中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Searches for the Higgs Boson Invisible Decay at the CEPC

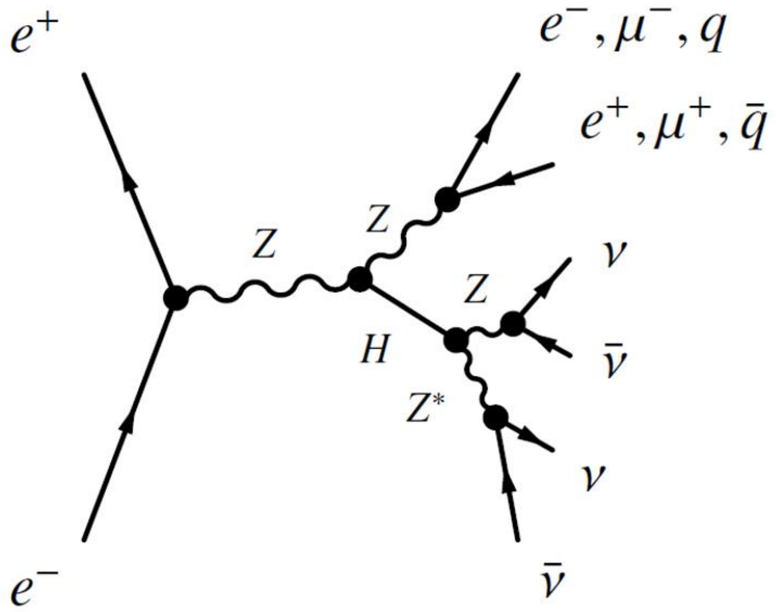
Geliang Liu（刘格良）

Nov. 5 – 10, 2025, CEPC Workshop 2025

Motivation

➢ Higgs boson invisible decays

- Higgs strahlung: ee→Z*→ZH
  ee→Z(→ee/μμ/qq)H(→invisible)
- In the SM: H→ZZ*→4v
  - BR(H→4v)=**0.105%**
- BSM: H→sparticles / dark matter / LLPs, ...



➢ Strong probe of BSM physics

- **Direct search** for new physics
- By nature **model-independent**

➢ Tests of detector performances

- The overall validity of the **particle flow (PF) algorithm**
- Large **phase space coverage**
- Reconstruction of **missing kinematics**

Previous studies

At the LHC

Experiments	Data	Expected results	Publication
ATLAS	13 TeV; 139 fb-1	UL on BR(H→inv): 10%	JHEP08(2022)104
CMS	13 TeV; 139 fb-1	UL on BR(H→inv): 10%	PRD 105 (2022) 092007

Future electron-positron colliders

Experiments	Data	Expected results	Publication
ILC	250, 350, 500 GeV; 250, 350, 500 fb-1	UL on BR(H→inv): 0.26%	arXiv:1909.07537
FCC-ee	240+365 GeV; 10.8, 3 ab-1	3.9 σ on BR(H→ZZ→4v)	Presentation
CEPC (CDR)	240 GeV; 5.6 ab-1	UL on BR(H→inv): 0.26%	Chinese Phys. C 44 123001

- **Superiorities of electron-positron colliders:**
  - Initial state well recognized
  - No pileup contamination
  - Better signal-over-background ratios for SM Higgs boson production.
- **Aim: estimate H→invisible searches based on CEPC reference detector full simulation**

Samples

➢ **Simulation samples produced with CEPCSW 25.3.6**

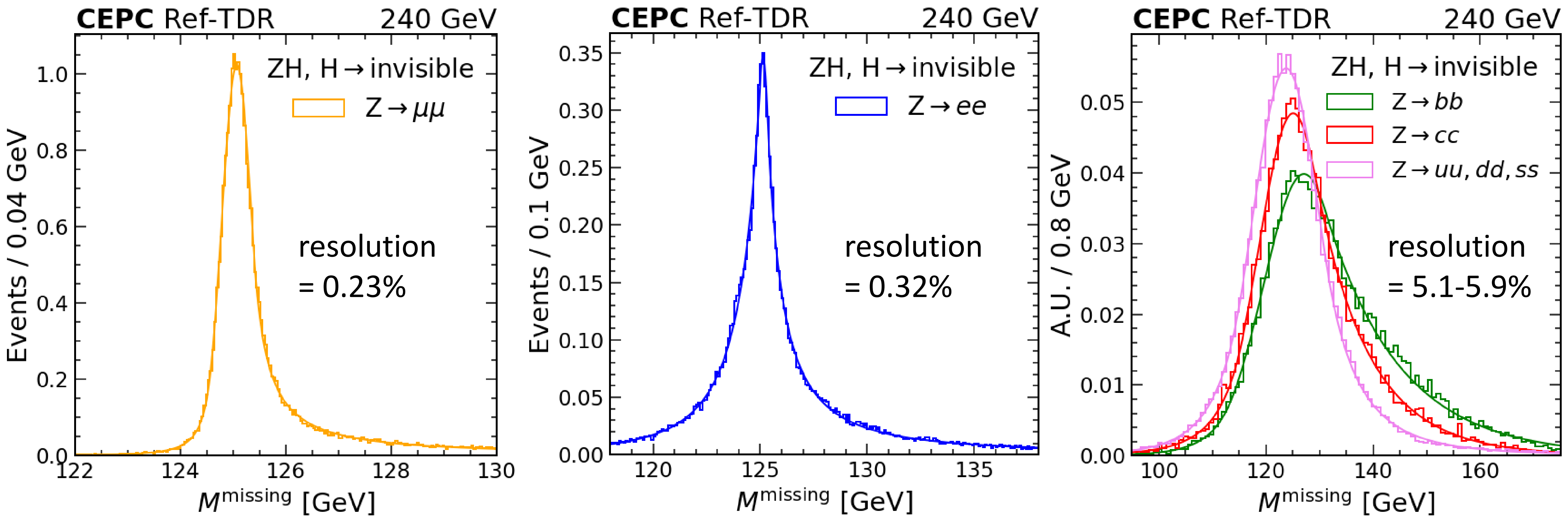
Experiments	Processes	Events
Signal	Z(→ee/μμ/qq)H(→4v)	100k per final state
4-fermion (4f) backgrounds	single-Z, single-W, Z-or-W ZZ, WW, ZZ-or-WW 41 different final states in total	400k per final state
2-fermion (2f) backgrounds	ee, μμ, ττ, qq	100k – 400 k per final state
H visible decay	(ee/μμ/ττ/vv/qq) + H→visible	1M per sample

