Introduction to Anomaly Detection at the LHC

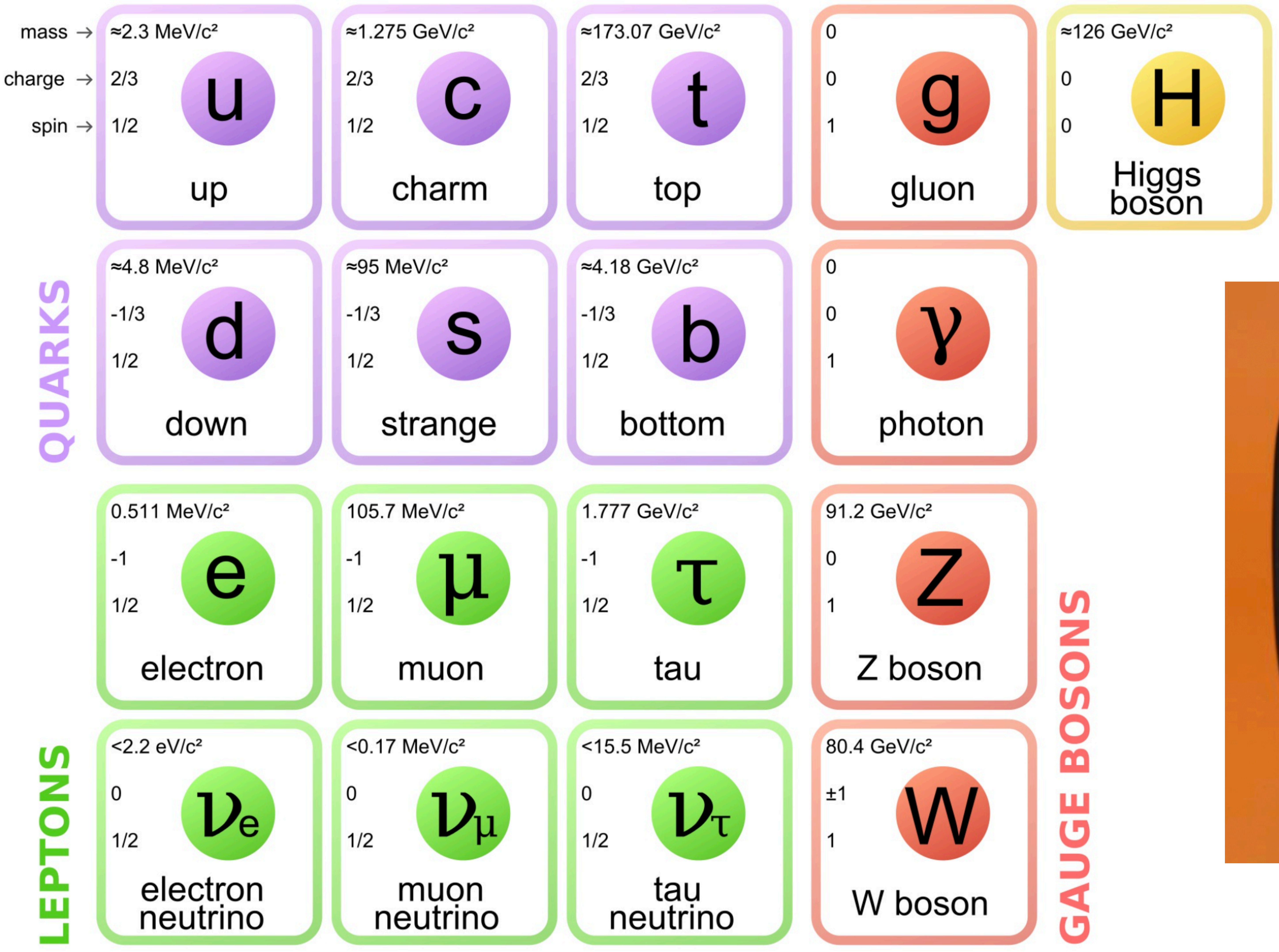
Dylan Rankin
November 23rd, 2024

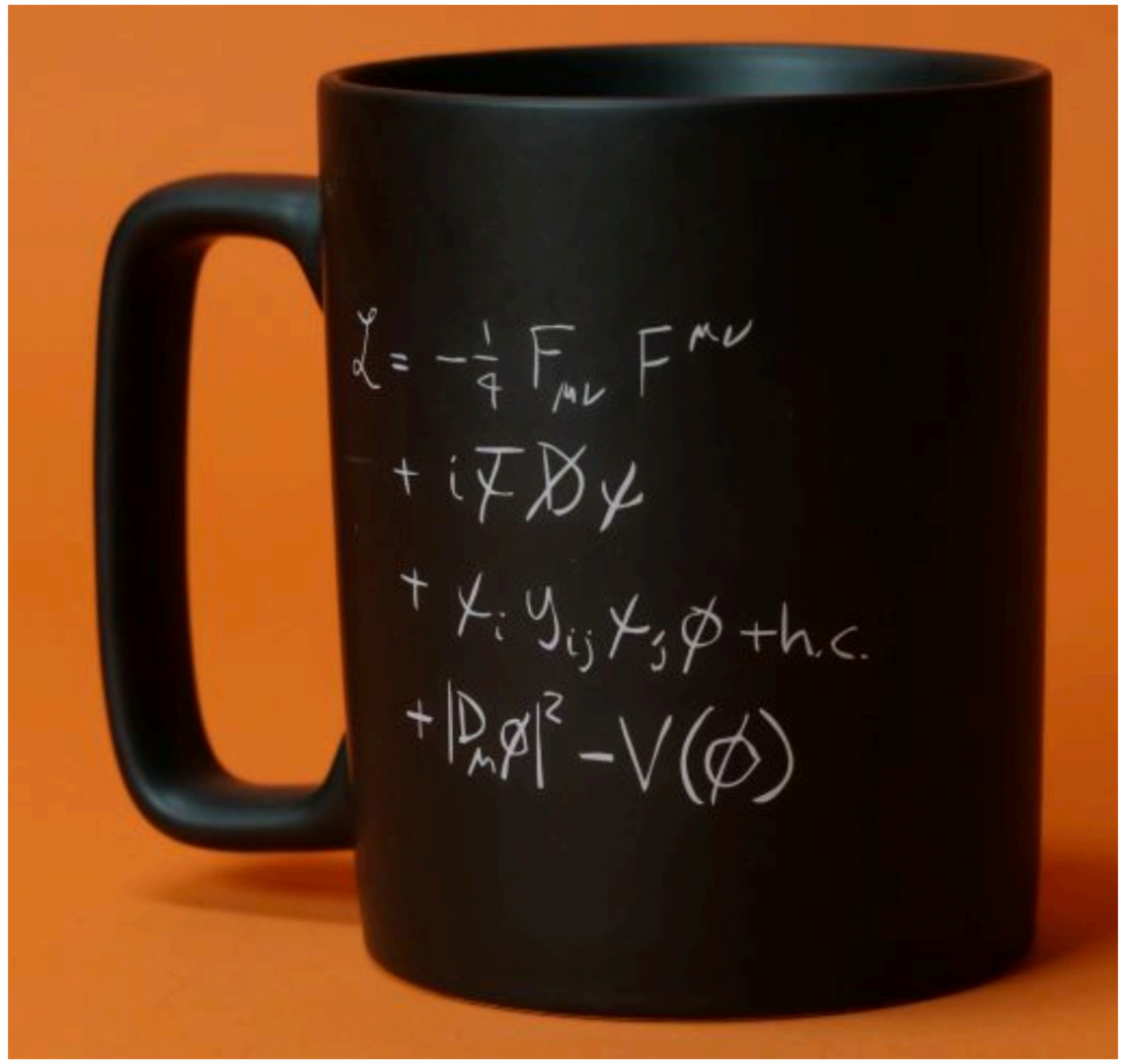
Large Hadron Collider (LHC) at CERN

Large Hadron Collider (LHC) at CERN

Large Hadron Collider (LHC) at CERN

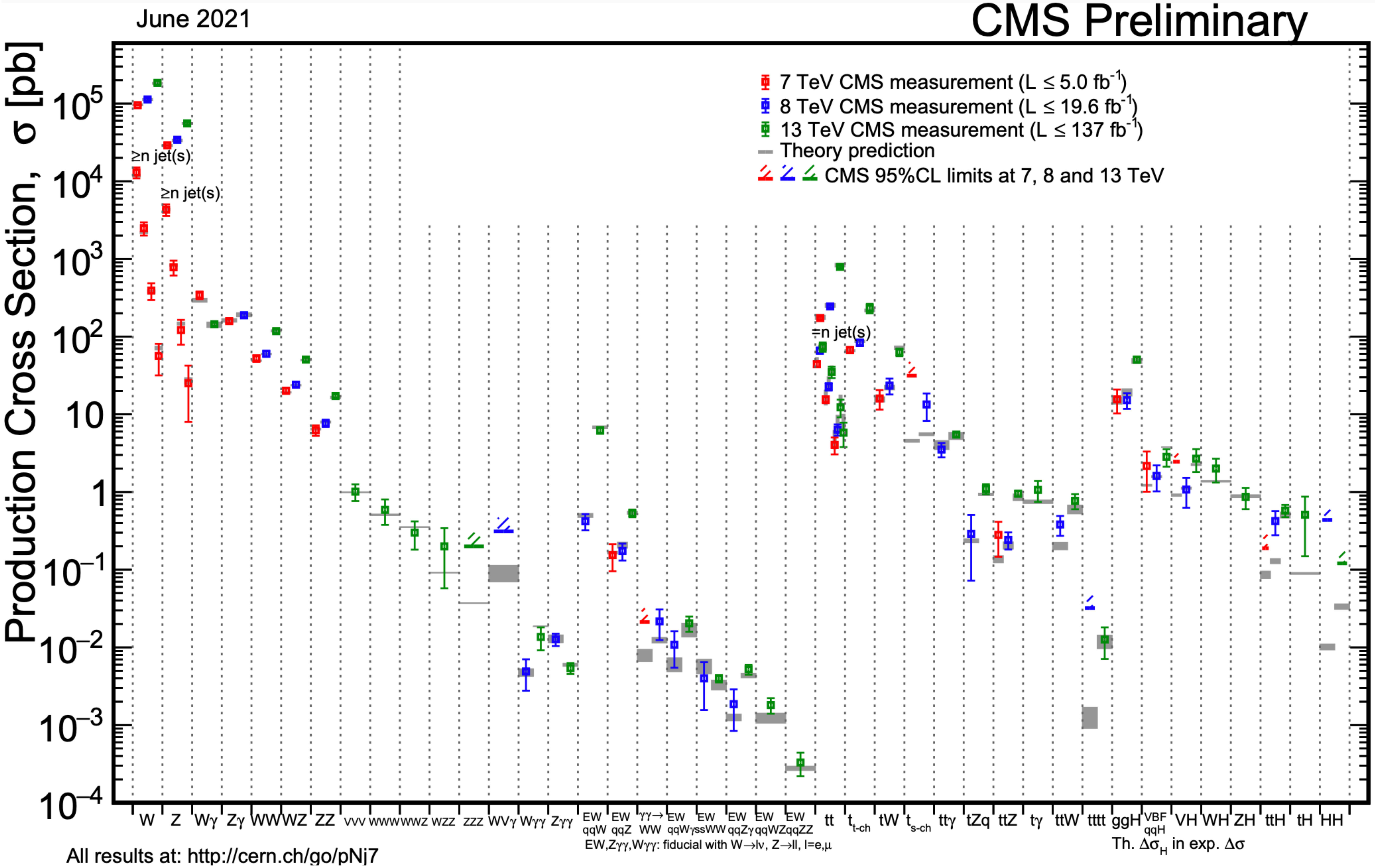
The Standard Model



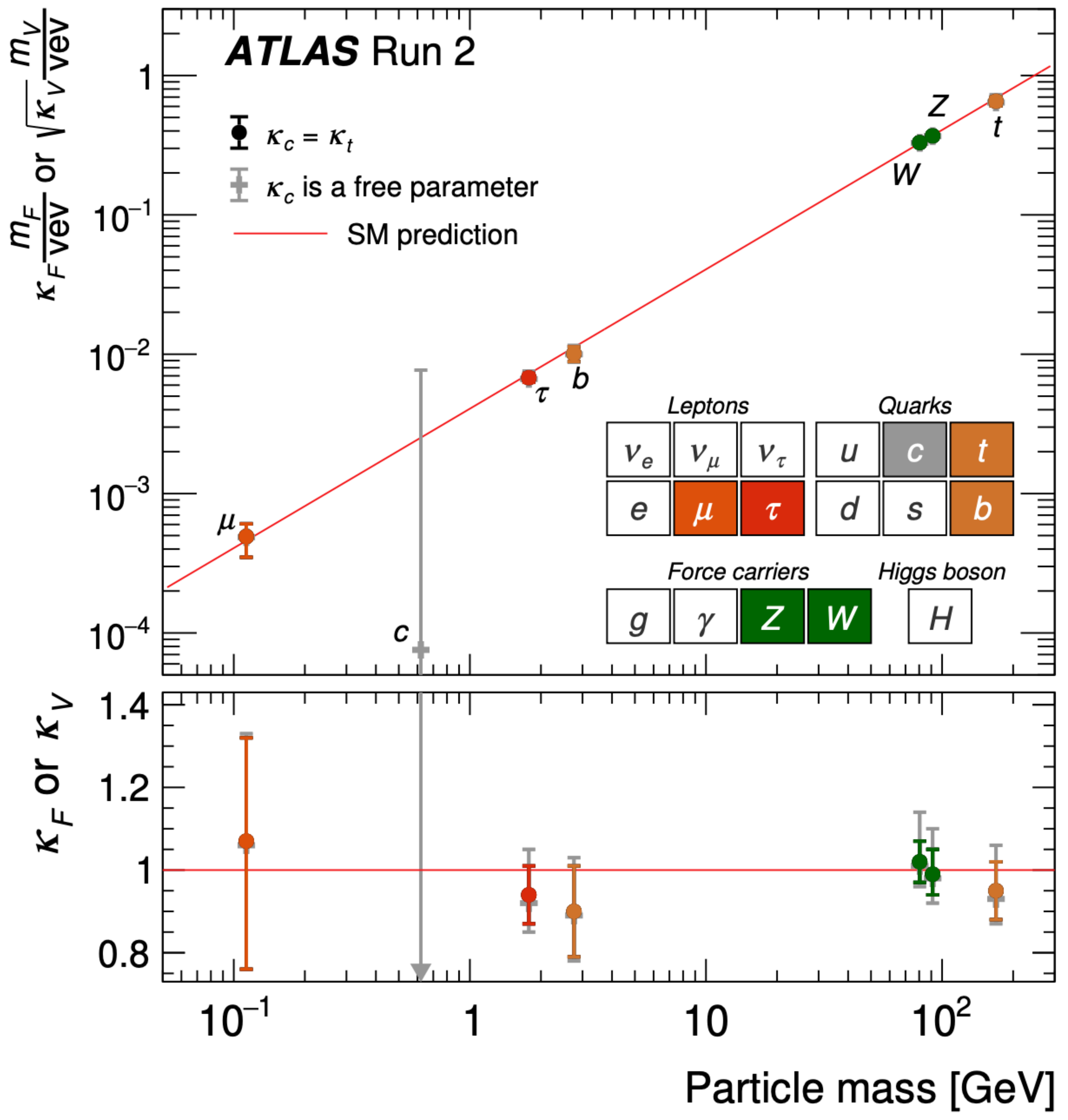


Standard Model Measurements

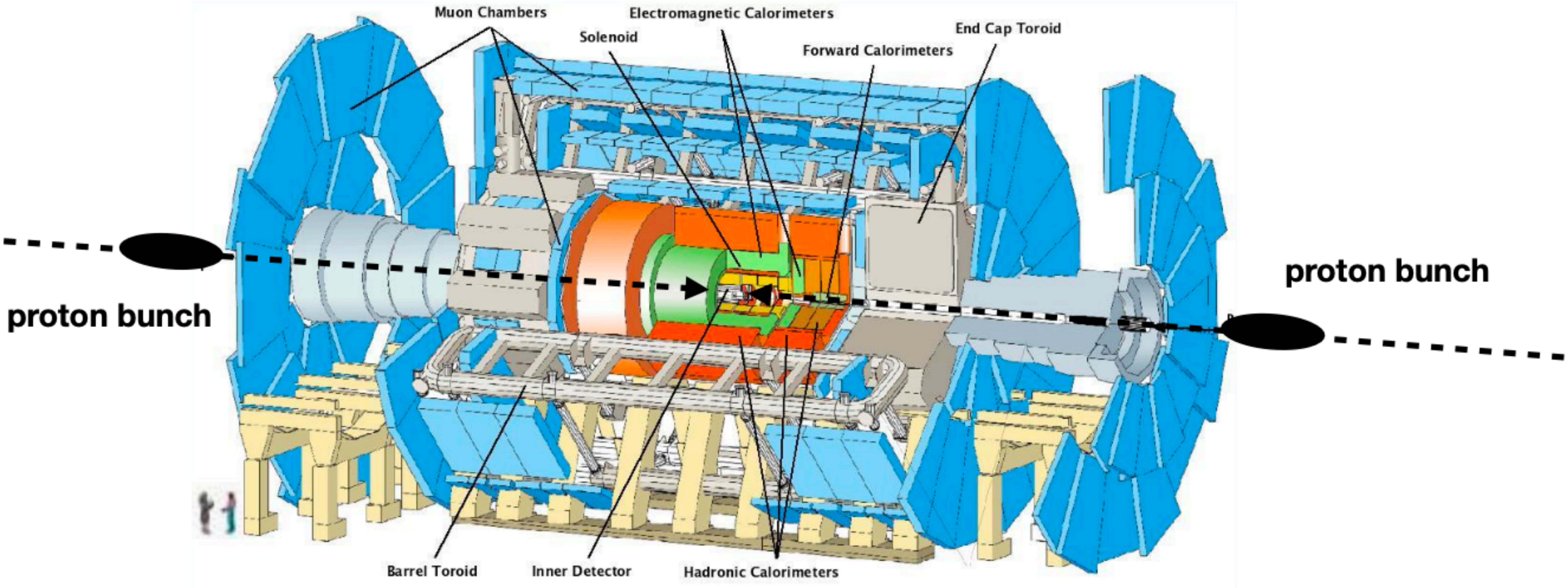
Standard Model cross sections



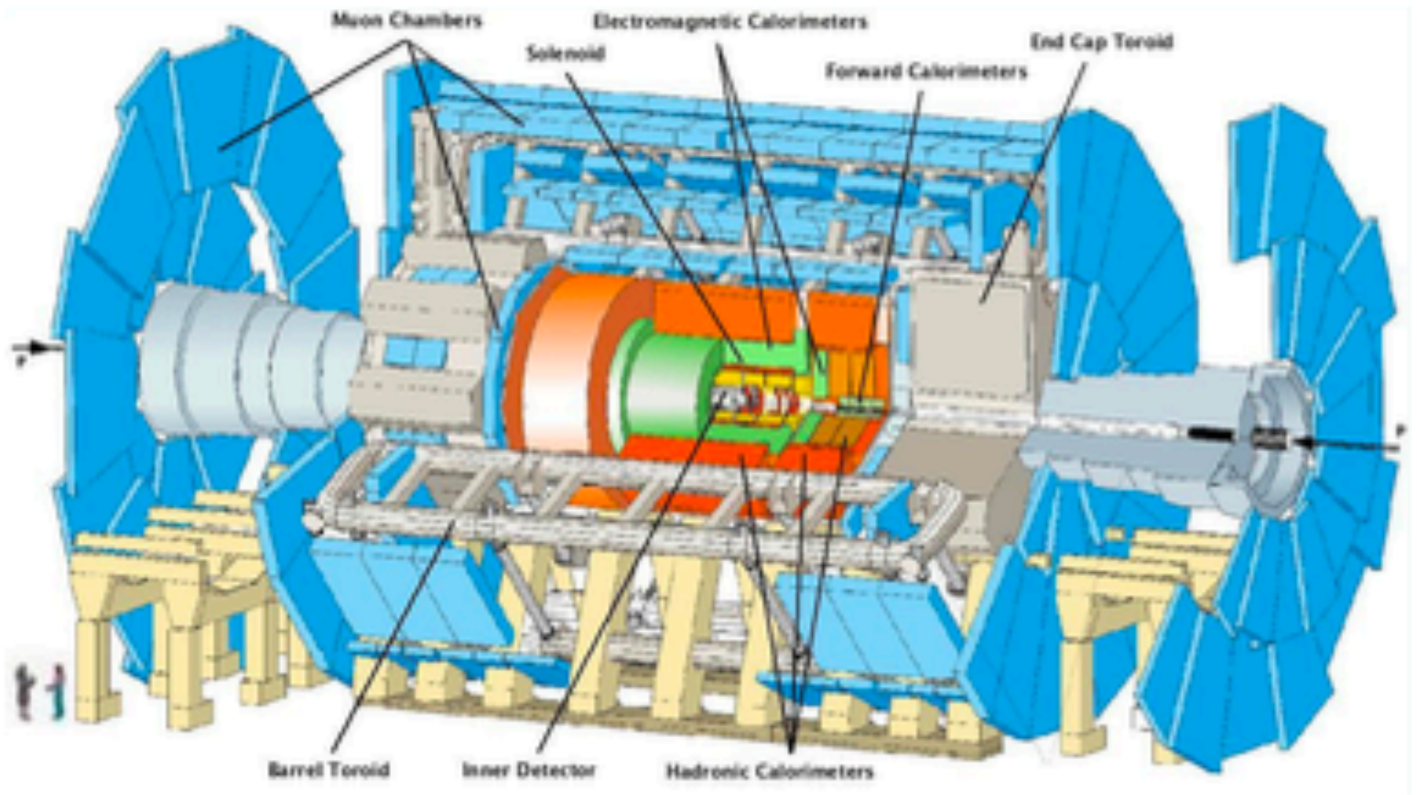
Higgs boson couplings

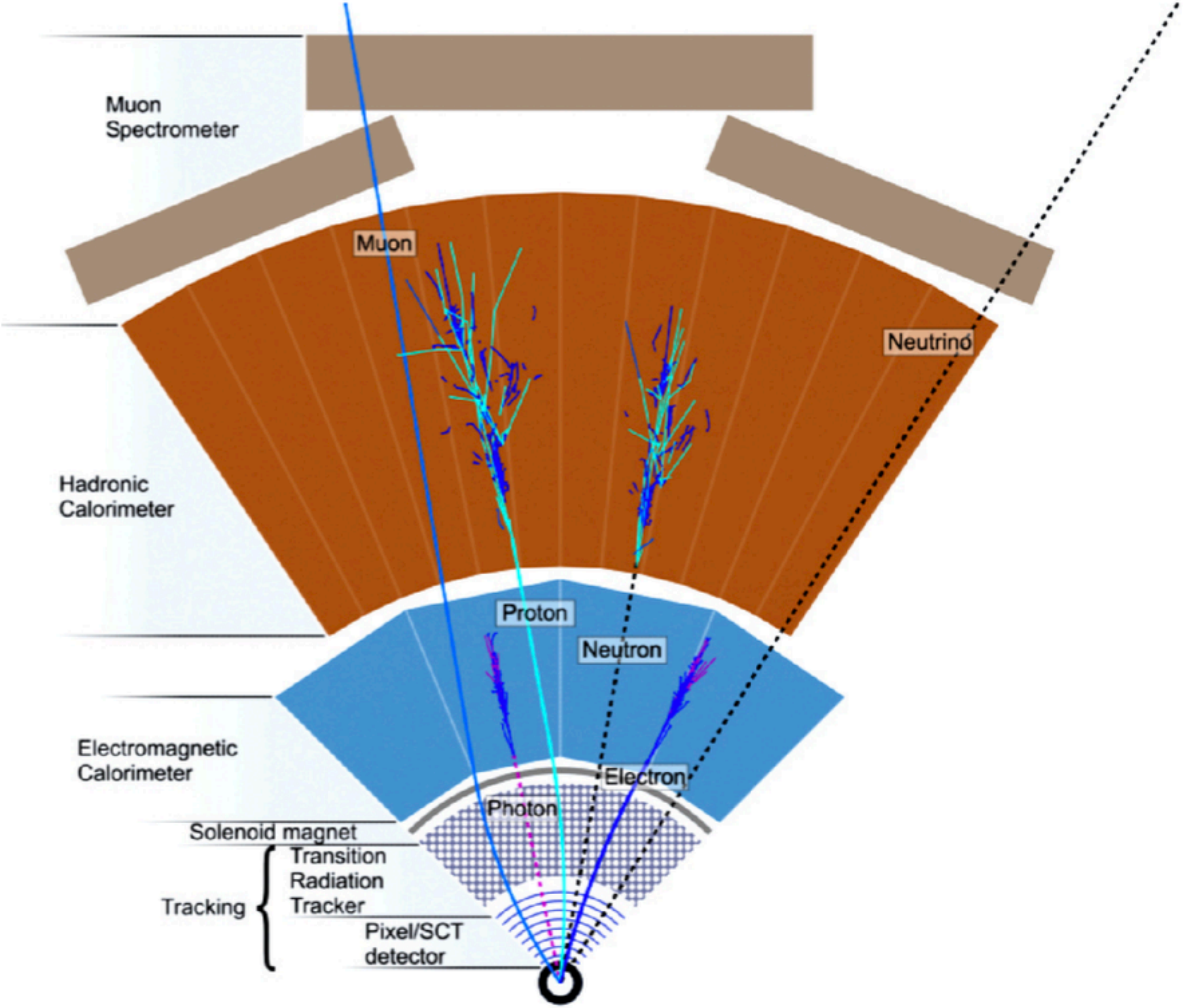


A Toroidal LHC ApparatuS (ATLAS)



