Coordinamento del software per EIC e per ATHENA

A. Bressan
UniTS e INFN Sezione di Trieste

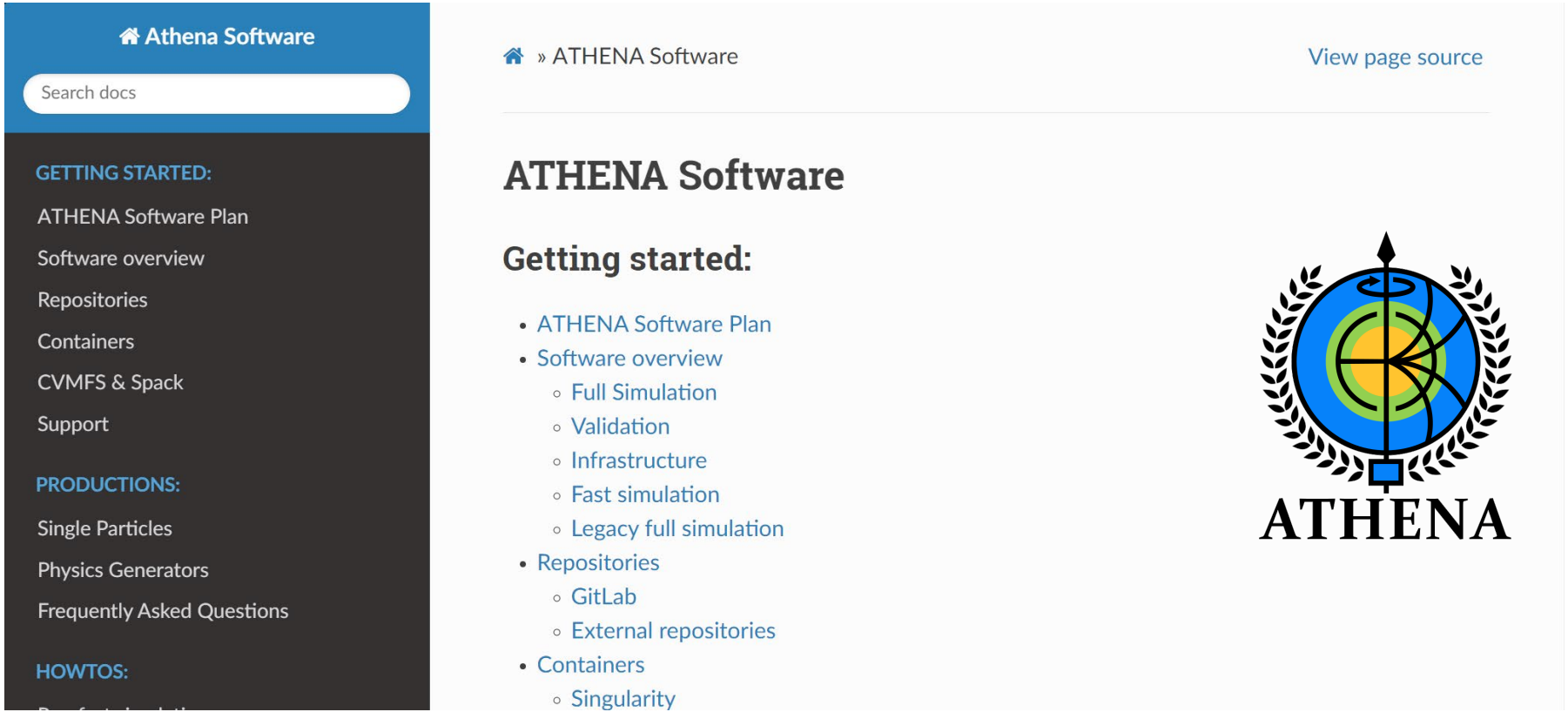
Outlook

Report sulle seguenti attività

- Il software WG di Athena, Software tools e produzioni
- Attività dell'EIC SWG
- Attività del Computing Coordination Group (CCG) formato recentemente
- CNAF e OSG