- All samples will be scaled to an integrated luminosity of 21.6 ab-1

Key property

➢ Missing mass

- $p^{visible} = \sum_i^{objects} p_i$
- $p^{missing} = p^{total} - p^{visible}$, $p^{total} = (0, 0, 0, 240\ \text{GeV})$



Powerful discriminant thanks to **full reconstruction of missing information** and **good detector resolution**

Event selection

Baseline selection

- Start from the objects reconstructed from the **particle flow algorithm** (PFO)
- **Object selection** to define different final states / channels
- Preliminary **kinematic selection**

2μ channel:
- Two muons, with $|\cos\theta|<0.99$
- Opposite charge.
- $M_{\mu\mu} \in [40, 120]$ GeV

2e channel:
- Two electrons, with $|\cos\theta|<0.99$
- Opposite charge.
- $M_{ee} \in [40, 120]$ GeV

2q channel:
- At least two PFOs
- $M^{visible} \in [30, 130]$ GeV
- $p^{visible} \in [10, 80]$ GeV

- **Lepton identification**
  - based on full detector information;
  - great performance with >90% efficiency and <1% (0.1%) for e (μ) at p > 2 GeV;
  - being used in many benchmark analyses in the CEPC ref-detector TDR
- **Orthogonality** guaranteed: 2μ > 2e > 2q

Event selection

Kinematic selection

- Further suppress backgrounds while keeping a high signal efficiency

2μ channel:

- $E_{recoil} > 125$ GeV
- $M_{recoil} > 110$ GeV
- $M_{missing} \in (110,150)$ GeV
- $E_{visible} \in (80,120)$ GeV
- $M_{visible} < 120$ GeV
- $P_{visible} \in (20, 70)$ GeV
- $E_{neutral} < 60$ GeV
- $N_{charged} < 7$
- $N_{neutral} < 10$
- $|D_0| < 0.05$ mm
- $|Z_0| < 0.1$ mm

2e channel:

- $E_{recoil} > 125$ GeV
- $M_{recoil} > 120$ GeV
- $M_{missing} \in (50,160)$ GeV
- $E_{visible} \in (80,190)$ GeV
- $M_{visible} < 160$ GeV
- $P_{visible} \in (20, 70)$ GeV
- $E_{neutral} < 90$ GeV
- $N_{charged} < 7$
- $N_{neutral} < 15$
- $|D_0| < 0.1$ mm
- $|Z_0| < 0.1$ mm

2q channel:

- $M_{missing} \in (100,170)$ GeV
- $E_{visible} \in (70,130)$ GeV
- $P_{visible} \in (20, 70)$ GeV
- $E_{charged} \in (15,100)$ GeV
- $E_{neutral} \in (5,90)$ GeV
- $N_{charged} > 5$
- $N_{neutral} > 10$
- Jet $N_{charged} > 0$
- Jet $N_{neutral} > 2$

* $p_{recoil} = p_{total} - p_{l_1} - p_{l_2}$
* D_0, Z_0: transverse, longitudinal impact parameter of the inner track w.r.t. to IP
* For 2q channel: jets are clustered with ee-kt algorithm, requiring 2 jets

