

# Data Science at LHCb

**Tim Head, for LHCb**

**November 2015  
Data Science @ LHC 2015**







# The Large Hadron Collider



**This is the right  
time:**

**Dark Matter,  
stuff vs anti-stuff**

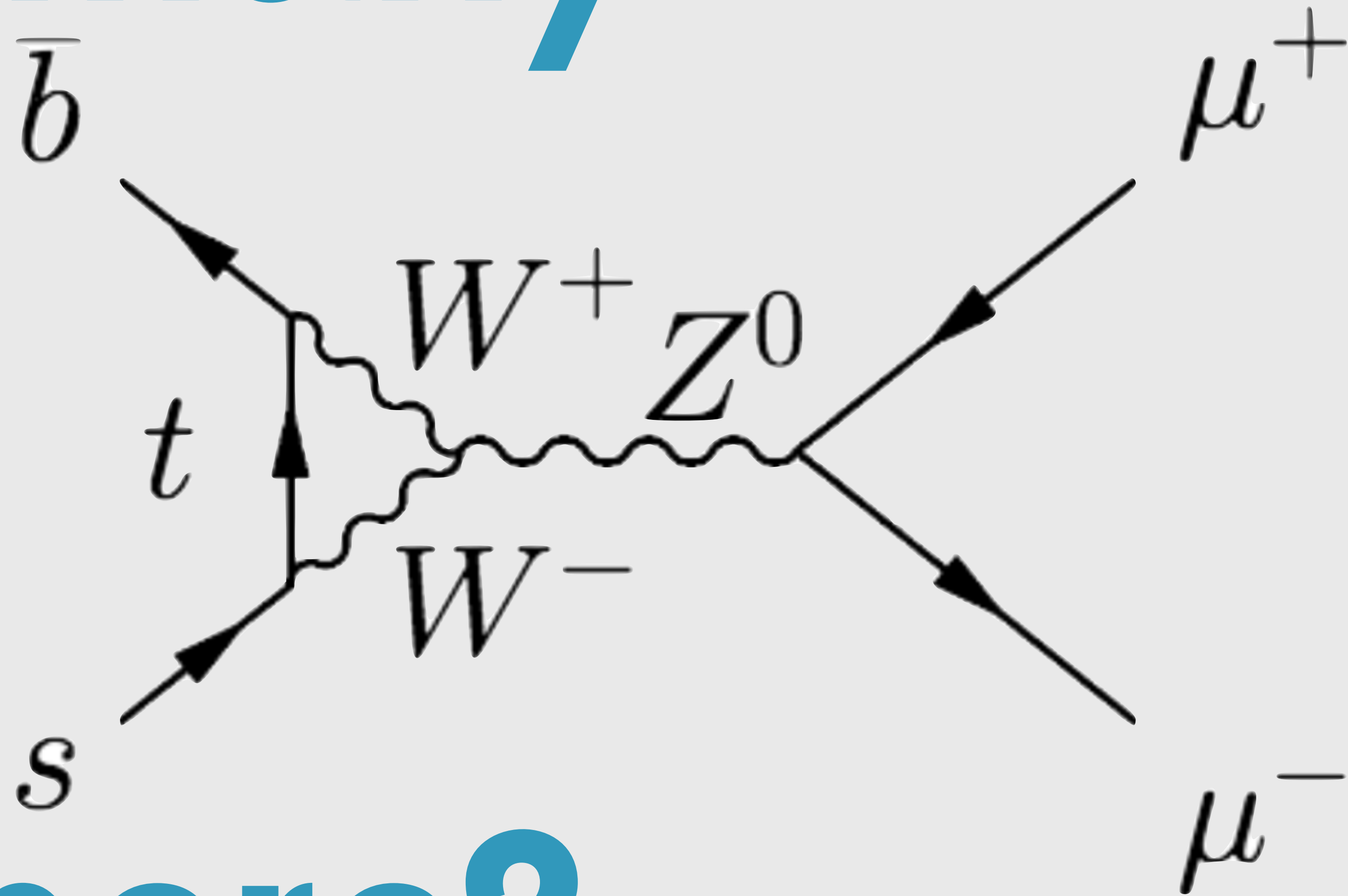
# New Physics 101





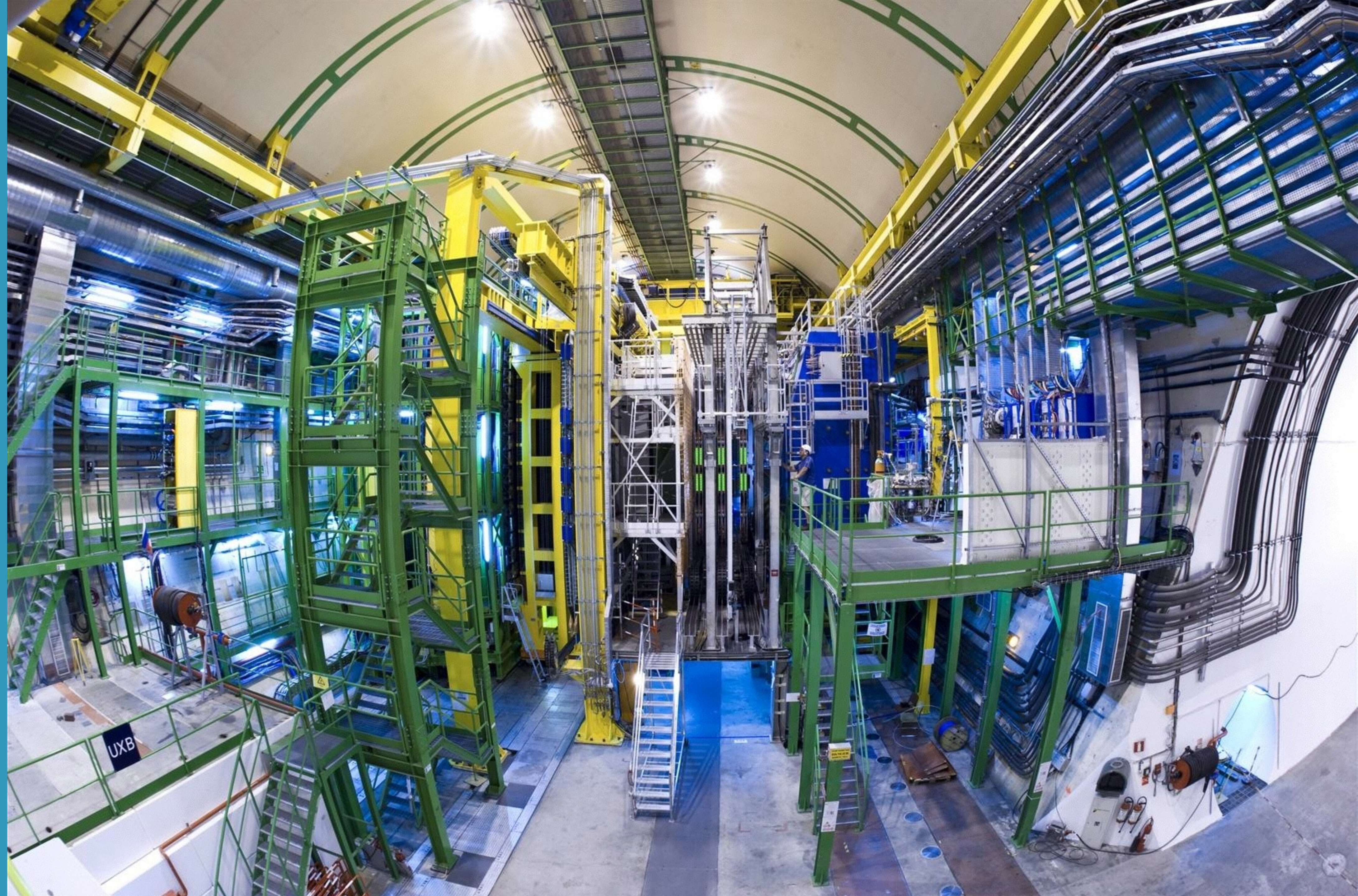