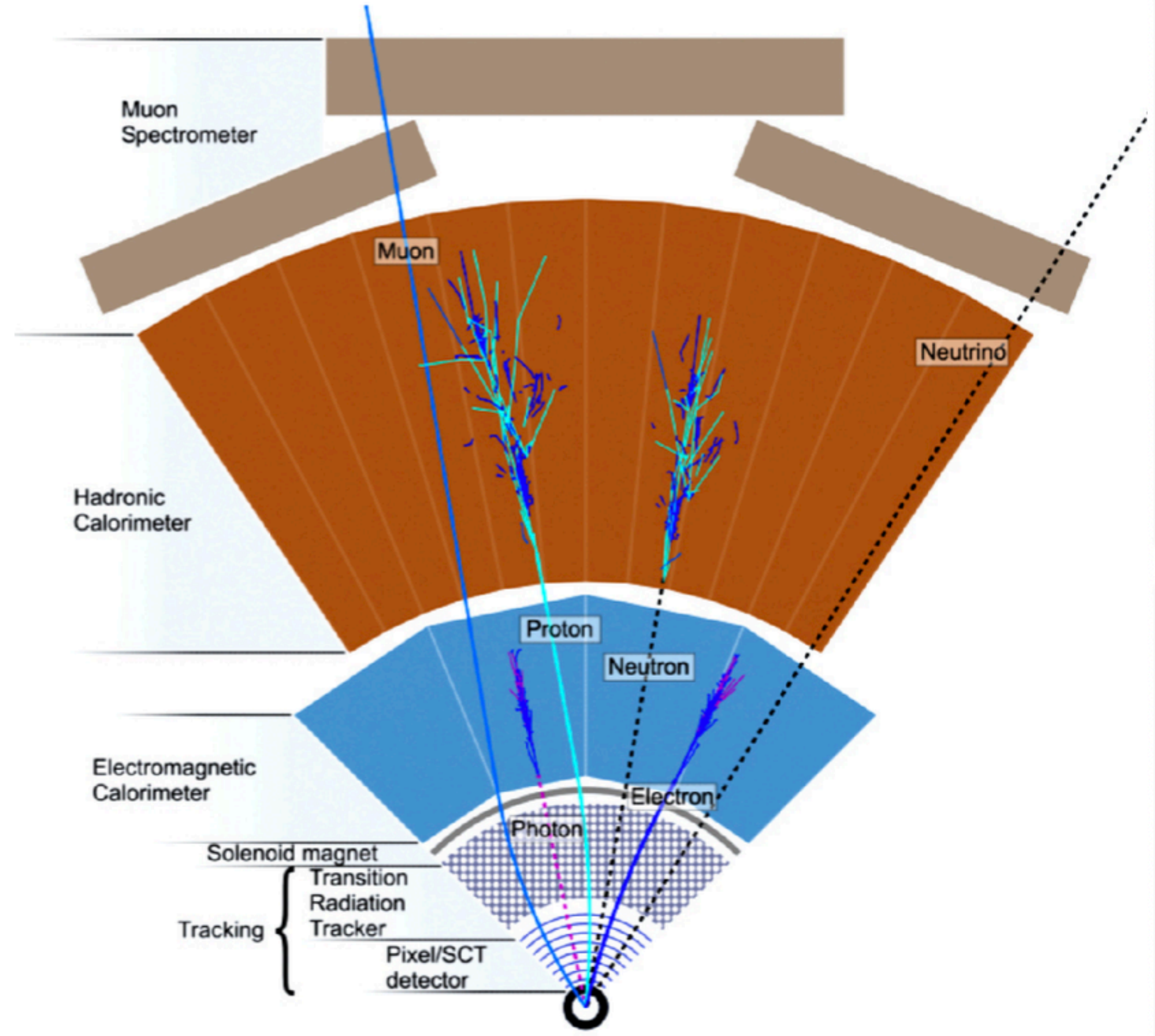
ATLAS Slice

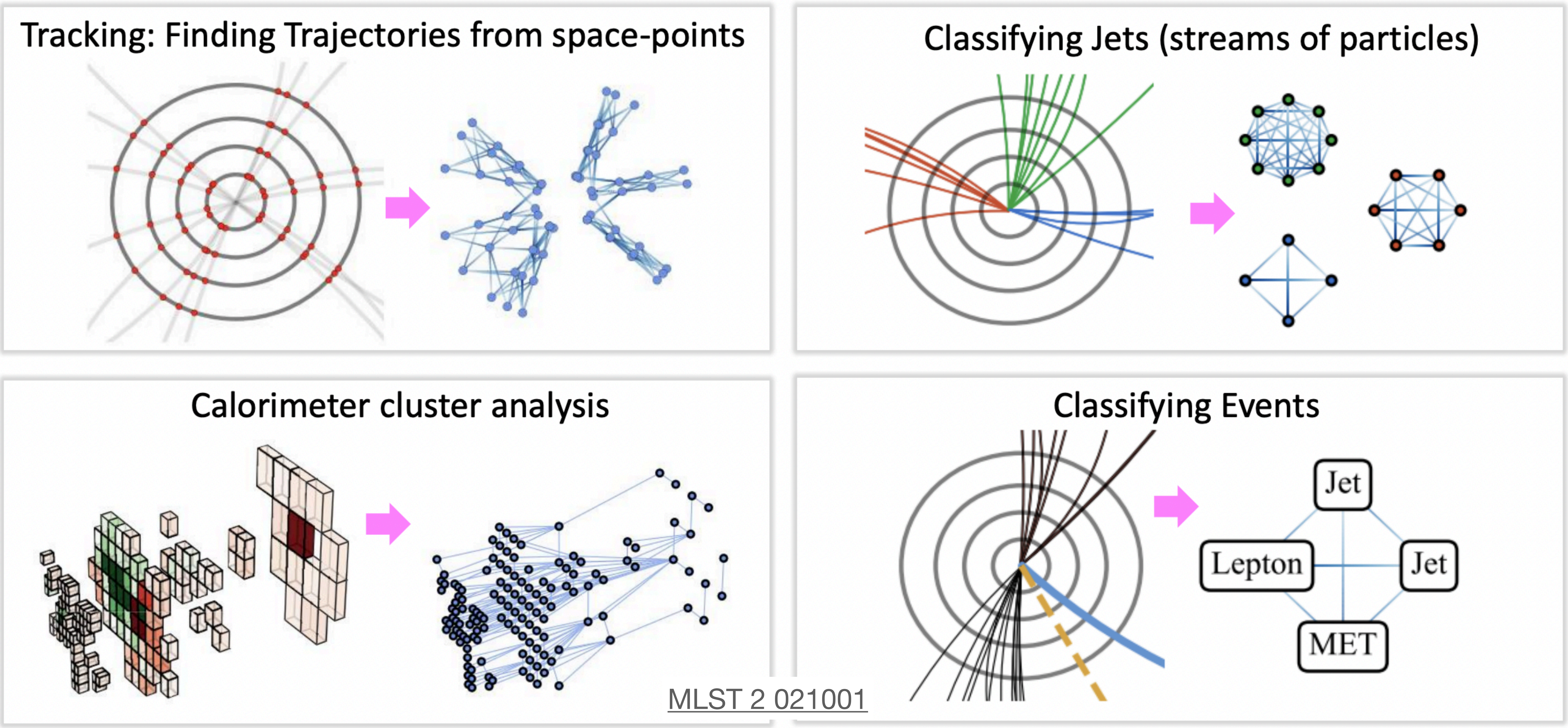

Muon Chambers
Electromagnetic Calorimeters
Solenoid
End Cap Toroid
Forward Calorimeters
Barrel Toroid
Inner Detector
Hadronic Calorimeters


Muon Spectrometer
Muon
Neutrino
Hadronic Calorimeter
Proton
Neutron
Electromagnetic Calorimeter
Electron
Photon
Solenoid magnet
Transition Radiation Tracker
Tracking
Pixel/SCT detector

ML in High Energy Physics (HEP)

- ML is becoming more and more popular, HEP/LHC no exception
- Better algorithms → improved performance