Website

https://eic.phy.anl.gov/ip6/

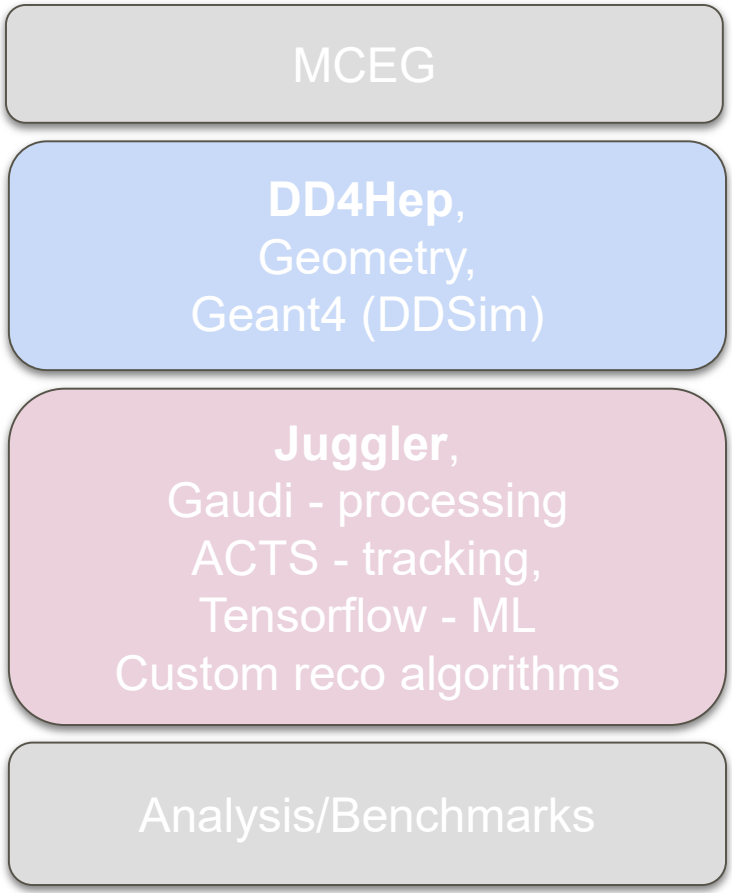


Sviluppo software di Athena

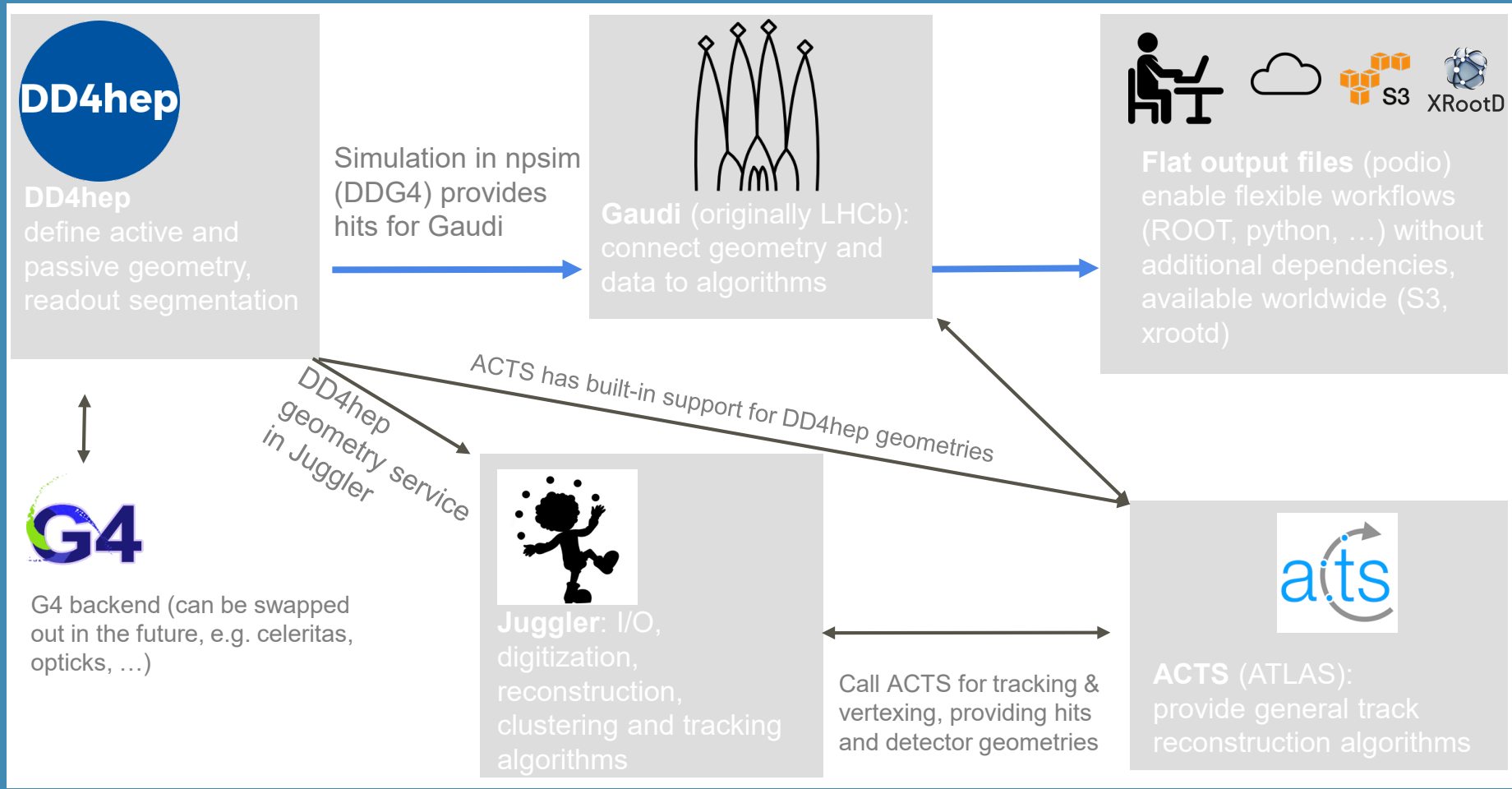
- Conveners del SWG: AB, Dmitry Romanov (Jefferson lab) , Sylvester Joosten (Argonne National Laboratory) , Whitney Armstrong (Argonne National Laboratory) , Wouter Deconinck (The University of Manitoba)
- Il software (costruito in 3 mesi) e' basato sulle competenze del gruppo di Argonne
- Sviluppo basato su un approccio moderno
  - Modularità
  - Integrazione con workflow per sistemi HTC/HPC,
  - Uso di tools di CI/CD
- Usare al massimo CERN-supported software quando possibile.

Software stack di Athena

- DD4hep: Geant4 geometry, detector plugin library, wrappers to run Geant4
- Juggler: Digitization and reconstruction software (based on Gaudi with Podio-based data model and ACTS for tracking)
- Gaudi: Generic open project for building event processing frameworks. Enables modern task-based concurrent execution in a heterogeneous computing environment. Used by ATLAS and LHCb.
- ACTS: Experiment-independent tracking toolkit (ACTS' geometry constructed from DD4hep via plugin)
- Podio: Robust data model definition to cross the boundaries between the tools

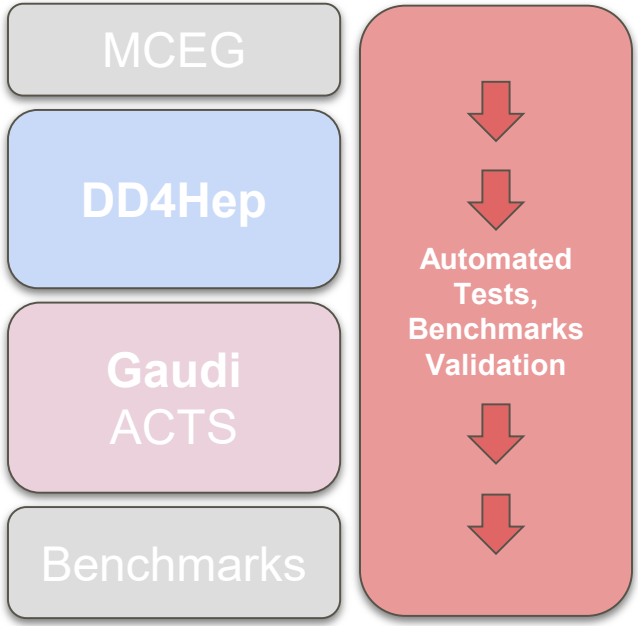


ATHENA Software Ecosystem:



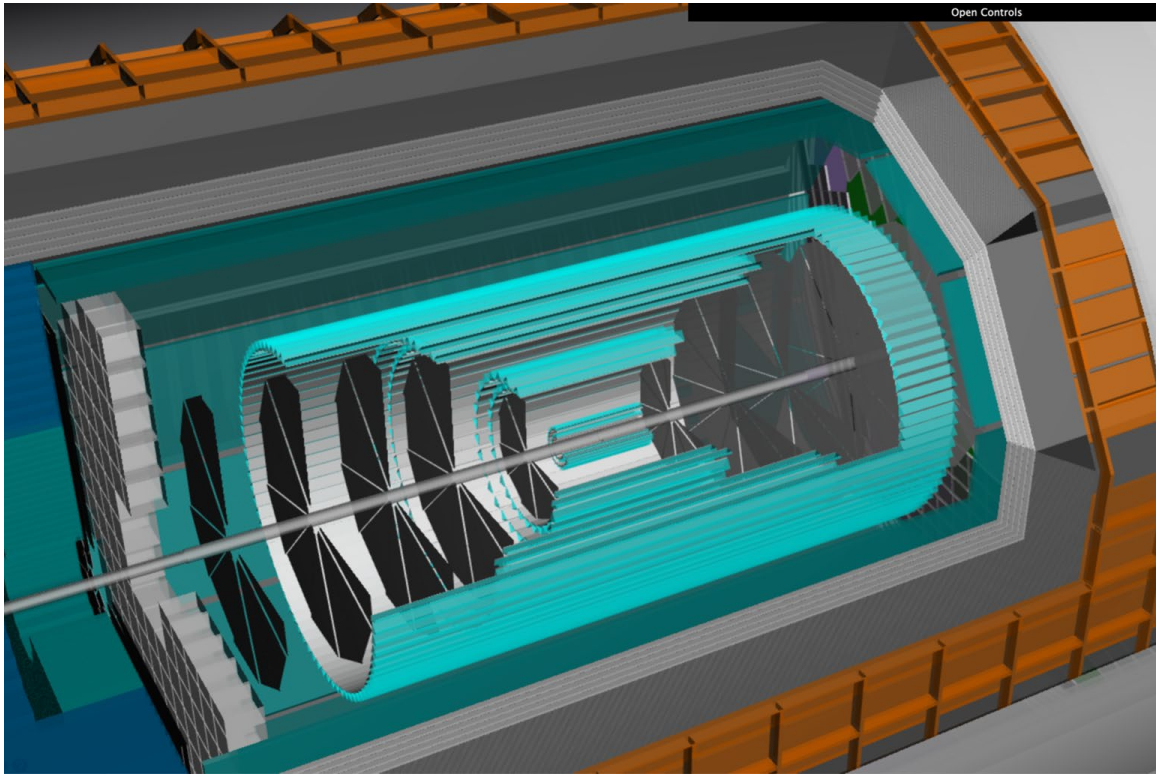
Automated workflows at eicweb

- GitLab server (eicweb.phy.anl.gov)
  - continuous integration
  - dedicated build cluster
  - Runs automatically on each user commit, executing workflows running multiple tests, benchmarks and analysis
- Automated containers
  Both Docker and Singularity images are created nightly or on demand (commit) providing:
  - reproducibility,
  - production level images
  - latest updates for those working locally

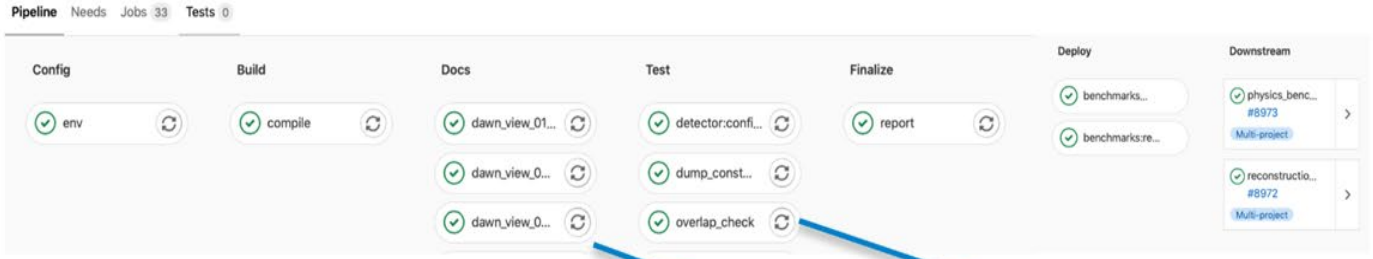


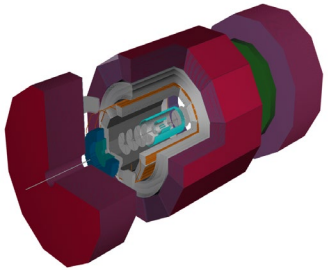
Detailed geometry implementation

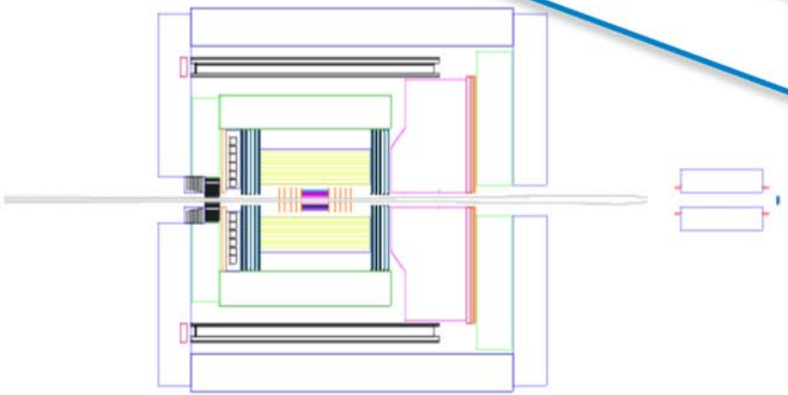
EXAMPLE: Tracking Systems



Benchmarks, documentation, conterization

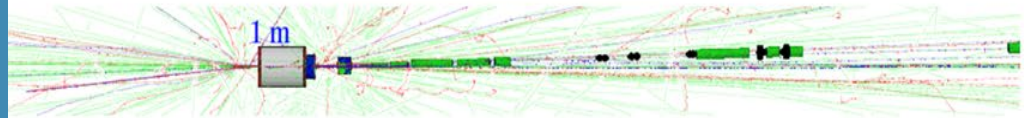


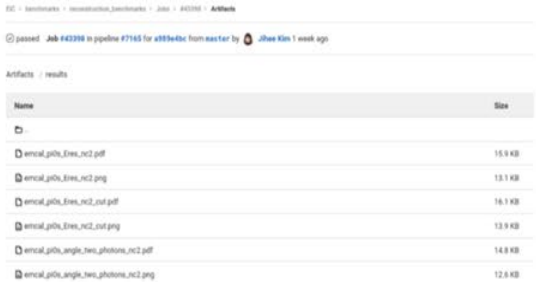




Geometry overlap checks running as part of every merge request

Automatic visualizations for detector geometries, saved as job artifacts (browsable!)





Large scale ATHENA data productions

- Input: HepMC files (mcconv developed for other formats)
- Full simulation with current detector model, all bells and whistles:
  - Typical: ~6 s/event, <500MB RAM RSS, 30 kB to 750 kB output size/event
  - Full ROOT files on S3 under ATHENA/FULL/ (but likely only need reco files)
- Full reconstruction (reconstruction_benchmarks/benchmarks/full):
  - Calorimetry clustering (Ecal, ScFi, Hcal), tracking (up to inner tracker), RICH hits/digi
  - Reco ROOT files on S3 under ATHENA/RECO/ and sci-xrootd.jlab.org
    - Working on jsroot and file browser support on sci-xrootd
- Full simulation: ~weekly repetition; reconstruction: every few days
  - So far simulated 285M events (excluding the CI system!)
- Written to work on any slurm batch system; performed at Compute Canada
  - Trial runs on OSG at the ~2k job scale for single particle events
  - Trial runs on HPC (THETA@ALCF)

First Larger-scale campaign to study detector variations just started

Software Working Group

Software Working Group

Conveners

Andrea Bressan	Andrea.Bressan@ts.infn.it
Markus Diefenthaler	mdiefent@jlab.org
Torre Wenaus	wenaus@gmail.com

Conveners' mailing list: eicug-software-conveners@eicug.org

The Group

The Software Working Group includes more than 100 motivated physicists and software aspects of EIC Software and Computing.

