Boost physics analysis at BESIII with Deep Learning

Yunxuan Song[1,2,3], **Yangu Li[1,2]**, Linxuan Zhu[2], Xiaoxuan Ding[1], Junpeng Zhao[2], Peirong Li[4], Jifeng Hu[5], Xiao-Rui Lyu[2], Dayong Wang[1], Yajun Mao[1], Yangheng Zheng[2]

[1]PKU [2]UCAS [3]EPFL [4]LZU [5]SCNU

Quantum Computing and Machine Learning Workshop

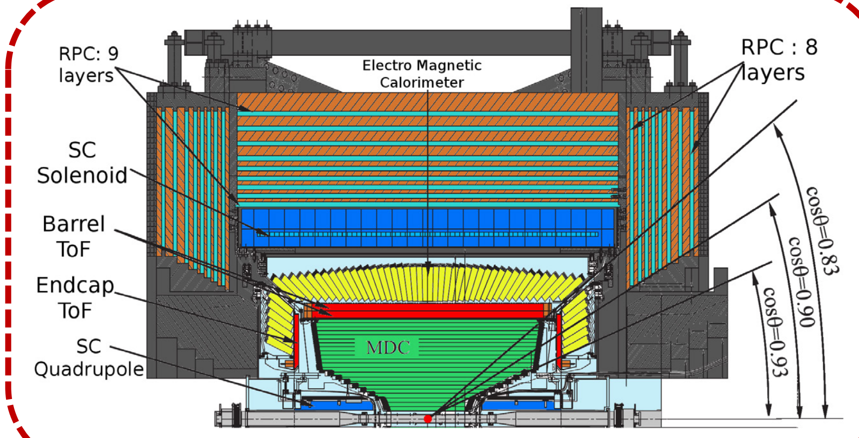
August 13, 2023

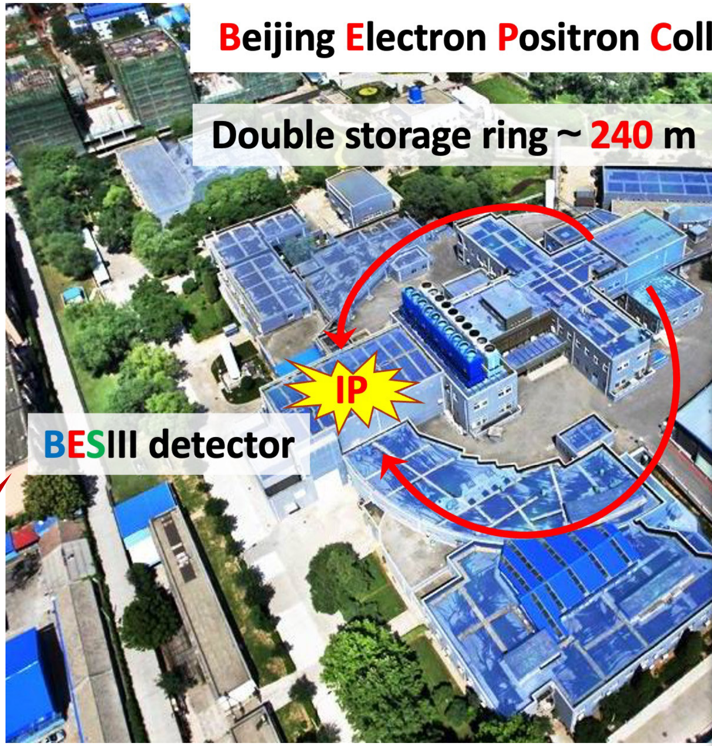
BESIII

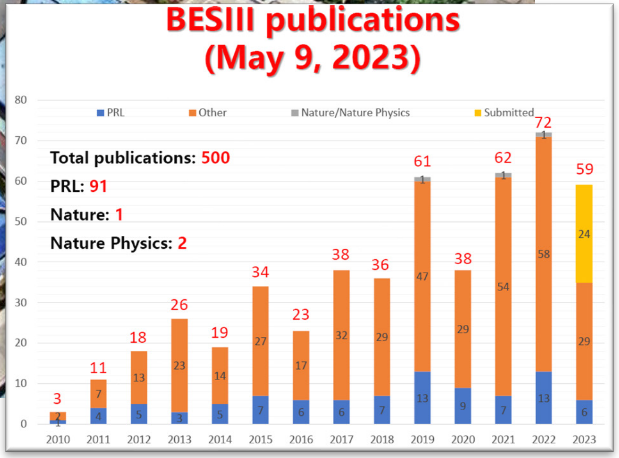
The BESIII experiment

- **BEPCII: an e^+e^- collider**
  - Operation at $\sqrt{s} = 2.0{\sim}4.95\ \mathrm{GeV}$
  - Peak luminosity: $1.1 \times 10^{33}\ \mathrm{cm^{-2}s^{-1}}$
  - More than 47 fb^{-1} of data taken since 2009

- **BESIII: a multi-purpose 4π detector with**
  - Good tracking
  - Calorimetry
  - PID & muon detection



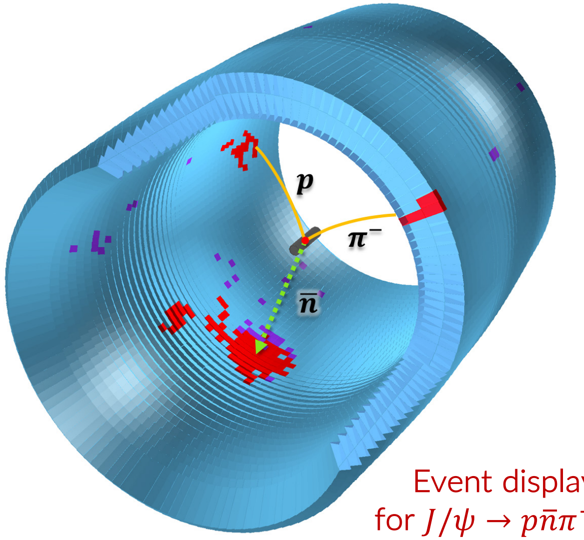




- **Over 500 publications achieved**
  - Improving data analysis technique is vital to fully explore the physics potential!

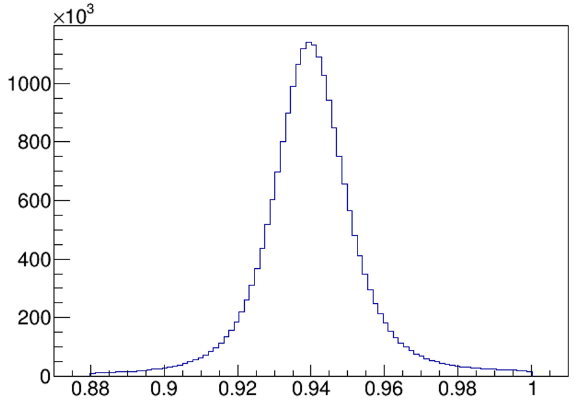
Issues in physics analysis: Neutral hadron reconstruction

- **Neutral hadron $(n \,\&\, K_L^0)$ study at BESIII**
  - HCAL is absent in BESIII detector
  - Scattering & energy leak when interact with EMC
  - Typical image:
    - An indefinite number of EMC showers
    - Showers have expanded cluster shape & low deposit energy



Event display for $J/\psi \to p\bar{n}\pi^-$

- **Recoil strategy**
  - Reconstruct all other particles in event, then calc the recoil 4-momentum
  - Works fine in most cases except more than one 'invisible' particles
    - Decays with both neutron & neutrino: $\Lambda_c^+ \to ne^+\nu_e$, ...
    - Decays with both neutron & photon: $\Lambda \to n\gamma$, ...
    - Decays with multiple neutrons: $e^+e^- \to n\bar{n}$, ...



