Introduction to Machine Learning in High Energy Physics

INTERNATIONAL CONFERENCE OF HIGH ENERGY PHYSICS

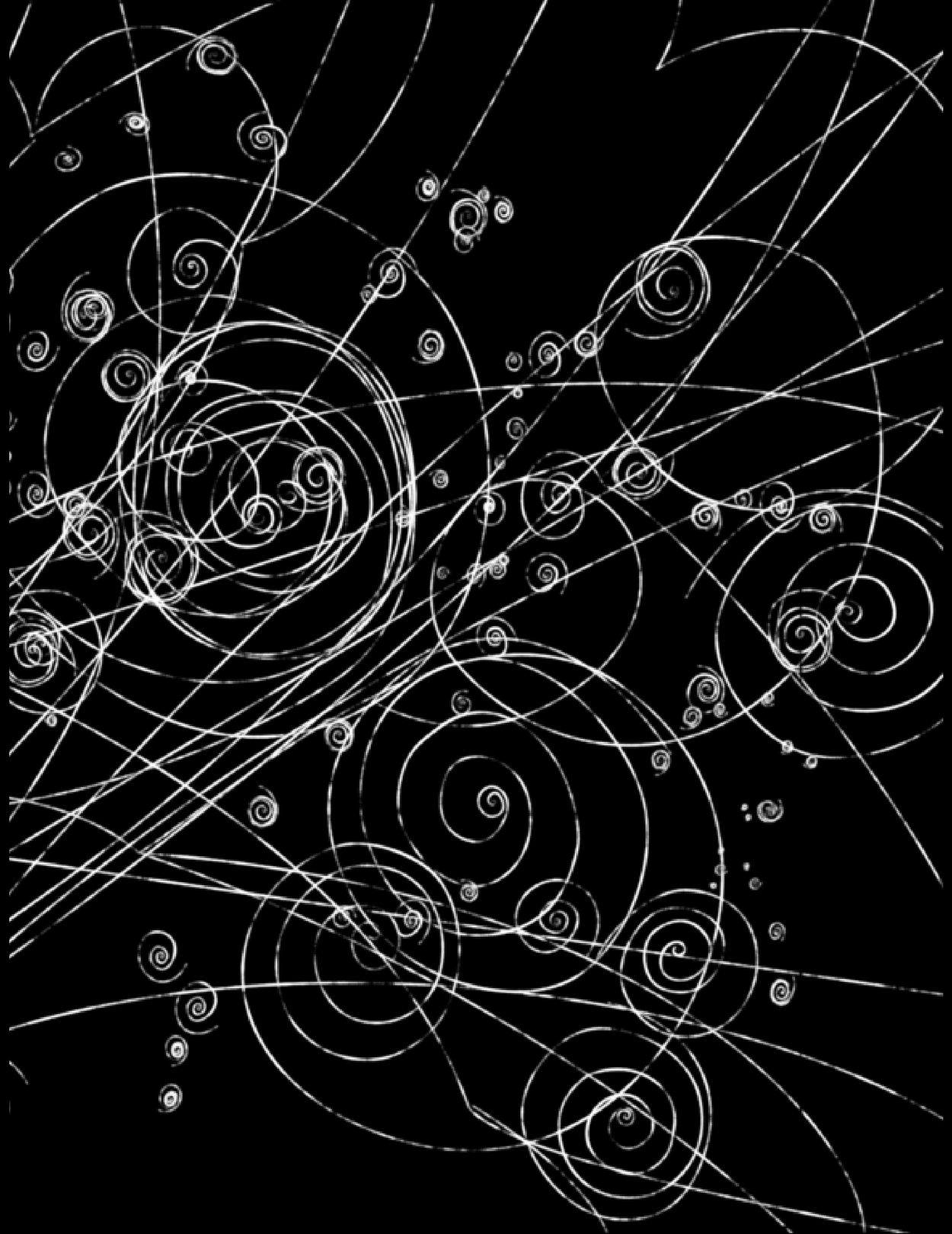
Sebastian Olivares | August 2021

Content:

"The saddest chapter in modern physics is and remains the Dirac theory...I regard it as learned trash which no one can take seriously."

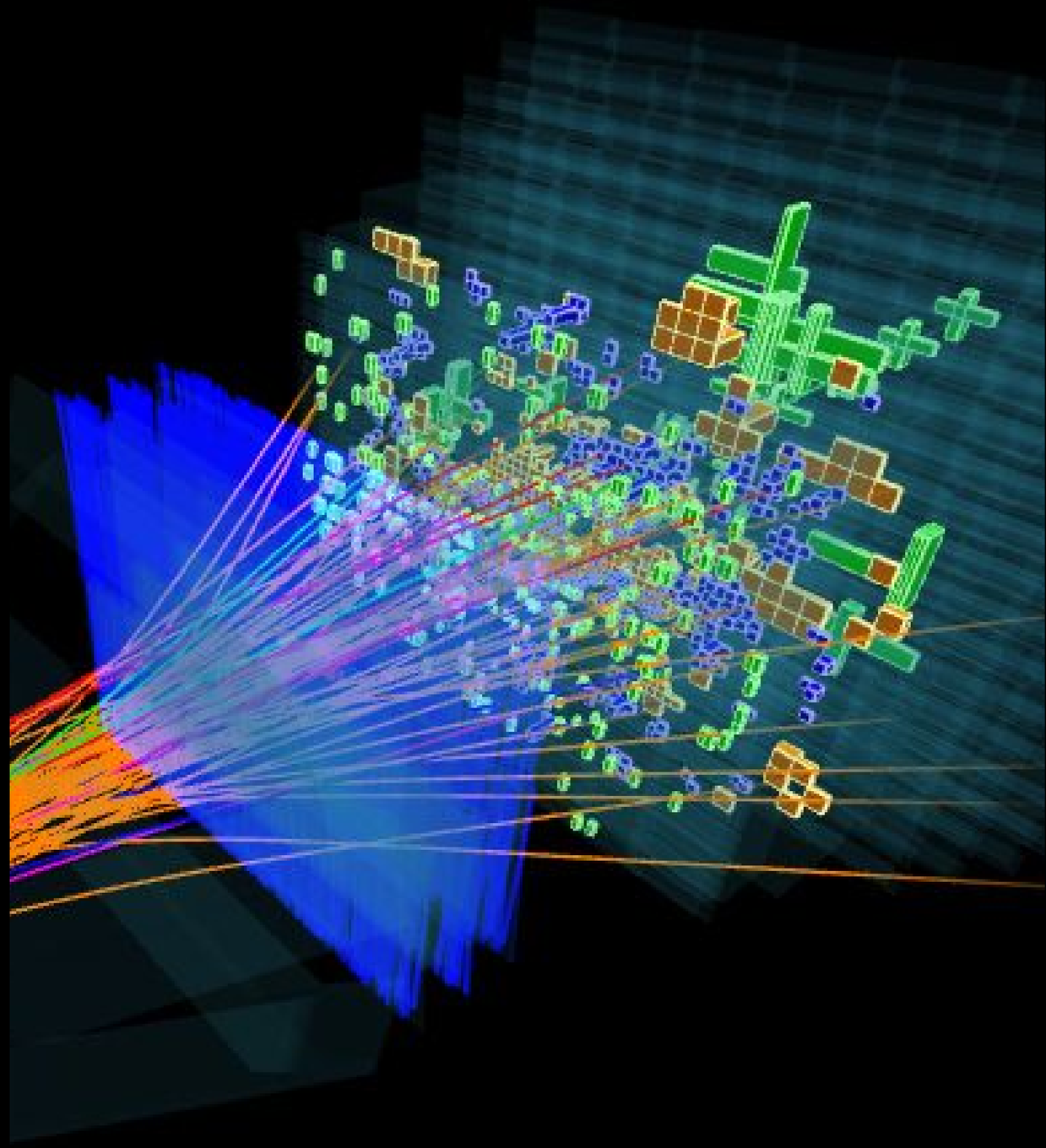
W. Heisenberg (1930)

