CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Gefördert durch:

Bundesministerium
für Forschung, Technologie
und Raumfahrt

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

New Dijet Anomaly Detection Search with Run3

Jovin Drews, Louis Moureaux, Karim El Morabit, Gregor Kasieczka,
Chitrakshee Yede, Tore von Schwartz

Exotica Jets+X Working Group Meeting

January 19, 2026

Anomaly Detection Introduction

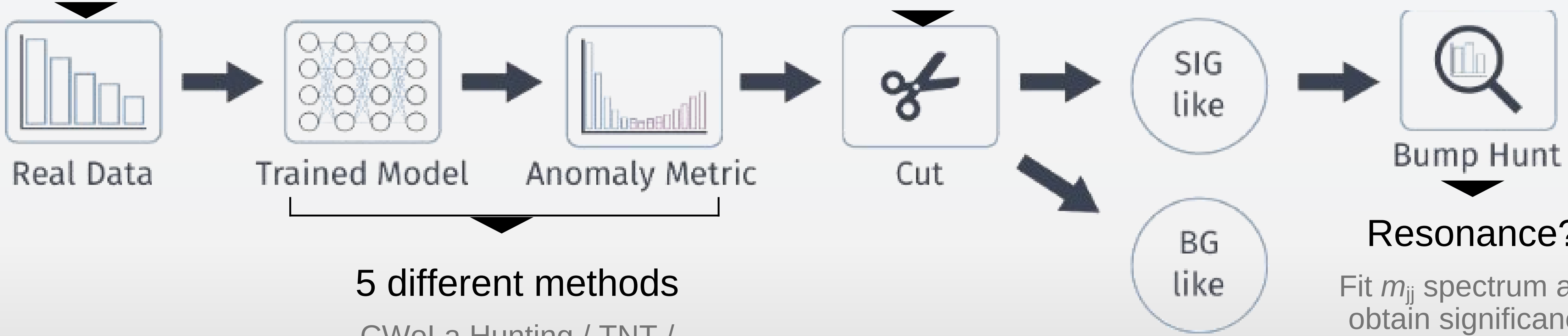
- New physics may appear in unknown signatures that traditional analyses miss
- Identify rare or unexpected events without explicit signal model
- If a new particle is produced, it appears as a peak in a kinematic distribution
- Choose a variable where the signal could appear (e.g., invariant mass)
- Estimate the background distribution
- Look for deviations/peaks above the expected background
- ML models filters events to signal-like ones using additional features
- Careful selection of features that are sensitive to the phenomena of interest

Analysis Strategy

Start from data

Anti-k_T jets with R = 0.8
Basic selection criteria

Keep ~1% most anomalous events



5 different methods

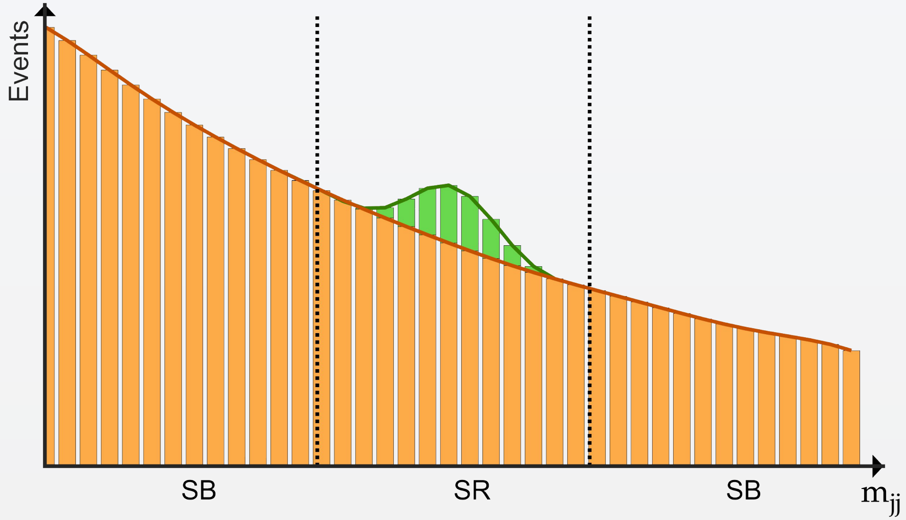
CWoLa Hunting / TNT / CATHODE(-b) / VAE / QUAK

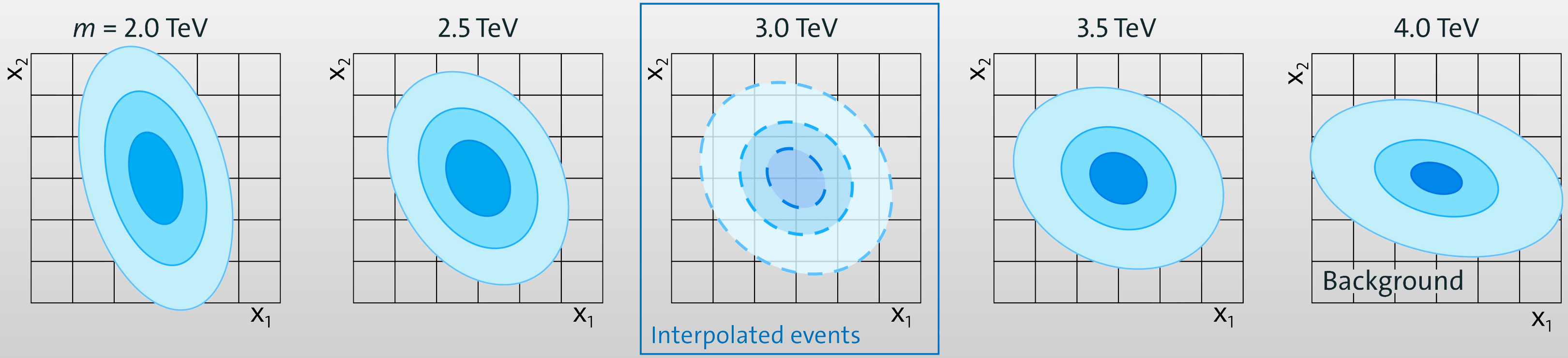
Resonance?

Fit m_{jj} spectrum and obtain significance

CATHODE

- Classifying Anomalies THrough Outer Density Estimation
- Select classification variables **x**
- Define signal region (SR) on the dijet mass m_{jj}
- Train generative model on sidebands
- Learn conditional density p(**x**|m)
- Generate new **x** samples using m values in the SR



