北京大學
PEKING UNIVERSITY

"Hadron-in-fat-jet" AI Tagging to Detect Rare Decays

ML4Jets 2026

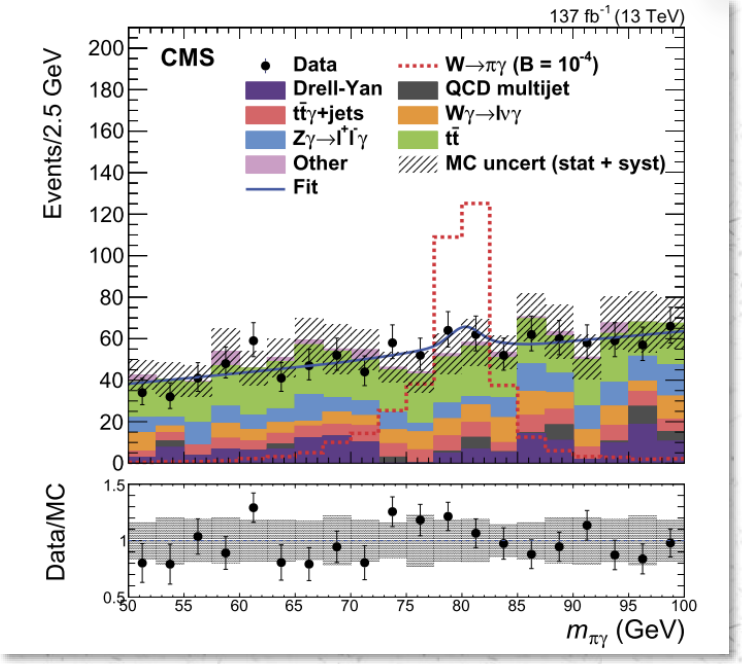
Linrui Chen, Tianyi Yang, Zixun Kou, Qiang Li

Sep 14, 2026

L. Chen et al., Phys. Rev. D, DOI: 10.1103/z3j5-12p5

1 Introduction

- **Rare Hadronic Decays of the W Boson**
  - Theoretical significance: A unique laboratory to test the SM at the interface of perturbative and non-perturbative QCD.
  - QCD factorization: Serves as a gold-standard "test bench" for meson form factors at large energy scales.
  - Extreme rarity: The SM predicted BF for $W^{\pm} \to \pi^{\pm}\gamma$ is as low as $O(10^{-9})$.
- **Experimental Searches at the LHC**
  - Explored by ATLAS, with current the most stringent upper limits constrained at $O(10^{-6})$.
  - CMS reported its latest result of the upper limit in 2021 as 1.50×10^{-5} with 137 fb^{-1} proton-proton collisions data at $\sqrt{s} = 13$ TeV with lepton trigger in $t\bar{t}$ events.
- **Machine Learning Gap**
  - Advanced AI taggers are rarely applied to exclusive/pure hadronic rare decays.



[arXiv:2011.06028v2]

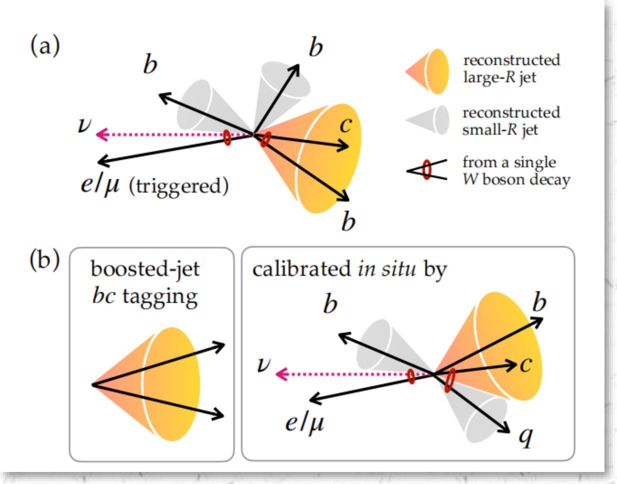
This work: Phys. Rev. D, DOI: 10.1103/z3j5-12p5

