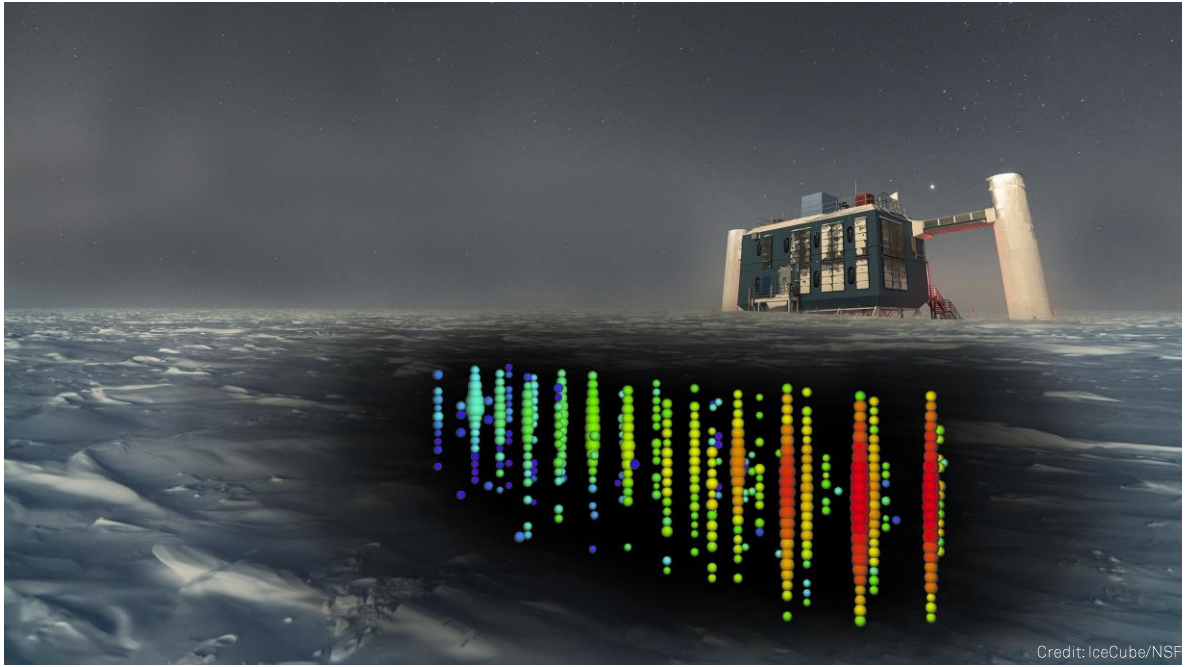
technische universität dortmund

ICECUBE SOUTH POLE NEUTRINO OBSERVATORY

SFB 876 Providing Information by Resource-Constrained Data Analysis

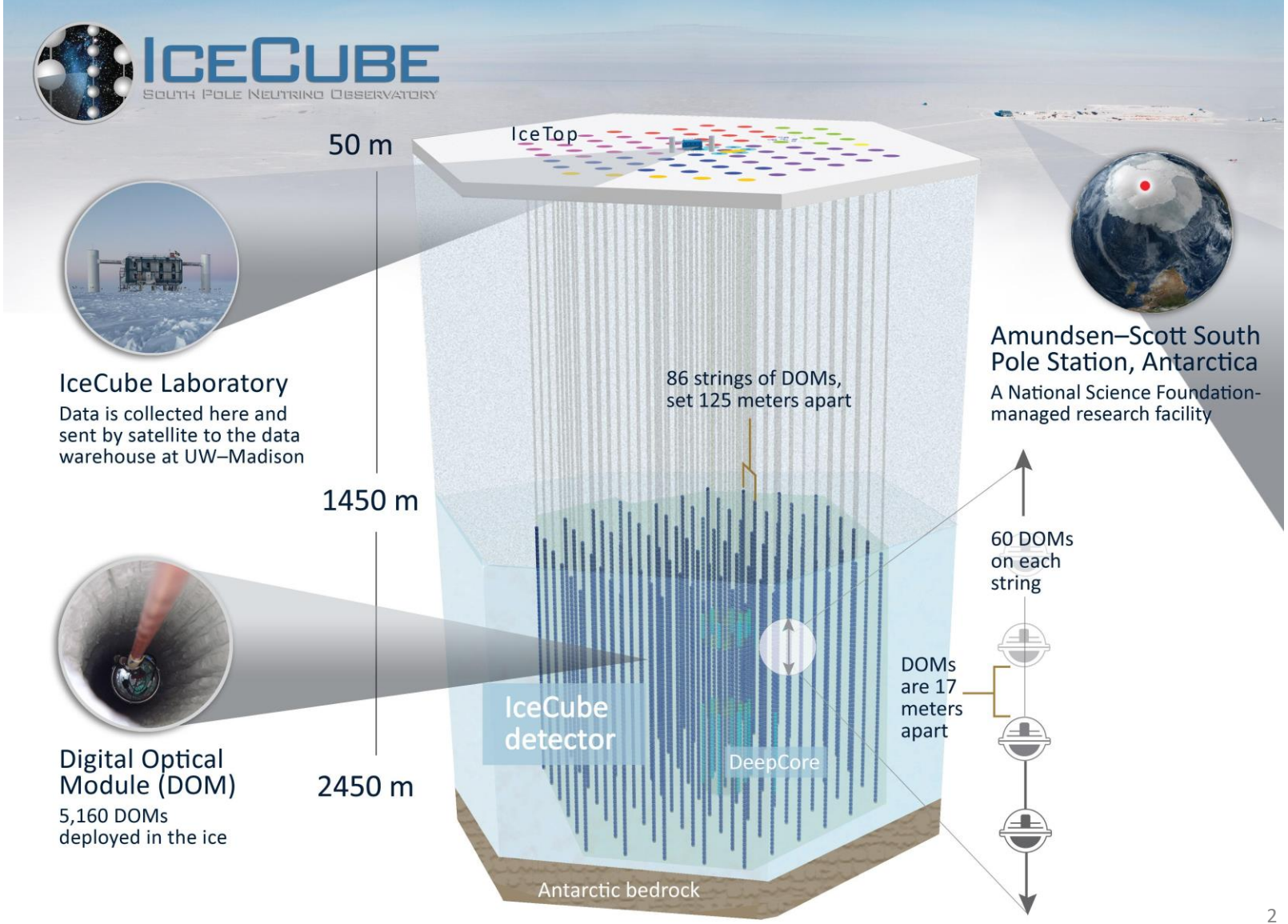
lehrstuhl physik e5



Cascade Reconstruction in IceCube using Generative Networks

Mirco Huennefeld
mirco.huennefeld@tu-dortmund.de

Deep Learning Workshop, Aachen
Feburary 19, 2019

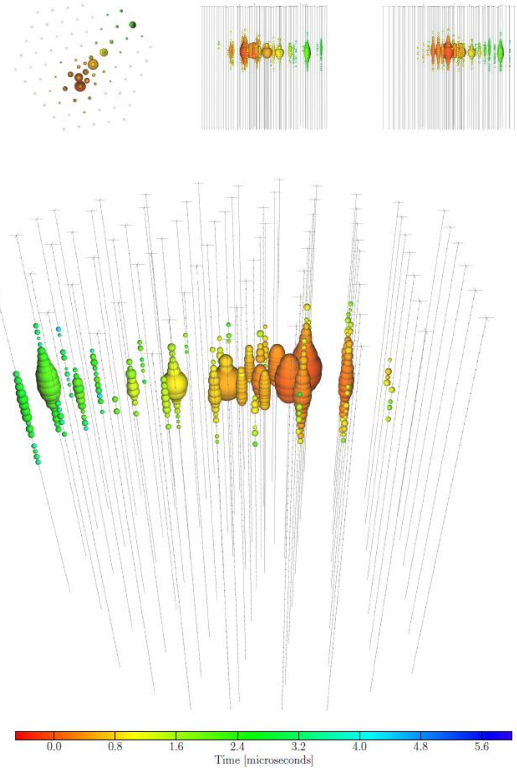

ICECUBE
SOUTH POLE NEUTRINO OBSERVATORY
50 m
IceTop
IceCube Laboratory
Data is collected here and sent by satellite to the data warehouse at UW–Madison
Amundsen–Scott South Pole Station, Antarctica
A National Science Foundation-managed research facility
86 strings of DOMs, set 125 meters apart
1450 m
60 DOMs on each string
DOMs are 17 meters apart
IceCube detector
DeepCore
Digital Optical Module (DOM)
5,160 DOMs deployed in the ice
2450 m
Antarctic bedrock

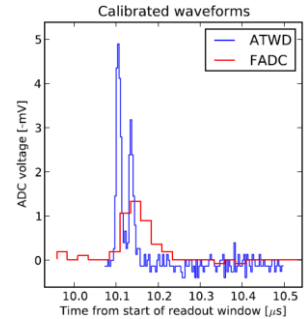
IceCube – Data Format

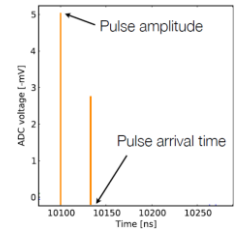
- 5160 DOMs
- 128 bins per Waveform
- Arbitrary number of waveforms per DOM

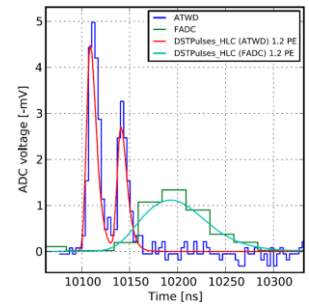
$$5160 \cdot 128 \cdot n = 660480 \cdot n$$

- MNIST: $28 \cdot 28 = 784$
- CIFAR: $32 \cdot 32 = 1024$
- ImageNet: $256 \cdot 256 = 65536$ (cropped)
- Readout window dependent on trigger
- 4-dimensional, highly variable data





