



Machine learning and particle physics

P. Zurita



Second IPARCOS workshop on Machine Learning and
Applications to Physics
20/12/2023, Madrid, Spain



Outline

- HEP for non-“*HEPsters*”: the Standard Model
- ML applications
 - theory
 - phenomenology
 - simulations
 - experiments
- Summary

The standard model

HEP is described by the ***Standard Model*** of particle physics

HEP is described by the ***Standard Model*** of particle physics

leptons: particles with no internal structure and spin $1/2$

electron, muon, tau: with electric charge
3 neutrinos

HEP is described by the ***Standard Model*** of particle physics

leptons: particles with no internal structure and spin $1/2$

electron, muon, tau: with electric charge
3 neutrinos

quarks: particles with no internal structure and spin $1/2$

up, down, strange, charm, bottom, top: electric and
colour charge

HEP is described by the ***Standard Model*** of particle physics

leptons: particles with no internal structure and spin $1/2$

electron, muon, tau: with electric charge
3 neutrinos

quarks: particles with no internal structure and spin $1/2$

up, down, strange, charm, bottom, top: electric and colour charge

bosons: particles with no internal structure and integer spin

photon, Z, H
gluon: colour charge
 W^+ , W^- : electric charge

QFT tell us how these particles interact with each other.

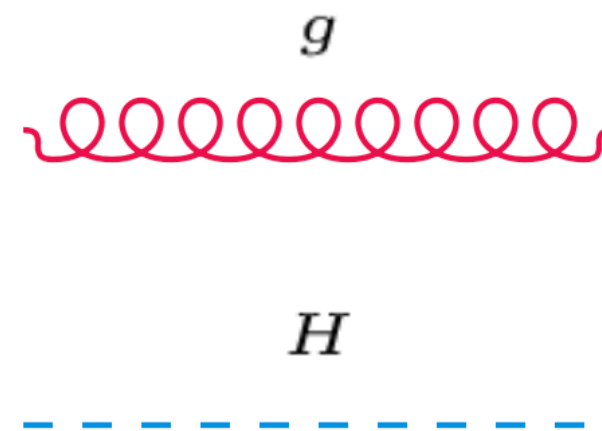
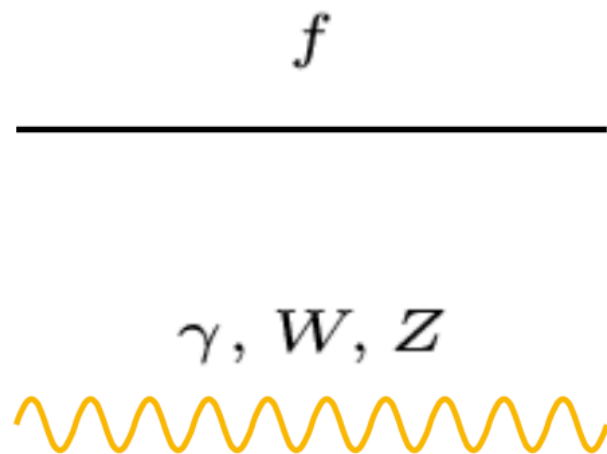
We can't solve the equations exactly, but in some regimes we can do perturbation theory, expanding in the ***running constants***.

And we depict the terms of the perturbative expansion pictorially with Feynman diagrams:

QFT tell us how these particles interact with each other.

We can't solve the equations exactly, but in some regimes we can do perturbation theory, expanding in the ***running constants***.

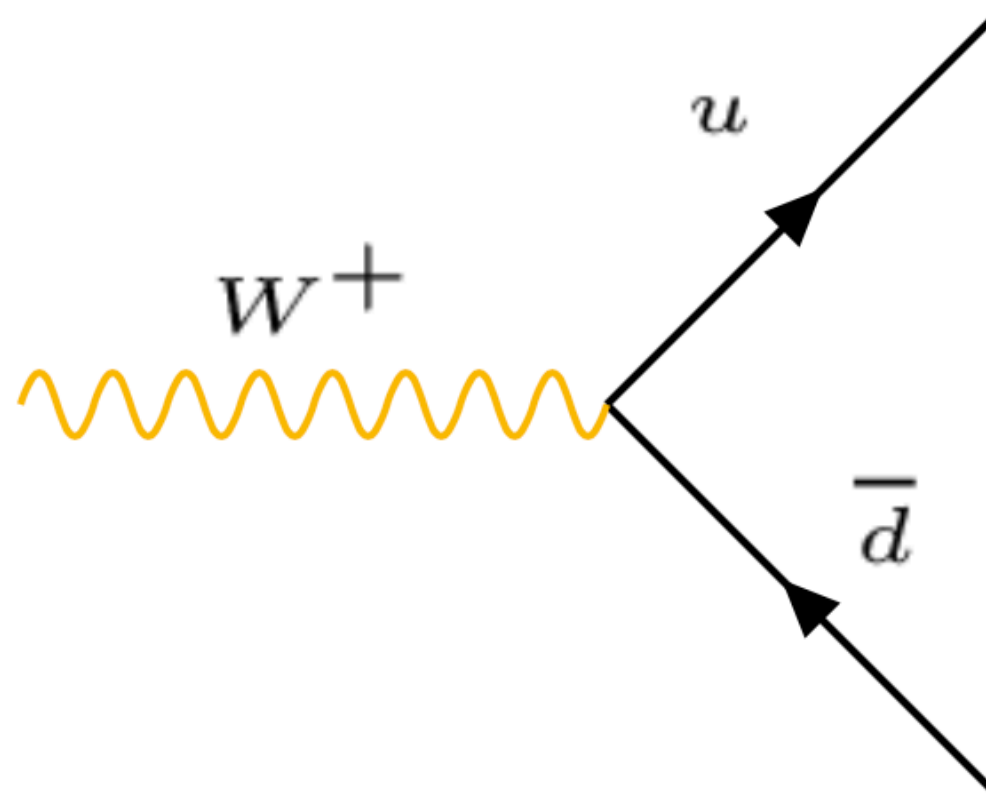
And we depict the terms of the perturbative expansion pictorially with Feynman diagrams:



QFT tell us how these particles interact with each other.

We can't solve the equations exactly, but in some regimes we can do perturbation theory, expanding in the ***running constants***.

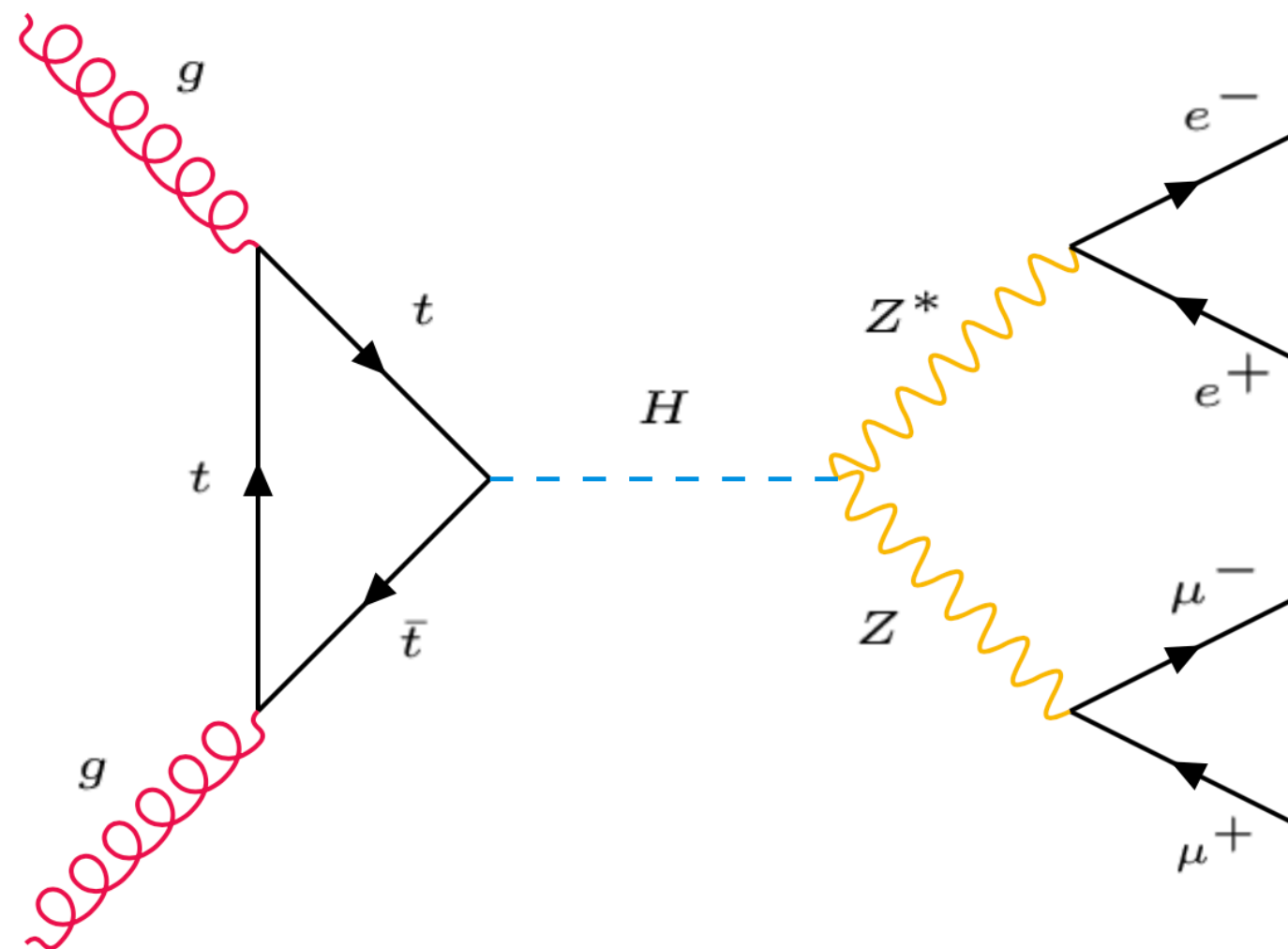
And we depict the terms of the perturbative expansion pictorially with Feynman diagrams:



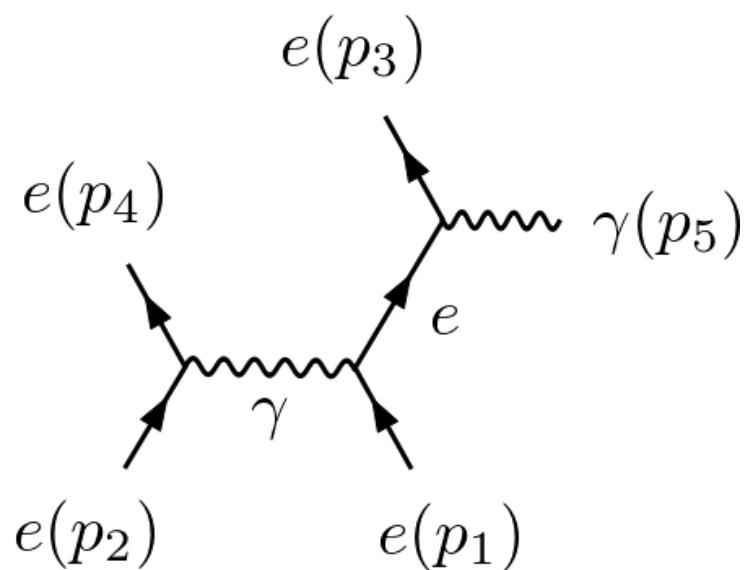
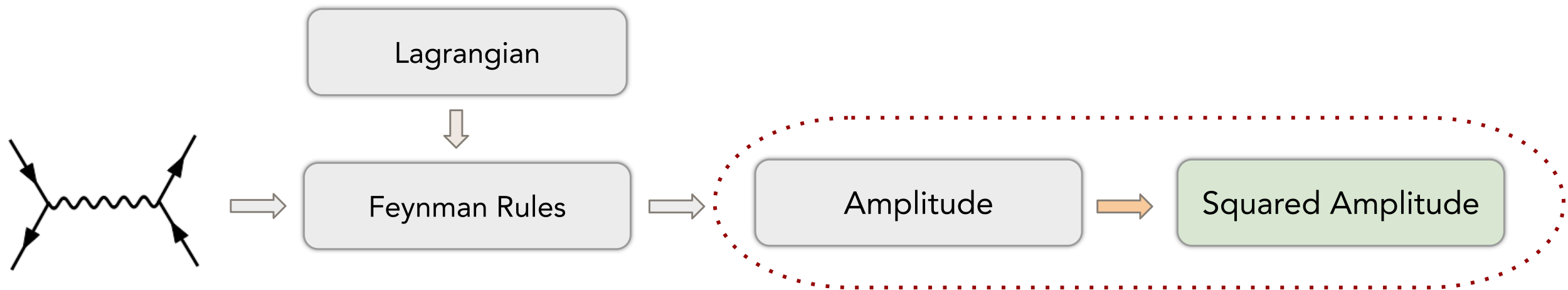
QFT tell us how these particles interact with each other.

We can't solve the equations exactly, but in some regimes we can do perturbation theory, expanding in the **running constants**.

