Plan and Validation of 240 GeV Samples

Xianghu Zhao, Gang Li, Manqi Ruan

June 28, 2018

Outline

- Production status
- Plan for 240 GeV samples production
- Deal user’s requirement for MC samples

Production Status

- Generation data are ready for 240, 250, 350 GeV
  - Higgs signal
  - 2 fermion and 4 fermion background
- Using WHIZARD generator version 1.95
- For detailed list, please refer to the web page
  - http://cepcsoft.ihep.ac.cn/guides/Generation/docs/ExistingSamples/

CEPC v4 MC Samples, 240 GeV

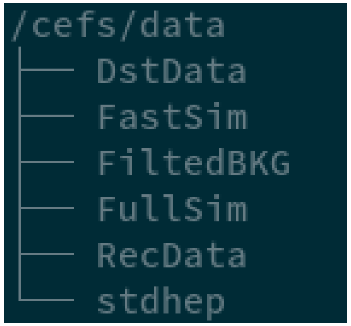
- Producing started, reconstructed with compressed dst format
- Existing samples
  - Higgs signal, samples with specified Higgs final states
  - qq background, 1M events
  - ww_l, ww_sl, zz_h, zz_l, zz_sl backgrounds, not all fully produced
- Data location
  - /cefs/data/FullSim/CEPC240/CEPC_v4
  - /cefs/data/DstData/CEPC240/CEPC_v4

CEPC v1 MC samples, 250 GeV

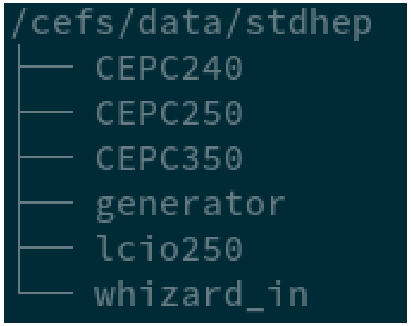
- Most samples are ready for use
- Simulation:
  - All signals, most 4 fermion backgrounds, part of 2 fermion backgrounds
- Reconstruction:
  - Mostly reconstructed corresponding to simulation
  - New data are being reconstructed with updated Arbor version, will soon be available
- Data location:
  - /cefs/data/FullSim/CEPC250/CEPC_v1
  - /cefs/data/RecData/CEPC250/CEPC_v1

How to Find the Samples

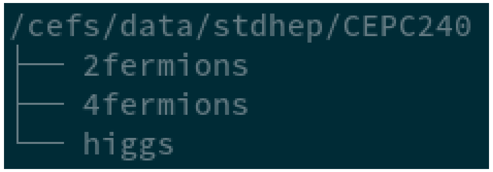
- Official samples are produced according to the generation data
  - http://cepcsoft.ihep.ac.cn/guides/Generation/docs/ExistingSamples/
- Data files could be found on the IHEP farm
  - All data are located under /cefs/data
- Sample data type
  - stdhep, FullSim, FastSim, RecData, DstData
- Each data type has samples of different energy
  - Simulation and reconstruction data also divided into different CEPC detector version



```
/cefs/data
├── DstData
├── FastSim
├── FiltedBKG
├── FullSim
├── RecData
└── stdhep
```



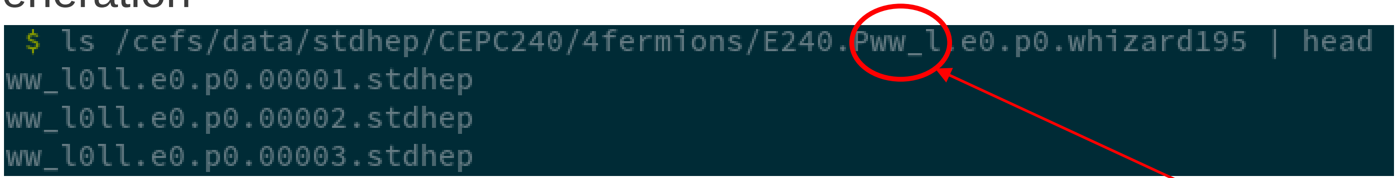
```
/cefs/data/stdhep
├── CEPC240
├── CEPC250
├── CEPC350
├── generator
├── lcio250
└── whizard_in
```



```
/cefs/data/stdhep/CEPC240
├── 2fermions
├── 4fermions
└── higgs
```

