Interpreting and Controlling Latent Space Parameters of Autoencoders

Yulia Dubenskaya[1], Stanislav Polyakov[4], Alexander Kryukov[1], Andrey Demichev[1], Pavel Volchugov[1], Elizaveta Gres[2], Dmitry Zhurov[3], Evgeny Postnikov[1], Alexander Razumov[1]

[1] Skobeltsyn Institute of Nuclear Physics, Moscow State University
[2] Research Institute of Applied Physics, Irkutsk State University
[3] Irkutsk State University
[4] Institute for Informatics and Automation Problems, National Academy of Sciences of the Republic of Armenia

The work was supported by RSF, grant no.24-11-00136.

07.2025

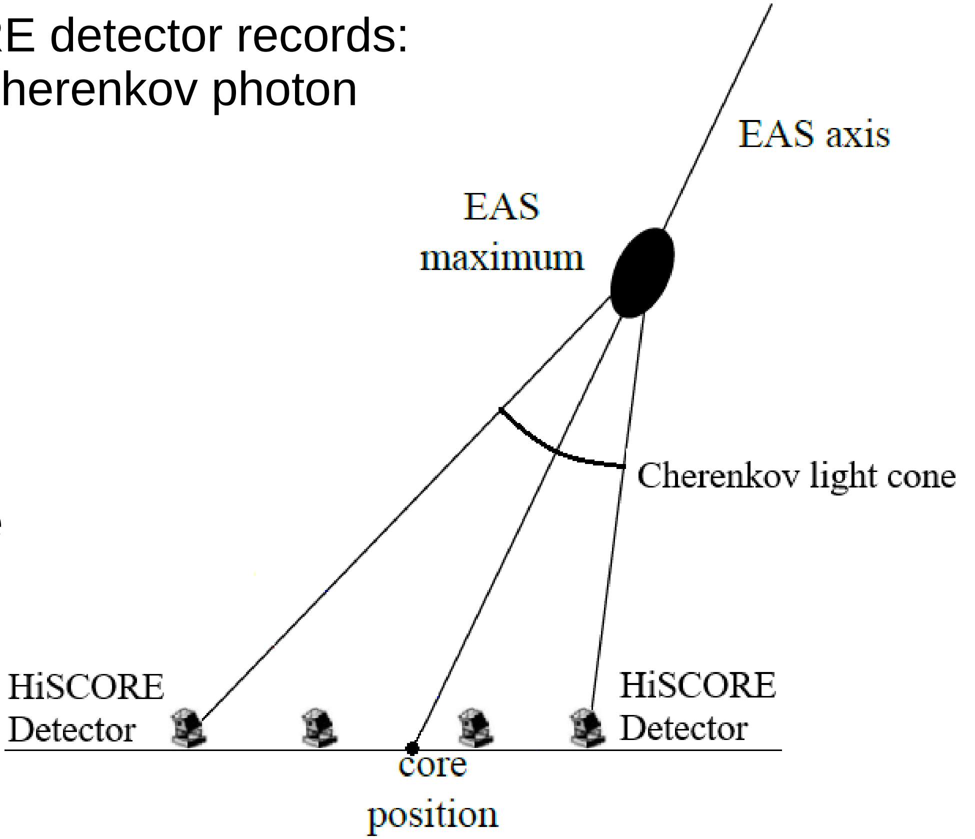
About TAIGA-HiSCORE

TAIGA-HiSCORE is a large-scale (1 km²) array of wide-angle non-imaging Cherenkov detectors, spaced ~100 m apart, designed to detect and study extensive air showers (EAS) from primary cosmic rays and gamma rays

Each TAIGA-HiSCORE detector records:
- Light amplitudes (Cherenkov photon density)
- Signal arrival times

The fundamental challenge

Reconstruct the EAS parameters (core position, arrival direction, and energy of the primary particle) from detector signals



The conventional approach to EAS parameter reconstruction

The current shower reconstruction method is based on approximating the Cherenkov light distribution as a function of distance from the estimated EAS axis

While effective, this method is simulation-dependent, having been empirically optimized and validated through Monte Carlo studies

Open questions:

- How optimal is the current method?
- Can we extract more information?
- What might we be missing?

The main question: Can we achieve better reconstruction accuracy?

