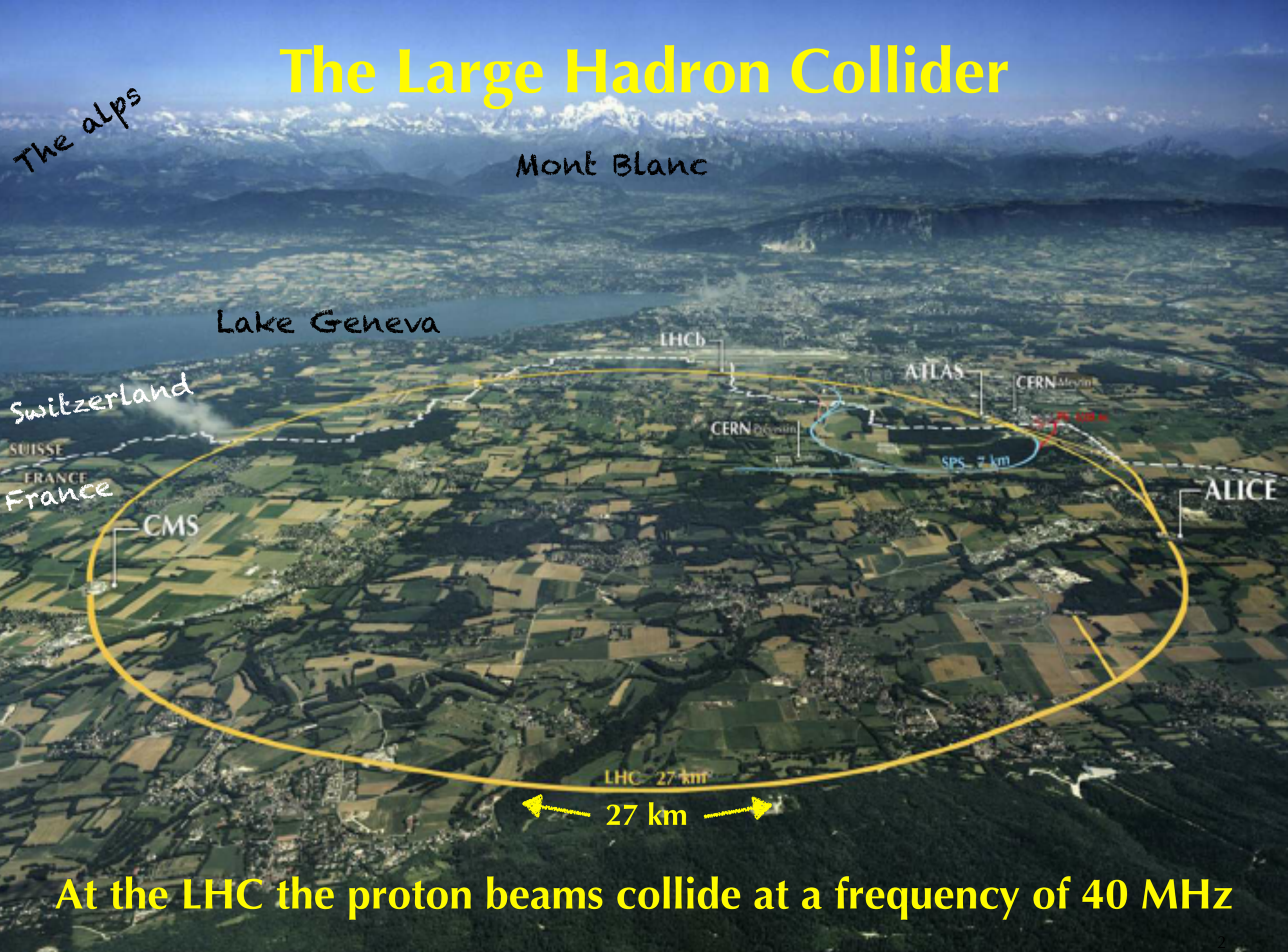
# Intro to LHC jet tagging task

Jennifer Ngadiuba (Fermilab)