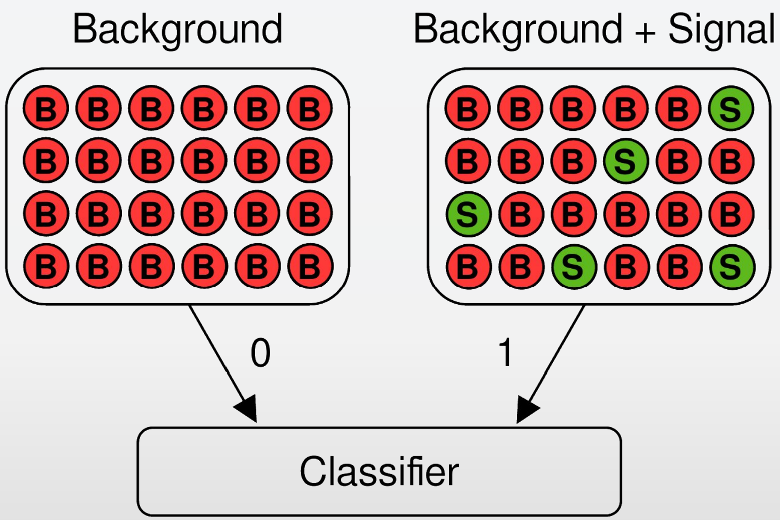


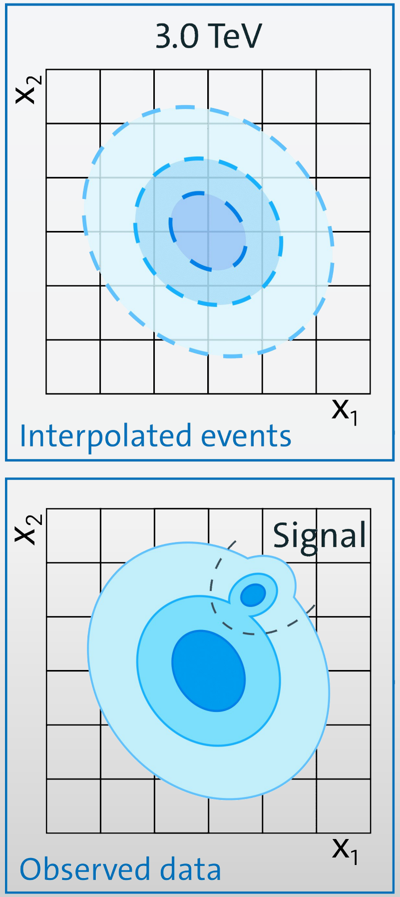
CATHODE

- Train a classifier between sampled bkg and data
- If there is signal it learns to classify it

$$R_{data}(x) = \frac{p_{data}(x)}{p_{bkg}(x)}$$

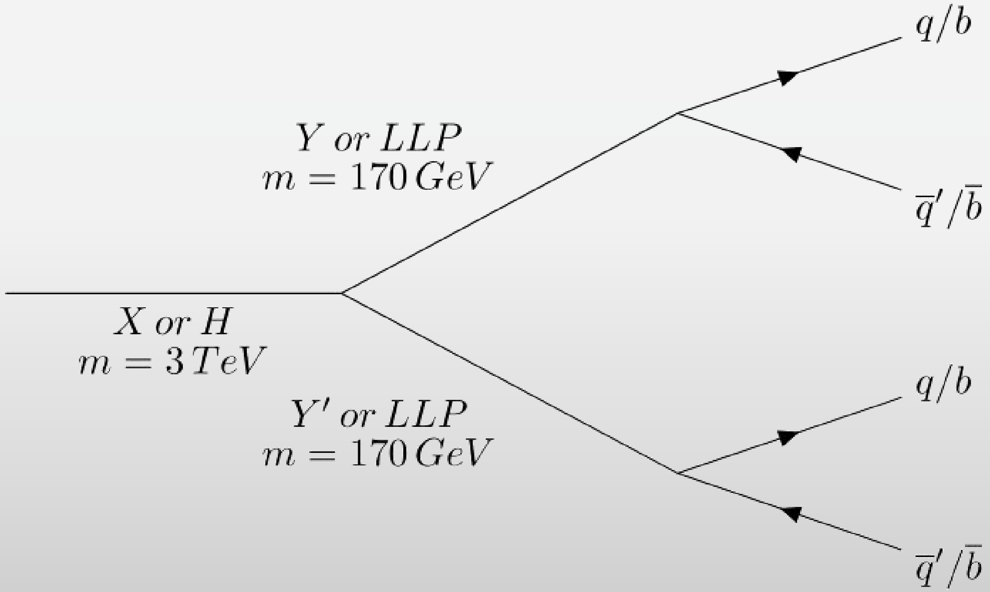
$$p_{data}(x) = (1-\varepsilon)p_{bkg}(x) + \varepsilon p_{sig}(x)$$

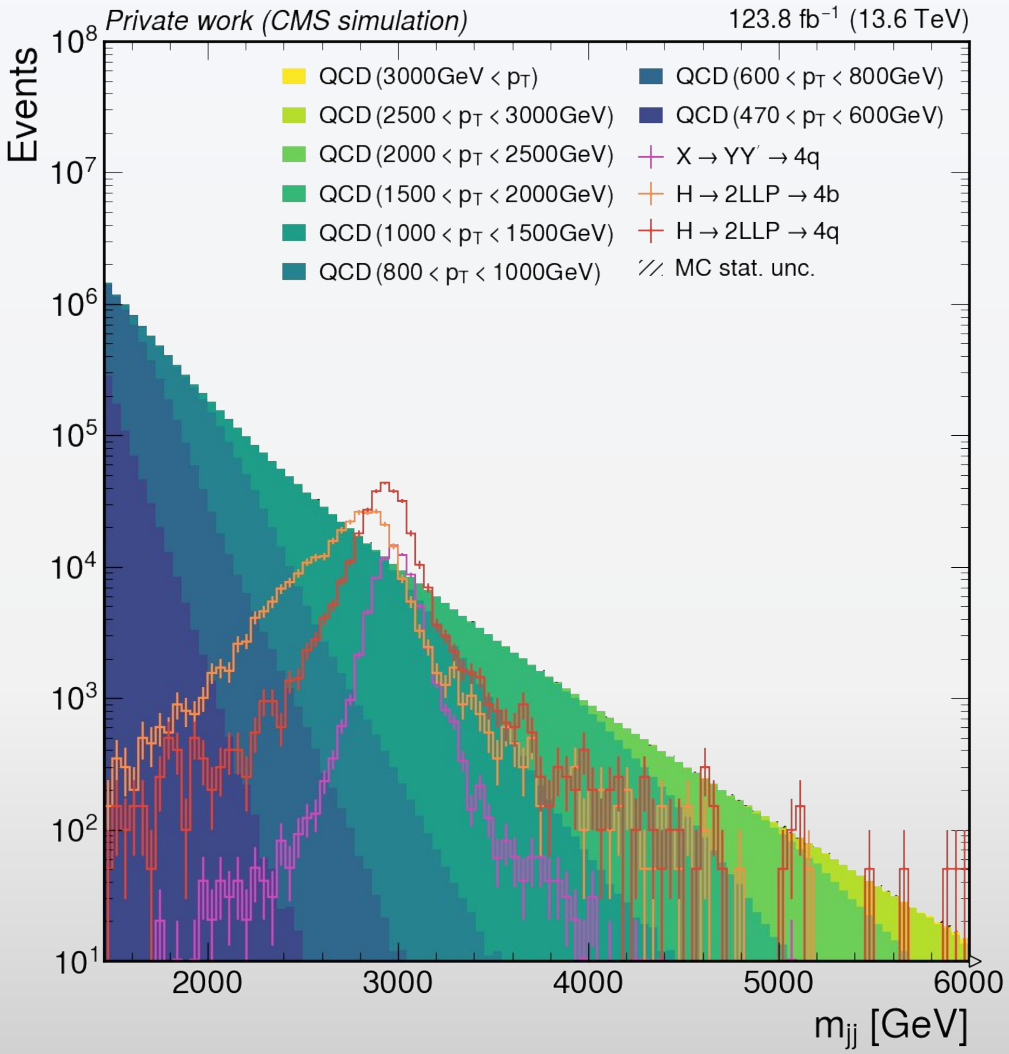




Datasets

- Simulated events only right now
- Focusing on 2024 LHC conditions
- Background: QCD Pythia pT-binned
- Prompt and long-lived signals