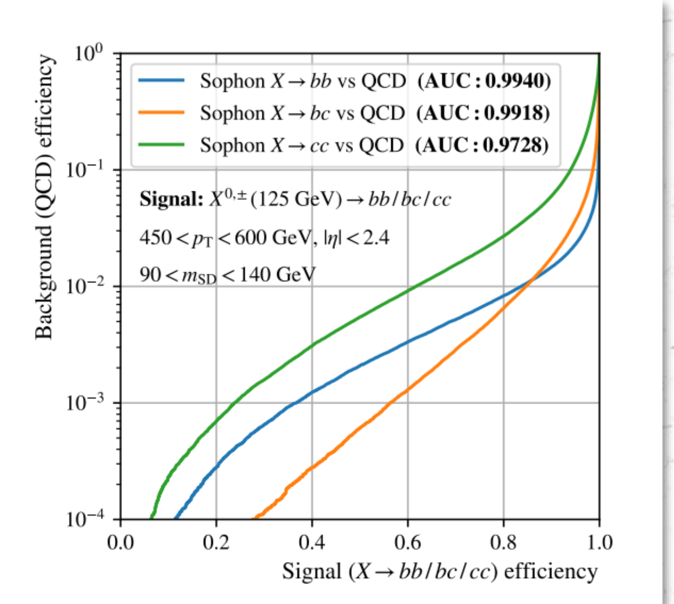
2

Introduction of Sophon: a Jet-tagging Model

[arXiv:2405.12972v1]

- **Overview of the Sophon Model**
  - A next-generation, deep-learning-based jet tagger designed for optimal performance across the vast majority of possible jet final-states.
  - Built upon the Particle Transformer (ParT) architecture and pre-trained on the comprehensive JetClass-II dataset [Hugging face].
  - Output a high-dimensional discriminant (188 categories), encoding both perturbative and non-perturbative jet substructures.
- **Proven Tagging Capabilities**
  - Validated to perform comparably to or even exceed top-tier CMS taggers (e.g. ParticleNet-MD and GloParT) in simulation.
  - Achieve excellent background rejection and high AUC in established channels like *bb* and *cc* tagging.
- **Extension to Unexplored Region**
  - Successfully covers exotic configurations (e.g., novel *bc*-tagging) without requiring proxy phase spaces for calibration.
  - Its generalized representation makes it an ideal backbone for identifying rare "hadron-in-fat-jet" topologies (like $W^\pm \to \pi^\pm\gamma$).





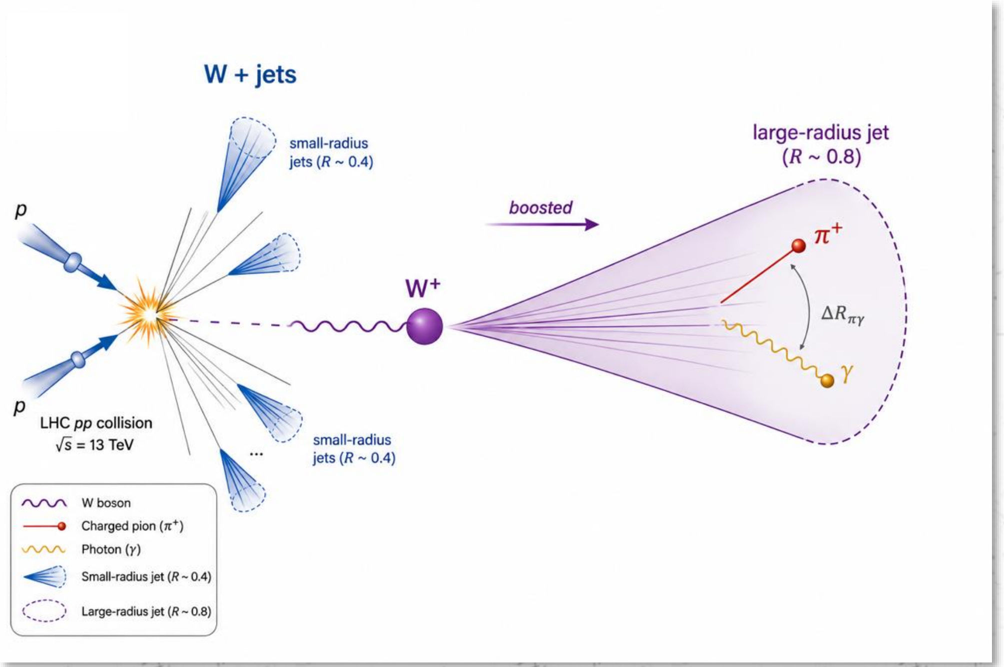
[arXiv:2503.00118v1]

