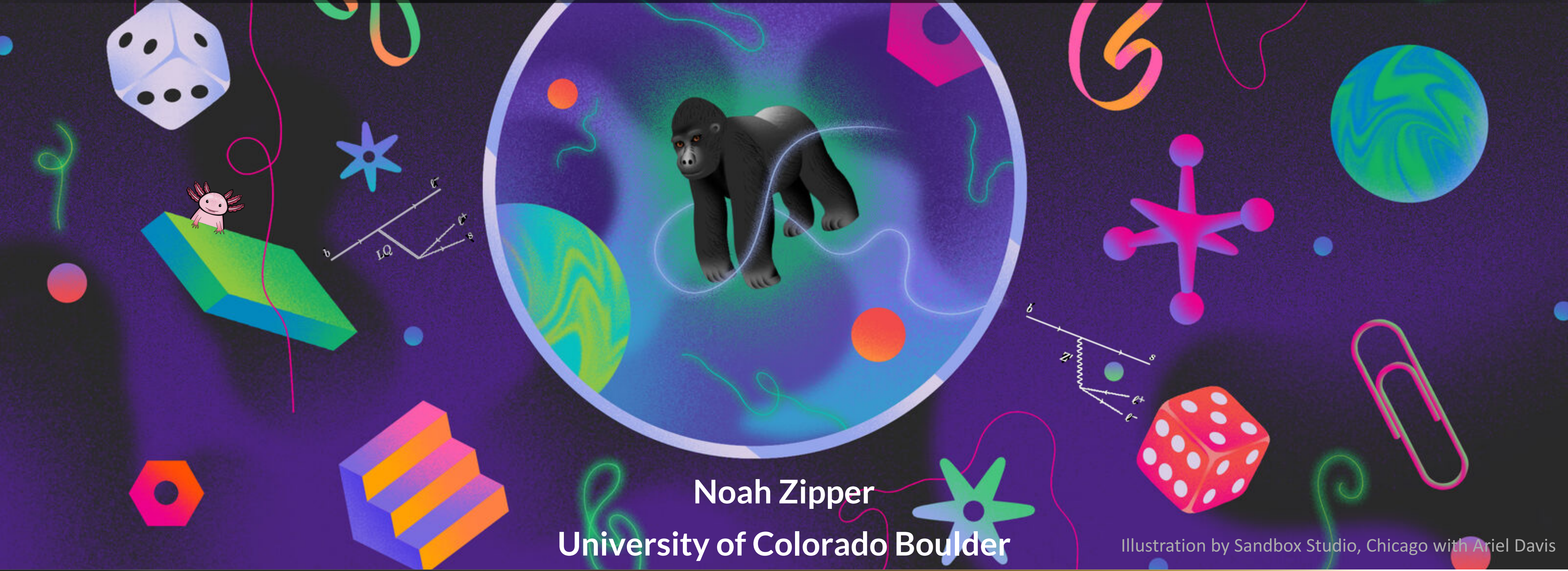
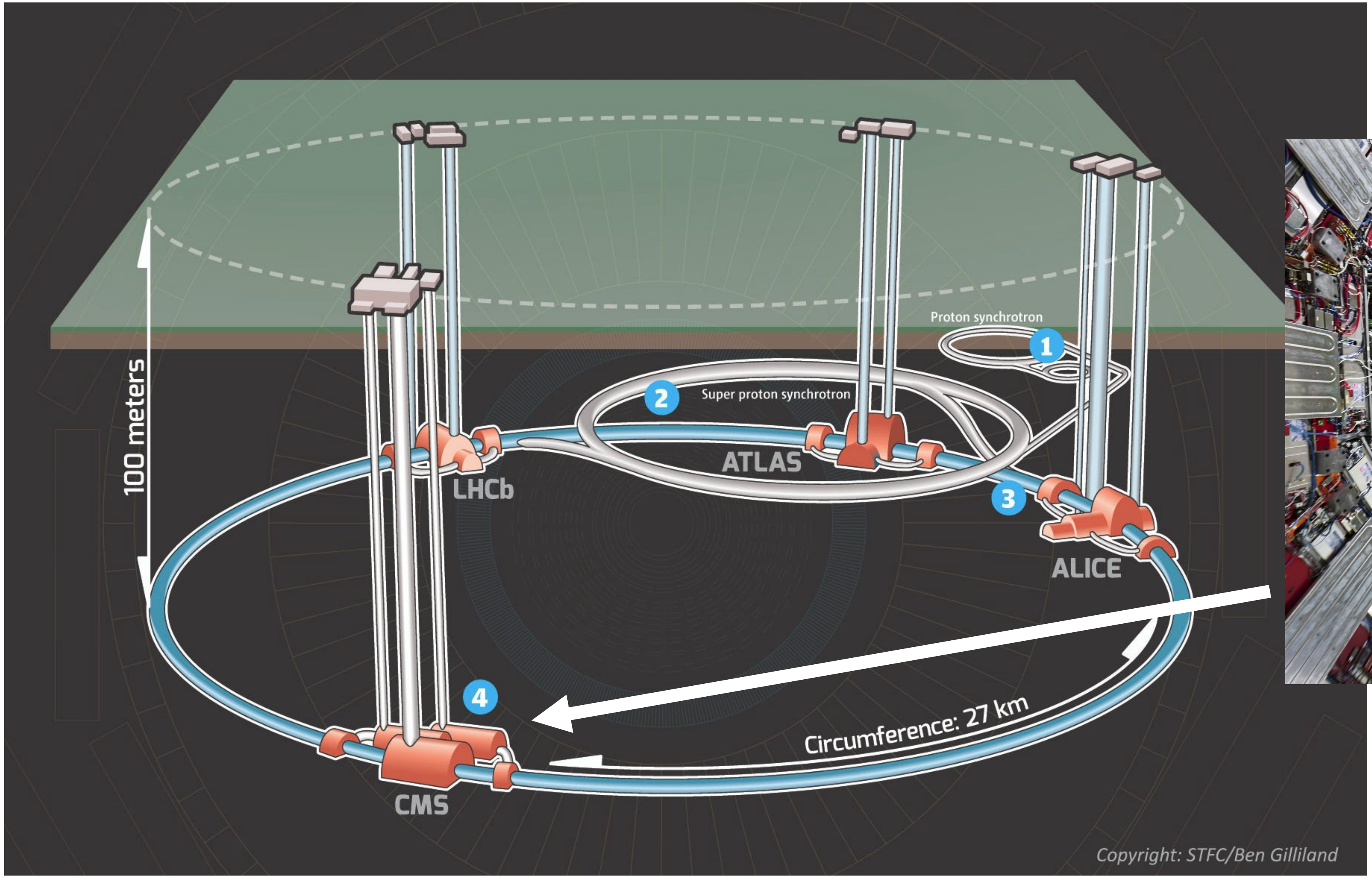
Precision Flavor Measurements and Real-Time Anomaly Detection at the CMS Detector

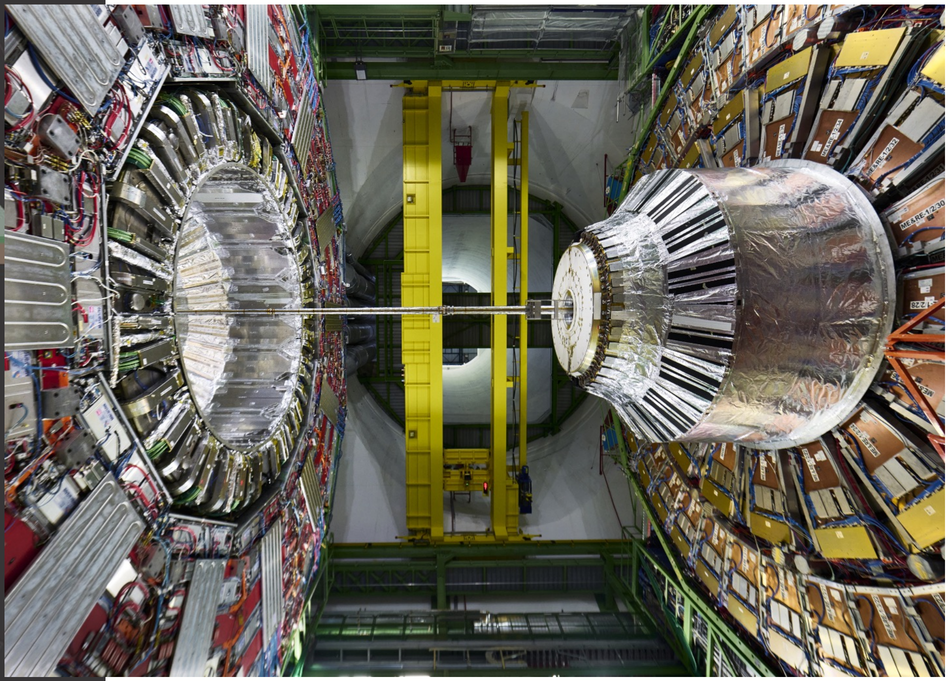


Noah Zipper
University of Colorado Boulder

Illustration by Sandbox Studio, Chicago with Ariel Davis

The Large Hadron Collider (LHC) @ CERN





https://home.cern/science/experiments/cms

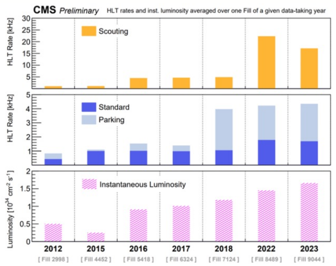
Why LHC Physics?

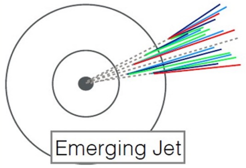
We have SOOO much data

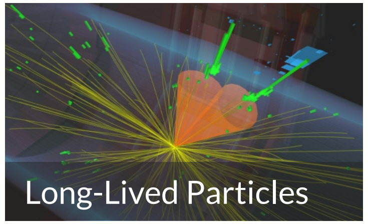
- It's been analyzed and over-analyzed
- Time to get creative ➔ new approaches to collect and analyze data

AXOL1TL

CICADA







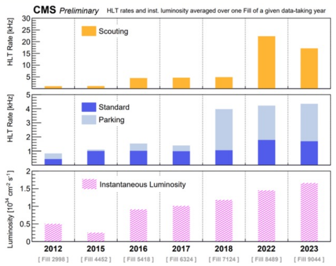
Why LHC Physics?

