Using Graph Neural Networks, for Cosmic-Ray Composition Analysis at IceCube Observatory

Paras Koundal

paraskoundal.com

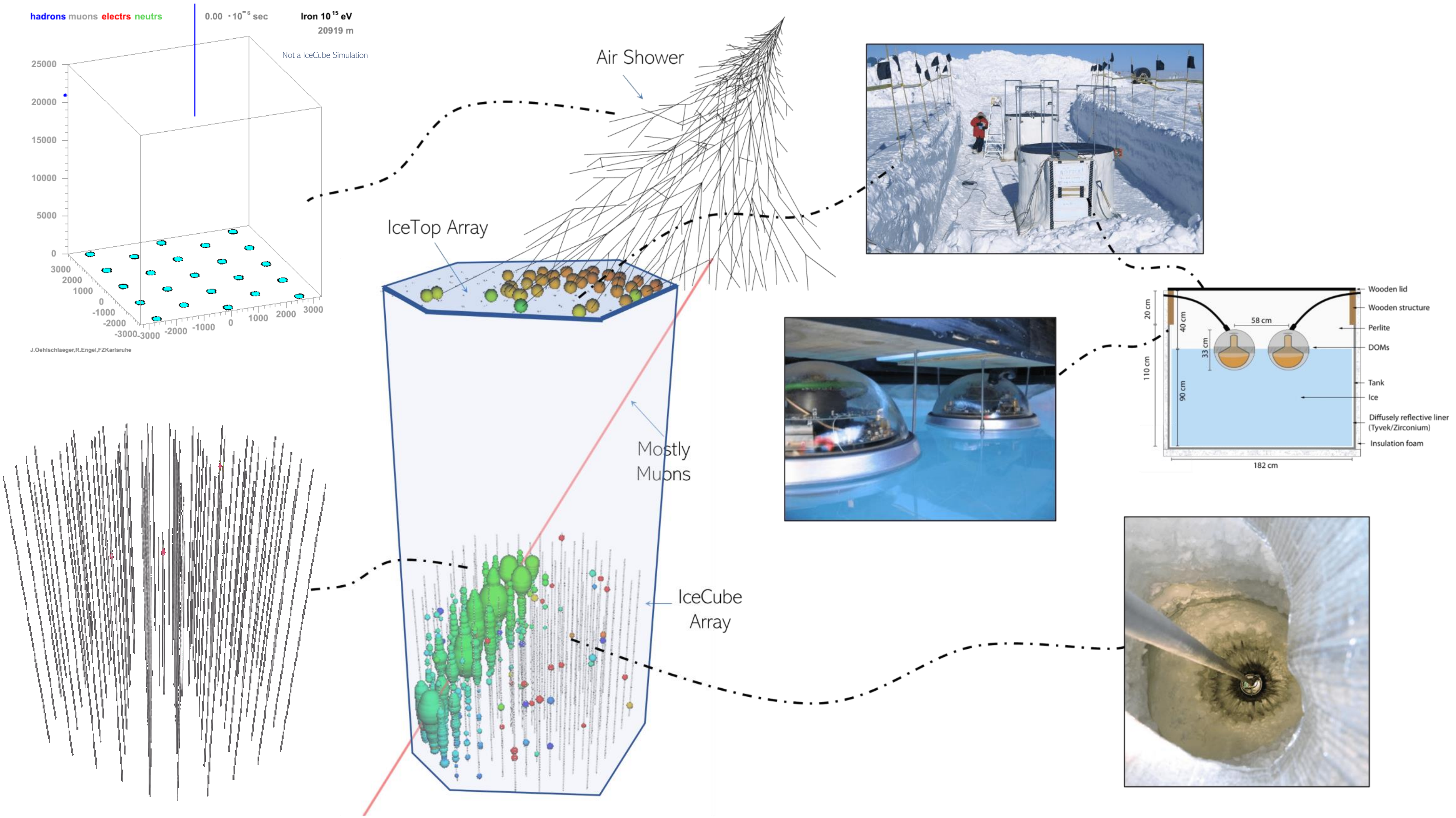
Karlsruhe Institute of Technology, Germany

/ParasKoundal

Talk Outline

- Introduction
    - CRays at IceCube
    - Previous Composition Analysis at IceCube
- New Composition Parameters
    - Muon-Spread Dependent Parameters
    - Muon Number Dependent Parameter
    - Muon Energy-Deposit Dependent Parameters
- Graph Neural Networks
    - Message Passing in GNNs
    - Improving Previous Architecture

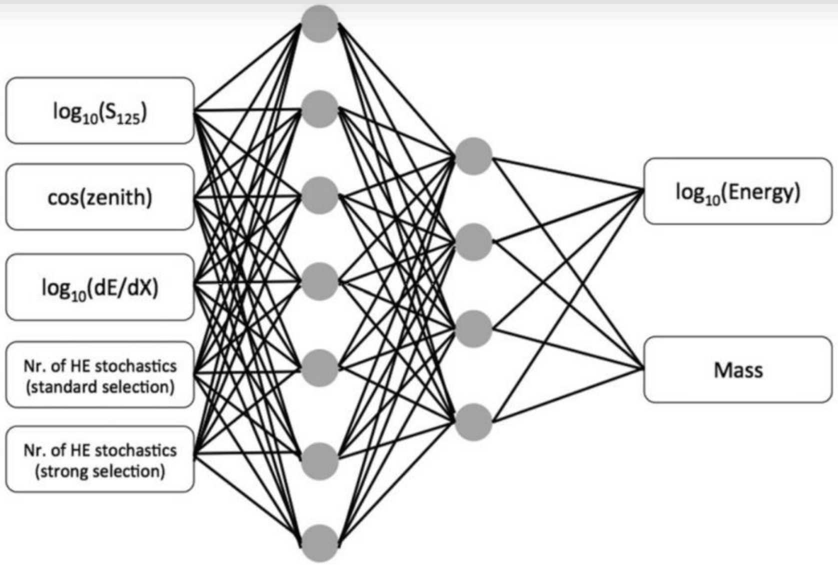
CR Analysis @ IceCube

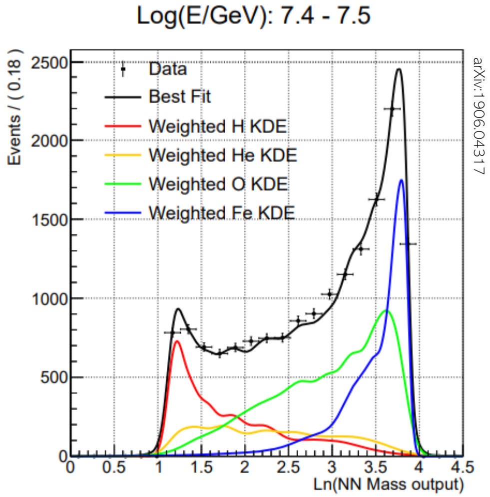


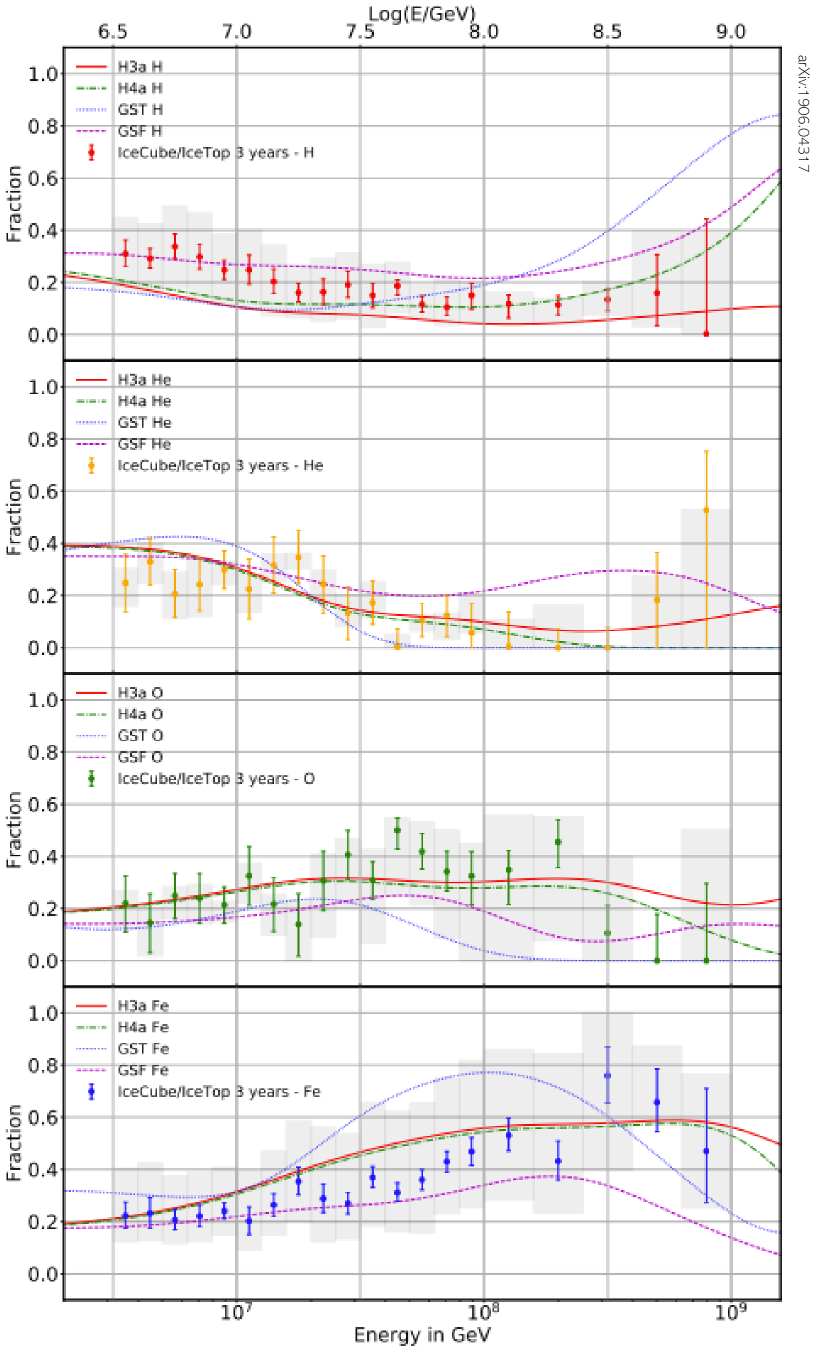
Previous Work

Cosmic ray spectrum and composition from PeV to EeV using 3 years of data from IceTop and IceCube

