New potentials in boosted-jet regimes through inclusive jet model pre-training

Yuzhe Zhao, Congqiao Li, Antonios Agapitos, Dawei Fu, Leyun Gao (Speaker), Yajun Mao, Qiang Li

State Key Laboratory of Nuclear Physics and Technology, Peking University

School of Physics, Peking University

AI+HEP in East Asia February 25, 2025

Based on: arXiv:2502.xxxxx and arXiv:2503.xxxxx.

Outline

Outline

How inclusive boosted jet models transform the HEP researches

Experimental impact

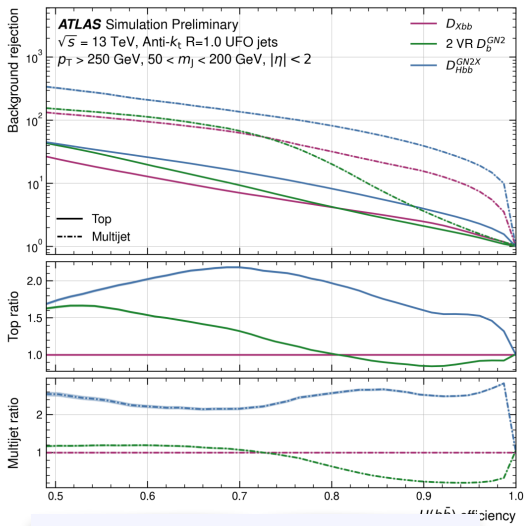
1

➔ Similar cases for large-R jet tagging

(Congqiao's talk yesterday)

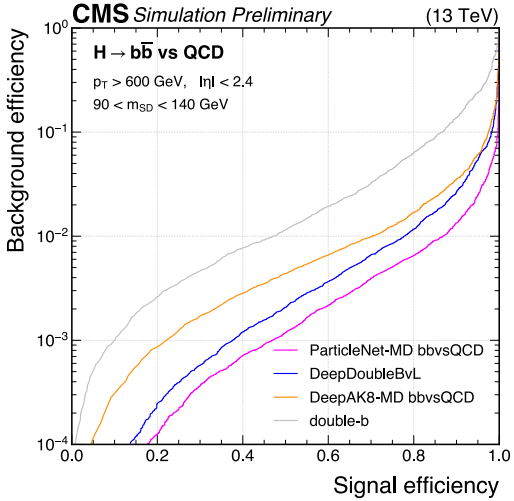
- more complex tasks! (o(30-100) particles within a large cone size)
- believed to have larger benefits from DNN algorithm improvements

ATL-PHYS-PUB-2023-021



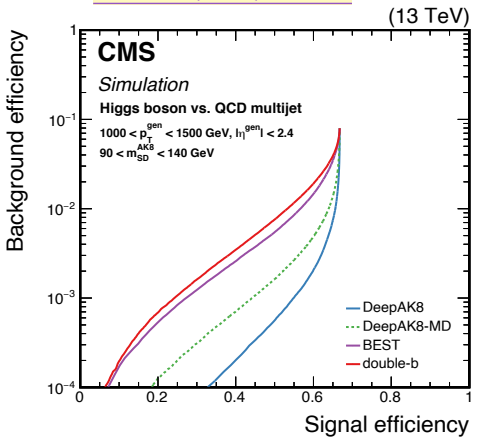
Transformer-based GN2X tagger:
~x3 📈 QCD and x2 top background rejection

CMS-PAS-BTV-22-001



DeepAK8 → ParticleNet:
x5 📈 QCD background rejection

JINST 15 (2020) P06005



Comparing with early approaches
Another **~x5** 📈 improvement achieved

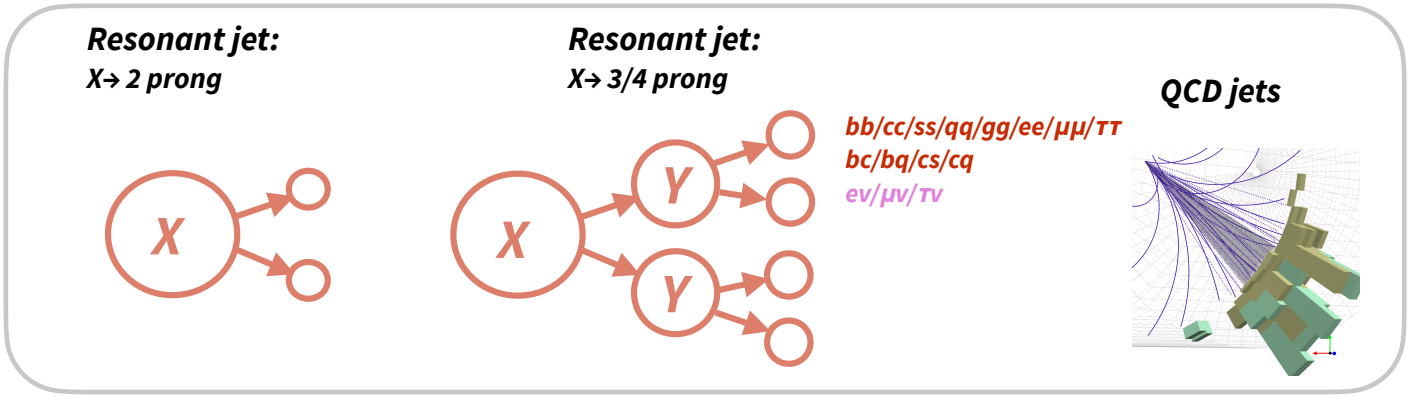
Sophon, a very recent inclusive boosted jet model

Introducing Sophon

arXiv:2405.12972

https://github.com/jet-universe/sophon

- ➔ **Signature-Oriented Pre-training for Heavy-resonant ObservatioN**
- ➔ the model is based on Particle Transformer (**ParT)** architecture
- ➔ a pre-trained model on a newly developed comprehensive dataset: **JetClass-II**
  - ***finely categorized labels:***



contributed final states: ***bb/cc/ss/qq/gg/ee/μμ/ττ bc/bq/cs/cq*** *all combination of Y decays, resulting to 4-prong or 3-prong*

(credit slides)

Key property: we do not focus on any specific *X* and *Y* masses
Their masses are variables: ranges from 20-500 GeV

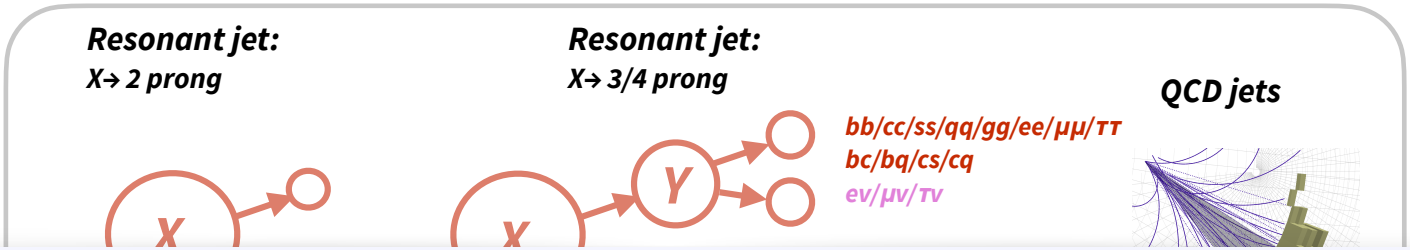
Sophon, a very recent inclusive boosted jet model

Introducing Sophon

arXiv:2405.12972

https://github.com/jet-universe/sophon

- ➔ **Signature-Oriented Pre-training for Heavy-resonant ObservatioN**
- ➔ the model is based on Particle Transformer (**ParT)** architecture
- ➔ a pre-trained model on a newly developed comprehensive dataset: **JetClass-II**
  - ***finely categorized labels:***



All final states!