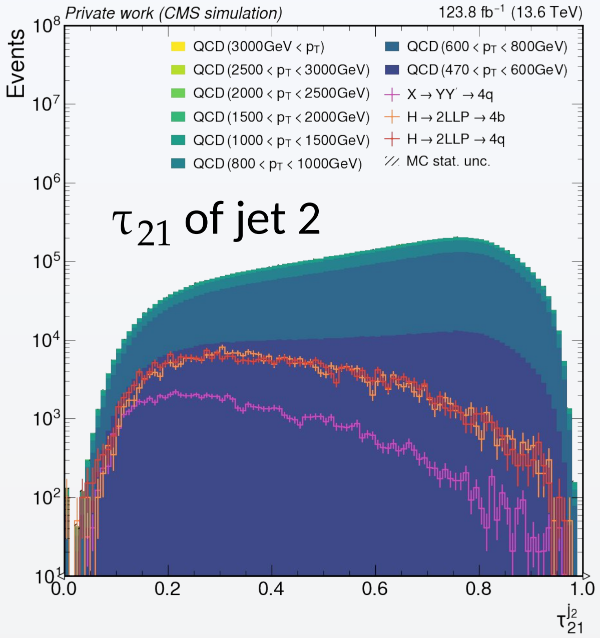


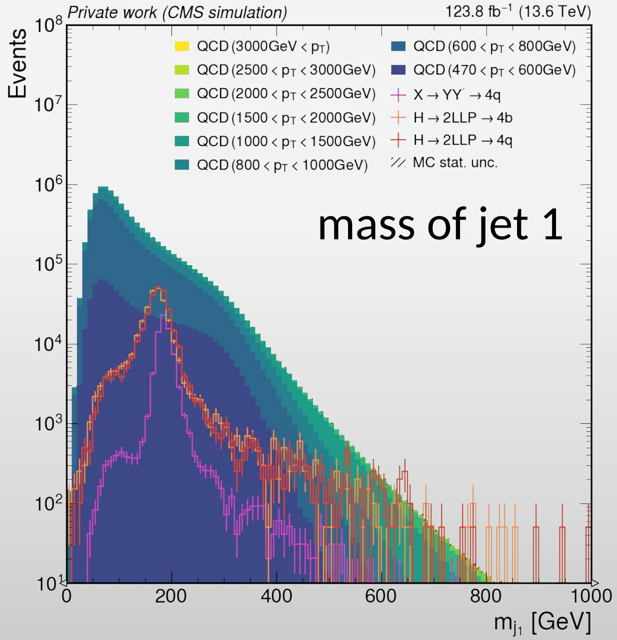
Basic Event Selection Criterion

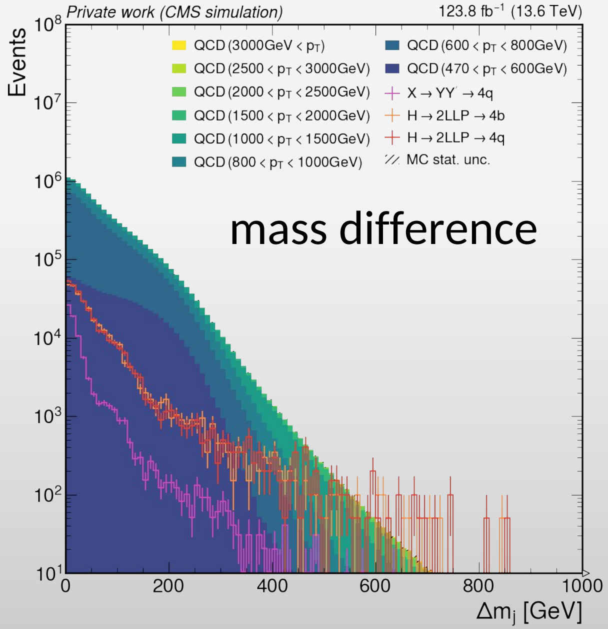
- Same selection as CASE analysis
- Two Anti-k_T fat jets with highest p_T
- Only jets with p_T > 300GeV and $|\eta|$ < 2.5
- $|\eta_{j1}$-$\eta_{j2}|$ < 1.3 and m_{jj} > 1455GeV
- One of the trigger HLT_PFHT1050 or HLT_AK8PFJet500
- Signal region is 2.7TeV < m_{jj} < 3.3TeV

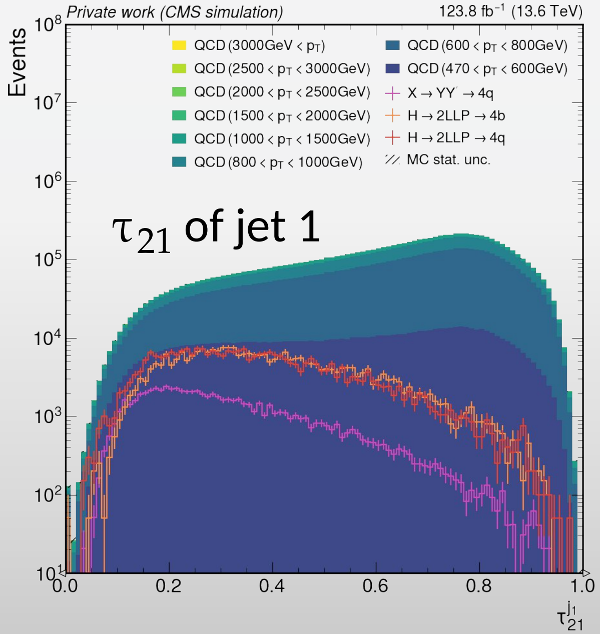
Baseline Features

- Four features already used in previous studies
- Can be well predicted by the flow
- No special sensitivity for LLPs





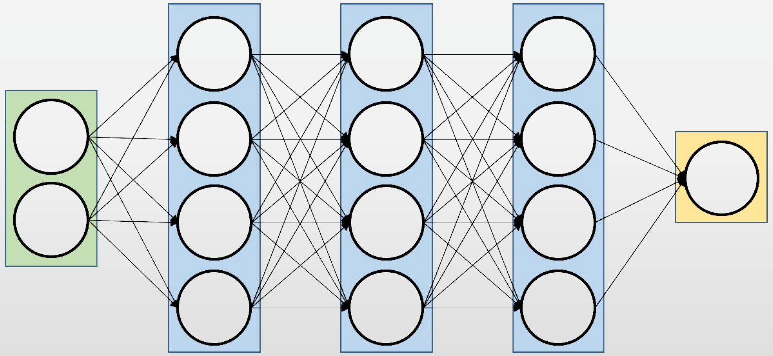




Classifier Training

- Simple DNN model
- Used 1% signal in SR
- A total of 145k bkg events, corresponding to a luminosity of $124\mathrm{fb}^{-1}$
- Train, validation, and test sets are split 60 % / 20 % / 20 %
- Trained for 100 epochs
- Average over best 10 epochs

set	Class 0	Class 1
train	43.5k inter. bkg	43.5k bkg + 870 sig
val	14.5k inter. bkg	14.5k bkg + 290 sig
test	29k bkg	290 sig

