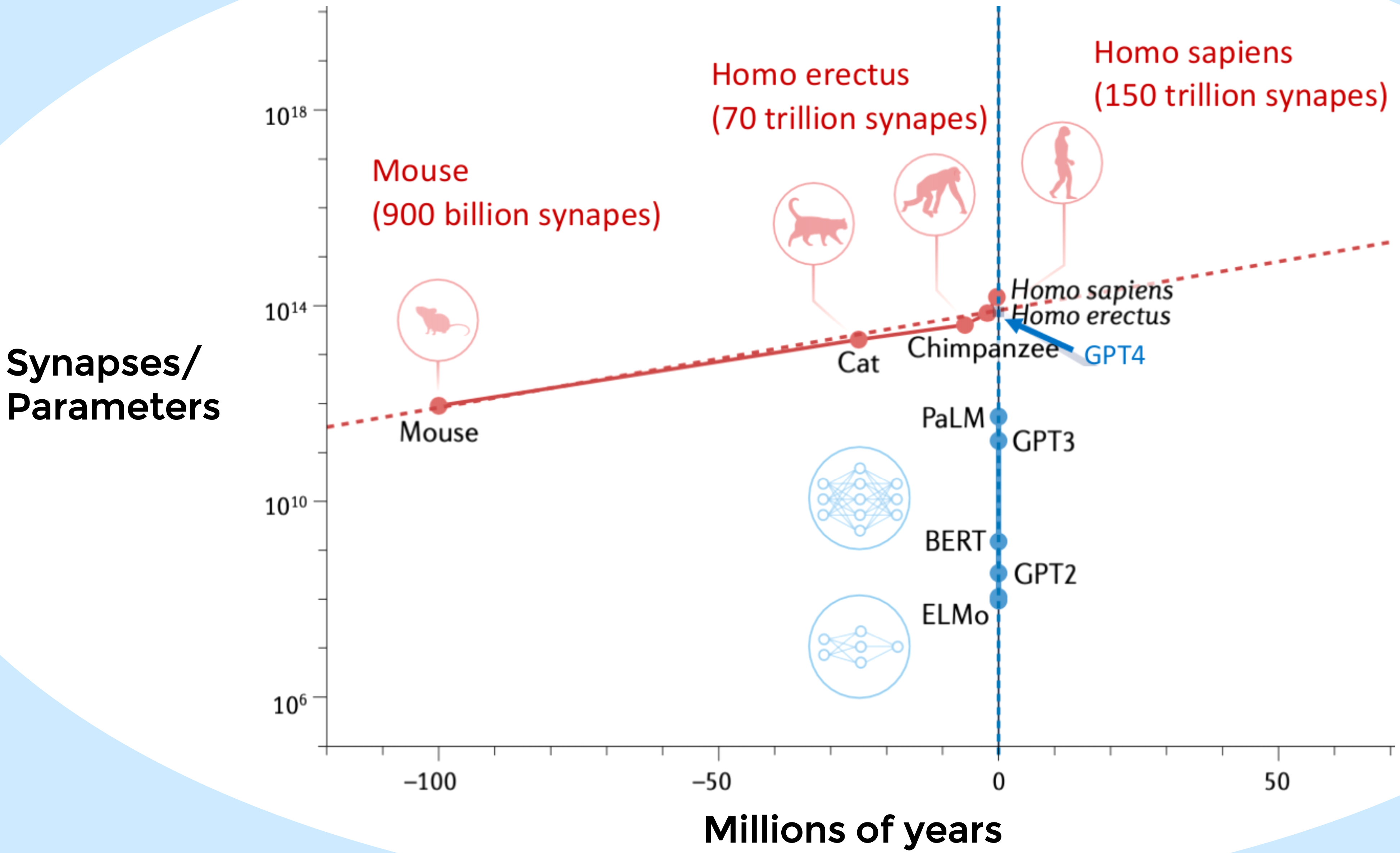
Deep Learning meets Physics

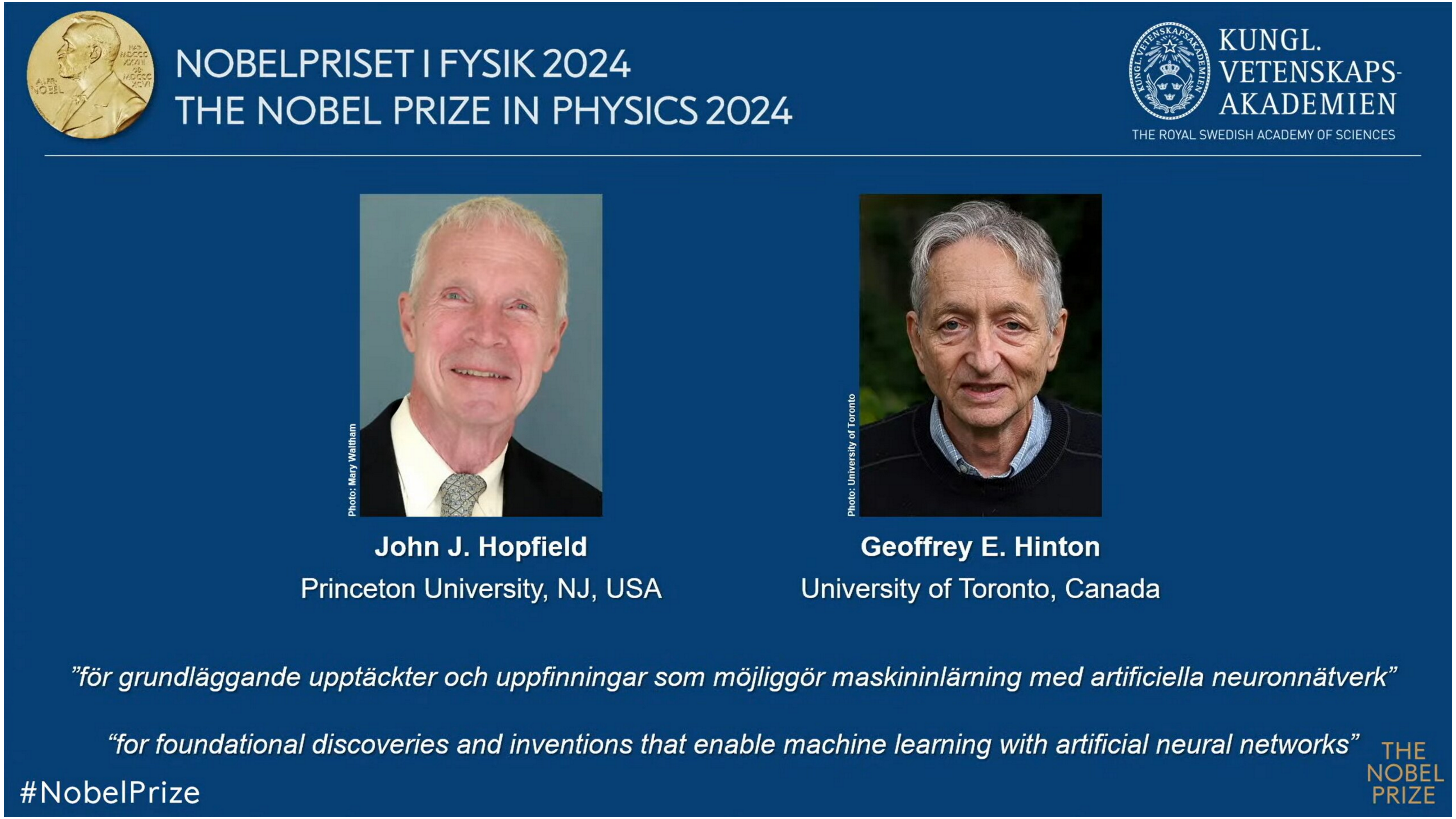


Matthew D. Schwartz,
Nat Rev Phys 4 (2022) 741,
adapted

Generative Modeling

https://thispersondoesnotexist.com

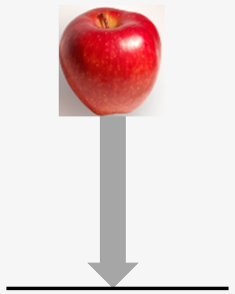
Nobelprize Physics 2024



Physics: Models for phenomena of nature

Classic: physics concept

$$F = m\,a$$



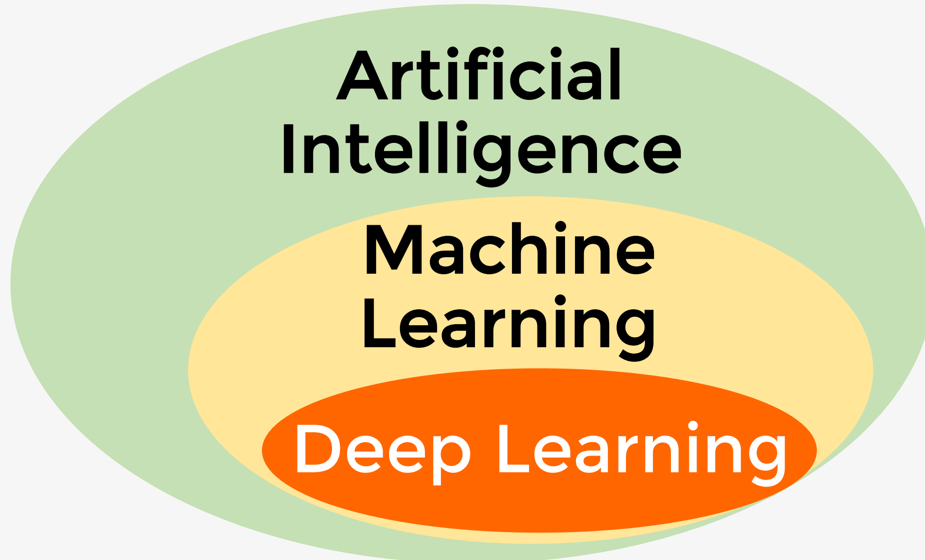
h

0

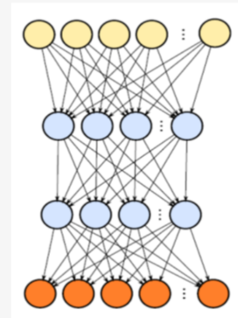
$$t = \sqrt{\frac{2h}{a}}$$

Mathematically readable model

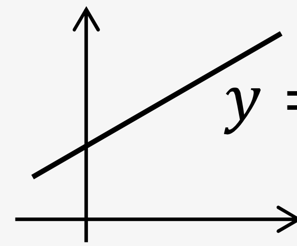
Data-driven concepts



h



$t = \ldots \leftrightarrow t_{true}$



$y = \boldsymbol{w}\,x + \boldsymbol{b}$

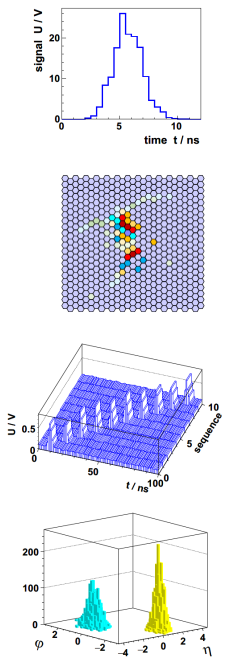
$y > 0$ positive

Highly complex physics models

Scientific research: Quality of the prediction & scope of validity

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

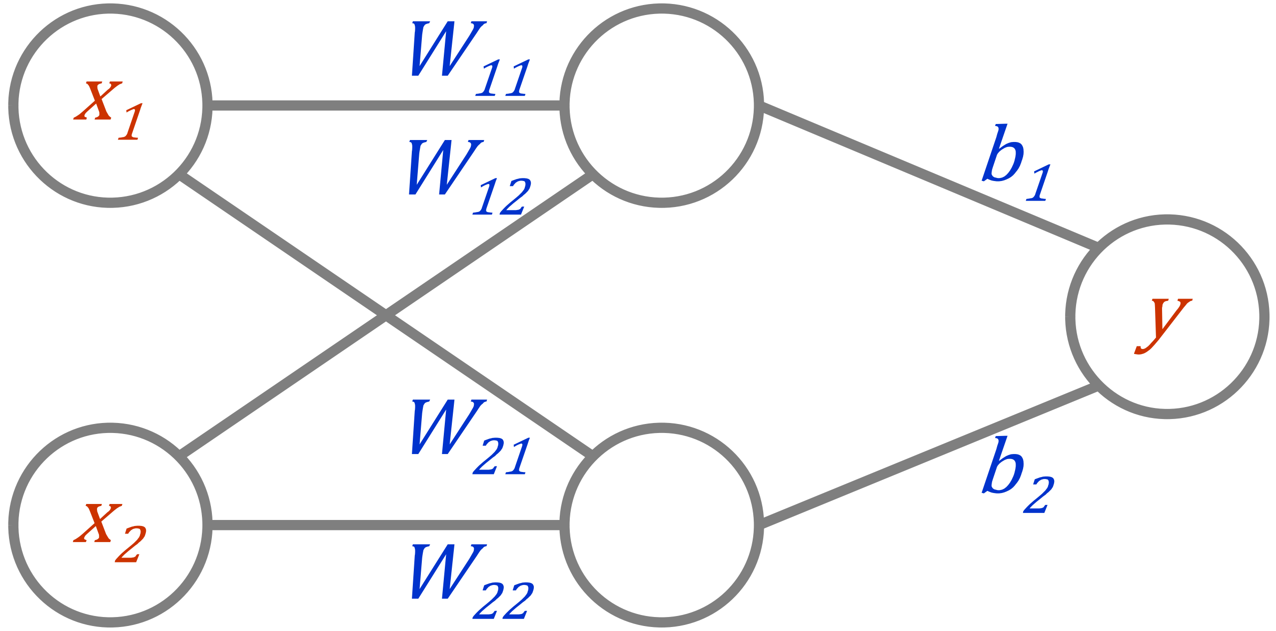
Neural Network Operations



x multi-dimensional input data

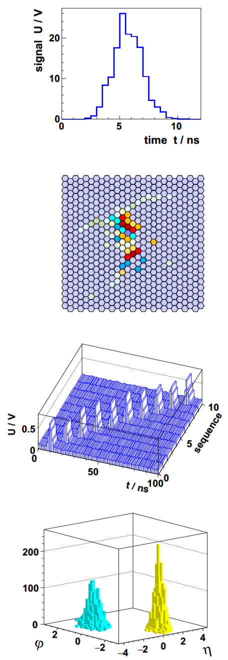
W, b to be trained

$$y = W x + b$$



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

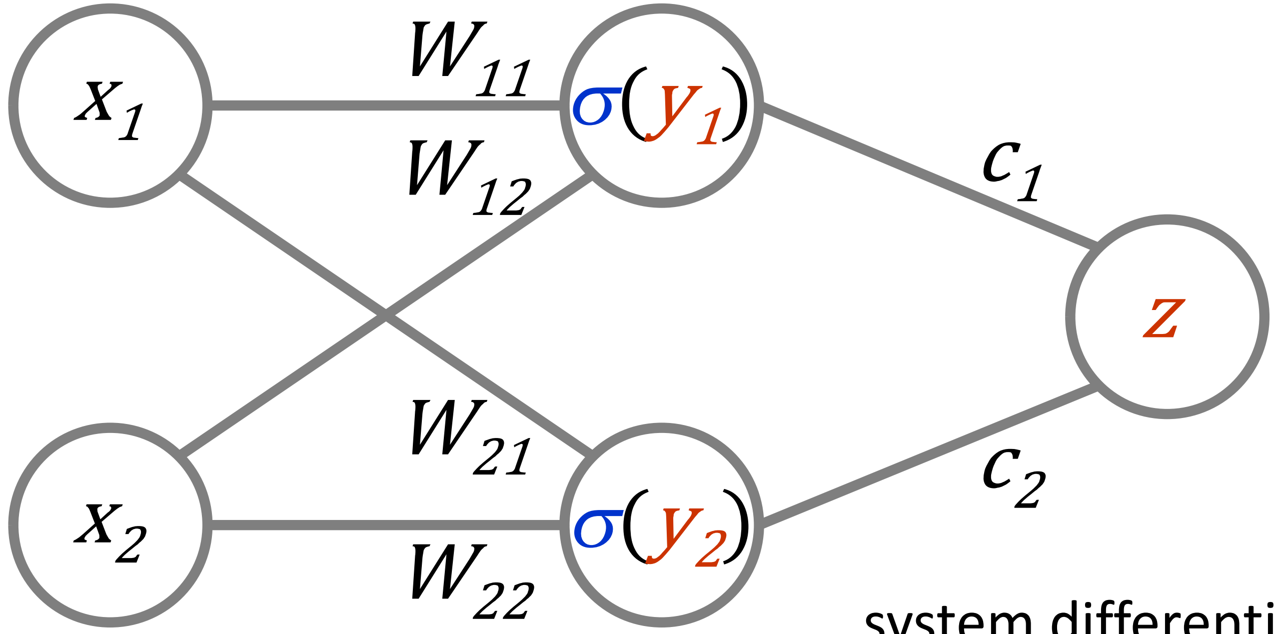


$\boldsymbol{x}$ multi-dimensional input data

W, b to be trained

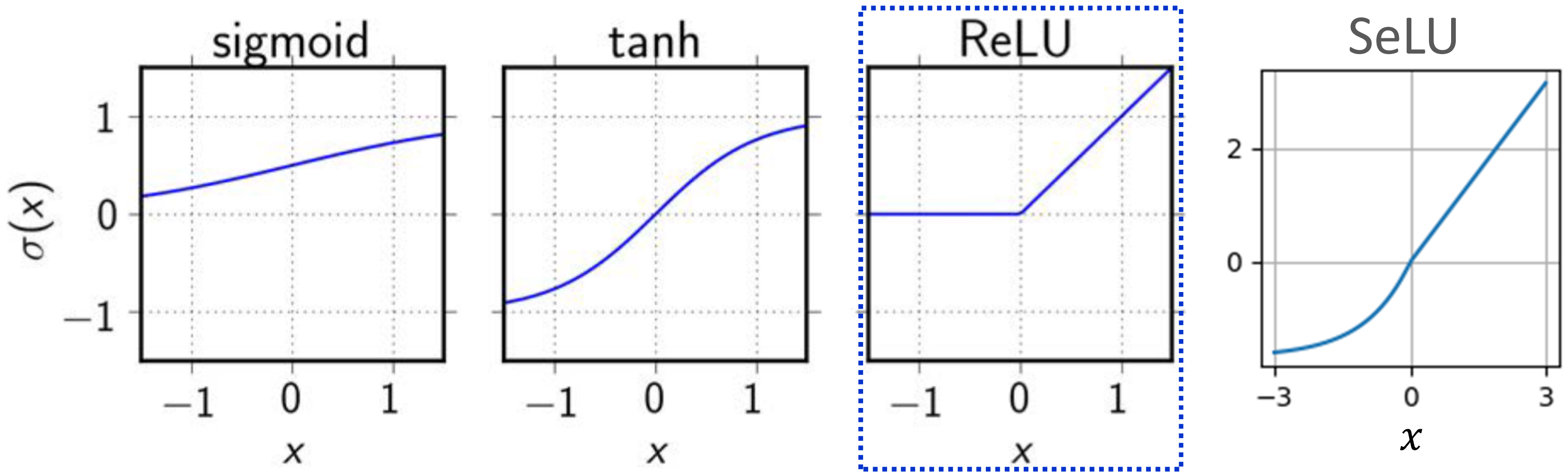
$$y = W\,x + b$$

$$z = \sigma(y)$$