Major types	Index range	Label names
Resonant jets: $X \to 2$ prong	0–14	bb, cc, ss, qq, bc, cs, bq, cq, sq, gg, ee, $\mu\mu$, $\tau_h\tau_e$, $\tau_h\tau_\mu$, $\tau_h\tau_h$
Resonant jets: $X \to 3$ or 4 prong	15–160	$bbbb$, $bbcc$, $bbss$, $bbqq$, $bbgg$, $bbee$, $bb\mu\mu$, $bb\tau_h\tau_e$, $bb\tau_h\tau_\mu$, $bb\tau_h\tau_h$, bbb, bbc, bbs, bbq, bbg, bbe, $bb\mu$, $cccc$, $ccss$, $ccqq$, $ccgg$, $ccee$, $cc\mu\mu$, $cc\tau_h\tau_e$, $cc\tau_h\tau_\mu$, $cc\tau_h\tau_h$, ccb, ccc, ccs, ccq, ccg, cce, $cc\mu$, $ssss$, $ssqq$, $ssgg$, $ssee$, $ss\mu\mu$, $ss\tau_h\tau_e$, $ss\tau_h\tau_\mu$, $ss\tau_h\tau_h$, ssb, ssc, sss, ssq, ssg, sse, $ss\mu$, $qqqq$, $qqgg$, $qqee$, $qq\mu\mu$, $qq\tau_h\tau_e$, $qq\tau_h\tau_\mu$, $qq\tau_h\tau_h$, qqb, qqc, qqs, qqq, qqg, qqe, $qq\mu$, $gggg$, $ggee$, $gg\mu\mu$, $gg\tau_h\tau_e$, $gg\tau_h\tau_\mu$, $gg\tau_h\tau_h$, ggb, ggc, ggs, ggq, ggg, gge, $gg\mu$, bee, cee, see, qee, gee, $b\mu\mu$, $c\mu\mu$, $s\mu\mu$, $q\mu\mu$, $g\mu\mu$, $b\tau_h\tau_e$, $c\tau_h\tau_e$, $s\tau_h\tau_e$, $q\tau_h\tau_e$, $g\tau_h\tau_e$, $b\tau_h\tau_\mu$, $c\tau_h\tau_\mu$, $s\tau_h\tau_\mu$, $q\tau_h\tau_\mu$, $g\tau_h\tau_\mu$, $b\tau_h\tau_h$, $c\tau_h\tau_h$, $s\tau_h\tau_h$, $q\tau_h\tau_h$, $g\tau_h\tau_h$, $qqqb$, $qqqc$, $qqqs$, $bbcq$, $ccbs$, $ccbq$, $ccsq$, $sscq$, $qqbc$, $qqbs$, $qqcs$, $bcsq$, bcs, bcq, bsq, csq, $bce\nu$, $cse\nu$, $bqe\nu$, $cqe\nu$, $sqe\nu$, $qqe\nu$, $bc\mu\nu$, $cs\mu\nu$, $bq\mu\nu$, $cq\mu\nu$, $sq\mu\nu$, $qq\mu\nu$, $bc\tau_e\nu$, $cs\tau_e\nu$, $bq\tau_e\nu$, $cq\tau_e\nu$, $sq\tau_e\nu$, $qq\tau_e\nu$, $bc\tau_\mu\nu$, $cs\tau_\mu\nu$, $bq\tau_\mu\nu$, $cq\tau_\mu\nu$, $sq\tau_\mu\nu$, $qq\tau_\mu\nu$, $bc\tau_h\nu$, $cs\tau_h\nu$, $bq\tau_h\nu$, $cq\tau_h\nu$, $sq\tau_h\nu$, $qq\tau_h\nu$
QCD jets	161–187	$bbccss$, $bbccs$, $bbcc$, $bbcss$, $bbcs$, bbc, $bbss$, bbs, bb, $bccss$, $bccs$, bcc, $bcss$, bcs, bc, bss, bs, b, $ccss$, ccs, cc, css, cs, c, ss, s, others

The CKM matrix and current measurement results

- The Cabibbo–Kobayashi–Maskawa (CKM) matrix quantifies the strengths of the flavour-changing weak interactions. According to the latest SM global fit result [1]:

$$|V_{\mathrm{CKM}}| = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97435 & 0.22501 & 0.003732 \\ 0.22487 & 0.97349 & 0.04183^{+0.00079}_{-0.00069} \\ 0.00858 & 0.04111 & 0.999118 \end{pmatrix},$$

  where the primary contribution to the precision of $|V_{cb}|$ ($\sim 2\%$) is from semileptonic decays of B mesons to charm.

- A persistent tension exists between $|V_{cb}|$ results from inclusive and exclusive B decay methods [1]:
  - Inclusive: $(42.2 \pm 0.5) \times 10^{-3}$.
  - Exclusive to D and D^*: $(41.1 \pm 1.2) \times 10^{-3}$.

- The precision of $|V_{cb}|$ is found to be significantly affected by experimental uncertainties in flavor-tagging and mistagging efficiencies [2–4], especially under high-luminosity conditions. Reducing these uncertainties will be valuable to enhancing the unique opportunity to probe $|V_{cb}|$ via $W \to cb$ decays.

Our proposal for precise $|V_{cb}|$ measurement in the energy frontier

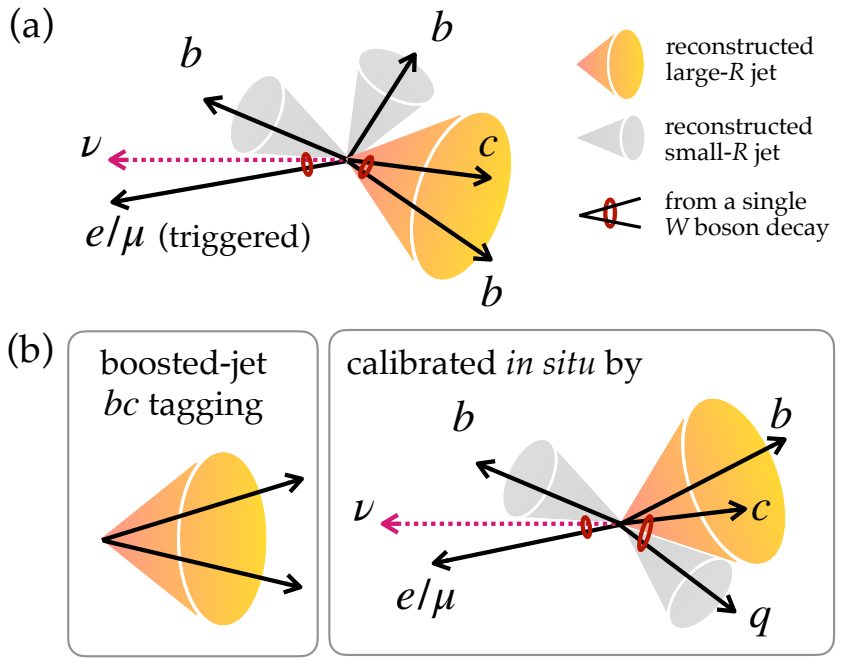
bc-tagging in boosted region

Benefits:

- Stronger background suppression: increasingly sophisticated deep learning techniques for boosted bb- and cc-tagging have brought substantial improvements in $H \to bb/cc$ measurements in CMS [5–8].

