Data Science at LHCb

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Tim Head - EPFL

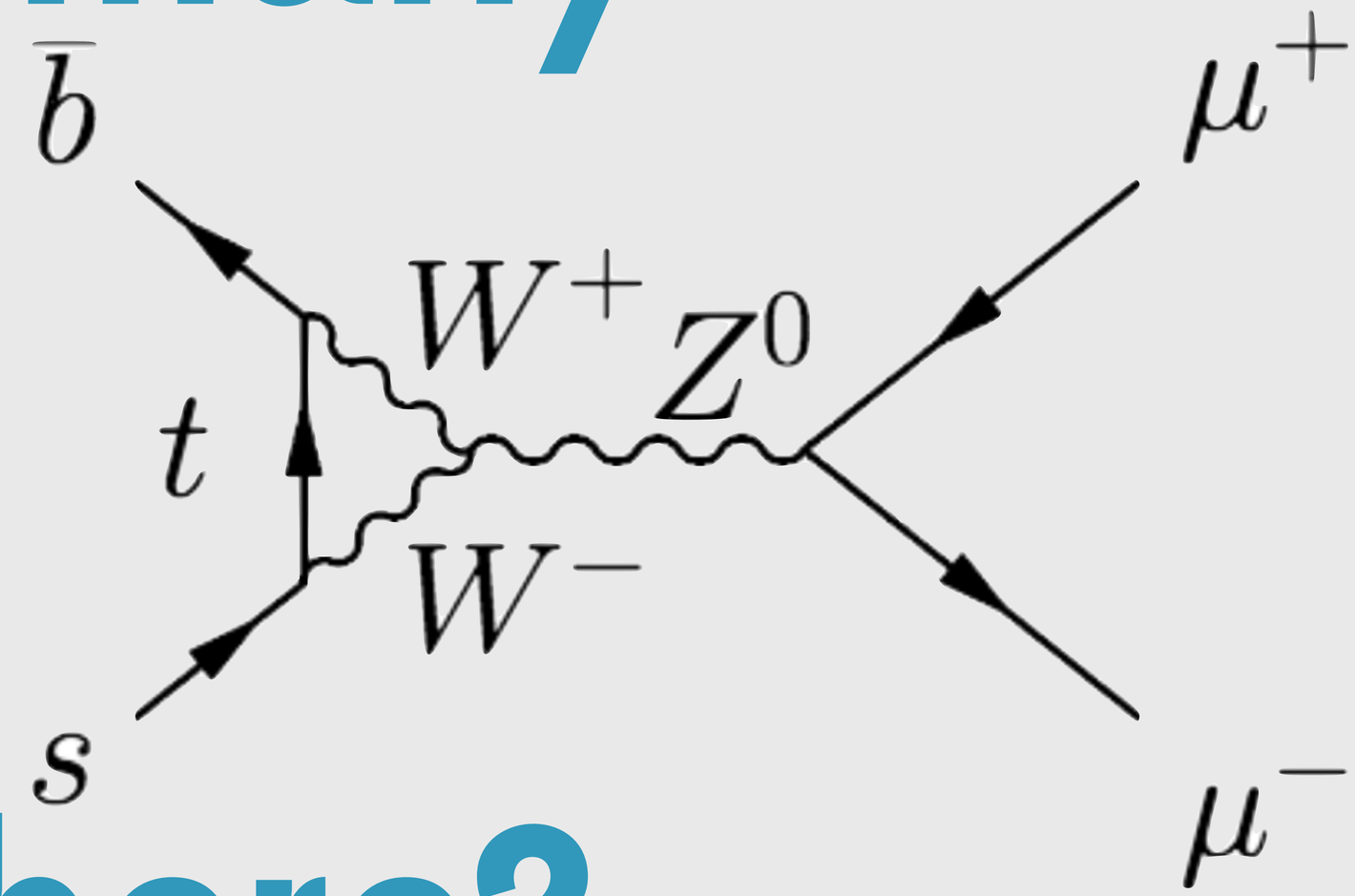
UZH
February 2016

The Large Hadron Collider

This is the right time: Dark Matter, stuff vs anti-stuff

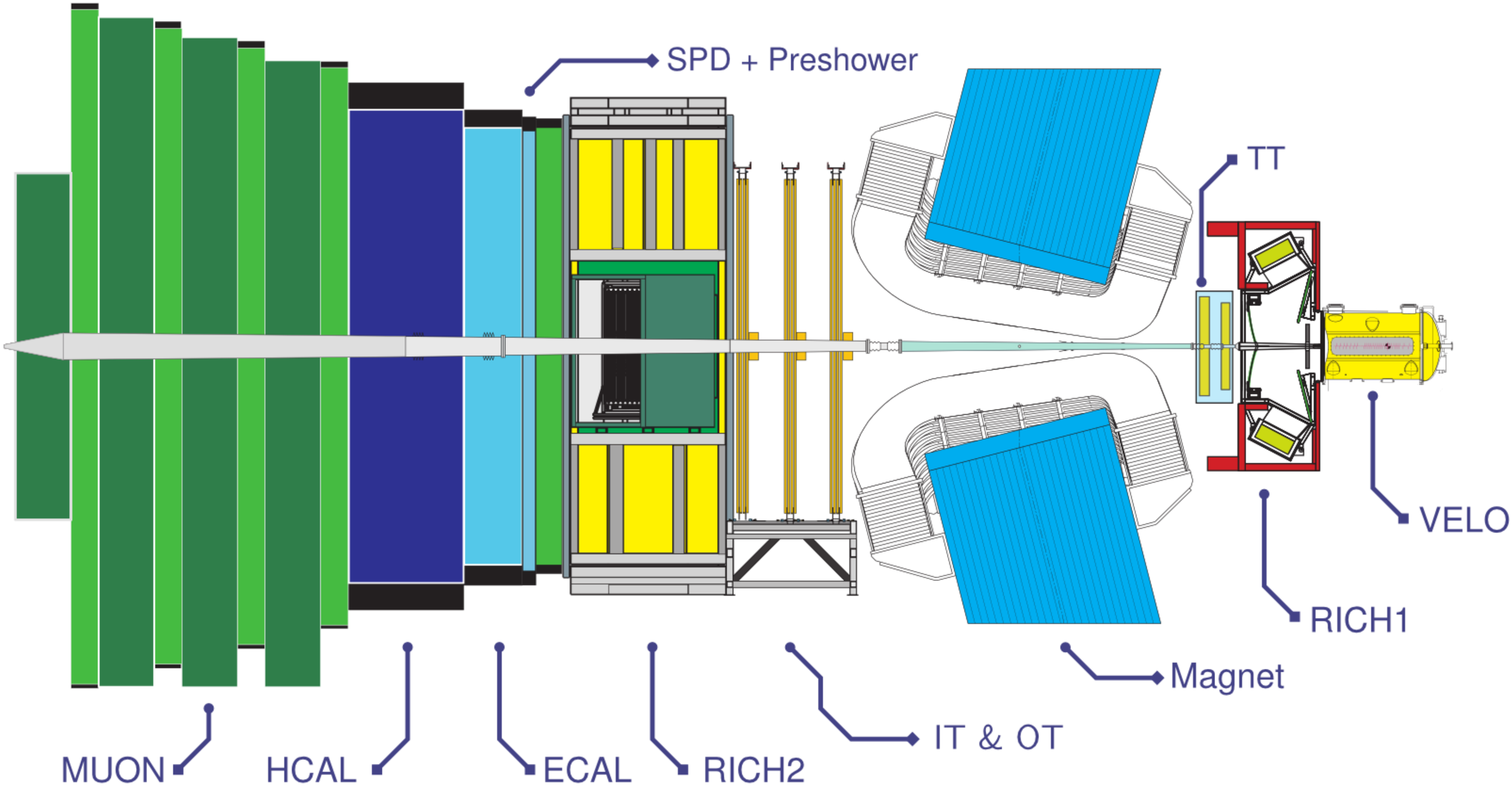
New Physics 101

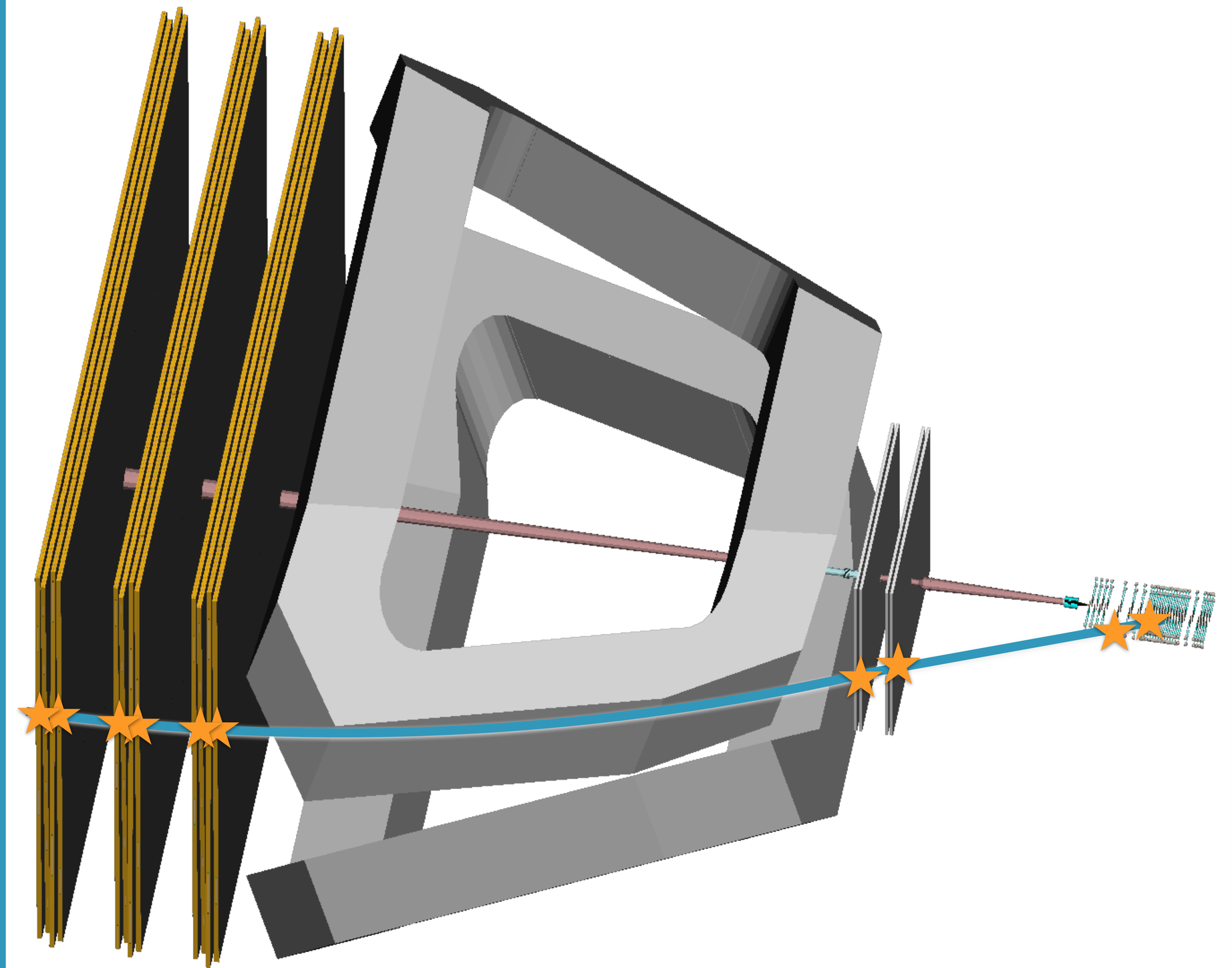
How many



are there?

