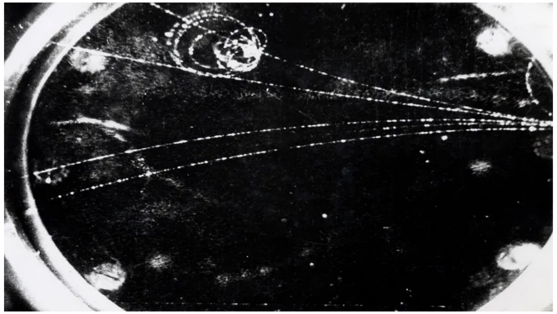
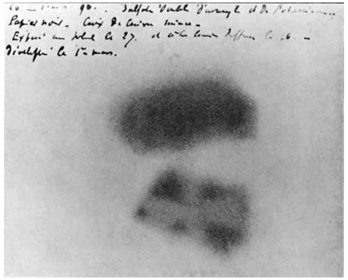
# Embracing the Unexpected:

## AI Methods for Surprise Discoveries in Particle Physics

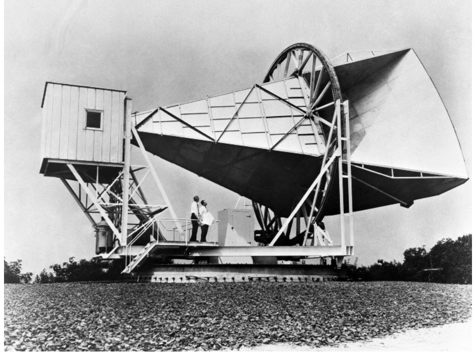
1936



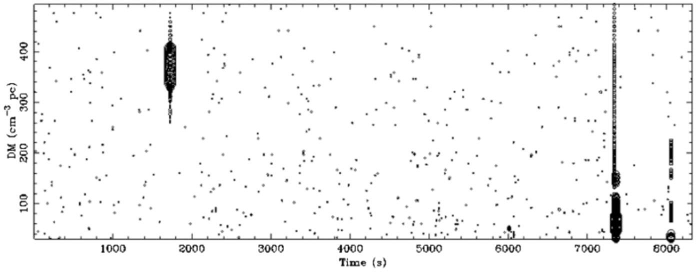
1896



1964



2006




**2026?**

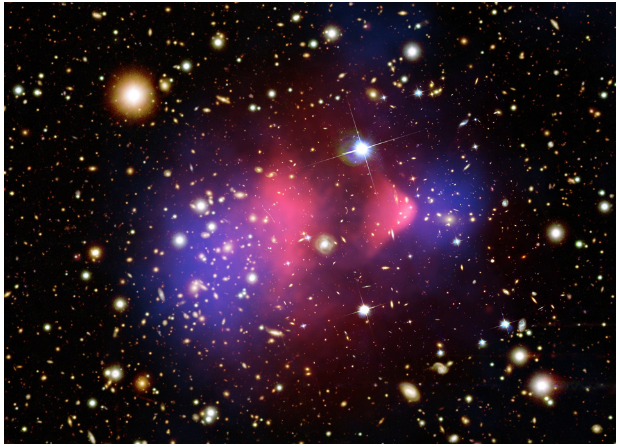

Oz Amram
Jan. 20th, 2026
SLAC Seminar

Fermilab

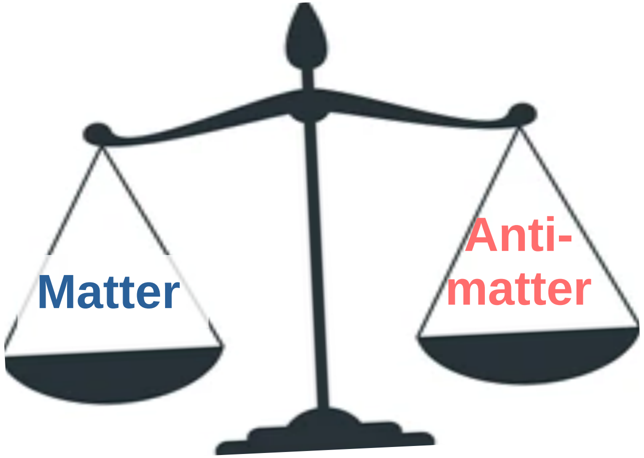
CMS

# Lots of Questions

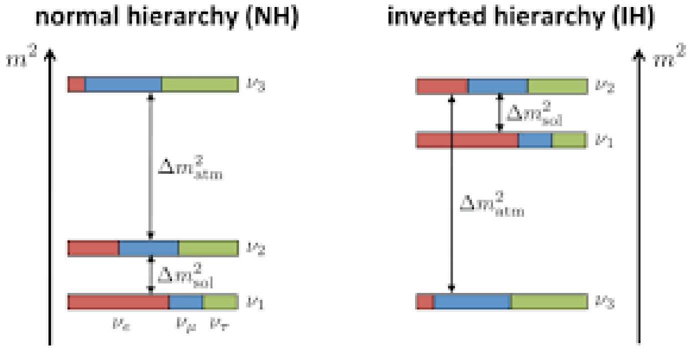
Dark Matter?



Baryogenesis?




Neutrino Mass?




Hierarchy Problem?

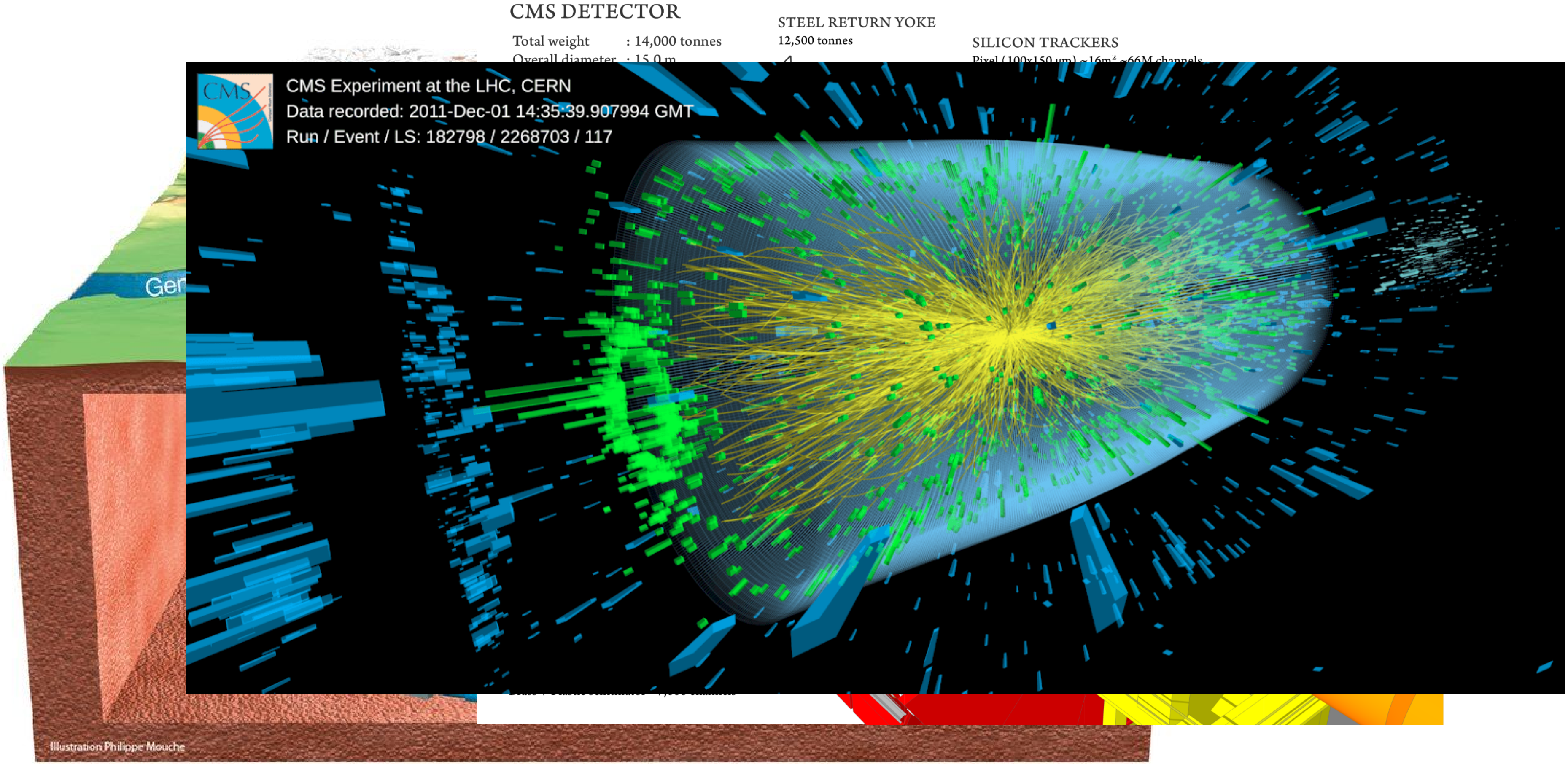
Grand Unification?

And many more…

Strong CP problem?

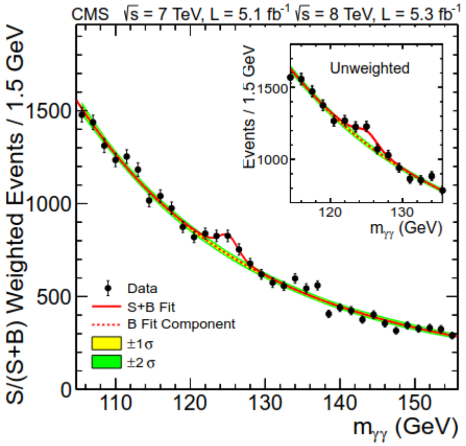
Origin of SM parameters?

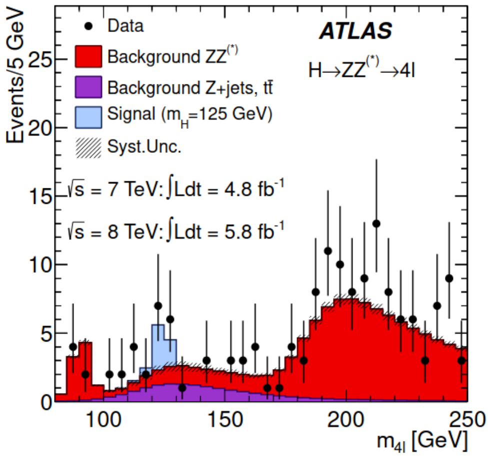
# LHC & CMS


CMS DETECTOR
Total weight : 14,000 tonnes
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
CMS Experiment at the LHC, CERN
Data recorded: 2011-Dec-01 14:35:39.907994 GMT
Run / Event / LS: 182798 / 2268703 / 117
Illustration Philippe Mouche

# 2012: A watershed

## Higgs Discovery

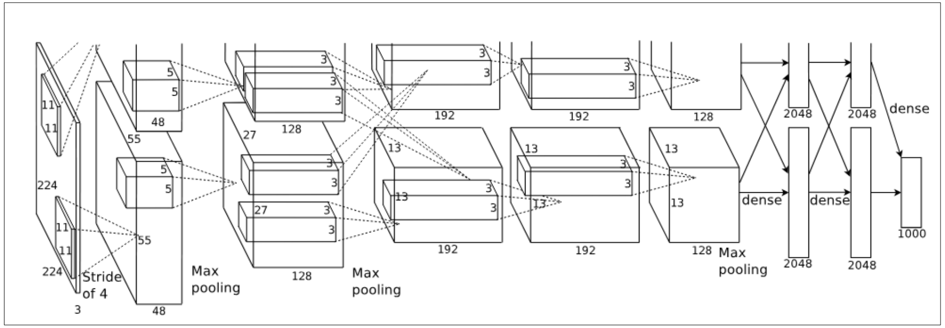



Fills in the last missing piece of the Standard Model

→ No more clear discovery targets!

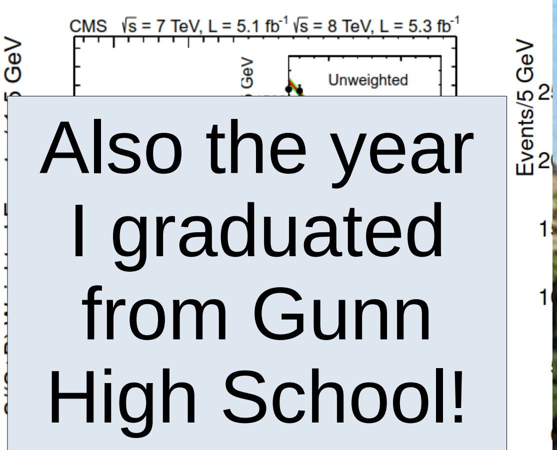
## AlexNet



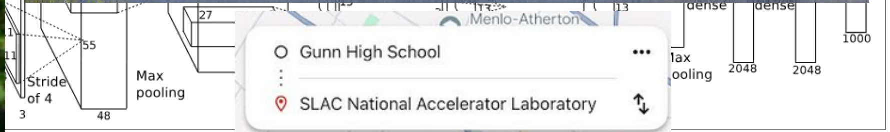

Kickstarts deep learning AI/ML revolution

# 2012: A watershed

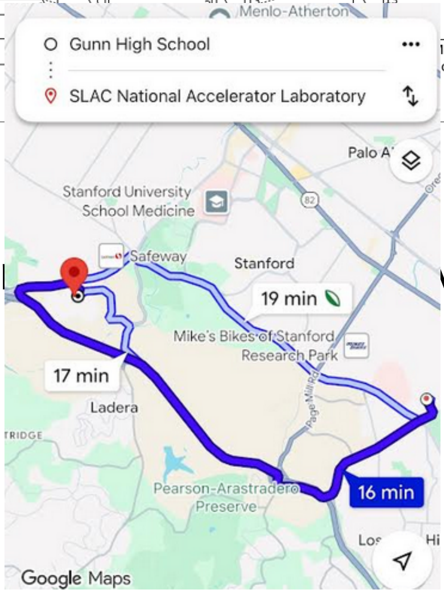
Higgs Disc




Also the year I graduated from Gunn High School!





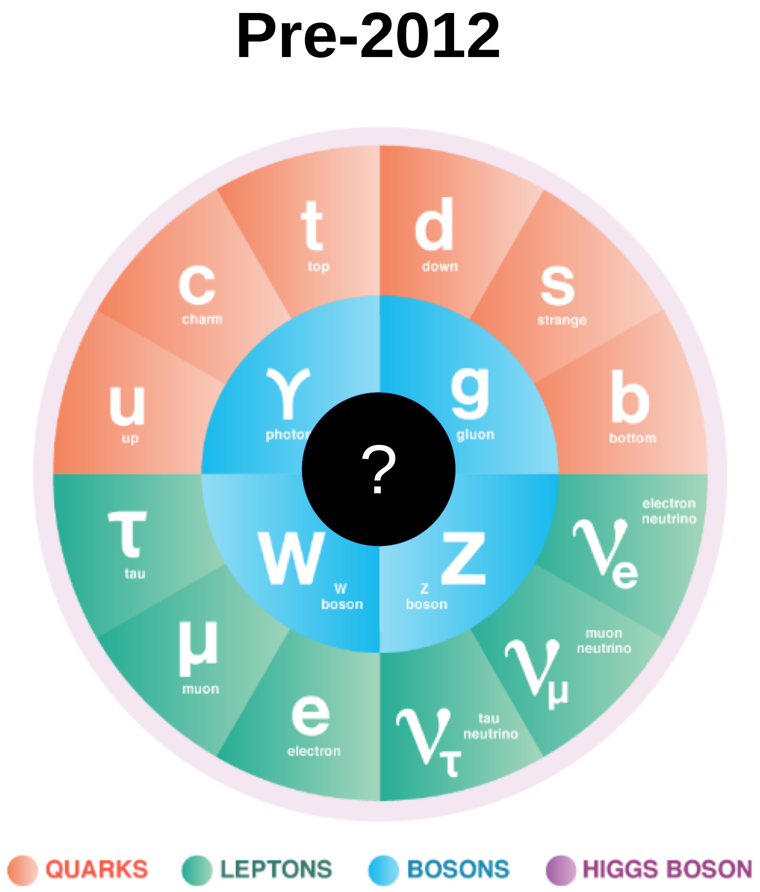


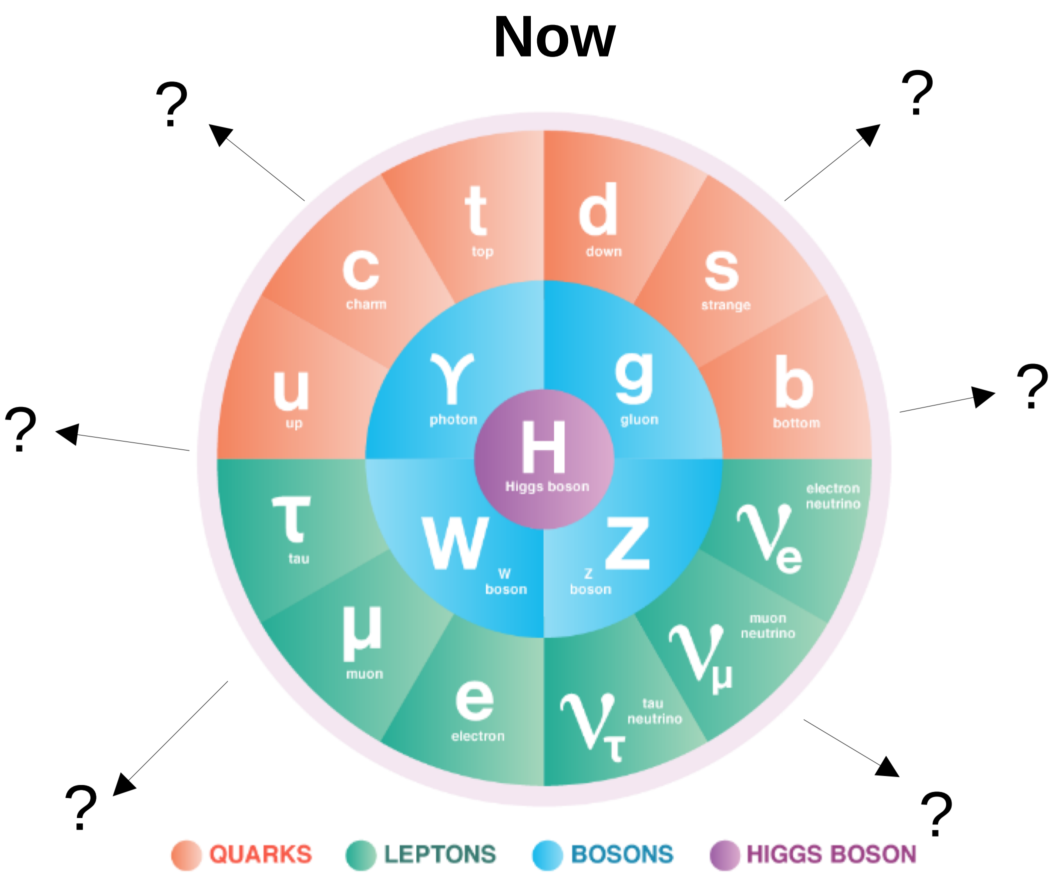

Fills in the last missing
Standard Mo

→ No more clear disco

Kicksta ML

# Before & After

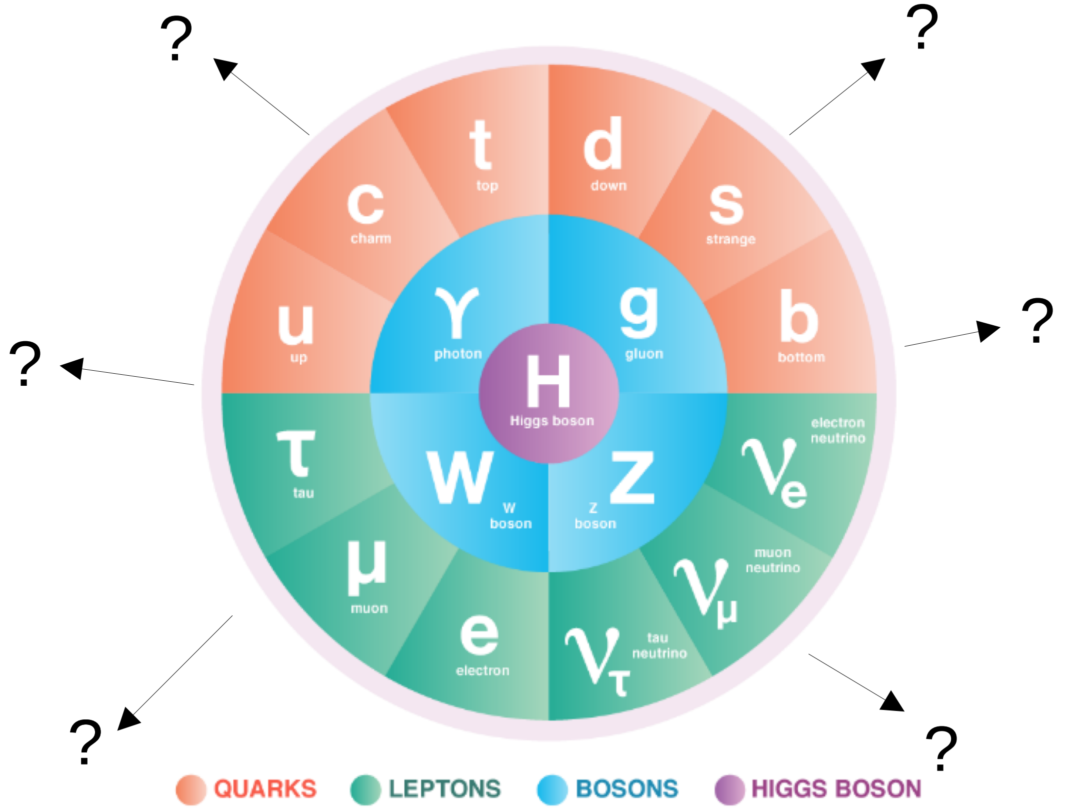


# Before & After


iml-wg.github.io/HEPML-LivingReview

Now



Lacking clear targets, we should cast our net as wide as possible

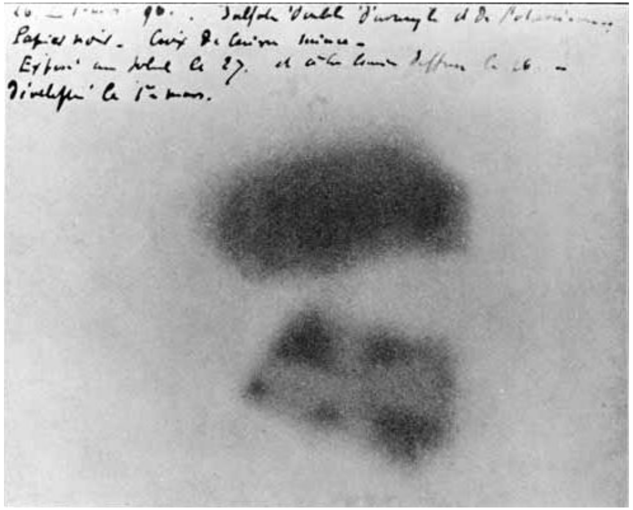
And hope to get ‘lucky’!

*“Luck is when preparation meets opportunity”*

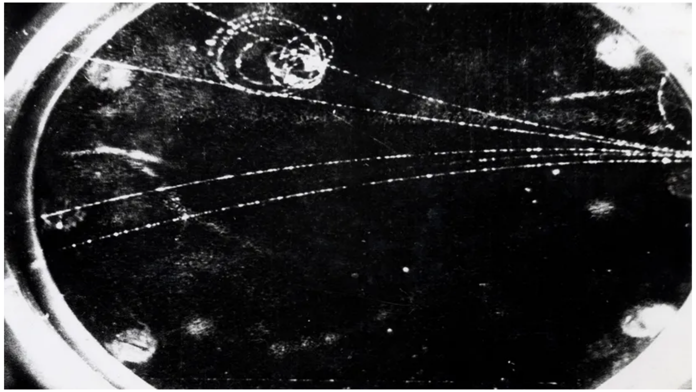
How can we prepare?

# Surprise Discoveries in Fundamental Physics

Radioactivity
(1896)

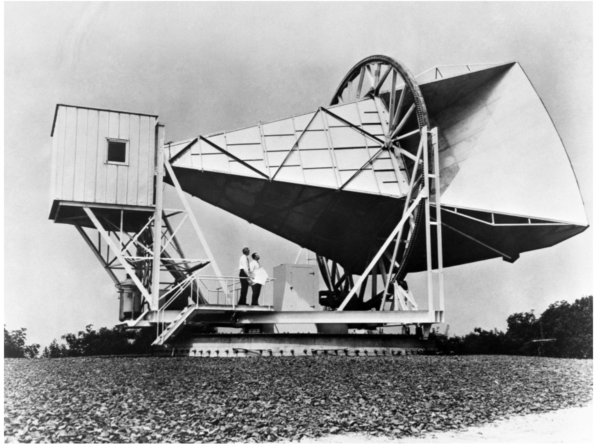


Muon
(1936)



“Who ordered that?” – I.I. Rabi

Cosmic Microwave
Background (1964)



Arguably many others as well!

*Kaon, Parity Violation, CP violation, neutrinos, neutrino oscillation, dark energy, ...*

# History of a 'lucky' discovery

*From 2024 historical reflections from discovery team (link)*

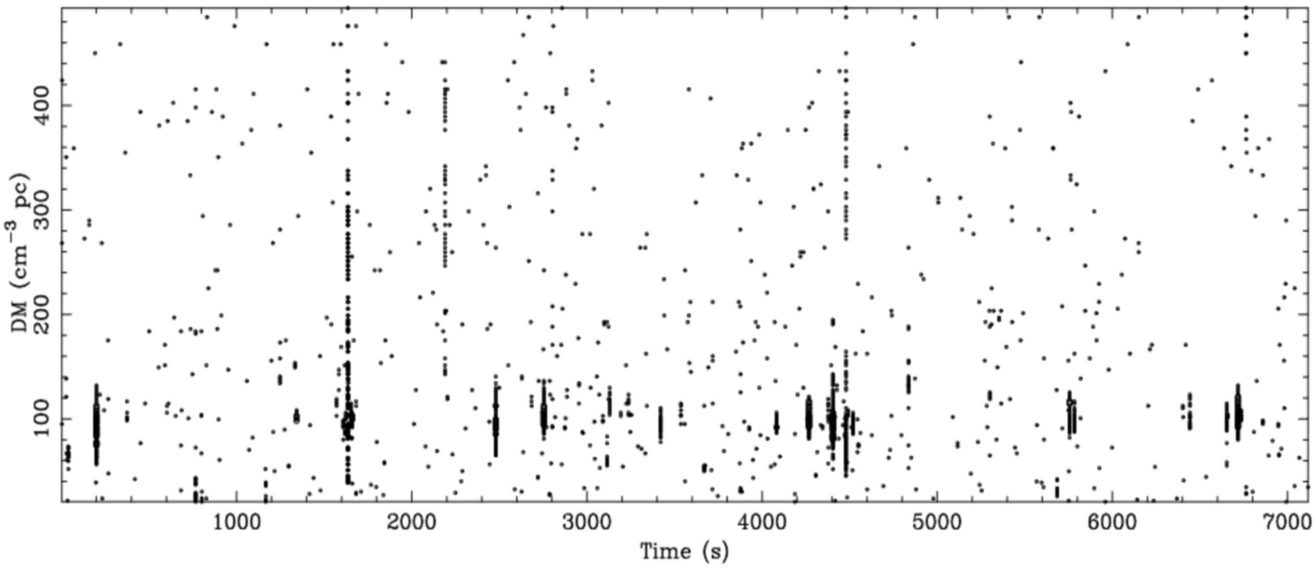
# History of a ‘lucky’ discovery

- In 2006 an undergrad was tasked with looking through archival data from the Parkes telescope to look for bright sources

*From 2024 historical reflections from discovery team (link)*

# History of a ‘lucky’ discovery

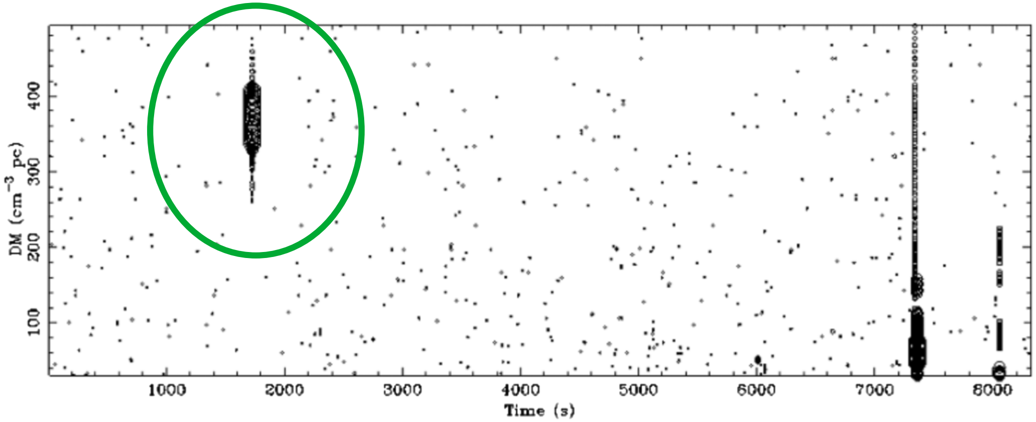
- In 2006 an undergrad was tasked with looking through archival data from the Parkes telescope to look for bright sources
- Computationally limited, they analyzed a portion of the data each week and inspected the results ‘by eye’
  - ~1 plot a week, then discussed with their advisor



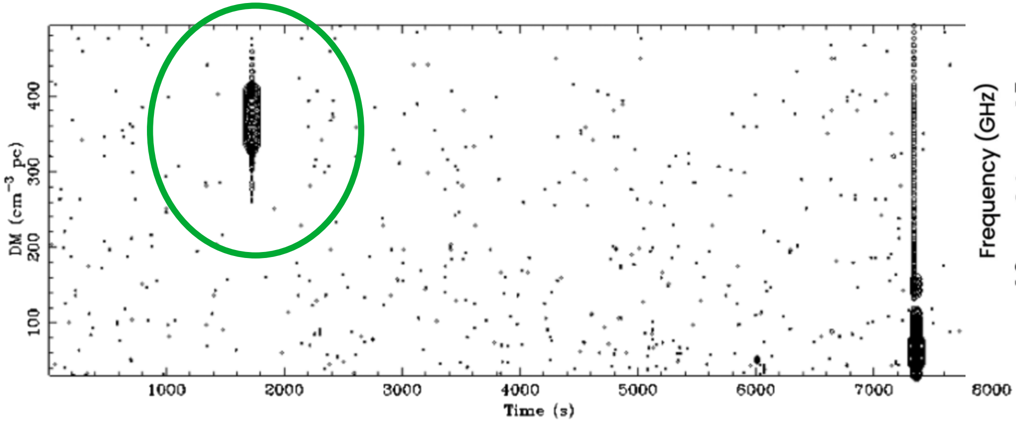

Features mostly known sources / backgrounds

*From 2024 historical reflections from discovery team (link)*

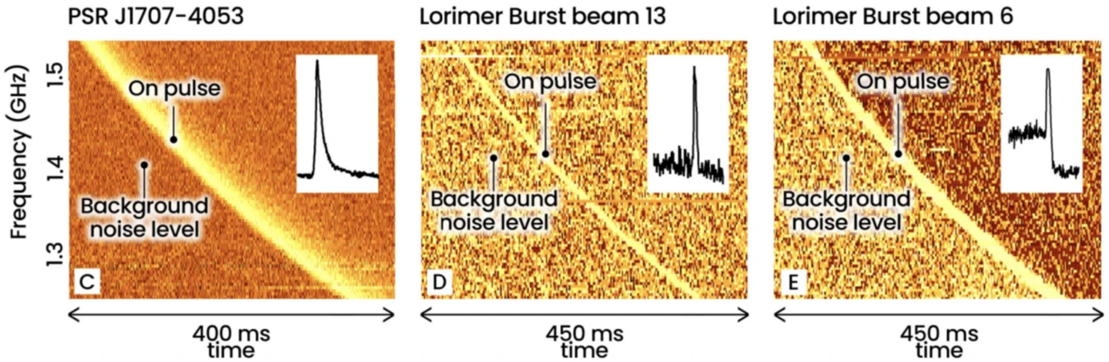
# An Anomaly!


