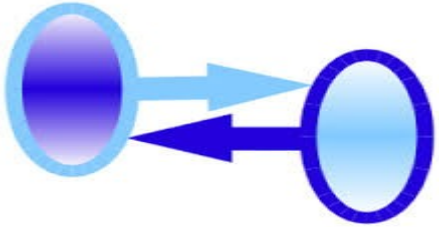


HepSim Monte Carlo repository and integration of its software with key4hep

S.Chekanov (HEP/ANL)

FCC-ee software meeting, CERN

May 30, 2023

What is HepSim?

https://atlaswww.hep.anl.gov/hepsim/

Repository with MC files & software for

- Physics (discovery potential, future precision measurements, etc.)
- Exploration of general aspects of detectors & interactions with material using fast and full Geant4 simulations

Experiment neutral: Can be used for any current & future experiment & phenomenological paper

Organize MC samples for download using collision energy, collision type & physics topics

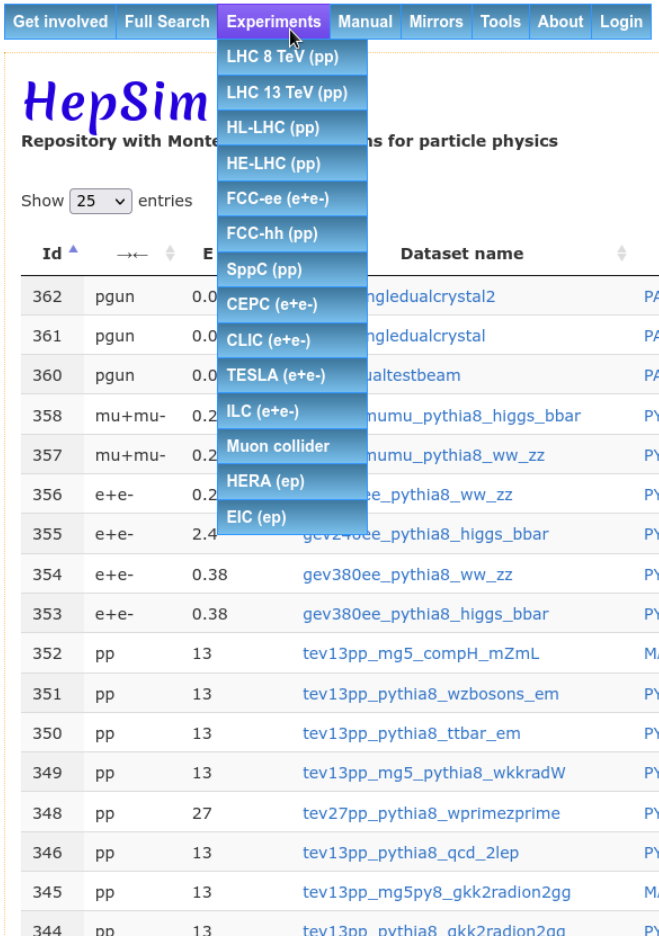
Event samples assigned to Digital Object Identifiers (DOI) in the form xx.yyyy/zzzzz (see osti.gov link)

Show all

$p \rightarrow\leftarrow p$
- 8 TeV
- 13 TeV
- 14 TeV
- 27 TeV
- 33 TeV
- 100 TeV

$e^+ \rightarrow\leftarrow e^-$
- 240 GeV
- 250 GeV
- 380 GeV
- 500 GeV
- 1 TeV
- 3 TeV

$\mu^+ \rightarrow\leftarrow \mu^-$
- 250 GeV
- 1 TeV
- 5 TeV
- 10 TeV
- 20 TeV
- 40 TeV



What is HepSim?

