# Monte Carlo simulations for future collider studies

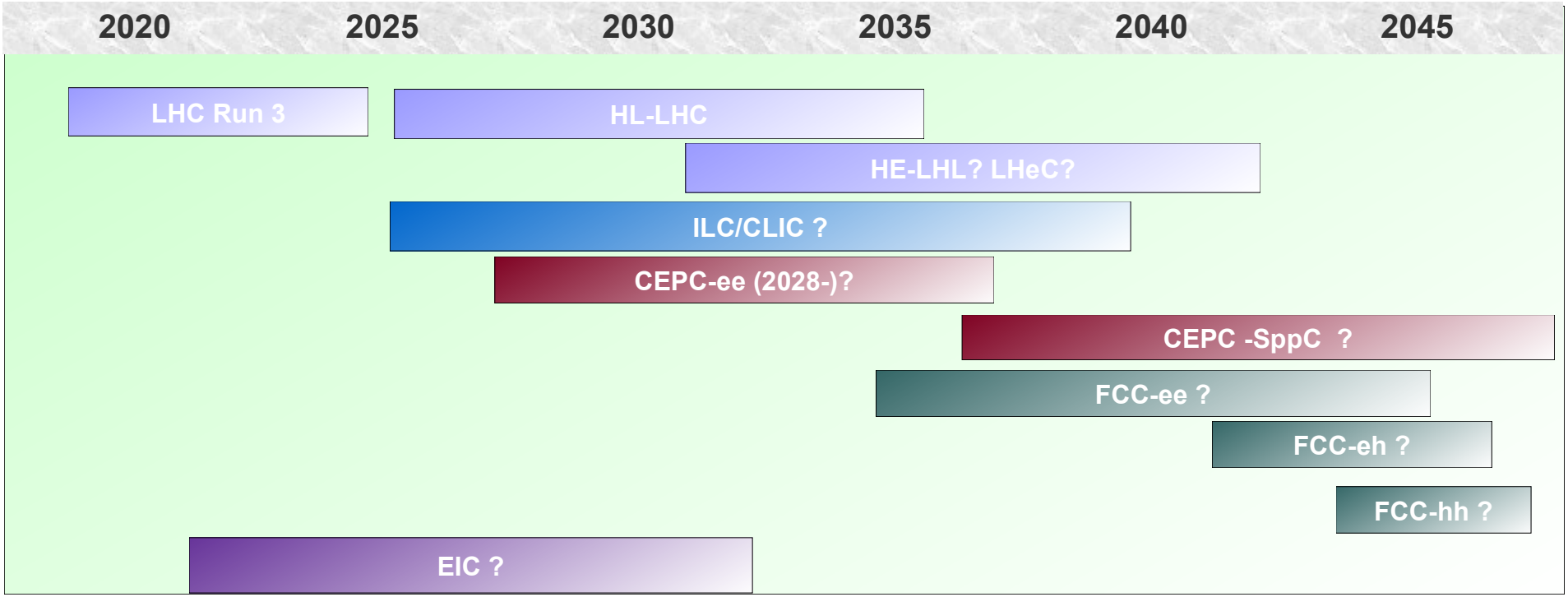
**S. Chekanov**
**+ contributors** (www link)

*HEP/ANL*

*ICHEP 2016. Chicago*

*Aug. 4, 2016*

# Timeline of particle collision experiments




In the next decade we will deal with explorations of physics reach, detector parameters and new technology options for post-LHC era

**Requires detailed simulation of physics processes and detector responses**

# Simulations for particle-collision experiments

**Detector geometry**

**Physics modeling**

- Known particle properties & established Standard Model (SM)
- Event generators for Standard Model and beyond (LO, NLO, NNLO, NLO matched to NLO)

computing
(HPC, grid)

**Performance studies (analysis)**

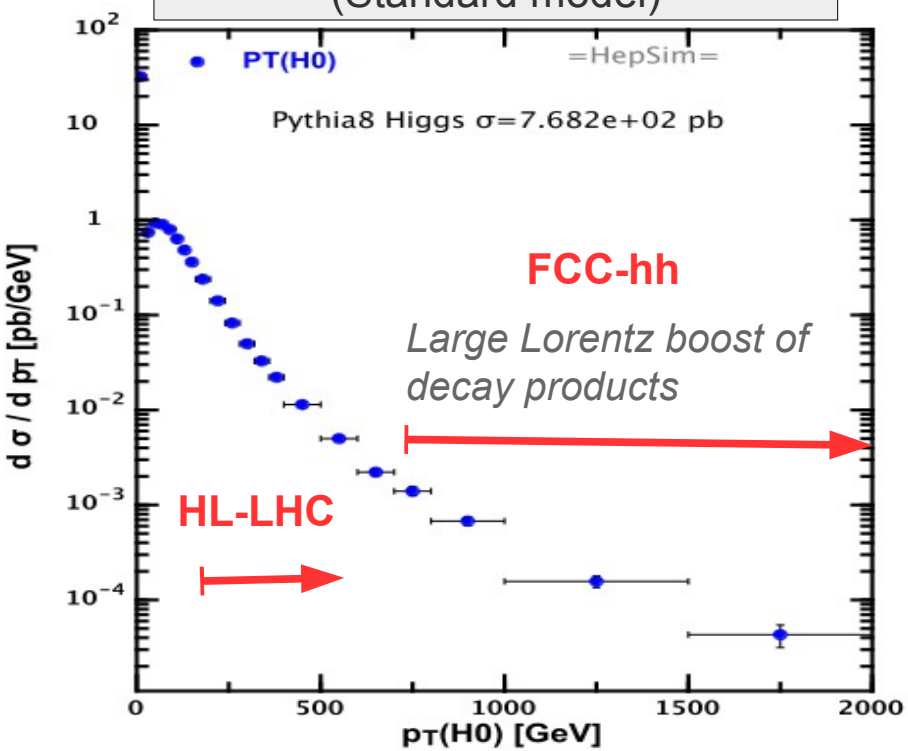
**Simulation of detector response**

- Interactions of particles with detector material
- Many parametrized cross sections (when exact theory is unknown)
- Simulation packages (Geant4,etc.)