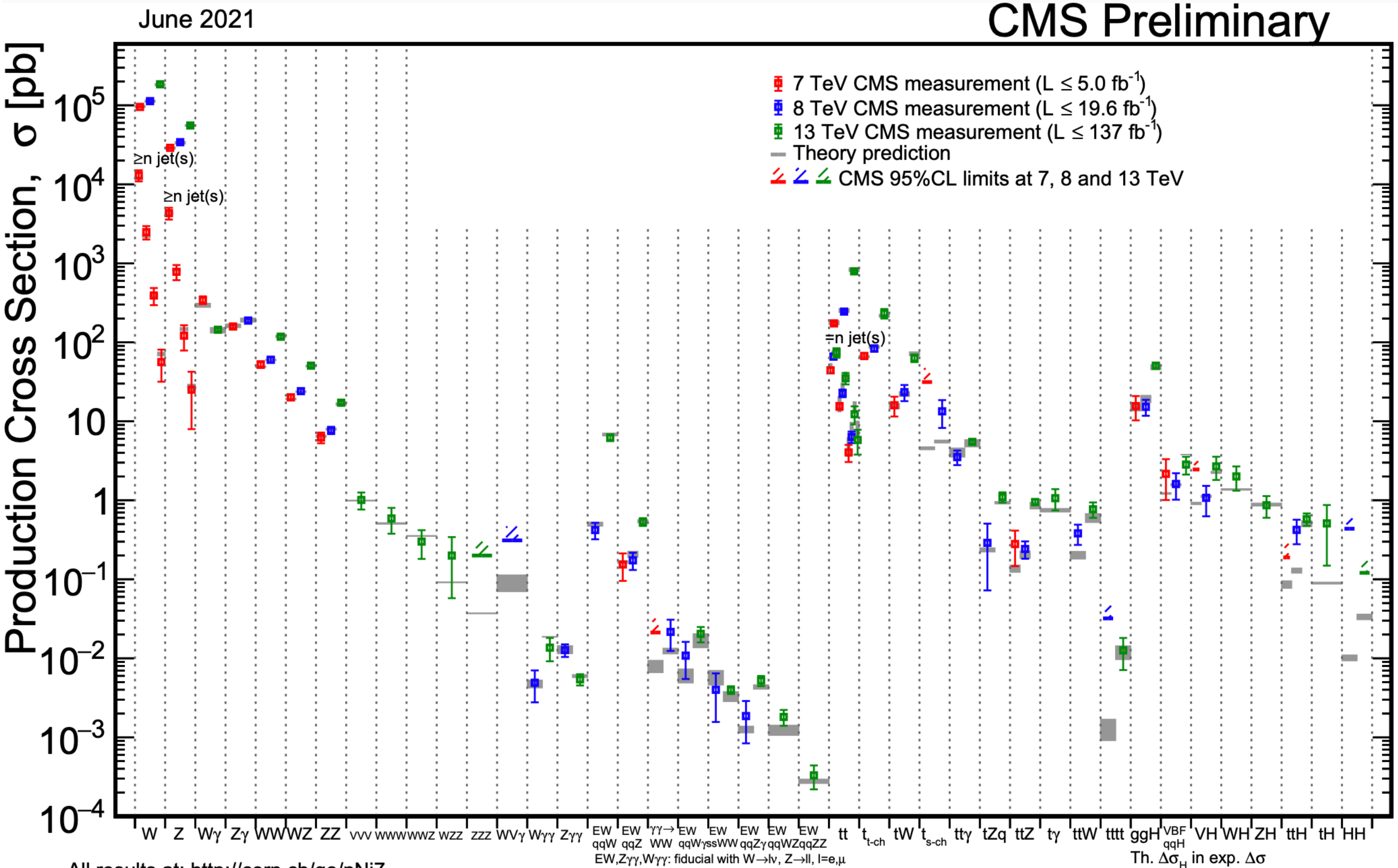




MLST 2 021001

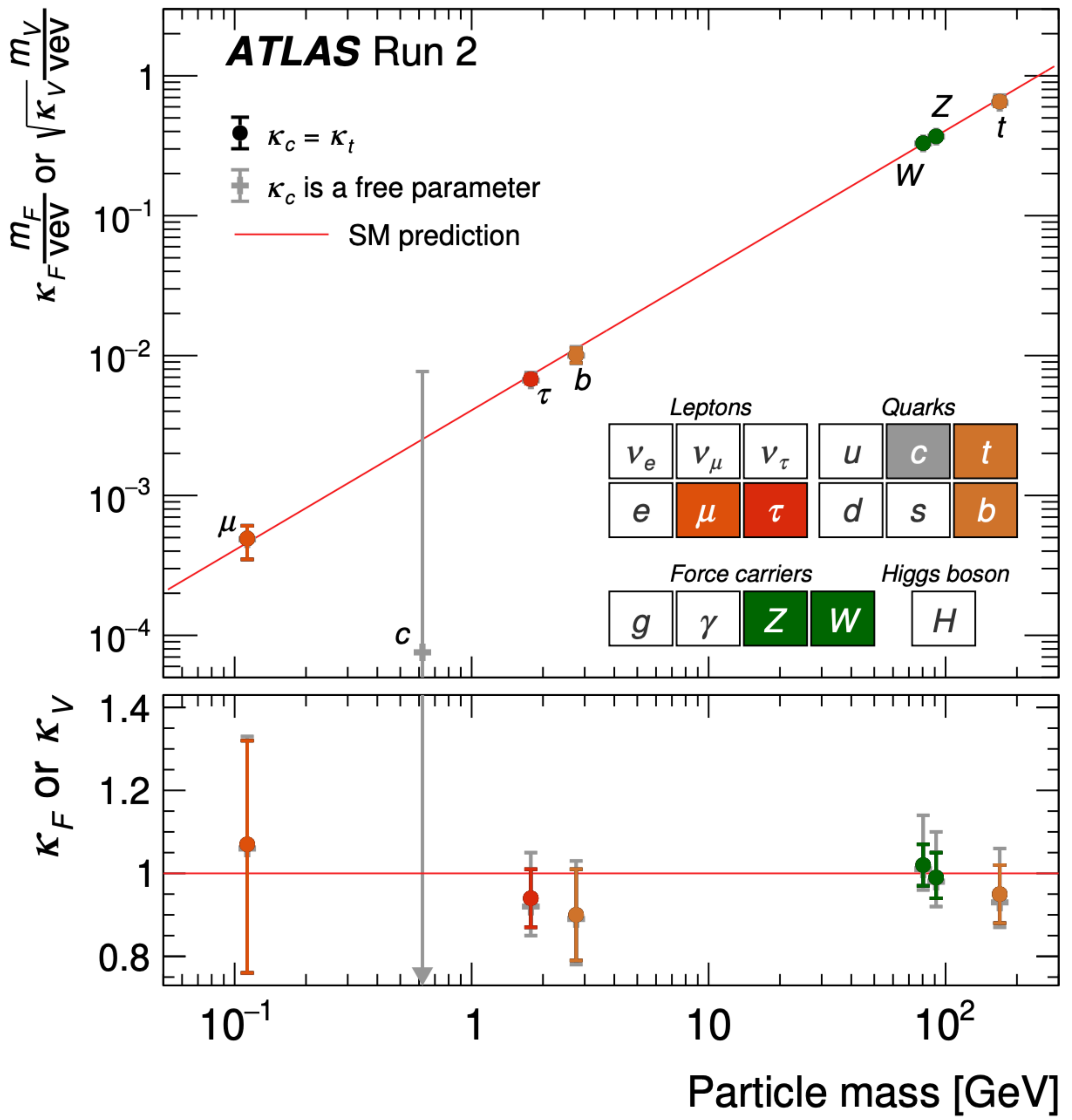
Standard Model Measurements

Standard Model cross sections

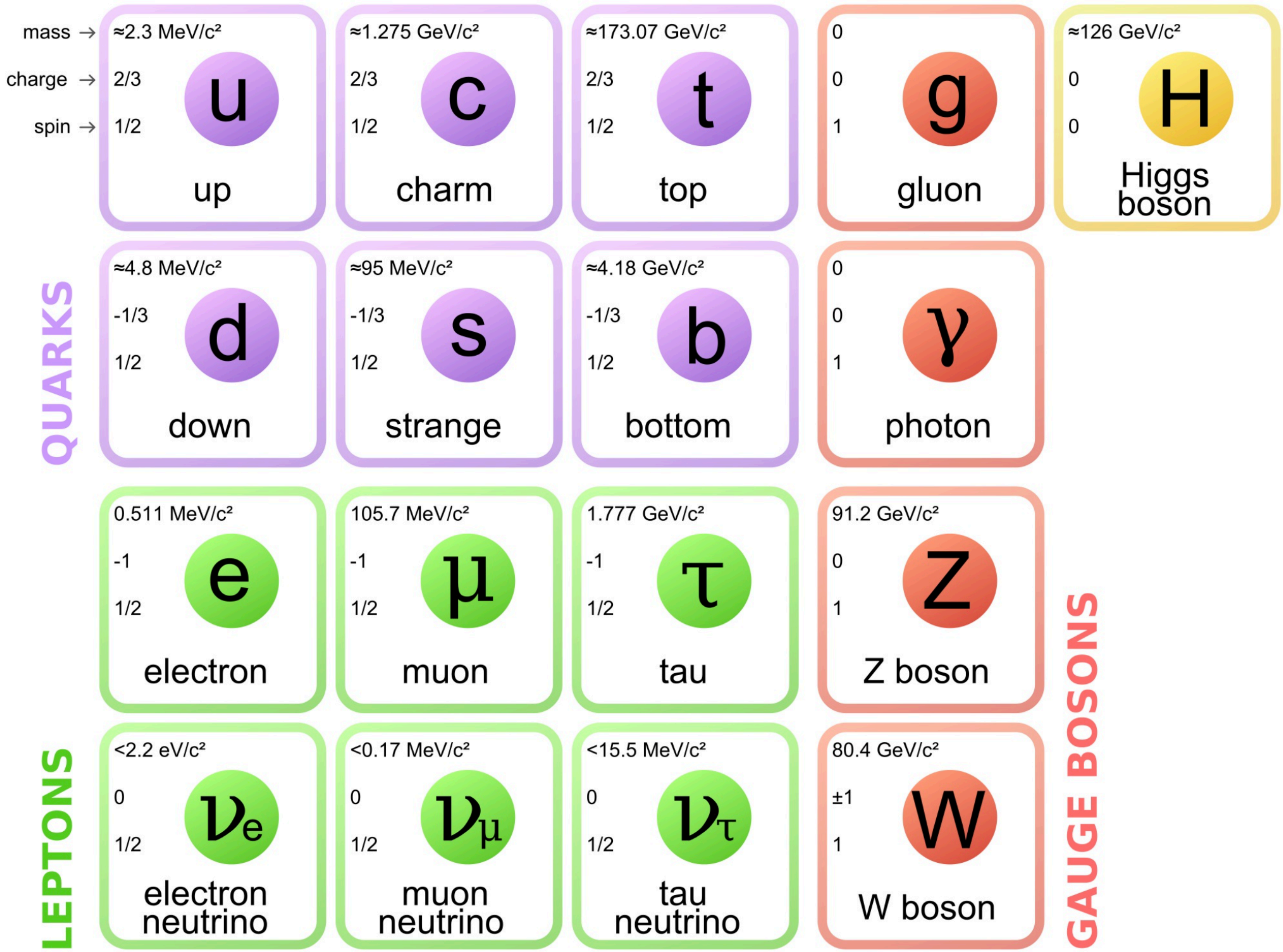


All results at: http://cern.ch/go/pNj7

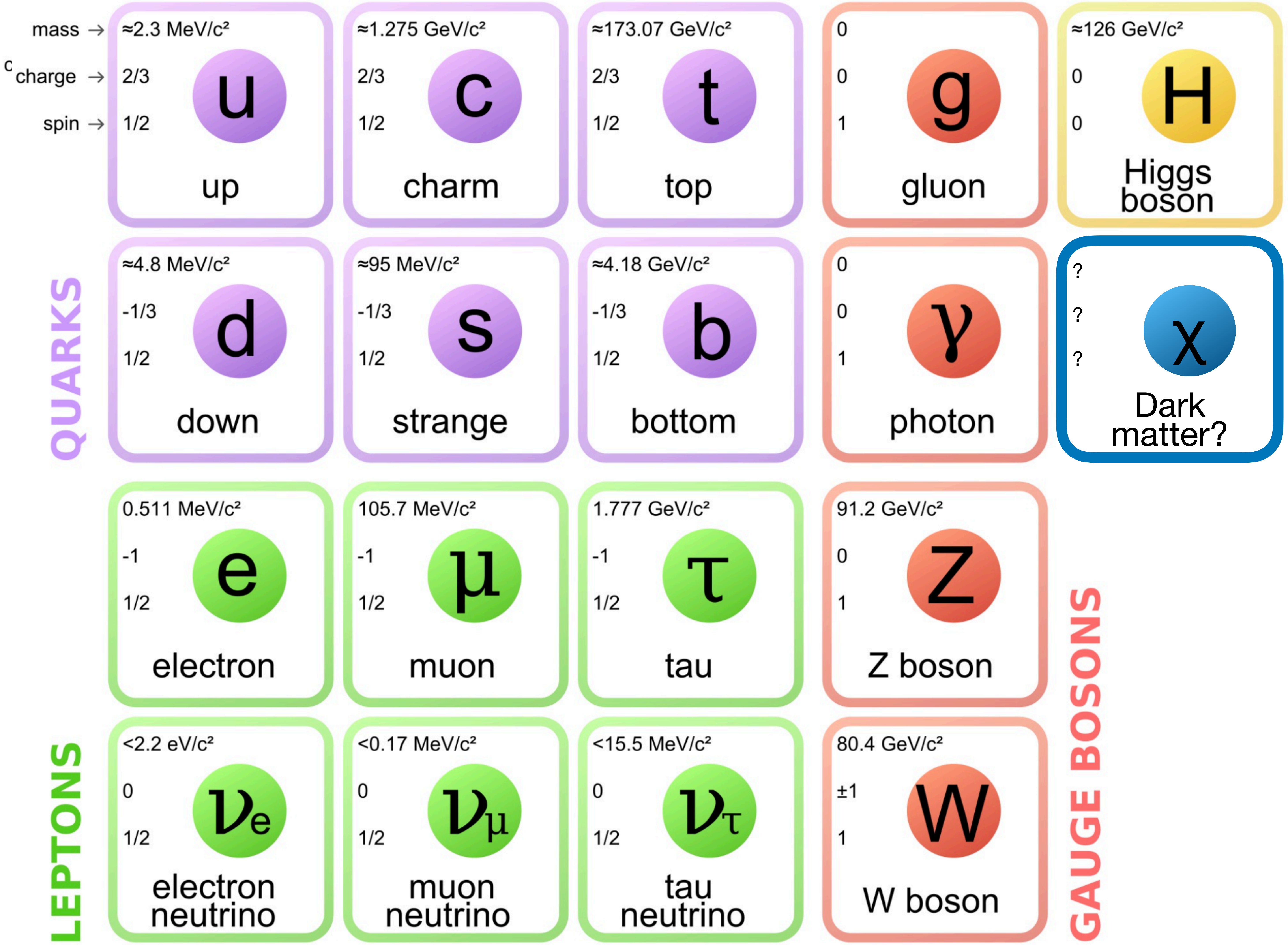
Higgs boson couplings



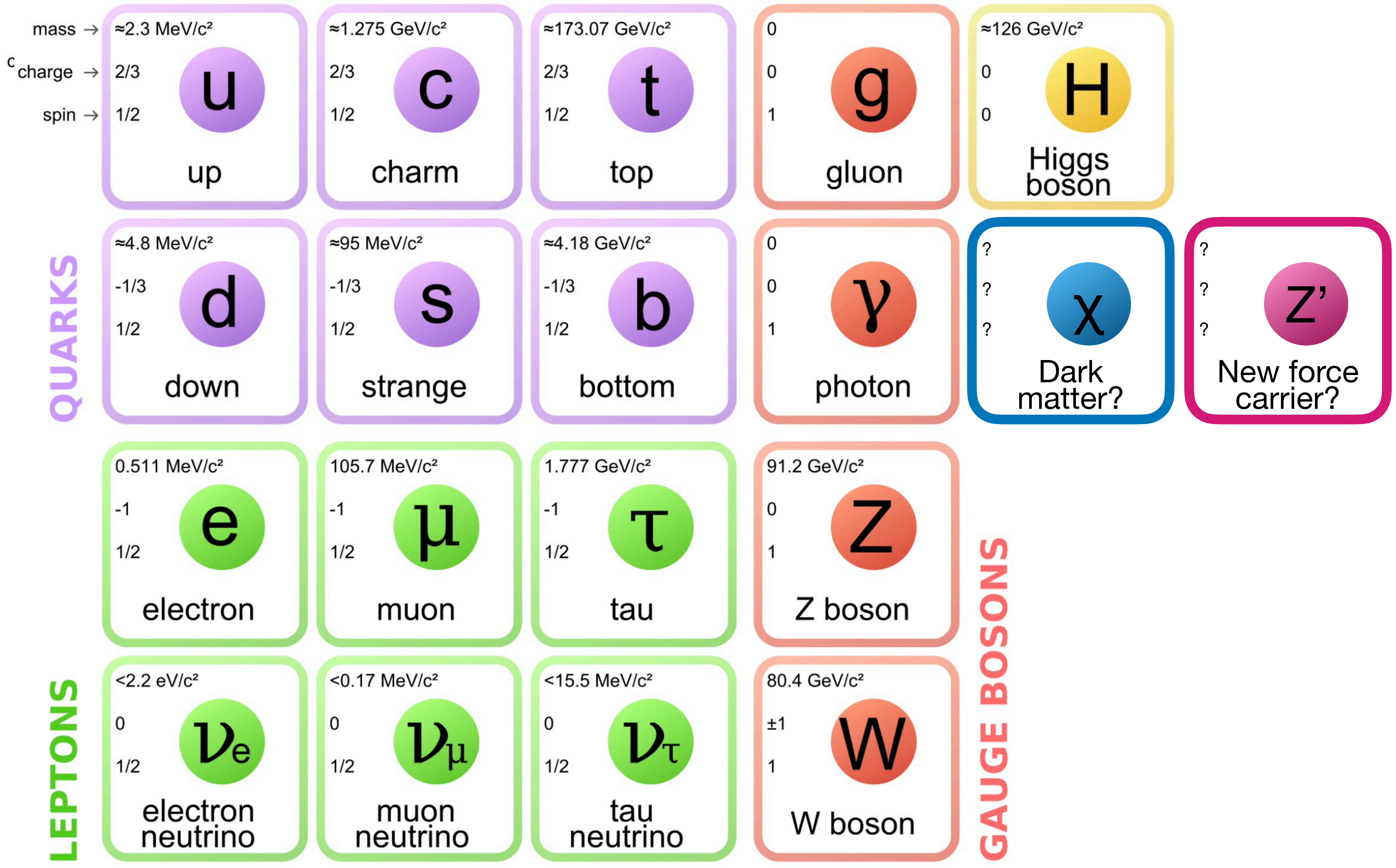
The Standard Model



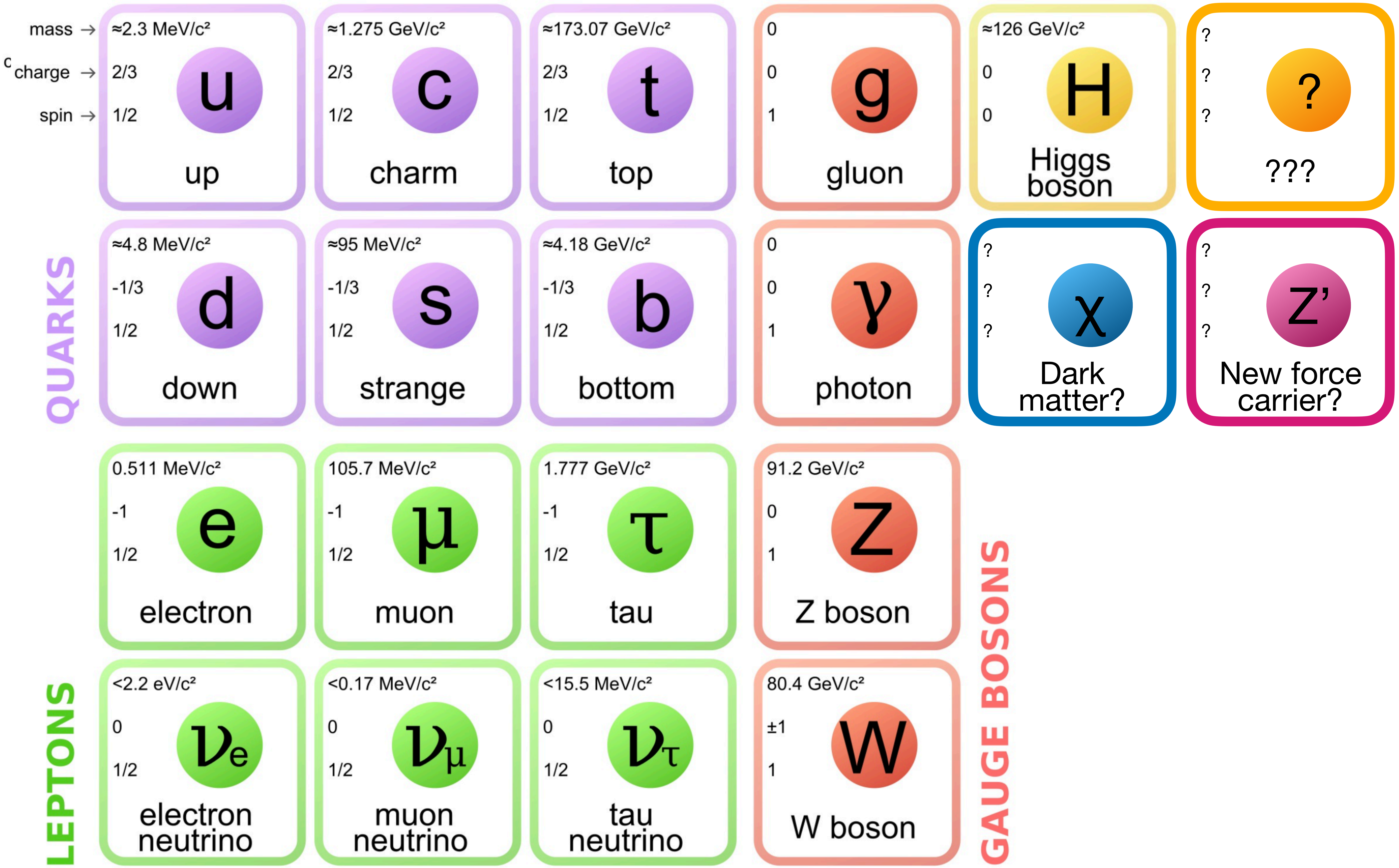
Beyond the Standard Model?



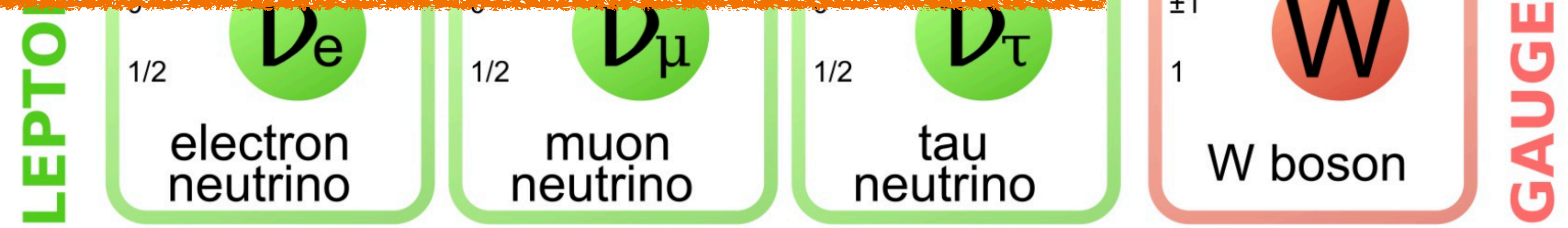
Beyond the Standard Model?

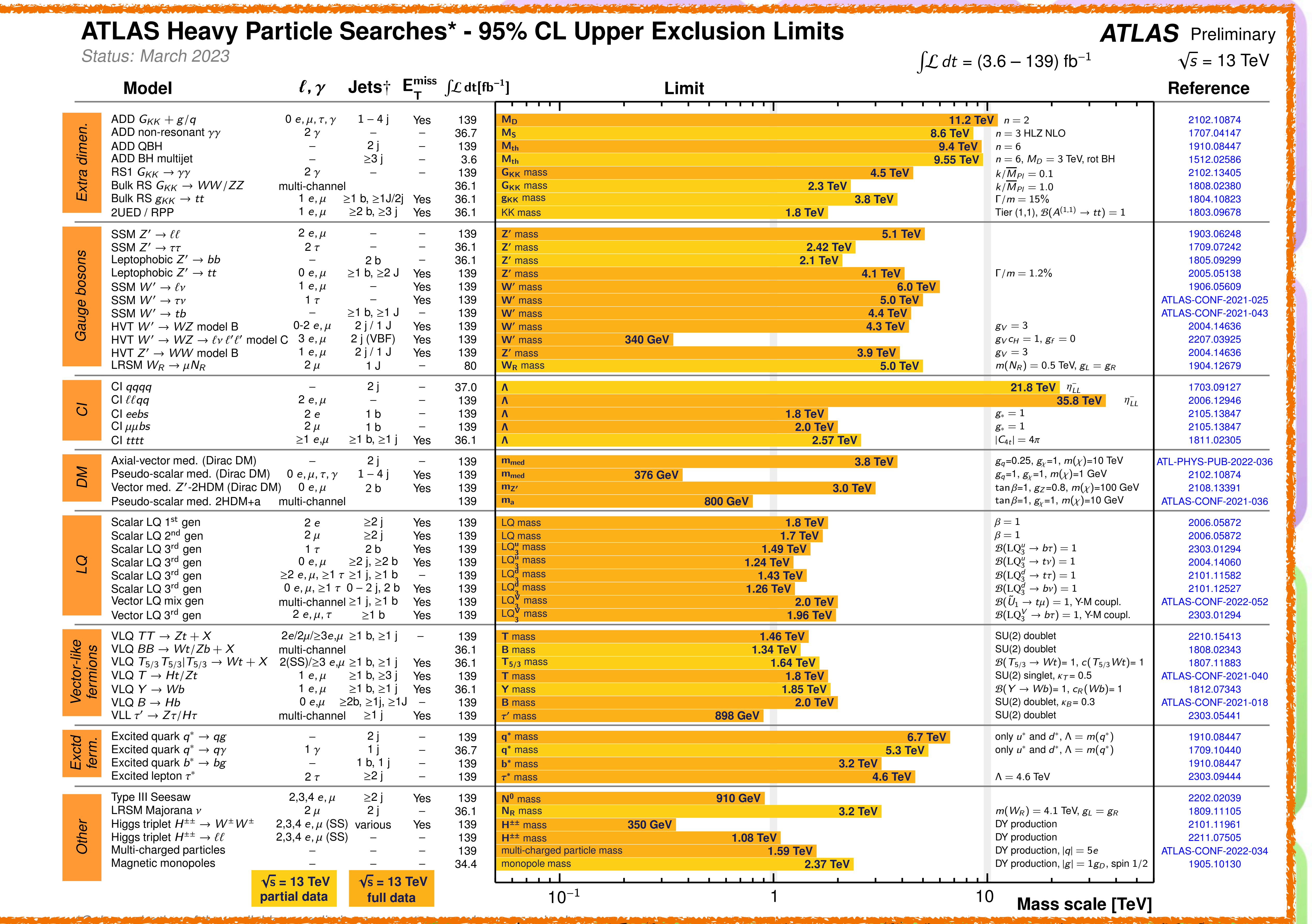


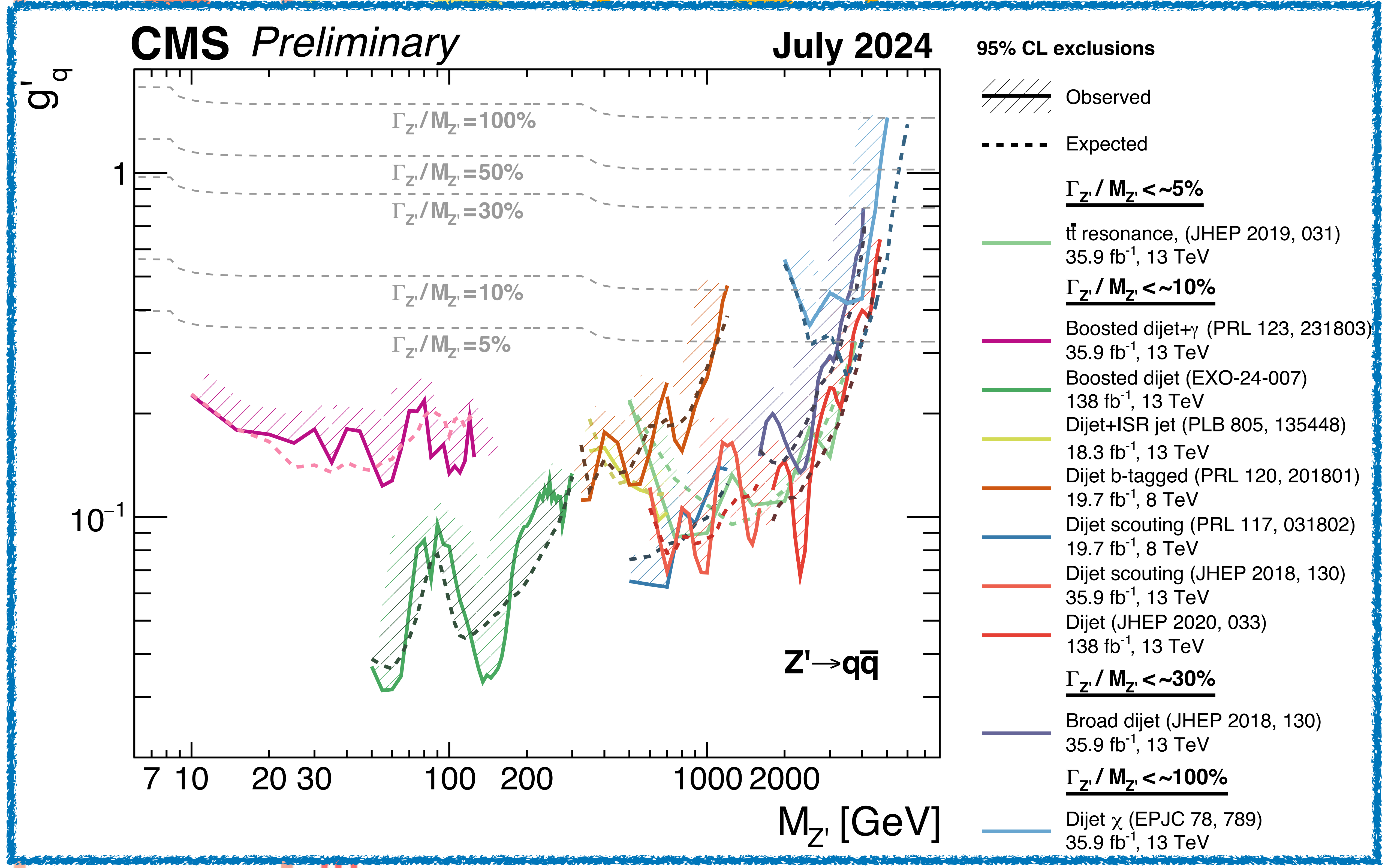
Beyond the Standard Model?



Beyond the Standard Model?

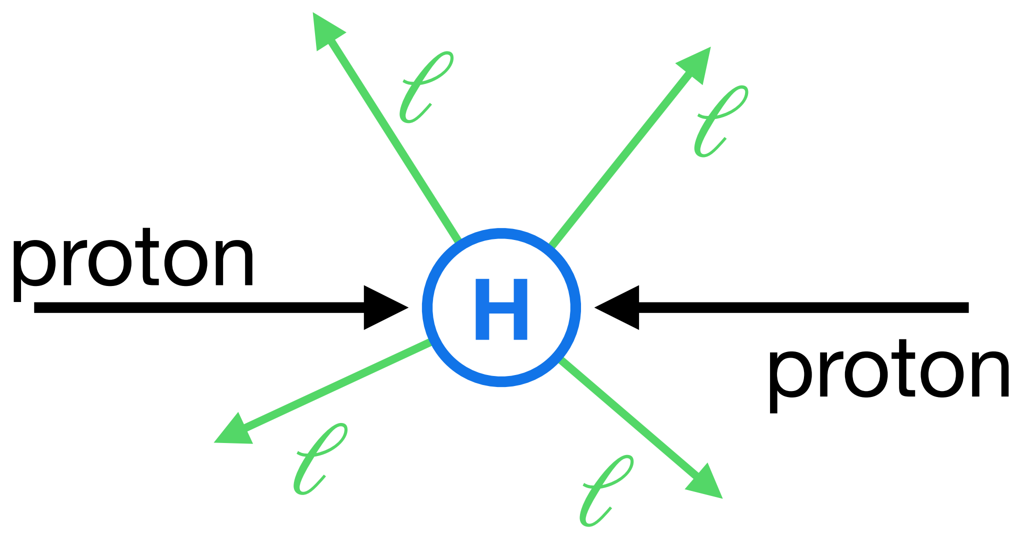


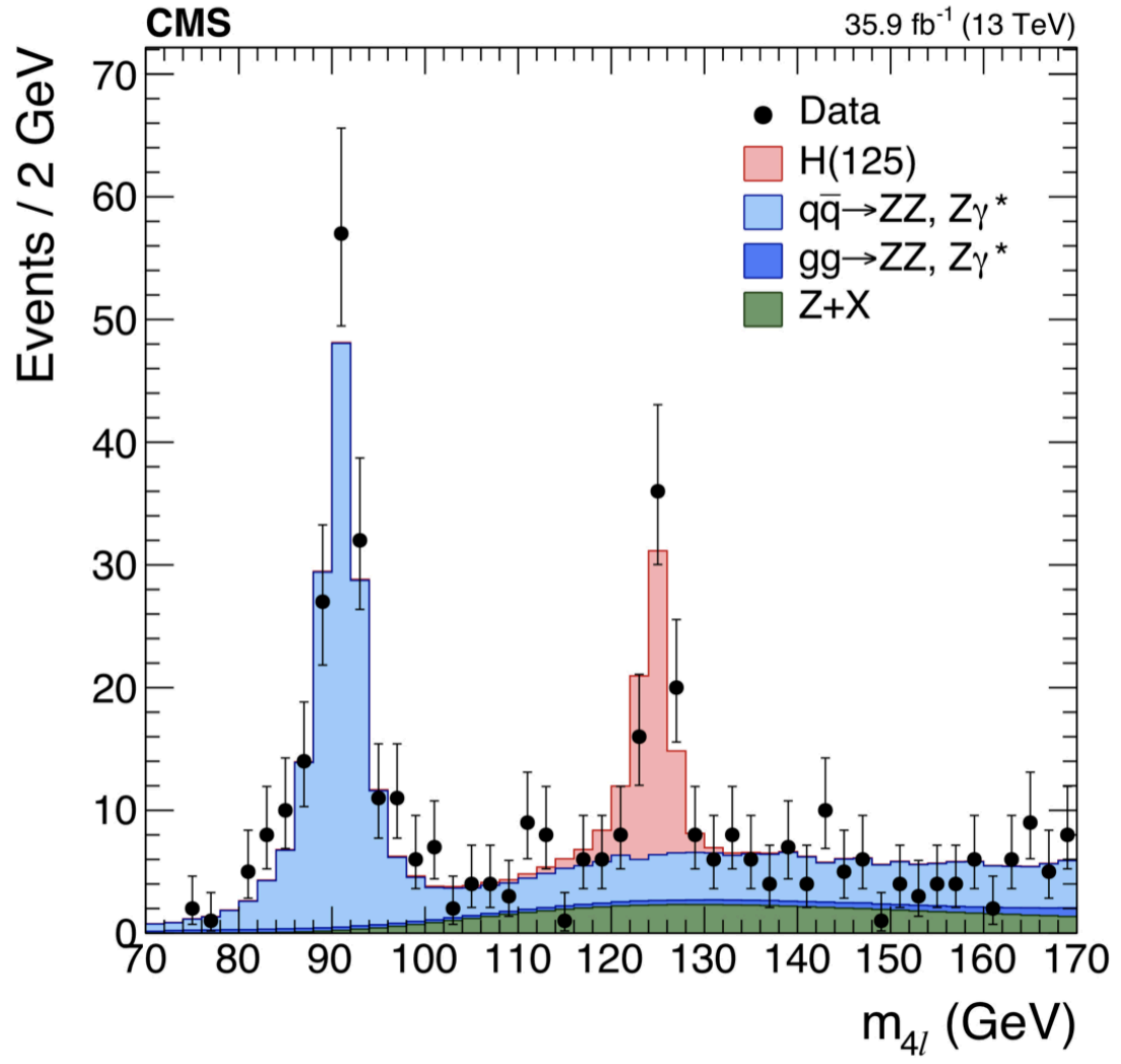




Searches have yet to find new particles…

Standard LHC Searches

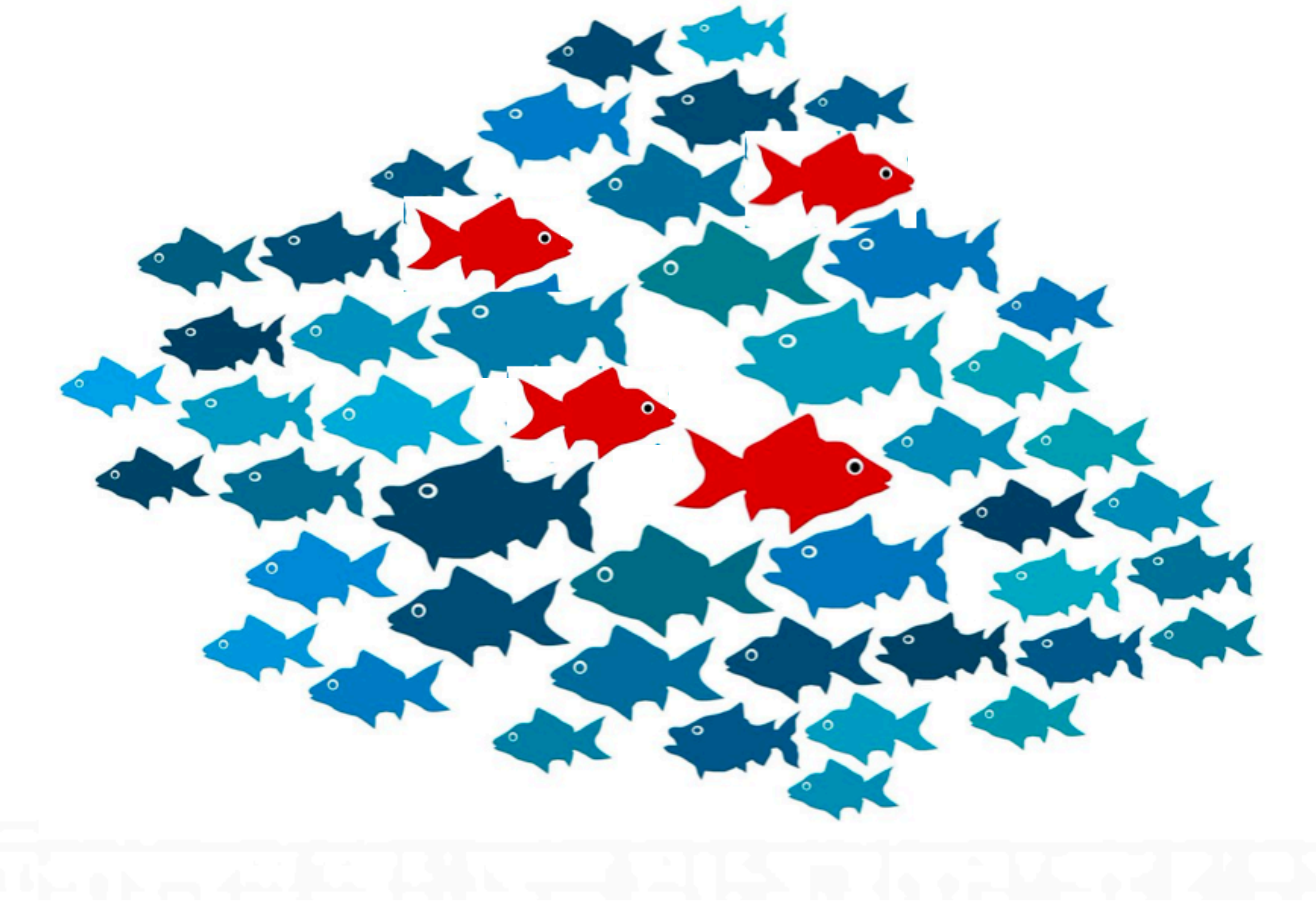




- Most searches have similar basic strategy
- Search for new particle decaying to specific particles ($H \to 4\ell$)
- Select set of events that match this (eg. have 4ℓ)
- Define some variable based on this ($f(\ell_1, \ell_2, \ell_3, \ell_4) \sim E_{\ell_1} + E_{\ell_2} + E_{\ell_3} + E_{\ell_4}$)
- Compare the # of events we expect to find based on Standard Model (simulation) to the # of events we actually measure (data) (■ + ■ + ■ vs. ●)

Anomaly Detection

- What if we don't know exactly what we are looking for?
  - *Select set of events that match what?*
  - *Define what variable?*
- ML offers unique solution to this challenge (no traditional alternative)
  - Broad field of ***anomaly detection (AD)***