Invariant mass spectrum for recoil $\bar{n}$

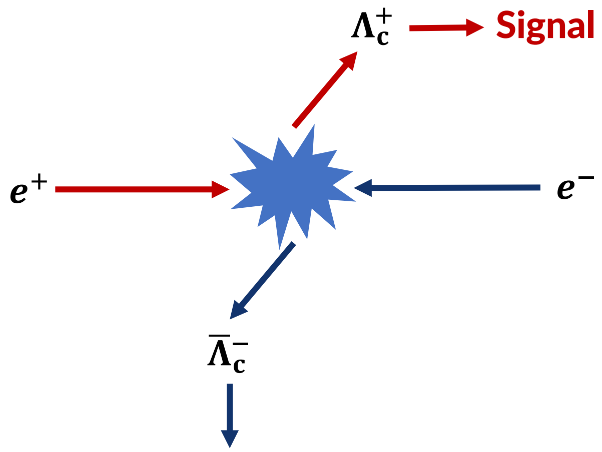
- **More straightforward & sensitive approach?**

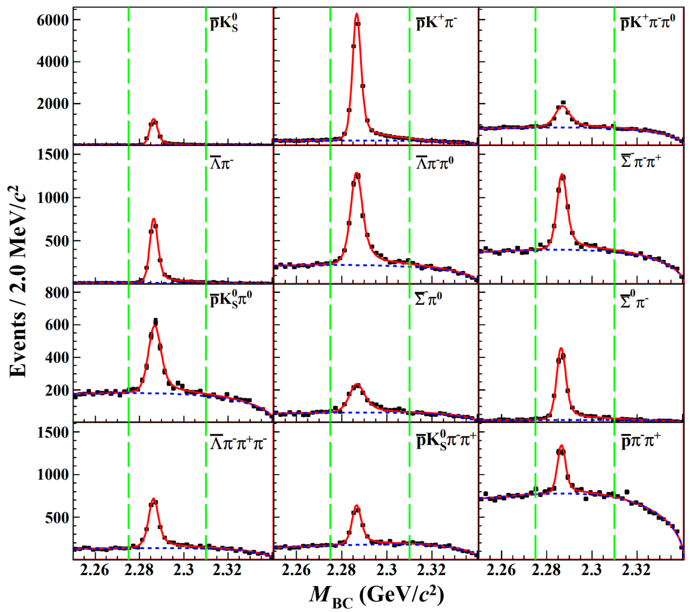
Issues in physics analysis: Charmed hadron tagging

- **Charmed hadron study at BESIII**
  - Mostly produced above $h\bar{h}$ threshold ($h = D^{0/+},\ D_s^+,\ \Lambda_c^+ \ldots$)
  - Two reconstruction strategy
    - Single-tag (ST): not constrain $\bar{h}$ decay
    - Double-tag (DT): constrain $\bar{h}$ decay exclusively
  - Tradeoff between signal efficiency & background level

- **Efficiency problem**
  - Charmed hadrons have $\mathcal{O}(100)$ decay modes
  - A dozen of primary decay modes are considered in DT
  - Total covered BF $\sim 30\%$, reconstructed $h\bar{h}$ pairs only $\sim 15\%$

- **Can we improve the tagging performance by doing it inclusively?**



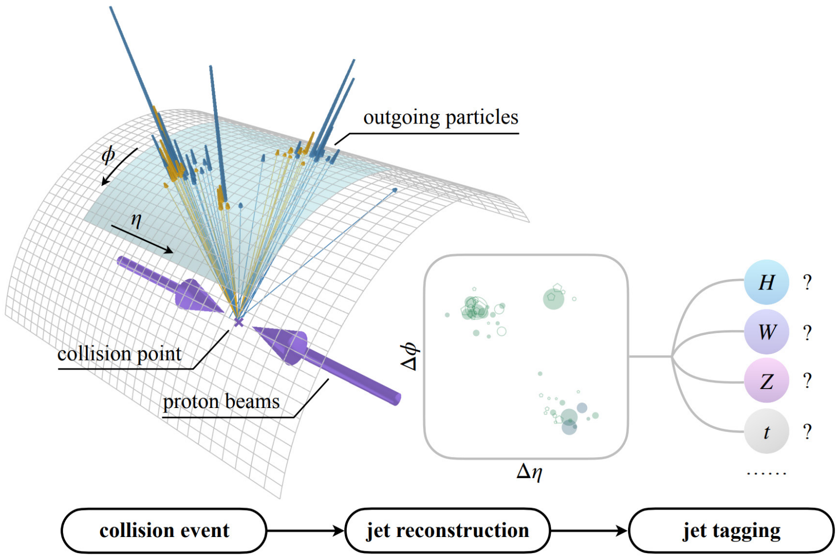


Issues in physics analysis

- **What's common between above two issues?**
  - Information embedded into an indefinite number of particles

- **Why are classical MVA tools not good solution?** (e.g., BDT)
  - Only process fixed number of features – loss of information anyway
  - Obsolete ML structure – less efficient to find signal process

- **Stones from other hills -** jet tagging at LHC
  - Jet: a collimated spray of particles (with indefinite number)
  - Jet tagging: identify the elementary particle that initiates a jet
  - Revolutionary changes thanks to deep learning
    - Continuous iteration using cutting-edge ML techniques
    - Officially recognized and supported by CMS & ATLAS Collaboration
    - International workshop: ML4Jets

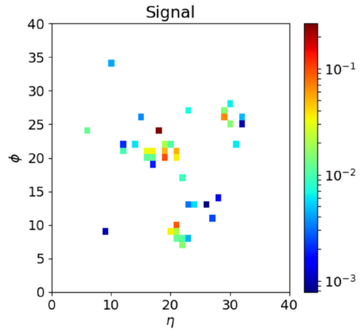


2202.03772

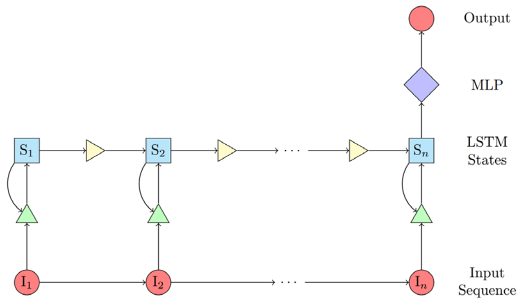
Our deep learning toolkit

- **Event representation – point cloud**
  - As image?
    - ✓ Natural idea by regarding calorimeter cells as pixels
    - ✓ Can benefit from **computer vision (CV)** applications
    - ! Most pixels remain blank
    - ! Nontrivial to combine non-additive features of particles
  - As sequence?
    - ✓ More compact data structure
    - ✓ Straightforward to include any kind of features
    - ✓ Can benefit from **natural language processing (NLP)** applications
    - ! Impose a sorting order manually
  - As **point cloud!**
    - Unordered, permutation-invariant set of particles
    - Each particle carries spatial coordinates + additional features (charge, momentum, track & shower parameters, etc.)
    - Symmetry-preserving, high expressiveness, low computational cost

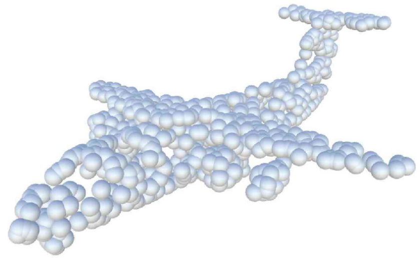
The physics result will remain same regardless of the coordinates entered!



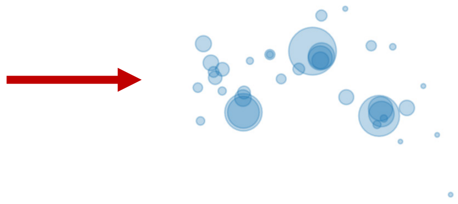
2012.09719



1607.08633



Point cloud of an aircraft generated by 3D scanning

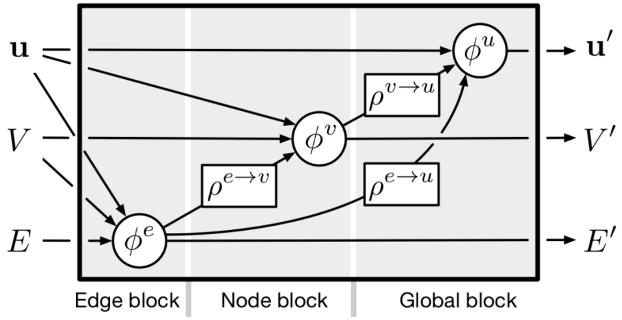


Point cloud of a HEP event

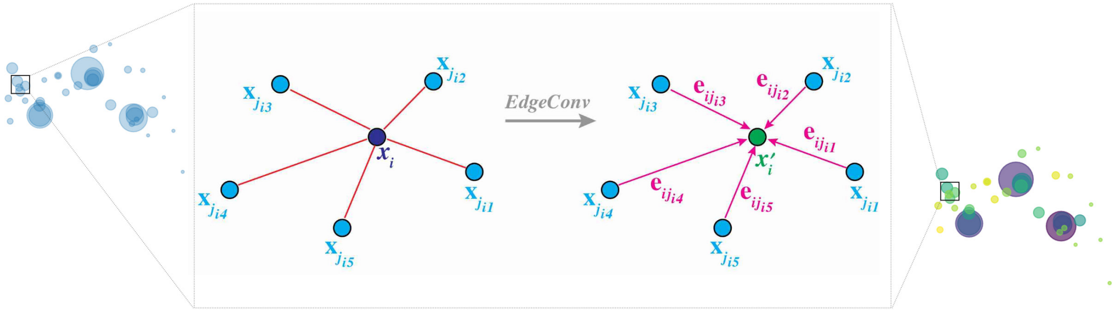
Our deep learning toolkit

- **Model structure – Graph Neural Network (GNN)**
  - Graph with vertex (node), edge (link) and global level features
  - "Message passing" framework
  - Points in cloud naturally becomes nodes, while edges remain to be defined

