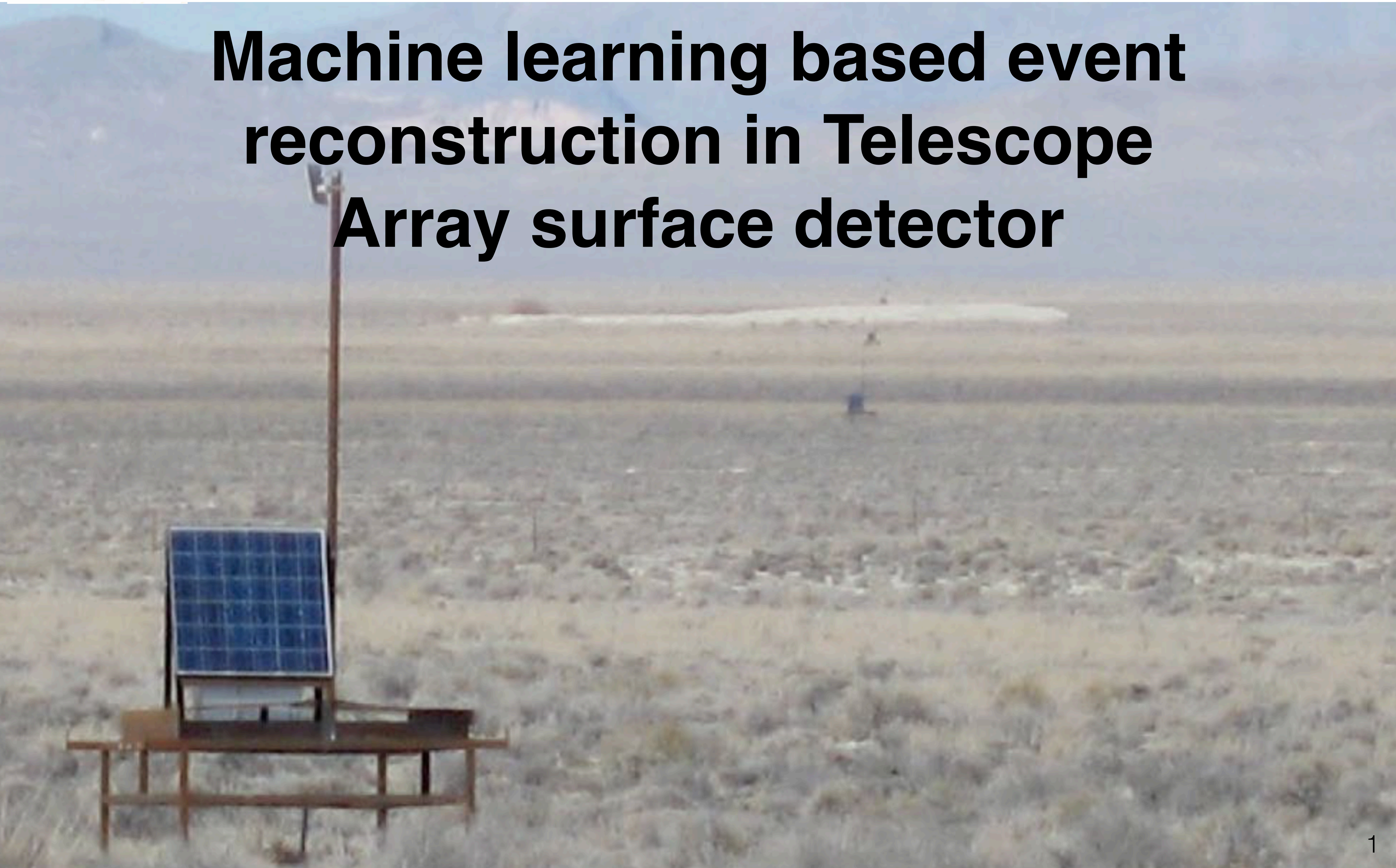
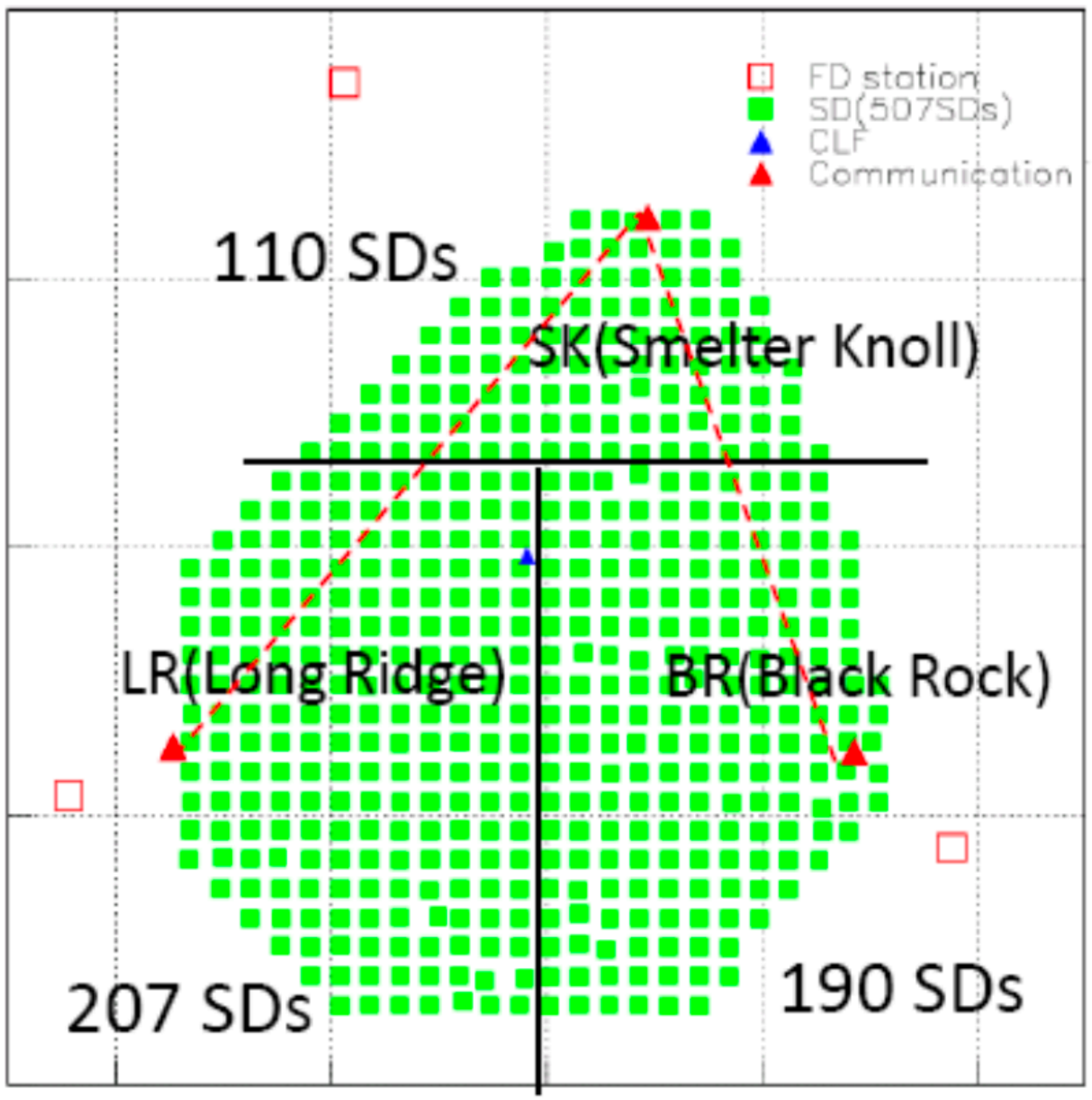
Workshop on ML for Cosmic Ray Air Showers, 02.02.2020
Univ. of Delaware Bartol Research Institute

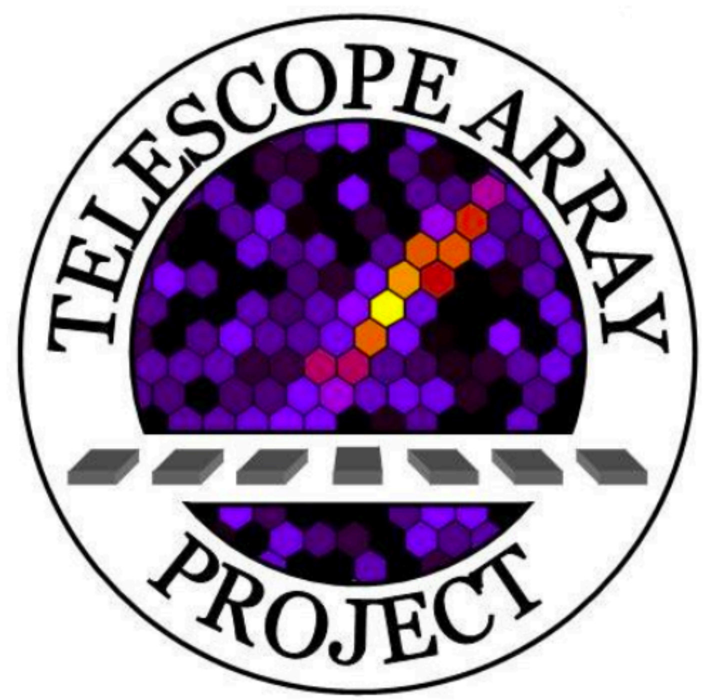
Oleg Kalashev
INR RAS
Moscow

Machine learning based event reconstruction in Telescope Array surface detector



Telescope Array Surface Detector





- 507 SD's, 3 m^2 each
- 680 km^2 area
- operating since May 2008

Largest UHECR statistics in the Northern Hemisphere

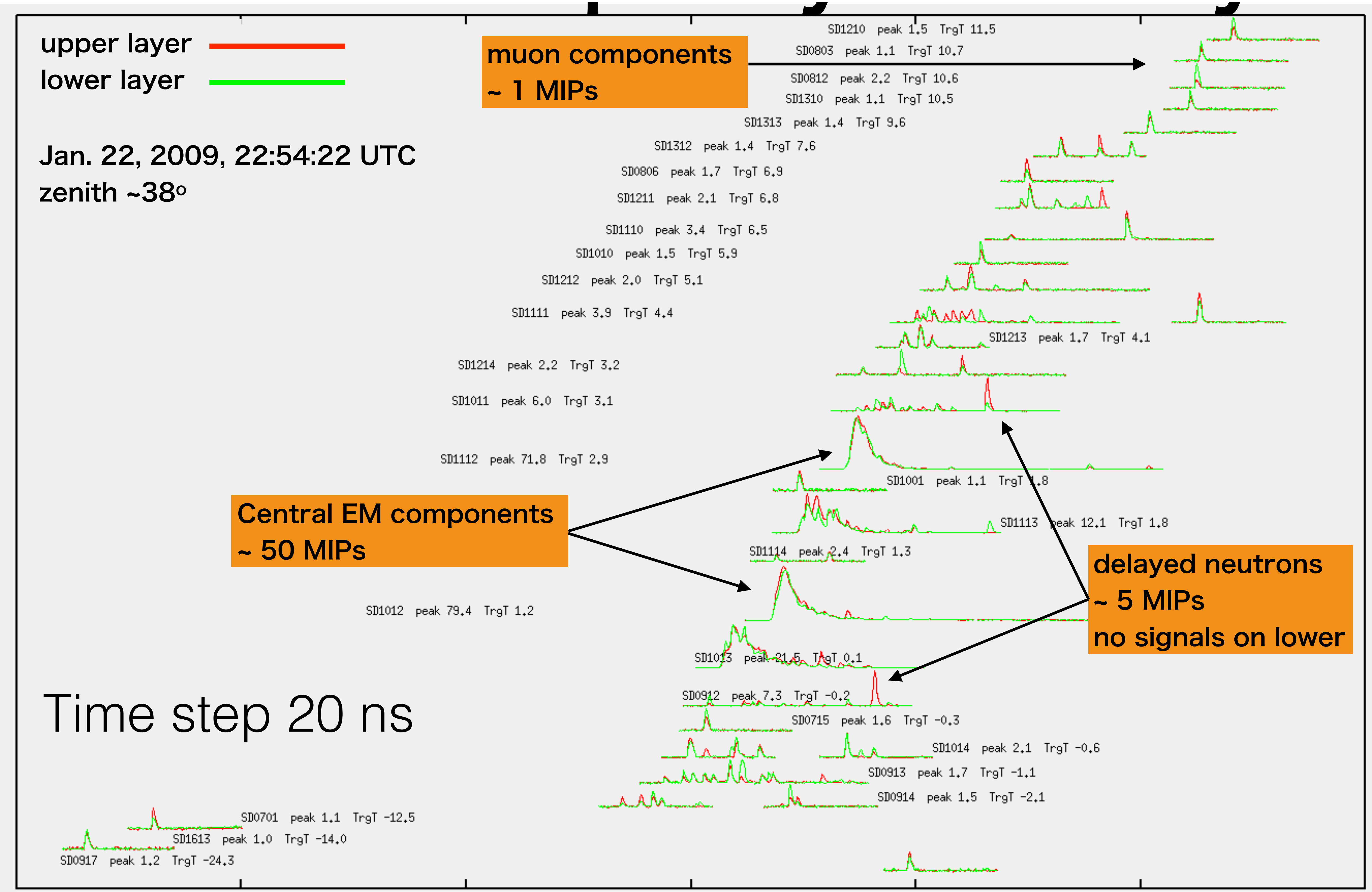
Outline

With ML-based reconstruction we will

- *recover UHE primary energy and direction*
- *study mass composition*
- *search for UHE photons*

O. Kalashev et al, PoS ICRC2019 (2020) 304
TA Collaboration,, J.Phys.Conf.Ser. 1525 (2020) 1, 012001
D. Ivanov et al, Mach.Learn.Sci.Tech. 2 (2021) 1, 015006
TA Collaboration, PoS ICRC2021 (2021) 384
TA Collaboration, PoS ICRC2021 (2021) 864
O. Kalashev et al, arXiv:2112.02072

Sample event



Event reconstruction

standard parametric approach

- LDF

$$f\left(r\right)=\left(\frac{r}{R_m}\right)^{-1.2}\left(1+\frac{r}{R_m}\right)^{-(\eta-1.2)}\left(1+\frac{r^2}{R_1^2}\right)^{-0.6}$$

$$R_m=90.0\ \mathrm{m},\ R_1=1000\ \mathrm{m},\ R_L=30\ \mathrm{m},\quad \eta=3.97-1.79\left(\sec\left(\theta\right)-1\right),$$

$$r=\sqrt{\left(x_{\mathrm{core}}-x\right)^2+\left(y_{\mathrm{core}}-y\right)^2},$$

