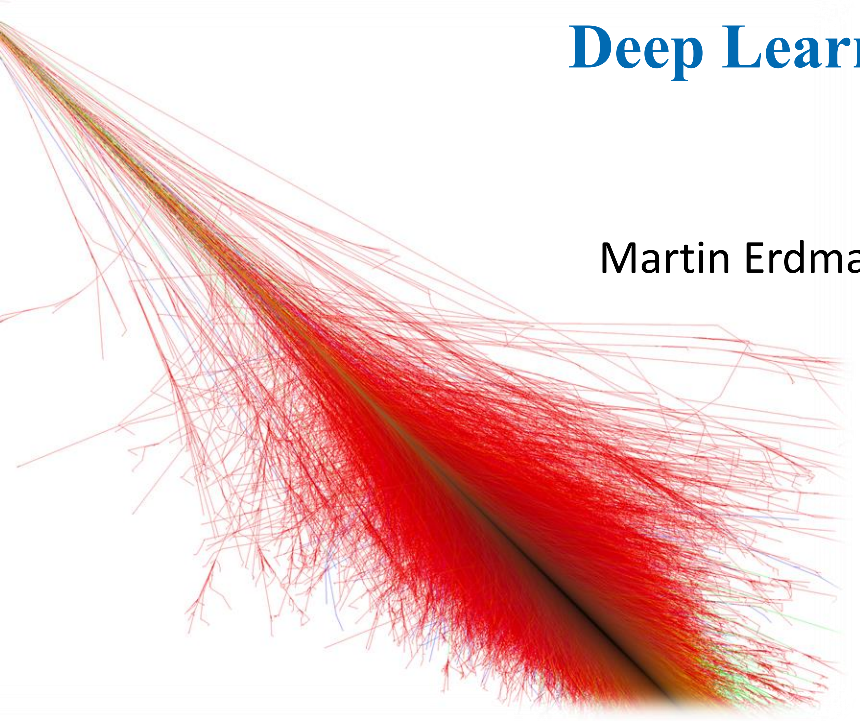


Deep Learning for the Pierre Auger Observatory and its Upgrade

Martin Erdmann, **Jonas Glombitza**, **Niklas Langner**, Dominik Steinberg

III. Physikalisches Institut A
RWTH Aachen University

DNN → **Air Shower Properties**

Shower Image Credit: https://www-zeuthen.desy.de/~jknapp/fs/proton-showers.html

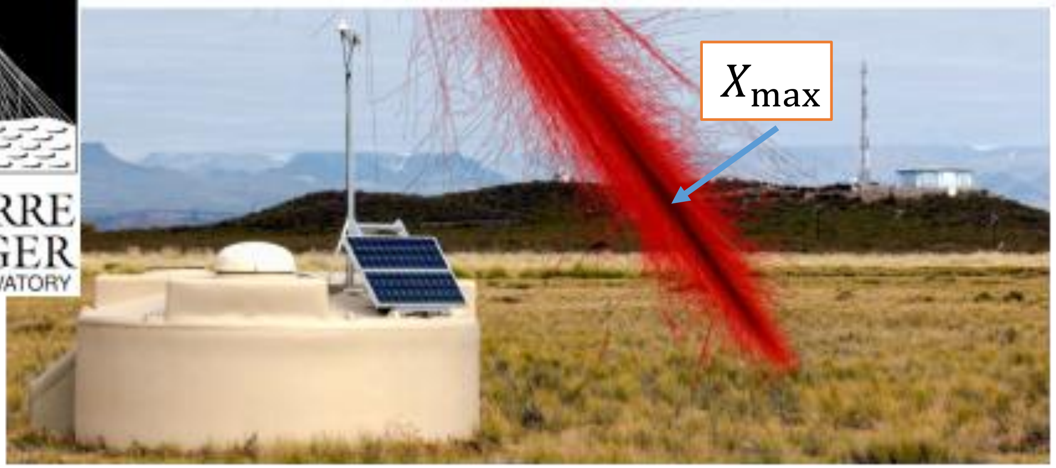
ErUM Data Collaboration Meeting February 2022

Quality assurance criteria

1. Defining a meaningful **challenge**
2. Development of a dedicated **model**
   - Exploit data symmetries
   - Careful Preparation of **data**, e.g., no extrapolation
3. Verification of the **training**
   - Ensure adequate model configuration
   - Test set (check for overtraining)
4. **Validation on simulations**
   - Verification on various simulations
   - Systematic uncertainties
5. **Validation on Data**
   - Cross calibration
   - Systematic uncertainties
6. Performance measures
   - physical and model **limits**
7. **Understanding** the model and data
   - Input relevance

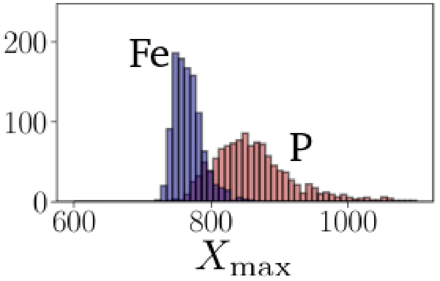
Challenge: $X_{\max}$ reconstruction at the Pierre Auger Observatory

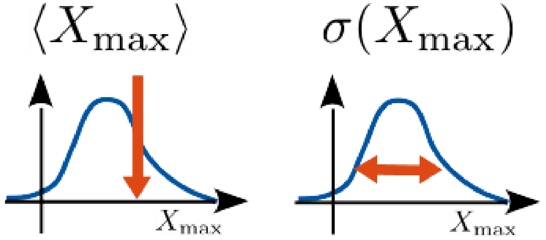
PIERRE AUGER OBSERVATORY

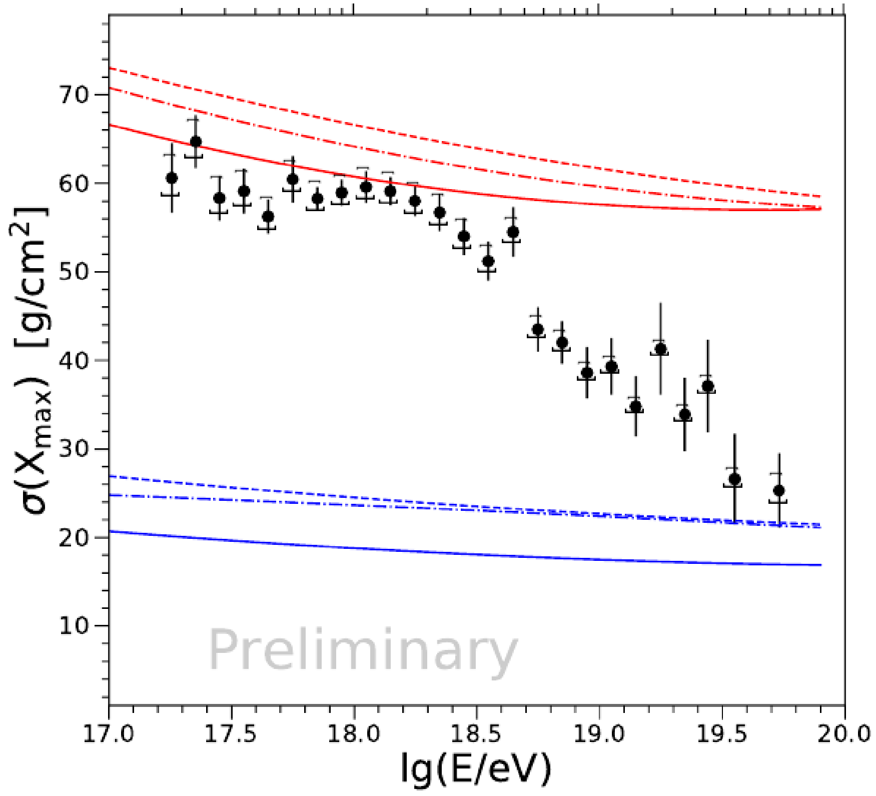


Pierre Auger Observatory

- Fluorescence Detector (15% duty cycle)
  - direct and precise observation of shower maximum $X_{\max}$
- Surface Detector (~100% duty cycle)
- reconstruction of shower maximum using deep learning
- verification using hybrid measurements







Measurement of $\sigma(X_{\max})$

- sensitive to composition mix
- reconstructed using FD
- fluctuations 20-60 g/cm²
- uncertainty statistically dominated

