PHY654

Machine learning (ML) in particle physics

Swagata Mukherjee • IIT Kanpur

26th September 2024

top–tagging using CNN

- Today we will run an example code of **CNN** which performs **top-tagging** using collider simulation.
- Signal: top-quark, Background: any other quark/gluons. Classification problem.
- We will start from basics and discuss several things needed to understand this problem. We will discuss in the context of proton-proton collision at LHC.

CNN code (jet images):
https://github.com/swagata87/IITKanpurPhy654/blob/main/jetImage_CNN.ipynb

Units

Our scale	
Length	m
Mass	kg
Time	s
Energy	kg m^2 s^{-2}

Particle Physics	
Length	fm
Mass	eV/c^2
Time	s
Energy	eV

Convert

1 eV = 1.6 x 10^{-19} J

1 GeV = 10^9 eV

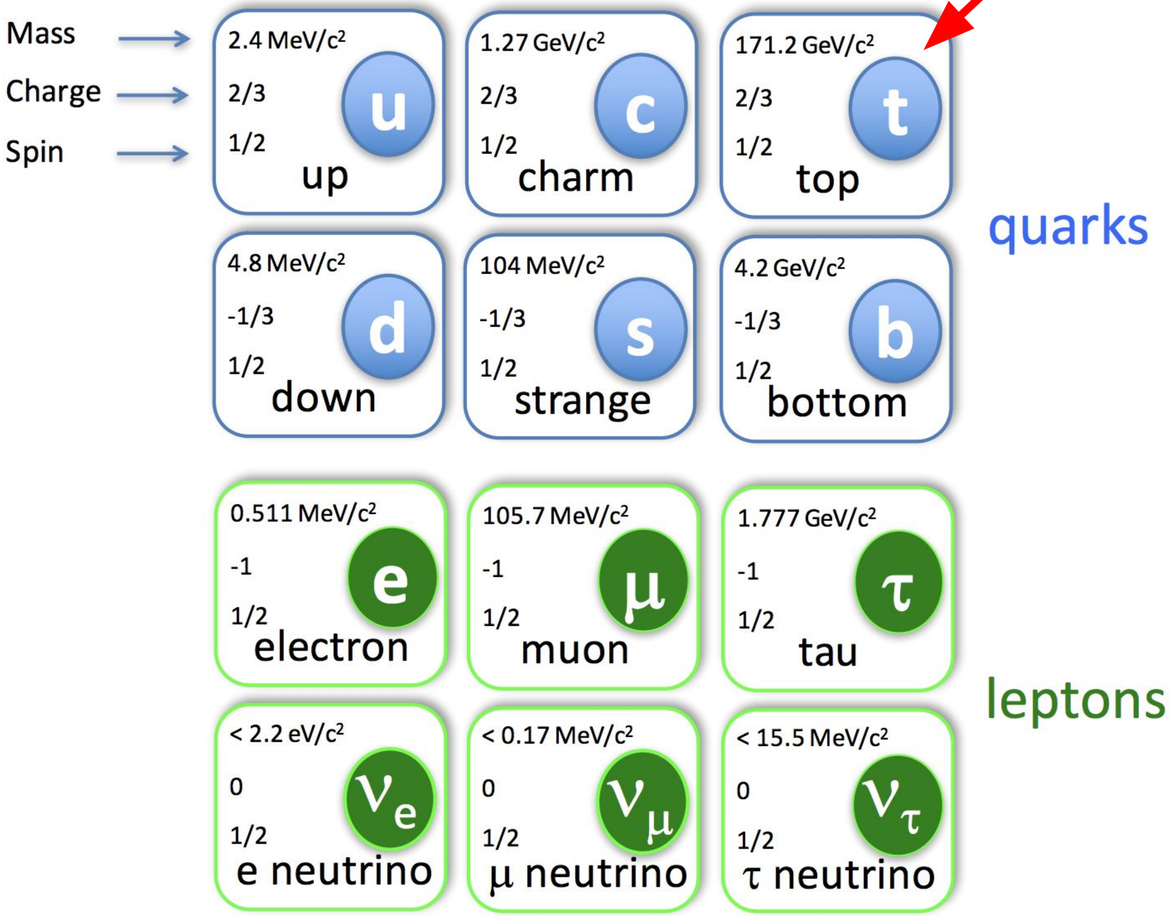
1 TeV = 10^3 GeV

1 fm = 10^{-15} m

Note: often set $\hbar = c = 1$

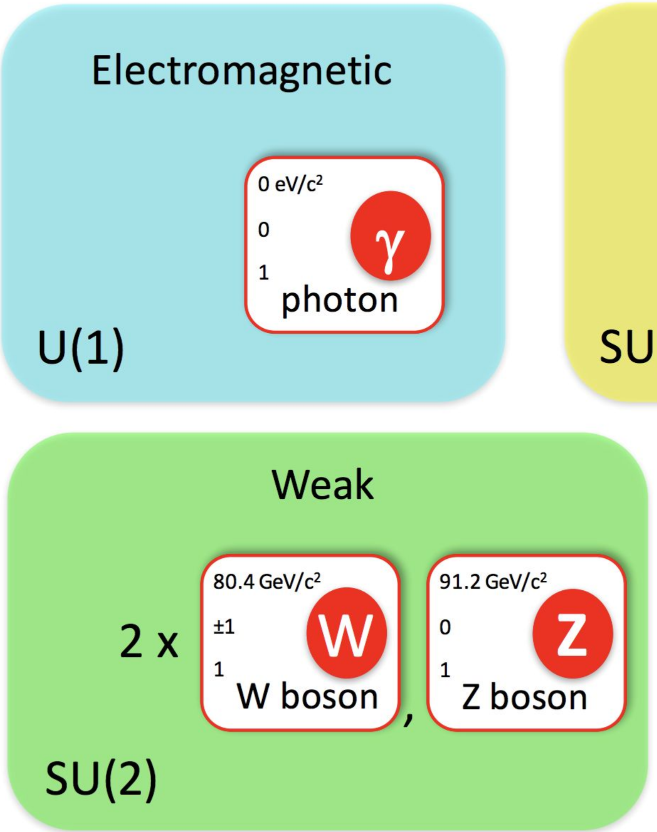
3 generations

Top (this is what we want to tag, i.e. find)



+ anti-particles

Matter is held together by forces. Mediated by force-carrying particles.

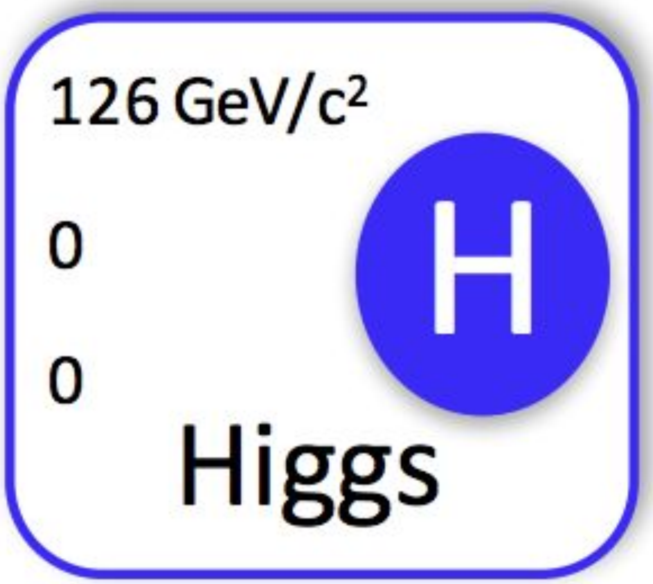


Note:
No gravity!!