DM (cm$^{-3}$ pc)
100
200
300
400
1000
2000
3000
4000
5000
6000
7000
8000
Time (s)

# An Anomaly!

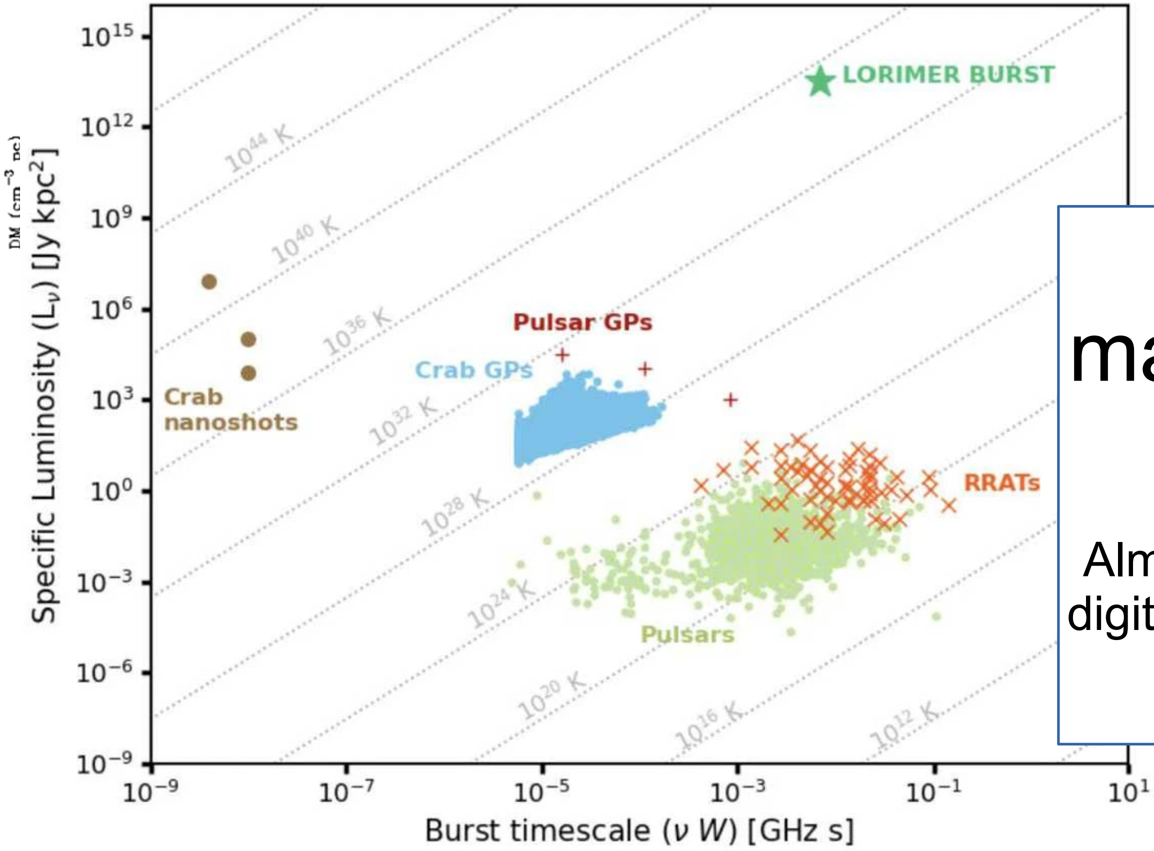


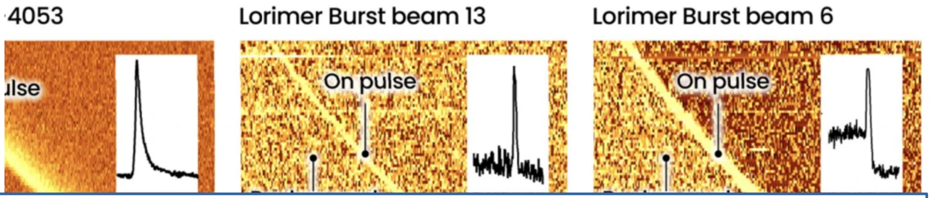
DM (cm$^{-3}$ pc)
100
200
300
400
1000
2000
3000
4000
5000
6000
7000
8000
Time (s)

PSR J1707-4053
Lorimer Burst beam 13
Lorimer Burst beam 6
Frequency (GHz)
1.3
1.4
1.5
On pulse
Background noise level
C
D
E
400 ms time
450 ms time
450 ms time

# An Anomaly!



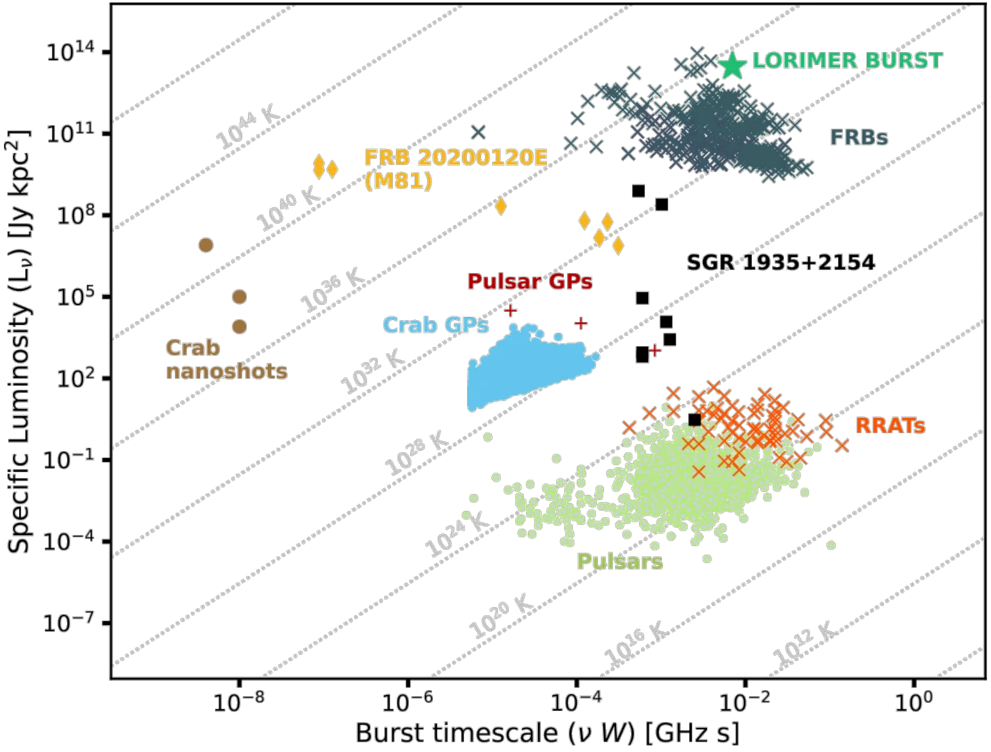




A **signal** 10 orders of magnitude larger than had ever been seen!

Almost **missed** because so bright it saturated digitization threshold of reconstruction software

# Confirmation

- This first discovery generated community interest & skepticism
- A few years later other ‘Fast Radio Bursts’ (FRB) were found by HTRU collaboration
- **Today**: several FRB’s per day recorded
- Their astrophysical origin is still poorly understood

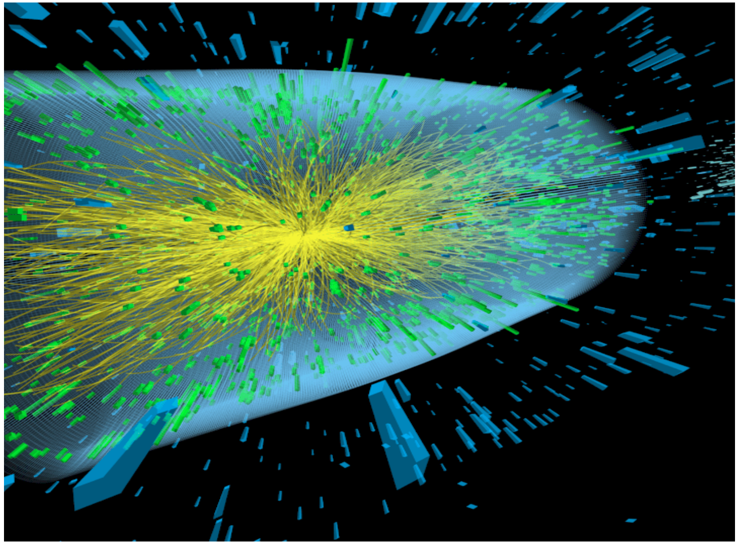


# Ingredients of a Serendipitous Discovery

- High quality scientific instrument/data
- Clear indication of outlier (luminosity)
- Understanding of backgrounds
- Short analysis timescale (weeks not years)
- A curious young scientist (undergrad) inspecting the data

# Ingredients of a Serendipitous Discovery

- High quality scientific instrument/data
- Clear indication of outlier (luminosity)
- Understanding of backgrounds
- Short analysis timescale (weeks not years)
- A curious young scientist (undergrad) inspecting the data



How can we scale these elements to modern massive, complex experiments??

# Ingredients of a Serendipitous Discovery

- High quality scientific instrument/data
- Clear indication of outlier (luminosity)
- Understanding of backgrounds
- Short analysis timescale (weeks not years)
- A curious young scientist (undergrad) inspecting the data


How can we scale these elements to modern massive, complex experiments??

## Ingredient

## AI Technology

- High quality scientific data
- Clear indication of outlier
- Understanding of backgrounds
- Short analysis timescale
- A curious young scientist inspecting the data

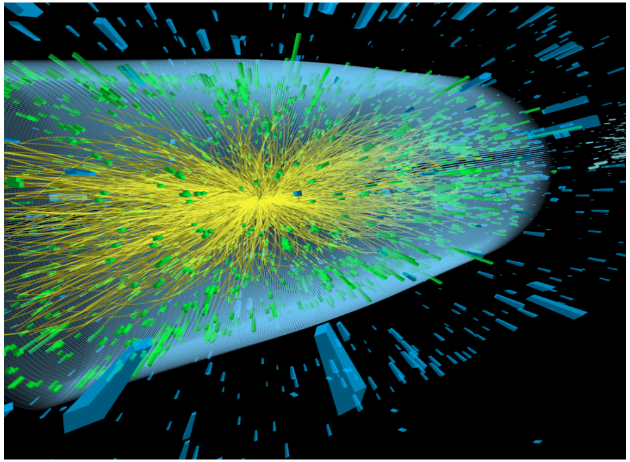
## Ingredient

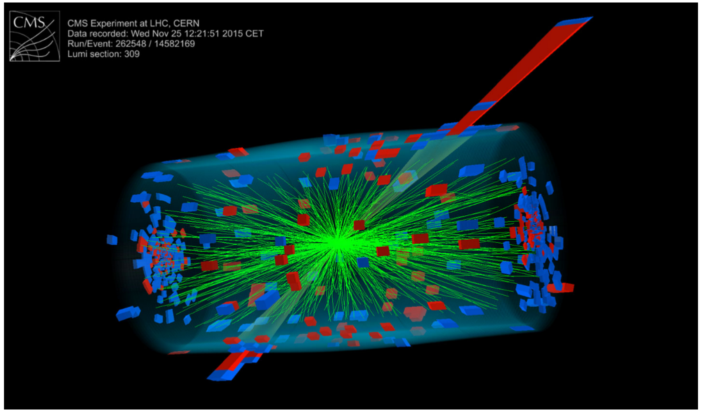
- High quality scientific data
- Clear indication of outlier
- Understanding of backgrounds
- Short analysis timescale
- A curious young scientist inspecting the data

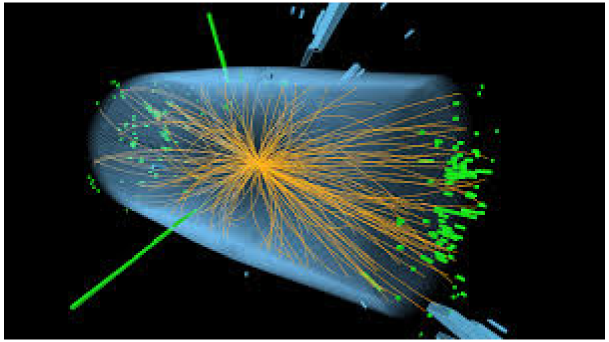
## AI Technology

Anomaly Detection

# Anomaly Detection





How do you know which of these is ‘anomalous’?

# Anomaly Detection

**The LHC Olympics 2020**

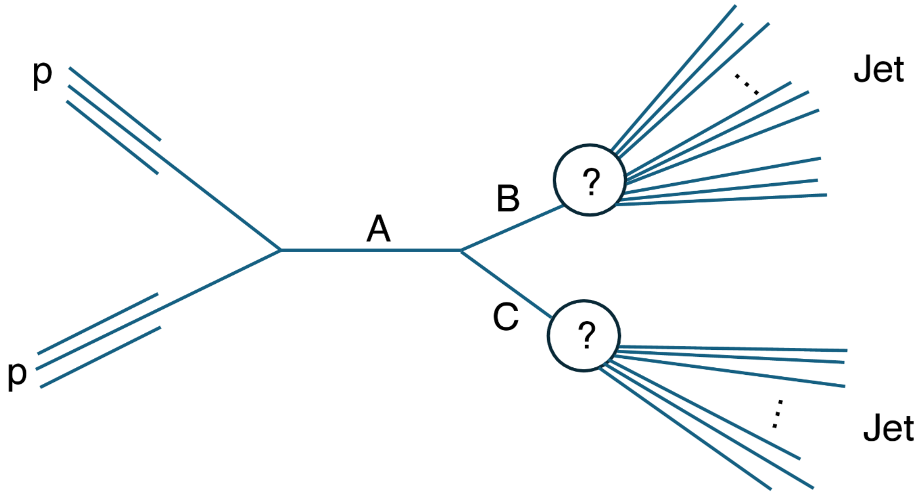
**A Community Challenge for Anomaly Detection in High Energy Physics**

- Focus on a single topology at a time
- Entirely **data-driven**
- Novel ML methods to reduce bkg

arXiv: 2101.08320

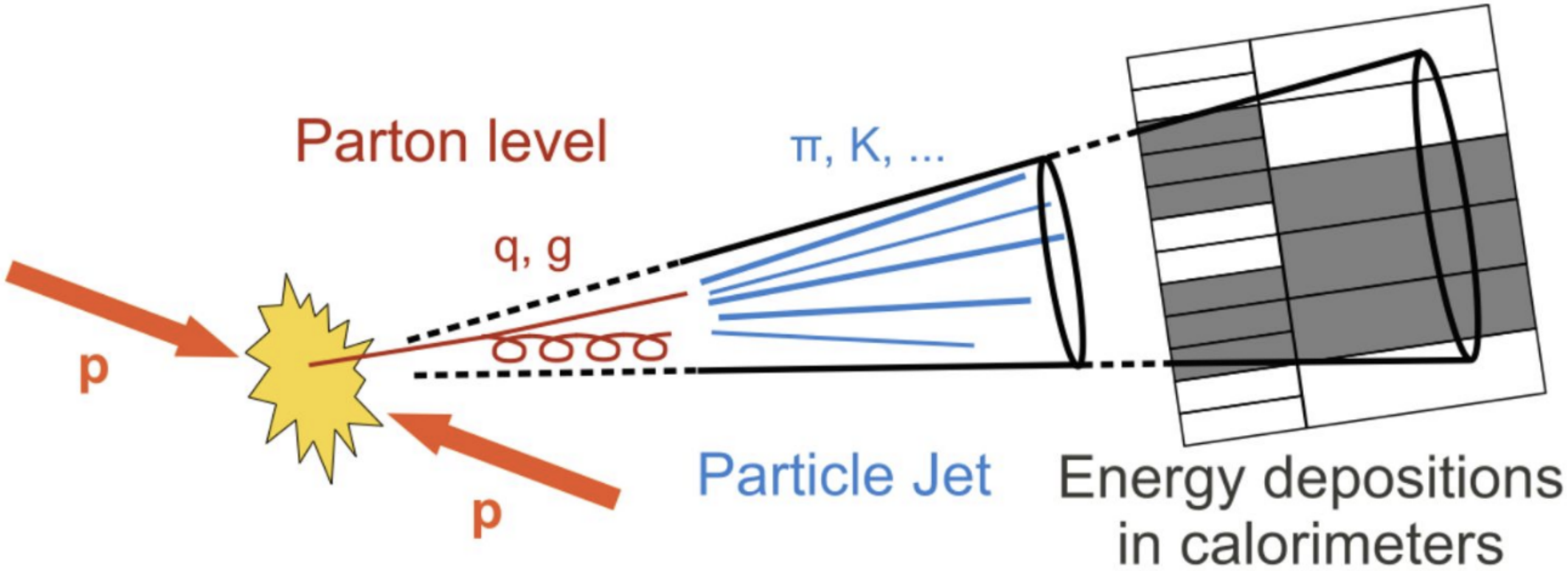
# CMS Anomaly Search

Physics result:
arXiv:2412.03747
ML details:
arXiv:2512.20395
Jet Uncertainties:
arXiv:2507.07775




- **First** CMS search to use anomaly detection
- Heavy resonance (A) → daughters B and C → 2 jets
- Employed **five** separate anomaly detection methods!

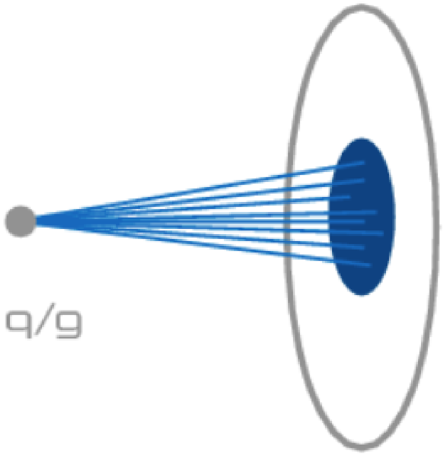
# Jets

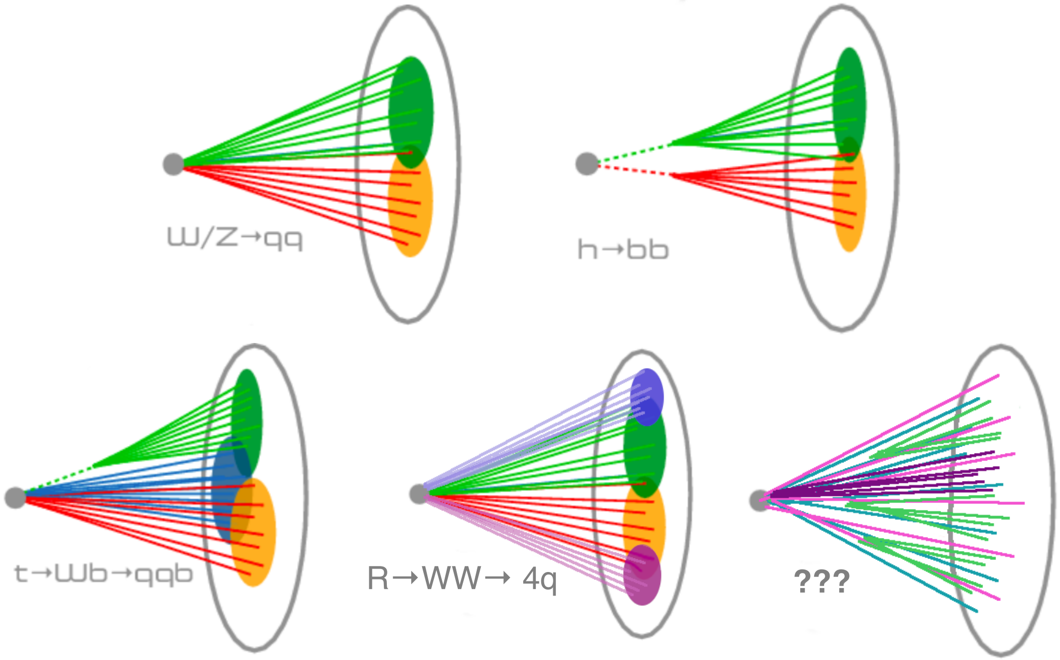



Sprays of O(50) particles produced via strong force

A playground for HEP-ML!

# Don't Judge a Jet by its Cover



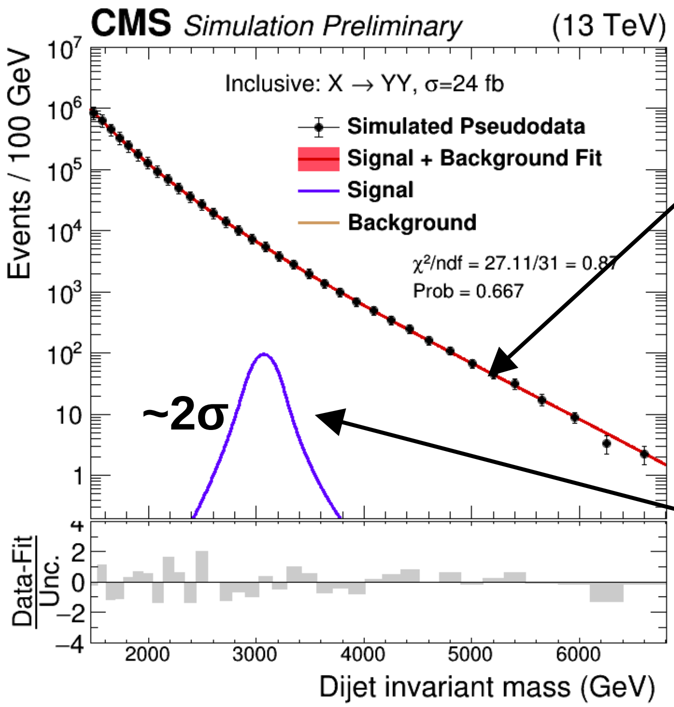

**Typical jet**

- One central axis (prong)
- From primary vertex
- ...

**Anomalous jets**

- Multiple prongs
- Displaced vertices
- ???

# The Bump Hunt

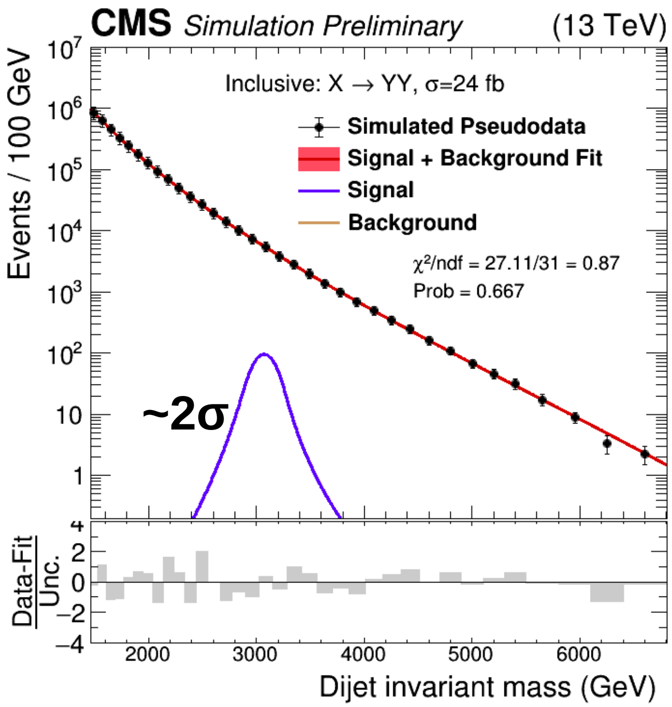



Smoothly falling background

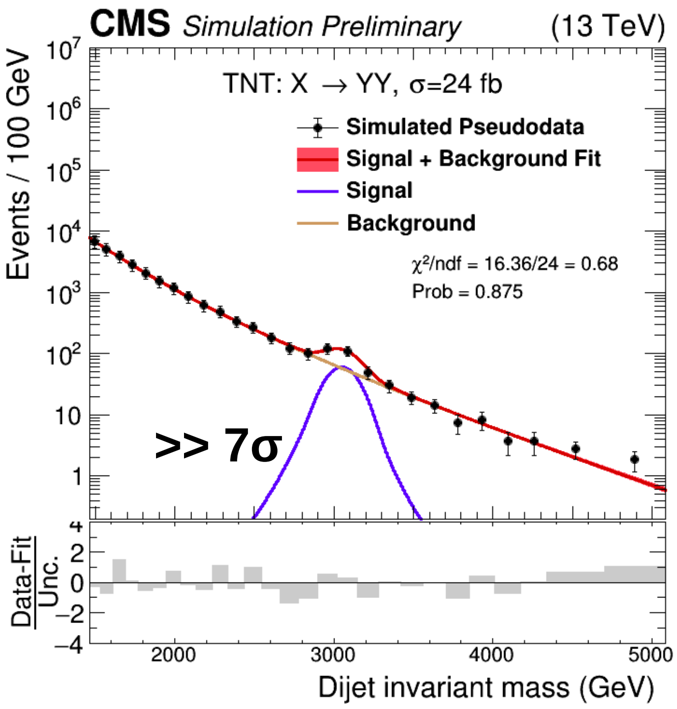
Localized resonance 'bump'

Without any substructure cuts →
Signal swamped by QCD background…

# The Bump Hunt



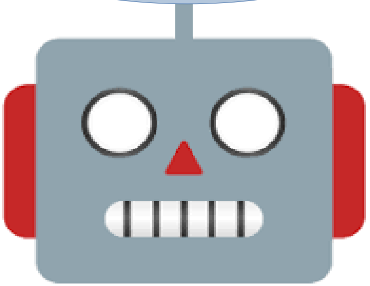

Cut on anomaly score




**Anomaly detection finds hidden resonance!**

How to identify anomalies?

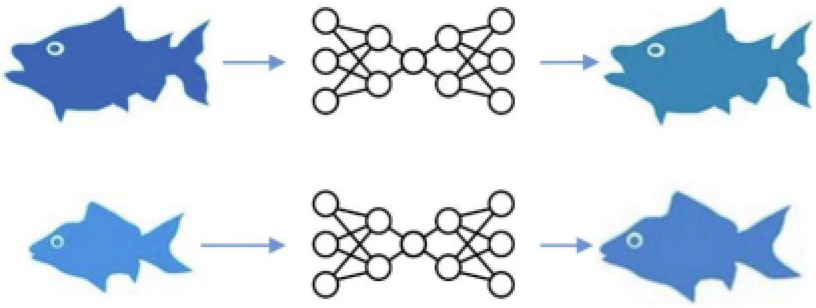
Learn your bkg → look for outliers



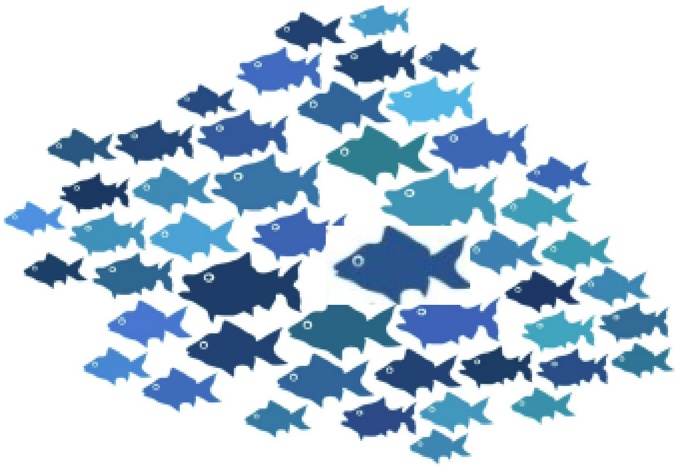
Increasing Model Dependence

# Looking for Outliers

Train 'Autoencoder'

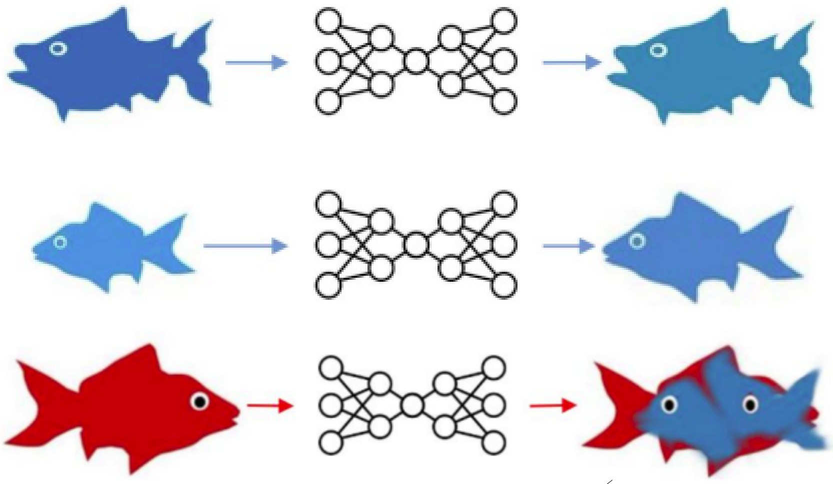


Training Sample from data sideband



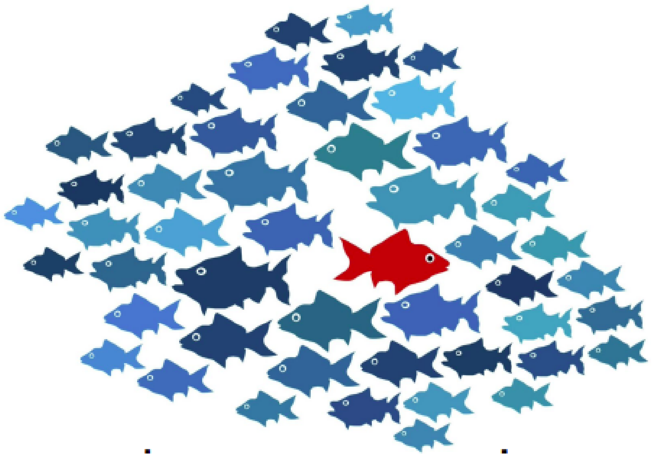
# Looking for Outliers

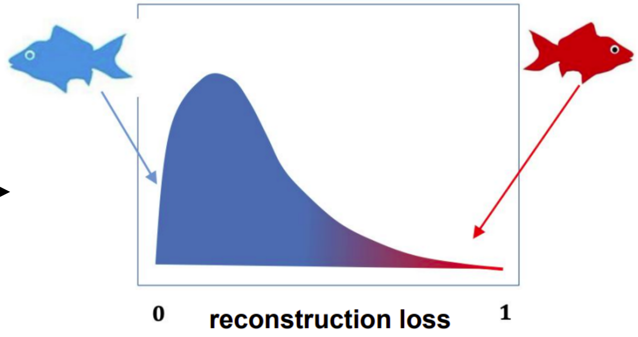
Apply Autoencoder



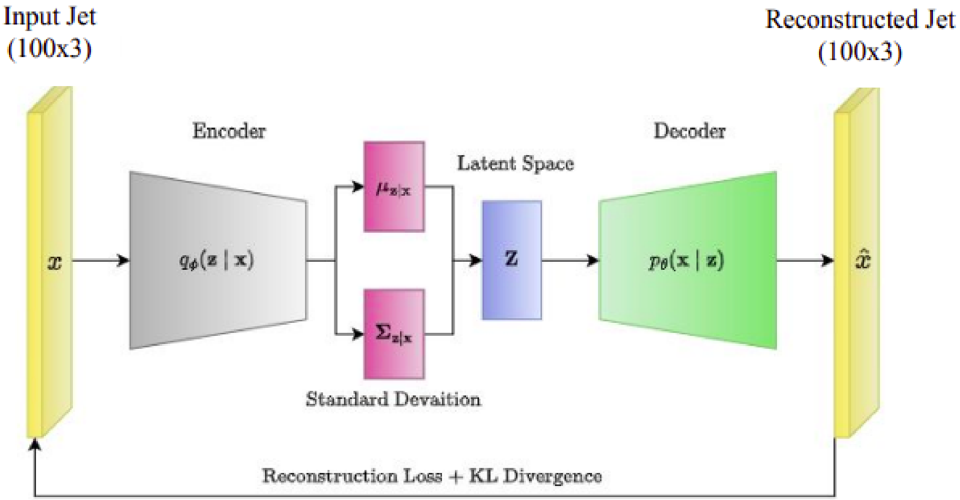
Take difference

Data from signal region





# Variational Autoencoder (VAE)

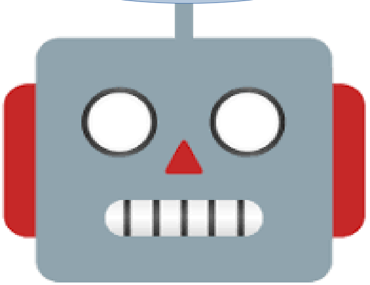



Latent space forced to be Gaussian thru additional term in loss

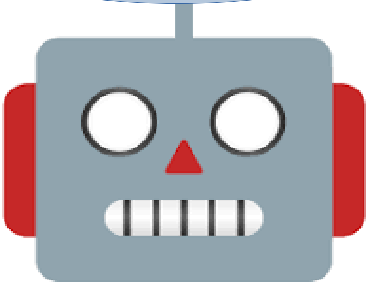
- Jet represented by up to 100 highest $p_T$ constituents ($p_x$, $p_y$, $p_z$)
- 100x3 matrix compressed to latent space of size 12
- Trained on jets from data control region

