A Review on Machine Learning in Neutrino Experiments

Karl Warburton – Iowa State University

Quantil Seminar

12th November 2020

https://arxiv.org/abs/2008.01242

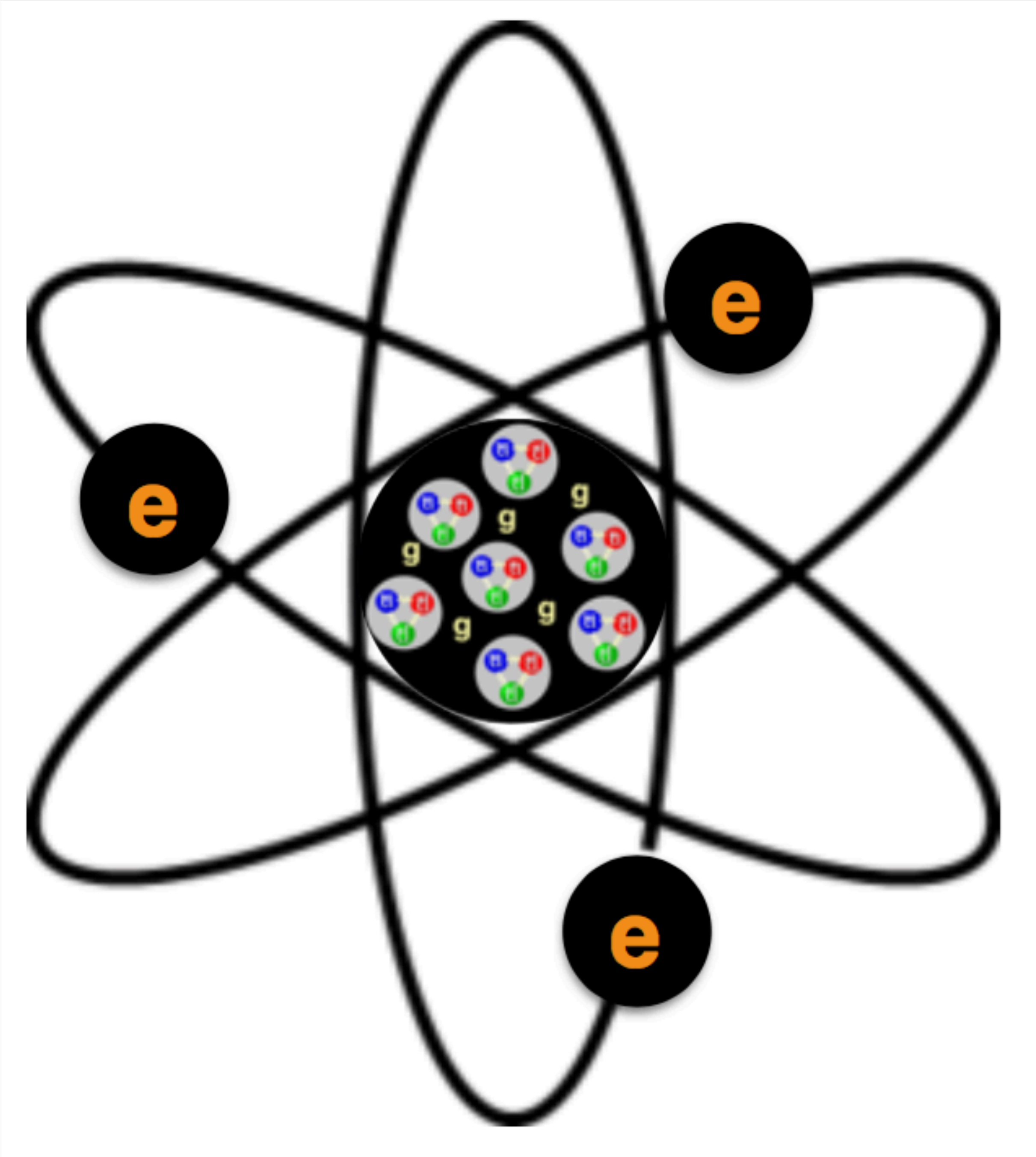
Who am I?

- From Stoke-on-Trent in the UK.
- I use Machine Learning to improve studies I perform looking for neutrino oscillations, a key driver of particle physics research.
- Did my undergraduate and PhD at the University of Sheffield, UK.
  - Worked on simulations and reconstruction techniques in a prototype detector for the next generation of neutrino experiments (DUNE).
- Currently working at Iowa State University, USA and based at Fermilab, in Illinois, USA.
  - Working on supernova triggering in DUNE. Will speak about how Machine Learning may be a better method of doing this later on in the seminar.
  - Lead the Reconstruction and Deep Learning on NOvA, which had the first physics result using a Machine Learning algorithm.

A Brief Outline

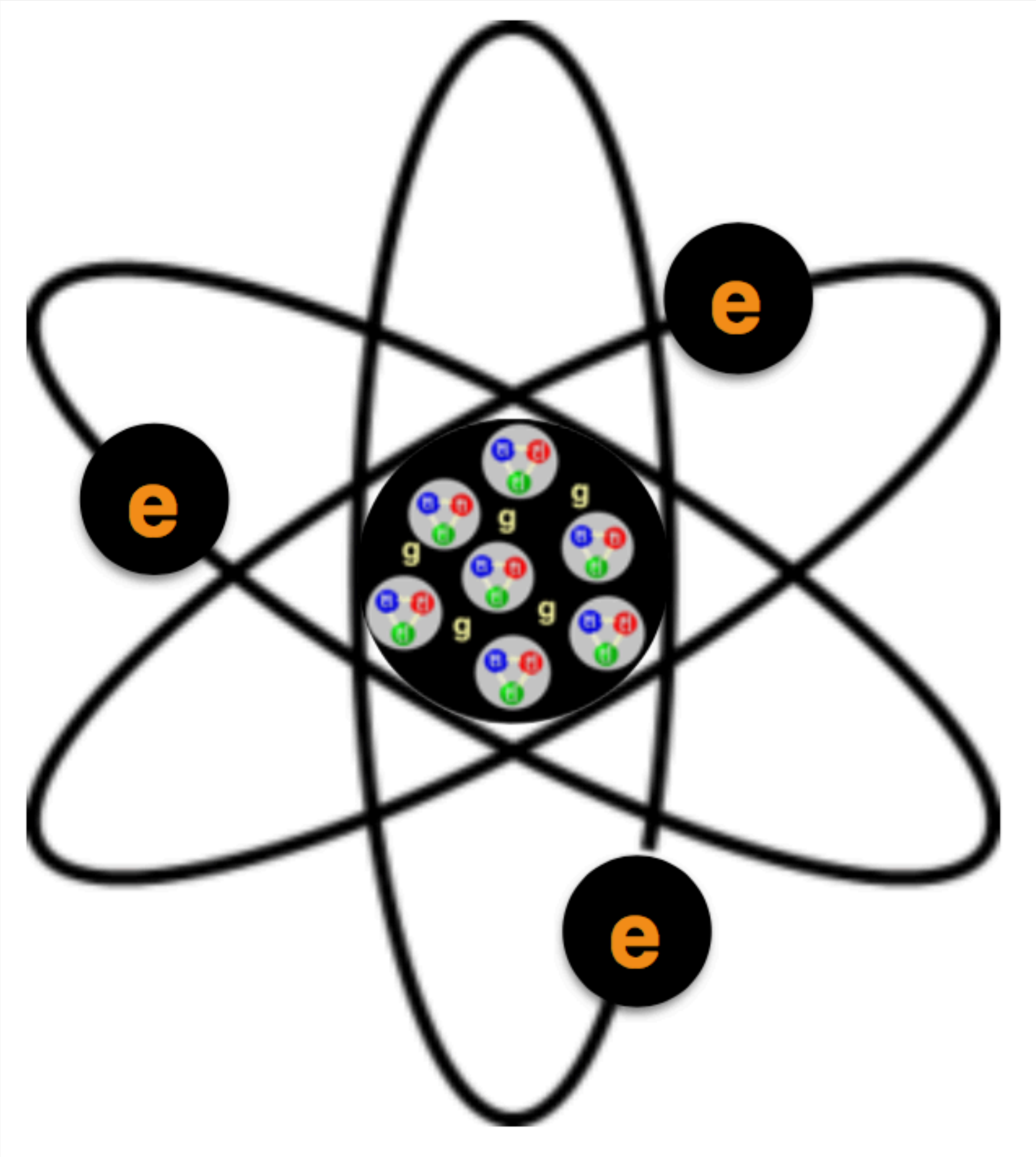
- What are neutrinos? What do we try to do in Neutrino Experiments?
- A brief introduction to Machine Learning.
- The challenges of applying Machine Learning in neutrino experiments.
- The future opportunities presented by Machine Learning techniques.

A crash course in Particle Physics

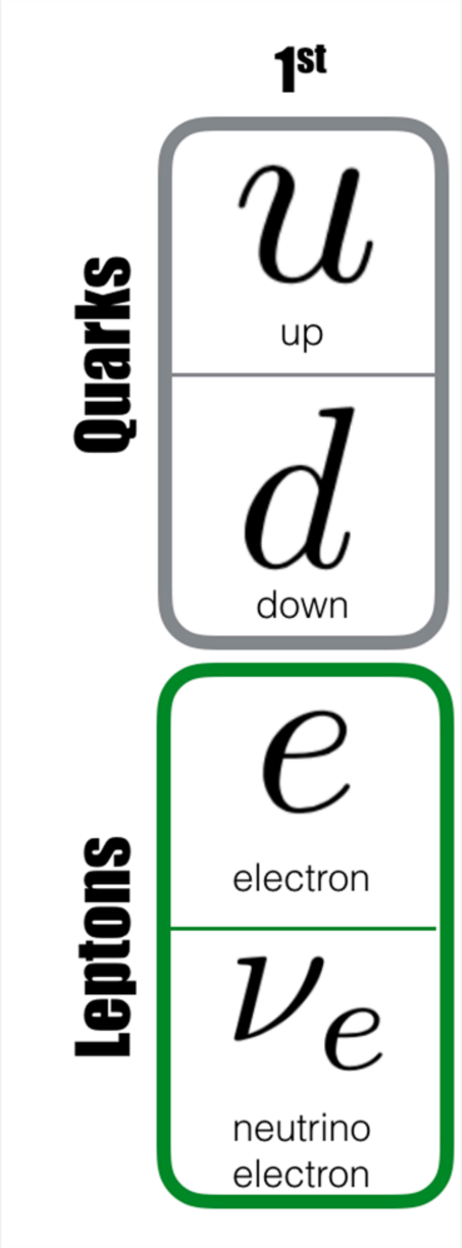


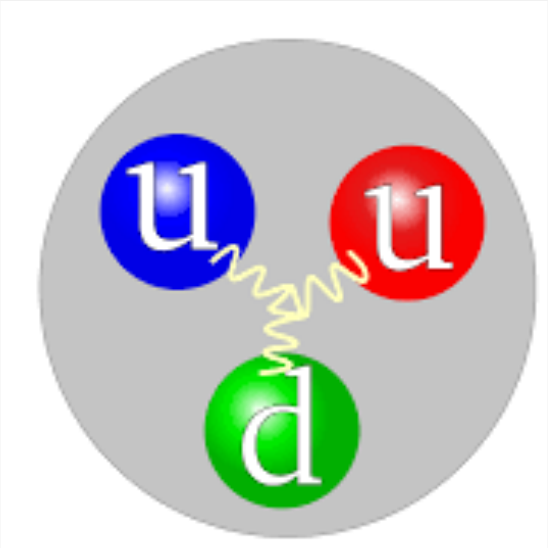
An atom has a nucleus (made of protons and neutrons) orbited by electrons.

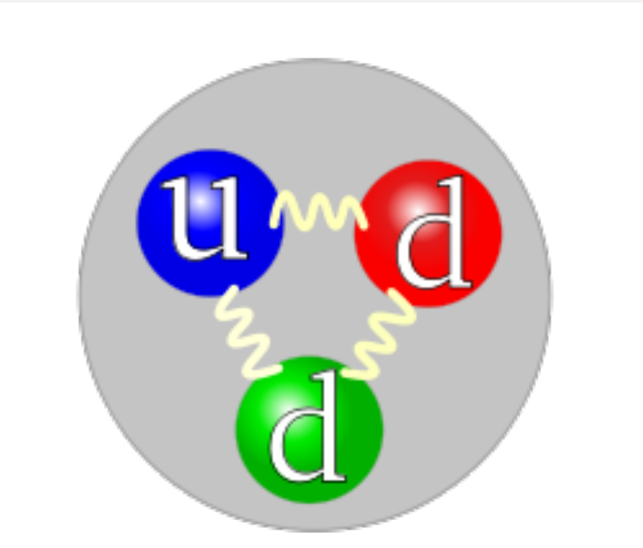
A crash course in Particle Physics



An atom has a nucleus (made of protons and neutrons) orbited by electrons.

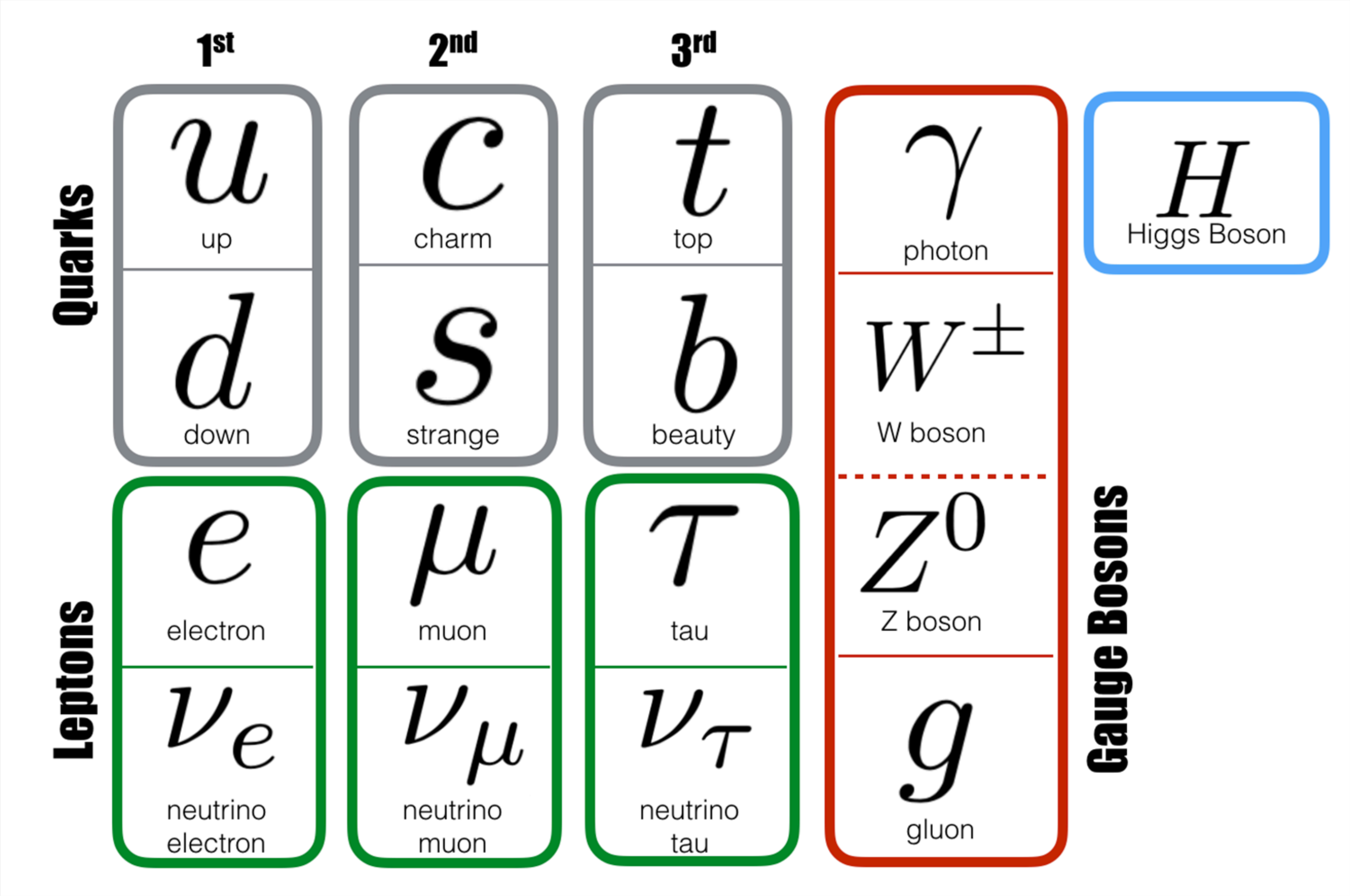






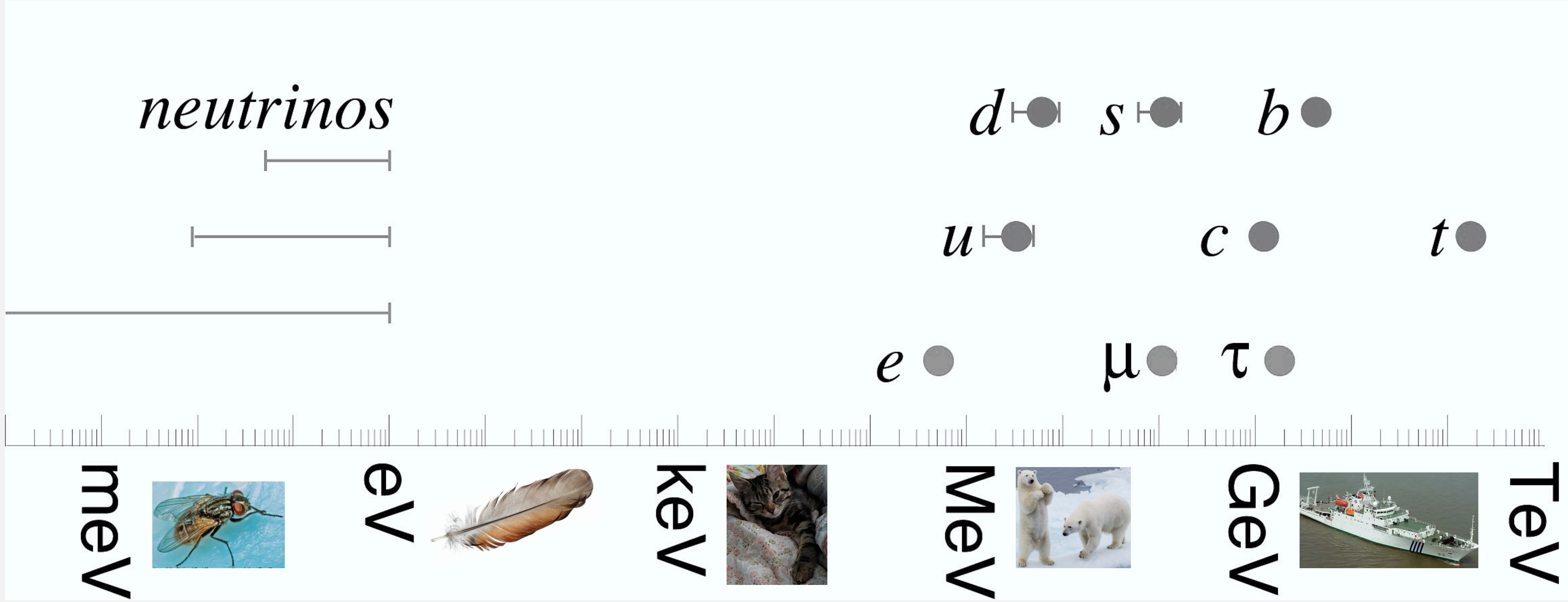
- Protons: 2 up and 1 down quarks.
- Neutrons: 1 up and 2 down quarks.
- Electrons are fundamental and stable.
- Neutrinos are counterparts to electrons, but are rarely discussed despite being extremely numerous and having played an important role in the evolution of the Universe.

A crash course in Particle Physics



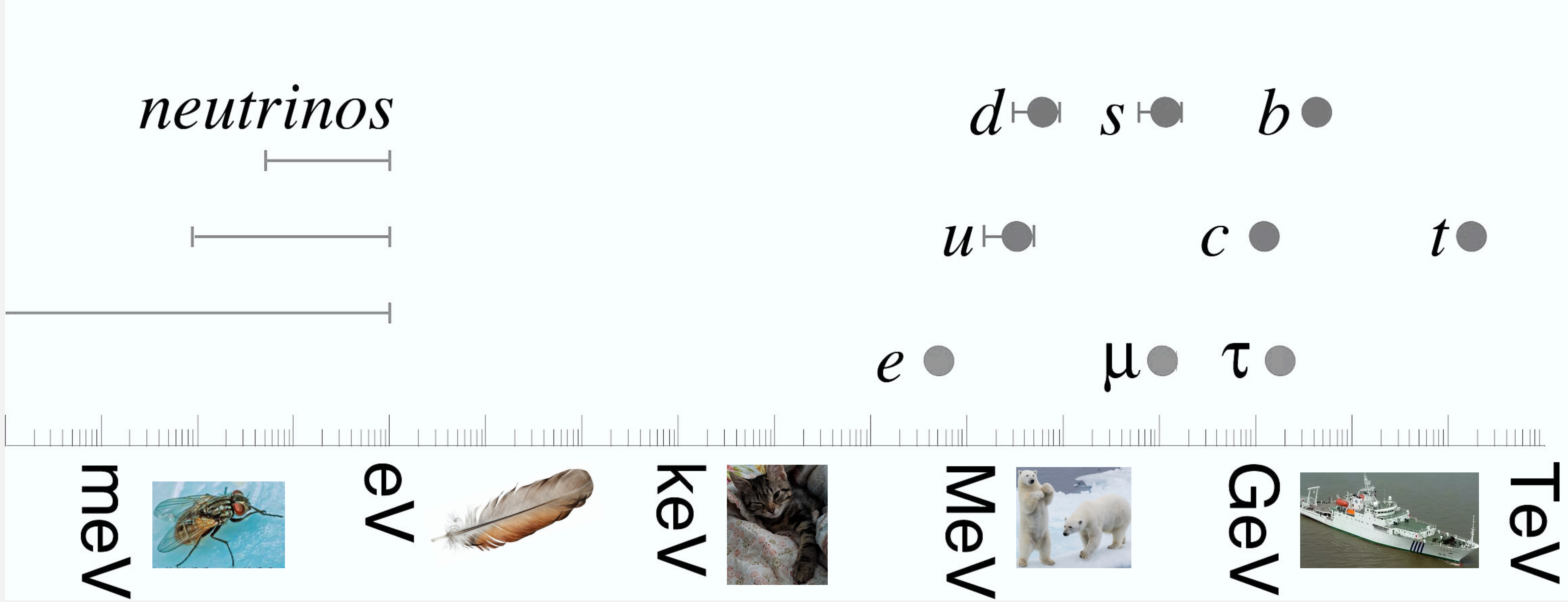
- The Standard Model of Particle Physics.
- Various hints that it isn't complete;
  - Doesn't explain dark matter,
  - Doesn't explain dark energy,
  - Doesn't the matter/anti-matter asymmetry in the Universe,
  - Doesn't explain "Grand Unified Theories" such as Super Symmetry,
  - *Struggles to explain the properties of neutrinos which we observe…*

A crash course in Particle Physics – Neutrinos

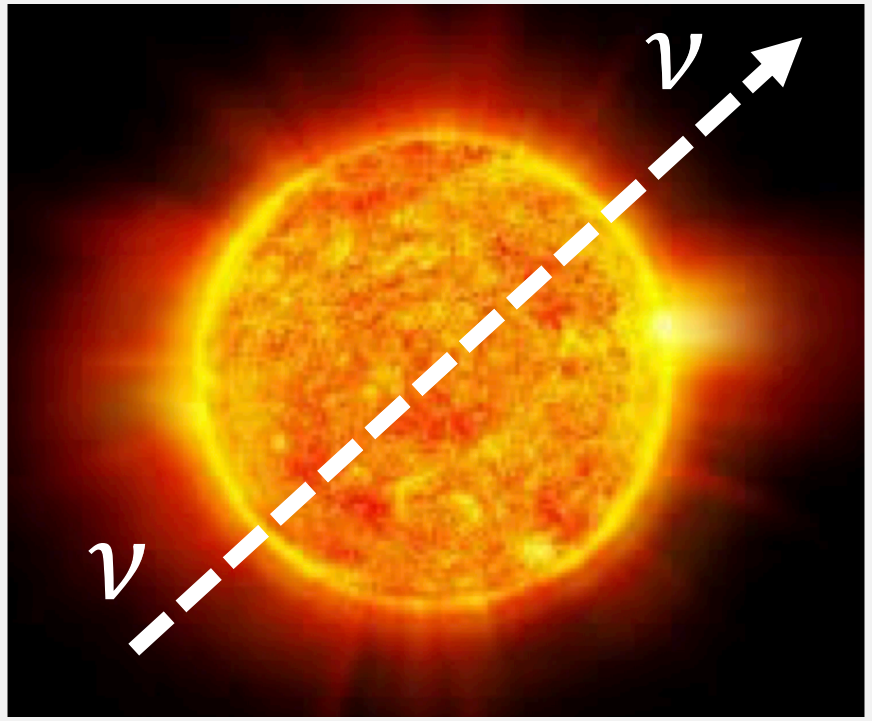


- Neutrinos are really light.
  - This "mass-gap" cannot currently be explained without some pretty extreme modifications to the Standard Model.
- They are also the only fundamental particle with no electric charge.

A crash course in Particle Physics – Neutrinos



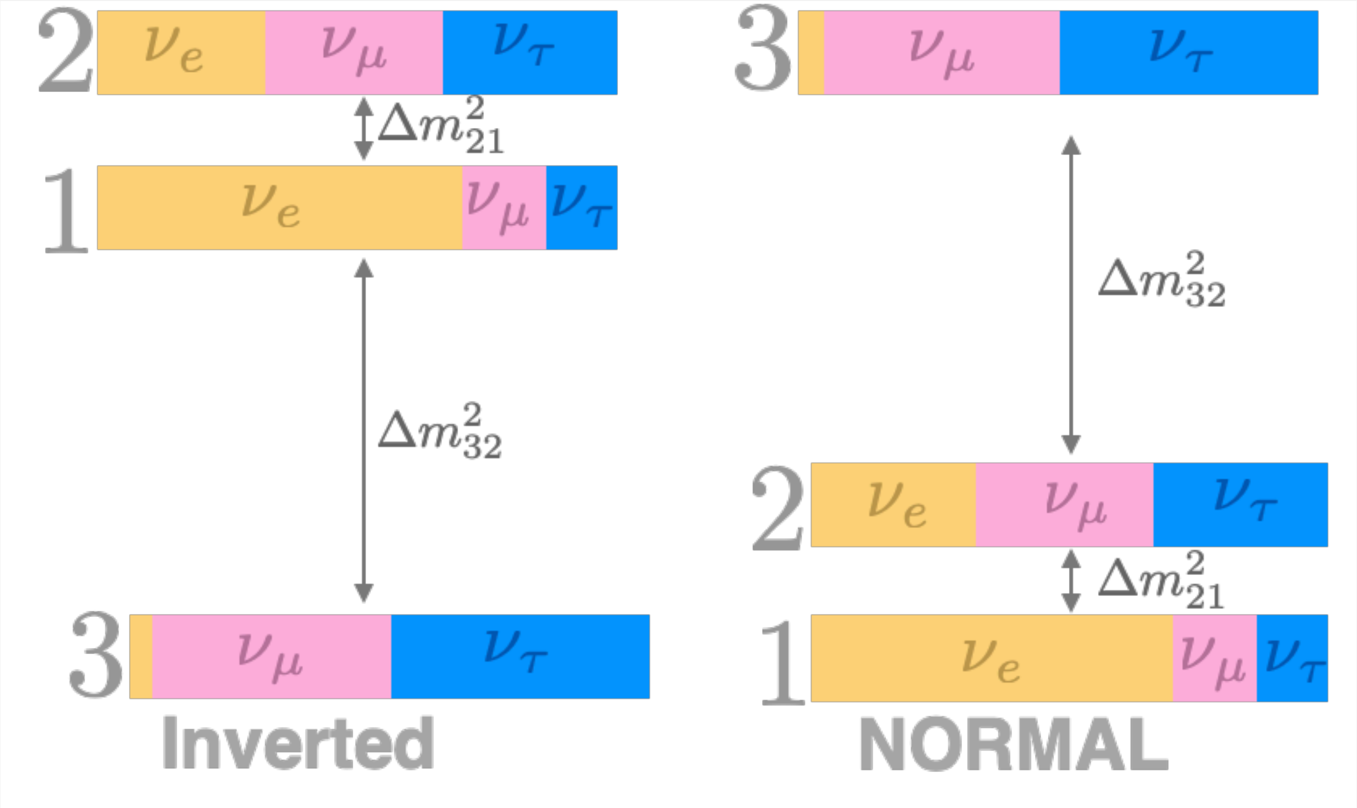
- Neutrinos are really light.
  - This "mass-gap" cannot currently be explained without some pretty extreme modifications to the Standard Model.
- They are also the only fundamental particle with no electric charge.