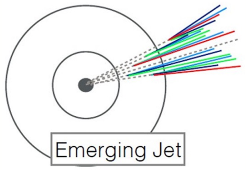
We have SOOO much data

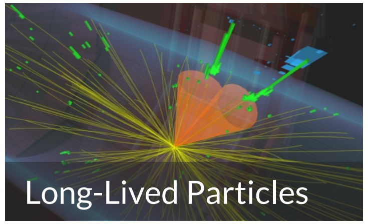
- It's been analyzed and over-analyzed
- Time to get creative ➔ new approaches to collect and analyze data

AXOL1TL

CICADA







CMS **Collaboration** @ CERN

- Complex interconnected detector systems
  - Tracking, calorimetry, and muon detection
  - Target vastly different searches and measurements
- Yet, we all contribute to maintaining and improving the detector for everyone's benefit

Coming Up…

We’ll talk about the trigger system

- Can we collect data in a smarter way?
- We think AXOL1TL can leverage machine learning to do it

Coming Up…

We’ll talk about the trigger system

- Can we collect data in a smarter way?
- We think AXOL1TL can leverage machine learning to do it

Introduction to precision measurements at CMS

- Confirming the Standard Model vs. new physics
- How do we actually *do* the analysis work?

Real-Time Anomaly Detection with an Unsupervised Autoencoder at the CMS Level-1 Trigger

The CMS Trigger

How can we deal with new collision data ~40 million times a second?

- We read in >60 TB/s from the detector!

The CMS Trigger

How can we deal with new collision data ~40 million times a second?

- We read in >60 TB/s from the detector!

The trigger cuts > 99.9% of incoming data, only picks interesting interactions

The CMS Trigger

How can we deal with new collision data ~40 million times a second?

- We read in >60 TB/s from the detector!

The trigger cuts > 99.9% of incoming data, only picks interesting interactions

We use a set of algorithms – the "trigger menu" – that looks at each event and decides to keep or toss data

The CMS Trigger

How can we deal with new collision data ~40 million times a second?

- We read in >60 TB/s from the detector!

The trigger cuts > 99.9% of incoming data, only picks interesting interactions

We use a set of algorithms – the "trigger menu" – that looks at each event and decides to keep or toss data

The trigger is broken up into two phases

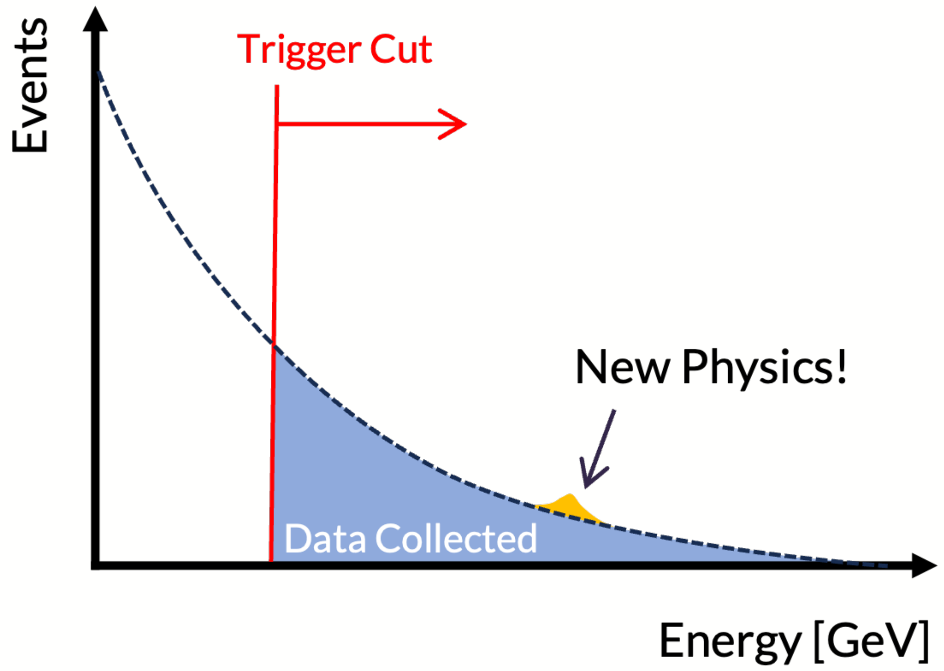
- Level-1 (L1T) – First step of real-time triggering, on hardware
- High-Level (HLT) – Data is passed from hardware to off-detector software

Why Anomaly Detection?

Currently, we use simple heuristics to define trigger algorithms

- Energy, charge, direction, momentum, etc.

In this approach, we need to know what we're looking for to target it



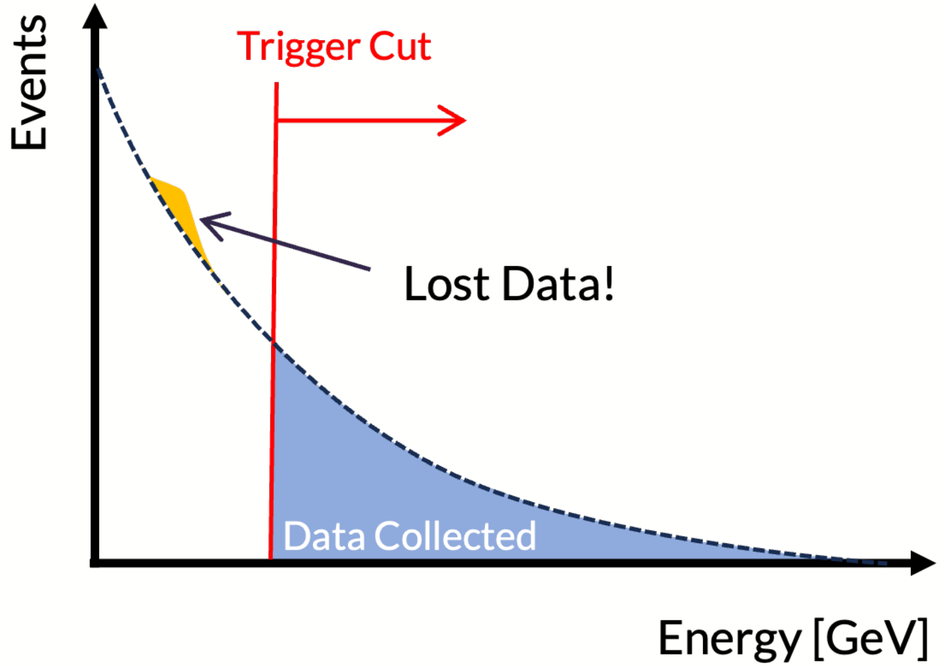
Why Anomaly Detection?

Currently, we use simple heuristics to define trigger algorithms

- Energy, charge, direction, momentum, etc.

In this approach, we need to know what we're looking for to target it

- **How do we stop rejecting data because we don't know what to look for?**



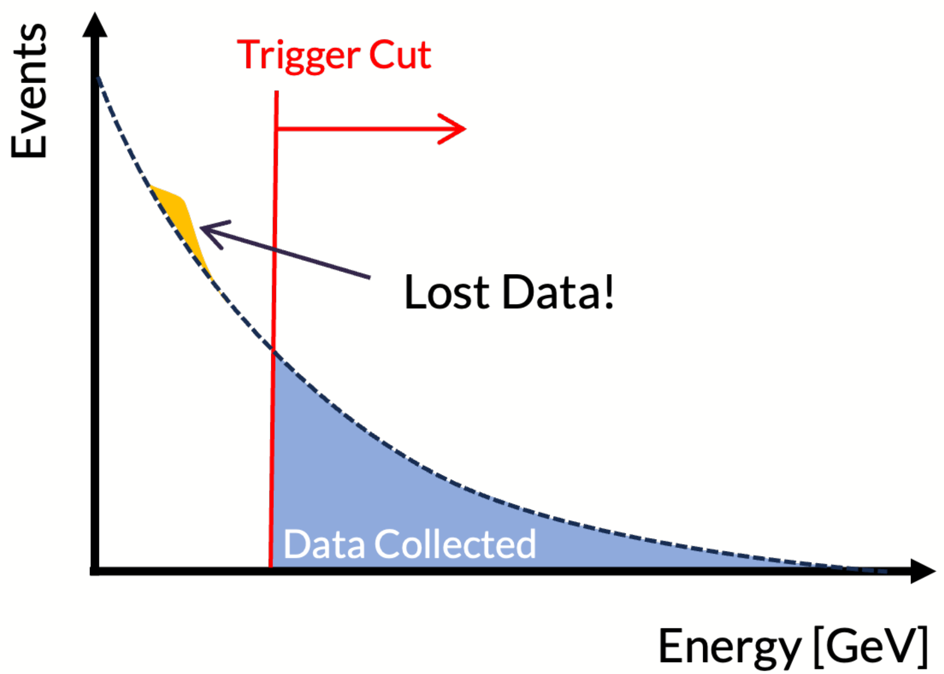
Why Anomaly Detection?

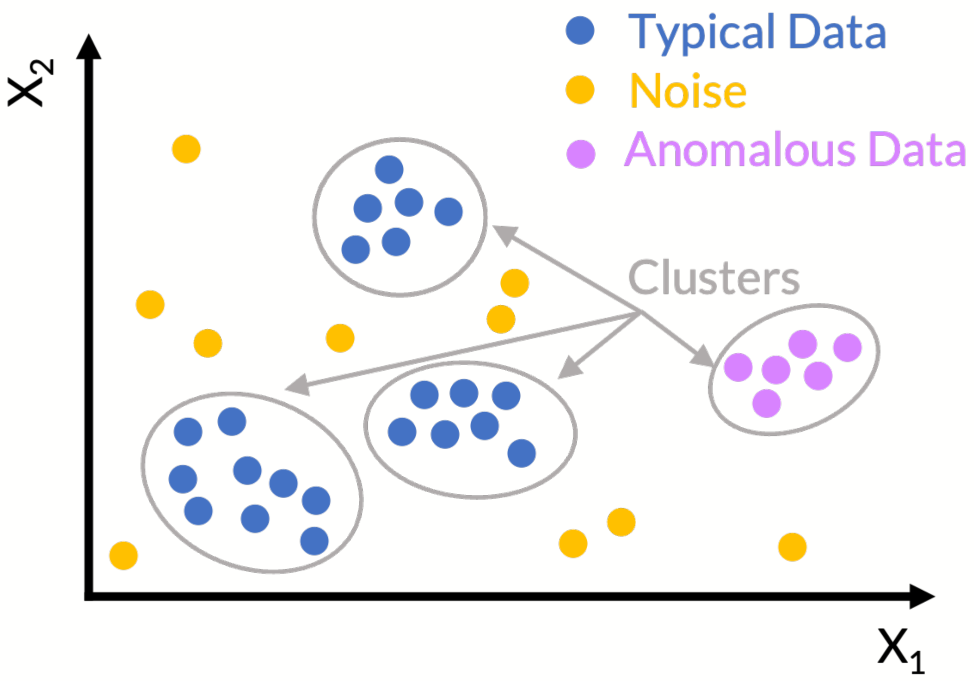
Currently, we use simple heuristics to define trigger algorithms

- Energy, charge, direction, momentum, etc.

