Jet Origin Identification
&
Quantum-based Jet Clustering

朱永峰(PKU) working with 周辰(PKU)、阮曼奇(IHEP) and others

2024.08.15

第十四届全国粒子物理学术会议@青岛

Motivation:

1, Quarks and gluons carry color charge and can not travel freely. Once generated in high-energy collisions, quarks, and gluon would fragment into numerous particles.

2, Deep learning can promote the development of high-energy physics.

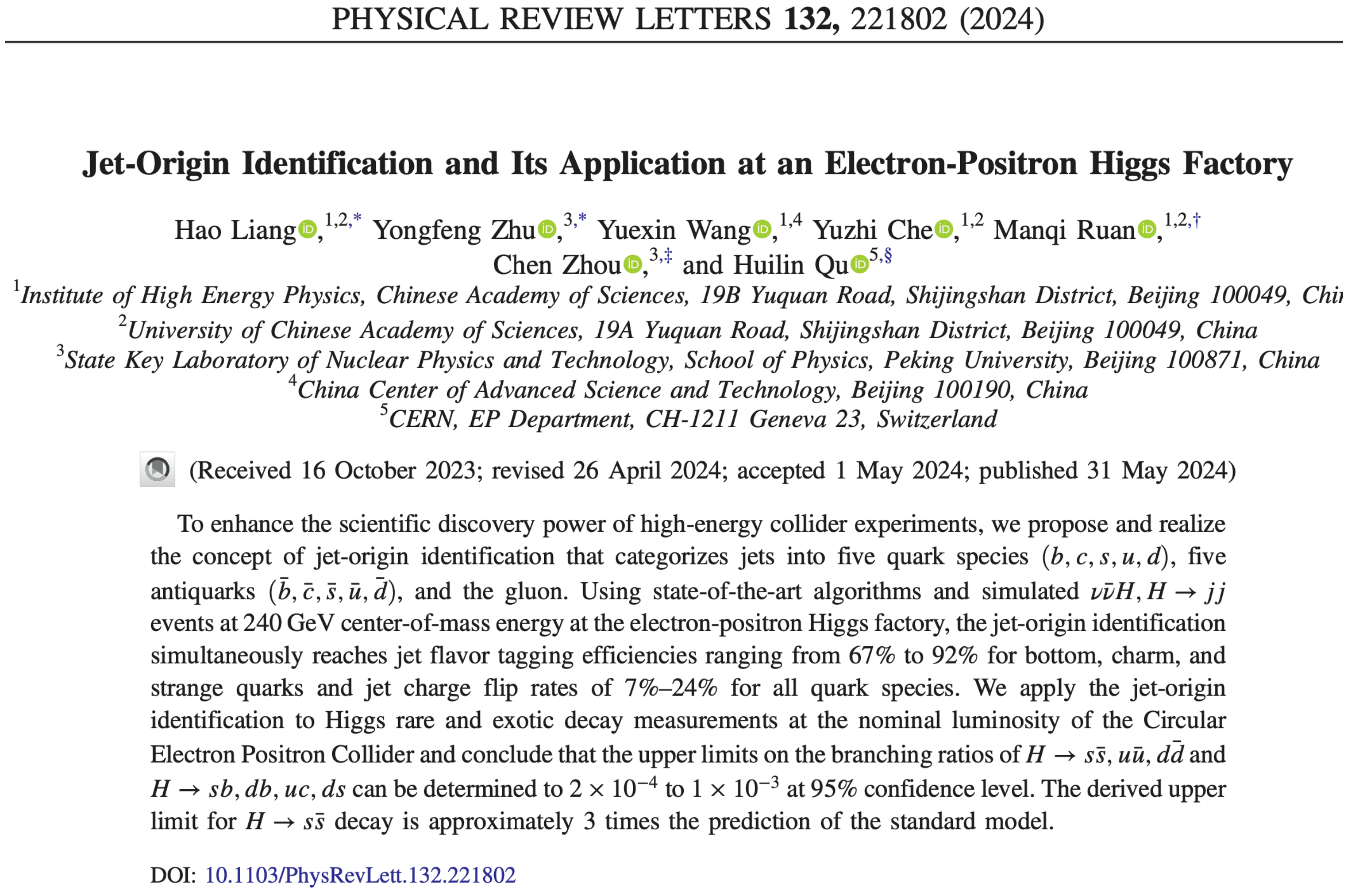
3, Exploring the application of quantum technologies to jet clustering is key to fostering innovation for quantum computing and high-energy physics.

Contents:

1. Jet Origin Identification (JOI)
2. Application of Quantum Approximate Optimization Algorithm (QAOA) to Jet Clustering

Jet Origin Identification

categorizes jets into 5 quarks (b, c, s, u, d), 5 anti-quark ($\bar{b}, \bar{c}, \bar{s}, \bar{u}, \bar{d}$), and gluon
= jet flavor tagging + jet charge measurement + s-quark tagging + gluon finding.



PHYSICAL REVIEW LETTERS **132,** 221802 (2024)

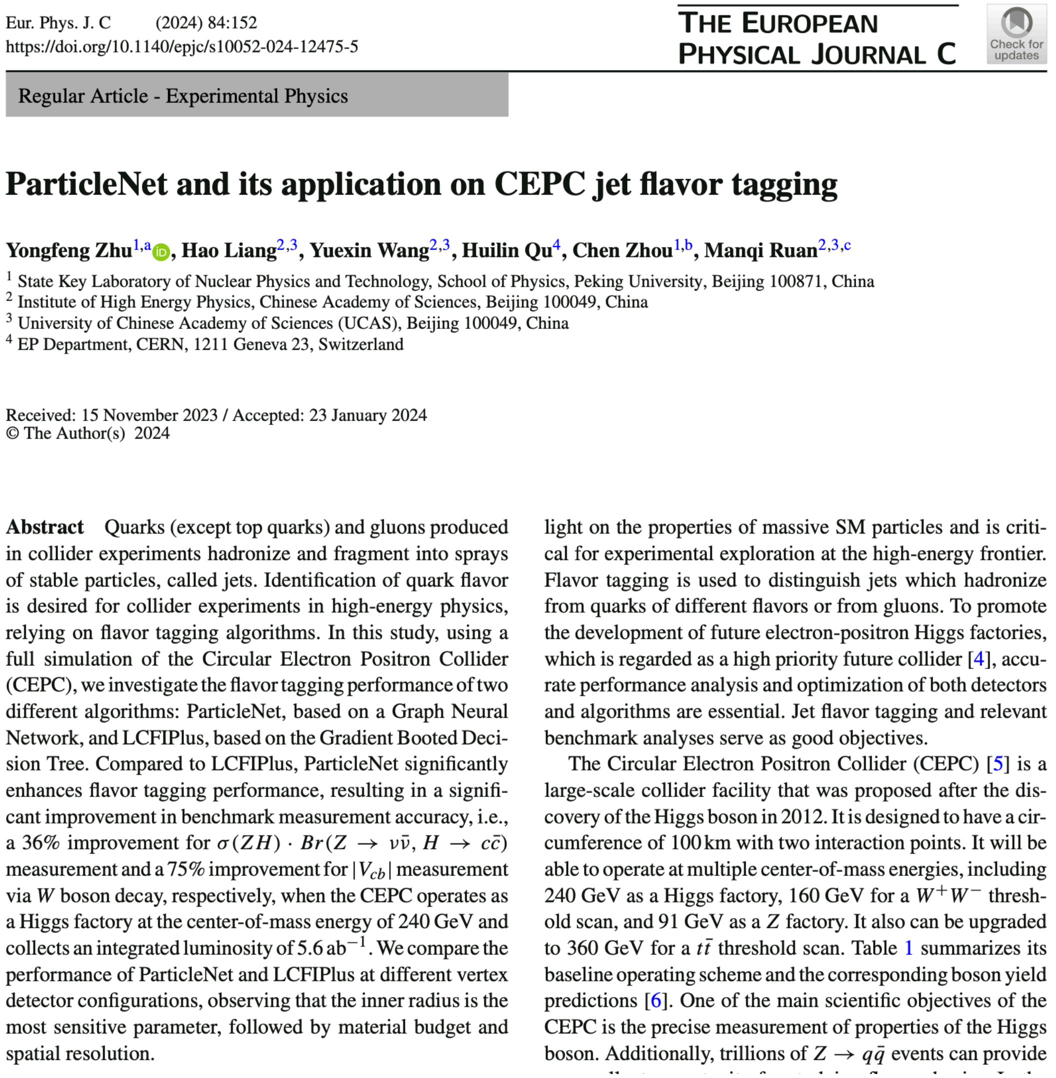
Jet-Origin Identification and Its Application at an Electron-Positron Higgs Factory

Hao Liang,[1,2,*] Yongfeng Zhu,[3,*] Yuexin Wang,[1,4] Yuzhi Che,[1,2] Manqi Ruan,[1,2,†]
Chen Zhou,[3,‡] and Huilin Qu[5,§]
[1]*Institute of High Energy Physics, Chinese Academy of Sciences, 19B Yuquan Road, Shijingshan District, Beijing 100049, Chi*
[2]*University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing 100049, China*
[3]*State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Beijing 100871, China*
[4]*China Center of Advanced Science and Technology, Beijing 100190, China*
[5]*CERN, EP Department, CH-1211 Geneva 23, Switzerland*

(Received 16 October 2023; revised 26 April 2024; accepted 1 May 2024; published 31 May 2024)

To enhance the scientific discovery power of high-energy collider experiments, we propose and realize the concept of jet-origin identification that categorizes jets into five quark species (b, c, s, u, d), five antiquarks $(\bar{b}, \bar{c}, \bar{s}, \bar{u}, \bar{d})$, and the gluon. Using state-of-the-art algorithms and simulated $\nu\bar{\nu}H, H \to jj$ events at 240 GeV center-of-mass energy at the electron-positron Higgs factory, the jet-origin identification simultaneously reaches jet flavor tagging efficiencies ranging from 67% to 92% for bottom, charm, and strange quarks and jet charge flip rates of 7%–24% for all quark species. We apply the jet-origin identification to Higgs rare and exotic decay measurements at the nominal luminosity of the Circular Electron Positron Collider and conclude that the upper limits on the branching ratios of $H \to s\bar{s}, u\bar{u}, d\bar{d}$ and $H \to sb, db, uc, ds$ can be determined to 2×10^{-4} to 1×10^{-3} at 95% confidence level. The derived upper limit for $H \to s\bar{s}$ decay is approximately 3 times the prediction of the standard model.

DOI: 10.1103/PhysRevLett.132.221802



Eur. Phys. J. C (2024) 84:152
https://doi.org/10.1140/epjc/s10052-024-12475-5

THE EUROPEAN
PHYSICAL JOURNAL C

Regular Article - Experimental Physics

ParticleNet and its application on CEPC jet flavor tagging

Yongfeng Zhu[1,a], **Hao Liang**[2,3], **Yuexin Wang**[2,3], **Huilin Qu**[4], **Chen Zhou**[1,b], **Manqi Ruan**[2,3,c]
[1] State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Beijing 100871, China
[2] Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China
[3] University of Chinese Academy of Sciences (UCAS), Beijing 100049, China
[4] EP Department, CERN, 1211 Geneva 23, Switzerland

Received: 15 November 2023 / Accepted: 23 January 2024

Abstract Quarks (except top quarks) and gluons produced in collider experiments hadronize and fragment into sprays of stable particles, called jets. Identification of quark flavor is desired for collider experiments in high-energy physics, relying on flavor tagging algorithms. In this study, using a full simulation of the Circular Electron Positron Collider (CEPC), we investigate the flavor tagging performance of two different algorithms: ParticleNet, based on a Graph Neural Network, and LCFIPlus, based on the Gradient Booted Decision Tree. Compared to LCFIPlus, ParticleNet significantly enhances flavor tagging performance, resulting in a significant improvement in benchmark measurement accuracy, i.e., a 36% improvement for $\sigma(ZH) \cdot Br(Z \to \nu\bar{\nu}, H \to c\bar{c})$ measurement and a 75% improvement for $|V_{cb}|$ measurement via W boson decay, respectively, when the CEPC operates as a Higgs factory at the center-of-mass energy of 240 GeV and collects an integrated luminosity of $5.6\,\text{ab}^{-1}$. We compare the performance of ParticleNet and LCFIPlus at different vertex detector configurations, observing that the inner radius is the most sensitive parameter, followed by material budget and spatial resolution.