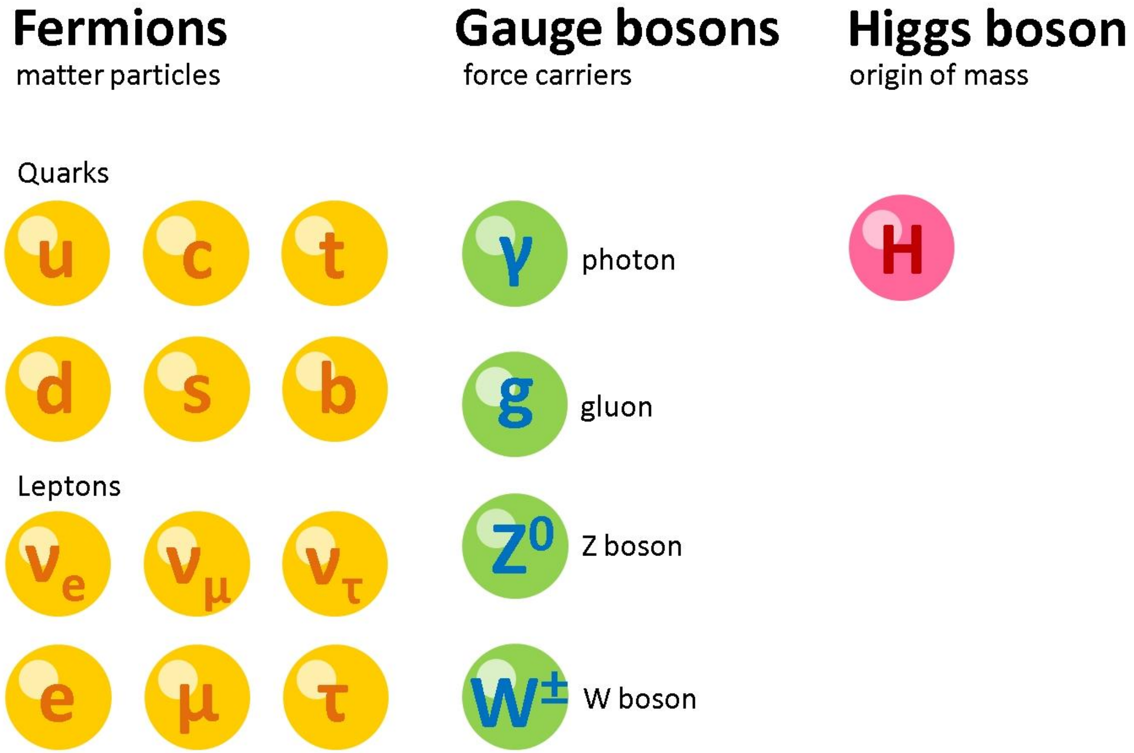
Strong:	$\alpha_s \sim 1$
Electromagnetic:	$\alpha_{em} \sim 1/137$
Weak:	$\alpha_W \sim 10^{-6}$

Note: these are low energy / large distance values. Coupling strength changes with energy.

And, the Higgs boson.
This was needed to write down mass-terms of W and Z.



The Standard model of particle physics



The strength of strong force grows with distance.

Implication → Confinement

No free quarks or gluons can be detected in a detector.

Quarks emit gluons.

Gluons split to quarks or gluons.

Showering

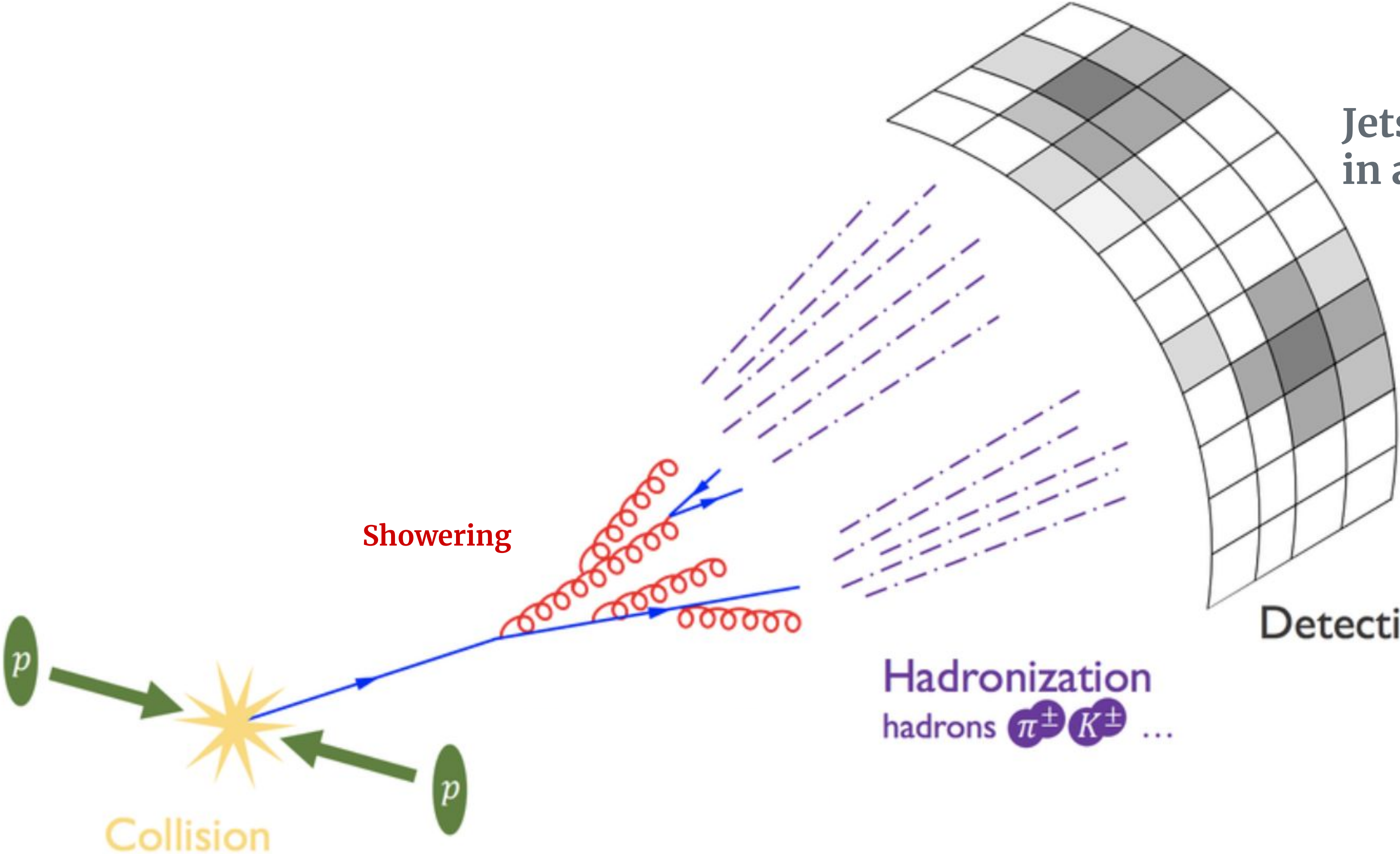
Finally, the quarks form colourless hadrons

- Baryons (3 quarks)
- Mesons (quark and anti-quark)

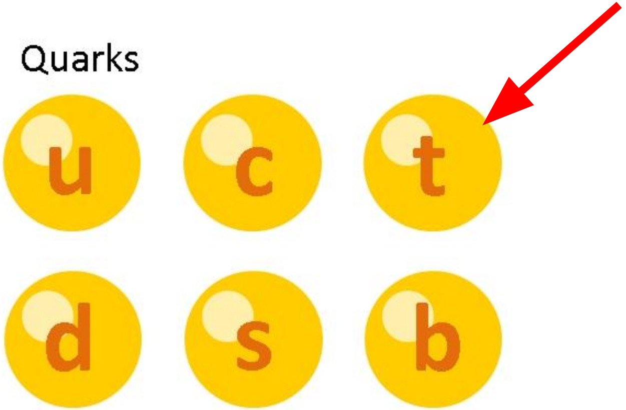
Hadronization

Spray of hadrons are detected in detectors as jets.

