Approval:
Search for Higgs boson invisible decays at CEPC

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

Geliang Liu (刘格良)

Jun. 13th, 2025

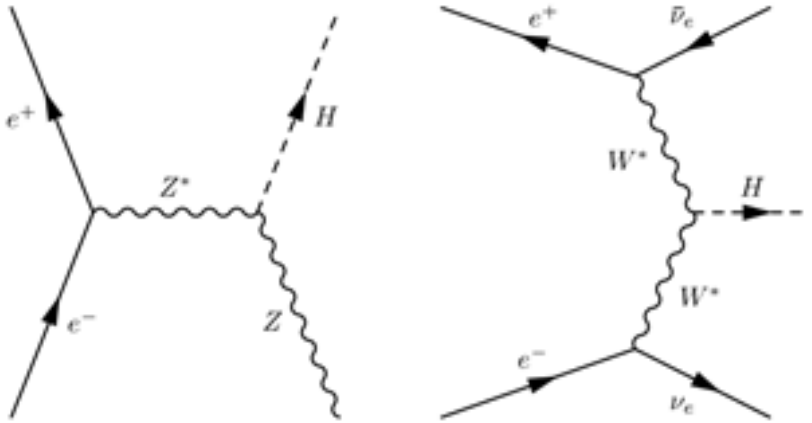
Outlines

- **Introduction**
- **Samples**
- **Key property**
- **Event selection**
  - Baseline selection
  - Kinematic selection
- **Multivariate-analysis discriminant**
- **Results**
- **Conclusion and future improvements**

Introduction

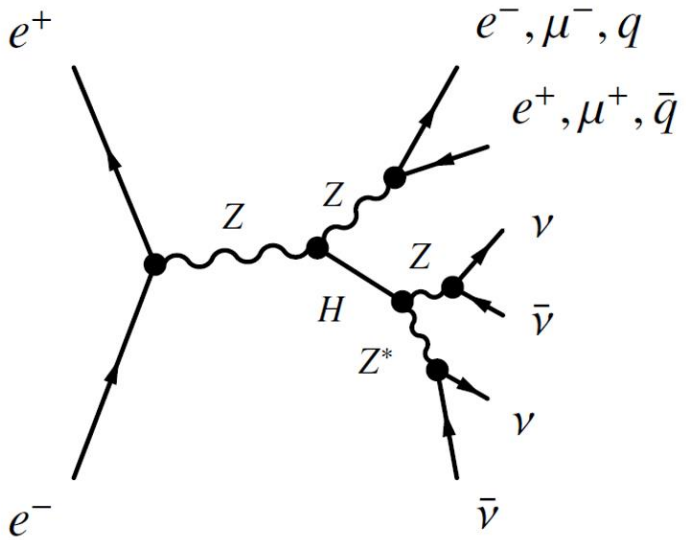
➢ Higgs boson production

- Higgs strahlung: ee→Z*→ZH (dominant)
- W fusion: ee→$\nu_e \nu_e$H (sub-dominant)



➢ Higgs boson invisible decays

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



➢ Tests of detector performances

- The overall validity of the particle flow (PF) algorithm
- Enough detector coverage
- Reconstruction of missing kinematics

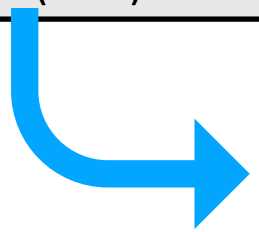
Previous studies

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→4ν)	Presentation
CEPC (CDR)	240 GeV; 5.6 ab-1	UL on BR(H→inv): 0.26%	Chinese Phys. C 44 123001



Will proceed the analysis with the full simulation of CEPC ref-detector for TDR !

Samples

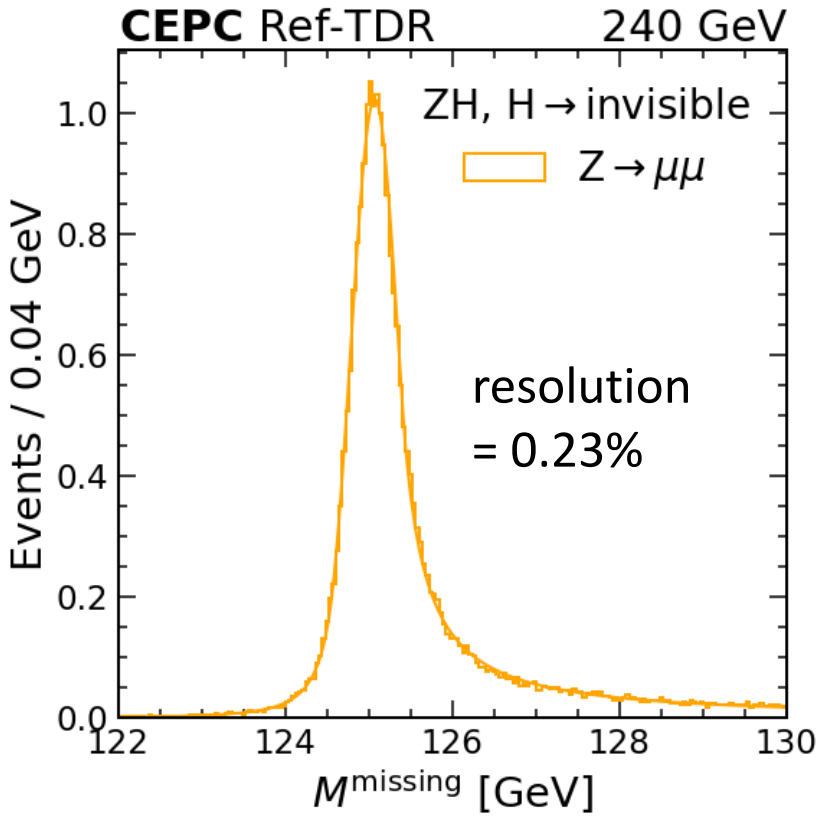
➢ Samples produced with CEPCSW 25.3.6

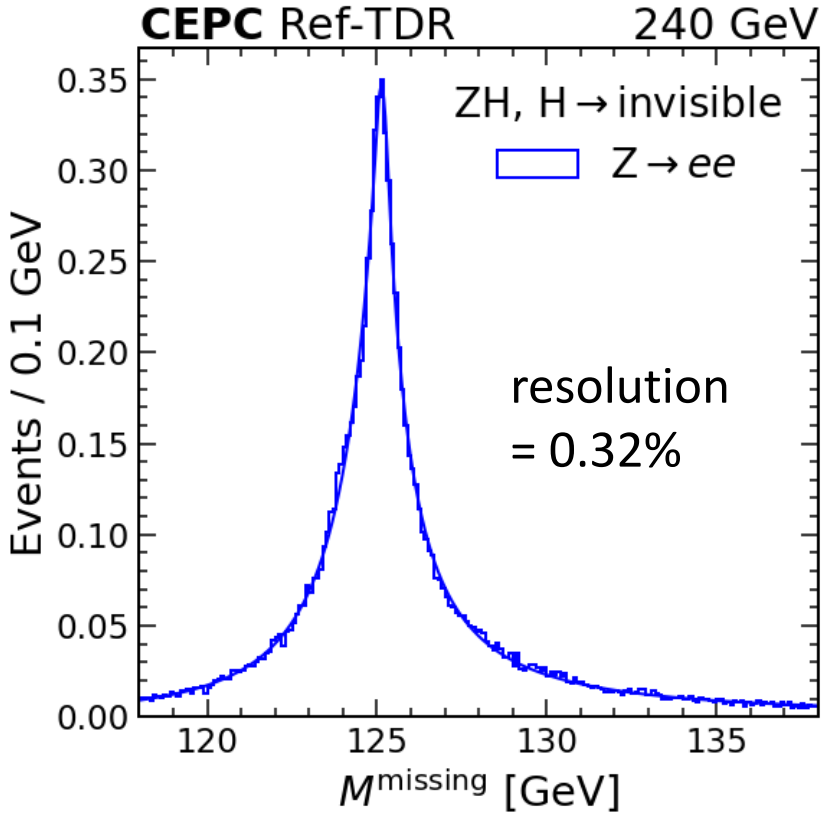
Experiments	Processes	Location at lxlogin	Events
Signal	Z(→ee/μμ/qq)H(→4ν)	/cefs/higgs/liugeliang/CEPC/202503/Production/Hinvi	100k per final state
4-fermion backgrounds	single-Z, single-W, Z-or-W, ZZ, WW, ZZ-or-WW (41 final states in total)	/cefs/higgs/zhangkl/Production/25036/4fermions	400k per final state
2-fermion backgrounds	ee, μμ, ττ, qq	/cefs/higgs/zhangkl/Production/25036/E240_e1e1(e2e2, e3e3, qq)	100k – 400 k per final state
ZH	Z(→ee/μμ/ττ/νν/qq) H(→incl)	/cefs/higgs/zhangkl/Production/25036/E240_*HX	1M per sample

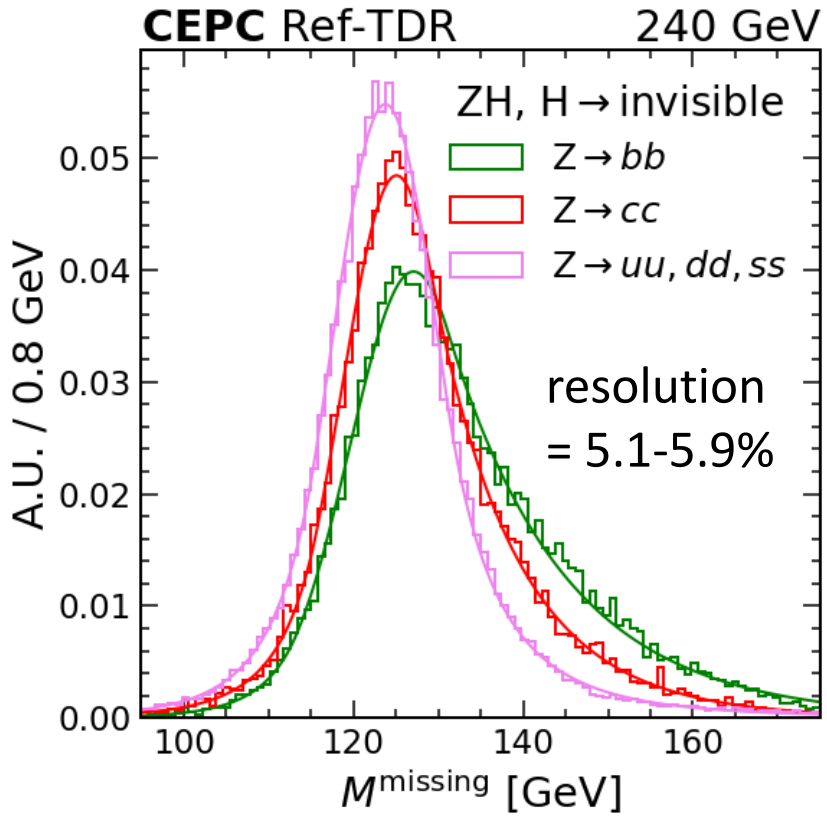
Key property

➢ Missing mass

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







Great discrimination thanks to **full reconstruction of missing information** and **good detector resolution**

Event selection

Baseline selection

- Selection criteria to have the very preselection of signals.
- Remain diagonal by selecting successively.

- 2μ channel:
  - Exactly two PFOs passing the BEST muon ID, with $|\cos\theta|<0.99$
  - Opposite charge.
  - $M_{\mu\mu} \in [40, 120]$ GeV
- 2e channel:
  - Exactly two PFOs passing the BEST electron ID, with $|\cos\theta|<0.99$
  - Opposite charge.
  - $M_{ee} \in [40, 120]$ GeV
- 2q channel:
  - Due to the exclusive ee-kt jet clustering algorithm, di-jet kinematics are equal to visible kinematics.
  - $M^{visible} \in [30, 130]$ GeV
  - $p^{visible} \in [10, 80]$ GeV

Event selection

Kinematic selection

- Further suppress backgrounds while keeping a high signal efficiency

- 2mu 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 $N_{neutral} > 2$

Selection cutflow

process		signal	2(μ/e/q)+2v	2f	visible H	others
2μ	total yield	1.44E+02	5.68E+06	1.78E+09	4.07E+06	3.79E+08
	Baseline sel	96.11%	32.01%	2.35%	2.55%	0.88%
	Kinematic sel	97.97%	19.78%	3.40%	0.44%	5.31%
	selected	1.35E+02	3.59E+05	1.42E+06	4.55E+02	1.78E+05
2e	total yield	1.49E+02	5.57E+06	1.78E+09	4.07E+06	3.79E+08
	Baseline sel	83.75%	41.68%	1.03%	1.96%	1.60%
	Kinematic sel	95.34%	23.04%	3.35%	2.19%	5.77%
	selected	1.19E+02	5.35E+05	6.13E+05	1.75E+03	3.49E+05
2q	total yield	2.90E+03	7.39E+06	1.78E+09	4.07E+06	3.77E+08
	Baseline sel	98.97%	66.07%	9.24%	19.76%	8.35%
	Kinematic sel	95.41%	38.05%	37.29%	37.82%	12.87%
	selected	2.74E+03	1.86E+06	6.13E+07	3.04E+05	4.05E+06

Machine learning algorithm

XGBoost model trained in each channel to distinguish signal v.s. background

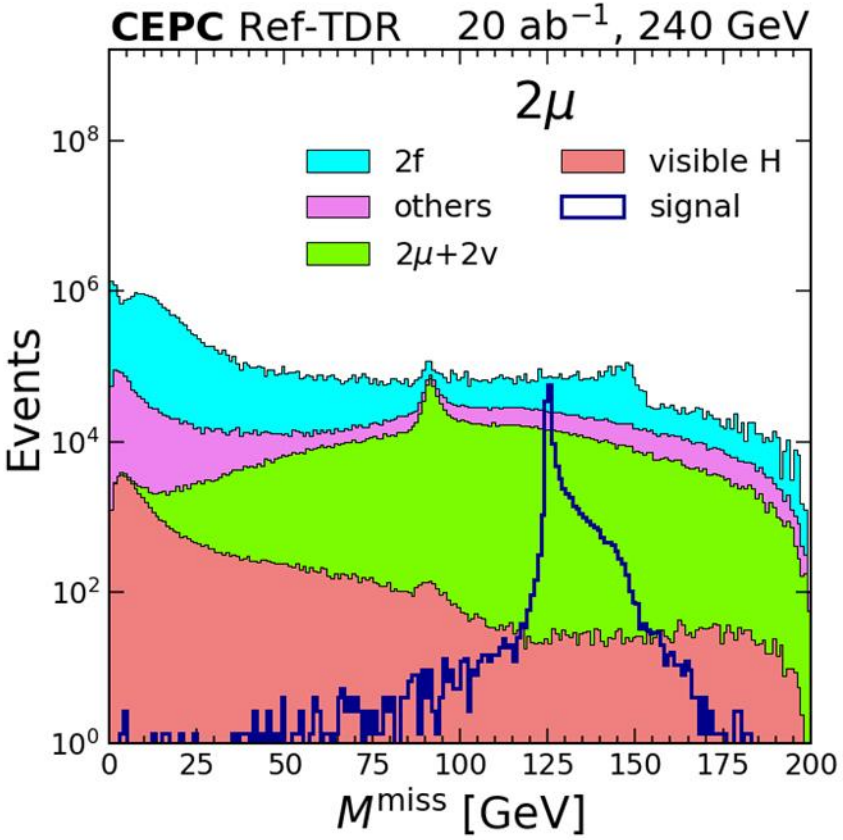
Input features

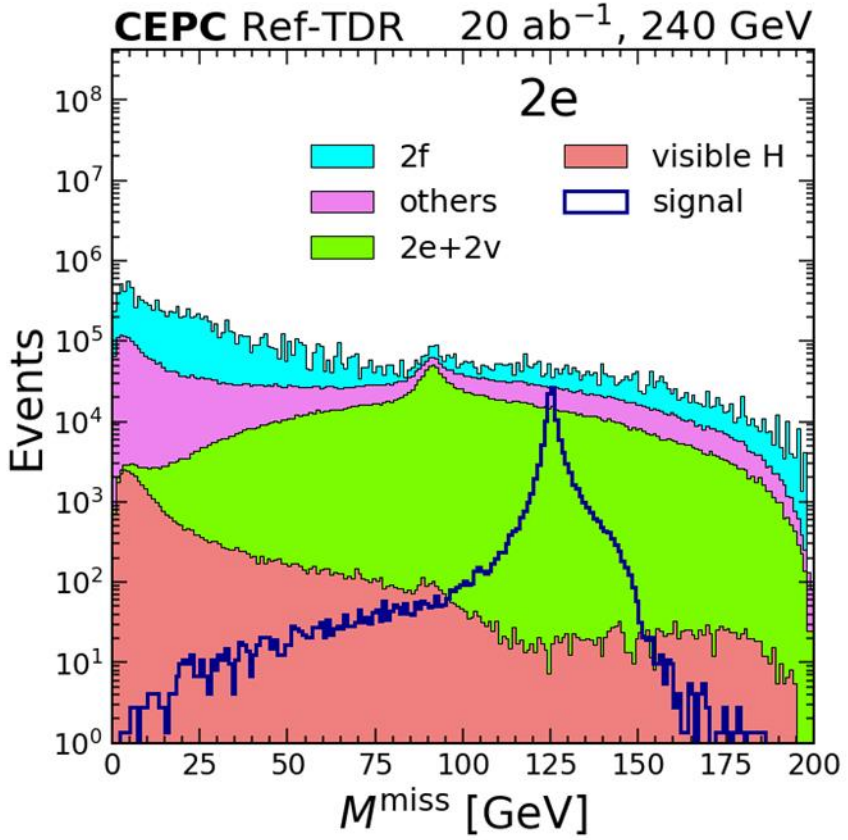
- 2mu channel:
  - $E_{\mu\mu}, M_{\mu\mu}, P_{\mu\mu}, P_{\mu\mu}^{T}$
  - M_{recoil}
  - $E_{visible}, M_{visible}, P_{visible}, P_{visible}^{T}$
  - $M_{missing}$
  - $\Delta\phi_{\mu\mu}, \Delta R_{\mu\mu}$
  - $D_0^{\mu_1}, D_0^{\mu_2}, Z_0^{\mu_1}, Z_0^{\mu_2}$
  - $N_{charged}, N_{neutral}, E_{neutral}$
- 2e channel:
  - Same as 2mu, except changing μ to e
- 2q channel:
  - $E_{visible}, M_{visible}, P_{visible}, P_{visible}^{T}$
  - $M_{missing}$
  - $\Delta\phi_{jj}, \Delta R_{jj}$
  - $N_{charged}, N_{neutral}$
  - $N_{charged}^{j_1}, N_{charged}^{j_2}$
  - $ECF_2^{j_1}, ECF_2^{j_2}$ (energy correlation function)
  - $\left(\frac{\tau 3}{\tau_1}\right)^{j_1}, \left(\frac{\tau 3}{\tau_1}\right)^{j_2}$ (N-subjettiness)