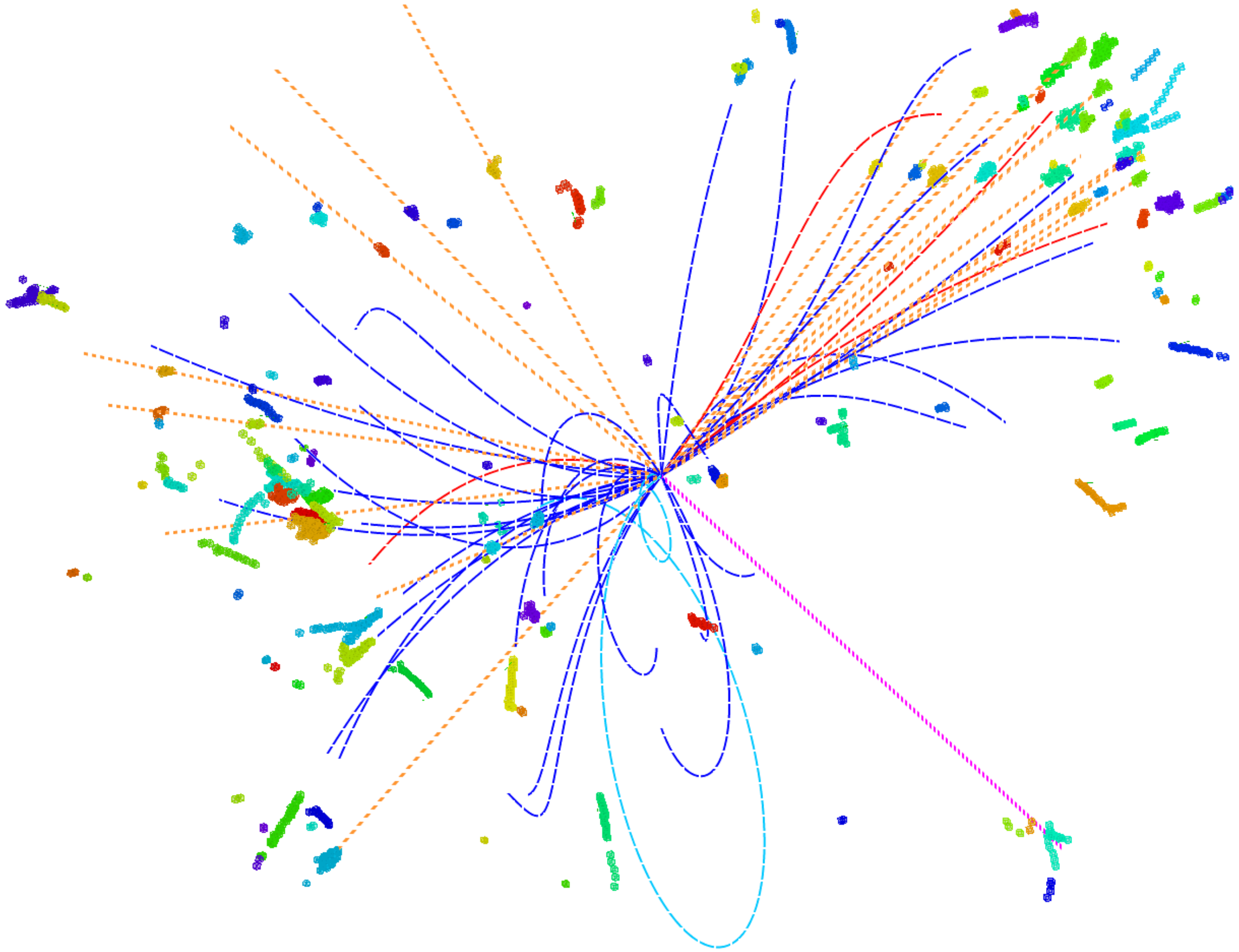
light on the properties of massive SM particles and is critical for experimental exploration at the high-energy frontier. Flavor tagging is used to distinguish jets which hadronize from quarks of different flavors or from gluons. To promote the development of future electron-positron Higgs factories, which is regarded as a high priority future collider [4], accurate performance analysis and optimization of both detectors and algorithms are essential. Jet flavor tagging and relevant benchmark analyses serve as good objectives.

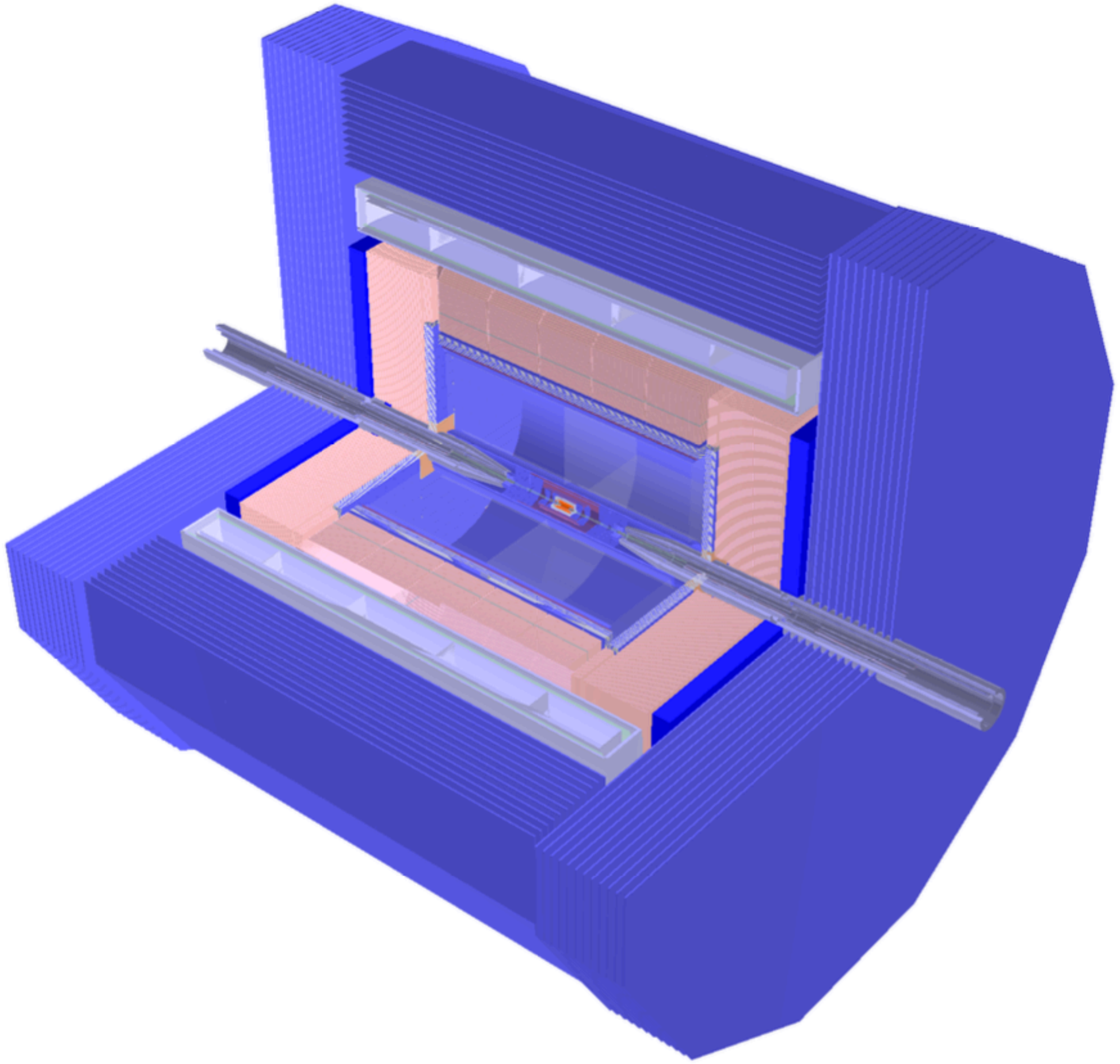
The Circular Electron Positron Collider (CEPC) [5] is a large-scale collider facility that was proposed after the discovery of the Higgs boson in 2012. It is designed to have a circumference of 100 km with two interaction points. It will be able to operate at multiple center-of-mass energies, including 240 GeV as a Higgs factory, 160 GeV for a W^+W^- threshold scan, and 91 GeV as a Z factory. It also can be upgraded to 360 GeV for a $t\bar{t}$ threshold scan. Table 1 summarizes its baseline operating scheme and the corresponding boson yield predictions [6]. One of the main scientific objectives of the CEPC is the precise measurement of properties of the Higgs boson. Additionally, trillions of $Z \to q\bar{q}$ events can provide

PhysRevLett.132.221802

Eur. Phys. J. C (2024) 84:152

Samples

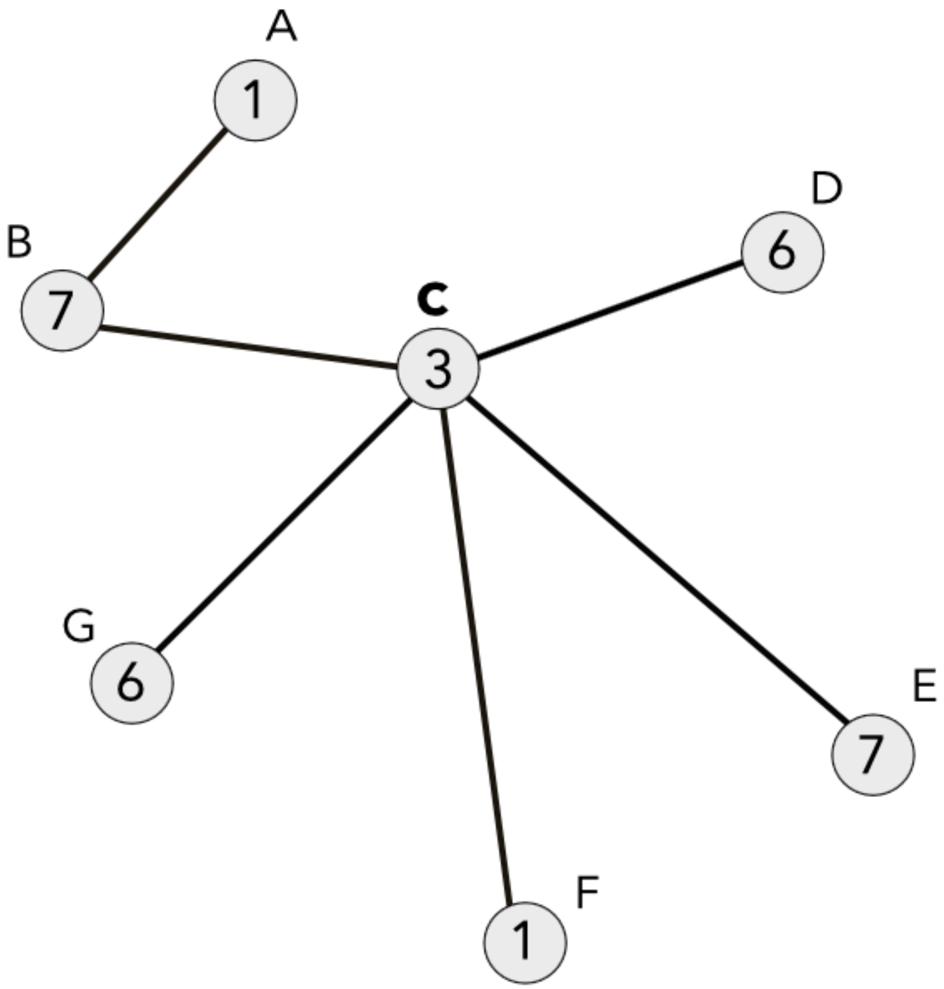




Full Simulated vvH, Higgs to two jets sample at CEPC baseline configuration: CEPC-v4 detector,

1 Million samples each, 60/20/20% for training, validation & test

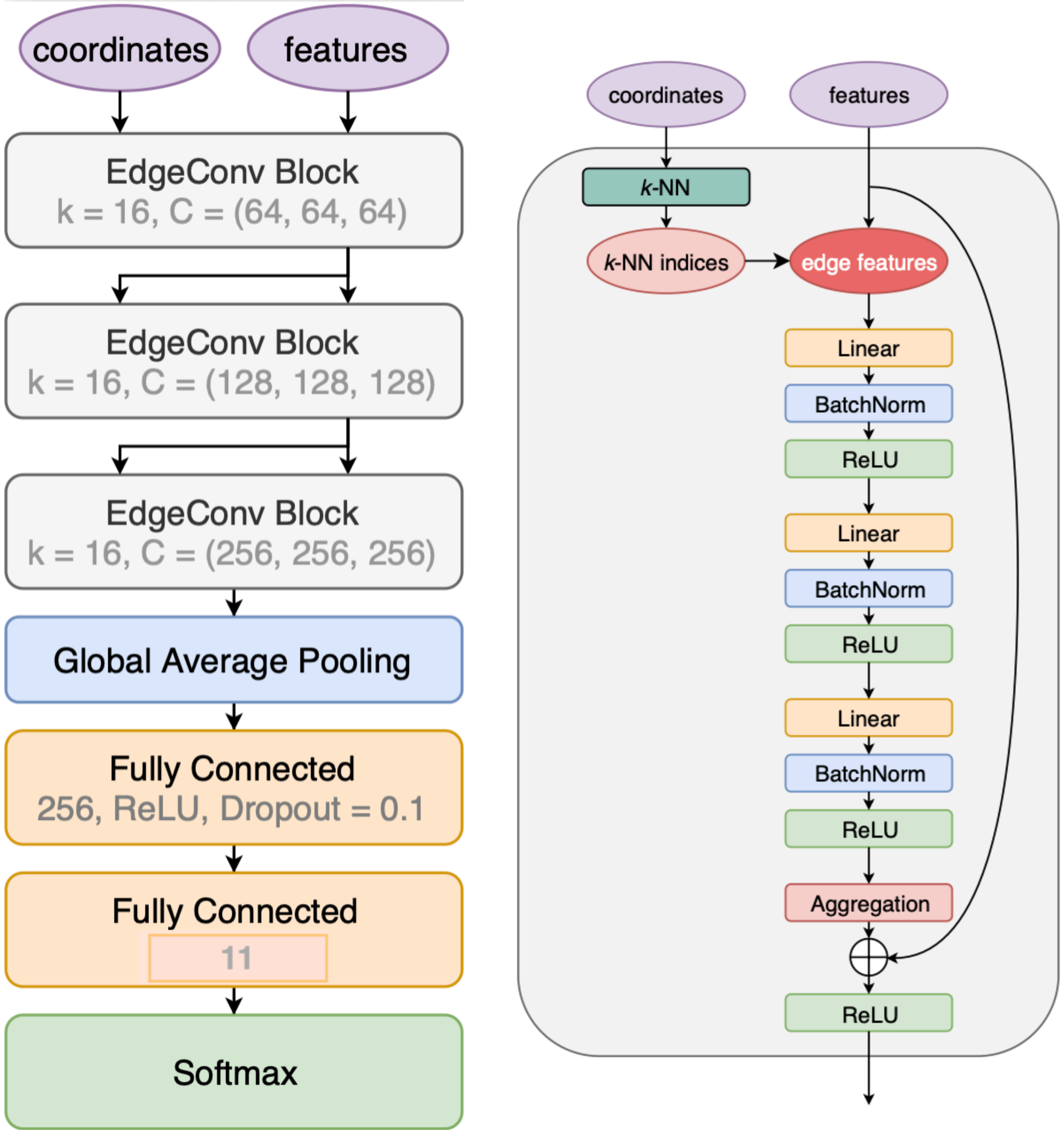
jet represented as :