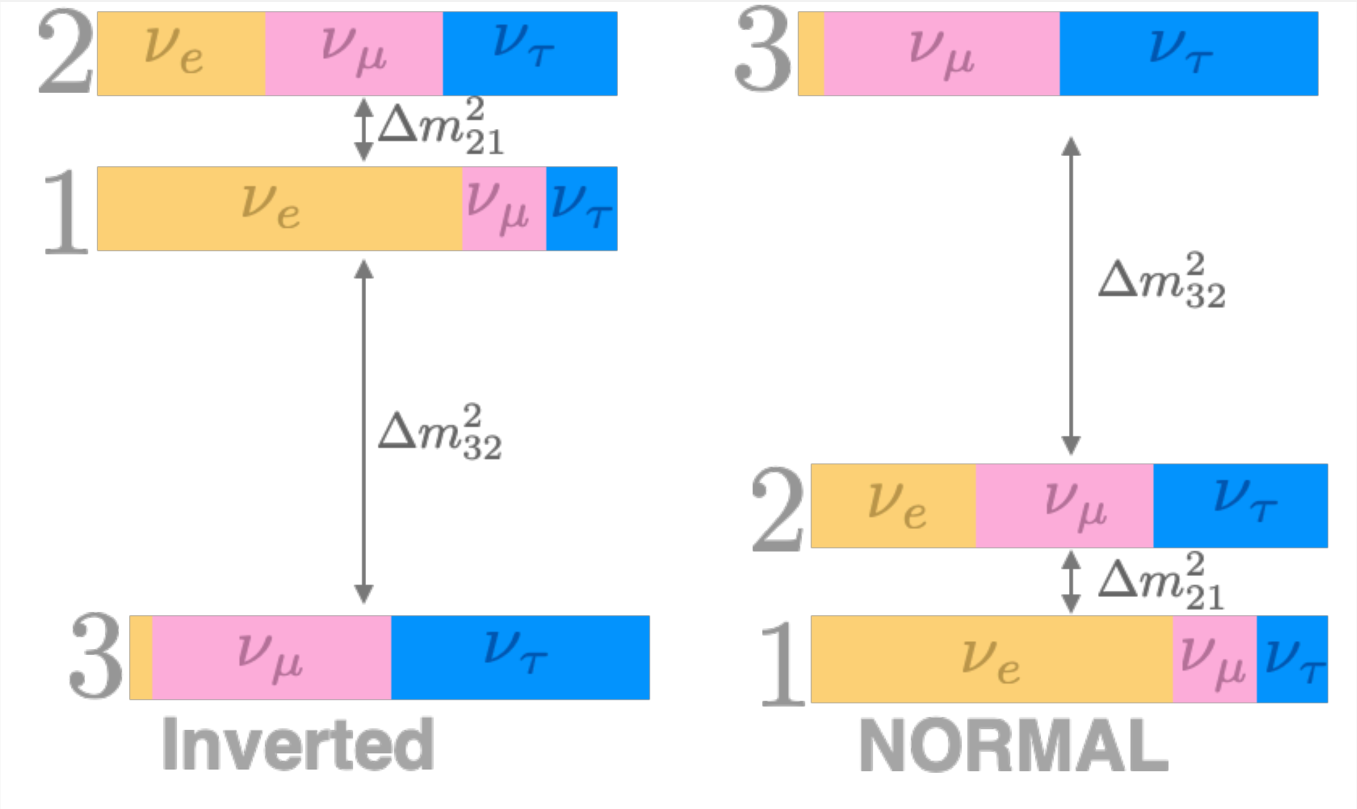
- Neutrinos are the most numerous matter particle in the Universe and are produced by pretty much everything.
- Neutrinos interact much less than any other particle.
  - This makes them very hard to study.

A crash course in Particle Physics – Neutrinos

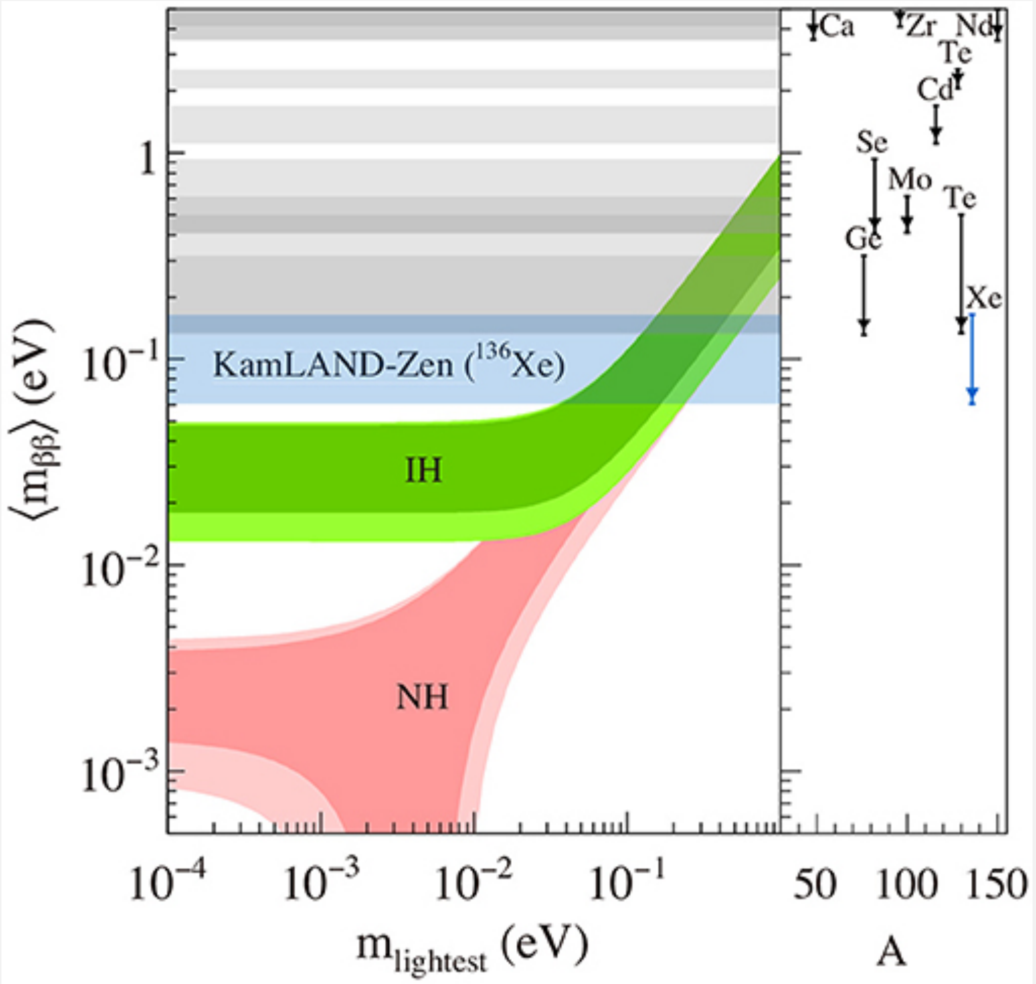
- Neutrinos don't exist as discrete flavour states, but instead as a combination of the three flavour states, called ν_1, ν_2, ν_3 which each have a different mass.
- We have handles on the differences of these masses, but do not know the exact structure of them.

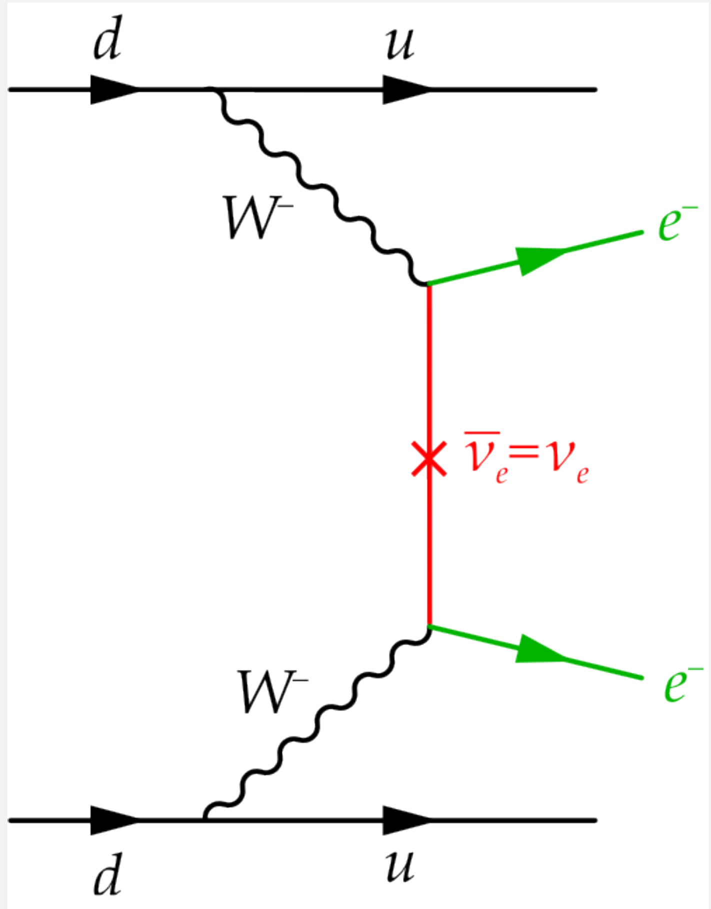


A crash course in Particle Physics – Neutrinos



- Neutrinos don't exist as discrete flavour states, but instead as a combination of the three flavour states, called ν_1, ν_2, ν_3 which each have a different mass.
- We have handles on the differences of these masses, but do not know the exact structure of them.





- It is possible that the nature of neutrino masses are different to that of all other particles.
- The observation of a forbidden process in the SM would shows that this is true, as well as possibly resolving the order of the neutrino masses.

A crash course in Particle Physics – Neutrino Oscillations

$\nu_\alpha \ \nu_\alpha \ \nu_\alpha$

$\nu_\alpha \ \nu_\beta \ \nu_\alpha \ \nu_\alpha$

$\nu_\alpha \ \nu_\alpha \ \nu_\alpha \ \nu_\alpha \ \nu_\alpha$

$\nu_\alpha \ \nu_\alpha \ \nu_\alpha \ \nu_\alpha$

$\nu_\alpha \ \nu_\alpha \ \nu_\alpha$

Imagine a cluster of neutrinos.

It is initially very pure, with almost all being of flavour ν_α, though there is a small contamination of ν_β.

A crash course in Particle Physics – Neutrino Oscillations

$\nu_\alpha\ \nu_\alpha\ \nu_\alpha$

$\nu_\alpha\ \nu_\beta\ \nu_\alpha\ \nu_\alpha$

$\nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha \xrightarrow[L_1]{} \nu_\alpha\ \nu_\beta\ \nu_\alpha\ \nu_\beta\ \nu_\beta$

$\nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha$

$\nu_\alpha\ \nu_\alpha\ \nu_\alpha$

$\nu_\beta\ \nu_\beta\ \nu_\alpha$

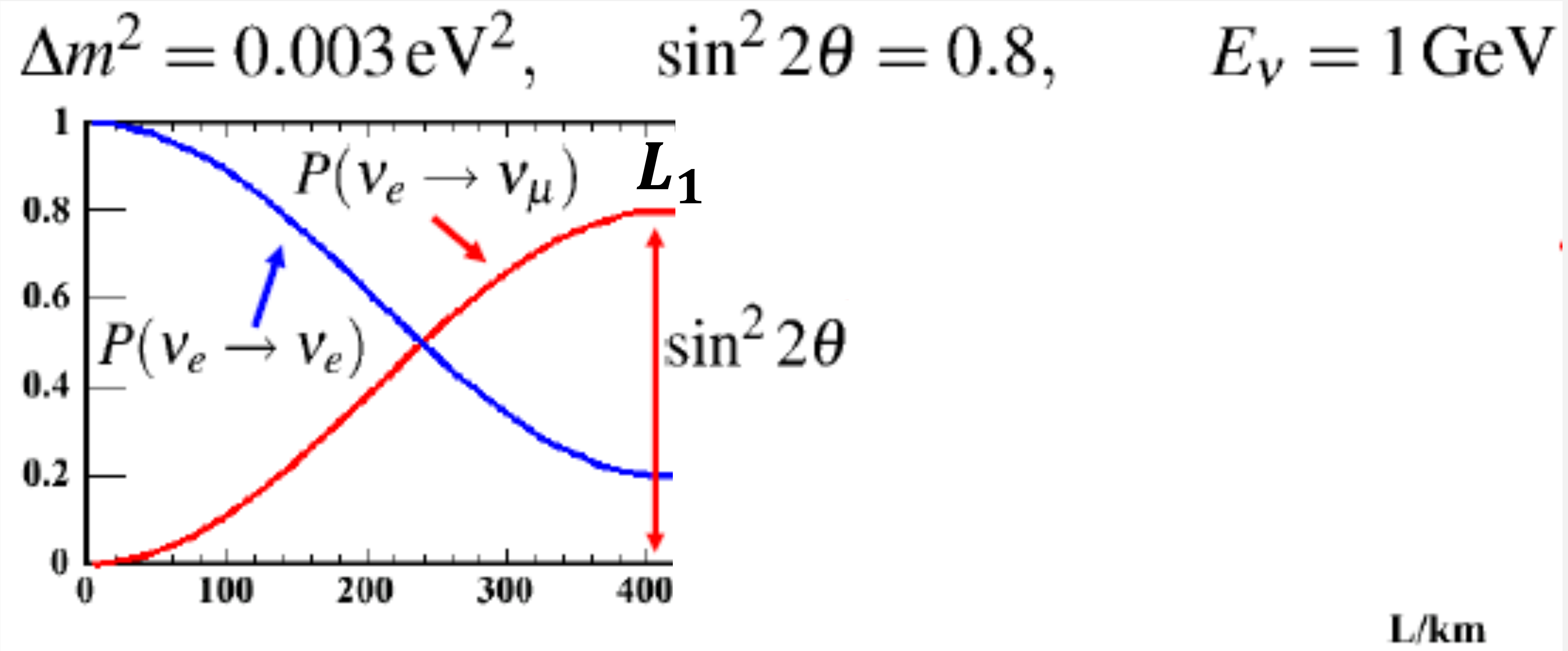
$\nu_\beta\ \nu_\beta\ \nu_\alpha\ \nu_\beta$

$\nu_\beta\ \nu_\alpha\ \nu_\beta\ \nu_\alpha$

$\nu_\beta\ \nu_\beta\ \nu_\beta$

The neutrinos travel a distance L_1, over a time *T*.

Many of the initially ν_α neutrinos will behave as a different flavour ν_β.

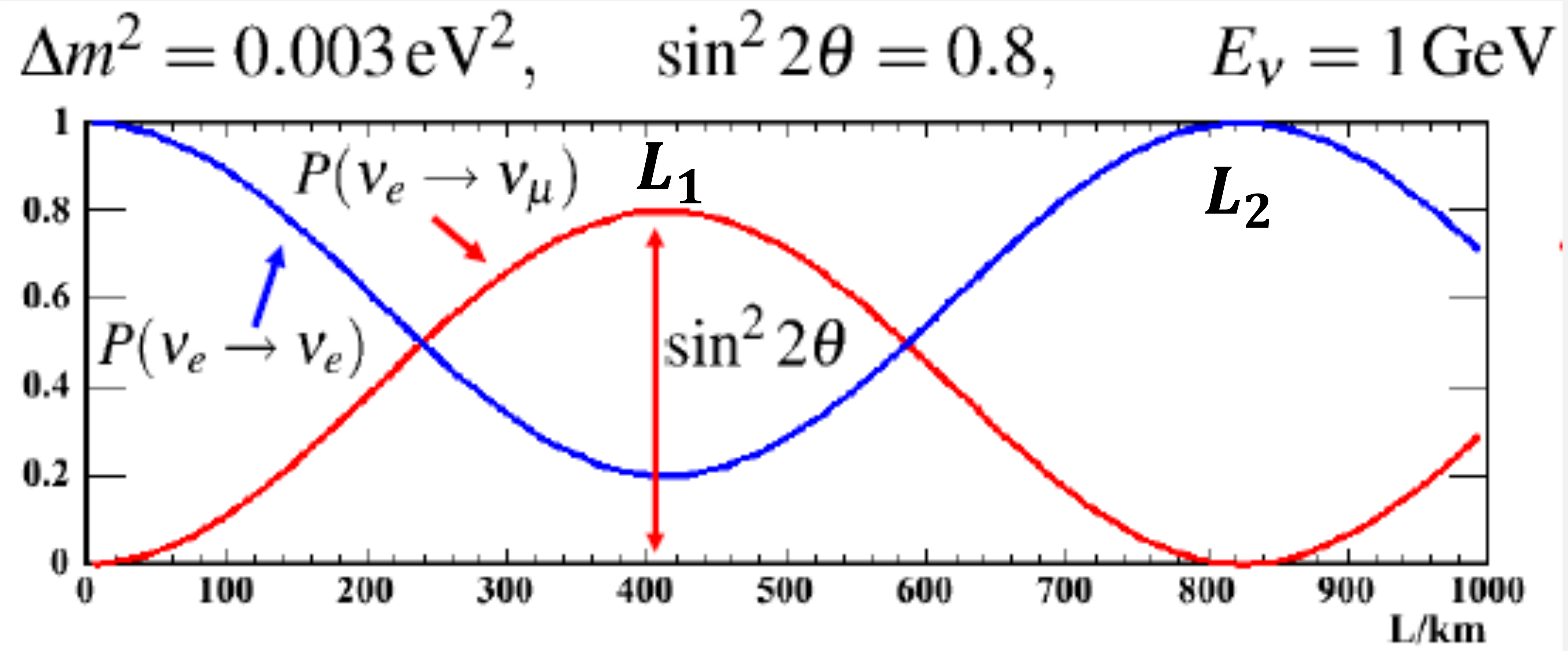


A crash course in Particle Physics – Neutrino Oscillations

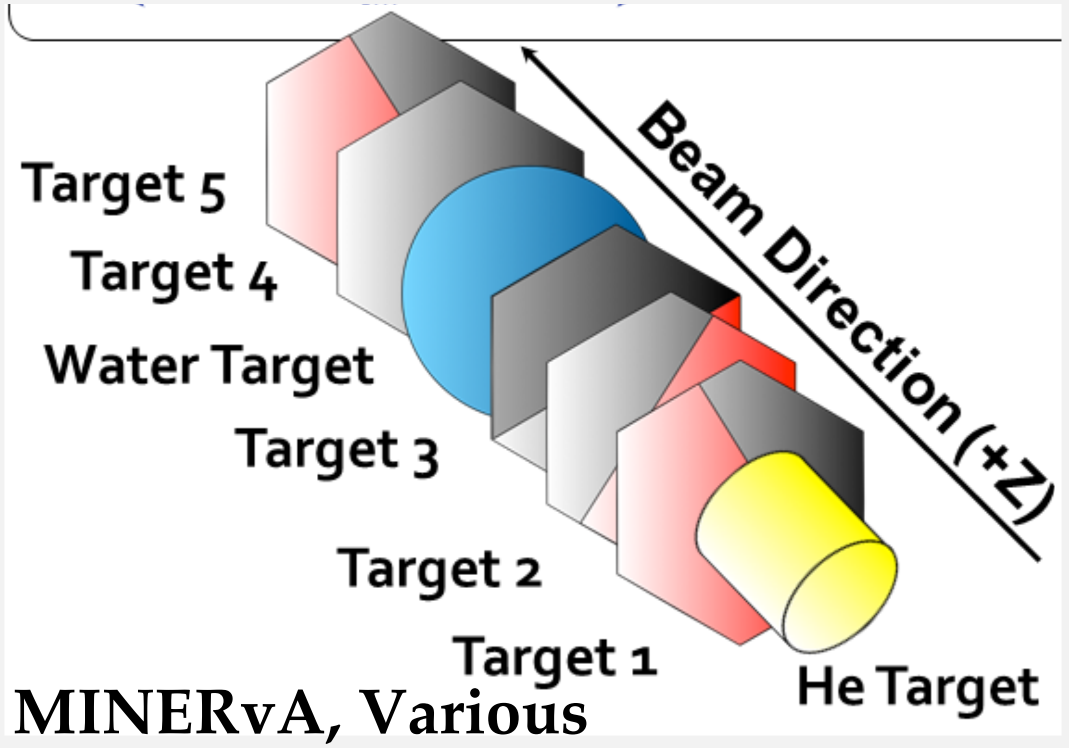
$\nu_\alpha\ \nu_\alpha\ \nu_\alpha$		$\nu_\beta\ \nu_\beta\ \nu_\alpha$		$\nu_\alpha\ \nu_\alpha\ \nu_\beta$
$\nu_\alpha\ \nu_\beta\ \nu_\alpha\ \nu_\alpha$		$\nu_\beta\ \nu_\beta\ \nu_\alpha\ \nu_\beta$		$\nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha$
$\nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha$	$\xrightarrow[L_1]{}$	$\nu_\alpha\ \nu_\beta\ \nu_\alpha\ \nu_\beta\ \nu_\beta$	$\xrightarrow[L_2]{}$	$\nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha$
$\nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha$		$\nu_\beta\ \nu_\alpha\ \nu_\beta\ \nu_\alpha$		$\nu_\alpha\ \nu_\alpha\ \nu_\alpha\ \nu_\alpha$
$\nu_\alpha\ \nu_\alpha\ \nu_\alpha$		$\nu_\beta\ \nu_\beta\ \nu_\beta$		$\nu_\alpha\ \nu_\alpha\ \nu_\alpha$

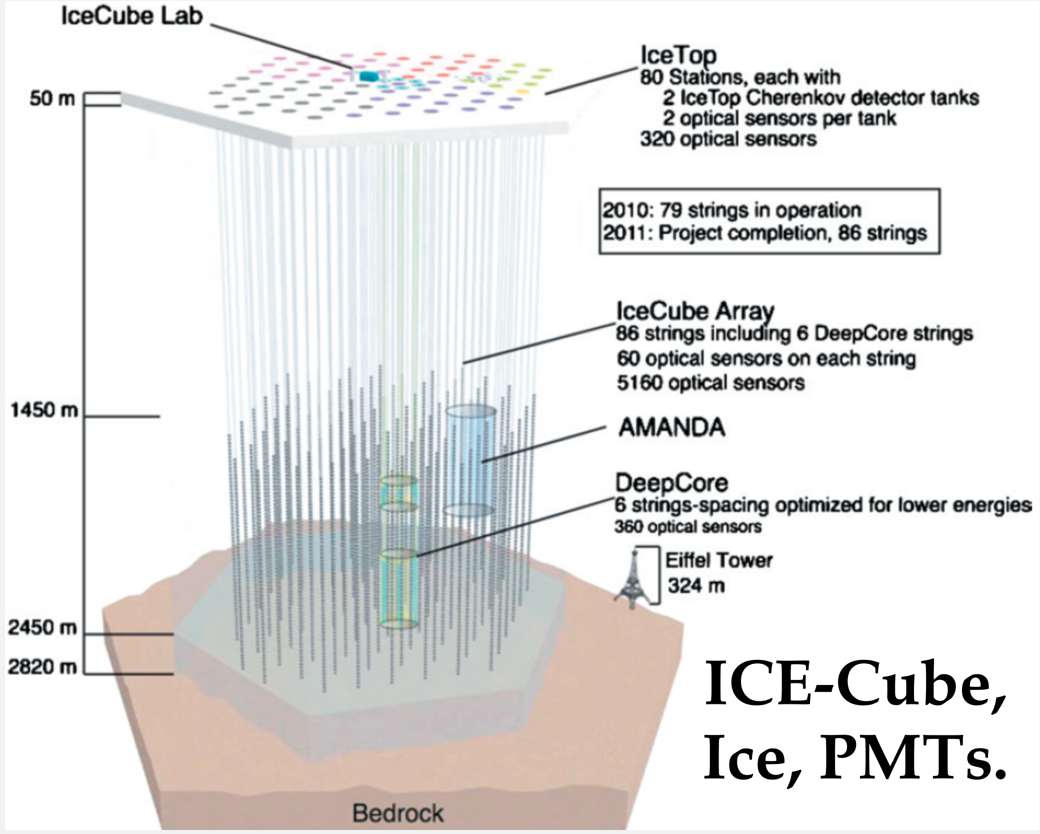
A distance L_2 later most of the ν_β neutrinos will have oscillated back to ν_α.

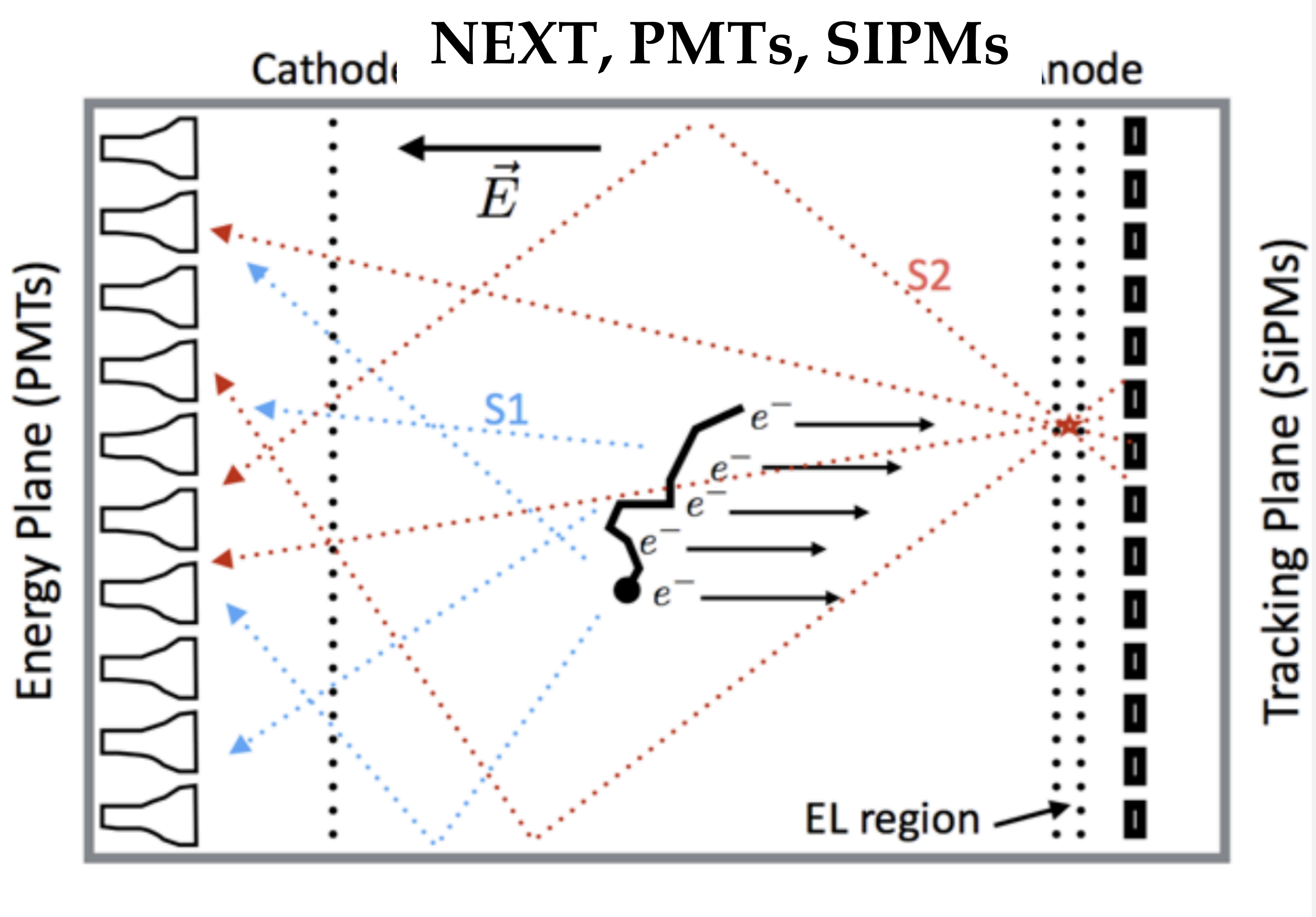
The values of the oscillation parameters affect the rate of oscillations, as does the presence of a 3rd neutrino mass state.

