U.S. DEPARTMENT OF ENERGY | Office of Science | JSA

Generic R&D for an EIC:

Developing Analysis Tools and Techniques for the EIC

Whitney Armstrong (ANL), Elke-Caroline Aschenauer (BNL), Franco Bradamante (INFN Trieste), Andrea Bressan (INFN Trieste), Andrea Dotti (SLAC), Sergei Chekanov (ANL), **Markus Diefenthaler (Jefferson Lab, co-PI)**, **Alexander Kiselev (BNL, co-PI)**, Anna Martin (INFN Trieste), Christopher Pinkenburg (BNL), Stefan Prestel (SLAC)

ESC Meeting
October 17th 2016

Jefferson Lab
Thomas Jefferson National Accelerator Facility

Agenda

Review of existing software

- What technology is used?
- What is available?
- How flexible?
- Examples

Talks by Alexander, Chris, Mauri, Sergei || Whit

Discussion about requirements

- What does the community need?
- What is urgently required?
- What long-term goals do we have?

Discussion about common goals and work plan

- focus on geometry and detector interface
- focus on unified tracking

Talks by Haiwang, Markus, Whit

PAST

PRESENT

FUTURE

Forming a software consortium for the EIC

September 2015 EIC Software Meeting
Workshop organized by Elke-Caroline Aschenauer and Markus Diefenthaler
https://www.jlab.org/conferences/eicsw/
review of existing EIC software frameworks and MCEG available for the EIC

January 2016 Generic R&D Meeting: **LOI for Software Consortium**

> **Review** "A robust software environment, compatible with the existing software frameworks, is very important for the development of the physics case for the EIC."

March 2016 Future Trends in NP Computing
Workshop organized by Amber Boehnlein, Graham Heyes, and Markus Diefenthaler
https://www.jlab.org/conferences/trends2016/
discussion of computing trends, e.g., Big Data, machine learning, Exascale Computing
incubator for ideas on how to improve analysis workflows in NP

July 2017 Generic R&D Meeting: **Proposal for Software Consortium**
consisting of scientists from ANL, BNL, JLab, INFN Trieste, and SLAC
R&D funds for workshop, travel, and students have been awarded (eRD20)

Global objectives

Interfaces and integration
- connect existing frameworks / toolkits
- identify the key pieces for a future EIC toolkit
- collaborate with other R&D consortia

Planning for the future with future compatibility
- workshop to discuss new scientific computing developments and trends
- incorporating new standards
- validating our tools on new computing infrastructure

Organizational efforts with an emphasis on communication
- build an active working group and foster collaboration
- documentation about available software
- maintaining a software repository
- workshop organization

building up on existing documentation:
https://wiki.bnl.gov/eic/index.php/Simulations and related pages

PAST

PRESENT

FUTURE

Immediate development in FY17

Interfaces and integration

- start the development of a library for simulating radiate effects
- **work towards a common geometry and detector interface**
- **work towards an unified track reconstruction**
- collaborate with **TMD MC** and **DPMJetHybrid** (eRD17) and other software projects that are essential for an EIC