M. G. Aartsen *et al.* (IceCube Collaboration)
Phys. Rev. D **100**, 082002 – Published 23 October 2019







Scope for Improvement

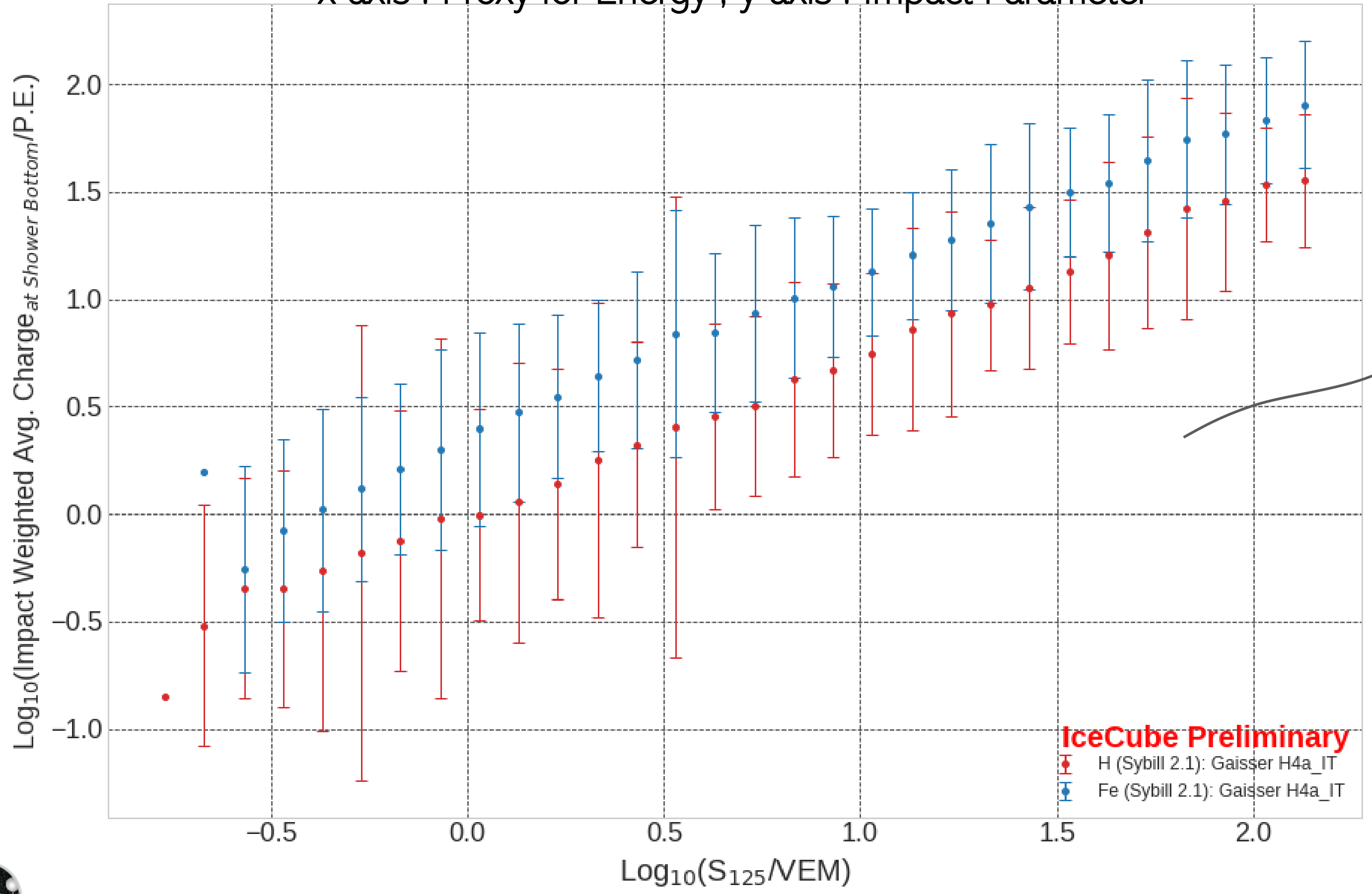
- Focus only on <u>per-event based composition</u> analysis
- Invent <u>newer composition-sensitive parameters</u>
  - Focus on in-ice deposit (primarily muon-deposit)
- Use and Improve SOTA Deep Learning Methods for composition-analysis
  - <u>Use full in-ice footprint for composition analysis</u>

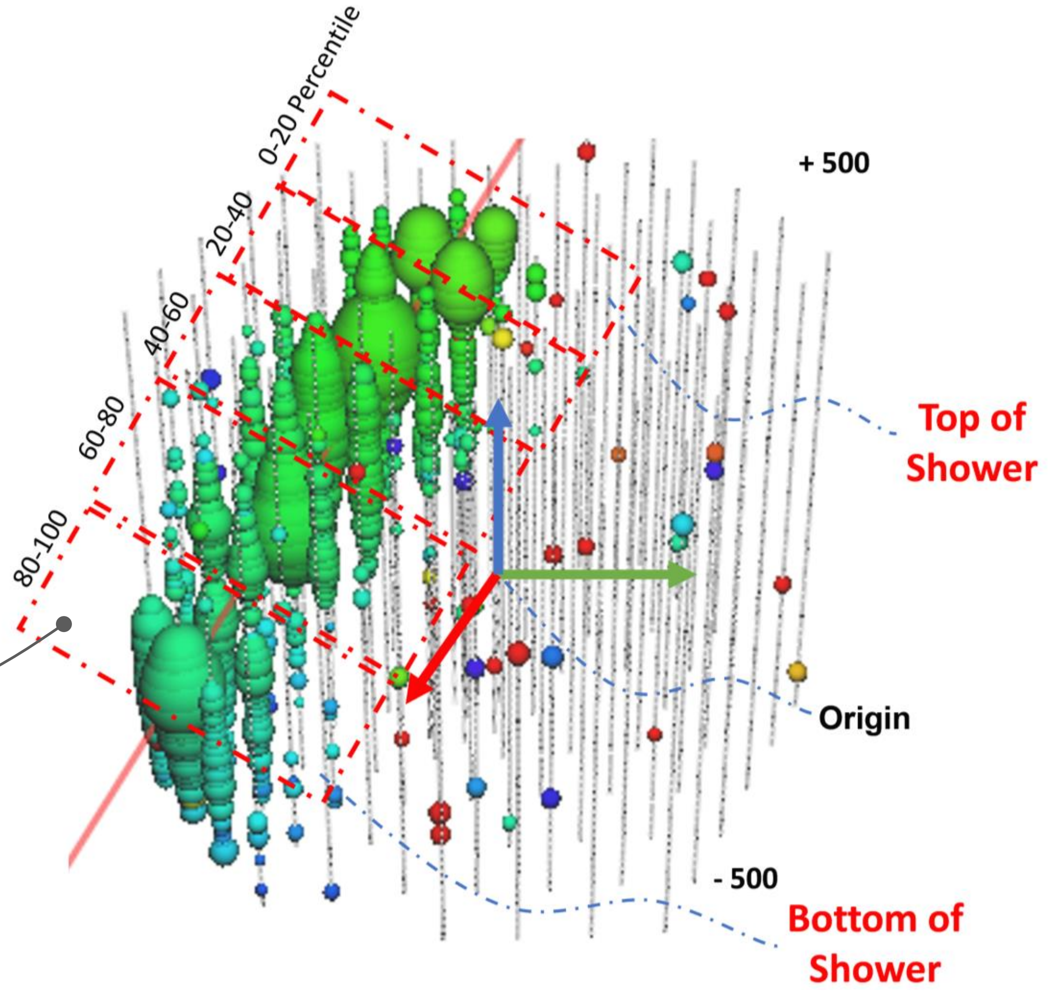
New Composition Parameters

Impact Weighted Charge $= \dfrac{\sum Charge_i * ri}{\sum r_i}$

- Help understand shower attenuation in-ice and a dynamic parameter
- Possibly help in understanding photon propagation in-ice
    - Ongoing work
- Good Separation Between Primaries

x-axis : Proxy for Energy ; y-axis : Impact Parameter



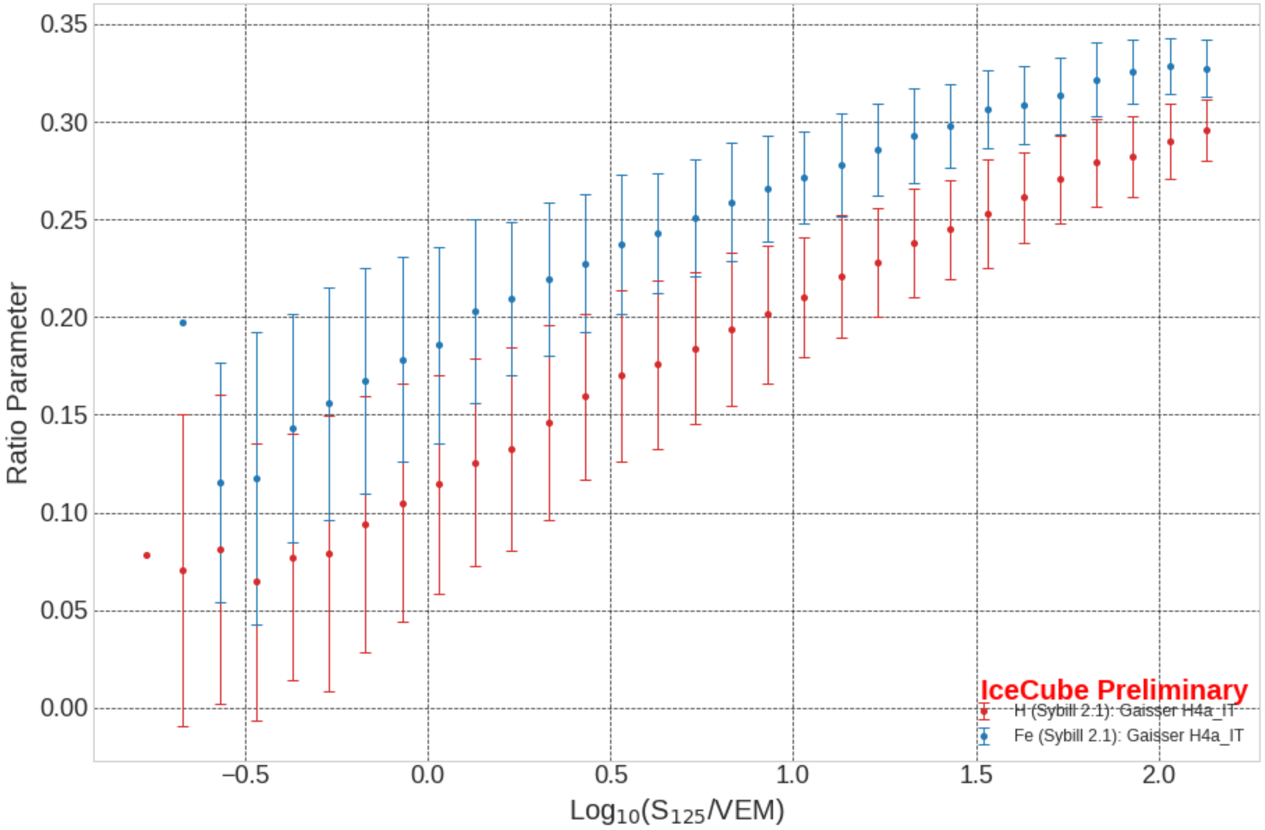


New Composition Parameters

Muons are the most promising candidates for cosmic-ray composition analysis

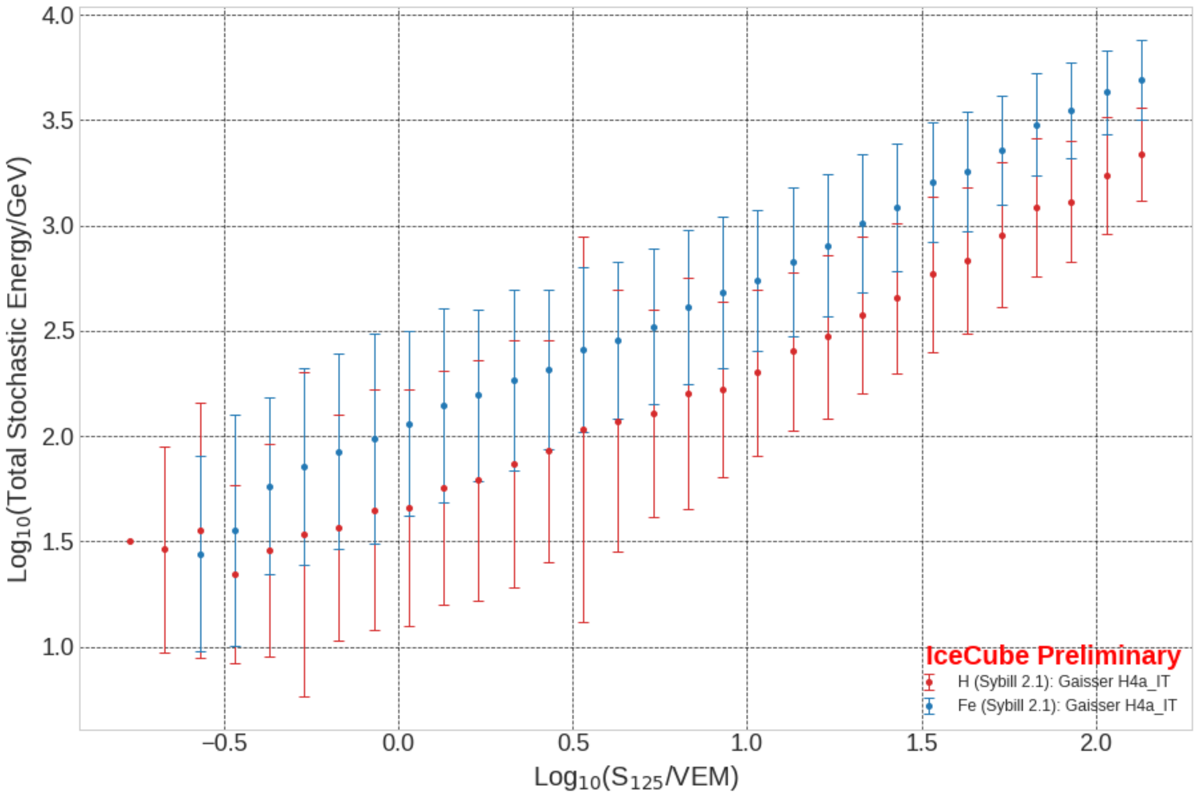
Ratio-Parameter

- Uses IceTop and In-Ice info.
- Proxy for muon-to-electron number ratio
  - Motivated by work of KASCADE-Grande (arXiv:1306.6283)



