Updates on Higgs boson invisible decay studies

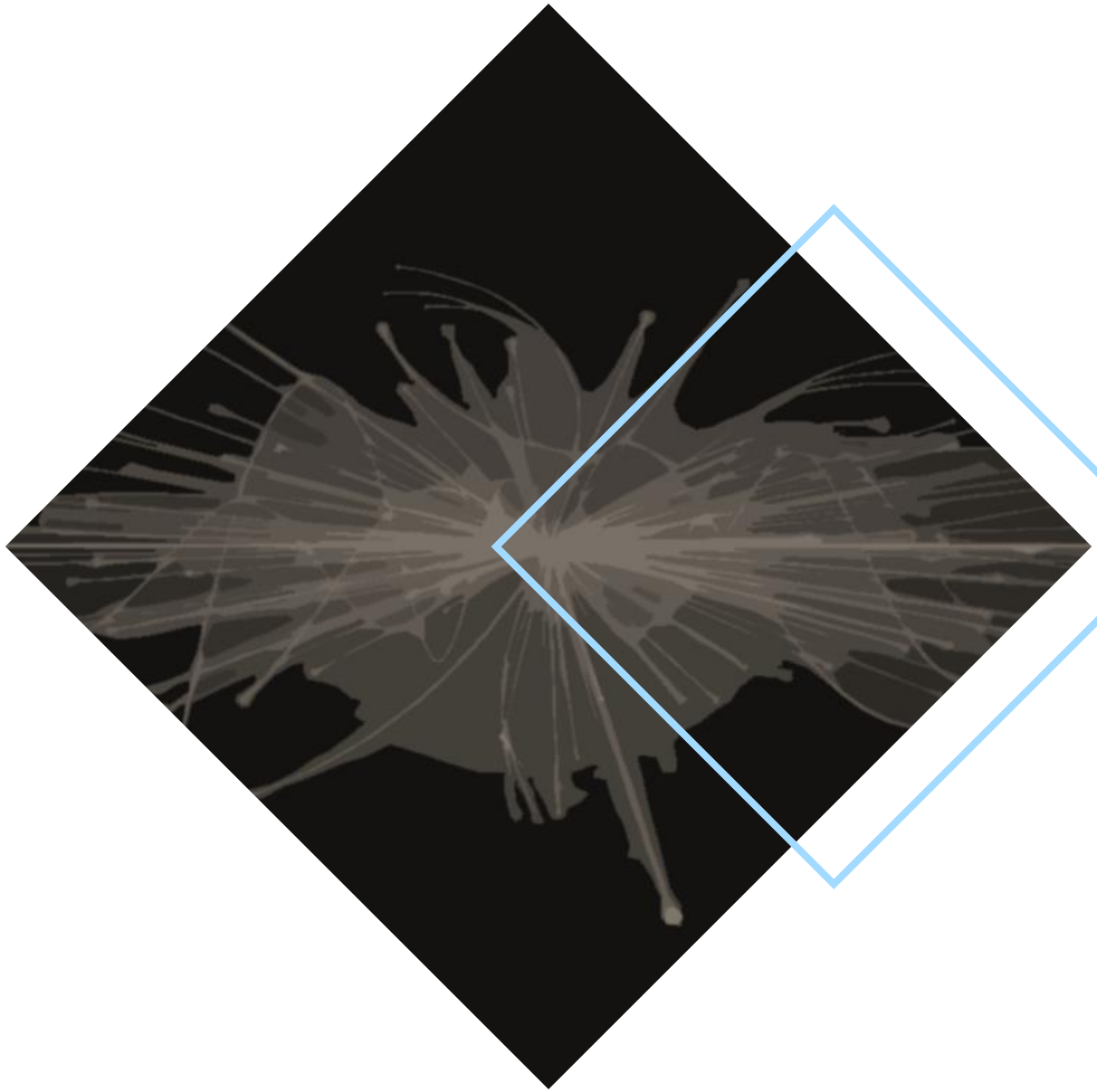
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Geliang Liu (刘格良)

May. 14th, 2025

Outlines

- **Studies with previous CEPCSW version**
  - Overall workflow
  - Preliminary results
- **Status with version ~25.3.6**
  - Signal features
  - Repeat the workflow
  - Possible improvements



Studies with previous versions

Workflow

- **Samples produced with the previous CEPCSW version**
  - Produced with 25.1.2 and additional updates
  - Limited statistics
- **Baseline selection for eeH(→inv) and μμH(→inv)**
  - Preliminary lepton ID
  - Two PFOs passing $|\cos\theta| < 0.99$, muon ID
  - Opposite charge
- **Kinematic selection**
  - Selection criteria based on Chinese Phys. C 44 123001
  - Use kinematic variables including Evis, Pvis, Mll, Mrecoil, etc.
- **Discriminating variable**
  - The missing mass, peaked at 125 GeV