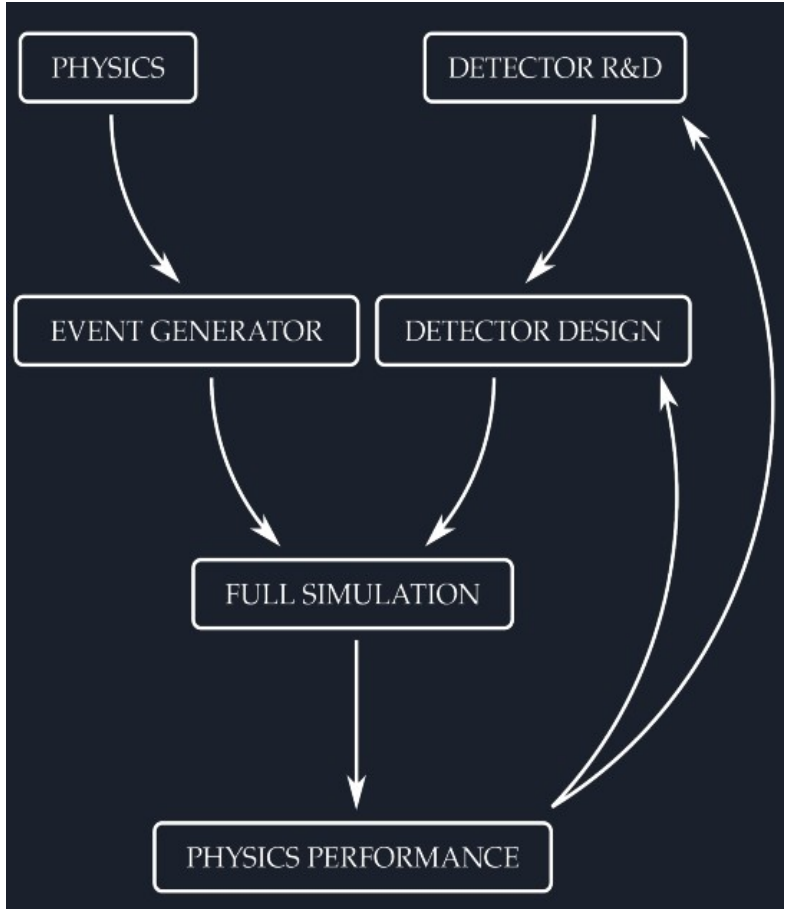
https://atlaswww.hep.anl.gov/hepsim/

- Consists of a web interface, distributed web storage, command-line tools, event browser, containerized software (docker/singularity image)
- Began at Snowmass DPF 2013 (Top/Higgs, see URL) and evolved to →
- Since 2015 used for physics and detector studies for future experiments (HL-LHC, HE-LHC, FCC, CLIC, CEPC, EIC, etc.) and several ATLAS/LHC papers
- 360 truth-level samples:
  - ~20% converted to Delphes ROOT files
  - ~5% with Geant4 simulation/reconstruction

Why HepSim?

https://atlaswww.hep.anl.gov/hepsim/

- **Open access**
  - No authentication for use of event files
- **Preservation of MC data, settings and detectors**
- **Mitigate reproducibility problem in publications**
  - Cite Monte Carlo data using DOI identifies
- **Cache for iterative experiment design process**
- **Analysis using platform-independent software on Linux/Mac/Windows (+ URL data streaming)**



Credits to W.Armstrong (Physics/ANL)

How it works

Leveraging large-scale computing

Event Generators

- PYTHIA6
- PYTHIA8
- HERWIG++
- Madgraph5
- MCFM
- JetPhox
- FPMC
- NLOjet++
- LEPTO/Ariadne
- . . . . . .

EVGEN

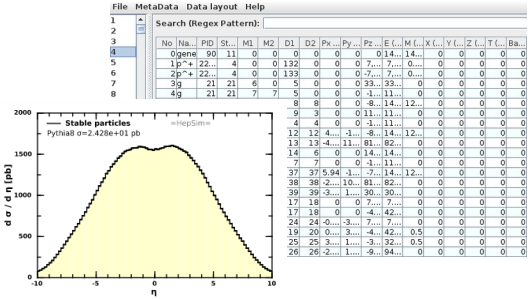
fast → Delphes fast simulation

full → Geant4 full simulation and reconstruction software

- **Integrated with Chicago-area computing:**
  - CPU: OSG Connect, UChicago, LCRC, ALCF
  - File storage: PETREL Data Management and Sharing Pilot
- **HepSim FrontEnd mirrored at JLab, NERSC, CERN**
- **Easy to link self-managed external file storage**

