



CEPC samples overview

Kaili Zhang

IHEP

zhangkl@ihep.ac.cn

CEPCSW samples production

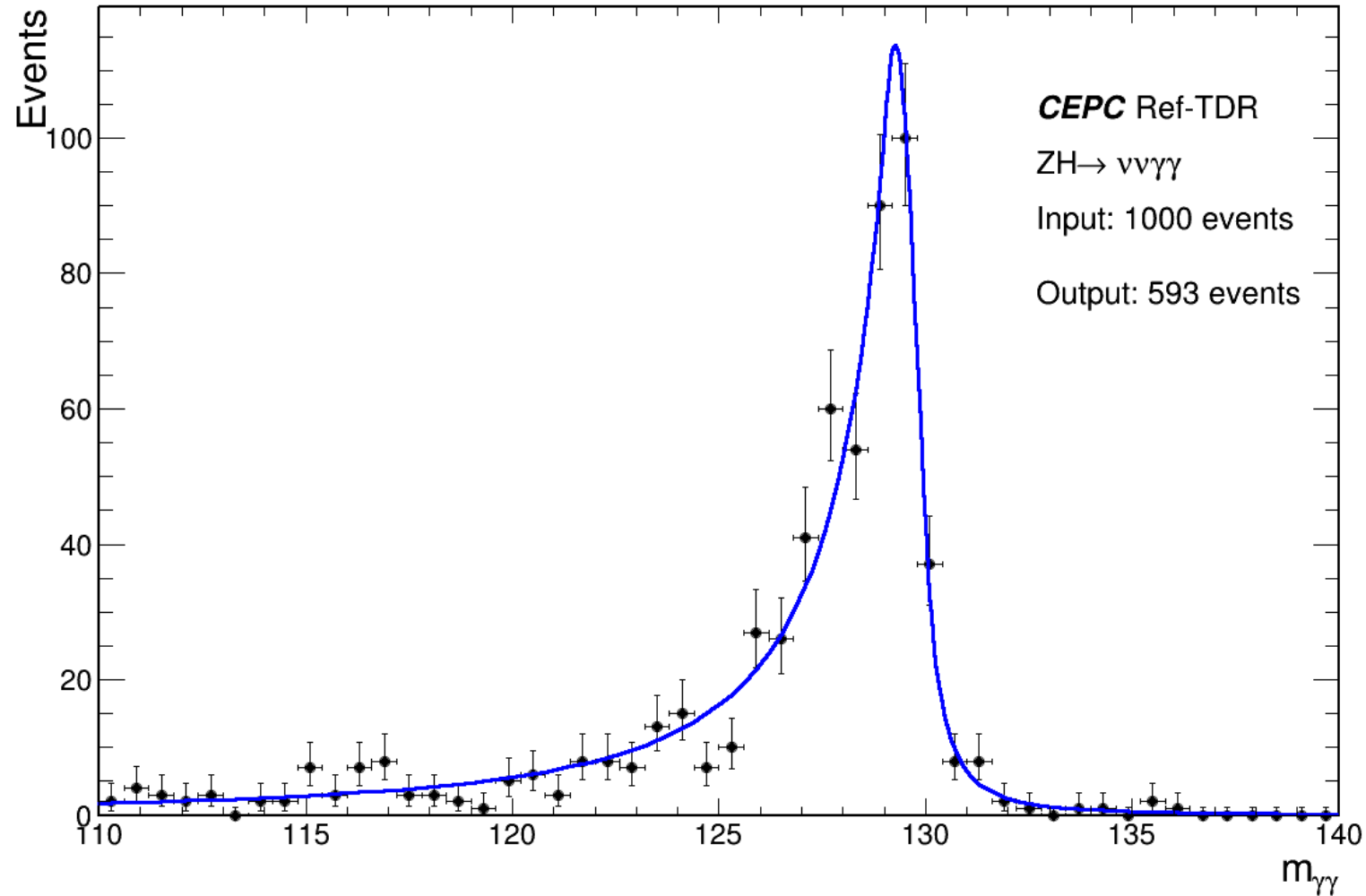


- Small amount 1k sample available for mm, yy and gg.
- Script for production, fitting, plotting ready.
- JetClustering implemented;
- Ecal bug fixed;
- Plan to use one universal script for production, and one small sample only containing MCParticle and PandoraPFO and JetClustering for analysis.
- Changes will merge to CEPCSW master soon.

