sPHENIX software and simulations status

Jin Huang (BNL)
Introduction
sPHENIX software

Open source @ GitHub
https://github.com/sPHENIX-Collaboration/
Core software: 200k line of code
Analysis: 59k line of code

Detector Design → G4 Simulation → Digitization → Reconstruction

Real Data

Hardware Design Decisions
without calibration
with calibration

p+p, 10 weeks

\( \sigma_{180} = 80 \pm 1.4 \text{ MeV} \)

Jin Huang
2nd sPHENIX Asia Meeting
Geant4 simulation and reconstruction are integrated in the sPHENIX software framework.

In production mode: run Geant4 sim in central production (CPU intensive), buffer the output file (DST) for reuse (require disk space), then run reconstruction in separated user sessions.

Ensure same configuration and geometry are used in simulation and reconstruction → embedding
  - For example, Geometry and magnetic field configuration in Geant4 is automatically passed down to reconstruction stage for use in alignment adjustment, in tracking Kalman filter and in calorimetry geometric presentation.
Repositories: https://github.com/sPHENIX-Collaboration
  ◦ **analysis**: which contains the analysis modules. Everyone who is a collaborator (request to join here) can update this repository directly.
  ◦ **coresoftware**: which contains the framework, the G4 simulation and the Event library to read raw data.
    • Pull request & review required for updates
  ◦ **macros**: which contains macros to run the show.
    • Pull request & review required for updates

Nightly build:
  ◦ Most recent software build for use on RCF daily (new builds)
  ◦ Weekly snapshot of software environment (ana builds)

Coding style convention
  ◦ Based on google code style with sPHENIX tweak: https://wiki.bnl.gov/sPHENIX/index.php/Codingconventions
  ◦ Automatic style format tool: clang-format

Further information:
Software QA

- Code management system: git
  - pull requests, followed by informal code review
- Nightly builds, including daily running of cpu expensive code analysis tools
- Continuous integration using short running checks and some standard analysis
- State of the art code analysis tools
  - Valgrind
  - Cppcheck
  - Scan-build
  - Coverity
  - Insure
  - Clang (produces identical results)
  - Include-what-you-use (clang based include file checker)
sPHENIX Continuous Integration

https://github.com/sPHENIX-Collaboration/coresoftware/pulls

Passed checks
Report available

Inner HCal QA plots for PR686, switching default inner HCal material

HCALIN fraction truth energy
- Reference
- New: KS-Test P=0.000

HCALIN visible sampling fraction
- Reference
- New: KS-Test P=0.000
RACF resources

- sPHENIX accounts are for collaborators who do not have an racf account
- Existing PHENIX, STAR (and likely all others) can be used to run sPHENIX software
  - needs the **sphenix** secondary group id, following day-1 checklist: [https://wiki.bnl.gov/sPHENIX/index.php/SPHENIX_software_day-1_checklist](https://wiki.bnl.gov/sPHENIX/index.php/SPHENIX_software_day-1_checklist)

- Disk resources (**user writable in bold**):
  - sPHENIX $HOME: 3GB
  - /sphenix/sim/sim01 200TB
  - /sphenix/sim/sim02 20TB
  - /sphenix/data/data01 10TB
  - /sphenix/data/data02 250TB
  - /sphenix/user 200TB
Portability and containers

- Simulation are suitable for distribute for opportunistic computing offsite, e.g. OSG
- Successful experience in distributing sPHENIX simulation via Singularity container
  - A light-weight virtual environment to reproduce RCF software environment offsite
  - Validated output to be consistent with RCF
  - Installation at Fudan and Sun Yat-Sen University

https://github.com/sPHENIX-Collaboration/Singularity

Singularity container for sPHENIX and EIC-sPHENIX

Singularity container for sPHENIX and EIC-sPHENIX allow collaborators to run sPHENIX RCF/SDCC environment with the nightly builds on your local computers or on external high-performance computing clusters.

This repository includes the instruction and local update macro for this Singularity container.