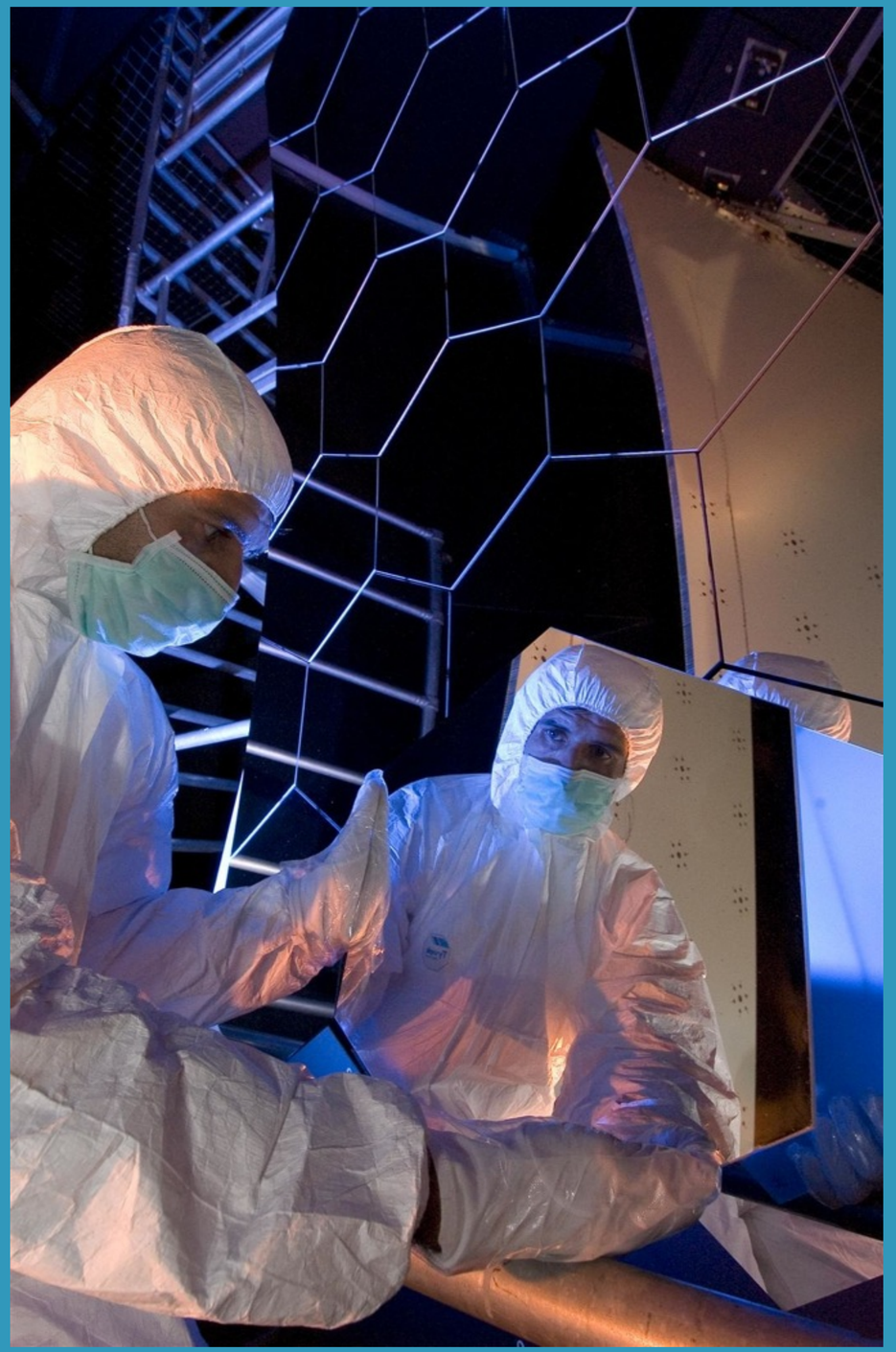

UXB

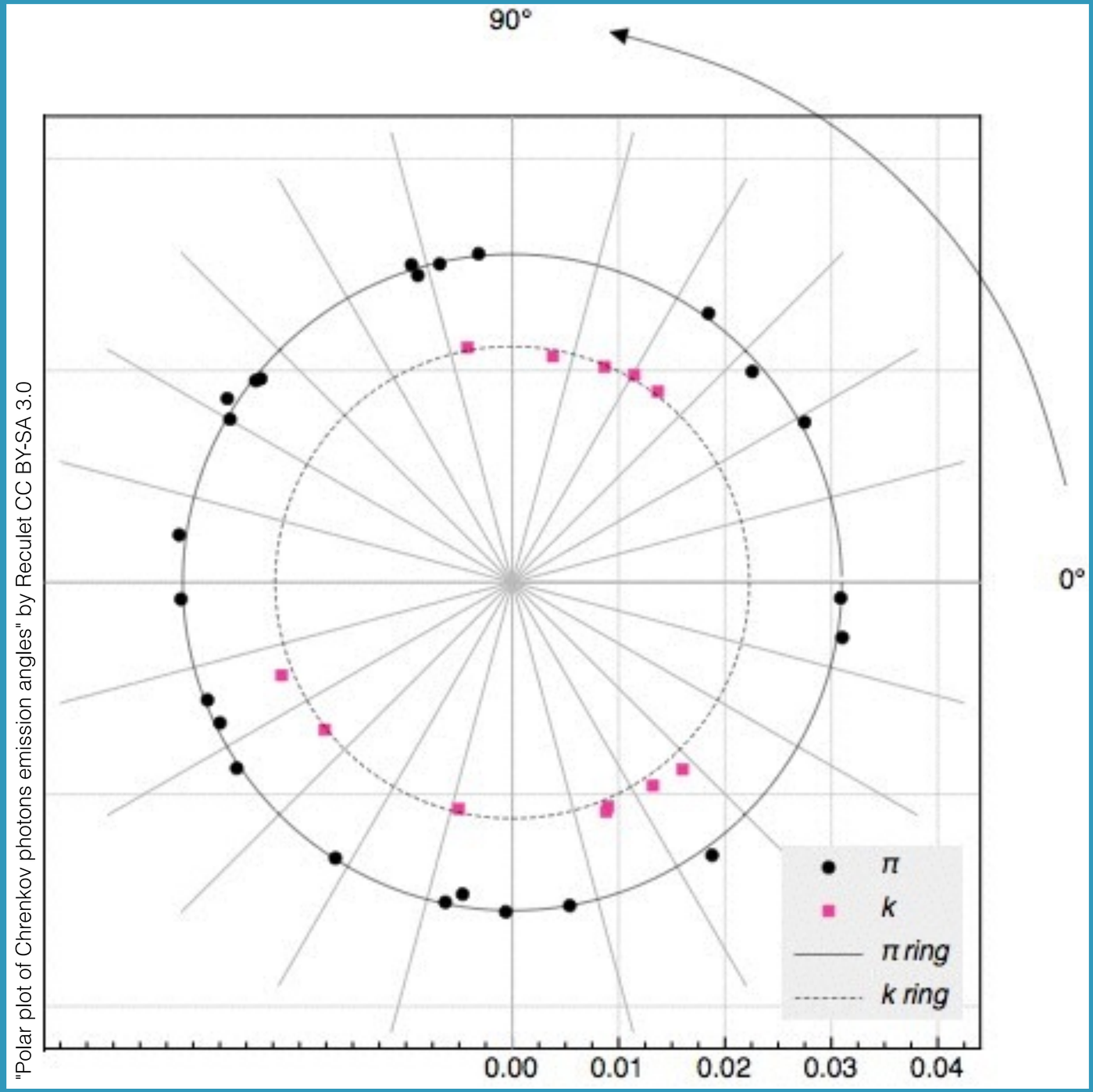

SPD + Preshower
TT
VELO
RICH1
Magnet
IT & OT
RICH2
ECAL
HCAL
MUON



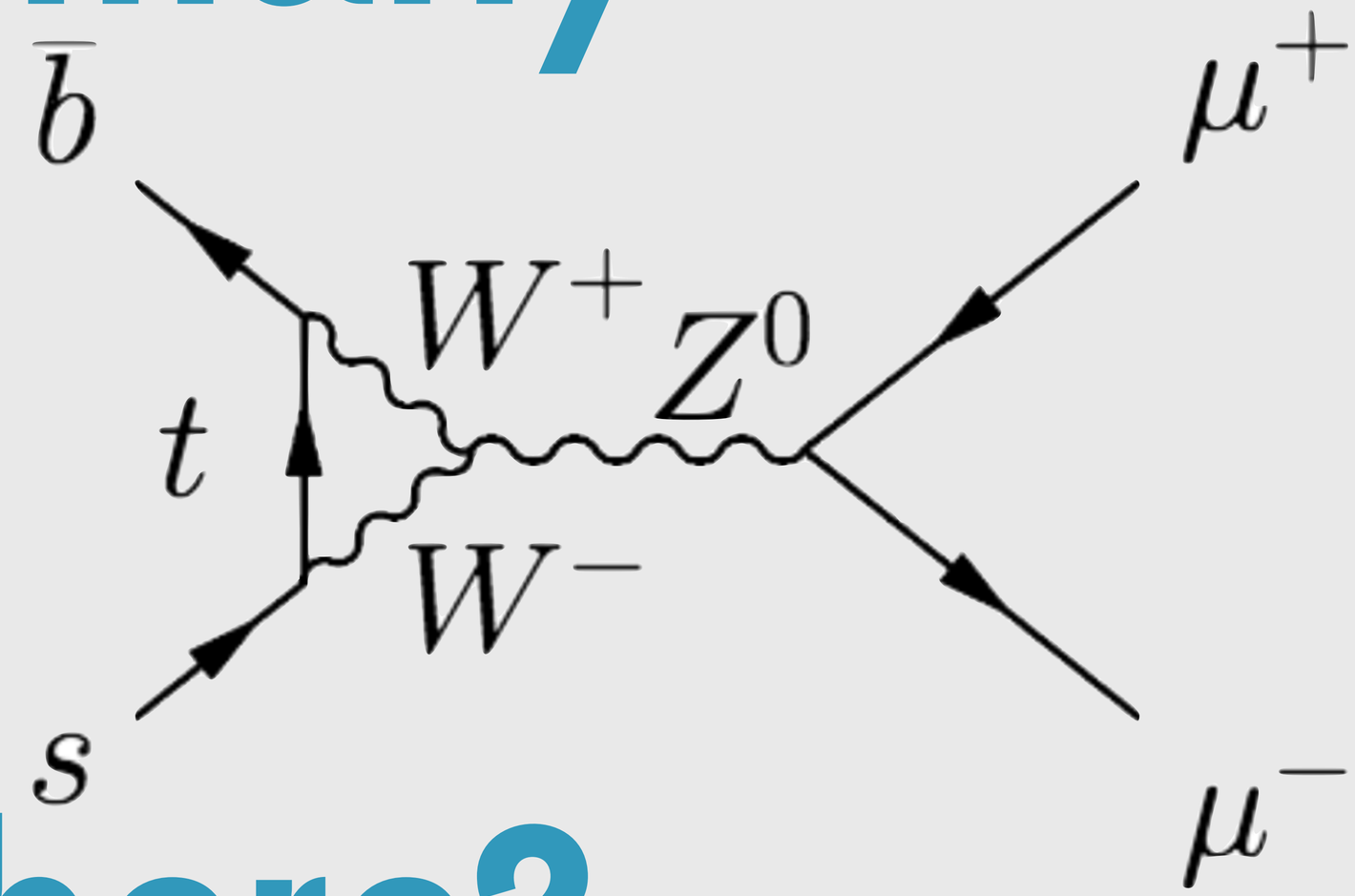
Who ya gonna call?

