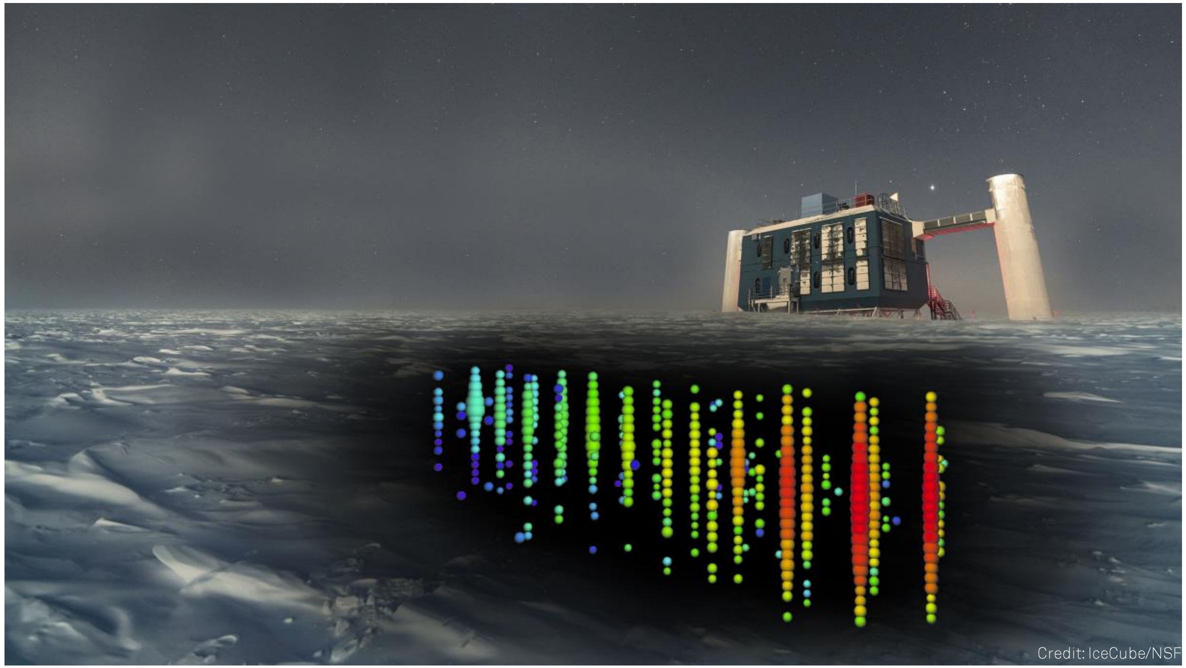
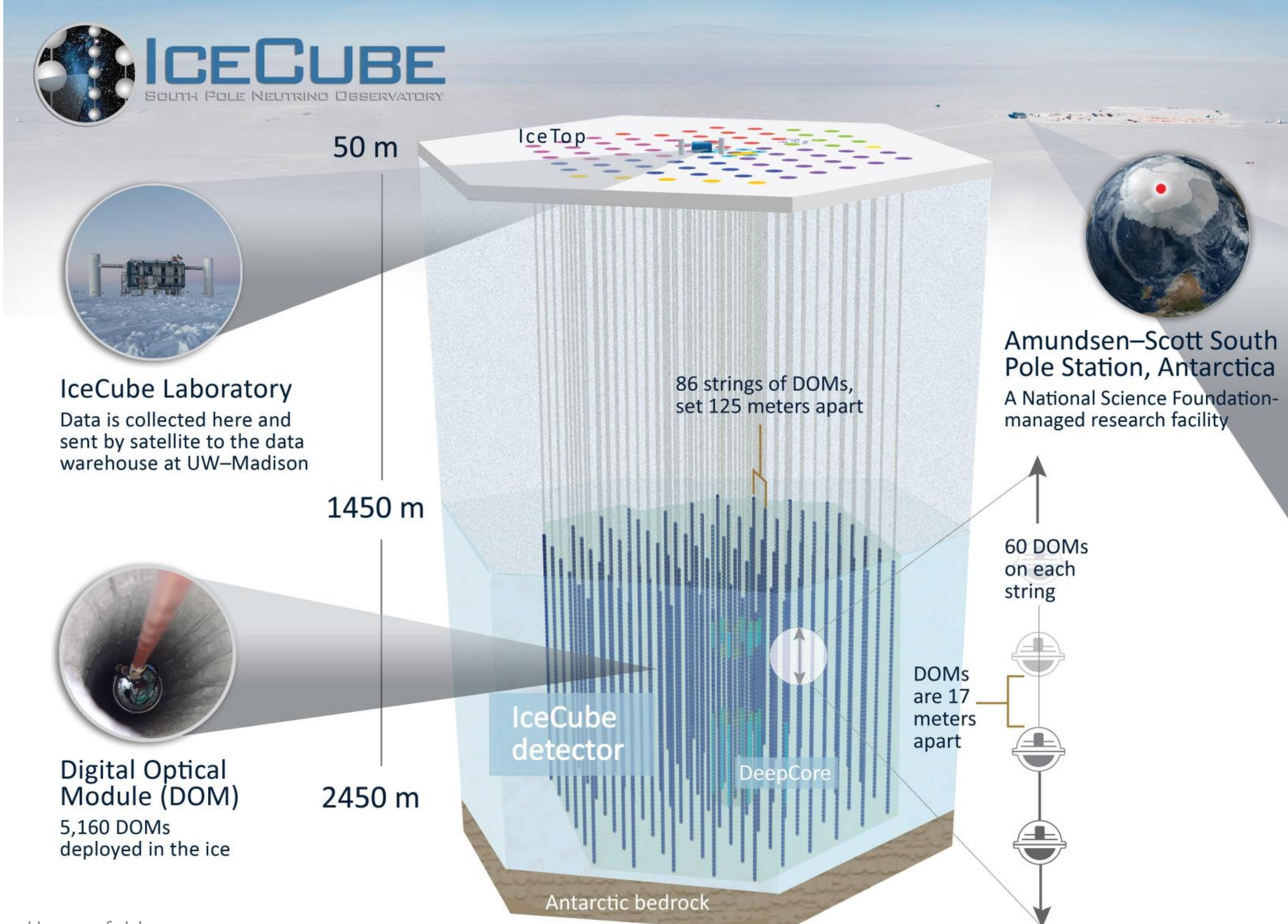