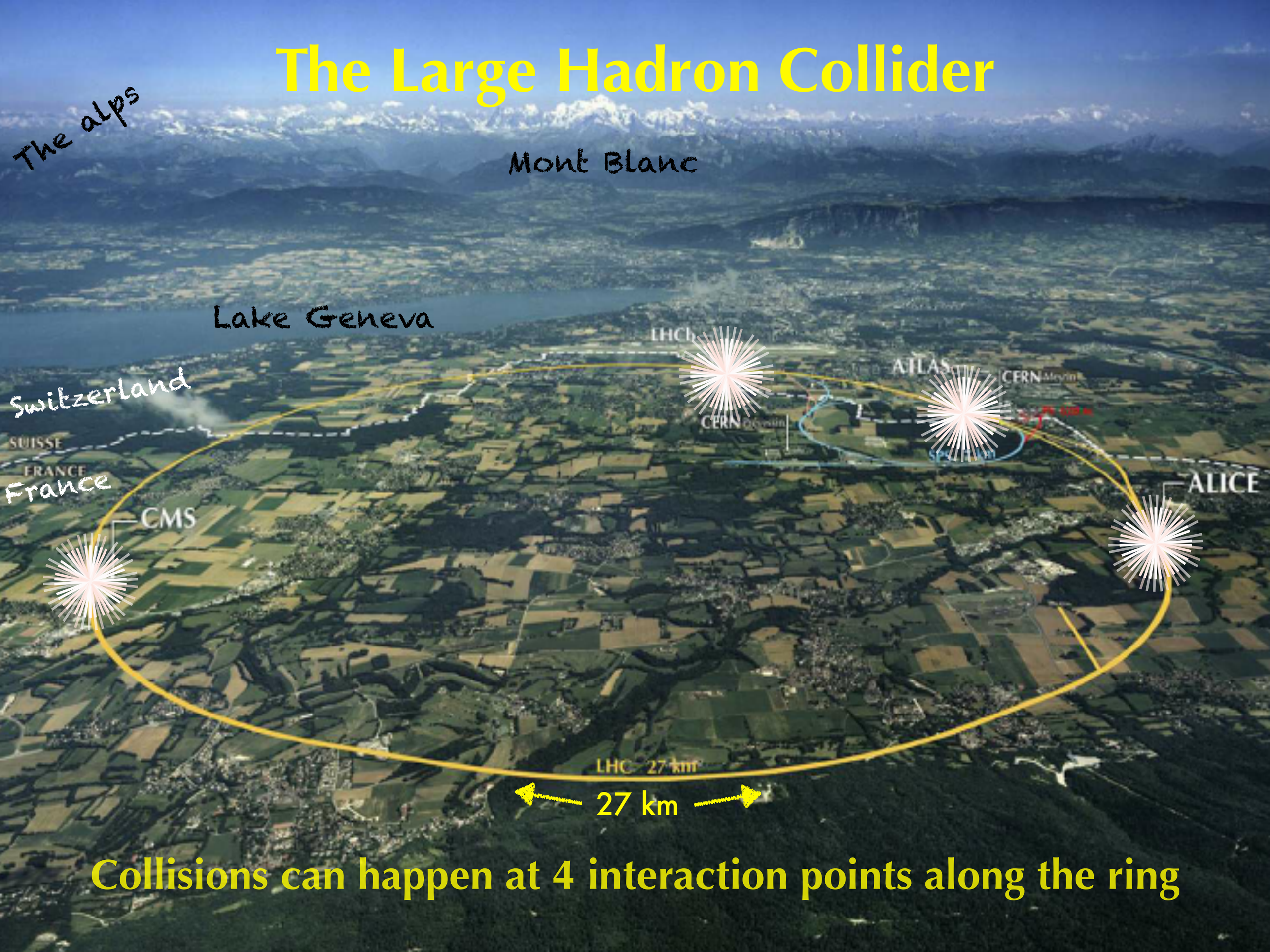
University of Pavia, May 8-12 2023

# The Large Hadron Collider



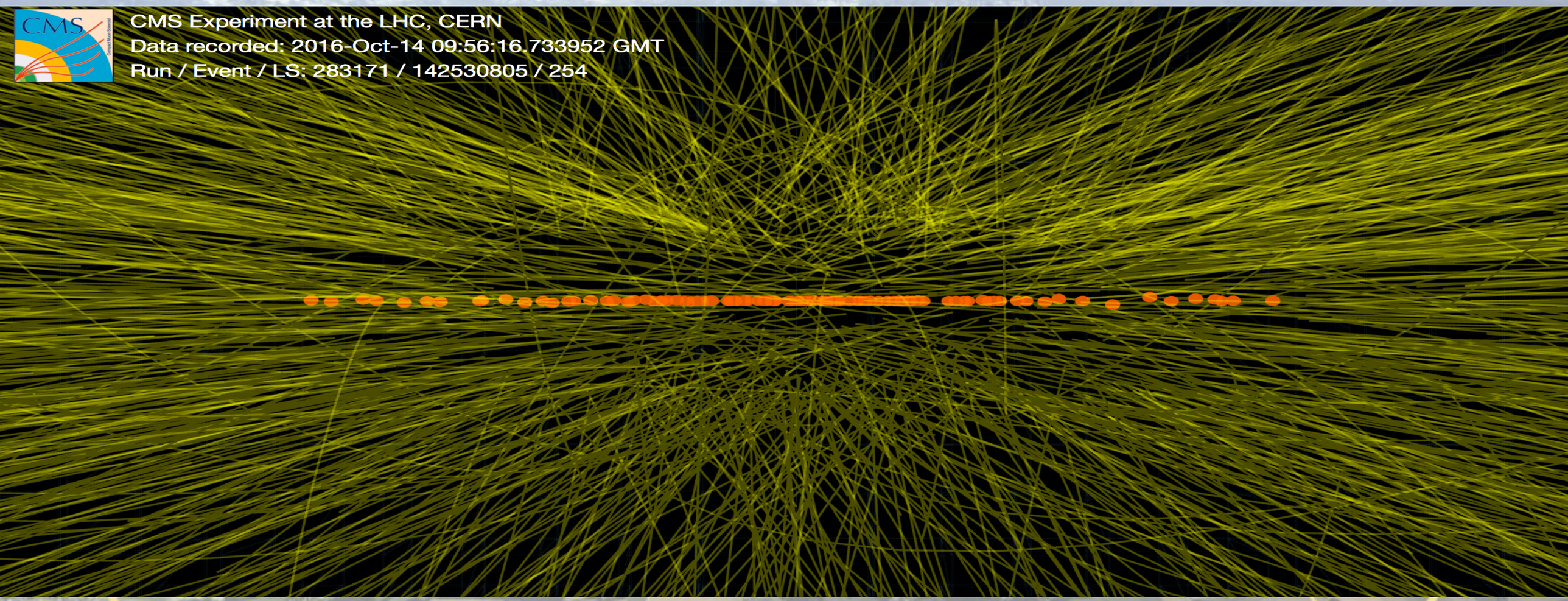

**At the LHC the proton beams collide at a frequency of 40 MHz**

# The Large Hadron Collider




**Collisions can happen at 4 interaction points along the ring**

# The Large Hadron Collider





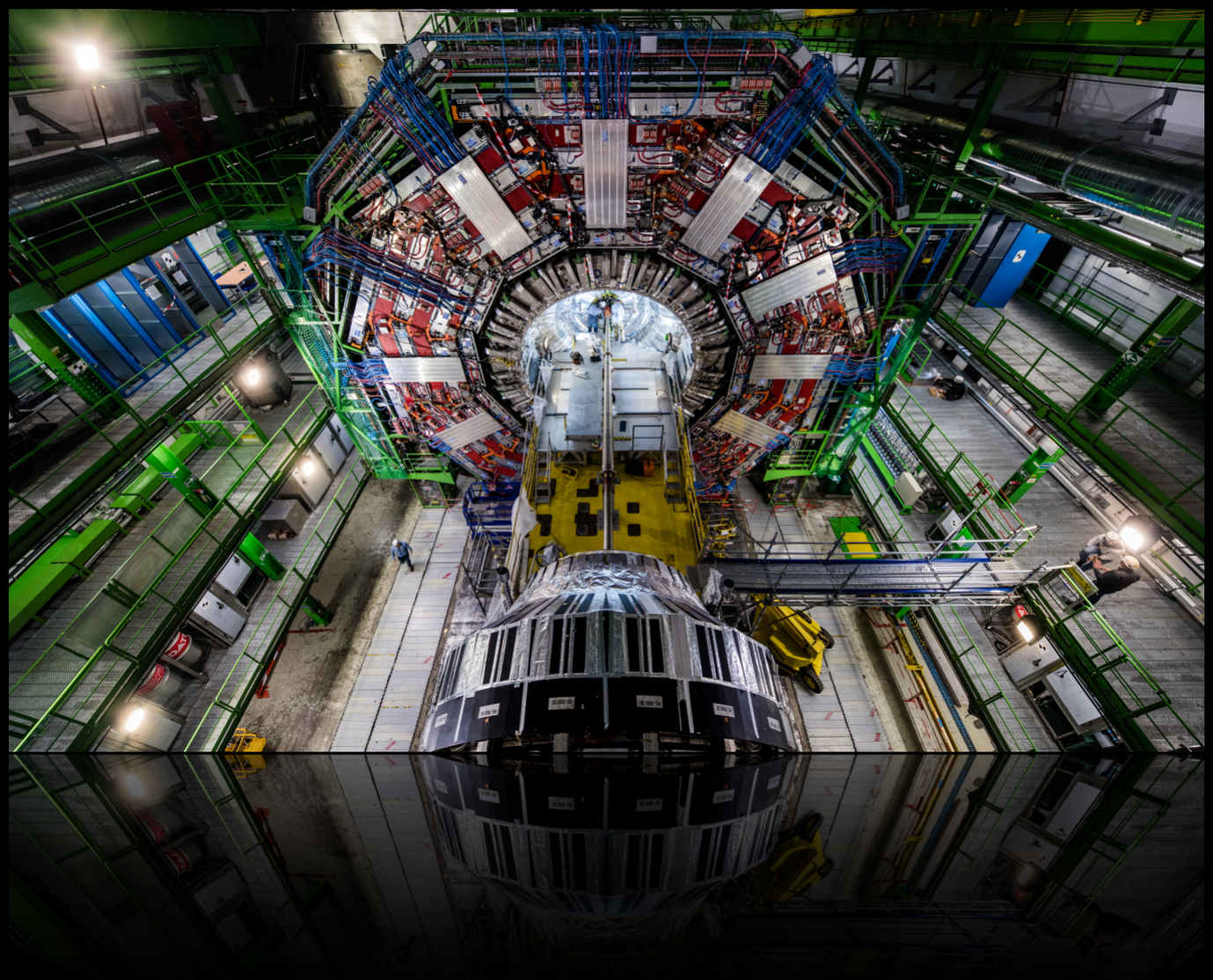
**Each collision produces O($10^3$) particles!**

# The Large Hadron Collider

**ex, the Compact Muon Solenoid**

The detectors have O($10^8$) sensors used to detect these particles

# The CMS detector



A giant, high-speed and high-resolution camera, taking 3D “photographs” of particle collisions.

With 100 million pixels can take 40 million pictures per second.