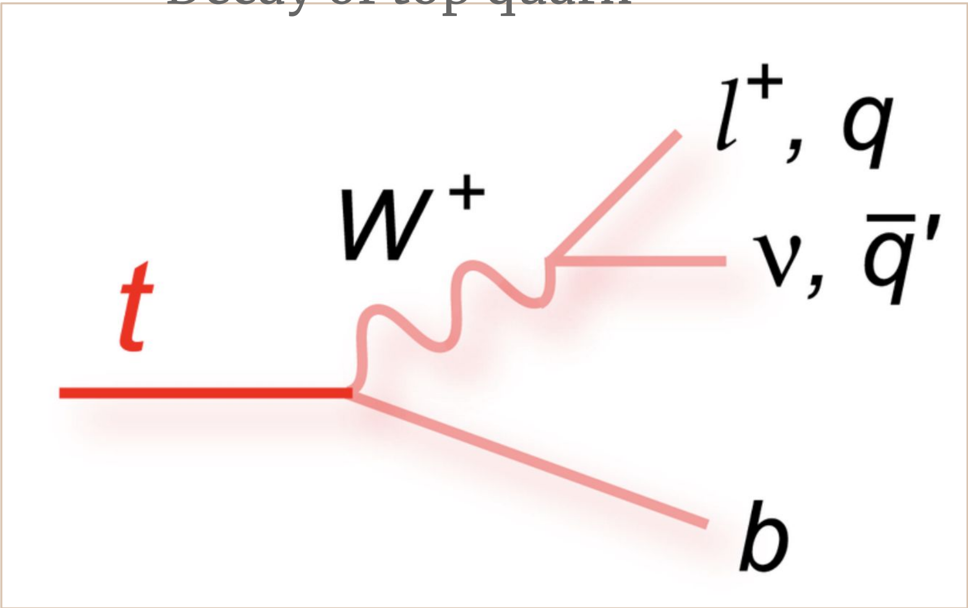
Jets are seen in a detector



Top quark is special.
It is the heaviest elementary particle known to us.
It decays before hadronization can occur.

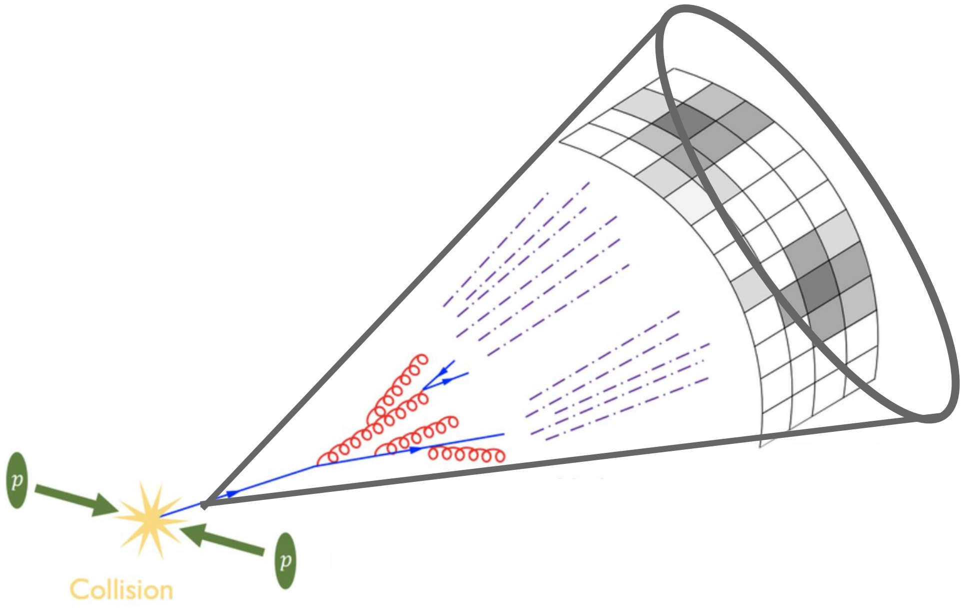


Decay of top quark



Jets

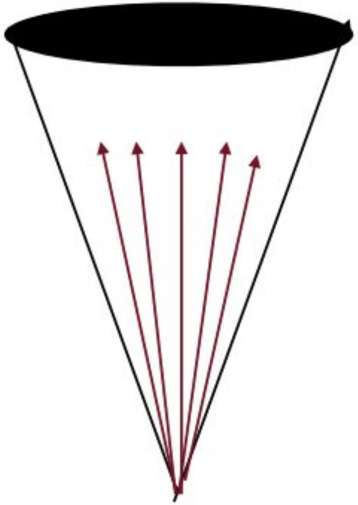
Jets: collimated spray of particles coming from showering and hadronization of quark/gluon
Basic idea: construct a cone which captures all the hadrons from the initial parton



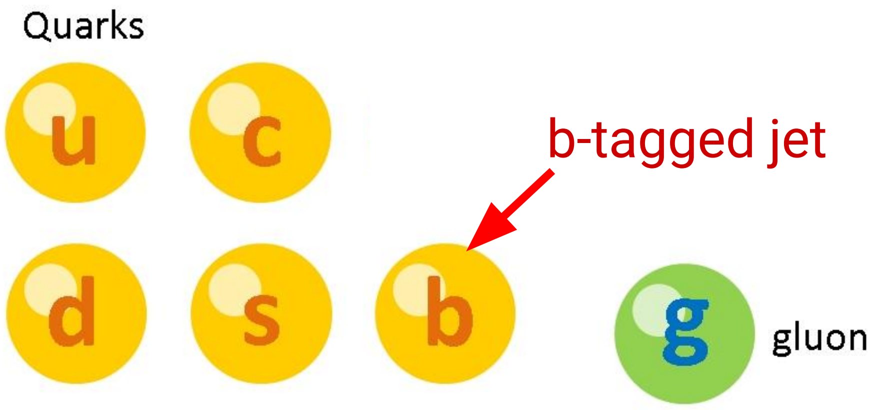
Jet: which cone size to choose?

May not
Capture
All the
particles

Capture
All the
particles



These particles are detected as jets in the detector



Some of the quarks decay, but they decay after hadronization

These particles are detected as it is in a typical collider detector.

These particles have very short lifetime, so they decay to other (stable) particles as soon they are produced. They can be indirectly seen via their decay products.

Neutrinos will escape collider detectors. They can be indirectly seen via missing transverse momentum → an imbalance in momentum in transverse plane

Particles that we directly see in detectors are...

Electron, **muon**, **photon**, and some hadrons (mostly charged pion, charged kaon, neutral kaon, proton, neutron)

$$\left.\begin{array}{lll} e^{\pm} & m_e = 0.511\ \text{MeV} & \\ \mu^{\pm} & m_\mu = 105.7\ \text{MeV} & \sim 200\, m_e \\ \gamma & m_\gamma = 0\ ,\quad Q = 0 & \end{array}\right\}\ \text{EM}$$

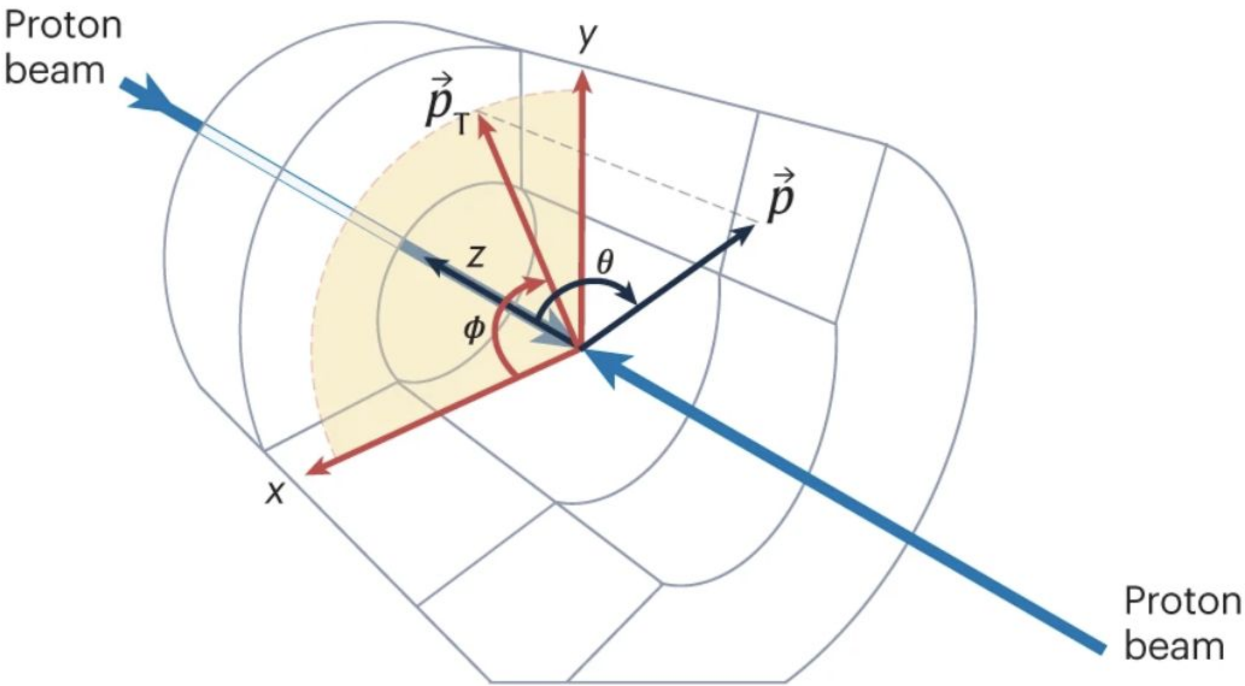
$$\left.\begin{array}{lll} \pi^{\pm} & m_\pi = 139.6\ \text{MeV} & \sim 270\, m_e \\ K^{\pm} & m_K = 493.7\ \text{MeV} & \sim 1000\, m_e \\ p^{\pm} & m_p = 938.3\ \text{MeV} & \sim 2000\, m_e \end{array}\right\}\ \begin{array}{l} \text{EM, Strong} \\ \sim 3.5\, m_\pi \\ \end{array}$$

