CERN 70: A Laboratory for AI Research

CERN 70th Anniversary, Bergen Norway
Thea Klæboe Årrestad (ETH Zürich)

Sep 29th 1954

La sixième session du Conseil fut organisée à Paris du 29 juin au
1er juillet 1953. C'est à cette occasion que la Convention établissant l'Organisation fut signée, sous réserve de ratification, par douze Etats membres.

For the German Federal Republic — Pour la République Fédérale d'Allemagne
subject to ratification

For the Kingdom of Norway — Pour le Royaume de Norvège
Subject to ratification

For the Kingdom of Belgium — Pour le Royaume de Belgique
sous réserve de ratification

For the Kingdom of the Netherlands — Pour le Royaume des Pays-Bas
Subject to ratification

For the Kingdom of Denmark — Pour le Royaume de Danemark
sous réserve de ratification

For the United Kingdom of Great Britain and Northern Ireland — Pour le Royaume-Uni de la Grande-Bretagne et de l'Irelande du Nord
Subject to ratification.

For the French Republic — Pour la République Française
sous réserve de ratification

For the Kingdom of Sweden — Pour le Royaume de Suède
Torsten Gustafson
Subject to ratification

For the Kingdom of Greece — Pour le Royaume de Grèce
sous réserve de ratification.

For the Confederation of Switzerland — Pour la Confédération Suisse
sous réserve de ratification

For Italy — Pour l'Italie
sous réserve de ratification

For the Federal People's Republic of Yugoslavia — Pour la République Fédérative Populaire de Yougoslavie
Pavle Savić
sous réserve de ratification

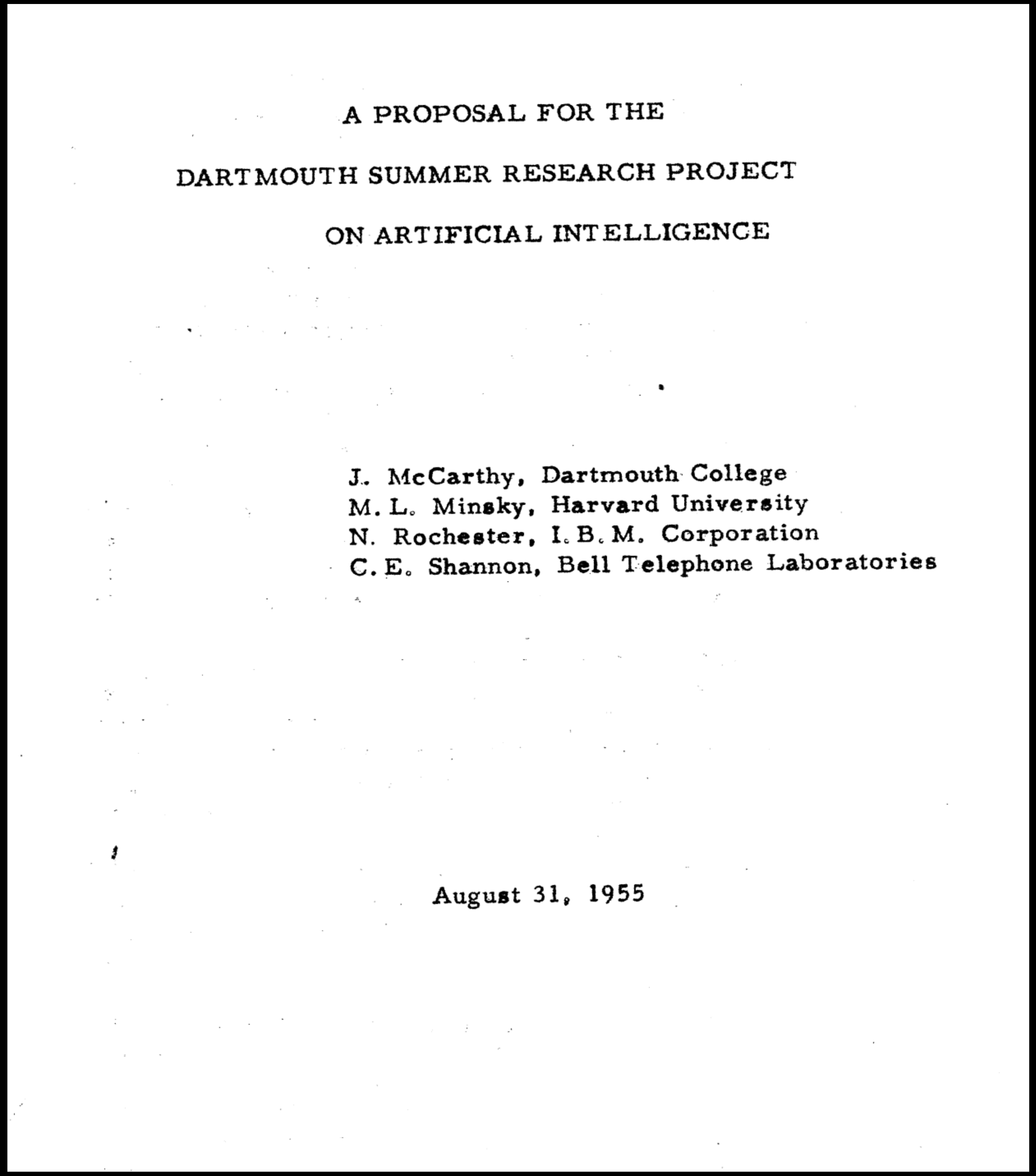
The Sixth Session of the CERN Council took place in Paris on 29 June—1 July 1953. It was here that the Convention establishing the Organization was signed, subject to ratification, by twelve States.

Sep 29th 1954

La sixième session du Conseil fut organisée à Paris du 29 juin au 1er juillet 1953. C'est à cette occasion que la Convention établissant l'Organisation fut signée, sous réserve de ratification, par douze Etats membres.

The Sixth Session of the CERN Council took place in Paris on 29 June—1 July 1953. It was here that the Convention establishing the Organization was signed, subject to ratification, by twelve States.

Sep 2nd 1955



A PROPOSAL FOR THE

DARTMOUTH SUMMER RESEARCH PROJECT

ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I. B. M. Corporation
C. E. Shannon, Bell Telephone Laboratories

August 31, 1955

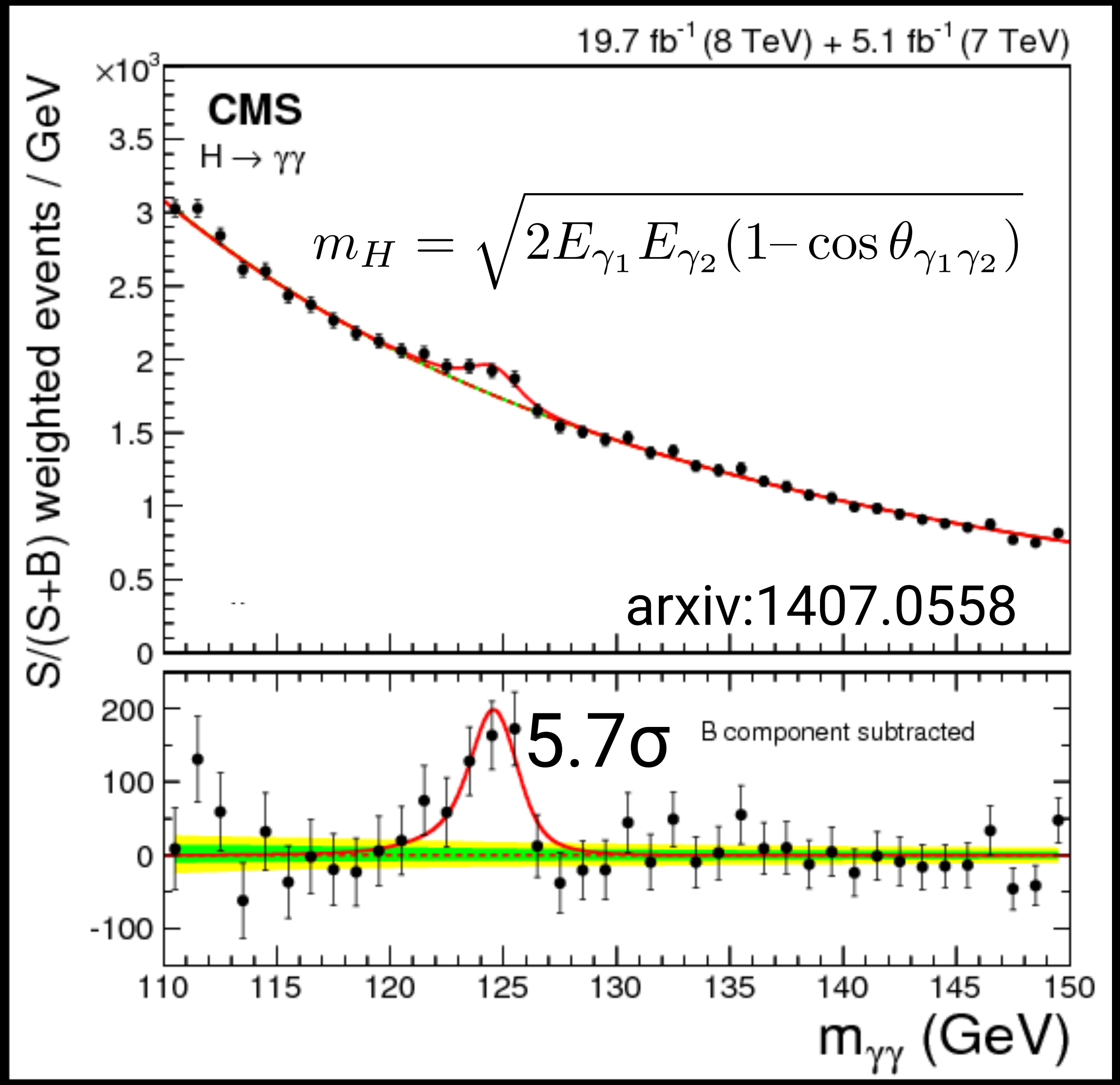
1957

1957

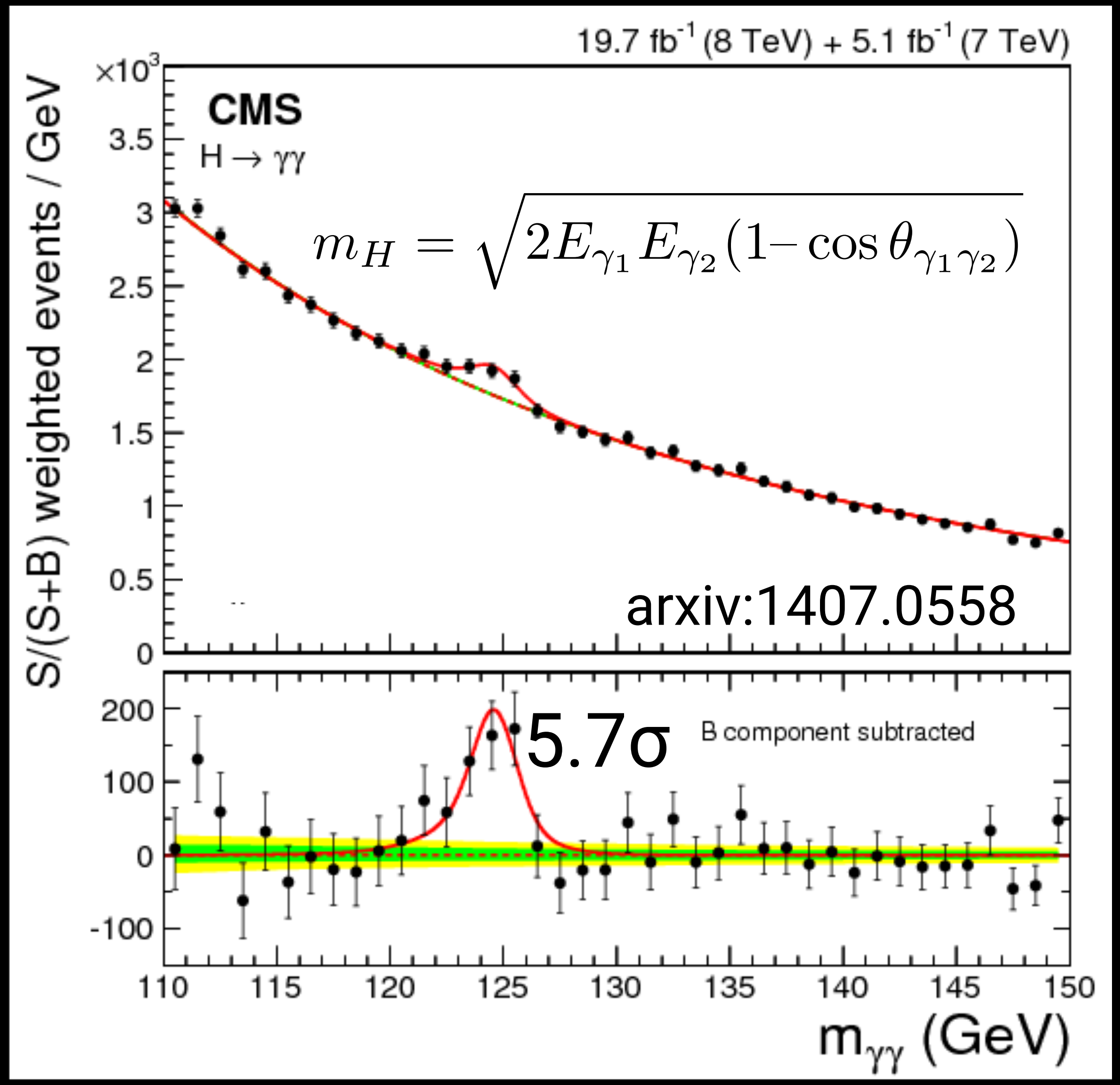
1958

a perceptron “may eventually be able to learn, make decisions, and translate languages.”

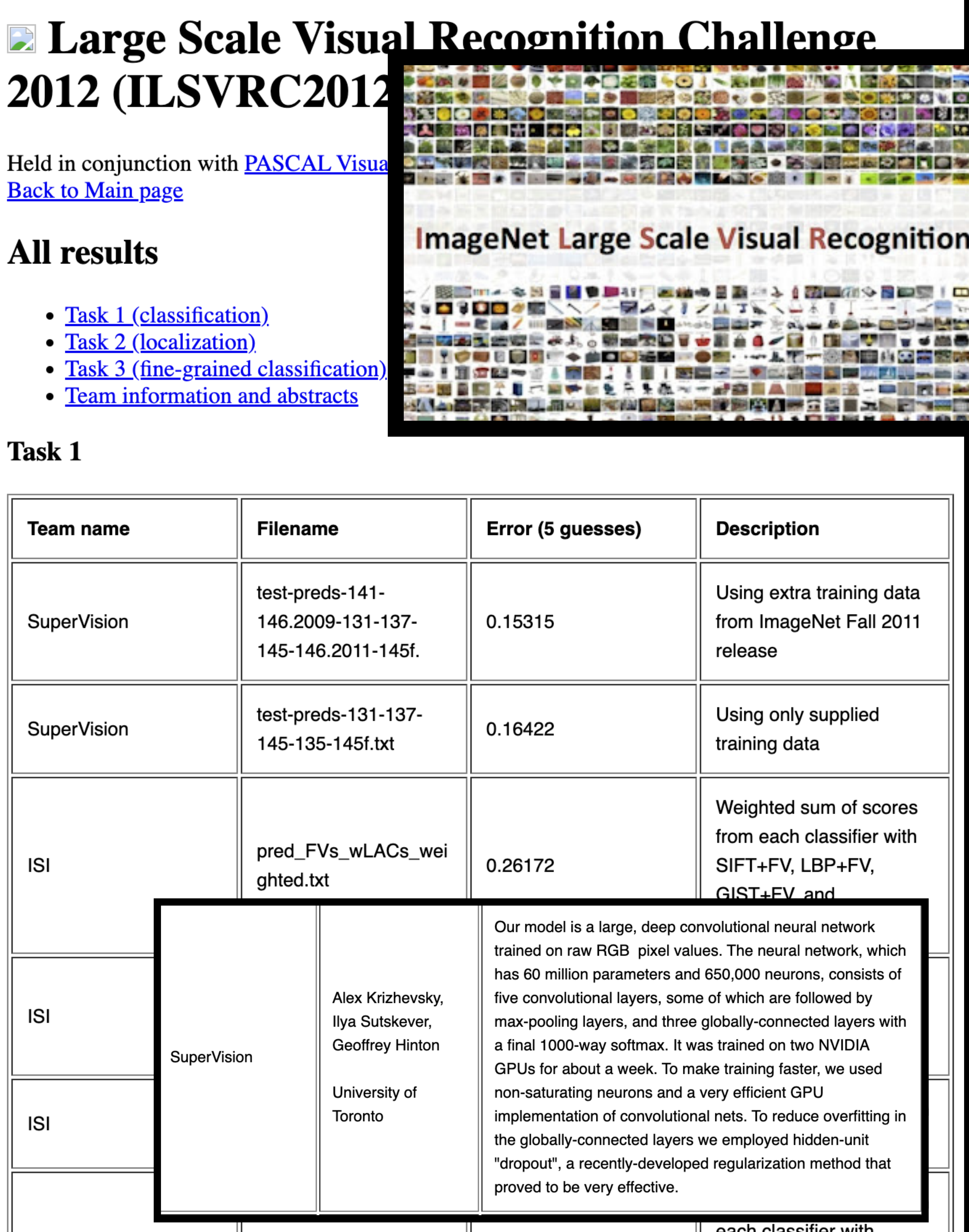
July 4th 2012



July 4th 2012



Sep 30th 2012



Large Scale Visual Recognition Challenge 2012 (ILSVRC2012

Held in conjunction with PASCAL Visua
Back to Main page

All results

- Task 1 (classification)
- Task 2 (localization)
- Task 3 (fine-grained classification)
- Team information and abstracts

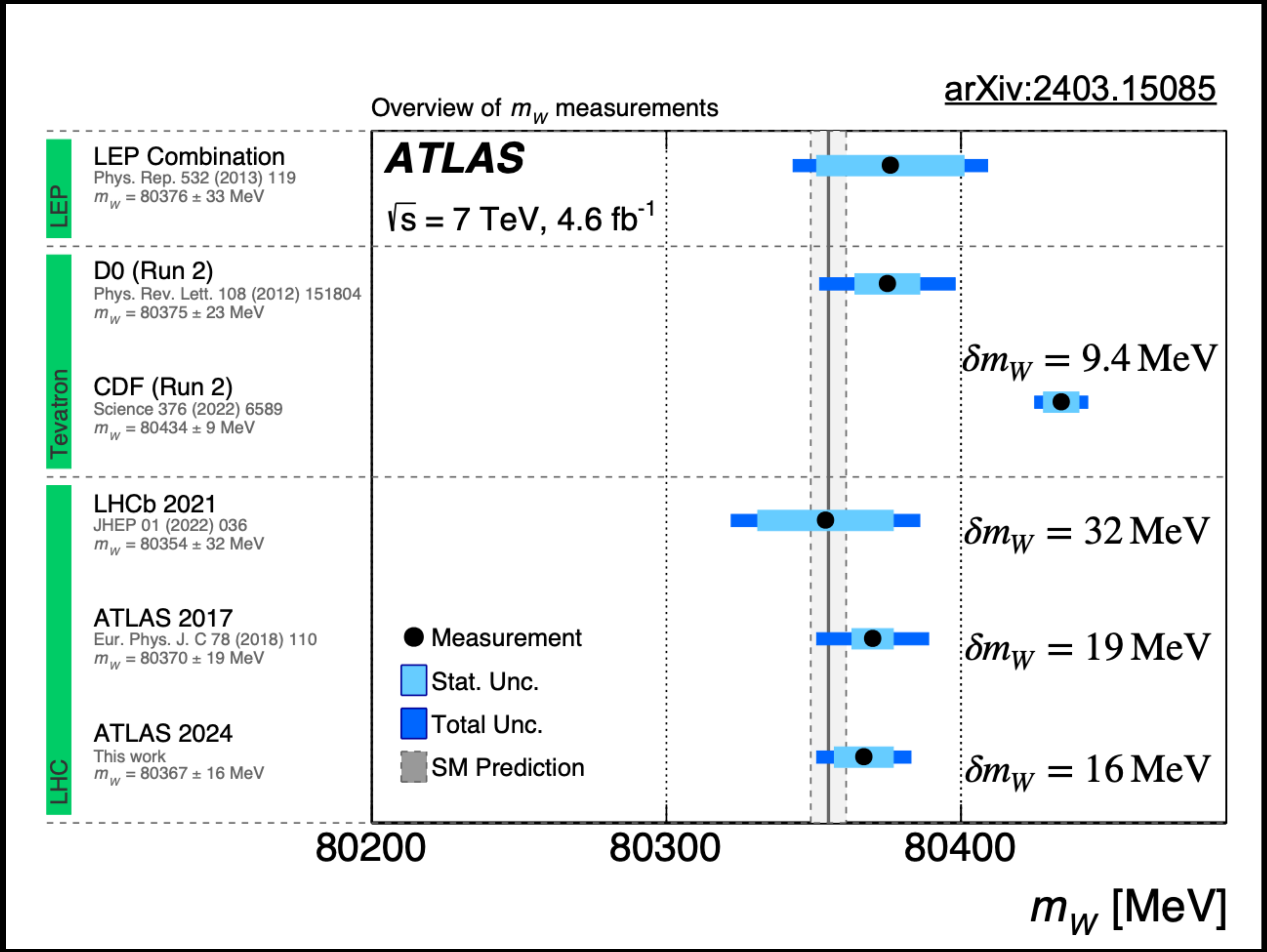
Task 1

Team name	Filename	Error (5 guesses)	Description
SuperVision	test-preds-141-146.2009-131-137-145-146.2011-145f.	0.15315	Using extra training data from ImageNet Fall 2011 release
SuperVision	test-preds-131-137-145-135-145f.txt	0.16422	Using only supplied training data
ISI	pred_FVs_wLACs_weighted.txt	0.26172	Weighted sum of scores from each classifier with SIFT+FV, LBP+FV, GIST+FV, and
ISI			
ISI			
			each classifier with

SuperVision	Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton University of Toronto	Our model is a large, deep convolutional neural network trained on raw RGB pixel values. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three globally-connected layers with a final 1000-way softmax. It was trained on two NVIDIA GPUs for about a week. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of convolutional nets. To reduce overfitting in the globally-connected layers we employed hidden-unit "dropout", a recently-developed regularization method that proved to be very effective.

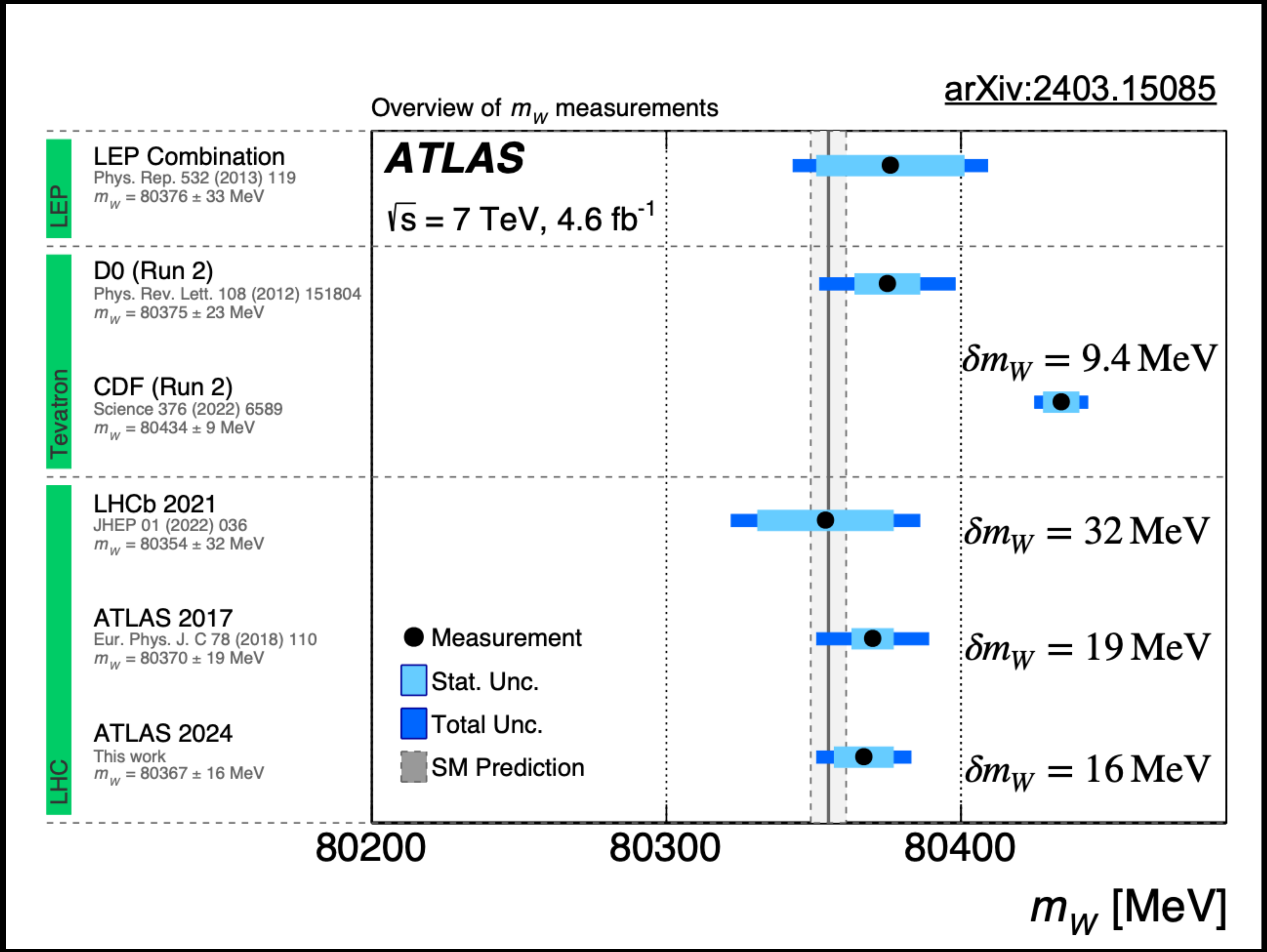
ImageNet Large Scale Visual Recognition Challenges

2024



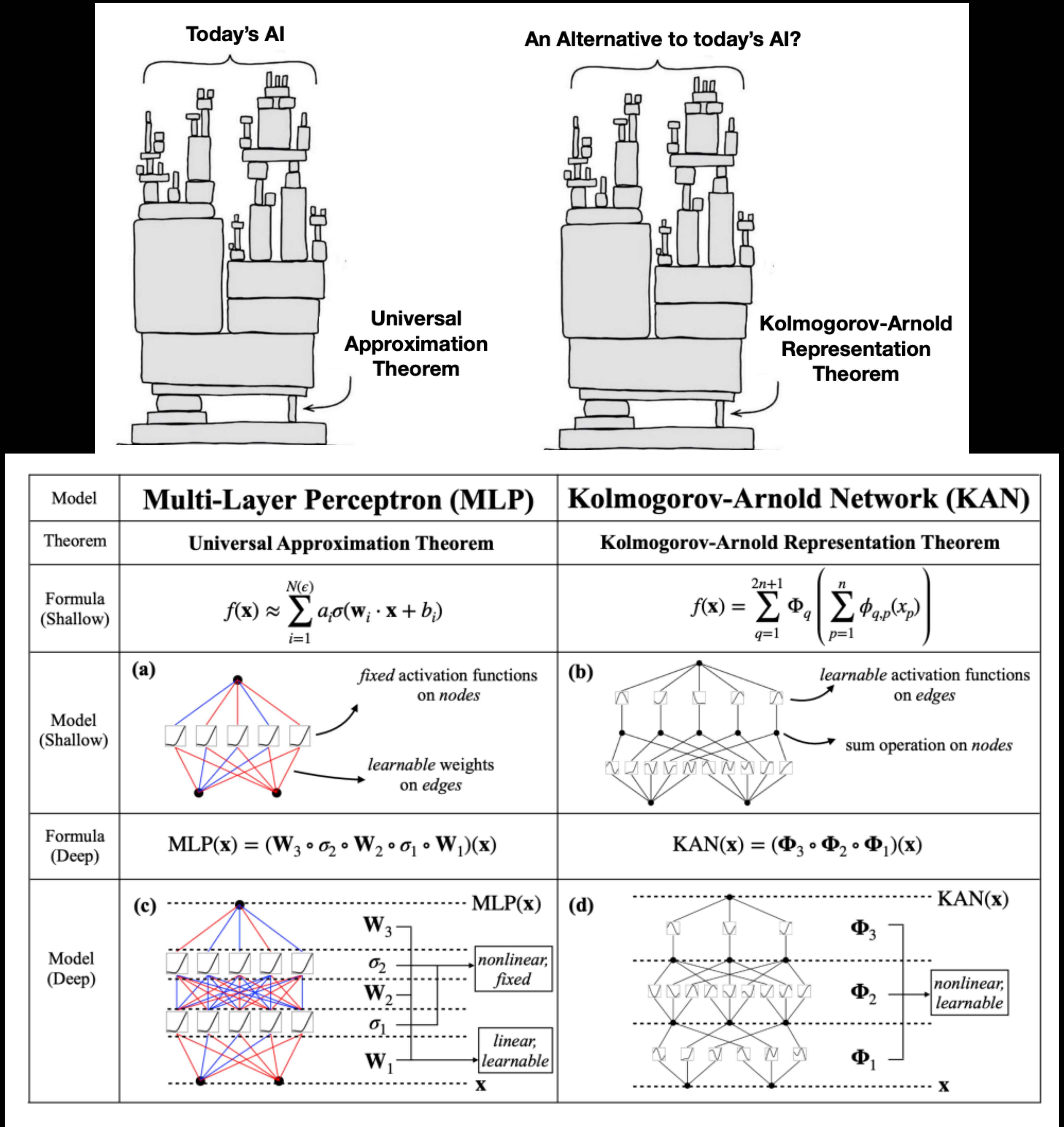
CERN Seminar: W-mass measurement in CMS

2024



[CERN Seminar: W-mass measurement in CMS]

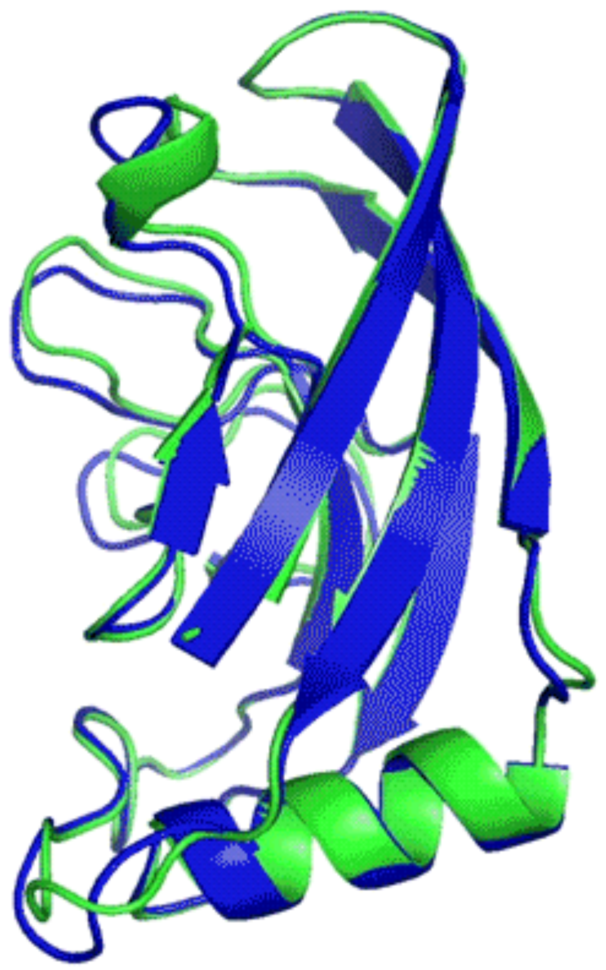
2024



[arxiv:2404.19756]

“AI for accelerated discovery”

T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

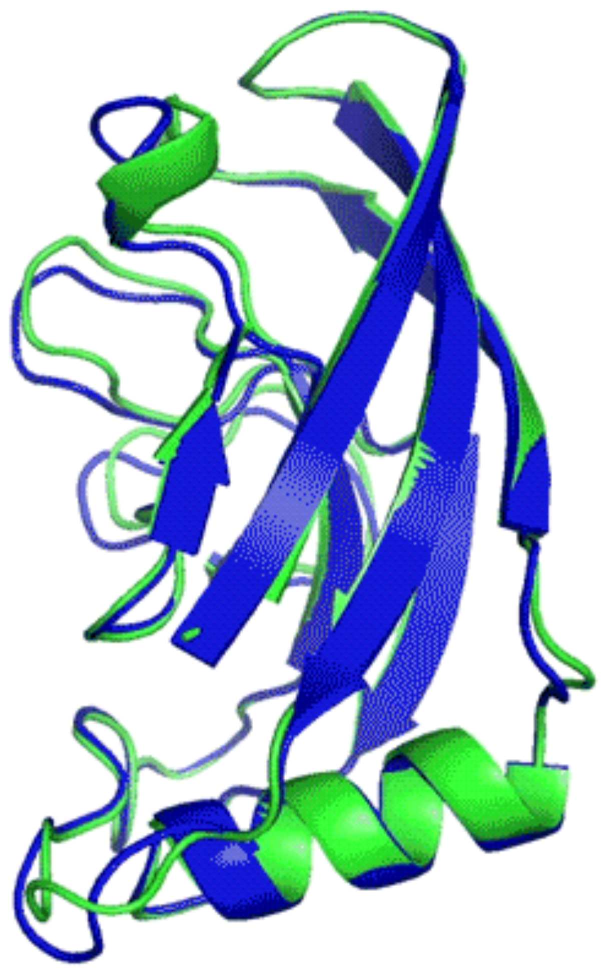


T1049 / 6y4f
93.3 GDT
(adhesin tip)

Experimental result
Computational prediction

“AI for accelerated discovery”

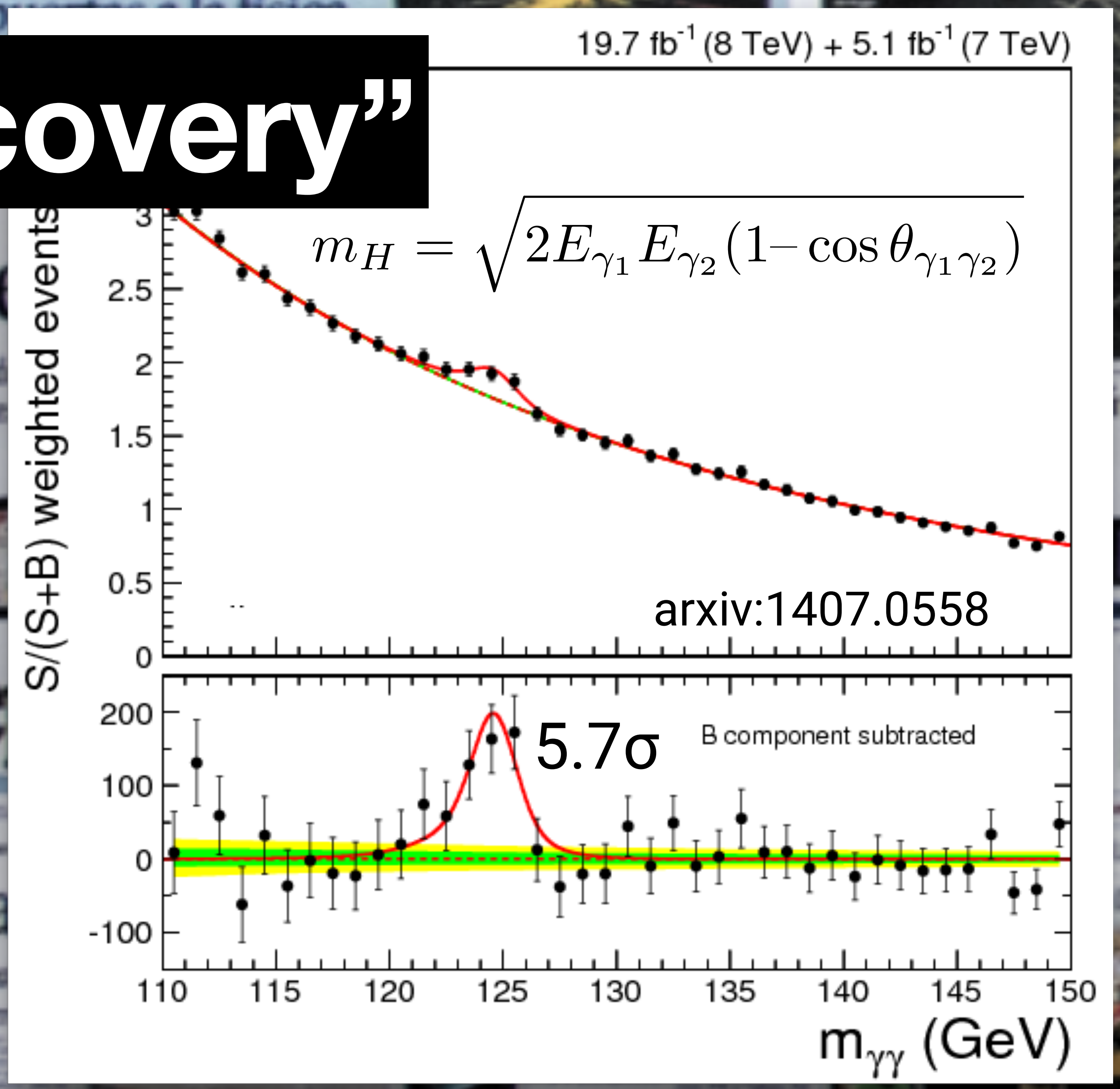
T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



T1049 / 6y4f
93.3 GDT
(adhesin tip)

- Experimental result
- Computational prediction

“ML for accelerated discovery”



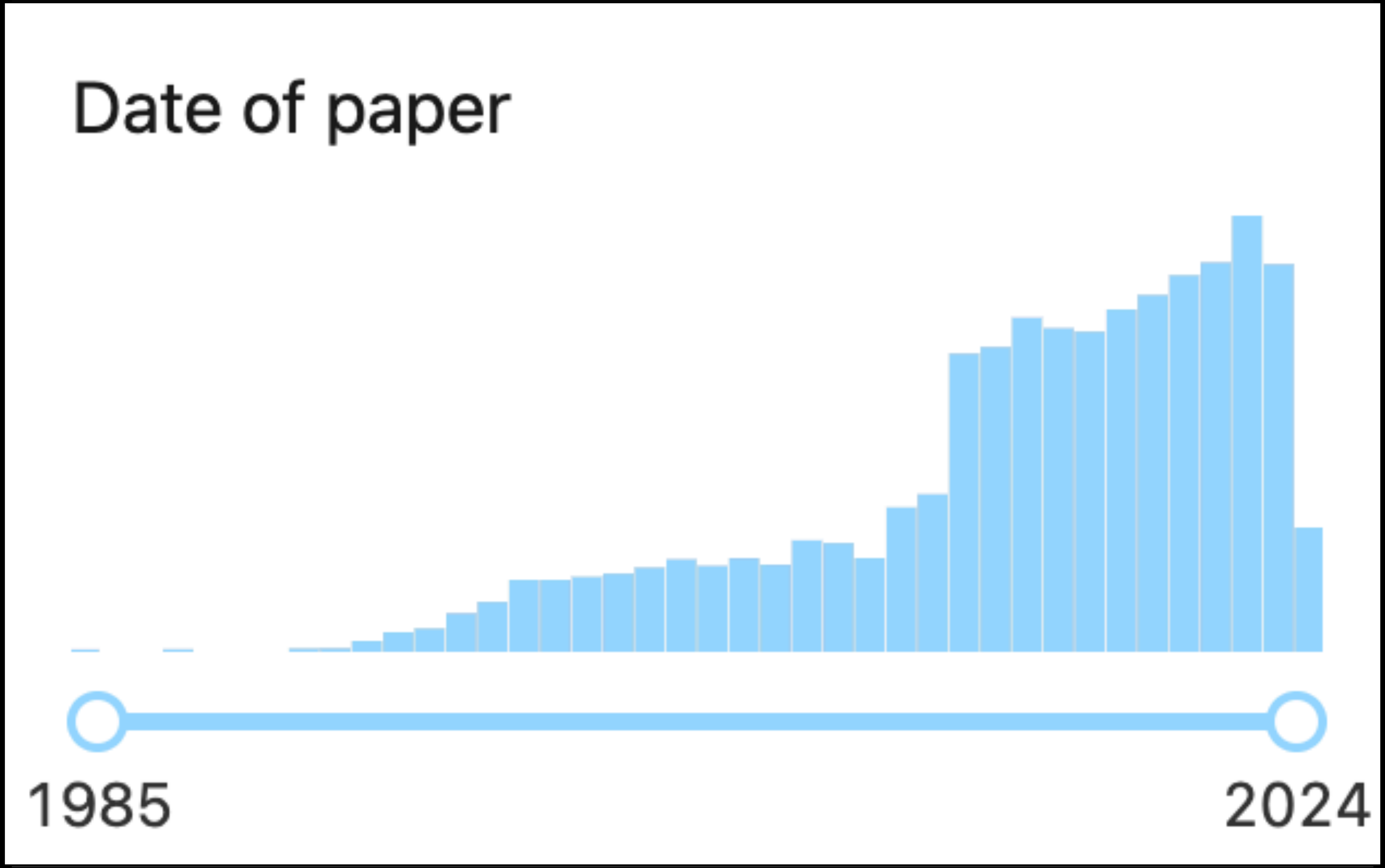

CERN Summer student 2012

Nature Review

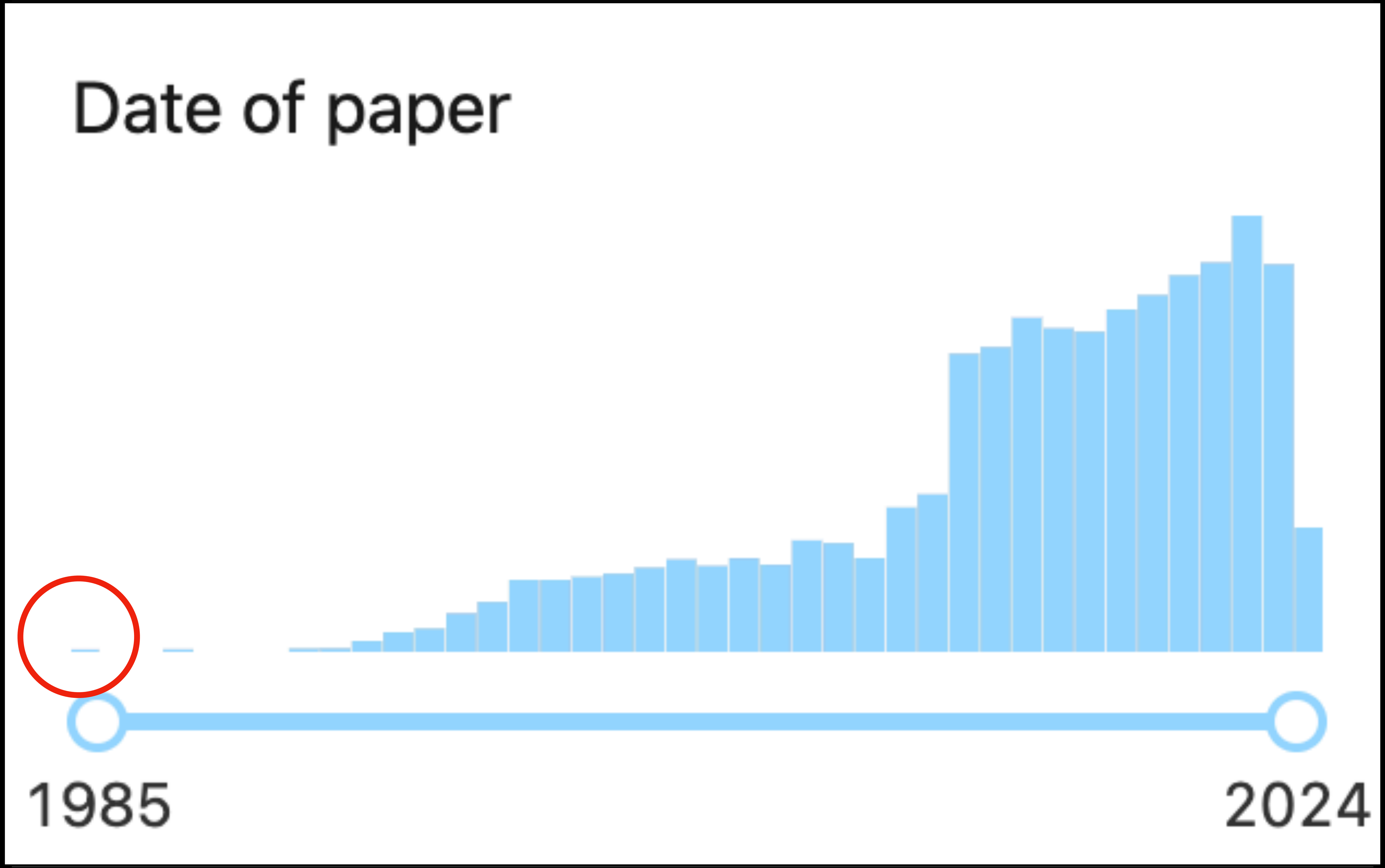
Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \to \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \to \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \to bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \to bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \to bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

We were using ML for discovery very early on

40 000 papers so far!



40 000 papers so far!



MPI-PAE/PTh 64/84
October 1984

An Evolutionary Procedure for Machine Learning

Leonard D. Mlodinow*

and

Ion O. Stamatescu**

Max-Planck-Institut für Physik und Astrophysik
- Werner-Heisenberg-Institut für Physik -
8000 Munich 40, West Germany

Abstract:

We discuss an evolutionary procedure for machine learning and present in detail an application of this procedure to the control of a robot TURTLE, which, beginning from a state of total ignorance, is able to develop the ability to circumnavigate a variety of obstacles. The procedure discussed is related to the strategy signature table method used in computer game playing.

NEURAL NETWORKS AND CELLULAR AUTOMATA IN EXPERIMENTAL HIGH ENERGY PHYSICS

B. DENBY

Laboratoire de l'Accélérateur Linéaire, Orsay, France

Received 20 September 1987; in revised form 28 December 1987

Within the past few years, two novel computing techniques, cellular automata and neural networks, have shown considerable promise in the solution of problems of a very high degree of complexity, such as turbulent fluid flow, image processing, and pattern recognition. Many of the problems faced in experimental high energy physics are also of this nature. Track reconstruction in wire chambers and cluster finding in cellular calorimeters, for instance, involve pattern recognition and high combinatorial complexity since many combinations of hits or cells must be considered in order to arrive at the final tracks or clusters. Here we examine in what way connective network methods can be applied to some of the problems of experimental high energy physics. It is found that such problems as track and cluster finding adapt naturally to these approaches. When large scale hard-wired connective networks become available, it will be possible to realize solutions to such problems in a fraction of the time required by traditional methods. For certain types of problems, faster solutions are already possible using model networks implemented on vector or other massively parallel machines. It should also be possible, using existing technology, to build simplified networks that will allow detailed reconstructed event information to be used in fast trigger decisions.

NEURAL NETWORKS AND CELLULAR AUTOMATA IN EXPERIMENTAL HIGH ENERGY PHYSICS

B. DENBY

Laboratoire de l'Accélérateur Linéaire, Orsay, France

Received 20 September 1987; in revised form 28 December 1987

Within the past few years, two novel computing technique
considerable promise in the solution of problems of a very high
processing, and pattern recognition. Many of the problems faced i
Track reconstruction in wire chambers and cluster finding in cell
and high combinatorial complexity since many combinations of hi
tracks or clusters. Here we examine in what way connective netv
experimental high energy physics. It is found that such probler
approaches. When large scale hard-wired connective networks becc
problems in a fraction of the time required by traditional methods.
possible using model networks implemented on vector or other ma
existing technology, to build simplified networks that will allow d
trigger decisions.

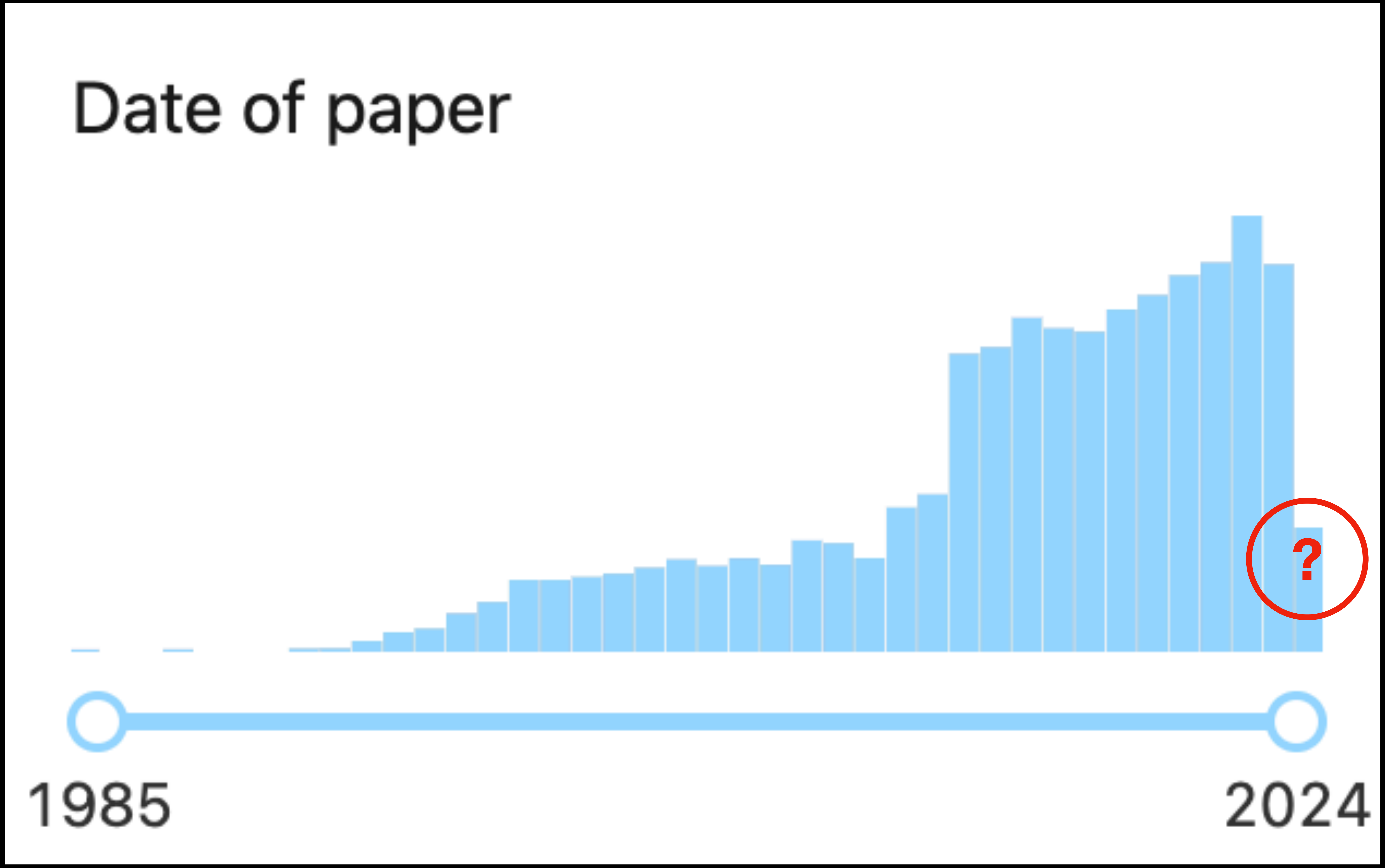
Institut Langevin
ONDES ET IMAGES

SORBONNE UNIVERSITÉ

What's in the article?

- Introduces **neural networks** to the HEP community for the first time
  - **Simple units** sum their inputs & apply an **activation function**
  - Outputs connect to other inputs via **weights**, and
  - **Perform a useful task** by mapping from inputs to outputs
- Proposes a recurrent neural network **algorithm** for **track finding** (*Denby-Peterson algorithm*)
- Highlights the **parallel nature** of calculating with neural networks and its interest for experimental **triggers**
- Also discusses **feed-forward neural networks** for **template matching**, and the possibility of using **learning**

40 000 papers so far!

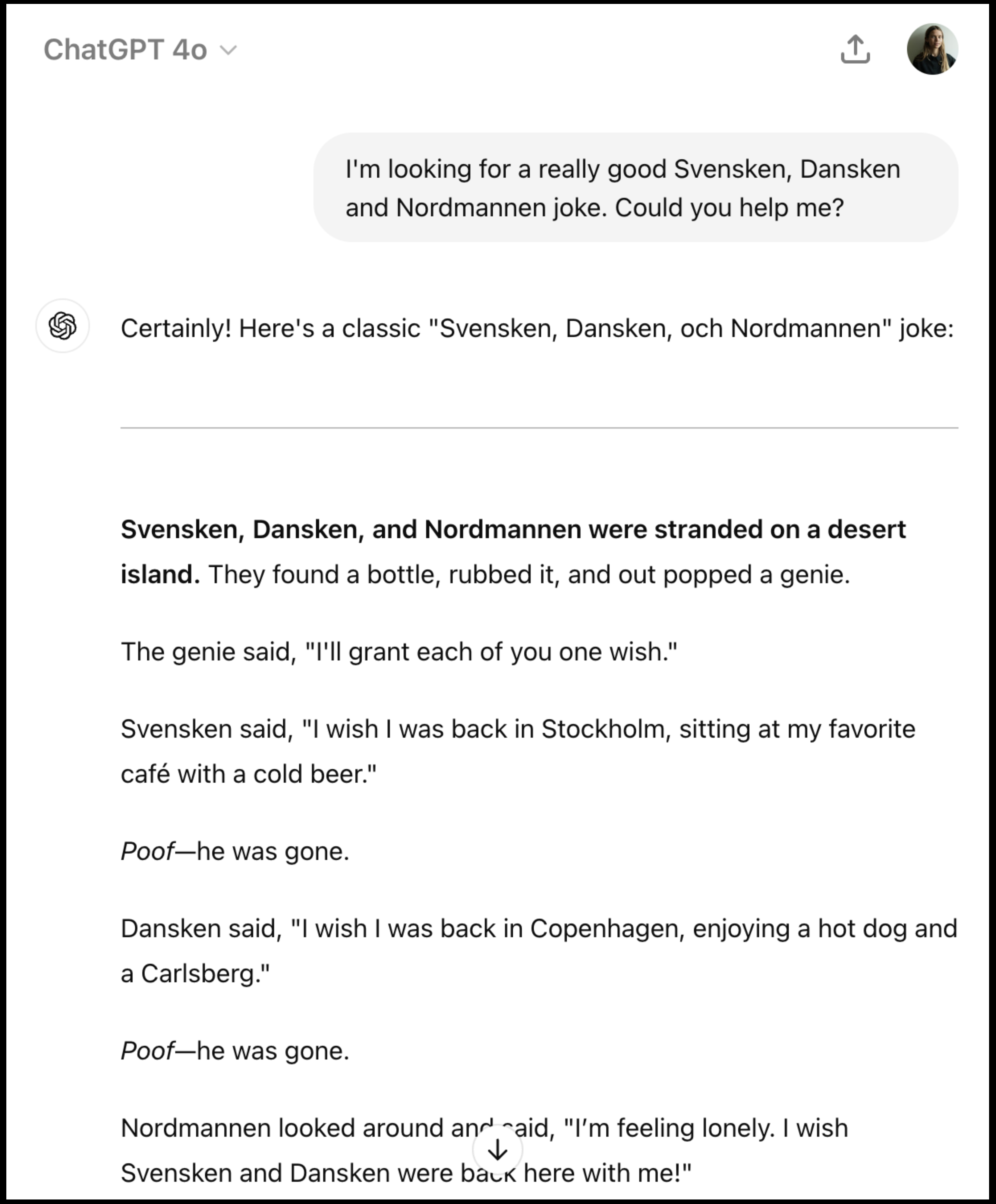


CERN 70:
A Laboratory for AI Research

?

Train (GPT-4):

- 2.15^{25} floating point operations
- ~25,000 A100 GPUs
- 90-100 days
- $63 million
- Trained on 13 trillion tokens



Train (GPT-4):

- 2.15^{25} floating point operations
- ~25,000 A100 GPUs
- 90-100 days
- $63 million
- Trained on 13 trillion tokens

Inference (GPT-4):

- Multiple clusters of 128 GPUs
- Model carefully mapped onto hardware

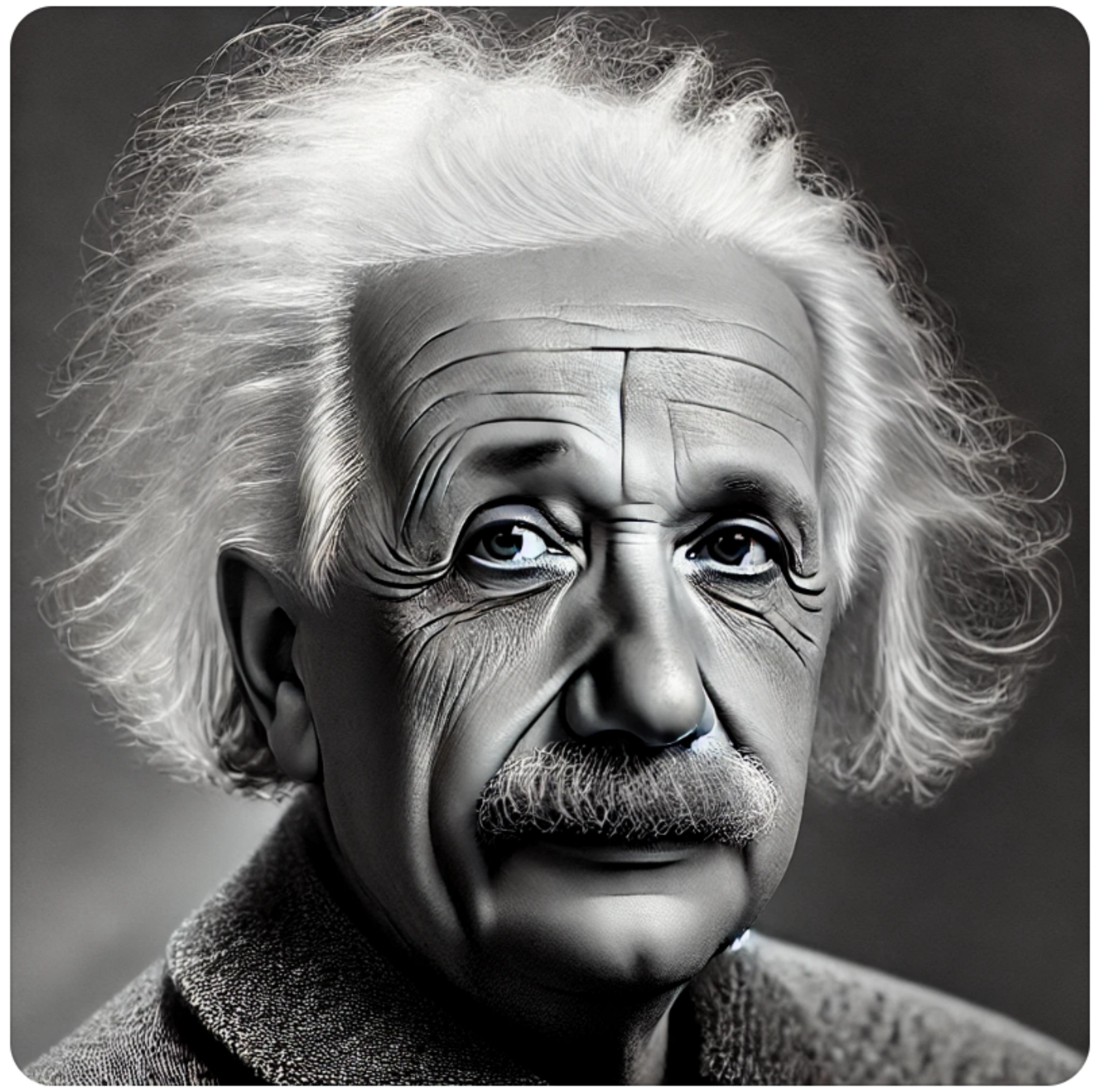
A personal take

- we have to be humble

- Cannot compete with industry
  - Loosing many of our best people to private companies (salaries, interesting problems)
- Where can we contribute?
  Where are we special?

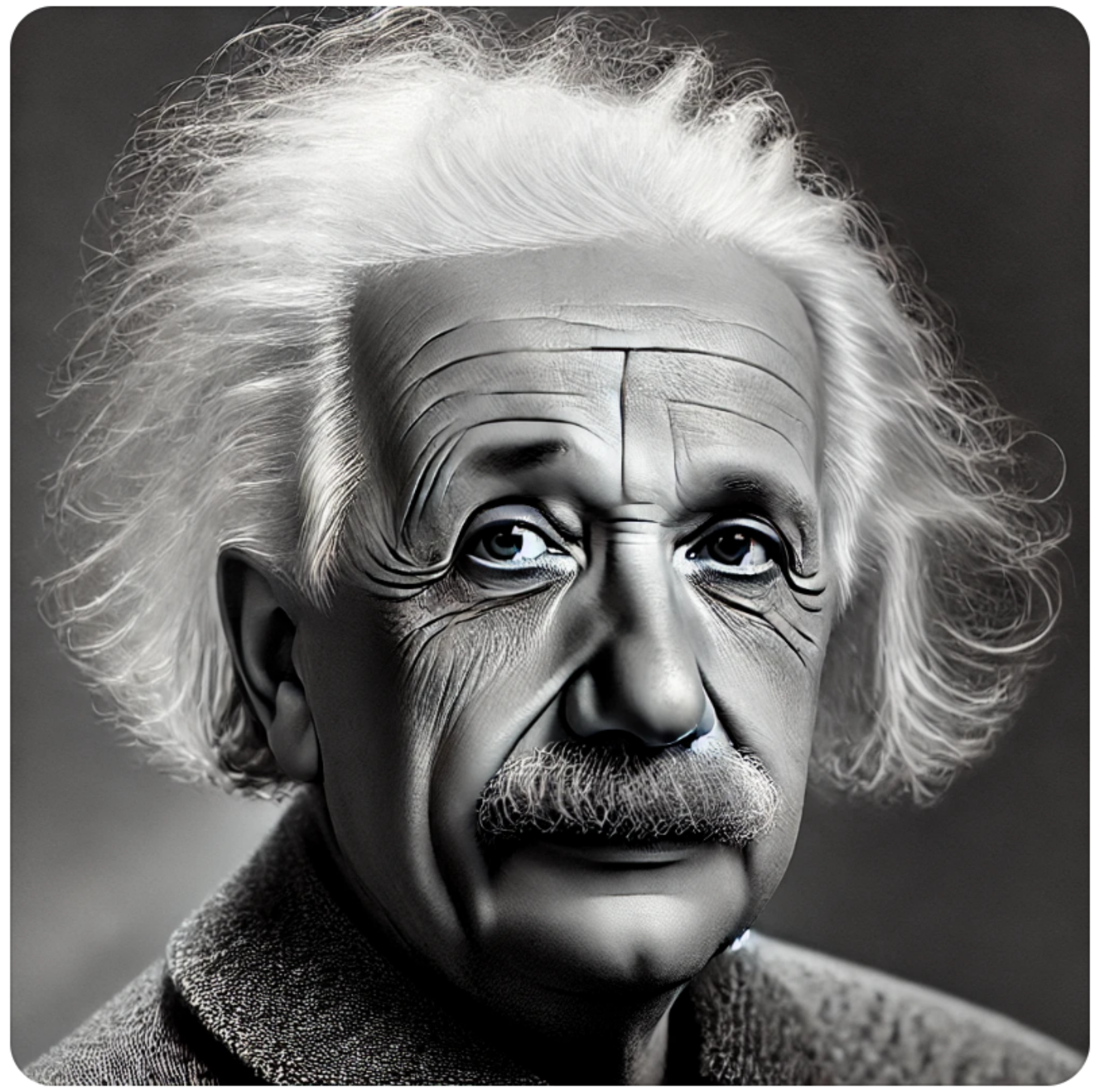
- Where can we contribute?
  Where are we special?

  - High fidelity

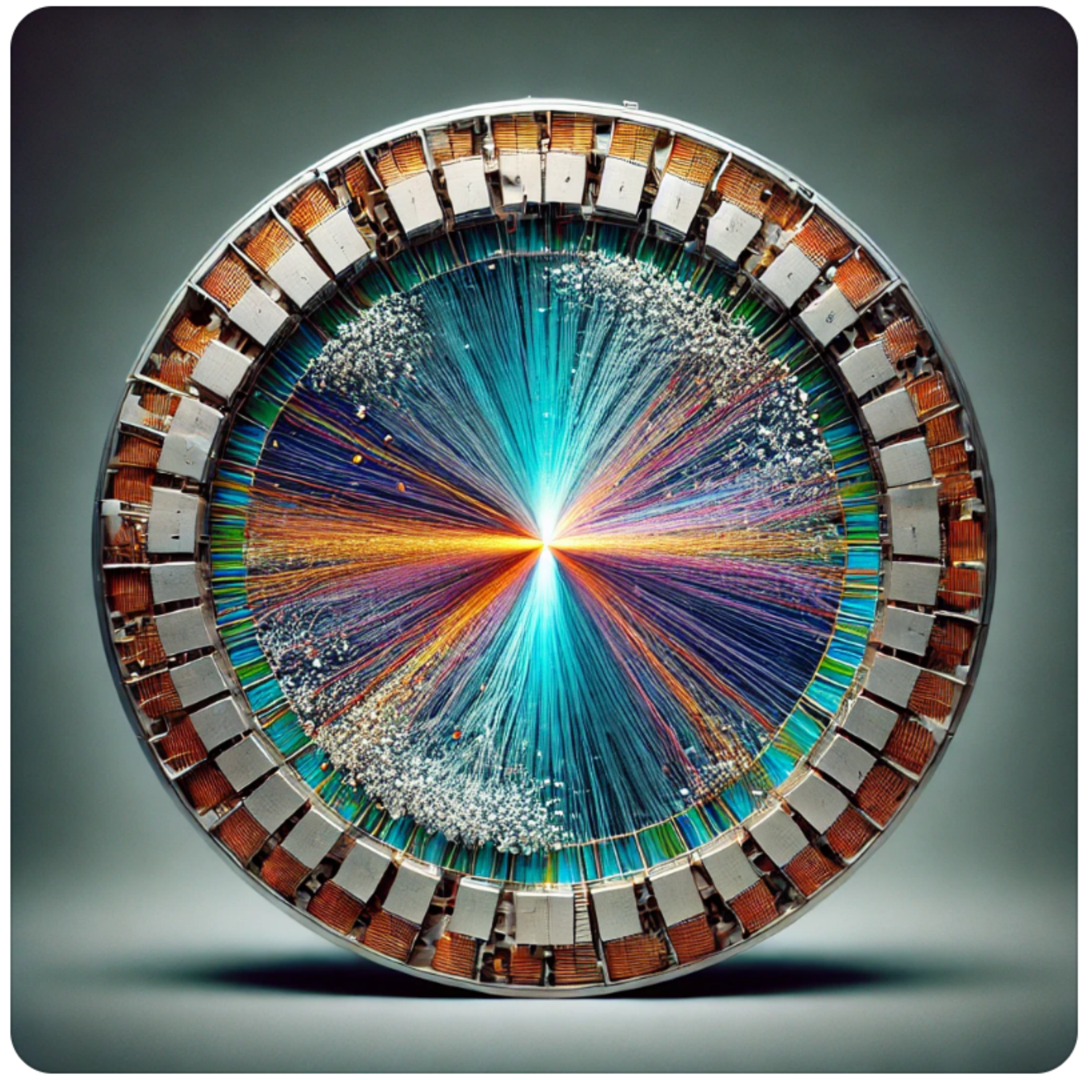


Here is the subtly flawed image of Albert Einstein, where a minor detail, such as his hair being parted on the wrong side or a slight facial proportion, creates a small but noticeable inaccuracy.