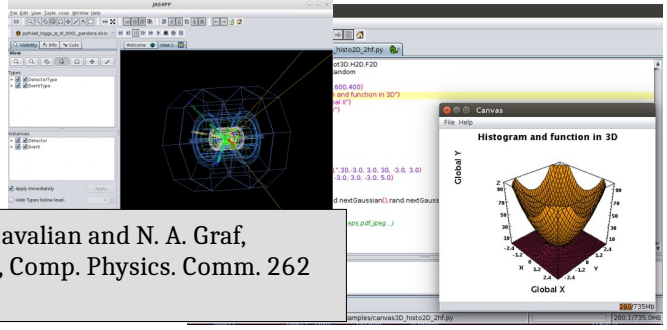
Software for end user

https://atlaswww.hep.anl.gov/hepsim/



Two OS-independent packages

- **(1) hs-toolkit (30 MB)**
  - Discovery and download Monte Carlo files from remote sites
  - Event browser for truth-level events
  - Processing truth-level files and full data-analysis with Lorentz vectors, jet algorithms, etc.
  - Histograms in 2D, 3D, X-Y plotting etc., → vector-graphics images
  - Includes full Python 2.7 API via Jython
- **(2) JAS4PP-** ***Java Analysis Studio for Particle Physics*** (https://atlaswww.hep.anl.gov/asc/jas4pp/) **(110 MB)**
  - hs-toolkit included
  - User friendly IDE
  - Analysis of detector-level files in LCIO file format
  - ROOT I/O + many physics libraries
  - Full experiment-independent event display



S.V. Chekanov, G. Gavalian and N. A. Graf, arXiv:2011.05329, Comp. Physics. Comm. 262 (2021) 107857

Event samples for e+e- studies

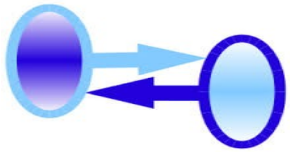
https://atlaswww.hep.anl.gov/hepsim/

- 250 GeV: 25 events samples (truth level), ~ 30% after full & Delphes simulation
- 380 GeV: 24 event samples (truth level), ~ 20% after full & Delphes studies
- Also some 240 GeV, 500 GeV, 1 TeV samples are available
- See full list:
  - https://atlaswww.hep.anl.gov/hepsim/index.php?c=epem&e=0&t=all

Note on iteration:

- Any external event samples can easily be integrated in HepSim
- Put an event sample on Apache web server and index it as explained in https://atlaswww.hep.anl.gov/hepsim/doc/doku.php?id=hepsim:dev_hepsim
- Low maintenance!

Tutorial using e+e- samples



- How to search and download a file with 10,000 e+e-: H→ bbar process
- What is inside the download file?
- How to fill a few histogram from the truth-level file

Require Java

- How to create Pythia8 events in the HepSim format
- Conversions to ROOT, STDHEP, LCIO, HEPMC formats
- How to create fast Delphes simulations

Require Linux / ROOT / GCC compilers /
/cvmfs/sw.hsf.org/key4hep/

URL link with this tutorial:
https://atlaswww.hep.anl.gov/hepsim/doc/doku.php?id=fcs:fccee:tutorial

FCC-ee HepSim tutorial - 1

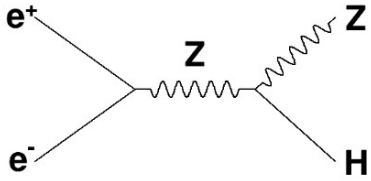
This part of the tutorial does not use any C++ specific libraries and can be done on any computers with Java installed. Check java:

```
java -version
```

Typically, it tells “openjdk version “1.8.0_352” or higher Java version.

```
wget https://atlaswww.hep.anl.gov/hepsim/soft/hs-toolkit.tgz -O - | tar -xz;
source hs-toolkit/setup.sh
```

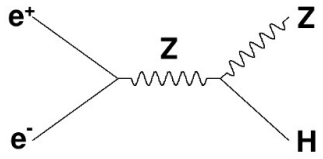
Look at few events: Z→ Z H, where Z→ nunu, and H decays to bbar.



The CM energy is 250 GeV.

The sample is described in https://atlaswww.hep.anl.gov/hepsim/info.php?item=353

FCC-ee HepSim tutorial – 2 part



First, print all files with Higgs processes: `# hs-find higgs`
Then grab the file with H to bbar at e+e-: `# hs-ls gev250ee_pythia8_zhiggs_nunubbar`
Download 10 files (in 2 threads): `# hs-get gev250ee_pythia8_zhiggs_nunubbar data 2 10`

We should have 10 files in the directory “data”. Take a look at a single file. We what to check how many events in the file etc. `# hs-info data/gev250ee_zh_nunubb_001.promc`

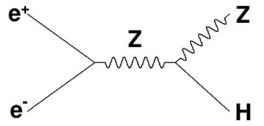
Do you want to print 1st event? Do this: `# hs-info data/gev250ee_zh_nunubb_001.promc 1`

Want to examine the log file? Do this: `# hs-log data/gev250ee_zh_nunubb_001.promc`

Let's study each event in the GUI mode (needs X-session!). Start this GUI and click each event number using the left panel: `# hs-view data/gev250ee_zh_nunubb_001.promc`