Graph = (Vertices, Edges)

Construct GNN
with “message passing”

the ParticleNet and input features



Phys. Rev. D 101, 056019

input features :

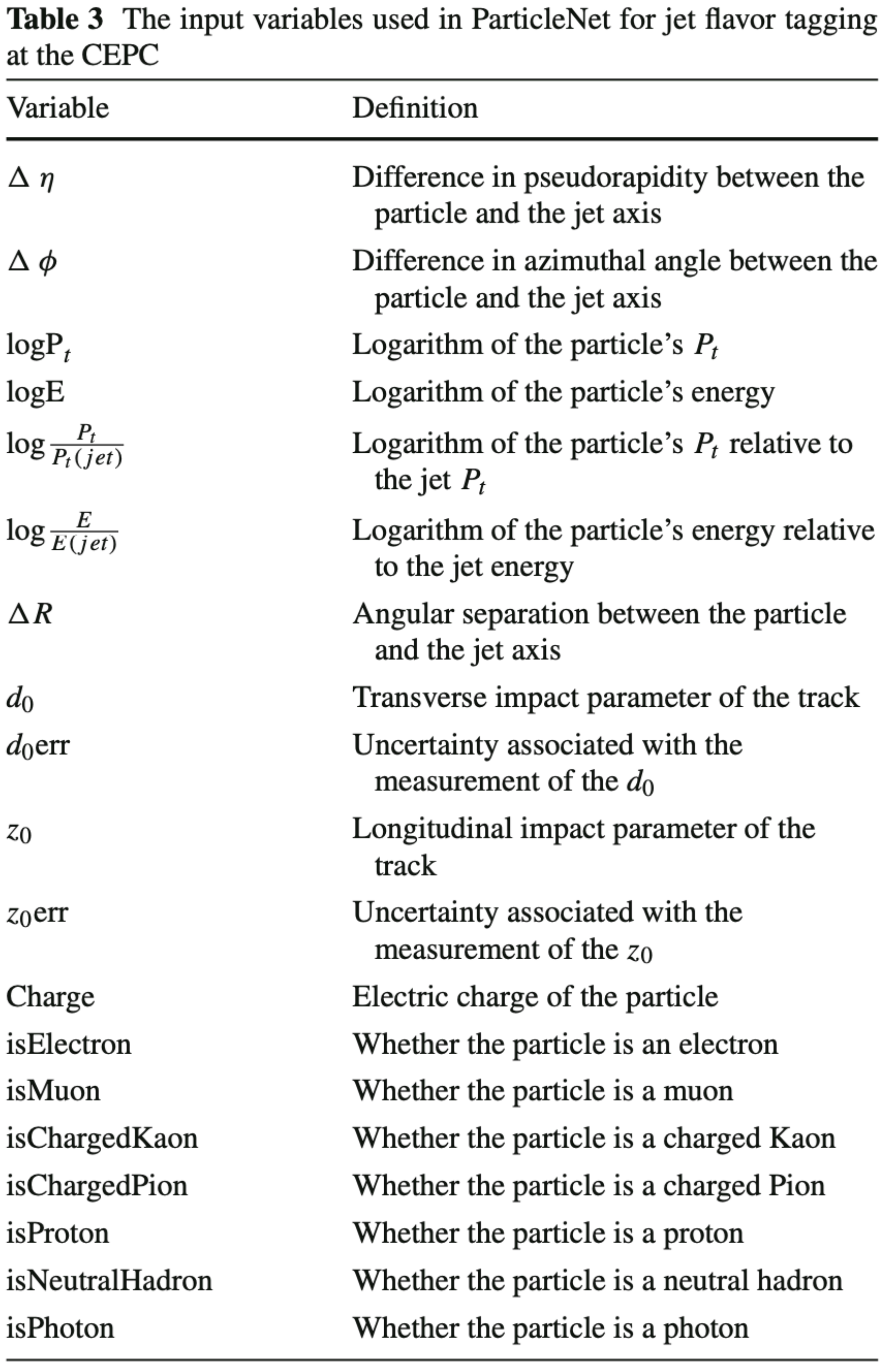
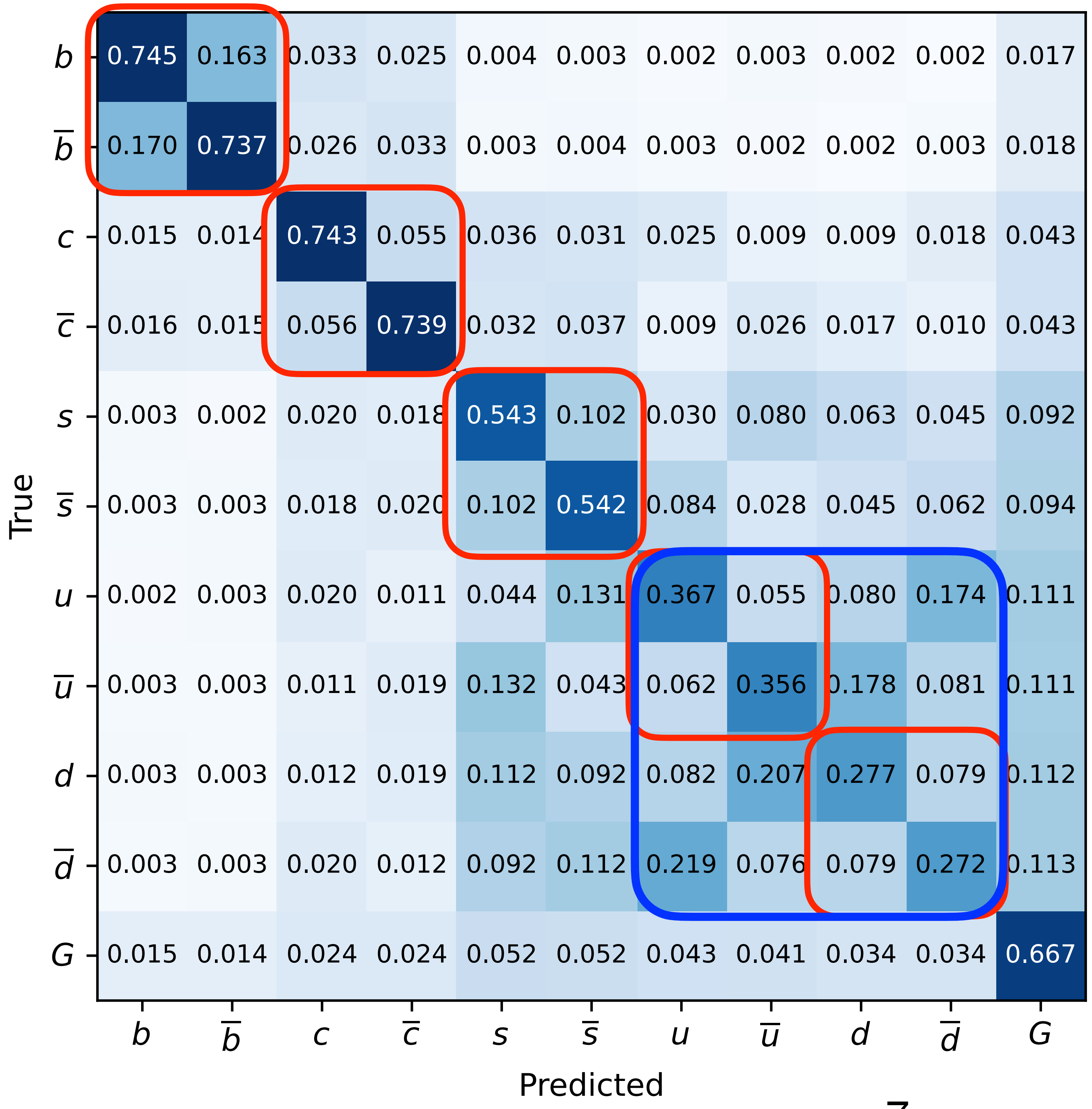


Table 3 The input variables used in ParticleNet for jet flavor tagging at the CEPC

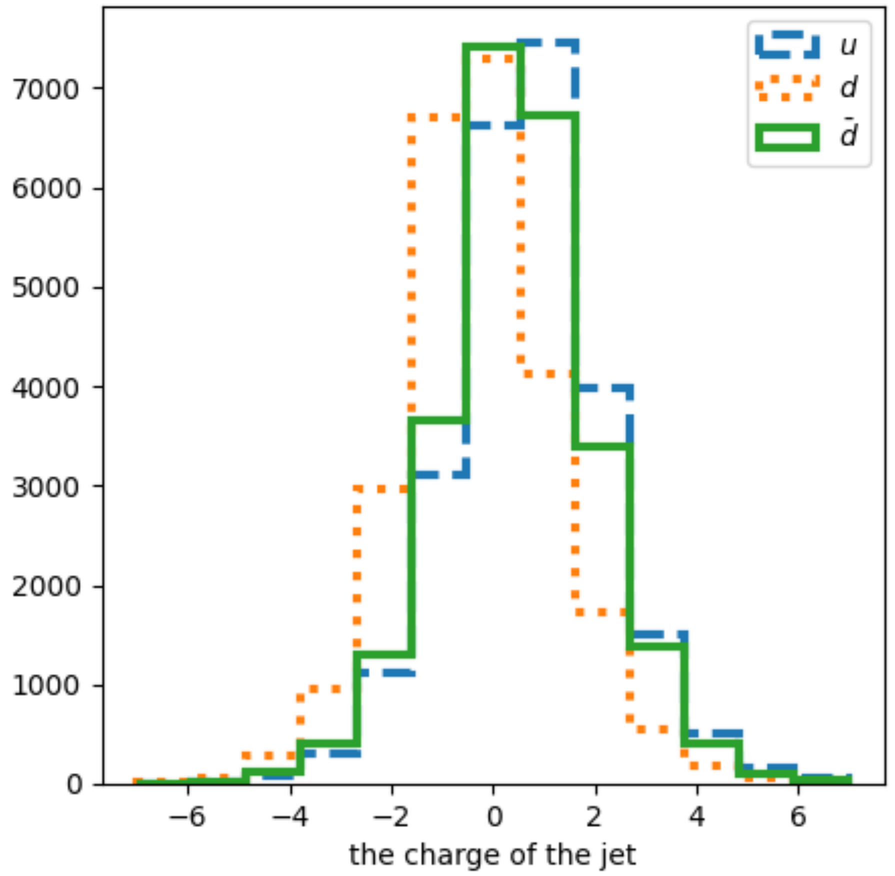
Variable	Definition
$\Delta\eta$	Difference in pseudorapidity between the particle and the jet axis
$\Delta\phi$	Difference in azimuthal angle between the particle and the jet axis
$\log P_t$	Logarithm of the particle's P_t
logE	Logarithm of the particle's energy
$\log\frac{P_t}{P_t(jet)}$	Logarithm of the particle's P_t relative to the jet P_t
$\log\frac{E}{E(jet)}$	Logarithm of the particle's energy relative to the jet energy
ΔR	Angular separation between the particle and the jet axis
d_0	Transverse impact parameter of the track
$d_0$err	Uncertainty associated with the measurement of the d_0
z_0	Longitudinal impact parameter of the track
$z_0$err	Uncertainty associated with the measurement of the z_0
Charge	Electric charge of the particle
isElectron	Whether the particle is an electron
isMuon	Whether the particle is a muon
isChargedKaon	Whether the particle is a charged Kaon
isChargedPion	Whether the particle is a charged Pion
isProton	Whether the particle is a proton
isNeutralHadron	Whether the particle is a neutral hadron
isPhoton	Whether the particle is a photon

the performance of jet origin identification

ParticleNet algorithm attaches each jet with 11 likelihoods corresponding to 11 types of jets. Then the jet type is determined according to the maximum likelihood.



