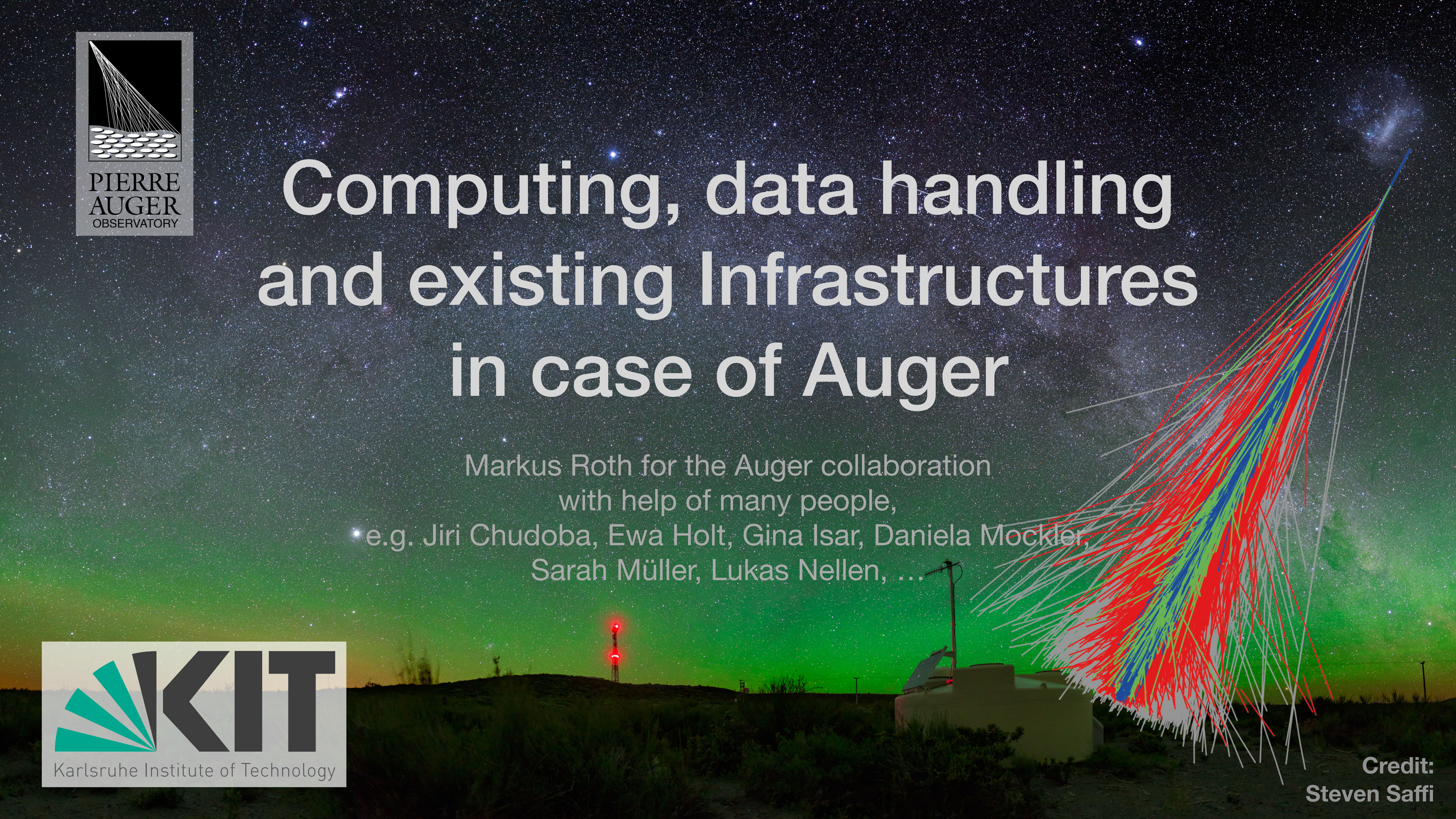
Computing, data handling and existing Infrastructures in case of Auger

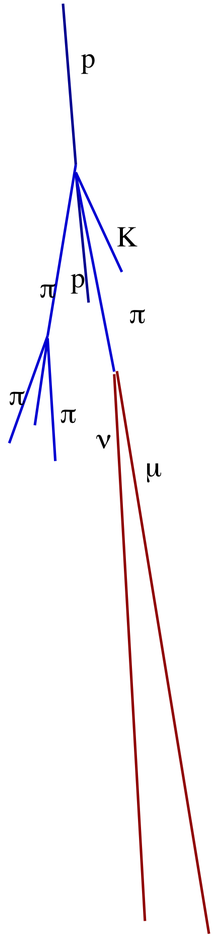
Markus Roth for the Auger collaboration
with help of many people,
e.g. Jiri Chudoba, Ewa Holt, Gina Isar, Daniela Mockler,
Sarah Müller, Lukas Nellen, …

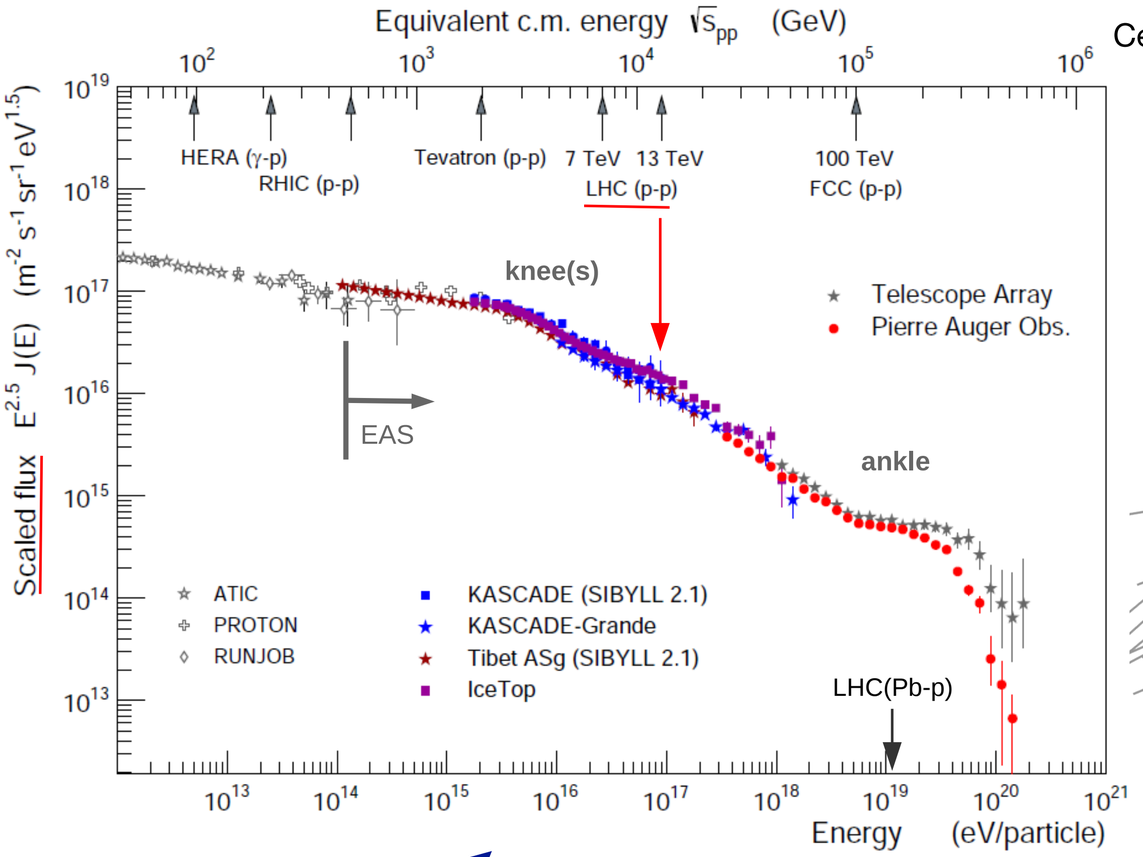
Karlsruhe Institute of Technology

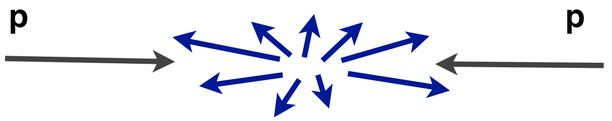


Credit:
Steven Saffi

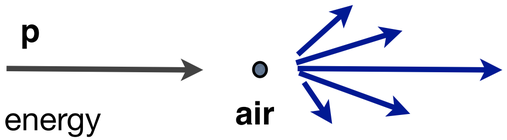
UHECRs: Flux of cosmic rays and interaction energies