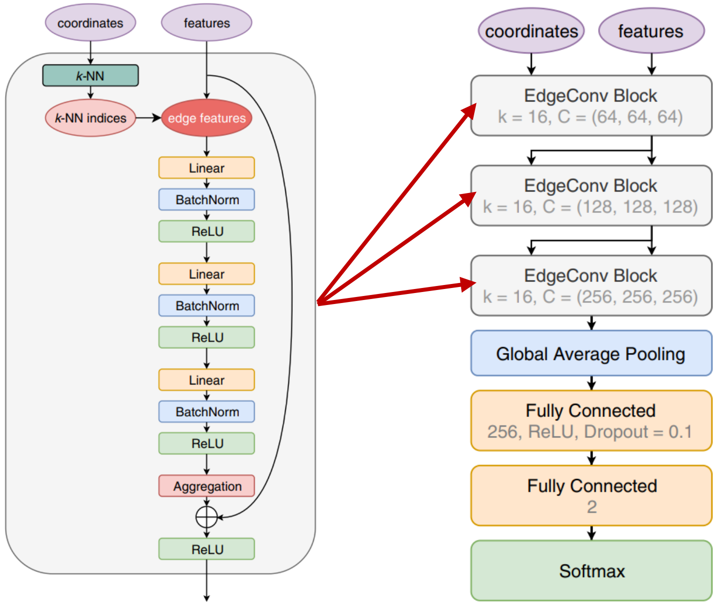
2007.13681



- **ParticleNet:** PRD 101 (2020), 056019
  - Build "edge features" between k-nearest neighboring points
  - Design a symmetric "convolution" function on edges
  - Dynamically update the graph



$e_{ij} = \mathrm{MLP}(x_i, x_j)$ $x_i' = \mathrm{mean}_j\, e_{ij}$

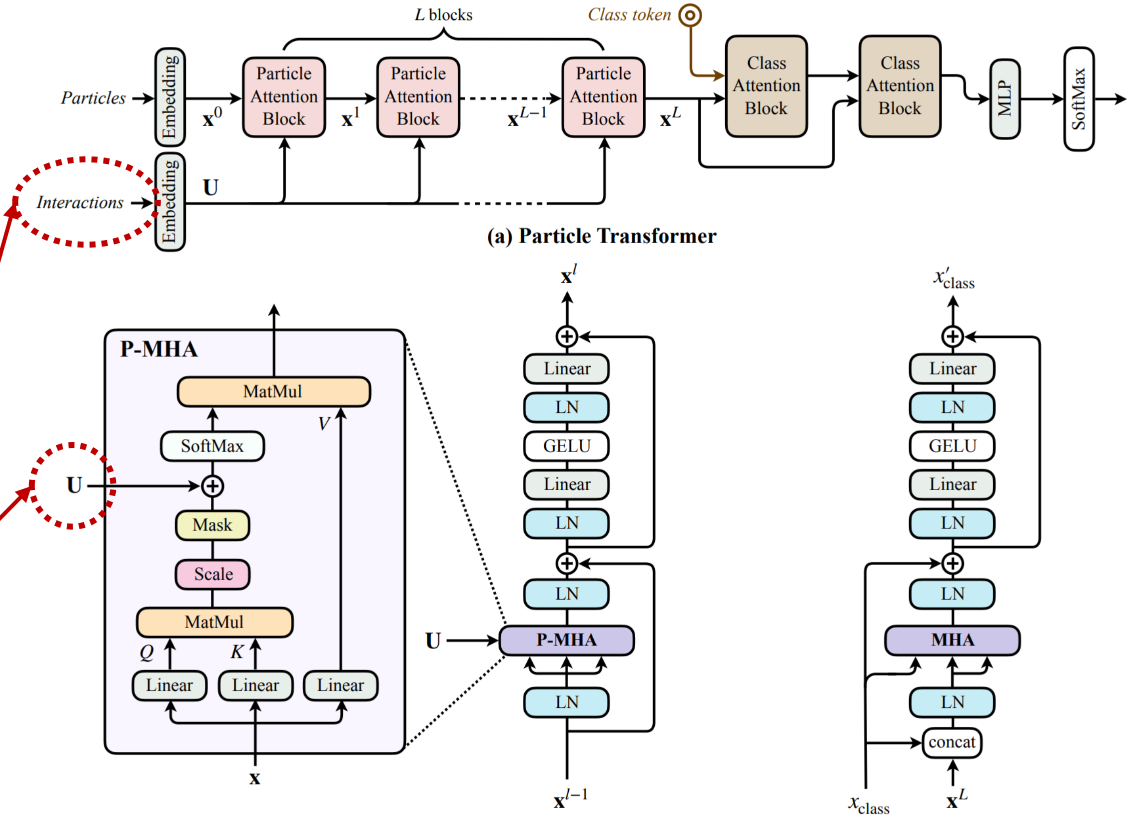


Our deep learning toolkit

- **Model structure – Transformer**
  - Foundation of Large Language Models like GPT
  - Core concept: self-attention mechanism

- **Particle Transformer:** [2202.03772](2202.03772)
  - A transformer model tailored for particle physics
  - Inject physics-inspired pairwise features as "bias" to the self-attention block
  - Can be also viewed as a fully-connected GNN

$$\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2},$$
$$k_T = \min(p_{T,a}, p_{T,b})\Delta,$$
$$z = \min(p_{T,a}, p_{T,b})/(p_{T,a} + p_{T,b}),$$
$$m^2 = (E_a + E_b)^2 - ||\mathbf{p}_a + \mathbf{p}_b||^2.$$