# How to identify anomalies?

Learn your bkg → look for outliers



Look for local overdensities of signal (collective anomalies)

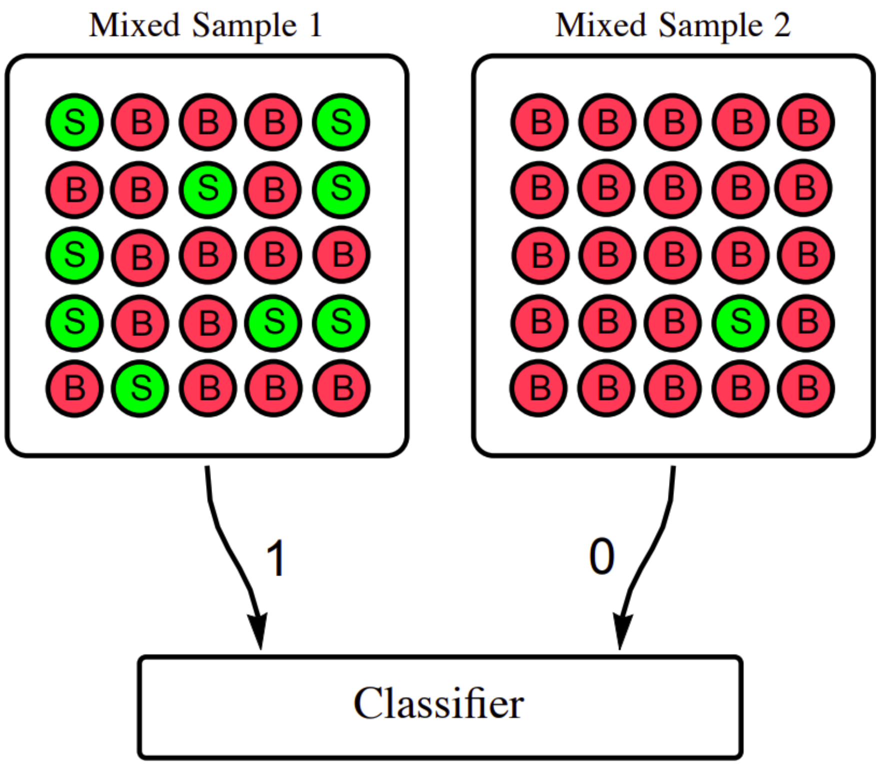


**Increasing Model Dependence**

# Weak Supervision

Aka 'Classification Without Labels' (CWoLa)

**Train on two mixed samples**




- Train a classifier between **signal-rich** and **background-rich** mixed samples

  → Learns to identify signal vs. bkg

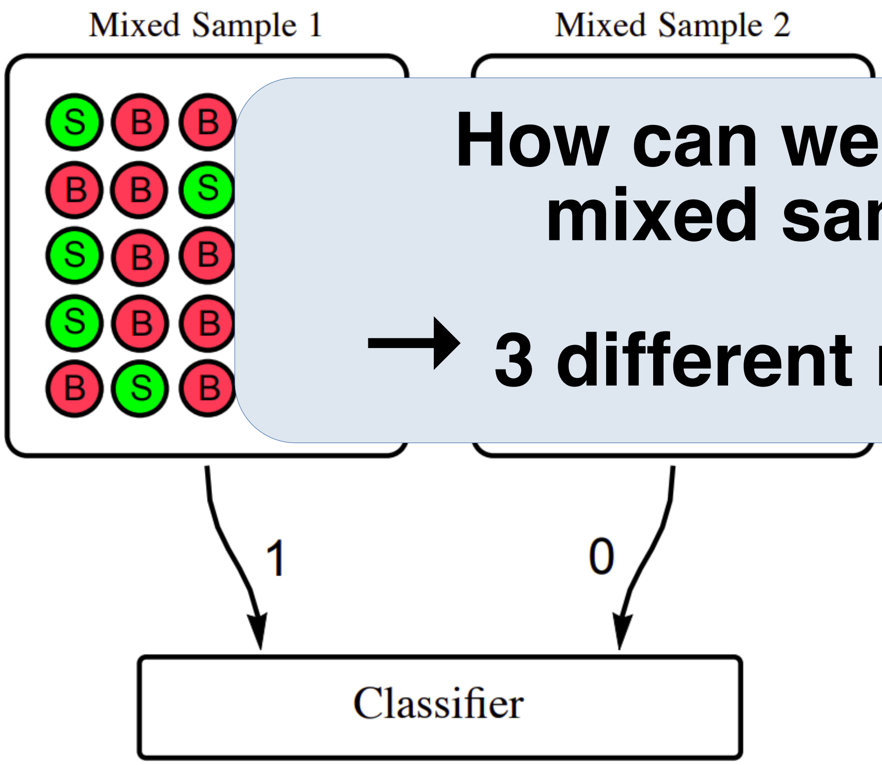
- Performance changes with amount of signal in training data

[Metodiev, Nachman, Thaler, 1708.02949]

# Weak Supervision

Aka ‘Classification Without Labels’ (CWoLa)

**Train on two mixed samples**

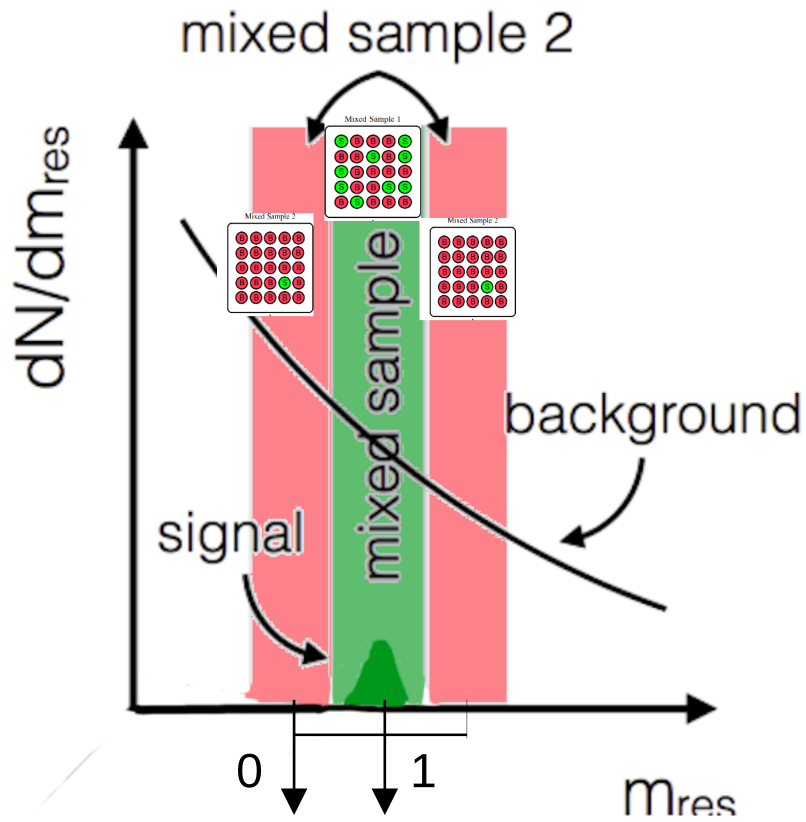



**How can we construct these mixed samples in data?**

**→ 3 different methods employed**

- Train a classifier between **signal-rich** … **nd-** … mples … dentify
- Performance changes with amount of signal in training data

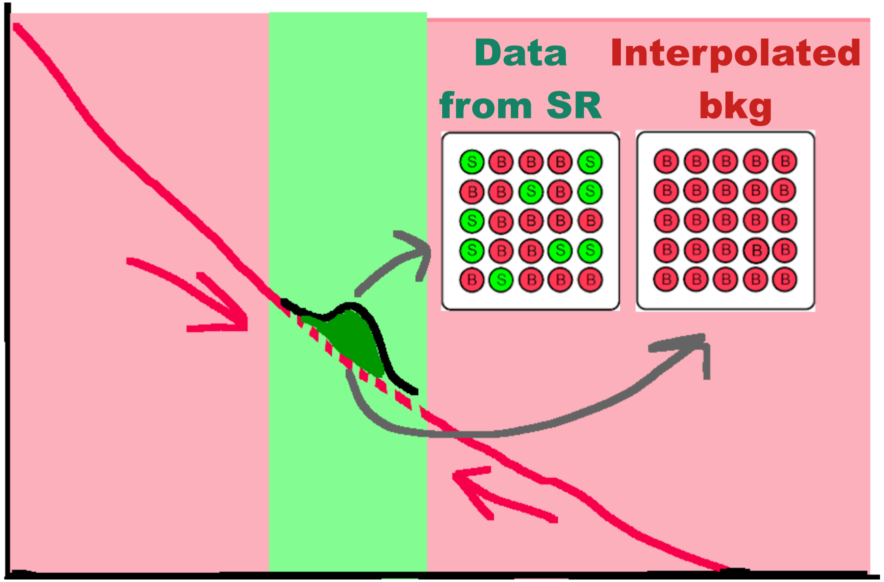
# CWoLa Hunting




- Assume signal is a **narrow** resonance
- **Guess** a mass window where it lives
  - Train signal window vs. narrow sidebands using weak supervision
- **Repeat procedure**, scanning over different mass windows
  - (2x6 windows used)
- Need to be careful about correlations btwn feats & mass

[Collins, Howe, Nachman 1902.02634]

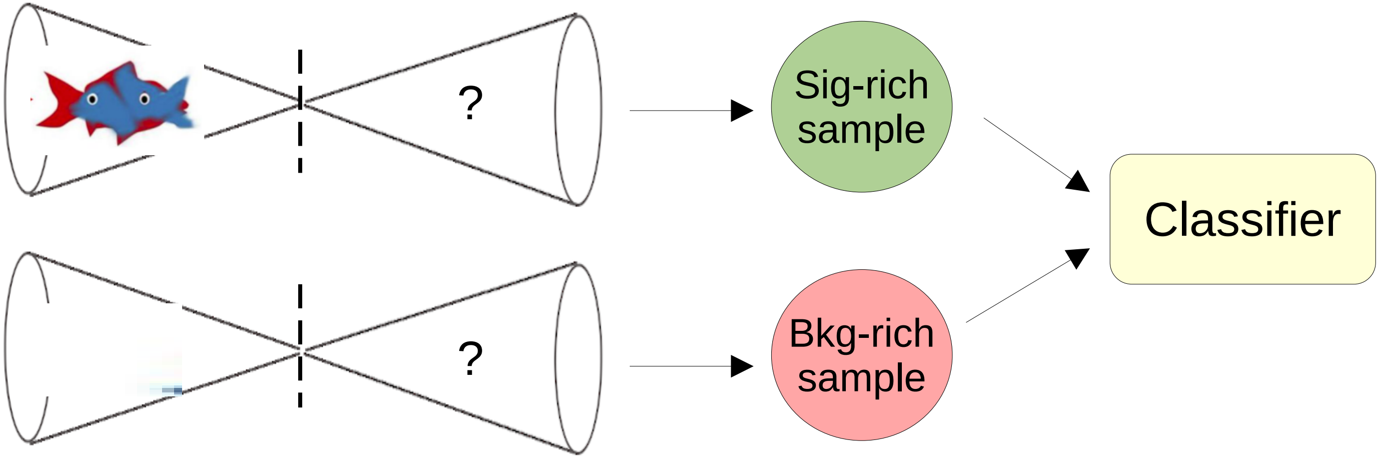
**CATHODE**
**Gen AI** to **interpolate** bkg
into SR to construct sample




[Hallin et al 2109.00546]

---

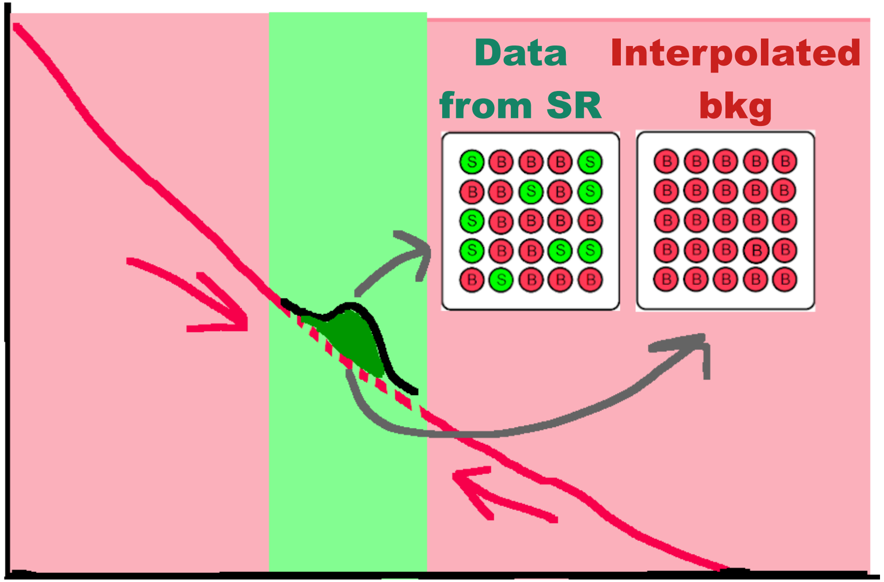
**Tag N' Train**
Looks for pairs of
anomalies,
purifies samples




[**OA** & Suarez 2002.12376]

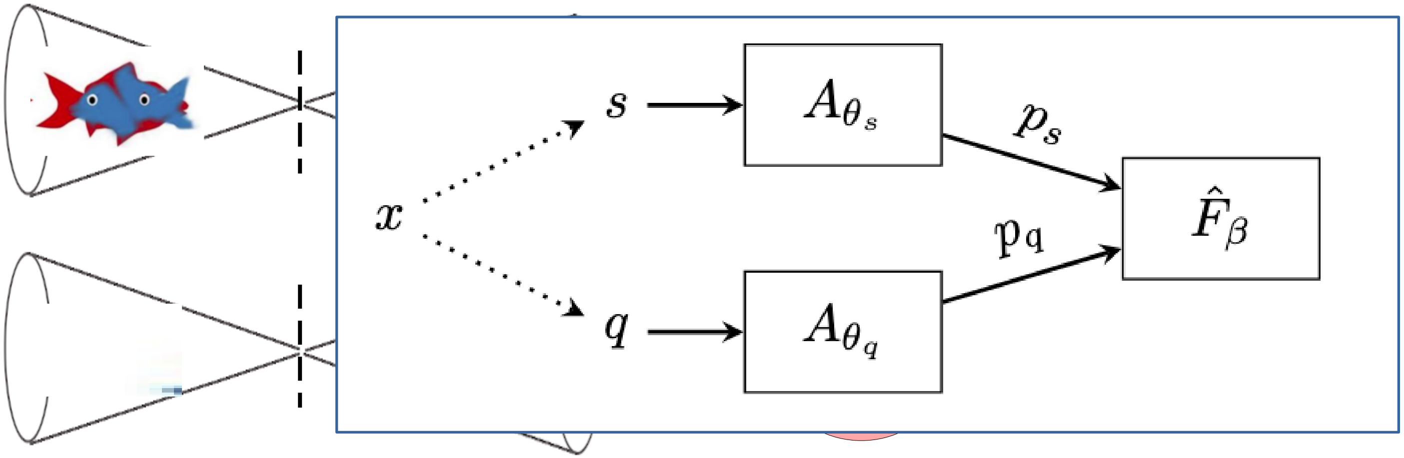
## CATHODE

**Gen AI** to **interpolate** bkg into SR to construct sample

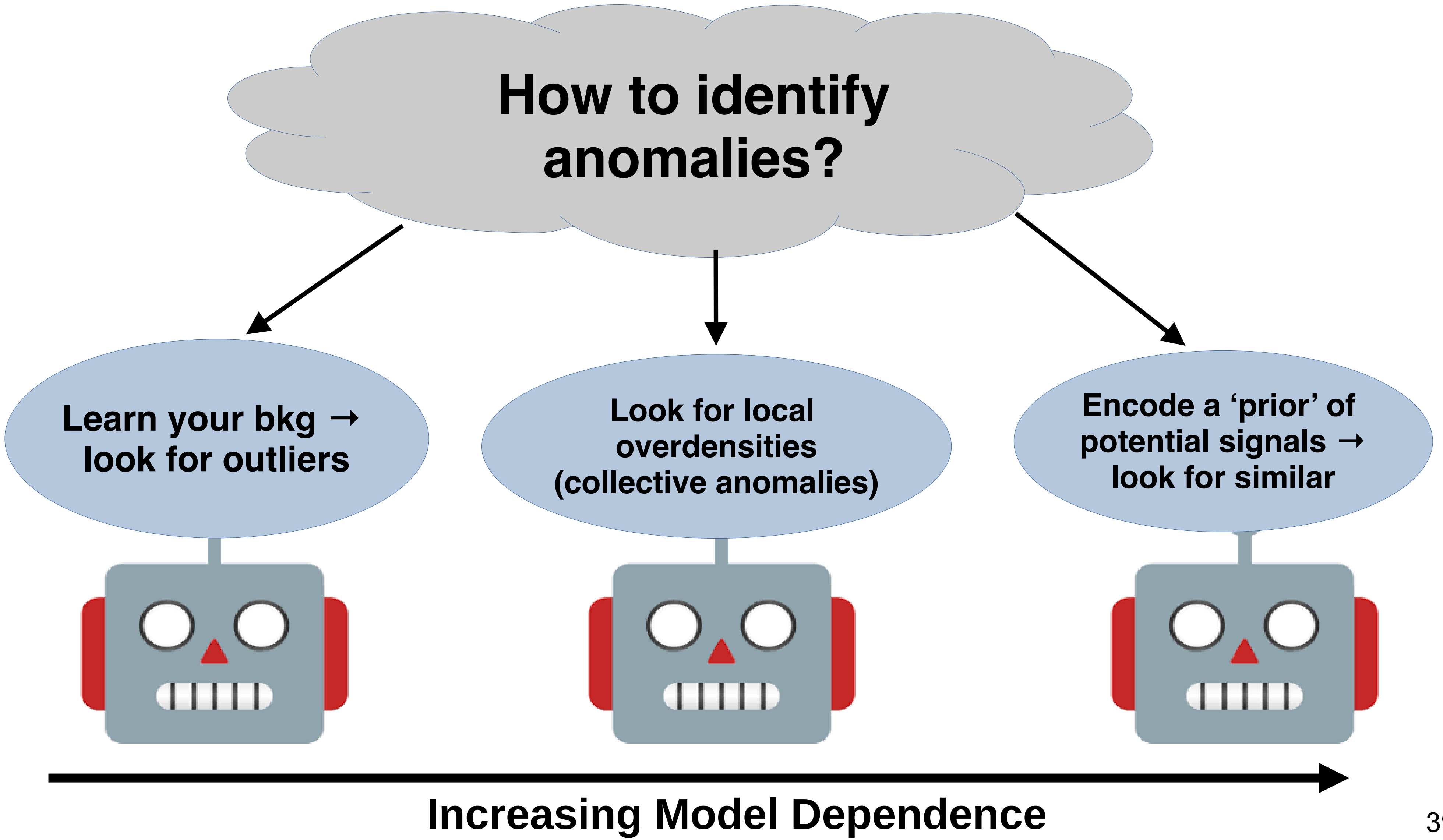



[Hallin et al 2109.00546]

Very similar to coincident AD approach developed at SLAC!



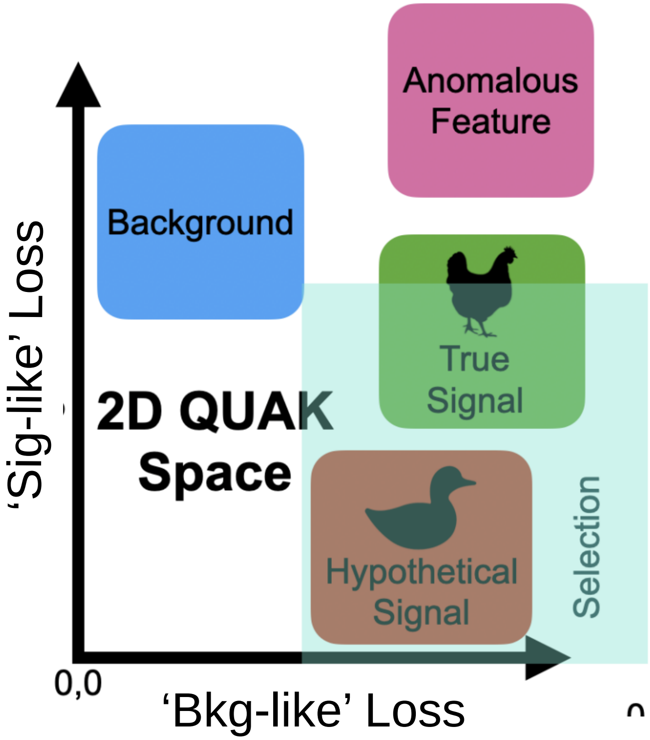

2301.11368


How to identify anomalies?
Learn your bkg → look for outliers
Look for local overdensities (collective anomalies)
Encode a ‘prior’ of potential signals → look for similar
Increasing Model Dependence

# Quasi Anomalous Knowledge (QUAK)

- **Hybrid approach** between fully model-indep. and standard search
- **Encode a prior** on what a potential signal may look like
  - Use an AE trained on a variety of different signal MC's
- Construct 'QUAK space':
  - Loss of signal AE vs bkg AE
- Select events with low sig loss and high bkg loss

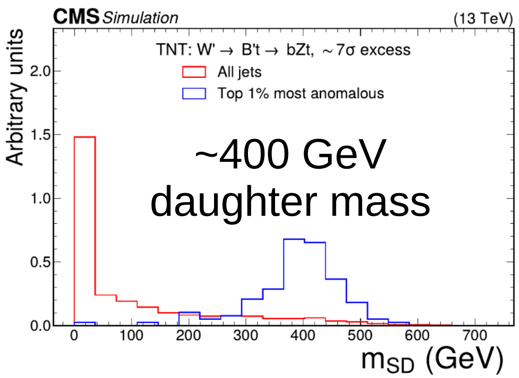
Hypothetical QUAK Space

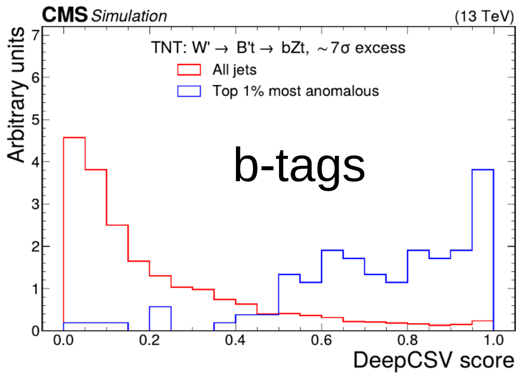


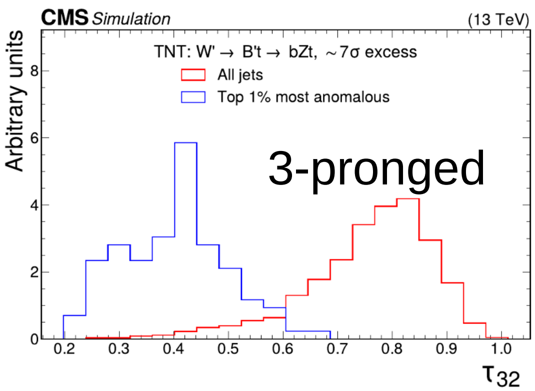

[Park et al 2011.03550]

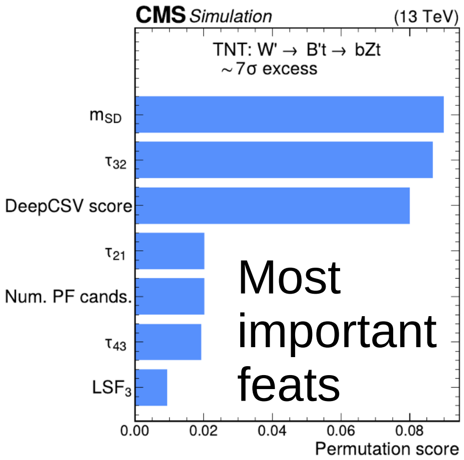
# Understanding Anomalies

Investigate features of most **anomalous events**!
Compare against **regular events**



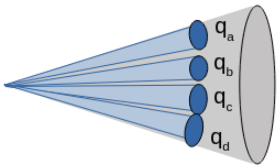



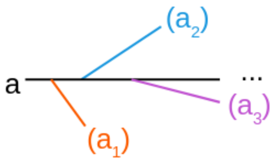



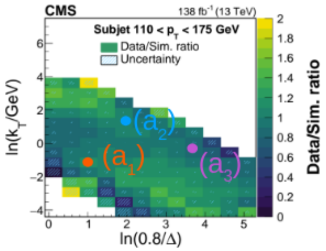

✔ Matches characteristics of injected signal

W' → B't, B' → bZ
$M_{B'}$ = 400 GeV

# Uncertainties

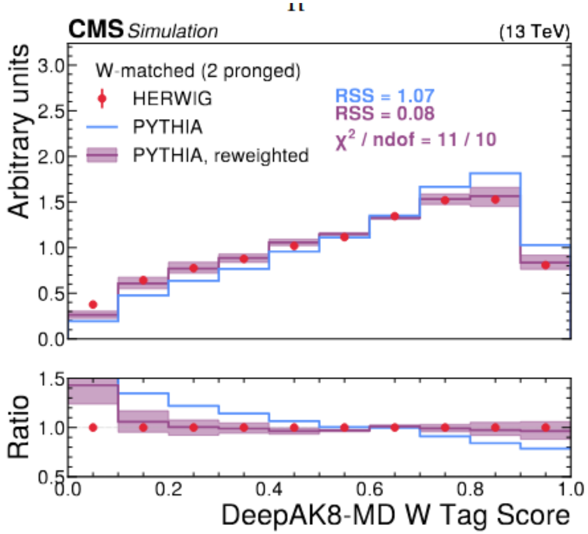



$$W_{\text{jet}} = \prod_{\text{subjets}} W_{\text{subjet}} = \prod_{\text{splittings}} LPR(\text{splitting})$$

Use physics domain knowledge
to factorize problem




Account for data/sim domain shift

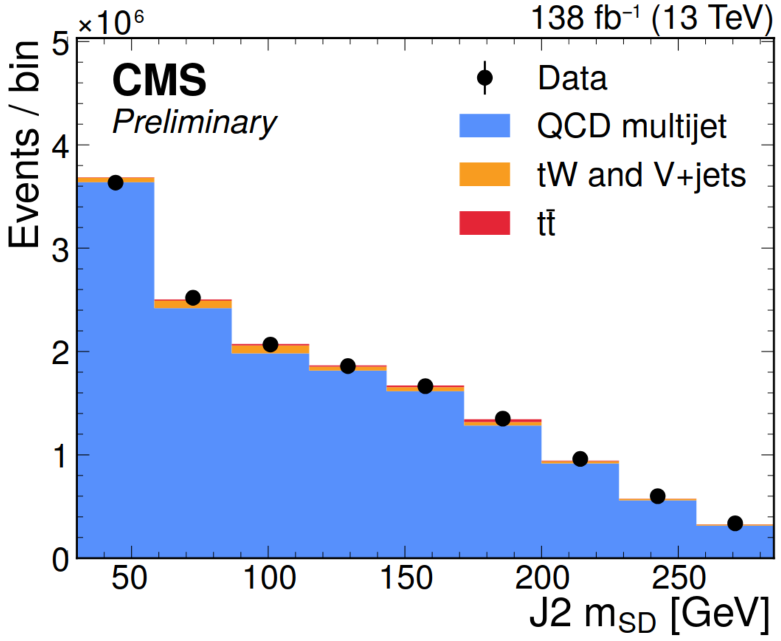
Developed novel procedure
to calibrate AI-classification
of anomalous jets!

Separate paper
arXiv:2507.07775

Method now standard within CMS, employed by multiple (5+) analyses!

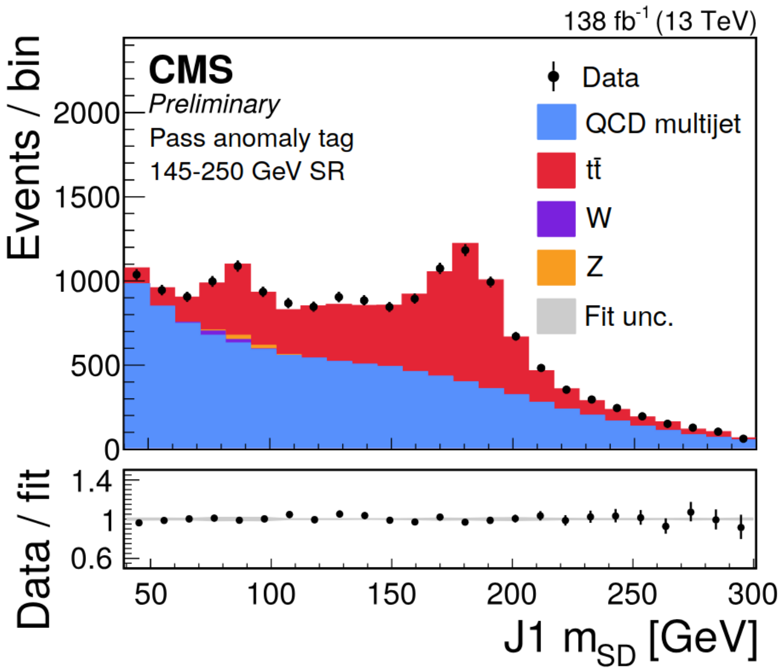
# Validation on a real 'anomaly'

'Rediscovering' boosted **top quarks** in data with anomaly detection!