Autoencoder for
EAS data analysis and control

Autoencoder (AE)

- Neural network that compresses input data into a latent space (essential features)
- Reconstructs data from this compact representation

Our AE Implementation

- Trained on TAIGA-HiSCORE events
- Latent space captures meaningful EAS features

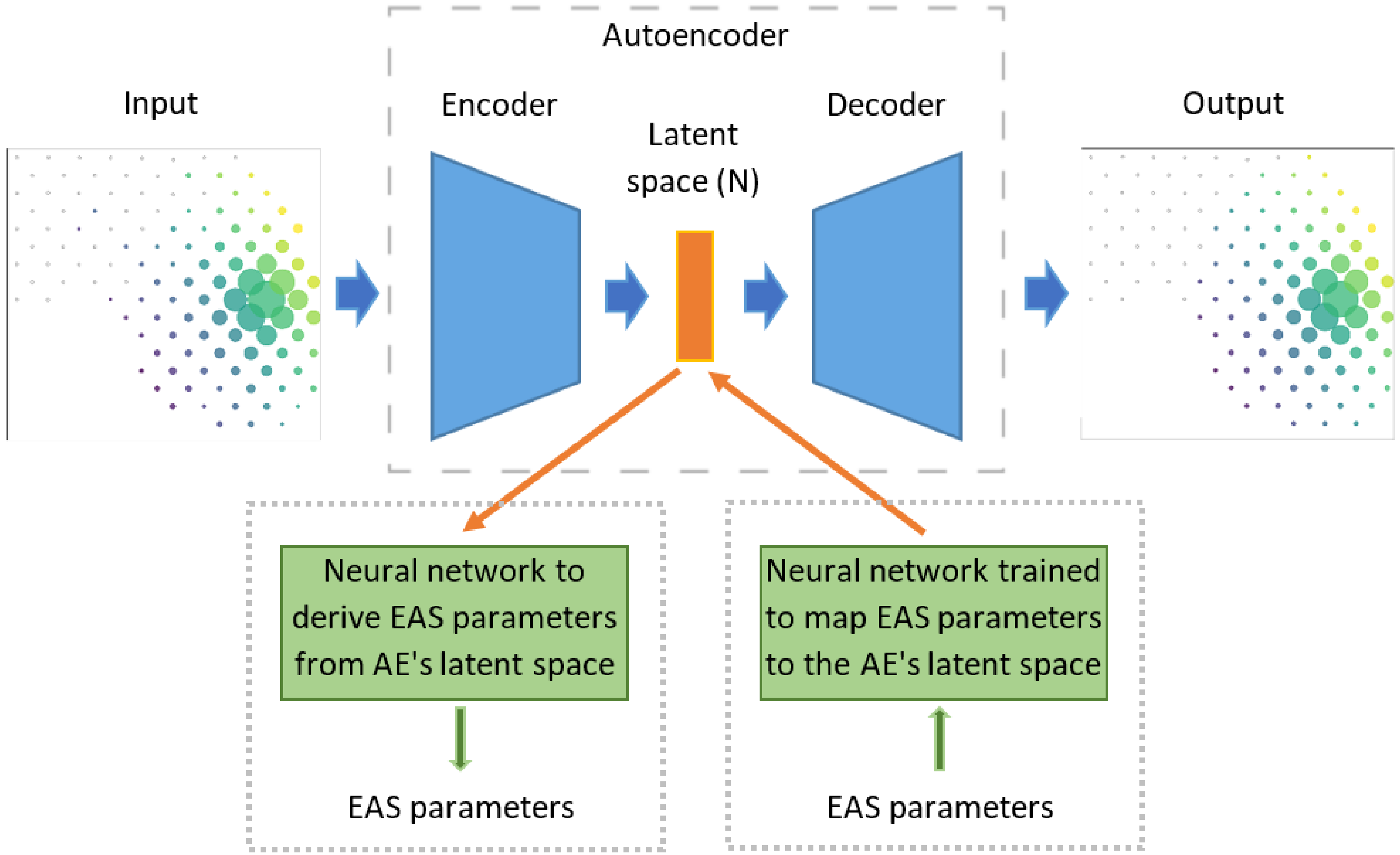
Key Applications

