

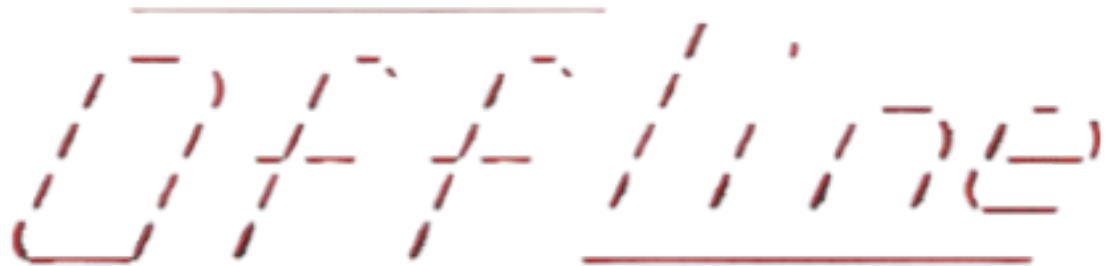
PIERRE  
AUGER  
OBSERVATORY

# Update on the Offline Analysis Framework for AugerPrime and integration of the AugerPrime Radio Detector reconstruction

**Marvin Gottowik**

for the *Pierre Auger Collaboration*,  
Karlsruhe Institute of Technology (KIT)

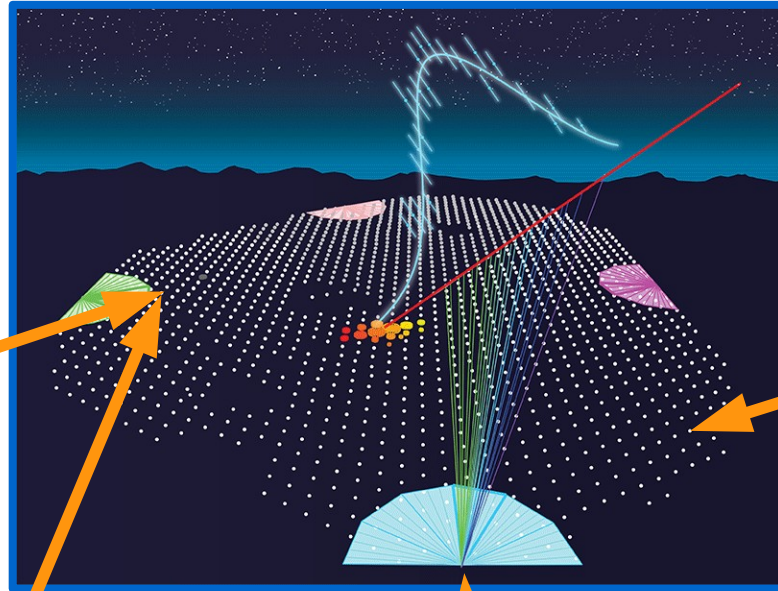
RICAP2024  
26.09.2024



Auger Offline Framework v3r99p99-  
Reference: Nucl. Instrum. Meth. A 580 (2007) 1485