Unsupervised Learning

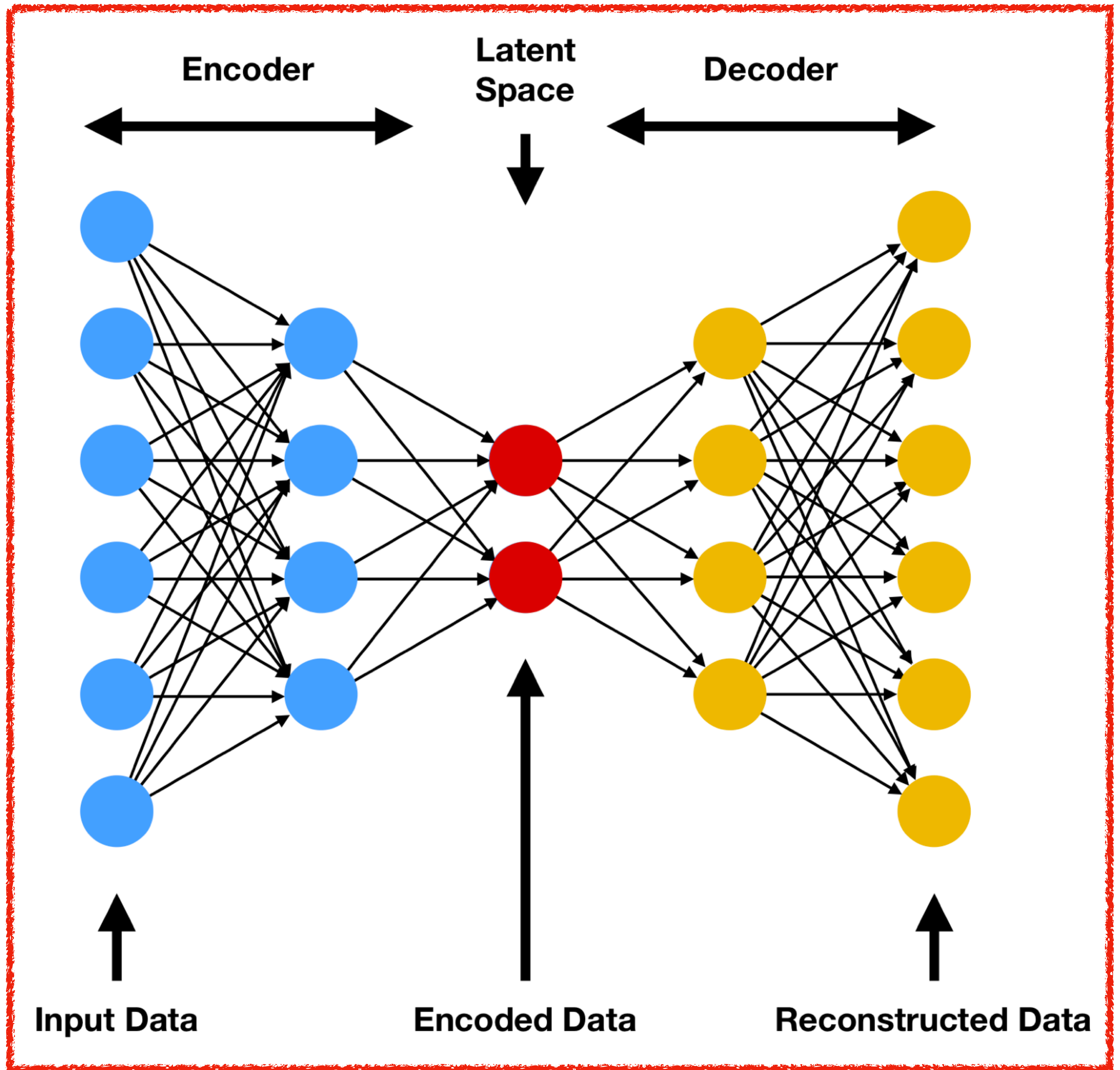
- What if we don't have/can't use labels? → "Unsupervised learning" or "self-supervised learning"

- **Autoencoder (AE):**

- $$\text{Loss} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2$$

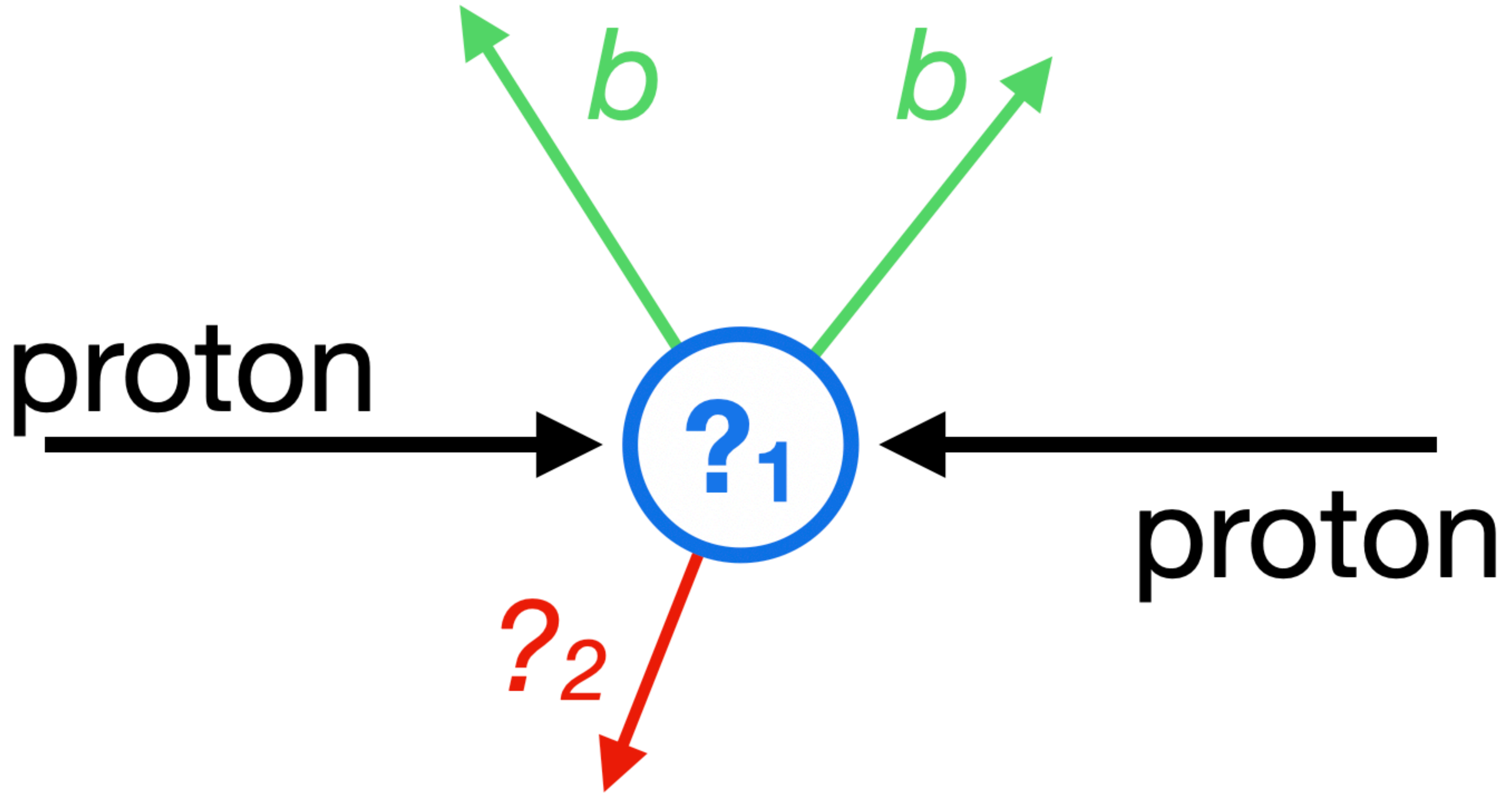
  mean squared error (MSE)

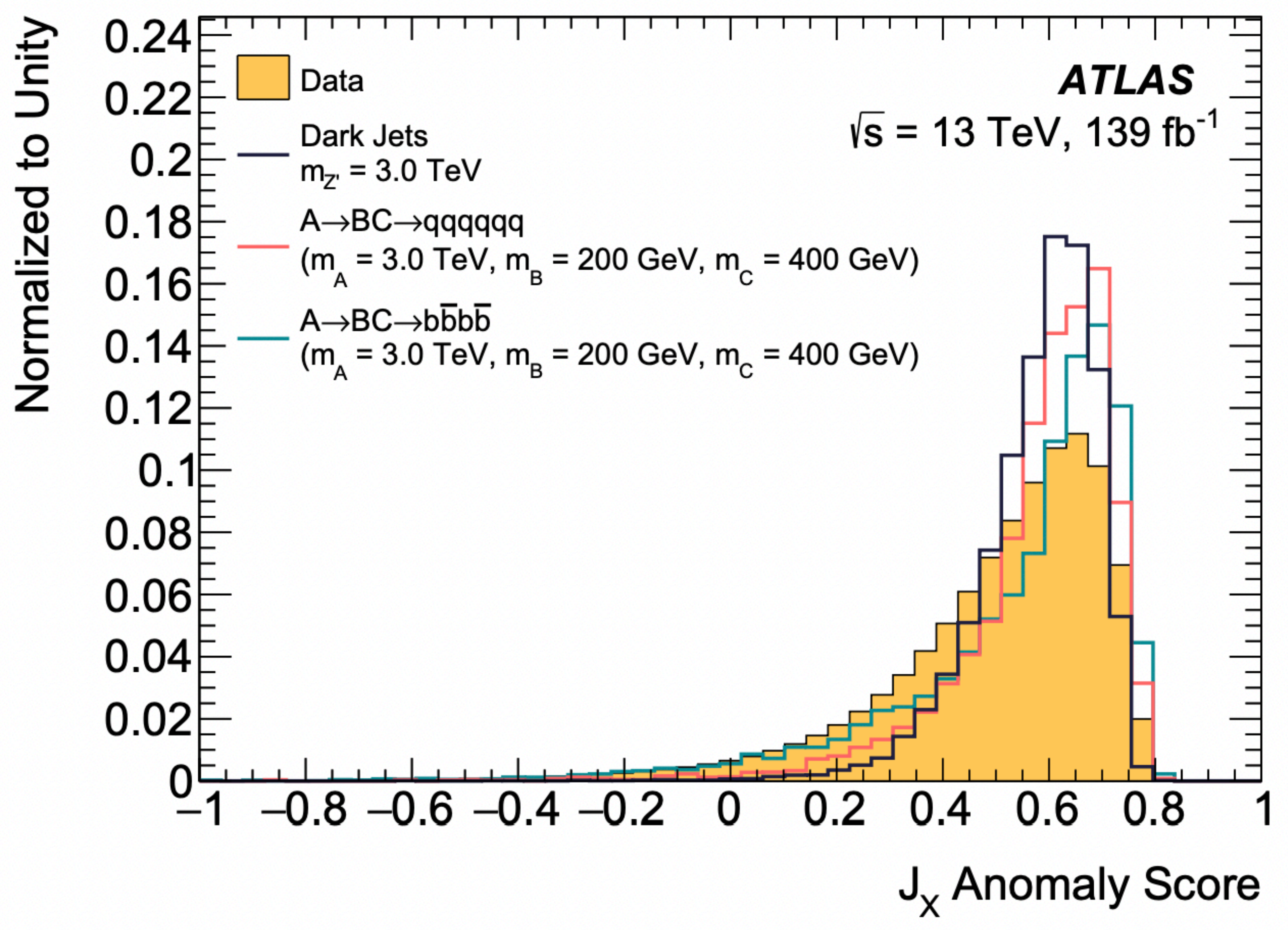
- AE is just simplest form of unsupervised learning



Anomaly Detection

- Train AE using known Standard Model processes
- Events with new particles may not reconstruct well
- *Select set of events that* don't reconstruct well
- $f(b_1, b_2, ?) \sim E_{b_1} + E_{b_2} + E_{?_1}$

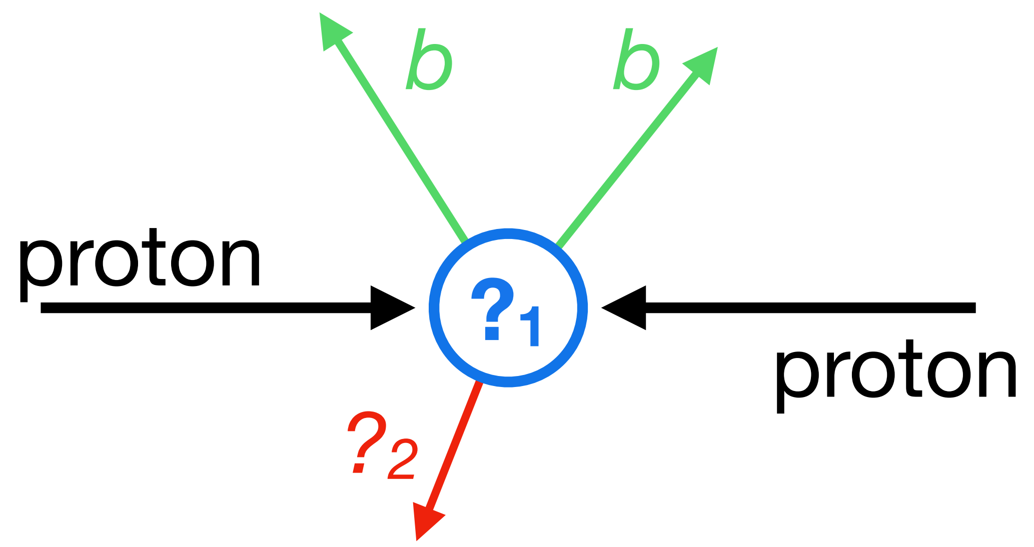


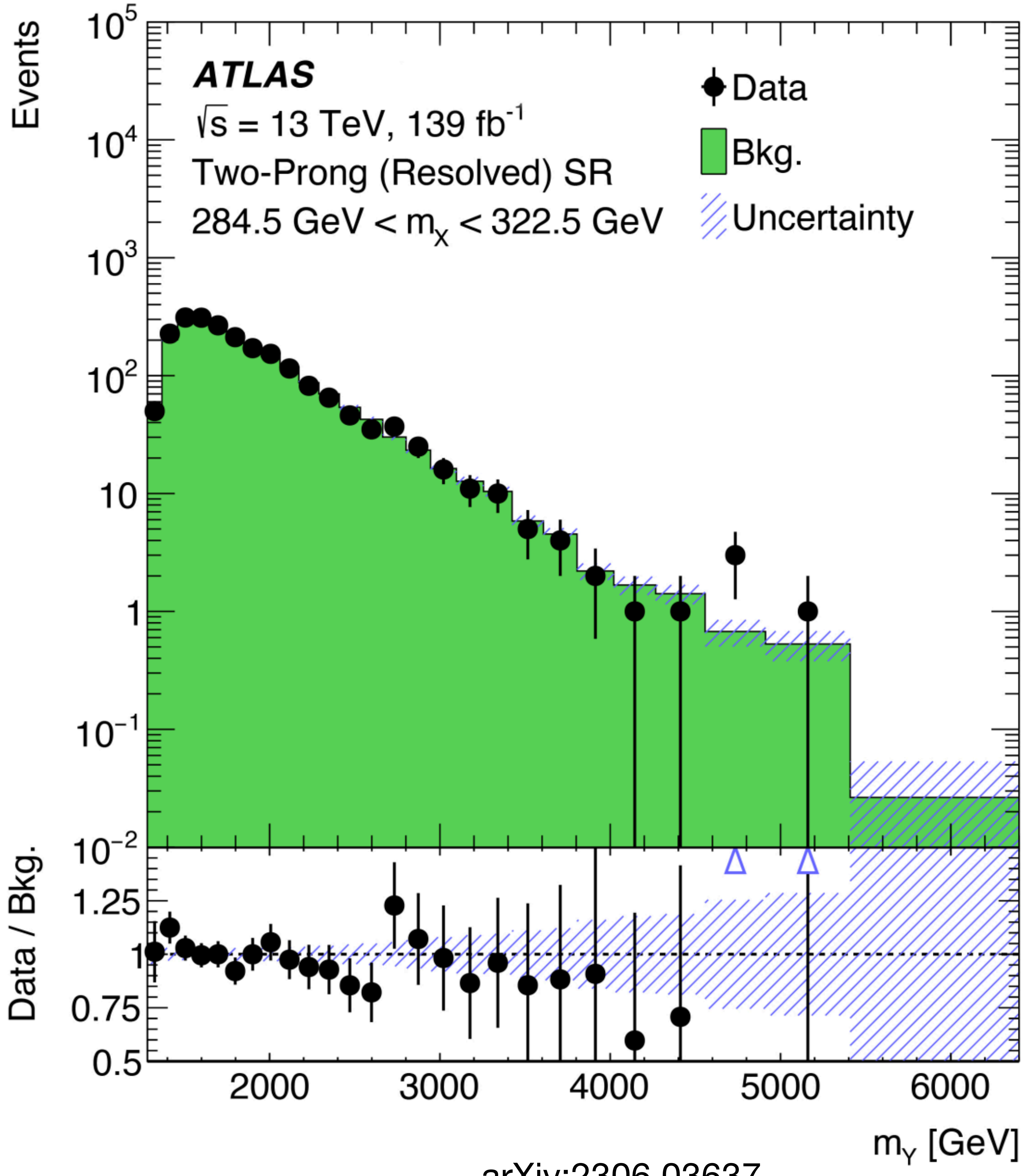


arXiv:2306.03637

Anomaly Detection

- Train AE using known Standard Model processes
- Events with new particles may not reconstruct well
- *Select set of events that* don't reconstruct well
- $f(b_1, b_2, ?) \sim E_{b_1} + E_{b_2} + E_{?_1}$

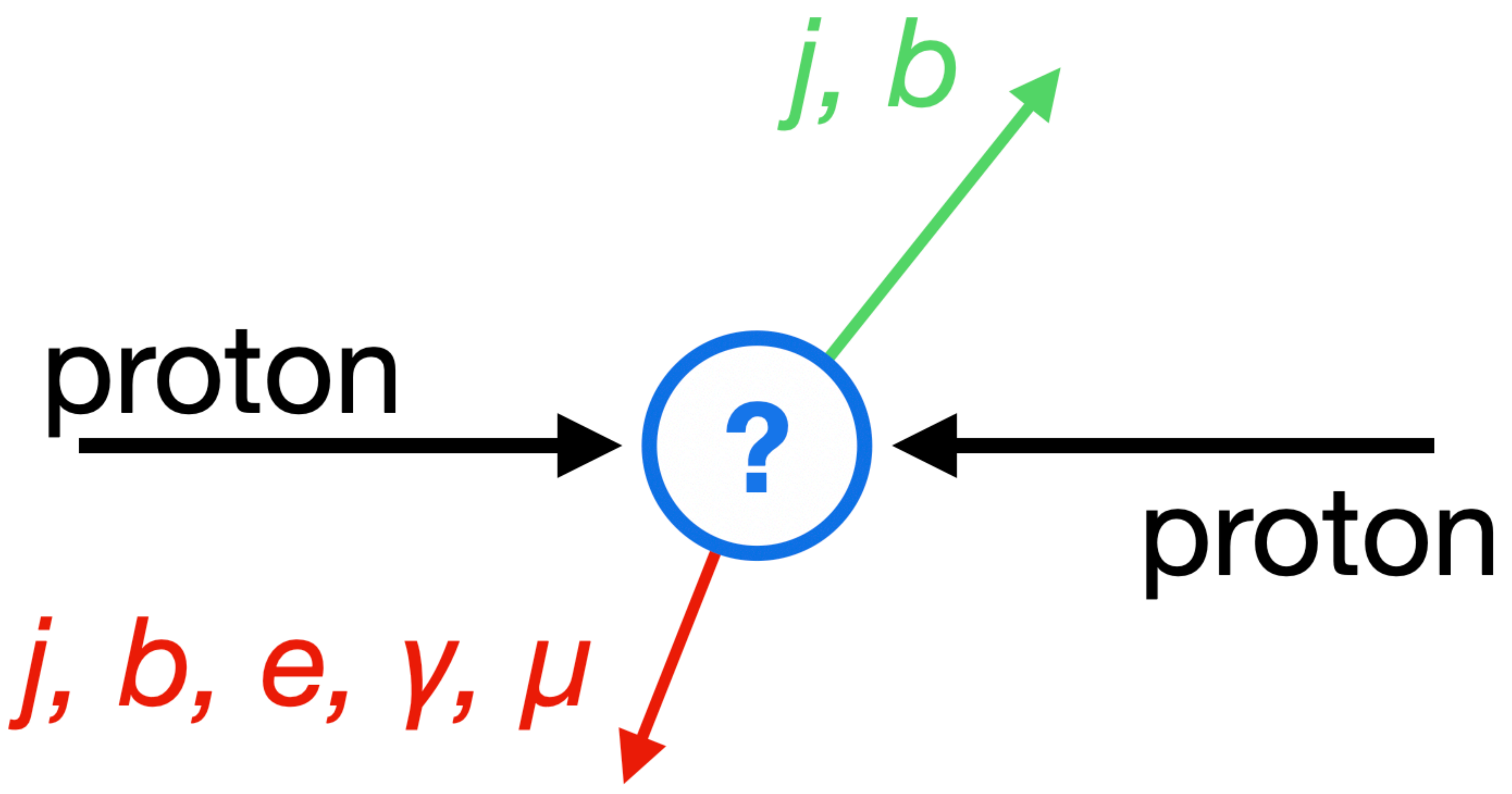




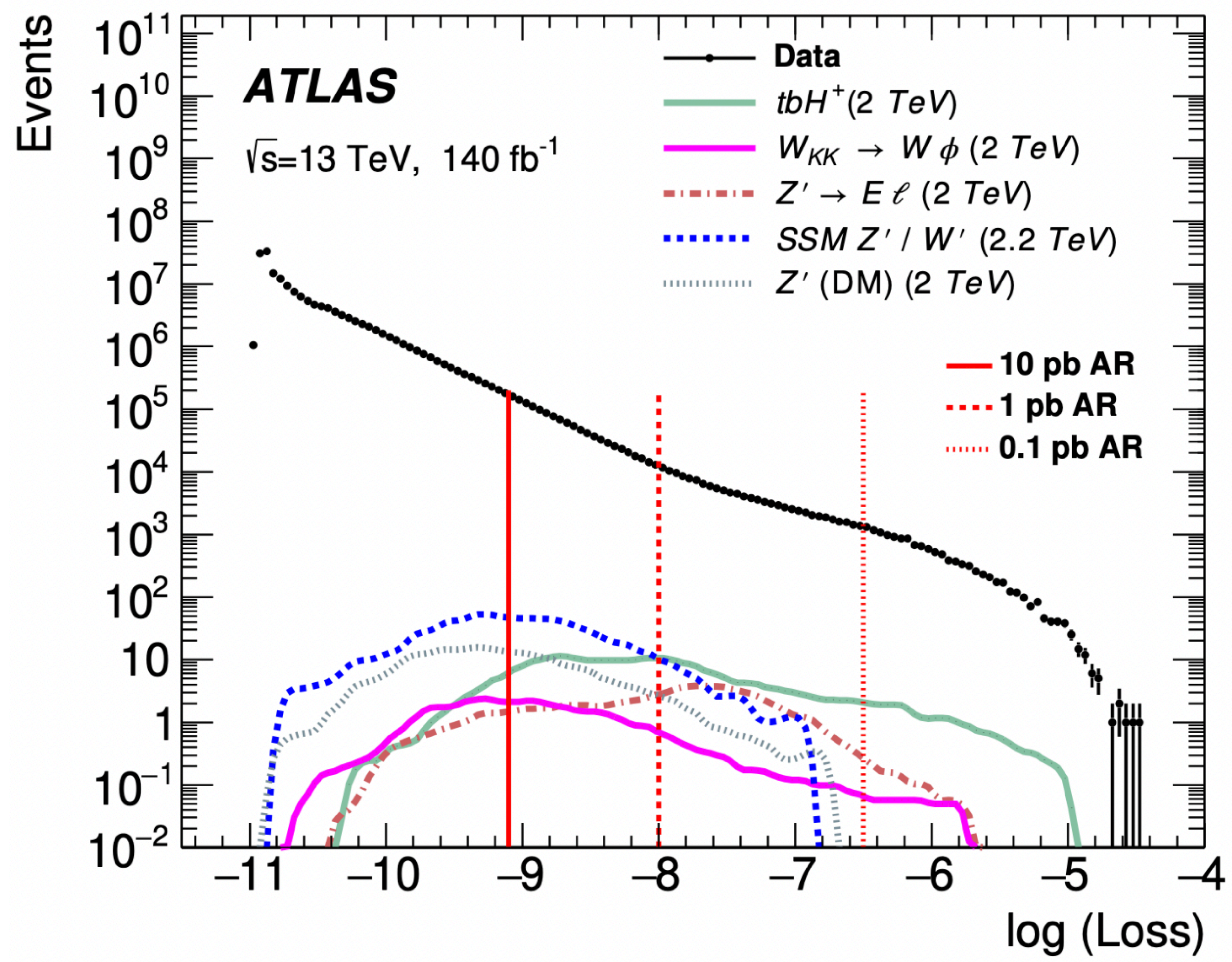
arXiv:2306.03637

Anomaly Detection

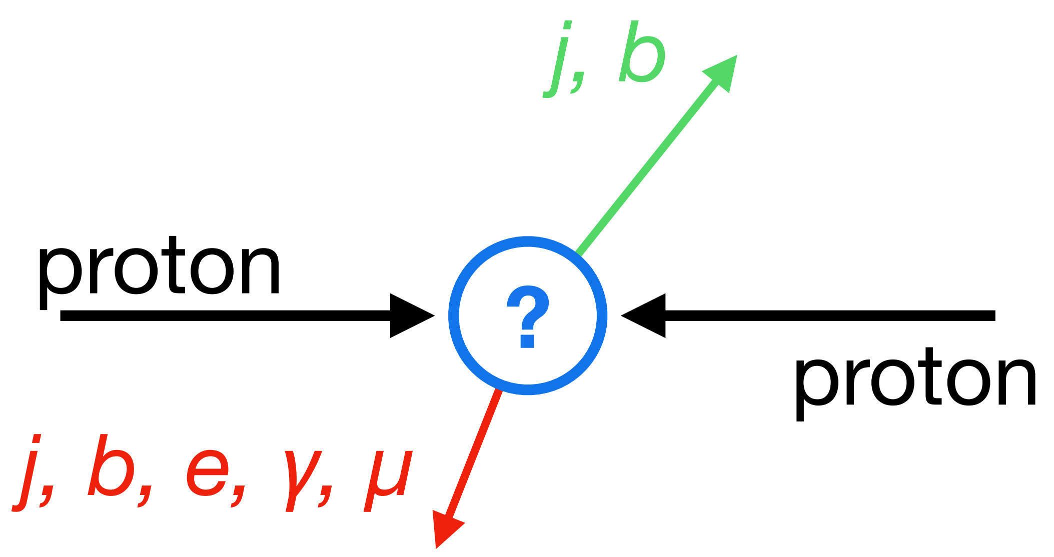
- Train AE using known Standard Model processes
- Events with new particles may not reconstruct well
- *Select set of events that* don't reconstruct well
- $f(\{j, b\}_1, \{j, b, e, \gamma, \mu\}_2) \sim E_{\{j,b\}_1} + E_{\{j,b,e,\gamma,\mu\}_2}$



