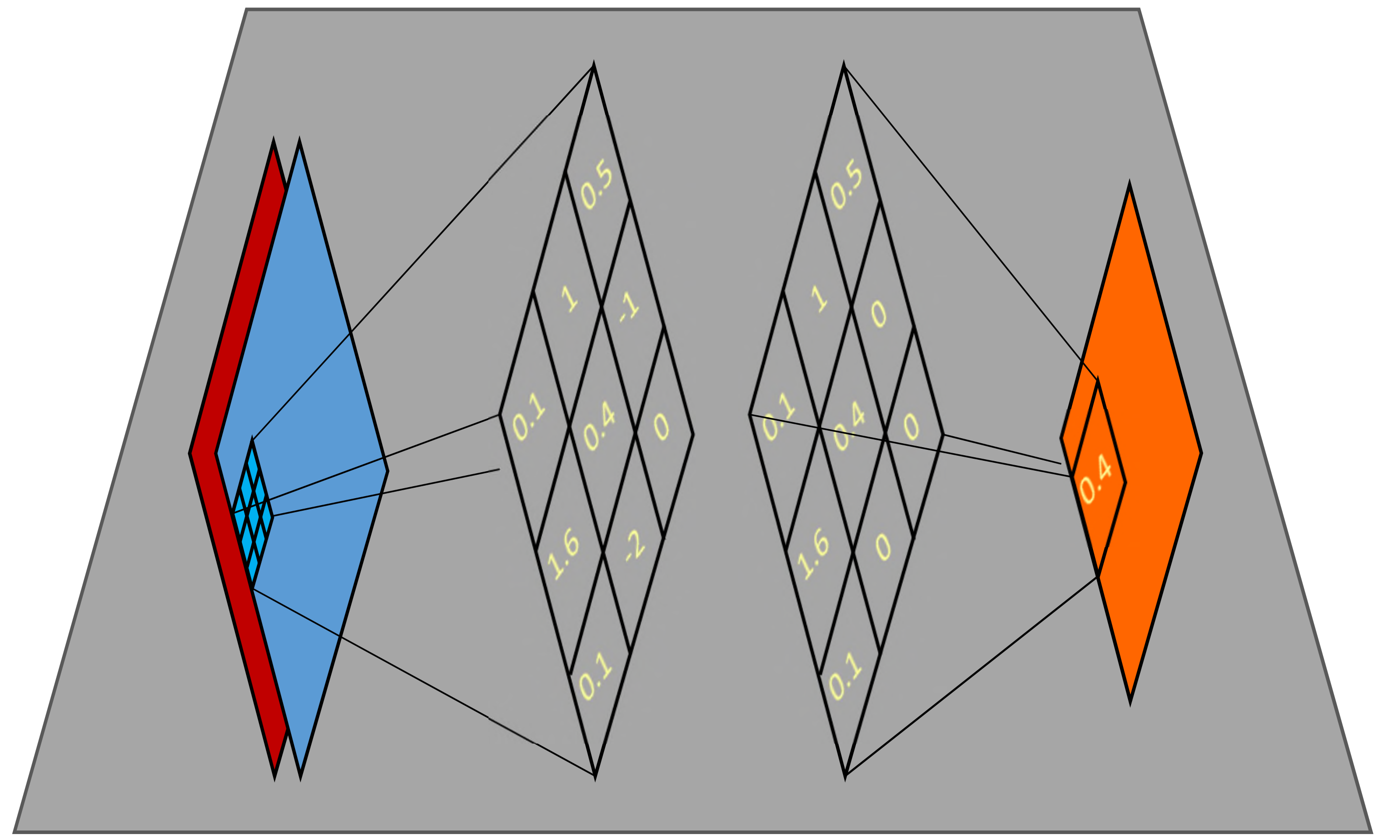
Deep Learning meets Physics



Prof. Dr. Martin Erdmann, RWTH Aachen University, 24-Nov-2020

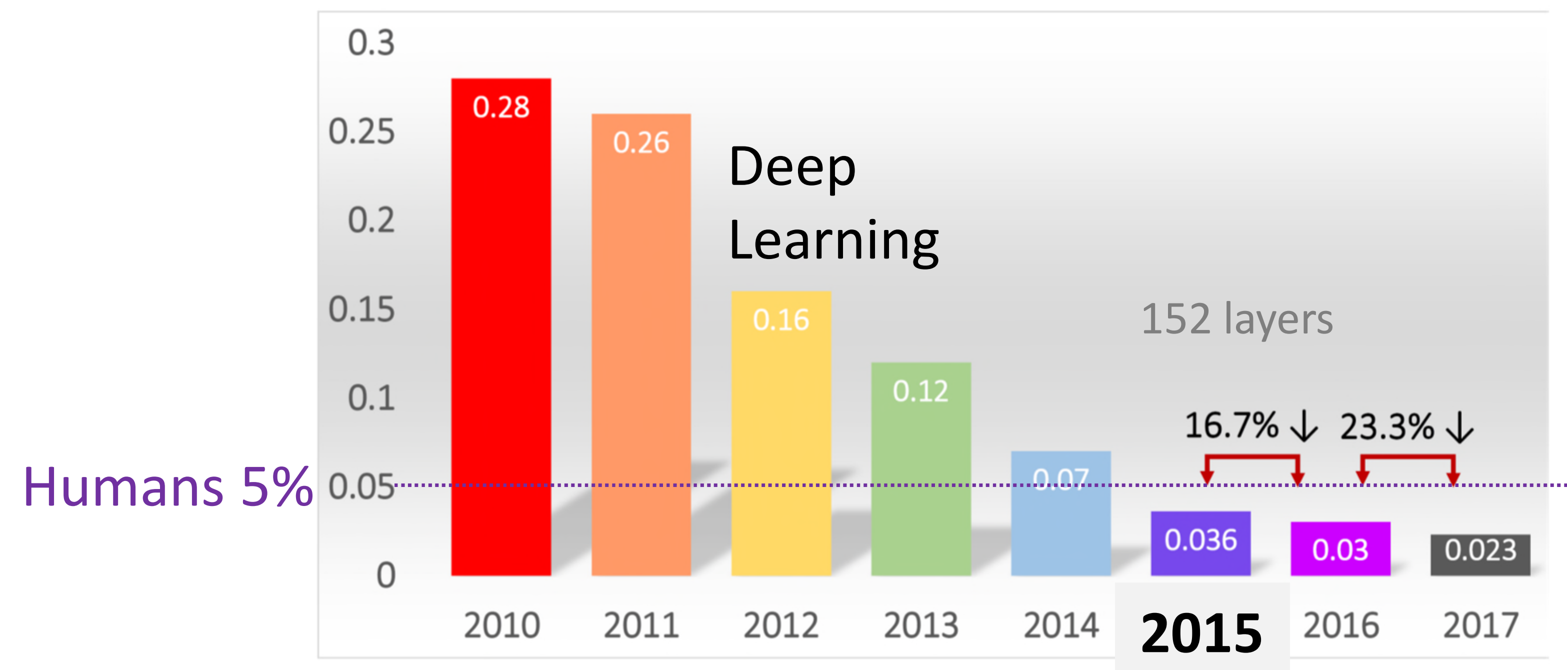
O. Russakovsky et al, arXiv:1409.0575; K. He, X. Zhang, S. Ren, J. Sunar, arXiv:1512.03385
WMW Jie Hu, Li Shen (Oxford), Gang Sun, 2017

Deep Learning spectacular success

Image recognition challenge

ImageNet: 1.2 million images in 1000 categories

Classification error rate



Deep learning errors < human errors

Generative Modeling

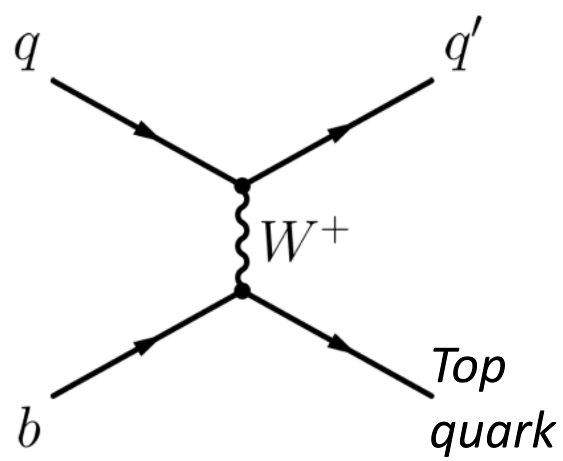
https://thispersondoesnotexist.com

Plan for today

- Machine learning accelerates physics research
- What deep learning is precisely: neural networks
- Examples of deep learning algorithms in (astro)particle physics

Electroweak Top Quark Production with CMS

CMS, Phys. Rev. Lett. 107, 091802 –25 Aug. 2011

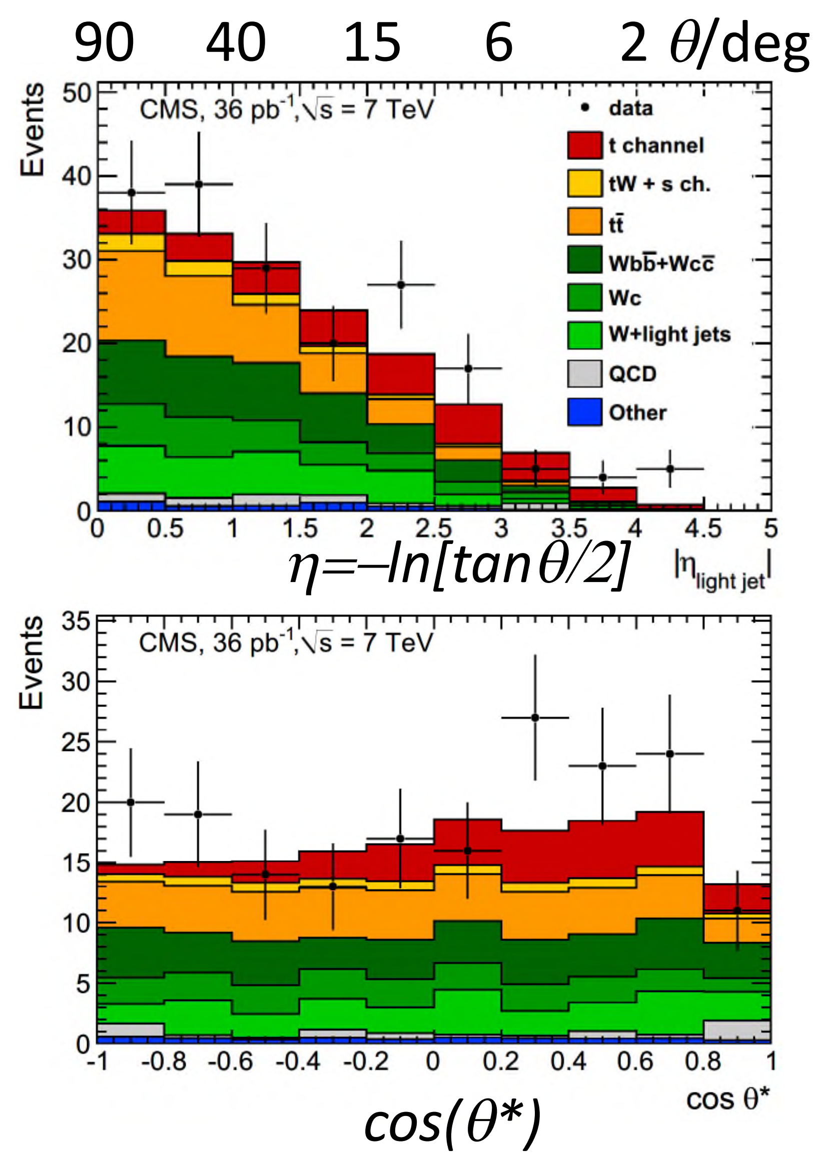


1) Rutherford-type small angle θ scattering

2) Top quark decays
$t \rightarrow W\,b \rightarrow e\,\nu\;\, b$
W spin=1 → particular lepton decay angular distribution *cos(θ*)*

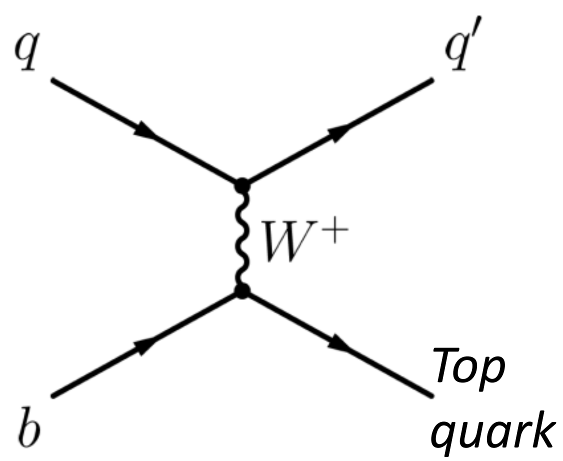
Expected at LHC
Luminosity = **1 fb^{-1}**

Luminosity = **36 pb^{-1}**



Electroweak Top Quark Production with CMS

CMS, Phys. Rev. Lett. 107, 091802 –25 Aug. 2011

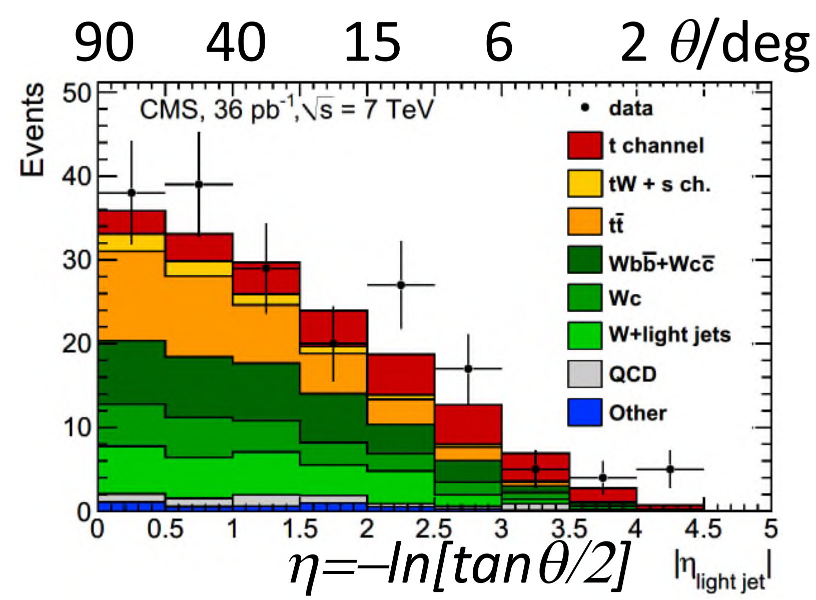


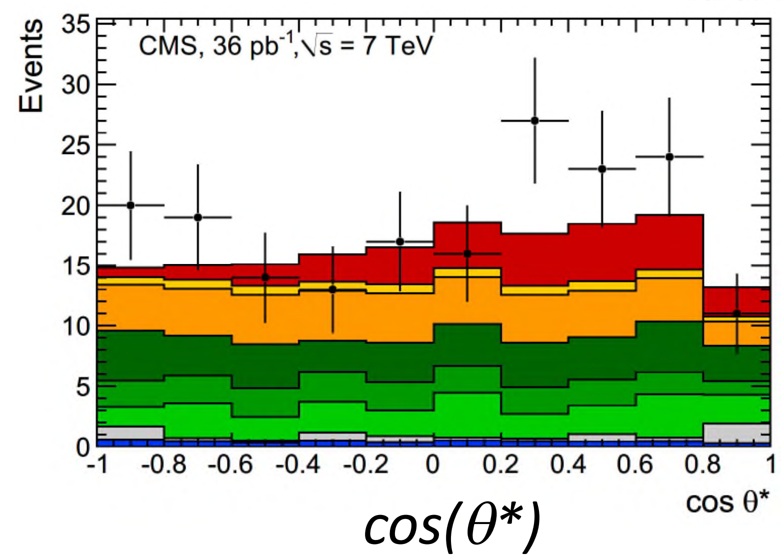
1) Rutherford-type small angle θ scattering

2) Top quark decays
$t \rightarrow W\, b \rightarrow e\nu\;\; b$
W spin=1 → particular lepton decay angular distribution $cos(\theta^*)$

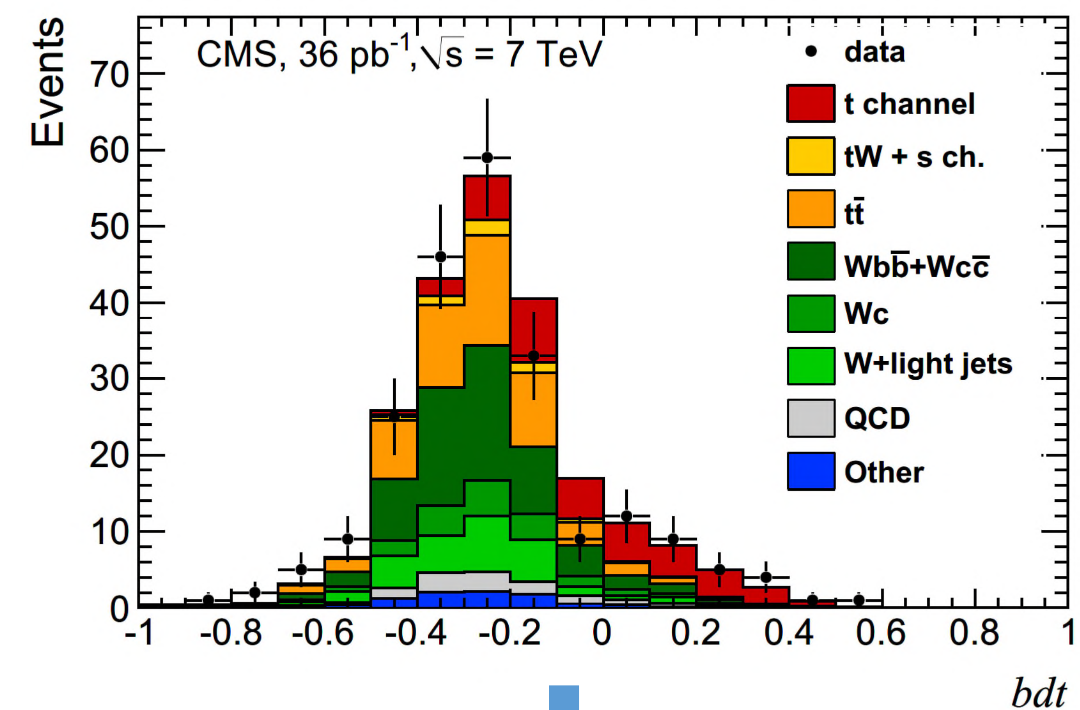
Expected at LHC
Luminosity = **1 fb^{-1}**

Luminosity = **36 pb^{-1}**





Machine Learning (37 variables):
Boosted Decision Tree



Classic

$124.2 \pm 33.8^{+30.0}_{-33.9}$ pb

Machine Learning

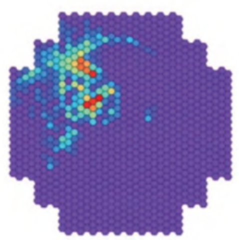
$78.7 \pm 25.4^{+13.2}_{-14.6}$ pb

Combined cross section $\sigma = 83.6 \pm 30.0$ pb

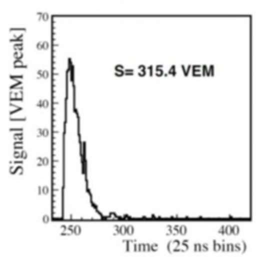
✱ Theorist: "I almost fell out of my seat when I saw this."

