Simulation, Reconstruction and Softwares @ CEPC

Manqi, on behalf of the simulation group...
Outline

- SCRAC Software Status
- Simulation
- Reconstruction
  - Tracking
  - PFA
  - Flavor Tagging
- Detector optimization
- Perspective & Roadmap
SCRAC

Generator
Whizard, PYTHIA
(ref samples
Available)

Full Simulation
Mokka, DD4HEP
(req: Geant4,
Database...)

Digitization
MarlinReco, etc.
eg: G2CD for
Calorimeter

Tracking
MarlinReco, etc.

Stdhep

LCIO

GDML/Root

Event Display: Druid

PFA
Pandora/Arbor

Flavor Tagging
LCFIPlus

Analysis
Marlin, Root,
FSClasser

Physics
result

Simulation Calibration Reconstruction Analysis Chain
Simulation - geometries

- Nankai University
  - **Xu yin**, Qiuyang, Yanli, etc
  - Release of official CEPC detector models (CEPC_v1, CEPC_o_v2, ...) & Hit map – tracking level validation, etc
  - Irradiation study & MDI design implementation

- IHEP
  - **Fu Chengdong**, Shuzheng, etc
  - Extension of Mokka
    - General HGC toolkit
    - SLCIO input plugin for generators
  - DD4HEP
Reconstruction:
How the physics objects are Served to analyzers?
PFA oriented Detector & Reconstruction
5-10% $X_0$ material before Calorimeter
Octagon structure of Calorimeter Barrel
Tracking

- Objective: for any geometry, produce & understand the track response:
  - Track efficiency & multiplicity
  - Differential resolution of D0, Z0, Phi, TanLambda(Theta), Omega(P_T)
- Team: Libo, Wu Linghui & Zhang Yao
  - Libo, Tracking Expert at ILC, voluntarily supporting CEPC study
- Status:
  - Kalman-Filter based technology
  - Good understand of TPC – Silicon based geometry;
  - Pure Silicon geometry – ongoing work
  - Digitization need profound understanding...
Arbor @ single particle

07/04/2016

Charged

Neutral
Arbor @ single particle

Numbers of ChiPFO for Muon

Numbers of ChiPFO for Electron

Numbers of ChiPFO for Pion

Numbers of NePFO for Gamma

Numbers of NePFO for Neutron

Charged Particles

Photons

Neutral Hadrons

07/04/2016
Separation

Tiny inefficiency: bridging effect by fragments
Lepton ID

Signal Classification & Background rejection.

Almost everything you want to measure at electron-positron collider...

Higgs: Even for $H\rightarrow bb$, $cc$, $gg$ measurement, the lepton number provides useful information for $b/c$ tagging
Lepton ID @ Arbor V3

- Developed by **BINSONG**
- E > 10 GeV, Efficiency for Lepton ID > 99%...
- Leads to elegant
  - Higgs recoil analysis with Z → leptons
  - Higgs-→ZZ/WW→leptonic/semi-leptonic final states
    - We are advanced comparing to ILC studies
    - Improve signal efficiency by 2 times comparing to PreCDR
    - More details: See Gang & Yuqian's talk
Lepton ID, next step

- Developed by **YUDAN** (Joint Ph.D)
- TMVA based
  - For each charged PFO, provide a electron-likeness and muon-likeness value
  - Typical working point:
    - eff > 99% for E > 20 GeV Lepton
    - eff > 97% for E > 5 GeV Pion
- To be polished, encapsulated & integrated
- More importantly: We understand
  - We know where/why the limit is;
  - We know how to improve
Bremsstrahlung photon recovery of electron/positron

CEPC
Preliminary

- Developed by Zhenxing, Binsong, Wanglei, etc

\[ M_{\text{recoil}}^{e^+e^-} (\text{GeV}/c^2) \]

Events/(1 GeV/c^2)