Selection cutflow table

process		signal	2(μ/e/q)+2v	2f	visible H	others
2μ	total yield	1.56E+02	6.13E+06	1.92E+09	4.40E+06	4.09E+08
	Baseline sel eff	96.1%	32.0%	2.35%	2.55%	0.88%
	Kinematic sel eff	98.0%	19.8%	3.40%	0.44%	5.31%
	selected	1.46E+02	3.88E+05	1.53E+06	4.91E+02	1.78E+05
2e	total yield	1.61E+02	6.02E+06	1.92E+09	4.40E+06	4.09E+08
	Baseline sel eff	83.8%	41.7%	1.03%	1.96%	1.60%
	Kinematic sel eff	95.3%	23.0%	3.35%	2.19%	5.77%
	selected	1.29E+02	5.78E+05	6.62E+05	1.89E+03	3.77E+05
2q	total yield	3.13E+03	7.98E+06	1.92E+09	4.40E+06	4.07E+08
	Baseline sel eff	99.0%	66.1%	9.24%	19.8%	8.35%
	Kinematic sel eff	95.4%	38.1%	37.29%	37.8%	12.9%
	selected	2.96E+03	2.01E+06	6.62E+07	3.28E+05	4.37E+06

Multi-variate analysis

XGBoost models

- trained in each channel to further distinguish signal and backgrounds

Input features

2μ channel:

- $E_{\mu\mu}, M_{\mu\mu}, P_{\mu\mu}, P^{T}_{\mu\mu}$
- M_{recoil}
- $E_{visible}, M_{visible}, P_{visible}, P^{T}_{visible}$
- $M_{missing}$
- $\Delta\phi_{\mu\mu}, \Delta R_{\mu\mu}$
- $D_0^{\mu_1}, D_0^{\mu_2}, Z_0^{\mu_1}, Z_0^{\mu_2}$
- $N_{charged}, N_{neutral}, E_{neutral}$

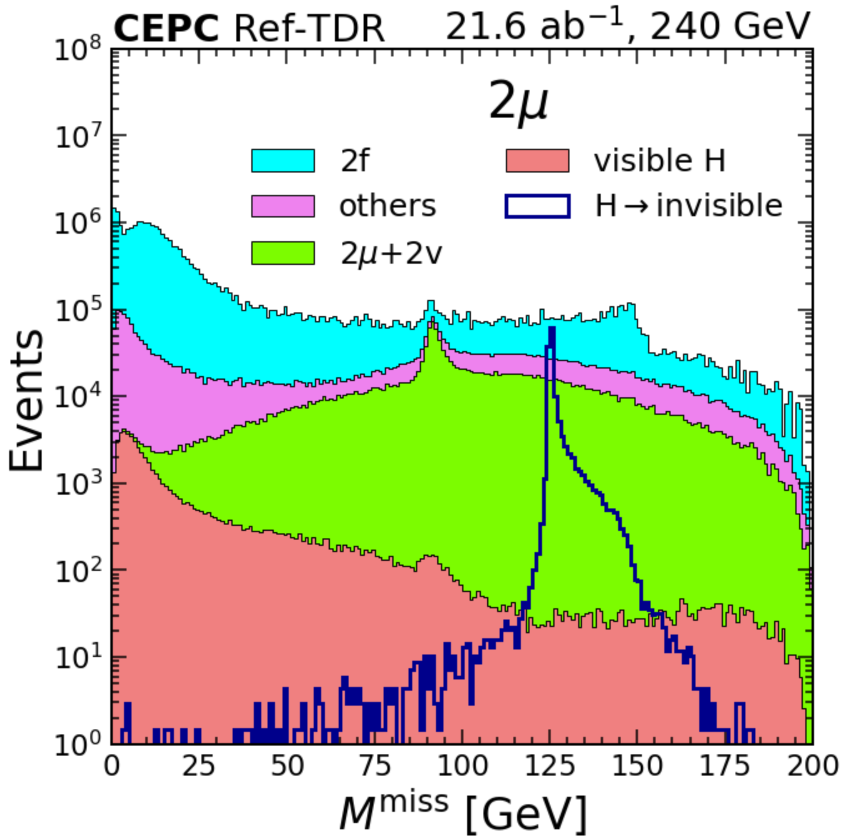
2e channel:

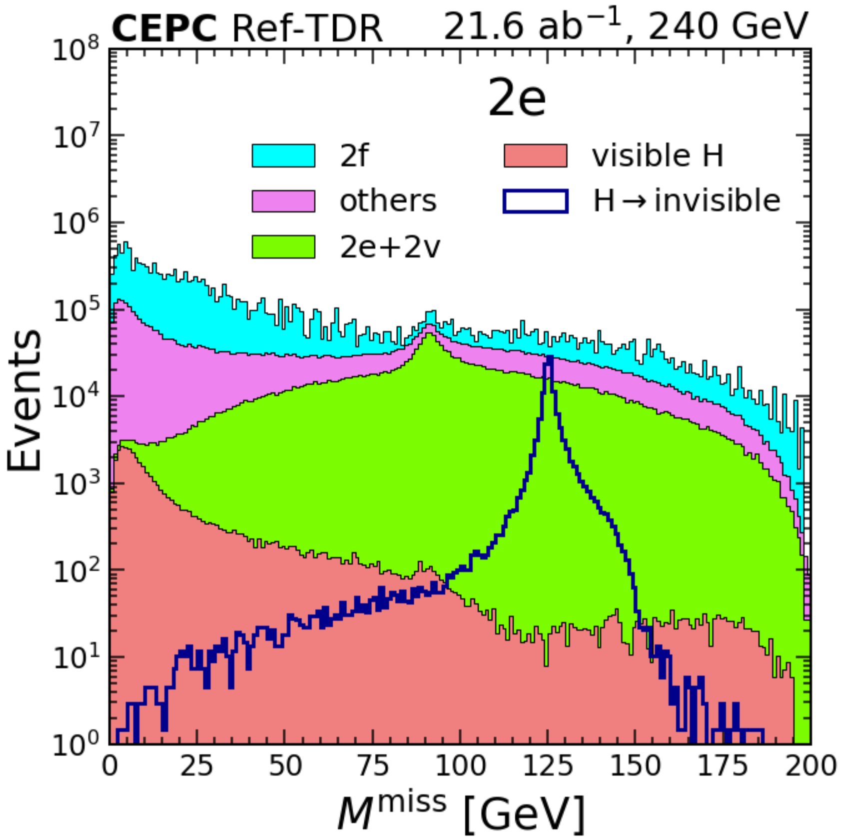
- $E_{\mu\mu}, M_{\mu\mu}, P_{\mu\mu}, P^{T}_{\mu\mu}$
- M_{recoil}
- $E_{visible}, M_{visible}, P_{visible}, P^{T}_{visible}$
- $M_{missing}$
- $\Delta\phi_{\mu\mu}, \Delta R_{\mu\mu}$
- $D_0^{e_1}, D_0^{e_2}, Z_0^{e_1}, Z_0^{e_2}$
- $N_{charged}, N_{neutral}, E_{neutral}$

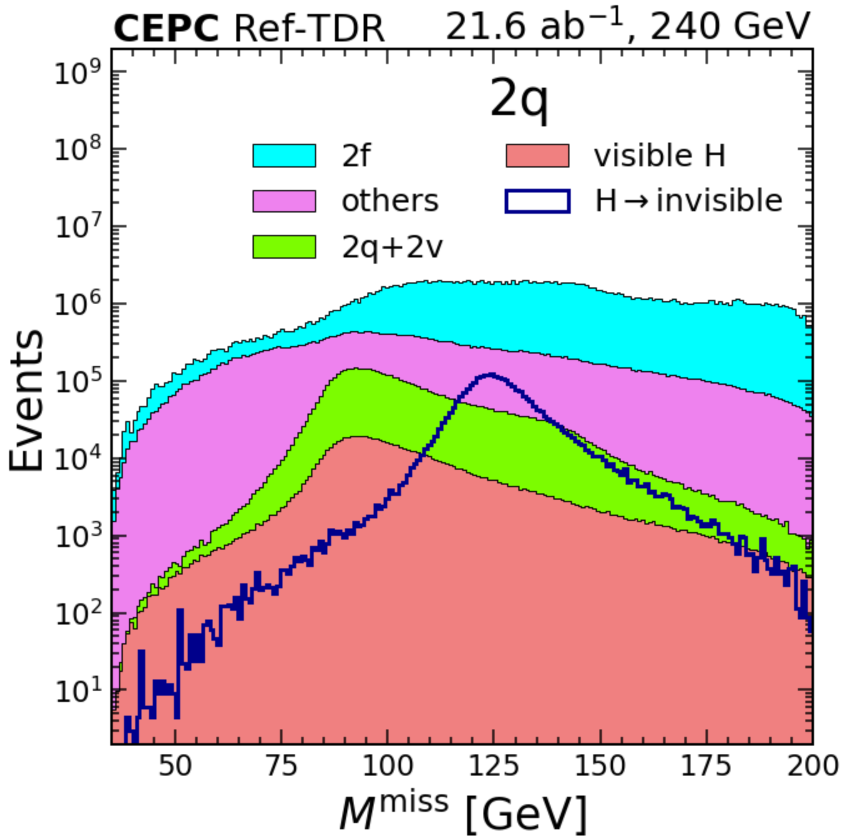
2q channel:

- $E_{visible}, M_{visible}, P_{visible}, P^{T}_{visible}$
- $M_{missing}$
- $\Delta\phi_{jj}, \Delta R_{jj}$
- $N_{charged}, N_{neutral}$
- $N^{j_1}_{charged}, N^{j_2}_{charged}$
- $ECF_2^{j_1}, ECF_2^{j_2}$ (energy correlation function)
- $\left(\frac{\tau 3}{\tau_1}\right)^{j_1}, \left(\frac{\tau 3}{\tau_1}\right)^{j_2}$ (N-subjettiness)