True \ Predicted	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$	G
b	0.745	0.163	0.033	0.025	0.004	0.003	0.002	0.003	0.002	0.002	0.017
$\bar{b}$	0.170	0.737	0.026	0.033	0.003	0.004	0.003	0.002	0.002	0.003	0.018
c	0.015	0.014	0.743	0.055	0.036	0.031	0.025	0.009	0.009	0.018	0.043
$\bar{c}$	0.016	0.015	0.056	0.739	0.032	0.037	0.009	0.026	0.017	0.010	0.043
s	0.003	0.002	0.020	0.018	0.543	0.102	0.030	0.080	0.063	0.045	0.092
$\bar{s}$	0.003	0.003	0.018	0.020	0.102	0.542	0.084	0.028	0.045	0.062	0.094
u	0.002	0.003	0.020	0.011	0.044	0.131	0.367	0.055	0.080	0.174	0.111
$\bar{u}$	0.003	0.003	0.011	0.019	0.132	0.043	0.062	0.356	0.178	0.081	0.111
d	0.003	0.003	0.012	0.019	0.112	0.092	0.082	0.207	0.277	0.079	0.112
$\bar{d}$	0.003	0.003	0.020	0.012	0.092	0.112	0.219	0.076	0.079	0.272	0.113
G	0.015	0.014	0.024	0.024	0.052	0.052	0.043	0.041	0.034	0.034	0.667



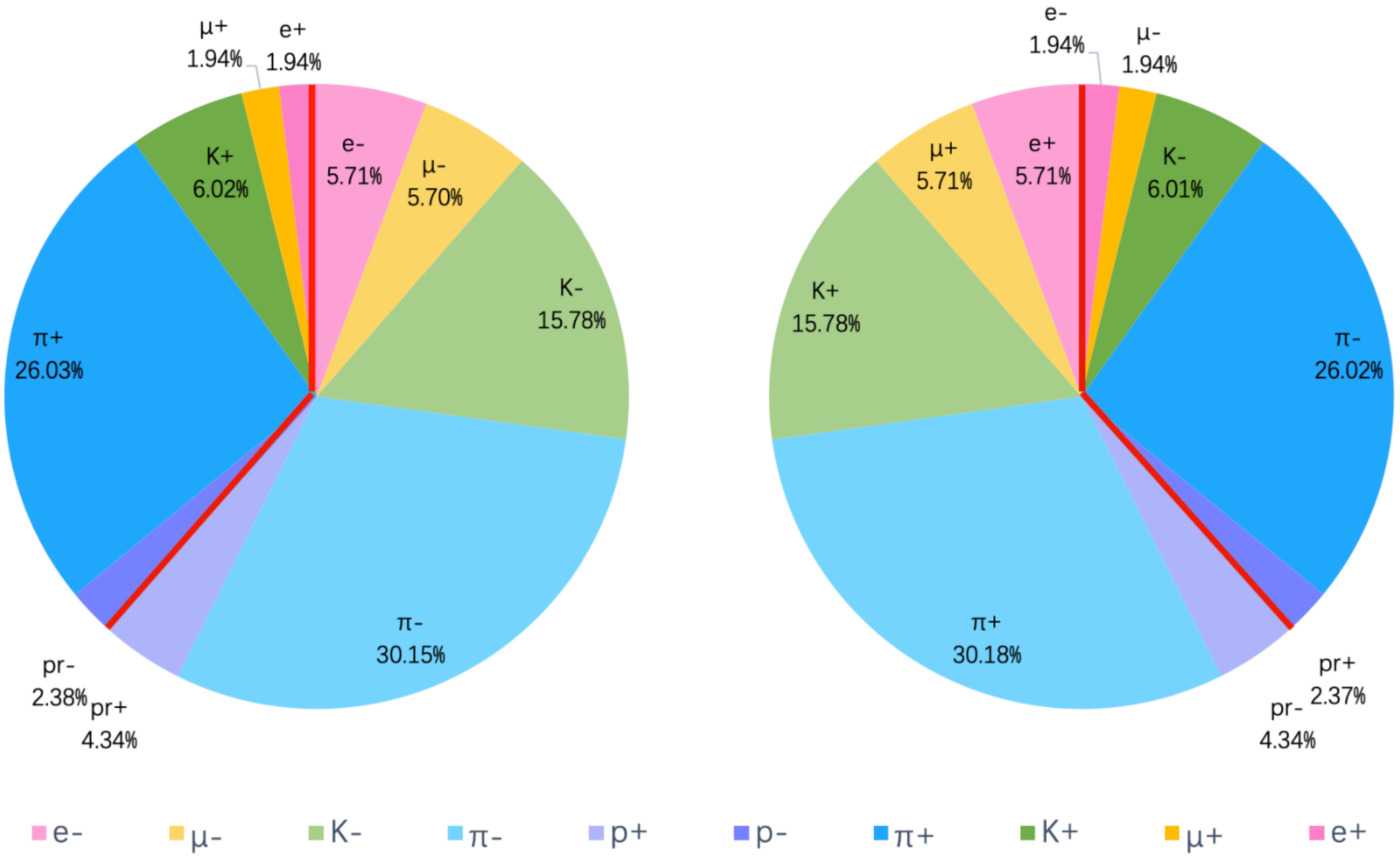
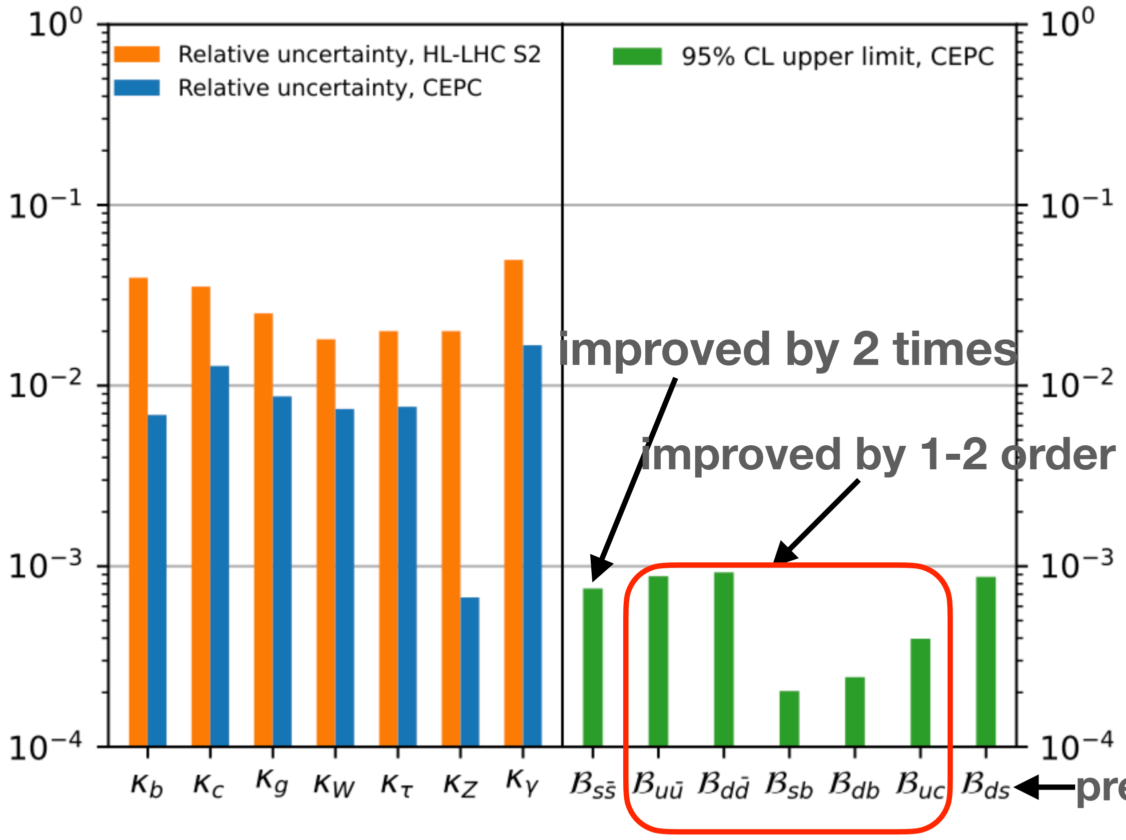


Figure 5. The percentages of species of final state leading charged particles within the b jet (left) and the $\bar{b}$ jet (right) by WHIZARD 1.95.

arXiv:2306.14089

Benchmark physics analyses

Begin with the existing analyses of $\nu\bar{\nu}H, H \to b\bar{b}/c\bar{c}/gg$, (arXiv:2203.01469) and combining the jet origin identification, we obtain the upper limits on branching ratios of seven Higgs rare and FCNC hadronic decay modes.



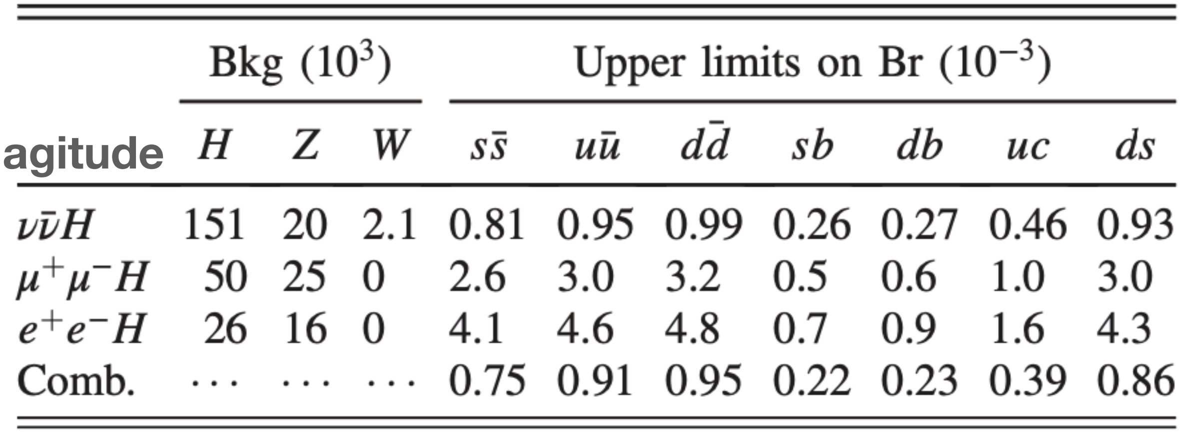
improved by 2 times

improved by 1-2 order of magnitude

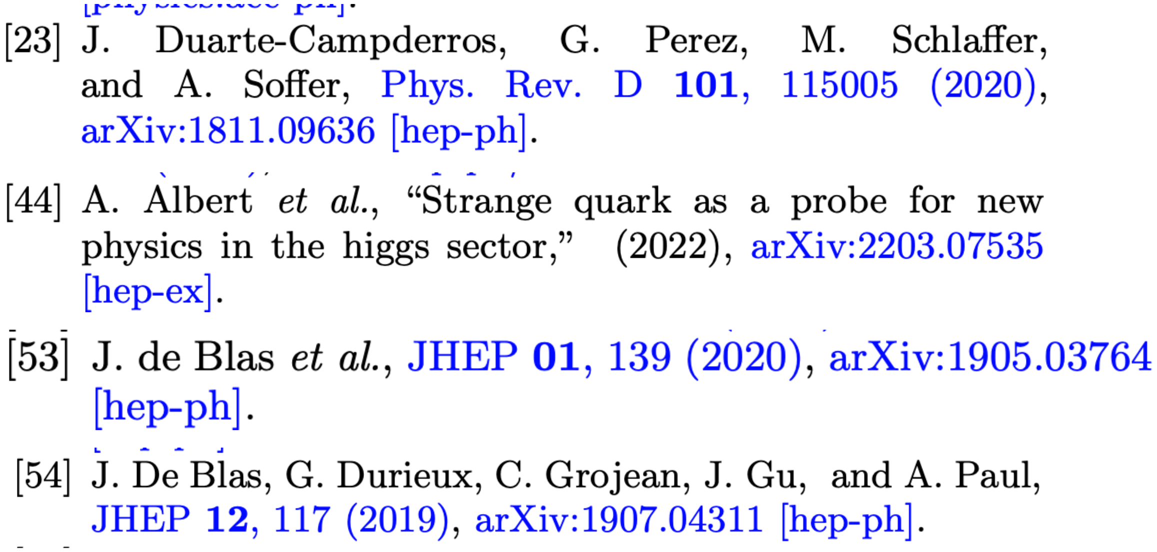
presumably first quantified

FIG. 5. Expected upper limits on the branching ratios of rare Higgs boson decays from this Letter (green) and the relative uncertainties of Higgs couplings anticipated at CEPC [19] (blue) and HL-LHC [43] (orange) under the kappa-0 fit scenario [54] and scenario S2 of systematics [55], as cited in Ref. [19]. The limit on $B_{s\bar{s}}$ corresponds to an upper limit of 1.7 on the Higgs-strange coupling modifier κ_s (not shown).

TABLE I. Summary of background yields from $H \to b\bar{b}/c\bar{c}/gg$, Z, and W prior to the flavor-based event selection, along with the expected upper limits on Higgs decay branching ratios at 95% CL under the background-only hypothesis.



	Bkg (10^3)			Upper limits on Br (10^{-3})						
	H	Z	W	$s\bar{s}$	$u\bar{u}$	$d\bar{d}$	sb	db	uc	ds
$\nu\bar{\nu}H$	151	20	2.1	0.81	0.95	0.99	0.26	0.27	0.46	0.93
$\mu^+\mu^-H$	50	25	0	2.6	3.0	3.2	0.5	0.6	1.0	3.0
e^+e^-H	26	16	0	4.1	4.6	4.8	0.7	0.9	1.6	4.3
Comb.	⋯	⋯	⋯	0.75	0.91	0.95	0.22	0.23	0.39	0.86



[23] J. Duarte-Campderros, G. Perez, M. Schlaffer, and A. Soffer, Phys. Rev. D **101**, 115005 (2020), arXiv:1811.09636 [hep-ph].

[44] A. Albert *et al.*, "Strange quark as a probe for new physics in the higgs sector," (2022), arXiv:2203.07535 [hep-ex].