- Where can we contribute?
  Where are we special?
  - High fidelity



Here is the subtly flawed image of Albert Einstein, where a minor detail, such as his hair being parted on the wrong side or a slight facial proportion, creates a small but noticeable inaccuracy.

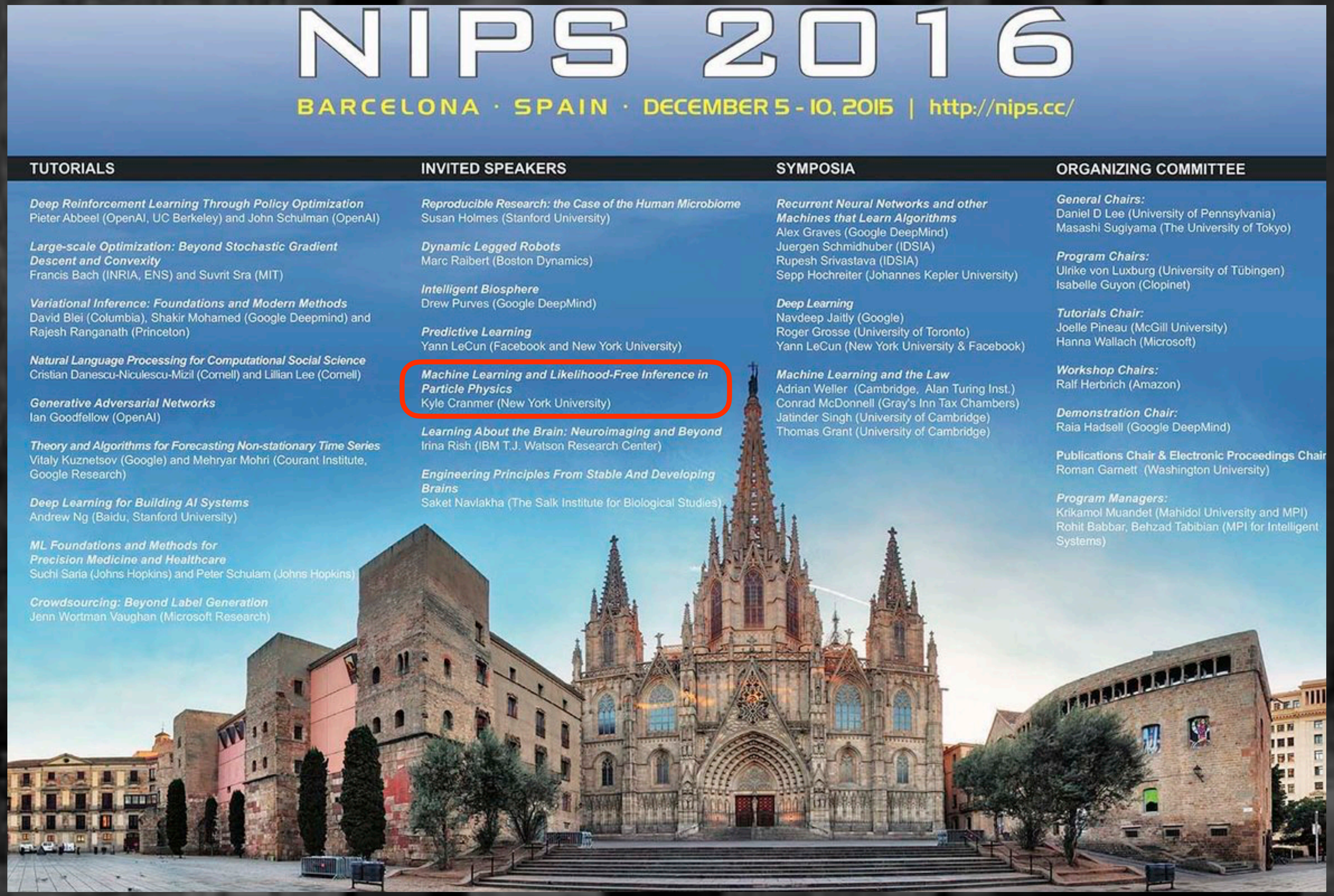


Here is the subtly flawed event display of a proton-proton collision in the CMS experiment. The image contains realistic particle tracks and detector layers, but with a small inaccuracy, such as incorrect track curvature or missing elements, noticeable to those familiar with high-energy physics experiments.

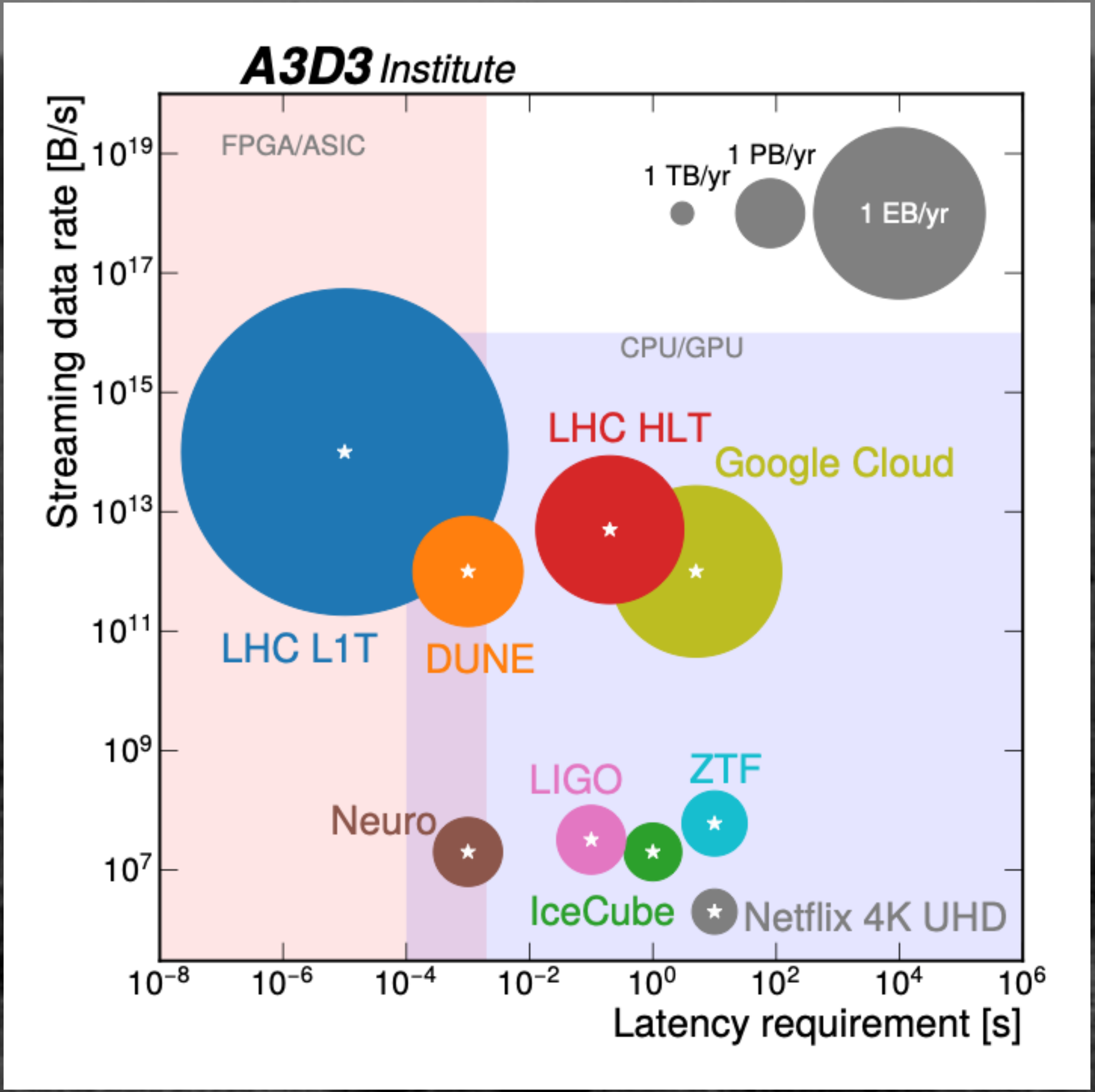
- Where can we contribute?
  Where are we special?
  - High fidelity
  - Uncertainty treatment

Approximating Likelihood Ratios with Calibrated Discriminative Classifiers

Kyle Cranmer[1], Juan Pavez[2], and Gilles Louppe[1]
[1]New York University
[2]Federico Santa María University



- Where can we contribute? Where are we special?
  - High fidelity
  - Uncertainty treatment
  - Extreme high throughput, low-latency constraints

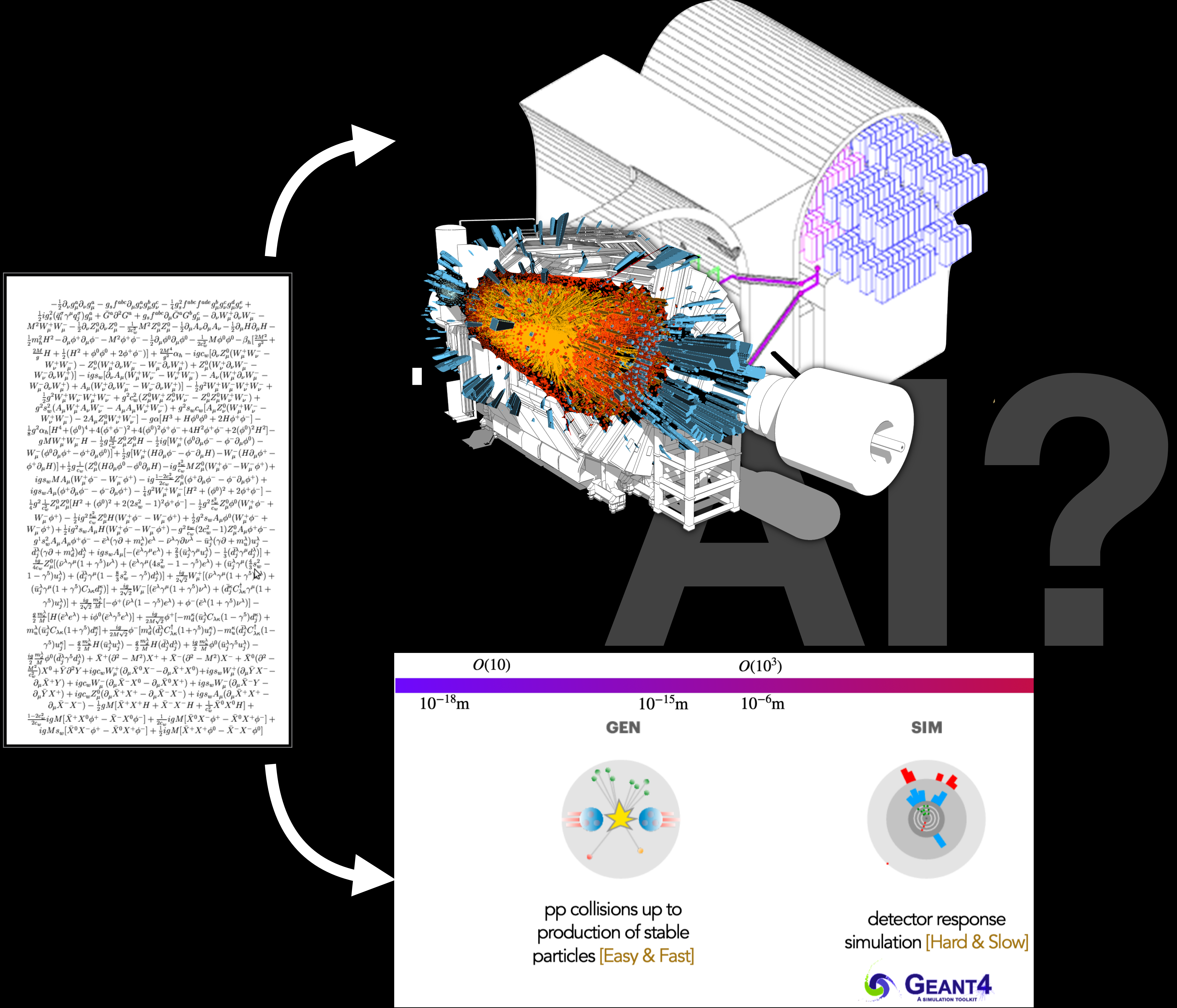


- Where can we contribute?
  Where are we special?
  - High fidelity
  - Uncertainty treatment
  - Extreme high throughput, low-latency constraints
  - And extremely complex and exciting problems!

AI?

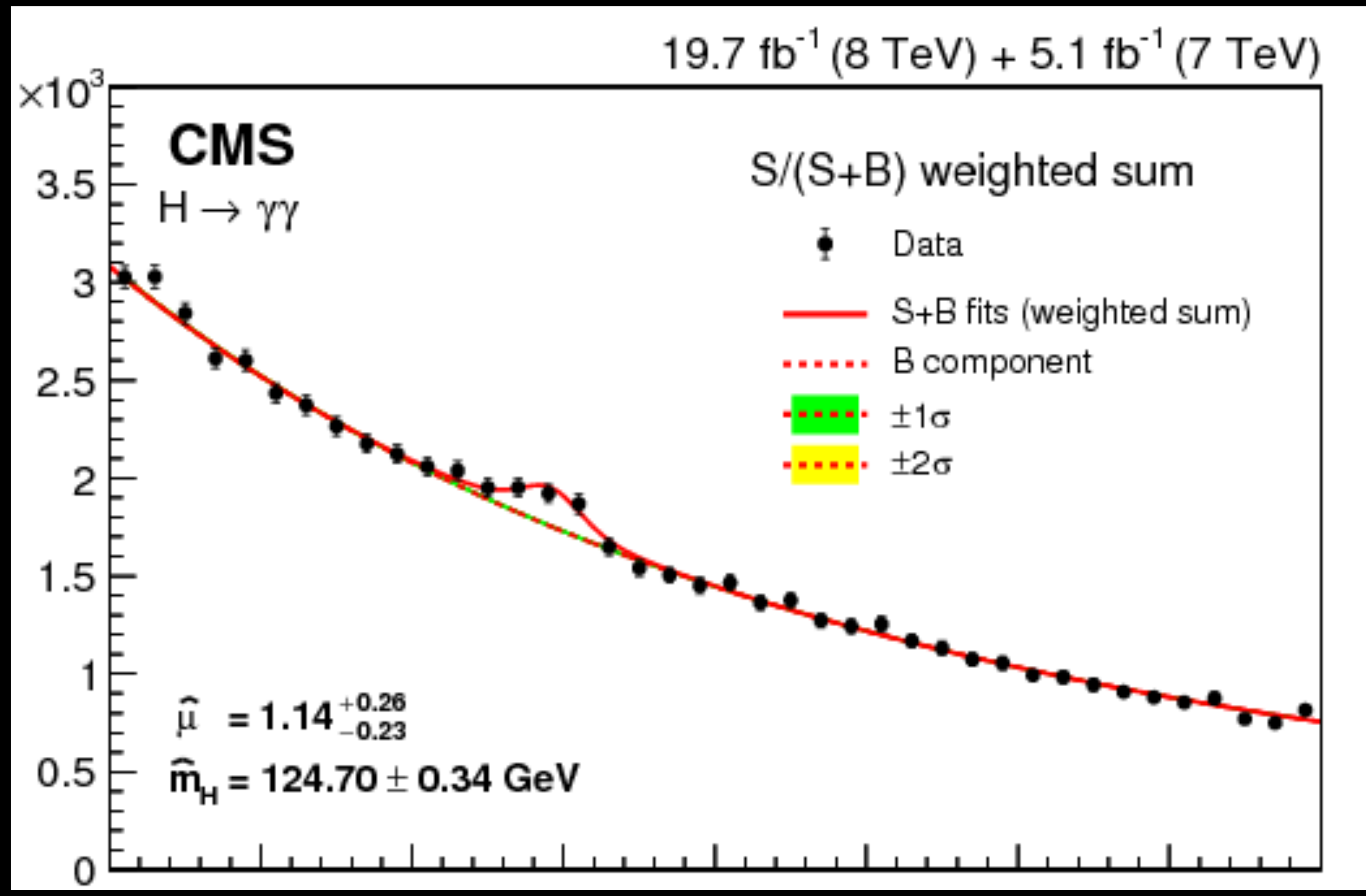
$$
\begin{gathered}
-\frac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \frac{1}{4}g_s^2 f^{abc} f^{ade} g^b_\mu g^c_\nu g^d_\mu g^e_\nu + \\
\frac{1}{2}ig_s^2(\bar{q}^\sigma_i \gamma^\mu q^\sigma_j)g^a_\mu + \bar{G}^a\partial^2 G^a + g_s f^{abc}\partial_\mu \bar{G}^a G^b g^c_\mu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu - \\
M^2 W^+_\mu W^-_\mu - \frac{1}{2}\partial_\nu Z^0_\mu \partial_\nu Z^0_\mu - \frac{1}{2c_w^2}M^2 Z^0_\mu Z^0_\mu - \frac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - \frac{1}{2}\partial_\mu H \partial_\mu H - \\
\frac{1}{2}m_h^2 H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - M^2 \phi^+\phi^- - \frac{1}{2}\partial_\mu \phi^0 \partial_\mu \phi^0 - \frac{1}{2c_w^2}M\phi^0\phi^0 - \beta_h[\frac{2M^2}{g^2} + \\
\frac{2M}{g}H + \frac{1}{2}(H^2 + \phi^0\phi^0 + 2\phi^+\phi^-)] + \frac{2M^4}{g^2}\alpha_h - igc_w[\partial_\nu Z^0_\mu(W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - Z^0_\nu(W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + Z^0_\mu(W^+_\nu \partial_\nu W^-_\mu - \\
W^-_\nu \partial_\nu W^+_\mu)] - igs_w[\partial_\nu A_\mu(W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - A_\nu(W^+_\mu \partial_\nu W^-_\mu - \\
W^-_\mu \partial_\nu W^+_\mu) + A_\mu(W^+_\nu \partial_\nu W^-_\mu - W^-_\nu \partial_\nu W^+_\mu)] - \frac{1}{2}g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \\
\frac{1}{2}g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2 c_w^2(Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu - Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + \\
g^2 s_w^2(A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + g^2 s_w c_w[A_\mu Z^0_\nu(W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - 2A_\mu Z^0_\mu W^+_\nu W^-_\nu] - g\alpha[H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-] - \\
\frac{1}{8}g^2\alpha_h[H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2 H^2] - \\
gMW^+_\mu W^-_\mu H - \frac{1}{2}g\frac{M}{c_w^2}Z^0_\mu Z^0_\mu H - \frac{1}{2}ig[W^+_\mu(\phi^0\partial_\mu\phi^- - \phi^-\partial_\mu\phi^0) - \\
W^-_\mu(\phi^0\partial_\mu\phi^+ - \phi^+\partial_\mu\phi^0)] + \frac{1}{2}g[W^+_\mu(H\partial_\mu\phi^- - \phi^-\partial_\mu H) - W^-_\mu(H\partial_\mu\phi^+ - \\
\phi^+\partial_\mu H)] + \frac{1}{2}g\frac{1}{c_w}(Z^0_\mu(H\partial_\mu\phi^0 - \phi^0\partial_\mu H) - ig\frac{s_w^2}{c_w}MZ^0_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) + \\
igs_w MA_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) - ig\frac{1-2c_w^2}{2c_w}Z^0_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) + \\
igs_w A_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) - \frac{1}{4}g^2 W^+_\mu W^-_\mu[H^2 + (\phi^0)^2 + 2\phi^+\phi^-] - \\
\frac{1}{4}g^2\frac{1}{c_w^2}Z^0_\mu Z^0_\mu[H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2\phi^+\phi^-] - \frac{1}{2}g^2\frac{s_w^2}{c_w}Z^0_\mu\phi^0(W^+_\mu\phi^- + \\
W^-_\mu\phi^+) - \frac{1}{2}ig^2\frac{s_w^2}{c_w}Z^0_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) + \frac{1}{2}g^2 s_w A_\mu\phi^0(W^+_\mu\phi^- + \\
W^-_\mu\phi^+) + \frac{1}{2}ig^2 s_w A_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) - g^2\frac{s_w}{c_w}(2c_w^2 - 1)Z^0_\mu A_\mu\phi^+\phi^- - \\
g^1 s_w^2 A_\mu A_\mu\phi^+\phi^- - \bar{e}^\lambda(\gamma\partial + m_e^\lambda)e^\lambda - \bar{\nu}^\lambda\gamma\partial\nu^\lambda - \bar{u}^\lambda_j(\gamma\partial + m_u^\lambda)u^\lambda_j - \\
\bar{d}^\lambda_j(\gamma\partial + m_d^\lambda)d^\lambda_j + igs_w A_\mu[-(\bar{e}^\lambda\gamma^\mu e^\lambda) + \frac{2}{3}(\bar{u}^\lambda_j\gamma^\mu u^\lambda_j) - \frac{1}{3}(\bar{d}^\lambda_j\gamma^\mu d^\lambda_j)] + \\
\frac{ig}{4c_w}Z^0_\mu[(\bar{\nu}^\lambda\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar{e}^\lambda\gamma^\mu(4s_w^2 - 1 - \gamma^5)e^\lambda) + (\bar{u}^\lambda_j\gamma^\mu(\frac{4}{3}s_w^2 - \\
1 - \gamma^5)u^\lambda_j) + (\bar{d}^\lambda_j\gamma^\mu(1 - \frac{8}{3}s_w^2 - \gamma^5)d^\lambda_j)] + \frac{ig}{2\sqrt{2}}W^+_\mu[(\bar{\nu}^\lambda\gamma^\mu(1+\gamma^5)e^\lambda) + \\
(\bar{u}^\lambda_j\gamma^\mu(1+\gamma^5)C_{\lambda\kappa}d^\kappa_j)] + \frac{ig}{2\sqrt{2}}W^-_\mu[(\bar{e}^\lambda\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar{d}^\kappa_j C^\dagger_{\lambda\kappa}\gamma^\mu(1 + \\
\gamma^5)u^\lambda_j)] + \frac{ig}{2\sqrt{2}}\frac{m_e^\lambda}{M}[-\phi^+(\bar{\nu}^\lambda(1-\gamma^5)e^\lambda) + \phi^-(\bar{e}^\lambda(1+\gamma^5)\nu^\lambda)] - \\
\frac{g}{2}\frac{m_e^\lambda}{M}[H(\bar{e}^\lambda e^\lambda) + i\phi^0(\bar{e}^\lambda\gamma^5 e^\lambda)] + \frac{ig}{2M\sqrt{2}}\phi^+[-m_d^\kappa(\bar{u}^\lambda_j C_{\lambda\kappa}(1-\gamma^5)d^\kappa_j) + \\
m_u^\lambda(\bar{u}^\lambda_j C_{\lambda\kappa}(1+\gamma^5)d^\kappa_j] + \frac{ig}{2M\sqrt{2}}\phi^-[m_d^\lambda(\bar{d}^\lambda_j C^\dagger_{\lambda\kappa}(1+\gamma^5)u^\kappa_j) - m_u^\kappa(\bar{d}^\lambda_j C^\dagger_{\lambda\kappa}(1 - \\
\gamma^5)u^\kappa_j] - \frac{g}{2}\frac{m_u^\lambda}{M}H(\bar{u}^\lambda_j u^\lambda_j) - \frac{g}{2}\frac{m_d^\lambda}{M}H(\bar{d}^\lambda_j d^\lambda_j) + \frac{ig}{2}\frac{m_u^\lambda}{M}\phi^0(\bar{u}^\lambda_j\gamma^5 u^\lambda_j) - \\
\frac{ig}{2}\frac{m_d^\lambda}{M}\phi^0(\bar{d}^\lambda_j\gamma^5 d^\lambda_j) + \bar{X}^+(\partial^2 - M^2)X^+ + \bar{X}^-(\partial^2 - M^2)X^- + \bar{X}^0(\partial^2 - \\
\frac{M^2}{c_w^2})X^0 + \bar{Y}\partial^2 Y + igc_w W^+_\mu(\partial_\mu\bar{X}^0 X^- - \partial_\mu\bar{X}^+ X^0) + igs_w W^+_\mu(\partial_\mu\bar{Y}X^- - \\
\partial_\mu\bar{X}^+ Y) + igc_w W^-_\mu(\partial_\mu\bar{X}^- X^0 - \partial_\mu\bar{X}^0 X^+) + igs_w W^-_\mu(\partial_\mu\bar{X}^- Y - \\
\partial_\mu\bar{Y}X^+) + igc_w Z^0_\mu(\partial_\mu\bar{X}^+ X^+ - \partial_\mu\bar{X}^- X^-) + igs_w A_\mu(\partial_\mu\bar{X}^+ X^+ - \\
\partial_\mu\bar{X}^- X^-) - \frac{1}{2}gM[\bar{X}^+ X^+ H + \bar{X}^- X^- H + \frac{1}{c_w^2}\bar{X}^0 X^0 H] + \\
\frac{1-2c_w^2}{2c_w}igM[\bar{X}^+ X^0\phi^+ - \bar{X}^- X^0\phi^-] + \frac{1}{2c_w}igM[\bar{X}^0 X^-\phi^+ - \bar{X}^0 X^+\phi^-] + \\
igMs_w[\bar{X}^0 X^-\phi^+ - \bar{X}^0 X^+\phi^-] + \frac{1}{2}igM[\bar{X}^+ X^+\phi^0 - \bar{X}^- X^-\phi^0]
\end{gathered}
$$

AI?

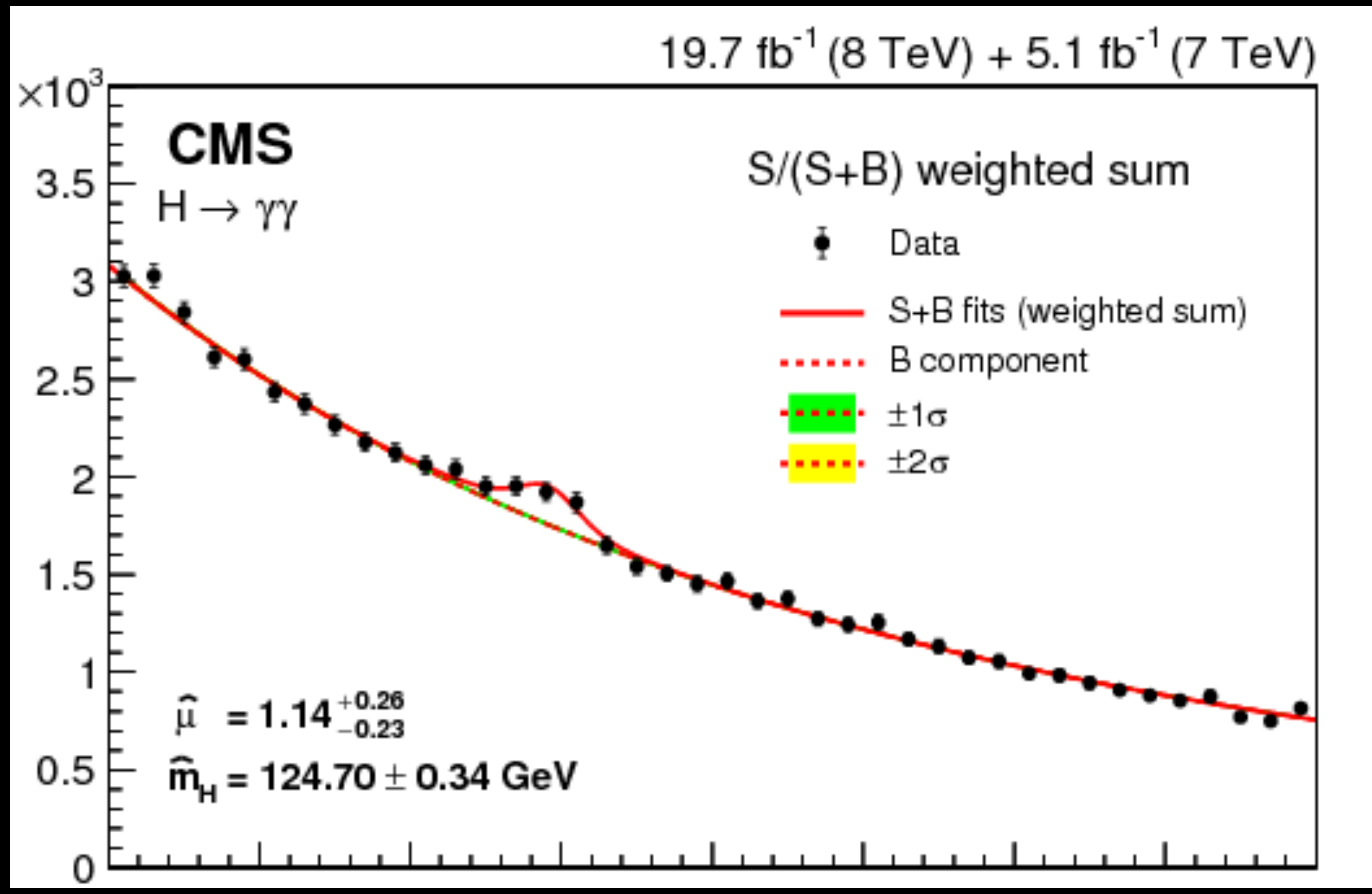

AI?
O(10)
O(10³)
10⁻¹⁸m
10⁻¹⁵m
10⁻⁶m
GEN
SIM
pp collisions up to production of stable particles [Easy & Fast]
detector response simulation [Hard & Slow]
GEANT4
A SIMULATION TOOLKIT

AI?
O(10)
O(10^3)
10^−18m
10^−15m
10^−6m
GEN
pp collisions up to production of stable particles [Easy & Fast]
SIM
detector response simulation [Hard & Slow]
GEANT4
A SIMULATION TOOLKIT
O(10^10)
100m
DIGI+RECO
Energy deposits→digital signals→reconstructed by the reconstruction software [Hard & Slow]

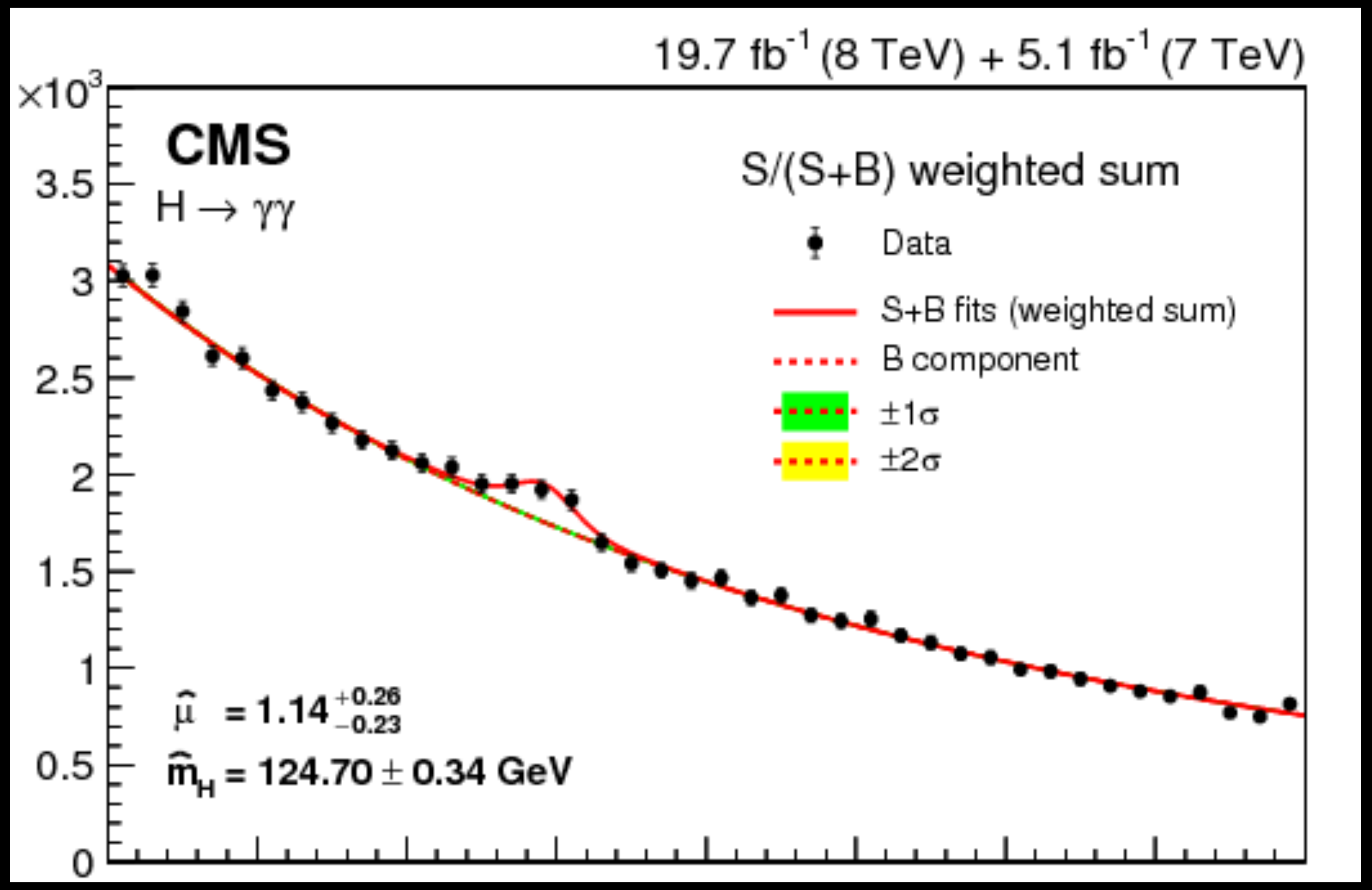
$O(10^{10})$
100m
DIGI+RECO
Energy deposits→digital signals→reconstructed by the reconstruction software [Hard & Slow]
19.7 fb^{-1} (8 TeV) + 5.1 fb^{-1} (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
±1σ
±2σ
$\times10^3$
$\hat{\mu} = 1.14^{+0.26}_{-0.23}$
$\hat{m}_H = 124.70 \pm 0.34$ GeV
$O(10)$
$O(10^3)$
10^{-18}m
10^{-15}m
10^{-6}m
GEN
pp collisions up to production of stable particles [Easy & Fast]
SIM
detector response simulation [Hard & Slow]
GEANT4



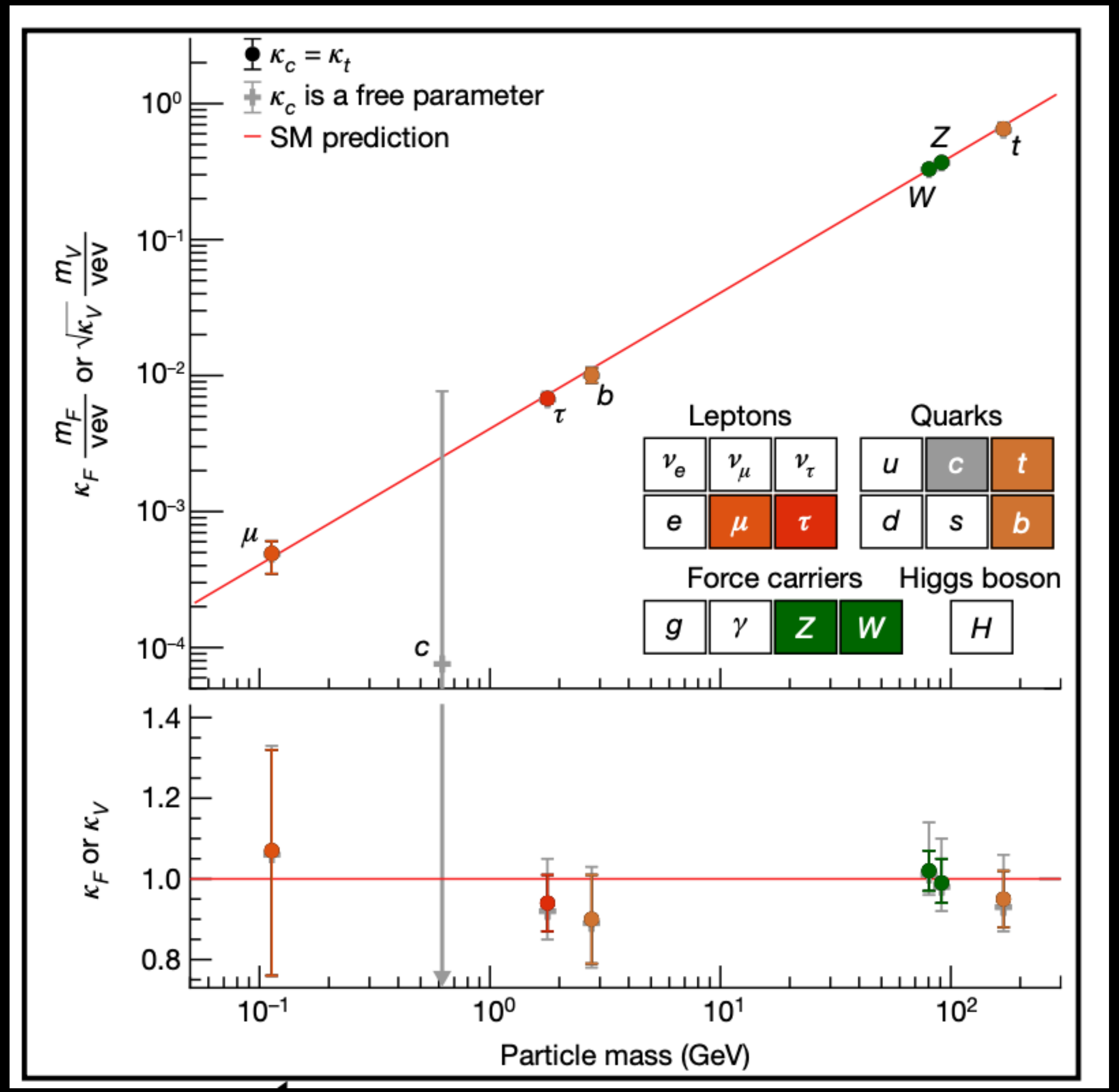
19.7 fb^{-1} (8 TeV) + 5.1 fb^{-1} (7 TeV)
$\times 10^3$
CMS
$H \to \gamma\gamma$
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
$\pm 1\sigma$
$\pm 2\sigma$
$\hat{\mu} = 1.14^{+0.26}_{-0.23}$
$\hat{m}_H = 124.70 \pm 0.34$ GeV
3.5
3
2.5
2
1.5
1
0.5
0



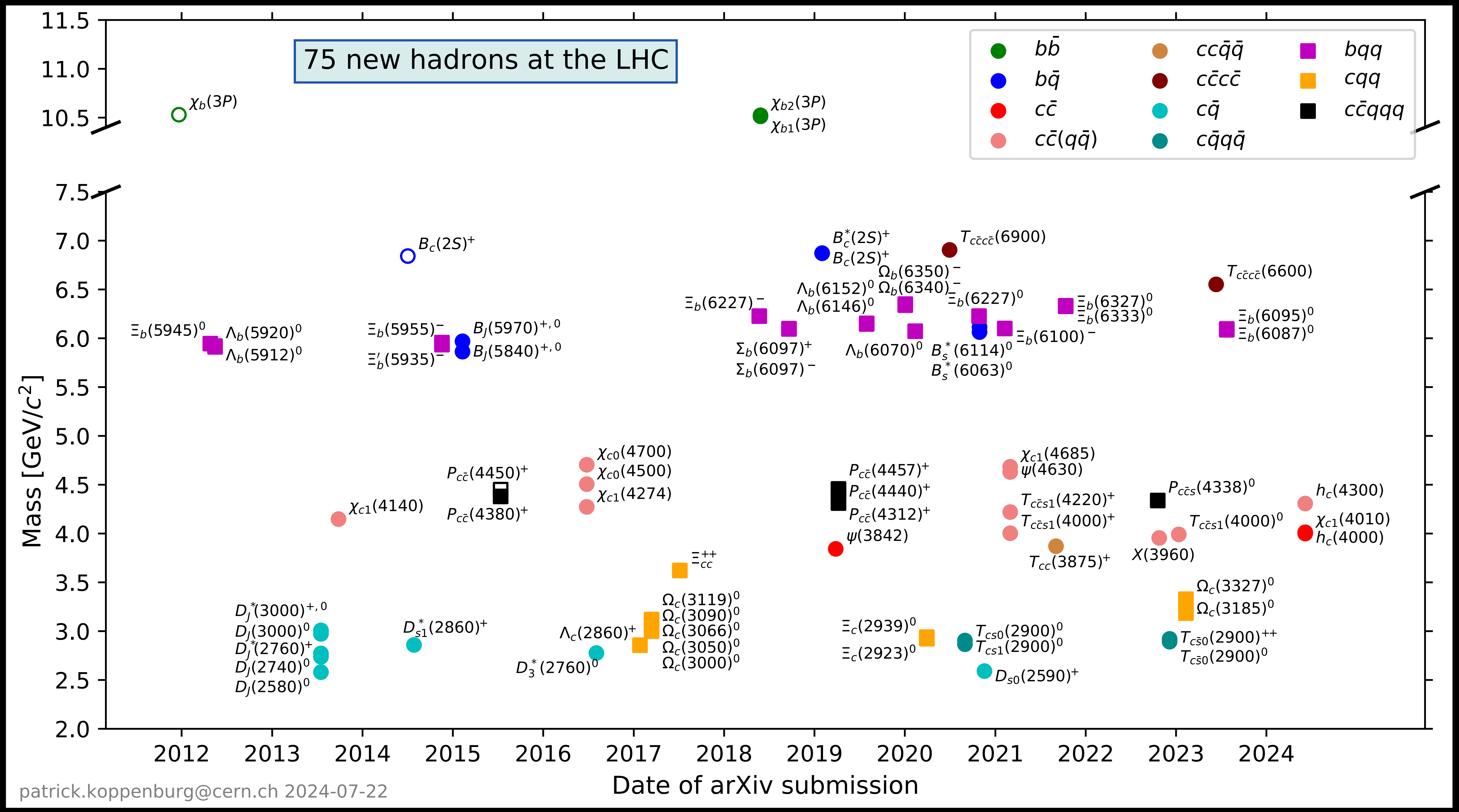
19.7 fb-1 (8 TeV) + 5.1 fb-1 (7 TeV)
×10³
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
±1σ
±2σ
μ̂ = 1.14 +0.26 −0.23
m̂H = 124.70 ± 0.34 GeV
3.5
3
2.5
2
1.5
1
0.5
0



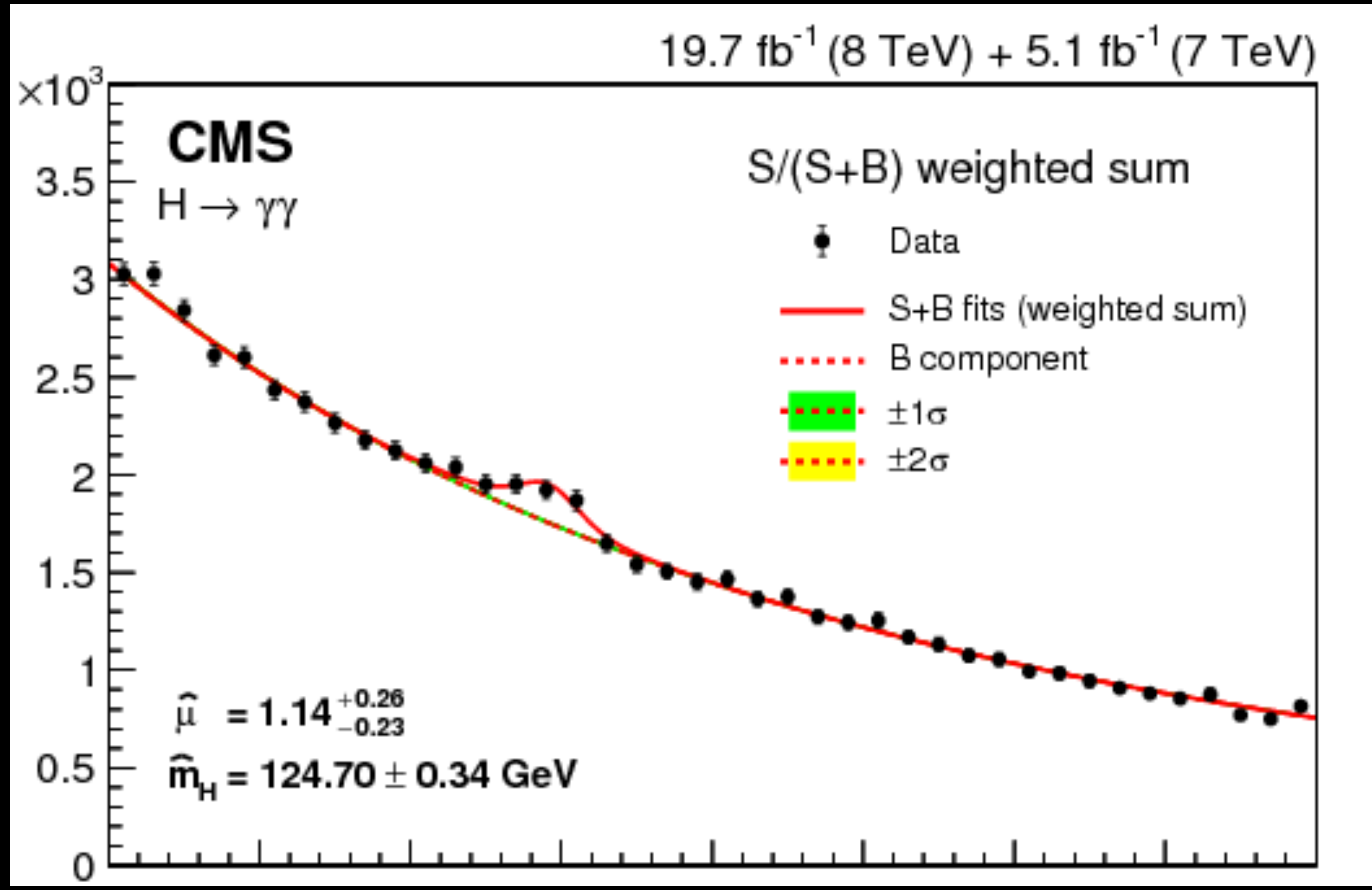
19.7 fb⁻¹ (8 TeV) + 5.1 fb⁻¹ (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
±1σ
±2σ
$\hat{\mu} = 1.14^{+0.26}_{-0.23}$
$\hat{m}_H = 124.70 \pm 0.34$ GeV

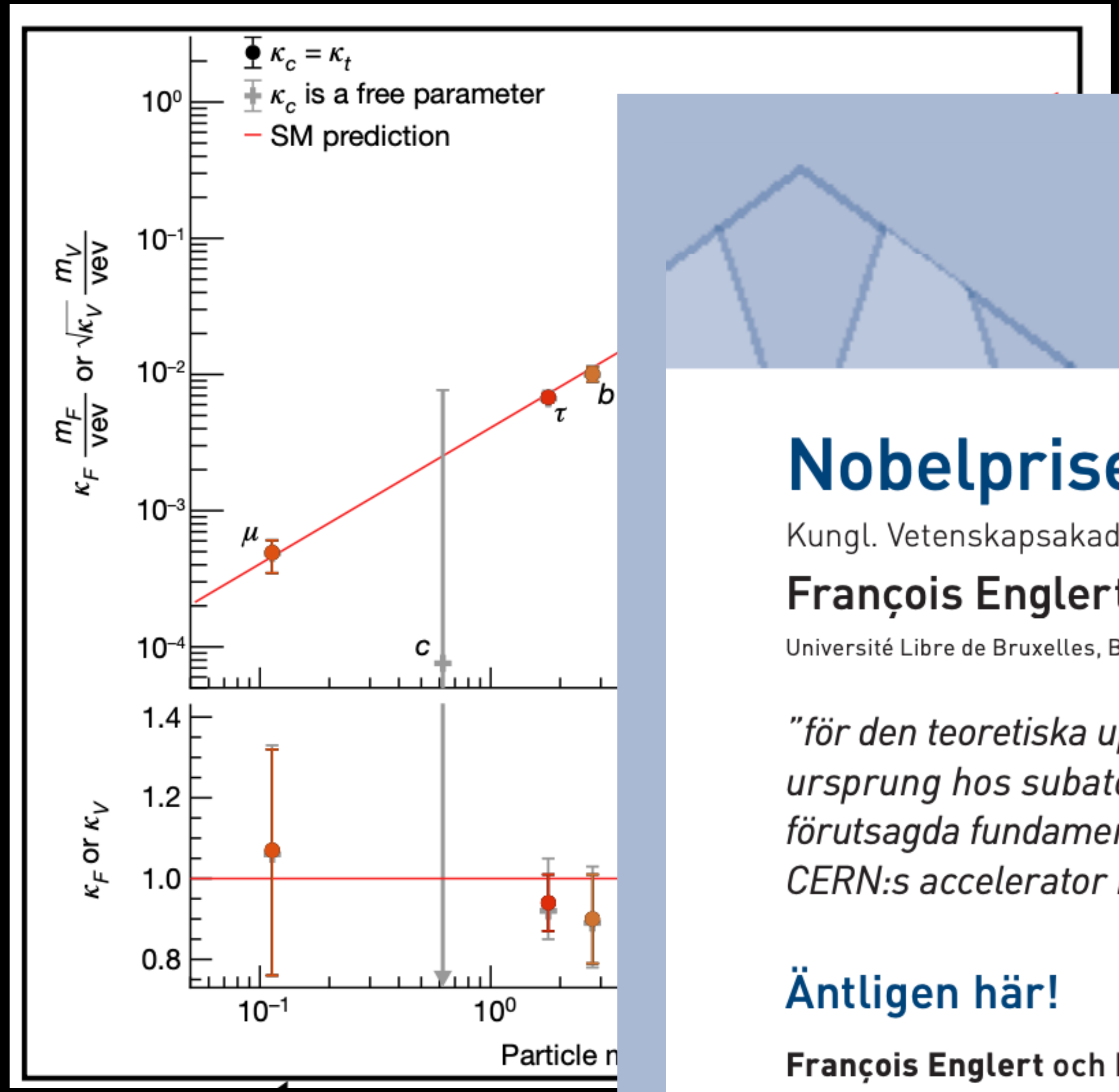


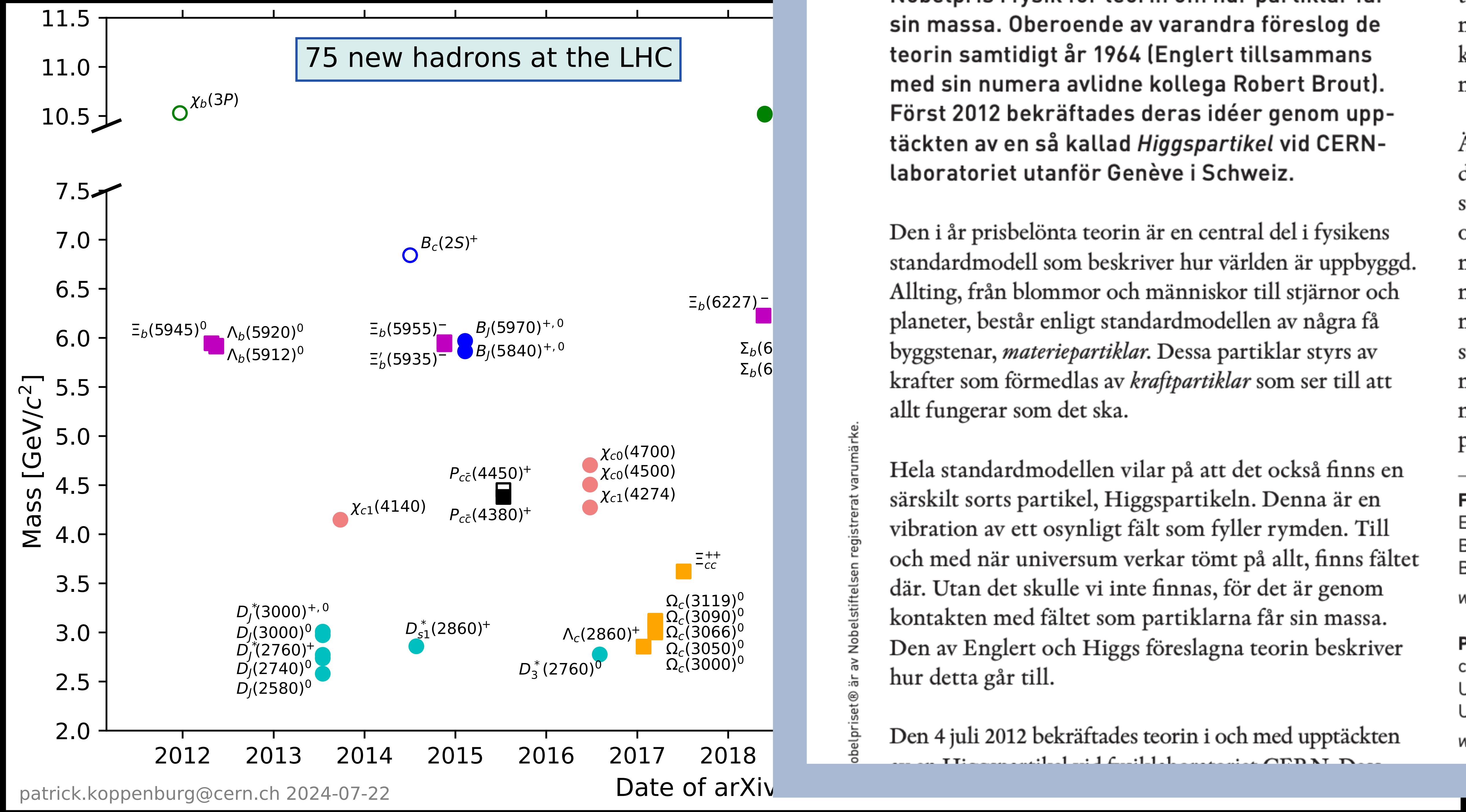
$\kappa_c = \kappa_t$
κ_c is a free parameter
SM prediction
$\kappa_F \frac{m_F}{vev}$ or $\sqrt{\kappa_V} \frac{m_V}{vev}$
κ_F or κ_V
Leptons
Quarks
Force carriers
Higgs boson
Particle mass (GeV)

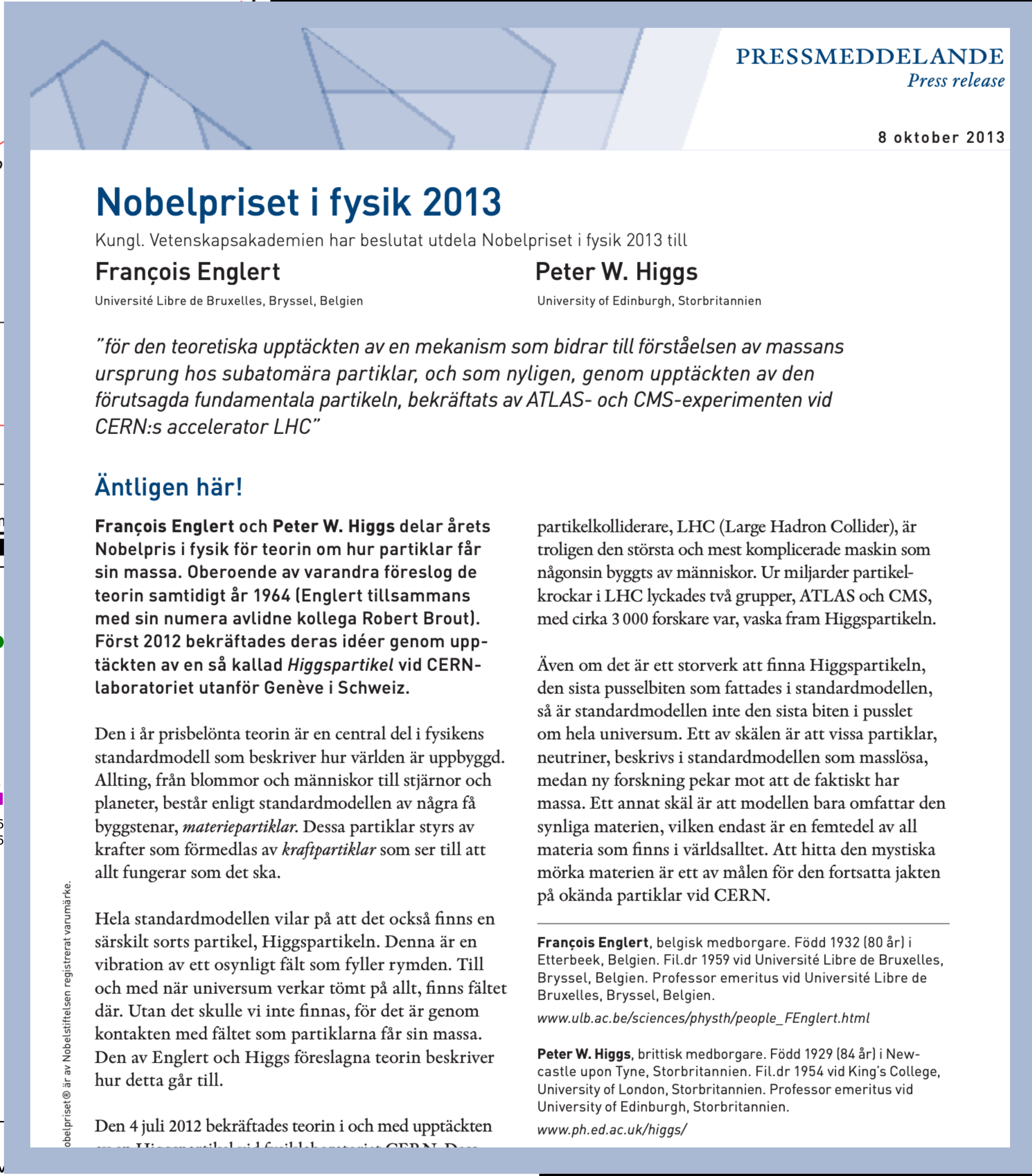


75 new hadrons at the LHC
$b\bar{b}$
$b\bar{q}$
$c\bar{c}$
$c\bar{c}(q\bar{q})$
$cc\bar{q}\bar{q}$
$c\bar{c}c\bar{c}$
$c\bar{q}$
$c\bar{q}q\bar{q}$
bqq
cqq
$c\bar{c}qqq$
Mass [GeV/c^2]
Date of arXiv submission
patrick.koppenburg@cern.ch 2024-07-22








PRESSMEDDELANDE
Press release

8 oktober 2013

Nobelpriset i fysik 2013

Kungl. Vetenskapsakademien har beslutat utdela Nobelpriset i fysik 2013 till

François Englert
Université Libre de Bruxelles, Bryssel, Belgien

Peter W. Higgs
University of Edinburgh, Storbritannien

"för den teoretiska upptäckten av en mekanism som bidrar till förståelsen av massans ursprung hos subatomära partiklar, och som nyligen, genom upptäckten av den förutsagda fundamentala partikeln, bekräftats av ATLAS- och CMS-experimenten vid CERN:s accelerator LHC"

Äntligen här!

François Englert och Peter W. Higgs delar årets Nobelpris i fysik för teorin om hur partiklar får sin massa. Oberoende av varandra föreslog de teorin samtidigt år 1964 (Englert tillsammans med sin numera avlidne kollega Robert Brout). Först 2012 bekräftades deras idéer genom upptäckten av en så kallad *Higgspartikel* vid CERN-laboratoriet utanför Genève i Schweiz.

Den i år prisbelönta teorin är en central del i fysikens standardmodell som beskriver hur världen är uppbyggd. Allting, från blommor och människor till stjärnor och planeter, består enligt standardmodellen av några få byggstenar, *materiepartiklar*. Dessa partiklar styrs av krafter som förmedlas av *kraftpartiklar* som ser till att allt fungerar som det ska.

Hela standardmodellen vilar på att det också finns en särskilt sorts partikel, Higgspartikeln. Denna är en vibration av ett osynligt fält som fyller rymden. Till och med när universum verkar tömt på allt, finns fältet där. Utan det skulle vi inte finnas, för det är genom kontakten med fältet som partiklarna får sin massa. Den av Englert och Higgs föreslagna teorin beskriver hur detta går till.

Den 4 juli 2012 bekräftades teorin i och med upptäckten

partikelkolliderare, LHC (Large Hadron Collider), är troligen den största och mest komplicerade maskin som någonsin byggts av människor. Ur miljarder partikelkrockar i LHC lyckades två grupper, ATLAS och CMS, med cirka 3 000 forskare var, vaska fram Higgspartikeln.

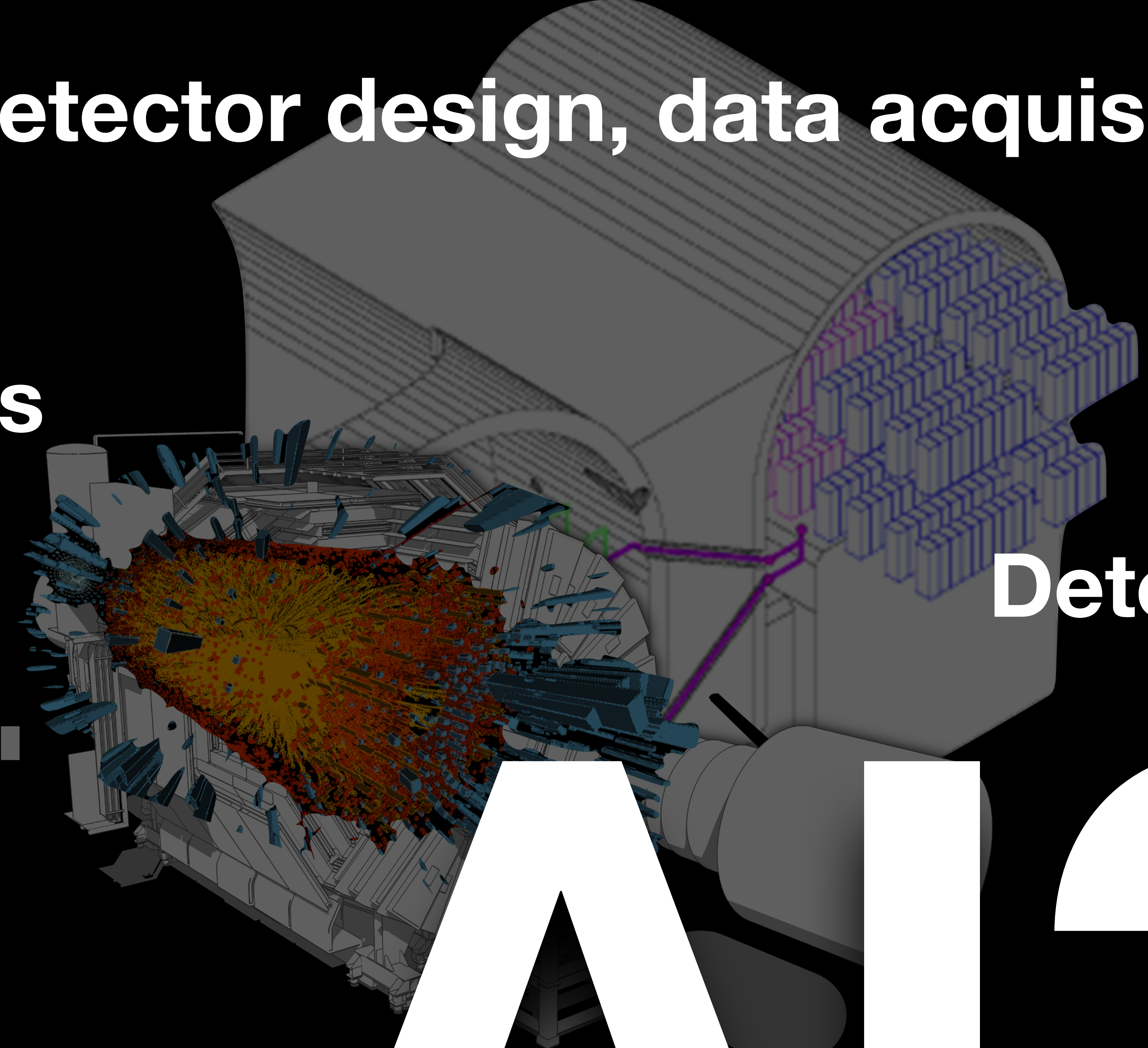
Även om det är ett storverk att finna Higgspartikeln, den sista pusselbiten som fattades i standardmodellen, så är standardmodellen inte den sista biten i pusslet om hela universum. Ett av skälen är att vissa partiklar, neutriner, beskrivs i standardmodellen som masslösa, medan ny forskning pekar mot att de faktiskt har massa. Ett annat skäl är att modellen bara omfattar den synliga materien, vilken endast är en femtedel av all materia som finns i världsalltet. Att hitta den mystiska mörka materien är ett av målen för den fortsatta jakten på okända partiklar vid CERN.