Exploitation of Symmetries and prior Knowledge in DL Architectures

Mirco Huennefeld
mirco.huennefeld@tu-dortmund.de

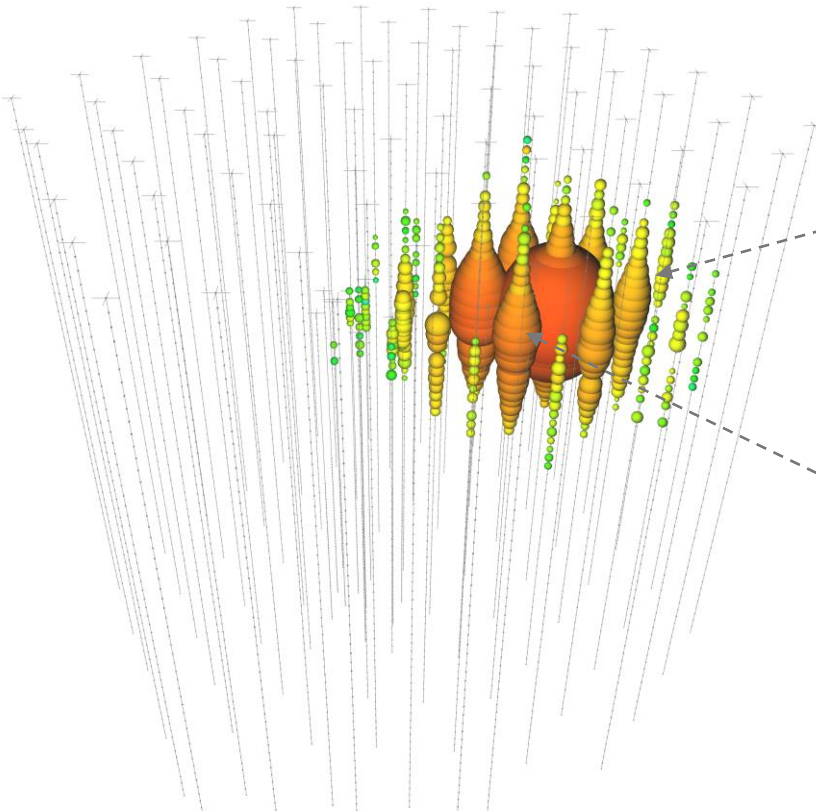
DL Workshop Aachen
February 18, 2020

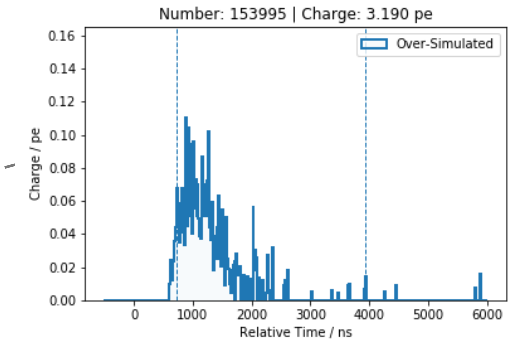


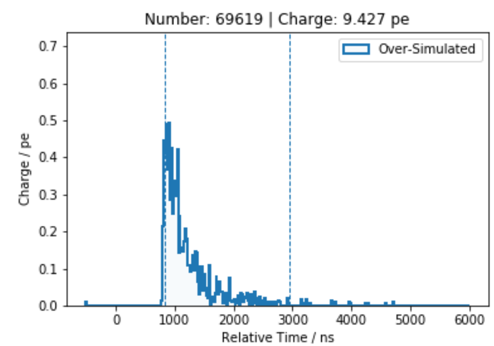
IceCube
South Pole Neutrino Observatory
50 m
IceTop
IceCube Laboratory
Data is collected here and sent by satellite to the data warehouse at UW–Madison
1450 m
Digital Optical Module (DOM)
5,160 DOMs deployed in the ice
2450 m
86 strings of DOMs, set 125 meters apart
IceCube detector
DeepCore
Antarctic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

Introduction – Cascade Reconstruction in IceCube

Goal: Cascade Reconstruction







Overview

- Define Goal: Cascade Reconstruction
- Convolutional Neural Network (CNN)
- Maximum Likelihood Estimation (MLE)
- Hybrid MLE and DL reconstruction method
- Summary

Credit: Pixabay.com

Overview

- Define Goal: Cascade Reconstruction
- Convolutional Neural Network (CNN)
- Maximum Likelihood Estimation (MLE)
- Hybrid MLE and DL reconstruction method
- Summary

Credit: Pixabay.com

Convolutional Neural Network

- What do I know about my data?
  - Physical processes are invariant in time and space
  - Coincidence between nearby digital optical modules (DOMs) is important (locality)
  - 3 spatial + 1 temporal dimension with variable size
  - Strings located on approximately triangular grid

Potential Solutions:

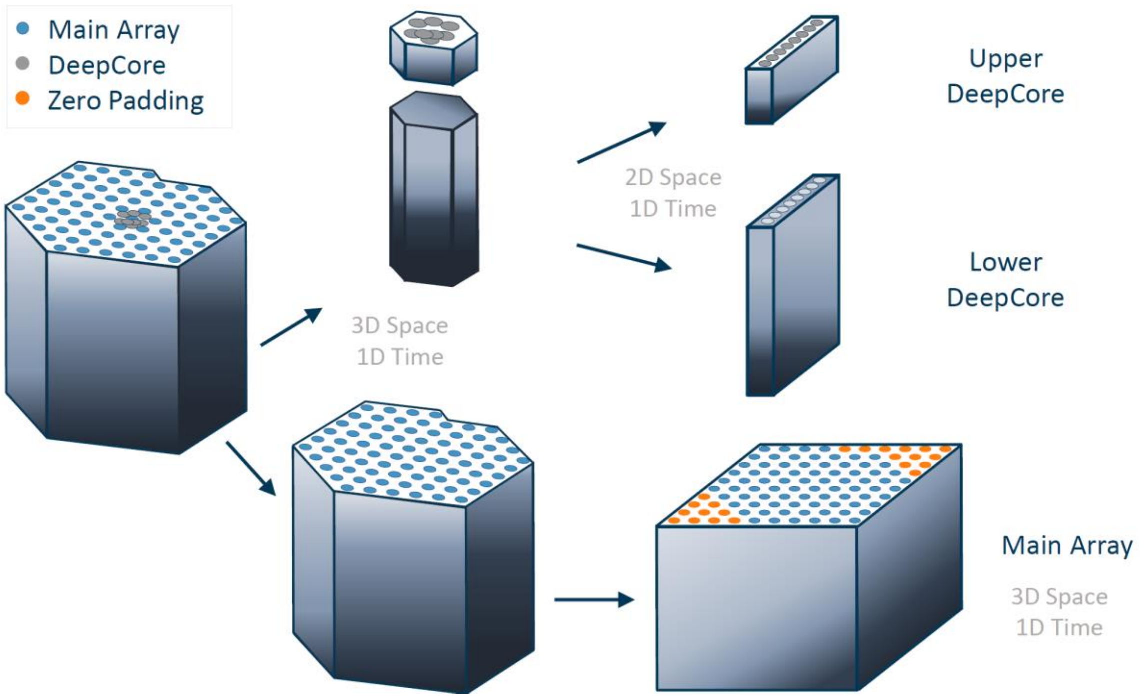
A Convolutional Neural Network (CNN) can utilize this

Reduce and unify dimensionality

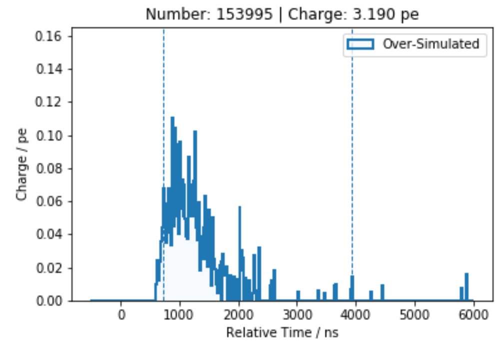
Transform data and use hexagonal convolution kernels