Interpretation: Reconstruct EAS parameters from latent space

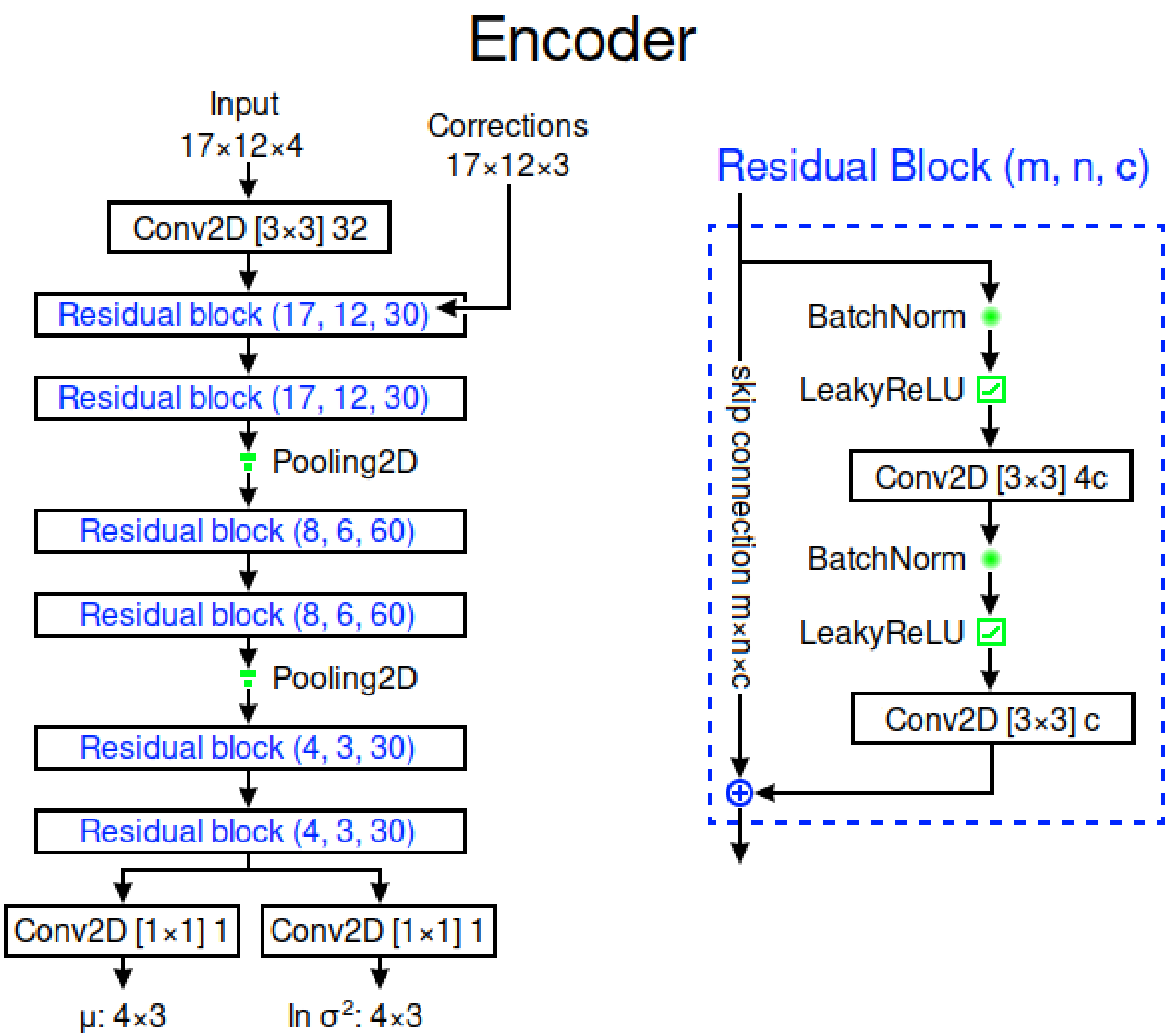
Control: Generate synthetic TAIGA-like data with tunable EAS parameters

General architecture of the system

AE acts as a data compressor: the dimensionality of the latent space (N) is significantly smaller than that of the input data. N is a parameter to be optimized