- Timing

$$t_r=t_o+t_{plane}+a\times\left(1+r/R_L\right)^{1.5}LDF\left(r\right)^{-0.5}$$

$$LDF\left(r\right)=f\left(r\right)/f\left(800\ \mathrm{m}\right)\qquad S\left(r\right)=S_{800}\times LDF\left(r\right)$$

Free parameters:

$x_{core},\ y_{core},\ \theta,\ \phi,\ S_{800},\ t_0,\ a$

Observables:

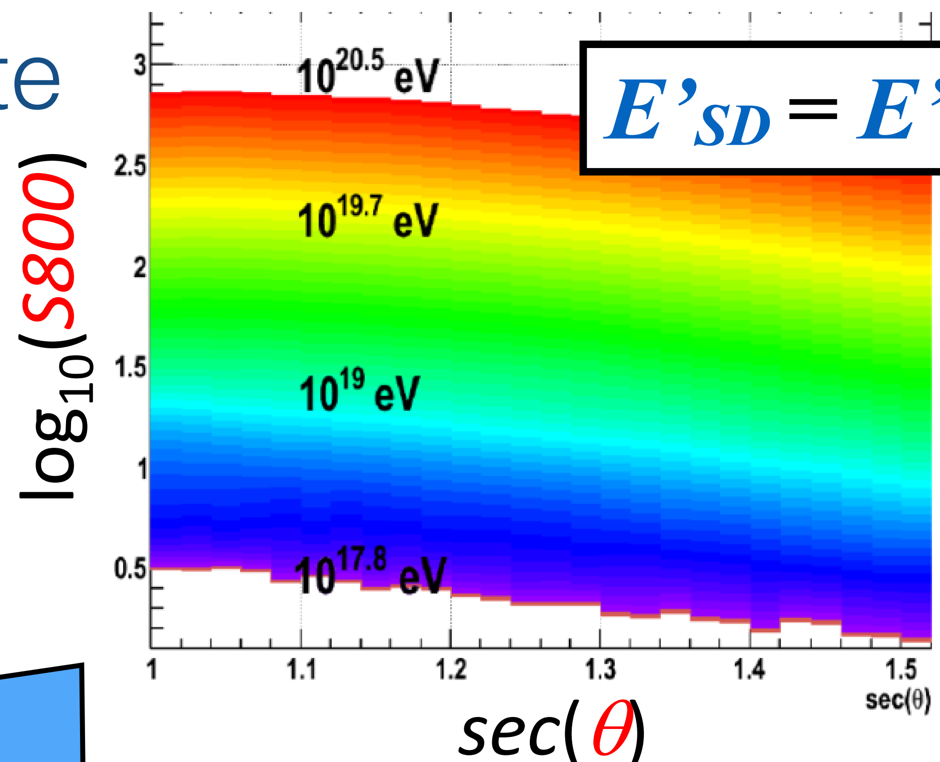
t_r - detector time

S_r - detector integral signal

Event reconstruction

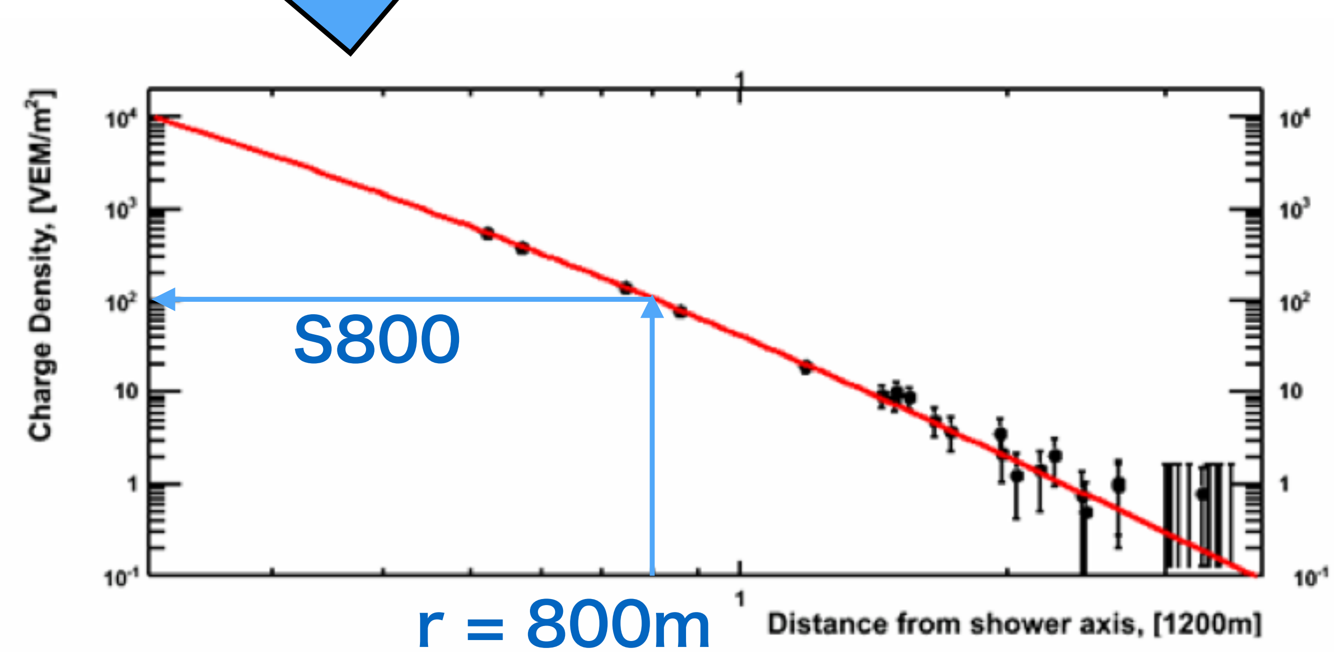
standard parametric approach

Energy estimate



$E'_{SD} = E'_{SD}(S800, \theta)$

- table function



Event reconstruction

Machine learning approach

Purpose (ideally): recover primary particle properties (arrival direction, energy, mass, …) as function of observables.

Direct observables in SD:

- Time series of the SD signals

Instruments:

- SD Monte-Carlo (EAS development and detector response)
- Artificial neural network (NN)
  - Can describe any continuous function of input data
  - Can be tuned using examples generated using Monte-Carlo

Event reconstruction

Machine learning approach

Purpose (ideally): recover primary particle properties (arrival direction, energy, mass, …) as function of observables.

In real life:

- observables depend on unknown/random factors
- NN function defines optimal test statistic
- obtain corrections to parametric reconstruction

Observables in SD:

- Time series of the SD signals

Instruments:

- SD Monte-Carlo (EAS development and detector response)
- Artificial neural network (NN)
  - Can describe any continuous function of input data
  - Can be tuned using examples generated using Monte-Carlo

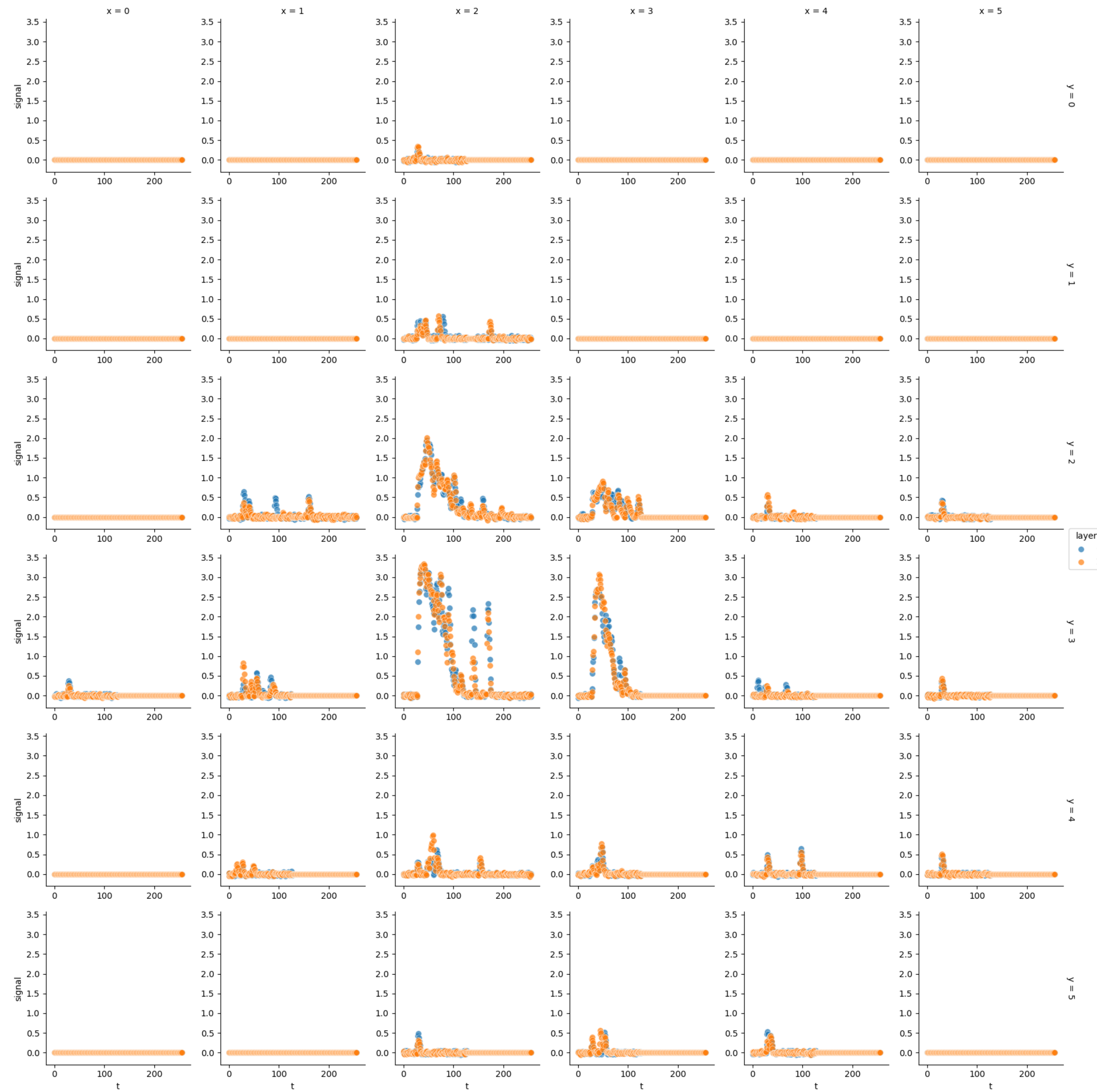
Method in nutshell

- Extract useful detector features using 1-D convolutions
- Treat detector network as a multichannel image using 2D convolution layers

M. Erdmann et al, Astropart.Phys. 97 (2018) 46-53

SD reconstruction NN architecture

event data



Dimensions:
(N,N,T,2)

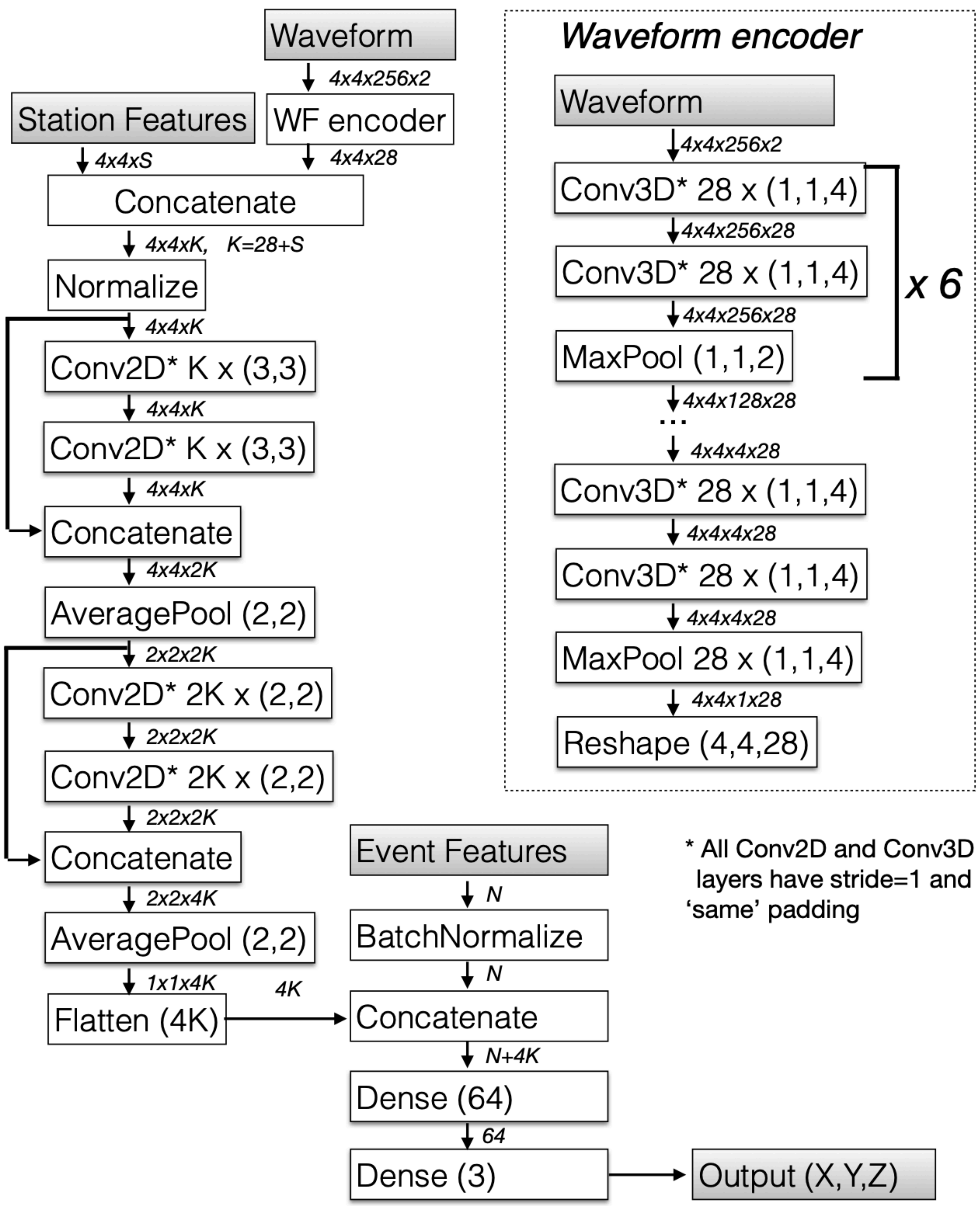
Waveform

detector layers

N=4-8, T=128-256

Standard SD reconstruction is used to center image around shower core

SD reconstruction NN architecture



- waveform encoder extracts useful features from readings of the two SD station layers
- the extracted features are passed to 2D-convolutional network along with SD station properties
- event features extracted by convolutional network are analysed along with 14 composition sensitive variables in the dense layer part of the model