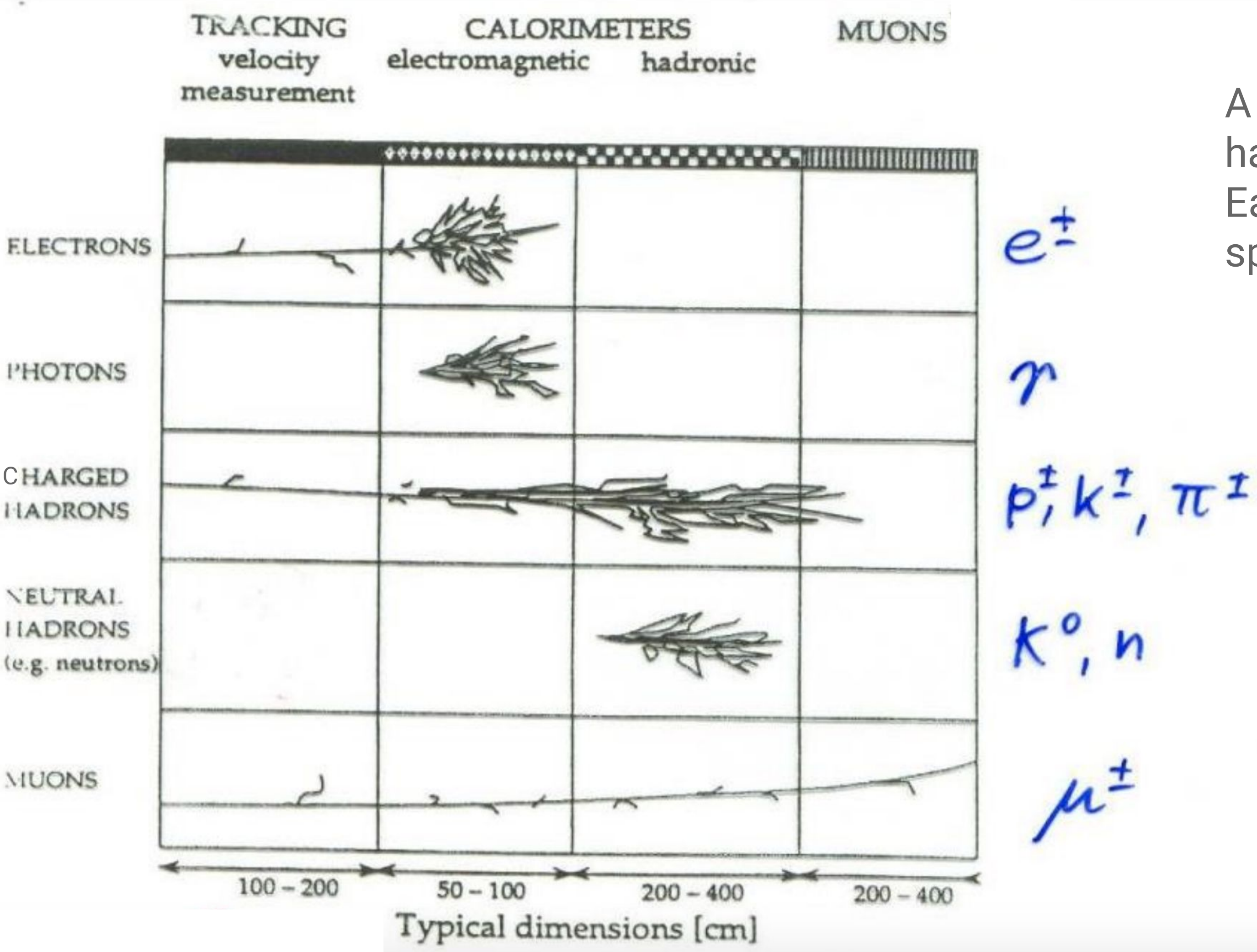
$$\left.\begin{array}{lll} K^{0} & m_{K^0} = 497.7\ \text{MeV} & Q = 0 \\ n & m_n = 939.6\ \text{MeV} & Q = 0 \end{array}\right\}\ \text{Strong}$$

A typical collider detector should be able to identify these particles, and measure the energy and momentum

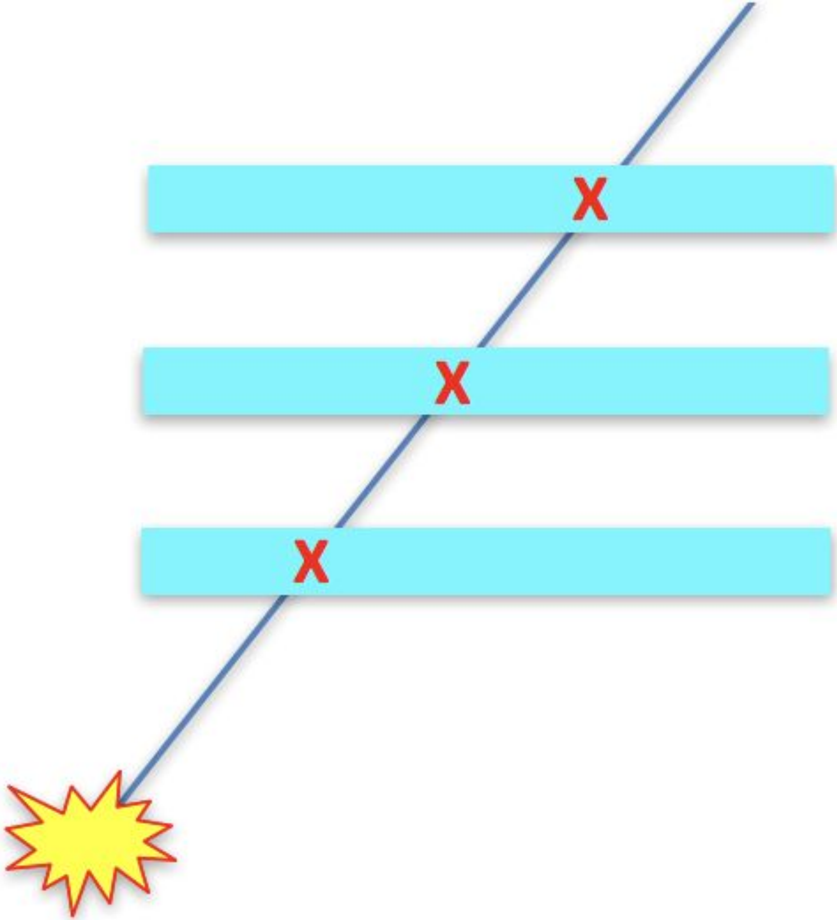
Their difference in mass, charge and interaction is the key to their identification.

Cylindrical detector. Almost 4Π coverage around collision-point.