Diphoton



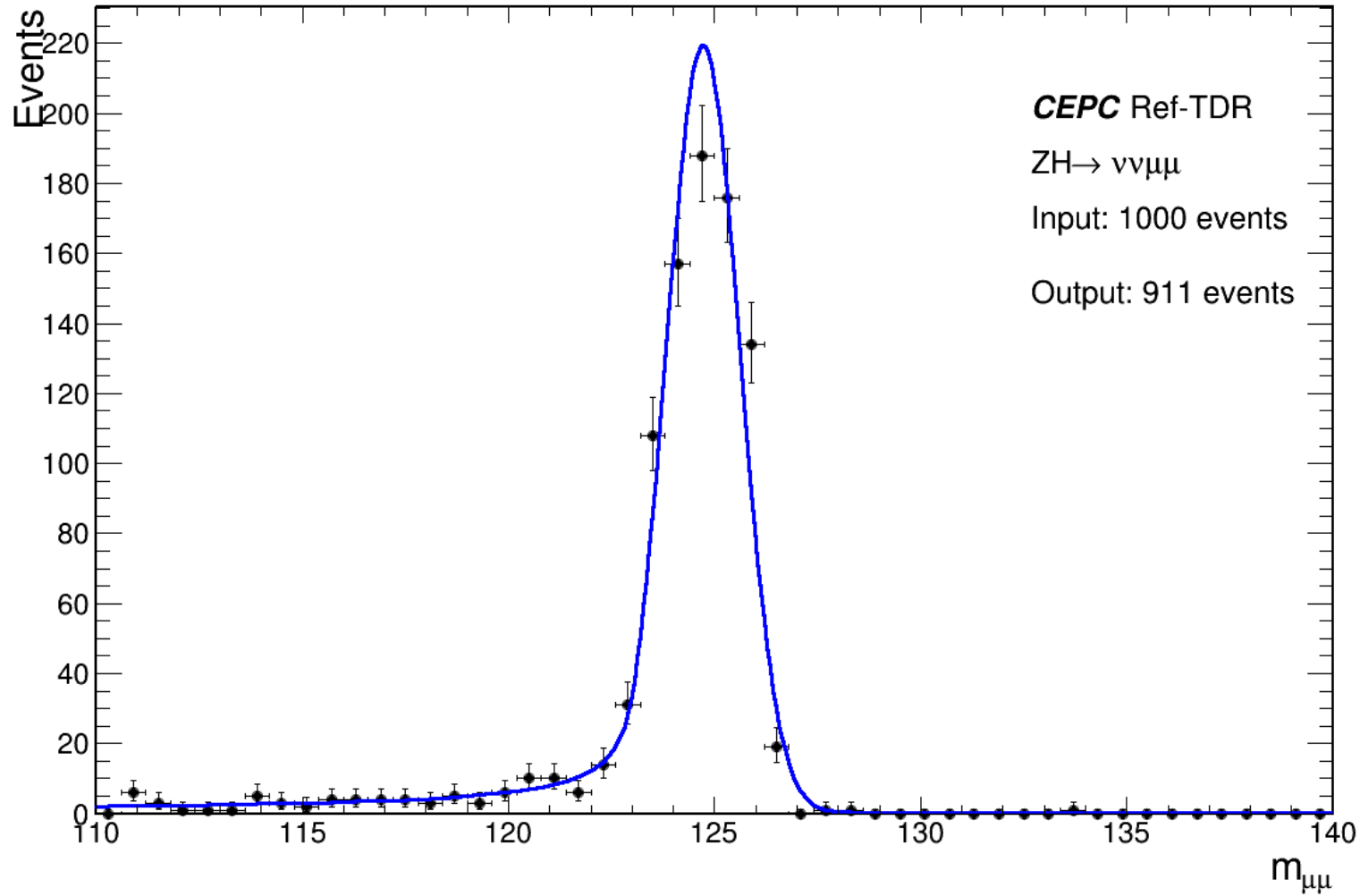
Choose 2 largest PFO.
Mean value: (High energy leakage) Calibration?
Resolution $\sim 2\%$. In expectation.