07/04/2016
Arbor: photon reconstruction

Angular Correlation of EM Shower energy response
Arbor: photon reconstruction

$M_{\gamma\gamma}$ without geometry correction

- Full
- Fast

$M_{\gamma\gamma}$ with $\theta$ & $\phi$ correction

- Full
- Fast

Histograms showing $M_{\gamma\gamma}$ distribution before and after $\theta$ & $\phi$ correction.
Photon conversion & recovery

Save ~ 10% of the H->di photon statistic
Arbor: Tau reconstruction

- In no-jet environment: counting number of charged particle – (pions & leptons), photons (pi0s) + restrict impact parameters leads to very high efficiency in Tau finding:
  - At inclusive Higgs decay sample: Efficiency ~ 98% for of H→tautau event finding, with llH and vvH final state. The remaining bkgrds are H→WW/ZZ→leptonic/tau final state
  - More detail: see Gang's talk

To be covered soon...
Arbor: JER/MET

- MET: usually no ambiguity;
- Jet: Highly depending on Jet clustering if #Jet > 2...
KD algorithm boost: $N^2 \rightarrow N \log(N)$
Flavor Tagging

TMVA based method from ILC Study:
http://indico.ihep.ac.cn/event/5592/contribution/16/material/slides/0.pdf
btag performance

\[ \text{btag} > 0.95 \]

\[ \text{btag} > 0.90 \]

\[ \text{btag} > 0.40 \]

\[ \text{btag} > 0.30 \]
ctag performance

- **ctag > 0.95**
  - b jet
  - c jet
  - uds jet

- **ctag > 0.90**
  - b jet
  - c jet
  - uds jet

- **ctag > 0.40**
  - b jet
  - c jet
  - uds jet

- **ctag > 0.30**
  - b jet
  - c jet
  - uds jet
Analysis tool

FSClassifier Vs self-wrote analysis code...
Pandora is great

2007, M. Thomson, Optimising GLDC for PFA

\[ \alpha = 0.315 \left( \frac{B}{4} \right)^{-0.19} \left( \frac{R}{1.68} \right)^{-0.49} (1 + 6.3e^{-\frac{N}{8.0}}) \]

\[ \alpha = 0.42 \left( \frac{B}{4} \right)^{-0.31} \left( \frac{R}{1.78} \right)^{-0.61} (1 + 21.6e^{-\frac{N}{7.1}}) \]

http://www.hep.ph.ic.ac.uk/calice/others/070530lcs07/thomson1.pdf
V1, used for preCDR

V2, May 2015, Improved PID

V3, Nov 2015, Improved JER

KD, Jan 2016, Boost Speed

Short of resource...
Optimisation
$\delta m_H = 36.286 \times (1 + 0.092 \times e^{-1.820 - R_{TPC}})$ MeV.

$\frac{\delta \sigma_{ZH}}{\sigma_{ZH}} = 0.485 \times (1 + e^{-0.094 \times R_{TPC}})$

$H \rightarrow \text{di photon branching ratio measurement}$
Calorimeter: Saturation & Leakage

Calo optimization effort: UCAS(陈石), IHEP(成栋, 赵航, 树正, 学正, etc)

Eg, Saturate at 175 GeV photon, 20mm cell size: 2500 MIP

单层结构：2mm塑闪+2mmPCB+3mmW
入射粒子为gamma

L 1sigma value = 0.87x - 0.24yy + 0.97y - 0.43z + 0.82
x = log10(energy) y = log10(cell size) z = log10(angle)

Energy

Resolution (%)

28/01/2016
Calorimeter: Linearity & Resolution

- 敏感层 $\delta_{Si} = 0.3, 0.5, 0.75$ (单位: mm)
- 吸收层 $W = 1.4, 2.8, 5, 10, 15$ (单位: mm)
Calorimeter: Energy Estimation of Hadronic object

Default Energy Cut: 0.15 MIP, time < 10000 ns

Using Time, Energy & Nature information, Hadron energy measurement can be Dramatically improved... (up to 40%)
Parameter put by hand, motivated by:
Saving the cost.
Closer VTX inner layer, better flavor tagging?