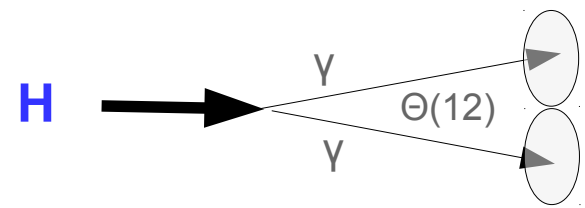
# Why do we need simulations? Higgs example

- 100 TeV collider will hunt for M~30 TeV particles decaying to Higgs/W/Z bosons
- Completely new kinematic regime → very challenging for detector designs
- The detector must be optimized to reconstruct Higgs with pT>1 TeV

100 TeV pp
Differential Higgs cross section
(Standard model)




**SM predictions:** ~100,000 Higgs / ab$^{-1}$ for pT>1 TeV

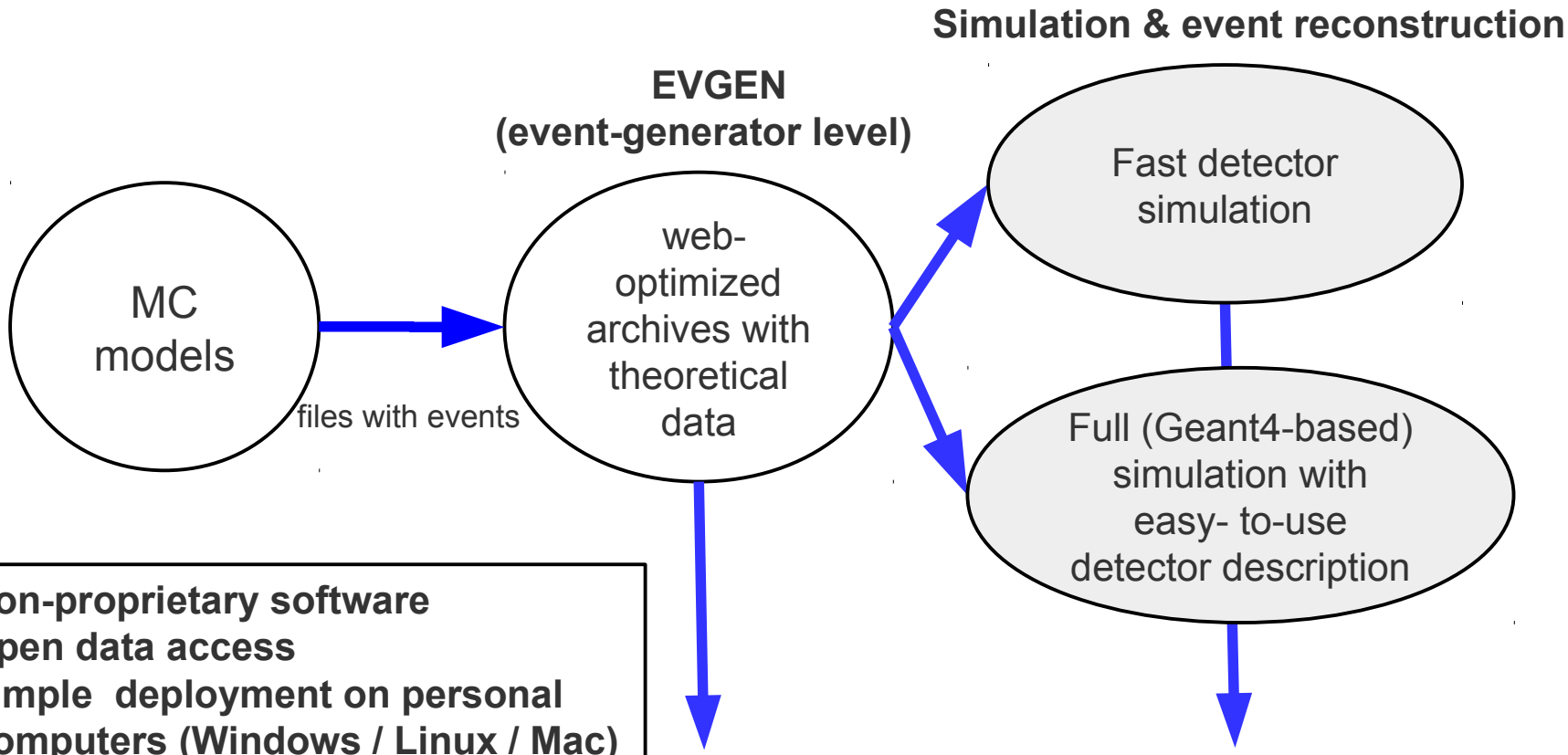



**Just kinematics:**
pT(H)>2 TeV → **~ 5 deg** between γ's
pT(H)>10 TeV → **~ 1 deg** between γ's

Instrumental challenges:
- identify 2 photons separated by 1 degree
- reject $\pi^0 \to \gamma\gamma$ background at the same time!
- similar problems for electron, b-jets decays

# Next step after Snowmass 2013:

## Public Repository with Fast and Full Monte Carlo simulations




→ **Non-proprietary software**
→ **Open data access**
→ **Simple deployment on personal computers (Windows / Linux / Mac)**

OPEN ACCESS **http://**

**Long-term availability & preservation**

# HepSim event simulations

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

NERSC, CERN mirrors

Get Involved | Full Search | Manual | About | Mirrors | Login

HEP.ANL.GOV

Show all

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

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

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

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

Misc.
- 1 particle
- 2 particles
- 1 jet

## HepSim

**Repository with Monte Carlo predictions for HEP experiments**

Feb.5, 2016: Single particles for ITK studies (ATLAS phase II upgrade) (link)
Feb.1, 2016: Z' with M=10,20,40 TeV decaying to qqbar, ttbar, WW for full simulations (link)
Jan.19, 2016: 10 TeV Z' using a full simulation with 40 and 64 HCAL layers (link)

Show 25 entries | Previous 1 2 3 4 5 ... 8 Next | Search:

| Id | →← | E [TeV] | Name | Generator | Process | Topic | Info | Link | Created |
|---|---|---|---|---|---|---|---|---|---|
| 1 | pp | 100 | | PYTHIA8 | Higgs production | Higgs | Info | URL | 2016/01/07 |
| 2 | pp | 100 | ng5 | MADGRAPH/HW6 | Higgs+ttbar (NLO+PS) | Higgs | Info | URL | 2015/11/13 |
| 5 | pp | 8 | | FPMC | Exclusive WW production | SM | Info | URL | 2015/03/23 |
| 6 | pp | 8 | p | HERWIG++ | Direct photons | SM | Info | URL | 2015/04/11 |
| 7 | pp | 100 | tev100_qcd_herwigpp_pt2700 | HERWIG++ | QCD dijets, pT>2700 GeV | SM | Info | URL | 2015/04/11 |
| 10 | pp | 100 | _pythia8 | PYTHIA8 | KKgluon to ttbar M=1-20 TeV | Exotic | Info | URL | 2015/03/23 |
| 11 | pp | 100 | tev100_qcd_pythia8_pt300 | PYTHIA8 | QCD dijets, pT>300 GeV | SM | Info | URL | 2015/04/10 |
| 12 | pp | 100 | t900 | PYTHIA8 | QCD dijets, pT>900 GeV | SM | Info | URL | 2015/10/03 |
| 13 | pp | 10 | t2700 | PYTHIA8 | QCD dijets, pT>2700 GeV | SM | Info | URL | 2016/01/07 |
| 14 | pp | 10 | t8000 | PYTHIA8 | QCD dijets, pT>8 TeV | SM | Info | URL | 2015/10/21 |
| 15 | pp | 10 | | MADGRAPH/HW6 | pp->ttbar at NLO | Top | Info | URL | 2015/11/13 |
| 16 | pp | 10 | mg5_lo | MADGRAPH/HW6 | pp->ttbar, pT>2500 GeV | Top | Info | URL | 2015/04/10 |

LHC run 1/2

HL-LHC

SPPC, FCC-hh

ILC, CEPC

samples for detector performance studies

Available: EVGEN files (LO,NLO, etc), fast simulations, full Geant4 simulations

# HepSim repository. How it works

large-scale computing resources

**Event Generators**

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

. . . . . .