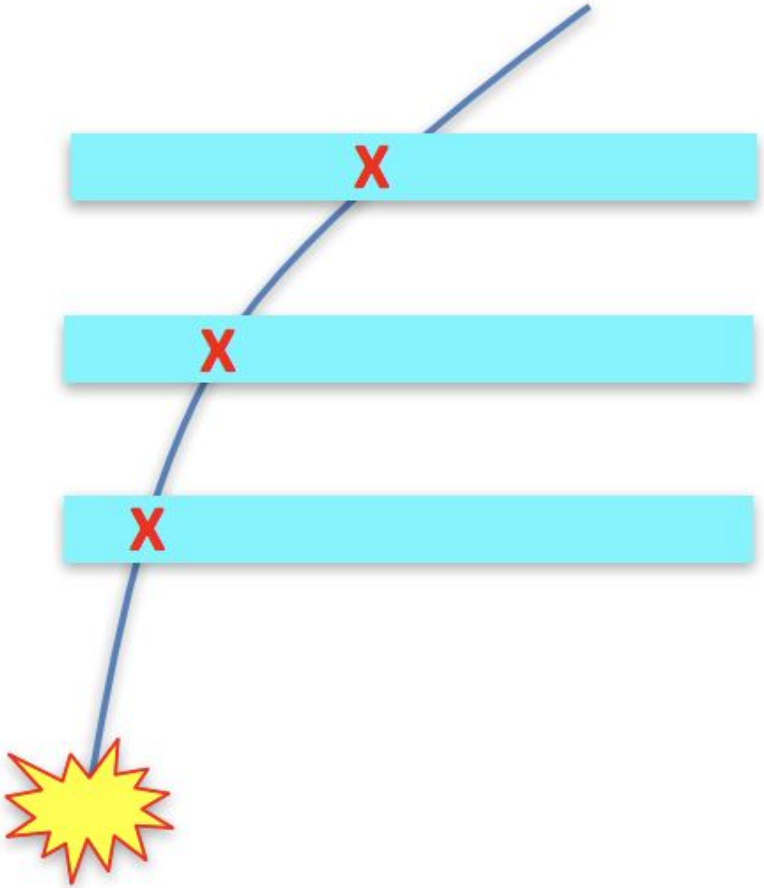




A typical detector has several layers, Each layer serve a specific purpose.



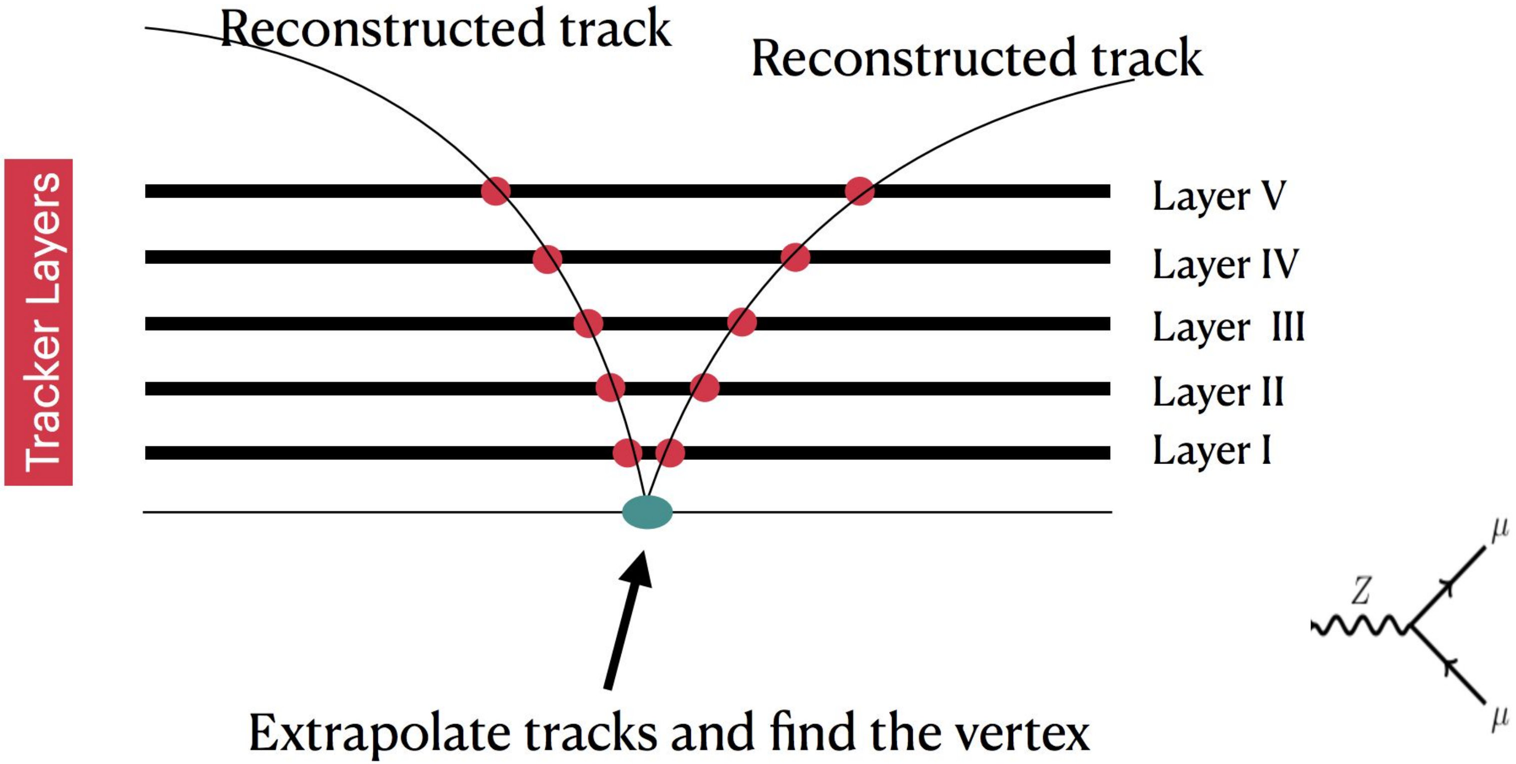
- Inner tracking detector.
- Detect charged particles.
- Determine location of charged particles.

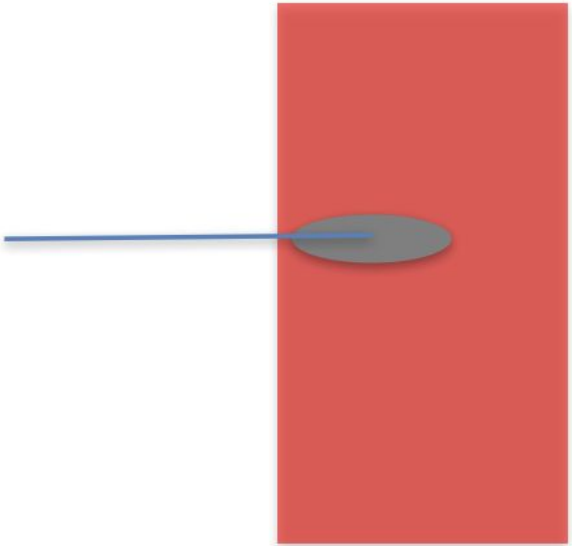


- Inner tracking detector placed inside a magnetic field (B).
- Measure momentum of charged particles.
- Track curvature, B and momentum are related.