Dimuon

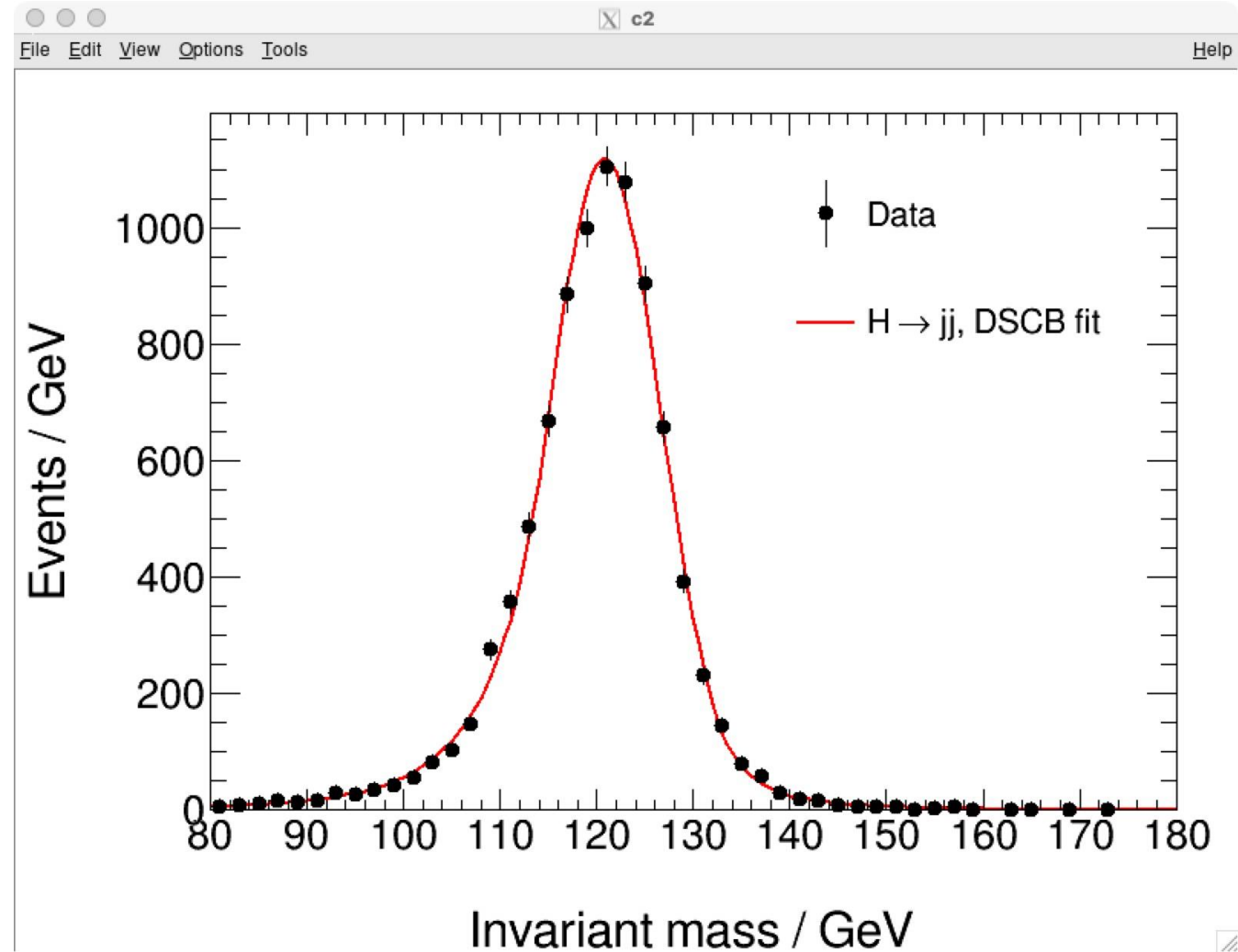


Choose 2 largest PFO and different charges.
One bug in PxPyPz with charge.



Dijet

BMR~5%.



backup

Software framework



- CEPCsoft (old) / CEPCSW(new)

The past sample, under /cefs/data, based on CEPC_v4 layout, mainly generated before 2018.

Need maintenance and update.

- Old: ILCsoft/Marlin/Whizard/Delphes.....

- Whizard 1.95 as default generator: <https://github.com/lhprojects/WhizardAis>
- Delphes card for fast simulation: [My talk @ Zhengzhou](#)
- Full simulation based on SiWECAL_GSHCAL;

- CEPCSW

- Need further work to migrate to new release.