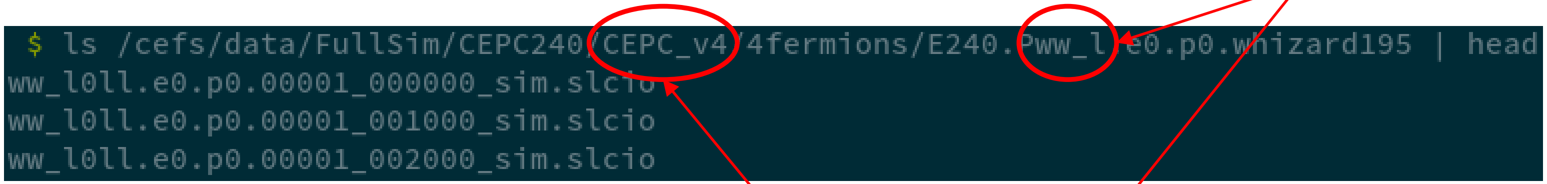
Data Sample Location Convention

- Similar structure for generator, simulation and reconstruction data
- Generation



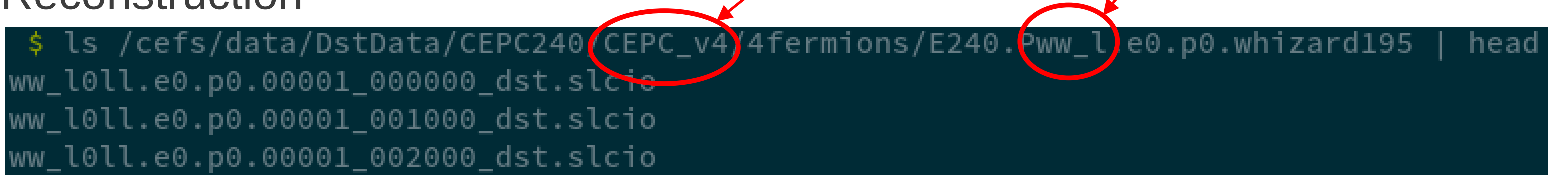
```
 $ ls /cefs/data/stdhep/CEPC240/4fermions/E240.Pww_l.e0.p0.whizard195 | head
ww_l0ll.e0.p0.00001.stdhep
ww_l0ll.e0.p0.00002.stdhep
ww_l0ll.e0.p0.00003.stdhep
```

- Simulation



```
 $ ls /cefs/data/FullSim/CEPC240/CEPC_v4/4fermions/E240.Pww_l.e0.p0.whizard195 | head
ww_l0ll.e0.p0.00001_000000_sim.slcio
ww_l0ll.e0.p0.00001_001000_sim.slcio
ww_l0ll.e0.p0.00001_002000_sim.slcio
```

- Reconstruction



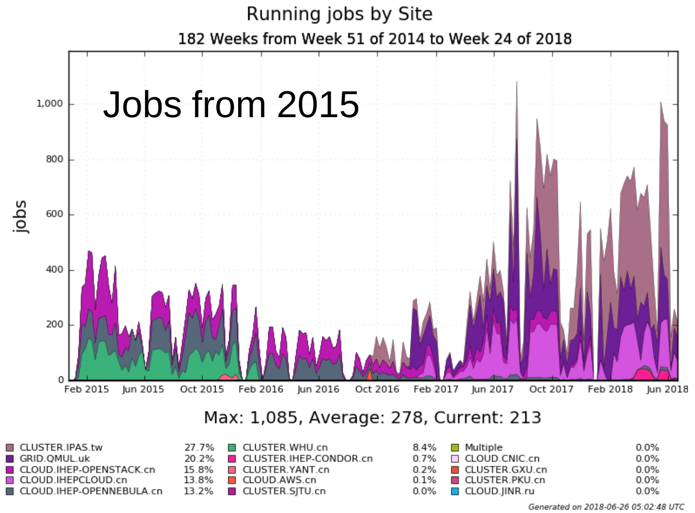
```
 $ ls /cefs/data/DstData/CEPC240/CEPC_v4/4fermions/E240.Pww_l.e0.p0.whizard195 | head
ww_l0ll.e0.p0.00001_000000_dst.slcio
ww_l0ll.e0.p0.00001_001000_dst.slcio
ww_l0ll.e0.p0.00001_002000_dst.slcio
```