Important features: missing mass

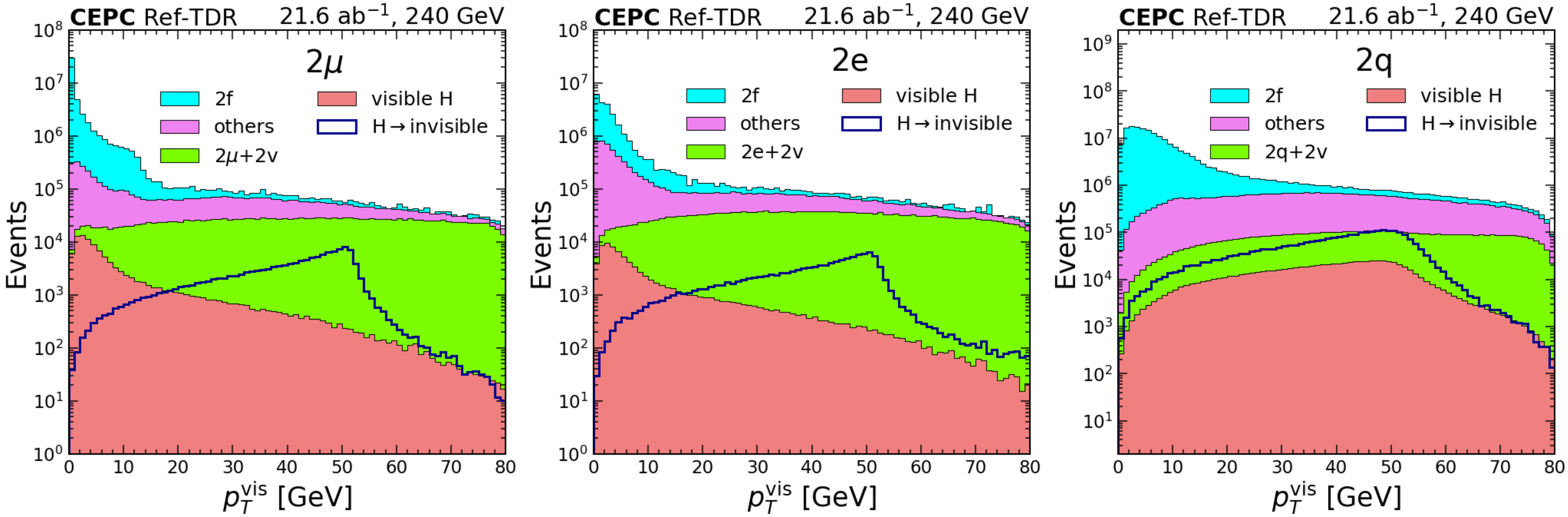






- The **signal** distributed around 125 GeV (BR(H→invisible) set to be 1).
- Reducible background 2(μ/e/q)+2ν distributed around 91 GeV.
- Other backgrounds distributed around 0 or flatly.

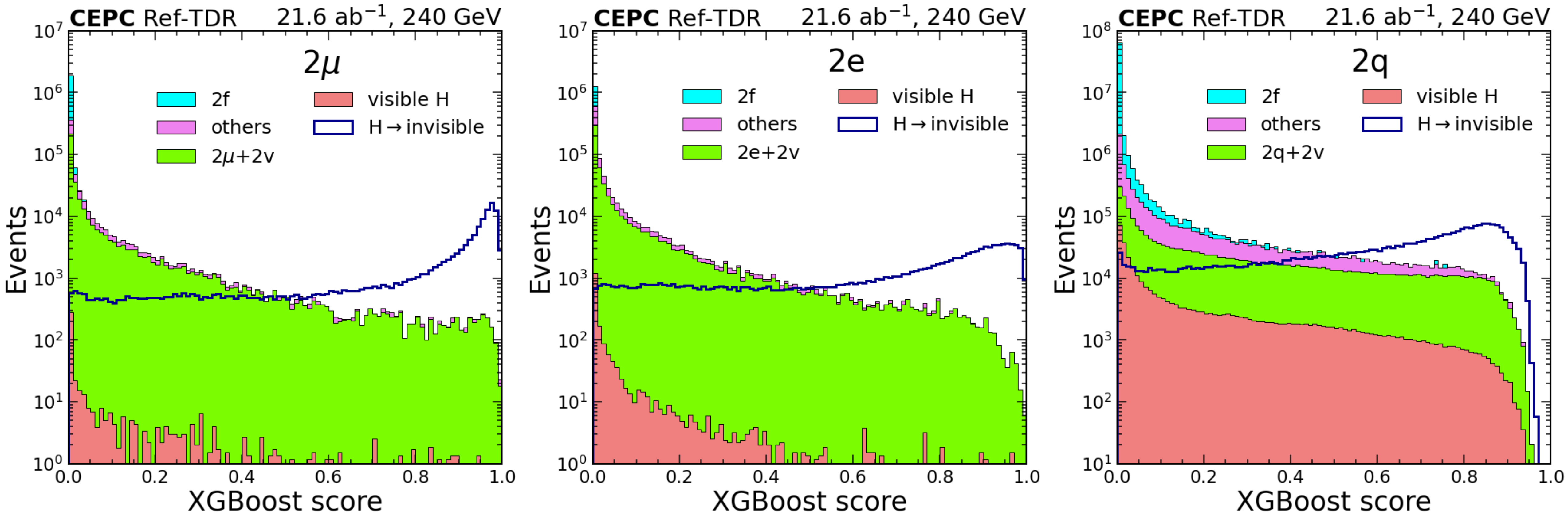
Important features: visible pt



- The **signal** distributed around 50 GeV.
- 2(μ/e/q)+2v distributed flatly.
- 2f distributed around 0: mostly in high |cosθ| region.