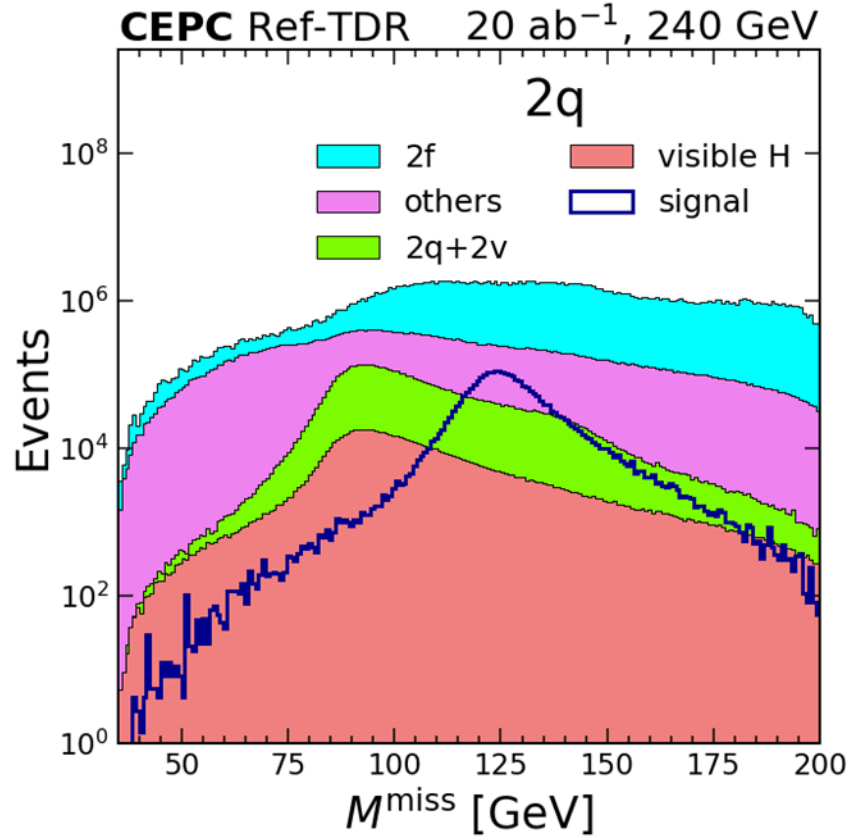
More input features than kinematic selections

- Some features not efficient enough but can provide discrimination.

Important features: missing mass

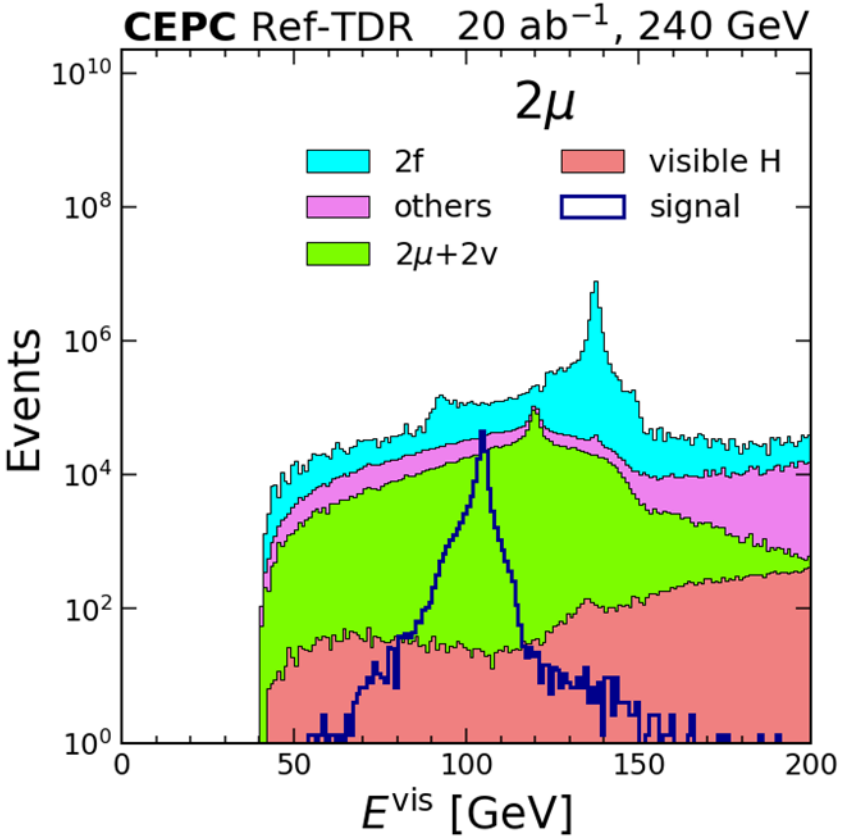


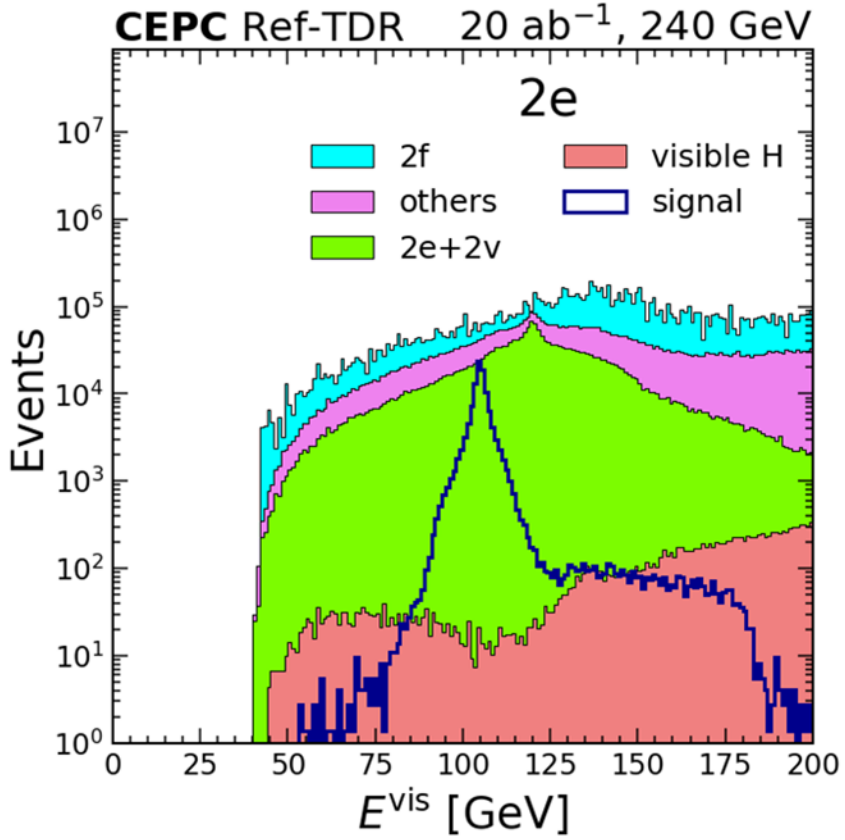


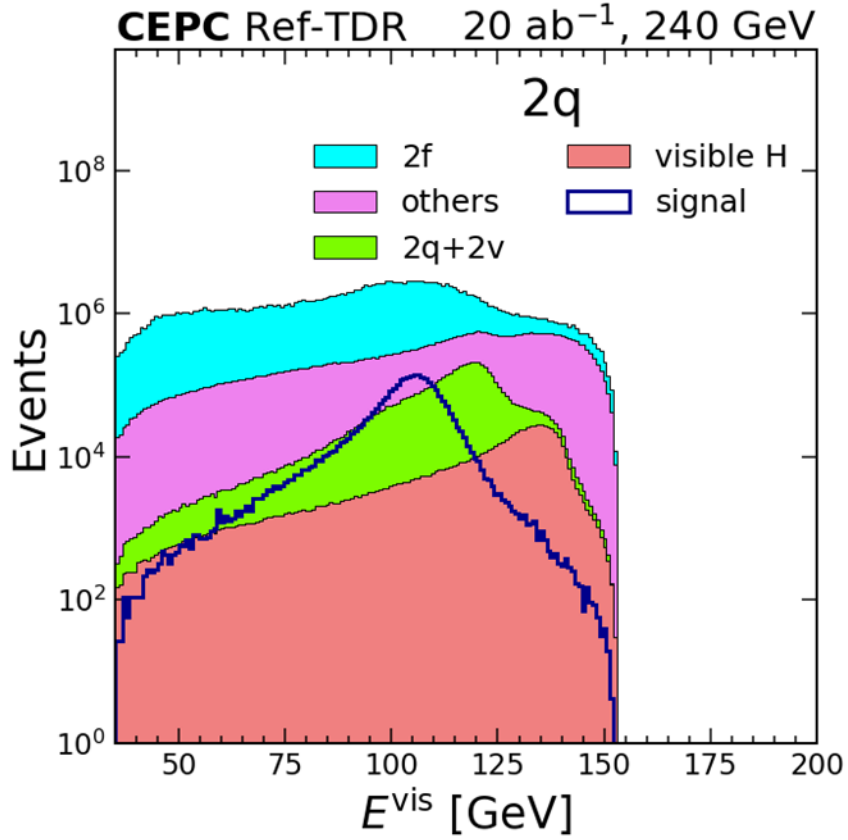


- The **signal** distributed around 125 GeV (BR(H→invisible) set to be 1).
- 2(μ/e/q)+2v distributed around 91 GeV.
- Other backgrounds distributed around 0 or flatly.

Important features: visible energy

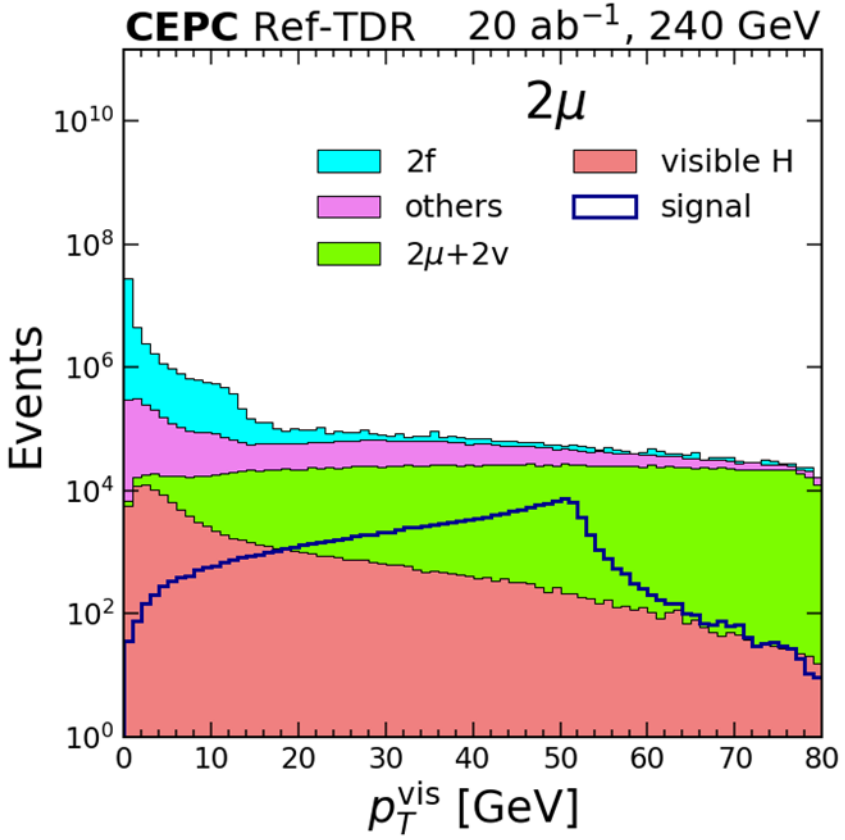


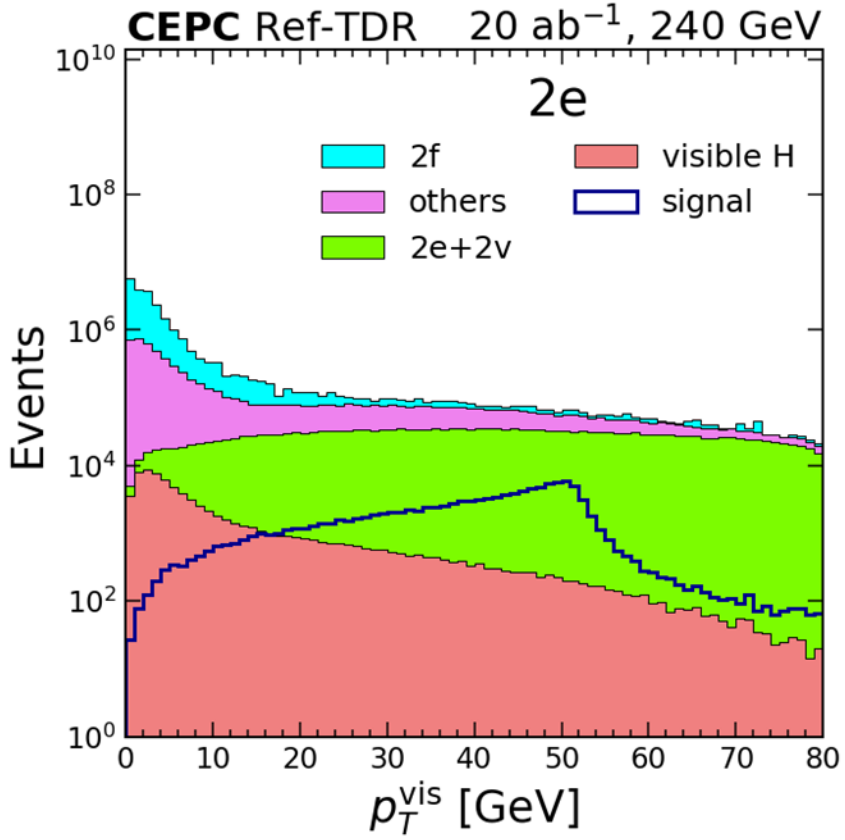


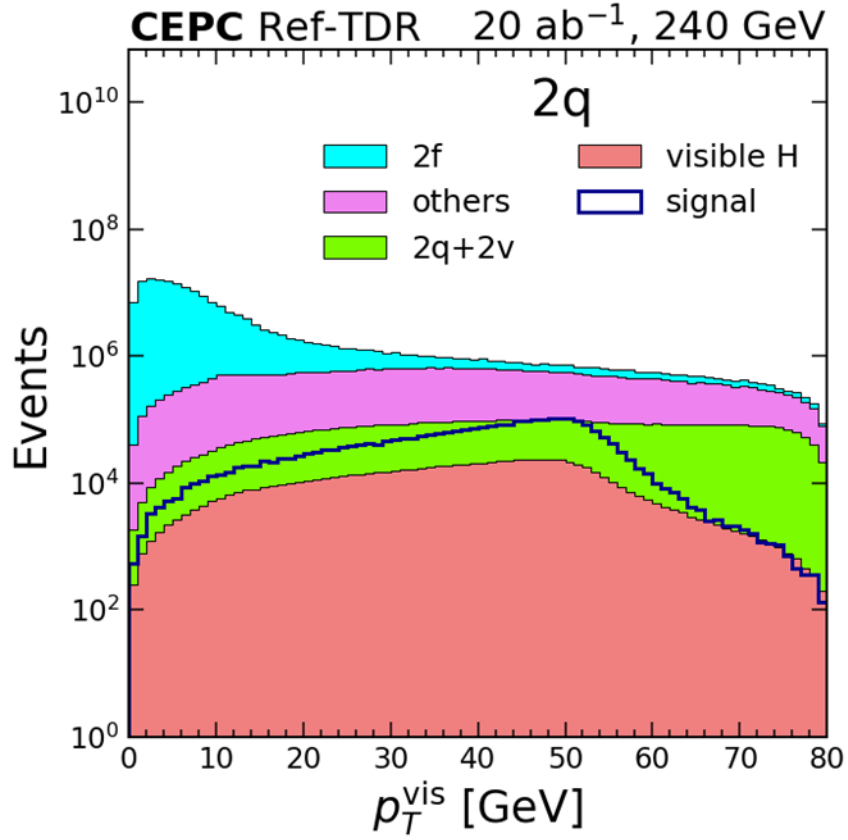


- The **signal** distributed around 105 GeV.
- 2(μ/e/q)+2v distributed around 120 GeV.
- Other backgrounds mostly distributed flatly.