Run over this file using Python syntax and make a few simple distributions:
https://atlaswww.hep.anl.gov/hepsim/doc/doku.php?id=fcs:fccee:tutorial#validation

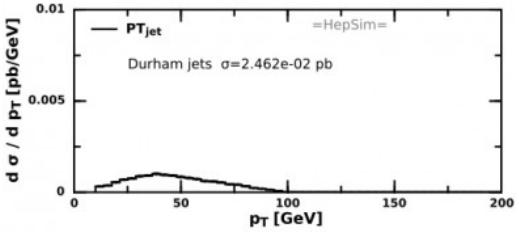
Making a first plot (truth-level events)

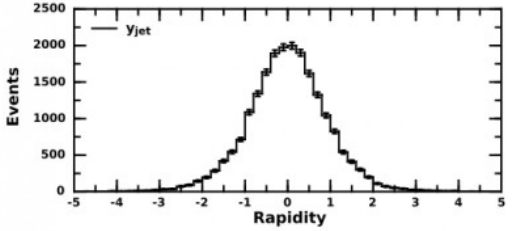


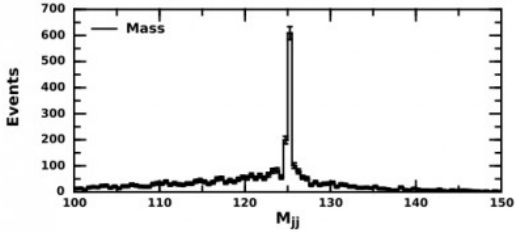
https://atlaswww.hep.anl.gov/hepsim/doc/doku.php?id=fcs:fccee:tutorial#validation

The example is based on ~30 MB *hs-toolkit* package and a script written in in Python (Jython/Java backend)

- Run over local PROMC files or streaming events over the network by specifying dataset name
- Fill histograms using physics libraries with jet algorithms, event shapes, Lorenz vectors etc..
- Export images to vector format

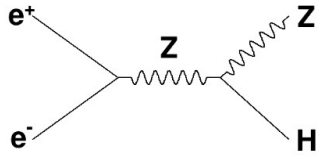






HepSim includes about ~100 scripts for data validation for truth-level samples:
https://atlaswww.hep.anl.gov/hepsim/macrolist.php

Delphes fast simulations



Fast simulations: Use the key4hep setup with gcc11 + ROOT:

```
# source /cvmfs/sw.hsf.org/key4hep/setup.sh

# wget http://atlaswww.hep.anl.gov/asc/promc/download/current.php -O ProMC.tgz
# tar -zvxf ProMC.tgz
# cd ProMC
# ./build.sh                    # build all source files
# ./install.sh lib              # install into the "lib" directory
# source lib/promc/setup.sh     # make it available
```

Install Delphes:

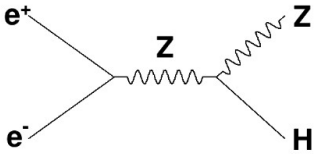
```
# wget http://cp3.irmp.ucl.ac.be/downloads/Delphes-3.5.0.tar.gz
# tar -zxf Delphes-3.5.0.tar.gz
# cd Delphes-3.5.0;  make
```

Run:

```
# ./DelphesProMC  ./cards/delphes_card_CircularEE.tcl \
  ../data/gev250ee_zh_nunubb_001.root  ../data/gev250ee_zh_nunubb_001.promc
```

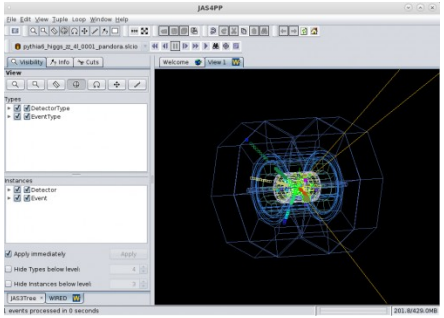
Using Jas4pp IDE

S.V. Chekanov, G. Gavalian and N. A. Graf, arXiv:2011.05329, Comp. Physics. Comm. 262 (2021) 107857



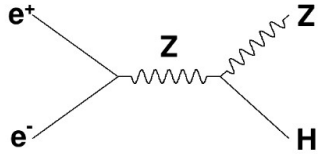
User-oriented desktop application. Download installer: https://atlaswww.hep.anl.gov/asc/jas4pp/ and install as usual desktop application. Click the icon and launch it