ML-based discriminant

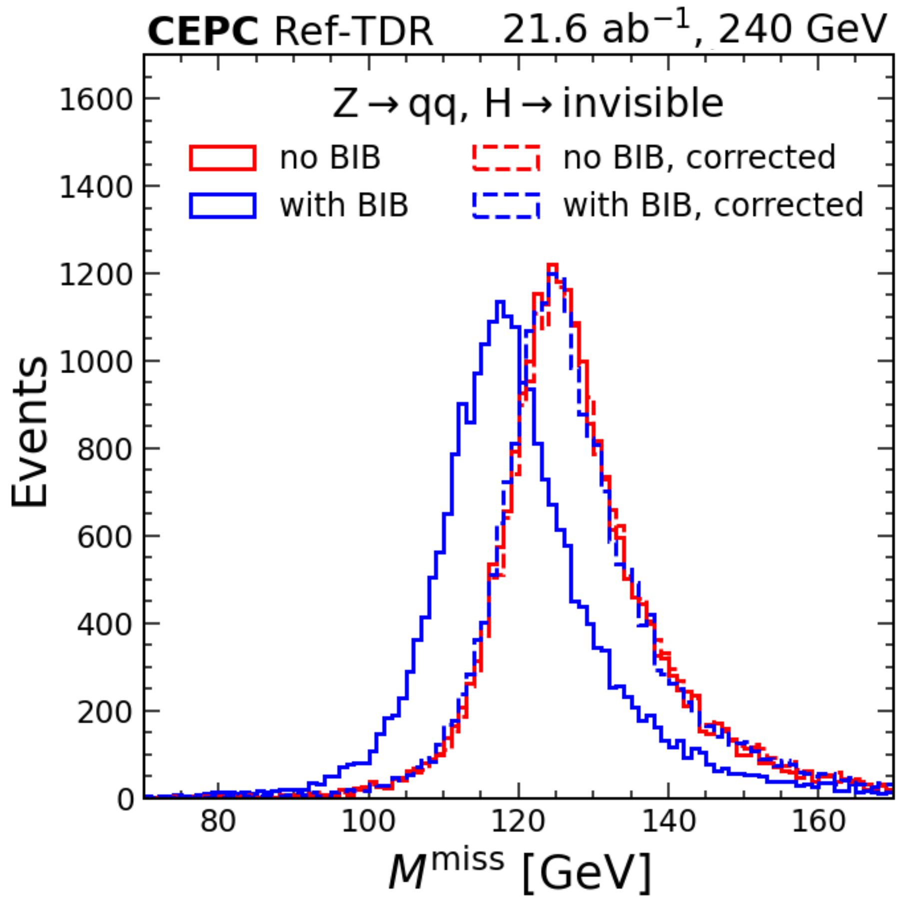
XGBoost score distributions



- Main backgrounds contaminating with signals: 2(μ/e/q)+2v
- See backup for overtraining tests, feature importance and correlations

Systematic uncertainties

- **Theoretical uncertainties**
  - Not considered, should have small impacts
- **Lepton efficiency, energy corrections, jet energy corrections**
  - Believed to be <1% thanks to the expected large datasets for calibration
- **Beam induced backgrounds (BIB)**
  - After filters of collimators, the main BIB is **pair production**: ee→eeγγ→eeee
  - **Very low energy**: not recognized by the tracker
  - **Very forward**: induce neutral objects with |cosθ| > 0.98
  - Treatment: **discard neutral objects with |cosθ| > 0.98**
  - **Negligible impacts** on energy scale and resolution



Statistical inferences

- **Statistical analyses are performed with the CMS combine tool.**
- **Parameter of interest: BR(H→invisible)**
- **Discriminating variable: XGBoost score**
  - Binned likelihood fits & asymptotic formulae
- **Statistical-uncertainty-only results**
- **Two scenarios:**
  - **SM** H→invisible as a signal: expected **uncertainty** and statistical **significance**
  - **BSM** H→invisible as a signal, while the SM one is a background: expected **upper limits** at 95% confidence level.

Statistical-only results

channel	5.6 ab-1				21.6 ab-1		
	uncertainty (SM)	CDR uncertainty	significance (SM)	UL (BSM)	uncertainty (SM)	significance (SM)	UL (BSM)
2μ	-80%/+84%	222%	1.2σ	0.18%	-42%/+43%	2.5σ	0.089%
2e	-100%/+124%	428%	0.9σ	0.27%	-60%/+62%	1.7σ	0.13%
2q	-58%/+58%	90%	1.7σ	0.12%	-29%/+29%	3.4σ	0.061%
combine	-44%/+45%	82%	2.3σ	0.092%	-22%/+22%	4.5σ	0.047%

- Much **better sensitivities** than the CDR studies !
- Thanks to the MVA algorithm

- Within SM: close to discovery level
- BSM: sensitive to any invisible decays with a branching ratio ~ 0.05%

Conclusion and prospects

Preliminary searches for Higgs boson invisible decays are performed

- Missing mass with good resolution as the key feature
- Much improved sensitivities thanks to the ML-based algorithm