And we depict the terms of the perturbative expansion pictorially with Feynman diagrams:

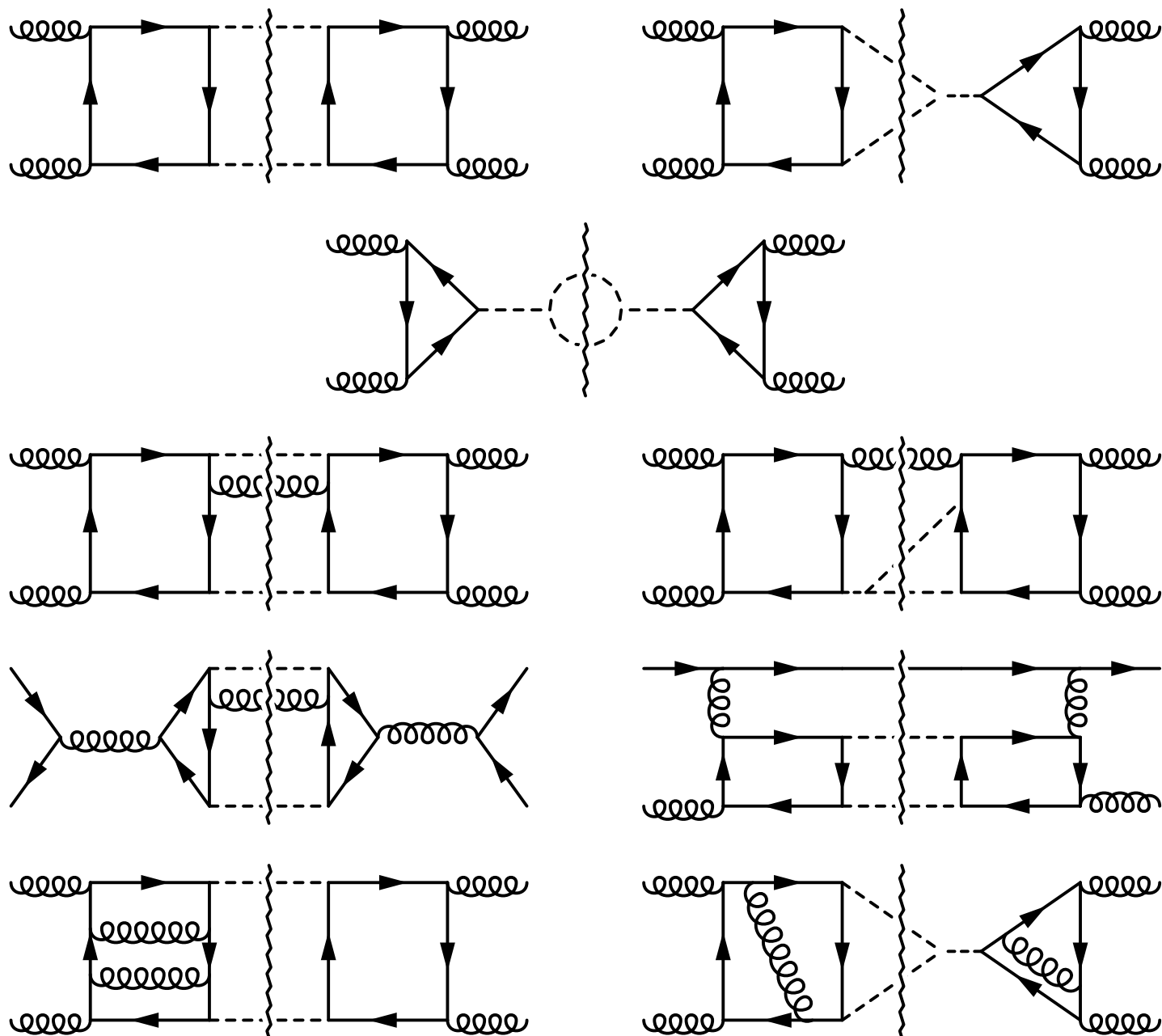
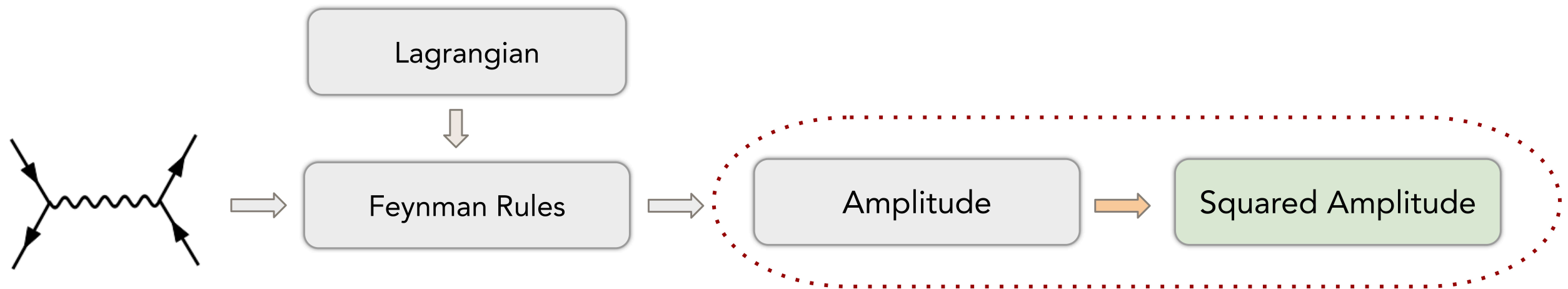


ML & theory



$$i\mathcal{M} = \frac{\frac{1}{2} ie^3 (p_{3\rho} \gamma_\epsilon^\rho \gamma_{\rho\eta} A_j^{\rho*}(p_5) \mathbf{e}_{i\eta}^*(p_4) \mathbf{e}_{l\epsilon}^*(p_3) \mathbf{e}_{k\delta}(p_2) \mathbf{e}_{i\delta}(p_1) - \frac{1}{2} p_{5\sigma} \gamma_{\rho\epsilon} \gamma_\epsilon^\rho \gamma_{\rho\eta} \gamma_\epsilon^\sigma A_j^{\rho*}(p_5) \mathbf{e}_{i\eta}^*(p_4) \mathbf{e}_{l\epsilon}^*(p_3) \mathbf{e}_{k\delta}(p_2) \mathbf{e}_{i\delta}(p_1))}{((m_e^2 - \vec{p}_2 \cdot \vec{p}_4) * \vec{p}_3 \cdot \vec{p}_5)}$$

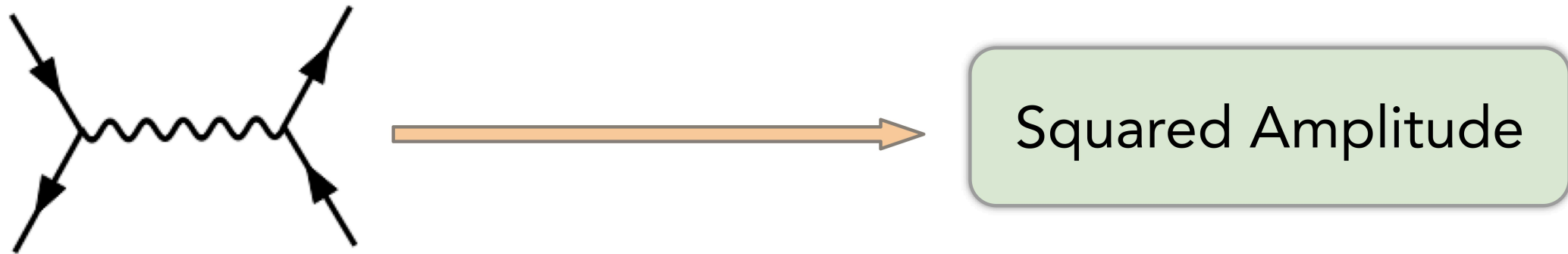
$$|\mathcal{M}|^2 = -\frac{e^6}{((\vec{p}_3 \cdot \vec{p}_5)^2 * (m_e^2 - \vec{p}_2 \cdot \vec{p}_4)^2)} (2m_e^6 + m_e^4 * (-\vec{p}_1 \cdot \vec{p}_3 - \vec{p}_1 \cdot \vec{p}_5 - \vec{p}_2 \cdot \vec{p}_4 + 2\vec{p}_3 \cdot \vec{p}_5) + m_e^2 * (\vec{p}_1 \cdot \vec{p}_2 * \vec{p}_3 \cdot \vec{p}_4 + \vec{p}_1 \cdot \vec{p}_2 * \vec{p}_4 \cdot \vec{p}_5 + \vec{p}_1 \cdot \vec{p}_4 * \vec{p}_2 \cdot \vec{p}_3 + \vec{p}_1 \cdot \vec{p}_4 * \vec{p}_2 \cdot \vec{p}_5 + \vec{p}_1 \cdot \vec{p}_5 * \vec{p}_3 \cdot \vec{p}_4 - \vec{p}_2 \cdot \vec{p}_4 * \vec{p}_3 \cdot \vec{p}_5) - \vec{p}_1 \cdot \vec{p}_2 * \vec{p}_3 \cdot \vec{p}_5 * \vec{p}_4 \cdot \vec{p}_5 - \vec{p}_1 \cdot \vec{p}_4 * \vec{p}_2 \cdot \vec{p}_5 * \vec{p}_3 \cdot \vec{p}_5)$$



a fraction of diagrams
for $gg \rightarrow gg$ scattering

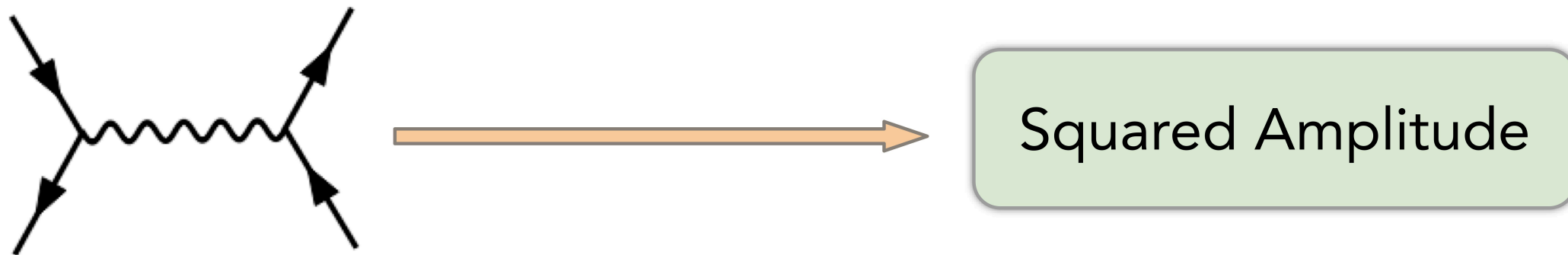
for some processes
we're talking about
thousands of diagrams

Proposal: use **transformers** (**arXiv:1706.03762v7**) to directly translate symbolic Feynman diagrams into the squared amplitudes.



Mach.Learn.Sci.Tech. 4 (2023) 1, 015007

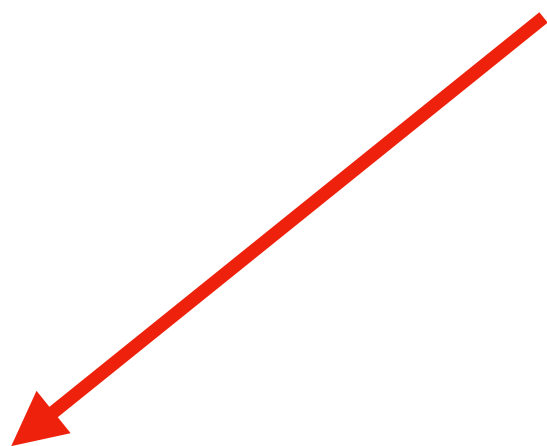
Proposal: use **transformers** (**arXiv:1706.03762v7**) to directly translate symbolic Feynman diagrams into the squared amplitudes.



Mach.Learn.Sci.Tech. 4 (2023) 1, 015007

- create symbolic expressions for interactions in QED/QCD
- simplify/organize expressions in a standard format
- map symbols to integers (assigning an integer to each one)
- convert symbol sequences into vectors built from these integers
- train

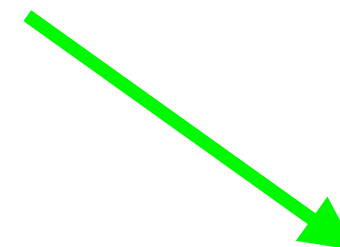
	Training Size	Sequence Acc.	Token Score	RMSE
QED (sequence)	251K	98.6%	99.7%	1.3×10^{-3}
QCD (sequence)	140K	97.4%	98.9%	8.8×10^{-3}
(QED + QCD) on QED	391K	99.0%	99.4%	2.5×10^{-3}
(QED + QCD) on QCD	391K	97.6%	98.8%	6.8×10^{-3}
QED (diagram)	258K	99.0%	99.7%	9.3×10^{-4}
QCD (diagram)	142K	73.4%	82.0%	0.3



number of input-
output pairs



how well a full sequence
is "translated"



how well a symbol
is "translated"

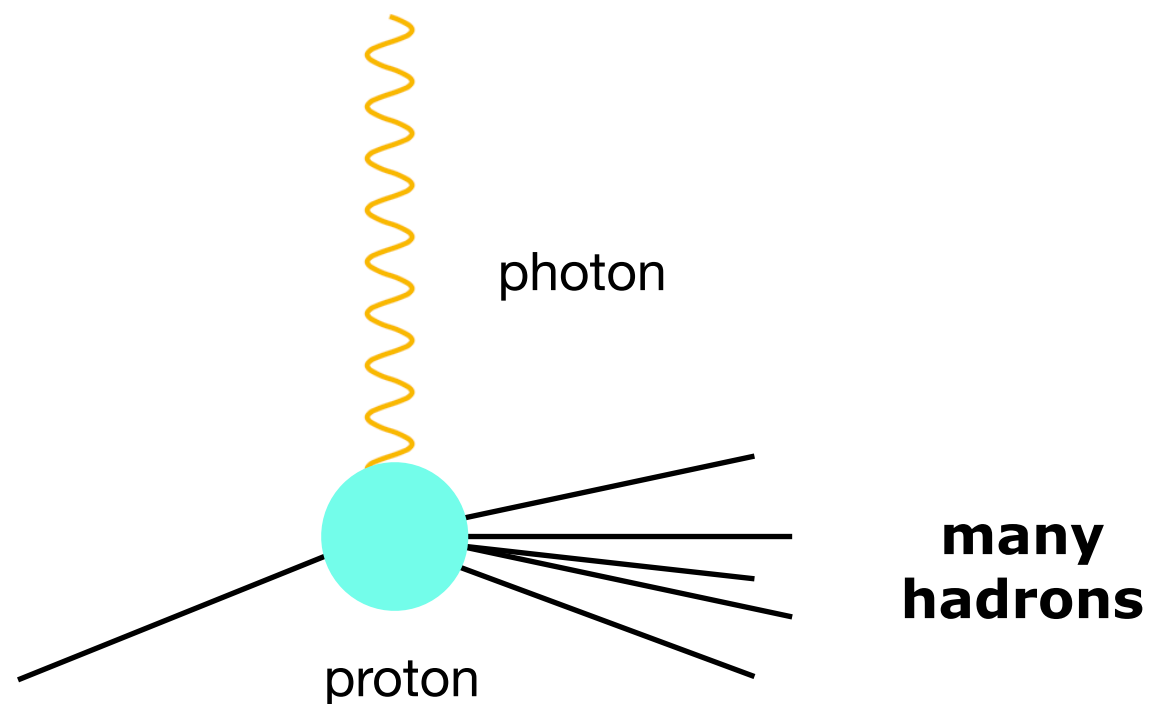
ML &

phenomenology

Particles with colour charge (a.k.a. **partons**) can't be seen alone. They are **bound together in hadrons** (protons, neutrons, pions, kaons, etc).