system differentiable

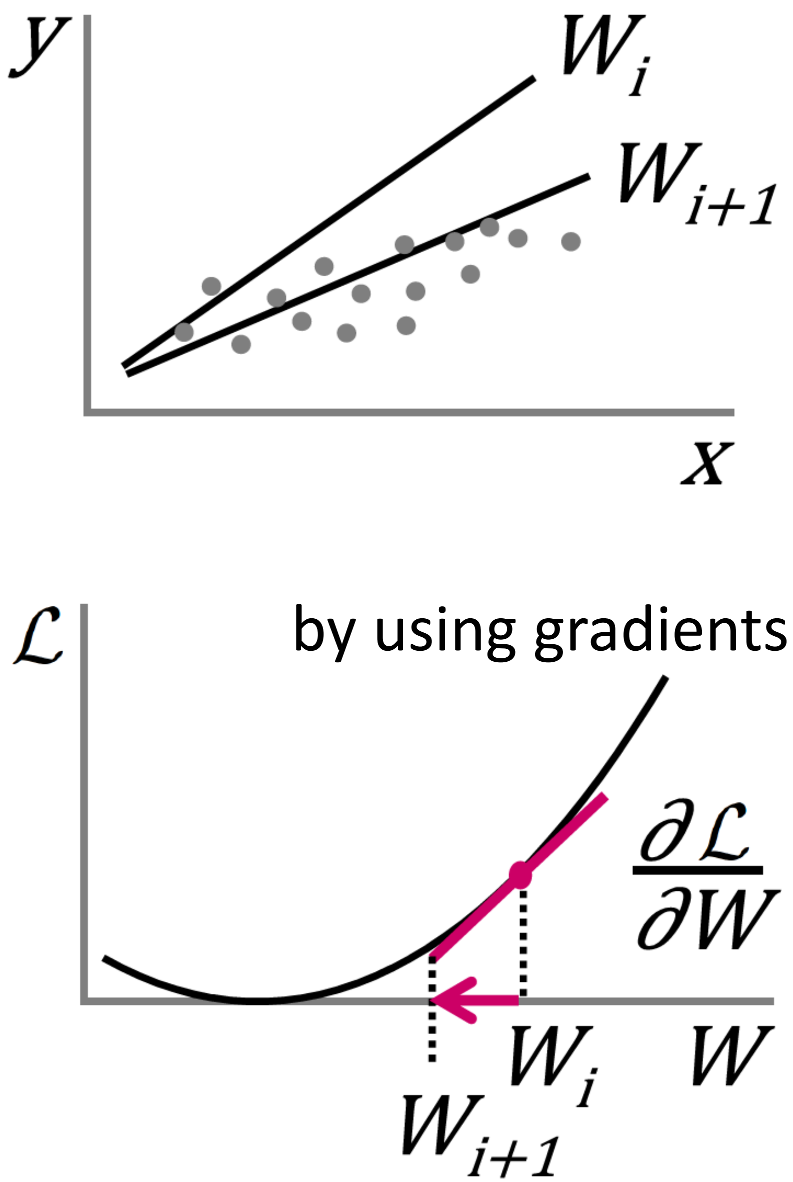
activation function: departure from linear system



Neural Network Training

('supervised')

iterative parameter improvement



- **Data** set

  $\{ x_j , y_j \} \quad j = 1 , \ldots , N$

- Define **model** = network prediction

  $y_m (x) = W x + b$

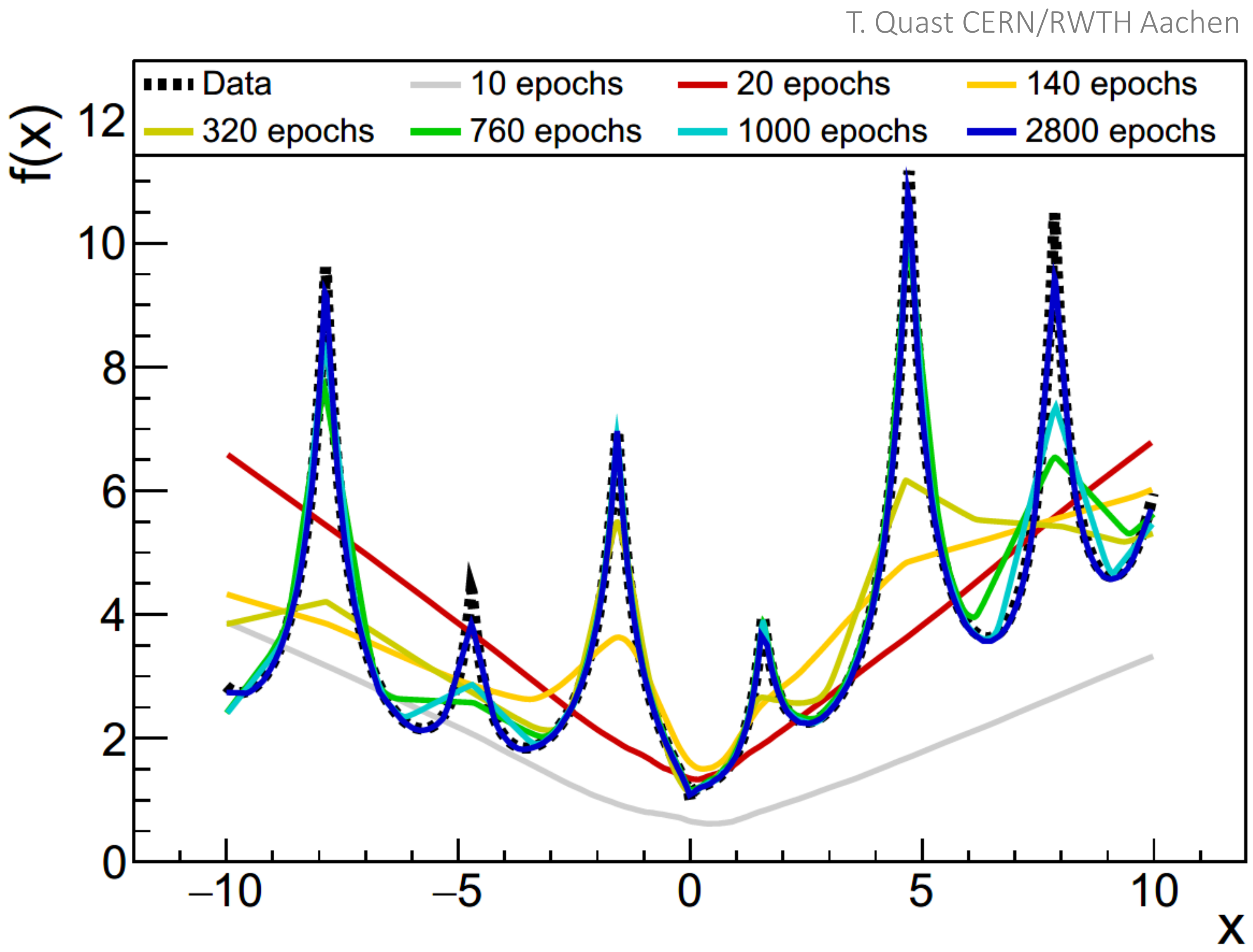
- Define **objective** function (=**loss, cost**)

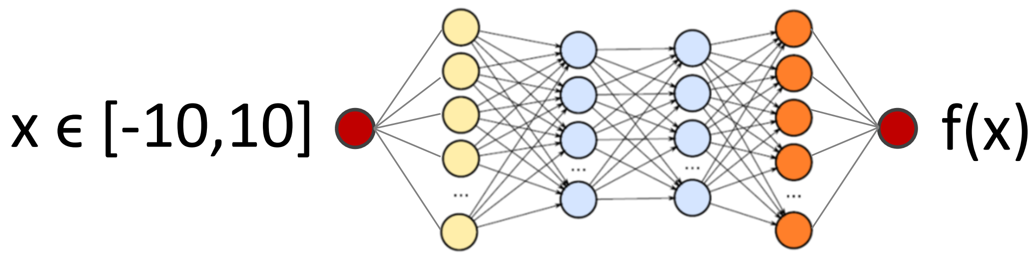
  $$\mathcal{L}(W, b) = \frac{1}{N} \sum_{j=1}^{N} \left[y_m(x_j) - y_j \right]^2$$

- **Train** model by optimizing parameters

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

Automated parameterization of arbitrary function





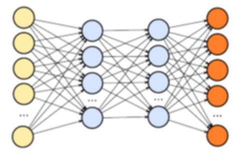
7 hidden layers
200 nodes each
ReLU activation function

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

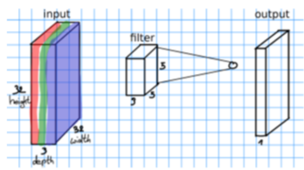
Deep Learning Progress

Concepts

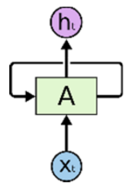
- Fully connected



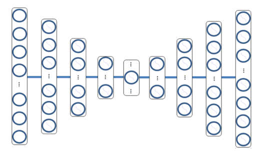
- Convolutional



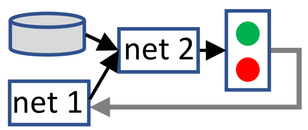
- Graph
- Recurrent



- Lorentz Boost Network
- Autoencoder



- Adversarial



- Reinforcement
- Invertible
- Transformer

Improved set of tools

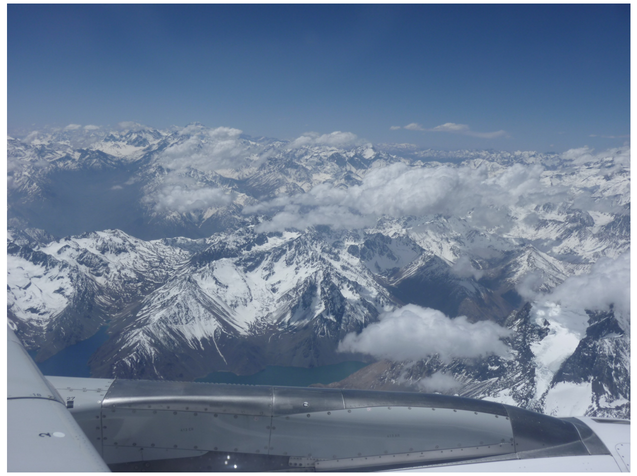
Train millions of parameters by:

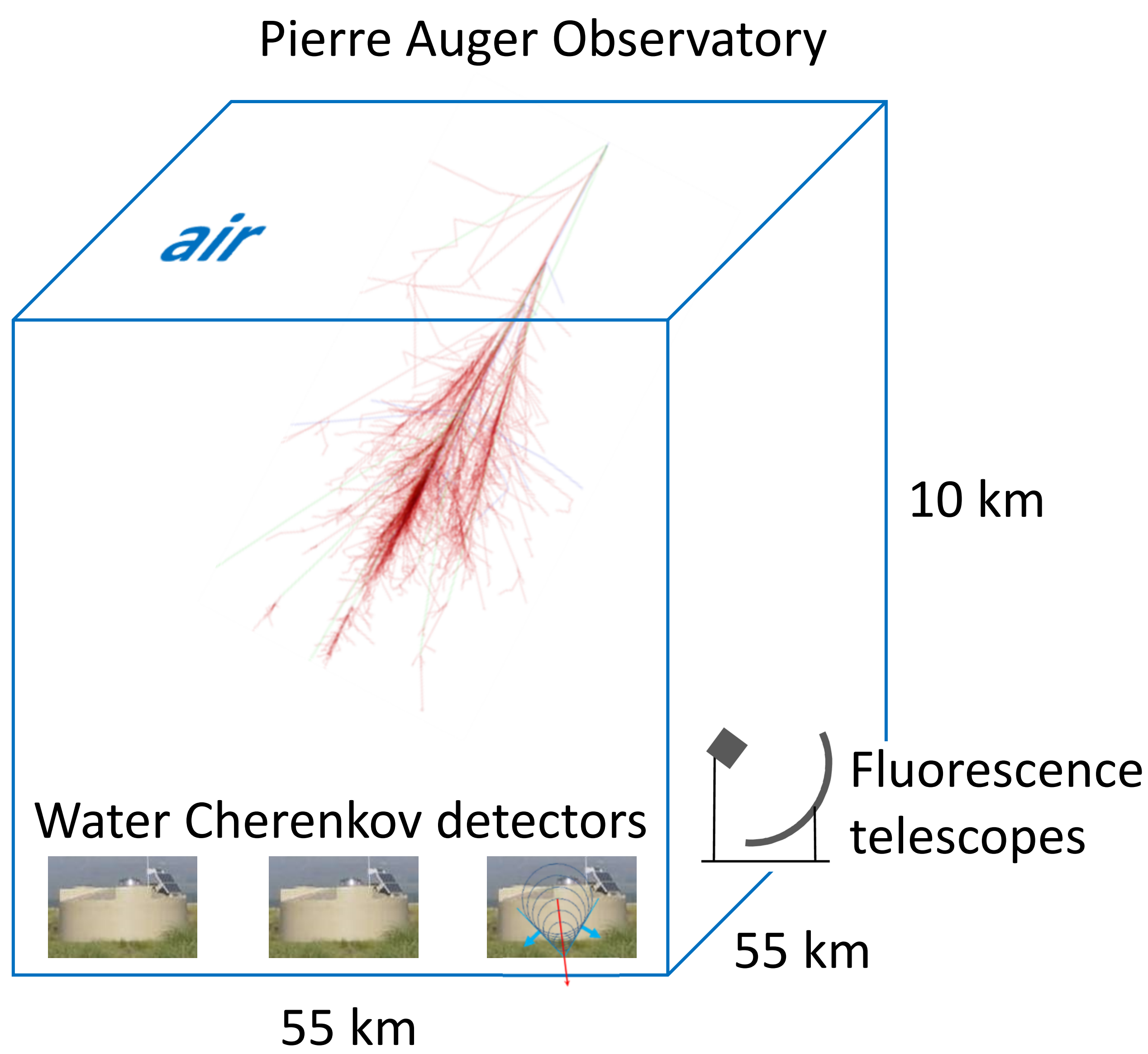
- Data preprocessing
- Normalization
- Regularization
- Short cuts
- Learning strategies
- ...

Computing

- Graphics Processing Unit (GPU)
- Software Libraries
  - TensorFlow Keras
  - PyTorch

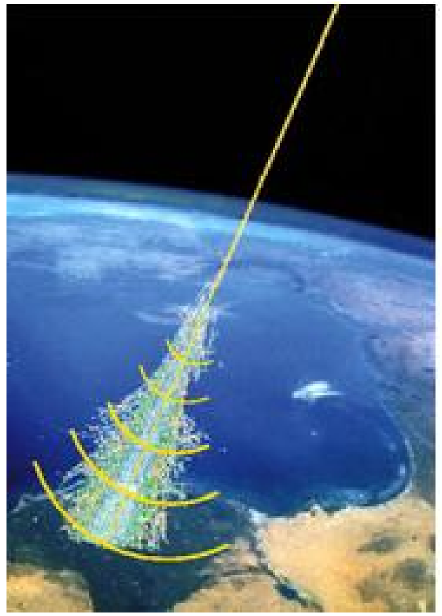
World's largest Calorimeter for Cosmic Rays

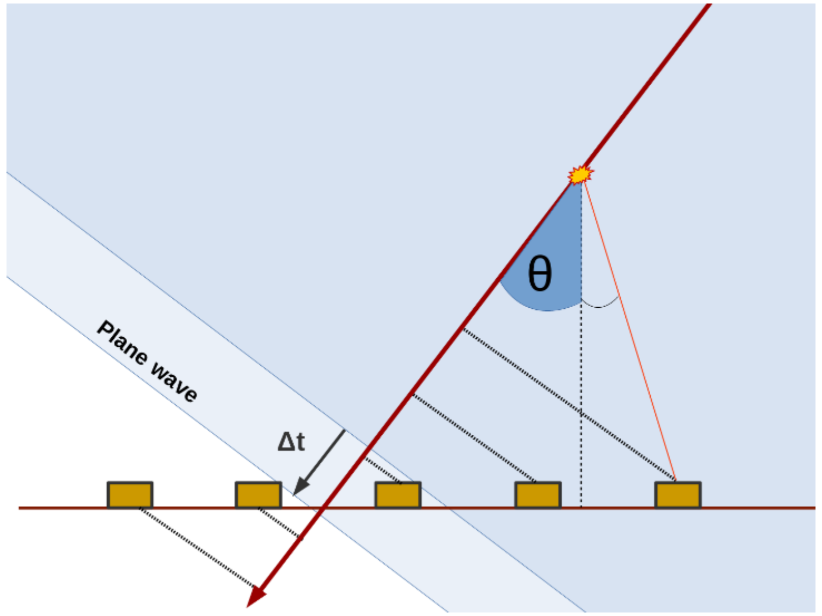


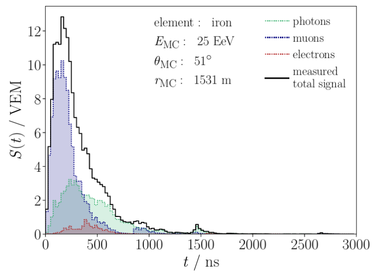


M. E., Jonas Glombitza, David Walz, 10.1016/j.astropartphys.2017.10.006
The Pierre Auger Collaboration, A. Abdul Halim, arXiv:2406.06319

Cosmic ray arrival directions by physicist or network





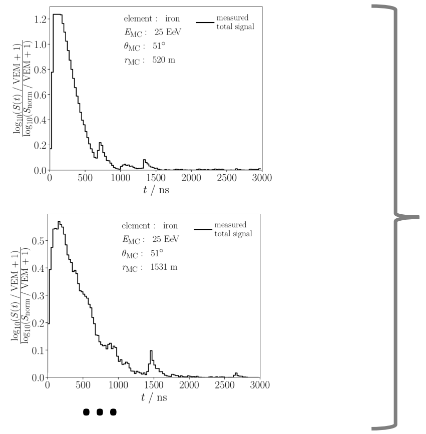


Photomultiplier simulted signal time traces

Educated physicist:

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

RAW input data:

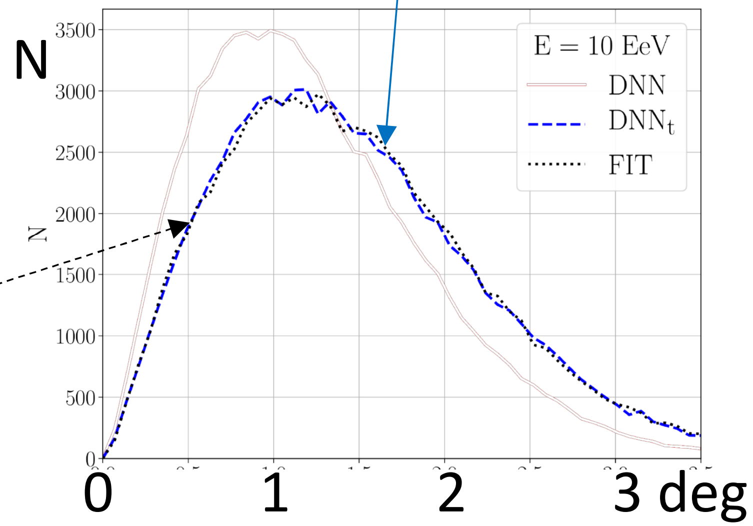


Deep Neural Network
No physics education
No explicit information about
- *locations of detectors*
- *speed of light*

Needs data with true target θ

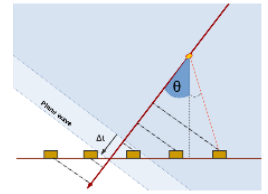
- Time offset only

Shower direction angular resolution



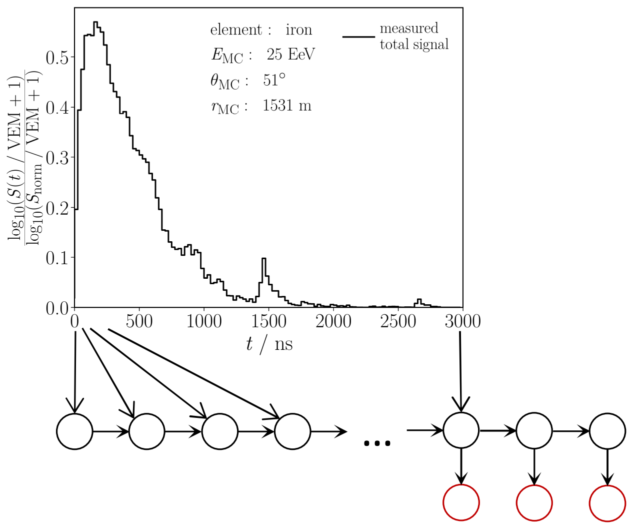
Deep Neural Network learns physics from data within 3h

M. E., Jonas Glombitza, David Walz, 10.1016/j.astropartphys.2017.10.006
The Pierre Auger Collaboration, A. Abdul Halim, arXiv:2406.06319



Neural network to characterize signal traces

Characterization of signal traces



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

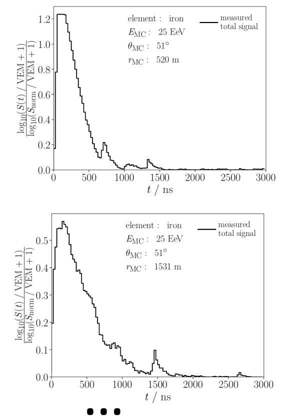
Time offsets & total signals

Recurrent Network: signal trace analysis

Convolutional Network combines these with detector locations

Predict shower direction θ

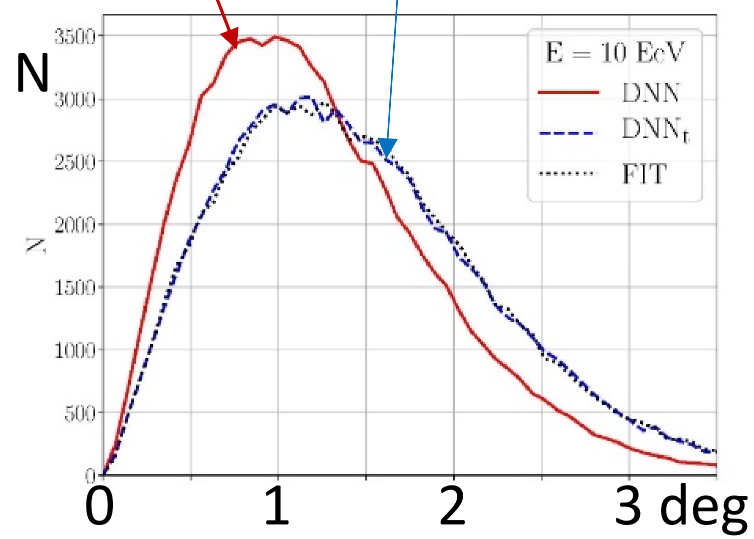
RAW input data:



Deep Neural Network

- Time offset only
- Signal traces added

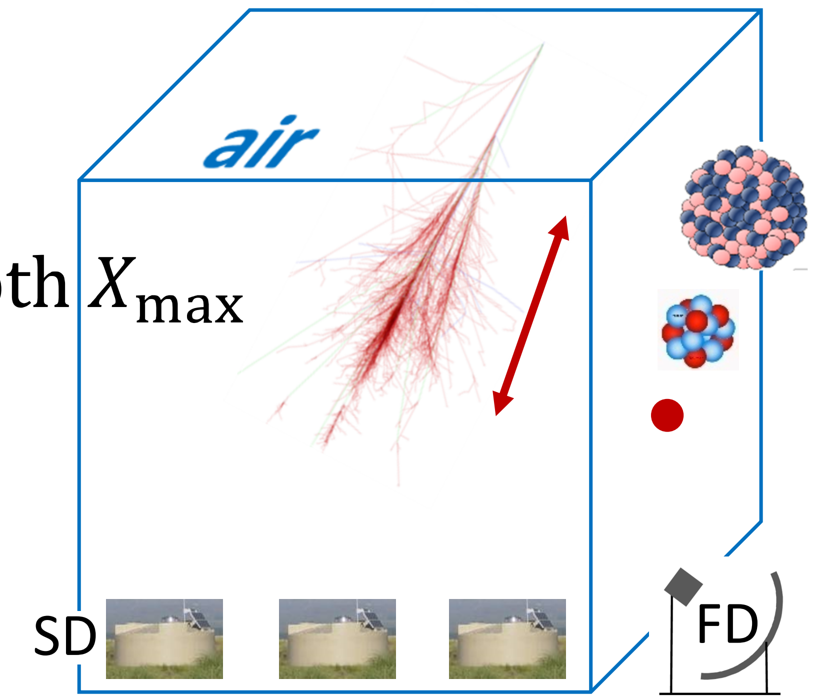
shower direction angular resolution



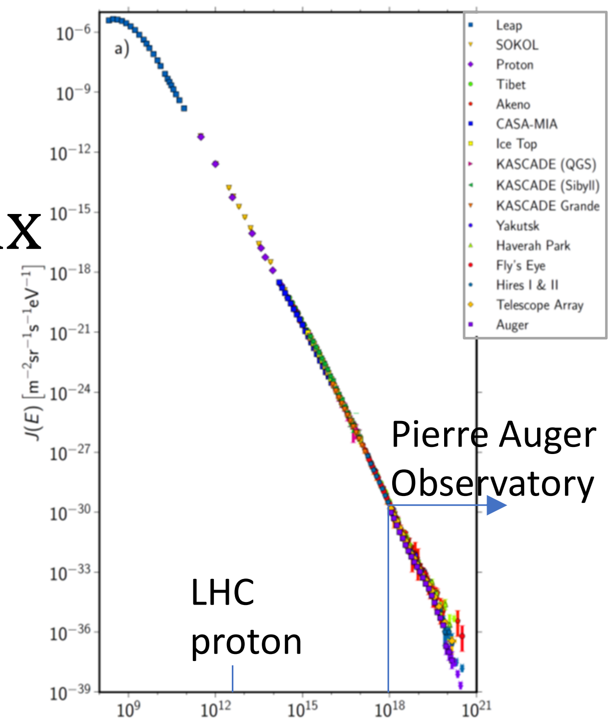
Cosmic-Ray Flux & Shower Depth

The Pierre Auger Collaboration, A. Abdul Halim, arXiv:2406.06315

Maximum of shower depth $X_{\max}$



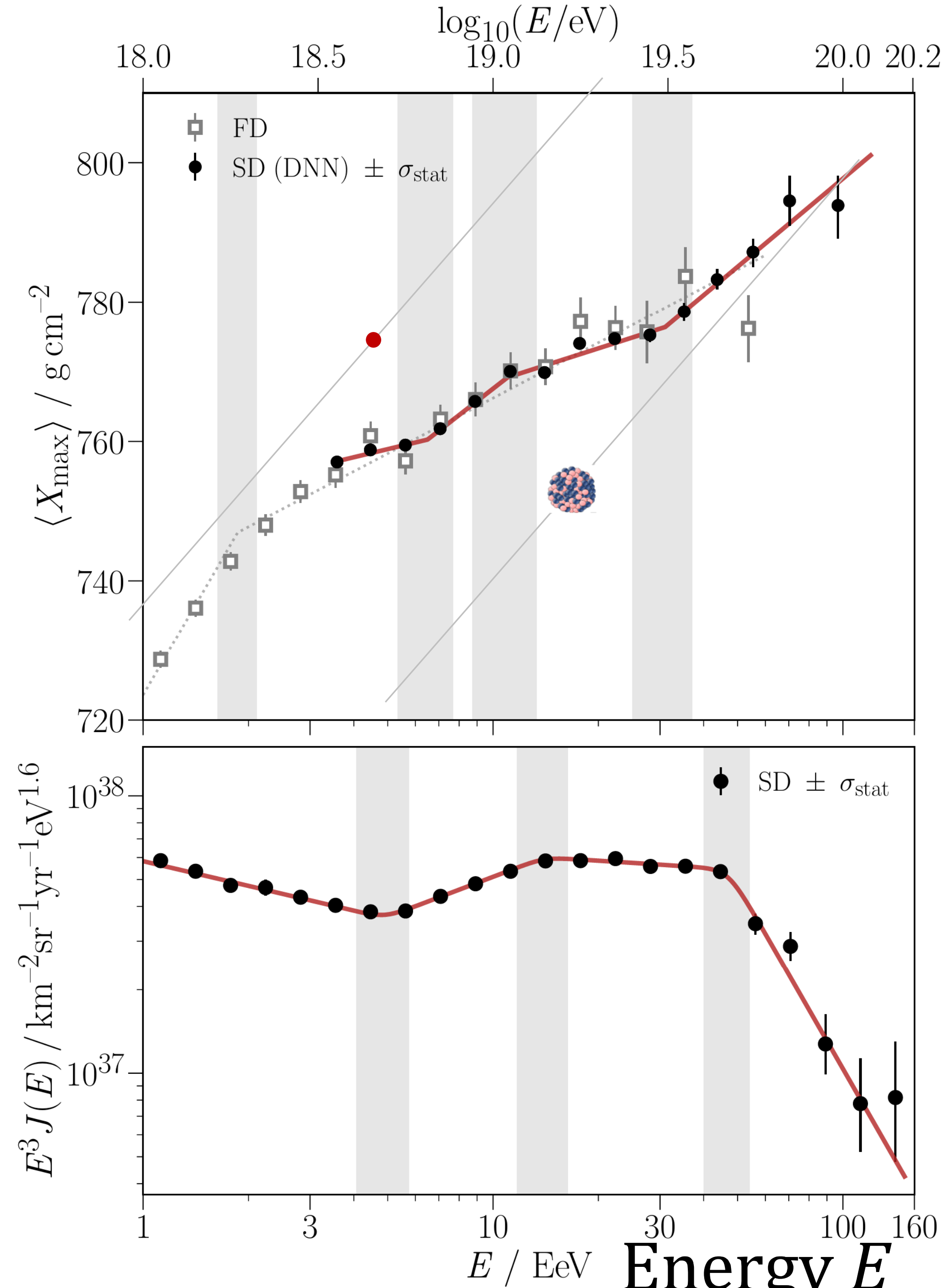
Flux



Energy E

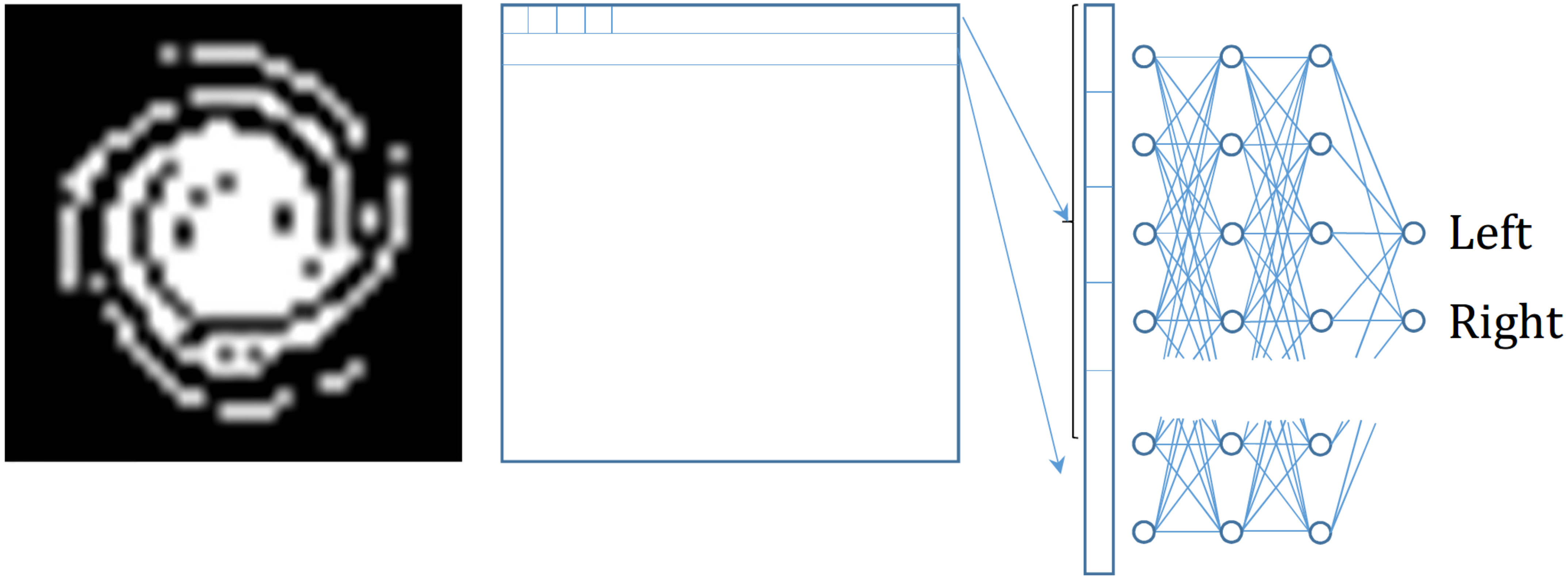
→ $\langle X_{\max} \rangle$

→ E^3Flux



Energy E

News: Common Changes in Energy and Mass Spectra



Convolutional Networks

…looking for better ways than 1 pixel = 1 network input node

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

Convolutional network to analyse image-like data

Hierarchy of ’features’ with increasing level of complexity

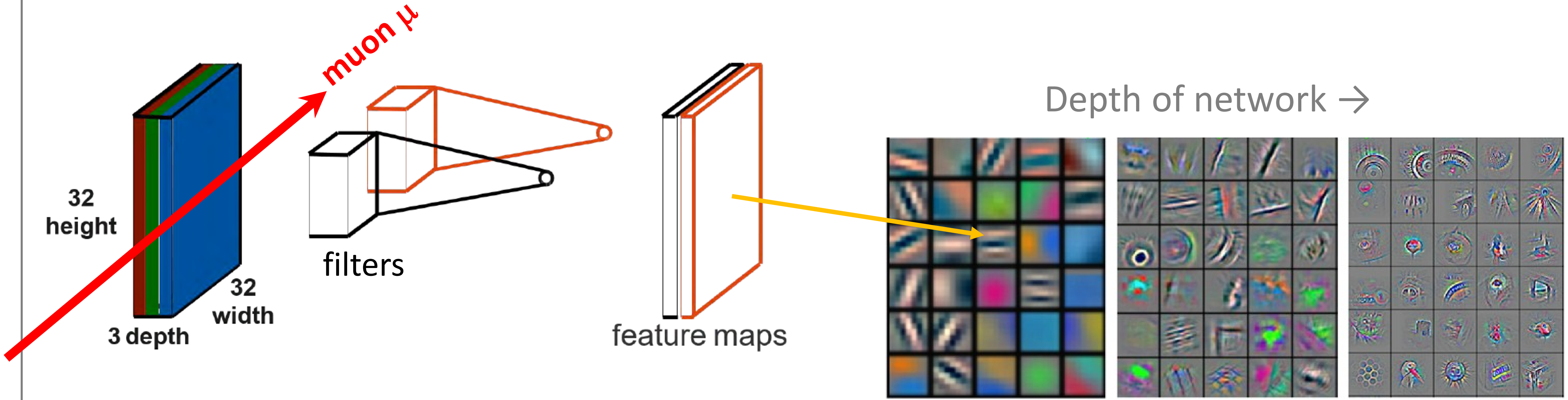
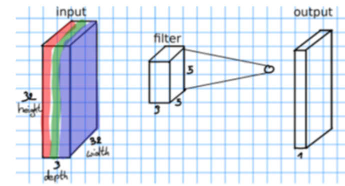


Image recognition

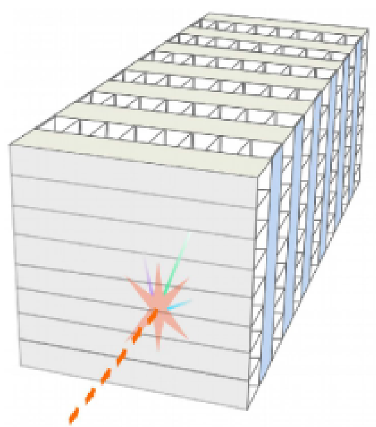
Pixel → edge → texton → motif → part → object

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

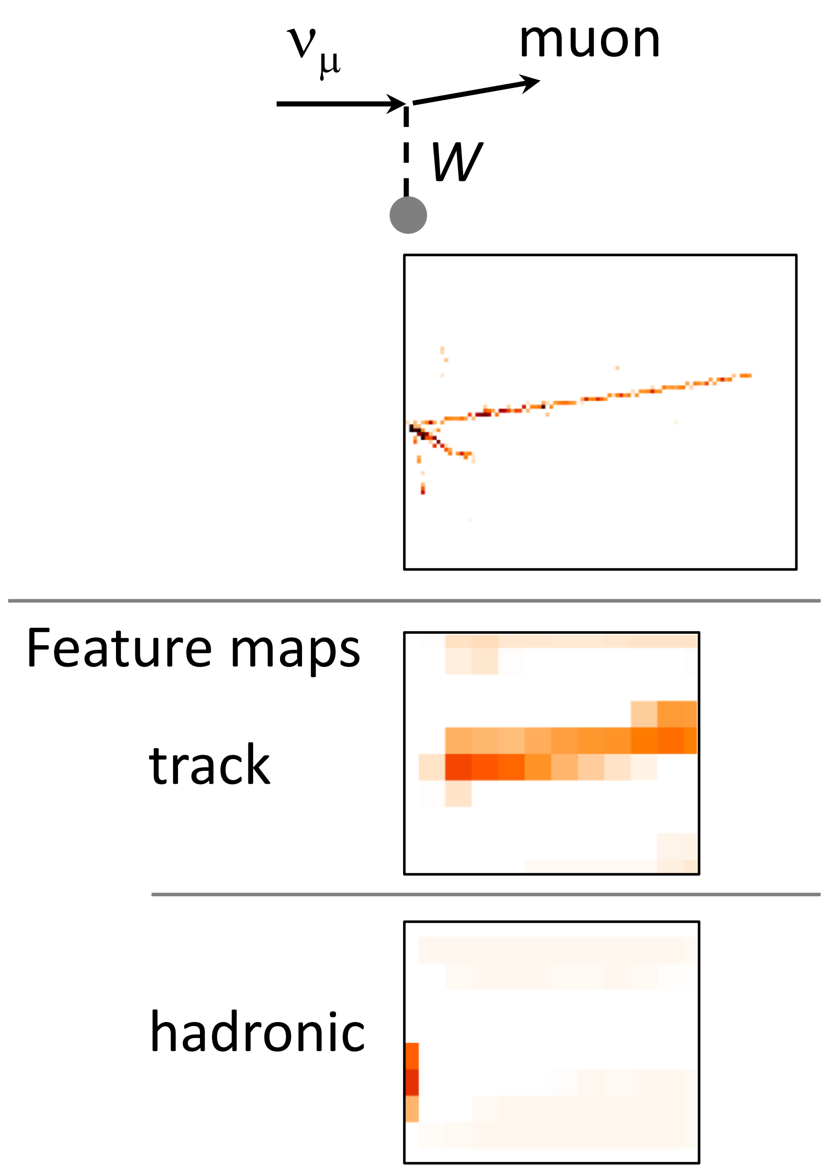
Convolutional network to identify electron neutrinos