Bubble chambers filled with superheated liquids that boil when charged particles pass through them transform the paths of the particles into visible tracks of bubbles, which can then be photographed and analised

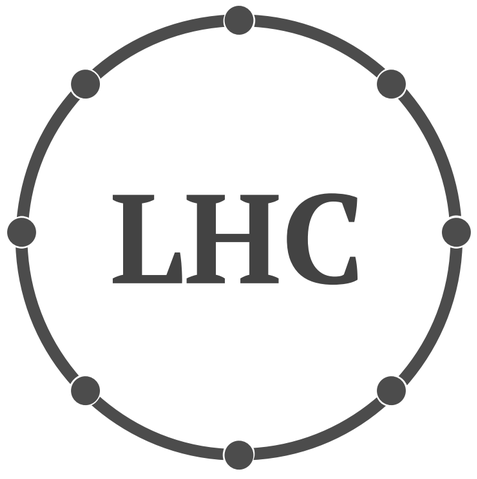


How complex are current particle detectors?

LHCb experiment analyses as many events every six seconds as the Big European Bubble Chamber recorded in its entire 11 years of operation (1973-1983)



Big data era!



Data at a rate of about: 1 petabyte/second

Higgs boson is produced only once every few billion proton-proton collisions

100 million detection elements

After data-reduction executed in real-time: 50 terabytes/second

Store as much data every hour as Facebook collect globally in a year

Machine Learning excels in the collection and analysis of big data samples...bring them on!

Geoffrey Hinton

Particle physics is governed by quantum mechanics

Was there a Higgs boson in this event?

Particle physics is governed by quantum mechanics

Probability distribution for signal and background production

$$dP^n_{\text{data}} = |\mathcal{M}_S + \mathcal{M}_B|^2 dp_1 \cdots dp_n$$

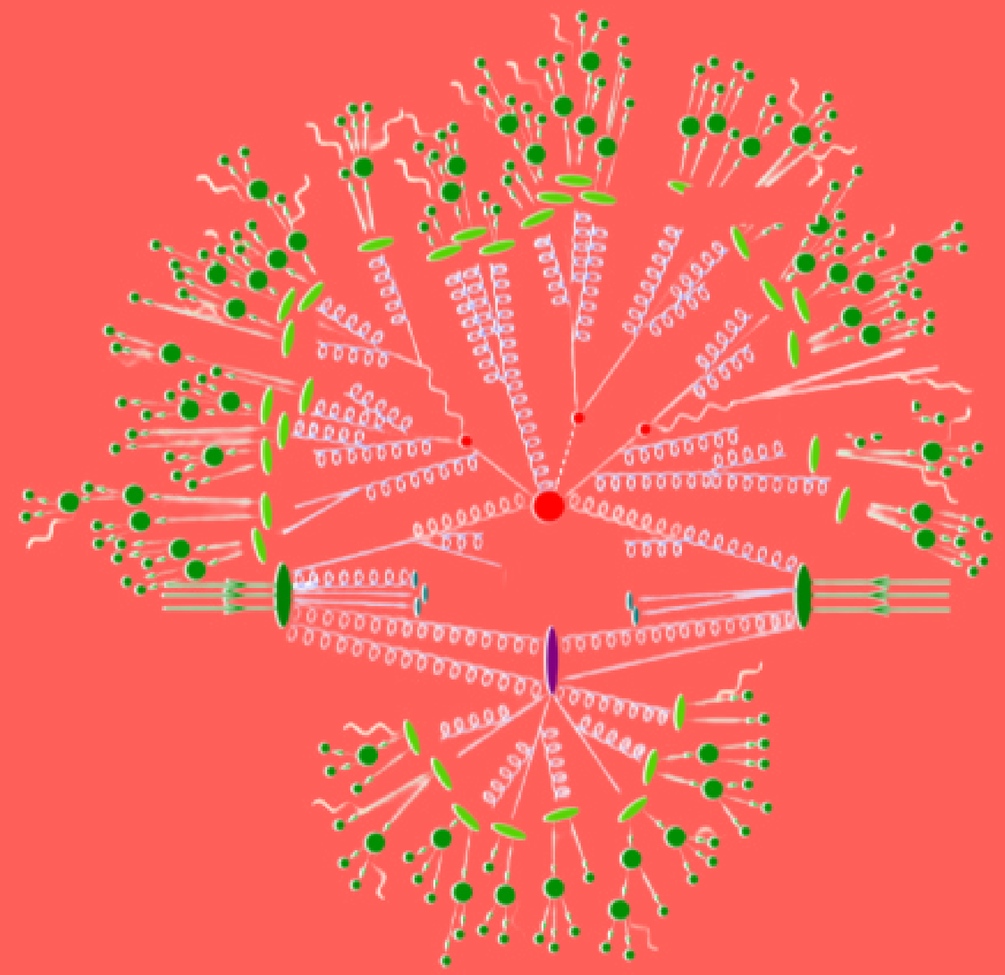
$$\mathcal{M}_S + \mathcal{M}_B \approx \mathcal{M}_S \qquad \mathcal{M}_S + \mathcal{M}_B \approx \mathcal{M}_B$$