Cherenkov - faster than the speed of light





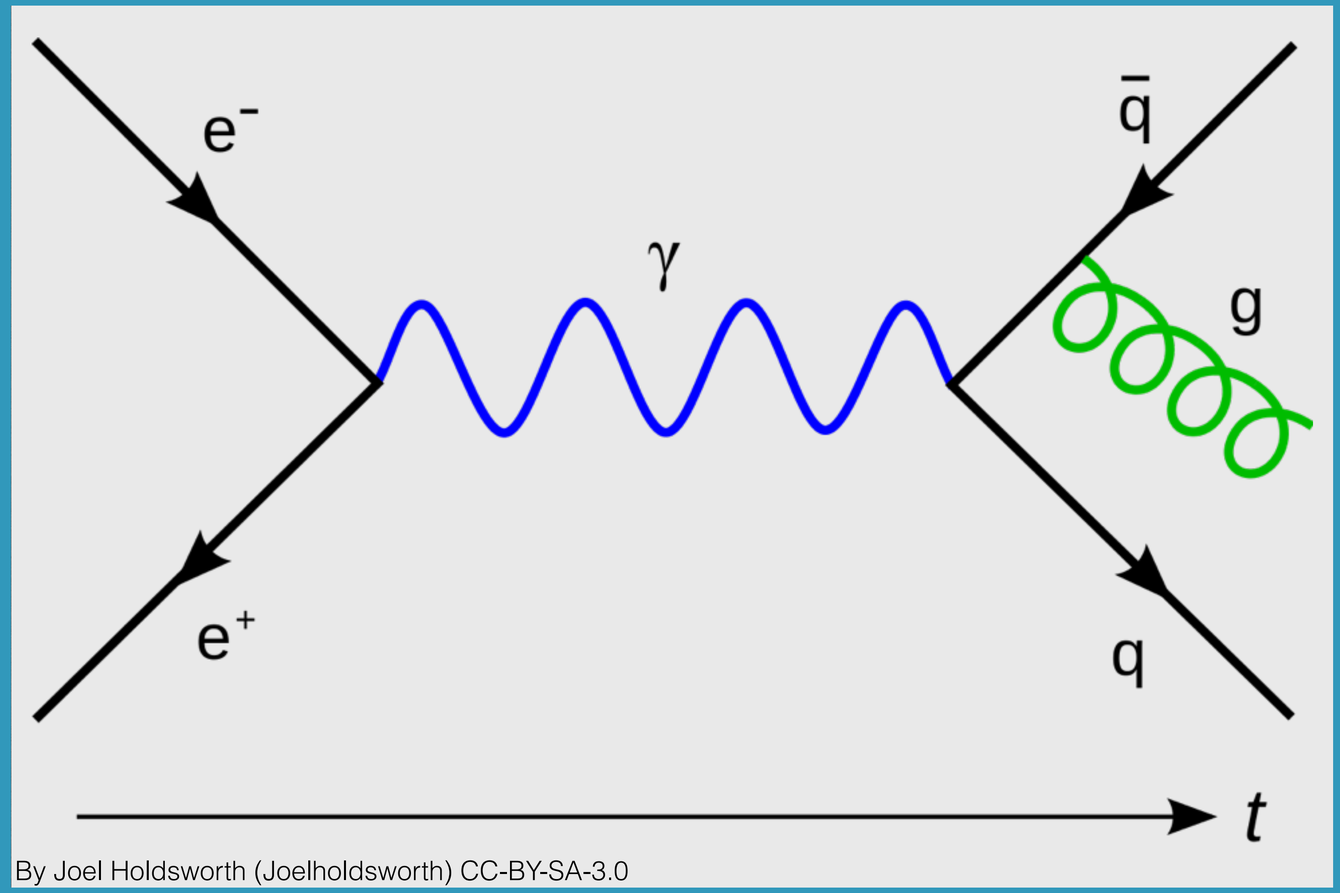
How many



are there?

What else does the LHC produce?

Simulation



Simulate all the things

Simulate all the things

#hard

What else does the LHC produce?