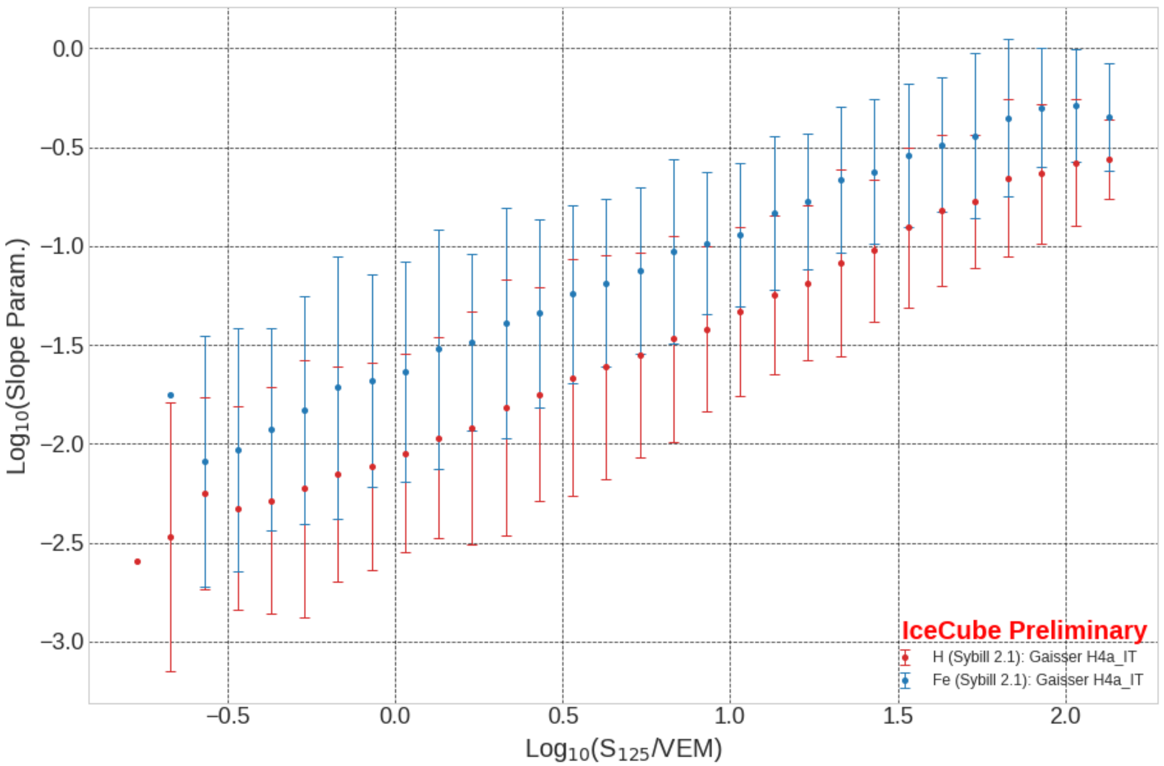
Total Stochastic Energy

- Uses In-Ice info.
- Captures local stochastic-deposits in-ice, primarily by muons.
- Good separation between primaries

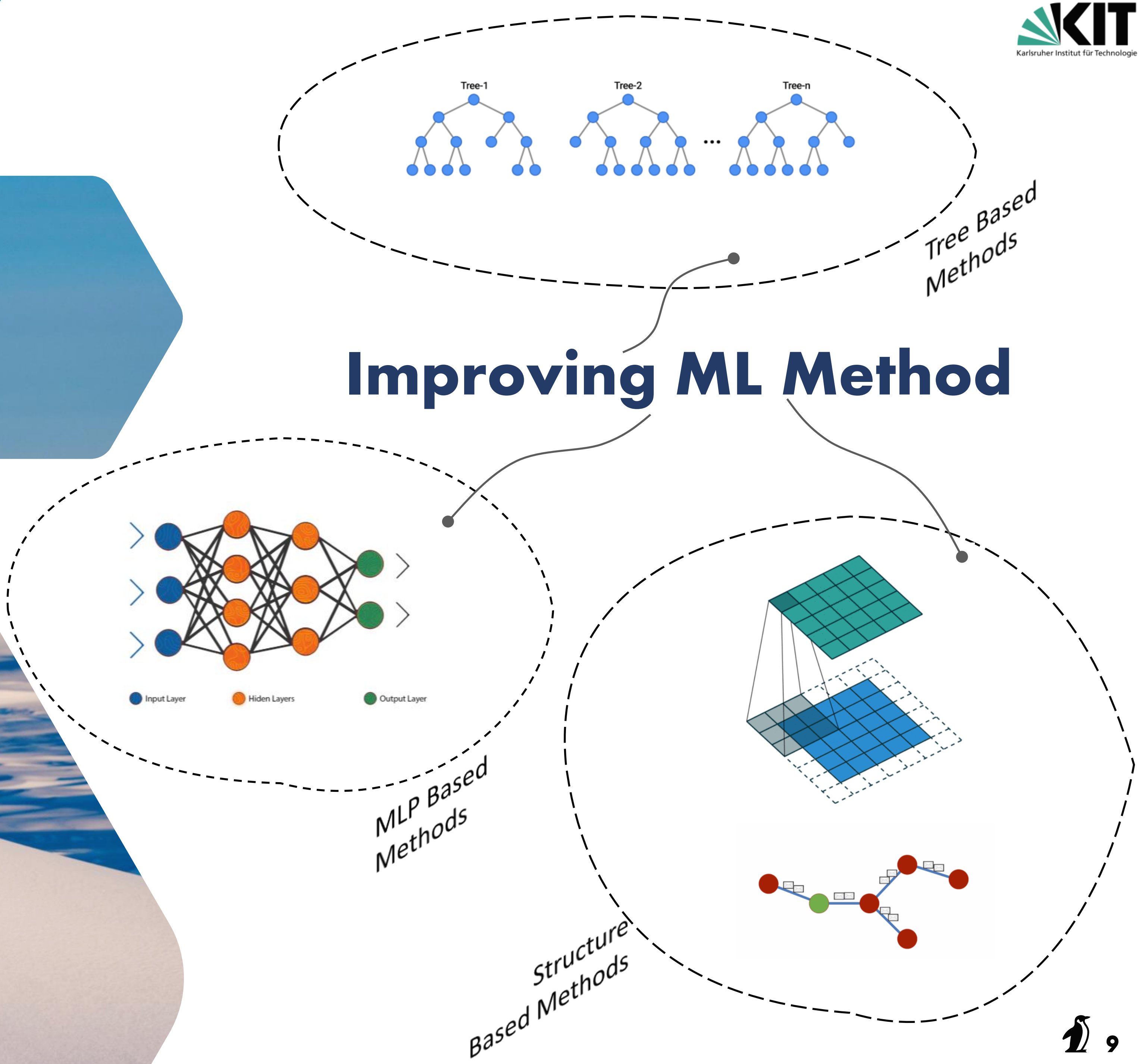


Slope-Parameter

- Uses In-Ice info.
- Captures the rate of in-ice charge deposit.
- Good separation between primaries



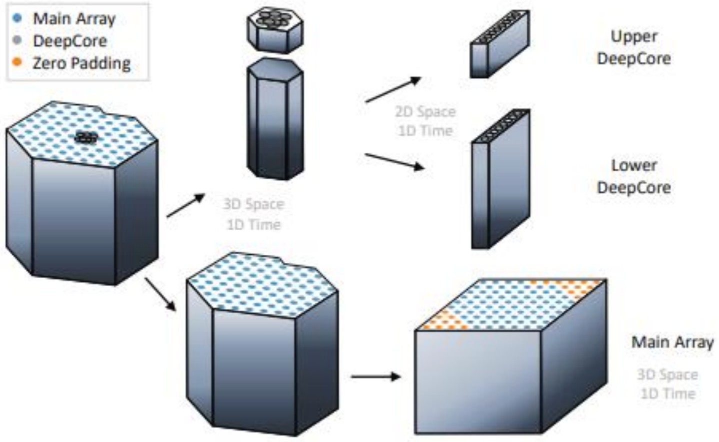
Improving ML Method

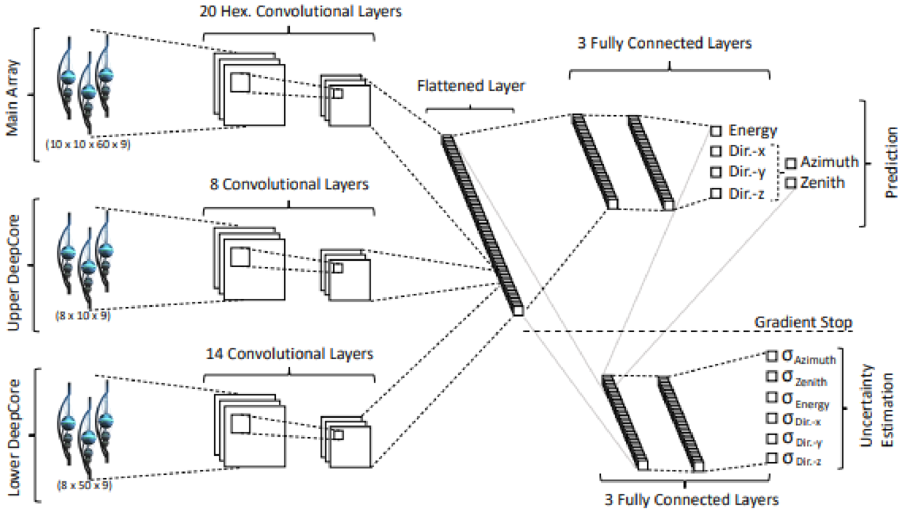


Using In-Ice Signal Footprint

Current Implementation at IceCube

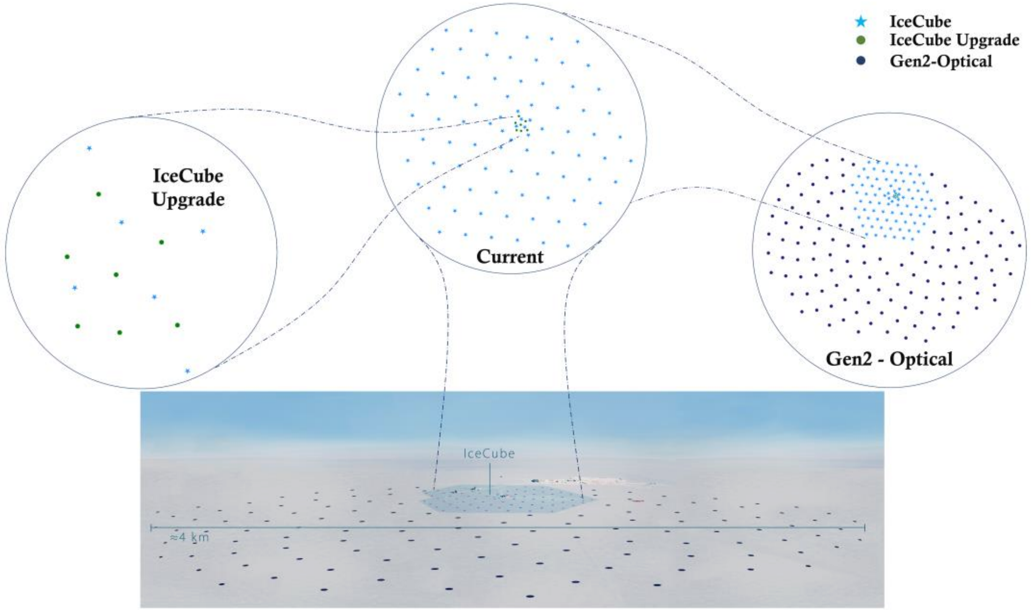
Credits: M. Huennefeld (arXiv:2101.11589)