$$P_{\text{data}} = \alpha_S \, P_{\text{S}} + \alpha_B \, P_{\text{B}}$$

$$\alpha_S + \alpha_B = 1$$

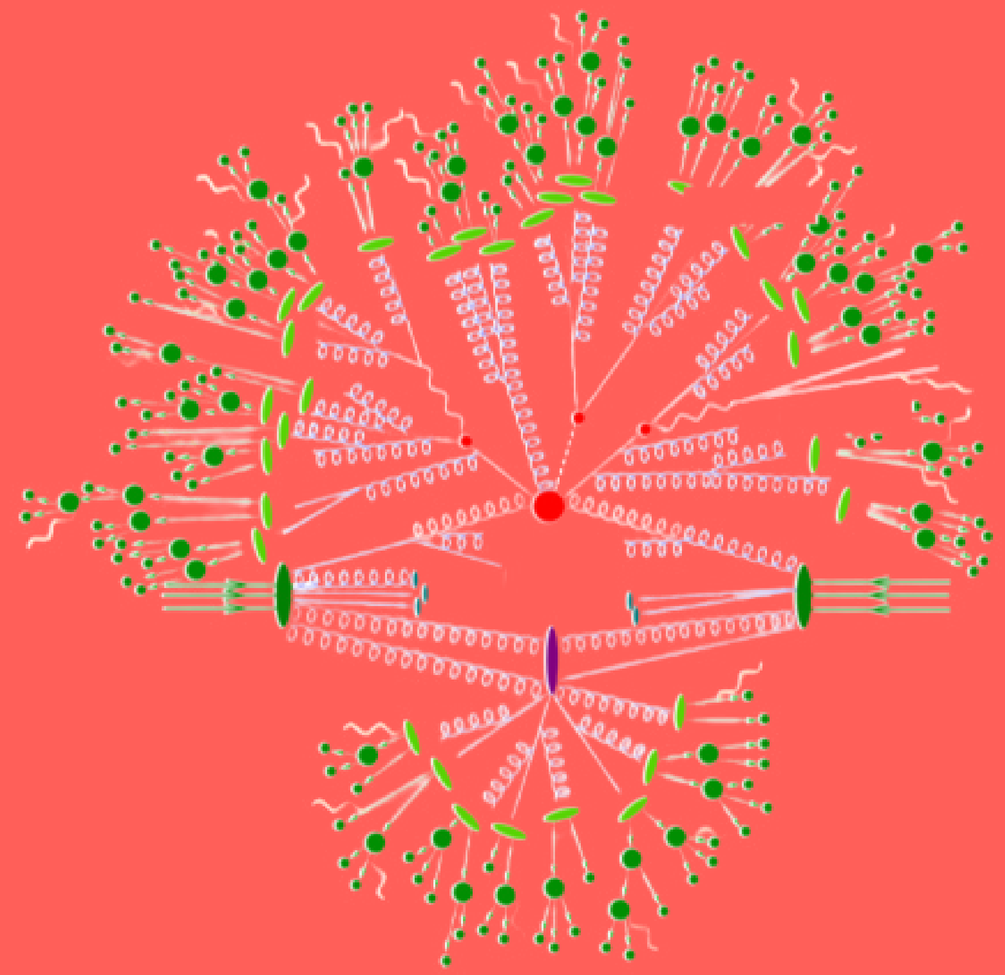
Remarkably accurate simulated data

- The smallest scale LHC probes is ~10^-18
- Higgs boson size ~10^-17 m
- Between 10^-18 to 10^-15 m, a semi-classical Markov model used to turn primordial particles to into hundreds of
- Between 10^-15 to 10^-6 m, quarks/gluons turn into metastable sub-atomic
- Detectors accurate from 10^-6 to 100 m



Remarkably accurate simulated data

- The smallest scale LHC probes is ~10^-18
- Higgs boson size ~10^-17 m
- Between 10^-18 to 10^-15 m, a semi-classical Markov model used to turn primordial particles to into hundreds of
- Between 10^-15 to 10^-6 m, quarks/gluons turn into metastable sub-atomic
- Detectors accurate from 10^-6 to 100 m