Sample Category

Type	Format	Software	Ratio
Generator	stdhep	Whizard	2
Full Simu	Slcio/root		100
Reco	Slcio/root		10
Fast Simu	Ntuple	Delphes	1
Analysis	Ntuple		1

Sample generation types followed by Mo Xin, Gang and CEPC higgs white paper.

For **one event**,

Disk quota:

full simulation dominate the event size and take **~1MB**.

Taken all parts together, 1 event need **~1.2MB** disk space.

Time occupancy

1 event full simulation take **~1 cpu** 1 minute.

In CEPCSW, PFA reconstruction now need more **~5 min**.

To meet the requirement for TDR, plan to generate a new series sample for **CEPC SM-complete sample set**.

Priority:

1. 240 GeV Higgs sample
2. 240 GeV Bkg
3. 360 GeV Higgs/Bkg/ttbar
4. 91.2 GeV Z-pole
5. 160 GeV WW threshold

Plan one small, urgent sample list to run first for RefTDR interview.

Generator Whizard stdhep file ready.

Prior sample

As currently the analysis chain is not complete, Samples for both analysis and performance study.



Energy	Z decay	H decay	Event size
240 GeV	$Z \rightarrow \nu\nu$	$H \rightarrow uu$	100k
		$H \rightarrow dd$	100k
		$H \rightarrow ss$	100k
		$H \rightarrow cc$	100k
		$H \rightarrow gg$	100k
		$H \rightarrow bb$	100k
		$H \rightarrow \mu\mu$	100k
	$H \rightarrow \gamma\gamma$	100k	
	$Z \rightarrow qq$	$H \rightarrow ZZ \rightarrow \nu\nu\nu\nu$	100k
	WWfusion ($\nu\nu$)	$H \rightarrow bb$	100k
$Z \rightarrow ee$	$H \rightarrow$ inclusive	100k	
$Z \rightarrow \mu\mu$	$H \rightarrow$ inclusive	100k	
360 GeV	$Z \rightarrow \nu\nu$	$H \rightarrow bb$	100k
	WWfusion ($\nu\nu$)	$H \rightarrow bb$	100k

1.4M events total.
 1 event 6min.
 ->
 1000cpu 144hours.
 ~6 days.

Physics Benchmarks @ RefTDR



	Processes @ c.m.s.	Domain	Relevant Det. Performance
$H \rightarrow ss/cc/sb$	vvH @ 240 GeV	Higgs	PFA + Jet Origin Id (Jol)
$H \rightarrow inv$	qqH	Higgs/NP	PFA
Vcb	$WW \rightarrow lvqq$ @ 240/160 GeV	Flavor	Jol + Pid (Lepton, tau)
W fusion Xsec	vvH @ 360 GeV	Higgs	PFA + Jol
α_s	$Z \rightarrow \text{tautau}$ @ 91.2 GeV	QCD	PFA: Tau & Tau final state id
CKM angle $\gamma - 2\beta$	$Z \rightarrow bb, B \rightarrow DK$ @ 91.2 GeV	Flavor	PFA + Jol + Pid (Kaon)
Weak mixing angle	Z @ 91.2 GeV	EW	Jol
Higgs recoil	llH	Higgs	Pid (Lepton), track dP/P
$H \rightarrow bb, gg$	$vvH + qqH$	Higgs	PFA + Jol + Color Singlet id
$H \rightarrow di$ muon	qqH	Higgs	PFA, Leptons id, Tracking
$H \rightarrow di$ photon	qqH	Higgs	PFA, Photons id, EM resolution
W mass & Width	W threshold scan @160 GeV	EW	Beam energy
Top mass & Width	Top threshold scan @360 GeV	EW	Beam energy
$B_s \rightarrow vv\phi$	91.2 GeV	Flavor	Object (ϕ) in jets; MET
$B_c \rightarrow \tau\nu$	91.2 GeV	Flavor	Object (τ) in jets; MET
$B_0 \rightarrow 2\pi^0$	91.2 GeV	Flavor	π^0 in jets; EM resolution

Summary



- Need 1.4M Higgs sample as prior for October;
 - For both analysis and performance study
 - ~1000CPU 1week.
- Need several prior task
 - Jet Clustering implementation.
 - PFA reconstruction optimization.
 - Particle PID.
 - More convenient output structure like ntuples for analyzers.
 -

Backup (SM-complete set)

240/360 GeV Cross Sections



20iab data in 240GeV, 1iab in 360GeV.