3 Novel W→πγ Region: "Hadron-in-fat-jet"

- "Hadron-in-jet" ⟶ $\Delta R_{\pi\gamma} < 0.8$ ⟶ Boosted W region
- Choose W+jets process since its significantly large cross-section at the LHC than $t\bar{t}$ process.
- A pion and a photon will be closer in a boosted W boson decay final-state due to Lorentz boost which provides a clean environment to apply novel "hadron-in-jet" tagging method with Sophon.
- **Fine-tune**: by treating W→πγ process as a single fat-jet (R=0.8).
  - Jet-matching with their inside substructure.
  - Expand the category of signal and confusing background (e.g., $\tau_h\nu$ final state).
  - Fine-tune Sophon model in two steps:

    (i). Frozen backbone with only classification head could be updated. (Large LR)

    (ii). Unfrozen the entire network. (Relatively low LR)



4 Experimental Set-up: MC Samples

北京大學 PEKING UNIVERSITY

- Generator: MadGraph5_aMC@NLO v2.9.18
- Parton-shower and hadronization: Pythia8
- Fast detector simulation: Delphes 3.5.1
- Requiring $H_T > 700$ GeV at generator-level and using $H_T > 1050$ GeV trigger at reconstruction-level for all MC samples which is consistent with CMS trigger.
- **Fat-jet reconstruction:**
  - Apply Pileup Per Particle Identification (PUPPI) algorithm to mitigate pileup contamination.
  - The PUPPI-weighted energy-flow objects are then clustered into large-radius (R=0.8) fat-jets using the anti-k_T algorithm with $p_T >$ 200 GeV and $|\eta| < 2.5$.

Table 1: Summary of MC samples

Background Processes	$N_{\text{gen}}(\times 10^4)$	Cross-Section (mb)
Signal ($W^\pm \to \pi^\pm \gamma$)	990	1.632×10^{-8}
QCD	99025	9.047×10^{-6}
$W^\pm \to \tau^\pm \nu$	200	1.629×10^{-8}
$p\,p \to t\bar{t}$	2010	1.523×10^{-8}
$p\,p \to W\gamma$ (inclusive: ISR & FSR)	10	2.587×10^{-11}
$p\,p \to W + \text{jets},\ W \to l\nu$	400	4.576×10^{-9}
$p\,p \to W + \text{jets},\ W \to q\bar{q}$	3010	3.461×10^{-8}
$p\,p \to Z + \text{jets},\ Z \to l^+l^-$	50	4.922×10^{-10}
$p\,p \to Z + \text{jets},\ Z \to \nu\bar{\nu}$	100	1.189×10^{-9}
$p\,p \to Z + \text{jets},\ Z \to q\bar{q}$	1000	1.407×10^{-8}

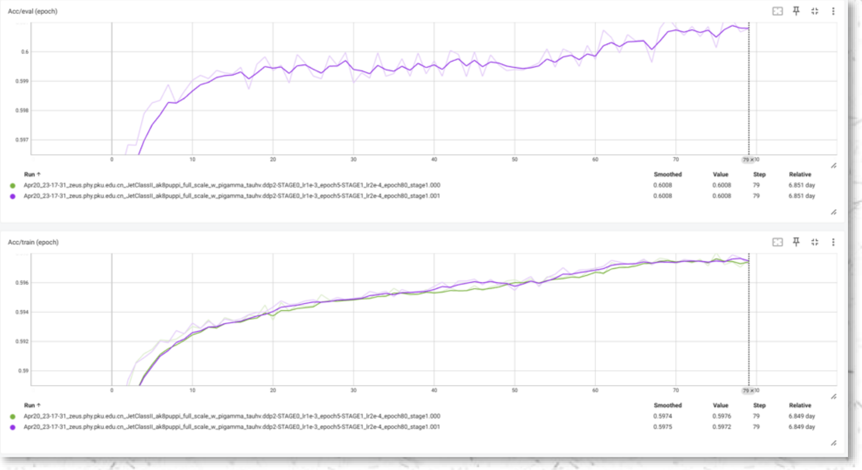
5 Sophon Model Fine-tune

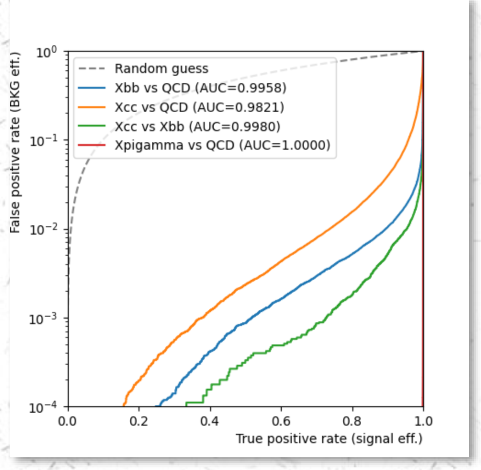
北京大學 PEKING UNIVERSITY

- **Training sample:**
  - Generated from the decays of a hypothetical spin-0 resonance, $X^{0,\pm}$, with its mass mX uniformly sampled between 15 and 500 GeV.
  - Three types of training samples:
    - $X \to \pi\gamma$ (2 prong): ~24 million
    - $X \to$ 3 or 4 prong: ~20 million
    - $X \to \tau_h \nu$ (2 prong): ~50 million