Important features: visible pt

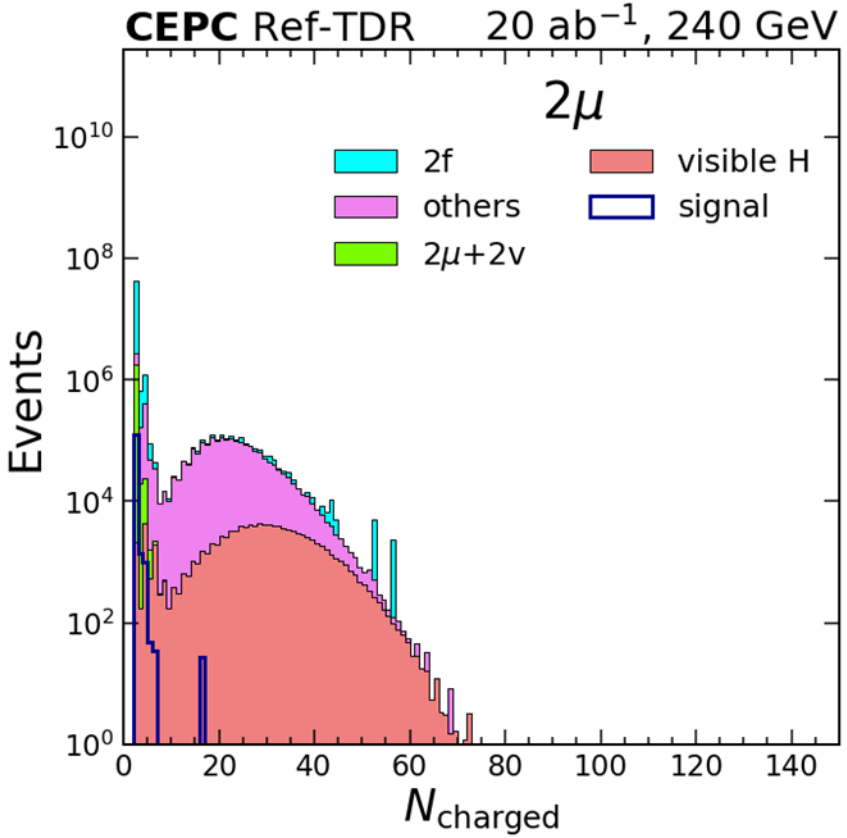


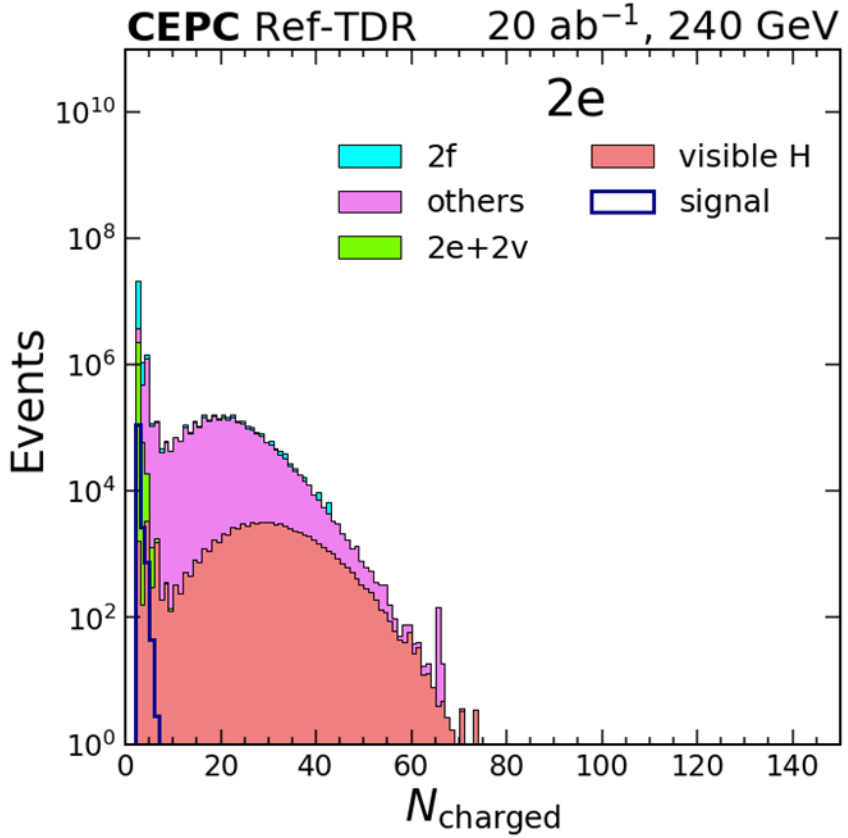


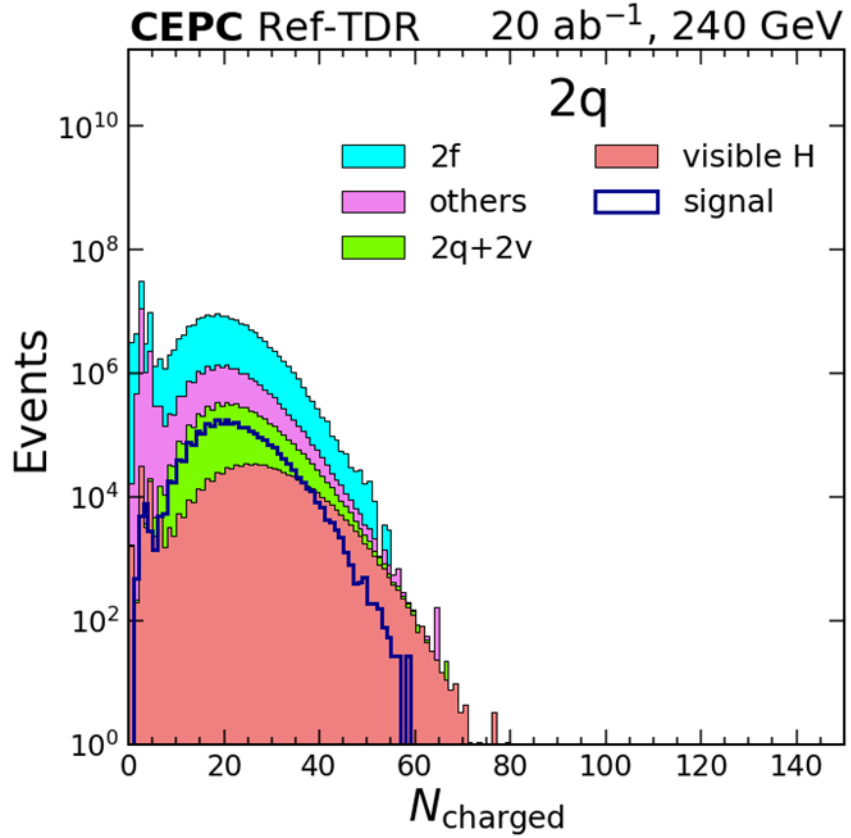


- The **signal** distributed around 50 GeV.
- 2(μ/e/q)+2v distributed flatly.
- 2f distributed around 0: mostly in high |cosθ| region.

Important features: number of charged PFOs

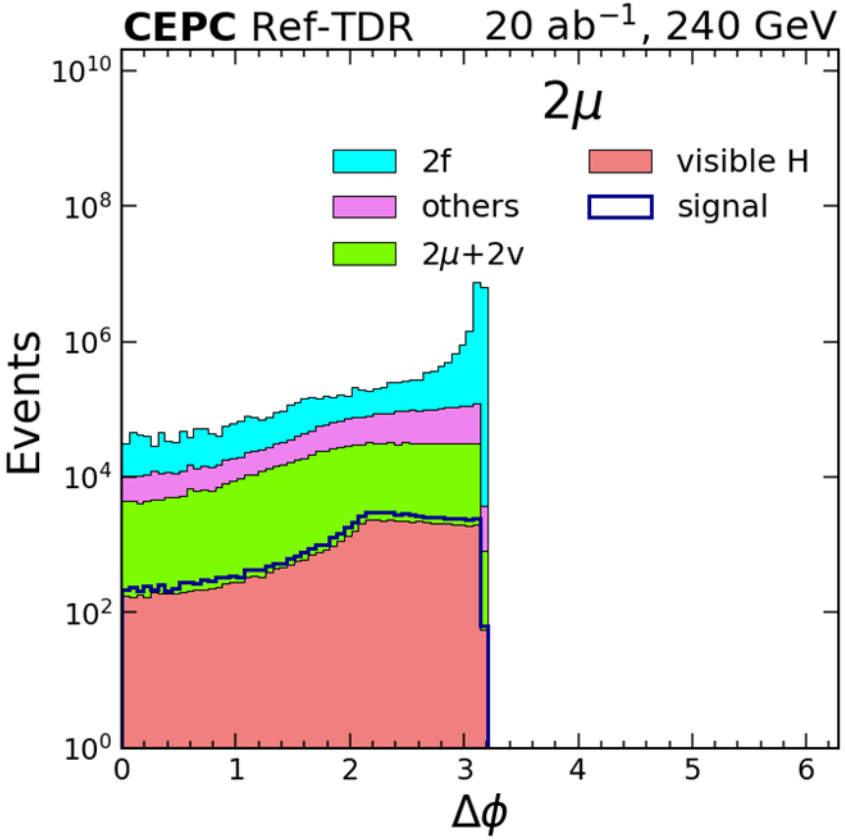


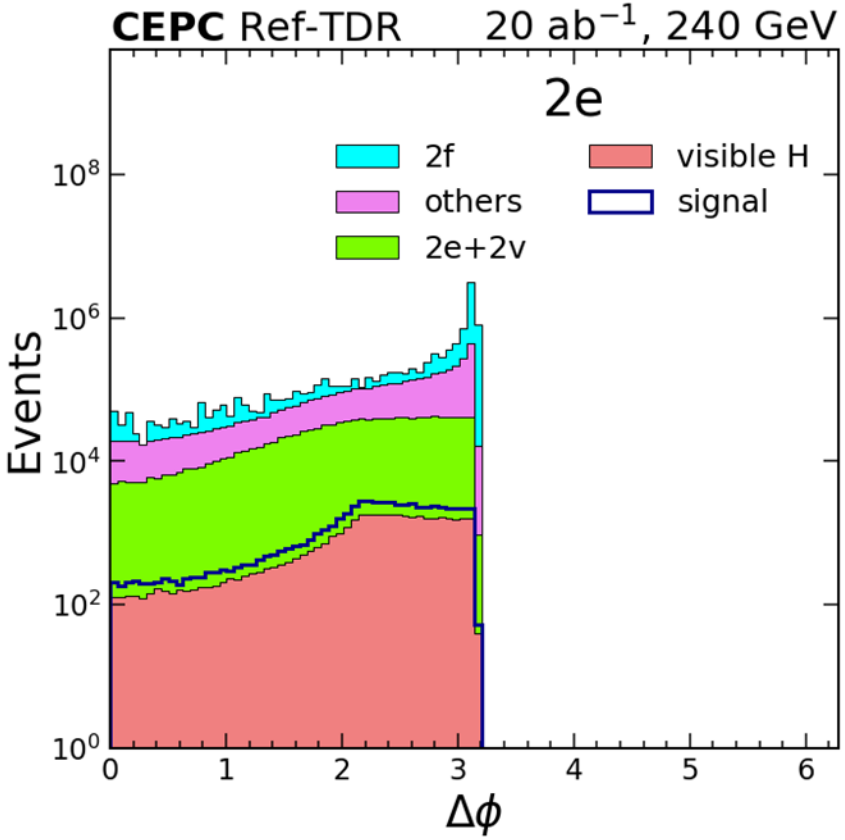


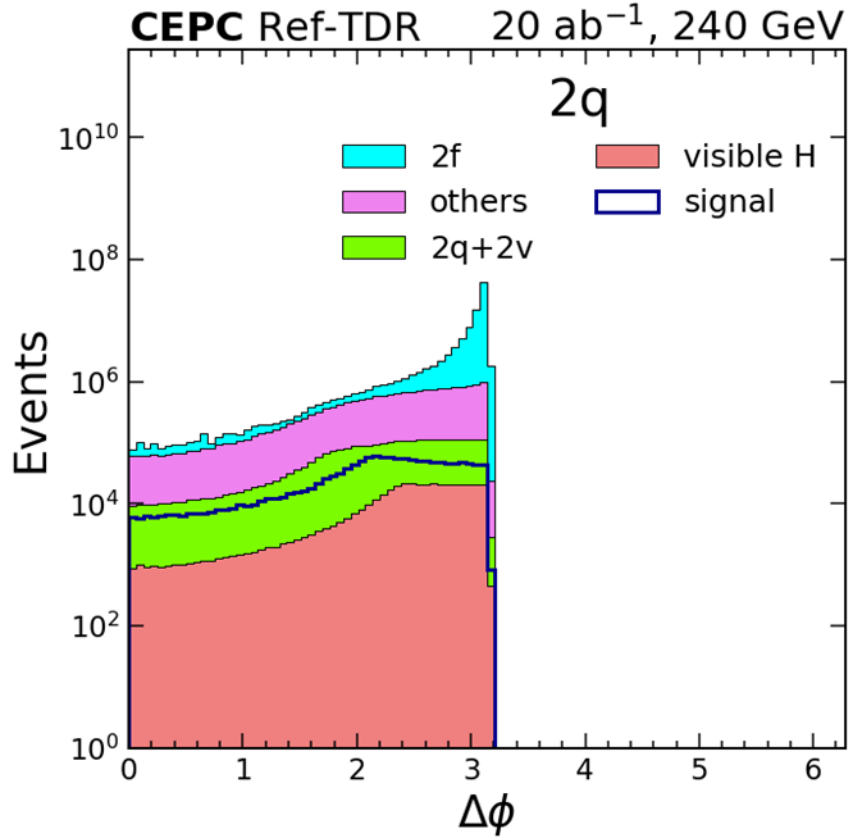


- The **signal** has different signatures in different channels.

Important features: ΔΦ(ll / jj)



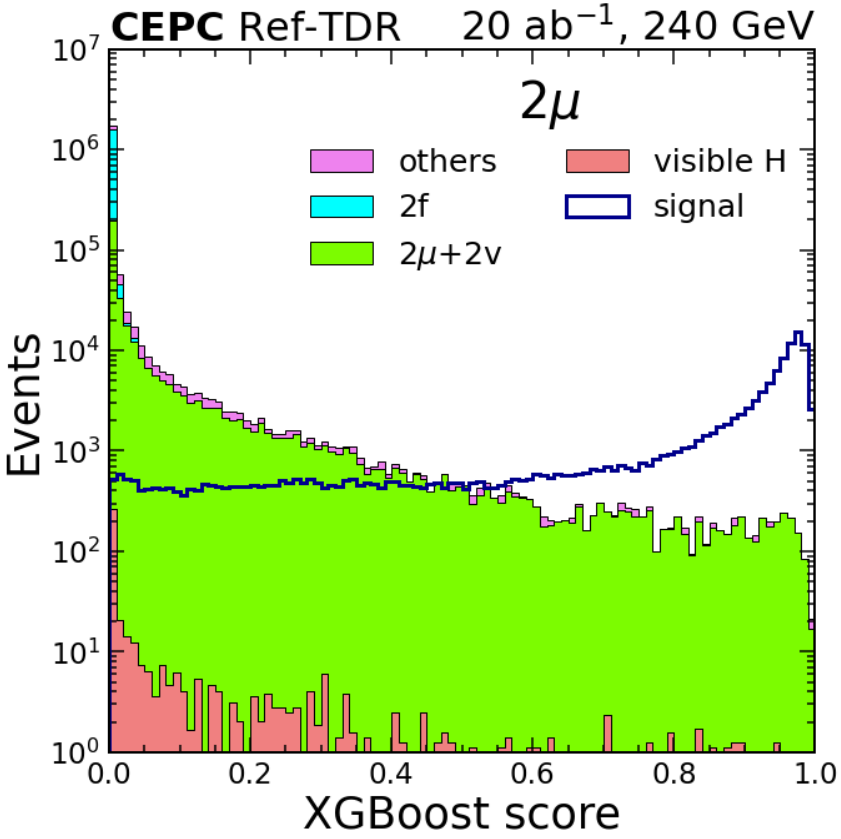


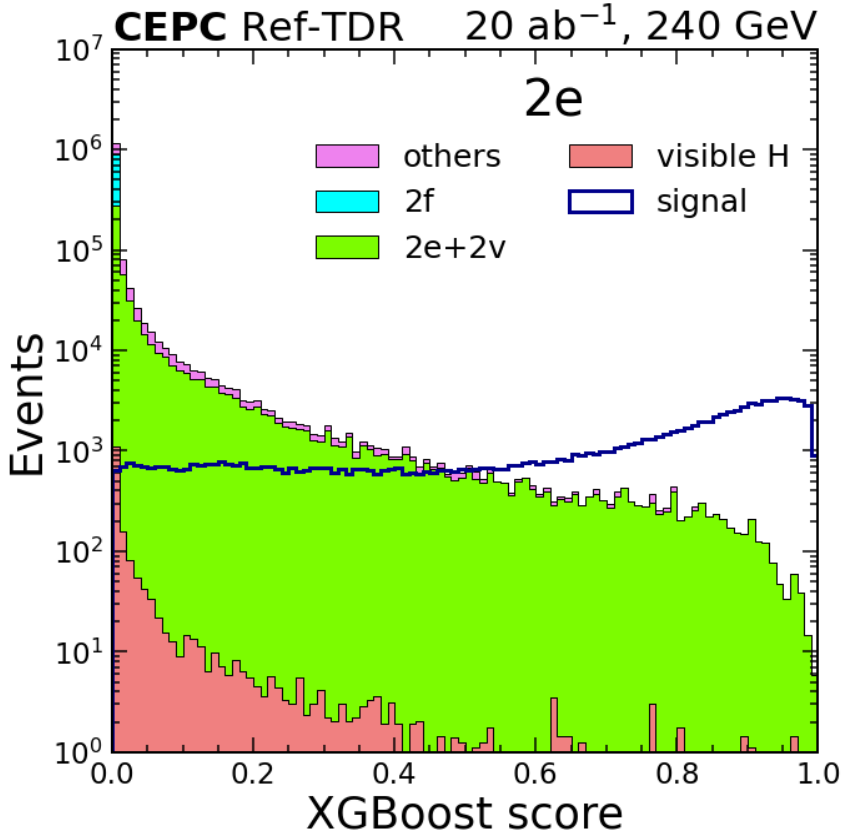


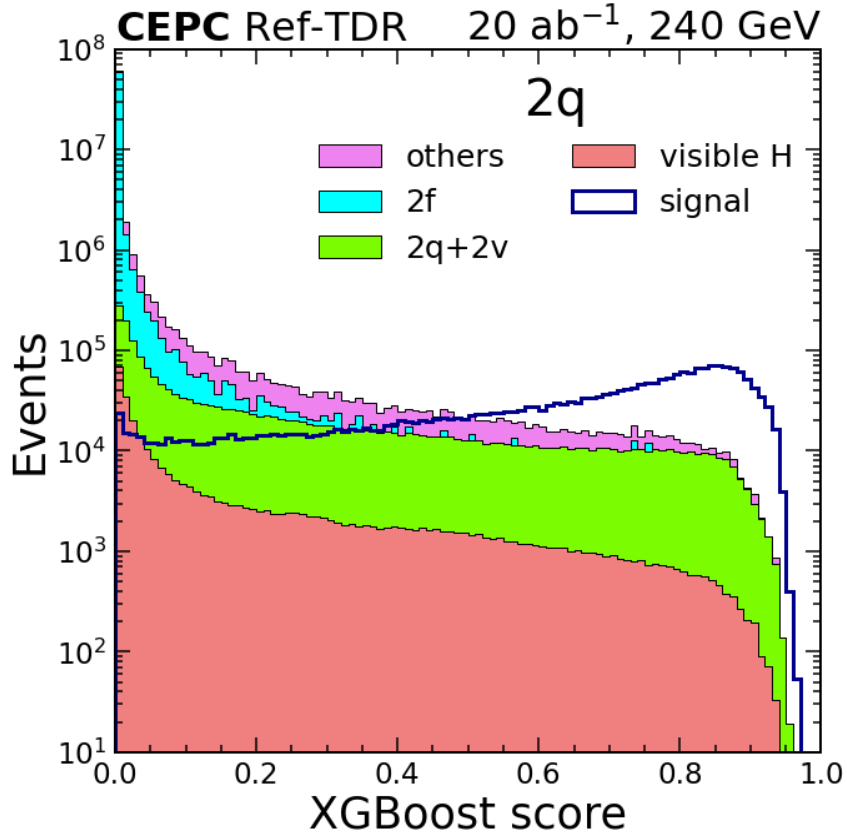
- 2f distributed around 3.14: mostly back-to-back.