In this approach, we need to know what we're looking for to target it

- **How do we stop rejecting data because we don't know what to look for?**





AXOL1TL Algorithm

"Zero Bias"

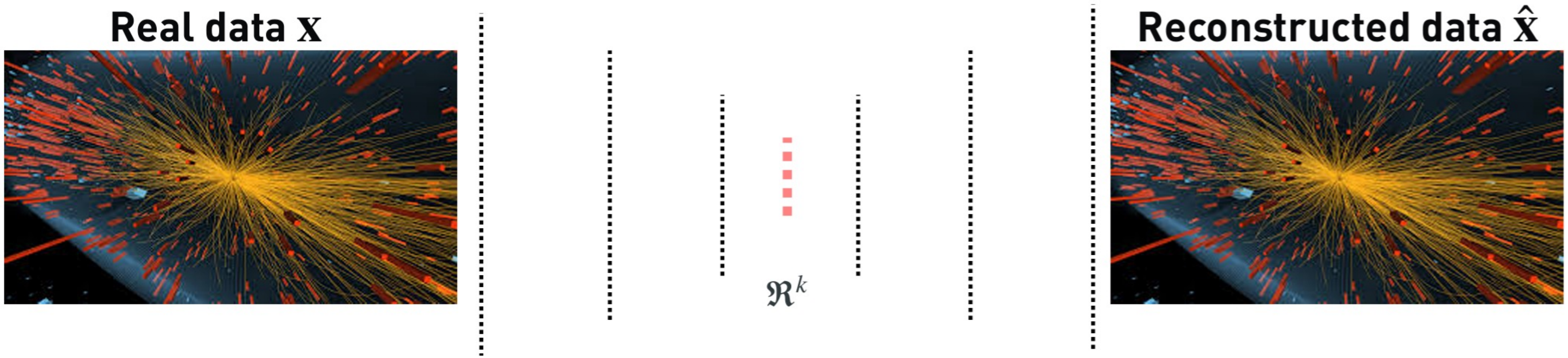
A dataset with no triggers, only turned on for small slices of time. Records events synched up with when collisions occur, saves everything.

We use an unsupervised Variational Autoencoder (VAE)

- Simple neural network(s), trained on real Zero Bias* data
- Basic trigger objects as vector inputs

VAE uses encoder & decoder to compress and reconstruct the input data

- Squeeze data into a small dimension "latent space"
  - Forces efficient information encoding ➔ network "learns"
- Network gets good at encoding + decoding typical data examples



AXOLITL Algorithm

"Zero Bias"

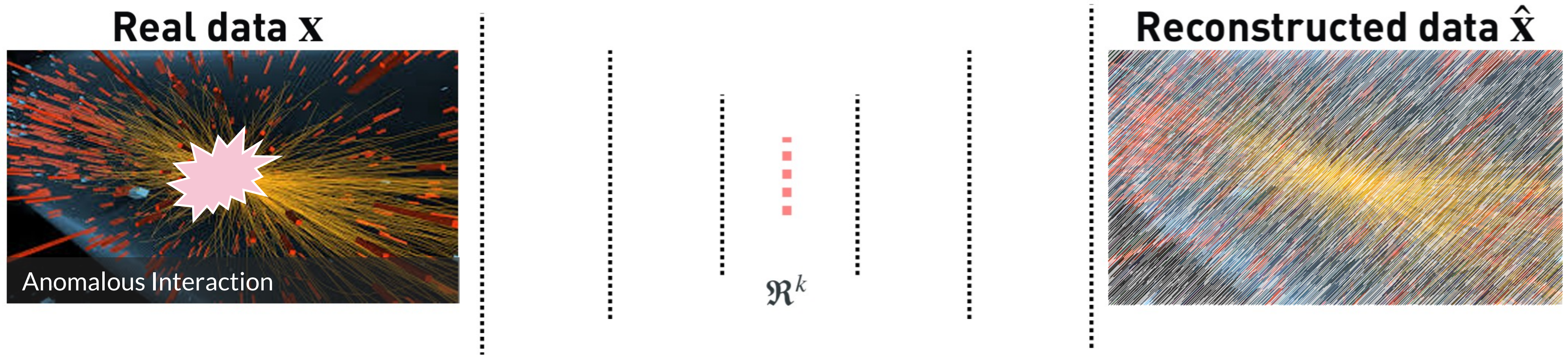
A dataset with no triggers, only turned on for small slices of time. Records events synched up with when collisions occur, saves everything.

We use an unsupervised Variational Autoencoder (VAE)

- Simple neural network(s), trained on real Zero Bias* data
- Basic trigger objects as vector inputs

VAE uses encoder & decoder to compress and reconstruct the input data

- Squeeze data into a small dimension "latent space"
  - Forces efficient information encoding ➔ network "learns"
- Network gets good at encoding + decoding typical data examples
- Much worse for atypical examples



AXOL1TL Algorithm

"Zero Bias"

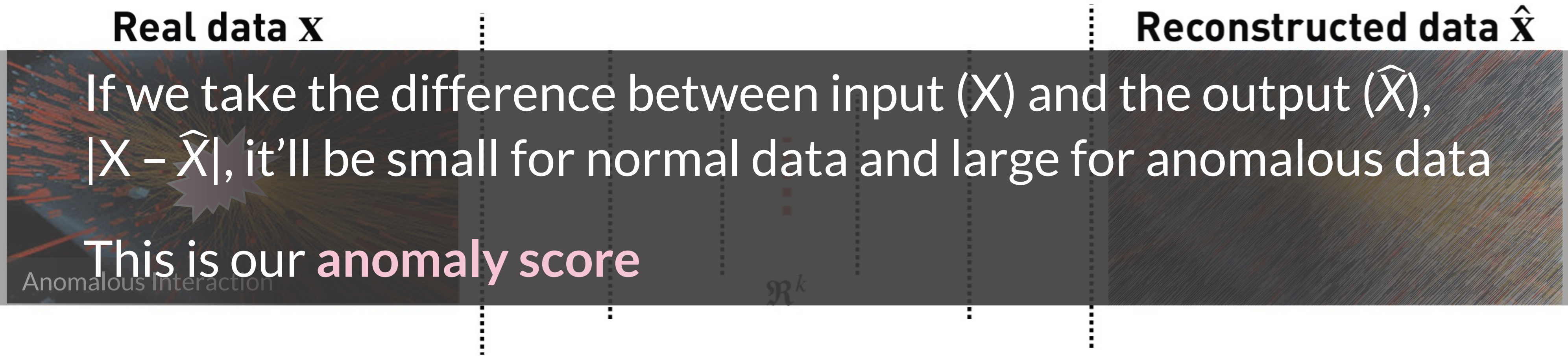
A dataset with no triggers, only turned on for small slices of time. Records events synched up with when collisions occur, saves everything.

We use an unsupervised Variational Autoencoder (VAE)

- Simple neural network(s), trained on real Zero Bias* data
- Basic trigger objects as vector inputs

VAE uses encoder & decoder to compress and reconstruct the input data

- Squeeze data into a small dimension "latent space"
  - Forces efficient information encoding ➔ network "learns"
- Network gets good at encoding + decoding typical data examples
- Much worse for atypical examples



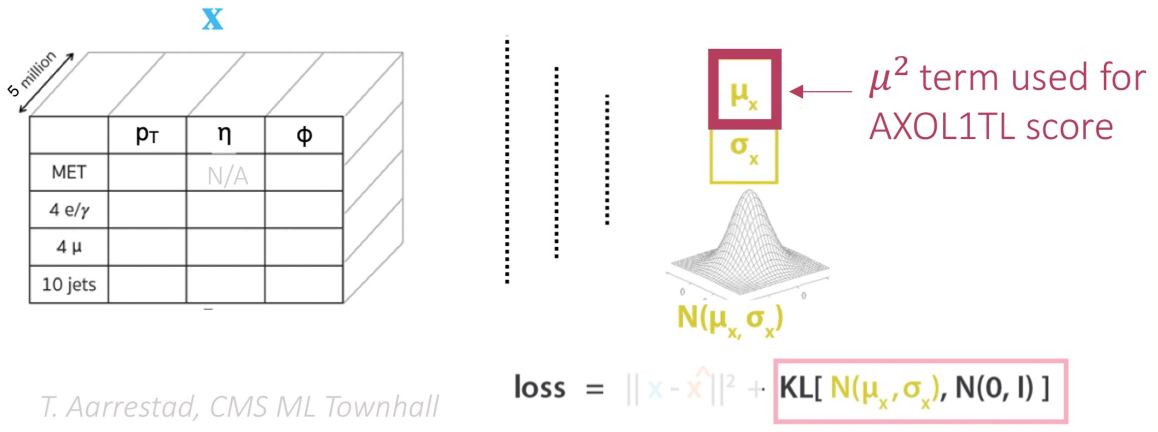
If we take the difference between input (X) and the output ($\widehat{X}$), $|X - \widehat{X}|$, it'll be small for normal data and large for anomalous data

This is our **anomaly score**

Integrating into the Trigger System

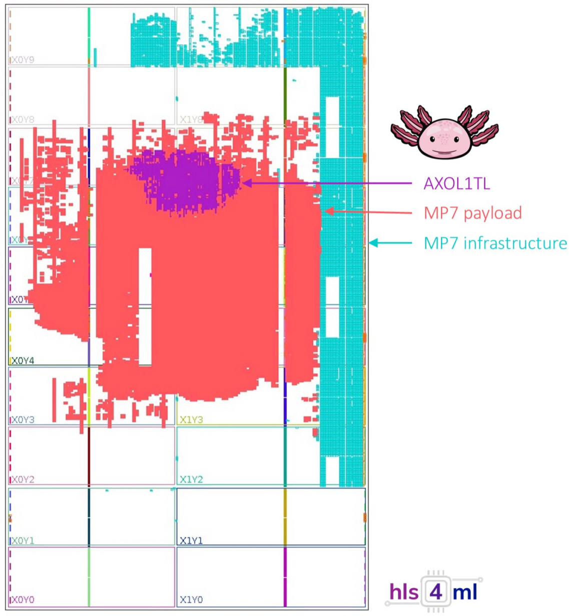
Algorithm must run on Field Programmable Gate Arrays (FPGAs)