Outlook and Perspectives After a Year of LHC - John Campbell DIS 2011

Data Analysis: towards deep learning

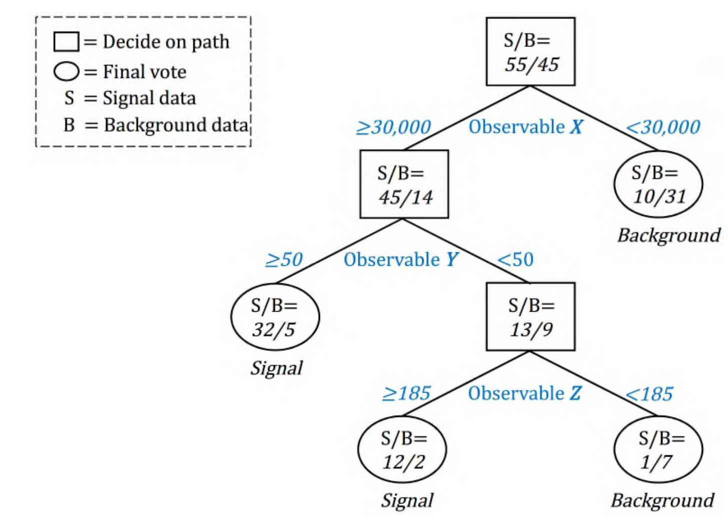


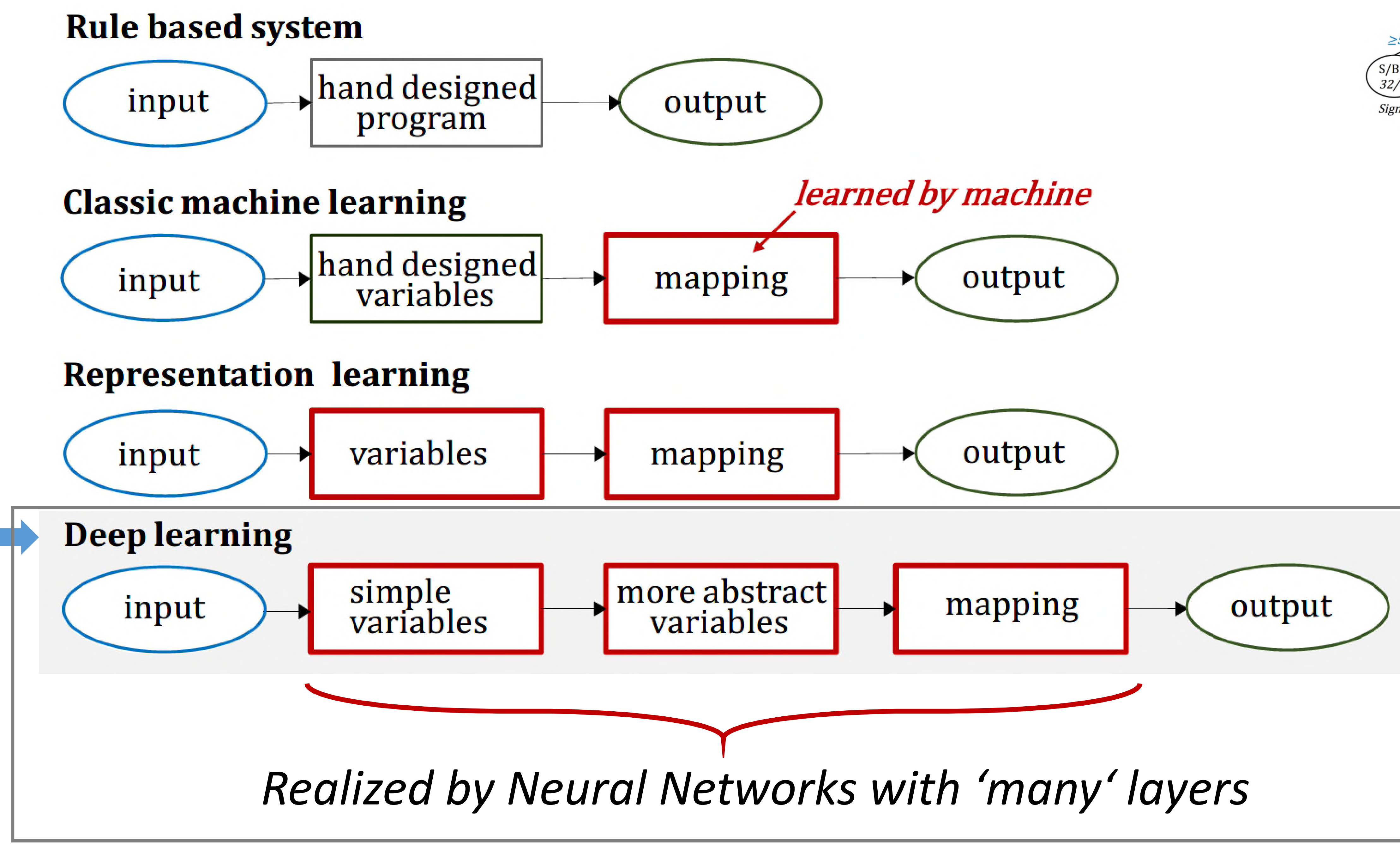
(a)



(c)

Boosted Decision Tree





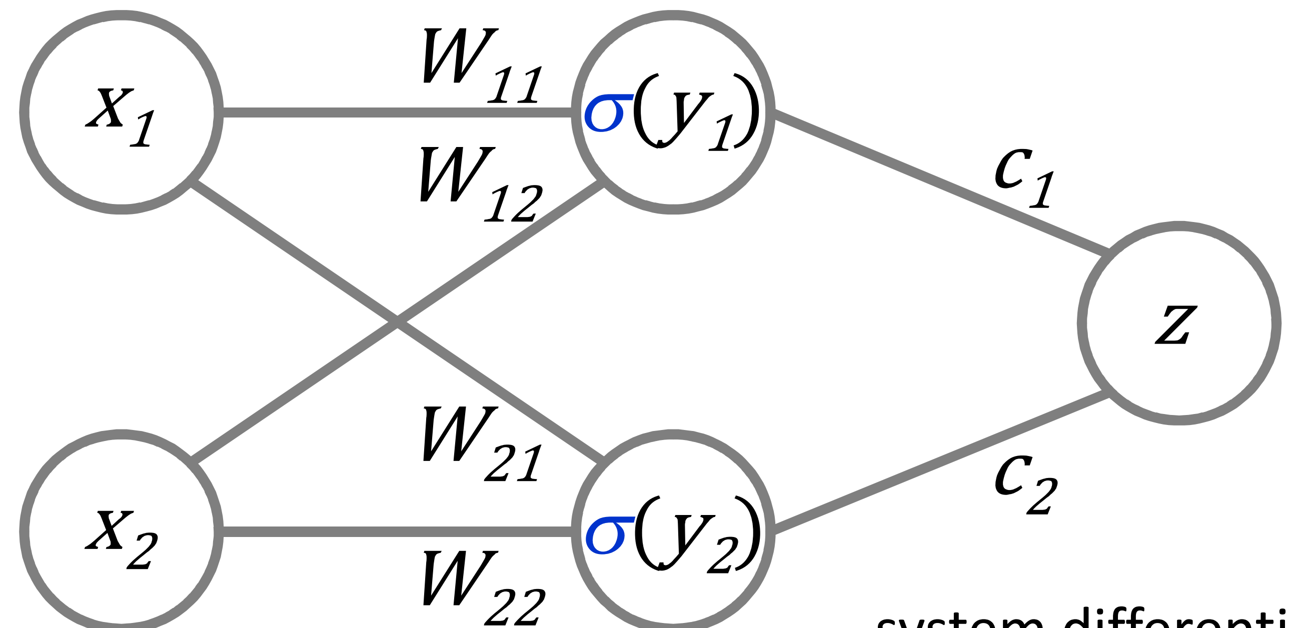
variable = *feature*

McCulloch, W.S., Pitts, W.: Bulletin of Mathematical Biophysics (1943) 5: 115.
Frank Rosenblatt, Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC, 1961

Neural Network Operations

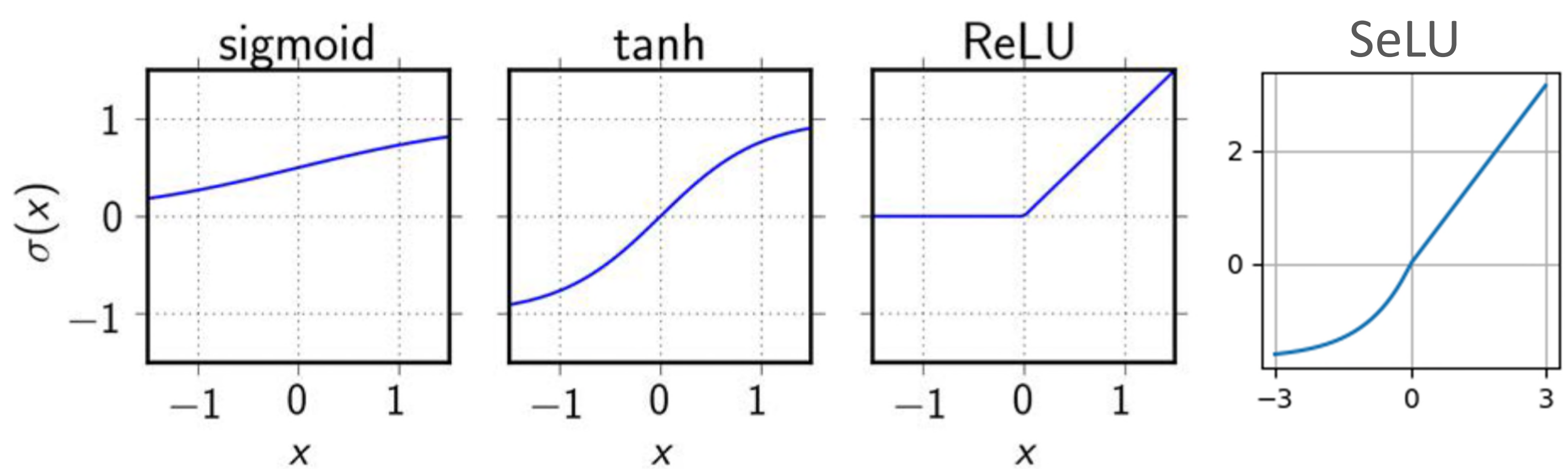
x multi-dimensional input data
W, b to be trained
successively apply 2 operations:

$$y = Wx + b$$
$$h = \sigma(y)$$



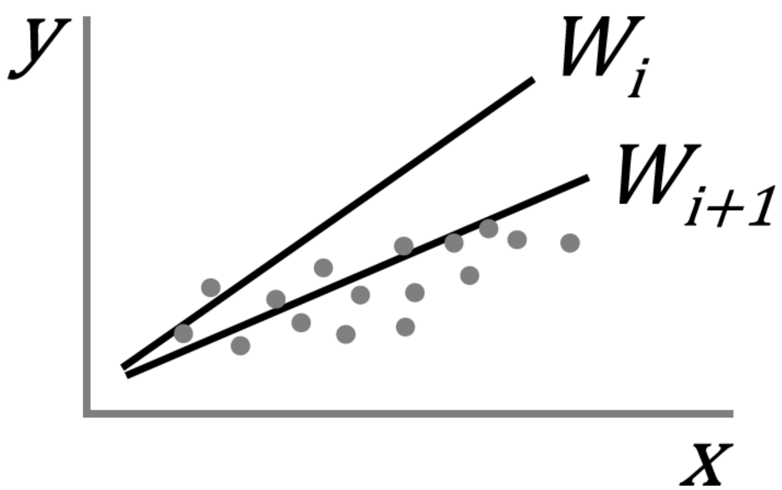
system differentiable

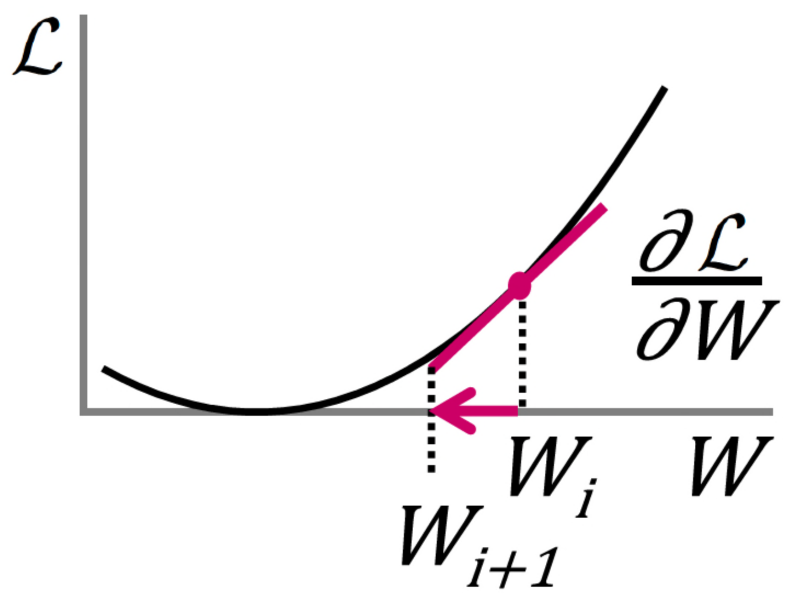
activation function: departure from linear system



Neural Network Training

('supervised')





- **Data** set

  $\{ x_i , y_i \} \quad i = 1 , \ldots , N$

- Define **model**

  $y_m (x) = W x + b$

- Define **objective** function (=**loss, cost**)

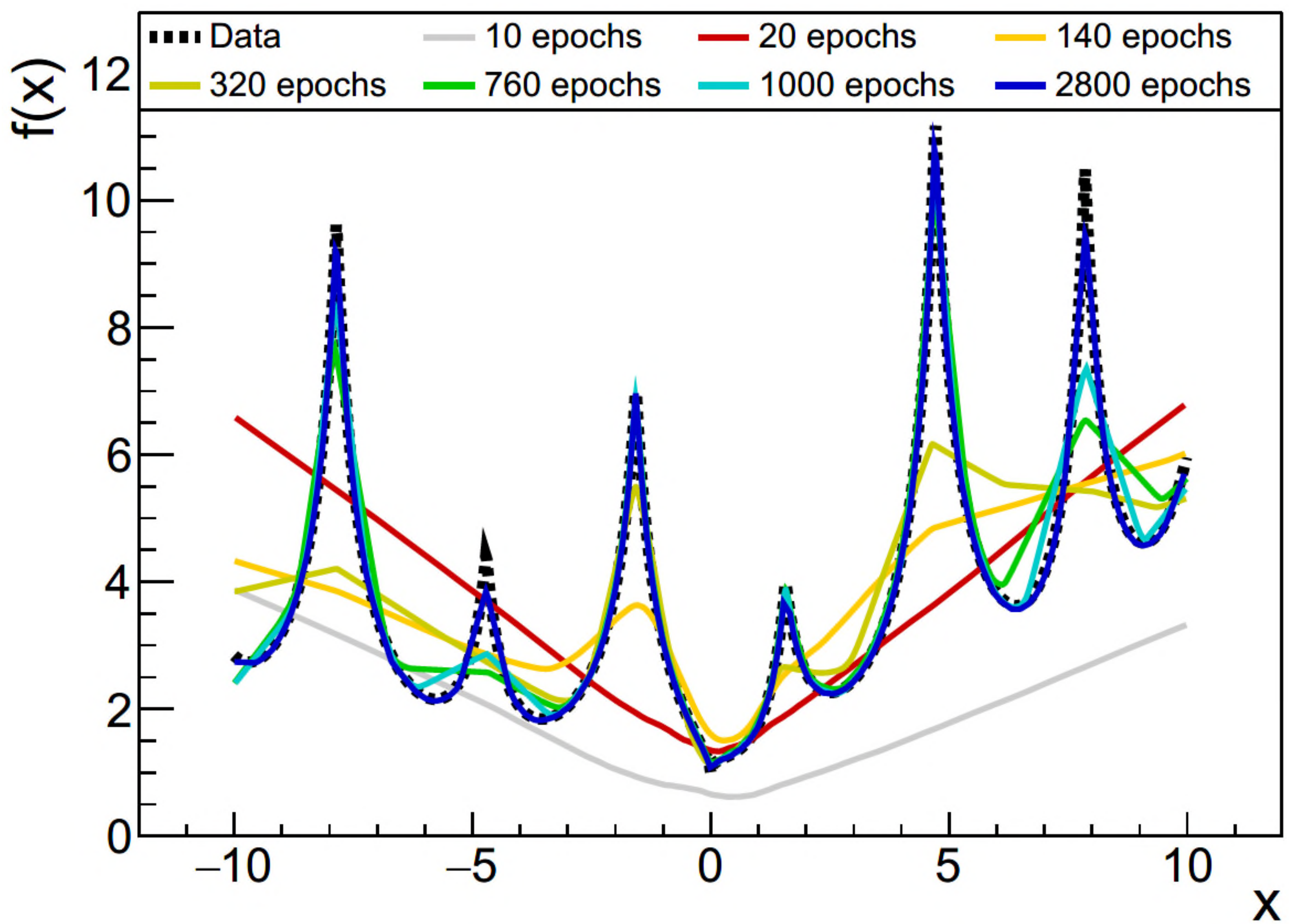
  $$\mathcal{L}(W, b) = \frac{1}{N} \sum_{i=1}^{N} [y_m (x_i) - y_i]^2$$

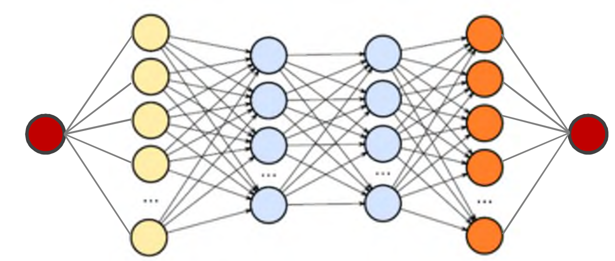
- **Train** model by optimizing the parameters

  $$(\hat{W} , \hat{b}) = \arg\min \mathcal{L}(W, b)$$

Automated parameterization of arbitrary function

T. Quast CERN/RWTH Aachen



x ∈ [-10,10]  f(x)

7 hidden layers
200 nodes each
ReLU activation function

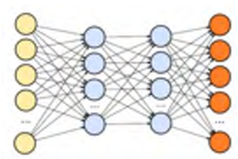
original function (black symbols):
fair description after 2800 training steps (purple)

- *Reality: function working in multi-dimensions*
  $$\vec{x} \in \mathbb{R}^n \rightarrow \vec{z} \in \mathbb{R}^m$$
- *Function: training is million-parameter fit*

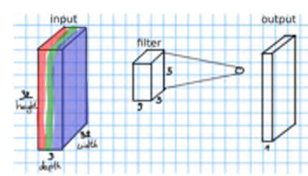
Deep Learning Progress

Concepts

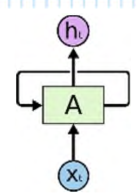
- Fully connected



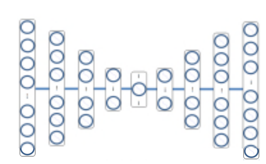
- Convolutional



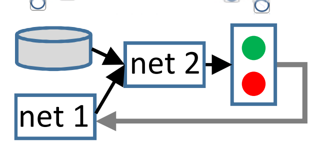
- Graph
- Recurrent



- Lorentz Boost Network
- Autoencoder



- Adversarial



- Reinforcement
- Invertible

Improved set of tools

Train millions of parameters by:

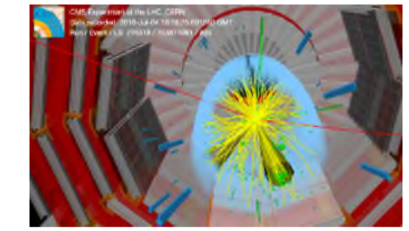
- Data preprocessing
- Normalization
- Regularization
- Short cuts …

Computing

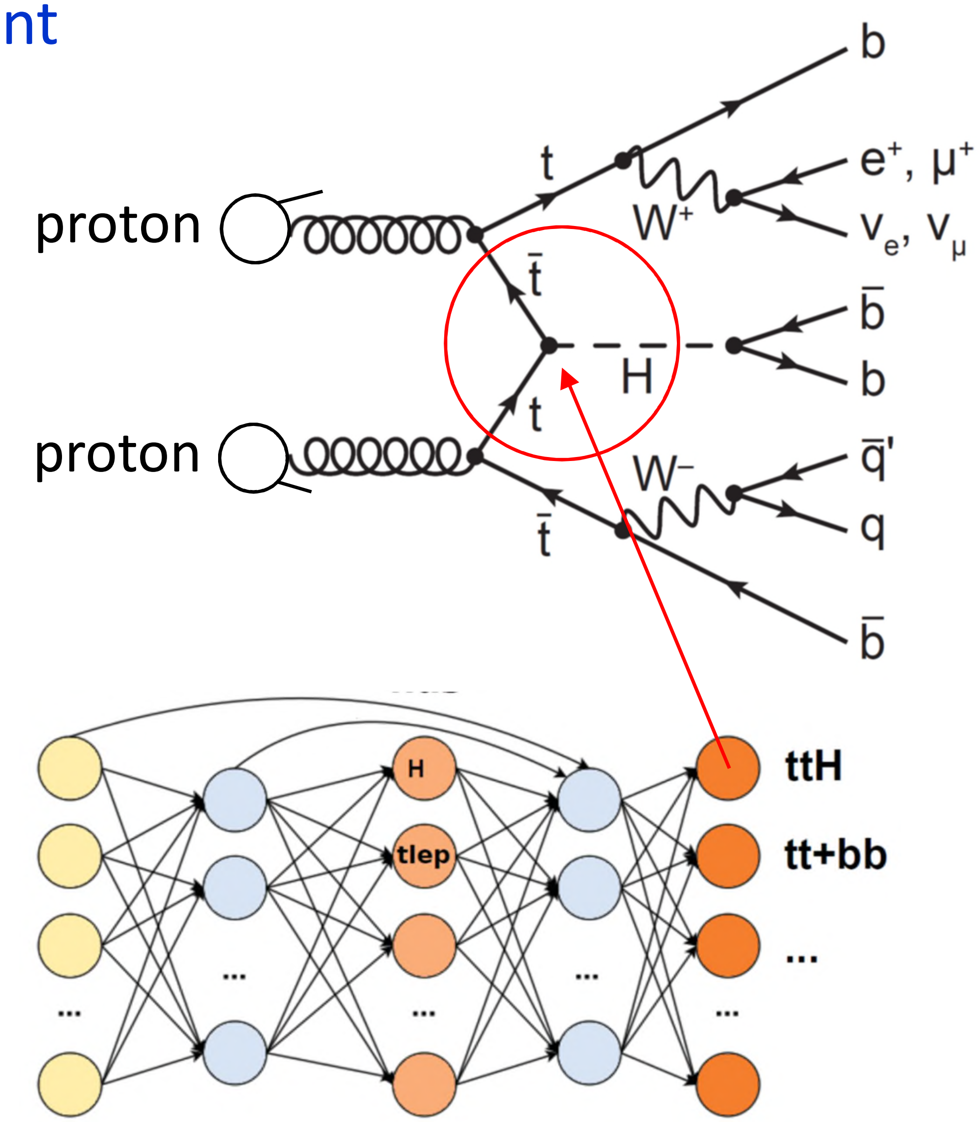
- Graphics Processing Unit (GPU)
- Software Libraries
  - TensorFlow
  - keras…