# How many

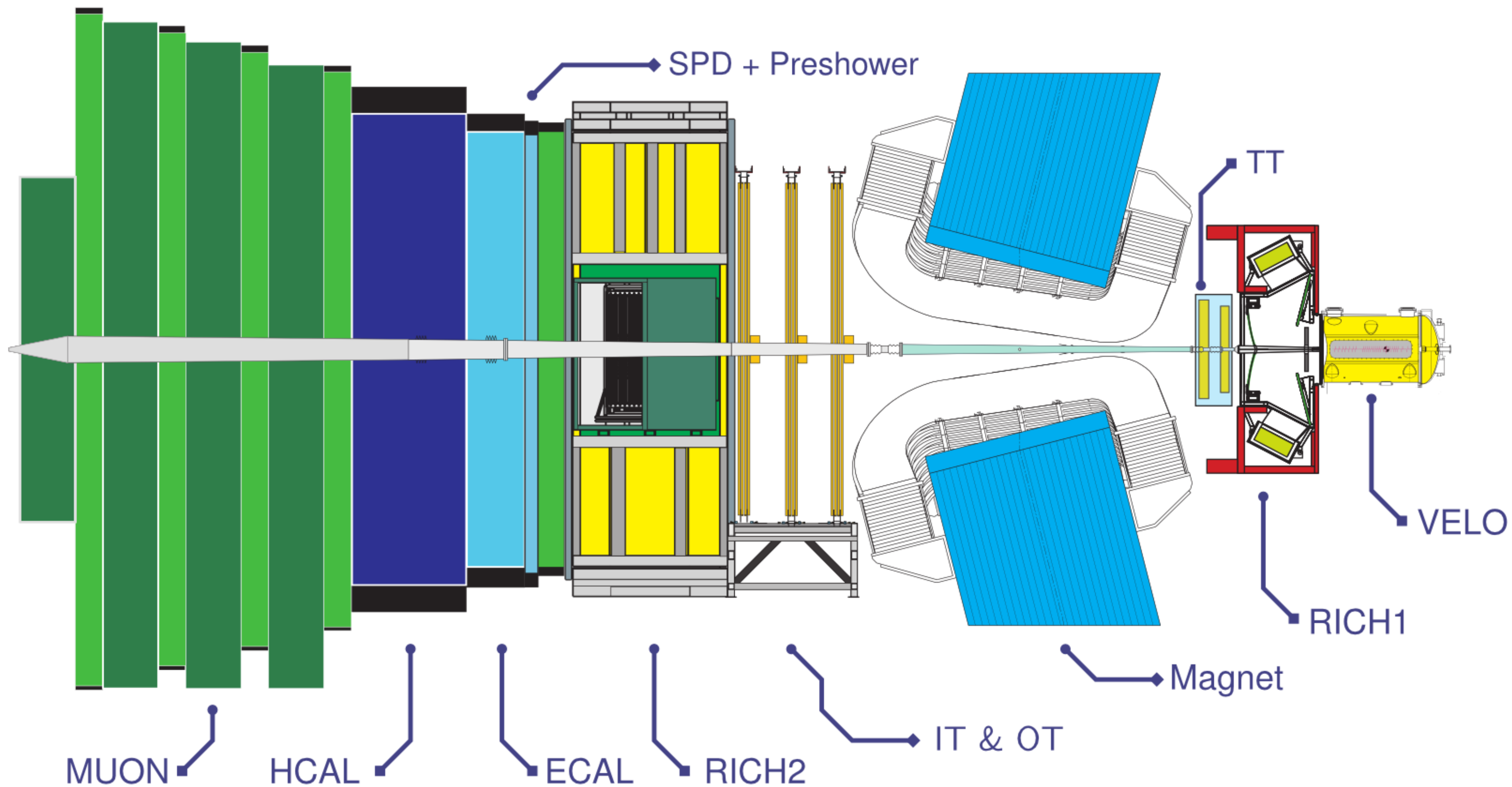


# are there?

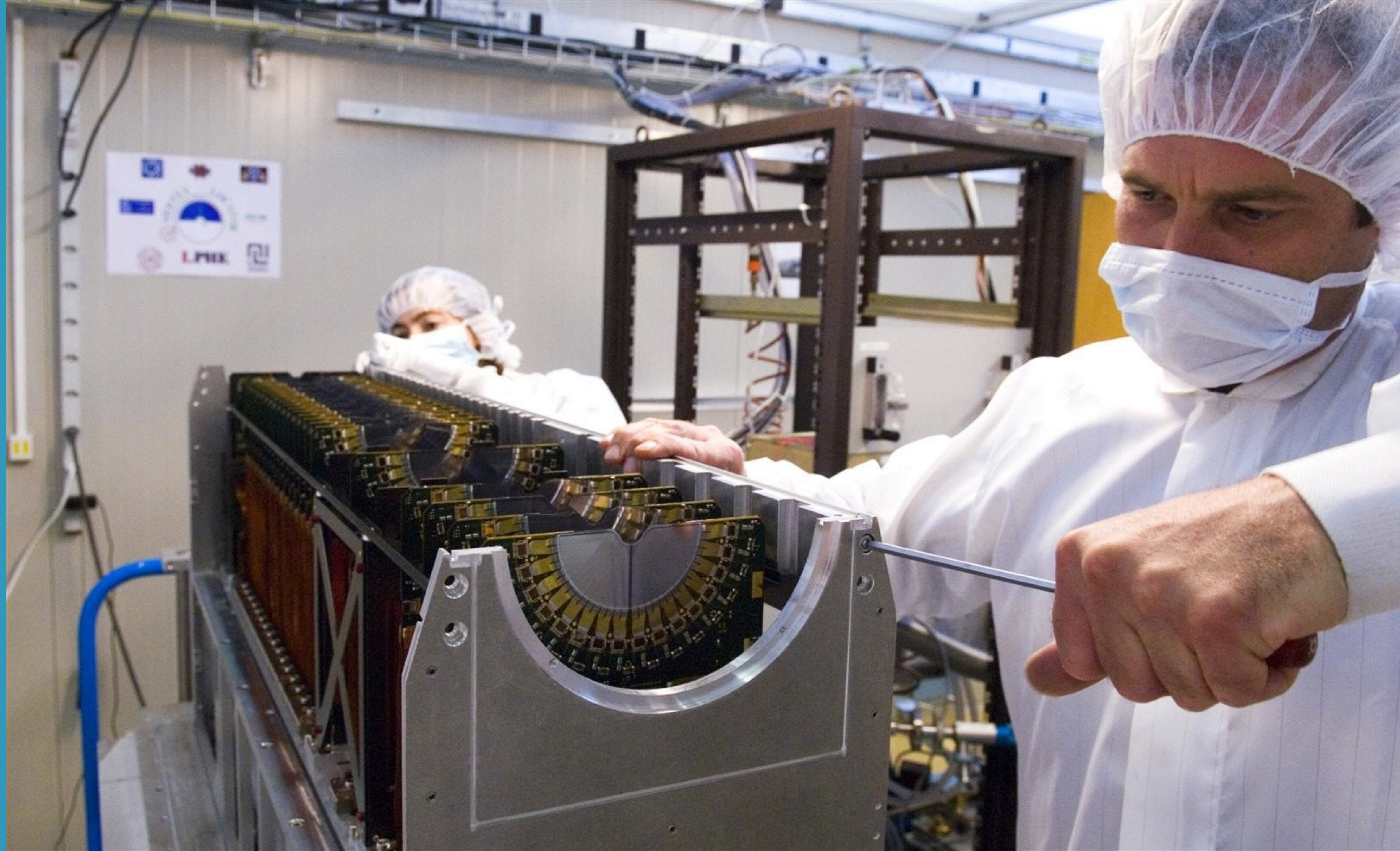




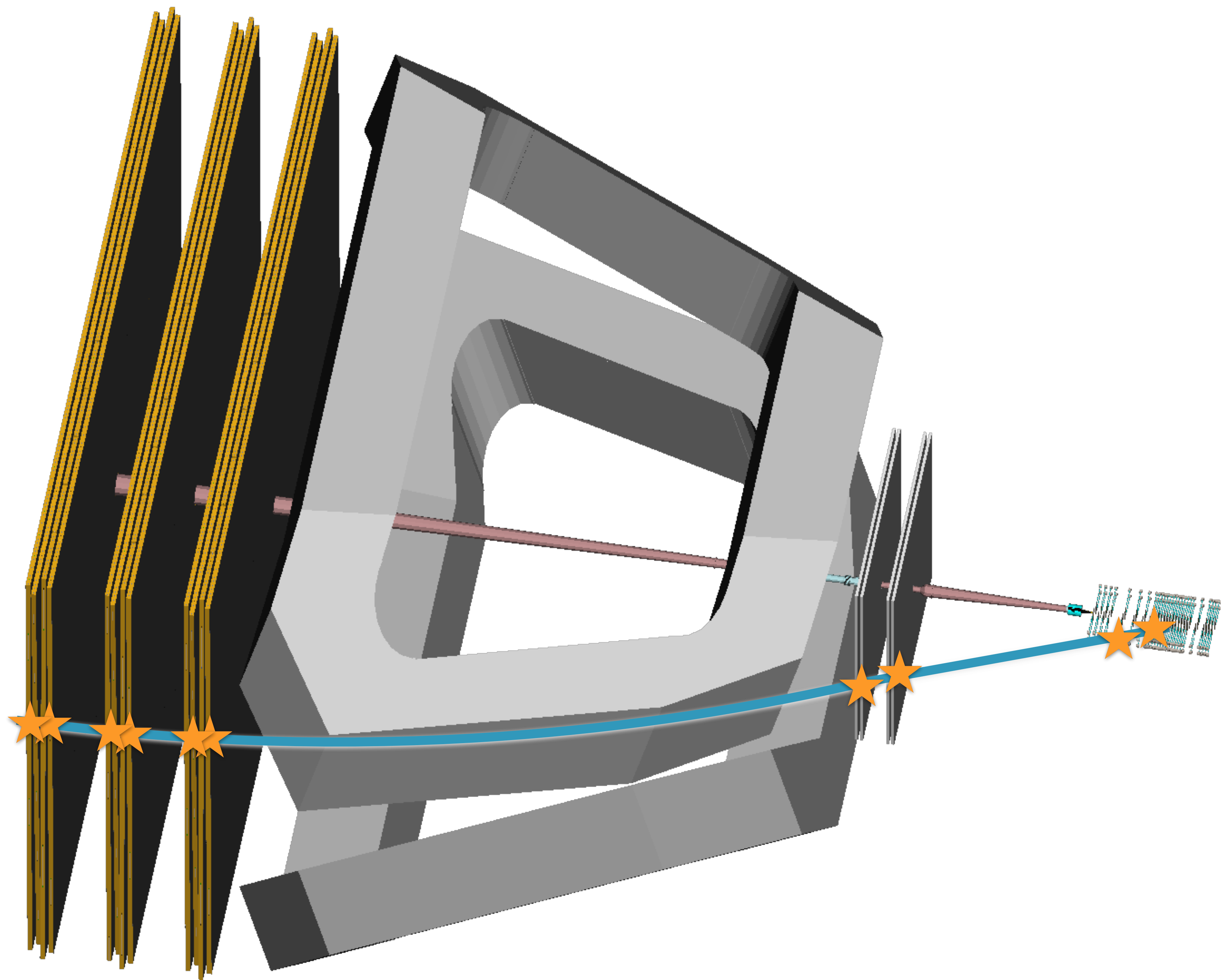














Who ya  
gonna call?