Focus

Planning for the future with future compatibility

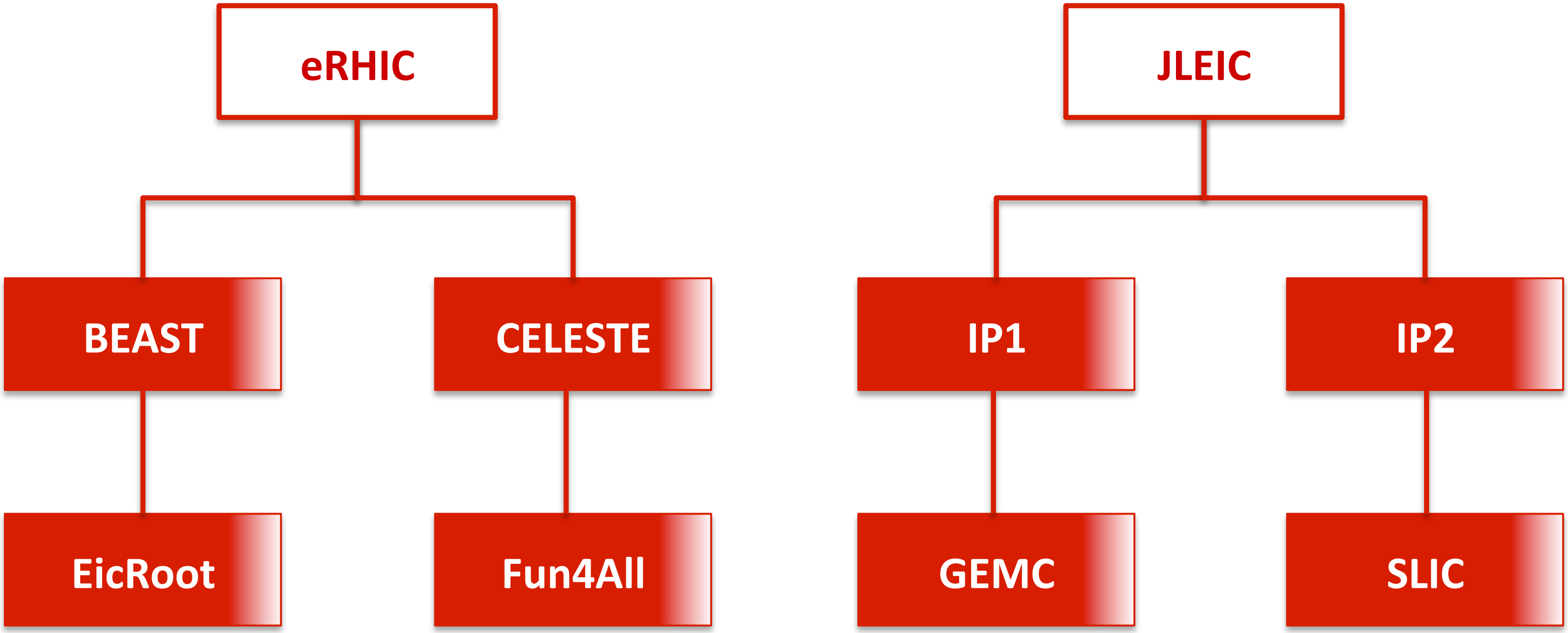
- validation of critical Geant4 physics in the energy regime of the EIC
- start the development of an universal event display for MC events
- promote open-data developments for efficient data-MC comparison from the beginning
- build interfaces to forward compatible, self-descriptive file formats

Organizational efforts with an emphasis on communication

- build a community website
- organize software repositories dedicated to the EIC
- organize a workshop

Existing software frameworks for the EIC

Talks by Alexander, Chris, Mauri, Sergei || Whit



Building on existing EIC software:

- build forward-compatible interfaces between existing frameworks / tools
- identify common tools and improve them (e.g. MCEG)
- add tools that are forward-compatible with existing frameworks