Validations: updatebuild.sh --build-new build passing --build-roots build passing

- standard macros
- git tutorials
- code reference
- Doxygen
- last commit

Counts

m_{ee} (GeV)

Weihu Ma (Fudan Univ)
Computing Plan - 2019
Computing plan - 2019


- Event processing time (MB HI)
  \[= 5_{\text{reco1}} + 5_{\text{cali}} + 5_{\text{reco2}} + 5_{\text{part\_flow}} + 4_{\text{cal}} = 24 \text{ sec}\]

- Prompt data processing ->
  \(~100-200k\) CPU cores

<table>
<thead>
<tr>
<th>CPU-cores</th>
<th>year-1</th>
<th>year-2</th>
<th>year-3</th>
<th>year-4</th>
<th>year-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>calorimetry</td>
<td>1.6E+04</td>
<td>2.4E+04</td>
<td>3.2E+04</td>
<td>2.4E+04</td>
<td>3.2E+04</td>
</tr>
<tr>
<td>tpc-calibration</td>
<td>1.9E+04</td>
<td>2.0E+04</td>
<td>4.0E+04</td>
<td>2.4E+04</td>
<td>4.0E+04</td>
</tr>
<tr>
<td>reconstruction</td>
<td>5.8E+04</td>
<td>7.9E+04</td>
<td>1.2E+05</td>
<td>7.9E+04</td>
<td>1.2E+05</td>
</tr>
<tr>
<td>analysis</td>
<td>1.9E+03</td>
<td>2.4E+03</td>
<td>4.0E+03</td>
<td>2.4E+03</td>
<td>4.0E+03</td>
</tr>
<tr>
<td>total (no ana)</td>
<td>9.3E+04</td>
<td>1.2E+05</td>
<td>1.9E+05</td>
<td>1.3E+05</td>
<td>1.9E+05</td>
</tr>
</tbody>
</table>
Simulation flow

Event generation
- Process
  - Support pile up and embedding generations
  - Generators: Pythia6/8, HIJING, (JetScape), EIC wrapper, anything HepMC-compatible
- Storage
  - HepMC objects in DST or standalone file
  - Particle at collision and immediate decay

Detector Simulation
- Process
  - Geant4 package
  - Detailed detector/field model
  - Store G4-particle interaction and truth-look up table
- Storage
  - Arrays of Geant4-hit objects
  - Field map configurations
  - Detailed geometry objects
  - Detector configuration parameters
  - Truth association table

Digitization
- Process
  - Convert Geant4 steps to ADC/TDC data
  - E.g. Light production, diffusion, noises, electronics, digitizers
  - Owned by detector groups, calibrated to hardware
  - Maintain truth association tables
- Storage
  - Arrays of digitized hits
  - Truth association table

Reconstruction
- Process
  - See the last few talks
  - Tracking, clustering, jet finding
- Storage
  - Reconstructed objects
  - See the last three talks
  - Truth association table

---

<table>
<thead>
<tr>
<th>100 ms / ev (HIJING)</th>
<th>1500 s / ev</th>
<th>500 s / ev</th>
<th>See reco. slide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7 MB / ev * 29%</td>
<td>210 MB / ev * 49%</td>
<td>Not saved</td>
<td>See reco. slide</td>
</tr>
</tbody>
</table>

- Above table: per-event resource for **0-30%** Au+Au event in full sPHENIX. (Expect order of **30M** Au+Au event sim total)
- Rare signal probes simulated in the embedding mode (fast 100s/ev, regularly carried out 0.1-1M sample studies in the past)
Reconstruction + Analysis Flow

Buffer boxes

135Gb/s

Calibration + Q/A

Conditions DB

Online monitoring

HPSS disk cache
Raw

Tape

HPSS disk cache
DST

DST Disk cache

Size reduction +
calibration pass

Analysis
Taxi

Data

1st pass TPC
tracking

2nd pass TPC
tracking +
reconstruction

1st pass TPC
tracking +
reconstruction

Data
Tracking reconstruction

- Tracking is the most challenging part of software
- Recent: complete rewrite of tracking infrastructure and data containers
  - Modulization, significantly reduce of CPU/memory cost
Current full track reconstruction performance:
- 9sec / event, 70% tracking efficiency

Target mid-2020 to achieve the projected reconstruction performance:
- 5 sec / event, > 90% tracking efficiency
- Realistic target based on ACTS projections
Looking forward
Coming simulation campaigns

- Building substantial centrally produced sample for each physics topical groups
- Need manpower for verification and analysis

<table>
<thead>
<tr>
<th>Year</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
<th>2023</th>
<th>2024</th>
<th>2025</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarter</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>sPHENIX Exp.</td>
<td>Construction</td>
<td>AA</td>
<td>pp/A</td>
<td>AA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p+p events [B]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A+A events [M]</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>p+A events [B]</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>p+p embed [B]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

1st campaign 2nd campaign sPHENIX Installation Year-1 ana. Year-2 ana. Year-3 ana.
**TODOs, milestones**