- Can search & download HepSim truth-level / reco files
- Supports Python, Groovy (faster!) syntax and Java
- Analyze data using the **ProMC (truth-level)** or **LCIO (detector level)** formats (or any native Java/Python IO supported too)
- ROOT-like graphics and large number of physics libraries



+ detector-independent event display (Wired3/Jas3)

Streaming e+e- data over the network



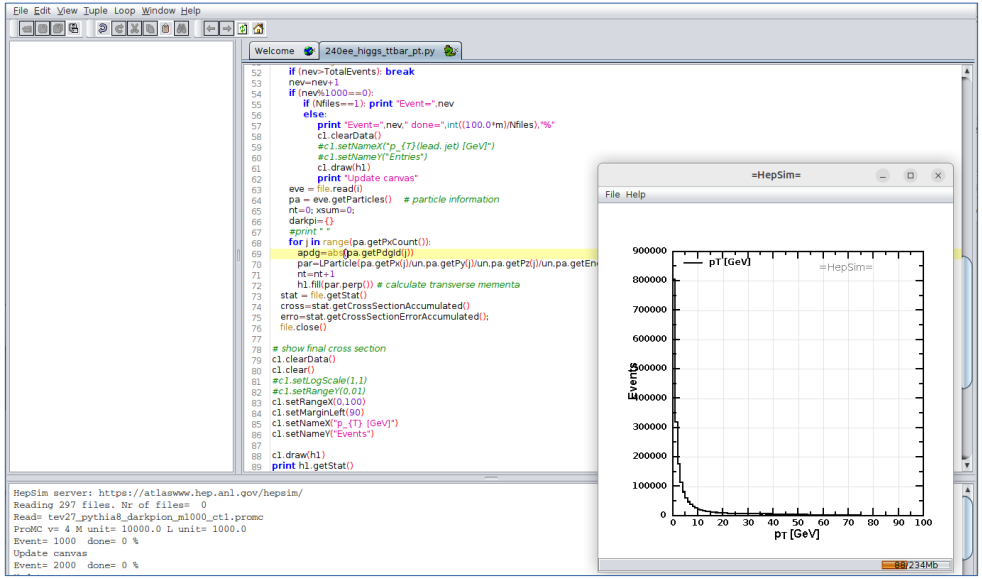
One can also specify HepSim data sample inside a script (without downloading files)

Example for H→ bbar at 240 GeV e+e-. Copy Python code and save as “example.py”
https://atlaswww.hep.anl.gov/hepsim/doc/doku.php?id=fcs:fccee:tutorial#data_streaming

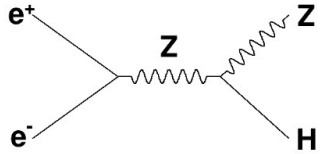
Open analysis macro inside Jas4pp and click “Run”

The e+e- data will be streamed, pT-distribution plot will be updated in real time.
→ Used for validation tasks

Note: Data streaming is also supported using minimalistic hs-toolkit package



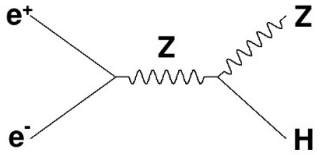
Analyzing LCIO files (detector-level)



LCIO truth and detector-level files (See the talk LCIO turns 20: N.Graf CHEP2023 [PDF talk]) can also be analyzed in Python scripts using Java backend (or Java)

- LCIO files can be download from HepSim using “detector tags”.
- Collection of Python scripts for LCIO input files listed in:
  - https://atlaswww.hep.anl.gov/hepsim/macrolist_lcio.php
- Many *.py and *.java scripts located in the directory “examples” of *Jas4pp* installation
  - Shows how to analyze tracks, hits, calorimeter clusters using LCIO files
- miniDST files from ILD detector (also SID and CLIC) can be used too:
  - Full analysis e+e- example running in jas4pp: https://atlaswww.hep.anl.gov/asc/wikidoc/doku.php?id=asc:jas4pp#reading_minidst_files