**EVGEN**

fast → Delphes fast simulation (ROOT files)

full → SLIC (Geant4) full simulation and reconstruction software (LCIO files)

ProMC files on several public web servers

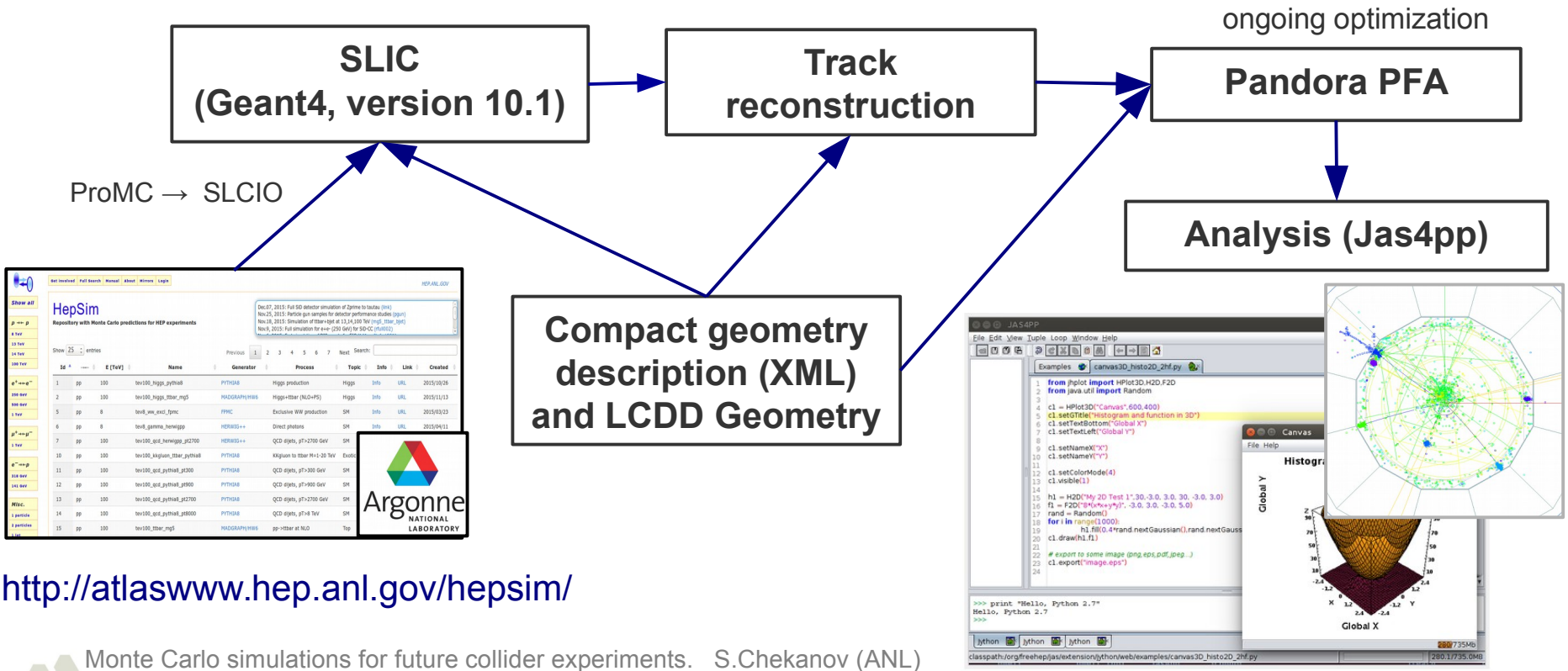
**HepSim**

- index files
- create metadata
- prepare for batch download
- validate with Jython scripts
- create search database

# Software for full simulations

## Simulator for the Linear Collider (SLIC) software

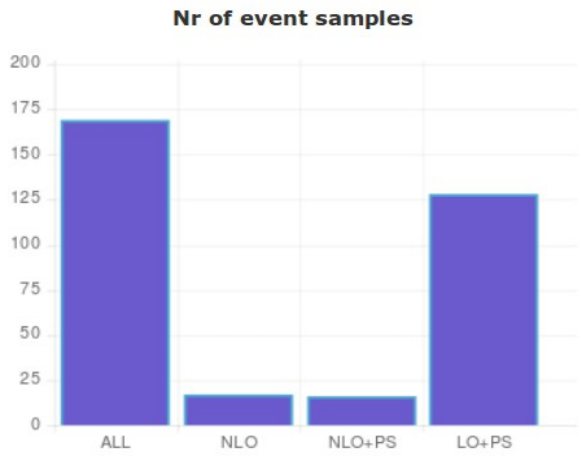
- Optimized for the SiD detector at SLAC (T.Johnson, N.Graf, J.McCormick, J.Strube)
- Re-purposed for future pp collider studies (S.C., A.Kotwal, J.Strube)
- Integrated with HepSim. Deployed on Open-Science Grid (OSG)
- Analysis: C++/Root or **Jas4pp** (ANL,S.C,E.May). Based on Jas3 (SLAC)




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

# HepSim event statistics

(excluding fast and Geant4 detector simulations)




| | |
|---|---|
| Number of public file servers | 3 |
| Number of event samples | 208 |
| Number of NLO samples | 17 |
| Number of NLO+PS samples | 17 |
| Number of LO (+PS) samples | 144 |
| Number of events | 1560741507 |
| NLO events | 583000000 |
| NLO+PS events | 32860595 |
| LO (+PS) events | 859498212 |
| Total size (GB) | 6897.468 |
| NLO size (GB) | 238.06 |
| NLO+PS size (GB) | 348.693 |
| LO (+PS) size (GB) | 6292.482 |
| Number of files | 334606 |