Architecture of the AE

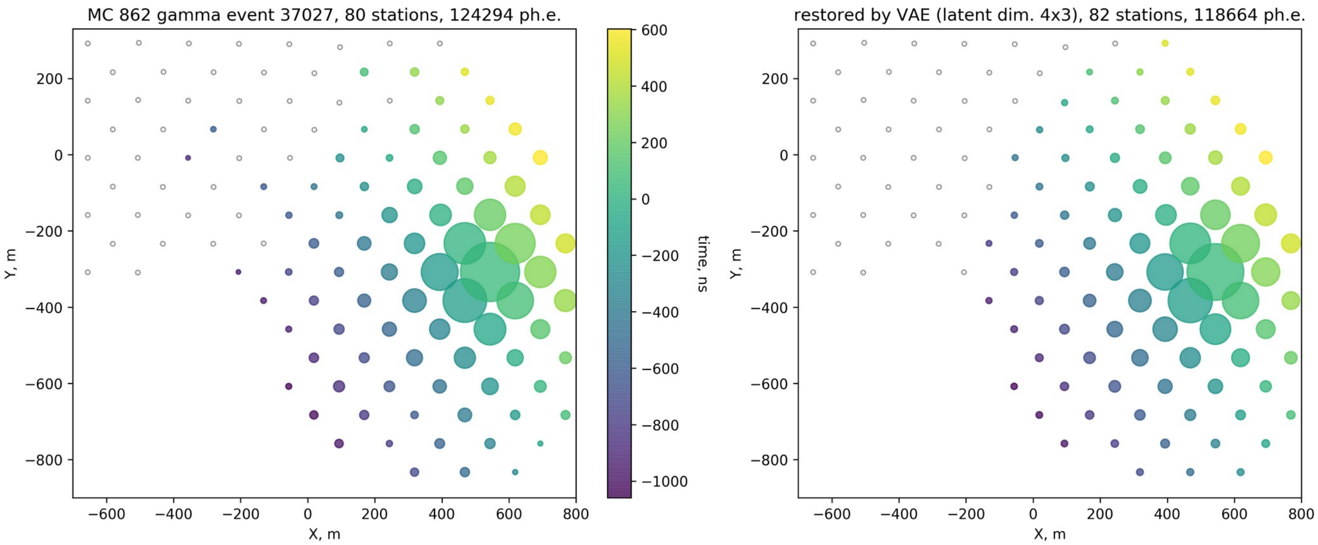


The decoder is mostly symmetrical with the encoder, with Conv2D layers replaced by Conv2DTranspose

We train our AE on augmented Monte Carlo simulation data for gamma quanta with 58,600 events

Results for the AE: Image reconstruction

A Monte-Carlo event restored by the AE with 12 latent parameters. Time is denoted by circle color and amplitude is denoted by size. Grey circles denote untriggered stations



Coefficients of determination (R^2) for restored quasi-images depending on the latent space dimension N of the AE (N = 4, 6, 8, 12, 16, 20)

	4	6	8	12	16	20
time	0.99922	0.99973	0.99973	0.99975	0.99976	0.99976
amplitude	0.83241	0.91355	0.96229	0.97174	0.97374	0.9735
s.d. of times	0.514	0.53314	0.54794	0.58936	0.60801	0.62623
trigger	0.71928	0.744	0.74836	0.75272	0.76456	0.78352

Interpretation of essential features: Neural network for EAS parameter reconstruction

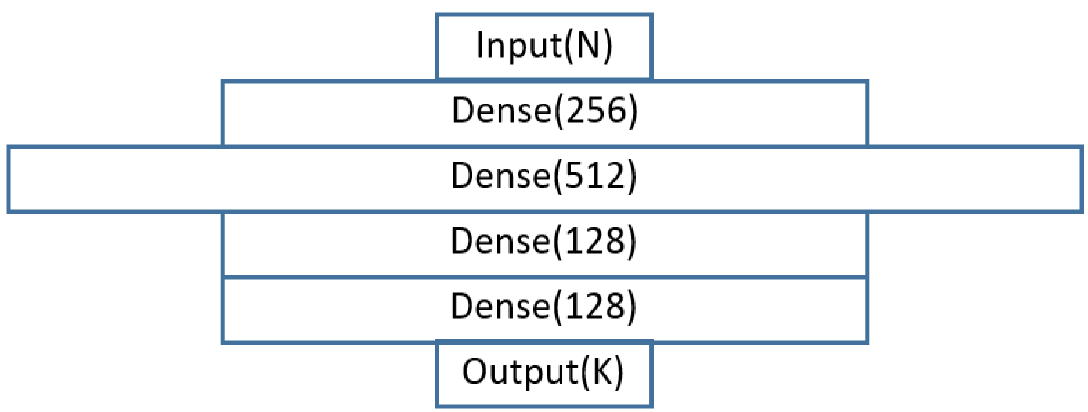
The neural network to derive EAS parameters from AE’s latents space is a multi-layer perceptron

Input: N=4, 6, 8, 12, or 16 values according to the dimensionality of the latent space

Output: a set of EAS parameters (K values)