Convolutional Neural Network – Data Preprocessing

Transform data to orthogonal grid



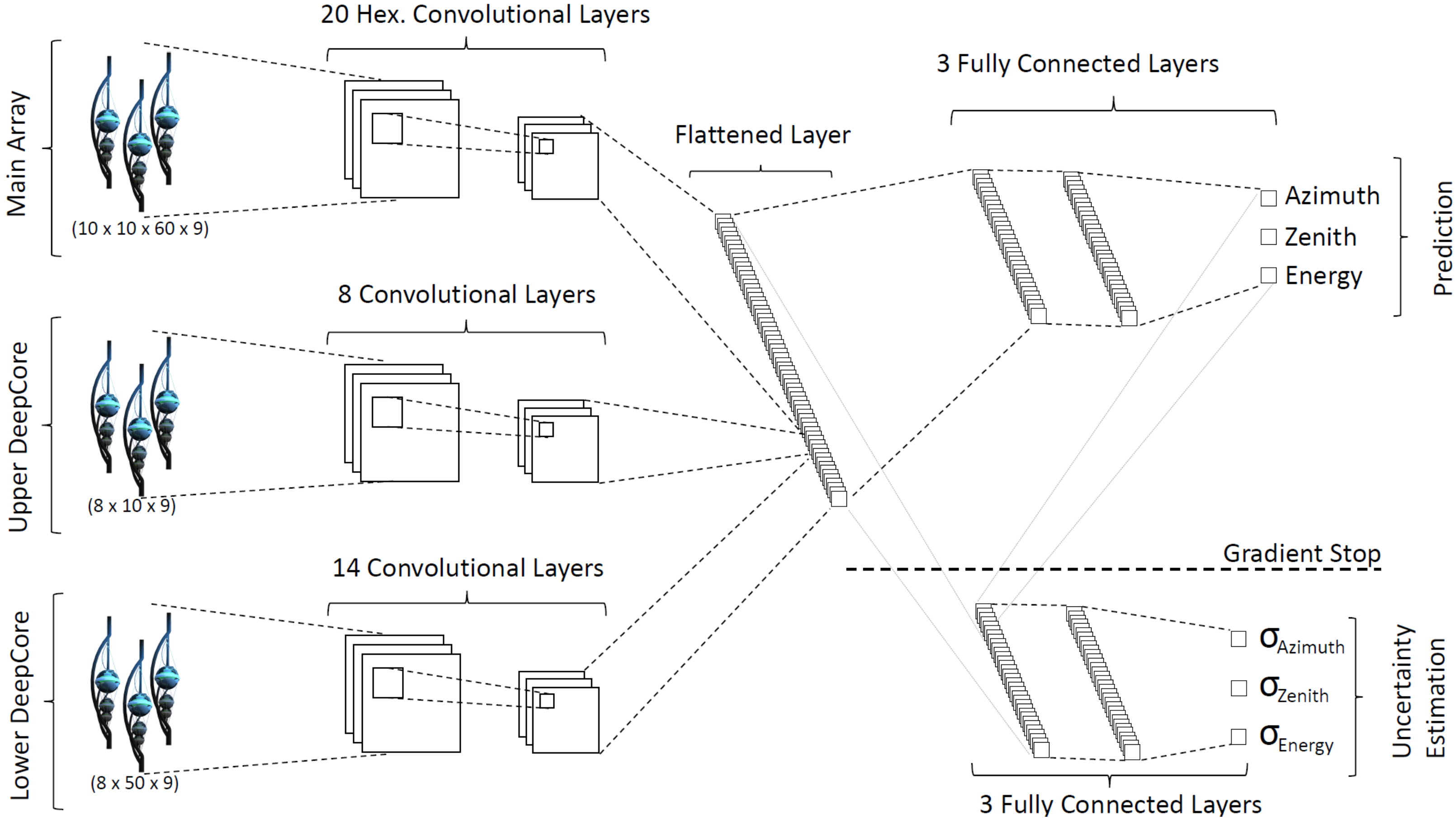
Reduce dimensionality



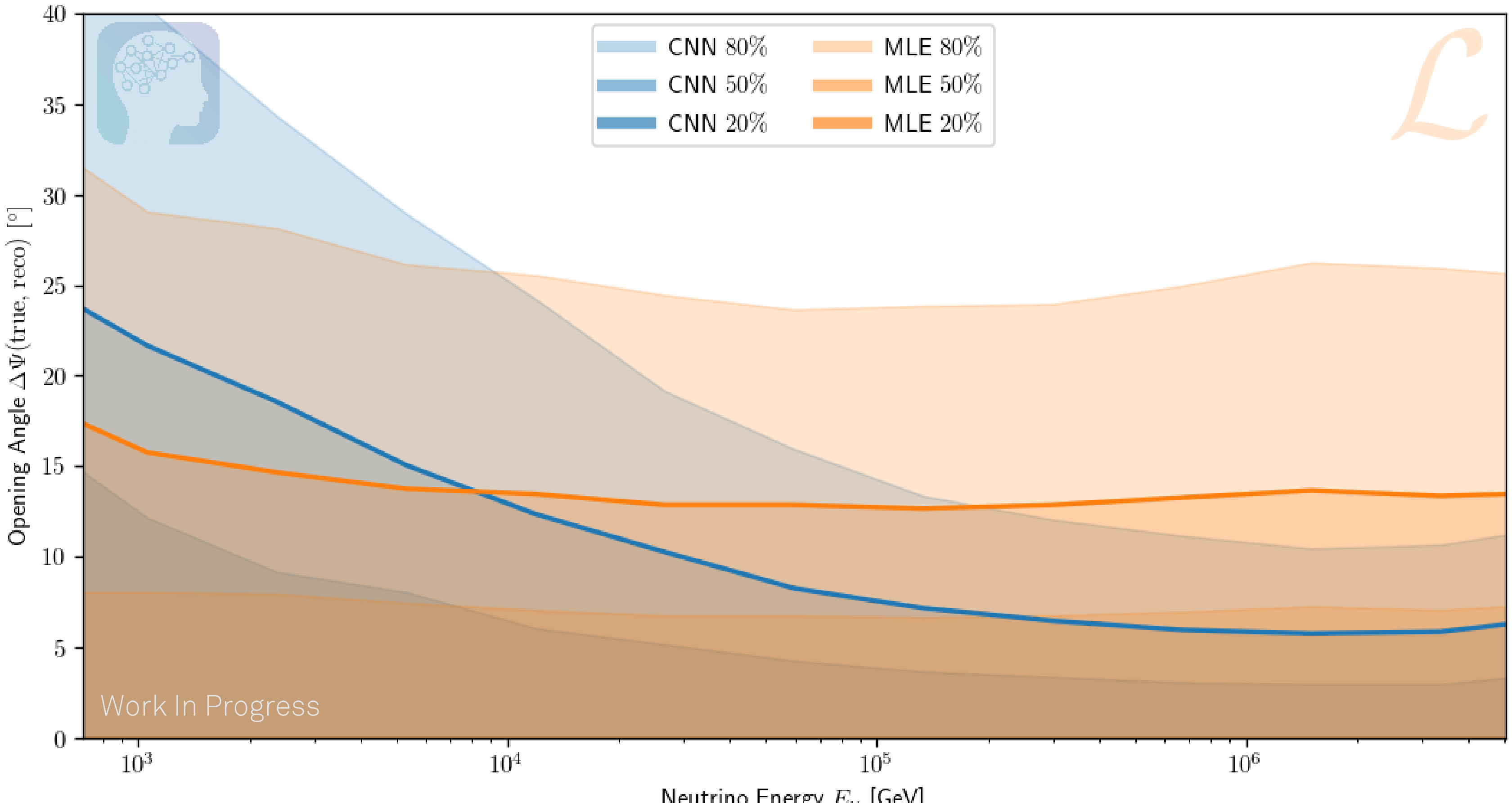
For each DOM:

- Total Charge
- Charge within 100 ns
- Charge within 500 ns
- Time of first pulse
- ...