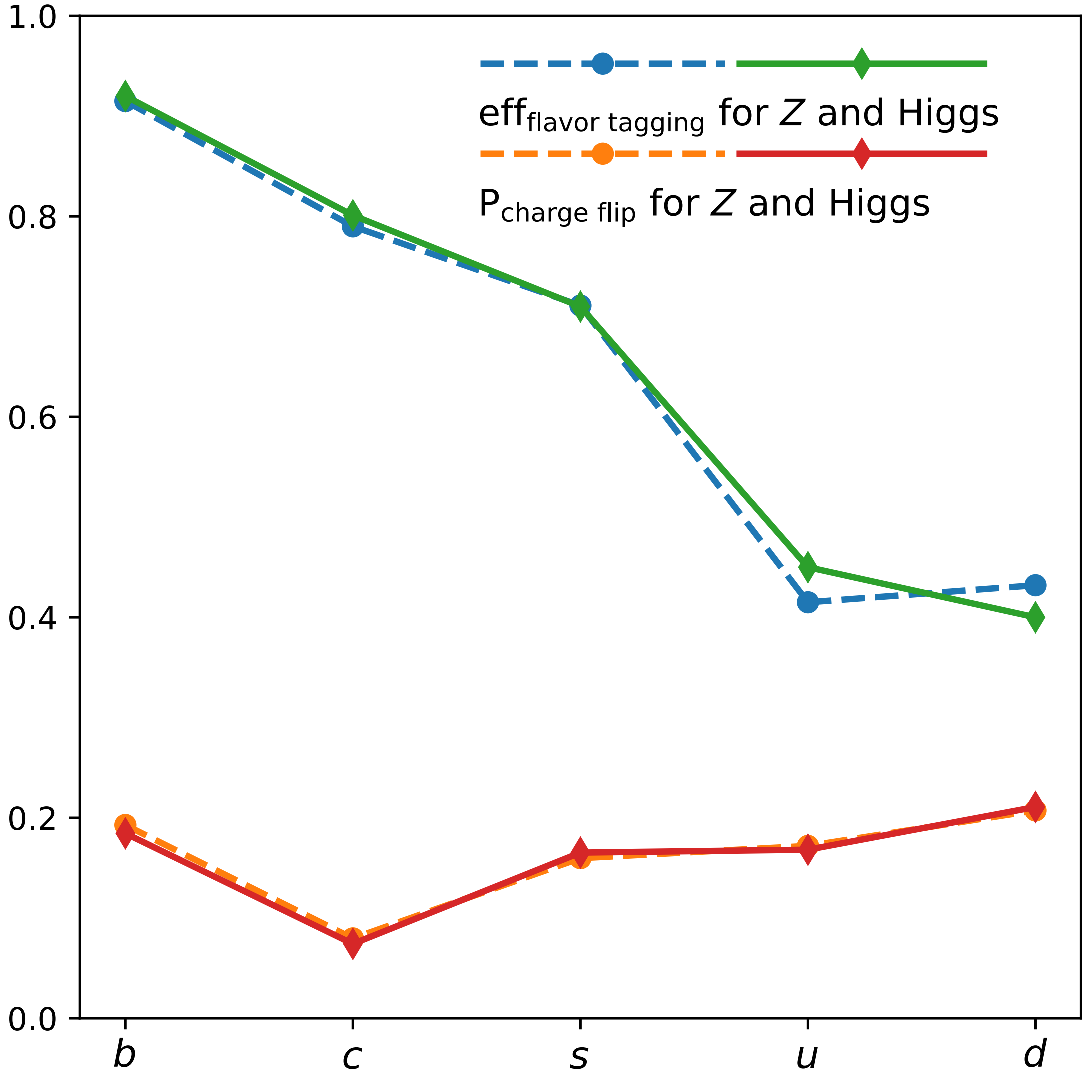
[53] J. de Blas *et al.*, JHEP **01**, 139 (2020), arXiv:1905.03764 [hep-ph].

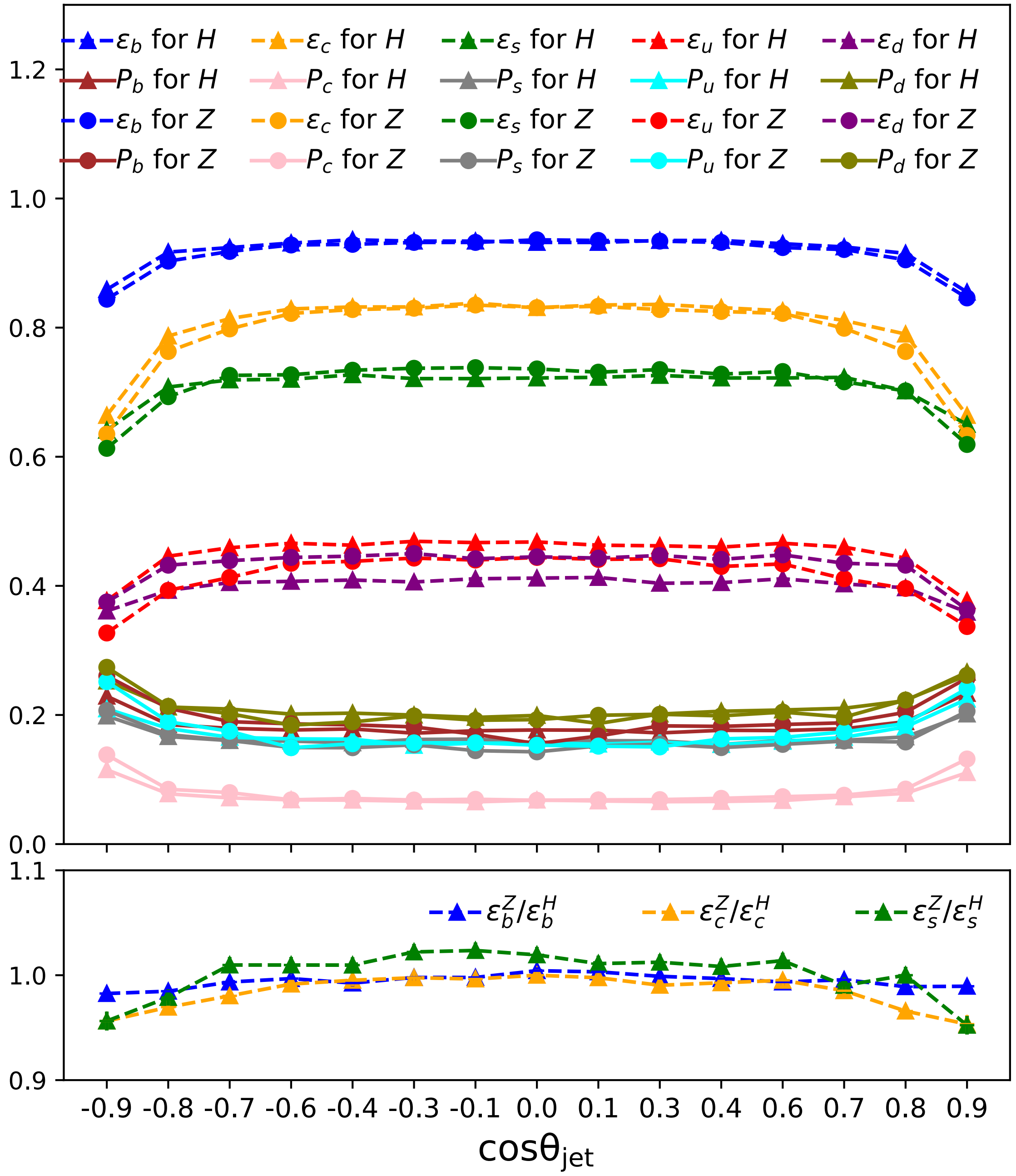
[54] J. De Blas, G. Durieux, C. Grojean, J. Gu, and A. Paul, JHEP **12**, 117 (2019), arXiv:1907.04311 [hep-ph].

Comparison between different physics processes

$Z \to q\bar{q}$ at $\sqrt{s} = 91.2\ GeV$

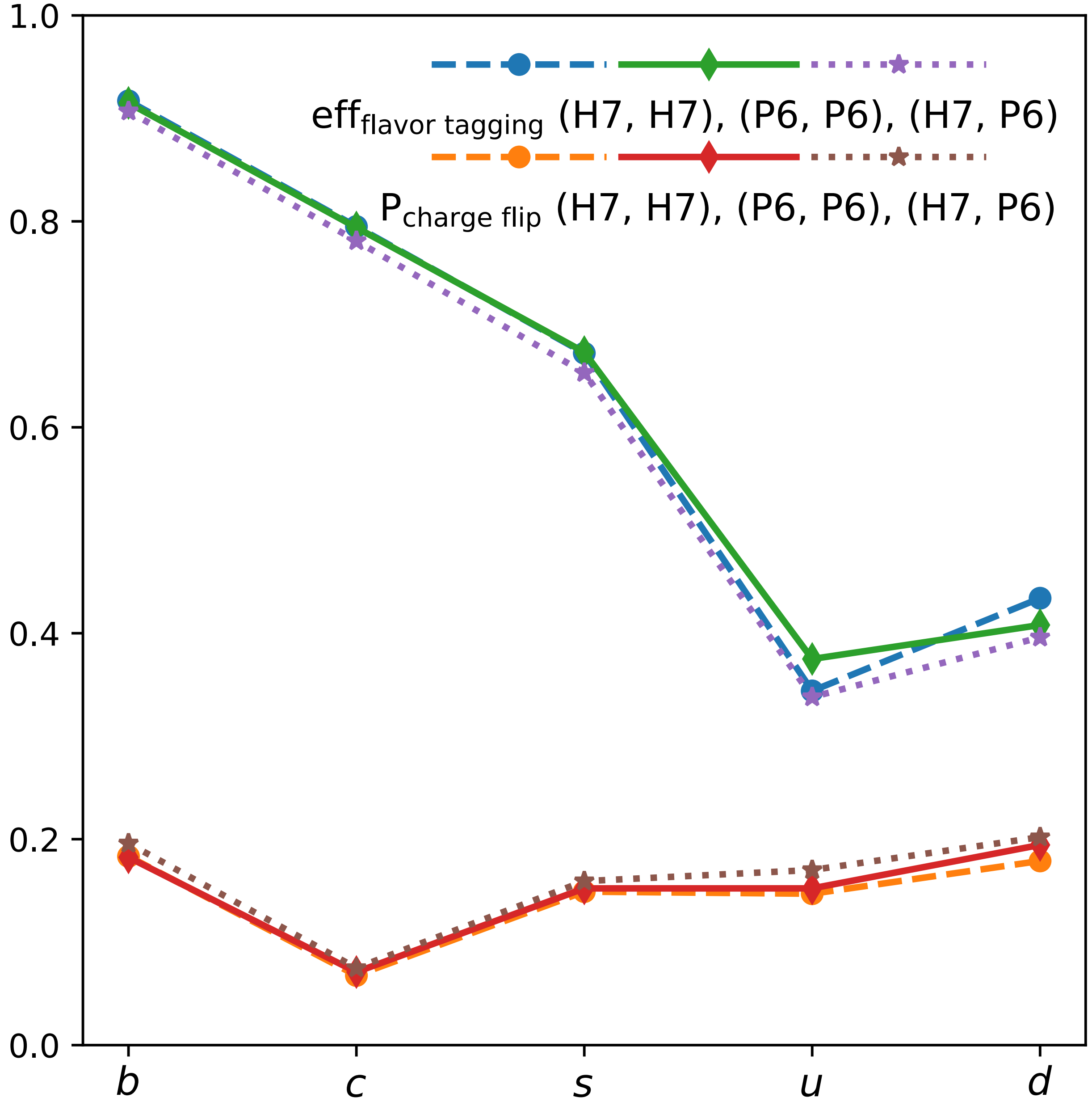
$\nu\bar{\nu}H, H \to q\bar{q}$ at $\sqrt{s} = 240\ GeV$





The jet origin identification performance agrees with each other, especially in the fiducial barrel region of the detector for the flavor tagging performance of b, c, and s.

Comparison between different hadronization models



Pythia-6.4
Herwig-7.2.2

$\nu\bar{\nu}H, H \to jj$ at $\sqrt{s} = 240\ GeV$

(A, B) means: training on A, test on B

The jet origin identificaion performanc agrees with each other, especially for b, c, and s jets, while exhibits small but visible differences for u and d jets.

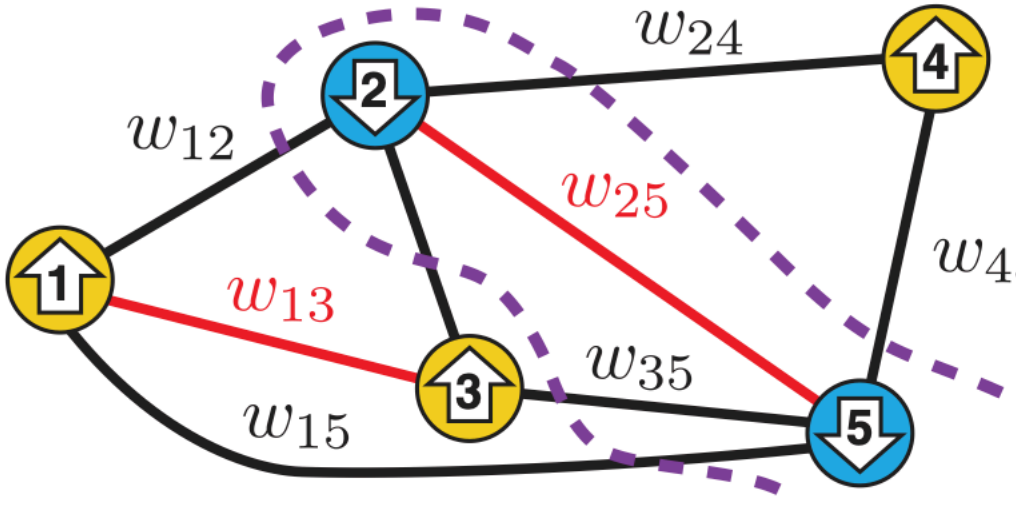
A Novel Quantum Realization of Jet Clustering in High-Energy Physics Experiments

Yongfeng Zhu[a], WeiFeng Zhuang[b], Chen Qian[b], Yunheng Ma[b], Dong E. Liu[b,c,*], Manqi Ruan[d,e,*], Chen Zhou[a,*]