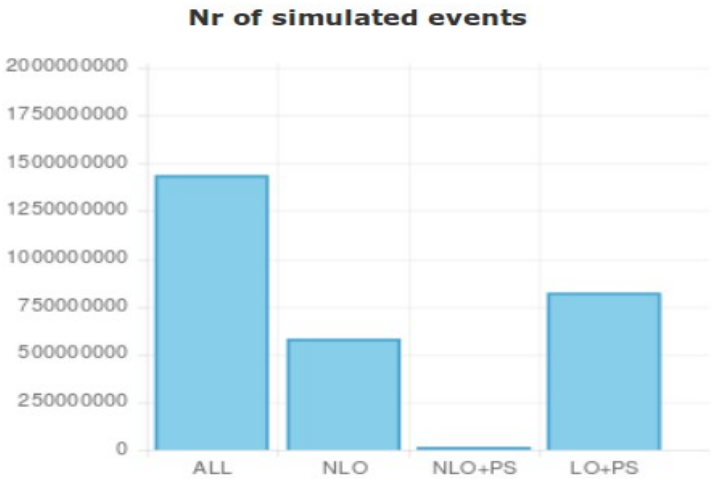
~210 Monte Carlo samples

~1.6 billion EVGEN events

- ~ 10% after fast simulations(Delphes)
- ~ 0.1% after Geant4 simulations

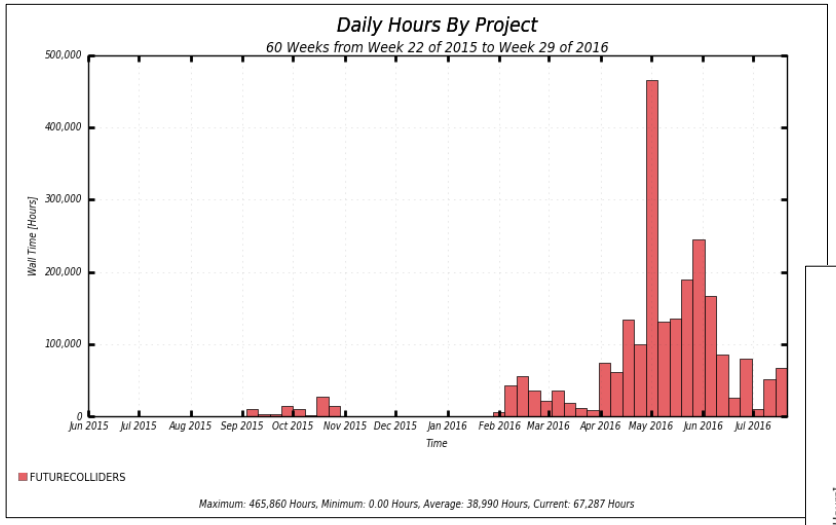
**Platforms for event generations (EVGEN)**

- 10% → BlueGene/Q (ANL/Mira) (Jetphox, MCFM)
- 50% → HEP-ANL (mainly Madgraph)
- 40% → OSG-CI grid and USATLAS CI (for phase II)

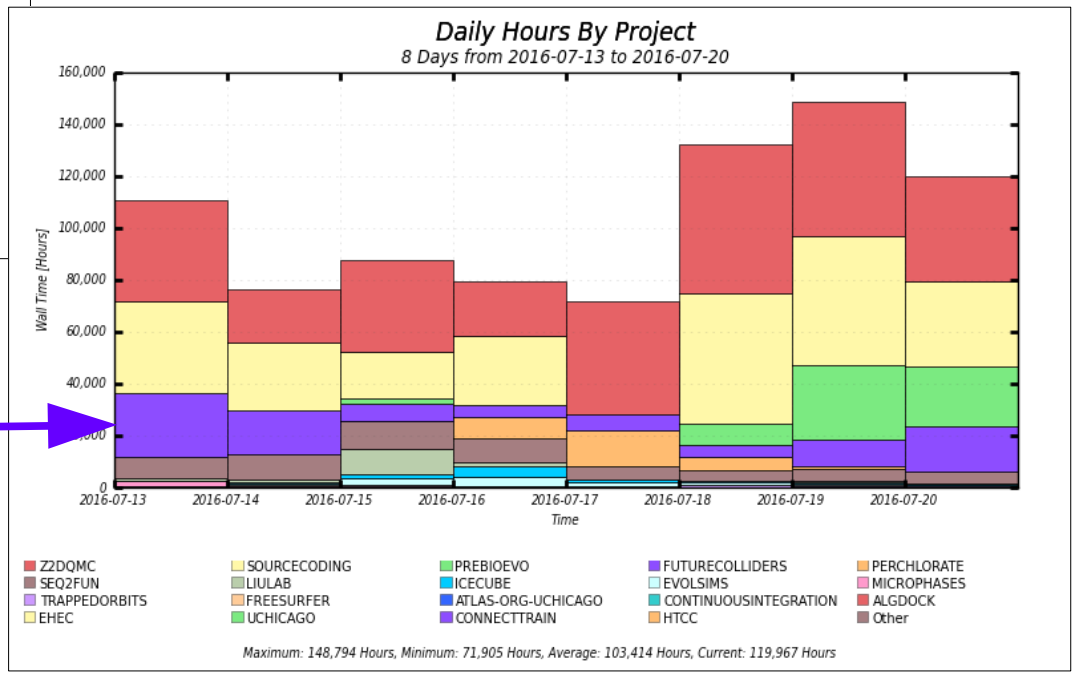


# CPU usage for SLIC (Geant4) simulations

OSG-Connect "FutureColliders" project for HepSim jobs




**CPU*h OSG-Connect**




**HepSim simulations (SLIC/Geant4) ~ 16 GB RAM per job**

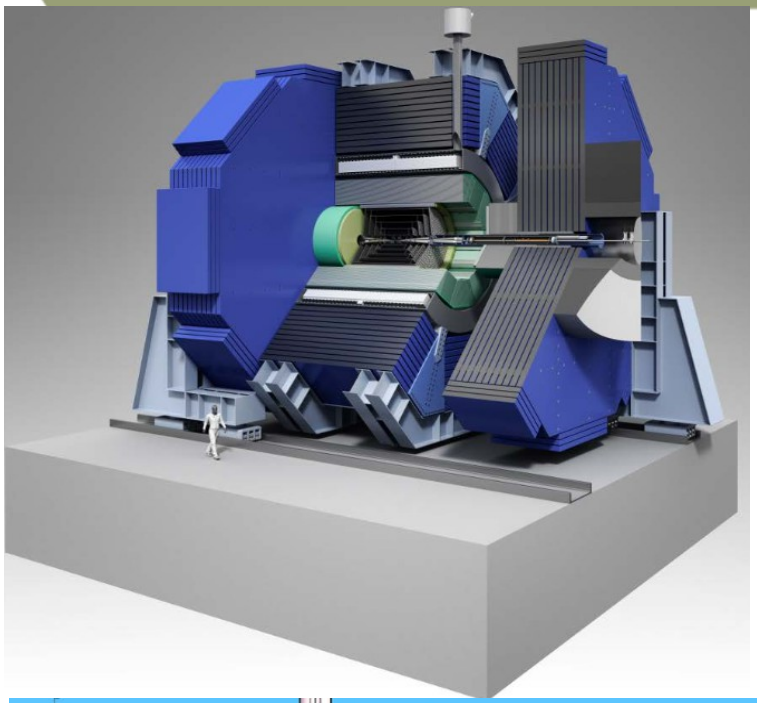
**2.5 million CPU*h in 2016 using OSG-grid for Geant4 simulations**

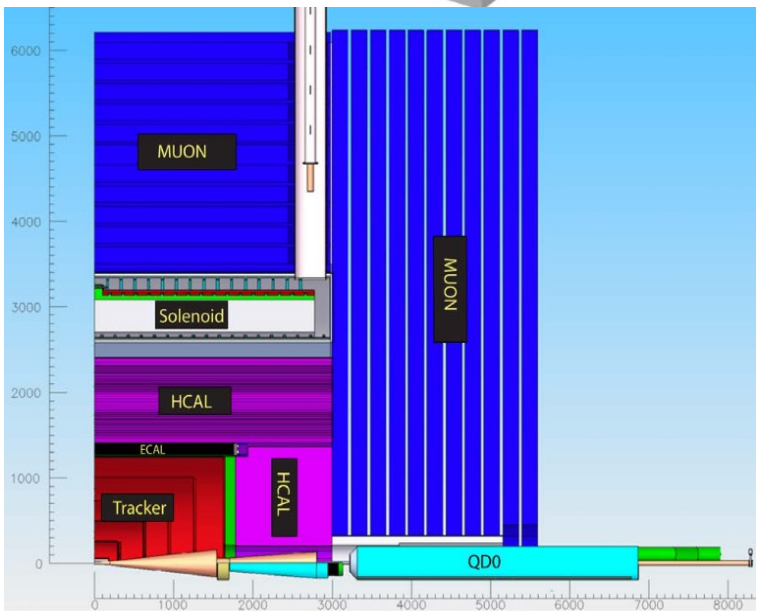
(equivalent to ~10 million CPU*h on HPC BlueGene/Q Mira)