ML-based discriminant

XGBoost score distributions







- Main backgrounds contaminating with signals: 2(μ/e/q)+2v

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
- **No systematic considered yet**
  - Systematic uncertainties < 1%, dominated by statistical uncertainties
- **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				20 ab-1		
	uncertainty (SM)	CDR uncertainty	significance (SM)	UL (BSM)	uncertainty (SM)	significance (SM)	UL (BSM)
2μ	-80.4%/+84.3%	222%	1.25σ	0.179%	-43.1%/+44.1%	2.36σ	0.093%
2e	-100.0%/+124.4%	428%	0.86σ	0.266%	-62.6%/+64.9%	1.62σ	0.137%
2q	-57.6%/+57.8%	90%	1.74σ	0.121%	-30.5%/+30.6%	3.28σ	0.064%
combine	-43.7%/+44.7%	82%	2.31σ	0.092%	-23.2%/+23.4%	4.36σ	0.049%

Close to discovery level!

Much better sensitivities than the CDR studies !

- Thanks to the MVA algorithm

Conclusion and prospects

Preliminary searches for Higgs boson invisible decays are performed

- **Full simulation** of the CEPC reference detector
- **Missing mass with good resolution** as the key feature
- Much improved sensitivities thanks to the **ML-based algorithm**
- Strong proof of the **overall detector performances**

Future progress is expected

- Further tune the **kinematic selection** (i.e. cut on visible pt)
- Categorization based on **jet flavor**
- Inclusion of **systematic** uncertainties

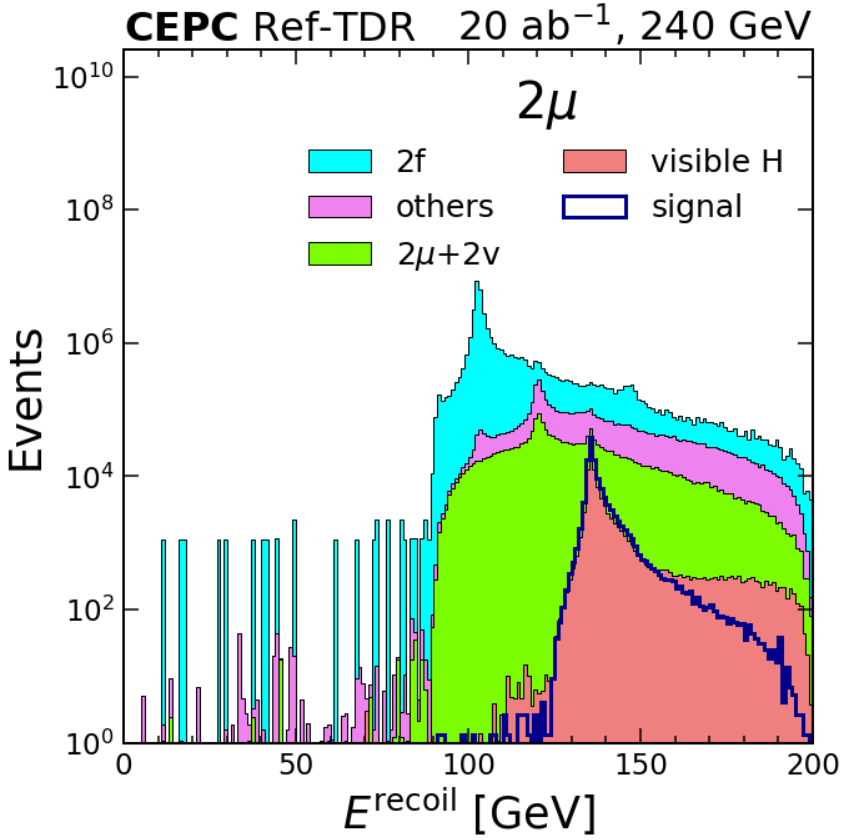
Can we reach 5 σ ?

Thanks for your attention!