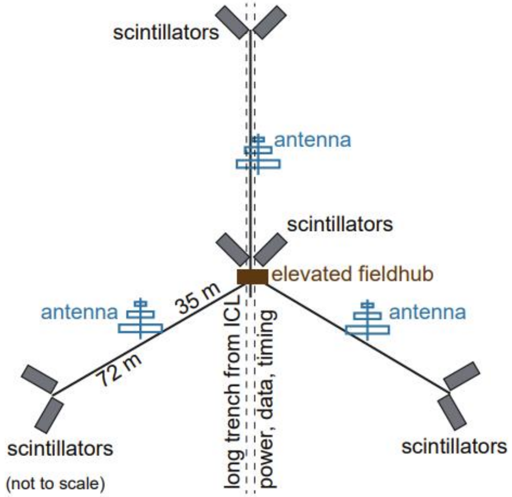


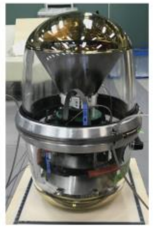


Moving Away from CNNs

Credits: IceCube-Gen2 (arXiv:2008.04323)







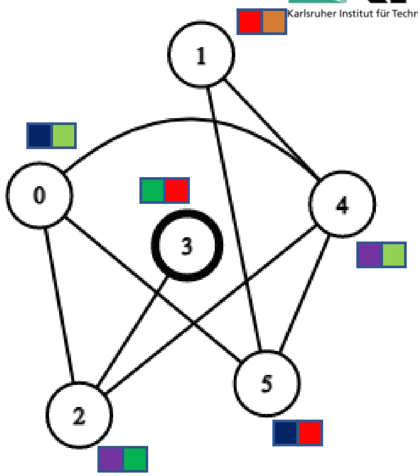
Learning on Graphs

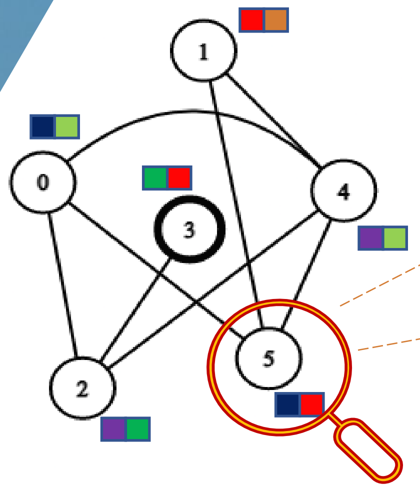
Defined by set of nodes (V) and set of edges (E) between the nodes

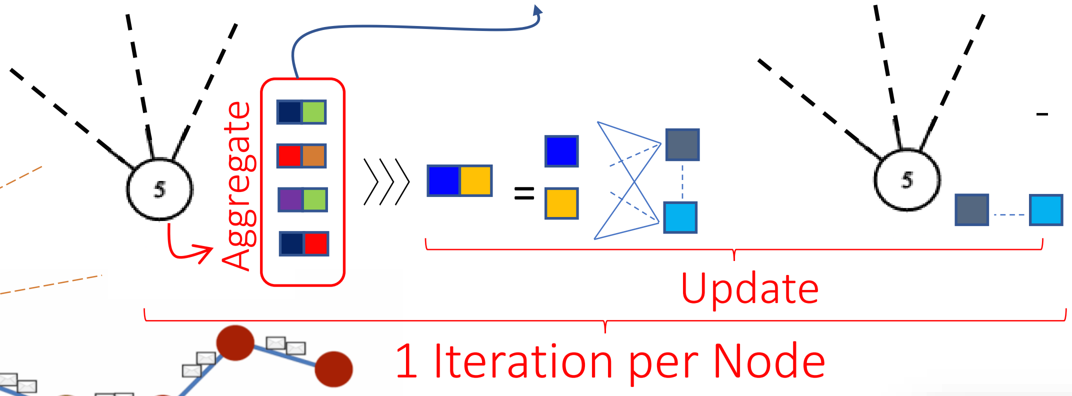
$$G = (V, E)$$

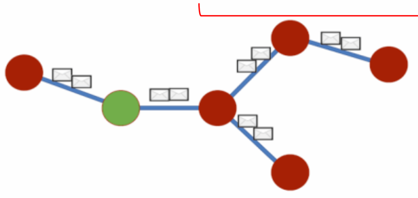
- Neighborhood and Connectivity & permutational invariance of Node Labelling

Undirected : Facebook Friends … ; Directed : Citation Graph … ; Bidirectional : Twitter Follows







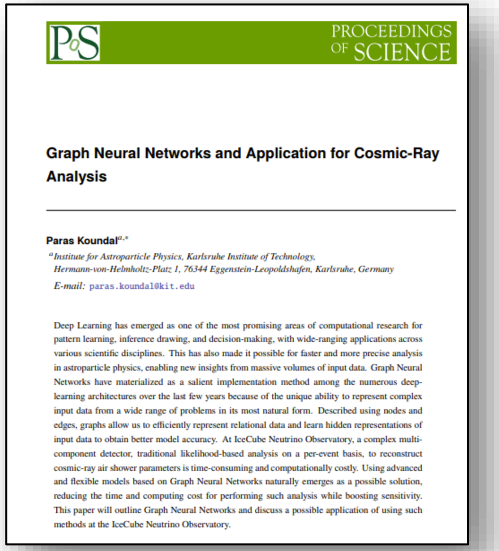


- Other Attributes
  - Node Features
  - Edge Features

Other Material

- paraskoundal.com/dlcp21
- My ML & AI Paper List
- My Twitter (Paras Koundal)

5th International Workshop on Deep Learning in Computational Physics, 2021



PROCEEDINGS OF SCIENCE

Graph Neural Networks and Application for Cosmic-Ray Analysis

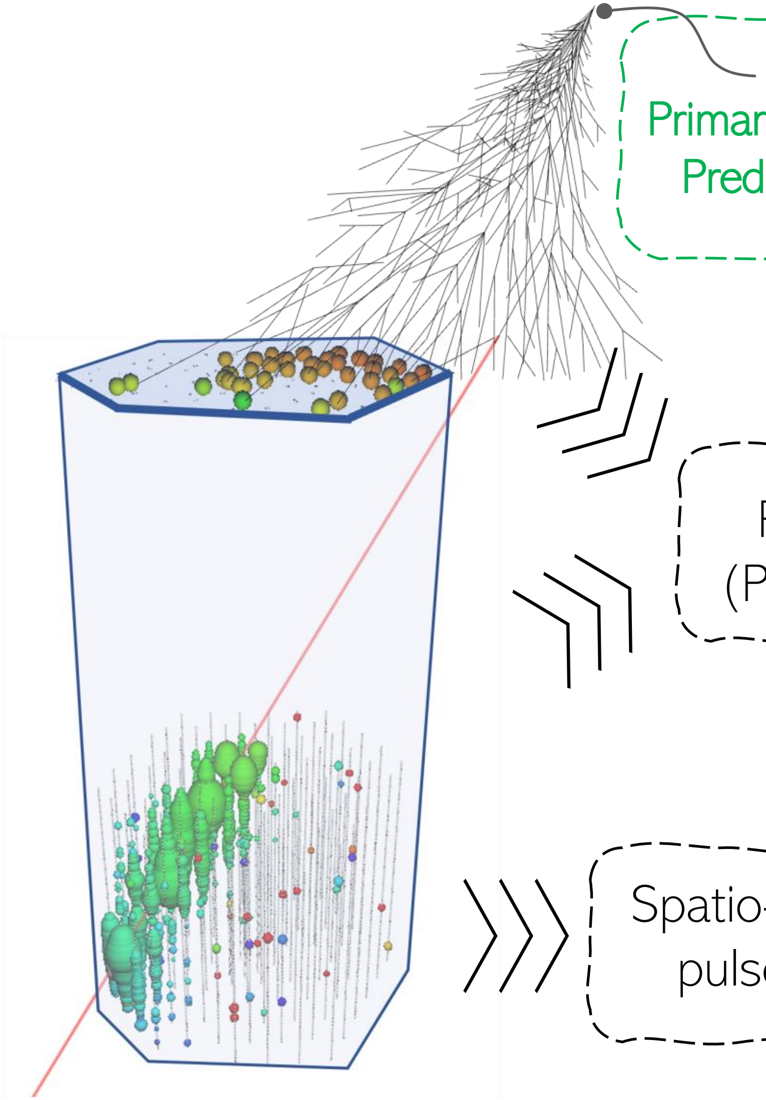
Paras Koundal[a,*]

[a] *Institute for Astroparticle Physics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Karlsruhe, Germany*

E-mail: paras.koundal@kit.edu

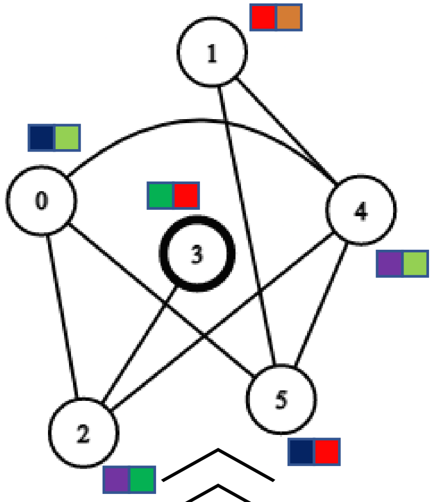
Deep Learning has emerged as one of the most promising areas of computational research for pattern learning, inference drawing, and decision-making, with wide-ranging applications across various scientific disciplines. This has also made it possible for faster and more precise analysis in astroparticle physics, enabling new insights from massive volumes of input data. Graph Neural Networks have materialized as a salient implementation method among the numerous deep-learning architectures over the last few years because of the unique ability to represent complex input data from a wide range of problems in its most natural form. Described using nodes and edges, graphs allow us to efficiently represent relational data and learn hidden representations of input data to obtain better model accuracy. At IceCube Neutrino Observatory, a complex multi-component detector, traditional likelihood-based analysis on a per-event basis, to reconstruct cosmic-ray air shower parameters is time-consuming and computationally costly. Using advanced and flexible models based on Graph Neural Networks naturally emerges as a possible solution, reducing the time and computing cost for performing such analysis while boosting sensitivity. This paper will outline Graph Neural Networks and discuss a possible application of using such methods at the IceCube Neutrino Observatory.

GNNs at IceCube (for CR Analysis)



Primary Mass Prediction

Input for Graph Neural Network



Pre-engineered Features (Primarily Mass-dependent)

Generate Edges between nodes; kNN/ Weighted/ Dynamic

Spatio-temporal Cleaning of pulses around the track

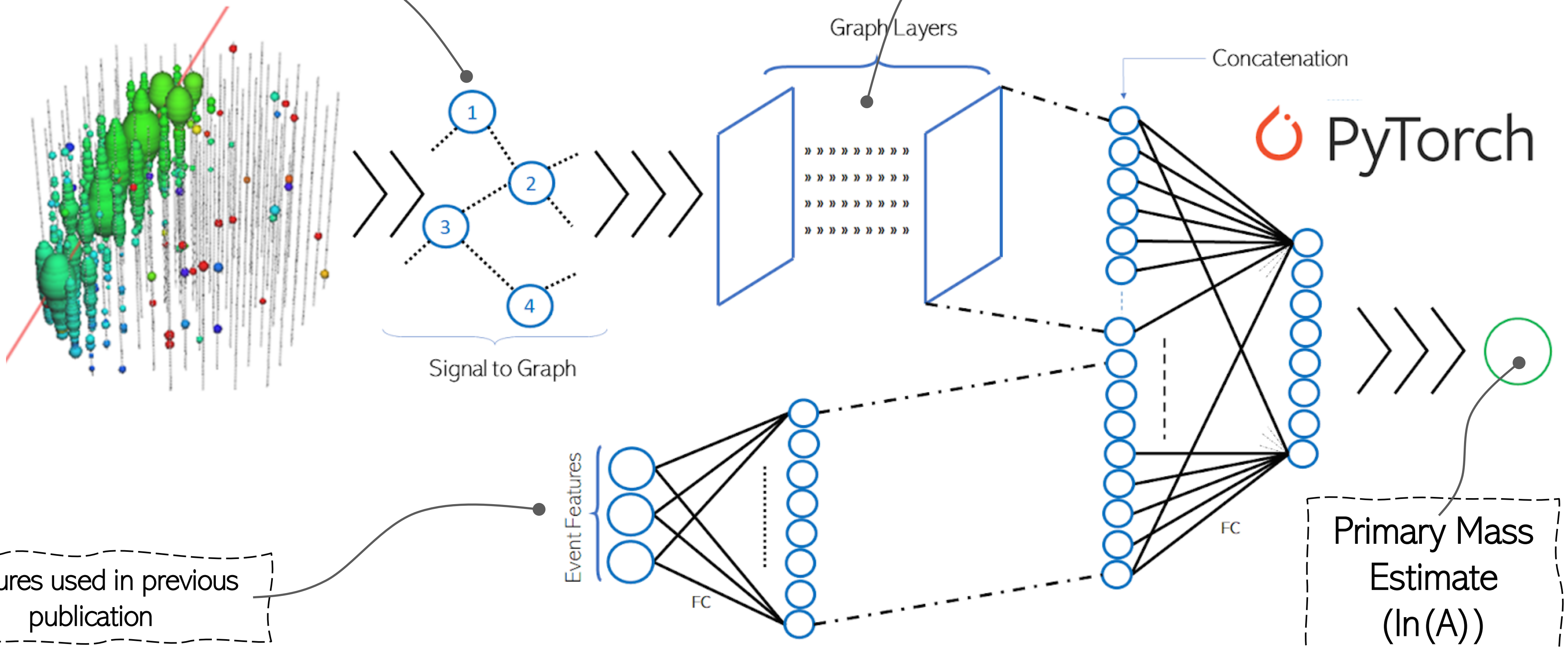
DOMs = Nodes
Node Features = Spatio-temporal measurements ; charge and others