# The Pierre Auger Observatory

**Auger Engineering  
Radio Array (AERA)**



**Water Cherenkov Detector (WCD)**

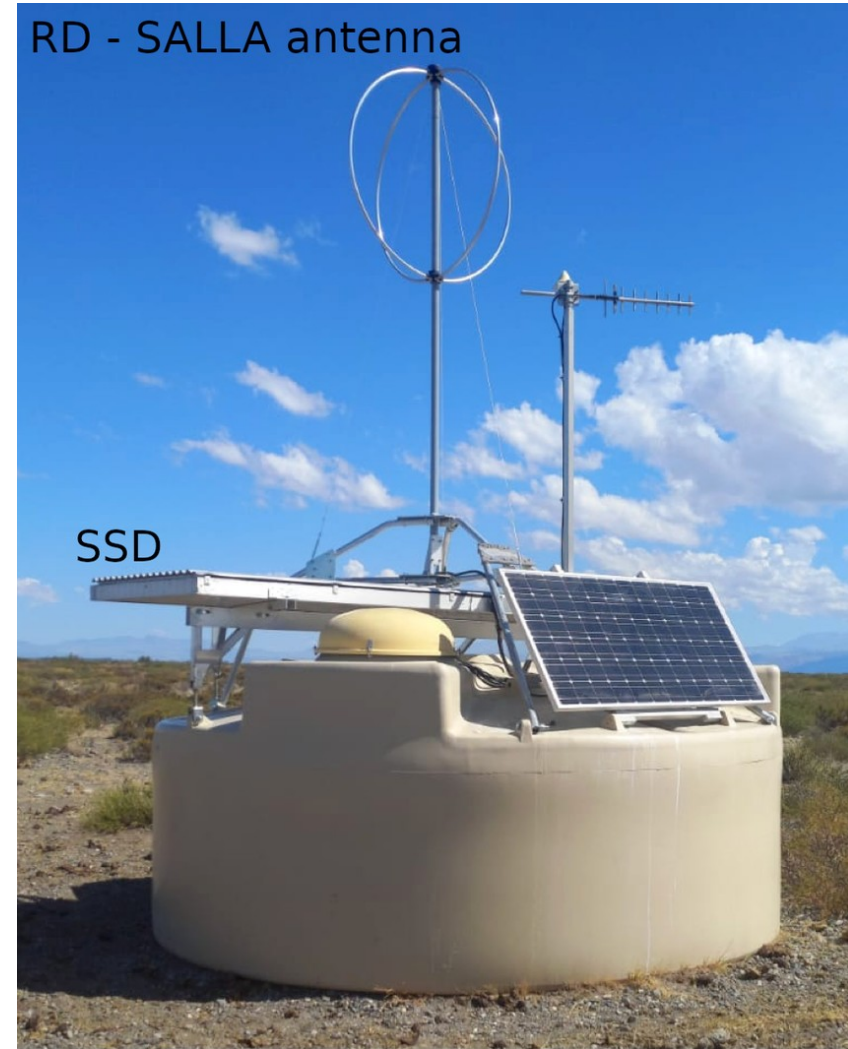
**Fluorescence Detector (FD)**

**Underground Muon Detector  
(UMD)**



# AugerPrime

- Enhanced sensitivity to the mass of cosmic rays
- New detector types (small PMT, SSD, RD) and new station electronics
- Requires changes to the Auger Analysis Framework



# Auger Offline framework

---

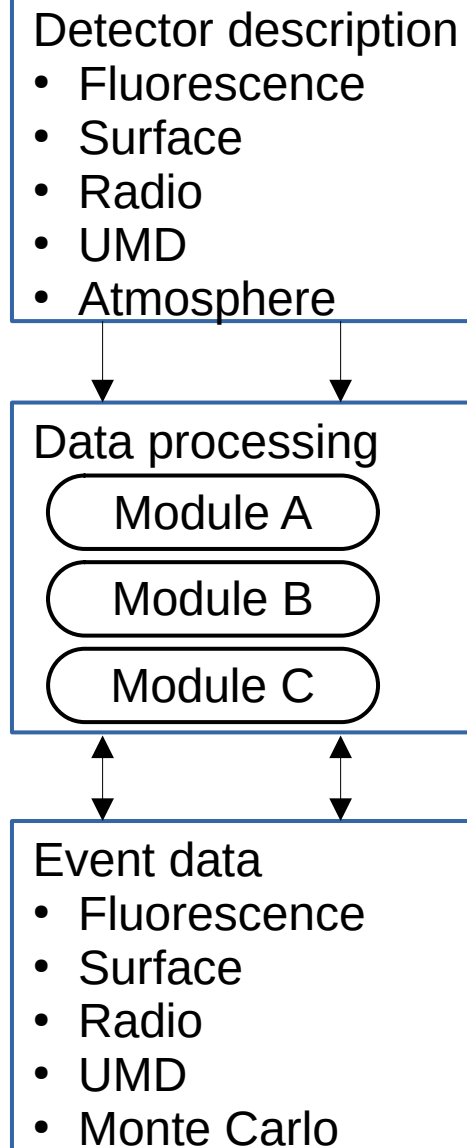
- Standard framework for:
  - Detector simulations
  - Calibration
  - Reconstruction
  - Data preparation
  - (some) analysis
- Flexible and extensible
- The Offline Framework stood the test of time. Could be extended to handle various extensions and a major upgrade
- Started over 20 years ago, first commit: 31<sup>st</sup> of January, 2003
  - CVS → SVN → git(lab)
  - C++98 → C++ 11/14 (→ C++ 17/20)
- Input:
  - Air shower simulation
  - Auger data
  - Offline format
- Output
  - Offline format
  - Auger Data Summary Trees (ADST) can be used stand-alone for further analysis



