Universität Siegen

HEP Experimental High Energy Physics

CPPS Center for Particle Physics Siegen

Scientific Computing for Astroparticle Physics / Auger

Marcus Niechciol
Center for Particle Physics Siegen, University of Siegen

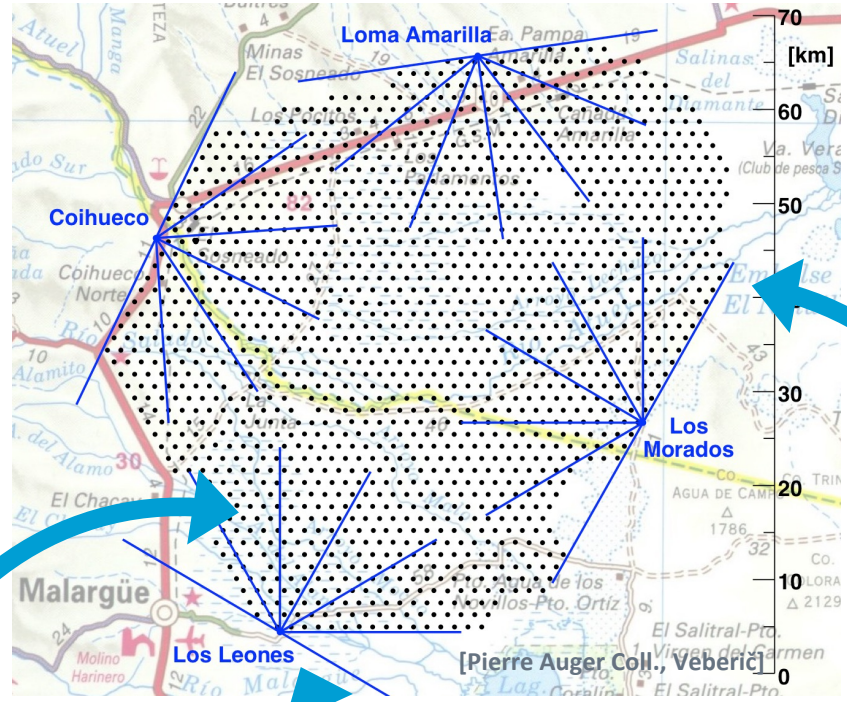
CPPS Retreat
15 February 2024

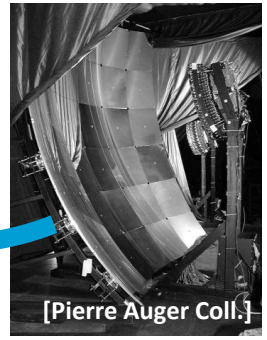
Photo by S. Saffi (Univ. of Adelaide)

Pierre Auger Observatory

- **Surface Detector (SD)**
  - ~1660 water Cherenkov detector stations, covering about 3000 km^2
- **Fluorescence Detector (FD)**
  - Four FD stations with 27 telescopes
- Data taking started in **2004**
- Detector upgrade (**AugerPrime**) ongoing

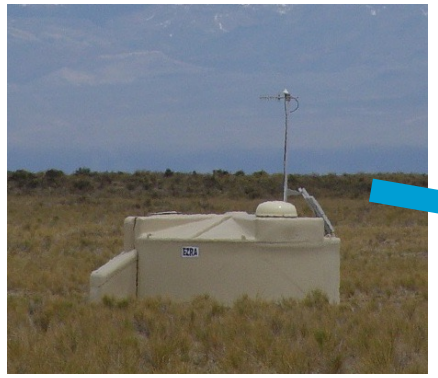
[Pierre Auger Coll., NIM A 798 (2015) 172]