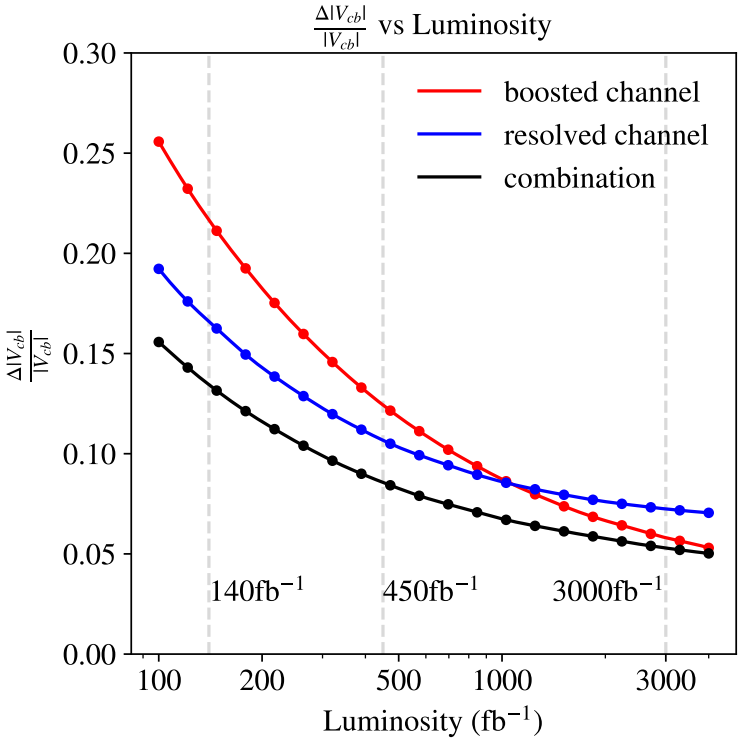
- Ability to facilitate *in-situ* calibration of the signal process, bypassing the calibration proxies [9, 10]:
  - The background is extremely dominated by "bc-matched" jets after tight bc-tagging.
  - The bc-tagger efficiency can be corrected using a shared unconstrained scale factor.

  This not only reduces uncertainty in boosted bc-tagging efficiency but also mitigates the dependence of remaining event selections on b/c flavor tagging, thereby enabling more precise extraction of $|V_{cb}|$ under high-luminosity conditions.



Outline

The high-lumi boosted regime surpasses the resolved one in $|V_{cb}|$ precision



Uncertainty contributions obtained by individually freezing $\vec{\nu}$ and λ in the fit.

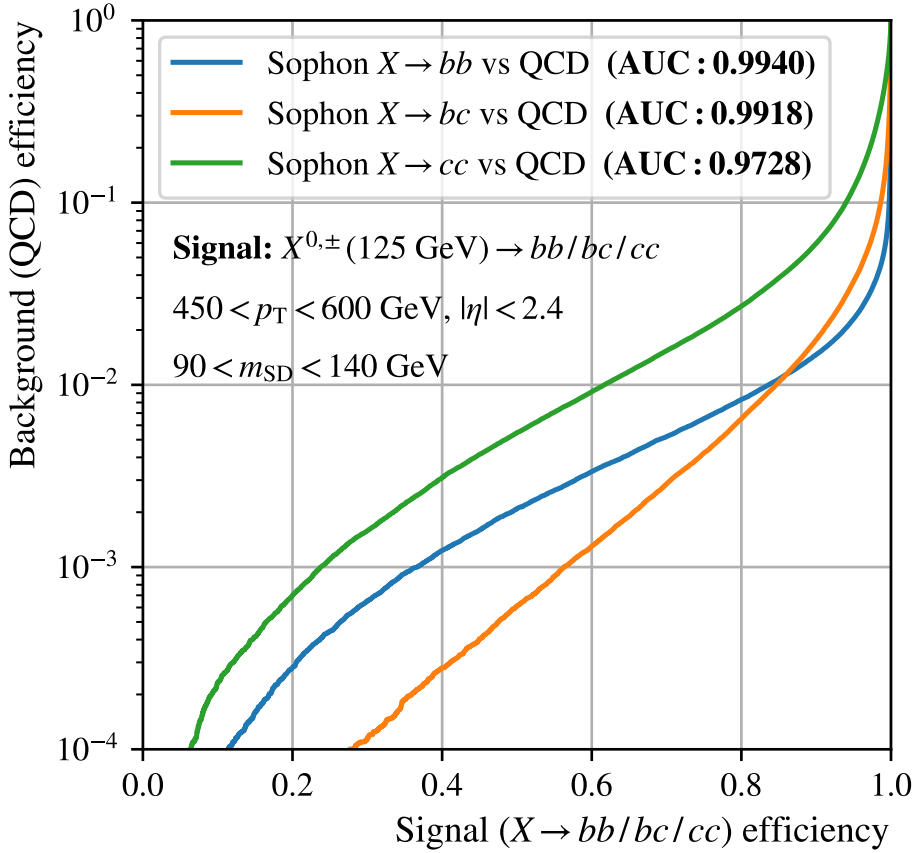
Lumi. (chan.)	b/c-tag.	D_{bc}-tag.	Stat.
140 fb^{-1} (boosted)	0.036	0.100	0.191
140 fb^{-1} (resolved)	0.065	—	0.154
450 fb^{-1} (boosted)	0.036	0.056	0.106
450 fb^{-1} (resolved)	0.065	—	0.086
3000 fb^{-1} (boosted)	**0.035**	0.022	**0.041**
3000 fb^{-1} (resolved)	0.065	—	0.033

- The uncertainty under $140\ \text{fb}^{-1}$ is consistent with the preliminary expected result from ATLAS [11] at around $0.13\,(\text{syst.}) \oplus 0.13\,(\text{stat.})$.
- The boosted channel is predominantly limited by statistical uncertainty, while the contribution from flavor-tagging-related uncertainty is smaller than the resolved channel. As the luminosity increases, the boosted channel demonstrates a significant advantage in the overall uncertainty, surpassing the traditional resolved approach.