- Cut out decoder and simplify score metric
- Minimal performance degradation
- **Runs in < 50 nanoseconds**



T. Aarrestad, CMS ML Townhall

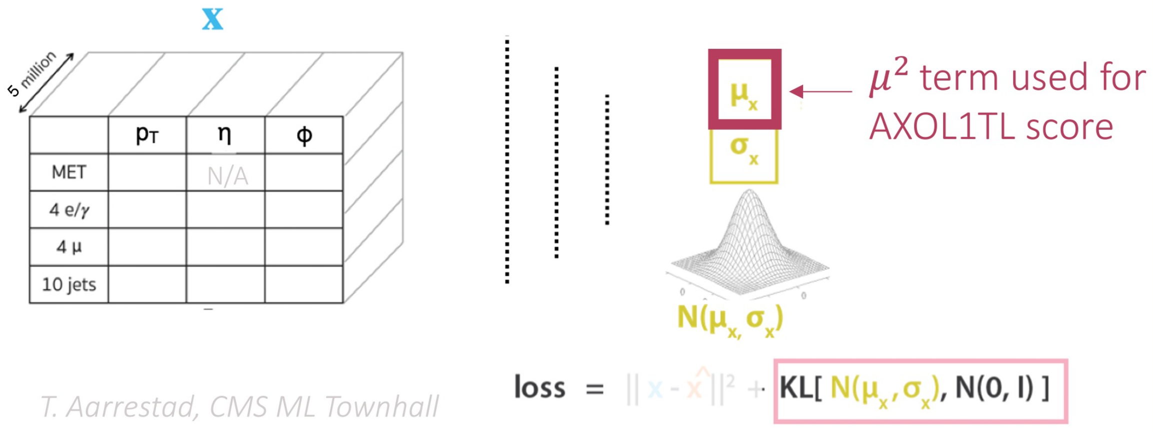
FPGA "Floorplan"



Integrating into the Trigger System

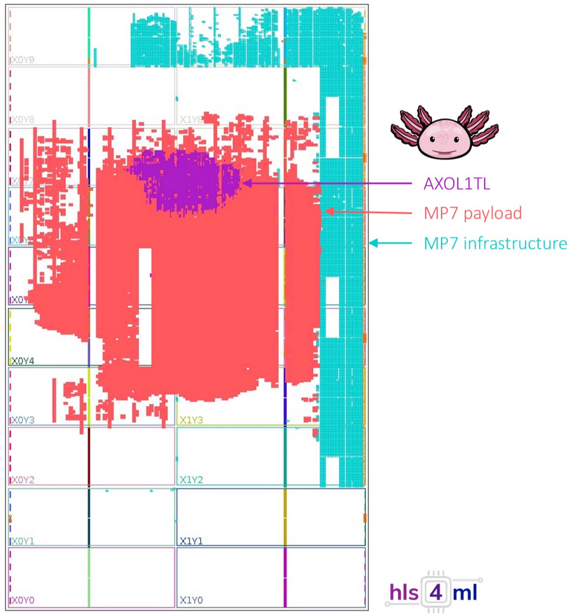
Algorithm must run on Field Programmable Gate Arrays (FPGAs) with constraints

- Cut out decoder and simplify score metric
- Minimal performance degradation
- **Runs in < 50 nanoseconds**

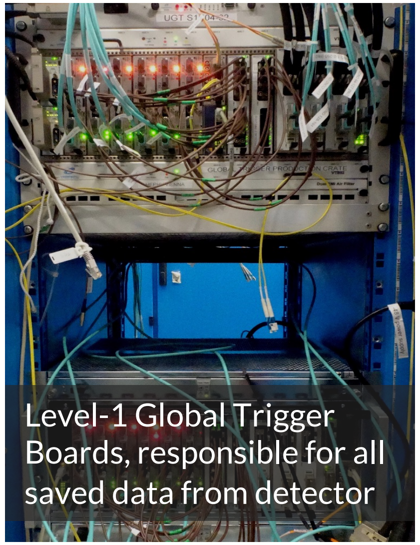


T. Aarrestad, CMS ML Townhall

FPGA "Floorplan"



AXO added into production system in May 2024 🎉



First Results from Real 2024 Data!

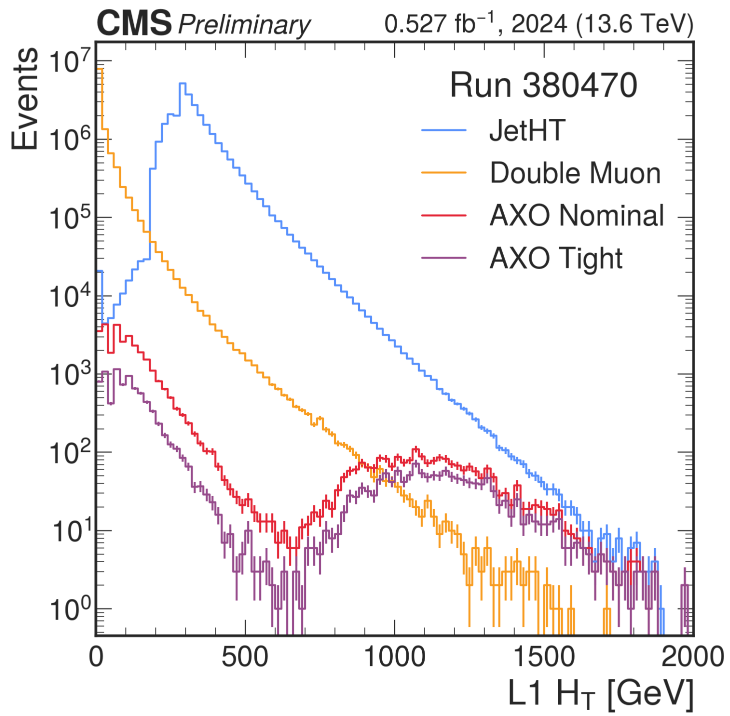
Still have lots of data to look through, but these are some first observations…

First Results from Real 2024 Data!

“H_T” or “Hadronic Energy Sum”

Quarks or gluons from collisions produce clusters of energy in the detector. We sum up all this energy in an event to get the H_T.

Still have lots of data to look through, but these are some first observations…



In some kinematic variables like H_T*, we see different shapes in AXO vs. other triggers

AXO decides certain known signals are too common

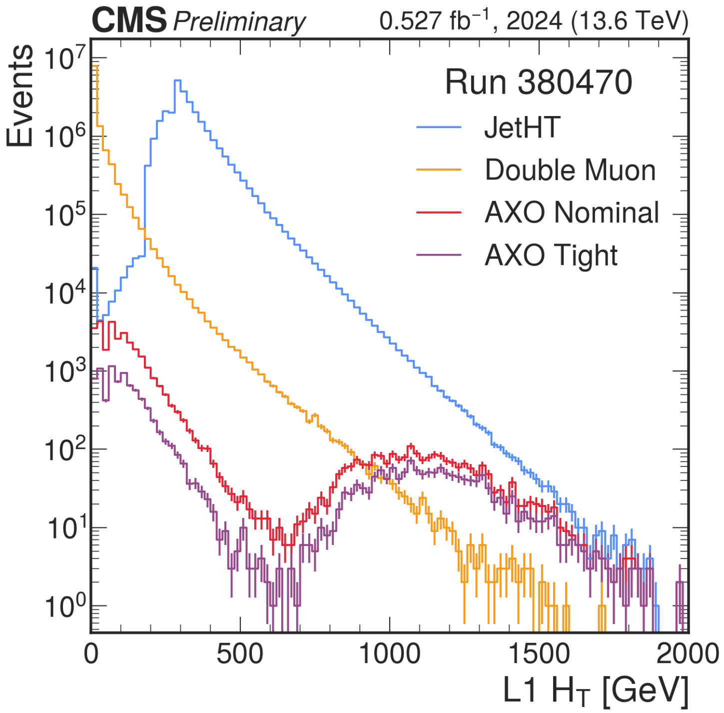
- Selects other, more anomalous, patterns

First Results from Real 2024 Data!

"H_T" or "Hadronic Energy Sum"

Quarks or gluons from collisions produce clusters of energy in the detector. We sum up all this energy in an event to get the H_T.

Still have lots of data to look through, but these are some first observations…



In some kinematic variables like H_T*, we see different shapes in AXO vs. other triggers

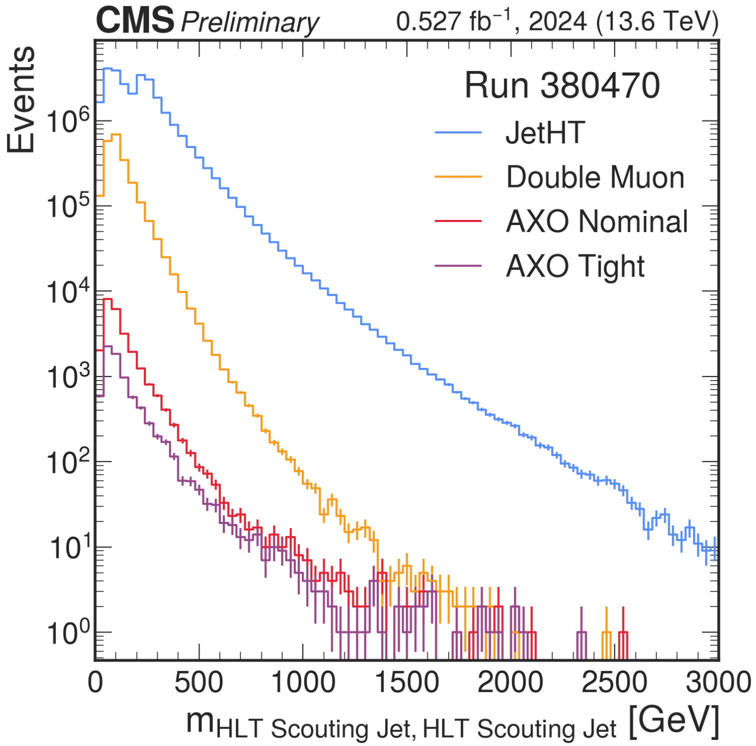
AXO decides certain known signals are too common

- Selects other, more anomalous, patterns

Invariant mass distributions

- Combine objects to find a decaying particle mass
- Smooth and falling shapes

Smooth shapes means easier backgrounds to characterize

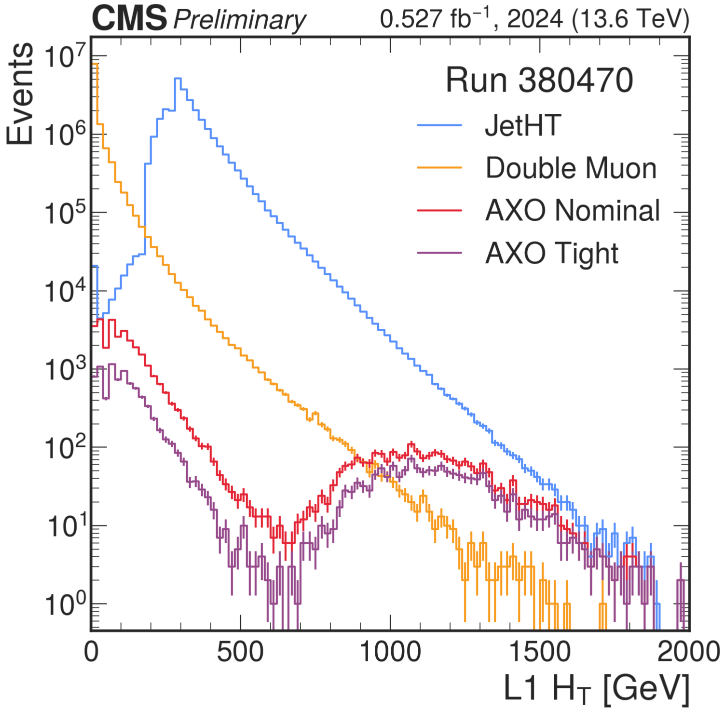


First Results from Real 2024 Data!