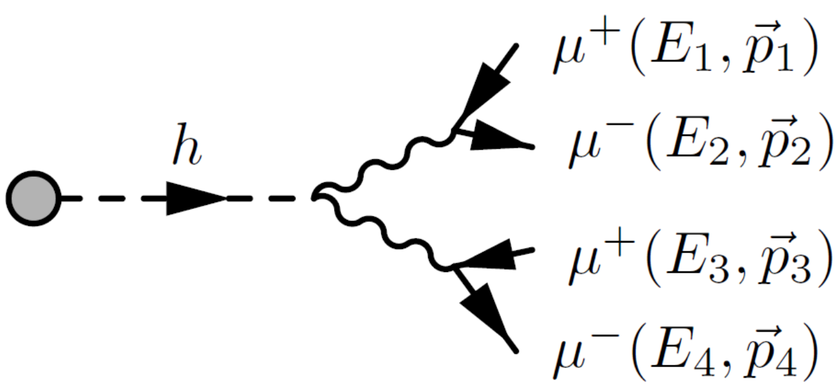
Remarkably accurate simulated data

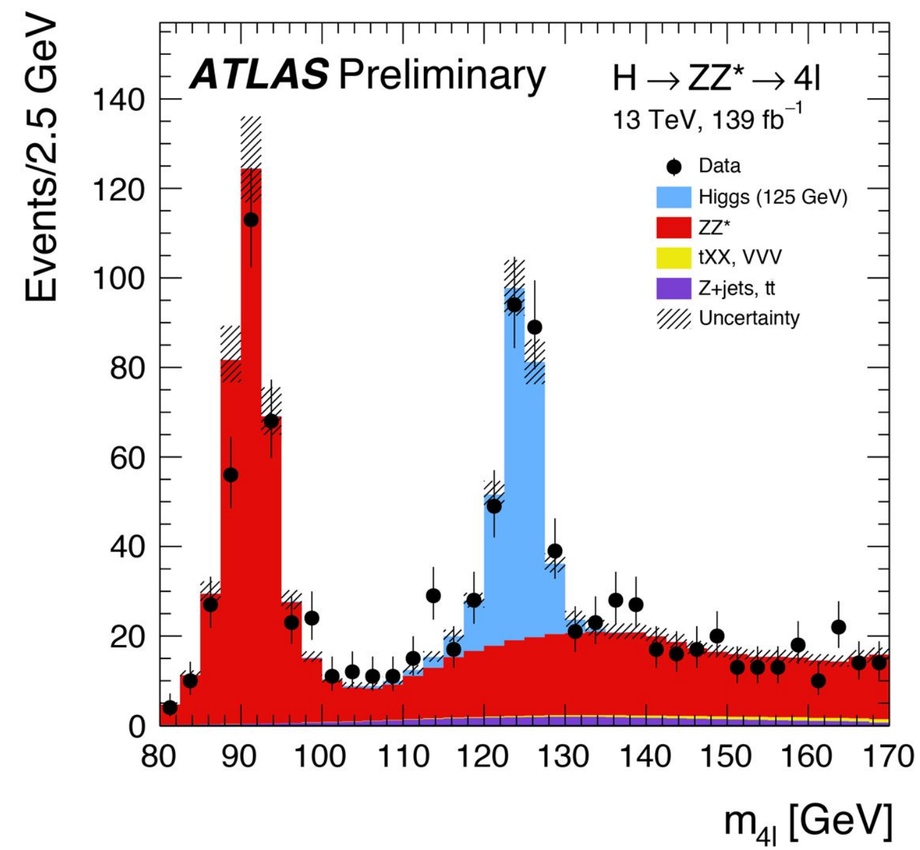
1. Process raw sensor data into low-level objects (e.g. calorimeter clusters and tracks)
2. Estimate the energy, momentum and identity of individual particles
3. Event selection algorithms selects subsets of the collision data for further analysis on the basis of the information associated to individual events
4. Cumulative product of these steps reduce the dimensionality of the problem to a number small enough to allow the missing statistical model to be estimated using simulated samples

Higgs boson discovery with the four-lepton channel

$$m = \sqrt{(E_1 + E_2 + E_3 + E_4)^2 - (\vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \vec{p}_4)^2}$$

Occurs roughly once every few billion proton-proton collisions





What feature is the optimal way of statistically discriminating a signal from its background?

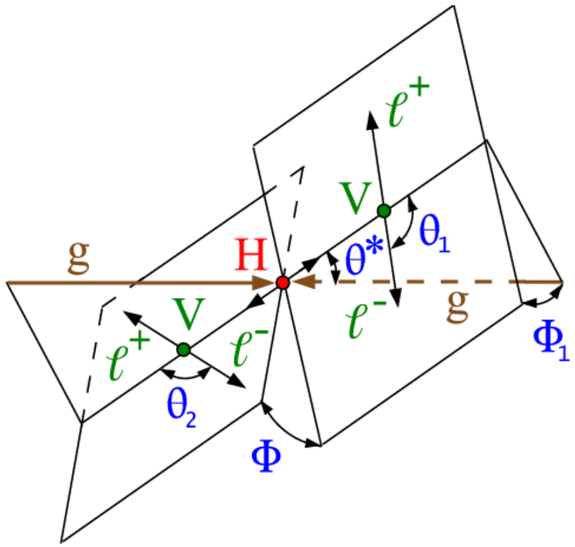
Higgs boson discovery with the four-lepton channel

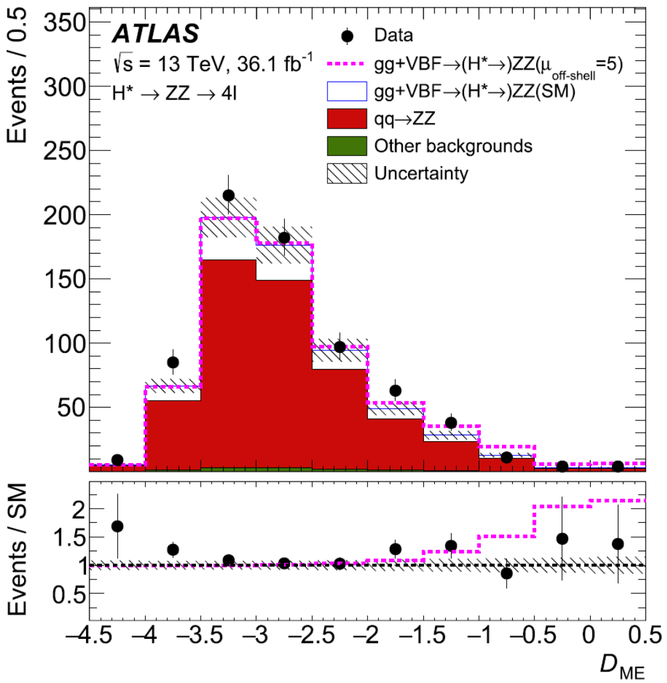
Particle physicists have sought to improve the power of their analysis by employing algorithms that utilize **mutiple variables** simultaneously

multi-variate analisis (MVA)