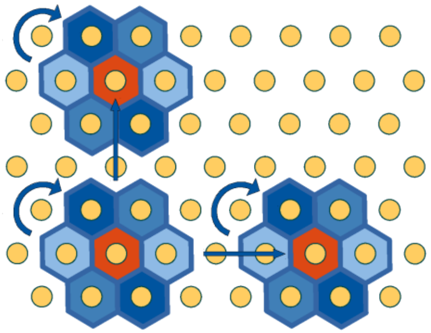
Data preparation and model design

Network for X_{max} reconstruction

- signals vary on exponential scale
    - apply logarithmic transformation
- normalize timing measurements
- exploit data symmetries

Bidirectional LSTMs analyze signal traces

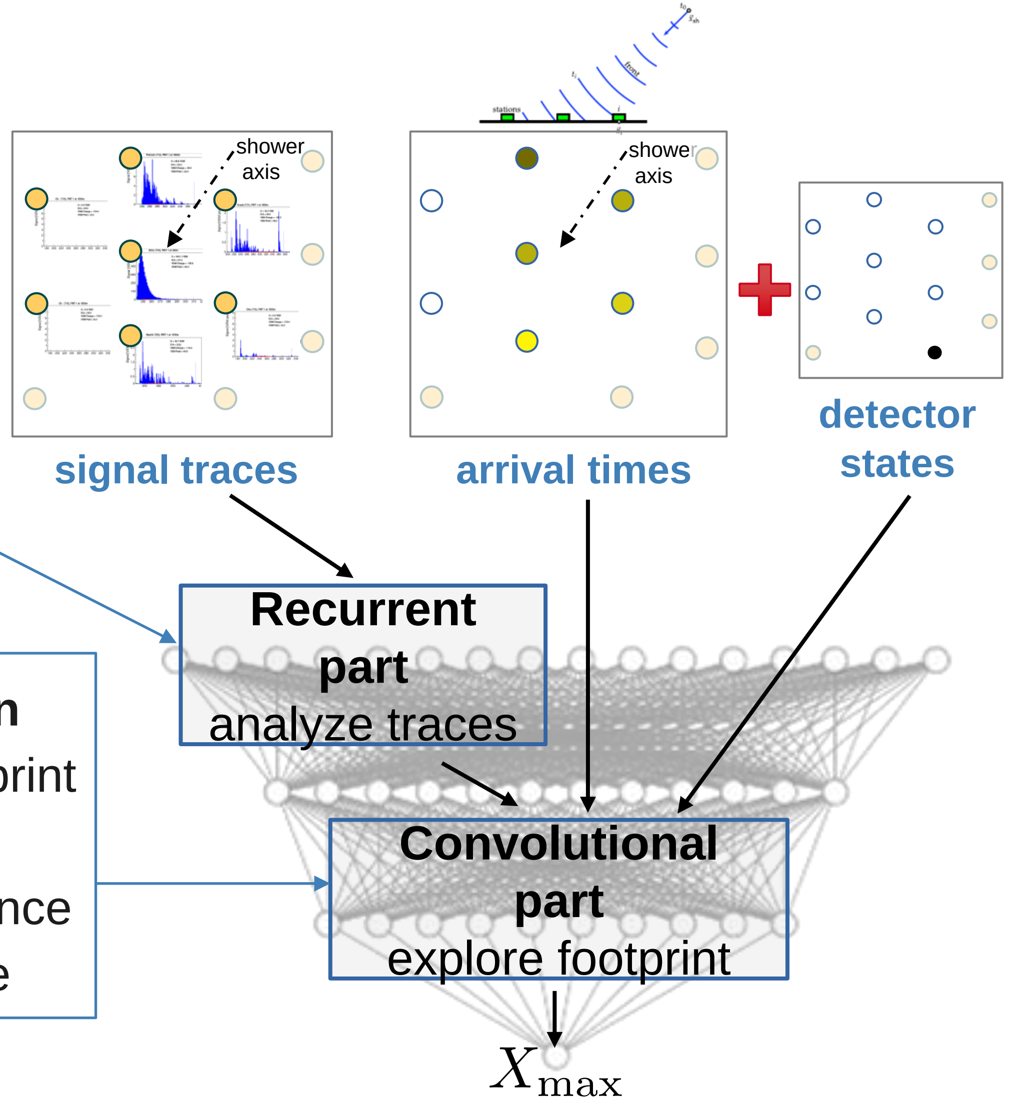
- network shared over stations



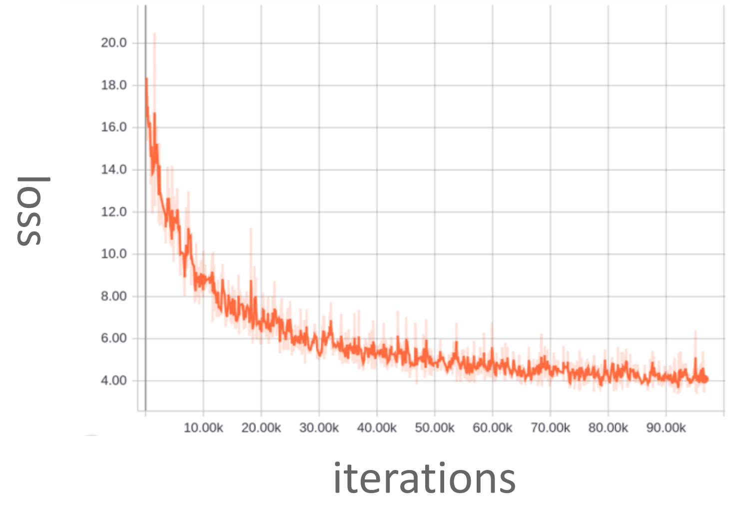
Hexagonal convolution

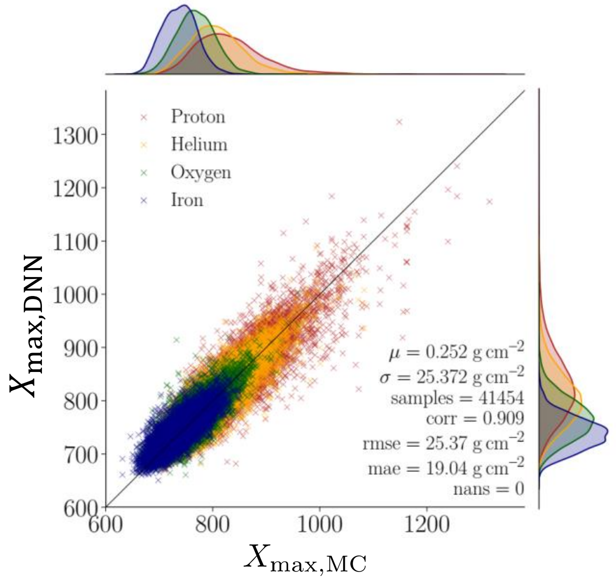
exploits hexagonal footprint

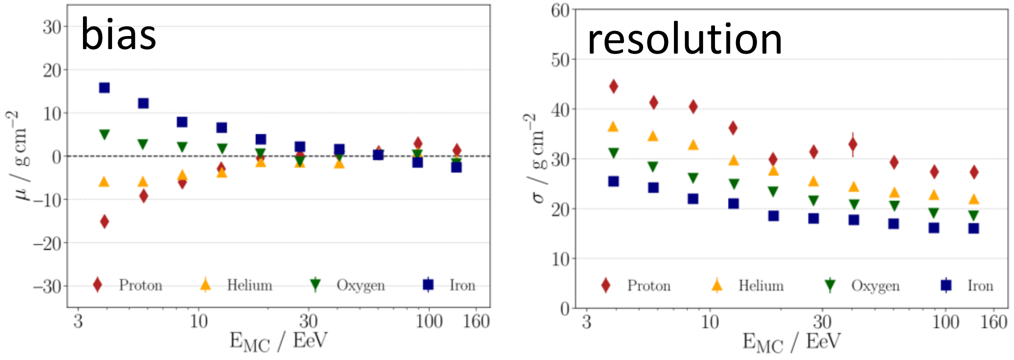
- hexagonal filter
- translational invariance
- rotational invariance



Network training and basic verification







- features ~1.5 million parameters
- train with augmented simulation data
  - *mimic various detector states: broken stations/PMTs, saturation*
  - *training on GPU ~ 1-2 days*