Radius of Curvature = $r = \frac{p_T}{0.3B}$

Location of the collision vertex





Calorimeters

Charged + neutral particles

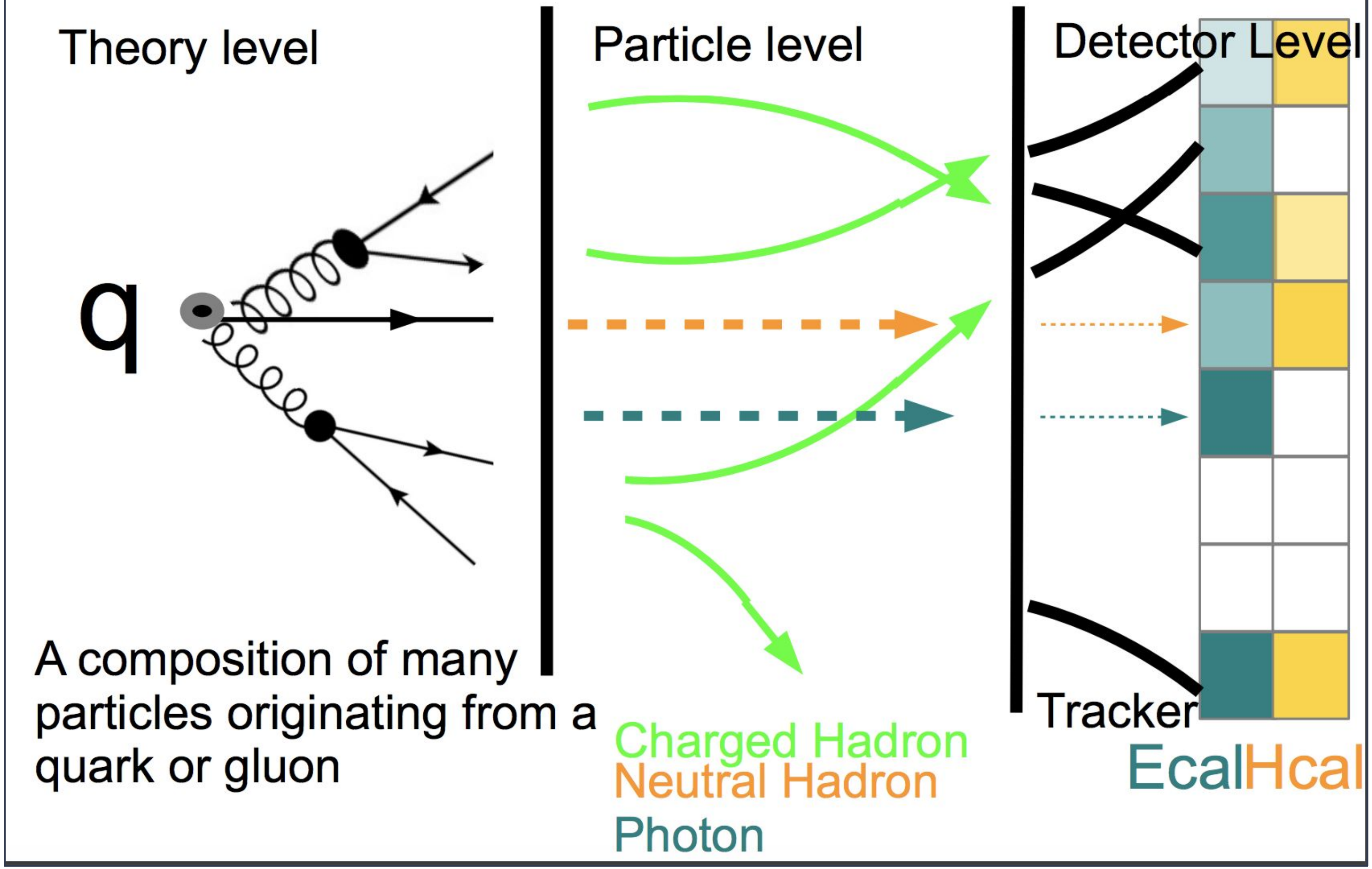
Two types:

- Electromagnetic
- Hadronic

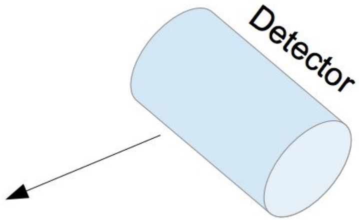
Absorb + measure energy

Calorimeters are destructive detectors, unlike tracker.

What is a jet?



Theory level
Particle level
Detector Level
q
A composition of many particles originating from a quark or gluon
Charged Hadron
Neutral Hadron
Photon
Tracker
EcalHcal



Detector

Jet formation algorithm

Starts with N stable particles

$$R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$$

Kt algorithm

$$d_{ij} = min(p_{Ti}^2, p_{Tj}^2)\,\frac{R_{ij}^2}{R^2}$$

$$d_{iB} = p_{Ti}^2$$

Anti-Kt algorithm

$$d_{ij} = min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2}\right)\frac{R_{ij}^2}{R^2}$$

$$d_{iB} = \frac{1}{p_{Ti}^2}$$

Cambridge Aachen algorithm

$$d_{ij} = \frac{R_{ij}^2}{R^2}$$

$$d_{iB} = 1$$

Find the minimum of $\{d_{ij}, d_{iB}\}$ for the entire set of N particles

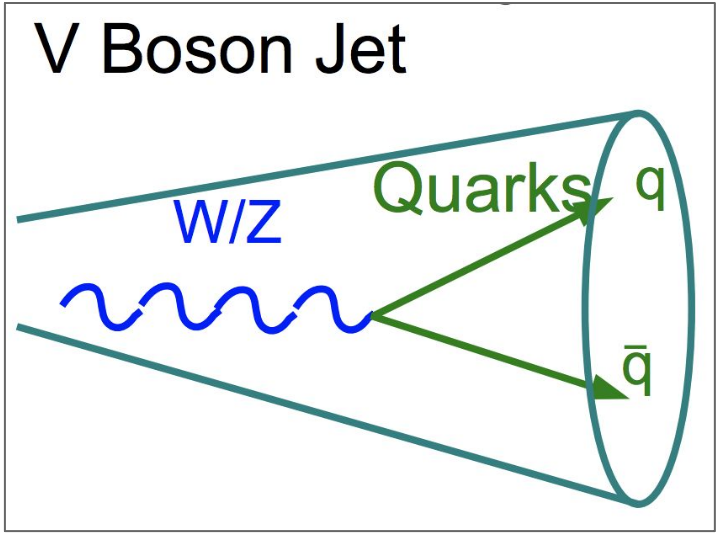
If some d_{ij} is minimum => combine i and j particle into one particle (number of particle is reduced to N-1)

If some d_{iB} is minimum => i^{th} particle is declared as a jet and removed from the list

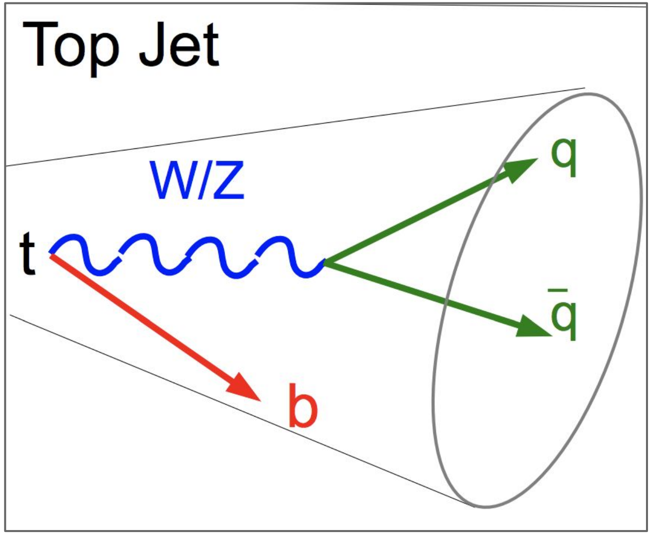
AK4 is a popular choice. Anti-kT, with cone size=0.4

Jet substructure

Large-radius jet (also called fat-jet).



2 sub-jets



3 sub-jets

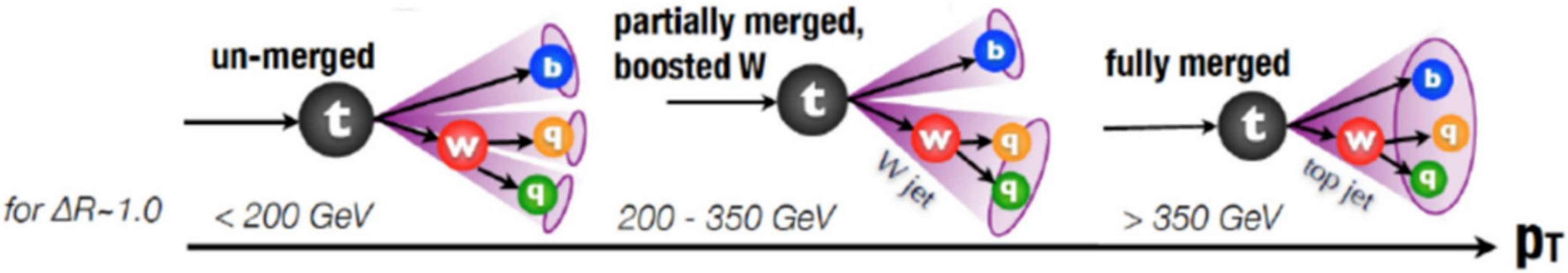
The term **boosted** applies to particles with p_T>2*mass.