Particles with colour charge (a.k.a. **partons**) can't be seen alone. They are **bound together in hadrons** (protons, neutrons, pions, kaons, etc).

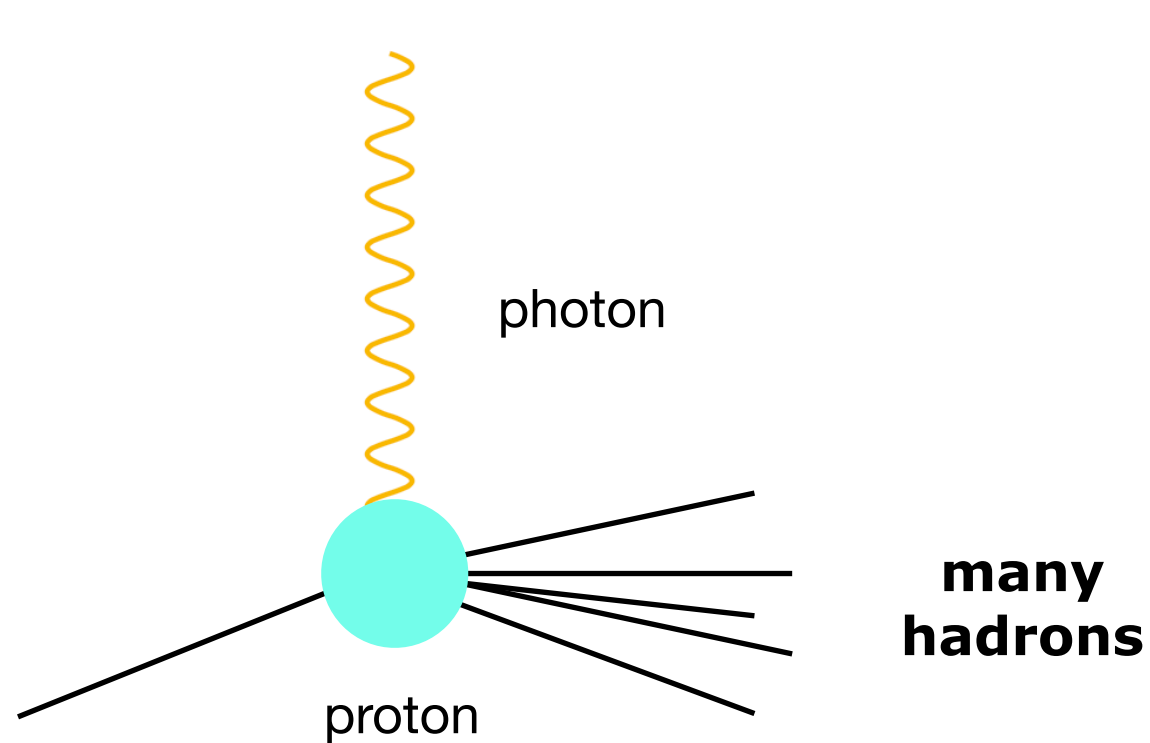
If we study a photon-proton interaction



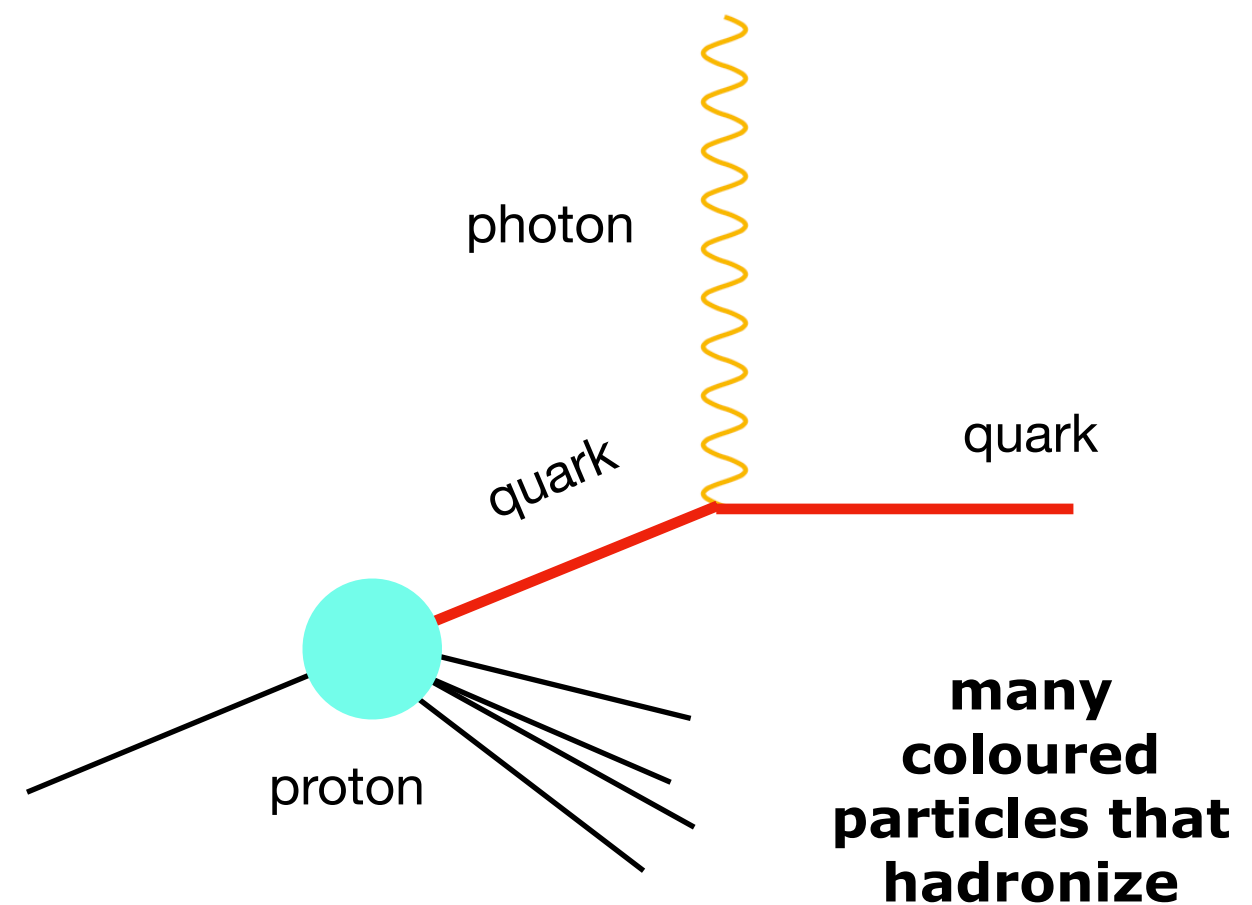
What we see

Particles with colour charge (a.k.a. **partons**) can't be seen alone. They are **bound together in hadrons** (protons, neutrons, pions, kaons, etc).

If we study a photon-proton interaction



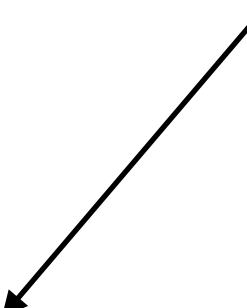
What we see



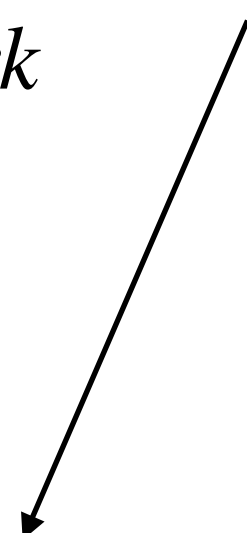
How we compute it

Which quark should I use? It could be any of them. So we add all contributions:

$$d\sigma^{\gamma+p}(Q) \propto \sum_{i=\text{quark}} d\sigma^{\gamma+i}(Q)$$



**probability of
photon-proton
interaction**



**probability of
photon-quark
interaction**

Which quark should I use? It could be any of them. So we add all contributions:

$$d\sigma^{\gamma+p}(Q) \propto \sum_{i=\text{quark}} d\sigma^{\gamma+i}(Q) \otimes f^{i/p}(Q)$$

**probability of
photon-proton
interaction**

**probability of
photon-quark
interaction**

**probability
density of
finding the quark
in the proton**

For many processes we can do this type of **factorization**.

The parton density/distribution functions (or **PDFs**) are a crucial part of the calculations in HEP.

There are many types of PDFs: initial state, final state, collinear, unpolarised, polarised, transverse momentum dependent, generalised, nuclear, etc.

The parton density/distribution functions (or **PDFs**) are a crucial part of the calculations in HEP.

There are many types of PDFs: initial state, final state, collinear, unpolarised, polarised, transverse momentum dependent, generalised, nuclear, etc.



PDFs are universal.



We know how PDFs evolve with the relevant scale(s) of the processes.



PDFs can't be computed from the theory.

PDFs using **regression**

Step 1: propose a functional form

$$f^{s/p}(x; \mu_0) = N_i x_i^\alpha (1 - x)_i^\beta$$

Step 2: evolve with known equations to the measured scale(s)

Step 3: convolute to compute the observable

PDFs using **NN**

Step 1: use as input the output of a NN

$$f^{i/p}(x; \mu_0) = NN(x; \theta)$$

PDFs using **regression**

Step 1: propose a functional form

$$f^{s/p}(x; \mu_0) = N_i x_i^\alpha (1-x)_i^\beta$$

Step 2: evolve with known equations to the measured scale(s)

Step 3: convolute to compute the observable

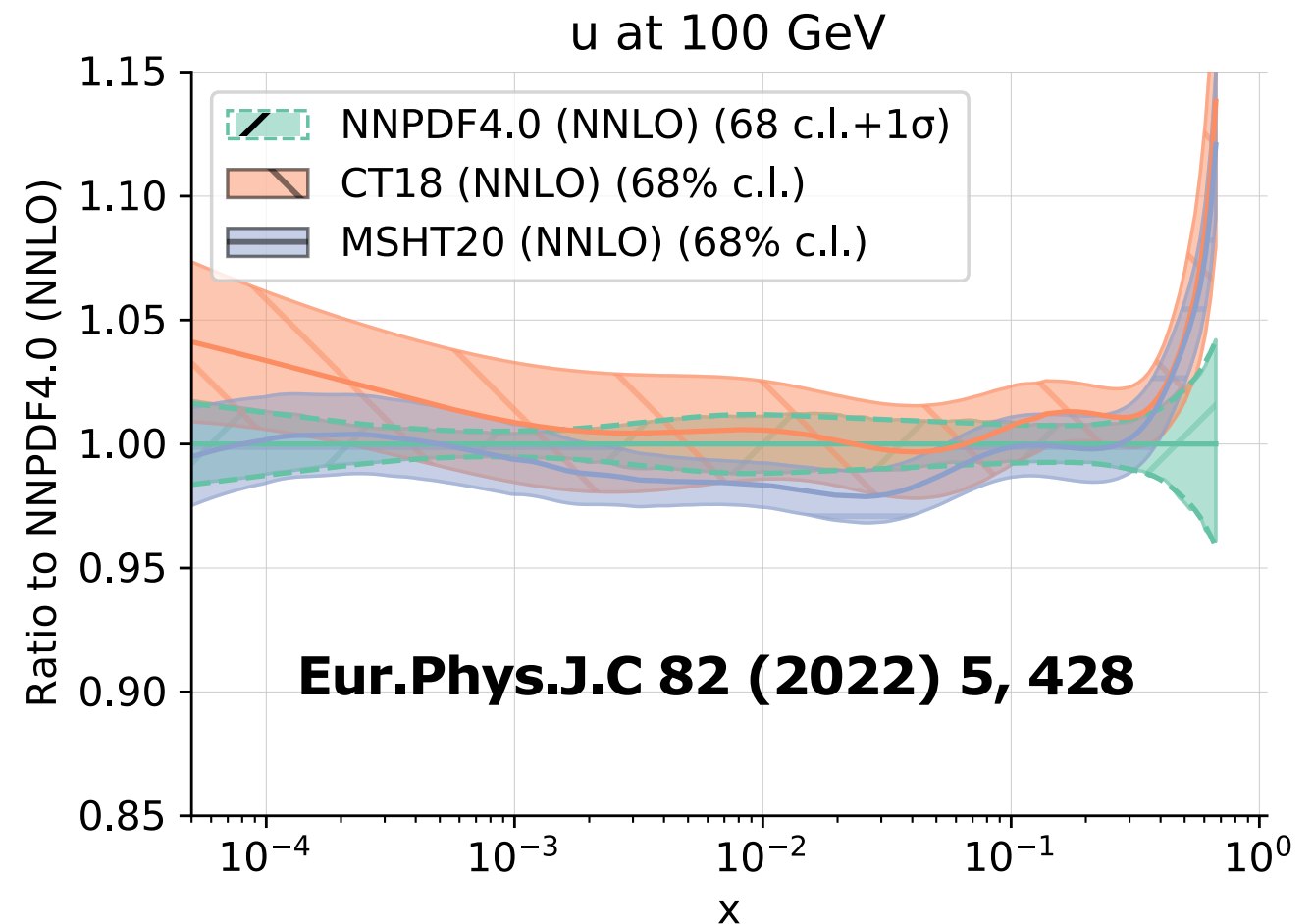
MSHT20: 67 parameters in
“classical” regression model

NNPDF4.0: 2-25-20-8
architecture (~ 750
parameters), hyperbolic
tangent activation