(a) Particle Transformer

(b) Particle Attention Block

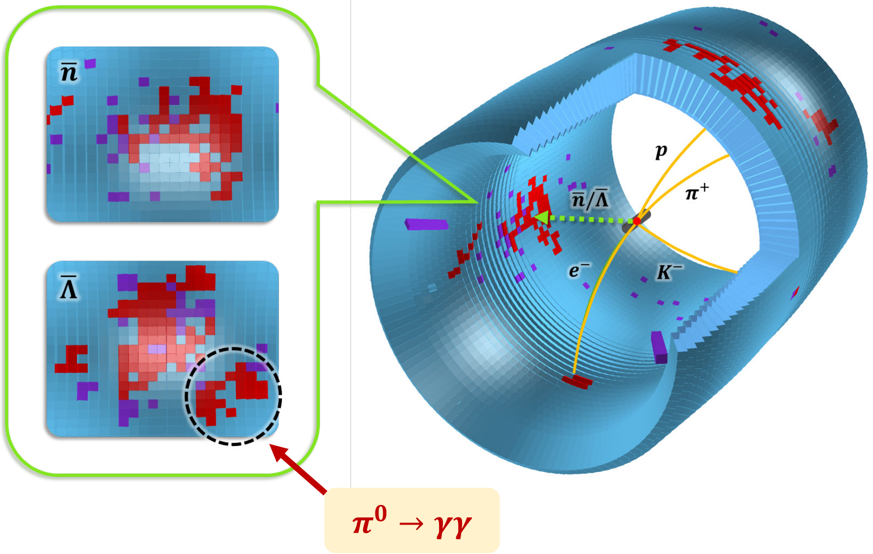
(c) Class Attention Block

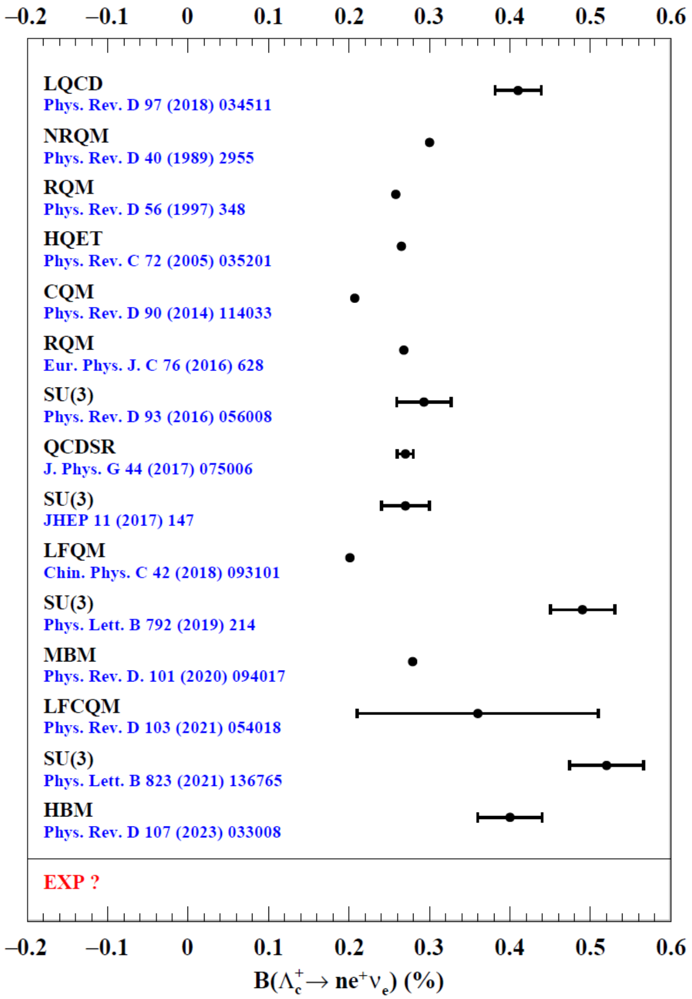
Application in Neutral hadron reconstruction: Study of $\Lambda_c^+ \to n e^+ \nu_e$

Study of $\Lambda_c^+ \to n e^+ \nu_e$

- **Physics motivation**
  - Λ_c^+ semi-leptonic decays are essential to test non-perturbative QCD
  - Tons of theoretical predictions, yet never experimentally observed

- **Experimental challenge**
  - Decays with both neutron & neutrino
  - Dominant background from $\Lambda_c^+ \to \Lambda(n\pi^0)e^+\nu_e$, yield ~10x signal
  - Unable to extract signal with conventional analysis strategy



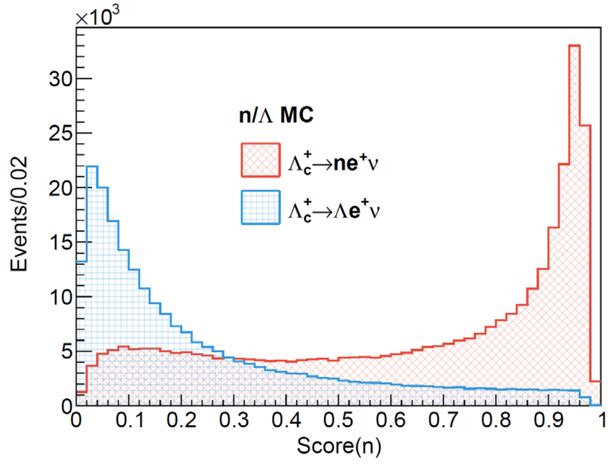


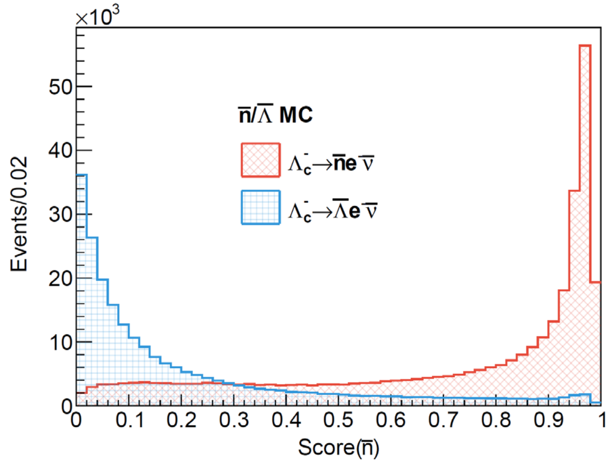
Deep learning solution for $\Lambda_c^+ \to n e^+ \nu_e$

- **Use GNN to identify $\Lambda_c^+ \to n e^+ \nu_e$ from $\Lambda e^+ \nu_e$**
  - Form point cloud with n/Λ induced EMC showers
  - Input features including energy deposition & cluster expansion information
  - Treat conjugate channels separately
    - Anti-neutron may annihilate while neutron won't
  - Firstly try training using MC simulated samples ->

- **Good separation! But…**
  - The Geant4-based MC simulation for neutron is not perfect
  - Up to 10~20% difference from real data
  - In HEP view: data-MC discrepancy can affect BF measurement
  - In ML view: domain shift if train with MC but predict with data

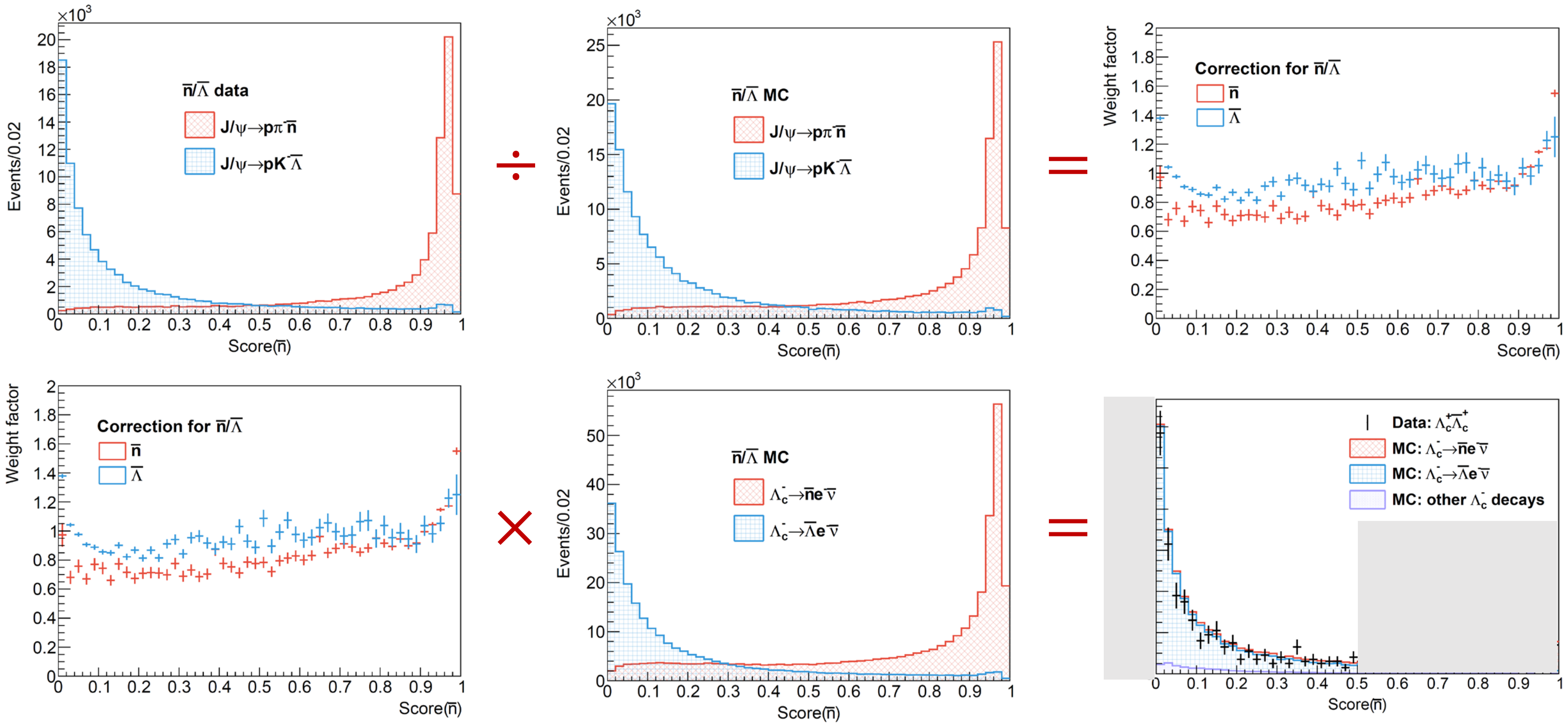
- **BESIII has various "control samples" for study**
  - Thanks to its large datasets & low background
  - Select 20M $J/\psi \to \bar{p} n \pi^+$ & $\bar{p} \Lambda K^+$ events from real data