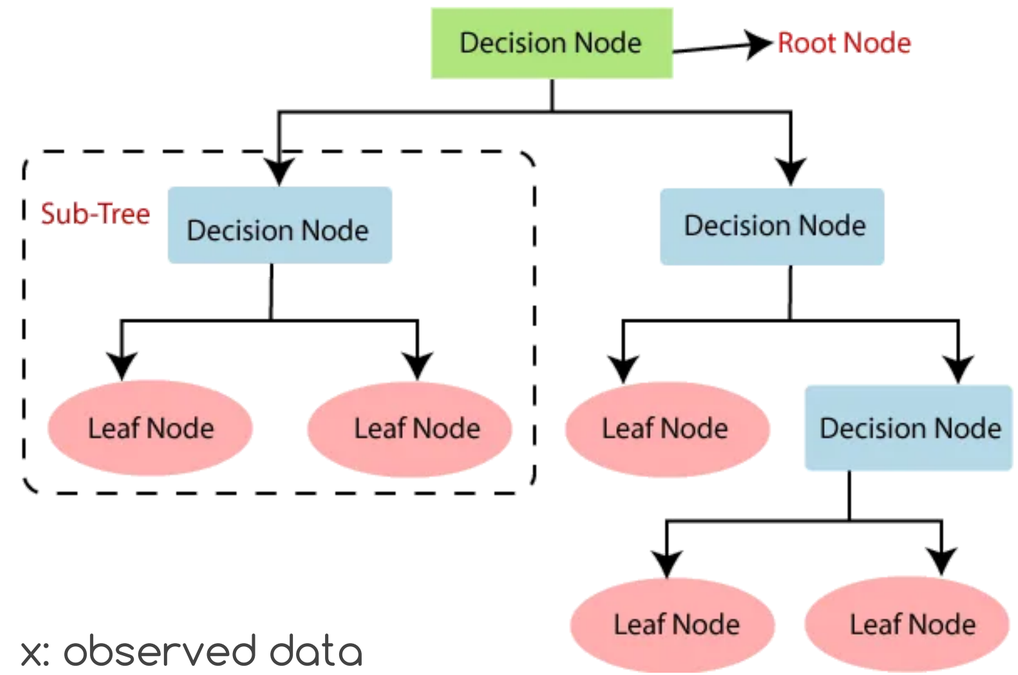
↕

machine learning (ML)





For several years the status quo of machine learning in HEP was to use boosted decision trees (BDT) implemented in the software package TMVA



Ensemble output: $\hat{y}(x) = \sum_t w_t h_t(x)$

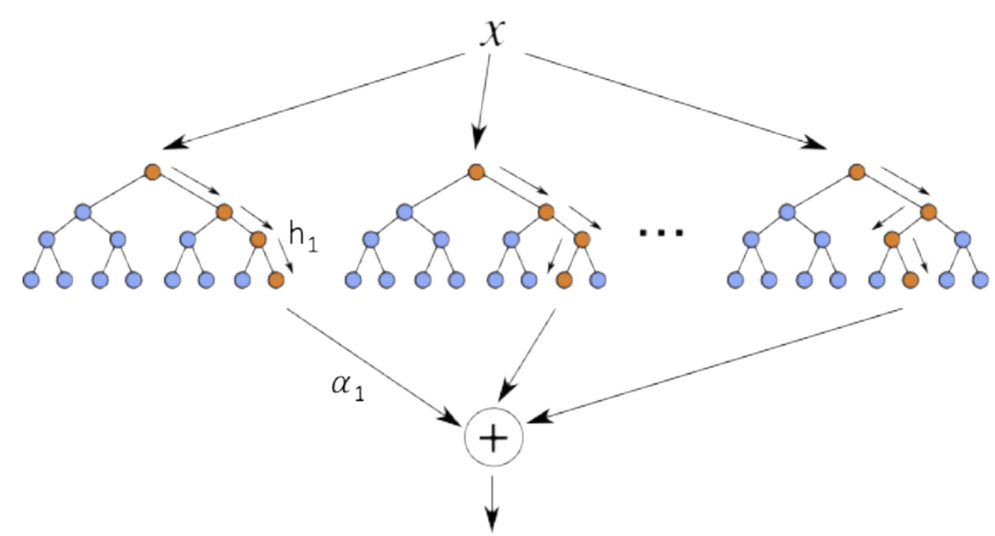
x: observed data
h(x): tree's output
w: weight
y: output

Objective function (goal is to minimise it):

$$O(x) = \sum_i l(\hat{y}_i, y_i) + \sum_t \Omega(f_t)$$

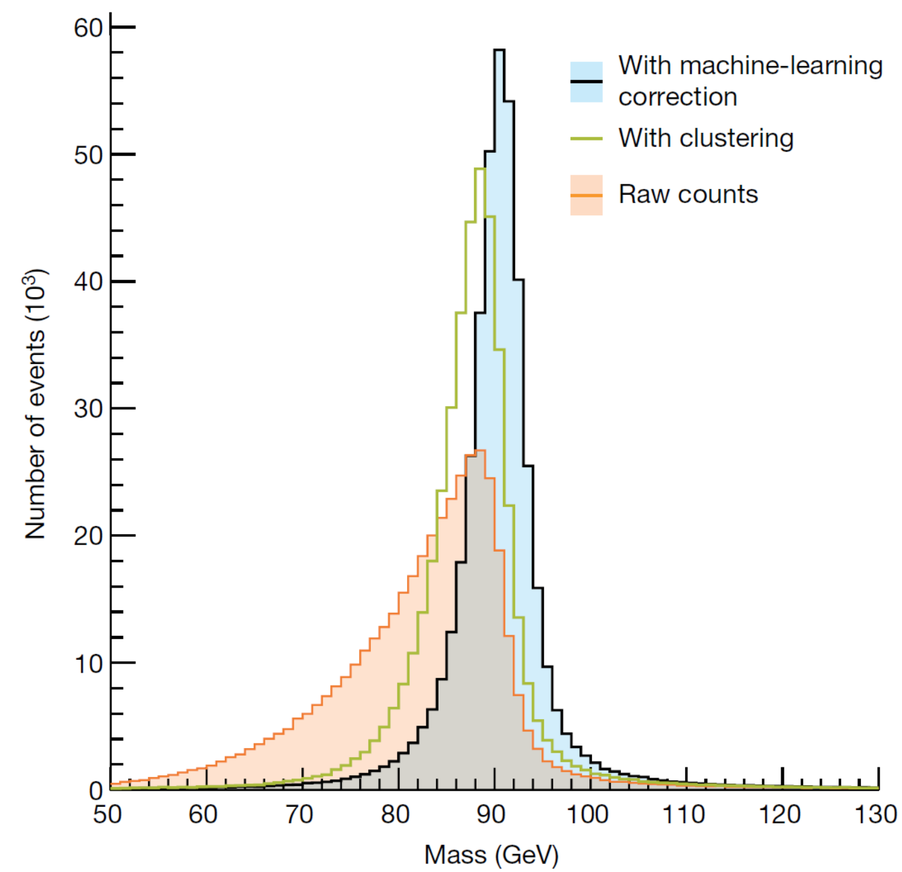
$l(\hat{y}_i, y_i)$ Loss function, which is the distance between the truth and the prediction

