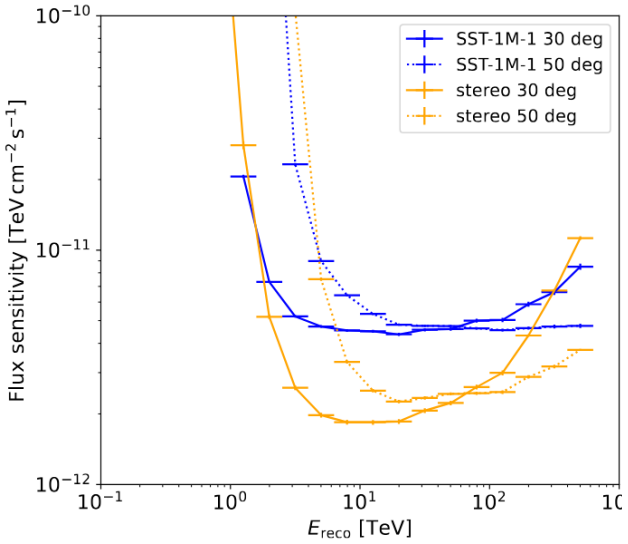


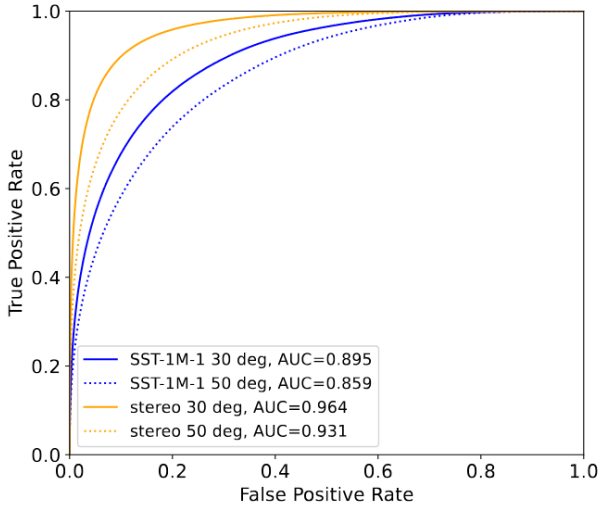


SST-1M: Stereoscopic trigger

- The two telescopes are synchronized using the White Rabbit protocol
- Currently, only software-based stereoscopic trigger is implemented
  - After independent PATCH7 triggers, each telescope sends data independently, which saturates readout at approximately 7 kHz
  - Stereoscopic observation improves the flux sensitivity (less time needed for observation) and improves ROC curve in the offline version of Random Forest gamma/proton classifier
- A hardware stereoscopic trigger would significantly improve the system performance
  - This would require real-time data exchange between the two telescopes