**Anomaly Detection!**

Modified version* of **TNT** method

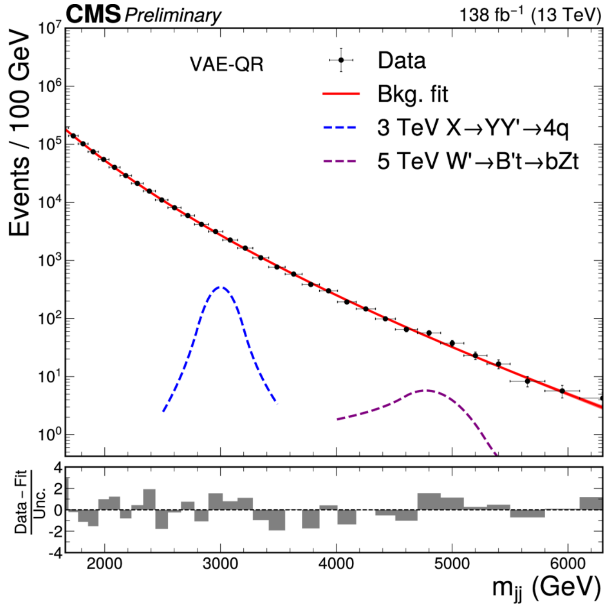


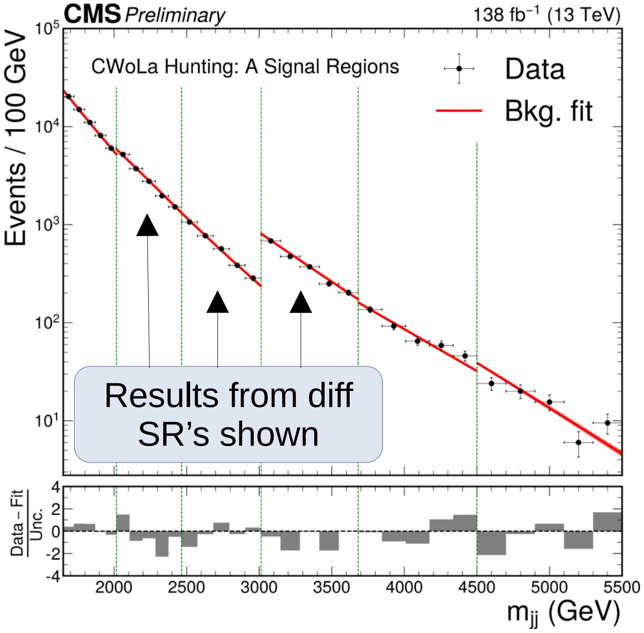


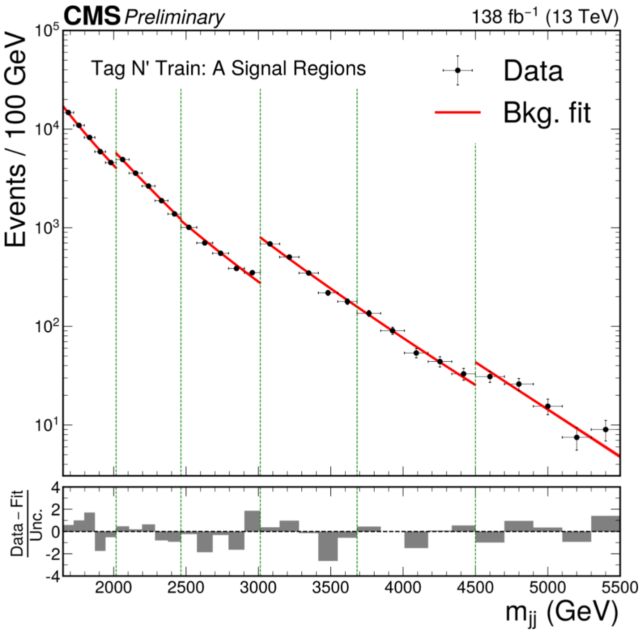

*Training setup modified to target pair production rather than a heavy resonance, everything else unchanged

# Search Results

## No significant anomalies from any of the five methods



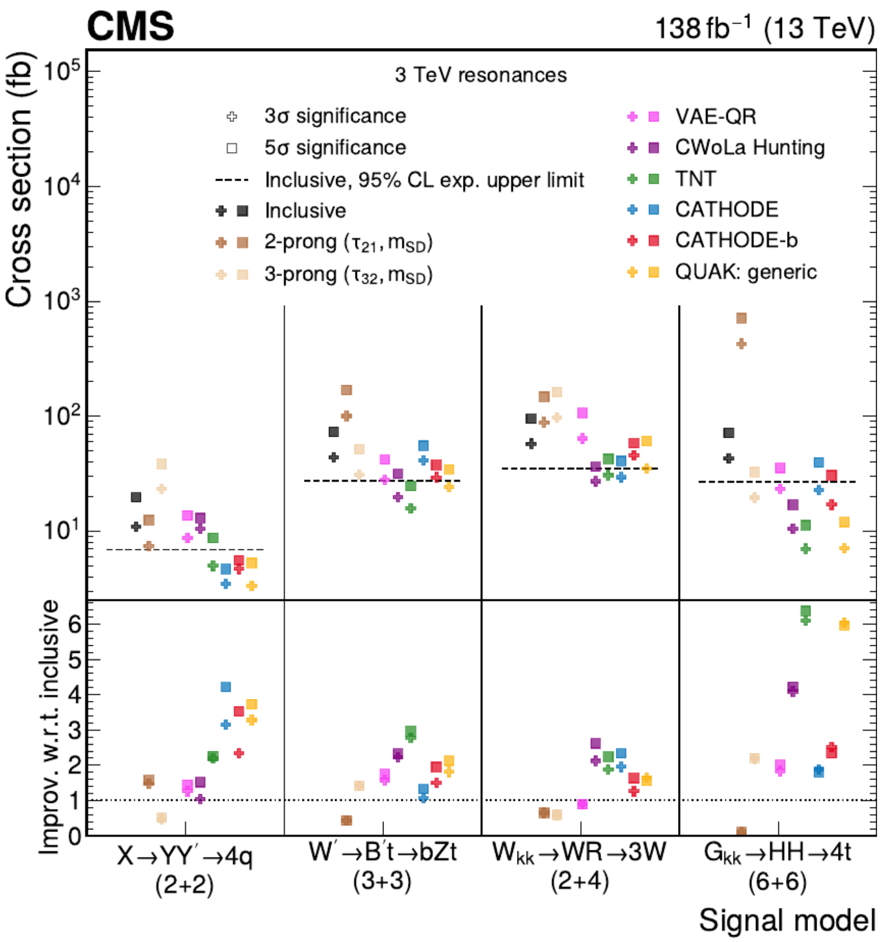

QUAK & CATHODE
results similar

# Improved Sensitivity

- "How strong of a signal do I need to get an expected 3σ/5σ excess?"

- **Anomaly detection** improves sensitivity by **~3-7x**!

  Compared to standard bump hunt



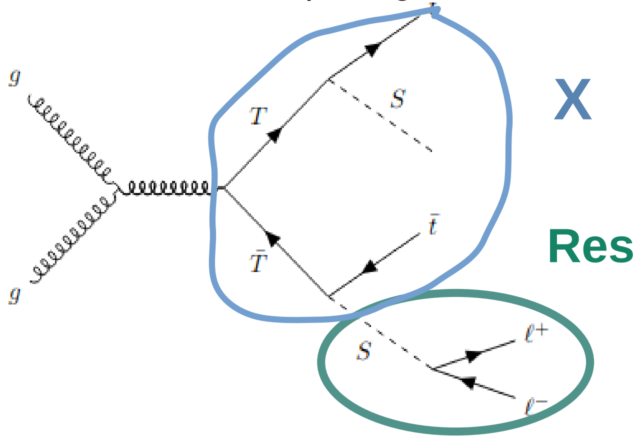
# What's next?

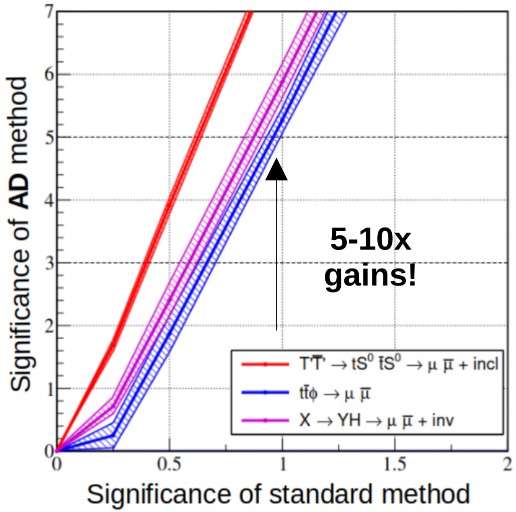
# New Topologies

arXiv:2504.13249
Brennan, Vami, **OA** et al

- Expand extend methods to new signatures **beyond jets**
- One program : **Resonance + X**
  - Motivated by 'non-minimal' dark/Higgs sector models
- Look for a resonance produced in association with other 'anomalous' stuff (X)
  - Eg di-tau+X or di-muon + X

Example signal

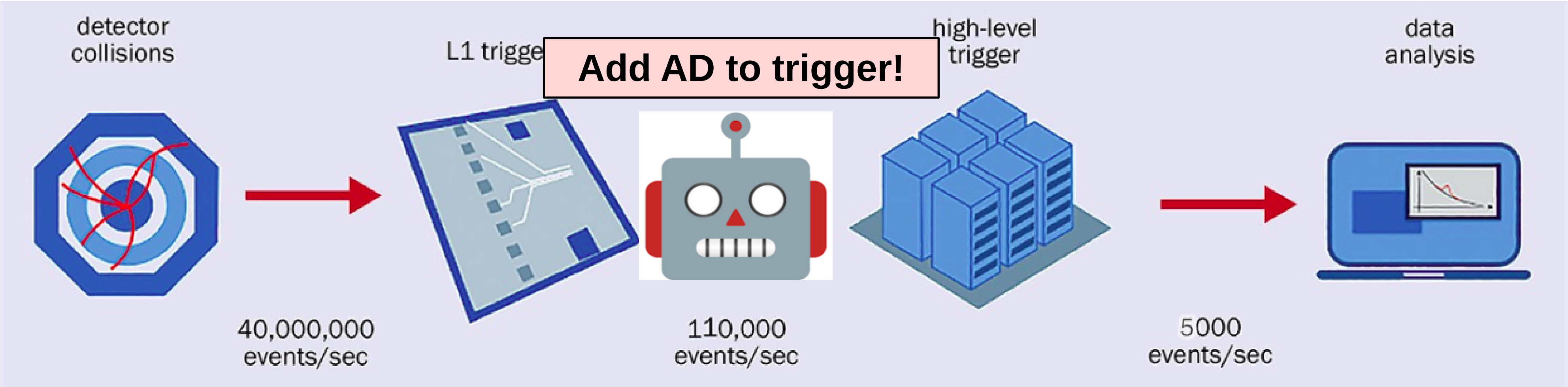



See also search for X → Higgs+X in backup

# Trigger

- Trigger rejects >99% of events
- What if we aren't saving the new particles ?



# Anomaly Detection in Trigger

- Recently CMS & ATLAS have deployed anomaly detection triggers for the first time!
- Significant resource constraints!
  - FPGA, operate at 40 MHz

AXOL1TL

Deployed 2024!

CICADA

GELATO

hls 4 ml

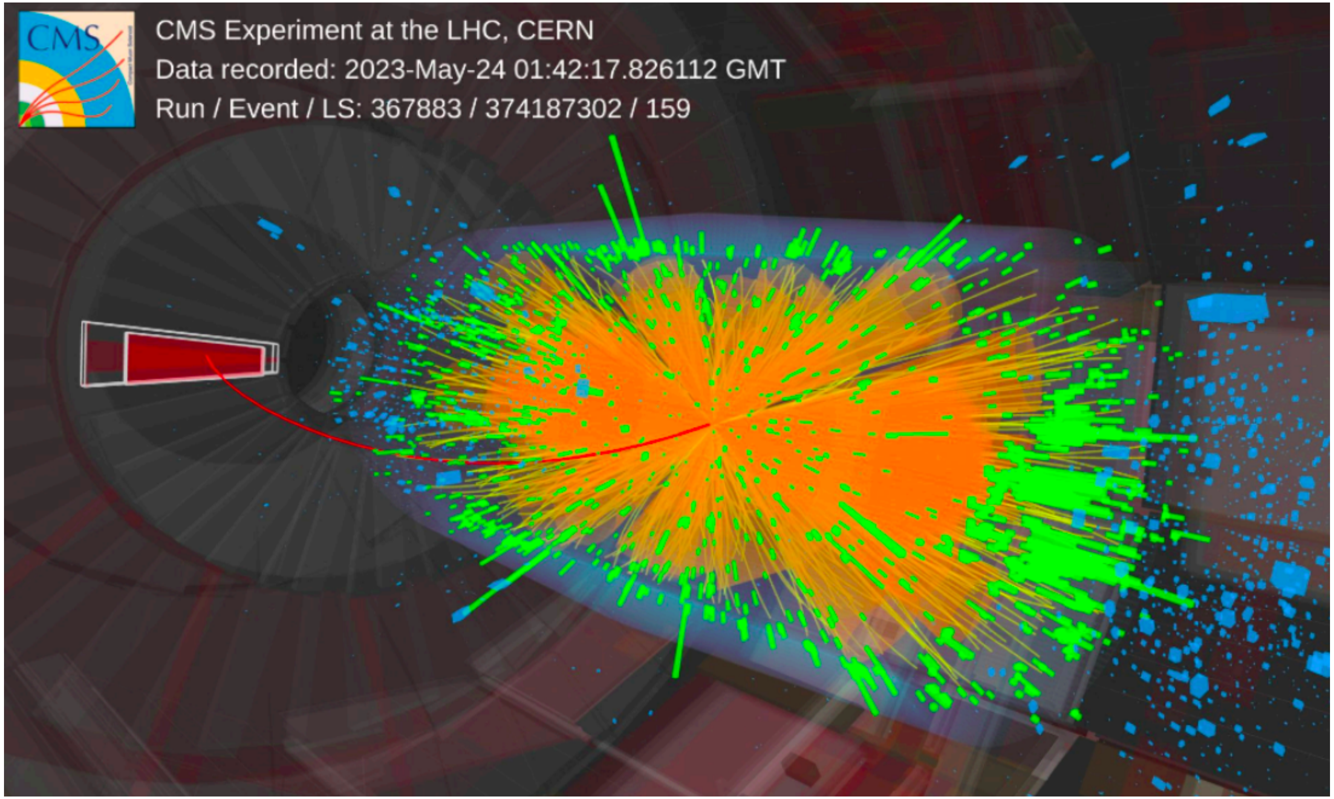
QKeras

AXOL1TL CMS-DP-2023-079
CICADA CMS-DP-2023-086
GELATO ATL-DAQ-SLIDE-2025-362

# An Anomalous Event

A unique **AXOL1TL** event!

Very busy, 11 jets + 1 muon



# Future AD Research Questions

**How can we broaden the scope of these AD searches in HEP?**

- beyond resonances, using AD triggered data, ...

**How can we maximize the physics reach of realtime AD systems?**

- AD robust to changing conditions, build invariances, …

**Where else can these AD methods be applied?**

- Other experiments, fault detection, data quality monitoring, ...

## Ingredient

- High quality scientific data
- Clear indication of outlier
- Understanding of backgrounds
- Short analysis timescale
- A curious young scientist inspecting the data

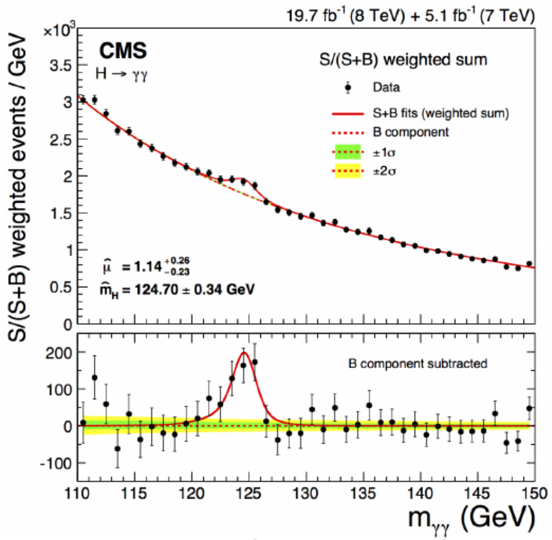
## AI Technology

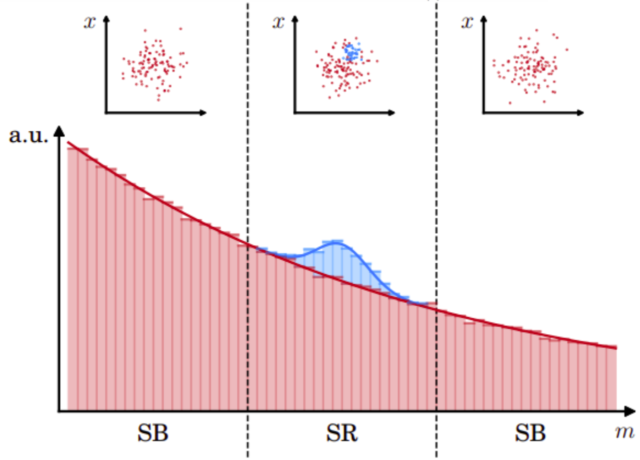
Anomaly Detection

Generative Models

# Extending the Bump Hunt

- Classic bump hunt
  - Smoothly falling background, localized resonance
  - 1D fit
- Generative models:
  - Extend this idea to higher dimensions

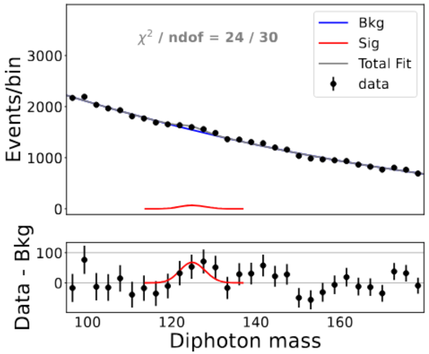


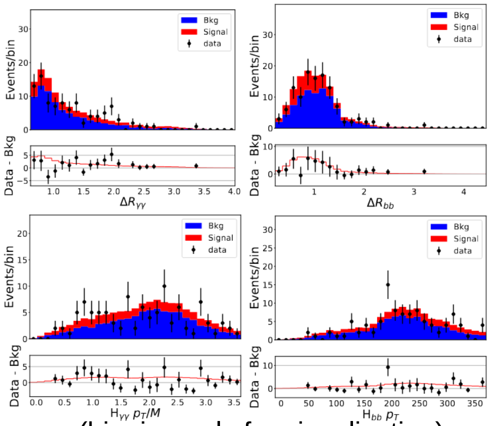



# Gen-AI Measurements

2506.06438
**OA** & Szewc

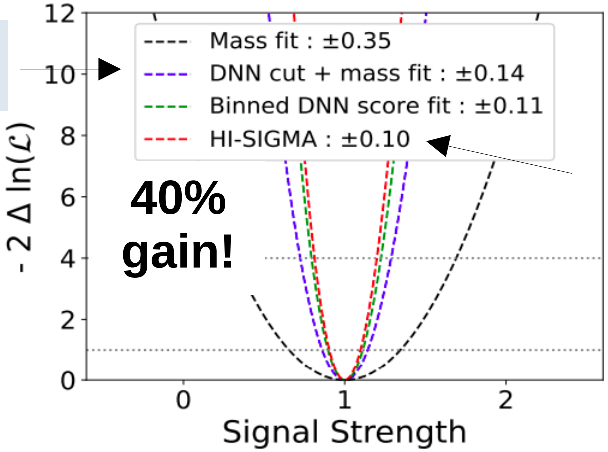
Example 5D (!) fit of **HH→ bbγγ**




(binning only for visualization)

Currently used




**Optimal and includes data-driven bkg!**

- ‘**HI-SIGMA**’
  - Bump hunt in high dimensions
- Extends ‘simulation based inference’ ideas to data-driven bkgs!
- More work on uncertainty quantification & robustness needed!

## Ingredient

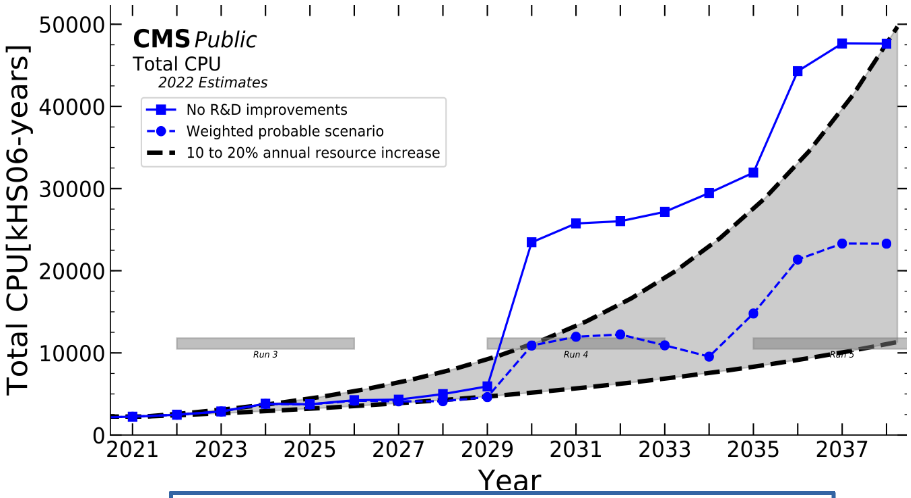
- High quality scientific data
- Clear indication of outlier
- Understanding of backgrounds
- Short analysis timescale
- A curious young scientist inspecting the data

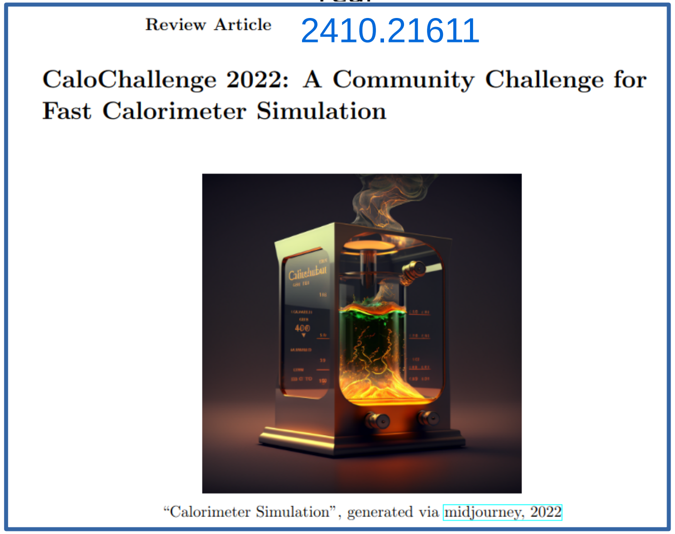
## AI Technology

Anomaly Detection

Generative Models

# The Need for Fast Simulation



Review Article 2410.21611

CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation

"Calorimeter Simulation", generated via midjourney, 2022

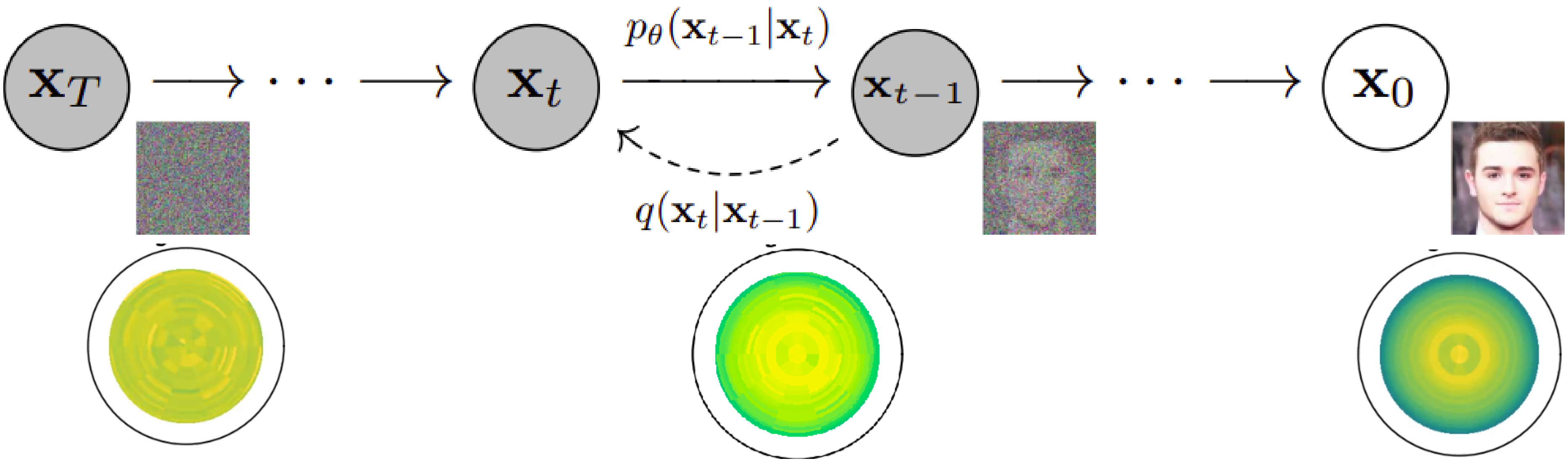
- Simulation crucial in all HEP analyses
- Computing budgets can't keep up with data rate
- ML-based fast sim. **needed** for HL-LHC!

How would our science change if simulations took a week/day/hour not ~6 months?

# ‘CaloDiffusion’

**Goal** : Train a generative ML model to mimic physics based simulation (Geant) with high accuracy & significant speedup

**Pure Noise** **Noise + Image** **Real Image**

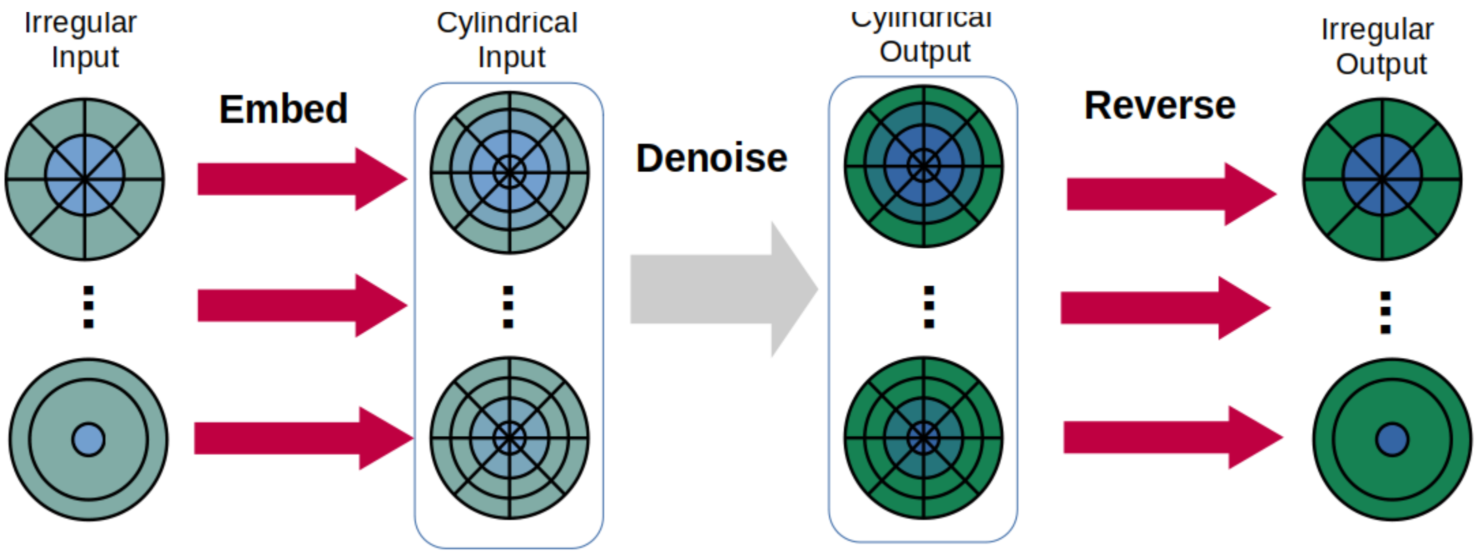



**OA** & Pedro
arXiv:2308.03876

# One Innovation : Irregular Geometries

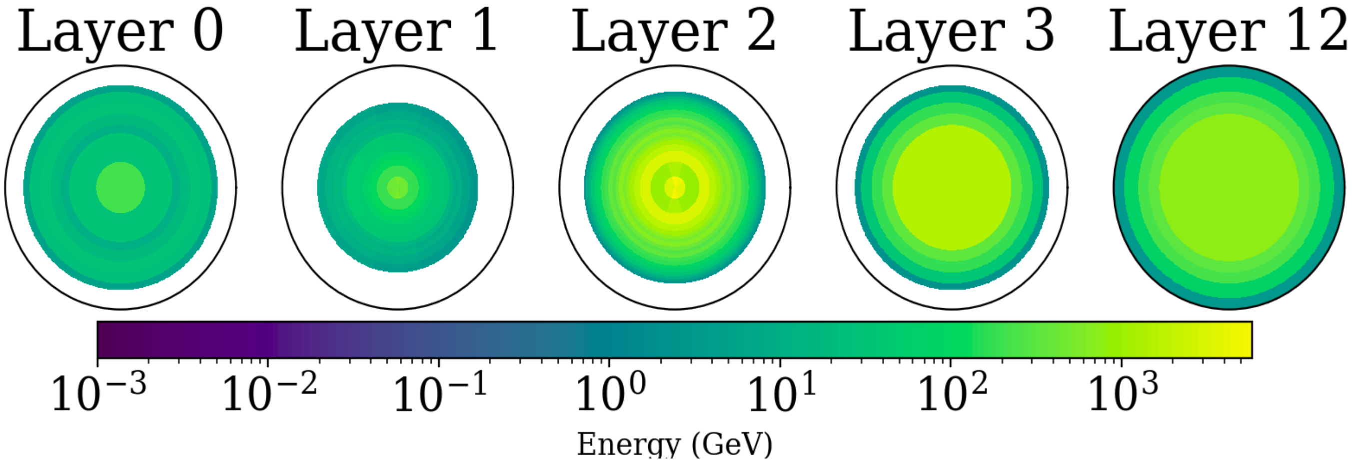
- Real scientific instruments unlike perfect images, have **irregular structure**
  - Can't natively apply your favorite ML tools (like convolutions)
- Learn an **GLaM-orous embedding** that maps input into **regular cylindrical structure**