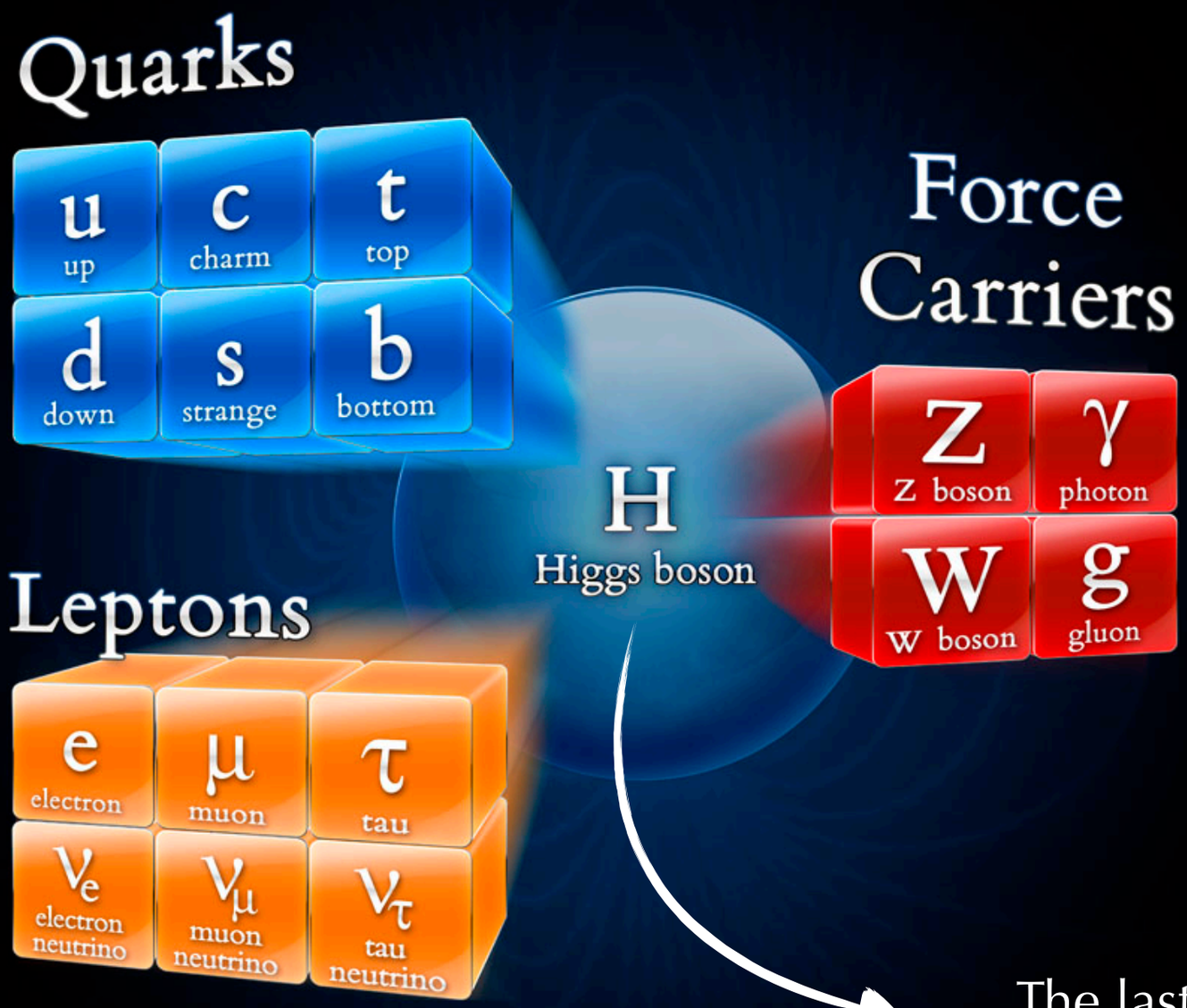
# Big data @ LHC

**Collision frequency: 40 MHz**
**Particles per collision: O($10^3$)**
**Detector resolution: O($10^8$) channels**
**Number of collaborators: O($10^3$)**
**Cost: O(10) B euros**


**Extreme data rates of O(100) Tb/s!**

# What do we know?

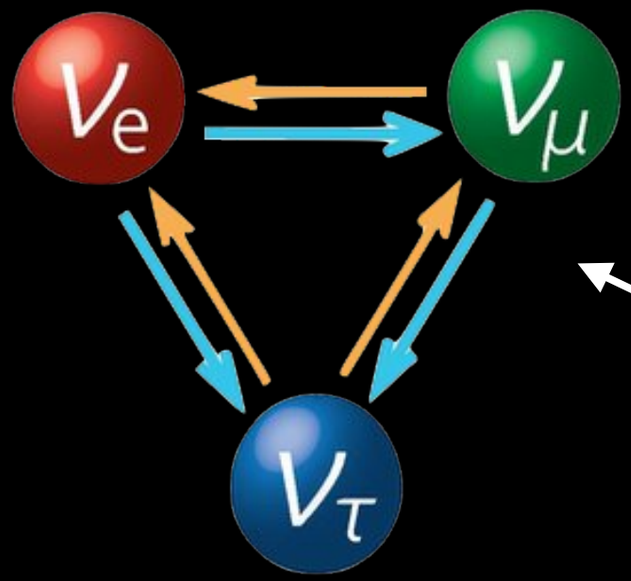
## The Standard Model of particle physics



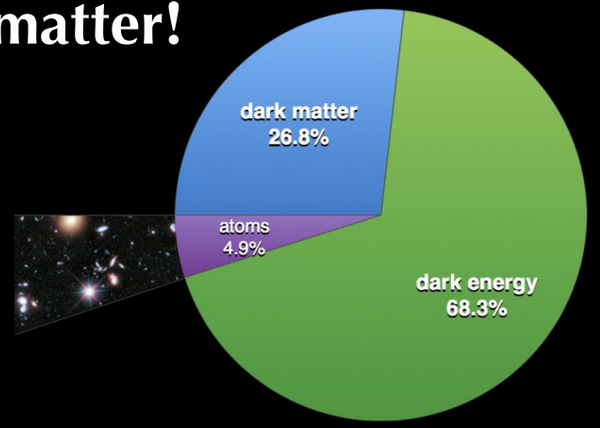

The last missing piece,
discovered in 2012 at the LHC!

# What we DON'T know?

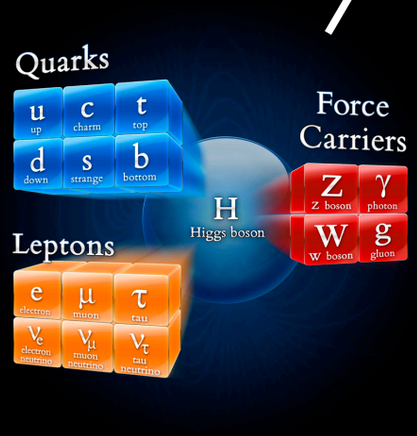
## Neutrino masses?