- For detailed naming rules of each process, please refer to note
  - “Generated sample status for cepc simulation studies”, CEPC-TLS-GEN-2015-001

Computing Resources

- The current computing resources are really limited
  - Most from distributed computing
- Thank QMUL, IPAS and IHEP cloud for their contributions
- Welcome more sites to contribute in distributed computing

CLUSTER.IHEP-CONDOR.cn	48
CLOUD.IHEPCLOUD.cn	200
GRID.QMUL.uk	1600
CLUSTER.IPAS.tw	500
CLUSTER.SJTU.cn	100

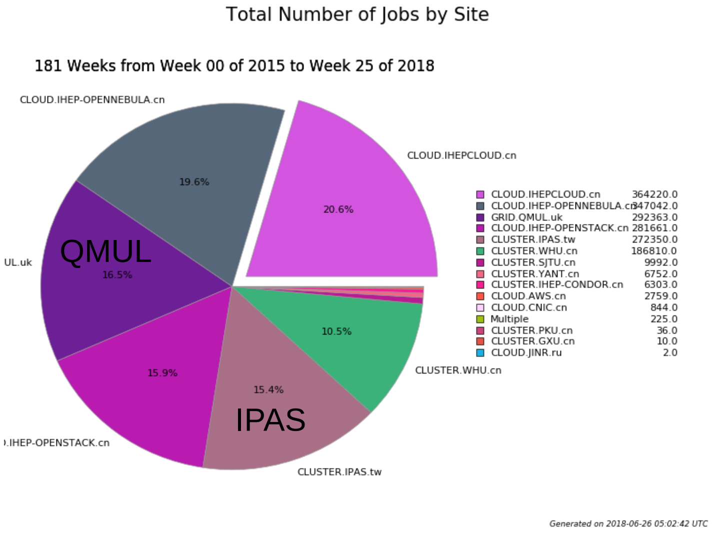

Running jobs by Site

182 Weeks from Week 51 of 2014 to Week 24 of 2018

Jobs from 2015

Max: 1,085, Average: 278, Current: 213

Generated on 2018-06-26 05:02:48 UTC


Total Number of Jobs by Site

181 Weeks from Week 00 of 2015 to Week 25 of 2018

QMUL

IPAS

Generated on 2018-06-26 05:02:42 UTC

Plan for 240 GeV Production

- Samples will be produced under CEPC_v4
  - Produce with CEPC software version 1.x
- Basic priority
  - Higgs signal
  - 4 fermion background
    - Leptonic
    - Semileptonic
    - Hadron
  - 2 fermion background
- Computing resources
  - Distributed computing will still be the main resource type
  - Try to utilize HPC resources

Timeline for 240 GeV Production

- Production speed vary a lot among different sites
- The speed is also related to event type
- Estimate with 1 minute per event on average, 500 cpu cores simultaneously

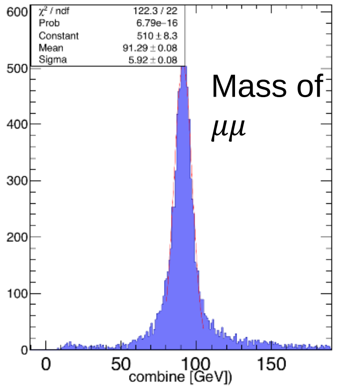
Process	Events	Time Estimated
Signal	~ 1M	2 days
4 fermion leptonic	~ 16M	1 month
4 fermion sl	~ 42M	2 months
4 fermion hadronic	~ 38M	2 months
2 fermion	~ 1250M (only produce part)	~ 2 months

Validation for Samples

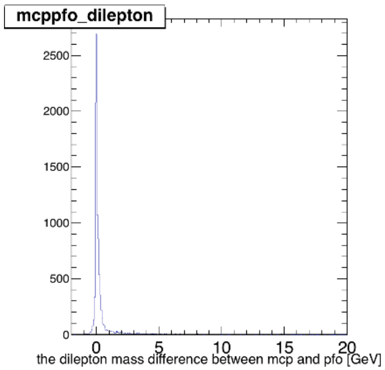
- Validation of samples is important to ensure the data are correctly produced
- Each sample need its own validation method
  - We will provide validation method for each sample
  - Apply validation after each sample finished
- Validations are implemented by checking
  - Physics results
  - Reconstruction performance obtained from comparing MC truth