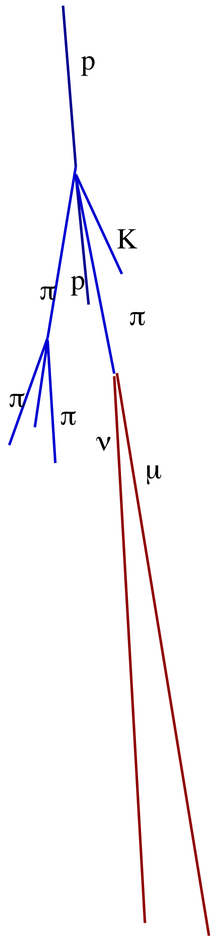
Center of mass energy

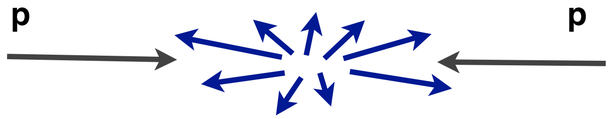


Laboratory energy

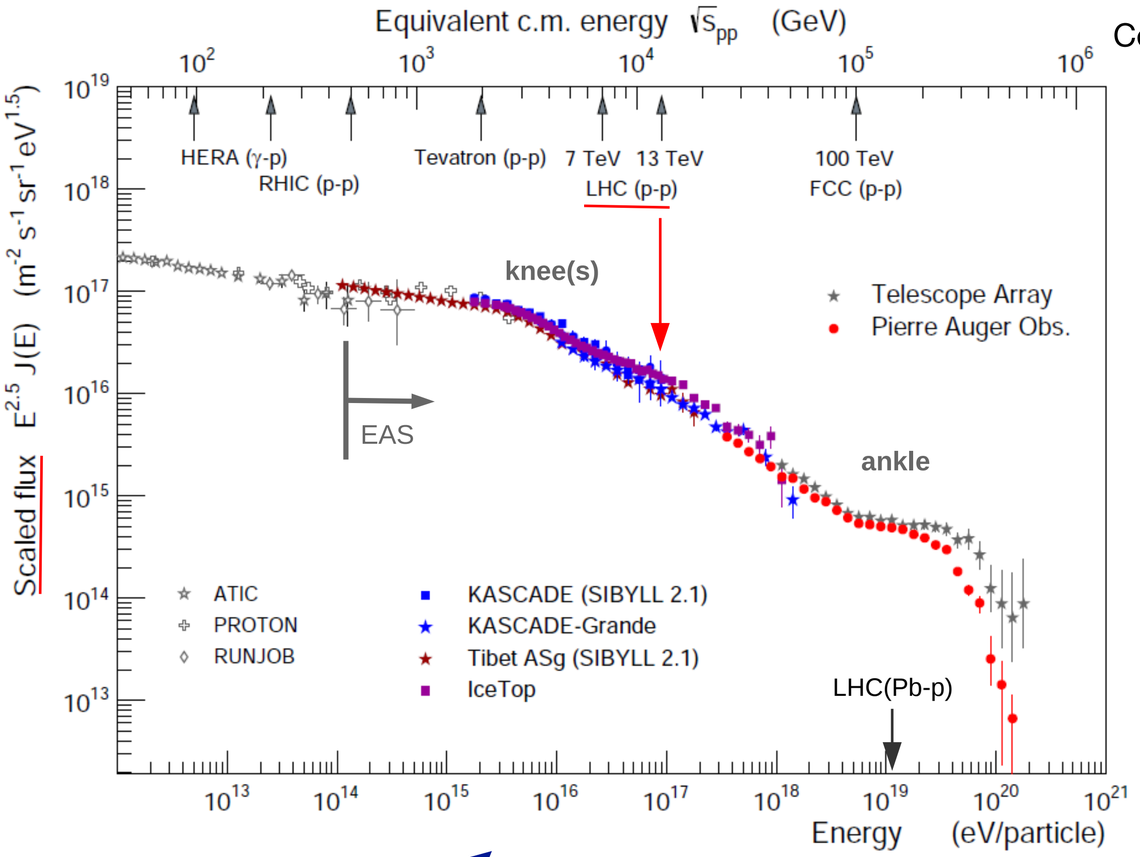
Courtesy R. Engel

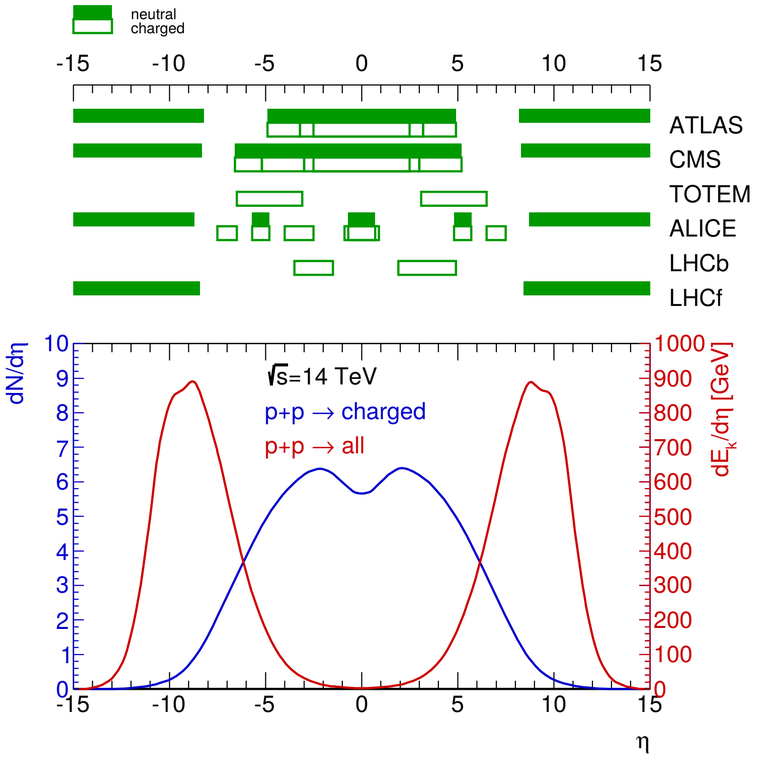
UHECRs: Flux of cosmic rays and interaction energies



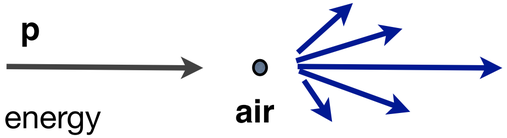


Center of mass energy





Courtesy R. Ulrich



Laboratory energy

Courtesy R. Engel

The Pierre Auger Observatory

Fluorescence detector (FD)
- 4 sites
  - 0-30°
  - **E>10^{18} eV**
- HEAT
  - 30°-60°
  - **E>10^{17} eV**

AERA array (radio)
- Grid with var. spacing
  - 17 km^2
  - >150 stations
  - **E>$10^{17.5}$ eV**

Surface detector array (SD)
- Grid of 1500 m
  - 3000 km^2
  - 1660 stations
  - **E>$10^{18.5}$ eV**
- Hexagon of 433 m
  - 0.5 km^2
  - 7 stations
  - **E>$10^{16.5}$ eV**
- Grid of 750 m
  - 24 km^2
  - 61 stations
  - **E>$10^{17.5}$ eV**

The Pierre Auger Observatory

Fluorescence detector (FD)

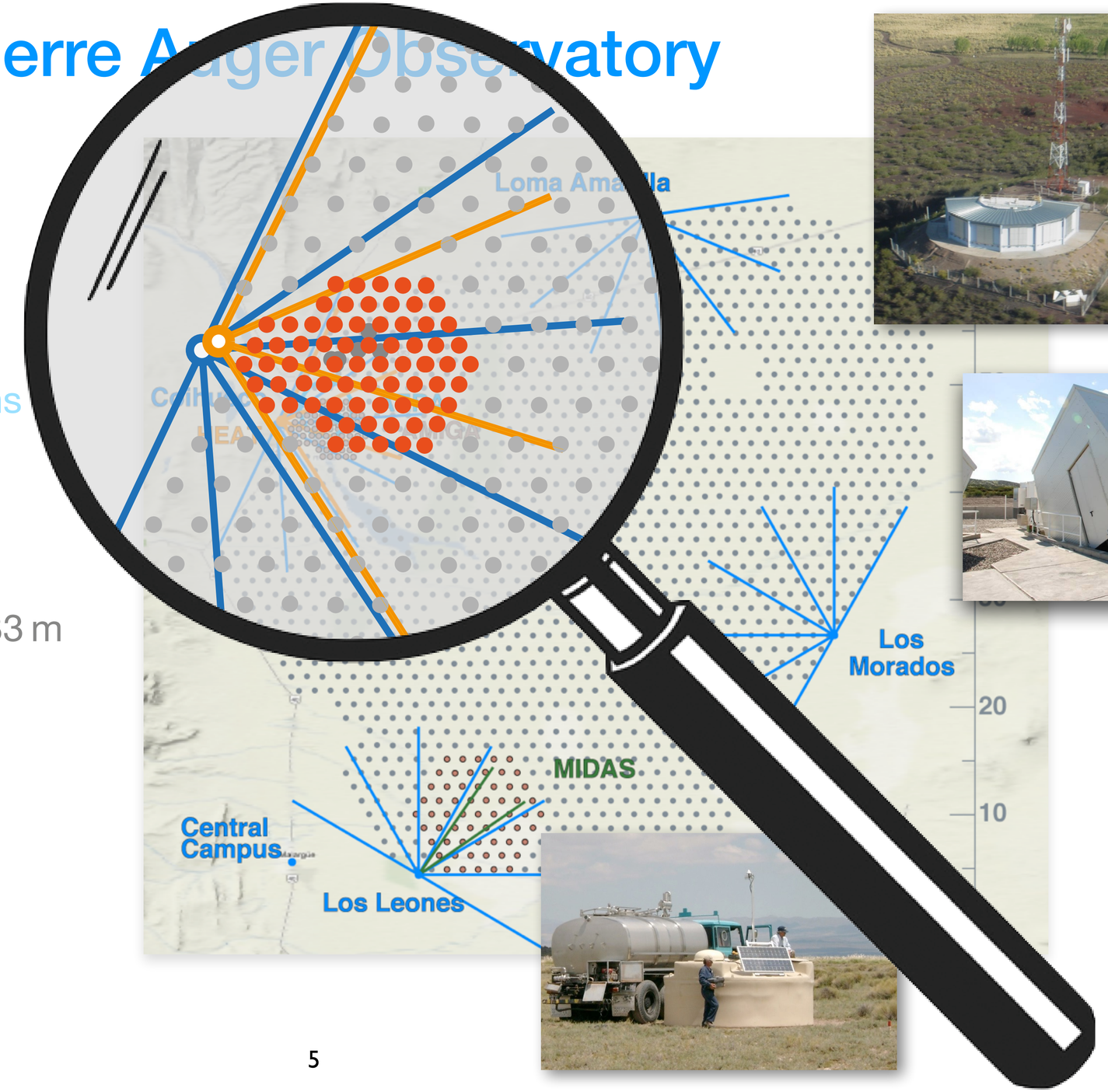
- 4 sites
  - 0-30°
  - **E>10^{18} eV**
- HEAT
  - 30°-60°
  - **E>10^{17} eV**

AERA array (radio)

- Grid with var. spacing
  - 17 km^2
  - >150 stations
  - **E>$10^{17.5}$ eV**

Surface detector array (SD)

- Grid of 1500 m
  - 3000 km^2
  - 1660 stations
  - **E>$10^{18.5}$ eV**
- Hexagon of 433 m
  - 0.5 km^2
  - 7 stations
  - **E>$10^{16.5}$ eV**
- Grid of 750 m
  - 24 km^2
  - 61 stations
  - **E>$10^{17.5}$ eV**