The Software Group mailing list: eicug-software@eicug.org

The Core Team

The core team consists of EICUG collaborators making frequent contributions to the effort.

Jaroslav Adam	jaroslavadam299@gmail.com
Makoto Asai	asai@slac.stanford.edu
Nathan Brei	nbrei@jlab.org
Andrea Bressan	Andrea.Bressan@ts.infn.it
Wouter Deconinck	wdconinc@jlab.org
Markus Diefenthaler	mdiefent@jlab.org
Sergei Furletov	furletov@jlab.org
Yulia Furletova	yulia@jlab.org
Sylvester Johannes Joosten	sjoosten@anl.gov
Kolja Kauder	kkauder@gmail.com
Alexander Kiselev	kisselev@mail.desy.de
Jerome Lauret	jeromel.bnl@gmail.com
David Lawrence	davidl@jlab.org
Chris Pinkenburg	pinkenburg@bnl.gov
Maxim Potekhin	potekhin@bnl.gov
Dmitry Romanov	romanov@jlab.org
Maurizio Ungaro	maureeungaro@gmail.com
Torre Wenaus	wenaus@gmail.com
Jinlong Zhang	jinlong.zhang@stonybrook.edu

The Core Team mailing list: eicug-software-core@eicug.org

Regular SWG Meetings

July 2021

- 28 Jul EICUG Software Working Group Meeting
- 07 Jul EICUG Software Working Group Meeting

June 2021

- 09 Jun EICUG Software Working Group Meeting
- 03 Jun Software Town Hall
- 02 Jun EICUG Software Working Group Meeting

May 2021

- 26 May EICUG Software Working Group Meeting
- 12 May EICUG Software Working Group Meeting
- 05 May EICUG Software Working Group Meeting

April 2021

- 21 Apr EICUG Software Working Group Meeting
- 07 Apr EICUG Software Working Group Meeting

March 2021

- 31 Mar EICUG Software Working Group Meeting
- 25 Mar EICUG Software Working Group Meeting
- 10 Mar EICUG Software Working Group Meeting
- 03 Mar EICUG Software Working Group Meeting

https://indico.bnl.gov/category/301/

Managers

- Andrea Bressan
- Markus Diefenthaler
- Torre Wenaus

February 2021

- 24 Feb Simulation tools for EIC project
- 24 Feb EICUG Software Working Group Meeting
- 10 Feb EICUG Software Working Group Meeting

January 2021

- 27 Jan Expression of Interest for Software

October 2020

- 28 Oct EICUG Software Working Group Meeting
- 21 Oct EICUG Software Working Group Meeting
- 07 Oct EICUG Software Working Group Meeting

September 2020

- 03 Sep Expression of Interest for Software
- 02 Sep EICUG Software Working Group Meeting

Main project for 2021: Project eAST

Project eAST

A Unified Fast/Full Simulation

Makoto Asai+ the eAST developer team of 18+ developers, with members from ATHENA and ECCE

As a result of the survey

The Software Working Group collected information on the community's specific software tools and practices during the Yellow Report Initiative.

Q7. Do you have any comments on your current experience with EIC Software?

There are too many generators and simulation tools used at the moment.

Project eAST: Unify the Simulation Effort

The SWG is launching **eAST**, a **common effort on next-generation simulations**:

- building on the work done in the existing simulations,
- a requirement for the common toolkit is that it integrates existing detector simulations in a modular way.
  - **Community Scenario** Members of ATHENA and ECCE involved. Possible migration to common tool after the current phase, in 2022.

eA Simulation Toolkit

Meetings on Project eAST with wider EIC community

- including many representatives from ATHENA, CORE, ECCE

03/25 Discussion of proposal and possible work plan

03/31 Discussion for anyone who could not attend on 03/25

- Introduction to Project eAST
- Writeup on Project eAST

04/05 Update at SWG meeting

05/05 Update at SWG meeting

05/20 Update at EIC Remote Meeting

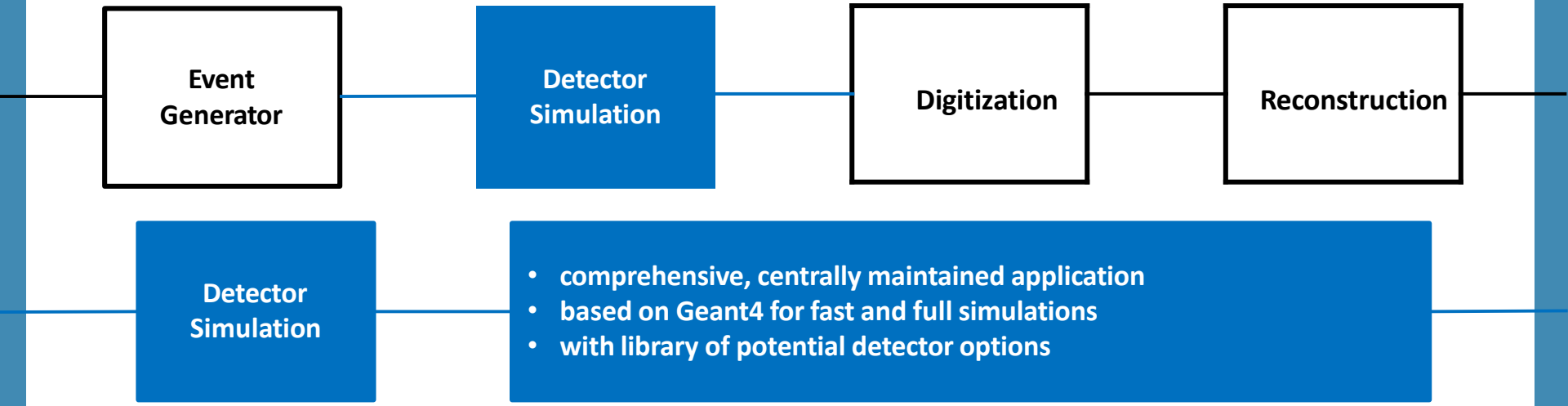
06-07 Regular development team meetings

Project eAST in a nutshell

Detector Simulation

- **comprehensive, centrally maintained application**
- **based on Geant4 for fast and full simulations**
- **with library of potential detector options**

- **Requirements**
  - ability to reuse existing simulation work
  - ease of switching detector options
  - ease of switching between detailed and coarse detector descriptions
  - ease of leveraging new and rapidly evolving
  - technologies,
    - e.g., AI/ML
    - computing hardware, e.g., heterogeneous architectures
- **Project Leader:**
  - Makoto Asai (SLAC), Geant4 project leader and deep technical expert for >20yrs.