Signals(fb)	240	360
ZH	196.9	126.6
WW fusion	6.2	29.61
ZZ fusion	0.5	2.80
Total	203.6	159.0
Total Events	4M	0.16M

Background(pb)		240	360
2fermion	$ee(\gamma)$	930	325
	$\mu\mu(\gamma)$	5.3	2.1
	$qq(\gamma)$	54.1	23.2
	$t\bar{t}$	\	0.317
4fermion	WW	16.7	10.0
	ZZ	1.1	0.63
	sZ	4.54	5.78
	sW	5.09	6.00

Higgs sample: 240GeV, Z->ll

Z decay	H decay	Event size
$Z \rightarrow ee$	$H \rightarrow u, d, s, c, g$	100k
	$H \rightarrow bb$	600k
	$H \rightarrow \tau\tau, \mu\mu, \gamma\gamma, Z\gamma$	100k
	$H \rightarrow ZZ$ full hardonic	100k
	$H \rightarrow ZZ$ semi leptonic	500k
	$H \rightarrow ZZ$ invisible	100k
	$H \rightarrow WW$ full hardonic	300k
	$H \rightarrow WW$ semi leptonic	500k

Sample generated exclusively, for each category, minimum sample set: **100k**.

Higgs sample generated this way and use weight to combine as one inclusive sample.
This event size make sure at least 1:1 to 20iab events.

$Z \rightarrow ee$ then corresponds **3M** events. So did $Z \rightarrow \mu\mu$ and $Z \rightarrow \tau\tau$.

Higgs sample: following

Generated 30M events corresponds to 4M, this set make sure at least 1:1 ratio.

$Z \rightarrow \nu\nu$ usually takes as the benchmark sample, so take **2X** of $Z \rightarrow ee$, corresponds **6M** events.

240GeV WWfusion ($\nu\nu h$) $\sim 17\%$ of $Z \rightarrow \nu\nu$.

Instead generate as $Z \rightarrow ee$, 3M.

240GeV ZZfusion ($ee h$) $\sim 7\%$ of $Z \rightarrow ee$.

Instead generate as $Z \rightarrow ee$, 3M.

For **interference**: need further study.

[See:Jun Ping's Whizard tutorial slides.](#)

$Z \rightarrow qq$ take **5X. 15M.**

example02: select Feynman diagrams in "whizard.prc"

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

$$e^+ e^- \rightarrow e^+ e^- H$$

Considering:

360GeV ZH Higgs	use 1/10 of 240 GeV Higgs.	3M
360GeV $\nu\nu H$ Higgs:	as 240GeV $Z \rightarrow ee$,	3M
360GeV $ee H$ Higgs:	1/10 of $Z \rightarrow ee$,	300k

In this way, CEPC 240GeV Higgs sample need **30M events** and 360GeV **6.3M.**

```
# Tag          In      Out          Method  Option
#=====
#####
#
nlnlh_o       e1,E1  n1,N1,h      omega   w:c,c
nlnlh_s_o     e1,E1  n1,N1,h      omega   w:c,c,r:3+4~Z
nlnlh_t_o     e1,E1  n1,N1,h      omega   w:c,c,r:1+3~W- && 2+4~W+

elelh_o       e1,E1  e1,E1,h      omega   w:c,c
elelh_s_o     e1,E1  e1,E1,h      omega   w:c,c,r:3+4~Z
elelh_t_o     e1,E1  e1,E1,h      omega   w:c,c,r:1+3~Z && 2+4~Z
```