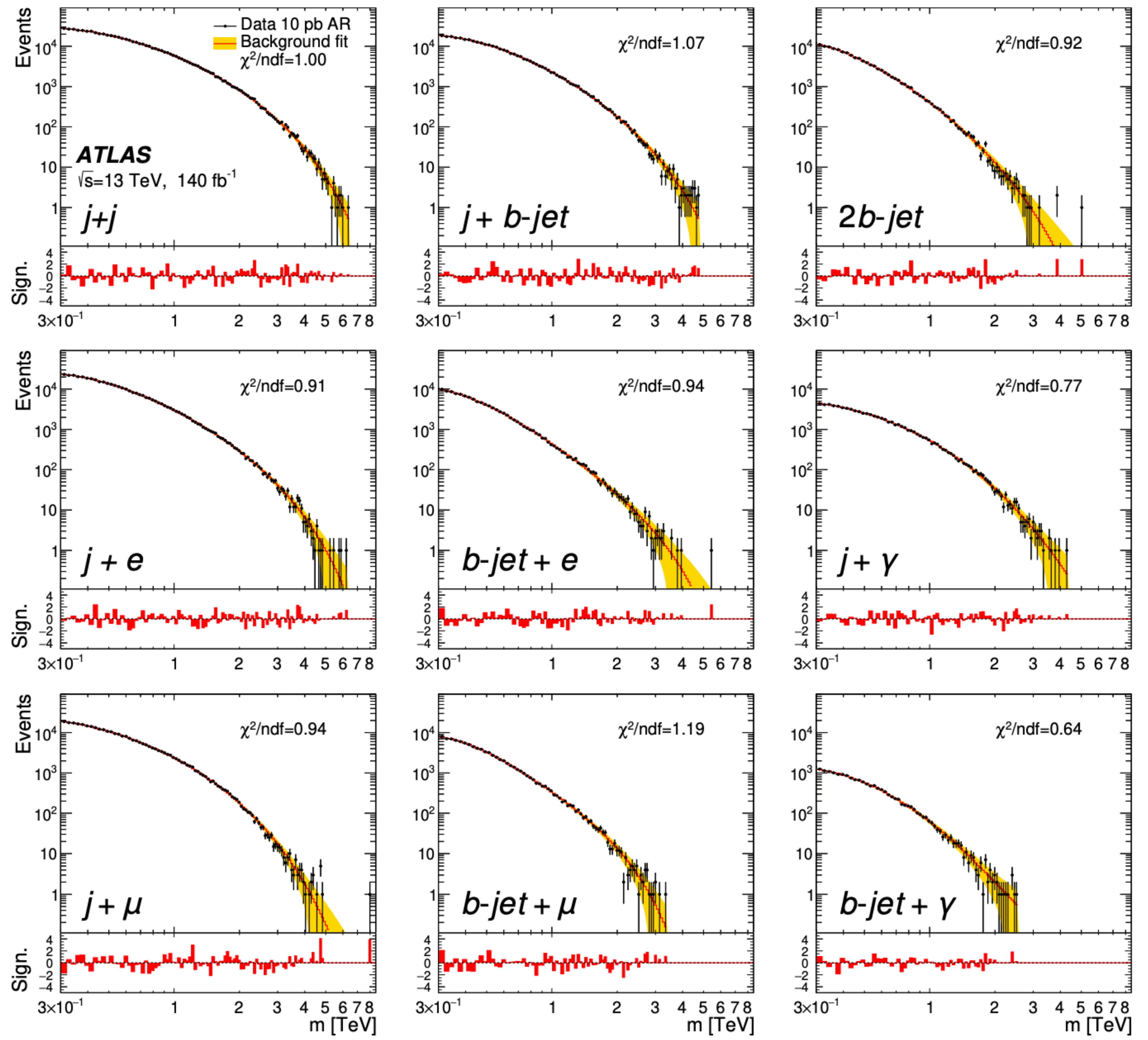
arXiv:2307.01612



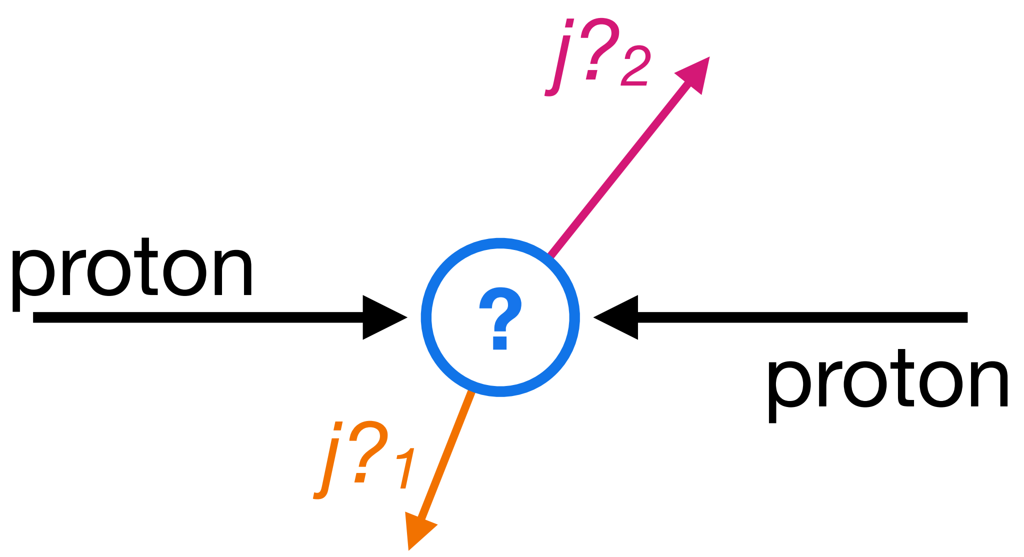
Anomaly Detection



- Train AE using known Standard Model processes
- Events with new particles may not reconstruct well
- *Select set of events that* don't reconstruct well
- $f(\{j,b\}_1, \{j,b,e,\gamma,\mu\}_2) \sim E_{\{j,b\}_1} + E_{\{j,b,e,\gamma,\mu\}_2}$



Anomaly Detection

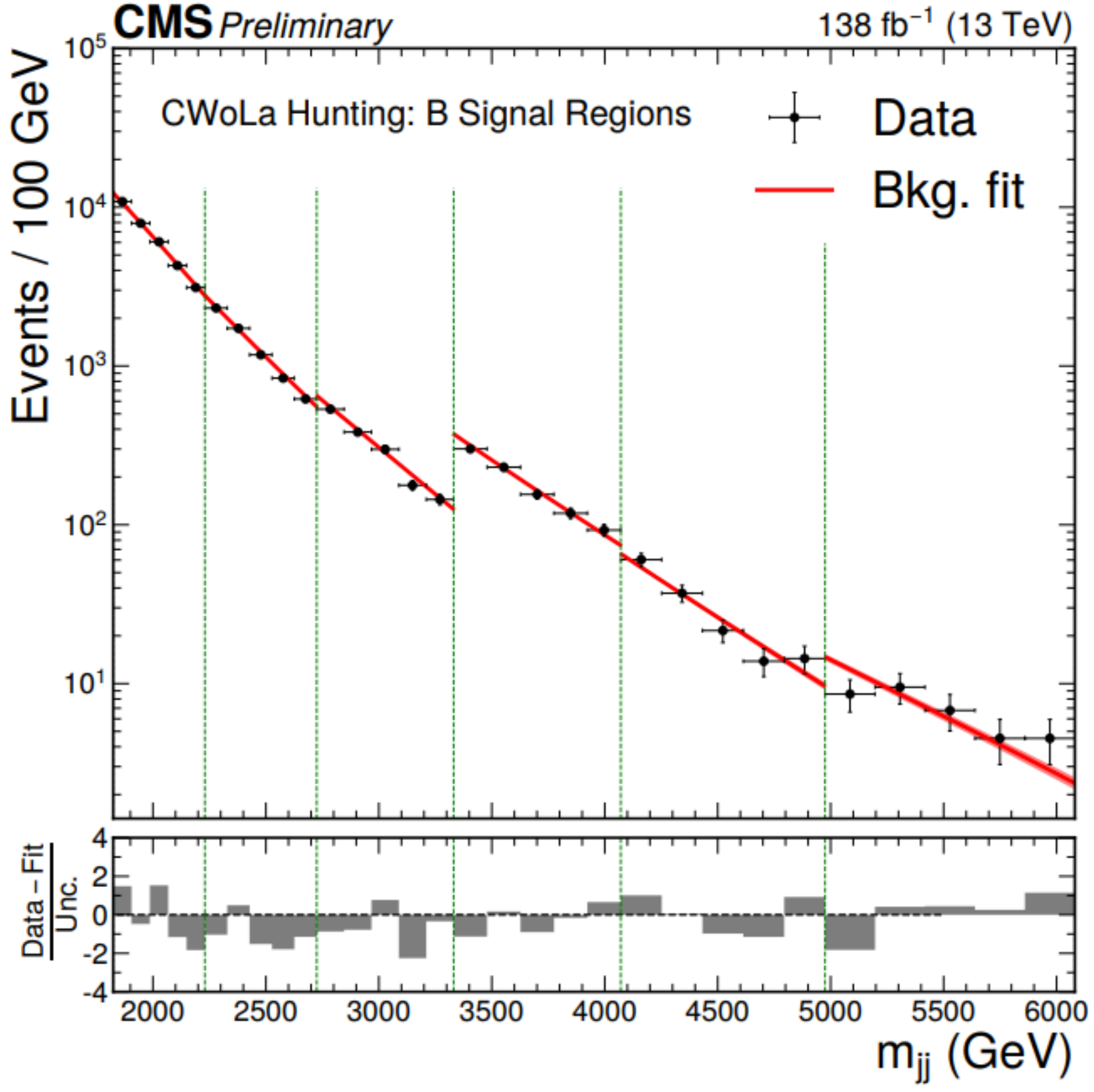


- General AD can be much worse than a dedicated search

- Some methods try to use some concept of signal, still remain insensitive to details

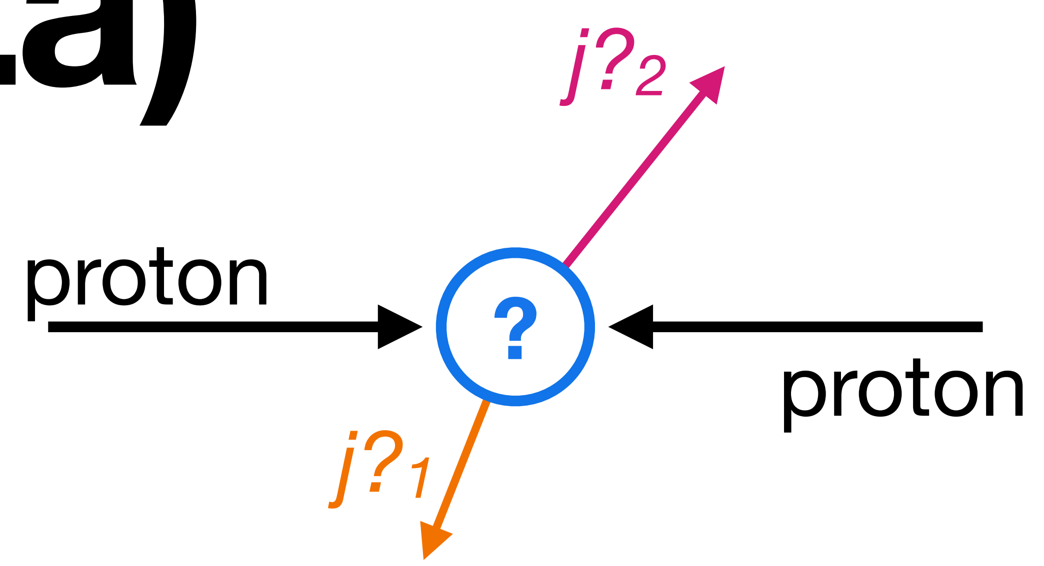
- “Semi-supervised” techniques

- $f(j?_1, j?_2) \sim E_{j?_1} + E_{j?_2}$

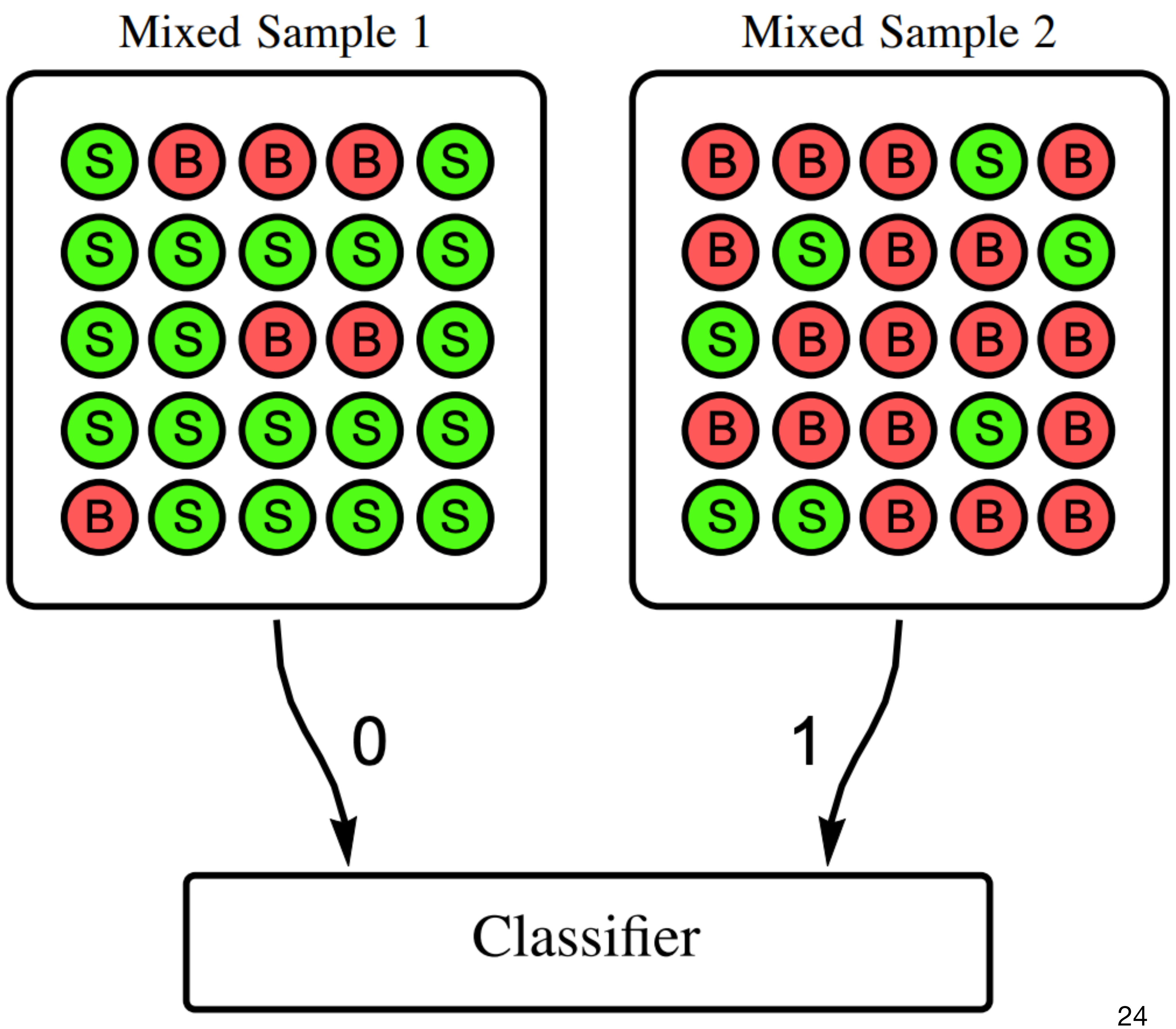


CMS-PAS-EXO-22-026

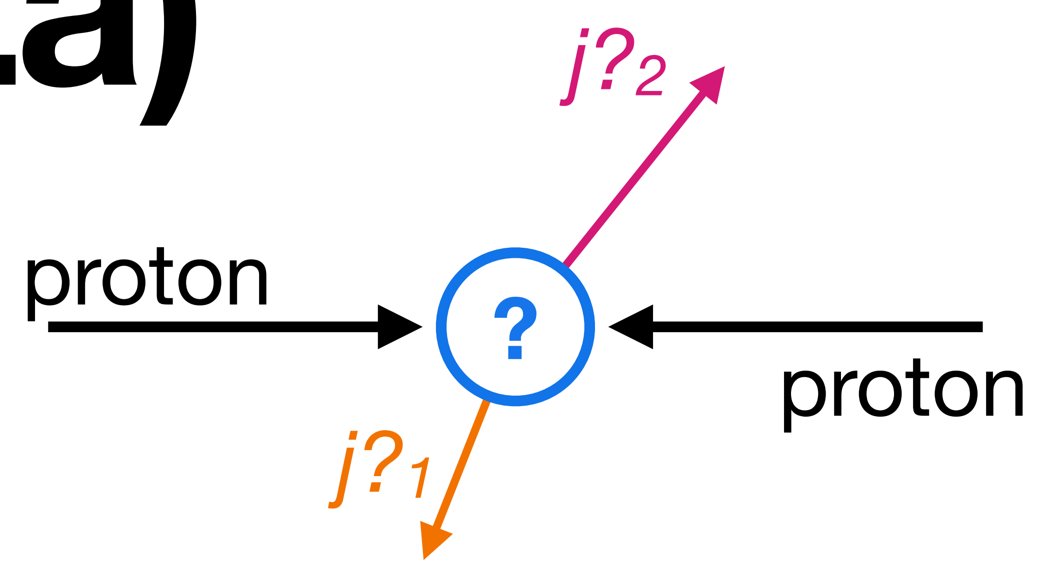
Classification Without Labels (CWoLa)



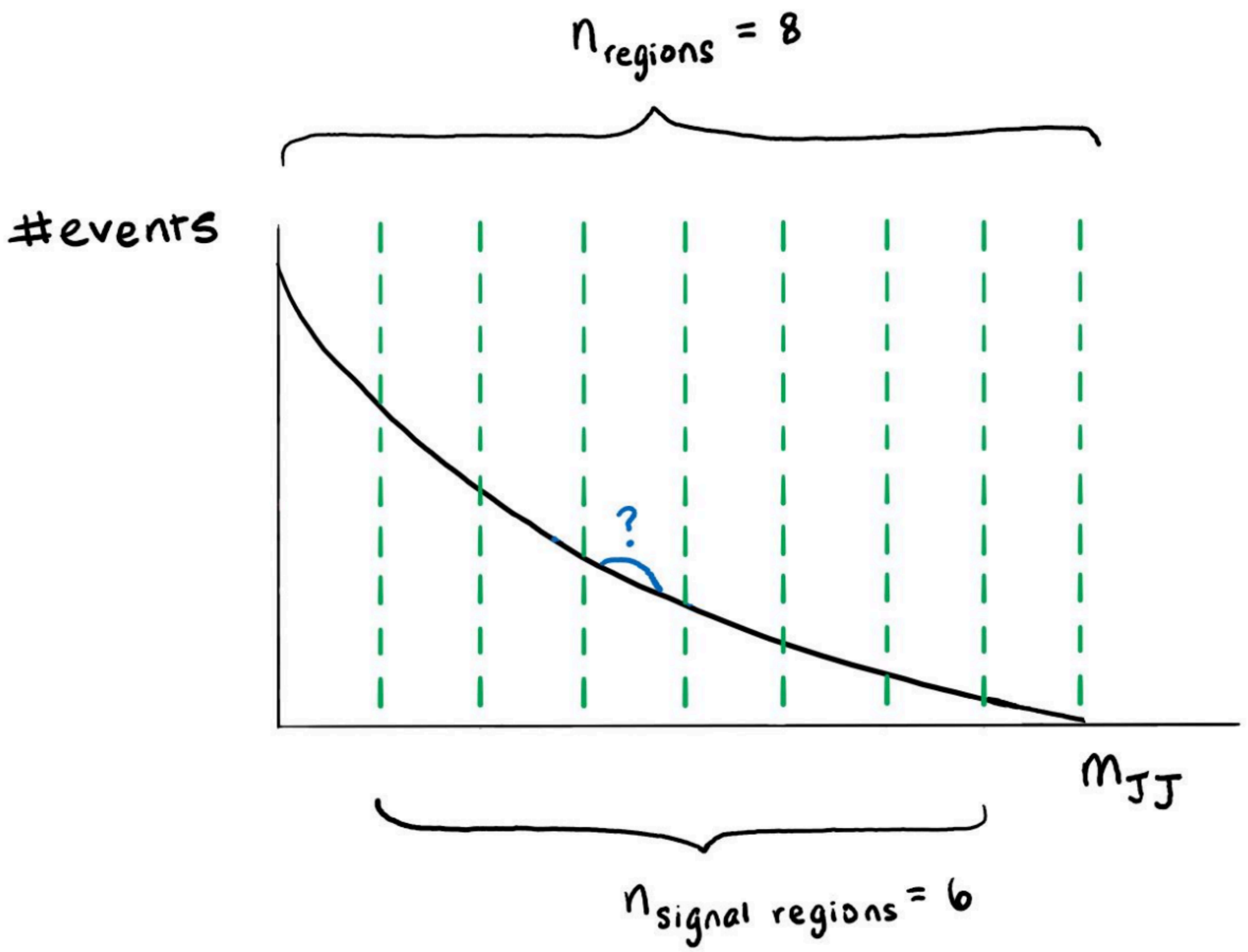
- Semi-supervised method [arXiv:1708.02949]
- Requires only $f(j?_1, j?_2) \sim E_{j?_1} + E_{j?_2}$
- If ? is localized in $f(\cdot)$, can train supervised classifier based on mixed samples
  - In the limit this will approach normal supervised performance: classifier learns find whats different about sample 1 and 2 (S vs. B)

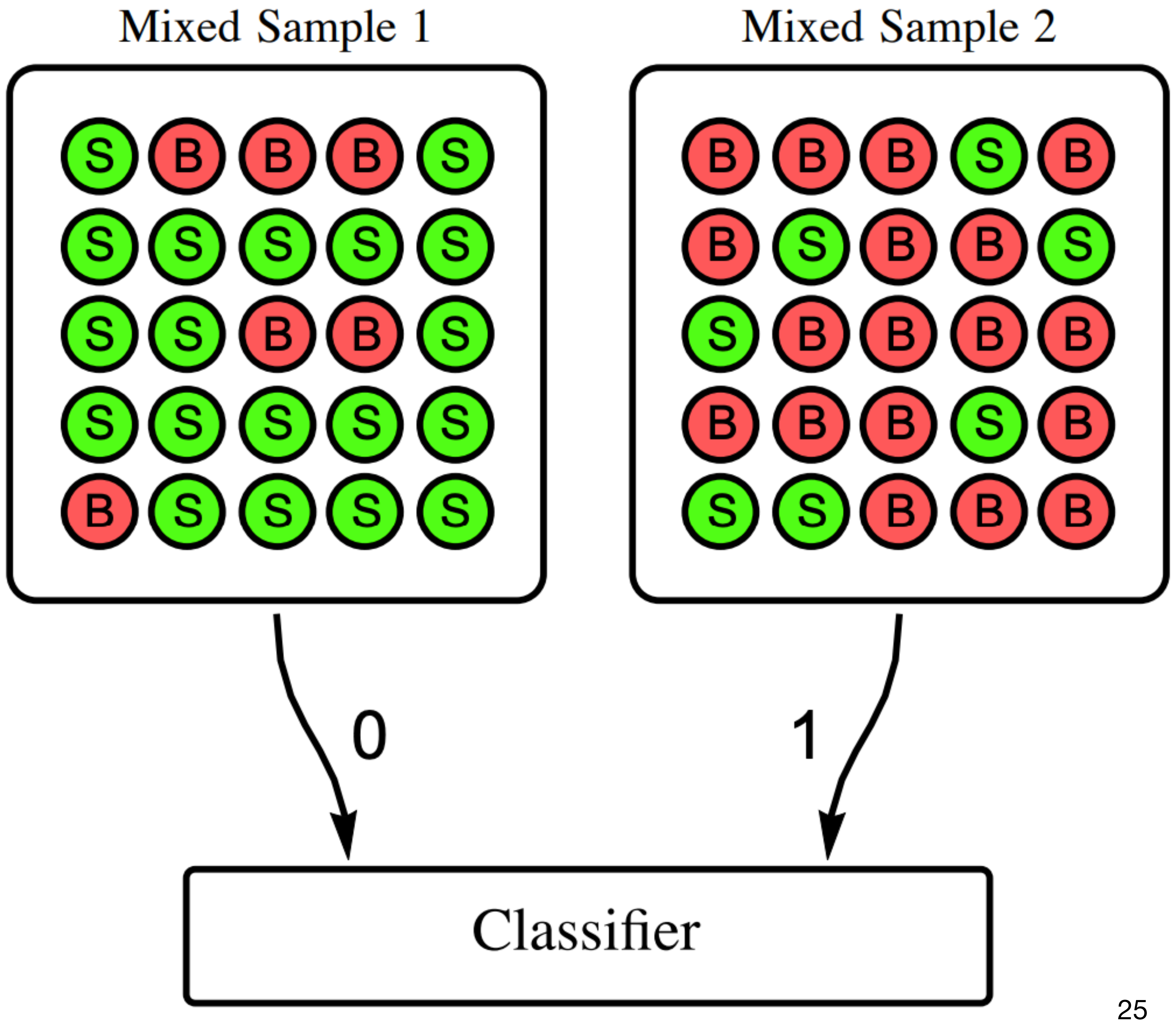


Classification Without Labels (CWoLa)

