CEPCSoft and Massive Production of the CEPC Simulation

Xianghu Zhao, Manqi Ruan, Gang Li

Nov 14, 2018

Outline

- CEPCSoft status

- Massive production of full simulation

- Future development of CEPCSoft

- Summary

CEPCSoft Release Status

- Official CEPCSoft release flow is established

- One release includes all sub packages of CEPCSoft

  - MokkaC, Arbor, ROOT, Geant4…

  - All packages and their versions are bind to a specific CEPC software release version

  - All release versions are defined in git repository

    - https://github.com/cepc/cepc-release

- Release version for CEPC_v4

  - Current version is 0.1.0

  - http://cepcsoft.ihep.ac.cn/releases/0.1.0/

CEPCSoft Management Tool

- CEPCEnv is developed for simplifying the management of CEPCSoft
    - Solve the package dependencies
    - Create software release definition
    - Installation of full software
    - Environment setup / clean / switch
    - Package backup
- Manage the CEPCSoft with cepcenv command
    - cepcenv install 0.1.0
    - cepcenv ls
    - cepcenv use 0.1.0
    - http://cepcsoft.ihep.ac.cn/guides/scratch/docs/cepcenv/
- Could also be easily adopted by other experiment or software

CEPCSoft Deployment

- CEPCSoft is deployed on CVMFS (CernVM File System)
  - All versions are located under /cvmfs/cepc.ihep.ac.cn/software/cepcsoft
  - Simple and efficient way for both local farm and distributed computing
  - Essential data are also deployed on CVMFS, e.g. LICH weight data
- 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/
- Use CEPCSoft with docker
  - Docker image is created and deployed on docker hub
    - https://hub.docker.com/r/cepc/cvmfs/
  - docker run --privileged -i -t cepc/cvmfs

Brief Usage for CEPCSoft

- Release 0.1.0 could be directly used on lxslc6 by the cepcenv command

    - source /cvmfs/cepc.ihep.ac.cn/software/cepcenv/setup.sh

    - cepcenv use 0.1.0

- A simple example of $\nu\nu H, H \rightarrow \mu\mu$ is provided to verify environment

    - http://cepcgit.ihep.ac.cn/cepcsoft/TestExample

    - This example includes simulation, reconstruction and a script to draw the $\mu\mu$ mass

    - Run this example to verify everything is working correctly

- More details could be found on the web

    - http://cepcsoft.ihep.ac.cn/guides/scratch/docs/quick_start/

CEPCSoft Web Site

■ Web site is established for publishing CEPCSoft

    ■ Release information

    ■ Documentation

    ■ Hold the software packages

■ Everyone is encouraged to contribute to the content

■ http://cepcsoft.ihep.ac.cn

Massive Production

- Massive MC samples are produced
    - Based on CEPCSoft
    - Including Higgs signal and backgrounds
    - Physics analysis and performance study
- Generation data are ready for 240, 250, 350 GeV
    - Higgs signal, 2 fermion and 4 fermion background
    - Using WHIZARD generator version 1.95
    - http://cepcsoft.ihep.ac.cn/guides/Generation/docs/Introduction/
- Simulation and reconstruction for massive production are applied for each detector version from the above generation data

CEPC v4 MC Production

- All samples produced with CEPCSoft version 0.1.0

- 240 GeV

- Full simulation with MokkaC 0.1.1

- Reconstruction output condensed dst format

- Data location on IHEP farm

    - /cefs/data/FullSim/CEPC240/CEPC_v4

    - /cefs/data/DstData/CEPC240/CEPC_v4

Currently Available Samples

	Energy	Data Type	Location
CEPC_v4	240 GeV	Simulation	/cefs/data/FullSim/CEPC240/CEPC_v4
		Reconstruction	/cefs/data/DstData/CEPC240/CEPC_v4
CEPC_v1	250 GeV	Simulation	/cefs/data/FullSim/CEPC250/CEPC_v1
		Reconstruction	/cefs/data/RecData/CEPC250/CEPC_v1

- Sub samples are located in different directories according to the generation data

  - http://cepcsoft.ihep.ac.cn/guides/Generation/docs/ExistingSamples/

- 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 in Massive Production

- The current computing resources are really limited
  - Most from distributed computing
  - Shared resources from the IHEP local farm
- 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
202 Weeks from Week 51 of 2014 to Week 44 of 2018