2, 4fermion bkg

```
# Fully hadronic
#
# ZZ : only one flavour, or two but not ud or cs
#
zz_h0utut e1,E1 up_type_q,up_type_q,up_type_q,up_type_q omega w:c,c
zz_h0dtdt e1,E1 down_type_q,down_type_q,down_type_q,down_type_q omega w:c,c
zz_h0uu_notd e1,E1 uq,uq,not_dq,not_dq omega w:c,c
zz_h0cc_nots e1,E1 cq,cq,not_sq,not_sq omega w:c,c
#
# WW: more than 2 flavours
#
ww_h0cuxx e1,E1 cq,down_type_q,uq,down_type_q omega w:c,c
ww_h0uubd e1,E1 uq,bq,uq,dq omega w:c,c
ww_h0uusd e1,E1 uq,sq,uq,dq omega w:c,c
ww_h0ccbs e1,E1 cq,bq,cq,sq omega w:c,c
ww_h0ccds e1,E1 cq,dq,cq,sq omega w:c,c
#
# ZZ/WW mix: two flavours ud or cs
#
zzorww_h0udud e1,E1 uq,dq,dq,uq omega w:c,c
zzorww_h0csos e1,E1 cq,sq,sq,cq omega w:c,c
```

```
# Semi-leptonic
#
# ZZ : qq + two charged or two neutral leptons
#
zz_s10nu_up e1,E1 not_nu_e,not_nu_e,up_type_q,up_type_q omega w:c,c
zz_s10tau_down e1,E1 tau,tau,down_type_q,down_type_q omega w:c,c
zz_s10mu_down e1,E1 mu,mu,down_type_q,down_type_q omega w:c,c
zz_s10nu_down e1,E1 not_nu_e,not_nu_e,down_type_q,down_type_q omega w:c,c
zz_s10tau_up e1,E1 up_type_q,up_type_q,tau,tau omega w:c,c
zz_s10mu_up e1,E1 up_type_q,up_type_q,mu,mu omega w:c,c
#
# WW : qq l nu
#
ww_s10tauq e1,E1 up_type_q,down_type_q,tau,neutrino omega w:c,c
ww_s10muq e1,E1 up_type_q,down_type_q,mu,neutrino omega w:c,c
#
# leptonic
#
# ZZ : four charged, or charged and neutral of differnt flavour
#
zz_l04tau e1,E1 tau,tau,tau,tau omega w:c,c
zz_l04mu e1,E1 mu,mu,mu,mu omega w:c,c
zz_l0tauuu e1,E1 tau,tau,mu,mu omega w:c,c
zz_l0mumu e1,E1 nu_tau,nu_tau,mu,mu omega w:c,c
zz_l0tautau e1,E1 nu_mu,nu_mu,tau,tau omega w:c,c
#
# WW : two charged of different flavour
#
ww_l011 e1,E1 nu_mu,mu,nu_tau,tau omega w:c,c
#
# ZZ/WW mix 1l nunu, all same flavour
#
zzorvw_l0mumu e1,E1 nu_mu,mu,mu,nu_mu omega w:c,c
zzorvw_l0tautau e1,E1 nu_tau,tau,tau,nu_tau omega w:c,c
```

```
# Single Z+ee: two electrons+ anything except two nu_e
#
sze_l0tau e1,E1 electron,electron,tau,tau omega w:c,c
sze_l0mu e1,E1 electron,electron,mu,mu omega w:c,c
sze_l0e e1,E1 electron,electron,electron,electron omega w:c,c
sze_sl0dd e1,E1 electron,electron,down_type_q,down_type_q omega w:c,c
sze_sl0uu e1,E1 up_type_q,up_type_q,electron,electron omega w:c,c
sze_l0nunu e1,E1 not_nu_e,not_nu_e,electron,electron omega w:c,c
#
# Single W e nu_e + anything except e nu_e
#
sw_sl0qq e1,E1 up_type_q,down_type_q,electron,nu_e omega w:c,c
sw_l0tau e1,E1 not_nu_e,tau,electron,nu_e omega w:c,c
sw_l0mu e1,E1 not_nu_e,mu,electron,nu_e omega w:c,c
#
# Single Z/Single W mix: ee e_nu_e_nu
#
szeorsw_l01 e1,E1 nu_e,electron,electron,nu_e omega w:c,c
#
# Single Z + nunu two nu_e + any pair except two electrons
#
sznu_l0mumu e1,E1 nu_e,nu_e,mu,mu omega w:c,c
sznu_l0tautau e1,E1 nu_e,nu_e,tau,tau omega w:c,c
sznu_sl0nu_up e1,E1 nu_e,nu_e,up_type_q,up_type_q omega w:c,c
sznu_sl0nu_down e1,E1 nu_e,nu_e,down_type_q,down_type_q omega w:c,c
```

- 2f bkg ~20B events for 240GeV.
- Generate 1M for each category in 240GeV (ee, mm, tautau, usdcb) and 100k for 360GeV.
- 1M for 360GeV ttbar. 100k for 240/360GeV nn.
- **10M for 2f.**
- Full 4f ~500M events.
- Plan: 11 types full hadronic channel, each 5M
30 semi-leptonic channel each 5M
- 360GeV 1:10.
- 205M for 4f 240GeV and 20.5M for 360GeV.
- Need ~**260TB** disk space.