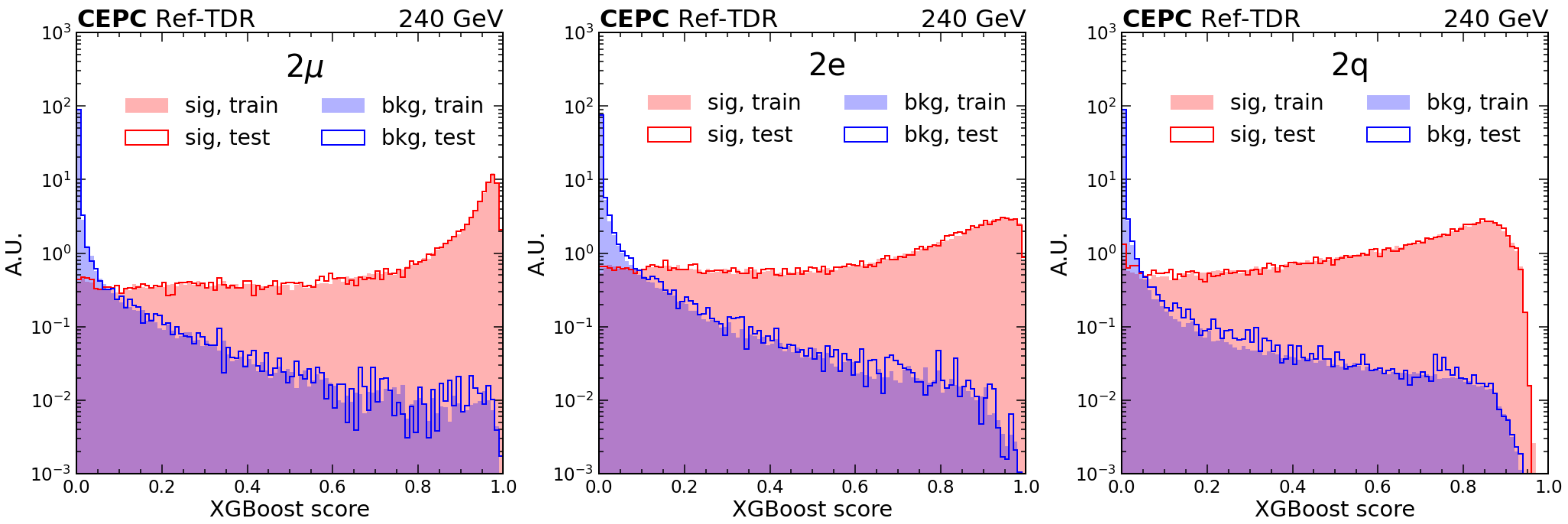
Ongoing developments

- **Jet-flavor tagging information**
  - **Jet-origin-ID** developed based on the CEPC ref-detector shows great performances.
  - Different jet flavors have **different resolution and background composition**: potential improvements on sensitivities.
  - **Categorization** based on jet flavors; add jet flavors as input features to XGBoost models.
- **Parameterization / smoothening of the XGBoost score distributions.**
- **Consideration of systematic uncertainties in more details.**

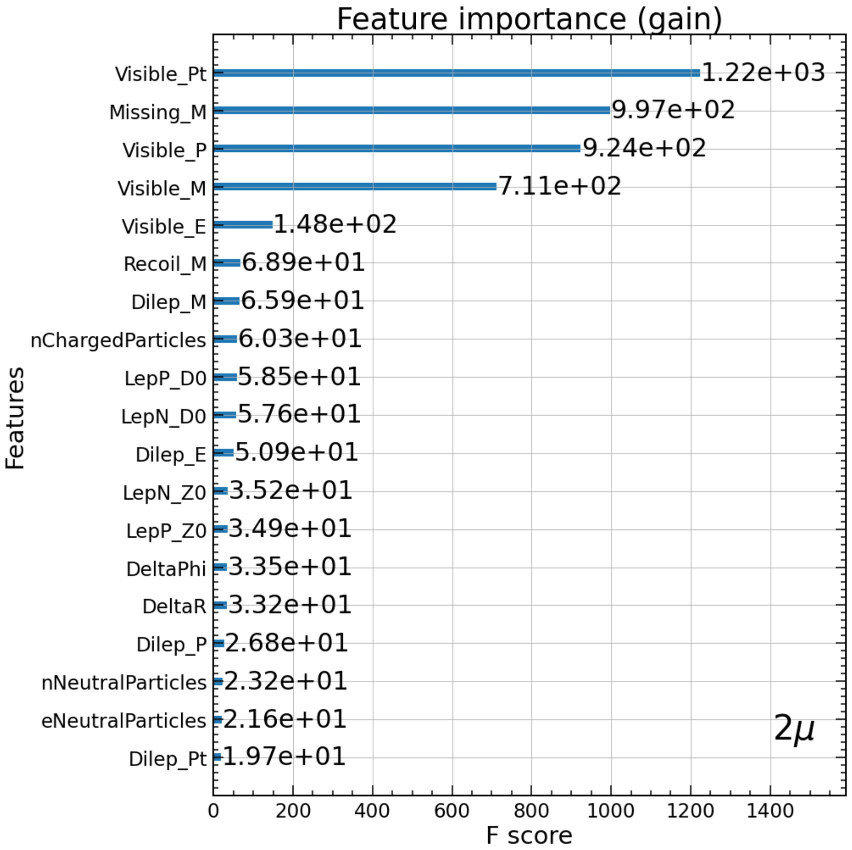
Thanks for your attention!