CMS Collaboration, Phys. Rev. Lett. 120, 231801 – Published 4 June 2018

1. Fully connected networks
LHC: Coupling Top-Quark – Higgs Boson

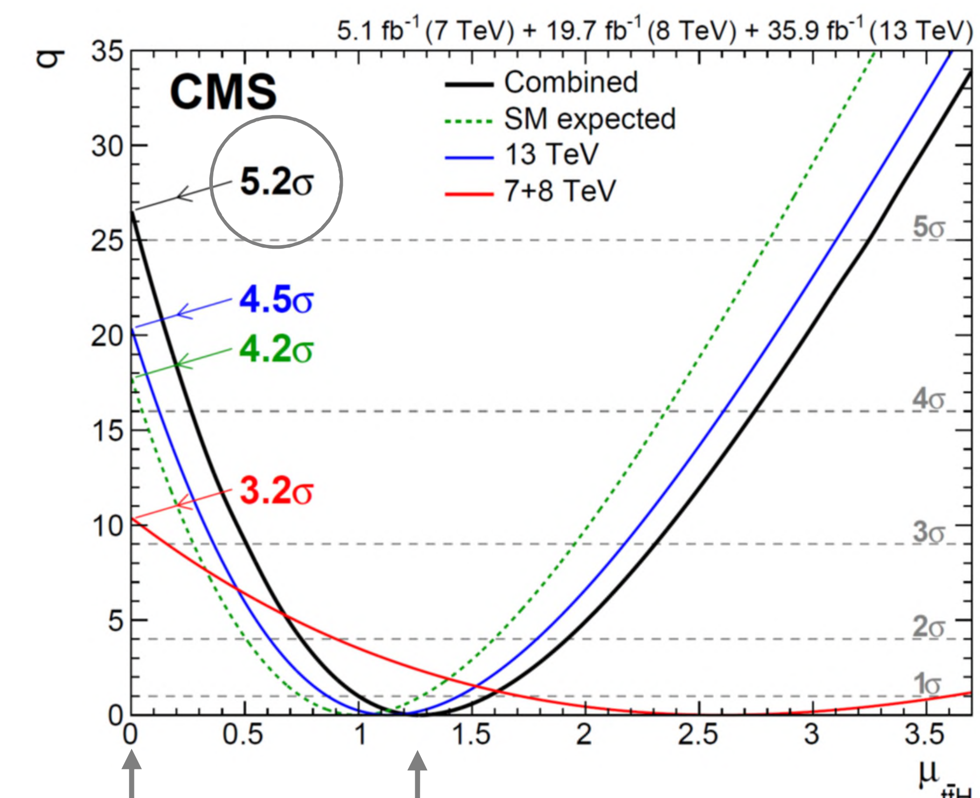


Deep Learning predicts physics process for each event



Observation of ttH production

H→ bb
H→ WW
H→ ZZ
H→ γγ
H→ ττ



excluded:
no coupling

Measured signal
close to Standard Model
of particle physics (μ=1)

M. Zeiler, R. Fergus, Stochastic pooling for regularization of deep convolutional neural networks. ICLR, 2013.

2. Convolutional networks to analyse image-like data

Hierarchy of ’features’ with increasing level of complexity

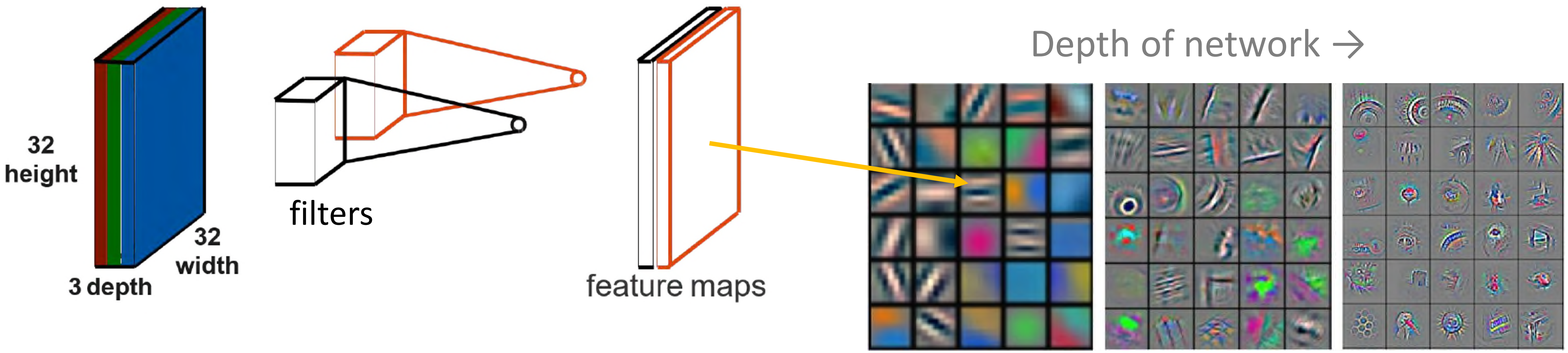


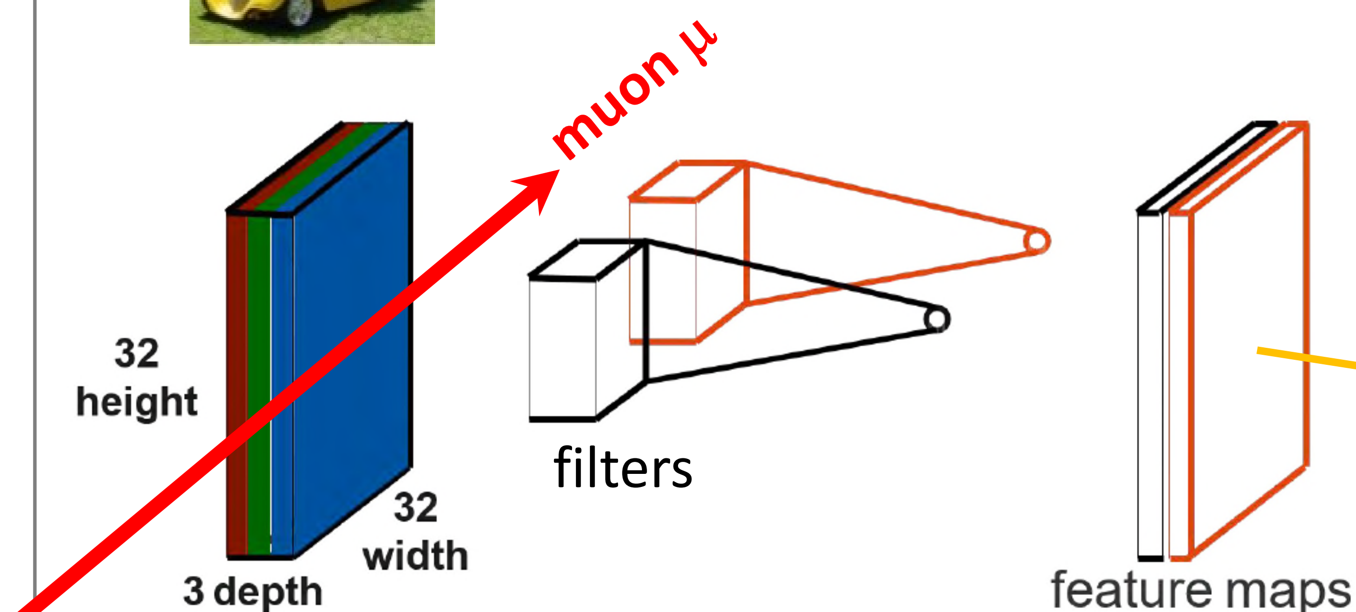
Image recognition

Pixel → edge → texton → motif → part → object

M. Zeiler, R. Fergus, Stochastic pooling for regularization of deep convolutional neural networks. ICLR, 2013.

2. Convolutional networks to analyse image-like data

Hierarchy of 'features' with increasing level of complexity



Depth of network →

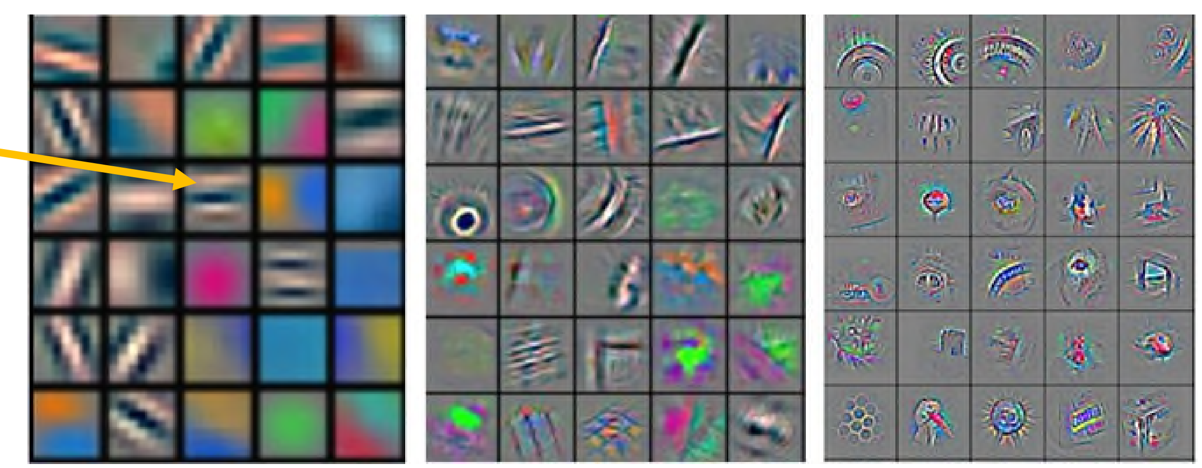
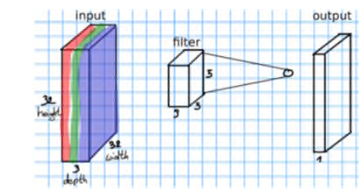


Image recognition

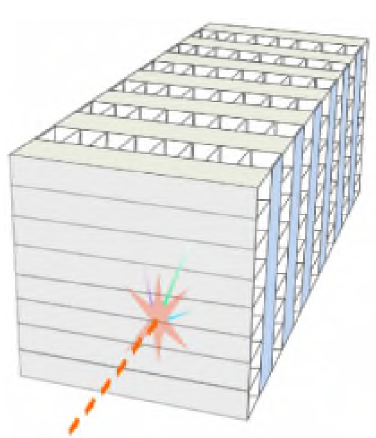
Pixel → edge → texton → motif → part → object

A. Aurisano et al., JINST 11 (2016) P09001

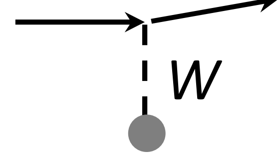


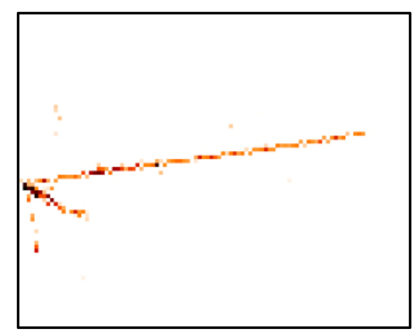
Electron neutrino identification

Fermilab (Chicago)
→NOvA experiment 810km



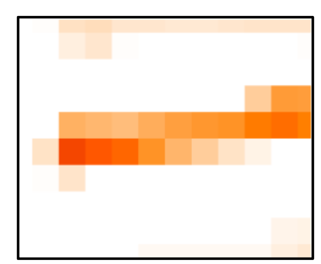
Neural network neutrino event classifier





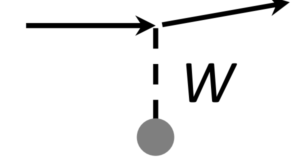
Feature maps

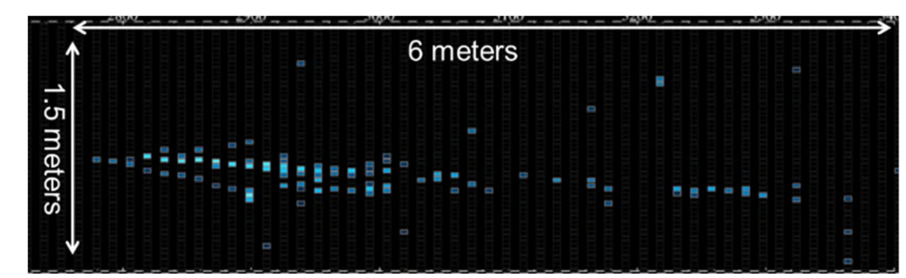
track



hadronic

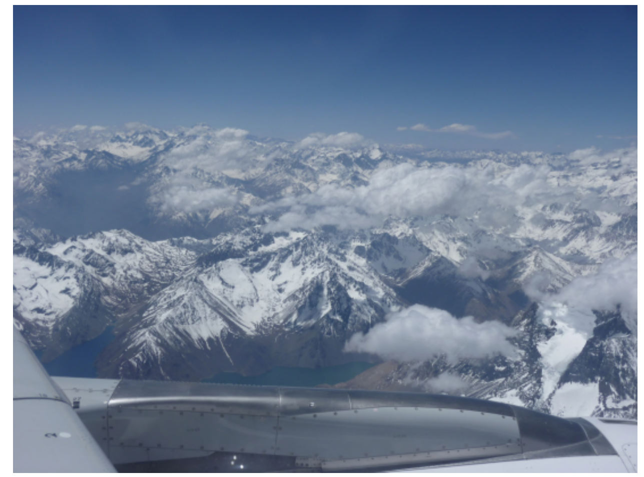
Challenge: electron-neutrinos

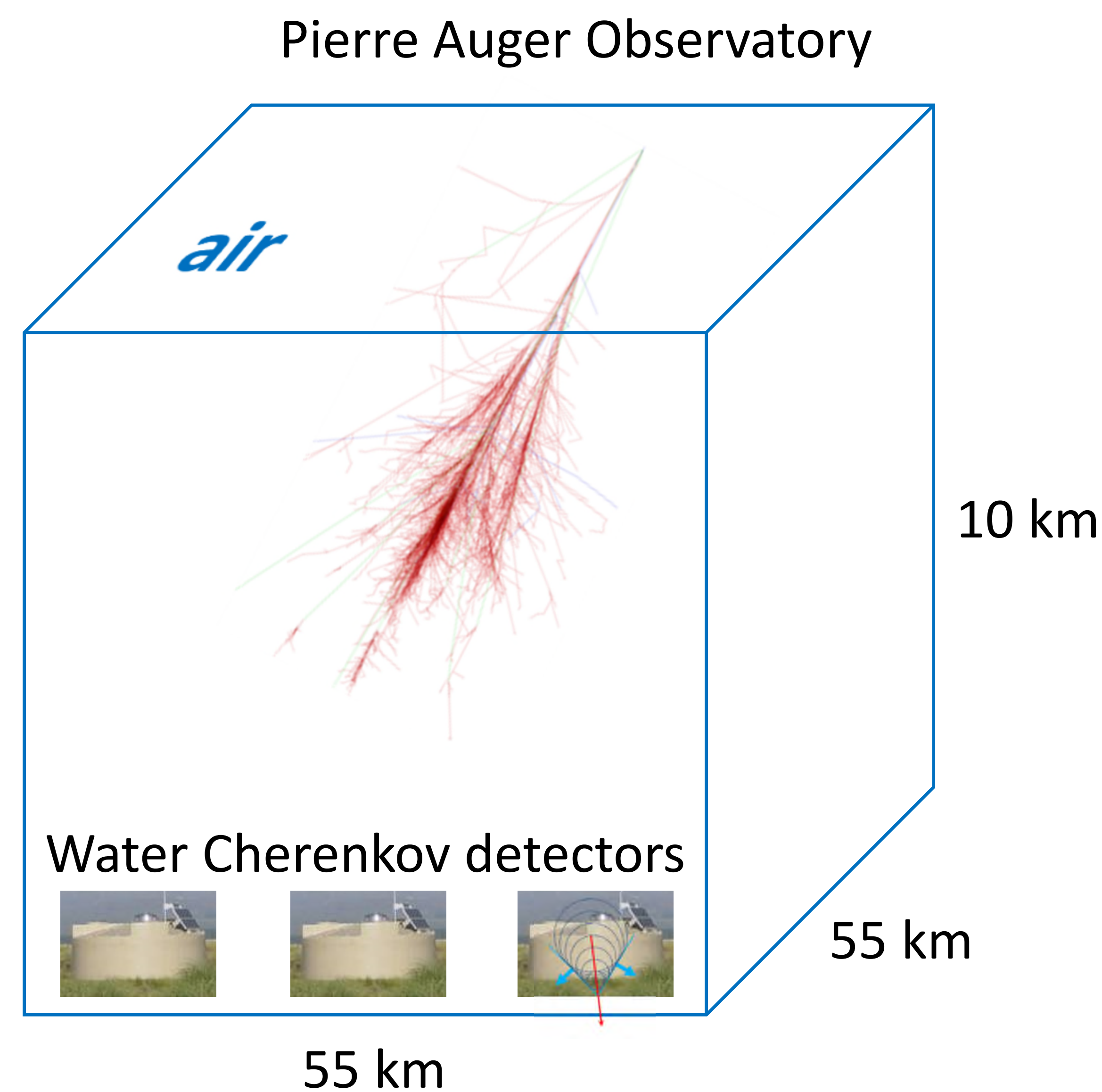




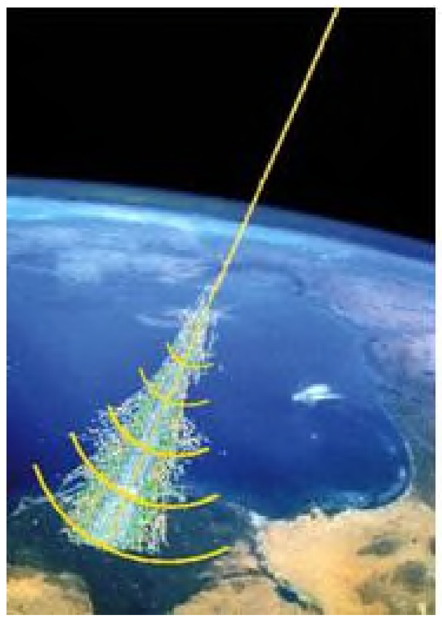
Method	**ν_e efficiency** (same purity)
Physicists algorithm	**35%**
Deep learning neural network	**49%**

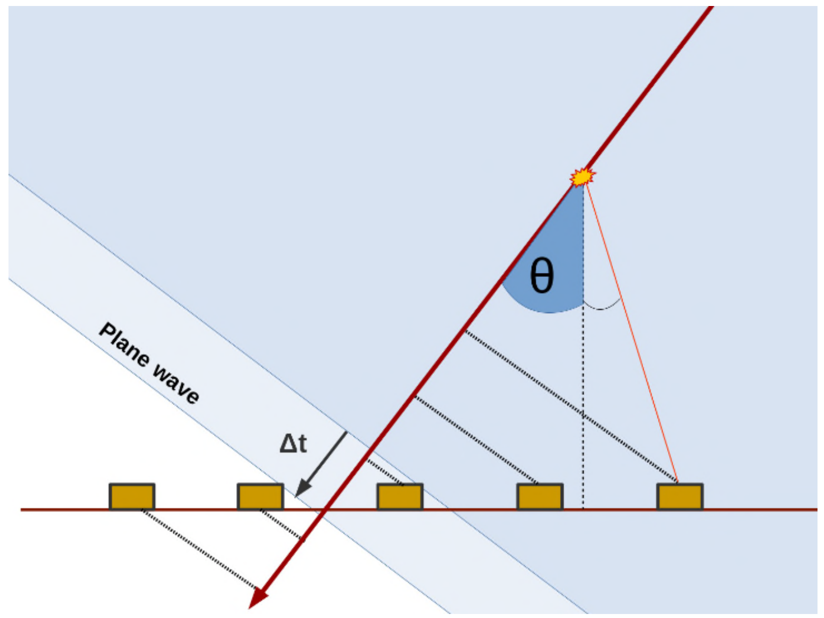
World‘s largest Calorimeter for Cosmic Rays

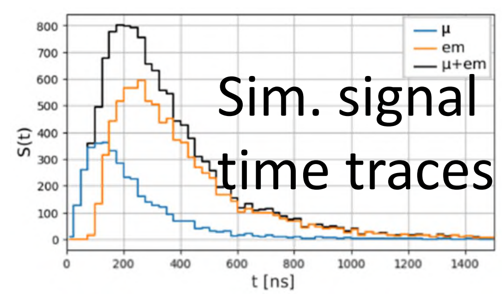




Calorimeter: cosmic ray induced air showers



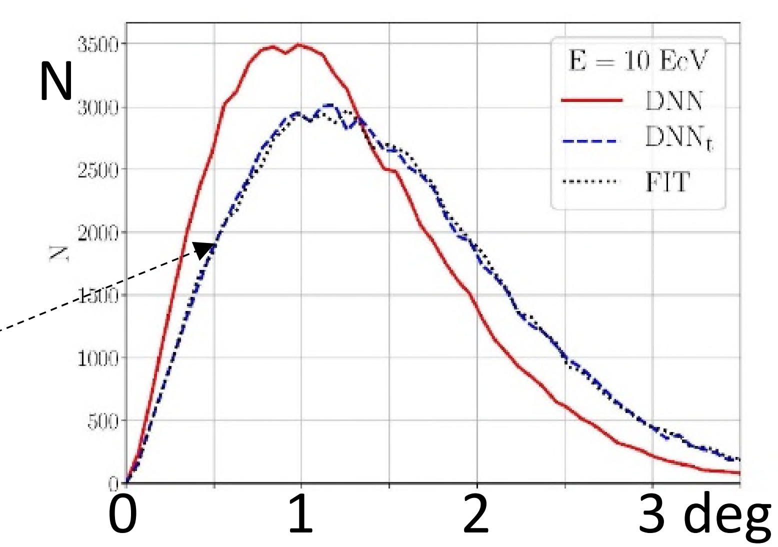




Educated physicist:

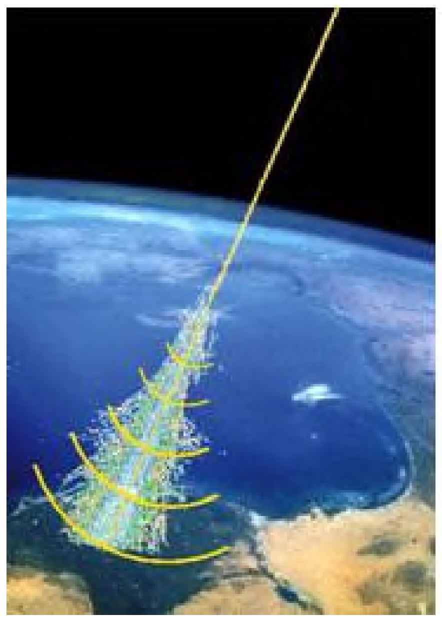
- Time offset between detectors
- Particle velocity
- Detector distance
- Plane fit of time residuals

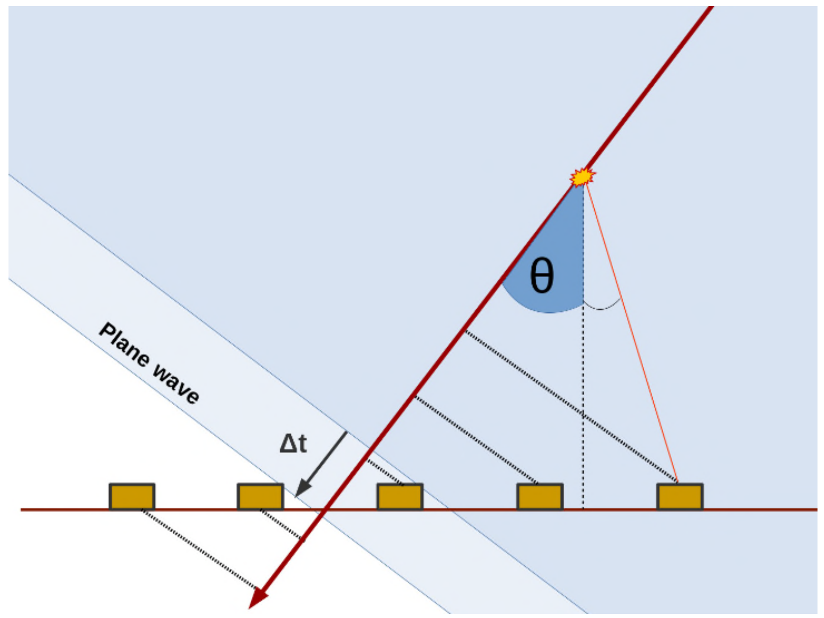
shower direction angular resolution

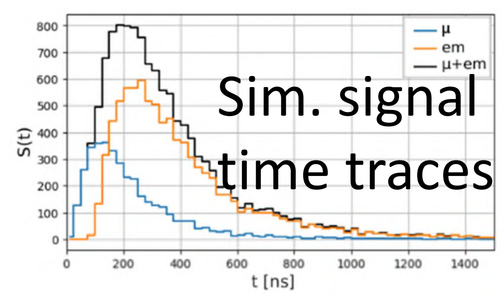


M.Erdmann, J. Glombitza, D. Walz, 10.1016/j.astropartphys.2017.10.006

Calorimeter: cosmic ray induced air showers



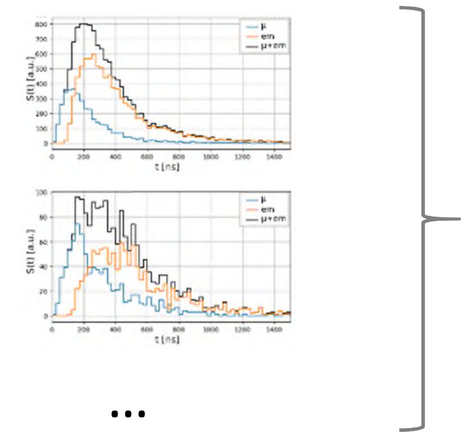




Educated physicist:

- Time offset between detectors
- Particle velocity
- Detector distance
- Plane fit of time residuals

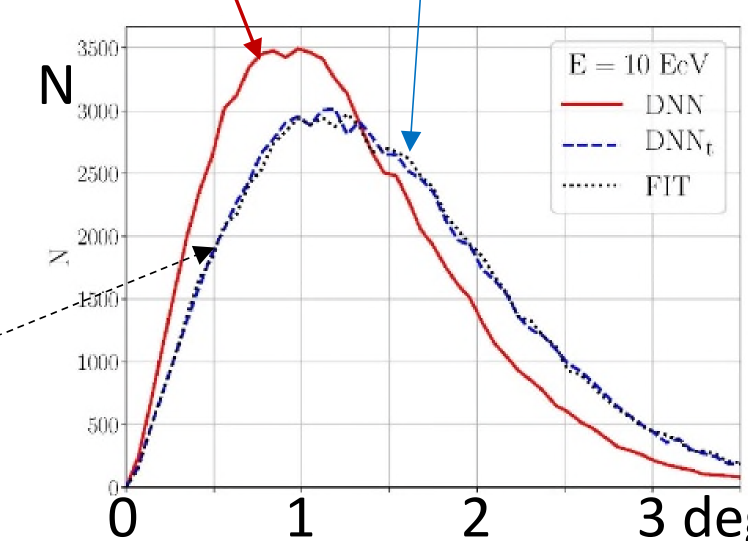
RAW input data:



Deep Neural Network

- Time offset only
- Black signal traces added

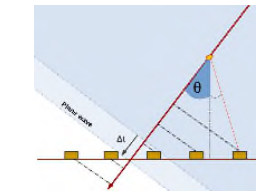
shower direction angular resolution



Deep Neural Network learns physics from data within 3h

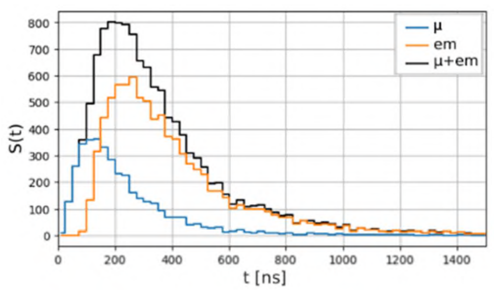
M.Erdmann, J. Glombitza, D. Walz, 10.1016/j.astropartphys.2017.10.006

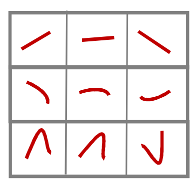
Calorimeter: cosmic ray induced air showers



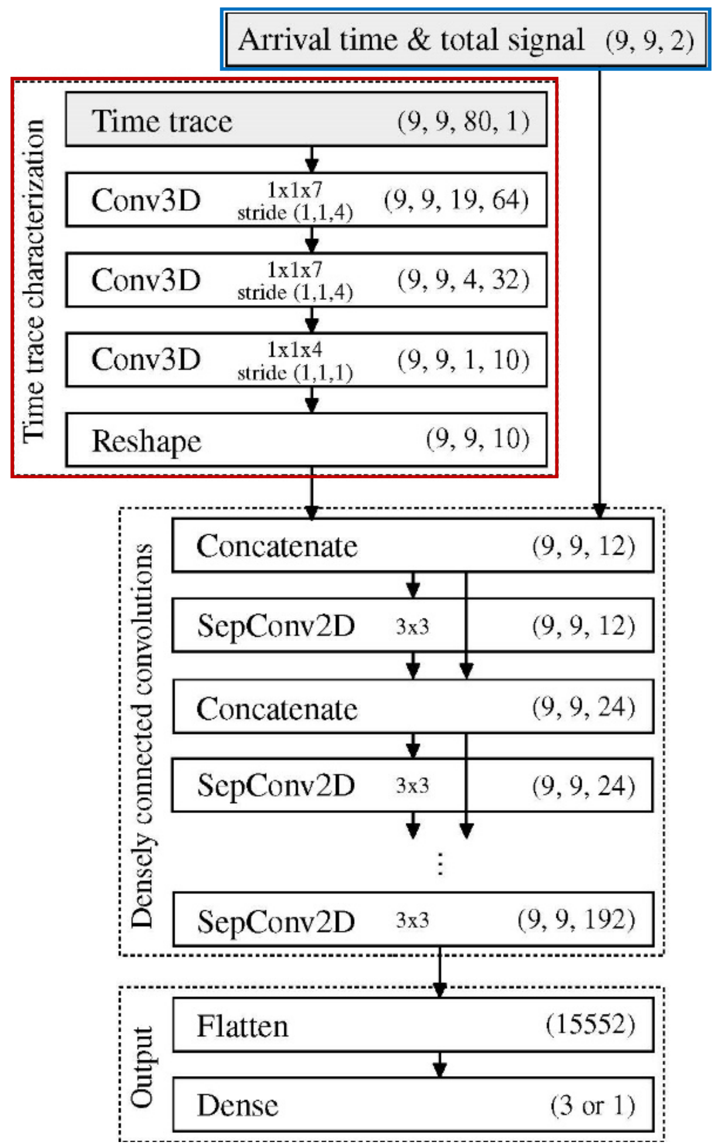
Time offset & total signal

Characterization
of signal traces





features

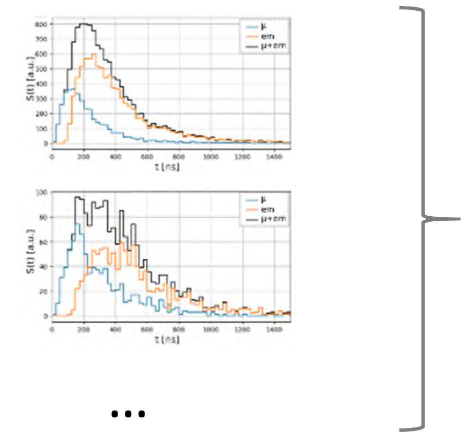


shower direction

Network extracts from training data *optimized intermediate variables* suited for shower direction

120k parameters

RAW input data:

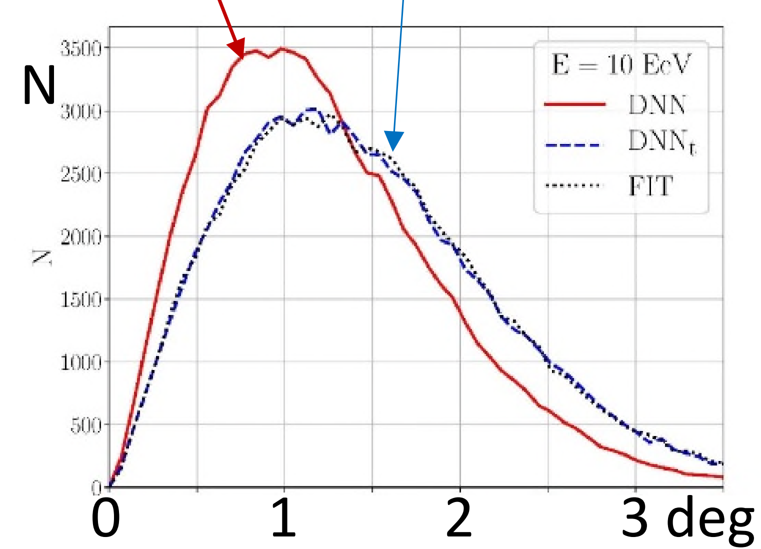


...

Deep Neural Network

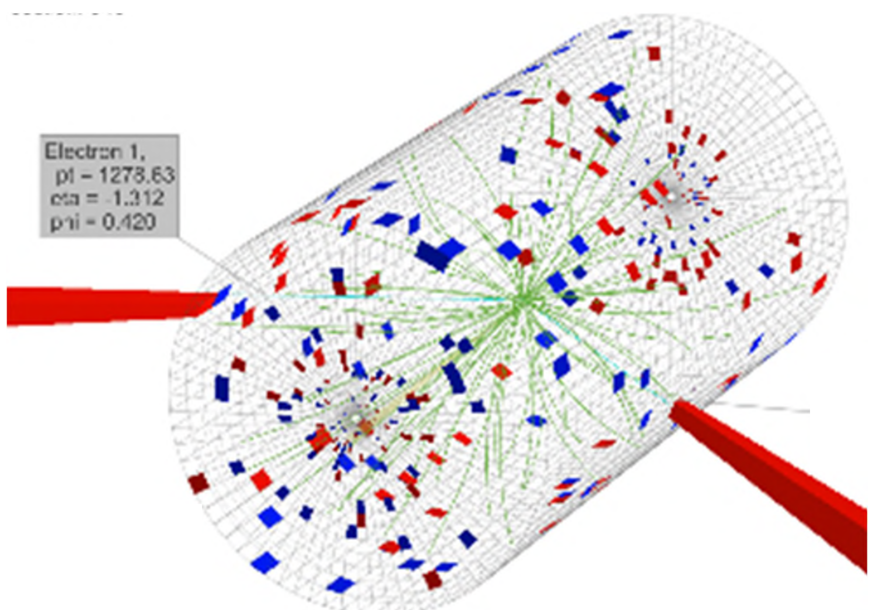
- Time offset only
- Black signal traces added

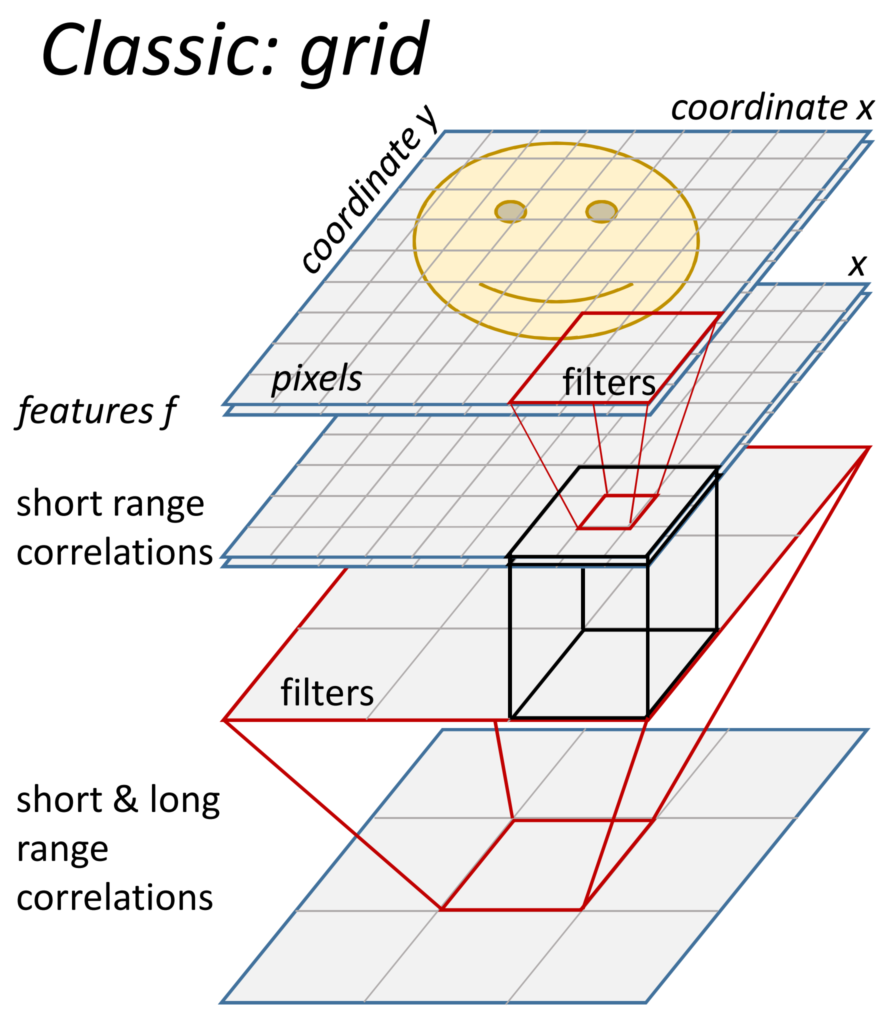
shower direction angular resolution



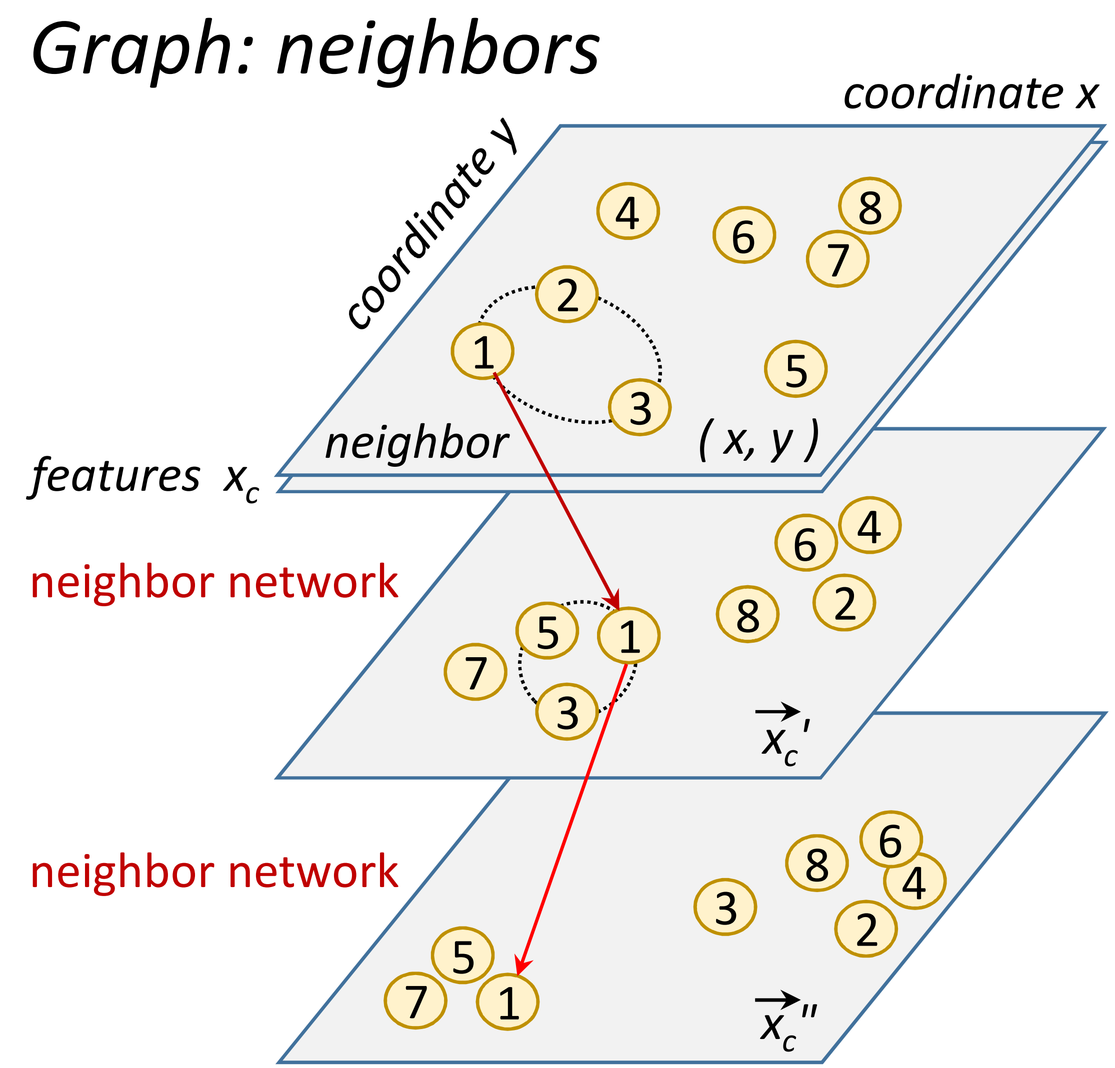
Deep Neural Network learns physics from data within 3h

3. Convolution: *Classic* versus *Graph* network





- dense pixel occupation
- object identification
- analysis: single pixel contributions



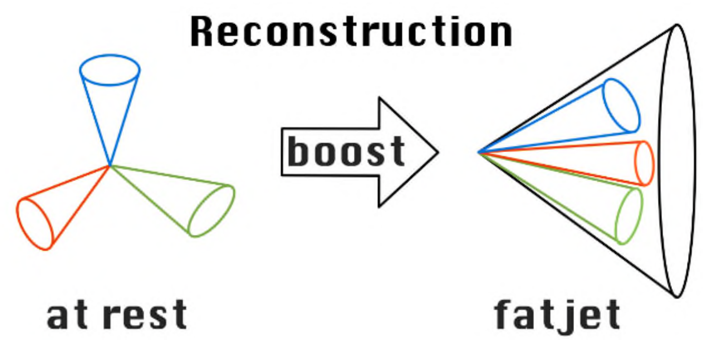
- sparse or dense neighbor occupation
- pattern recognition
- single objects 1 identification