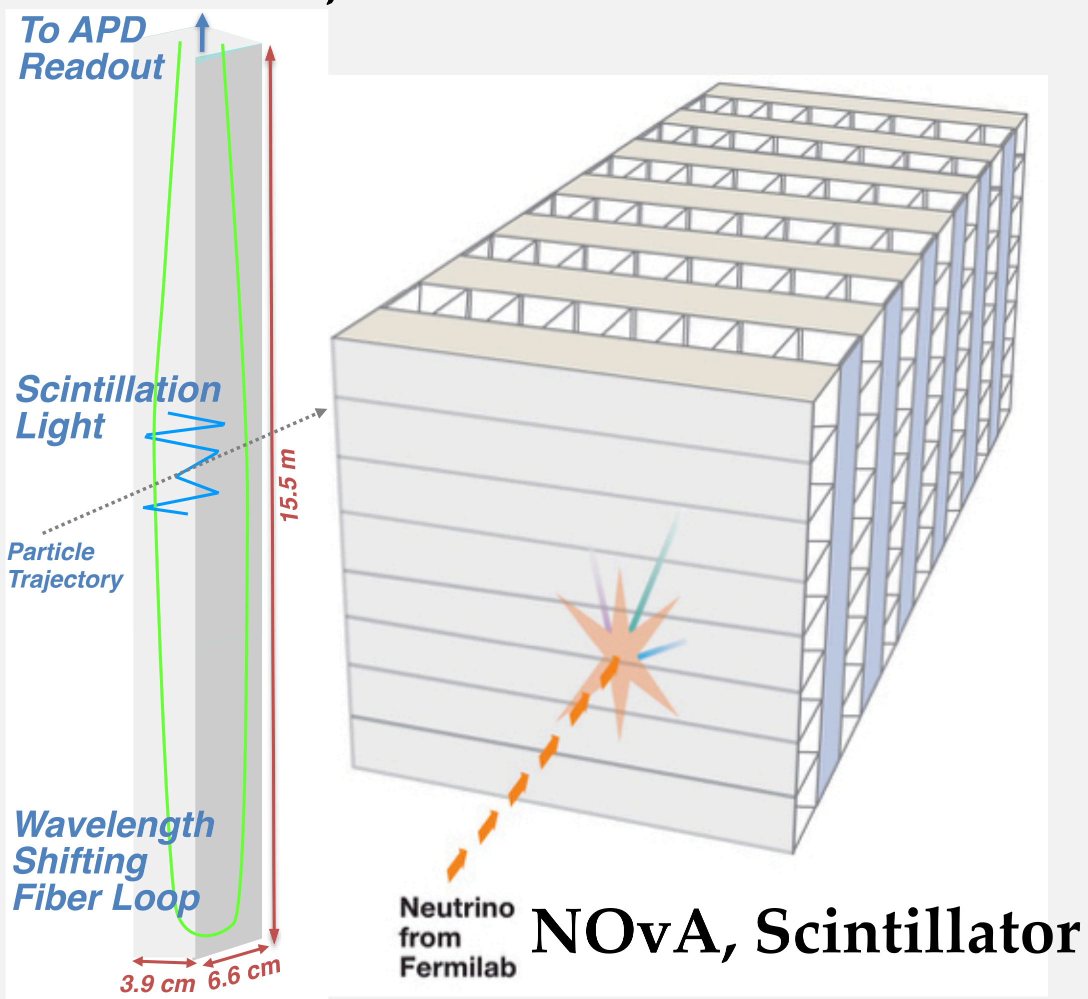


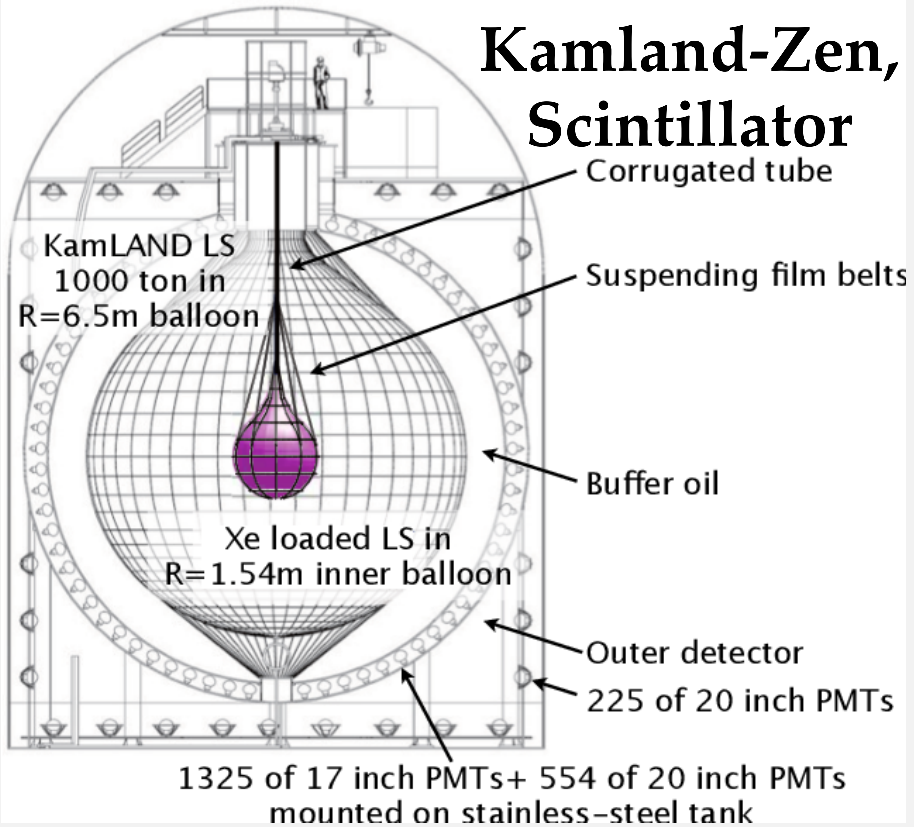
How we Build Neutrino Experiments – LOTS of Different Ways

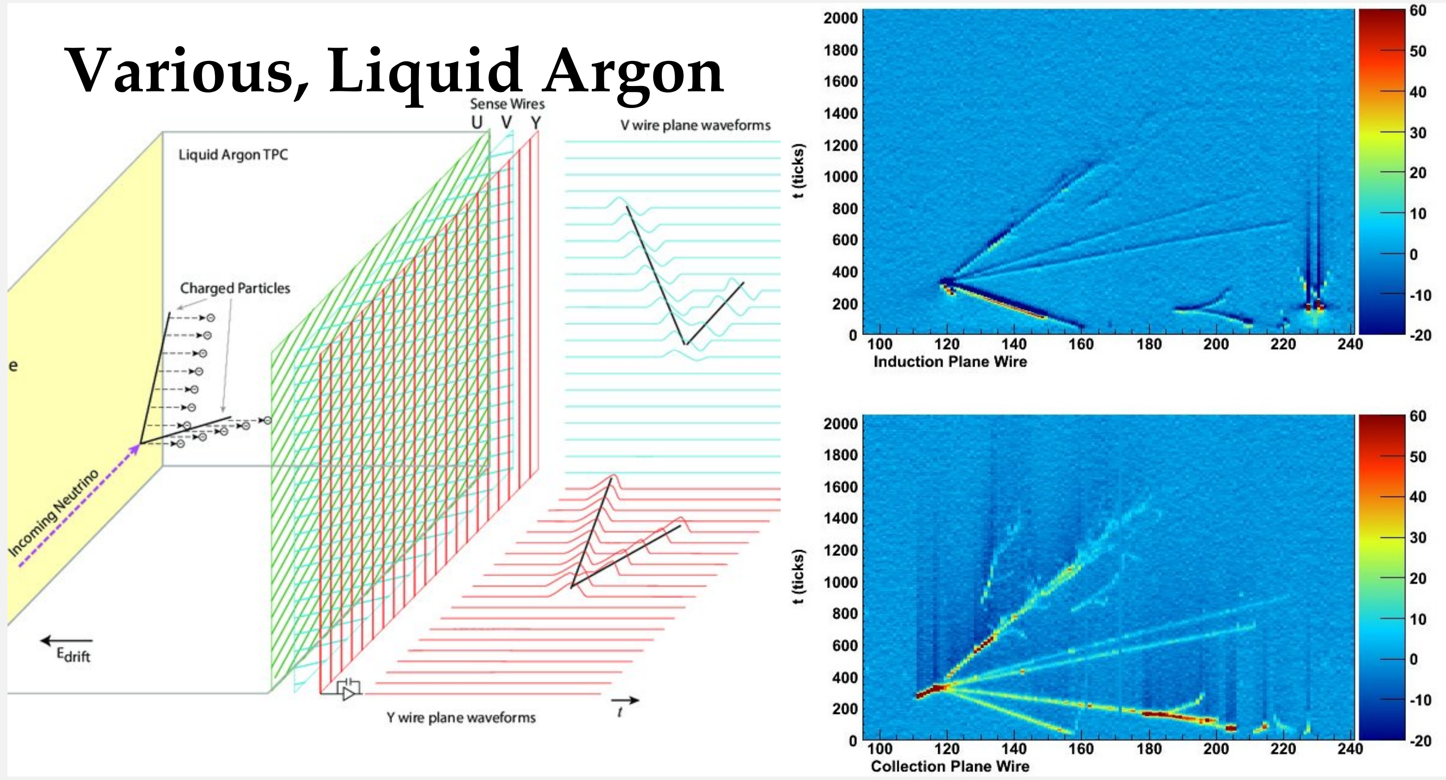






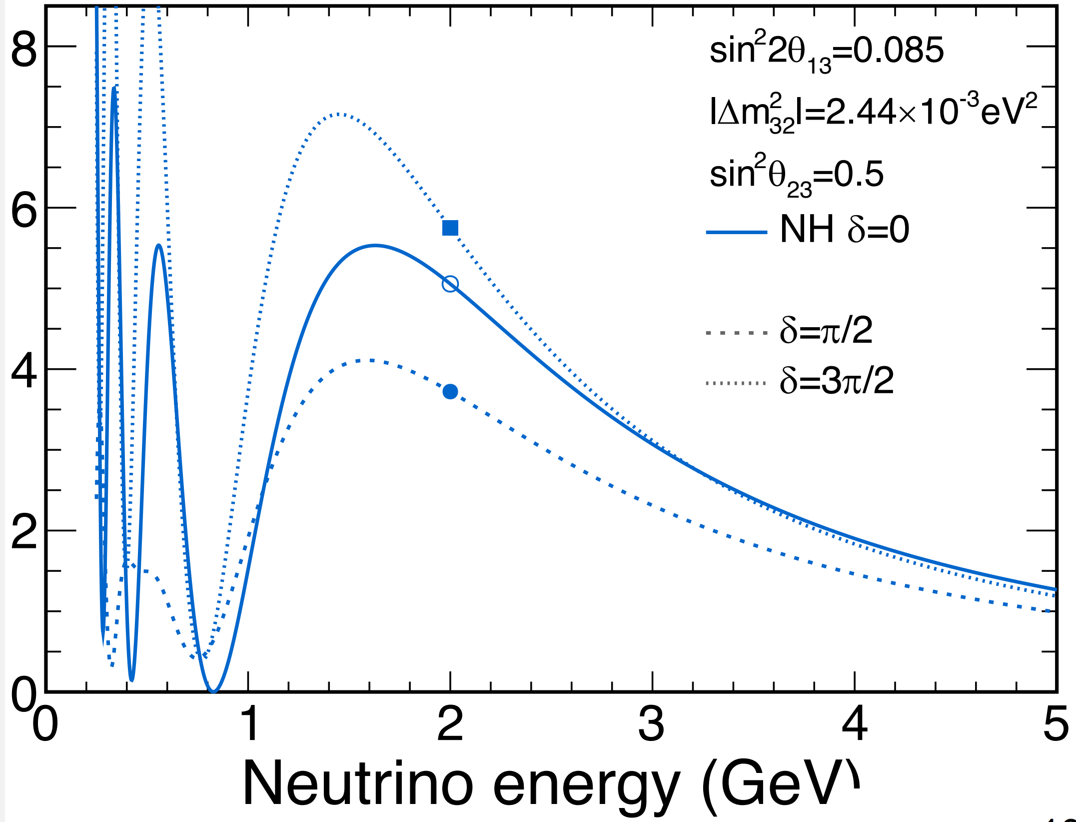






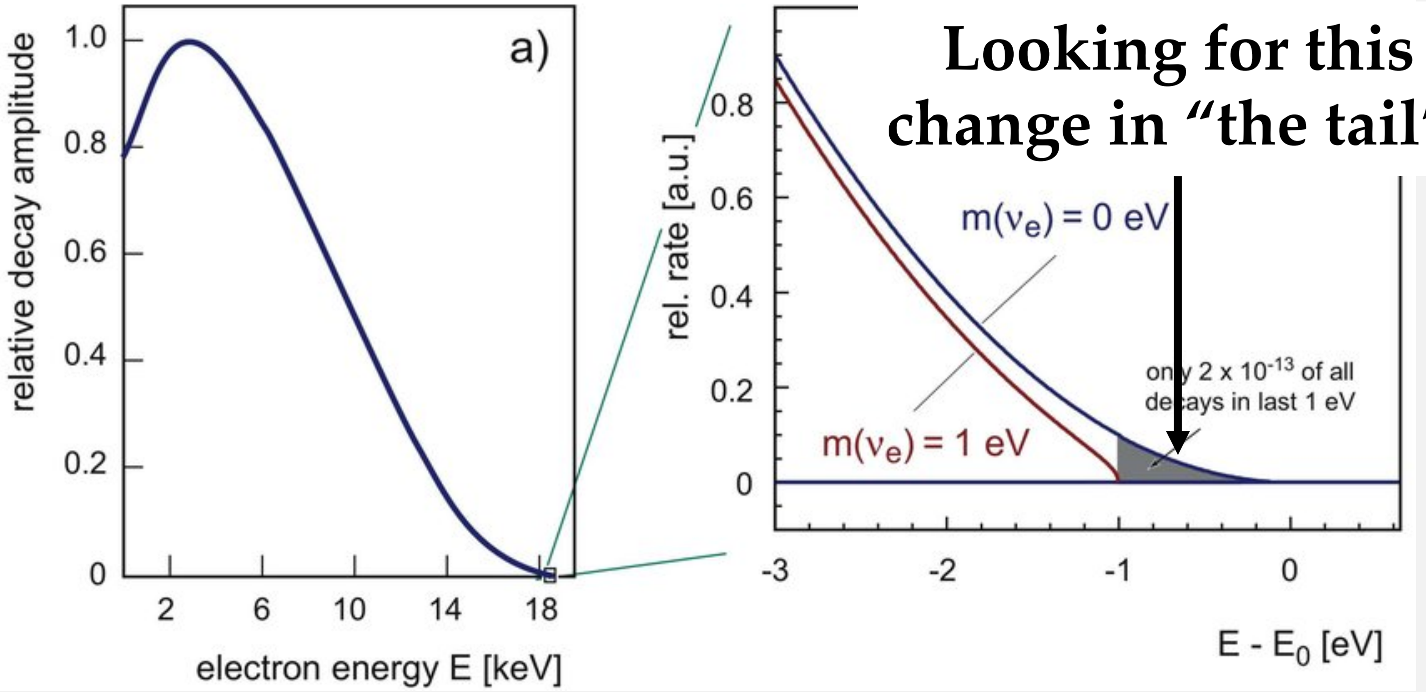
What we aim to do in Neutrino Experiments – Energies

NOvA, neutrino oscillation



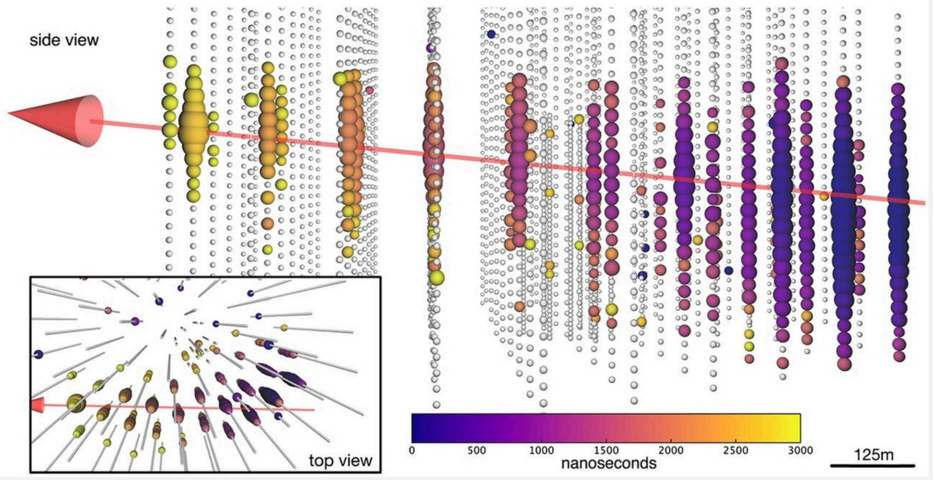
Neutrino interactions occur over a wide range of energies, almost 10 orders of magnitude!

KATRIN, neutrino mass measurements

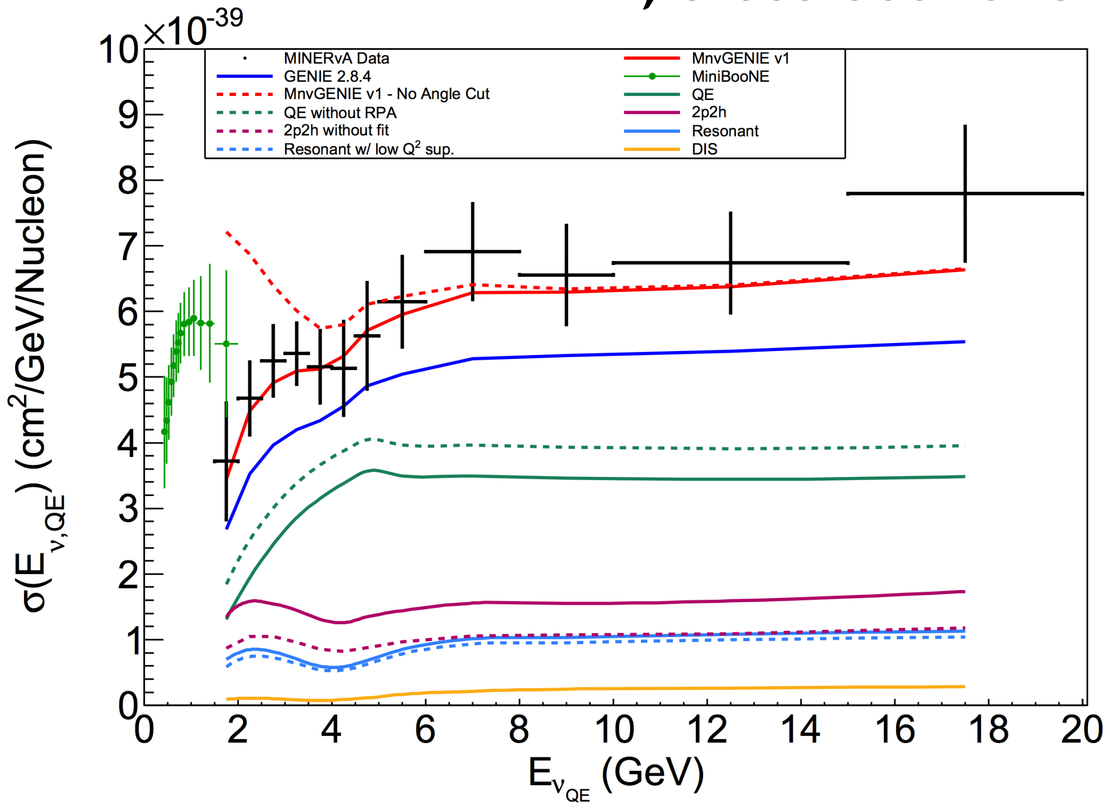


Looking for this change in "the tail"!

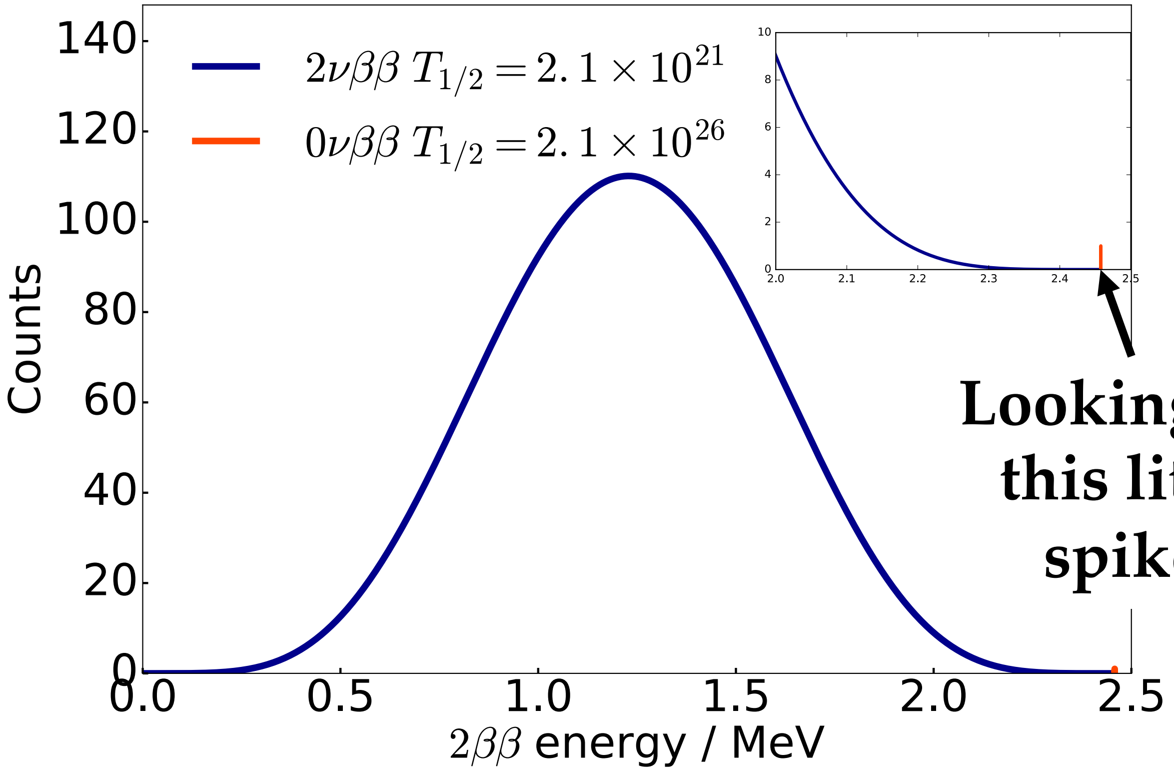
ICE-Cube, PeV energy neutrino!



MINERvA, cross-sections



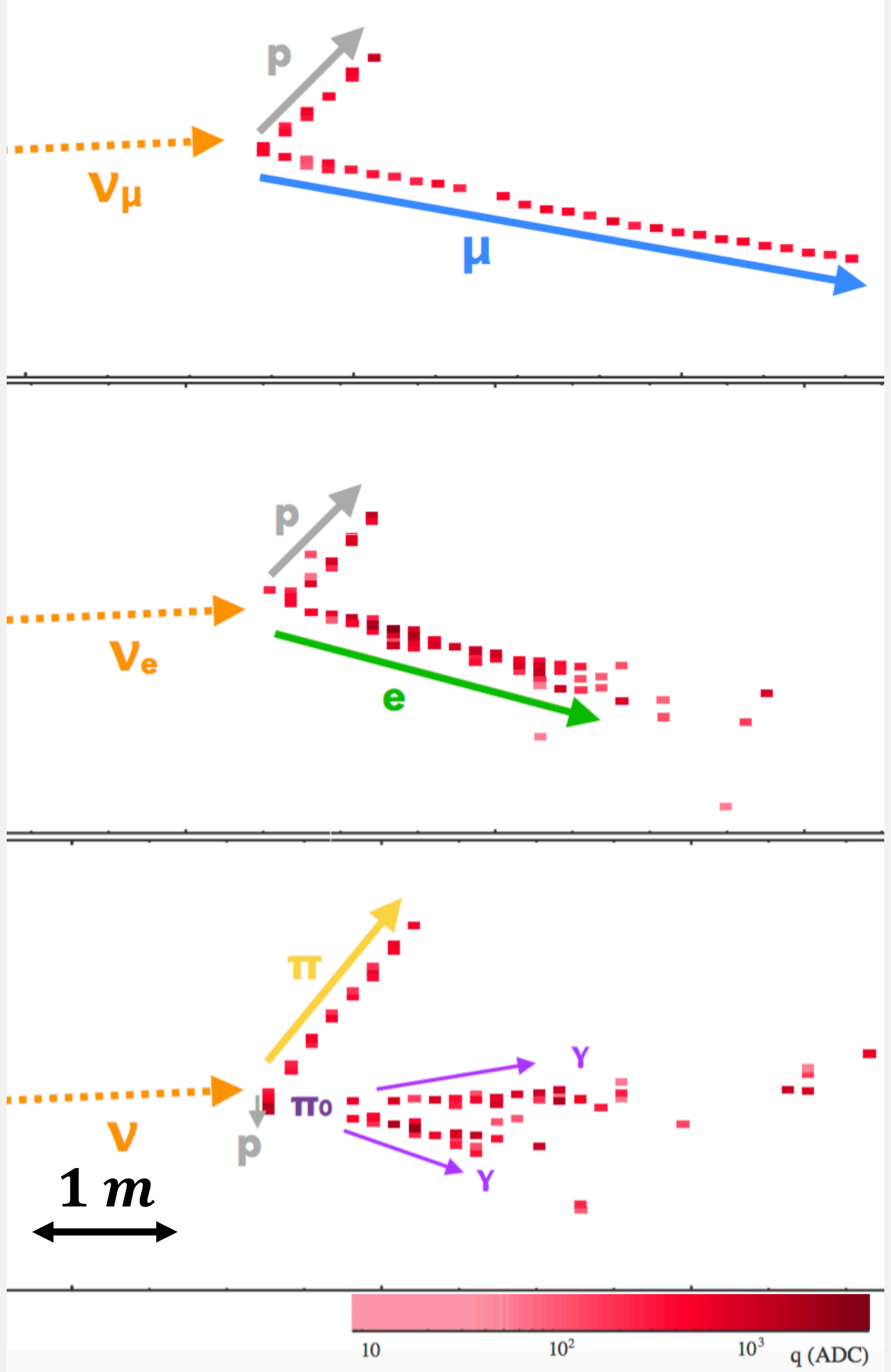
NEXT, neutrino-less double beta decay



Looking for this little spike!