Major types	Index range	Label names
$X \to$ 2 prong	188	$\pi\gamma$
$X \to$ 3 or 4 prong	189 – 204	$\pi\gamma qq,\ \pi\gamma qs,\ \pi\gamma bq,\ \pi\gamma cq,\ \pi\gamma cs,\ \pi\gamma bc,\ \pi\gamma b,\ \pi\gamma c,\ \pi\gamma s,\ \pi\gamma q,\ \pi\gamma e\nu,\ \pi\gamma\mu\nu,\ \pi\gamma\tau_e\nu,\ \pi\gamma\tau_\mu\nu,\ \pi\gamma\tau_h\nu, \pi\gamma\pi\gamma$
$X \to$ 2 prong	205	$\tau_h\nu$

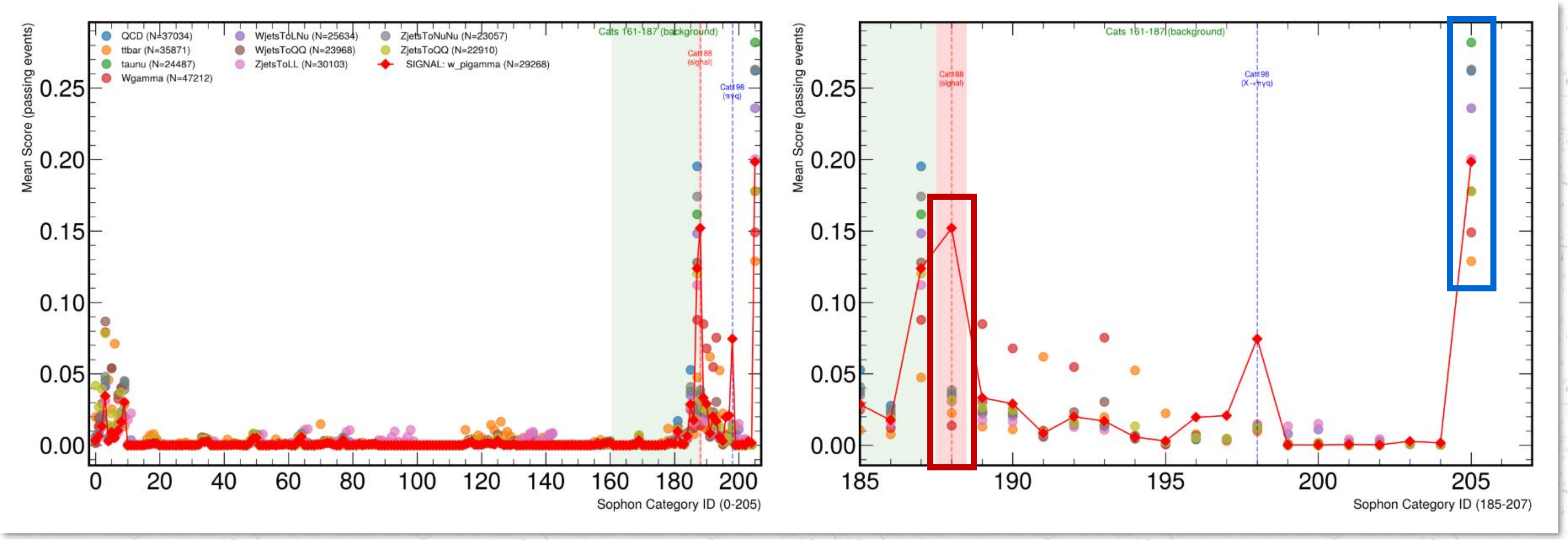
- New Labels: Sophon original 188 categories do not contain any hadronic final-state. 18 NEW categories were added to Sophon model.
- Two steps of fine-tune:
  - Step1: 5 epochs, LR: 1e-3, Freeze the whole network except for classification head.
  - Step2: 80 epochs, LR: 2e-4 (gradually decayed after 70% epochs), unfreeze the entire network.