"H_T" or "Hadronic Energy Sum"

Quarks or gluons from collisions produce clusters of energy in the detector. We sum up all this energy in an event to get the H_T.

Still have lots of data to look through, but these are some first observations…



In some kinematic variables like H_T*, we see different shapes in AXO vs. other triggers

AXO decides certain known signals are too common

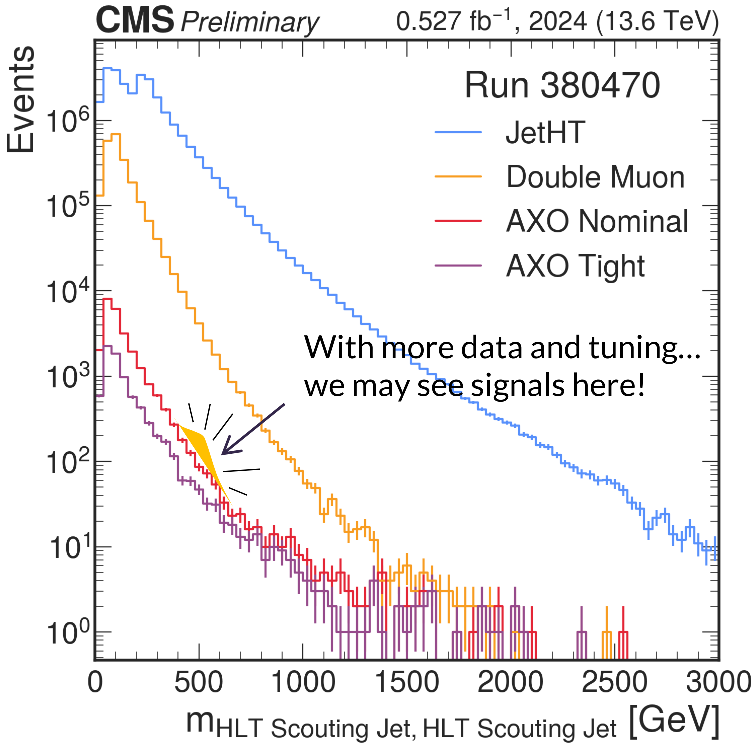
- Selects other, more anomalous, patterns

Invariant mass distributions

- Combine objects to find a decaying particle mass
- Smooth and falling shapes

Smooth shapes means easier backgrounds to characterize

- We can find new particles!



Ongoing Work

Dig more into the data, figure out what patterns AXO is finding

Design analysis strategies with anomaly data

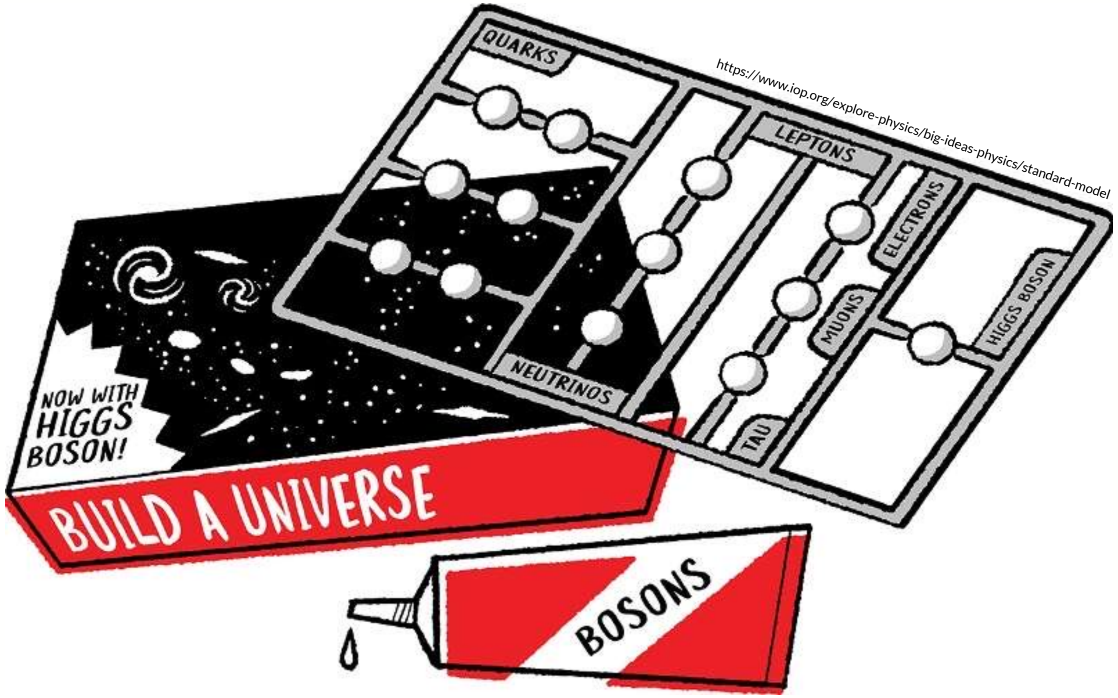
Update and upgrade algorithm

A Precision Measurement of Lepton Flavor Universality with the R(K) Ratio at the CMS Detector

Lepton Flavor Universality (LFU)

The Standard Model (SM) of particle physics is built on symmetries

- Particles and interactions are constructed so they obey these symmetries



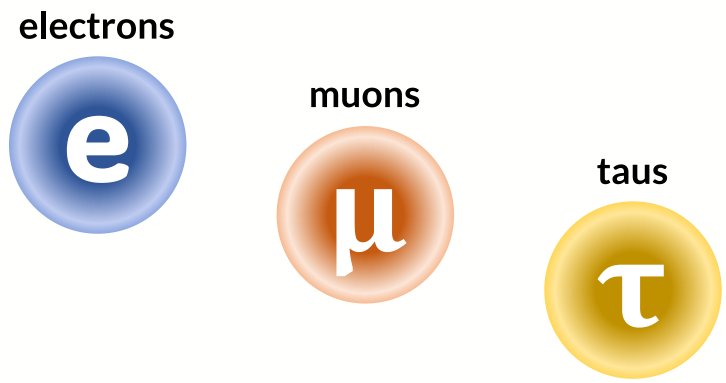
Lepton Flavor Universality (LFU)

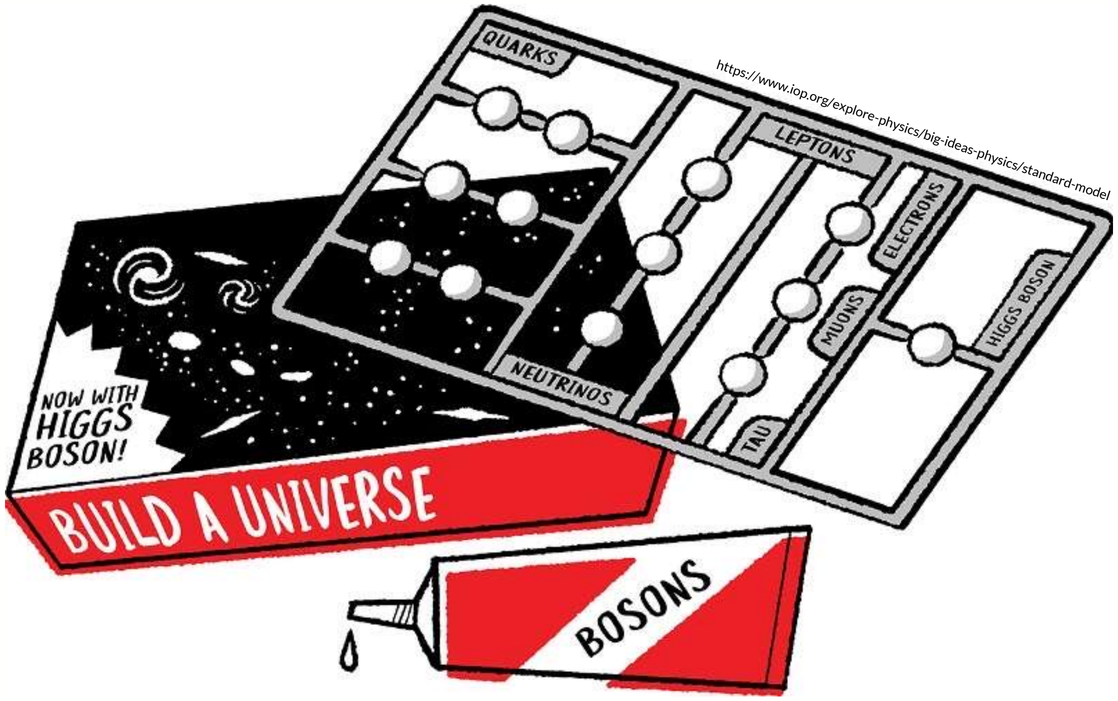
The Standard Model (SM) of particle physics is built on symmetries

- Particles and interactions are constructed so they obey these symmetries

One implicit symmetry is LFU

- We have 3 lepton flavors (+ neutrinos)





- LFU states these flavors of leptons must behave identically, aside from their different masses

The R(K) Measurement

To test LFU, we want an identical measurement for electrons and muons

At the LHC, we can find B meson decays that are really rare

- B decays with a kaon and non-resonant lepton pair (< 1 out of 2 million)
- Suppressed at tree-level by the standard model ➔ extra sensitive to new physics