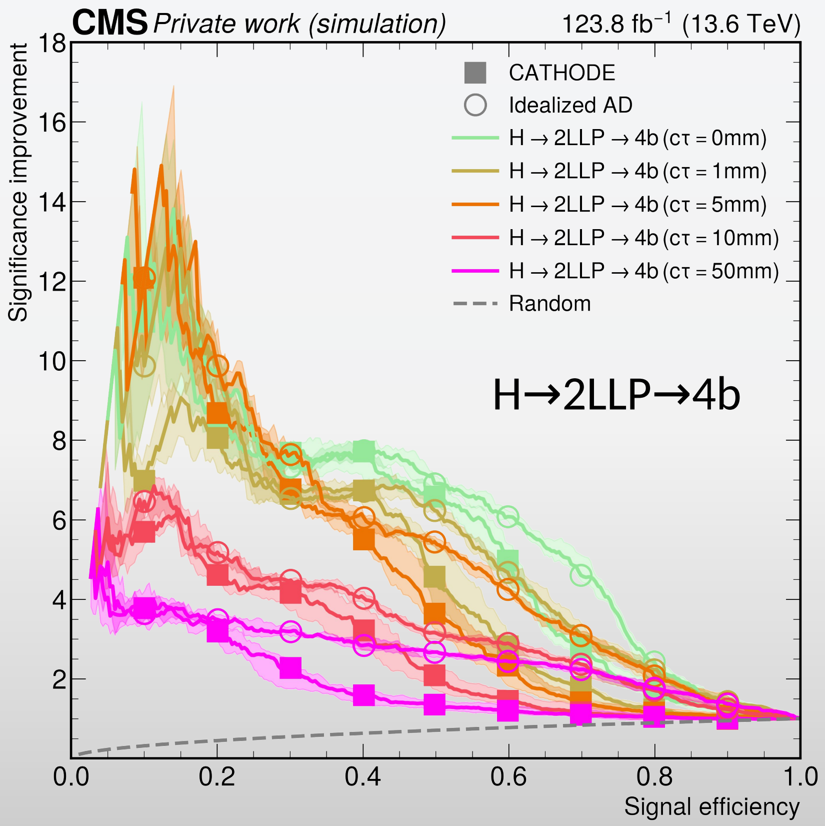


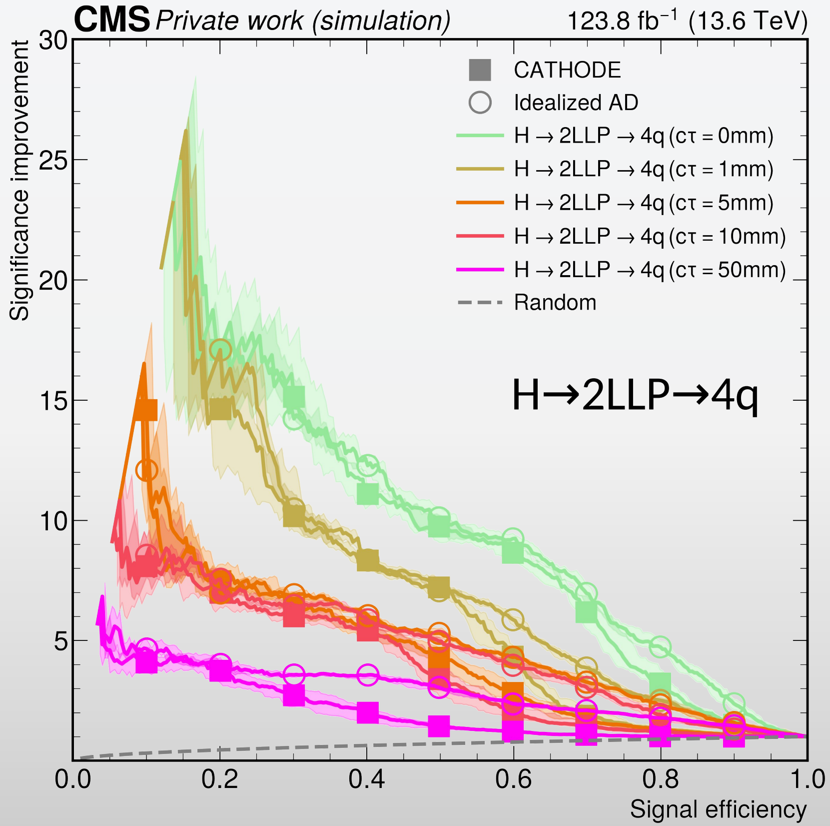
64 nodes per layer

Baseline Trainings

$$\mathrm{SIC} = \frac{\varepsilon_S}{\sqrt{\varepsilon_B}}$$

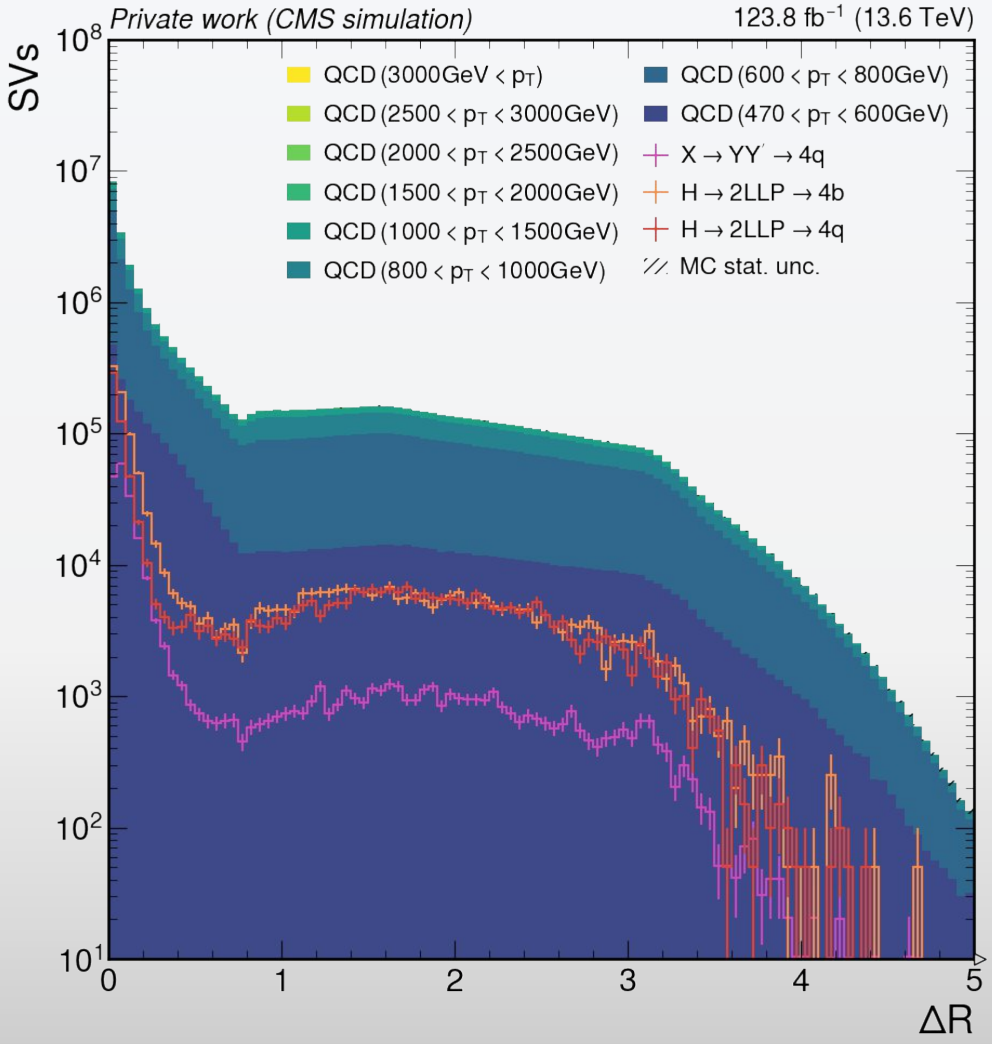
- Dropdown due to bad jet reconstruction