D. Ivanov et al, Mach.Learn.Sci.Tech. 2 (2021) 1, 015006

SD station features

Produced by standard reconstruction procedure

- Detector signal saturation flag
- If detector is excluded from geometry fit (affected by random muon)
- x relative to shower core
- y relative to shower core
- z detector position altitude relative to common level
- detector signal, MIP
- time of the plane front arrival
- time of the waveform relative to the plane front

Event features

Produced by standard reconstruction procedure

Signal density at 800 m from the shower core, S_{800};
Linsley front curvature parameter, a;
Area-over-peak (AoP) of the signal at 1200 m;

Pierre Auger Collaboration, Phys.Rev.Lett. 100 (2008) 211101

AoP LDF slope parameter;
Number of detectors hit;
N. of detectors excluded from the fit of the shower front;
$\chi^2/d.o.f.$;
$S_b = \sum S_i \times r_i^b$ parameter for $b = 2.5, 3.0, 3.5, 4.0$ and $b = 4.5$;

Ros, Supanitsky, Medina-Tanco et al. Astropart.Phys. 47 (2013) 10

The sum of signals of all detectors of the event;
Asymmetry of signal at upper and lower layers of detectors;
Total n. of peaks within all FADC traces;
N. of peaks for the detector with the largest signal;
N. of peaks present in the upper layer and not in lower (and vice versa);

Training the model

- Minimizing mean square error
- Adaptive learning rate (adadelta optimizer arxiv 1212.5701)
- Number of training samples ~ 10^6 (100 GB data) - do not fit into RAM). hdf container is used и generator API in keras
- Number of weights to learn 10^5 - 10^6
- Regularization to avoid overfitting:
  - L2
  - dropout
  - noise layers
- Optimizing network architecture hyper-parameters (ray.tune & hyperopt)
- Hardware: NVIDIA GTX-2080 GPU
- Instruments: python, numpy, tensorflow, keras, h5py

EAS modelling

T. Abu-Zayyad et al. *Astrophys. J.*, 768:L1, 2013

- MC: CORSIKA
- HE hadronic interactions: QGSJETII-03 and QGSJETII-04 (test set only)
- LE hadronic interactions: FLUKA
- EM processes: EGS4
- Detector response: GEANT4
- Event sampling:
  - Energy sampling E^{-1}
  - Mass composition: H, He, N, Fe (1:1:1:1)
  - Isotropic primary flux with zenith angles < 45 degrees
  - Standard energy spectrum reconstruction cuts applied

Reconstruction comparison

in presence of unavoidable uncertainty

Explained variance score

$$EV(y, \hat{y}) = 1 - \frac{Var(y - \hat{y})}{Var(y)}$$

y - true value of quantity being predicted (in our case, error of parametric reconstruction)

$\hat{y}$ - model estimate of y

How to see that model does job

in presence of unavoidable uncertainty

Explained variance score

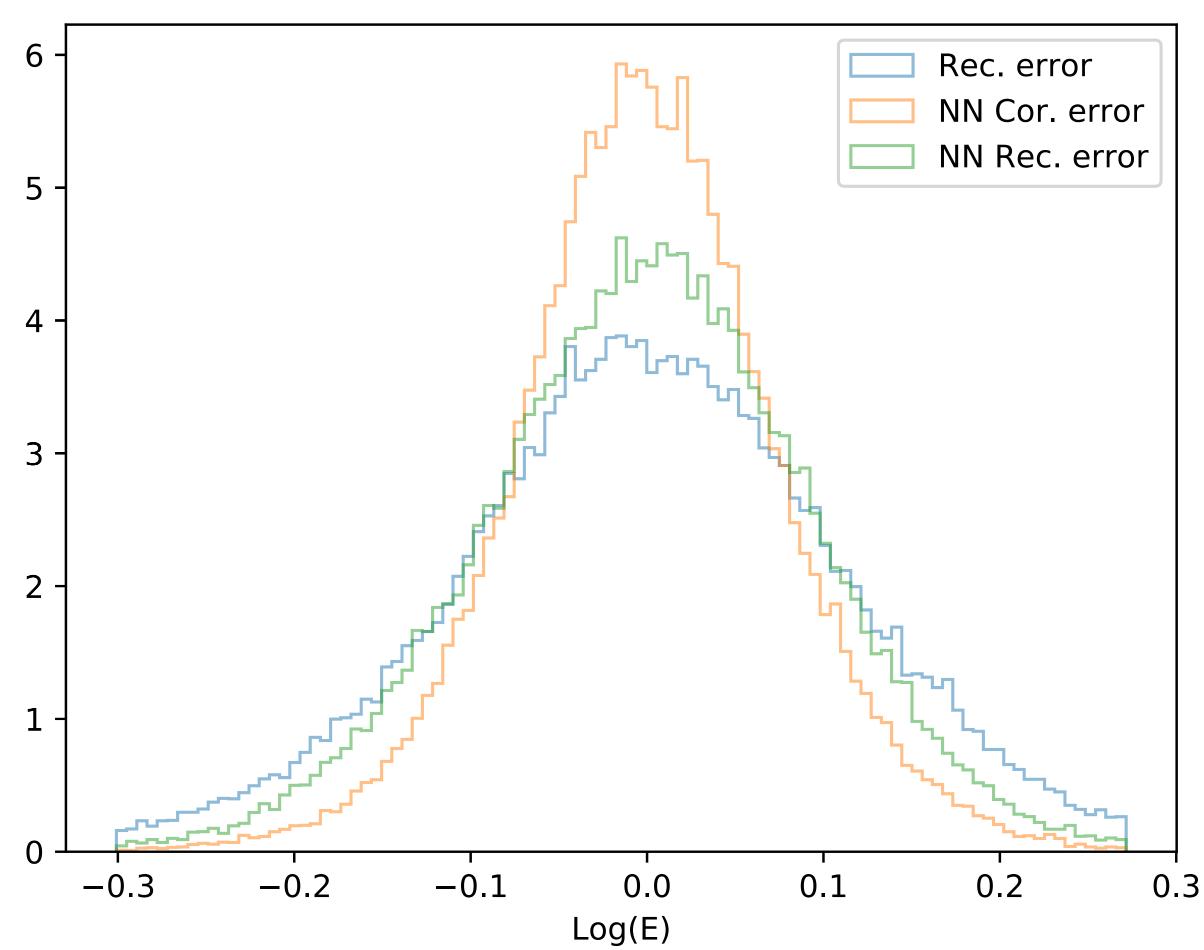
$$EV(y, \hat{y}) = 1 - \frac{Var(y - \hat{y})}{Var(y)}$$

y - true value of quantity being predicted (in our case, error of standard reconstruction)

$\hat{y}$ - model estimate of y

More visually:

Compare error distribution in two approximations



How to see that model does job

in presence of unavoidable uncertainty

Explained variance score

$$EV(y,\hat{y}) = 1 - \frac{Var(y-\hat{y})}{Var(y)}$$

y - true value of quantity being predicted (in our case, error of standard reconstruction)

$\hat{y}$ - model estimate of y

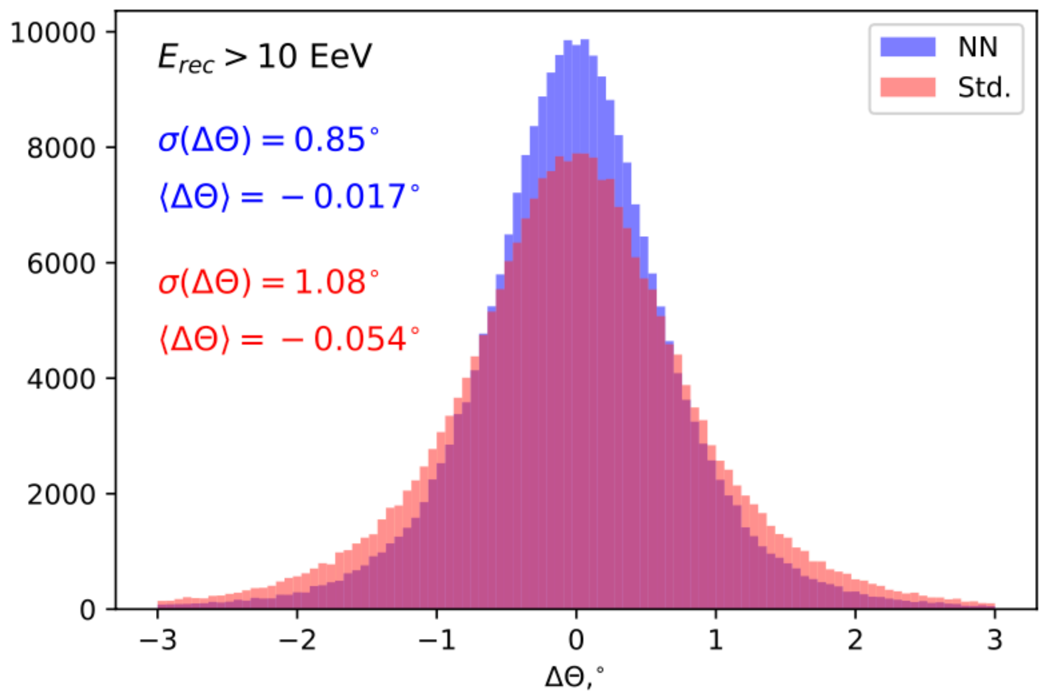
Explained variance for directional vector reconstruction

$(EV_X + EV_Y + EV_z)/3$

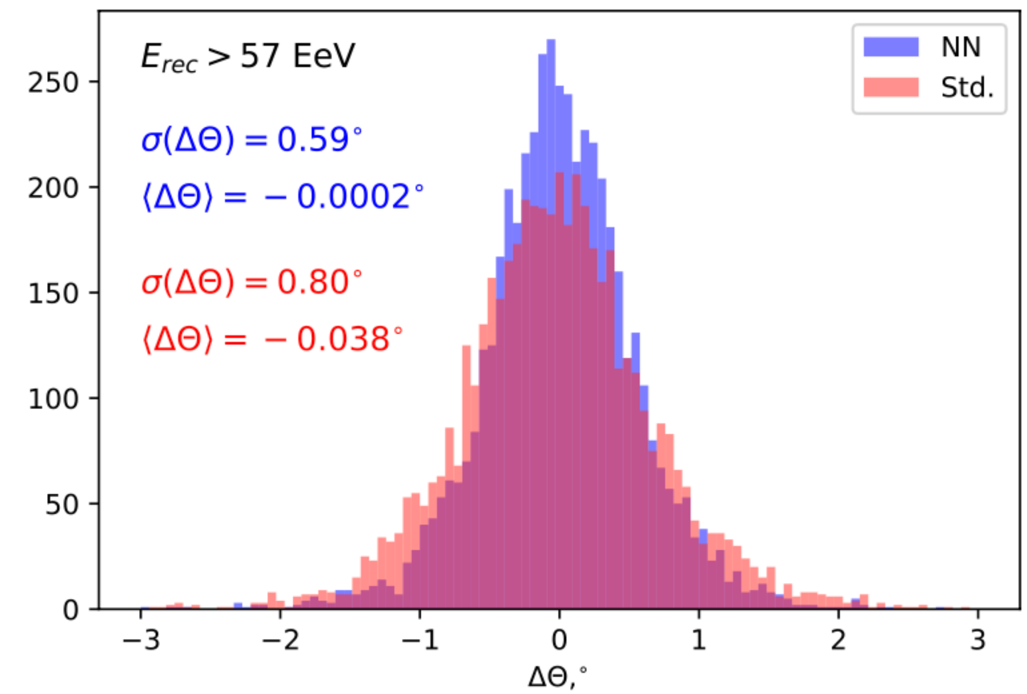
Event Features \ Waveform	Included	Excluded
Included	0.37	0.22
Excluded	0.30	0.11