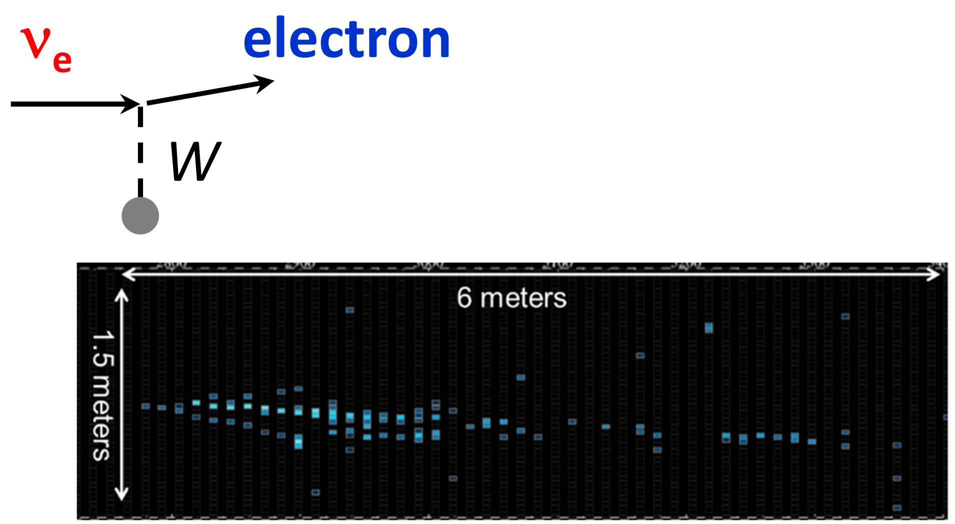
Fermilab (Chicago)
→NOvA experiment 810km



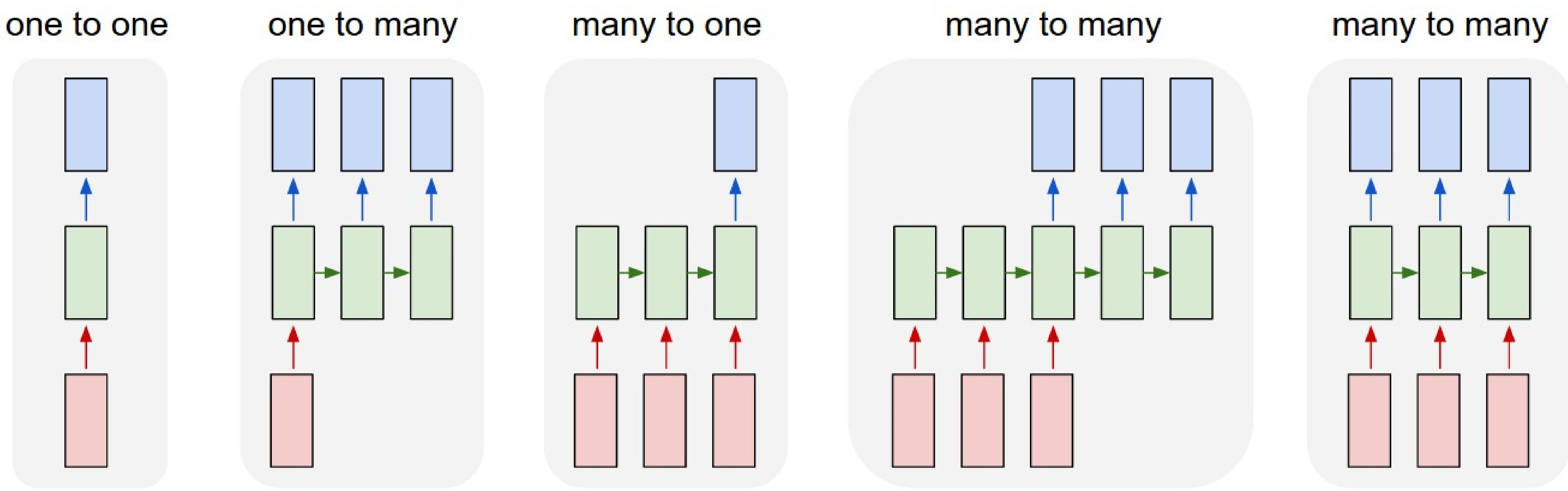
Neural network neutrino event classifier



Challenge: electron-neutrinos



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



Fully connected,
Convolutional
networks

Image: Kaparthy

From Recurrent Networks to Transformers

to analyse input data of variable size

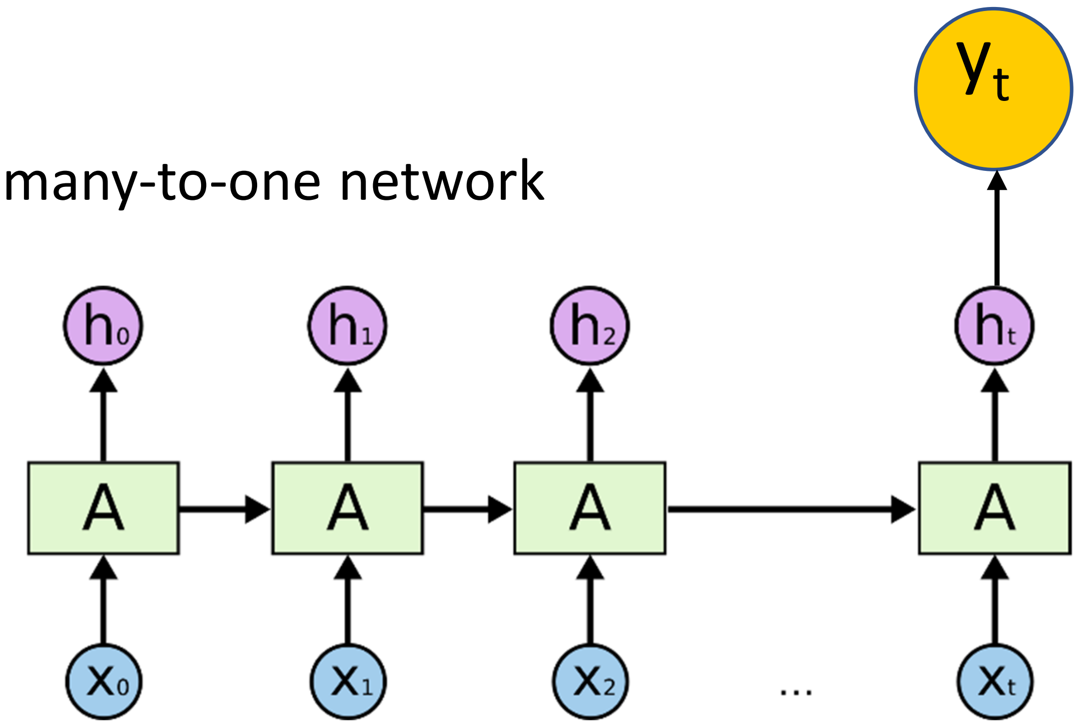
Recurrent networks

D.E. Rumelhart, G.E. Hinton, R.J. Williams, Nature 323(9-Oct-1986)533
S. Hochreiter, J. Schmidhuber, Neural Computation 9(1997)1735

Hidden state depends on previous state

$$h^{(t)} = \tanh\left(U\, h^{(t-1)} + W\, x^{(t)} + b \right)$$

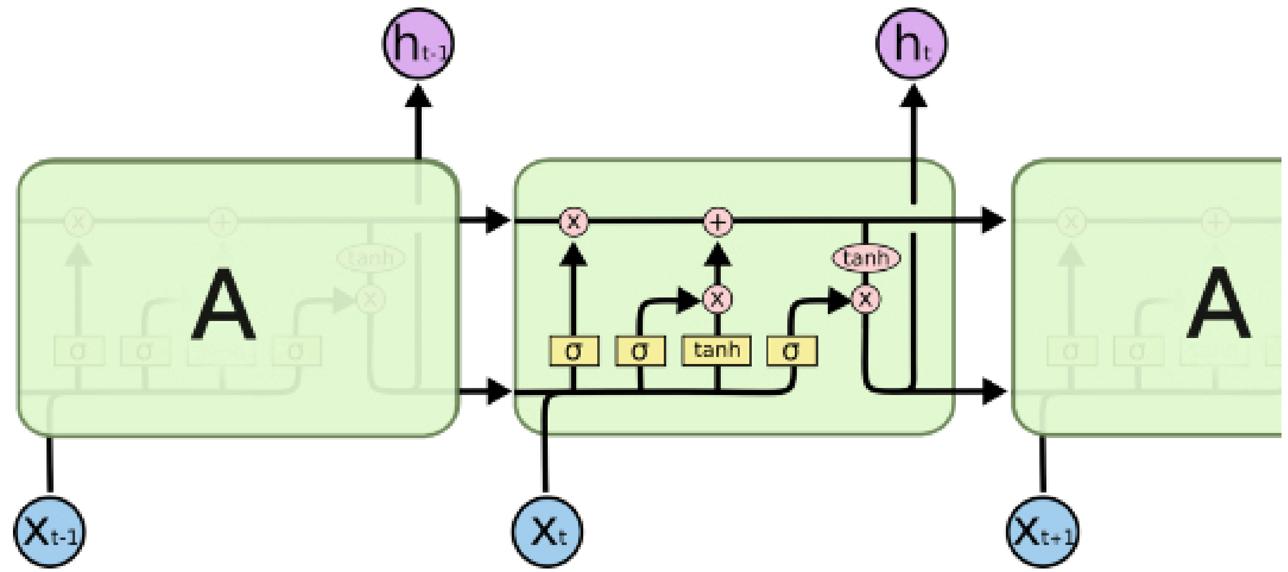
e.g. many-to-one network



Efficient parameter sharing: same *W, U, b* for all input data

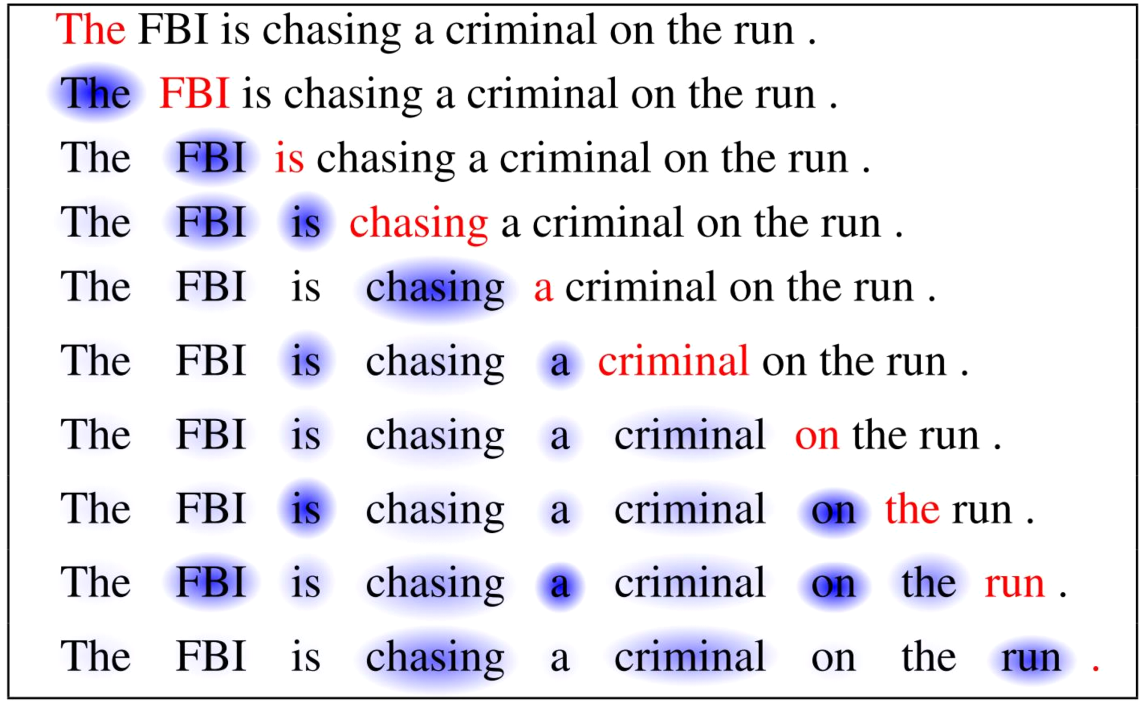
Long **S**hort **T**erm **M**emory

E.g. detecting correlations between tracks on input far away from each other through a **cell memory**