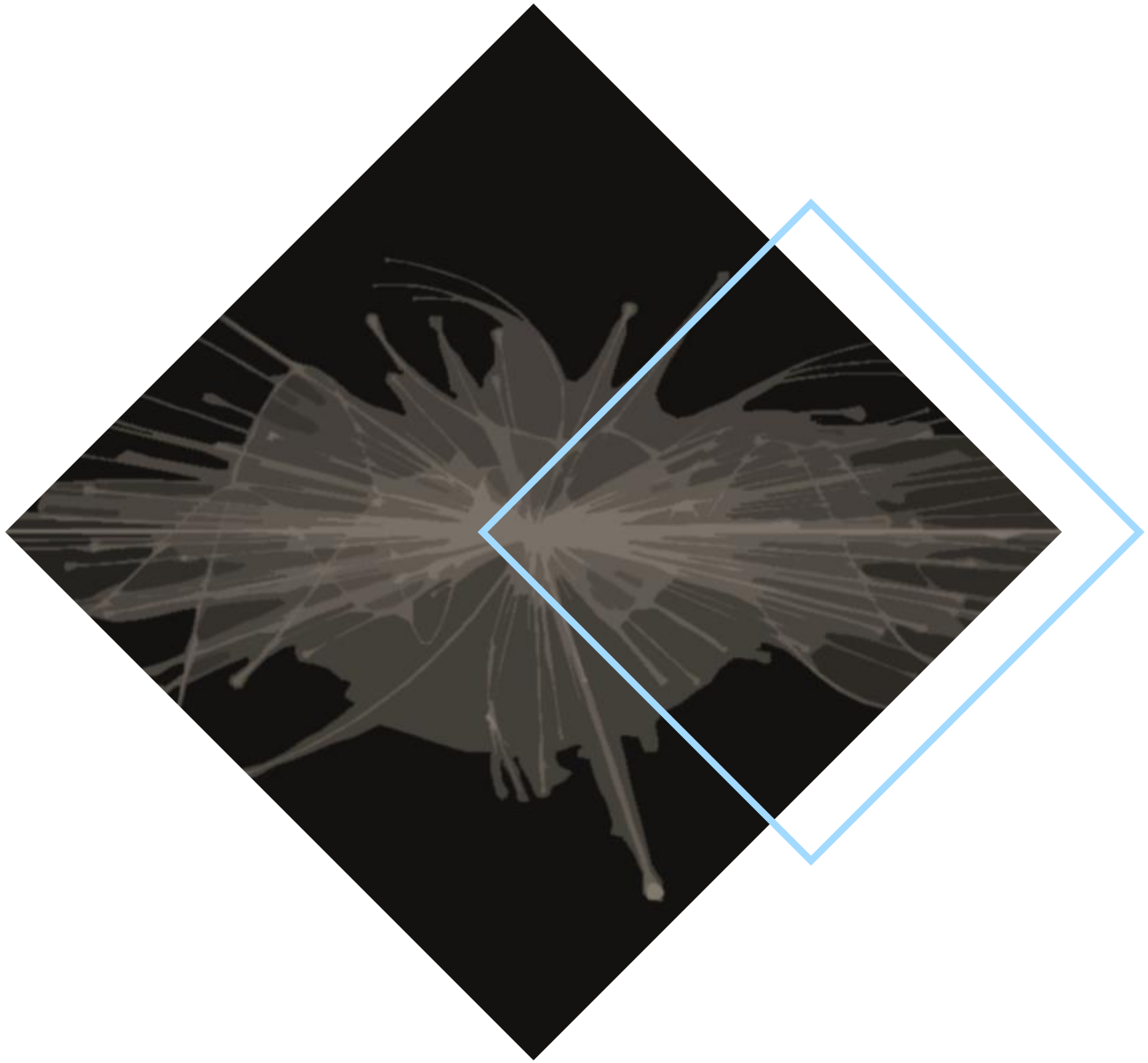
Results

5.6 ab-1

Final state	Uncertainty	Uncertainty (CDR)	Upper limit
ee	$(1.00^{+4.68}_{-1.00}) \cdot 0.1\%$	$\pm 4.54 \cdot 0.1\%$	0.921%
μμ	$(1.00^{+2.53}_{-1.00}) \cdot 0.1\%$	$\pm 2.36 \cdot 0.1\%$	0.495%
qq	$(1.00^{+1.21}_{-1.00}) \cdot 0.1\%$	$\pm 0.95 \cdot 0.1\%$	0.237%
All	$(1.00^{+1.05}_{-0.92}) \cdot 0.1\%$	$\pm 0.87 \cdot 0.1\%$	0.202%

20 ab-1

Final state	Uncertainty	Upper limit
ee	$(1.00^{+2.46}_{-1.00}) \cdot 0.1\%$	0.483%
μμ	$(1.00^{+1.28}_{-0.86}) \cdot 0.1\%$	0.234%
qq	$(1.00^{+0.64}_{-0.64}) \cdot 0.1\%$	0.125%
All	$(1.00^{+0.55}_{-0.53}) \cdot 0.1\%$	0.102%