Hybrid detection

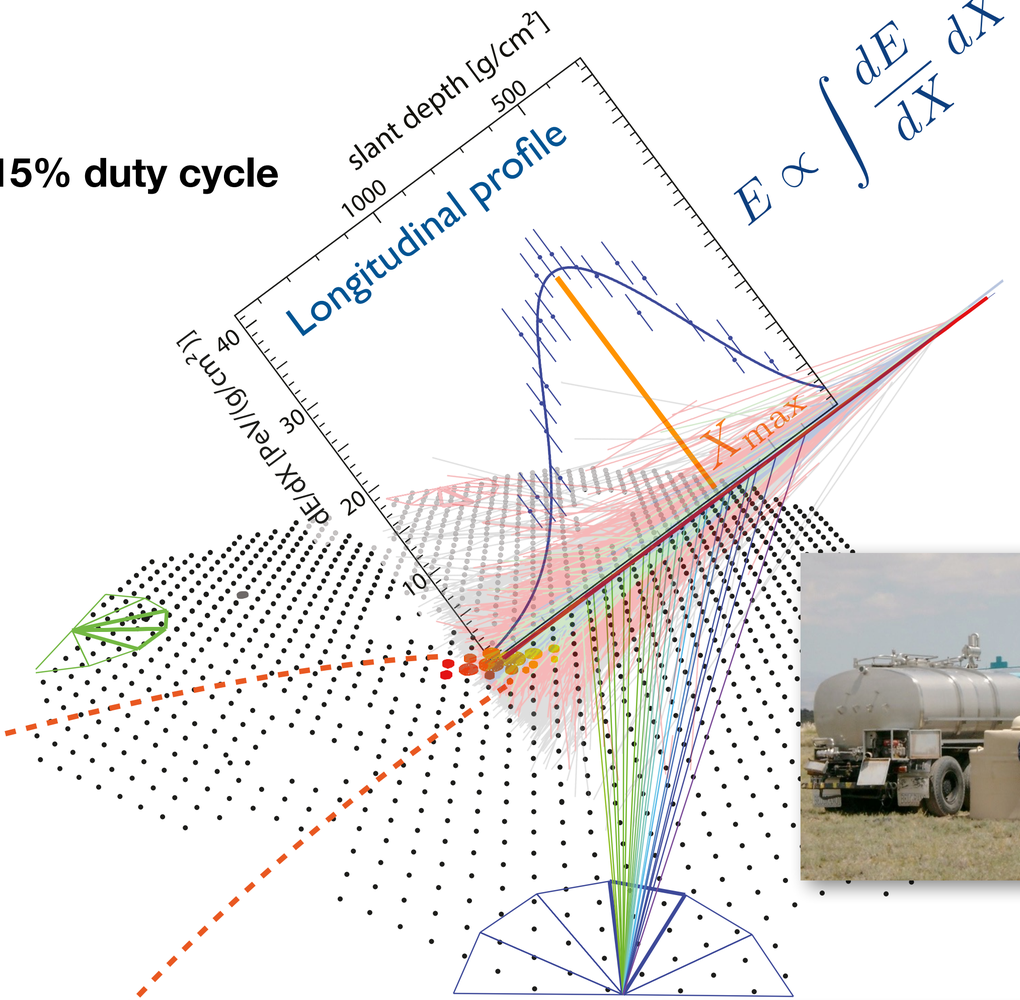
FD:

- calorimetric measurement of energy
- ca.15% duty cycle

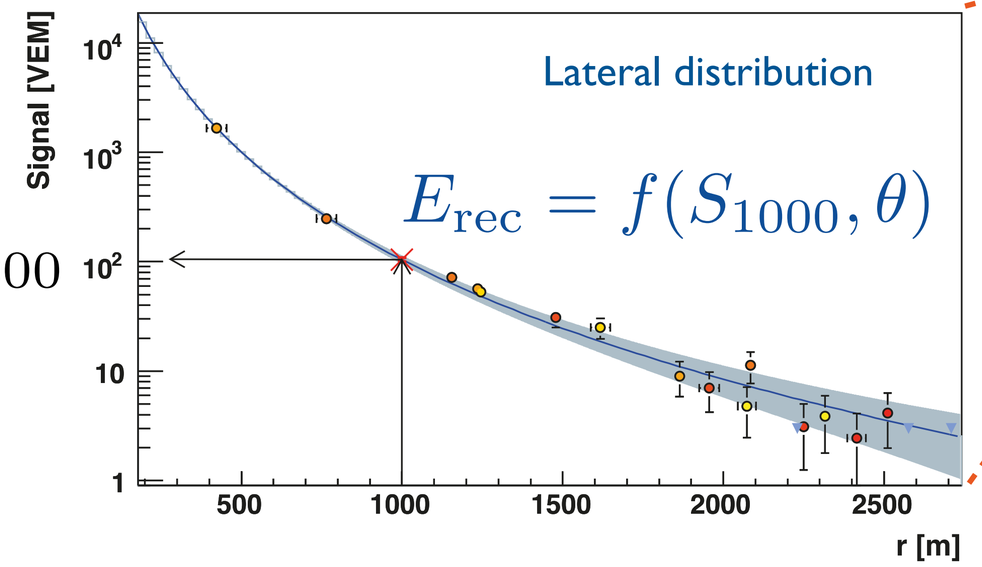
SD:

- data driven shape of LDF
- optimal distance at 1000 m
- ca. 100% duty cycle

15% duty cycle



100% duty cycle

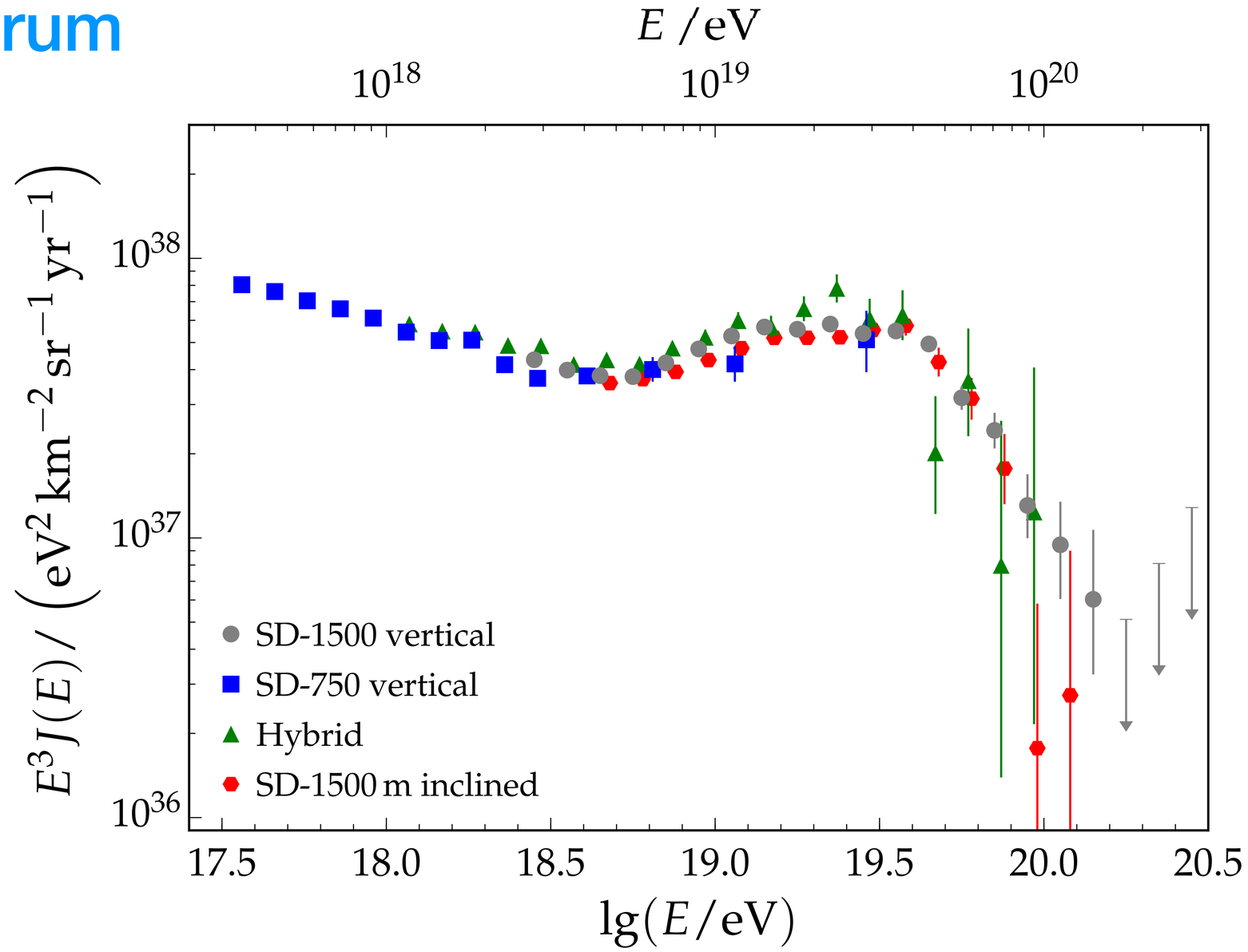


Example:
Event observed with
Auger Observatory

All particle spectrum

Systematic uncertainty of energy scale

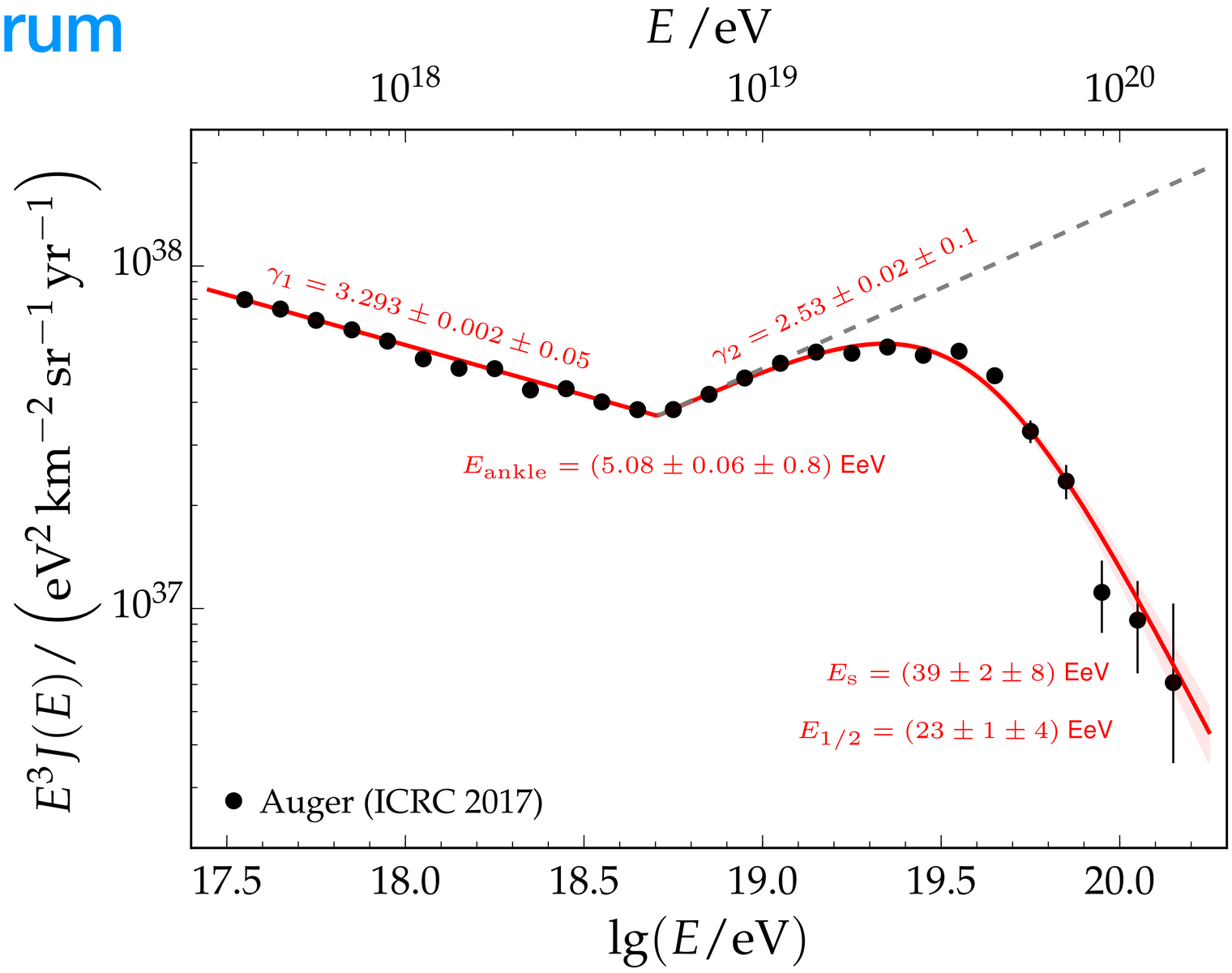
Auger: ΔE/E = 14%



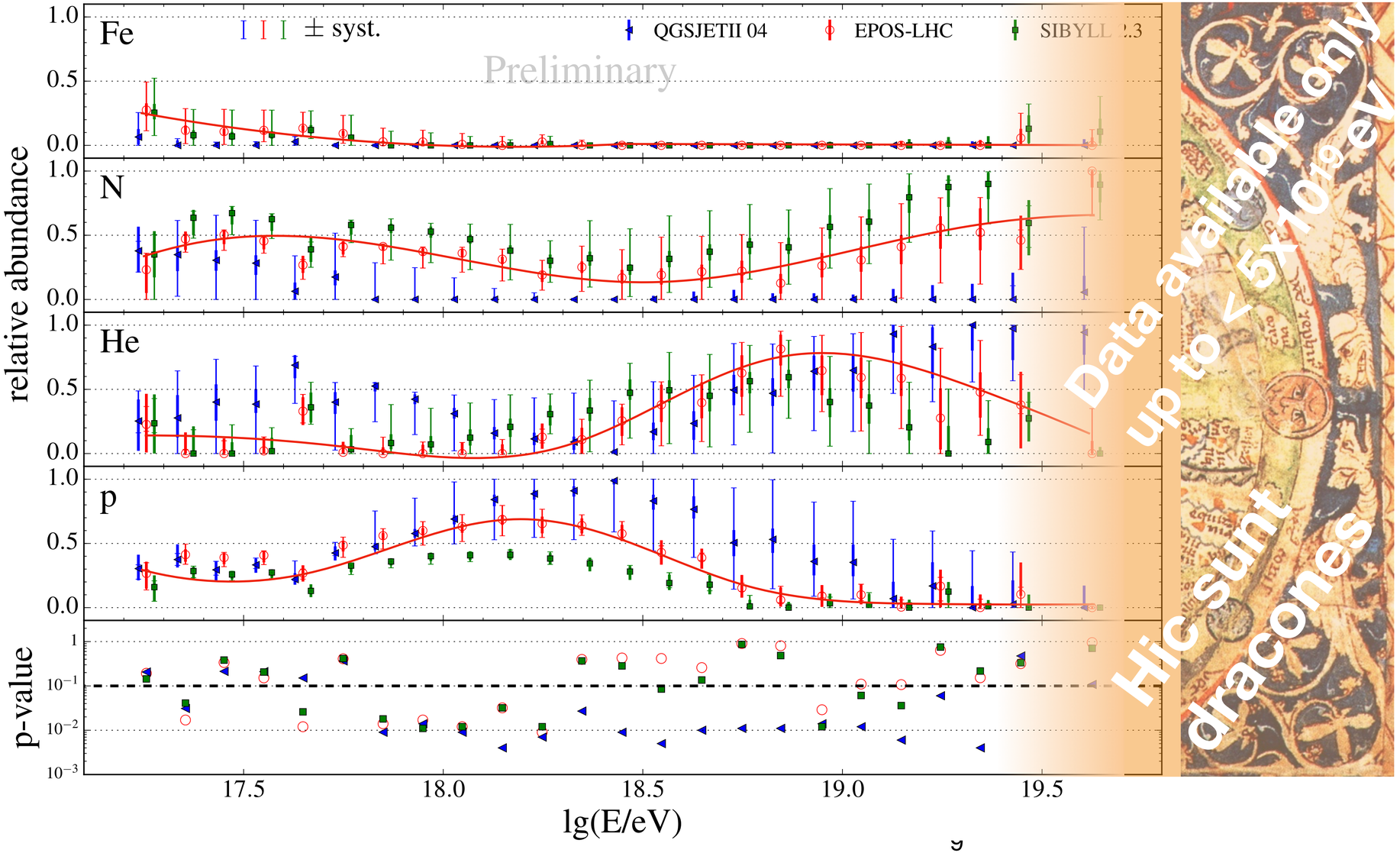
All particle spectrum

Systematic uncertainty of energy scale

Auger: ΔE/E = 14%



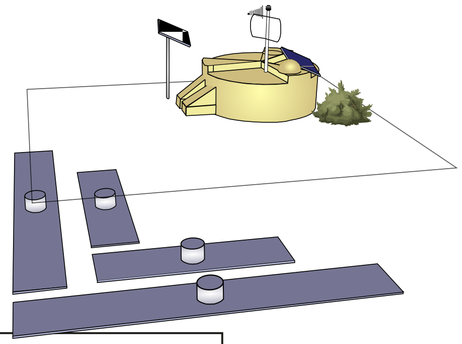
Mass composition at top of the atmosphere

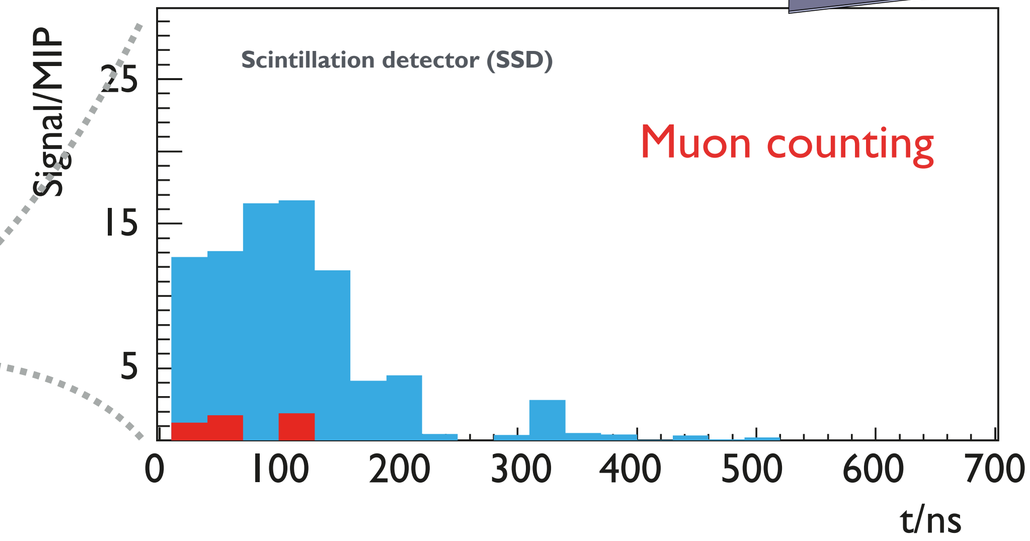


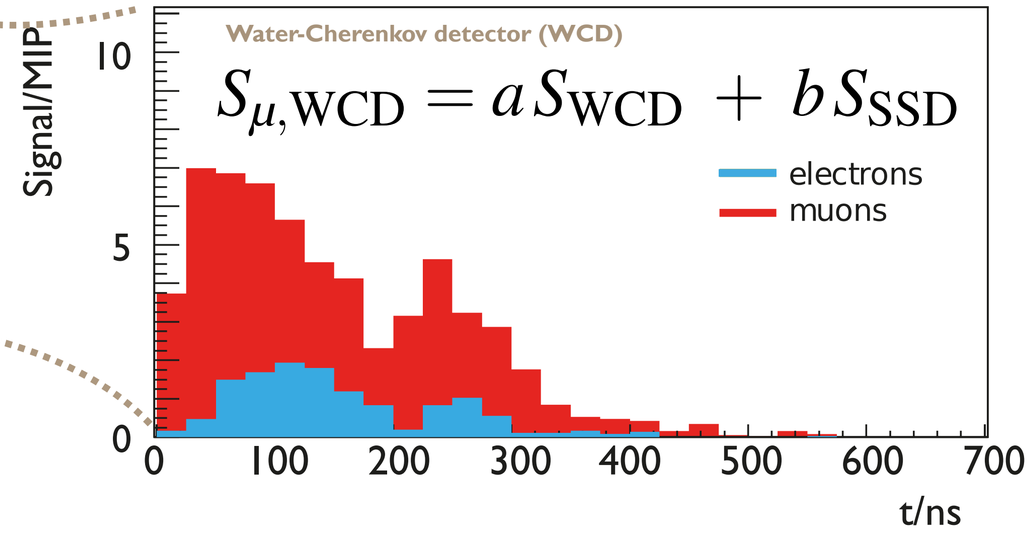
No composition data at and above onset of suppression

Surprises are expected to happen here!