Check bias and variance/resolution

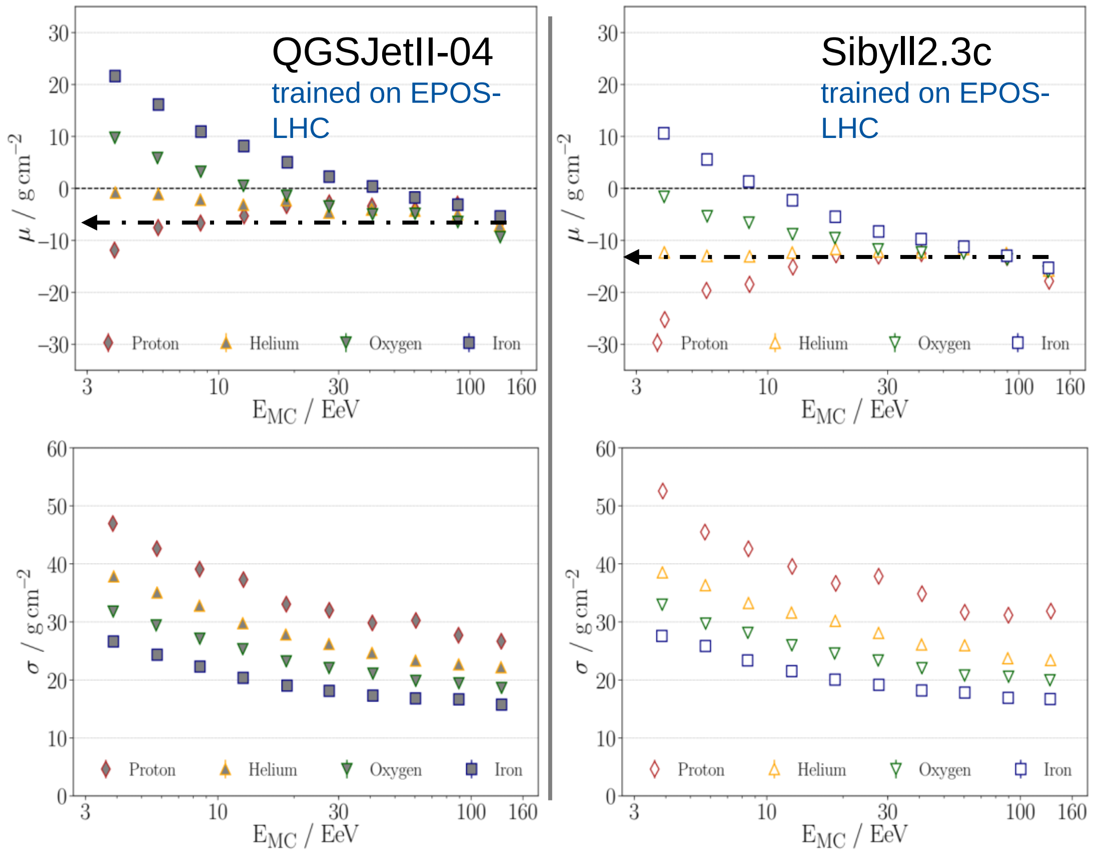
Validation on simulation: additional interaction models

Simulation: phenomenological modeling of air showers → various interaction models
DNN trained using EPOS-LHC model
Evaluated on simulations with different hadronic interaction models

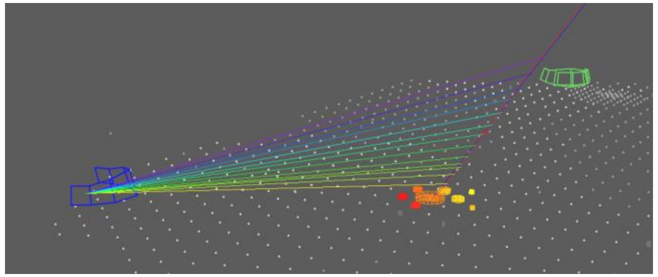
- QGSJET-II.04
- SIBYLL2.3c

- similar resolution
  - interaction model independent

- bias depends on model
  - absolute scale shifted (negative)
  - $X_{\max}$ scale of the DNN depends on interaction model



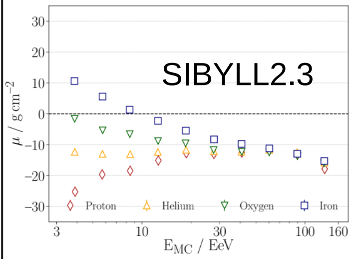
Validation on data: Auger Hybrid Measurements



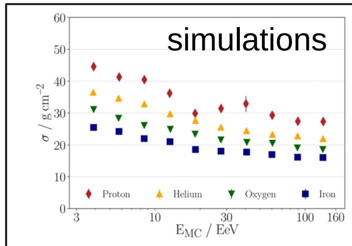
Calibrate method using hybrid data (SD measurement and FD observation)

- calibrate for -30 g/cm² offset
  - shortcomings of (detector-)/simulation
- validate resolution

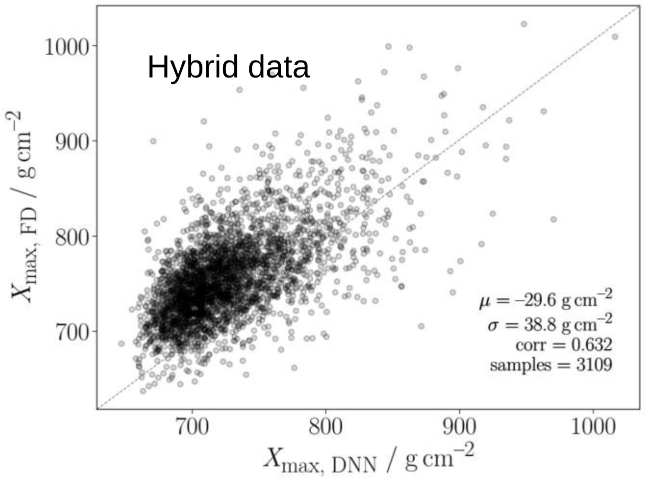
→ promising results to measure UHECR composition using SD statistics