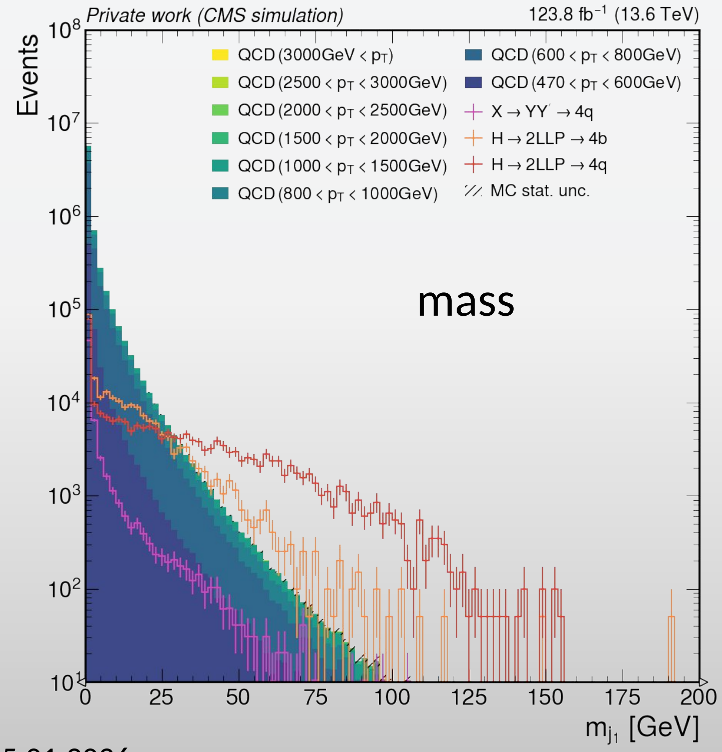
Secondary Vertex Criterion

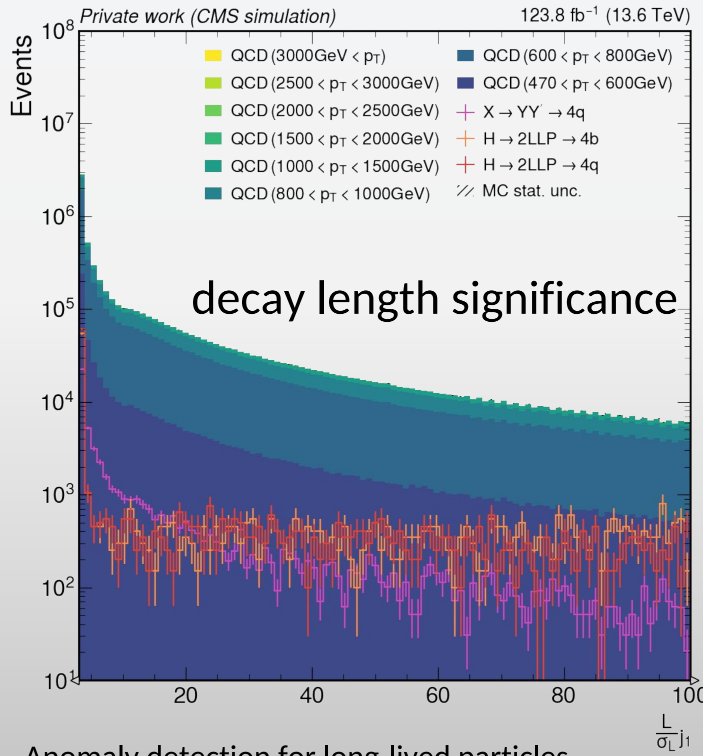
- Use one secondary vertex (SV) per jet
- Require the distance ΔR(SV, jet) < 0.8
- Select the SV with the highest L_{xy}/σ_L

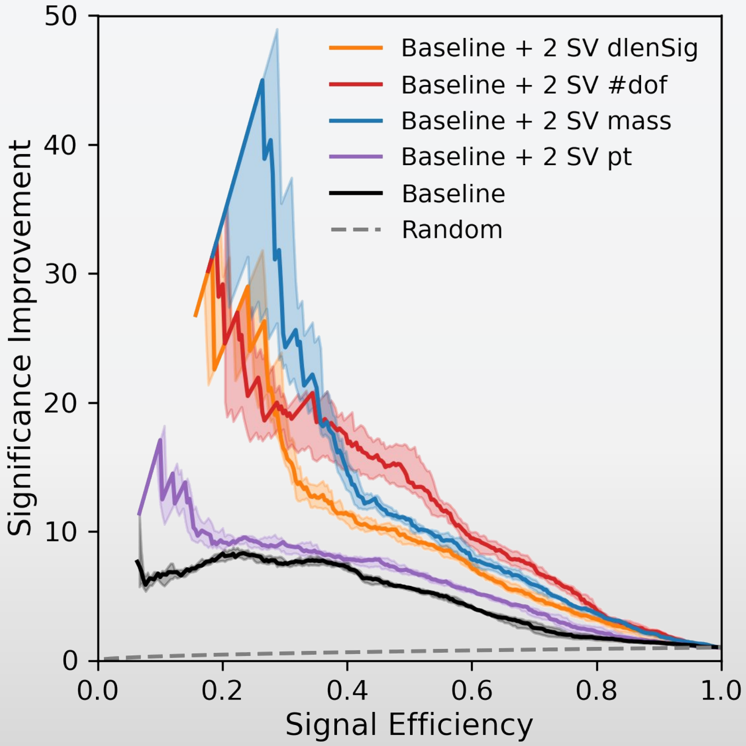


Selected SV Features

- Identified useful SV features by adding the SV features to baseline features
- Tested all SV features available in nanoAOD.
- Best combination with mass and decay length significance