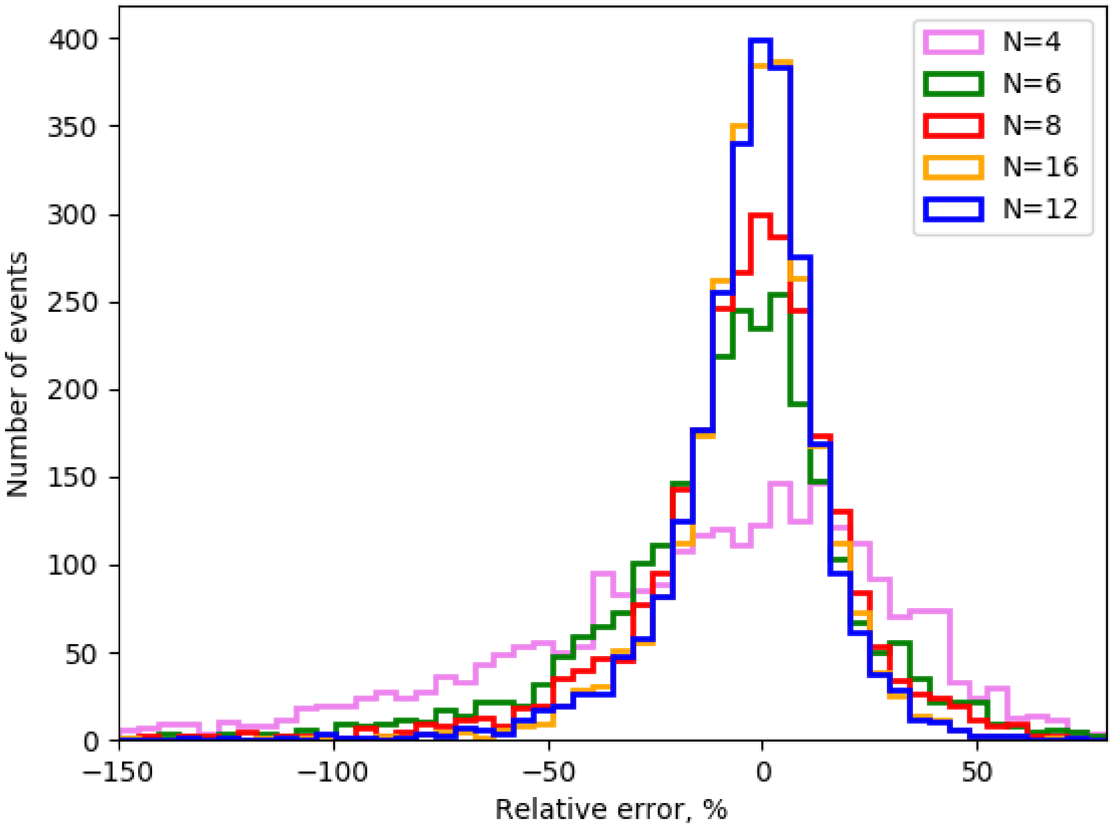
We validate the method specifically for energy reconstruction (K=1 test case)

Network architecture



We train this network on 35,500 sets of latent features along with the energy values of the corresponding Monte Carlo gamma quanta events

Selecting the dimension of the latent space



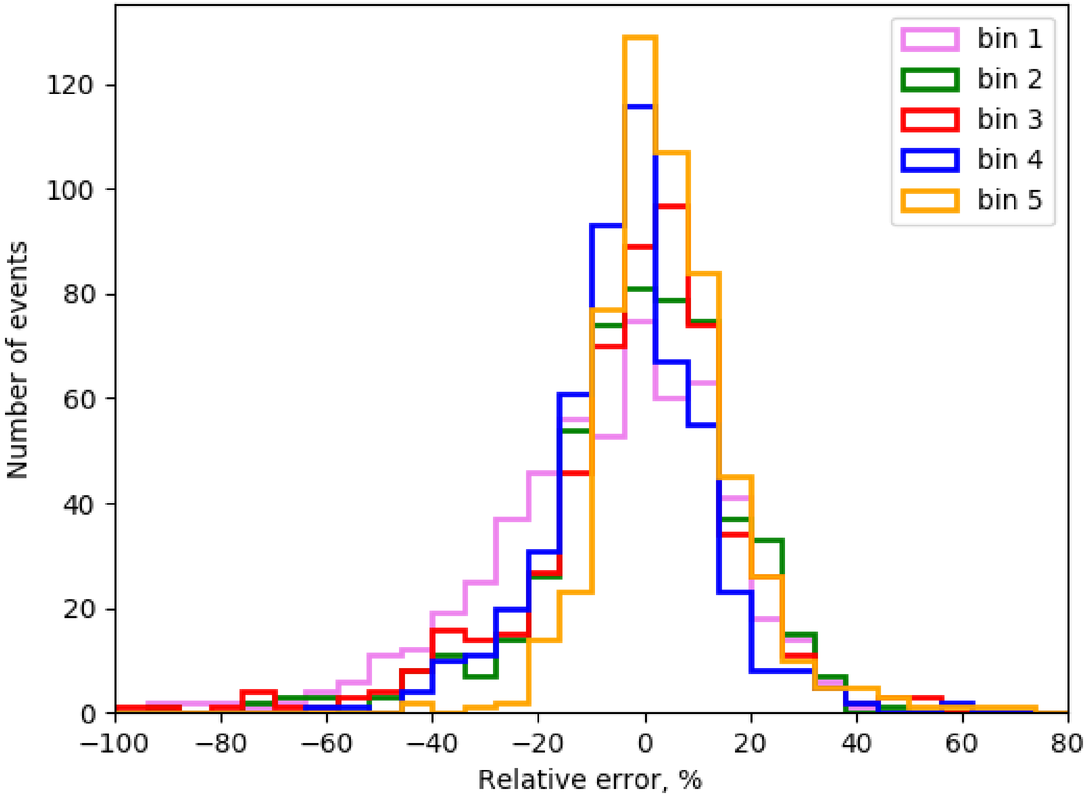
The distribution of the relative error of energy reconstruction for different values of N (dimension of the latent space):