Convolutional Neural Network – Architecture



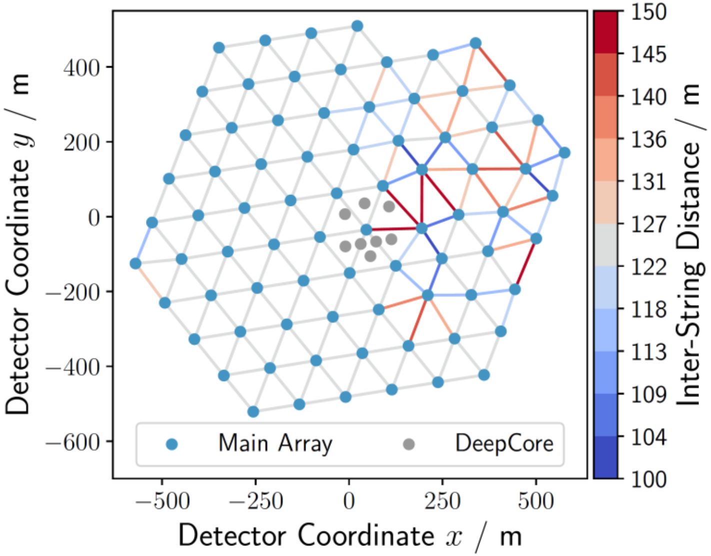
Convolutional Neural Network – Performance

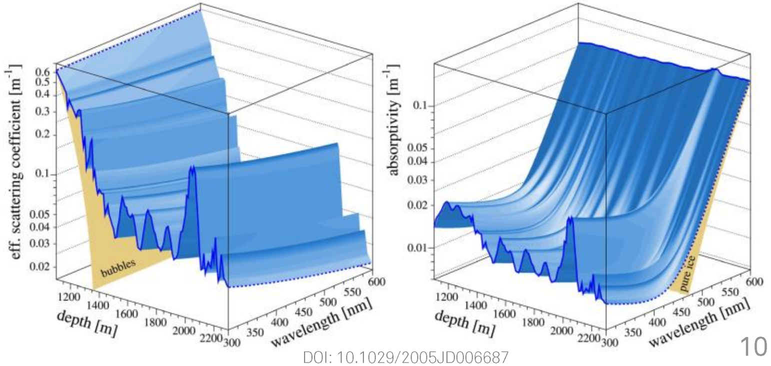


Convolutional Neural Network – Summary

- ⊕ Translational invariance and locality is used
- ⊖ More information and symmetries are available
- ⊖ Detector parts must be separated
- ⊖ Assumptions imposed by CNNs are only approximately met in IceCube
  - ⊖ Irregularities in detector grid
  - ⊖ No real translational invariance in observable space

Graph and recurrent neural networks can help in some areas, but also generally fail to include available information





DOI: 10.1029/2005JD006687

Overview

- Define Goal: Cascade Reconstruction
- Convolutional Neural Network (CNN)
- Maximum Likelihood Estimation (MLE)
- Hybrid MLE and DL reconstruction method
- Summary

Credit: Pixabay.com

Maximum Likelihood Estimation

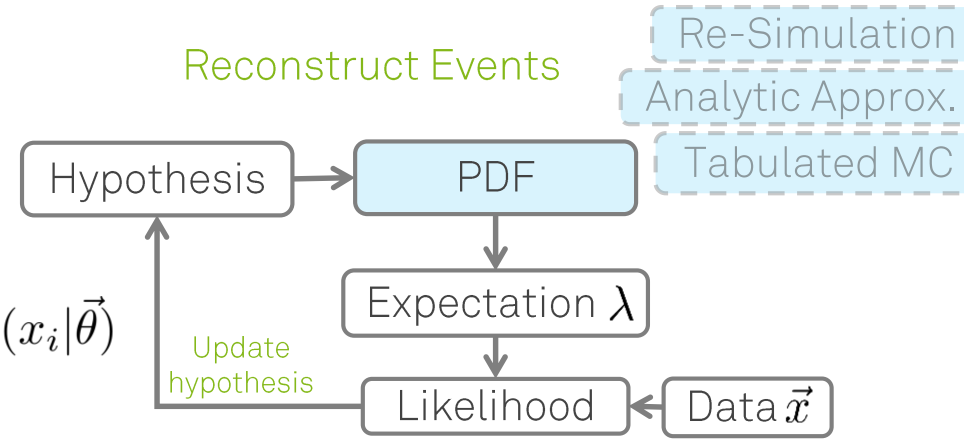
Alter event hypothesis until it matches data
Cascade Hypothesis:

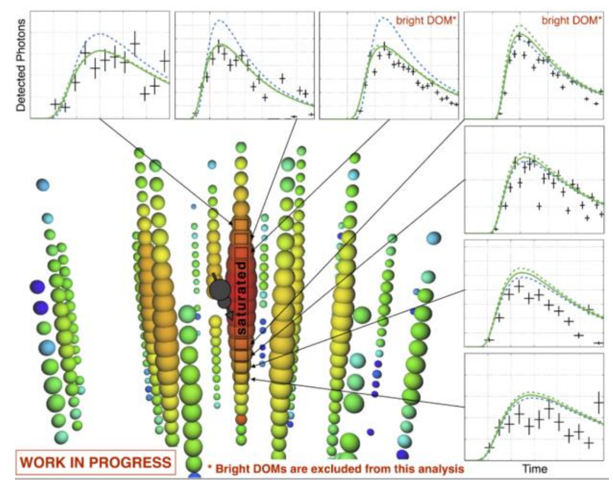
$$\vec{\theta} = (x, y, z, \varphi, \theta, E, t)$$ 7 free parameters

Properties:

$$\mathcal{L}(\vec{x}|\vec{\theta}) = \prod_i p(x_i|\vec{\theta})$$

- ⊕ Physics knowledge is incorporated into the Likelihood and PDF
- ⊕ Exact detector geometry can be used
- ⊕ In theory: optimum of what we can do
- ⊖ Often forced to simplify PDF and Likelihood
- ⊖ Difficult to find global minimum





Credit: HESE team

Overview

- Define Goal: Cascade Reconstruction
- Convolutional Neural Network (CNN)
- Maximum Likelihood Estimation (MLE)
- Hybrid MLE and DL reconstruction method
- Summary

Credit: Pixabay.com

Hybrid Method – General Idea

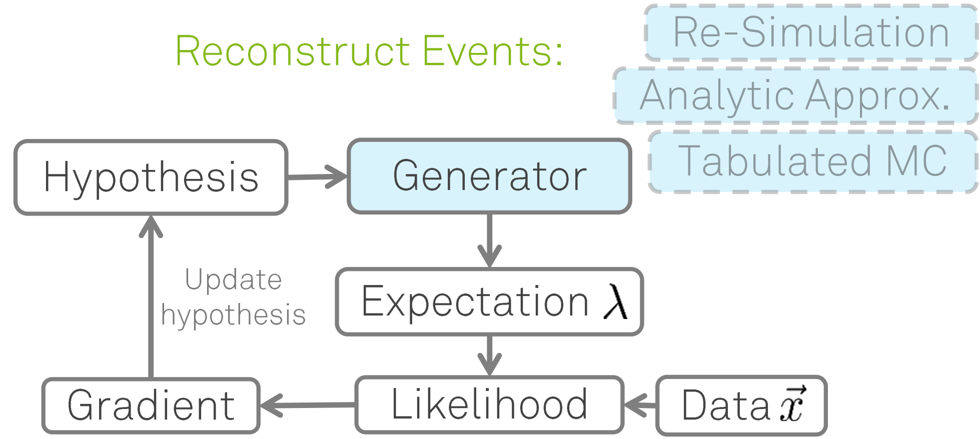
Combine strengths of neural networks and maximum-likelihood methods