Discard events as early as possible

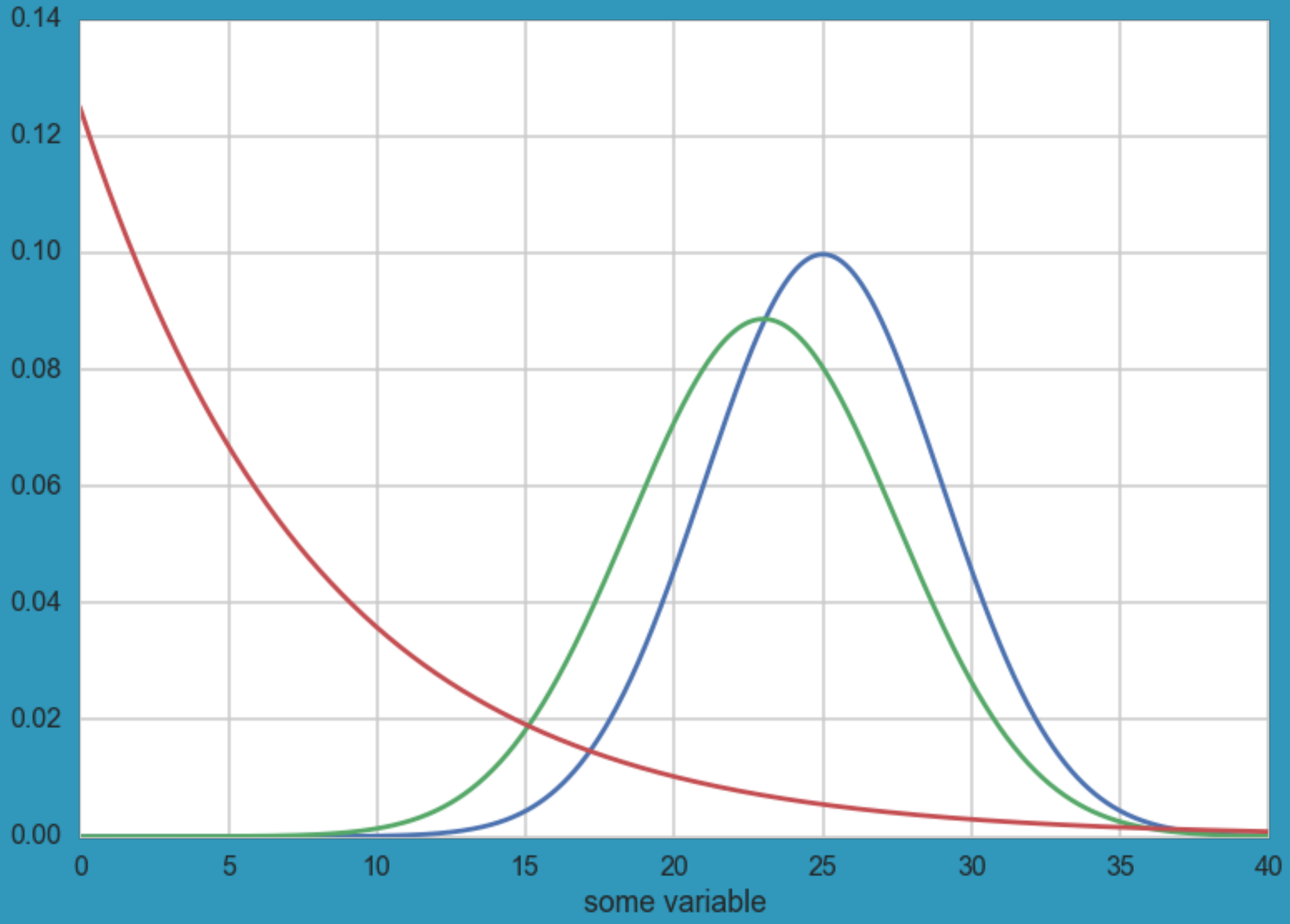
Real Time Stream Processing

“Real” Time Stream Processing

The Topological Trigger

Fast
Robust
General purpose

Gligorov, Wiliams, arXiv:1210.6861


0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
25
30
35
40
some variable

“I’ve mostly figured out what the BDT does. In easy regions it does easy to understand things and in hard regions it does hard to understand things.”