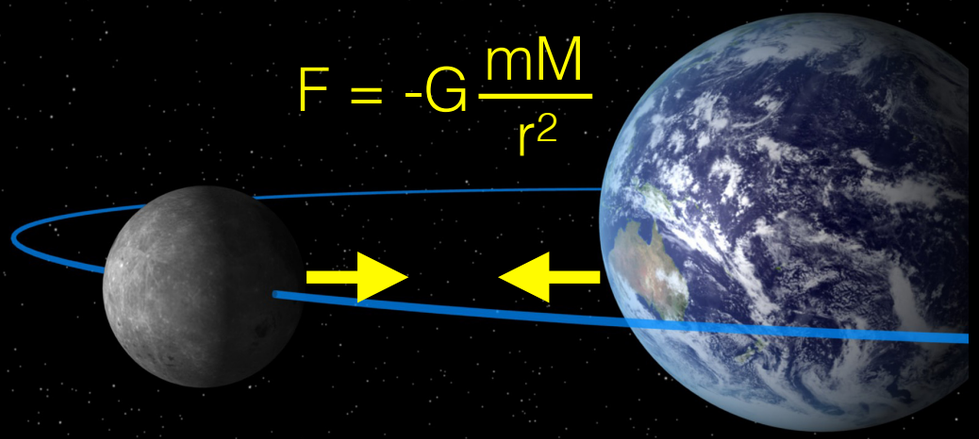

## Explains only 5% of the universe. No SM candidate for dark matter!



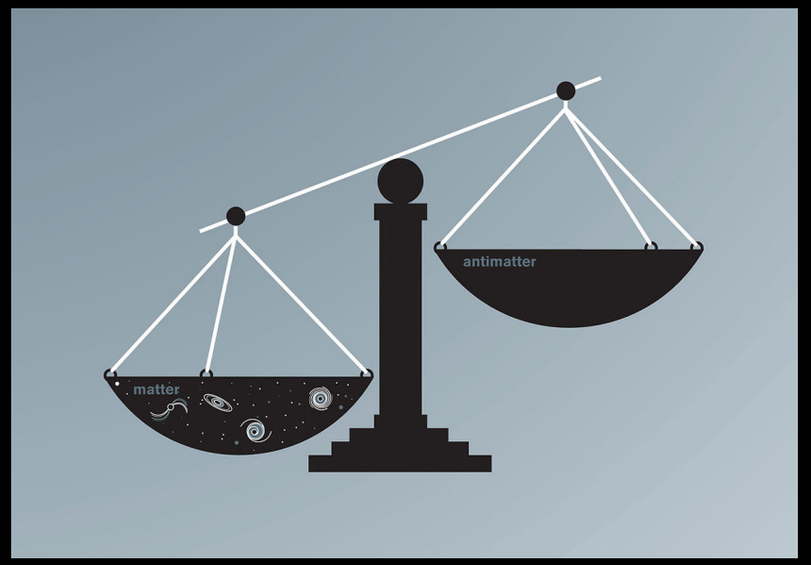

@AstroKatie/Planck13




## Gravity not included in the theory

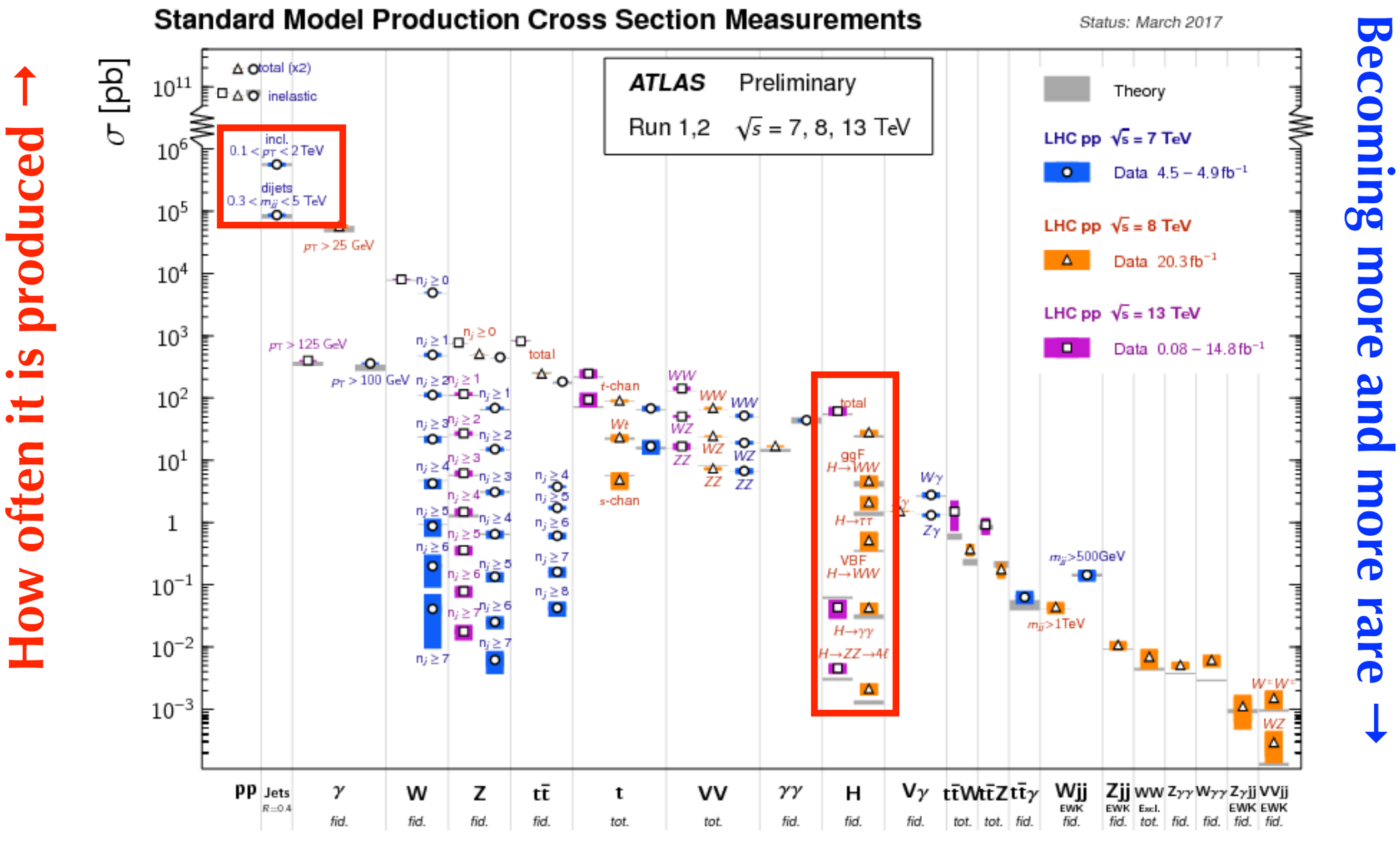



## Matter-antimatter asymmetry?



# Collisions which produce interesting products (ex: Higgs boson) are typically very rare

The probability of producing a Higgs boson is 5-9 orders of magnitude smaller than producing only jets



# Heavy jet tagging

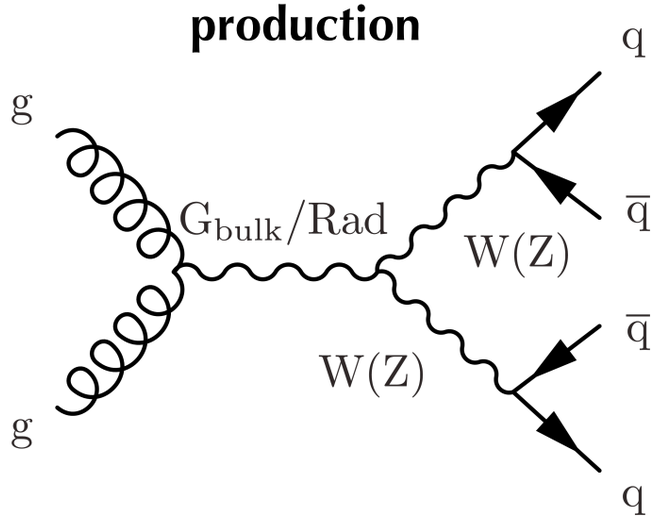
- **Identification of jets arising from hadronization of boosted W/Z/H/top is a key task in LHC physics:**

  - new physics searches, standard model measurements, higgs sector

  - **unique signature** from hadrons merging in single jet with substructure

  - exploit to suppress **overwhelming background from multijet** processes in most sensitive all-hadronic and semi-leptonic channels

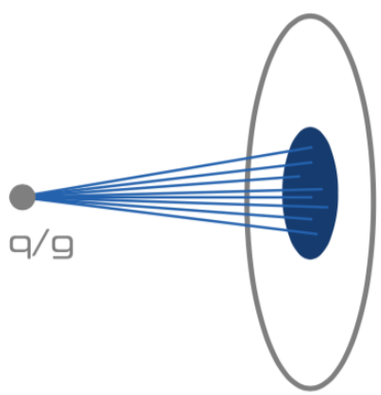
**ex: Graviton or Radion production**



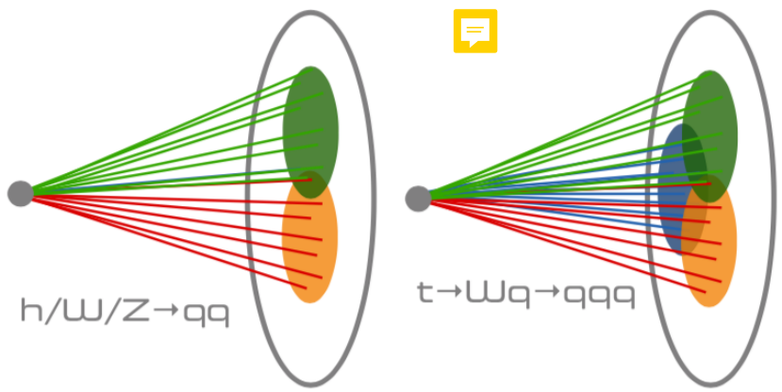

- A topic of interest in both theory and experiment communities since ~ 30 years

- Recent years advancement in ML enabled more powerful algorithms

BACKGROUND JET (single q/g)