GLaM : Geometry Latent Mapping

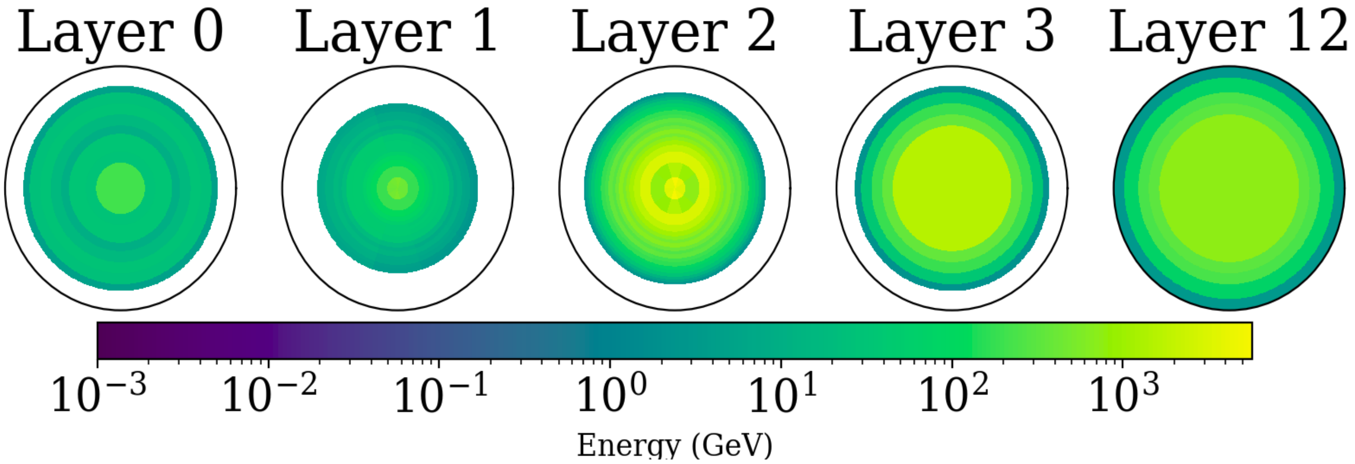


# Average Showers

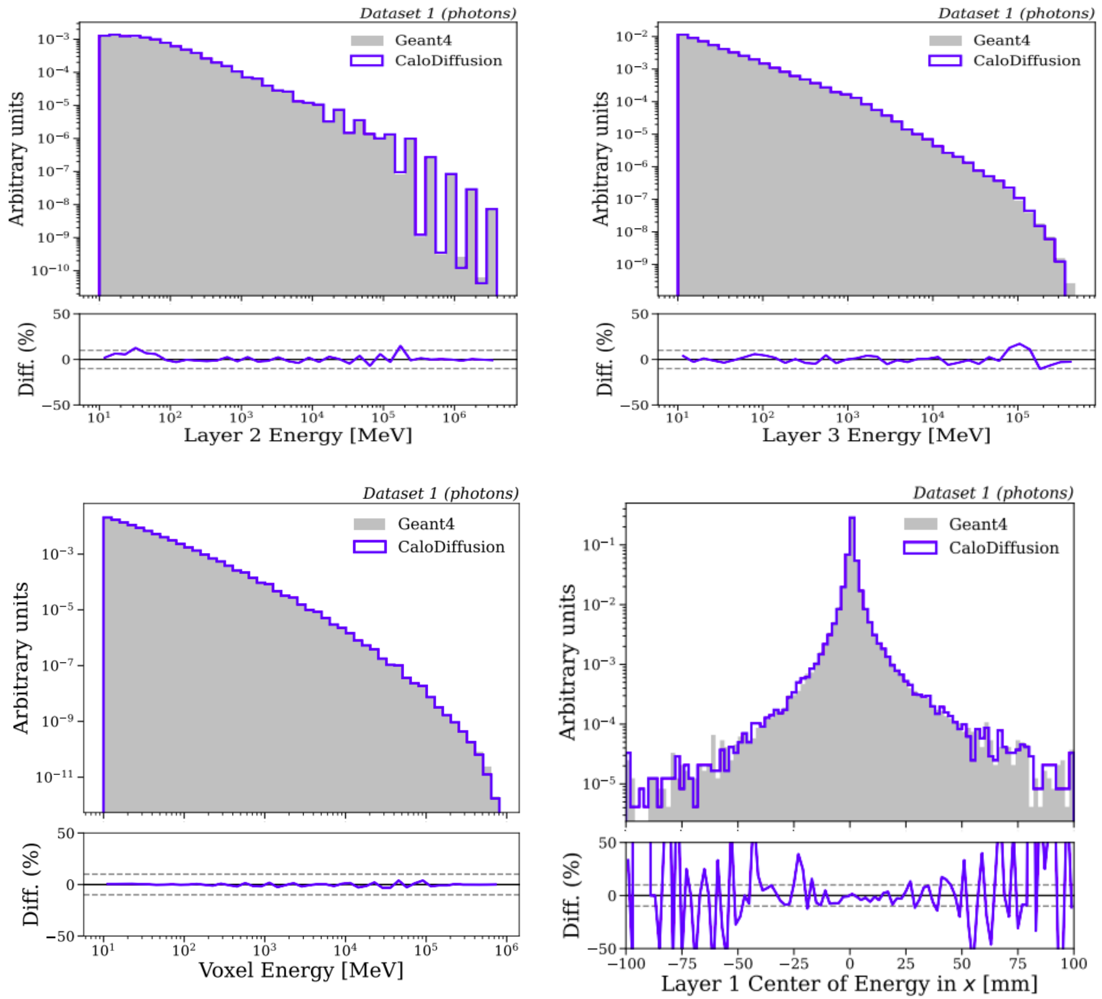
**Geant**




**Calo Diffusion**

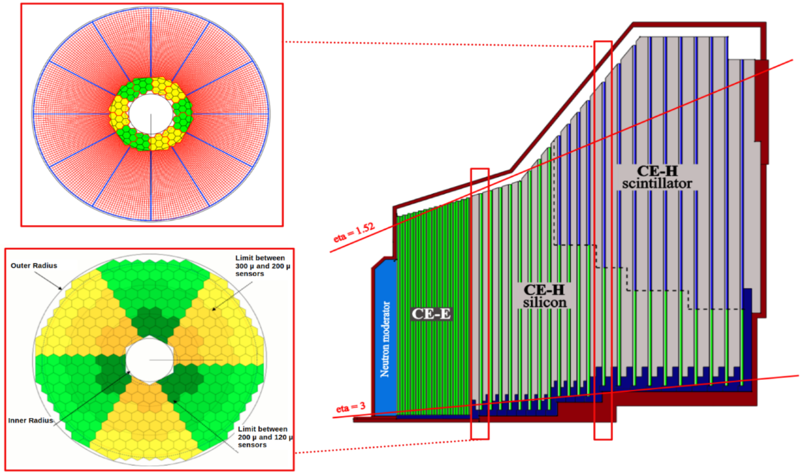


# CaloChallenge Results




- **Calodiffusion** matches **Geant** to very high fidelity
  - Up to 1000x speed up
- Top-2 in quality across all 4 CaloChallenge datasets (50 total submissions)

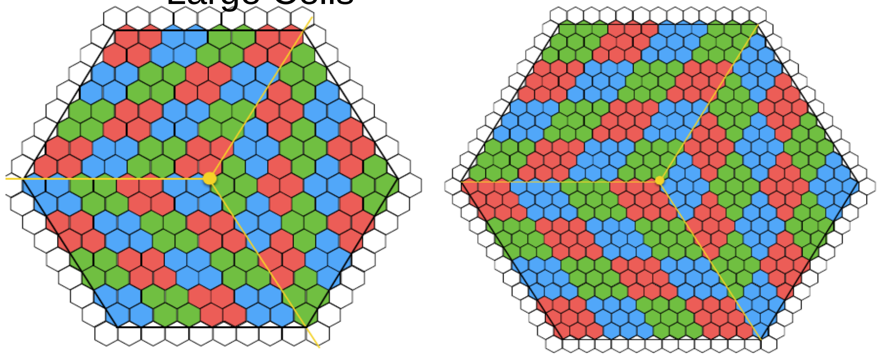
# CMS High Granularity Calorimeter




Individual Hexagon Wafers

Large Cells

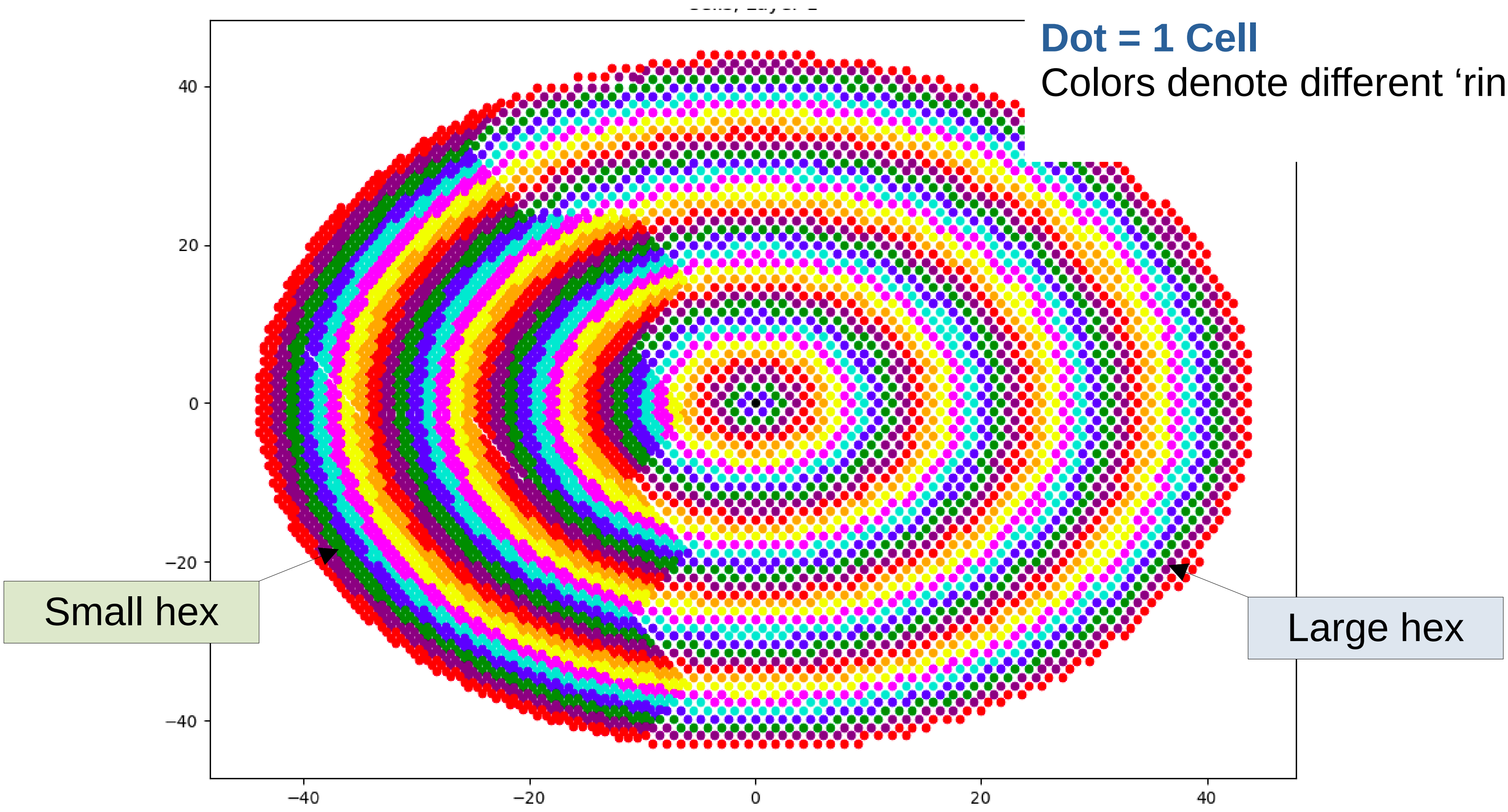
Small Cells



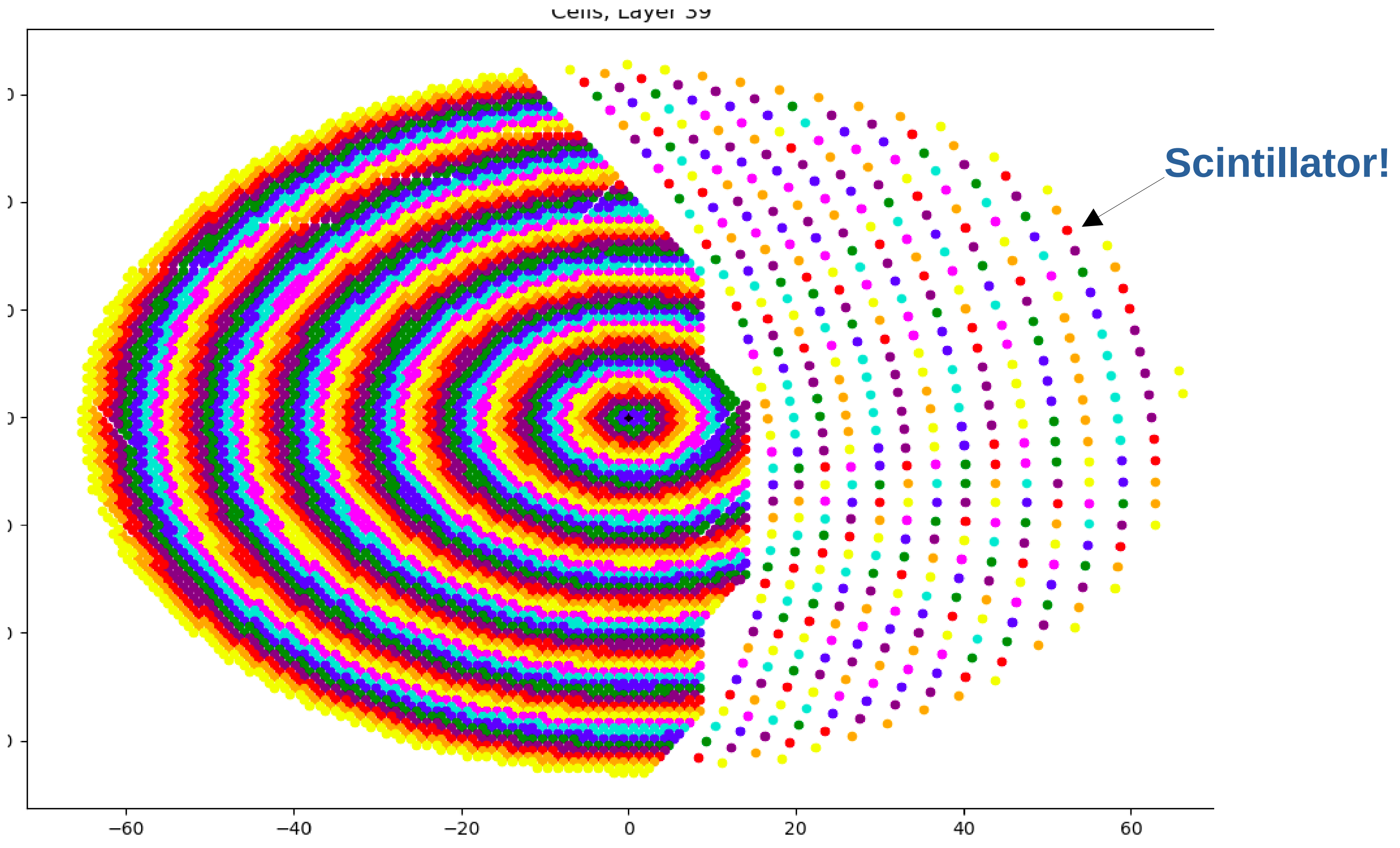
- Complicated hexagonal geometry!
- Very high granularity
  - Up to **1 million** cells in region of shower
- Extremely sparse & irregular
  - Novel methods to do sparsity-based sampling!

# Example HGCal Region : Layer 1

**Dot = 1 Cell**
Colors denote different ‘rings’



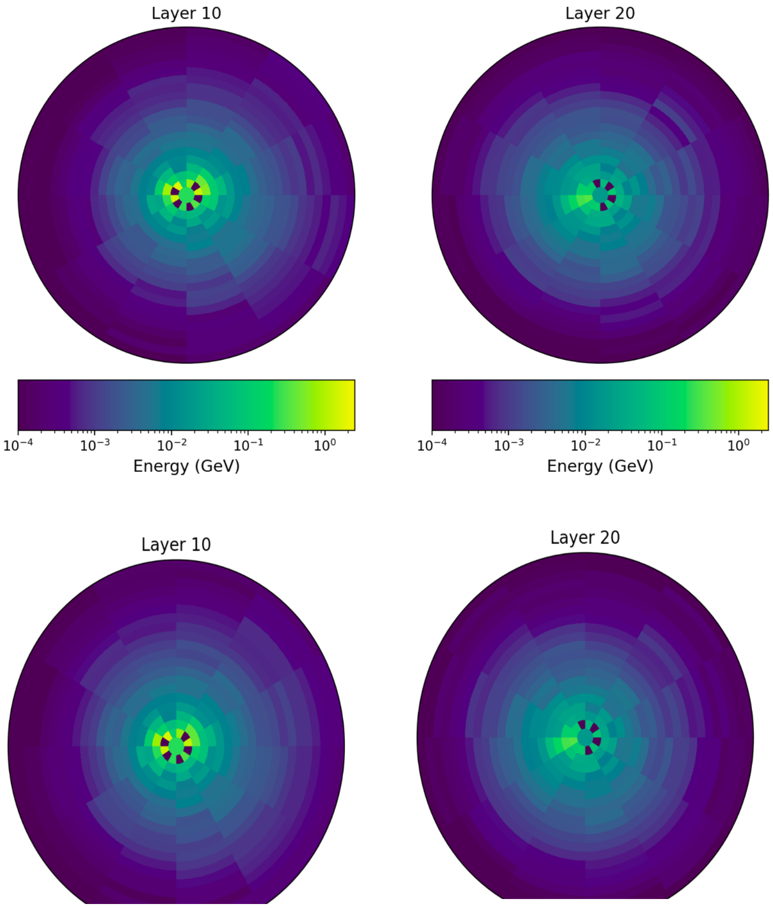
# Example HGCal Region : Layer 39

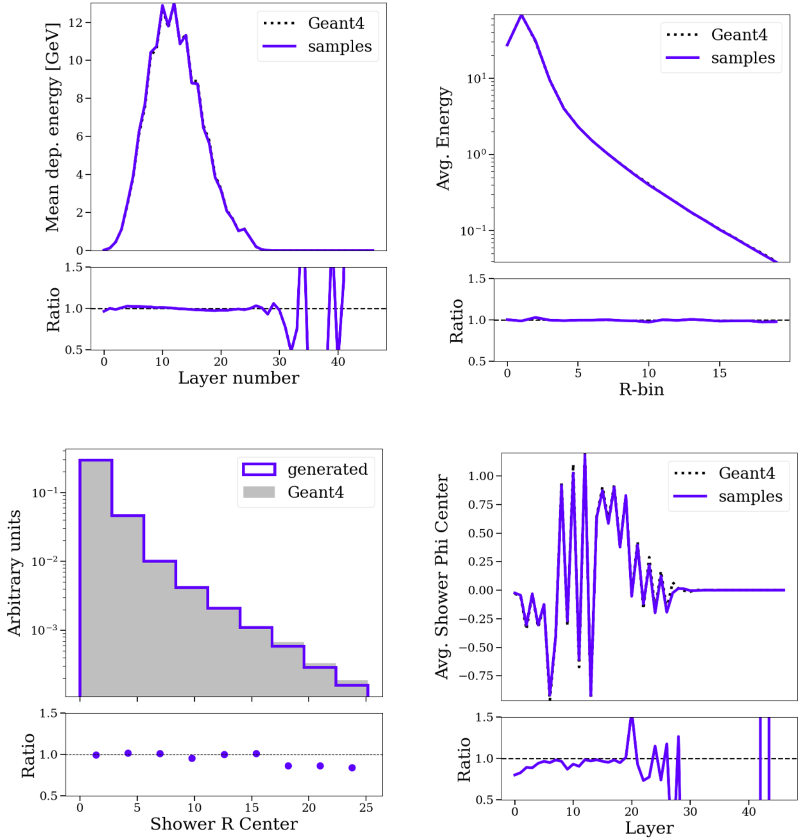


# Preliminary Results

**Calo Diffusion**

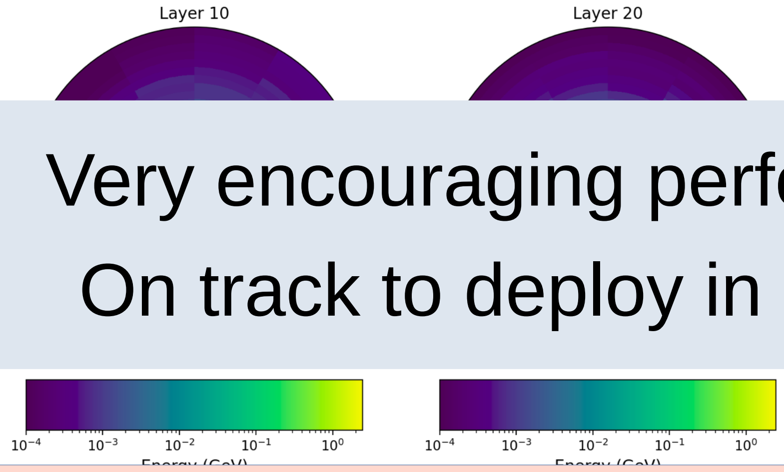
**Geant (CMS Simulation)**

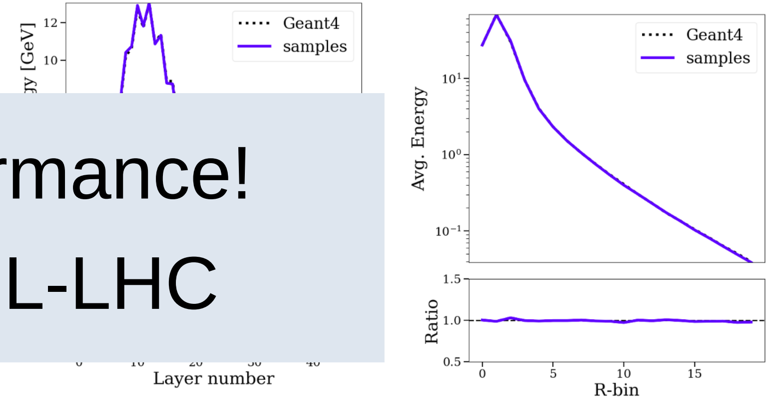


# Preliminary Results

## Calo Diffusion



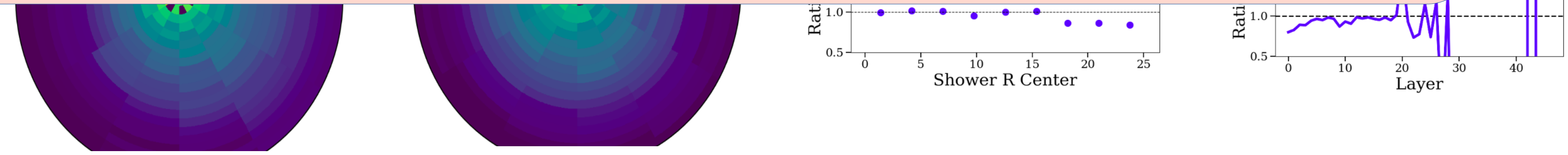




**Very encouraging performance!**

**On track to deploy in HL-LHC**

**Similar methods applicable to other experiments!**

*Fixed target experiments, future colliders, surrogate models for optimization, ...*




**(CMS Simulation)**

## Ingredient

- High quality scientific data
- Clear indication of outlier
- Understanding of backgrounds
- Short analysis timescale
- A curious young scientist inspecting the data

## AI Technology

Anomaly Detection

Generative Models

AI Design Optimization

# We have excellent experiments now / near future!

HiLumi HL-LHC PROJECT

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

...

# But what if the particles are just at higher energy?

# Beyond the TeV Scale

2023 P5 Report

## 2.3 The Path to a 10 TeV pCM

Exploring the Quantum Universe

Realization of a future collider will require resources at a global scale and will be built through a world-wide collaborative effort where decisions will be taken collectively from the outset by the partners. This differs from current and past international projects in particle physics, where individual laboratories started projects that were later joined by other laboratories. The proposed program aligns with **the long-term ambition of hosting a major international collider facility in the US, leading the global effort** to understand the fundamental nature of the universe.

…

In particular, a muon collider presents an attractive option both for technological innovation and for bringing energy frontier colliders back to the US. The footprint of **a 10 TeV pCM muon collider is almost exactly the size of the Fermilab campus**. A muon collider would rely on a powerful multi-megawatt proton driver delivering very intense and short beam pulses to a target, resulting in the production of pions, which in turn decay into muons. This cloud of muons needs to be captured and cooled before the bulk of the muons have decayed. Once cooled into a beam, fast acceleration is required to further suppress decay losses.

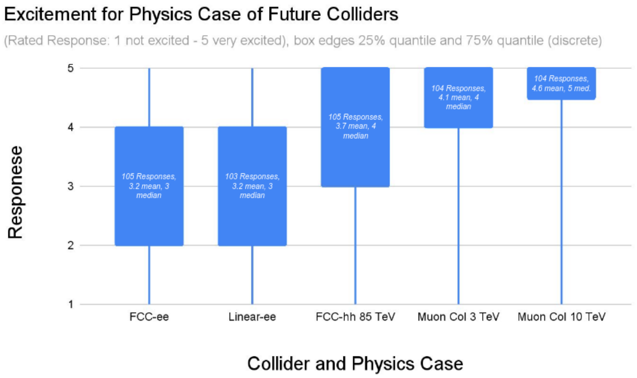
…

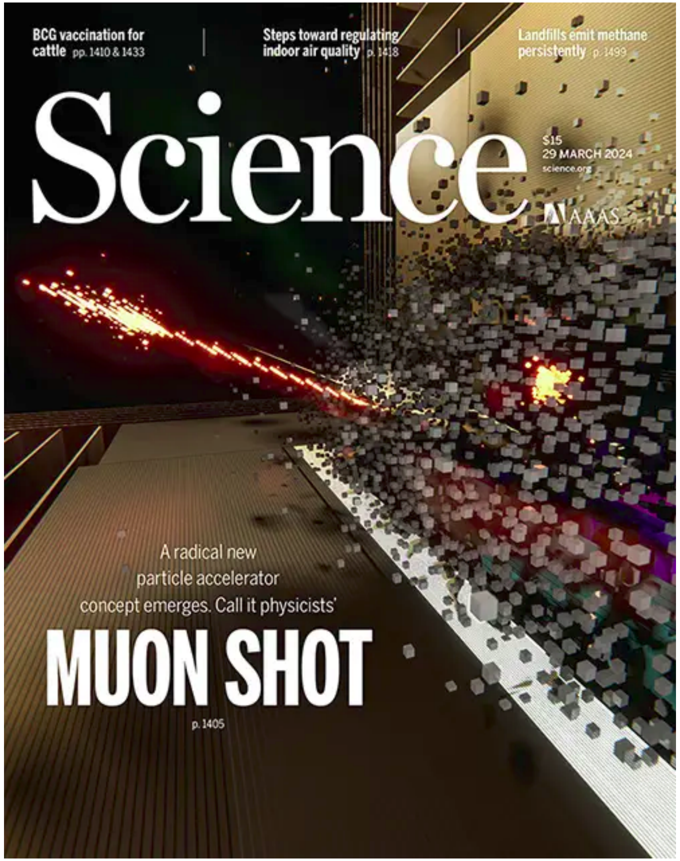
Although **we do not know if a muon collider is ultimately feasible**, the road toward it leads from current Fermilab strengths and capabilities to **a series of proton beam improvements and neutrino beam facilities**, each producing world-class science while performing critical R&D towards a muon collider. At the end of the path is an unparalleled global facility on US soil. **This is our Muon Shot.**

*Slide from Hitoshi Murayama* 39

“Most exciting” future collider option in survey of US Early Career HEP researchers