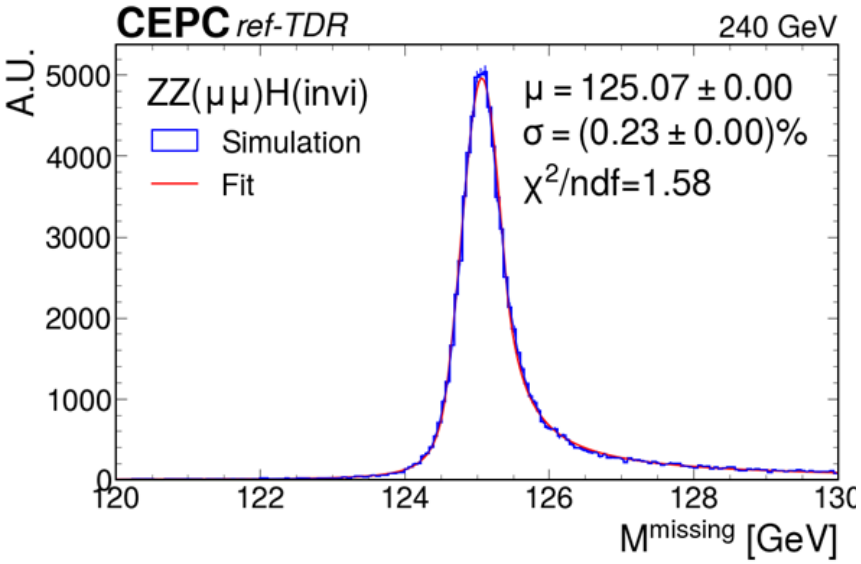
Status with the new CEPCSW version

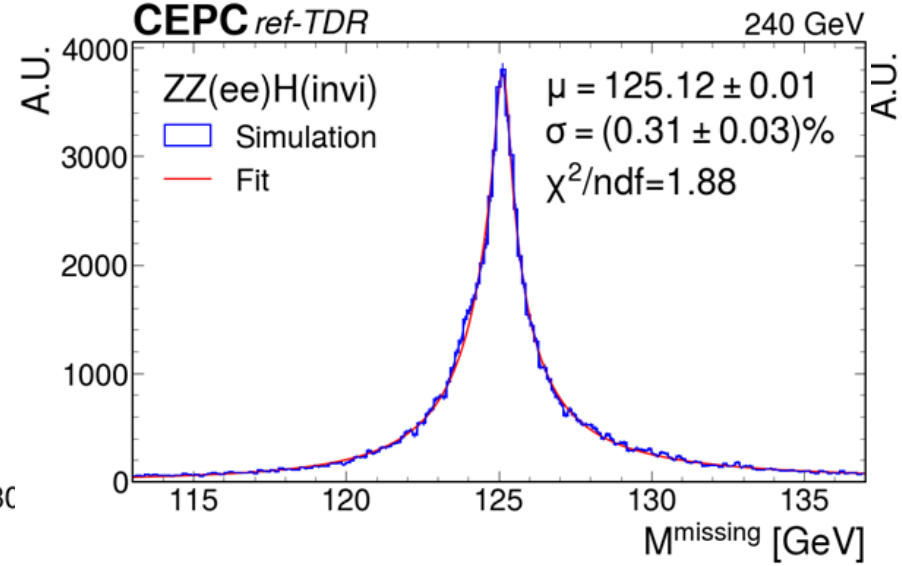
Samples

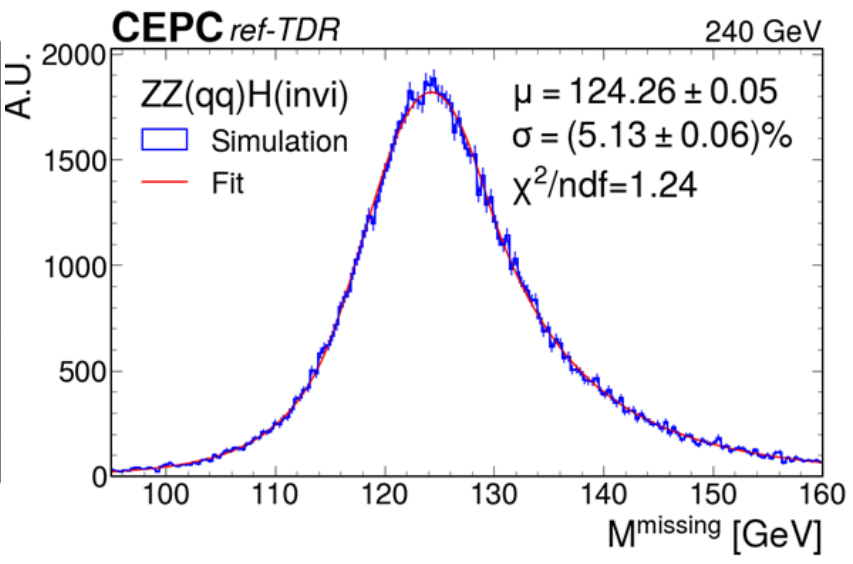
➢ Samples produced with CEPCSW 25.3.6

	Processes	Location	Events
Signal	Z(→ee/μμ/qq)H(→4v)	/cefs/higgs/liugeliang/CEPC/202503/Production/Hinvi	100k per final state
4-fermion bkg	single-Z, single-W, Z-or-W, ZZ, WW, ZZ-or-WW	/cefs/higgs/zhangkl/Production/25036/4fermions	400k per final state
2-fermion bkg	ll, qq	/cefs/higgs/zhangkl/Production/25036/E240_e1e1(e2e2, e3e3, qq)	100k – 400 k per final state
ZH	Z(→ee/μμ/ττ/vv/qq) H(→incl)	/cefs/higgs/zhangkl/Production/25036/E240_*HX	1M per sample