# SiD detector for ILC

- Multi-purpose detector for the ILC
- Conceived at SLAC (USA LC Physics Group)
- The key characteristics:
  - 5 Tesla solenoid
  - Silicon tracker: 25/50 um readout pitch
  - ECAL: (0.35 cm cell size, W / silicon)
  - HCAL:
    - 1x1 cm cell size (RPC for LOI3*)
    - 40 layers for barrel (HCAL) ~4.5 $\lambda_I$
- Optimized for particle-flow algorithms (PFA)
- Fully configurable using SLIC software

→ **See talk by Andrew White**

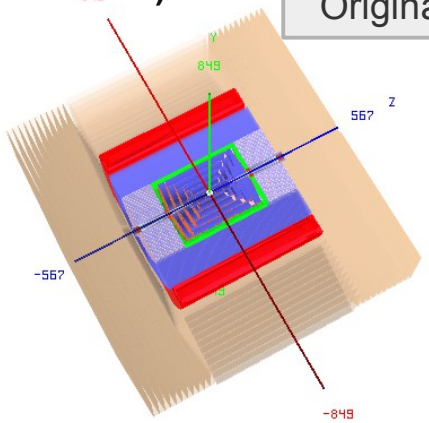




# ‘All-silicon’ design concepts supported in HepSim

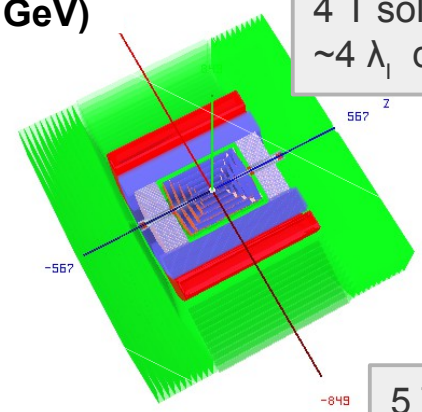
**SiD**
**($e^+ e^-$ up to 1 TeV)**

Original SiDLO3



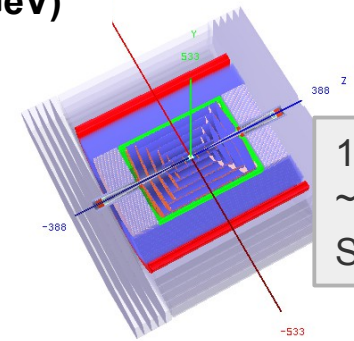

**SiCPEC**
**( $e^+ e^-$ 250 GeV)**

4 T solenoid
~4 $\lambda_I$ digital HCAL



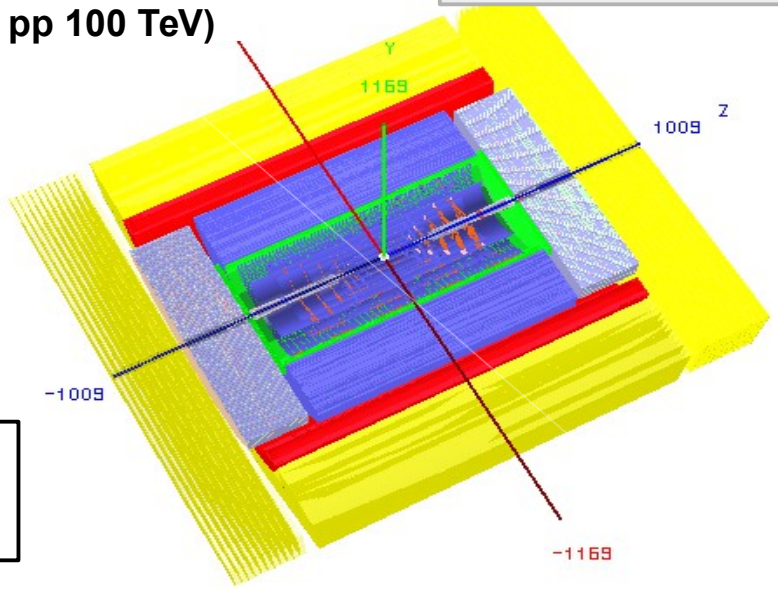

5 T solenoid
~12 $\lambda_I$ Scint/Fe HCAL
Larger size

**SiEIC**
**(ep, 141 GeV)**




1 T solenoid
~3 $\lambda_I$ HCAL
Smaller size

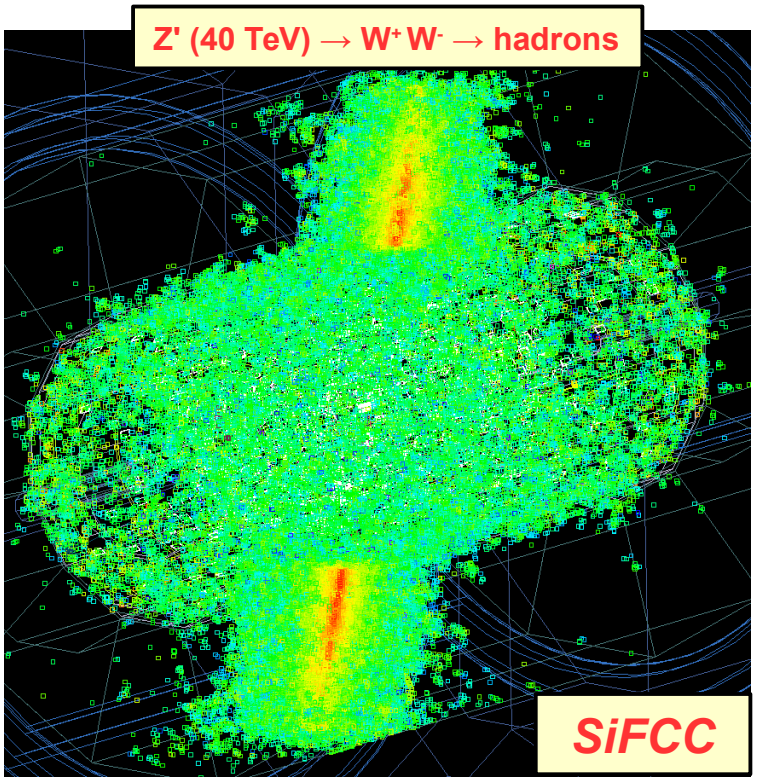
**SiFCC**
**(FCC-hh, pp 100 TeV)**




**HepSim provides single-particle and physics events after simulation & reconstruction**

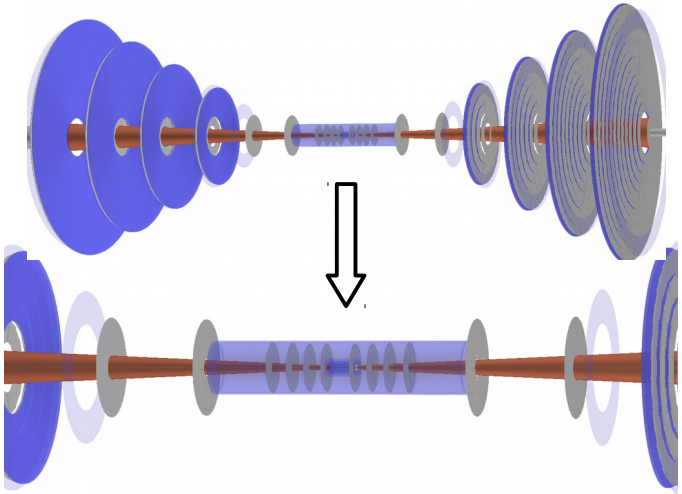
# Example I: FCC-hh detector studies

- See the ICHEP presentations by Shin-Shan Yu
- ICHEP16 posters by Nhan. V Tran and Sourav Sen