Z peak created using miniDST LCIO input



File Edit View Tuple Loop Window Help

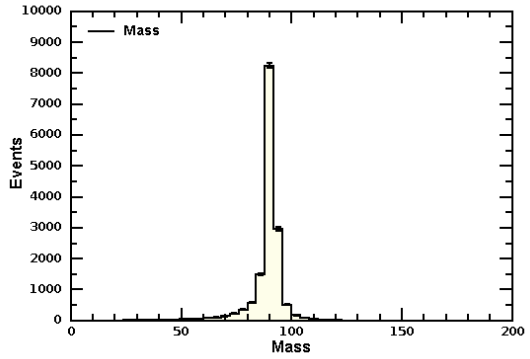
test.py × Welcome

```
1  from hep.lcio.implementation.io import LCFactory
2  from hephysics.particle import LParticle
3  from jhplot import  HPlot,H1D
4  from math import sqrt
5  import glob
6
7  files=["rv01-16-p10_250.sv01-14-01-p00.mILD_o1_v05.E250-TDR_ws.I106479.Pe2e2h.eL.pR-00001-ILDminiDST.slcio"]
8  factory = LCFactory.getInstance()
9
10 h1= H1D("Mass",50,0,200)   # create a histogram
11 h1.setFill(True)
12 nEvent=0
13 for f in files:
14   print "Open file=",f
15   reader = factory.createLCReader()
16   reader.open(f)
17   while(1):
18      evt=reader.readNextEvent()
19      if (evt == None): break
20      nEvent=nEvent+1
21      # print " file event: ",evt.getEventNumber(), " run=",evt.getRunNumber()
22      if (nEvent%100==0): print "# Event: ",nEvent
23      strVec = evt.getCollectionNames()
24      if nEvent == 1:
25             for col in  strVec:
26                          print col
27      col = evt.getCollection("IsolatedMuons")
28      muons=[]
29      for i in range( col.getNumberOfElements() ):
30           track=col.getElementAt(i)
31           mom=track.getMomentum()
```

```
# Event:  15200
# Event:  15300
# Event:  15400
# Event:  15500
# Event:  15600
# Event:  15700
# Event:  15800
# Event:  15900
# Event:  16000
# Event:  16100
# Event:  16200
# Event:  16300
# Event:  16400
# Event:  16500
# Event:  16600
# Event:  16700
```

=HepSim=

File Help



List with SLCIO miniDST files from the ILD project

Link to manual with Python code

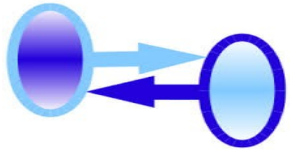
Ideas related to integration with FCC-ee (key4hep) software. Part I

- ProMC C++ library (https://atlaswww.hep.anl.gov/asc/promc/) can be put under key4hep. It allows reading HEPSIM files in C++ and develop the needed converters ProMC→ EDM (truth) and back, i.e. promc2edm and edm2promc
- Delphes should be recompiled after setting up $PROMC environmental variable. This variable is set automatically after running ProMC “setup.sh” scrip.
  - Compiling Delphes with $PROMC env variable creates “**DelphesProMC”** binary file
- Few other converters should be created:
  - LCIO→ EDM and EDM→LCIO
  - STDHEP→EDM, HEPMC2→EDM (to support legacy files from ILC)
- If EDM files are publicly stored (under Apache web server), HepSim Web interface can link such files and make them searchable & downloadable

Ideas related to integration with FCC-ee (key4hep) software. Part II

- The rest of HepSim software is Java-centric. Does not require compilation.
  To make it fully available in ke4hep, we need:
  - JDK18 (or above)
  - hs-toolkit (https://atlaswww.hep.anl.gov/hepsim/doc/doku.php?id=hepsim:hs-tools) used to discover files, validate files and process
- Jas4pp is optional (it is a desktop application). Requires X-window and thus it is easier to run it on a desktop

Thanks!



For more information, see the HepSim web manual and **hs-help** on the command line.

HepSim manual: https://atlaswww.hep.anl.gov/hepsim/doc/

HepSim contributors:
https://atlaswww.hep.anl.gov/hepsim/doc/doku.php?id=hepsim:contributions

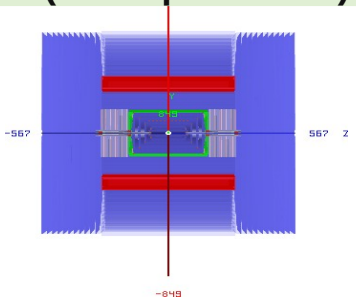
HepSim public results:
https://atlaswww.hep.anl.gov/hepsim/doc/doku.php?id=hepsim:public