Signal features







Previously

resolution = 0.23%	resolution = 0.66%	resolution = 4.57%

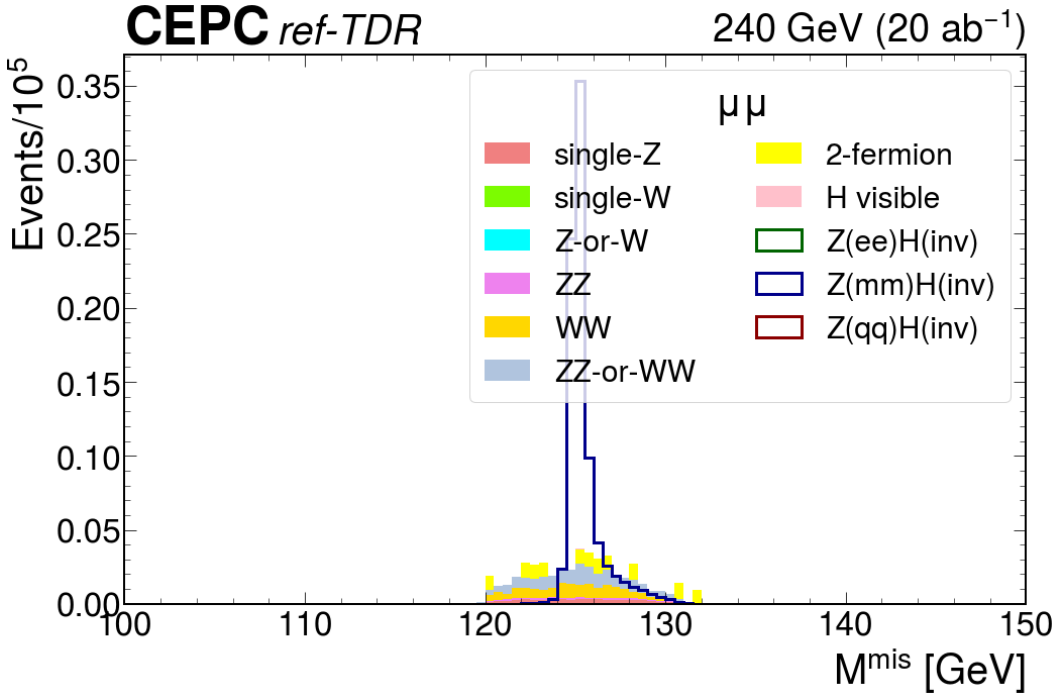
- Resolution in the ee channel improved: updates on tracker and tracking algorithm.
- Resolution in the qq channel worse: potentially due to ECAL geometry changes.

Event selection

- Baseline selection can be improved by using the XGBoost lepton ID at 98% WP.
- Other selection criteria remain the same for comparison.

Event selection: μμ

	mmH	sz	sw	szw	zz	ww	zzww	2f	H visible	All bkg
Total yield	135397	32046835	69704759	4989450	22625137	181521998	72958479	1682305536	4069166	2070221360
Base sel eff (%)	94.454	20.527	2.188	0.115	3.34	2.286	5.901	7.082	0.705	
Evis ∈ [102, 107]	73.54	0.288	0.01	0.004	0.315	0.063	0.203	0.047	0.015	
Mll ∈ [85,97] (%)	66.491	0.02	0	0	0.013	0.01	0.033	0.014	0.001	
Ptll > 12 (%)	64.773	0.019	0	0	0.01	0.01	0.031	0.001	0.001	
Ell/Pll < 2.4 (%)	63.938	0.018	0	0	0.009	0.008	0.024	0.001	0.001	
Mrecoil∈ [120,150]	**63.938**	**0.017**	**0**	**0**	**0.009**	**0.007**	**0.022**	**0.001**	**0.003**	**0.000**
eff 25.1.2 (%)	**56.133**	**0.017**	**0.000**	**0.000**	**0.010**	**0.007**	**0.019**	**0.000**	**0.001**	**0.002**