Attention Mechanism

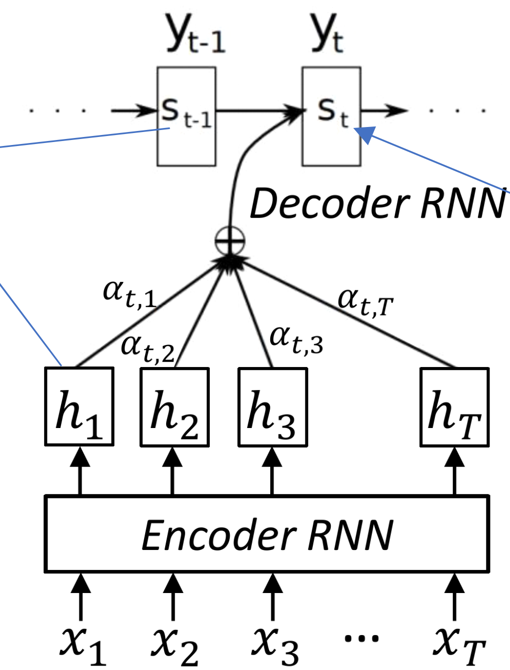
J. Cheng, L. Dong, M. Lapata, arXiv:1601.06733
D. Bahdanau, K. Cho, Y. Bengio, arXiv:1409.0473
BLEU (bilingual evaluation understudy)



score function

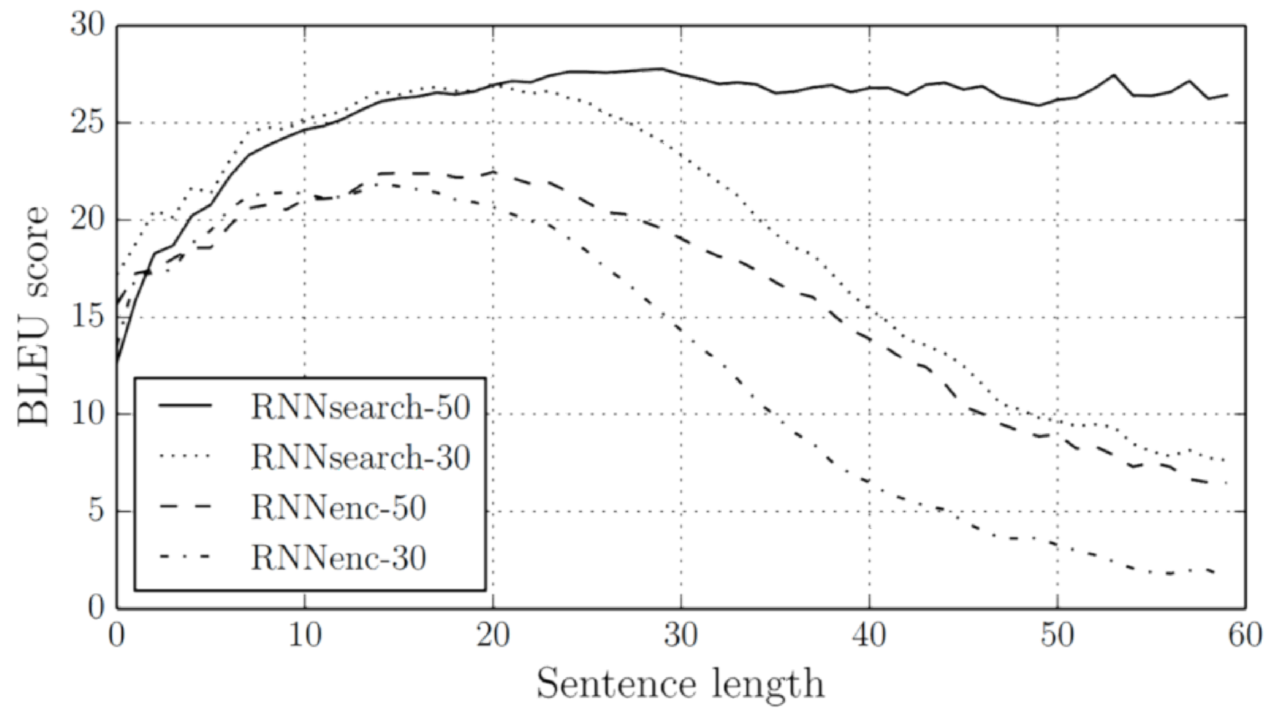
$$e_{tj}(s_{t-1}, h_j) = \frac{s_{t-1}^{T}\, h_j}{\sqrt{n}}$$

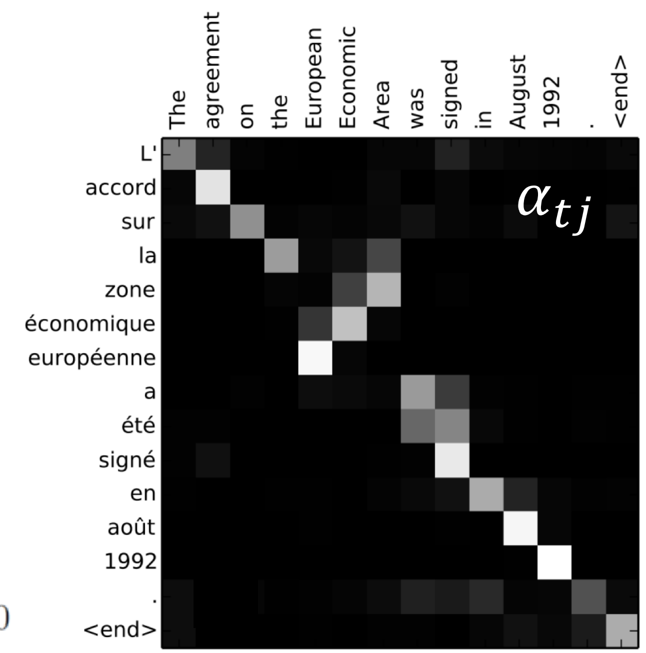
normalized weight α_{tj}



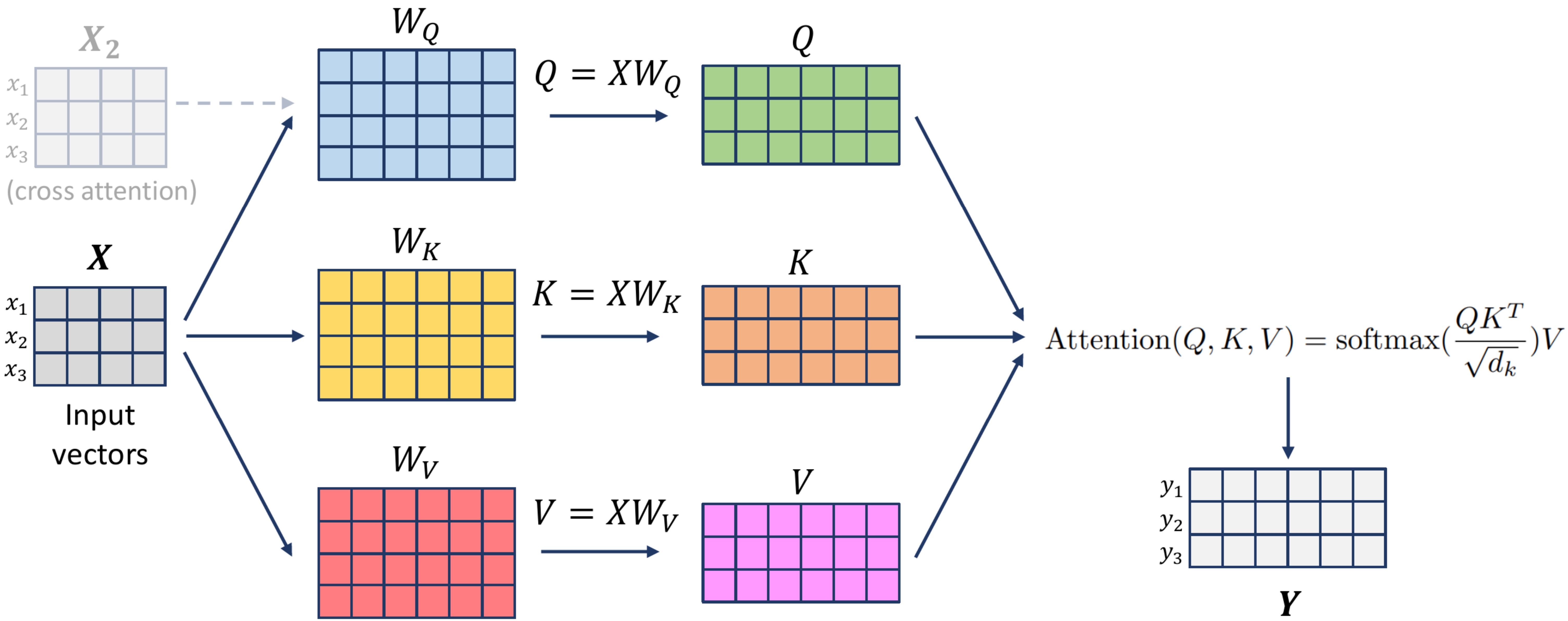
Combine next state s_t

$$c_t = \sum_{j=1}^{T} \alpha_{tj}\, h_j$$

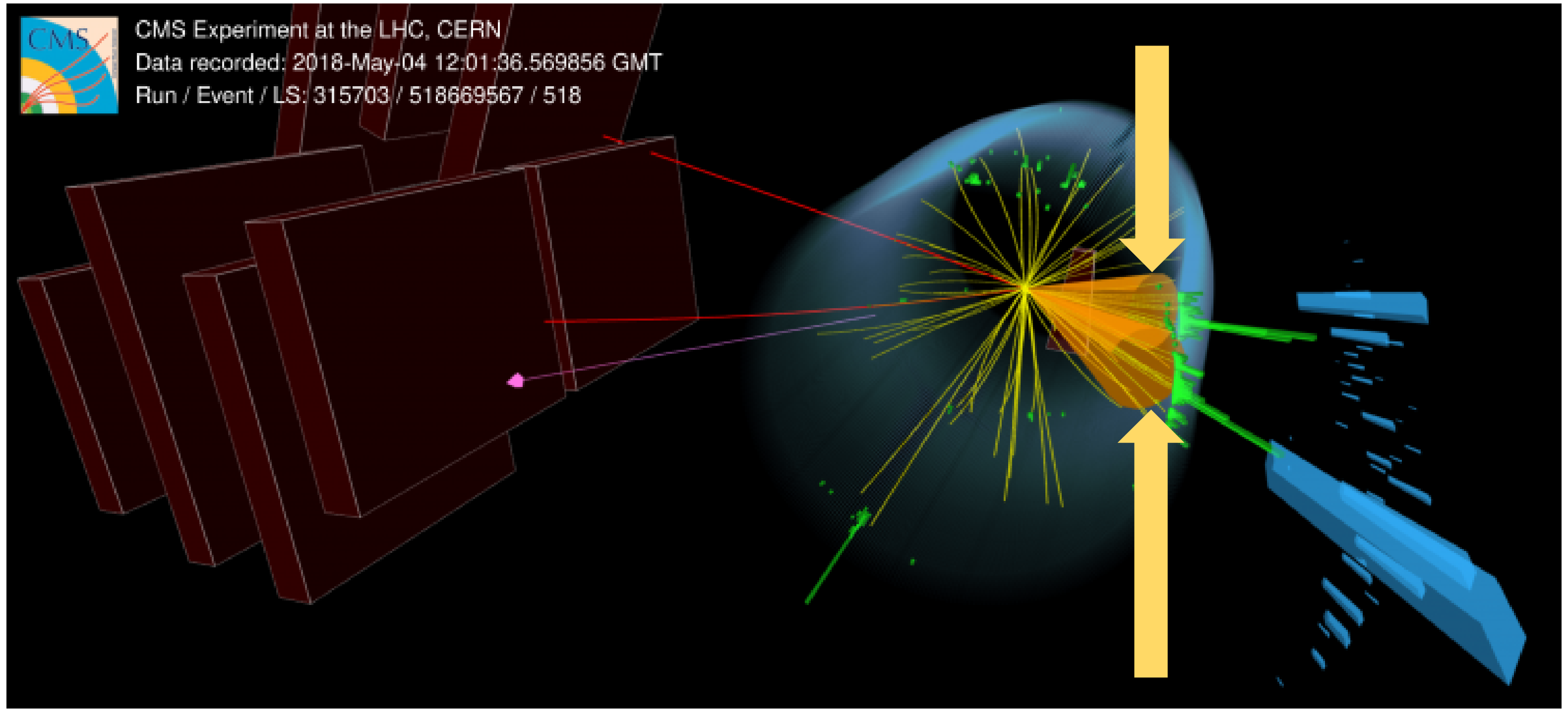




Transformer: `Attention Is All You Need´



Identifying Higgs bosons at the LHC



1 Higgs-Decay → 2 Bottom-Quarks

Probability identifying

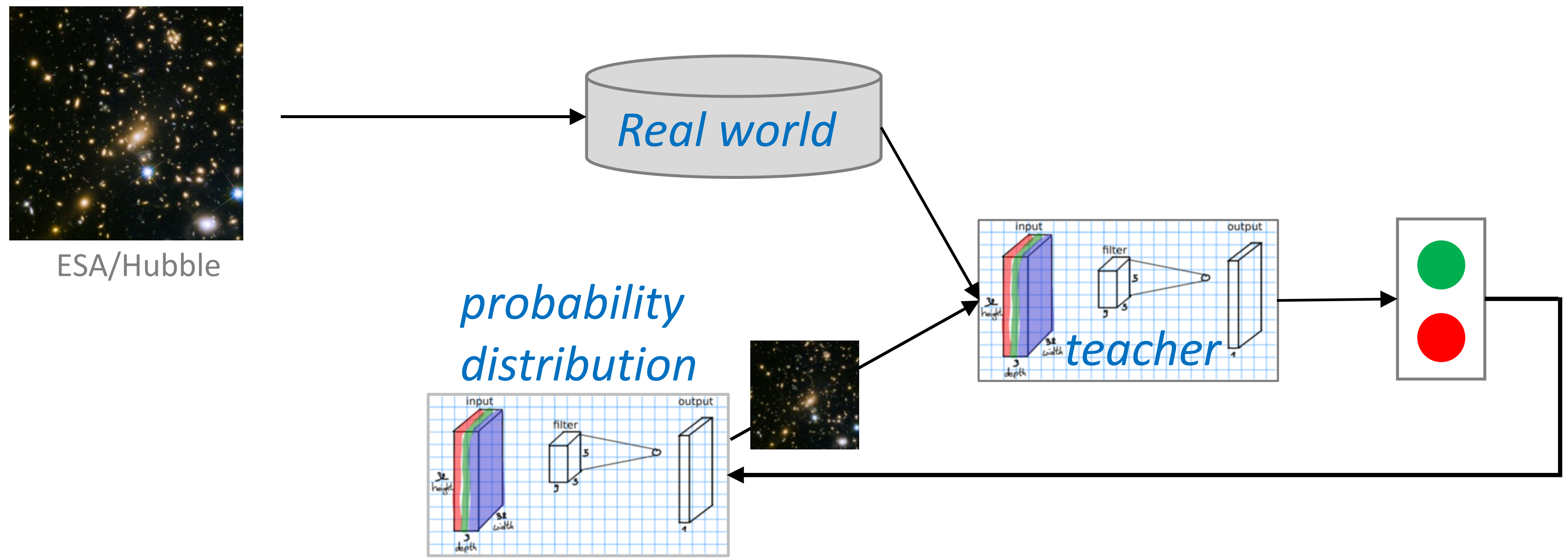
Transformer

1 Bottom-Quark $p = 65\% \rightarrow 80\%$

Transformer

2 Higgs-Bosons → 4 Bottom-Quarks $p^4 = 18\% \rightarrow 41\%$

AI: same signal after half the LHC operating time (>2029)