The Topological Trigger

Old

Fast

Robust

General purpose

Gligorov, Wiliams, arXiv:1210.6861

The Final Selection

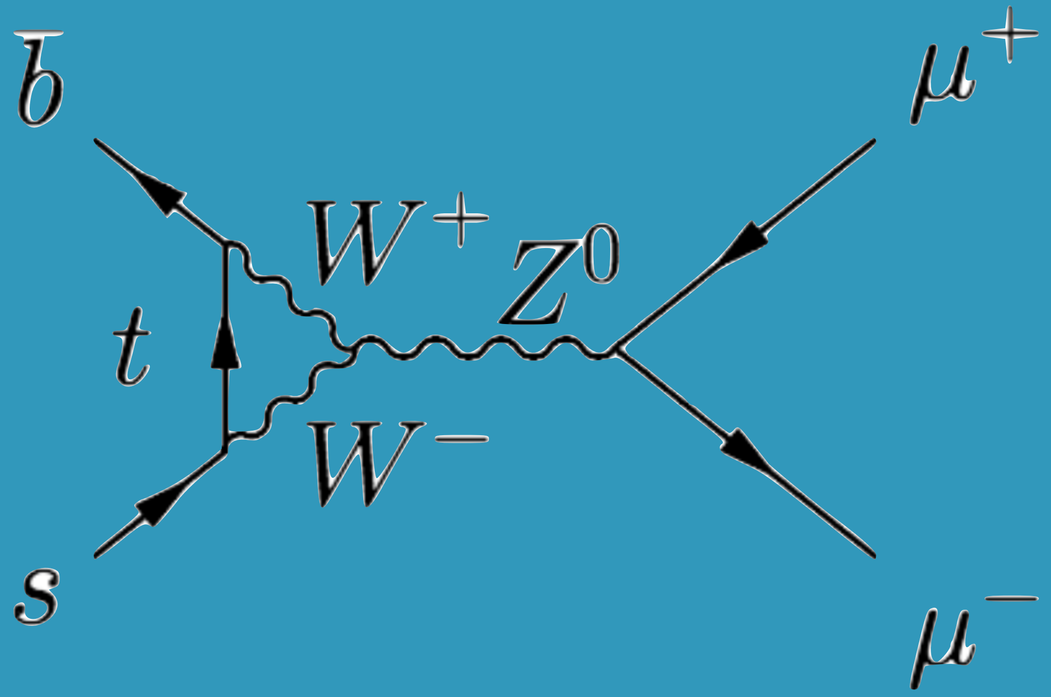
The Final Selection

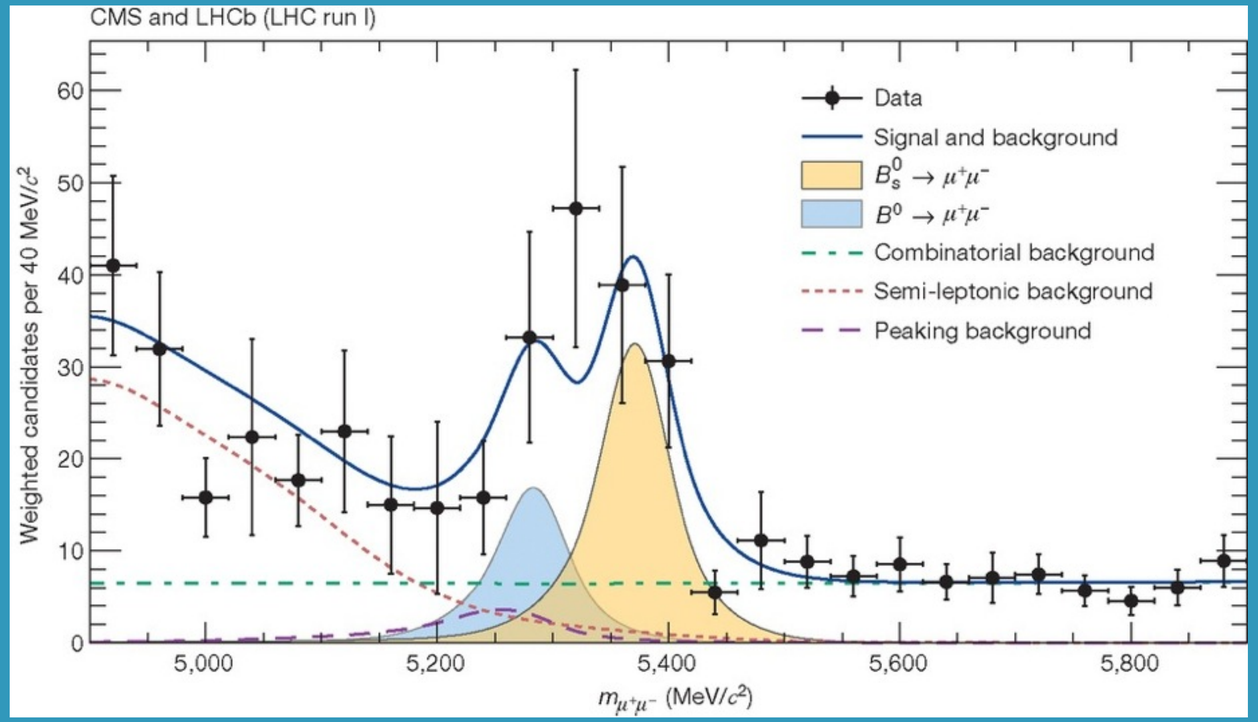
This is where kaggle starts

Power Users

Performance!!