Reconstruction: be aware of TPC boundary & B-Field Strength
*Implemented by Xuyin (NanKai U)*
Cost estimation: extrapolate from ILD

ILD -> CEPC_v1: reduced by 13%
CEPC_v1 -> CEPC_v2: reduced by further 25%
Performances: Tracks

D0

Z0

- cepc_v2
- cepc_v1
Performances: Tracks

Banana Shape emerged at high energy... \( \sim o(10^{-3}) \) level
Single particle: clustering

CEPC_v1

CEPC_o_v2

Splitting is more severe in new geometry; need adjustment
Performances: Lepton ID

Graph

\[ |\cos(\theta)| < 0.98 \]

- \( \mu \)
- \( e \)
- \( \pi \)

\( P \) (GeV/c)

PID Efficiency

\[ 1 \quad 0.98 \quad 0.95 \quad 0.8 \quad 0.6 \]

\[ 10^{-10} \quad 100 \]

a bit crazy – mainly due to the parameter shift due to the Calo change...
Performances: Overlap particles

Efficiency vs Distance/mm

- Eff of 1 Charged and 1 Neutral PFO
- Eff of Tot Reco En within 2 σ
- Eff of Position within 30mm

Anomaly in Matching observed

07/04/2016
Performances: Jets

- Without optimization... similar tendency observed at both Pandora & Arbor
Performances: Flavor Tagging

- c background
- uds background
- CEPC_v1 refs.

"1 - \varepsilon_{bkg}
\varepsilon_{sig}"

"1 - \varepsilon_{bkg}
\varepsilon_{sig}"
CEPC_v1 → CEPC_o_v2

W.R.T CEPC_v1, Reduce:

- Total cost ~ 25%
- ECAL power/FEE: 75%
- HCAL thickness/channels ~ 20%
- B-Field to 17% (3.5 → 3)
- VTX inner radius: 25%

Qualitatively: everything goes into the Expected direction

Quantitatively: ???

Reconstruction: Adapted, lots of effects needed for OPTIMIZATION, especially the PFA

<table>
<thead>
<tr>
<th>Performance</th>
<th>adapted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking: D0, Z0</td>
<td>20% ↑ @ E &lt; 20 GeV (VTX); 5% ↓ @ E &gt; 20 GeV (B-Field effect);</td>
</tr>
<tr>
<td>Lambda, Phi</td>
<td>worse</td>
</tr>
<tr>
<td>Omega</td>
<td>worse</td>
</tr>
<tr>
<td>PFA: Clustering</td>
<td>Slightly worse</td>
</tr>
<tr>
<td>Matching</td>
<td>worse</td>
</tr>
<tr>
<td>Separation</td>
<td>~10% ↓</td>
</tr>
<tr>
<td>PID</td>
<td>3-5% ↓ @ E &gt; 10 GeV; 10% ↓ @ E &lt; 10 GeV;</td>
</tr>
<tr>
<td>JER</td>
<td>20% ↓</td>
</tr>
<tr>
<td>Flavor Tagging</td>
<td>Improved up to 5%↑</td>
</tr>
</tbody>
</table>
CEPC_v1 → CEPC_o_v2

W.R.T CEPC_v1, Reduce:

- Total cost ~ 25%;
- ECAL power/FEE: 75%;
- HCAL thickness/channels ~ 20%;
- B-Field to 17% (3.5 → 3);
- VTX inner radius: 25%;

Qualitatively: everything goes into the expected direction

Quantitatively: ???