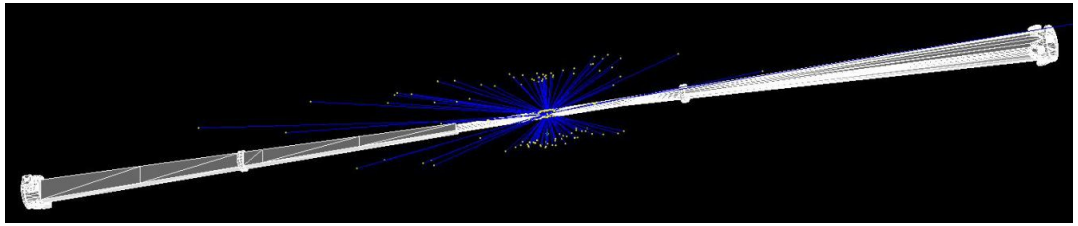


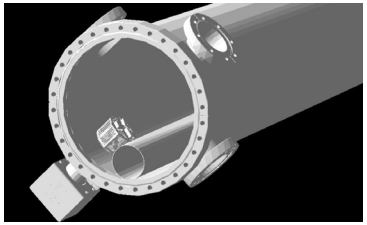
- **Requirements**
  - ability to **reuse existing simulation work**
  - ease of **switching detector options** with comparable levels of detail
  - ease of switching between **detailed and coarse** detector descriptions
  - ease of **leveraging new and rapidly evolving**
    - technologies, e.g., AI/ML
    - computing hardware, e.g., heterogeneous architectures

- **Initial Focus**
  - interface to MCEG output (but no further work on MCEGs)
  - clear separation of detector effects and responses (digitization)
  - common geometry interface between detector simulation tool and reconstruction tools

Project eAST: First Achievements

CAD Interface





EIC Project

- Simulation based (in part) on CAD files provided by EIC project engineering teams, rather than a bottoms-up reliance on constructive solid geometry
- Present “best choice”: MRADSIM – STEP to GDML Converter from INFN

High priority request of community:

Physics List

“Baseline” physics list available https://github.com/eic/east/tree/main/PhysicsList

- It can be used for every EIC detector component as a baseline.
- It needs tuning / optimization / specialization for each detector component.

- It works with Geant4 version 10.7-patch01. Can be integrated in other simulation efforts.
- Basic validation within Geant4 validation suite under preparation. Test beam data needed.

MRADSIM – STEP to GDML Converter

From Asai, Makoto <asai@slac.stanford.edu>

Reply | Reply All | Forward | Archive | Junk | Delete | More

Subject **Re: MRADSIM Converter** 4/7/2021, 7:48 PI

To behcet.alpat@pg.infn.it <behcet.alpat@pg.infn.it>, Luciano Pandola <pandola@lns.infn.it>, behcet alpat <behcet.alpat@iradet: **1 more**

Cc Wouter Deconinck <wdconinc@jlab.org>, jan.bernauer@stonybrook.edu, andrea.bressan@ts.infn.it

Dear Behcet,

We confirmed that the name ambiguity issue has been completely and nicely solved. Thanks a lot for your prompt fix.

We are happy to inform you that, at the same time, volume overlaps disappeared. We believe what happened is the following. There are some volumes in the original CAD STEP file that have the same name but different sizes. Then, because the original MRADSIM did not disambiguate the volume names, converted logical volumes were associated to incorrect solids and caused volume overlaps. Now, since you have addressed the name ambiguity issue, this won't happen any more.

We are now proceeding to trying some more realistic and detailed CAD files. I will keep you updated.

Thank you once again,
Makoto

http://mc-infn.lns.infn.it/?action=Geant4/MRADSIM_Converter

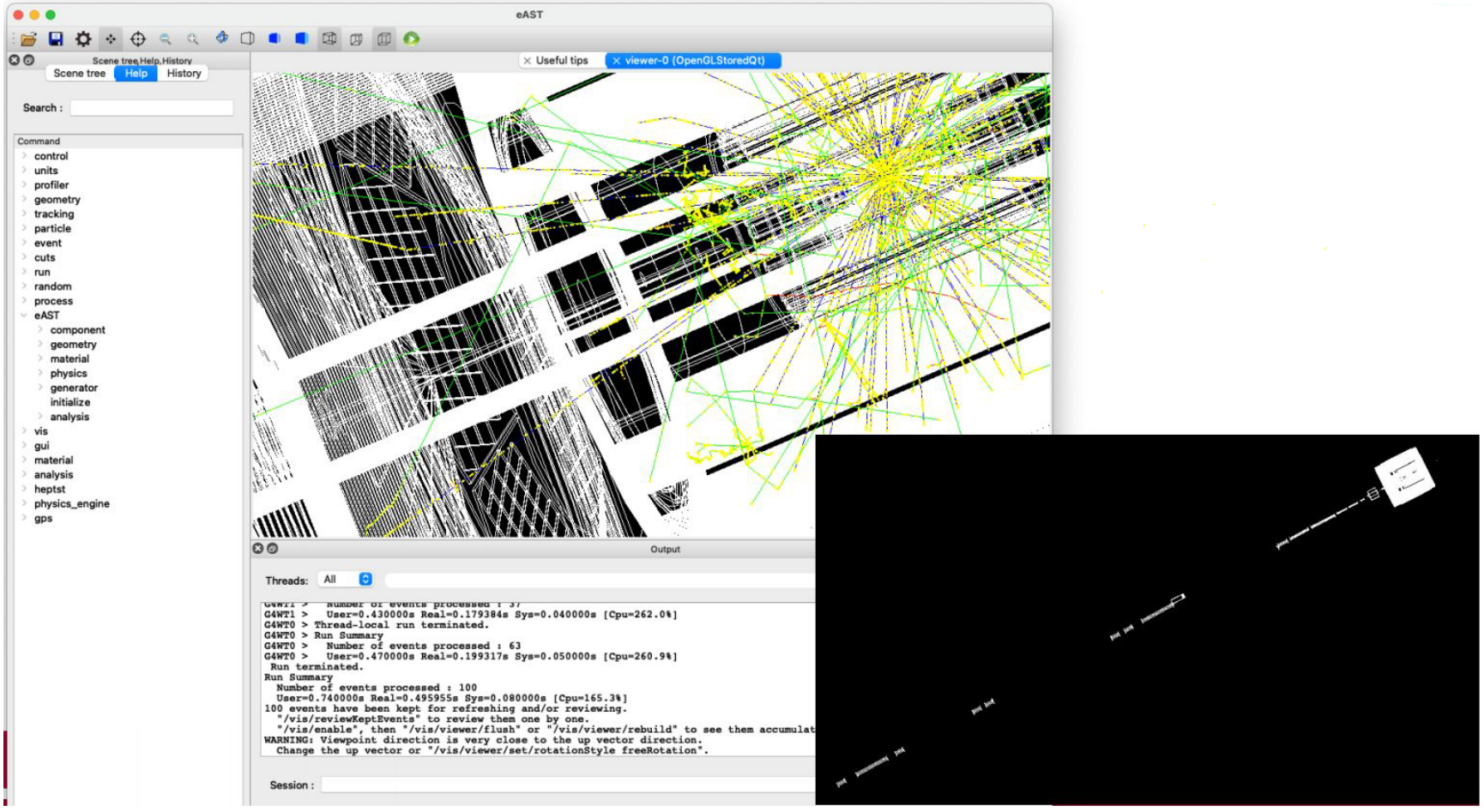
Detector Simulation

- **comprehensive, centrally maintained application**
- **based on Geant4 for fast and full simulations**
- **with library of potential detector options**

- **Deliverables**
  - Create CAD interface to the simulation (a distinct project that occupied June into July).