The opening angle of the decay products is: → $\Delta R = \frac{2m}{p_T}$

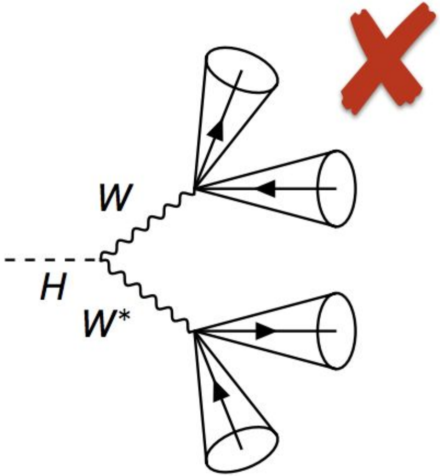
For H—>bb to be captured in a AK8 jet (jet cone of radius 0.8)
0.8 = 2*125 / pT , where Higgs mass ~ 125 GeV
=> pT = 2*125 / 0.8 ~ **312** GeV → Higgs pT have to be this high or more.

For top to be captured in a AK10 jet (jet cone of radius 1.0)
1.0 = 2*173 / pT , where top mass ~ 173 GeV
=> pT = 2*173 / 1.0 ~ **346** GeV → Top pT have to be this high or more.

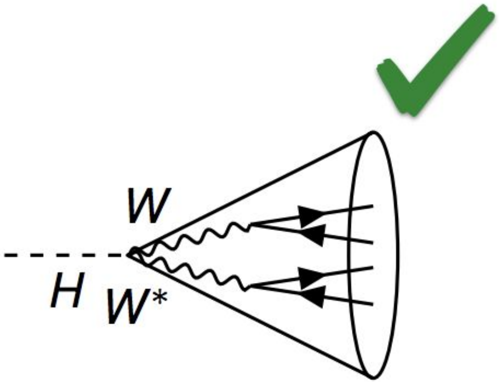
For top to be captured in a AK10 jet (jet cone of radius 1.0)
1.0 = 2*173 / pT , where top mass ~ 173 GeV
=> pT = 2*173 / 1.0 ~ **346** GeV → Top pT have to be this high or more.



Another example: H→WW*→4q tagging



Overwhelming background

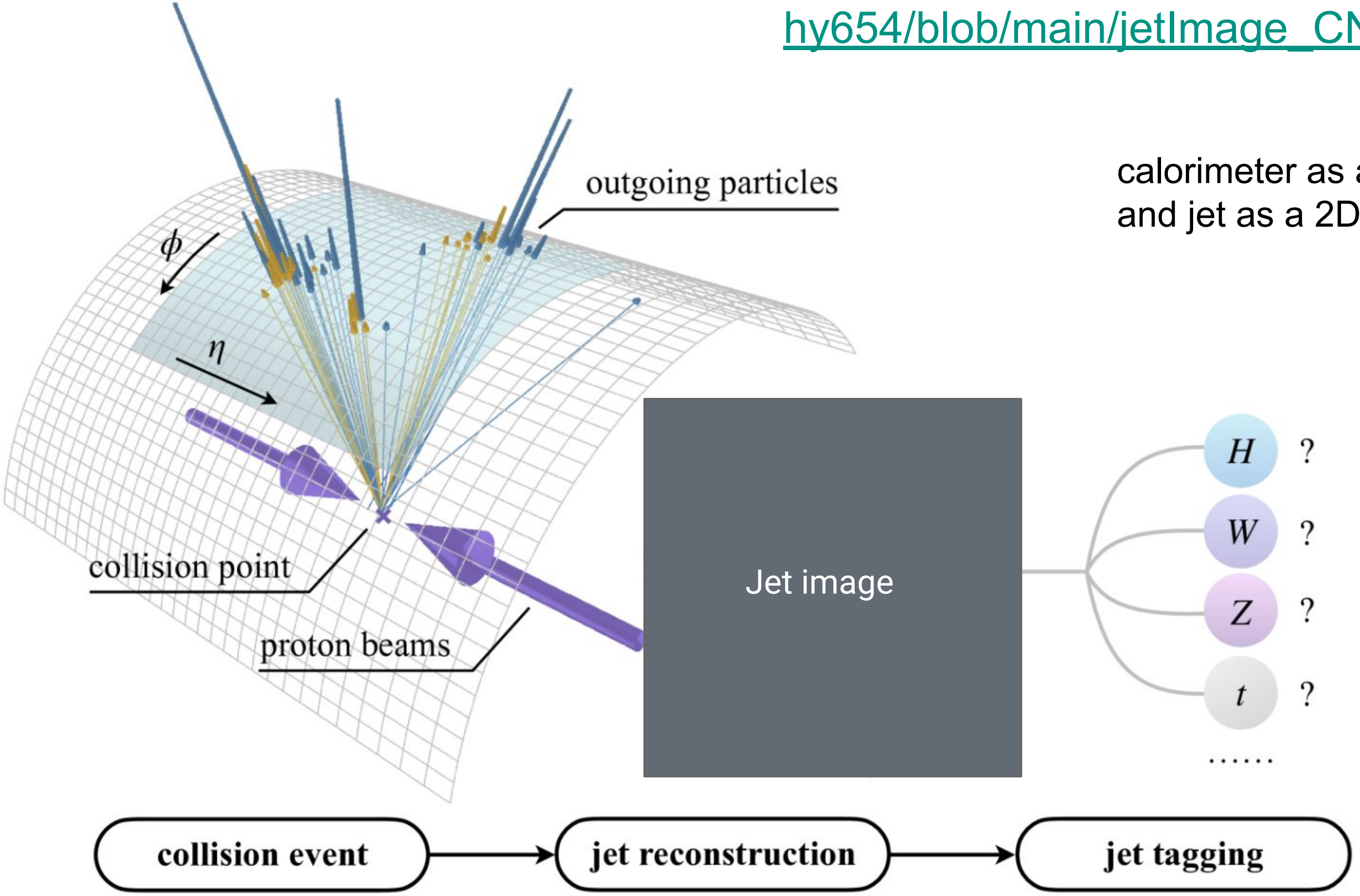


Potentially useful

Jet as an image?

Coming back to the CNN code (jet images): https://github.com/swagata87/IITKanpurPhy654/blob/main/jetImage_CNN.ipynb

calorimeter as a camera
and jet as a 2D image



References

- CERN summer school lectures
  - https://indico.cern.ch/category/345/
- Lectures on Collider Physics by Biplob Bhattacherjee (IISc Bengaluru) in Sangam school at HRI, 2024.
  - https://www.hri.res.in/~sangam/sangam24/lectures/BB/collider1.pdf
  - https://www.hri.res.in/~sangam/sangam24/lectures/BB/collider2.pdf
  - https://www.hri.res.in/~sangam/sangam24/lectures/BB/collider3.pdf

Where do we use ML in High energy physics?

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

Regression

- Pileup
- Calibration
- Recasting
- Matrix elements
- Parameter estimation
- Parton Distribution Functions (and related)
- Lattice Gauge Theory
- Function Approximation

Formal Theory and ML

- Theory and physics for ML
- ML for theory

Reviews

- Modern reviews
- Specialized reviews
- Classical papers
- Datasets

Another useful website https://paperswithcode.com/

paperswithcode.com/method/resnet

Search

Browse State-of-the-Art Datasets Methods More

Convolutional Neural Networks

Residual Network

Introduced by He et al. in Deep Residual Learning for Image Recognition

Residual Networks, or **ResNets**, learn residual functions with reference to the layer inputs, instead of learning unreferenced functions. Instead of hoping each few stacked layers directly fit a desired underlying mapping, residual nets let these layers fit a residual mapping. They stack residual blocks ontop of each other to form network: e.g. a ResNet-50 has fifty layers using these blocks.