Reconstruction: Adapted, lots of effects needed for **OPTIMIZATION**, especially the PFA

<table>
<thead>
<tr>
<th>Performance</th>
<th>adapted</th>
<th>optimized*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking: D0, Z0</td>
<td>10% ↑ @ E &lt; 20 GeV (VTX); 5% ↓ @ E &gt; 20 GeV (B-Field);</td>
<td></td>
</tr>
<tr>
<td>Theta, Phi</td>
<td>worse</td>
<td>-</td>
</tr>
<tr>
<td>Omega</td>
<td>worse</td>
<td>-</td>
</tr>
<tr>
<td>PFA:Clustering</td>
<td>Slightly worse</td>
<td>same</td>
</tr>
<tr>
<td>Matching</td>
<td>~10% ↓</td>
<td>~5% ↓</td>
</tr>
<tr>
<td>Separation</td>
<td>~10% ↓</td>
<td>~2% ↓</td>
</tr>
<tr>
<td>PID</td>
<td>3-5% ↓ @ E &gt; 10 GeV; 10% ↓ @ E &lt; 10 GeV;</td>
<td>~1% ↓</td>
</tr>
<tr>
<td>JER</td>
<td>20% ↓</td>
<td>~10%↓</td>
</tr>
<tr>
<td>Flavor Tagging</td>
<td>Improved up to 5%↑</td>
<td>?</td>
</tr>
</tbody>
</table>

*My personal expectation...
CEPC\_v1 $\rightarrow$ CEPC\_o\_v2

W.R.T CEPC\_v1, Reduce:

- Total cost $\sim 25\%$;
- ECAL power/FEE: 75%;
- HCAL thickness/channels $\sim 20\%$;
- B-Field to 17% (3.5 $\rightarrow$ 3);
- VTX inner radius: 25%;

Qualitatively: everything goes into the expected direction

Quantitatively: ???

Reconstruction: Adapted, lots of effects needed for OPTIMIZATION, especially the PFA

<table>
<thead>
<tr>
<th>Performance</th>
<th>adapted</th>
<th>optimized*</th>
<th>Manpower/ people*month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking: D0, Z0</td>
<td>20% ↑ @ E &lt; 20 GeV (VTX); 5% ↓ @ E &gt; 20 GeV (B-Field);</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Theta, Phi</td>
<td>worse</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Omega</td>
<td>worse</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PFA: Clustering</td>
<td>Slightly worse</td>
<td>same</td>
<td>-</td>
</tr>
<tr>
<td>Matching</td>
<td>~10% ↓</td>
<td>~5% ↓</td>
<td>6</td>
</tr>
<tr>
<td>Separation</td>
<td>~10% ↓</td>
<td>~2% ↓</td>
<td>2</td>
</tr>
<tr>
<td>PID</td>
<td>3-5% ↓ @ E &gt; 10 GeV; 10% ↓ @ E &lt; 10 GeV;</td>
<td>~1% ↓</td>
<td>4</td>
</tr>
<tr>
<td>JER</td>
<td>20% ↓</td>
<td>~10% ↓</td>
<td>4</td>
</tr>
<tr>
<td>Flavor Tagging</td>
<td>Improved up to 5%↑</td>
<td>?</td>
<td>3</td>
</tr>
</tbody>
</table>

Given the current status, 23 people\*month is needed to fully optimize the performance.
For the next geometry, the needed manpower will be half
Activities & Cooperations

- Physics, optimization & PFA working group...
- Supported by IHEP innovation funding...
  - Short term visit;
  - Training;
  - Physics Software workshop(s);
  - Computing resource;
- Cooperations:
  - Chicago University & Wuhan University: Photon reconstruction & H->di photon analysis
  - SLAC: Irradiation studies, etc
  - LLR, France: PFA & Tau finder
  - Shandong University, ZhongShan University: Software framework
  - Nankai University: simulation toolkit
  - ...
Perspective at CDR

- **Benchmark detector(s)**
  - **Optimized PFA detector**
    - Scan key para (size; B-field)
    - Compare different technology: develop adequate digitizer, validate at prototype test...
    - Understand the MDI constrain
  - Other concepts/option?