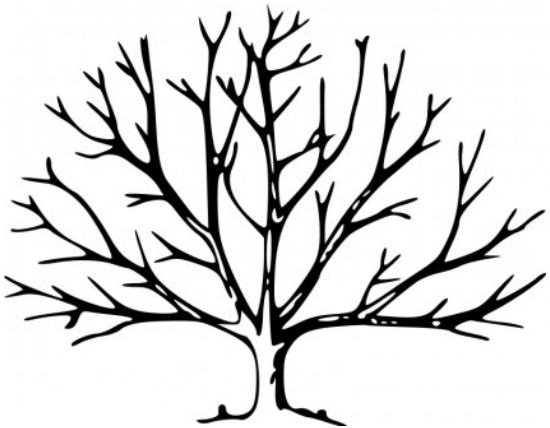
  - Create an initial version of the fast and full simulation tool.
  - Produce a documented prototype of an existing simulation integrated in the new tool.
  - Develop and deliver a documented common physics list for EIC detectors.
  - Deliver a framework extensible to heterogeneous architectures using Geant4's new task-based concurrency and sub-event level parallelism

Athena geometry/event in eAST



The EIC Computing Coordination Group (CCG)

- EIC Computing (initially)
- Evolution: the EIC Computing Coordination Group (CCG)
  - Composition, meetings and communication
  - Computing “resources”
- Requirements and resources



EIC Computing (initially)

- [Computing Report May 2021]
- EIC Computing Organization = a representative from BNL and Jlab
  - Both labs charged a representative to be their POC: Jerome Lauret (BNL), Graham Heyes (JLab)
  - Synergies in meeting organization - friendly exchange of knowledge: Computing round tables, Technology Meetings, initially EIC computing discussions (bi-weekly) at BNL, participation in the EIC SWG
  - Weekly meetings to coordinate progress and ideas, share work and services
- Discussion with the EIC UG in early March 2021- concerns:
  - WLCG model has benefited the LHC - need a structure to build a common infrastructure ; connecting to a common fabric rendered easy - what of the EIC?
  - Need to coordinate and monitor contributions and usage from the collaboration(s)
  - Need to develop and own “a” computing model (shared? Pledged? Hybrid? Distributed? federated?)
  - Computing and SWG working together - possible representation from the SWG (Andrea B.)
- (Scientific) Computing Infrastructure and Software dev at National Labs are separated but interconnected via S&C projects of the experiments - EIC/UG recognizes this

Evolution: the EIC CCG
The Charge and Composition

- The EIC UG addressed forming the ECG at the end of May
- Charges
  - Coordination of *resources* among EIC computing efforts
  - Keep a record of required and available resources as well as their usage
  - Access point for institutions that intend to contribute computing resources

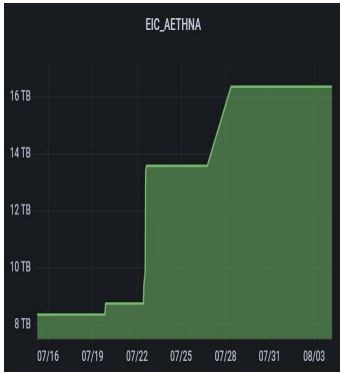
- Membership
  - Contact persons/liaisons for EIC computing resources from both host labs.
  - The liaisons of the two host labs will be the contact persons for the group
  - Representatives of computing groups of the three *proto-collaborations*, one or two each, to be nominated by *ATHENA, CORE and ECCE.*
  - Liaison to EICUG Software Working Group. Andrea Bressan agreed to take on this position
- Regular meetings of the CCG

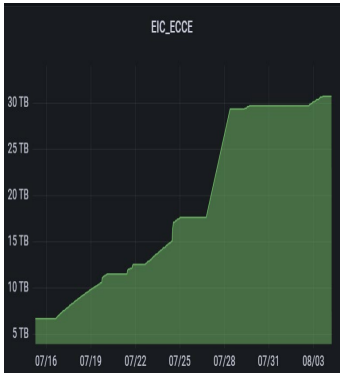
Current composition, meetings, …

- Composition
  - Graham Heyes and Jerome Lauret as host lab POC
  - Andrea Bressan confirmed to serve on the CCG
  - Representation from the proto-collaborations. ATHENA: Wouter Deconinck ; CORE: TBD ; ECCE: Cameron Dean
- Meetings & communication
  - Weekly meeting with the CCG core members
    … but meetings with rep from ATHENA and ECCE only started recently
  - Bi-weekly with ATHENA/ECCE proto-collaborations to help and sustain the proposal effort
  - Engagement with the respective laboratories and the OSG to provide the resources needed
  - Mailing list: eic-comp-l (12) but most used is Mattermost channel "EIC Computing" (33)
- The ECCG is discussing starting a monthly common meeting (TBD)
  - Dedicated meetings with the proto-collaborations still required / desired

Requirements and resources

- Resource requirements - feedback from ATHENA and ECCE
  - We assumed ½ CPU and 30% less storage for CORE & contingency
  - CPUs requested: 7k cores average, 11k max
  - CPUs provided: 2k slots from each JLab / BNL with possible opportunistic use of a larger pool and a commitment of +4k slots from the OSG => a 8k+ slots baseline was covered
  - Storage requested: numbers indicated 1.5 PBytes of storage x 2 (assumed 1 PBytes for CORE + contingency margin) i.e. 4 PBytes needed total
  - Storage provided: 1 PBytes from each lab (different technologies for access) +0.5 PBytes at BNL later. 2.5 PBytes (but …)





- Resource available summarized here (Web site planned later)
- S3 storage at BNL (monitoring example) and Xrootd at JLab + user local space of several 100 TB worth - BOTH storage are accessible with standard ROOT access for interactive analysis
- S3 is accessible in WRITE mode from a handful of accounts
- OSG submit host available at both labs - S3 writes integrated to workflows as well as CI process ...