Auger Data Stream

Array trigger rate: 0.1 Hz



Data requested from the stations as needed via radio to build an event

Trigger rate per FD building: 0.012 Hz

Data stored directly at the FD buildings and copied once per day

Malargüe Central Campus

Copy all raw data once per day

1 year of (merged) raw data: 300 GB

Lyon Computing Center + mirrors

Siegen

Transfer as needed

Monitoring data, calibration…

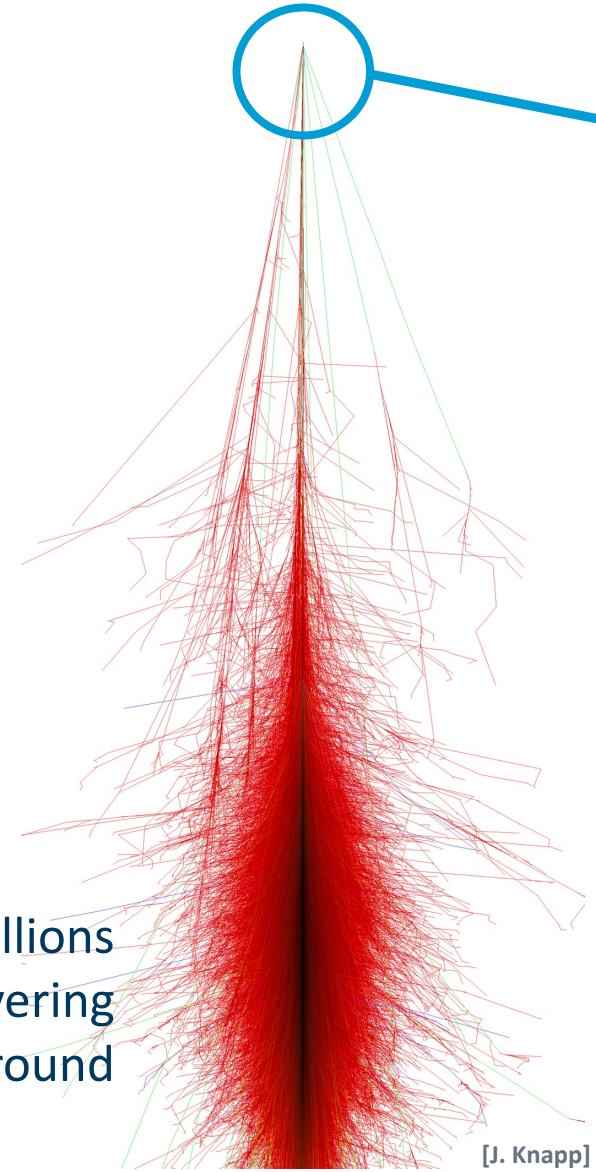
All databases: 150 GB (SQLite format)

Whenever ready…

Official data reconstruction at KIT

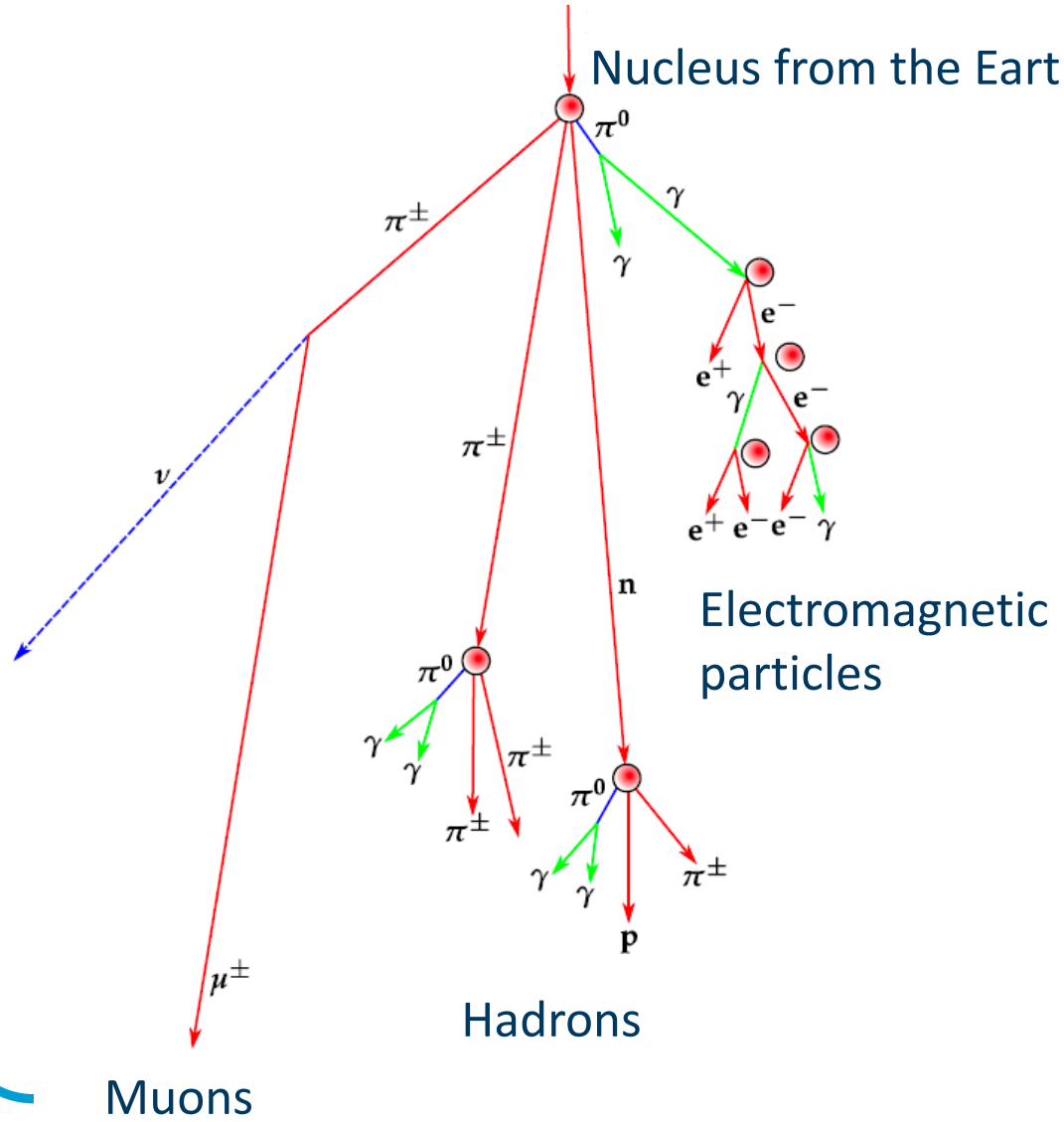
1 year of reconstructed data: 40 GB (SD-only + Hybrid)