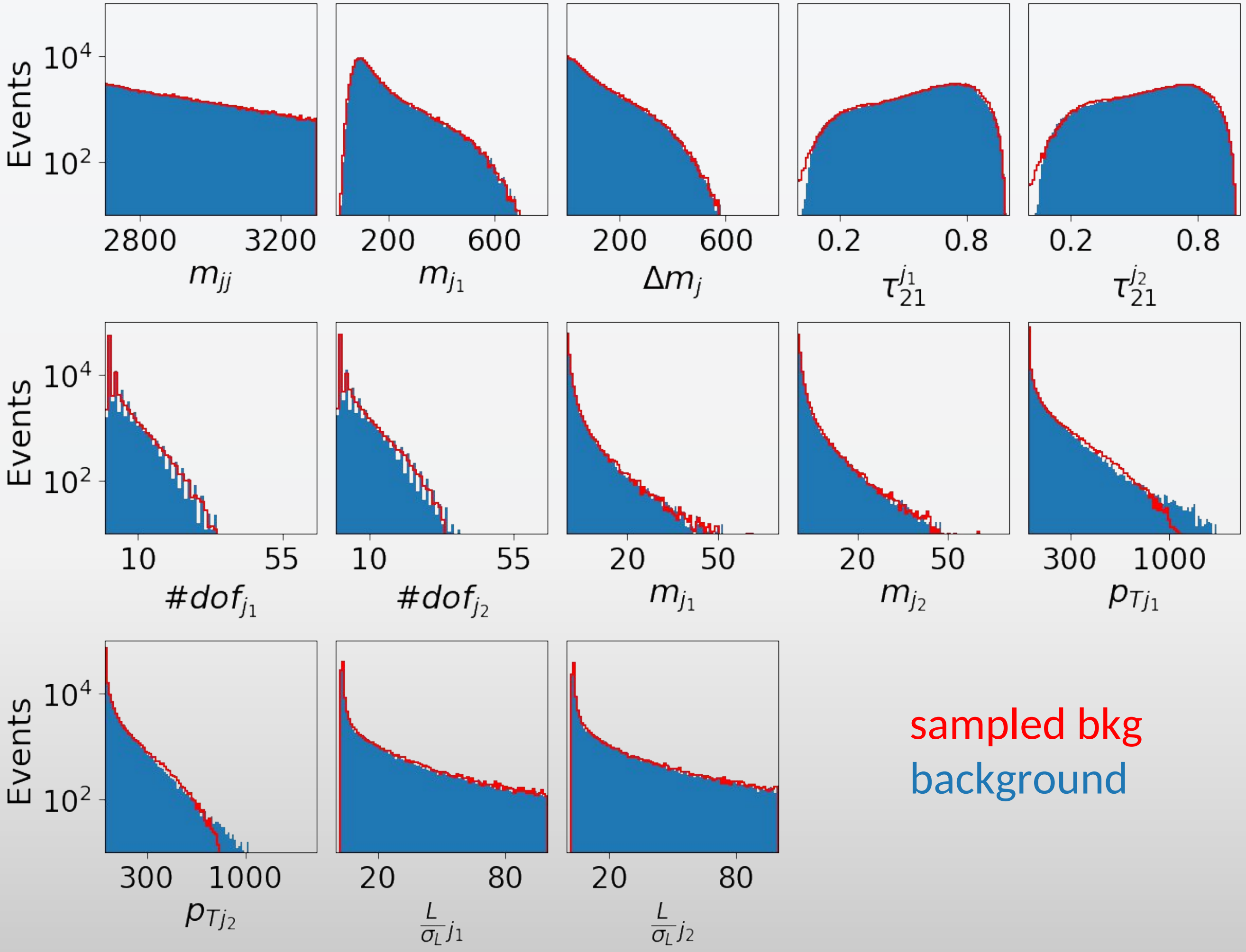


H→2LLP→4q

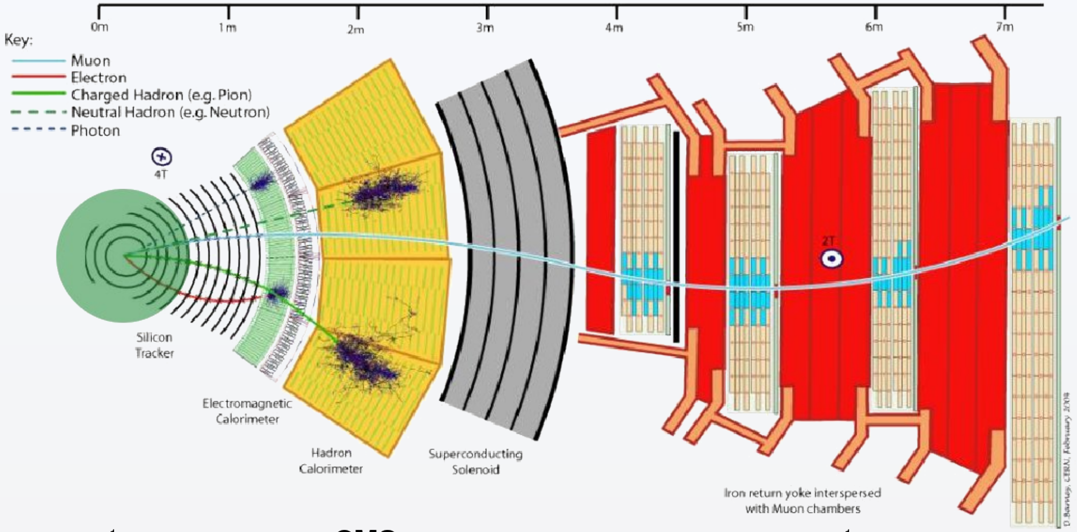
Background Interpolation

- Trained with 300k bkg and 100k for validation
- Sampled bkg equally from best 10 epochs
- Flow matching trained for 30k epochs
- Normalized Flow trained for 100 epochs

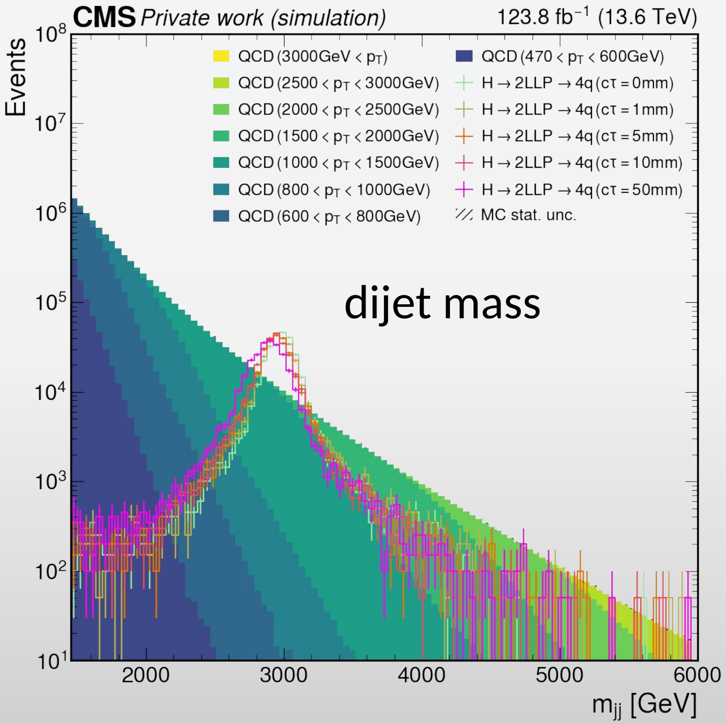
Flow matching with SV features

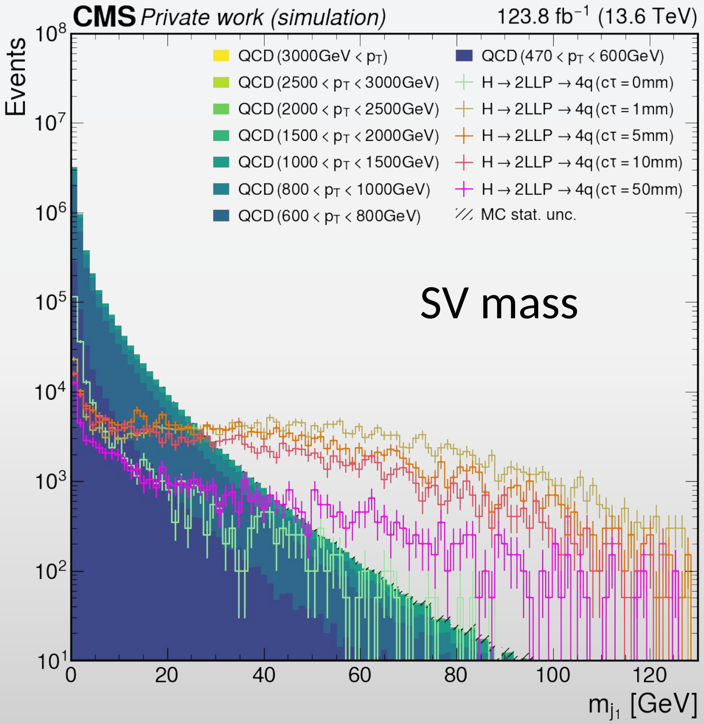


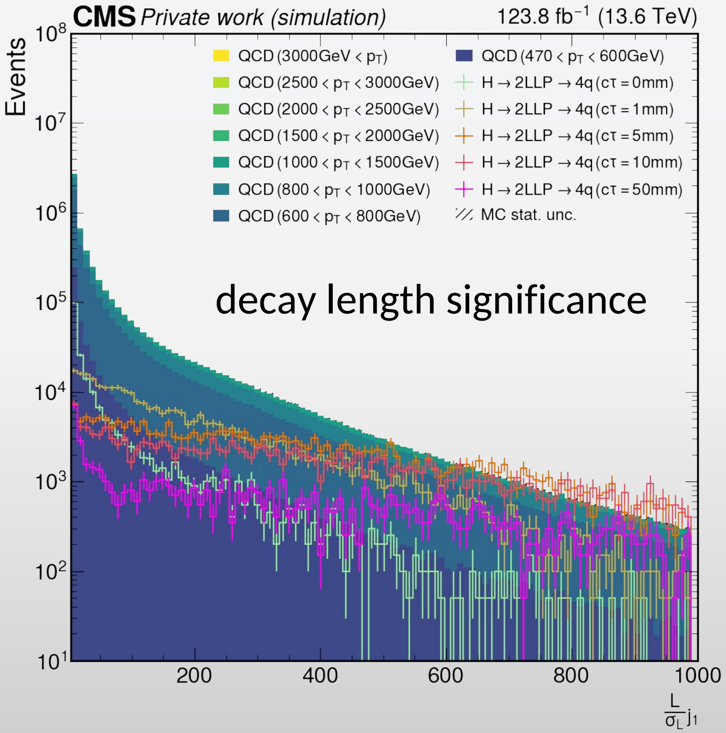
Different Lifetime



- Decay inside the tracker
- Changes in the lifetime lead to shifts across all features