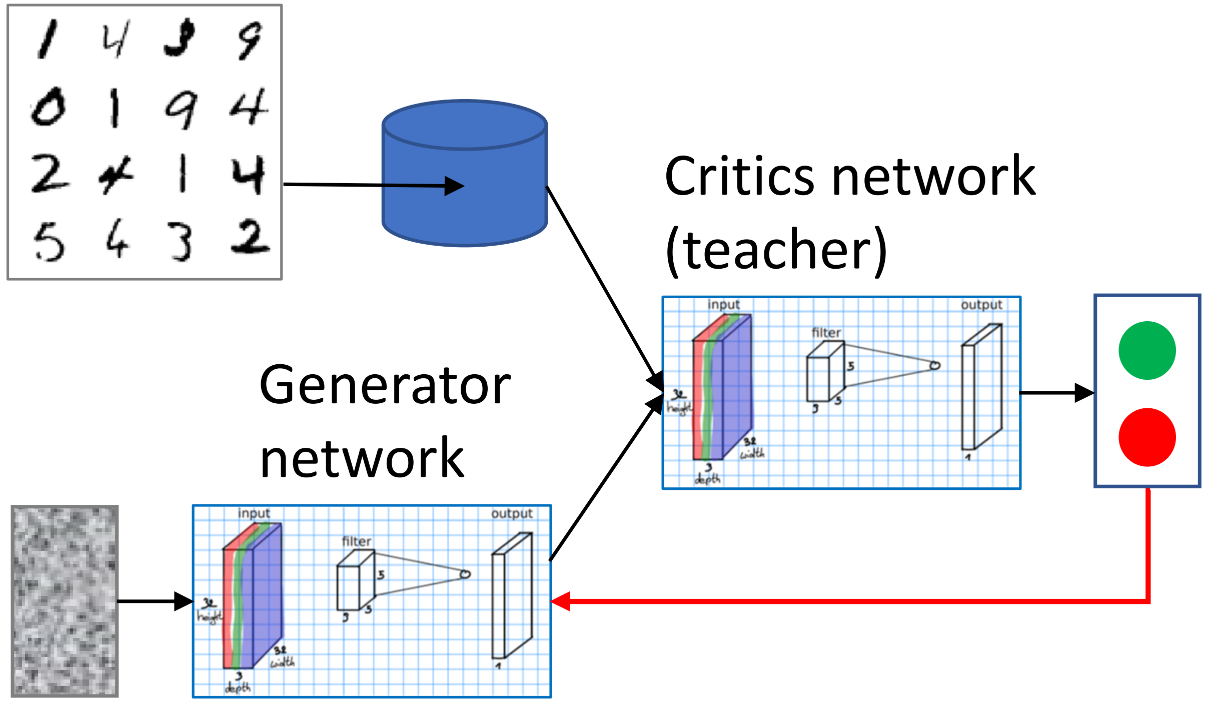
Autonomous model building

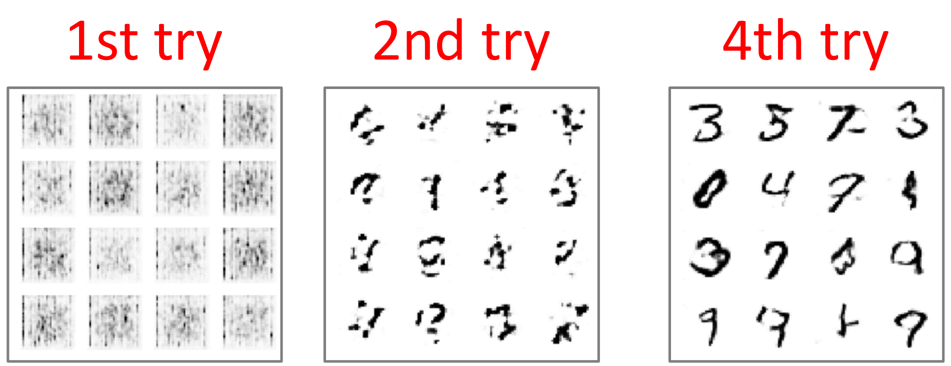
Assign functional target → training data optimize network (`unsupervised´)

Generative Modeling to simulate particle showers

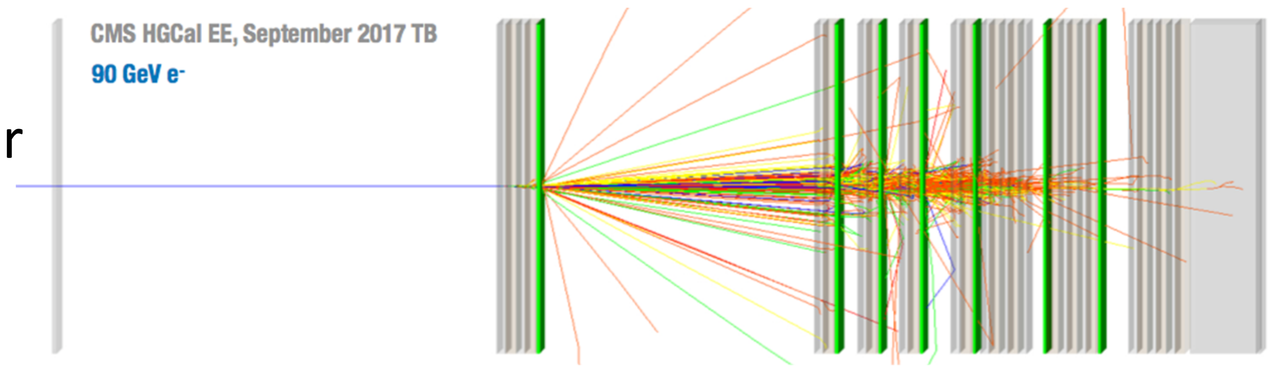
Real world images

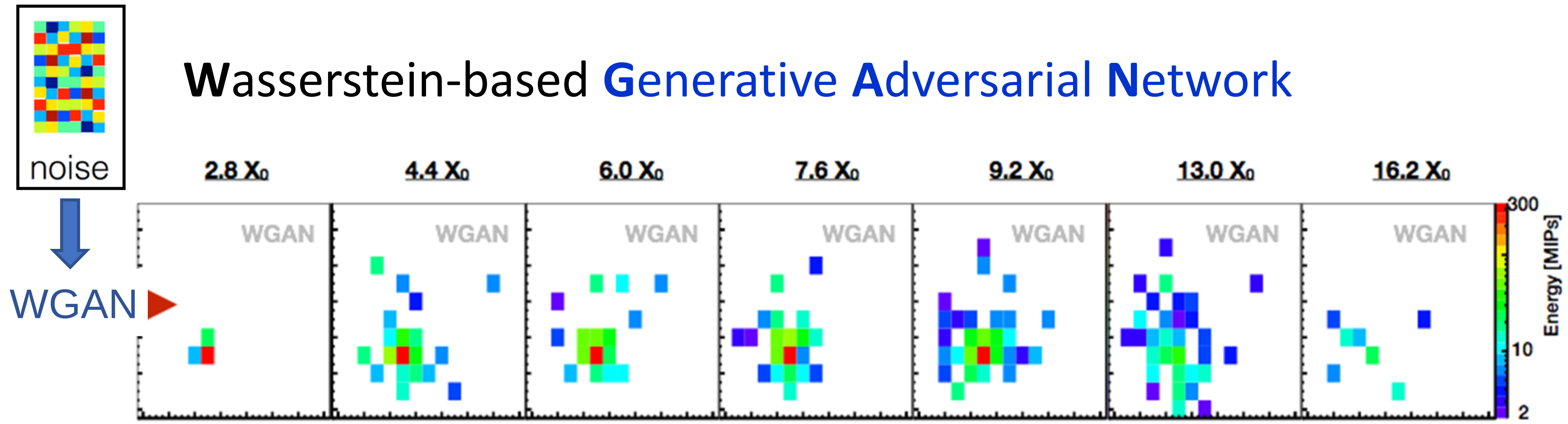
(‘unsupervised‘)





Electron calorimeter in CERN test beam





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

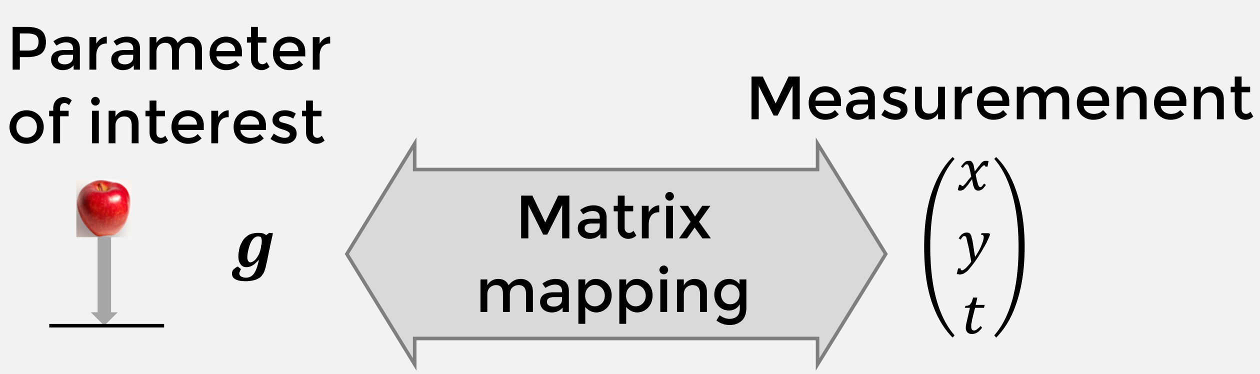
R&D: Ultrafast detector simulations of high quality.
Huge effort ahead for production versions

A wealth of possibilities: Invertible Problems

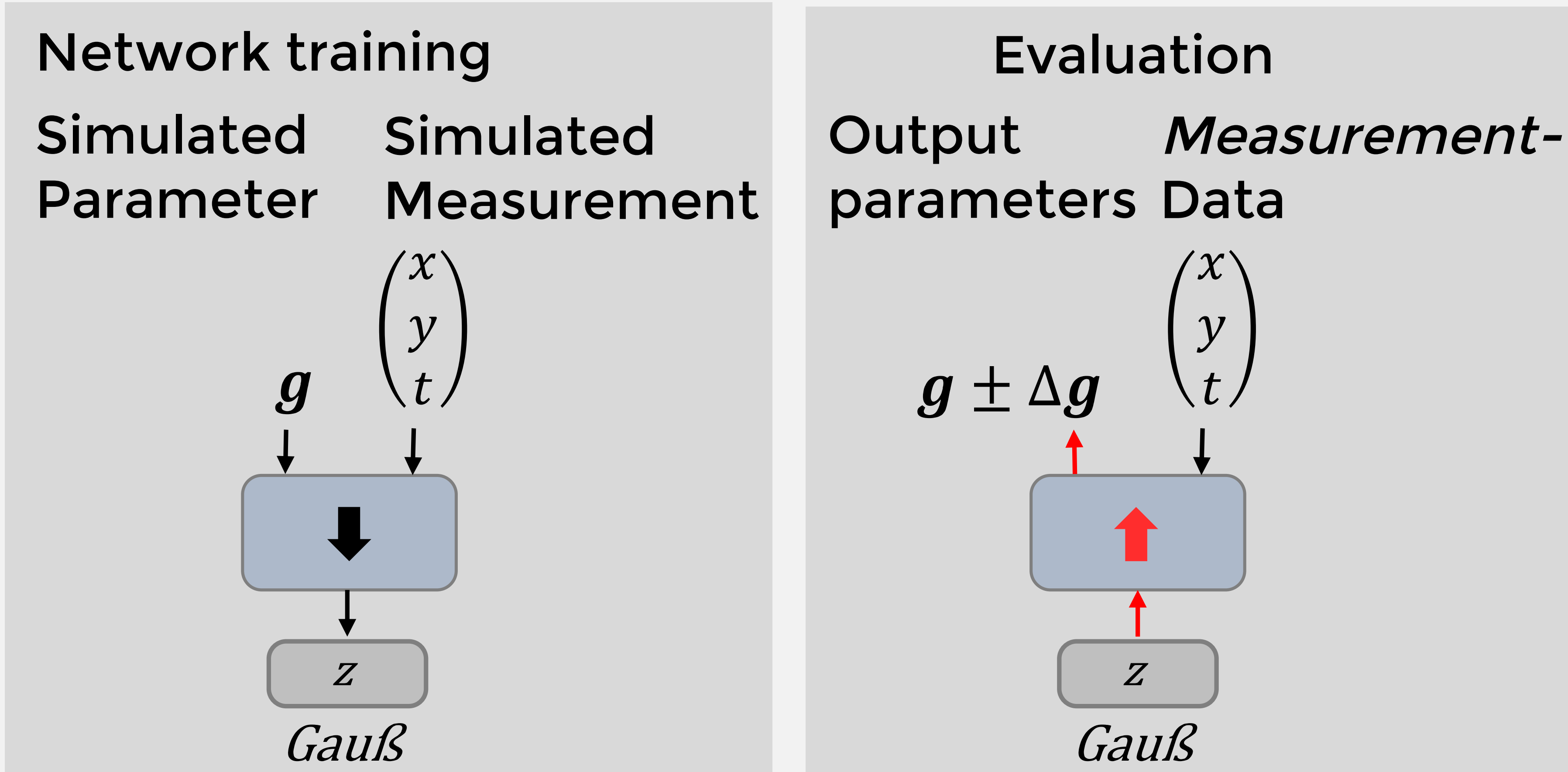
Phenomenon of nature → Measurement

← Invert

So far: Mathematics matrix inversion



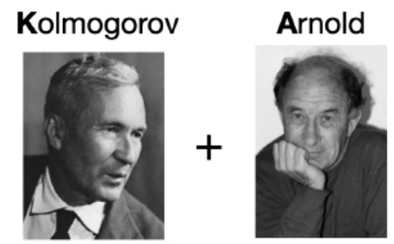
AI: Invertible networks



Further impact & developments

'Reconciliation' of AI with classical physics

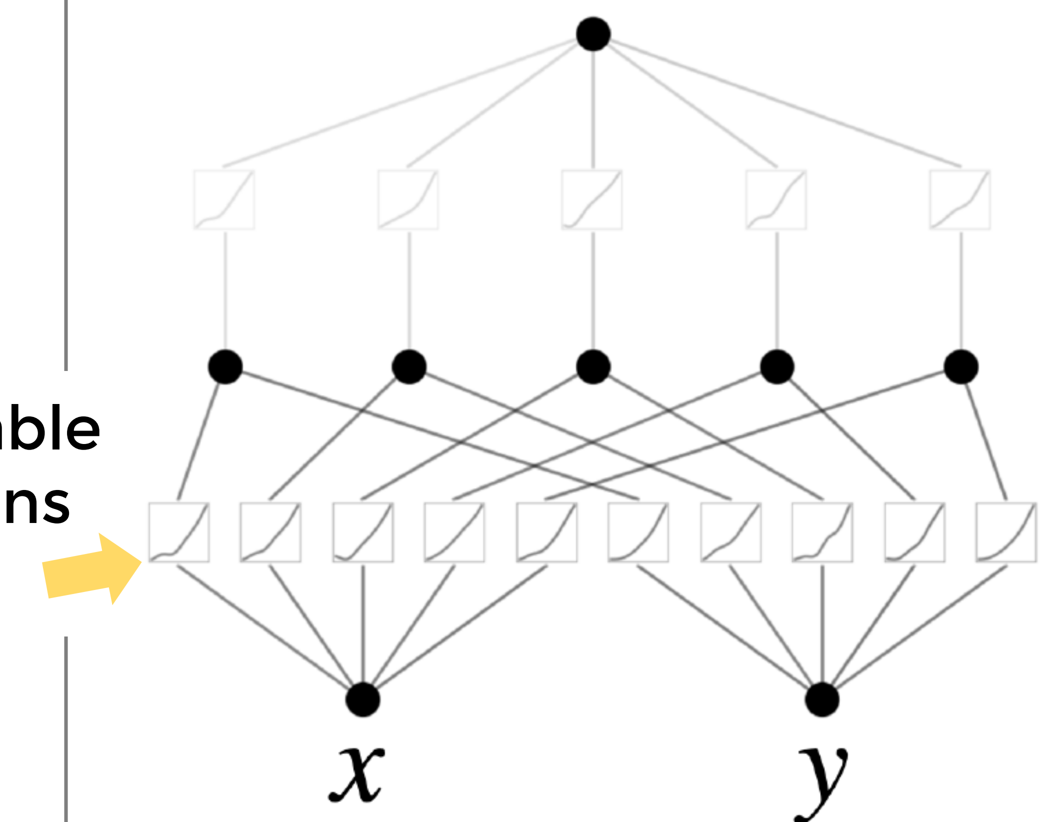
arXiv:2404.19756v2, 2-May-2024



$$f(\mathbf{x}) = f(x_1, \cdots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$$