Extensive Air Showers



[J. Knapp]

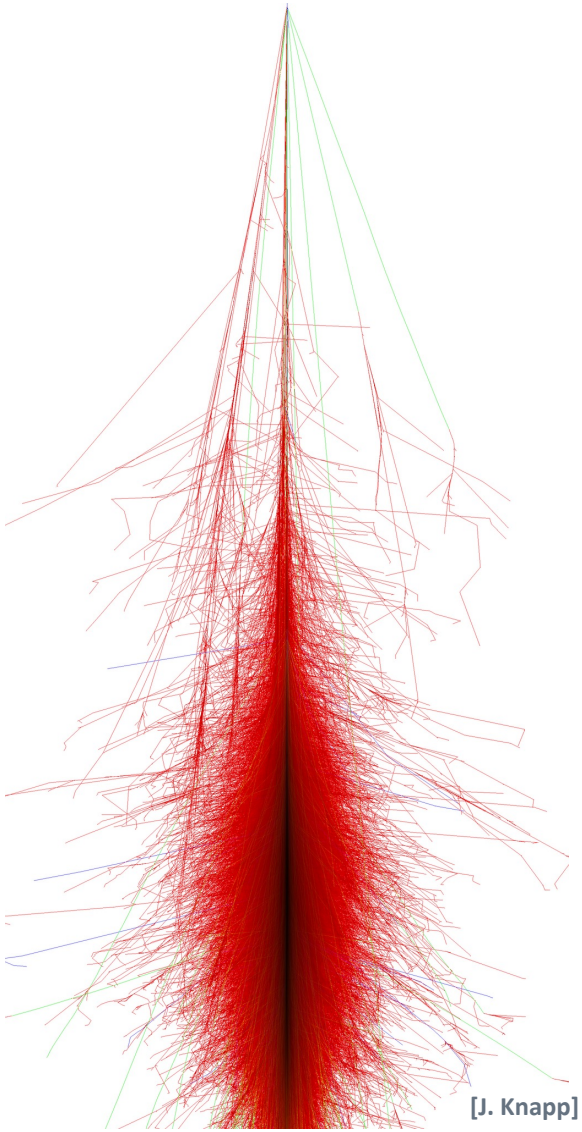
Cascade of billions
of particles, covering
a large area on ground



Simulating Extensive Air Showers

[D. Heck et al., FZKA Report 6019 (1998)]