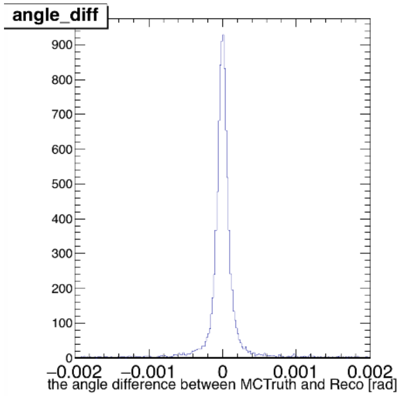
Validation Examples

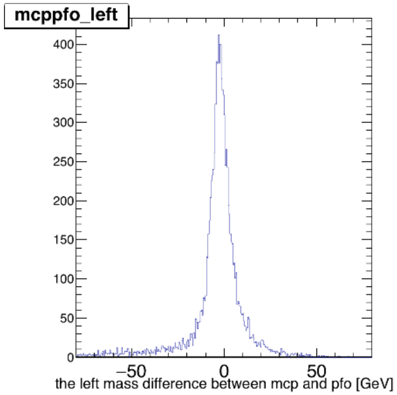
- zz_sl0mu_up



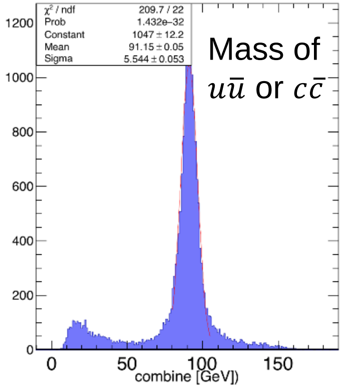
Mass of $\mu\mu$



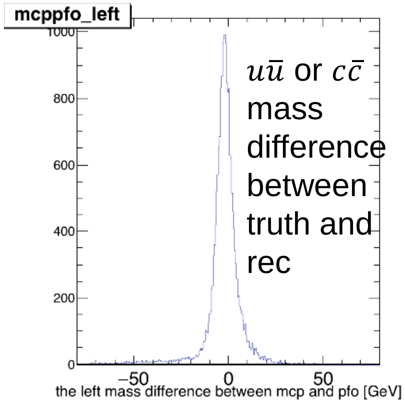




- zz_sl_nu_up



Mass of $u\bar{u}$ or $c\bar{c}$

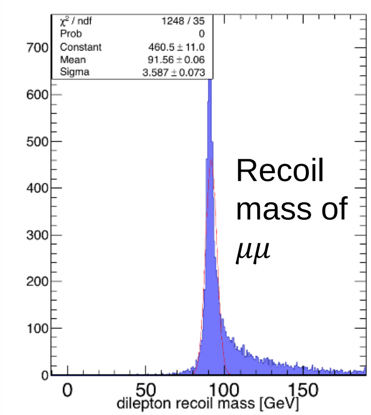


$u\bar{u}$ or $c\bar{c}$ mass difference between truth and rec

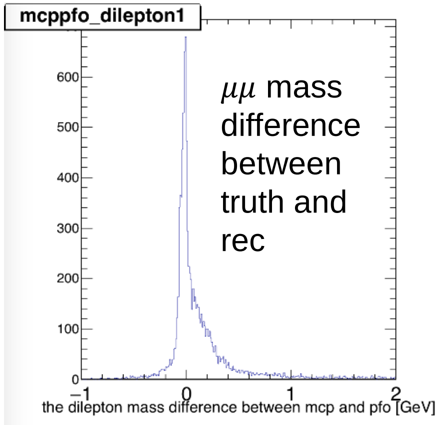
From Yongfeng's result

Validation Examples

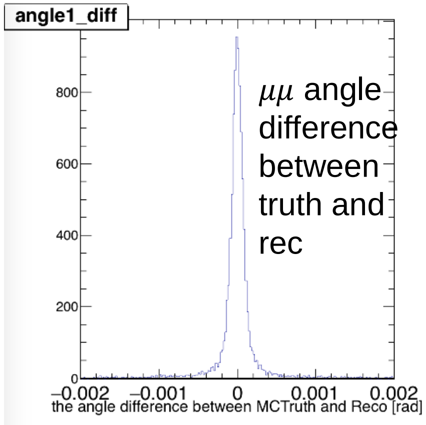
- zz_l0_mumu



Recoil mass of $\mu\mu$

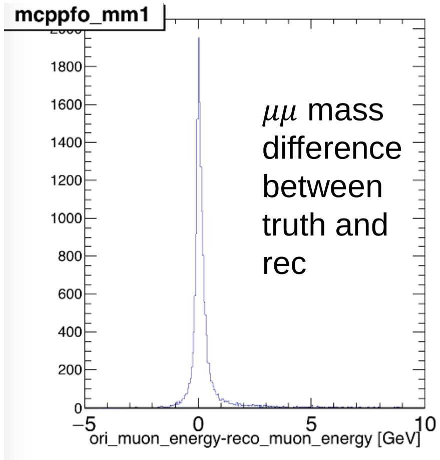


$\mu\mu$ mass difference between truth and rec

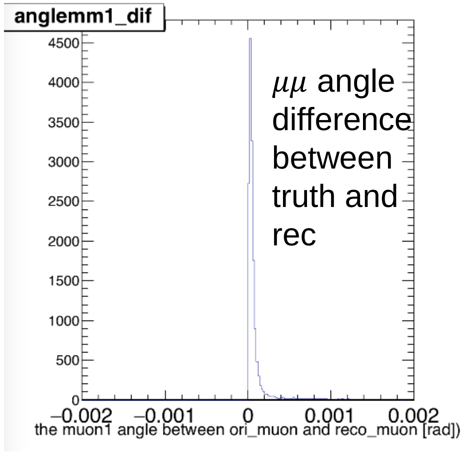


$\mu\mu$ angle difference between truth and rec

- ww_l



$\mu\mu$ mass difference between truth and rec



$\mu\mu$ angle difference between truth and rec

From Yongfeng’s result

User Requirement for Data Samples

- If some samples are urgent for your analysis and not present, please send your request
- If the data are in the official list, but not yet produced
  - Raise the priority if sample size is small
  - For large samples, we could first produce part of them
- Not in the list, and may be common sample
  - Like the smart final state processes
  - Add to the production list and produce
- For small and uncommon samples, you could also produce it yourself

CEPC Software Release Status

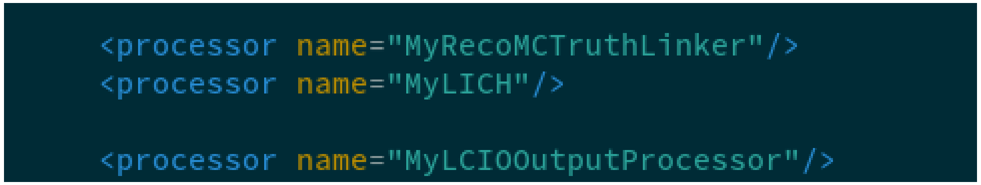
- Official CEPC software release flow is established
- One release includes all sub packages of CEPC software
  - MokkaC, Arbor, ROOT, Geant4…
  - All packages and their versions are bind to a specific CEPC software release version now
- Release version for CEPC_v4
  - Current version is 0.1.0-rc9, this version is in testing status