cq,sq,cq,sq

uq,dq,uq,dq

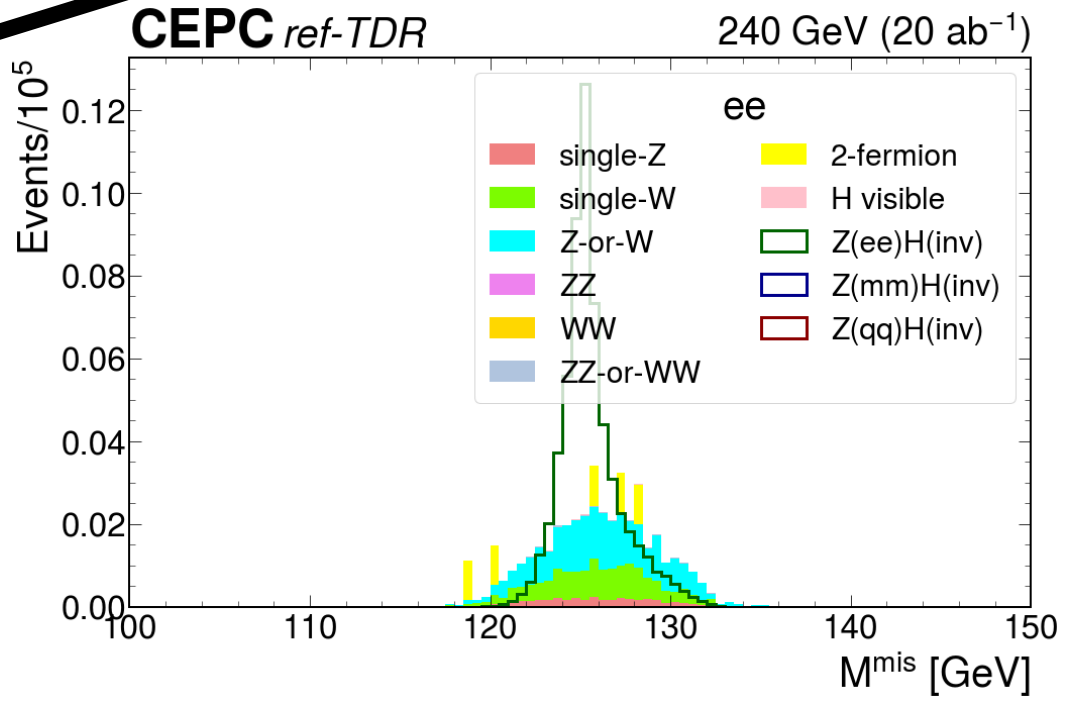
mu, mu, nu_μ, nu_μ

$tau, tau, nu_\tau, nu_\tau$

Event selection: ee

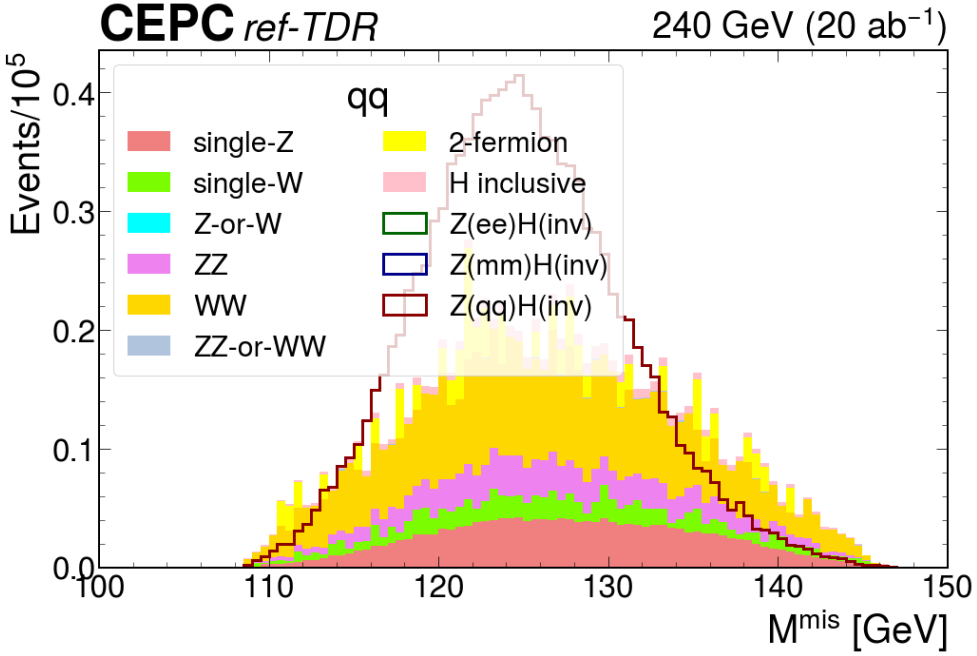
	eeH	sz	sw	szw	zz	ww	zzww	2f	H visible	All bkg
Total yield	140799	32046835	69704759	4989450	22625137	181521998	72958479	1682305536	4069166	2070221360
Base sel eff (%)	84.948	17.287	2.984	83.034	0.334	0.221	0.388	9.712	0.369	
Evis ∈ [101, 107]	56.120	0.304	0.167	3.317	0.017	0.004	0.018	0.021	0.009	
Mll ∈ [71,99] (%)	49.692	0.042	0.056	1.263	0.000	0.000	0.002	0.003	0.001	
Ptll ∈ [12,55] (%)	47.593	0.018	0.044	1.009	0.000	0.000	0.002	0.001	0.001	
Ell/Pll ∈ [1,8,2.4]	42.939	0.011	0.018	0.425	0.000	0.000	0.001	0.000	0.000	
Mrecoil∈ [120,170]	**42.939**	**0.011**	**0.018**	**0.425**	**0.000**	**0.000**	**0.001**	**0.000**	**0.000**	**0.000**
eff 25.1.2 (%)	**39.429**	**0.011**	**0.020**	**0.359**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.002**

$e^-, e^+, \nu_e, \bar{\nu}_e$



Event selection: qq

	qqH	sz	sw	szw	zz	ww	zzww	2f	Hincl
Total yield	2736200	32046835	69704759	4989450	22625137	181521998	72958479	1.68E+09	4069166
Base sel eff (%)	100	97.85	99.85	99.534	99.755	99.968	99.822	98.014	99.988
Evis ∈ [90, 117] (%)	87.219	8.793	5.767	15.756	9.296	2.822	2.059	6.141	4.003
Mvis ∈ [85, 102]	71.049	3.14	1.831	4.706	4.515	1.14	0.766	1.97	2.075
Pvis ∈ [30, 58] (%)	64.843	1.138	1.031	2.401	1.136	0.601	0.435	0.802	1.452
Ptvis ∈ [18, 60] (%)	60.767	0.842	0.919	2.017	0.818	0.526	0.371	0.07	1.314
Mmis ∈ [100,150]	60.767	0.842	0.919	2.017	0.818	0.526	0.371	0.07	1.314
Nc>5, Ec>1 GeV	**43.285**	**0.543**	**0.123**	**0**	**0.557**	**0.218**	**0.002**	**0.005**	**0.947**
eff 25.1.2 (%)	**59.460**	**0.754**	**0.184**	**0.000**	**0.800**	**0.332**	**0.028**	**0.009**	**1.351**