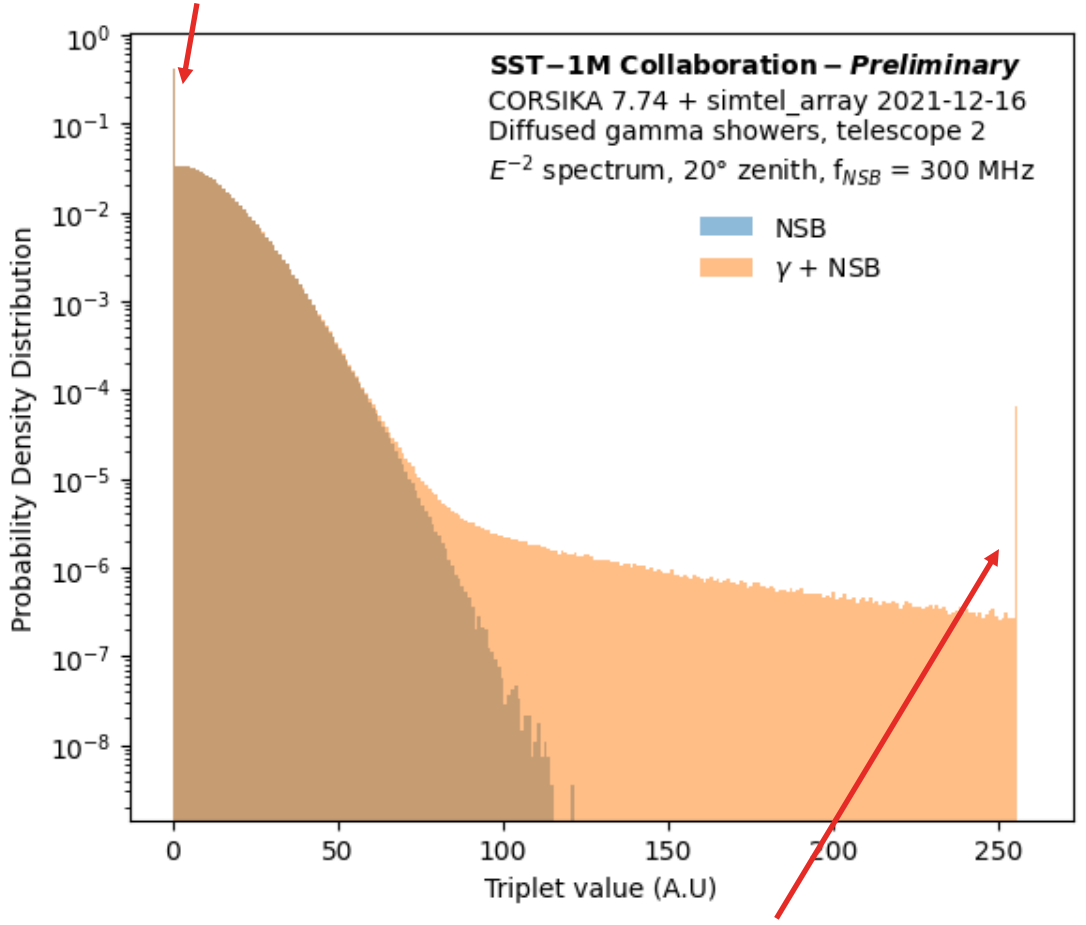


DOI: 10.1051/0004-6361/202555292

What is signal and what is background?

- The signal are the γ-induced showers (order of Hz)
  - Proton-induced shower (~1 kHz) that creates similar patterns are not rejected at trigger level and are handled offline
  - Photon bunches < 10 ns
- The background is mainly from Night Sky Background (~300 MHz)
  - Starlight, moonlight, airglow, Zodiacal light, artificial light pollution
  - Single photons
  - The readout samples at 250 MHz: ~1.2 NSB photoelectron per pixel per 4 ns sample