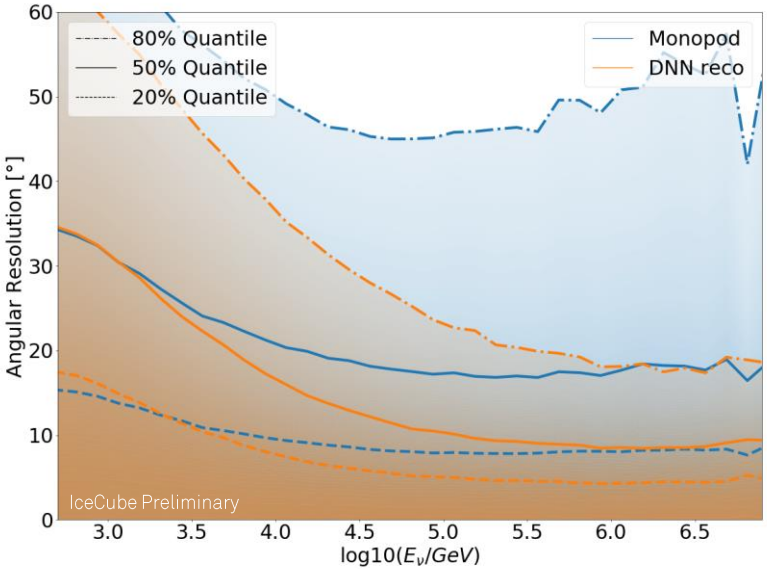


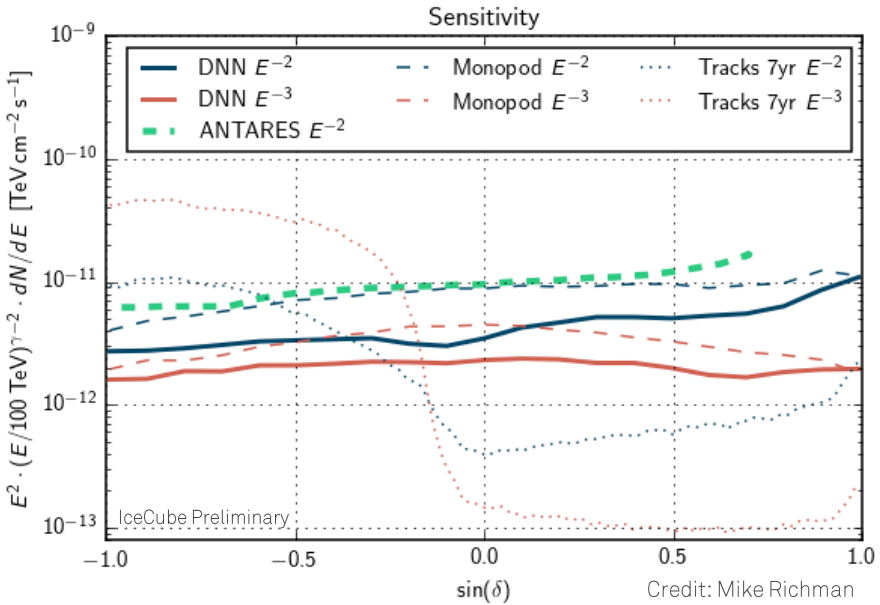


Goal: reconstruction of particle track and energy

Convolutional Neural Networks – DNN Reco

Cascade Angular Resolution



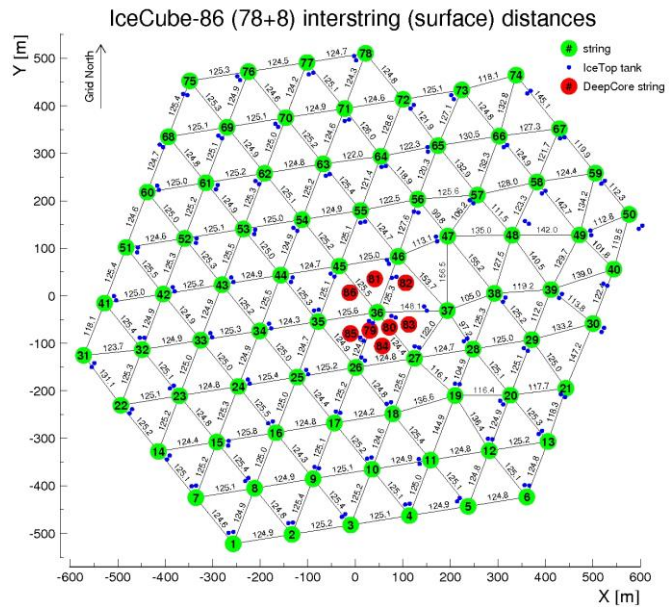


Credit: Mike Richman

- Application of CNN-based architectures can greatly improve resolution
- Example: cascade angular resolution and point source sensitivity increased by factor ~2
- Systematic checks: method is very robust, can marginalize over systematics

Convolutional Neural Networks – Limitations

- Only translational invariance and locality is used – More information and symmetries available
- Assumptions imposed by CNNs are only approximately met in IceCube
  - Irregularities in detector grid
  - No real translational invariance in observable space
- CNNs "wash out" data:
  - Great for robustness
  - Bad if exact information is needed

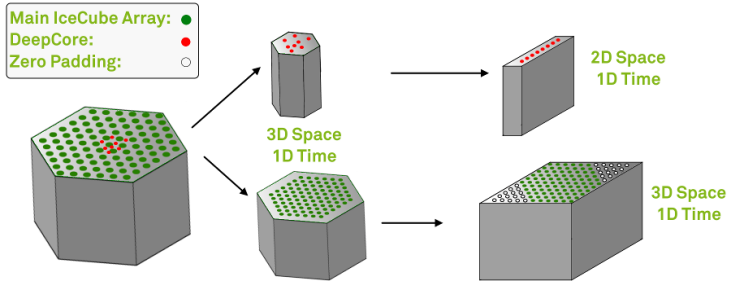


How can we improve?

Requirements for Next Generation

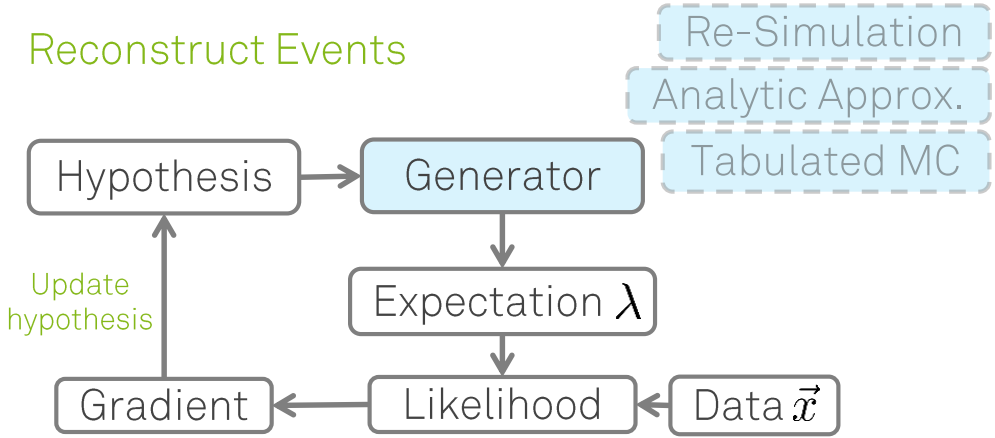
- Need for better suited network architectures:
  - Graph Neural Networks (GNNs) can help with irregularities in detector grid and uniformly handle IceCube + DeepCore
  - Combination of CNNs and recurrent neural networks can better handle time information
- Need to exploit more a priori knowledge:
  - Data generation process is known very well, but not exploited
  - Likelihood methods use this information

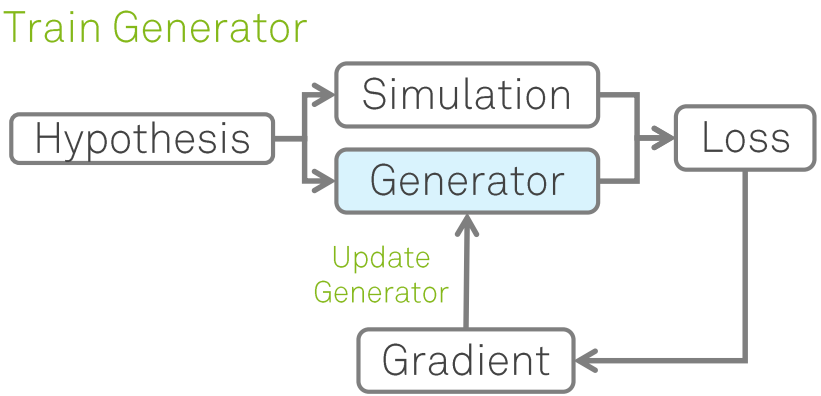
Can these approaches be combined?



Cascade Generator – General Idea

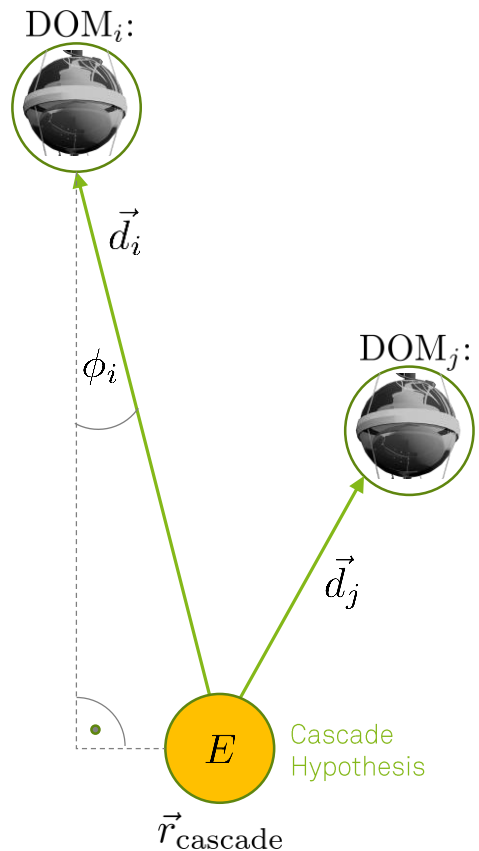
- Combine strengths of neural networks and maximum-likelihood methods
- Generative network to obtain fast approximation of simulation
- Possible in between of spline-based reconstruction and direct re-simulation
- Once generator is trained, it can be used in reverse mode for reconstruction





Cascade Generator – Network Architecture

- No adversarial or randomness needed as usually required for Generative Adversarial Networks
- Knowledge of exact detector geometry is included
- Translational invariance in physics variable space is exploited as opposed to observable space



Cascade Hypothesis:

$(x, y, z, \varphi, \theta, E, t)$ 7 free parameters

Input per DOM:

$\vec{d}, \phi, \ldots$ } Include geometry

$\varphi_{\mathrm{cascade}}$
$\theta_{\mathrm{cascade}}$
$\vec{r}_{\mathrm{cascade}}$

Weight sharing over DOMs:
Exploit invariances in latent variable space

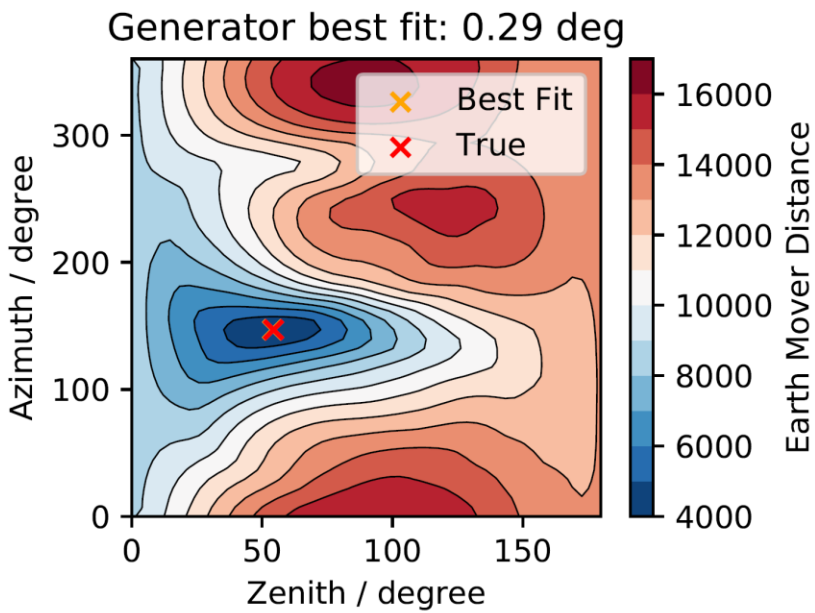
$E, \ldots$

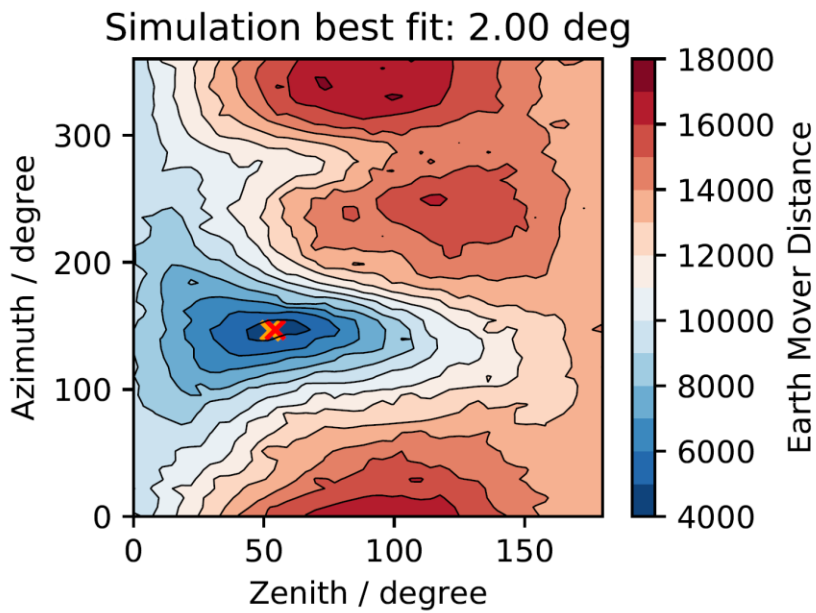
Additional independent weights per DOM:
Account for asymmetries and inhomogeneities

Cascade Generator vs. Direct Re-Simulation

- Neural Network is extremely good at interpolating and smoothing
- Landscape very well behaved, smooth gradients available
- Much faster per-event runtime

Potential to surpass direct re-simulation

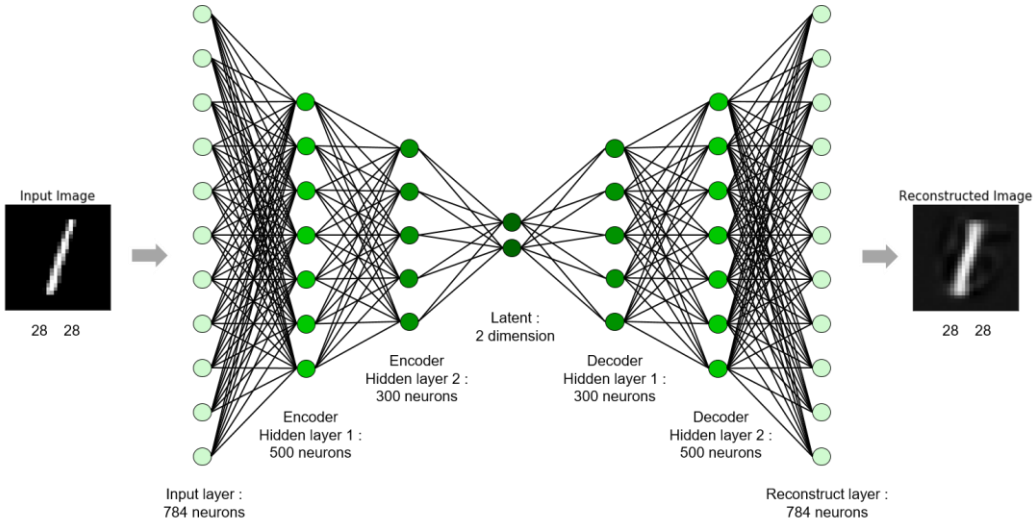


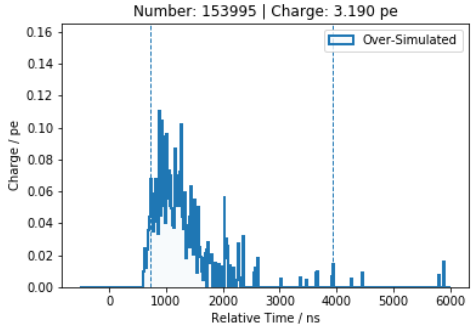


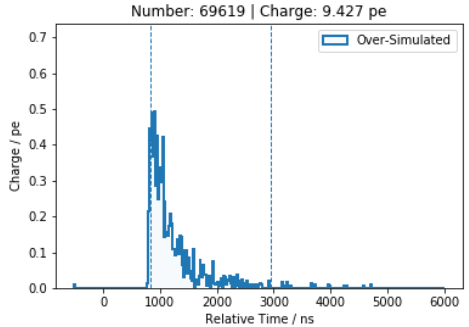
Cascade Generator – Time Parameterization

How can the time dimension be included?