AugerPrime: Detection principle







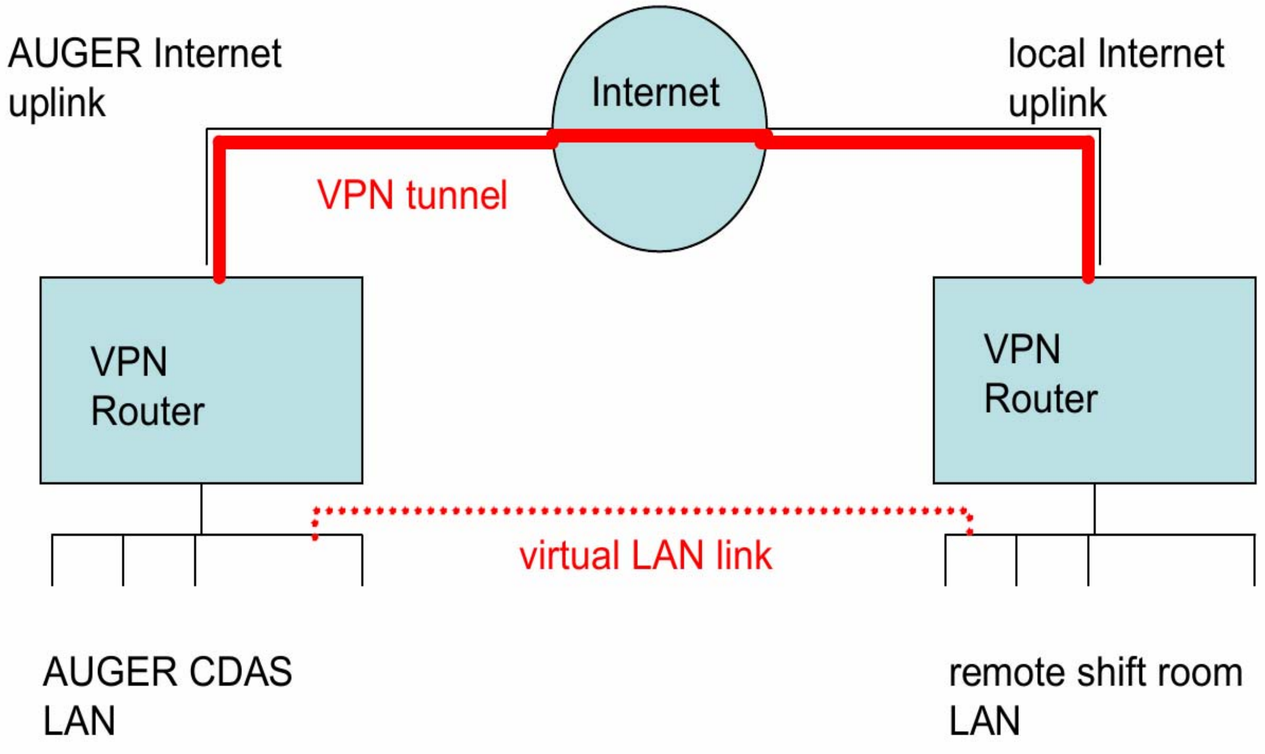
Signals of both detectors used to extract muons and electrons/photons

@100% duty cycle

Auger Access

Remote control

5 remote control rooms installed
7 more underway



Computing and data storage at CC Lyon

Download/upload to/from CC Lyon:
iRODS (integrated Rule-Oriented Data System)

Raw data transferred from Malargüe

Event merging done at and stored at CC Lyon

Personal accounts at Lyon for data processing and simulation purposes

PIERRE AUGER OBSERVATORY

Raw data storage

- **Raw data streams centrally stored at CC Lyon: 46 TB**
  - Fluorescence detector data: 20 TB/10 yr
  - Surface detector data: 14 TB/10 yr
  - Radio data: 22 TB/ 5 yr
  - AMIGA: < 1 TB/ 3 yr

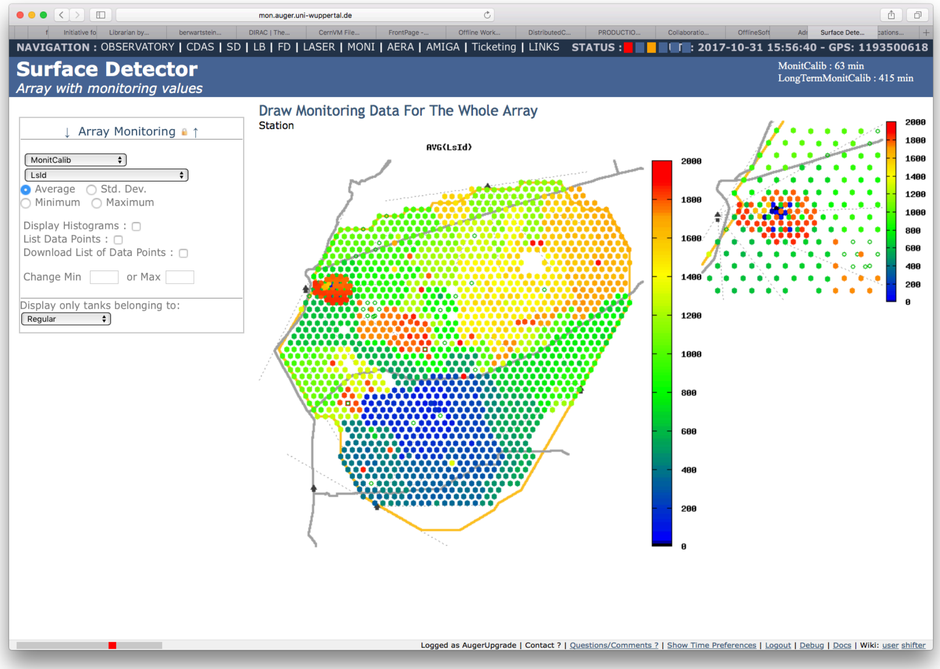
- **Offline event merging for various data streams at CC Lyon**
  - **Merging ~8 days after data taking**
  - **Fluorescence detector data: 1 TB**
    - Eye centric
    - Stereo
    - HEAT (+ Coiheco)
  - **Surface detector data: 4.3 TB**
    - *Vertical* data set $\theta \in [0,60°)$: 750 m & 1500 m array
    - *Inclined* data set $\theta \in [60°,80°]$: 750 m & 1500 m array
    - AERALet (AMIGA & Radio & SSD test bed)
  - **Radio data: 1.2 TB**
  - **AMIGA: 10 GB**

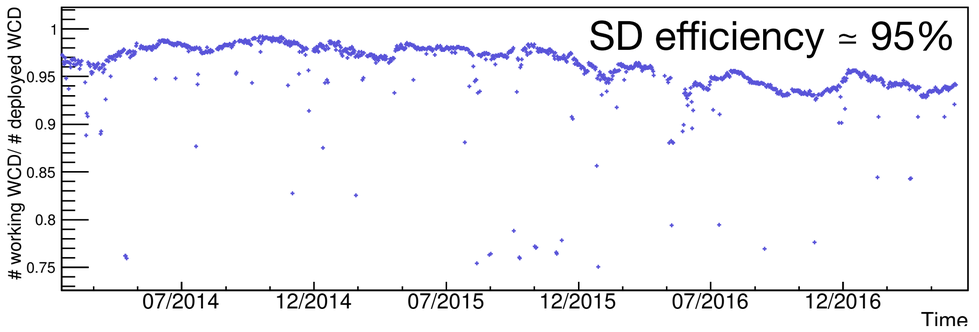
PIERRE AUGER OBSERVATORY

Monitoring data

- Amalgamation of different sources at Wuppertal, Aachen, KIT **5 TB**
  - Detector stations
  - Additional devices, e.g.
    - Weather stations
    - Field mill (b-field)
    - Lightning detectors
  - External data (GDAS)
- Collected in data streams at
  - Malargüe (PMS)
  - Lyon

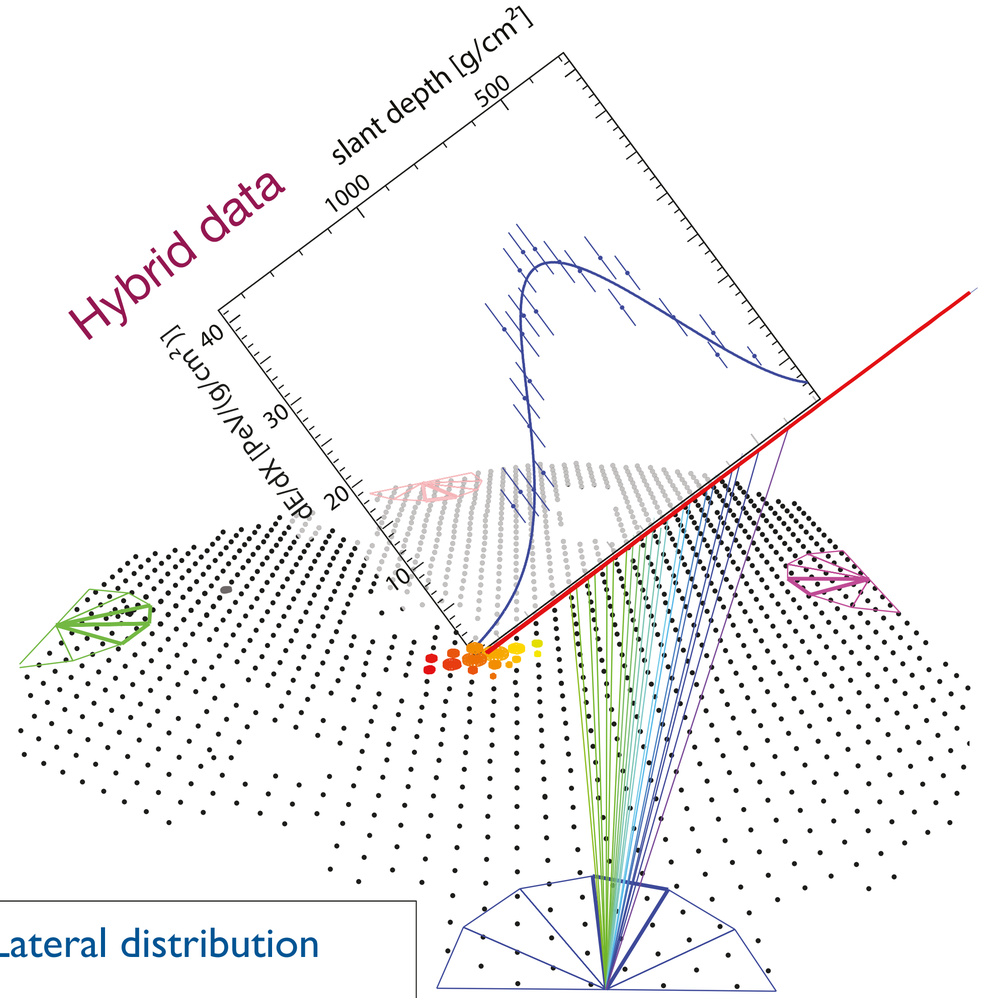
- Event-based data
  - Calibration
  - Detector status
  - …
- Non-event-based data
  - Weather stations
  - Atmosphere (LIDAR, LASER)
  - Batteries
  - Communication
  - …
- Accessible via Sqlite, Mysql, …