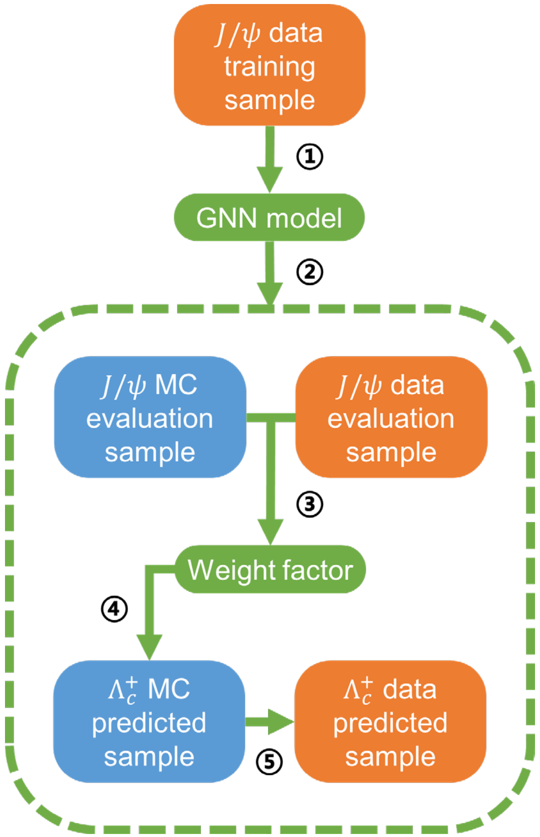




Deep learning solution for $\Lambda_c^+ \to n e^+ \nu_e$

- **A data-driven procedure to extract signal**
  1. Train GNN model using real data from $J/\psi \to \bar{p} n \pi^+$ vs. $\bar{p} \Lambda K^+$
  2. Predict 4 datasets
  3. Weight model responses between J/ψ data / MC datasets
  4. Correct Λ_c^+ MC shape with the weight factor
  5. Fit to Λ_c^+ data using Λ_c^+ MC shape



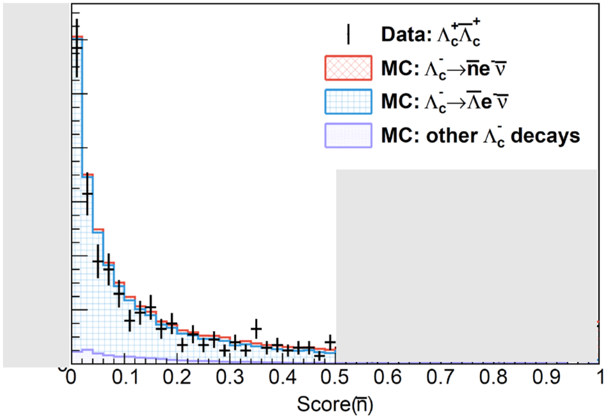


Good separation & data-MC consistence now!

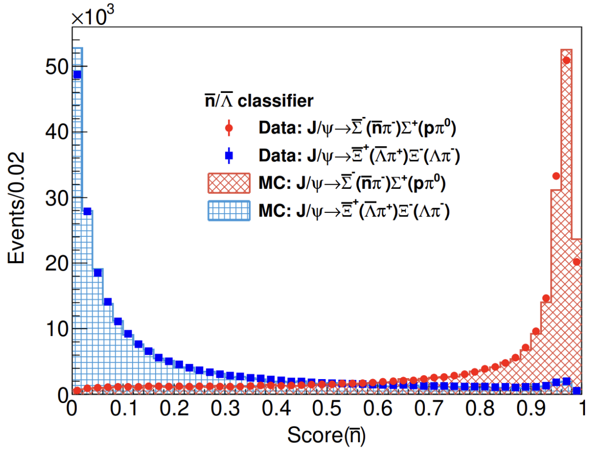
Deep learning solution for $\Lambda_c^+ \to n e^+ \nu_e$

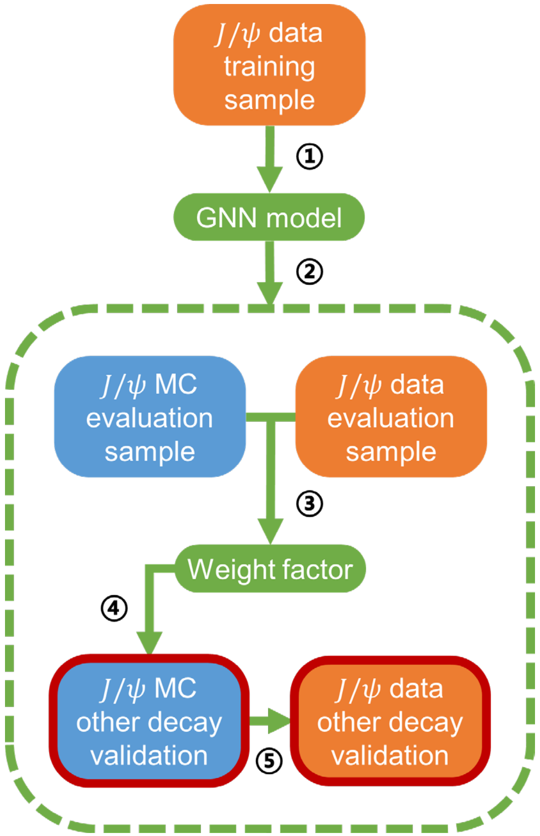
- **Good separation & data-MC consistence now! But…**
  - The phase space & background environment is different between J/ψ & Λ_c^+ samples
  - In HEP view: should prove the weight factors have no bias
  - In ML view: another source of domain shift

- **Apply another control sample for prediction**
  - $J/\psi \to \Sigma^+(n\pi^+)\bar{\Sigma}^-(\bar{p}\pi^0)$ vs. $J/\psi \to \Xi^-(\Lambda\pi^-)\bar{\Xi}^+(\bar{\Lambda}\pi^+)$
  - Data & MC still well consistent in large statistics after weighting



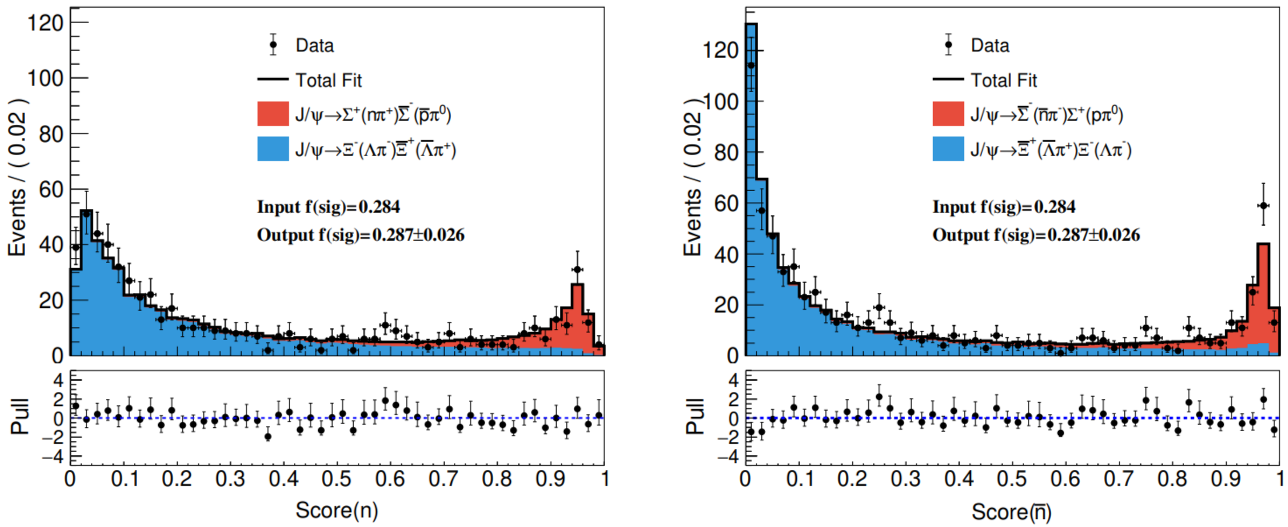
=





Deep learning solution for $\Lambda_c^+ \to n e^+ \nu_e$

- **Finally…**
  - Assume $\mathcal{B}(\Lambda_c^+ \to n e^+ \nu_e) = 3 \times 10^{-3}$ to construct pseudo data
  - We can extract the signal with $> 10\sigma$ significance & no bias