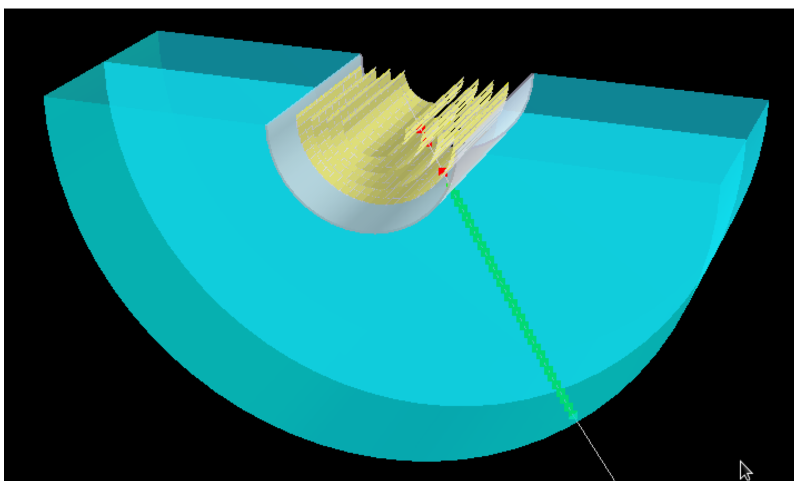
Unified track reconstruction library

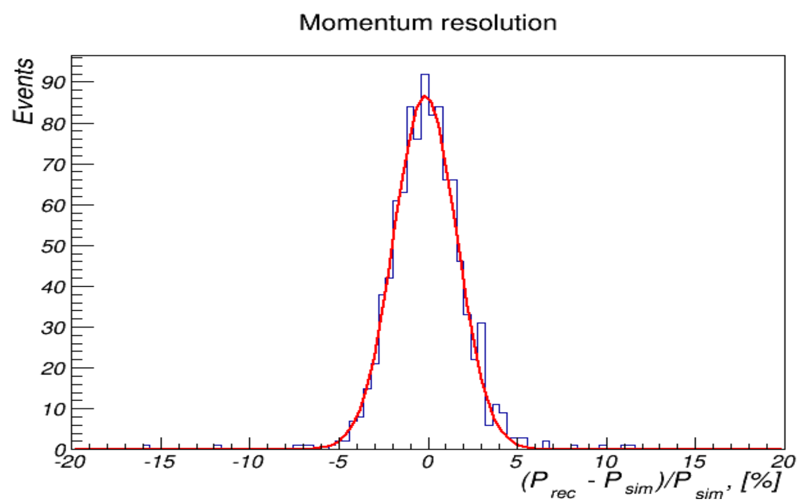
Talk by Haiwang

Pre-conditions

- Similar requirements for and similar tracker outline of all proposed EIC detectors
- Similar analysis dataflow from simulation to event reconstruction
- Existence of powerful generic libraries for track and vertex fitting (genfit, rave)
- Expertise in the EIC community
- Well-advanced EIC-related set of tracking R&D tools exists already (EicRoot):

Consider a basic example: a vertex tracker + a TPC in a realistic ~3T magnetic field; what is the momentum resolution for pions at p=10 GeV/c and θ=75°?





Distance between the above question and the momentum resolution plot is only ~200 lines of trivial ROOT scripts

But: the tool is at present software-framework-bound!

Unified track reconstruction library

The proposal

- Pull the relevant fraction of tracking-related tools out of the EicRoot framework
- Complement and/or upgrade them with up-to-date libraries (genfit2, rave, etc)
- Provide a suitable unified track finder code for the EIC tracker geometry

- Make use of EIC-specific and framework-independent geometry definition format
- Decide on flexible detector hit formats (raw; digitized; suitable for reconstruction)