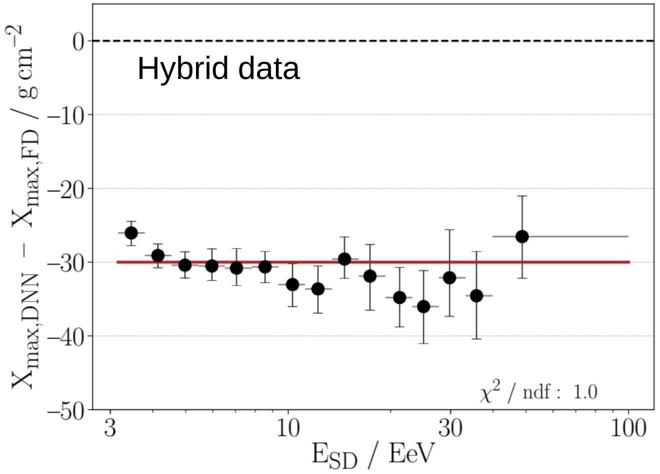


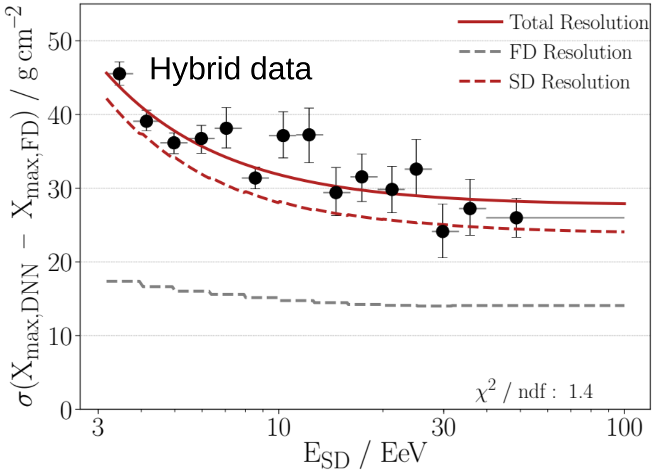
bias depends on hadr. model



agreement with simulations

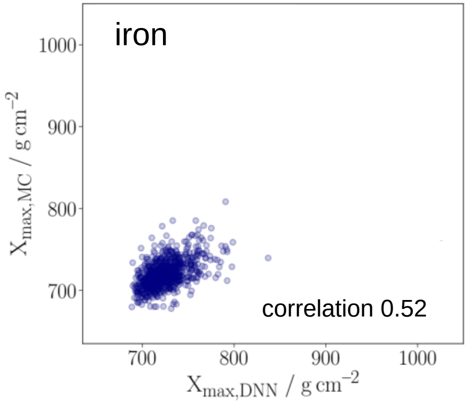


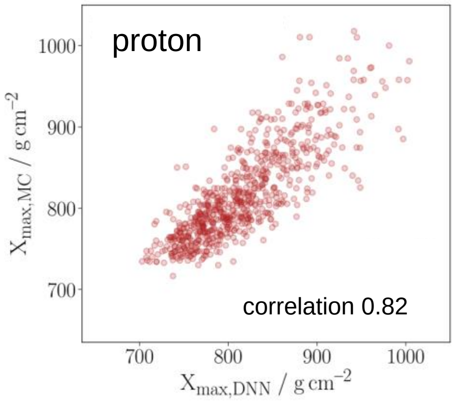


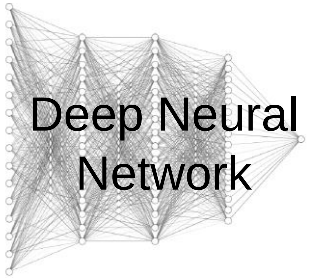


Validation on data: comparison to traditional methods

Simulations

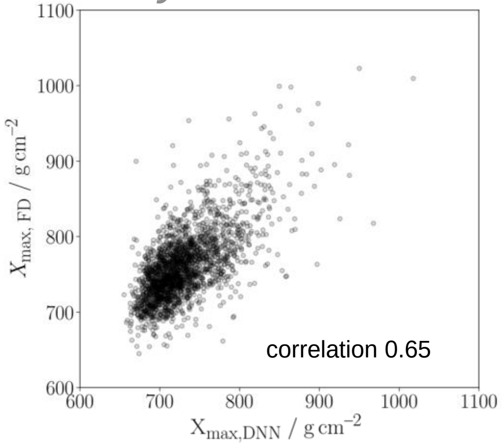






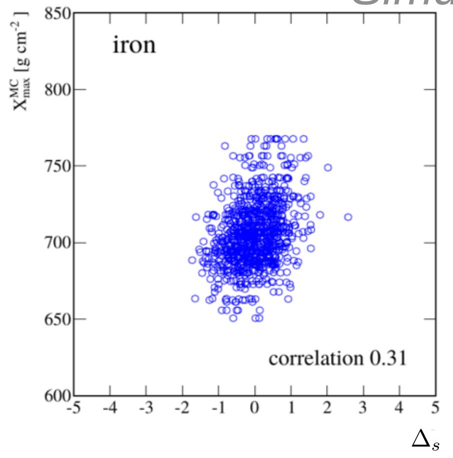
improvement of DNN
by almost 40% (correlation)
enable to measure $\sigma(X_{\mathrm{max}})$

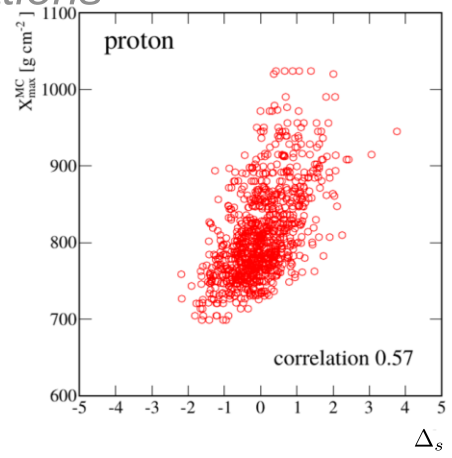
Hybrid Data



A. Aab et al. (Pierre Auger Collaboration), JINST 16 P07019 (2021)

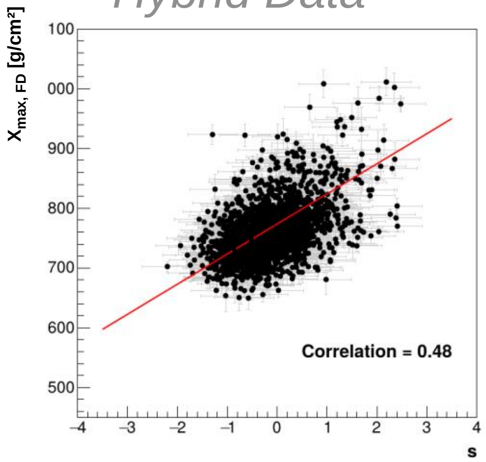
Simulations





Delta Method
determines composition
via rise time of signal

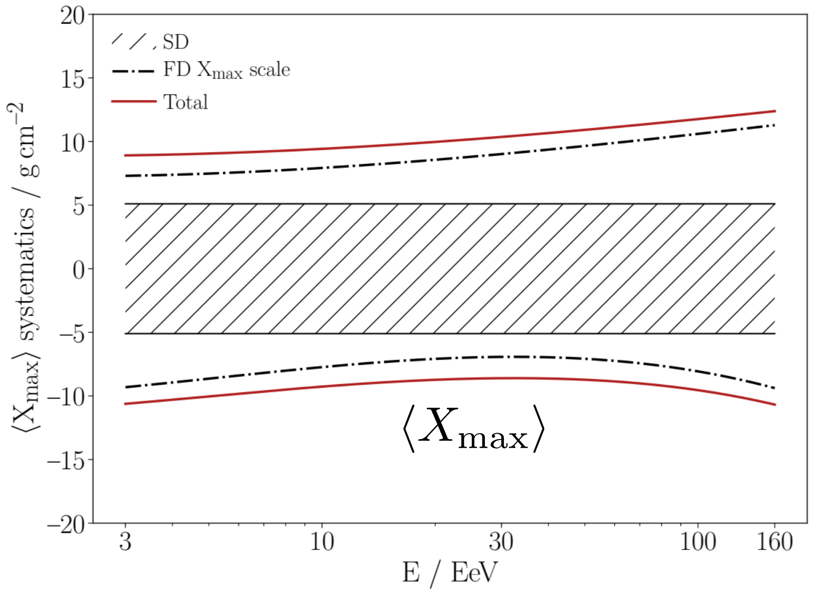
Hybrid Data



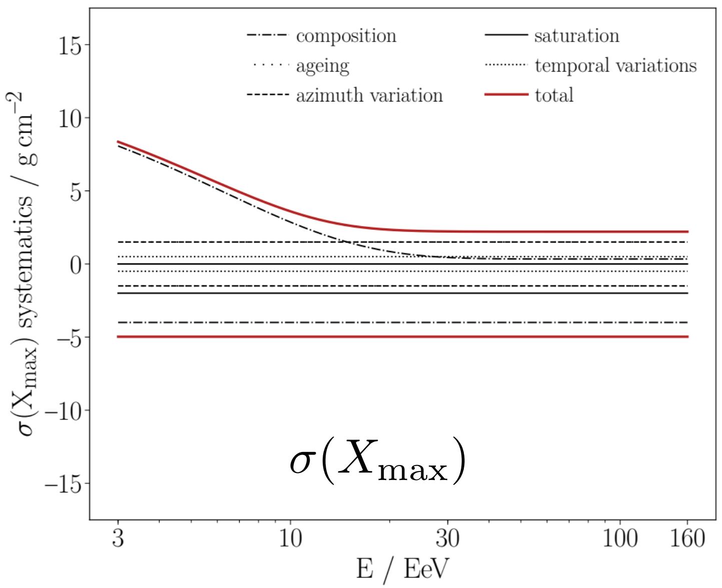
A. Aab et al. (Pierre Auger Collaboration), Phys. Rev. D 96, 122003, 2017

Estimate systematic uncertainties

Estimation of systematic uncertainties (using data and simulation)
- highly dependent on analysis



- method calibrated using the FD
  - sys. uncertainty 10 – 15 g/cm²
- uncertainty similar to delta method



- no calibration using FD performed
  - sys. uncertainty 5 – 10 g/cm²
- **1st measurement beyond 80 EeV**
- **can provide new insights into cosmic-ray composition**