Outline

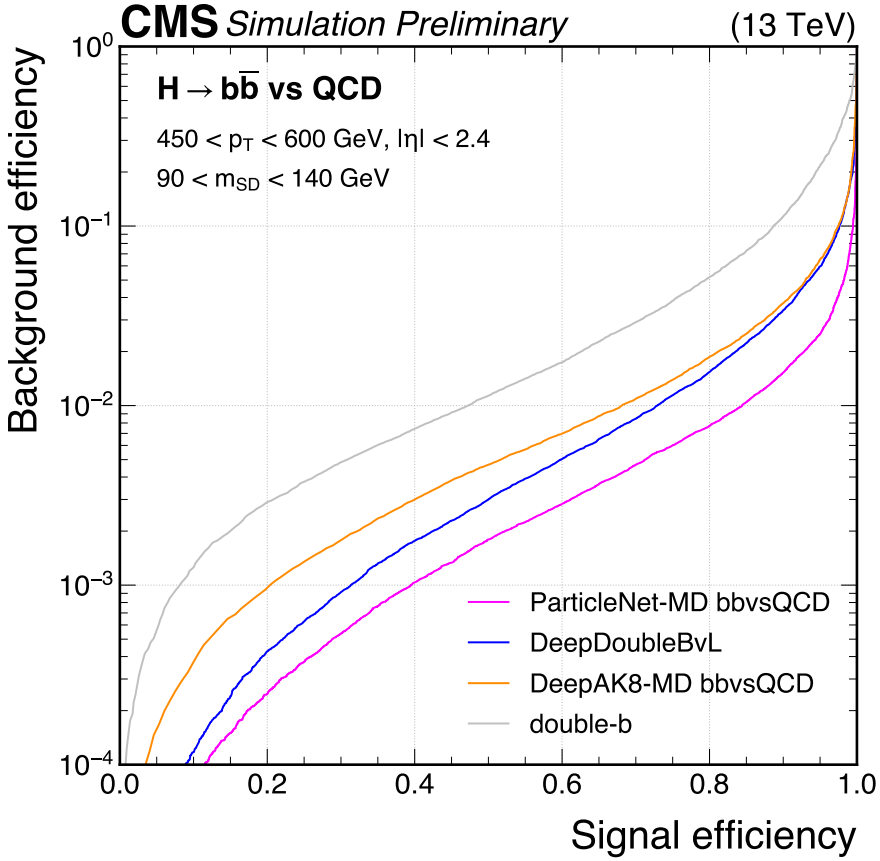
Performance of Sophon in various flavor tagging tasks

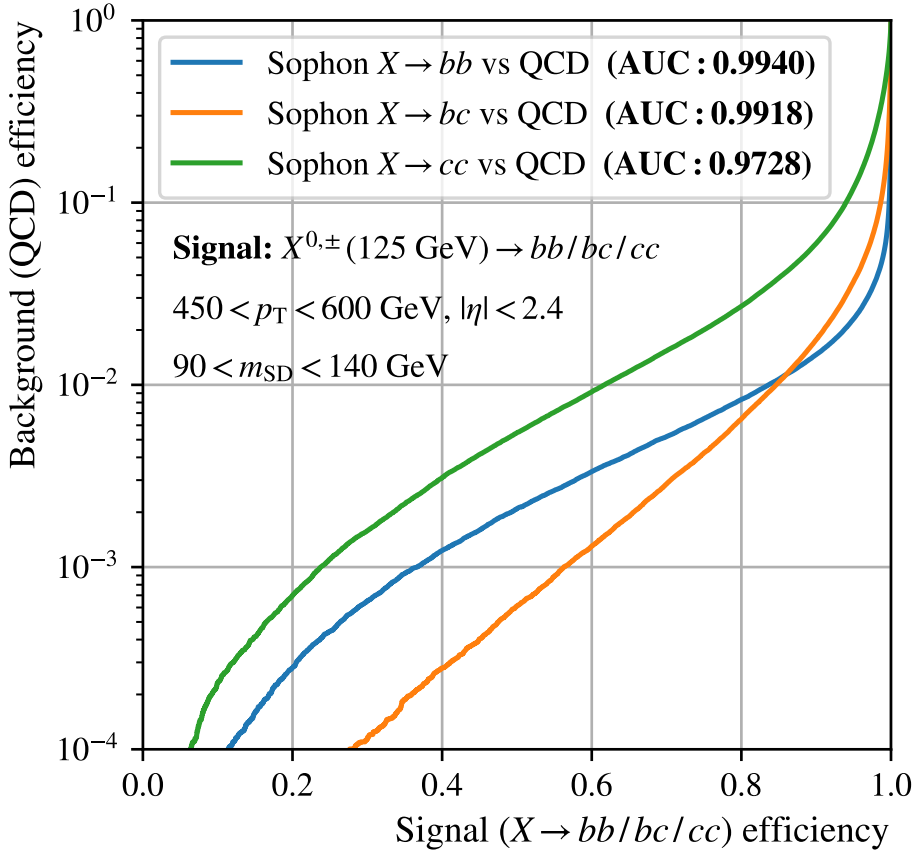
- bc tagging achieves an AUC between those of bb and cc tagging.
- At a tight working point, bc-tagging provides stronger QCD jet suppression compared to bb and cc, as QCD processes do not produce bc from gluon splitting.
- bb- and cc-tagging results are found compatible with CMS's state-of-the-art taggers ParticleNet-MD and GloParT.



Compared against state-of-the-art CMS taggers [12, 13] under consistency in signal and background definitions, jet selections, and discriminant definitions.

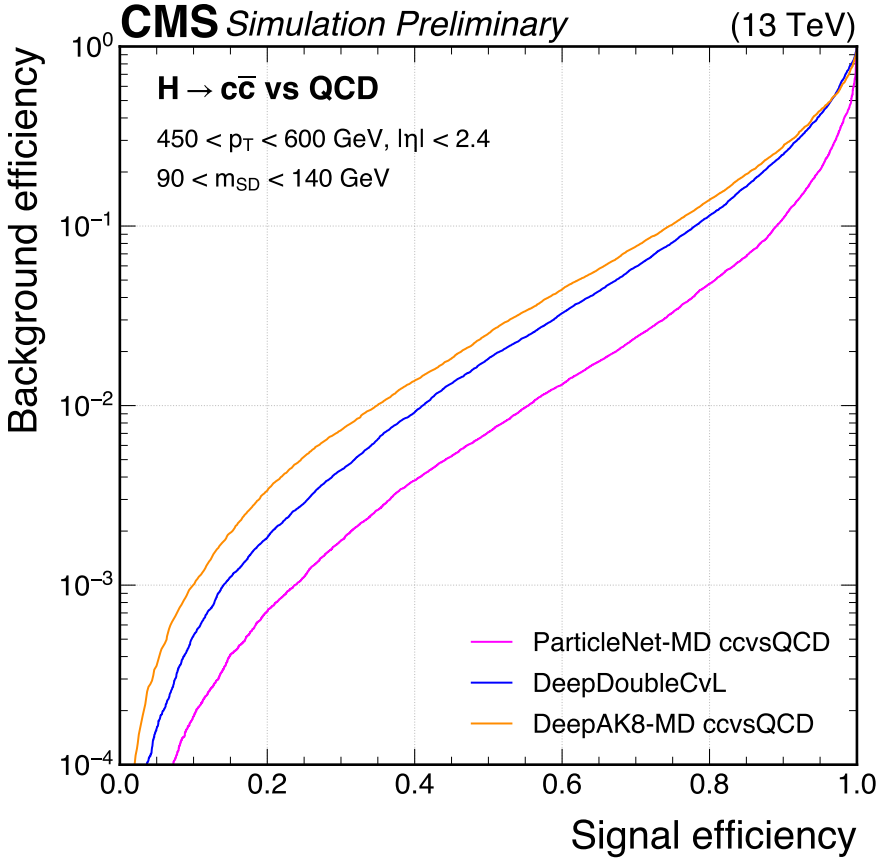
Performance of Sophon in various flavor tagging tasks