Advanced techniques

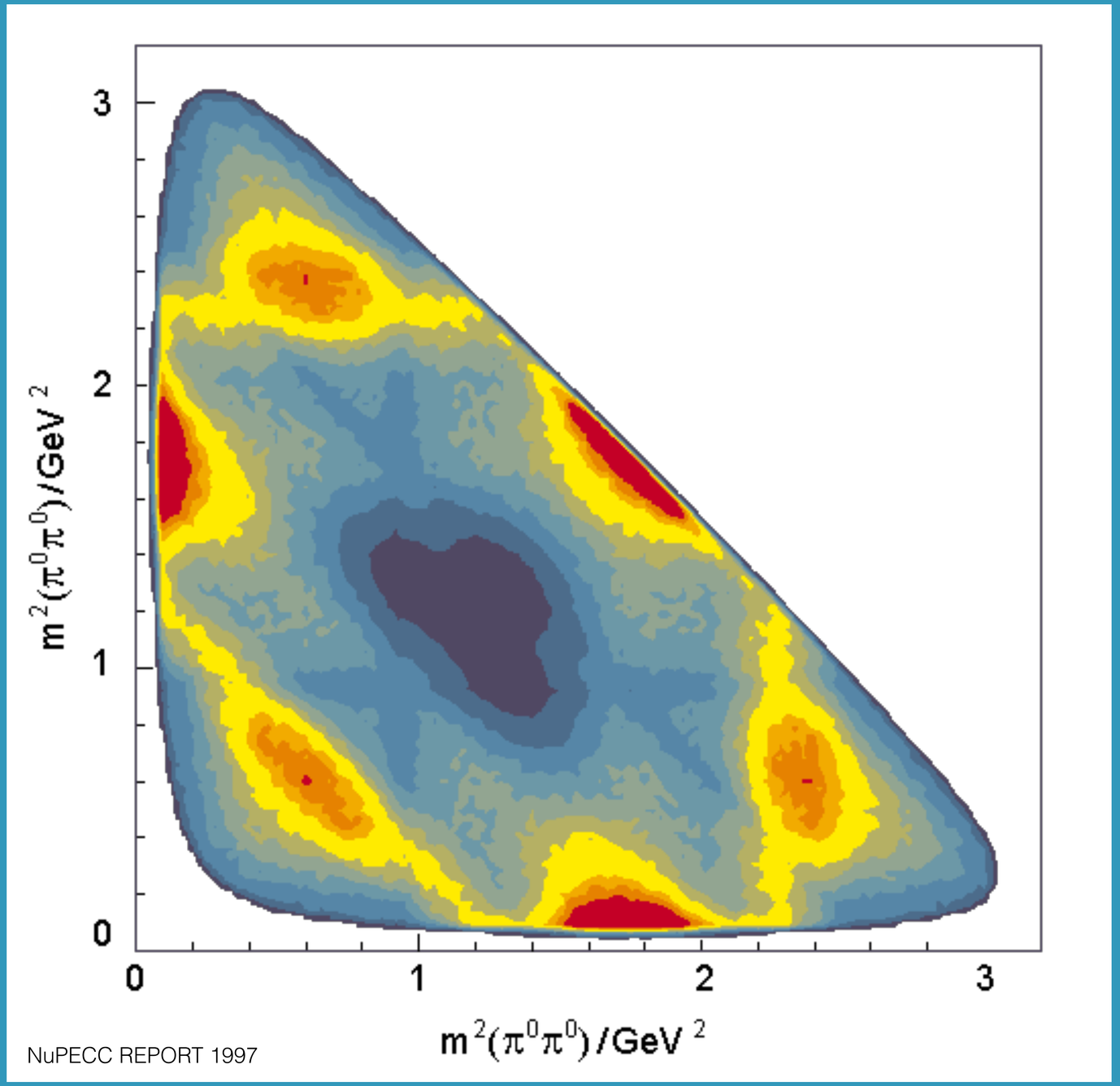




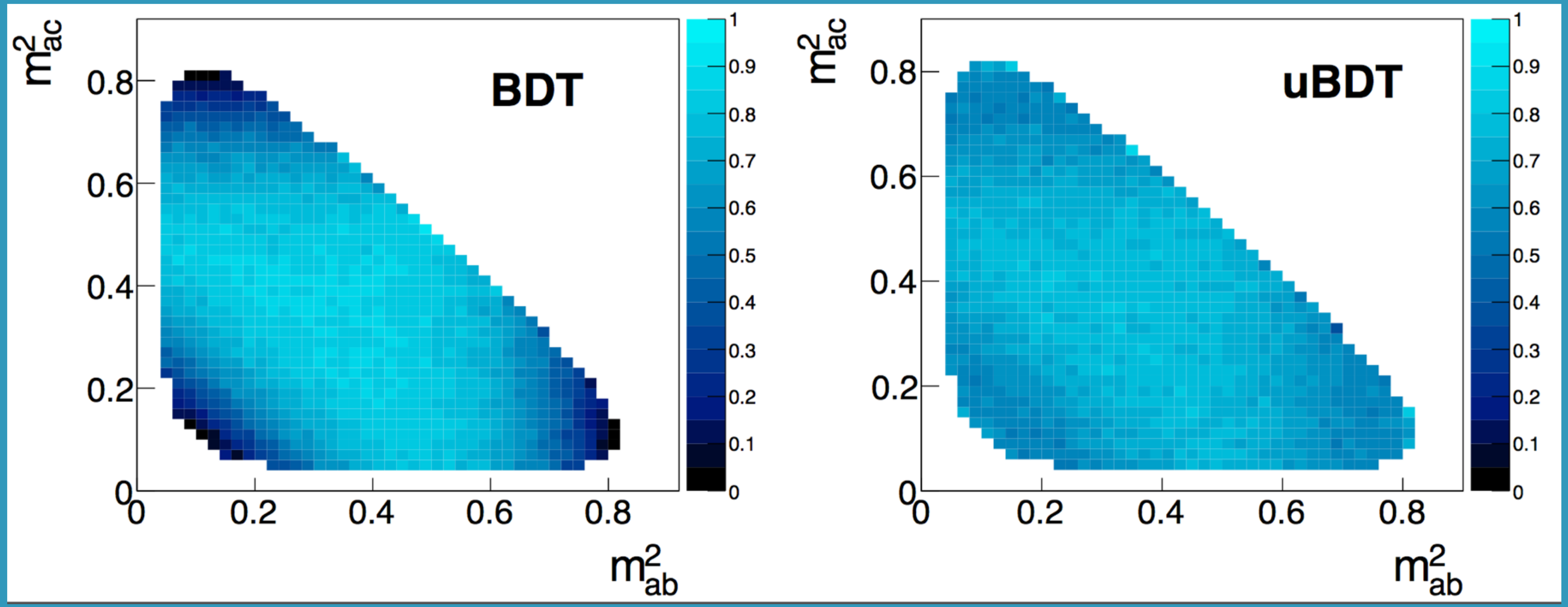
Normal People

Simplicity
What is it doing?