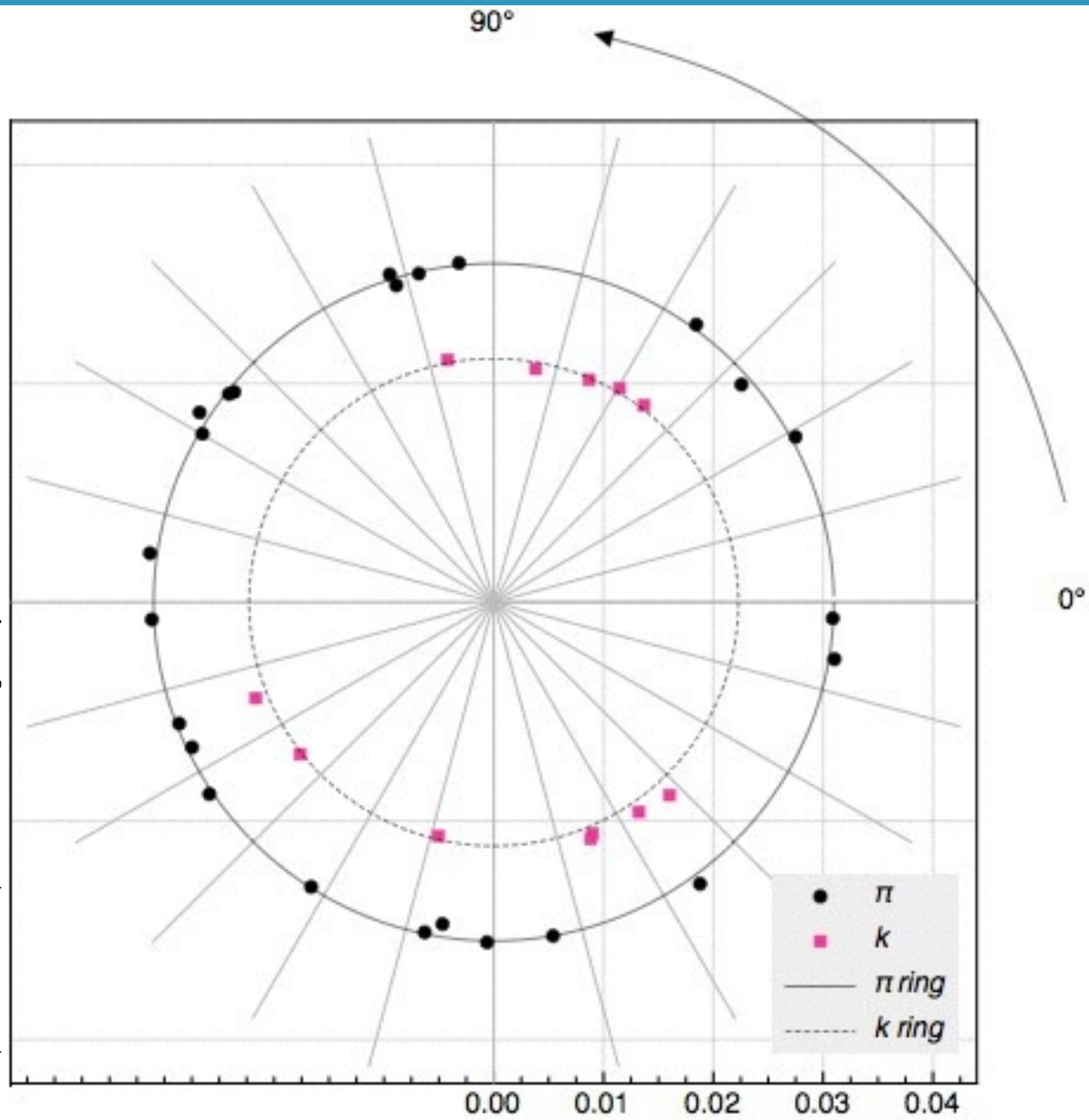


Cherenkov -  
faster than  
the speed  
of light



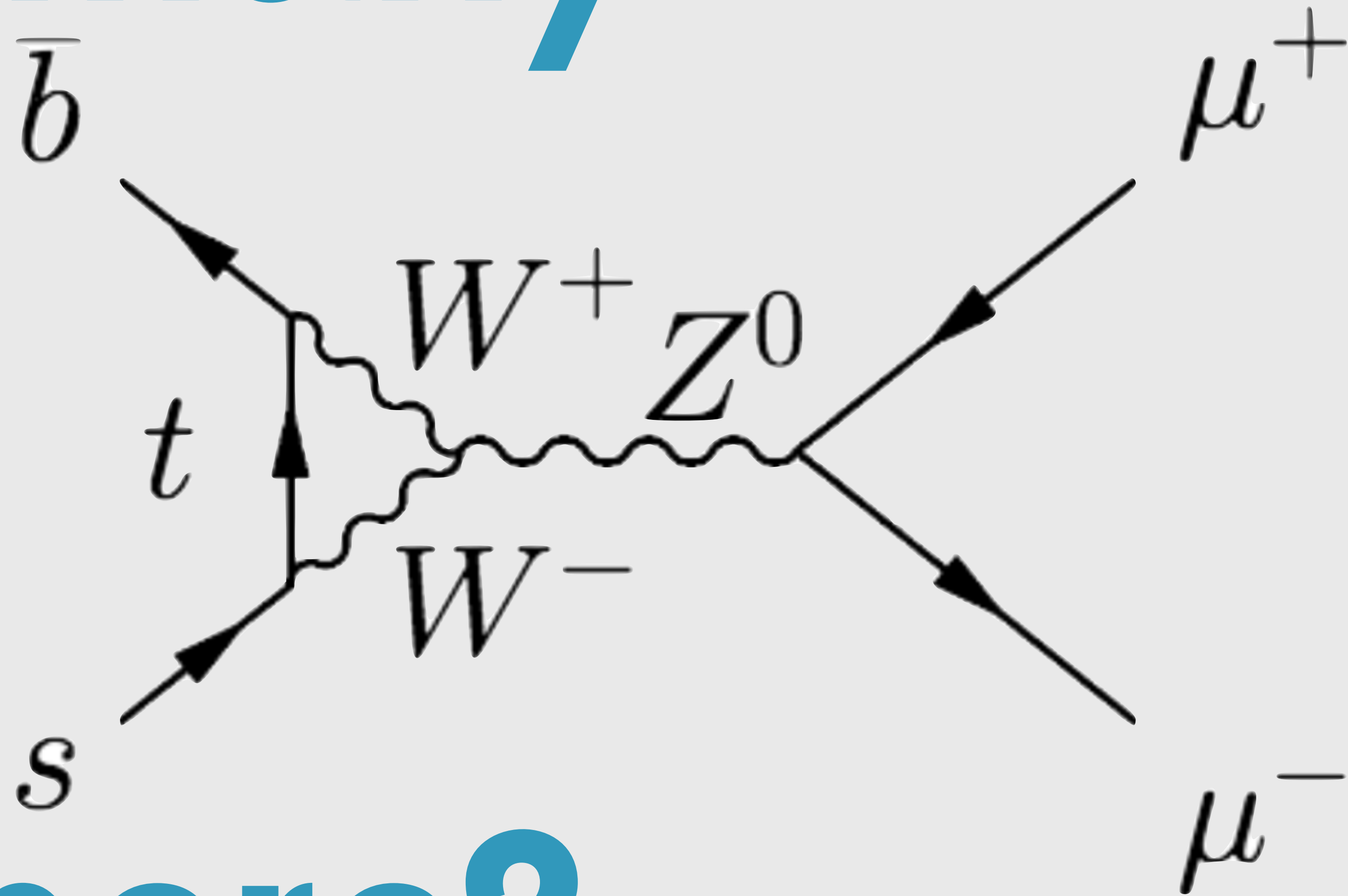


"Polar plot of Chrenkov photons emission angles" by Reculet CC BY-SA 3.0





# How many



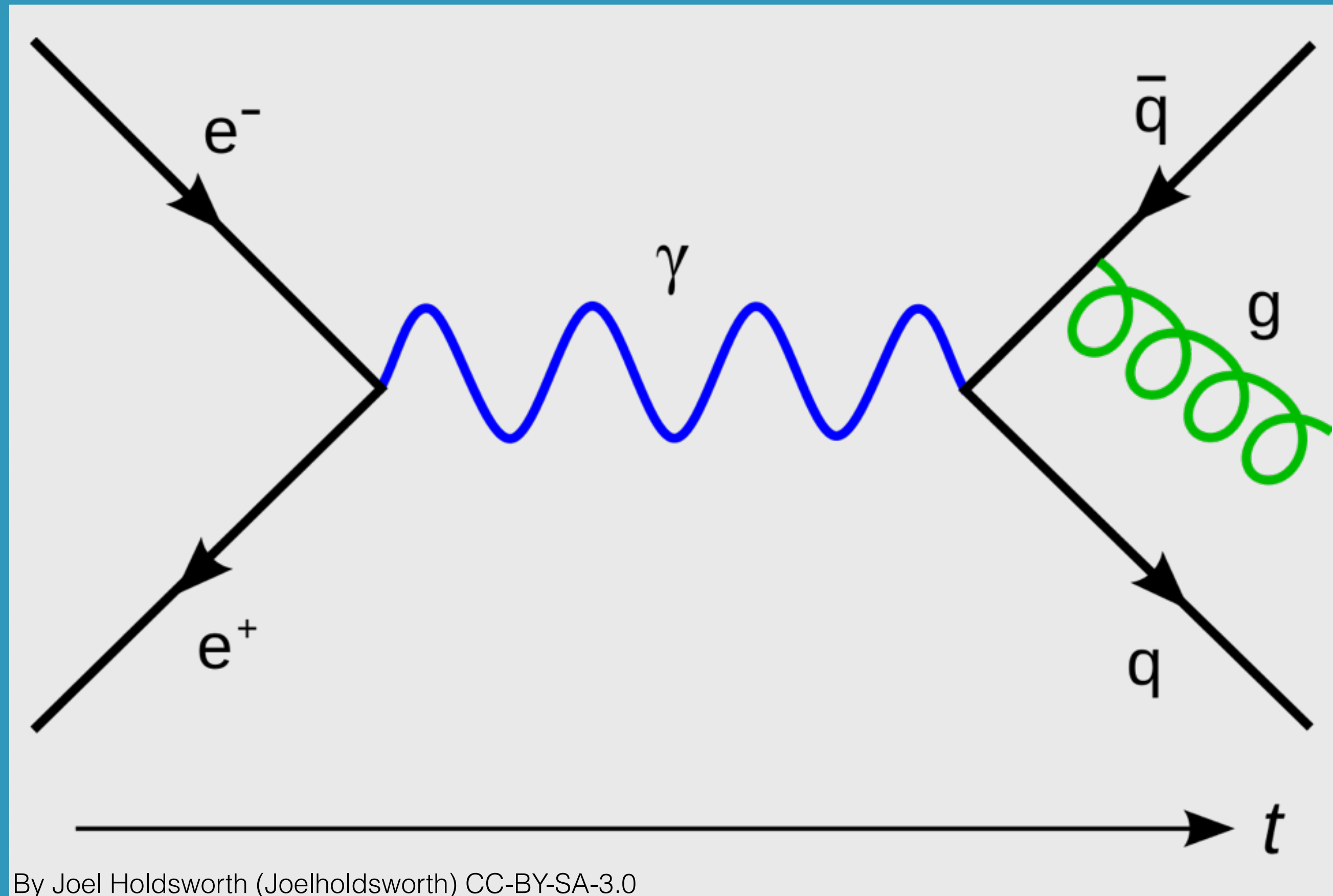
# are there?