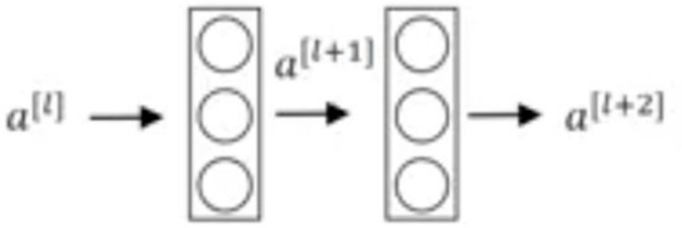
Formally, denoting the desired underlying mapping as $\mathcal{H}(x)$, we let the stacked nonlinear layers fit another mapping of $\mathcal{F}(x) := \mathcal{H}(x) - x$. The original mapping is recast into $\mathcal{F}(x) + x$.

There is empirical evidence that these types of network are easier to optimize, and can gain accuracy from considerably increased depth.

Source: Deep Residual Learning for Image Recognition

Read Paper See Code

Reminder of a neural net



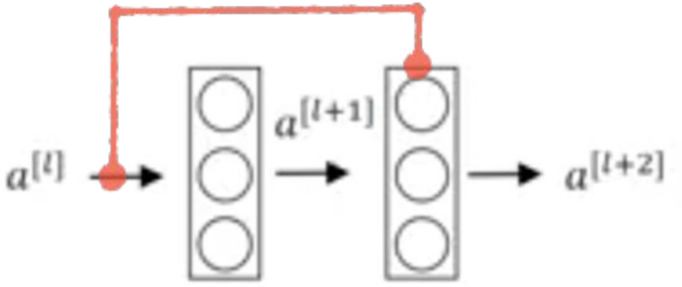
Example with 2 layers of
a neural network (NN)

a[l] → Linear → ReLU → Linear → ReLU → a[l+2]

z[l+1] = W[l+1]a[l] +b[l+1]   a[l+1] = g(z[l+1])   z[l+2] = W[l+2]a[l+1] +b[l+2]   a[l+2] = g(z[l+2])

For information to flow from a[l] to a[l+2] it needs to go through all of these steps in case of a usual NN.

ResNet



ResNets are built out of Residual blocks

a[l] → Linear → ReLU → Linear → ReLU → a[l+2]

z[l+1] = W[l+1]a[l] +b[l+1]

a[l+1] = g(z[l+1])

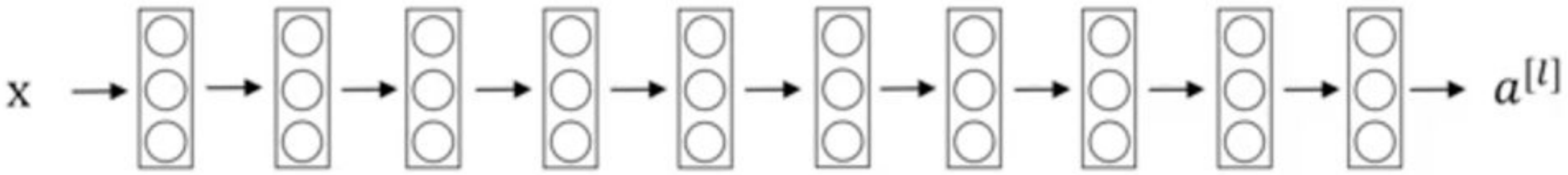
z[l+2] = W[l+2]a[l+1] +b[l+2]

~~a[l+2] = g(z[l+2])~~

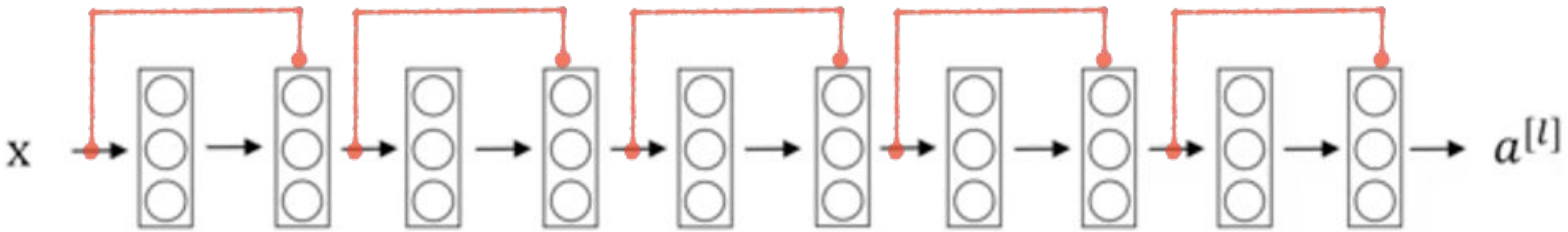
a[l+2] = g(z[l+2] +a[l])

ResNet

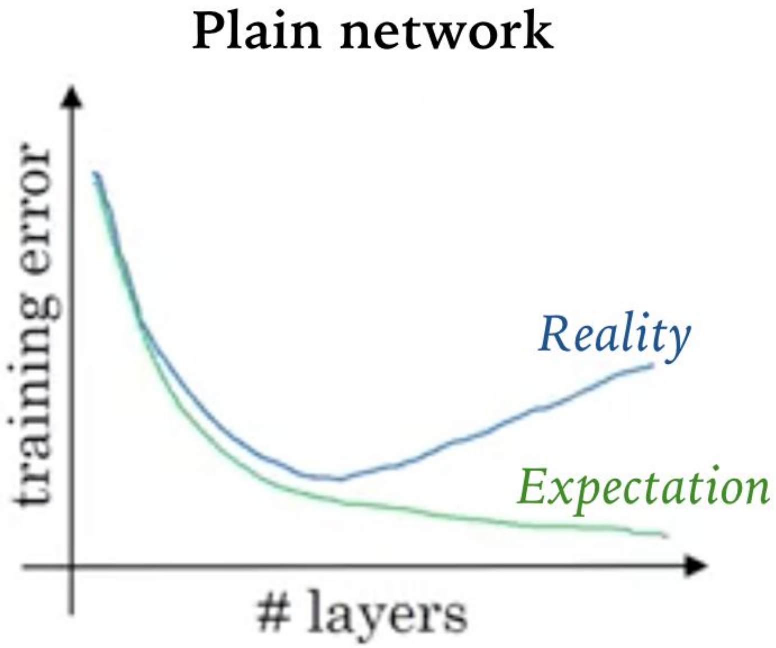
Take many residual blocks and stack them together to build a ResNet.

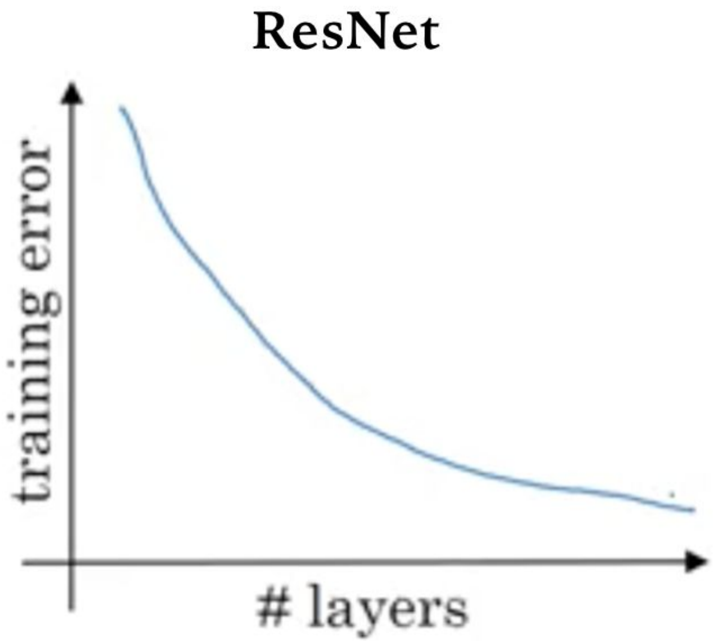


Plain Network



ResNet





Deep Residual Learning for Image Recognition, Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, https://arxiv.org/pdf/1512.03385.pdf