Previous Results @ ICRC,2021

Nodes = Spatio-temporal and Charge Features
Edges = Each node connected to each other, weighted by Gaussian-kernel with learnable width

Simple Graph Aggregation:
Weighted Mean of neighbors

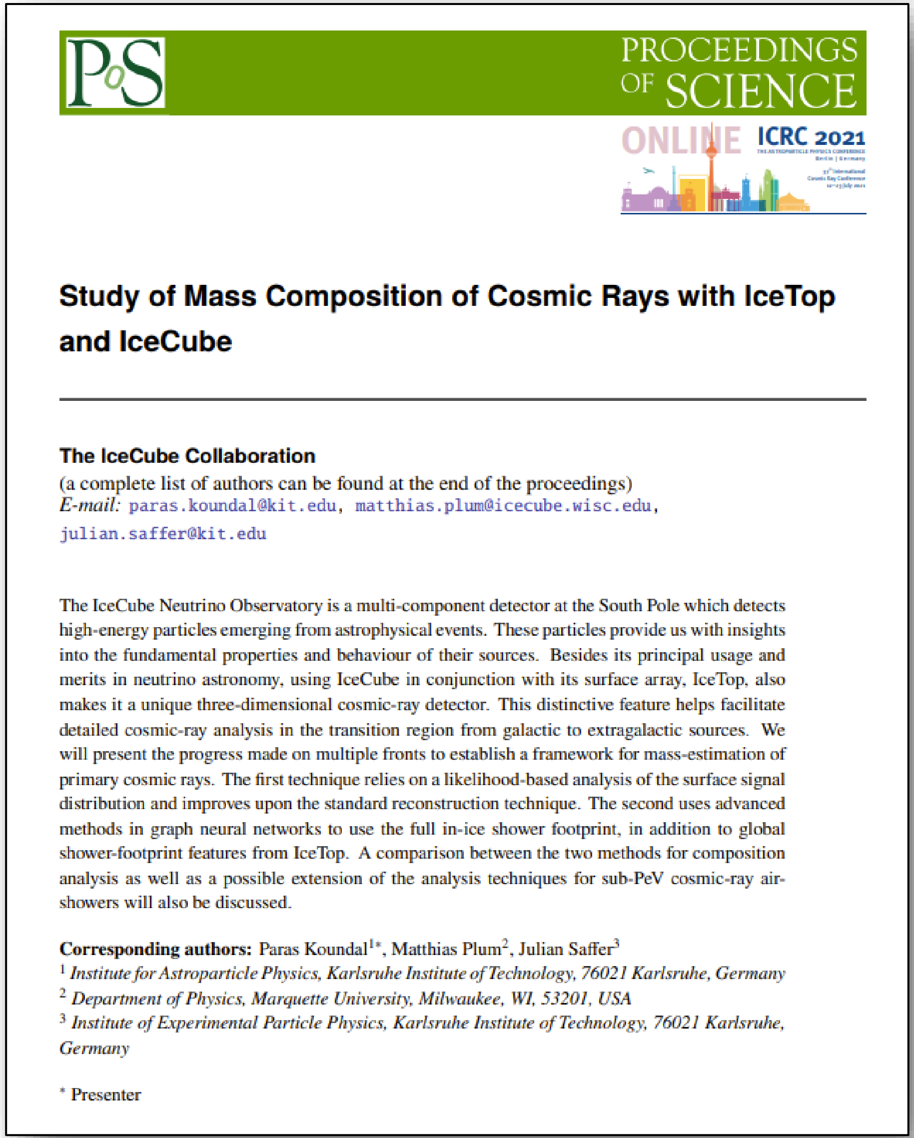
- Ignored True-Composition Spectrum:
  - Simulations follow approximately E^{-1} power law. However, real spectrum decays much faster. Moreover, we already have some prior knowledge about variation of composition with energy.
  - Equal number of elements were taken (for work in next slides too) - Focus is to improve per-event based composition estimate.



Features used in previous publication

Primary Mass Estimate (ln(A))

Previous Results @ ICRC,2021



PROCEEDINGS OF SCIENCE

ONLINE ICRC 2021

Study of Mass Composition of Cosmic Rays with IceTop and IceCube

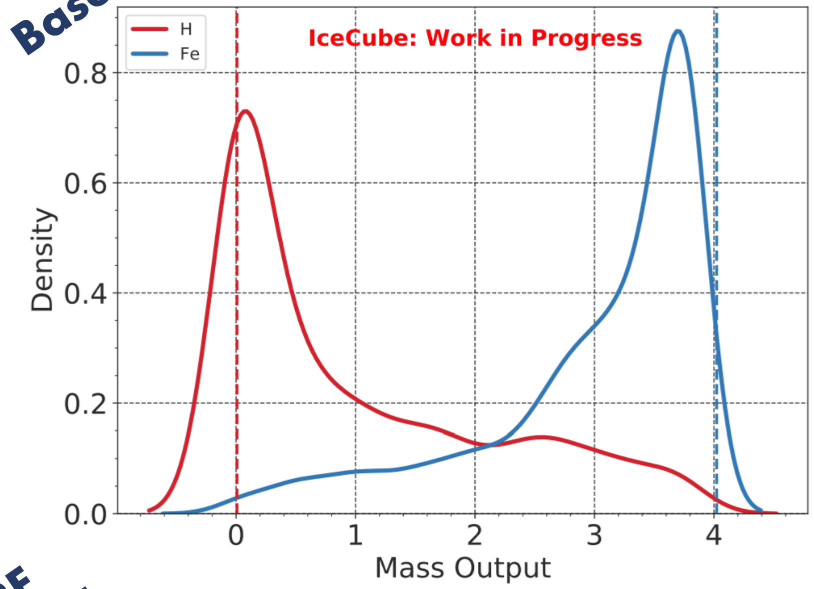
The IceCube Collaboration
(a complete list of authors can be found at the end of the proceedings)
E-mail: paras.koundal@kit.edu, matthias.plum@icecube.wisc.edu, julian.saffer@kit.edu

The IceCube Neutrino Observatory is a multi-component detector at the South Pole which detects high-energy particles emerging from astrophysical events. These particles provide us with insights into the fundamental properties and behaviour of their sources. Besides its principal usage and merits in neutrino astronomy, using IceCube in conjunction with its surface array, IceTop, also makes it a unique three-dimensional cosmic-ray detector. This distinctive feature helps facilitate detailed cosmic-ray analysis in the transition region from galactic to extragalactic sources. We will present the progress made on multiple fronts to establish a framework for mass-estimation of primary cosmic rays. The first technique relies on a likelihood-based analysis of the surface signal distribution and improves upon the standard reconstruction technique. The second uses advanced methods in graph neural networks to use the full in-ice shower footprint, in addition to global shower-footprint features from IceTop. A comparison between the two methods for composition analysis as well as a possible extension of the analysis techniques for sub-PeV cosmic-ray air-showers will also be discussed.

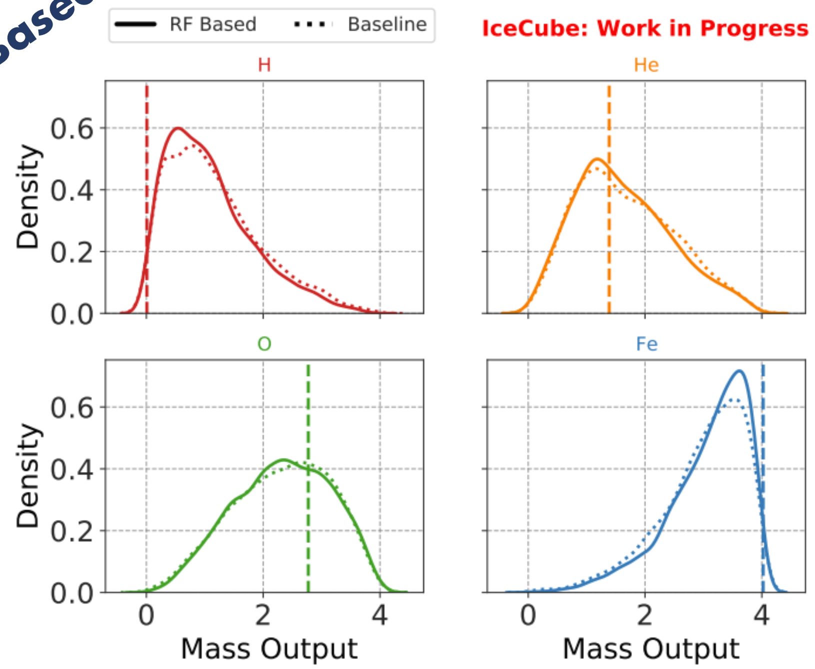
Corresponding authors: Paras Koundal[1*], Matthias Plum[2], Julian Saffer[3]
[1] *Institute for Astroparticle Physics, Karlsruhe Institute of Technology, 76021 Karlsruhe, Germany*
[2] *Department of Physics, Marquette University, Milwaukee, WI, 53201, USA*
[3] *Institute of Experimental Particle Physics, Karlsruhe Institute of Technology, 76021 Karlsruhe, Germany*

* Presenter

GNN Based



RF Based



Issues Resolved with Previous Work

- **Minor Issues Resolved**
  - Adding More Mass-Dependent Features
  - Correcting Feature normalizations

- **Inherently introduced simplification for training – Using PyTorch**
  - Batch Size (subset of training data) = 1 graph
    - For Better Generalization: BS (not too big , not too small)

- **Graph architecture was over-simplified**
  - Number of nodes were made equal
    - **Now:** Flexibility of choice.
  - Message aggregate and update in graph was over-simplistic → Generalized to implement any architecture
    - **Now:** k-nn connection (can choose between based on feature-vector or spatial coordinates)
  - For global-aggregation: Flattened-mean was used
    - **Now:** Implement any SOTA global-aggregation

Major Update

- Global Features also included in graph message-passing framework
  - Global attributes are same features shared over all nodes

- **Small Dataset**
  - CR-MC Simulations are limited.
    - Simulations are costly.

  - **Working Solution for increasing dataset size:** Randomly drop (.1% % & .2%) in-ice DOMs to increase data 5-fold
    - **Speculation -** Our in-ice track and energy reconstructions are resilient to these changes : Will be tested in future
    - **Food for Thought –** Change introducing hadronic-interaction model dependence.

Major Update

- New Graph Message-Passing framework
  - Adapted from the work "**Hierarchical Multi-View Graph Pooling with Structure Learning**"
    - Improvement in Graph pooling (or downsampling) to learn hierarchical representations
    - Attention mechanism utilized to generate robust node ranks.
    - Preserve the underlying graph topological information, using a structure learning mechanism.