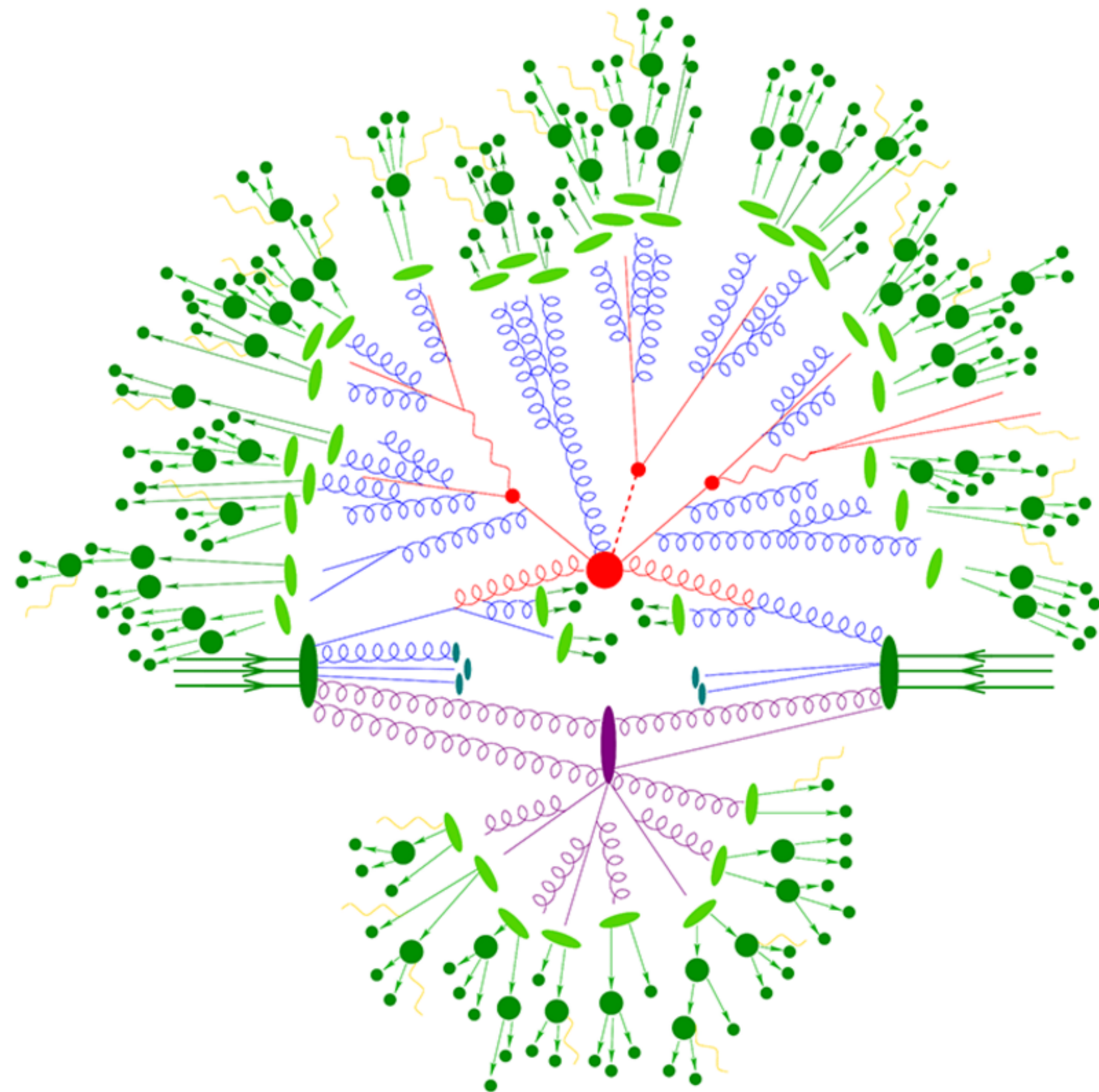
PDFs using **NN**

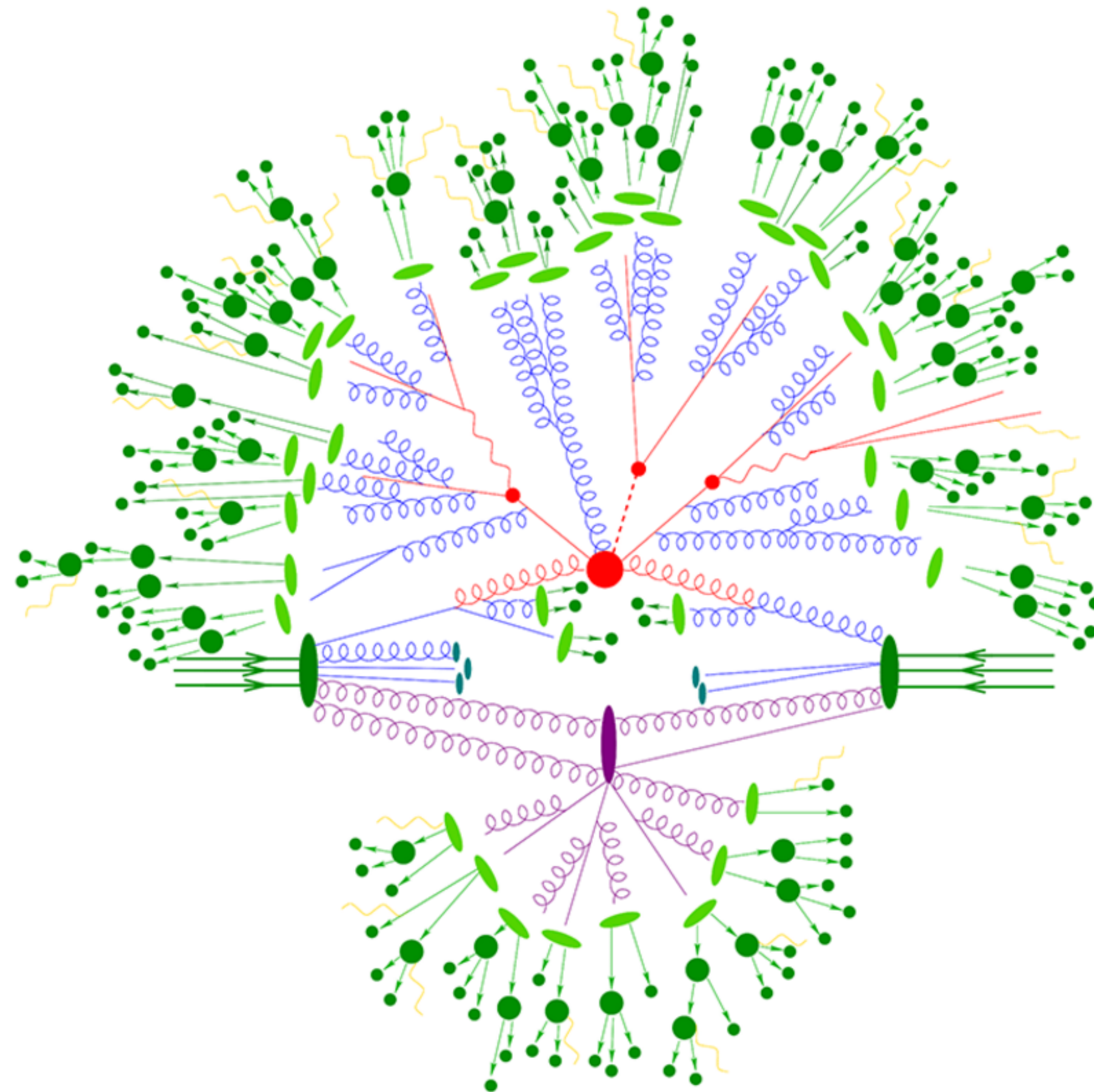
Step 1: use as input the output
of a NN

$$f^{i/p}(x; \mu_0) = NN(x; \theta)$$



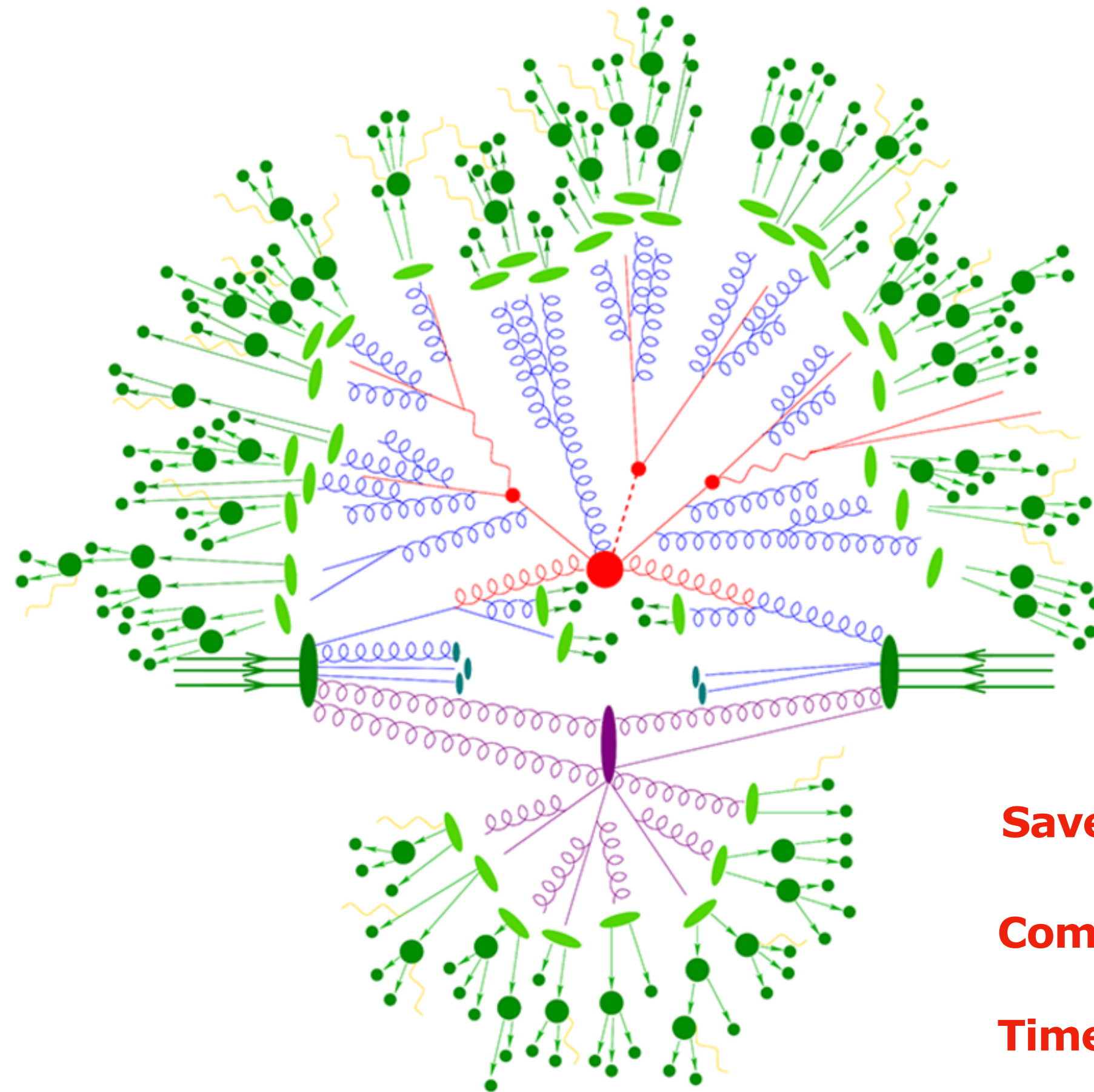
ML & simulations





Monte Carlo event generators:

simulate a process
(event) from start
to end, including
the detector effects.



Monte Carlo event generators:
simulate a process
(event) from start
to end, including
the detector effects.