[a] *State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Haidian District, Beijing, 100871, China*
[b] *Beijing Academy of Quantum Information Sciences, Haidian District, Beijing, 100193, China*
[c] *Department of Physics, Tsinghua University, Haidian District, Beijing, 100084, China*
[d] *Institute of High Energy Physics, Chinese Academy of Sciences, Shijingshan District, Beijing, 100049, China*
[e] *University of Chinese Academy of Sciences, Shijingshan District, Beijing, 100049, China*

arXiv:2407.09056

MaxCut problem



Phys. Rev. X 10, 021067

undirected graph $G = (V, E)$

V: set of vertices

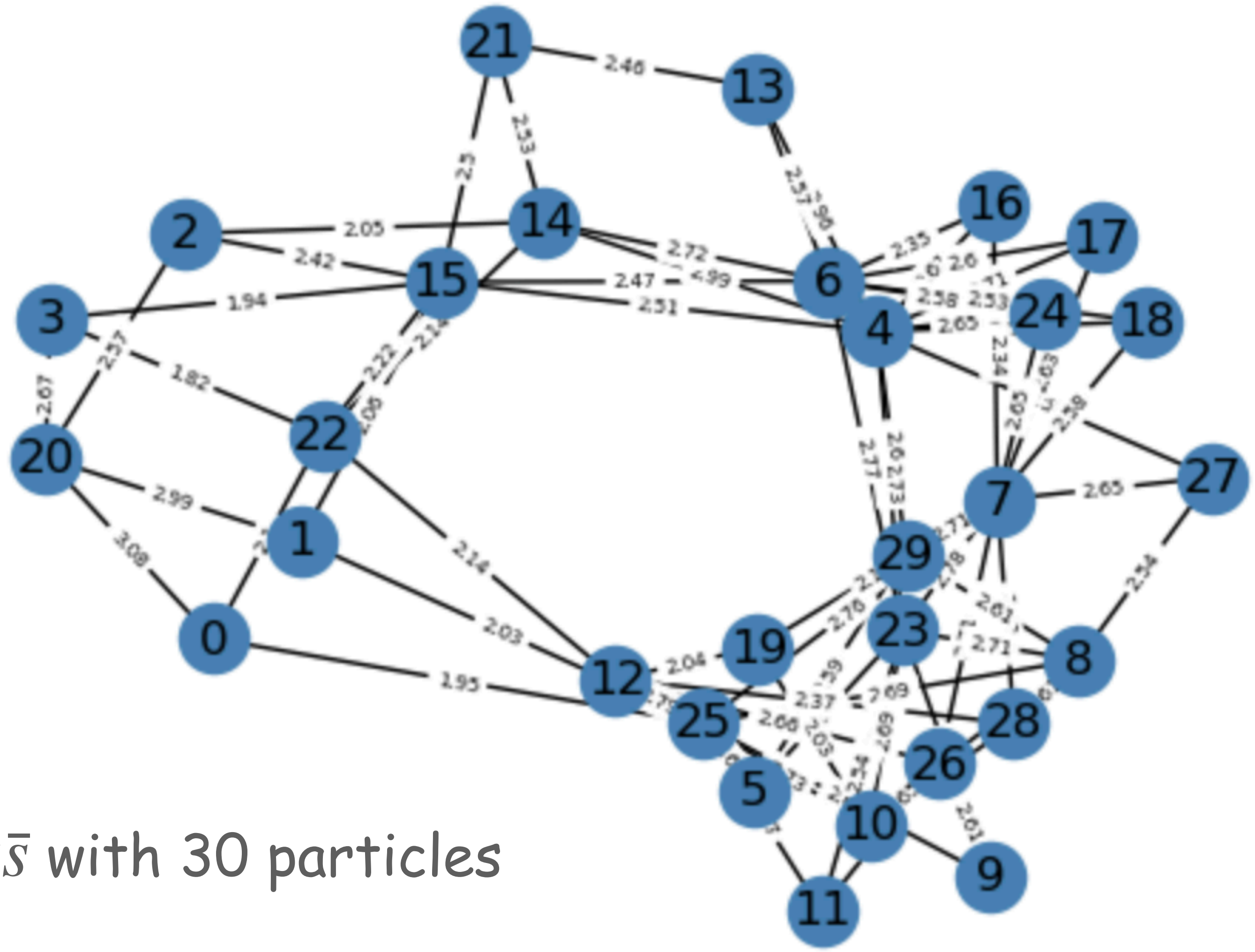
E: set of edges

$w_{ij} > 0$: the weight of the edge $(ij) \in E$

goal: partition the graph vertices into two complementary subsets to maximize the total weight of edges with two vertices belong to two subsets $C(x) = \sum_{i,j=1}^{|V|} w_{ij} x_i (1 - x_j)$

A collision event can be represented as a graph

particles as vertices
the angle of two particles as the edge weight
only the k leading large edges are retained (k-regular graph)



$e^+e^- \to ZH \to \nu\bar{\nu}s\bar{s}$ with 30 particles

the process of AQC :

1, define the Hamiltonian: $H(t) = (1 - s(t))H_D + s(t)H_C$ and let our quantum system evolve under it, $U(t) = \tau e^{\frac{-i}{h}\int_o^t H(T)dT}$.

2, We discretize $U(T)$ into intervals Δt small enough that the Hamiltonian is approximately constant over each interval.

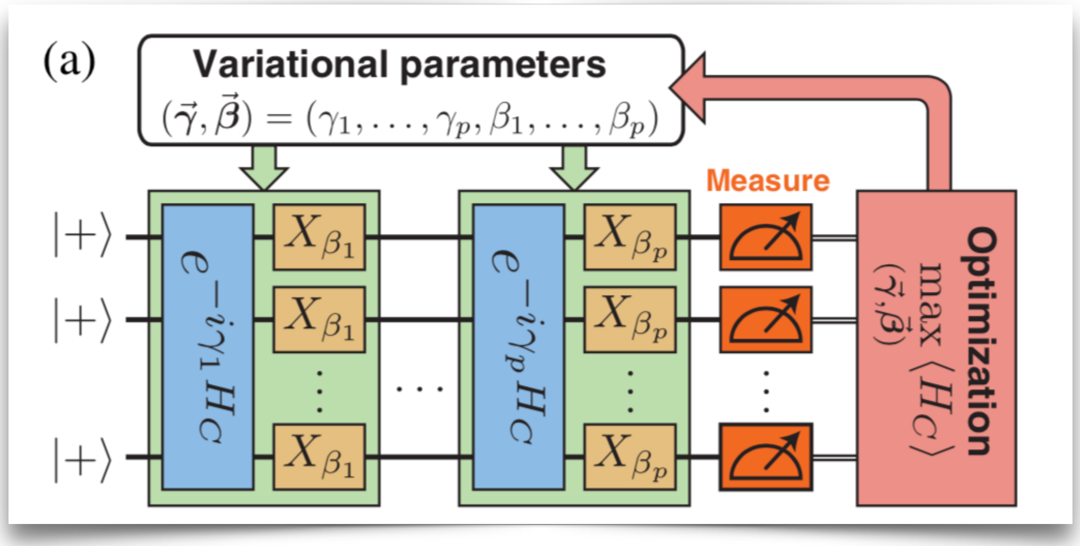
3, Let $U(b, a)$ represent time evolution from time a to time b

$$U(T,0) = U(T, T - \Delta t)U(T - \Delta t, T - 2\Delta t)\dots U(\Delta t,0) = \prod_{j=1}^{P} U(j\Delta t, (j-1)\Delta t) \approx \prod_{j=1}^{P} e^{-iH(j\Delta t)\Delta t}$$

$$U(T,0) \approx \prod_{j=1}^{P} e^{-i(1-s(j\Delta t))H_D\Delta t} e^{-is(j\Delta t)H_C\Delta t} = \prod_{j=1}^{P} e^{-i\beta_P H_D} e^{-i\gamma_P H_C} = \prod_{j=1}^{P} \hat{U}_D(\beta_j)\hat{U}_C(\gamma_j)$$

Thus we can approximate AQC by repeatedly letting the system evolve under H_C for some time γ_j and then H_D for some time β_j.

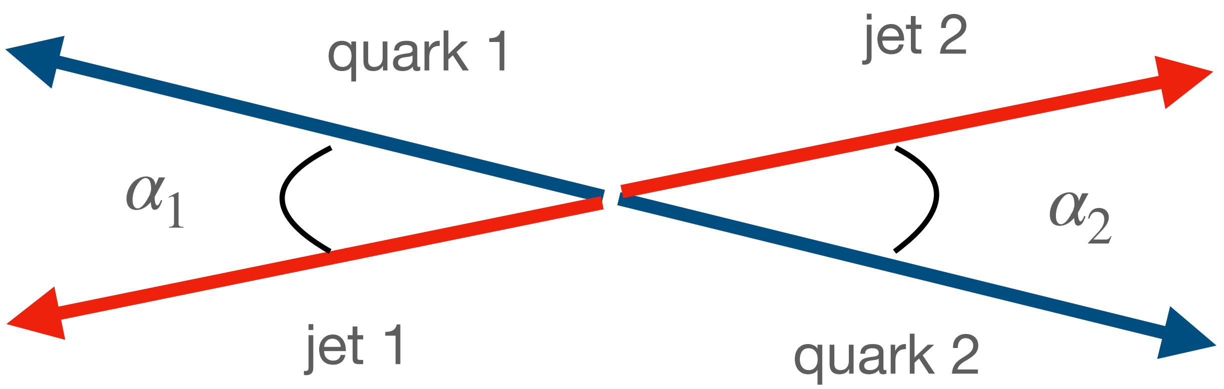
QAOA for MaxCut problem



1. $U(T,0) \approx \prod_{j=1}^{P} e^{-i\beta_P H_D} e^{-i\gamma_P H_C} = \prod_{j=1}^{P} \hat{U}_D(\beta_j)\hat{U}_C(\gamma_j)$
2. $H_D : B = \sum_{j=1}^{n} \sigma_j^x$
3. $H_C : C = \frac{1}{2} \sum_{(i,j)\in E} W_{ij}(I - \sigma_i^z \sigma_j^z)$
4. Initialize the system in the state $|s\rangle = |+\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x\in(0,1)^n} |x\rangle$
5. Construct the circuit (ansatz) by applying the unitaries $\hat{U}_C(\gamma_j)$ and $\hat{U}_D(\beta_j)$ repeatedly P times
6. The final state output by the circuit is $|\psi_P(\gamma,\beta)\rangle = \hat{U}_D(\beta_P)\hat{U}_C(\gamma_P)\ldots\hat{U}_D(\beta_1)\hat{U}_C(\gamma_1)|s\rangle$
7. The expectation value $\hat{H}_C$ with respect to the state $|\psi_P(\gamma,\beta)\rangle$ is calculated through repeated measurements (1024 times in this analysis) of the final state on the computational basis, $F_P(\gamma,\beta) = \langle\psi_P(\gamma,\beta)|\hat{H}_C|\psi_P(\gamma,\beta)\rangle$
8. A classical optimization algorithm is employed to iteratively update the parameters γ and β to find the optimal set of parameters (γ^*,β^*) such that the expectation value $F_P(\gamma,\beta)$ is maximized.

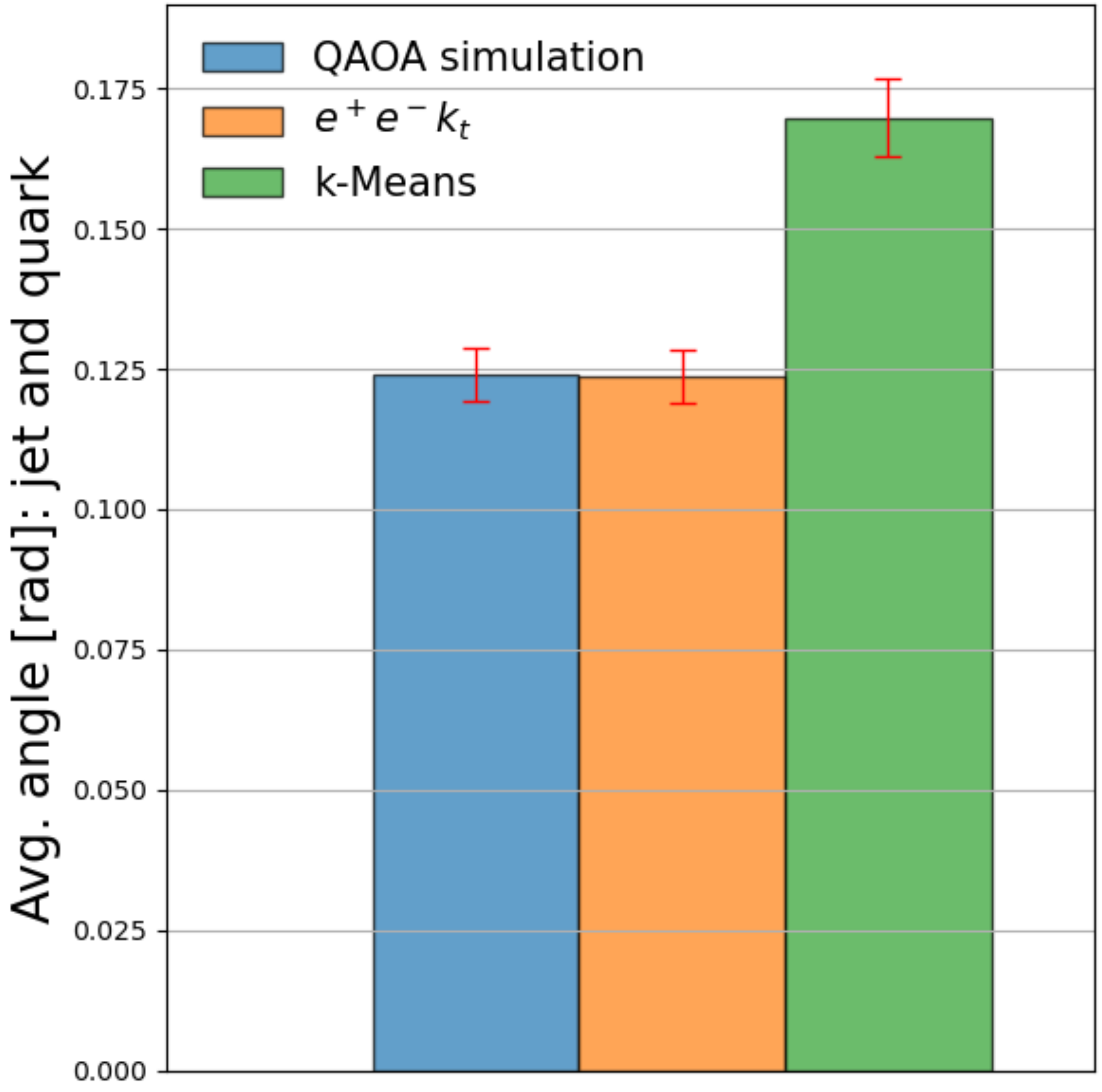
samples and the criteria of jet clustering performance