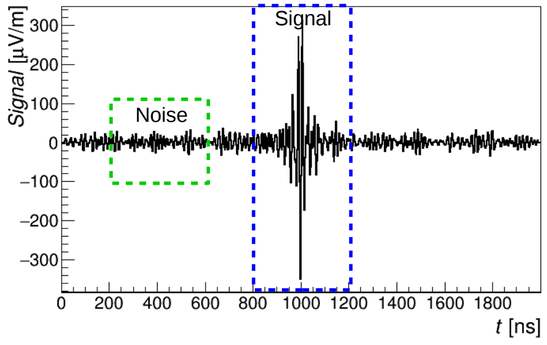




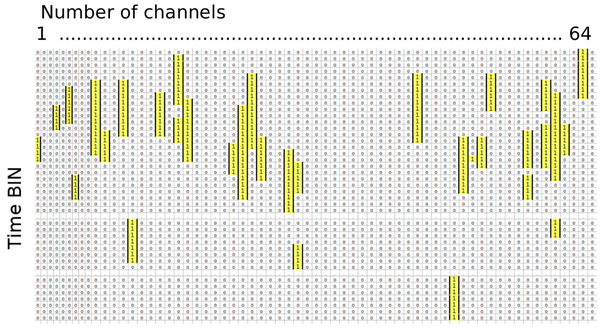
Reconstruction Offline

- **High-level information split in several data streams**
  - SD data: 15 GB/yr ~200 kB/event
  - Hybrid data : 27 GB/yr ~500 kB/event
  - Merged radio data: 3.5 TB/yr 2MB/event (slave mode)
  - AMIGA data: 10 GB/yr 4kB/event

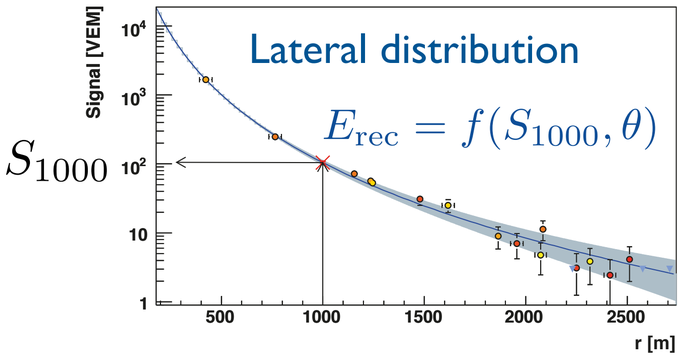




radio data



AMIGA data



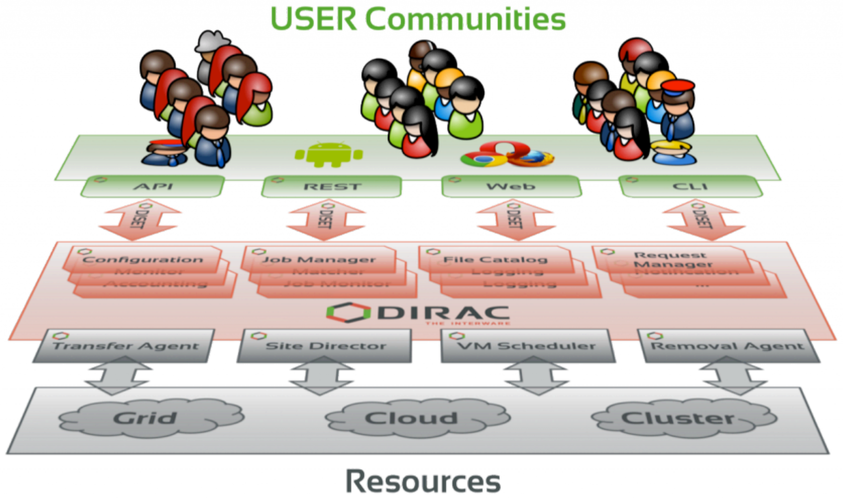
SD data

VO Auger

- **Computing on the GRID**
  - **VO organized by Czech collaborators**
  - **EGI – Operation Centers in Auger VO, 2014-2017**
    - NGI_CZ:prague_cesnet_lcg2,praguelcg2
    - NGI_DE:FZK_LCG2,RWTH_Aachen
    - NGI_FRANCE:GRIF,IN2P3-CC,M3PEC,OBSPM
    - NGI_IBERGRID:BIFI,CESGA,IFCA-LCG2,NGC-INGRID-PT
    - NGI_IT:INFN-CATANIA,INFN-LECCE,INFN-T1
    - NGI_NL:NIKHEF-ELPROD
    - NGI_RO:RO-ISS-13
    - NGI_SI:ARNES,SiGNET
    - ROC_LA:CBPF,ICN-UNAM
  - **Total number of jobs 1,008,604**
  - **Total of 700 TB of files produced**
- **DIRAC (Distributed Infrastructure with Remote Agent Control) INTERWARE for job submission**
- **CVMFS distributed, global, read-only filesystem used for software distribution**
- **Docker/Container tested**

MetaCentrum

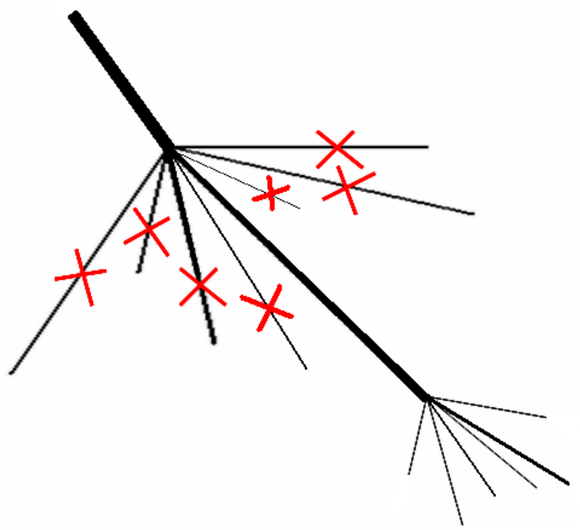
EGI

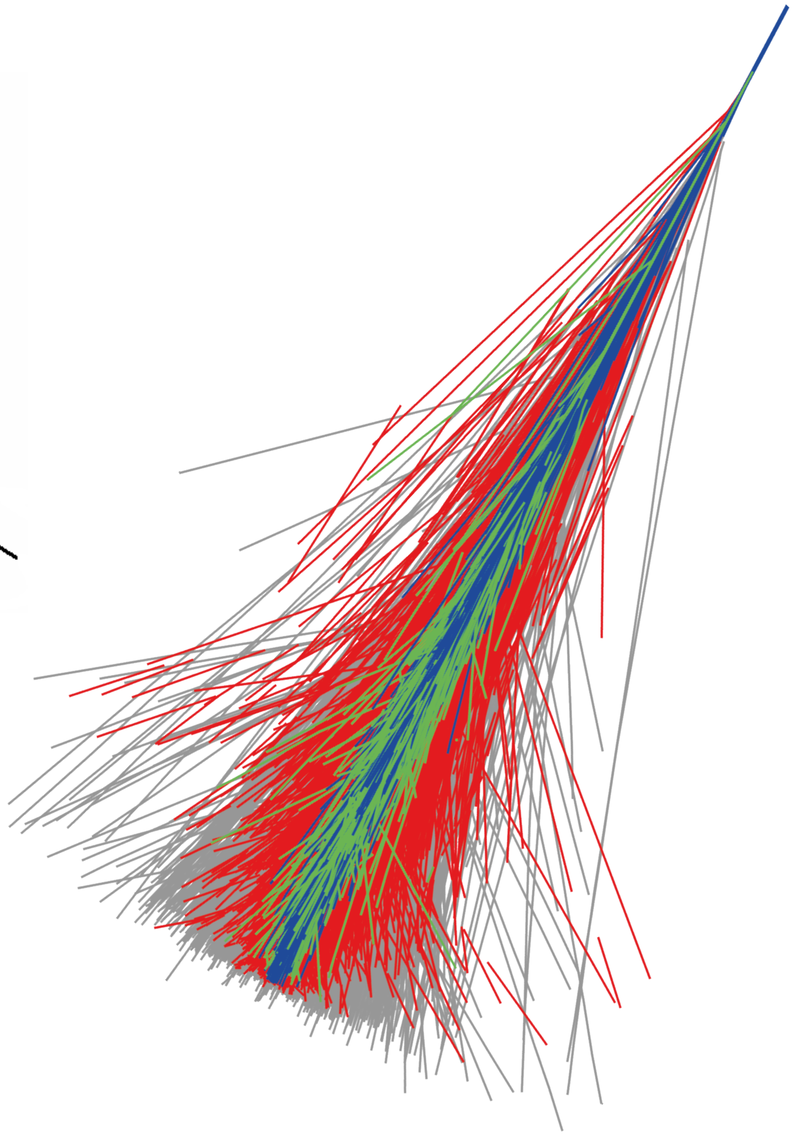


CernVM File system

Simulation production Offline

- **Typical simulation sets**
  - Contain some 10,000 CORSIKA showers (# of events for SD ~6,000,000)
  - 1 hadronic interaction generator
  - ~2GB/shower at 10^{19}eV
  - Thinning algorithm
  - Total 5-10 TB

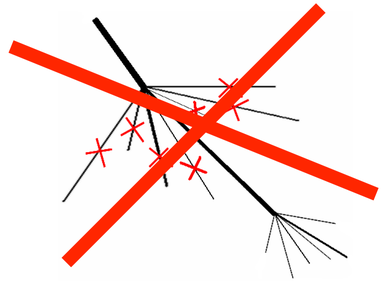


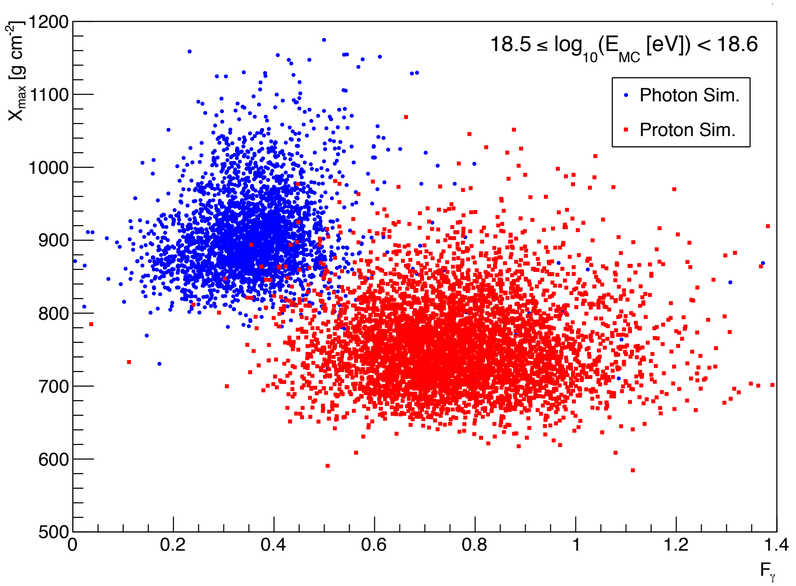


- **Or sets simulating**
  - time-dependent
    - atmosphere
    - detector status
  - Using
    - CONEX (fast 1D simulation of long. development)
      Millions of showers with no detailed particle info at ground
    - CoREAS (Radio extension of CORSIKA)
      Computational expensive
  - some 100 GB to 1TB