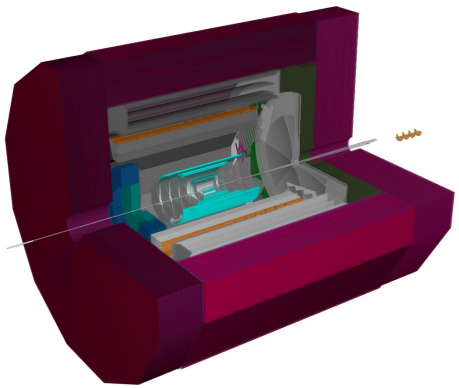
CNAF

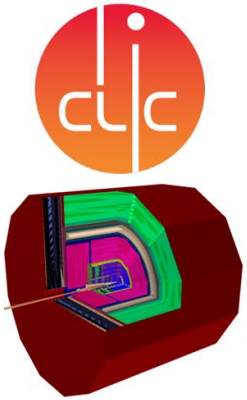
- PLEDGE 2021:
  - 250 HS06
  - 10 TB servizio
  - VO OSG in definizione (marian.zvada@cern.ch new OSG link)
  - Work in progress: configurazione della VO sui CE

- RICHIESTE 2022
  - 1000 HS06
  - 100 TB di spazio disco
  - In previsione delle attività per il TDR

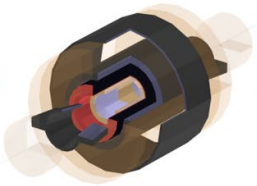
SPARES

The DD4hep community

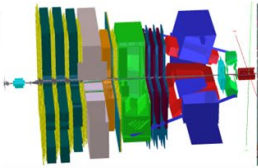




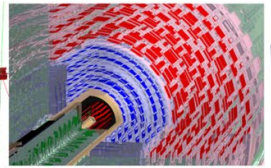
Production



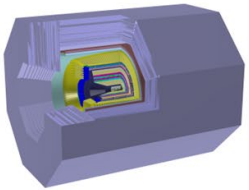
Production



Production
Run 3



Under
investigation
Run 3

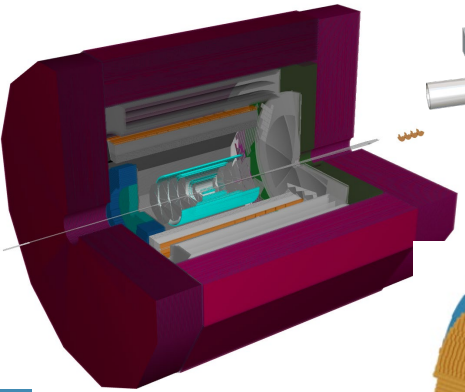
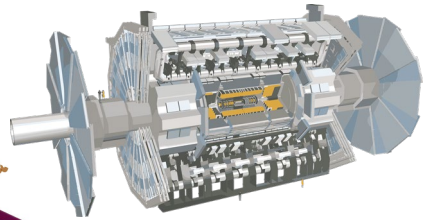
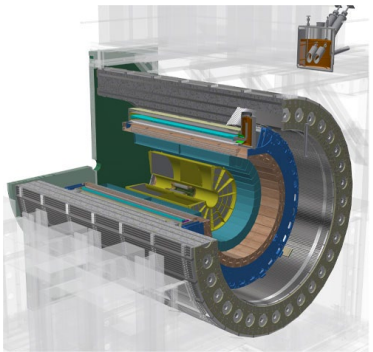
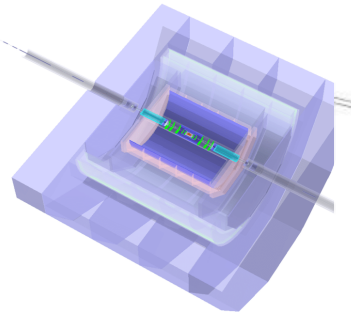
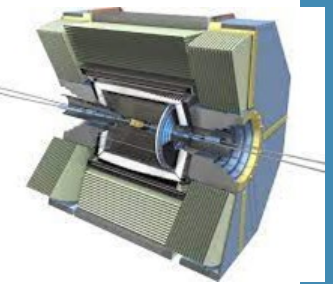
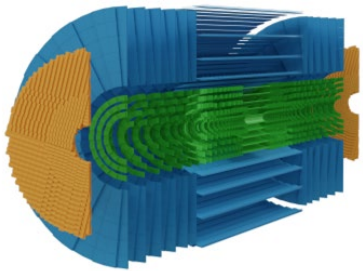
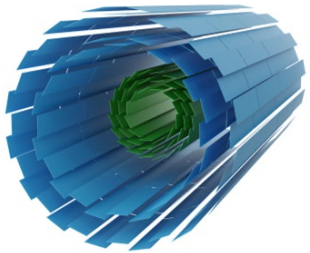
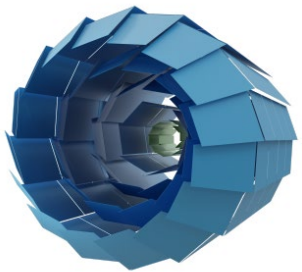


Production

Super Charm
Tau Factories

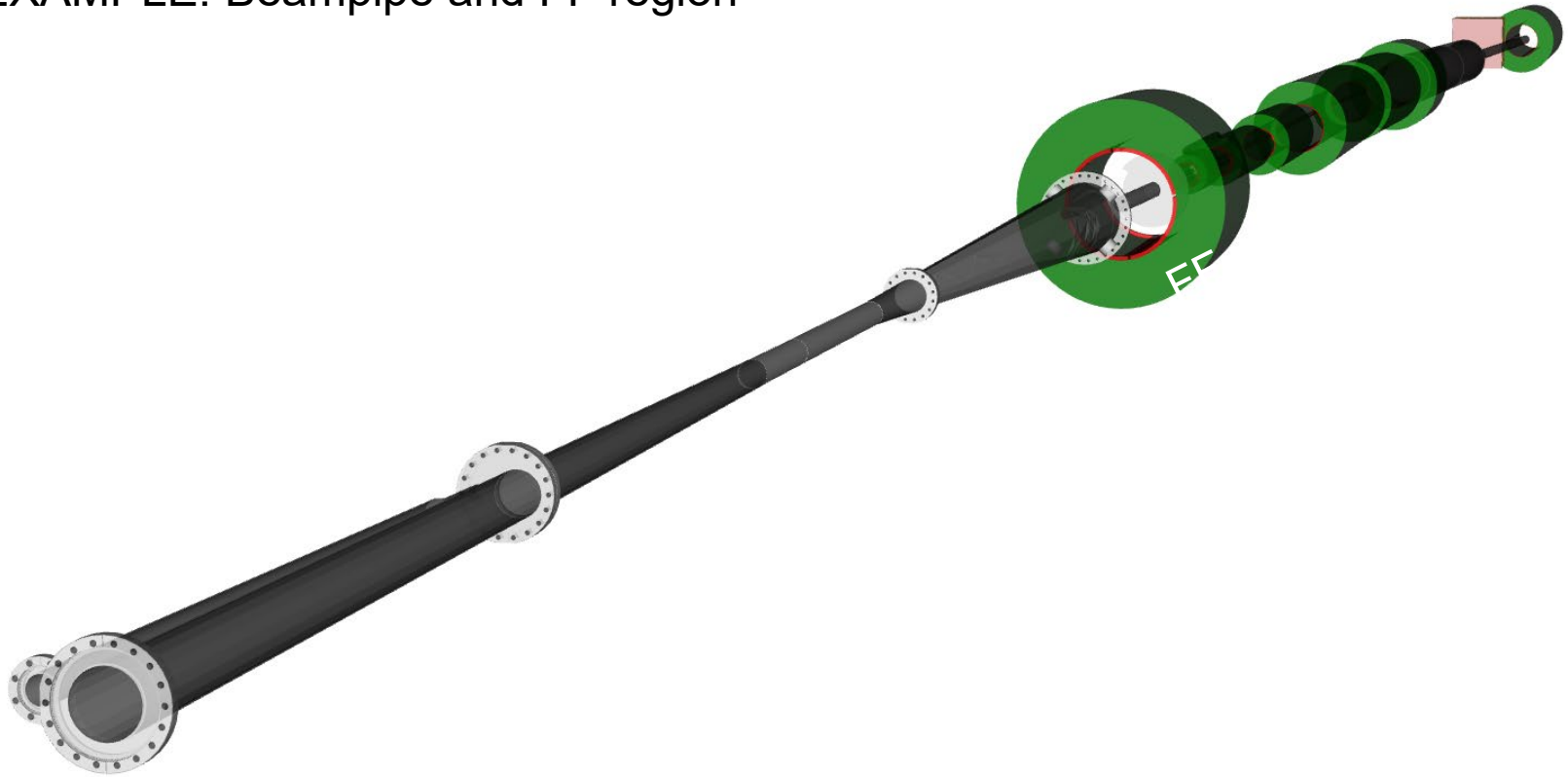
"framework for providing a complete solution for full detector description (geometry, materials, visualization, readout, alignment, calibration, etc.)"