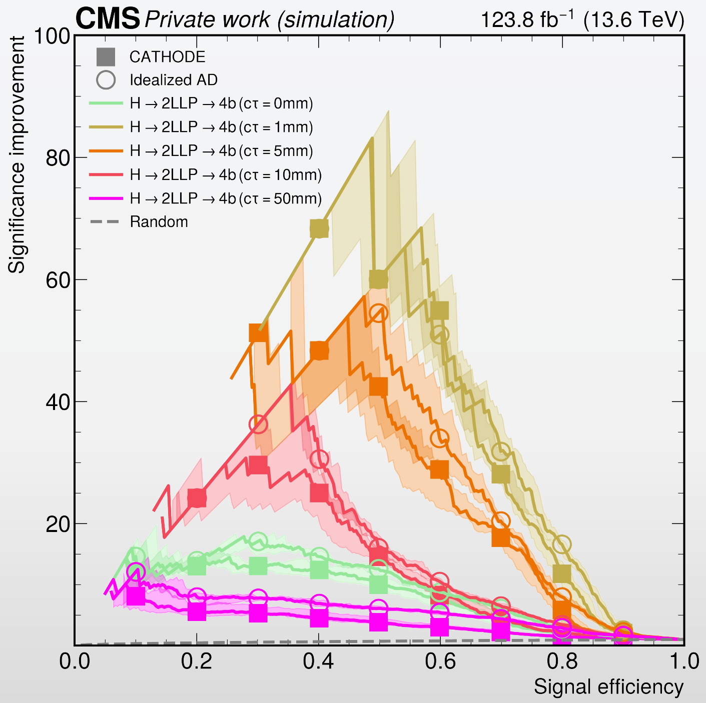




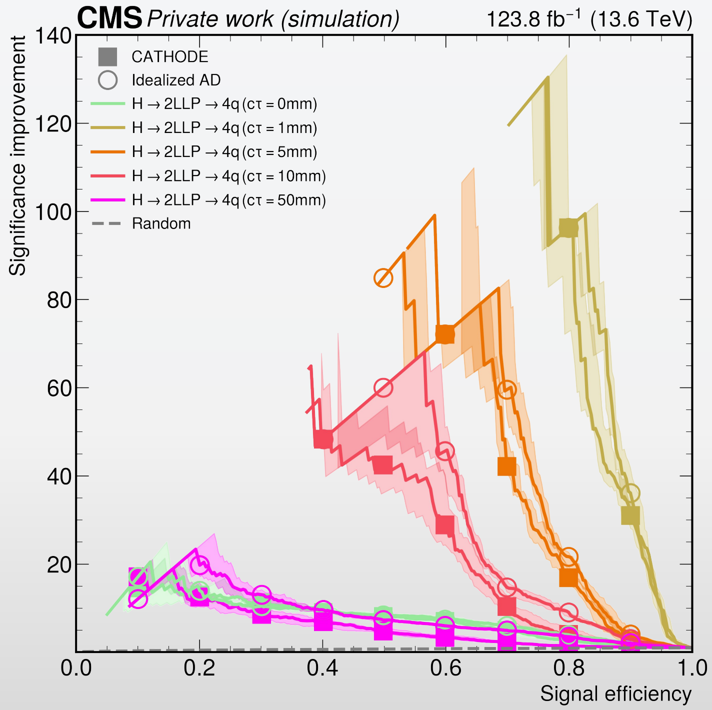


Trainings with SV Features

- Huge improvement for signals with medium cτ



H→2LLP→4b



H→2LLP→4q

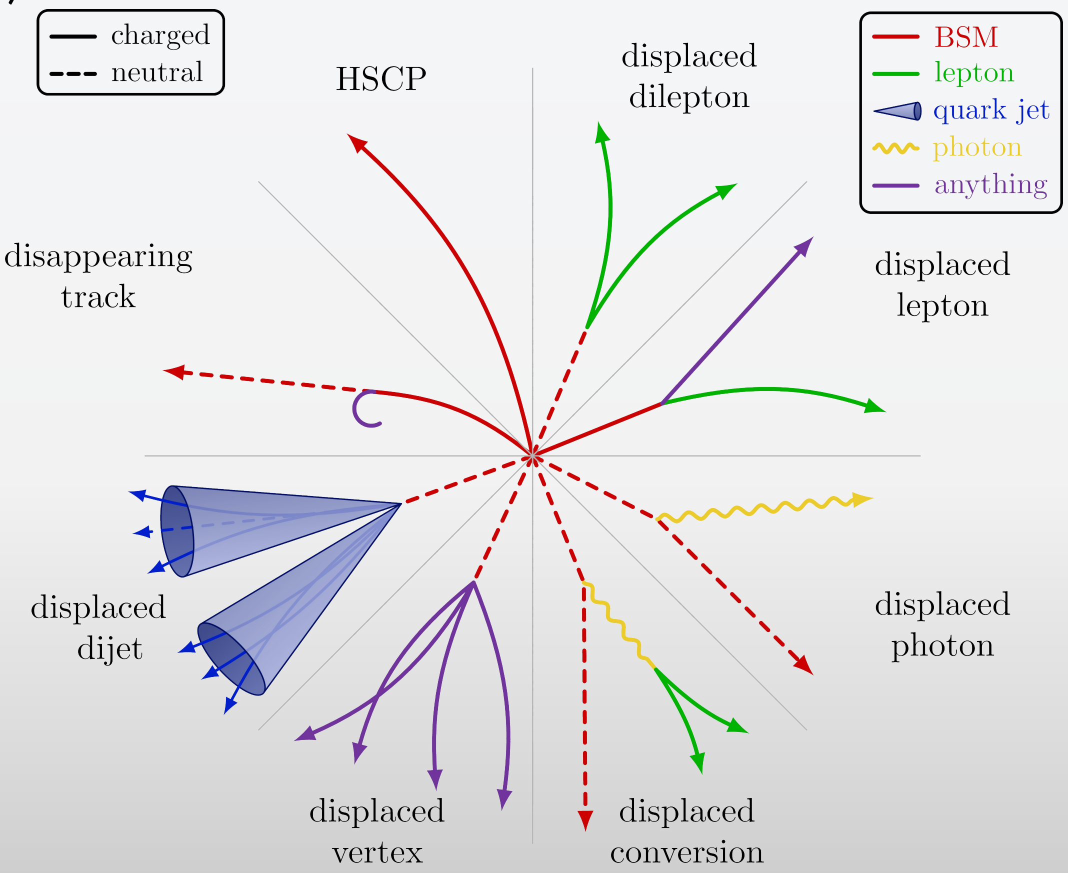
Conclusion

- Incorporating SV features enhances LLP signal identification
- Best performance achieved with SV mass and decay length sig
- Future plans:
  - Add minor backgrounds
  - Testing signals with different properties
  - Implement the full analysis workflow
  - Start with real data by adding to the histograms