- **2019 (Q4)**
  - Demonstrate the ability to read time-aligned data from multiple systems
  - Deploy 2D & 3D vertex reconstruction

- **2020**
  - Demonstrate the ability to build events offline
  - First simulation campaign (realistic TPC drift & ExB effects)
  - Modeling and Simulation of TPC space-charge distortions
  - Implement ACTS tracking, optimize tracking for 5-sec/ev
  - Detailed Calorimeter simulation validation
  - ADC time-series signal fitting optimization
  - Develop calibration strategy, workflow, simulation tools
  - Define post-DST data formats
  - Implement PANDA/RUCIO for distributed job submission and data management
  - Implement Multi-Threading
  - Demonstrate ability to run on OSG

- **2021**
  - Demonstrate the ability to read time-aligned data from multiple systems
  - Second simulation campaign (with optimized tracker)
  - Develop/Adopt particle flow algorithms and Optimize Calorimeter Clustering
  - Database integration of alignment, detector parameters, and calibrations
  - Deploy space-charge correction framework
  - Mock data challenge(s) from simulated raw data thru calibration and reconstruction

- **2022**
  - Full chain test and QA for all subsystems
  - Cosmic running and analysis
  - Readiness testing and commissioning
How to get involved

- Discussion group:
  - sPHENIX simulation meeting: [https://indico.bnl.gov/categoryDisplay.py?categId=88](https://indico.bnl.gov/categoryDisplay.py?categId=88)
    - Pre-announced discussion topic
    - Review of recent core-software changes
  - Software and repository email list:

- Documentation
    - Always good to start with [day-1 checklist](https://wiki.bnl.gov/sPHENIX/index.php/Software)
    - Please search your email for read password/open writable account registration with ITD

- Resource
  - RHIC computing facility
    - Works with both PHENIX and STAR (in testing) existing RCF account
    - New: also support sPHENIX-dedicated accounts
    - 10 TB per user base disk associated with sPHENIX group tag: Register with [this form](https://indico.bnl.gov/categoryDisplay.py?categId=88).
  - sPHENIX code repository: [https://github.com/sPHENIX-Collaboration](https://github.com/sPHENIX-Collaboration)
Extra information
Radiation dose at sPHENIX

- Many detector parts are sensitive to radiation background
  - Ionization radiation (Radiation damage, flip bits, ...)
  - MeV neutron damage (Si sensor lattice damage, SiPM noise, ...)
  - Dominated by AuAu dose. Expect lower but comparison dose from pp and pA systems

- Strategy
  - Geant4 physics list QGSP_BERT_HP and FTFP_BERT_HP
  - Reproduce the PHENIX Run14 AuAu Radmon data
    - Same collision system, and closest accelerator configuration
    - Radiation sensor
  - Use the same procedure to predict sPHENIX radiation map

- Analysis interface to G4ScoringManager: [https://github.com/sPHENIX-Collaboration/coresoftware/pull/465](https://github.com/sPHENIX-Collaboration/coresoftware/pull/465)

sPH-TRG-2018-001: ~20% pp collision in |z|<10cm

Table 2: Summary of integrated samples summed for the entire five-year scenario.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Au+Au</td>
<td>200</td>
<td>35 nb(^{-1}) (239 billion)</td>
<td>80 nb(^{-1}) (550 billion)</td>
<td>214 nb(^{-1}) (1.5 trillion)</td>
</tr>
<tr>
<td>p+p</td>
<td>200</td>
<td>—</td>
<td>197 pb(^{-1}) (8.3 trillion)</td>
<td>1.0 fb(^{-1}) (44 trillion)</td>
</tr>
<tr>
<td>p+Au</td>
<td>200</td>
<td>—</td>
<td>0.33 pb(^{-1}) (0.6 trillion)</td>
<td>1.46 pb(^{-1}) (2.6 trillion)</td>
</tr>
</tbody>
</table>
PHENIX experiment in sPHENIX simulation

Calibration – PHENIX data
Calibration – PHENIX data

PHENIX data, from Eric Mannel

- Channel 1 (r=3.5cm)
- Channel 2 (r=8.5cm)
- Channel 0 (r=16.2cm)

Run 16

Run 15

Run 14

Run 13

1 MeV Equivalent Fluence x 10^6 n/cm²

- p + p 510 GeV
- Au+Au 200 GeV
- p + p/A 200 GeV

Date

01/15/13 02/25/13 03/14/13 04/23/13 05/11/13 06/21/13 07/31/13

PHENIX experiment in sPHENIX simulation

- RadFET TID [Gy]
- RADMON carrier board
- pin diode \( \Phi_{eq} [n_{eq}/cm^2] \)
Calibration – PHENIX data