The R(K) Measurement

To test LFU, we want an identical measurement for electrons and muons

At the LHC, we can find B meson decays that are really rare

- B decays with a kaon and non-resonant lepton pair (< 1 out of 2 million)
- Suppressed at tree-level by the standard model ➔ extra sensitive to new physics

Build a ratio:

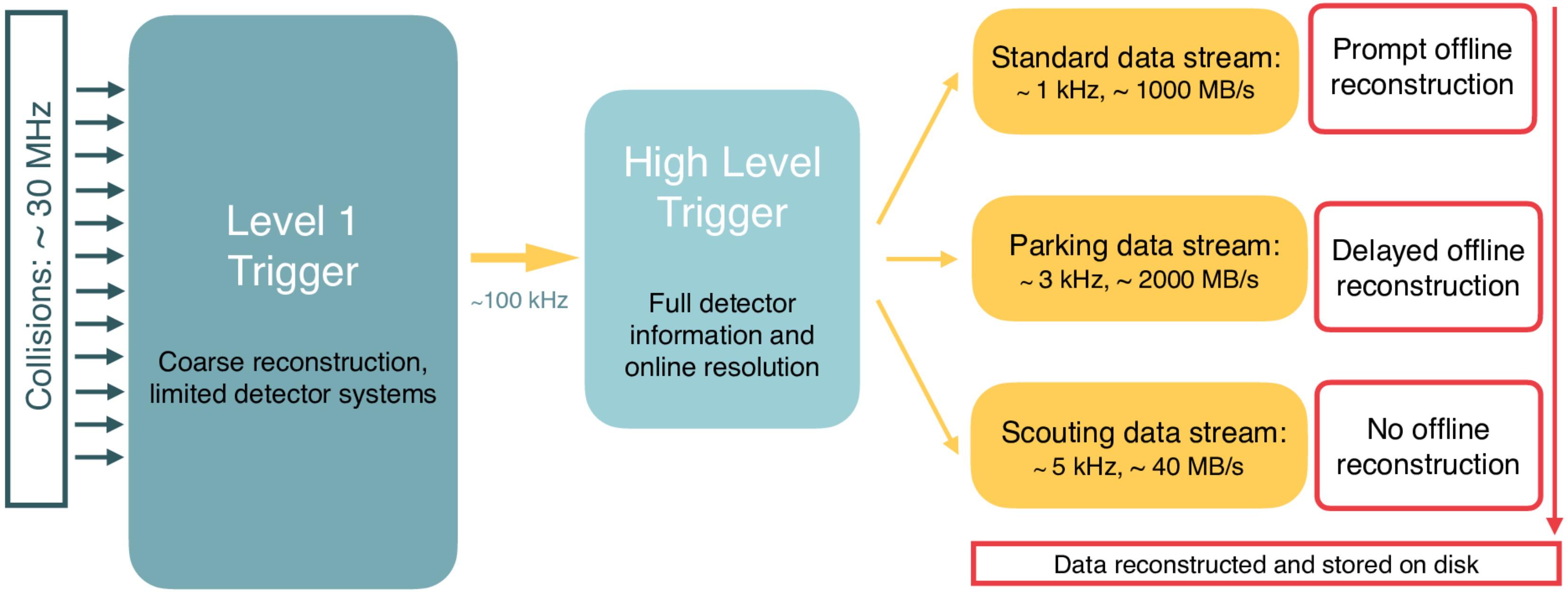
$$R(K) = \frac{\#\text{ of } B^{\pm} \rightarrow K^{\pm}\,\mu\,\mu \text{ decays}}{\#\text{ of } B^{\pm} \rightarrow K^{\pm}\,e\,e \text{ decays}}$$

R(K) $\begin{cases} = 1 \text{ means a confirmation of the SM} \\ \neq 1 \text{ could mean new physics Beyond the Standard Model (BSM)} \end{cases}$

Our Measurement – Unique Data-Taking Strategies

B Parking

- There is a data bottleneck during offline reconstruction*
- We can save more B decays by "parking" the data on separate storage, waiting to reconstruct it



"Offline Reconstruction"

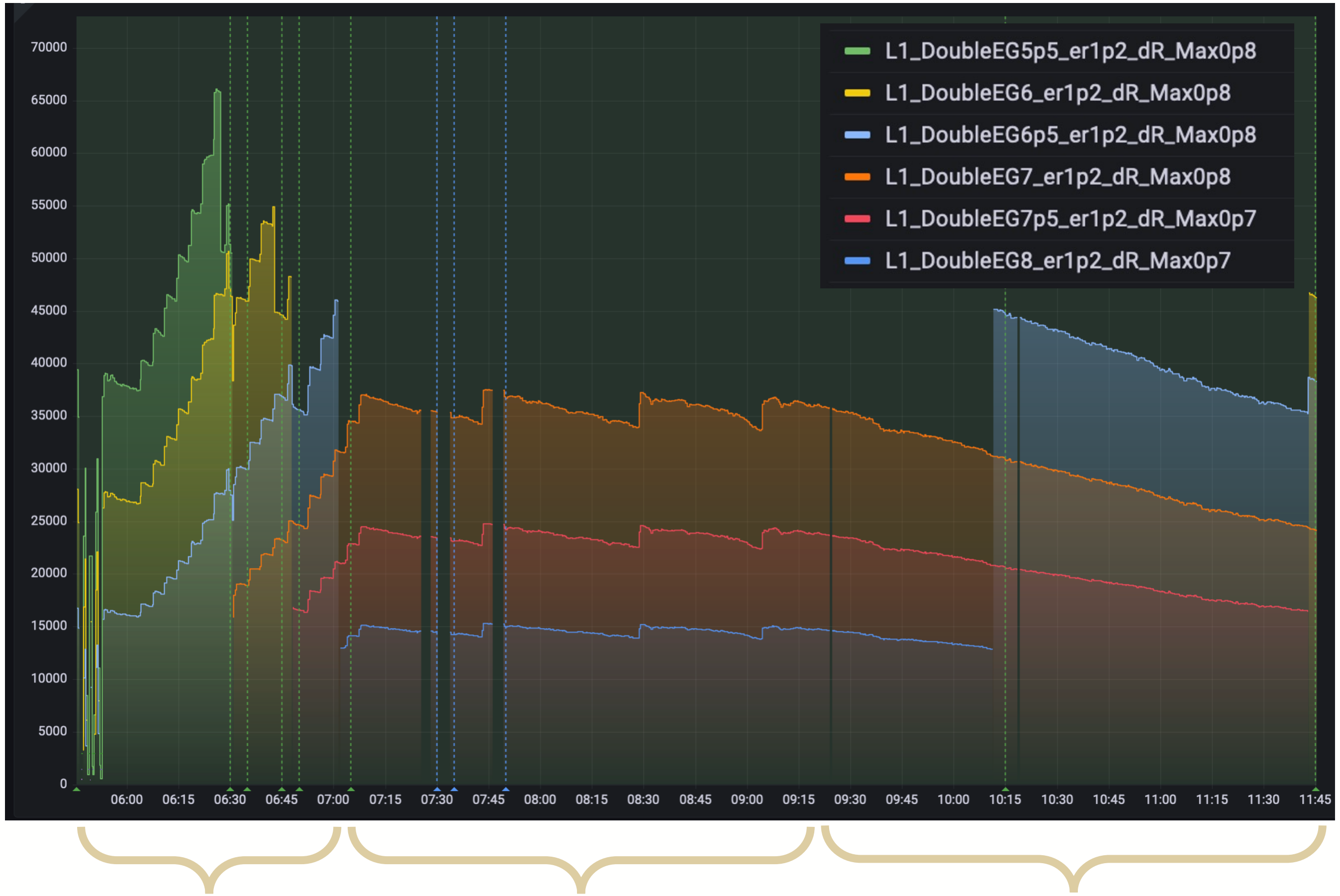
Data is sent to the global computing grid, where it's processed from simplified objects to fully calibrated physics objects

Our Measurement – Unique Data-Taking Strategies

Dynamic trigger scaling

- Need data with loose energy thresholds
- Always keeping thresholds loose saves too many events
- Use full L1T bandwidth by shifting thresholds as the luminosity* changes

Screenshot of a 2022 CMS Data-Taking Monitor (rate vs. time)



"Luminosity"

The number of collisions happening over time. This changes based on how many protons are in the beams and how "head-on" the beams are colliding.

How Do We Measure R(K)?

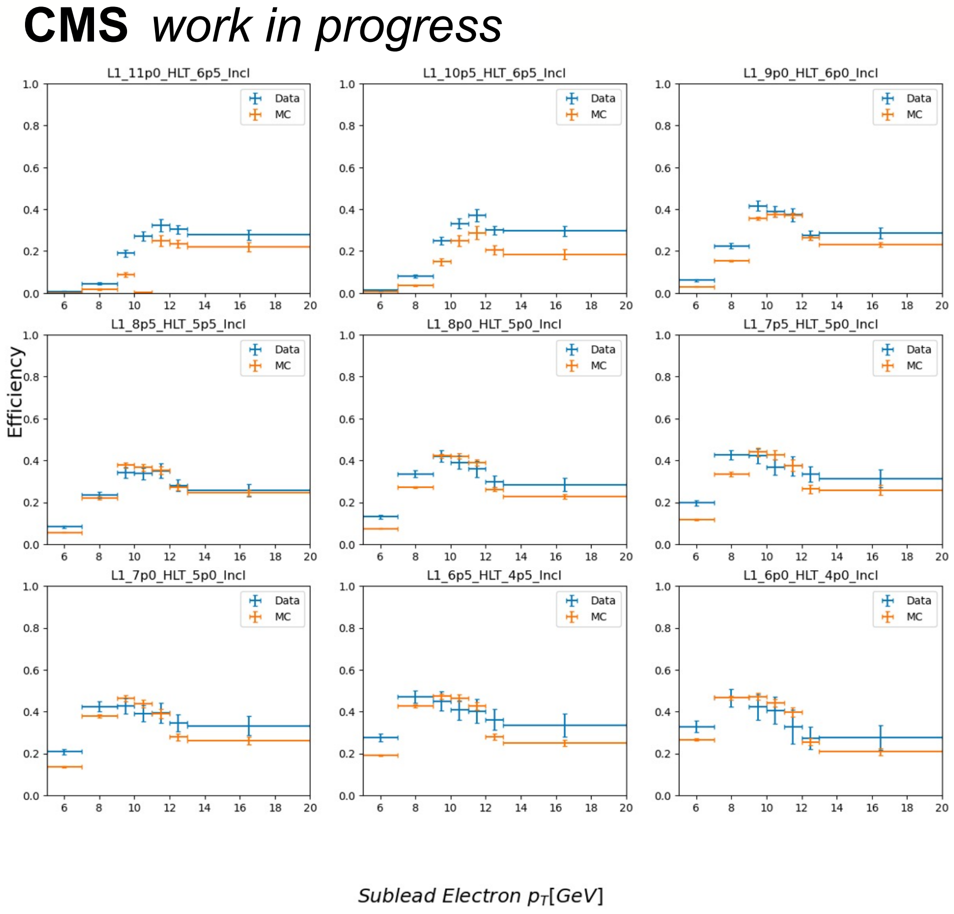
Simplified Analysis Steps

How Do We Measure R(K)?