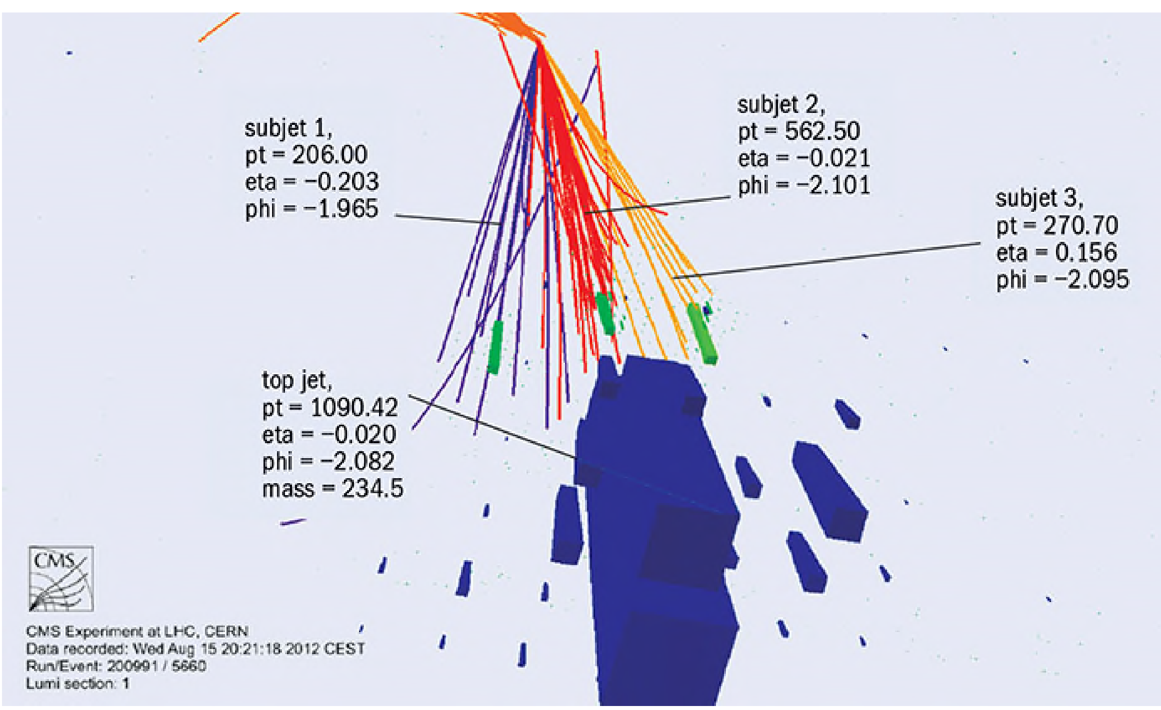
Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, Justin M. Solomon, arXiv:1801.07829

The 2019 Top Quark Recognition Challenge

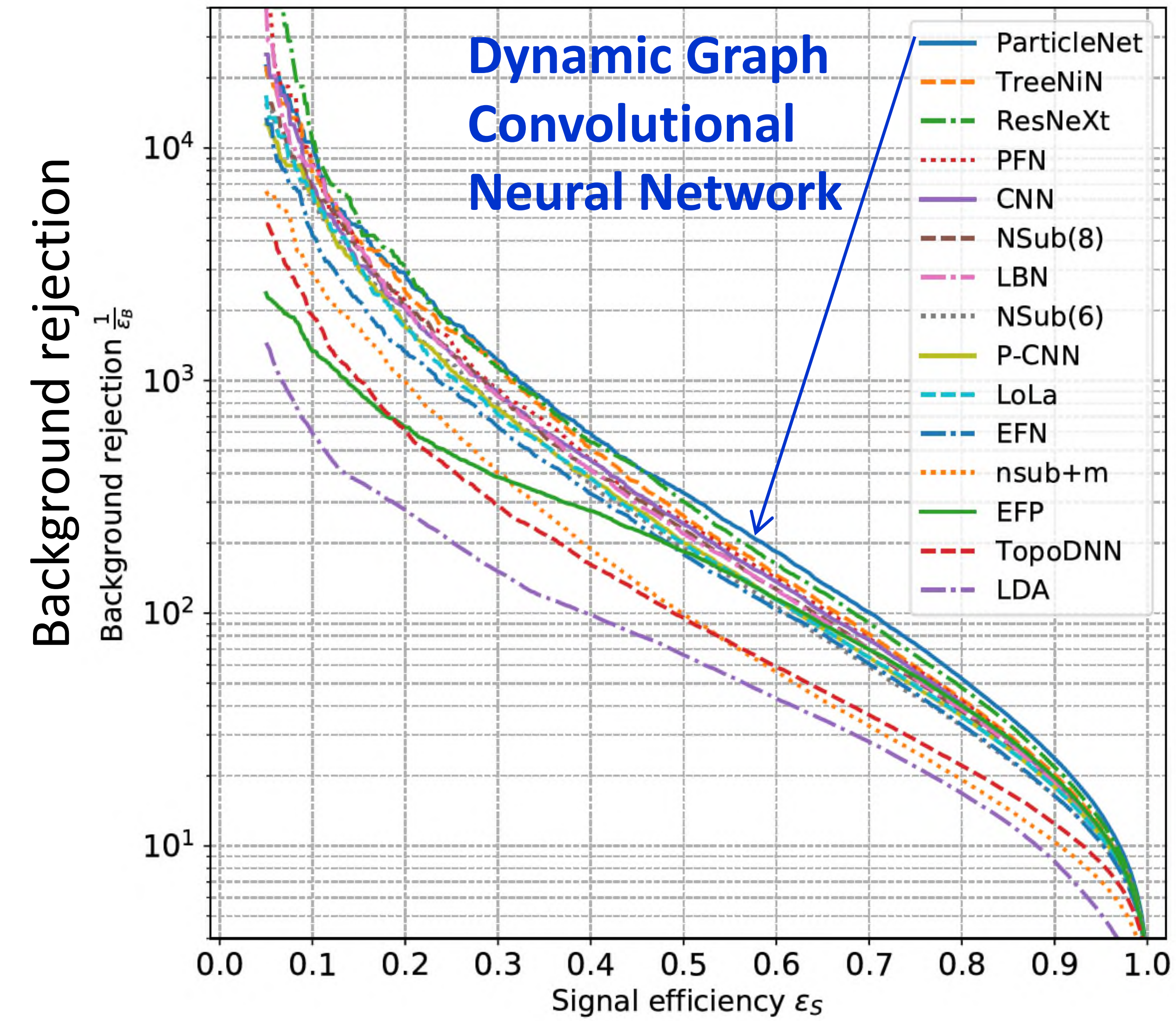
G. Kasieczka, T. Plehn et al, SciPost Phys. 7, 014 (2019)   Huilin Qu, L. Gouskos, arXiv:1902.08570

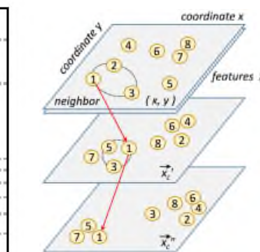
Challenge: identify top quark decay





Evaluation of top quark finders

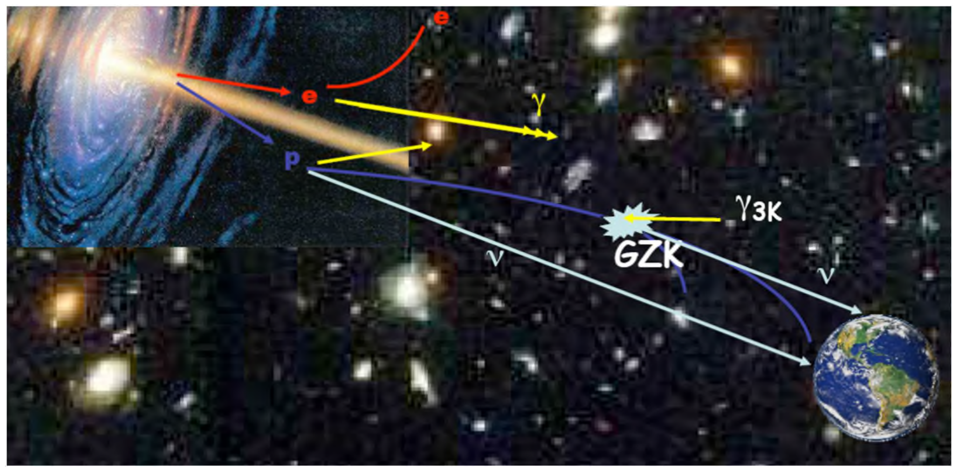




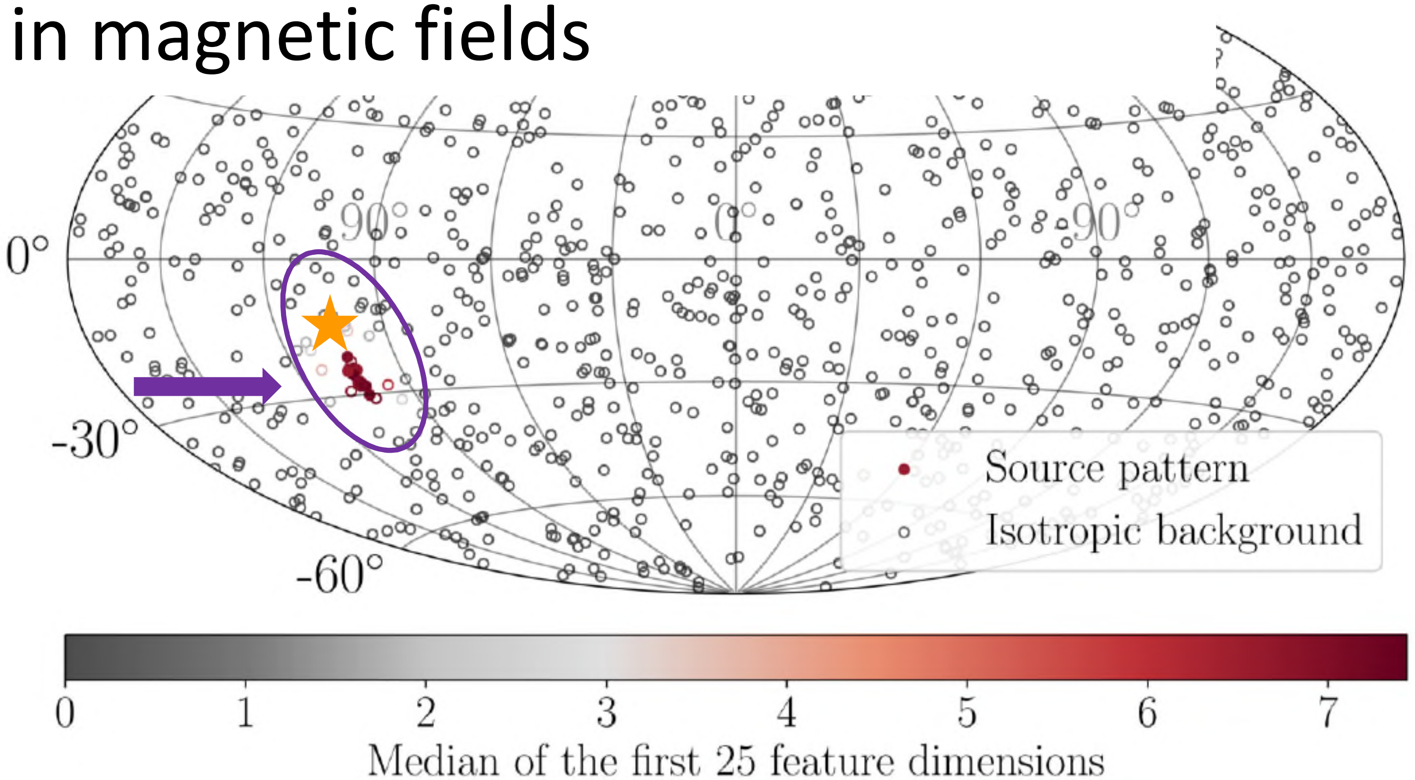
Top quark signal detection efficiency

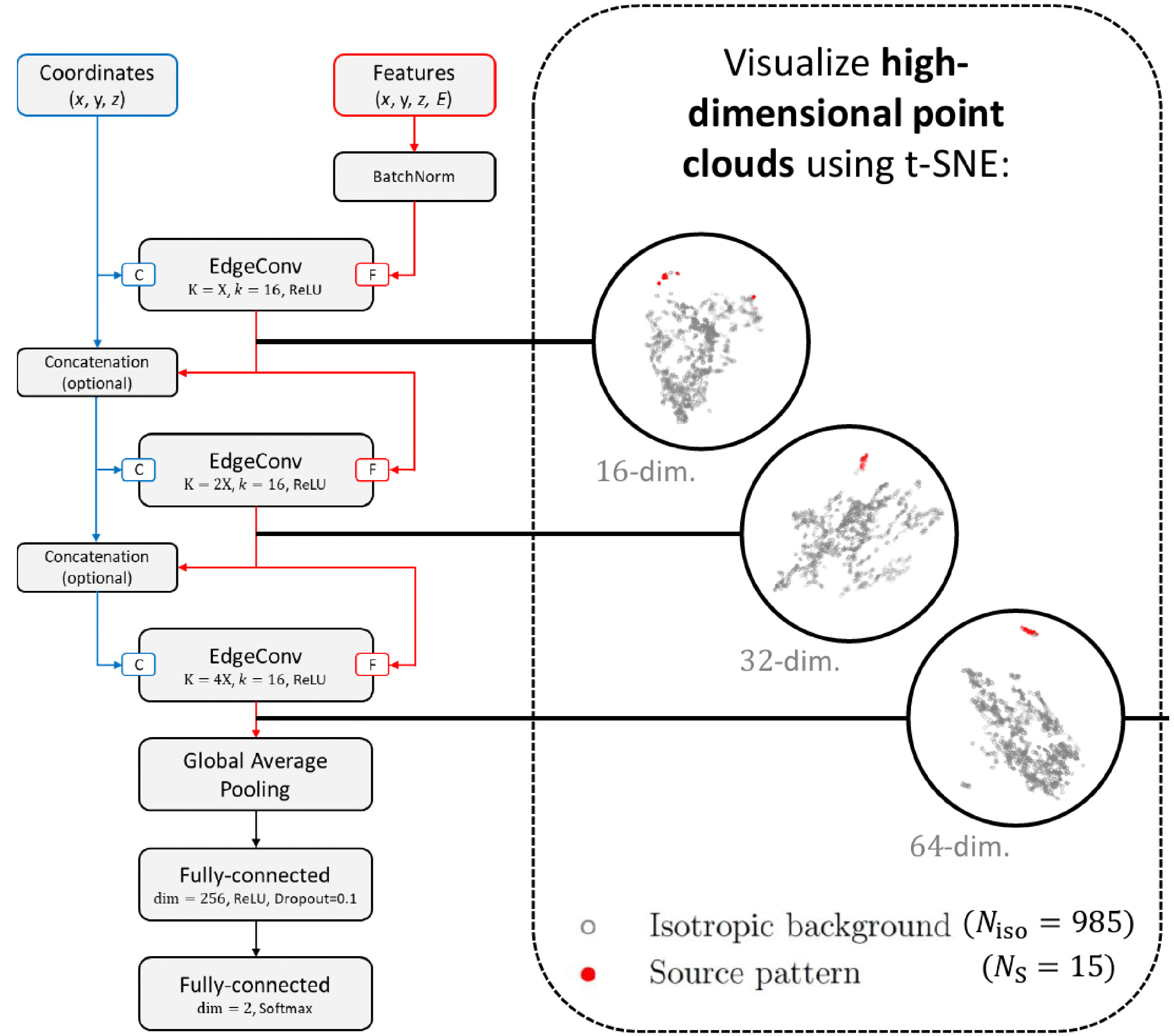
T. Bister, M. Erdmann, J. Glombitza, N. Langner, J. Schulte, M. Wirtz, arXiv:2003.13038

Deflections of cosmic rays in cosmic magnetic fields



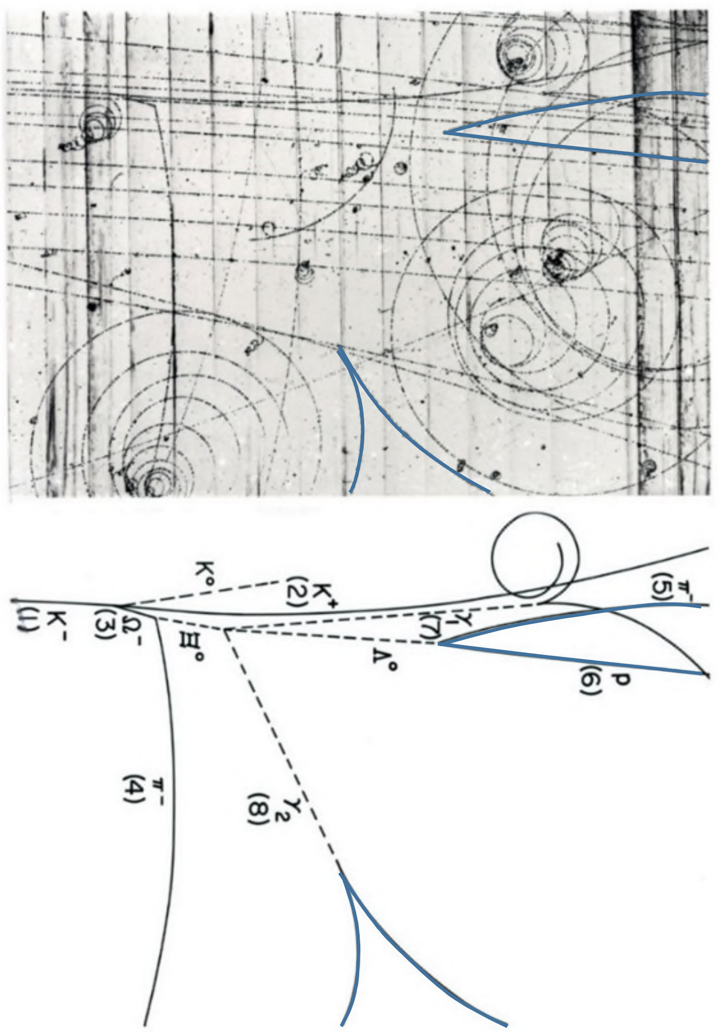
pattern: deflections of cosmic rays in magnetic fields





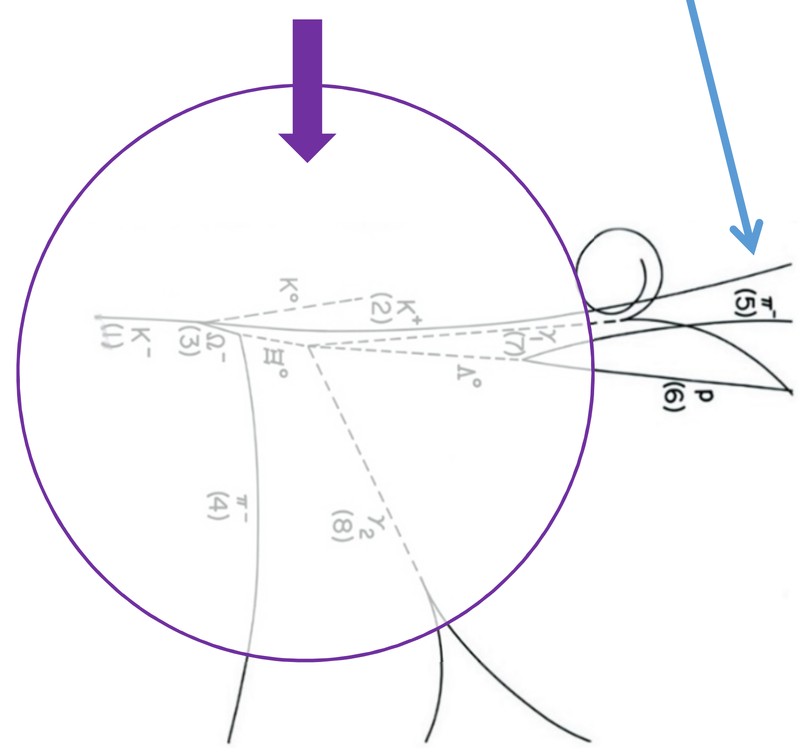
4. Recover interaction: **Lorentz Boost Network**

In classic bubble chamber experiments most of the interaction was visible

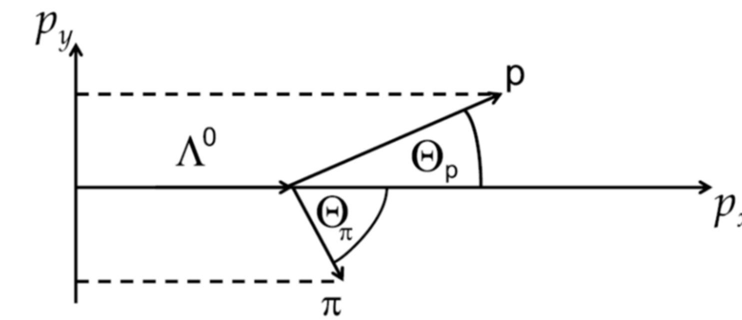


Todays experiments: reconstruct interaction from visible particles

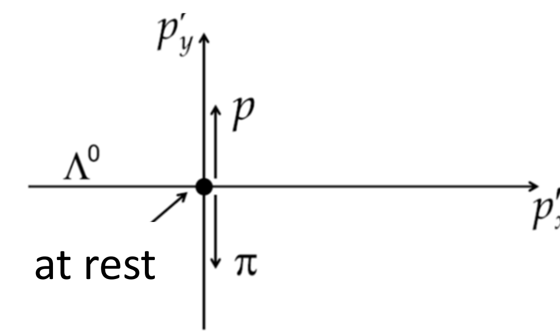
interaction at very small scales < 10^{-15} m remain invisible