$$Relative\ error = \frac{Energy_t - Energy_r}{Energy_t} * 100\%$$

Energy_t – real energy value
Energy_r – predicted (reconstructed) energy value

	N=4	N=6	N=8	N=12	N=16
mean(Relative Error), %	-20.4	-9.4	-4.7	-2.8	-2.4
std(Relative Error), %	52.5	34.4	25.1	18.9	19.1

Energy reconstruction for different energy bins



For N=12, we compare the distributions of relative errors of energy reconstruction for 5 different energy ranges (bins)

The energy bins are intervals used to group energy values, the bin widths are chosen so that each bin contains approximately the same number of events

	bin 1 100-160 TeV	bin 2 160-260 TeV	bin 3 260-420 TeV	bin 4 420-650 TeV	bin 5 650-1000 TeV
mean(Relative error), %	-8.1	-1.2	-2.3	-2.7	4.7
std(Relative error), %	24.2	19.2	20.4	14.7	13.1

Control of essential features:
Neural network to map EAS parameters to AE's latent space

Input: K= 8 EAS parameters*

*primary particle type and energy, shower direction (zenith and azimuth angles), first interaction height, shower maximum depth (X_{max}), and core position coordinates (X, Y)

Output: N = 12 latent features

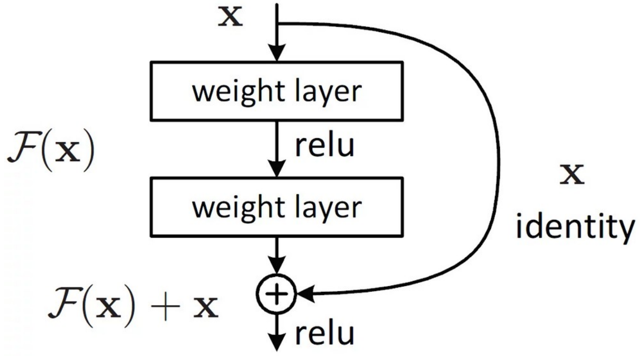
We train this network on 34,816 sets of EAS parameters along with corresponding latent features for Monte Carlo gamma quanta events

Network architecture

The neural network is a multi-layer perceptron with ResNet-like skip connections: one fully connected layer followed by 10 residue learning blocks of 2 fully connected layers