Residual from baseline subtraction

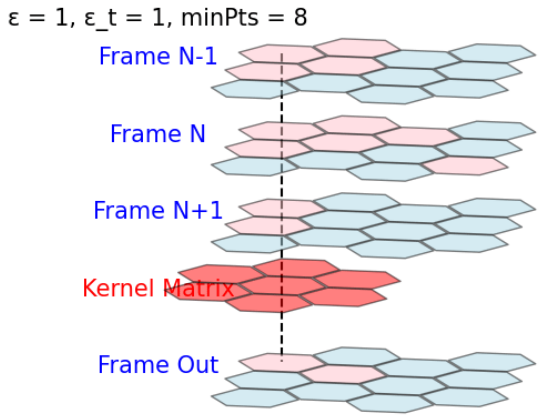


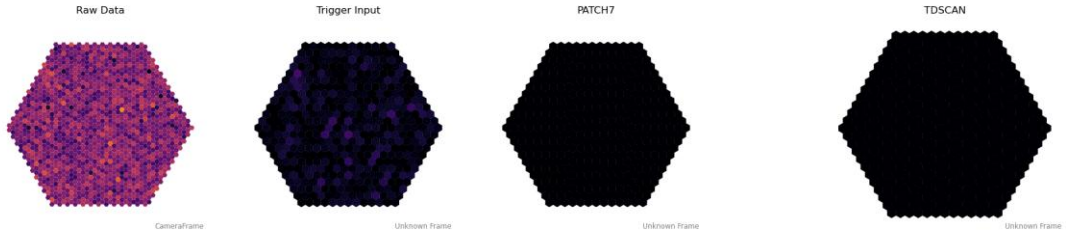
Residual from saturation and clipping

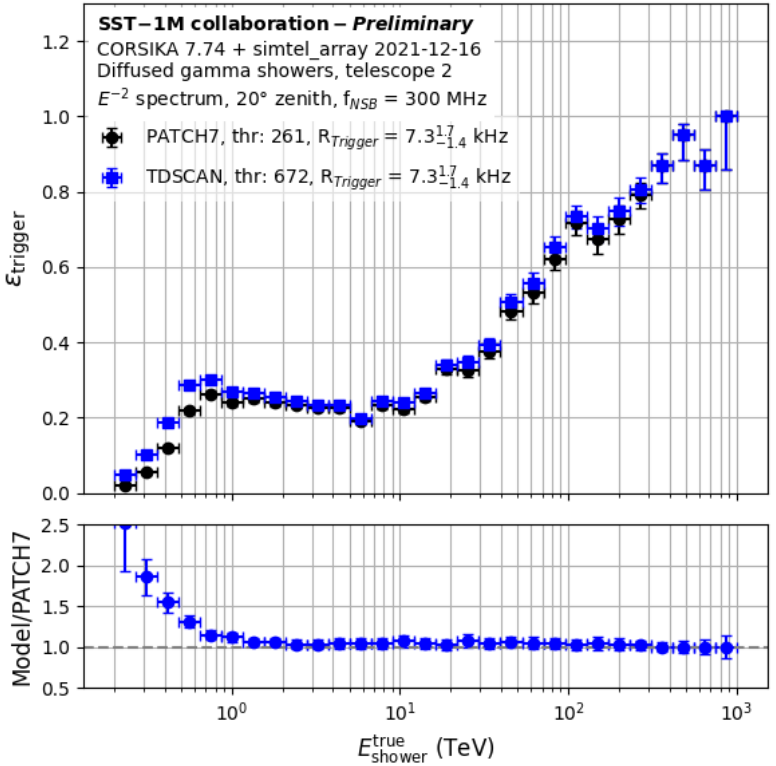
SST-1M: TDSCAN development

- DBScan (Density Based Spacial Clustering of Application with Noise)
  - Standard software/based clustering tool
  - Variable latency, non-parallelizable, not designed for continuous images
- TDScan (Trigger Distributed Spatial Convolution Accelerator Network)
  - Bottom-up FPGA-driven approach
  - Spatio-temporal clustering of camera waveforms
  - FPGA implementation
    - Fully parallelizable, no BRAM, complexity of $O(N)$
    - Fixed latency and high speed – tested up to 1 GHz
- Reduces random NSB while keeping shower structure

doi:10.5555/3001460.3001507



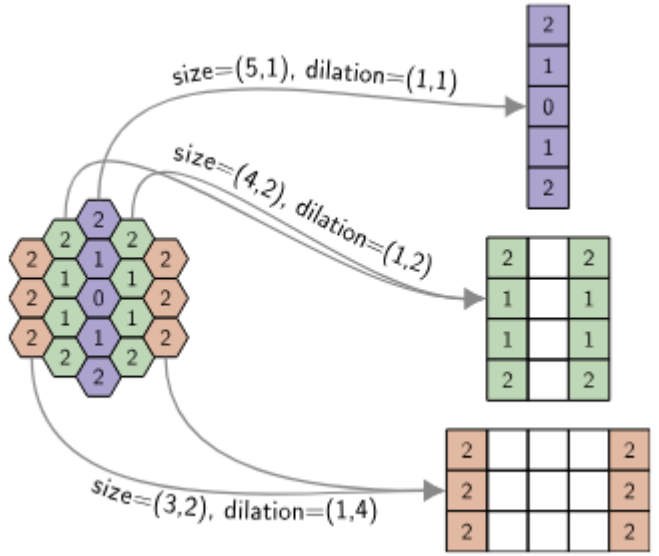


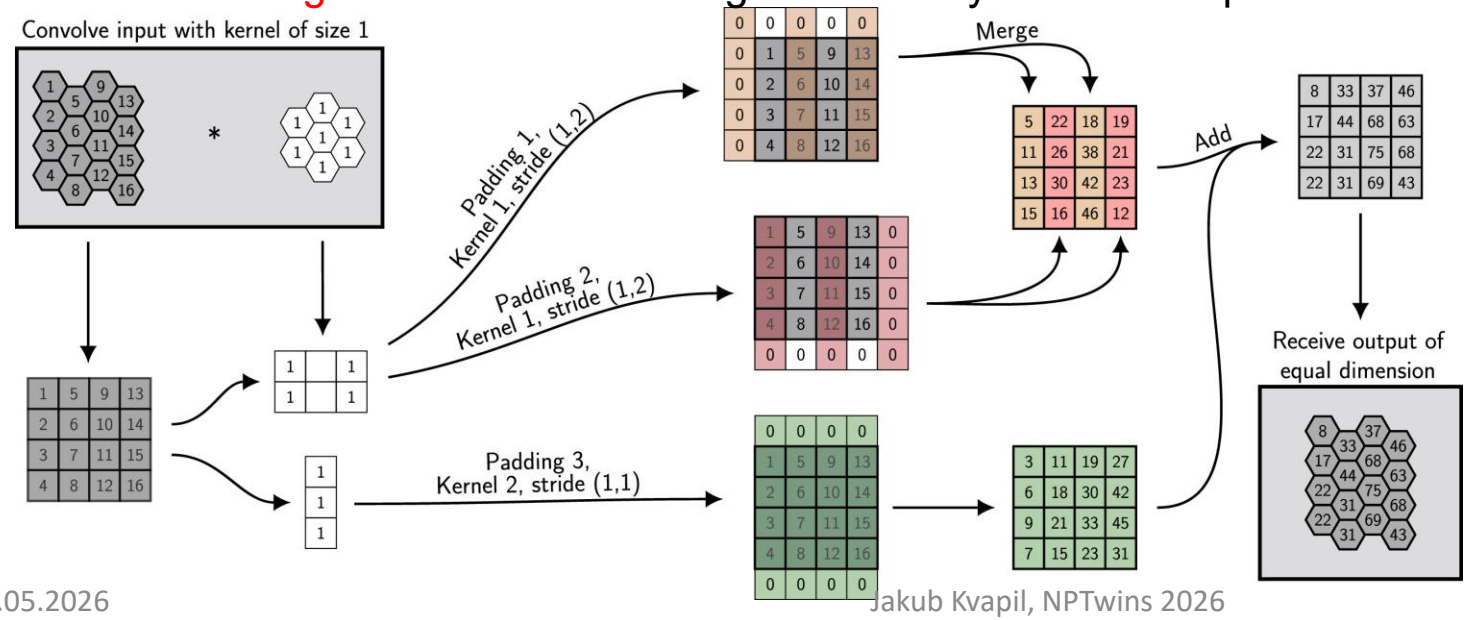


- Will be deployed either in parallel or as a replacement for the PATCH7 trigger
- Achieves a 31% efficiency improvement for 500 GeV showers!