Artisans



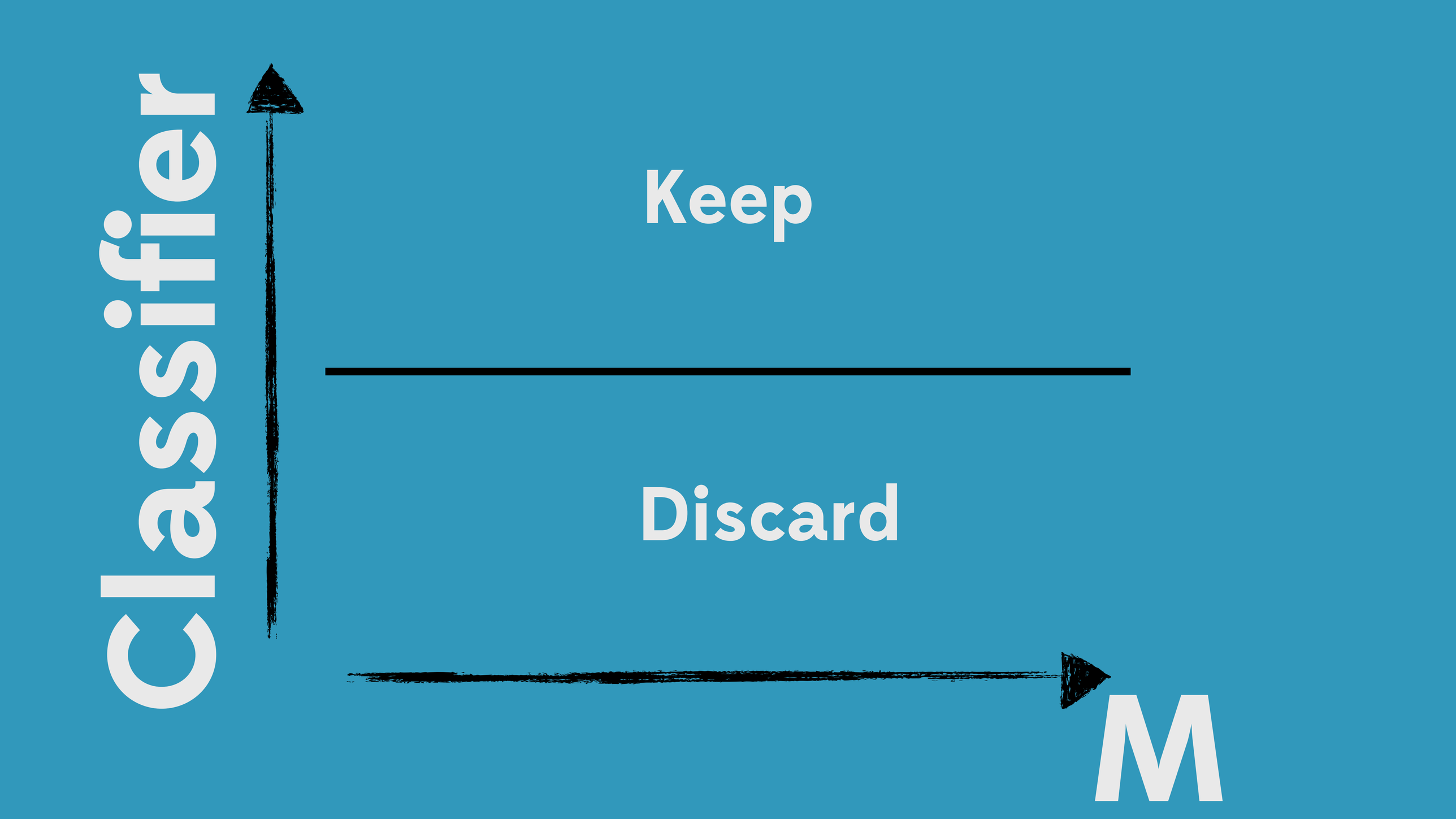
NuPECC REPORT 1997

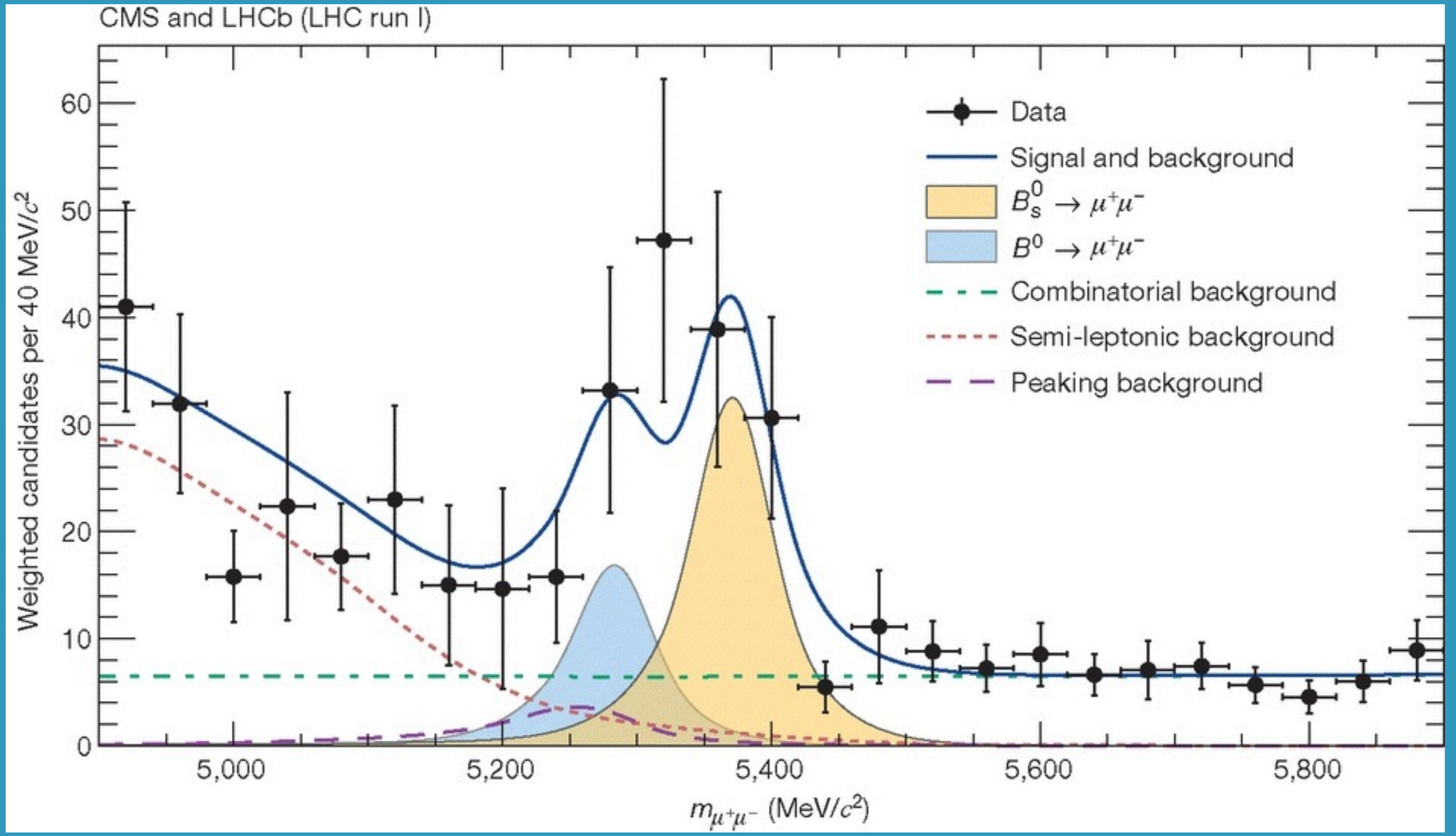


Artisans

Uniformity, or other special loss functions

The Final Selection


Classifier
Keep
Discard
M


CMS and LHCb (LHC run I)
Data
Signal and background
$B_s^0 \to \mu^+\mu^-$
$B^0 \to \mu^+\mu^-$
Combinatorial background
Semi-leptonic background
Peaking background
Weighted candidates per 40 MeV/c^2
0
10
20
30
40
50
60
5,000
5,200
5,400
5,600
5,800
$m_{\mu^+\mu^-}$ (MeV/c^2)