Saves a lot of money. 🎉

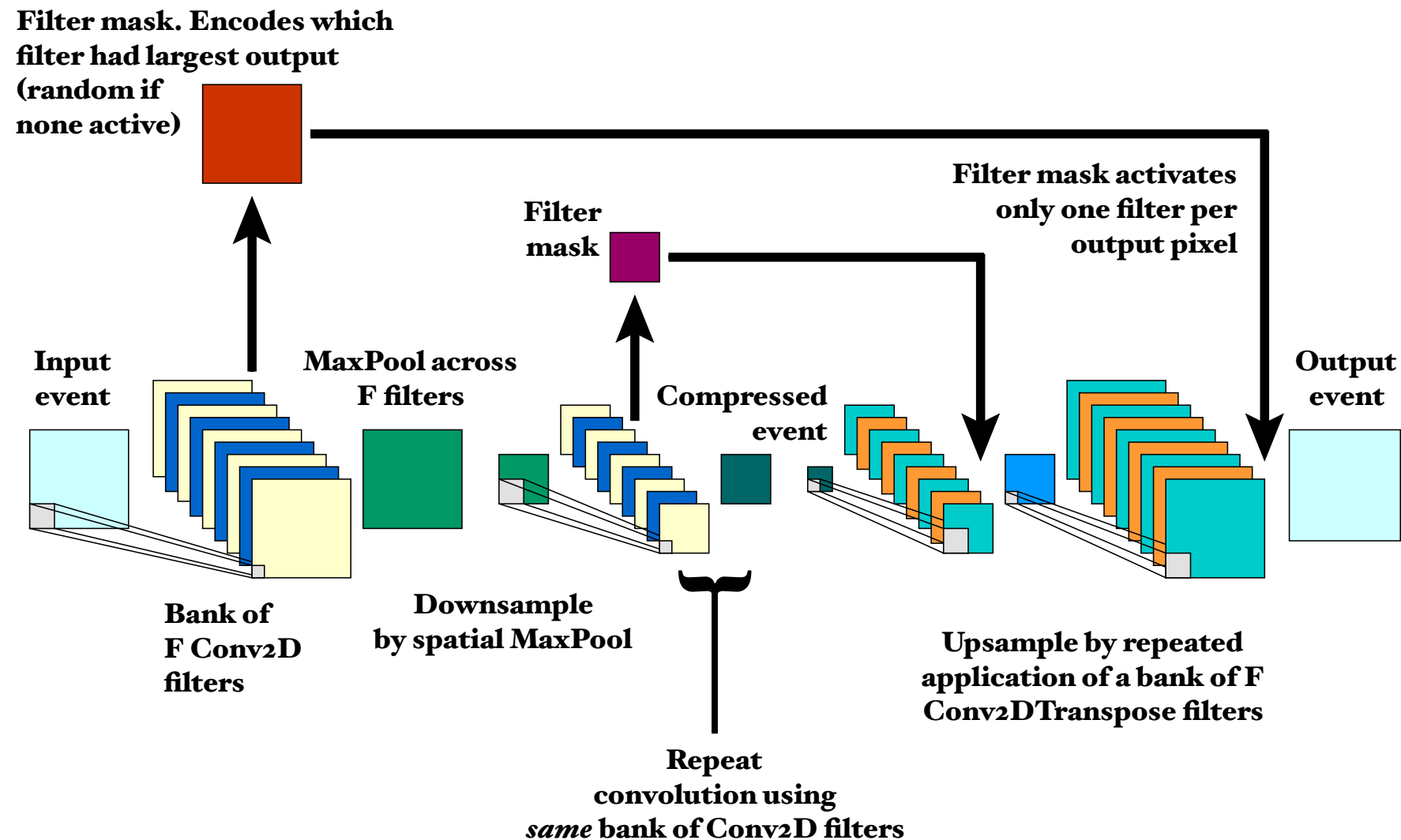
Computationally costly. 😭

Time costly. 😭

Must be tuned to data. 😭

Use deep learning to construct parton showers: autoencoding CNN

JHEP 12 (2018) 021



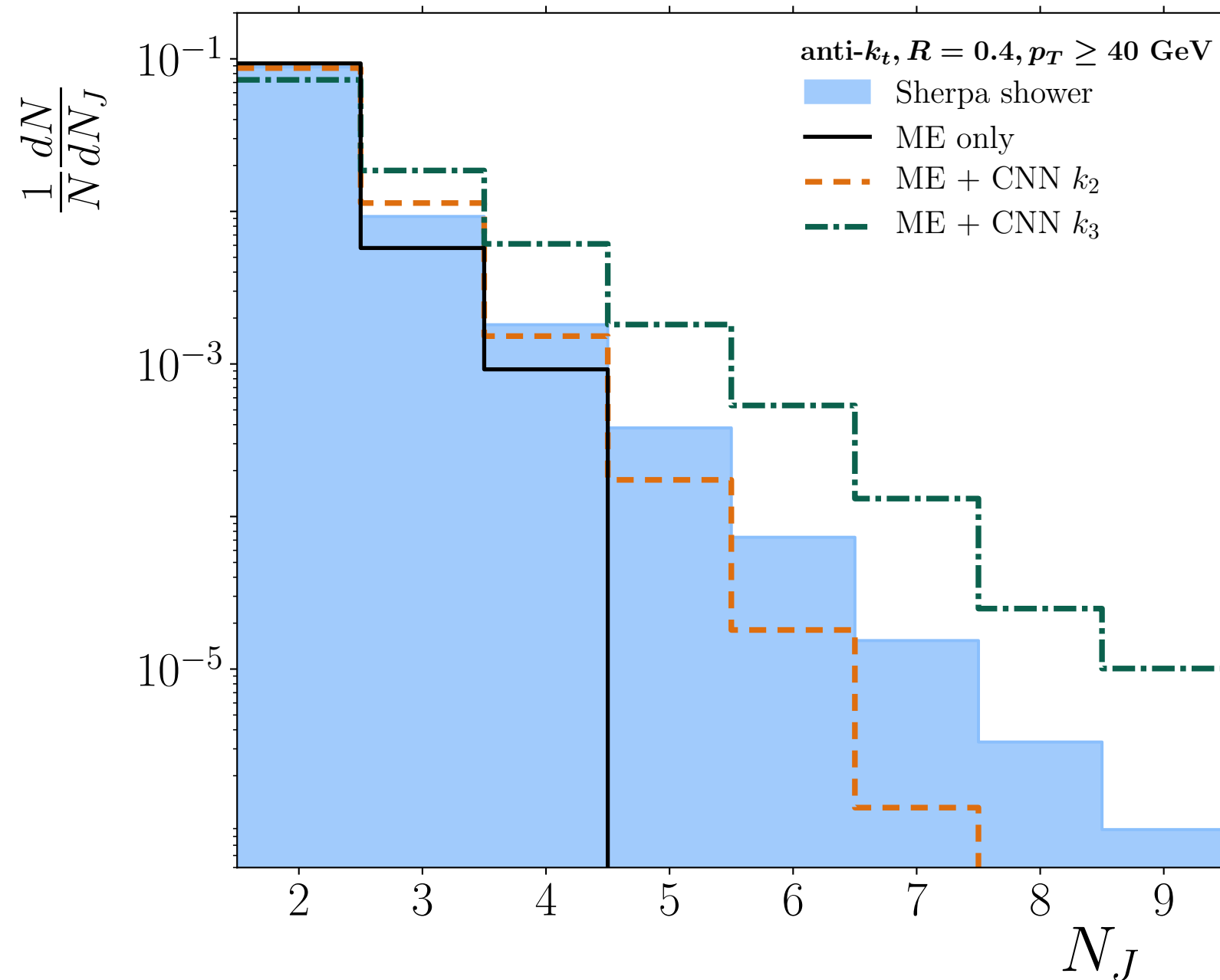
Event data: image of pixel arrays that represent the emissions of energy from each proton-proton collision.

The autoencoding uses several convolutional layers to reduce the size of the image to the size of the filters.

Sample of 8.5 million QCD proton-proton collision events.

Hadronisation and MPI turned off (require complicated deep learning model).

Each event is converted to $N \times N$ pixel-array images in rapidity-azimuth, identifying the pixel in the $y-\phi$ and adding the particle p_T .



ML & experiments

A standard detector in a collider:

CMS DETECTOR

Total weight : 14,000 tonnes
 Overall diameter : 15.0 m
 Overall length : 28.7 m
 Magnetic field : 3.8 T

STEEL RETURN YOKE
 12,500 tonnes

SILICON TRACKERS

Pixel ($100 \times 150 \mu\text{m}$) $\sim 1\text{m}^2 \sim 66\text{M}$ channels
 Microstrips ($80 \times 180 \mu\text{m}$) $\sim 200\text{m}^2 \sim 9.6\text{M}$ channels

SUPERCONDUCTING SOLENOID
 Niobium titanium coil carrying $\sim 18,000\text{A}$

MUON CHAMBERS

Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
 Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers

PRESHOWER

Silicon strips $\sim 16\text{m}^2 \sim 137,000$ channels

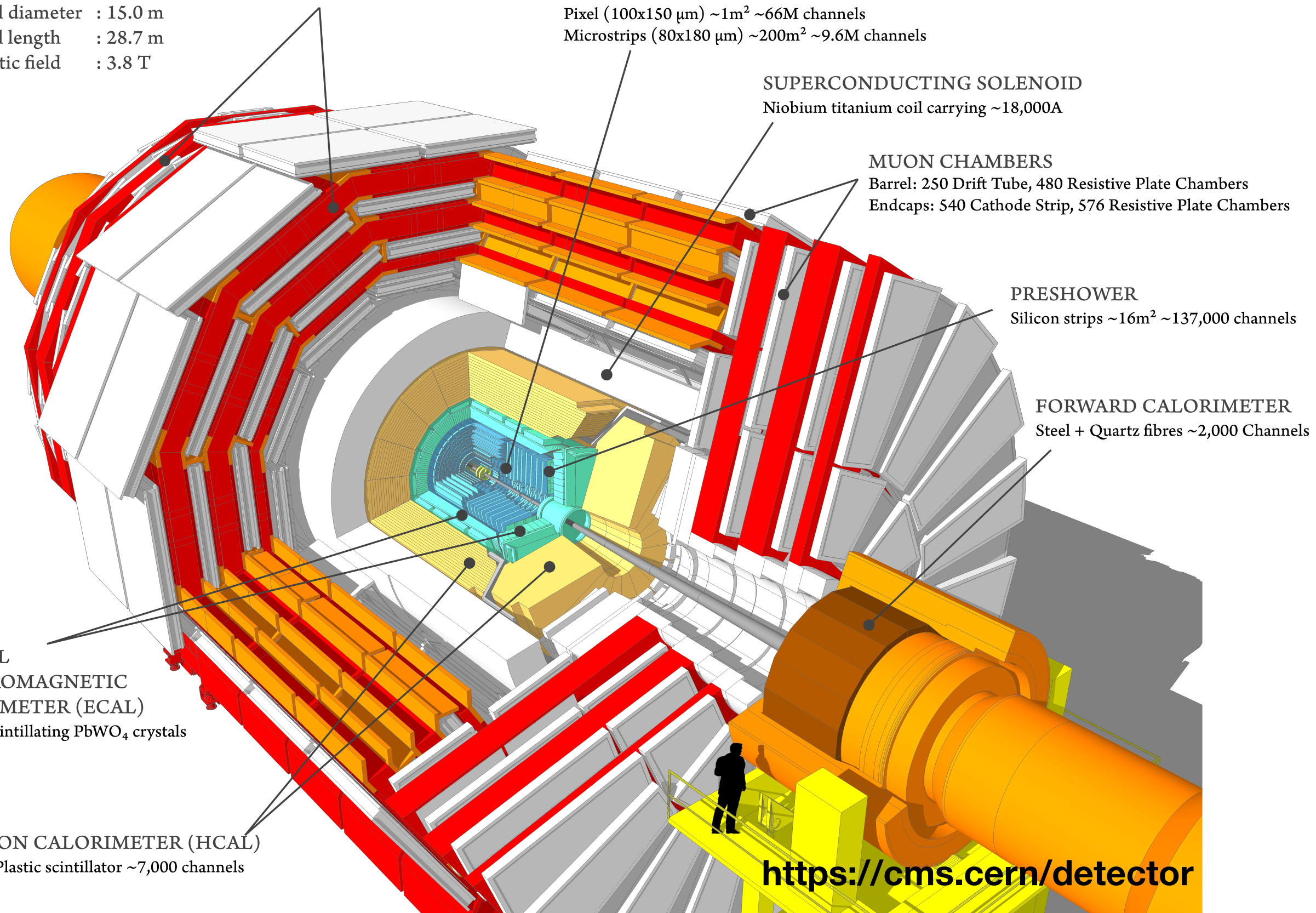
FORWARD CALORIMETER

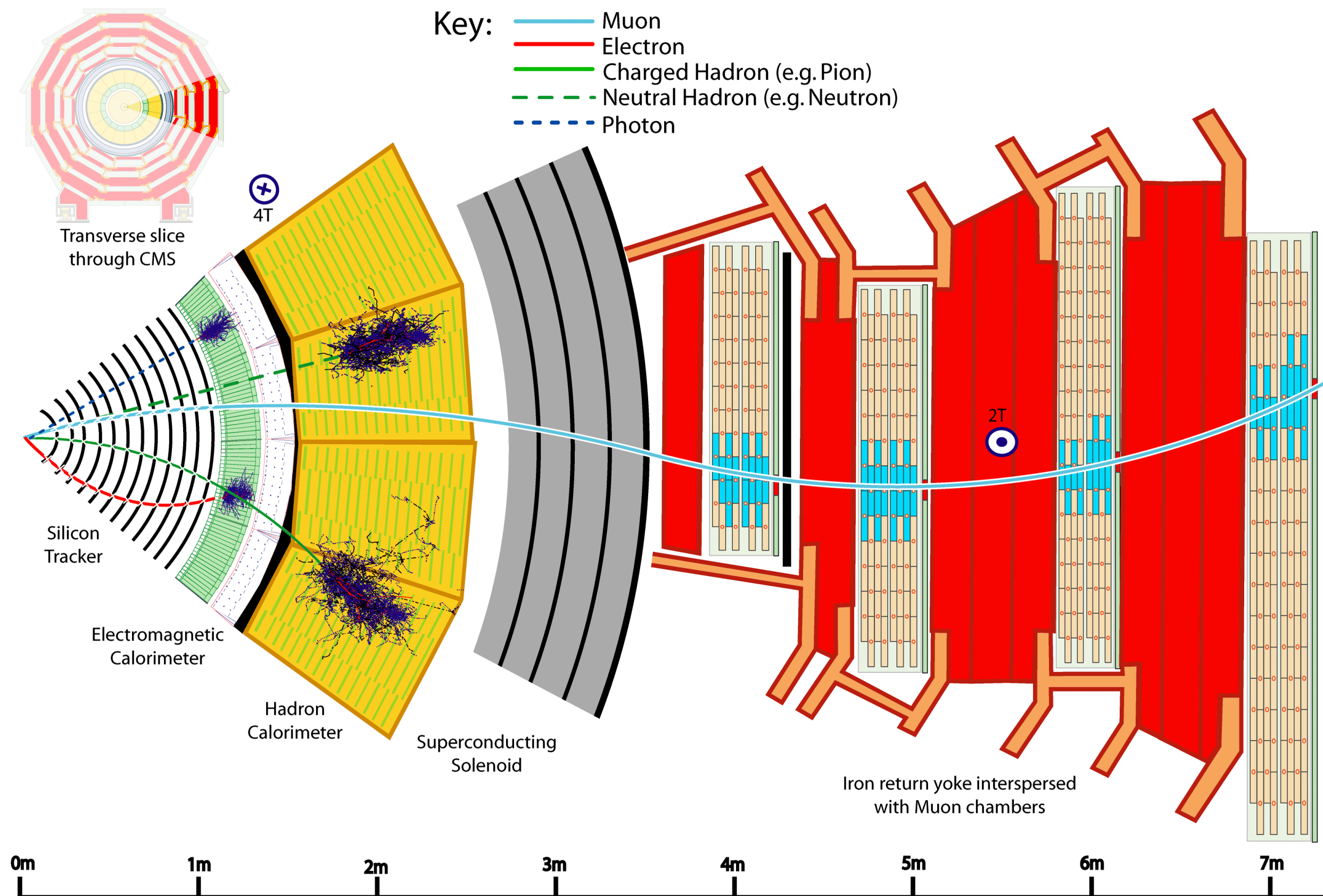
Steel + Quartz fibres $\sim 2,000$ Channels

CRYSTAL
 ELECTROMAGNETIC
 CALORIMETER (ECAL)
 $\sim 76,000$ scintillating PbWO_4 crystals

HADRON CALORIMETER (HCAL)
 Brass + Plastic scintillator $\sim 7,000$ channels

<https://cms.cern/detector>





With a large number of events ($\sim 6 \cdot 10^8$ collisions/s at LHC), it is impossible to record all that happens. The ***trigger system*** decides if one event is interesting and if it should be stored.