The need for more simulations

- **AugerPrime: Wealth of new data (SSDs, AMIGA, Radio, FD)**
  - Top-down simulation for high-quality events (1000/yr), e.g.
    - 100 p and 100 Fe showers per event
      (exercised already with Radio and Hybrid subsets)
  - Unthinned showers
    - Get detailed view of different components, e.g.
    - Proton: 10^{19} eV (120 CPU days)
      - Total: 1.2 TB
      - Sampling at detector positions only: 20 GB
  - „*Exotic*" searches: Neutrinos and photons
    - More stringent cuts and less systematics
      by scanning the tails

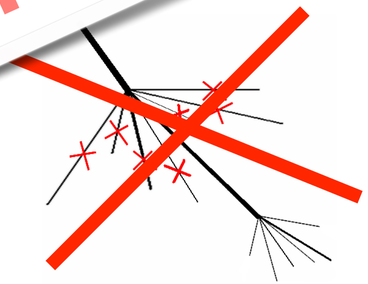


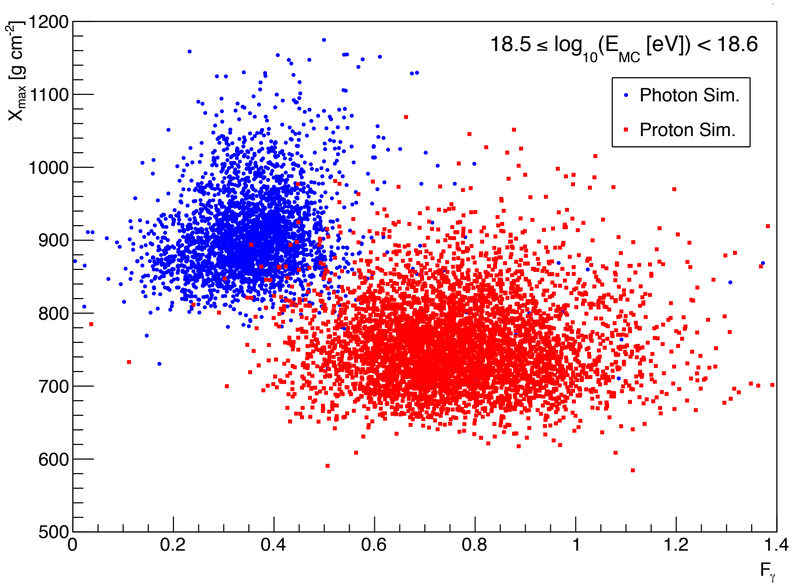


The need for more simulations

- **AugerPrime: Wealth of new data (SSDs, AMIGA, Radio, FD**
  - Top-down simulation for high-quality events (1000
    - 100 p and 100 Fe showers per event
  - Unthinned showers
    - Get detailed view
    - Proton: 10
      - T
      - ons only: 20 GB
  - „*Exot* Neutrinos and photons
    - Mo ngent cuts and less systematics by scanning the tails

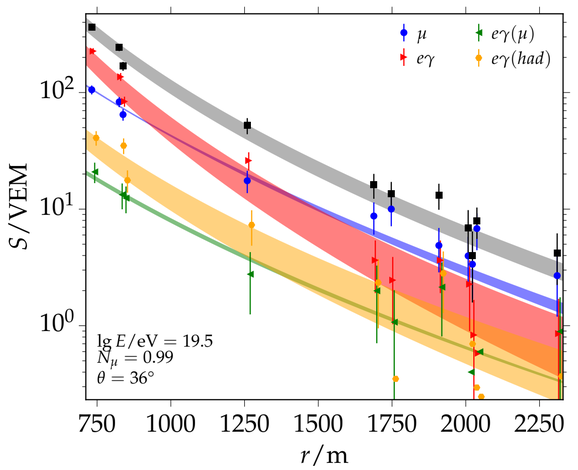
10-20 times more simulations and thus additional resources needed

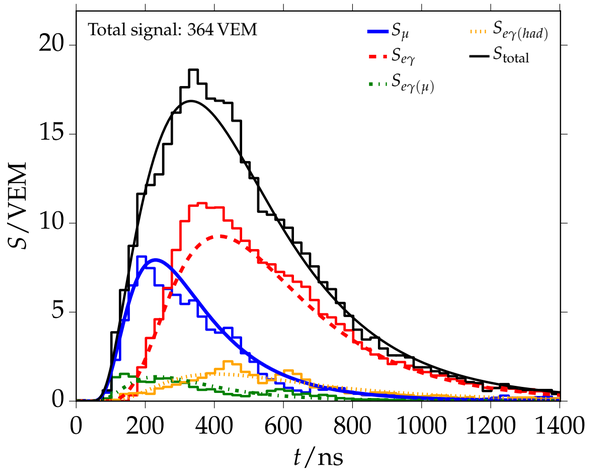




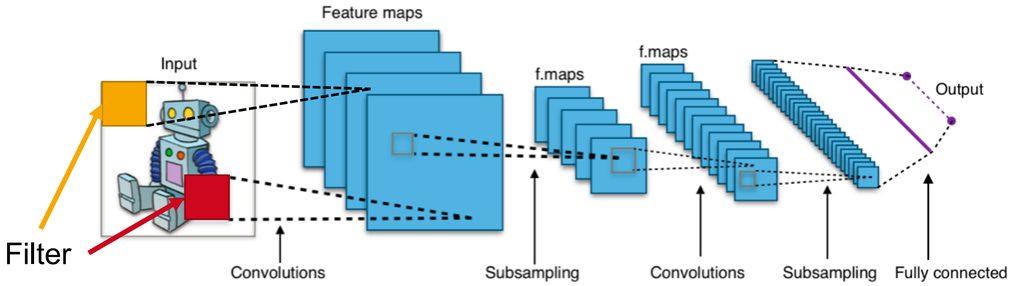
New analysis tools and paradigms

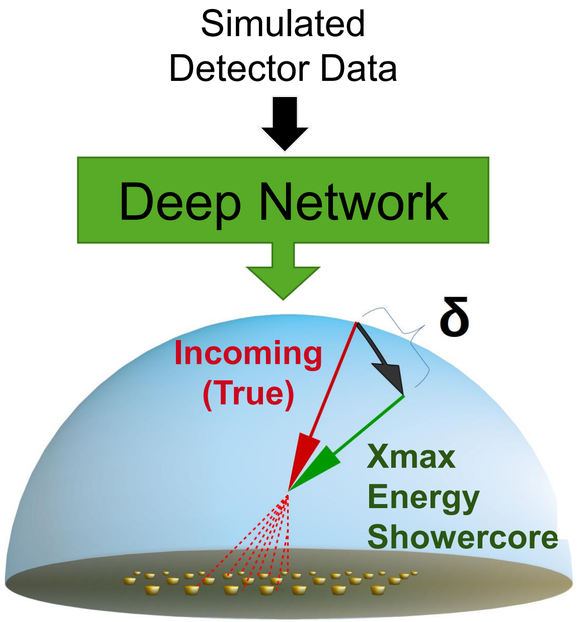
- Physics based approaches investigating details of the shower development: Shower universality