Neural Network based triggering with hexagons

- Initial focus on single telescope triggering
- Approach: Convolutional Neural Network on FPGA
  - Deployment strategy depends on inference latency/rate
    - Either as a standalone trigger network or with PATCH / TDSCAN pre-trigger
- The challenge: Camera pixels are hexagonal (6 neighbors)
- Use of hexagdly package (https://github.com/ai4iacts/hexagdly)
  - Emulates hexagonal convolution using standard Pytorch CNN operations





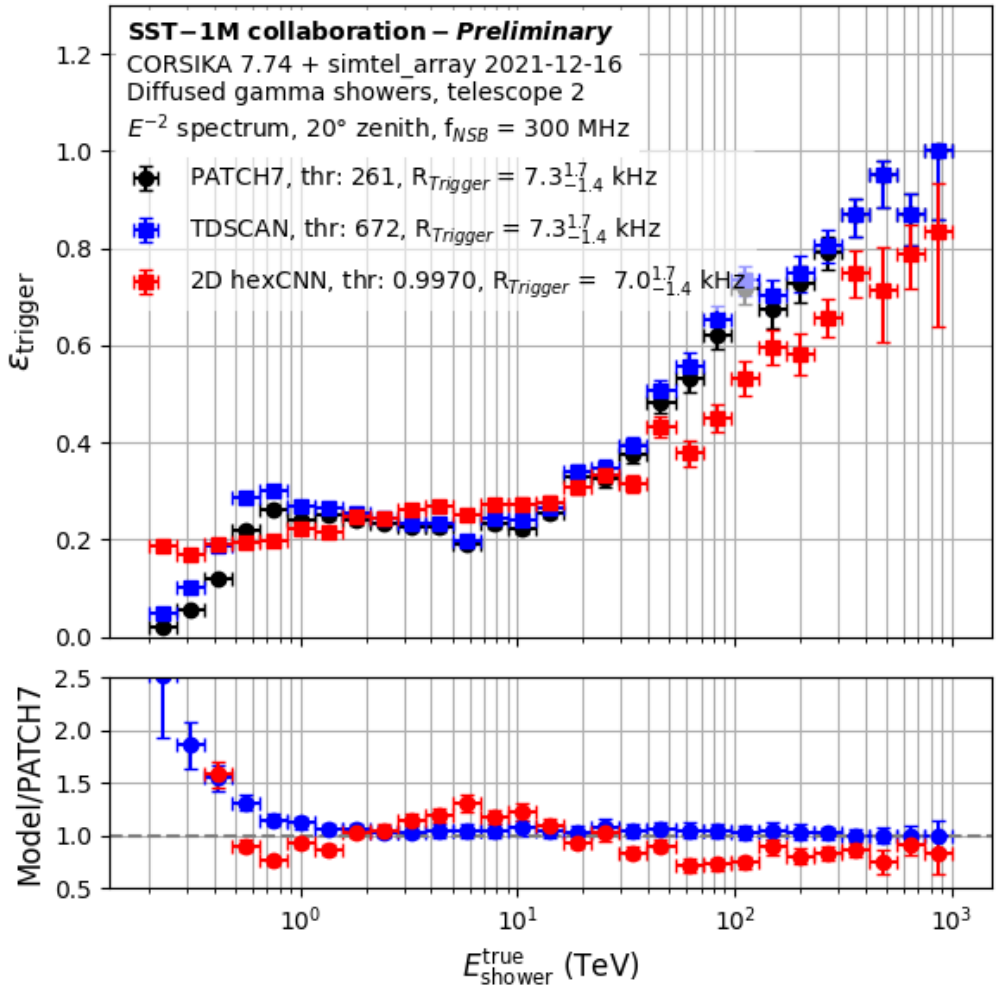
Steppa, C., & Holch, T. L. (2019). *HexagDLy—Processing hexagonally sampled data with CNNs in PyTorch*. SoftwareX, 9, 193–198. https://doi.org/10.1016/j.softx.2019.02.010

Image-like CNN model development

- Hexagonal CNN using hexagdly package
- Image-like approach (2D convolution)
  - Integrate the 50-time samples in a frame
- Image network:

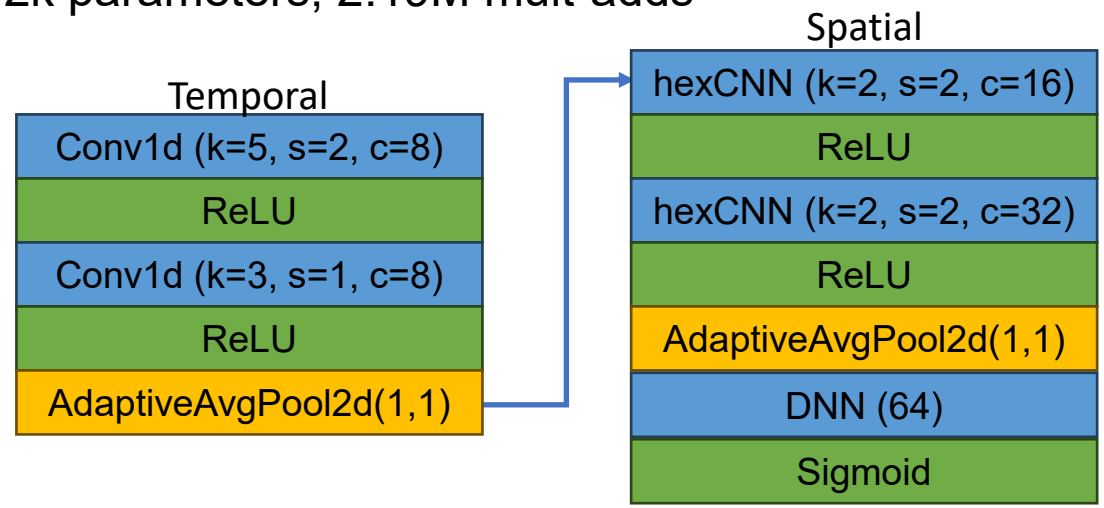
hexCNN (k=2, s=2, c=32)
ReLU
hexCNN (k=2, s=2, c=64)
ReLU
AdaptiveAvgPool2d(1,1)
DNN (64)
Sigmoid

- Best hypersampled model
  - 39.4k parameters, 0.17M mult-adds
  - Improvement in low energy, but worse in large energies compared to PATCH7 or TDSCAN triggers

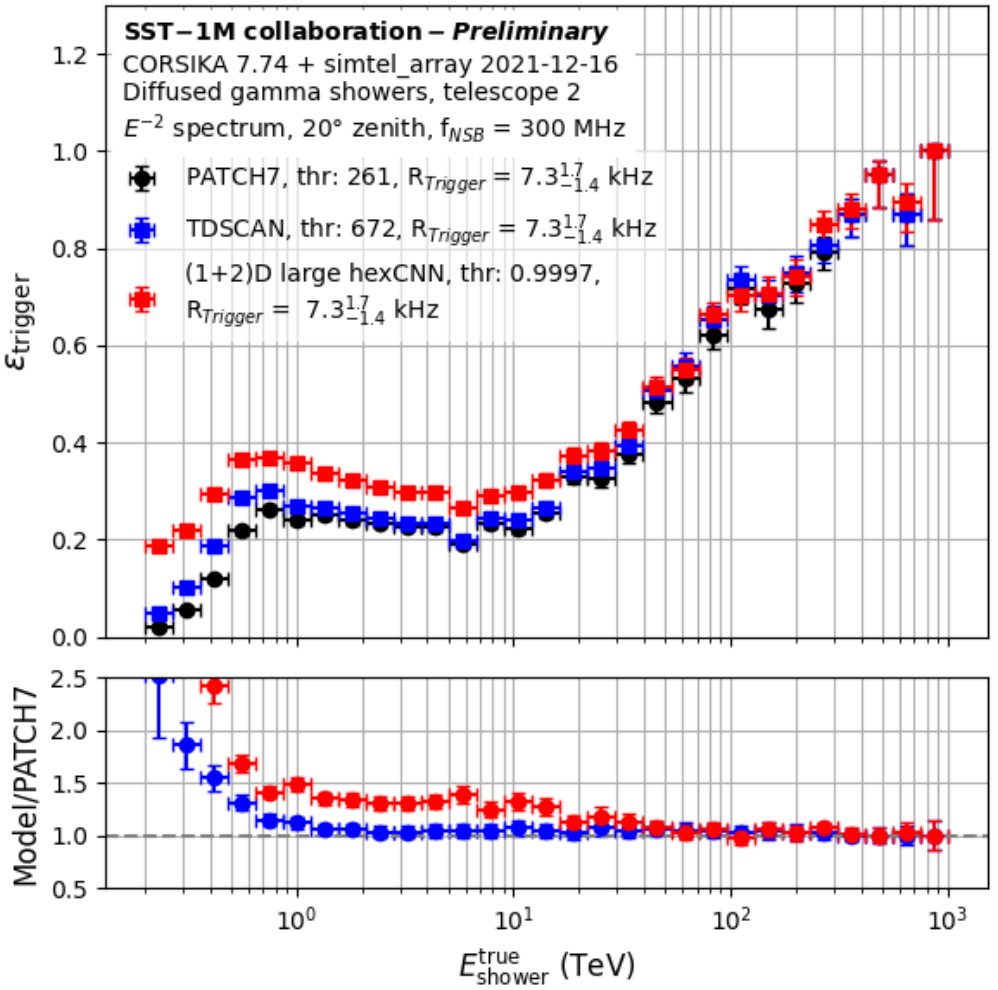


Adding temporal dimension

- Full 3D convolution will be computational heavy for FPGA
- Separate temporal and spatial convolution (1+2)D
- Initial architecture prototype - not yet optimized!
  - 12k parameters, 2.19M mult-adds