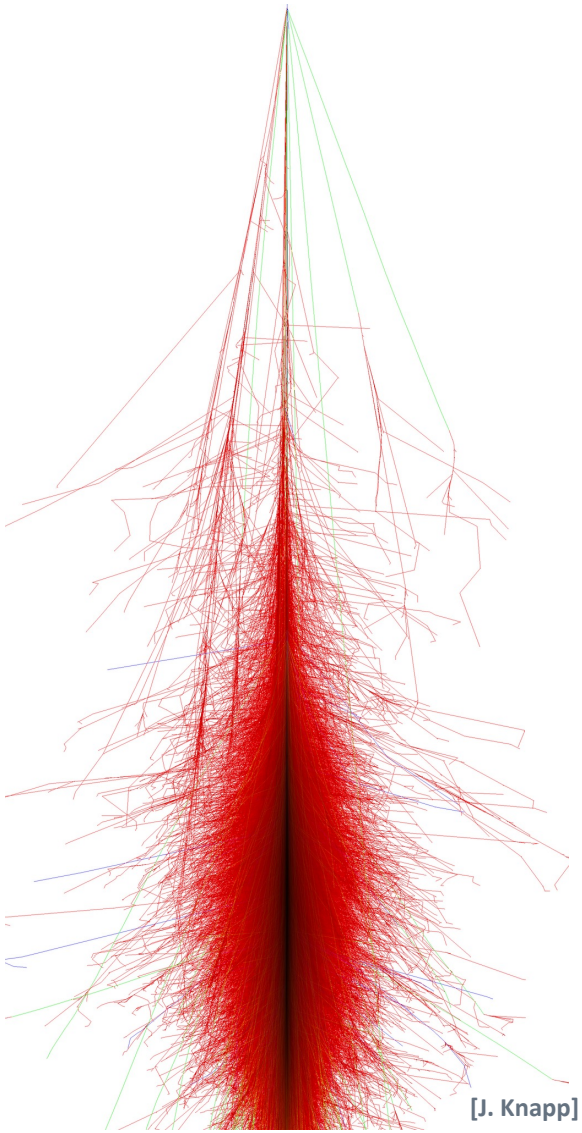
- Gold standard: **MC simulations** using CORSIKA
  - However: full MC simulation of an air shower at UHE **very resource-demanding**: CPU time needed on the SiMPLE cluster $\mathcal{O}(10^4\text{-}10^5$ h), disk space needed $\mathcal{O}(1$ TB) – **per shower**!
  - Need strategies to **drastically reduce** these demands:
    - **Parallelization** (e.g., Open MPI) will reduce the effective wall-clock time, but doesn‘t help with disk space
    - Not tracking every single particle will reduce both (**thinning algorithms**, cutting away very low-energy particles)
    - Typical/practical for one shower at UHE : **CPU time $\mathcal{O}$(1 d), disk space $\mathcal{O}$(1 GB)** – still not easy to produce large samples, but manageable
  - Side note: current **CORSIKA v7.7500** based on FORTRAN, new **CORSIKA v8** will be based on C++



[J. Knapp]

Simulating Extensive Air Showers

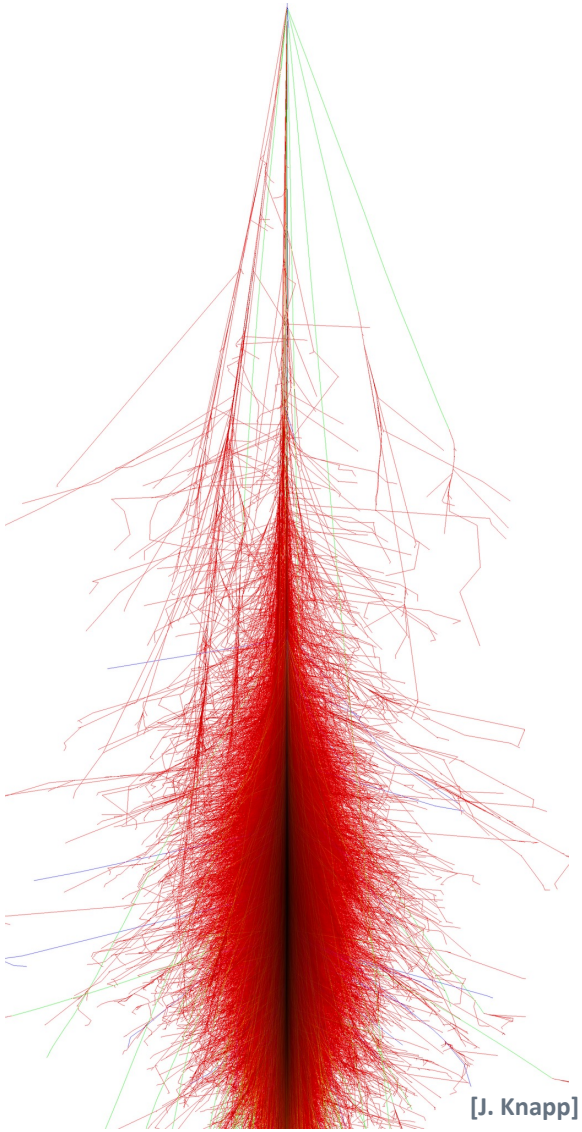
- **Complementary approach** when only the longitudinal development of the shower is of interest: **hybrid simulations** using CONEX [Bergmann et al.; Astropart. Phys. 26 (2007) 420] [Pierog et al.; Nucl. Phys. B, Proc. Suppl. 151 (2006) 159]
  - Full MC simulation only in the early part of the shower (most relevant for the overall development), numerical solution of **cascade equations** for the lower-energy part
  - Typical for one shower at UHE : **CPU time $\mathcal{O}$(1 min), disk space $\mathcal{O}$(10 KB)** – can produce huge samples quickly, but only limited information