Readable neural networks : Kolmogorov–Arnold (KAN)

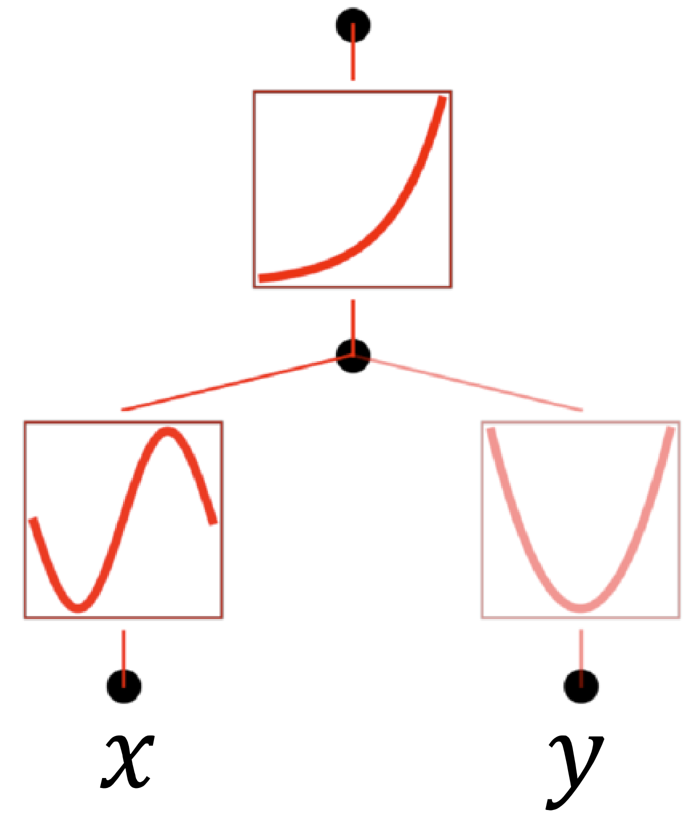
Adaptable functions



Data model $f = e^{\sin(\pi x)+y^2}$

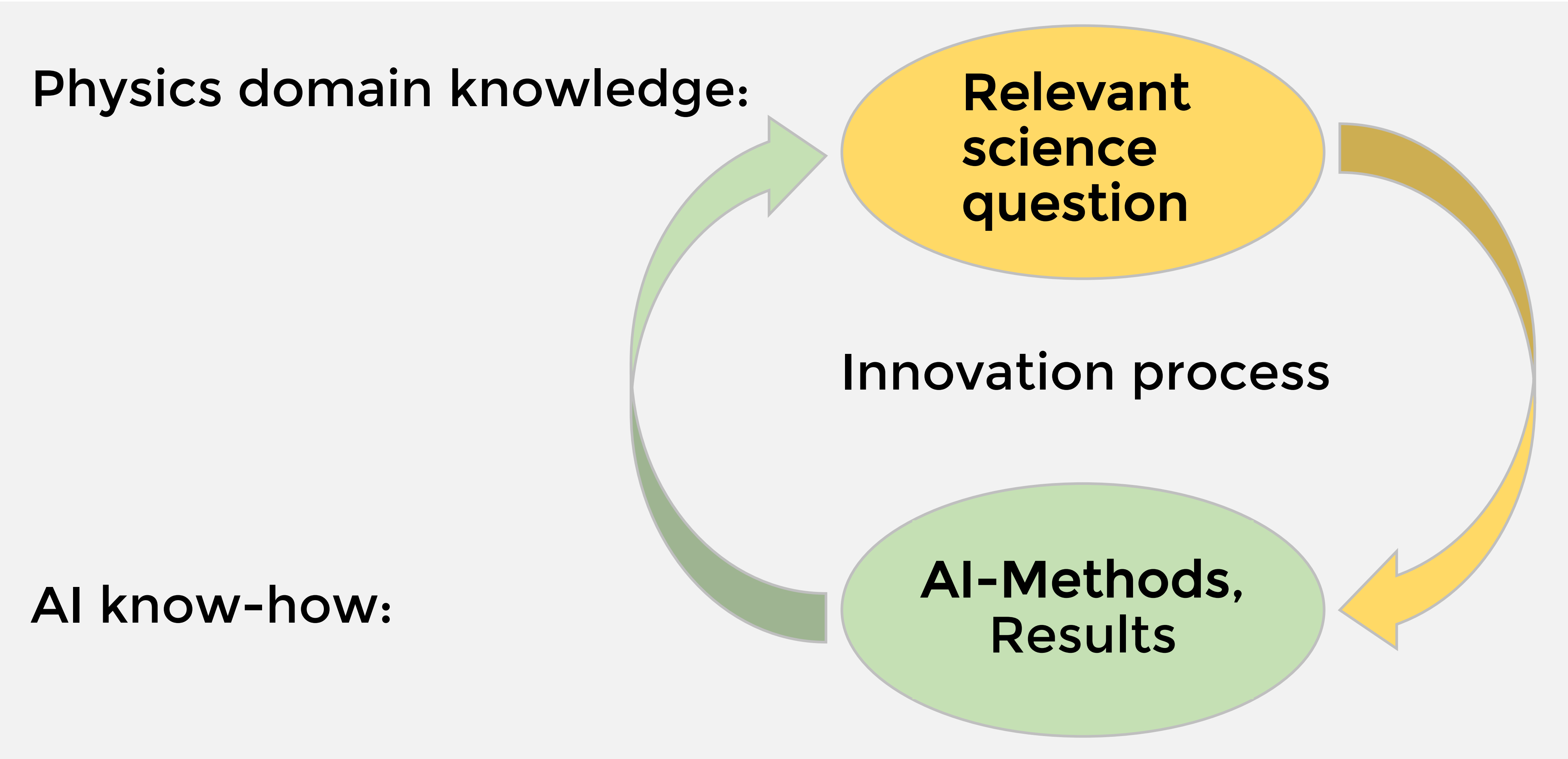
Network training

$f = e^{\sin(\pi x)+y^2}$



Summary: High Research & Innovation Potential

Physics expertise *and* AI knowledge



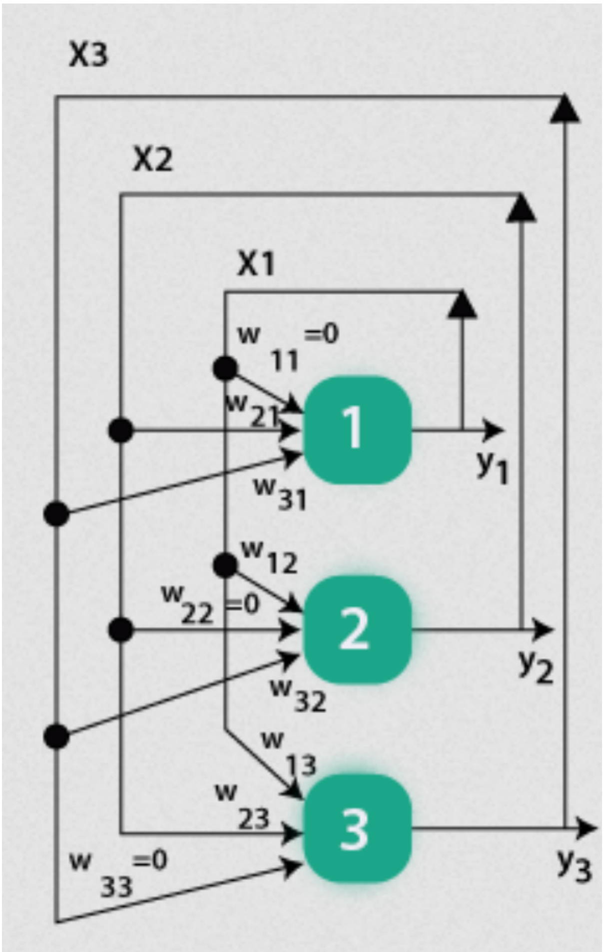
Computer Science Methods adapted, e.g. Deep Learning

- Fully connected
- Convolutional
- Graph
- Recurrent
- Autoencoder
- Generative Adversarial
- Normalizing Flows
- Diffusion
- Reinforcement
- Transformer
- Kolmogorov-Arnold
- Lorentz Boost

backup

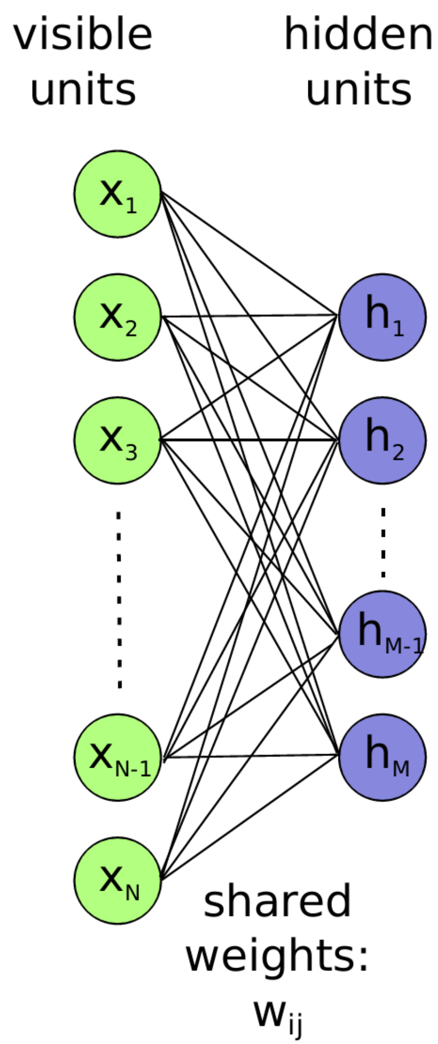
Early Networks

Hopfield Network

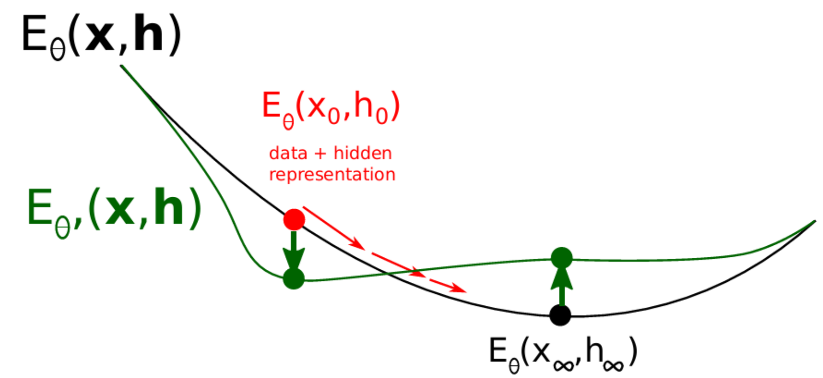


$$E = \frac{1}{2}\sum_{i,j=1}^{N} W_{ij}X_iX_j$$

Hinton: Restricted Boltzmann Machine



Unsupervised learning
→ Reconstruct the input
Minimize scalar energy



Organisations

National BMBF Action Plan „ErUM-Data“ Community Self-organization „DIG-UM“

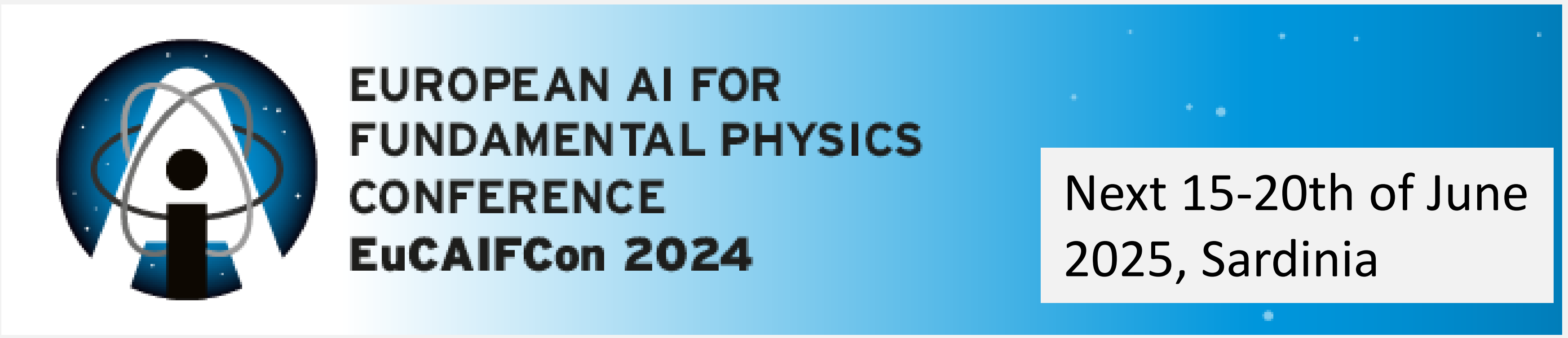
ERUM DATA HUB

Astronomy, accelerator, particles, astroparticles, hadrons+nuclei, synchrotron, neutrons, ions

DIG-UM

NFDI Helmholtz-Initiativen Datenkompetenzzentren

International Self-organization „EuCAIF“



EuCAIF is an European initiative for advancing the use of Artificial Intelligence (AI) in Fundamental Physics.

EOSC WLCG USA: HEP-Software Foundation