4000 $e^+e^- \to ZH \to \nu\bar{\nu}s\bar{s}$ with 30 particles



criterion is $\alpha = \alpha_1 + \alpha_2$

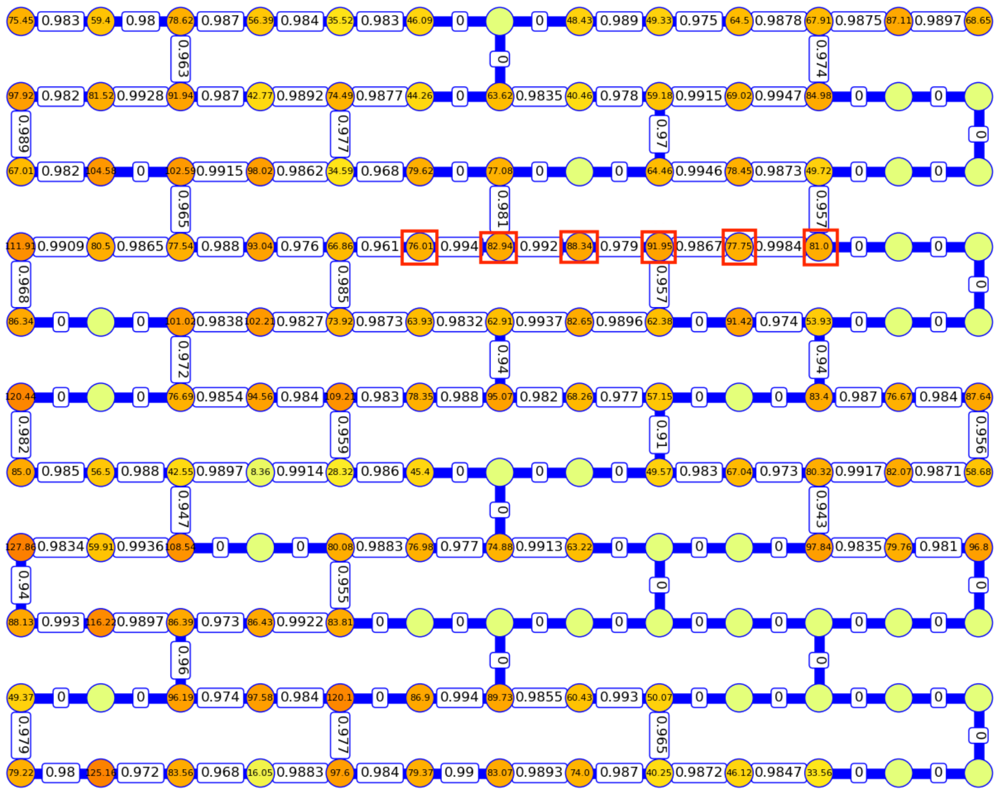
compare jet clustering performance obtained by QAOA, $e^+e^-k_t$ and k-Means



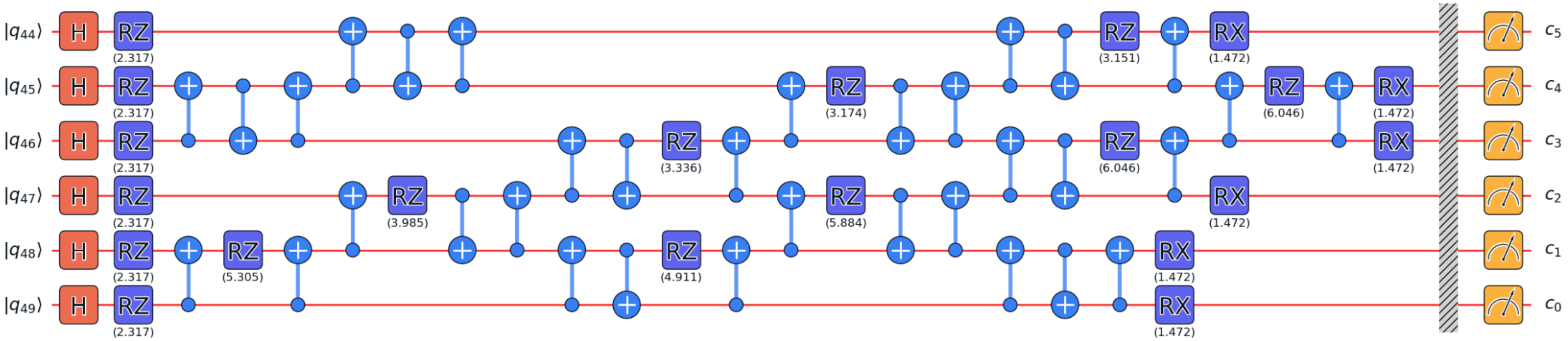
depth=3
k=7

This comparison highlights the potential of the QAOA in the jet clustering problem.

Conduct on the Baihua processor



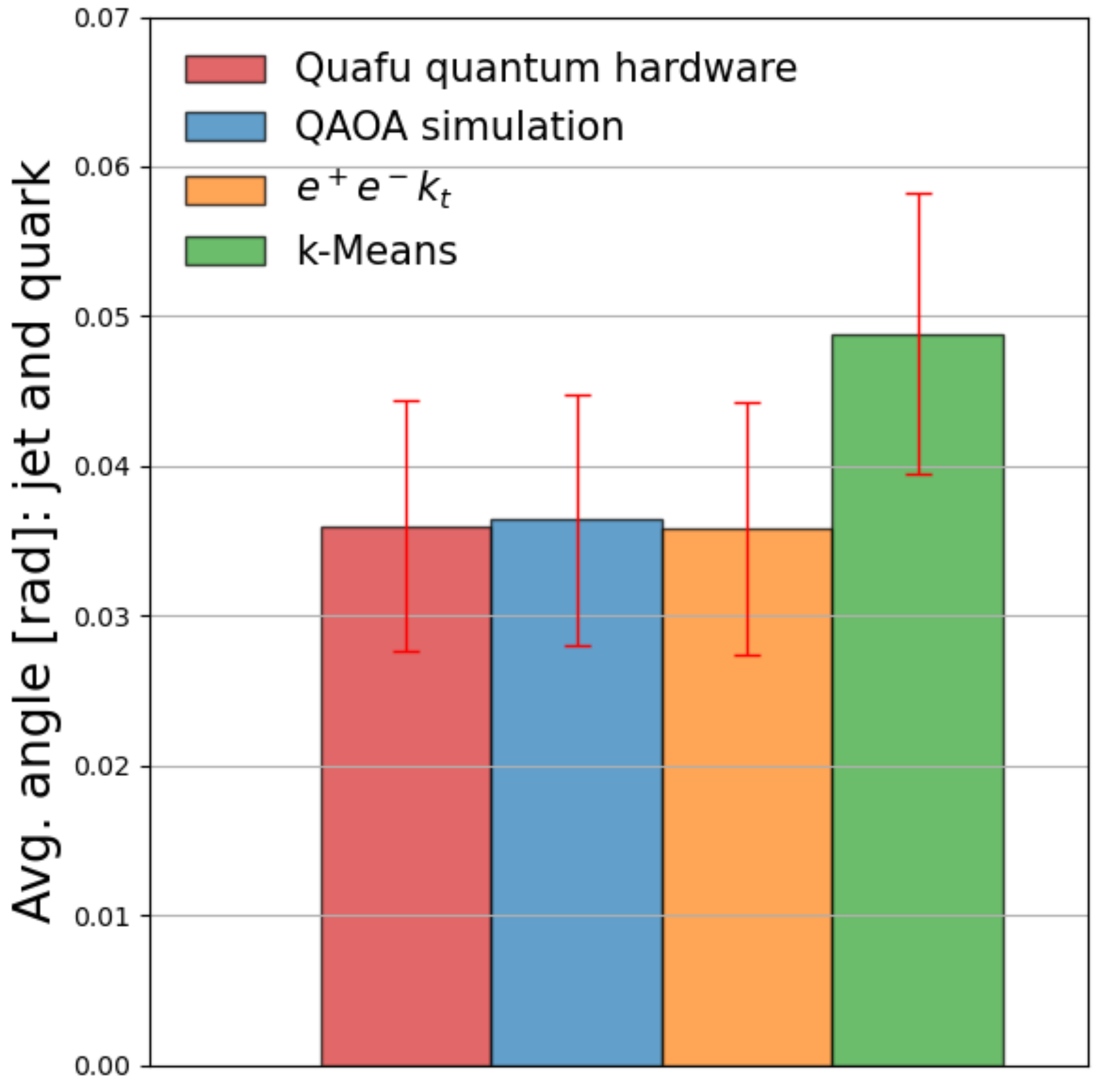
- 123 operational qubits
- relaxation time T_1 of 73.994 us
- dephasing time T_2^* of 29.02 us
- fidelity of single-qubit gate 99.9%
- fidelity of two-qubit gates (CZ) 98.9%



$e^+e^- \to ZH \to \nu\bar{\nu}s\bar{s}$ with 6 particles

compiled QAOA circuit on Baihua processor reaches a depth of 26 with 34 CNOT gates and 27 single-qubit gates.

180 $e^+e^- \rightarrow ZH \rightarrow \nu\bar{\nu}s\bar{s}$ with 6 particles



For this small-sized problem, the quantum hardware can achieve similar performance to a noiseless quantum computer simulator.

Summary

- We proposed and developed a jet origin identification algorithm for jet flavor tagging and jet charge measurement, and achieved significant improvement in the measurement of Higgs rare and FCNC decay.
- The rapid development of quantum algorithms and hardware devices enables the execution of small-scale but representative applications on quantum computers for fundamental sciences. By mapping collision events into graphs, we have obtained promising results in applying QAOA to jet clustering.

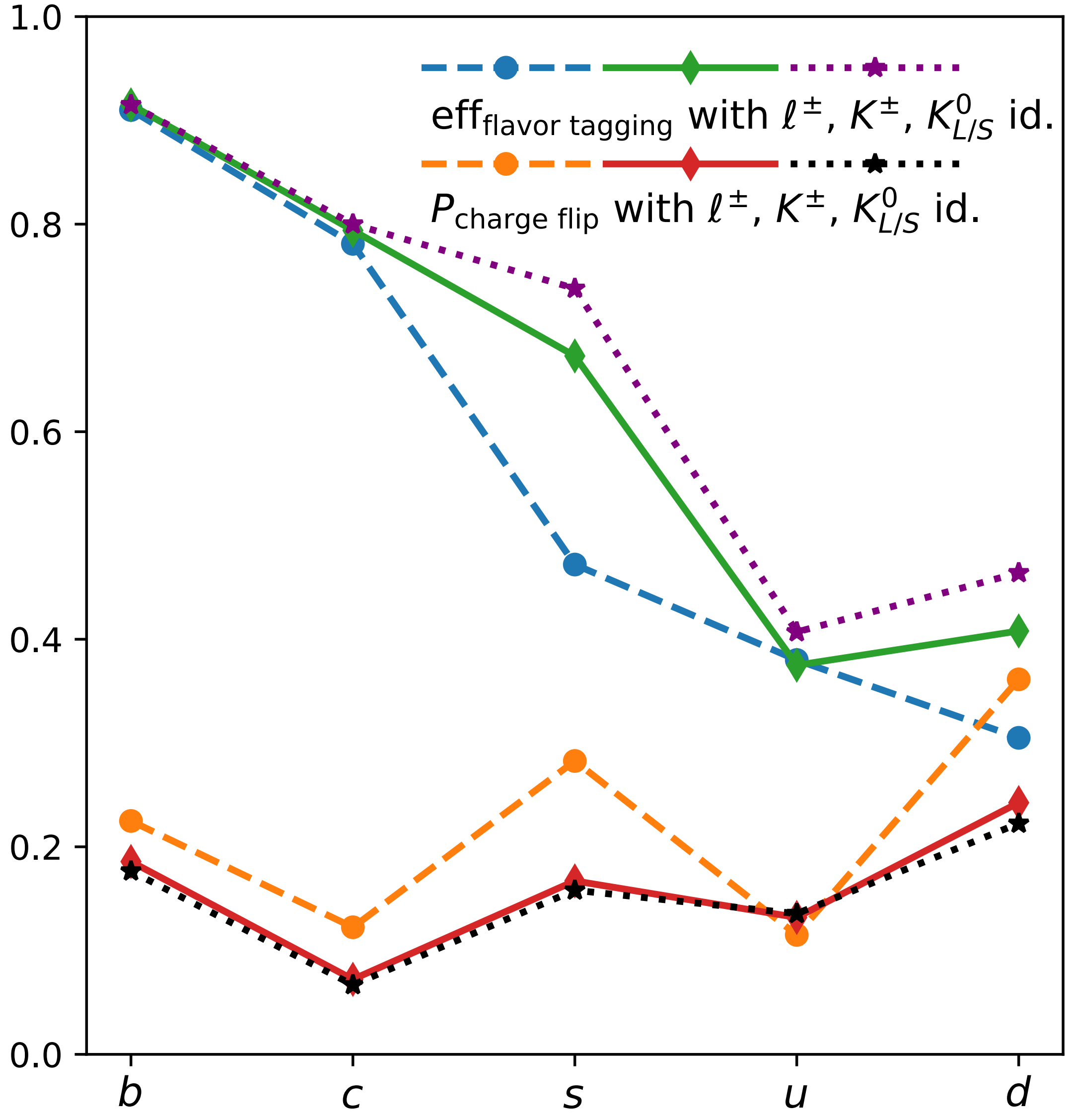
Many thanks !