$\Omega(f_t)$ Regularisation function, which penalises the complexity of the tree

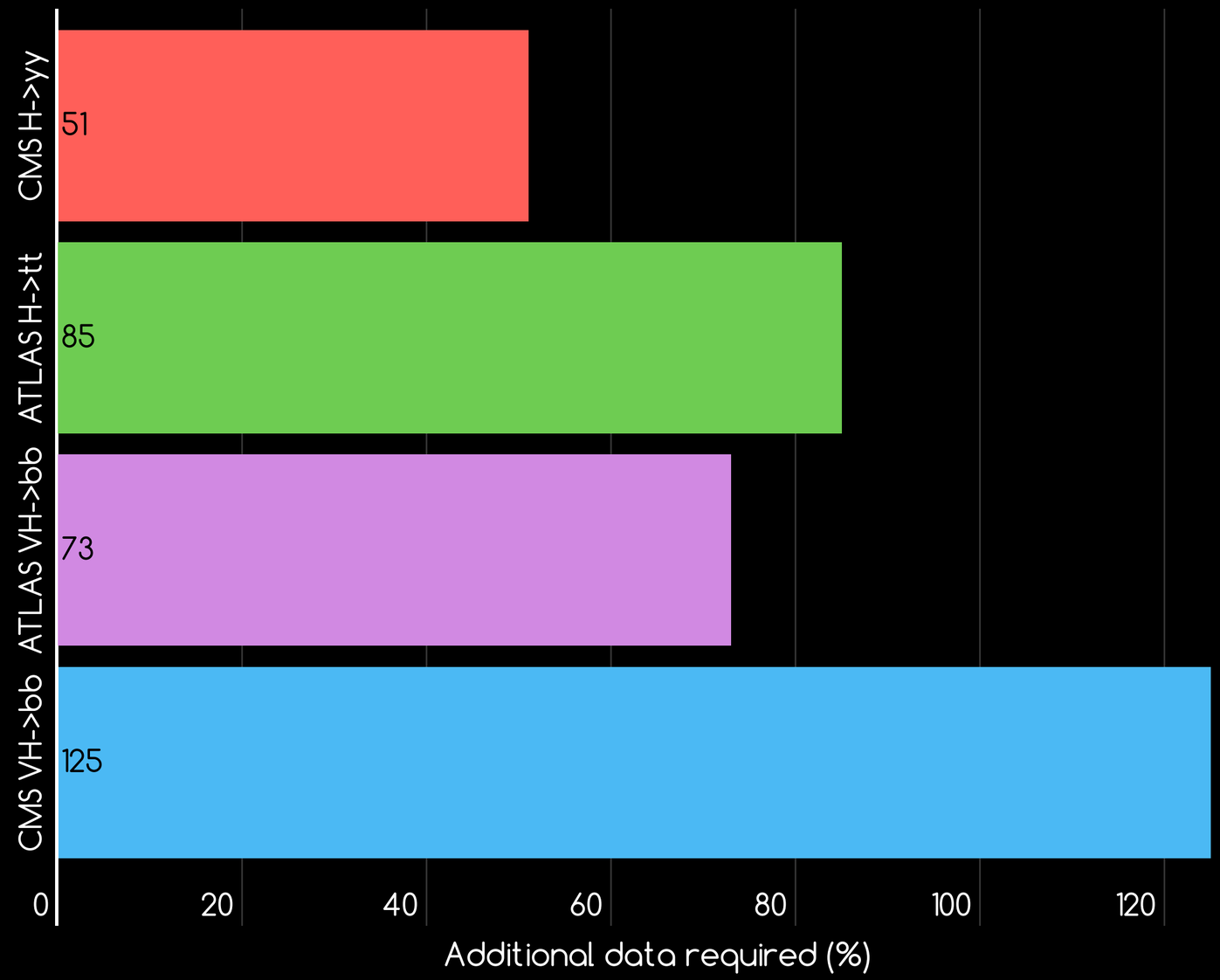


- For the first years, LHCb primary algorithm for classification was BDT
- The sensitivity achieved by a LHCb dark matter searches analysis with BDT using data collected just in 2016, would have required 10 years of data collection withut the use of machine learning

- BDTs are used to increase the resolution of the CMS electromagnetic calorimeter
- Deposited energy in different sensors are clustered together to recover the original energy of the particle
- Applying this energy correction to the Z boson decay into electron-positron



Data collected between 2011-2012



Particle physics has traditionally been using physically-motivated classifiers

Over the last several years this has been replaced with "modern" machine learning

The modern approach is to feed raw, minimally-processed data, rather than high-level physically-motivated variables

Bs-meson decay into muon-antimuon

SM predicts that only three out of every billion Bs mesons decay into muon-antimuon final state

CMS and LHCb were the first to find evidence of this decay

A human-designed tracking algorithm first reconstructs the paths taken by the muon and the antimuon, from these paths the momenta of the particles are inferred