~30 articles. Contributions to CEPC, CLIC, FCC-hh etc. conceptual design reports

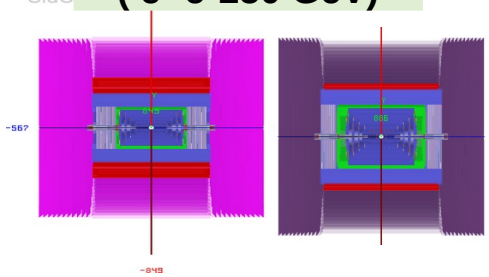
Backup

‘All-silicon’ design concepts supported in HepSim

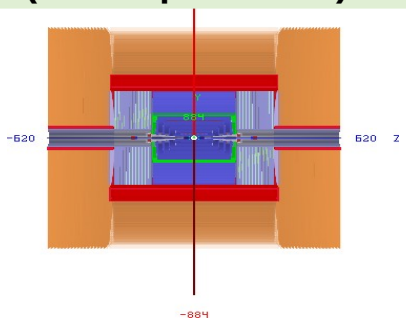
SiD (SiD LO3)
(e^+ e^- up to 1 TeV)



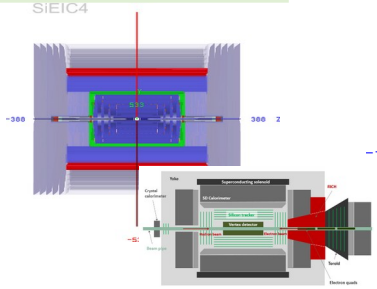
SiCPEC, SiDB
(e^+ e^- 250 GeV)



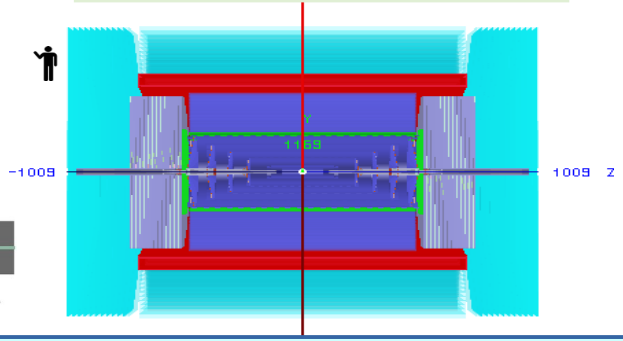
CLIC-SiD (CDR)
(e^+ e^- up to 3 TeV)



SiEIC, TopSide
(ep, 35-141 GeV)



SiFCC + 7 variations
(FCC-hh, pp 100 TeV)



Performance detectors:

- Physics reach studies using Geant4 simulations & full reconstruction
- Playground for various technologies and detector optimizations
- Fast turnover to modify detector & create events samples

Share similar design, but differ in sizes, calorimeter readouts etc

Interfaced with common Monte Carlo samples

13 TeV
14 TeV
27 TeV
33 TeV
100 TeV

$e^+ \rightarrow \leftarrow e^-$
250 GeV
380 GeV
500 GeV
1 TeV
3 TeV

$\mu^+ \rightarrow \leftarrow \mu^-$
1 TeV
5 TeV
10 TeV
20 TeV
40 TeV

$e^- \rightarrow \leftarrow p$
318 GeV
141 GeV
45 GeV
35 GeV

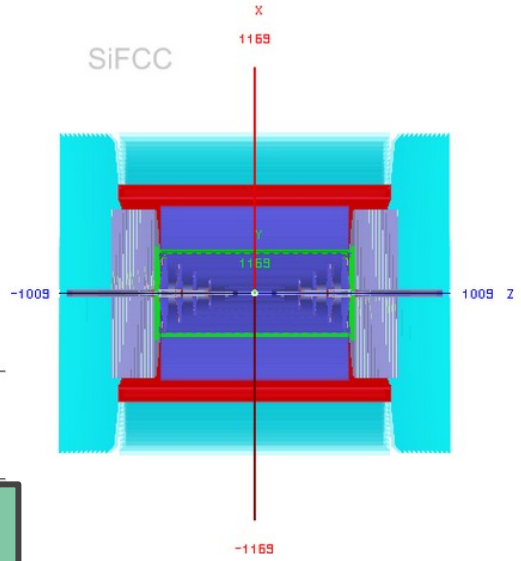
Misc.