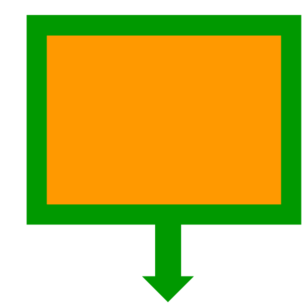
boosted particles



undo boost by Lorentz Transformation



Particle physics deep network architecture

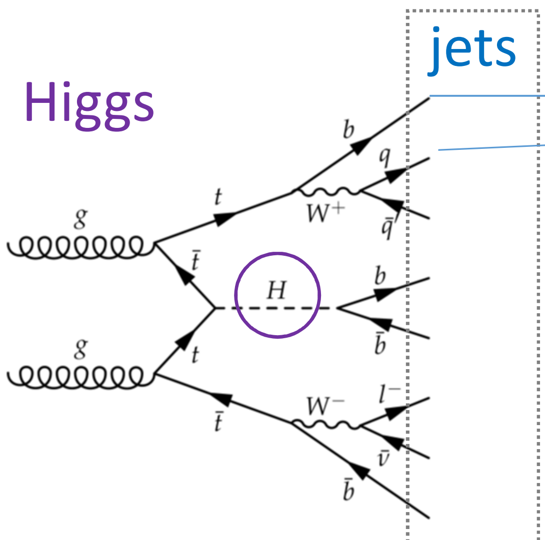


output: `high-level data' more smart data

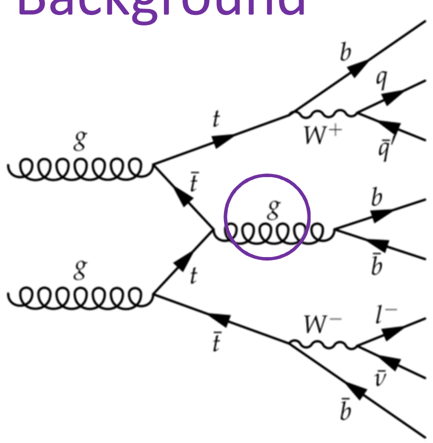
M. Erdmann, E. Geiser, Y. Rath, M. Rieger, JINST 14 (2019) P06006

Autonomous engineering of discriminating variables

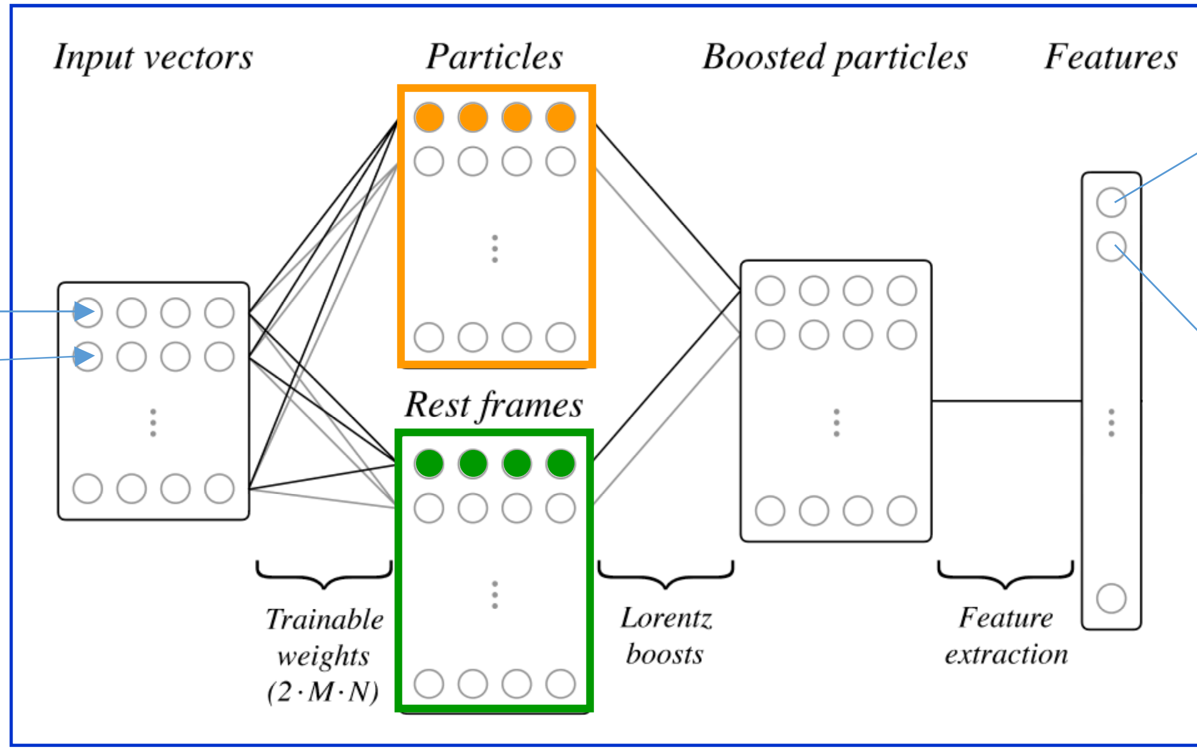
Higgs

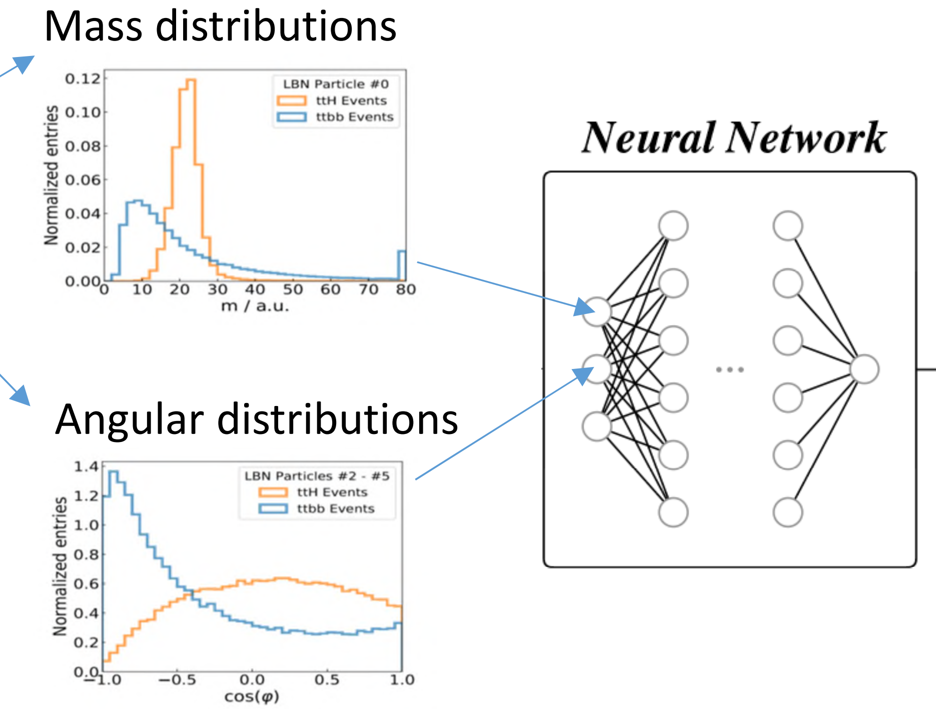


Background

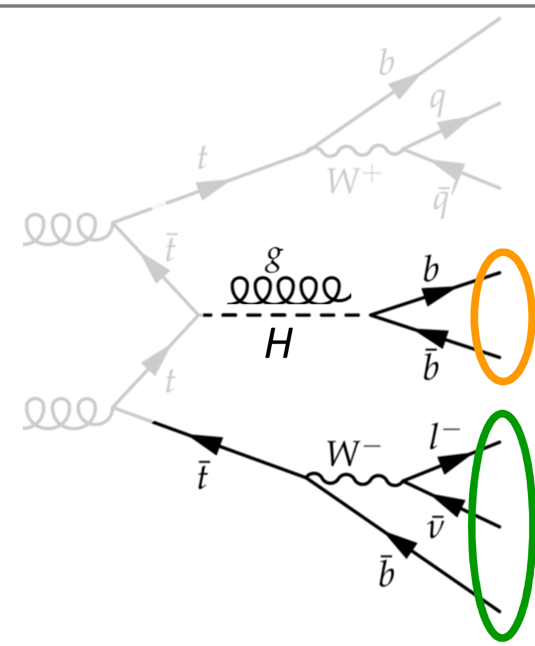


Lorentz Boost Network





Classification
Signal or
Background



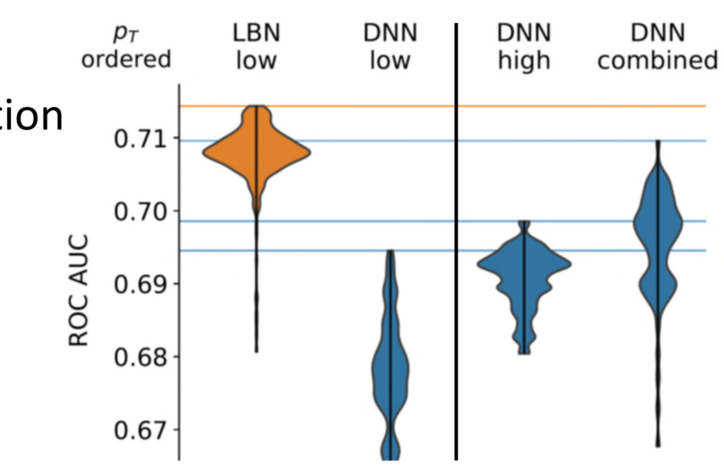
after training on simulated data:

network builds Higgs/gluon and

boosts it into rest frame of top quark

For data analysis: ***autonomous variables surpass human-made variables***

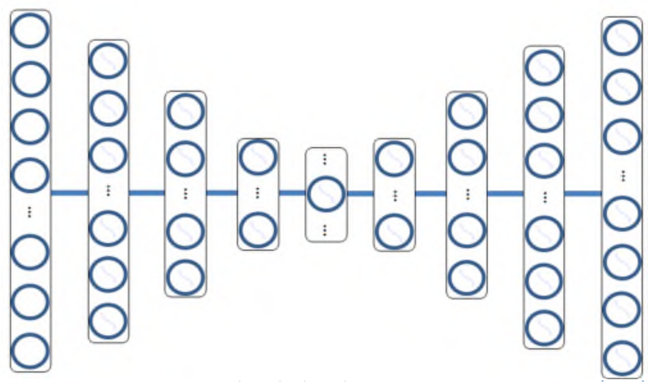
Classification
efficiency
& purity

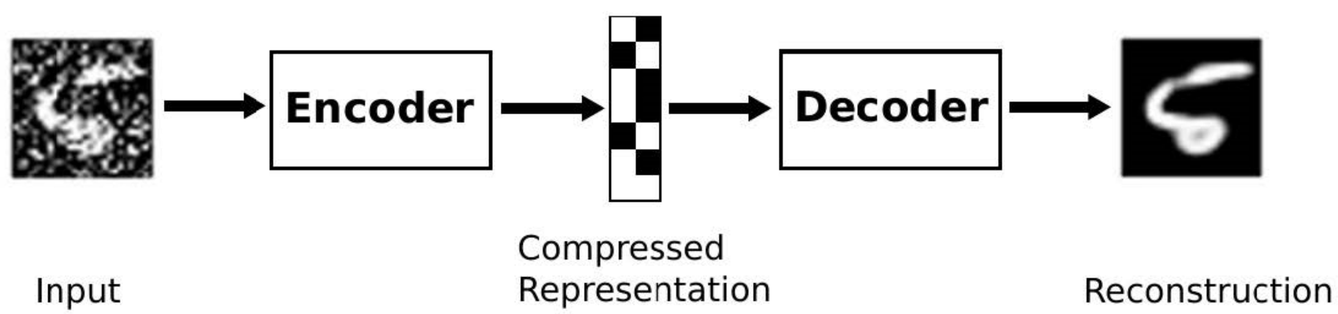


5. Autoencoder networks: inflate compressed signal

T. Heimel, G. Kasieczka, T. Plehn, J.M. Thompson, SciPost Phys. 6, 030 (2019)

Denoising

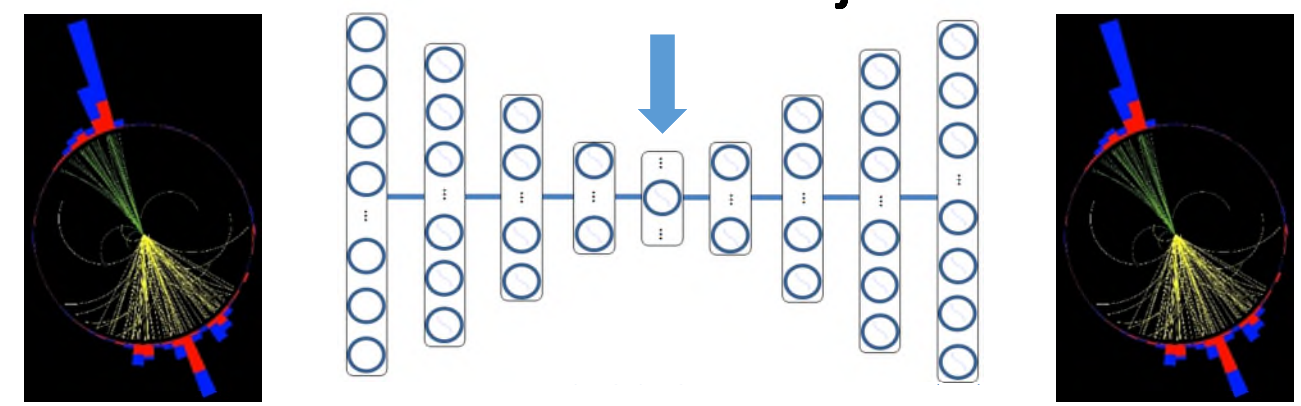




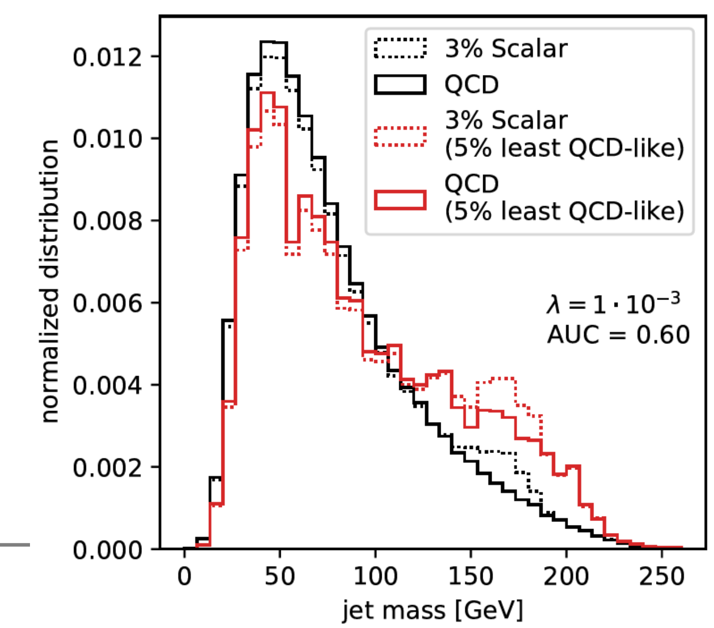
For training
compare
with input

Search for new physics

Train compressed representation:
Standard Model jets



Evaluation of jets:
***Disagreement** with*
input jet indicates
new physics

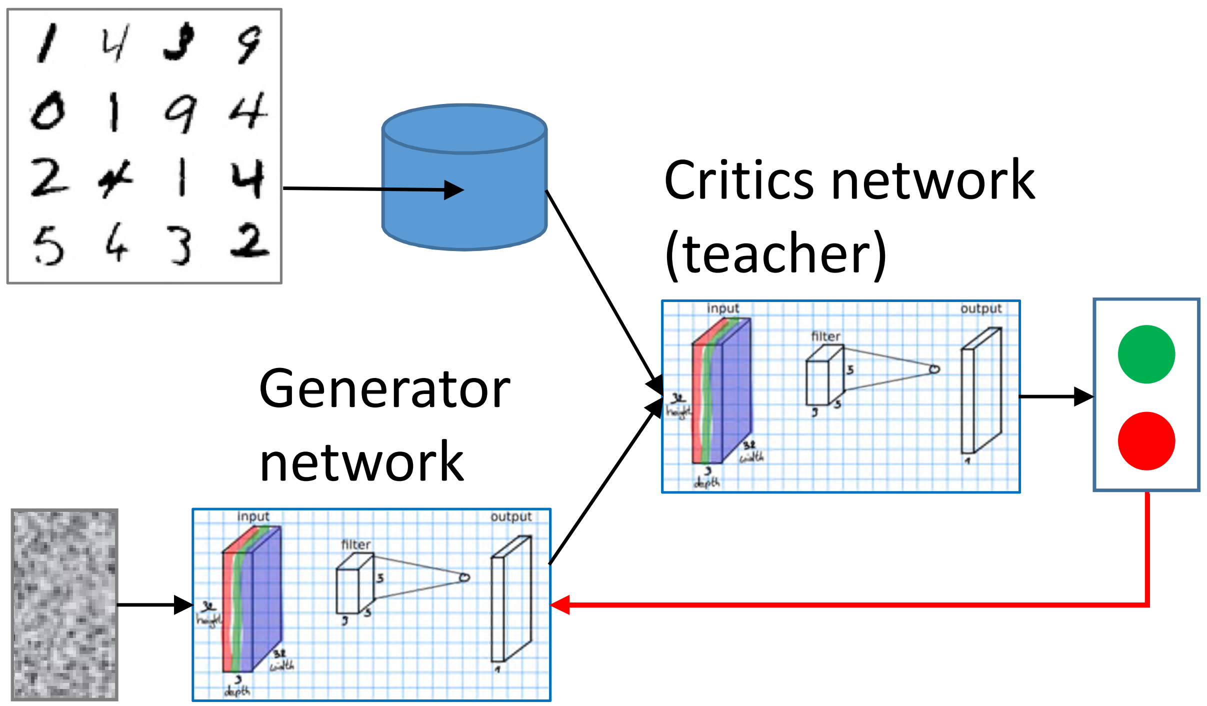


Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua BengioarXiv:1406.2661
M. Erdmann, J. Glombitza, T. Quast, Comput Softw Big Sci (2019) 3: 4

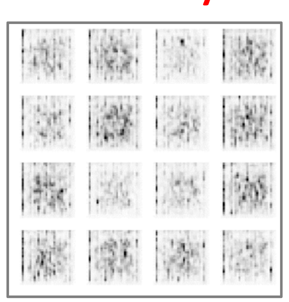
6. Simulations: Generative Modeling

('unsupervised')

Real world images

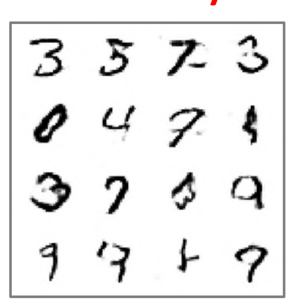


1st try

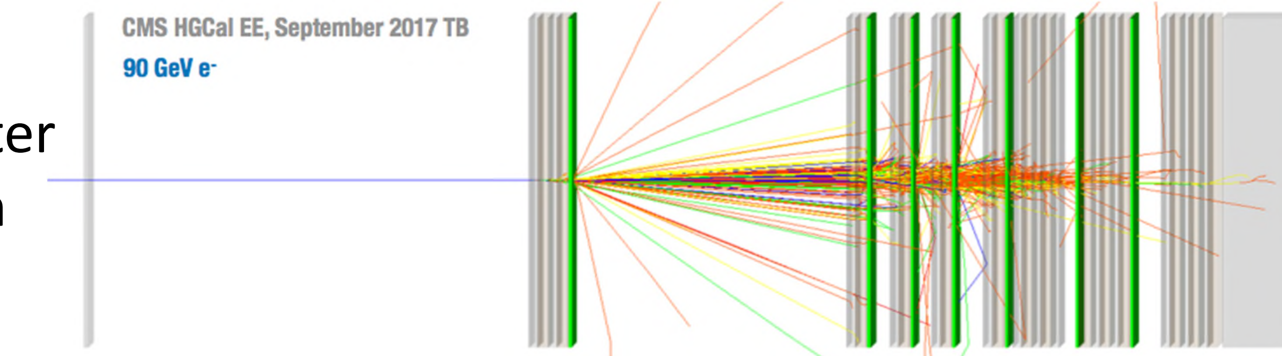


2nd try

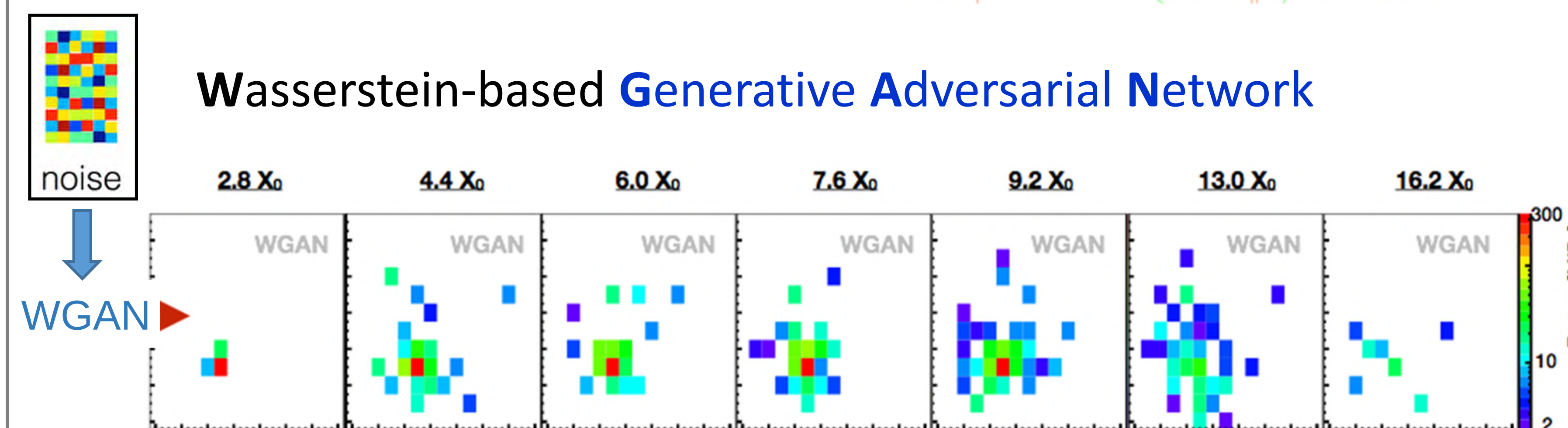
4th try



Electron calorimeter in CERN test beam



Wasserstein-based **G**enerative **A**dversarial **N**etwork



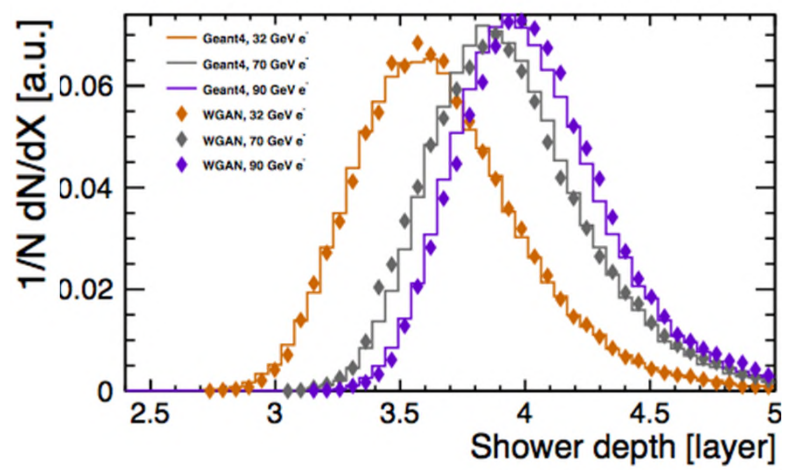
Generation Method	Hardware	milliseconds/ shower
GEANT4	CPU	**2000**
WGAN	CPU	52
	GPU	**0.3**