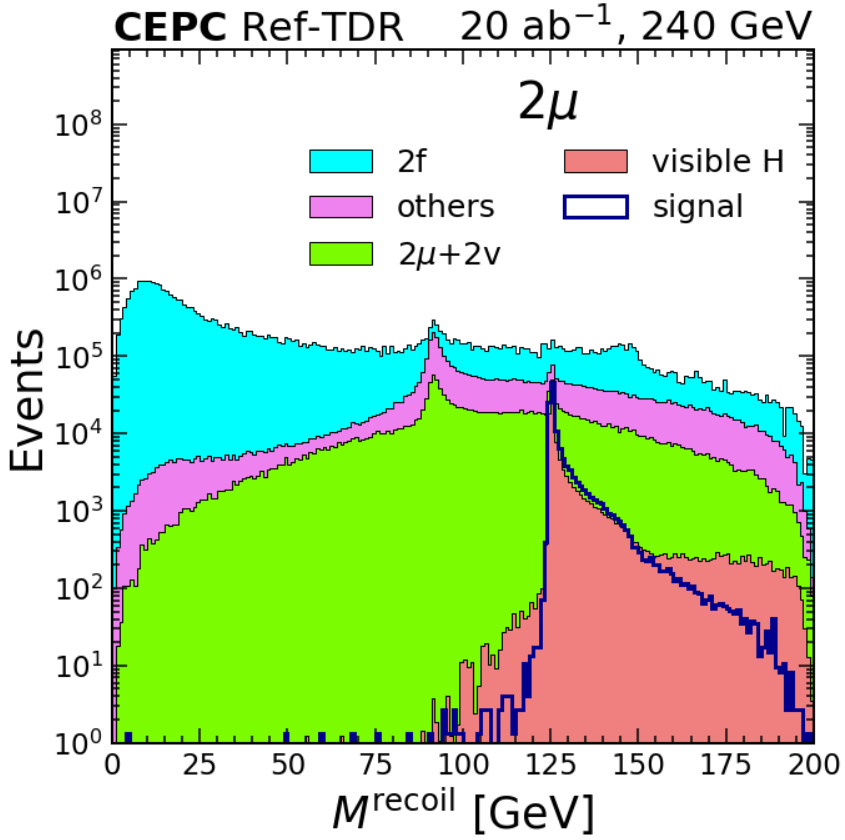
Backup

Kinematic variables: 2μ

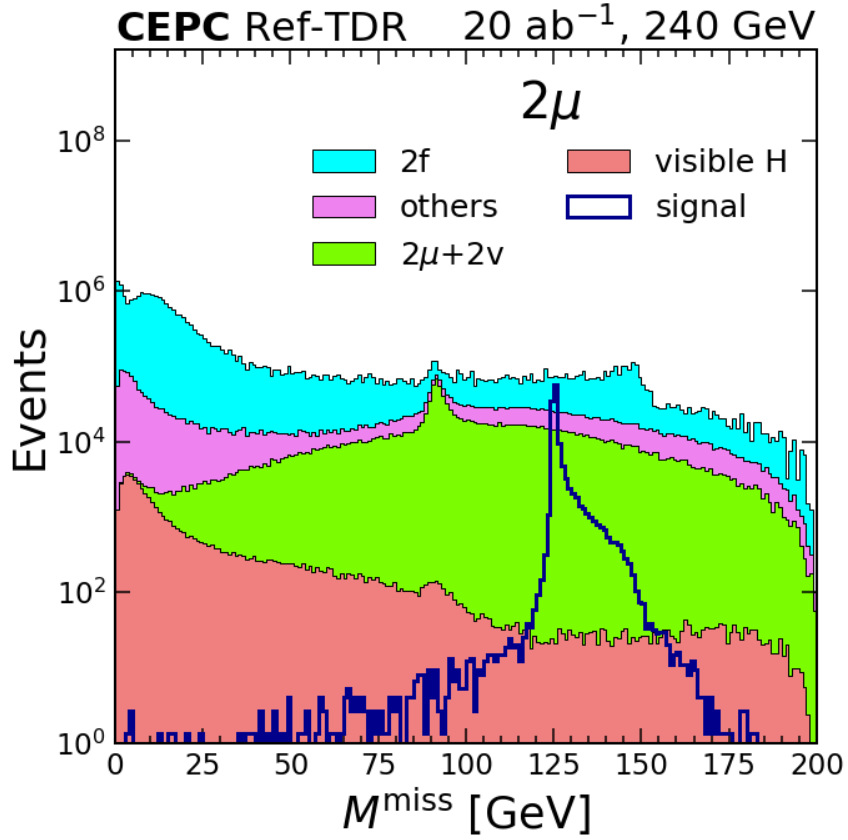
$E_{recoil} > 125$ GeV



$M_{recoil} > 110$ GeV

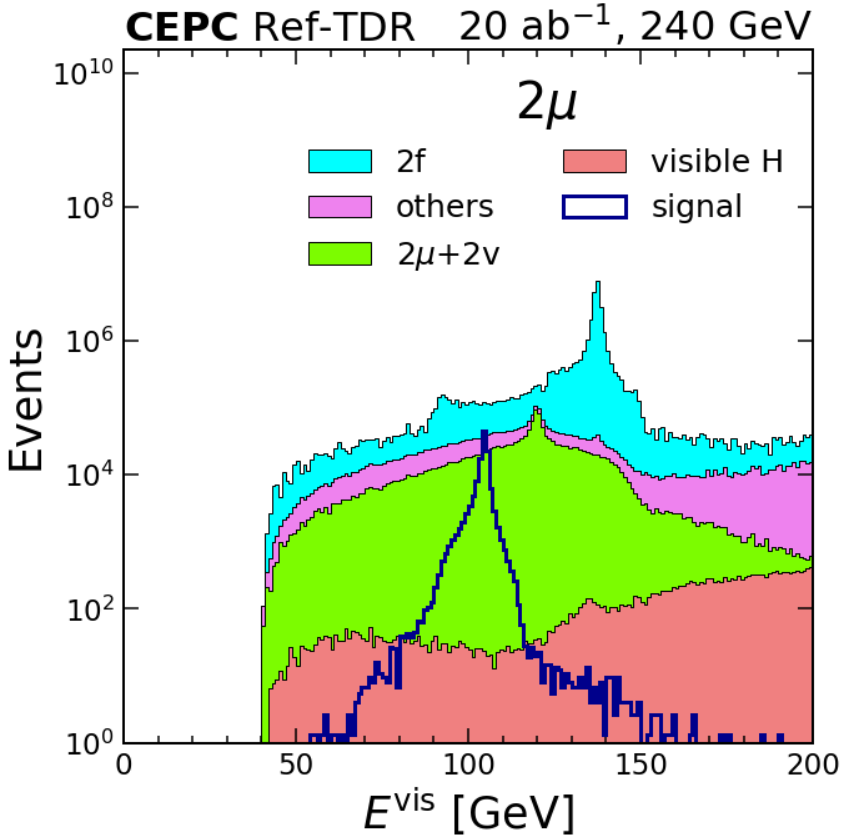


$M_{missing} \in (110,150)$ GeV

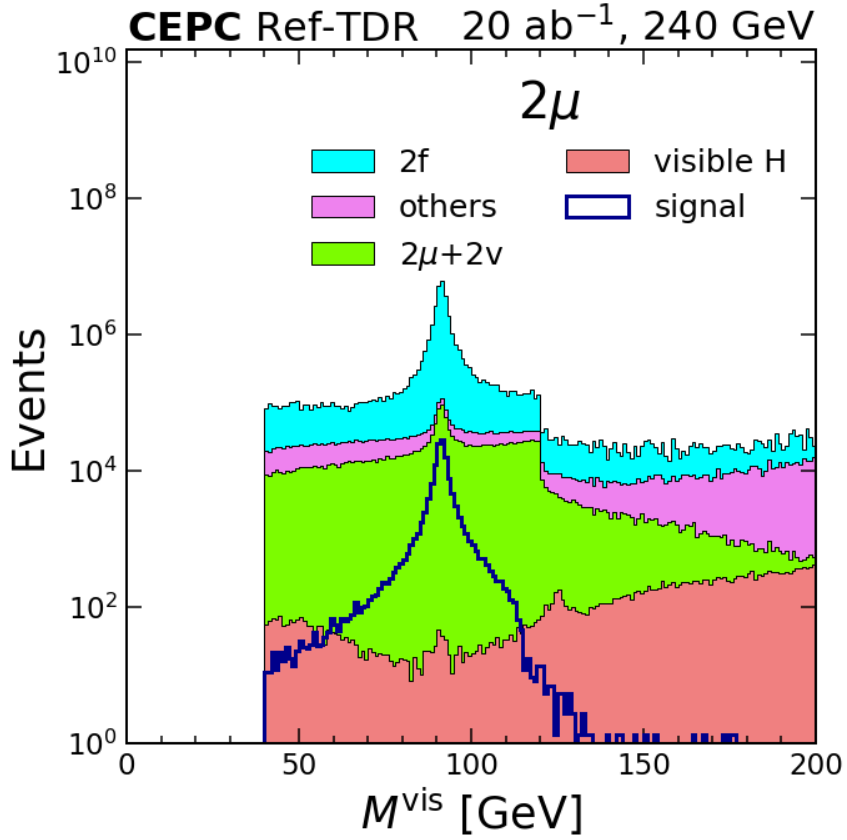


Kinematic variables: 2μ

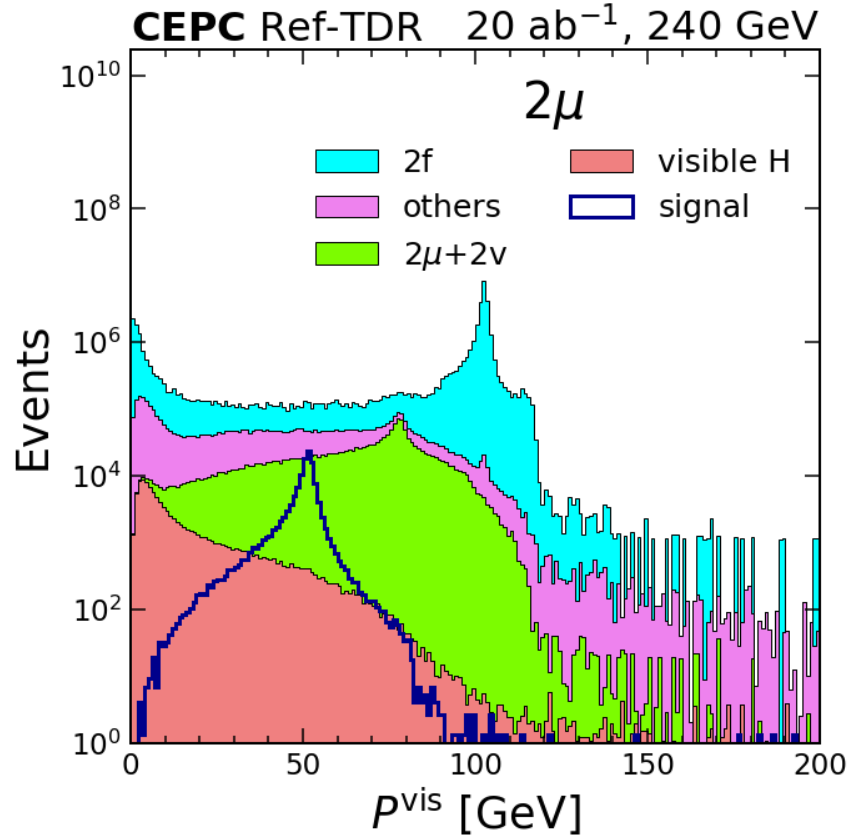
$E_{visible} \in (80,120)$ GeV



$M_{visible} < 120$ GeV

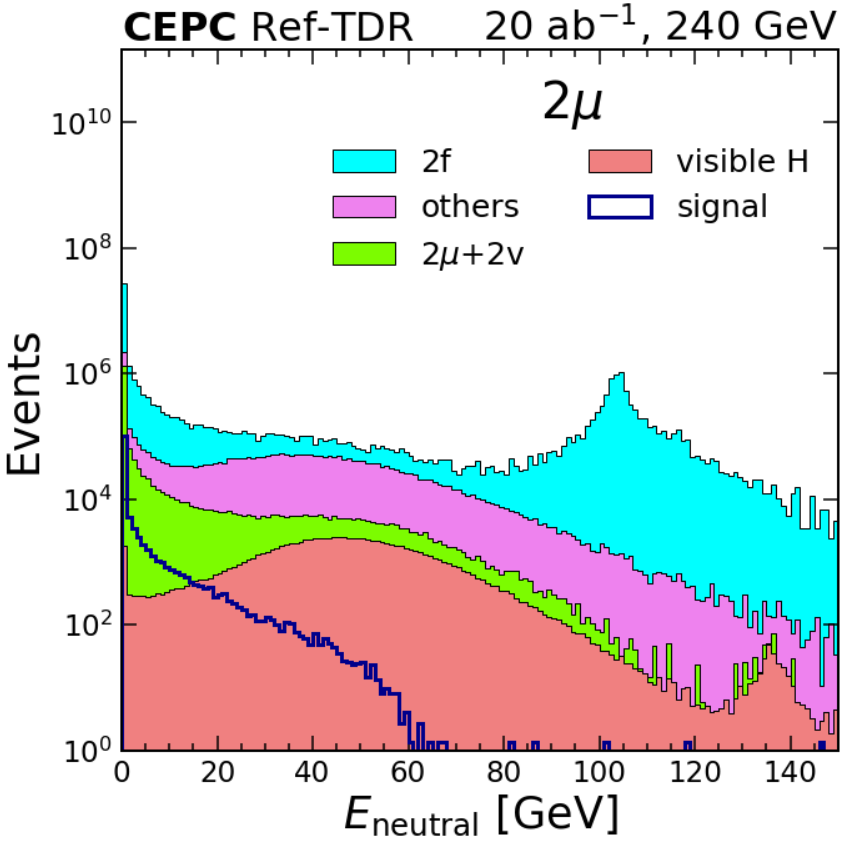


$P_{visible} \in (20,70)$ GeV

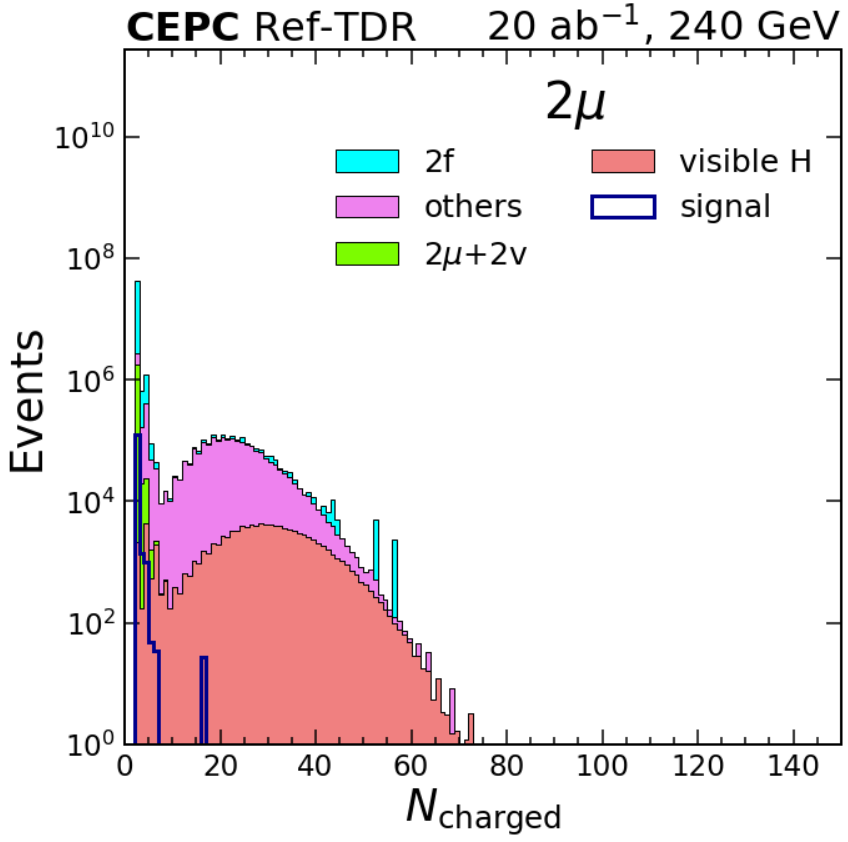


Kinematic variables: 2μ

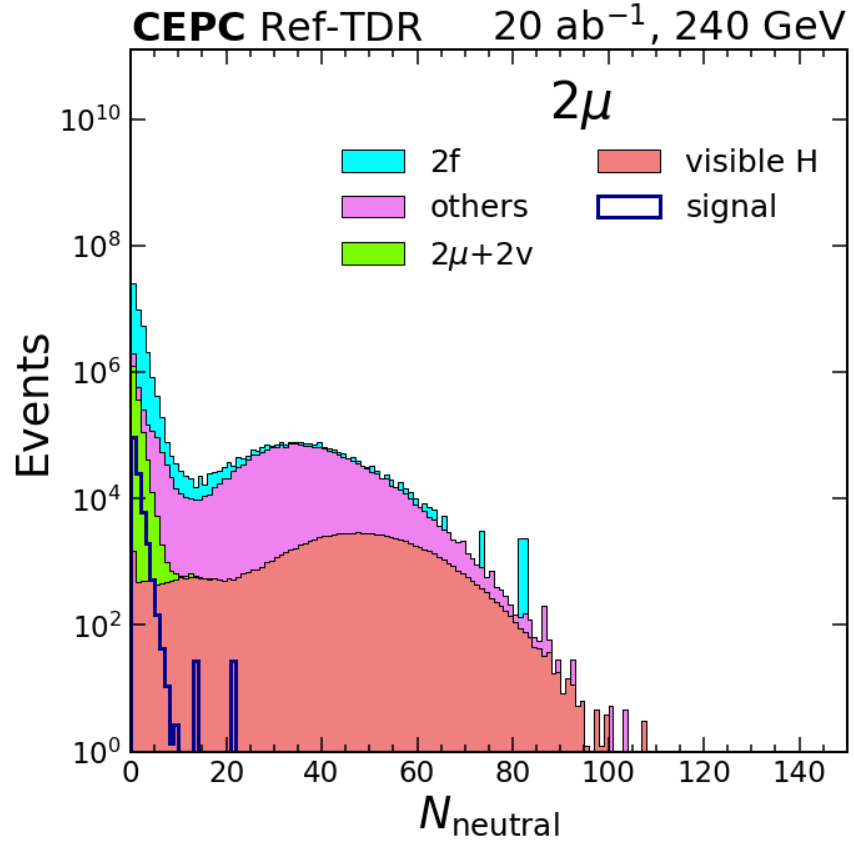
$E_{neutral} < 60$ GeV



$N_{charged} < 7$

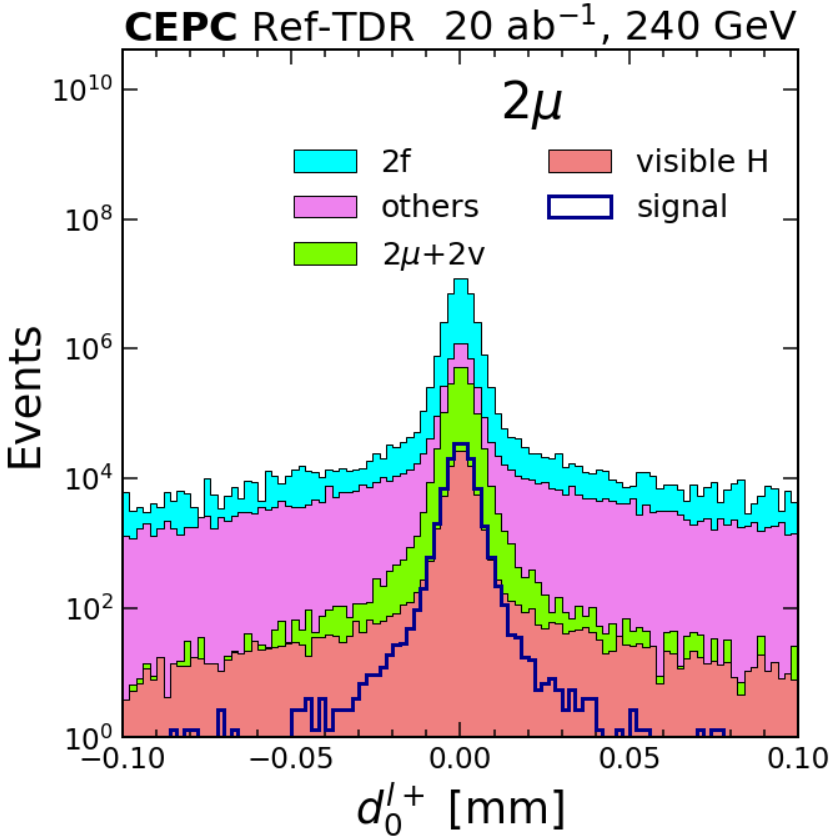


$N_{neutral} < 10$

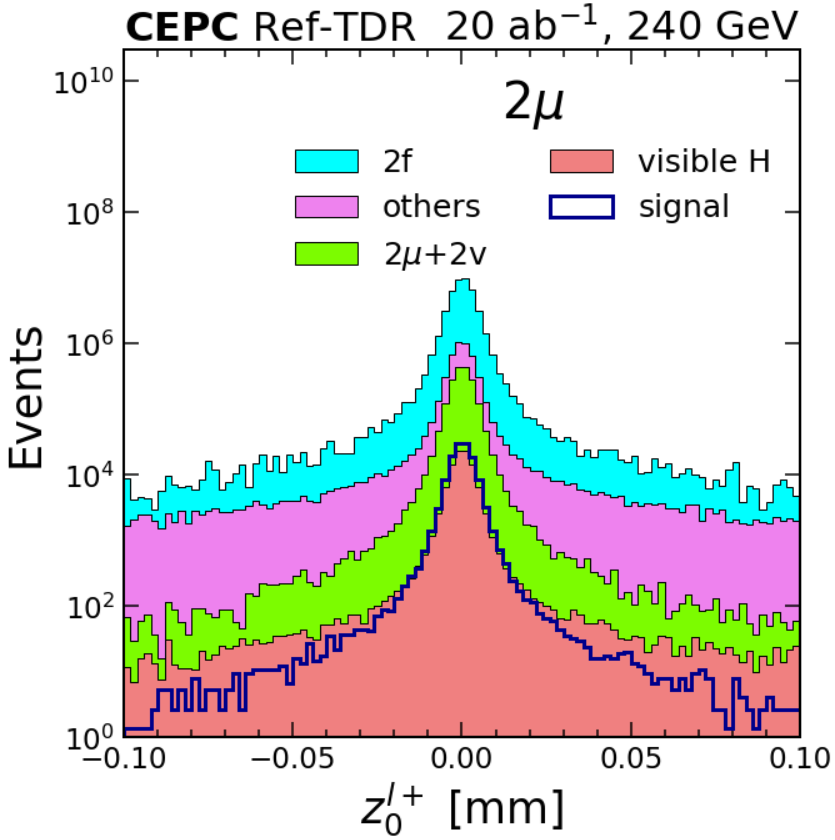