François Englert, belgisk medborgare. Född 1932 (80 år) i Etterbeek, Belgien. Fil.dr 1959 vid Université Libre de Bruxelles, Bryssel, Belgien. Professor emeritus vid Université Libre de Bruxelles, Bryssel, Belgien.
www.ulb.ac.be/sciences/physth/people_FEnglert.html

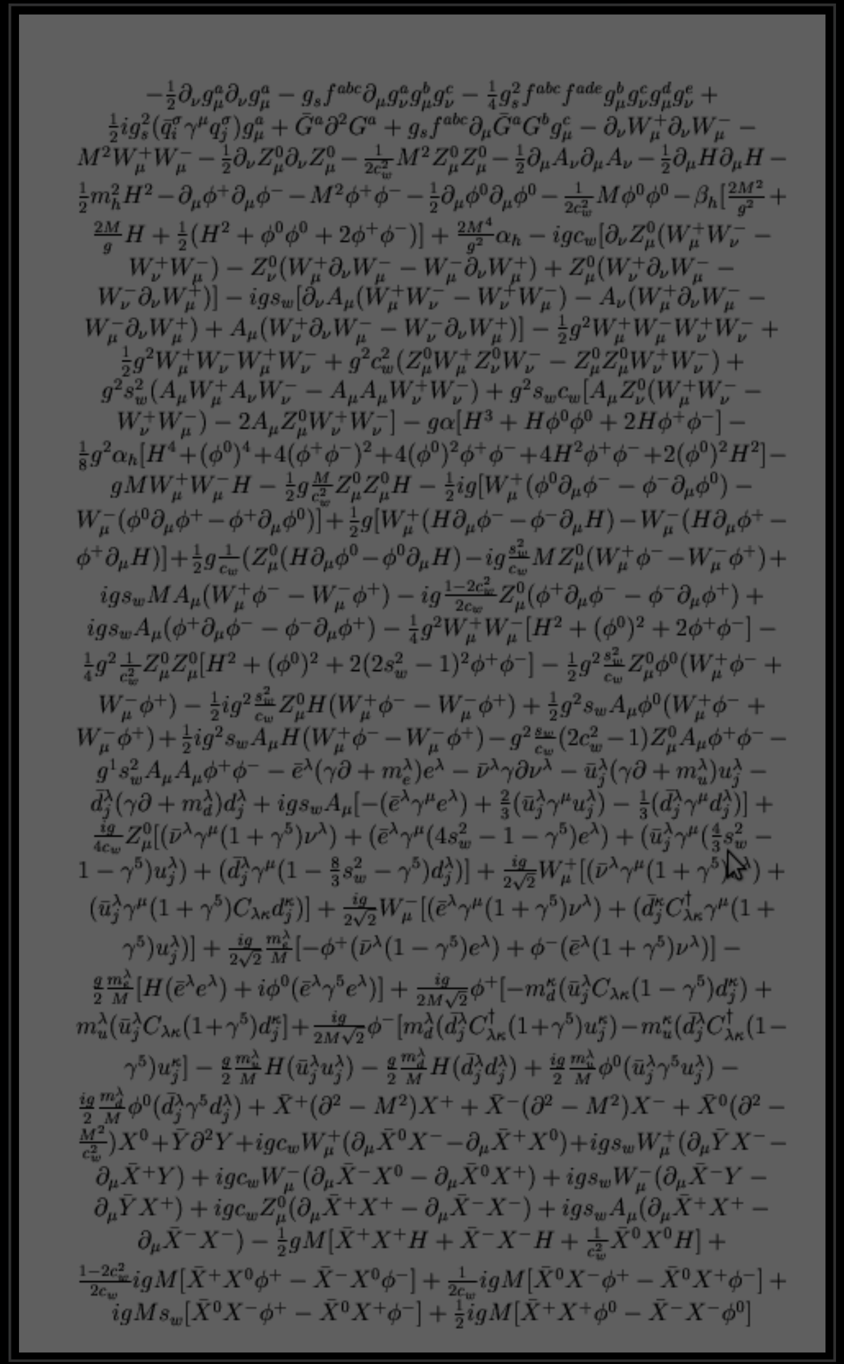
Peter W. Higgs, brittisk medborgare. Född 1929 (84 år) i Newcastle upon Tyne, Storbritannien. Fil.dr 1954 vid King's College, University of London, Storbritannien. Professor emeritus vid University of Edinburgh, Storbritannien.
www.ph.ed.ac.uk/higgs/


AI?
O(10)
O(10^3)
O(10^10)
10^-18m
10^-15m
10^-6m
100m
GEN
SIM
DIGI+RECO
pp collisions up to production of stable particles [Easy & Fast]
detector response simulation [Hard & Slow]
GEANT4
A SIMULATION TOOLKIT
Energy deposits→digital signals→reconstructed by the reconstruction software [Hard & Slow]
19.7 fb^-1 (8 TeV) + 5.1 fb^-1 (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
±1σ
±2σ
μ̂ = 1.14 +0.26 −0.23
m̂H = 124.70 ± 0.34 GeV
×10^3

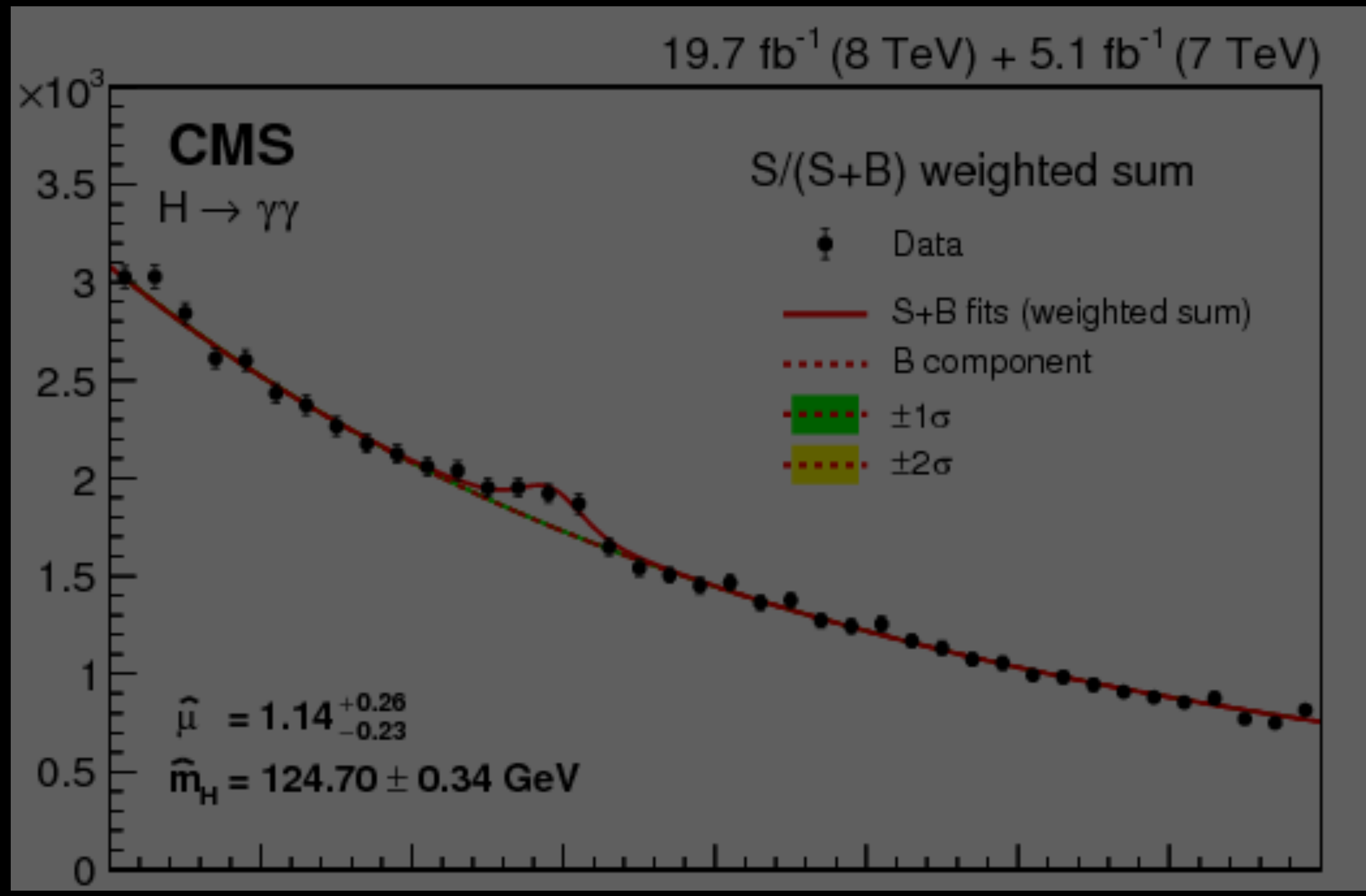
Detector design, data acquisition and triggering



Mainly fast integrals

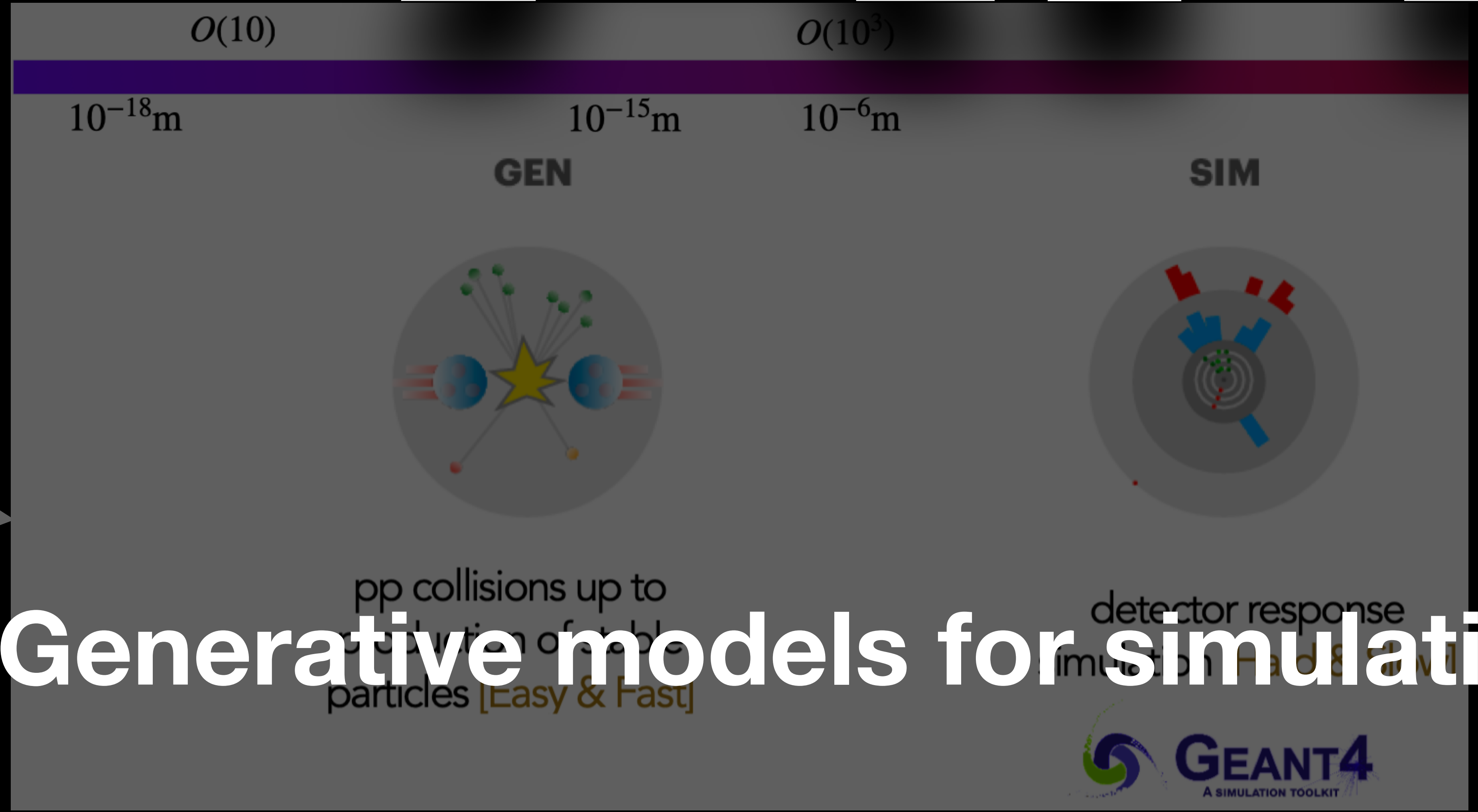


Detector reconstruction and tagging

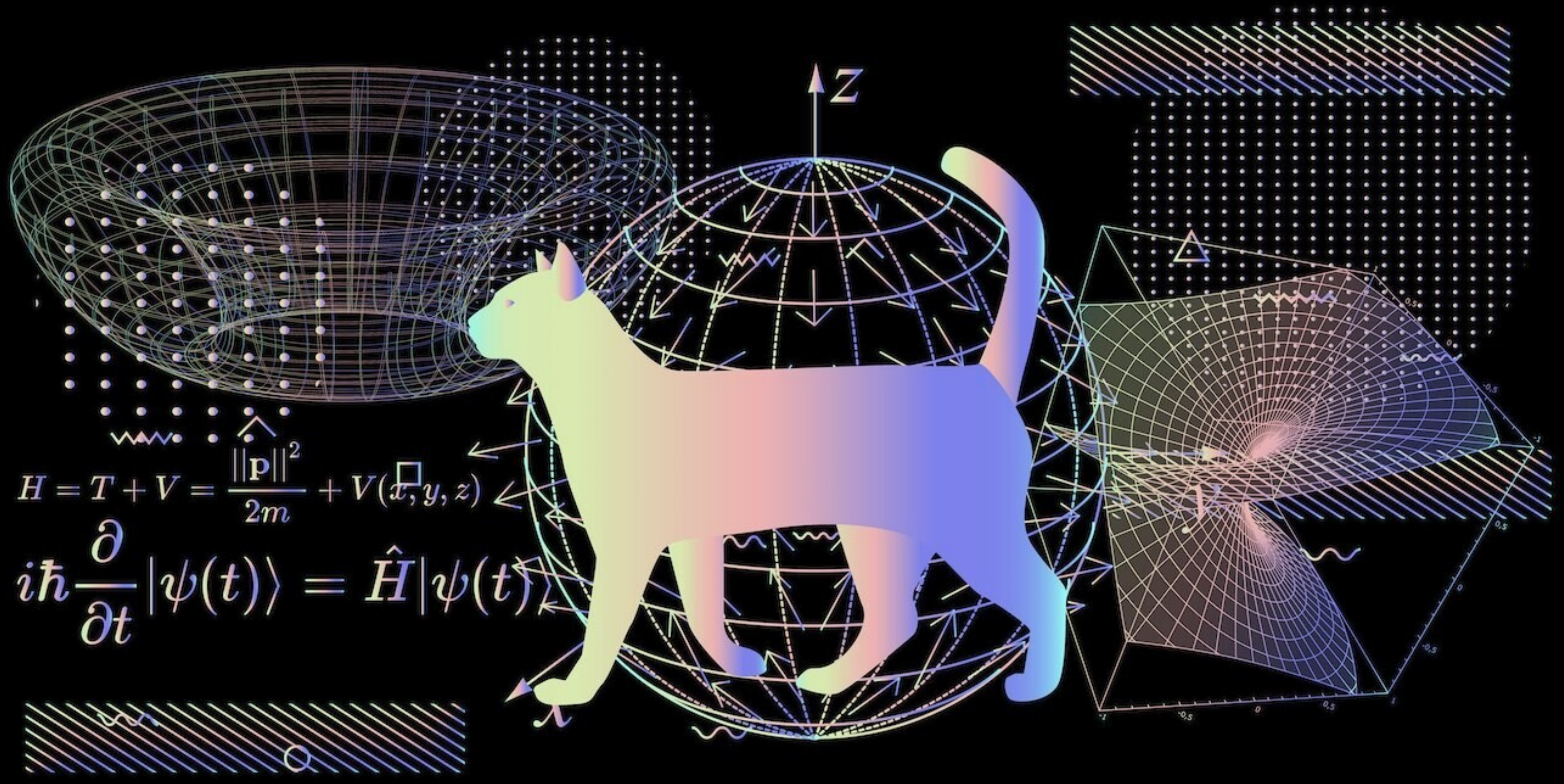


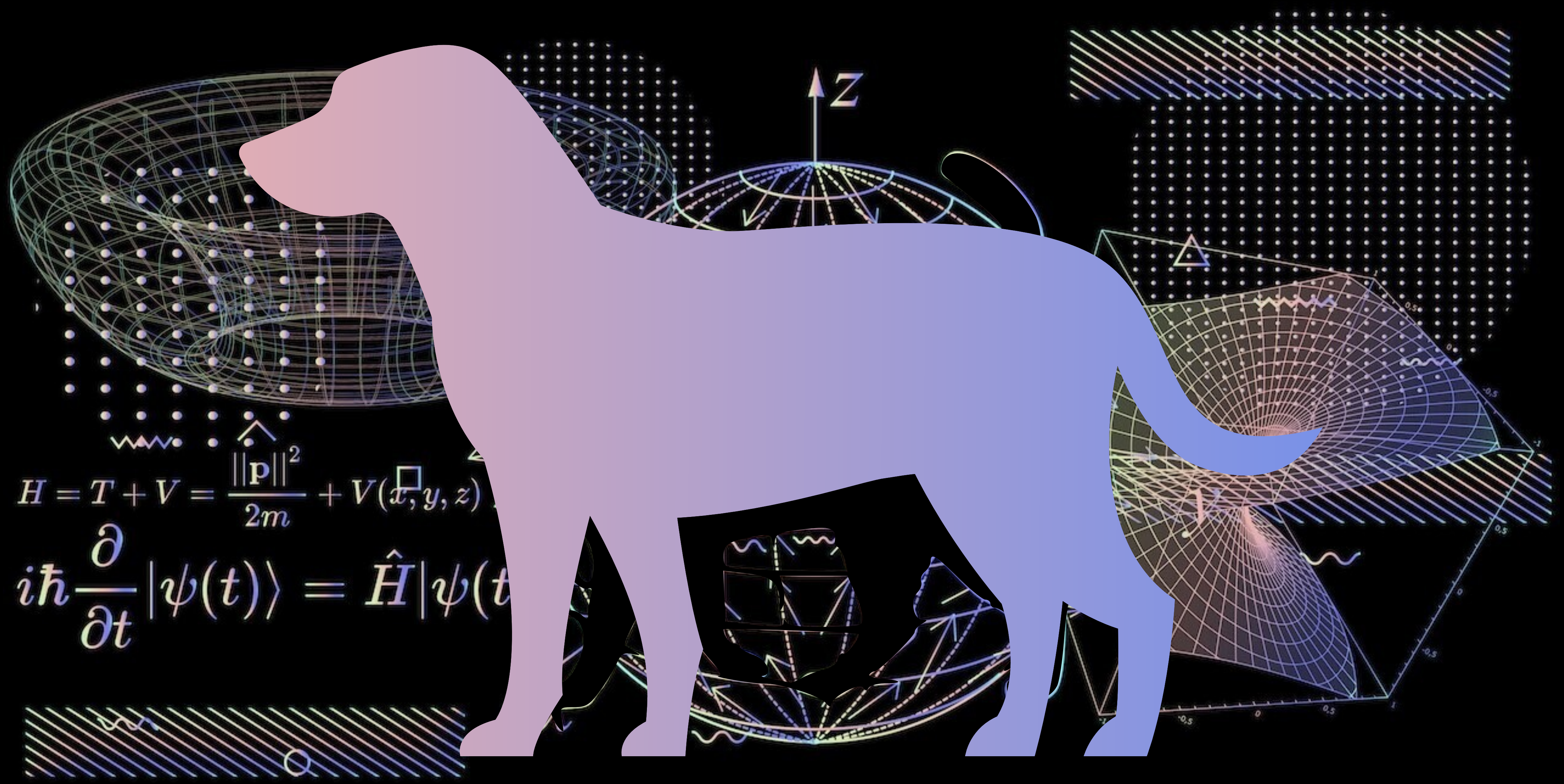
AI?

Data analysis

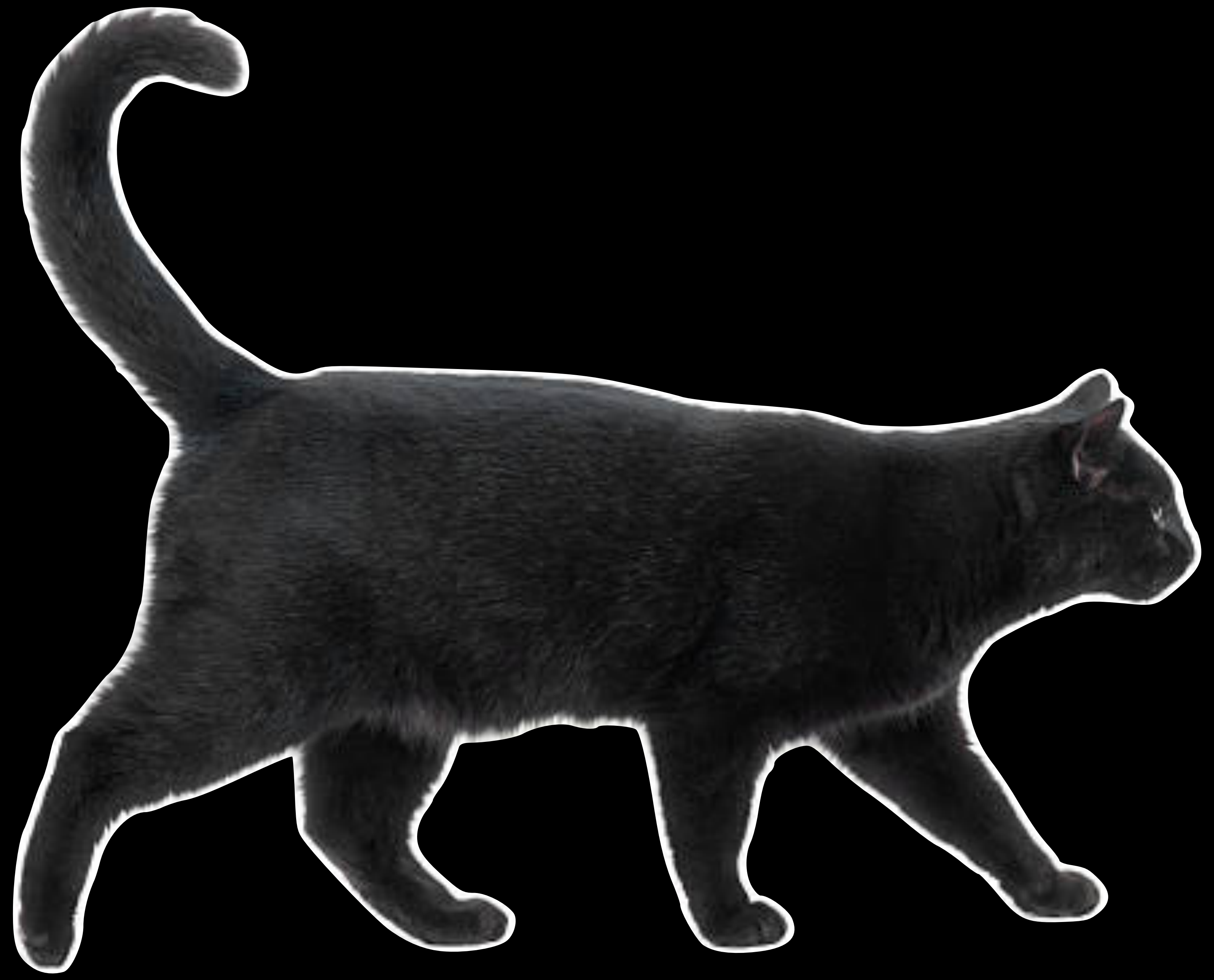


Generative models for simulation


z
$H = T + V = \frac{\|\mathbf{p}\|^2}{2m} + V(x, y, z)$
$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle$

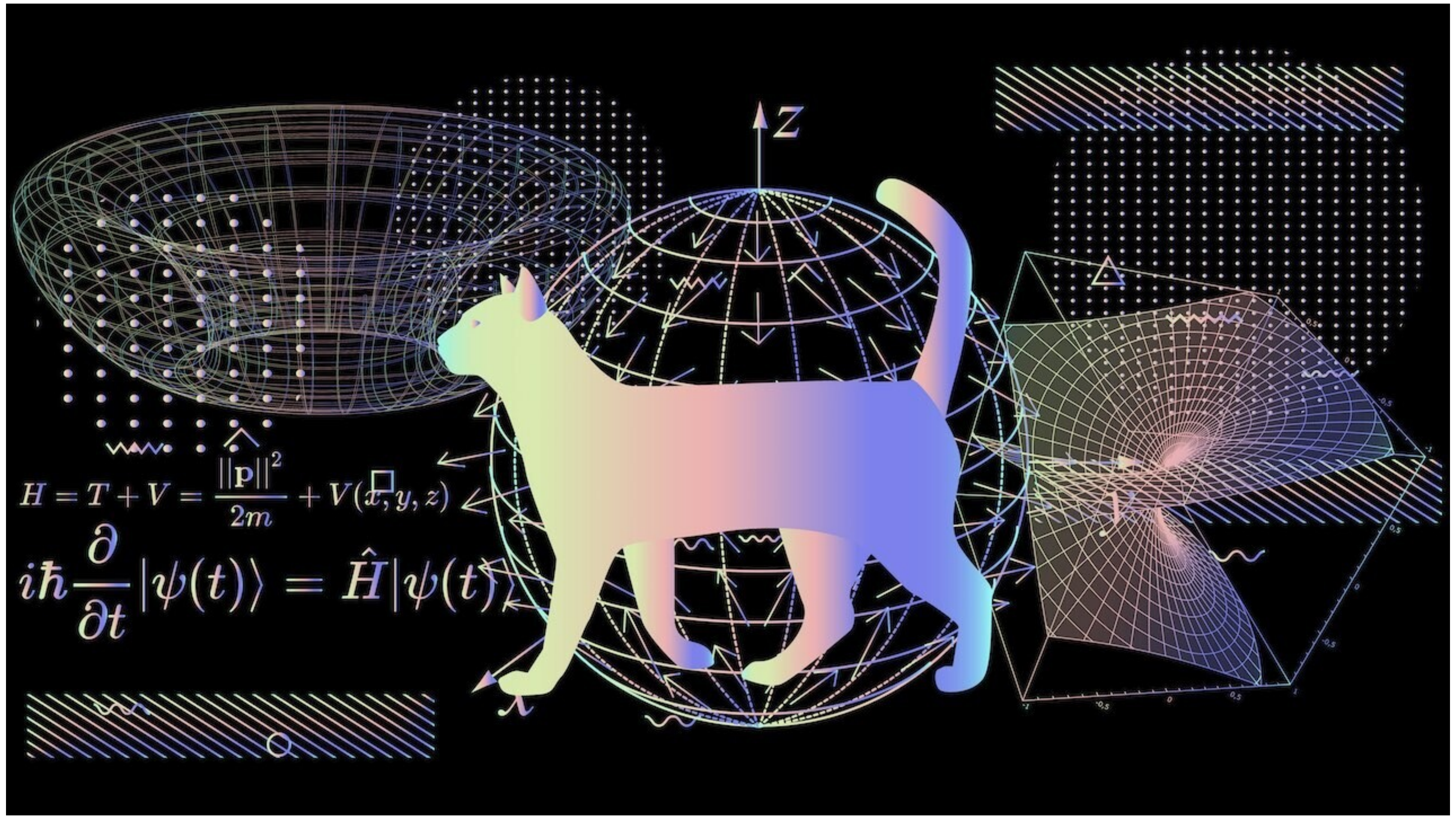

z
$H = T + V = \frac{\|\mathbf{p}\|^2}{2m} + V(x, y, z)$

!=



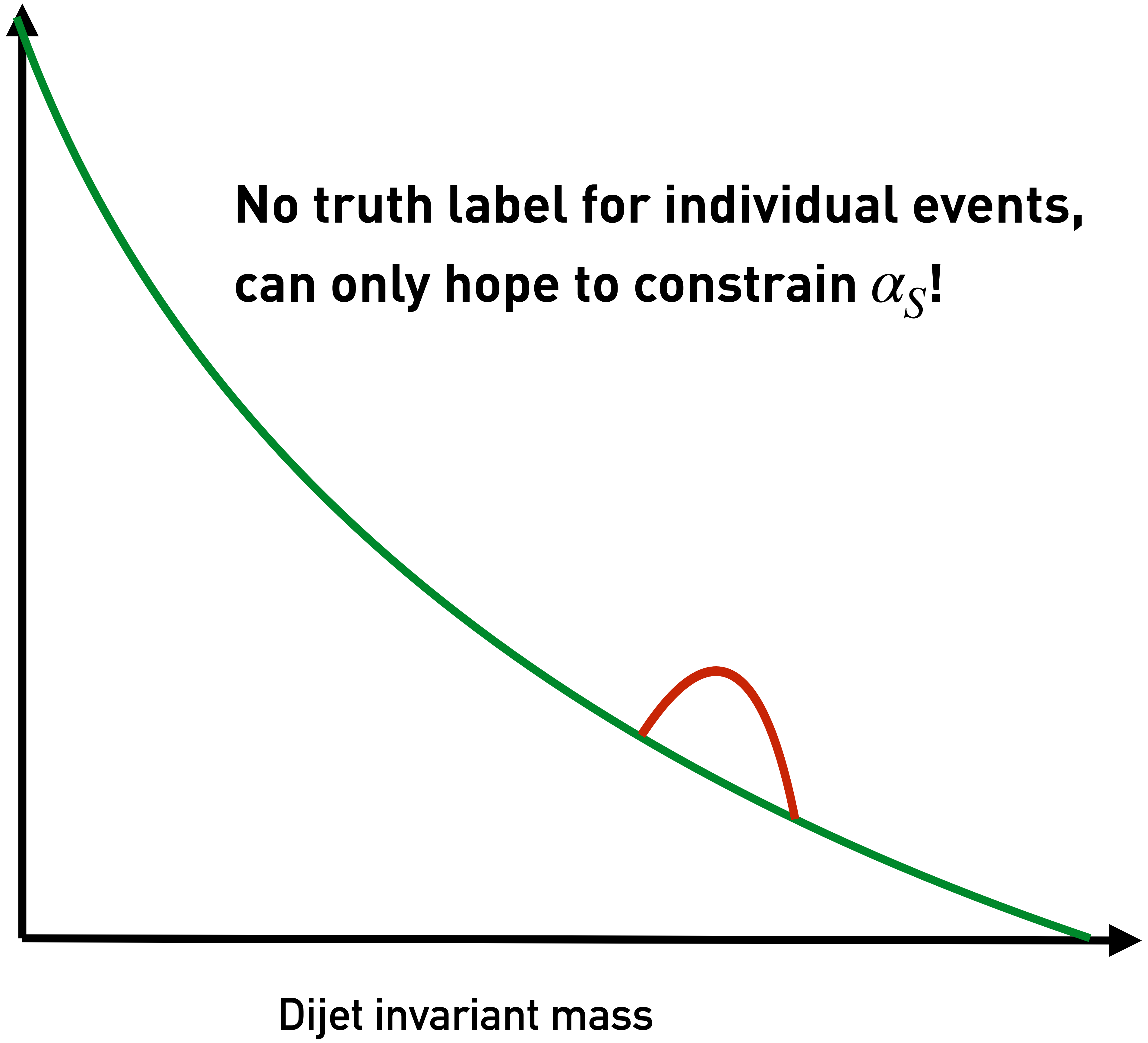
No labeled data!

$$dP^n_{data} = |M_S + M_B|^2 dp_1 dp_2 \ldots dp_n$$



$$M_S M_B{}^* + M_B M_S{}^*$$

$$P_{data} = \alpha_S P_S + \alpha_B P_B$$



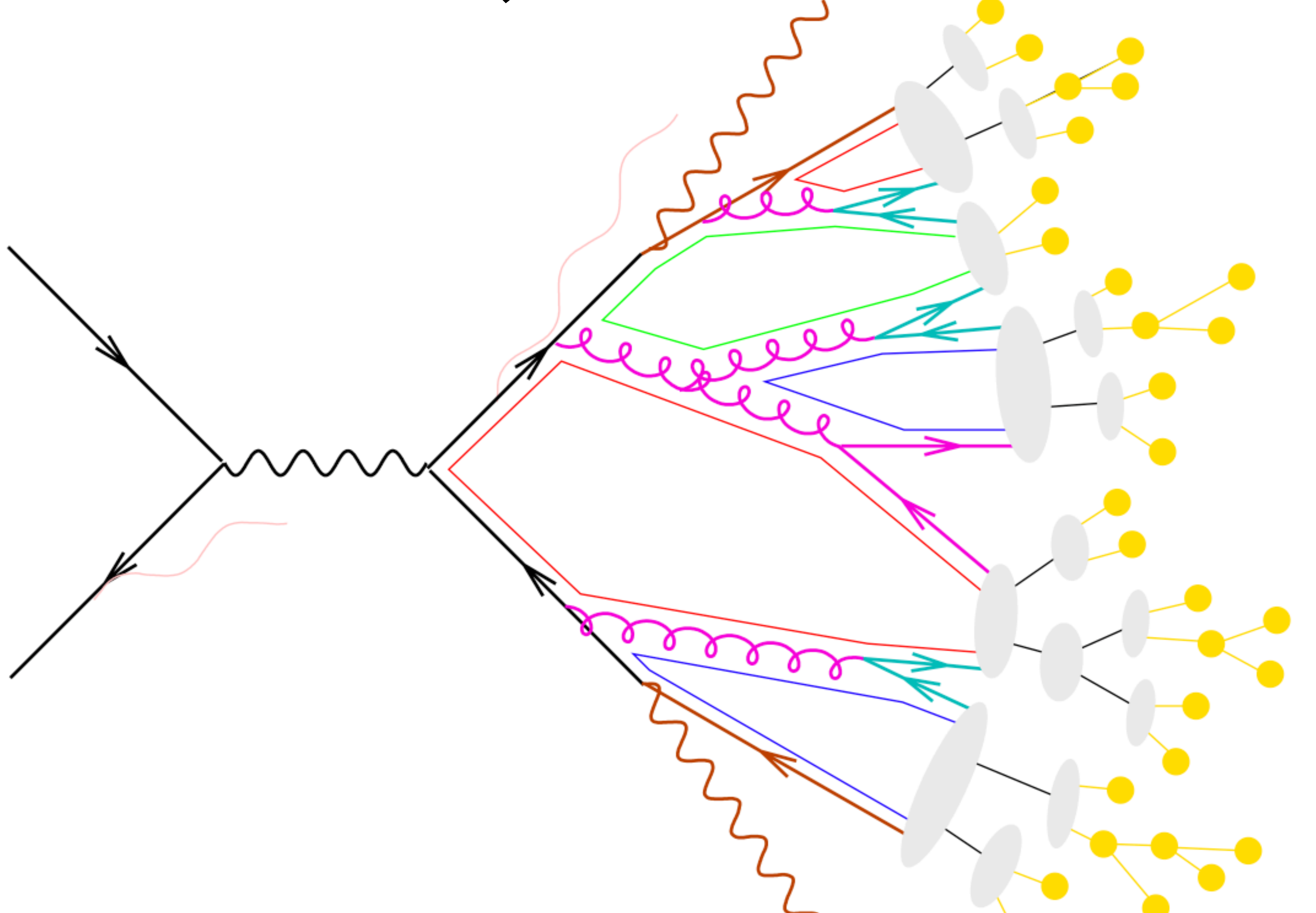
Monte Carlo simulation takes us over 20 orders of magnitude in length!

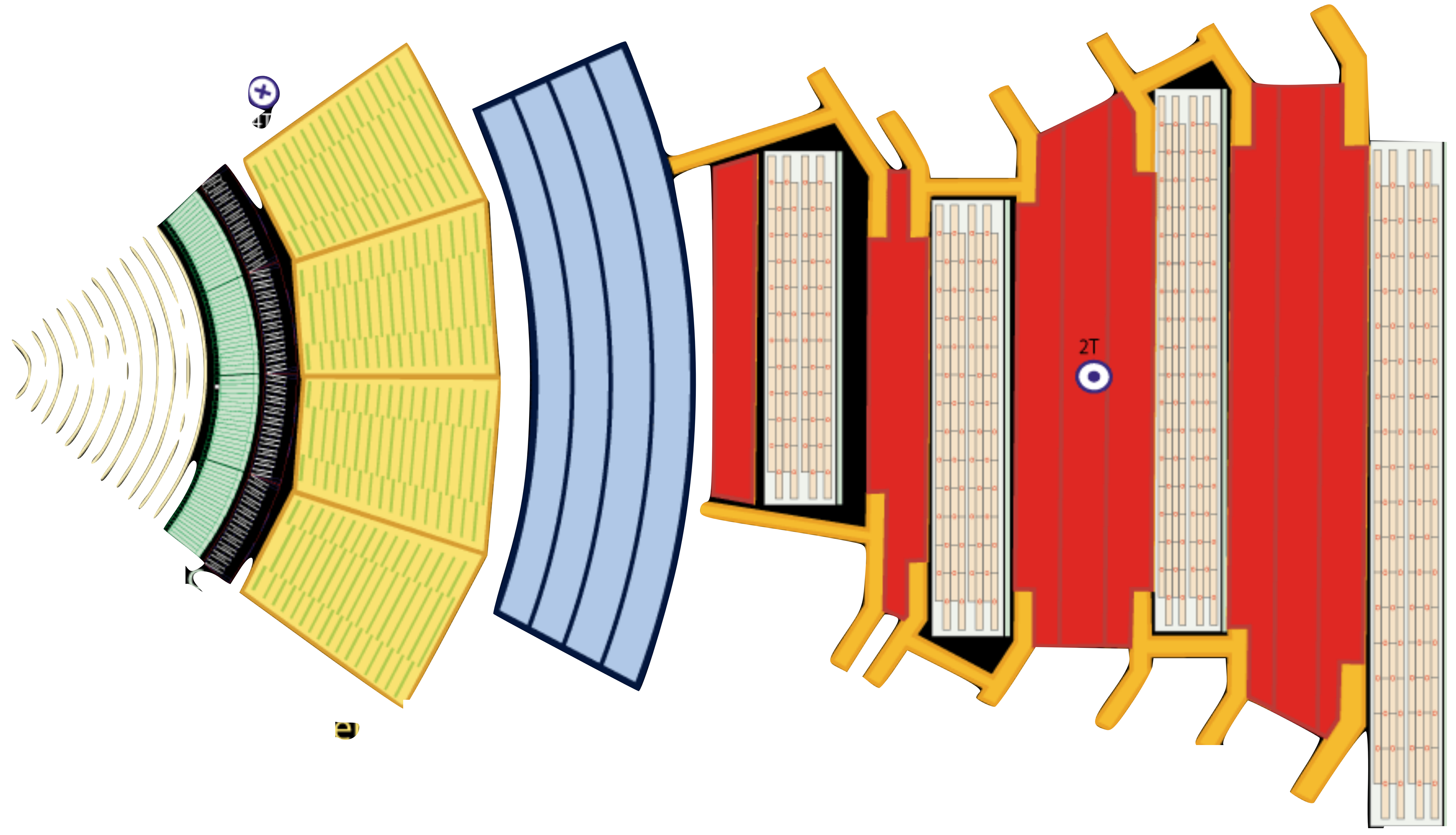
Dimensions

$\mathcal{O}(10)$ $\mathcal{O}(10^3)$ $\mathcal{O}(10^{10})$

10^{-18}m 10^{-15}m 10^{-6}m 100m

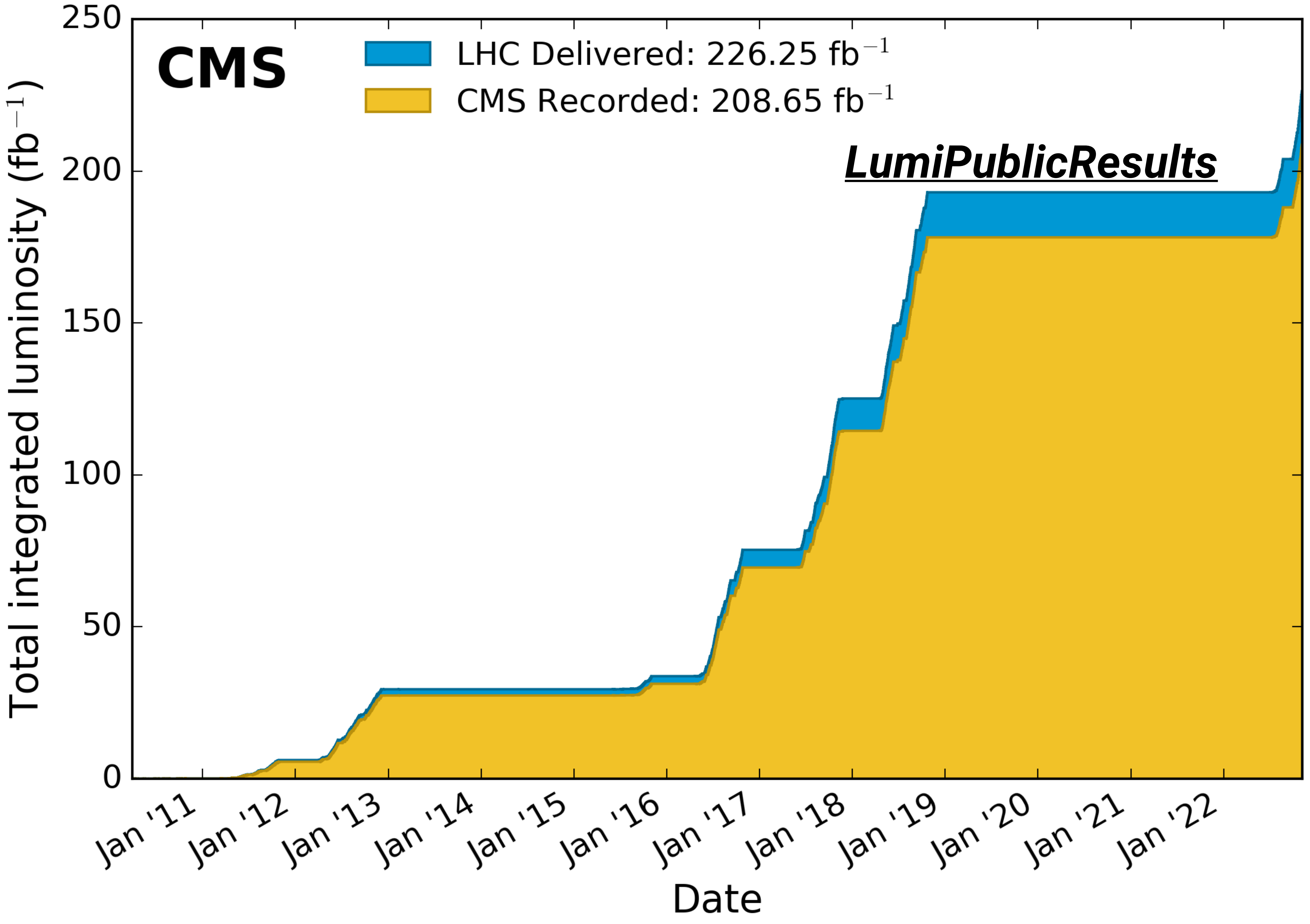
Perturbative QCD ***Markov model*** ***Hadronization*** ***Detection***





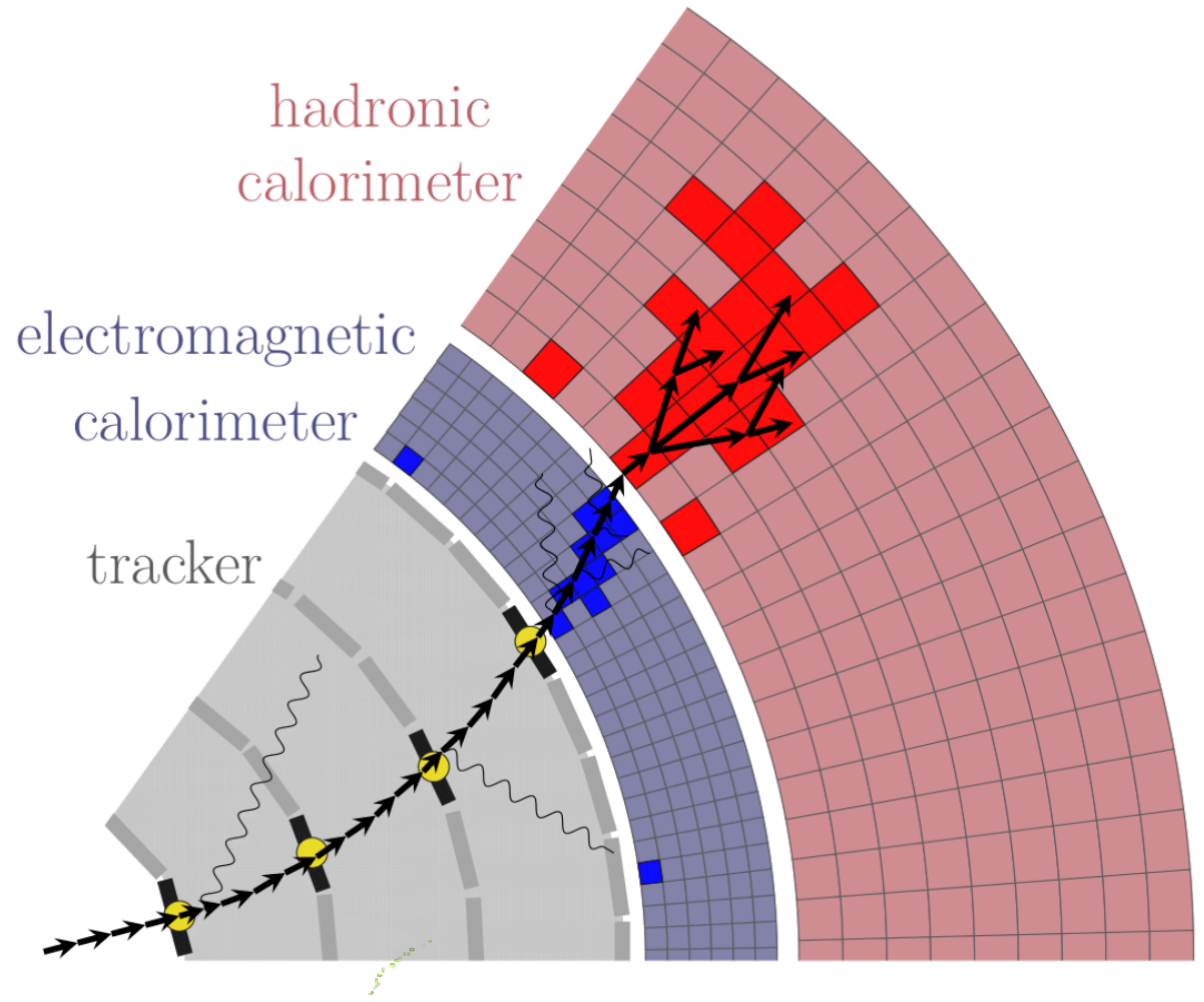
CMSOfflineComputingResults

~40 quadrillion collisions recorded at LHC (1 fb^{-1} ~ 100 trillion collisions)

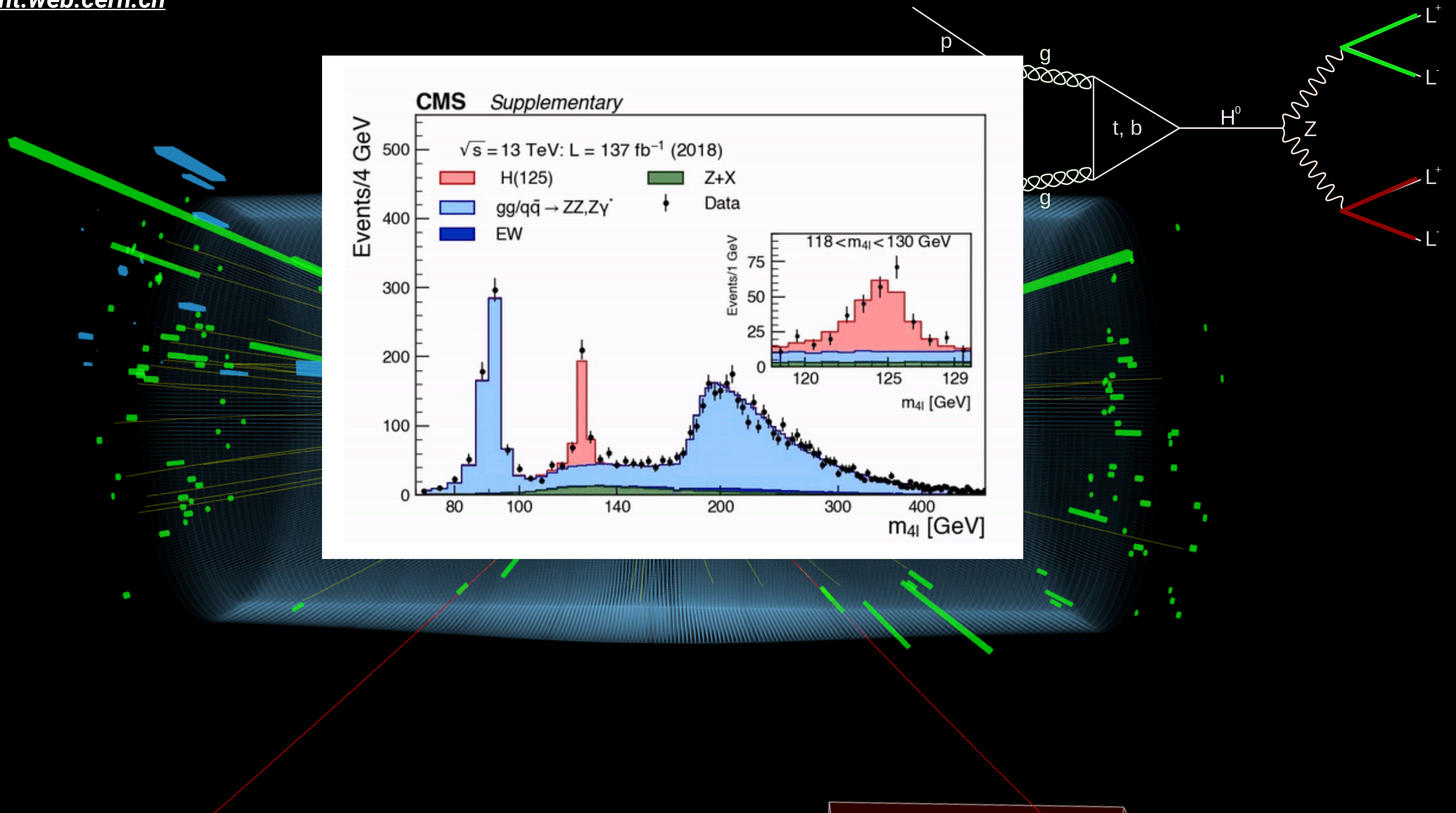


278 petabytes of data (Netflix 24/7 for more than 15,000 years)

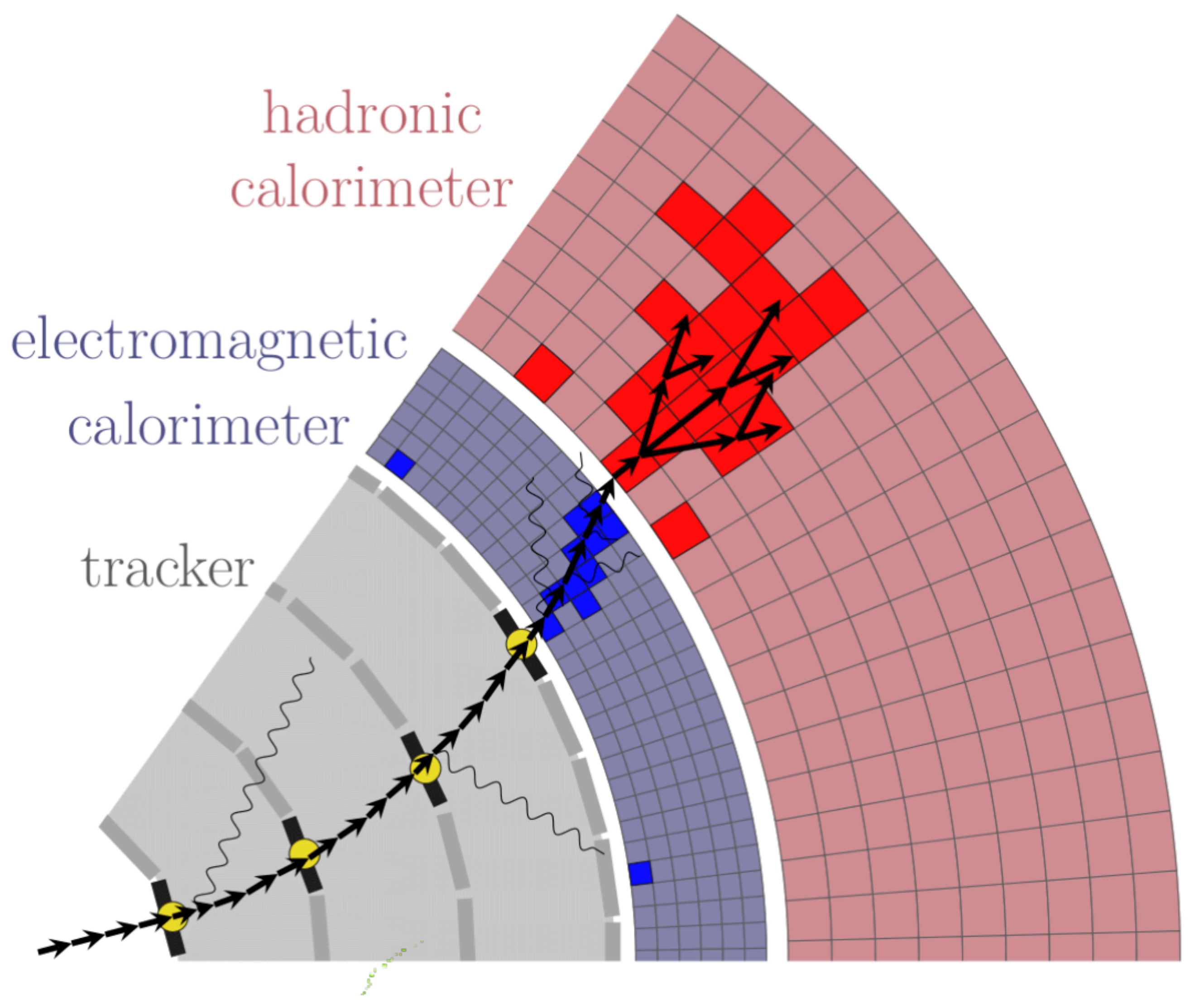
O(1) trillion simulated events

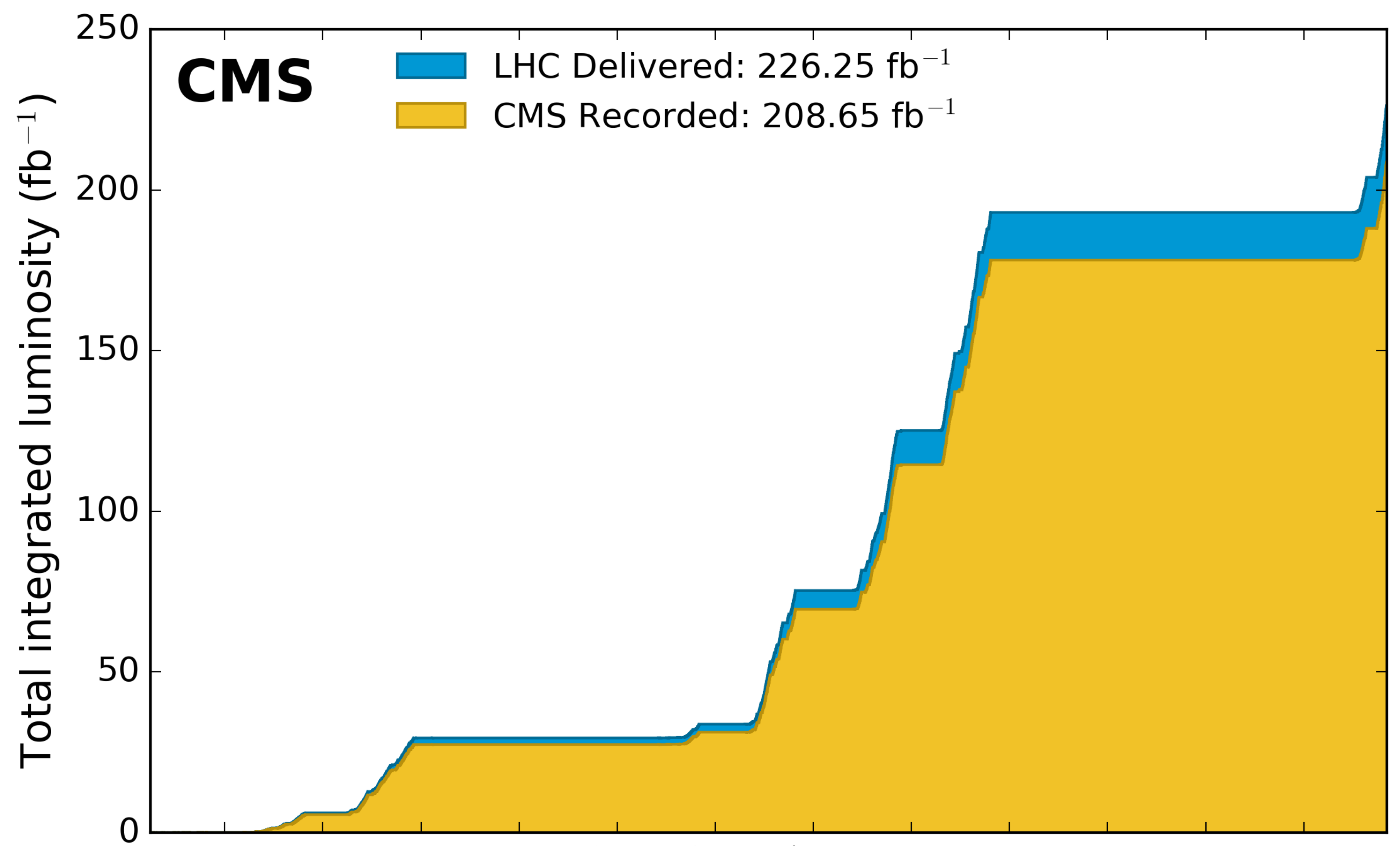


cmsexperiment.web.cern.ch



We had to collide billions of protons,
only around 10 signal events were needed to claim discovery!



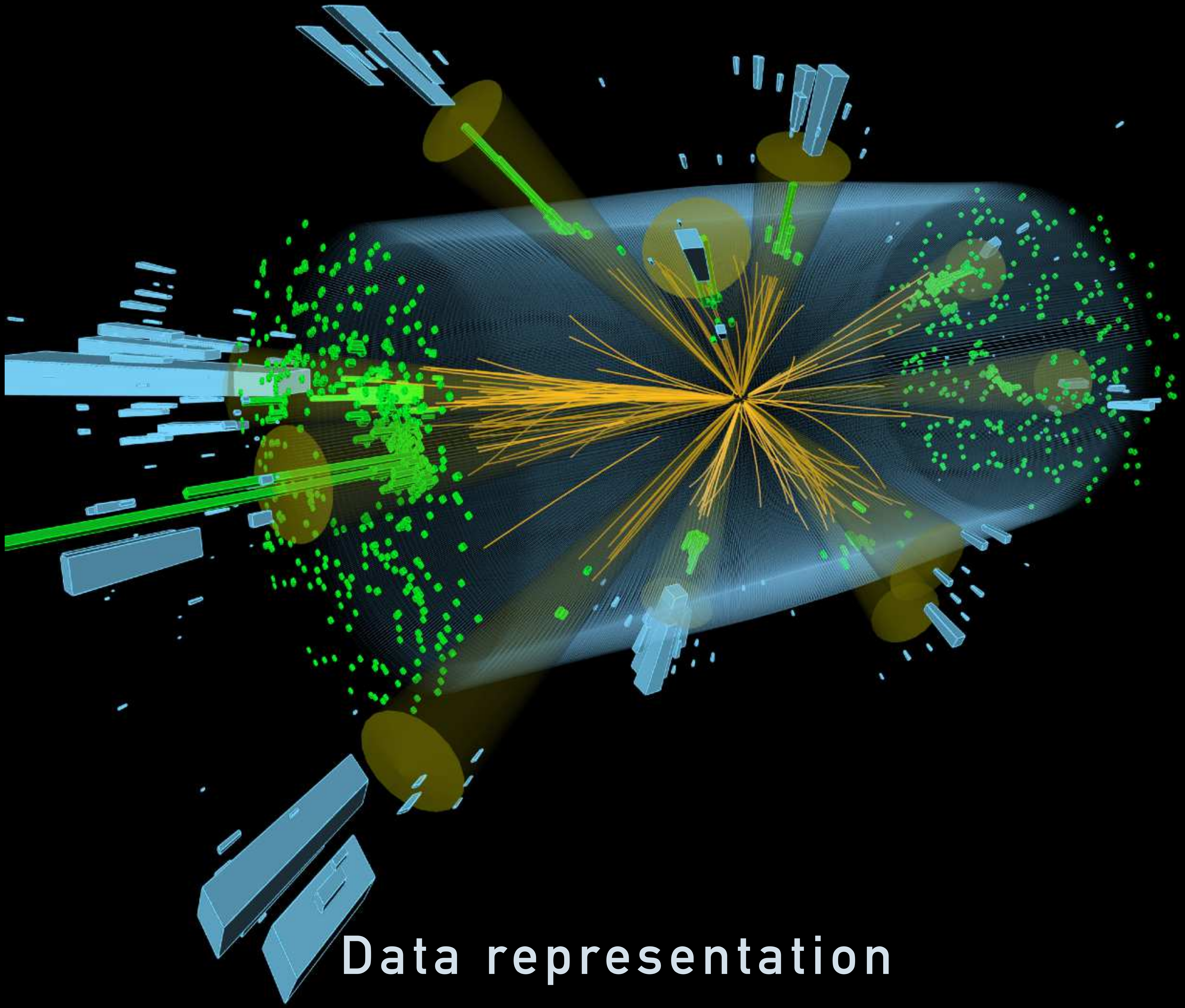


We have a lot of high quality simulated data that we want to use to train AI algorithms!

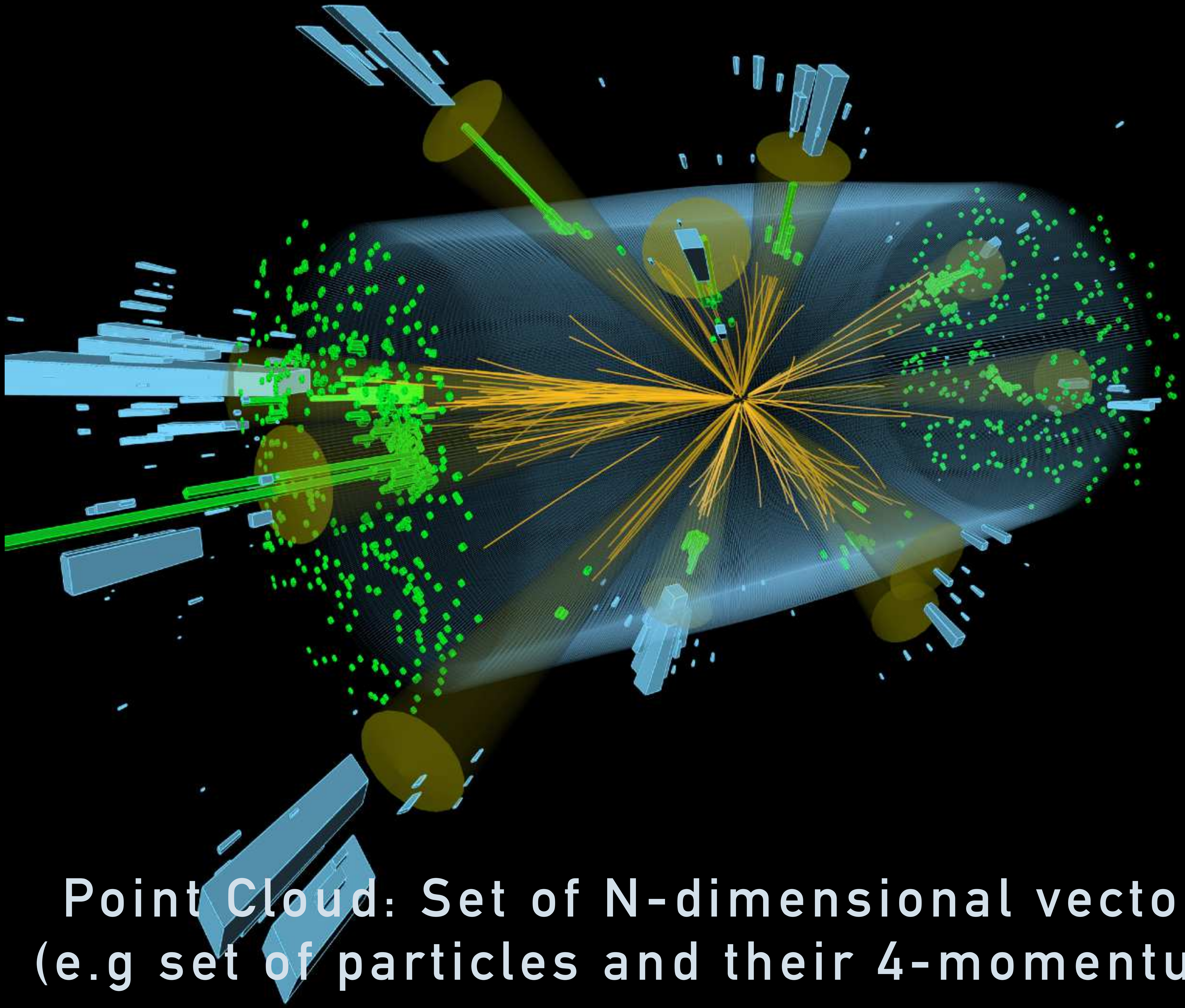
But we have even more unlabelled data we'd like to use!