Arrival direction reconstruction

Zenith angle reconstruction error distribution



E > 10 EeV

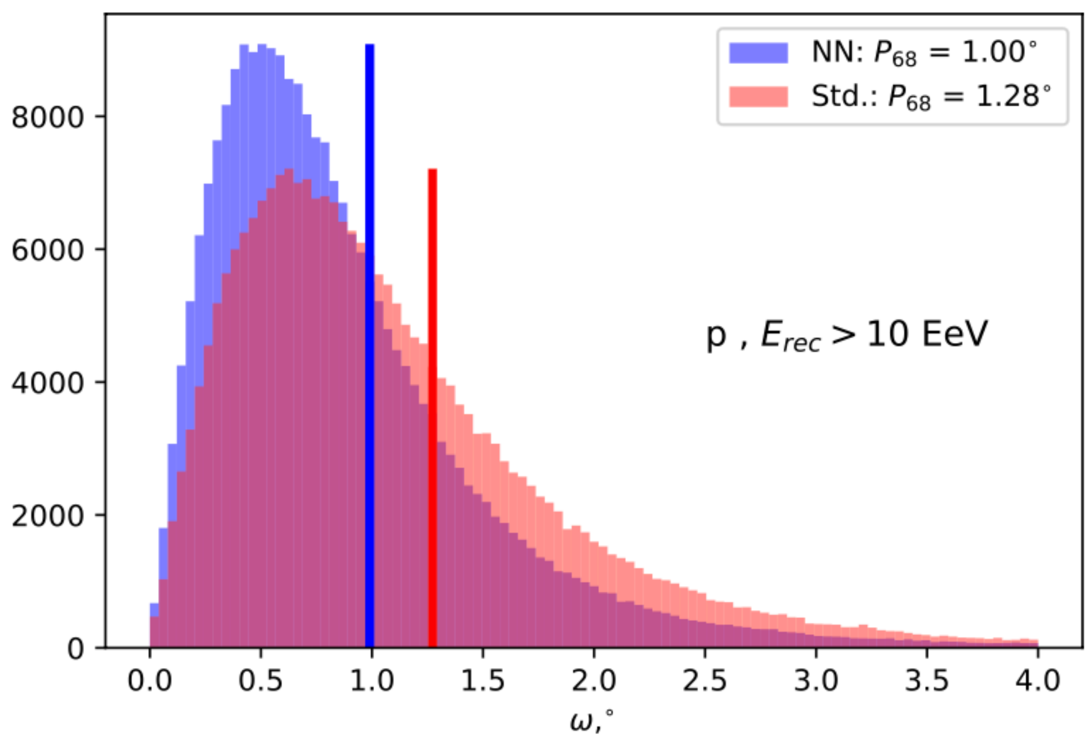


E > 57 EeV

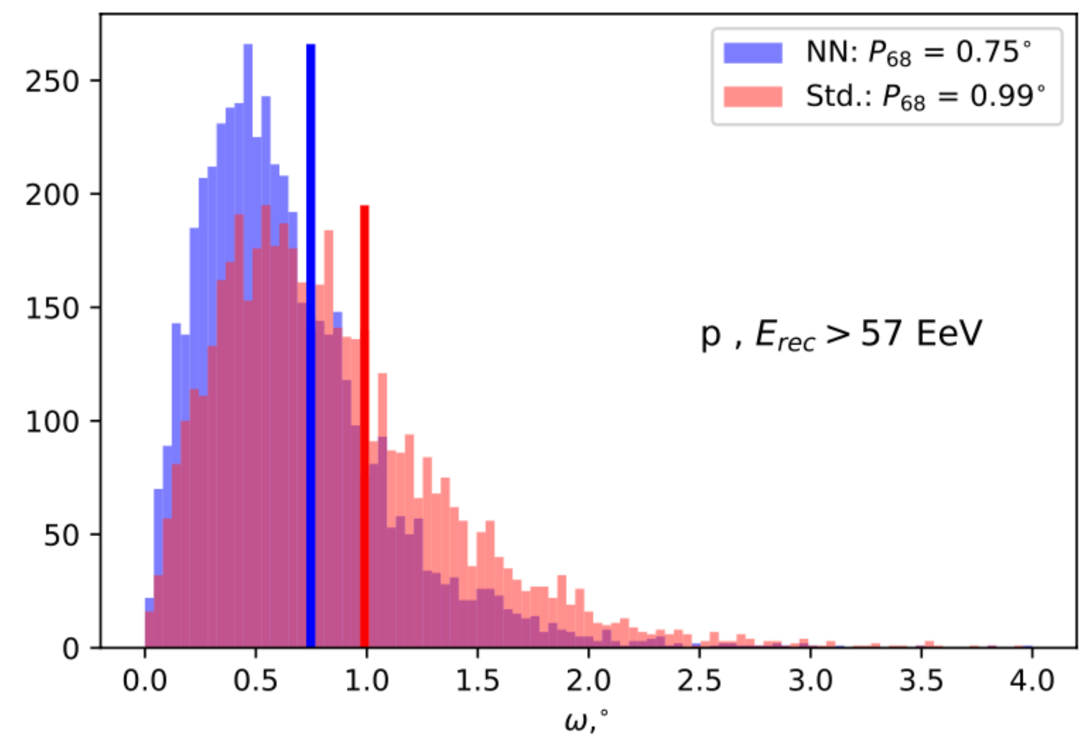
Proton Monte Carlo event set, QGSJETII-03 hadronic interaction model

Arrival direction reconstruction

Angular distance between true and reconstructed arrival direction



E > 10 EeV

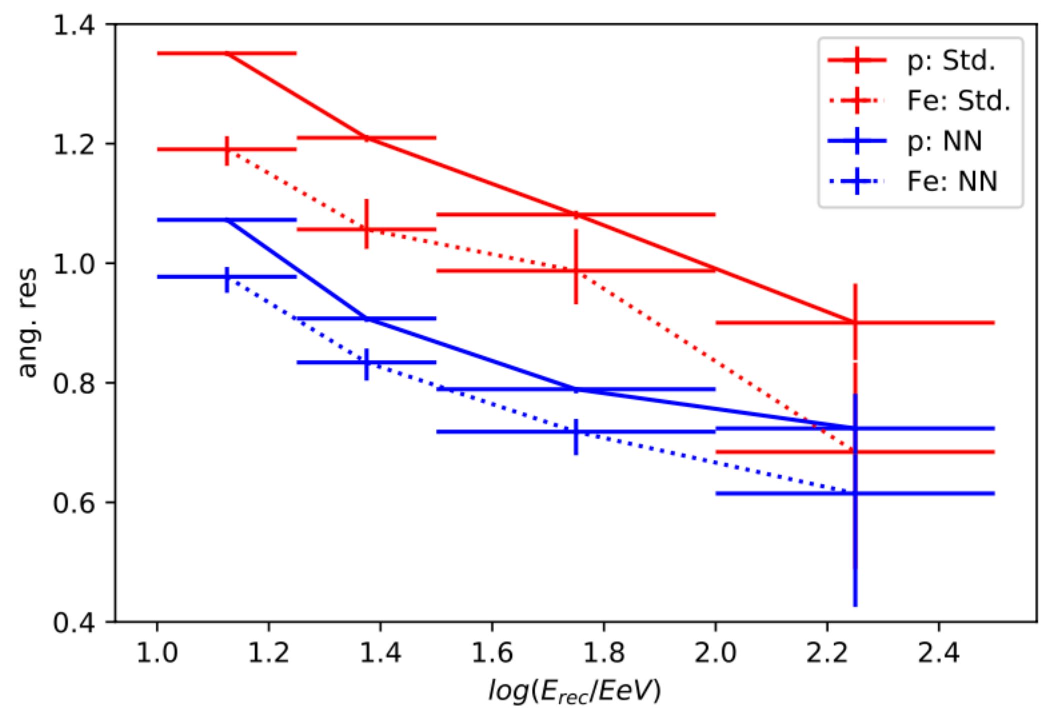


E > 57 EeV

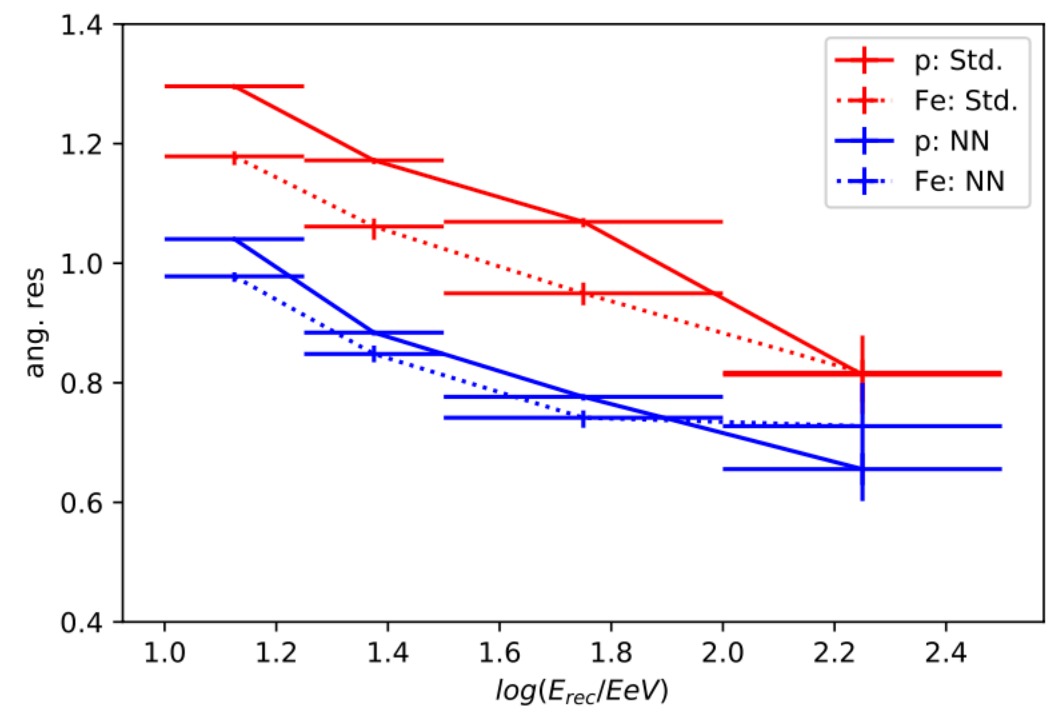
Proton Monte Carlo event set, QGSJETII-03 hadronic interaction model