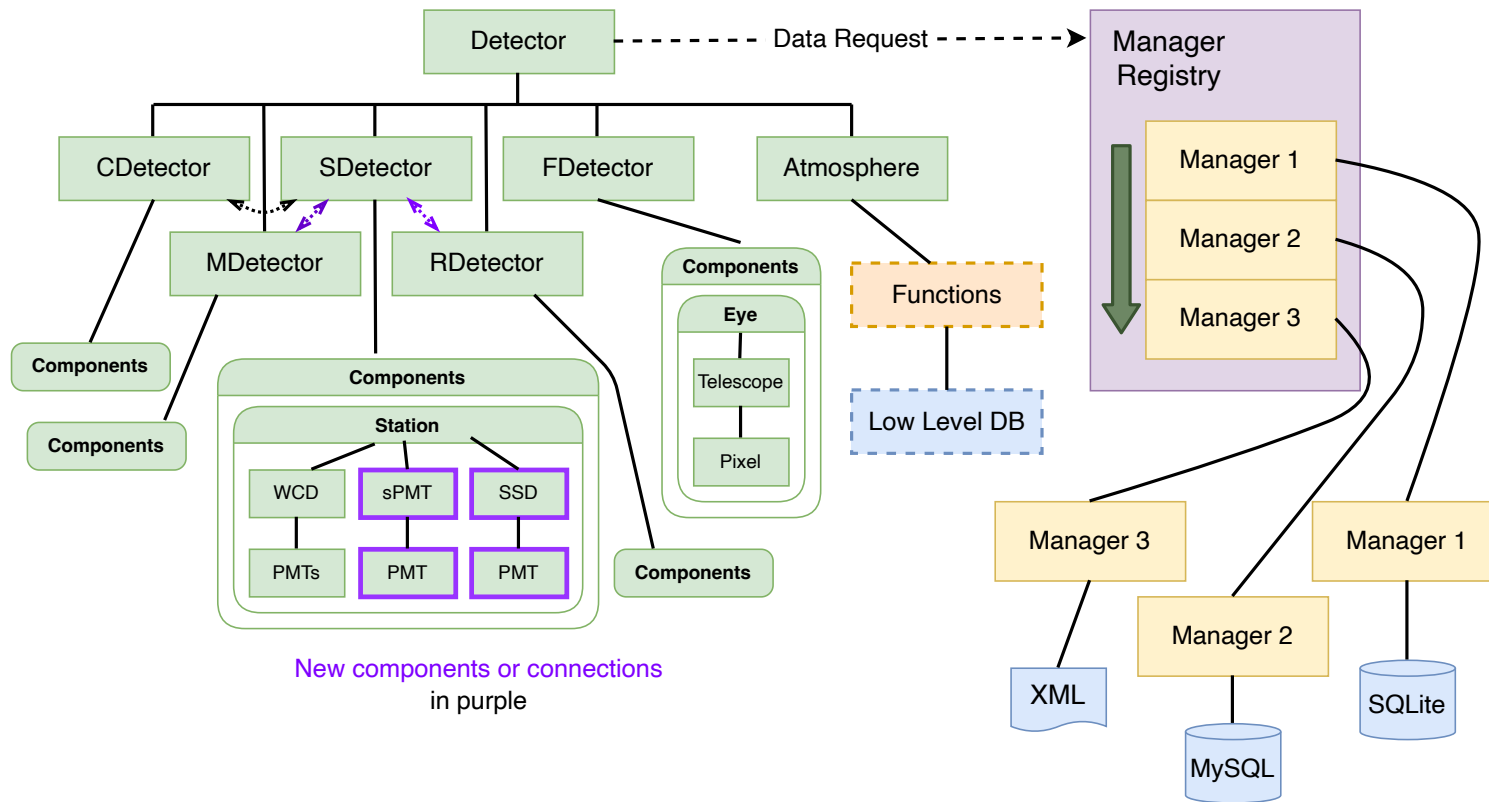
# Framework hierarchy

---

- Software components structured in non-cyclic hierarchy
  - access only to components lower in hierarchy
  - clean dependencies for parallel building
- Separate data holders from algorithms
- Framework: Detector, Event, RunControl
- Modules: simulation and reconstruction algorithms
- Utilities: used by modules
- Tools
- Validation tests
- ....