Only the di-muon mass and the angle between them are used in the BDT

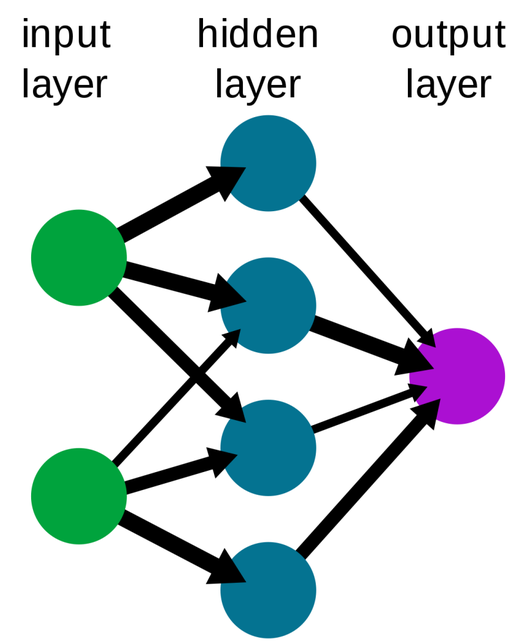
Information can be lost when these human-designed tools are used to extract features that fail to fully pacture the complexity of the problem

In neural networks, the space of functions searched is defined by the structure of the networks, which defines a series of transformations

These transformations map the input *x* onto internal or "hidden" states *hi*, until the final transformation maps these hidden states onto the function output *y*

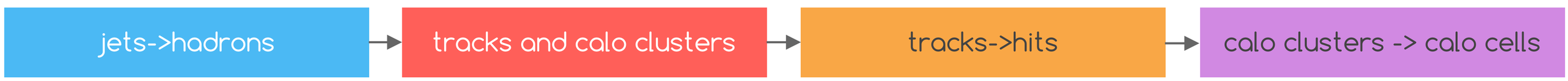
gi: activation function
Wi: weight matrix
b: bias vector

$$\boldsymbol{h}_{i+1} = g_i(W_i \boldsymbol{h}_i + \boldsymbol{b}_i)$$

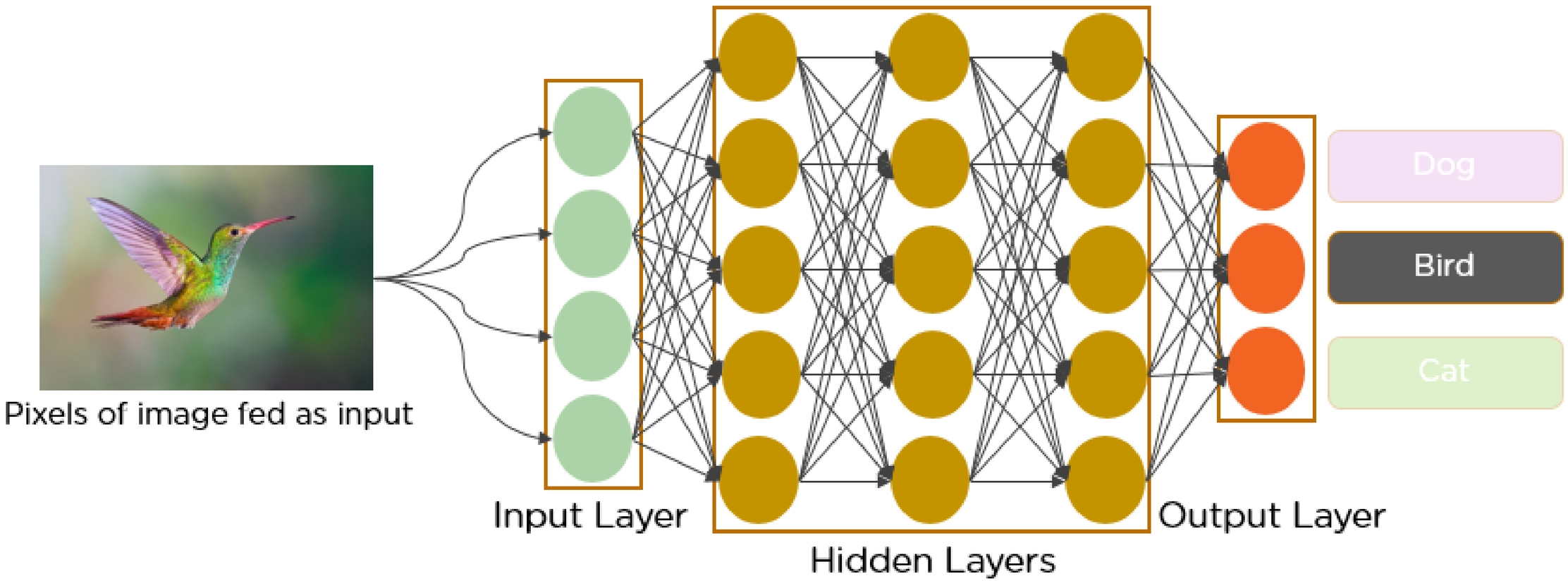


First embedding is simply the input vector *h0=x*, and the final embedding is the output of the network *y*