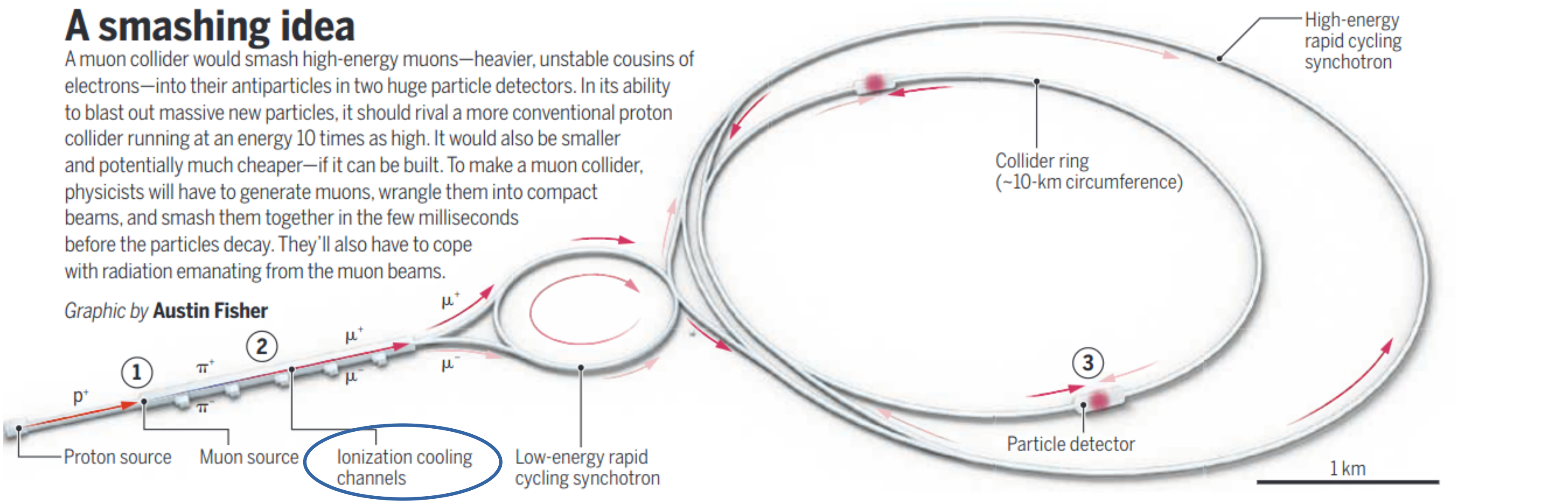
**OA** & Cummings
2503.22834

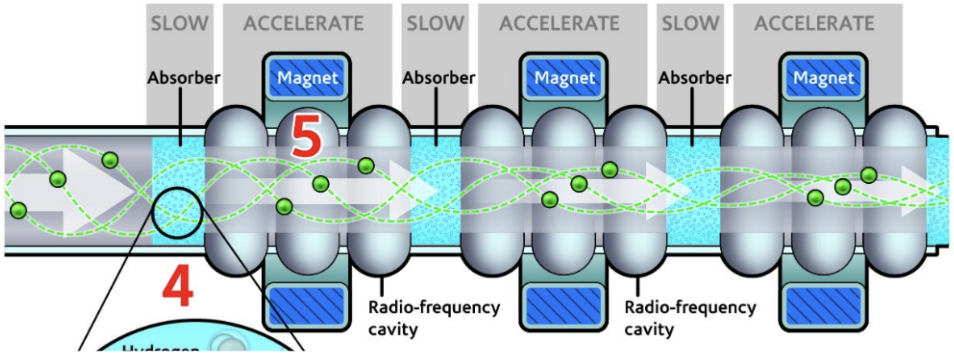


# Technical Challenges

Science Magazine



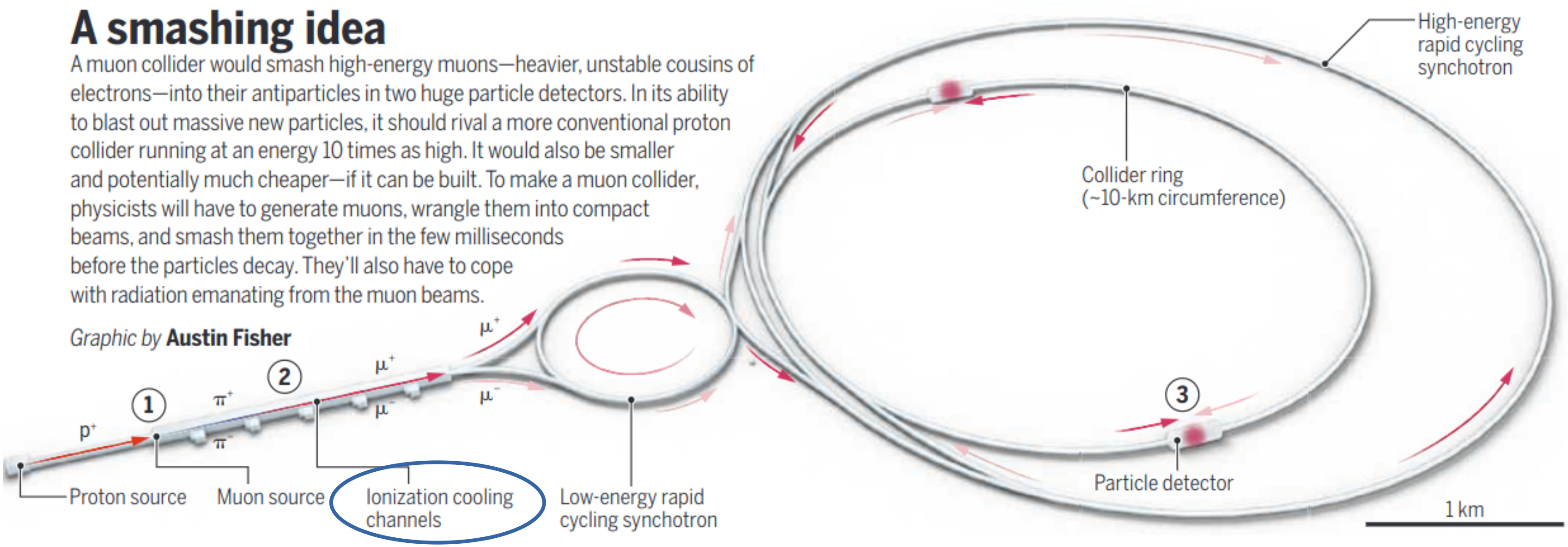



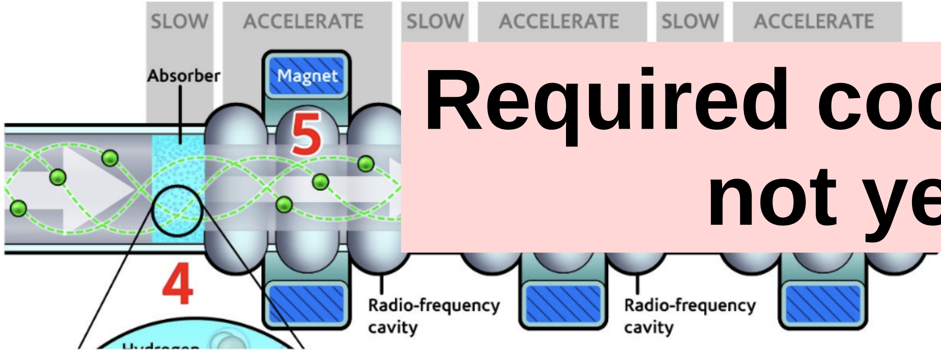

- Cooling step : absorb + accelerate
  - Each achieves few % cooling
  
  → Need O(100) steps

**Complex designs with many parameters**

# Technical Challenges

Science Magazine



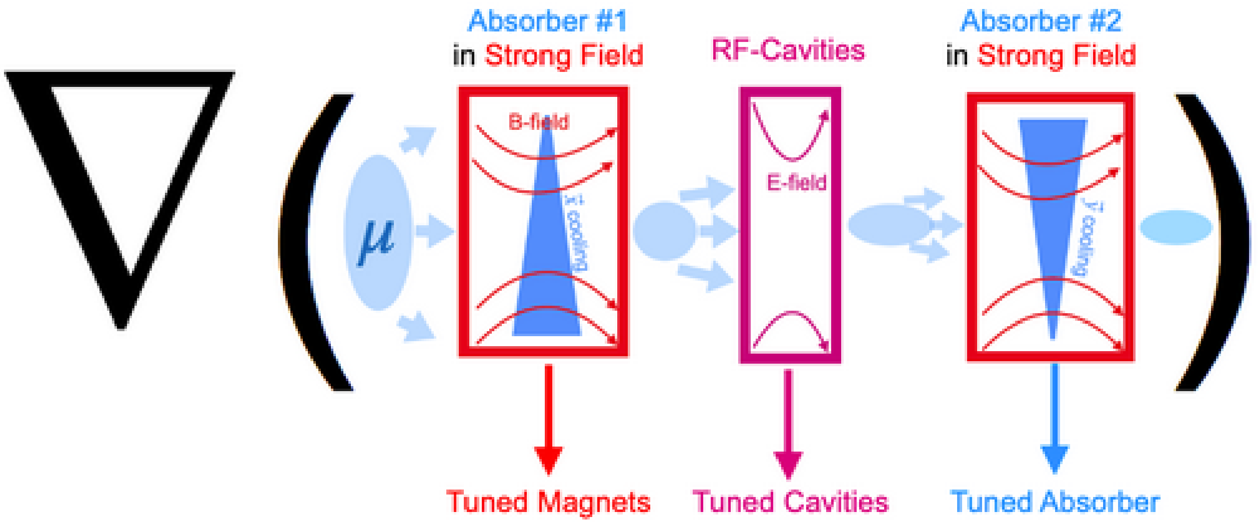

- Cooling step : absorb + accelerate

cooling

**Required cooling performance not yet achieved!**

**Complex designs with many parameters**

# AI-Optimized Cooling Designs




$$\frac{\partial\epsilon}{\partial B} = \frac{\partial\epsilon}{\partial A_2}\frac{\partial A_2}{\partial B} + \frac{\partial\epsilon}{\partial A_2}\frac{\partial A_2}{\partial RF}\frac{\partial RF}{\partial A_1}\frac{\partial A_1}{\partial B} + \dots$$

~100s param cooling designs difficult to hand optimize

→ use Diff. Prog. Methods!

*Will benefit from collaboration with SLAC Accelerator & Diff. Prog. Experts!*

Potential for field-shaping AI contribution!

## Ingredient

- High quality scientific data
- Clear indication of outlier
- Understanding of backgrounds
- Short analysis timescale
- A curious young scientist inspecting the data

## AI Technology

Anomaly Detection

Generative Models

AI Design Optimization

Foundation Models

# Foundation Models

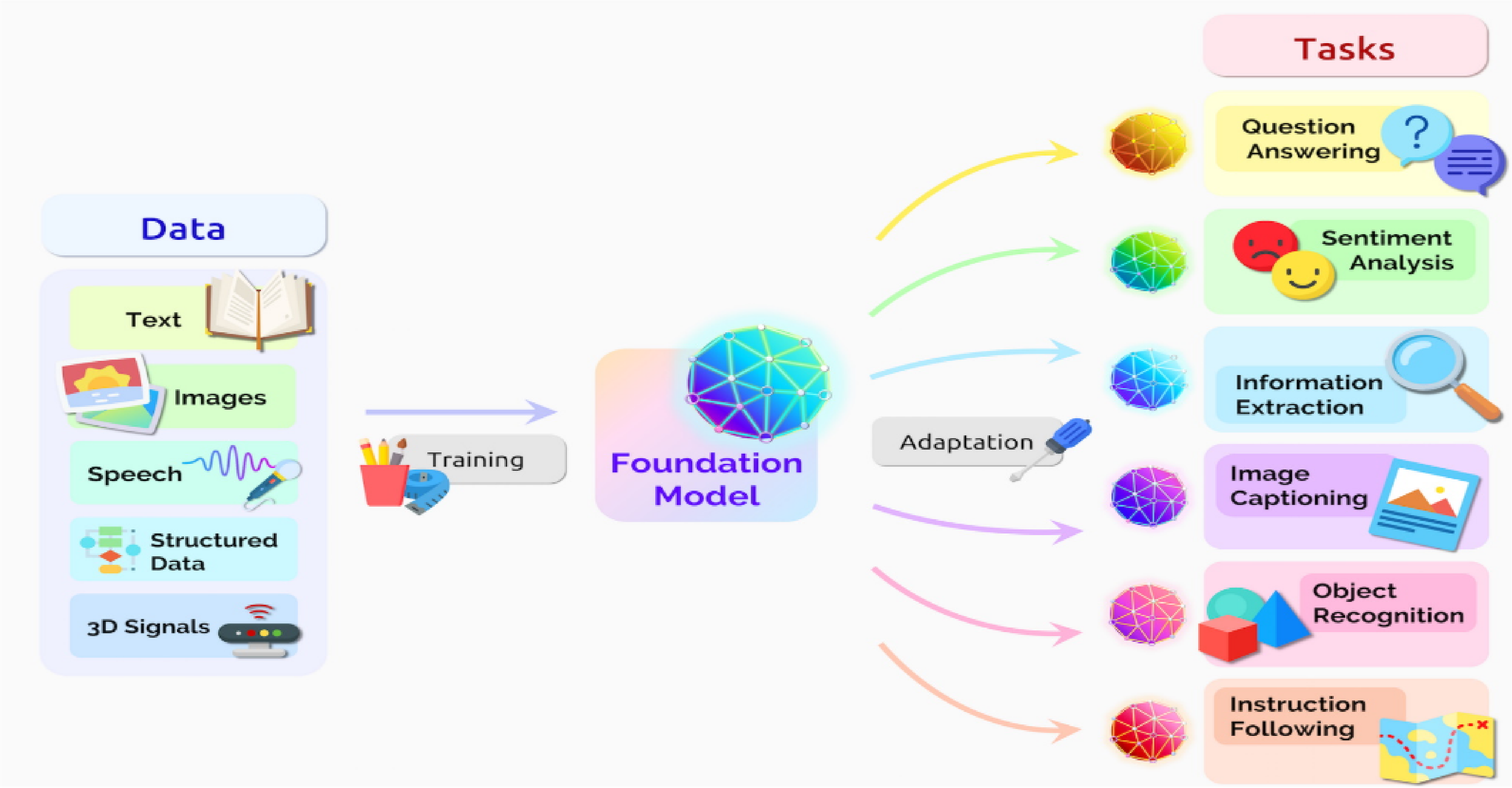
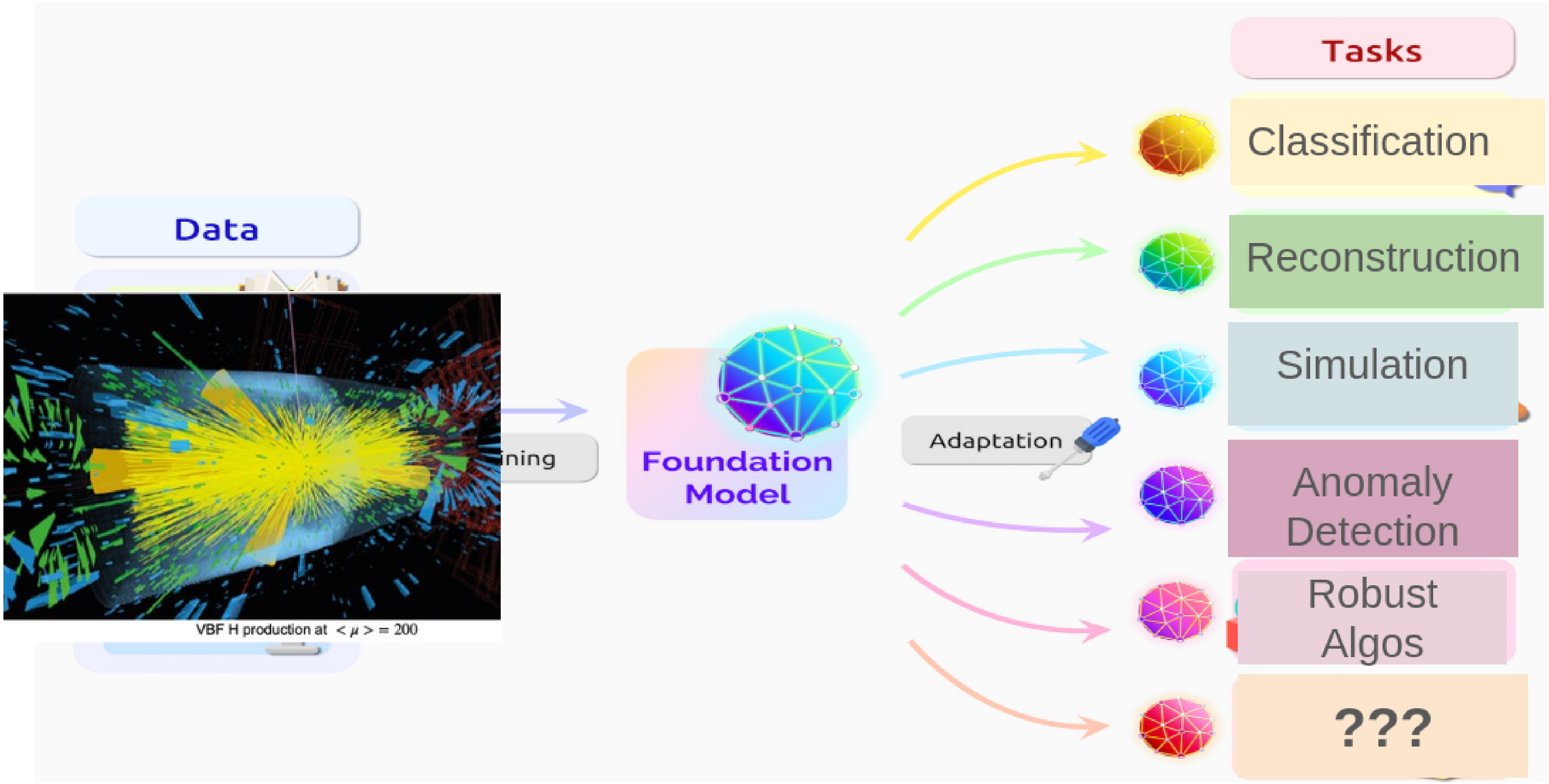
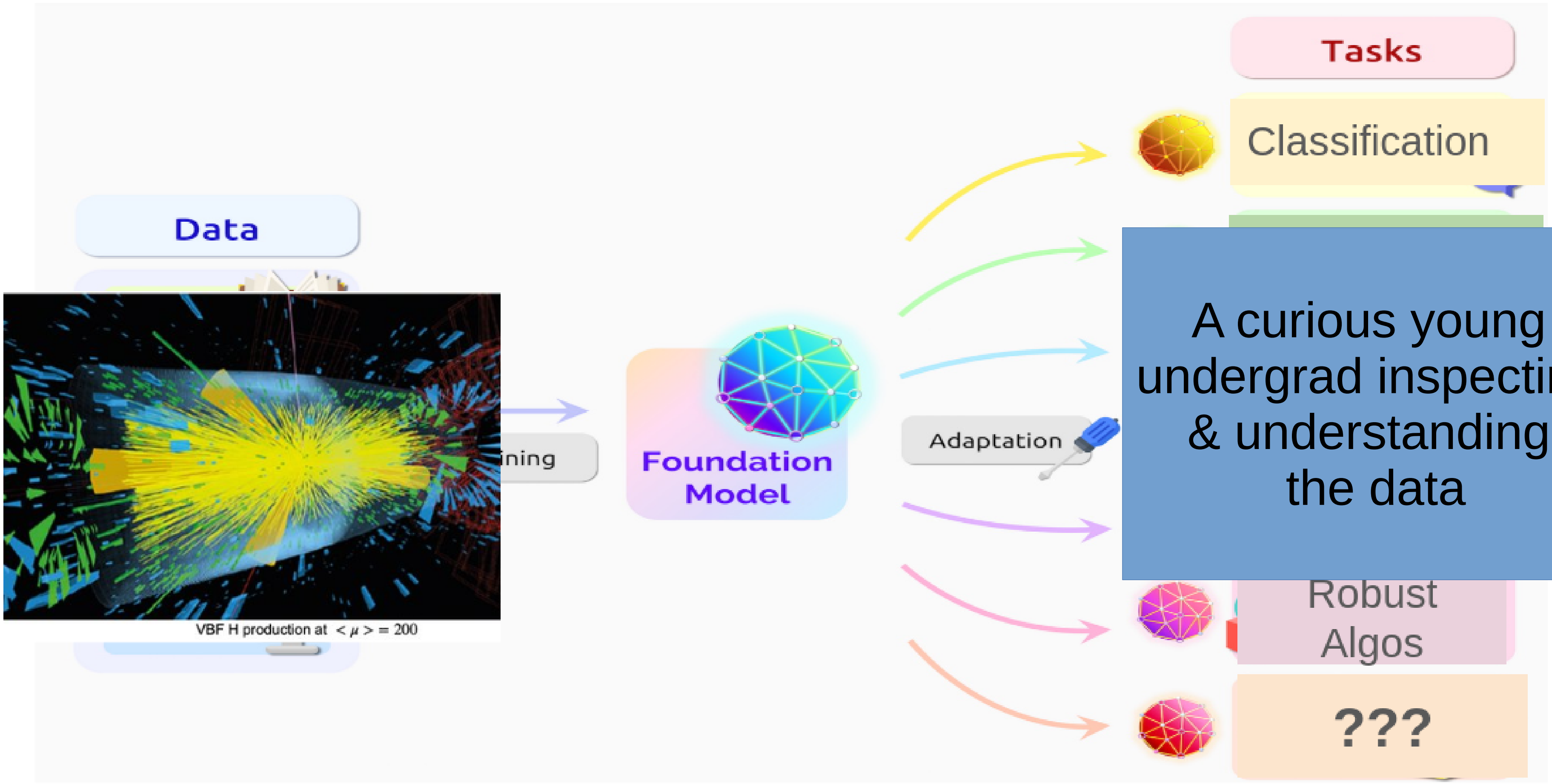



Image source: arXiv:2108.07258

# Foundation Models



# Foundation Models




A curious young undergrad inspecting & understanding the data

# How do we get there?

## LLM Agents

ChatGPT

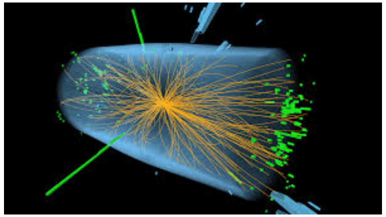
Gemini

Claude

Understands ‘**the world**’(?)

Given tasks, iterate, interpret

## Physics Understanding



Models that understand **our data**

Can ‘manually inspect’ at scale

# How do we get there?

## LLM Agents

ChatGPT

Gemini

Claude

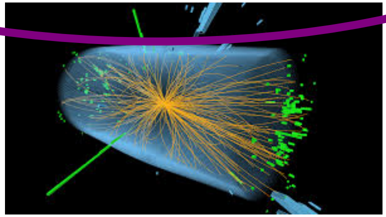
Understands ‘**the world**’(?)

Given tasks, iterate, interpret

?

Bridge?

## Physics Understanding



Models that understand **our data**

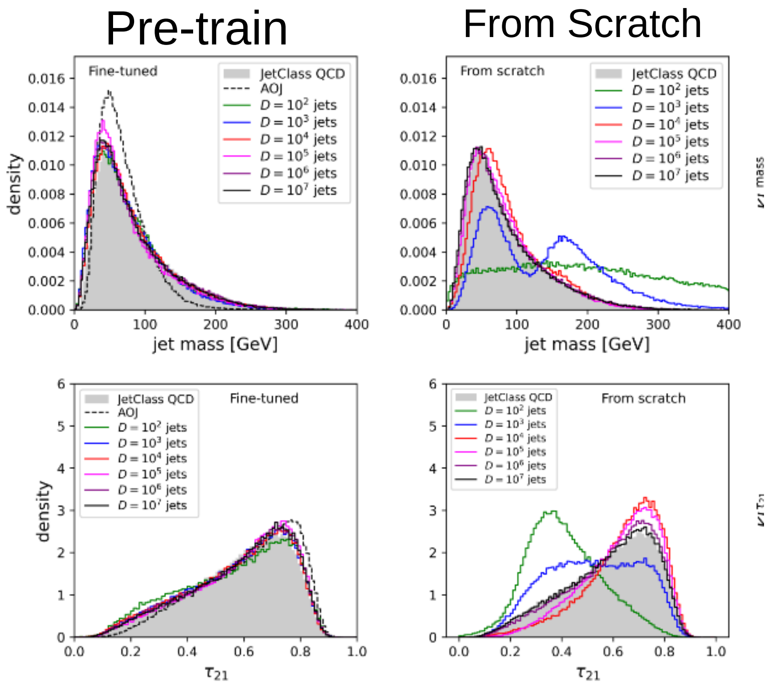
Can ‘manually inspect’ at scale

# How to develop understanding?

- One approach: Generation
- “What I cannot create I do not understand” – Feynman
- Many other possibilities!

Masking particles (2401.13537) , large supervised (2405.12972), mix gen. & sup. (2510.24066), …

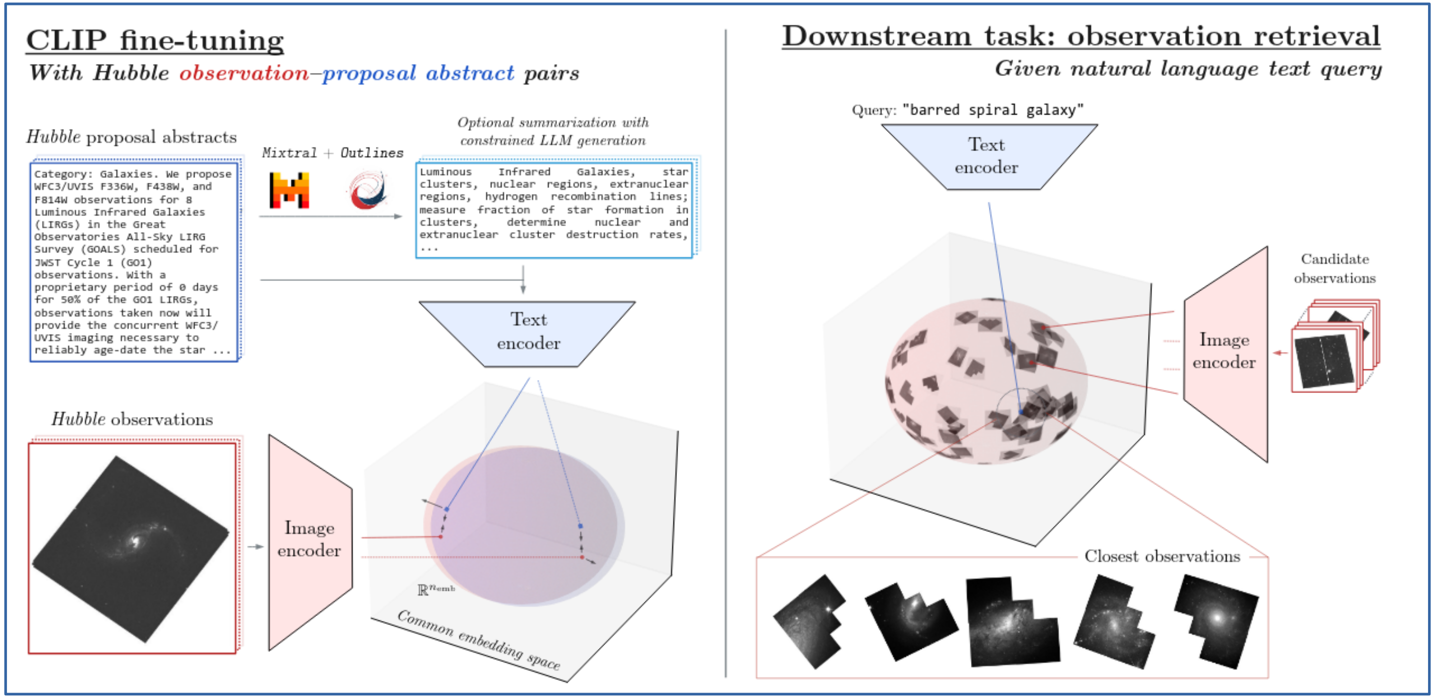
Pre-train on 180M jets from CMS open data → fine tune on 100-1M Delphes Sim. jets




**OA** et al arXiv:2412.10504

# How to build the bridge?

PAPERCLIP (2403.08851)




How would we do something similar for HEP/? data?

What capabilities could it unlock?

Built a text ↔ data (image) mapping with a contrastive pre-training

# Ingredient

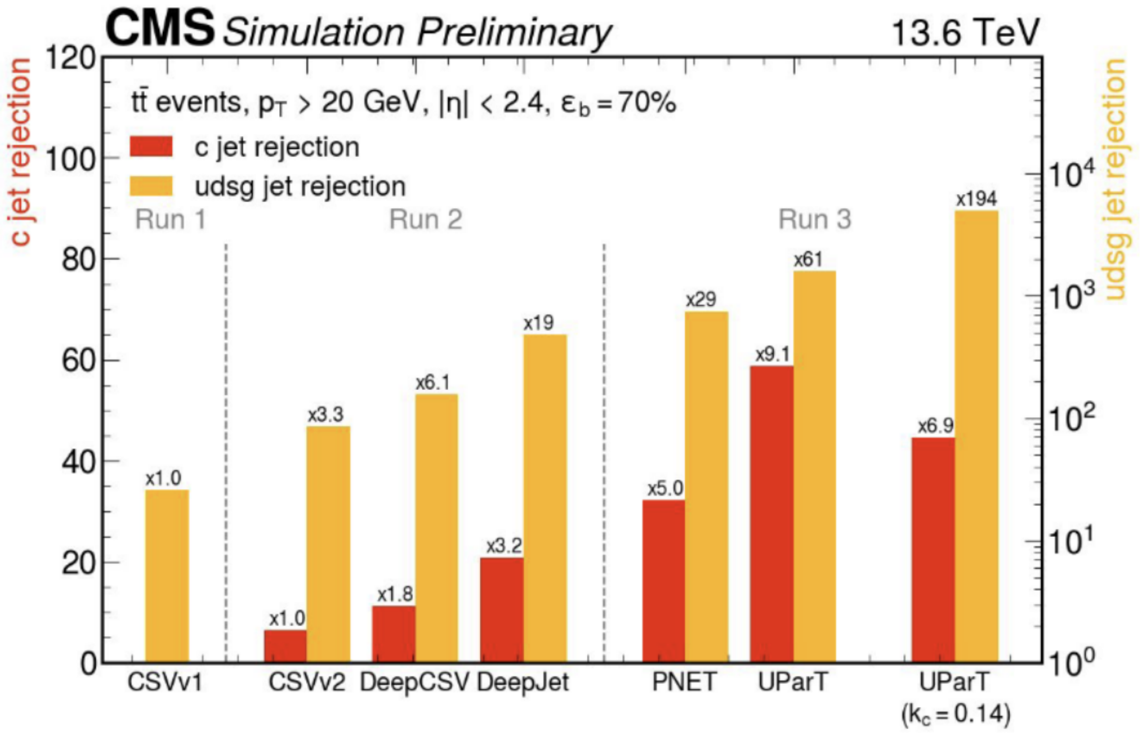
- High quality scientific data
- Clear indication of outlier
- Understanding of backgrounds
- Short analysis timescale
- A curious young scientist inspecting the data

# AI Technology

- Anomaly Detection
- Generative Models
- AI Design Optimization
- Foundation Models

# Previous Era: AI as Tool

‘Simple’ questions, complex answers



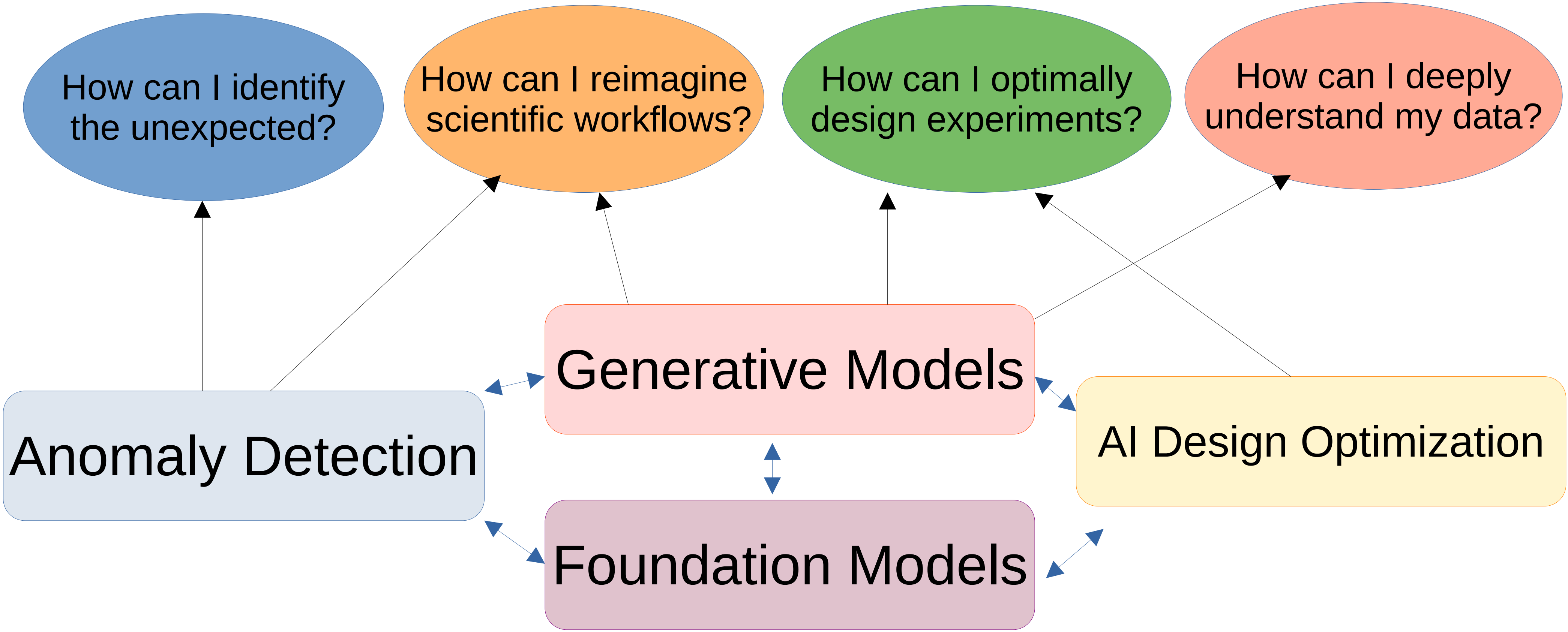

Fully connected → CNN → Graph → Transformer

Asking same question for the last ten years “Is this a bottom quark jet?”

Now 200x better!