What else  
does the  
LHC  
produce?



# Simulation







Simulate  
all the  
things

UXB





Simulate  
all the  
things  
#hard



What else  
does the  
LHC  
produce?



Discard events  
as early as  
possible



A photograph of a server room with multiple rows of server racks. The racks are filled with server units, and many of the units have glowing green and blue indicator lights. The perspective is from a low angle, looking down the aisles of the server racks. The text "Real Time Stream Processing" is overlaid on the left side of the image in a large, white, sans-serif font.

# Real Time Stream Processing



A photograph of a server room with multiple rows of server racks. The racks are filled with server units, and many of them have glowing green and blue indicator lights. The perspective is from a low angle, looking down the aisles of the server racks. The text "Real Time Stream Processing" is overlaid on the left side of the image in a large, white, sans-serif font.

# “Real” Time Stream Processing



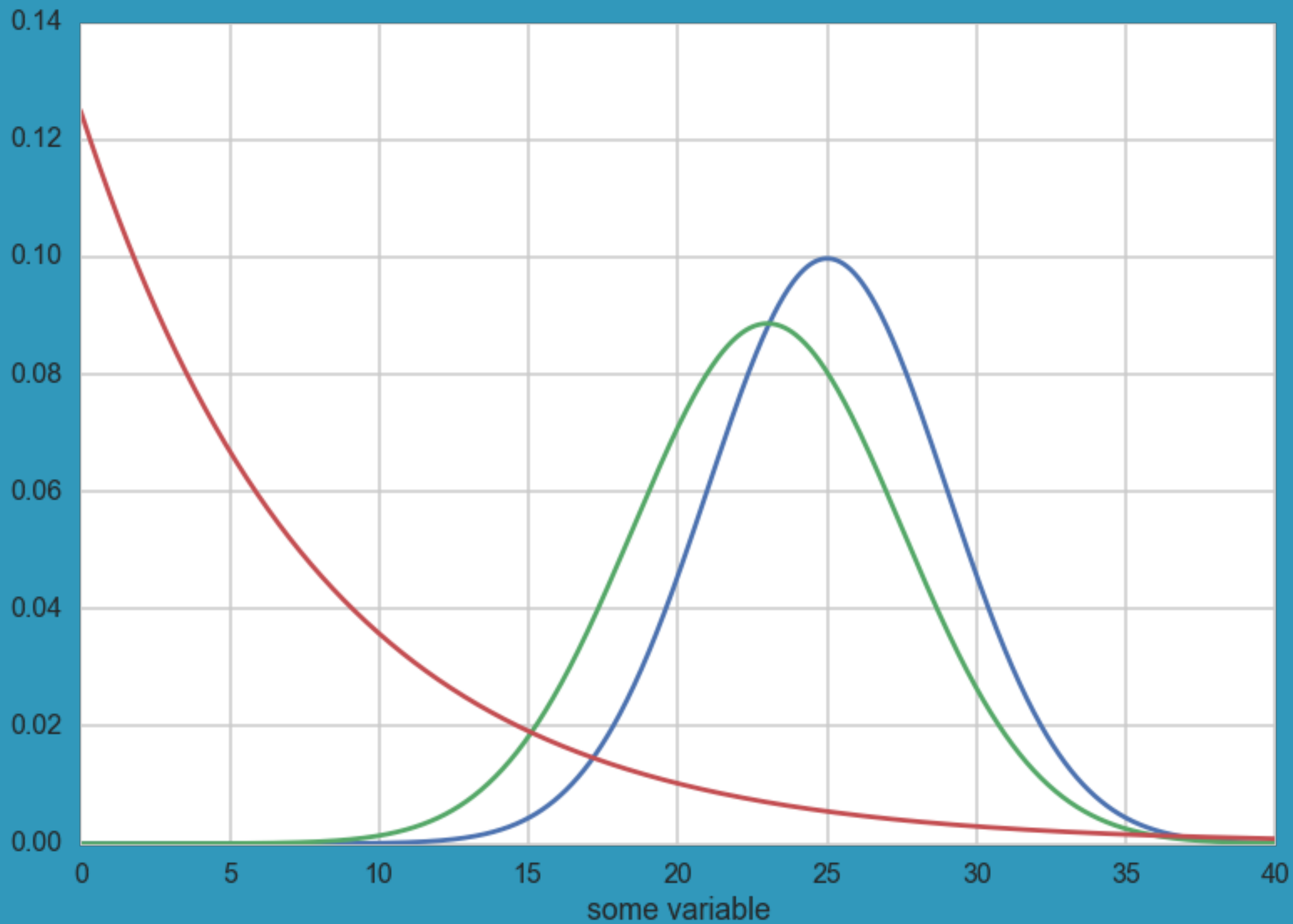
# The Topological Trigger

**Fast**

**Robust**

**General purpose**







“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**



# The Final Selection



# The Final Selection

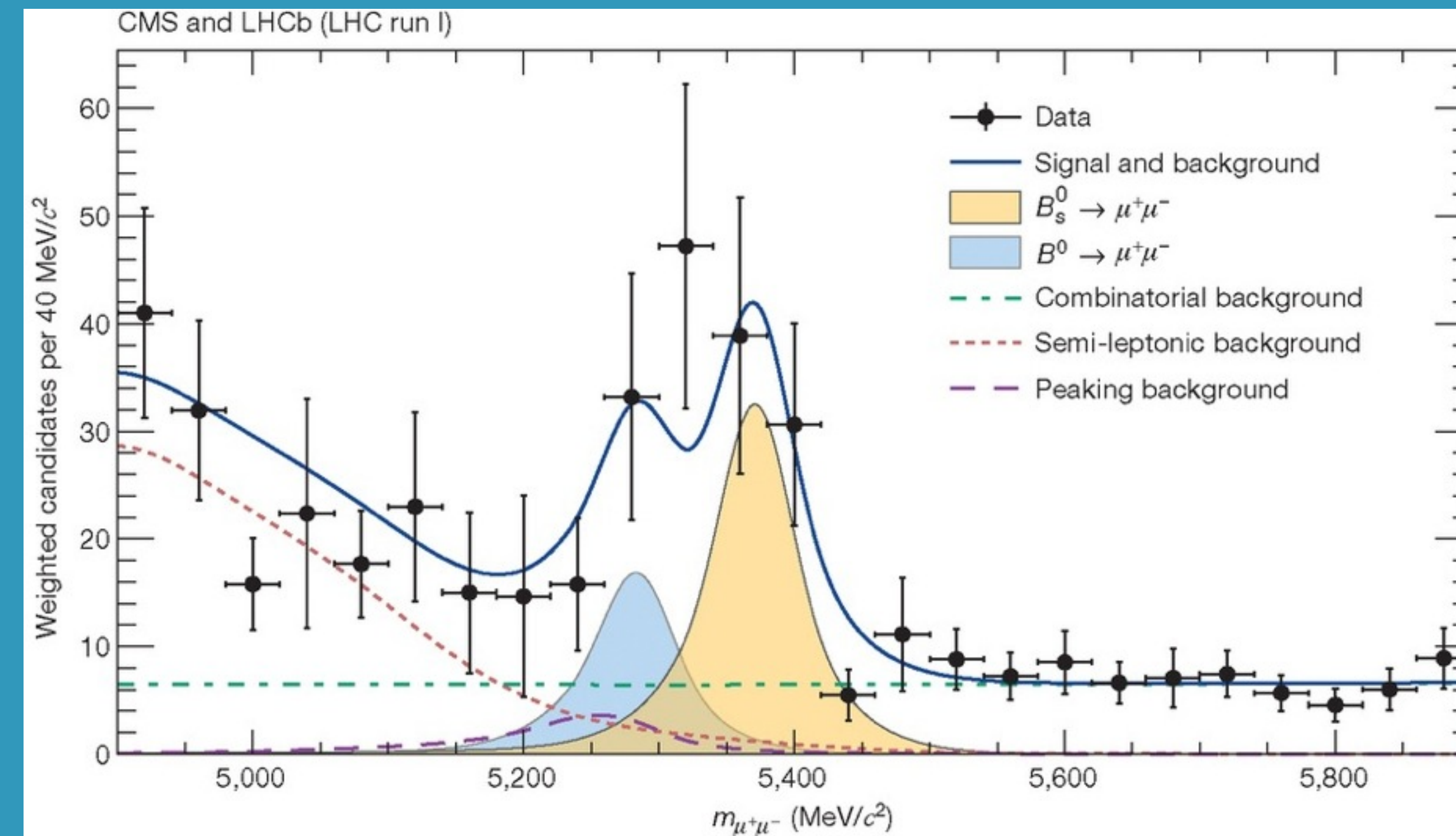
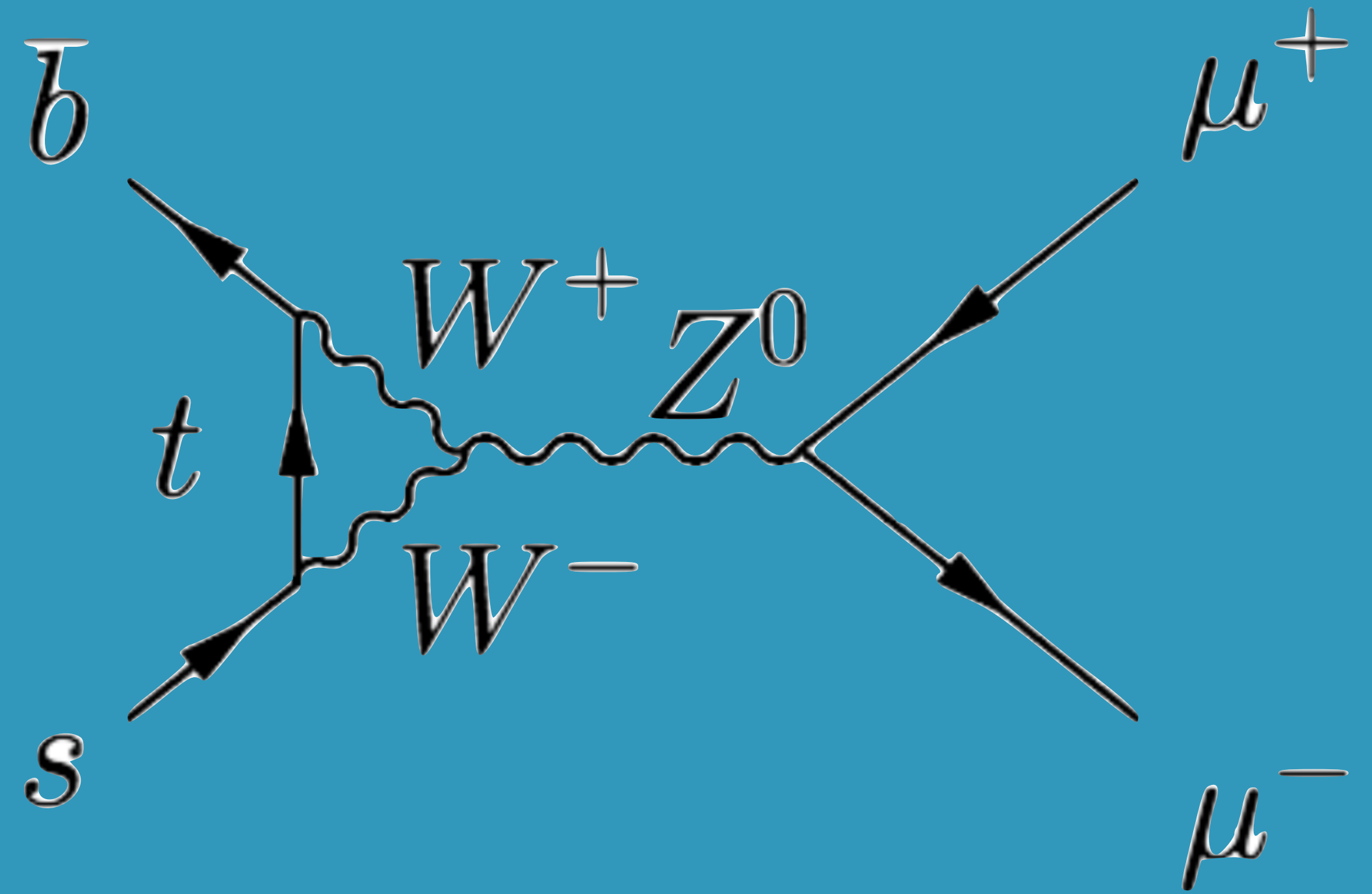
**This is where kaggle starts**



# Power Users

Performance!!