Event display

Comparison to sensor data

Reasonable agreement in x2 ranges
Physics list dependence is small

QGSP_BERT_HP

FTFP_BERT_HP
(Few % higher dose)
Next, apply to full sPHENIX detector
Radiation dose

sPHENIX Simulation, FTFP_BERT_HP, Au+Au √s_{NN}=200 GeV, sHIJING 0-20fm
Total energy deposition [MeV] for 5-year run plan (1.5 Trillion Collisions)

Radiation dose [rad] for 5-year run plan (1.5 Trillion Collisions)
Fluence

**sPHENIX** Simulation, FTFP_BERT_HP, Au+Au $\sqrt{s_{NN}}=200$ GeV, sHijing 0-20fm
Min-1-MeV Charged particle fluence [$N_p$/cm$^2$] for 5-year run plan (1.5 Trillion Collisions)

**sPHENIX** Simulation, FTFP_BERT_HP, Au+Au $\sqrt{s_{NN}}=200$ GeV, sHijing 0-20fm
Min-100-keV Neutron fluence [$n$/cm$^2$] for 5-year run plan (1.5 Trillion Collisions)
More quantitatively...

- Averaged in $|z| < 100$ cm
- Plot against $R$
- Dose peaked at 10kRad
- MeV-Neutron $\sim 10^{10}$ n/cm$^2$ for inner detectors
More quantitatively ...

- Averaged in R~5 cm
- Plot against z
- Dose peaked at 10kRad
- MeV-Neutron \( \sim 10^{10} - 10^{11} \) n/cm\(^2\) for inner detectors
Radiation map

**sPHENIX-EIC** Simulation, Collision only, e+p, 20+250 GeV/c, eRHIC Pythia6

Total energy deposition [MeV] for 10 fb^{-1}, collision-originated fluence only

**sPHENIX-EIC** Simulation, Collision only, e+p, 20+250 GeV/c, eRHIC Pythia6

Min-1-MeV Charged particle fluence [N./cm²] for 10 fb^{-1}, collision-originated fluence only

**sPHENIX-EIC** Simulation, Collision only, e+p, 20+250 GeV/c, eRHIC Pythia6

Radiation dose [rad] for 10 fb^{-1}, collision-originated fluence only

**sPHENIX-EIC** Simulation, Collision only, e+p, 20+250 GeV/c, eRHIC Pythia6

Min-100-keV Neutron fluence [n/cm²] for 10 fb^{-1}, collision-originated fluence only
KPP simulation towards CD-2

- Since KPP describe the hardware we would be contracted to build for sPHENIX, we may need to incorporate some version of KPPs in default simulation.

- Suggest a version of KPP (sKPP?) in next default setup, e.g.
  - 1% dead-area in calorimeter
  - 3% tower-by-tower mis-calibration
  - Until we can demonstrate calibration capability to do better
As we have various simulation configuration, code-tagging and plot labeling become important.

In KPP simulations, introduced a tagging scheme:
- Macros are tagged in branch
- Same tag is applicable to plots
- Example tags:
  - CD-1 reference Objective KPP
  - CD-1 reference Threshold KPP
Physics List

- Geant4 provides factory packaged “Physics List”, which gives sets of parametrization of particle interaction with matters
- Tuned for LHC energy ranges
- Strong effects for e/π and jet energy scale
- Need new tune for sPHENIX.
- Such a list will be applicable to EIC too, connection with EIC software consortium

EMCal – π rejection VS cut

HCal - e/π ration VS energy

arXiv 1704.01461

Jin Huang
2nd sPHENIX Asia Meeting
Continuous Integration

- We use pull requests (PR) gate keep core-software (398 PR merged as of today)
- Continuous Integration tools well applied in software industry can help us automating the checking progress
- Pilot program of sPHENIX-Jenkins:
  - Now: automatically perform CPPCheck on commits to sPHENIX core-software. Report result to GitHub
  - Can be further developed to run automatic QA, and build based on pull requests and report problem automatically prior to merge

sPHENIX-Jenkins
Evaluation site
RCF support
Standardized QA

- Standardized calorimeter QA introduced to check multiple aspects in machinery of event generation, calorimeter simulation, digitization and reconstruction
- Now only used in manual check of pull requests
- Nice to be integrated to Jenkins automatic pull request checks
- Nice to have a standardized tracking QA too (an older module developed by Mike M.)