Arrival direction reconstruction

Angular resolution dependence on energy



QGSJETII-03

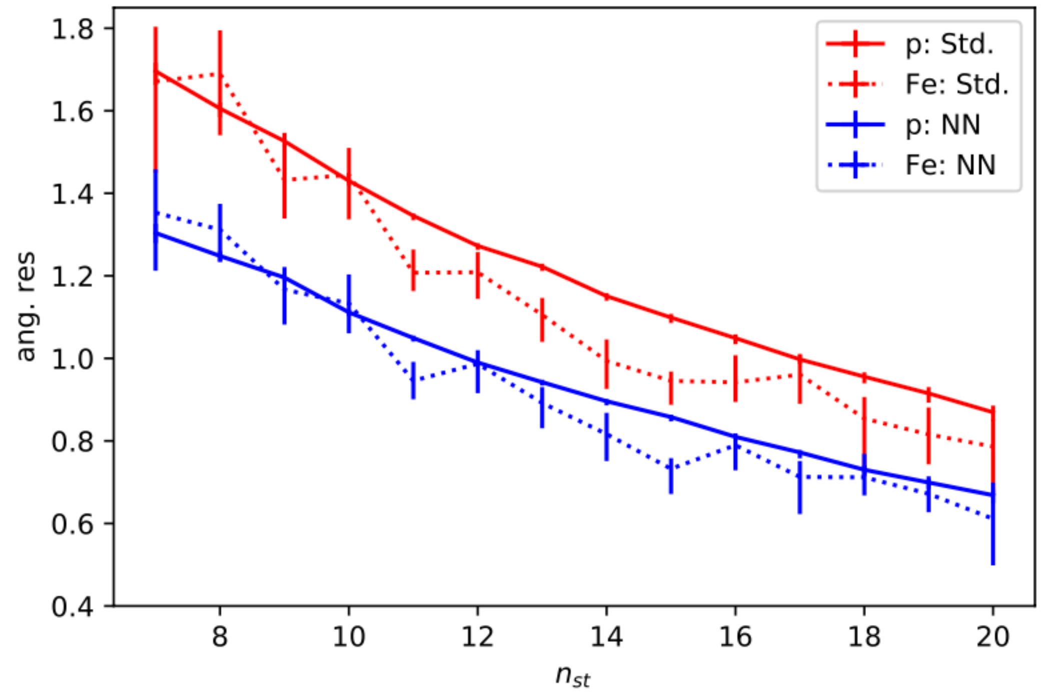


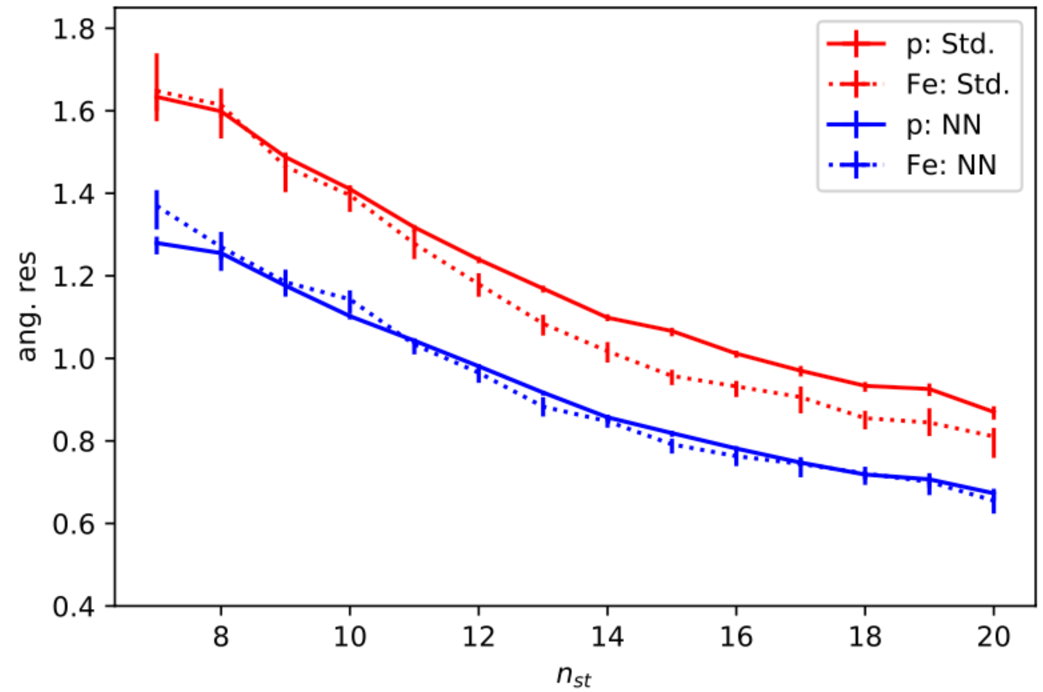
QGSJETII-04

Proton and iron Monte Carlo event sets

Arrival direction reconstruction

Angular resolution dependence on the number of detector stations triggered



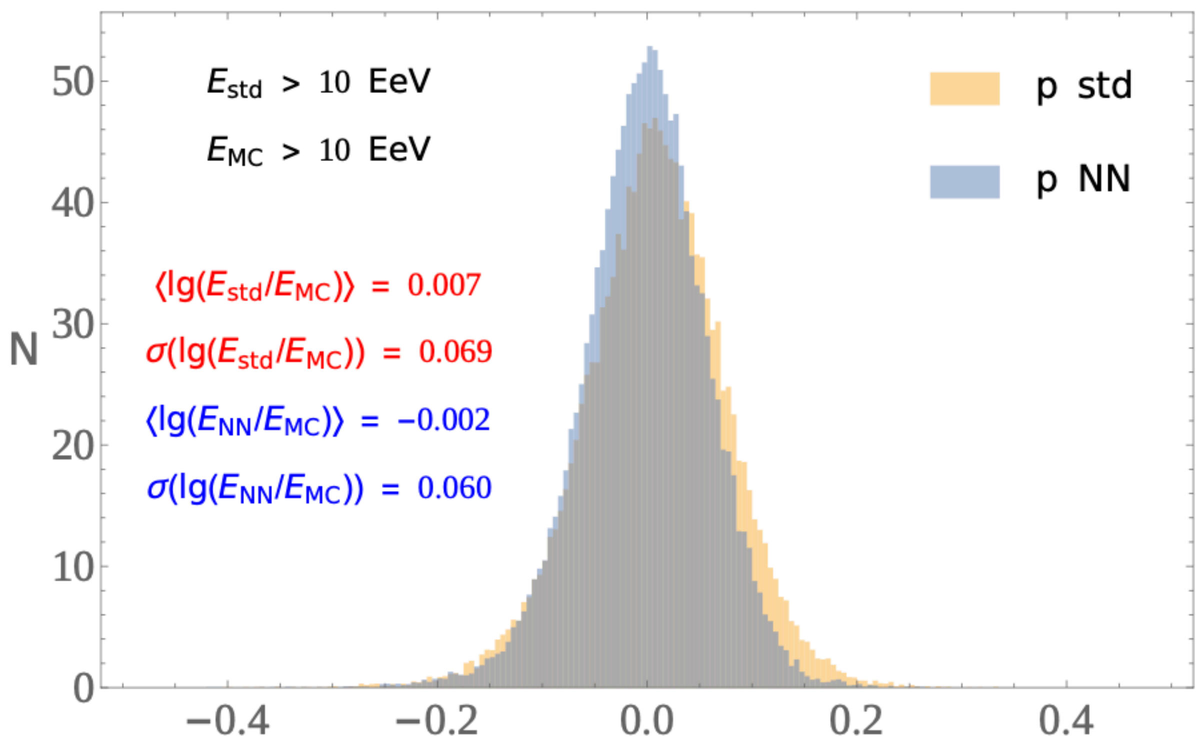


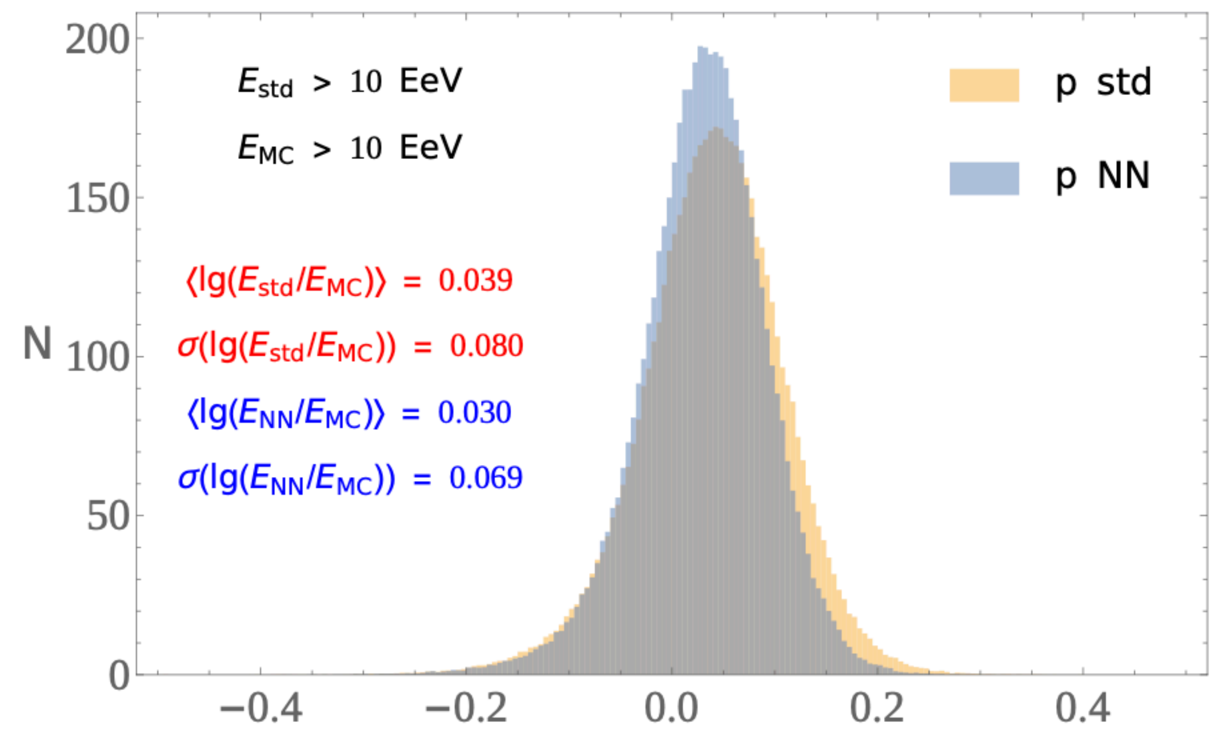
QGSJETII-03 QGSJETII-04

Proton and iron Monte Carlo event sets

Energy reconstruction

Energy reconstruction error distribution





proton Monte Carlo event set

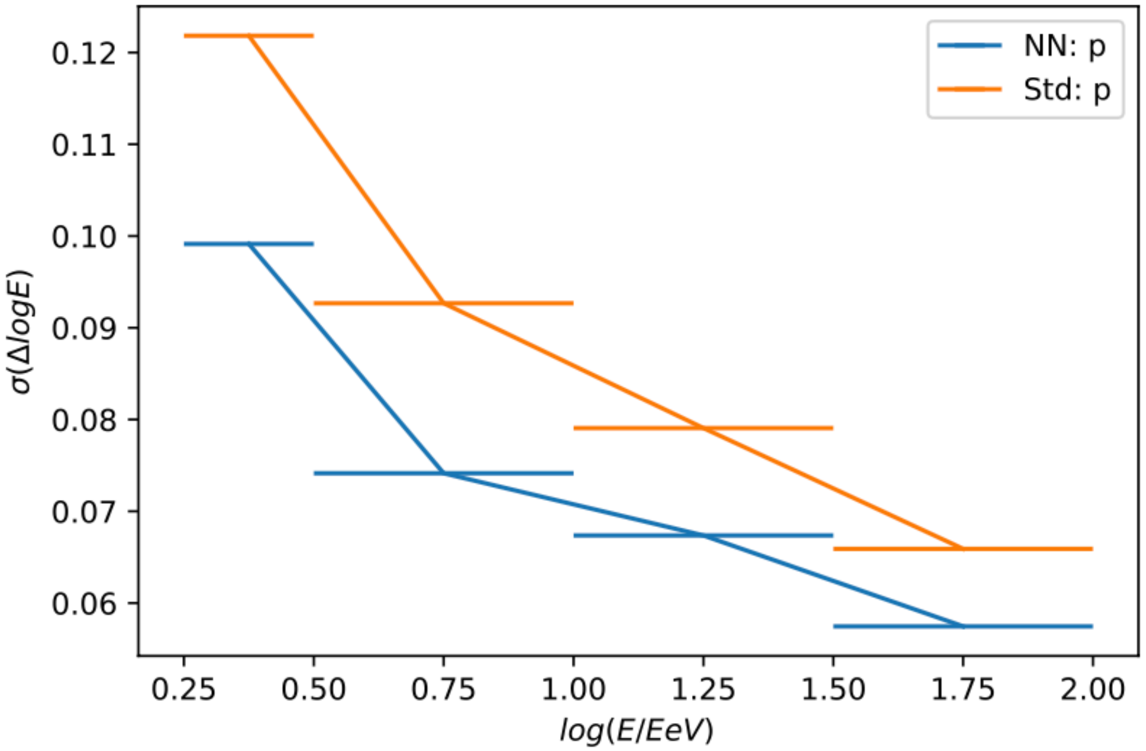
QGSJETII-03 QGSJETII-04

Both bias and variance decreased

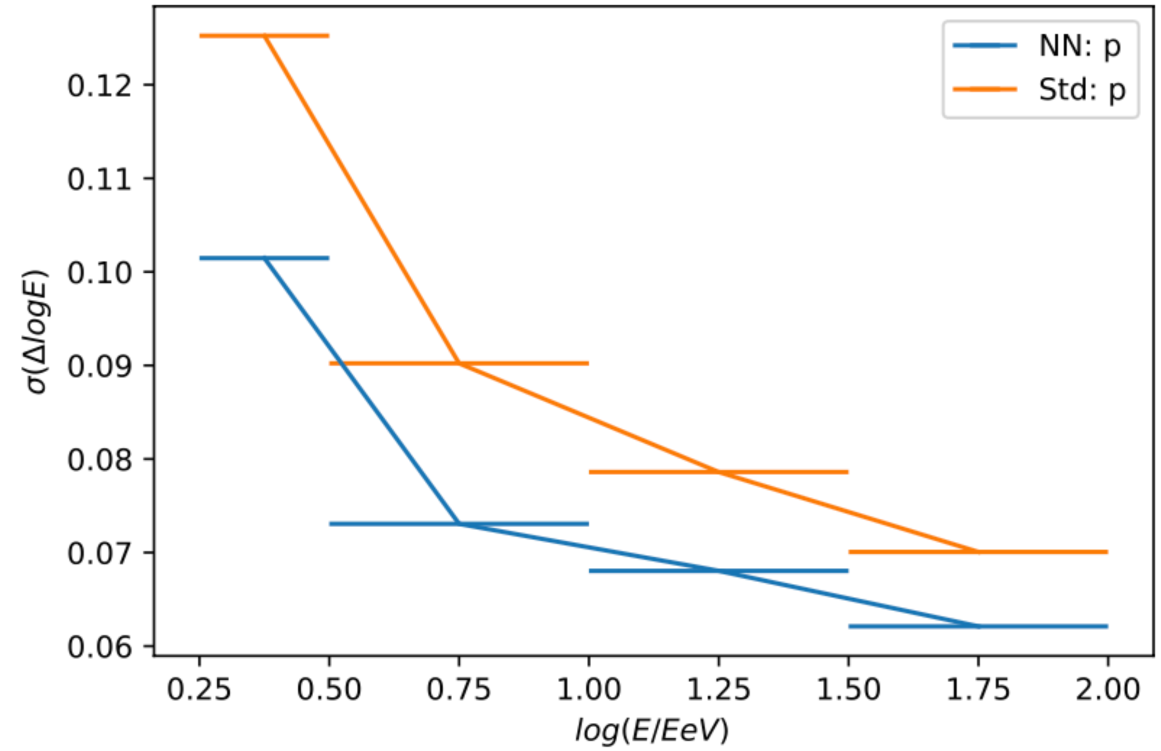
Systematic uncertainty due to hadronic model choice is comparable

Energy reconstruction

$\sigma(LogE - LogE_{MC})$



QGSJETII-03



QGSJETII-04

proton Monte Carlo event set

Conclusions

Energy reconstruction

+ log(E) resolution relative improvement is about 20%

— systematic bias due to uncertainty in the hadronic interaction model (or primary particle mass) is comparable

Arrival direction reconstruction

+ 20-25% improvement in angular resolution

+ we don’t see any systematic bias due to uncertainty in the hadronic interaction model

Conclusions

Energy reconstruction

+ log(E) resolution relative improvement is about 20%

— systematic bias due to uncertainty in the hadronic interaction model (or primary particle mass) is comparable

Arrival direction reconstruction

+ 20-25% improvement in angular resolution

+ we don't see any systematic bias due to uncertainty in the hadronic interaction model

Mass composition study

Mass composition study

Mass reconstruction on event by event basis:

- Classic regression task or classification task for N nuclei

- With p,He,N,Fe nuclei and NN described above we get accuracy ~ 35% (25% is random model result)

Mass composition study

Mass reconstruction on event by event basis:

- Classic regression task or classification task for N nuclei

- With p,He,N,Fe nuclei and NN described above we get accuracy ~ 35% (25% is random model result)

- Showers initiated by different nuclei are similar and highly stochastic. Primary particle mass mainly affects the expectation of the first interaction point depth (random quantity)

- with given observables there is no way to reconstruct primary particle mass on event by event basis

Mass composition study

Aim: estimate mass composition for a set of events:

- Train NN to reconstruct primary particle
- Use NN-classifier output as “optimal” composition sensitive observable

Mass composition study

Aim: estimate mass composition for a set of events:

- Train NN to reconstruct primary particle
- Use NN-classifier output as "optimal" composition sensitive observable

How to recover composition from model predictions:

- apply classifier to each individual event and count fractions
- correct fractions using model confusion matrix (linear model)

$$y_{pr} = C y_{true}$$

y_{pr} - predicted fractions of nuclei in control set

y_{true} - true fractions of nuclei in control set

$$y_{corrected} = C^{-1} y_{pr}$$

Mass composition study

Aim: estimate mass composition for a set of events:

- Train NN to reconstruct primary particle
- Use NN-classifier output as “optimal” composition sensitive observable

How to recover composition from model predictions:

- apply classifier to each individual event and count fractions
- correct fractions using model confusion matrix (linear model)

$$y_{pr} = C y_{true}$$

y_{pr} - predicted fractions of nuclei in control set

y_{true} - true fractions of nuclei in control set

$$y_{corrected} = C^{-1} y_{pr}$$

Problems:

- strong dependence of C^{-1} on control dataset
- negative predictions

Mass composition study

Aim: estimate mass composition for a set of events:

- Train NN to reconstruct primary particle
- Use NN-classifier output as “optimal” composition sensitive observable

How to recover composition from model predictions:

- apply classifier to each individual event and count fractions
- correct fractions using model confusion matrix (linear model)
- use some nonlinear model

Input: $E(p_A), Var(p_A)$ - average predicted “probabilities” for class A and their variance for $A = p, He, N, Fe$

Output: y_{true} - true fractions of nuclei

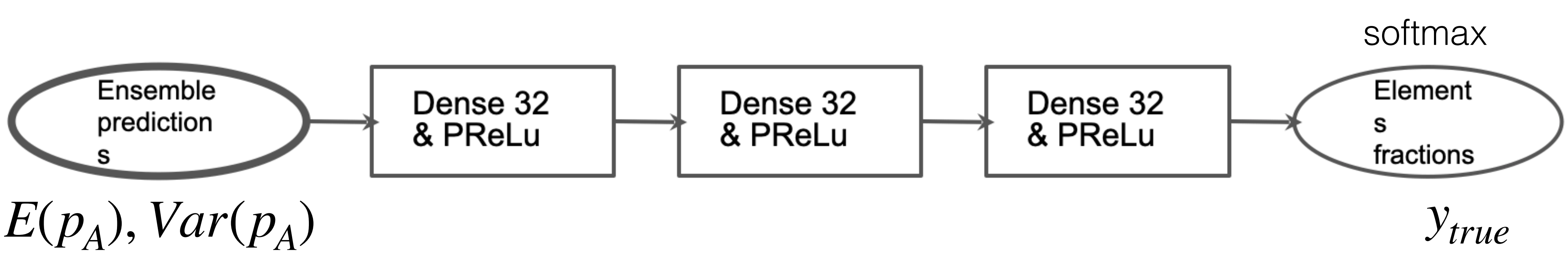
Mass composition study

Aim: estimate mass composition for a set of events:

- Train NN to reconstruct primary particle
- Use NN-classifier output as “optimal” composition sensitive observable

How to recover composition from model predictions:

- apply classifier to each individual event and count fractions
- correct fractions using model confusion matrix (linear model)
- use some nonlinear model



10 000 ensembles of events, consisting of 5 000 samples each used for training/validation

Mass composition study

Aim: estimate mass composition for a set of events:

- Train NN to reconstruct primary particle
- Use NN-classifier output as “optimal” composition sensitive observable

How to recover composition from model predictions:

- apply classifier to each individual event and count fractions
- use model normalised confusion matrix (linear model)
- use nonlinear model (another neural network ‘converter’)

Final result accuracy strongly depend on classifier accuracy

- try to enhance classifier

Mass composition study

Enhanced classifier

Uses

- Reconstruction parameters

+

- Spatial detectors data (without waveforms)