(Simulation != test data)


Detector reconstruction and tagging
AI?
$O(10^{10})$
100m
IGI+RECO
Energy deposits→digital signals→reconstructed by the reconstruction software [Hard & Slow]
19.7 fb^{-1} (8 TeV) + 5.1 fb^{-1} (7 TeV)
CMS
$H \to \gamma\gamma$
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
$\pm 1\sigma$
$\pm 2\sigma$
$\hat{\mu} = 1.14^{+0.26}_{-0.23}$
$\hat{m}_H = 124.70 \pm 0.34$ GeV
$O(10)$
$O(10^3)$
10^{-18}m
10^{-15}m
10^{-6}m
GEN
SIM
pp collisions up to production of stable particles [Easy & Fast]
detector response simulation [Hard & Slow]
GEANT4

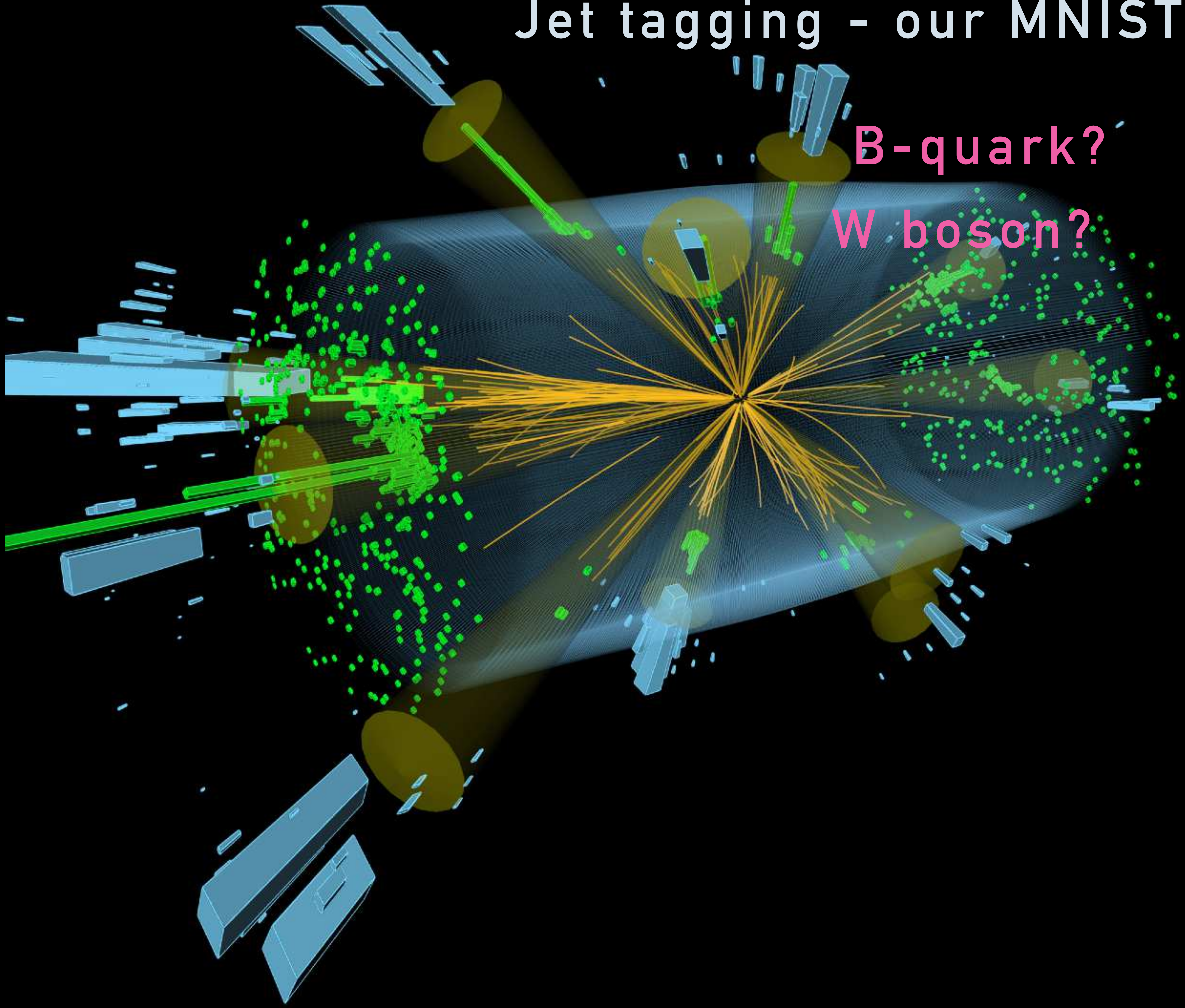


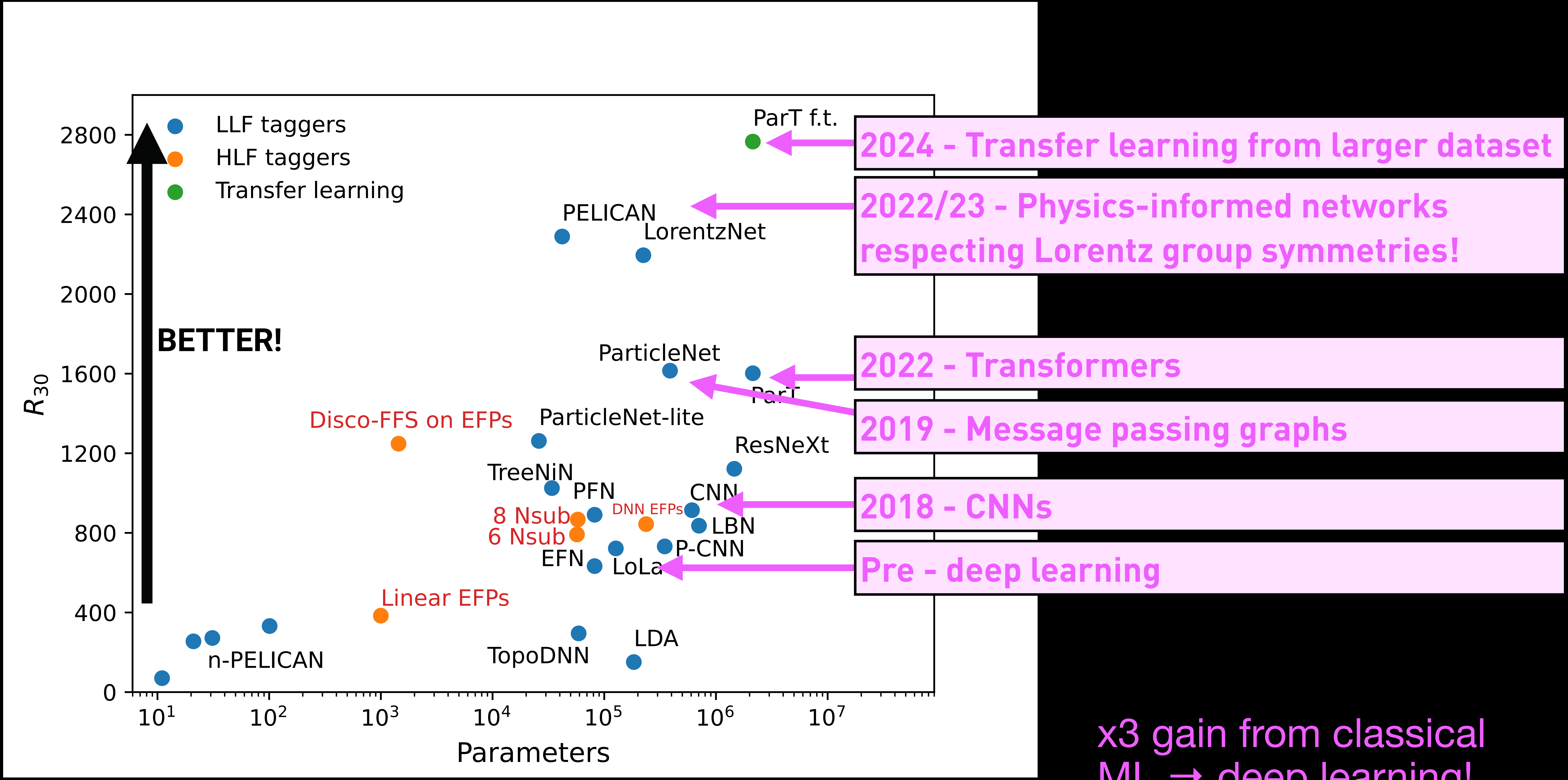
Data representation



Point Cloud: Set of N-dimensional vectors
(e.g set of particles and their 4-momentum)

Jet tagging - our MNIST!





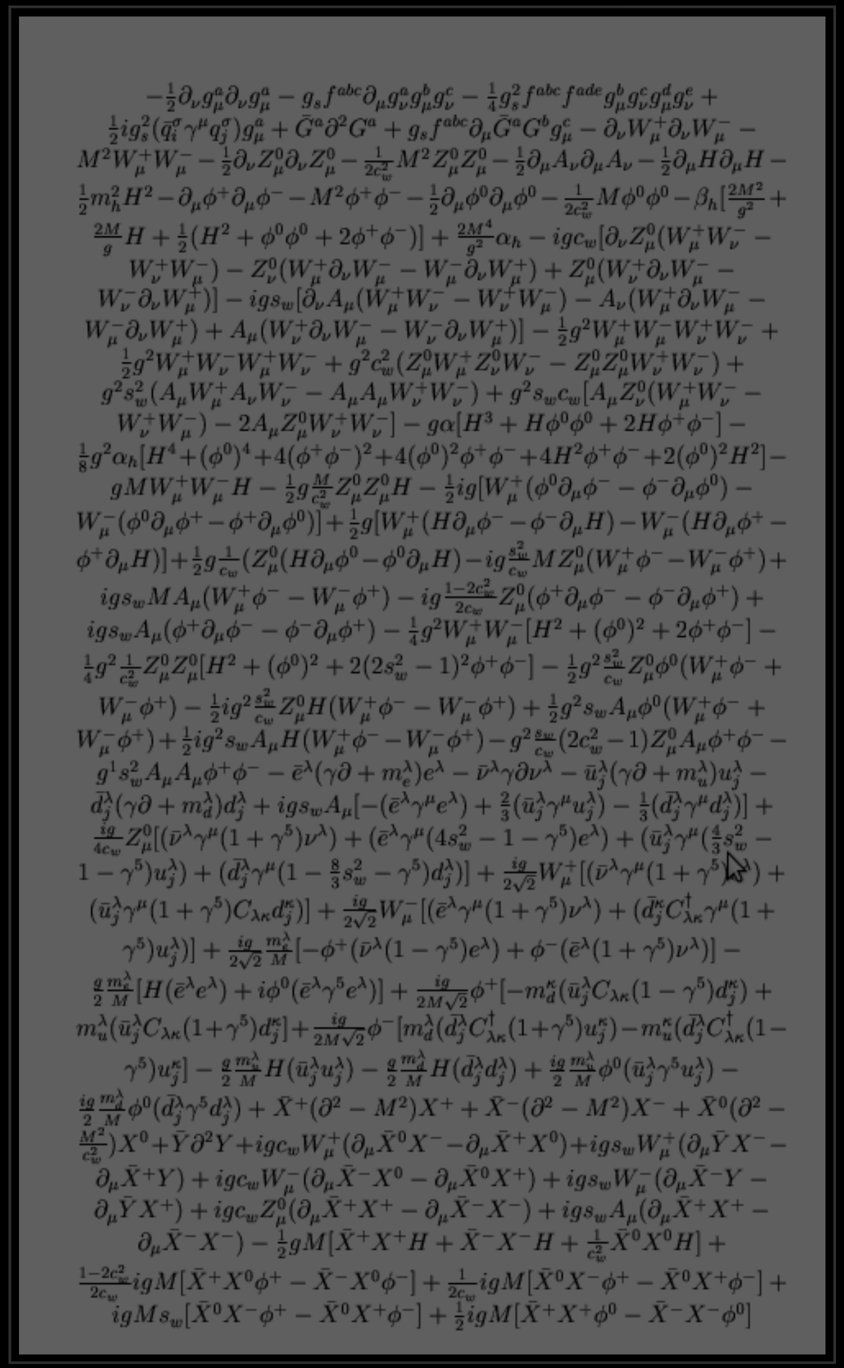
x3 gain from classical ML → deep learning!

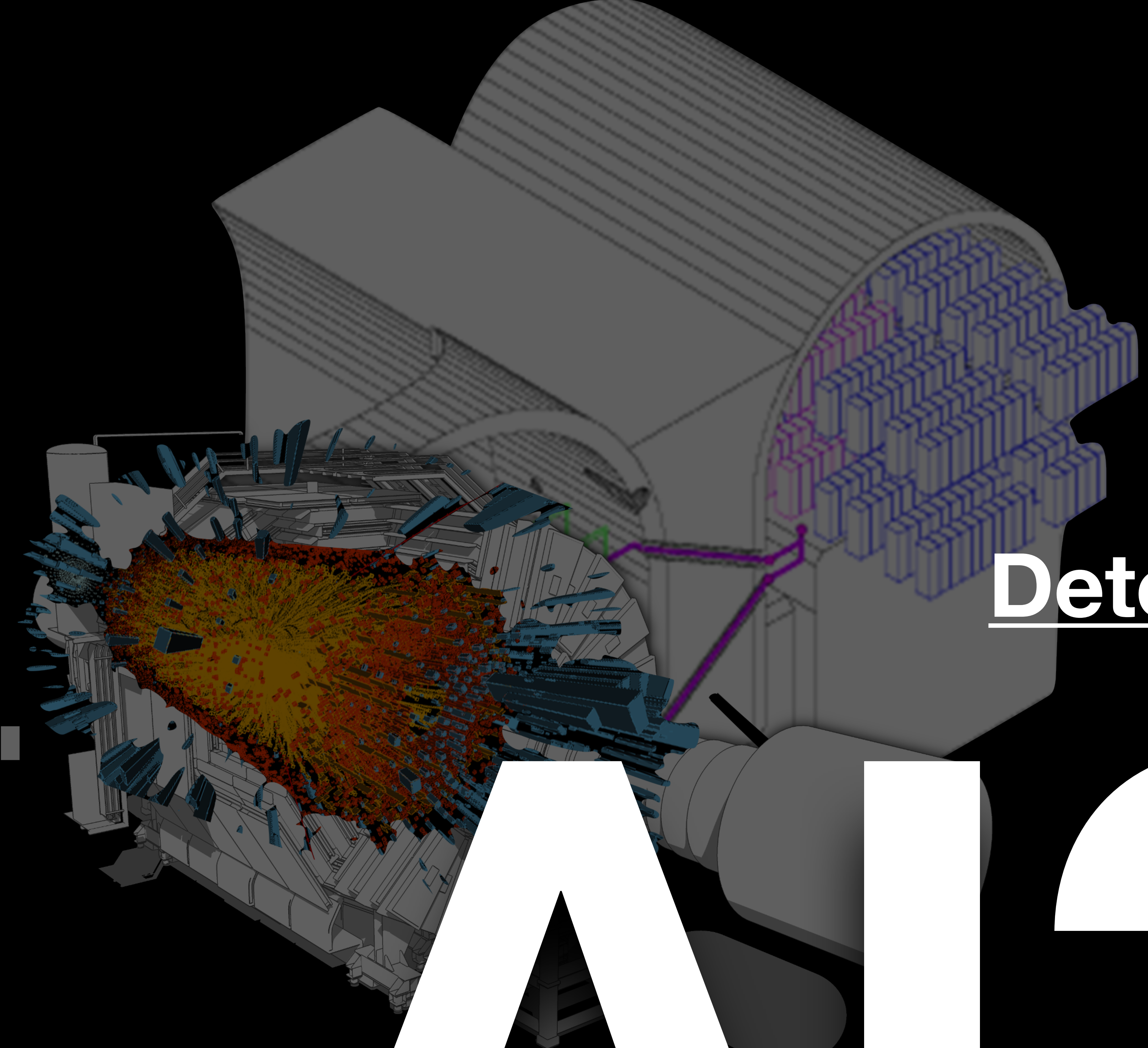
ABCNet:

Pixel intensity = particle importance w.r.t most energetic particle in jet, from attention weights

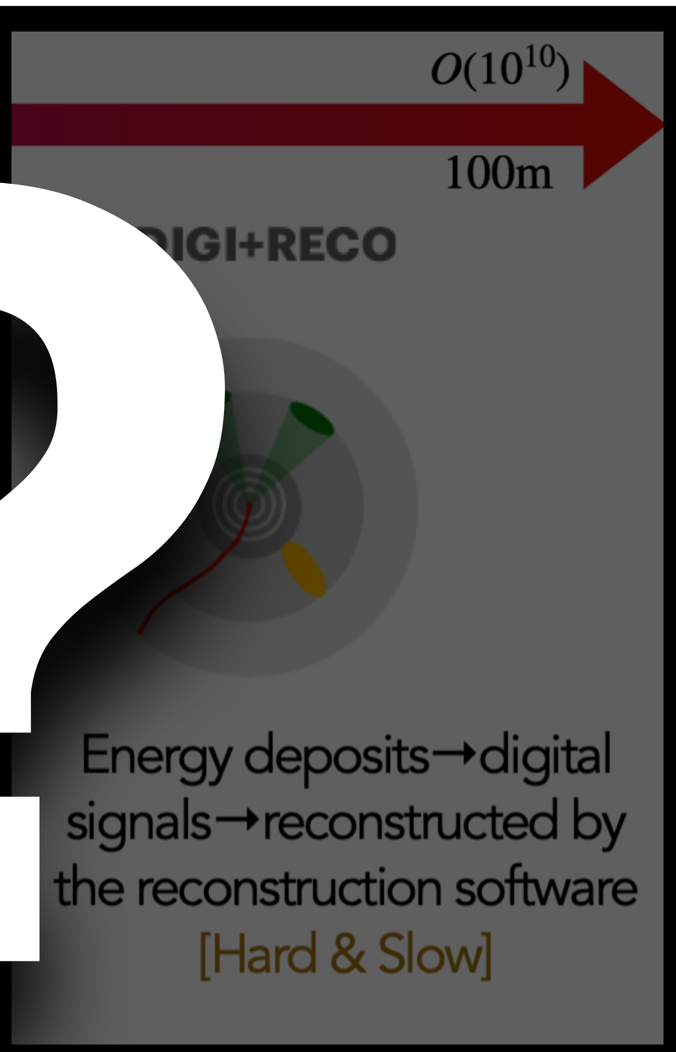
Learned through attention!

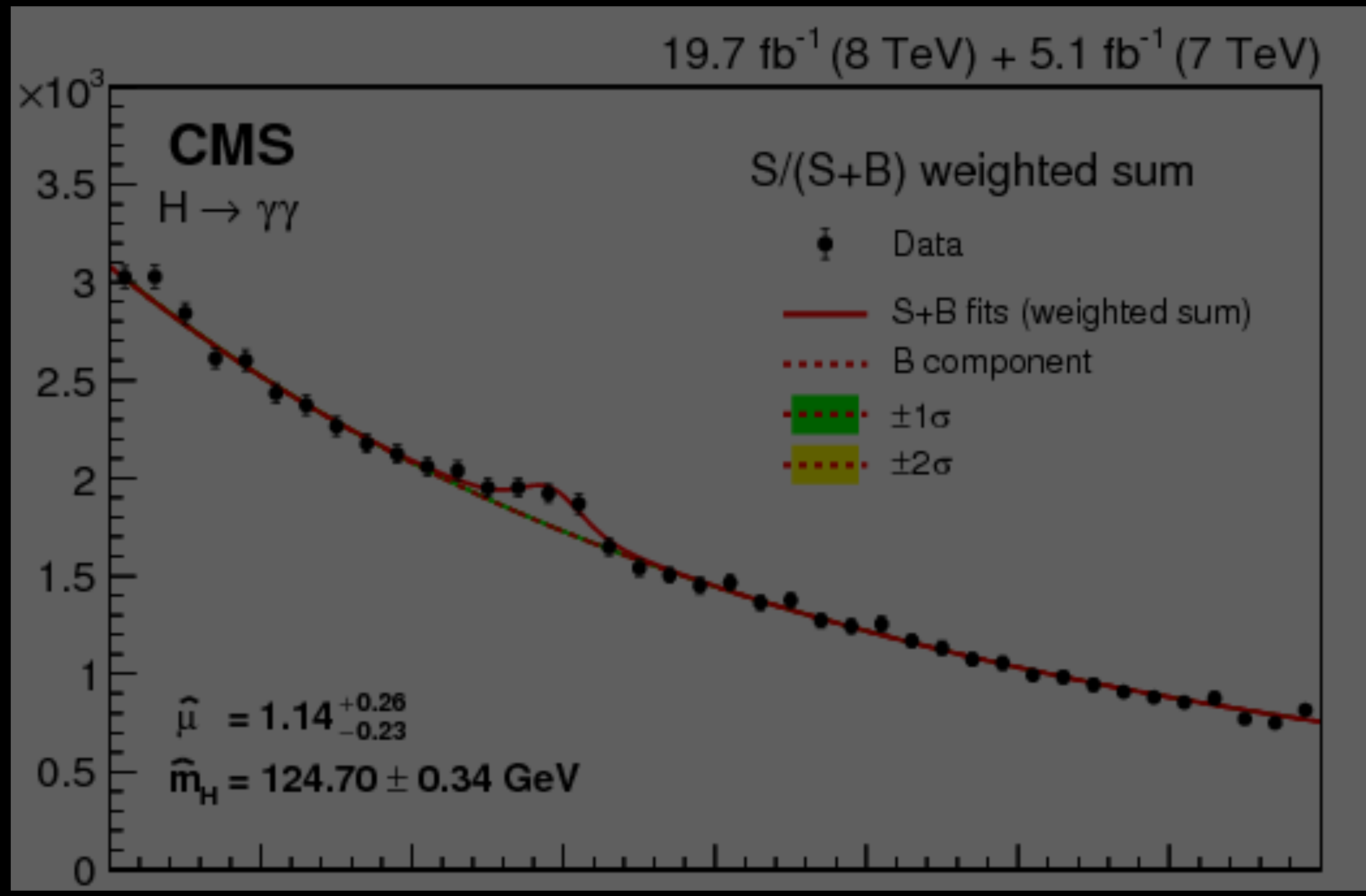
Detector reconstruction and tagging

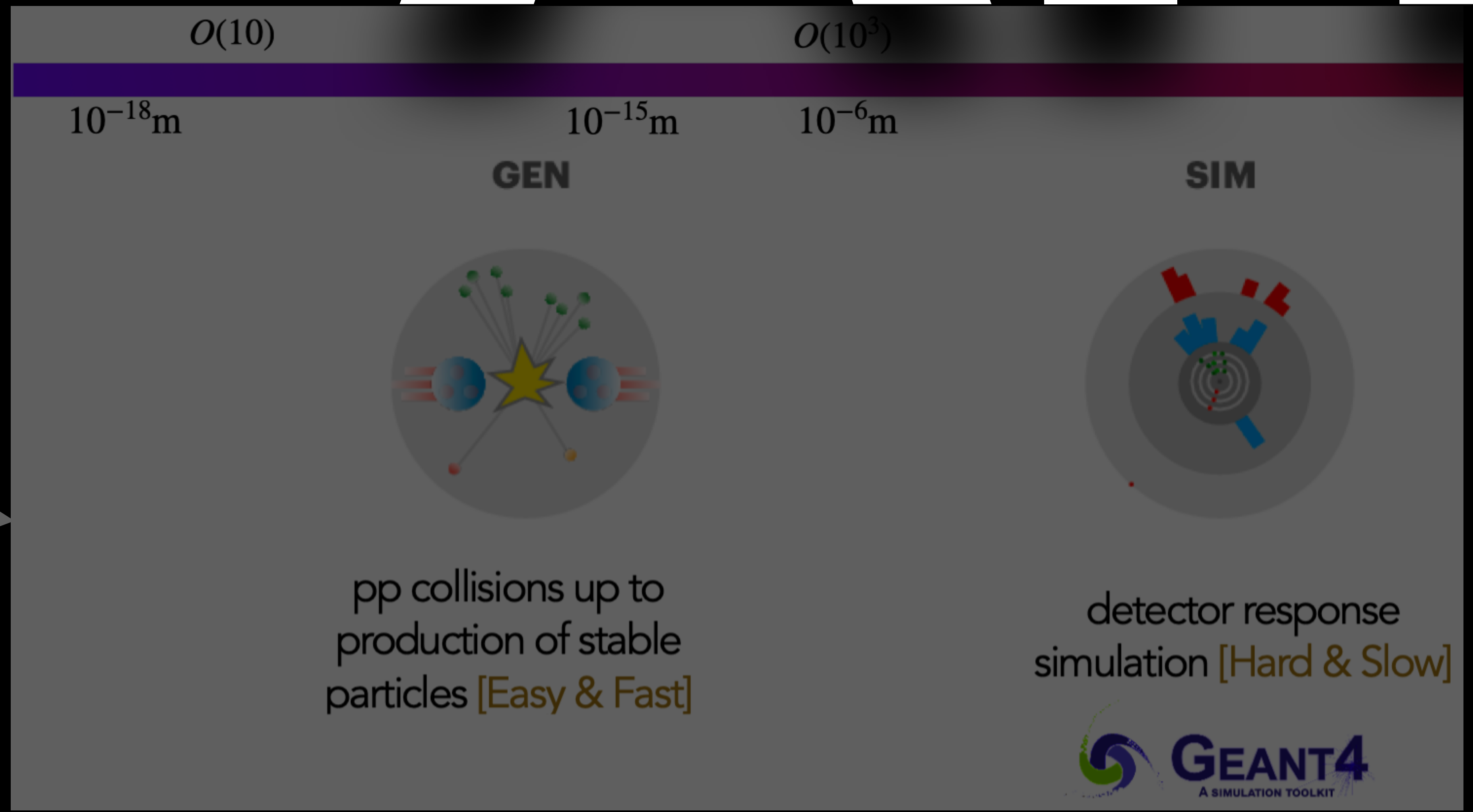




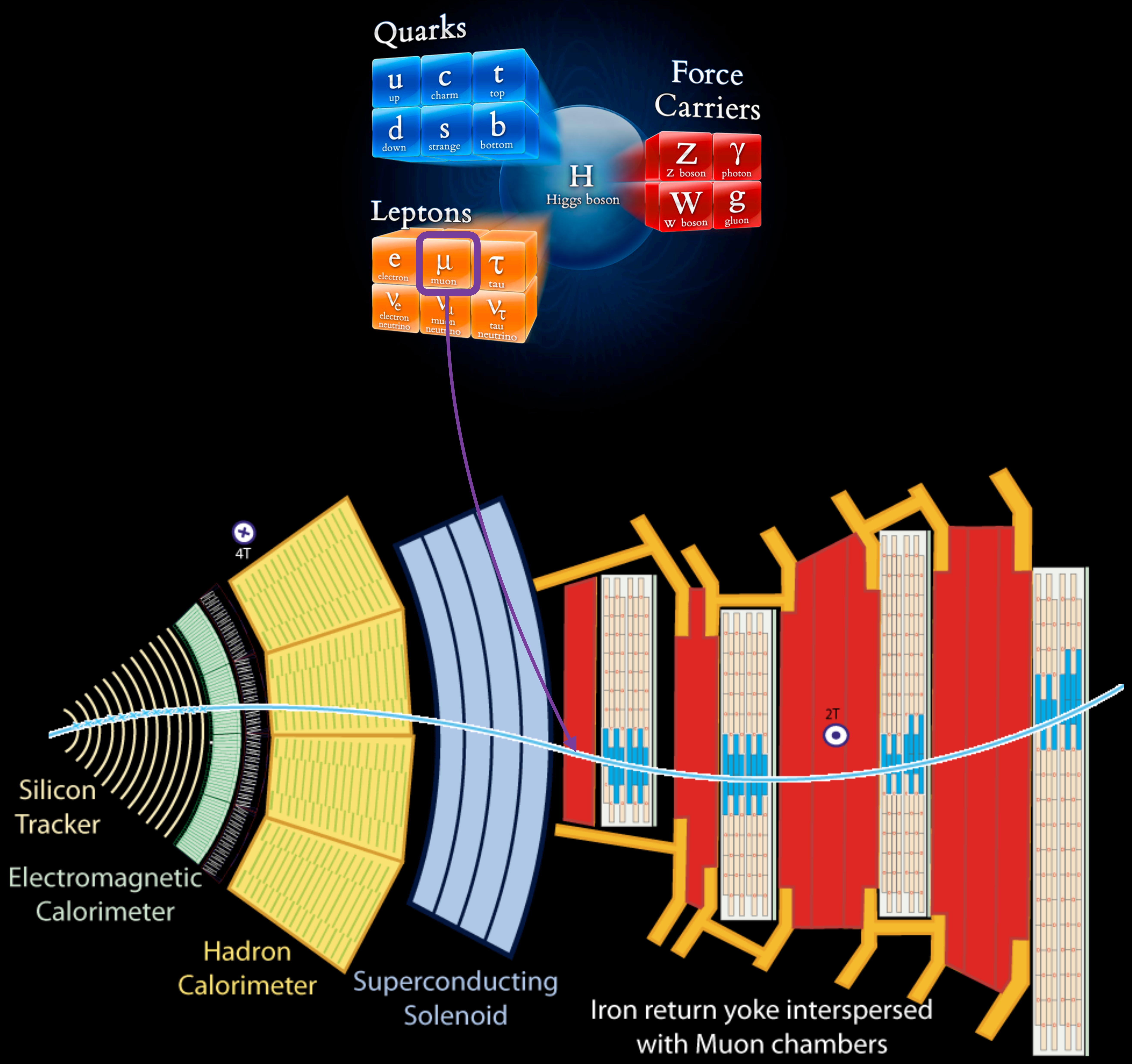
AI?

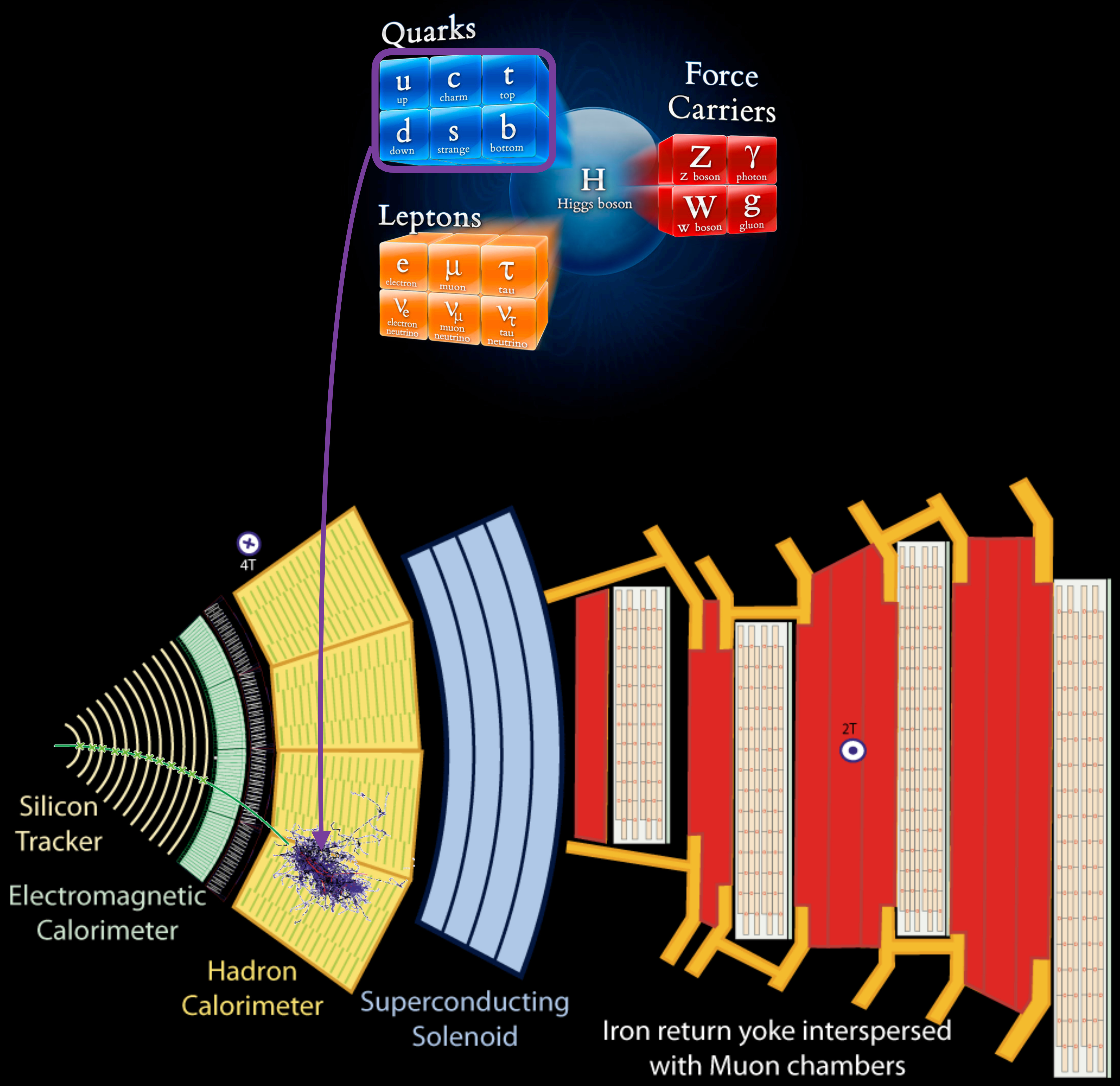






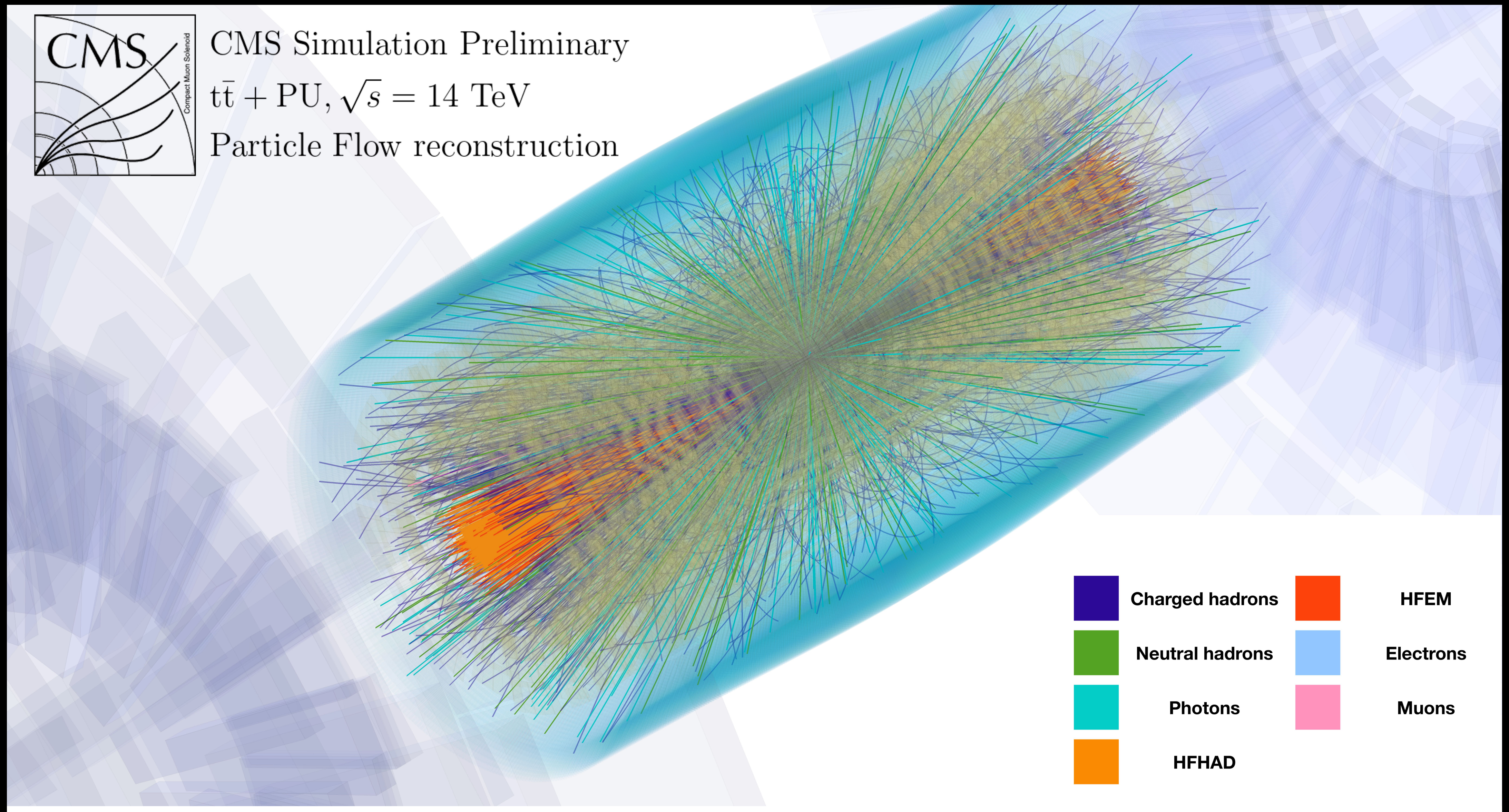
From billions of sensors to particles?


Quarks
u up
c charm
t top
d down
s strange
b bottom
Leptons
e electron
μ muon
τ tau
ν_e electron neutrino
ν_μ muon neutrino
ν_τ tau neutrino
H Higgs boson
Force Carriers
Z Z boson
γ photon
W W boson
g gluon
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers

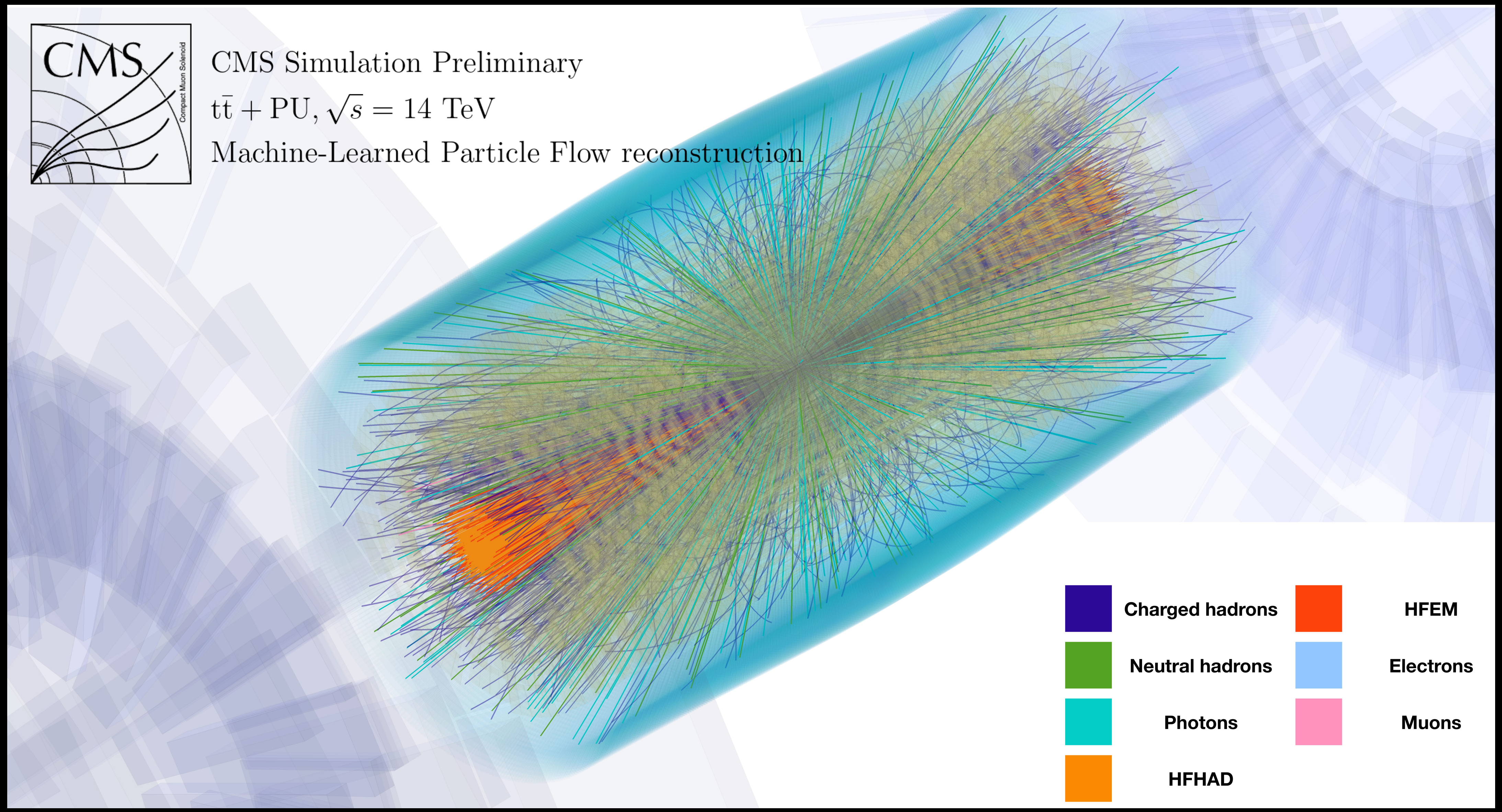

Quarks
u up
c charm
t top
d down
s strange
b bottom
Force Carriers
Z Z boson
γ photon
W W boson
g gluon
H Higgs boson
Leptons
e electron
μ muon
τ tau
νe electron neutrino
νμ muon neutrino
ντ tau neutrino
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers

arxiv:2309.06782

Classical Particle Flow

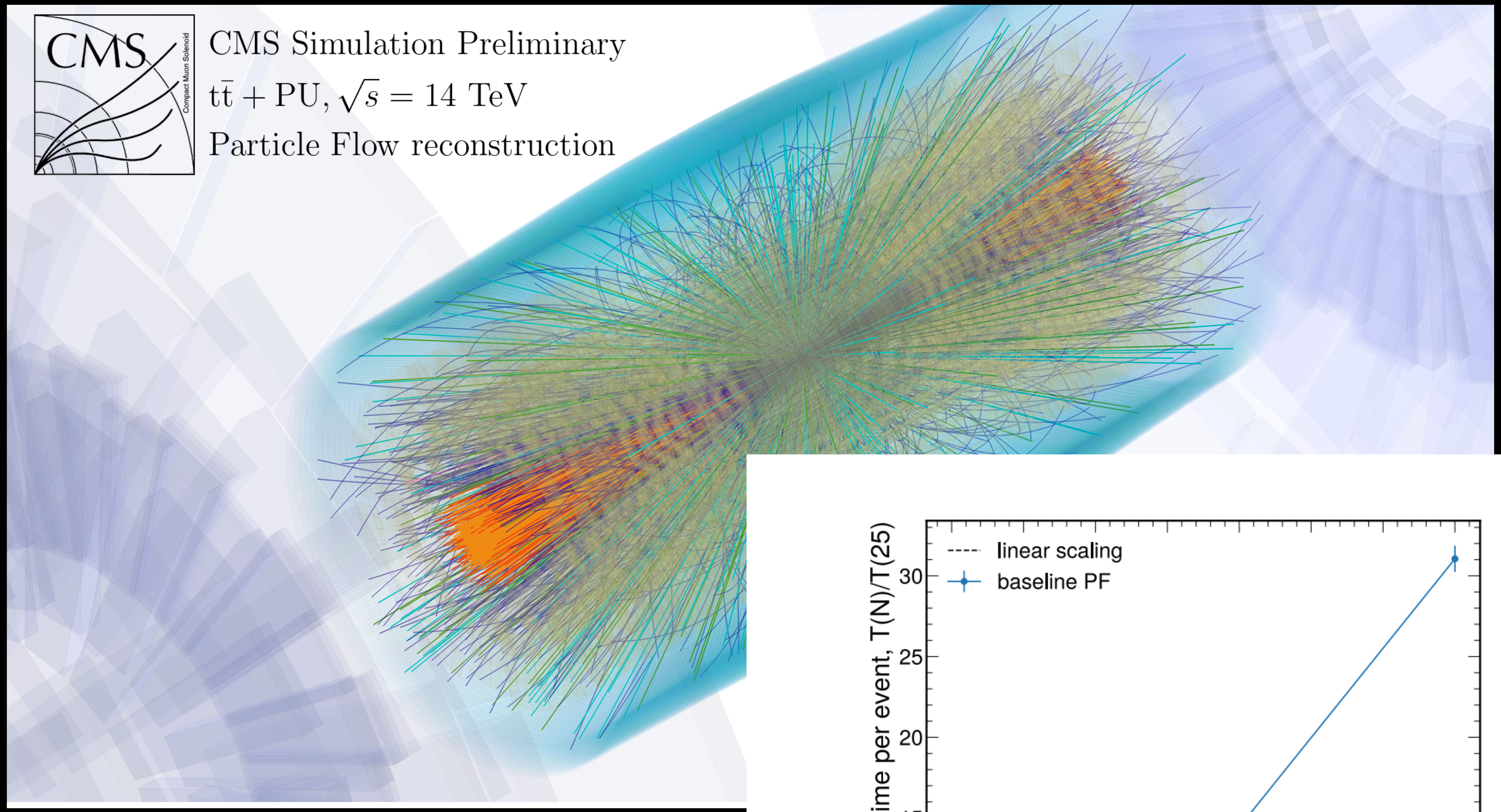


Graph Neural Network

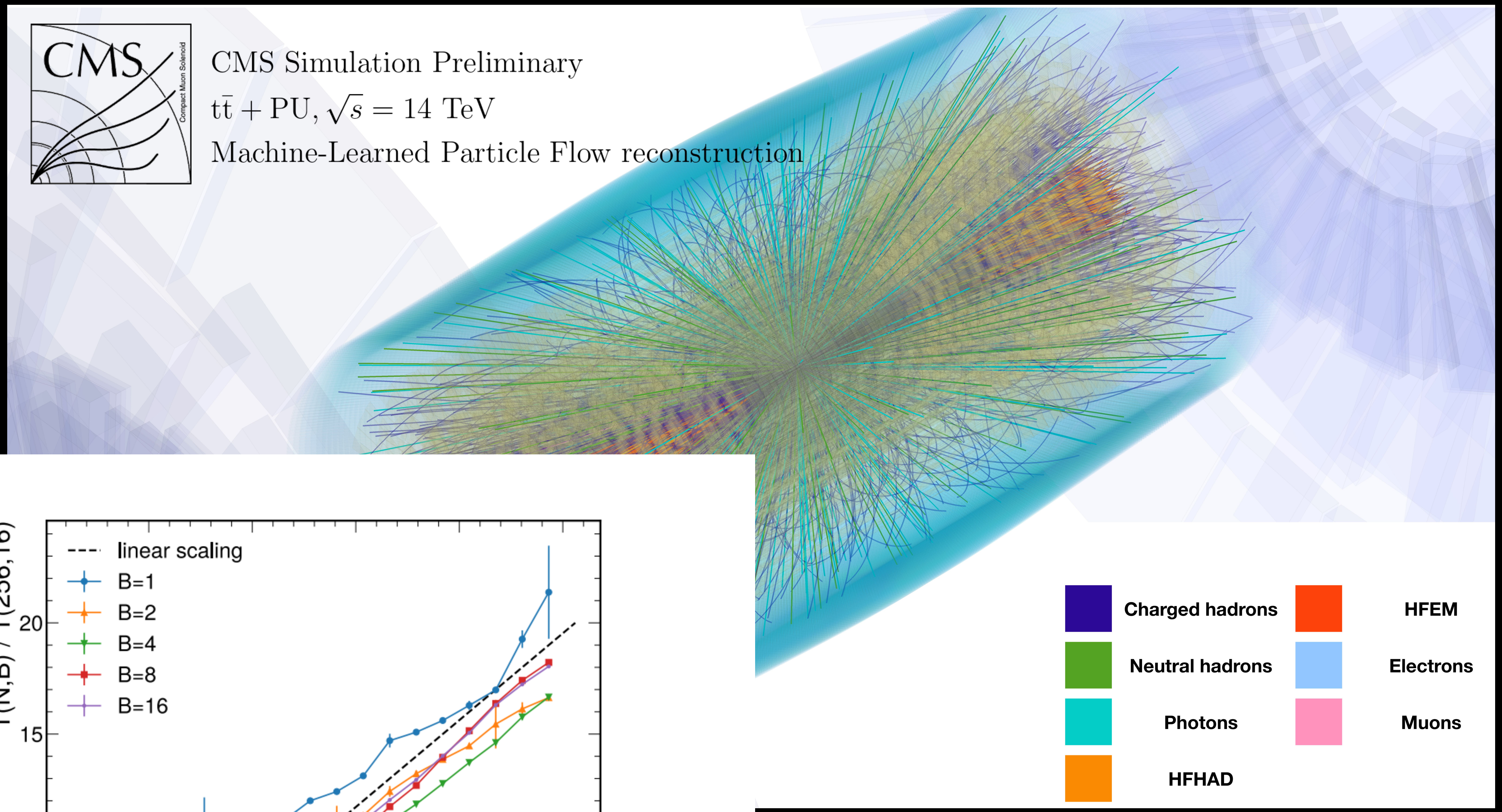


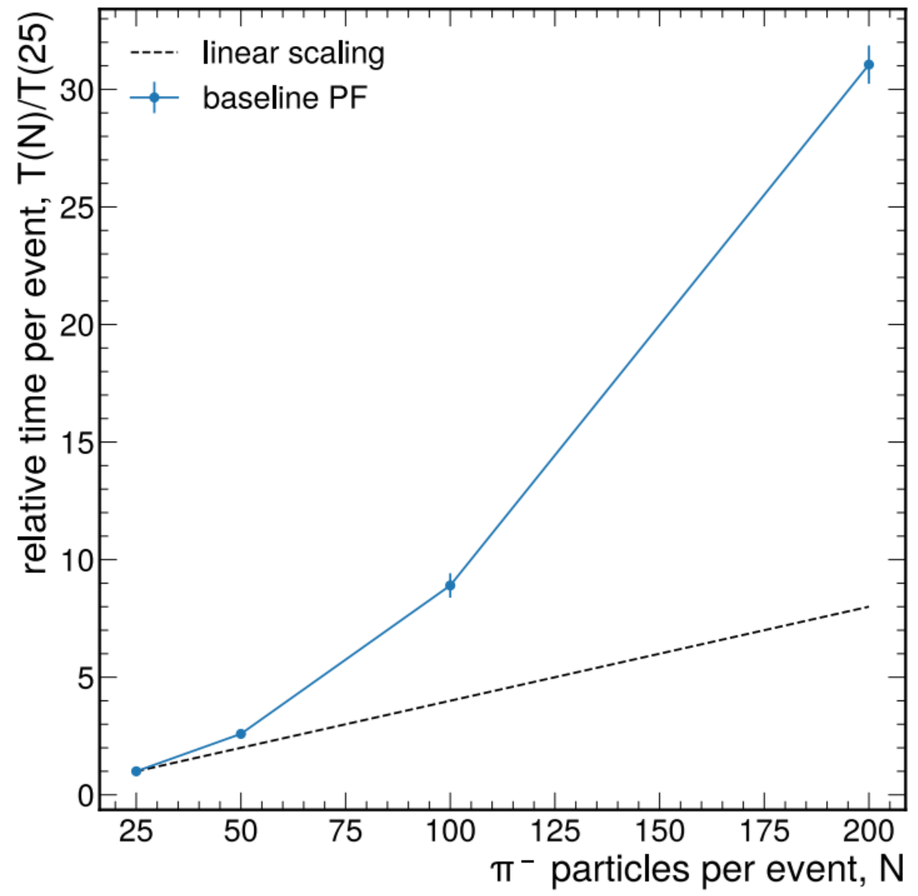
arxiv:2309.06782

Classical Particle Flow

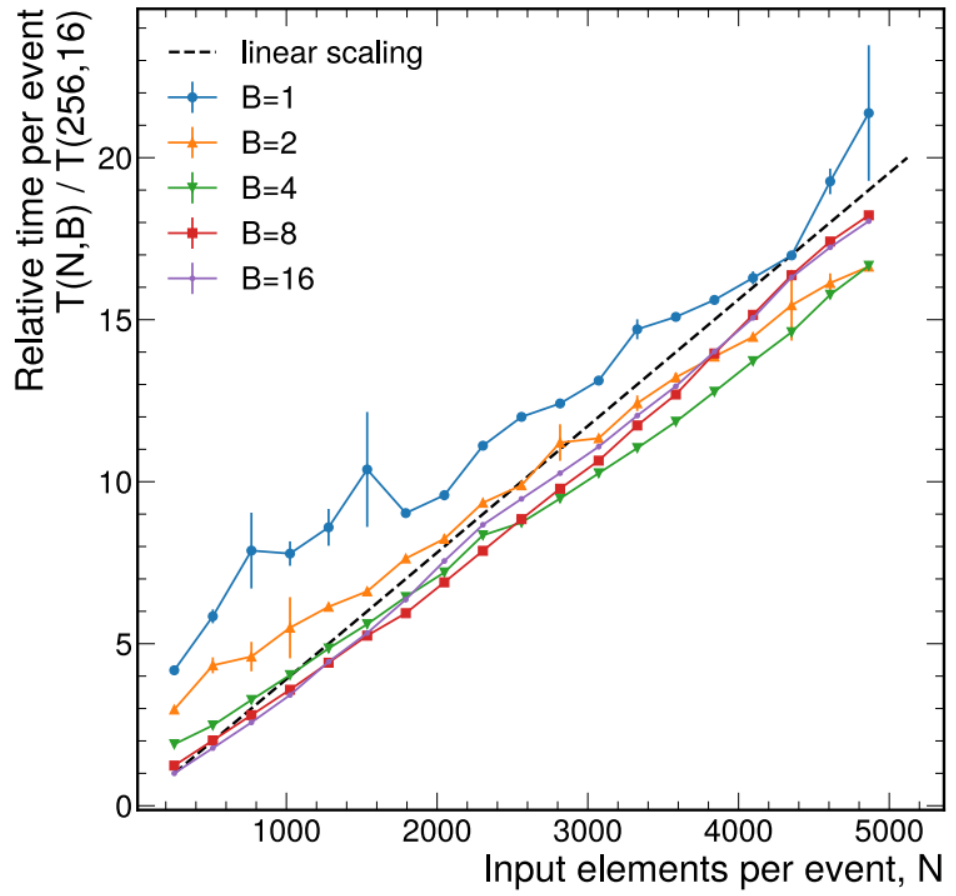


Graph Neural Network





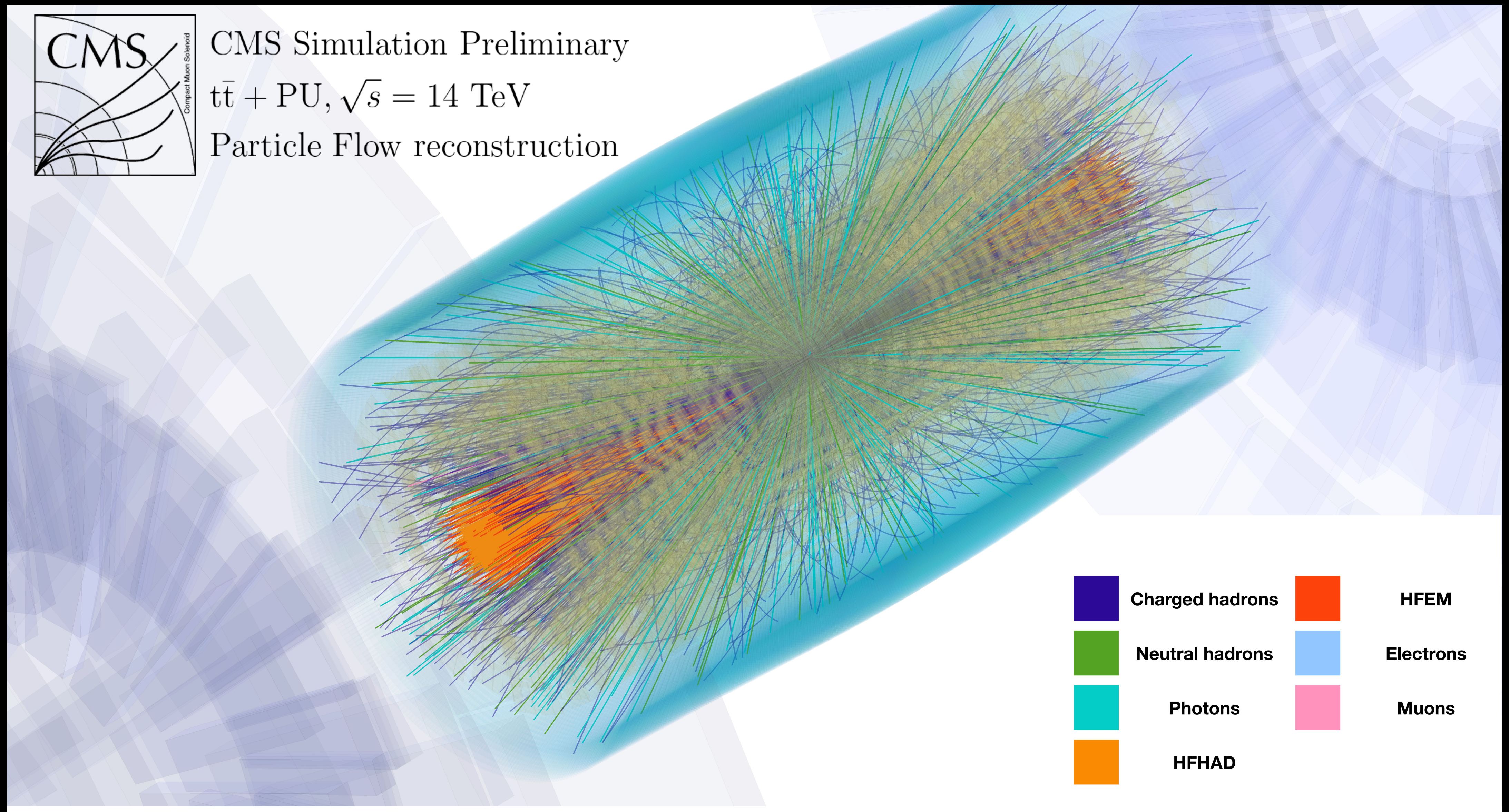
PF baseline scales non-linearly with increasing input size

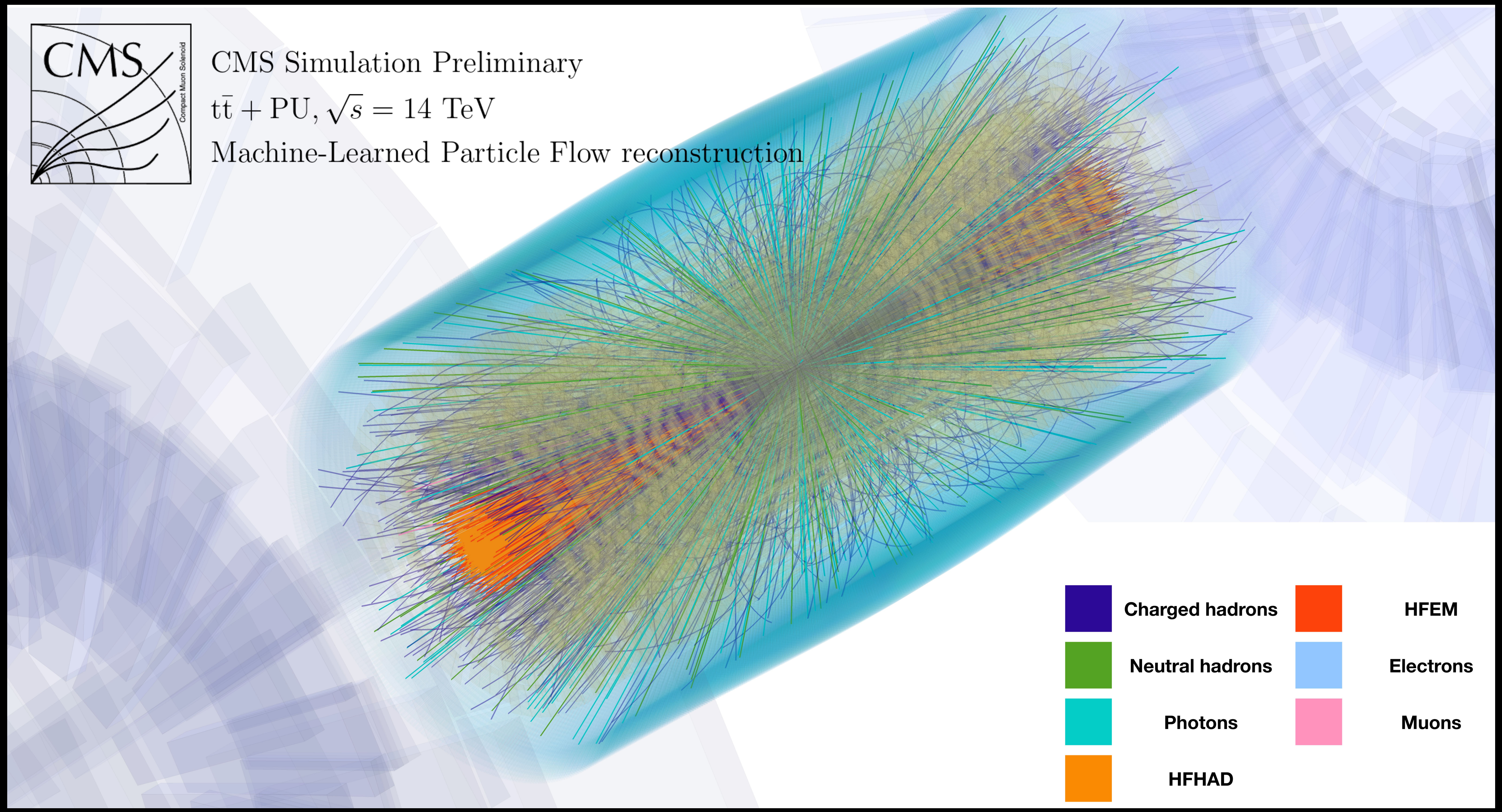


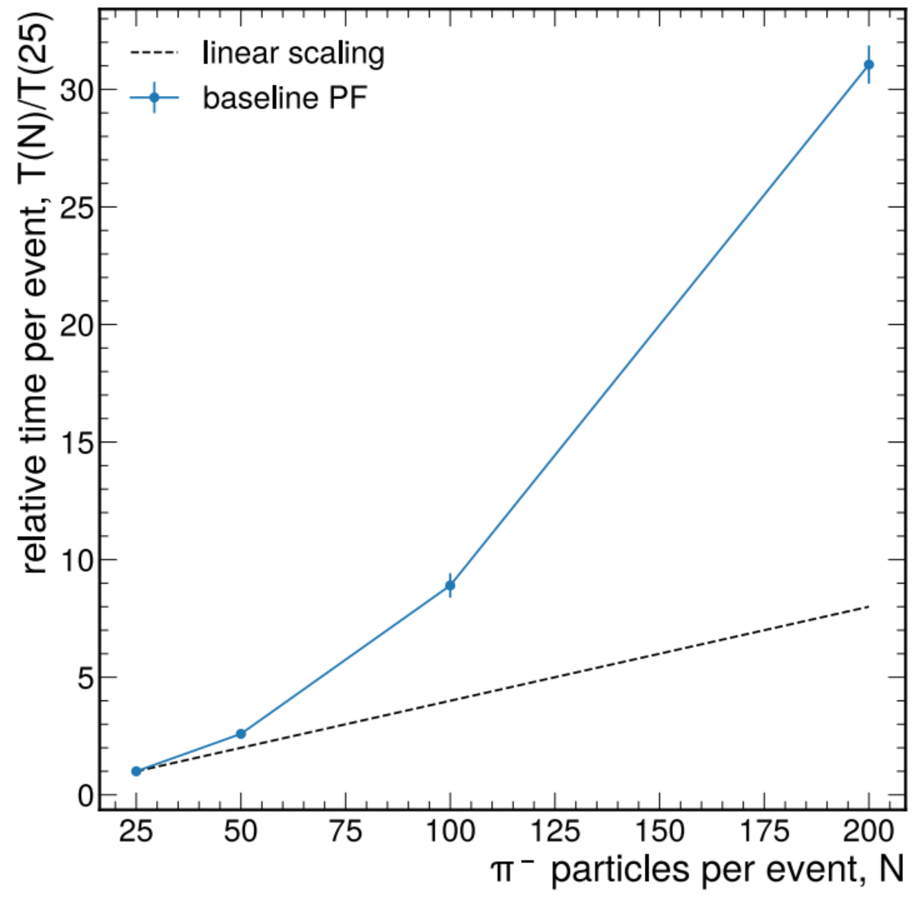
GNN-based model inference time scales approximately linearly with increasing input size

Classical Particle Flow | Graph Neural Network

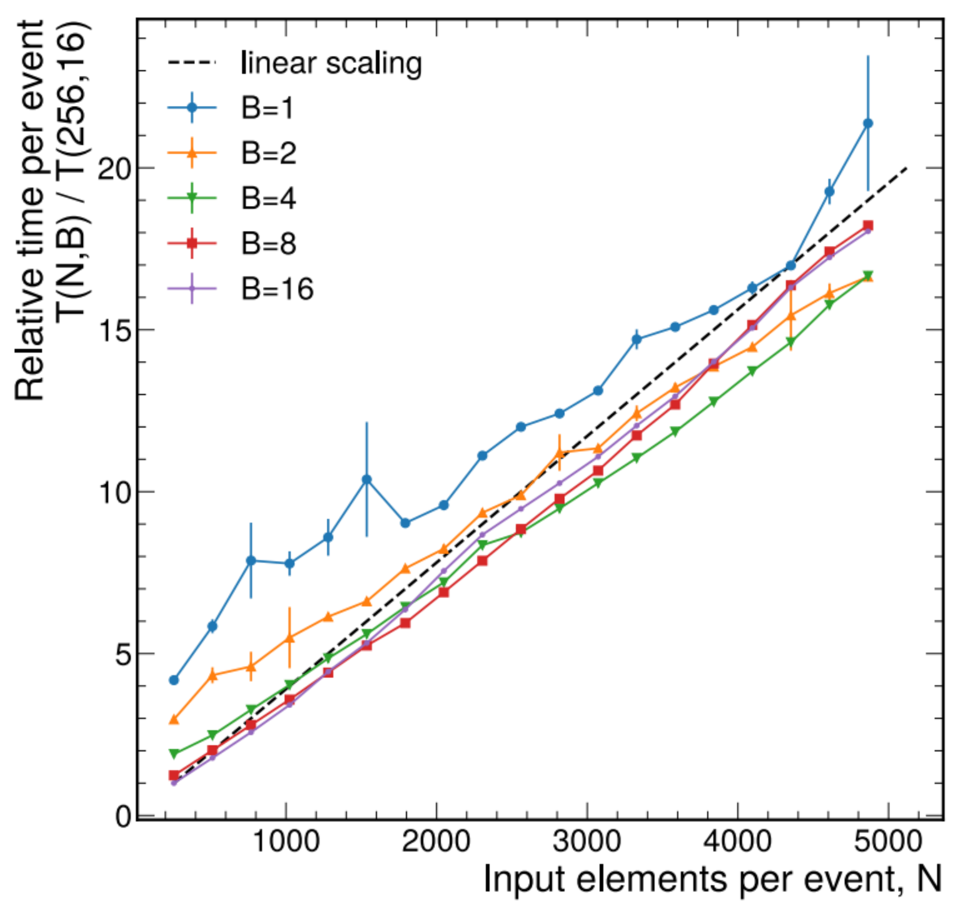
arxiv:2309.06782







PF baseline scales non-linearly with increasing input size

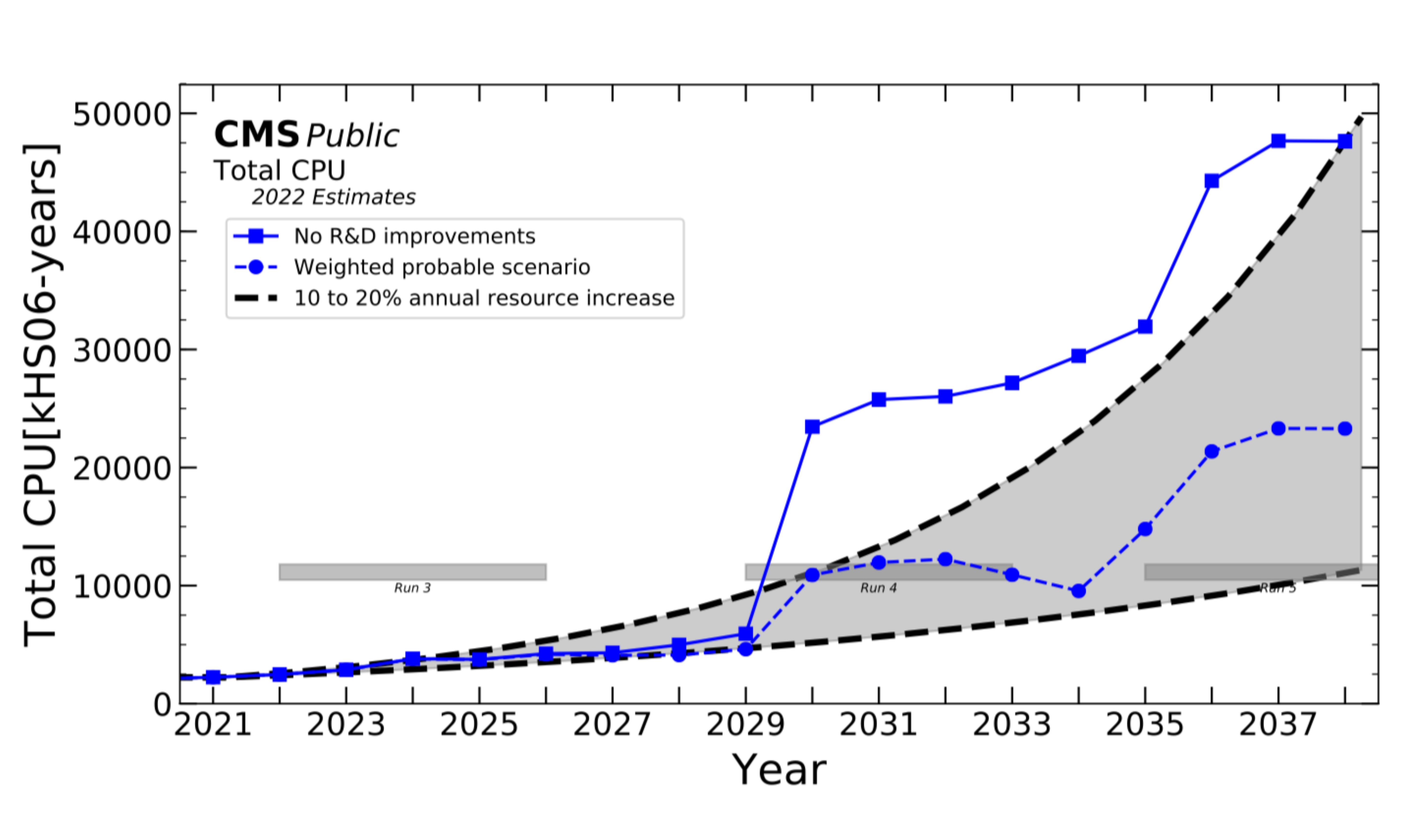


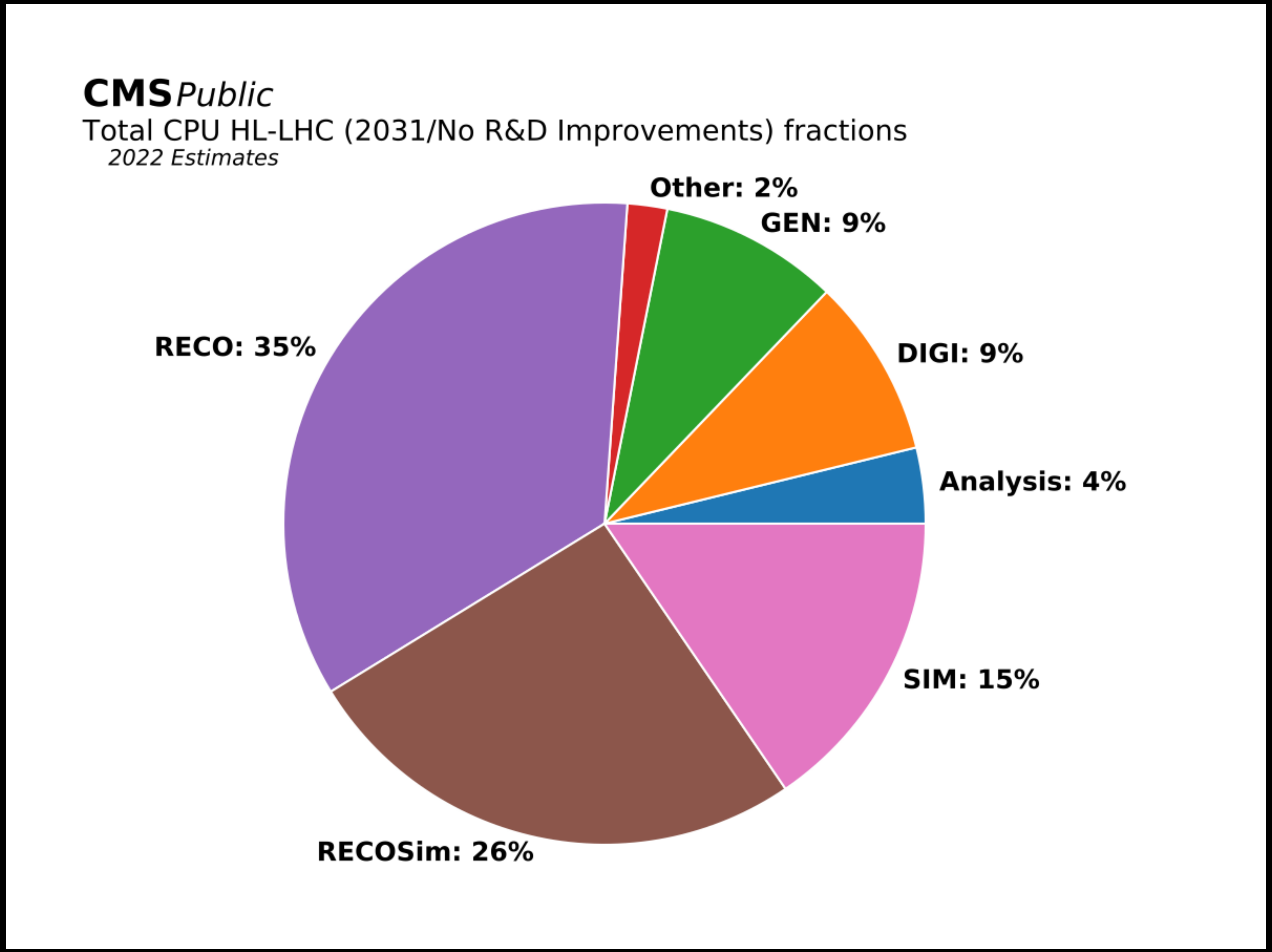
GNN-based model inference time scales approximately linearly with increasing input size

AI?