5 Sophon Model Fine-tune

- Each fat-jet passing the preselection is assigned 206 Sophon model scores, corresponding to the 206 categories, with values ranging from 0 to 1.
- Signal: closer to 1.
- Background: closer to 0.



6 Analysis Strategies

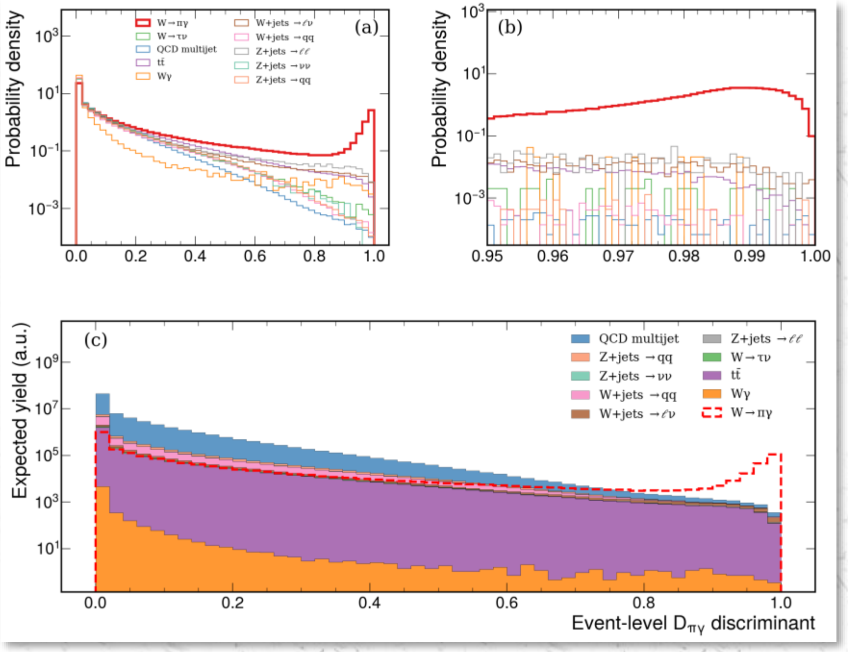
PEKING UNIVERSITY

- To better distinguish signal and background, we define a Discriminant as follows:

$$\boxed{D_{\pi\gamma} = \frac{g_{X\to\pi\gamma}}{g_{X\to\pi\gamma} + g_{X\to\pi\gamma q} + g_{X\to\tau_h\nu} + g_{\mathrm{QCD}} + \sum_{P_1} g_{P_1}}}$$

First Priority labels P_1			
$X \to$2 prong	$X \to ee$	$X \to \mu\mu$	
$X \to$ 3 or 4 prong	$X \to \pi\gamma qq$	$X \to \pi\gamma qs$	$X \to \pi\gamma cq$
	$X \to \pi\gamma cs$	$X \to qqee$	$X \to qq\mu\mu$
	$X \to see$	$X \to qee$	$X \to gee$
	$X \to q\mu\mu$	$X \to bqe\nu$	$X \to sqe\nu$
	$X \to qqe\nu$	$X \to bq\mu\nu$	$X \to sq\mu\nu$
	$X \to qq\mu\nu$		

- To maximize the separation between W→πγ signal and various background categories