The Known Unknowns

Publish!

CMS and LHCb collaborations, doi:10.1038/nature14474, arXiv:1411.4413

Preempting the Trigger

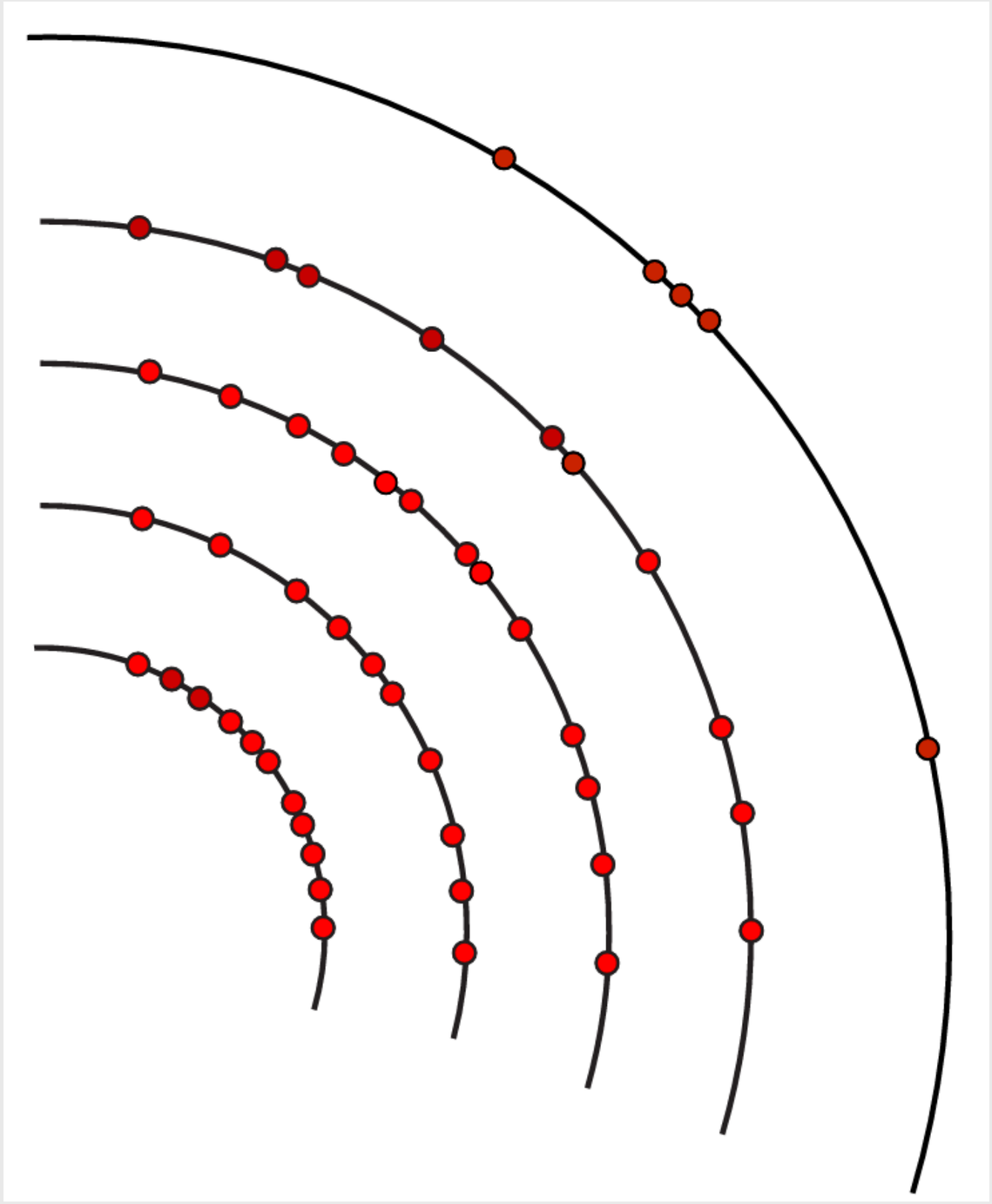
Learn the decision function of the whole LHCb trigger

Preempting the Trigger

No work
is faster than
some work.

Track Finding

Reconstruction
Extracting properties



Generative Models

Simulation is the largest user of CPU time!

Black Box Optimisation

Optimising the whole chain

Details matter

??!

Tim Head

thead@cern.ch

@betatim