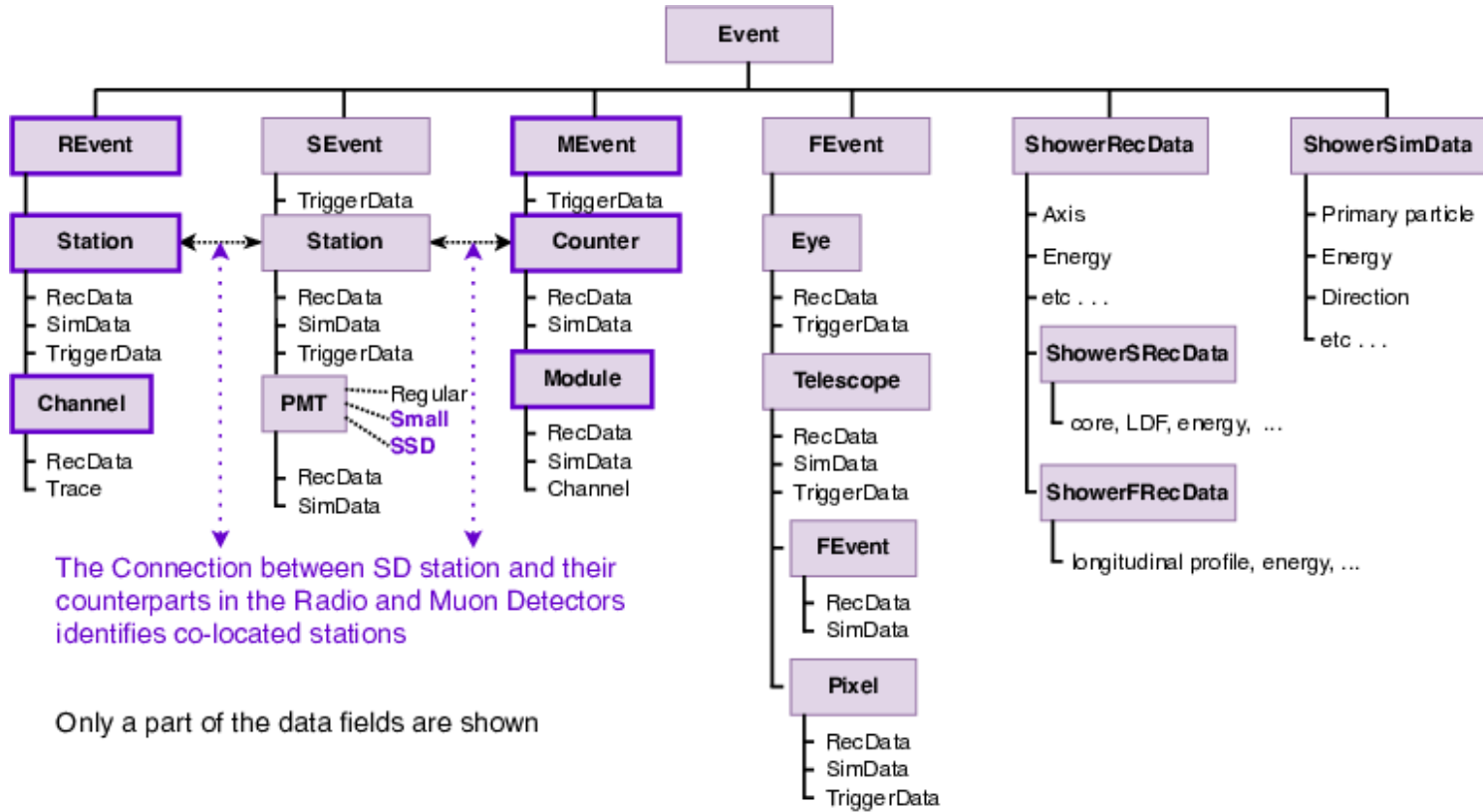


# The Detector



- Structure follows detector hierarchy
- Atmosphere is part of the detector
- Managers as abstraction for data access

# The event




The Connection between SD station and their counterparts in the Radio and Muon Detectors identifies co-located stations

Only a part of the data fields are shown


- Structure parallel to detector
- Contains raw, MC, and reconstructed data

# Some changes we had to do

- Upgraded electronics: different digitizer frequency (was hard-coded)
- Additional channels, PMTs (was hard-coded)
- Time delays:
  - PMT characteristics and change of FPGA
  - Have to match for comparison of Phase I and II
- In progress: Machine Learning support
  - Need to be able to use inference in regular modules
  - Evaluating Open Neural Network Exchange (ONNX)



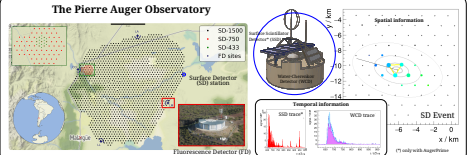
**Overview of Machine Learning Applications at the Pierre Auger Observatory**  
 Ezequiel Rodriguez<sup>a,b</sup> for the Pierre Auger Collaboration<sup>c</sup>  
 a. Instituto de Tecnologías en Detección y Astropartículas, San Martín, Argentina  
 b. Karlsruher Institut für Technologie, Karlsruhe, Germany  
 c. Observatorio Pierre Auger, Av. San Martín Norte 304, Malargüe, Argentina



**Abstract**

Running since 2004, the Pierre Auger Observatory remains the largest detector for extensive air showers induced by ultra-high-energy cosmic rays. The complex spatio-temporal information from shower footprints, comprised of particle arrival times and traces measured by water-Cherenkov detectors, is challenging to analyse with traditional methods but well-suited for machine learning based analyses. In this contribution, we provide an overview of the ML applications developed to leverage the high event statistics acquired by the Observatory.