Simulation and evaluation of SiPM noise and zero-suppression 20 GeV energy in EMCal just from SiPM noise and zero suppression

Fast simulation:
General packages e.g. DELPHES?

Continuous Integration for sPHENIX software
Testing a Jenkins for automatic software verification

Containers, virtual machines:
Distribute sPHENIX software to your laptop
Based on PHENIX software framework, a.k.a. Fun4All
- Modular design, C++ based, 1 PB data / week PHENIX data analysis
- Naturally supports pause analysis at any reconstruction stage, embedding, Naturally supports event filtering, Event mixing via analysis code
- Built-in Geant4 support Common macro run the simulation and standard analysis chain
- Easy access for user modules in analysis:
Event generator: event pile up

- Recent addition, simulation of pile up
- More details: Sanghoon Lim (Fri, afternoon)

Simulated TPC data stream

Collisions: +/- 35 µs  TPC: +/- 35 µs  MVTX: +/- 2 µs  INTT: [-20 ns, +80 ns]
Calorimeters: in Simulation

- **EM calorimeter** (EMCal): \(18 \lambda_0\) W-SPACAL
- **Inner hadron calorimeter** (Inner HCal): \(1 \lambda_0\) SS-Scint. sampling
- **sPHENIX coil and cryostat.** (Magnet): \(1.4 \lambda_0\) Coil & Cryostat
- **Outer hadron calorimeter** (Outer HCal): \(4 \lambda_0\) SS-Scint. sampling

Beam view of **full calorimeters**
Calorimeters: EMCal details as an example

10 GeV, e+

SPACAL Tower
w/ fibers displayed

EMCal Half Sector
(fibers simulated but hidden from display)

8 GeV photons in **full EMCal**
Side view/beam view

Jin Huang

2nd sPHENIX Asia Meeting
Calorimeters: EMCal verification

EMCal energy resolution for EM shower in tower center (Joe Obsborn)

Test beams @ FermiLab

EMCal rejection for $\pi^-$

-arXiv:1704.01461-

$sPHENIX$ Preliminary

4x4 cm region within one SPACAL block

- $10^\circ$ Incident Angle, Hodoscope Corrected
- Fit, $\Delta E / E = 2\% (\delta p/p) \oplus 1.6\% \oplus 13\%/\sqrt{E}$

- $10^\circ$ Incident Angle, Position Corrected
- Fit, $\Delta E / E = 2\% (\delta p/p) \oplus 1.3\% \oplus 13.6%/\sqrt{E}$
- GEANT4 Simulation
- $\Delta E / E = 2\% (\delta p/p) \oplus 2.5\% \oplus 12.7%/\sqrt{E}$

$sPHENIX$ beam test data
Calorimeters : HCal verification

HCal shower line shape (Data/Sim)

**π⁺**

<table>
<thead>
<tr>
<th>Energy</th>
<th>DATA</th>
<th>SIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 GeV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 GeV</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**π⁻**

<table>
<thead>
<tr>
<th>Energy</th>
<th>DATA</th>
<th>SIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 GeV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 GeV</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HCal resolution (Data/Sim)

\[
\text{Resolution} (\sigma_E/E) = \frac{\Delta E}{E} = 2(\delta p/p) \oplus 12.4\% \oplus 88.2\%/\sqrt{E}
\]

\[
\text{Resolution} (\sigma_E/E) = \frac{\Delta E}{E} = 2(\delta p/p) \oplus 10.8\% \oplus 87.5\%/\sqrt{E}
\]

\[
\text{Resolution} (\sigma_E/E) = \frac{\Delta E}{E} = 12.3\% \oplus 83.5\%/\sqrt{E}
\]

\[
\text{Resolution} (\sigma_E/E) = \frac{\Delta E}{E} = 13.0\% \oplus 80.0\%/\sqrt{E}
\]
Calorimeter study highlights

- Discussion in topical group updates (Fri PM, Sat AM)

### Photon energy resolution

- **Photons**: $14\%/\sqrt{E} \oplus 4.6\%$

### Jet efficiency and resolution in central Au+Au

- Reconstruction efficiency ($p_T > 10$ GeV)

### EMCal electron-ID

- **sPHENIX simulation**
  - $\pi^-$ embedded in HIJING Au+Au 0-4.4 fm
  - Hadronic rejection in EMCal at 90% electron efficiency cut

---

2nd sPHENIX Asia Meeting 44
Beyond sPHENIX baseline

Main discussion: Christin, Nils (Sat AM)