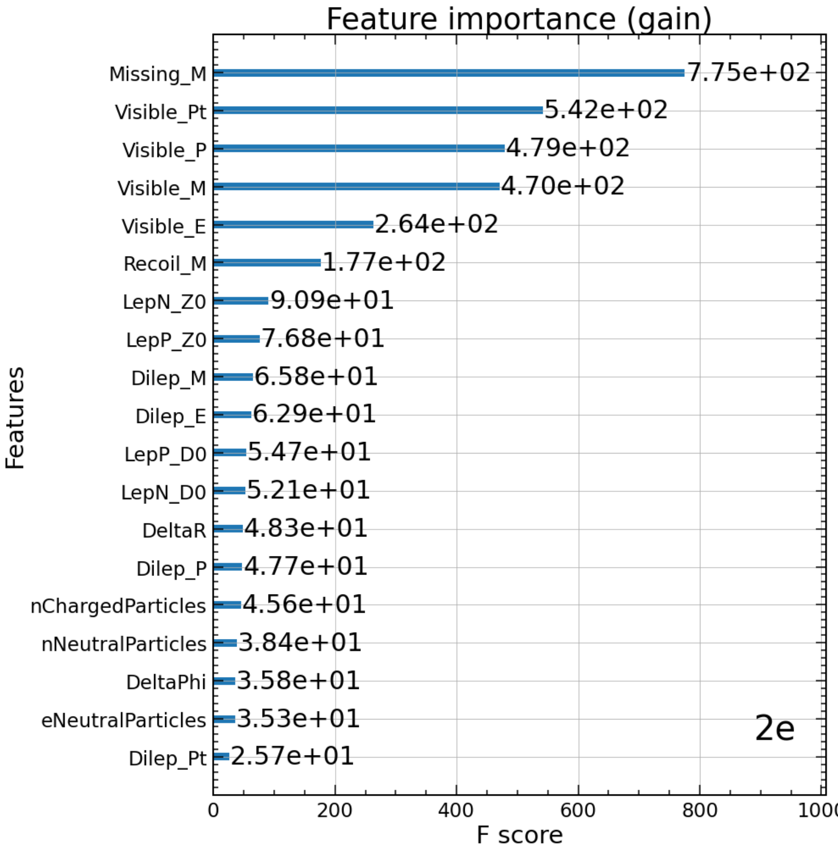
Backup

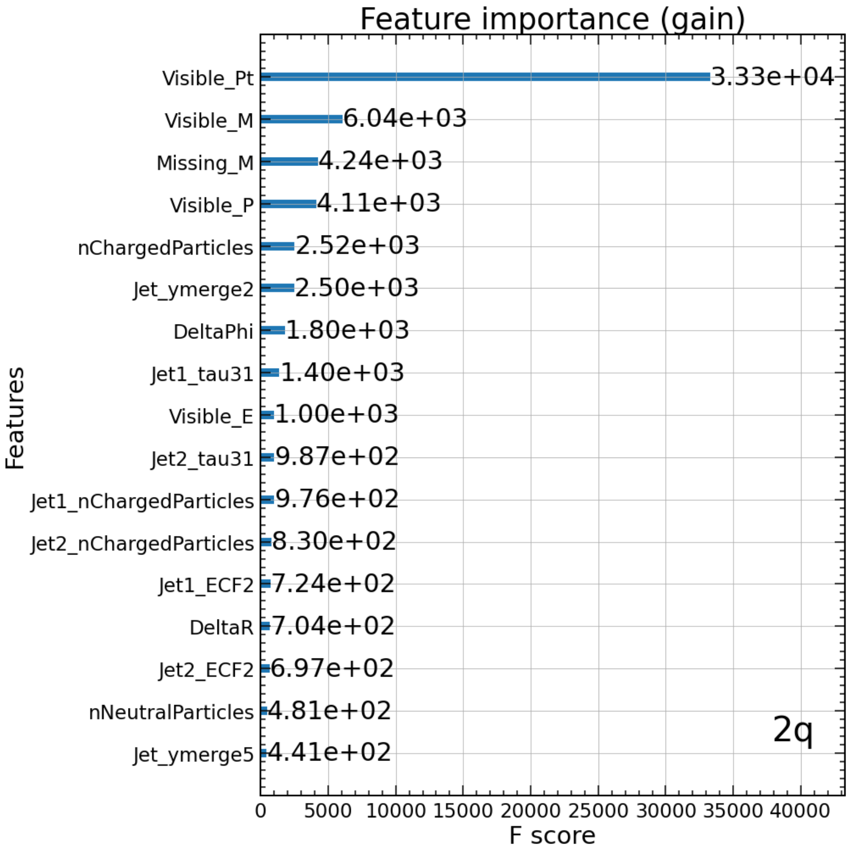
XGBoost overtraining



XGBoost feature importance







XGBoost feature correlation