R&D: Ultrafast detector simulations of high quality.
Huge effort ahead for production versions

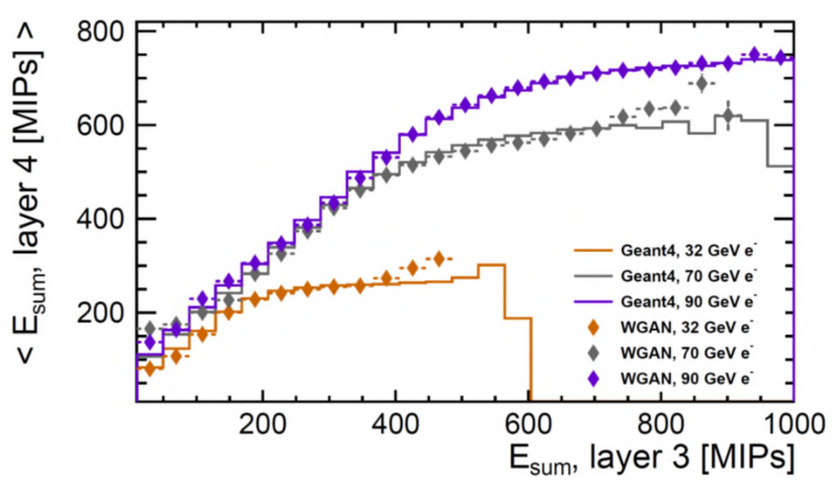
Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua BengioarXiv:1406.2661
M. Erdmann, J. Glombitza, T. Quast, Comput Softw Big Sci (2019) 3: 4

5. Simulations: Generative Modeling

Shower depth

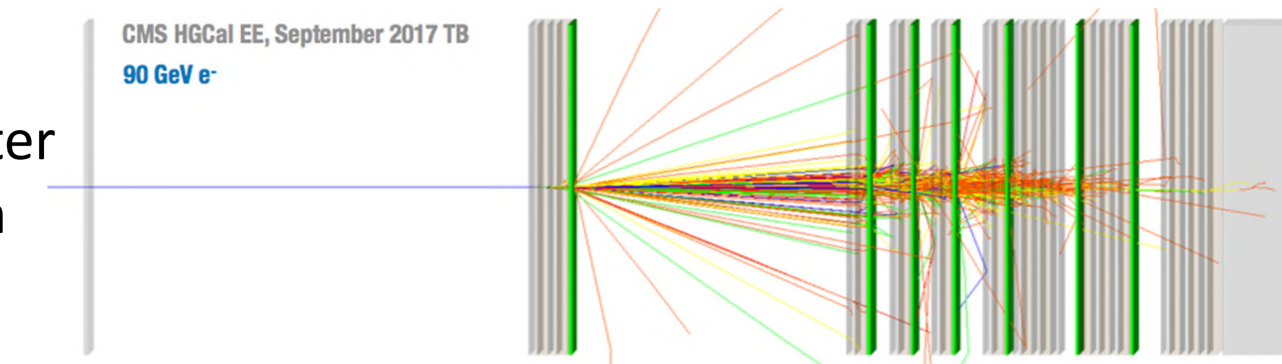


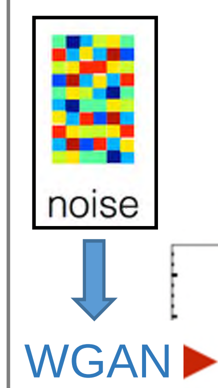
Correlations between layers



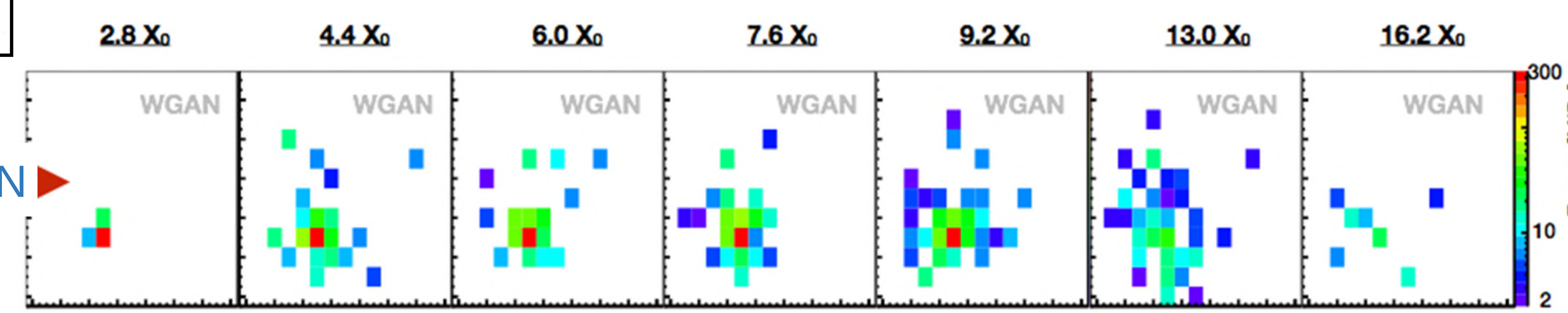
Deficit of low-energy depositions to be solved

Electron calorimeter in CERN test beam





Wasserstein-based **G**enerative **A**dversarial **N**etwork



Generation Method	Hardware	milliseconds/ shower
GEANT4	CPU	**2000**
WGAN	CPU	52
	GPU	**0.3**

R&D: Ultrafast detector simulations of high quality.
Huge effort ahead for production versions

Messages Machine Learning

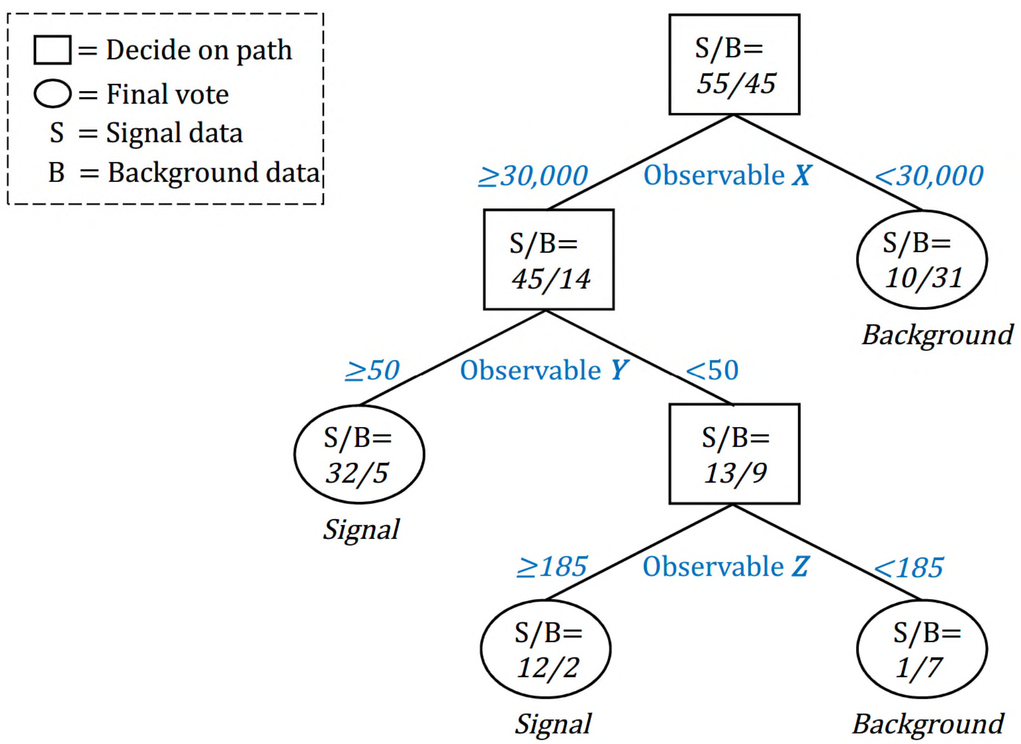
- Using machines physicists exploit physics contained in data deeper than before
- Construct neural networks to
  - Autonomously find optimal solutions & variables, reduce uncertainties
  - Transport probability distributions for ultra-fast detector simulations
  - Search for new physics phenomena in data
- Continuous quality challenge
  - Causality, stability, uncertainty of predictions
  - Experiment operation reality

We ought to prepare for fundamental change to include machines in our daily work

backup

Machine Learning: Boosted Decision Trees

Decisions at each node of tree



Train by minimizing *Gini* criterion

$$G \equiv (N_S + N_U)\ p\ (1-p)$$

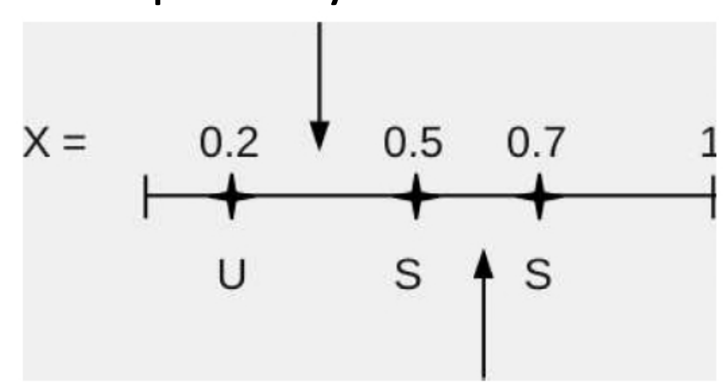
Signal purity $\quad p \equiv \dfrac{N_S}{N_S + N_U}$

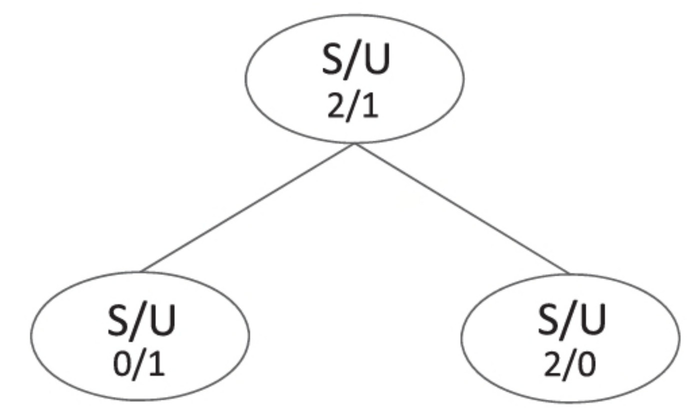
$$G_{\text{Go LEFT}} = (0+1)\,\frac{0}{1}\left(1-\frac{0}{1}\right) = 0$$

YES

$$G_{\text{Go RIGHT}} = (2+0)\,\frac{2}{2}\left(1-\frac{2}{2}\right) = 0$$

optimally cut here ?





or cut here ?

$$G_{\text{Go LEFT}} = (1+1)\,\frac{1}{2}\left(1-\frac{1}{2}\right) = \frac{1}{2}$$

NO

Causality: analysis of network predictions

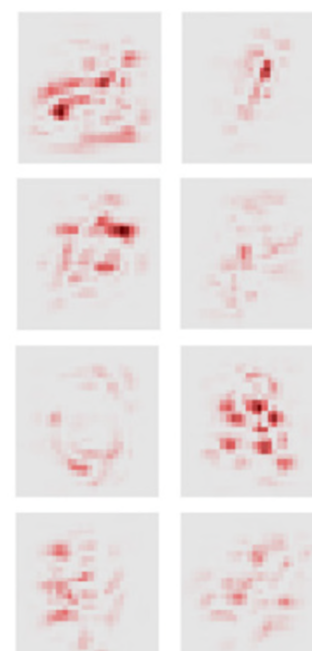
Measure impact: input x_i → overall prediction

G. Montavon, W. Samek, K.-R. Müller, Digital Signal Processing 73 (2018) 1

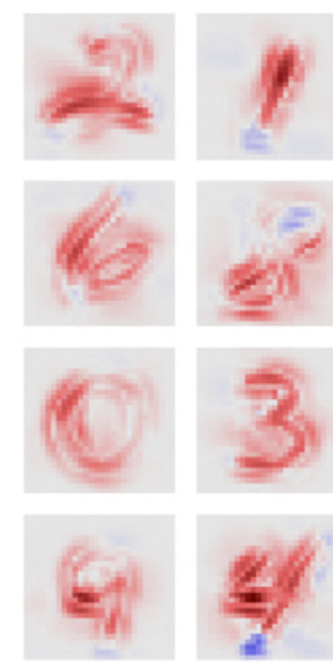
Original

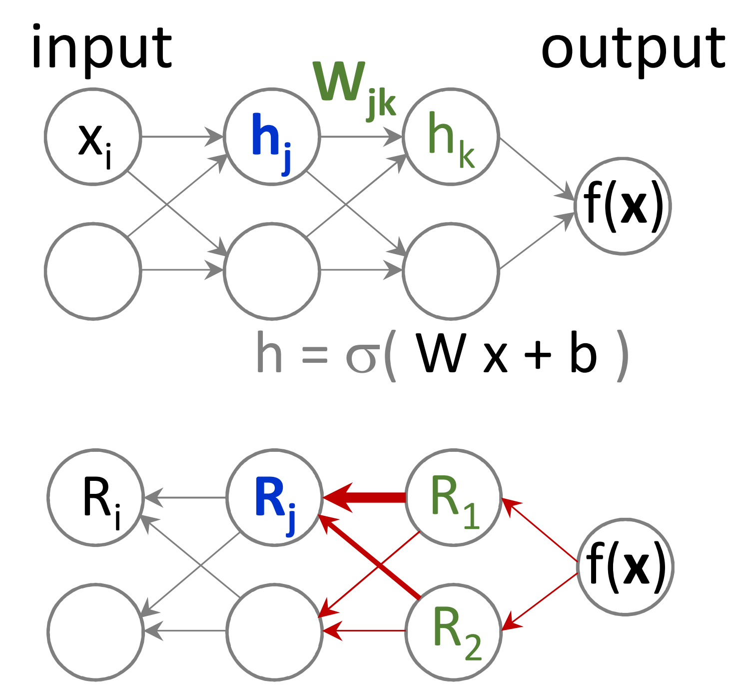
Sensitivity by
partial derivative

$\left.\frac{\partial f(\boldsymbol{x})}{\partial x_i}\right|_{\boldsymbol{x}_o}$



LRP=Layer-wise Relevance Propagation
separate sums: in favor / against prediction





$$a_k = \sigma\left(\sum_j a_j w_{jk} + b_k\right)$$

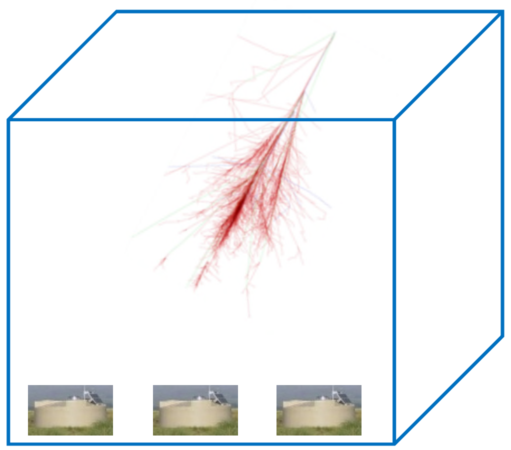
$$R_j = \sum_k \left(\alpha \frac{a_j w_{jk}^+}{\sum_j a_j w_{jk}^+} - \beta \frac{a_j w_{jk}^-}{\sum_j a_j w_{jk}^-}\right) R_k$$

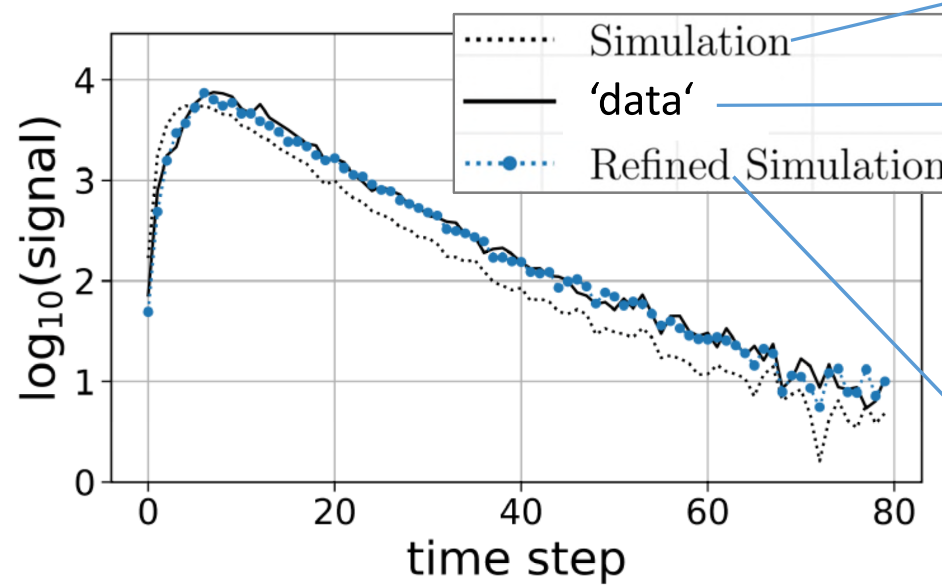
M. Erdmann, Lukas Geiger, Jonas Glombitza, David Schmidt, Comput Softw Big Sci (2018) 2: 4
M. Erdmann, Jonas Glombitza, David Walz, 10.1016/j.astropartphys.2017.10.006

For network training: adapt simulation to data

Long-standing problem ➡ Refine simulated traces

data ≠ simulation

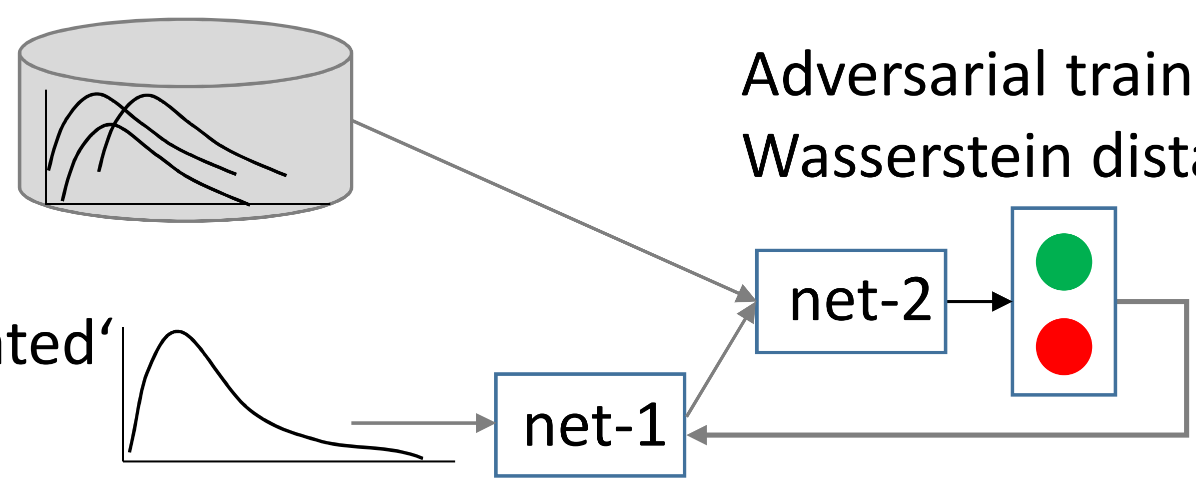




'data' traces
(simulated)

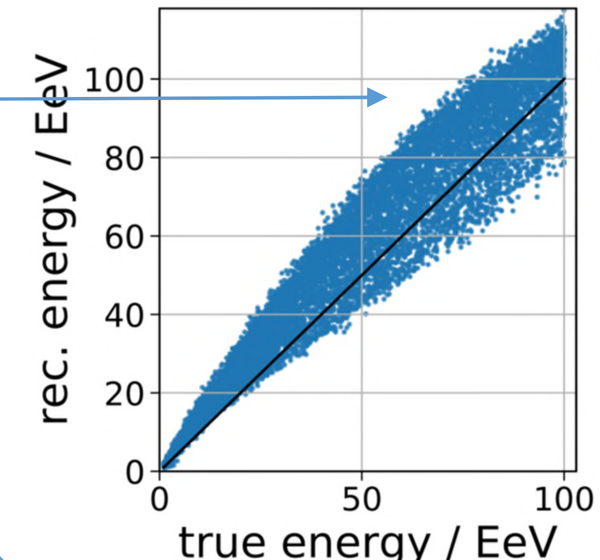
Adversarial training
Wasserstein distance