Large number of back-splash interactions in ECAL/HCAL
~4 CPU*h to simulate/reconstruct, 16 GB RAM
→ **CPU intensive!**

*Studies of tracking at multi-TeV scale:*
(J.Zuzelski, S.C., J.Proudfoot)



*Collaboration to study a high-granularity calorimeters:*
A.Kotwal (Fermilab/Duke), L.Gray (Fermilab), J.Strube (PNNL), N.Tran (Fermilab), S. Yu (NCU), S.Sen (Duke), J.Repond (ANL), J.McCormick (SLAC),J.Proudfoot (ANL), A.M.Henriques Correia (CERN), C.Solans (CERN), C.Helsens (CERN), S.C, etc.

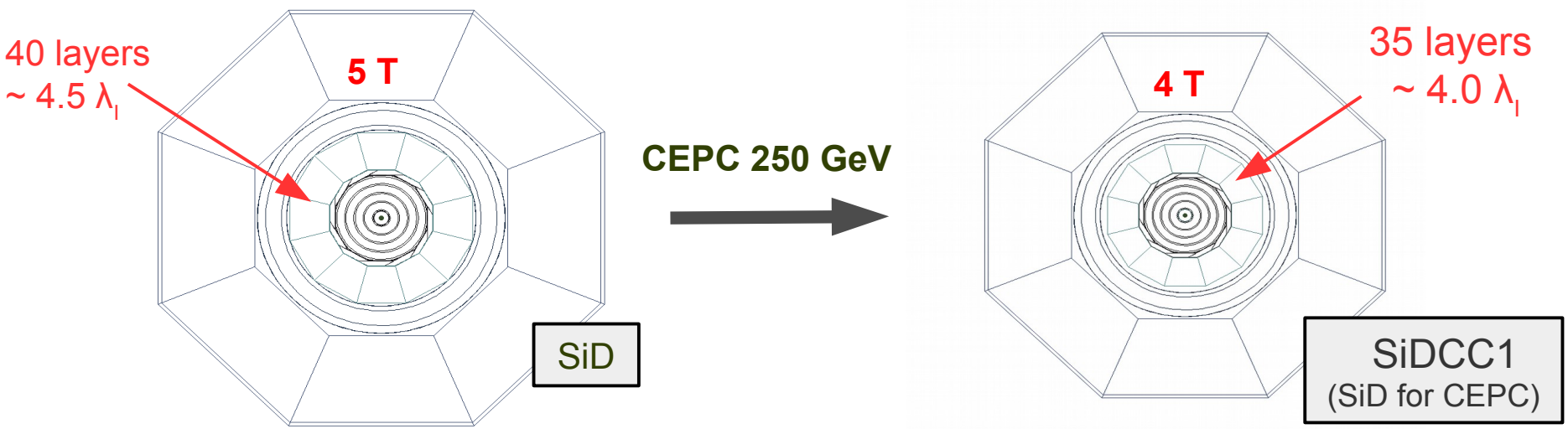
**Studies of jets and tracks up to ~20 TeV!**

# Example II: Designing a detector for CEPC (e+e- 250 GeV)

SiD detector is designed for ~500 GeV particles/jets (0.5-1 TeV CM energy)
But CEPC will measure particles/jets up to 125 GeV (250 CM energy)

Possible optimizations:

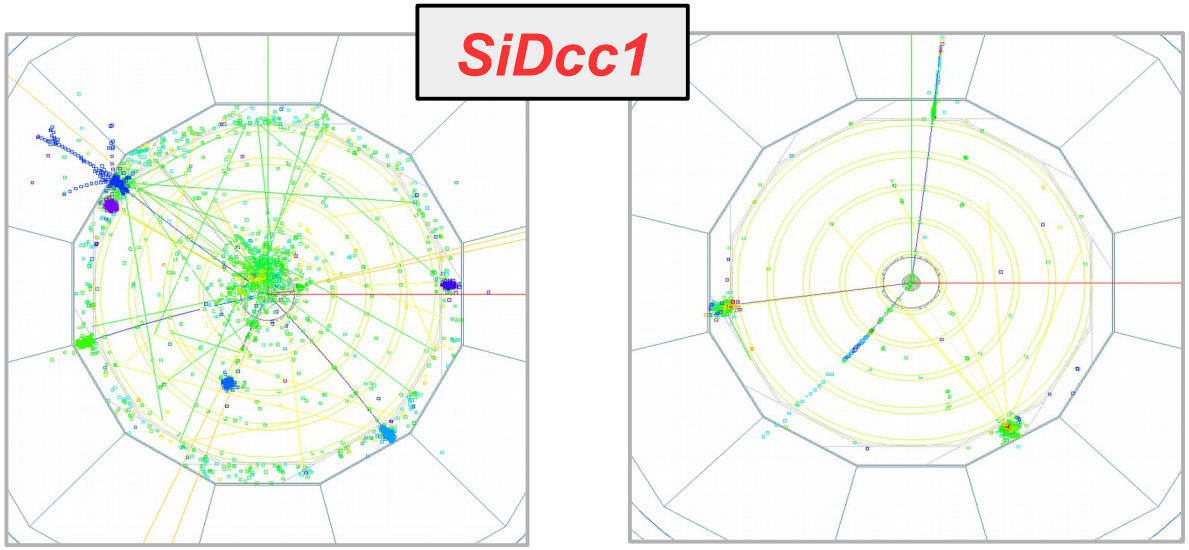
| | | |
|---|---|---|
| **- HCAL**: | barrel: 4.5 $\lambda_I$ (40 layers) | → 4.0 $\lambda_I$ (35 layers) |
| | endcap: 5 $\lambda_I$ (45 layers) | → 4.0 $\lambda_I$ (35 layers) |
| **- Tracking:** | 5 Tesla | → 4 Tesla |




**Design a light, cost-optimized version of the SiD detector for CEPC and use several physics processes to benchmark its performance**

S.C. and M.Demarteau, Conceptual Design Studies for a CEPC Detector.
arXiv:1604.01994. HKUST IAS 2016 proceeding