- **Software**
  - Adopt & **Optimize** to Benchmark detector
  - Simulation: more realistic...
  - Framework & computing: prototype
  - **Software team: regular release & validation** (need at least 2 professional software experts...)

- **Physics**
  - Higgs, EW & BSM @ Benchmark Geometry
  - Scan Benchmark Physics analyses at different geometry

- **Decent Documentation & Publication**
5-year perspectives & resource demand

• Perspective:
  - Physics: understand the physics requirement to CEPC detector, and demonstrate the physics potential from Higgs, EW & BSM
  - Software: develop, maintain and optimize the full set of SCRAC tools, develop future software framework and computing tools
  - Detector Design **GLOBAL** optimization: taken into account the physics goal, constraints (collision environments, detector hardware technologies), detector performance and cost

• Resources:
  - Computing & Storage: 1 PB storage & 1k CPU for Higgs Run.
    • 1 M/evt; 1 evt/CPU*min;
    • Higgs Run: $o(10^9)$ Physics events: 1M Higgs; Z pole: $o(10^{10})$ Physics events
  - Manpower: 12 FTE 、 9 PostDoc 、 30 Students
    • Analysis: 2 FTE + 2 PostDoc + 14 Students
    • Software:
      - **SCRAC**: 5 FTE + 3 PostDoc + 6 Students;
      - Framework & Computing: 3 FTE + 2 PostDoc + 4 Students
    • **Optimization**:
      - 2 FTE + 2 PostDoc + 8 Students

*Personally I hope we also have outreach experts...*
Summary

- Status of SCRAC: healthy & lots of progress
  - Simu: mastered existing tool, freely edit geometry
  - Tracking & Flavor Tagging: mastered the ILC tools
  - PFA reconstruction: leptons, photons, taus, Jets are reconstructed at high efficiency & accuracies...

- Optimization:
  - Explored at many different P.o.V
  - The REAL game: SCRAC@CEPC_o_v2, adaption is straight forward, optimization demands lots of manpower & expertise

- Lots of Activities & Global cooperations

- Toward CDR: optimized SCRAC & benchmark geometries
  - One Benchmark geometry: PFA detector. Optimization is secured
  - Open to other concepts...

- In 5 years, personal vision:
  - Global optimization: to achieve the physics goal in a feasible & most efficient way – Joint efforts between Theory-Pheno study, Simulation, Detector hardware development and Accelerator Study...
  - Strong team: Software Frame & SCRAC, Analysis...
Thank you!

- Special thanks to Gang, Xuyin, Chengdong, Libo, Binsong, Dan, and the full analysis team – who not only produce the physics results, but also valid/polishes the reconstruction tools!

- Apologize for being not able to cover the discussion & progress made in Software framework & computing
Back up
Software framework consideration

Use an existing one 😊 vs Develop from beginning 😞

• Consideration of the choice for CEPC
  – Enough services and functionalities
  – Easy to use
  – Future supports

• Almost all widely used frameworks can satisfy our requirements

• Several potential candidates are investigated and compared
Framework candidates investigation

- Marlin: currently used by CEPC (with uncertain official support)
- Gaudi: very popular for collider physics experiments, most familiar to us, very comprehensive but a bit heavy
- ROOT: very flexible and powerful, but need more manpower for some service functionalities development
- ART: optimized for high intensity physics experiments and a little complex
- SNiPER: lightweight and optimized for non-collider experiments

<table>
<thead>
<tr>
<th></th>
<th>Marlin</th>
<th>Gaudi</th>
<th>ROOT</th>
<th>ART</th>
<th>SNiPER</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Interface</td>
<td>XML</td>
<td>Python, TXT</td>
<td>Root script</td>
<td>FHiCL</td>
<td>Python</td>
</tr>
<tr>
<td>Adoption</td>
<td>ILC</td>
<td>Atlas, BES3, DYB</td>
<td>Phenix, Alice</td>
<td>Mu2e, NOVa, LArSoft, LBNF</td>
<td>JUNO, LHAASO</td>
</tr>
</tbody>
</table>
Computing considerations