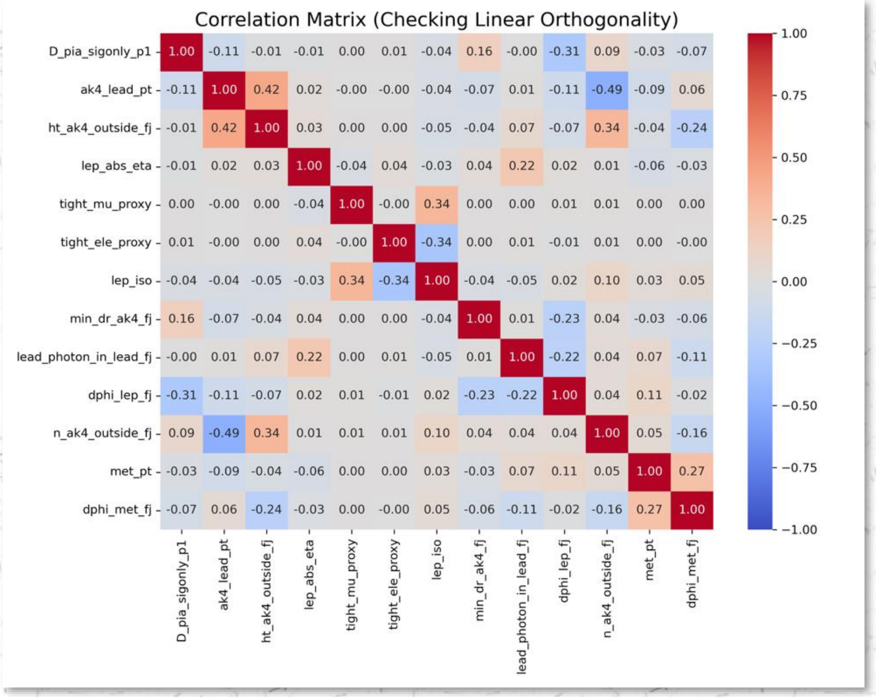


- Event-level selection: Fat-jet with highest p_T within each event.
- Signal heavily concentrates at $D_{\pi\gamma} \approx 1$, while inclusive SM backgrounds are robustly pushed down.

6 Analysis Strategies

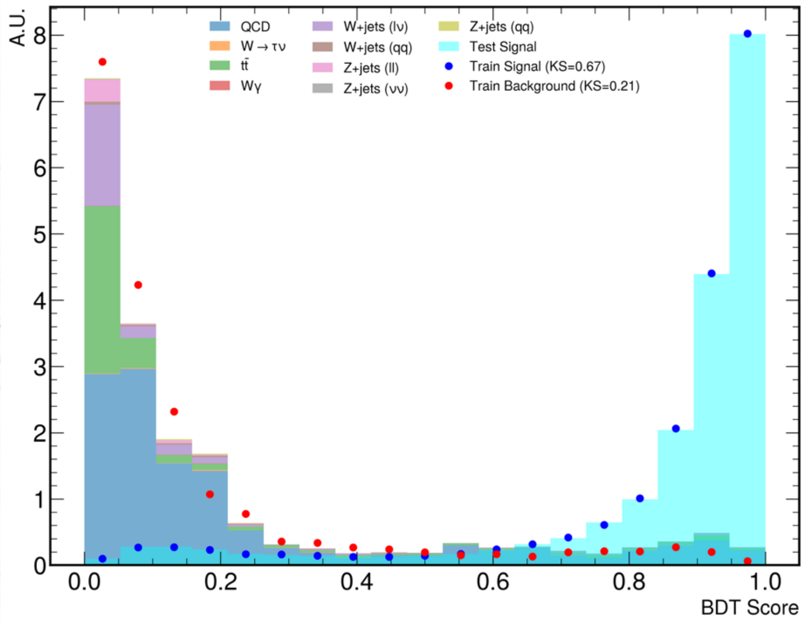
PEKING UNIVERSITY 北京大學

- **BDT TRAINING:**
- Exploit global event kinematics to enhance signal-to-background separation beyond the fat-jet level.
- Framework: XGBoost classifier (Sig vs. Inclusive Bkg)
- Input Variables:
  - Uses 12 macroscopic features, including missing transverse energy, ak4-jet kinematics, lepton proxies and angular correlations.
- **Orthogonality**
  - All BDT input variables must be explicitly orthogonal to the jet-level Sophon discriminant ($D_{\pi\gamma}$).
- Event Weighting Strategy
  - Physical weight: Scaled to physical cross-section with an expected luminosity of 450 fb^{-1}. Testing phase
  - Training weights: Re-weighted so that the sum of weights for signal and background both equal 1.0. Training phase