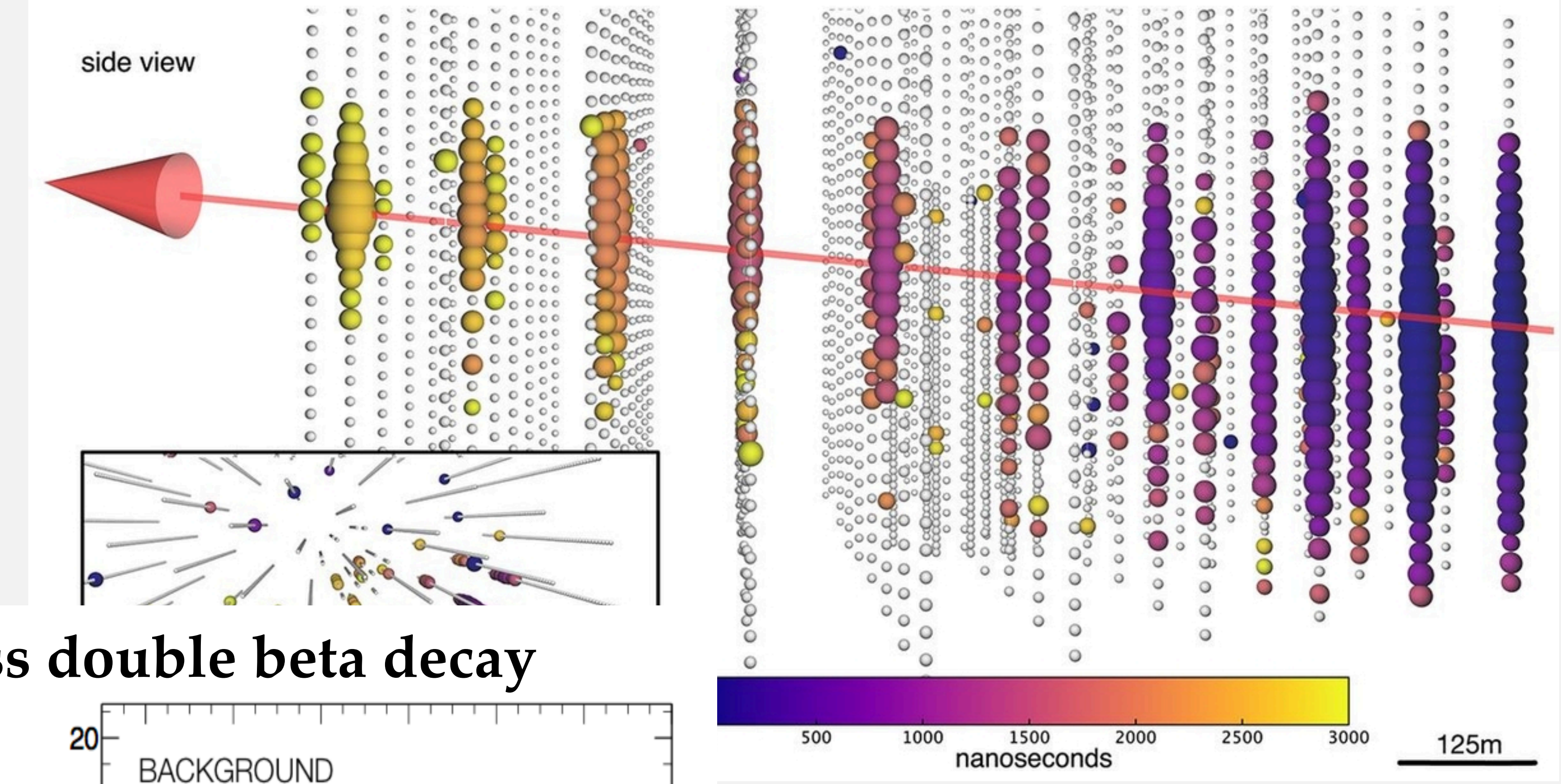
What we aim to do in Neutrino Experiments – Topologies

NOvA, neutrino oscillation

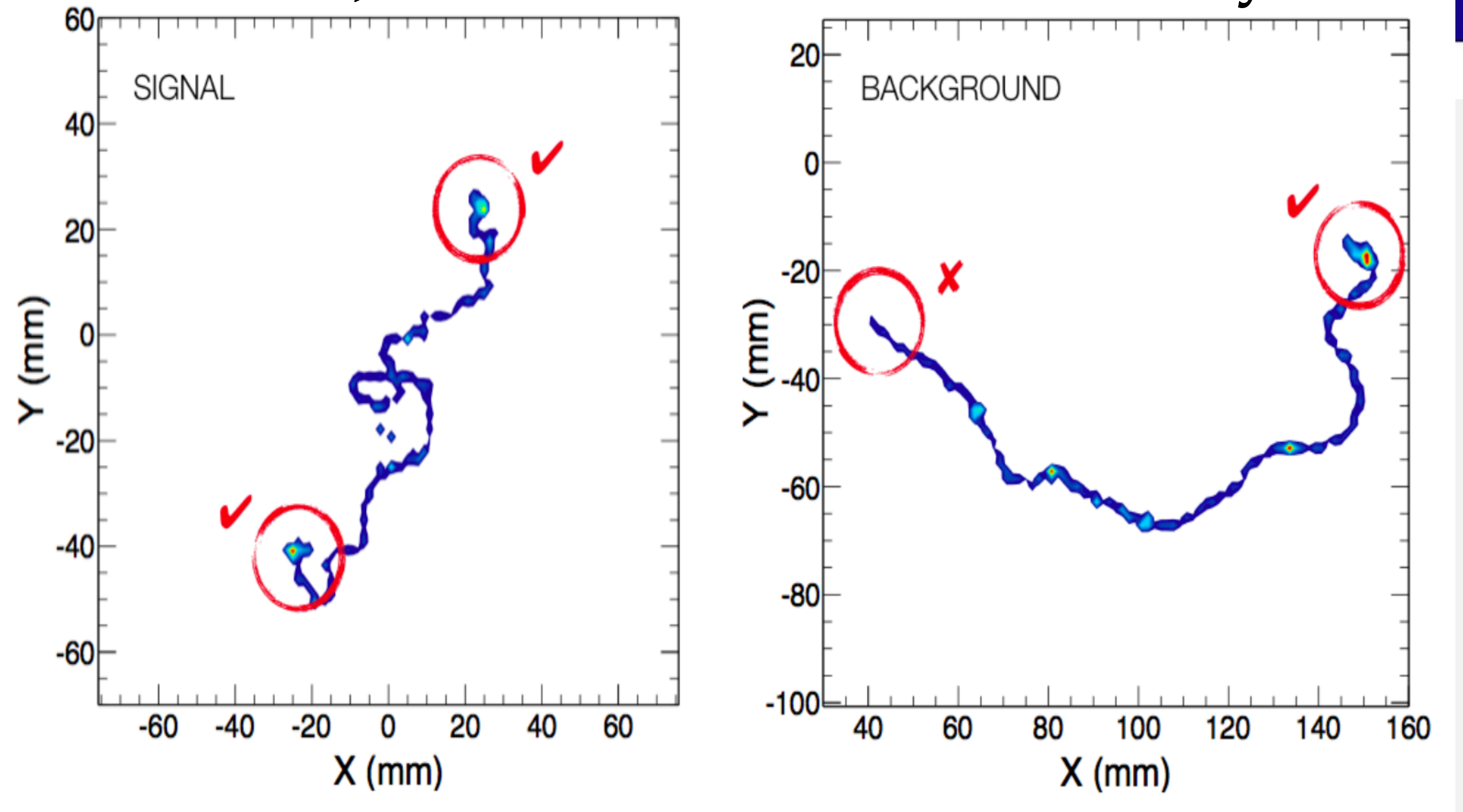


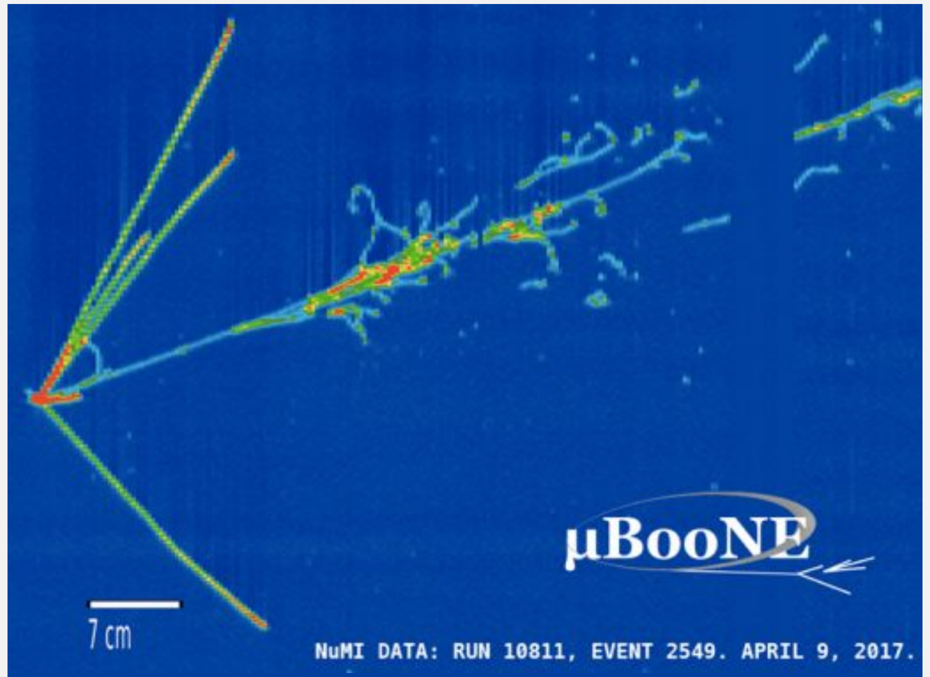
Neutrino interactions can look very different in different detectors.

ICE-Cube, oscillations and high energy neutrinos.



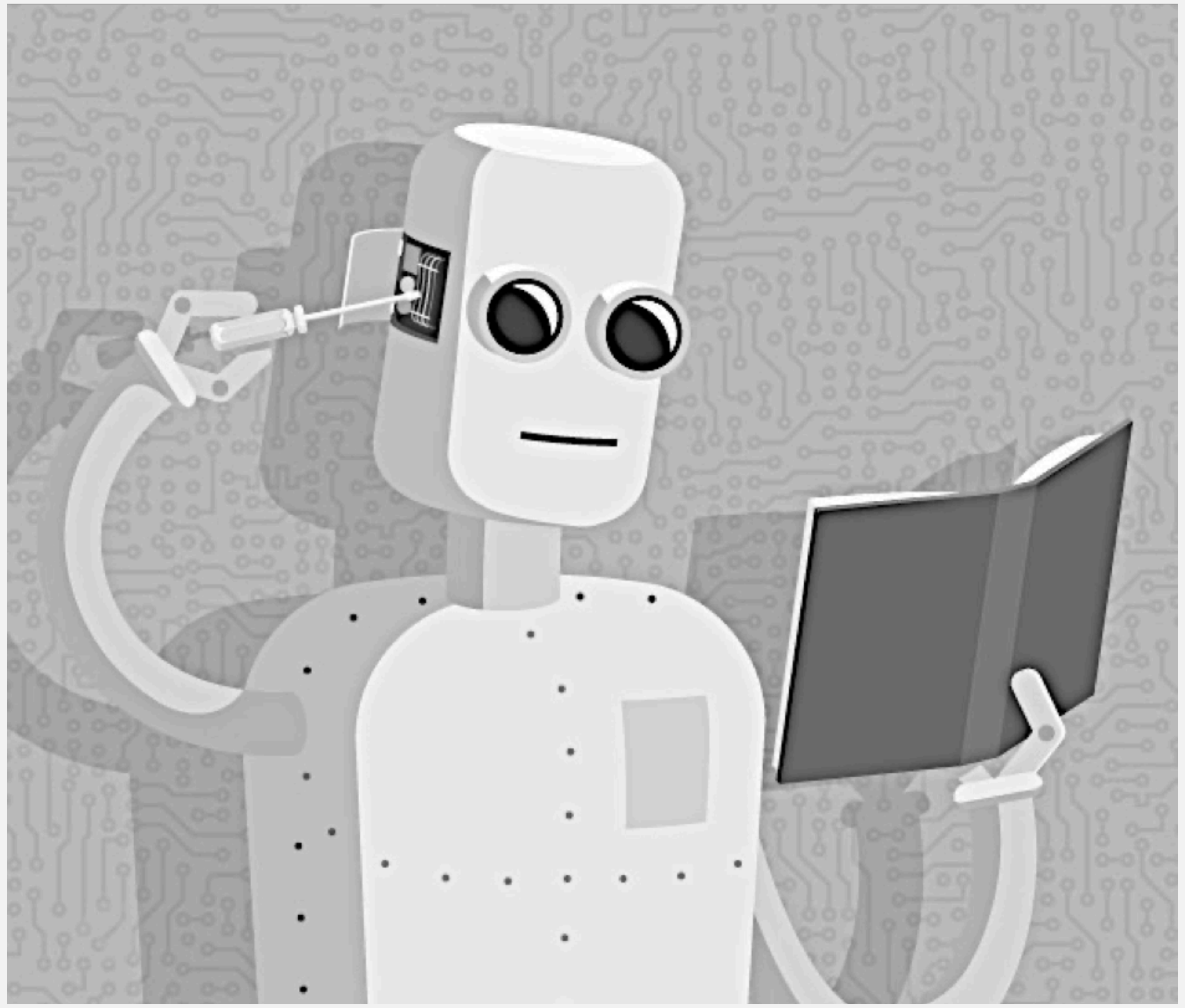
NEXT, neutrino-less double beta decay



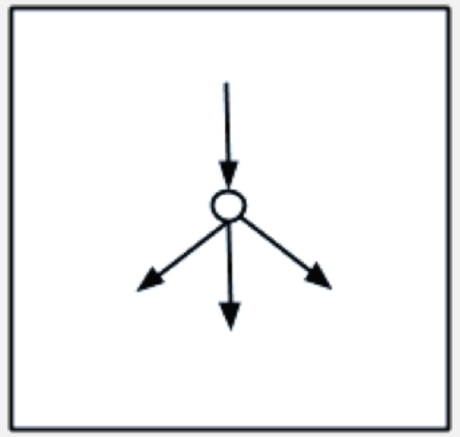


Cross-sections & oscillations

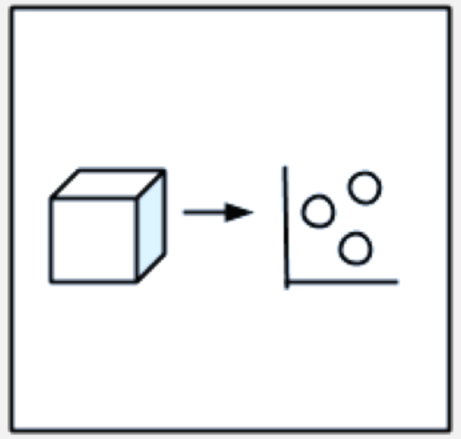
A Brief Introduction to Machine Learning



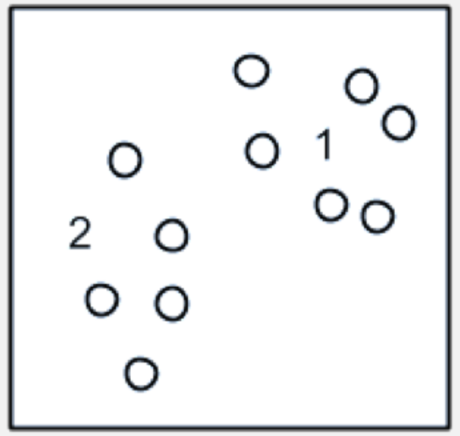
Algorithms whose performance for a given task improves with experience



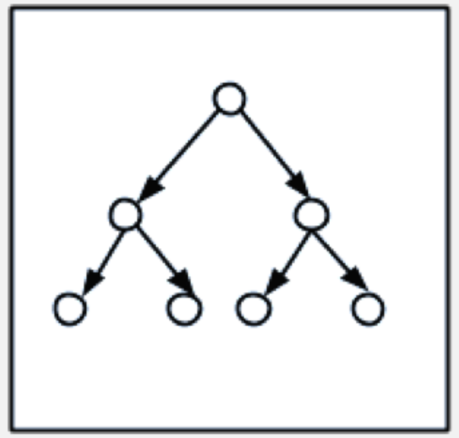
Artificial Neural Network Algorithms



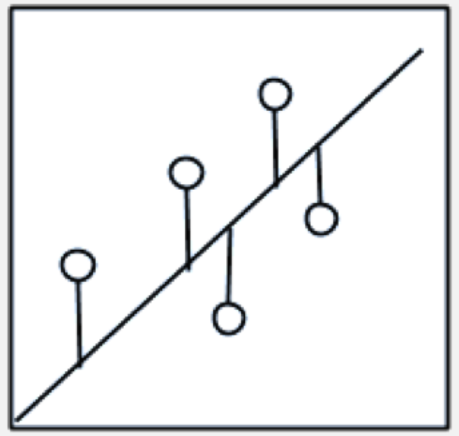
Dimensional Reduction Algorithms



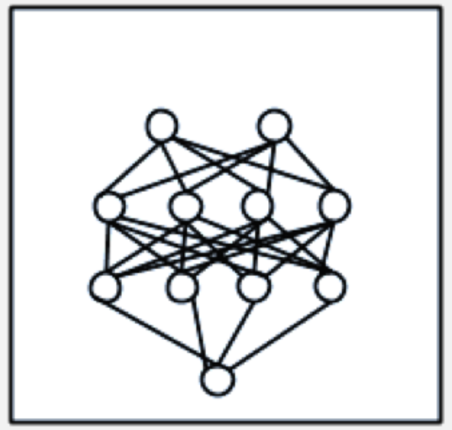
Clustering Algorithms



Decision Tree Algorithms



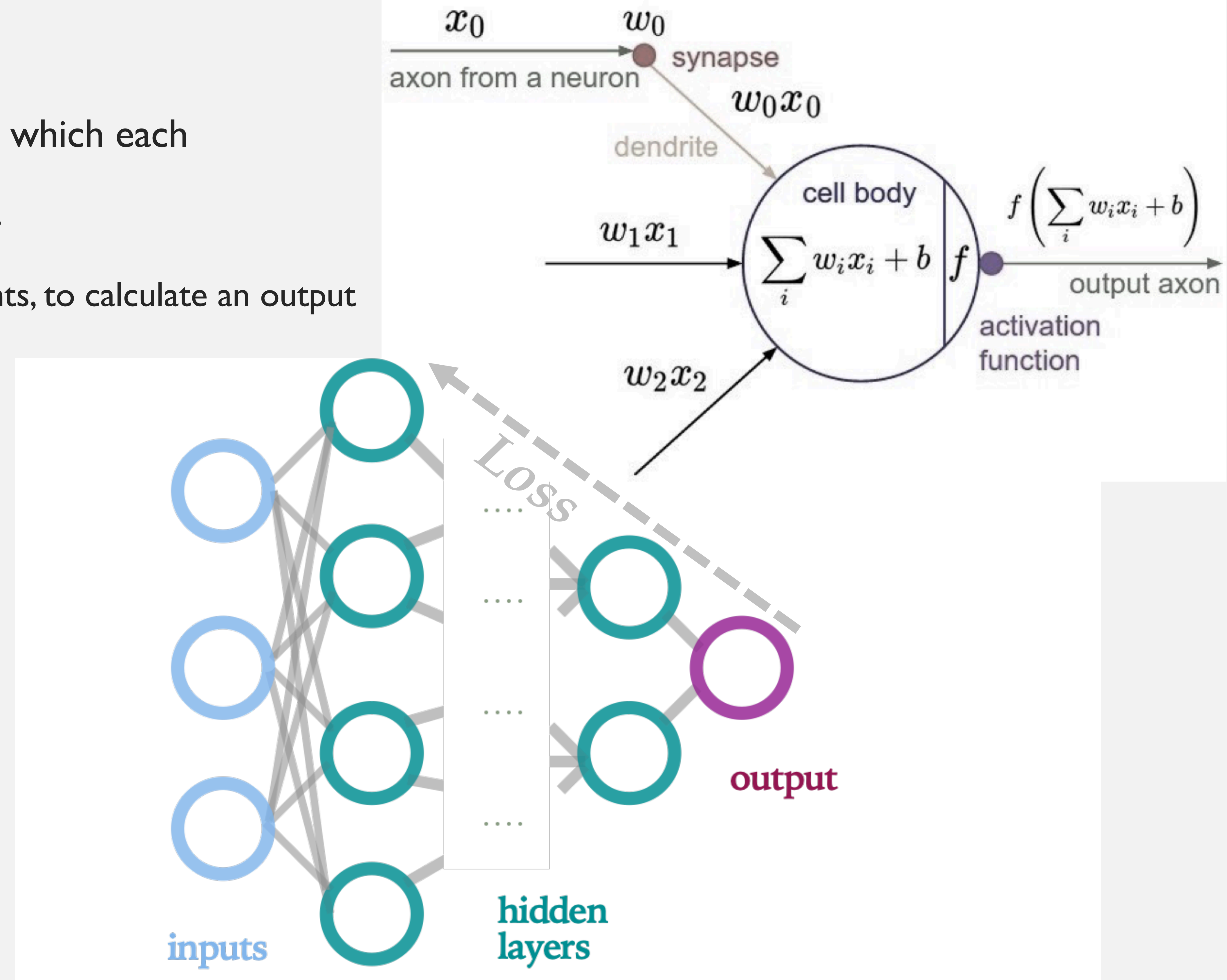
Regression Algorithms



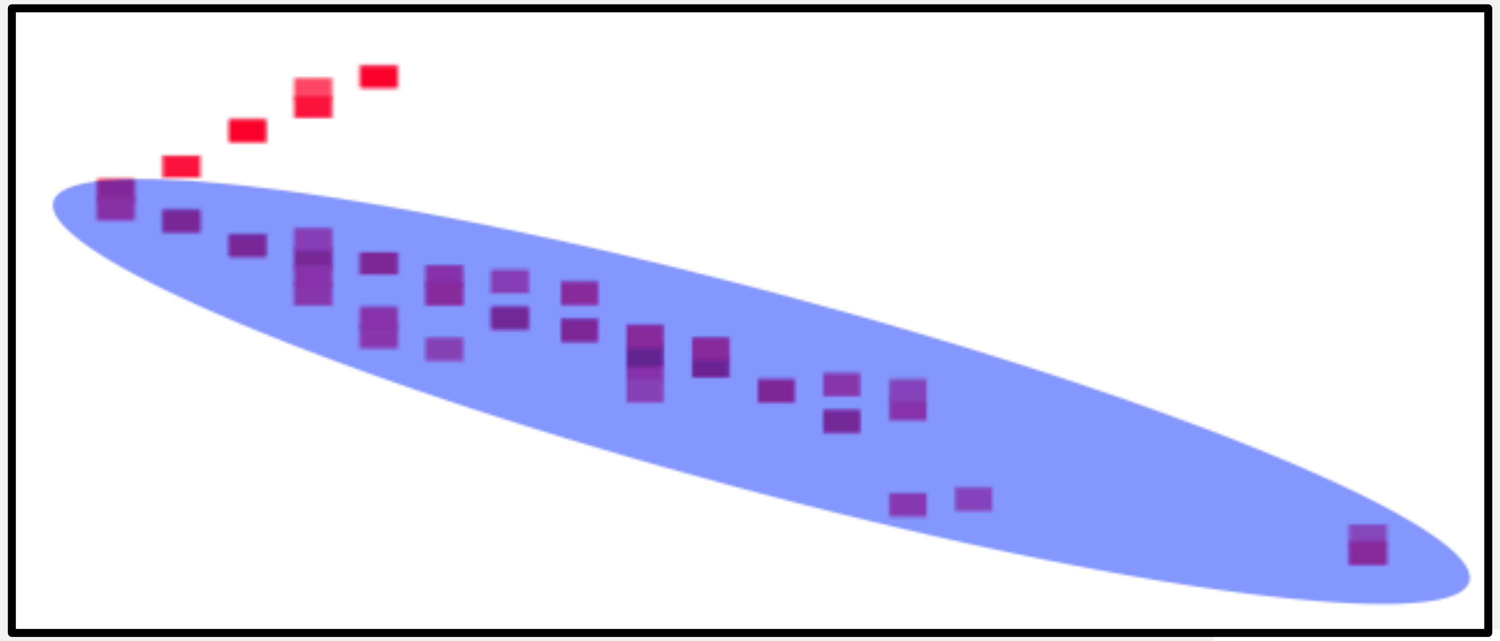
Deep Learning Algorithms

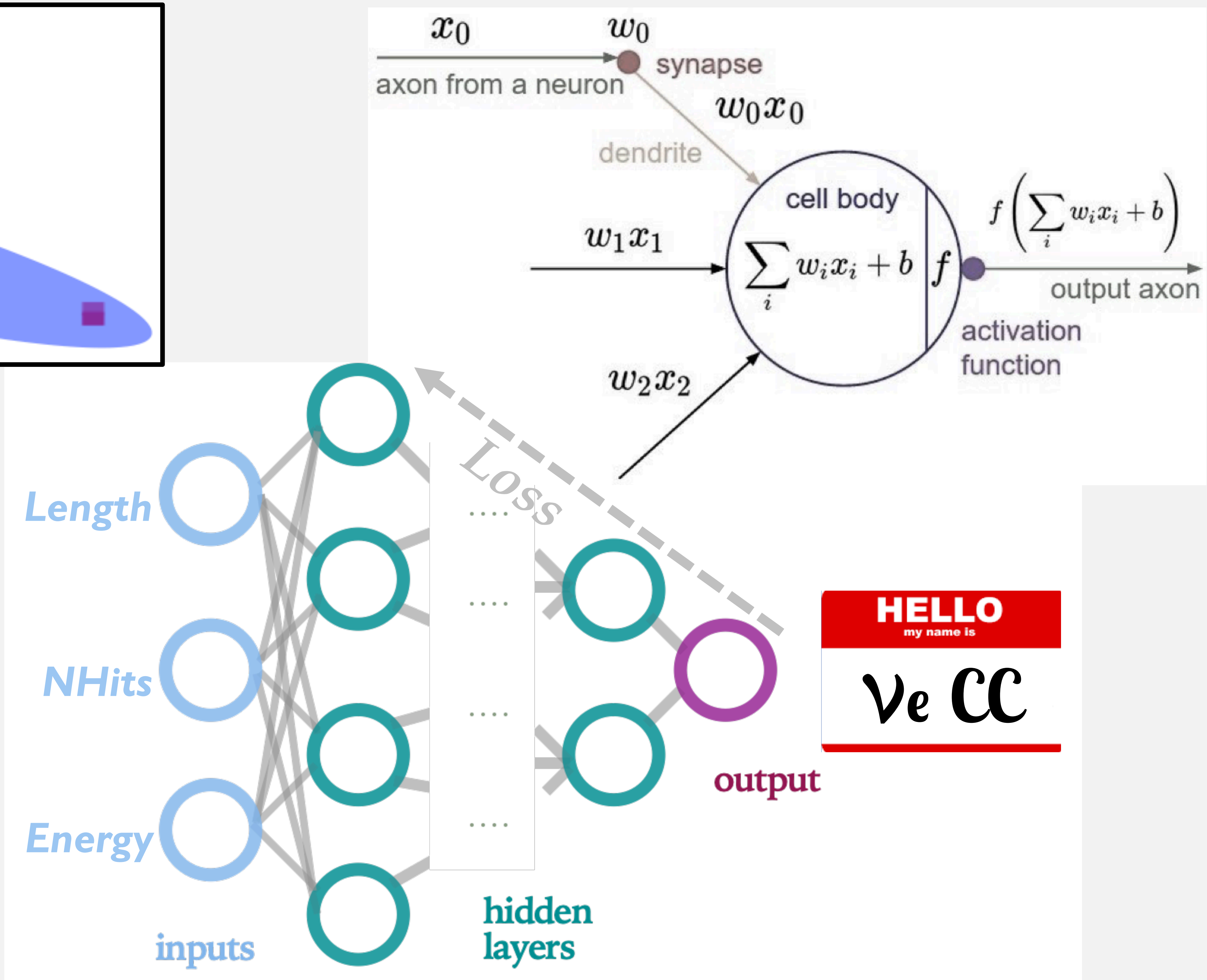
The Structure of Artificial Neural Networks (ANNs/DNNs)

- Simplest way to visualise a neural network.
- A large collection of interconnected neurons which each
  - Take in a number of individually weighted inputs
  - Employ mathematical functions, activation weights, to calculate an output weight which is passed to future neurons.
- Neurons are connected in layers allowing the network to learn about the inputs.
  - Arbitrary number of hidden layers.
  - Neurons are ultimately connected to form an output score which the network is trained to achieve.



The Structure of Artificial Neural Networks (ANNs/DNNs)





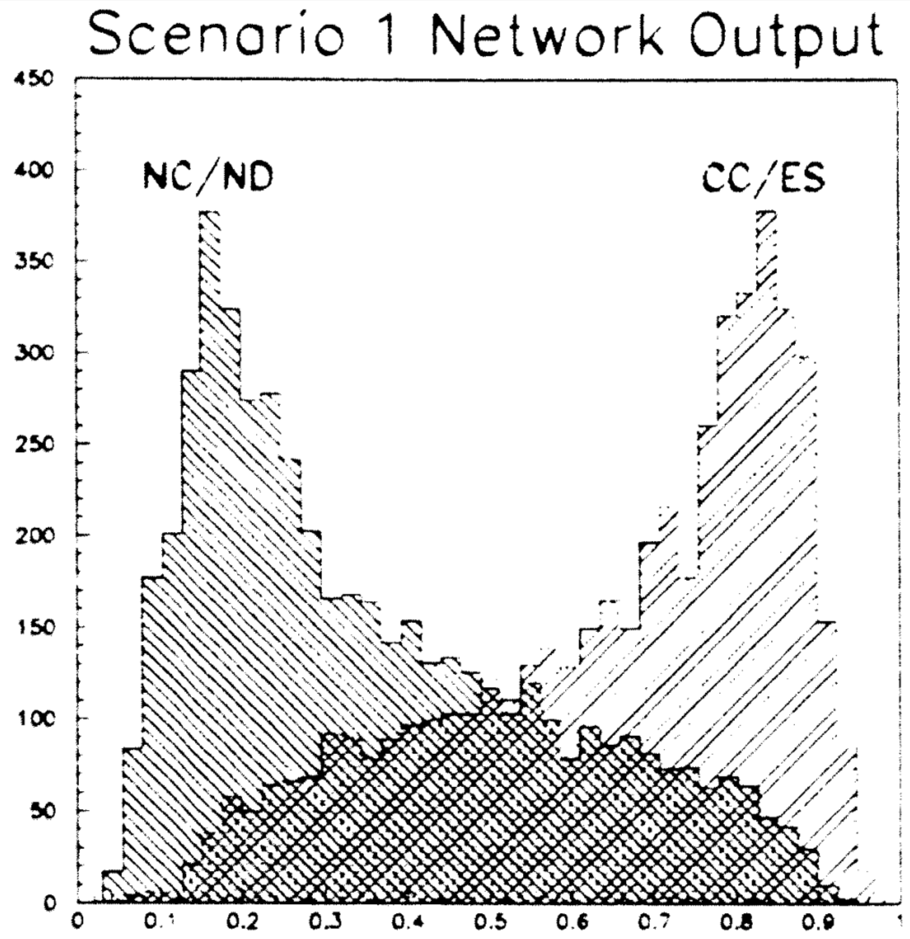
The inputs to the network are often extracted using traditional reconstruction methods.

Note, that they do not necessarily learn *why* certain inputs are correlated, just that they *are*.

Physics examples are things like decay kinematics.

The First Use of ANNs in Neutrino Physics

- Classified events using hit patterns.
- It did not achieve better separation than traditional methods.
- It showed that it was possible to do though, and lay the groundwork for these techniques to be explored.
- Modern analyses feature NNs across a wide range of applications.