Backup

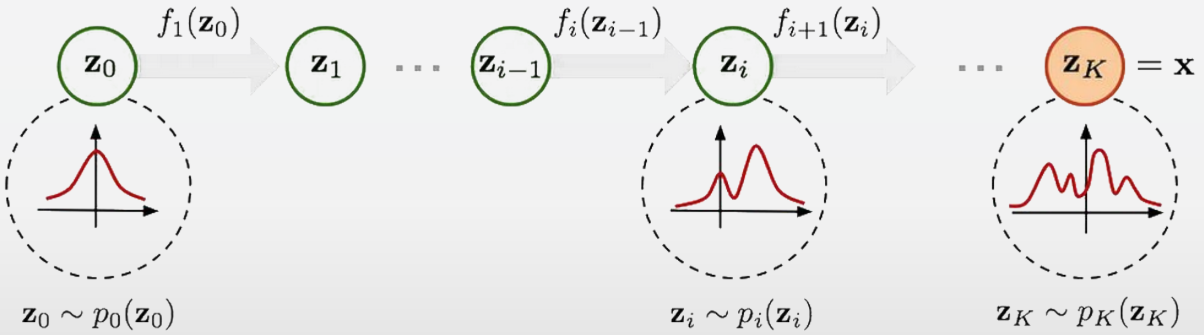
Long-lived Particles (LLPs)

- LLPs appear in many BSM theories (SUSY, dark matter…)
- Particles with measurable lifetimes before decay
- Can travel significant distances inside the detector
- Produce displaced vertices
- Secondary vertices provide good features for LLPs

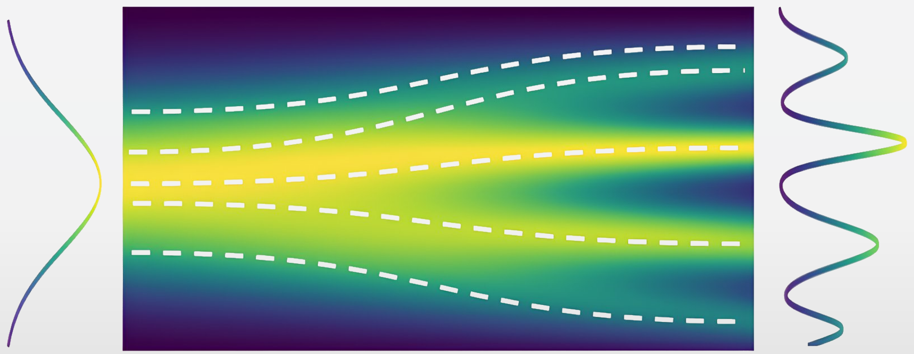


Generative Models

- Normalizing Flow uses arbitrarily complex bijective transformations f_θ
- Flow matching uses continuous vector field v_θ that transforms points along a trajectory
- Conditional flow models features conditioned on given variables (e.g. $\mathrm{m_{jj}}$), allowing generation
- The model learn the weights of f_θ or v_θ



Normalizing flow

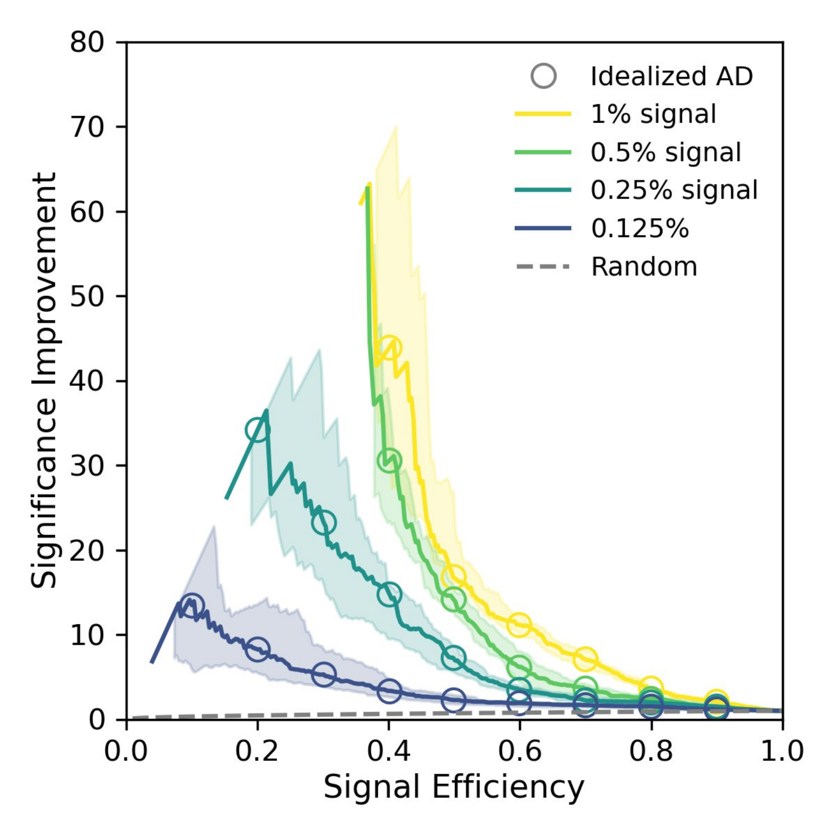


Flow matching

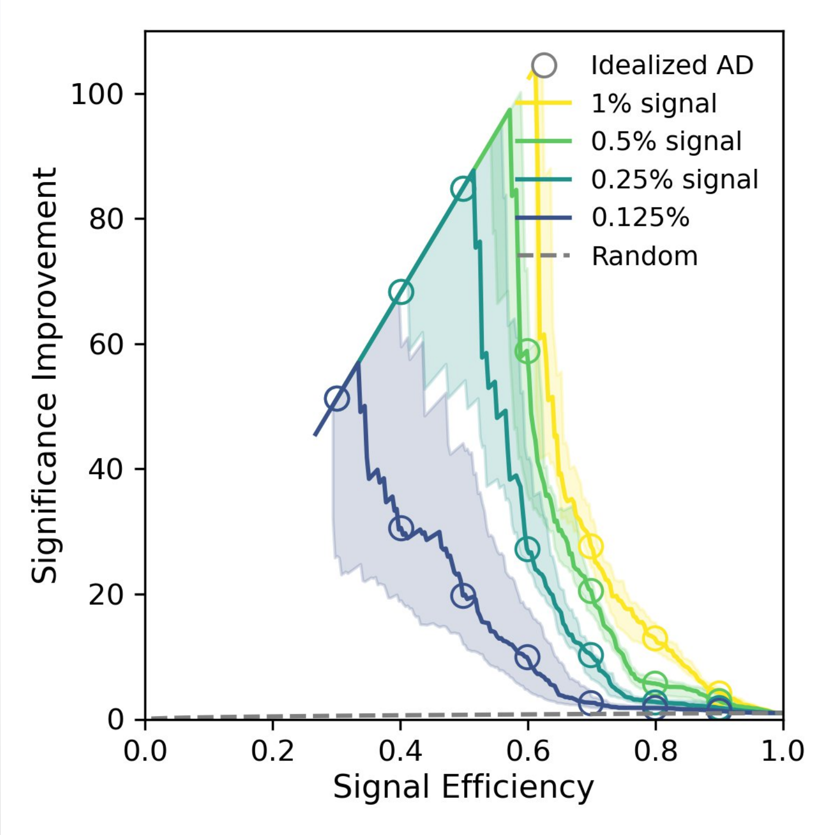
NanoAOD Features

SV_charge	sum of the charge of the SV tracks
SV_chi2	reduced chi2, i.e. chi/ndof
SV_dlen	decay length in cm
SV_dlenSig	decay length significance
SV_dxy	2D decay length in cm
SV_dxySig	2D decay length significance
SV_eta	eta
SV_mass	mass
SV_ndof	number of degrees of freedom
SV_ntracks	number of tracks
SV_pAngle	pointing angle, acos(p_SV*(SV-PV))
SV_phi	phi
SV_pt	pt
SV_x	X position, in cm
SV_y	Y position, in cm
SV_z	Z position, in cm
nSV	secondary vertices from IVF

Different Number of Signals



2LLP->4b



2LLP->4q