Advanced techniques

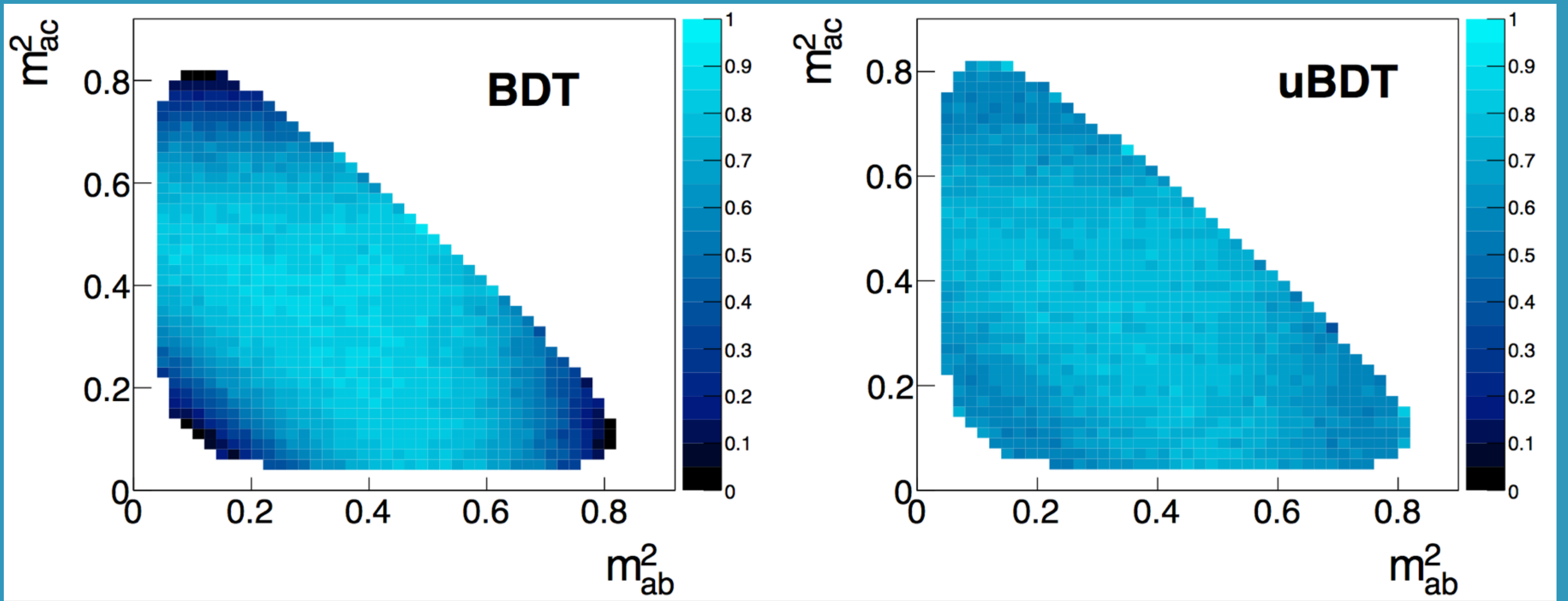




# Normal People

**Simplicity**  
**What is it doing?**





# Artisans

## Uniformity, or other special loss functions



# The Final Selection



# Building Models

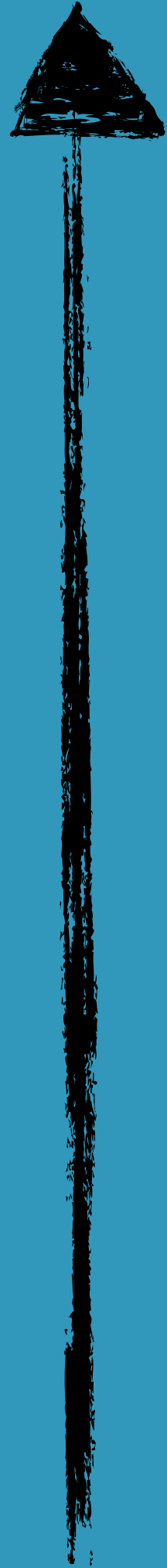


Classifier

Keep

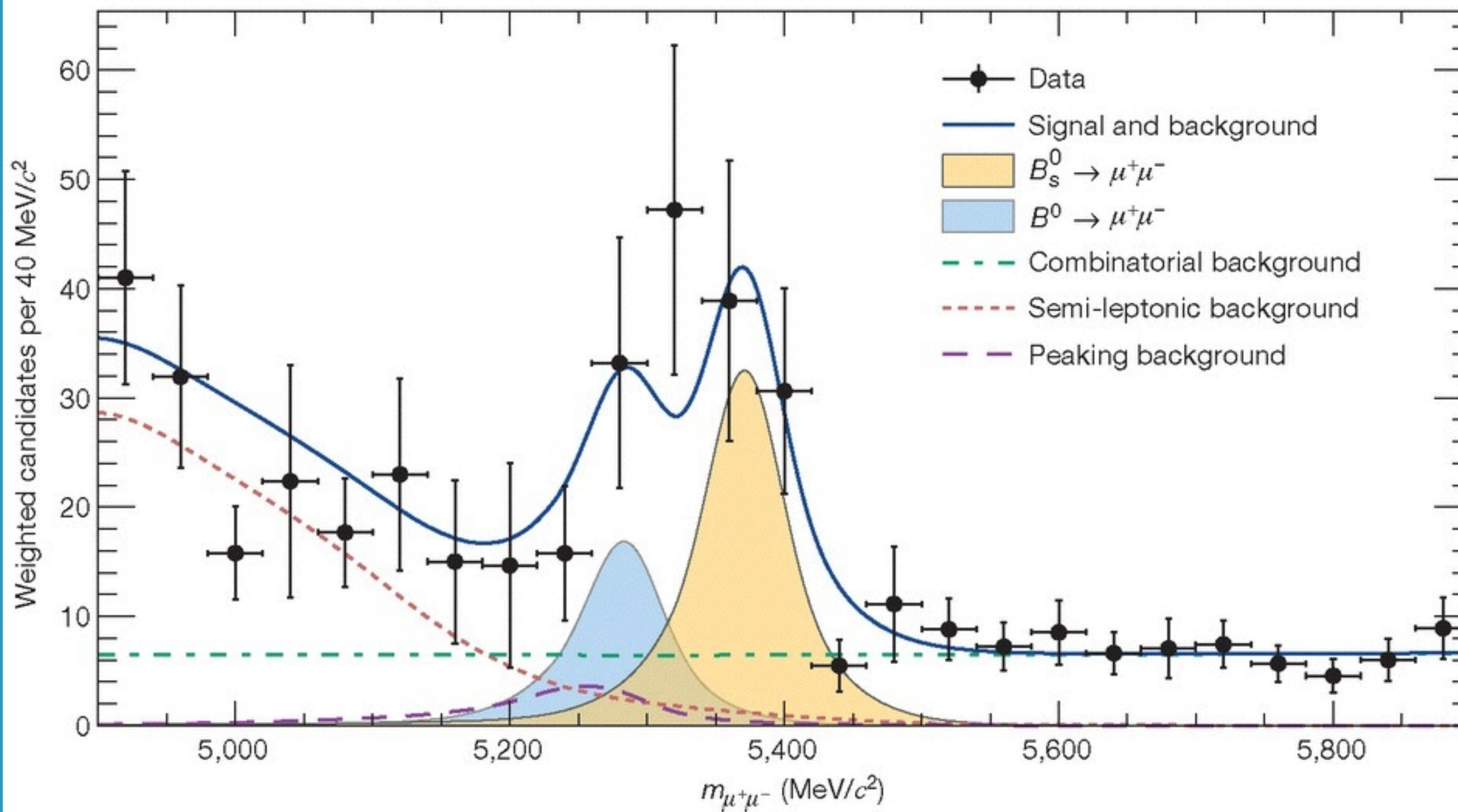
Discard

M





CMS and LHCb (LHC run I)