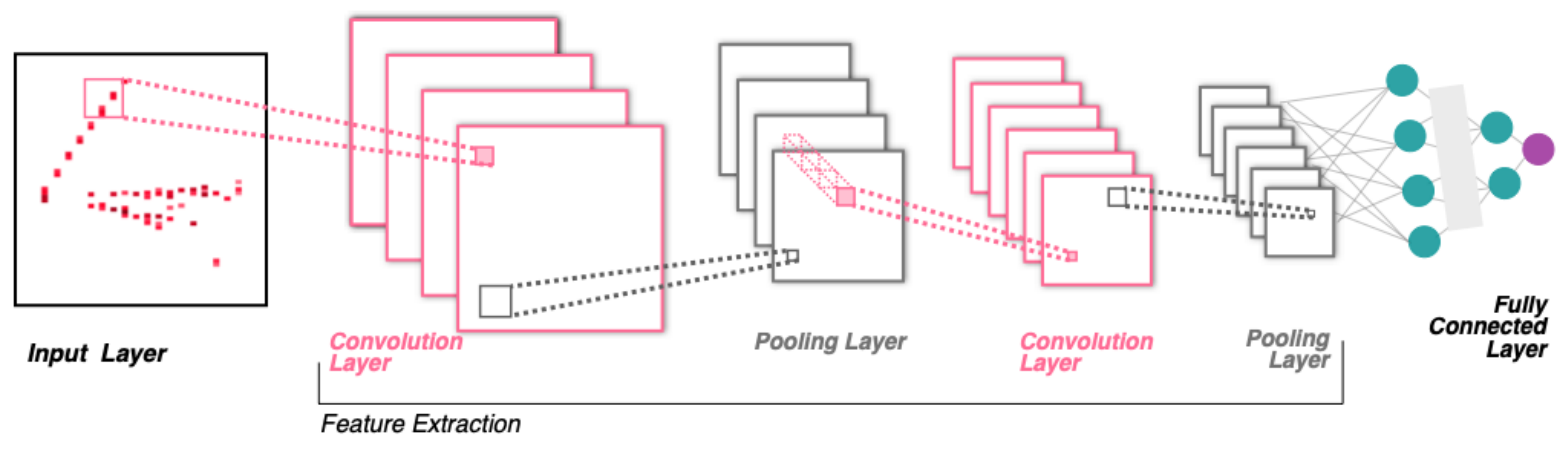


		Network Assignment	
		CC/ES	NC/ND
True Class	CC/ES	3809	1191
	NC/ND	1295	3705

Overall Purity $= (75.1 \pm 0.4)\%$

The SNO Experiments first use of a DNN in the 1990s.

Convolutional Neural Nets (CNNs)



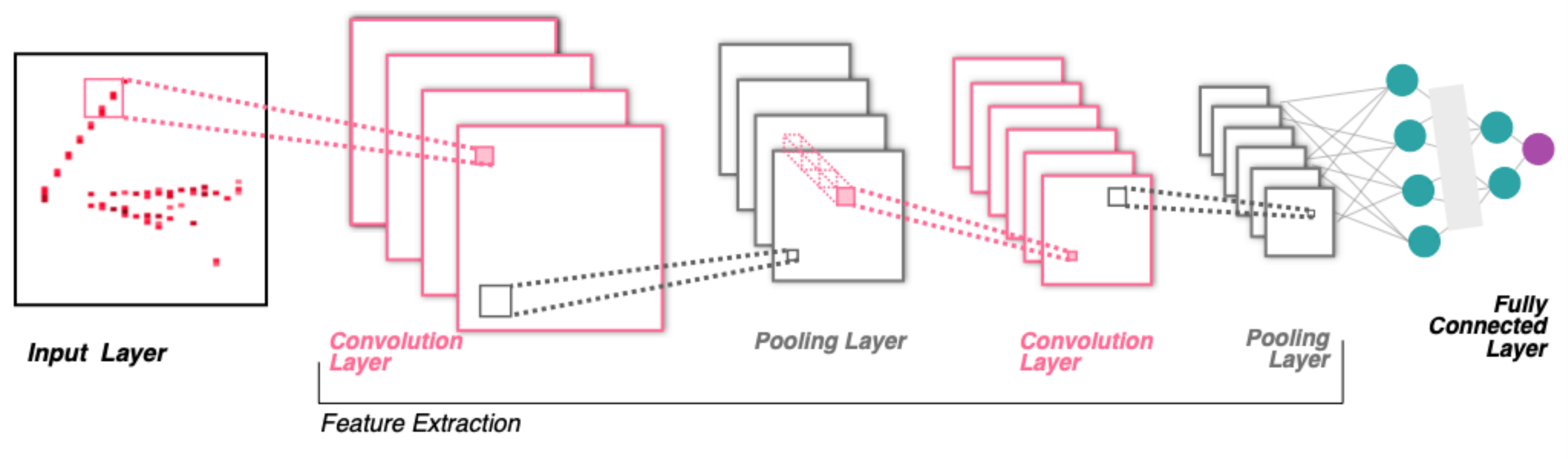
Premise: *Allow the network to extract features rather selecting them a-priori.*

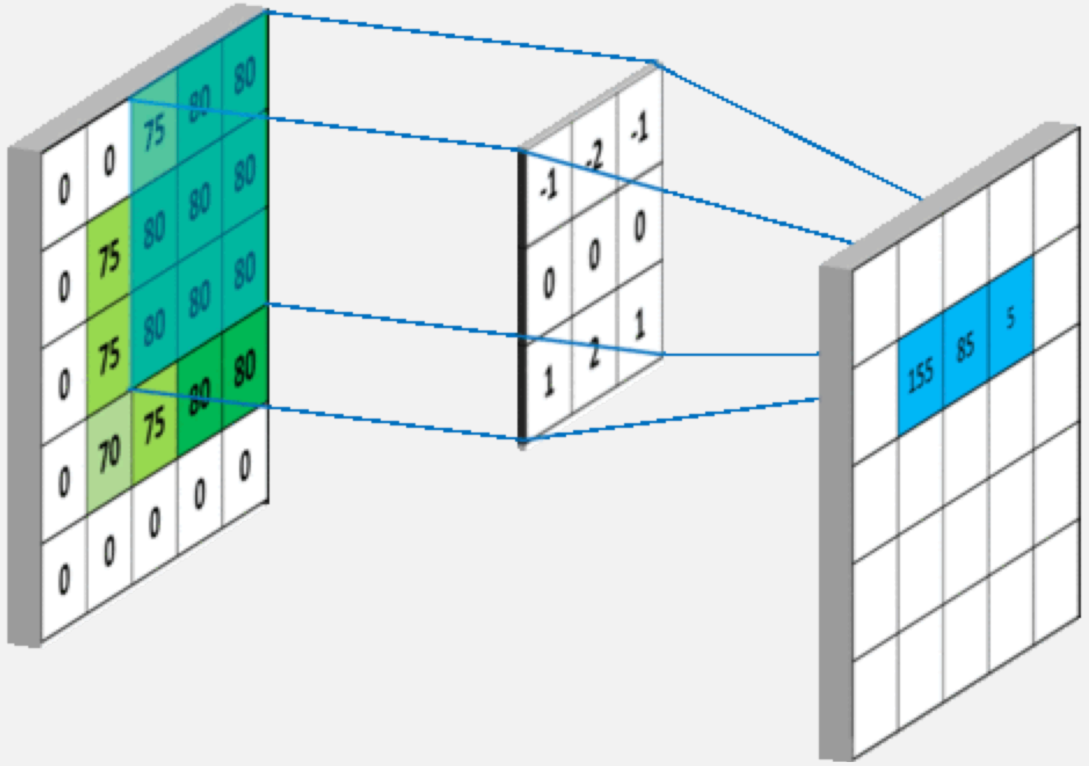
- Removes any biases which may be introduced from the traditional reconstruction algorithms.

In Practice: *Cast detector signals into maps and use CNNs to classify interactions in the style of image recognition.*

- Use image kernels to do this from 2D arrays.
- Traditionally use an image-to-RGB tensor strategy.

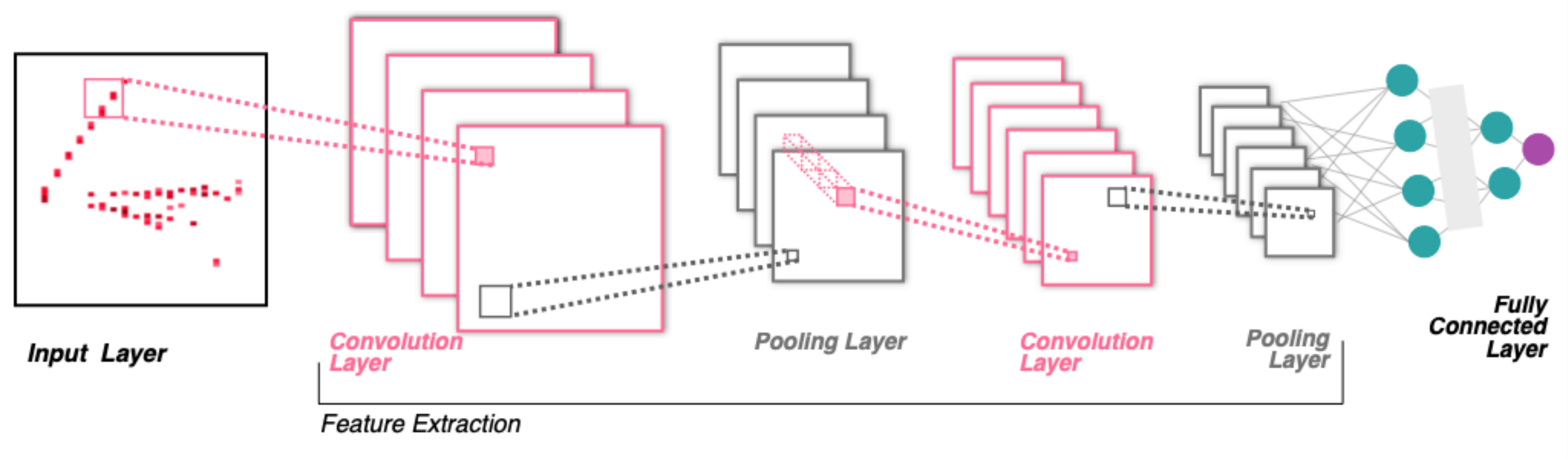
Convolutional Neural Nets (CNNs)

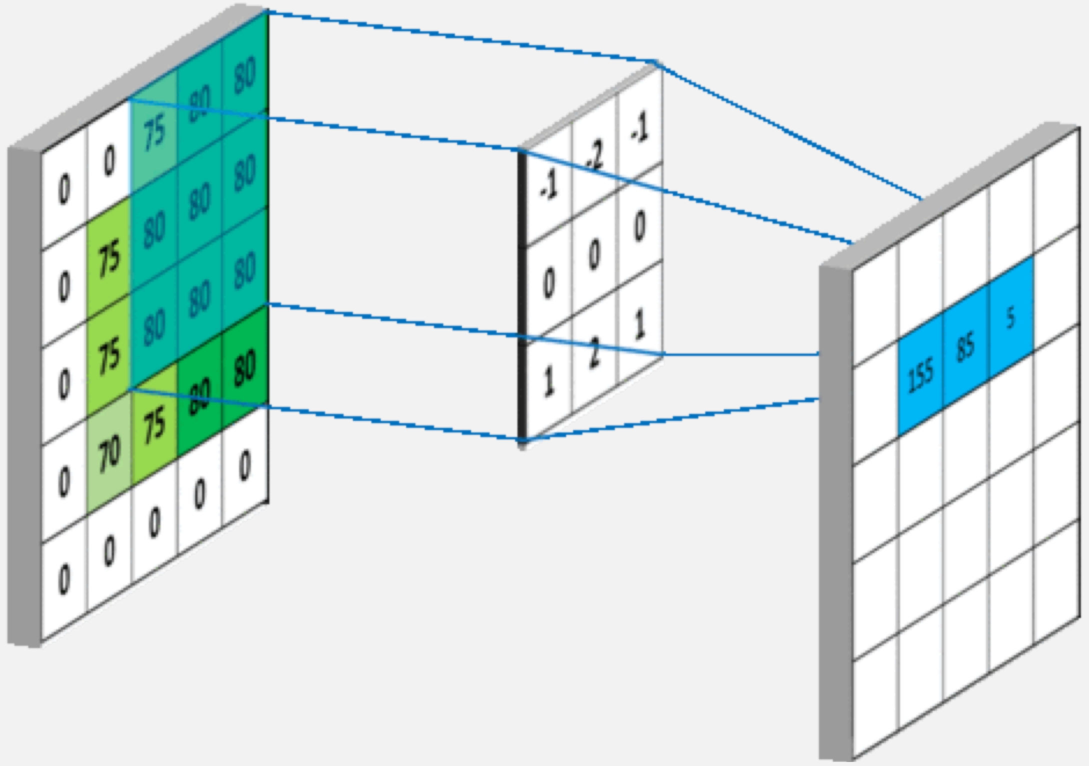




The convolution layers use image kernels for feature extraction.

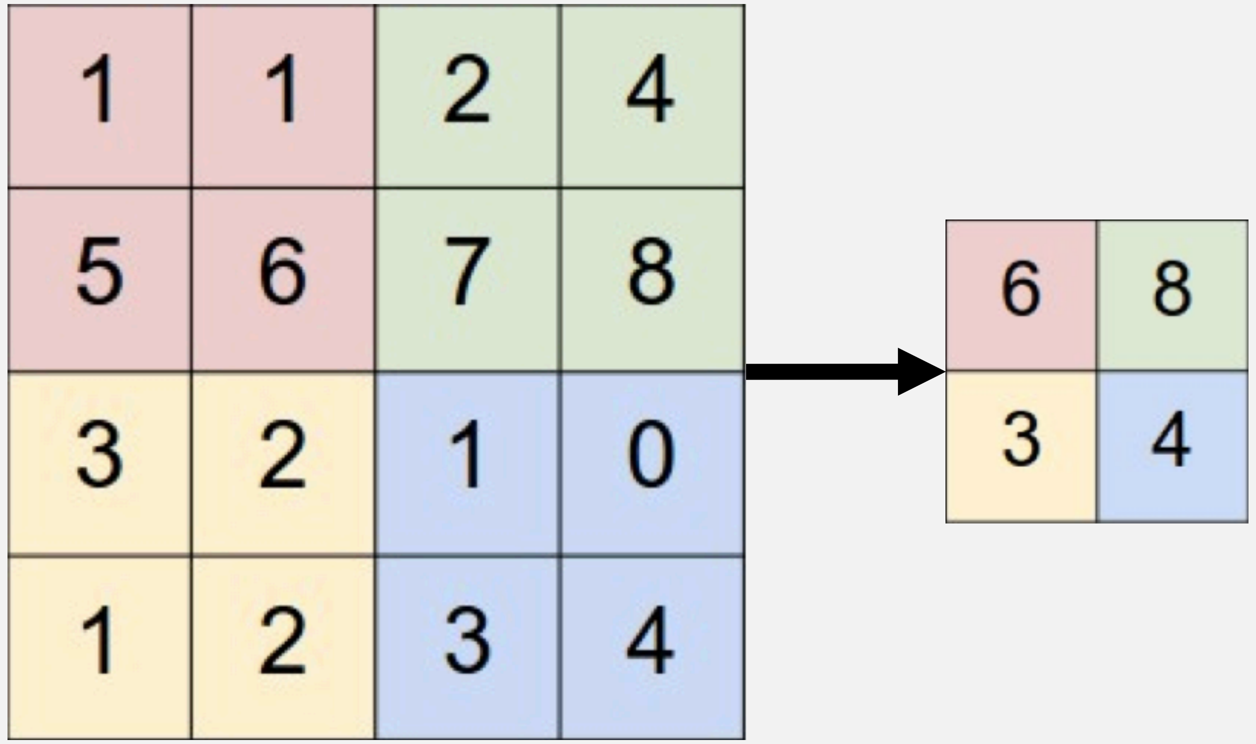
Convolutional Neural Nets (CNNs)



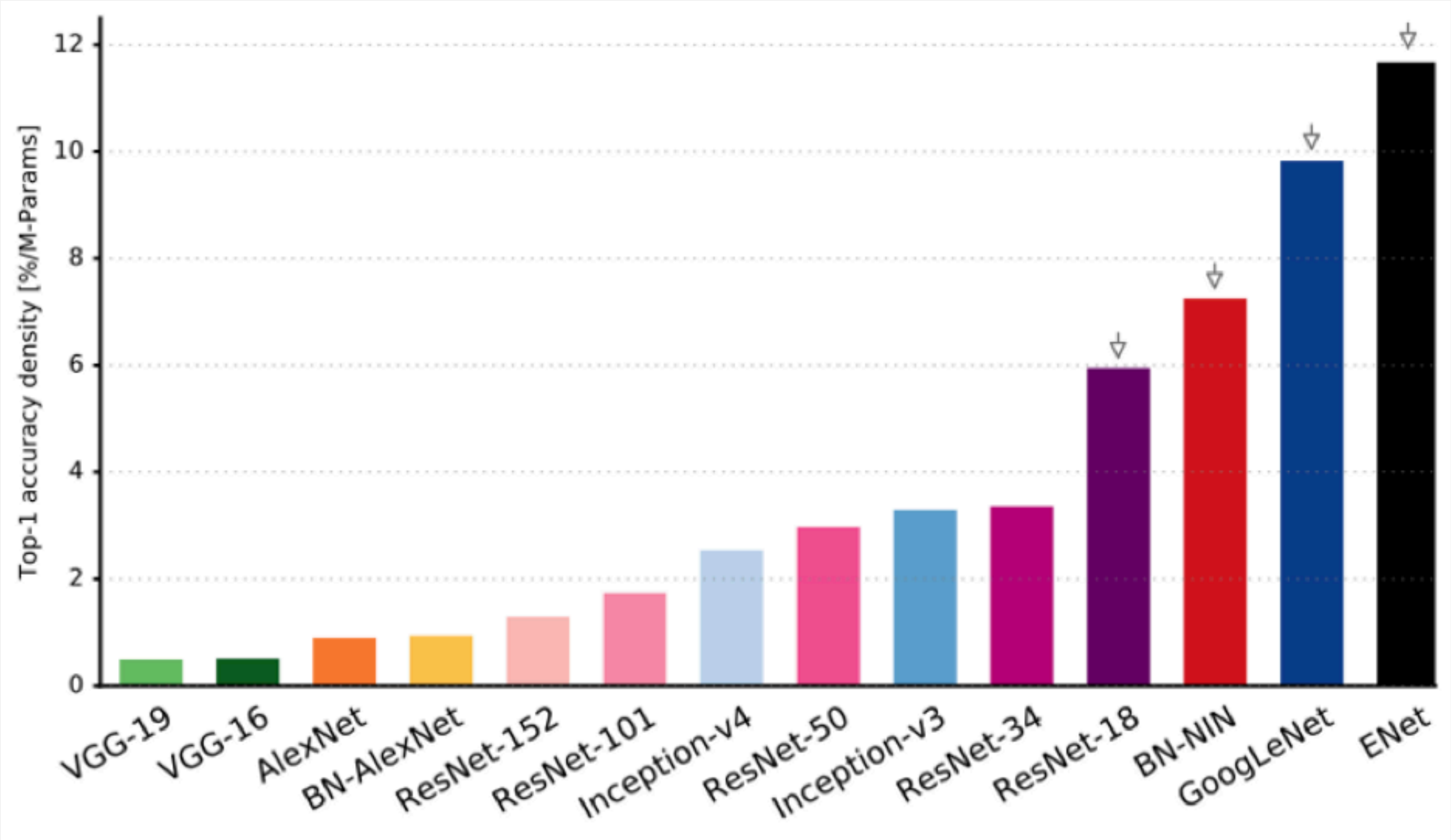


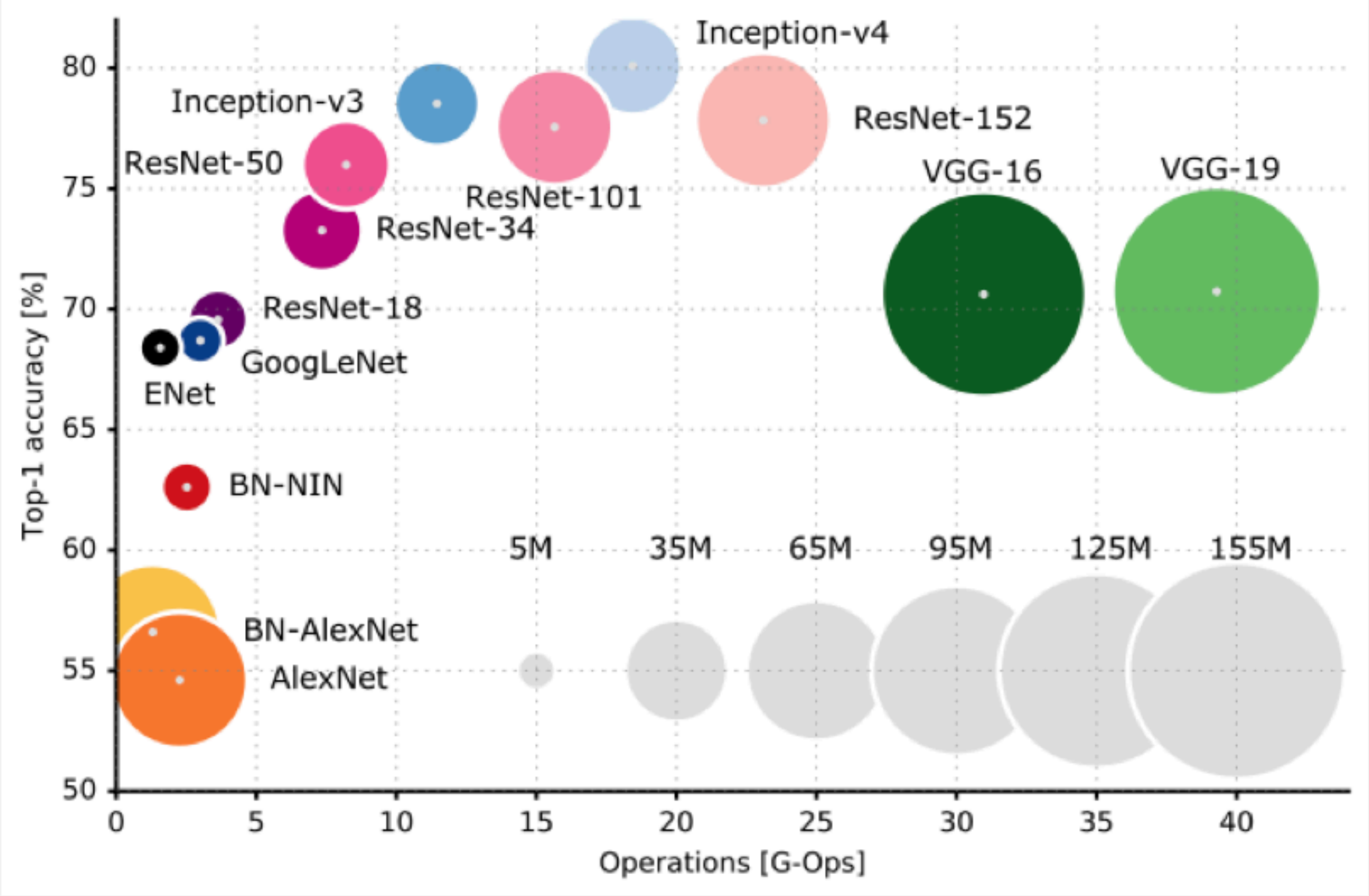
The convolution layers use image kernels for feature extraction.

The pooling layers down-sample the image, reducing computational cost and emphasizing features.



Why is Machine Learning Becoming Popular Now?





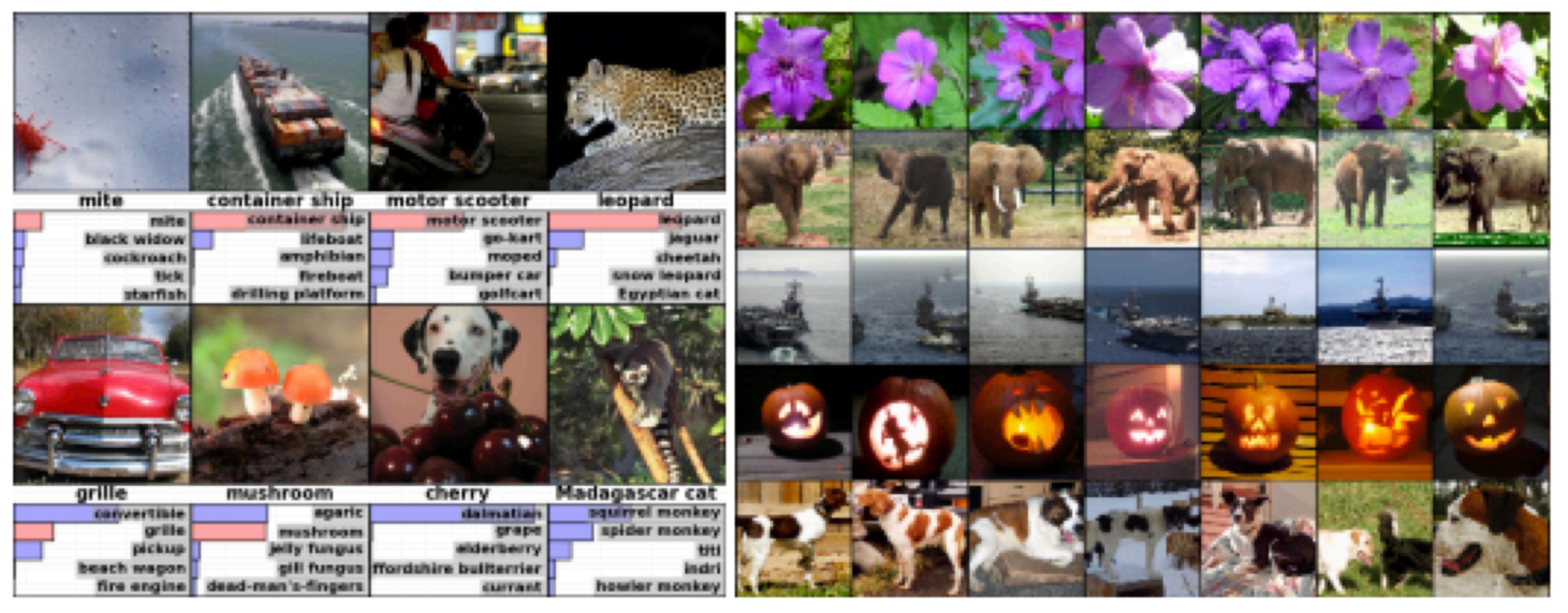
- A huge growth in computing power over the last decade.
  - Particularly GPU technology.
- At the same time, networks are becoming much more accurate, whilst requiring fewer computing resources.
- Modern experiments produce enormous datasets.
  - Some future experiments are ~30 PB per year!
- Applications from industry are applicable to many physics problem sets.

Challenges involved in applying Machine Learning in Neutrino Experiments

#1: The Adaptability of methods.

GoogLeNet and The First Applications of CNNs

- Developed for the 2014 ImageNet Challenge (ILSVRC 2014).
- The first creatively non-sequential implementation of convolutional layers in CNNs.
  - Had significantly higher accuracy and performance improvements compared to its competitors.
- Neutrino experiments attempted to use it with few modifications.



Going Deeper with Convolutions

Christian Szegedy[1], Wei Liu[2], Yangqing Jia[1], Pierre Sermanet[1], Scott Reed[3],
Dragomir Anguelov[1], Dumitru Erhan[1], Vincent Vanhoucke[1], Andrew Rabinovich[4]
[1]Google Inc. [2]University of North Carolina, Chapel Hill
[3]University of Michigan, Ann Arbor [4]Magic Leap Inc.
[1]{szegedy,jiayq,sermanet,dragomir,dumitru,vanhoucke}@google.com
[2]wliu@cs.unc.edu, [3]reedscott@umich.edu, [4]arabinovich@magicleap.com

Abstract

We propose a deep convolutional neural network architecture codenamed Inception that achieves the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main hallmark of this architecture is the improved utilization of the computing resources inside the network. By a carefully crafted design, we increased the depth and width of the network while keeping the computational budget constant. To optimize quality, the architectural decisions were based on the Hebbian principle and the intuition of multi-scale processing. One particular incarnation used in our submission for ILSVRC14 is called GoogLeNet, a 22 layers deep network, the quality of which is assessed in the context of classification and detection.

1. Introduction

In the last three years, our object classification and detection capabilities have dramatically improved due to advances in deep learning and convolutional networks [10]. One encouraging news is that most of this progress is not just the result of more powerful hardware, larger datasets and bigger models, but mainly a consequence of new ideas, algorithms and improved network architectures. No new data sources were used, for example, by the top entries in the ILSVRC 2014 competition besides the classification dataset of the same competition for detection purposes. Our GoogLeNet submission to ILSVRC 2014 actually uses 12 times fewer parameters than the winning architecture of Krizhevsky et al [9] from two years ago, while being significantly more accurate. On the object detection front, the biggest gains have not come from naive application of bigger and bigger deep networks, but from the synergy of deep architectures and classical computer vision, like the R-CNN algorithm by Girshick et al [6].