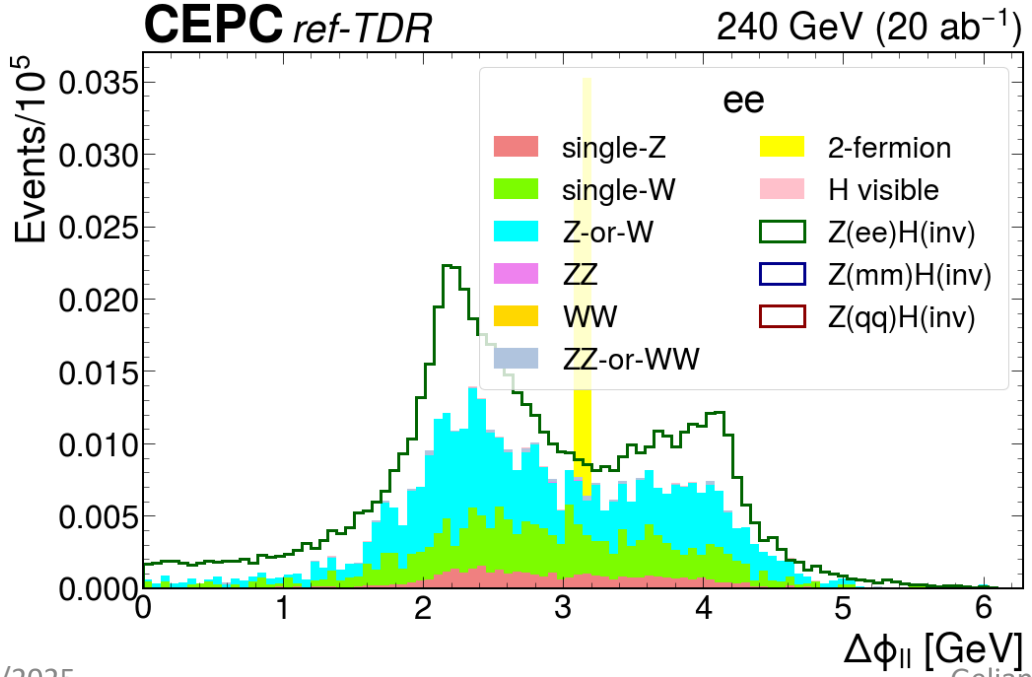


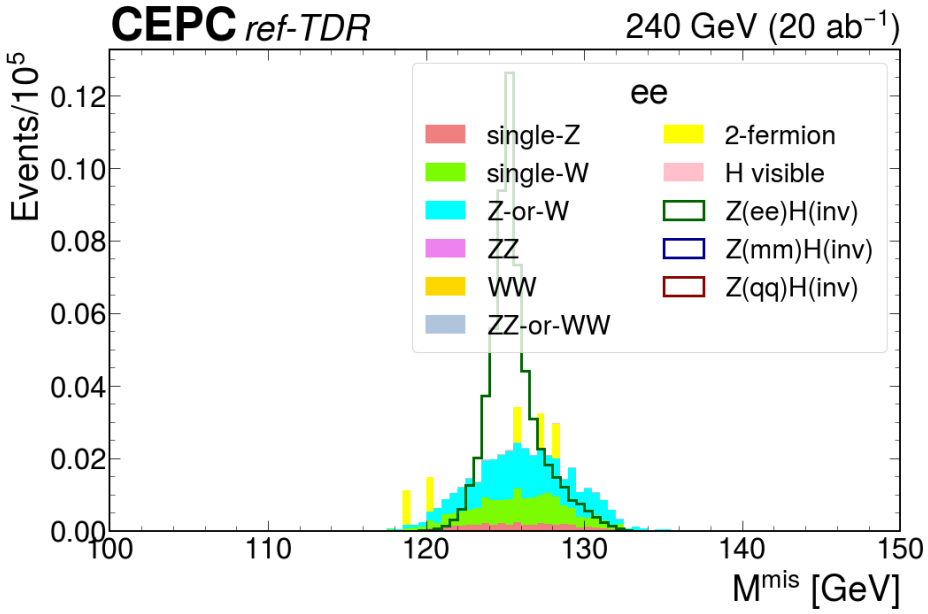
- The efficiencies are lower due to a mistake when computing Nc: it seems I only considered leptons.
- Will correct it soon.

Issues and improvements

➢ Low statistics of 2 fermion backgrounds

1. Use fast simulation to improve statistics
2. Use gen-filter
3. Use the shape from other backgrounds as the shape of 2-f backgrounds and put into the statistical model
4. Apply more cuts to completely remove 2-f backgrounds
   - E.g. $\Delta\phi$ of the two leptons. For 2-f backgrounds the two leptons are always back-to-back.