Properties of Generator NN:

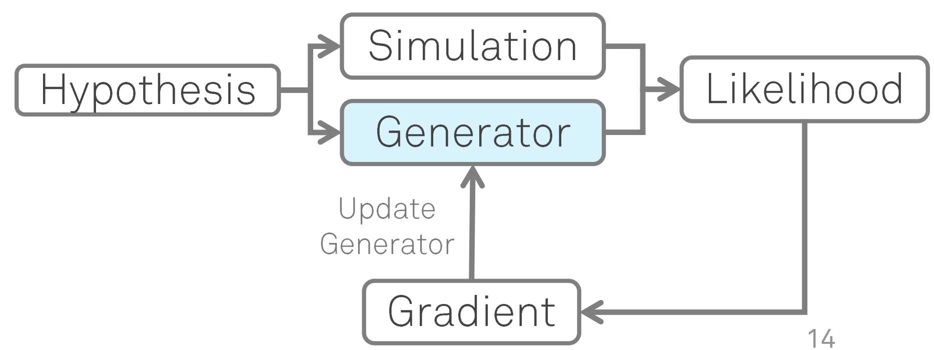
- Fast approximation of MC simulation
- No critic NN or randomness necessary
- Physics knowledge & symmetries can easily be included
- Exact detector geometry can be used
- Use in reverse mode for reconstruction

  → Fully differentiable: Gradient Descent

Reconstruct Events:



Train Generator:



Hybrid Method – Generator Architecture

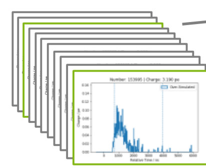
Cascade Hypothesis

$$\underbrace{\vec{\theta} = (x, y, z, \varphi, \theta, E, t)}$$

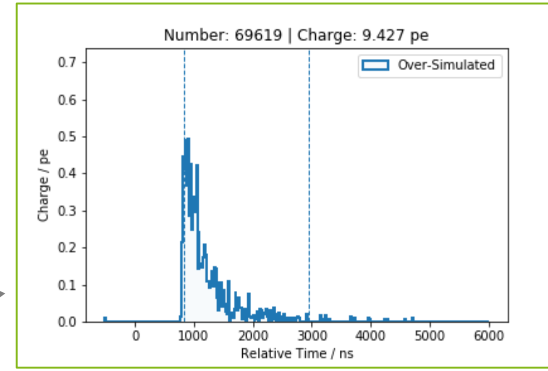
→ Generator →

Neural Network that maps input [?, 7] to output [?, 5160, n] with number of time bins n

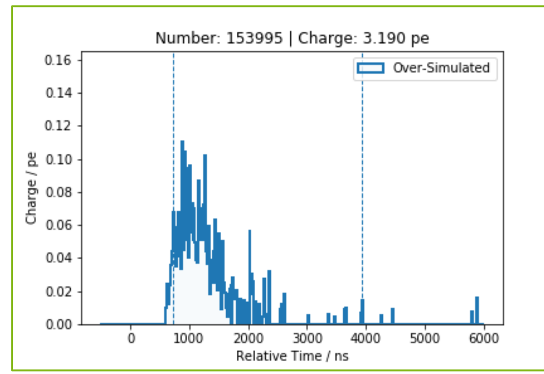
5160 Waveforms for each of the DOMs



DOM #42

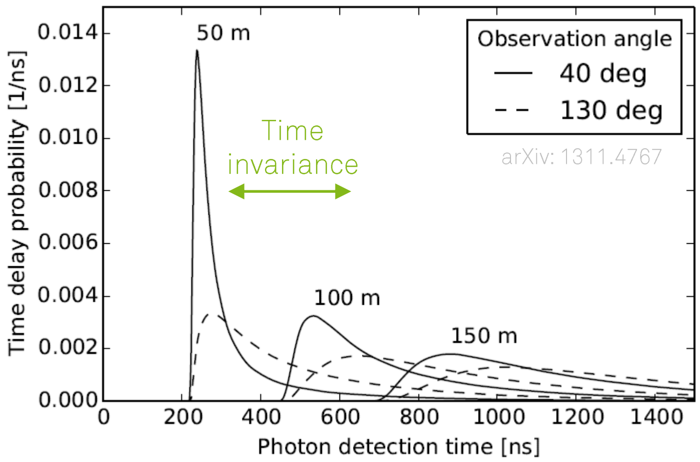


DOM #1337



Hybrid Method – Inclusion of Information

- Examples how specific information can be included:
  - Parameterization of waveforms at each DOM
  - Light yield scales linearly with cascade energy: Apply this scaling before adding noise and detector response
  - Time invariance simply shifts waveforms in time
- Can decouple physics and detector effects



→ Easier to include information in forward direction when not convolved with detector response yet

→ We know how to do this (we simulate the data!)

Hybrid Method – Inclusion of Information

- Light yield at each DOM (prior to local ice effects) only depends on relative displacement and orientation to cascade:
  - Weight sharing between DOMs
  - Translational invariance

→ Exact detector geometry and translational invariance can be exploited

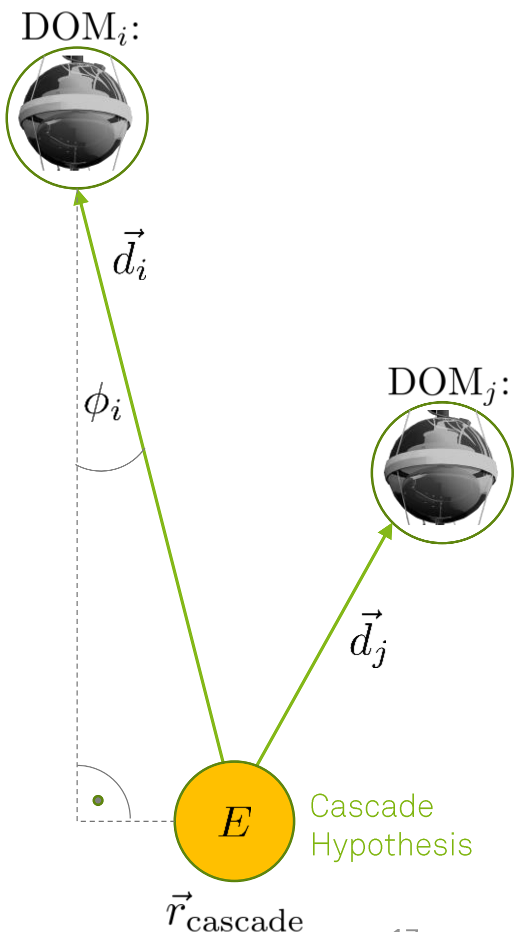
Input values per DOM:

$\vec{d}, \phi, \ldots$ } Relative parameters: include geometry and translational invariance

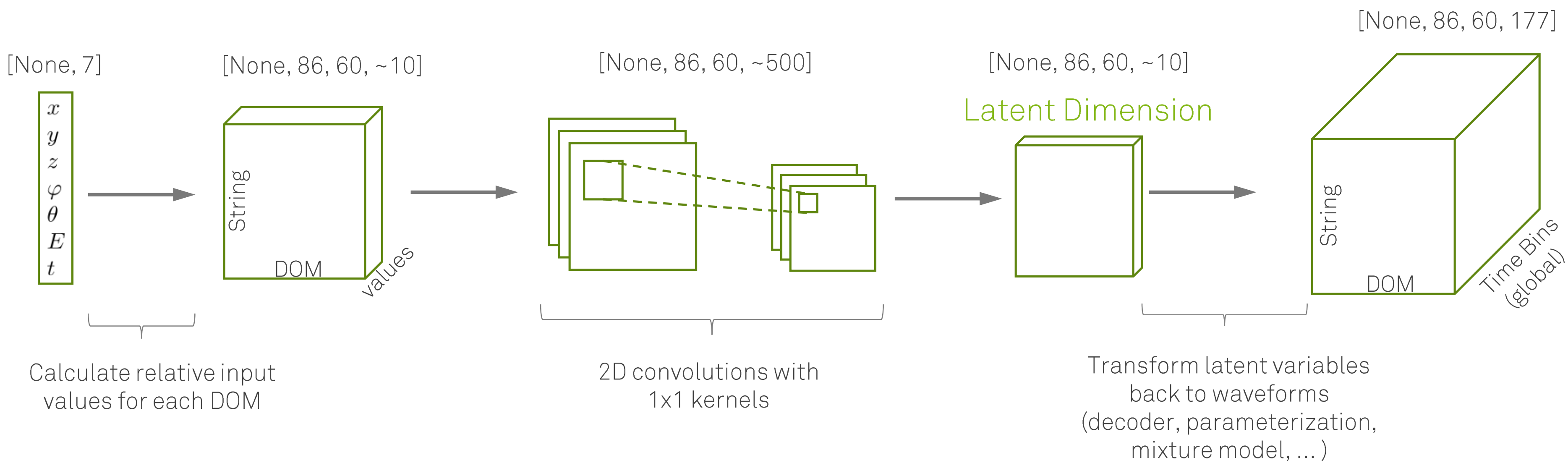
φ_{cascade}
θ_{cascade}
$\vec{r}_{\text{cascade}}$
$E, \ldots$

Weight sharing over DOMs:
Exploit invariances in physics variable space

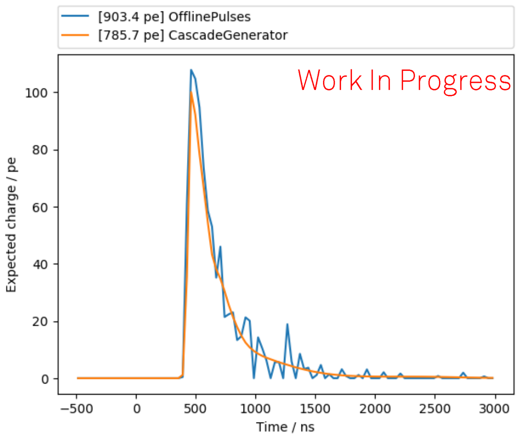
Additional independent weights per DOM:
Account for asymmetries and inhomogeneities