With a large number of events ($\sim 6 \cdot 10^8$ collisions/s at LHC), it is impossible to record all that happens. The ***trigger system*** decides if one event is interesting and if it should be stored.

The triggering happens in several stages:

- Level 1 (L1) is implemented on special electronic circuits and field-programmable gate arrays (FPGAs).
- The high-level trigger (HLT), based on CPUs and GPUs, uses more sophisticated algorithms. Recent developments allow for the use of ML methods on FPGAs. This allows for better cuts at the L1 level.

With a large number of events ($\sim 6 \cdot 10^8$ collisions/s at LHC), it is impossible to record all that happens. The **trigger system** decides if one event is interesting and if it should be stored.

The triggering happens in several stages:

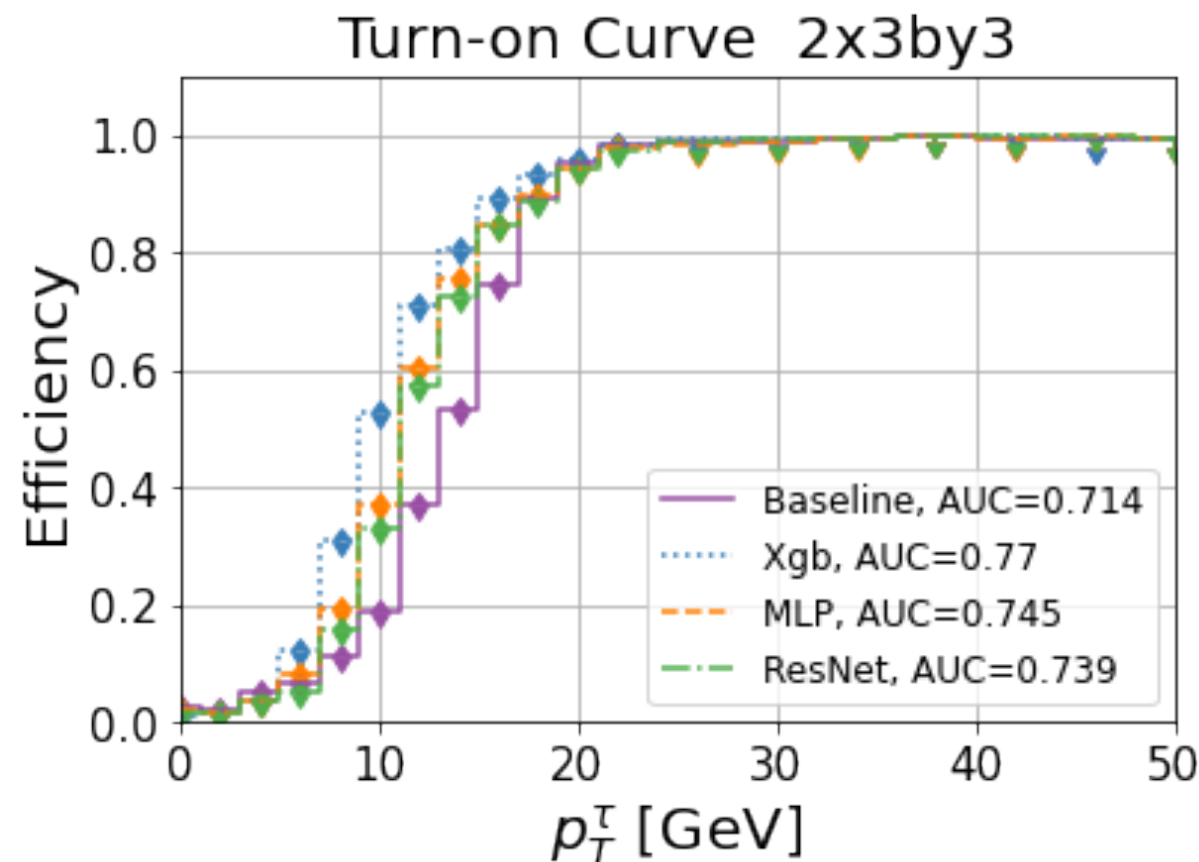
- Level 1 (L1) is implemented on special electronic circuits and field-programmable gate arrays (FPGAs).
- The high-level trigger (HLT), based on CPUs and GPUs, uses more sophisticated algorithms. Recent developments allow for the use of ML methods on FPGAs. This allows for better cuts at the L1 level.

A goal is to develop fast, simple algorithms to detect interesting processes at the L1 level (this is a strong constraint).

Example: $H \rightarrow \tau^+ \tau^- \rightarrow \text{hadrons}$ happens $\sim 68\%$ of the time. And they look very similar to jets $\rightarrow$ hadrons. So they might be thrown away by L1.

arXiv:2306.06743 [hep-ex]

- Extreme Gradient Boosting (XGBoost) algorithm: gradient boosting tree model with less parameters than NN.
- Multi-layer perceptron (MLP) NN: two hidden layers (5, 4 units).
- Residual Network: treats the inputs as images.

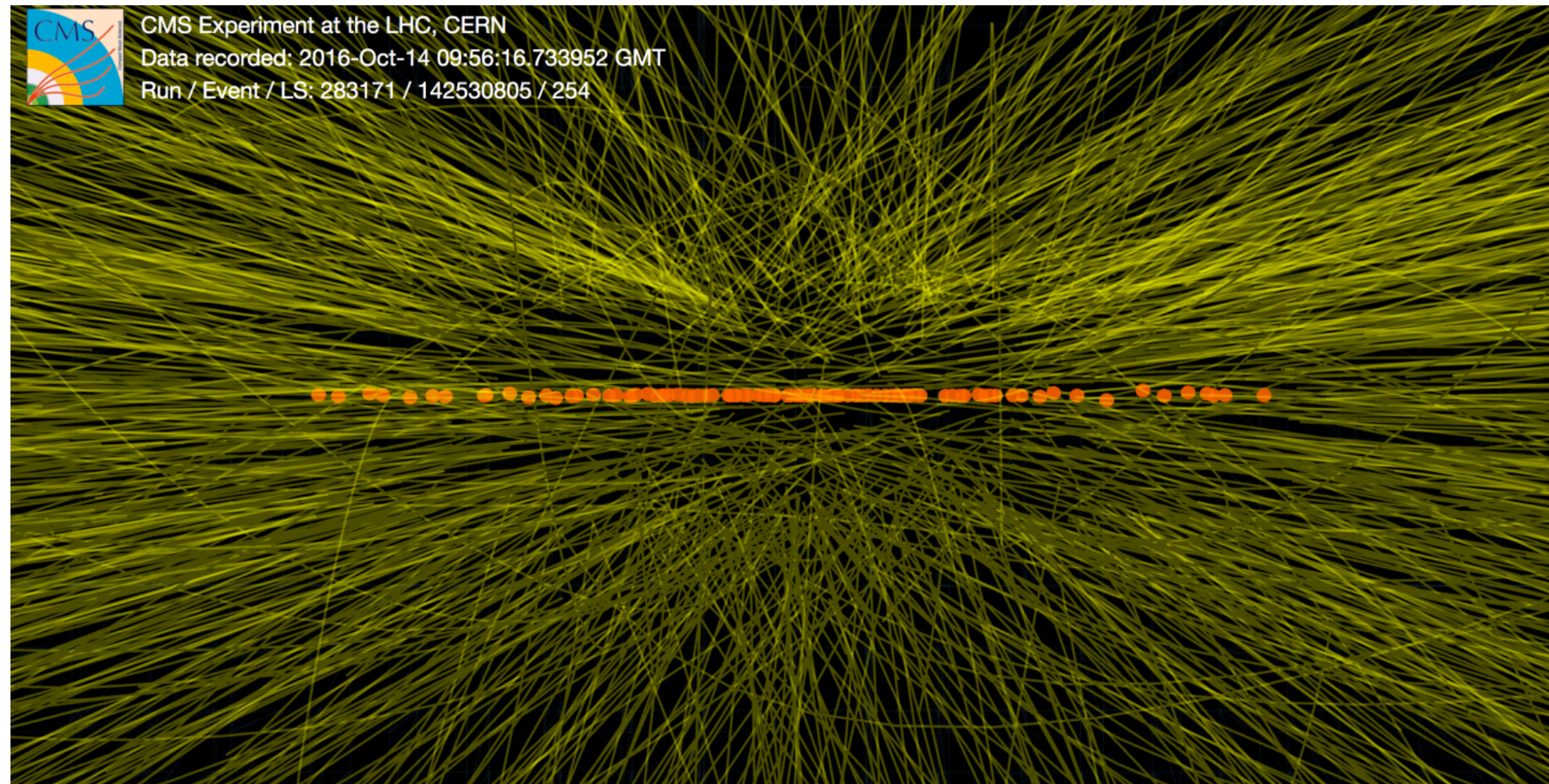


Another thing to deal with is the *pileup*: the nearly simultaneous occurrence of interesting (high p_T) and not-interesting events (low p_T).

The transverse coordinate of the pileup collisions are aligned to those of the interesting collision.

Another thing to deal with is the *pileup*: the nearly simultaneous occurrence of interesting (high p_T) and not-interesting events (low p_T).

The transverse coordinate of the pileup collisions are aligned to those of the interesting collision.

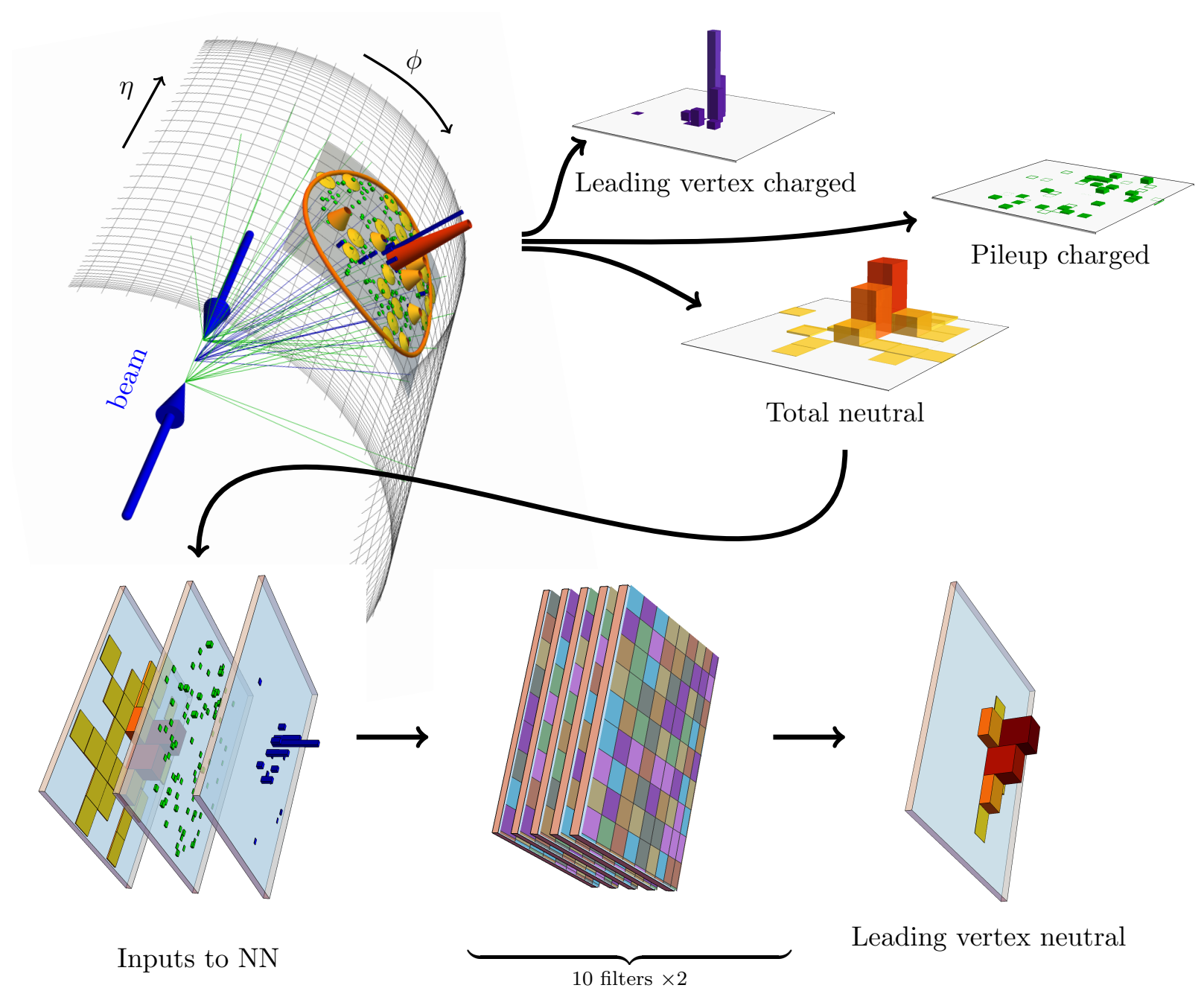


Pileup vertices can be displaced by several cm from the interesting vertex in the longitudinal direction.

Pileup from charged particles can be easily filtered because the tracks can be used to reconstruct the vertexes.