CEPC Production Jobs from 2015

Max: 1,085, Average: 277, Current: 8.07

CLUSTER.IPAS.tw	27.4%	CLUSTER.NEU.tr	2.4%	Multiple	0.0%
GRID.QMUL.uk	21.3%	CLUSTER.IHEP-CONDOR.cn	0.7%	CLOUD.CNIC.cn	0.0%
CLOUD.IHEP-OPENSTACK.cn	14.3%	DIRAC.Client.cn	0.2%	CLUSTER.GXU.cn	0.0%
CLOUD.IHEPCLOUD.cn	13.7%	CLUSTER.SJTU.cn	0.2%	CLUSTER.PKU.cn	0.0%
CLOUD.IHEP-OPENNEBULA.cn	11.9%	CLUSTER.YANT.cn	0.2%	CLOUD.JINR.ru	0.0%
CLUSTER.WHU.cn	7.6%	CLOUD.AWS.cn	0.1%		

Generated on 2018-11-13 04:18:15 UTC

Total Number of Jobs by Site

201 Weeks from Week 00 of 2015 to Week 45 of 2018

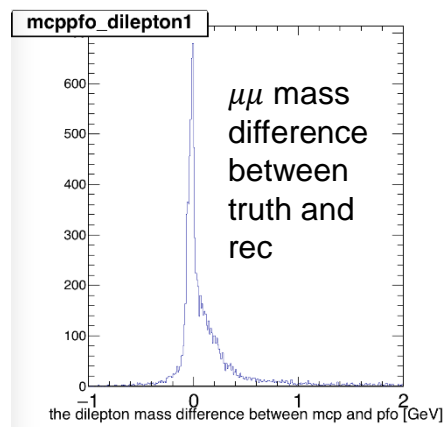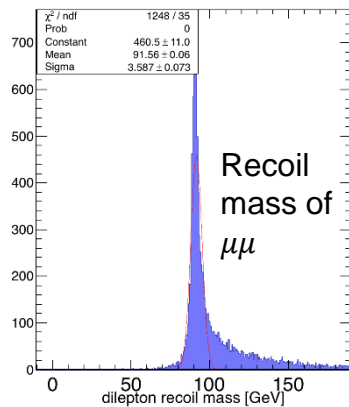|	
CLOUD.IHEPCLOUD.cn	592734.0
GRID.QMUL.uk	569774.0
CLUSTER.IPAS.tw	541526.0
CLOUD.IHEP-OPENNEBULA.cn	364372.0
CLOUD.IHEP-OPENSTACK.cn	281661.0
CLUSTER.WHU.cn	186810.0
CLUSTER.NEU.tr	160815.0
CLUSTER.SJTU.cn	32102.0
CLUSTER.IHEP-CONDOR.cn	31910.0
CLUSTER.YANT.cn	6752.0
DIRAC.Client.cn	3682.0
CLOUD.AWS.cn	2759.0
CLOUD.CNIC.cn	844.0
Multiple	271.0
CLUSTER.PKU.cn	36.0
CLUSTER.GXU.cn	10.0
CLOUD.JINR.ru	2.0

Generated on 2018-11-13 04:20:10 UTC
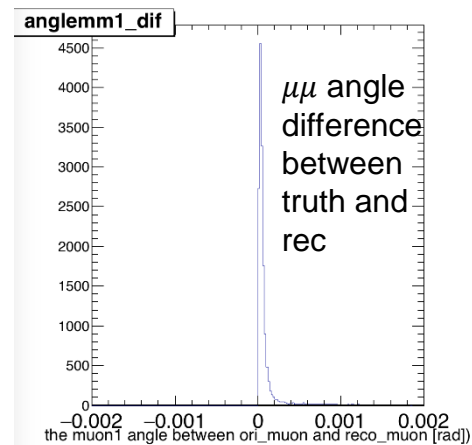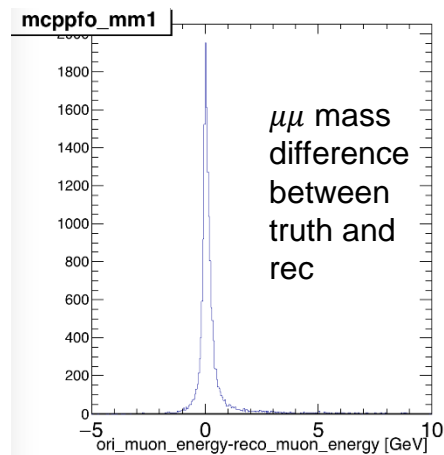
Validation for Samples

- Validation of samples is important to ensure the samples are correctly produced

- Each process 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 with MC truth