  - 1.0 version will be officially released in the next month
  - All release version for CEPC v4 will be 1.x.y

Brief Usage for CEPC Software

- Release 0.1.0-rc9 could be directly used on lxslc6
  - source /cvmfs/cepc.ihep.ac.cn/software/cepcenv/setup.sh
  - cepcenv use 0.1.0-rc9
- CEPC software could also be installed on your local PC
  - Detailed guides on installation could be found on web
  - http://cepcsoft.ihep.ac.cn/guides/scratch/docs/local/
- A simple example of $\nu\nu H, H \to \mu\mu$ is provided
  - wget http://cepcsoft.ihep.ac.cn/file/example/cepc_example.tar.gz
  - This example includes simulation, reconstruction and a script to draw the $\mu\mu$ mass
  - Run this example to verify everything is working correctly

About LICH Processor

- LICH processor is developed by Dan Yu in order to improve particle identification efficiency
- This algorithm is already integrated in release 0.1.0-rc9
- Some configurations are needed in the Marlin steering file
  - MyLICH processor must be enabled before MyLCIOOutputProcessor
  - MyRecoMCTruthLinker should be enabled before MyLICH if not
  - More parameters could be set in the MyLICH section
- A new collection “typedPFOs” will be added to the slcio file
  - “typedPFOs” could be safely used to replace “ArborPFOs” in the analysis with better PID



```
<processor name="MyRecoMCTruthLinker"/>
<processor name="MyLICH"/>

<processor name="MyLCIOOutputProcessor"/>
```

Issues related to “typedPFOs”

- Is it suitable to just rename typedPFOs to ArborPFOs?
  - This could simplify analysis, but may cause inconsistency with old reconstructed files
  - Or let them coexist?
- Dst files should also include typedPFOs?
- Will it affect fast simulation?