Expectations of network behavior

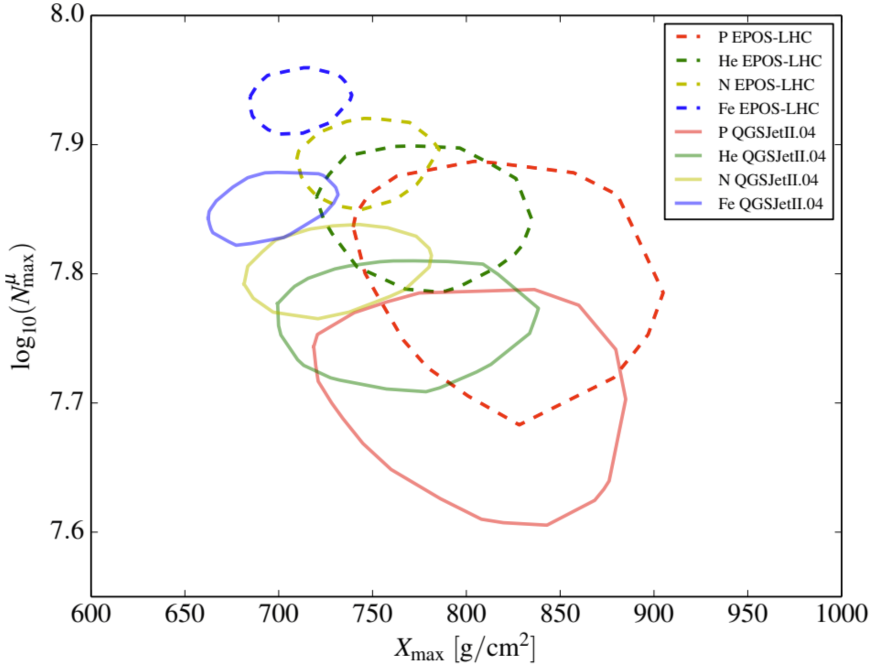


Image Credit: https://arxiv.org/abs/1604.03637

Rising edge: Muons

Falling edge: Electromagnetic

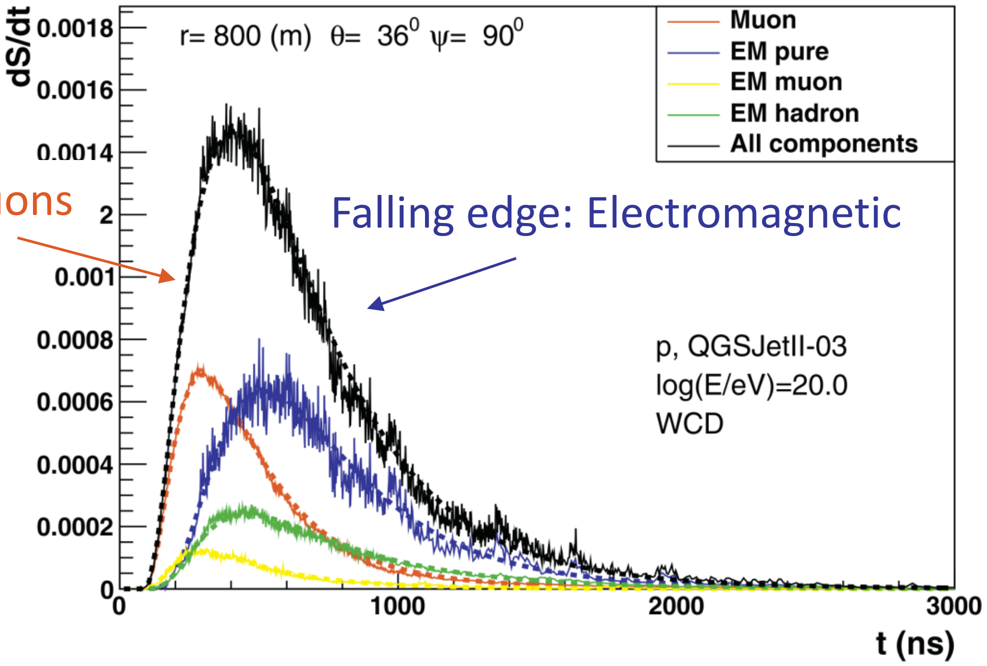
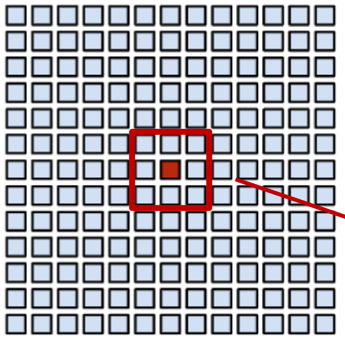


Image Credit: https://www.sciencedirect.com/science/article/abs/pii/S0927650517300105

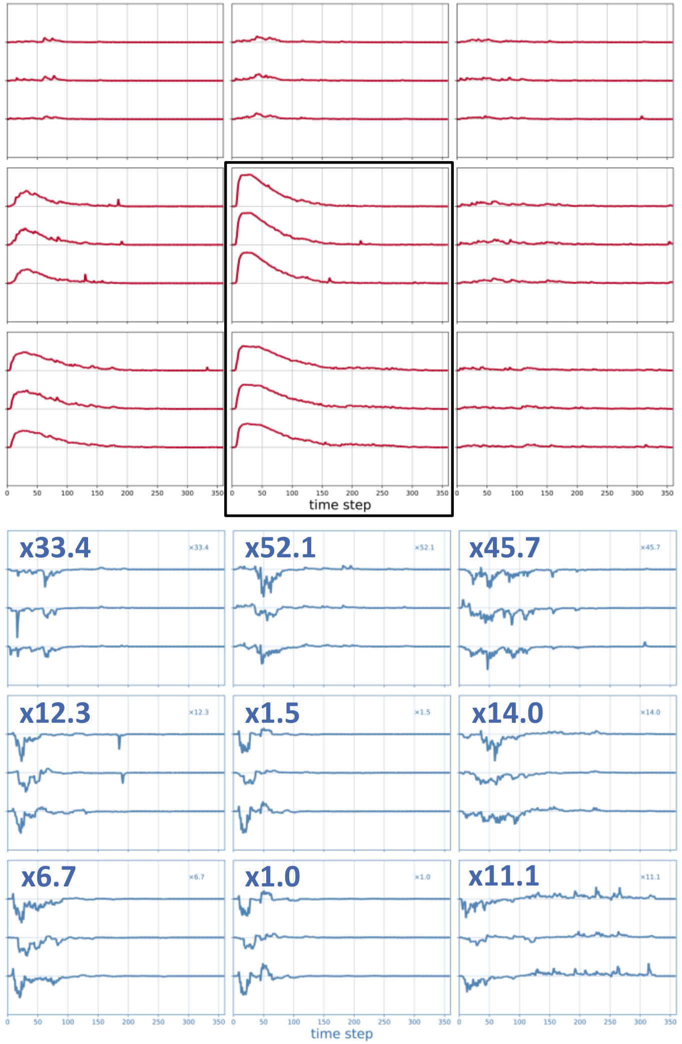
N_μ increases with A $\Rightarrow$ Stronger **rising edge** should **decrease** network output (heavier particle, smaller $X_{\max}$)
Stronger **falling edge** should **increase** network output (lighter particle, larger $X_{\max}$)

ϵ-LRP Examples with Respect to X_{max}-Prediction



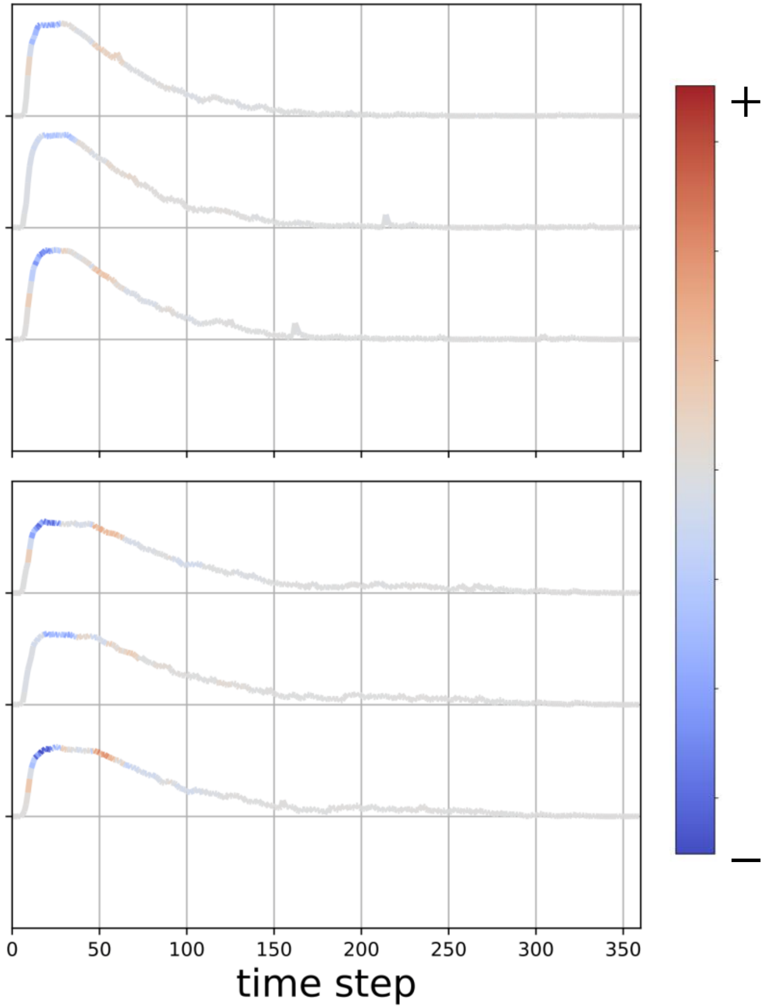
Look at 3×3 stations around shower center