**The Pierre Auger Observatory**



**Energy Estimation<sup>[1]</sup>**

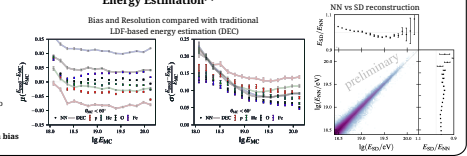
• CNN-based architecture

• Training dataset consists of detector simulations using p, He, O, and Fe

• Correction of detector ageing effects

• Calibration of network output with simultaneous SD and FD events due to mismatch between MC and data

• Potential reduction of composition bias



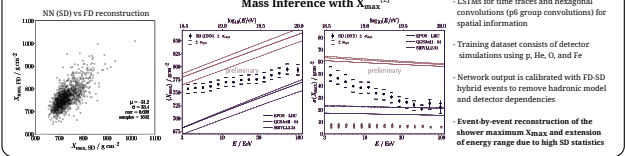
**Mass Inference with  $X_{max}$ <sup>[2]</sup>**

• LSTMs for time traces and hexagonal convolutions (p6 group convolutions) for spatial information

• Training dataset consists of detector simulations using p, He, O, and Fe

• Network output is calibrated with FD-SD hybrid events to remove hadronic model and detector dependencies

• Event-by-event reconstruction of the shower maximum  $X_{max}$  and extension of energy range due to high SD statistics



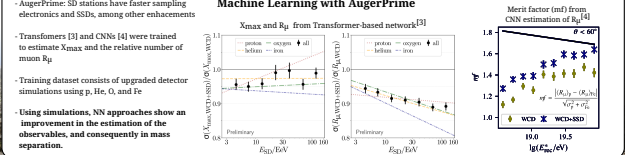
**Machine Learning with AugerPrime**

• AugerPrime: SD stations have faster sampling electronics and SSDs, among other enhancements

• Transformers [3] and CNNs [4] were trained to estimate  $X_{max}$  and the relative number of muon  $R_{\mu}$

• Training dataset consists of upgraded detector simulations using p, He, O, and Fe

• Using simulations, NN approaches show an improvement in the estimation of the observables, and consequently in mass separation.



[1] F. Eilwanger. PoS ICRC2023 (2023) 275.

[2] J. Glombitza. PoS ICRC2023 (2023) 278.

[3] N. Langner. PoS ICRC2023 (2023) 371.

[4] S. T. Hahn. PoS ICRC2023 (2023) 318.

spokespersons@auger.org



## RD reconstruction



# Calibration

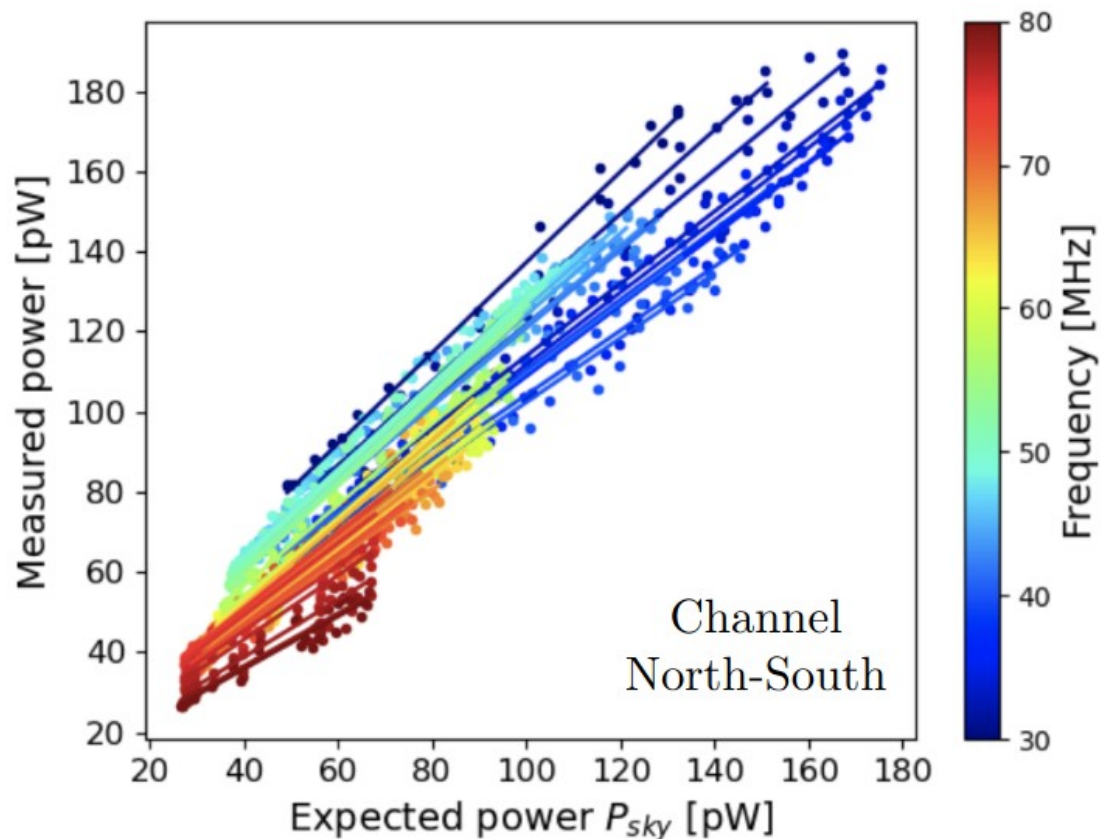
Relative directional calibration with reference antenna mounted on a drone