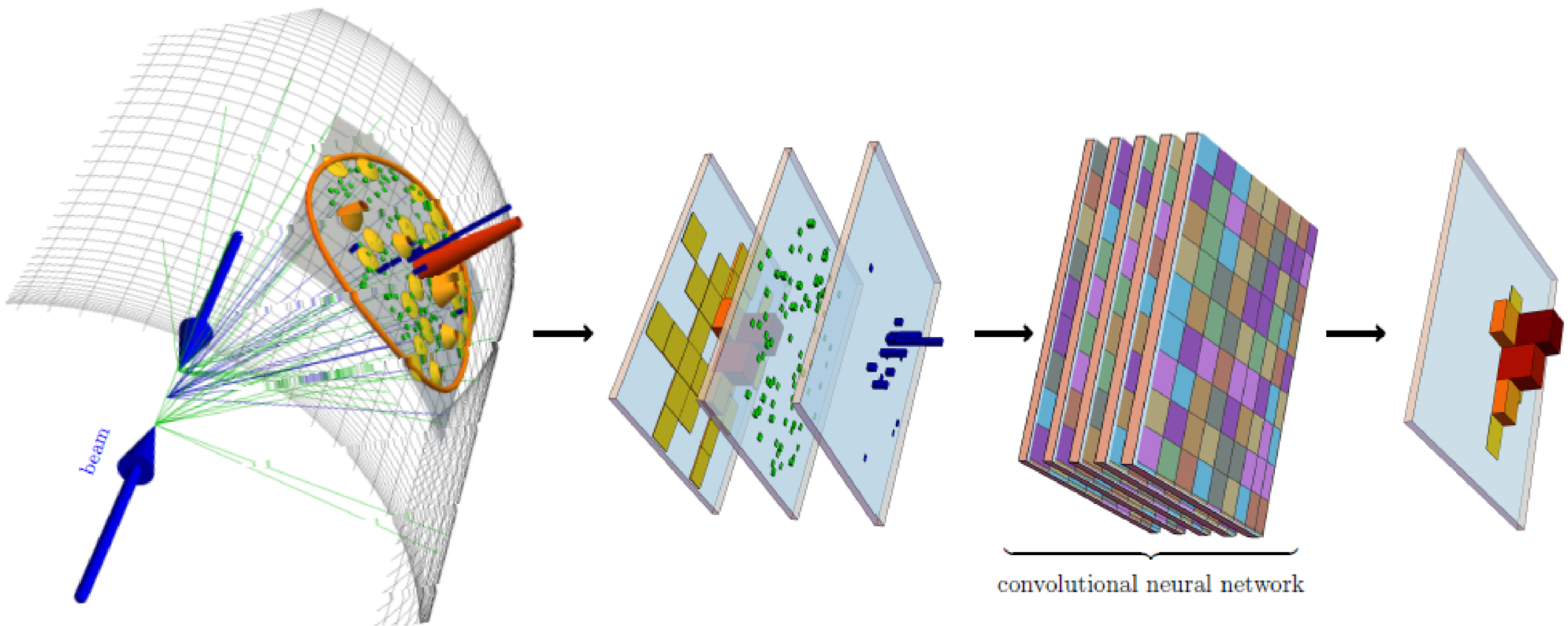
Deep networks not only have more expressive capacity, but also the layers can be interpreted as building up a **hierarchical** representation of the data



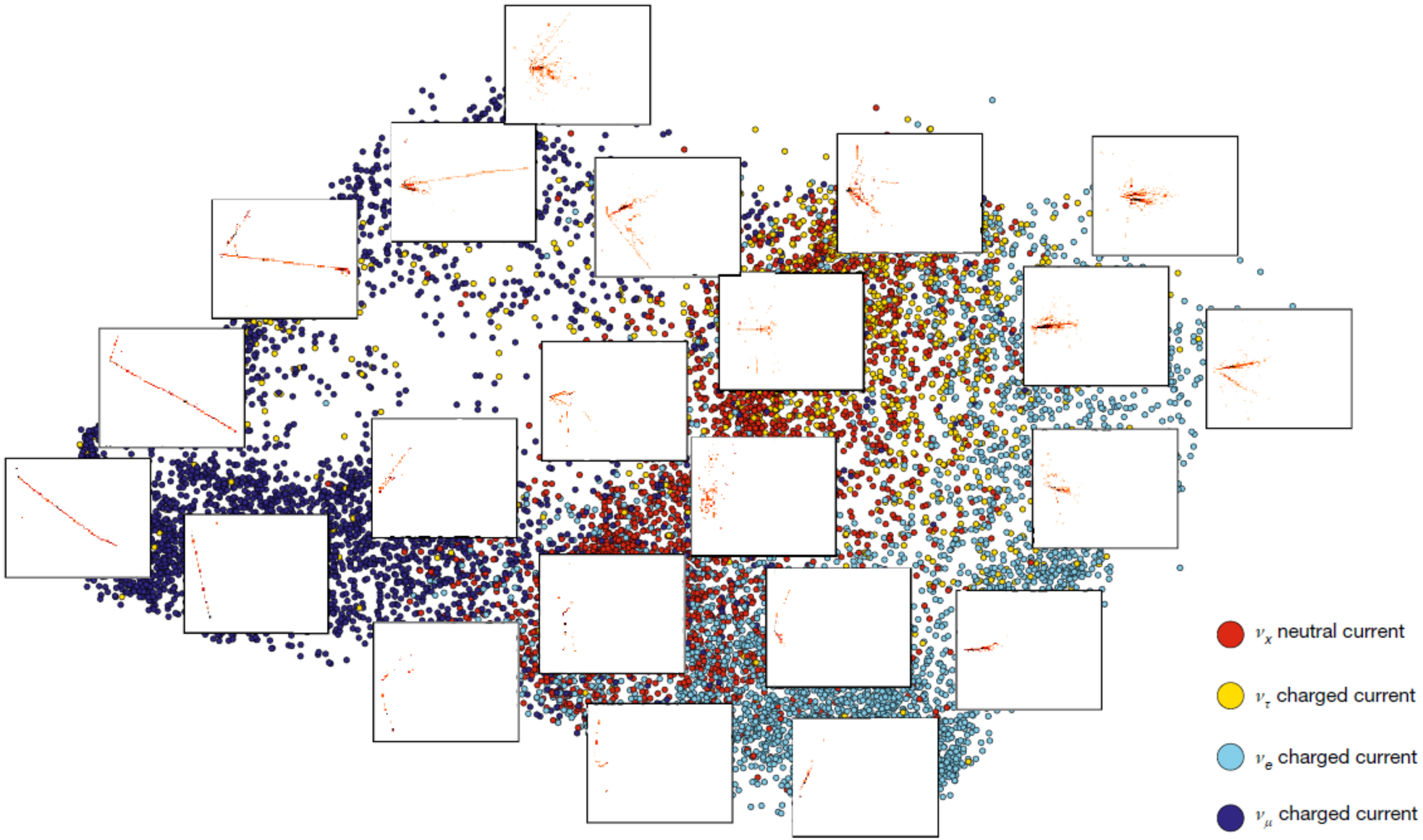
Convolutional Neural Network was invented for image



PUMML (PileUp Mitigation with Machine Learning) algorithm



t-SNE Nova event-selection using CNN



- Machine learning has the potential to improve the way lattice QCD calculation can be done with learned approximations or by more efficienctly sampling the configuration space

- AI Feynman, symbolic regression to learn 100 equations from the Feynman Lectures on Physics from noisy numerical sampling of

"I feel more optimistic about the possibility of transcendental progress in fundamental physics now that at any other time in my career, even if I myself may not be able to comprehend the final theory."

M. D. Schwartz (2021)