Pileup mitigation for neutral particles with two different approaches:
regression and classification

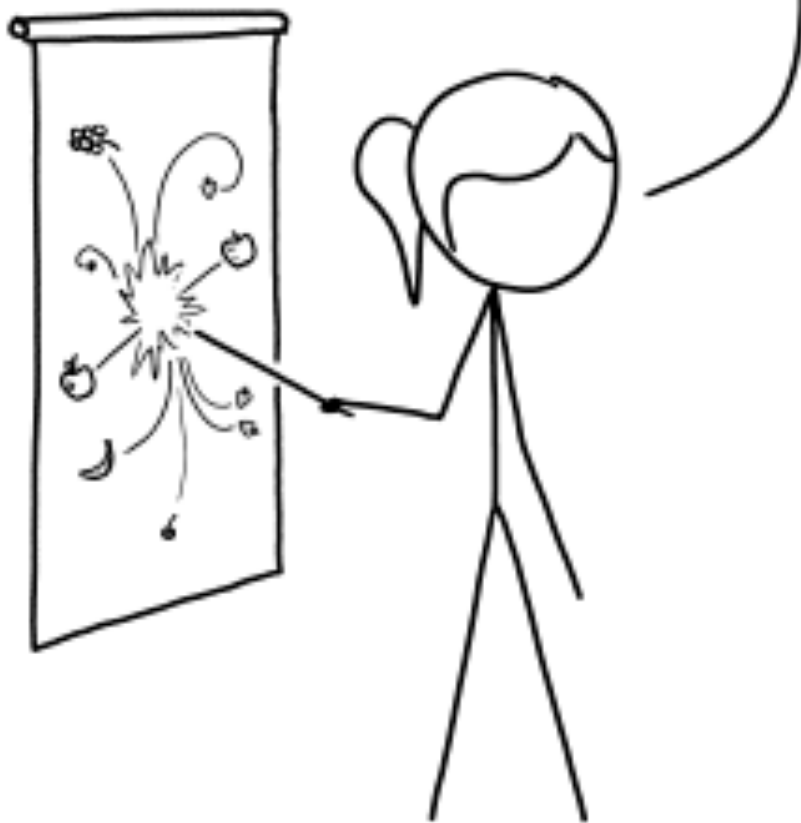
regression: energy distribution of particles as the intensity of pixels in an image, and use CNN to extract as output the neutral particles from the leading vertex.



WHEN TWO APPLES COLLIDE, THEY CAN BRIEFLY FORM EXOTIC NEW FRUIT. PINEAPPLES WITH APPLE SKIN. POMEGRANATES FULL OF GRAPES. WATERMELON-SIZED PEACHES.

THESE NORMALLY DECAY INTO A SHOWER OF FRUIT SALAD, BUT BY STUDYING THE DEBRIS, WE CAN LEARN WHAT WAS PRODUCED.

THEN, THE HUNT IS ON FOR A STABLE FORM.



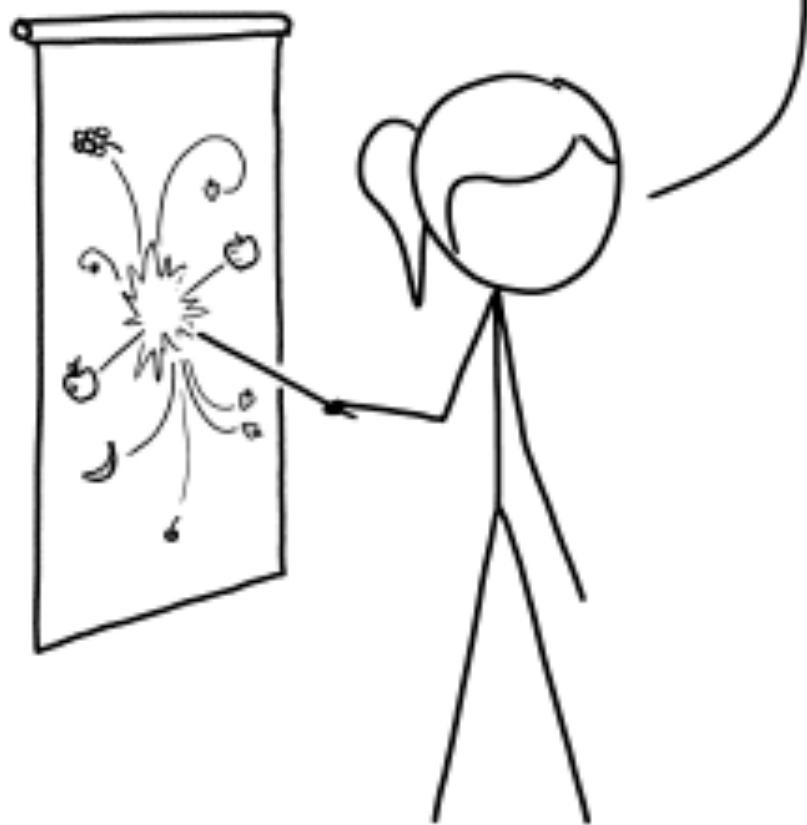
Once we have stored some data, we have to **classify** them.

HOW NEW TYPES OF FRUIT ARE DEVELOPED

WHEN TWO APPLES COLLIDE, THEY CAN BRIEFLY FORM EXOTIC NEW FRUIT. PINEAPPLES WITH APPLE SKIN. POMEGRANATES FULL OF GRAPES. WATERMELON-SIZED PEACHES.

THESE NORMALLY DECAY INTO A SHOWER OF FRUIT SALAD, BUT BY STUDYING THE DEBRIS, WE CAN LEARN WHAT WAS PRODUCED.

THEN, THE HUNT IS ON FOR A STABLE FORM.



HOW NEW TYPES OF FRUIT ARE DEVELOPED

Once we have stored some data, we have to **classify** them.

The collisions are dominated by interactions that produce quarks and gluons that turn into jets of mostly pions.

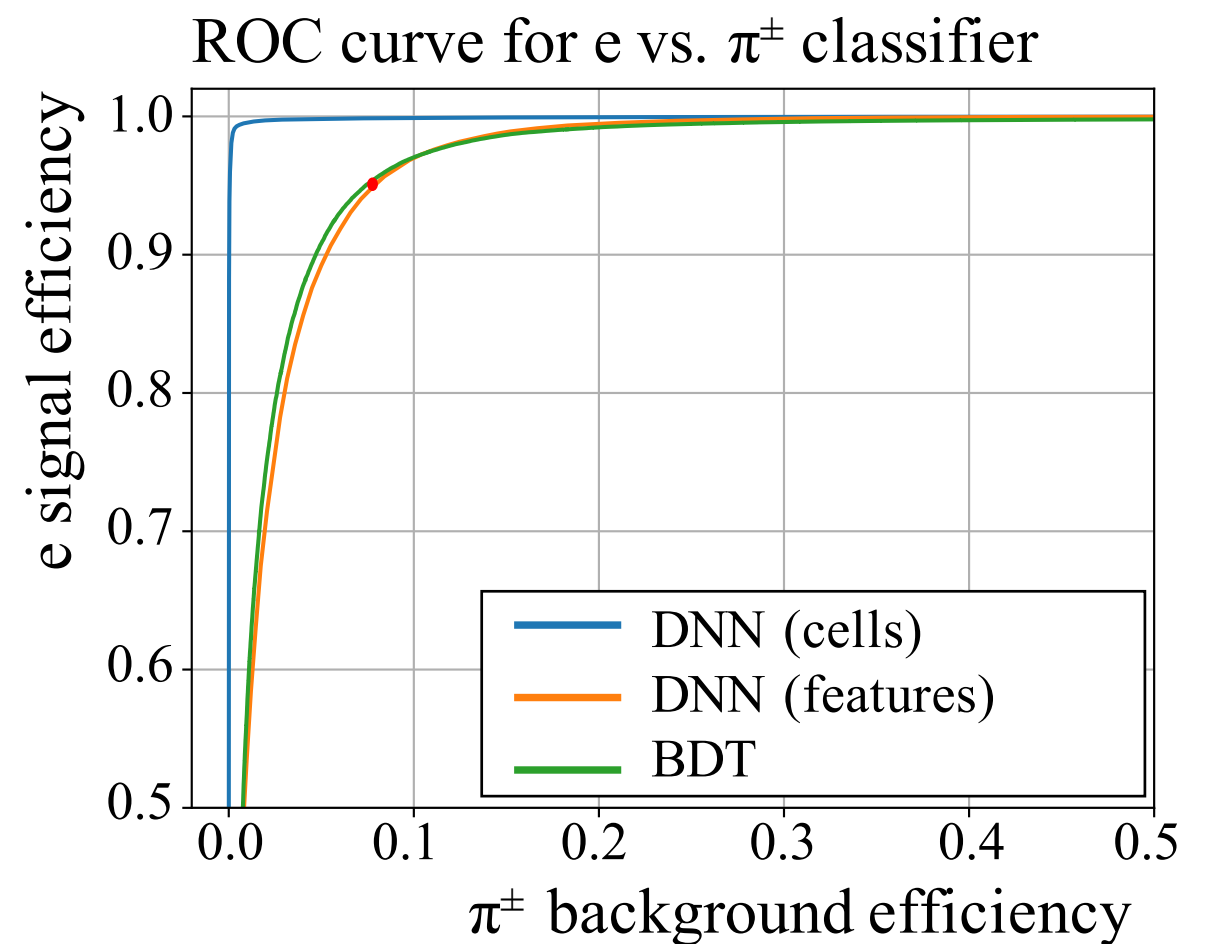
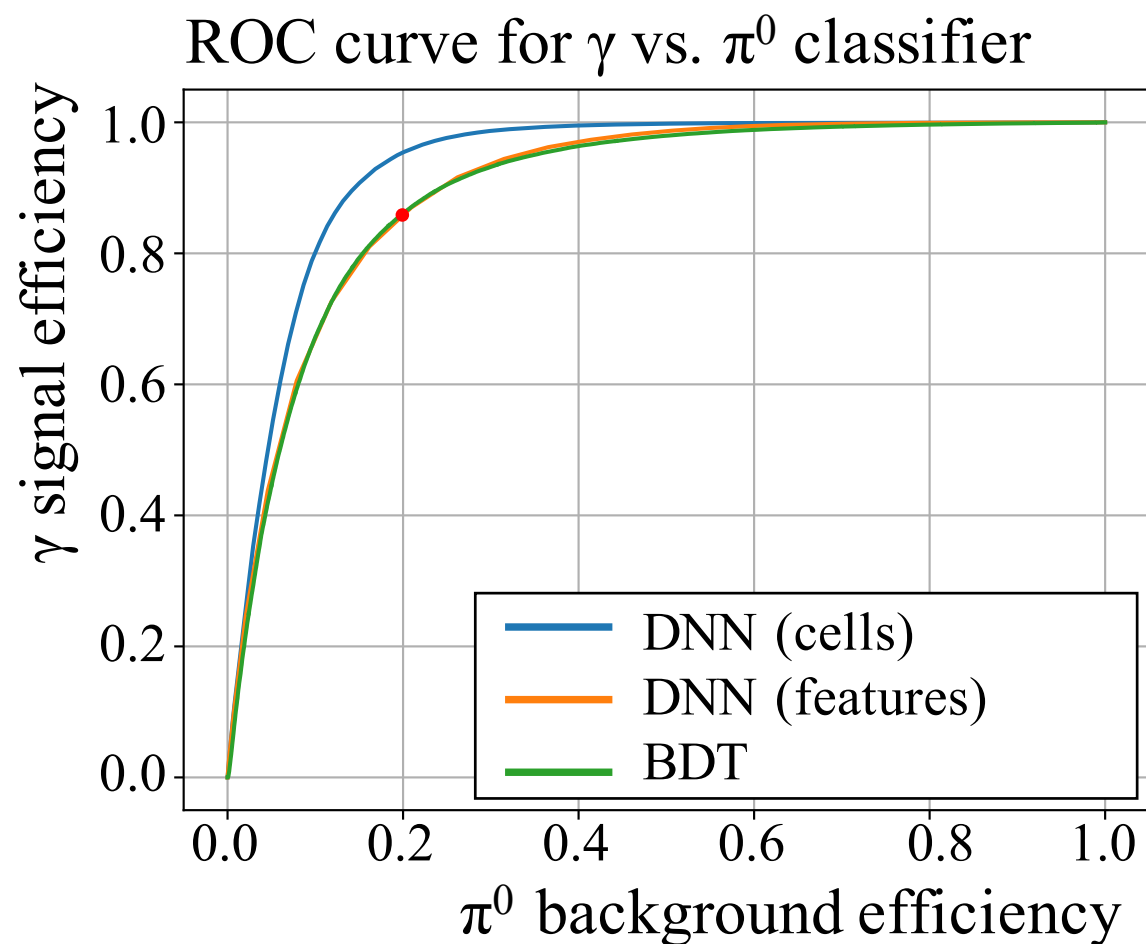
These jets are background to the identification of photons and leptons.

Calorimeters are required to only mis-identify roughly 1 in 10000 jets as a photon or electron.

Classify two cases under study:

- $\pi^\pm$ mis-identified as $e^\pm$ due to early showering in the ECAL
- $\pi^0 \rightarrow 2\gamma$ mis-identified as $\pi^0 \rightarrow \gamma$

Using cell-based DNNs, feature-based NN and Boosted Decision Trees:



Practical use for hardware: quality control of new silicon sensors

- High-Granularity Calorimeter to replace endcap calorimeters at CMS
- more than 27,000 hexagonal 8" silicon sensor wafers
- 5% of each batch will undergo quality control (search for anomalies)

Practical use for hardware: quality control of new silicon sensors

- High-Granularity Calorimeter to replace endcap calorimeters at CMS
- more than 27,000 hexagonal 8" silicon sensor wafers
- 5% of each batch will undergo quality control (search for anomalies)

Traditionally, anomalies were detected by taking hundreds of microscope images using a scan program, and having them inspected by a human operator.