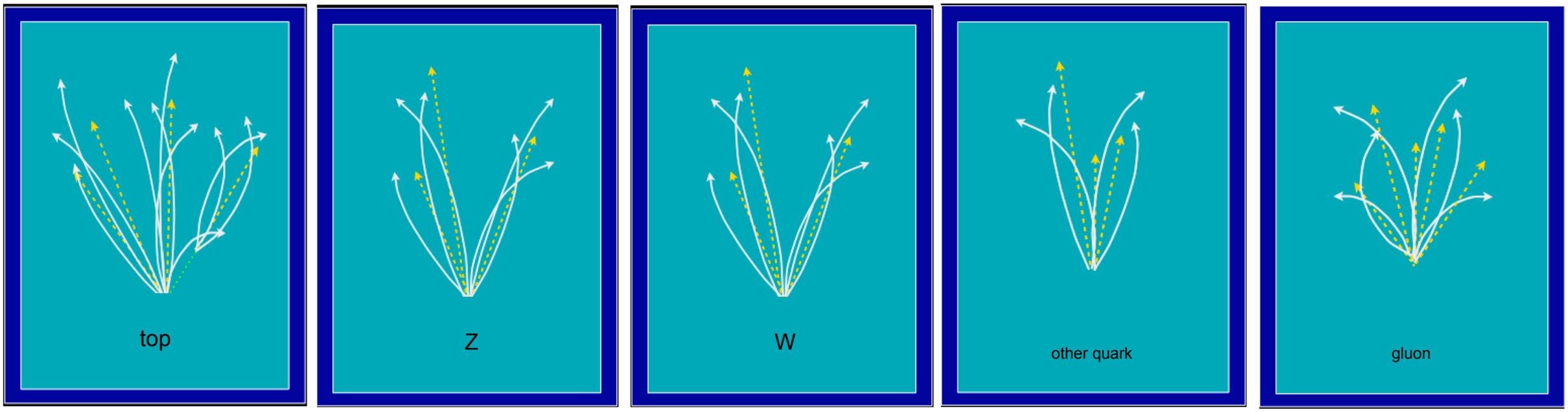



SIGNAL JETS



# Physics case: jet tagging

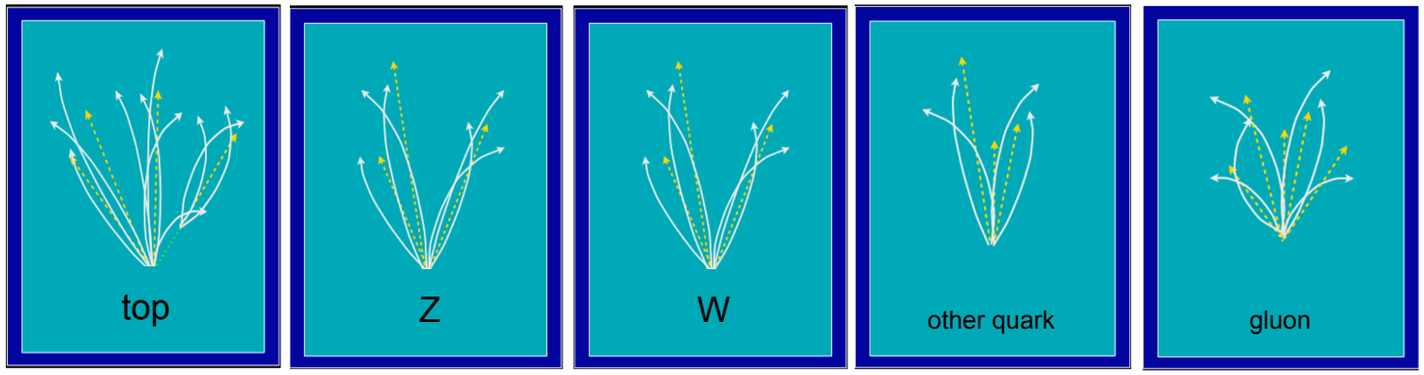
Study a **multi-classification task to be implemented on FPGA:** discrimination between highly energetic (boosted) ***q, g, W, Z, t*** initiated *jets*

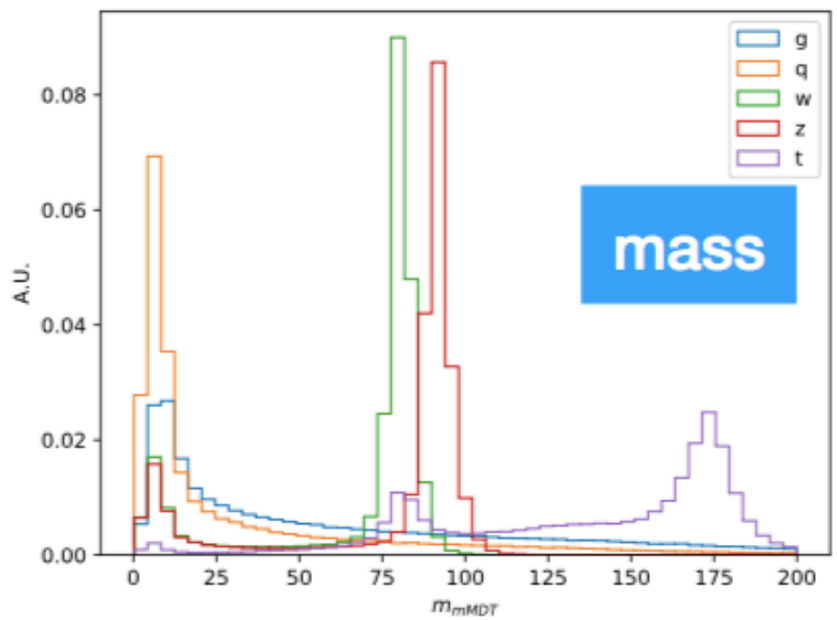


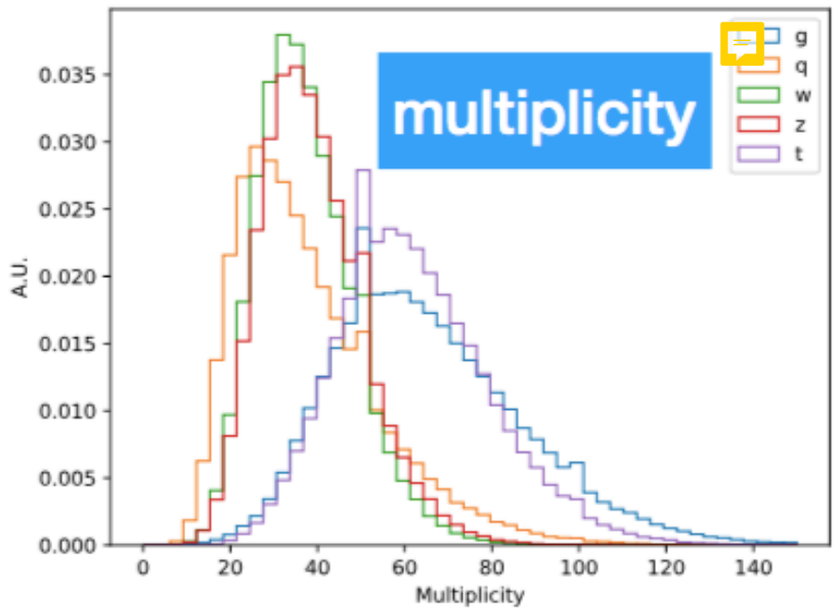

| **t→bW→bqq** | **Z→qq** | **W→qq** | **q/g background** |
|---|---|---|---|
| 3-prong jet | 2-prong jet | 2-prong jet | no substructure and/or mass ~ 0 |

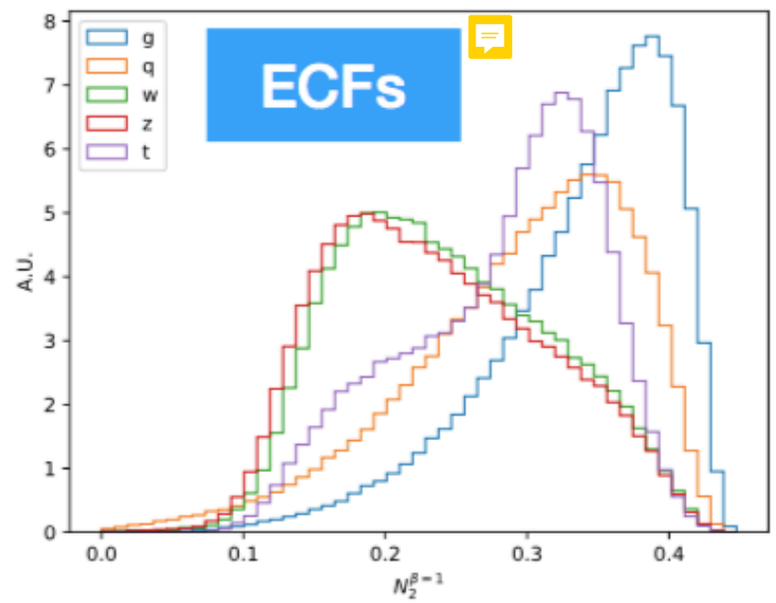
Reconstructed as one massive jet with substructure