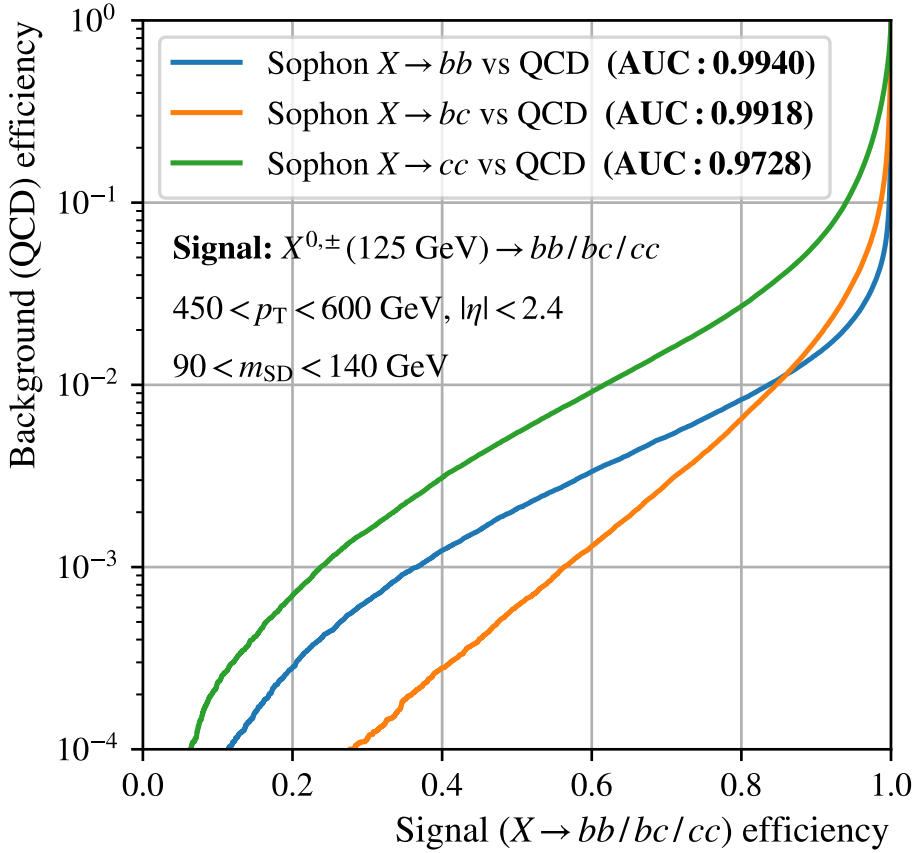




Compared against state-of-the-art CMS taggers [12, 13] under consistency in signal and background definitions, jet selections, and discriminant definitions.

Performance of Sophon in various flavor tagging tasks





Compared against state-of-the-art CMS taggers [12, 13] under consistency in signal and background definitions, jet selections, and discriminant definitions.

MC simulation

Simulated datasets for LHC pp collision at $\sqrt{s} = 13$ TeV:

- Trigger: single isolated lepton ($p_{\mathrm{T}} > 24$ GeV for e or $p_{\mathrm{T}} > 32$ GeV for μ).
- Other SM processes: $W + \mathrm{jets}$ ($W \to \ell\nu$), tW ($W \to \ell\nu$), and semileptonic WW.
- Hard process: MG5_aMC@NLO v2.9.18 with SM at LO $\to$ scaled to higher-order calculations [14–17].
- Parton shower: Pythia 8.3 [18] with an NNLO PDF in NNPDF 3.1 [19].
- Fast detector simulation: Delphes 3.5 with the same configuration as JetClass-II:
  - Based on the default CMS card.
  - Account for track smearing according to CMS tracker resolution.
  - Include PU with an average of 50 vertices and apply the PUPPI algorithm [20].
  - Cluster anti-k_{T} jets with $R = 0.4$ and $R = 0.8$, with $p_{\mathrm{T}} \geq 25$ GeV and 200 GeV, respectively.
  - Yield 1.8×10^4 inclusive $W \to cb$ events for 140 fb^{-1}.

Event selection

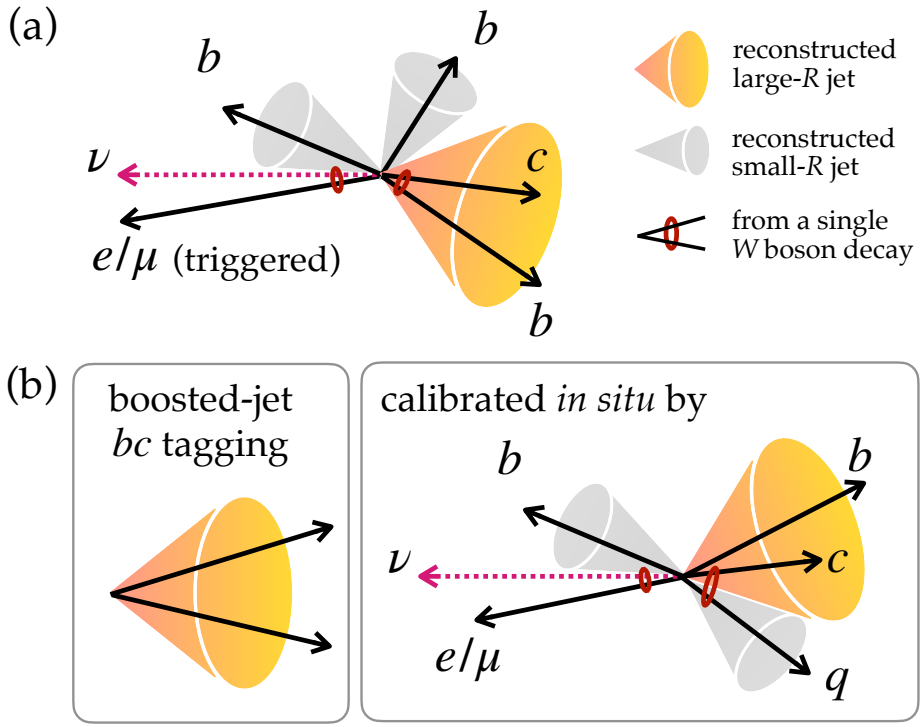
The boosted regime

Pre-selection:

- At least one $R = 0.8$ jet isolated from the trigger lepton.
- W candidate jet: the one with highest p_{T}, with $60 \leq m_{\mathrm{SD}}/\mathrm{GeV} \leq 110$ [21, 22].
- Only about 6% of signal events survives.

Event categorization:

- $t(bqq')$-matched (0.7%),
- $t(bc)$-matched (3.1%),
- $t(bq, q \neq c)$-matched (12.3%),
- $W(qq')$-matched (23.3%),
- non-matched (14.7%),
- QCD-originated (45.9%).



Event selection

The boosted regime

Two independent event selection strategies:

- For bc-content purification: define a discriminant

$$D_{bc} = \frac{g_{X\to bc}}{g_{X\to bc} + g_{X\to bq} + g_{X\to cs} + g_{X\to bqq} + g_{\mathrm{QCD}}}$$

  based on Sophon's 188 output scores $\vec{g}$.