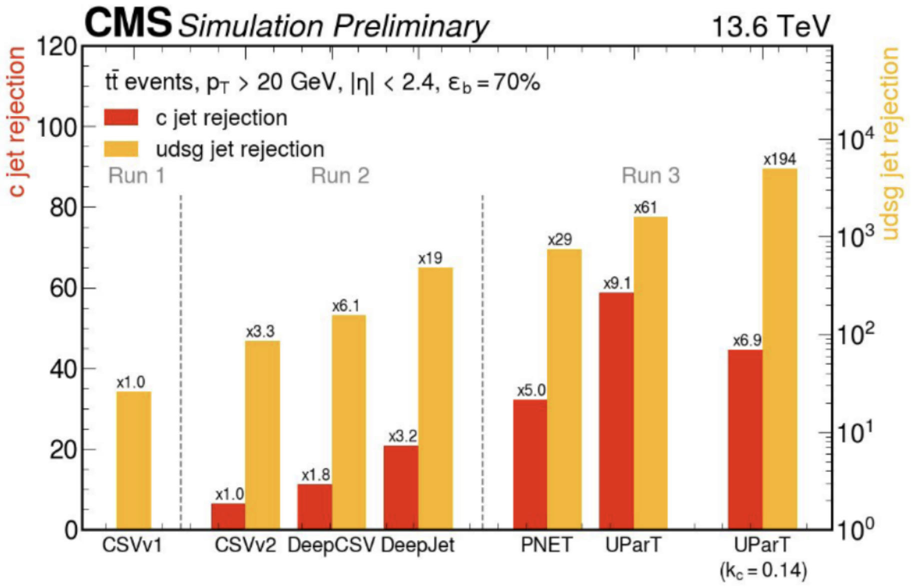
# AI as a Paradigm Shift

Ask **new questions**, only possible with AI



# AI as a tool

‘Simple’ questions, complex answers




Fully connected → CNN → Graph → Transformer

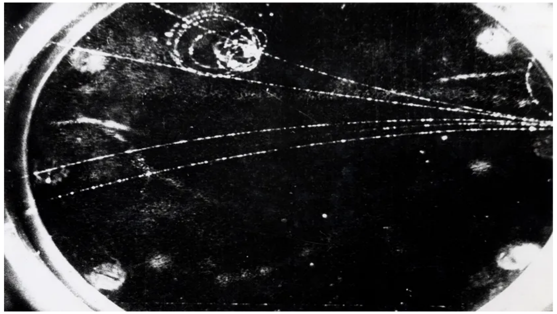
# AI as a paradigm shift

Ask **new questions**, only possible with AI

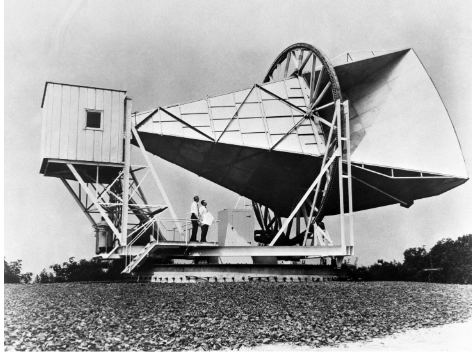

iml-wg.github.io/HEPML-LivingReview

# Whatever comes next in fundamental physics will be surprising

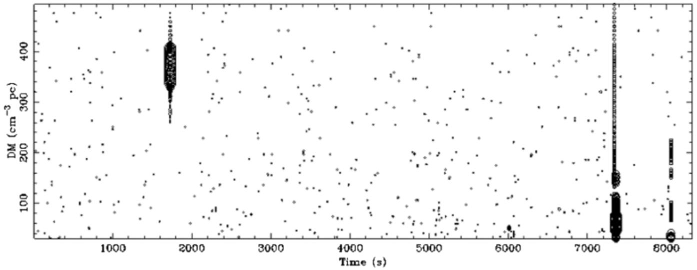
1936



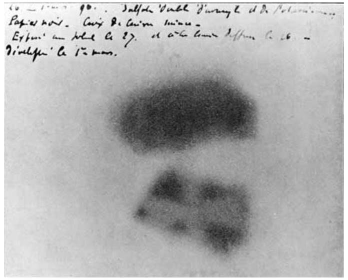
1964



2006




1896



**2026?**

## Lets be ready for it!

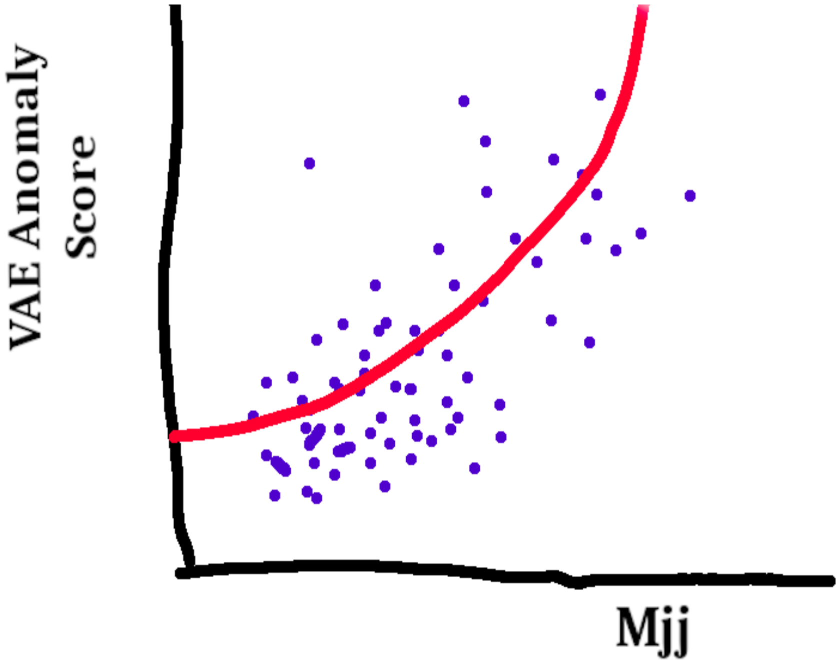
# Conclusions

Whatever comes next in particle physics will be surprising

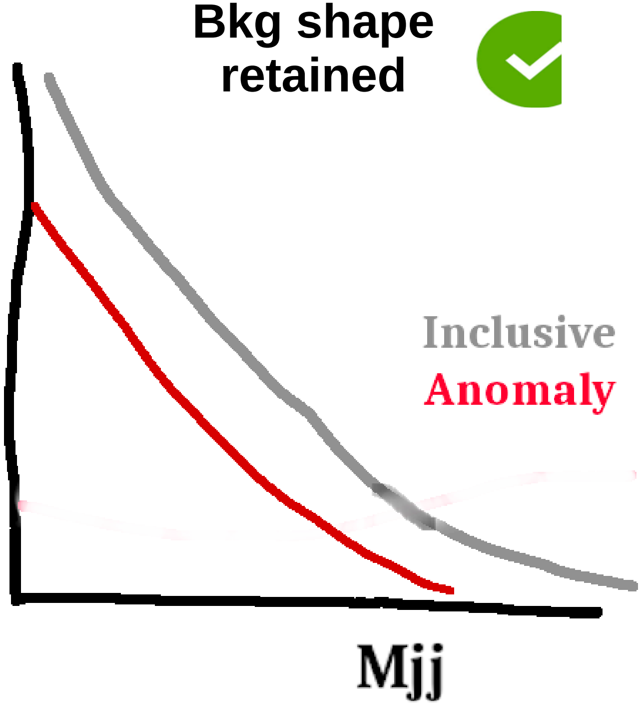
Lets let ourselves be surprised again!

# Backup

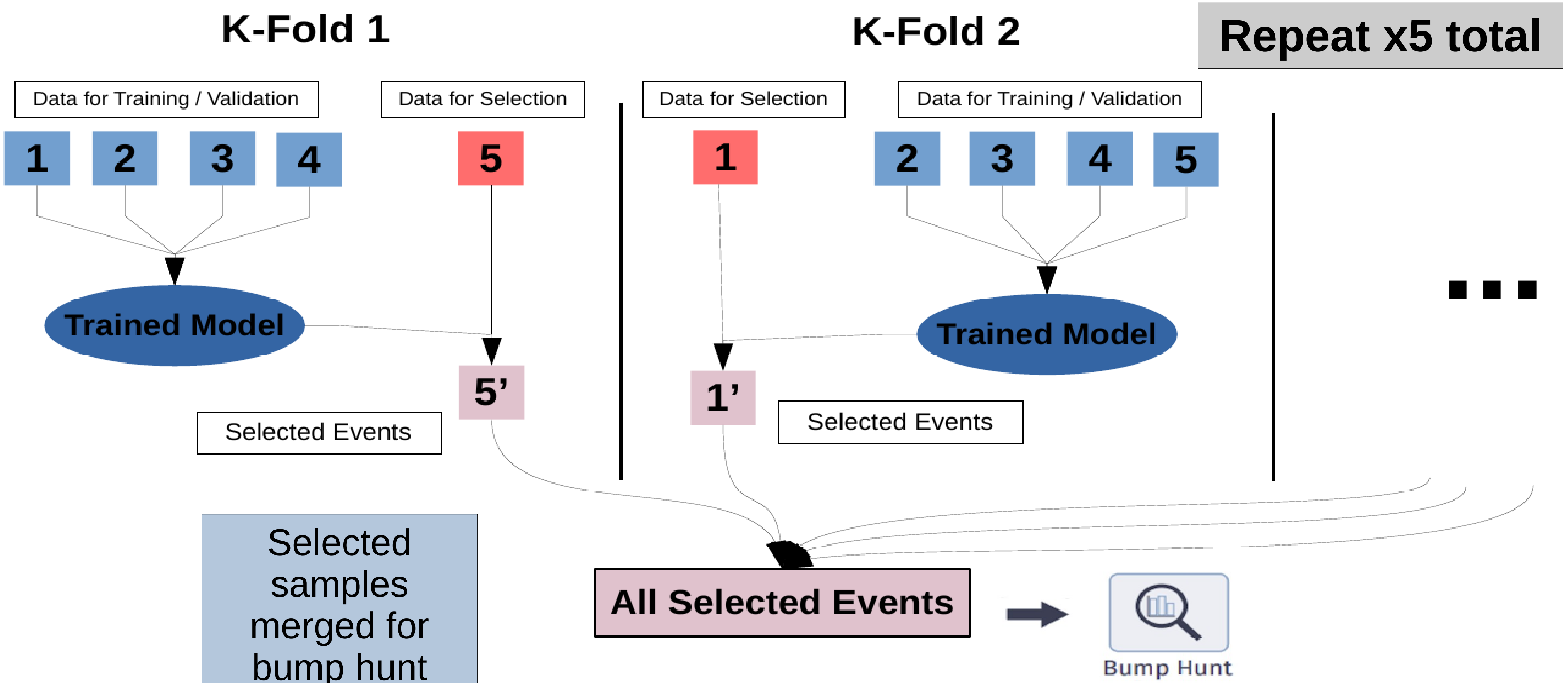
# Decorrelate with Mjj




**Cut with Flat Eff.**



# Cross Validation

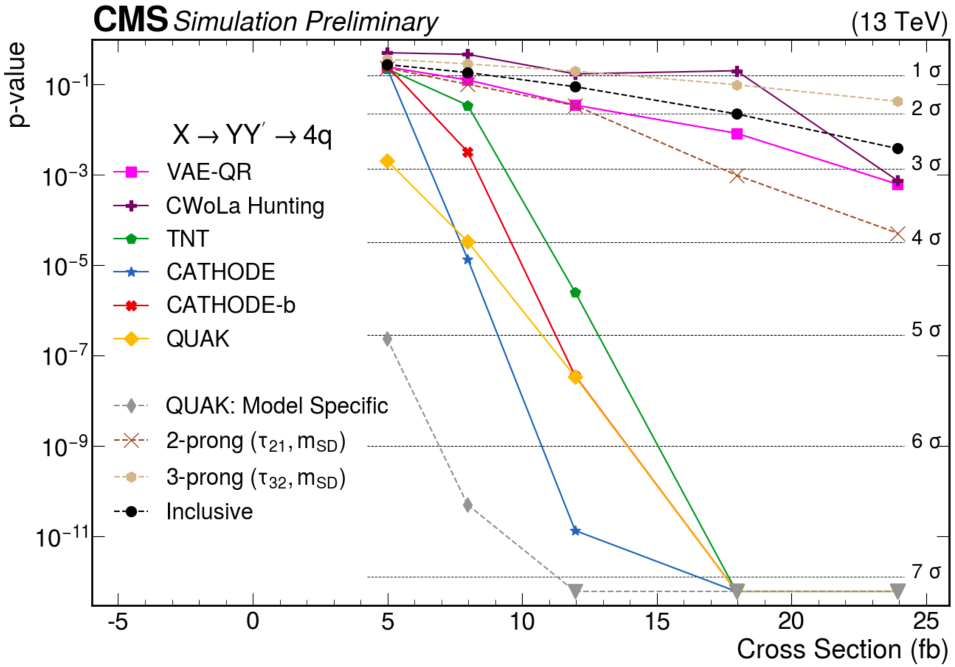



Selected samples merged for bump hunt

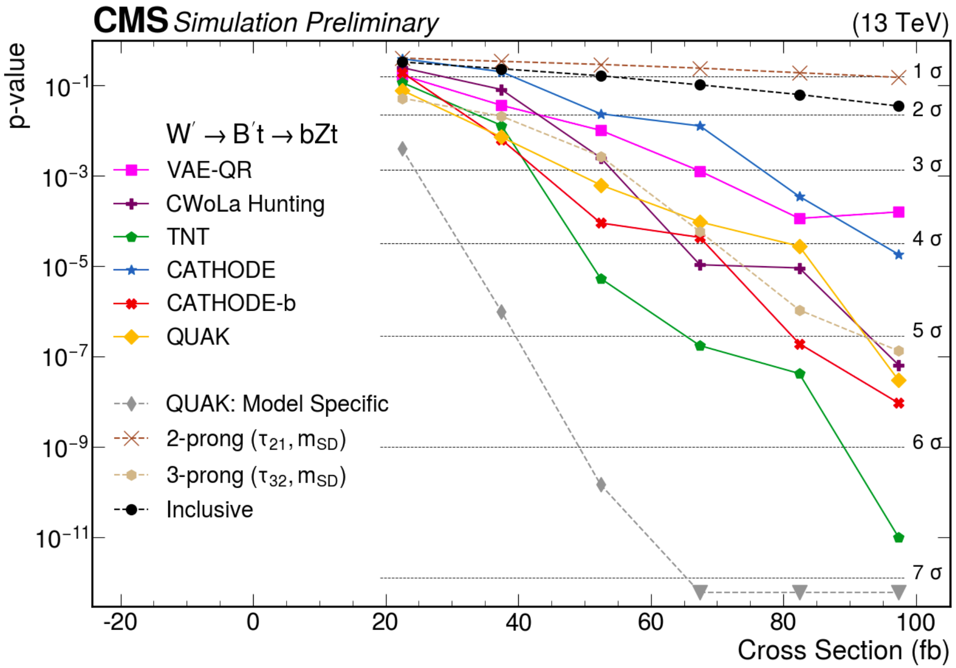
*Weakly supervised methods use additional layer of cross val for stability (see backup)

# Sensitivity

## 2 Pronged Signal

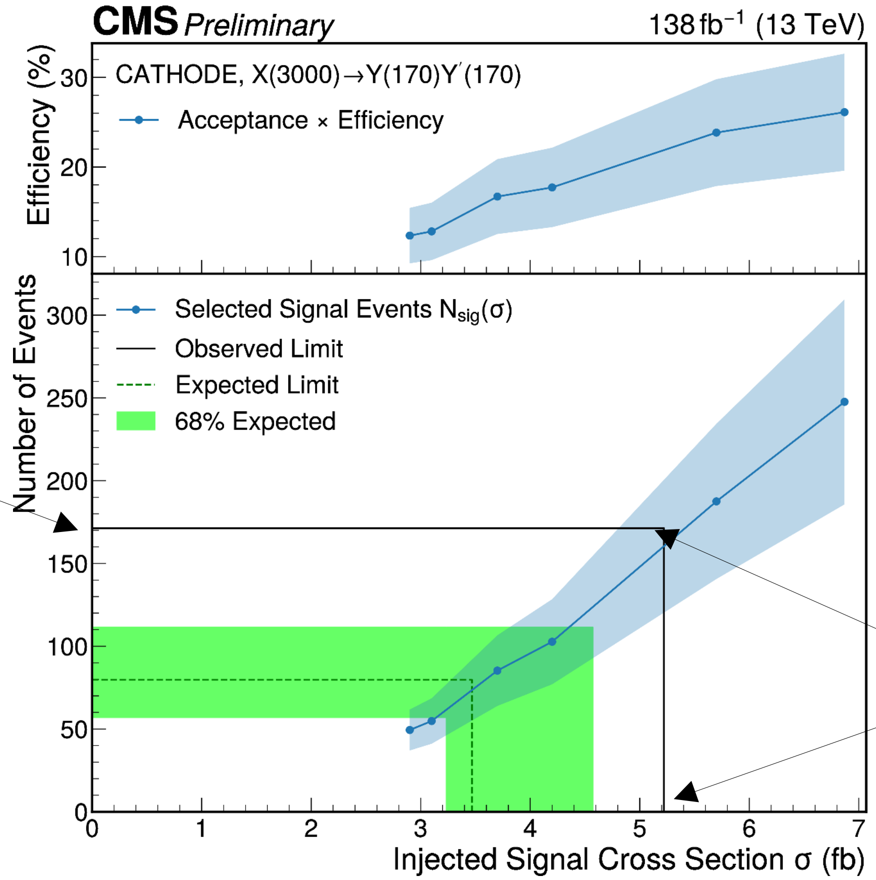



## 3 Pronged Signal



# Limits with Changing Eff.

Limit on # of signal events in SR from fit ($N_{exc}$)




Find $N_{sig}(\sigma)$= $L*\sigma*\varepsilon(\sigma)$ that matches $N_{exc}$

→ σ is limit

# Input Features

Low-level features

Hand-picked high-level features

**VAE**

Jet Constituents
$p_x$, $p_y$, $p_z$

**CWoLa Hunting**

Jet mass
$\tau_{21}$
$\tau_{32}$
$\tau_{43}$
$N_{const}$

Leptonic energy frac.

Sub-jets b-tag score

**TNT**

Same as CWoLa Hunting

**CATHODE**

Jet masses
$\tau_{41}$'s

-------------------

**CATHODE-b**

**+** Subjet b-tag scores

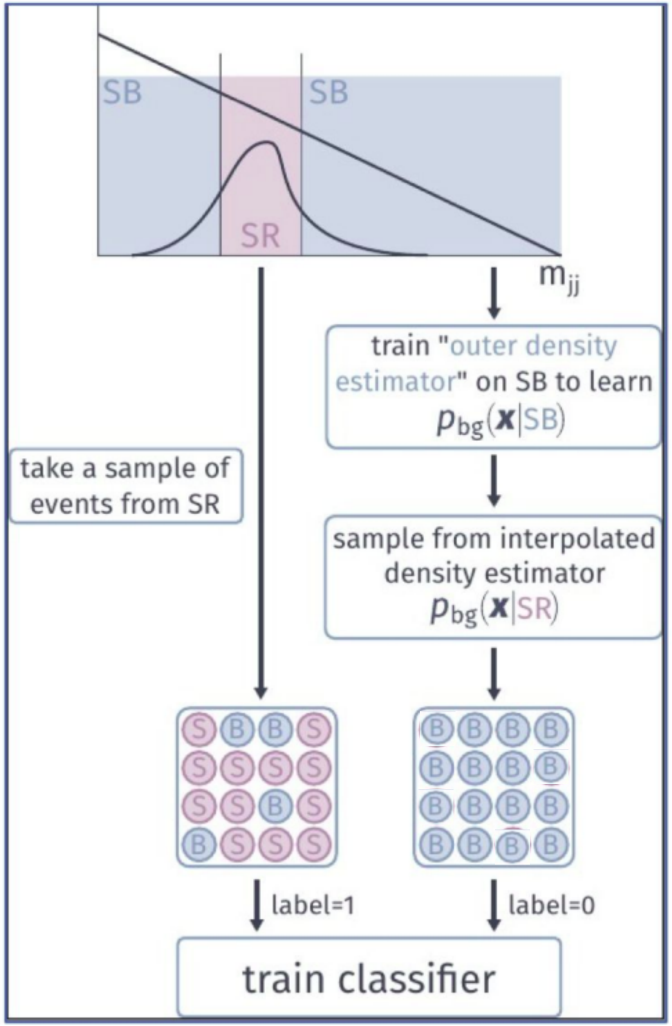
**QUAK**

$\rho$ = jet mass / $p_T$
$\tau_{21}$'s
$\tau_{32}$'s
$\tau_{43}$'s
$N_{const}$'s

$\sqrt{\tau_{21}}/\tau_1$

Sub-jets b-tag scores

# CATHODE




- Learn full multi-dim density $P_{bkg}( x \mid Mjj )$ from sidebands & **interpolate** into SR
  - 'Normalizing Flow'
- Draw samples to construct **bkg-rich sample**
- Weak supervision btwn data in SR and interpolated bkg samples

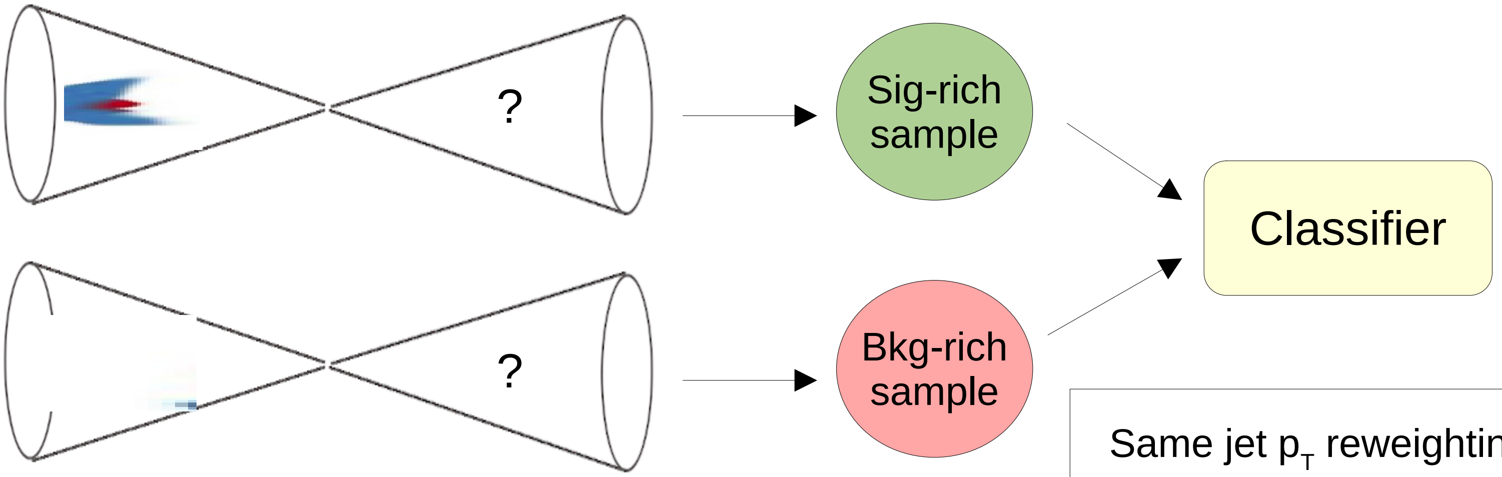


[Hallin et al 2109.00546]

# Tag N' Train (TNT)

- Similar to CWoLa Hunting, but additional assumption that for signal **both jets are anomalous**
- Enhance purity of mixed samples by first tagging one jet each SR event with an autoencoder

AE is a CNN based on jet images

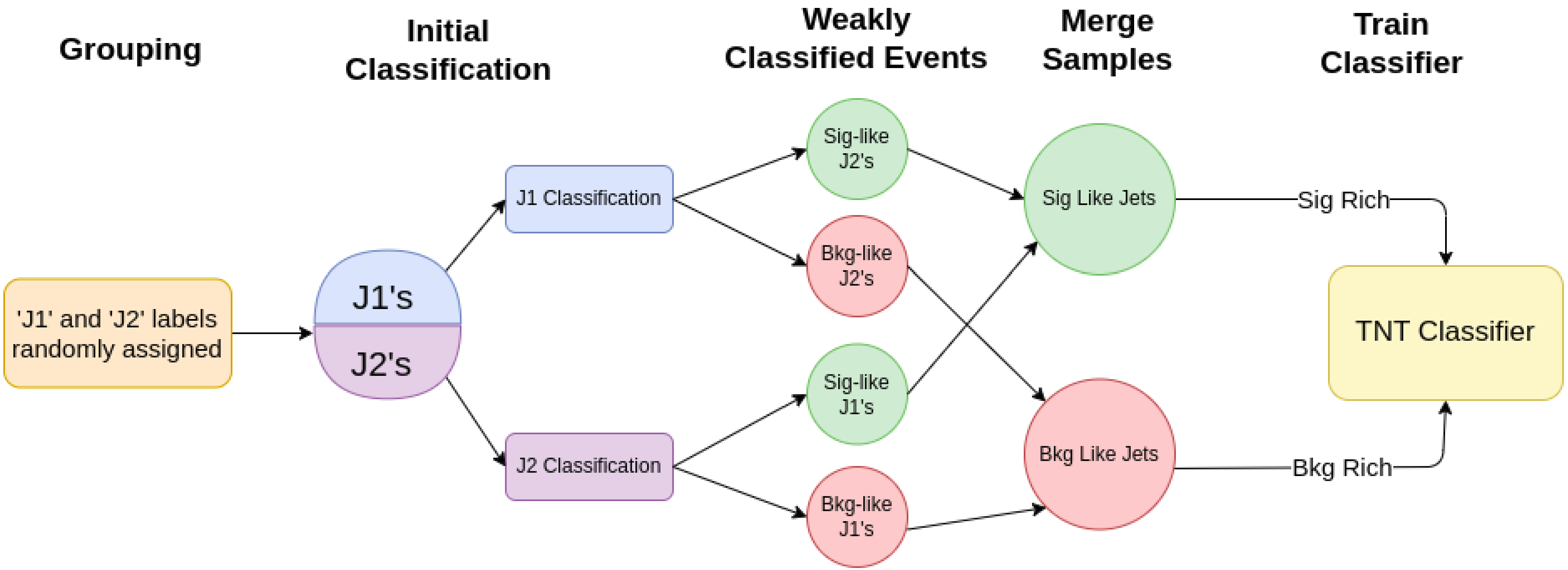



(full algo in backup)

Same jet $p_T$ reweighting method to ensure no correlation with Mjj

[**OA** & Suarez 2002.12376]

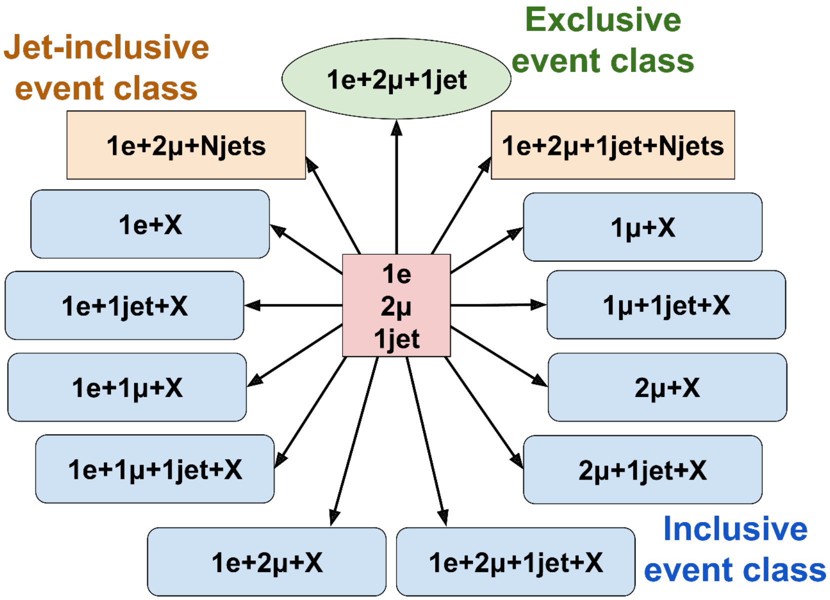
# TNT Diagram


Grouping
Initial Classification
Weakly Classified Events
Merge Samples
Train Classifier
'J1' and 'J2' labels randomly assigned
J1's
J2's
J1 Classification
J2 Classification
Sig-like J2's
Bkg-like J2's
Sig-like J1's
Bkg-like J1's
Sig Like Jets
Bkg Like Jets
Sig Rich
Bkg Rich
TNT Classifier