Date of Publication: 18 June 2021

Hierarchical Multi-View Graph Pooling with Structure Learning

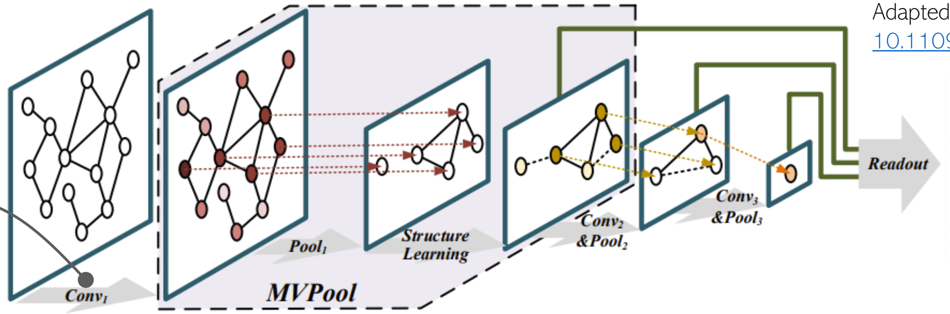
Zhen Zhang, Jiajun Bu*, Member, IEEE, Martin Ester, Senior Member, IEEE, Jianfeng Zhang, Zhao Li*, Chengwei Yao, Huifen Dai, Zhi Yu, Can Wang, Member, IEEE

Date of Publication: 22 Feb 2017

SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS

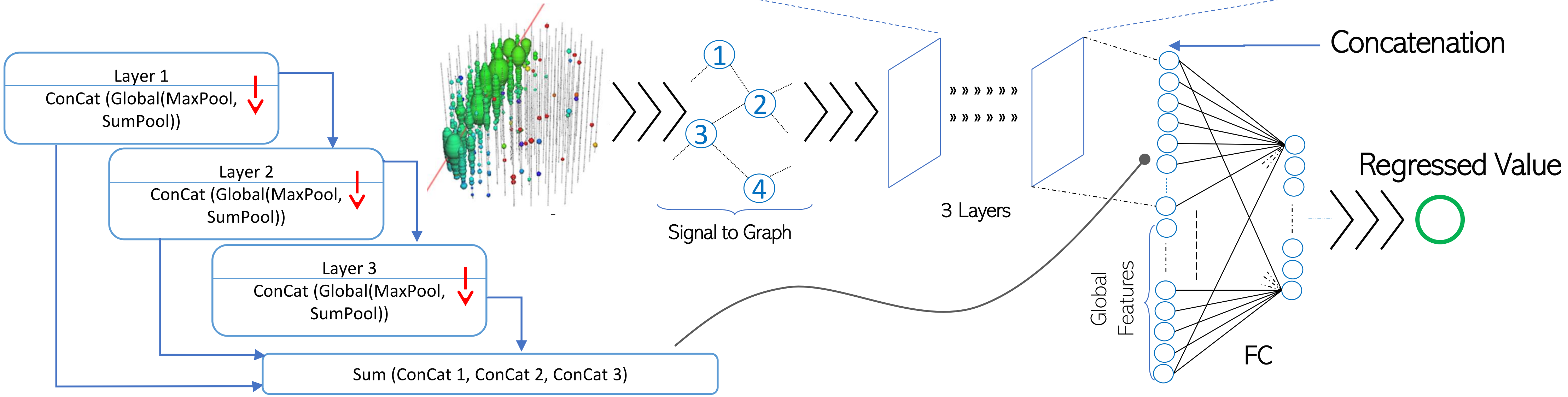
Thomas N. Kipf
University of Amsterdam
T.N.Kipf@uva.nl

Max Welling
University of Amsterdam
Canadian Institute for Advanced Research (CIFAR)
M.Welling@uva.nl



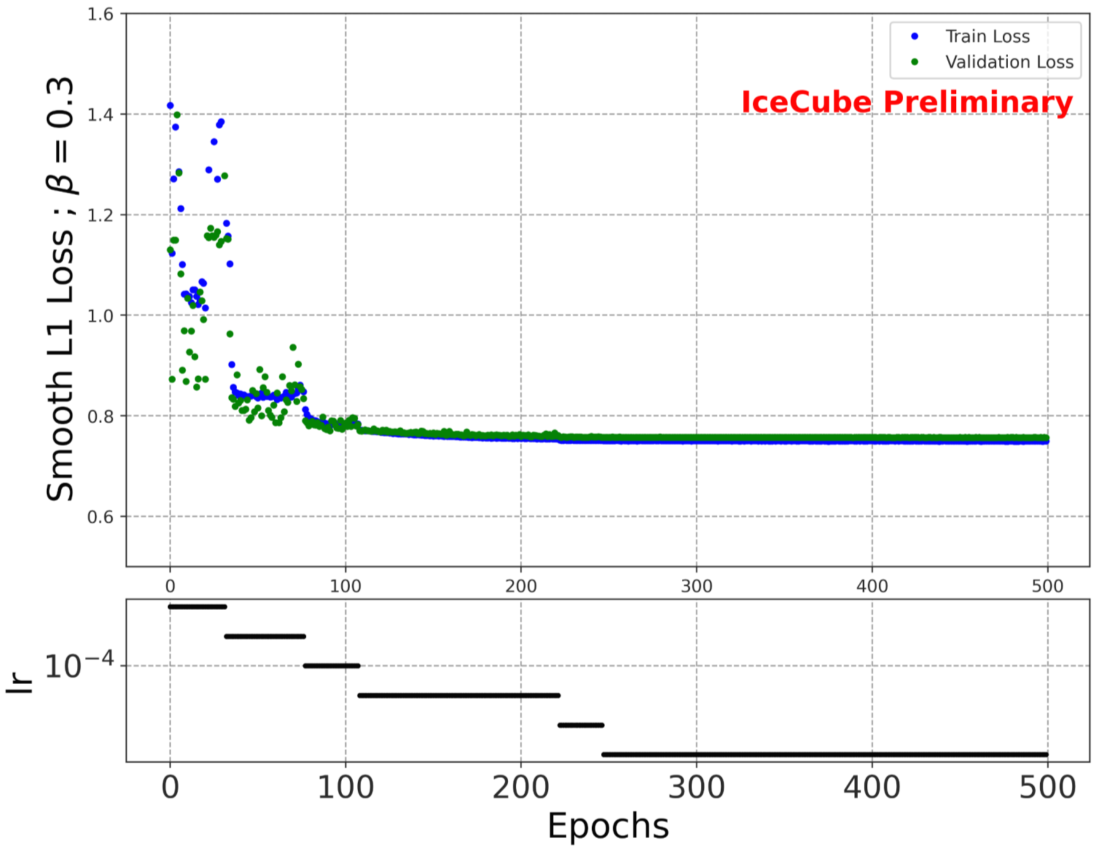
Adapted from DOI: 10.1109/TKDE.2021.3090664

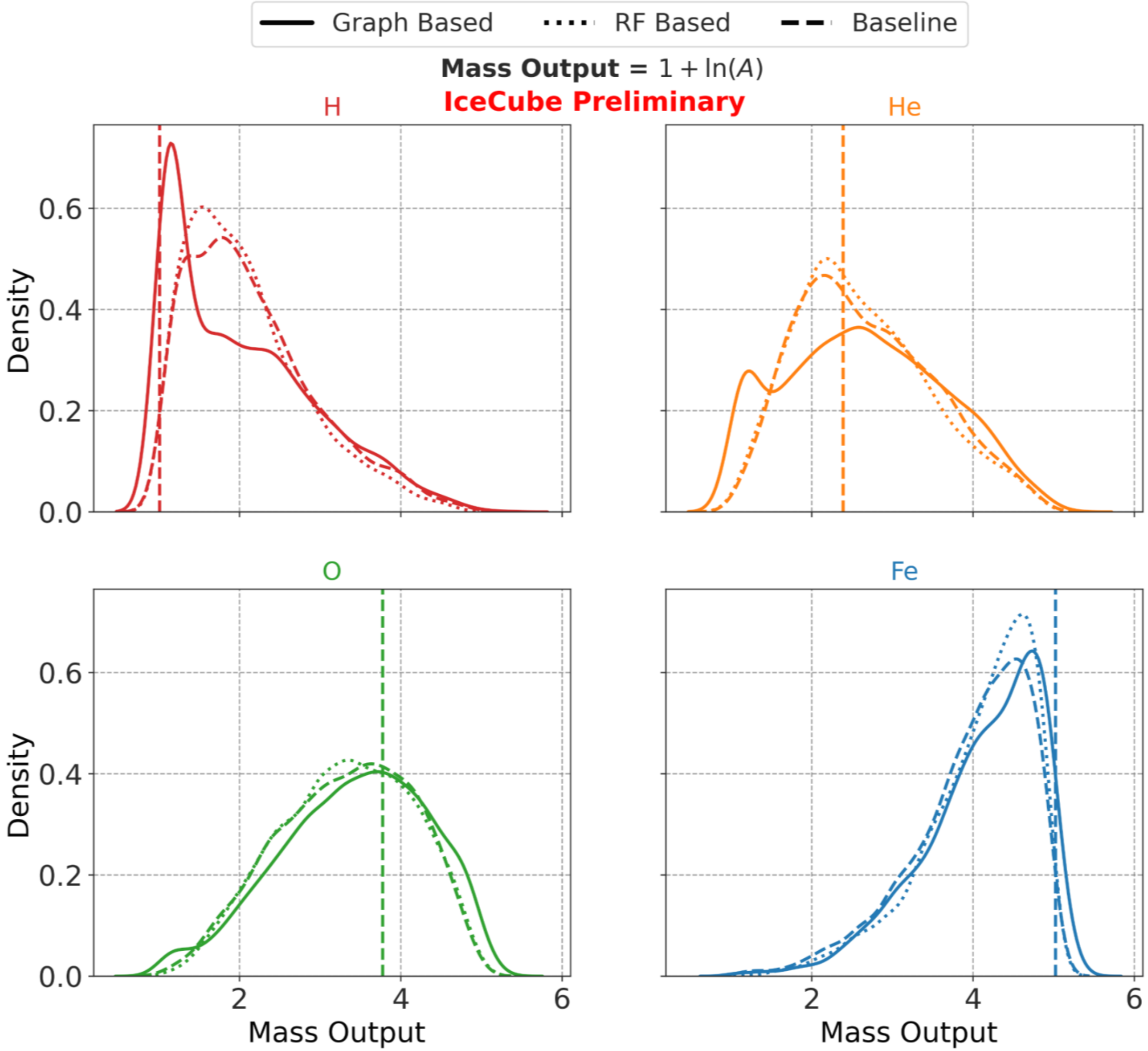
*3



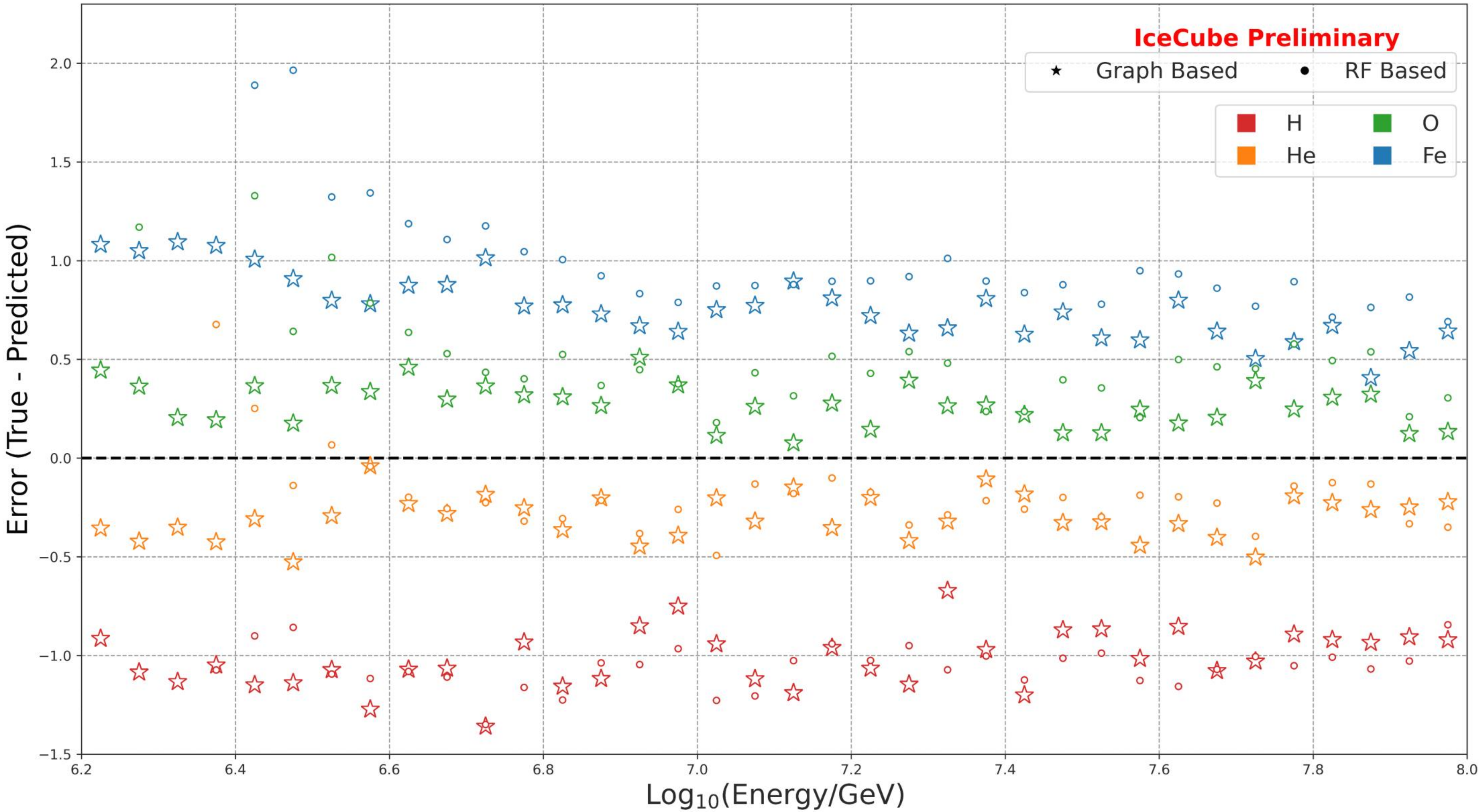
Results

- Target Variable: 1+ln(A)
- Major Improvements for all Primary Types
  - Maximum at True value
  - Shift towards lighter elements for H and He
  - Shift towards heavier for O and Fe
- Loss:
  - Adaptive Learning rate
  - Very Gradual decrease in error
  - No-overfitting most of the times