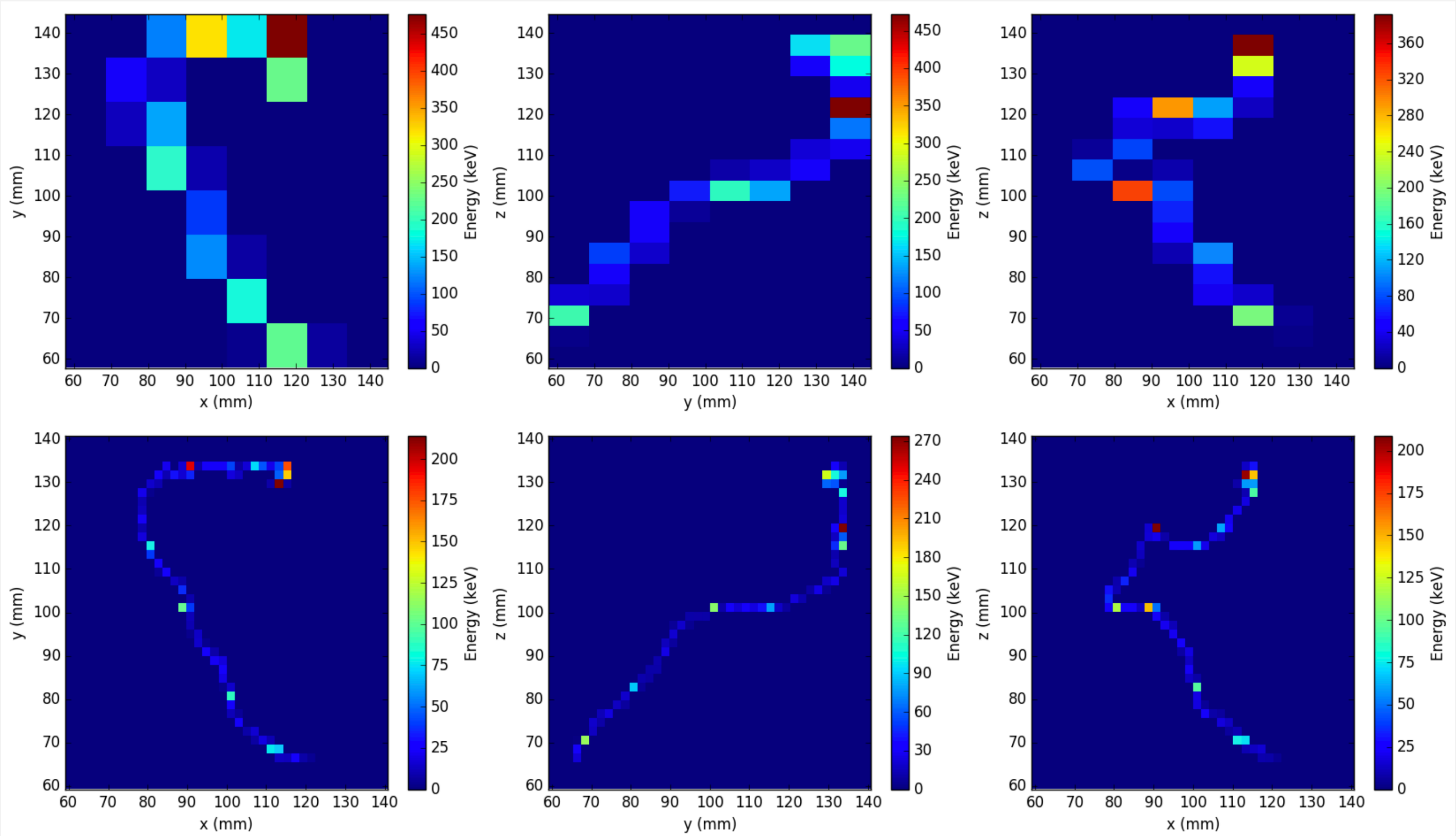
Another notable factor is that with the ongoing traction of mobile and embedded computing, the efficiency of our algorithms – especially their power and memory use – gains importance. It is noteworthy that the considerations leading to the design of the deep architecture presented in this paper included this factor rather than having a sheer fixation on accuracy numbers. For most of the experiments, the models were designed to keep a computational budget of 1.5 billion multiply-adds at inference time, so that the they do not end up to be a purely academic curiosity, but could be put to real world use, even on large datasets, at a reasonable cost.

In this paper, we will focus on an efficient deep neural network architecture for computer vision, codenamed Inception, which derives its name from the Network in network paper by Lin et al [12] in conjunction with the famous "we need to go deeper" internet meme [1]. In our case, the word "deep" is used in two different meanings: first of all, in the sense that we introduce a new level of organization in the form of the "Inception module" and also in the more direct sense of increased network depth. In general, one can view the Inception model as a logical culmination of [12] while taking inspiration and guidance from the theoretical work by Arora et al [2]. The benefits of the architecture are experimentally verified on the ILSVRC 2014 classification and detection challenges, where it significantly outperforms the current state of the art.

2. Related Work

Starting with LeNet-5 [10], convolutional neural networks (CNN) have typically had a standard structure – stacked convolutional layers (optionally followed by con-

GoogLeNet and The First Applications of CNNs – NEXT



Pixel maps used by the NEXT experiment.

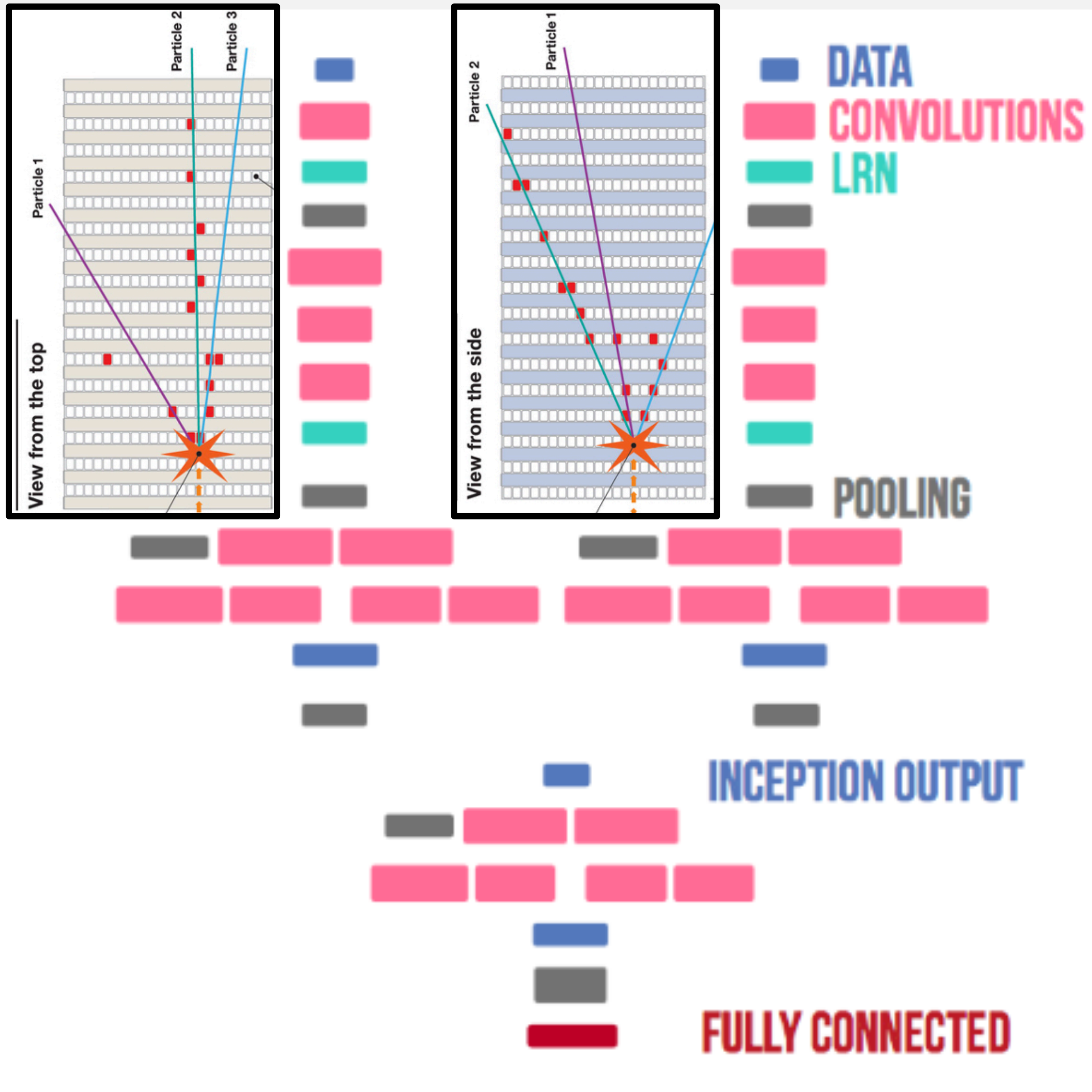
Top, a coarse voxelation (10 mm voxels) where structure at the end of the track is lost.

Bottom, a fine voxelation (2 mm voxels) where this structure is still visible.

- A network designed to perform background rejection for neutrinoless double beta decay.
  - Equal numbers of signal and background events are used in training.
- 2D projections of the detector readout (XY, YZ, XZ) are used as the RGB input for the network.
- Outperforms traditional reconstruction by between 20% and 60%.

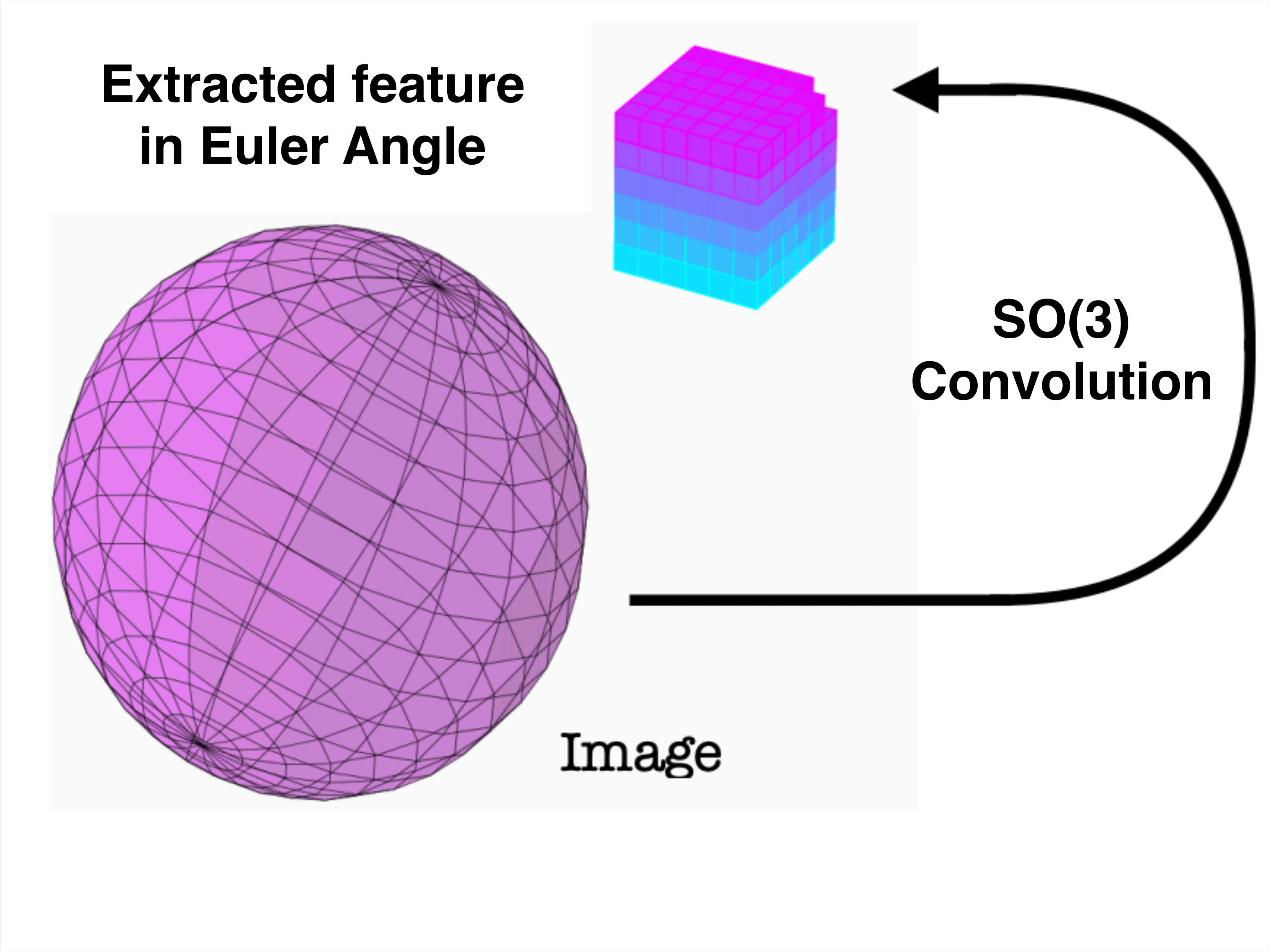
GoogLeNet and The First Applications of CNNs – NOvA

- Network designed to perform interaction classification.
  - Subsequently extended to perform the identification of individual particles.
- The detector has two decoupled views, meaning that they cannot be combined into a single RGB tensor.
  - Therefore employs a Siamese architecture, with an input for each view.
- When first used in 2017, it increased effective exposure by 30%.
  - Was the first CNN to be used in a published particle physics result.
- Found that training sign-dependant networks increased effectiveness.



$\bar{\nu}$ **Efficiency Improvement** Training Sample (ID $>$ 0.9)		
$\bar{\nu}_e$ CC Signal	$\bar{\nu}_\mu$ CC Signal	$\bar{\nu}$ NC Signal
14%	6%	10%

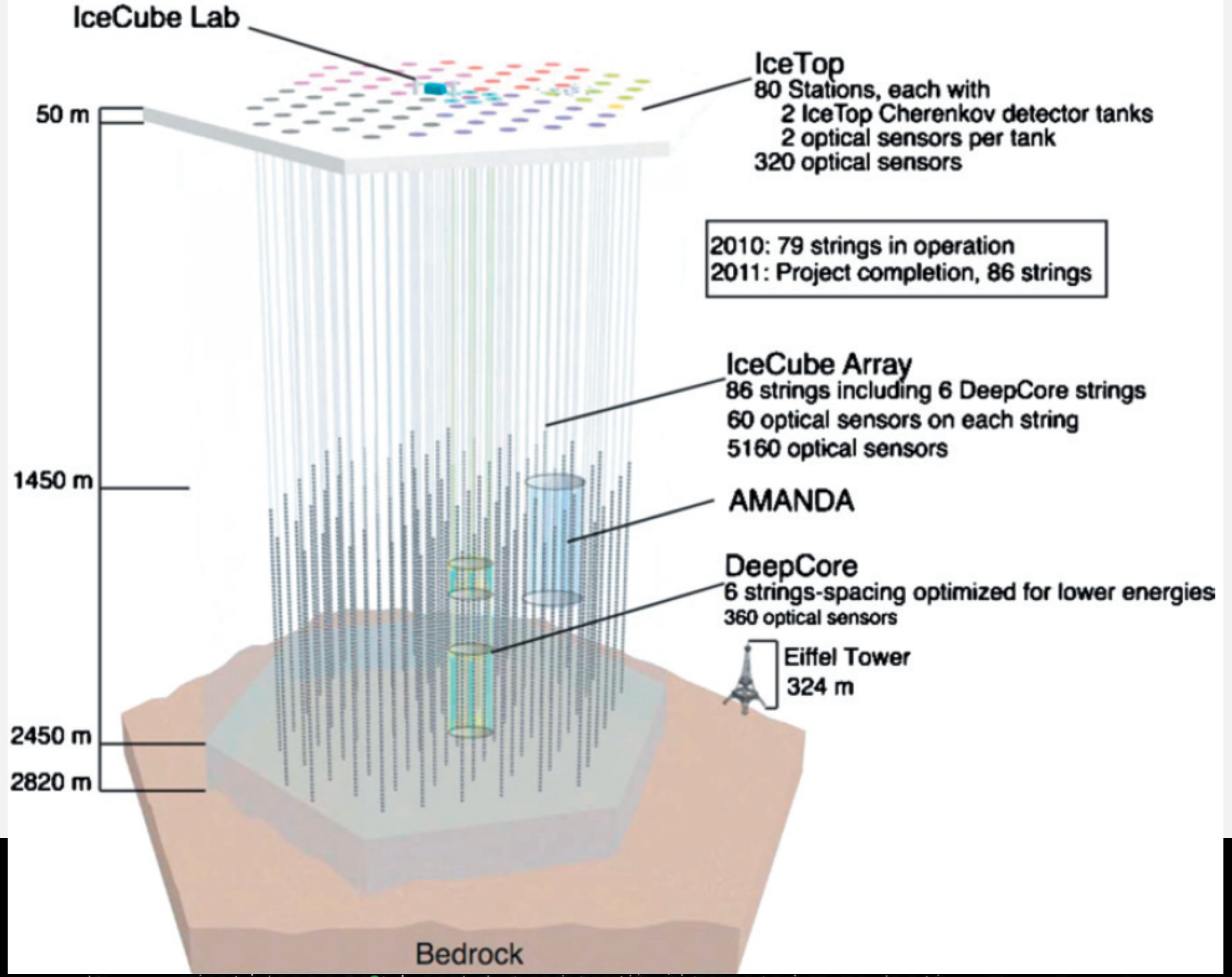
Spherical CNNs – Kamland-Zen

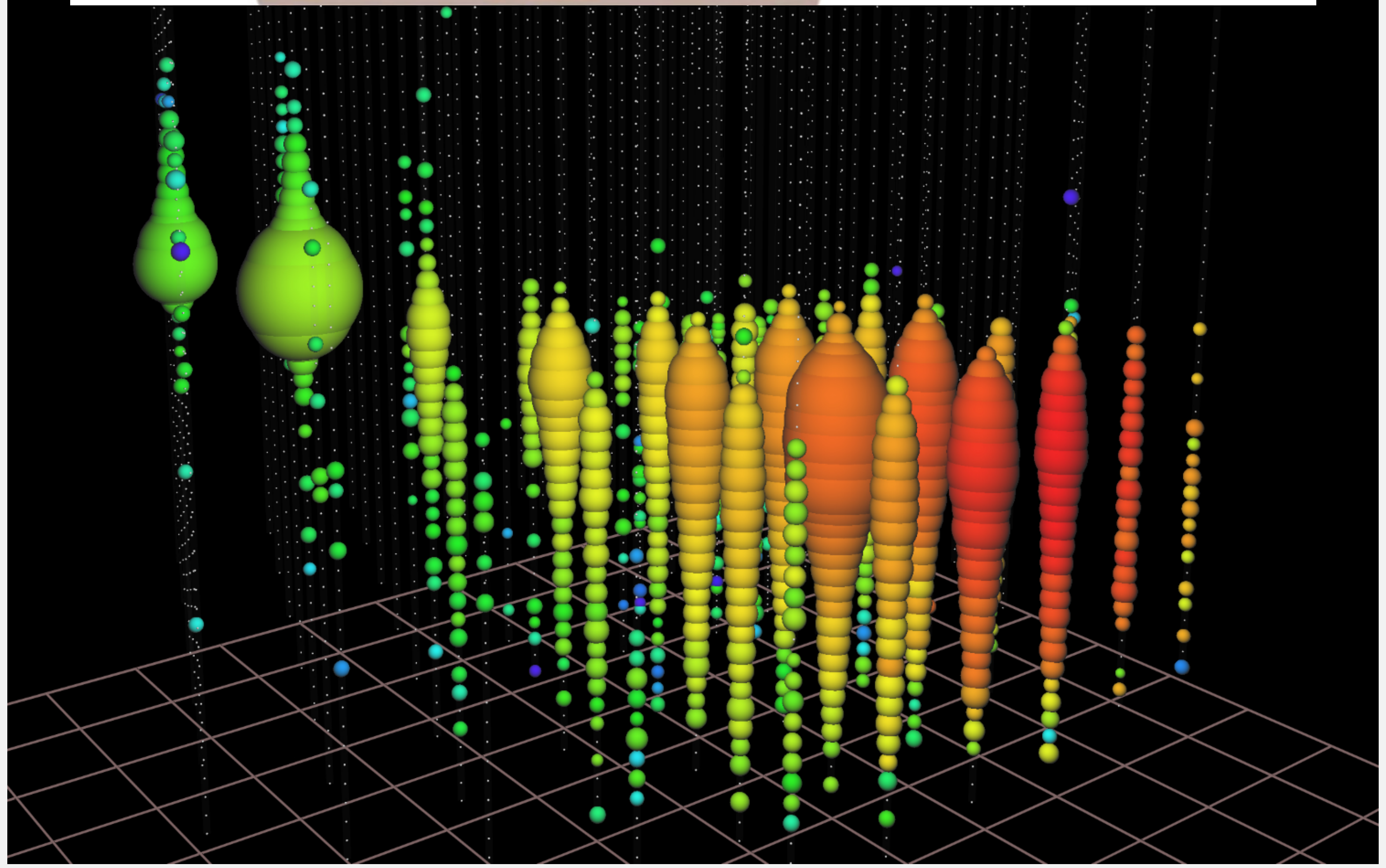


- Kamland-Zen is a near-spherical detector.
  - Translating that into a 2D space, can cause distortions.
- As such, the use of spherical CNNs is required.
- The kernel covers the whole phase-space by scanning Euler angles, avoiding such distortions.
- A Spherical network achieves better background than their original network (71% vs 61%).

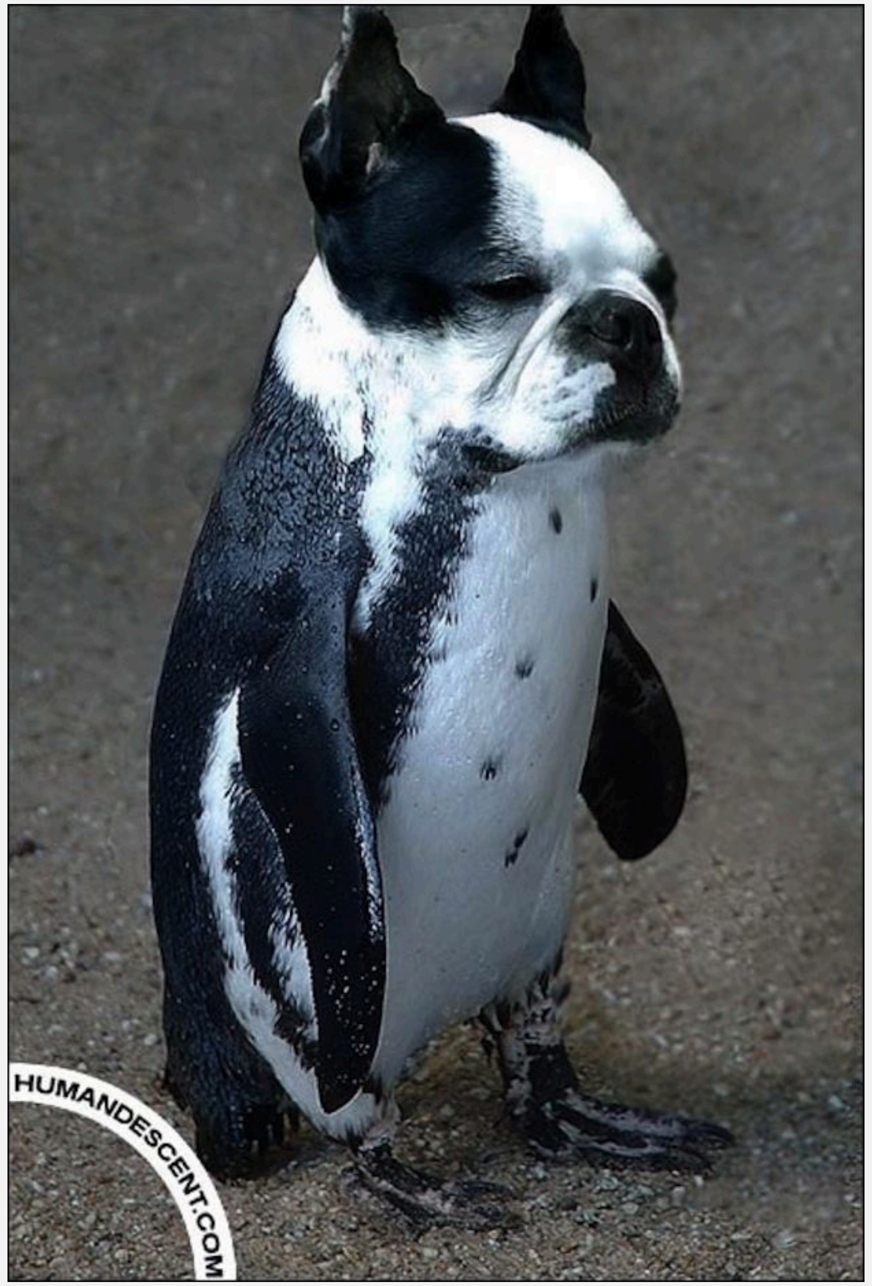
Graphical Neural Nets – IceCube

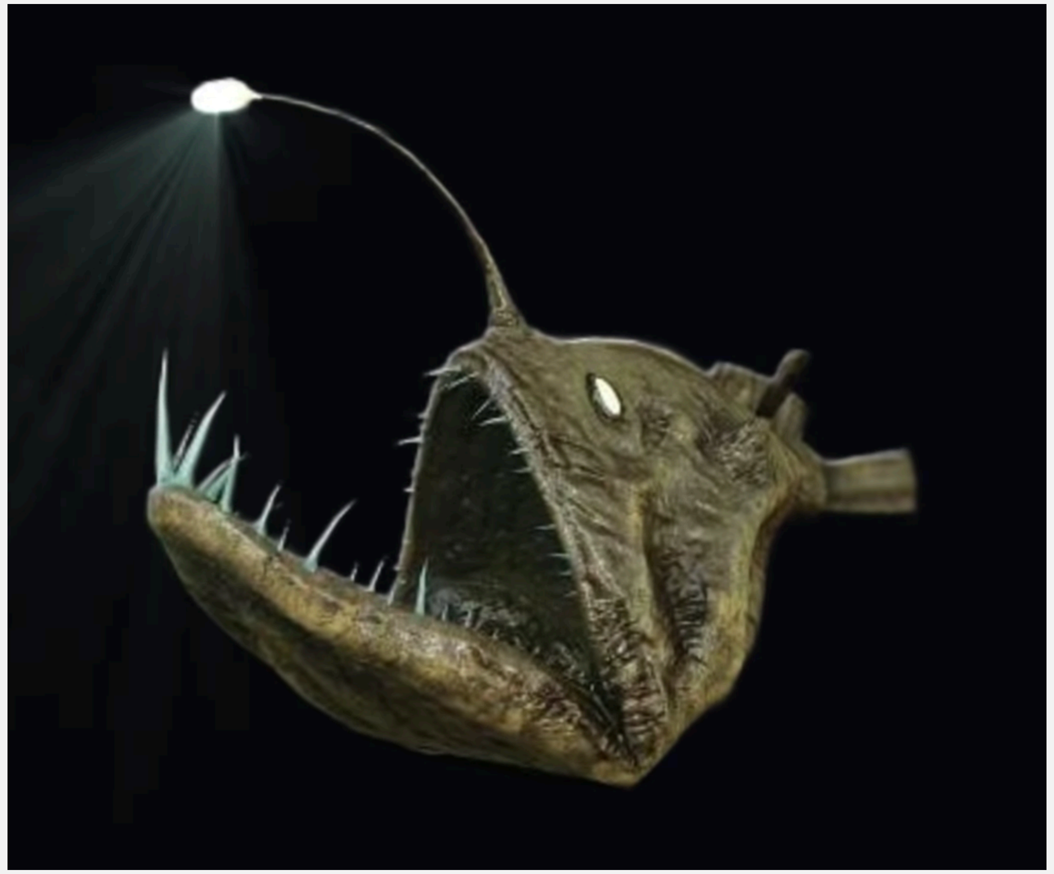
- IceCube is a very non-uniform detector which takes data that is very sparse.
  - It is therefore not well suited to CNNs.
- GNNs are designed to classify graphs, where the nodes define an element of the detector, and the edges show connections between elements.
  - Ideally places to mitigate the difficult aspects of data of IceCube.
- Developed a GNN to separate neutrino and cosmic induced events;
  - Identifies 630% more signal events than a CNN/traditional algorithm with a Signal-to-Noise ratio which is 3 times larger.





Challenge #2: Quantifying Network Bias and Uncertainties





Humans know features which animals do and don't have.

We need to make sure that our algorithms do too.

The Dangers…

https://arxiv.org/pdf/1807.04975.pdf

(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

The composition of training samples can constrain network performance by containing in-built model assumptions which may bias the results.