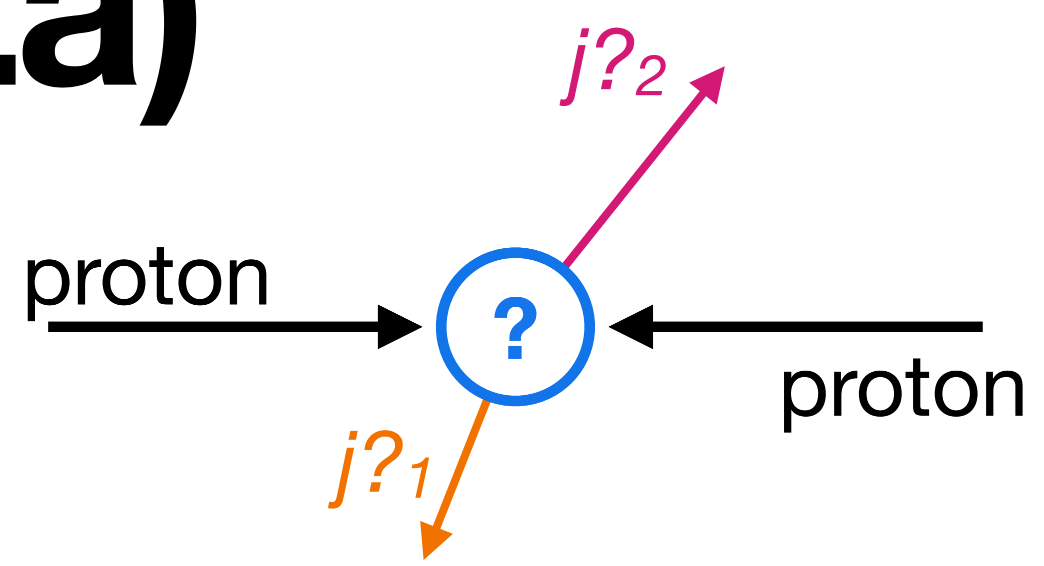


- Semi-supervised method [arXiv:1708.02949]
- Requires only $f(j?_1, j?_2) \sim E_{j?_1} + E_{j?_2}$
- If **?** is localized in $f(\,\cdot\,)$, can train supervised classifier based on mixed samples

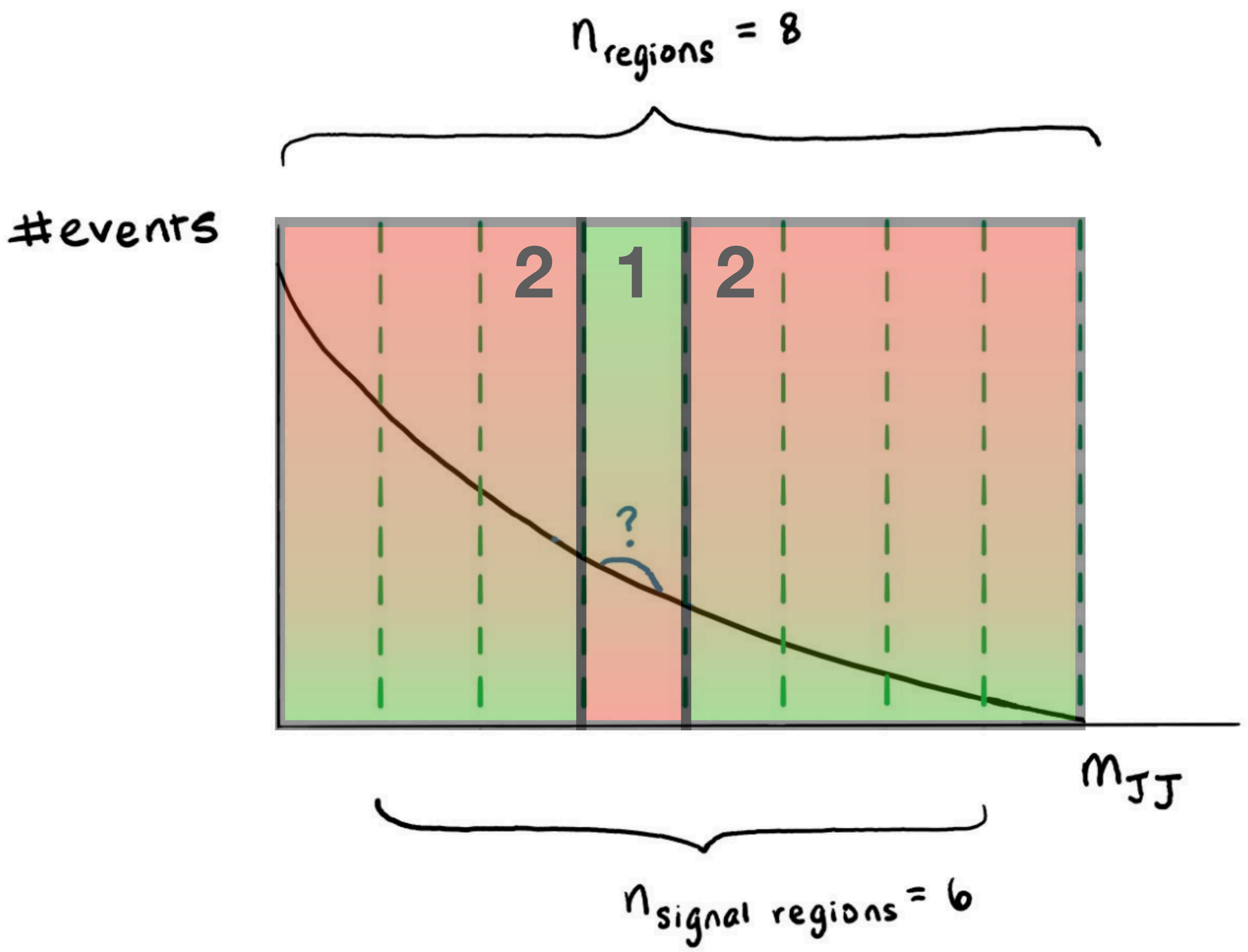


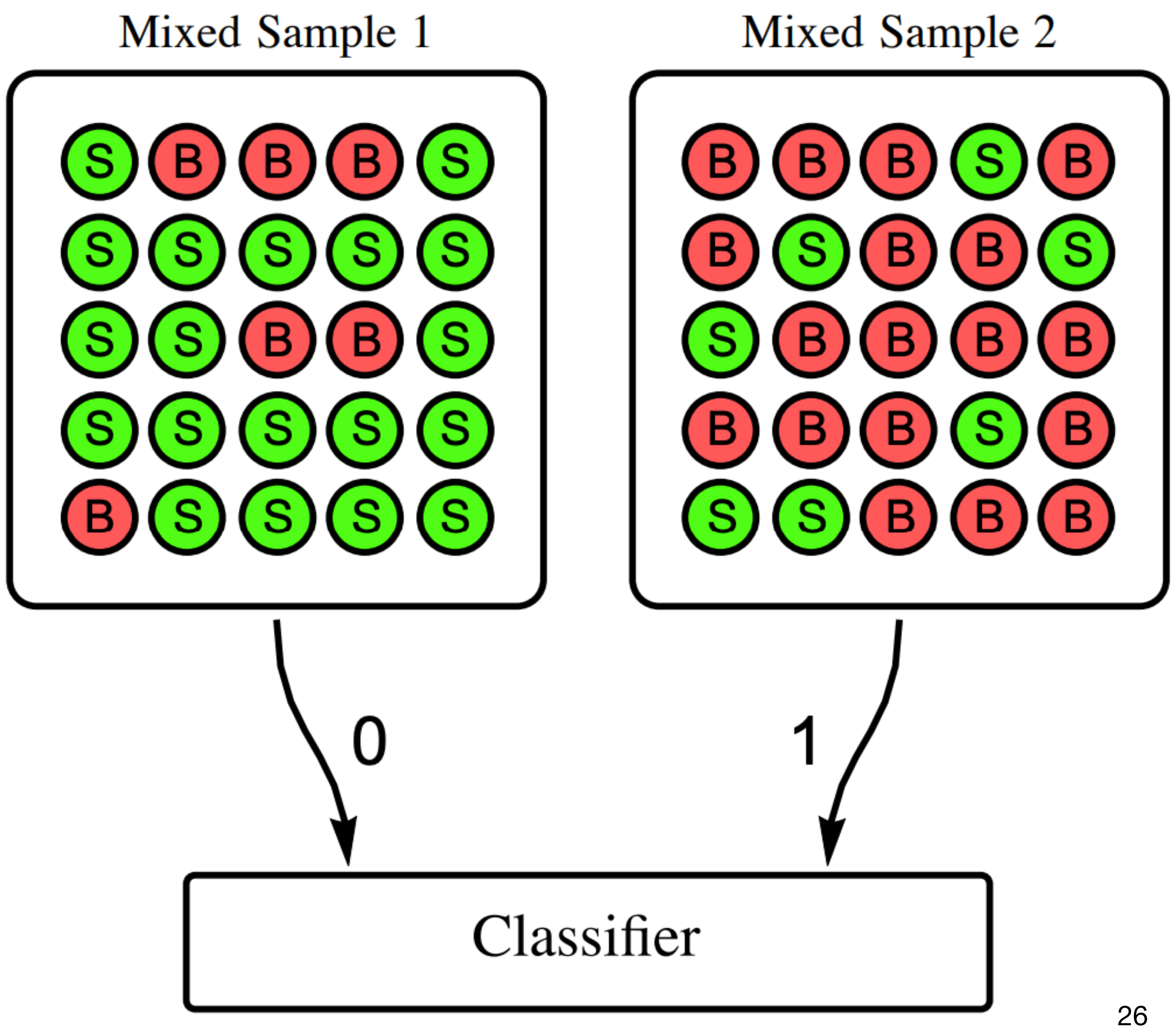


Classification Without Labels (CWoLa)

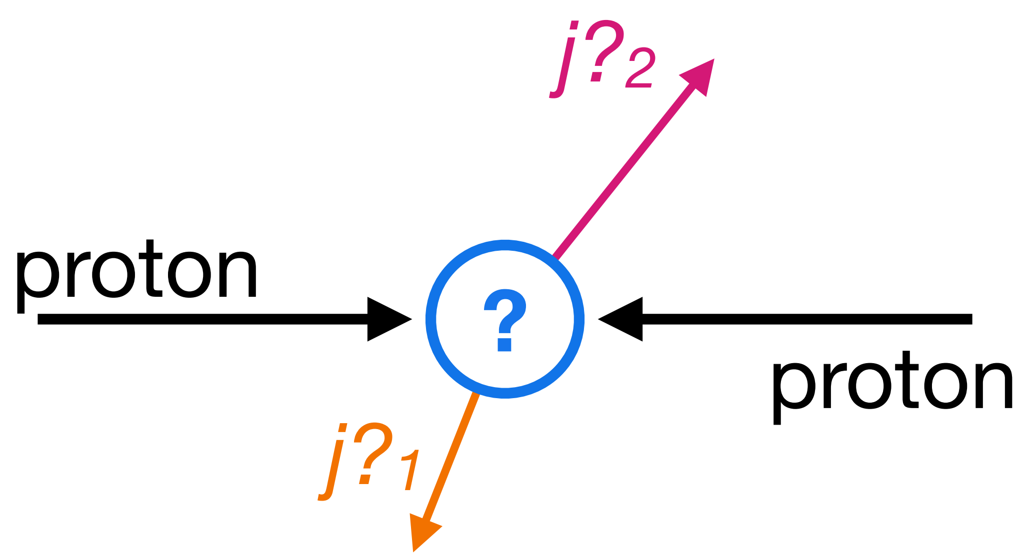


- Semi-supervised method [arXiv:1708.02949]
- Requires only $f(j?_1, j?_2) \sim E_{j?_1} + E_{j?_2}$
- If ? is localized in $f(\cdot)$, can train supervised classifier based on mixed samples





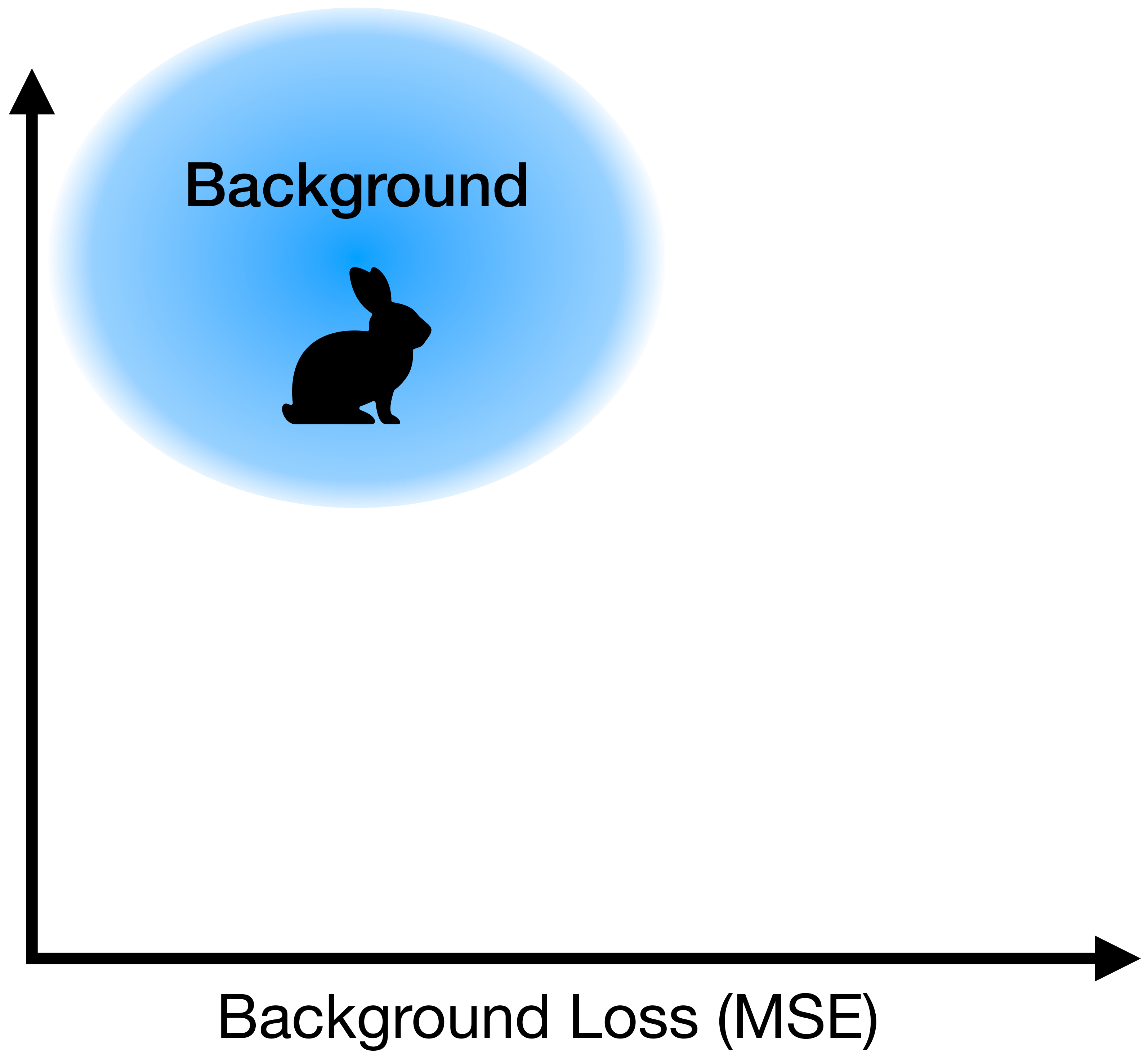
Quasi-Anomalous Knowledge: [2011.03550]



- What if you has some general idea of what signals might look like, but you aren't sure?
- Also possible to set some criteria on roughly what signal should look like, what features it might have
    - Want to point AD in the right direction
- Ideally want method that doesn't break down if the hypothesis turns out to be wrong (like supervised classifiers typically do)

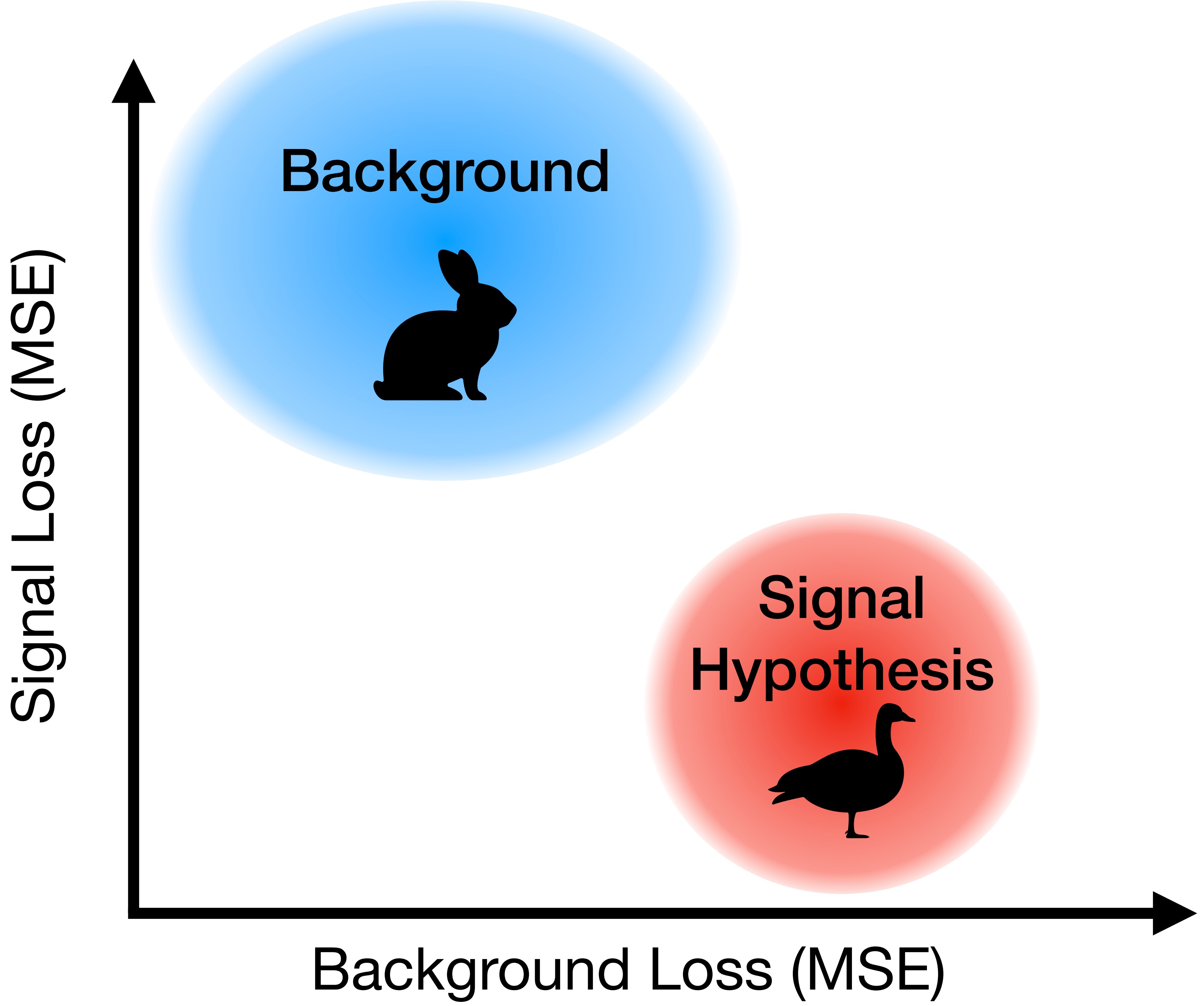
QUAK

- Train an AE for background
- Train an AE for hypothetical signal(s)
- Construct N-dimension QUAK space from losses



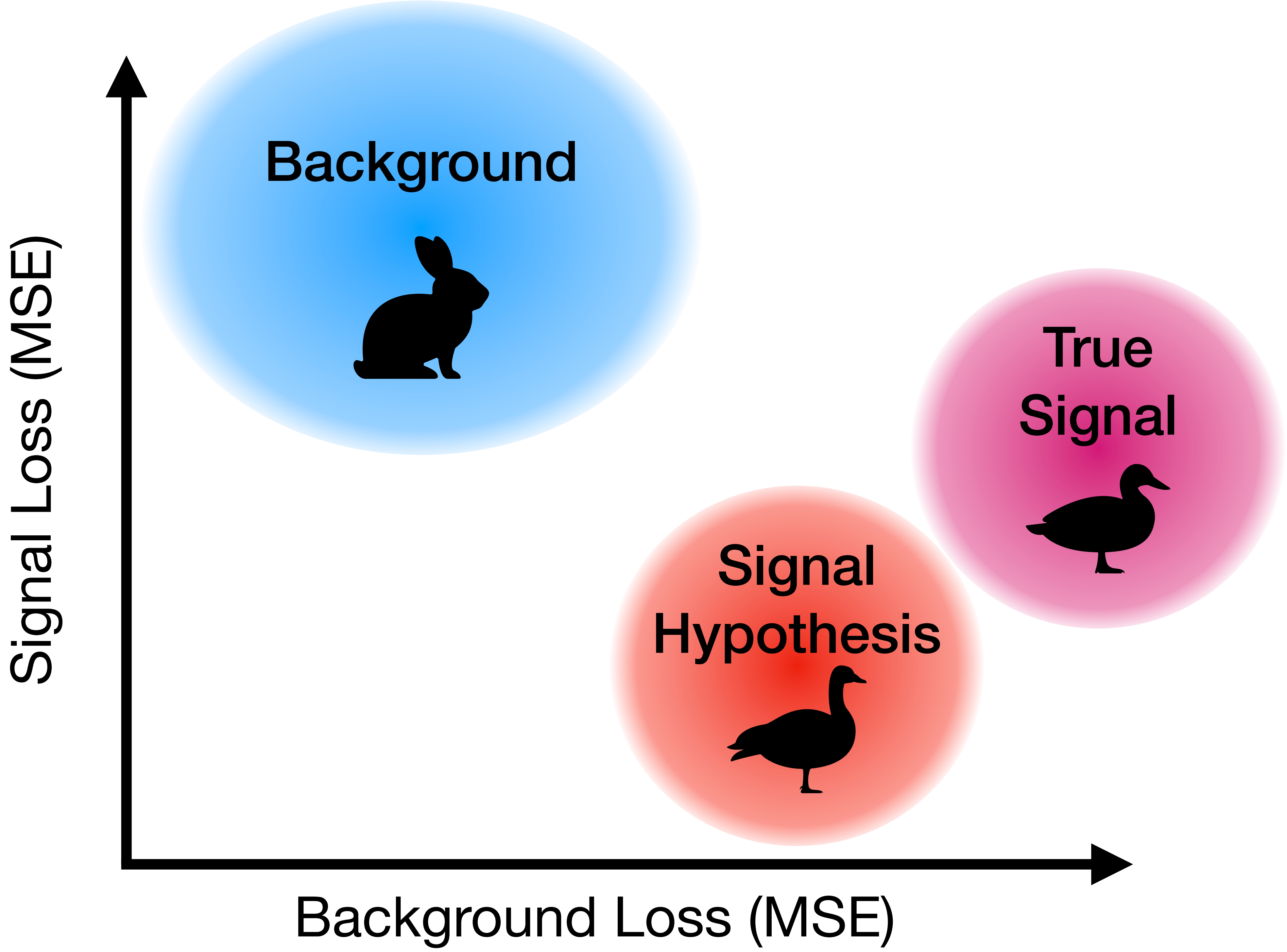
QUAK

- Train an AE for background
- Train an AE for hypothetical signal(s)
- Construct N-dimension QUAK space from losses