Possible first year deliverables

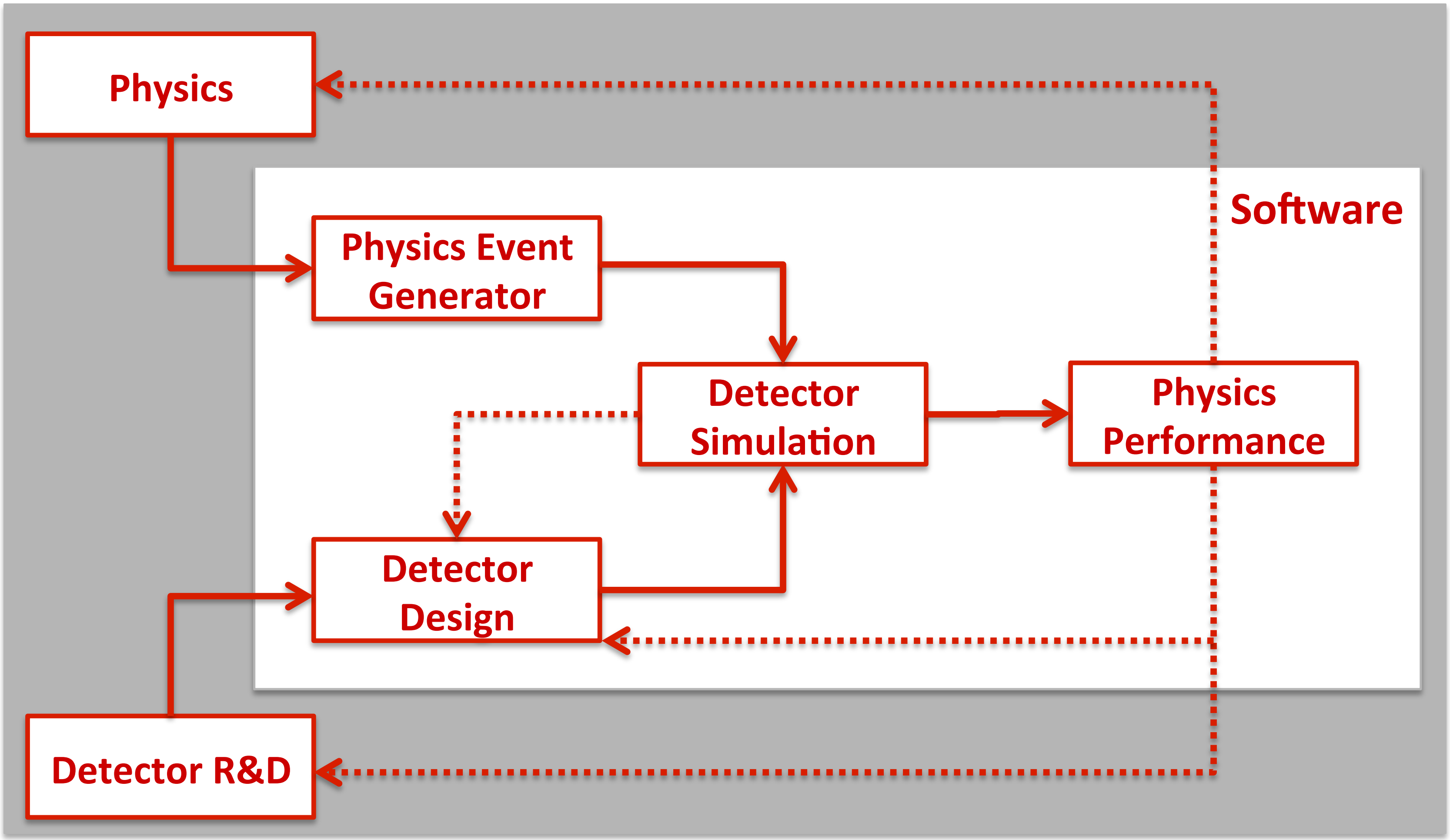
- Perform a detailed feasibility study of the above plan
- Should the task look doable, start code development with a universal standalone library of track *fitting* tools for a typical EIC tracker geometry

Potential benefits

- Provide a unified track reconstruction library which can be used in *any* EIC framework

- Leverage proposed geometry exchange procedure between different implementations
- Simplify detector performance comparisons between site-specific implementations

EIC R&D and software development



User Cases

- **User Case 1:** Requirements for studying a physics process at the EIC:
  - interface to MCEG
  - open access to accelerator specifications
  - open access to accelerator geometry || detector simulation
  - documentation
- **User Case 2:** Requirements for studying a detector at the EIC
  - open access to physics simulations || interface to MCEG
  - open access to accelerator specifications
  - open access to geometry && detector simulation
  - documentation
- **User Cases 1 and 2** might involve comparison of eRHIC and JLEIC:
  - eRHIC settings / geometry might be used in JLEIC software
  - JLEIC settings / geometry might be used in eRHIC software

What have I forgotten?

Unity via common data structures

Talk by Whit

- Common format for MC files? ProMC (next files)?
- Common format for simulation (generated events, reconstructed events):
  - mRun: settings (<-> self-descriptive data)
  - (m)Event, Event: event information
  - mProcess
  - mParticle
  - (m)Track
  - (m)Hit

Proposal: Let's start with the simple ones:

- Event: ID, x, Q2, …
- mEvent, ID, process
- mParticle
- Track / Particle: ID, charge, px, py, pz, E, theta, phi, particle type