6 Analysis Strategies

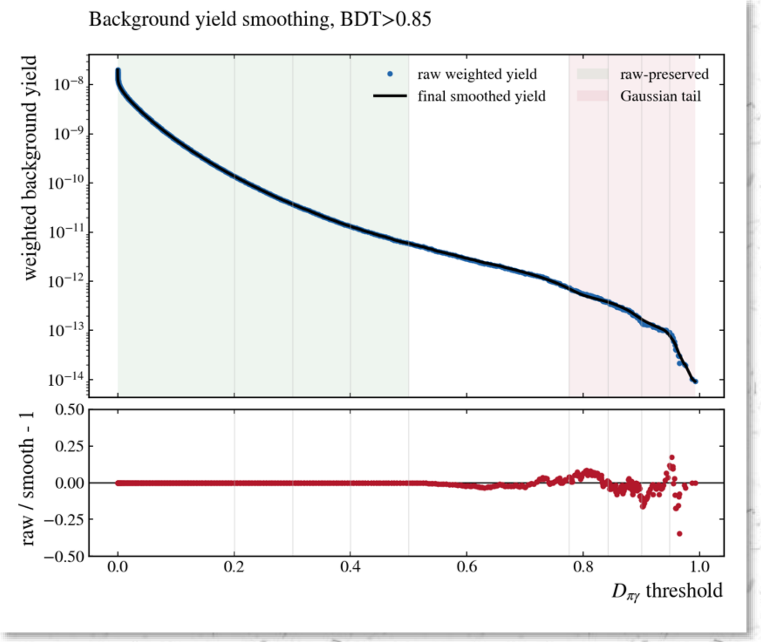
- Robustness Against Overtraining
  - Kolmogorov-Smirnov (KS) Test
    - Compares the BDT score distributions between training and testing datasets.
    - Excellent agreement observed for both sig and bkg processes.
- Overall Performance
  - Achieves strong background rejection while maintaining high signal efficiency.
  - The combined strategy (jet-level $D_{\pi\gamma}$ + event-level BDT) securely isolates the rare $W^{\pm} \rightarrow \pi^{\pm}\gamma$ topology.

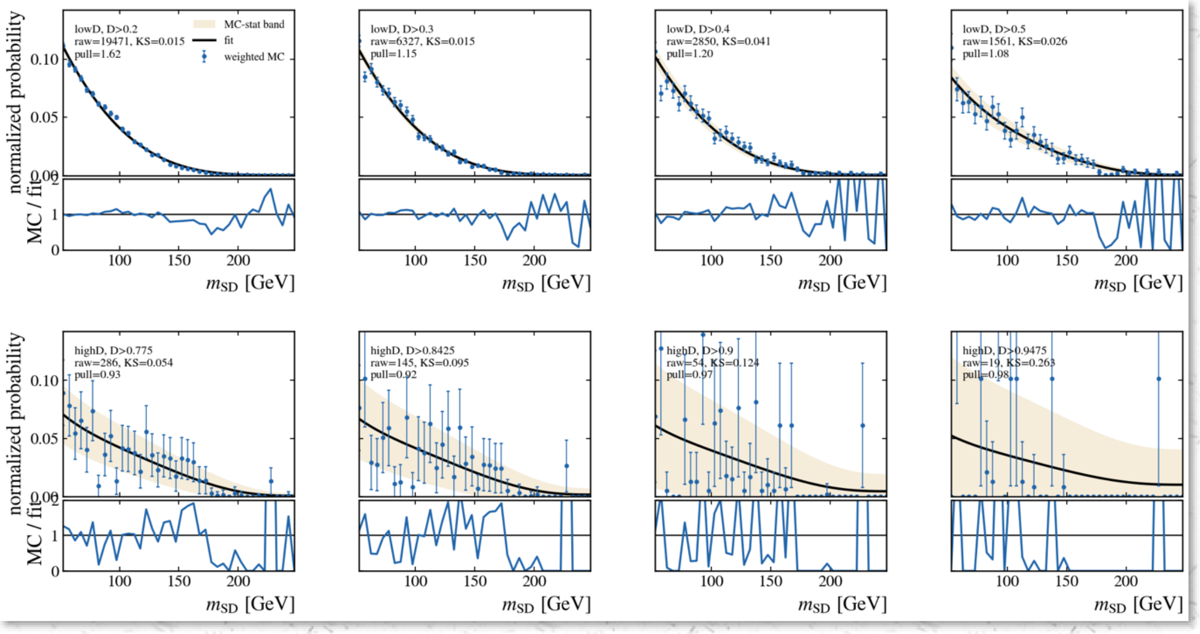


6 Analysis Strategies

- Background Yield Smearing $x = -\ln(1 - D_{\pi\gamma})$
  - Keep raw yield for $D_{\pi\gamma} < 0.5$ to preserve high-statistics region.
  - Transition to fully smoothed tail for $D_{\pi\gamma} > 0.775$ using a cubic smoothstep.
- Enforce monotonic projection to guarantee physically consistent background suppression.