Traces



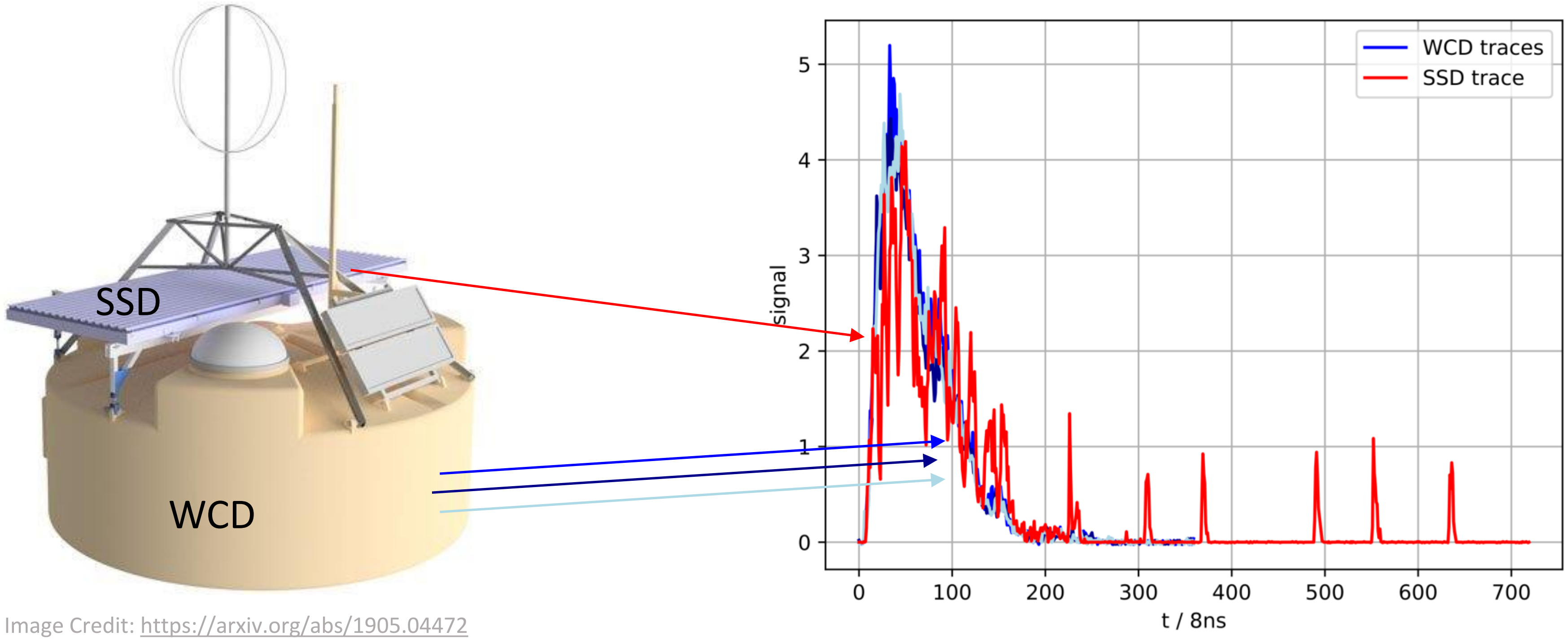
ϵ-LRP

How does the input have to be changed to increase the X_{max} output of the network?



AugerPrime: Surface Scintillation Detector upgrade

- Pierre Auger Observatory is currently being upgraded
- New **Surface Scintillation Detector (SSD)**: additional time trace
- Improved sensitivity to cosmic-ray composition expected (better seperation between muon and electromagnetic components)
- Electronics upgrade: sampling rate three times higher

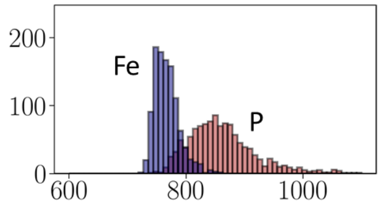


Incorporate into network!

Image Credit: https://arxiv.org/abs/1905.04472

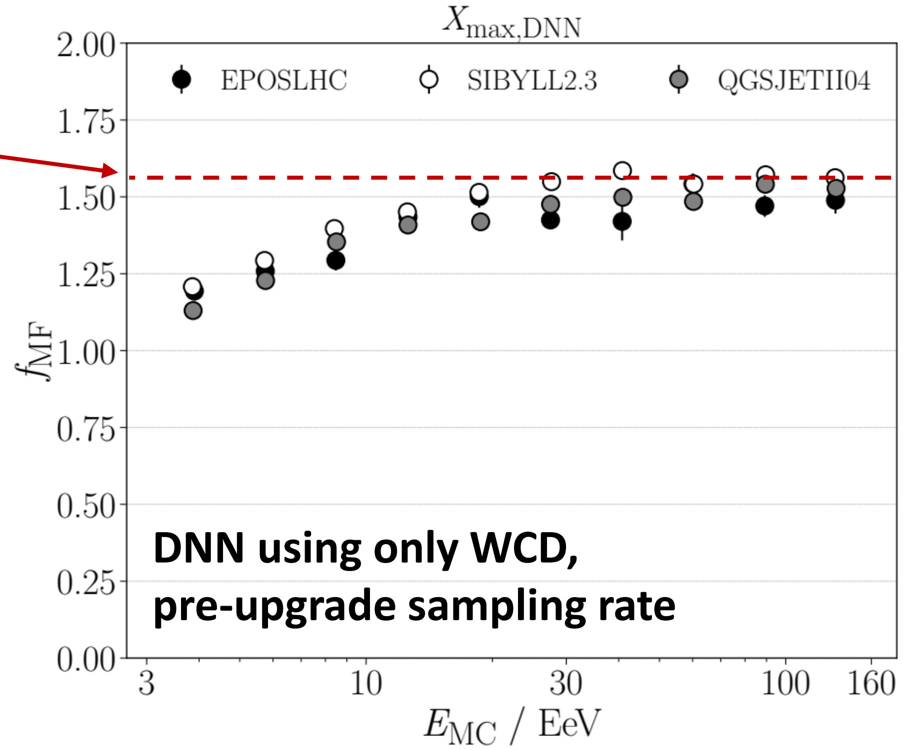
Limits of physics observables

Composition merit factor for discriminating between proton and iron



$$f_{\mathrm{MF}} = \frac{|\langle X_{\mathrm{max,P}}\rangle - \langle X_{\mathrm{max,Fe}}\rangle|}{\sqrt{\sigma^2(X_{\mathrm{max,P}}) + \sigma^2(X_{\mathrm{max,Fe}})}}$$

- merit factor of simulated $X_{\mathrm{max,MC}}$: ~ 1.5
- DNN merit factor increases with energy
  - above 10 EeV, merit factor = 1.5
  - good separation for all interaction models



Pre-upgrade DNN reconstructing X_{max} already reaches physical limit of mass separation power (of $X_{\mathrm{max,MC}}$)