# Classic Strategy

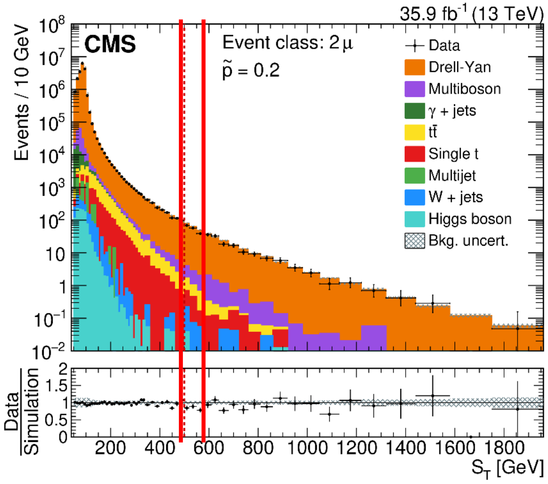
Using CMS MUSiC Search as an example

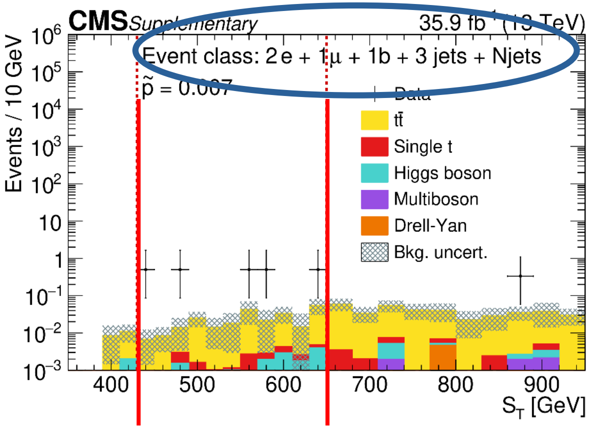
## Categorize

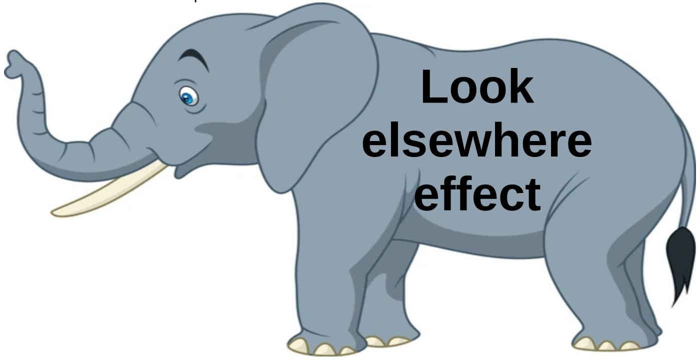



~1.5k event classes

## Data-MC Comparison



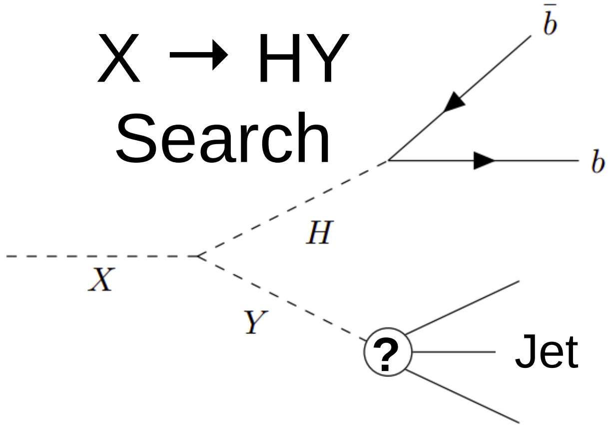



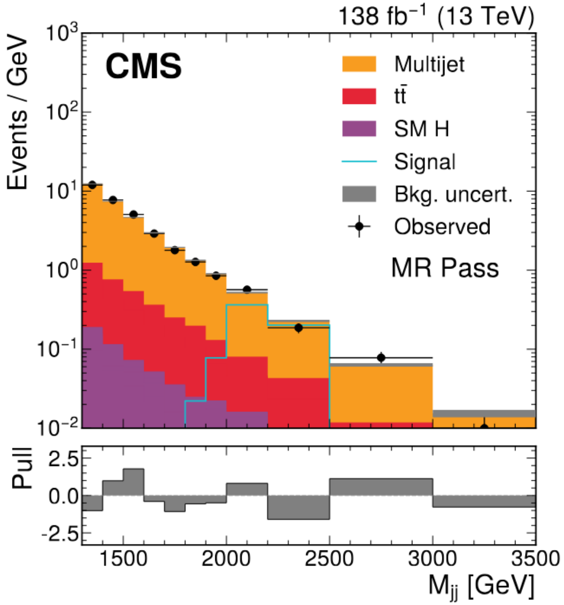



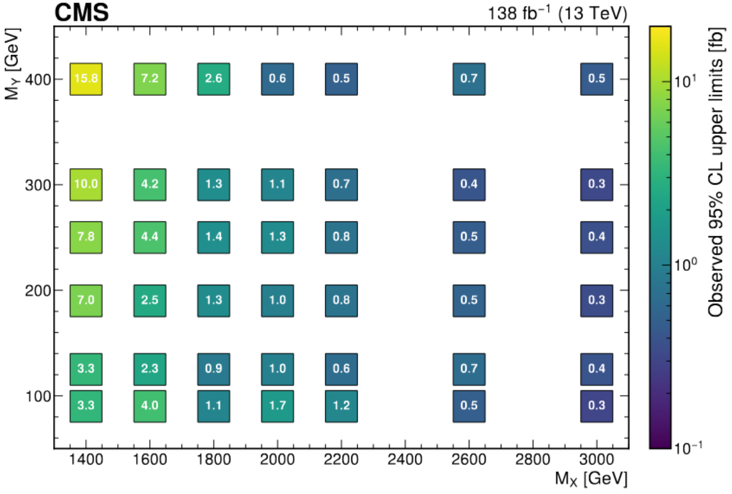
# More Anomalous Jet Searches

**New!**
2509.13635

X → HY Search



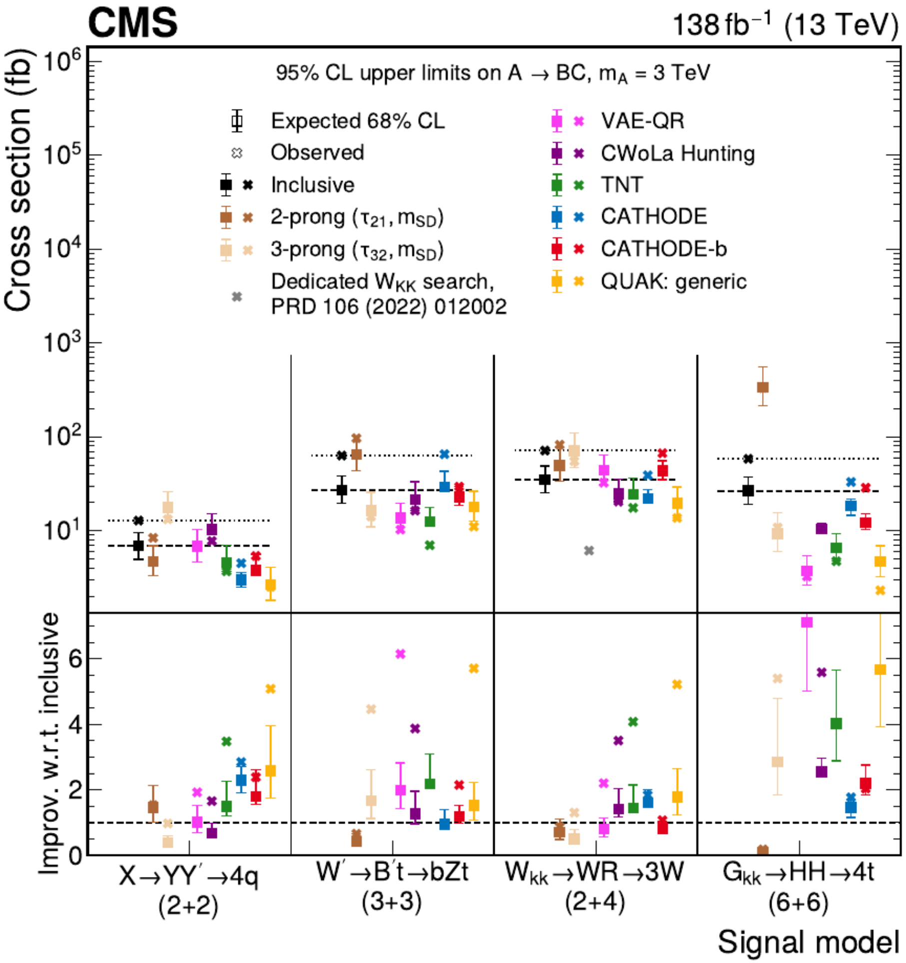




- Re-using an AE from the dijet search
- Limits on Y→WW and Y→bqq
  - ~Close to dedicated search for Y→ WW ! B2G-23-007

# Limits

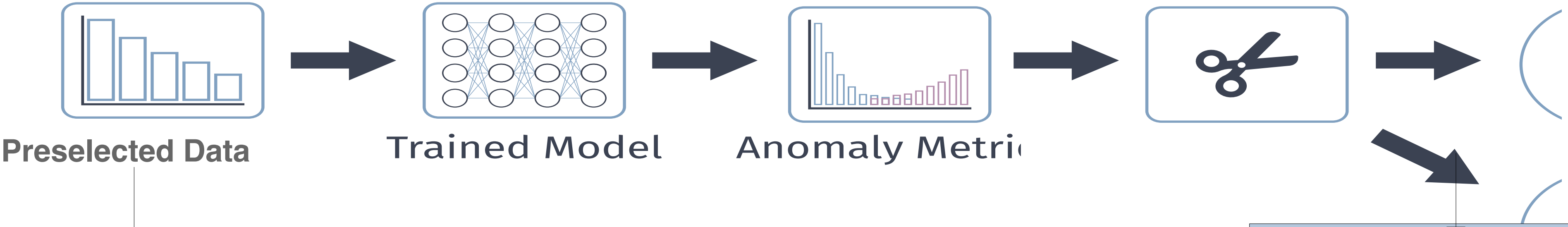
- Compute limits on benchmark from all **anomaly methods** on variety of signal models
  - Compare against **inclusive** & traditional **model-specific** approaches
  - First-ever limits on several models!

- **Anomaly detection** improves limits by **~2-6x**!
  - Does not reach sensitivity of **dedicated search**




Very different sig. models!

# Analysis Overview




**2 large radius (AK8) jets**

**Signal Region**
$|\Delta\eta| < 1.3$

**Control region**
$2.0 < |\Delta\eta| < 2.5$ *

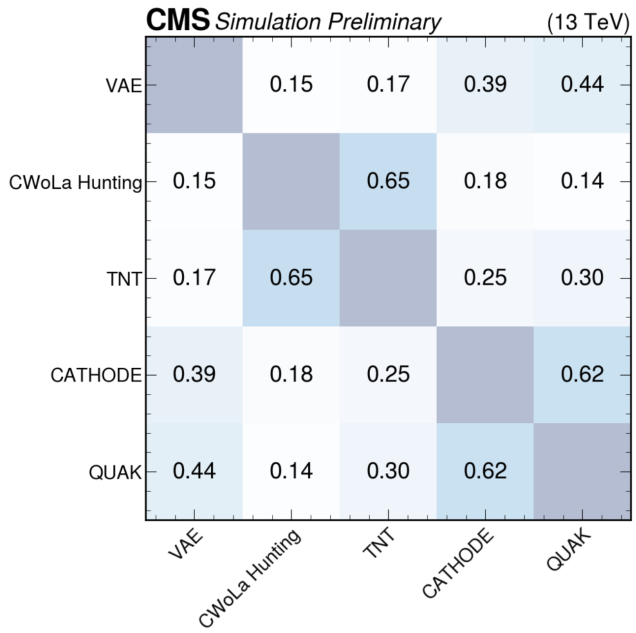
**Bkg**: Standard Dijet Parameterization
**Signal**: Double Crystal Ball

**Bumps?**

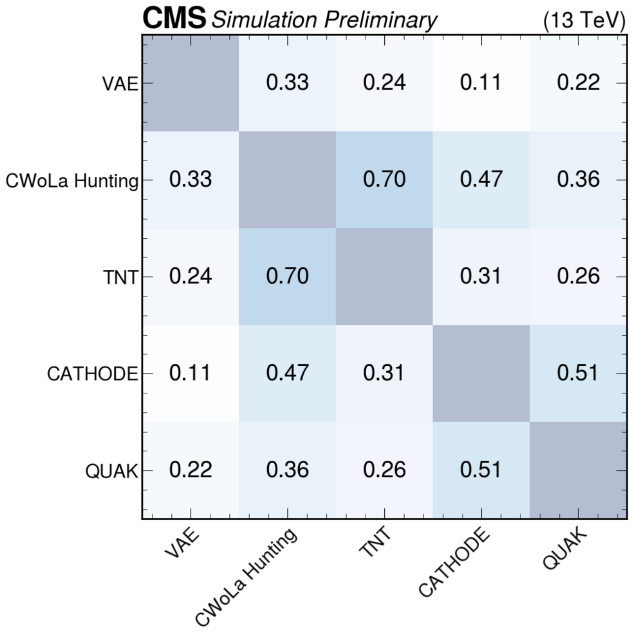
**Limits** on signal models

# “Are these five methods just learning the same thing?”

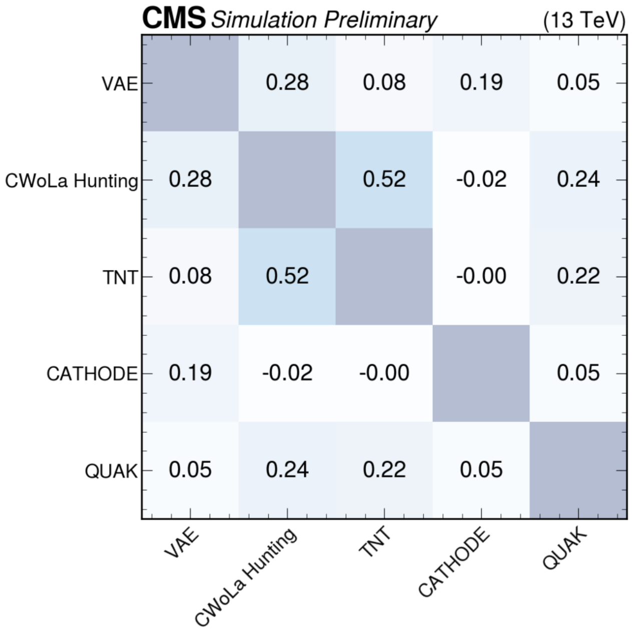
$X \rightarrow YY \rightarrow qq\,qq$




$W' \rightarrow B't \rightarrow bqq\,bqq$



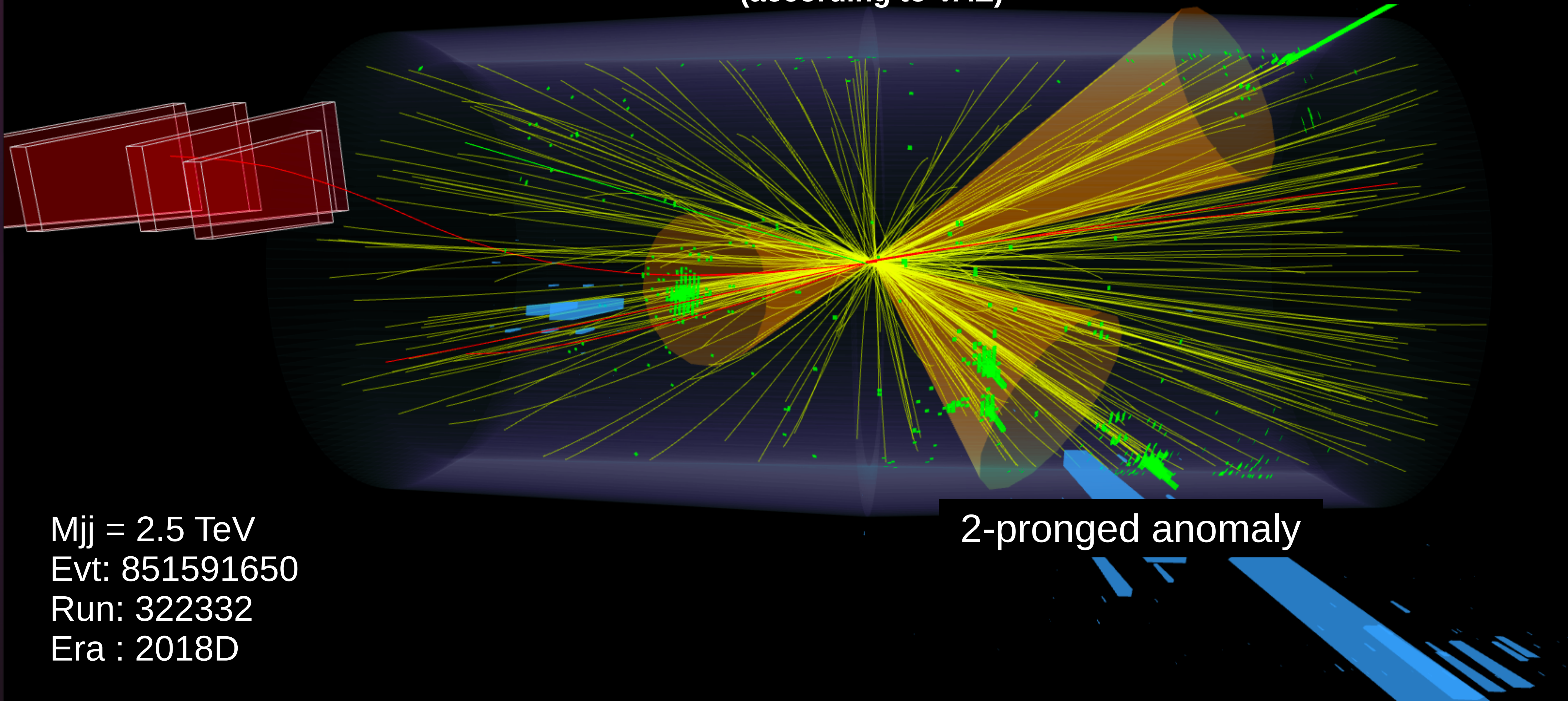

QCD Bkg.




- Compute **correlation coefficients** between different anomaly scores
- Relatively **low** correlations → methods are **complementary**!

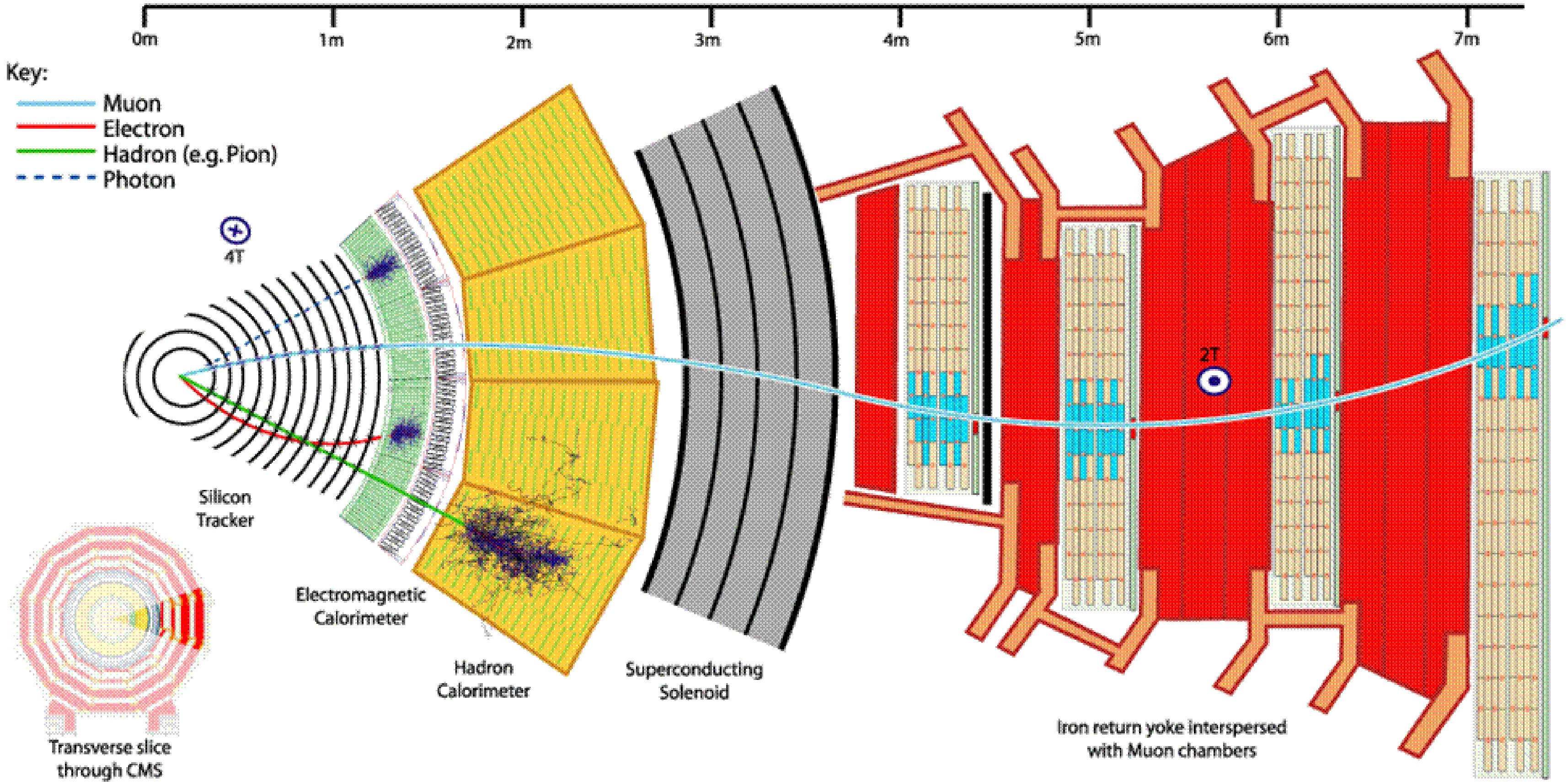
# One of our most anomalous events!

**(according to VAE)**




Mjj = 2.5 TeV
Evt: 851591650
Run: 322332
Era : 2018D

# CMS Reconstruction



# Standard Measurements

We typically formulate a search/inclusive measurement as:

$$P_b(x) + s * P_s(x)$$

Bkg probability distribution

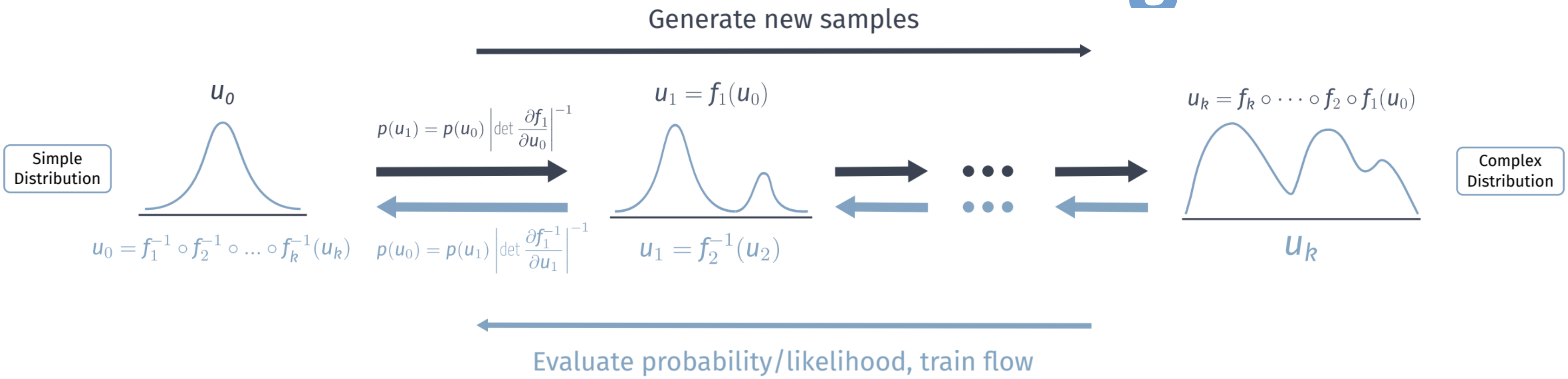
Observable(s) you are fitting

Signal strength

Signal probability distribution

NB: I'm will be sloppy with formalism, normalizations, nuisances, etc. for simplicity
Full formalism in paper / backup

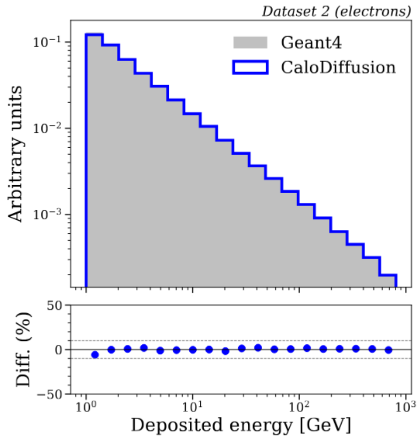
# ML Model : Normalizing Flows

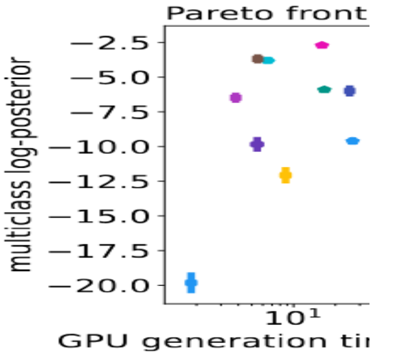


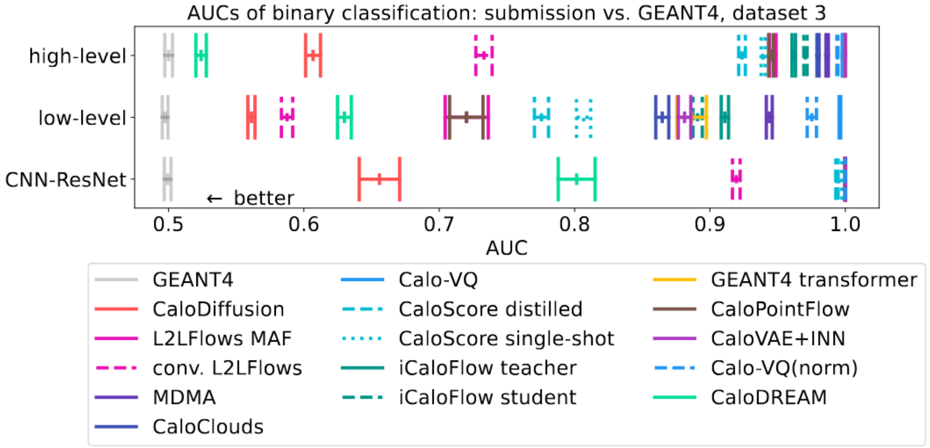

- Maps a complex multivariate dist. to a **standard multivariate Gaussian** via series of learnable **invertible** maps
- **Generate**: Sample from the Gaussians → apply maps 'forward'
- **Evaluate the density:** Apply inverse maps to data, evaluate likelihood of Gaussian
- Density for fitting, generation for visualization

# CaloChallenge Results

- Significant advance in SotA
  - First time AI showers not ~100% distinguishable from Geant
- Diffusion / Flow Matching models have best performance
- Tradeoff between quality and generation time




2410.21611

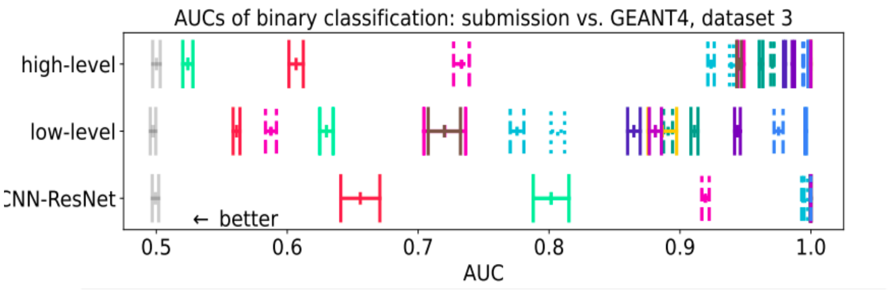
# Performance Comparison

- Calodiffusion performed very well in CaloChallenge
- New benchmark in quality at time of publication
- Top-2 in quality on all datasets in final evaluation
  - Out of 50 total submissions

Comparisons at time of publication

| Dataset | Classifier AUC (low / high) CaloDiffusion | CaloFlow | CaloScore v2 |
|---|---|---|---|
| 1 (photons) | **0.62** / 0.62 | 0.70 / **0.55** | 0.76 / 0.59 |
| 1 (pions) | **0.65** / **0.65** | 0.78 / 0.70 | - / - |
| 2 (electrons) | **0.56** / **0.56** | 0.80 / 0.80 | 0.60 / 0.62 |
| 3 (electrons) | **0.56** / **0.57** | 0.91 / 0.95 | 0.67 / 0.85 |

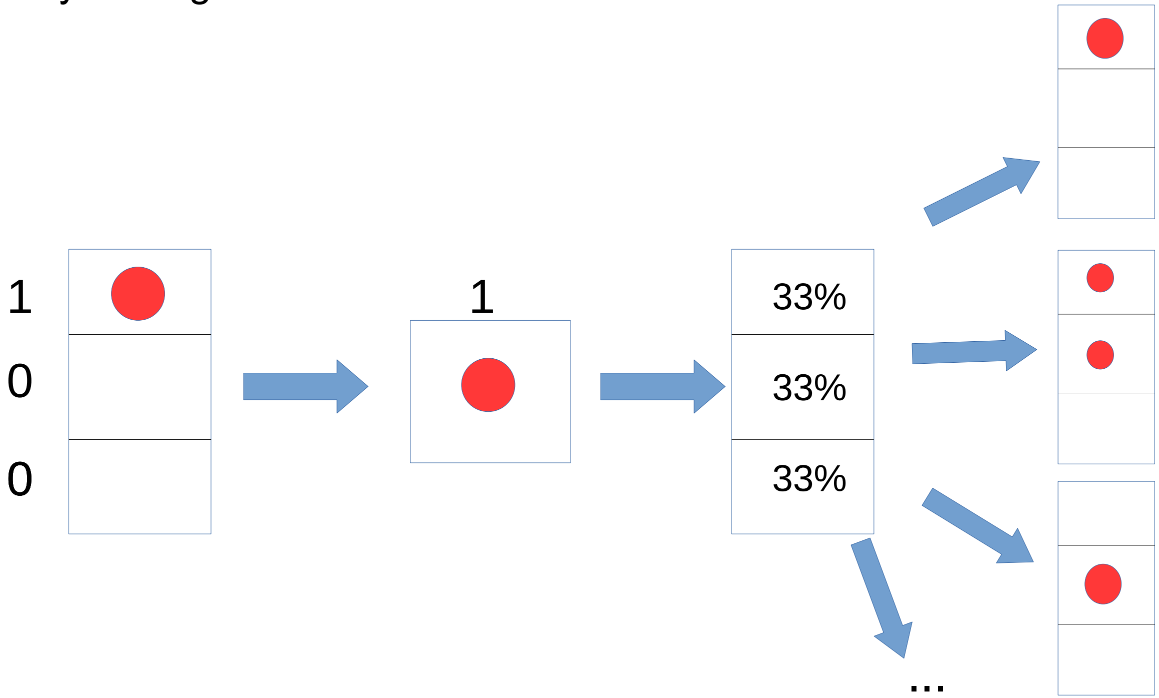
Final CaloChallenge Results




**CaloDiffu**

# Sparsity-Based Sampling

- Compressing data naturally results in ‘smearing’
- Instead of fractional sharing of energy, use fractions as **probabilities** for each cell to be non-zero
- Random sample from these probabilities to pick non-zero cells
  - Require at least one to be non-zero (avoid energy loss)
- Split energy evenly among chosen cells



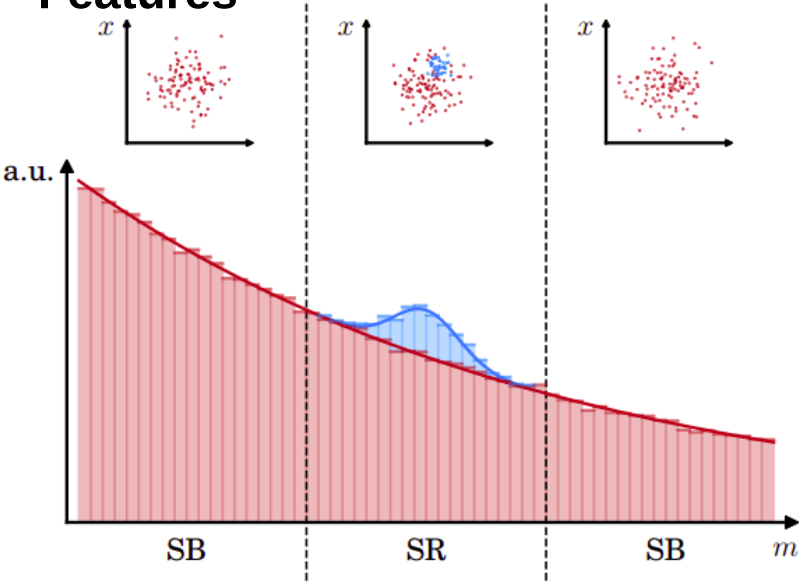
# HI-SIGMA

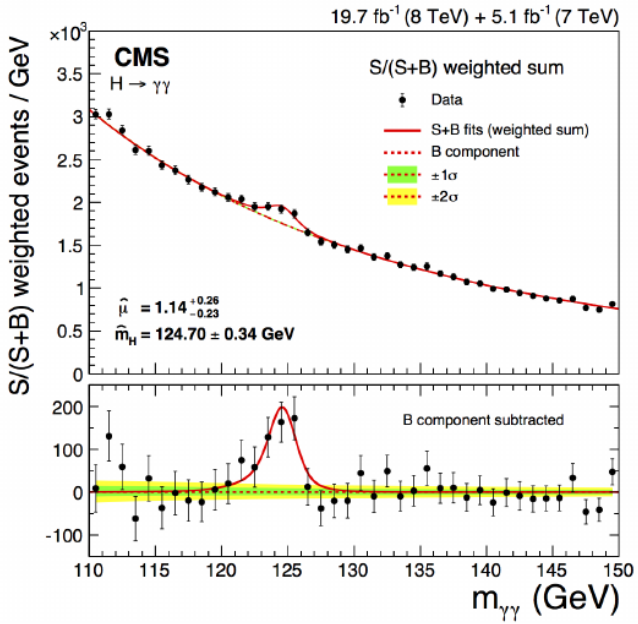
$$P_b(\vec{x}) = P_b(\vec{x}|m)P_b(m)$$

(CATHODE)

**1D Parametric Function**

**Features**




*We have been doing this anomaly detection for a while, but mostly haven't been using the learned density*

2109.00546

# HI-SIGMA

$$\underbrace{P_b(\vec{x}|m)}\underbrace{P_b(m)} + s * \underbrace{P_s(\vec{x}|m)}\underbrace{P_s(m)}$$

ML model trained on data sidebands, interpolated into SR

Standard parametric functional form (eg polynomials)

ML model trained on signal MC

Standard parametric function (eg Gaussian)

***And then you fit!***

*Resonant / non-smooth bkg's can be learned from MC*