[J. Knapp]

MC Shower Production for Auger

- **How many** CORSIKA showers are needed for an Auger analysis?
- **Example:** a search for photons with the Auger low-energy enhancements (SD-750 and HEAT) [Pierre Auger Coll., ApJ 933 (2022) 125]
  - **72,000** photon-induced showers, **42,000** proton-induced showers for the main analysis
  - **90,000** additional air showers for systematic checks (iron primaries, other hadronic interaction models)
  - Only a part was simulated on the SiMPLE cluster, the bulk of these simulations were **produced centrally** (by the Auger MC task) on the **Grid** and/or on local clusters in Napoli and Praha and were only copied to Siegen as needed (no longer-term storage, $\mathcal{O}$(1 TB) are needed temporarily for a typical sample of 5000 showers)



[J. Knapp]

Detector Simulations

- Second part of the MC production: simulating the **detector response** (including transmission through the atmosphere)
- Done locally on the SiMPLE cluster with the **Auger Offline Software Package**, which also encompasses the reconstruction code (i.e., same **output format** based on ROOT)
  [S. Argirò et al., NIM A 580 (2007) 1485]
- **Far less demanding** w.r.t. computing resources than shower simulations

Proton simulations	$\log_{10}(E/\mathrm{eV})$					
Properties	16.5–17.0	17.0–17.5	17.5–18.0	18.0–18.5	18.5–19.0	19.0–19.5
status	Completed	Completed	Completed	Completed	Completed	Completed
total size of dir. [GB]	28.37	45.20	79.14	138.81	267.02	491.20
total number of jobs run	5000	5000	5000	5000	5000	5000
avg. runtime per job [h:mm:ss]	0:19:55	0:29:34	0:46:55	1:15:44	2:16:18	3:40:47
total number of events simulated	25 000	25 000	25 000	25 000	25 000	25 000
avg. size of output file [MB]	0.88	1.54	2.92	5.36	10.60	19.77

[T. Fehler, Master's thesis (2024)]

Other Computing-Intensive Activities in our Group

- Impact of **Lorentz invariance** violation on air showers
  - Only looking at the impact on X_{max} so far, so CONEX can be used
  - Need to **modify the MC part of CONEX** and then set the threshold for the change to cascade equations accordingly – individual simulations take a little bit longer w.r.t. standard settings, but a **fine scan of the phase space** in primary energy and LV parameter is needed [Klinkhamer, MN, Risse; Phys. Rev. D 96 (2017) 116011] [Duenkel, MN, Risse; Phys. Rev. D 104 (2021) 015010] [Duenkel, MN, Risse; Phys. Rev. D 107 (2023) 083004]
  - Moving now to CORSIKA to also study the impact on the particles on ground: need to check how the **available resources can be used efficiently**

- Simulating the **propagation of cosmic particles through the Universe**
  - Relevant software: **CRPropa3** [Batista et al., JCAP 09 (2022) 035]
  - Similar to CONEX: **individual simulations** don‘t need much CPU time or disk space (can comfortably performed on a laptop), but in the end you need **a lot of them**
  - Activity **just starting up**, will occupy more computing resources in the near future