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
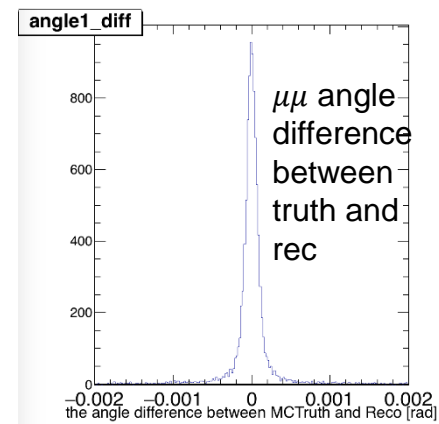
From Yongfeng's result

Future Development for CEPCSoft

- Describe CEPC detector geometry with DD4hep
    - No more need for database which is used in MokkaC
    - Better support for HPC and distributed computing
    - Need to update several packages in CEPCSoft and require intensive validation
- Extend the functionality of CEPCSoft management tool
- Support for volunteer computing resources (BOINC)
    - See Ran Du's poster CEPC@home for details
- Production system including job and data management
- Automate validation for both release and production

Summary

■ CEPCSoft release flow and management tool is established

■ Version 0.1.0 for CEPC v4 is released

■ Massive production are processing smoothly based on CEPCSoft

■ MC samples for CEPC v4 and v1 are now available for analysis

Thanks

Backup

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

Process

```
$ 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

Detector version

```
$ 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 by Xin Mo

  - "Generated sample status for cepc simulation studies", CEPC-TLS-GEN-2015-001

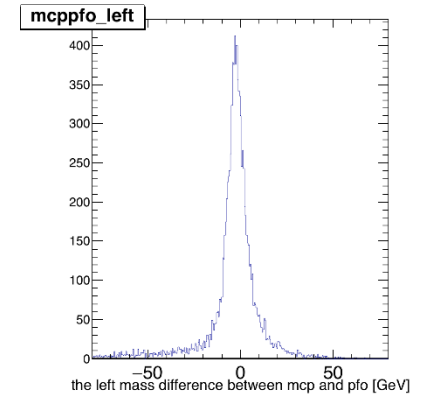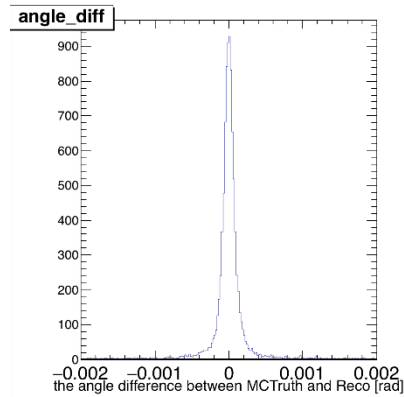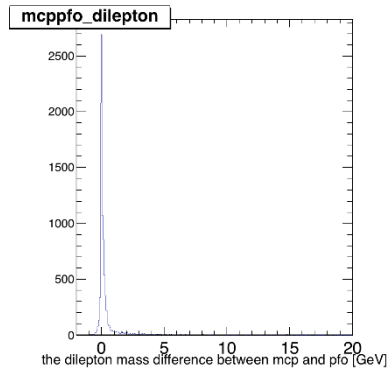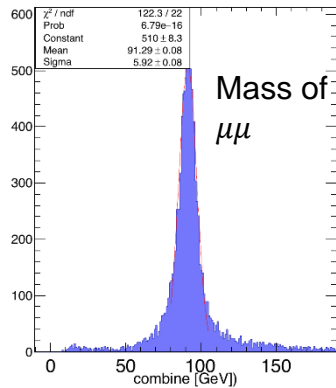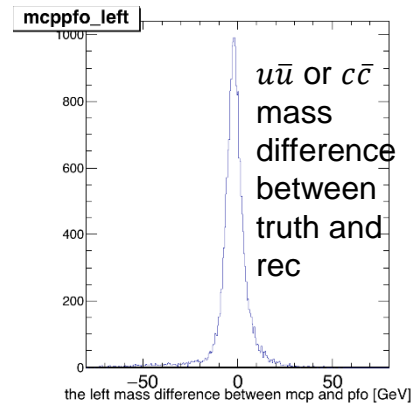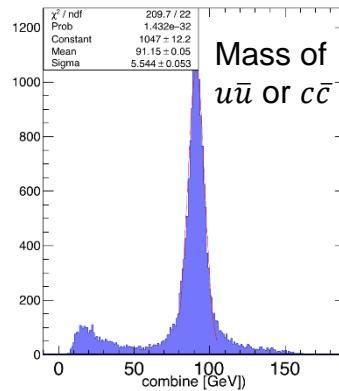# User Requirement for Data Samples

- 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

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