Interfaces to self-descriptive file formats

New data format for EVGEN: ProMC
baseline in addition to ROOT

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

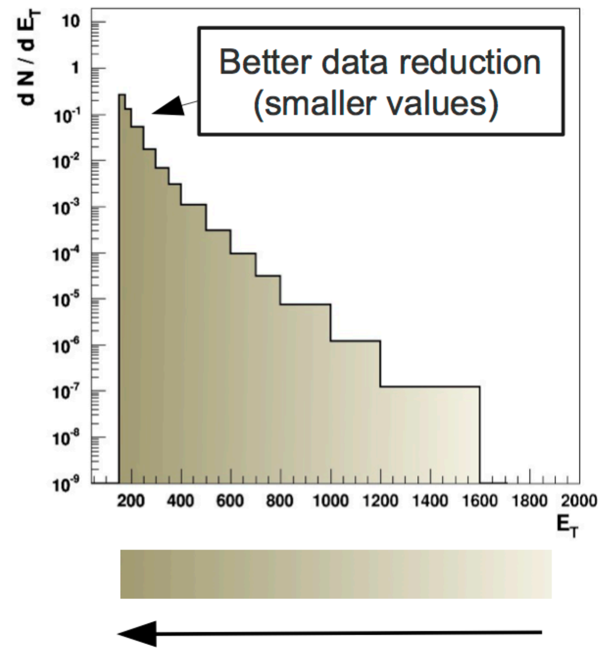
Number of used bytes depends on values.
Small values use small number of bytes

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).
- Supports C++/Java/Python
- Separate events can be streamed over the Internet:
  - similar to avi frames (video streaming)