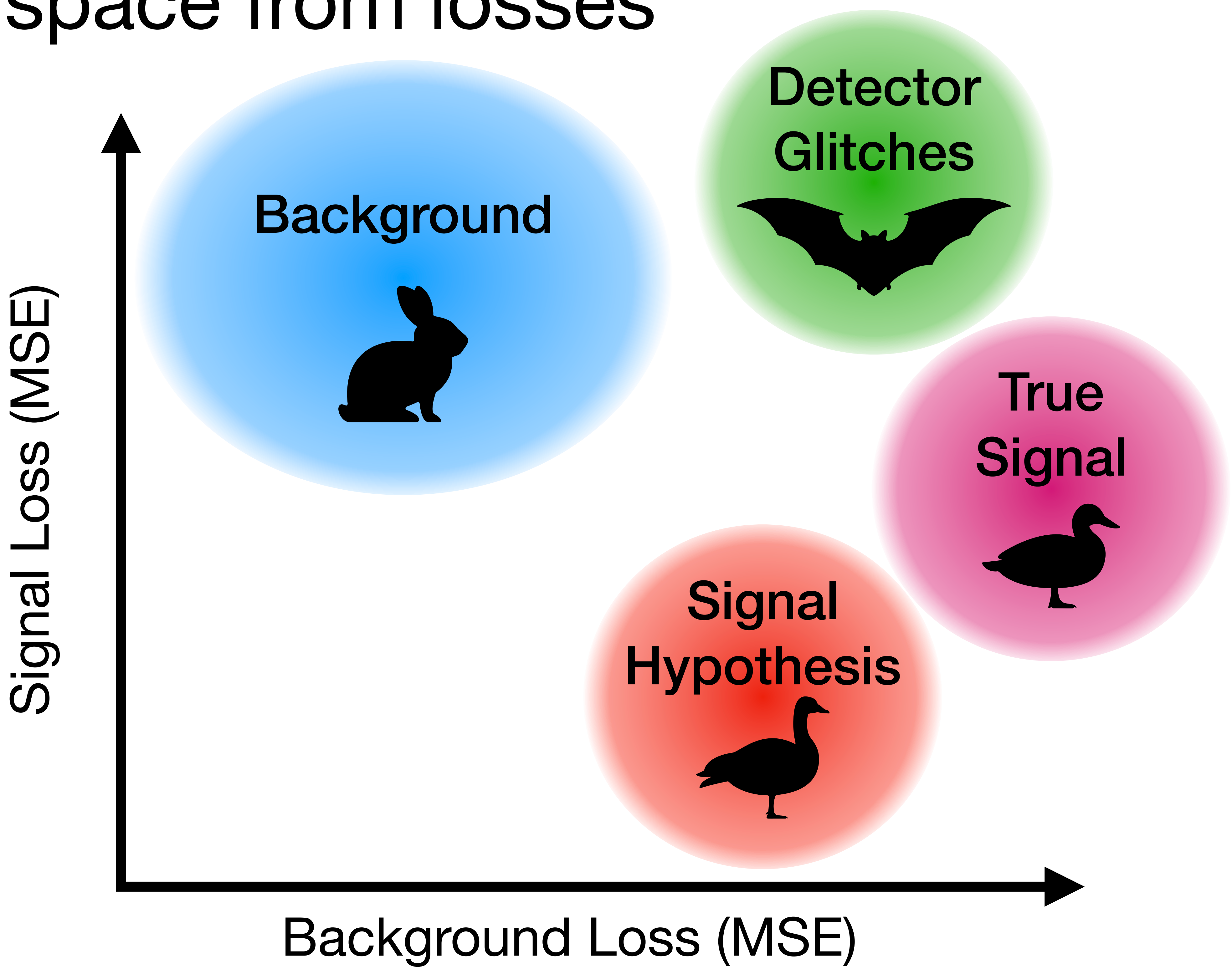
QUAK

- Train an AE for background
- Train an AE for hypothetical signal(s)
- Construct N-dimension QUAK space from losses



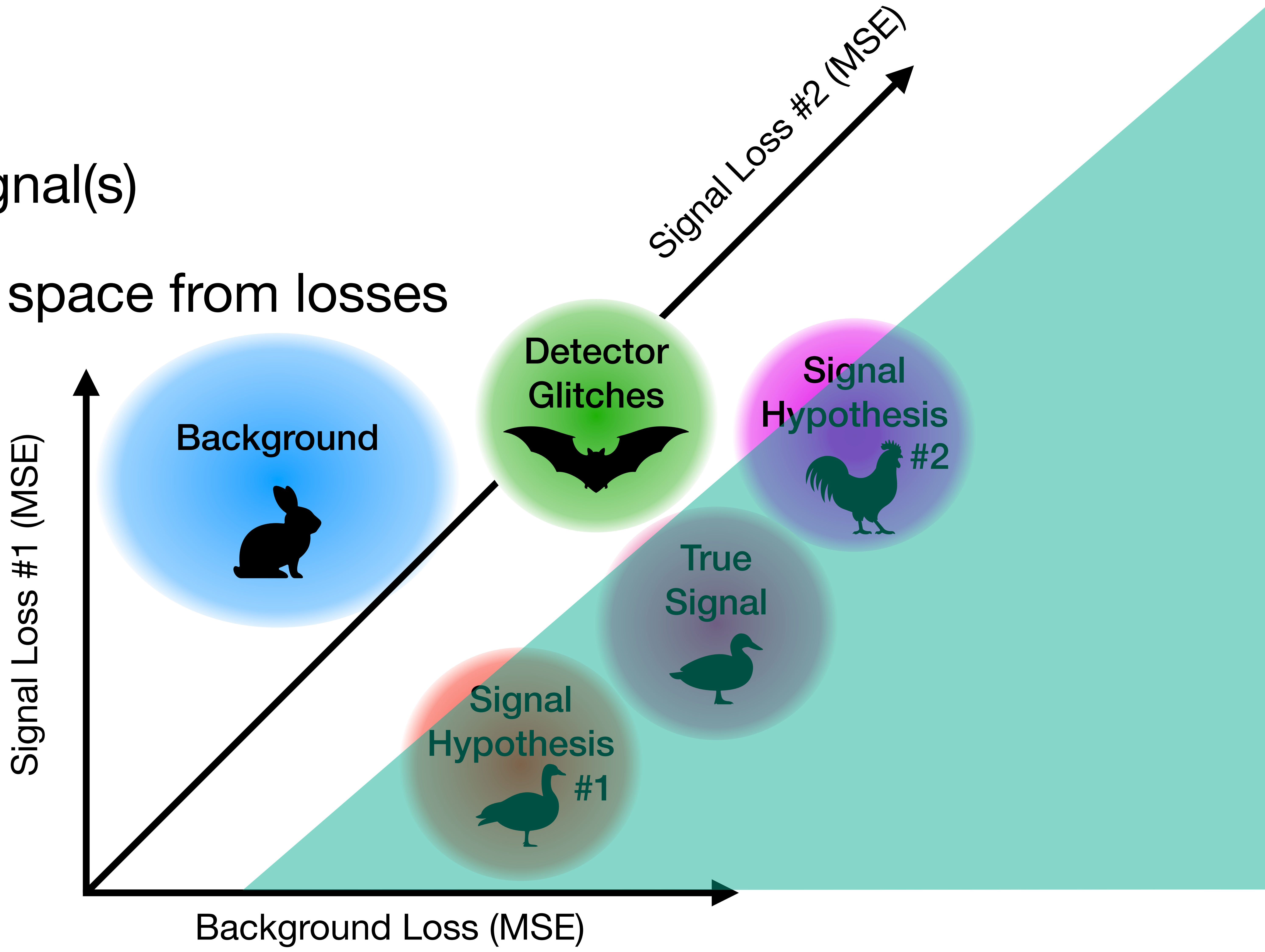
QUAK

- Train an AE for background
- Train an AE for hypothetical signal(s)
- Construct N-dimension QUAK space from losses



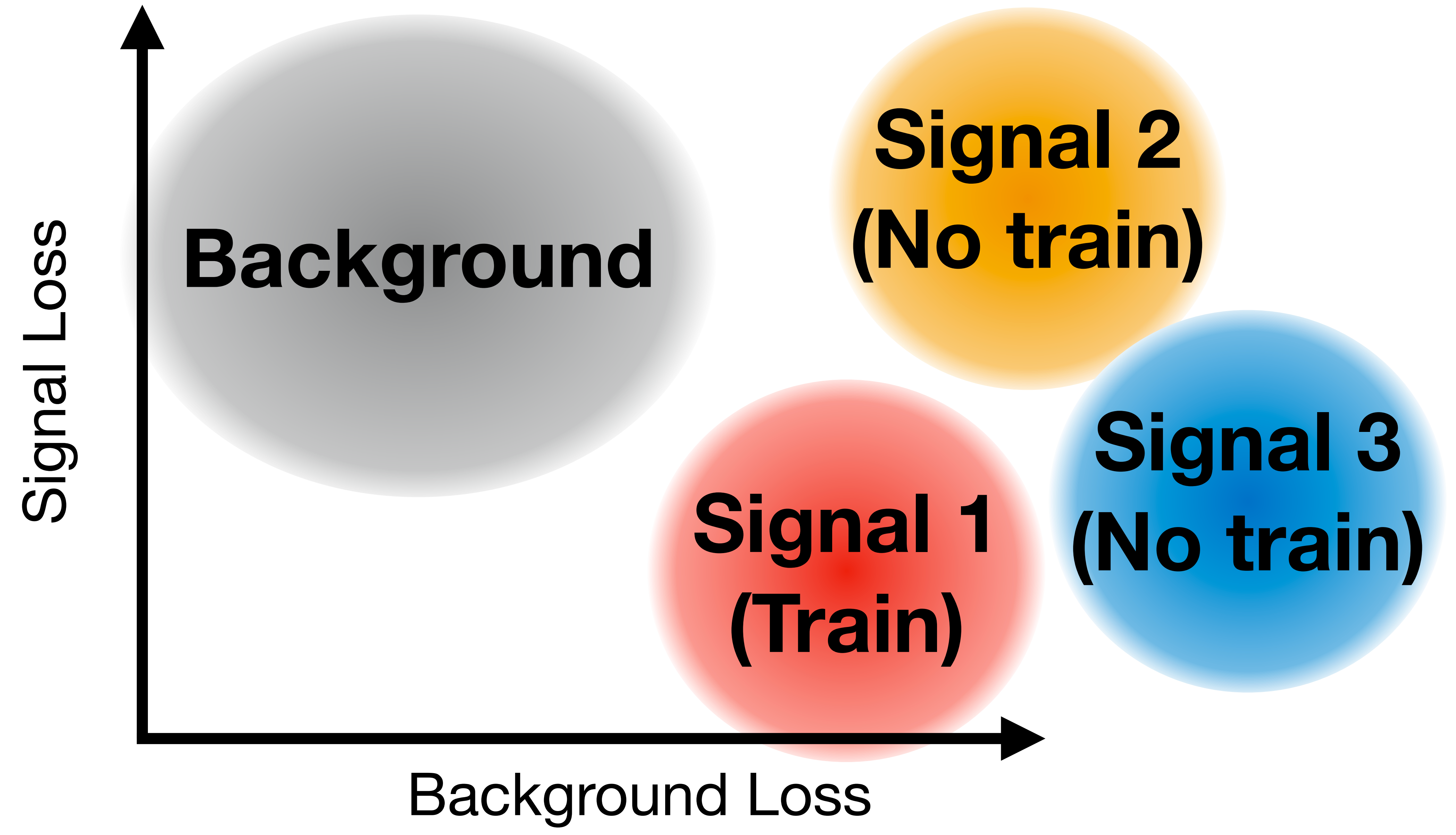
QUAK

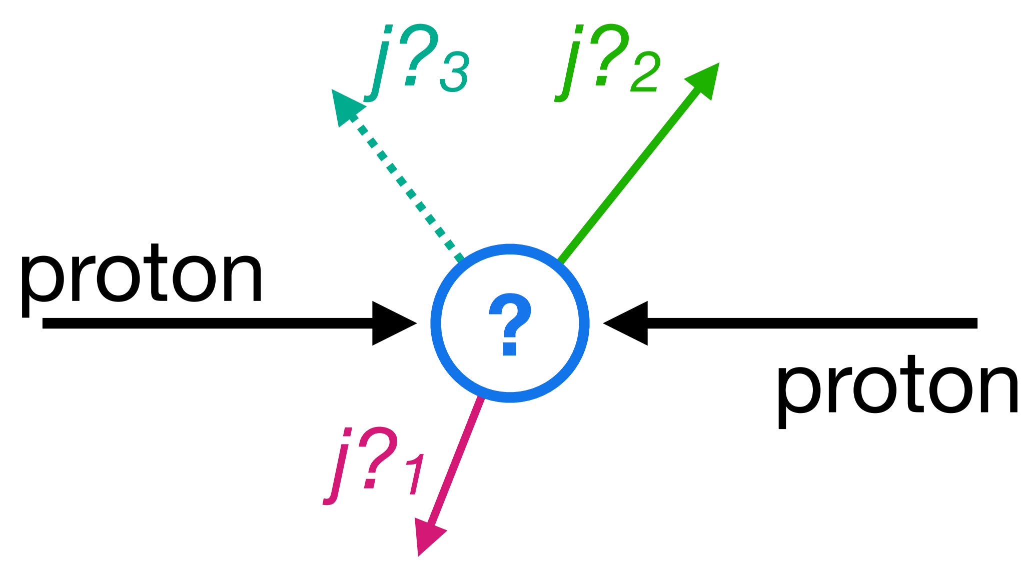
- Train an AE for background
- Train an AE for hypothetical signal(s)
- Construct N-dimension QUAK space from losses

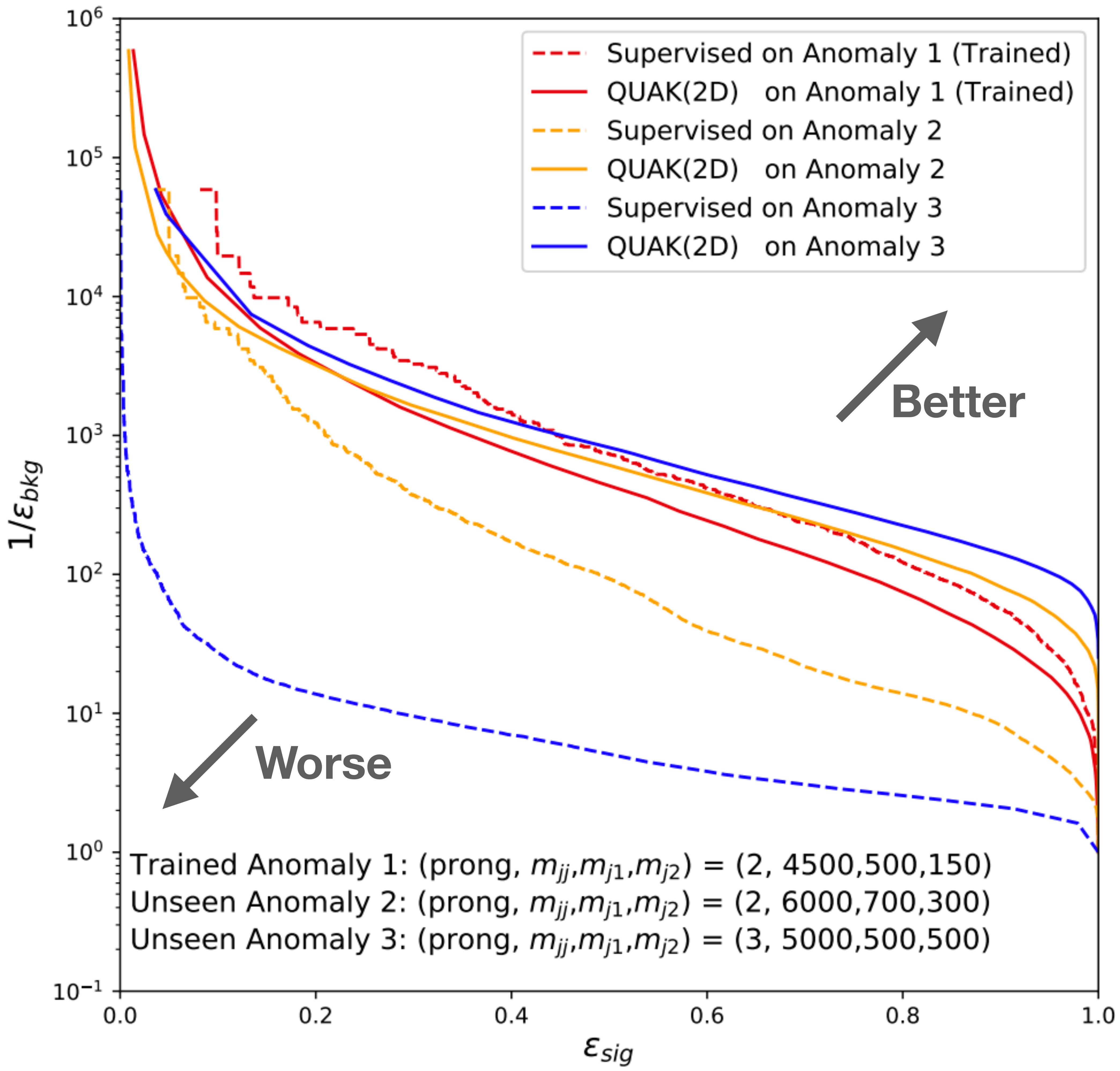


QUAK

- QUAK can perform similarly to supervised methods when given correct signal
- QUAK greatly outperforms supervised methods when signal hypothesis is wrong







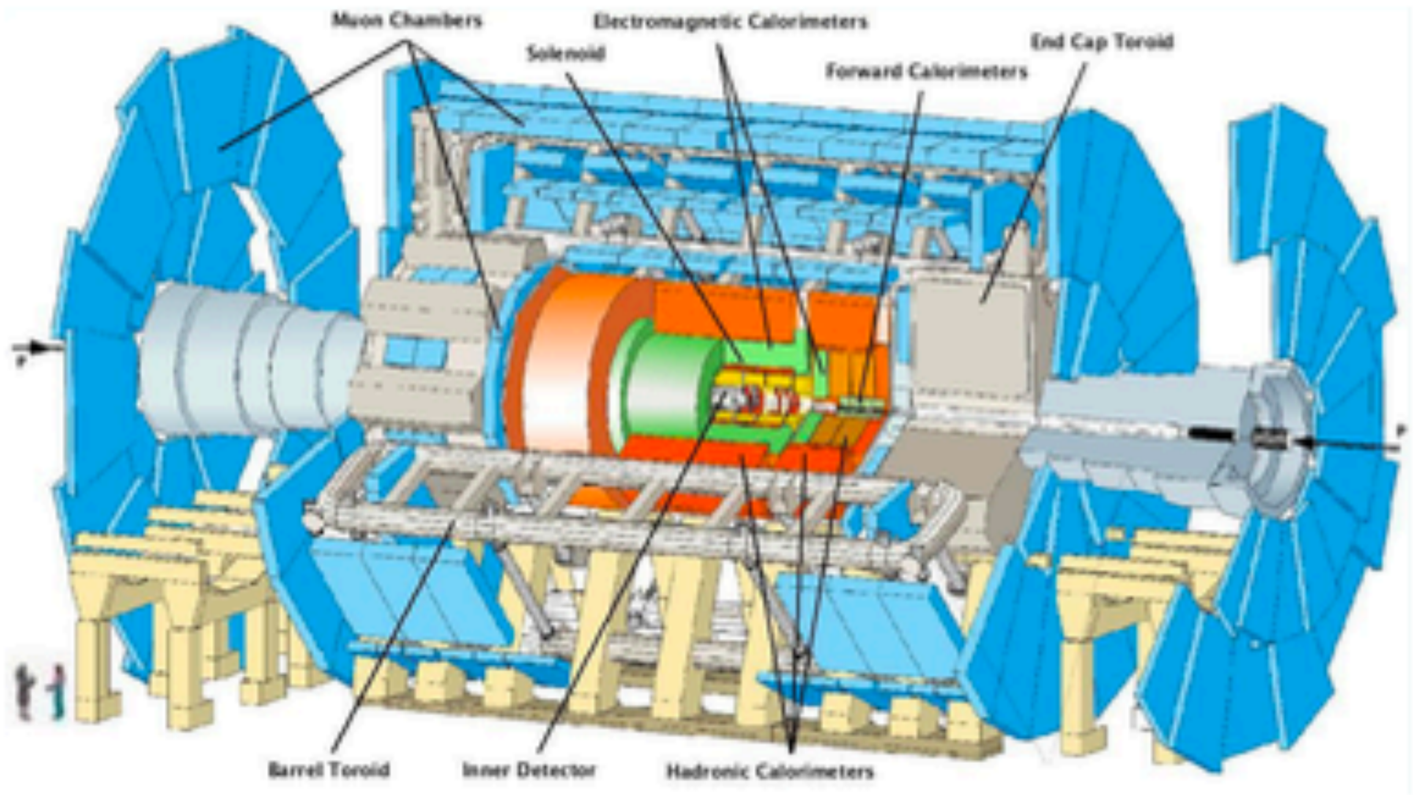
Conclusions

- I hope you learned a little bit about LHC research and anomaly detection
- Many more AD methods I didn't have time to mention
  - TNT, CATHODE, SALAD, CURTAINS, …
- Very active area of research at the LHC!
- Recent AD sessions at ML4Jets2024 conference give a nice flavor of cutting edge ideas [1][2]

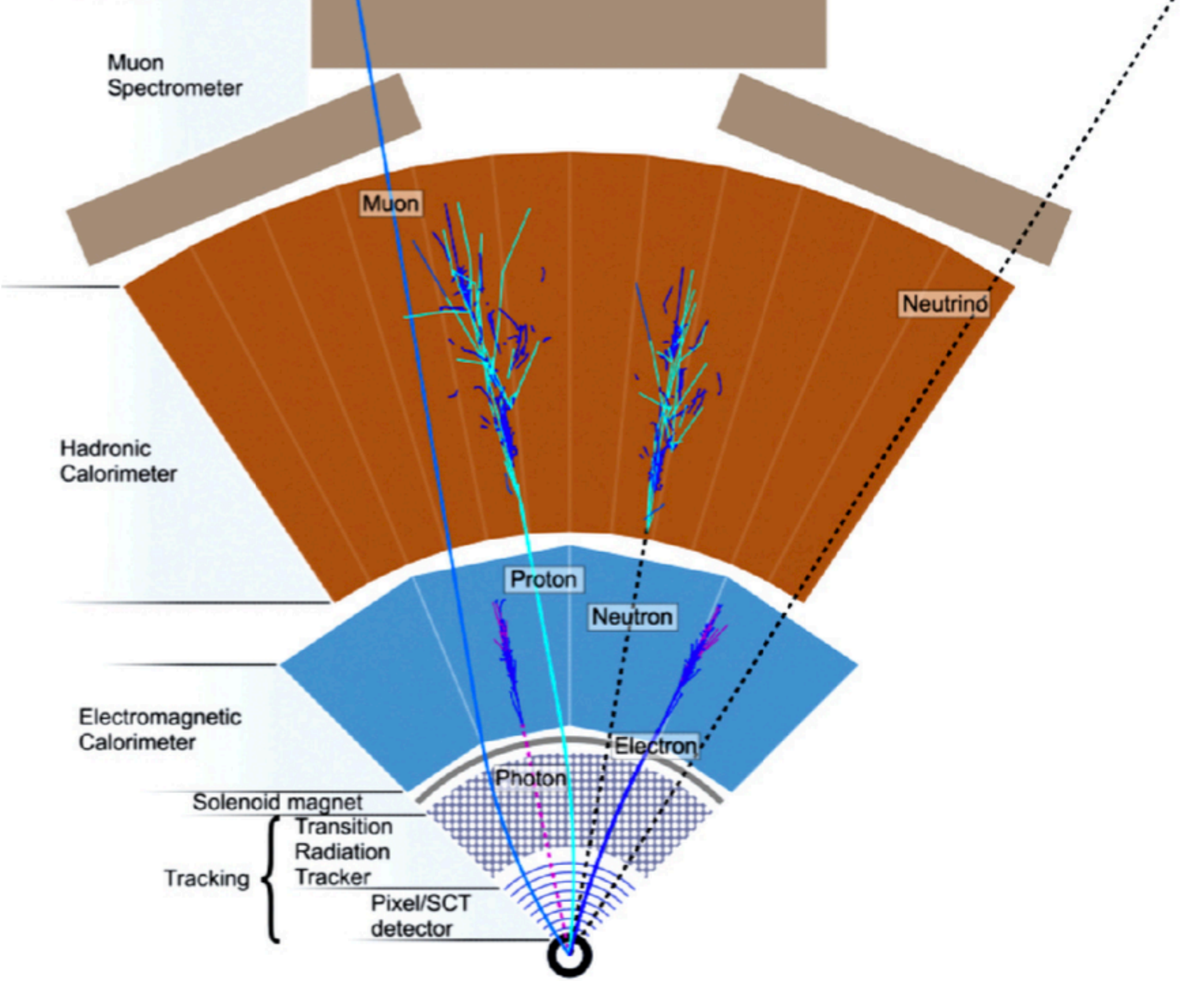
[1] https://indico.cern.ch/event/1386125/timetable/?view=standard#b-587345-anomaly-detection
[2] https://indico.cern.ch/event/1386125/timetable/?view=standard#b-587346-anomaly-detection