ACTS is rapidly becoming the standard across particle and nuclear physics

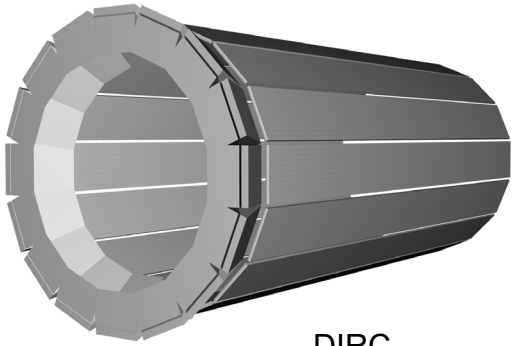
Detailed geometry implementation

EXAMPLE: Beampipe and FF region

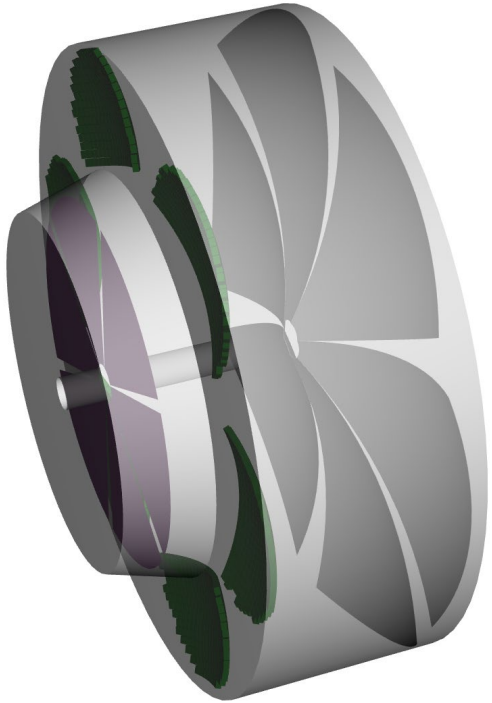


Detailed geometry implementation

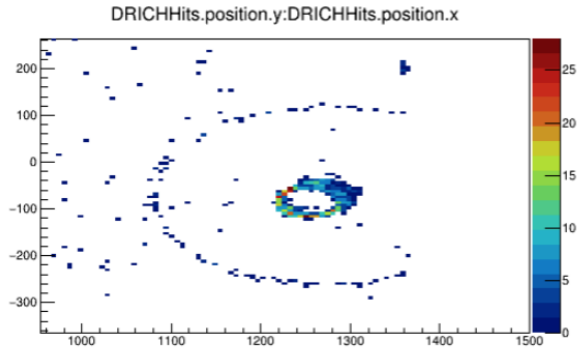
EXAMPLE: PID Systems



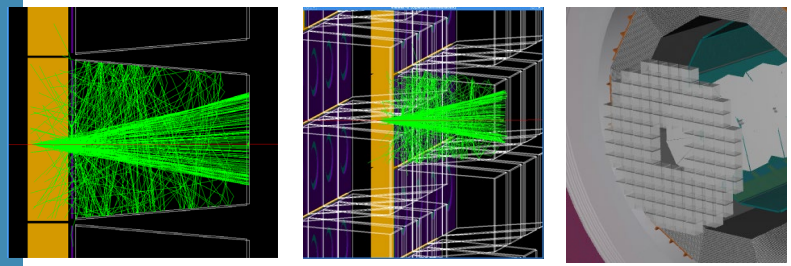
DIRC



dRICH



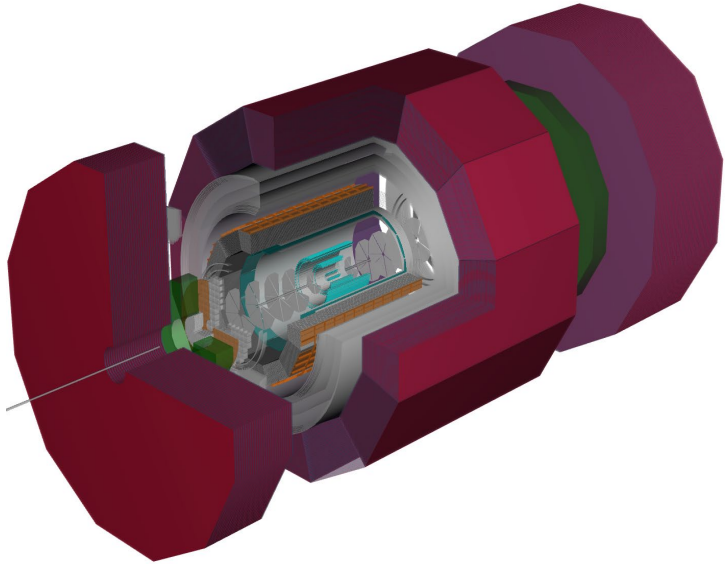
mRICH



EIC AI/ML in ATHENA

- Current use of AI/ML
  - ✓ e/π PID with 3D shower profiles from imaging calorimeter in center barrel region.
- Near-term anticipated use:
  - 🚧 ACTS: Track finding
  - 🚧 PID: Pattern recognition in RICH, DIRC
  - ✓ Calorimetry clustering (2D, 2+1D and 3D clustering)
  - 🚧 DNN-based fast simulation
  - 🚧 DNN-based detector optimization (Bayesian optimization)
  - 🚧 DNN-based reconstruction
- Implications on computing infrastructure:
  - 🚧 Many exascale GPU accelerators, but lack of support in current software tools limited by IO/memory bandwidth

✓ = working, 🚧 = in progress, ✕ = planned



ATHENA central detector (longitudinally expanded)