# Example II: Simulations for CEPC ($e^+e^-$ 250 GeV CM energy)




H(125) → 4 e

H(125) → γγ

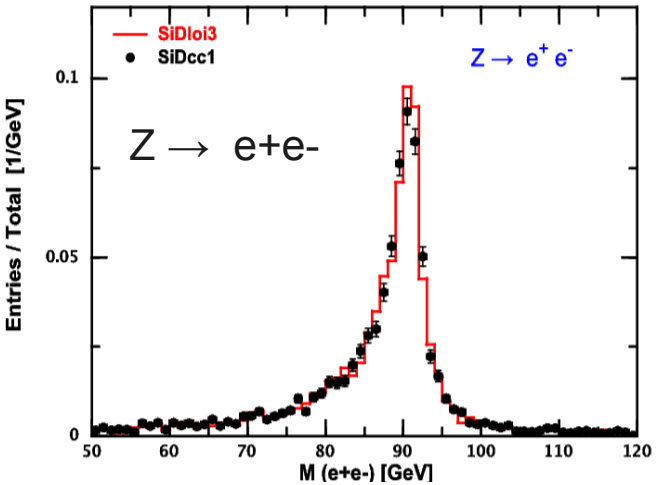
~ 10,000 reconstructed events for each physics channel after PFA (Pandora) reconstruction

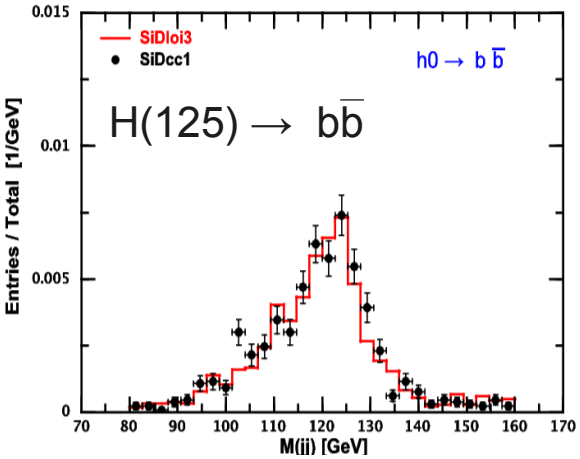
- **Available full simulations for the SiD and SiDCC (for CPC) detectors:**
  - Z → e+e-, Z → tau tau, Z → mu+mu-, Z → $b\bar{b}$
  - H(125) → $b\bar{b}$ H(125) → γγ, H(125) → ZZ* → 4l, H(125) → tau+tau-

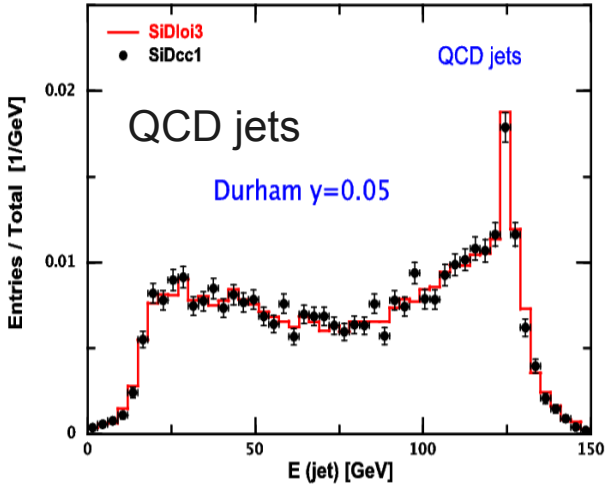
# Comparing SiD with SiDCC1

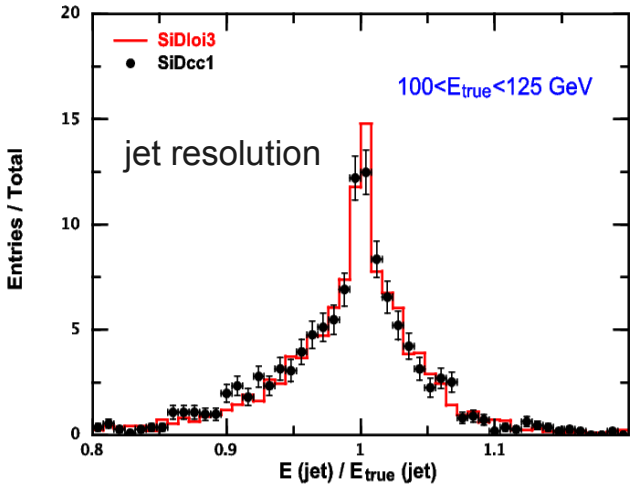
*Done with Jas4pp*

- Benchmark processes for e+e- (250 GeV)
  - Z → e+e-, Z → $b\bar{b}$, Z → tau+tau-, H → γγ
  - H → 4l, H → $b\bar{b}$, QCD jets
- Particle flow objects to reconstruct invariant masses and jet energy resolutions (Durham jets)




Simplification of the SiD detector does not compromise physics performance

S.C. and M.Demarteau. arXiv:1604.01994. HKUST IAS 2016 proceeding

# Summary

- **First public Monte Carlo repository with fast and full detector simulations**
- **Enable physics & detector-performance studies for current & future colliders + community outreach**
  - 1.6 billion events at the EVENT level for public downloads
  - Significant number of fast and fully reconstructed events for ep, μμ, ee, pp collisions (13-100 TeV) & single-particle samples for detector studies
- **14 articles, ~25 presentations since 2014 (linked to WWW):**
  - Physics reach studies for HL-LHC, HE-LHC, FCC-hh etc.
  - Calorimeter studies (cell granularity)
  - Tracking optimization at multi-TeV scale
  - Software development
- **2.5 million CPU*h from OSG-grid. OSG-Connect support from UChicago**
- **Contributions from 17 students/scientists**
- **Your contribution is welcome!**

Some studies presented at this meeting!

# Backup

# Long-term preservation of theoretical calculations

- **Storing Monte Carlo predictions in files makes sense if:**

$$\frac{\text{time to download \& analyze on commodity computer}}{\text{CPU*h needed to create the prediction}} \equiv \varepsilon << 1$$

| | | |
|---|---|---|
| $\varepsilon \sim 0.01\text{-}1$ | - | for LO MC |
| $\varepsilon << 0.01$ | - | for NLO etc. |

- **$\varepsilon << 1$:**
  - Madgraph5 etc. (NLO+PS+hadronisation), ALPGEN
  - Some fast-converging NLO calculations (MCFM, jetPHOX etc)
  - MC with $\varepsilon \sim 1$ but after mixing with pile-up (CPU intensive)
- **$\varepsilon \sim 1$:** Less appropriate approach for:
  - LO simulations (Pythia)
  - Some NLO programs with slow convergence
    - requires too large data volumes to keep weighted events