Generative models for simulation

60% of CPU used for simulation!

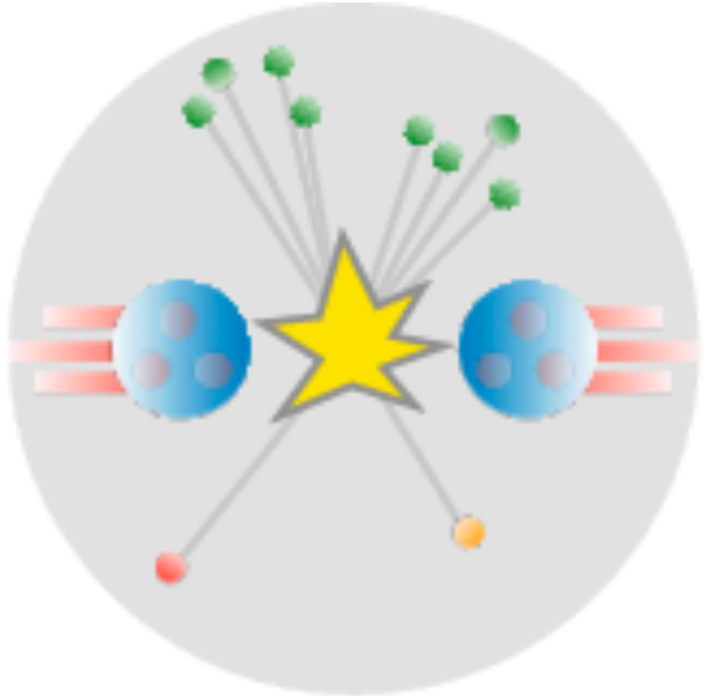




$\mathcal{O}(10)$ $\mathcal{O}(10^3)$ $\mathcal{O}(10^{10})$

10^{-18}m 10^{-15}m 10^{-6}m 100m

GEN



pp collisions up to production of stable particles [Easy & Fast]

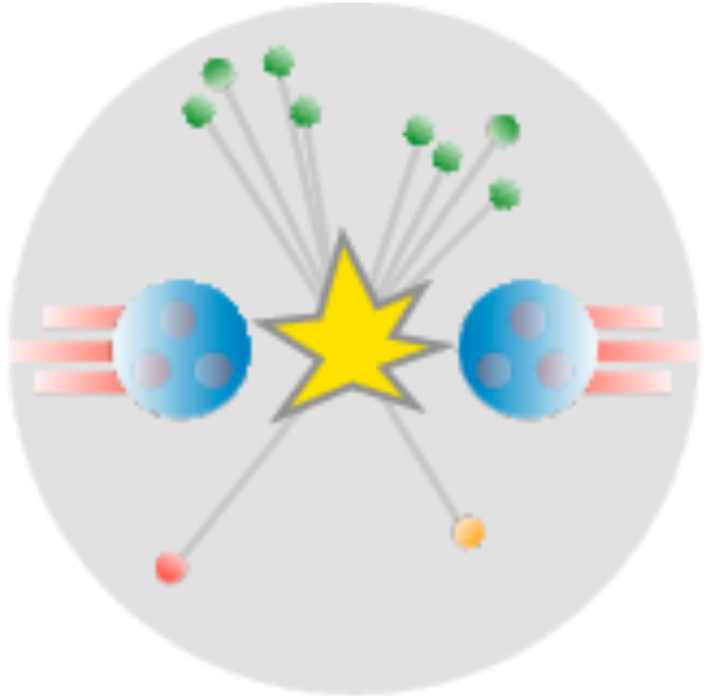
SIM

detector response simulation [Hard & Slow]

DIGI+RECO

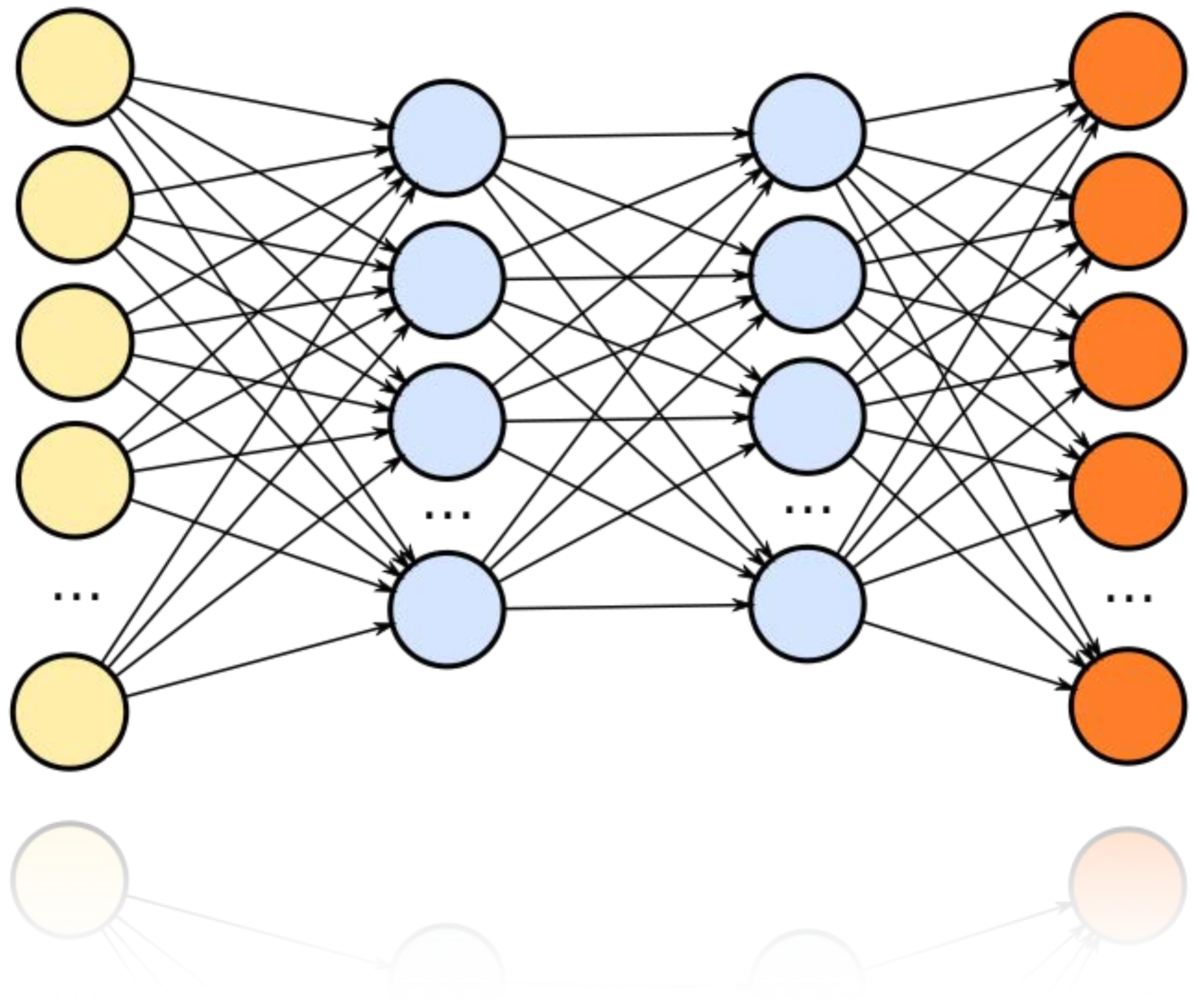
Energy deposits→digital signals→reconstructed by the reconstruction software [Hard & Slow]

GEN



pp collisions up to production of stable particles [Easy & Fast]

SIM

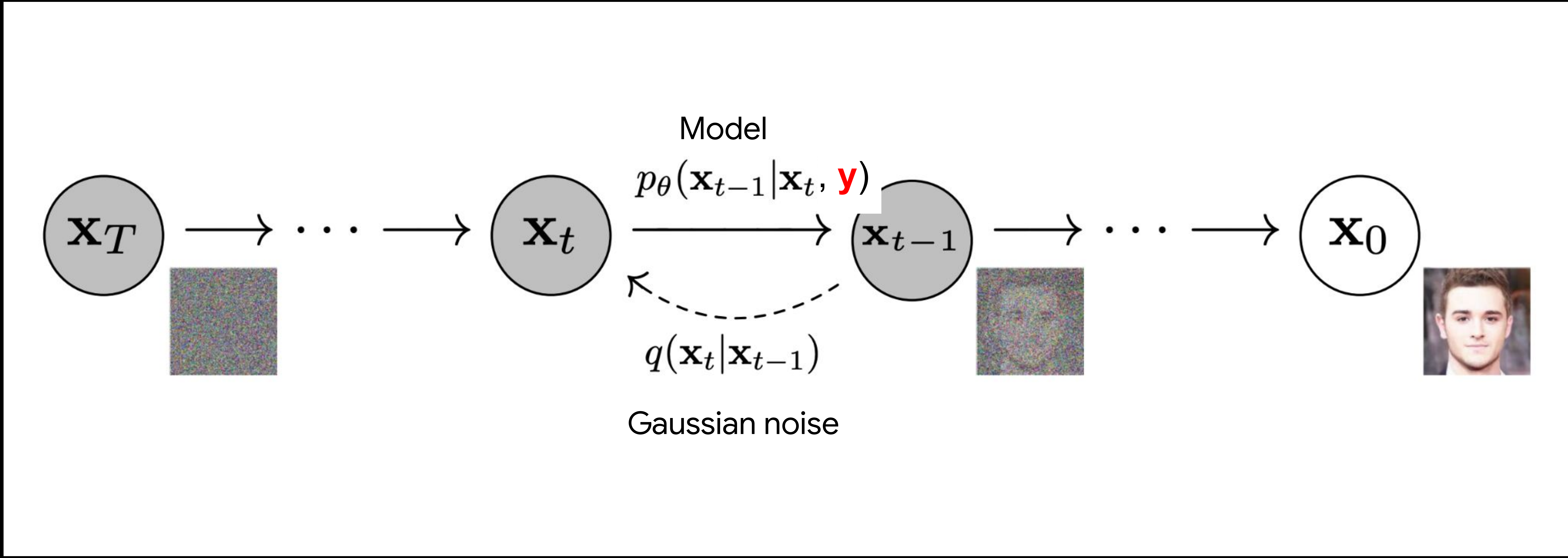


DIGI+RECO

Energy deposits→digital signals→reconstructed by the reconstruction software [Hard & Slow]


$O(10)$
$O(10^3)$
$O(10^{10})$
10^{-18}m
10^{-15}m
10^{-6}m
100m
GEANT4
10^{10} events
SLOW but ACCURATE
GEANT4
10^5 events
Surrogate model
10^{10} events
(GAN, VAE, Normalizing Flow, …)
Learn underlying distribution of GEANT4 events
FAST and ACCURATE?
ML methods can provide fast and accurate “surrogate models” for GEANT4 etc

Diffusion models

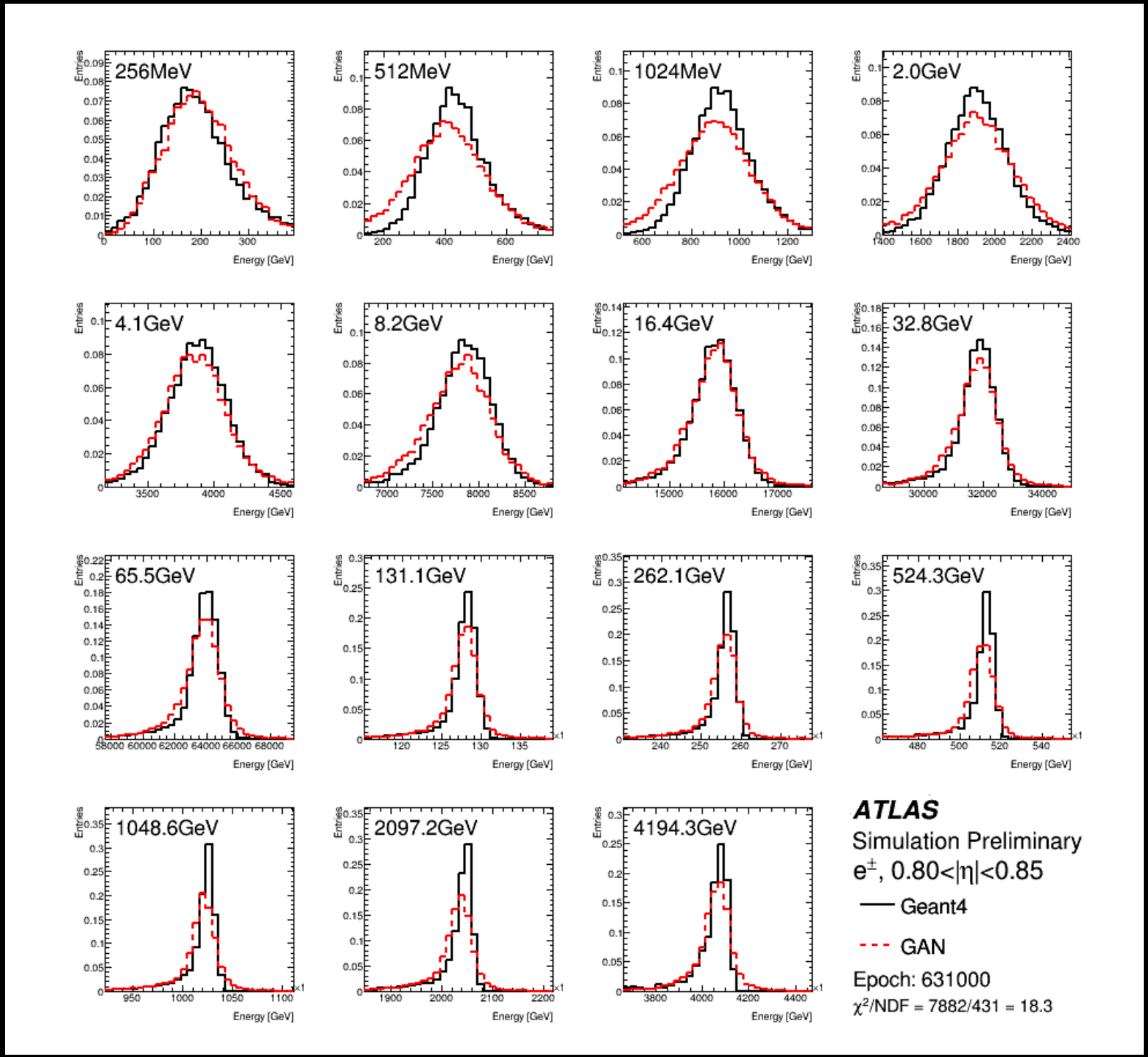


Learn systematic decay of information due to noise, then reverse process and recover the information back from the noise.

FastCaloGAN Being used in ATLAS!

100 networks (slices in η)

O(500) voxels



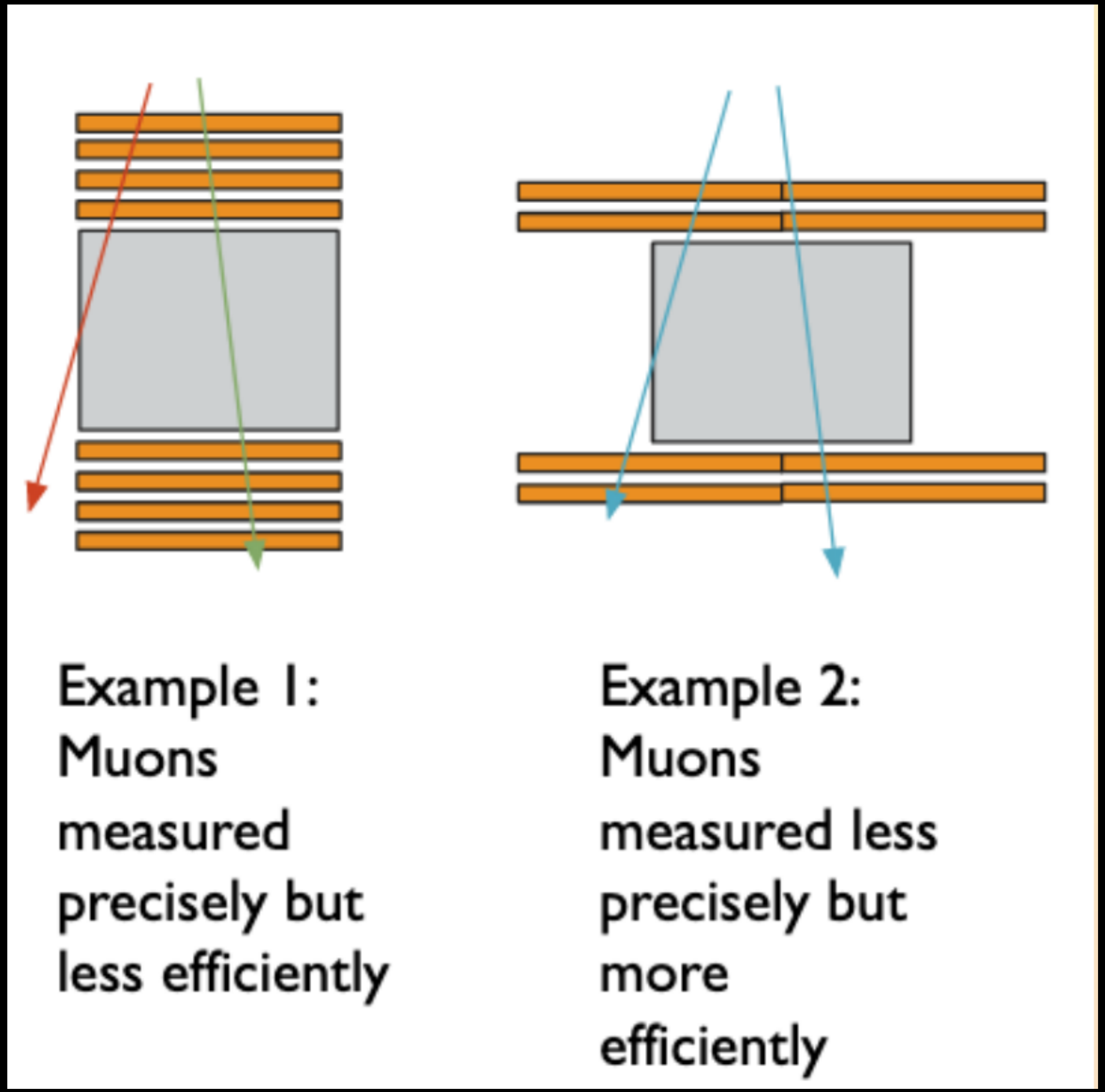
ATL-SOFT-PUB-2020-006; ATLAS 2109.02551

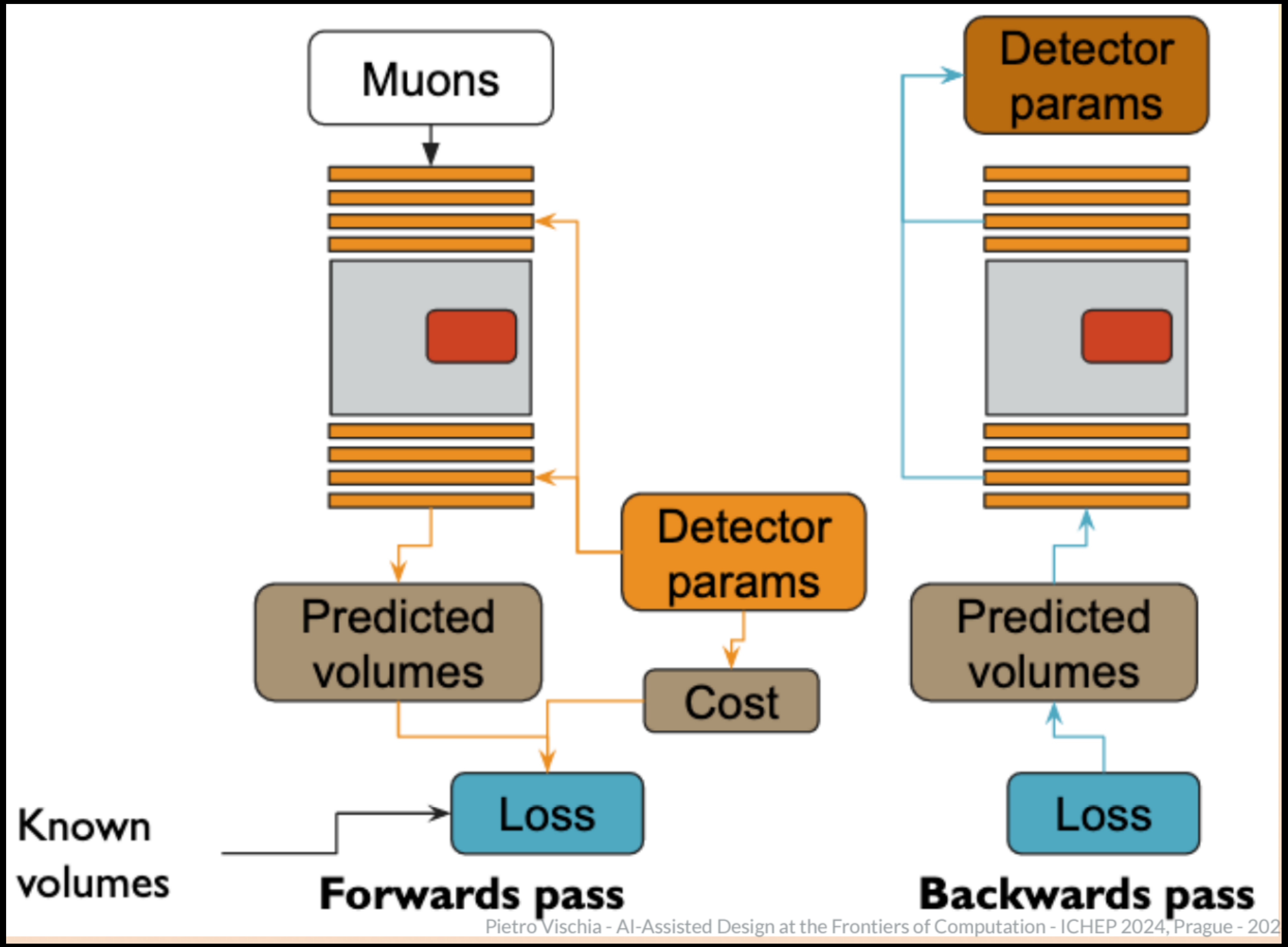
Detector design, data acquisition and triggering

AI-assisted design of experiments

doi:10.1016/j.revip.2023.100085

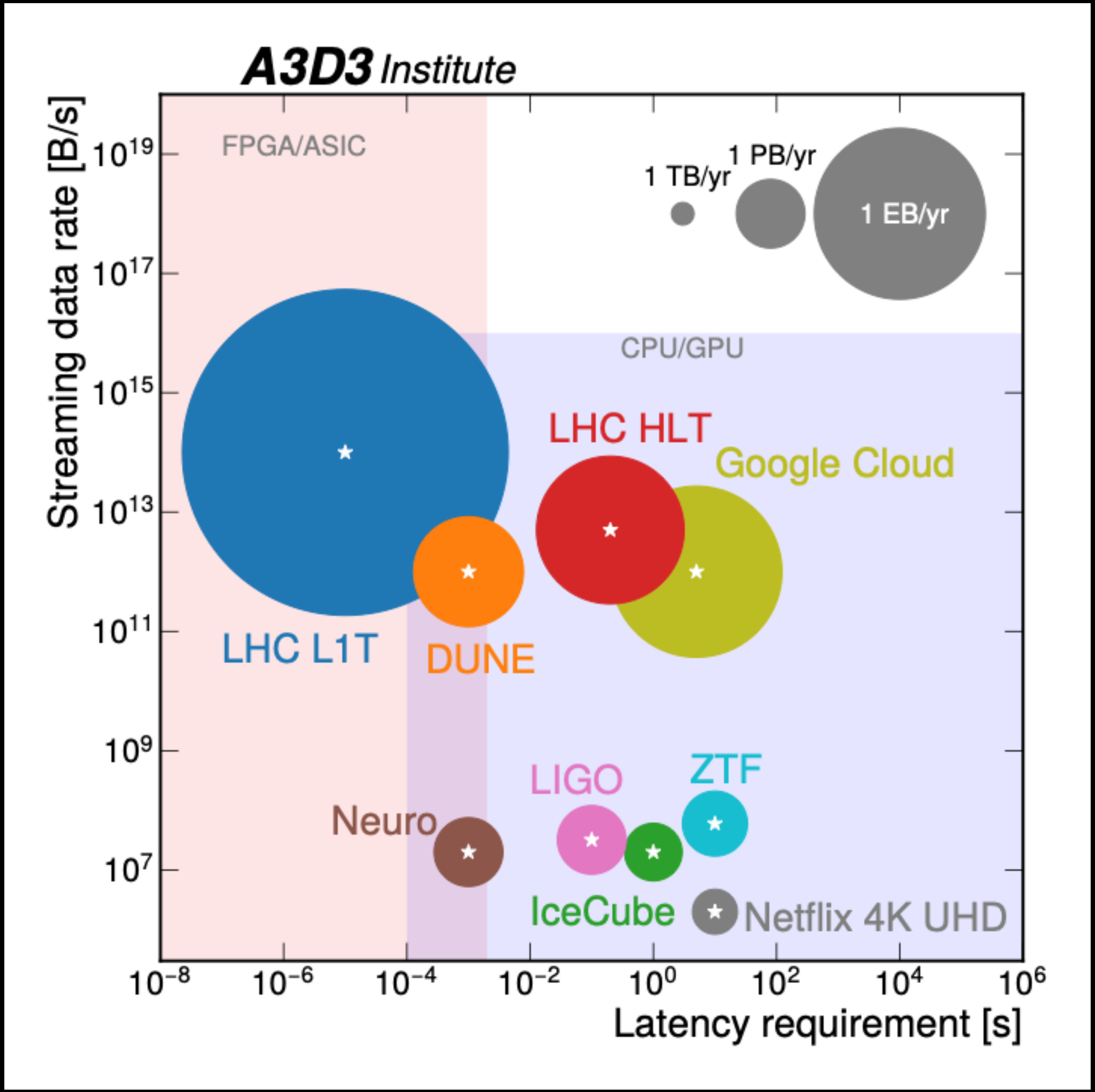
- Make everything differentiable!
- Joint optimization of design parameters w.r.t. inference made with data



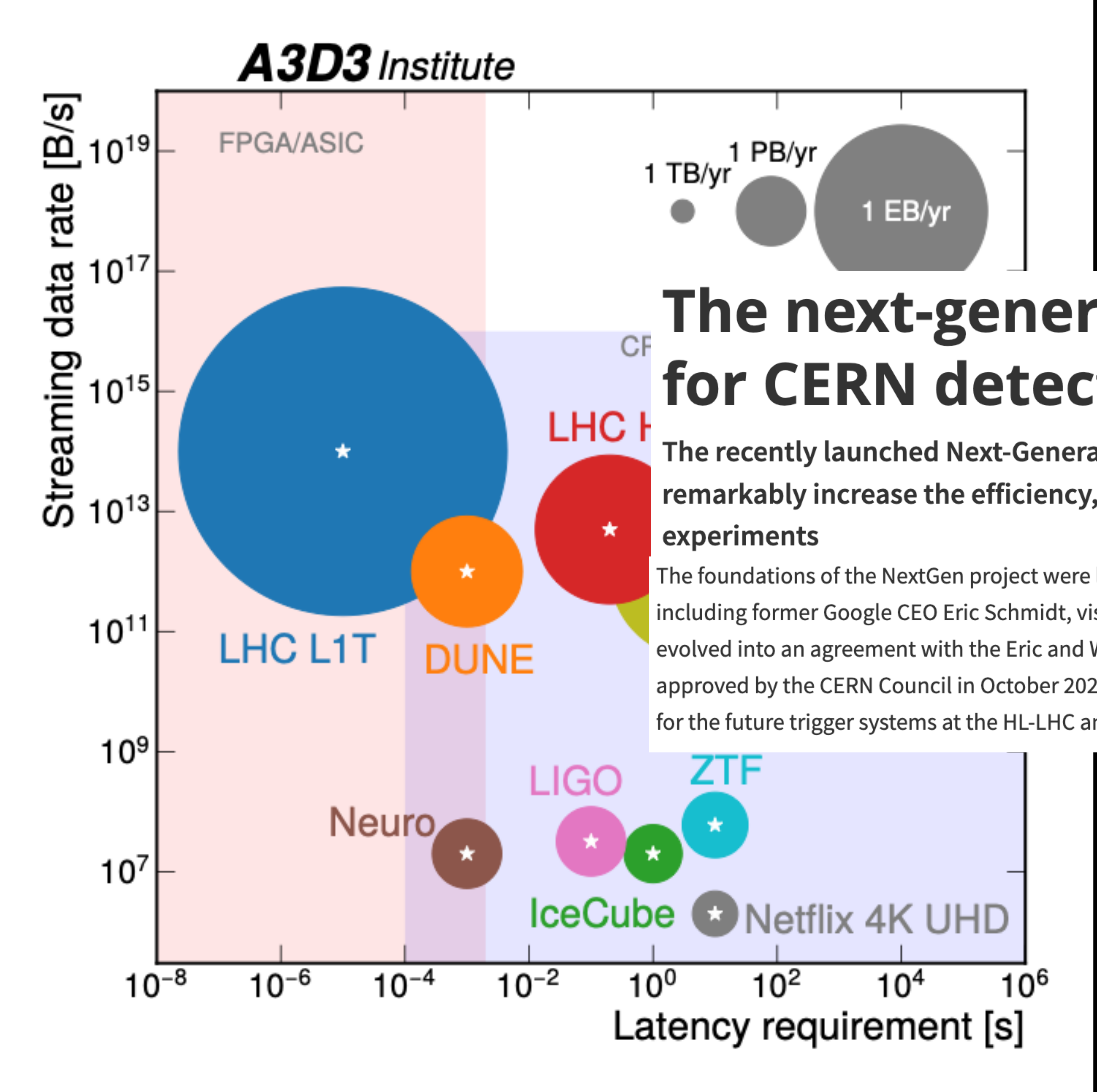


Detector design, data acquisition and triggering


CMS Experiment at the LHC, CERN
Data recorded: 2010-Nov-14 18:37:44.420271 GMT(19:37:44 CEST)
Run / Event: 151076 / 1405388
1 billion collisions /s
~1 MB of data / collision
~1 PB of data / s



A3D3 Institute
Streaming data rate [B/s]
Latency requirement [s]
FPGA/ASIC
CPU/GPU
1 TB/yr
1 PB/yr
1 EB/yr
LHC L1T
DUNE
LHC HLT
Google Cloud
Neuro
LIGO
IceCube
ZTF
Netflix 4K UHD
10^{19}
10^{17}
10^{15}
10^{13}
10^{11}
10^{9}
10^{7}
10^{-8}
10^{-6}
10^{-4}
10^{-2}
10^{0}
10^{2}
10^{4}
10^{6}



The next-generation triggers for CERN detectors

The recently launched Next-Generation Triggers project is set to remarkably increase the efficiency, sensitivity and modelling of CERN experiments

The foundations of the NextGen project were laid in 2022 when a group of private donors, including former Google CEO Eric Schmidt, visited CERN. This first inspiring visit eventually evolved into an agreement with the Eric and Wendy Schmidt Fund for Strategic Innovation, approved by the CERN Council in October 2023, to fund a project that would pave the way for the future trigger systems at the HL-LHC and beyond: NextGen was born.


Geneva
Geneva Lake
LHCb
ATLAS
ALICE
CMS
1 PB of data / second

Geneva
Geneva Lake
LHCb
CMS
Data temporarily stored
in detector electronics for 12(4) μs
LHC


Geneva
Geneva Lake
LHCb
AT
CMS
5% internet traffic to L1
(63 Tb/s HL-LHC)
LHC


Geneva
Geneva Lake
LHCb
CMS
LHC
5% internet traffic to L1
(63 Tb/s HL-LHC)
L1 trigger:
~1000 AMD FPGAs
Decide which event to keep within ~12 μs latency
Discard >99% of collisions!


Geneva
Geneva Lake
LHCb
AT
CMS
LHC
L1 bit:
Accept = 1
Reject = 0


Geneva
Geneva Lake
LHCb
ATLAS
ALICE
CMS
DATA
99.72% of events rejected!
110 kHz
LHC

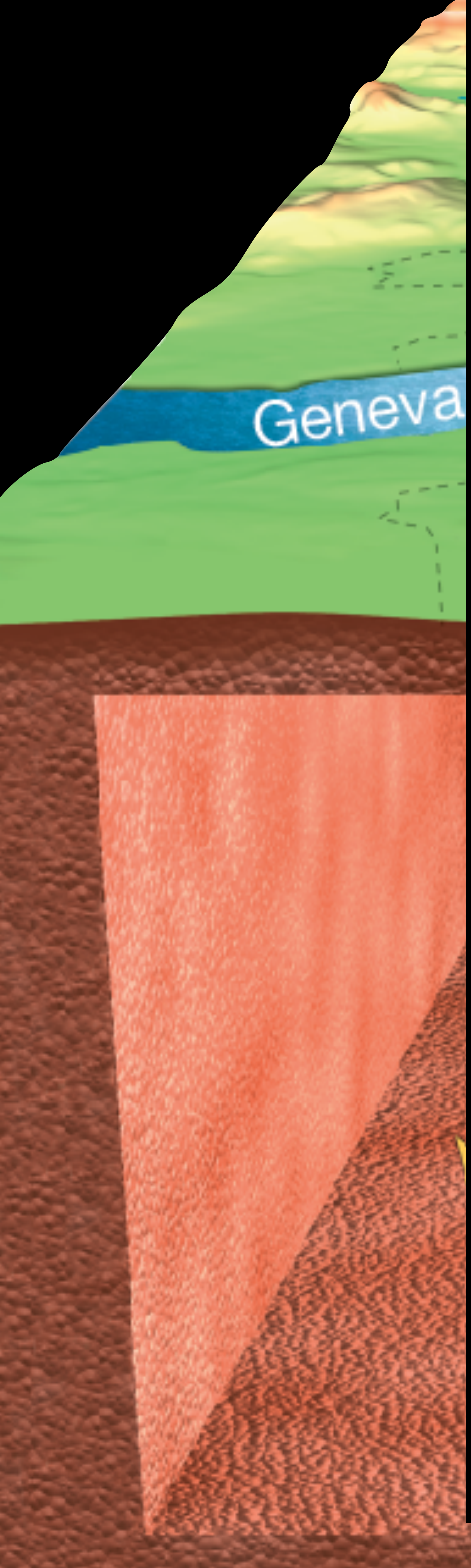

Geneva
Geneva Lake
LHCb
ATLAS
ALICE
CMS
DATA
99.72% of events rejected!
110 kHz
LHC

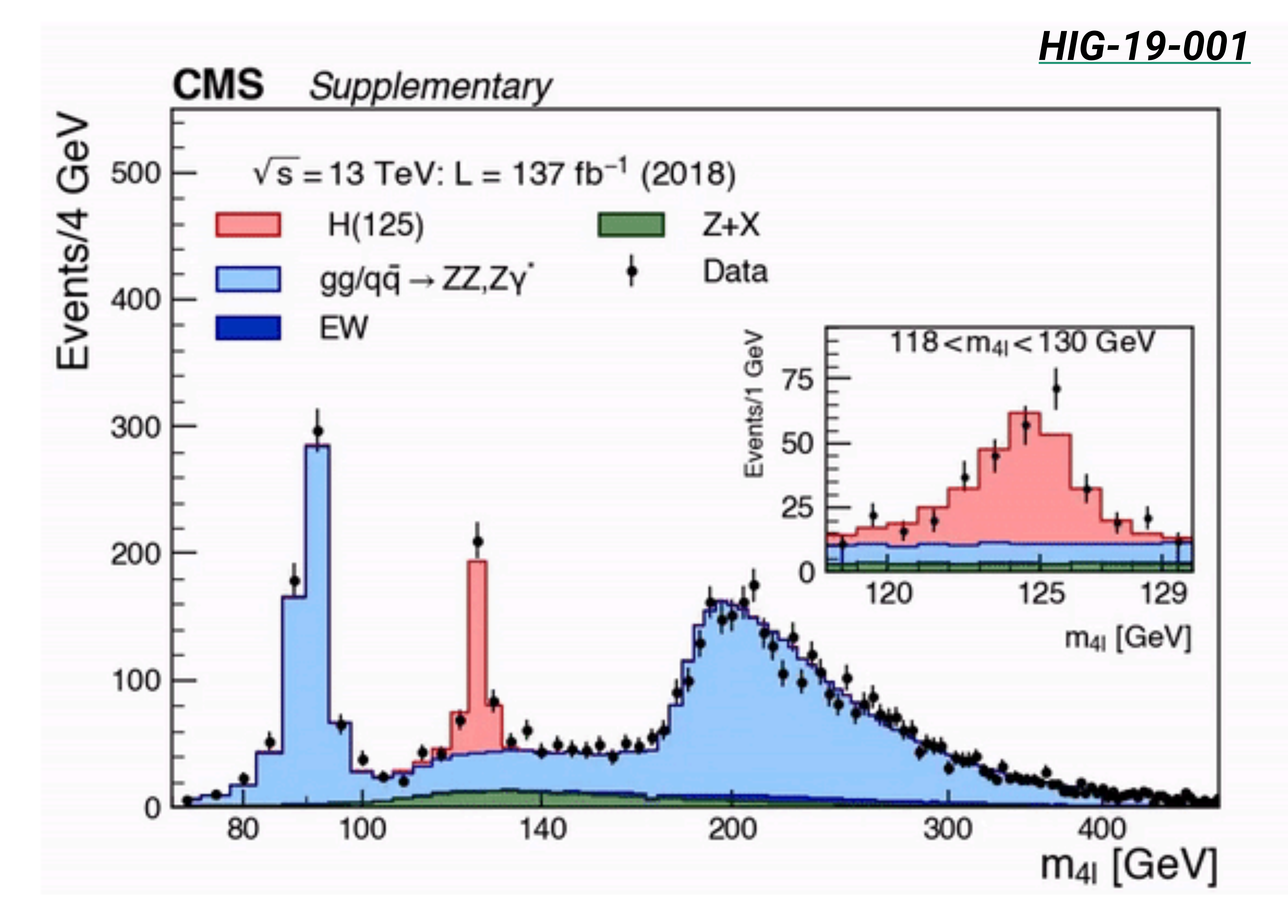

High Level Trigger:
25’600 CPUs / 400 GPUs
Latency: 3-400 ms
Reject further 99%!
eva
LHCb
ATLAS
ALICE
LHC

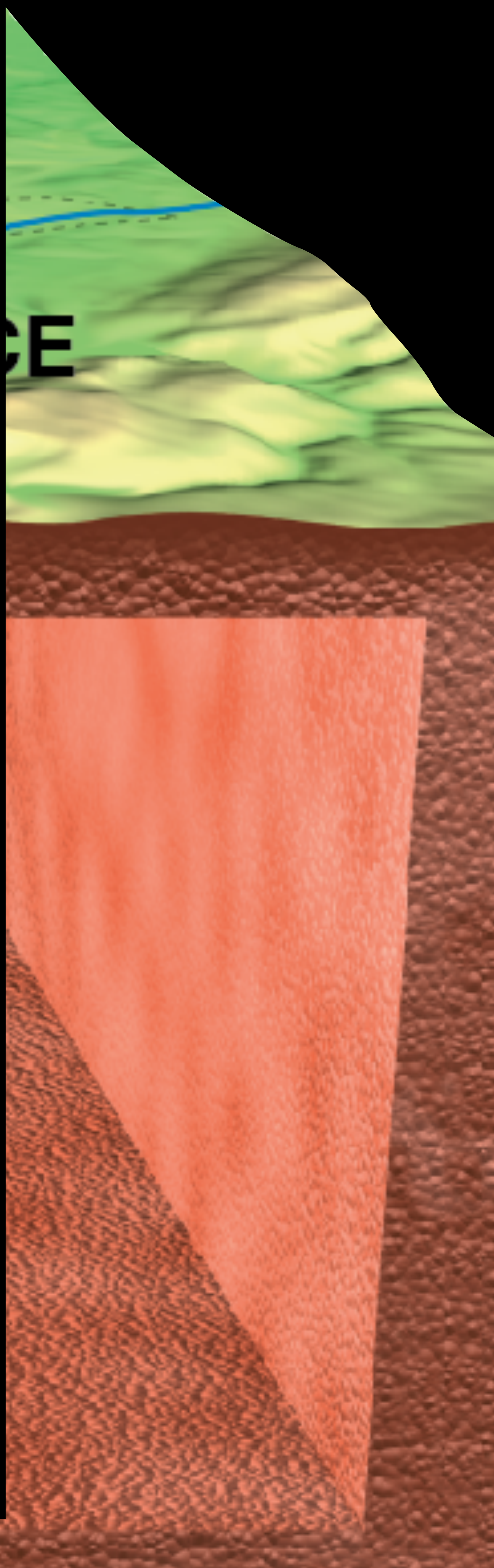

TIER 0: ∞
Geneva
High Level Trigger:
Latency O(100) ms
LHCb
ATLAS
CMS
DATA
99.9975% of events rejected!
1000 events/second
~5 GB/s
LHC


TIER 0: ∞
Geneva
High Level Trigger:
Latency O(100) ms
LHCb
ATLAS
CMS
DATA
99.9975% of events rejected!
1000 events/second
~5 GB/s
LHC


TIER 0: ∞
0.0025% of collision events remaining
Geneva
High Level Trigger:
Latency O(100) ms
LHCb
ATLAS
CMS
DATA
99.9975% of events rejected!
1000 events/second
~5 GB/s
LHC

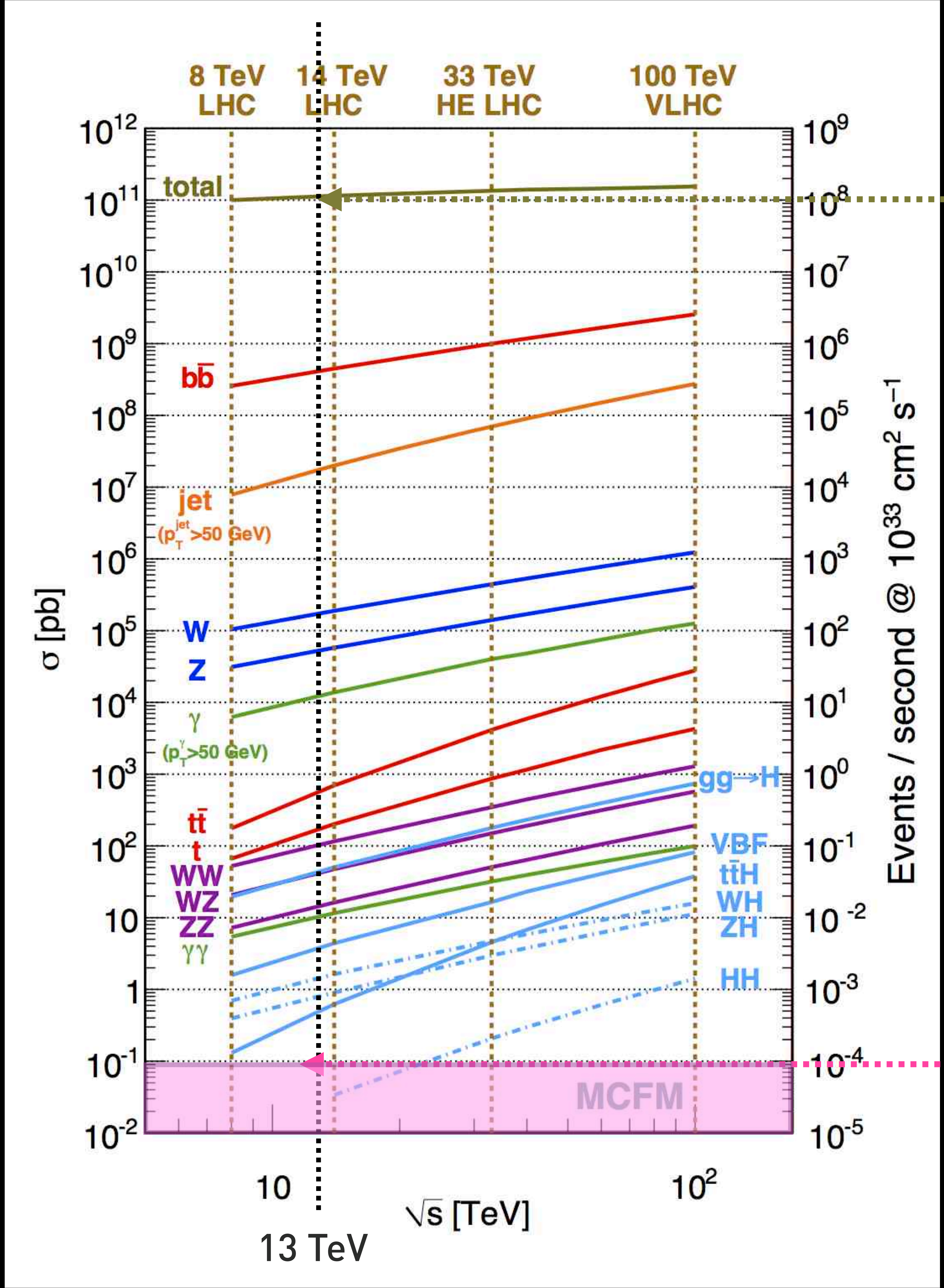






To make sure we select "the right" 0.0025%, algorithms must be

- Fast (get more data through)
- Accurate (select the right data)



"Probability" of producing "anything"

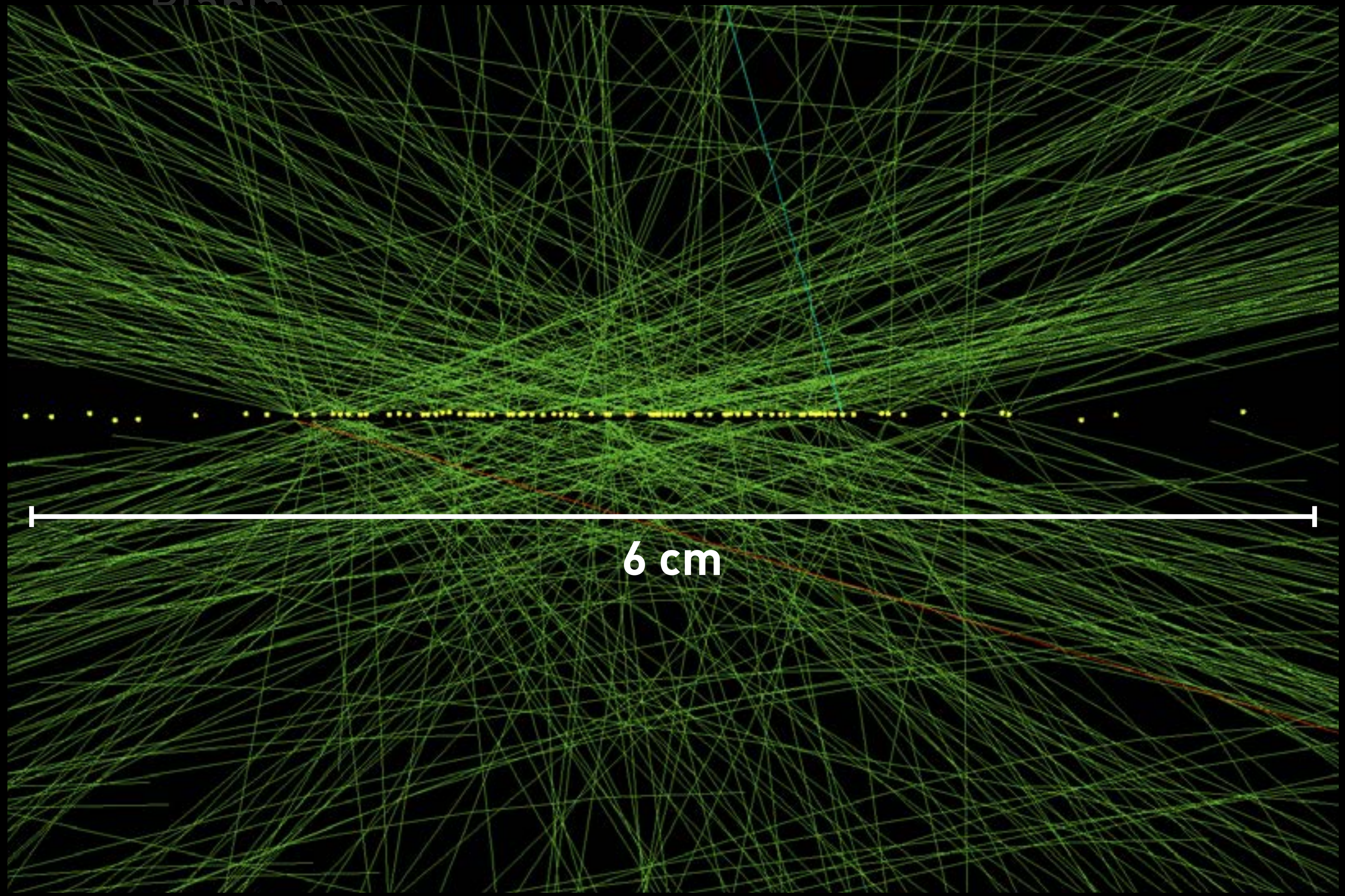
New Physics?

New Physics is produced less than 1 in a trillion (if at all)

Need <u>more</u> data!

LHC

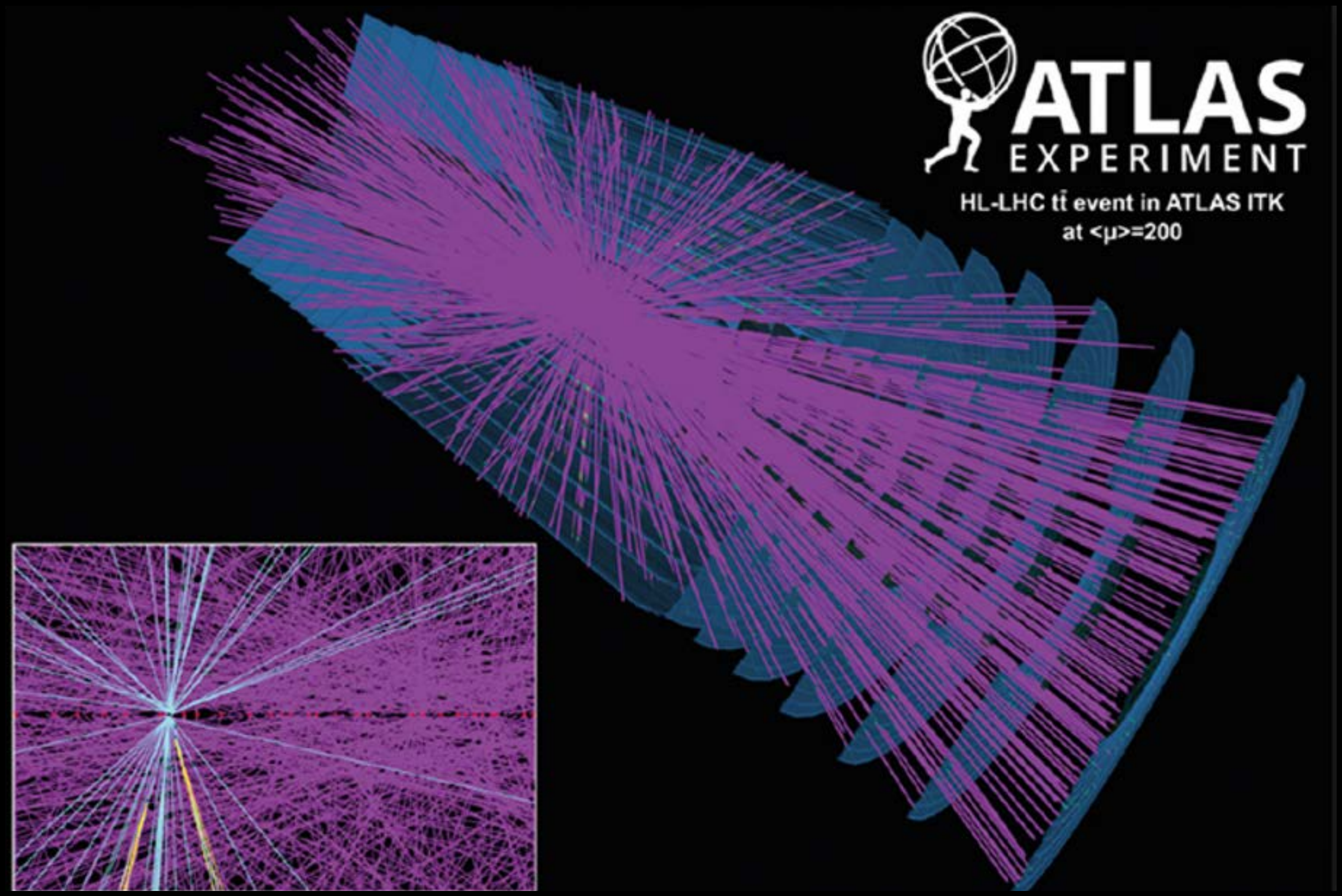
78 vertices
(average 60)



Run 3

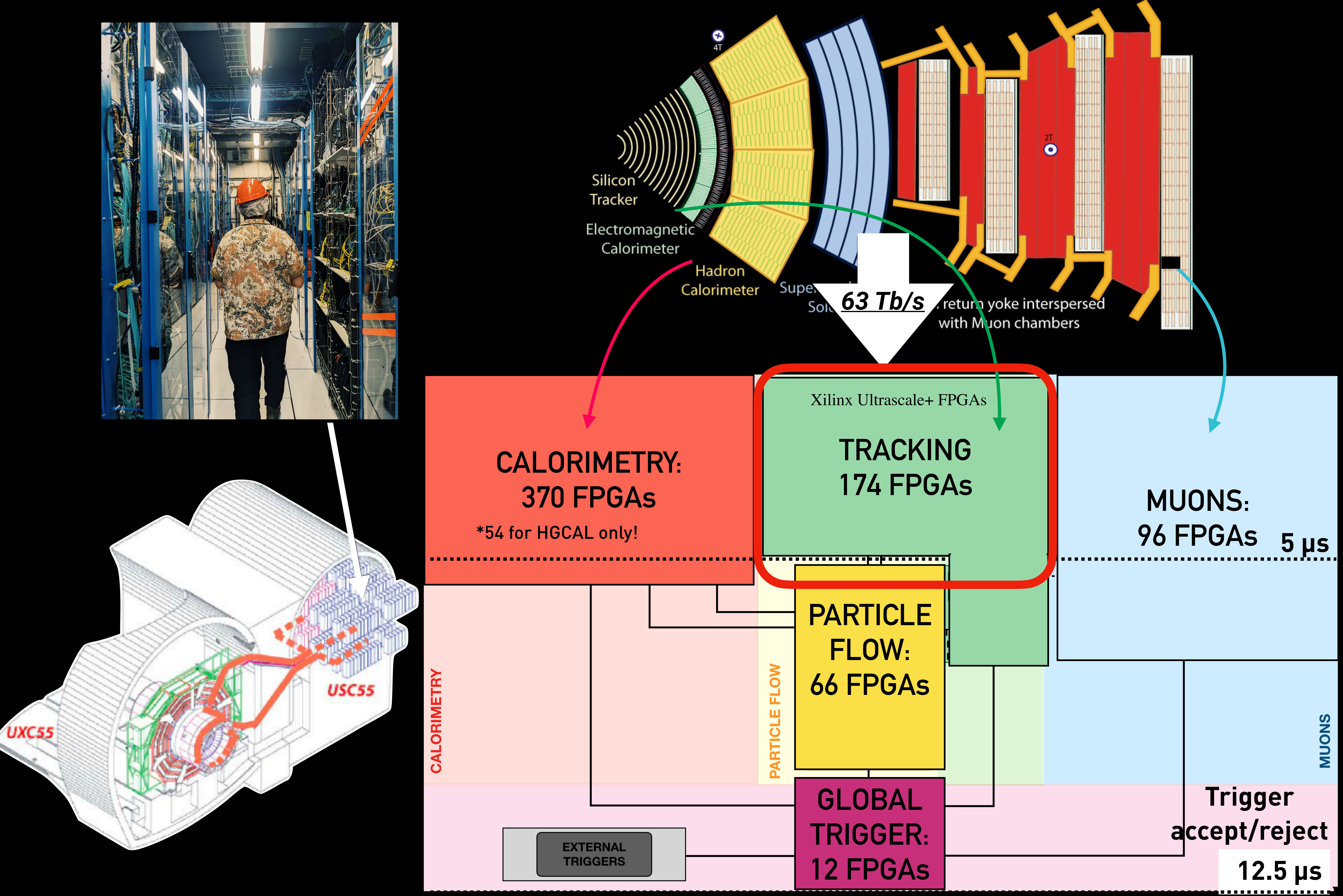
High Luminosity LHC

200 vertices
(average 140)

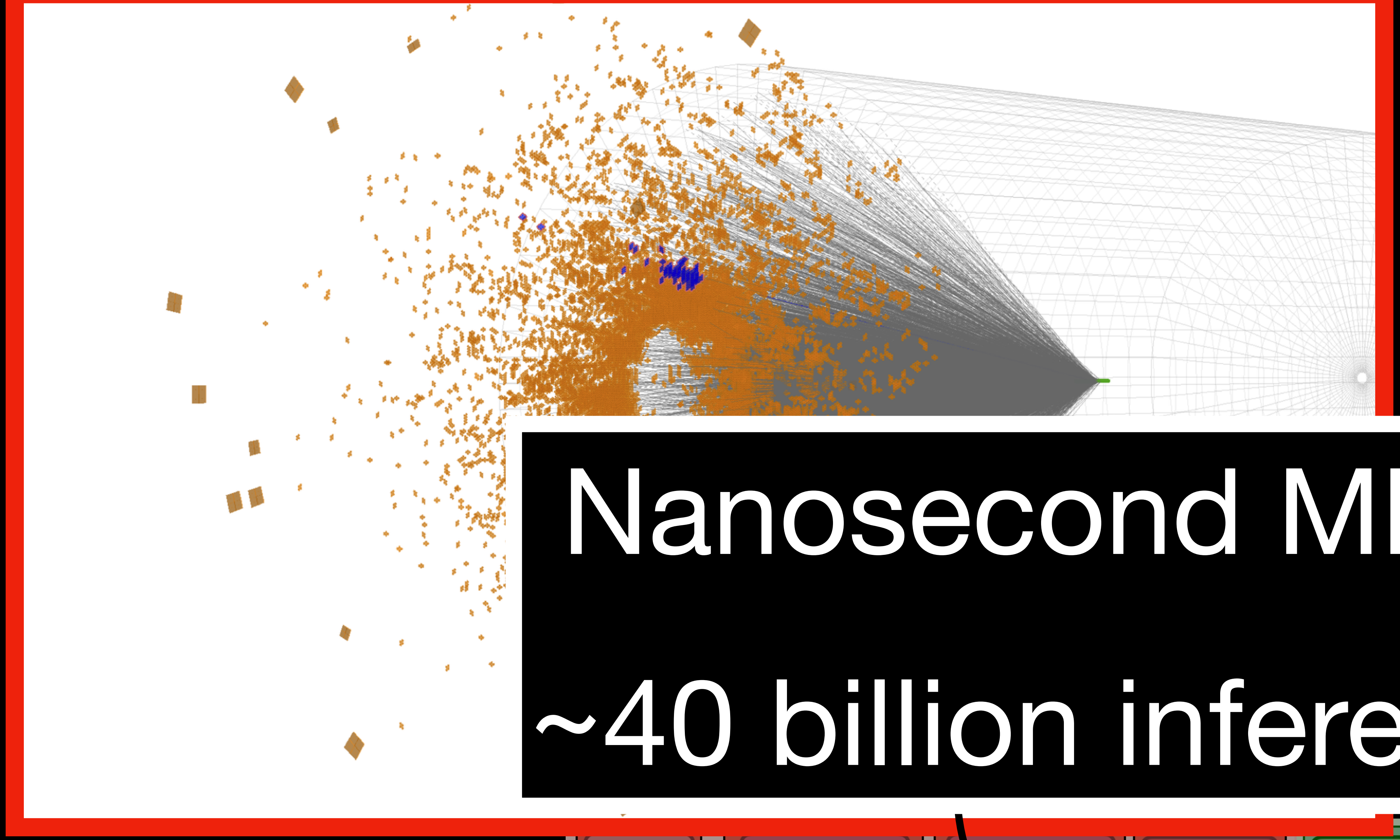


Run 4+5

Geneva
LHCb
ATLAS
ALICE
GPU inference
HLT trigger:
Latency O(100) ms
Fast inference on specialised hardware
ASIC inference
FPGA inference
Level-1 trigger:
Latency O(1) μs
Detector:
40 MHz
~Pb/s

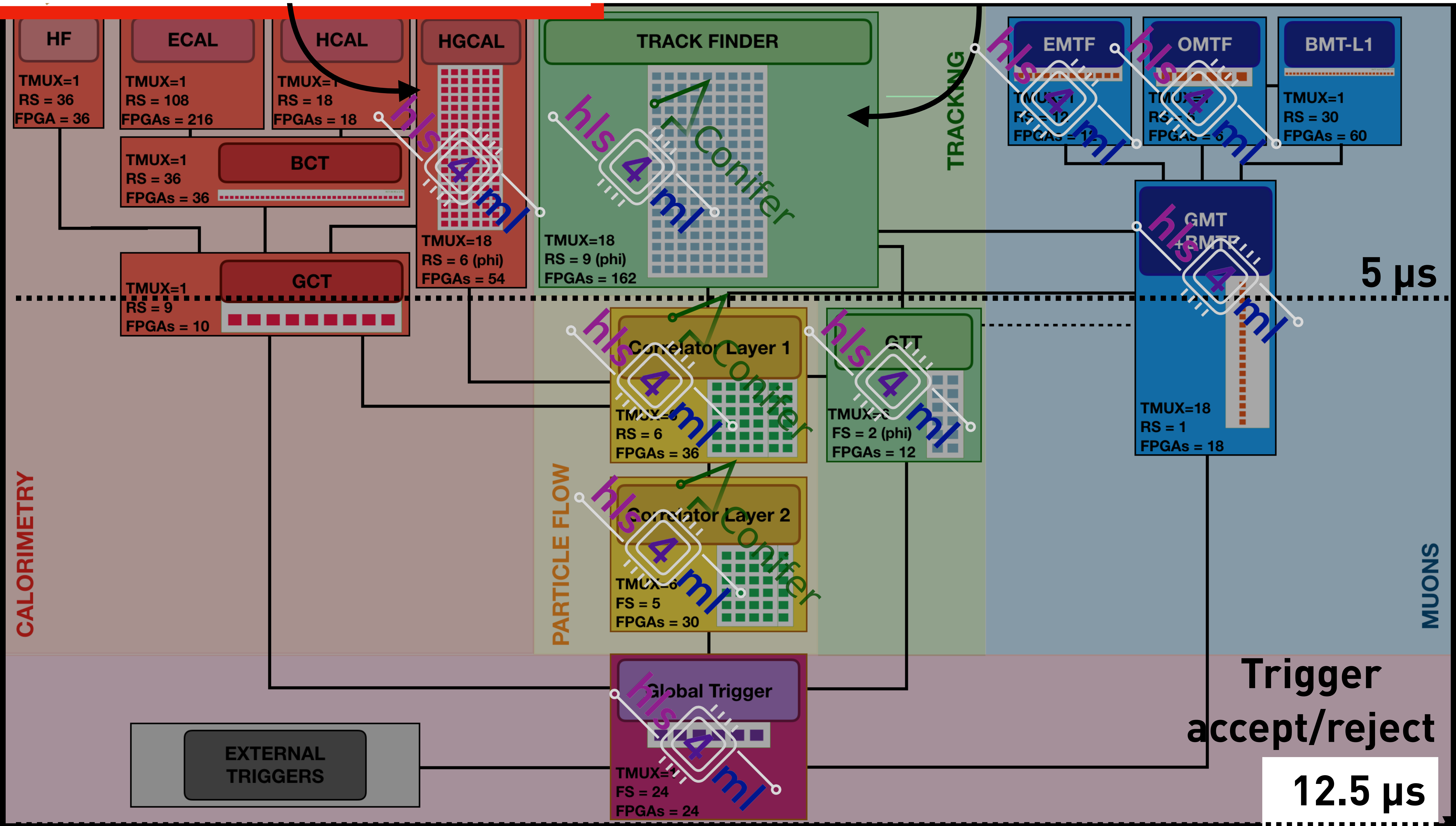


Silicon
Tracker
Electromagnetic
Calorimeter
Hadron
Calorimeter
4T
2T
return yoke interspersed
with Muon chambers
63 Tb/s
Xilinx Ultrascale+ FPGAs
CALORIMETRY:
370 FPGAs
*54 for HGCAL only!
TRACKING
174 FPGAs
MUONS:
96 FPGAs
5 μs
PARTICLE
FLOW:
66 FPGAs
GLOBAL
TRIGGER:
12 FPGAs
EXTERNAL
TRIGGERS
Trigger
accept/reject
12.5 μs
CALORIMETRY
PARTICLE FLOW
MUONS
USC55
UXC55



Nanosecond ML inference on FPGAs!