cf. PoS(ARENA2024)029



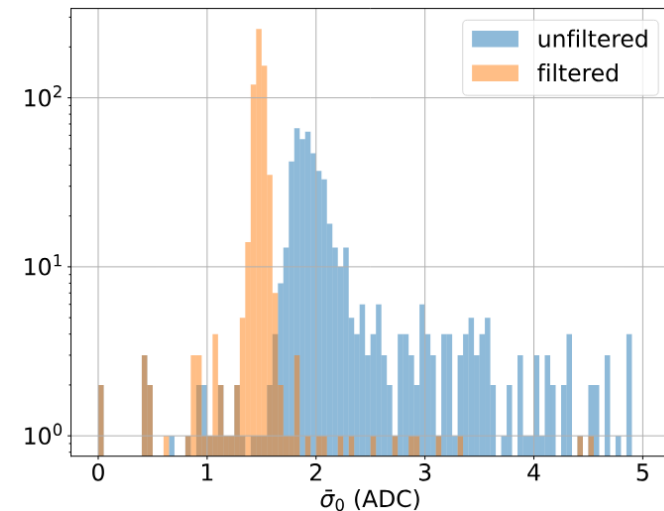
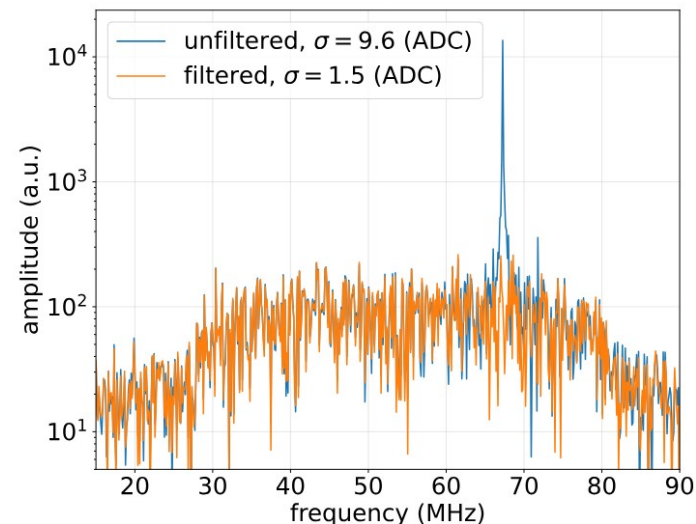
Absolute calibration using sky models of the galactic radio emission