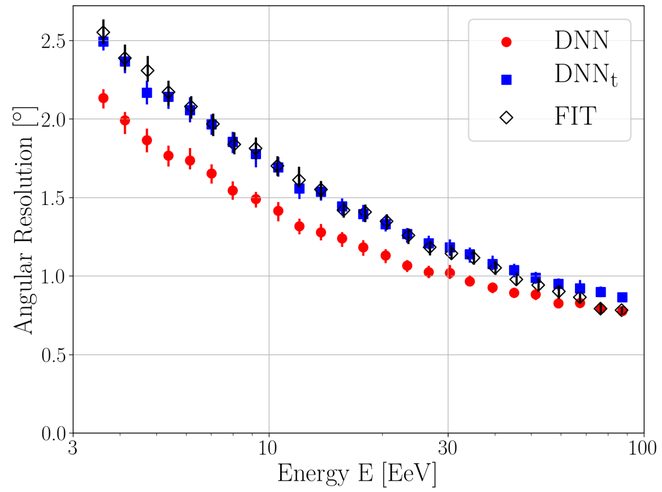


- Feature extraction using Machine learning / deep learning: GPU farms



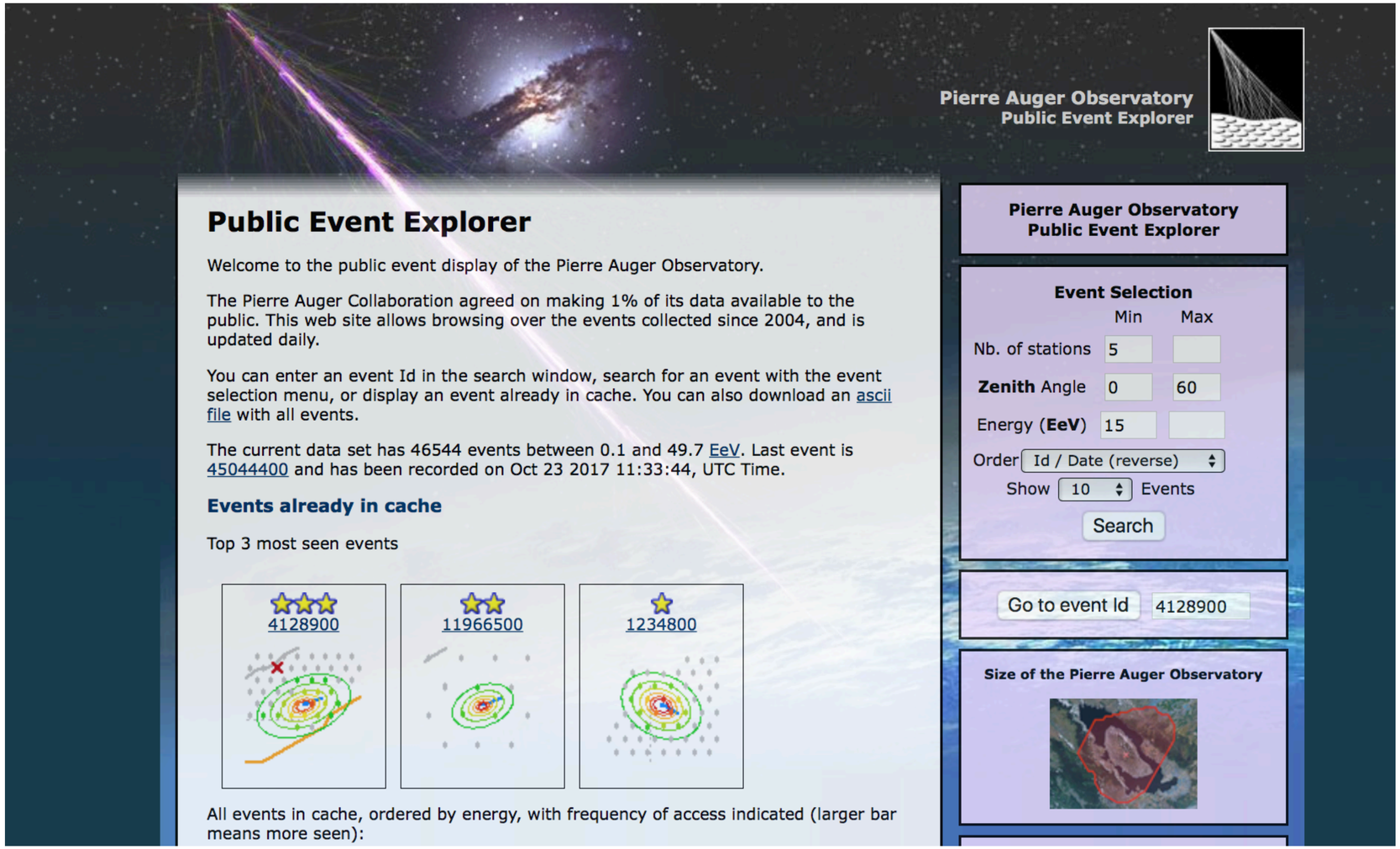


M. Erdmann, J. Glombitza, D. Walz



(Public) data

- 1% of data publicly available
- Publication policy under revision
- Simulations will have to be provided too
- Current data storage driven by active groups of the collaboration
- No long-term data preservation concept established beyond life-cycle of the experiment (WG in place)



Summary

- Resource saving approach: ~50 TB of raw data
  Will at least double until 2025

- Simulations on the GRID: 700 TB so far

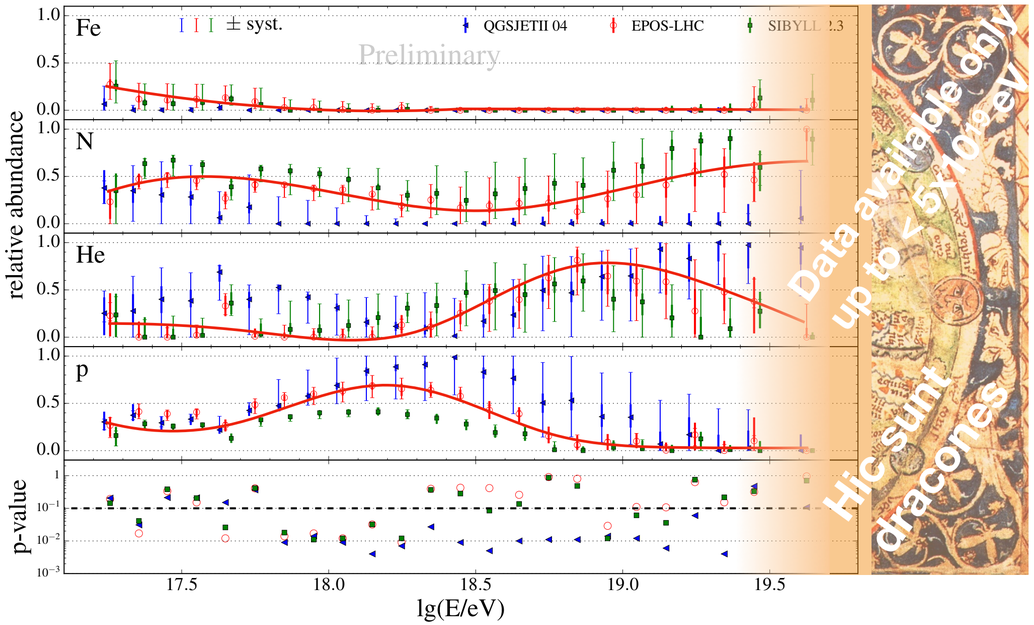
- New level of data quality
  ⇒ New simulation and storage requirements

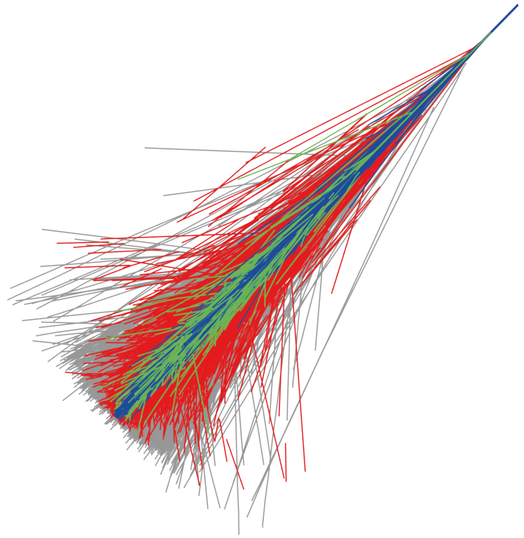
  (increased demand by factor of 10-20)

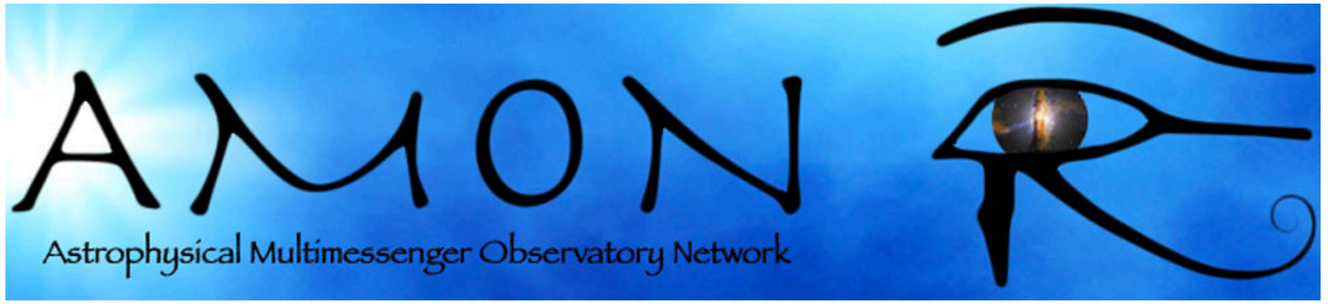
- New hardware resources needed
  (e.g. GPUs)

- No long-term data preservation concept yet

- Funding for post-docs and phd-students
  in new fields of work needed
  (currently even lack of funding for data manager)





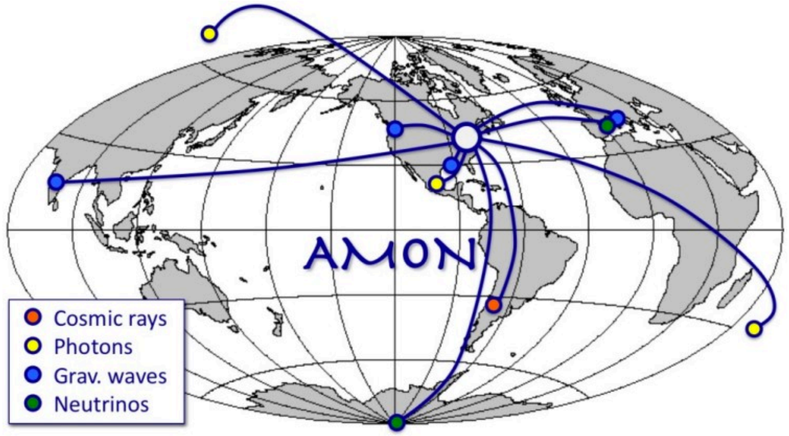


- Real-time correlation analysis of the high-energy signals across all known astronomical messengers – photons, neutrinos, cosmic rays, and gravitational waves – for *triggering* and *follow-up*

- Auger: Cluster of CRs in close temporal and spatial proximity (<3° and <100 s) Bears signature of burst of neutral particles produced in astrophysical transient.

  ⇒ AMON follow-up alert

- Auger will provide AMON with measure of "photon-likeness" of each event transmitted to AMON in real-time.



The Astrophysical Journal Letters, 848:L12 (59pp), 2017 October 20 https://doi.org/10.3847/2041-8213/aa91c9

OPEN ACCESS

Multi-messenger Observations of a Binary Neutron Star Merger

LIGO Scientific Collaboration and Virgo Collaboration, Fermi GBM, INTEGRAL, IceCube Collaboration, AstroSat Cadmium Zinc Telluride Imager Team, IPN Collaboration, The Insight-Hxmt Collaboration, ANTARES Collaboration, The Swift Collaboration, AGILE Team, The 1M2H Team, The Dark Energy Camera GW-EM Collaboration and the DES Collaboration, The DLT40 Collaboration, GRAWITA: GRAvitational Wave Inaf TeAm, The Fermi Large Area Telescope Collaboration, ATCA: Australia Telescope Compact Array, ASKAP: Australian SKA Pathfinder, Las Cumbres Observatory Group, OzGrav, DWF (Deeper, Wider, Faster Program), AST3, and CAASTRO Collaborations, The VINROUGE Collaboration, MASTER Collaboration, J-GEM, GROWTH, JAGWAR, Caltech-NRAO, TTU-NRAO, and NuSTAR Collaborations, Pan-STARRS, The MAXI Team, TZAC Consortium, KU Collaboration, Nordic Optical Telescope, ePESSTO, GROND, Texas Tech University, SALT Group, TOROS: Transient Robotic Observatory of the South Collaboration, The BOOTES Collaboration, MWA: Murchison Widefield Array, The CALET Collaboration, IKI-GW Follow-up Collaboration, H.E.S.S. Collaboration, LOFAR Collaboration, LWA: Long Wavelength Array, HAWC Collaboration, The Pierre Auger Collaboration, ALMA Collaboration, Euro VLBI Team, Pi of the Sky Collaboration, The Chandra Team at McGill University, DFN: Desert Fireball Network, ATLAS, High Time Resolution Universe Survey, RIMAS and RATIR, and SKA South Africa/MeerKAT
(See the end matter for the full list of authors.)

Received 2017 October 3; revised 2017 October 6; accepted 2017 October 6; published 2017 October 16