~40 billion inferences/s during HL-LHC

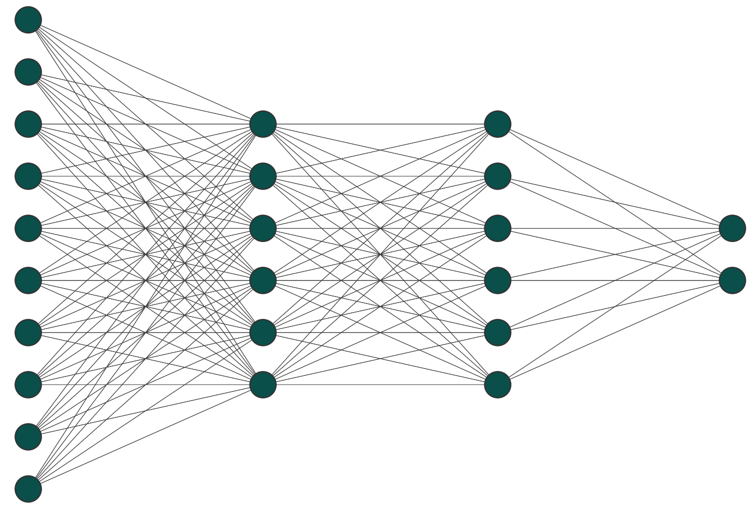


Conifer hls4ml

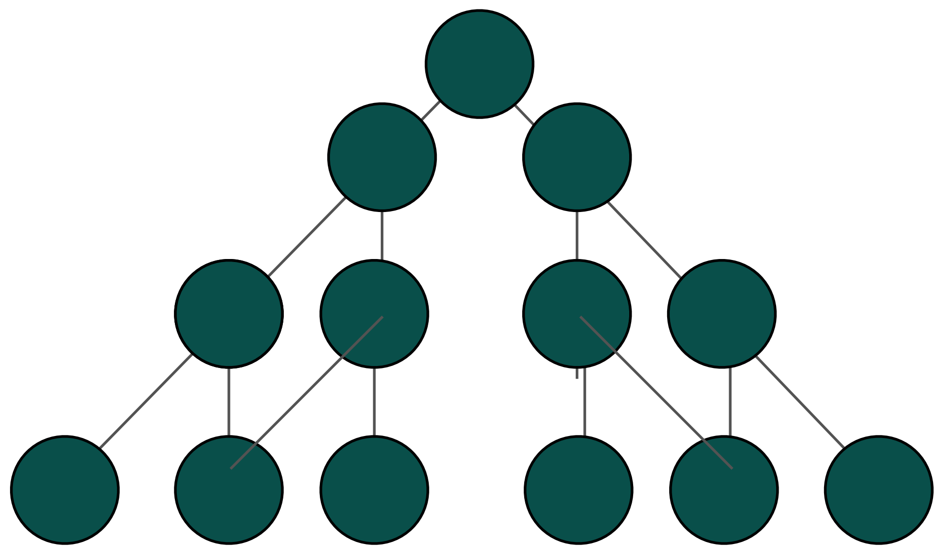
HEP developed libraries for fast ML on FPGAs

https://github.com/fastmachinelearning/hls4ml
https://fastmachinelearning.org/hls4ml/

KERAS / PyTorch / ONNX



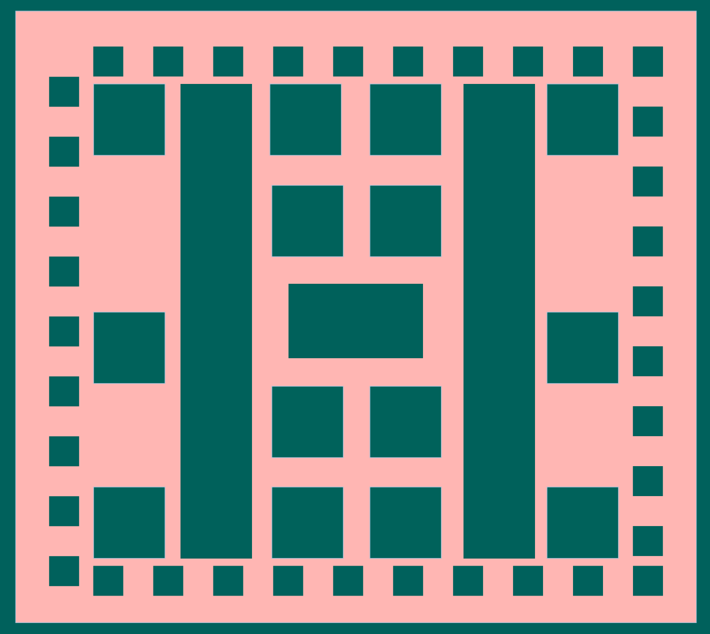
TensorFlow DF / scikit-learn / XGBoost

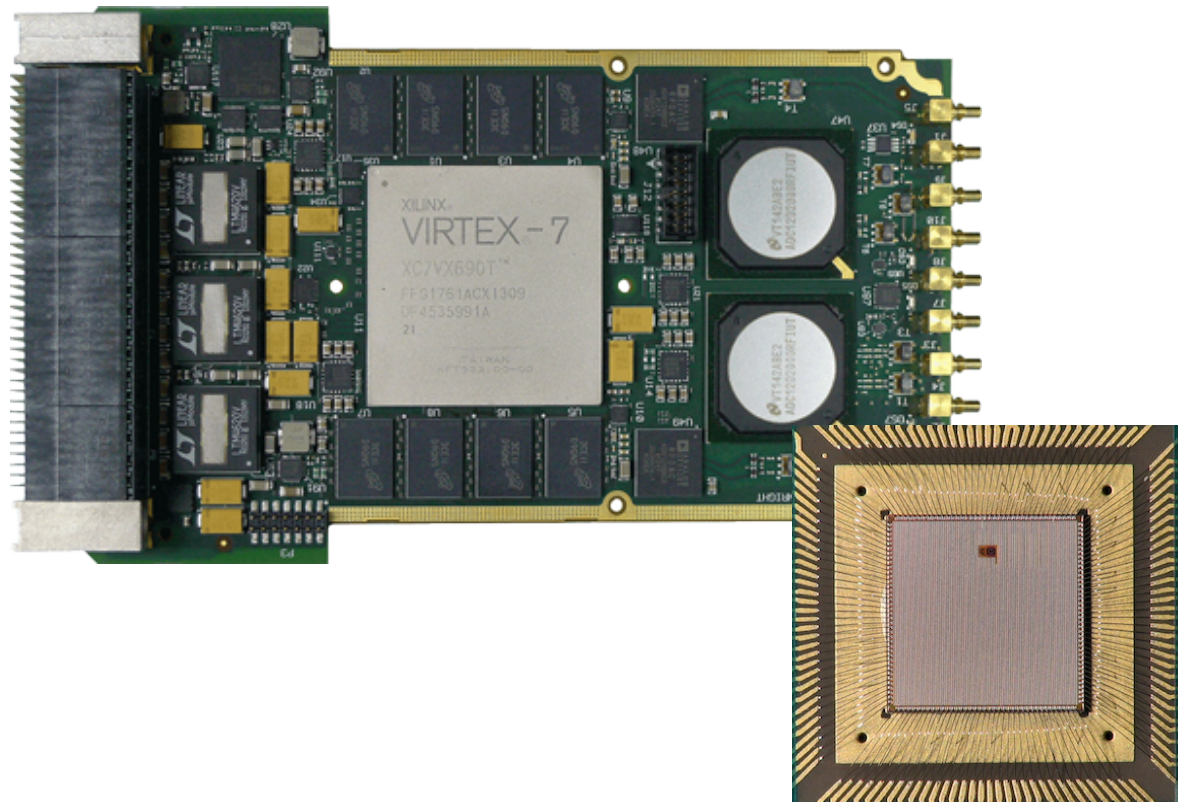


→ hls4ml →

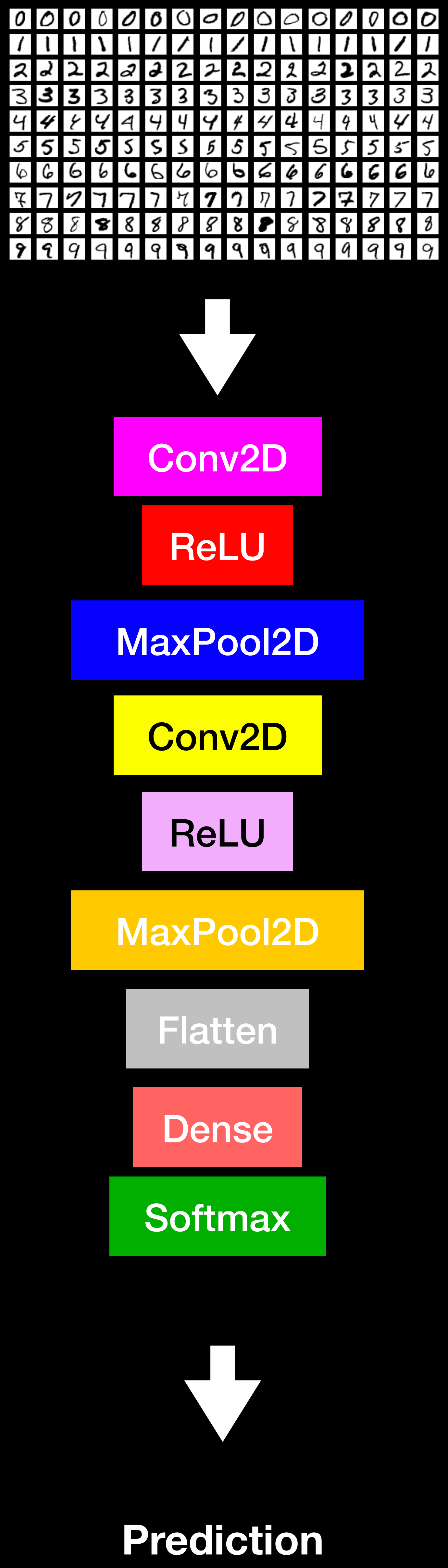
→ Conifer →

HLS project:
Vivado / Vitis / Intel Quartus / IntelOne API / Catapult





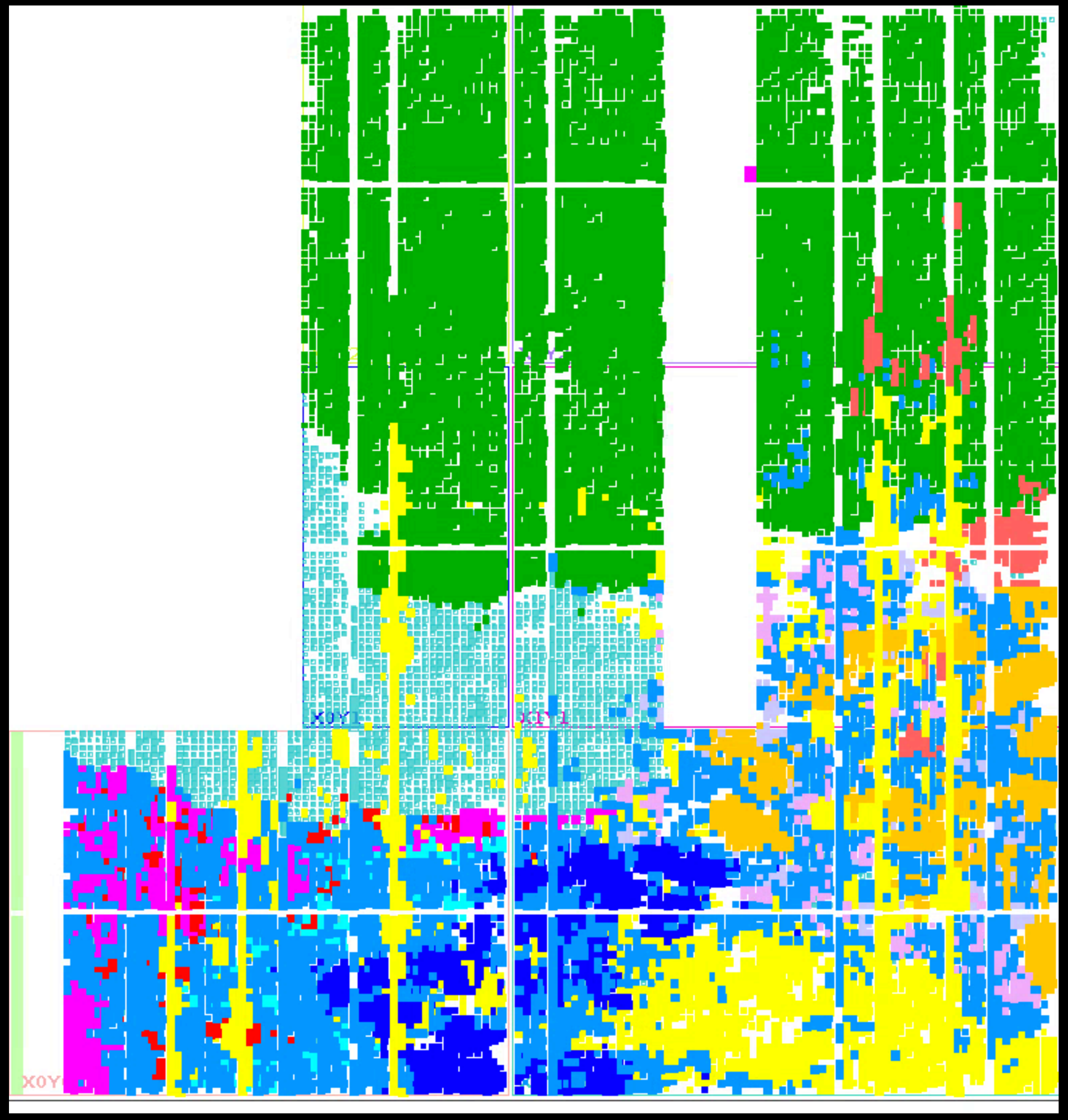

```
pip install hls4ml
pip install conifer
```

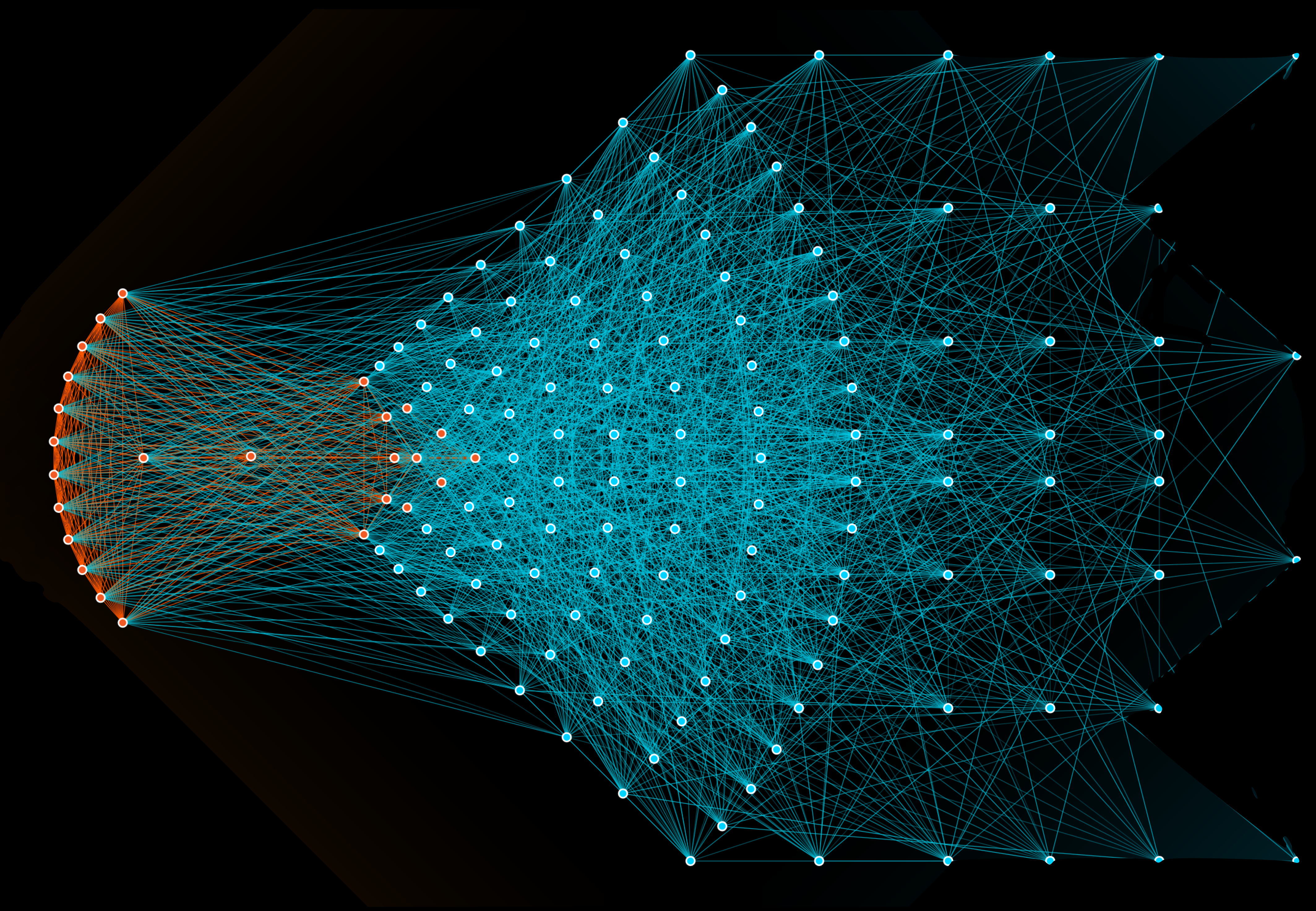


hls4ml

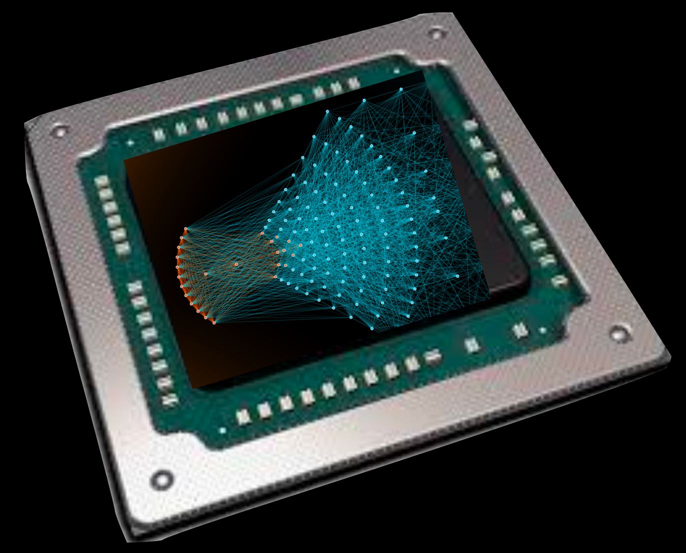
Data flow architecture

- Tailored hardware for a model
- Each layer is separate compute unit
- Stay on-chip
- “Decisions are design time”

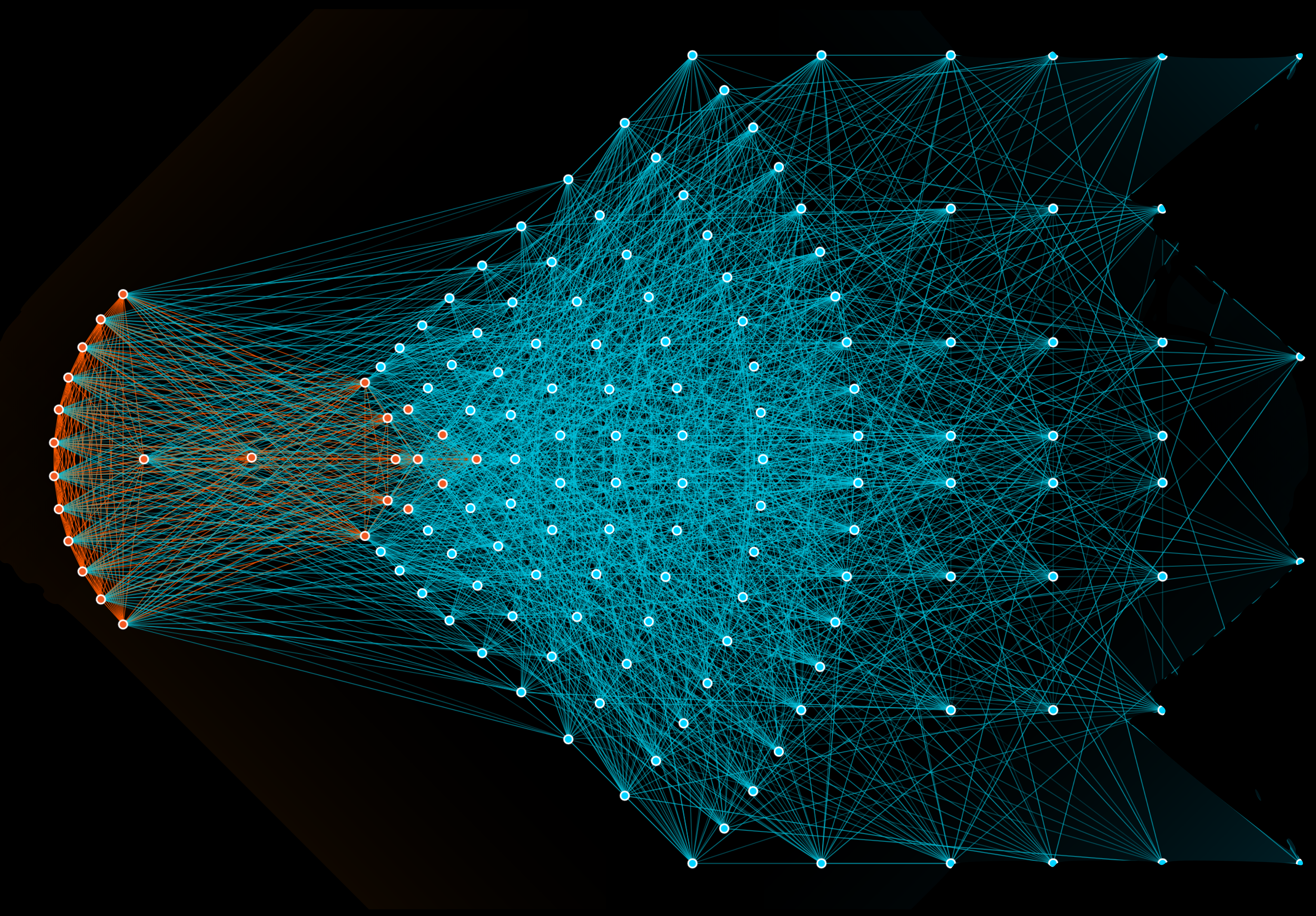




Ideally

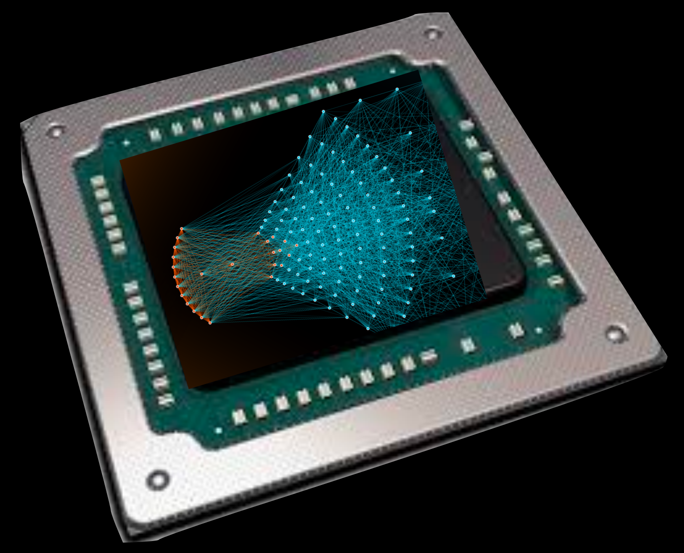


Reality



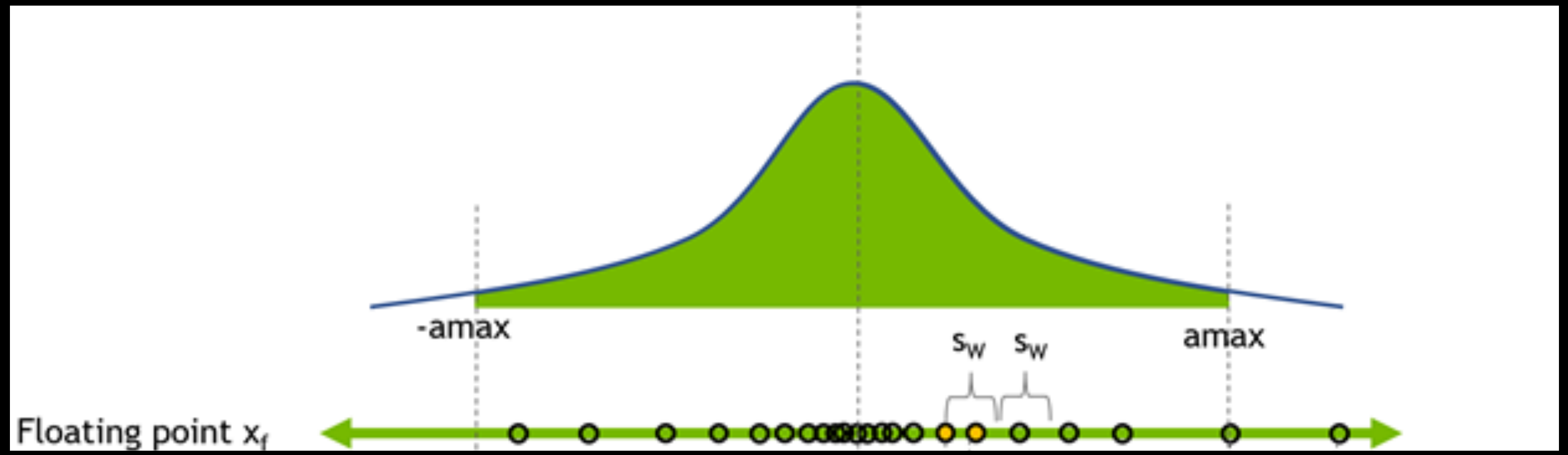
Ideally

- Quantization
- Pruning
- Parallelisation
- Knowledge distillation



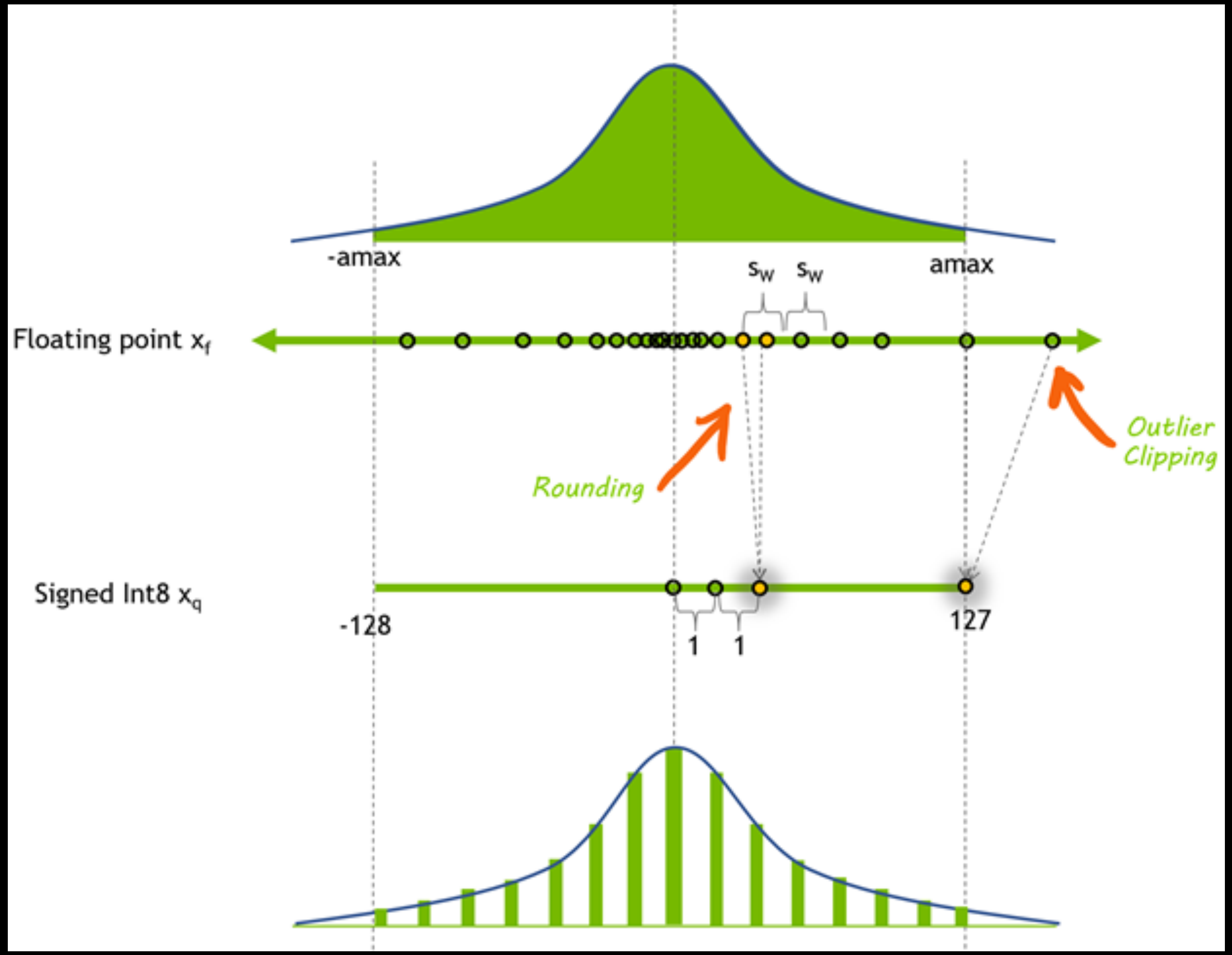
Reality

Quantization



Floating point 32:
4B numbers in [-3.4e38, +3.4e38]

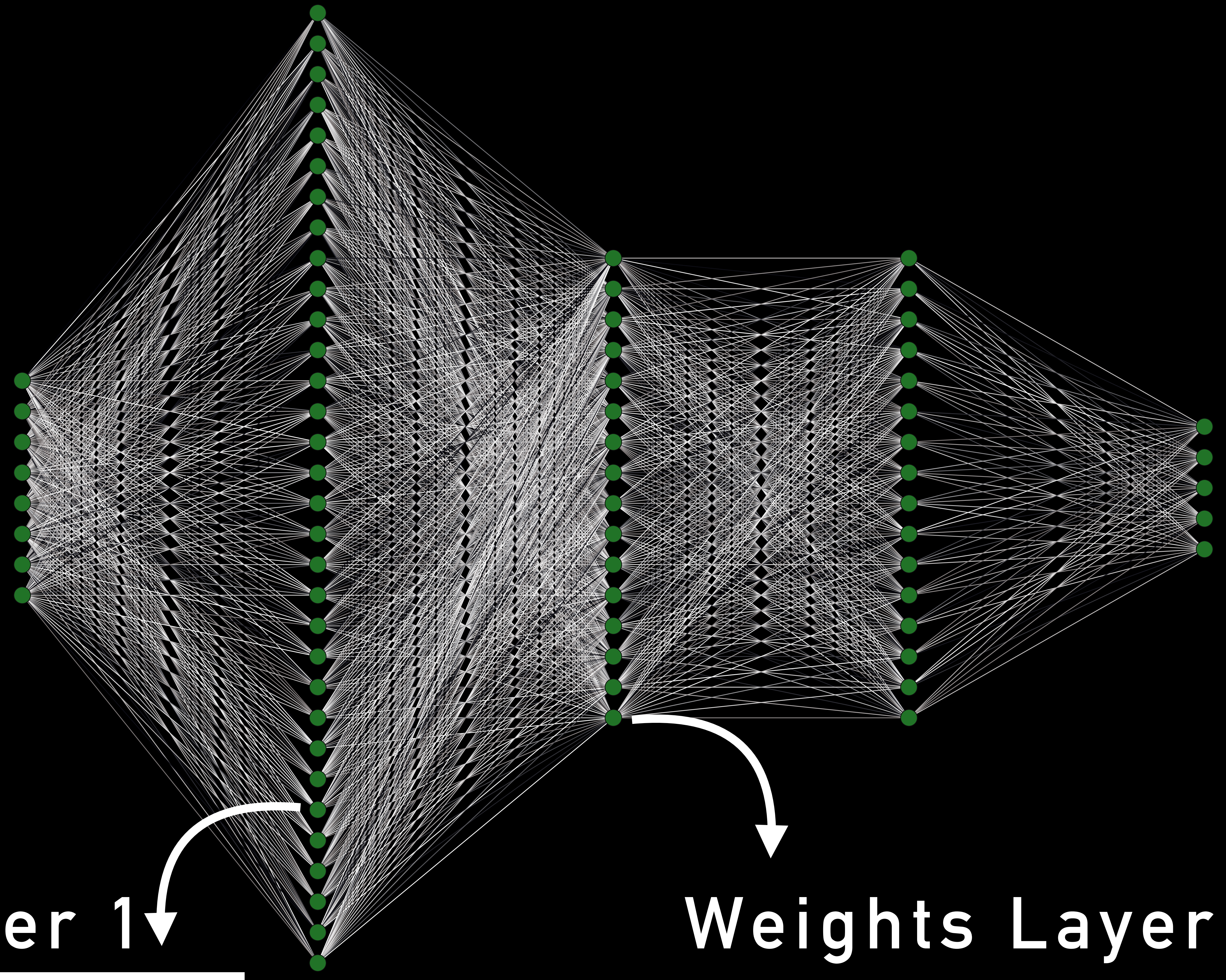
Quantization



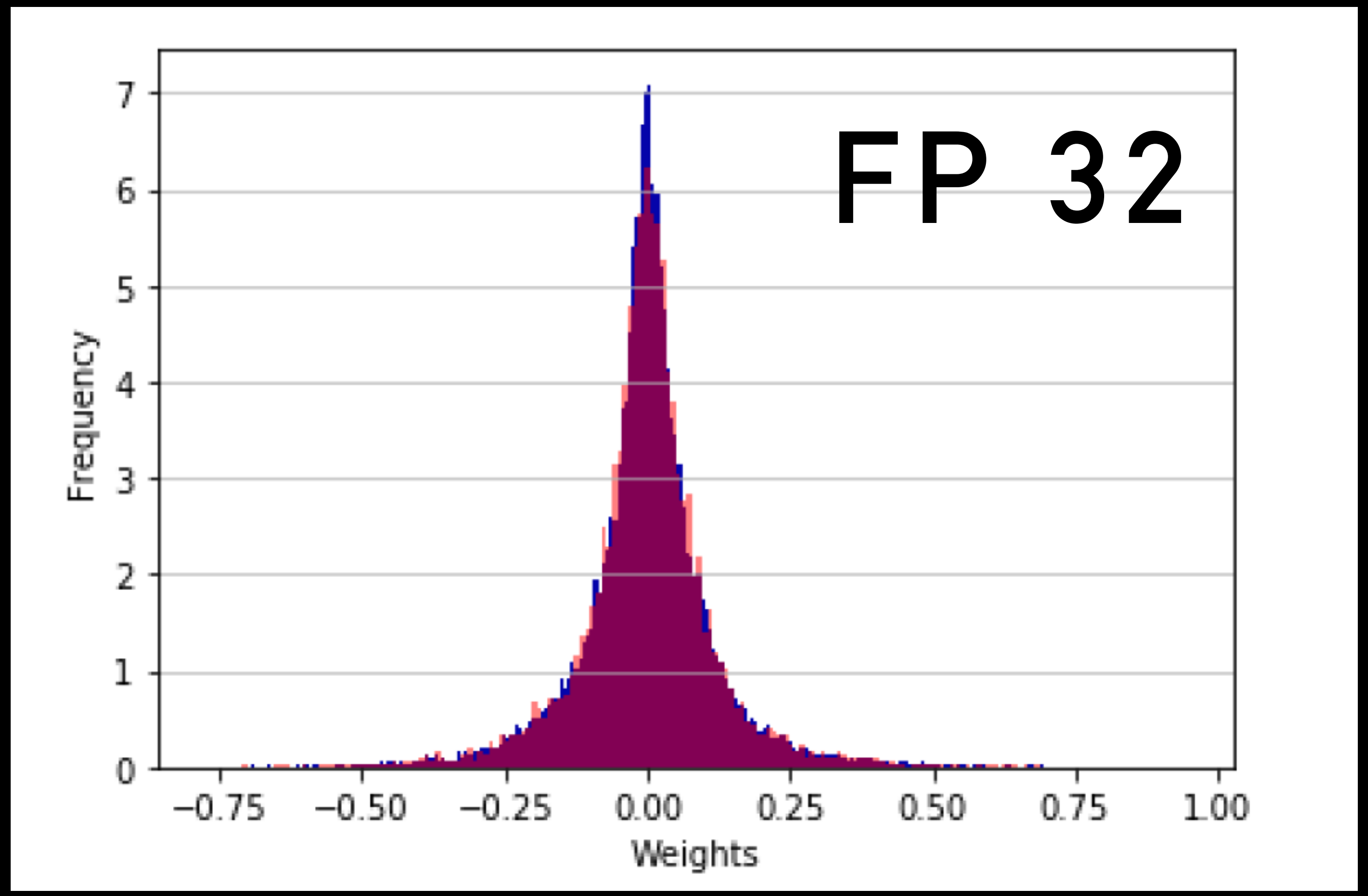
Quantising:

int8 2^8=256 numbers in [-128,127]

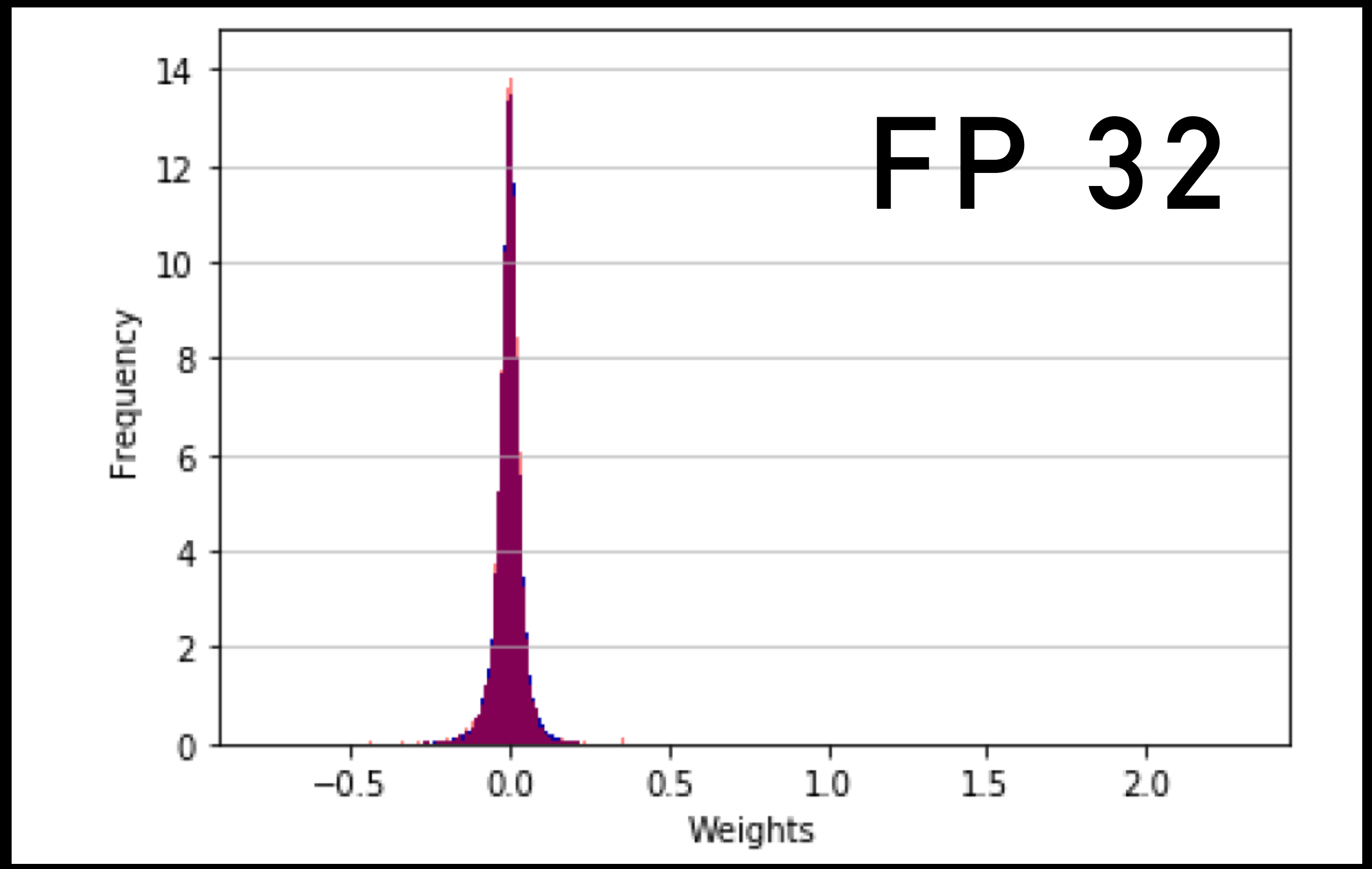
$$x_q = Clip(Round(\frac{x_f}{scale}))$$