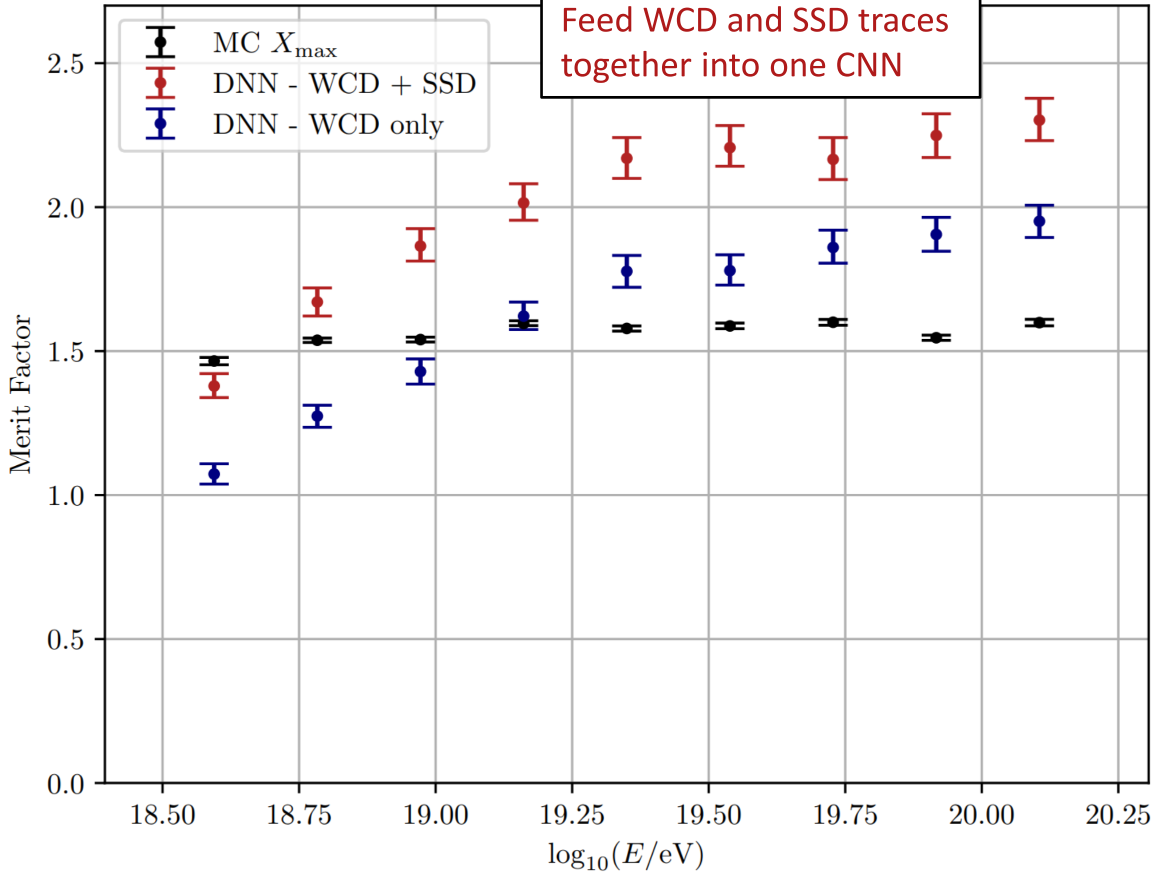
⇒ **Another observable is needed to benefit from upgrade**

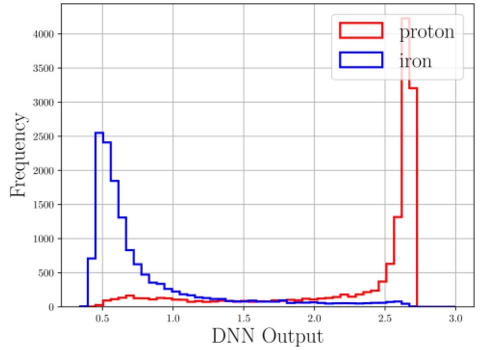
Composition merit factor increase by combined measurement

Train DNN to form its own observable, with objective to maximize the merit factor

→ Useful to assess performance limits on simulations but cannot be verified on data

Preliminary approach:
Feed WCD and SSD traces together into one CNN





- WCD measurements alone already able to surpass merit factor of X_{max}
- Clear improvement by SSD upgrade
- Combined WCD and SSD measurements have a large potential for improved mass separation

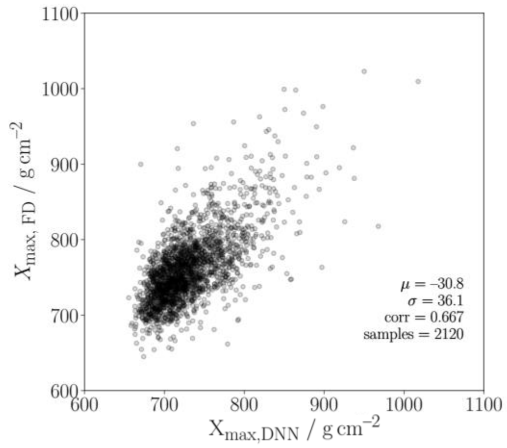
⇒ **Challenge:** Define a suitable (motivated by physics) observable that can harvest this potential

Summary

Deep Learning at the Surface Detector of the Pierre Auger Observatory

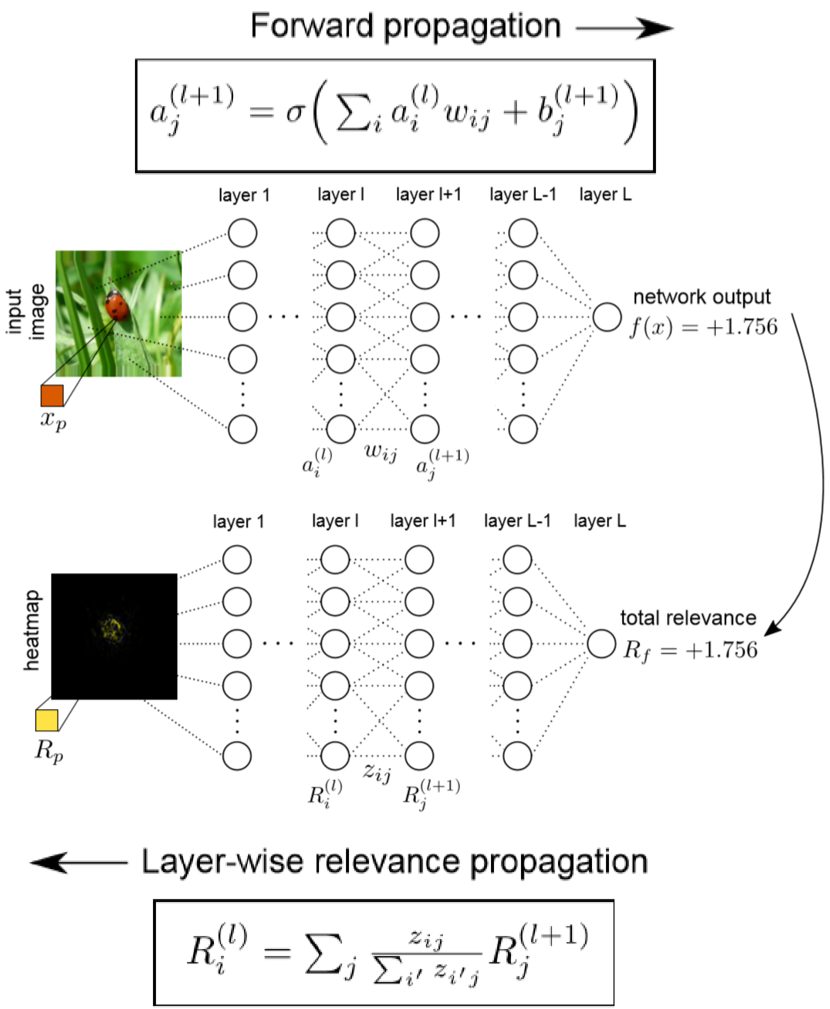
- extract mass-sensitive information, exploits symmetry of data (RNN + CNN)
- event-wise reconstruction of X_{max}
  - performance validated on simulations and data (hybrid events)
  - extensive study on systematic uncertainties
  - expected uncertainties for X_{max}, $\sigma(X_{\mathrm{max}})$ measurements are small
    - raise in statistics of a factor 10 → new insights into UHECR composition

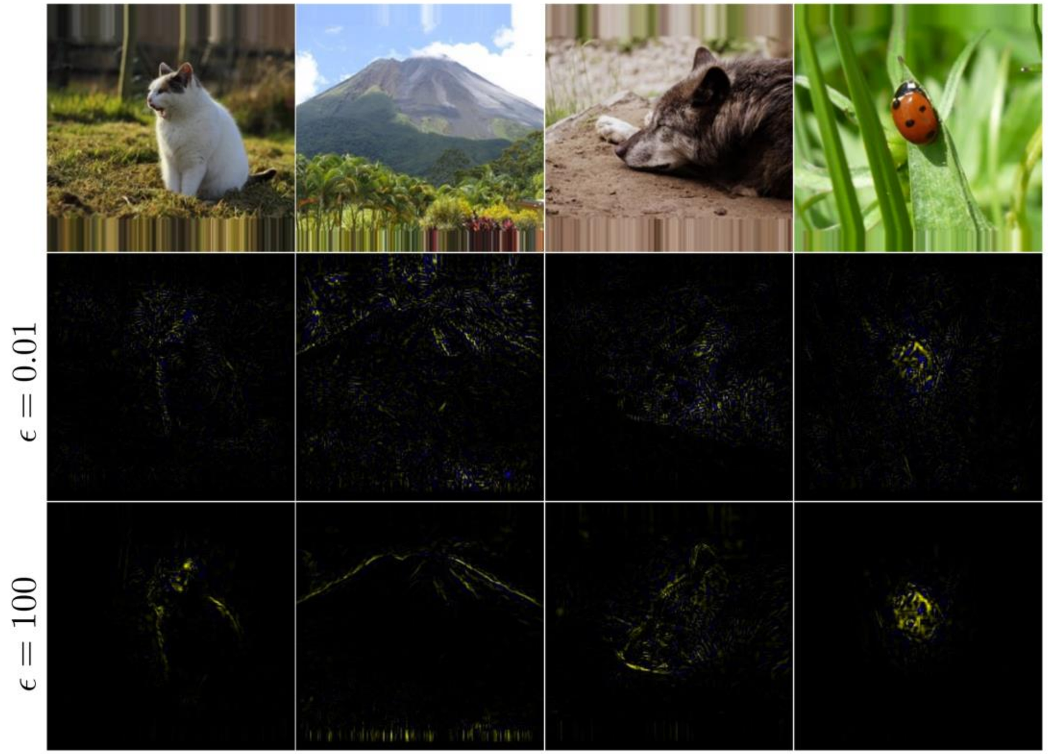
- First steps towards DNN introspection
- AugerPrime Upgrade will enable additional insights and cross checks!
  - development of dedicated architecture
  - increased performance at the event level expected
  - challenge: find powerful observable that can be validated by data