# High-Level Features



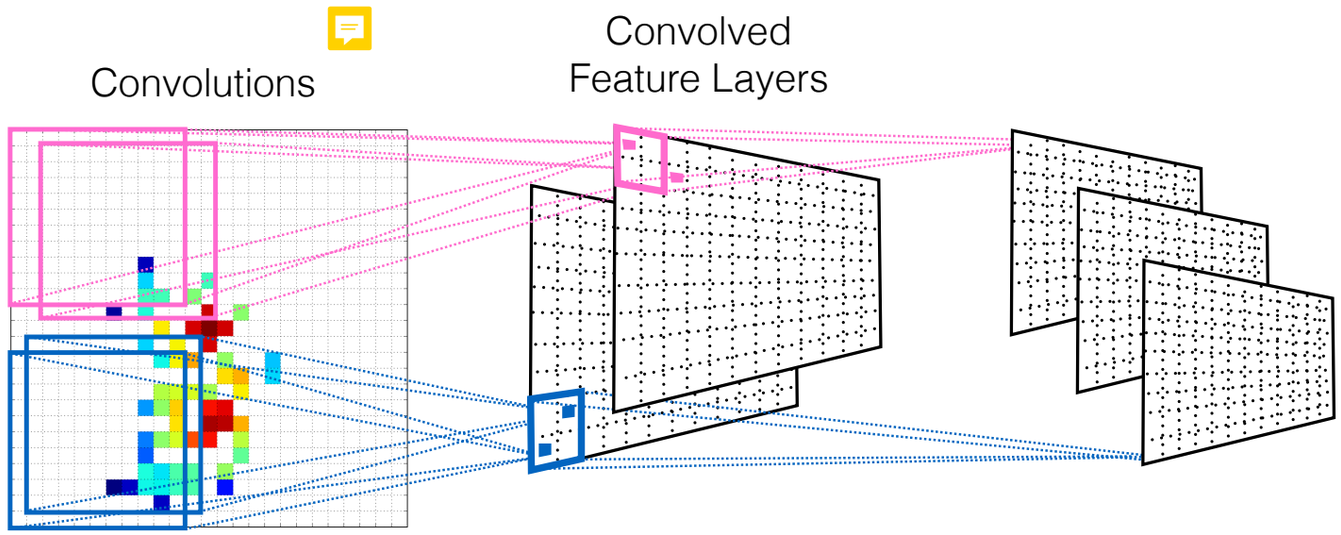

**Input variables: several observables known to have high discrimination power from offline data analyses and published studies [*]**

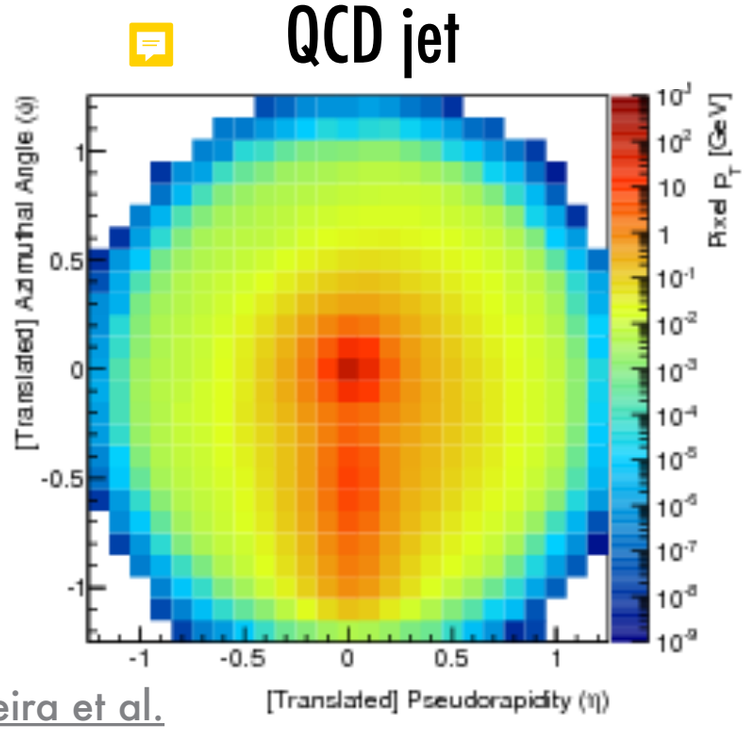
$m_{\mathrm{mMDT}}$
$N_2^{\beta=1,2}$
$M_2^{\beta=1,2}$
$C_1^{\beta=0,1,2}$
$C_2^{\beta=1,2}$
$D_2^{\beta=1,2}$
$D_2^{(\alpha,\beta)=(1,1),(1,2)}$
$\sum z \log z$
Multiplicity

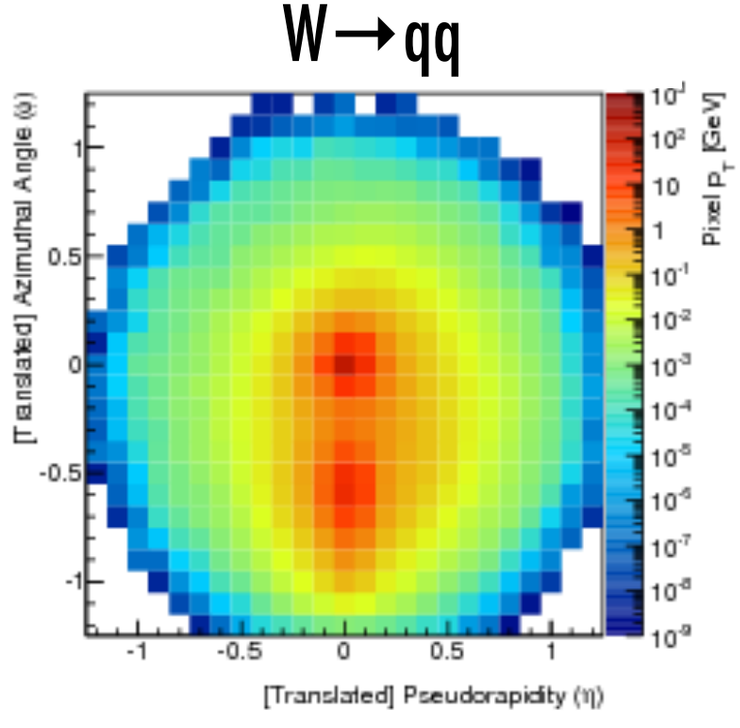
[*] D. Guest at al. PhysRevD.94.112002, G. Kasieczka et al. JHEP05(2017)006, J. M. Butterworth et al. PhysRevLett.100.242001, etc..