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

8-bytes (int64) → varint

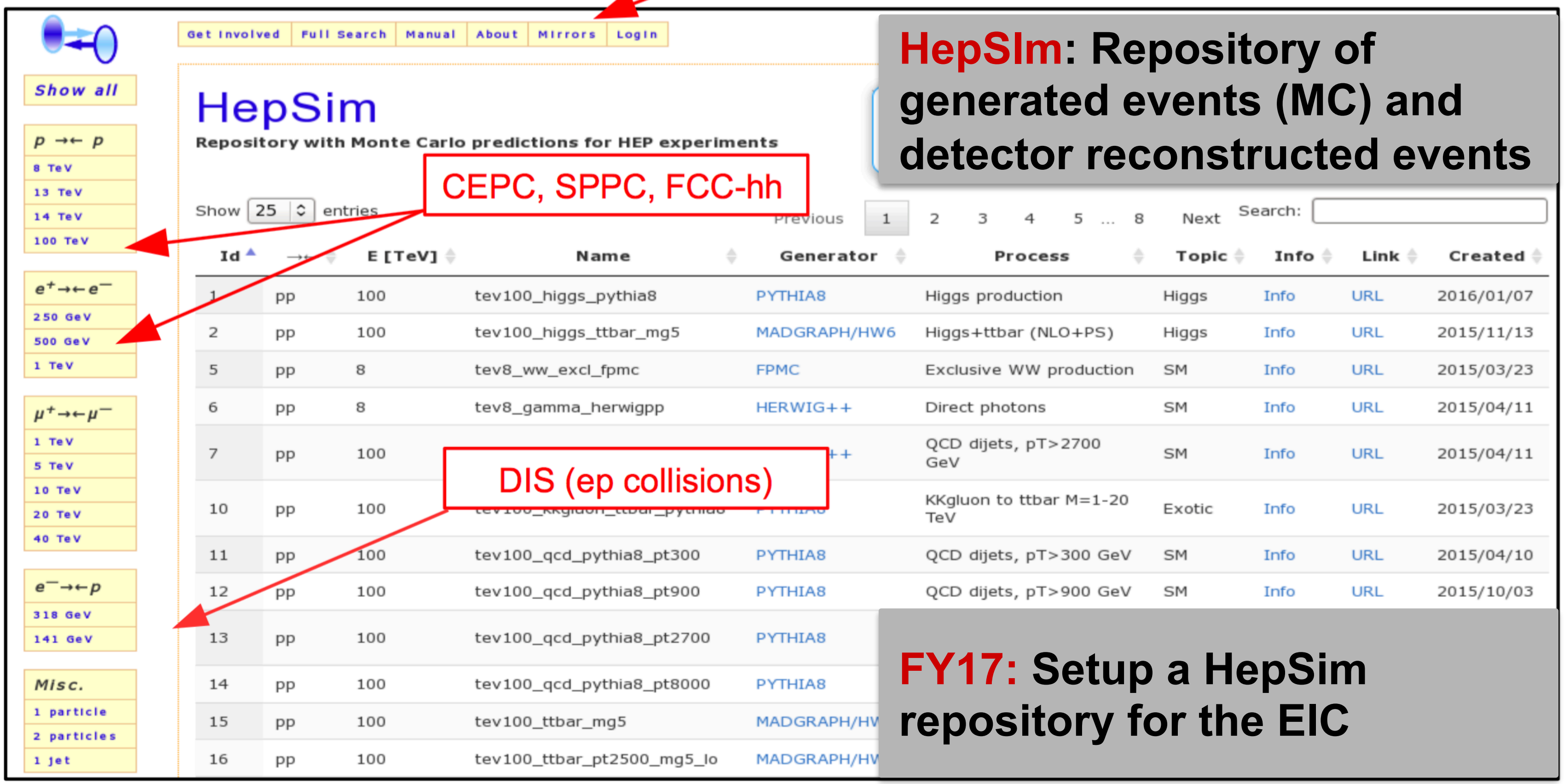


compression strength keeping precision of constant

HepSim repository for the EIC

HepSim simulation uses ProMC

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



Id	→←	E [TeV]	Name	Generator	Process	Topic	Info	Link	Created
1	pp	100	tev100_higgs_pythia8	PYTHIA8	Higgs production	Higgs	Info	URL	2016/01/07
2	pp	100	tev100_higgs_ttbar_mg5	MADGRAPH/HW6	Higgs+ttbar (NLO+PS)	Higgs	Info	URL	2015/11/13
5	pp	8	tev8_ww_excl_fpmc	FPMC	Exclusive WW production	SM	Info	URL	2015/03/23
6	pp	8	tev8_gamma_herwigpp	HERWIG++	Direct photons	SM	Info	URL	2015/04/11
7	pp	100		++	QCD dijets, pT>2700 GeV	SM	Info	URL	2015/04/11
10	pp	100			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	tev100_qcd_pythia8_pt900	PYTHIA8	QCD dijets, pT>900 GeV	SM	Info	URL	2015/10/03
13	pp	100	tev100_qcd_pythia8_pt2700	PYTHIA8					
14	pp	100	tev100_qcd_pythia8_pt8000	PYTHIA8					
15	pp	100	tev100_ttbar_mg5	MADGRAPH/HW					
16	pp	100	tev100_ttbar_pt2500_mg5_lo	MADGRAPH/HW					

HepSim stores EVGEN files (LO,NLO, etc), fast simulations, full Geant4 simulations

PAST

PRESENT

FUTURE

Analysis environments

Developments of analysis environments:

- new projects starting (JLab 12 GeV) and on the horizon (EIC)
- likely explosion of data even at the small nuclear experiments
- think about the **next generation(s) of analysis environments** that will **maximize** the **science output**

LHC experiments: tremendous success in achieving their analysis goals and producing results in timely manners

Lesson learned at LHC experiments:

- as the complexity and size of the experiments grew
- the complexity of analysis environment grew
- time dealing with the analysis infrastructure grew

Anecdote from LHC
a typical LHC student or post-doc spends up to 50 % of his/her time dealing with computing issues

New analysis environments

User centered design

- understand the user requirements first and foremost
- engage wider community of physicists in design whose primary interest is not computing
- make design decisions solely based on user requirements
- web-based user interfaces, e.g. interactive analysis in Jupyter Notebook

Future compatibility (both hardware and software)

- most powerful future computers will likely be very different from the kind of computers currently used in NP (Exascale Computing)
- structures robust against likely changes in computing environment
- apply modular design: changes in underlying code can be handled without an entire overhaul of the structure

Think out of the box

- the way analysis is done has been largely shaped by kinds of computing that has been available
- computing begins to grow in very different ways in the future, driven by very different forces than in the past (Exascale Computing)
- think about new possibilities and paradigms that can and should arise