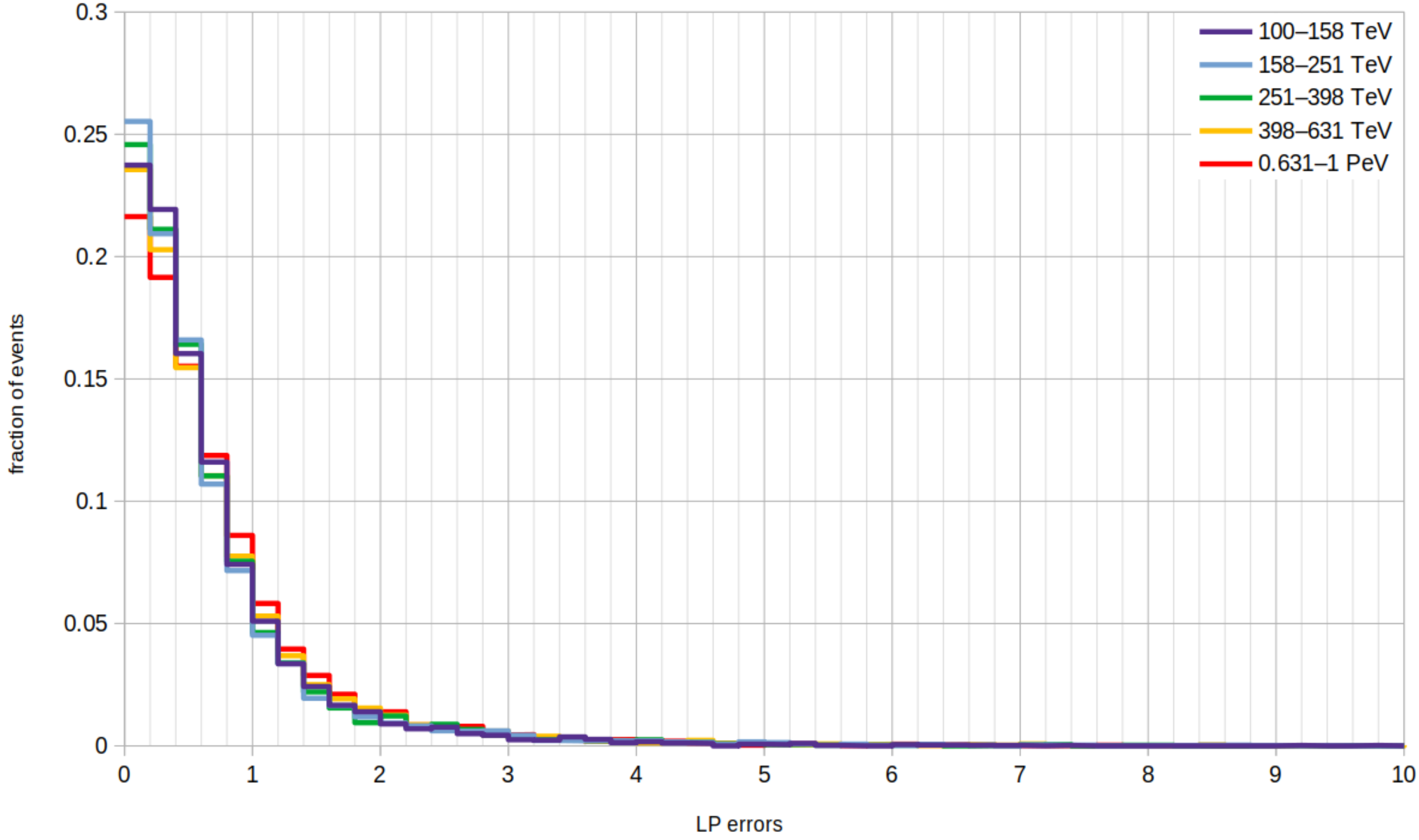
Total hidden layers: 21

Neurons per layer: 180



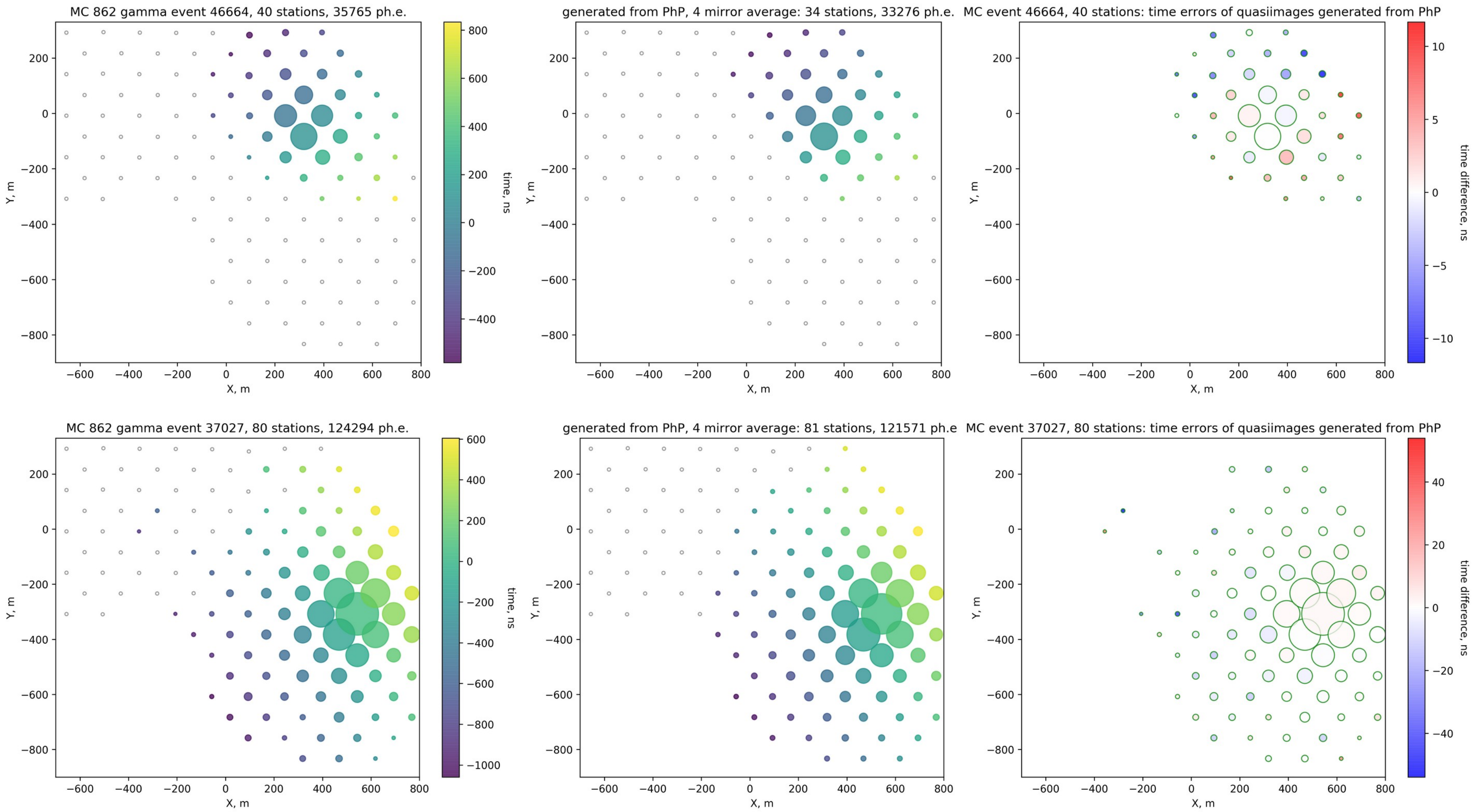
Validating Generated Latent Features

Histograms of absolute errors between generated and target latent feature values, binned by primary energy range (in total 5 bins)



Parameter-Controlled Shower Image Synthesis

Two Monte Carlo vs generated events; the right panels show timing errors



Conclusion

We demonstrate that an autoencoder can be trained to compress data from 121 TAIGA HiSCORE stations into 12 latent parameters with good reconstruction fidelity. The latent space of the autoencoder can be used to reconstruct the parameters of the primary particle as well as to generate synthetic TAIGA-like data with tunable parameters

Interpretation of essential features:

- Using the example of determining the energy of the primary particle for TAIGA-HiSCORE data, it is shown that the latent space dimension of 12 features is a reasonable choice
- The standard deviation of the relative energy reconstruction error is 20-25% for energies from 100 to 400 TeV and 13-15% for energies from 400 to 1000 TeV, which is comparable with the results obtained by the conventional methods

Control of essential features:

- Regarding the generation of essential features and corresponding images for specified physical parameters, the method has shown promising results with relatively low error rates
- The synthetic data constructed from EAS parameters approximate the respective Monte Carlo simulated data reasonably well but less accurately than their reconstruction by the autoencoder