- In parallel, train a multivariate classifier to distinguish $W(qq')$-matched-like W candidate jets from others without utilizing the Sophon's tagging information:
  - Model: Particle Transformer [23] (with pair-wise feature support)
  - Supposed categories: $W(qq')$-matched, "$t(bc)+t(bq)$"-matched, $t(bqq')$-matched, non-matched cases, and $W+\mathrm{jets}$ background
  - Input variables:
    - The triggered lepton, with its 4-vector;
    - Missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$;
    - Up to 5 $R=0.4$ jets exclusive to the triggered lepton and the W candidate jet, with their 4-vectors and flavor-tagging labels from SophonAK4.

Event selection

The boosted regime

For the 2nd classifier, before the Sophon D_{bc} selection:

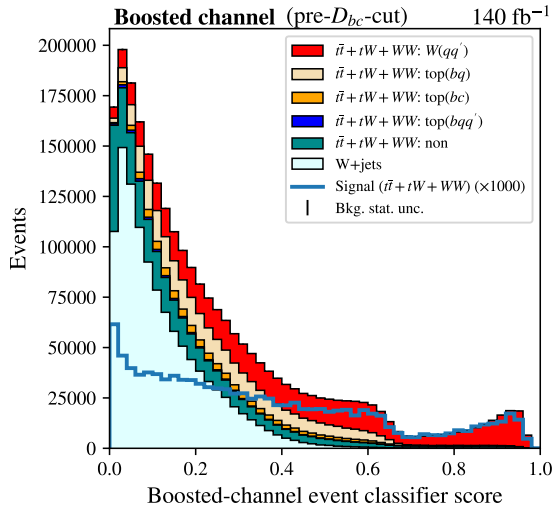
- Contributions from all background components are comparable (top figure).

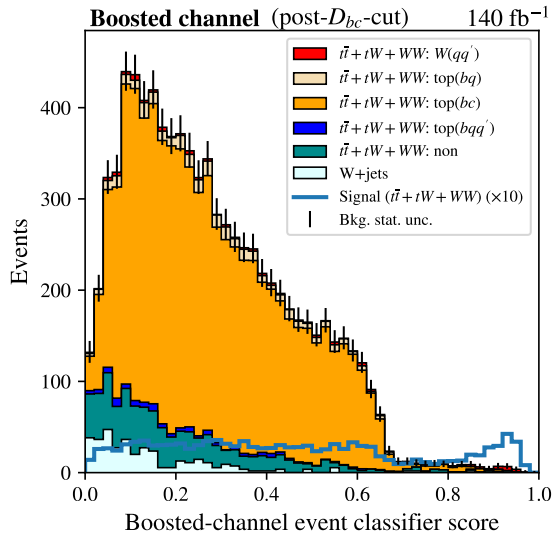
After a stringent Sophon D_{bc} selection:

- The background is predominantly composed of $t(bc)$-matched jets (bottom figure).

Main idea:

- Use event-level information fully independent of the W candidate jet content to construct another classifier.
- Pair the bc-matched background with the signal to calibrate the D_{bc} tagging efficiency: correct both yields using a shared, unconstrained scale factor.



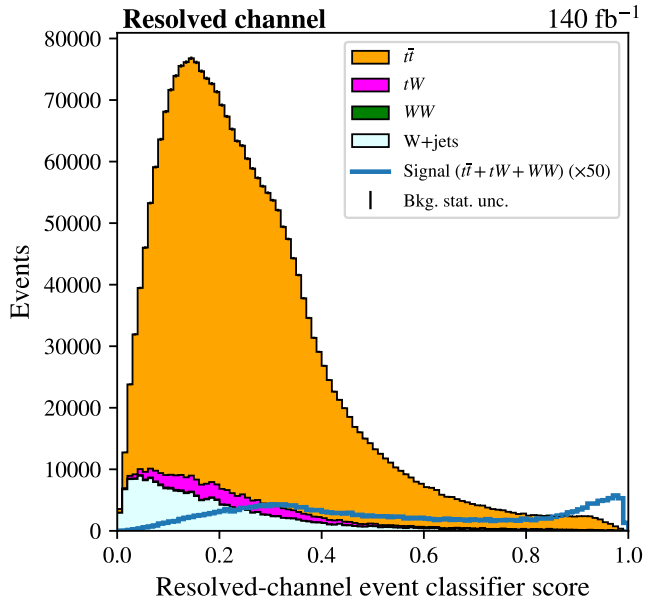


Event selection

The resolved regime (in comparison with the boosted regime)

For the resolved regime, the strategy is conventional, similar as Ref. [2]:

- Require exactly 1 lepton and at least 4 $R = 0.4$ jets exclusive to the triggered lepton, with at least 3 of them tagged as b/c.
- About 28% of the triggered signal events survive.
- A particle-transformer-based classifier similar to the boosted regime is trained to distinguish signal versus background events. The input variables are:
  - The trigger lepton (its 4-vector);
  - $p_{\mathrm{T}}^{\mathrm{miss}}$;
  - Up to 6 $R = 0.4$ jets exclusive to the triggered lepton (their 4-vectors and five SophonAK4 tagging labels).



Outline

The counting analysis for $|V_{cb}|$ extraction

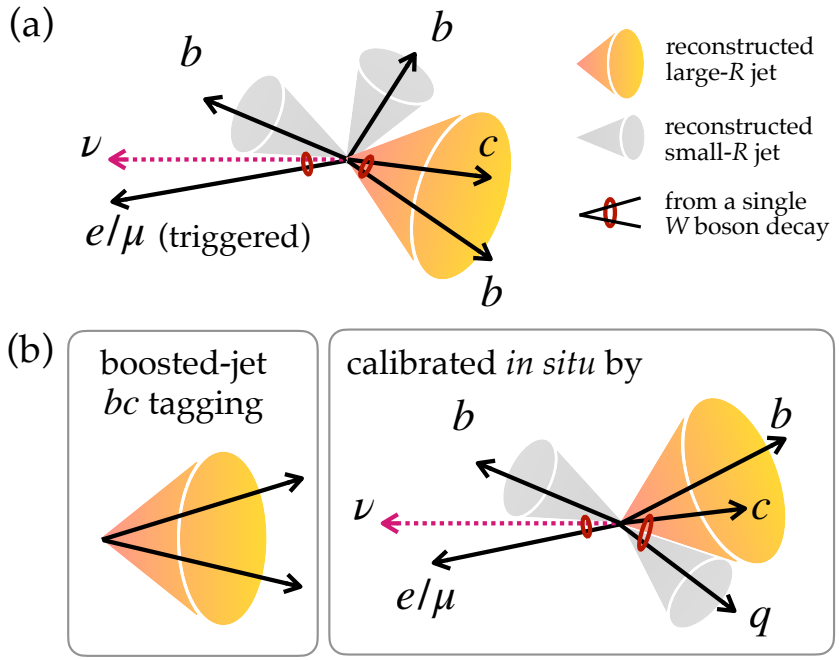
With uncertainty estimation and in-situ calibration