# New data format for EVGEN: ProMC

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

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

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

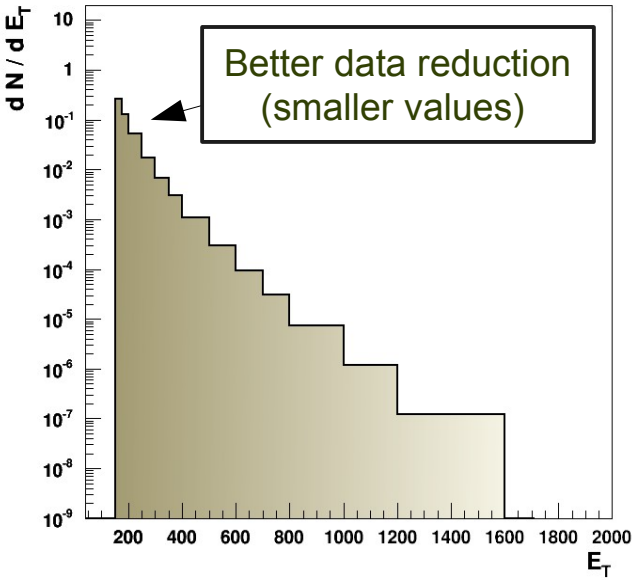
Google's Protocol buffers

protobuf
Protocol Buffers - Google's data interchange format

- Effective file size reduction for pile-up events
  - Particles with small momenta → less bytes used
- Installed on Mira (BlueGene/Q)
- Separate events can be streamed over the Internet:
  - similar to avi frames for web video players

http://atlaswww.hep.anl.gov/asc/promc/

8-bytes → varint




compression strength keeping precision of representation constant

# Dataset entry: e+e- collisions (CM energy = 250 GeV ). Z → e+e-

- $p \rightarrow\leftarrow p$
  - 8 TeV
  - 13 TeV
  - 14 TeV
  - 100 TeV
- $e^+ \rightarrow\leftarrow e^-$
  - 250 GeV
  - 500 GeV
  - 1 TeV
- $\mu^+ \rightarrow\leftarrow \mu^-$
  - 1 TeV
  - 5 TeV
  - 10 TeV
  - 20 TeV
  - 40 TeV
- $e^- \rightarrow\leftarrow p$
  - 318 GeV
  - 141 GeV
- Misc.
  - 1 particle
  - 2 particles
  - 1 jet

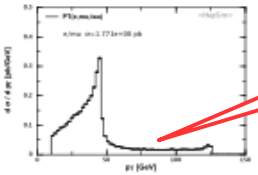
Repository with Monte Carlo predictions for HEP experiments

Feb.1, 2016: Z' with M=10,20,40 TeV decaying t
Jan.19, 2016: 10 TeV Z' using a full simulation w
Jan.14, 2016: TTbar+N jet process (pp, 14 TeV, M

## Information about *"gev250ee_pythia6_zpole_ee"* dataset

| | |
|---|---|
| Name: | *gev250ee_pythia6_zpole_ee* |
| Collisions: | e+e- |
| CM Energy: | 0.25 TeV |
| Entry ID: | 146 |
| Topic: | SM |
| Generator: | PYTHIA6 |
| Calculation level: | LO+PS+hadronisation |
| Process: | Z boson to e+e- |
| Total events: | 2000000 |
| Number of files: | 100 |
| Cross section (σ): | 1.7765 ± 0.0126 pb |
| Luminosity (L): | 1.126E+06 $pb^{-1}$ *(or)* 1125.7948 $fb^{-1}$ *(or)* 1.1258 $ab^{-1}$ |
| Format: | ProMC |
| Submission date: | Tue Oct 13 14:28:55 CDT 2015 |
| Download URL: | http://mc.hep.anl.gov/asc/hepsim/events/ee/250gev/pythia6_zpole_ee |
| Mirrors: | |
| MC truth size: | 0.826 GB |
| Fast simulation: | rfast001 (info) \| |
| Full simulation: | rfull002 (info) \| rfull001 (info) \| |
| Record slimmed: | No |
| Events weighted: | No |
| User description: | PYTHIA version 6.4. Z production (Zpole) with decays to e+e-. Other details in the |
| File metadata: | **ProMC version:** 4; **Nr events:** 1000; **Varint E:** 1000000; **Varint L:** 10000; **Lo** logfile.txt; **Last modified:** 2015-10-15 20:31:08; **Settings:** PYTHIA-6.4.28; MSE mix events; NTOT 0 0 1000 ! Number of events; ECME 0 0 250.0 ! CM energy (Ge 0 0 839264 ! random seed; MSEL 0 0 0 ! all mixed events; PMAS 6 1 172.5 !; PMA 91.1876 ! Z boson mass; PMAS 24 1 80.3850 ! W boson mass; PMAS 25 1 125. ! Hi mass; MSUB 1 0 1 ! ffbar to Z; MSTP 43 0 2 ! Z only, no gamma; MDME 174 1 0 ! MDME 175 1 0 ! U U~; MDME 176 1 0 ! S S~; MDME 177 1 0 ! C C~; MDME 178 1 B~; MDME 179 1 0 ! T T~; MDME 182 1 1 ! E- E+; MDME 183 1 0 ! NU_E NU_E~; M 1 0 ! MU+ MU-; MDME 185 1 0 ! NU_MU+ NU_MU-; MDME 186 1 0 ! TAU- TAU+; M 1 0 ! NU_TAU- NU_TAU+; PARJ 71 0 10 ! ctau=10mm; MSTJ 22 0 2 !; |

Validation:

| Nr | Analysis code | Output plot (SVG) | Output (XML) |
|---|---|---|---|
| 1 | pythia6_zpole_ee.py<br>Launch<br>Desktop: **hs-ide** [URL] |  | pythia6_zpole_ee.jdat |

URL for EVGEN files (download or streaming)

URL with fast (DELPHES) or Geant4 (SLIC) simulations

Validation distributions created using Python scripts on the Java platform

Run via JavaWeb start by streaming data over the Web

# Benchmarks for EVGEN files

ProMC files:
- 12 times smaller than HEPMC
- 30% smaller than ROOT
  30% faster to process (C++/Java)

File sizes for 10,000 $t\bar{t}$ events for pp at LHC

| File format | File Size (MB) | C++ (sec) | CPython (sec) | Java (sec) | Jython (sec) |
|---|---|---|---|---|---|
| ProMC | 307 | 15.8 | 980 | 11.7 (12.1 +JVM startup) | 33.3 (35 +JVM startup) |
| ROOT | 423 | 20.4 | 66.7 (PyROOT) | - | - |
| LHEF | 2472 | 84.7 | 30.4 | 9.0 (9.6 +JVM startup) | - |
| HEPMC | 2740 | 175.1 | - | - | - |
| LHEF (gzip) | 712 | - | - | - | - |
| LHEF (bzip2) | 552 | - | - | - | - |
| LHEF (lzma) | 513 | - | - | - | - |
| HEPMC (gzip) | 1021 | | | - | - |
| HEPMC (bzip2) | 837 | | | - | - |
| HEPMC (lzma) | 802 | | - | - | - |

ASCII text files
(after compression)

*Table 1. Benchmark tests for reading files with 10,000 ttbar events stored in different file formats. For each test, the memory cache on Linux was cleared. In case of C++, the benchmark program reads complete event records using appropriate libraries. CPython code for ProMC file is implemented in pure CPython and does not use C++ binding (unlike PyROOT that uses C++ libraries). In case of LHEF files. JAVA and CPYTHON benchmarks only parse lines and tokenize the strings, without attempting to build an event record, therefore, such benchmarks may not be accurate while comparing with ProMC and ROOT.*

https://atlaswww.hep.anl.gov/asc/wikidoc/doku.php?id=asc:promc:introduction