- It is still far to confirm the computing technology now used for 30 years more
- But we believe the technology is evolving step by step
- Now the main computing task is to study and follow the latest computing technology to prepare for the future, including
  - Cloud computing
  - Distributed computing
  - Multi-cores computing
  - High performance computing
  - Unified distributed data management and access
  - “Smart” network, high bandwidth future network
  - ......
Step by step: performance

Nvtx=0

Nvtx=1

Nvtx>1

2016-03-27  bbar  1st CEPC Physics & Software Workshop  ccbabar
Interesting/crucial topics to explore

- **Jet Clustering**
  - Distinguish 2 jets, 4 jets and 6 jets events from each other;
  - Identify the boson (color singlet) origin of different jets;
- **Analysis**
  - $\sigma(ZH)$ determination from qqH recoil
  - Data driven method to determine the Higgs observables
  - Systematic estimation for Higgs/Z pole runs
# Reconstruction

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking</td>
<td>Sub-detector Tracks</td>
</tr>
<tr>
<td></td>
<td>Merged Tracks</td>
</tr>
<tr>
<td>PFA: Clustering</td>
<td>Calorimeter Clusters</td>
</tr>
<tr>
<td>Matching</td>
<td>Final State Particles</td>
</tr>
<tr>
<td>PID</td>
<td>Final State Particles with Type flag/information</td>
</tr>
<tr>
<td>PFA: Advanced Object finder</td>
<td>Isolated Lepton</td>
</tr>
<tr>
<td></td>
<td>Converted Photon</td>
</tr>
<tr>
<td></td>
<td>Electron with BS photon recovery</td>
</tr>
<tr>
<td>Jet Clustering</td>
<td>Tau</td>
</tr>
<tr>
<td></td>
<td>MET</td>
</tr>
<tr>
<td>Flavor Tagging</td>
<td>Jets</td>
</tr>
<tr>
<td></td>
<td>Jets with b, c, uds, (gluon), multiple flags...</td>
</tr>
</tbody>
</table>
Detector optimization: general receipt

- Understand the motivation and carefully modify/edit the geometry accordingly
- Verify the detector geometry
  - Hit map
  - Object construction at sub-detector level: tracks, clusters & vertex
- Adjust/optimize the reconstruction & Understand the detector performance
  - Single particle level: reconstruction/id efficiency
  - Overlap particle level: separation performance, essential for PFA
  - Multi-particle object: Tau & Jets
    - *Tech. oriented, Time consuming & need strong expert (see Gang's talk)*
- Re-process the benchmark physics analysis
Theory-Pheno

Detector-Analysis
Theory-Pheno

- Model Building
- Feynman Rules
- Interpretation

Detector-Analysis

- Sample Production Validation
- Full/Fast Simulation
- Reconstruction Det. optimization
- Analysis result Conclusion
- Conclusion

Interpretation
Det. optimization
• **Team work...**
  - Theory-Phenology: Model Building - Interpretation
    - Description of Physics model & motivations
    - Propose newly observable/measurement
  - Detector-Analysis: Common SM background sample
    - Mutual: Maintain the interface
  - Phenomenology Generator development, NP sample production & Validation
    - Detector: Integration into the full chain
  - *Standard CEPC generator format should be discussed*

• **Vision: operational chain: MM->Sample->Simu/reco->Analysis->Interpretation**
  - Urgently needed: Devoting researcher with background from both sides
  
  • **Proposition:**
    - Pheno-Detector Forum,
    - At CEPC Physics-Software meeting (April, Aug & Nov-Dec, 3 times/year)
      - Phenomenology Generator School/Workshops
      - ...
        - Manpower allocation: Recruit Joint PostDoc/Ph.D
        - Support relevant works: short term visit, travel, etc