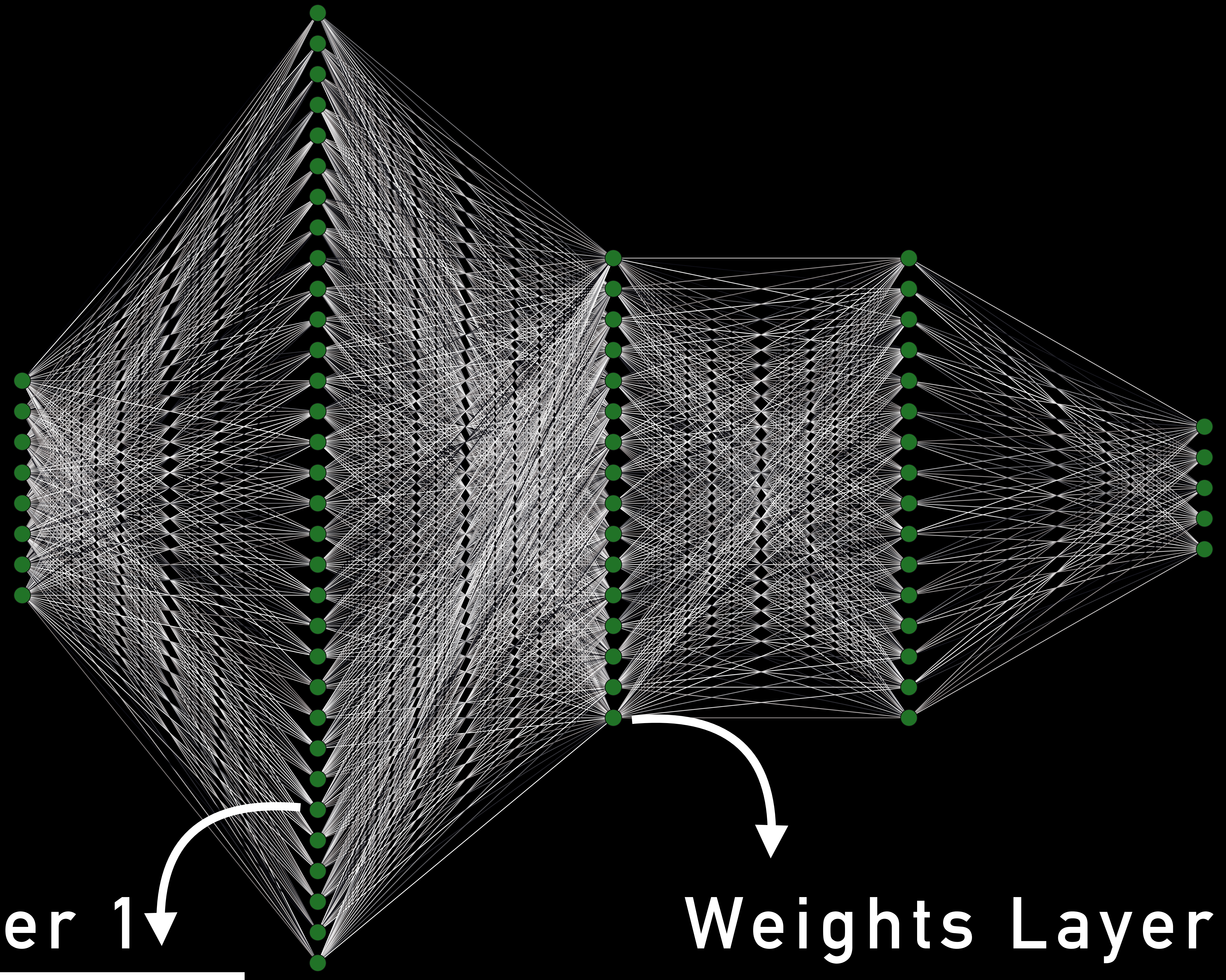
Weights Layer 1



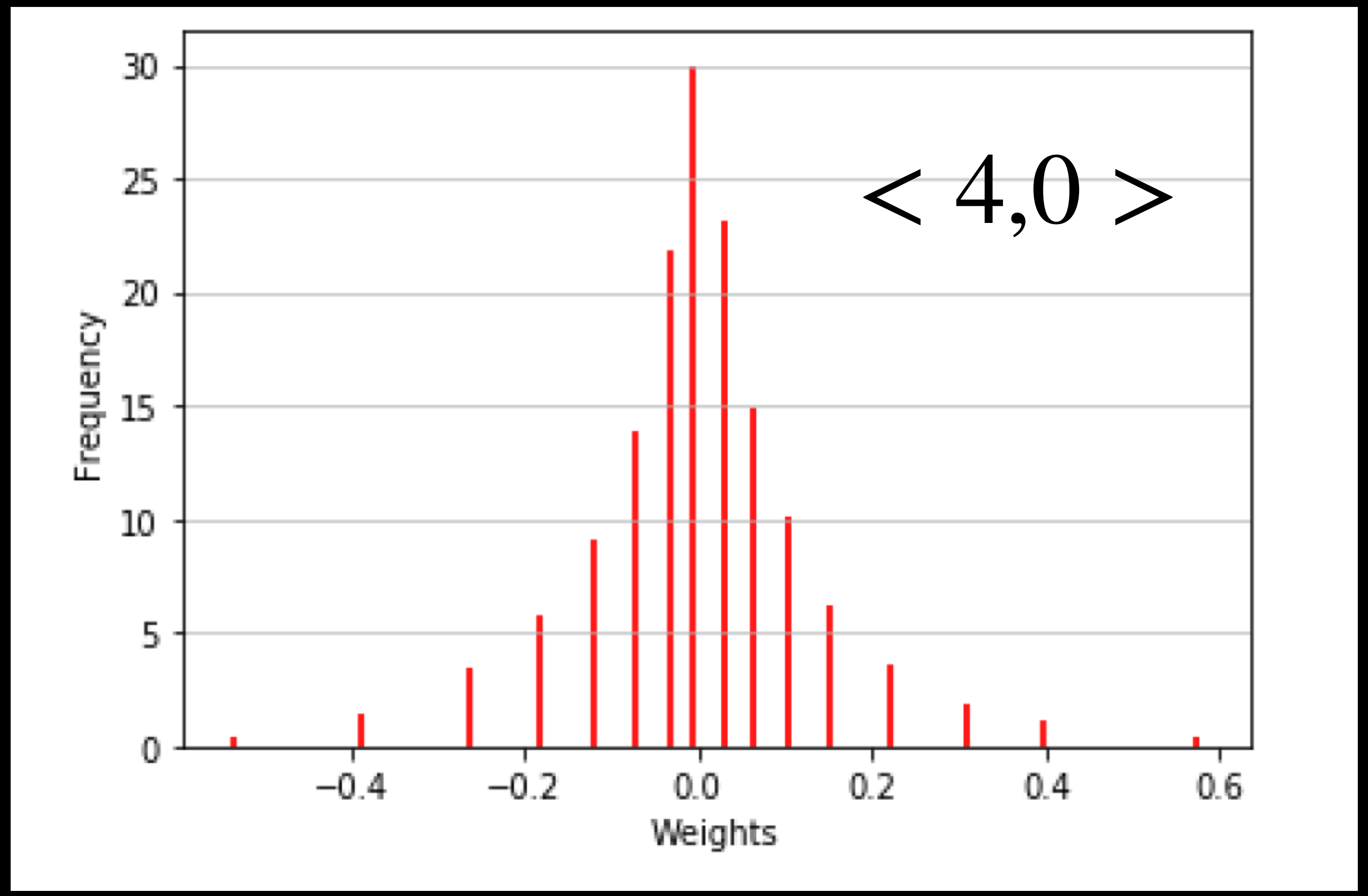
Weights Layer 2



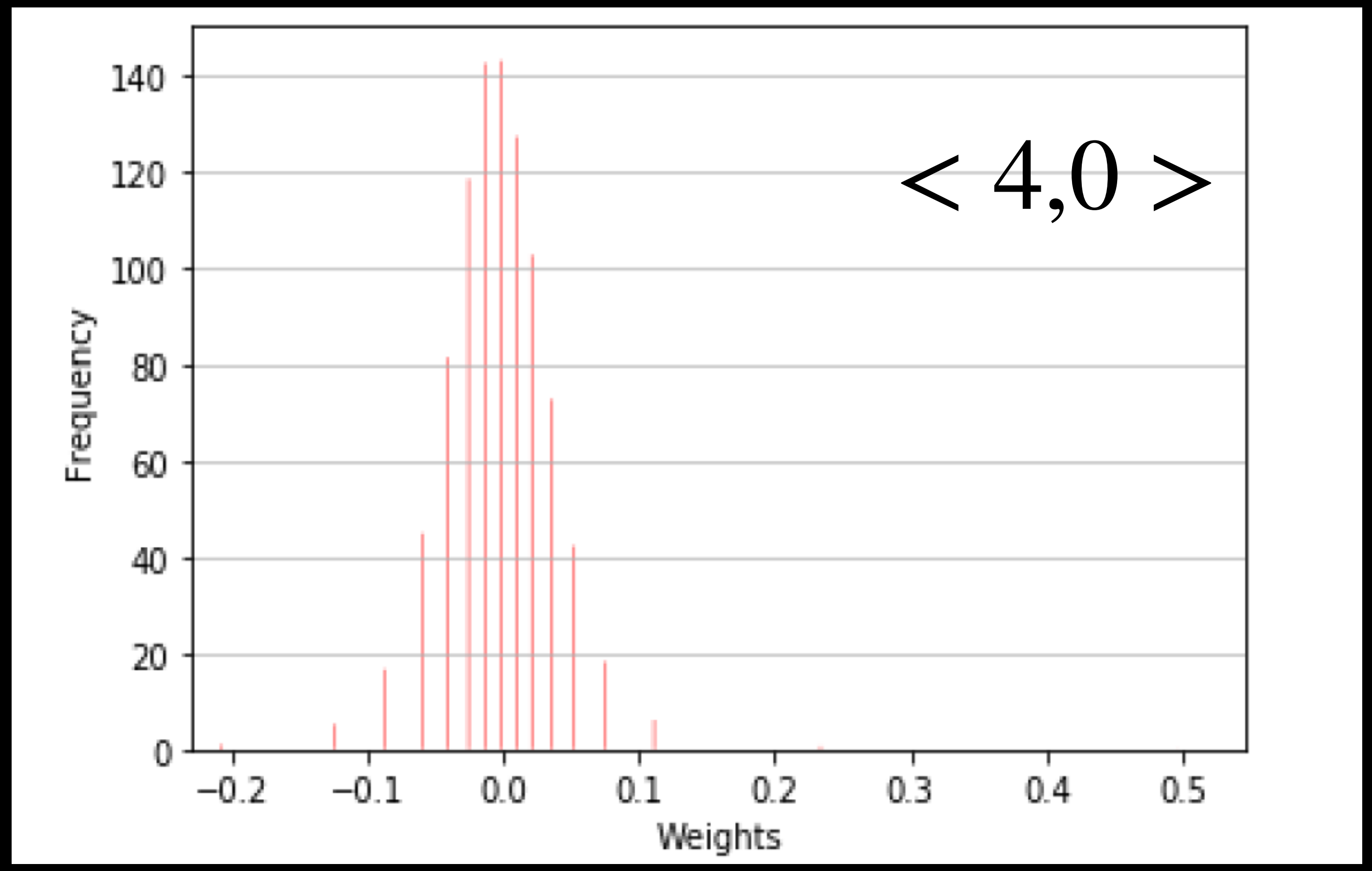
Fixed point

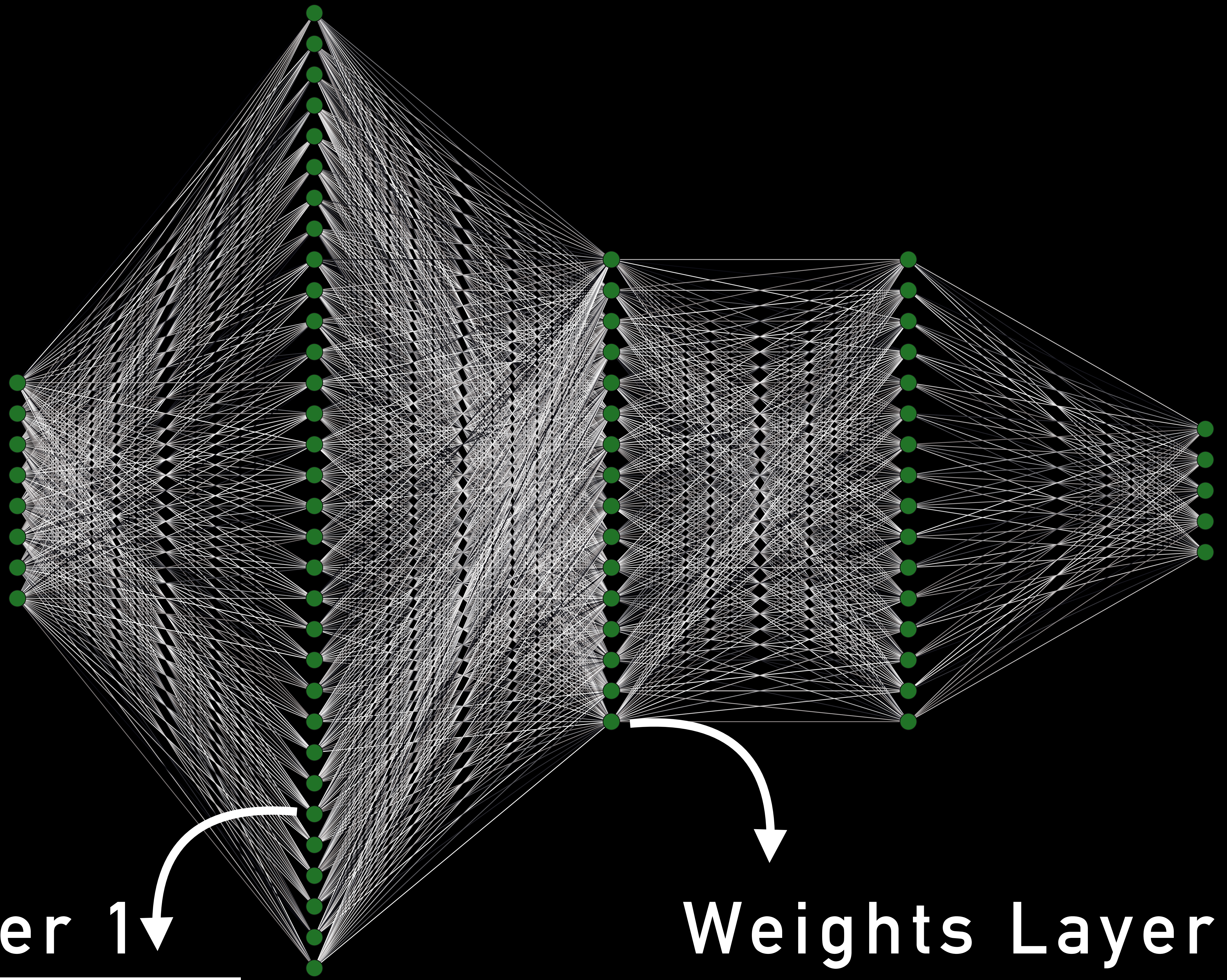


Weights Layer 1

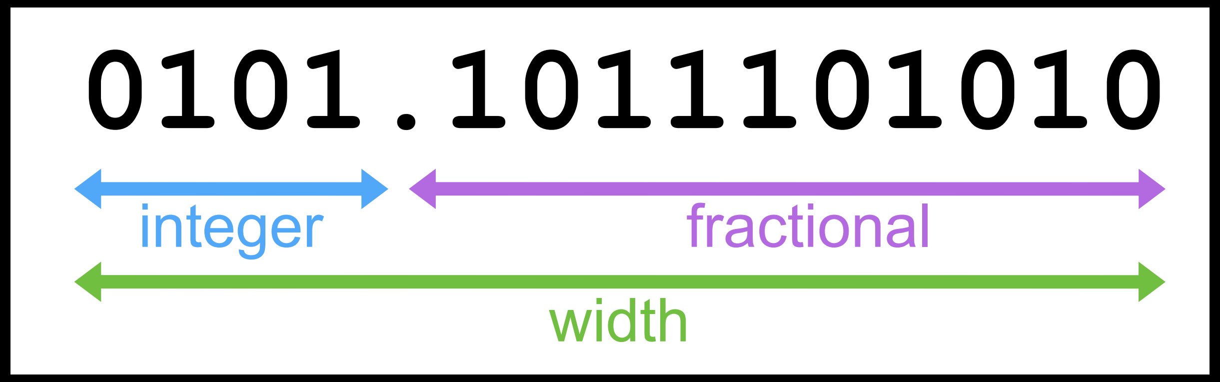


Weights Layer 2

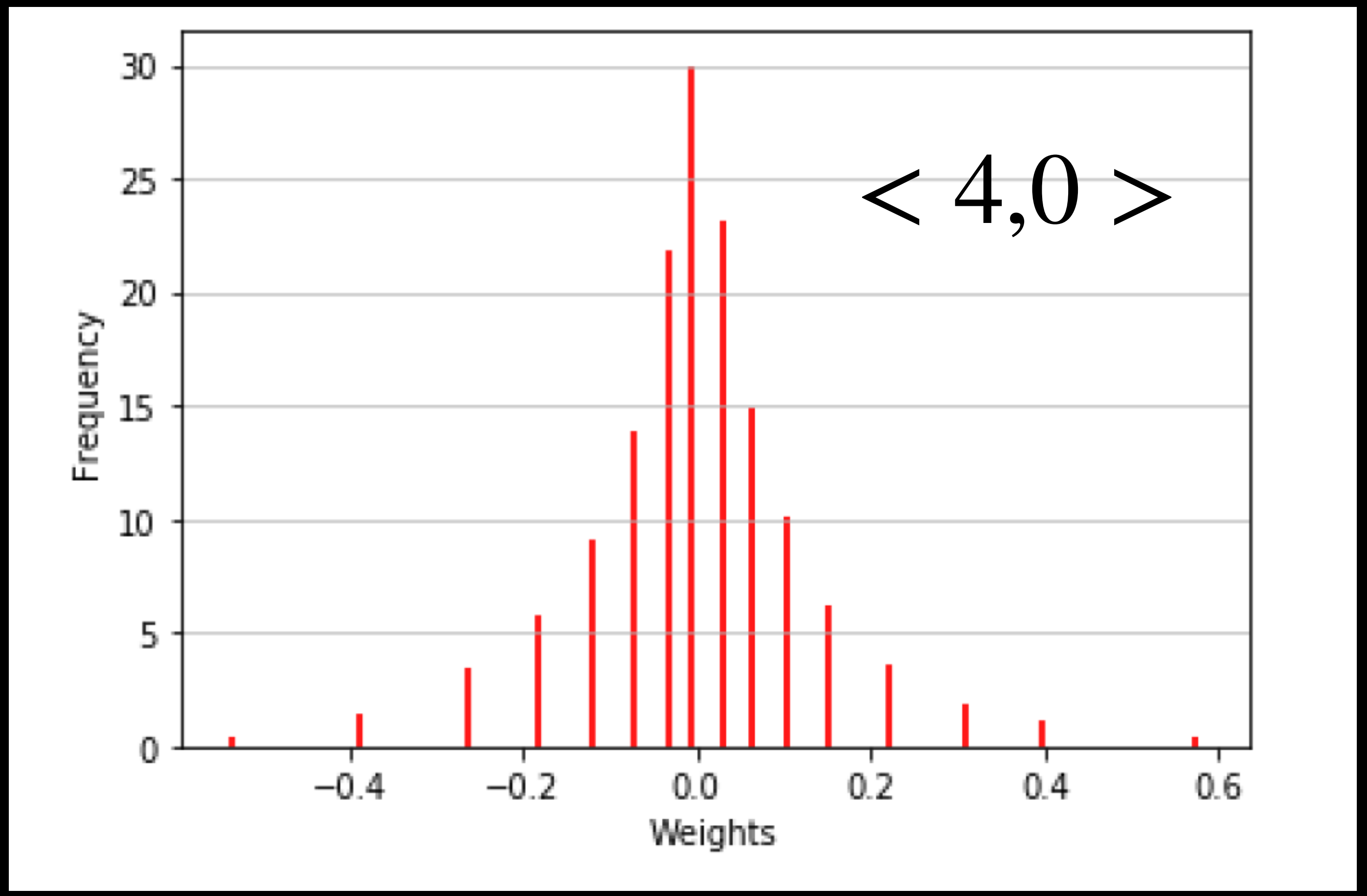




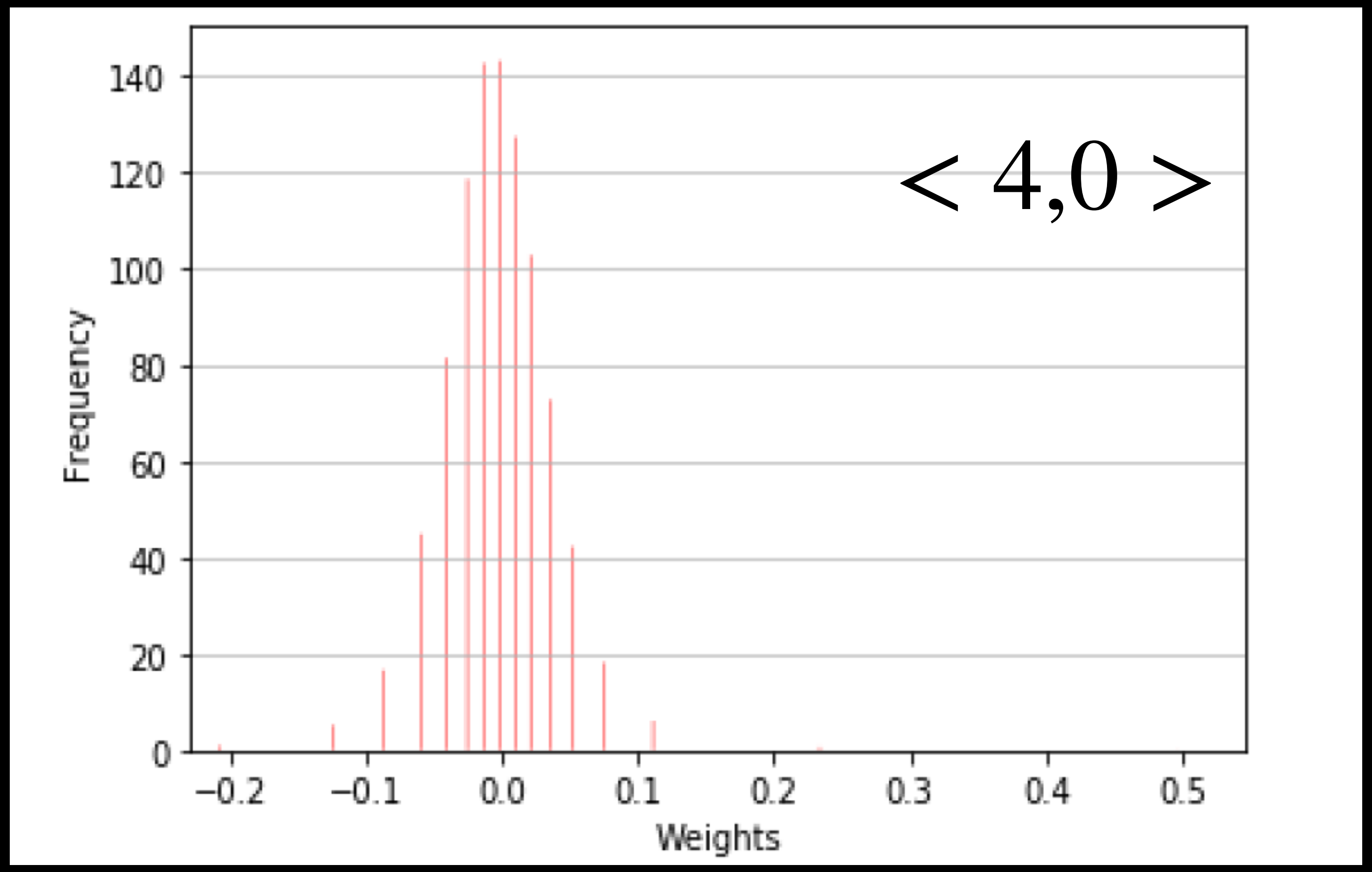
Fixed point



Weights Layer 1

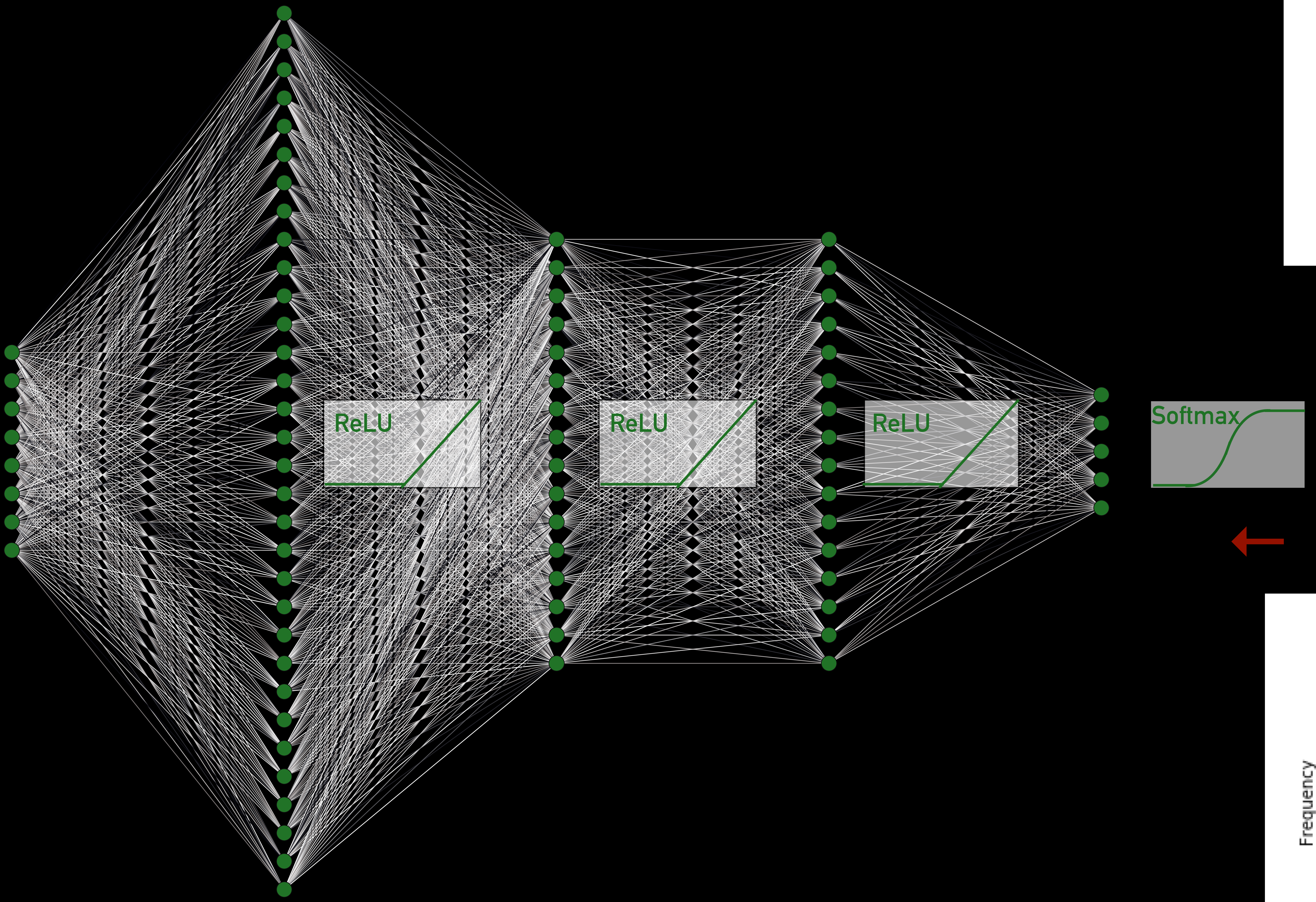


Weights Layer 2

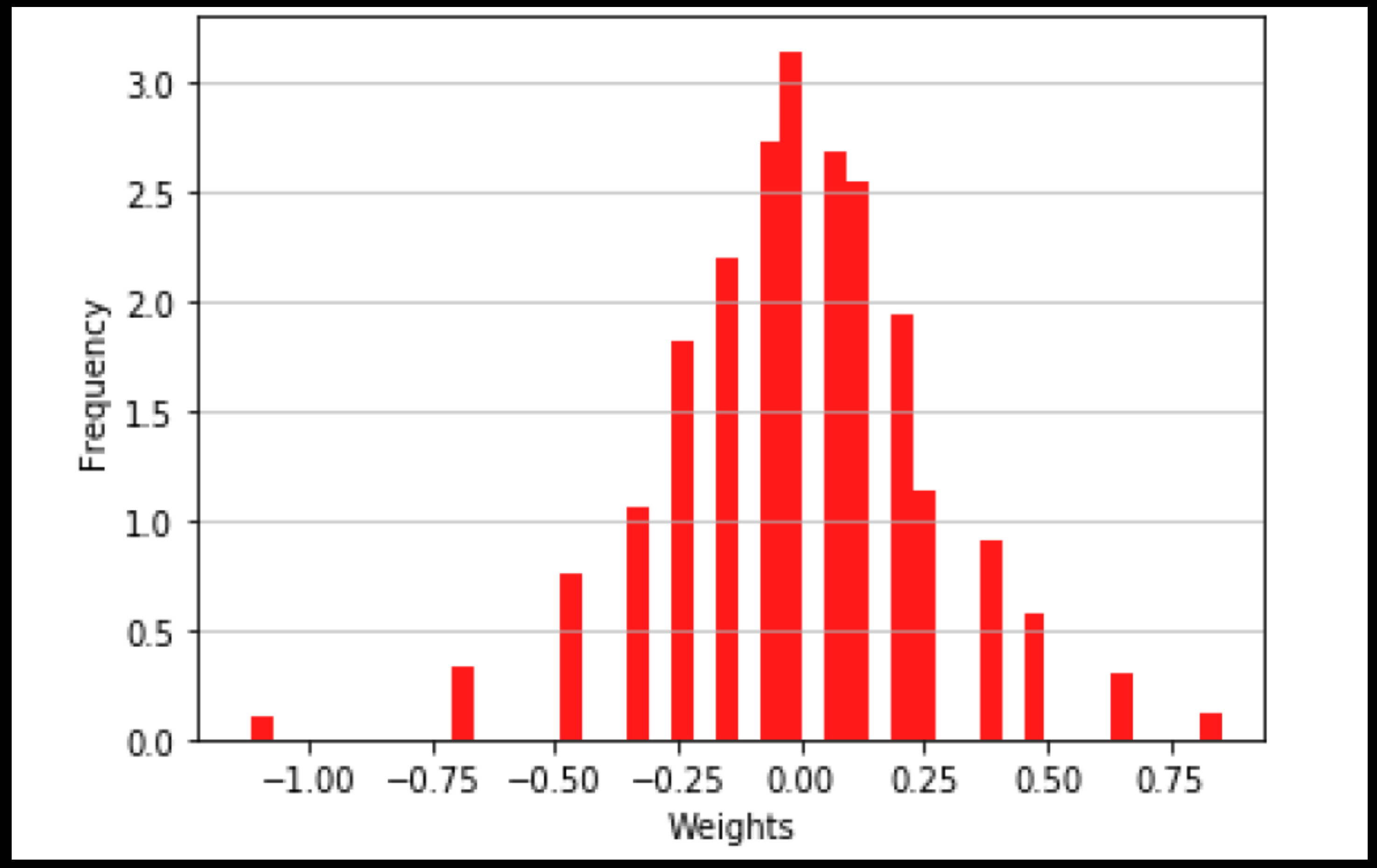


hls4ml + Google

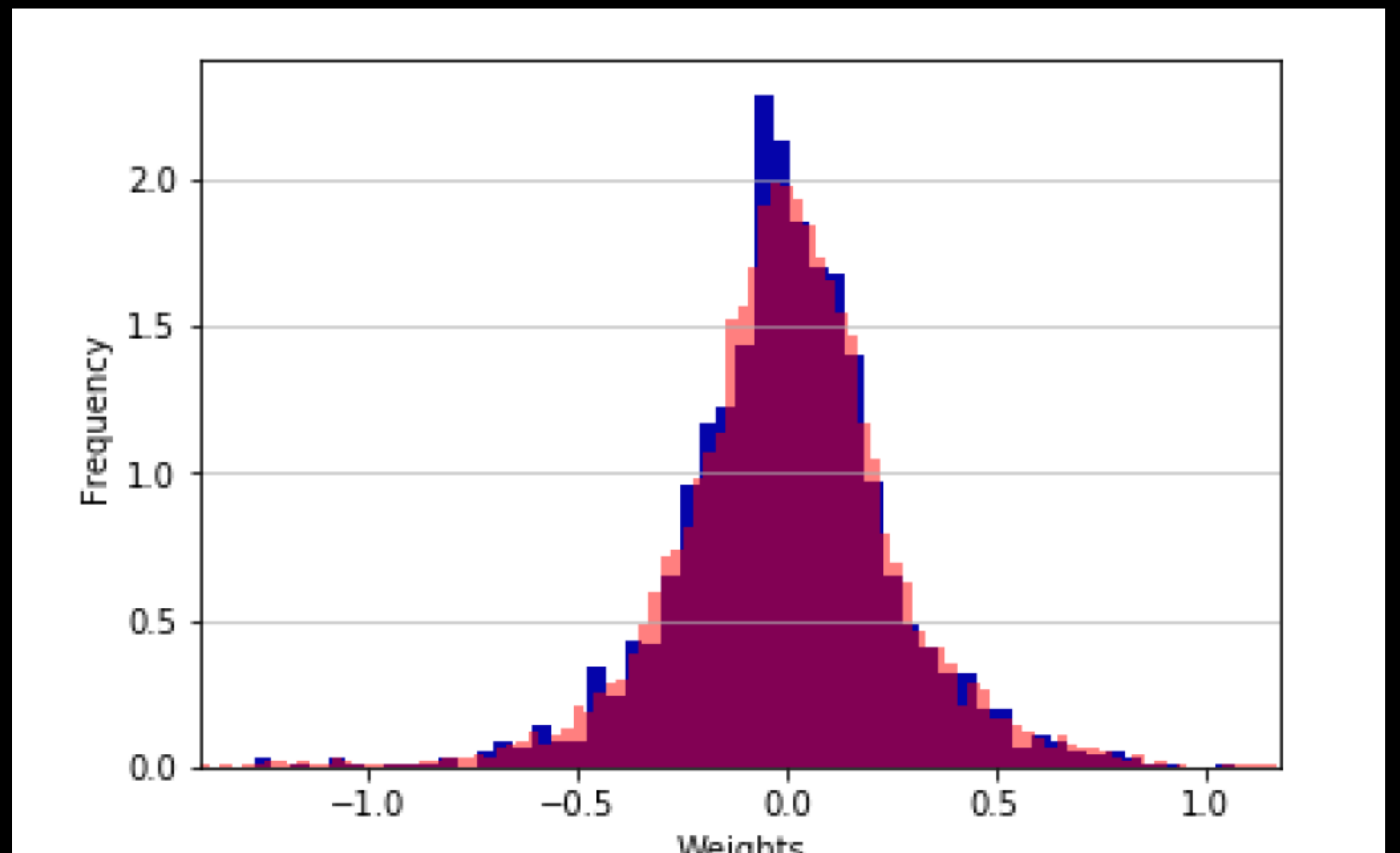
Quantization-aware training



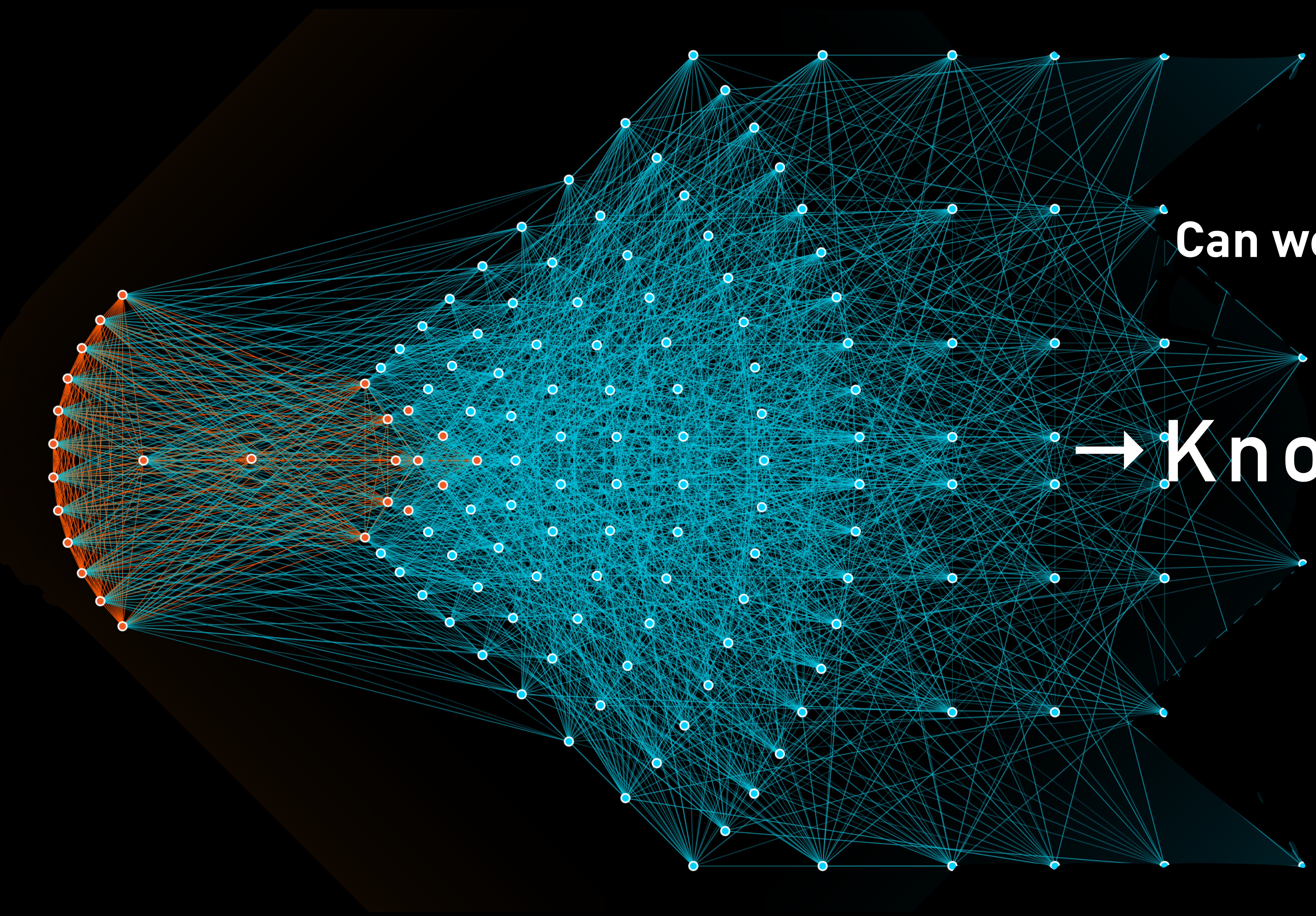
Forward pass →



← Back propagation

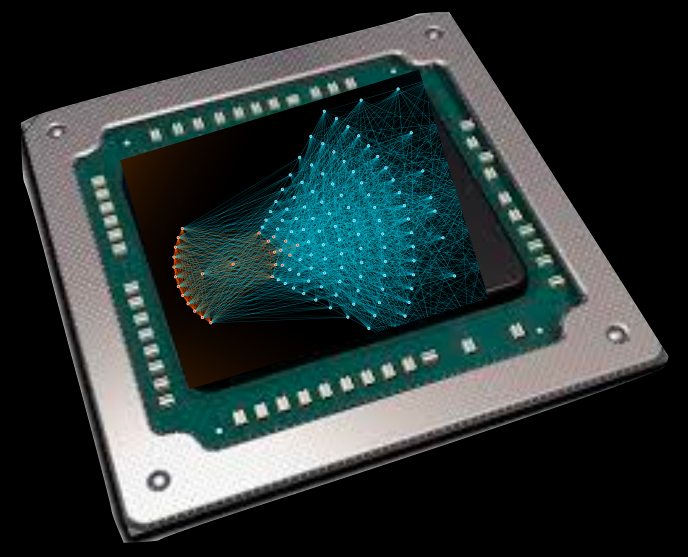


Nature Machine Intelligence 3 (2021)



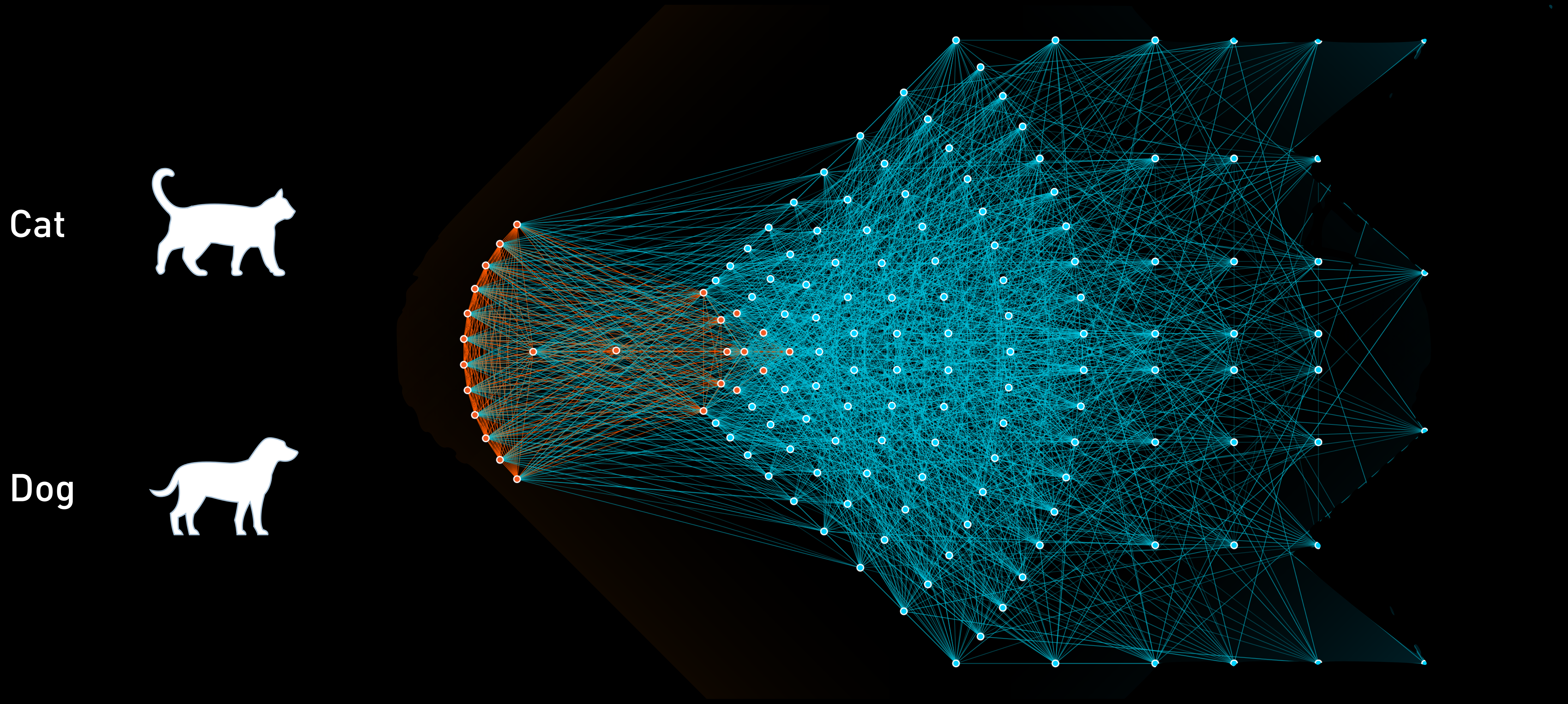
Can we have the best of both worlds?

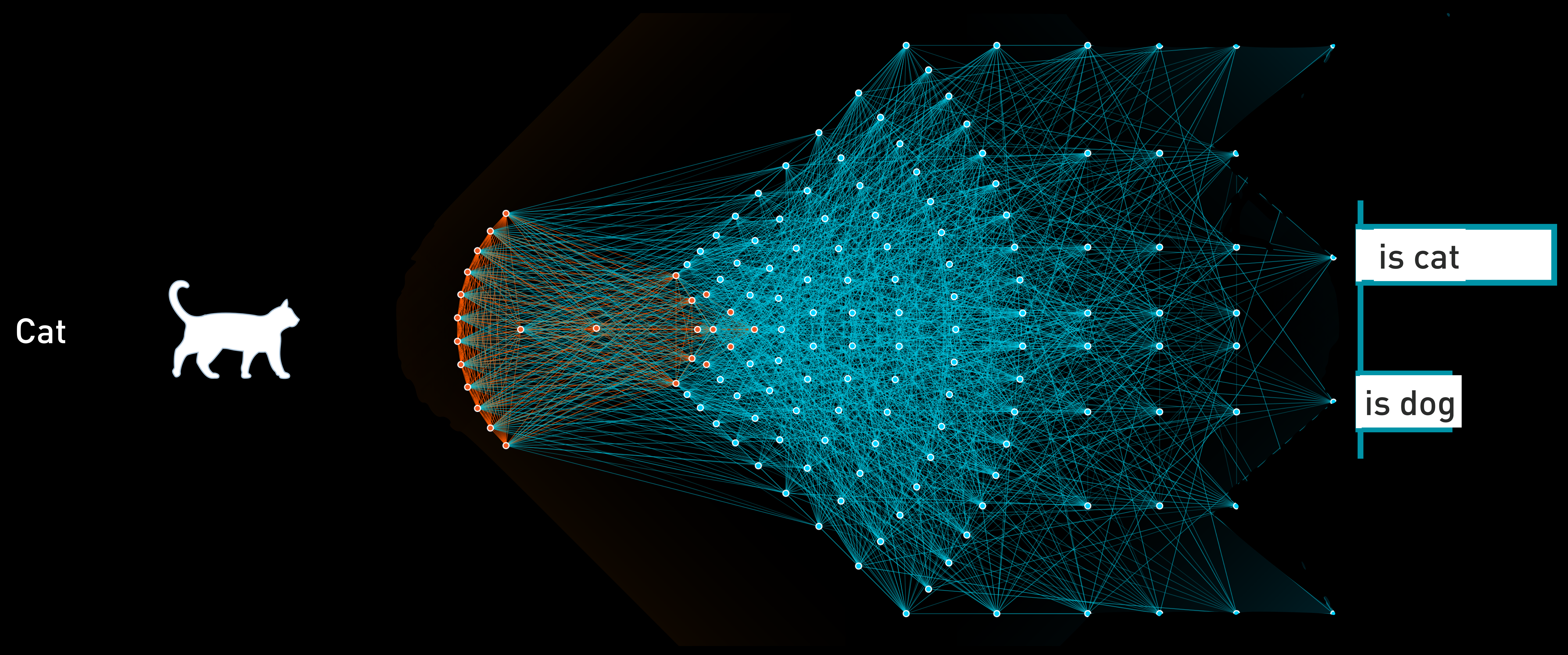
→Knowledge Distillation

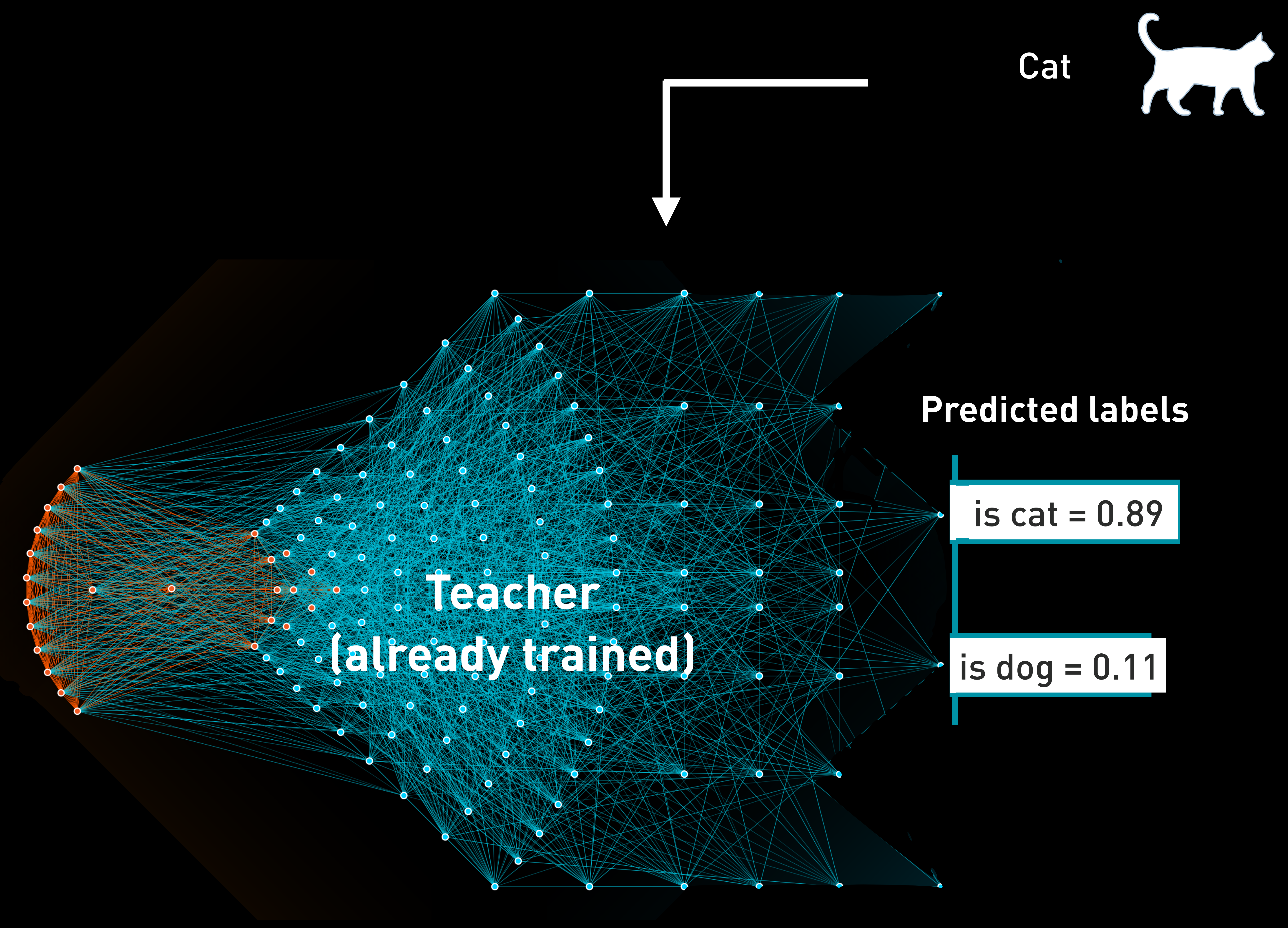


Train

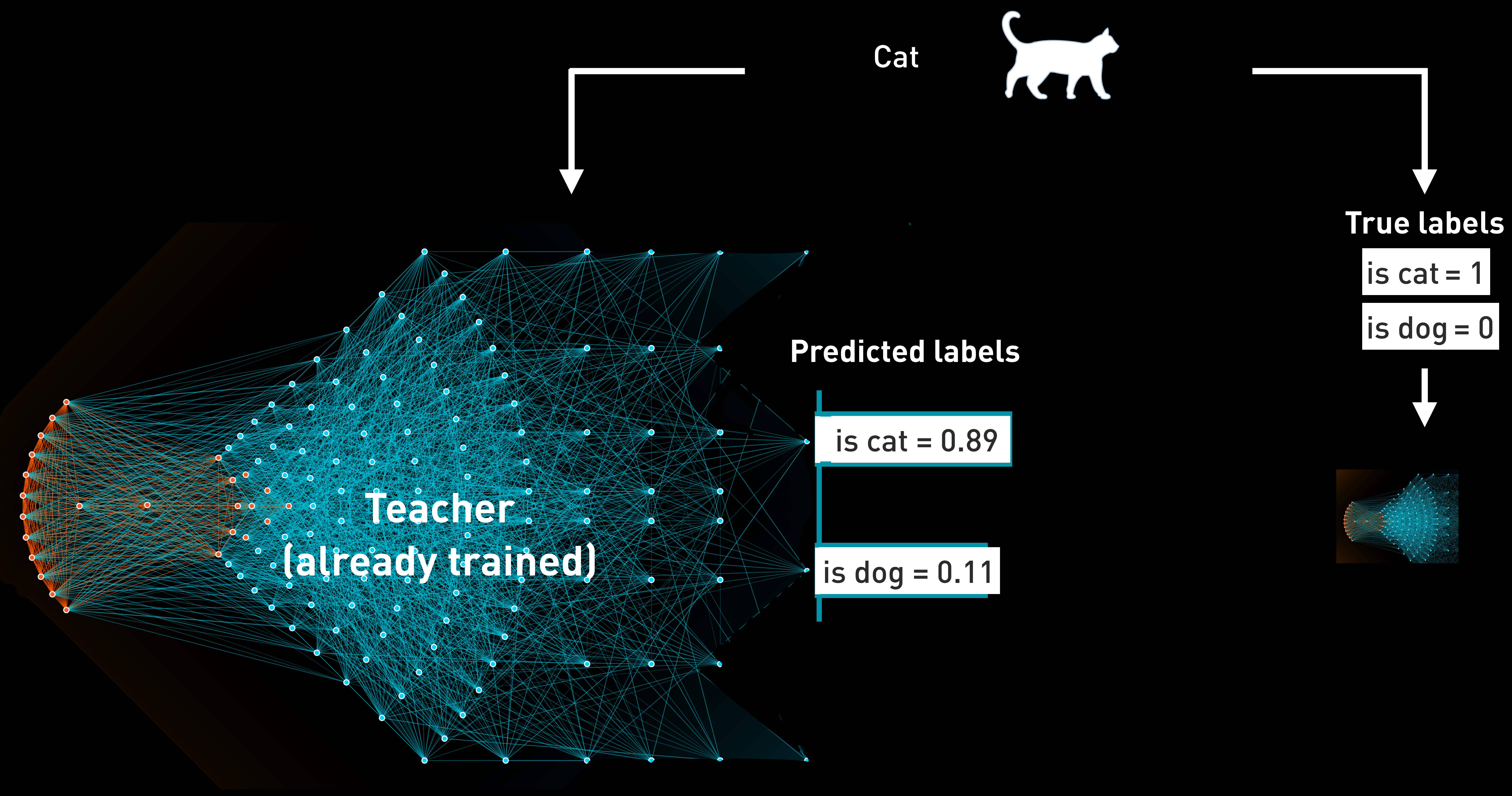
Inference


Cat
Dog


Cat
is cat
is dog



Cat
Teacher
(already trained)
Predicted labels
is cat = 0.89
is dog = 0.11



Cat
True labels
is cat = 1
is dog = 0
Predicted labels
is cat = 0.89
is dog = 0.11
Teacher
(already trained)

Predicted labels

is cat = 0.46

is dog = 0.54

True labels

is cat = 0

is dog = 1

Soft labels contain information!!

Predicted labels

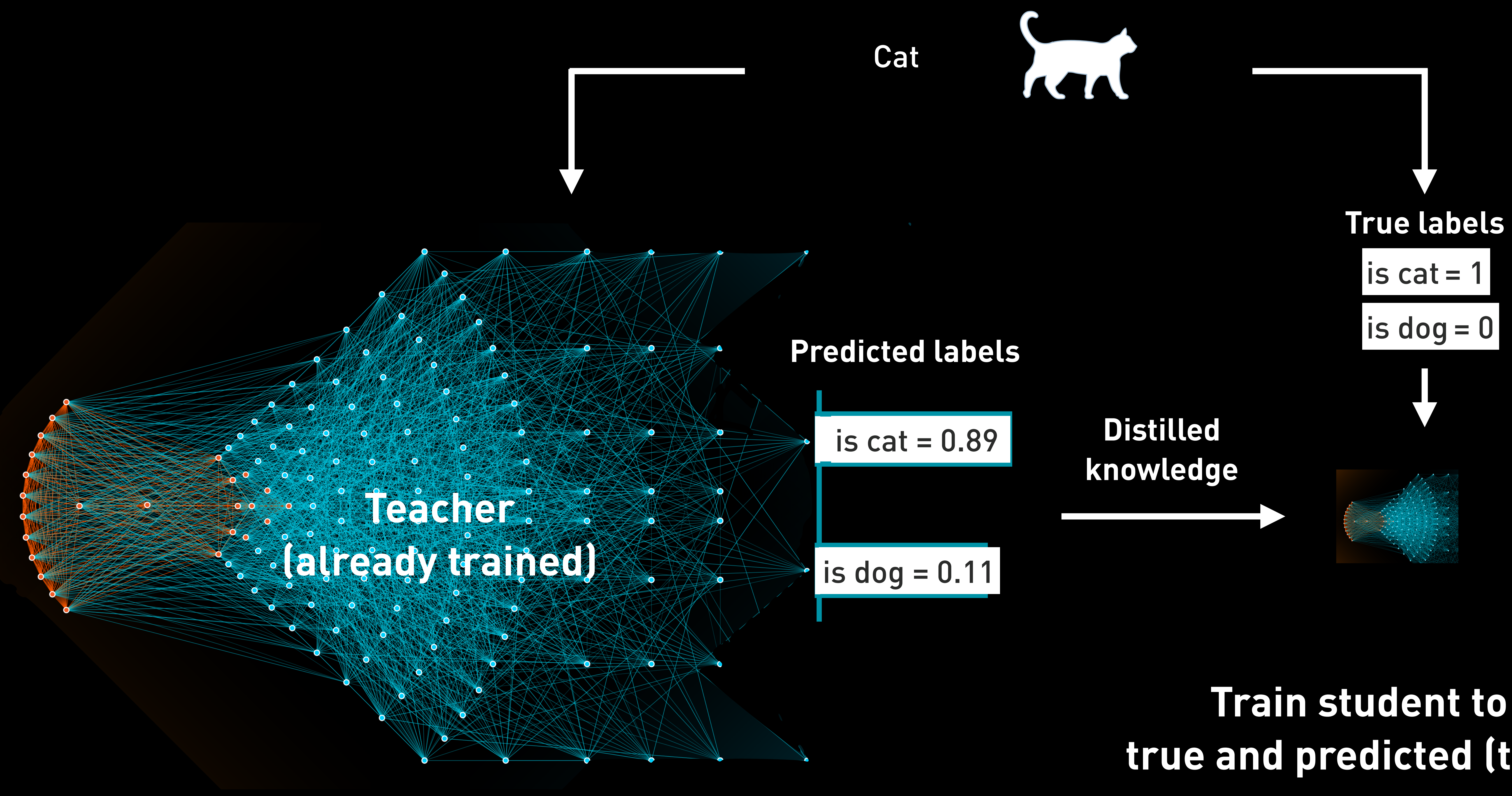
is cat = 0.03

is dog = 0.97

True labels

is cat = 0

is dog = 1



Train student to learn both true and predicted (teacher) labels!

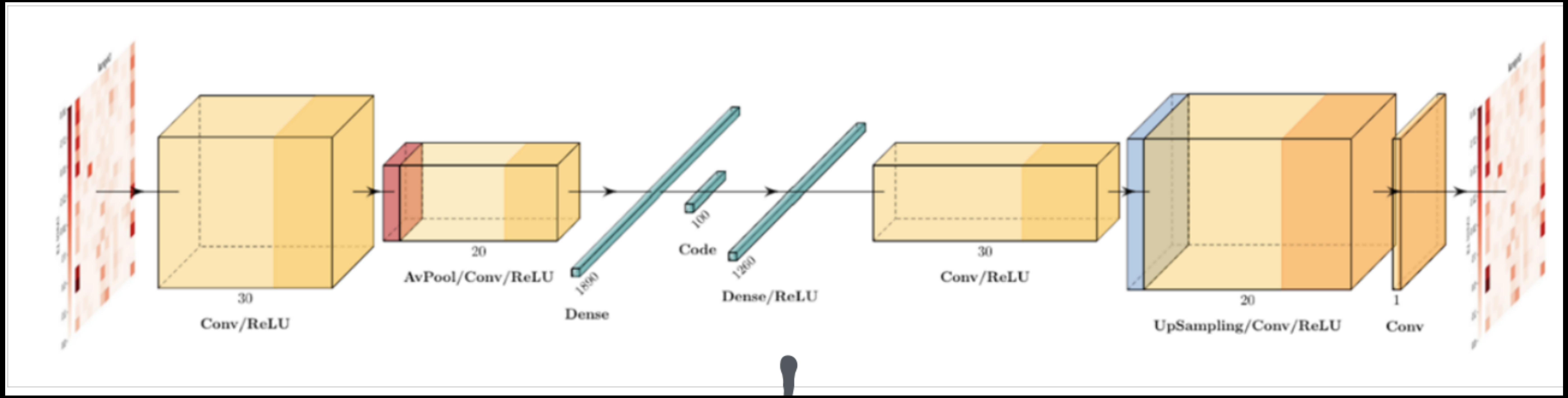
$$L_{total} = \beta \times L_{Distillation} + \alpha \times L_{student}$$

Student learns subtle learned features from teacher!

CICADA

Using knowledge distillation for CNN in hardware Calorimeter Trigger!

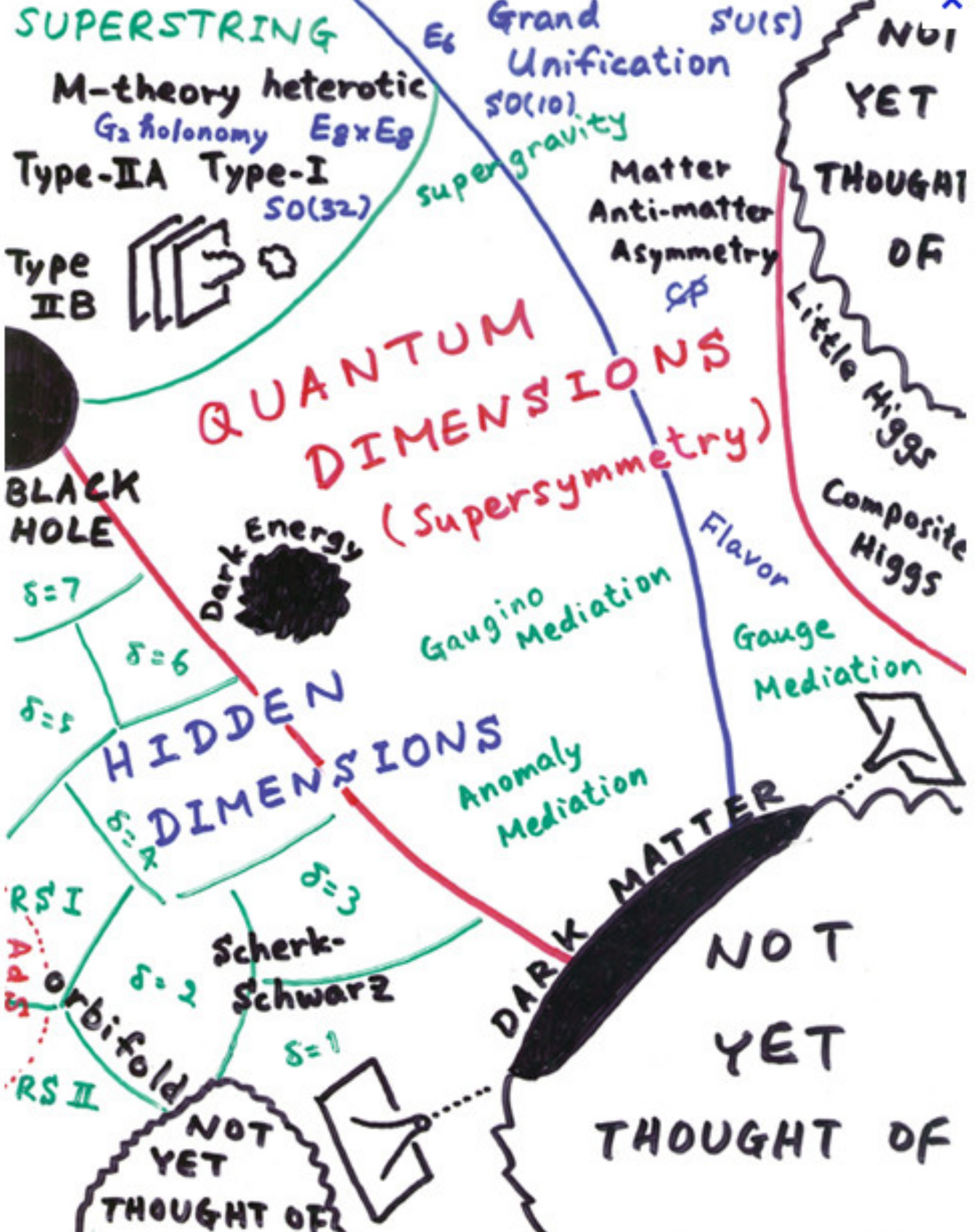
input



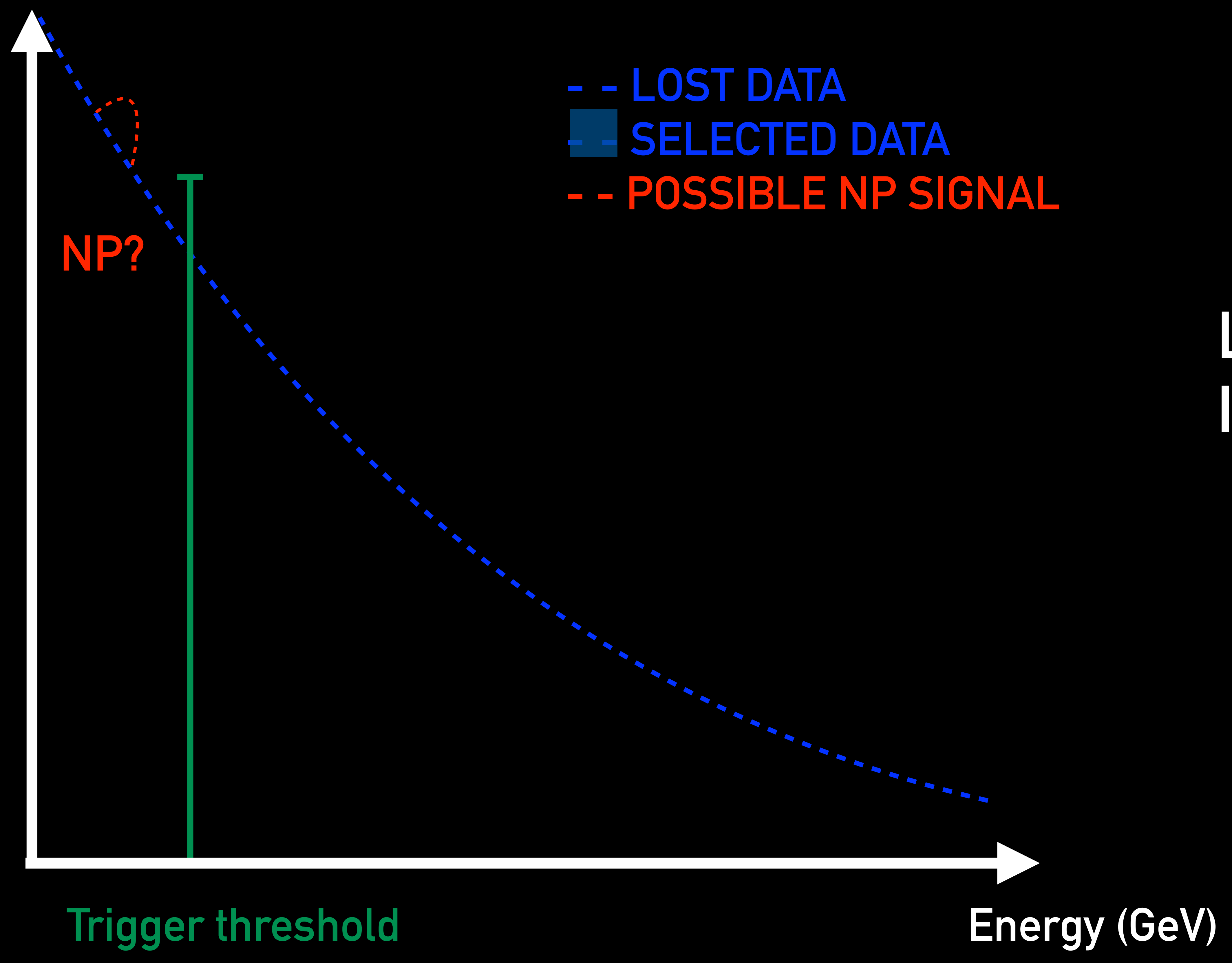
output

Anomaly score = input - output


CMS Preliminary
LHCP 2017
Vector mediator
Dirac DM
EXCLUDED
(a) CMS
Dark matter mass m_{DM} [GeV]
Mediator mass M_{med} [GeV]
DM Simplified Model Exclusions
ATLAS Preliminary March 2017
35.9 fb^{-1} (13 TeV)
(b) ATLAS
2016 diboson combination
CMS
Combination
HVT model B
$m_{W'}$ (GeV)
CMS Preliminary
LHCP 2020
Axial-vector mediator
Fall 2020
132-140 fb^{-1} (13 TeV)
$c\tau_0$ (mm)
Z' mass (GeV)
Pseudoscalar, Dirac, g_q=1, g_χ=1
$Z' \to q\bar{q}$
$M_{Z'}$ [GeV]

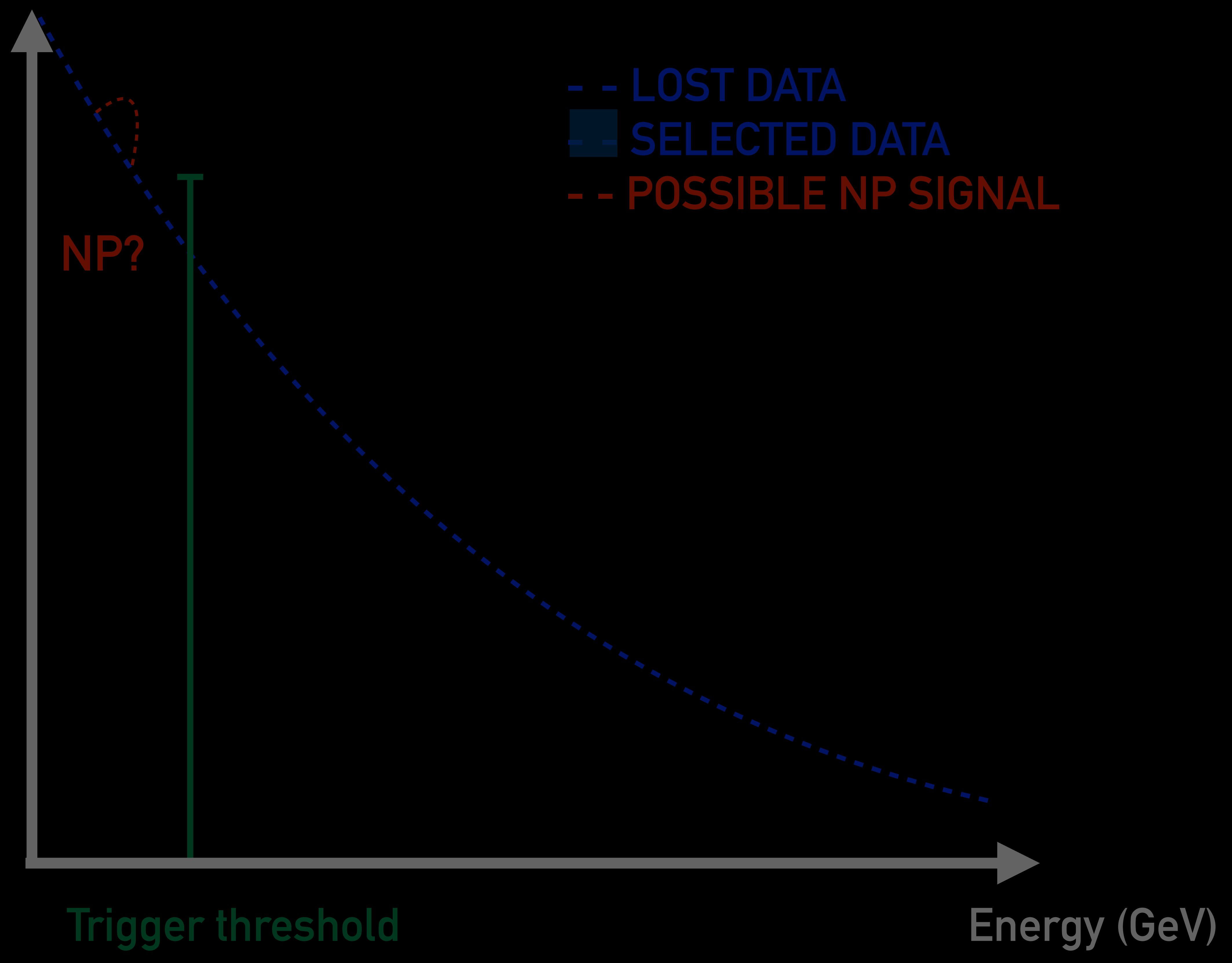

SUPERSTRING
M-theory
heterotic
G2 holonomy
E8xE8
Type-IIA
Type-I
SO(32)
Type IIB
BLACK HOLE
E6
Grand Unification
SU(5)
SO(10)
supergravity
Matter Anti-matter Asymmetry
CP
NOT YET THOUGHT OF
Little Higgs
Composite Higgs
QUANTUM DIMENSIONS
(Supersymmetry)
Dark Energy
Gaugino Mediation
Flavor
Gauge Mediation
δ=7
δ=6
δ=5
HIDDEN DIMENSIONS
δ=4
Anomaly Mediation
δ=3
RS I
AdS
orbifold
δ=2
Scherk-Schwarz
δ=1
RS II
DARK MATTER
NOT YET THOUGHT OF
NOT YET THOUGHT OF

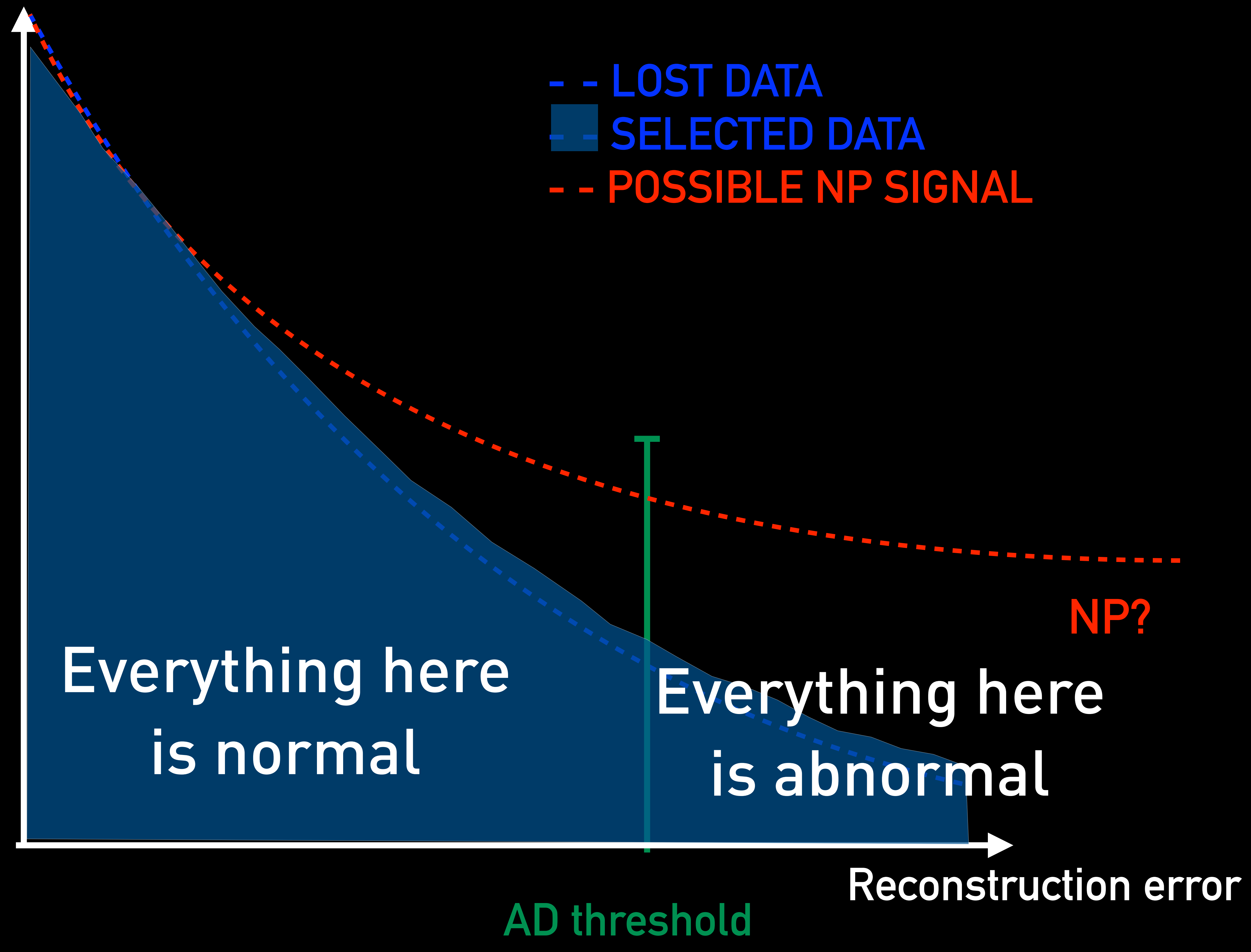
Anomaly Detection triggers



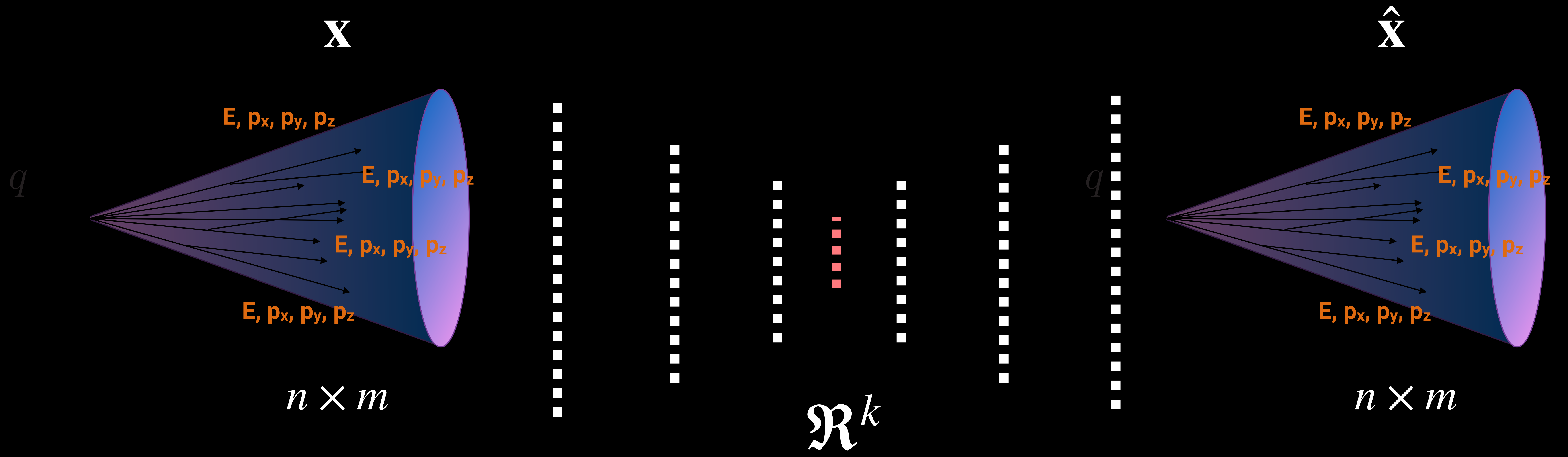
Level-1 rejects >99% of events!
Is there a smarter way to select?

Anomaly Detection triggers





Outlier detection

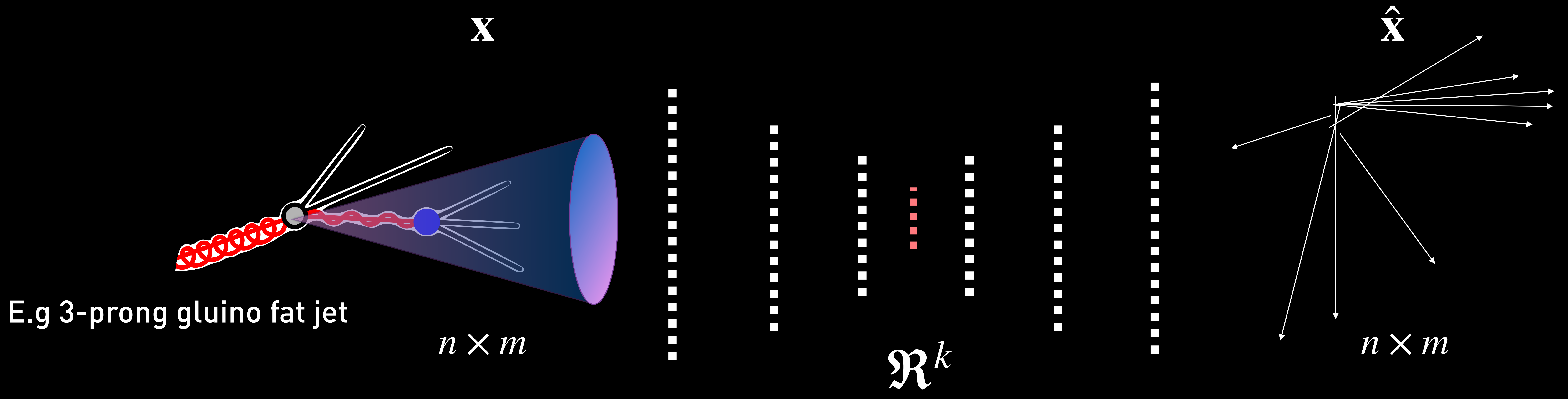


Compressed representation of x.

Latent space $\mathfrak{R}^k$, k < m×n

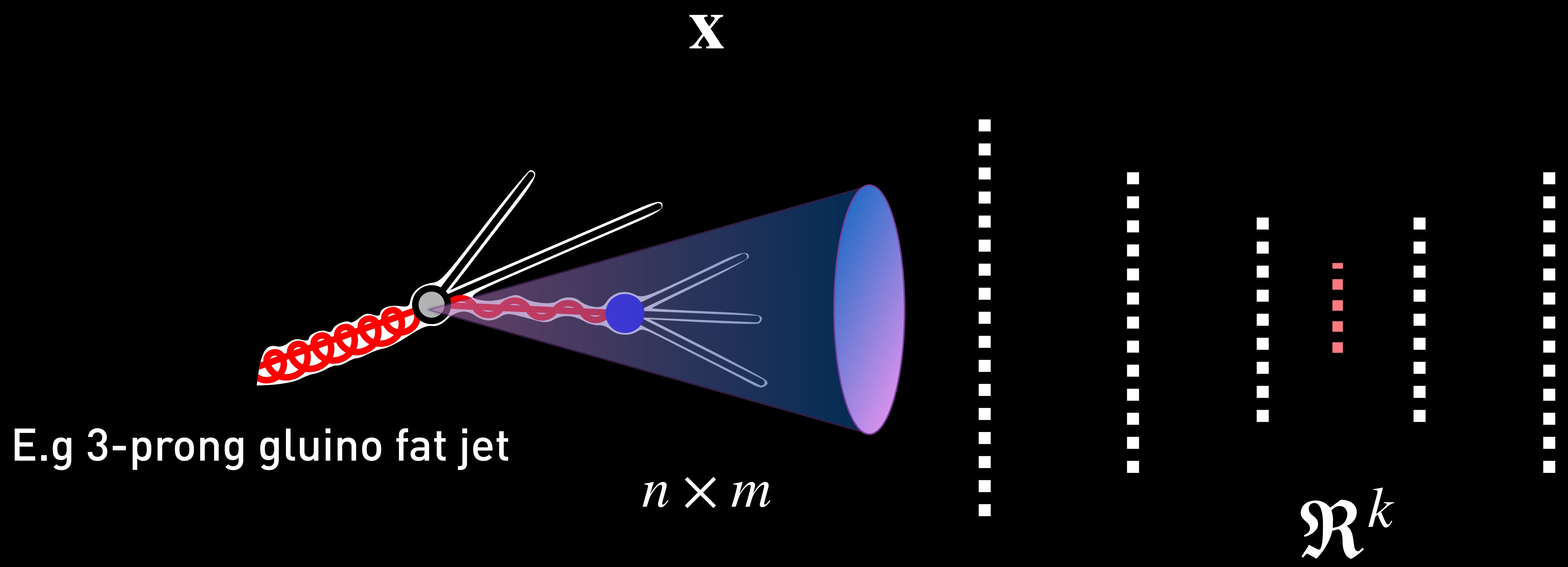
prevents memorisation of input, must learn

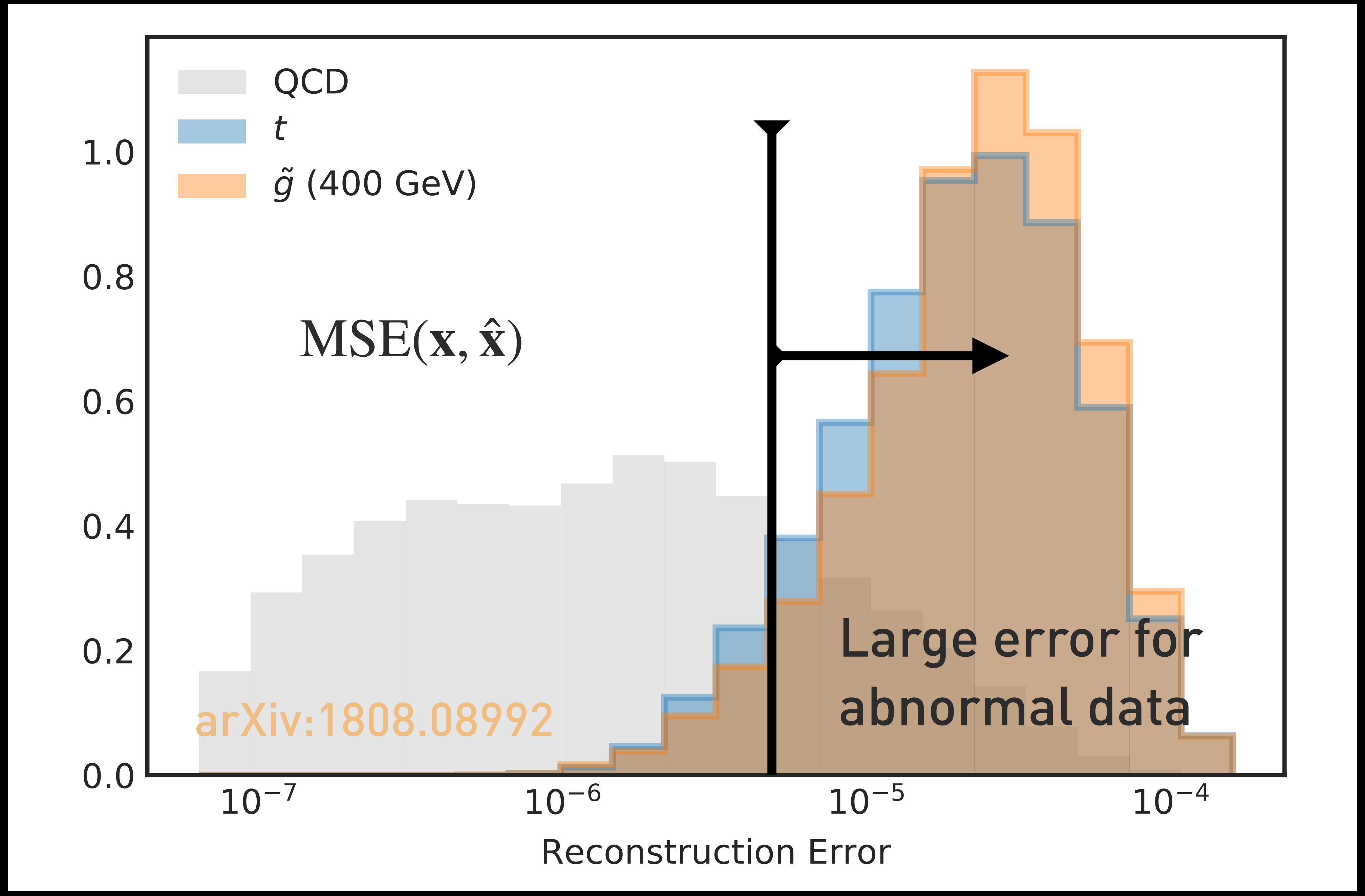
Outlier detection

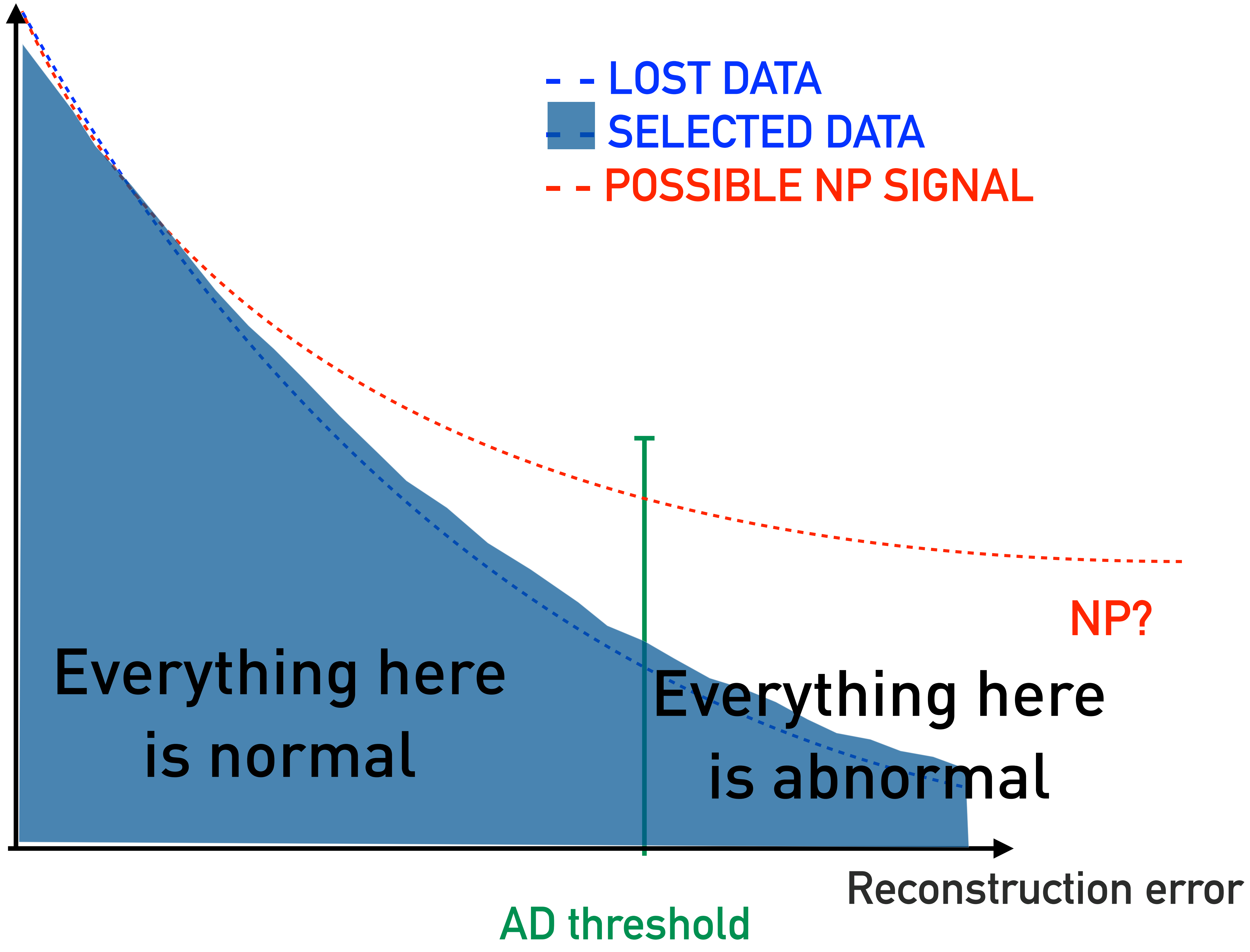


$\mathscr{L}(\mathbf{x}, \hat{\mathbf{x}})$ is Mean Squared Error$(\mathbf{x}, \hat{\mathbf{x}})$, "high error events" proxy for "degree of abnormality"

Outlier detection







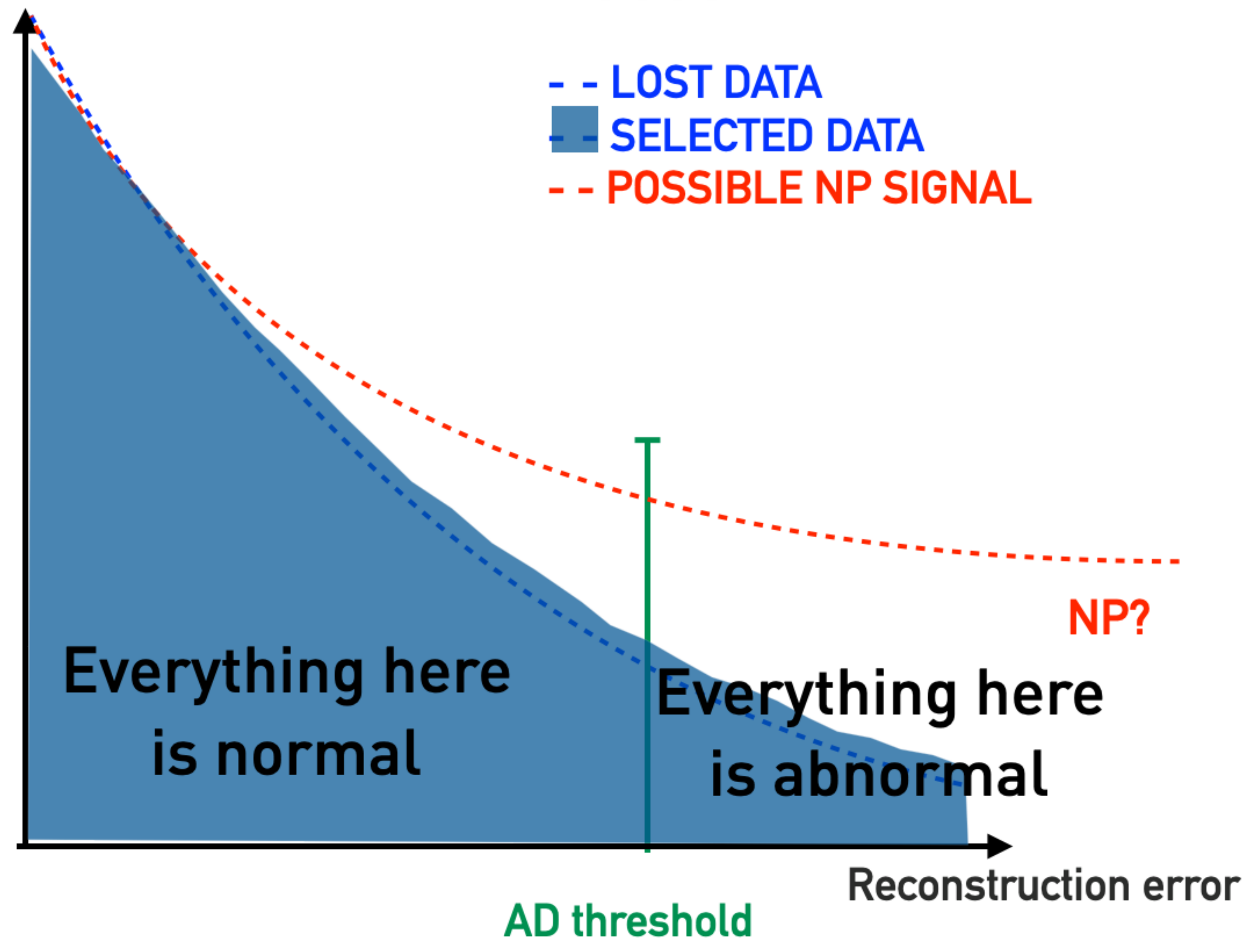
AXOL1TL

....in 50 nanoseconds!
Currently recording 300 collisions per second in CMS!