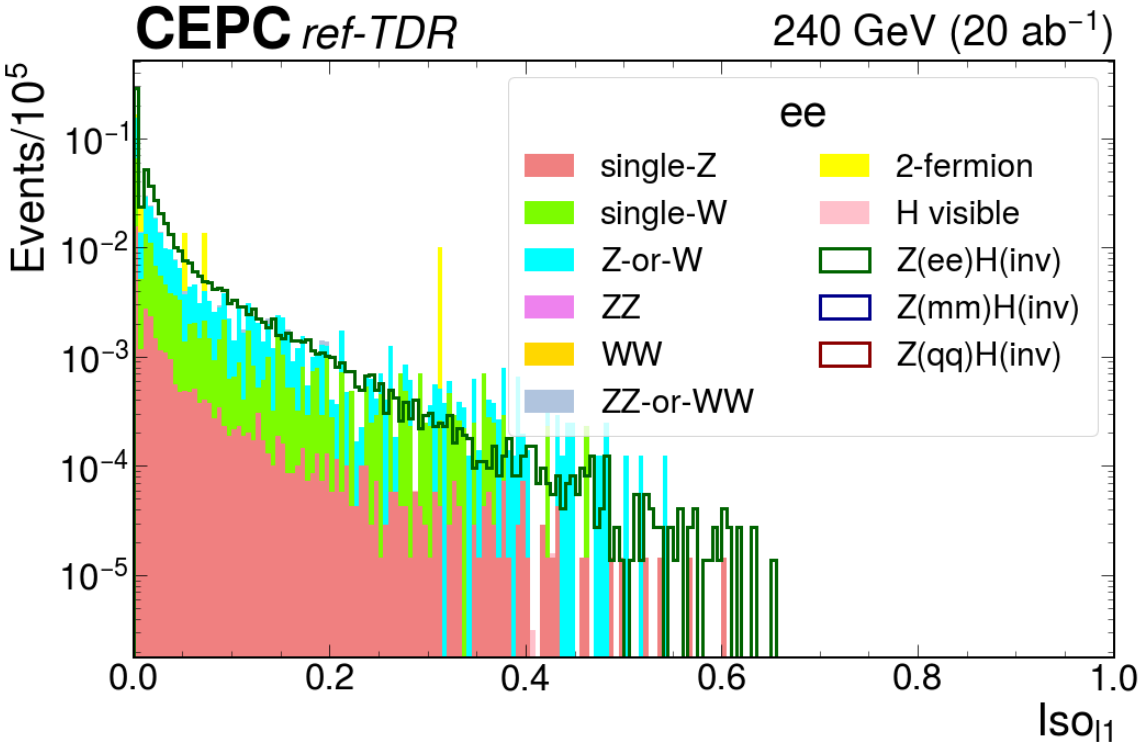


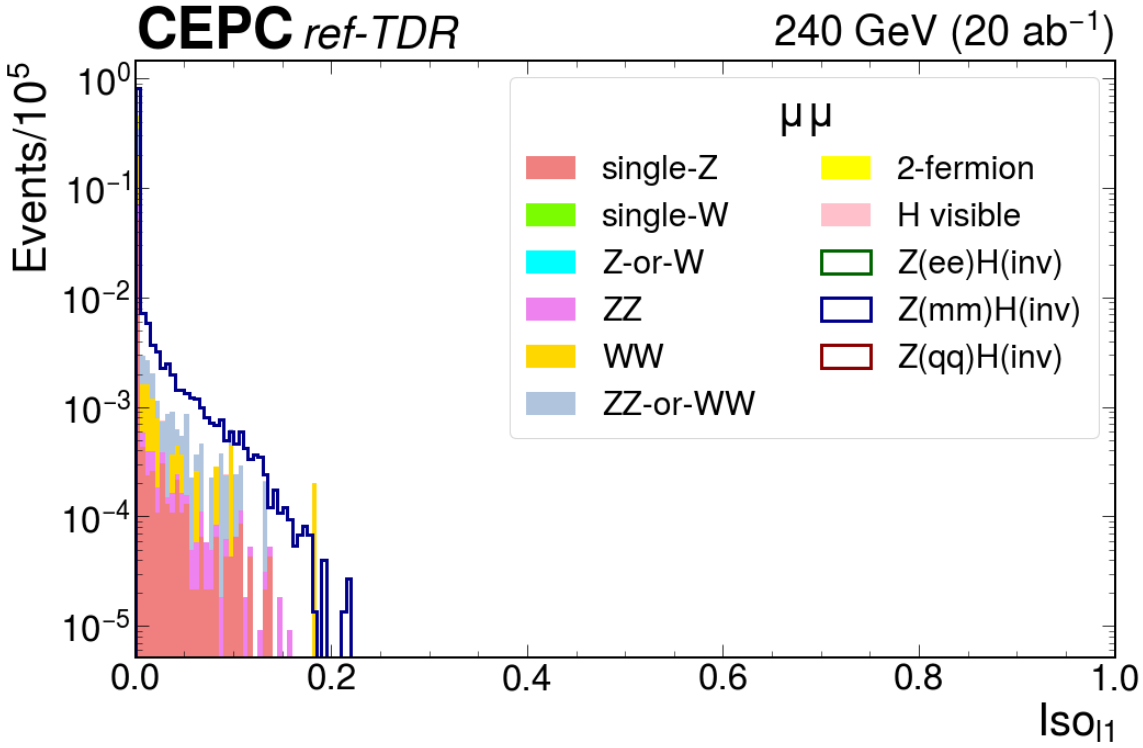


Issues and improvements

➢ Usage of other variables in ee and μμ channels

- Lepton isolation (with $\Delta R < 0.4$) ?