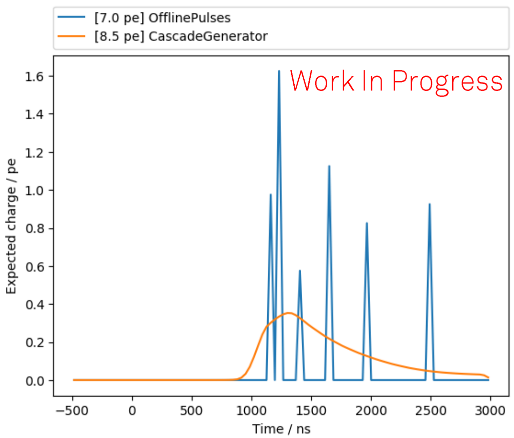


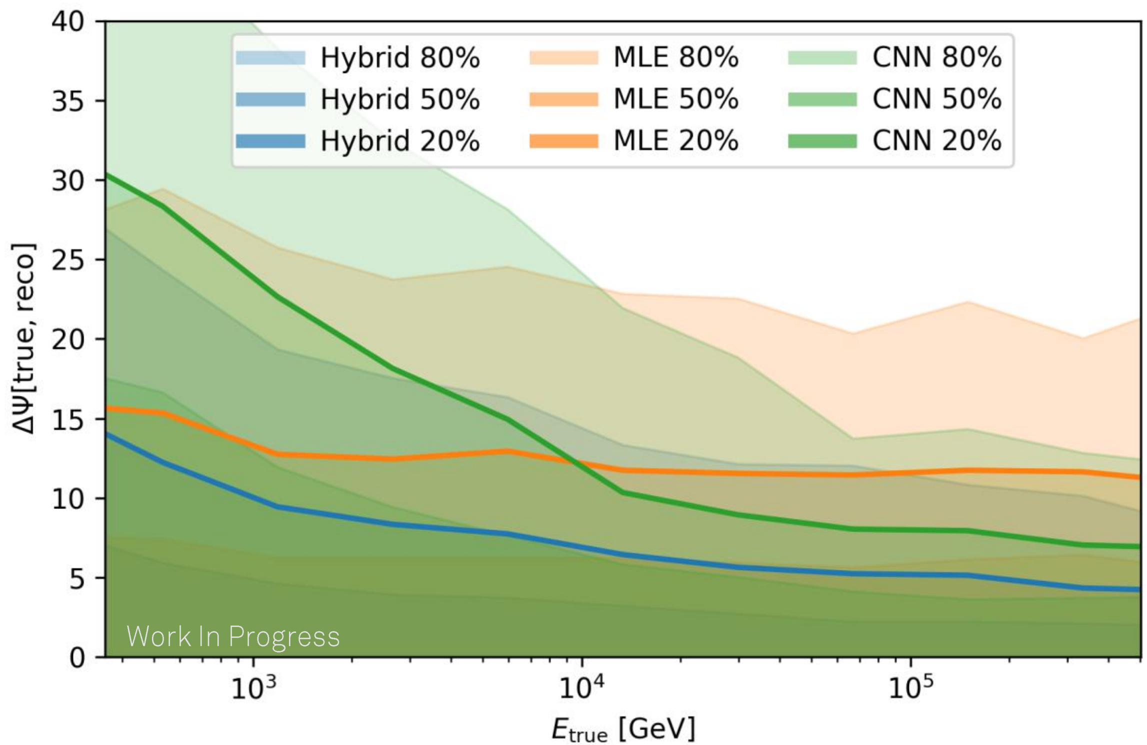
Hybrid Method – Generator Architecture



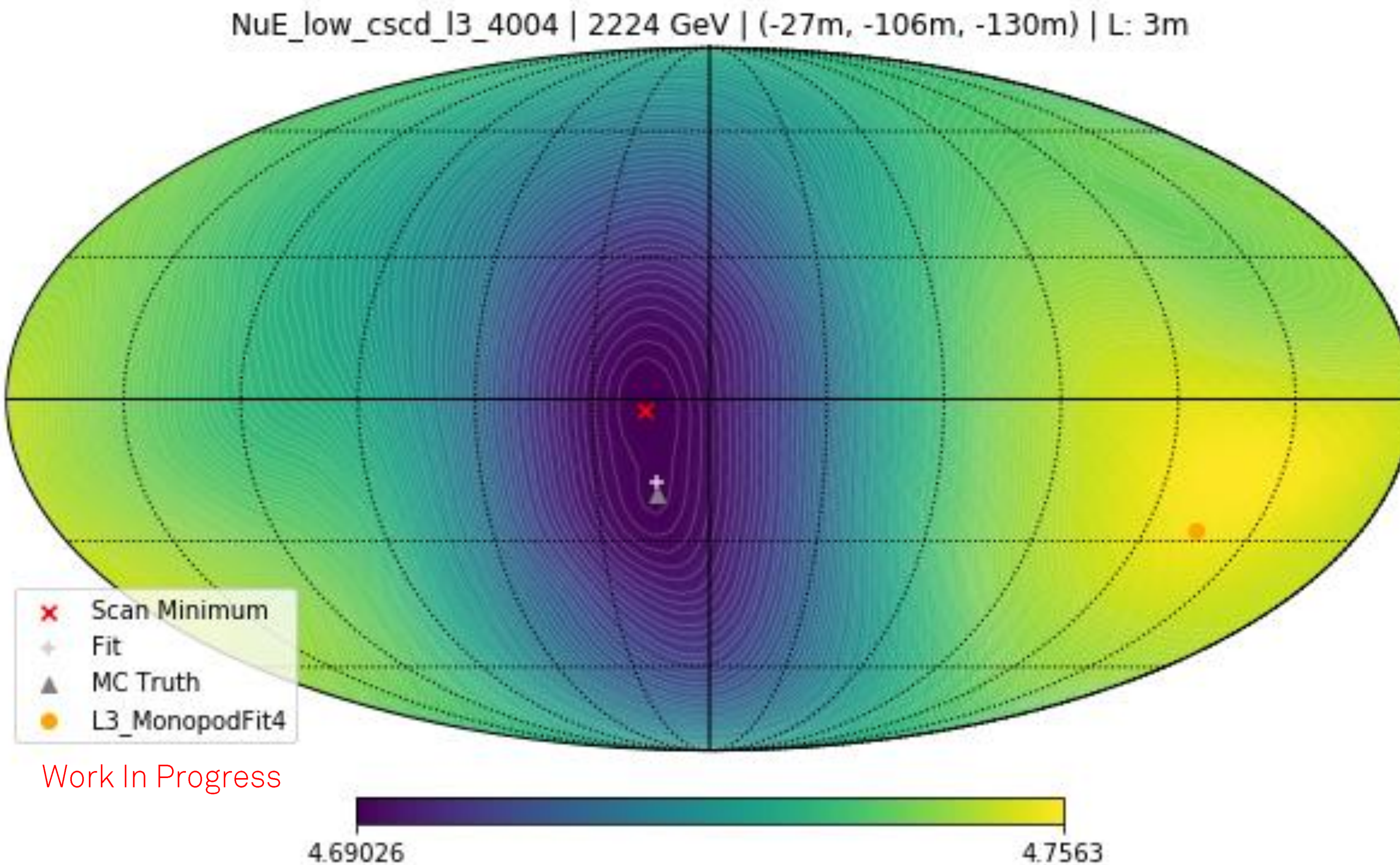
Hybrid Method – Results



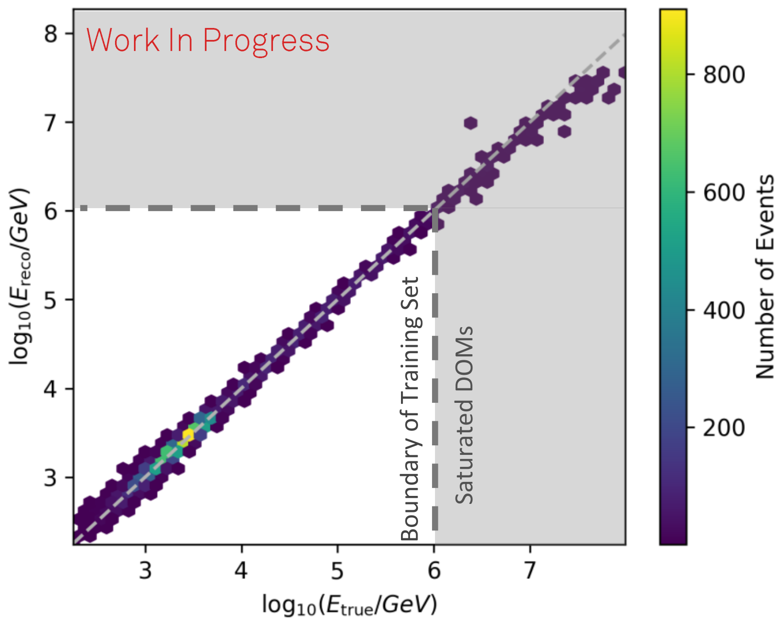


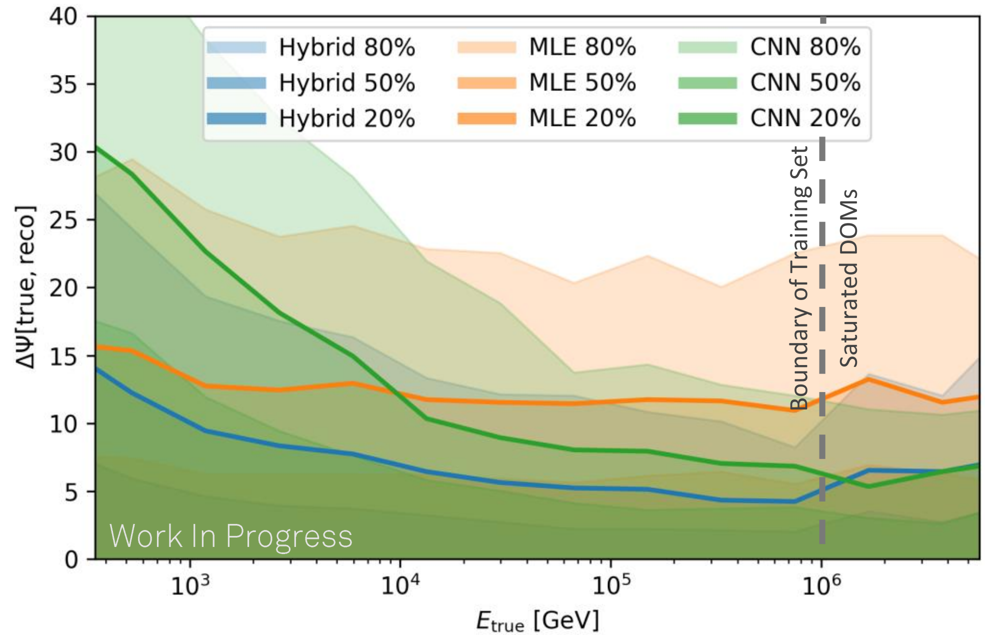


Hybrid Method – Likelihood Scan



Hybrid Method – Extrapolating beyond Training Data Coverage





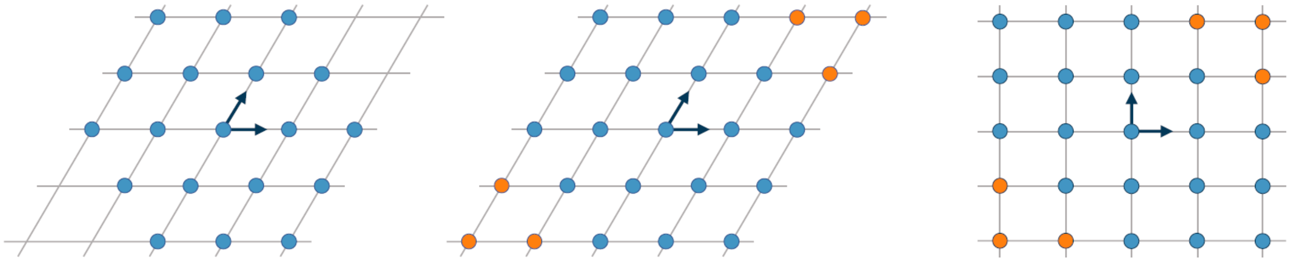
Model can extrapolate (until DOMs start to saturate) due to incorporation of linear light scaling into architecture

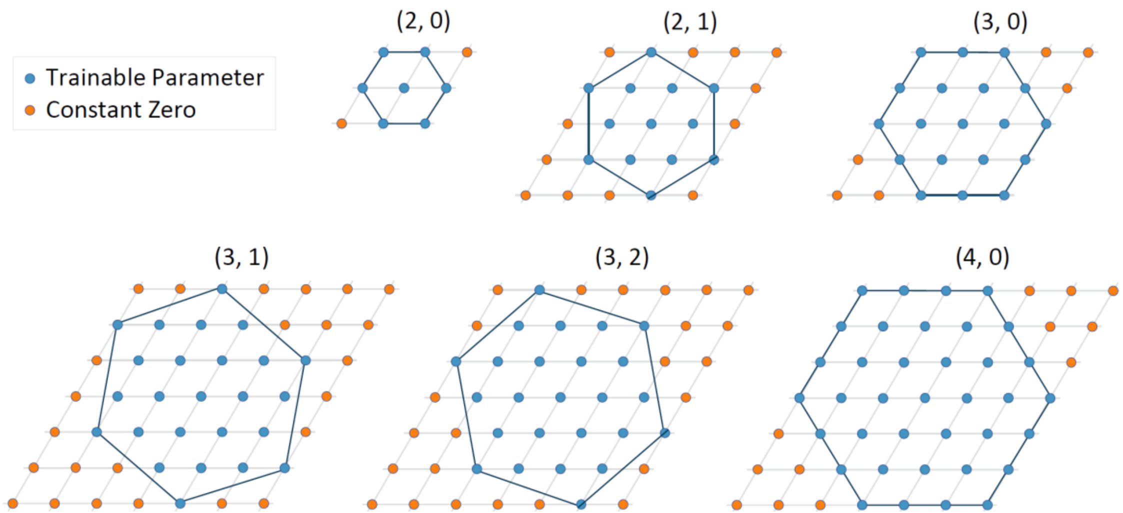
Summary

- Goal: cascade reconstruction
- Convolutional Neural Network (CNN)
  - Can exploit (approximate) translational invariance in time and space of measured data
  - Detector geometry and irregularities generate problems