- **The physics result is under internal review. Stay tuned!**

Application in Charmed hadron tagging: Study of $\Lambda_c^+ \to p\pi^0$

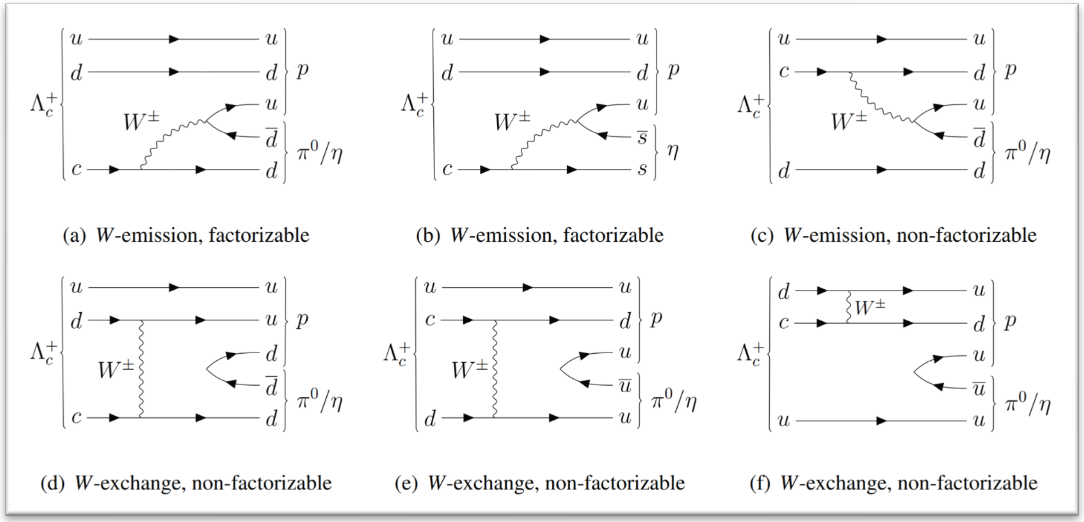
Study of $\Lambda_c^+ \to p\pi^0$

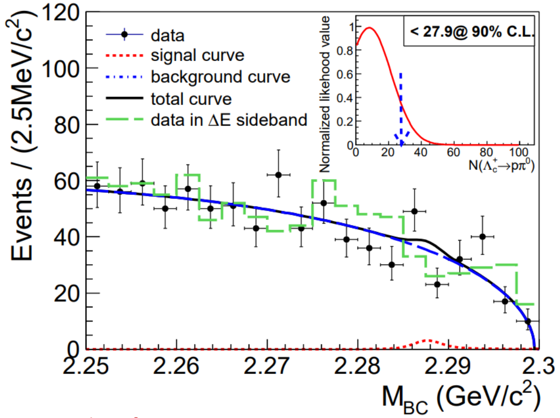
Physics motivation

- Non-leptonic, Cabibbo-Suppressed weak decay
- $\Lambda_c^+ \to p\eta$ well measured, $\Lambda_c^+ \to n\pi^+$ also observed
- $\Lambda_c^+ \to p\pi^0$ not yet observed

Experimental challenge

- High background for ST, mostly $q\bar{q}$ $(q = u, d, s)$
- Neither ST nor DT can achieve sufficient signal sensitivity



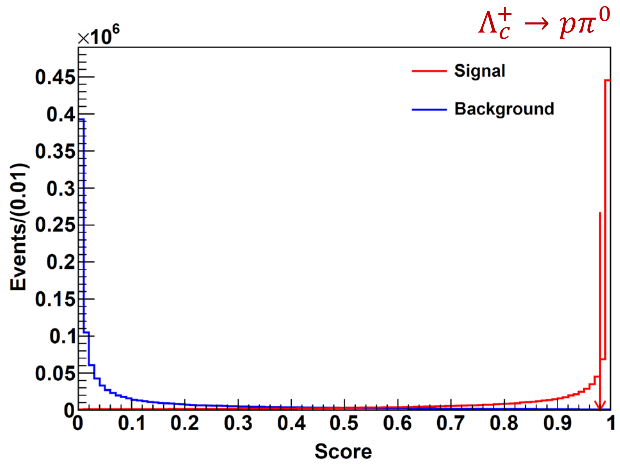


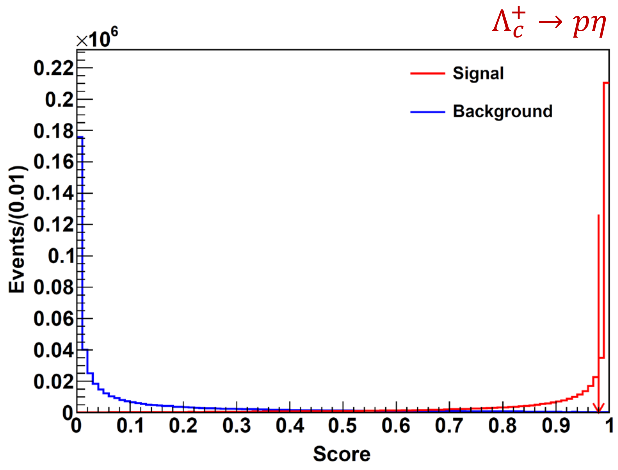
Single-tag
Phys. Rev. D 95, 111102 (2017)

Model	$\mathcal{B}(\Lambda_c^+ \to p\pi^0) \times 10^4$	$\mathcal{B}(\Lambda_c^+ \to p\eta) \times 10^4$	$\mathcal{B}(\Lambda_c^+ \to n\pi^+) \times 10^4$	$\frac{\mathcal{B}(\Lambda_c^+ \to n\pi^+)}{\mathcal{B}(\Lambda_c^+ \to p\pi^0)}$
Constituent quark model [7]	(1, 2)	3	(8, 9)	(8, 4.5)
Heavy quark effective theory [8]	1.1 - 3.6	-	1.0 - 2.1	0.5
Dynamic calculation [9, 10]	(0.75, 1.3)	12.8	2.66	(3.5, 2.1)
Topological diagram [11]	$0.8^{+0.9}_{-0.8}$	11.4 ± 3.5	7.7 ± 2.0	9.6
Topological diagram [12]	$(0.3^{+1.0}_{-0.3},\ 0.4^{+1.7}_{-0.4})$	(14.2 ± 2.3, 14.7 ± 2.8)	(7.6 ± 1.7, 8.3 ± 2.6)	(25.3, 20.8)
SU(3) flavor symmetry [13]	2	-	4	2
SU(3) flavor symmetry [14]	4.8	-	9.7	2.0
SU(3) flavor symmetry [15]	5.7 ± 1.5	-	11.3 ± 2.9	2.0
SU(3) flavor symmetry [16]	1.3 ± 0.7	13.0 ± 1.0	6.1 ± 2.0	4.7
SU(3) flavor symmetry [17]	$1.1^{+1.3}_{-1.1}$	11.2 ± 2.8	7.6 ± 1.1	6.9
BESIII experiment	< 2.7 [18]	12.4 ± 3.0 [18]	6.6 ± 1.3 [21]	-
Belle experiment	< 0.8 [20]	14.2 ± 1.2 [20]	-	> 7.2