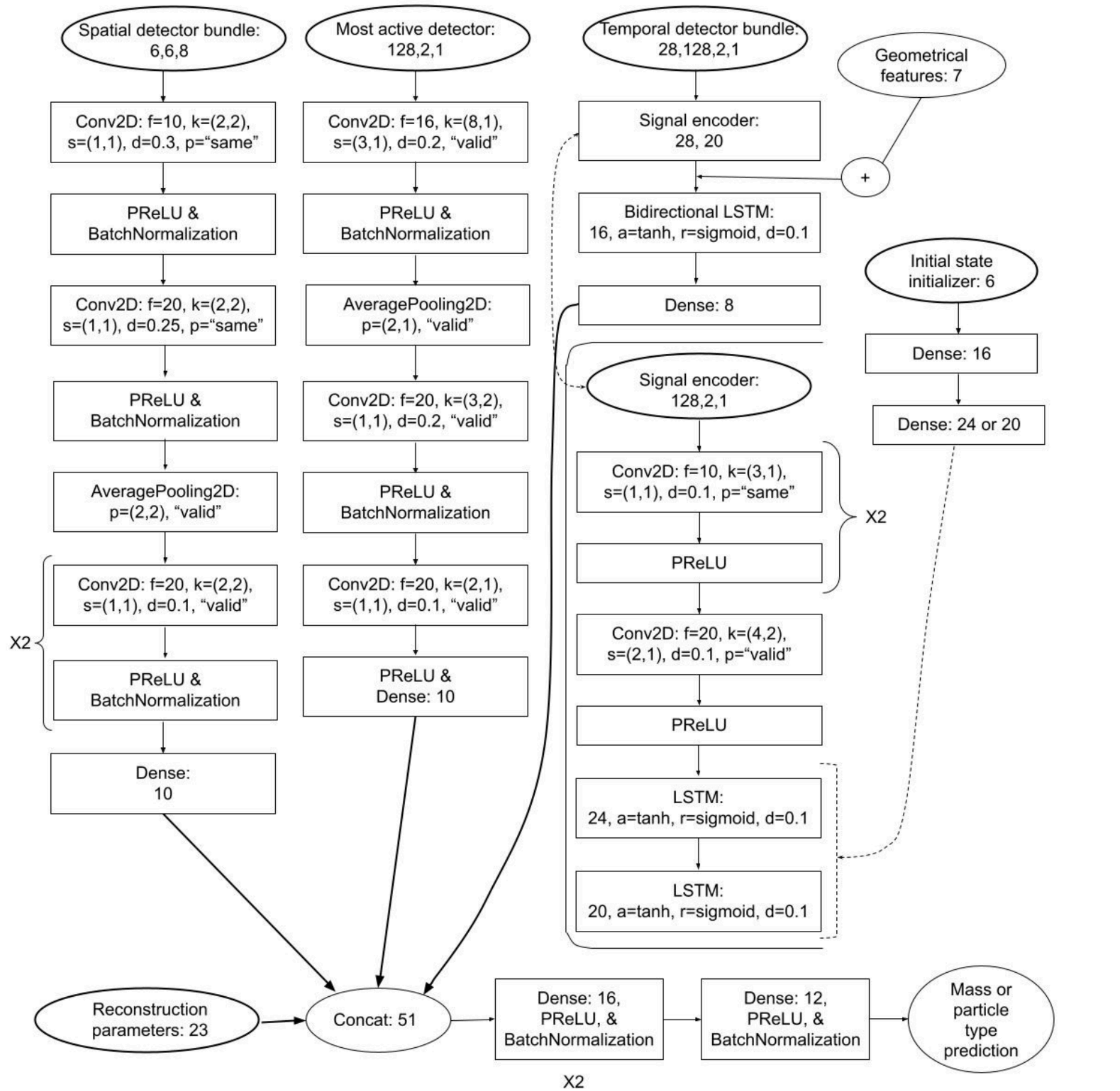
+

- Waveforms from the detector with largest signal

+

- Temporal detector data (waveforms only)

~35K trainable params



Mass composition study

Classifier accuracy (4 components): 40.2%

Separate block performance:

	temporal detector bundle	reconstruction parameters	spatial detector bundle	most active detector
accuracy	39.5%	33.8%	32.9%	31.0%

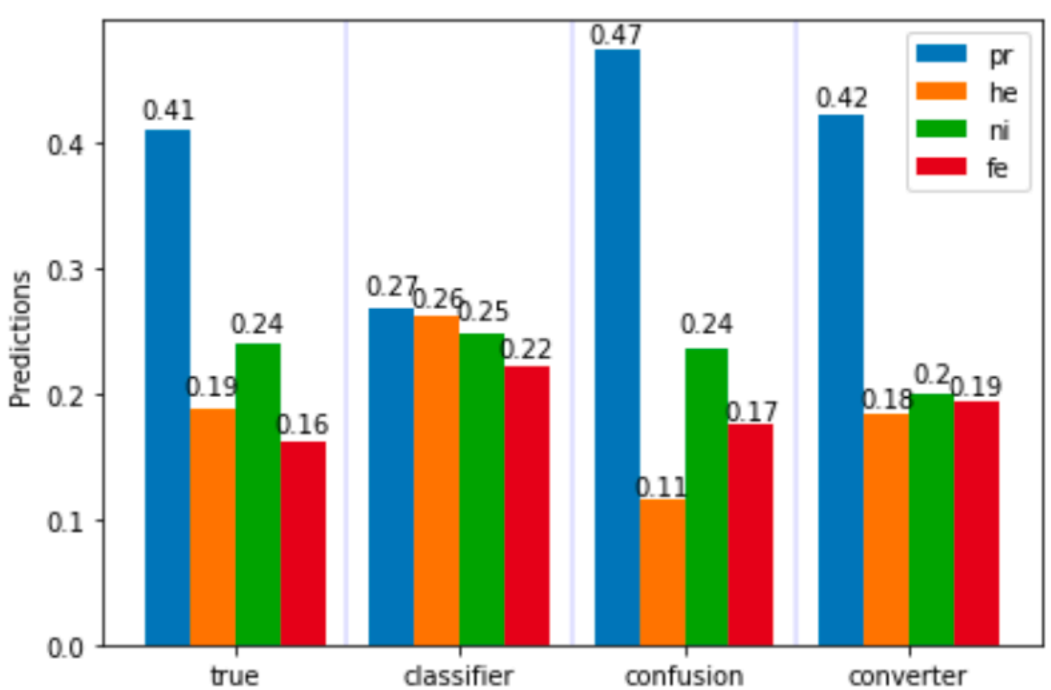
Mass composition study

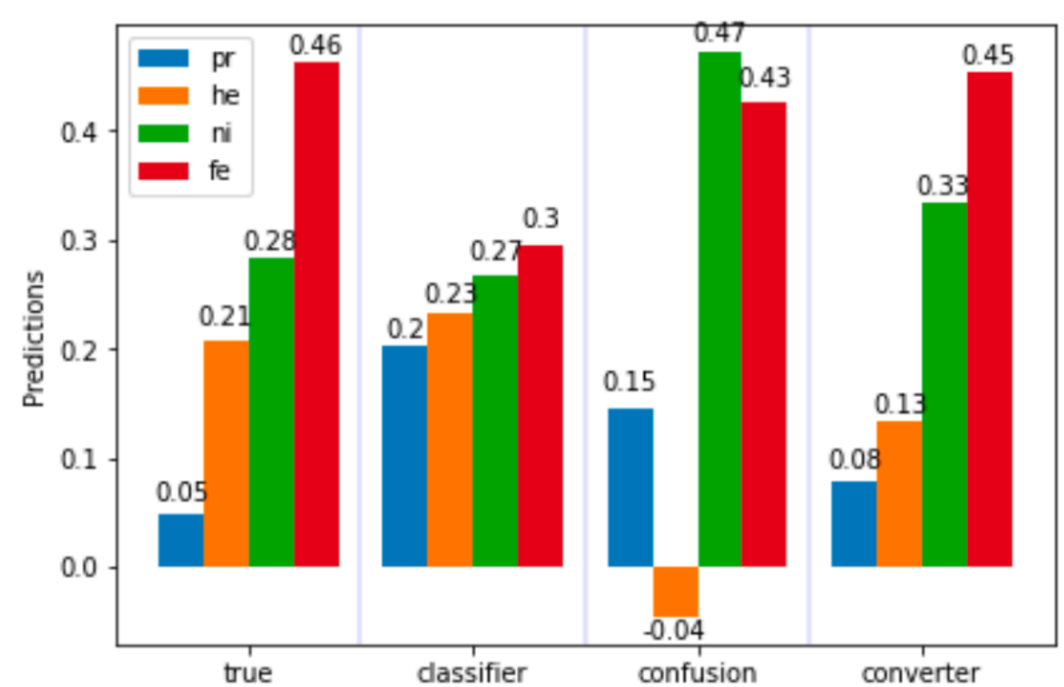
Classifier accuracy (4 components): 40.2%

Separate block performance:

	temporal detector bundle	reconstruction parameters	spatial detector bundle	most active detector
accuracy	39.5%	33.8%	32.9%	31.0%

Mass fraction calculation for two random compositions





Mass composition study

Classifier accuracy (4 components): 40.2%

Separate block performance:

	temporal detector bundle	reconstruction parameters	spatial detector bundle	most active detector
accuracy	39.5%	33.8%	32.9%	31.0%

Mass fraction calculation accuracy (4 components)

Average absolute error

based on 2000 test ensembles

	proton	helium	nitrogen	iron
classifier	0.10	0.11	0.11	0.09
confusion	0.06	0.14	0.12	0.04
converter	0.03	0.07	0.06	0.02

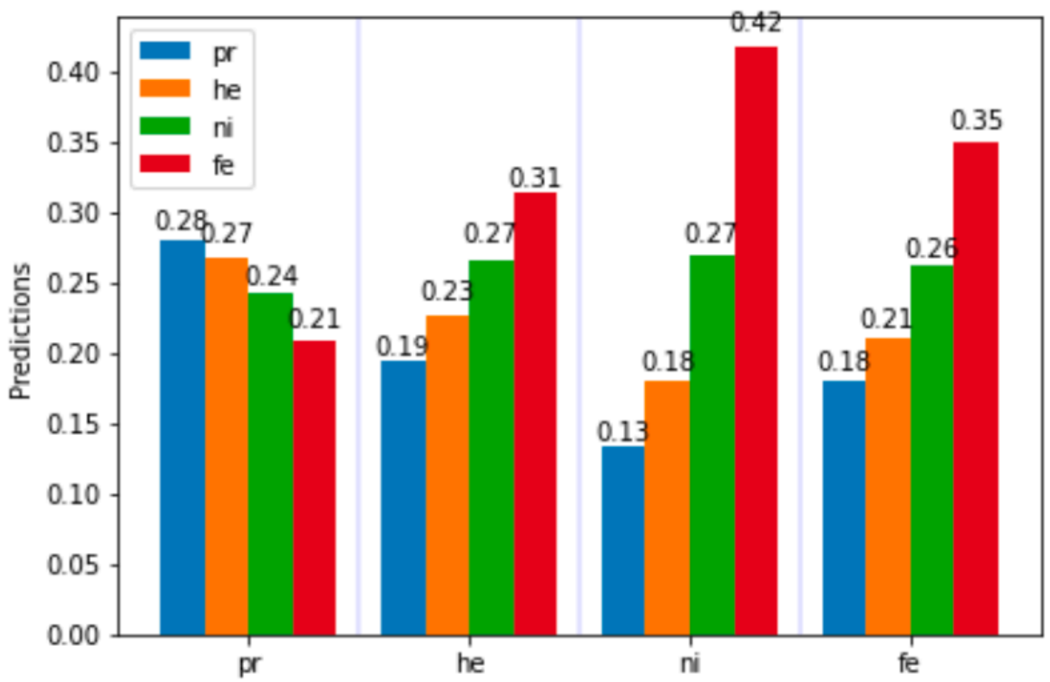
Average absolute error dependence on energy (with 'converter' model)

energy bin, log scale	18–18.25	18.25–18.5	18.5–18.75	18.75–19	> 19
averaged MAE	0.072	0.053	0.048	0.043	0.038

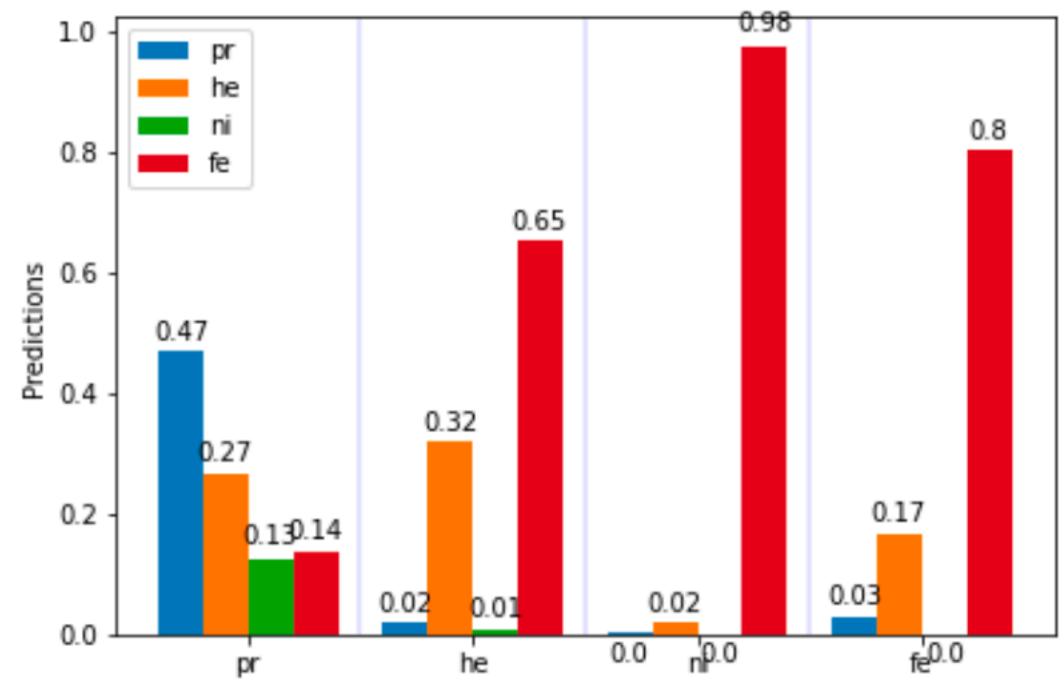
Mass composition study

Systematic uncertainty

Predictions on mono-composition data simulated using QGSJET-II 04 hadronic interaction model



averaged predictions of the classifier



‘converter’ predictions

primary particle	proton	helium	nitrogen	iron	averaged
reconstruction error	0.53	0.68	1.00	0.20	0.60

Method application is limited by systematics

Conclusions

Energy reconstruction

+ log(E) resolution relative improvement is about 20%

— systematic bias due to uncertainty in the hadronic interaction model (or primary particle mass) is comparable