Kinematic variables: 2μ

$|D_0| < 0.05$ mm

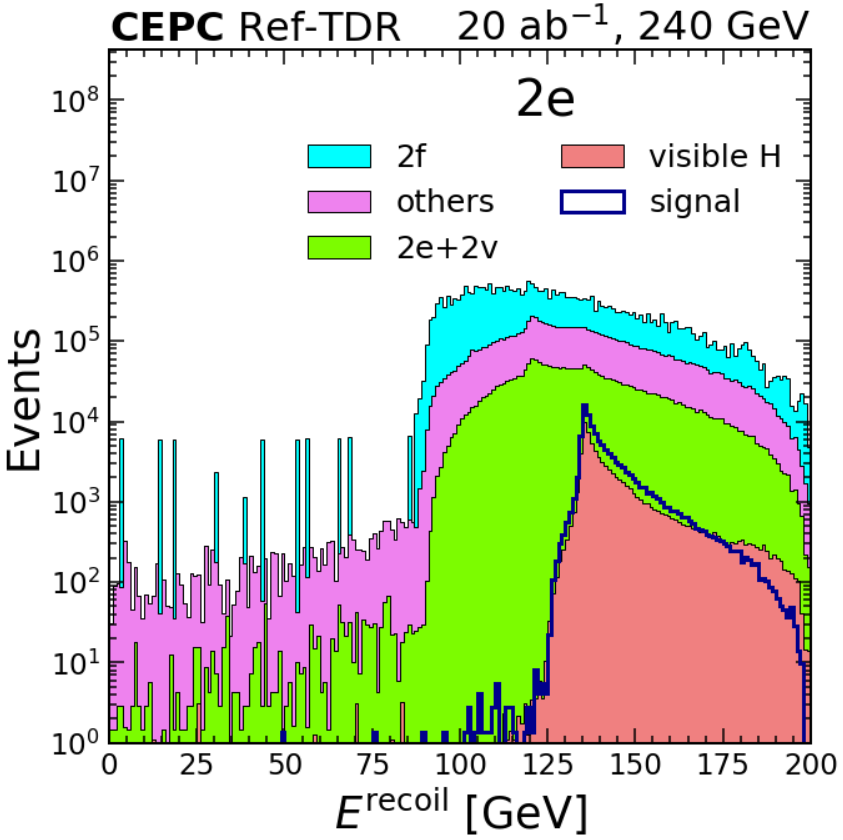


$|Z_0| < 0.1$ mm

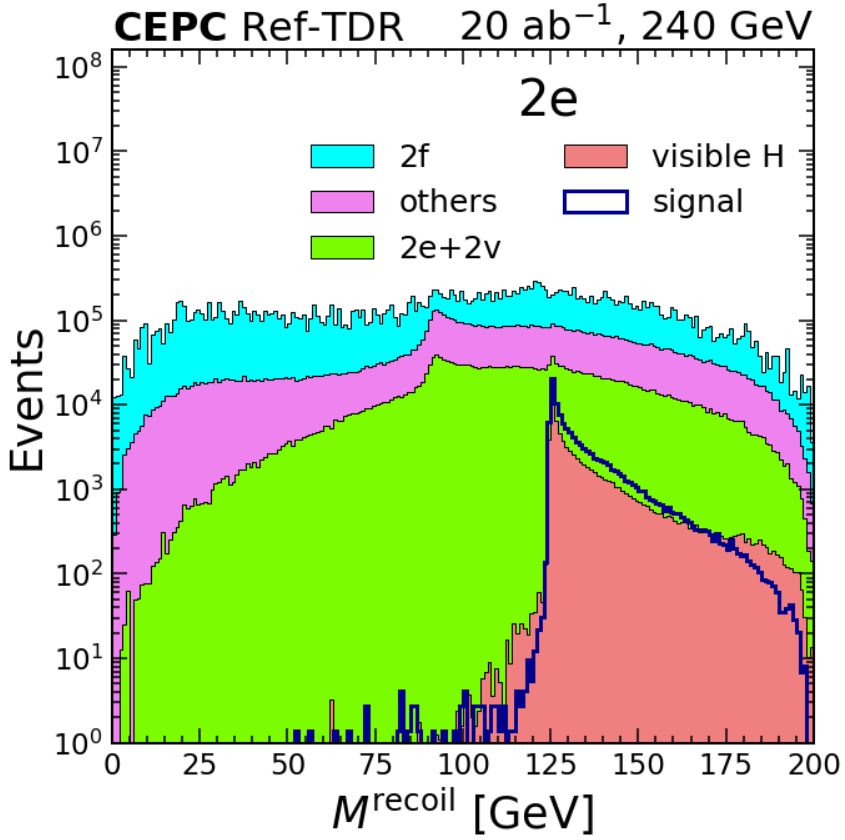


Kinematic variables: 2e

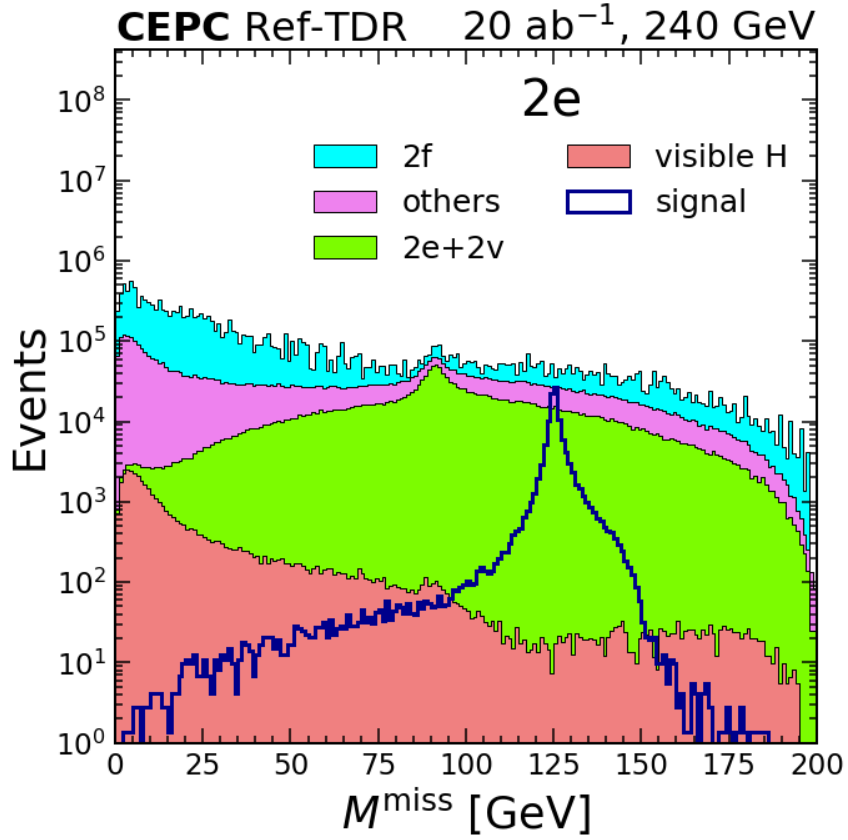
$E_{recoil} > 125$ GeV



$M_{recoil} > 120$ GeV

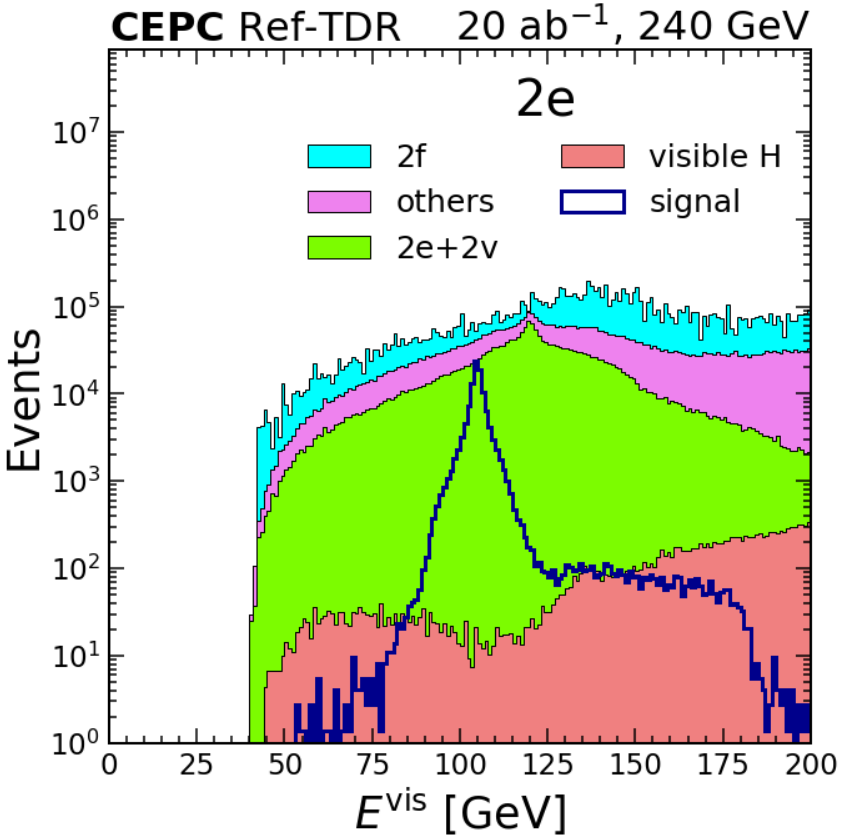


$M_{missing} \in (50,160)$ GeV

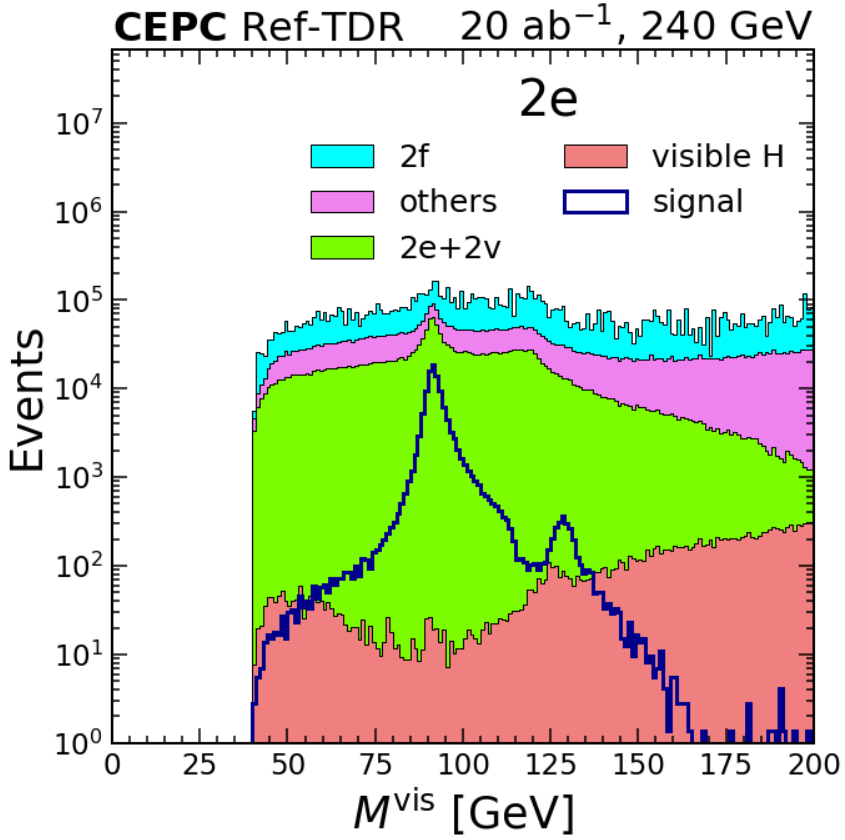


Kinematic variables: 2e

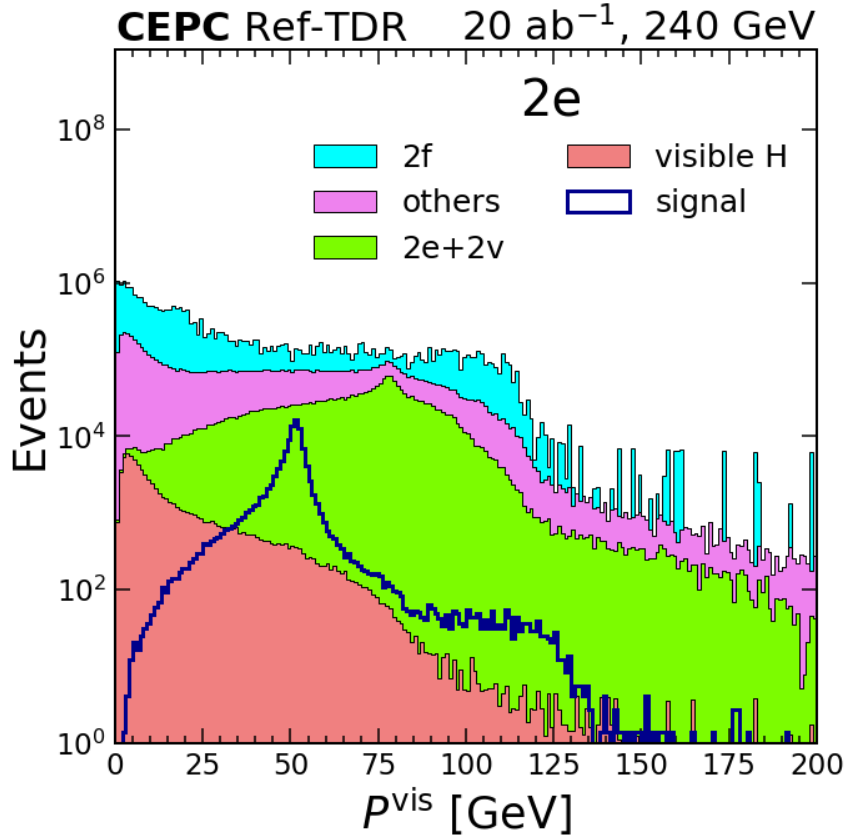
$E_{visible} \in (80,190)$ GeV



$M_{visible} < 160$ GeV

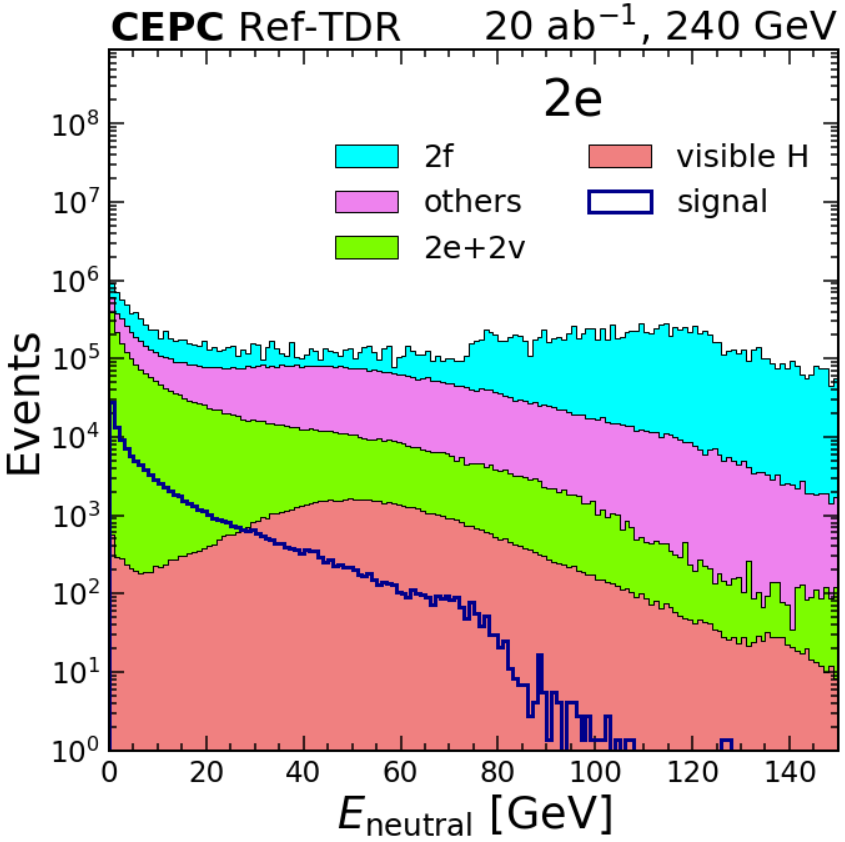


$P_{visible} \in (20,70)$ GeV

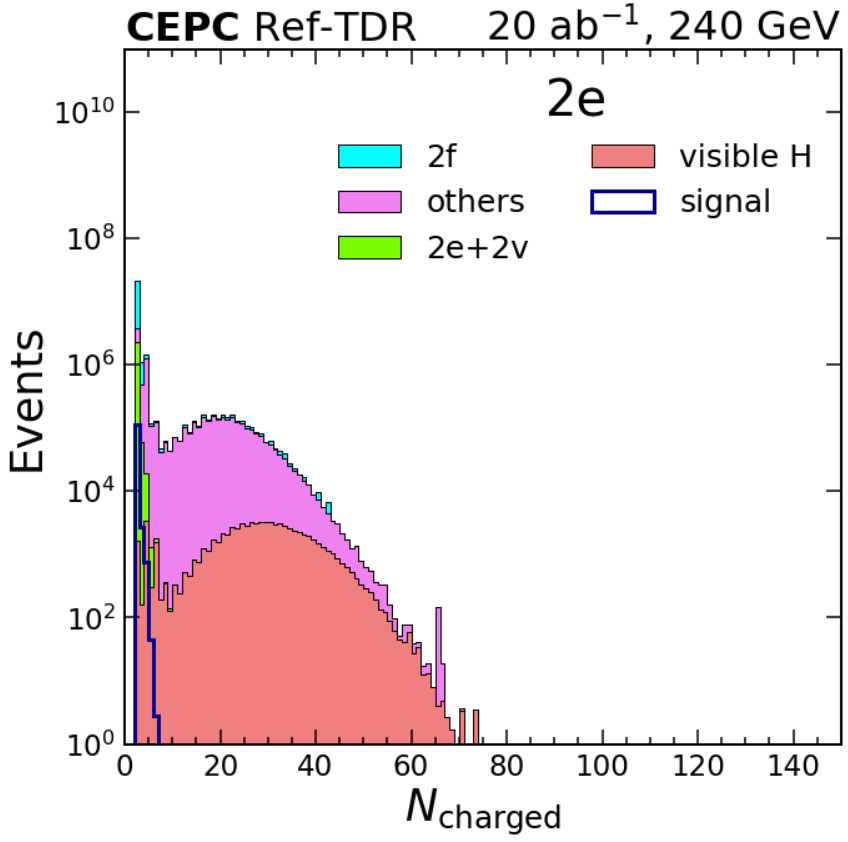


Kinematic variables: 2e

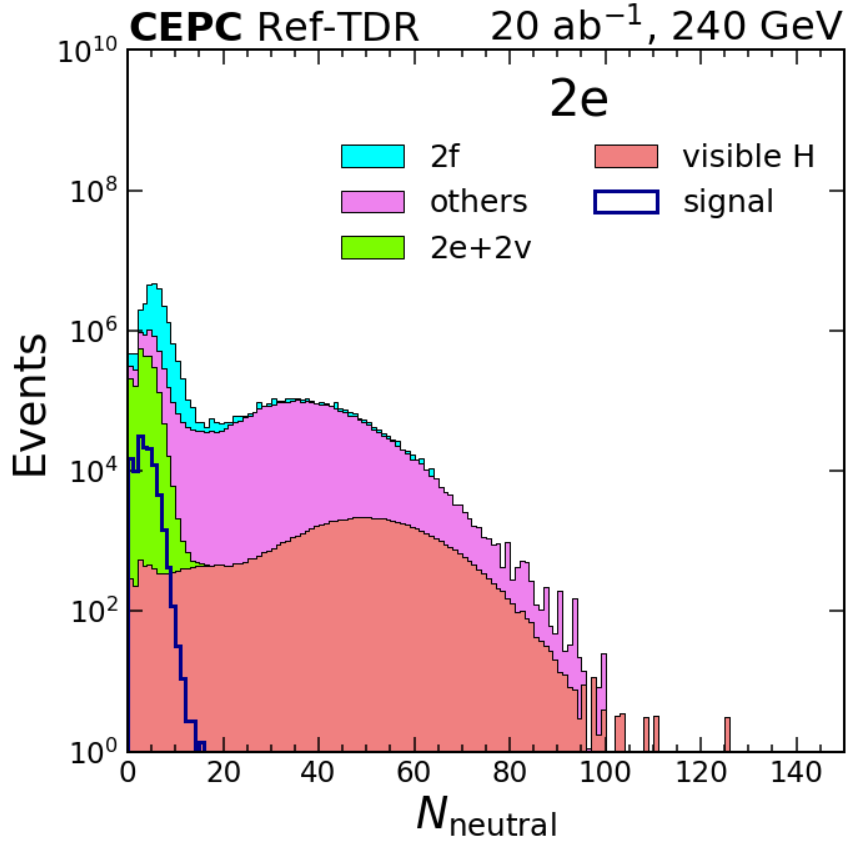
$E_{neutral} < 90$ GeV