- Bin charge in time bins: too many parameters!
- Calculate quantiles of cumulated charge
- Find a parameterization for the expected arrival times
- Train an autoencoder to reduce the dimensionality



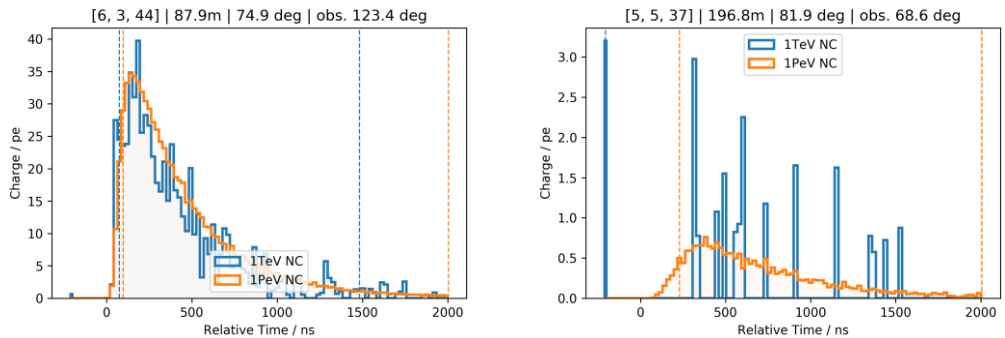




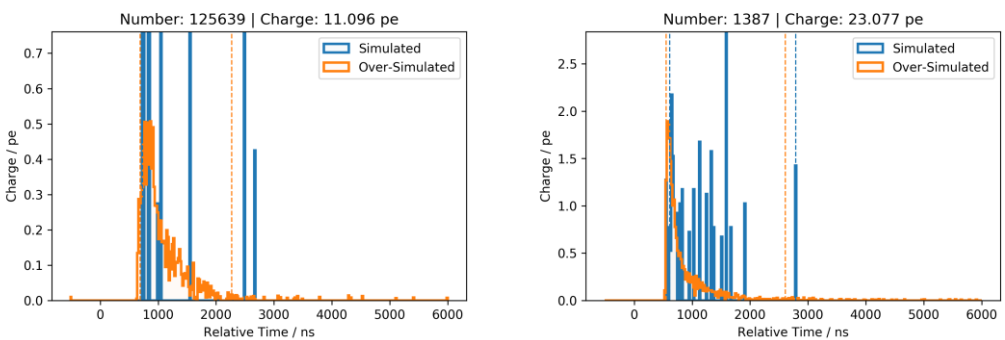
Cascade Generator – Autoencoder

How to define pulse arrival time pdf when only a few hits are present?

- Shape of arrival time pdf not well defined if only a few hits are there
- Need to over-simulate events: simulate same event n times and average hits



Increase cascade energy:
Light yield is proportional to energy

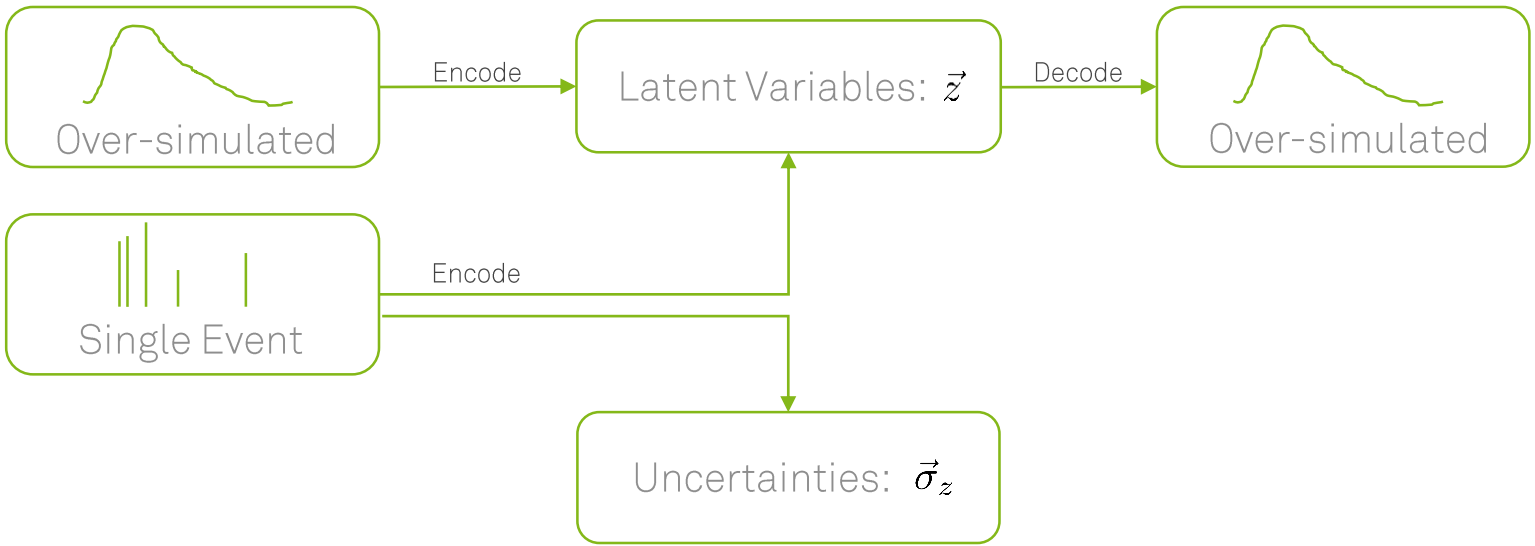


Over-simulate events:
Hits averaged over 100 simulations

Cascade Generator – Autoencoder

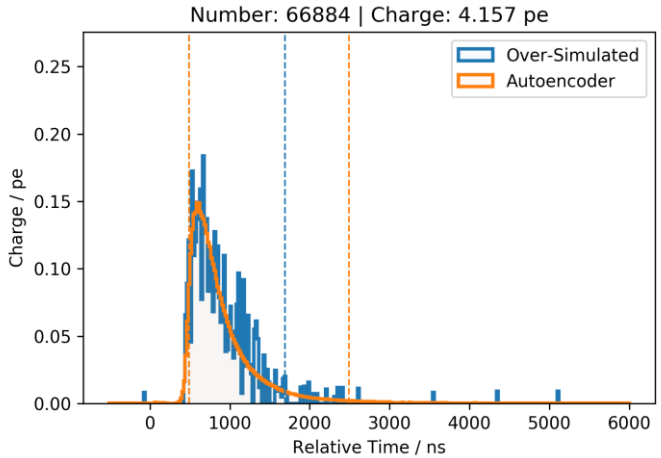
How to define pulse arrival time pdf when only a few hits are present?

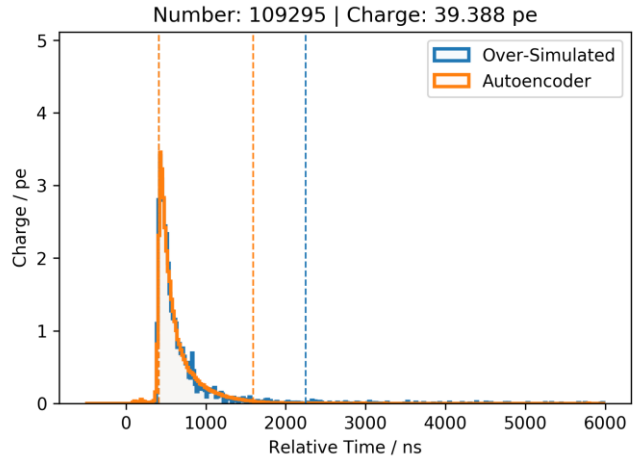
1. Train autoencoder on over-simulated events
2. Train second encoder on single events
   - Also learn to estimate uncertainty on latent variables



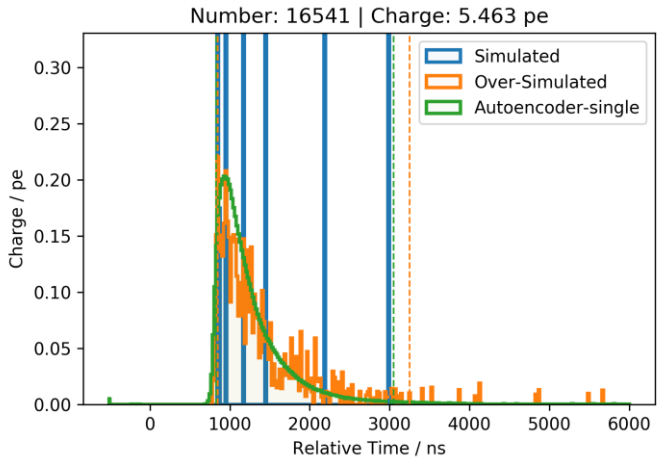
Cascade Generator – Autoencoder

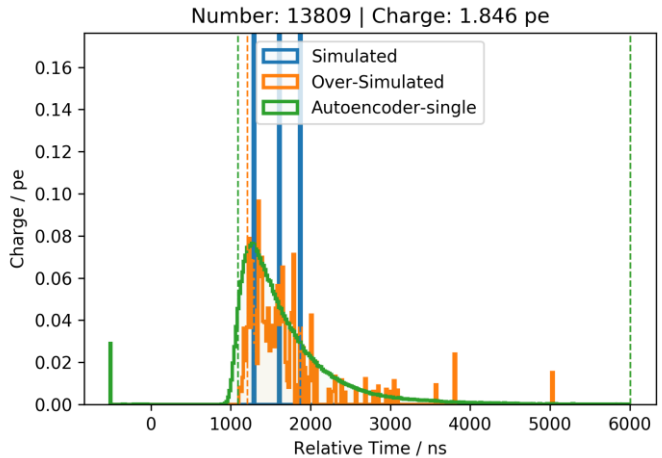
N = 100





N = 1



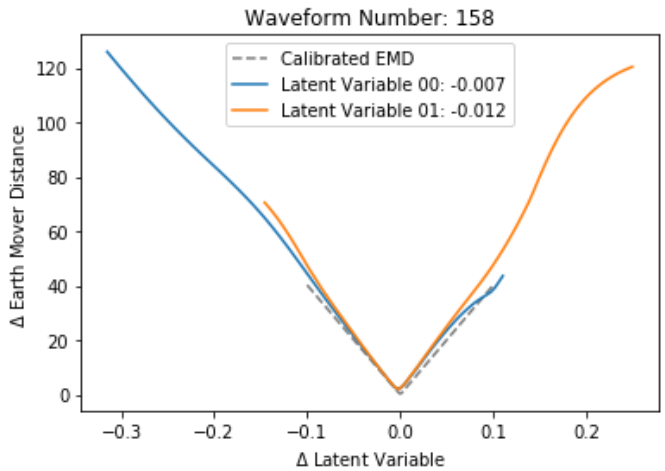


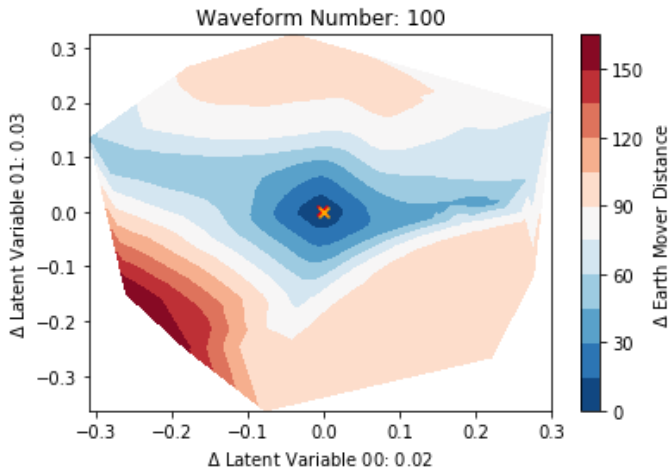
Cascade Generator – Autoencoder

Additional Requirements can be imposed on latent variables:

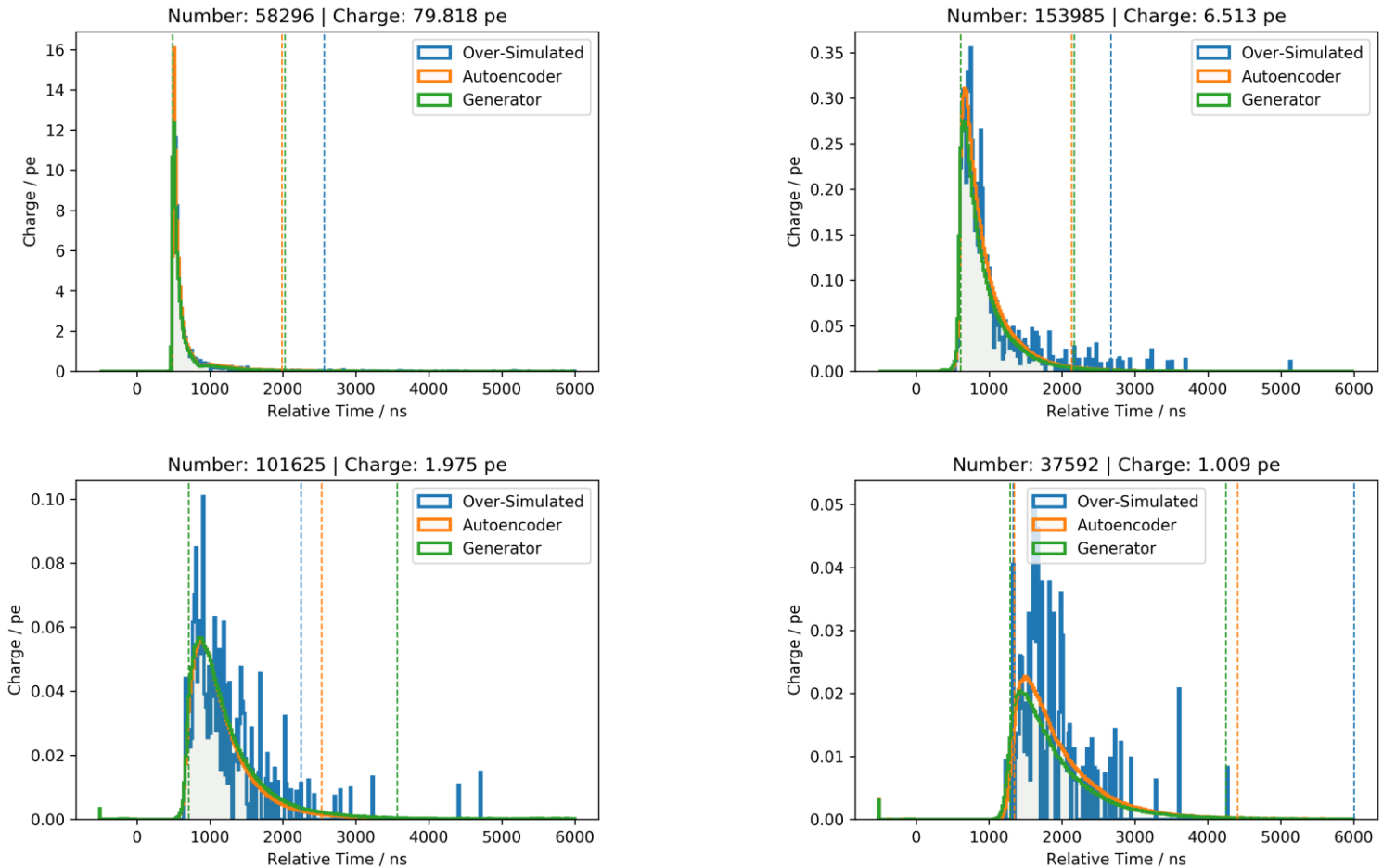
- Penalize correlation between latent variables
- Calibrate latent variables: distance in latent variable space corresponds to distance of two waveforms as defined by some distance measure (e.g. MSE, EMD, Chi square, etc.)

→ Reduces and simplifies dimensionality while retaining all relevant information!



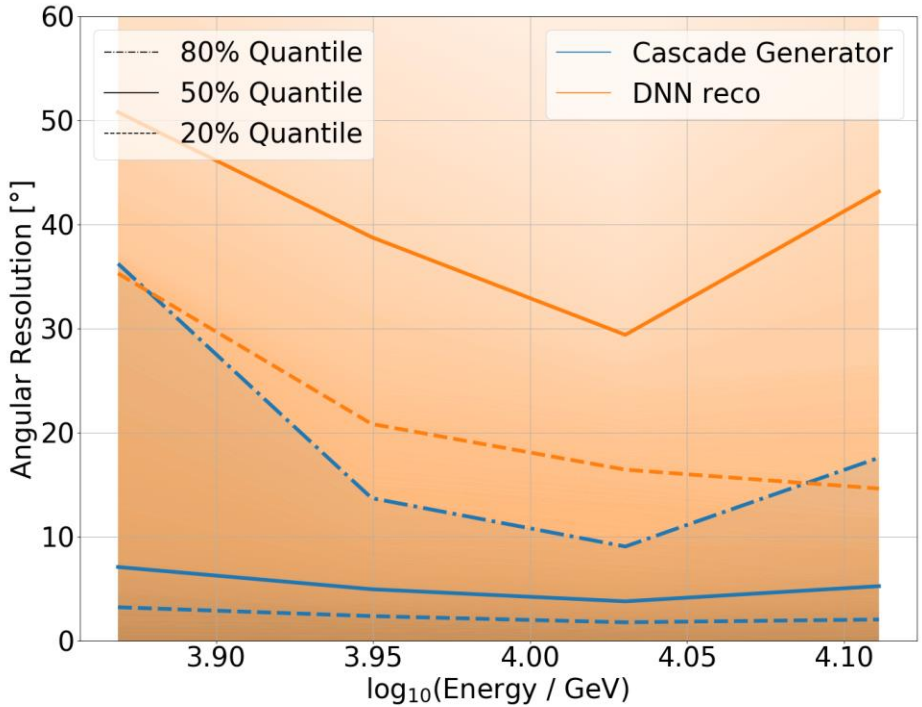


Cascade Generator – Generated Pulse Arrival Time PDFs



Cascade Generator – Performance

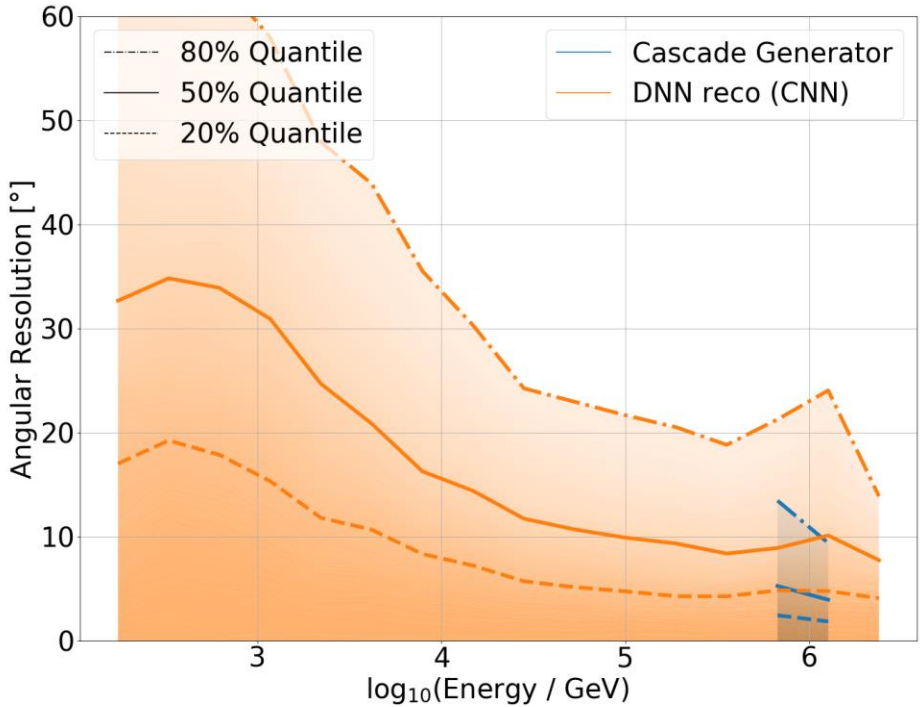
Over-simulated events (n=100):



- Works very well for over-simulated events
- Current limitations:
  - ⊕ DeepCore is not included
  - ⊕ Time resolution for AE: 20ns
  - ○ Only trained for mono-energetic events
  - ⊖ Only for over-simulated events
- Caveats in plot:
  - DNN reco is a very basic CNN that is used as seed for Cascade Generator
  - DNN reco reconstructs single event (n=1) whereas Cascade generator uses (n=100)

Cascade Generator – Performance

Comparison to state-of-the-art



- Works very well for over-simulated events
- Current limitations:
  - ⊕ DeepCore is not included
  - ⊕ Time resolution for AE: 20ns
  - ○ Only trained for mono-energetic events
  - ⊖ Only for over-simulated events
- Caveats in plot:
  - Cascade generator which uses over-simulated (n=100) events at 1e4 GeV is scaled to equivalent light yield (shifted in energy by a factor of 100)

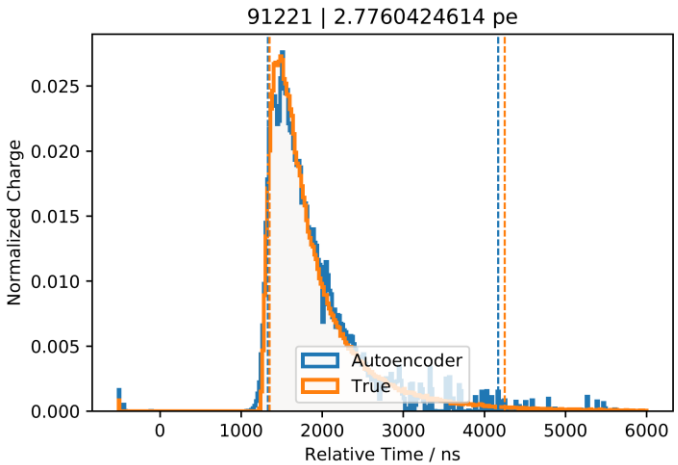
Summary & Outlook

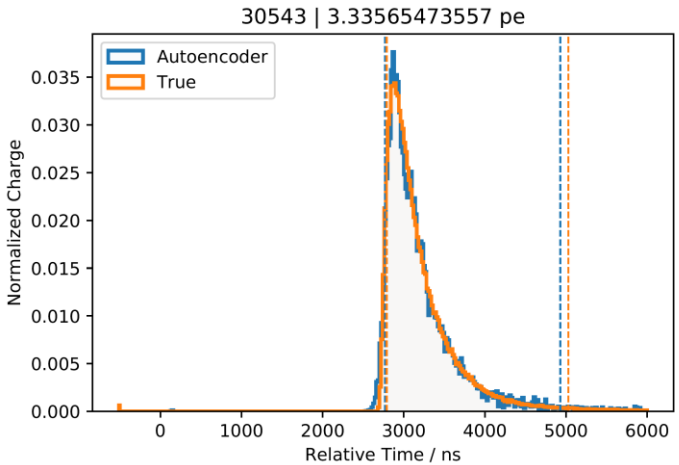
- Conventional NN architectures provide good results, but have limitations
- Necessity to develop dedicated techniques for the physics use-case
- Cascade Generator as a combination of maximum-likelihood methods and neural networks:
  - Able to combine strengths of both methods
  - Inclusion of symmetries and a priori knowledge
- Outlook:
  - Fix outstanding issues and limitations
  - Extension to reconstruction of other topologies such as muon tracks
  - Same approach can be used to fit Monte Carlo parameters (see afternoon talk)