Simplified Analysis Steps

Figure out how to collect the data

- Trigger strategy and characterization

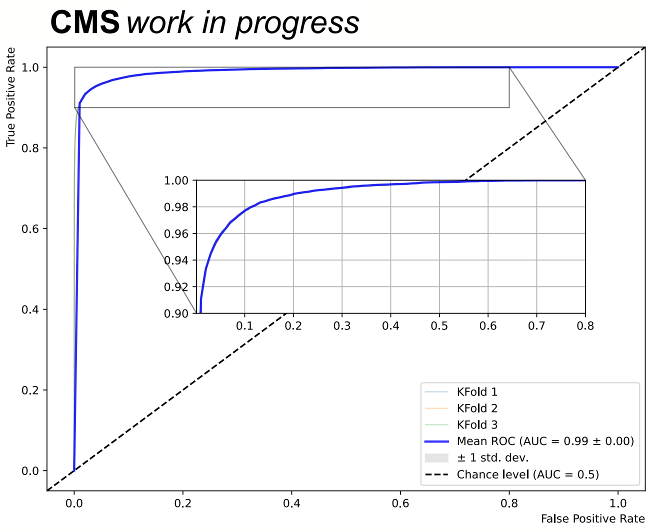


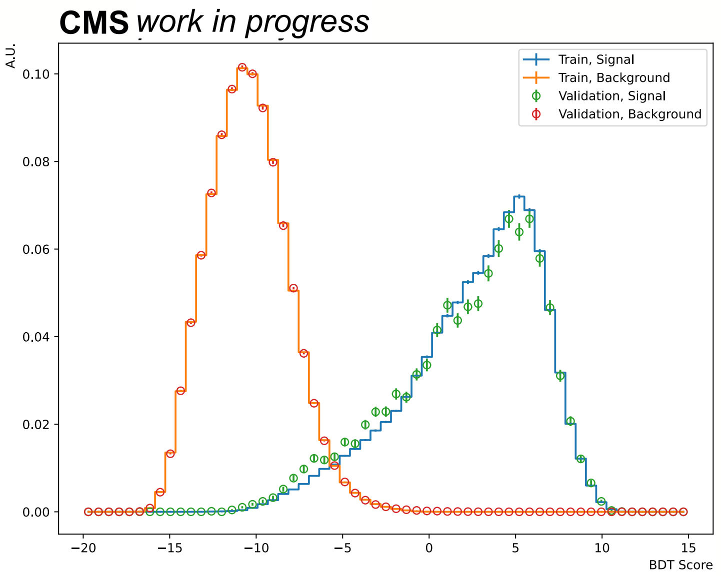
How Do We Measure R(K)?

Simplified Analysis Steps

Figure out how to collect the data
- Trigger strategy and characterization

Cut out as much noise as possible
- Object selection, kinematic fitting, and cuts
- Event selection Boosted Decision Tree (BDT)





How Do We Measure R(K)?

Simplified Analysis Steps

Figure out how to collect the data

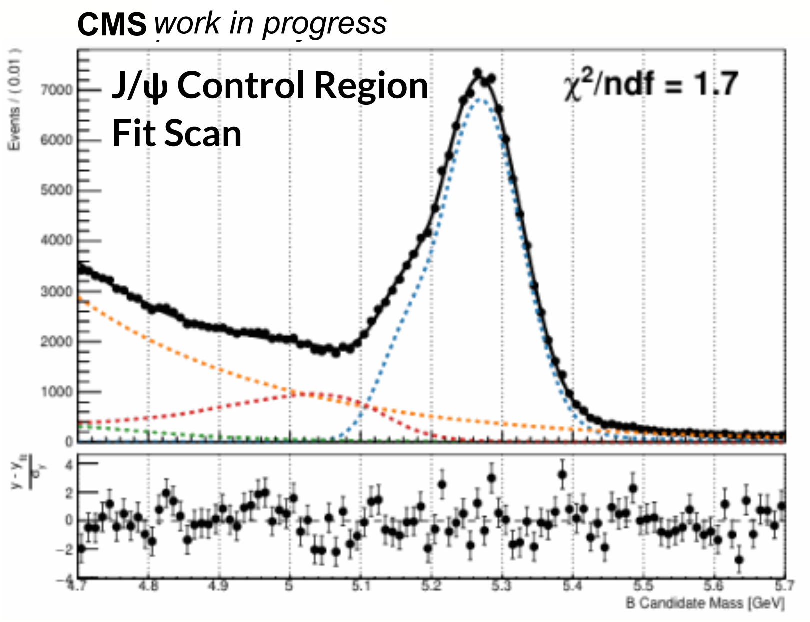
- Trigger strategy and characterization

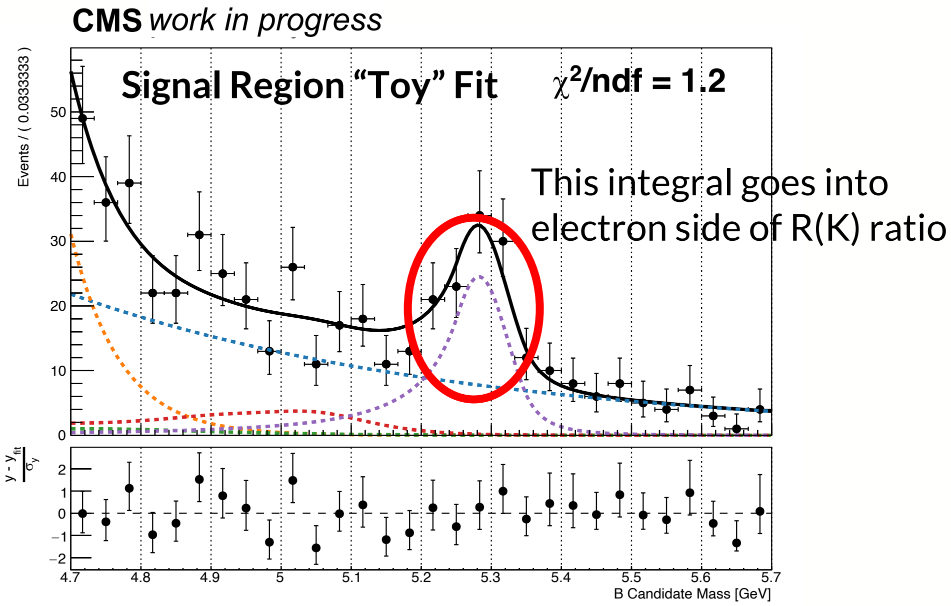
Cut out as much noise as possible

- Object selection, kinematic fitting, and cuts
- Event selection Boosted Decision Tree (BDT)

Plot data and fit signal + background model

- Signal yield from fit goes into R(K) ratio





How Do We Measure R(K)?

Simplified Analysis Steps

Figure out how to collect the data

- Trigger strategy and characterization

Cut out as much noise as possible

- Object selection, kinematic fitting, and cuts
- Event selection Boosted Decision Tree (BDT)

Plot data and fit signal + background model

- Signal yield from fit goes into R(K) ratio

Identify systematic uncertainties

Uncertainty Table from 2018 Analysis

Source	Impact on the $R(\mathrm{K})$ ratio [%]
Background description, low-q^2 bin	1.8
Trigger turn-on	1.3
Reweighting in p_{T} and rapidity	0.9
Background description, J/ψ CR	0.6
J/ψ meson radiative tail description	0.5
Pileup	0.4
Signal shape description	0.3
Trigger efficiency	0.2
J/ψ resonance shape description	0.1
Nonresonant contribution to the J/ψ CR	0.1
Total systematic uncertainty	2.6
Statistical uncertainty in MC samples	1.7
Statistical uncertainty in data	7.5
Total uncertainty	8.1

The CMS Collaboration 2024 *Rep. Prog. Phys.* **87** 077802

Our Uncertainty Calculations

$$r_{K^{*0}/K^{+}} = \frac{BR(B^0 \to J/\psi(e^+e^-)K^{*0}) * \varepsilon_{J/\psi(e^+e^-)K^{*0}}}{BR(B^+ \to J/\psi(e^+e^-)K^{+}) * \varepsilon_{J/\psi(e^+e^-)K^{+}}}$$

$$\varepsilon_X = \frac{\text{Cut and count sum}}{\text{Total \# of MC toys}}$$

$$\text{Data/MC Fit Ratio} = \frac{\left(\frac{\varepsilon(\text{data}, \mathrm{J}/\psi)}{\varepsilon(\text{MC}, \mathrm{J}/\psi)}\right)}{\left(\frac{\varepsilon(\text{data}, \psi(2s))}{\varepsilon(\text{MC}, \psi(2s))}\right)} = 0.93 \pm .02$$

WORK IN PROGRESS

How Do We Measure R(K)?

Simplified Analysis Steps

Figure out how to collect the data

- Trigger strategy and characterization

Cut out as much noise as possible

- Object selection, kinematic fitting, and cuts
- Event selection Boosted Decision Tree (BDT)

Plot data and fit signal + background model

- Signal yield from fit goes into R(K) ratio

Identify systematic uncertainties

Publish!

How Do We Measure R(K)?