Deep learning solution for $\Lambda_c^+ \to p\pi^0$

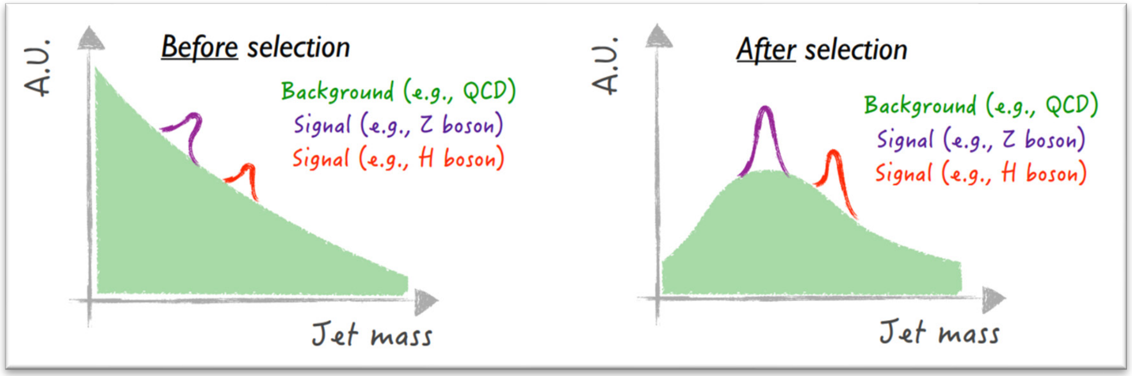
- **Use GNN to identify $\Lambda_c^+ \to p\pi^0$ from hadronic background**
  - Form point cloud with all recorded particles in event
  - Input features including momentum & low-level detector responses
  - Embedded information for training:
    - Kinematics, reconstruction quality & PID of each particle
    - Topology & dynamics of full e^+e^- decay
- **We can only use MC for training now!**
  - Study $\Lambda_c^+ \to p\eta(\gamma\gamma)$ as a reference channel
  - MC should be co-directional biased from data for $p\pi^0$ & $p\eta$
  - Report their relative BF ratio can greatly cancel the systematics
  - To achieve consistent training performance, data augmentation is applied
    - Mix their training samples
- **Dramatic separation!**
  - Two orders of magnitude BKG suppression @ 50% signal efficiency
  - Can be further enhanced via model ensemble





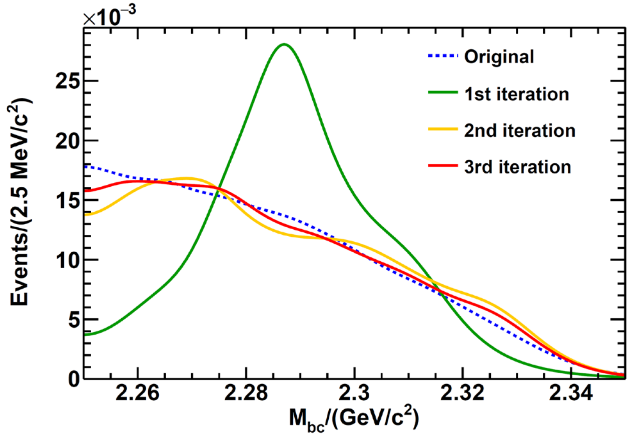
Deep learning solution for $\Lambda_c^+ \to p\pi^0$

- **Dramatic separation! But…**
  - We extract the signal by fitting to Λ_c^+ mass spectrum
  - It's somehow coupled with GNN response
  - Background shape peaks in signal region after veto
  - In jet tagging, such phenomenon is called mass sculpting



- **CMS colleagues have various experiences to prevent that**
  - Transform model response, design adversarial networks, reweight training samples…
  - But none can fit our case well

- **We establish an iterative reweight method**
  - Assign a weight $\omega(M_{BC})$ for each BKG event in loss function
    - $\omega_0(M_{BC}) = 1, \omega_i(M_{BC}) = \omega_{i-1}(M_{BC}) \cdot \frac{p_{i-1}^{BKG}(M_{BC})}{p_{orig}^{BKG}(M_{BC})}$
    - $p_{orig}^{BKG}(M_{BC})$: the original BKG shape before veto
    - $p_{i-1}^{BKG}(M_{BC})$: the BKG shape after veto in $(i-1)$th iteration
  - Repeat the training and iteration would converge when $p_{i-1}^{BKG}(M_{BC}) \simeq p_{orig}^{BKG}(M_{BC})$

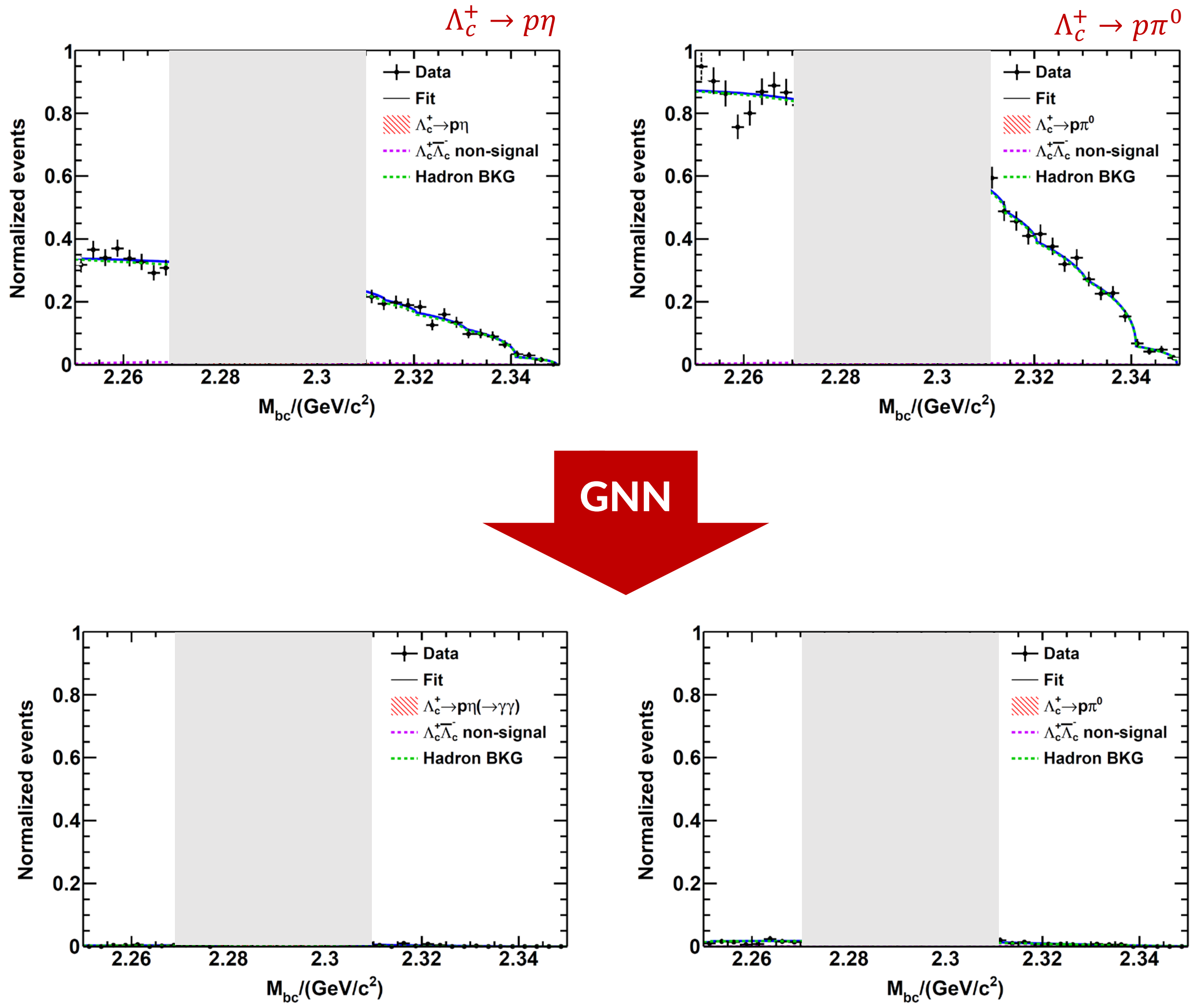


Deep learning solution for $\Lambda_c^+ \to p\pi^0$

- **Finally…**
  - Two orders of magnitude BKG suppression
  - Signal sensitivity greatly improved
  - Conventional ST method needs 5x datasets to achieve such sensitivity

"Boost" physics analysis with less time & budget!