$N_{charged} < 7$

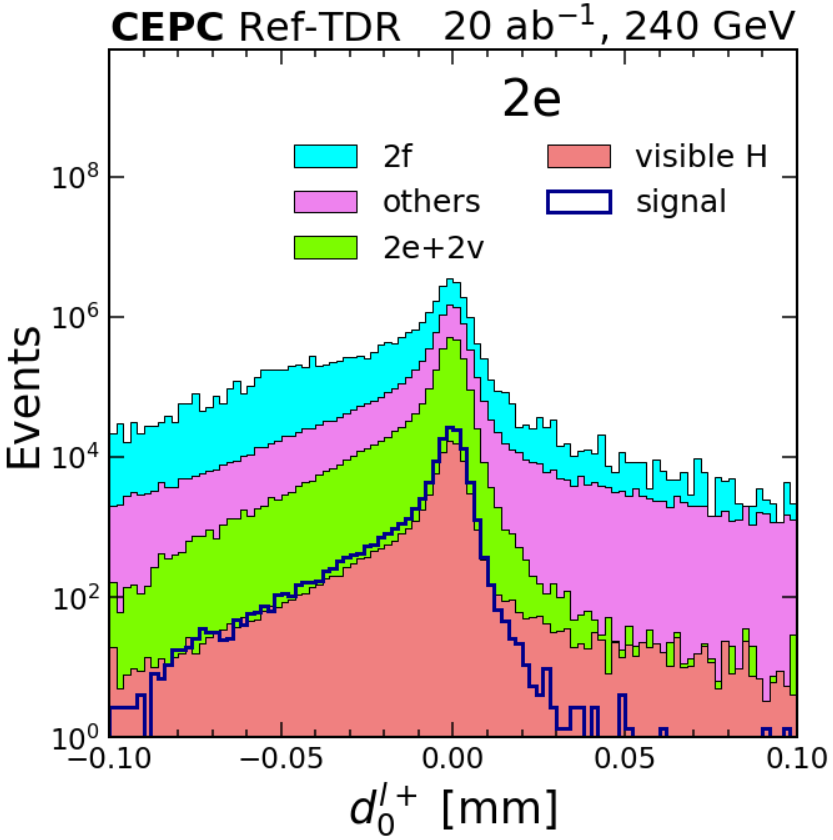


$N_{neutral} < 15$

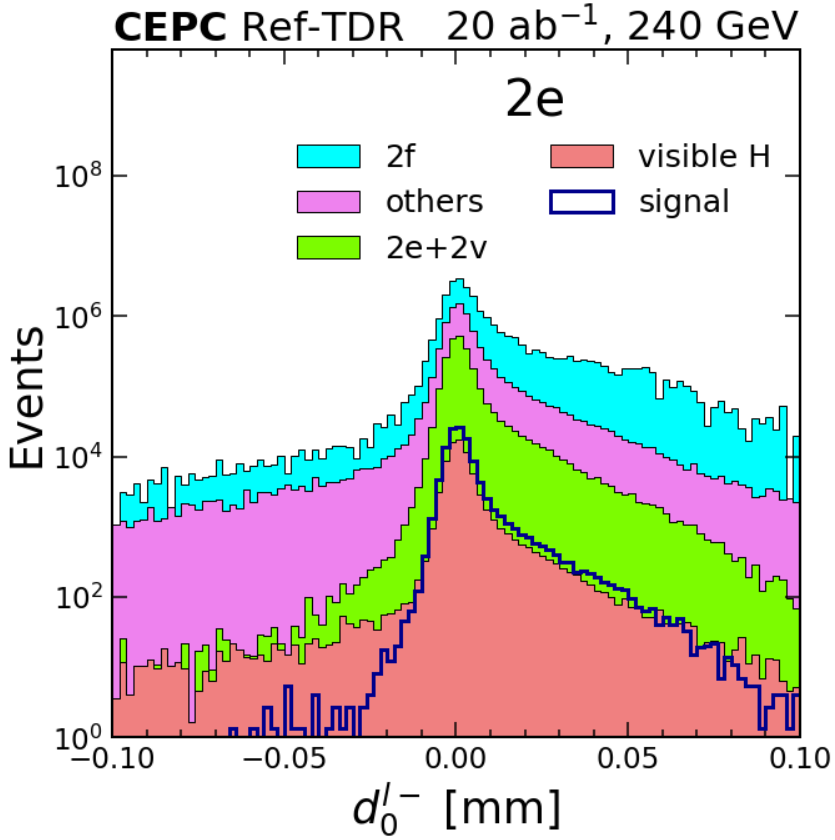


Kinematic variables: 2e

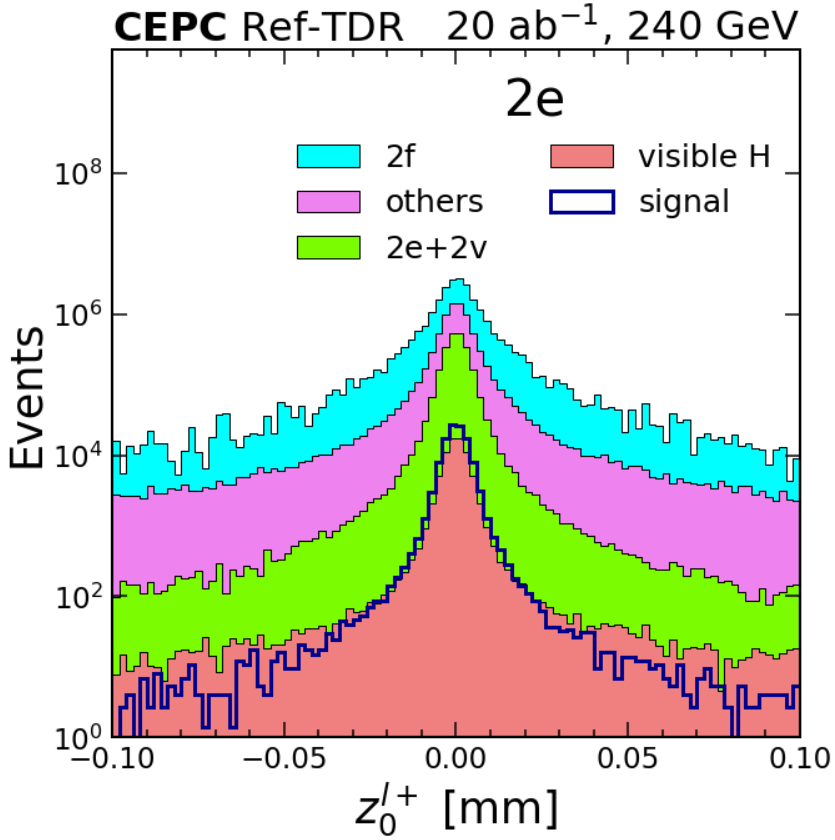
$D_0 \in [-0.1, 0.05]$ mm



$D_0 \in [-0.05, 0.1]$ mm

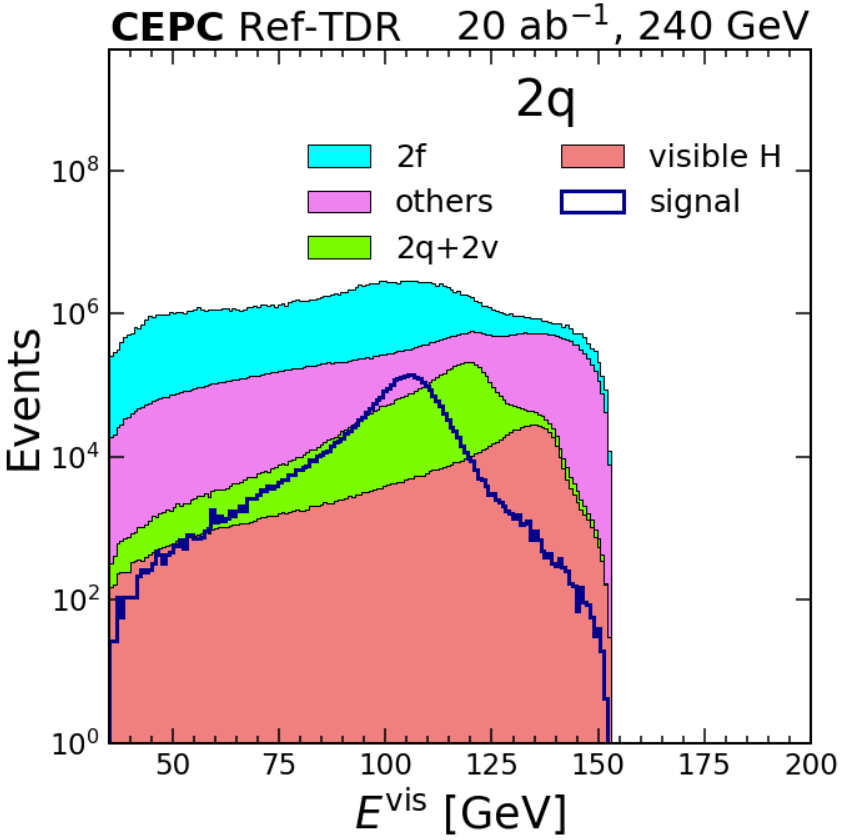


$|Z_0| < 0.1$ mm

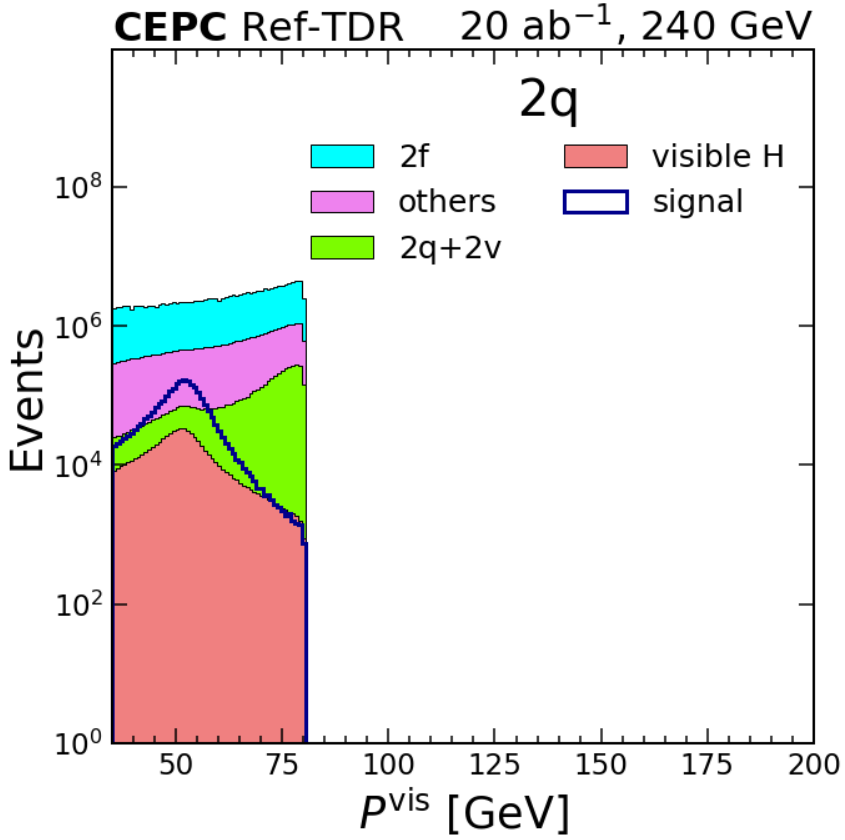


Kinematic variables: 2q

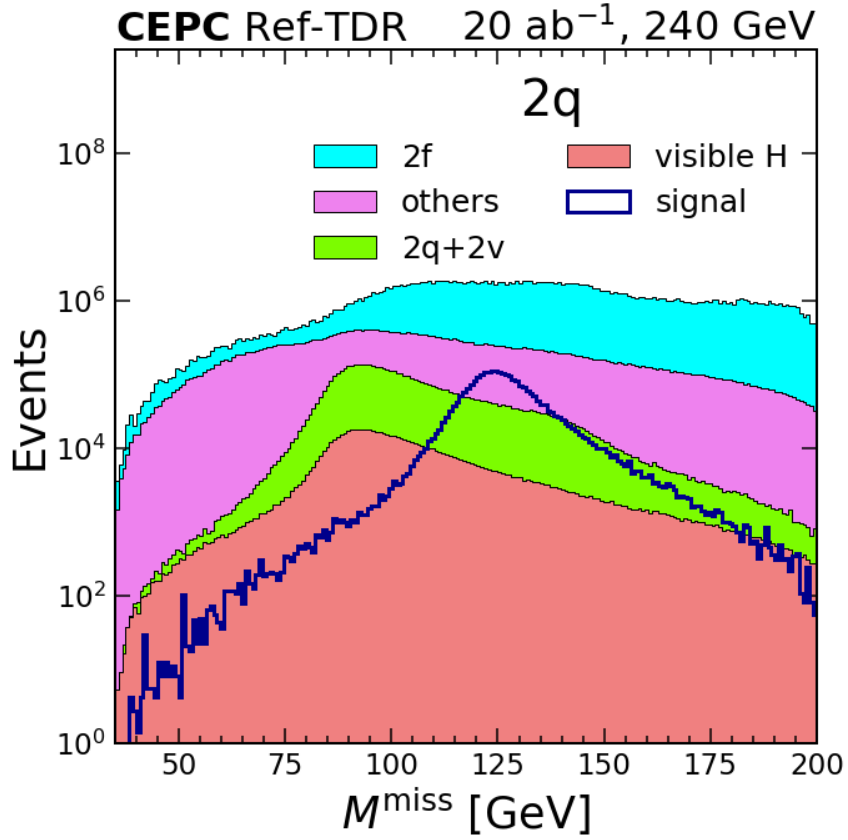
$E_{visible} \in (70,130)$ GeV



$P_{visible} \in (20,70)$ GeV

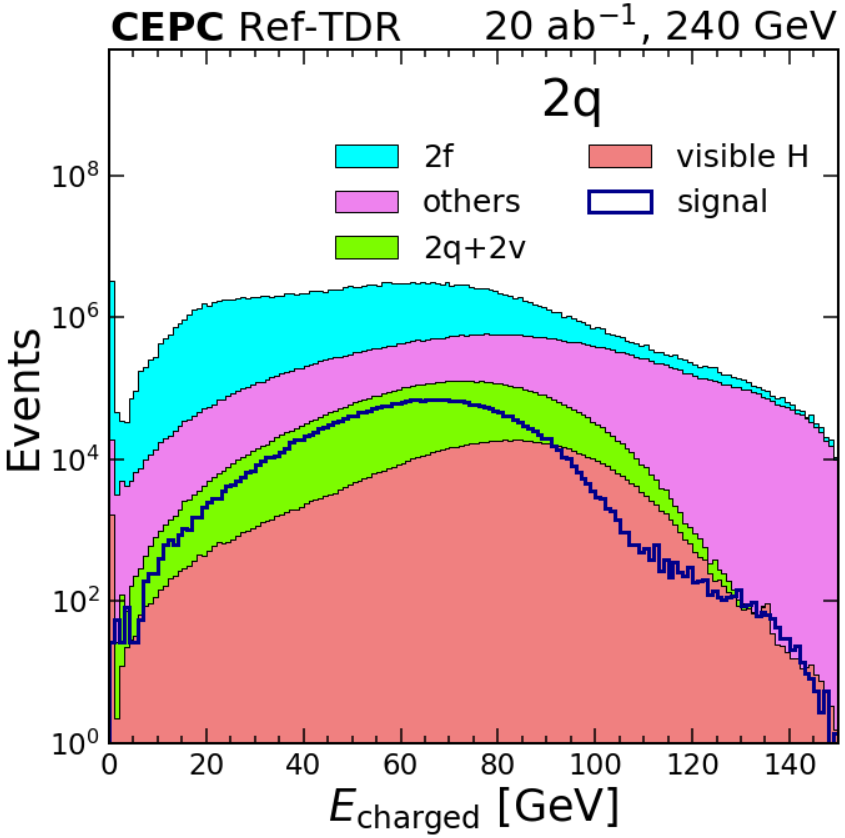


$M_{missing} \in (100,170)$ GeV

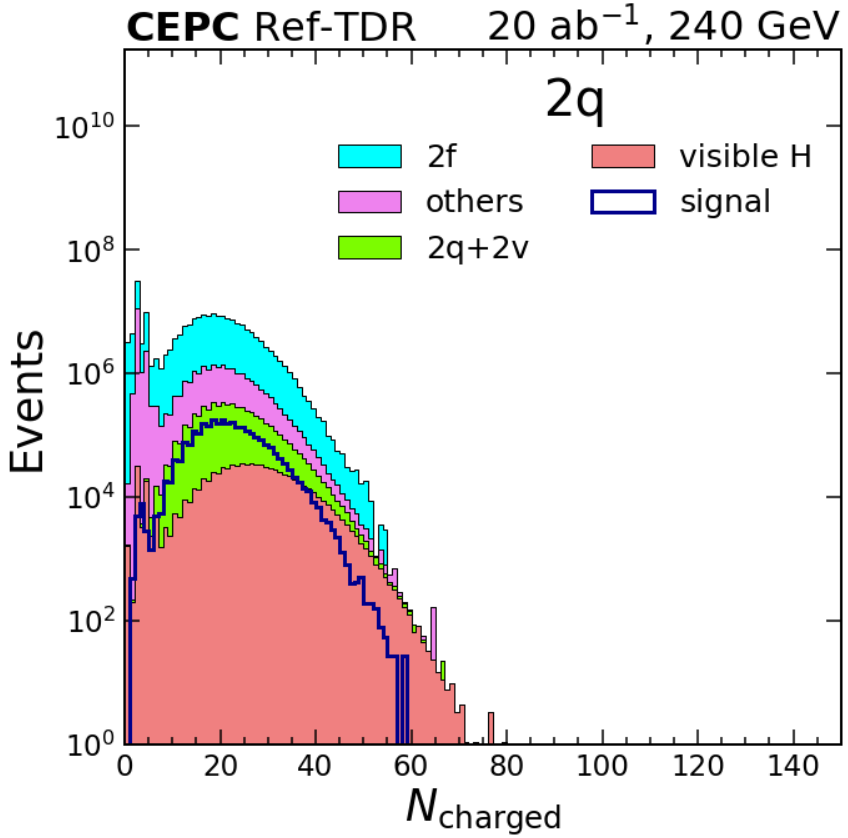


Kinematic variables: 2q

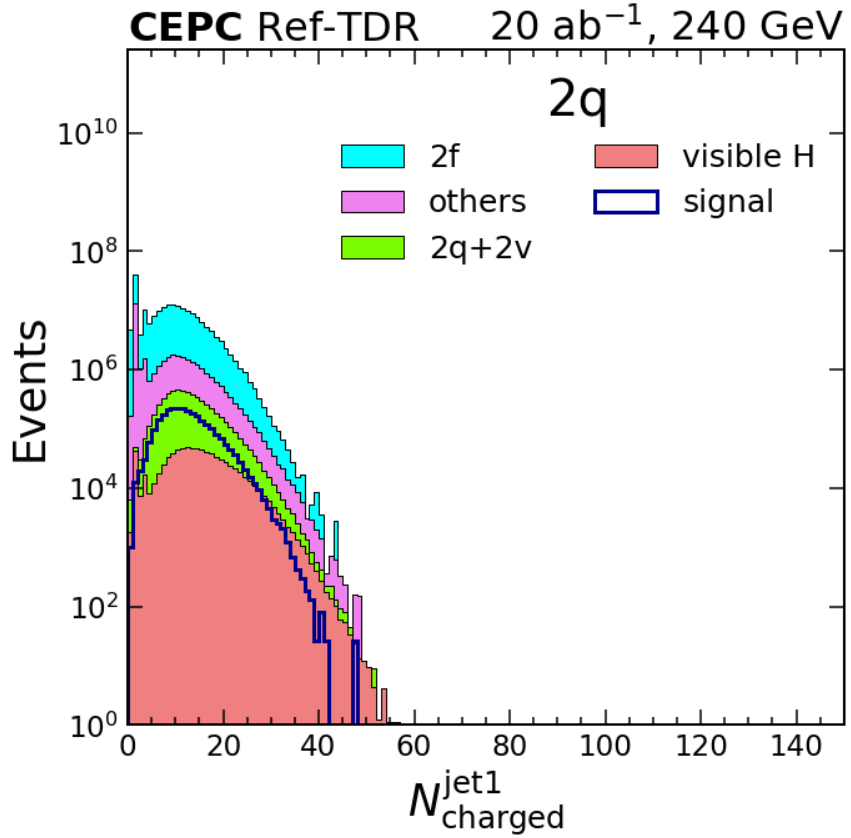
$E_{charged} \in (15,100)$ GeV



$N_{charged} > 5$