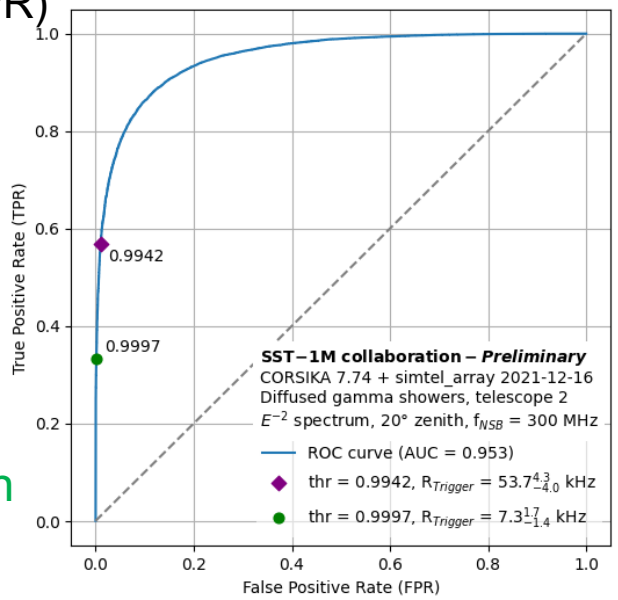
- There is a significant improvement in efficiency bellow 20 TeV!
  - 68% for 500 GeV showers



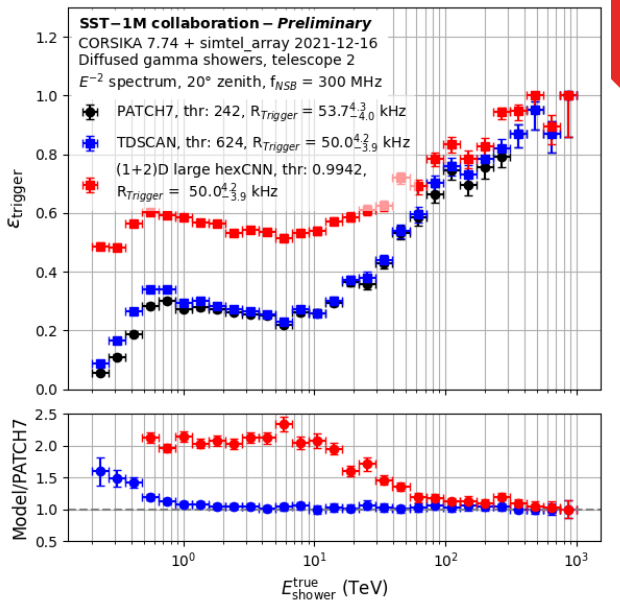
The problem of trigger rate

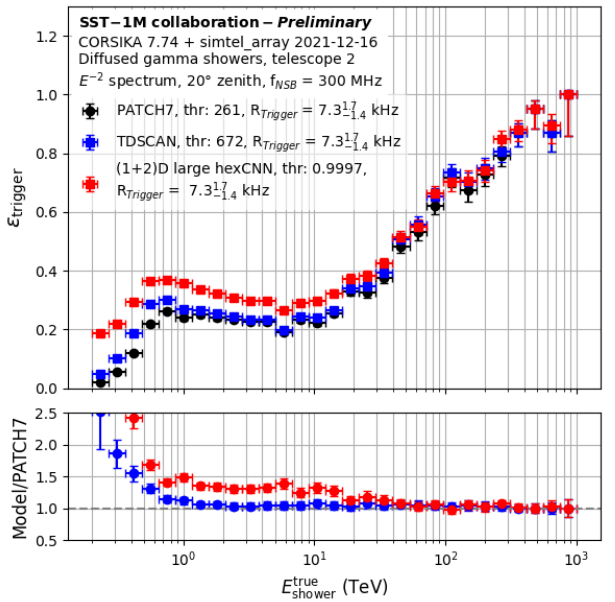
- Background dominated regime
  - Requires operation at very low false positive rate (FPR)
    - Directly translates to trigger rate
- Currently trained on semi-balanced dataset (signal dominated) with binary cross-entropy loss
  - Optimizes global ROC, not low FPR region
  - Possible improvement by adapting the loss function
- Performance gains are significant at low energy with higher rate!
  - Higher rate operation would be possible using hardware stereoscopic trigger