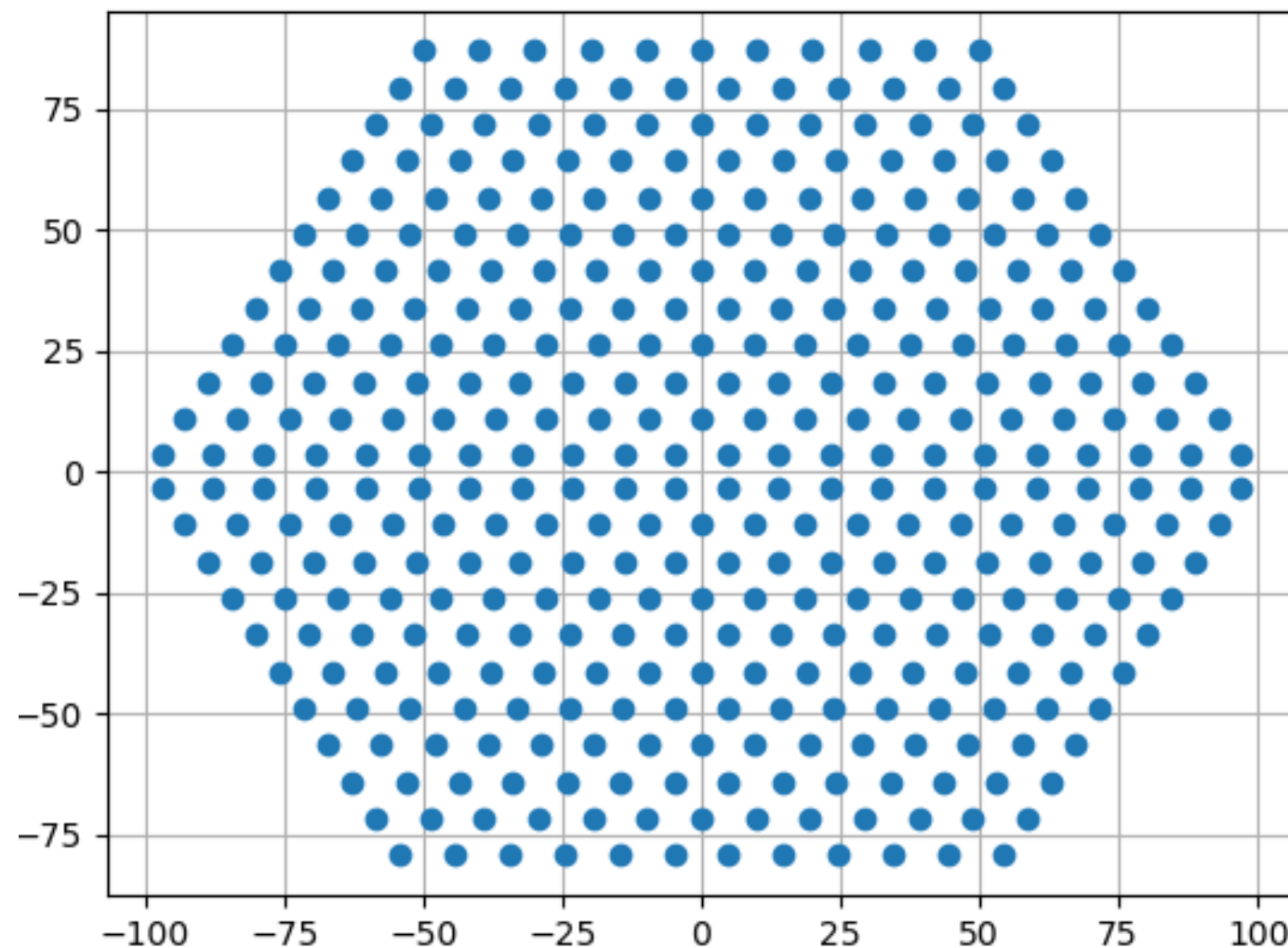
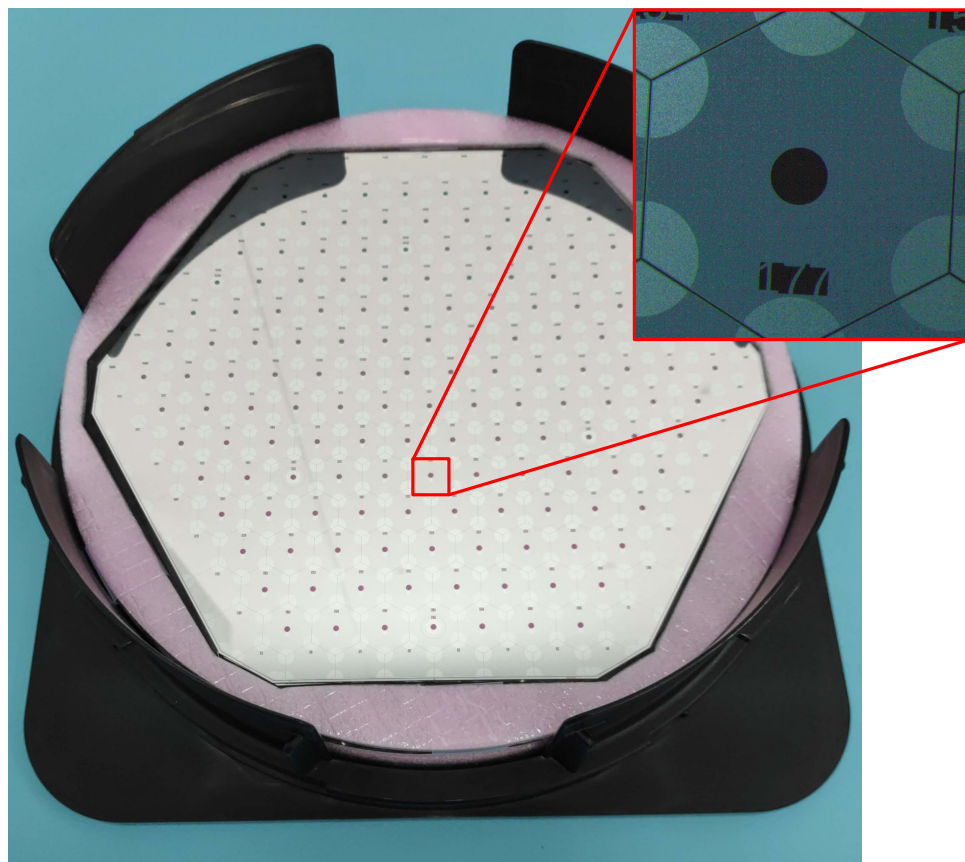
Practical use for hardware: quality control of new silicon sensors

- High-Granularity Calorimeter to replace endcap calorimeters at CMS
- more than 27,000 hexagonal 8" silicon sensor wafers
- 5% of each batch will undergo quality control (search for anomalies)

Traditionally, anomalies were detected by taking hundreds of microscope images using a scan program, and having them inspected by a human operator.

The idea is to reduce biases and detect anomalous scan images via:

- a deep convolutional autoencoder that acts on each image
- a deep convolutional classifier NN that acts on patches of the image, allowing for localization of anomalous areas



Scan map of a detector: 385 images are taken.

Scan image: contains 2720×3680 pixels.

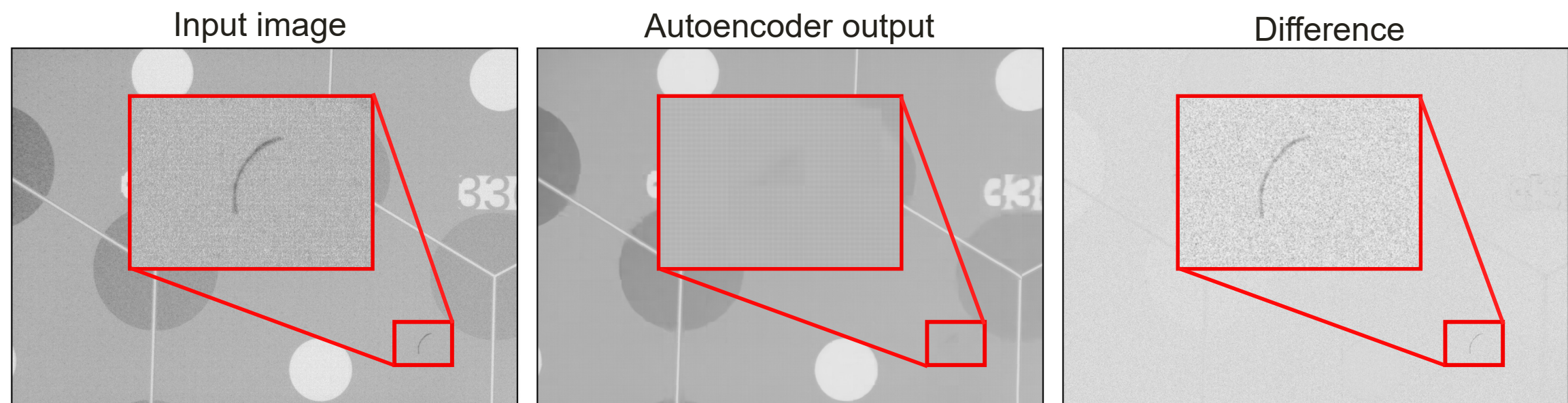
More than 25,000 images analysed (50 detectors).

15 new detectors used for testing.

1. Patching: the whole images are split into patches using a fixed grid.
2. Background detector applied to patches of the whole image.

CNN with four convolutional layers, ReLU activation, followed by dropout layers with a rate of 20% (84,401 trainable parameters), test accuracy of over 99%

3. Anomaly enhancement using an autoencoder:
 - Five convolutional layers; compression factor of 1,600.
 - 126,353 trainable parameters.



1. Patching: the whole images are split into patches using a fixed grid.
2. Background detector applied to patches of the whole image.

CNN with four convolutional layers, ReLU activation, followed by dropout layers with a rate of 20% (84,401 trainable parameters), test accuracy of over 99%

3. Anomaly enhancement using an autoencoder:

- Five convolutional layers; compression factor of 1,600.
- 126,353 trainable parameters.

4. Anomaly detecting classifier applied to the patches of the Difference:

CNN with 23 layers and 2,847,777 trainable parameters.

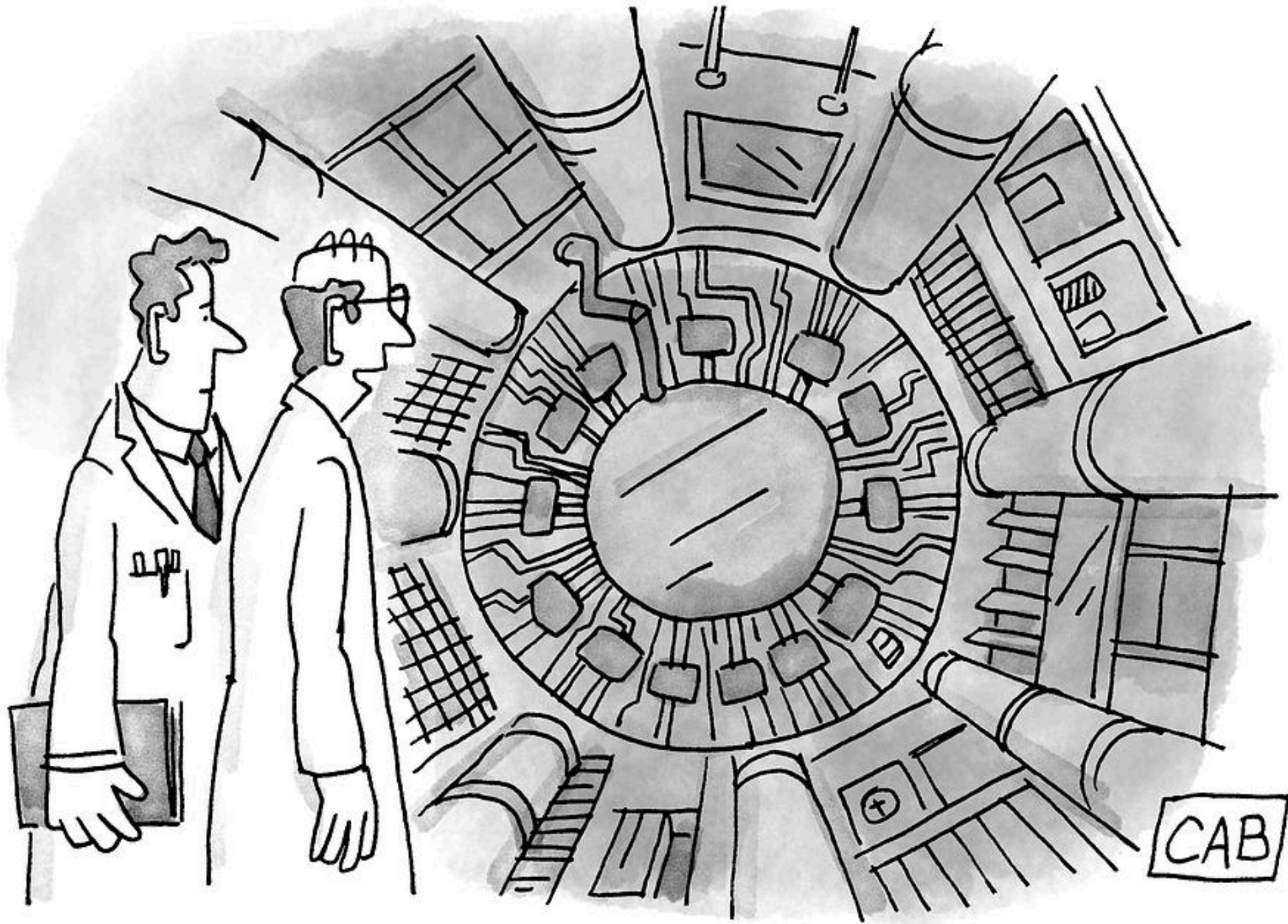
Metric	Value [%]	
	Original	Retrained
Recall	94.9	96.3
Specificity	85.0	87.8
Precision	53.3	58.7
False negative rate	5.2	3.7
False positive rate	15.0	12.2
F1-score	68.0	72.9
Balanced accuracy	90.0	92.1

Summary

- ML has become an omnipresent tool in HEP.
- It is key in the experimental domain, as they have the most financial resources.
- Plans for full AI MCEGs, data acquisition and data analysis have been proposed/discussed for future colliders (e.g. EIC).
- Its use has extended to phenomenology and, lately, to theory.
- You can find an incomplete list of articles in

<https://iml-wg.github.io/HEPML-LivingReview/>

- LHC will undergo an upgrade to become the HL-LHC, and searches for NP will go on.



“Once you have a collider, every problem starts to look like a particle.”