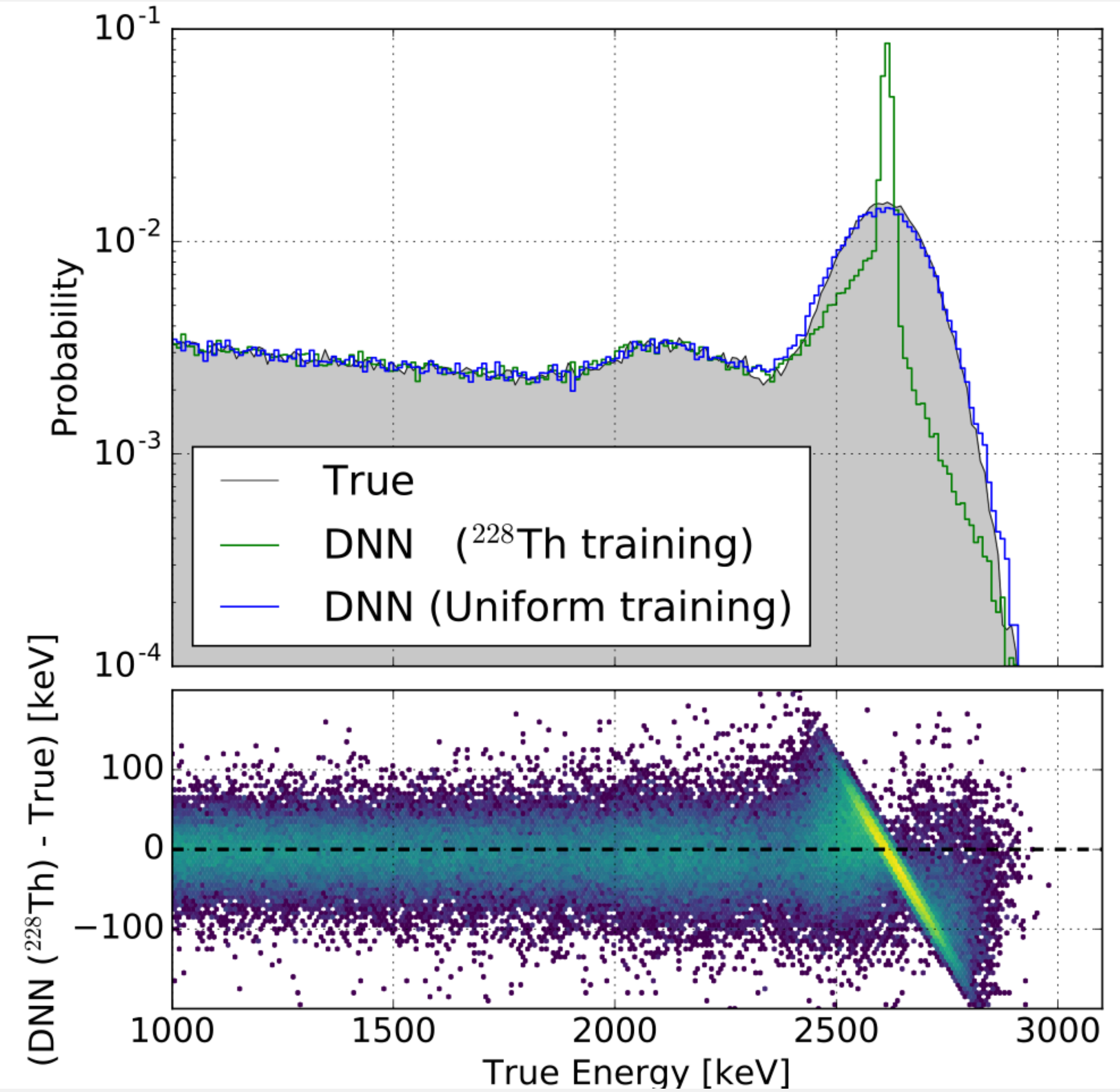
- Model bias is a problem for all Machine Learning.
  - As the extracted features are abstract, CNNs are particularly susceptible to underlying model bias though.
- Models are trained using simulated data;
  - Assumptions are made in generating such data on both the model and the detector performance.
  - Assumptions are made in selecting training dataset.
  - Simulation will never fully reproduce real data.
- Though this is a well known problem, no standard and complete techniques exist to address it.

Carefully Constructed Training Datasets – EXO-200.

- A network designed to perform charge-only energy reconstruction which was initially trained on a specific calibration source – ^{228}Th.

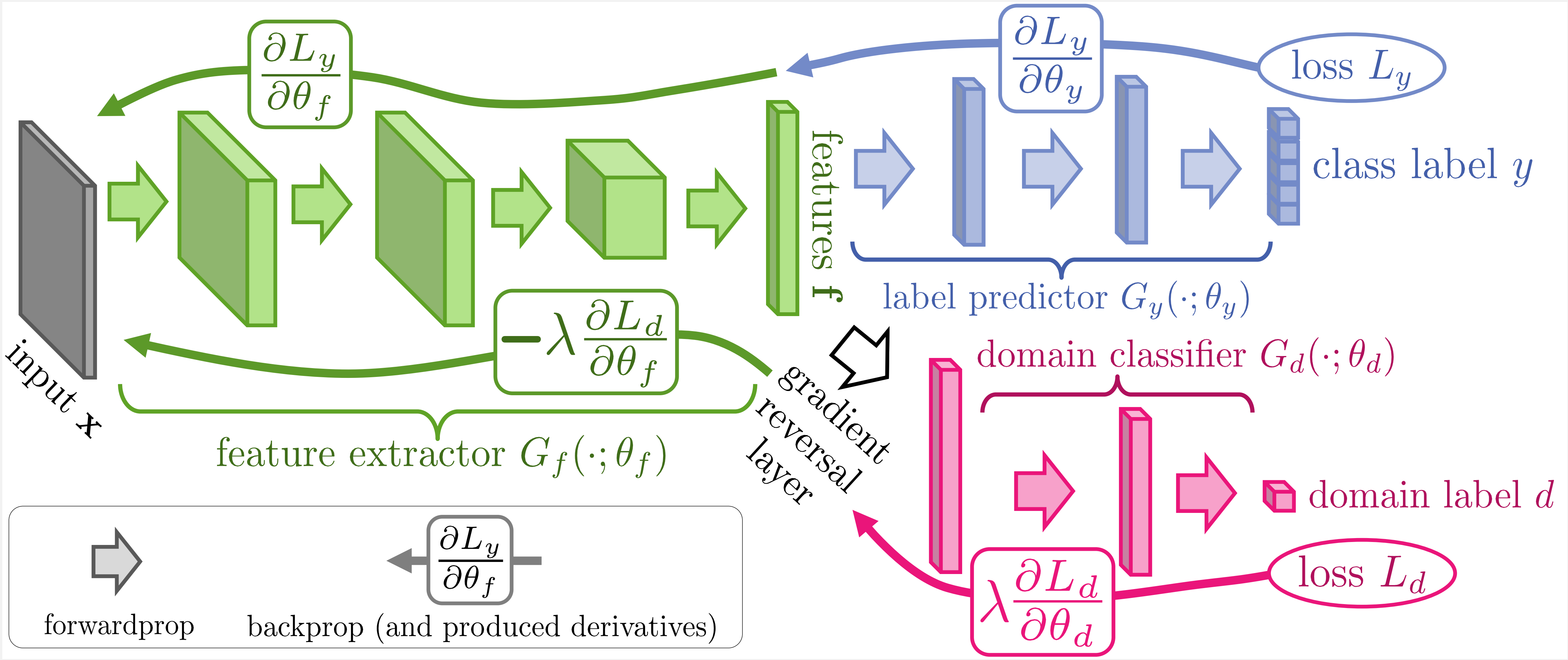
- A systematic study found an unusually large improvement in the E_{res} for events in the ^{208}Tl peak.

- Training on calibration data using a gamma-ray source located in the centre of the detector removed this unusually large improvement.
  - Studies repeated using a range of calibration sources at various locations yielded consistent results.



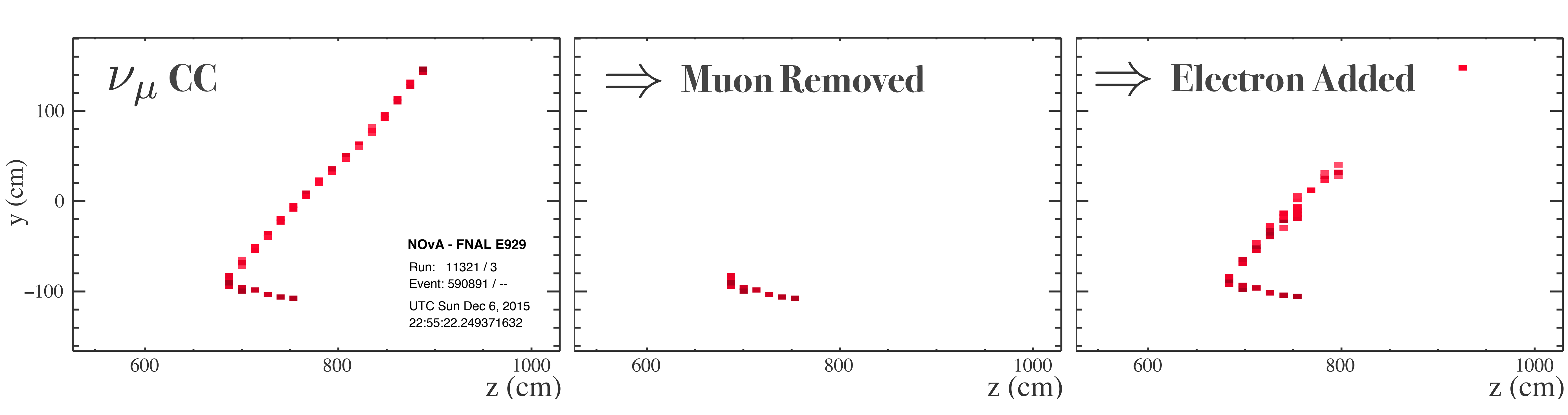
An unusually large improvement in the Energy Resolution is seen in the green line, which is not present in the Uniform Training.

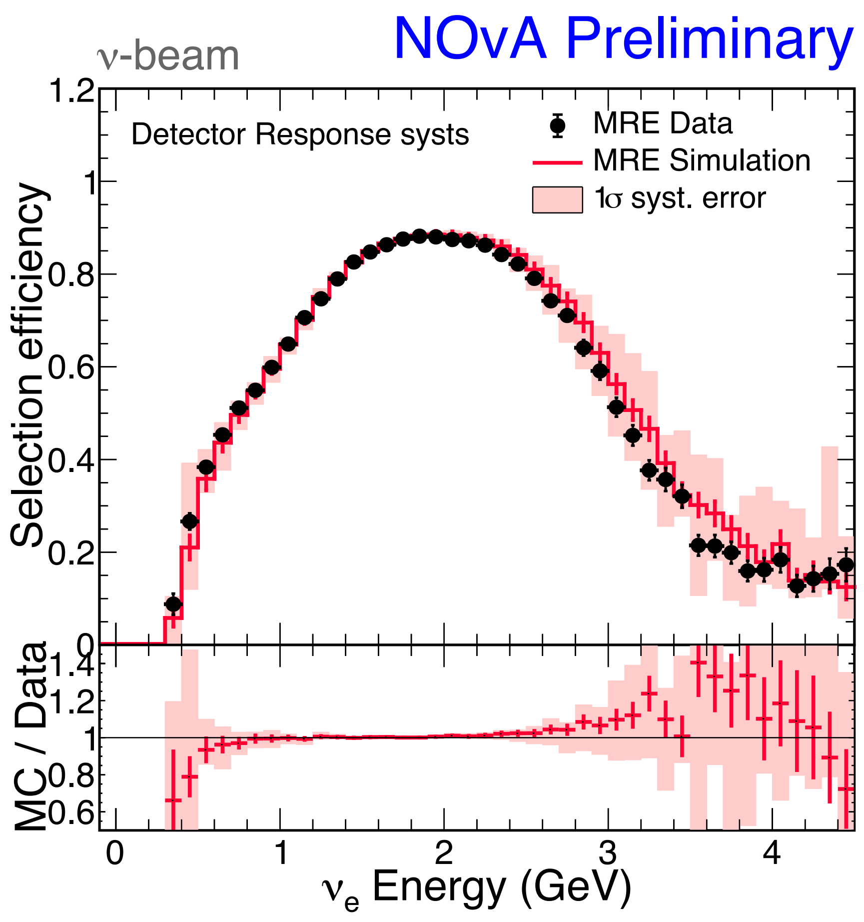
Domain Adversarial Neural Networks – MINERvA



- It is difficult to correct for, or quantify unknown biases.
  - It is possible to minimize their impacts though.
- The DANN used in MINERvA is a prime example.
  - A main network perform classification, whilst a second sub-network is incorporated for bias reduction.
- The domain sub-network incorporates real data into the training to identify simulation/data differences.
- The gradient reversal layer discourages the classification network from learning any of the differences between the domains.

Quantifying Bias in Simulation vs Data – NOvA





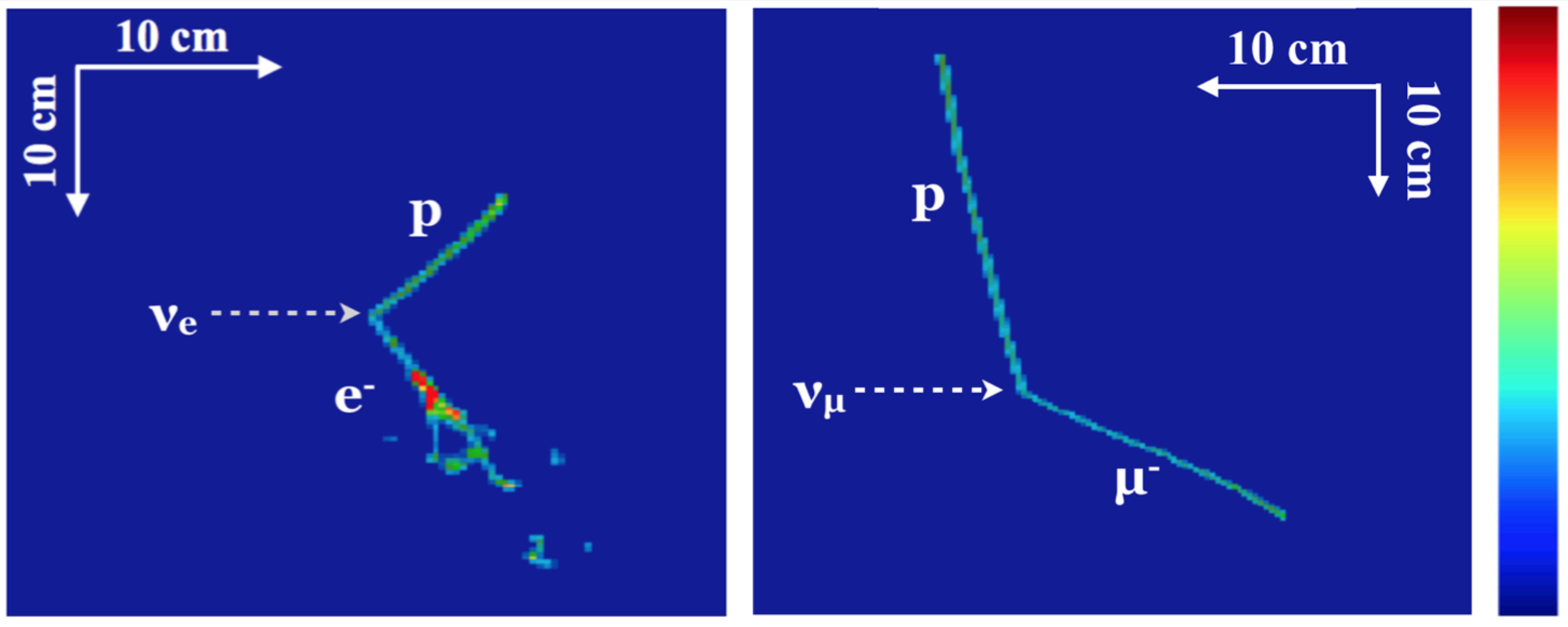
Select a Muon Neutrino interaction from Data/Monte Carlo.

Remove the hits associated with the muon.

Simulate an electron with the same energy as the removed muon.

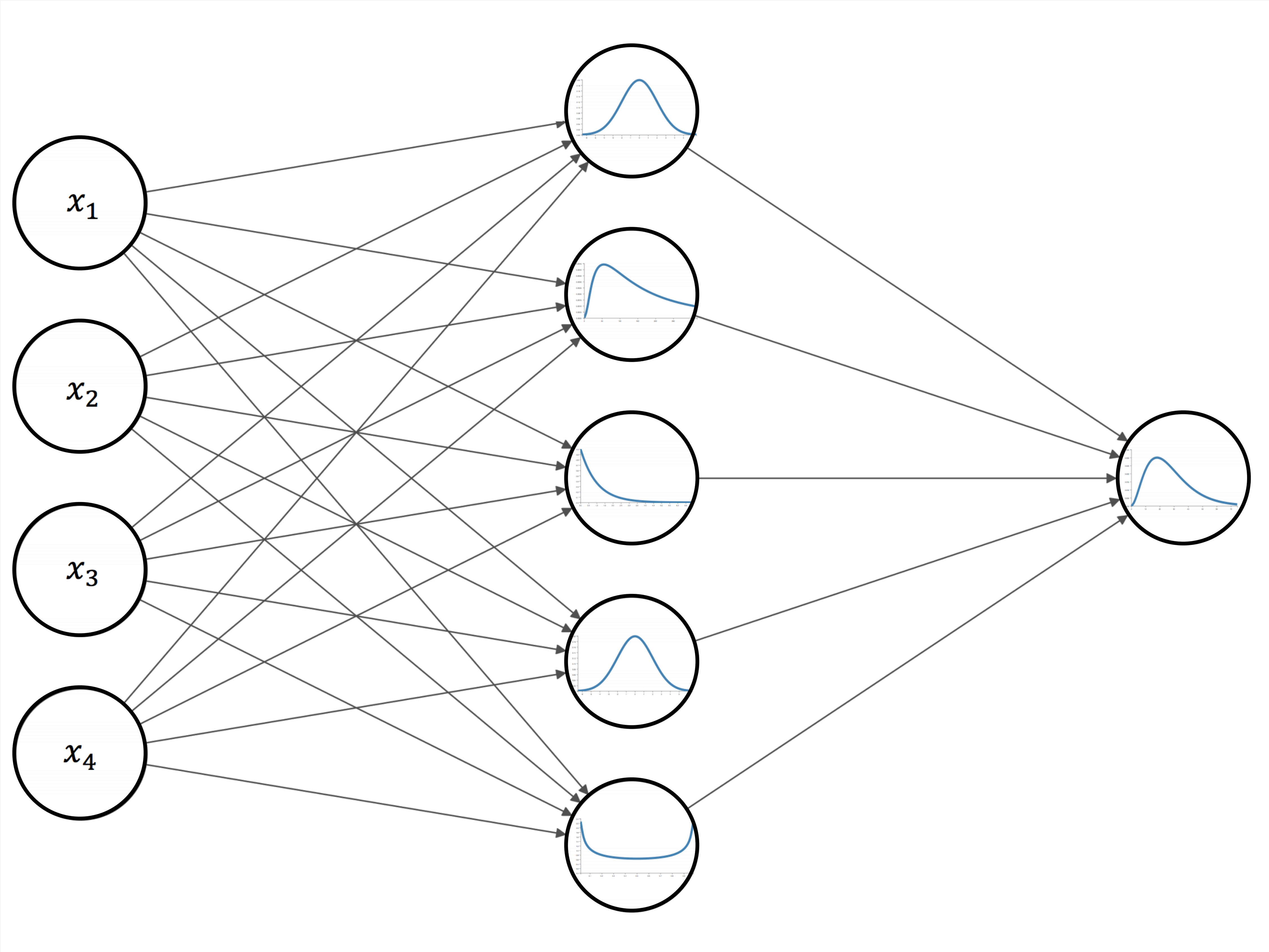
- A process of Data Augmentation, known as Muon Removed, Electron Added seeks to quantify bias in NOvA.
  - Good agreement is found between comparisons of augmented data and simulation.
- A second process aims to further quantify bias, by studying muons from cosmic rays which decay in the detectors.

Human Labelled Datasets – MicroBooNE



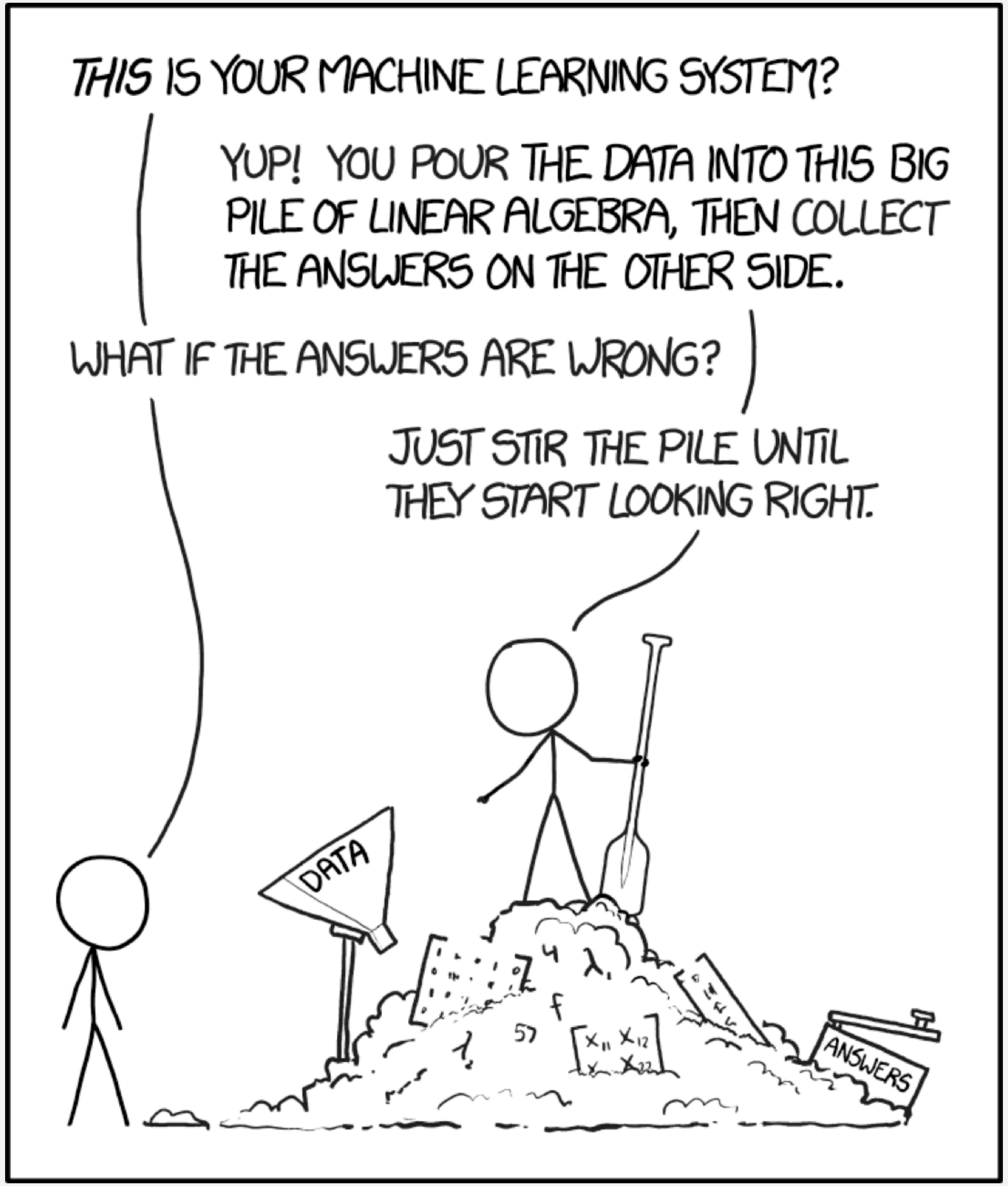
- Impossible to know the true identity of real data, though humans can often correctly identify it.
- Interestingly, when given identical datasets there will be differences between human and NNs.
  - Suggests unknown biases between humans to NNs.
- MicroBooNE created a human-labelled dataset to validate a semantic-segmentation network.
  - Trained on simulated events, with 5+ particles originating from a common vertex.
- Humans and the NN disagreement: ~2% of particles.

Bayesian Neural Networks



- Replace fixed value weights with probability distribution functions.
- The output is thus a probabilistic function which can be interpreted as most probable value.
- Thus able to convey a scale of uncertainty related to predictions which are outside of the scale of the training dataset.

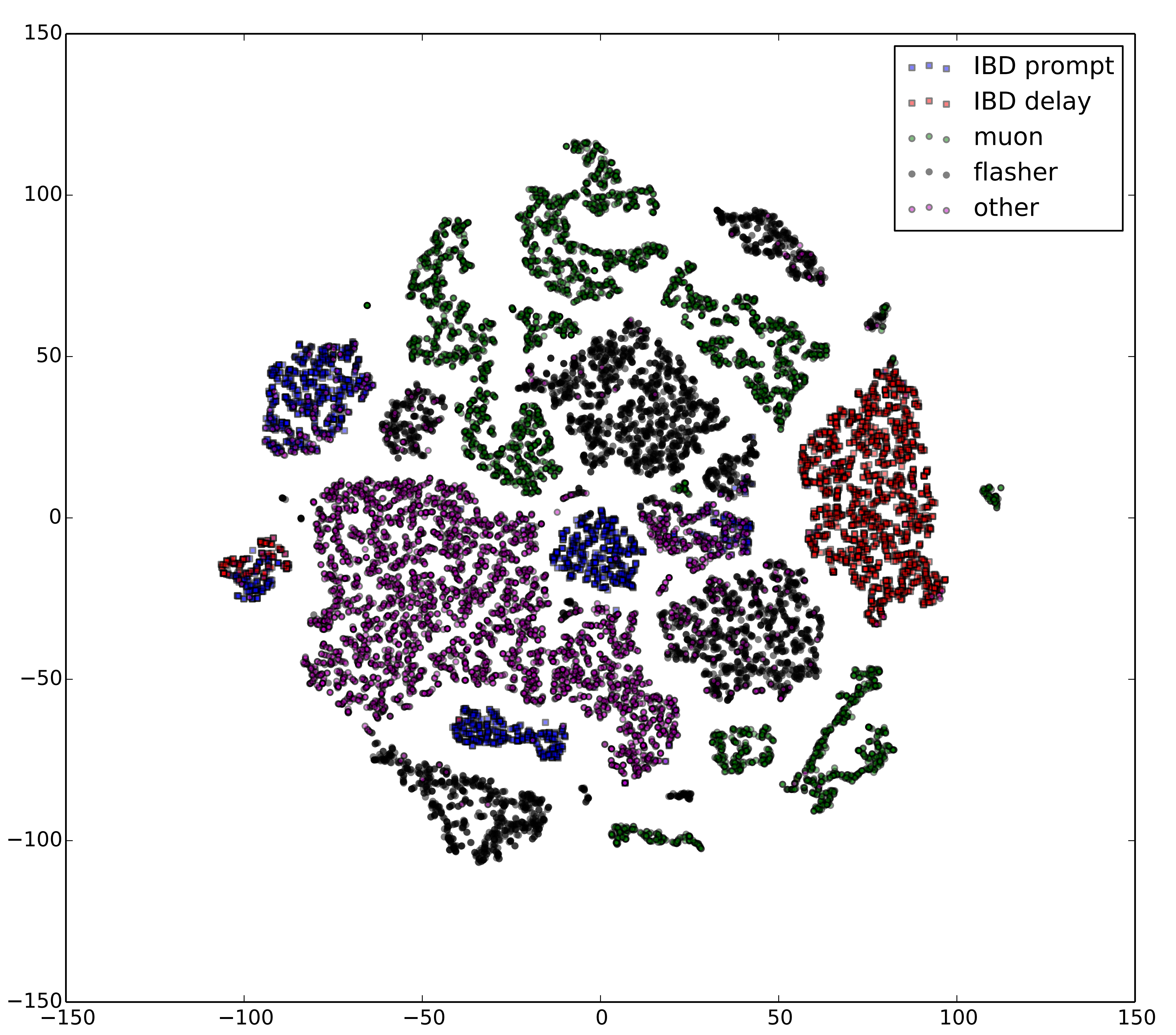
Challenge #3: Network Interpretability



- Deeper networks are more difficult to interpret.
  - The exact behaviour of individual kernels and how they combine to form the weights of a CNN is almost impossible to deduce.
  - This is much easier to do with boosted-decision-trees.
- Particularly troublesome for physicists who want to relate network features to underlying physics phenomena.
  - Knowing this could minimize or correct network inefficiencies.
  - Could also hint at specific features which are important to train against.

t-Distributed Stochastic Neighbour Embedding

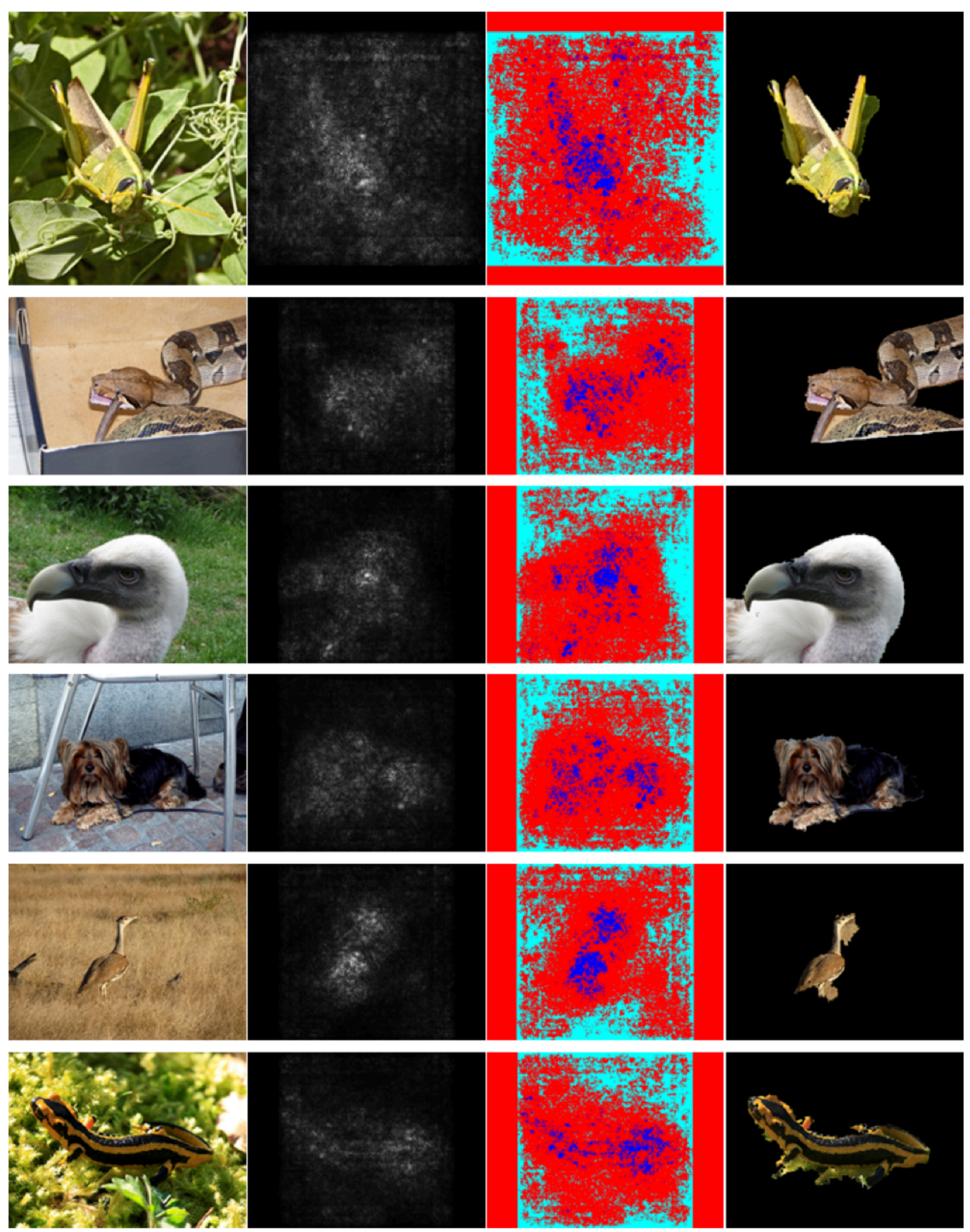
- Widely used technique to interpret features, and establish a proxy for classification separation.
- Network features are transformed down to 2D.
  - Uses a non-linear transformation to do this.
  - Preserves separation between points, but with lower dimensionality.
- A t-SNE from the Daya Bay Experiment to separate anti-neutrinos from nearby reactors with experimental backgrounds.