Backup

ϵ-Layer-Wise Relevance Propagation



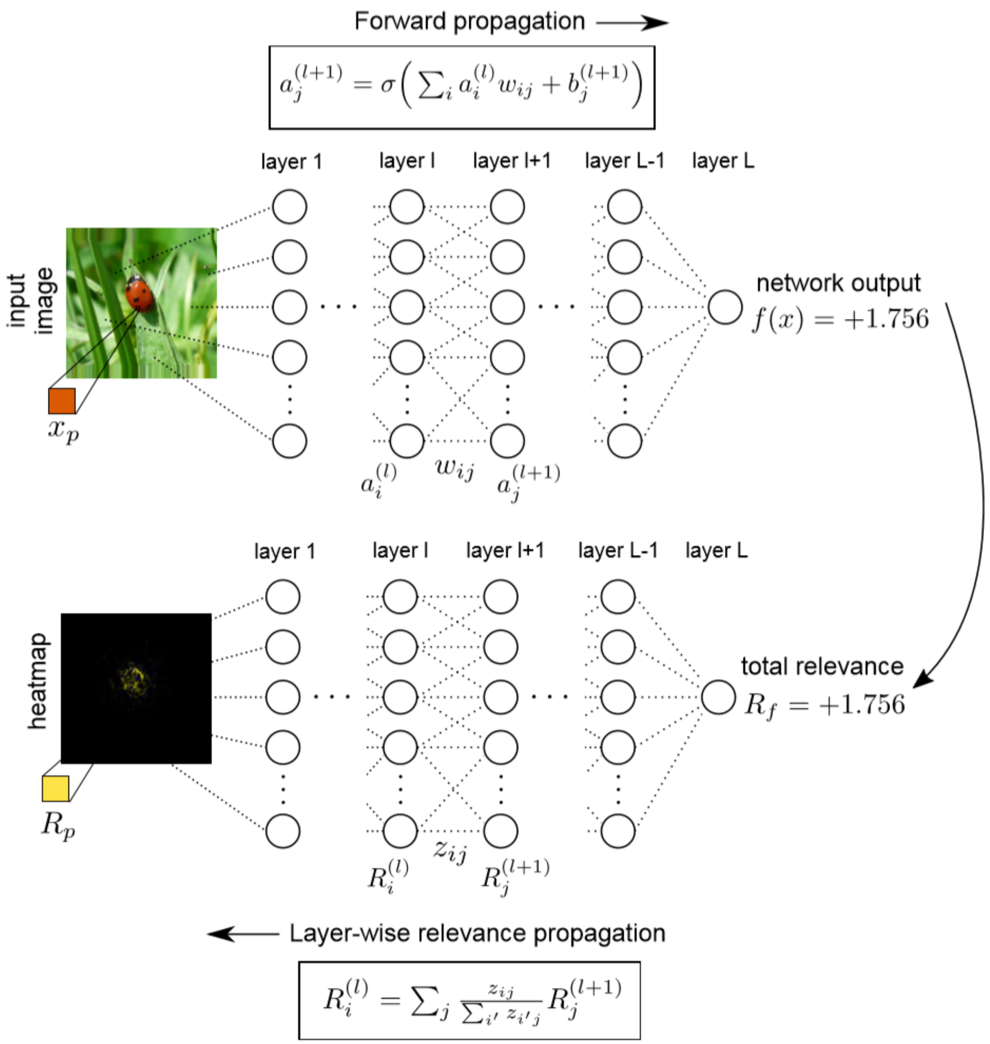


$$\epsilon\text{-LRP: } R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j} + \epsilon \cdot \mathrm{sign}\left(\sum_{i'} z_{i'j}\right)} R_j^{(l+1)}$$

Image Credit: http://iphome.hhi.de/samek/pdf/BinICISA16.pdf

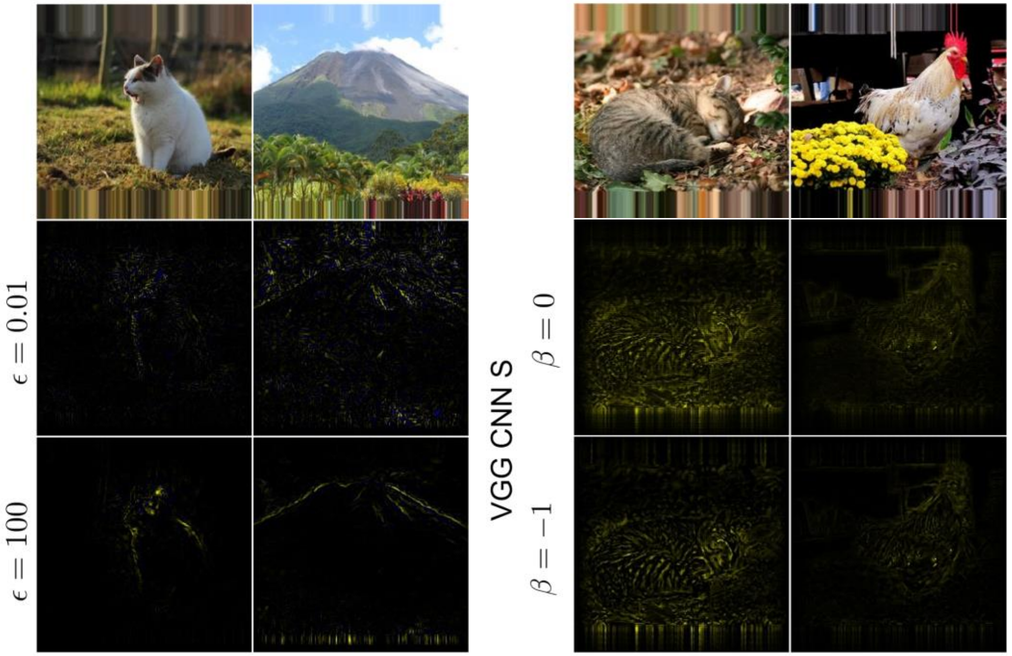
Layer-Wise Relevance Propagation

- Deep neural network: feed-forward graph of neurons:
  - $x_j^{(l+1)} = g\left(\sum_i x_i^{(l)} w_{ij}^{(l,l+1)} + b_j^{(l+1)}\right)$
- Use the networks output $f(\boldsymbol{x})$ and a backward pass of same graph to calculate relevance scores
  - $R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j}} R_j^{(l+1)}$ with $z_{ij} = x_i^{(l)} w_{ij}^{(l,l+1)}$
  - i: index of neuron at layer l; Σ_j sums over al upper-layer neutrons to which neuron i contributes
- Conservation property:
  - $\sum_p R_p^{(1)} = f(\boldsymbol{x})$



Layer-Wise Relevance Propagation

- Standard definition:
  - $R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j}} R_j^{(l+1)}$
    - $z_{ij} = x_i^{(l)} w_{ij}^{(l,l+1)}$
- ϵ-LRP (better numerical properties)
  - $R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j} + \epsilon \cdot \mathrm{sign}\left(\sum_{i'} z_{i'j}\right)} R_j^{(l+1)}$
- β-LRP (conserving relevance)
  - $R_i^{(l)} = \sum_j \left(\alpha \cdot \frac{z_{ij}^+}{\sum_{i'} z_{i'j}^+} + \beta \cdot \frac{z_{ij}^-}{\sum_{i'} z_{i'j}^-} \right) R_j^{(l+1)}$
    - $z_{ij}^+ + z_{ij}^- = z_{ij}$ (positive and negative part)
    - $\alpha + \beta = 1, \alpha > 0, \ \beta \leq 0$



- Higher ϵ: Less sensitivity to noise, but potential dampening
- Higher values for β: Only keep strongest regions (more weight on inhibitors), but potential supression