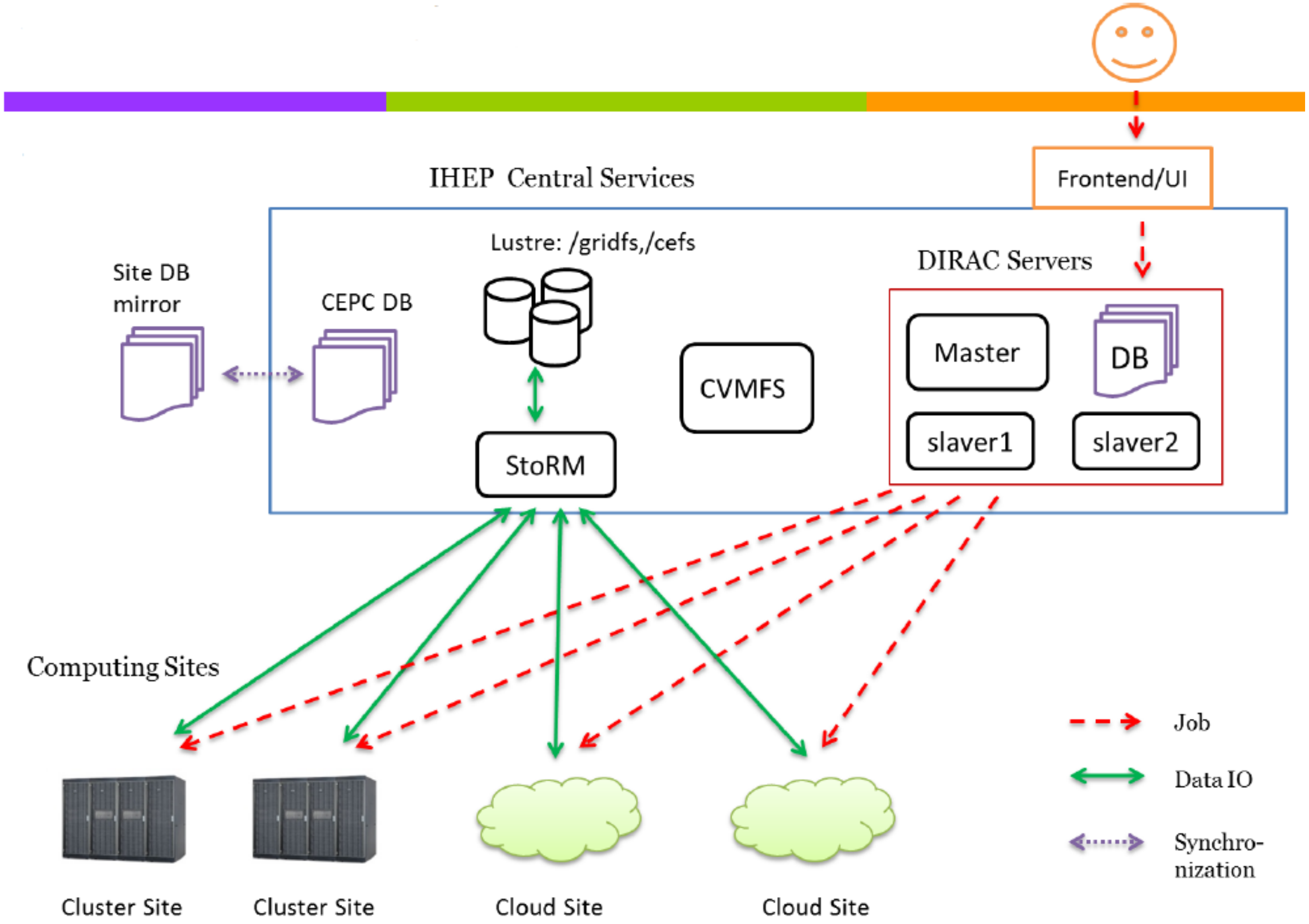
Summary

- MC samples are being produced with various resources steadily
- Some CEPC v4 240 GeV samples are already available
- CEPC v1 250 GeV samples are almost done
- User’s requirements could be considered in the production procedure