Backup

the performance of jet origin identification

jet flavor is defined as $max(b+\bar{b},\ c+\bar{c},\ s+\bar{s},\ u+\bar{u},\ d+\bar{d},\ g)$

jet charge is assigned by comparing the quark and anti-quark likelihoods of the corresponding flavor



To understand the impact of PID, three scenarios are compared.

1, assumes perfect identification of charged leptons ($\ell^\pm$)

2, further assumes perfect identification of the charged hadrons ($K^\pm$)

3, on top of the second scenario, assumes perfect identification of K_L and K_S.

default scenario: 2 scenario, based on:

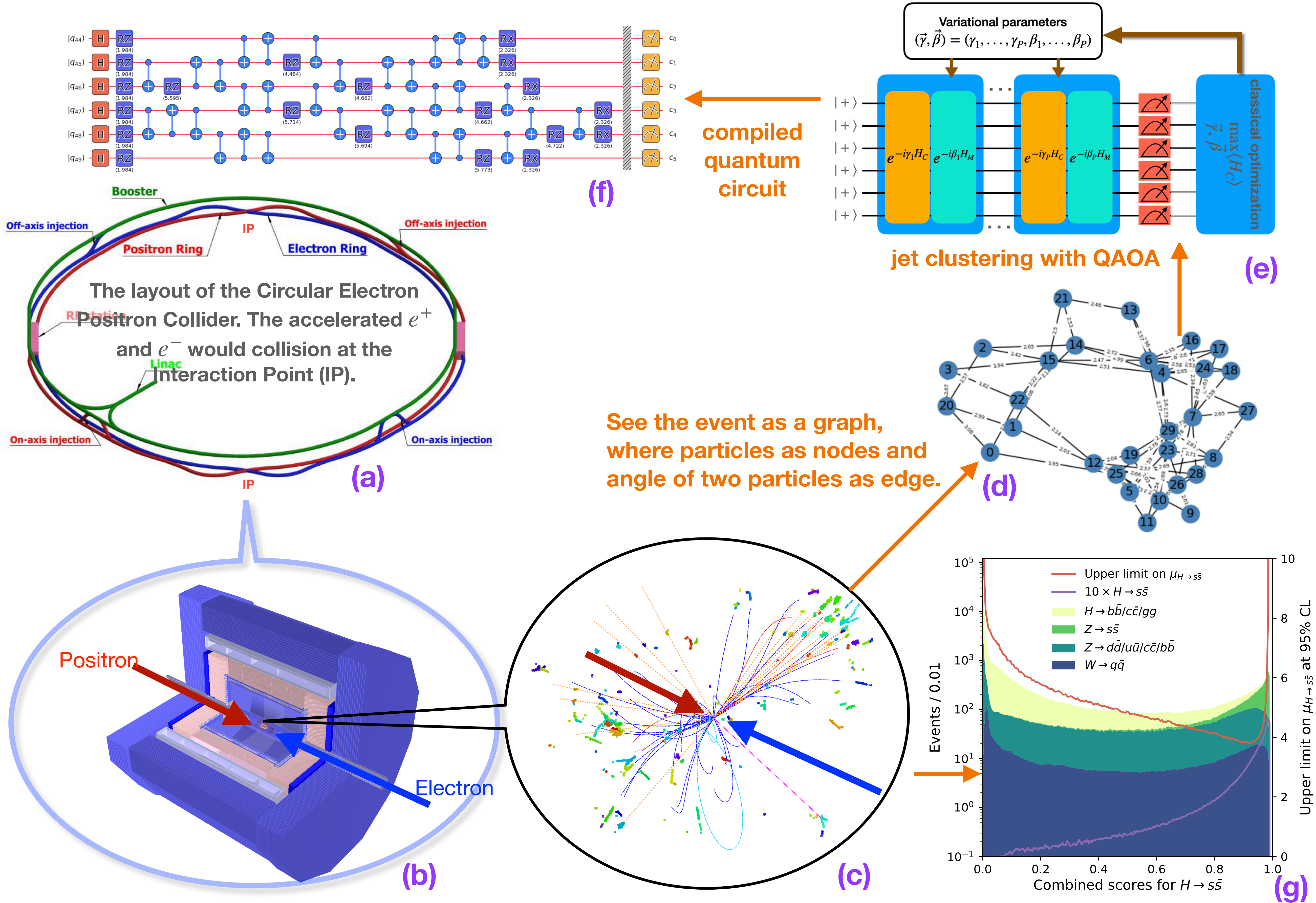
Eur. Phys. J. C 80, 7 (2020)

Journal of Instrumentation 16, P06013 (2021)

Eur. Phys. J. C 78, 464 (2018)

Eur. Phys. J. C 83, 93 (2023)

Nucl. Instrum. Meth. A 1047, 167835 (2023)



compiled quantum circuit

jet clustering with QAOA

The layout of the Circular Electron Positron Collider. The accelerated e^+ and e^- would collision at the Interaction Point (IP).

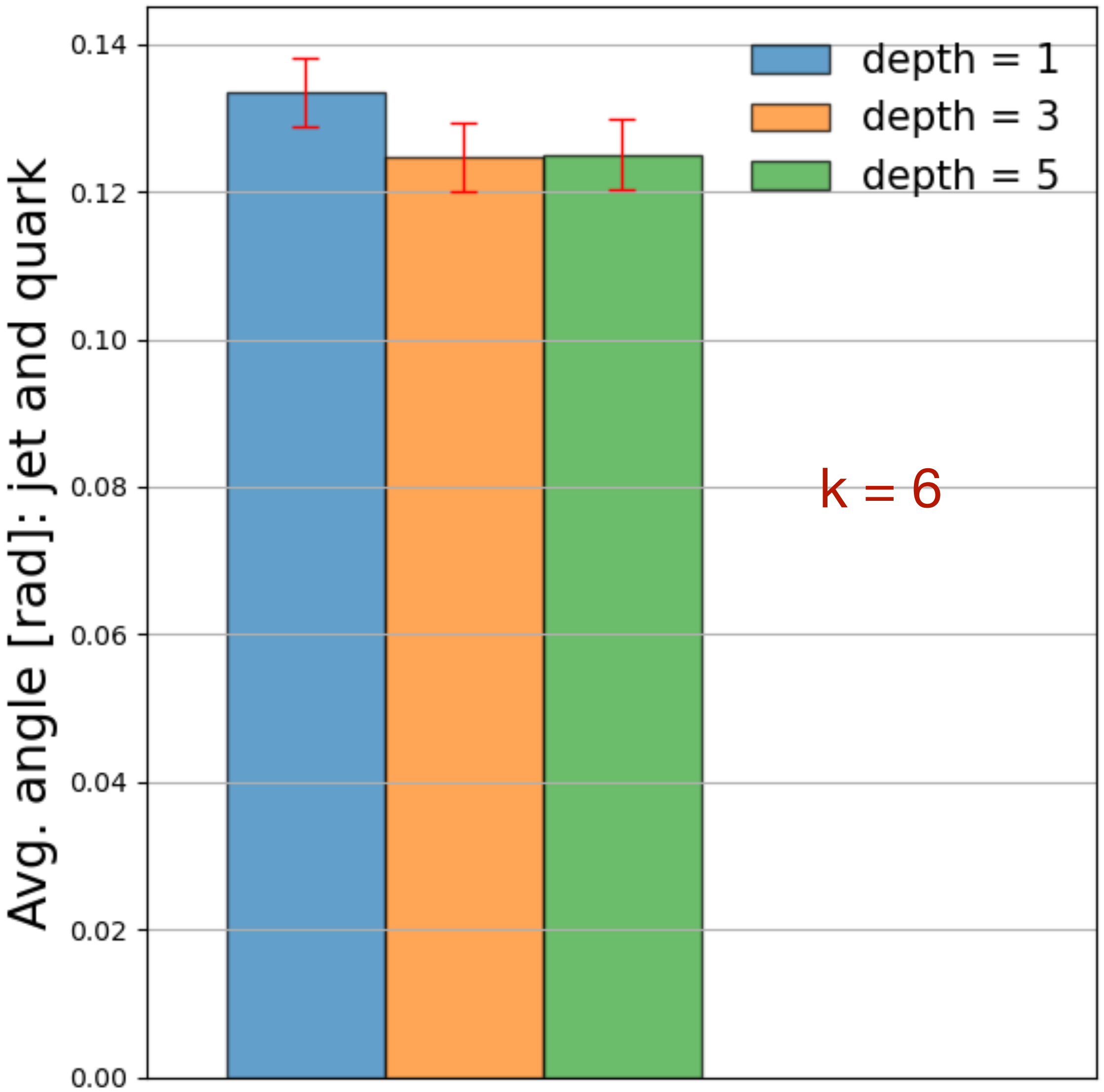
See the event as a graph, where particles as nodes and angle of two particles as edge.

The collision of e^+ and e^- can generate quarks, gluons, and leptons.

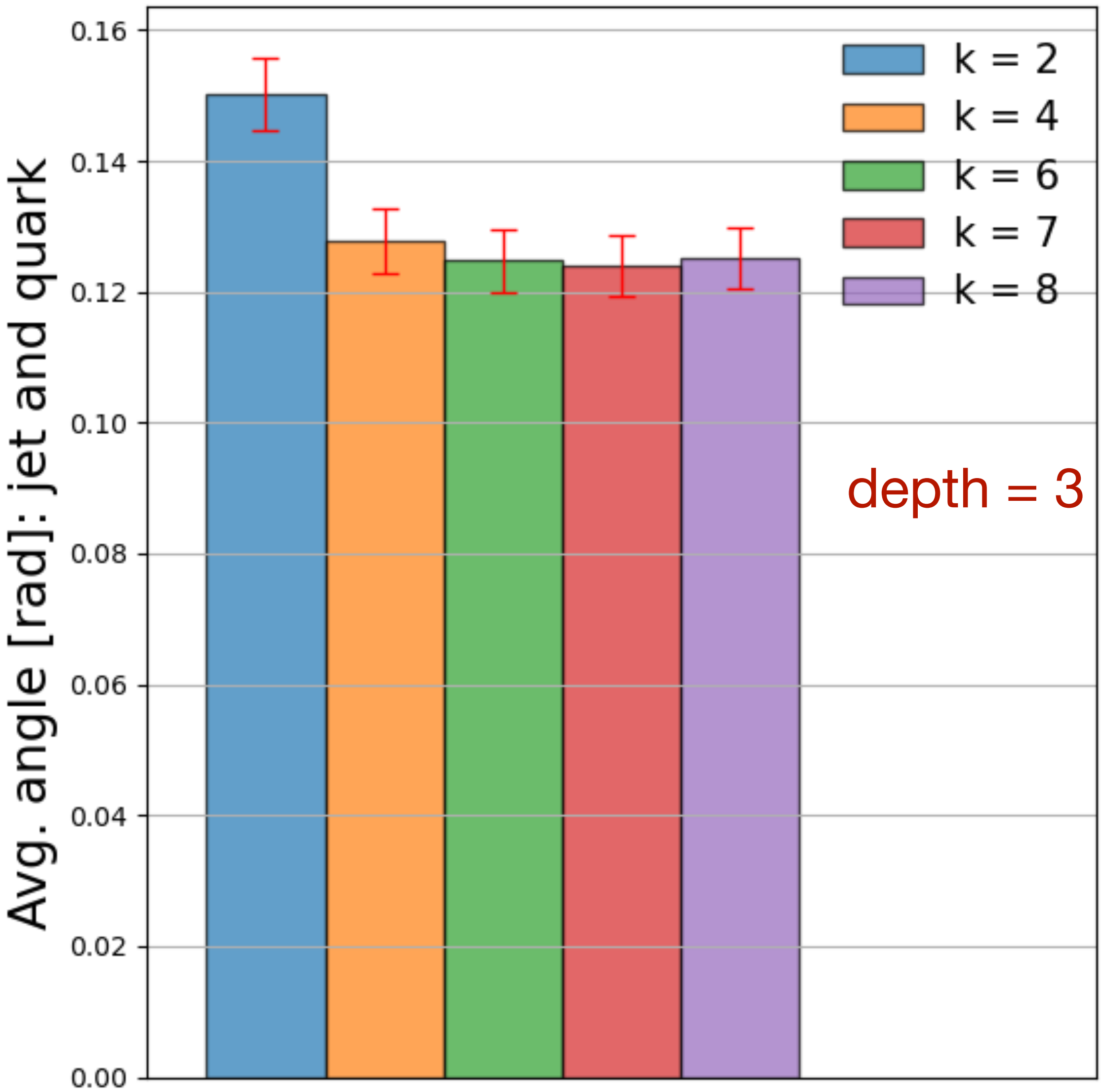
The quarks and gluons would immediately transform into collimated particle sprays known as jets.

With jet clustering and other techniques, the related physics analyses can be performed.

jet clustering performace v.s. QAOA depth (P) and k-value



k = 6



depth = 3