  - Inclusion of further information is hard in backwards direction (measurement -> physics)
- Maximum Likelihood Estimation (MLE)
  - In theory: optimum of what we can achieve, in practice: forced to apply simplifications
- Hybrid Method
  - Exploitation of **exact** detector geometry, spatial and time invariance
  - Easy to include further information in forward direction (physics -> measurement)
  - Combines strengths of and surpasses both MLE and CNN reconstruction methods

Backup

Convolutional Neural Network – Hexagon Kernels

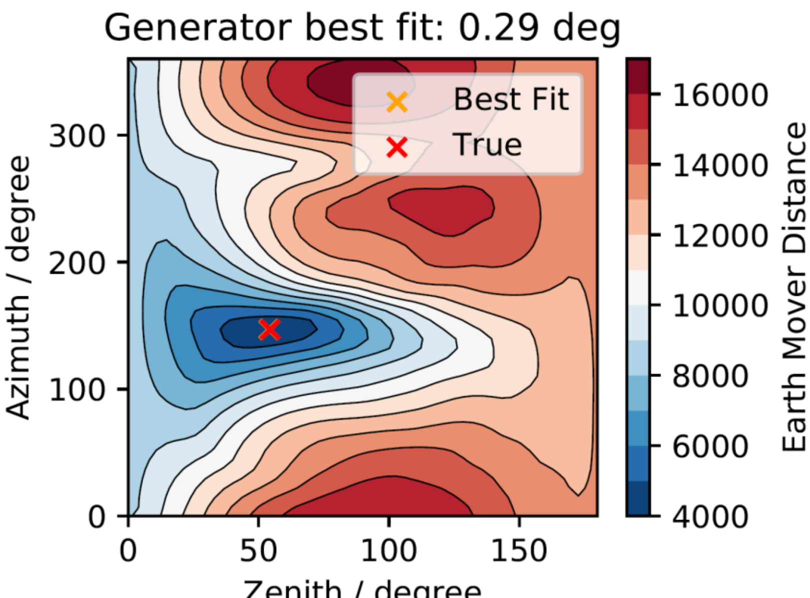


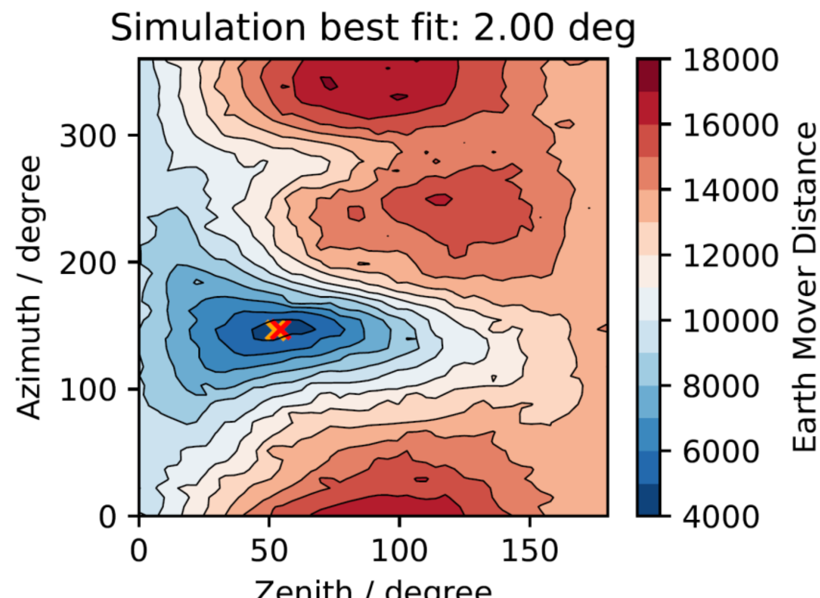


Cascade Generator vs. Direct Re-Simulation

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

$\Big\}$ Potential to surpass direct re-simulation





$\mathcal{O}(s)$ vs $\mathcal{O}(10-100\mathrm{h})$

Hybrid Method– Angular Resolution (without DeepCore)