Z-pole/WW-threshold sample

Table 3.2: CEPC operation plan (@ 50 MW)

Particle	$E_{c.m.}$ (GeV)	L per IP ($10^{34} \text{ cm}^{-2} \text{ s}^{-1}$)	Integrated L per year (ab^{-1} , 2 IPs)	Years	Total Integrated L (ab^{-1} , 2 IPs)	Total no. of events
H	240	8.3	2.2	10	21.6	4.3×10^6
Z	91	192*	50	2	100	4.1×10^{12}
W	160	26.7	6.9	1	6.9	2.1×10^8
$t\bar{t}$ **	360	0.8	0.2	5	1.0	0.6×10^6

* Detector solenoid field is 2 Tesla during Z operation.

** $t\bar{t}$ operation is optional.

- Expect 210M WW and 4T Z-pole.
- Instead, for the first step, plan to generate:
 - $Z \rightarrow ee/\mu\mu/\tau\tau$ @91.2GeV : each 2M
 - $Z \rightarrow u\bar{d}sc\bar{b}$ @91.2GeV : each 1M
 - WW@160GeV: 7 types full hadronic: each 1M
 - 5 times semi-leptonic each 1M
- In total 23M events.
 - Extendable like 10x in the future.

CEPC SM-complete set

Generate Order	Sample set	Event amount	Disk Space(*1.2MB)
1	240 Higgs	36M	43.2T
2	240 bkg	200M	240T
3	360 Higgs	6.3M	8T
4	360 bkg(ttbar)	22M	27T
5	91.2 Z-pole	11M	14T
6	160 WW-threshold	12M	15T

- CEPC SM-complete set generate **287.3M events** in total and need **340T storage**.
 - /cefs can handle.
 - 1 event full simulation take ~1 cpu 1 minute.
 - 300M->**5M cpu*h**. (**1000cpu*200day**)
 - The fast simulation set will be generated first. (Also due to CEPCSW need further validation)
- Requirements and contributions are welcome. Any number can change.

Extra sample for CEPC benchmark

- Additional:

- Z->vv usdcbg each 1M
- Z->vv dimuon +1M
- Z->vv diphoton +1M

Special signals like:

Z->bb B->DK@91.2 500k

.....

	Processes @ c.m.s.	Domain	Relevant Det. Performance
H→ss/cc/sb	vvH @ 240 GeV	Higgs	PFA + Jet Origin Id (Jol)
H→inv	qqH	Higgs/NP	PFA
Vcb	WW→lvqq @ 240/160 GeV	Flavor	Jol + Pid (Lepton, tau)
W fusion Xsec	vvH @ 360 GeV	Higgs	PFA + Jol
α_s	Z→tautau @ 91.2 GeV	QCD	PFA: Tau & Tau final state id
CKM angle $\gamma - 2\beta$	Z→bb, B→DK @ 91.2 GeV	Flavor	PFA + Jol + Pid (Kaon)
Weak mixing angle	Z@ 91.2 GeV	EW	Jol
Higgs recoil	llH	Higgs	Pid (Lepton), track dP/P
H→bb, gg	vvH + qqH	Higgs	PFA + Jol + Color Singlet id
H→di muon	qqH	Higgs	PFA, Leptons id, Tracking
H→di photon	qqH	Higgs	PFA, Photons id, EM resolution
W mass & Width	W threshold scan @160 GeV	EW	Beam energy
Top mass & Width	Top threshold scan @360 GeV	EW	Beam energy
Bs→vv ϕ	91.2 GeV	Flavor	Object (ϕ) in jets; MET
Bc→ $\tau\nu$	91.2 GeV	Flavor	Object (τ) in jets; MET
B0→2 π^0	91.2 GeV	Flavor	π^0 in jets; EM resolution

- CEPC SM-complete set need **281.3M events** in total and need **340T storage**.
 - Disk space can handle but cpu hours is a challenge.
 - Other crucial benchmarks need <20M events.
- Migration/Validation to new framework CEPCSW.
 - Also cross checks with other generators like Pythia, Herwig, Madgraphs.....?