# Jets as images

- Pixelate the surface covered by the jet and fill each pixel summing up the momenta of the jet constituents pointing to it
- Pixel size determined by calorimeter resolution
- Process the images through **state-of-the-art industry computer vision methods**
- Innovation: ***convolutional NN***
  - apply same matrix of weights to consecutive subsets of the image
  - reduced # parameters
  - able to learn features like edges, corners, etc… from raw data



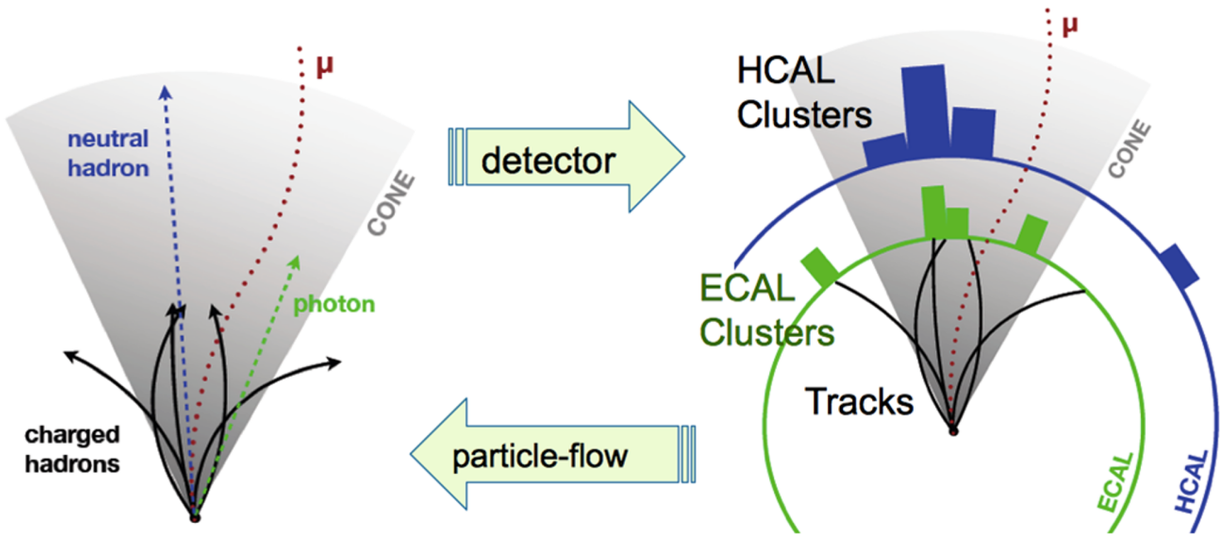



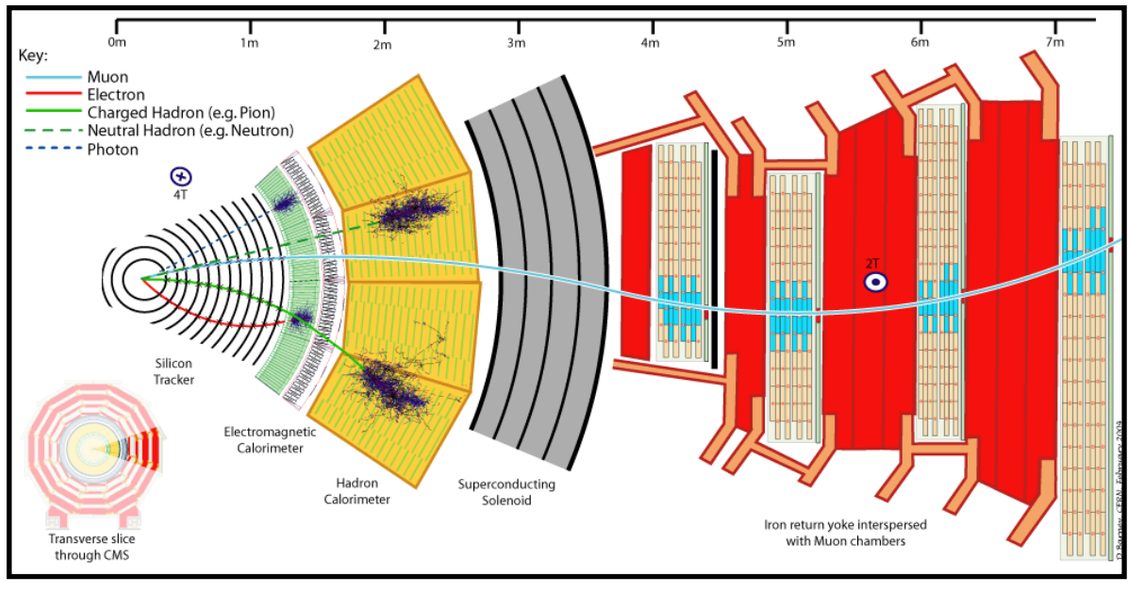




L. de Oliveira et al. JHEP 07 (2016) 069

# Is it a good representation?

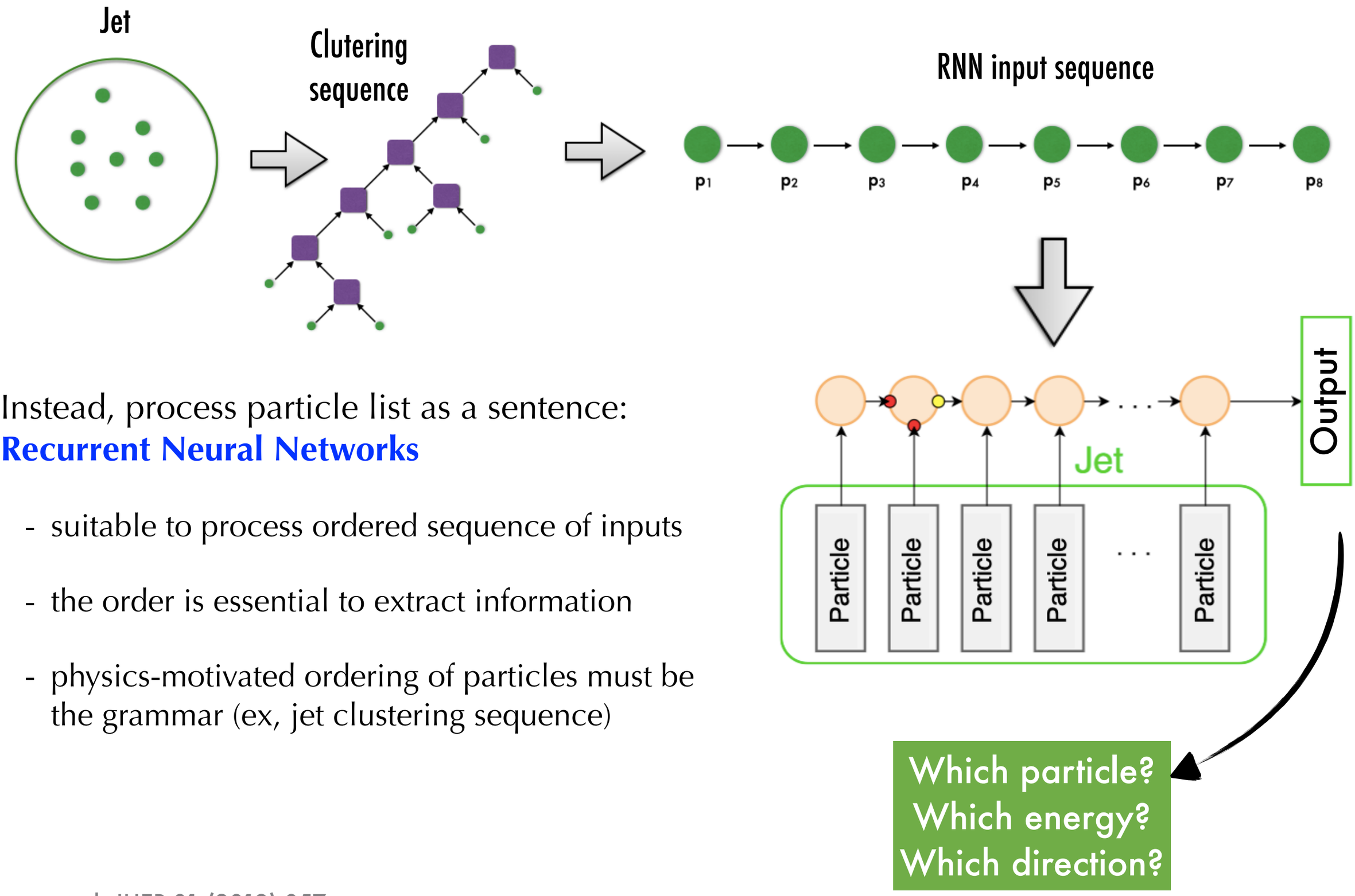
- Using images works good IF individual particles forming the jet were measured in the calorimeter only
- This is not always the case: the particle flow reconstruction makes **extensive use of track information** that cannot be captured by the image → **loss in information and resolution**




| Detector | $p_T$-resolution | η/Φ-segmentation |
|---|---|---|
| Tracker | 0.6% (0.2 GeV) – 5% (500 GeV) | 0.002 x 0.003 (first pixel layer) |
| ECAL | 1% (20 GeV) – 0.4% (500 GeV) | 0.017 x 0.017 (barrel) |
| HCAL | 30% (30 GeV) – 5% (500 GeV) | 0.087 x 0.087 (barrel) |