Thanks

Backup

Distributed System Architecture

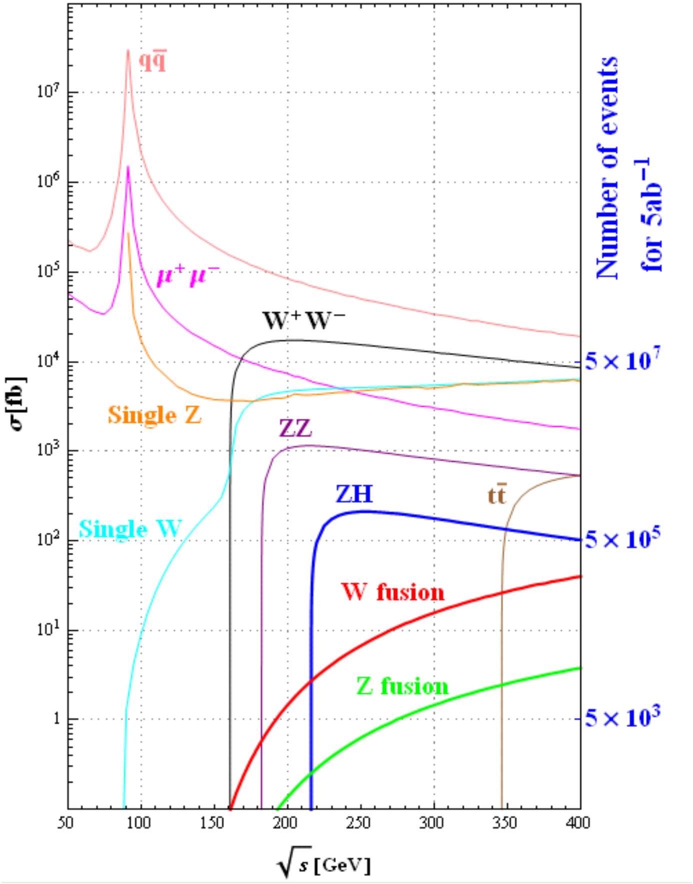


Problems in Production

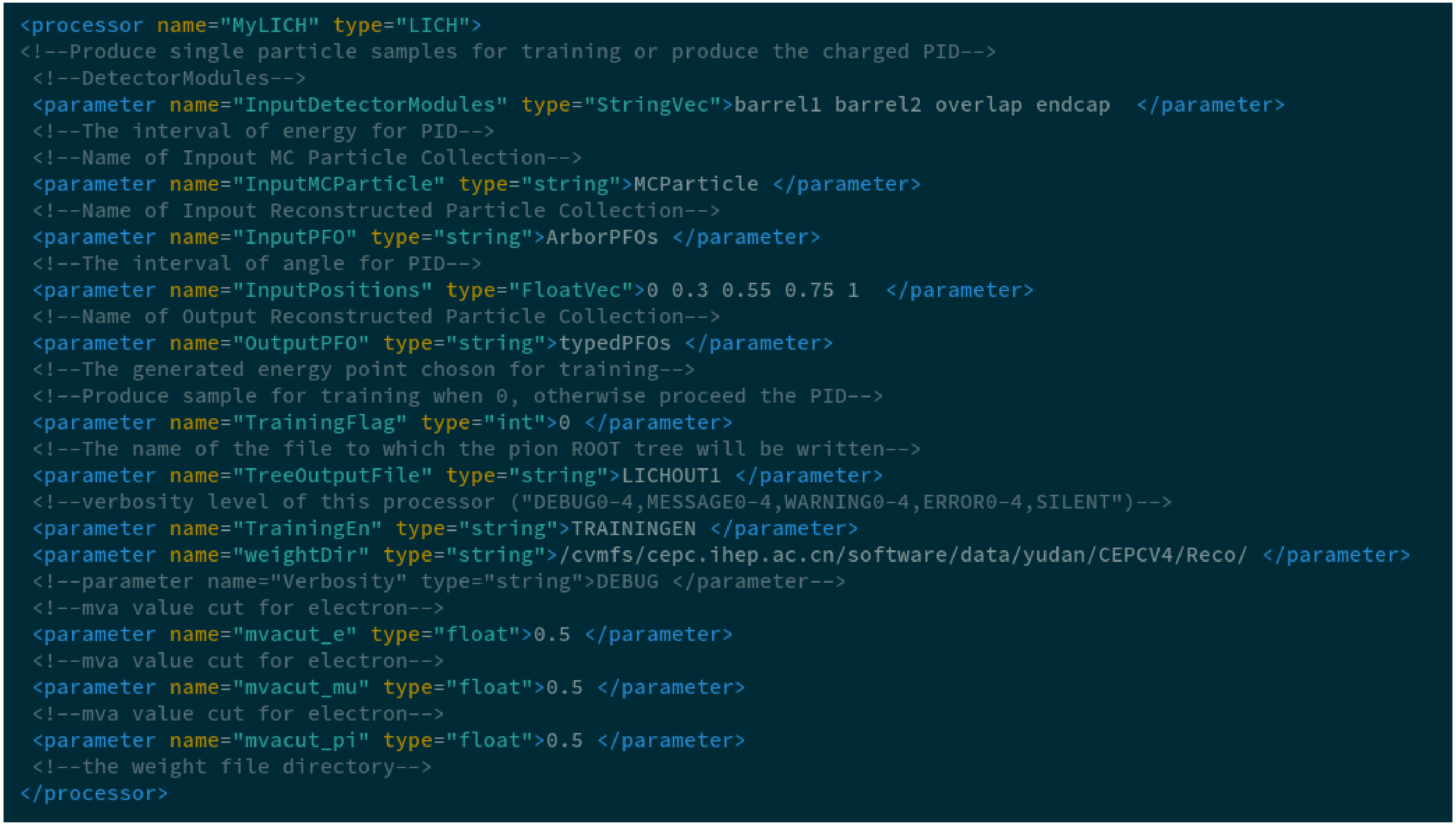
- Database access and synchronize
  - Access remote database could always be
- Data transfer
- Distributed system stability

Cross Sections

- Cross Sections of the leading Standard Model processes



LICH Parameters



```
<processor name="MyLICH" type="LICH">
<!--Produce single particle samples for training or produce the charged PID-->
 <!--DetectorModules-->
 <parameter name="InputDetectorModules" type="StringVec">barrel1 barrel2 overlap endcap  </parameter>
 <!--The interval of energy for PID-->
 <!--Name of Inpout MC Particle Collection-->
 <parameter name="InputMCParticle" type="string">MCParticle </parameter>
 <!--Name of Inpout Reconstructed Particle Collection-->
 <parameter name="InputPFO" type="string">ArborPFOs </parameter>
 <!--The interval of angle for PID-->
 <parameter name="InputPositions" type="FloatVec">0 0.3 0.55 0.75 1  </parameter>
 <!--Name of Output Reconstructed Particle Collection-->
 <parameter name="OutputPFO" type="string">typedPFOs </parameter>
 <!--The generated energy point choson for training-->
 <!--Produce sample for training when 0, otherwise proceed the PID-->
 <parameter name="TrainingFlag" type="int">0 </parameter>
 <!--The name of the file to which the pion ROOT tree will be written-->
 <parameter name="TreeOutputFile" type="string">LICHOUT1 </parameter>
 <!--verbosity level of this processor ("DEBUG0-4,MESSAGE0-4,WARNING0-4,ERROR0-4,SILENT")-->
 <parameter name="TrainingEn" type="string">TRAININGEN </parameter>
 <parameter name="weightDir" type="string">/cvmfs/cepc.ihep.ac.cn/software/data/yudan/CEPCV4/Reco/ </parameter>
 <!--parameter name="Verbosity" type="string">DEBUG </parameter-->
 <!--mva value cut for electron-->
 <parameter name="mvacut_e" type="float">0.5 </parameter>
 <!--mva value cut for electron-->
 <parameter name="mvacut_mu" type="float">0.5 </parameter>
 <!--mva value cut for electron-->
 <parameter name="mvacut_pi" type="float">0.5 </parameter>
 <!--the weight file directory-->
</processor>
```