'badly simulated'
trace



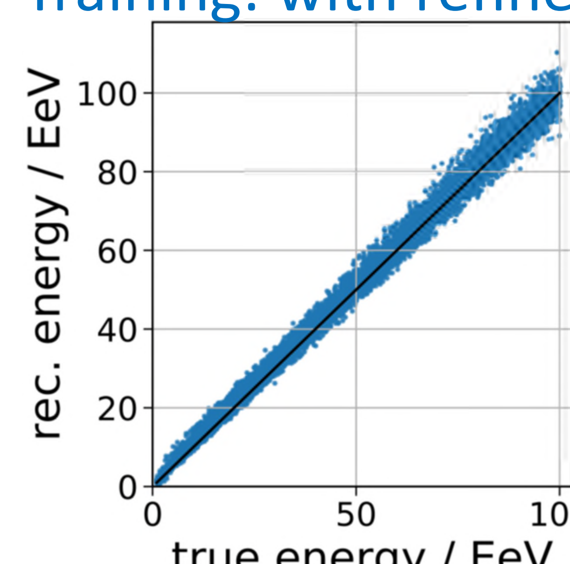
Energy reconstruction

Training: badly simulated traces



Deep Network
net-3: cosmic
ray energy from
'data' traces

Training: with refined simulation



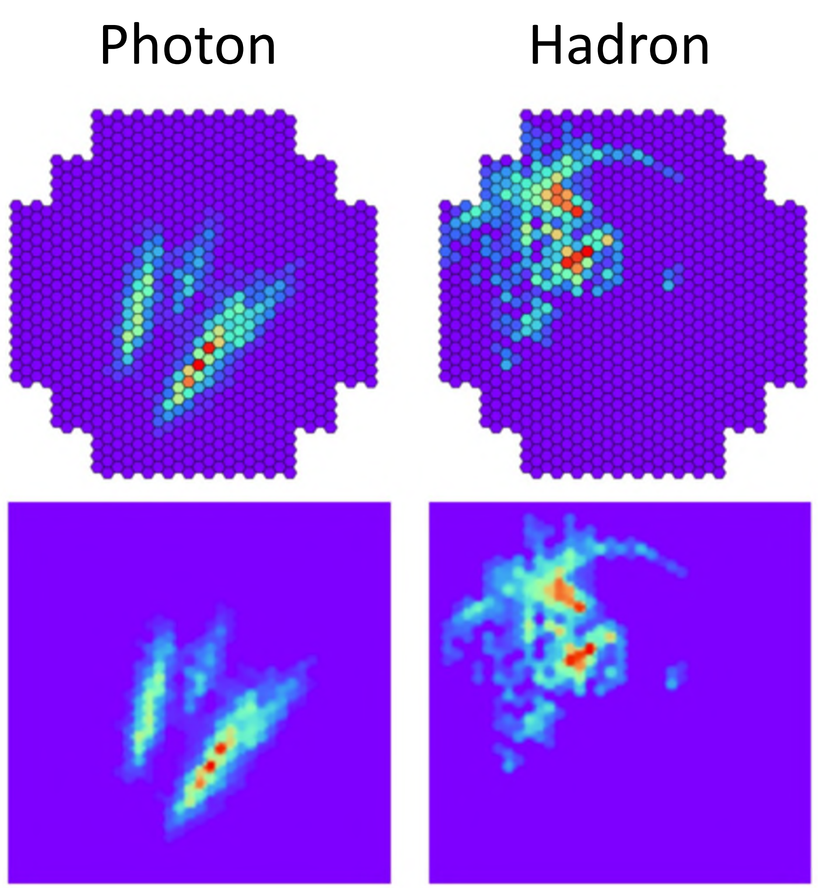
improved energy resolution

I. Shilon et al, arXiv:1803.10698

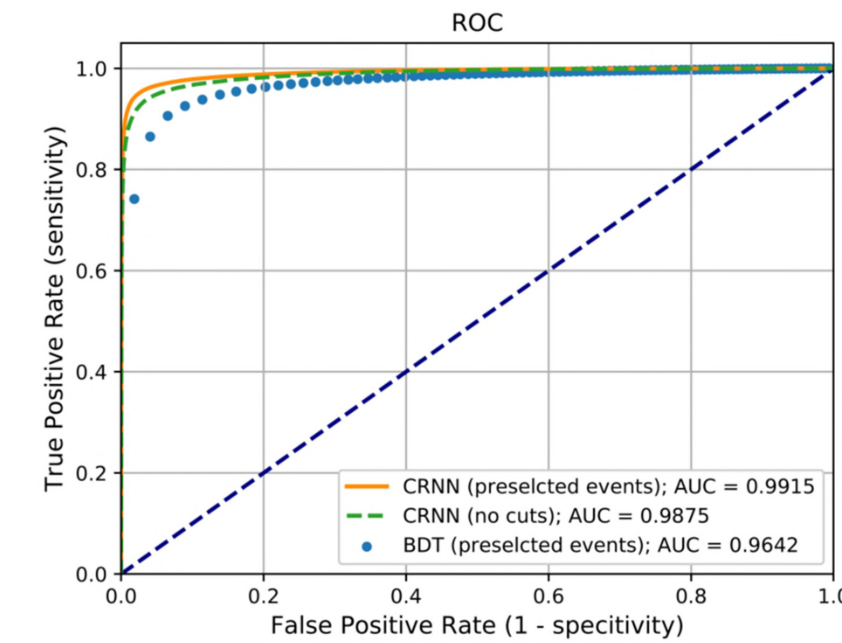
Imaging Atmospheric Cherenkov Telescopes

Telescope image (hexagonal structure)

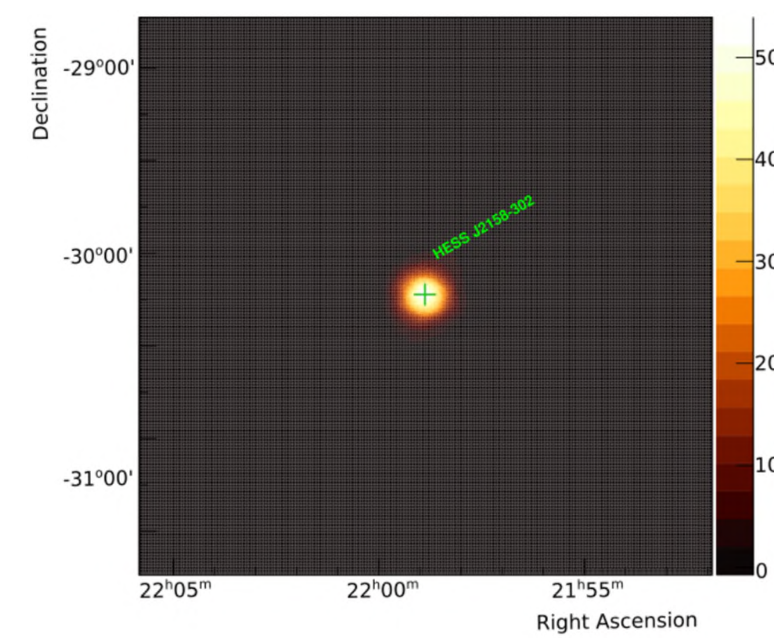
Preprocessing example for convolutional network (rebinning)



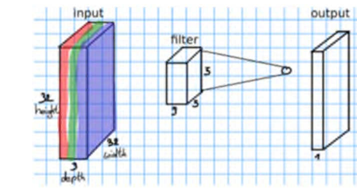
Simulation:
deep network better to reject background compared to Boosted Decision Tree

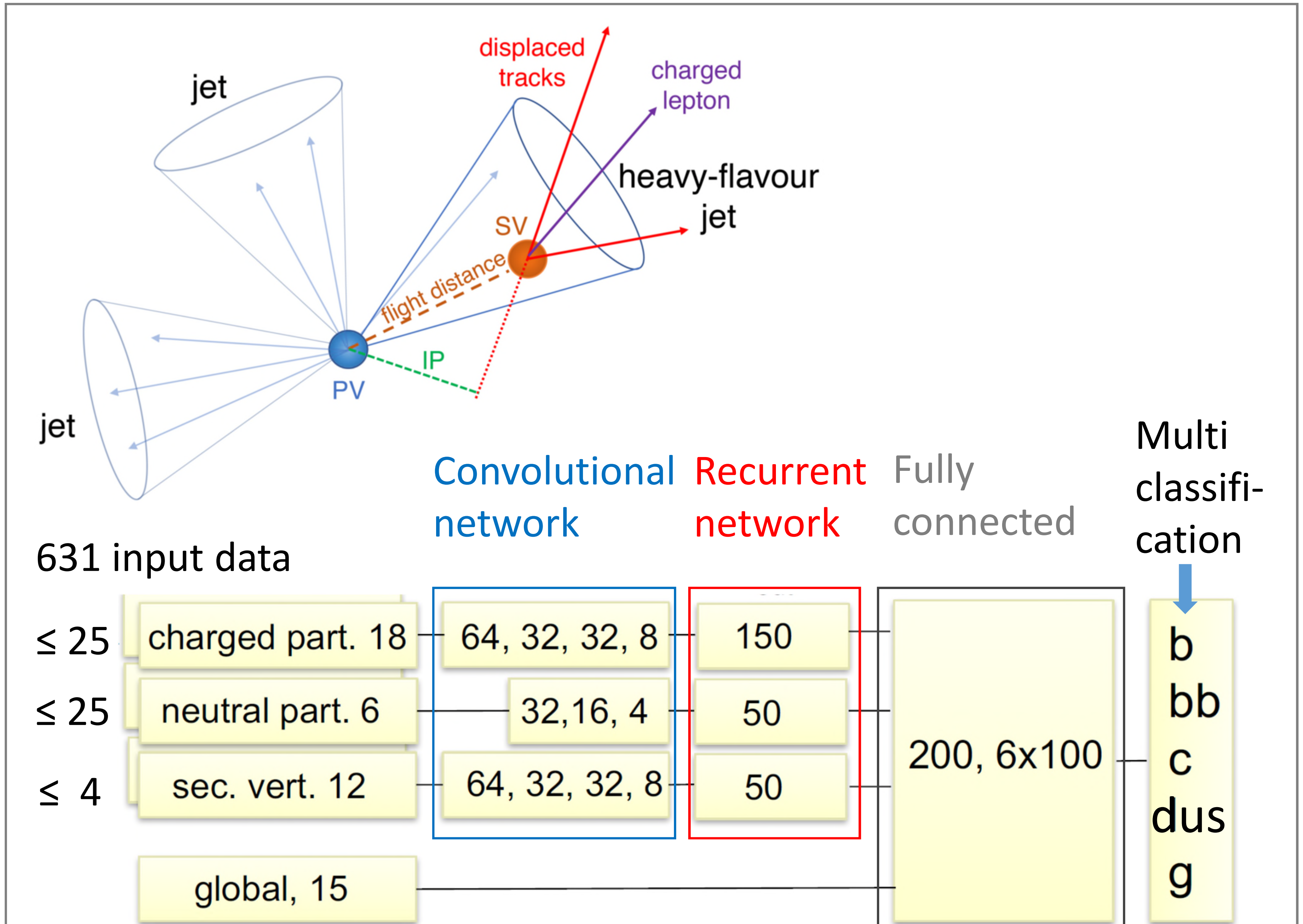


Data:
excess events observed in direction of PKS 2155-304 for one flare observation

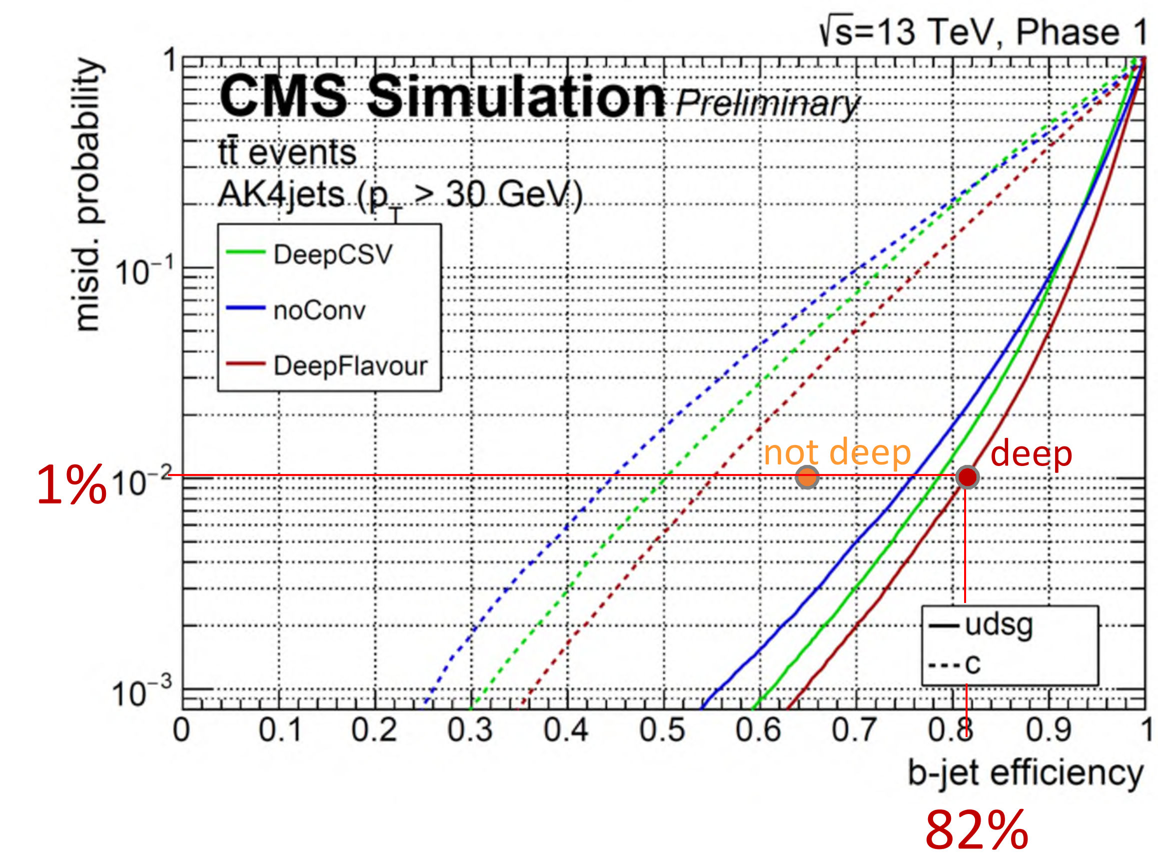


CMS jet flavor tagging





The up to 25 tracks have 18 properties px,py,pz,eta,quality,... the 1-dim Convolution adds up all properties of a track. In the RNN the summed properties of the tracks are merged.



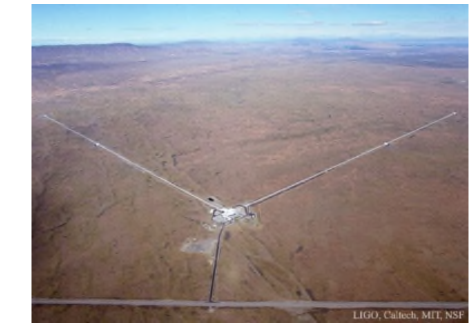
Brilliant improvement with high impact on data analyses

2 Higgs → 4 bottom quarks:

$\varepsilon^4_{bottom}: 0.18 \rightarrow 0.45$

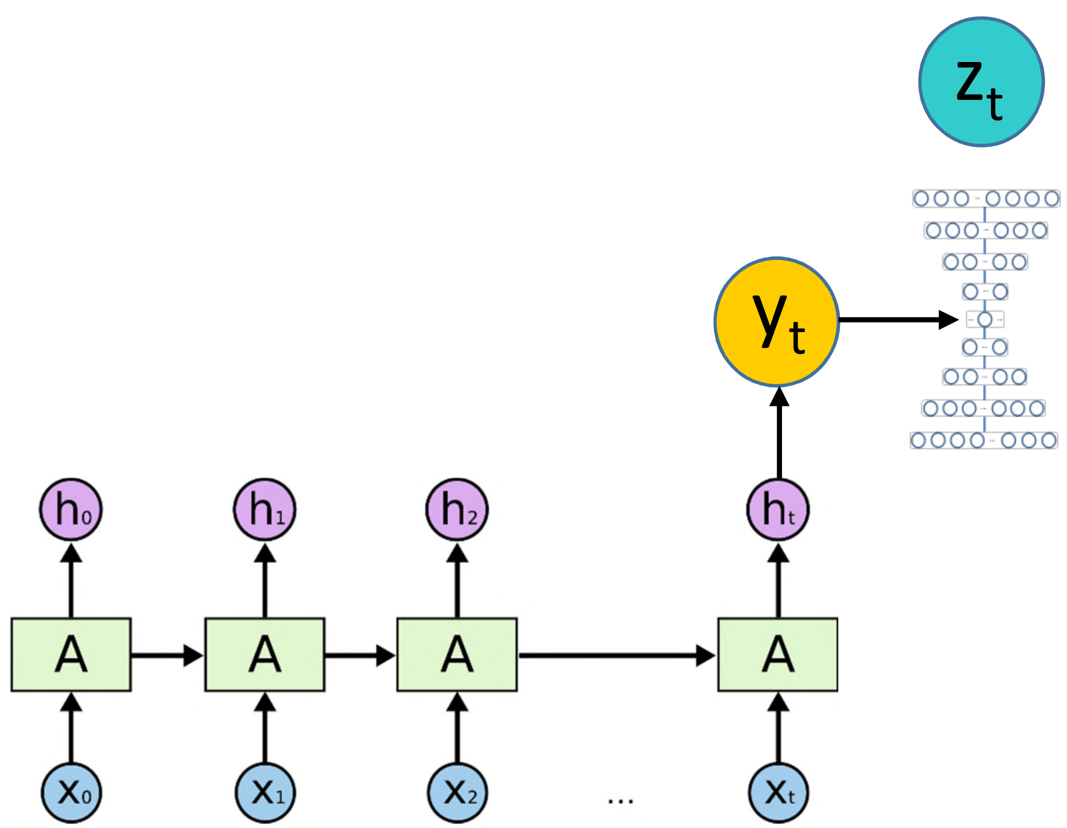
Hongyu Shen, Daniel George, E. A. Huerta, Zhizhen Zhao1, arXiv:1711.09919

Denoising Gravitational Waves with Recurrent Denoising Autoencoder



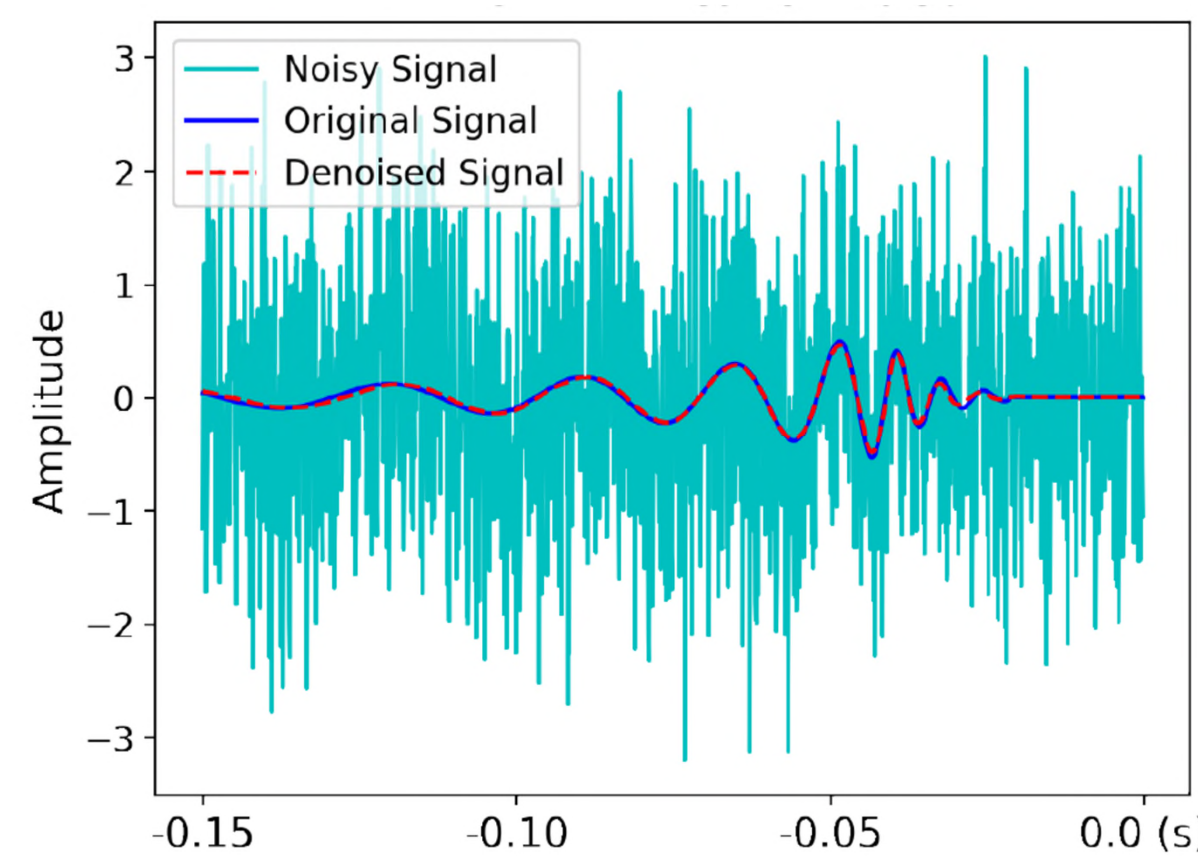
Recurrent Network

Autoencoder Network



many-to-one network

Gravitational wave (simulated)



binary black hole merger

Excellent recovery of original signal