- Based on: events passing an optimized classifier score threshold.
- Three contributions to the total event count:
  - N_{s}: the predicted signal events count;
  - N_{b0}: the predicted background count without a hadronically decayed W boson;
  - N_{b1}: the predicted background count with a hadronically decayed W boson.
- Statistics: the signal strength $\mu = N_{\mathrm{s}}/N_{\mathrm{s}}^{\mathrm{SM}}$, theoretically $\mu = \left(\frac{|V_{cb}|^{\mathrm{obs}}}{|V_{cb}|^{\mathrm{SM}}}\right)^2$ in LO, and $N_{\mathrm{b1}} = \frac{1-\mu r}{1-r} N_{\mathrm{b1}}^{\mathrm{SM}}$, where $r := \frac{\Gamma(W\to bc)}{\Gamma(W\to qq')} = \frac{1}{2}(|V_{cb}|^{\mathrm{SM}})^2$, due to the CKM unitarity.
- Likelihood function: $\mathrm{Poisson}(\lambda = N_{\mathrm{s}}(\mu) + N_{\mathrm{b0}}(\mu) + N_{\mathrm{b1}})$ before integrating the nuisance.
- Nuisance parameters 1/2: $\vec{\nu}$ ($R = 0.4$ jet tagging uncertainty): The impacts of the $b/c/j$-jets tagged as B1/B2/C1/C2/N ($3 \times 5 = 15$ independent parameters) on the event count, with the tagging efficiencies varied but globally normalized to a constant event yield, delicately measured in $b/c/j$-enriched regions following a latest ATLAS work [24].

The counting analysis for $|V_{cb}|$ extraction

With uncertainty estimation and in-situ calibration

- Nuisance parameters 2/2: λ ($R = 0.8$ jet tagging efficiency in-situ calibration): An unconstrained scale factor λ is applied to both the signal and the $t(bc)$-matched background in the post-D_{bc}-cut region.
- Perform simultaneous fit for μ, $\vec{\nu}$, and λ.
- This shared-efficiency assumption is being validated by altering the parton shower model (Pythia → Pythia with Vincia and Herwig) and examining the stability of the post-D_{bc}-cut ratio signal to $t(bc)$-matched background.



The counting analysis for $|V_{cb}|$ extraction

Impacts of the nuisance parameters

The boosted channel:

- b-tagging efficiency is a predominant factor.
- The $W \to bc$ tagging is integrated within $R = 0.8$ jet techniques.
- The result is less affected by the $R = 0.4$ jet flavor tagging uncertainties.

The resolved channel:

- Multiple factors contribute significantly to the overall uncertainty.
- The classifier relies more heavily on identifying multiple b and c jets.

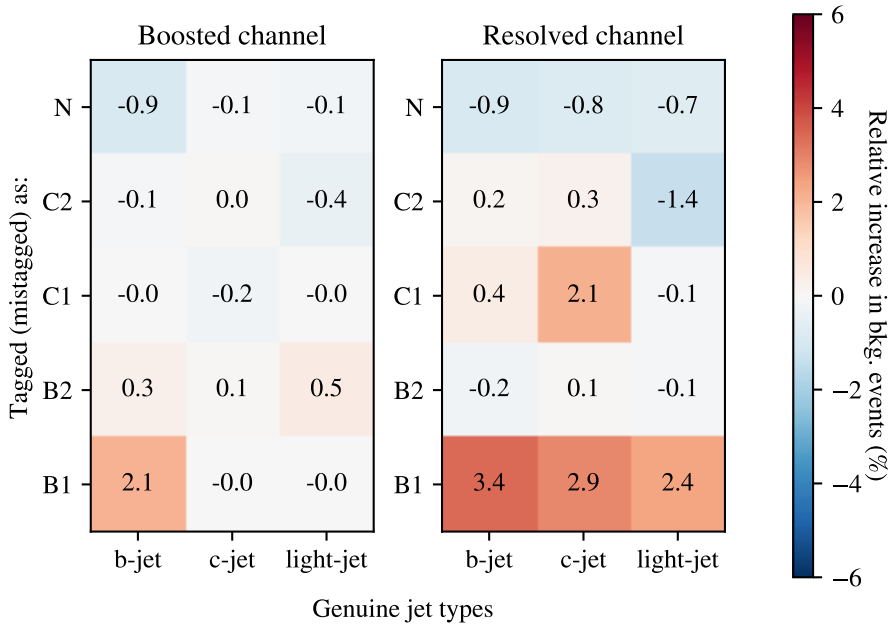
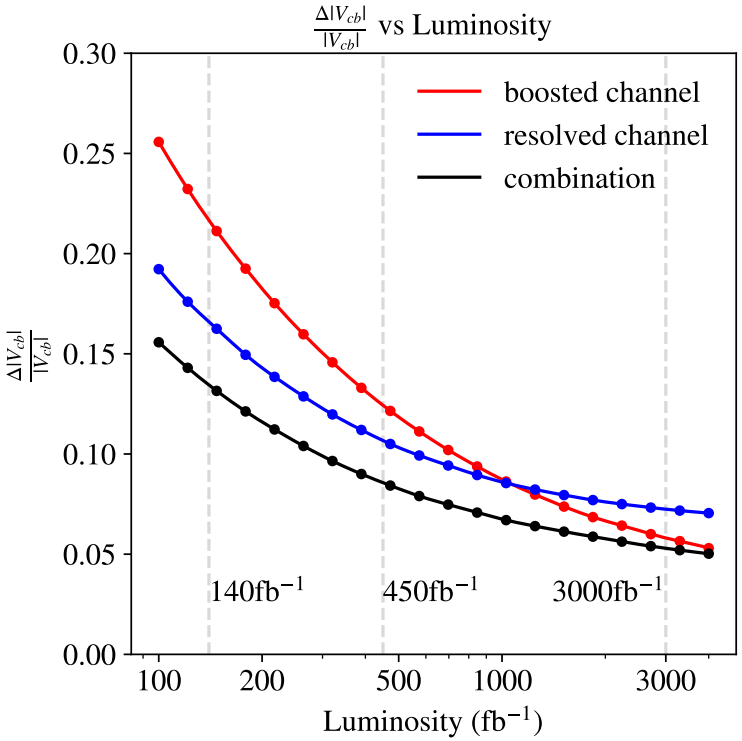


FIG. The relative increases in the background counts due to $+1\sigma$ variations in the 15 sources.

The high-lumi boosted regime surpasses the resolved one in $|V_{cb}|$ precision



Uncertainty contributions obtained by individually freezing $\vec{\nu}$ and λ in the fit.

Lumi. (chan.)	b/c-tag.	D_{bc}-tag.	Stat.
140 fb^{-1} (boosted)	0.036	0.100	0.191
140 fb^{-1} (resolved)	0.065	—	0.154
450 fb^{-1} (boosted)	0.036	0.056	0.106
450 fb^{-1} (resolved)	0.065	—	0.086
3000 fb^{-1} (boosted)	**0.035**	0.022	**0.041**
3000 fb^{-1} (resolved)	0.065	—	0.033

(Not a complete but a representative set of contributions considered.)

- The uncertainty under $140\ \text{fb}^{-1}$ is consistent with the preliminary expected result from ATLAS [11] at around $0.13\,(\text{syst.}) \oplus 0.13\,(\text{stat.})$.
- Under $3000\ \text{fb}^{-1}$, the orthogonalized combined precision reaches $\Delta|V_{cb}|/|V_{cb}| = \mathbf{0.051}$, representing a 30% reduction compared to resolved-only approach. A combination of ATLAS and CMS measurements can further reduce it to $\mathbf{0.036}$, sufficient to provide critical insights into resolving the $|V_{cb}|$ puzzle.