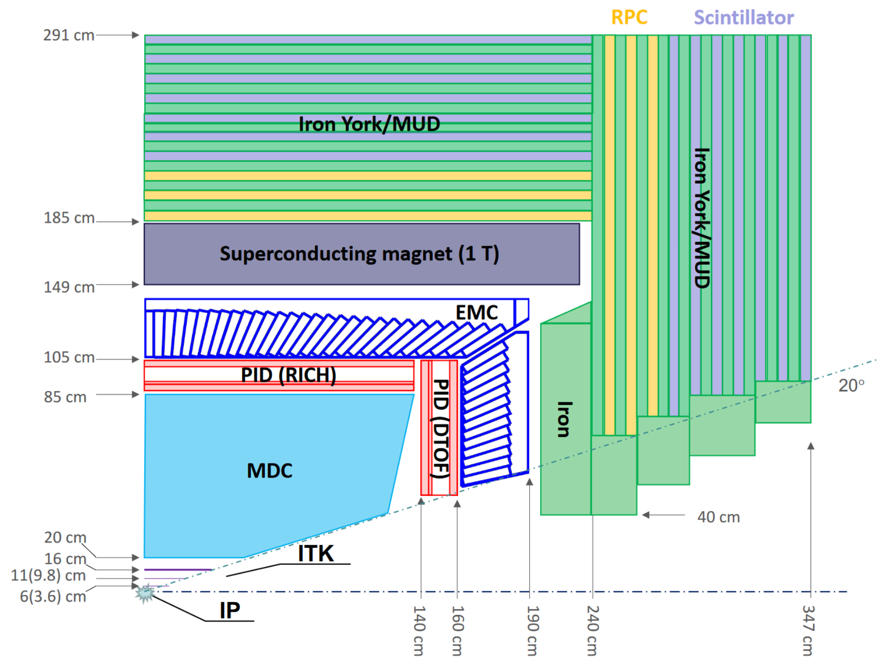
- **The physics result is under internal review.**
- **We can expand this paradigm to more charm studies!**



Outlook

R&D for future collider experiments

- **Neutron hadron reconstruction @ e.g. STCF**
  - EMC maintains BESIII spec, but with faster time response
    - Pure CsI crystal offers $\mathcal{O}(100)$ ps time resolution
  - MUD serves as auxiliary detector for neutral hadron
    - Capture leaked hadrons outside EMC
    - Inner layers use RPC, outer layers use scintillator
  - Complex topology, ideal for ML application

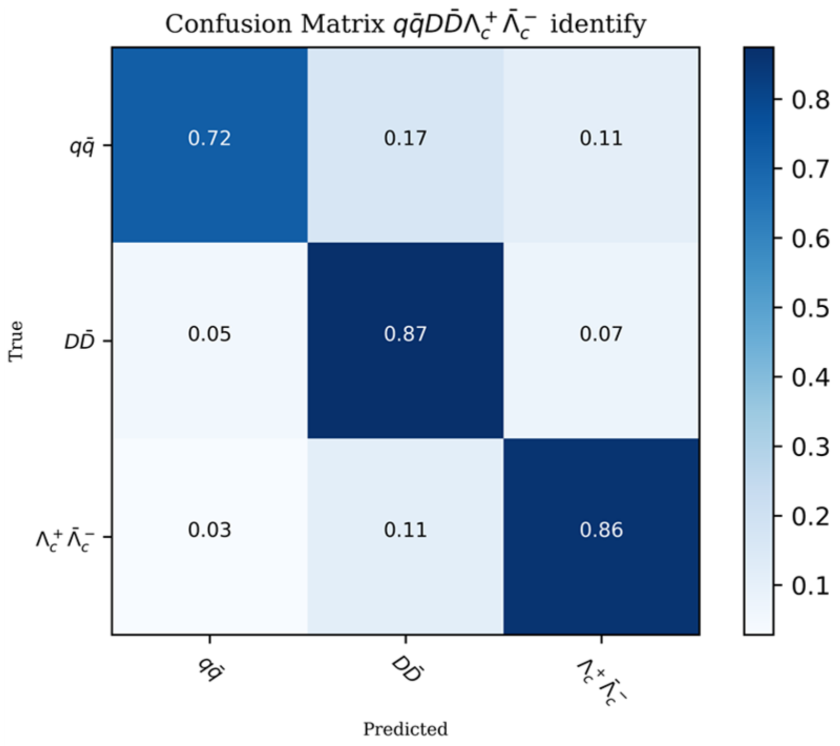


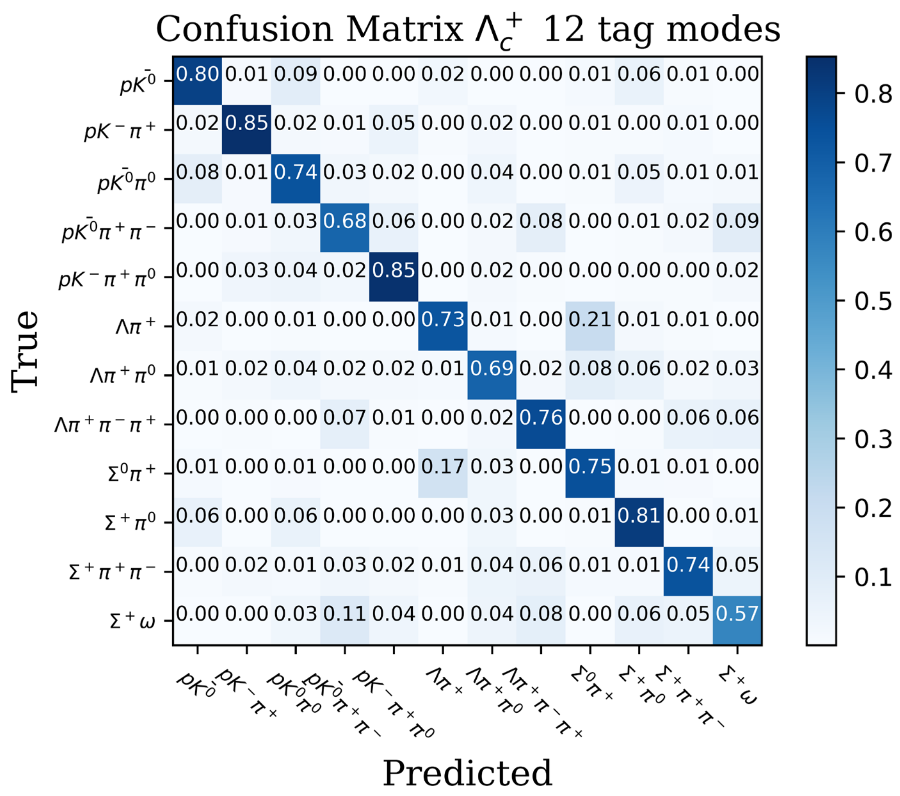
STCF CDR, 2303.15790

- **Charmed hadron tagging @ e.g. STCF**
  - Luminosity ~100x BESIII, data rate ~EB/year
  - High pressure on data storage & offline processing
  - A ML-based software trigger for charmed hadron study?
    - Assign a classification label for events in user-accessible dataset
    - Users can skip the background events to accelerate job running

R&D for future collider experiments

- **Feasibility study of a "software trigger"**
  - Classification between $\Lambda_c^+\bar{\Lambda}_c^-$ & background
  - Classification between Λ_c^+ tag modes





Advanced ML study @ BESIII

- **From analysis tools to real intelligence**
  - Fast simulation
  - Trigger: hardware / software
  - Reconstruction: tracking, PID, vertex, clustering...
  - Analysis: signal identification, likelihood approximation, anomaly detection...
  - Interpretability, uncertainty...

- **BESIII as an ideal laboratory for advanced ML study**
  - Large, labeled, background-free datasets
    - e.g., 10 billion J/ψ events @ BESIII
  - High-quality, fully simulated MC
  - Rich topology
    - Low-level detector response -> particle 4-momentum -> full decay tree
  - Energy-momentum conservation in event
    - Hidden symmetry can inspire new ML structures

- **We welcome ML experts for cooperation!**

Summary

- **Issues in physics analysis motivate our investigation**
  - Direct reconstruction of neutral hadron
  - Inclusive tagging for charmed hadron
- **Powerful deep learning toolkit**
  - Point cloud representation
  - GNN & Transformer-based model architecture
- **Rich physics results in production**
- **Bright future outlooks**
  - R&D for future Tau-Charm Factory
  - Support advanced ML study

Thanks for your attention!