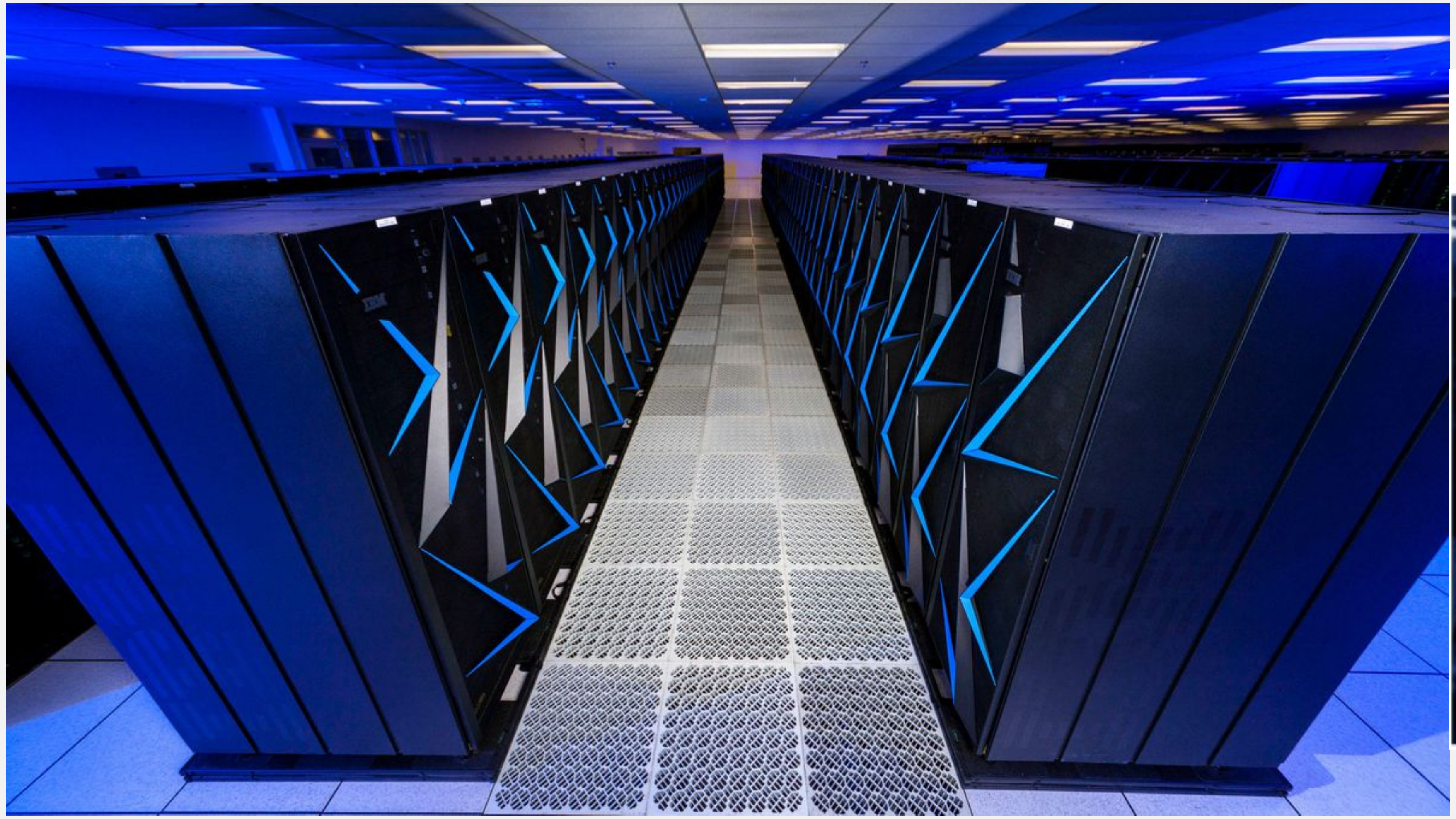
Principal Component Analysis (PCA) and Salience Maps

- PCA produces vectors along dimensions of maximal variation in the data
  - Vectors are orthogonal, and often performed on input data to reduce the number of inputs required.
  - Can also be performed on extracted features to reduce dimensionality for visualisation, similar to a t-SNE.

- Also possible to identify which features are important to CNNs.
  - Occlusion tests obstruct regions of an input image to find important features.
  - Salience maps show what features the network uses to make determinations.
  - Salience maps sometimes show that the network uses contextual information to make it's determinations.



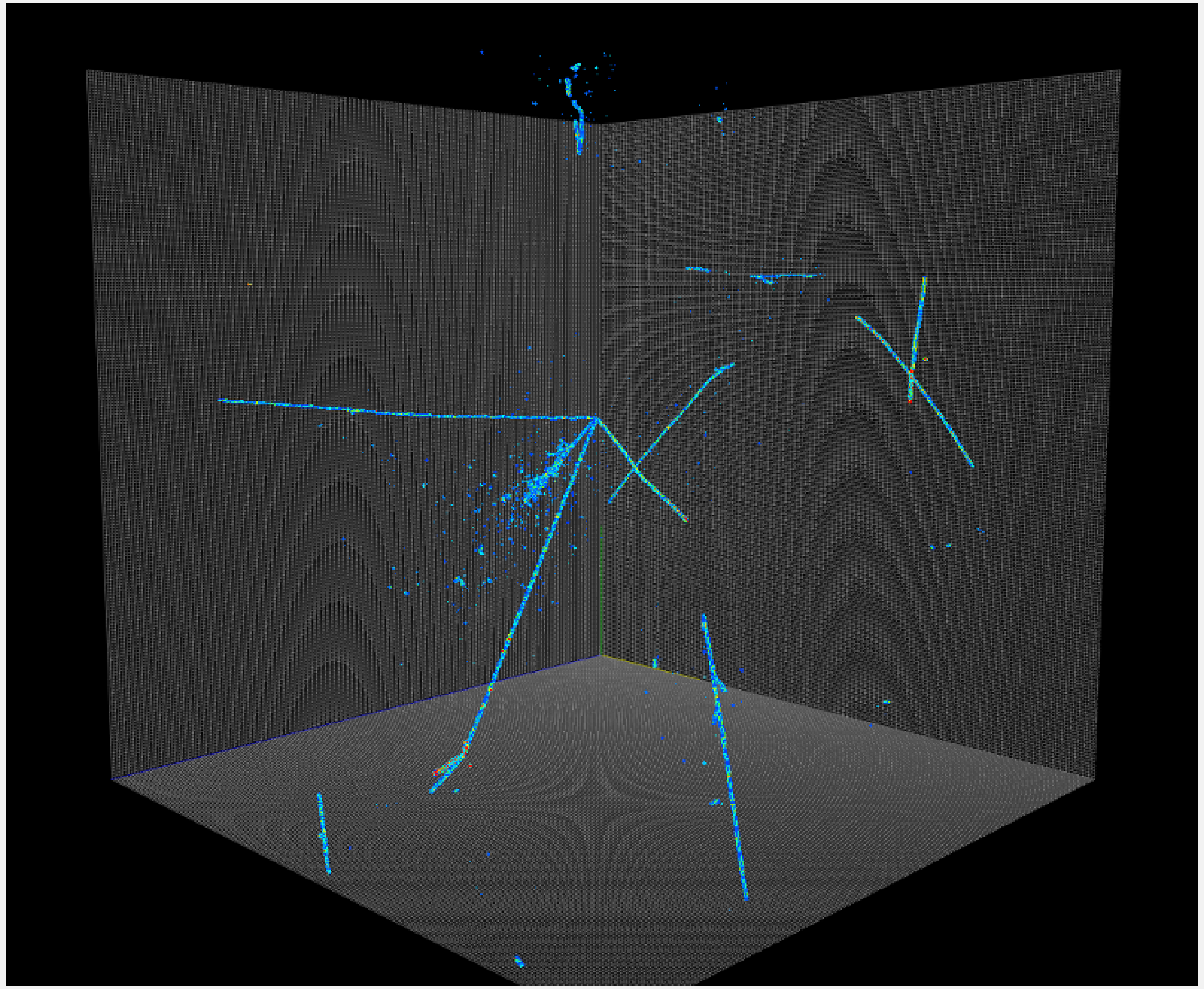
Challenge #4: Computational & System Constraints



- ML in neutrino experiments uses huge resources;
  - Neutrino experiments record billions of events a year.
  - Neural nets perform $>10^9$ floating point operations.
  - Widespread use of large-spread computing clusters such as the Open Science Grid to perform evaluation on CPUs.

- Small-scale GPU clusters often used for training.
- Large-scale GPU required for evaluation.
  - Currently no GPU computing clusters similar to the CPU OSG exists.

Improving Efficiency of Computing Resources



Liquid Argon Time Projection Chambers (LArTPCs) are globally sparse, but locally dense.

- Multi-task networks can reduce overall computational load.
  - Networks that identify neutrino flavour, can also be trained to identify neutrino sign, type of interaction and final state particles.
- Smaller networks often use fewer resources, but at the cost of reduced performance for high complexity applications.
- It is also possible to reduce the number of operations;
  - The structure of LArTPC data means that many CNNs multiply or sum together zeros.
  - Using submanifold sparse CNNs can reduce inference times by a factor of 30, and the memory cost by a factor of 300.

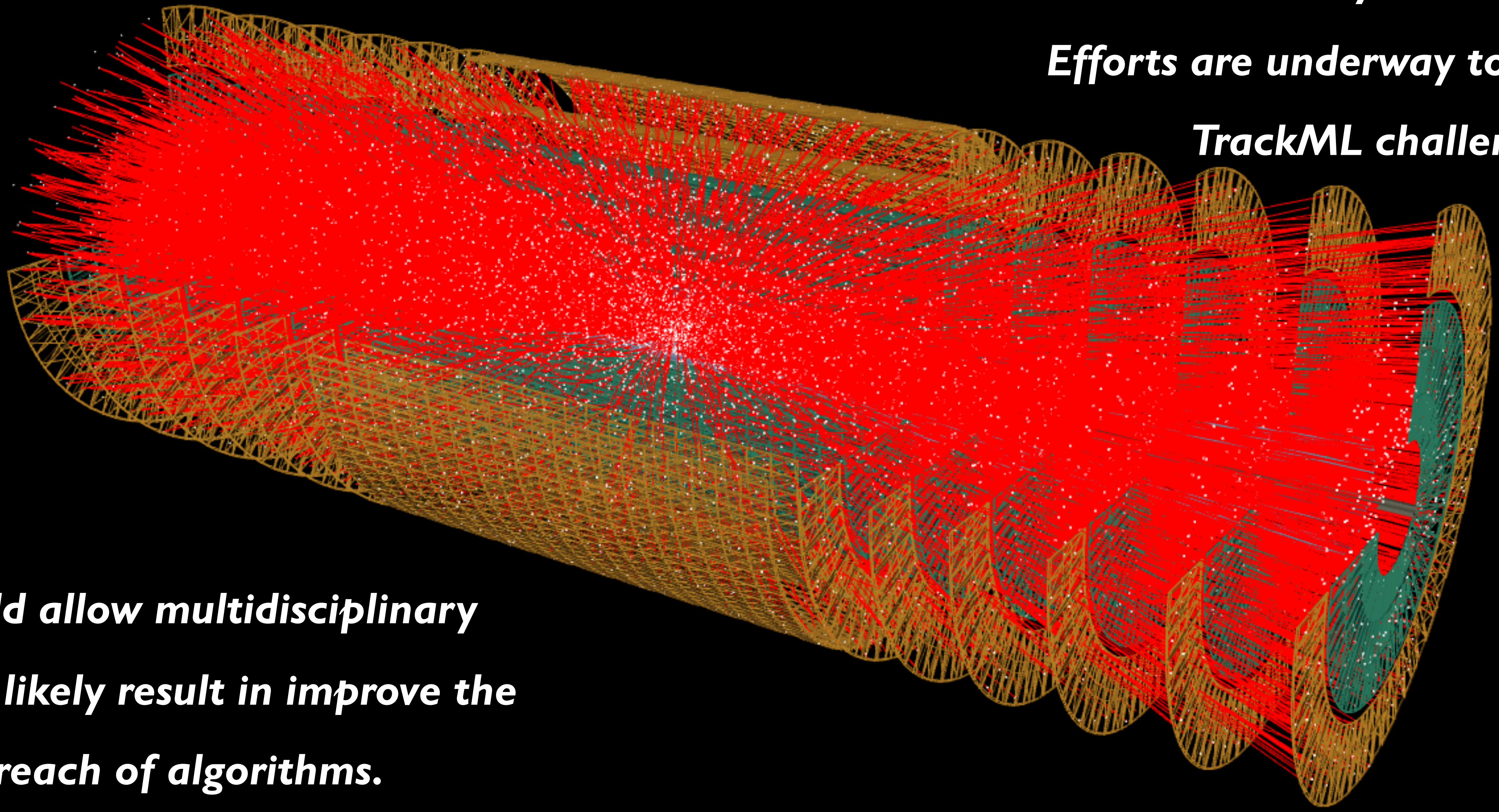
Challenge #5: Dataset Availability

Open datasets are commonplace in data science applications.

However, they are still not widely used in particle physics.

This is constrained by strict data-sharing restrictions.

Efforts are underway to change this, such as the TrackML challenge by LHC experiments.



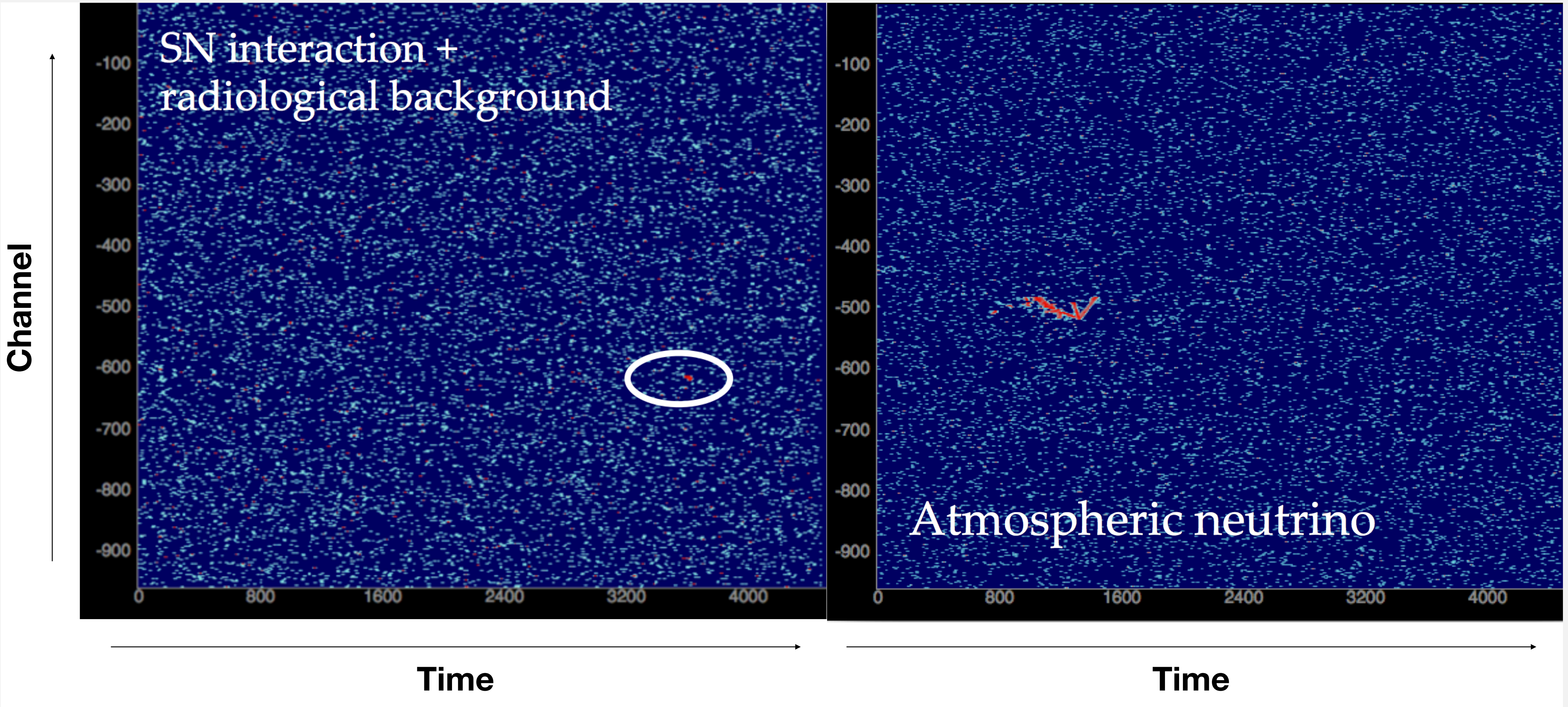
10,000 tracks from the TrackML challenge.

Open datasets would allow multidisciplinary research and would likely result in improve the quality and physics reach of algorithms.

This would require revisions to both authorship and data-sharing policies of experiments.

The Future Opportunities Offered by Machine Learning

Applications to Control and Manage Data Rates

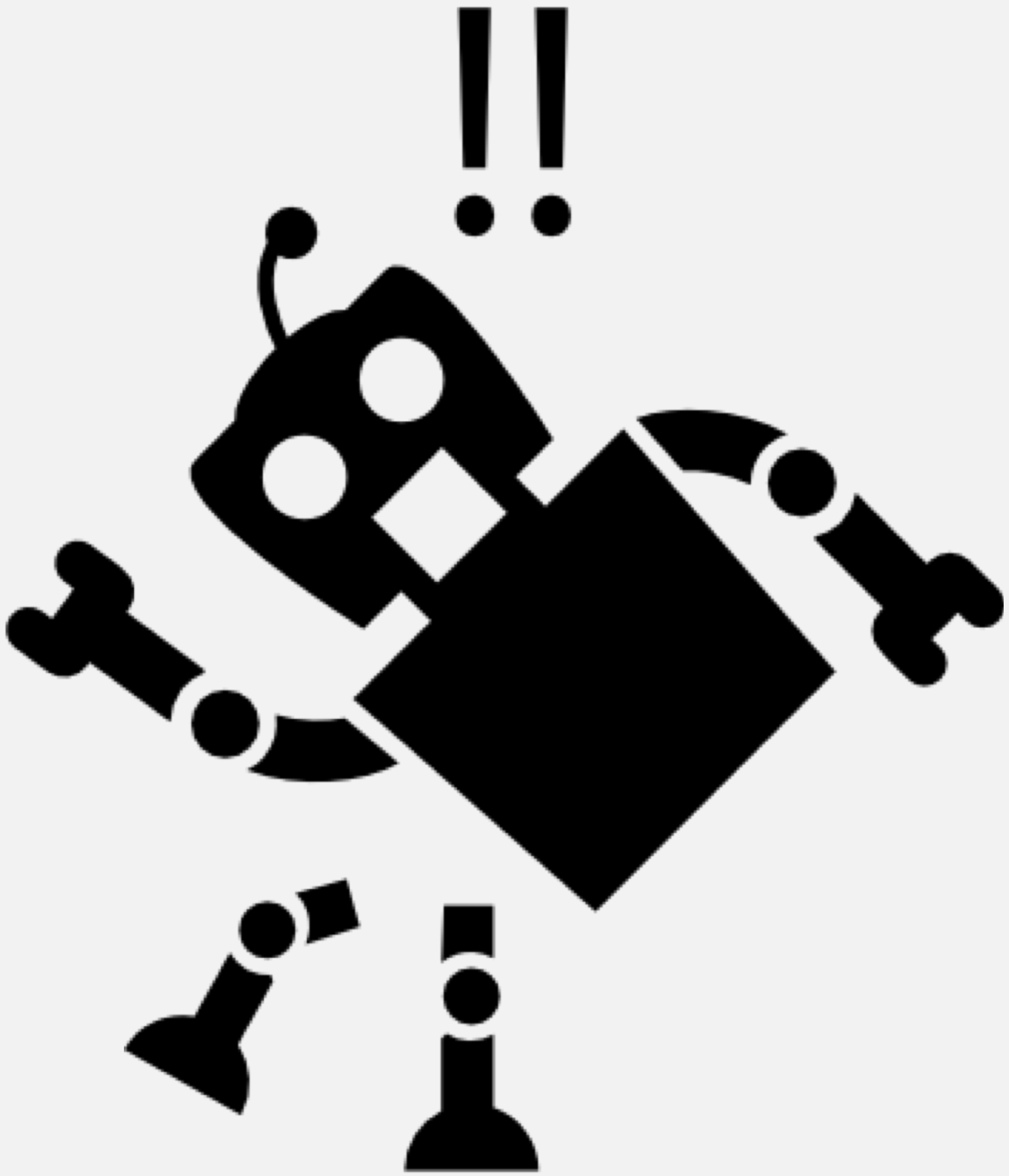


DUNE will search for rare, low energy interactions which are dominated by the constant backgrounds from electronics and radioactivity.

- FPGAs in future detectors offer the chance to manage the huge data rates of future experiments using CNNs.
  - Will require online data processing of terabits per second continuously for more than 20 years.
- The use of Machine Learning in triggers offer the chance to expand the physics reach of experiments.
  - Already seen in LHC experiments, and may reduce energy thresholds in LArTPC experiments by an order of magnitude.

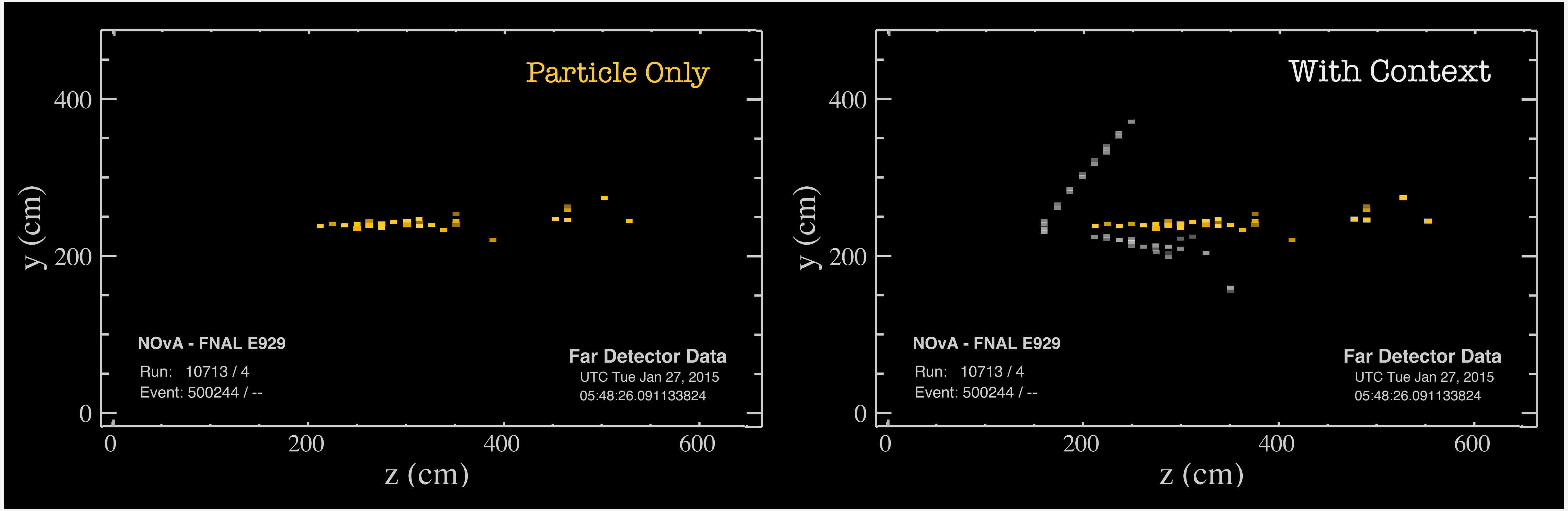
How Neutrino Physics Can Contribute to Machine Learning

- ***Quantitative results and careful statistical analyses.***
  - Consideration of systematic effects and bias is crucial to particle physics.
  - As ML becomes more widespread in physics, there will be increased efforts in understand this.
  - Key overlaps with ethical, and security concerns in wider ML applications.



- ***Use of simulated datasets corresponding to real data.***
  - Unlike most industry applications, particle physics trains on simulations.
  - These simulations can be tuned at will, making it possible to study networks behaviour under controlled modifications.
  - Comparing data and simulations can improve studies in domain transfer.

Physics Studies That Can Benefit Wider ML Applications



- NOvA found that adding contextual information to a particle identification network improved performance.
  - Has four Siamese towers, two for particle-only cluster, and two with the full event for contextual information.
  - Improves performance by 11%.
- First technique to employ a Siamese architecture to add context to a network.
- Bountiful scenarios for synergy between academic and industry applications.
  - Exploration will undoubtedly improve both fields.

Summary

- Neutrino experiments using a myriad of technologies to study numerous physics processes applying ML in the process.

- ML algorithms have been adapted for a range of applications including classification, energy reconstruction, Monte Carlo generation and bias reduction.

- Many of the challenges arising from applying ML have been overcome, though many challenges still remain.

- Future applications of neutrino physics can contribute to the development of ML in wider applications.

https://arxiv.org/abs/2008.01242