We anticipate these results will benefit multimodal analysis (joint processing of data from different detector types), enabling seamless integration of latent parameters and improving overall reconstruction accuracy

Thank you for attention!

Input data and loss function for the AE

The positions of 121 HiSCORE stations are approximated by a 17x12 rectangular grid. The input data for the AE have 4 channels:

- amplitude A (number of photoelectrons)
- average time t of photoelectrons detected by the station
- standard deviation of photoelectron detection times
- trigger indicator

The AE reconstructs all these values using the loss function:

$$L = C_t L_t + C_A L_A + C_{\text{s.d.t.}} L_{\text{s.d.t.}} + C_{\text{trigger}} L_{\text{trigger}}$$

where L_{trigger} is binary cross-entropy, L_A and $L_{\text{s.d.t}}$ are masked MSE, and L_t is masked MSE multiplied by lg A

The coefficients are: $C_t = 10$, $C_A = 1$, $C_{\text{s.d.t.}} = 0.5$, $C_{\text{trigger}} = 0.0005$

Masked loss function components let AE keep nonzero values for the stations with the detected signal below the threshold (100 ph.e.) or missing stations, theoretically allowing AE to make physically plausible estimates for would-be detection times

Additionally, 17x12x3 station coordinate corrections are given as auxiliary inputs to both encoder and decoder

R^2 for generated images

	autoencoder	latent feature generator + decoder
time	0.99975	0.99938
amplitude	0.97174	0.95289
s.d. of times	0.58936	0.50363
trigger	0.75272	0.74624