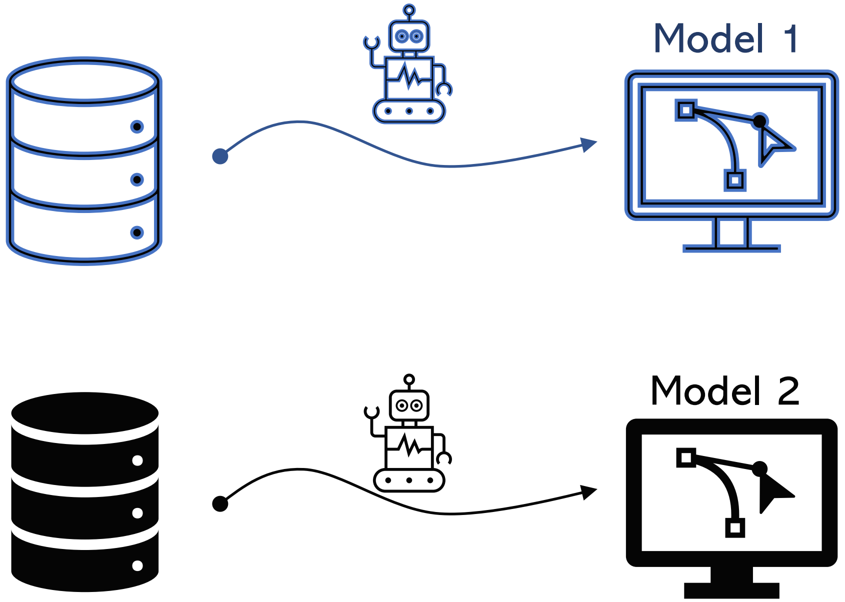


Prediction Error : Binned over Energy

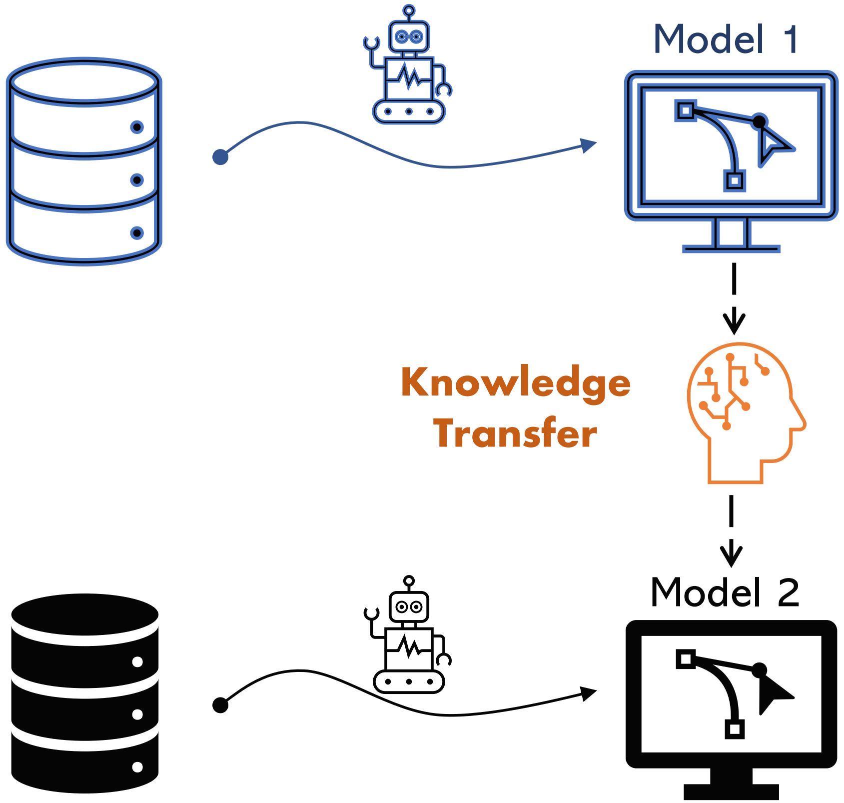


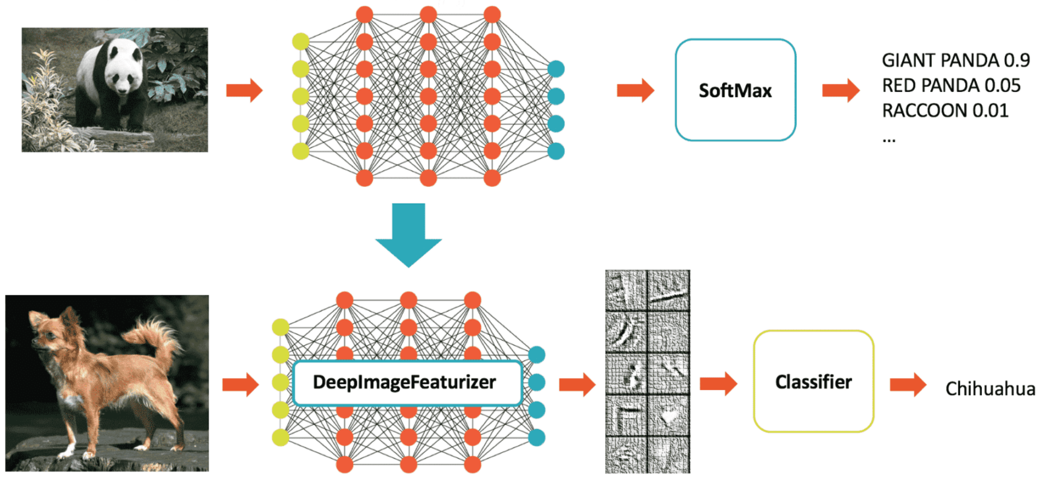
Transfer Your Learning

Traditional Machine Learning



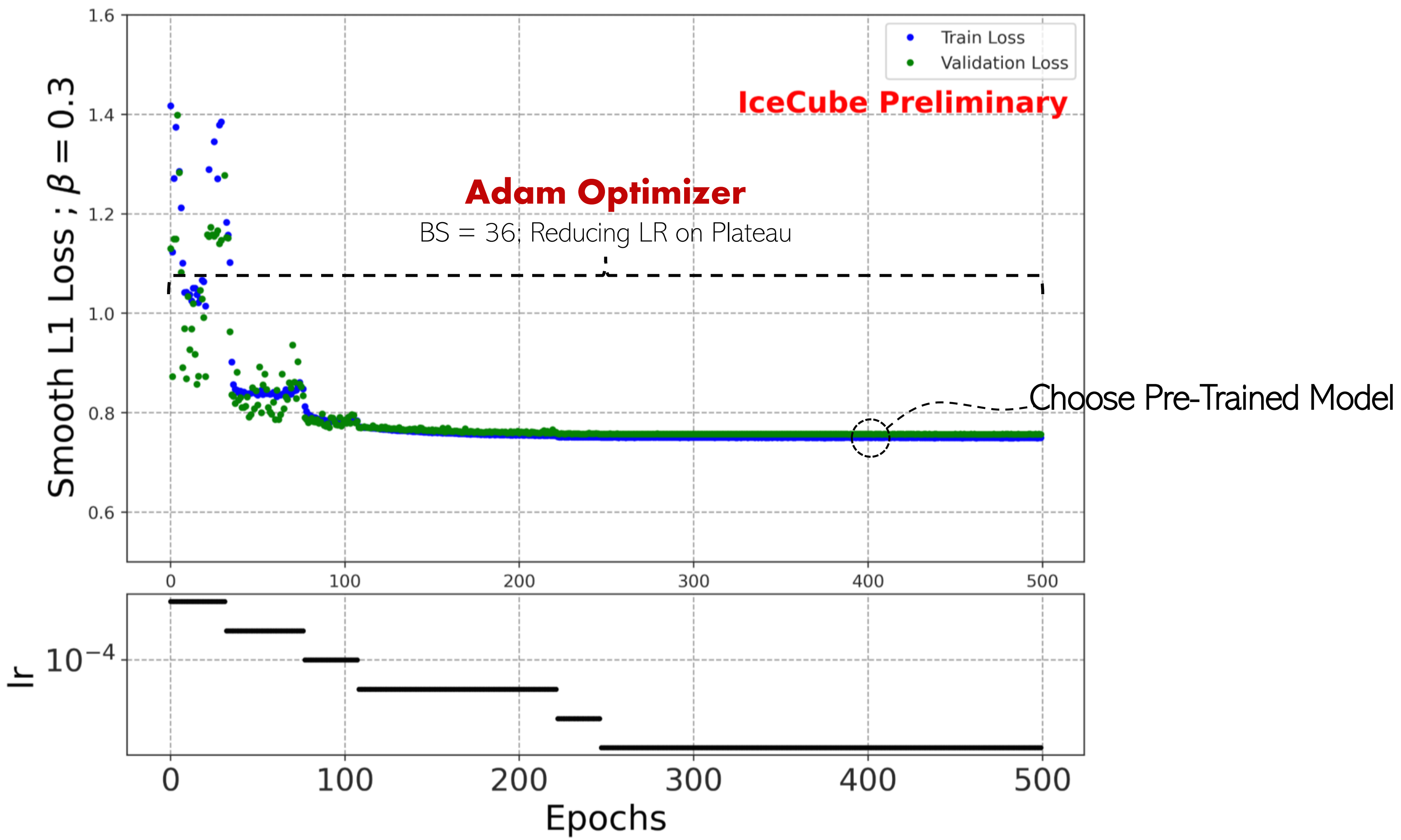
Transfer Machine Learning





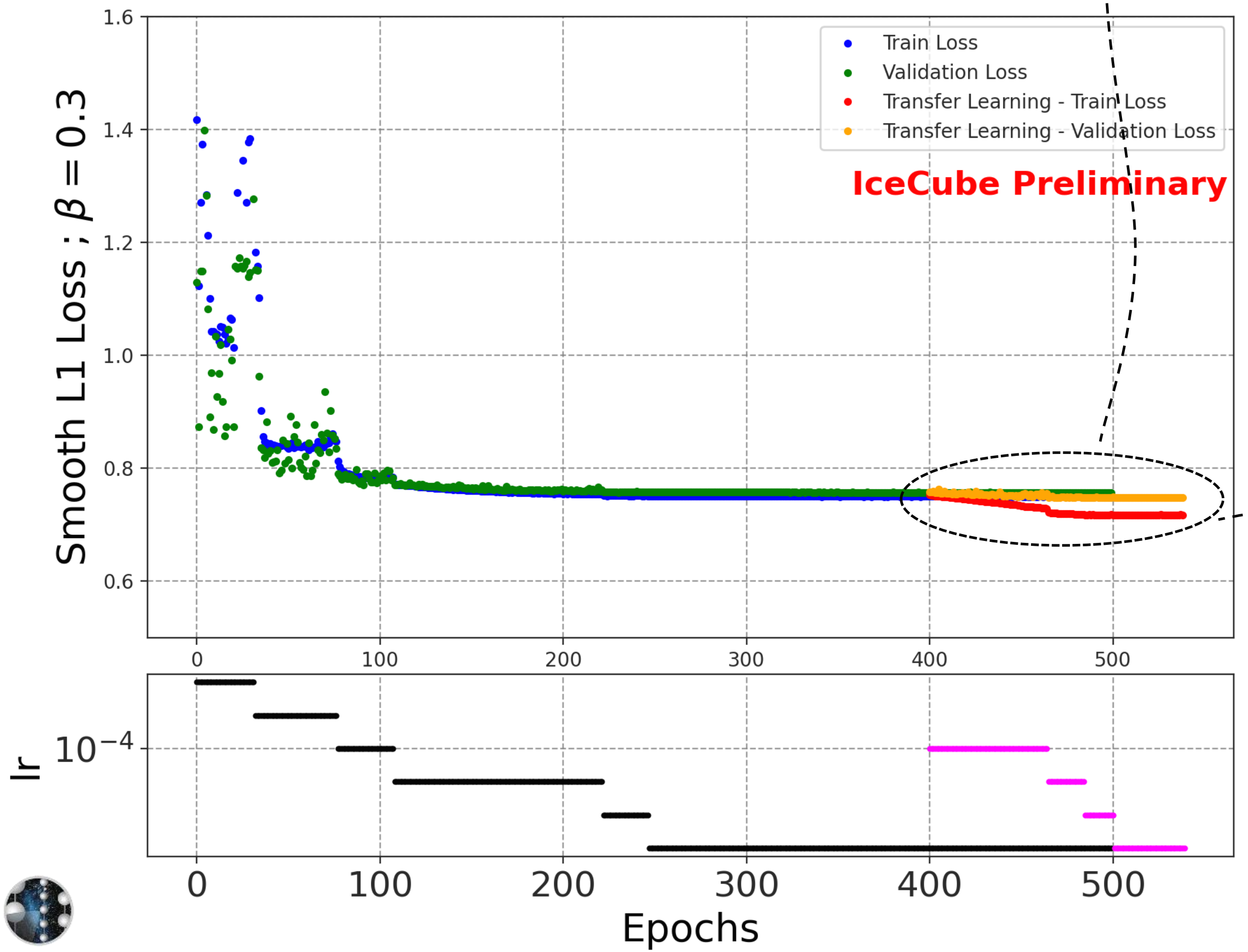
Bottom Image: Borrowed from kdnuggets.com

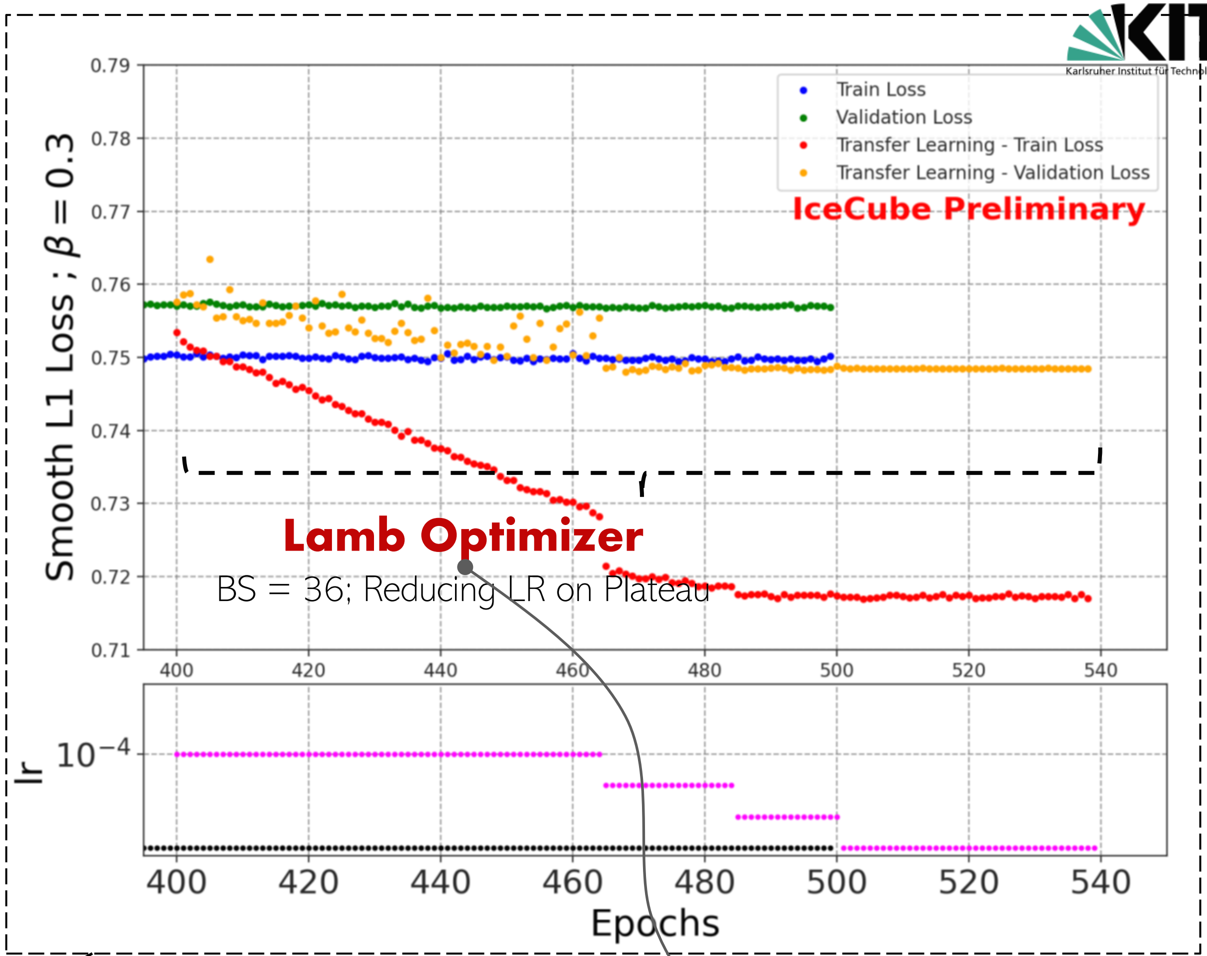
Transfer Learning @ IceCube



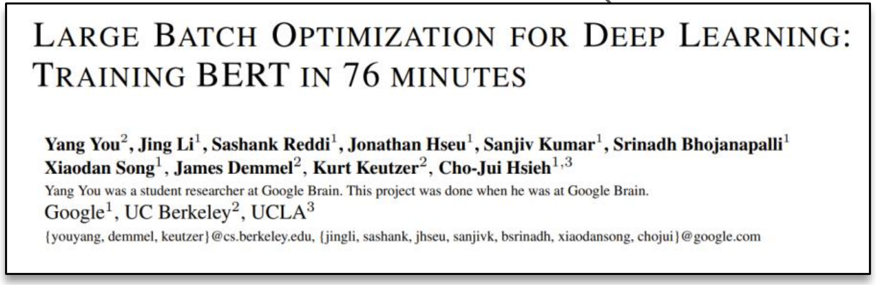
Transfer Learning @ IceCube

- The validation loss of the finetuned transferred-learned model is already lower than the training loss of the previous model. Hopefully, improvement if future.
  - Caution: Do, I see signs of overfitting





Date of Publication: 3 Jan 2020

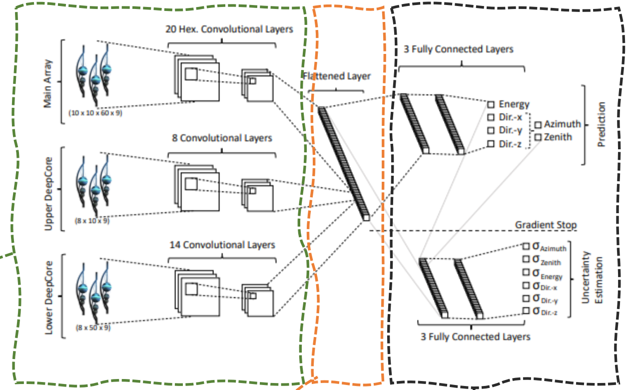


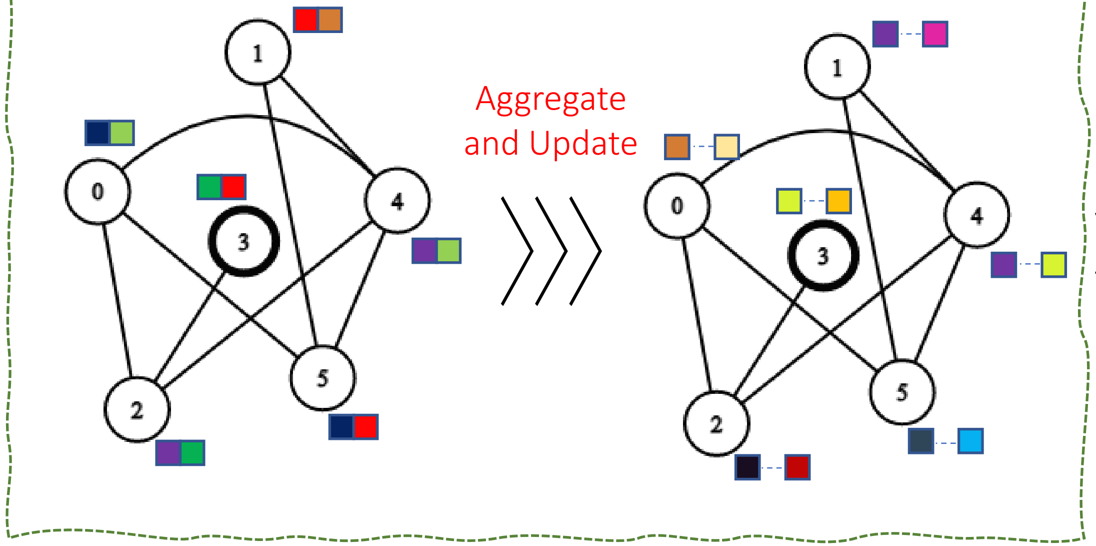
LARGE BATCH OPTIMIZATION FOR DEEP LEARNING: TRAINING BERT IN 76 MINUTES

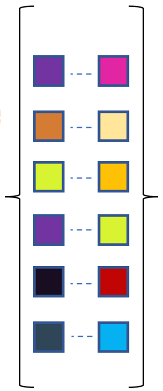
Yang You[2], Jing Li[1], Sashank Reddi[1], Jonathan Hseu[1], Sanjiv Kumar[1], Srinadh Bhojanapalli[1]
Xiaodan Song[1], James Demmel[2], Kurt Keutzer[2], Cho-Jui Hsieh[1,3]
Yang You was a student researcher at Google Brain. This project was done when he was at Google Brain.
Google[1], UC Berkeley[2], UCLA[3]
{youyang, demmel, keutzer}@cs.berkeley.edu, {jingli, sashank, jhseu, sanjivk, bsrinadh, xiaodansong, chojui}@google.com

Graph Level Prediction