BACKUP

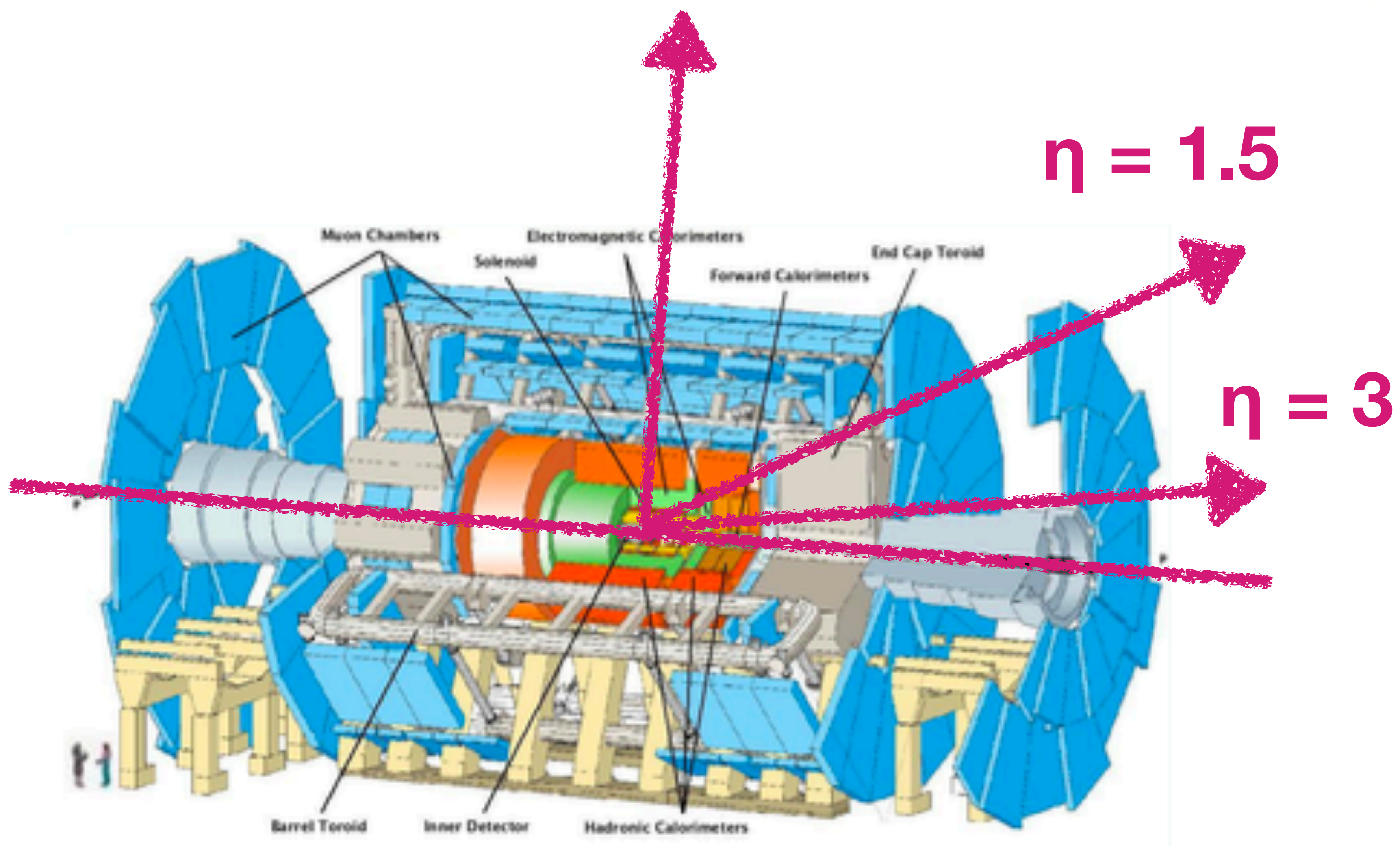
ATLAS Slice

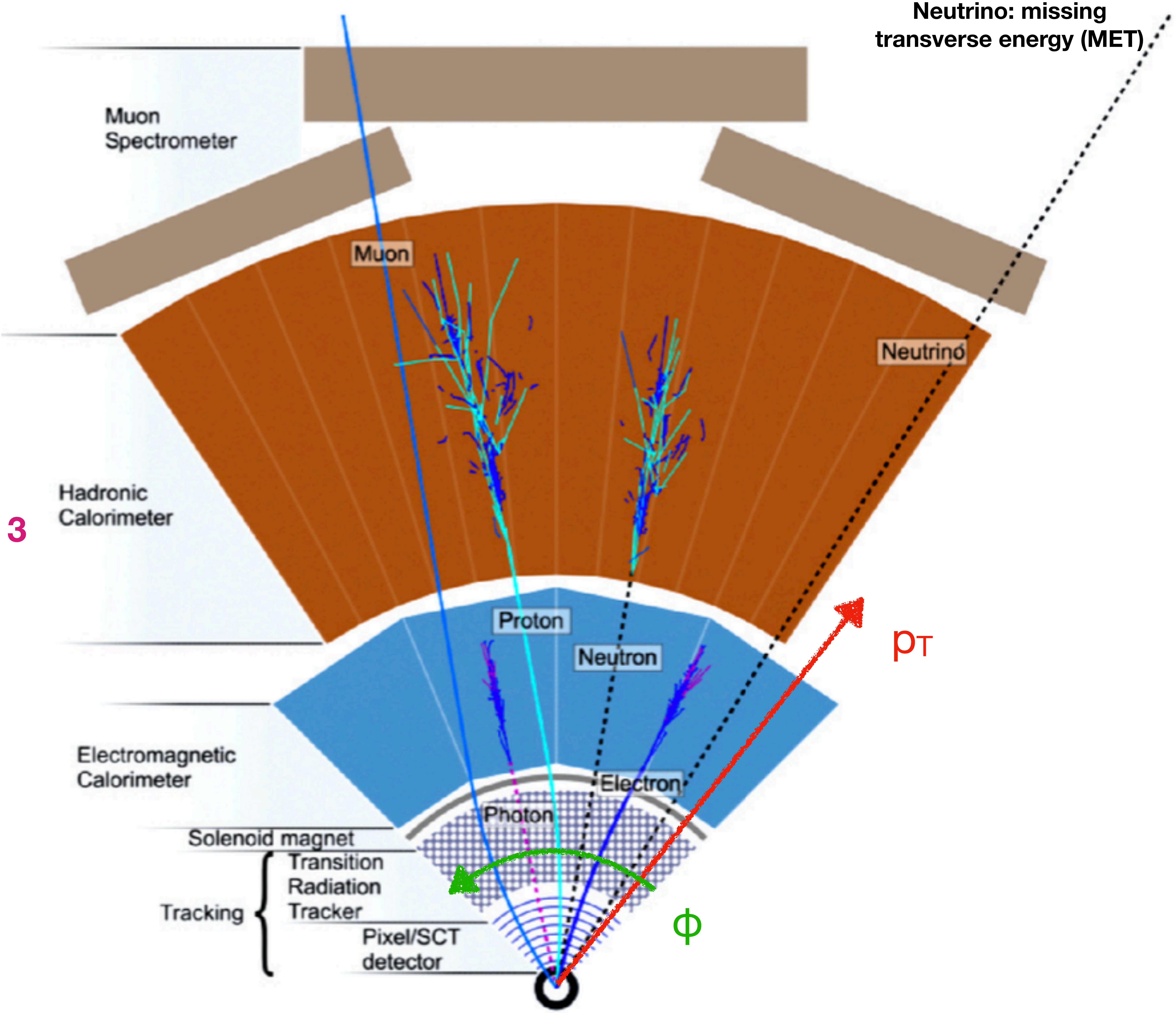


Neutrino: missing transverse energy (MET)



ATLAS Slice

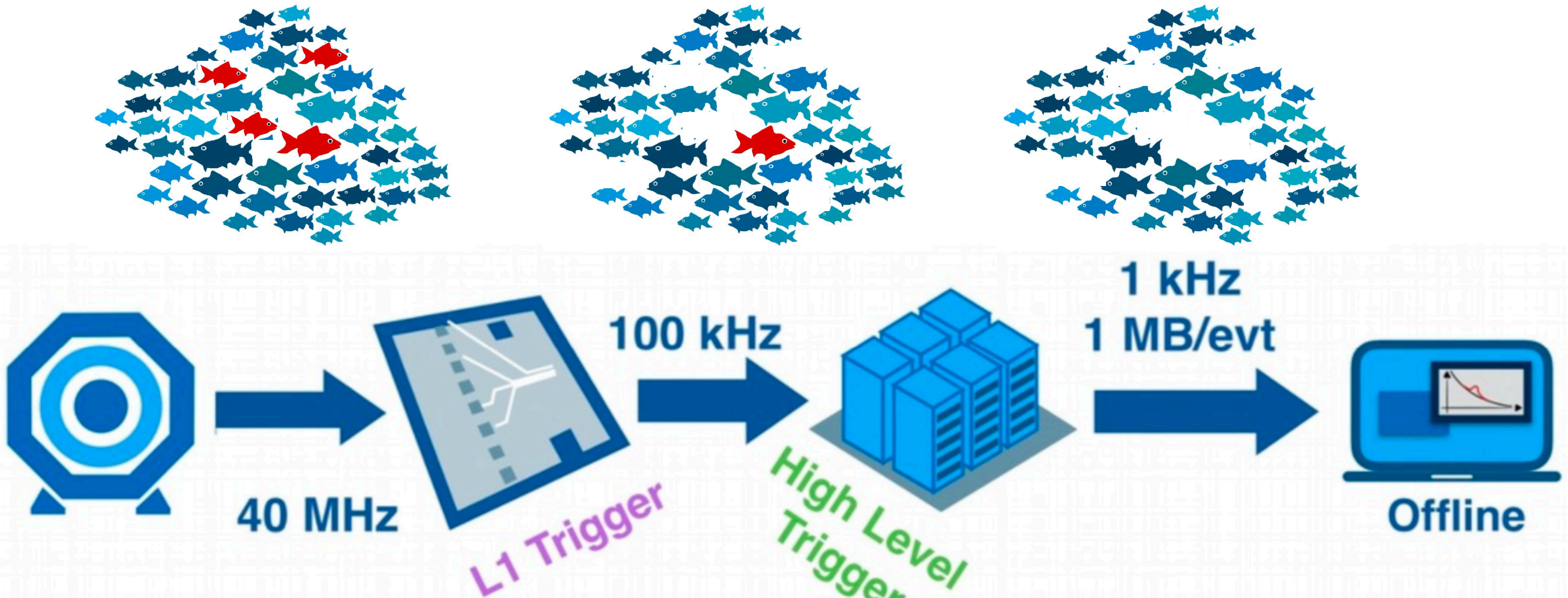




$p_4 = (p_T, \eta, \phi, E)$

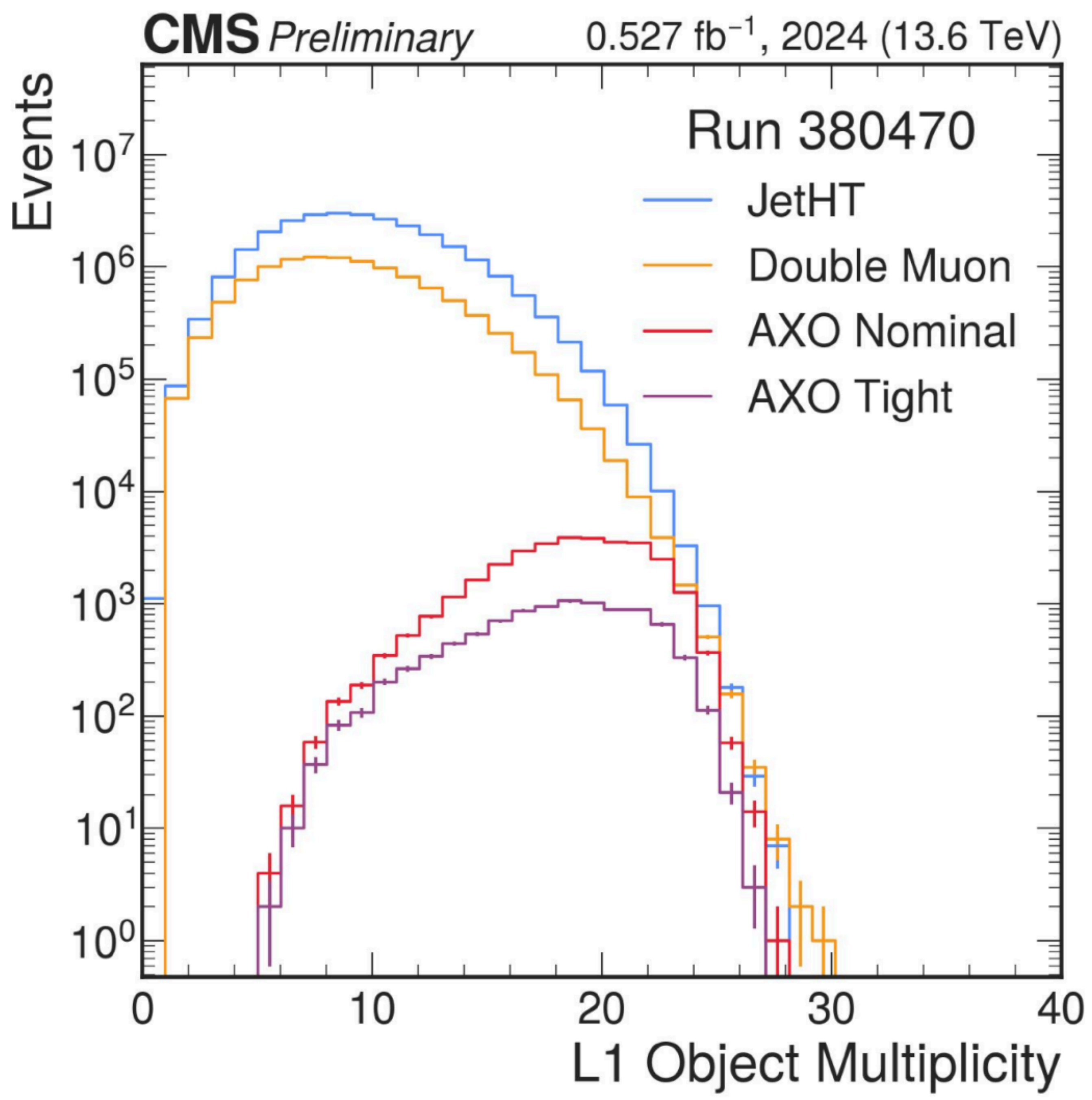
L1 Trigger AD

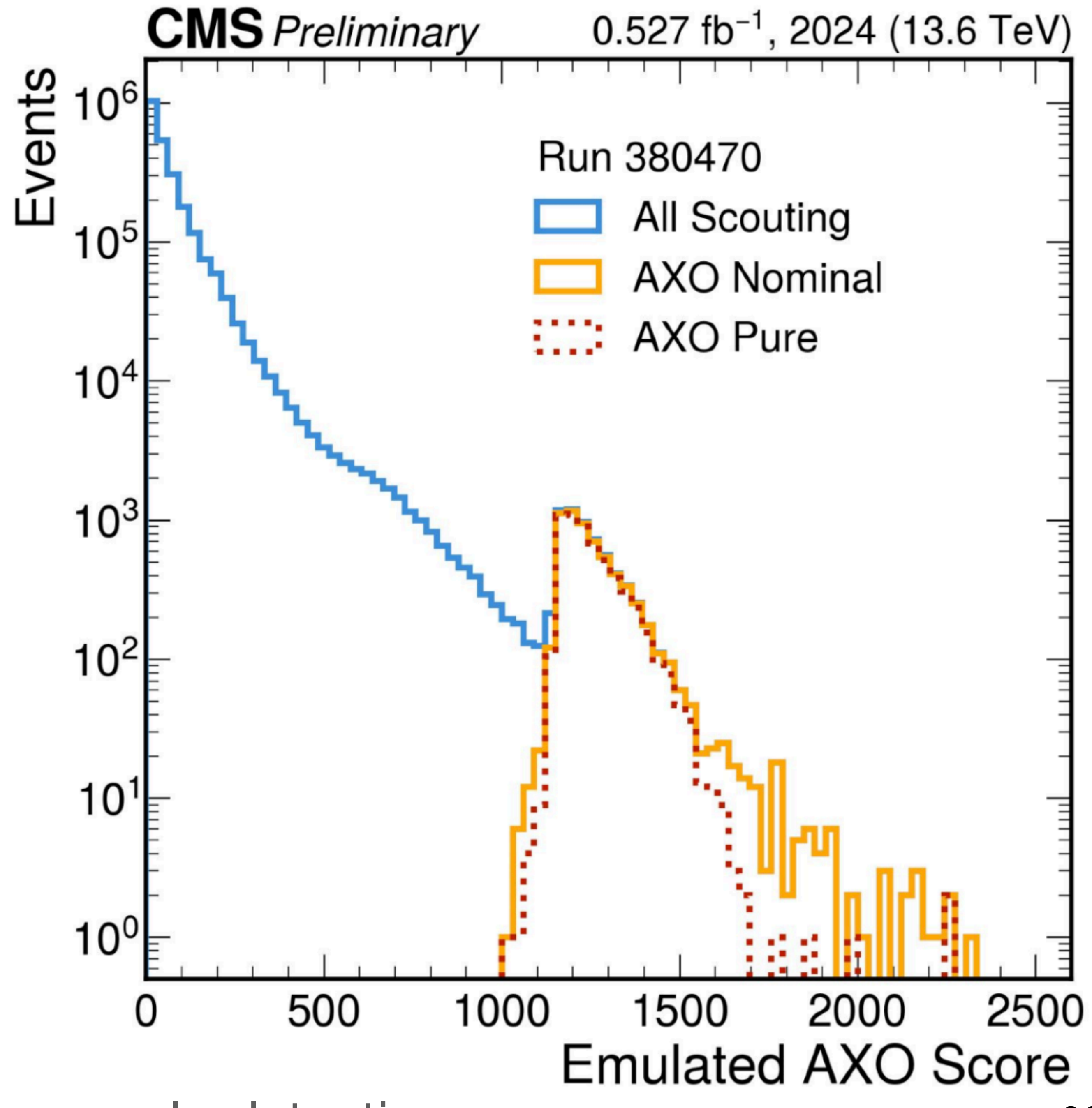
- Depending on anomaly, we could have none left in recorded data
- Low-latency ML is the only option! (eg. autoencoders)

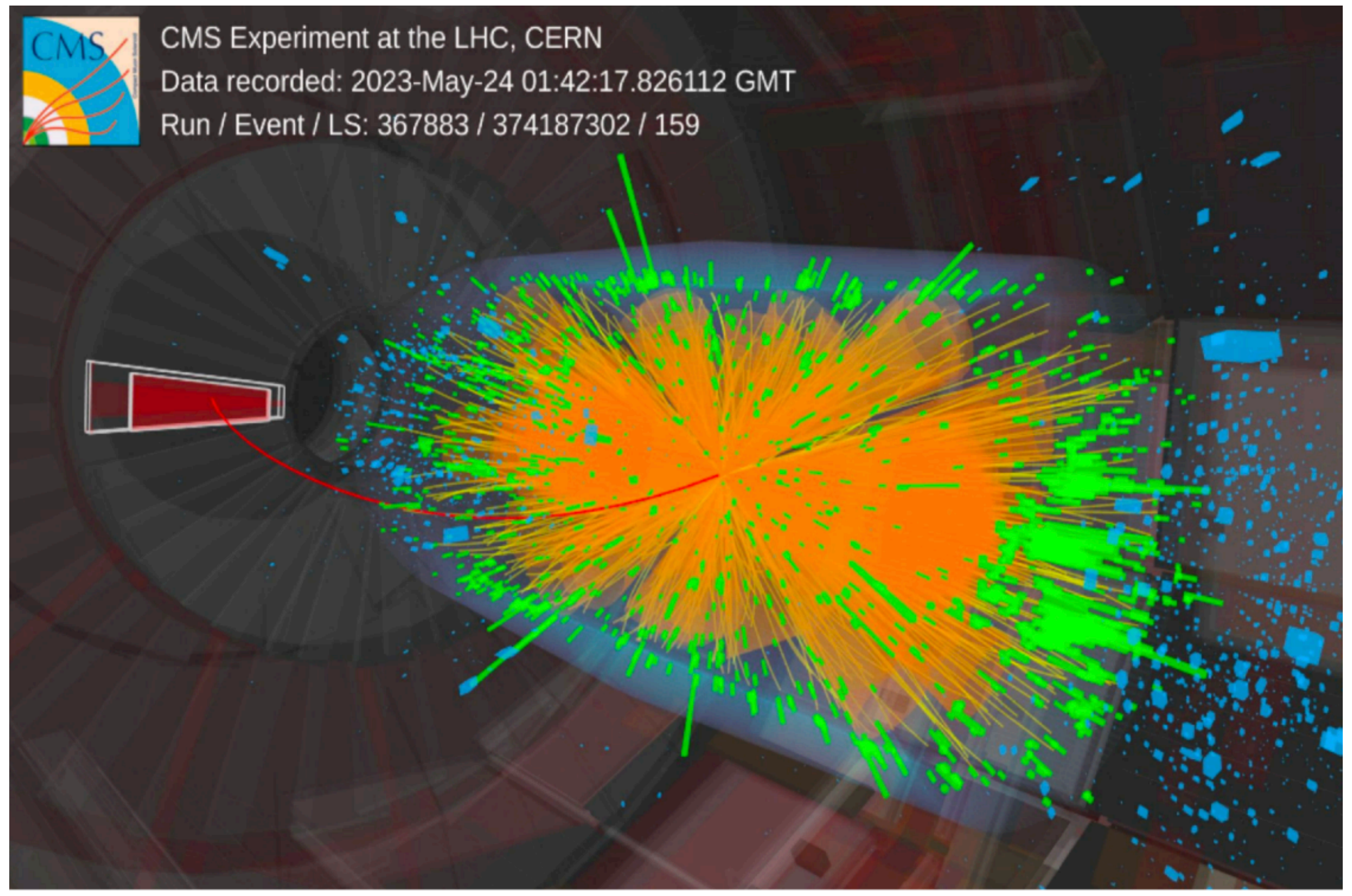


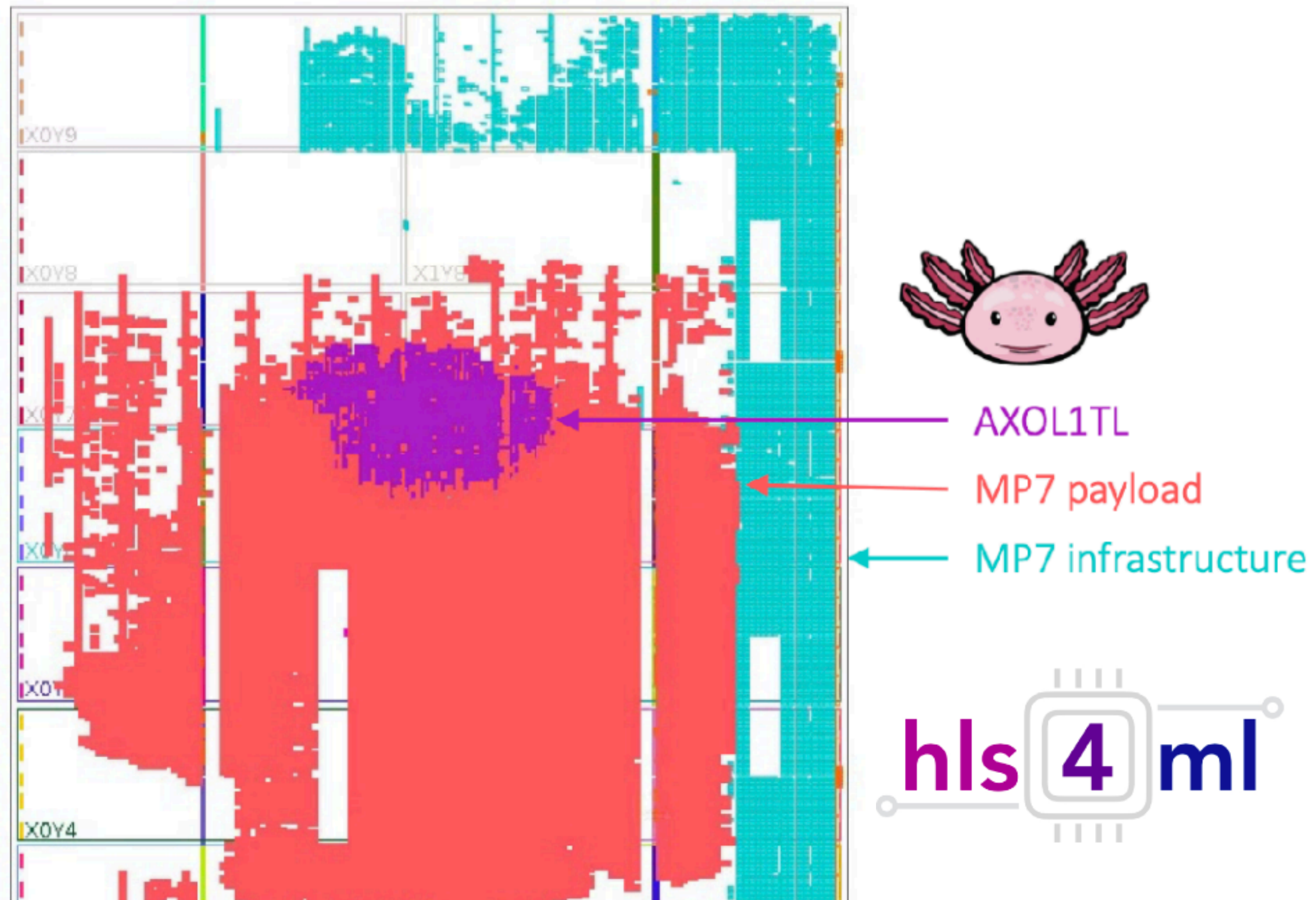
L1 Trigger AD

- CMS has already deployed multiple AD algorithms in trigger
  - AXOL1TL [CMS DP-2023/079, CMS DP-2024/059] & CICADA [CMS DP-2023/086] (see Noah's talk later [1])
- Currently collecting interesting events that would have been missed
  - Network preferentially identifies large multiplicity events, potentially large gains in new physics acceptance
- Development ongoing in ATLAS as well









hls4ml

[1] https://indico.cern.ch/event/1387540/timetable/?view=standard#66-realtime-anomaly-detection