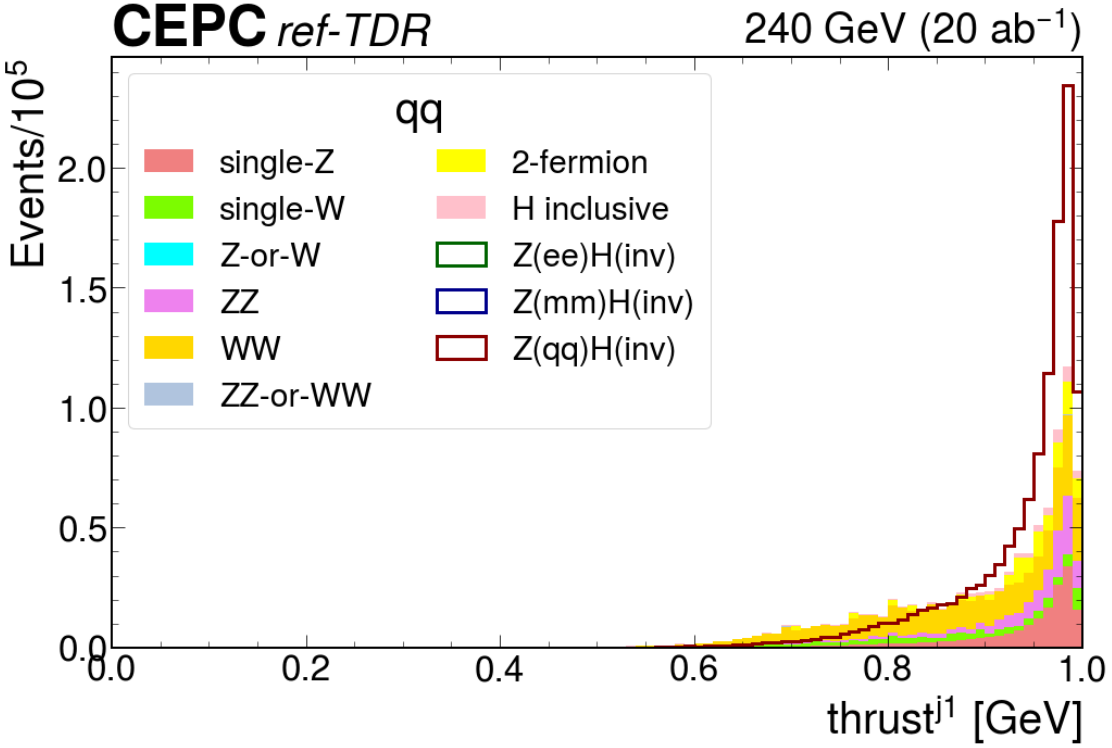


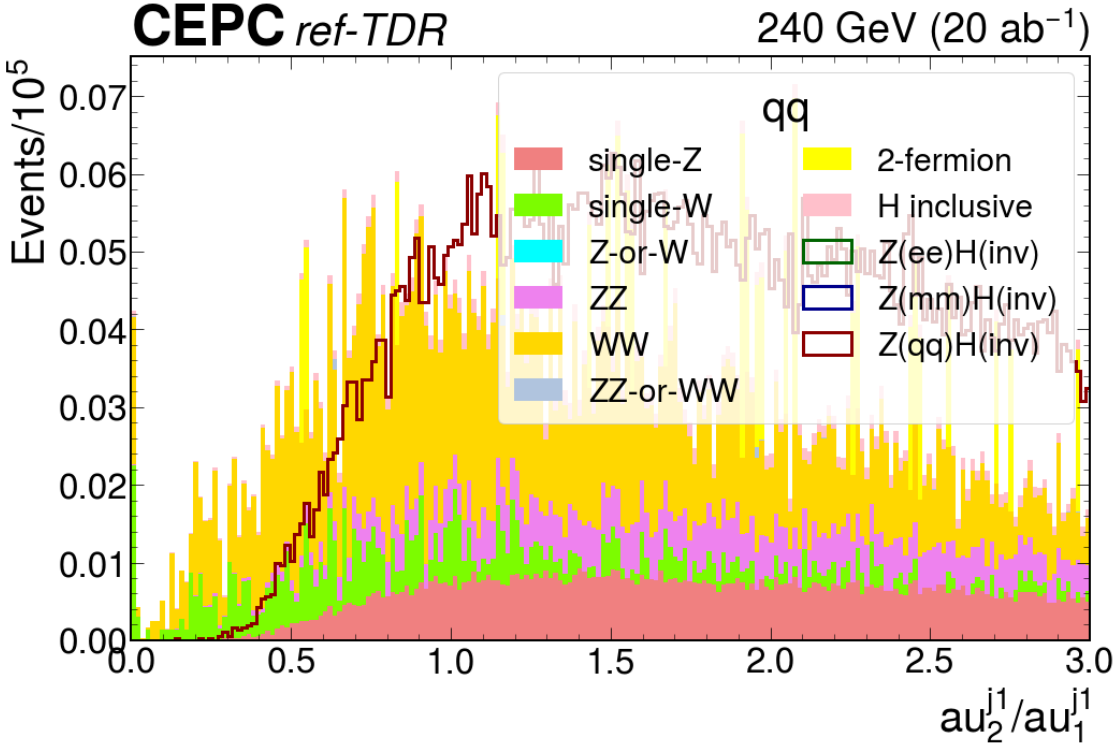
- Considering D0 and Z0 to see if we can distinguish between leptons from IP and leptons from tau decay. Will have them soon.

Issues and improvements

➢ Usage of other variables in the qq channel

- Some jet-substructure variables are considered





- Need to consider distinguishment between hadronic tau lepton and jets. Probably number of hadrons in the jet?

Issues and improvements

- **MVA-based method**
  - Since we have many variables with cuts applied on them, and these cuts need to be tuned, it is natural to think about MVA methods instead.
  - Train a BDT model to distinguish signals with backgrounds. We can directly fit on the BDT scores afterwards or apply cuts on BDT and then fit on missing mass.
  - All variables used before will be input features. Potentially more variables will be added.

Issues and improvements

- **Background illustration (4f)**
  - The conventions inherited from CDR (single-Z, single-W, Z-or-W, ZZ, WW, ZZ-or-WW) are not easily understandable and sometimes can be misleading.
  - Suggest to re-categorize the 4f backgrounds based on final states, depending on the analysis target.
  - E.g., for ll channels in the Higgs boson invisible decays, the main backgrounds with final states of llvv and lτvv can be picked up individually, while other 4f backgrounds are merged together and called “others”.
  - Also propose to other analyses.