Backup

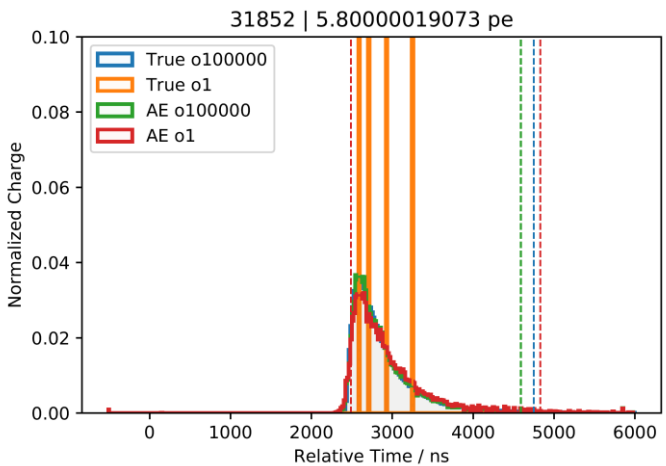
Cascade Generator – Autoencoder

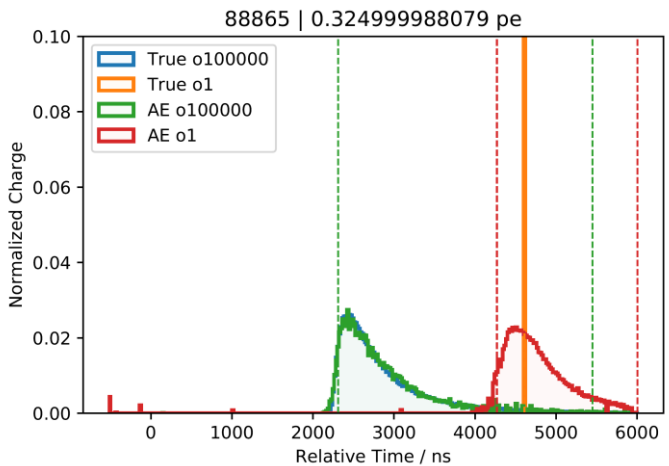
N = 100000



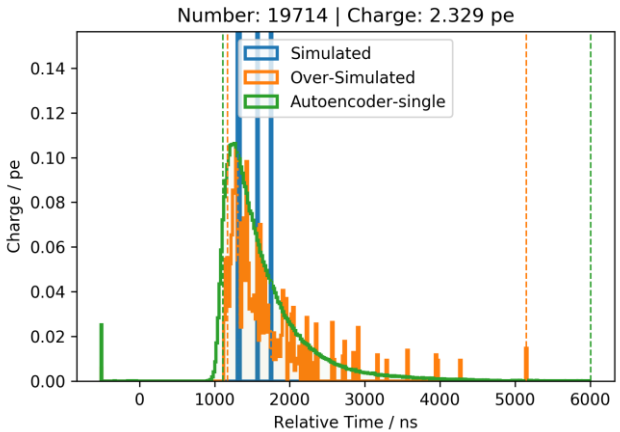


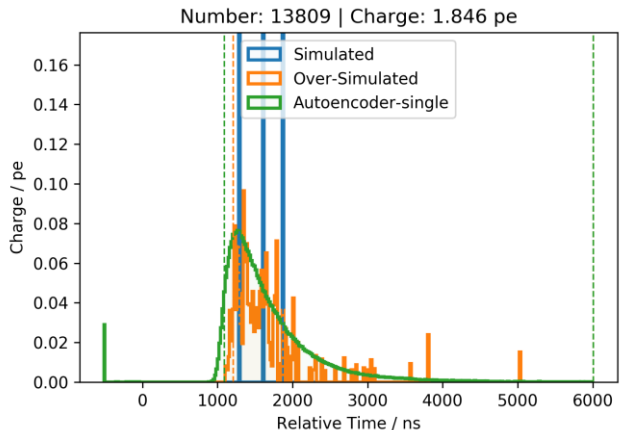
N = 1

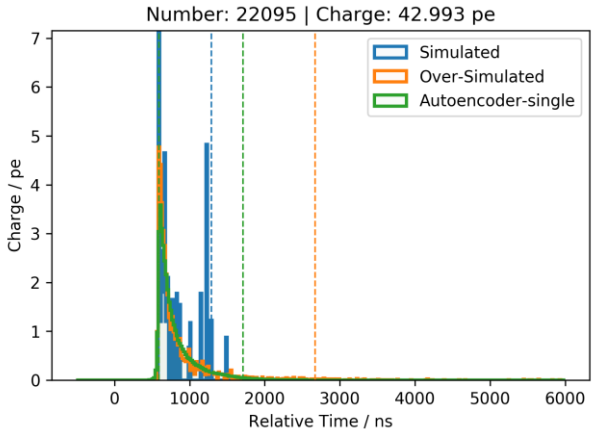


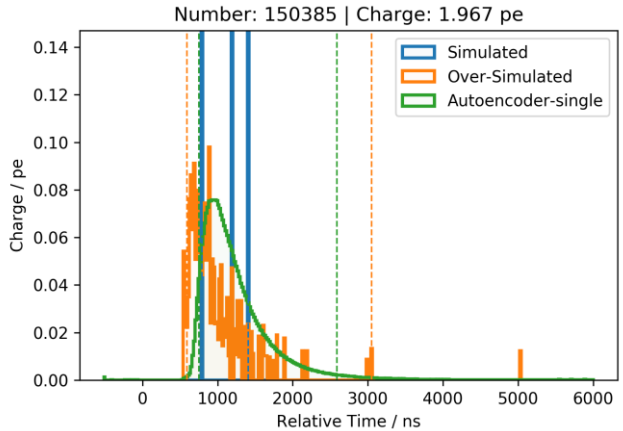


Cascade Generator – Autoencoder



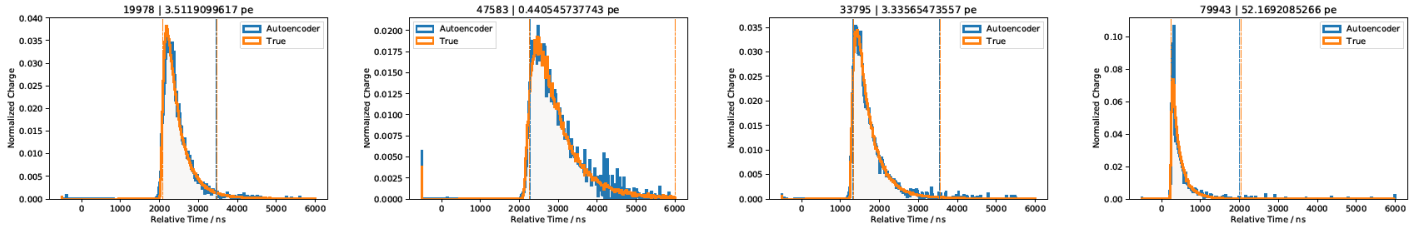




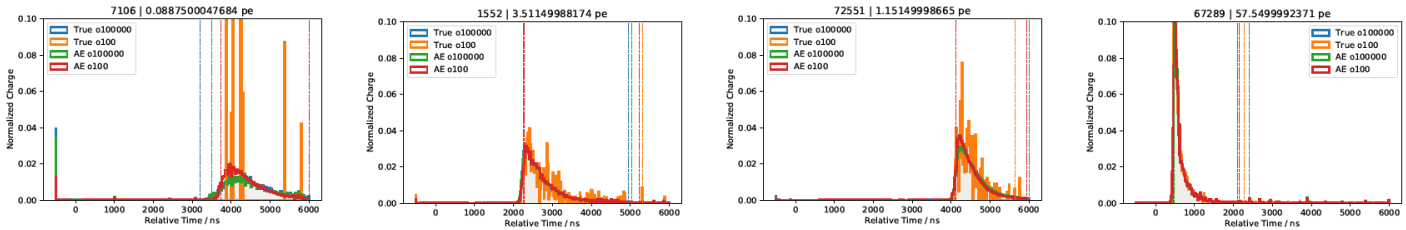


Cascade Generator – Autoencoder

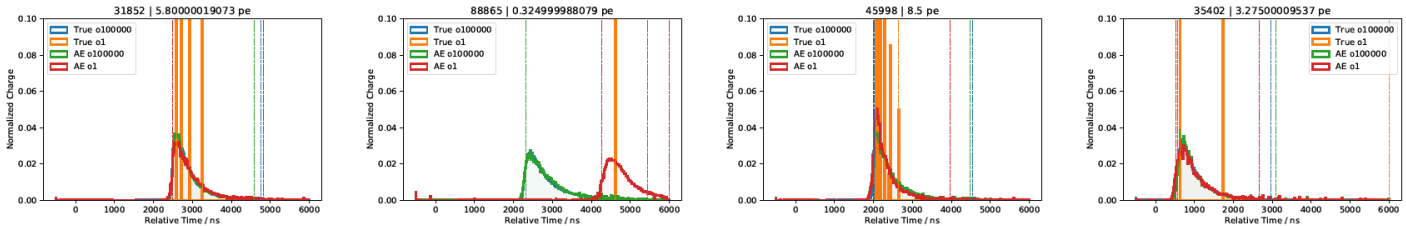
Oversampling Factor: 100 000



Oversampling Factor: 100

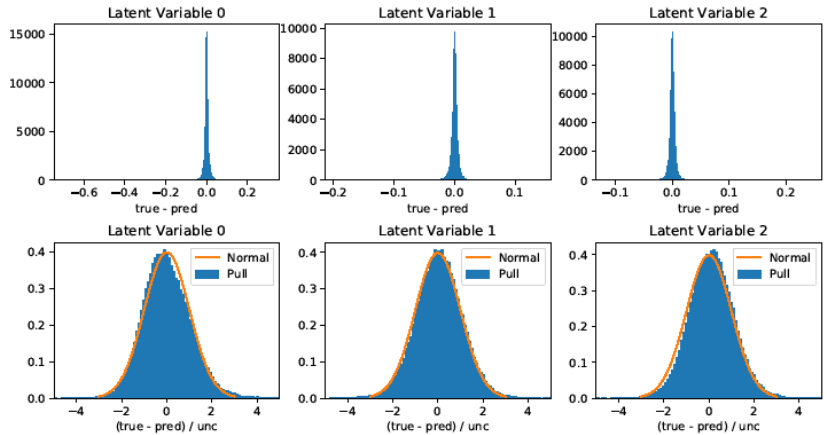


Oversampling Factor: 1

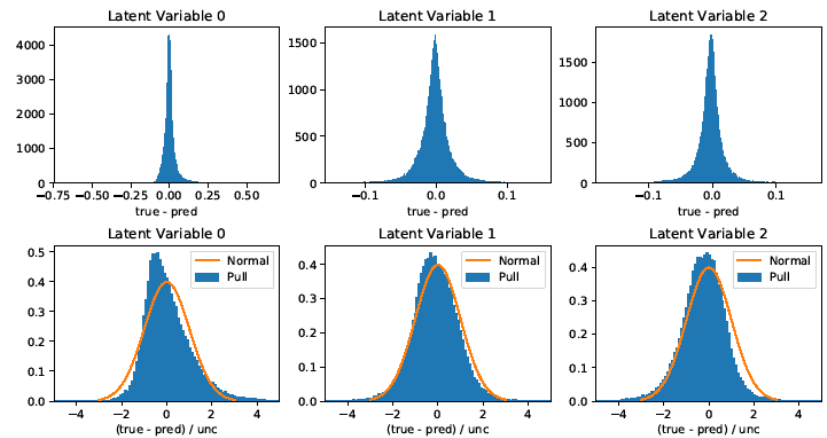


Cascade Generator – Autoencoder

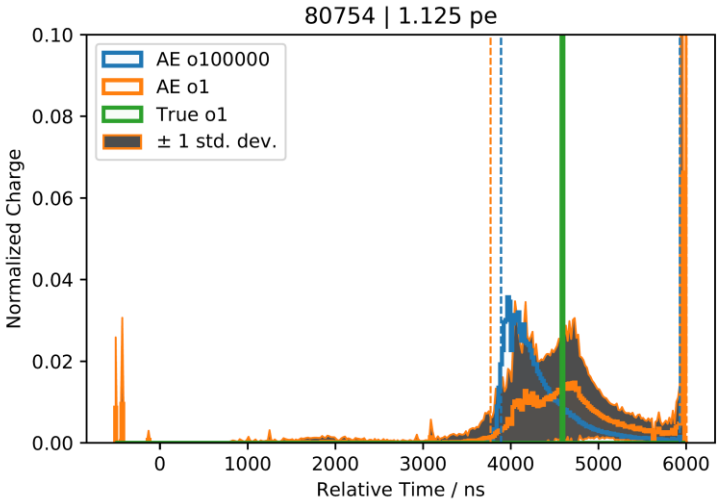
Oversampled events:

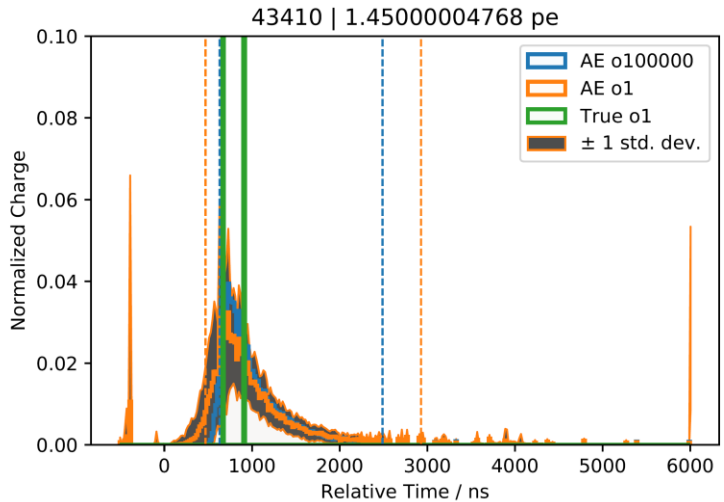


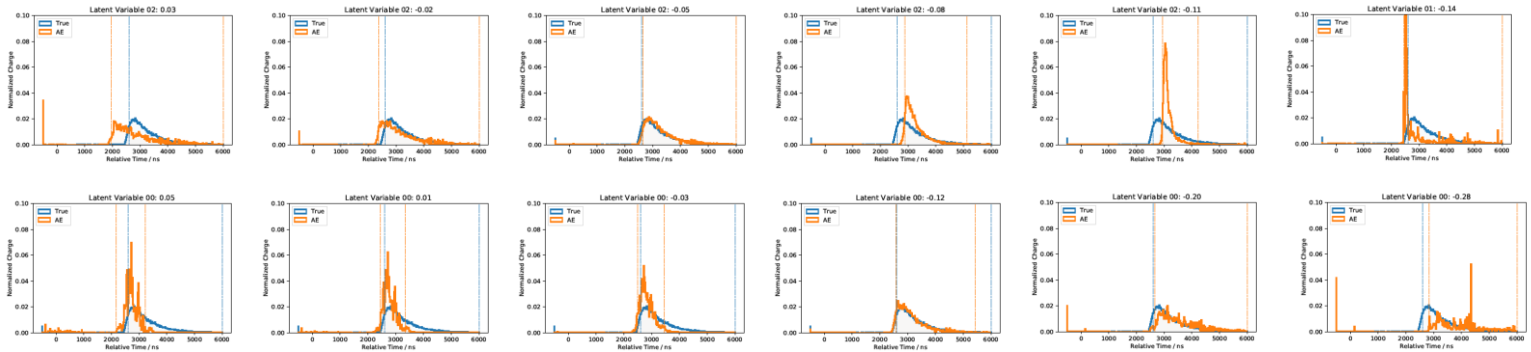
Single events:



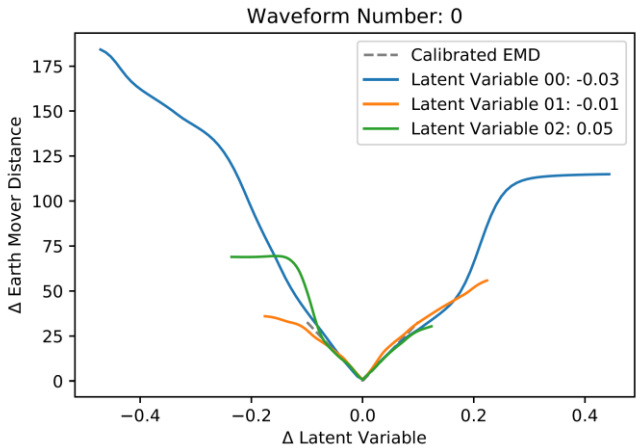
Cascade Generator – Autoencoder

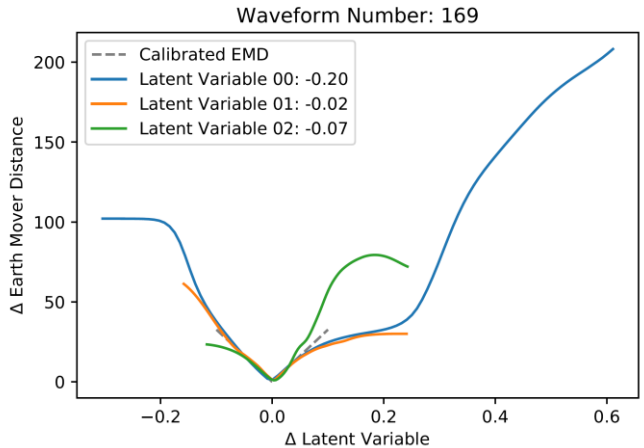


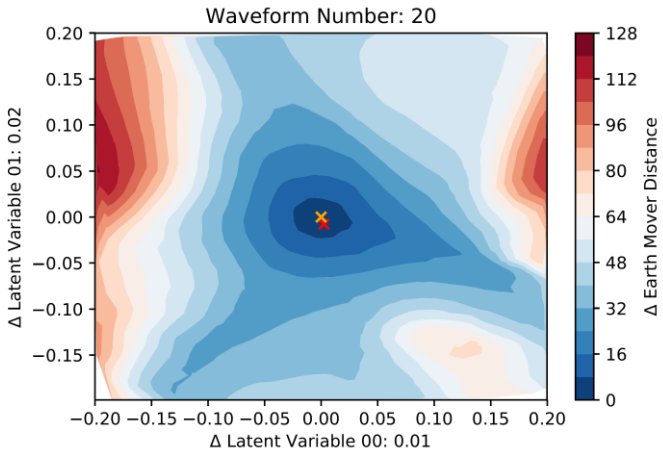


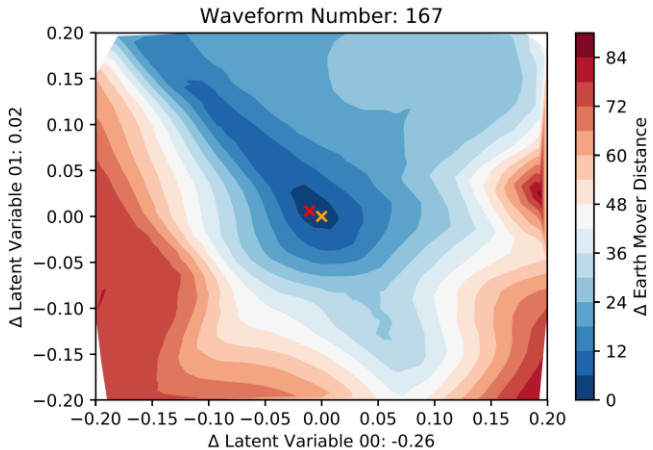


Cascade Generator – Autoencoder

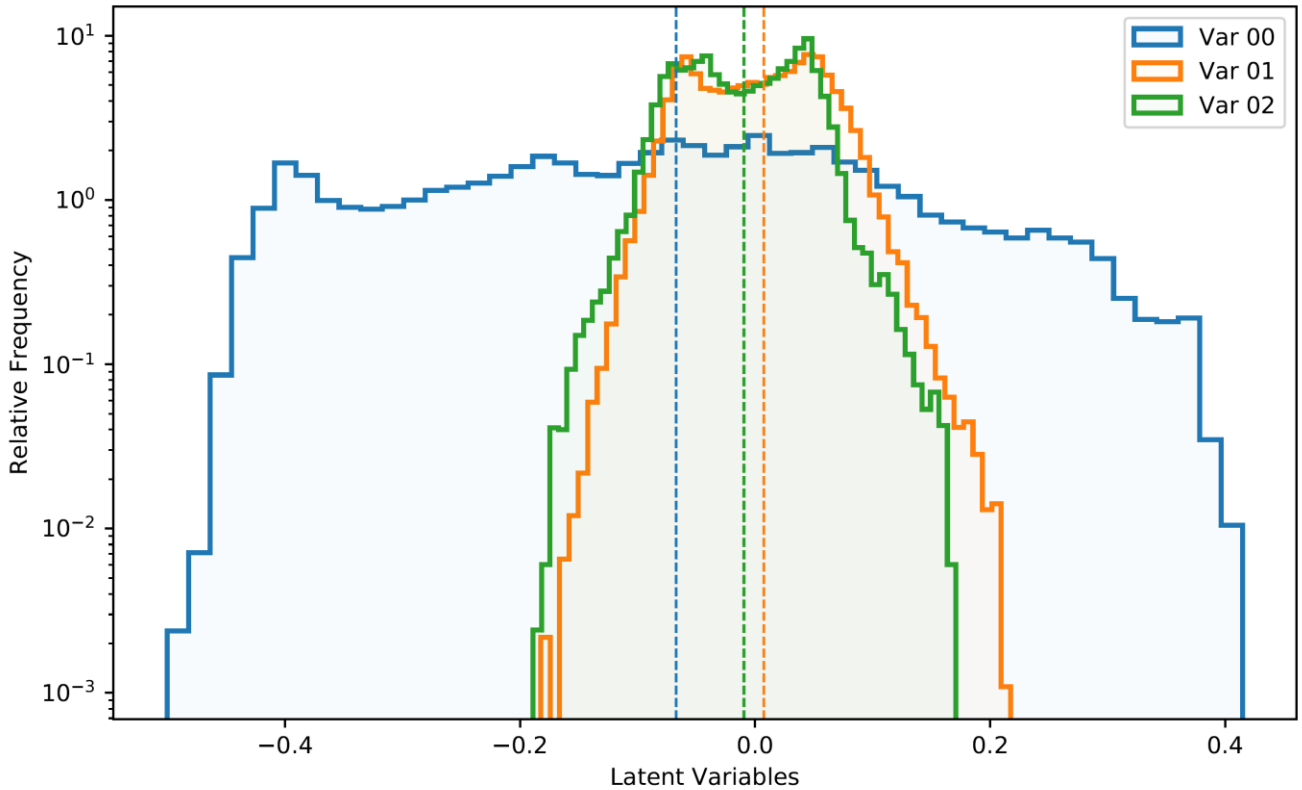




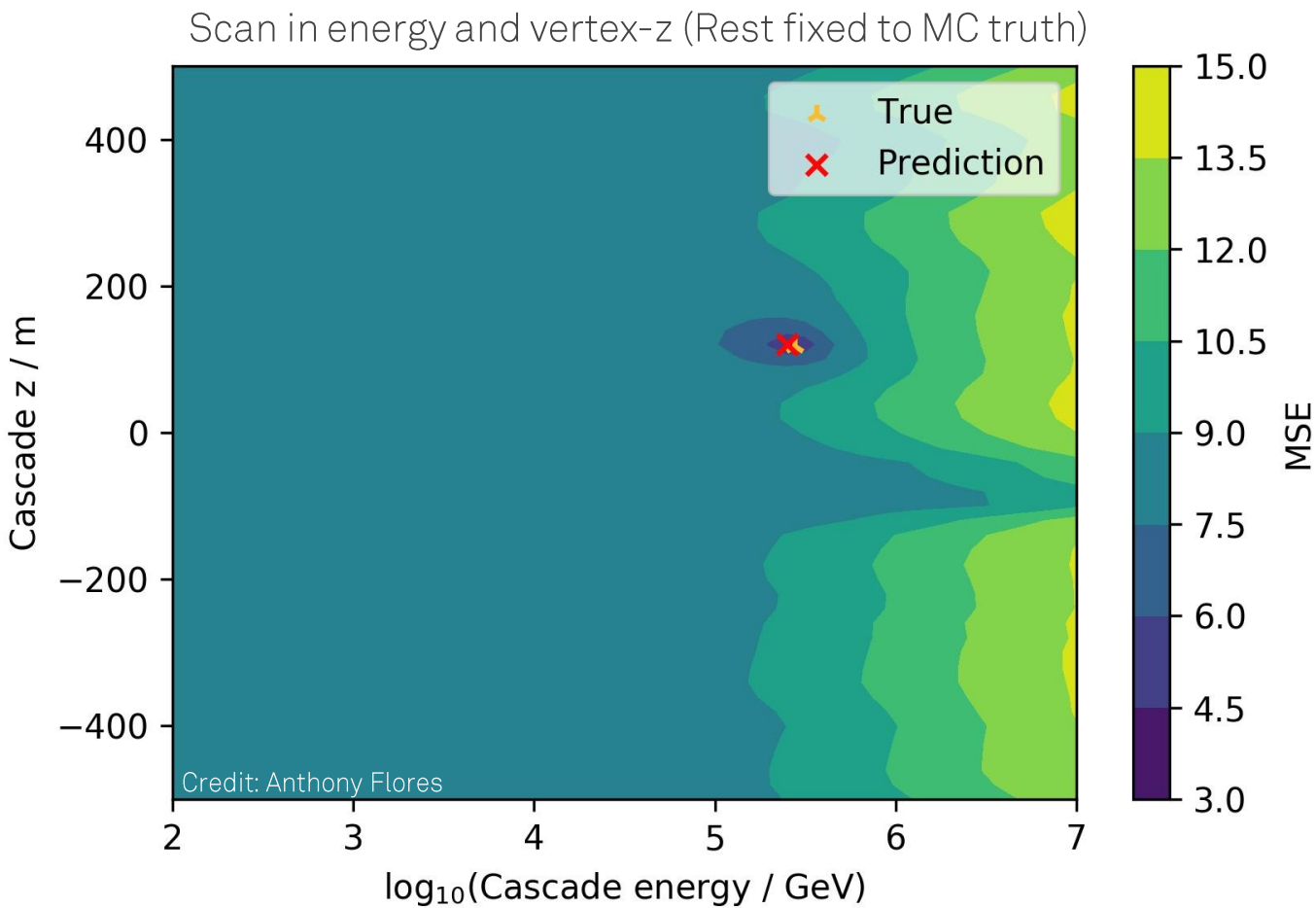




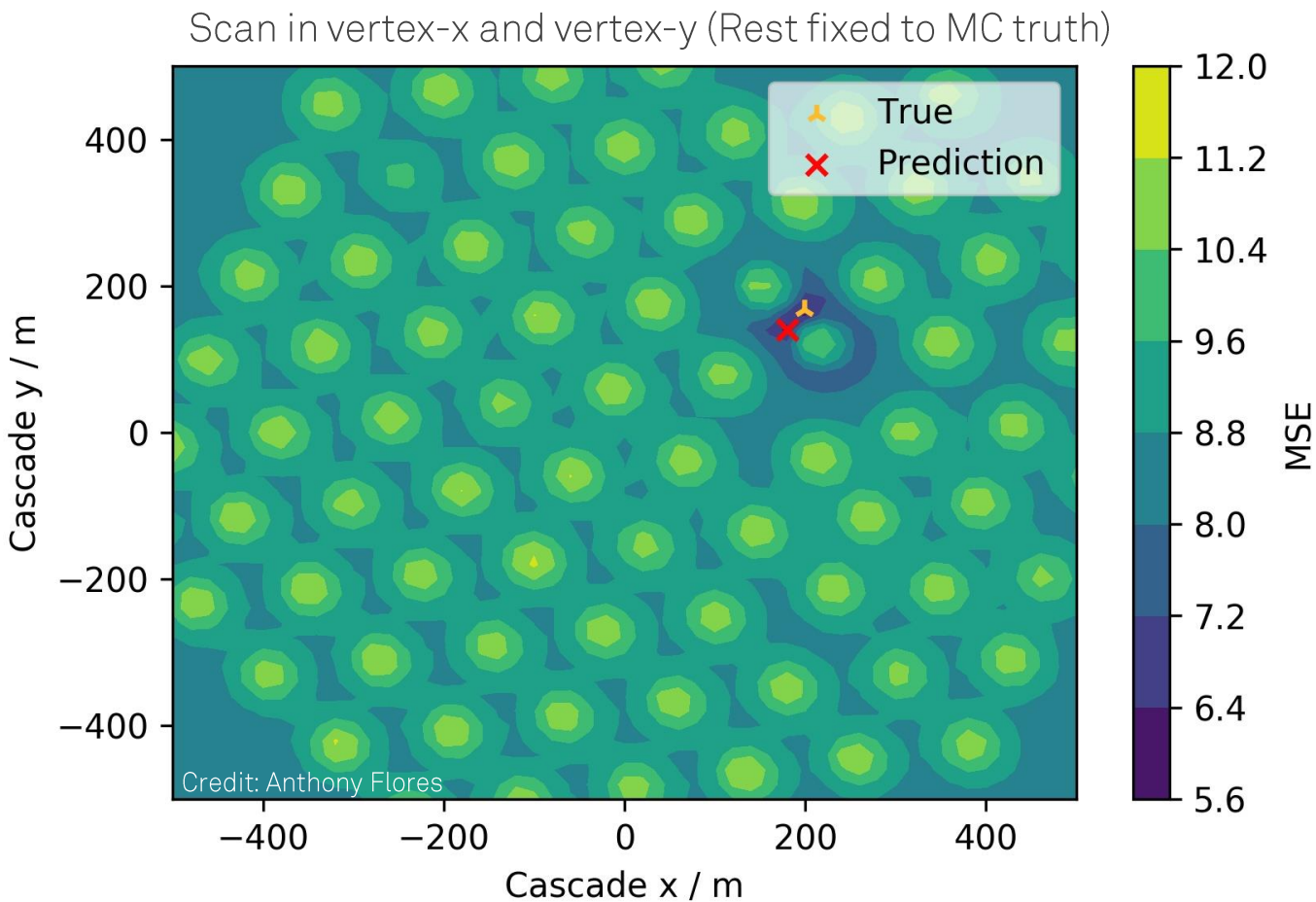
Cascade Generator – Latent Variables



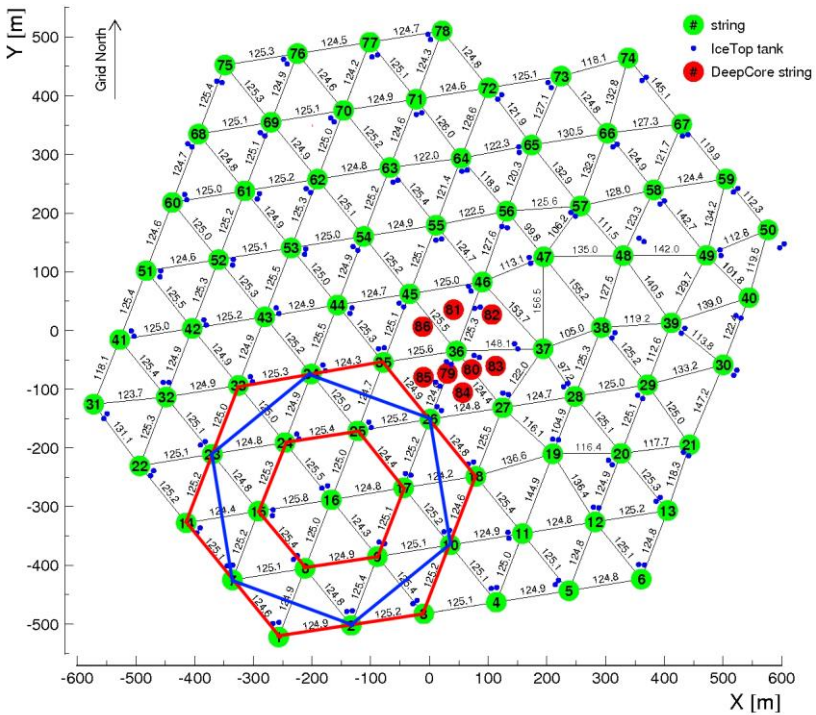
Cascade Generator – Scans

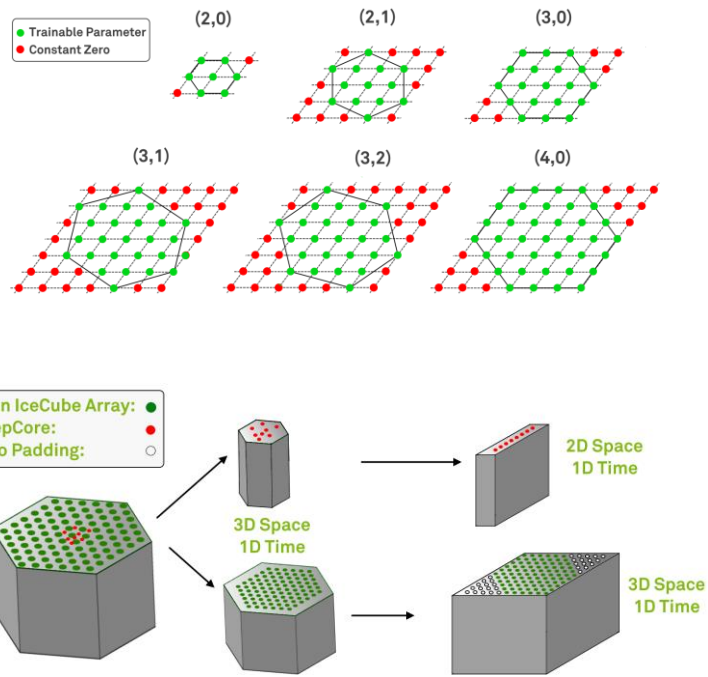


Cascade Generator – Scans

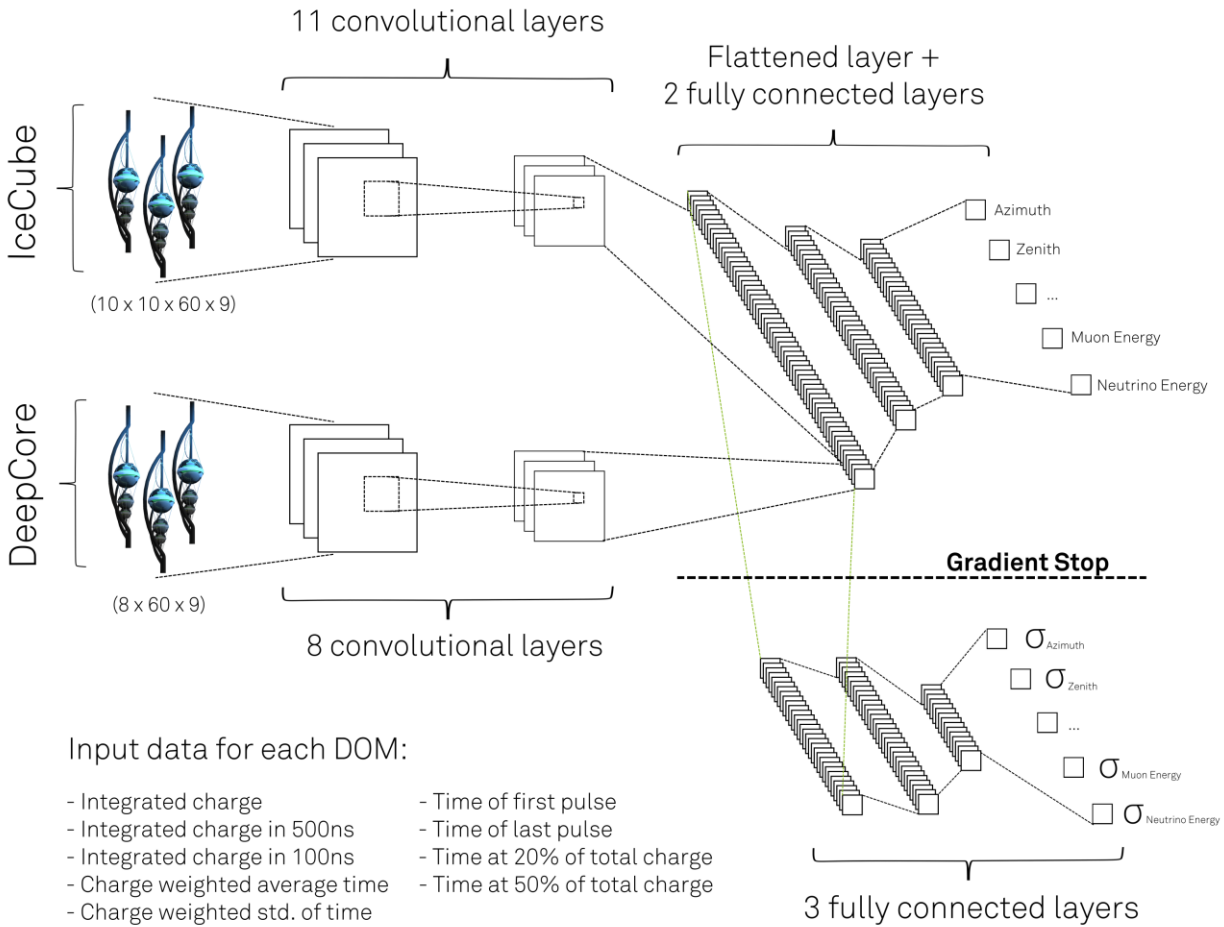


IceCube – Data Format





DNN Reco – Architecture

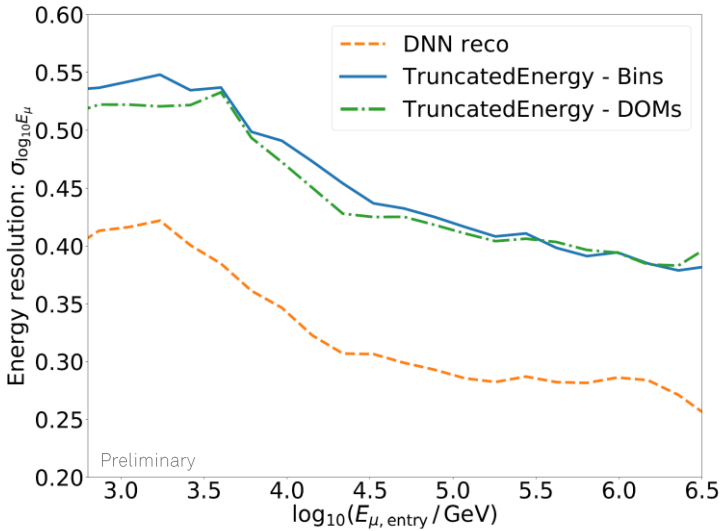


Input data for each DOM:

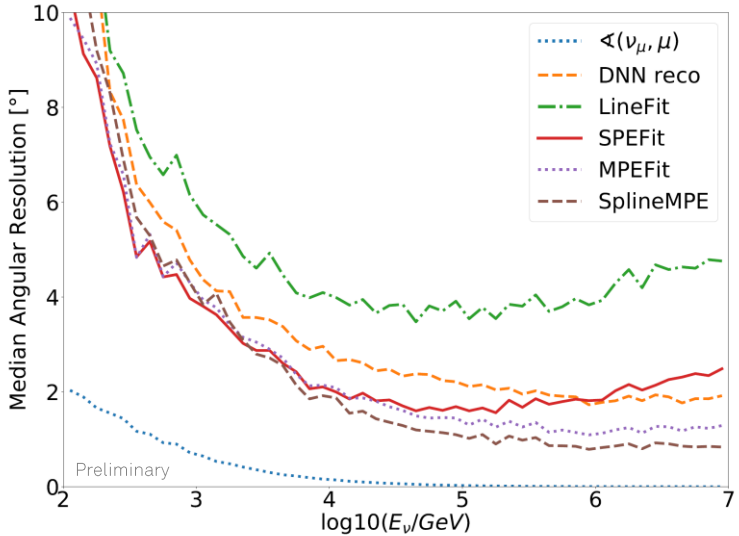
- Integrated charge
- Integrated charge in 500ns
- Integrated charge in 100ns
- Charge weighted average time
- Charge weighted std. of time
- Time of first pulse
- Time of last pulse
- Time at 20% of total charge
- Time at 50% of total charge

DNN reco – Track Performance

Muon Energy at Detector Entry

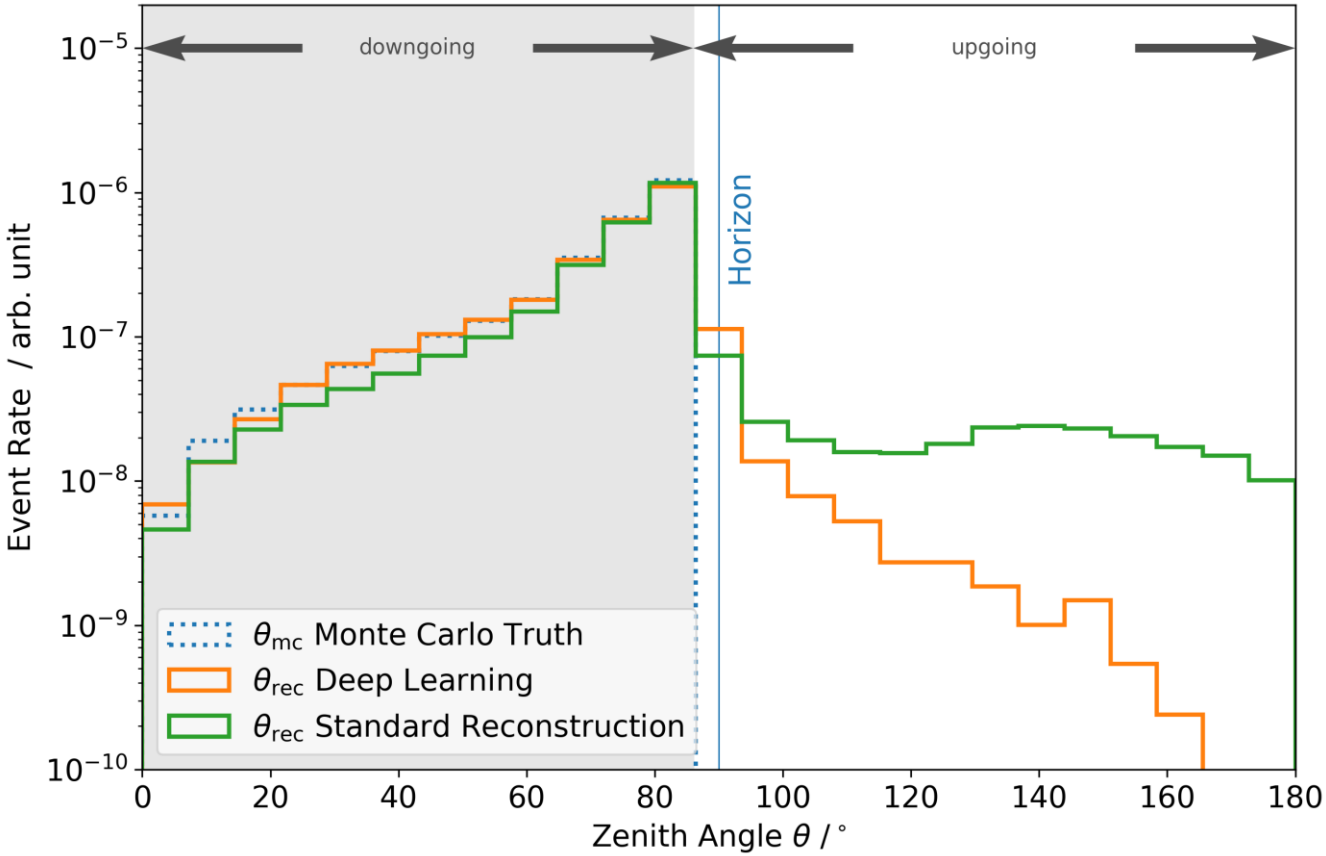


Median Angular Resolution



- Level 2 data: Final samples apply additional quality cuts
- Systematic uncertainties not included

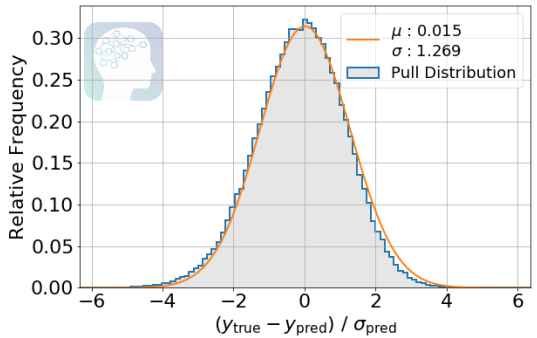
DNN reco – Track Performance

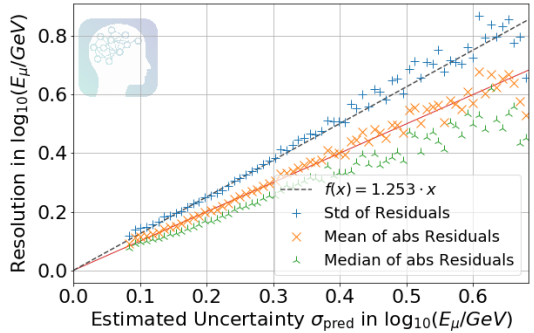


DNN reco – Uncertainty Estimation

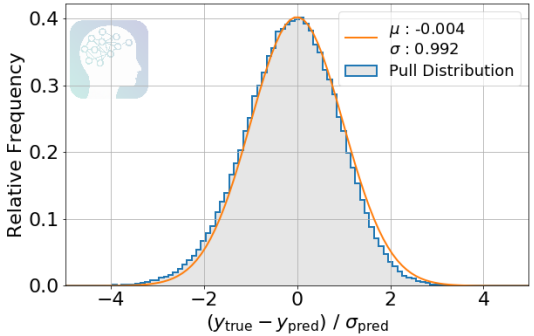
Muon Energy at Detector Entry

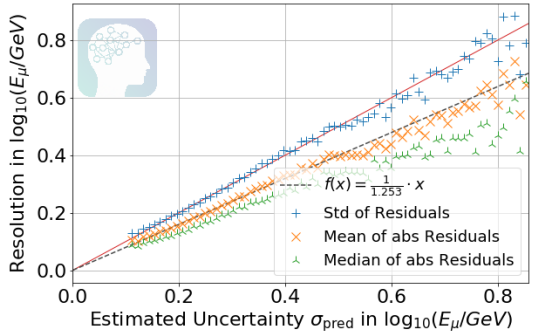
MSE of absolute Residuals



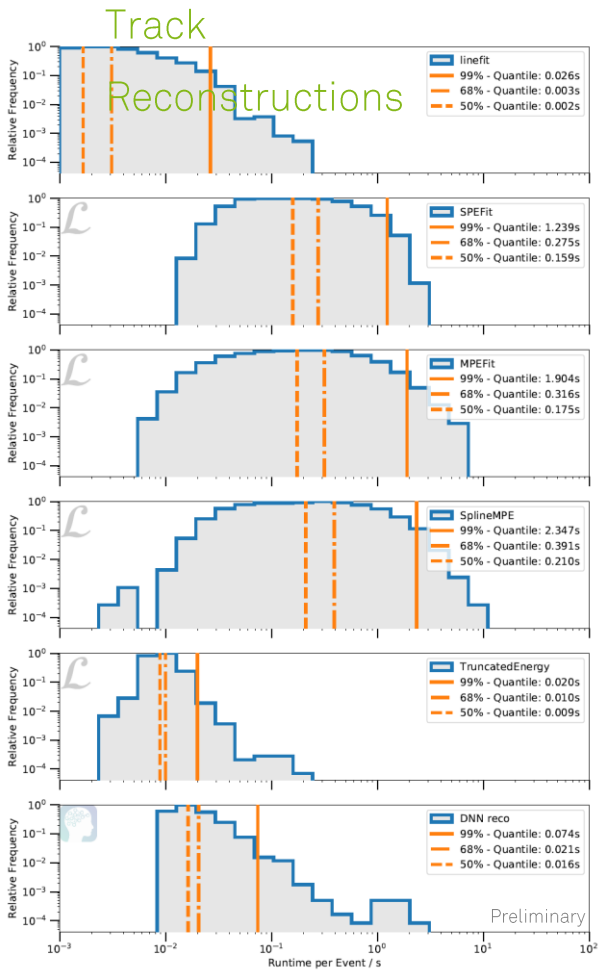
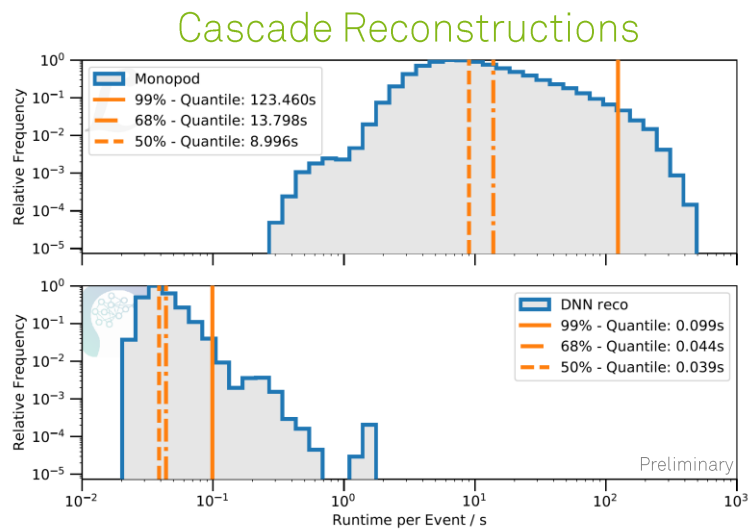


Gaussian Likelihood





DNN reco – Runtime



- Per event runtime for reconstruction
- More sophisticated reconstructions excluded in plots
- Smaller neural network architecture used for track reconstruction