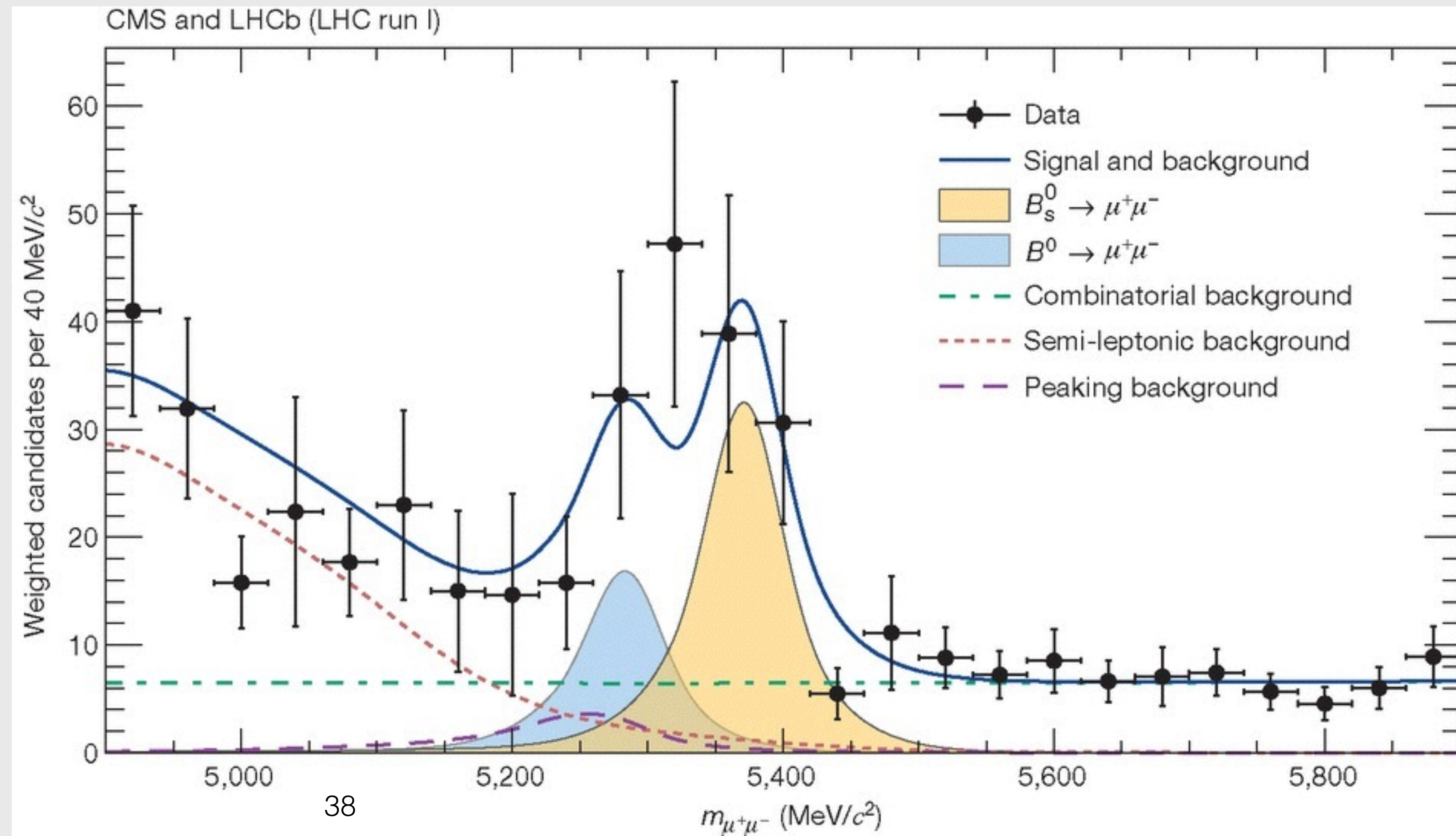




# The Known Unknowns



# Hypothesis Testing





# Publish!

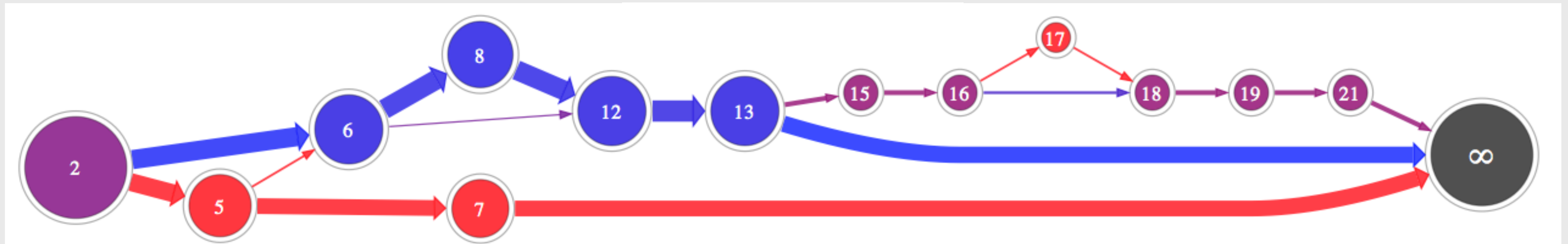


# Preempting the Trigger

**No work  
is faster than  
some work.**



# Preempting the Trigger





# Generative Models

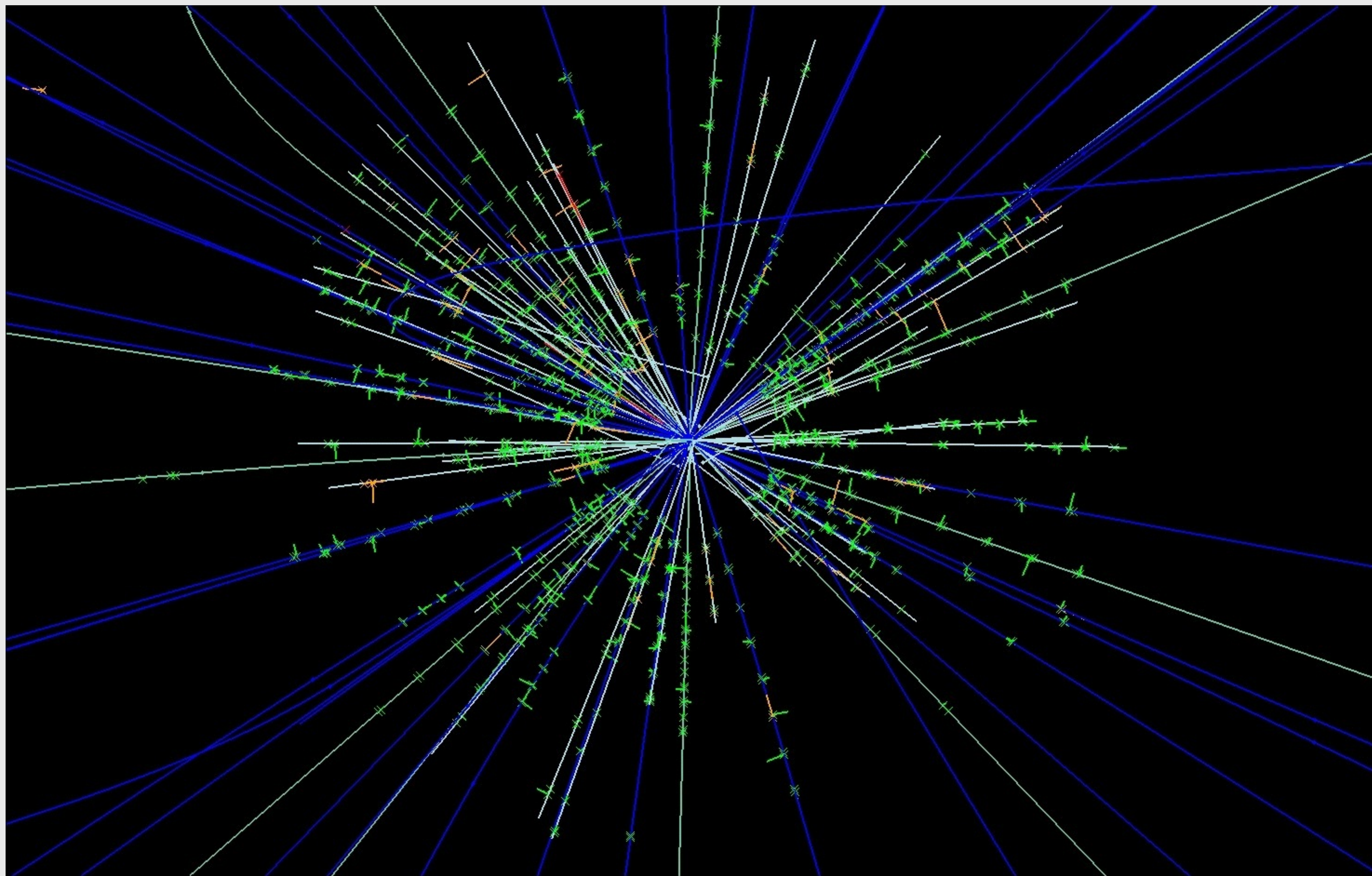
**Simulation is the largest  
user of CPU time!**



# Track Finding

**Reconstruction  
Extracting properties**







# Black Box Optimisation

**Optimising the whole chain**



??!



# Recap



Details matter







# Tim Head

 thead@cern.ch

  @betatim