Simplified Analysis Steps

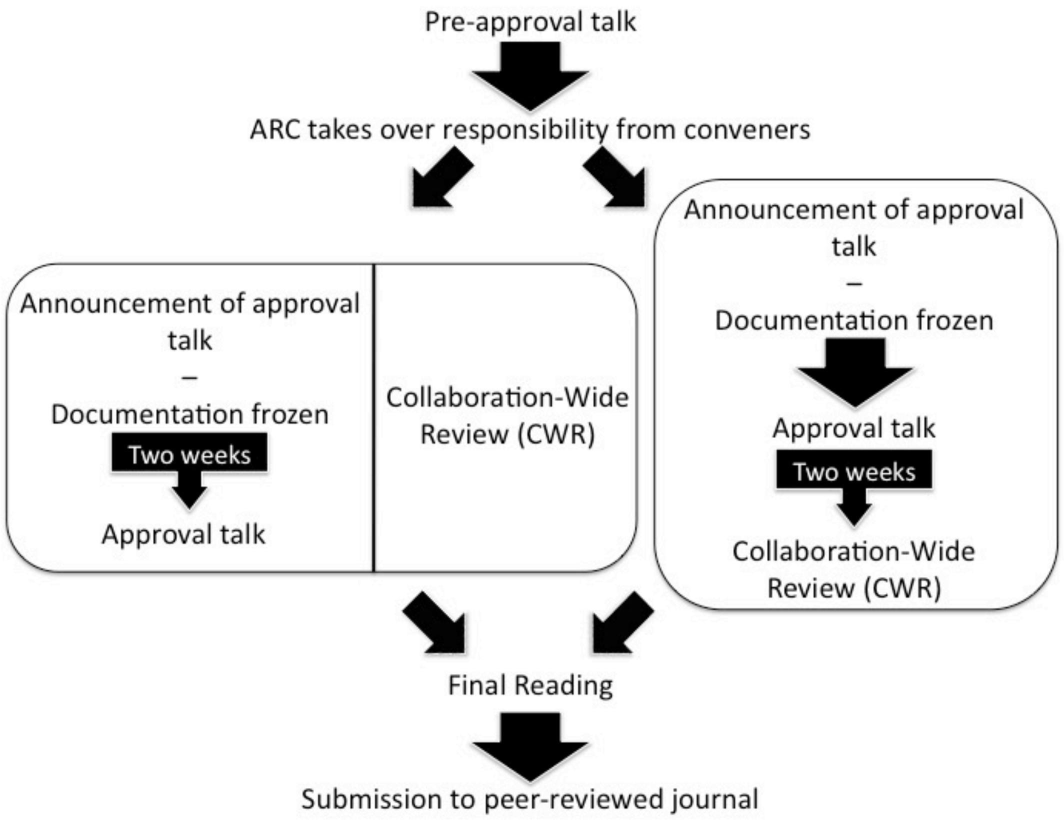
Figure out how to collect the data
- Trigger strategy and characterization

Cut out as much noise as possible
- Object selection, kinematic fitting, and cuts
- Event selection Boosted Decision Tree (BDT)

Plot data and fit signal + background model
- Signal yield from fit goes into R(K) ratio

Identify systematic uncertainties

Publish! After Review!



Thanks for Listening

Potential Takeaways

Why there’s still plenty of interesting physics at the LHC

How the CMS Level-1 Trigger works

The power of leveraging machine learning for data collection

How to test the Standard Model by probing rare decays

How a CMS analysis works

Backup

The Standard Model

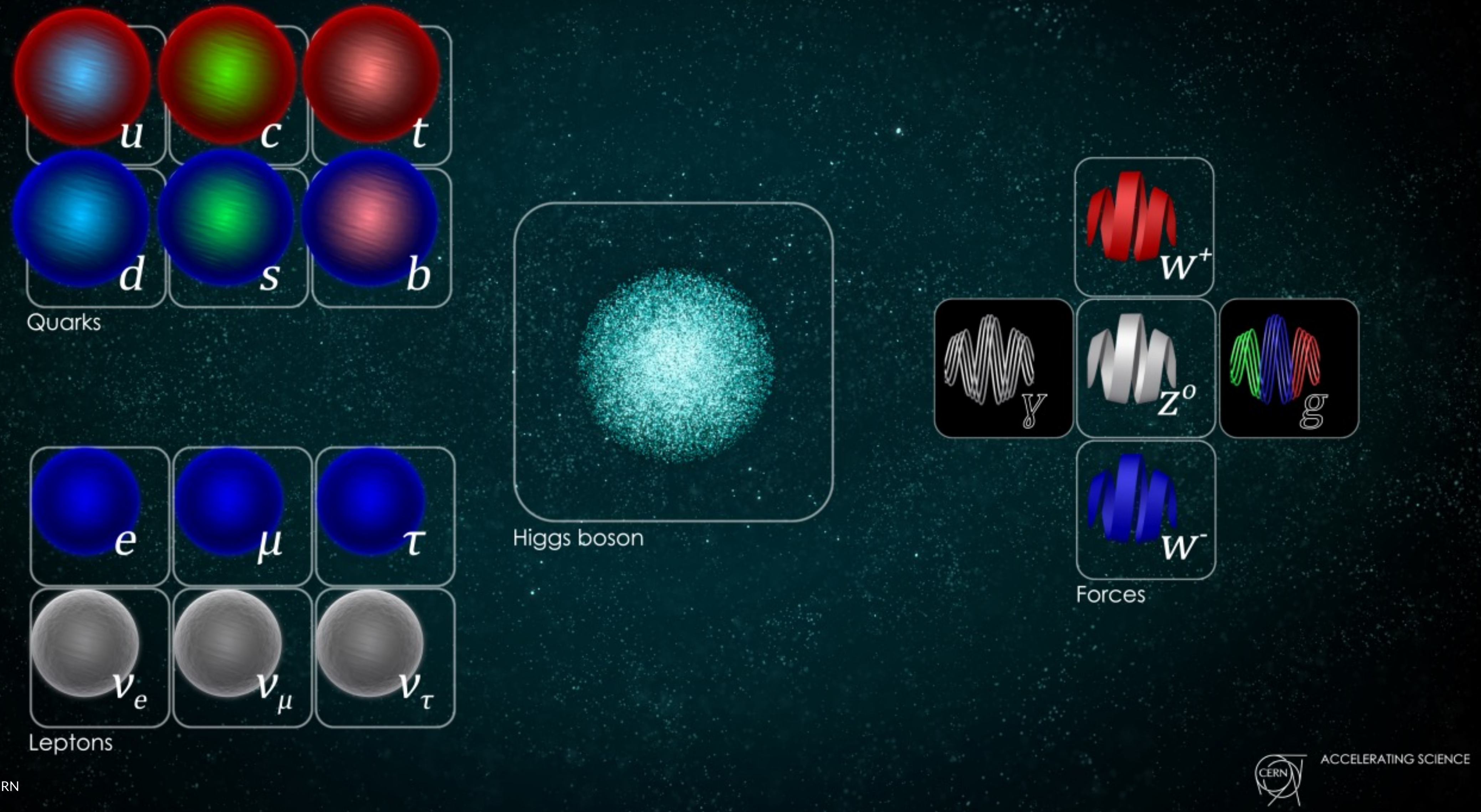


Image: Daniel Dominguez/CERN

The Full R(K) Story

Use a double-ratio

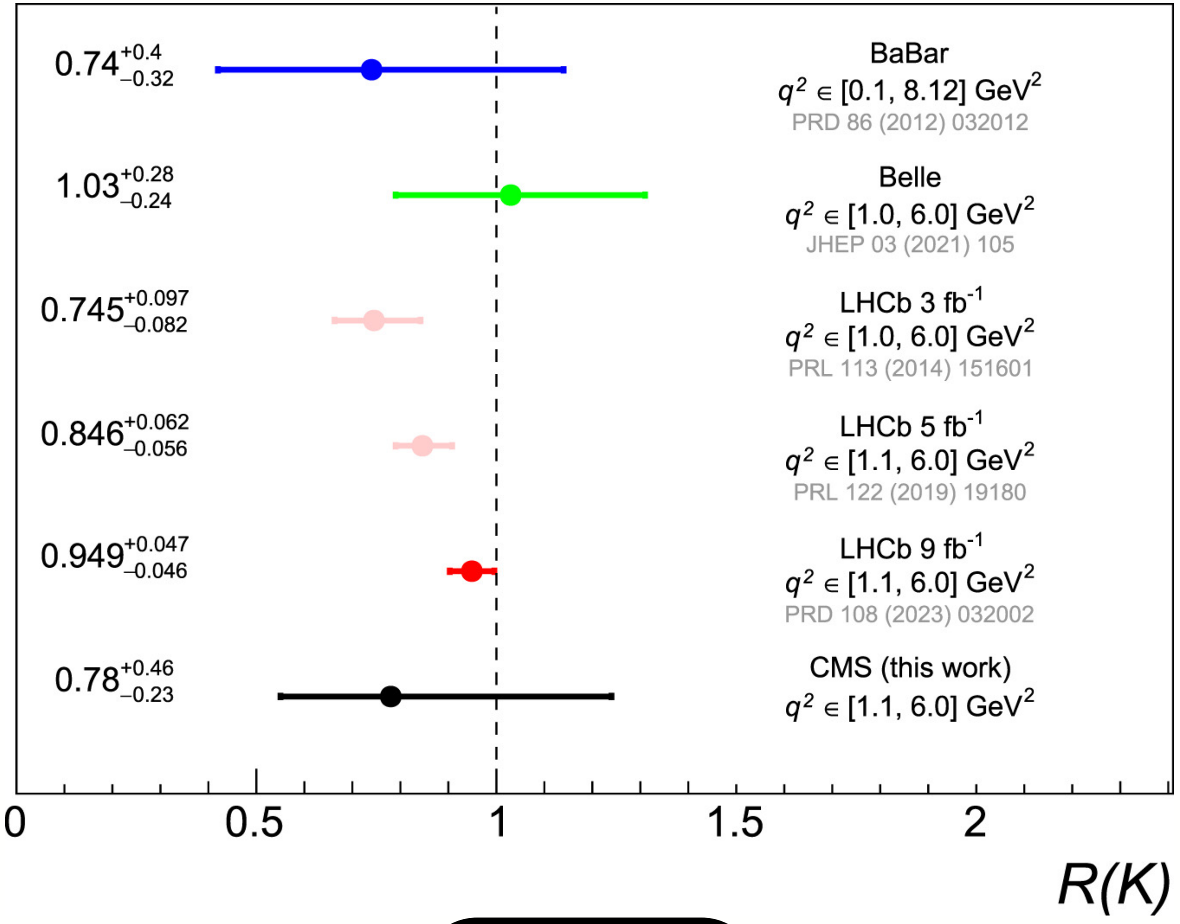
- J/ψ resonant decay ($B^+ \to J/\psi(\to e^+e^-)K^+$) is an ideal control channel
    - Similar kinematics, more events, better understood systematics
- Use the J/ψ to control for systematic uncertainty

$$R(K) = \frac{\dfrac{B^+ \to \mu^+\mu^-K^+}{B^+ \to J/\psi(\to \mu^+\mu^-)K^+}}{\dfrac{B^+ \to e^+e^-K^+}{B^+ \to J/\psi(\to e^+e^-)K^+}}$$

History of the R(K) Measurement

Been measured many different times from different experiments

Previous (anomalous) results have been superseded



CMS Run 2 RK Paper