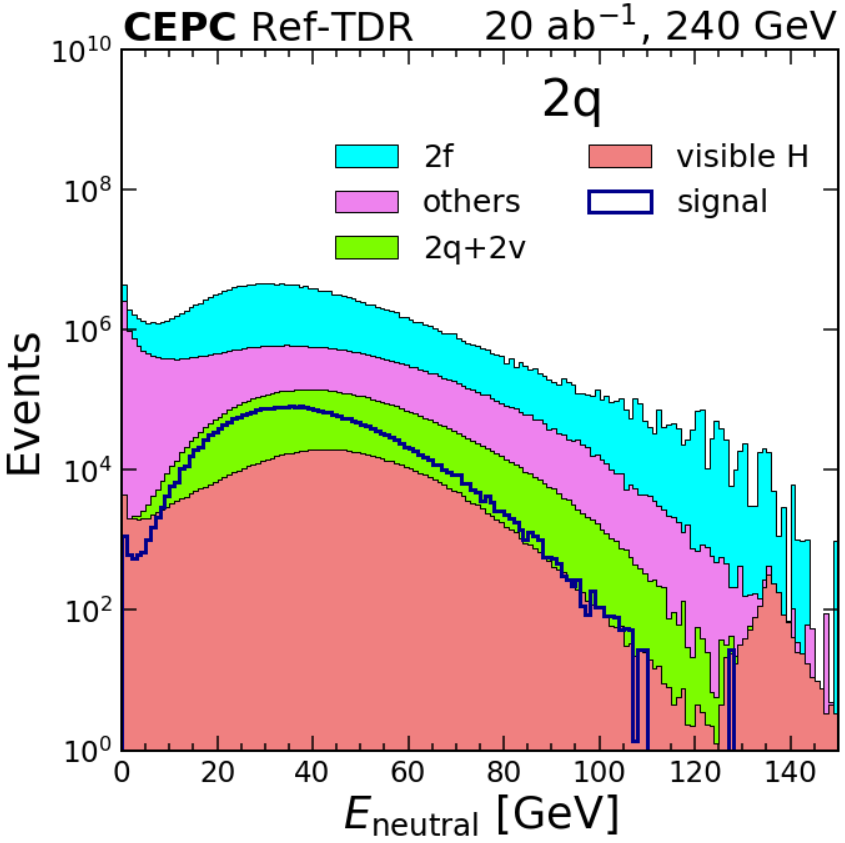


Jet $N_{charged} > 0$

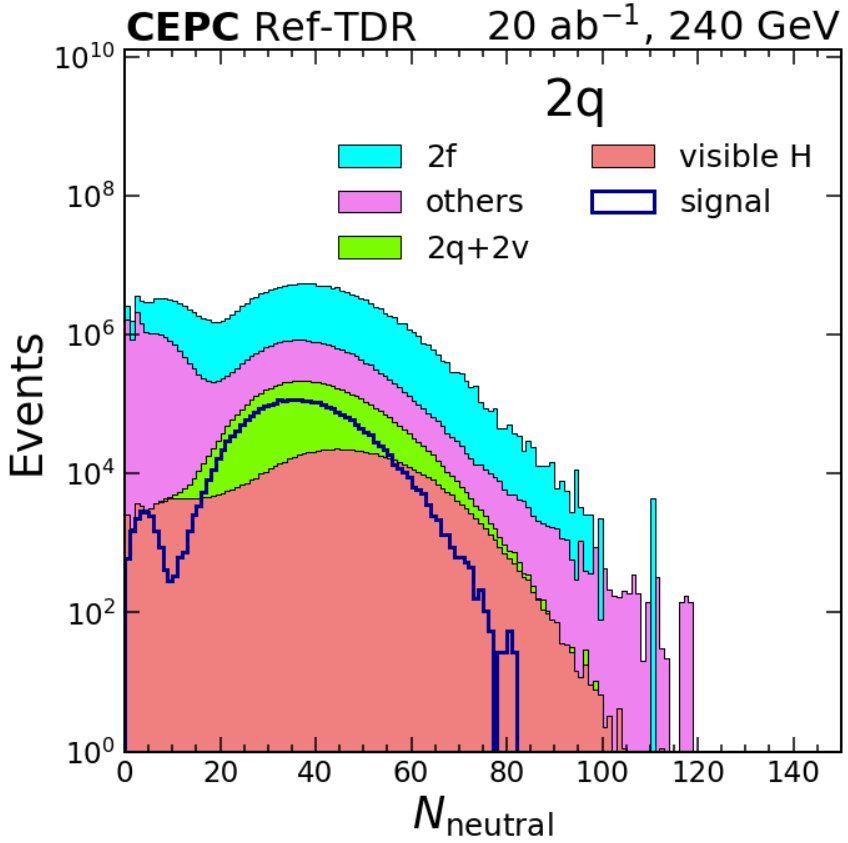


Kinematic variables: 2q

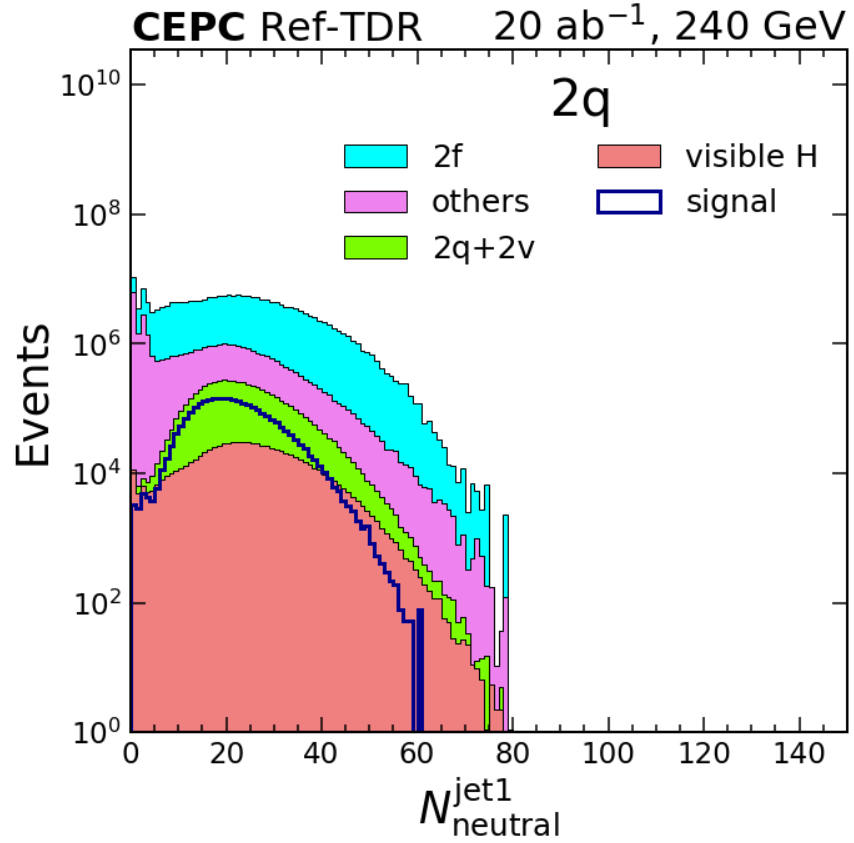
$E_{neutral} \in (5,90)$ GeV



$N_{neutral} > 10$

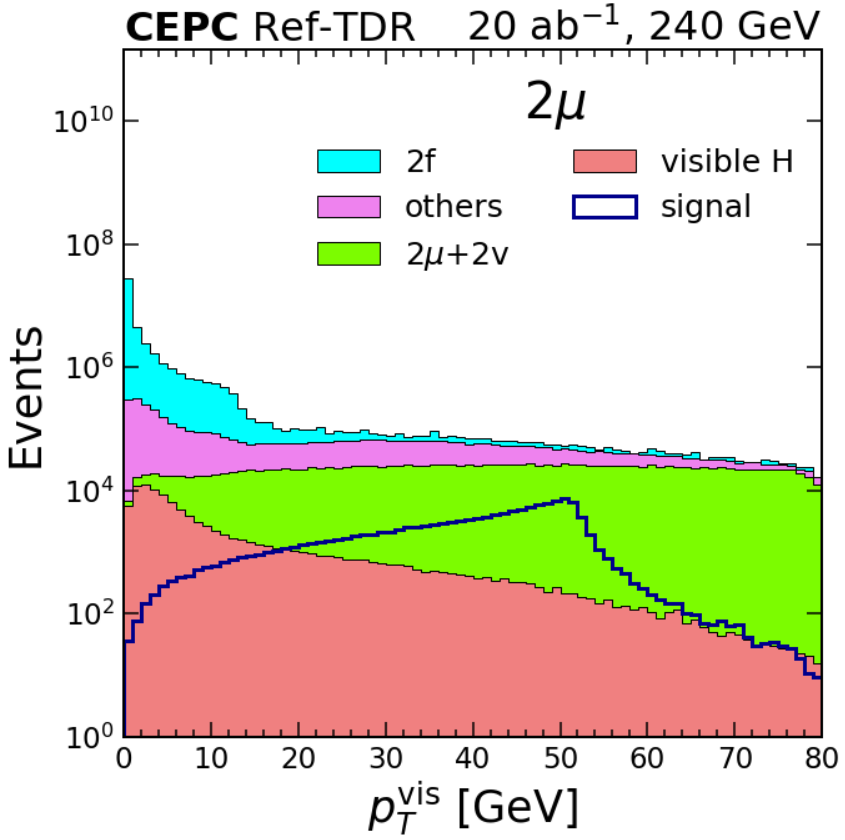


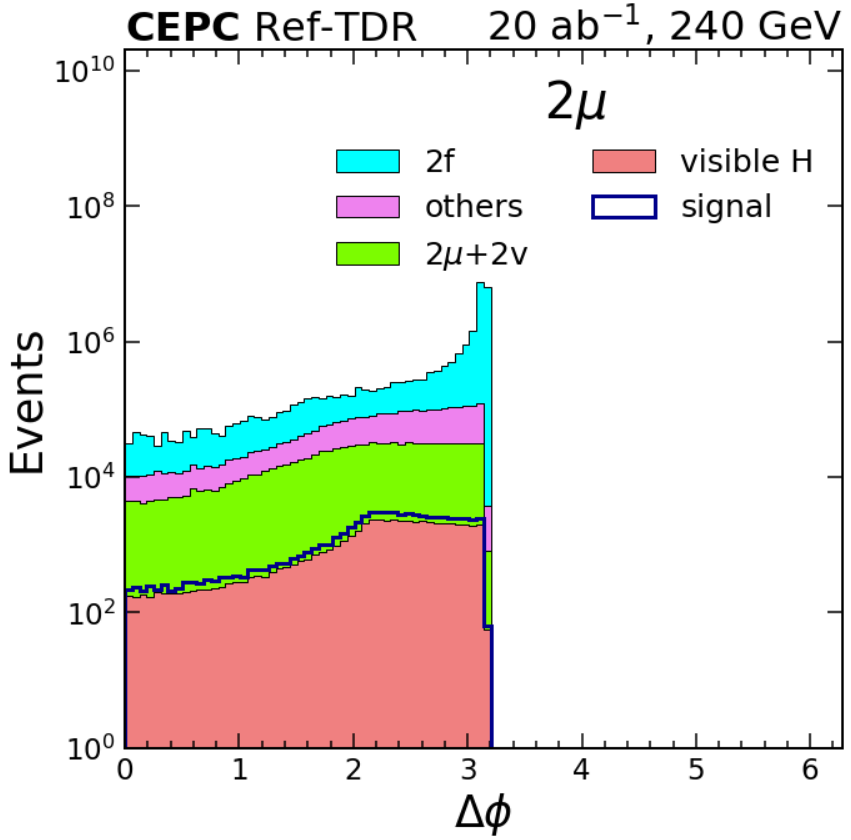
Jet $N_{neutral} > 2$

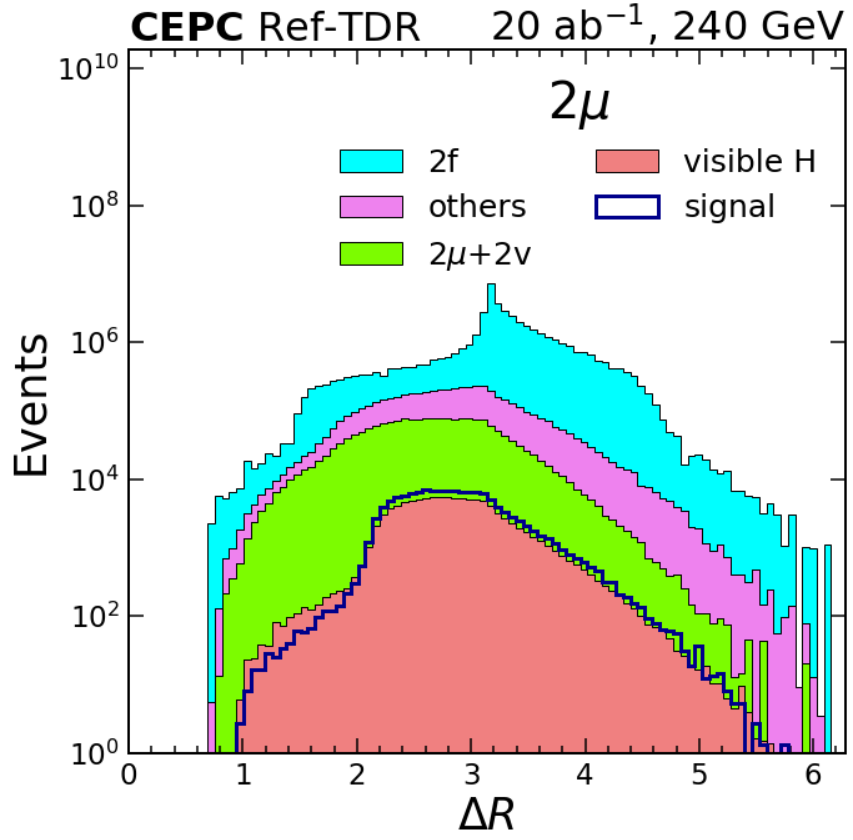


Input features: 2μ

Those shown in previous slides will not be shown again.

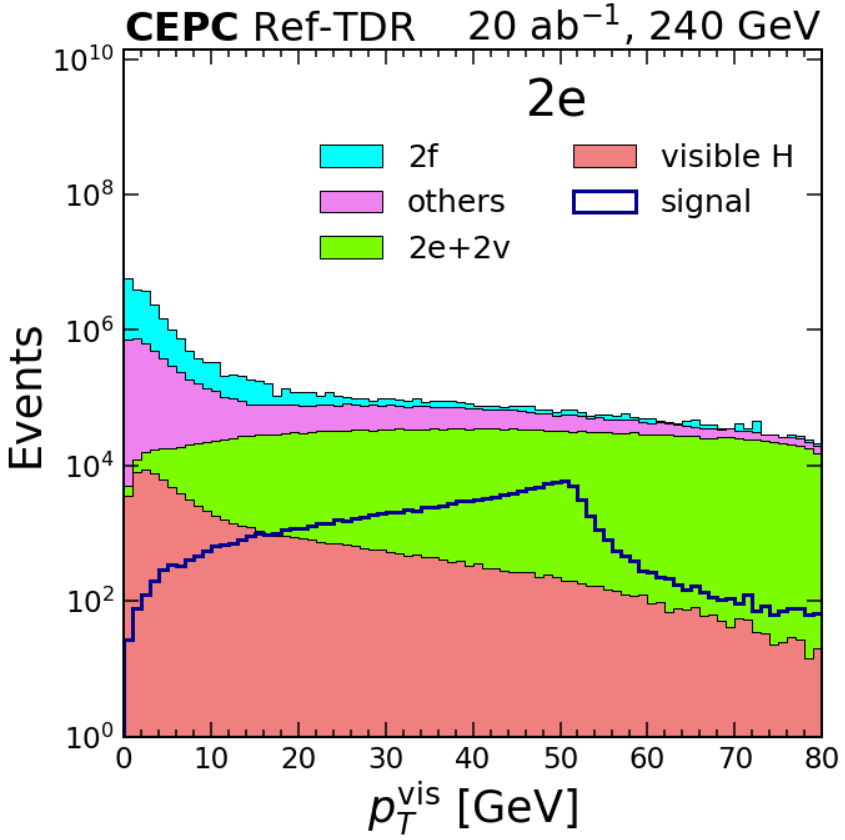


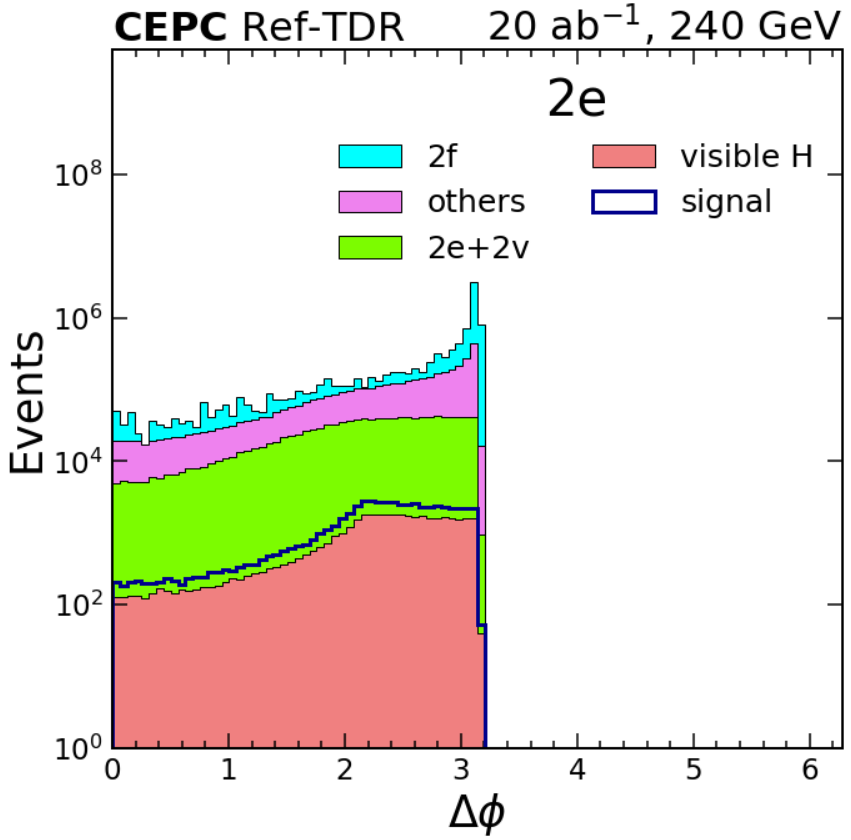


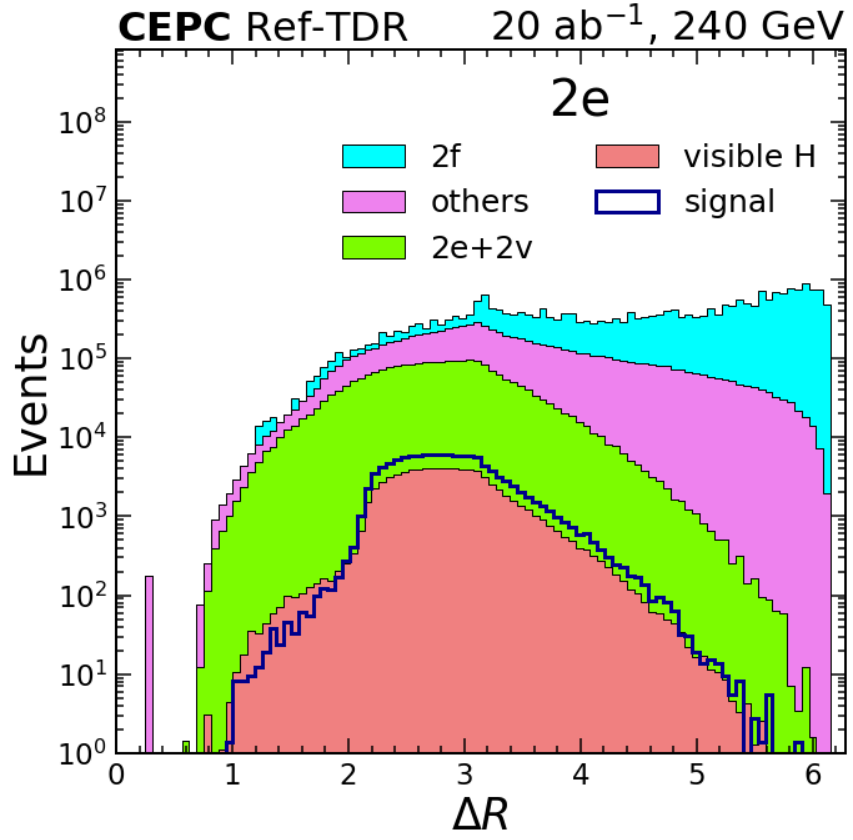


Input features: 2e

Those shown in previous slides will not be shown again.







Input features: 2q

Those shown in previous slides will not be shown again.