First ML triggers in ATLAS and in CMS in 2024

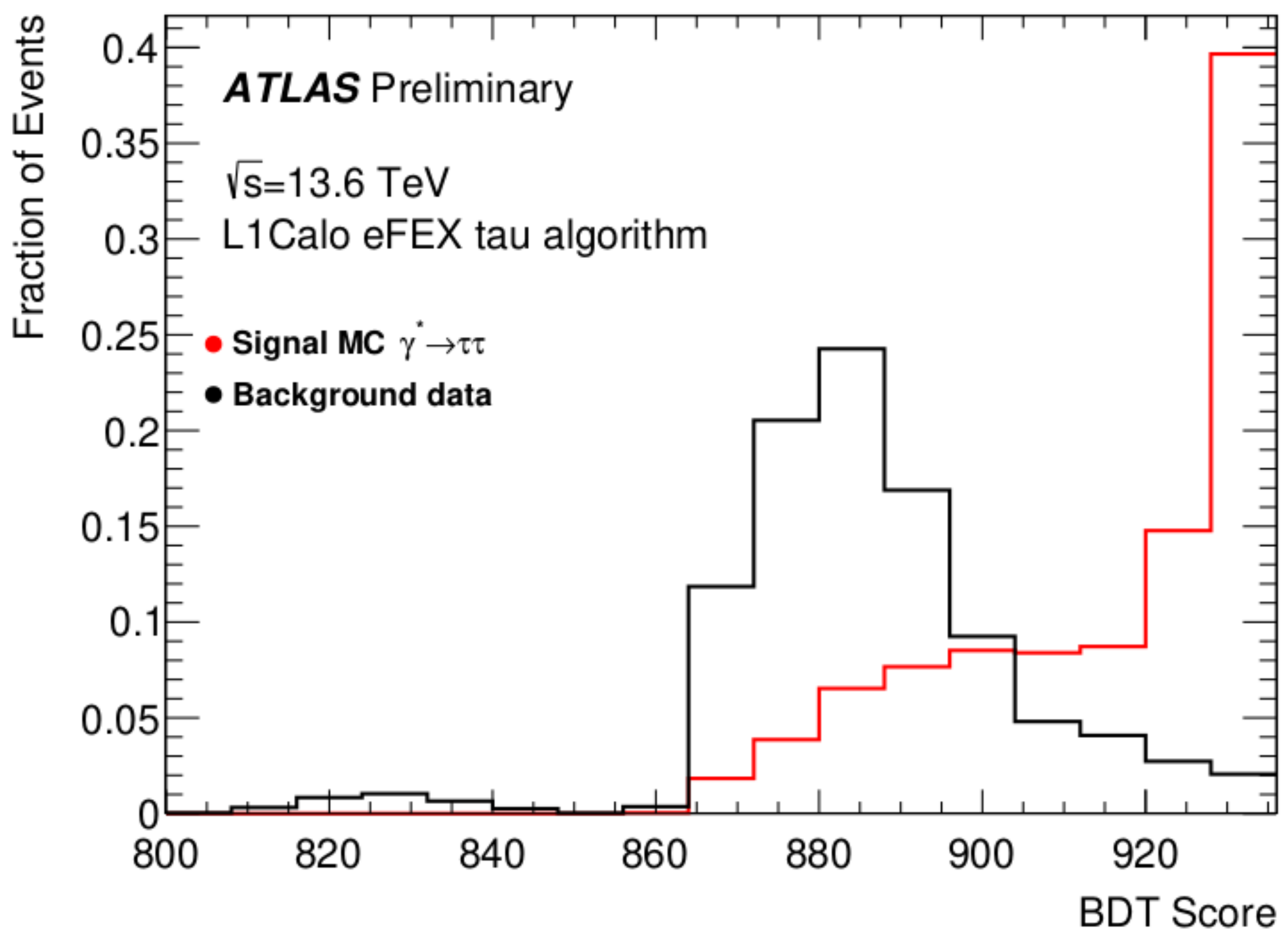
DNN anomaly detection in 50 ns

hls4ml



BDT selecting τ's <100 ns

Conifer

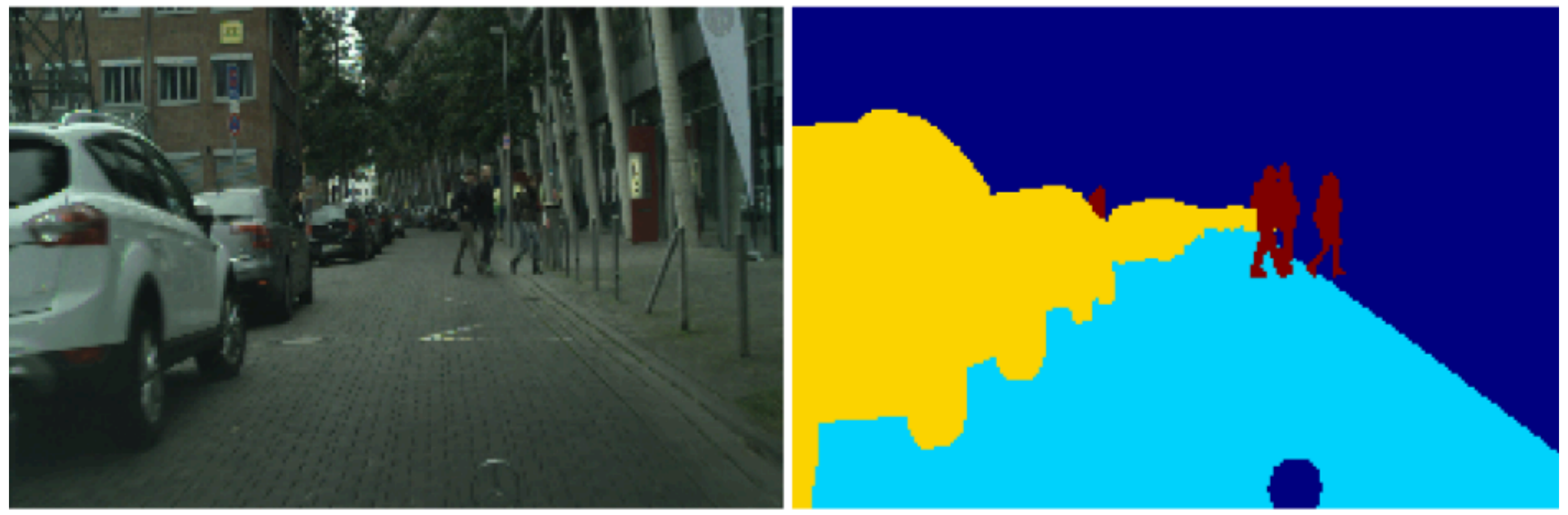


CMS DP2023_079

L1CaloTriggerPublicResults

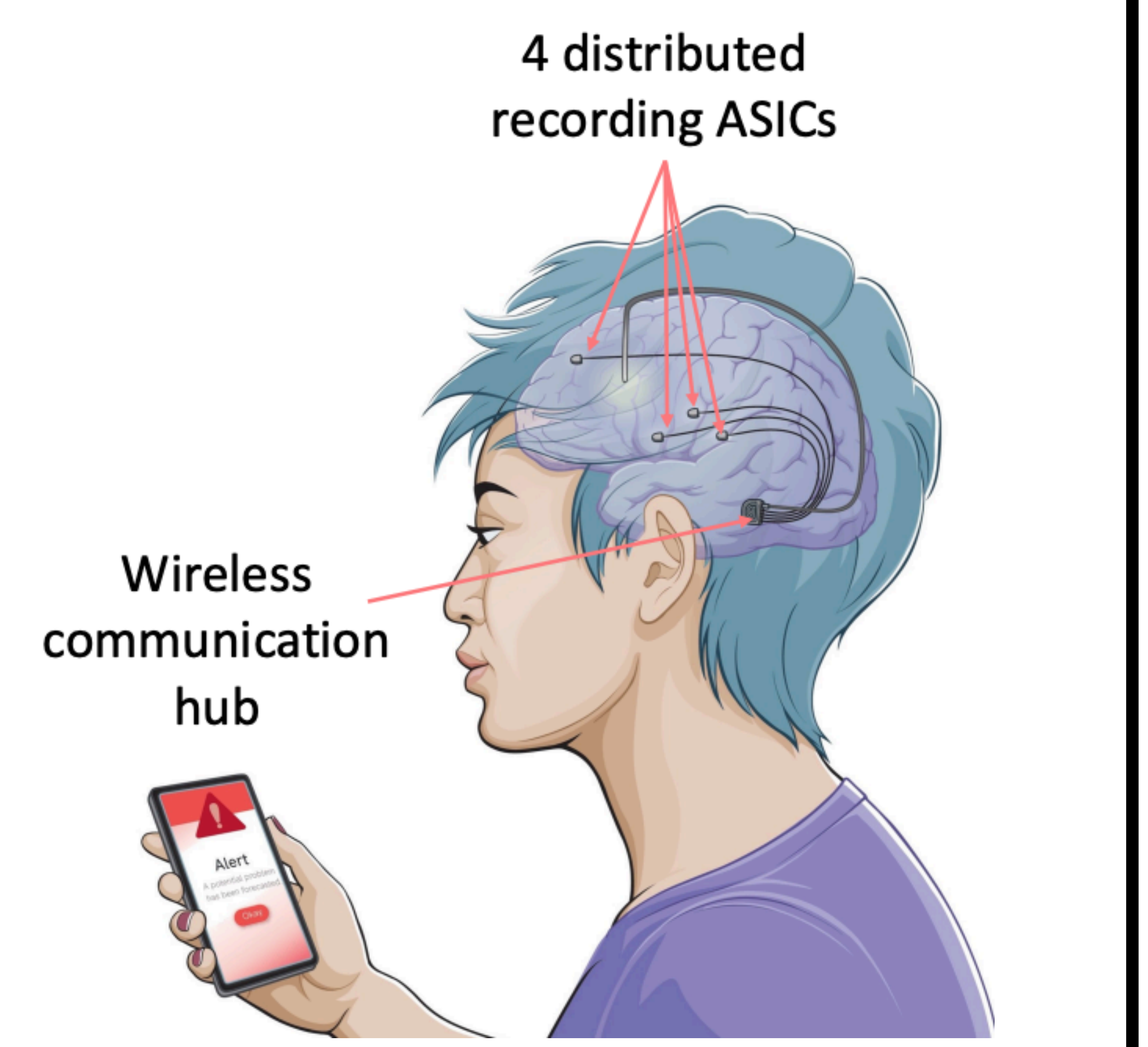
...and outside

Semantic segmentation for autonomous vehicles



N. Ghielmetti et al.

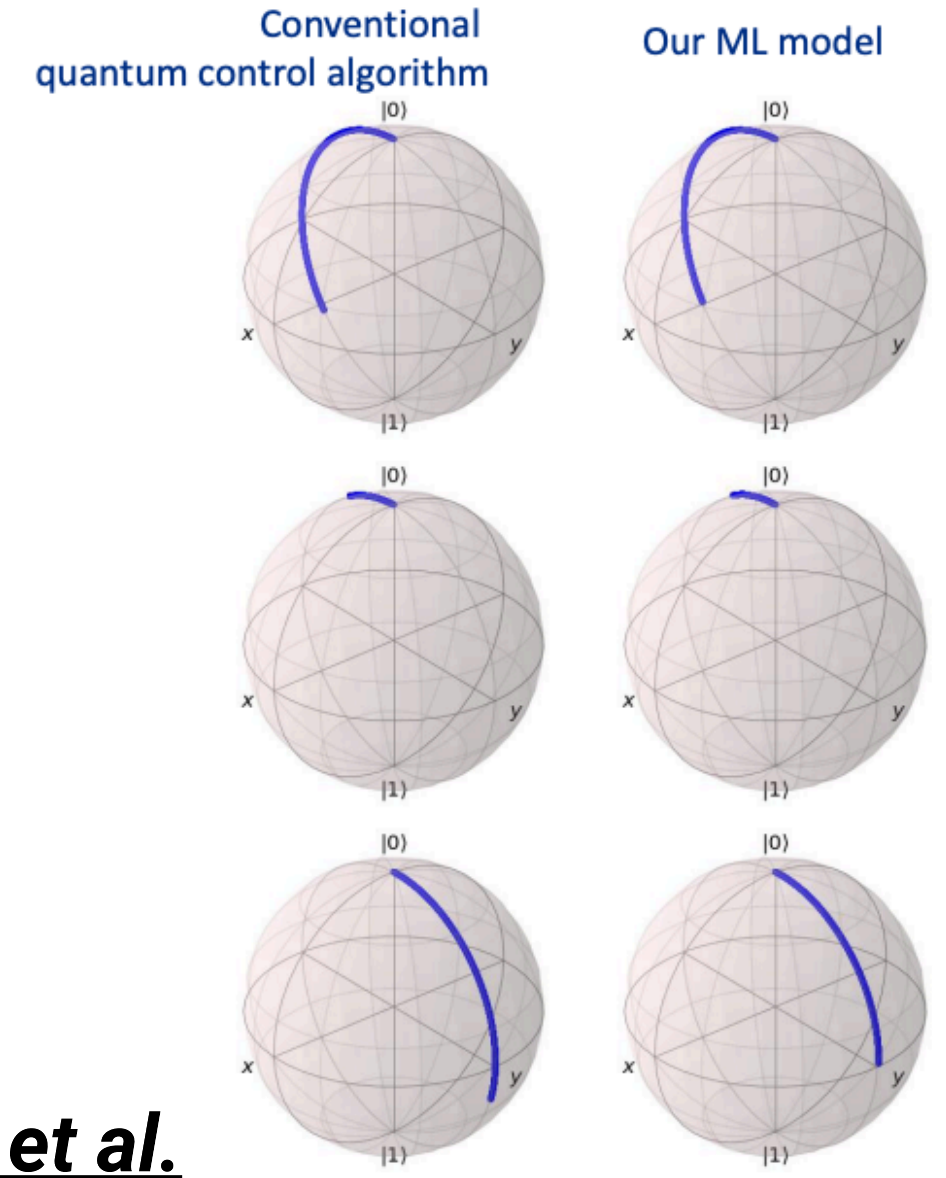
Seizure Predicting Brain Implant



W. Lemaire et al.

NN accelerator for quantum control

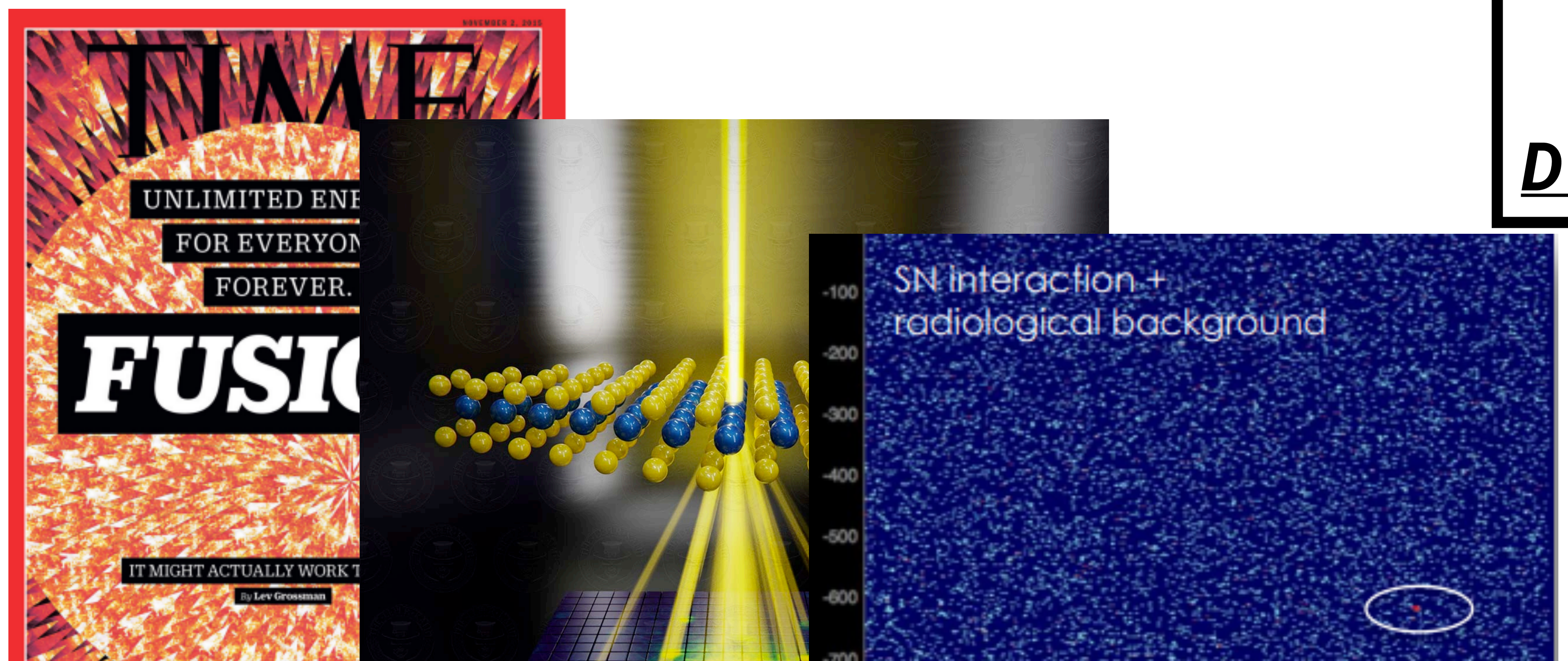
- Putting control in cryostat (e.g optimal pulse parameters)



D Xu et al.

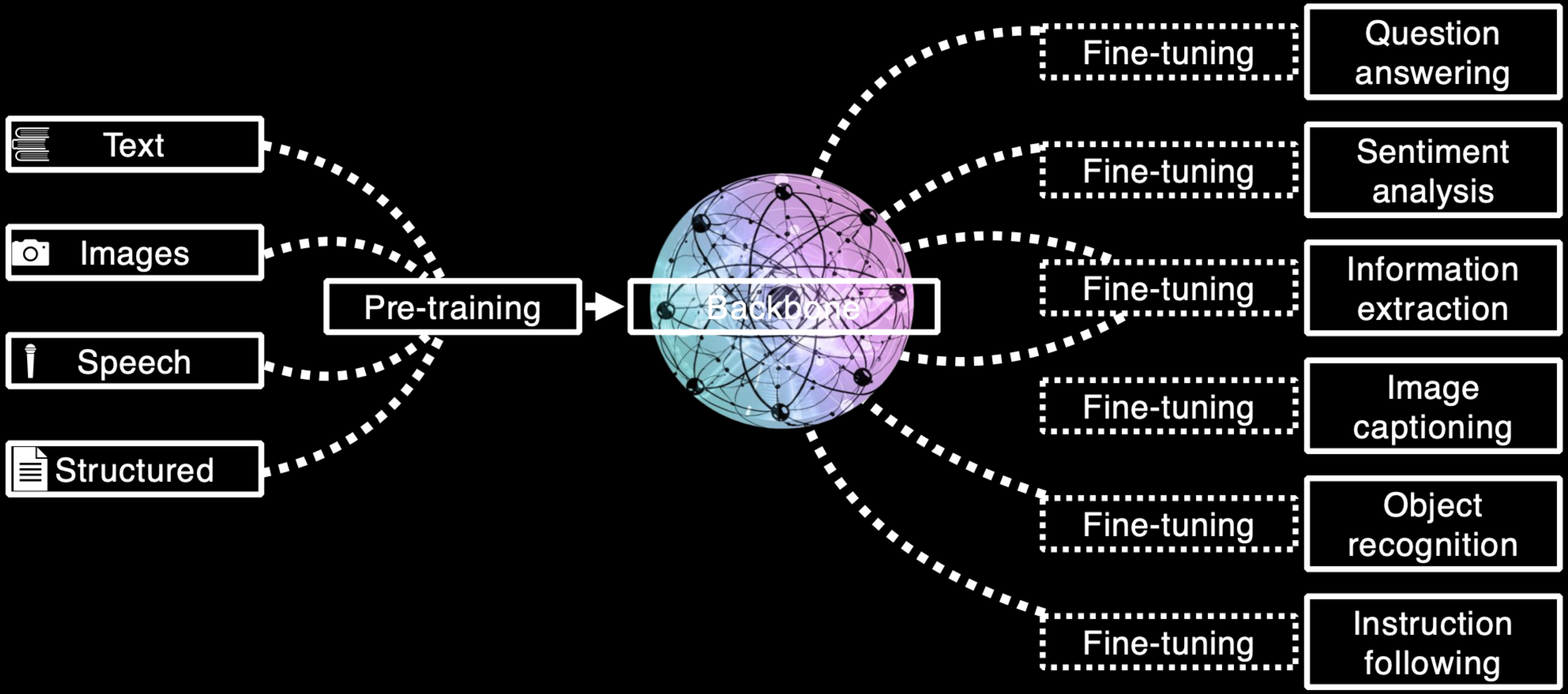
Other examples

- ***For fusion science phase/mode monitoring***
- ***Crystal structure detection***
- ***Triggering in DUNE***
- ***Accelerator control***
- ***Magnet Quench Detection***
- ***MLPerf tinyML benchmarking***
- ***Food contamination detection***
- etc….



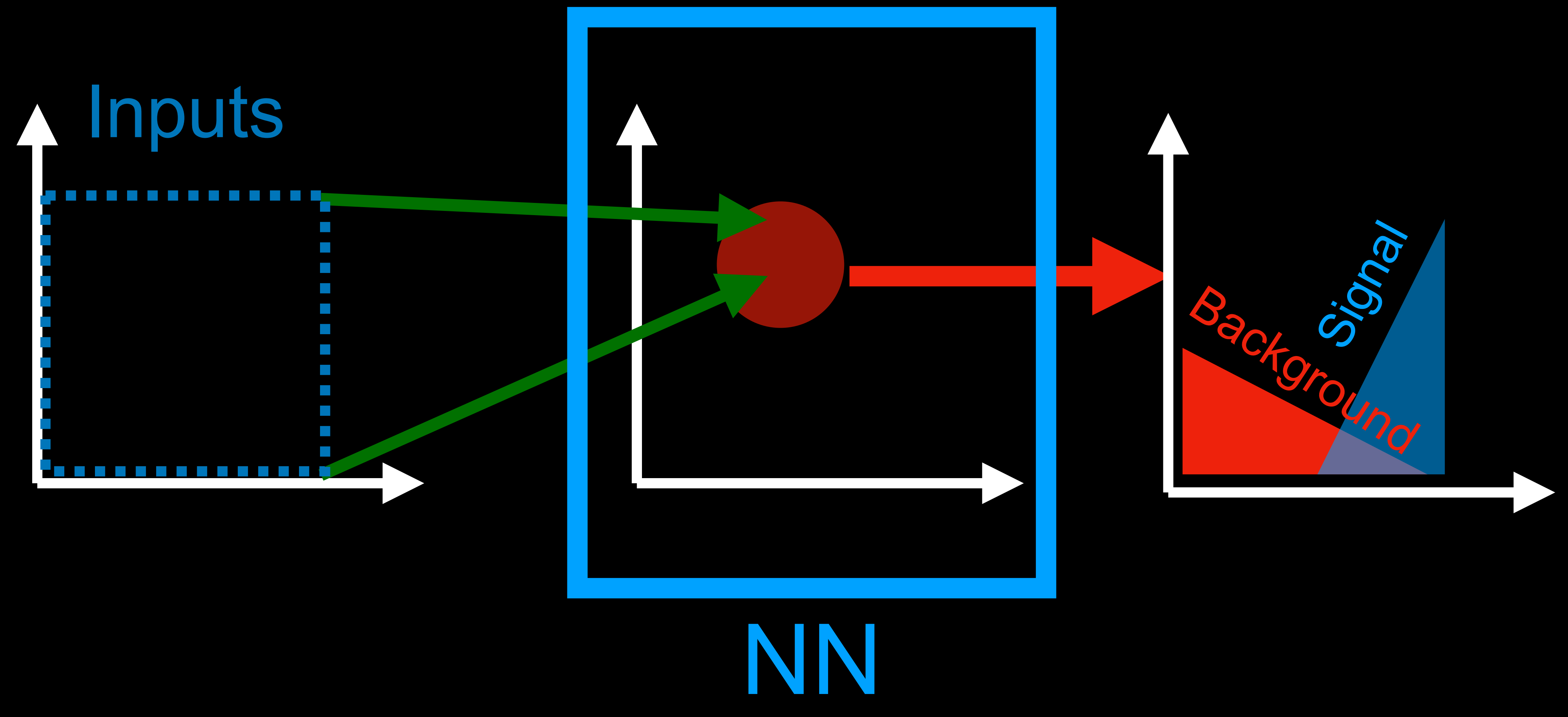
Foundation models

Foundation models


Text
Images
Speech
Structured
Pre-training
Backbone
Fine-tuning
Question answering
Fine-tuning
Sentiment analysis
Fine-tuning
Information extraction
Fine-tuning
Image captioning
Fine-tuning
Object recognition
Fine-tuning
Instruction following

Heterogeneous detector
Multi-modal input!
Pre-training
Backbone
Fine-tuning
Jet reconstruction
Fine-tuning
Electron reconstruction
Fine-tuning
Pile-up removal
Fine-tuning
Missing energy computation
Fine-tuning
Anomaly Detection
Fine-tuning
0/1?
Generate simulation?

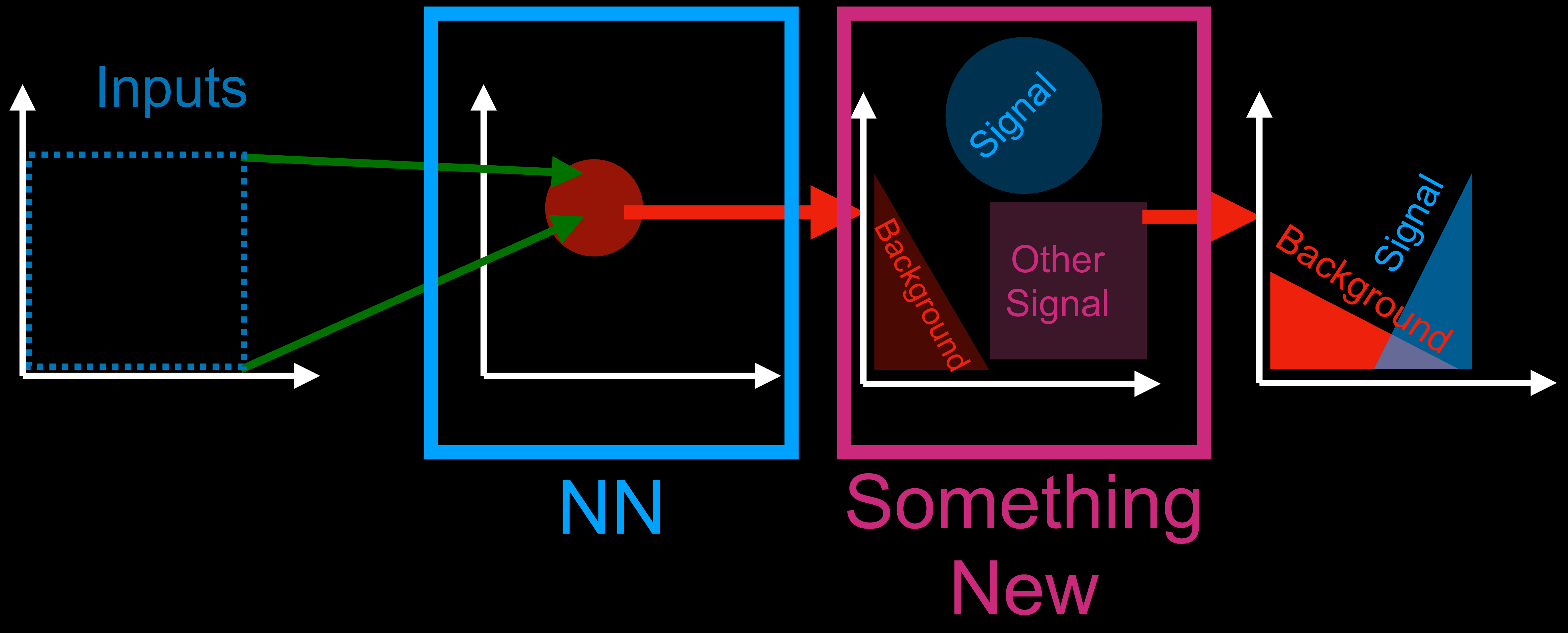
Too many models, too little learning?



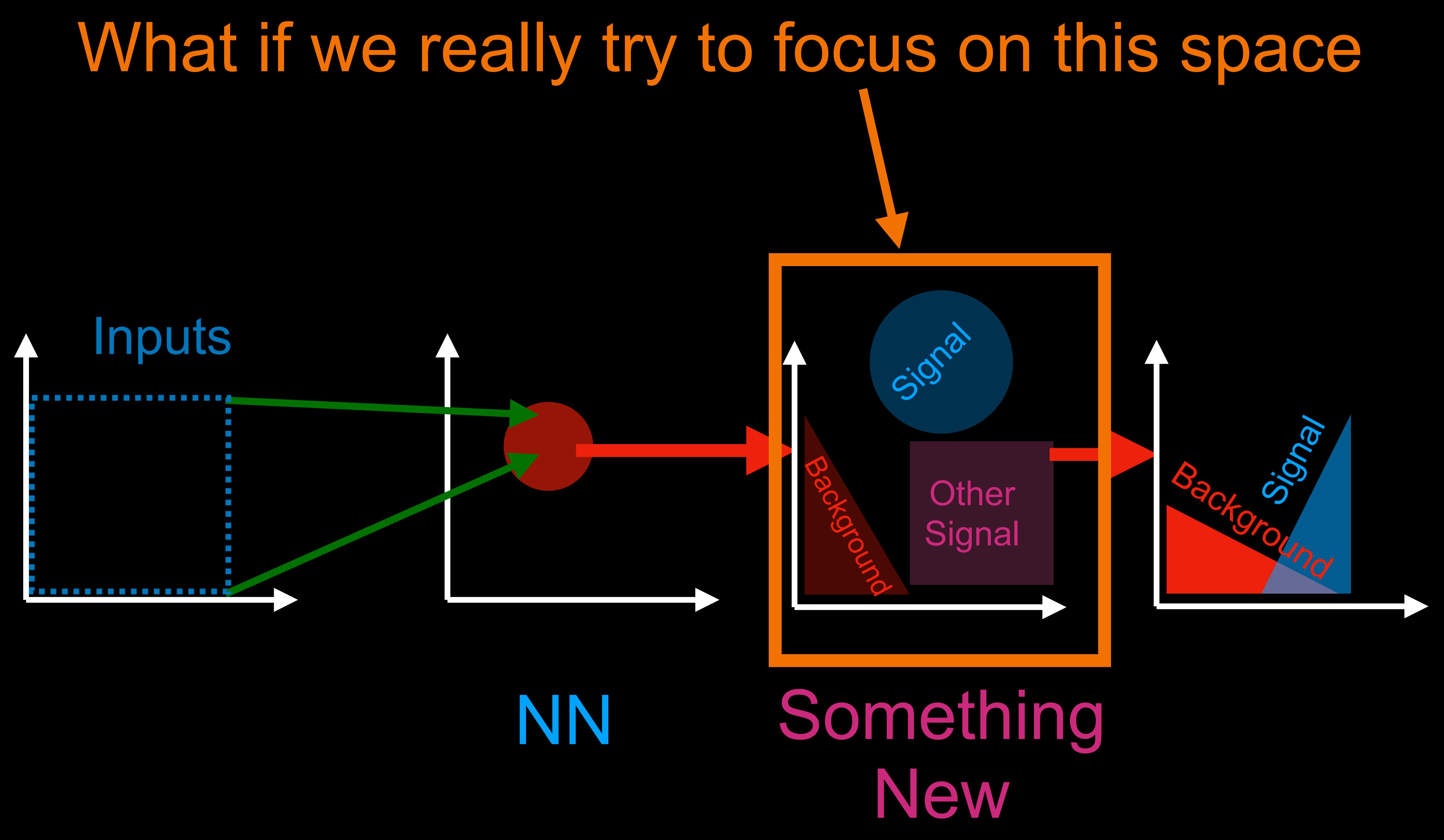
$x = (x_1, x_2)$ $f(x; w^*)$ $\hat{y}$

Discrimination

Instead of features like “says meow”, can we make new and better features?

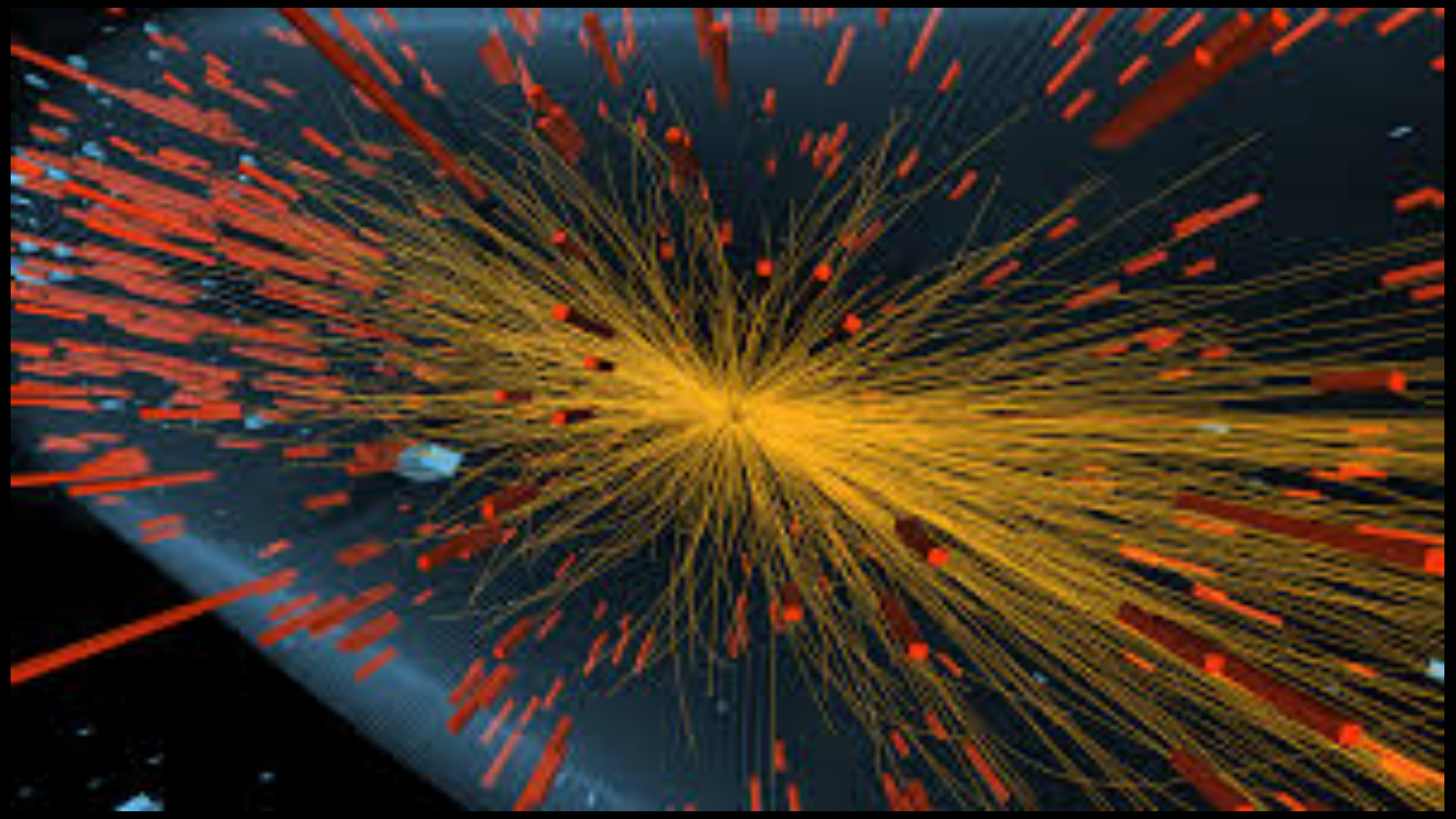


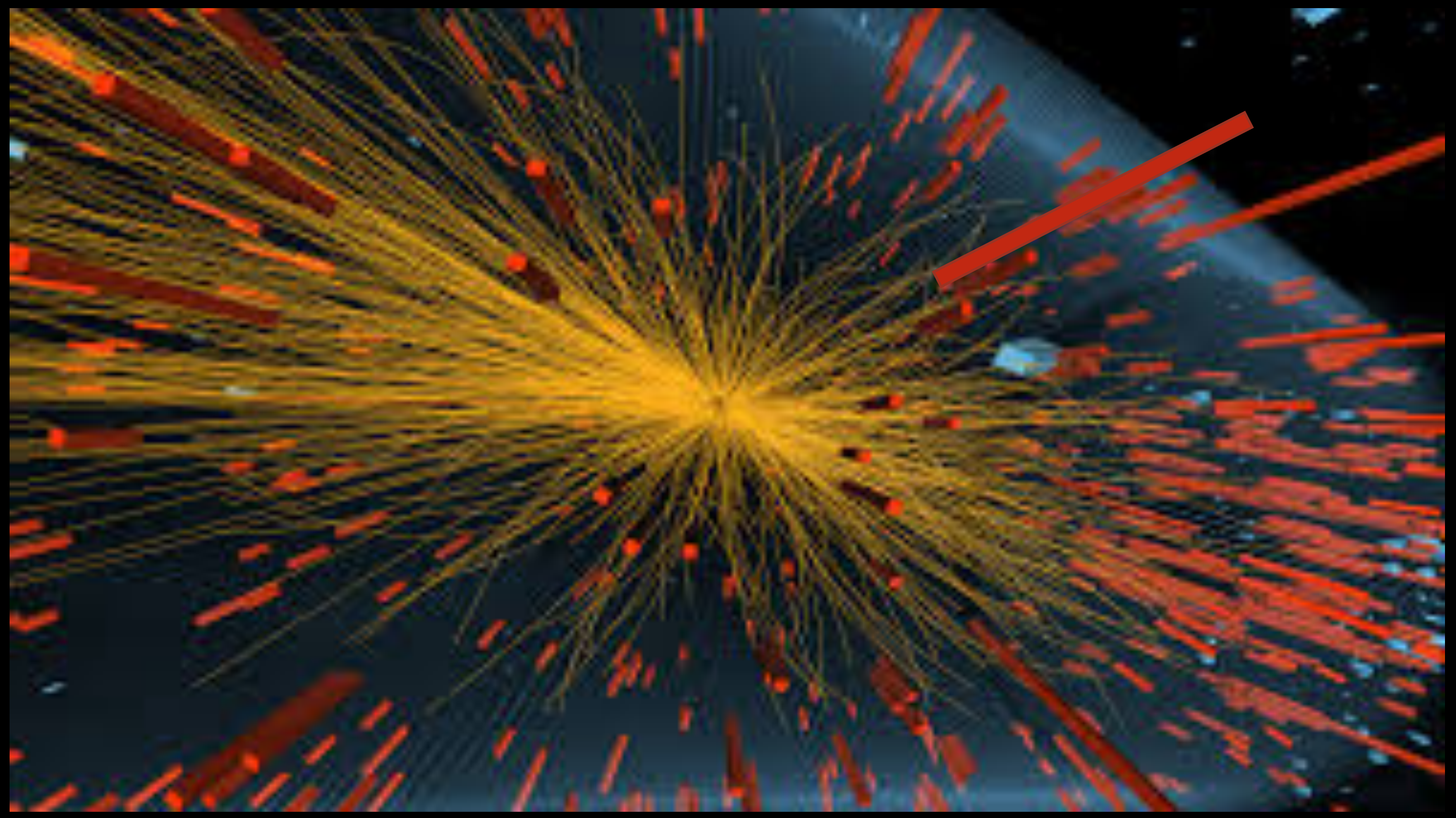
Metric Learning



Neural embedding

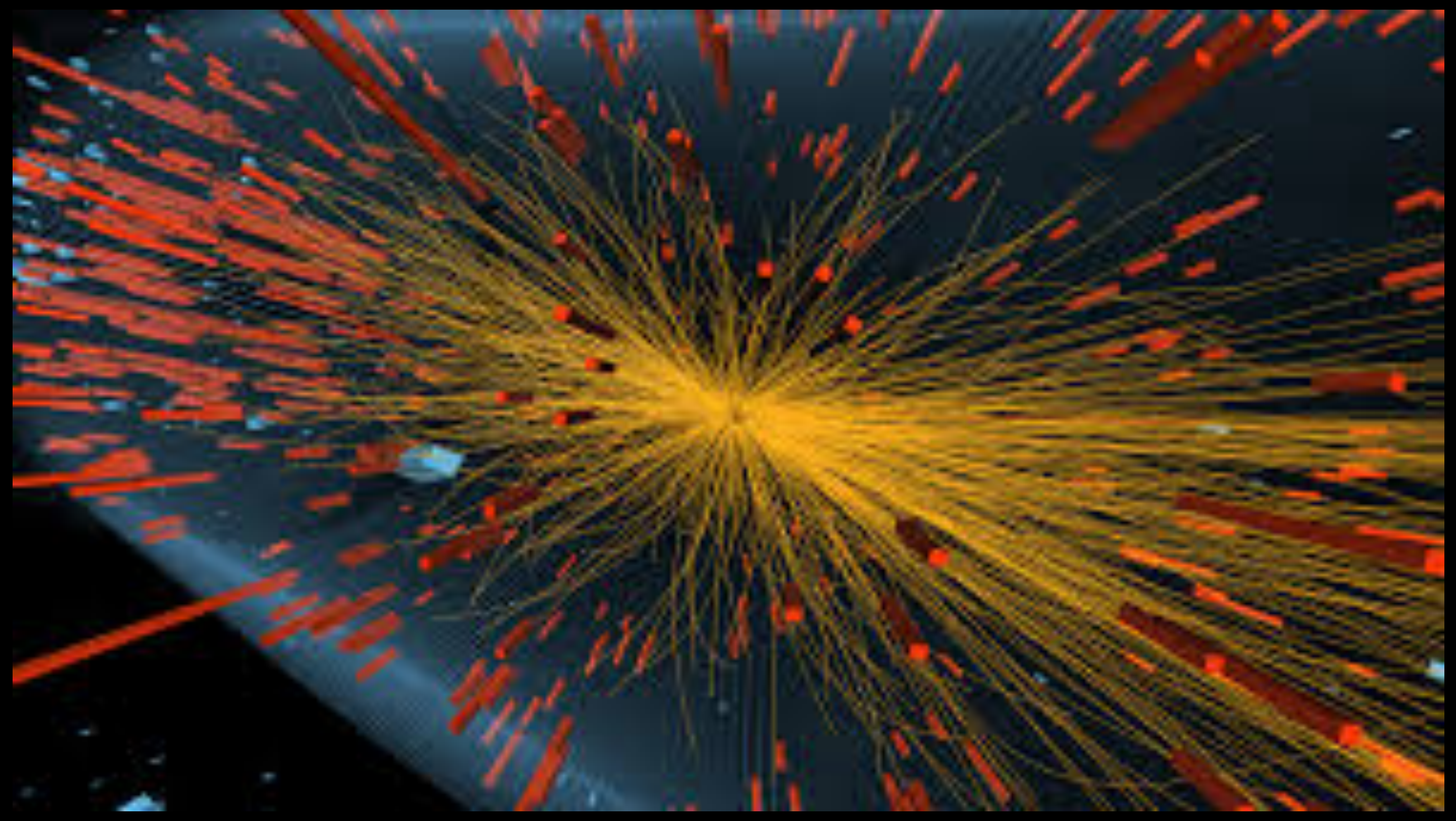
Learning the space

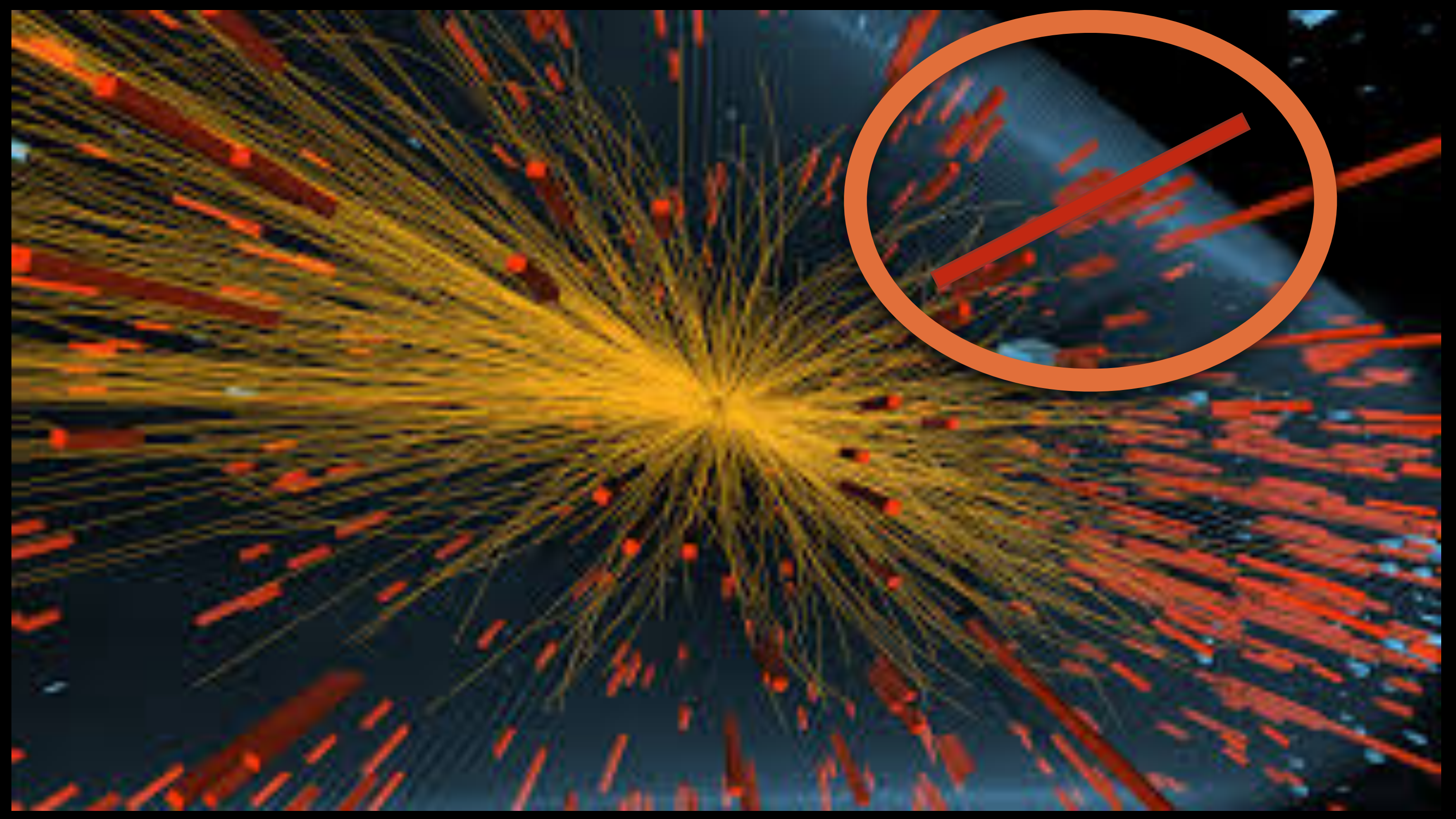


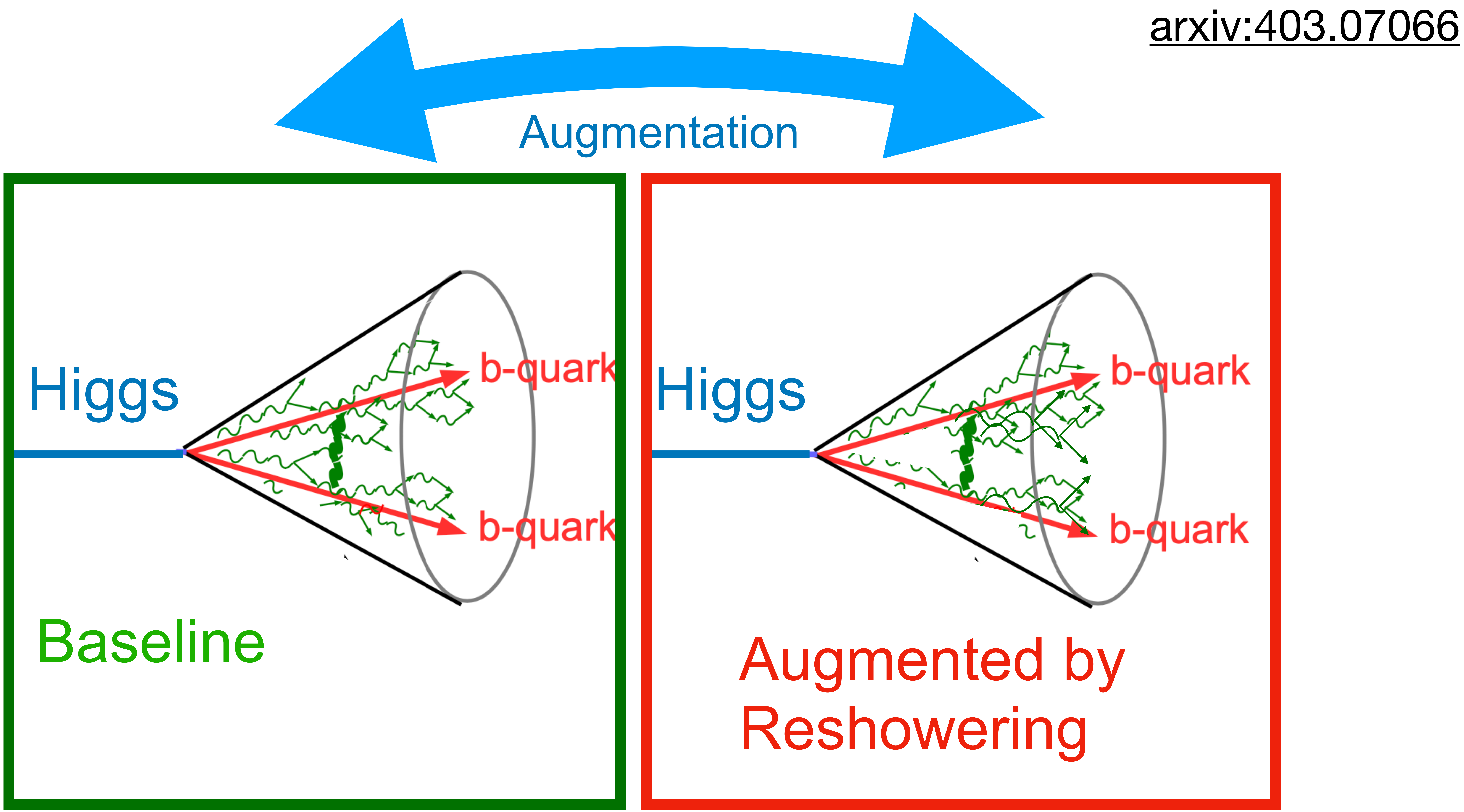


Learning the space

- By looking at data, we can learn a lot
  - Go over input piece by piece
  - Analyze every aspect
  - Compare every feature
- Find distinctive style of the input
  - can be done e.g by looking for a deviation



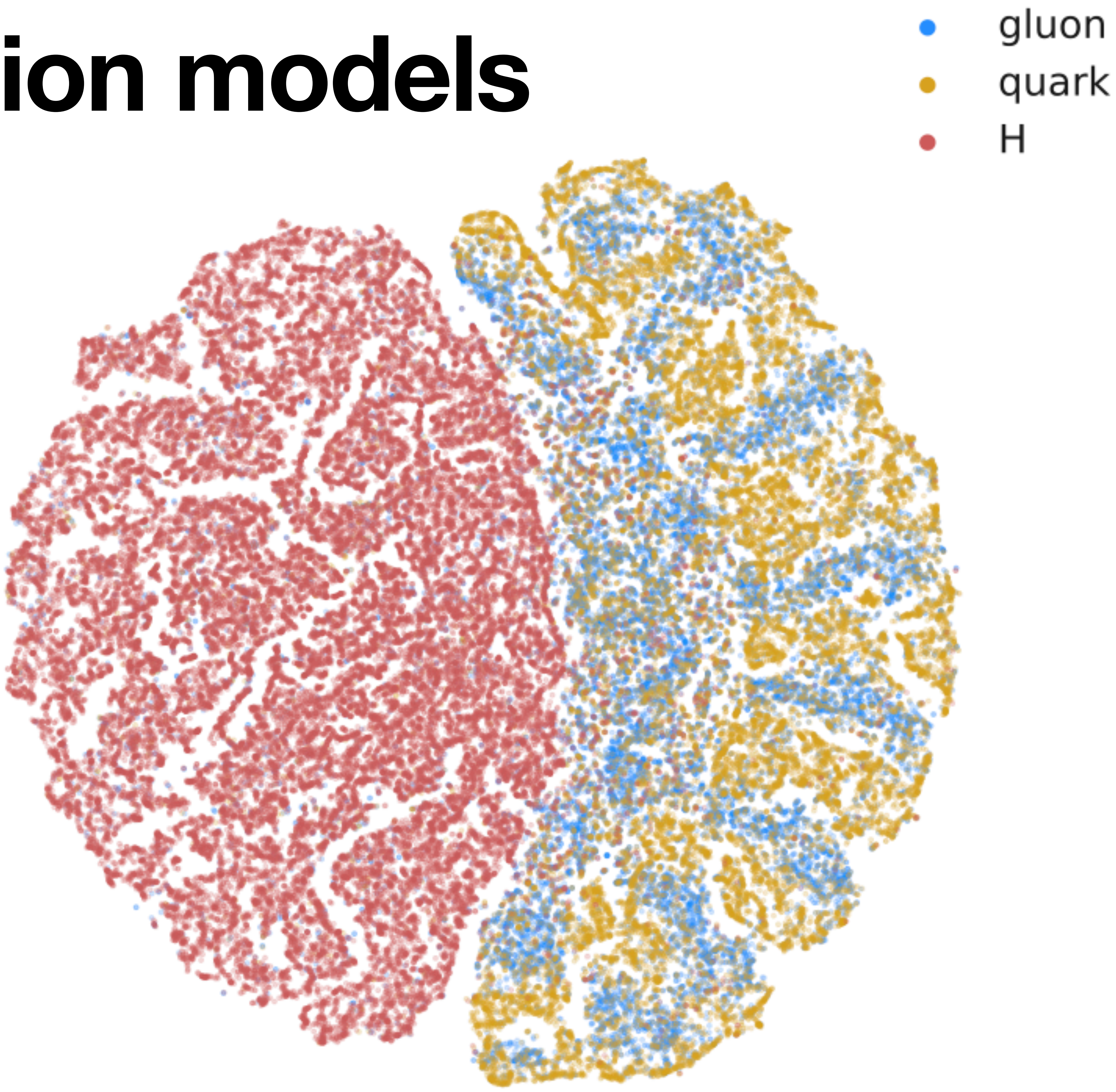




Embedded Space can use any NN to embed

arxiv:403.07066

QM foundation models



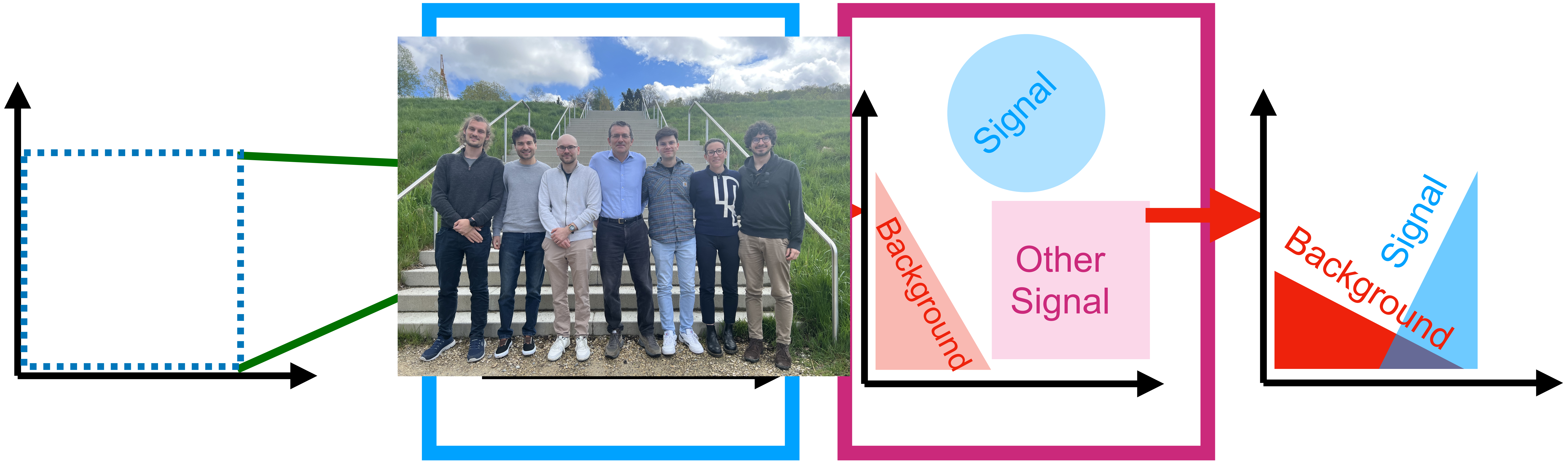
→ embedding quantum mechanics into AI algorithm

arxiv:403.07066

Physics feature
(says meow)

Theorists

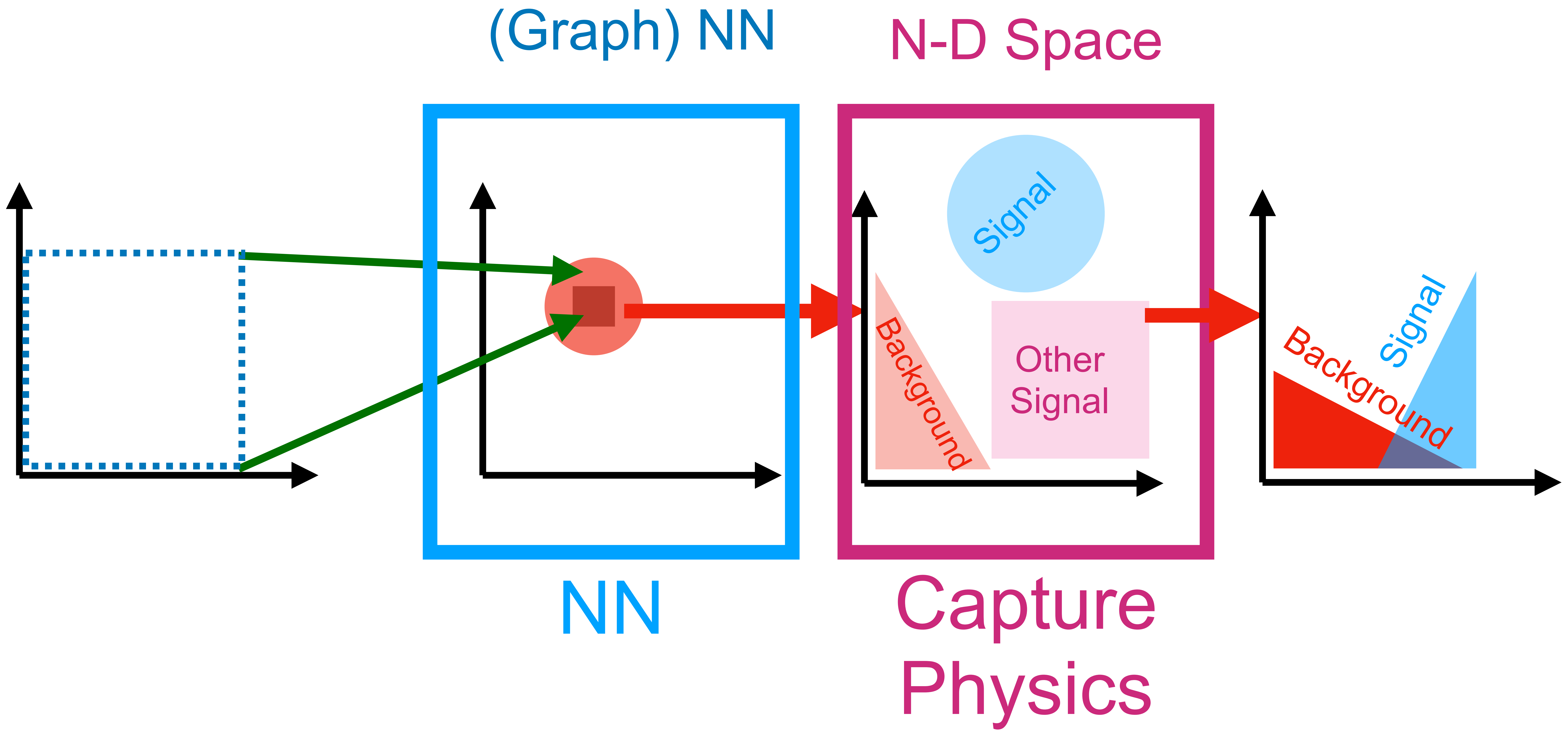
N-D Space



Capture
Physics

arxiv:403.07066

We can replace the QCD theorist with a NN
(And it works better)



Detector design, data acquisition and triggering
Detector reconstruction and tagging
AI!
Foundation models
Data analysis
Generative models for simulation
O(10^10)
100m
DIGI+RECO
[Hard & Slow]
19.7 fb^-1 (8 TeV) + 5.1 fb^-1 (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
±1σ
±2σ
μ̂ = 1.14 +0.26 −0.23
m̂_H = 124.70 ± 0.34 GeV
O(10)
O(10^3)
10^-18 m
10^-15 m
10^-6 m
GEN
pp collisions up to
[Easy & Fast]
detector response
GEANT4