cf. PoS(ARENA2024)030



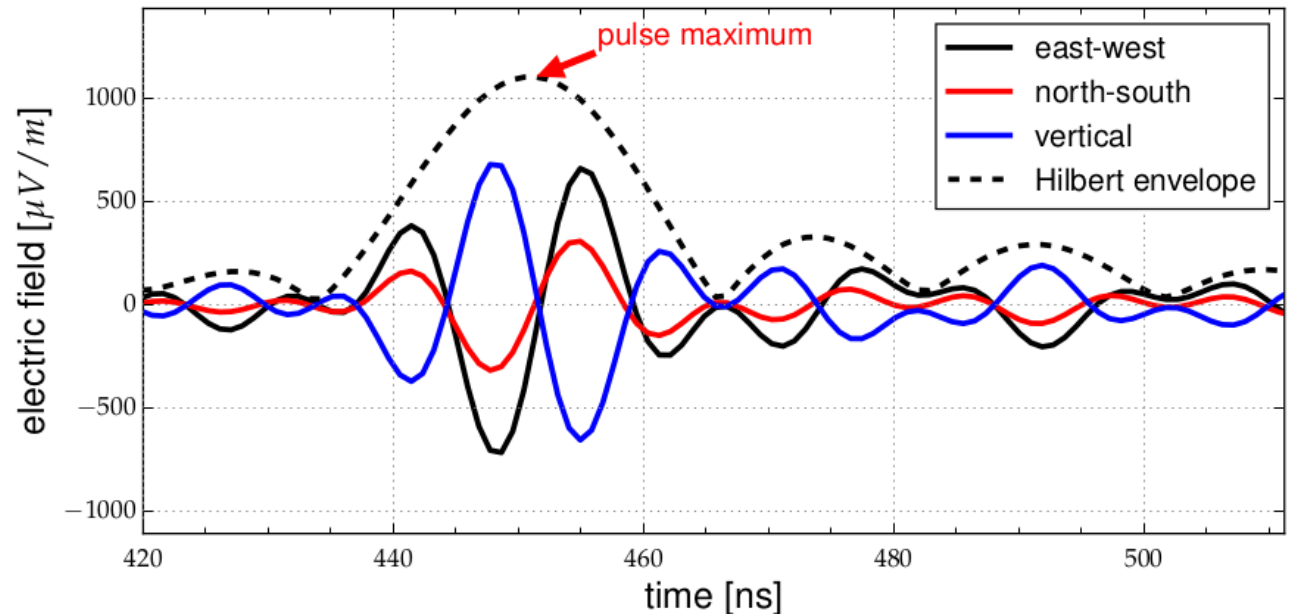
# Improved RFI removal

- Removing of strong narrowband RFI in the measurement challenging
  - Find frequency and phase of largest amplitude
  - Subtract corresponding sin wave in time domain
  - Continue as long as significant reduction of std or max number of removed frequencies
- Improved reconstruction of low-signal stations and increased number of low-energy events



# Station signal reconstruction

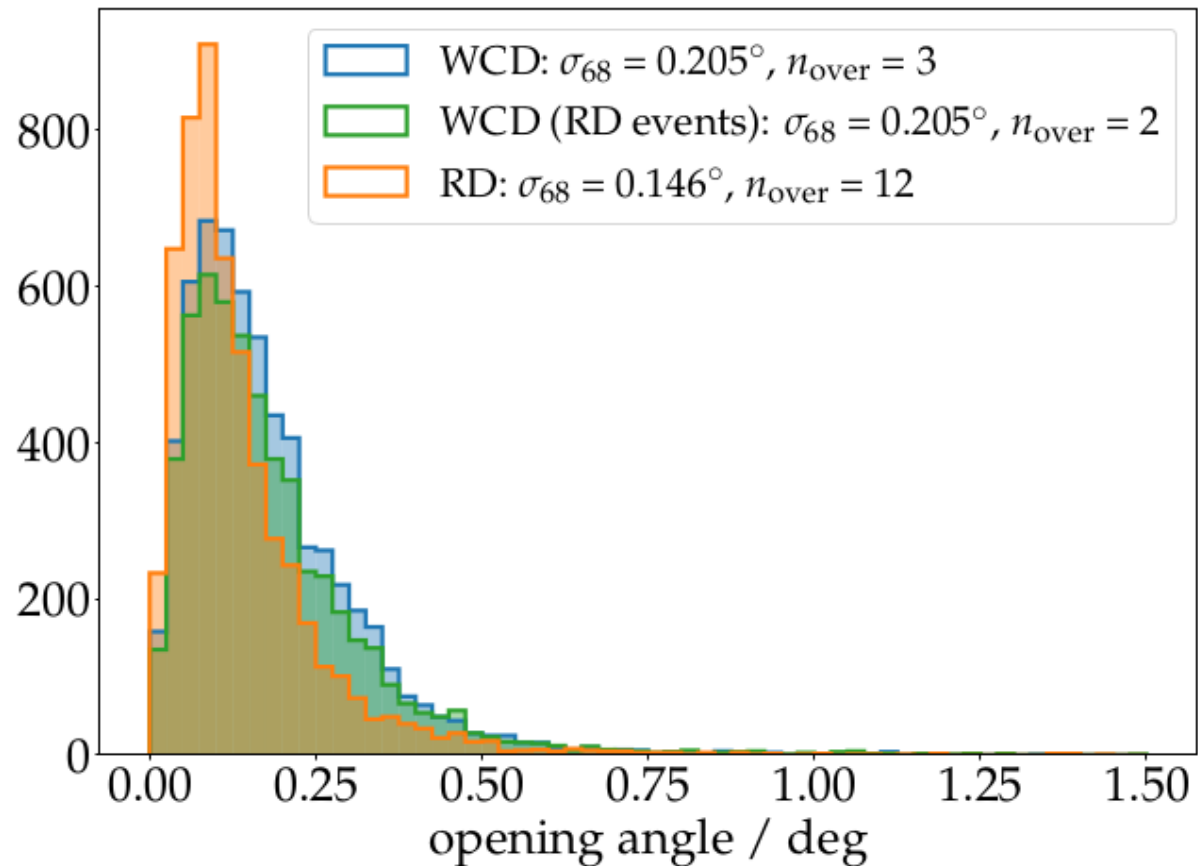
- 3D electric field reconstruction from 2 channel traces
- Signal arrival time given by maximum of the hilbert envelope
- Energy fluence estimated by the time integral in a signal window subtracting a noise contribution
- Working on an improved signal estimation in the presence of noise based on a robust and rigorous statistical background cf. [arXiv:2407.18654](https://arxiv.org/abs/2407.18654)
  - reduced bias and correctly estimated uncertainty
  - even at low SNR, no/lower SNR cut needed
  - lowering energy threshold for CR detection