BDT > 0.85

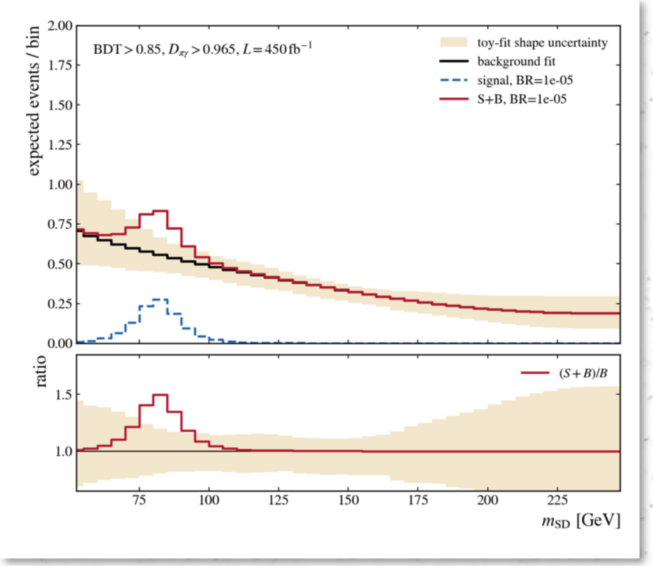


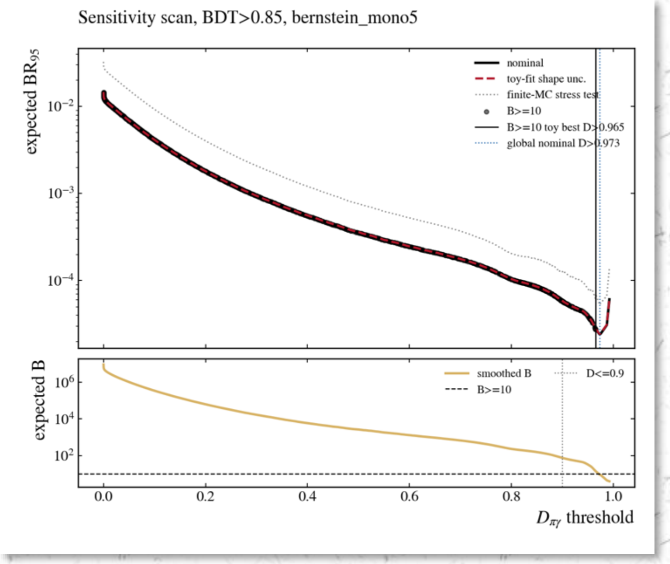


- Mass shape Modeling (m_{SD})
  - Model the background m_{SD} shape using a 5-degree monotone Bernstein polynomial.
  - Coefficients are strictly constrained via Sigmoid to prevent unphysical fake peaks.
  - Shape parameters $\theta_j(D_{\pi\gamma})$ are modeled as quadratic functions, globally fitted to selected MC anchors

7 Results

BDT > 0.85





- **Shape Uncertainty Estimation**
  - Generate 80 Poisson pseudo-data toys from the nominal background template.
  - Refit each toy with the mono5 model to extract a 1σ (16% to 84% quantiles) shape uncertainty band (σ_{toy}).
- **Sensitivity & Optimization**
  - Minimize the expected 95% CL branching ratio upper limit.
  - Incorporate shape systematic uncertainty into the significance calculation:

$$\boxed{Z_{\text{toy}} = \sqrt{\sum_i \frac{S_i^2}{B_i + \sigma_{\text{toy},i}^2}}}$$

  - Constrain the expected background events to $B \geq 10$ for statistical robustness.
- **Optimal cutting results**
  - Recommended optimal cut: $D_{\pi\gamma} > 0.965$.
  - **Yields:** $B = 14.6$, $S = 0.115$ (at benchmark $\mathcal{B}(W^{\pm} \to \pi^{\pm}\gamma) = 1 \times 10^{-6}$).
  - **Expected limit at 95% CL:** $\mathcal{B}(W^{\pm} \to \pi^{\pm}\gamma) < 2.78 \times 10^{-5}$.

8 Summary and Future Prospects

PEKING UNIVERSITY

- Successfully developed a novel "hadron-in-fat-jet" tagger based on Sophon to detect the rare $W^\pm \to \pi^\pm\gamma$ decay in the boosted region.
- Combined jet-level $D_{\pi\gamma}$ with event-level BDT to achieve strong QCD background rejection while maintaining robust signal efficiency.
- Established a rigorous parametric background model (monotonic yield smearing & monotone Bernstein shape) to control systematic uncertainties in high-$D_{\pi\gamma}$ tails.
- Achieved an expected upper limit of $\mathcal{B}(W^\pm \to \pi^\pm\gamma) < 2.78 \times 10^{-5}$ (including toy-fit shape systematics) at the optimal cut of $D_{\pi\gamma} > 0.965$.

- Low background statistics & limited sensitivity $\Longrightarrow$ Photon trigger?

Thanks for your attention!