Arrival direction reconstruction

+ 20-25% improvement in angular resolution

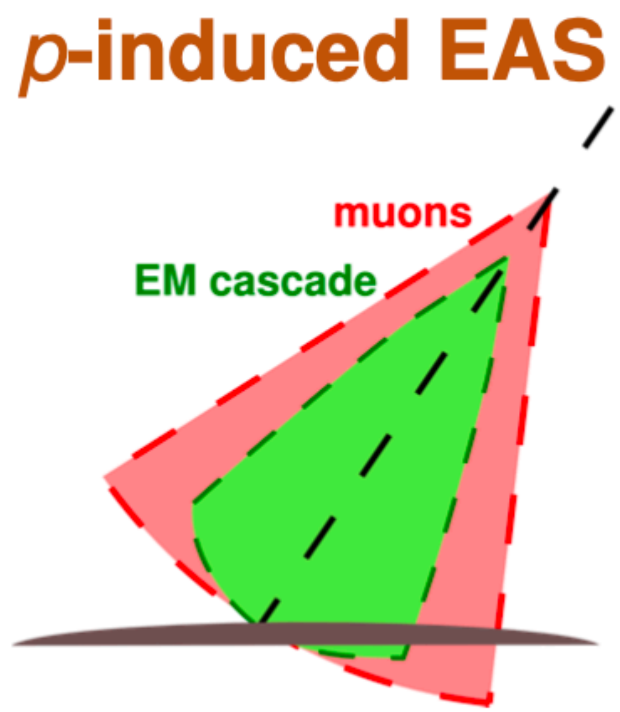
+ we don't see any systematic bias due to uncertainty in the hadronic interaction model

Mass composition study

+ Accuracy 7% (for p-He-N-Fe mixtures)

— Method application is limited by systematics

Search for UHE photons



Photon-induced showers:

- develop deeper in the atmosphere $\Rightarrow$ arrive younger
- contain less muons $\Rightarrow$ SD waveforms are less compressed

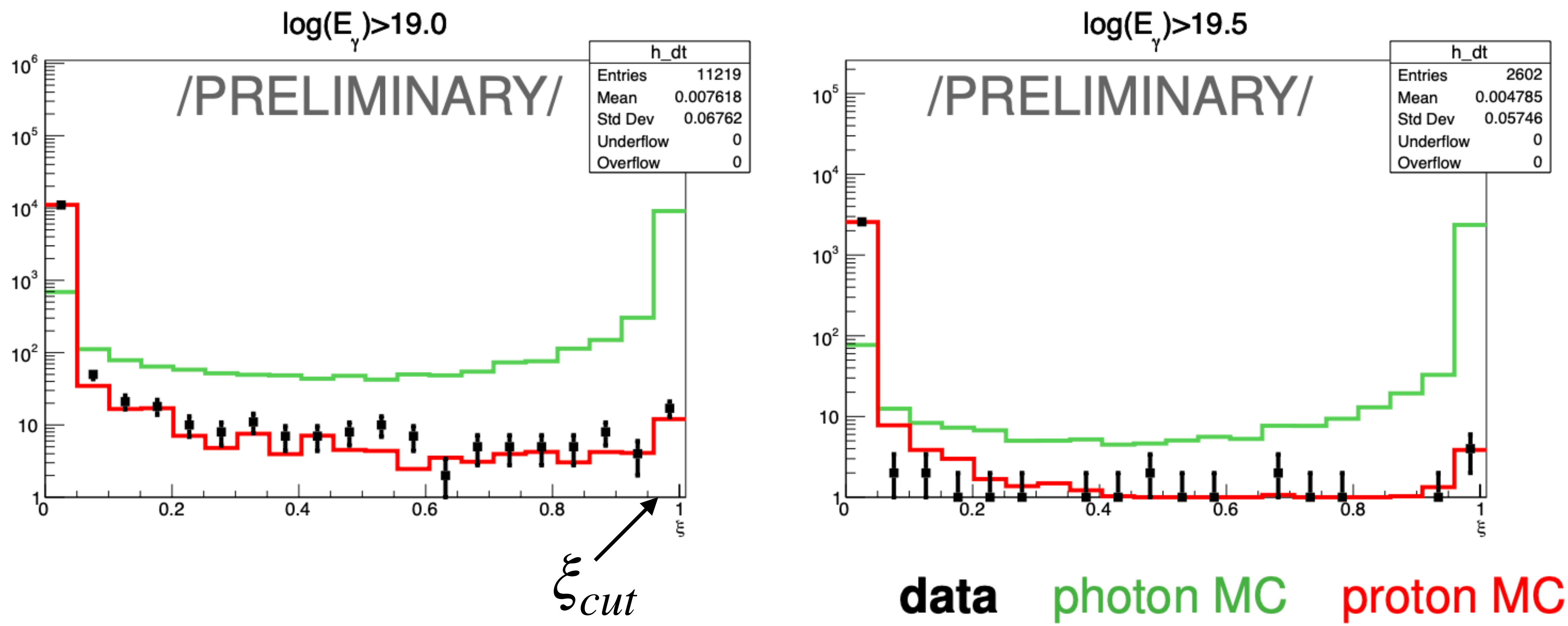
We use the neural-network classifier trained on both the

- **time-resolved waveforms**
- and derived features: front curvature, Area-over-peak, number of FADC signal peaks, $\chi^2/d.o.f.$, S_b

Search for UHE photons

- The p-γ classifier is trained with two Monte-Carlo sets:
  - γ-induced events (Signal)
  - proton-induced events (Background)
- The output of the classifier for each event is a number $\xi \in [0:1]$: 1 – pure signal (γ), 0 – pure background (p).
- We call "photon-candidates" events with $\xi > \xi_{cut}$.
- The optimal value of ξ_{cut} is obtained by the requirement of the strongest sensitivity in case null-hypothesis is valid, i.e. all events are protons.

Search for UHE photons



False positive errors have stronger effect on sensitivity than false negative

We adjust sample weights: $w_p / w_\gamma \simeq 10$

Search for UHE photons

- Data collected by TA surface detector for the 11 years:
  2008-05-11 – 2019-05-10
- *p* and γ Monte-Carlo sets with CORSIKA and dethinning

Stokes et al, Astropart.Phys.35:759,2012

Cuts for both data and MC:

- 7 or more detectors triggered
- core distance to array boundary is larger than 1200m
- $\chi^2/\text{d.o.f.} < 5$
- $\theta < 55°$
- $E_\gamma > 10^{19.0}$ eV (E_γ is estimated with photon Monte-Carlo)
  - or $E_\gamma > 10^{18.5}$ eV for training Monte-Carlo sets

11327 events after cuts

MC set is split into 3 parts: (I) 80% of events, for training the classifier, (II) for testing and cut optimization, (III) for exposure estimate.

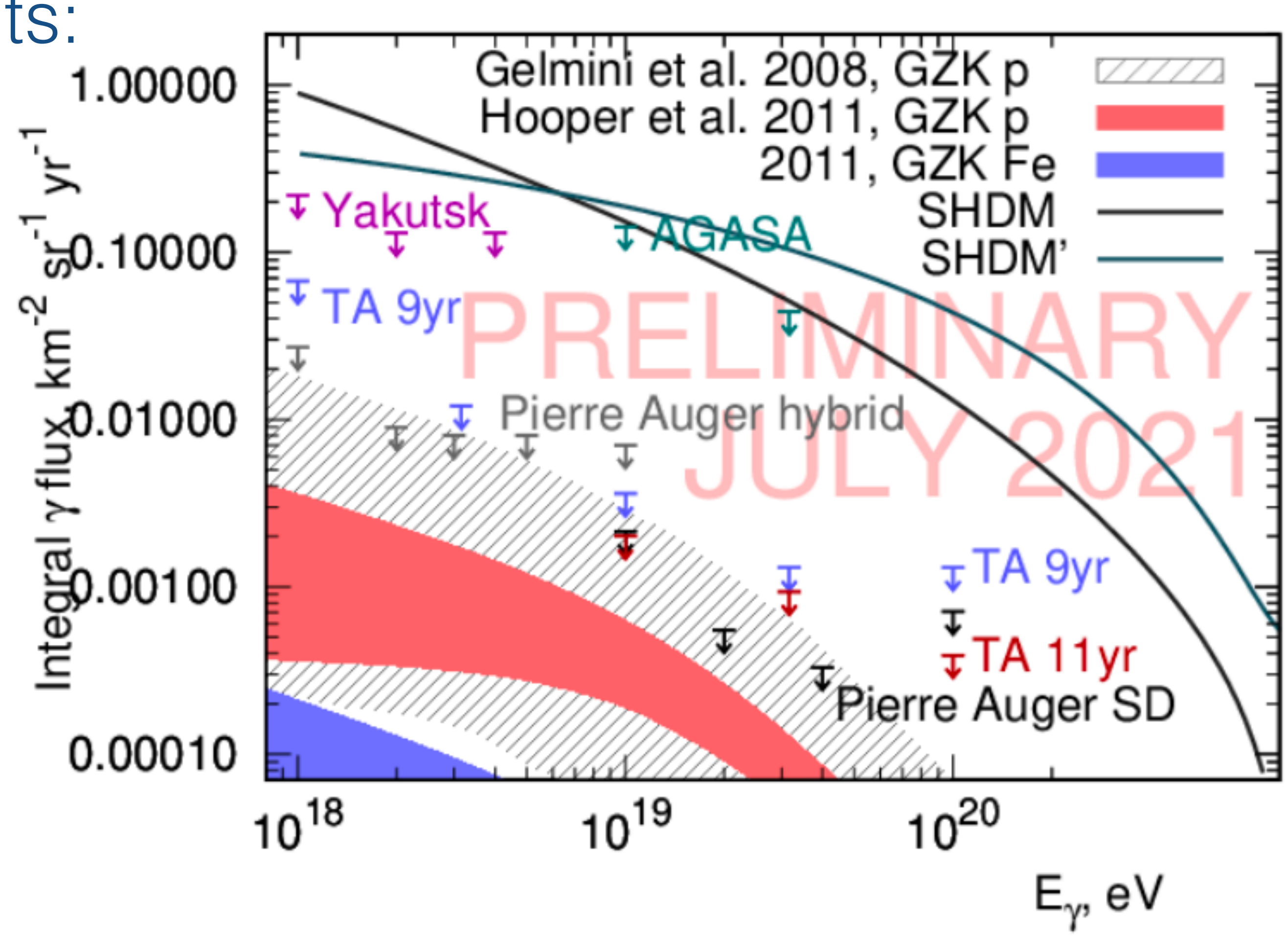
Search for UHE photons

E_0, eV	$10^{19.0}$	$10^{19.5}$	$10^{20.0}$
γ candidates	~~16~~ 2	~~11~~ 1	~~5~~ 0
$\bar{n}$ <	6.72	5.14	3.09
A_{eff}	3428	5546	7875
F_γ <	2.0×10^{-3}	9.3×10^{-4}	3.9×10^{-4}

Terrestrial Gamma-Ray Flashes candidate events are time correlated with the lightnings registered by National Lightning Detection Network.

TA collaboration, JGR Atmospheres (2020)

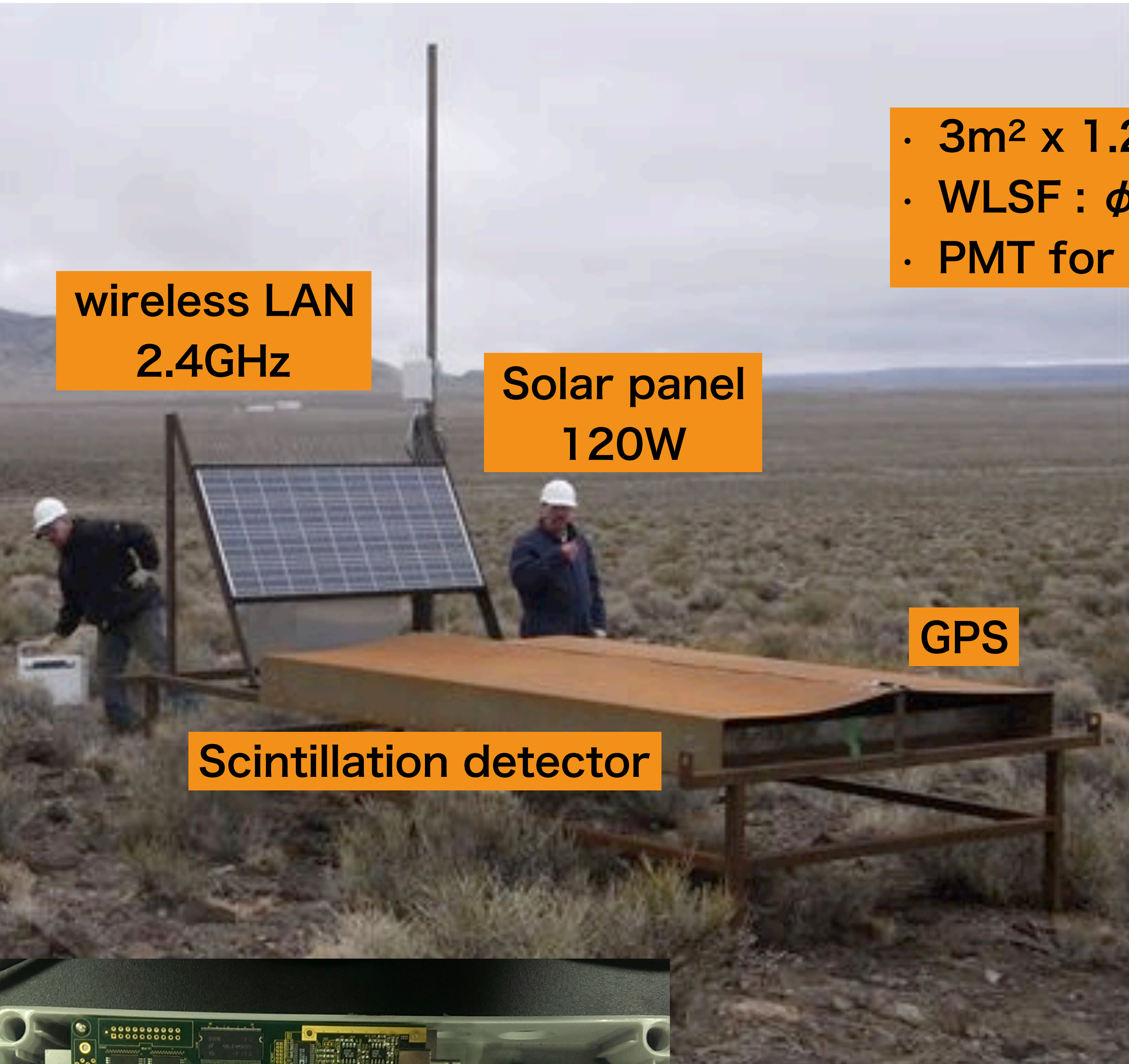
95% CL Limits:

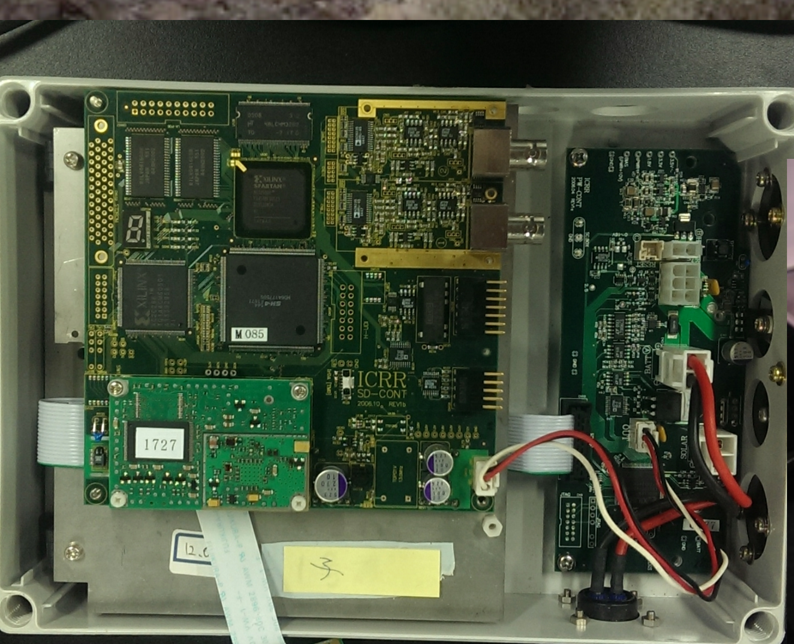


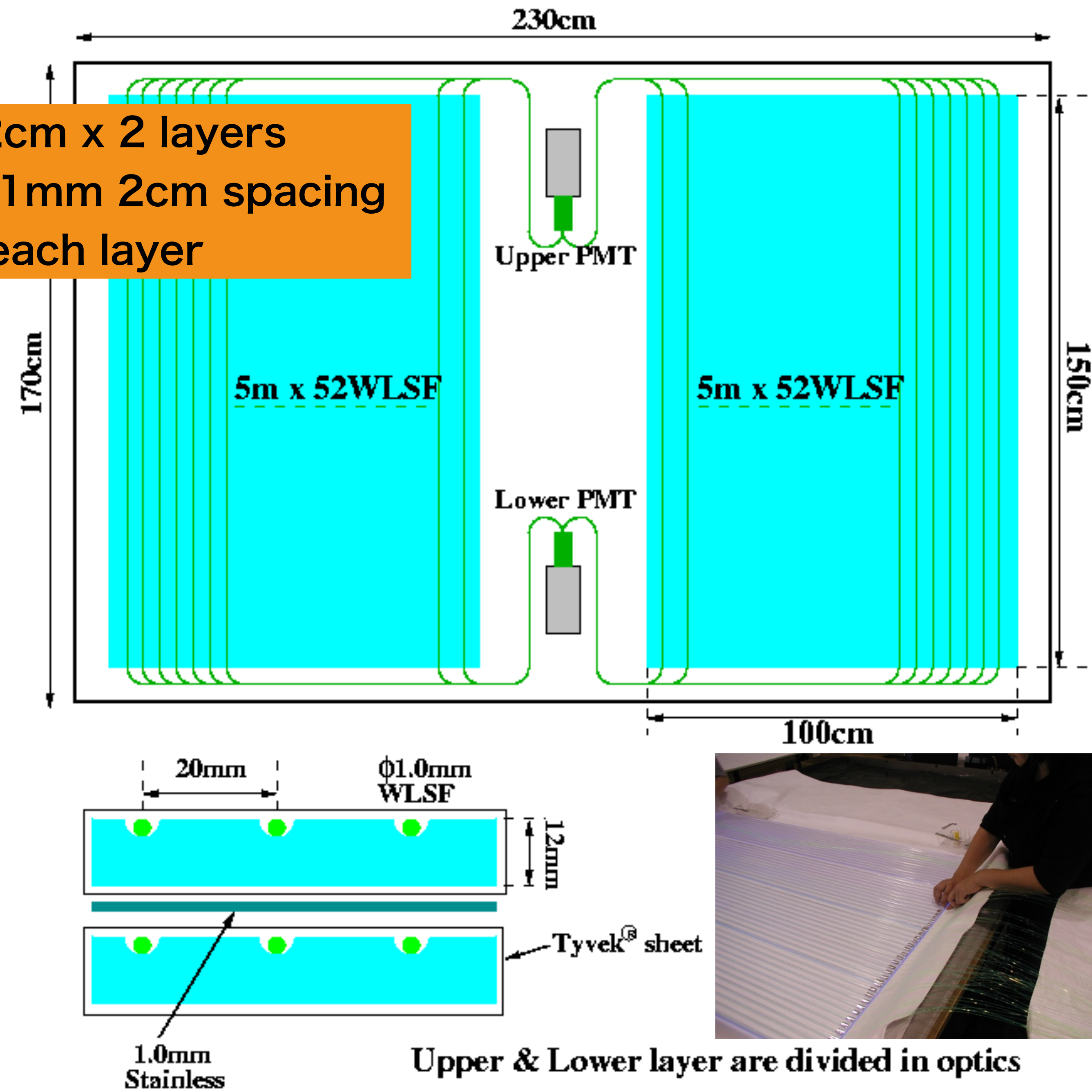
Appendix

Surface Detectors

- 3m² x 1.2cm x 2 layers
- WLSF : ϕ1mm 2cm spacing
- PMT for each layer







- 12bit 50MHz FADC x 2 layers
- CPU : Renesas SH4(25MHz)
- GPS, WLAN-modem
- Charge controller

SD reconstruction NN architecture

Problem: how to better take absent/not functioning detectors

- The event may occur close to detector network boundary
- Part of detectors may be turned off

SD reconstruction NN architecture

Problem: how to better take absent/not functioning detectors

- The event may occur close to detector network boundary
- Part of detectors may be turned off

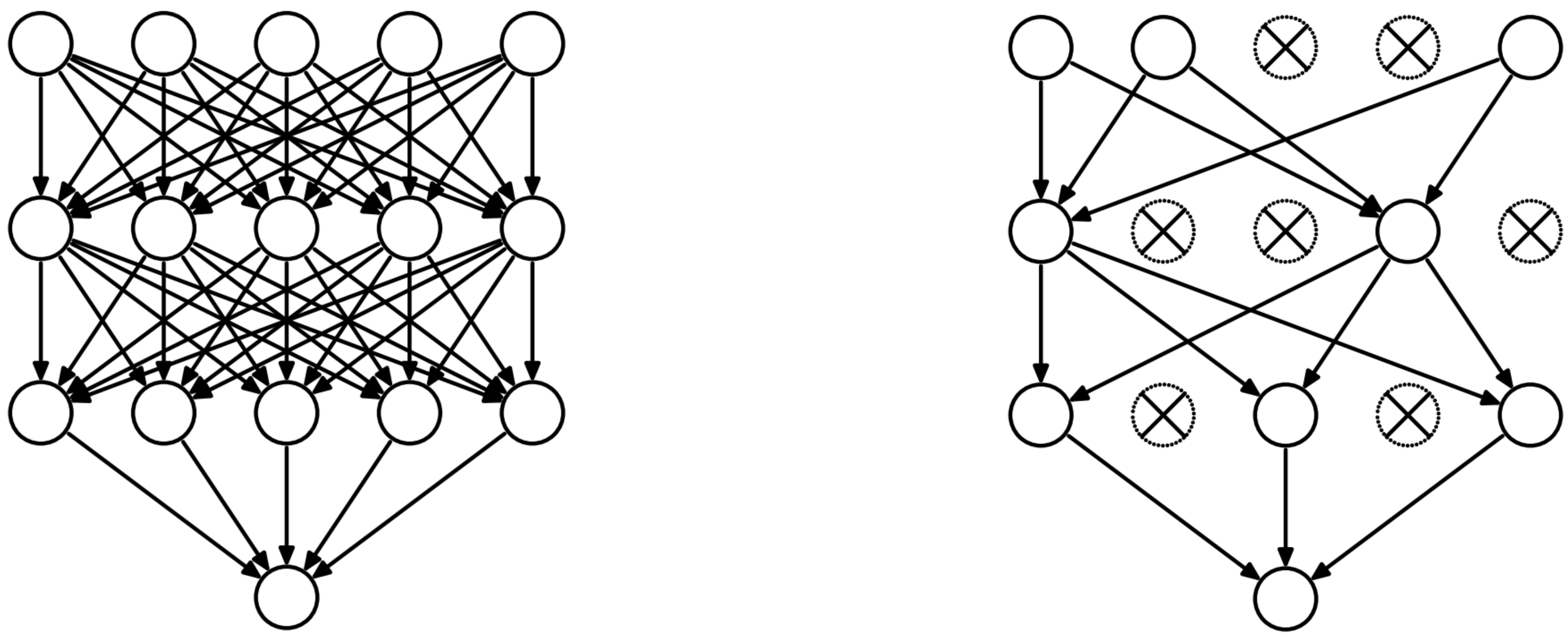
Dropout, the regularisation method in NN, simulate this situation

SD reconstruction NN architecture

Dealing with absent/not working detectors:

Dropout, the regularisation method in NN

Srivastava et. al JMLR 15 (2014)

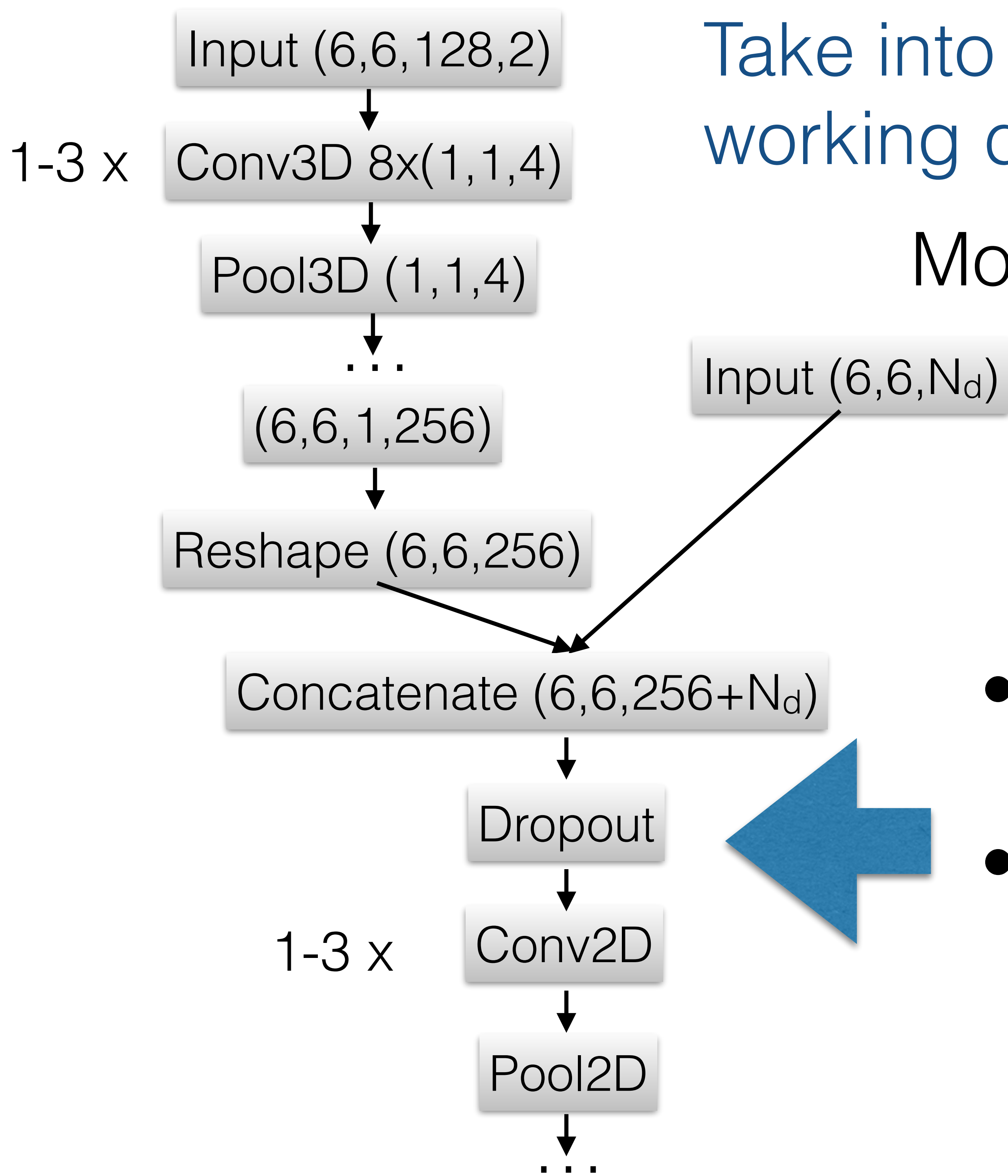


- In training mode neurons are switched off with probability *p*
- *For p=0.5* we train simultaneously 2^n thinned neural networks
- In prediction mode neurons are on but their output weights are multiplied by *p (we average predictions of thinned nets)*

SD reconstruction NN architecture

Take into account absent/not working detectors

Modified dropout



- Weights are corrected using fraction of working detectors
- In training mode part of the detectors may be switched off as in conventional dropout method

Cuts applied on MC samples.

- 1. Each event must include at least 5 counters.
- 2. The reconstructed primary zenith angle must be less than 45° .
- 3. The reconstructed event core must be more than 1200 m from edge of the array.
- 4. Both the timing and lateral distribution fits must have χ 2/degree of freedom value less than 4.
- 5. The angular uncertainty estimated by the timing fit must be less than 5° .
- 6. The fractional uncertainty in S(800) estimated by the lateral distribution fit must be less than 25%.

Search for UHE photons

- Geometric exposure for $\theta \in (0^\circ, 55^\circ)$: **13221 km^2 sr yr**
- Effective exposure is estimated using photon MC assuming E^{-2} primary spectrum

E_0	quality cuts	$\xi > \xi_{cut}$	A_{eff} km^2 sr yr
$10^{19.0}$	43.7%	**59.4%**	**3428**
$10^{19.5}$	52.0%	**80.7%**	**5546**
$10^{20.0}$	64.3%	**92.7%**	**7875**

- Efficiency of photon candidate selection ($\xi > \xi_{cut}$) has substantially grown compared to the previous analysis with BDT classifier – 16.2%, 37.2% and 52.3% for $\log_{10} E_0 = 19.0, 19.5$ and 20.0, correspondingly.

TA Collaboration, Astroparticle Physics 110 (2019) 8