# Jets as sentences




- Instead, process particle list as a sentence:
  **Recurrent Neural Networks**
  - suitable to process ordered sequence of inputs
  - the order is essential to extract information
  - physics-motivated ordering of particles must be the grammar (ex, jet clustering sequence)

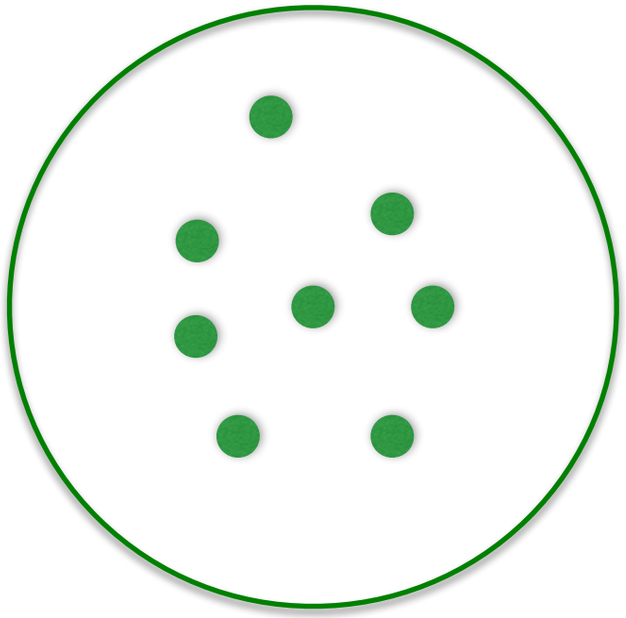
G. Louppe et al: JHEP 01 (2019) 057

# Jets as graphs

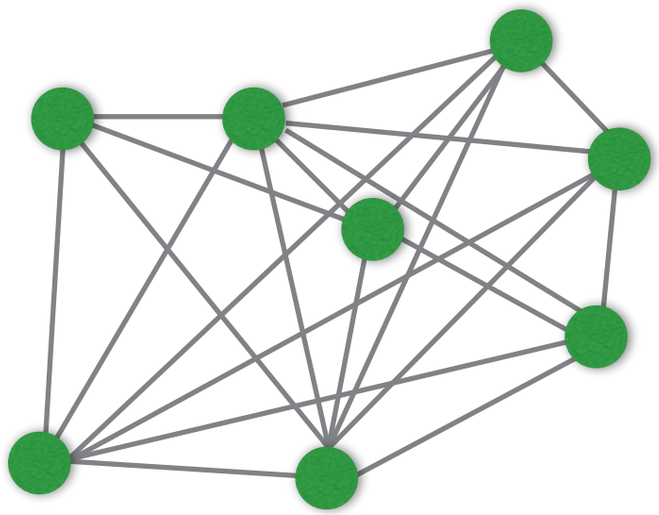
- What if the assumed clustering sequence is not the best ordering?
  → impose **less constraints** and let the NN learn the jet structure

- Represent particles in the jet by **a set of points in space**, where the points themselves are also **sparse and unordered** → graphs as a more natural representation (rather than forcing data into a CNN or RNN)

A Jet



A Graph

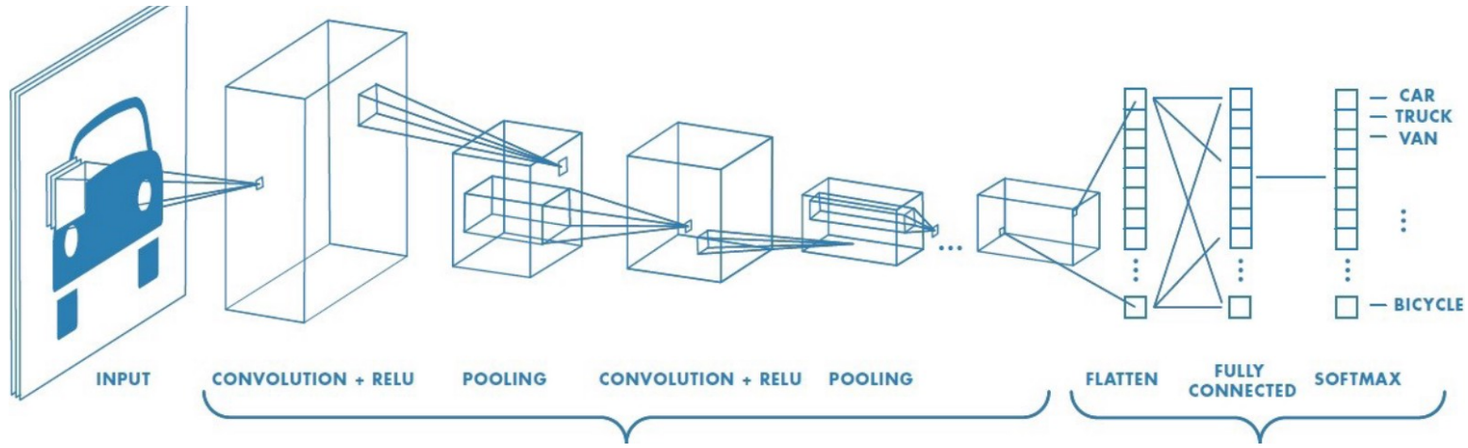



I. Henrion et al. NeurIPS ‘17
E. Moreno et al.: arxiv.1909.12285
E. Moreno et al.: Eur. Phys. J. C 80, 58 (2020)
H. Qu et al.: Phys. Rev. D 101, 056019 (2020)

# One problem, many solutions

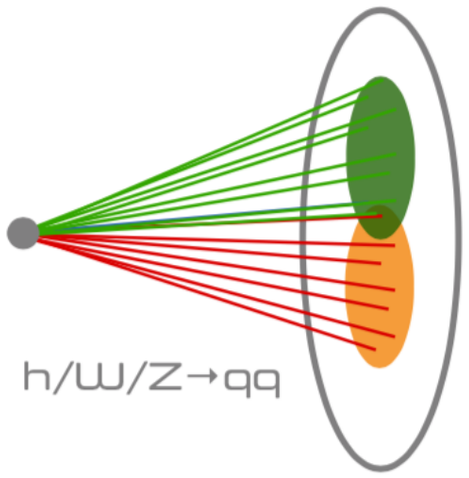
- **Jets as images:**

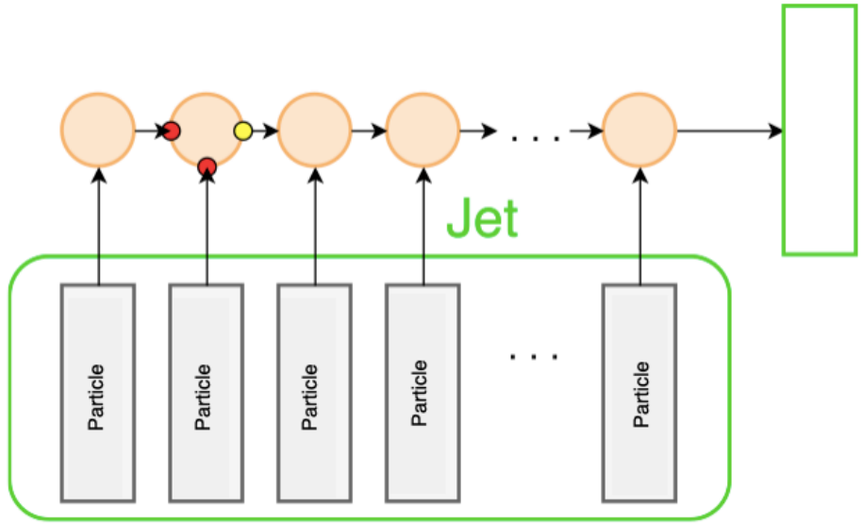
  - can use computer vision techniques




- **Jets as sentences:**

  - can use natural language processing techniques




- **Jets as graphs:**

  - can use graph networks as in social media analyses