Information about the "*sifcch7*" detector

Summary

Name:	*sifcch7*
Title:	*A silicon Detector for FCC-hh studies. Described in JINST 12 (2017) P06009 (arXiv:1612.07291)*
Author:	*S.Chekanov, A.Kotwal, J.Zuzelski, etc.*
Status:	*development*
Version:	*$Id: compact.xml,v3.0 2016/09/09 23:46:56 Sergei Chekanov Exp $*
Level:	Geant4 simulation and full event reconstruction
Summary:	view
3D View:	
GeoManager:	
Calibrations:	view
Tracking:	view
Last modified:	September 07, 201

Interactive 3D Visualization



Reconstruction tags

Tag lists: rfull009 | rfull015 | rfull016 | rfull017

Detector geometry files

HEPREP:	sifcch7.heprep
GDML:	sifcch7.gdml.gz
JSON:	sifcch7.json.gz
LCDD:	sifcch7.lcdd
Pandora:	sifcch7.pandora

Geometry in various formats

Download of complete detector

Download: sifcch7.zip

Comment

Main event file format ProMC and ProIO

http://atlaswww.hep.anl.gov/asc/promc/ and https://github.com/proio-org

- Archive-style self-described format to keep MC events - **ProMC**:
  - Event records, NLO,original logfiles, PDG tables etc.
- 30% smaller files than any existing formats after compression

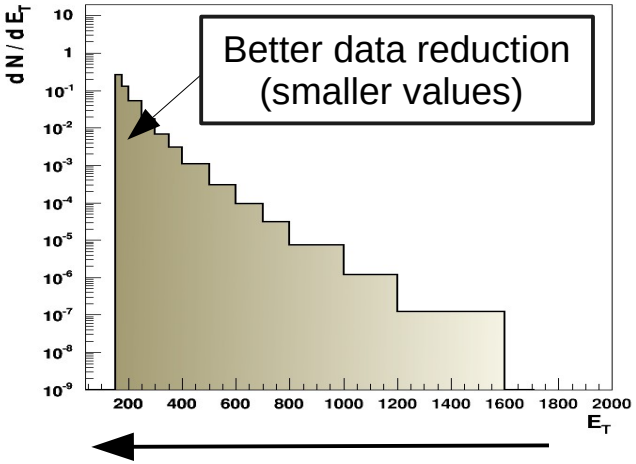
protobuf
Protocol Buffers - Google's data interchange format

Google's Protocol buffers

Number of used bytes depends on values.
Small values use small number of bytes

- Effective file size reduction for pile-up events where particles with small momenta use less bytes
- Separate events can be streamed over the Internet
- Other HepSim formats: **ROOT, LHE** and **LCIO** (full simulation)

8-bytes → varint



compression strength keeping precision of representation constant

ProMC: S.C., E.May, K. Strand, P. Van Gemmeren, Comp. Physics Comm. 185 (2014), 2629

NLO QCD calculations as “ntuples”

Theorists can use it too!

S.C. Adv. High Energy Physics, vol. 2015, 13609

- Stored several NLO QCD calculations (MCFM, JETPHOX, NLOjet++)
- Data structure is different compared to full parton-shower MC
  - “Particle record”: Usually 4-momenta of 3-4 particles per events
  - “Event record”: includes “weights” and deviations from central weights for different PDF eigenvector sets for calculations of PDF uncertainties

$$w_n = \left[1000 \times (1 - \frac{PDF(n)}{PDF(0)})\right]$$

N=1...51, for CT10 PDF

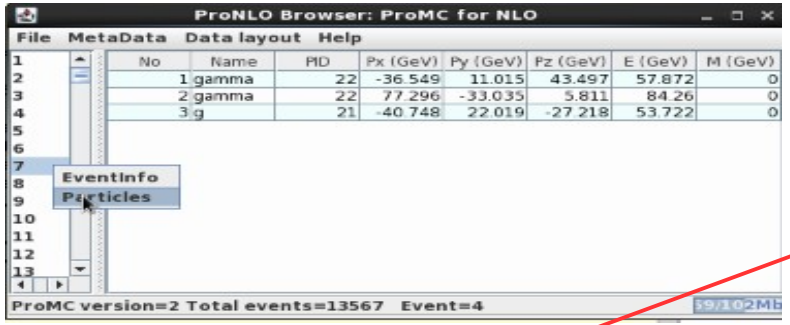
Weighted NLO events can be compactly stored using Google ProtocolBuffers:
→ *double precision “weights” → int64 varint (deviations) → 2 bytes per weight*
→ *Large deviations are stored using 4 or 8 bytes (rarely)*

NLO QCD calculations as “ntuples” for HEP experiments

MCFM prediction for H(→ γγ)+jet (pp 100 TeV)
“higgsjet_gamgam_mcfm” sample

Some NLO samples using MCFM
have been created on Mira supercomputer
(BlueGene/Q)

4-momenta of particles



Event weights

PDF variations for CT10 (51)