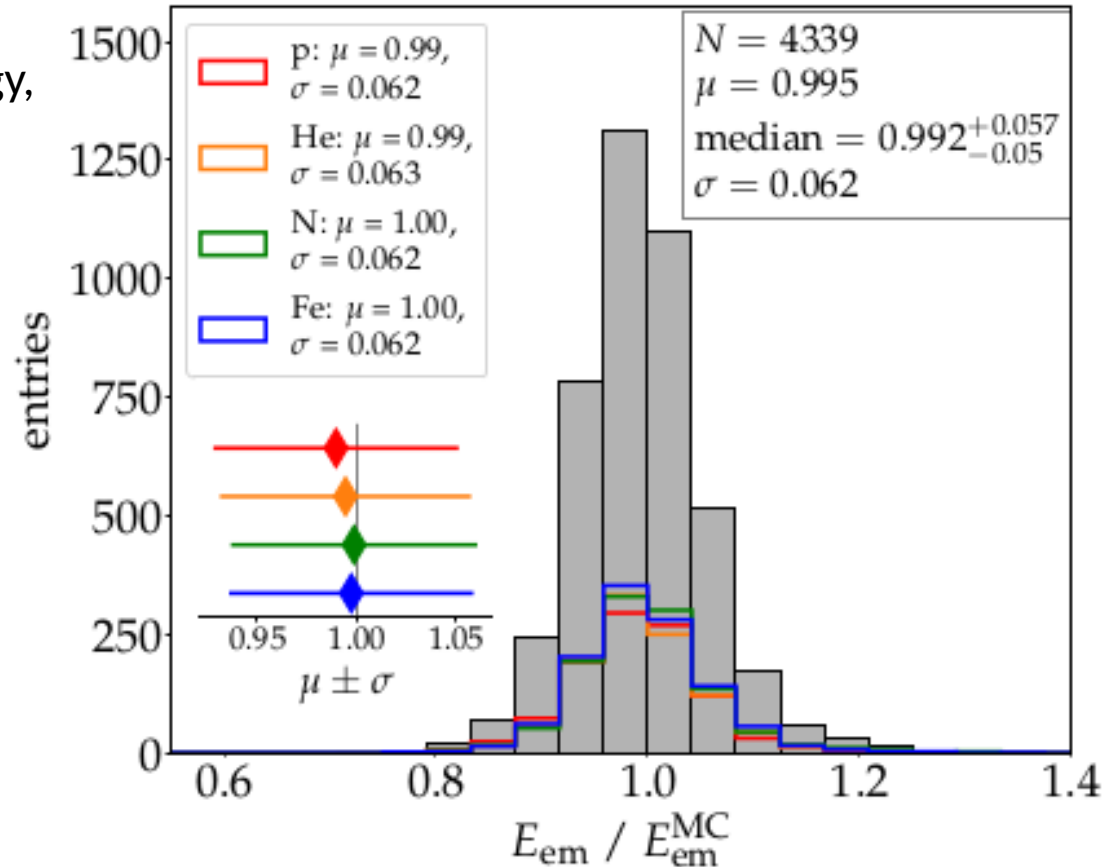
# Directional reconstruction

- Source of the radio emission close to Xmax, far away from detector for inclined showers  
→ spherical wavefront model
- Needs at least 4 signal stations, i.e. zenith  $> \sim 68^\circ$
- RD angular resolution of  $0.15^\circ$ , better than WCD resolution



# Lateral signal distribution

- Dedicated model for the lateral signal distribution on ground for inclined showers:  
cf. F. Schlüter JCAP01(2023)008
- Reconstruction of electromagnetic energy, distance to  $X_{\max}$ , and core
- 5 station needed for independent RD core estimation
- Bias-free reconstruction of  $E_{\text{EM}}$ , resolution of 6.2%









# Control Flow

- Application
  - sequence of steps, encapsulated as modules
- RunController
  - Configures sequence
  - Schedules execution
- CentralConfig configures
  - Detector
  - Modules
  - RunController

- Detector is read-only
- Event transports information between modules