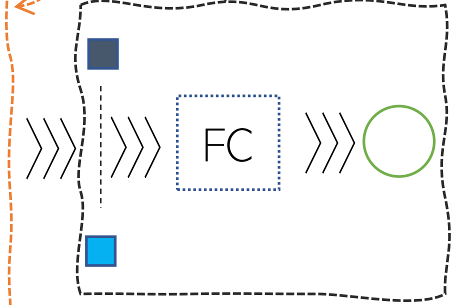
GNNs adapt the learning structure of MLPs as well as CNNs. The aggregate and update step of GNNs is similar to convolution layers of CNN. To make a graph-level prediction, we need to find a permutation-invariant graph-level aggregation method.







- Global Max(/Mean) Pooling
- Global Sum Pooling
- Global Sort Pooling
- Attention Based Pooling



Issues & Status of Current GNNs

- **Oversmoothing:** CNNs excel by their ability to use deeper architectures to improve accuracy. However, GNNs face with loss(or accuracy) saturation once the number of layers increase.
- Currently, the number of standard-datasets and correspondingly methods are limited in GNN.
- In most of the standard datasets, node number variation is small (not true for this analysis).

- To capture structural info, for smooth-graphs with big node number variation, clearly global max/mean pooling will not be very useful. Same for sort-pooling.
- Sum-pooling is size-dependent. However, not well suited for inverse problems.
- Attention Based Methods: Newer methods. Not very much explored.

- The input-from GNN message-passing can't be huge: Curse of dimensionality.

Conclusion and Outlook

- New cosmic-ray composition analysis seem promising.
- Overall Improvement in Cosmic-Ray Composition using Graph Neural Network
  - Improvement over full-energy range
- Future: PointNet++, Removing learning from graphs (arXiv:1905.04579)