Trigger rate 50 kHz →



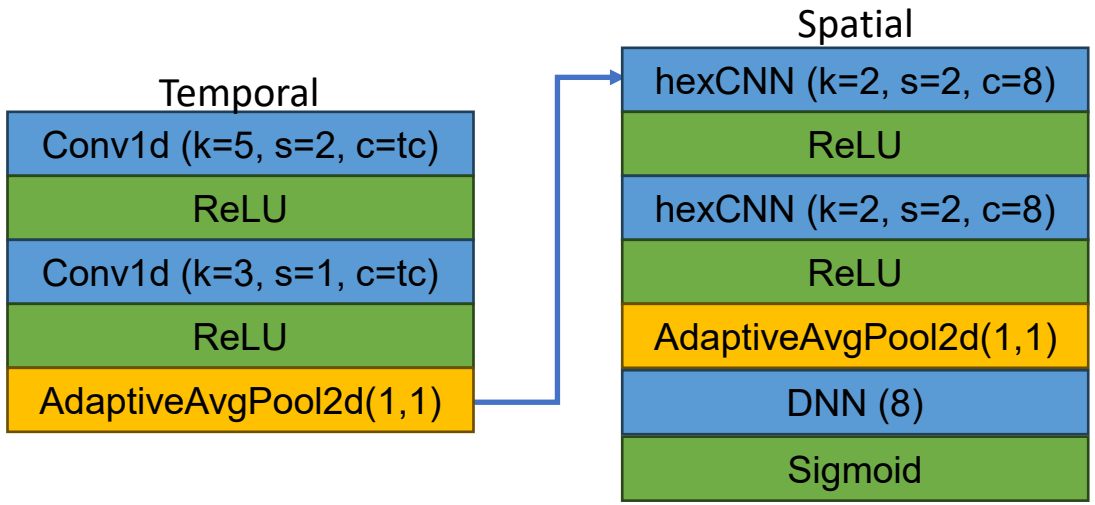
Trigger rate 7 kHz →



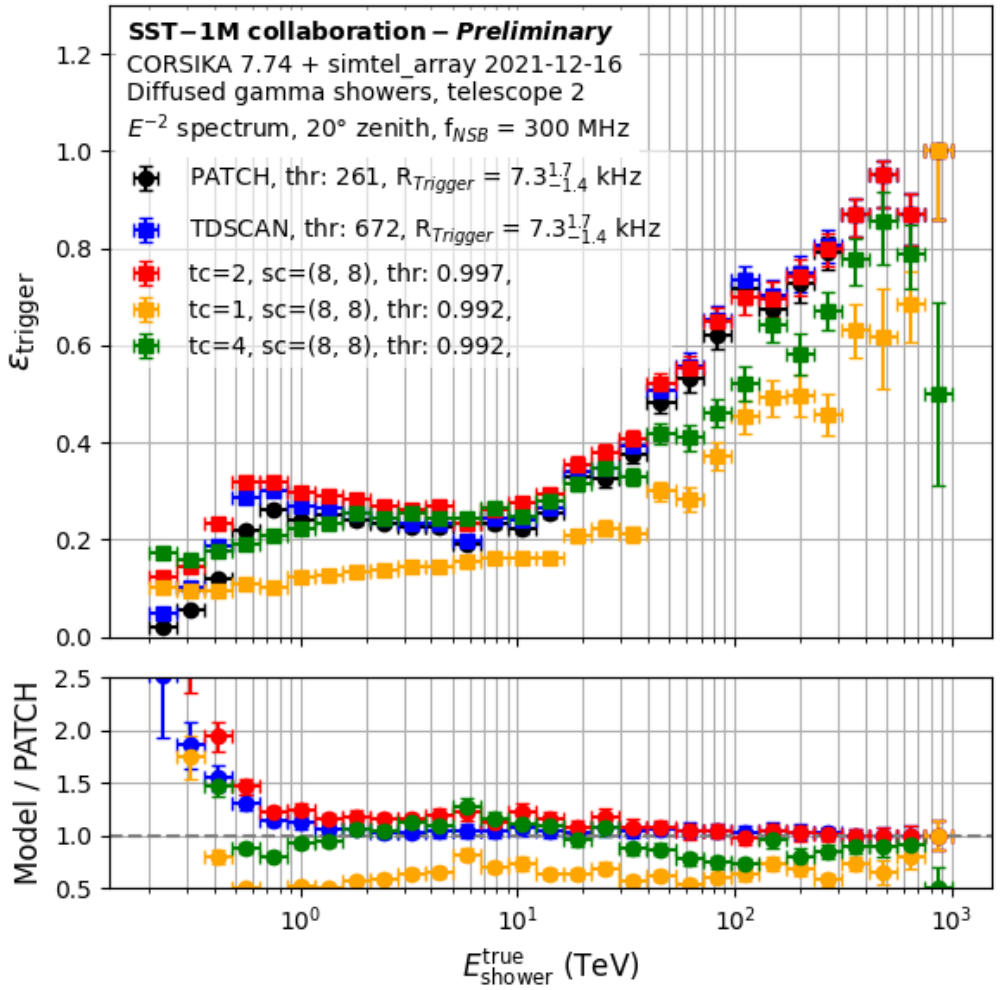


Smaller temporal model

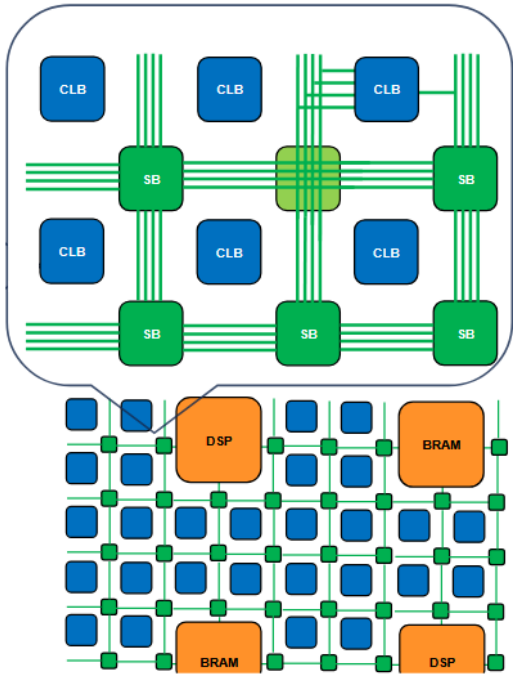
- First look at reduced complexity
- Variation of temporal depth (tc) and spatial channels reduction
  - Shrinks the model ~1.5k parameters and 0.18M mult-adds



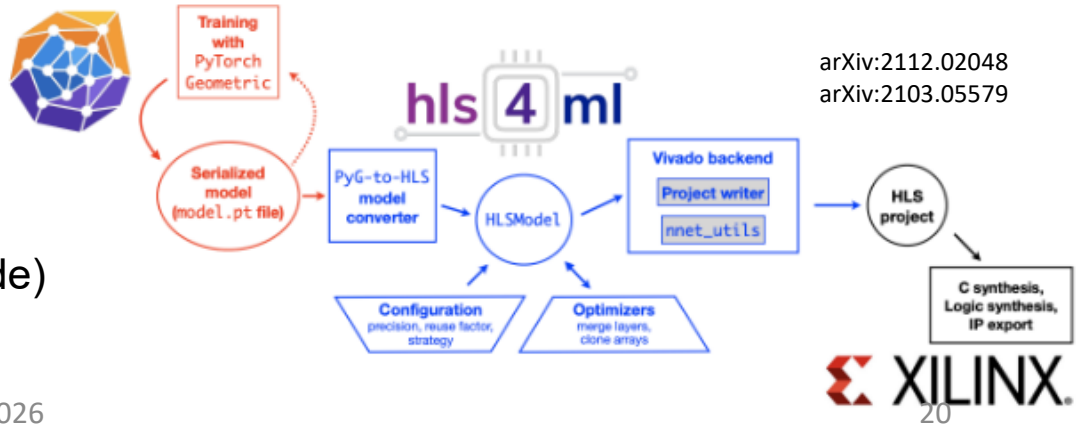
- Achieves comparable performance to baseline models
  - One configuration outperforms PATCH trigger
- Next steps:
  - Interpret the latent space to guide architecture design



Triggering on FPGA – the workflow



1. Model development and optimization (ongoing)
   - Design a compact model architecture – balance between accuracy, latency, and size
     - Goal: reduce computational complexity without large accuracy hit
2. Pruning
   - Removal of weights/neurons with negligible contribution
     - Usually small improvement on GPU, but significant resource saving on FPGA
3. Quantization
   - Converting weights and activations to lower precision e.g. FP32 → INT(2-8)
     - Aim for best compromise between accuracy and hardware performance
4. Model translation to hardware language
   - Can be done with tools like hls4ml
   - However, some functions from Pytorch are not ported into the Vitis backend (multidimensional cat, hexagdly special stride) and must be implemented



arXiv:2112.02048
arXiv:2103.05579

Summary

- SST-1M telescopes are successfully taking data in Ondřejov Observatory
  - Ongoing discussions on relocation to Argentina or India
- The readout firmware is being upgraded to support TDSCAN and CNN inference via hls4ml
- Initial development of a hexagonal convolutional-like trigger architecture
  - Current work focuses on trade-offs between architecture design and FPGA constraints
  - Once the trigger is designed, it will be necessary to verify that low-energy events remain reconstructable in the analysis pipeline
    - A heavier offline CNN reconstruction model could also be implemented and trained jointly with the trigger network
  - Efficiency improvement for 500 GeV showers ranging from 68% to 112% depending on the output rate
- The CNN deployment strategy will depend on the achievable inference latency
  - Either as a standalone trigger network, or chained operation with PATCH / TDSCAN pre-trigger