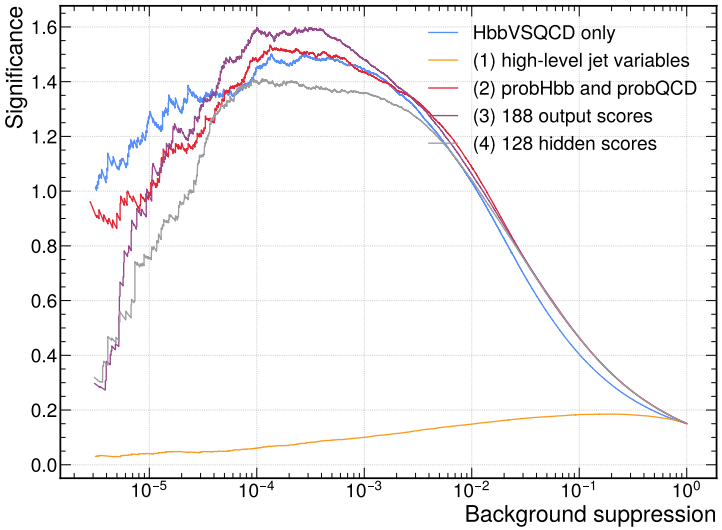
Outline

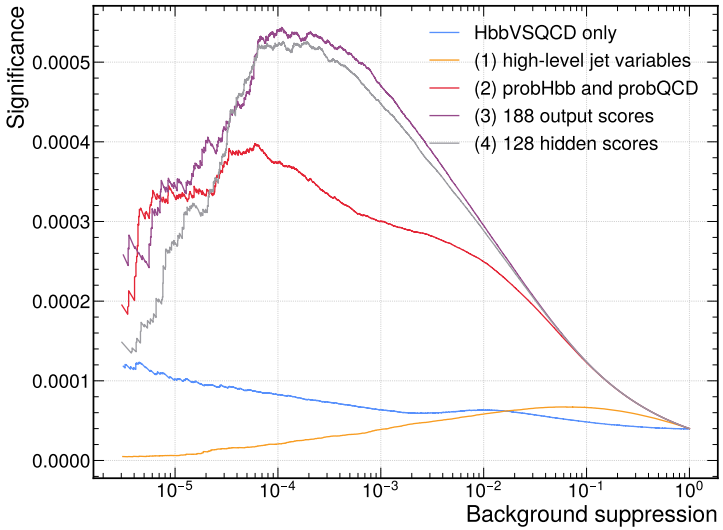
Summary and outlook

- We propose a novel approach for precise $|V_{cb}|$ extraction at the LHC, utilizing the boosted regime:
  - Employing an advanced boosted-jet bc-tagger via the Sophon model.
  - Creating an innovative *in-situ* calibration technique.
- This approach substantially improves the measurement precision over the conventional method.
- Recent advancements in deep learning algorithms for particle physics have enhanced sensitivity through well-established boosted channels such as bb and cc, revealing potential in exploring a broader range of boosted final states.
- This highlights the Sophon's philosophy: extending boosted-jet techniques to previously unexplored regions.
- It also suggests a more important role for boosted-regime searches in future LHC explorations.

Another possibility: boosted jet model fine-tuning for event selection

- Modern deep learning algorithms haven't been directly used for event-level selections, with less sophisticated jet information being utilized compared to what is available from advanced jet tagging models.
- By encoding the events incorporating a high-dimensional, comprehensive jet representation from successful pre-trained models with event-level physics objects (leptons, photons), a new-trained event-level classifier can show a more powerful background suppression capability under the same signal efficiency levels.
- This approach maintains the modularity of conventional analysis strategies, ensuring the feasibility of per-object calibrations.





Thanks for your attention!

And, welcome to Peking University for recent HEP workshops:

- Larger than Larger: Large AI Models at the Frontiers of Experimental High-Energy Physics. (1st on Jan 7, 2025, 2nd upcoming).
- Workshop on Quantum Entanglement at the Energy Frontier. (Apr 25–28, 2025; reports are welcome).

Outline

References I

[1] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024.

[2] P. F. Harrison and V. E. Vladimirov. A Method to Determine $|V_{cb}|$ at the Weak Scale in Top Decays at the LHC. *JHEP*, 01:191, 2019.

[3] P. Azzi et al. Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC. Technical report, 2019.

[4] Hao Liang, Lingfeng Li, Yongfeng Zhu, Xiaoyan Shen, and Manqi Ruan. Measurement of CKM element $|V_{cb}|$ from W boson decays at the future Higgs factories. *JHEP*, 12:071, 2024.

[5] Albert M Sirunyan et al. Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair. *Phys. Rev. Lett.*, 120:071802, 2018.

[6] Armen Tumasyan et al. Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks. *Phys. Rev. Lett.*, 131:041803, 2023.

[7] Armen Tumasyan et al. Search for Higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 131:061801, 2023.

[8] Armen Tumasyan et al. Search for Higgs boson and observation of Z boson through their decay into a charm quark-antiquark pair in boosted topologies in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 131:041801, 2023.

[9] Albert M Sirunyan et al. Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques. *JINST*, 15:P06005, 2020.

[10] Georges Aad et al. Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV. *Eur. Phys. J. C*, 79:836, 2019.

References II

[11] Bryn Arthur Roberts. *Progress towards the first measurement of the V_{cb} element of the CKM matrix in semi-leptonic $t\bar{t}$ decays*. PhD thesis, Warwick U., 2022. Presented 2022.

[12] CMS Collaboration. Performance of heavy-flavour jet identification in boosted topologies in proton-proton collisions at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-BTV-22-001, 2022.

[13] Congqiao Li. *Modern deep learning for large-R jet tagging—algorithms, calibration methods, and applications in the CMS experiment*. PhD thesis, Peking U., Beijing, 2024. Presented 24 May 2024.

[14] Michal Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[15] Kirill Melnikov and Frank Petriello. Electroweak gauge boson production at hadron colliders through $O(\alpha_s^2)$. *Phys. Rev. D*, 74:114017, 2006.

[16] Nikolaos Kidonakis. NNLL threshold resummation for top-pair and single-top production. *Phys. Part. Nucl.*, 45:714, 2014.

[17] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi. W^+W^- production at hadron colliders in next to next to leading order QCD. *Phys. Rev. Lett.*, 113:212001, 2014.

[18] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

References III

[19] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77:663, 2017.

[20] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup Per Particle Identification. *JHEP*, 10:059, 2014.

[21] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *JHEP*, 09:029, 2013.

[22] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014.

[23] Huilin Qu, Congqiao Li, and Sitian Qian. Particle Transformer for jet tagging. In *Proceedings of the 39th International Conference on Machine Learning*, pages 18281–18292, 2022.

[24] Georges Aad et al. Measurements of WH and ZH production with Higgs boson decays into bottom quarks and direct constraints on the charm Yukawa coupling in $13\,\mathrm{TeV}pp$ collisions with the ATLAS detector. 10 2024.