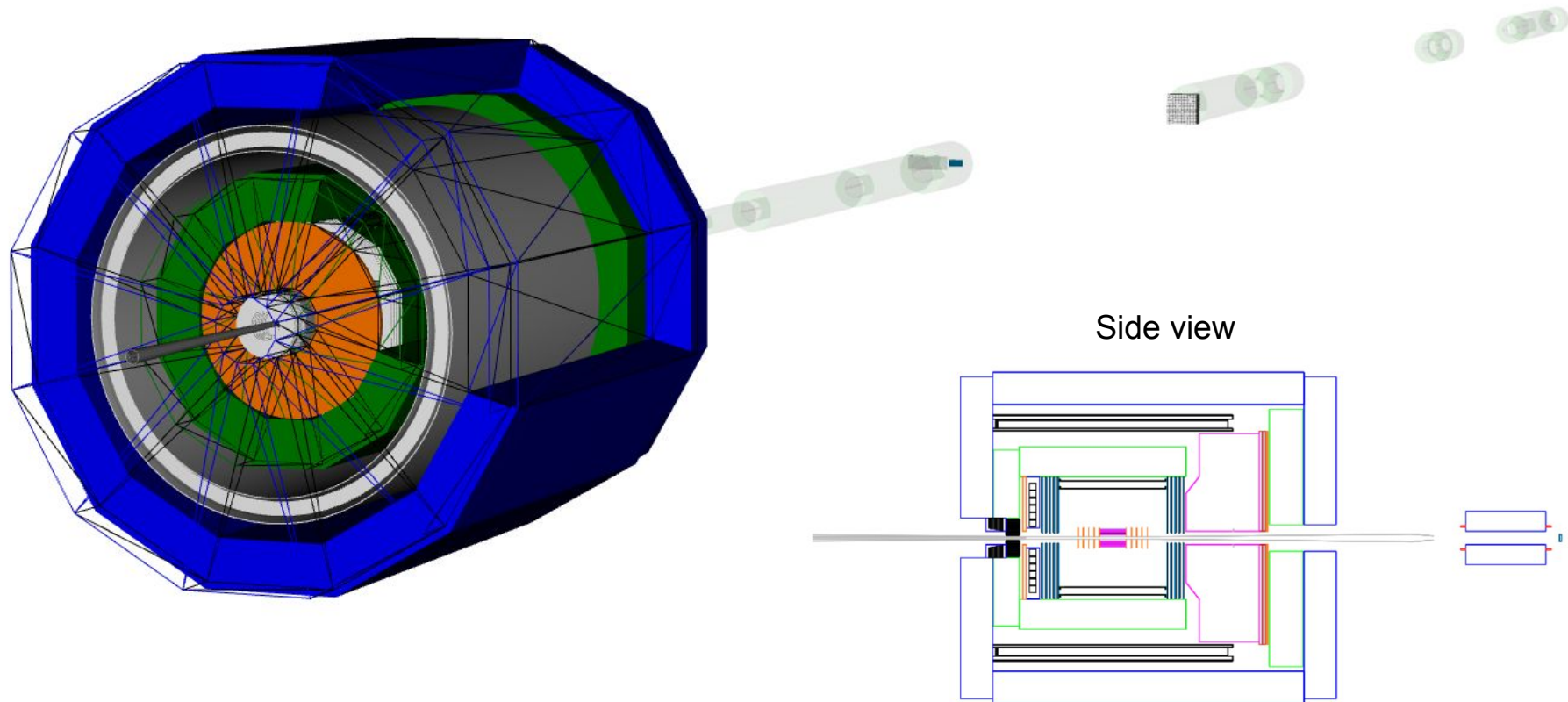# EIC@IP6 Software
2021-05-06 meeting

Andrea Bressan (University of Trieste and INFN) ,
Dmitry Romanov (Jefferson lab) ,
Sylvester Joosten (Argonne National Laboratory) ,
Whitney Armstrong (Argonne National Laboratory) ,
Wouter Deconinck (The University of Manitoba)

Software Updates & Milestones

Reference detector implementation well underway

Side view

Geometry status and next steps

- Reference detector implementation almost complete (misses DIRC)
- Tracking system:
  - progressively add detail and options, start with tracking detectors to enable work on reconstruction.
  - Implemented MPGD detectors (including some infrastructure) obtained from eRD6 group, geometry will need tweaking
  - Together with tracking WG: identify and implement all-silicon variant
- Next steps:
  - In concert with detector WGs, add realistic candidates for other subsystems. Good places to start would be calorimetry, DIRC, dRICH, …
  - Add more detail in far forward/far backward regions
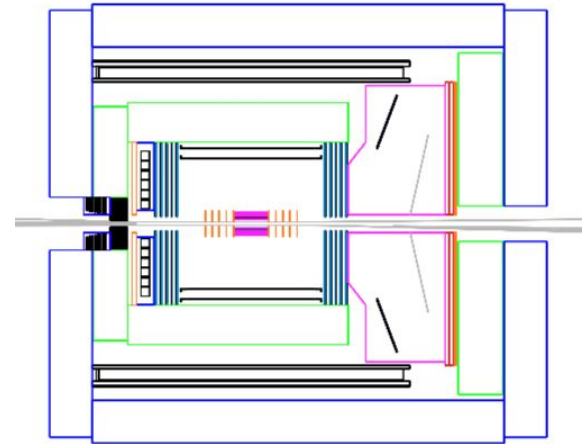
Side view of current central detector implementation

Software development

Digitization & Reconstruction ([Project Juggler](#))

- In good working state (using Gaudi v34)
- Next steps:
  - Revisit digitization of tracking detectors, and optimize tracking performance
  - Update to Gaudi v35 (introduces many breaking changes)

[EIC@IP6 Container](#)

- Ongoing: Refactor existing container into eic_base image for CI usage, and "dressed" eic-image for local development and deployment
- Next: Improved deployment mechanism and utility scripts

[Automized benchmarks](#)

- Ongoing: Refactor to combine common code between detector, reconstruction and physics benchmarks into a central benchmark repository
- Next: Implement detector & reconstruction benchmarks for tracking system
- Document benchmark workflow (see later)

Documentation and tutorials

Central documentation

https://eic.phy.anl.gov/ip6/   (mirror ip6soft.readthedocs.io)

Central documentation website. Goal: once a user is there, she/he should be able to find/navigate to the information needed

Documentation should include 4 different concepts:

1. High level overview
2. Concepts explained
3. Reference documentation (API's, data format description)
4. **How-to, tutorials, workflows**

Imminent actions:

1. How to run fast simulation (if generator output is given)
2. How to run full simulation (if generator output is given)
3. How to add/modify your detector
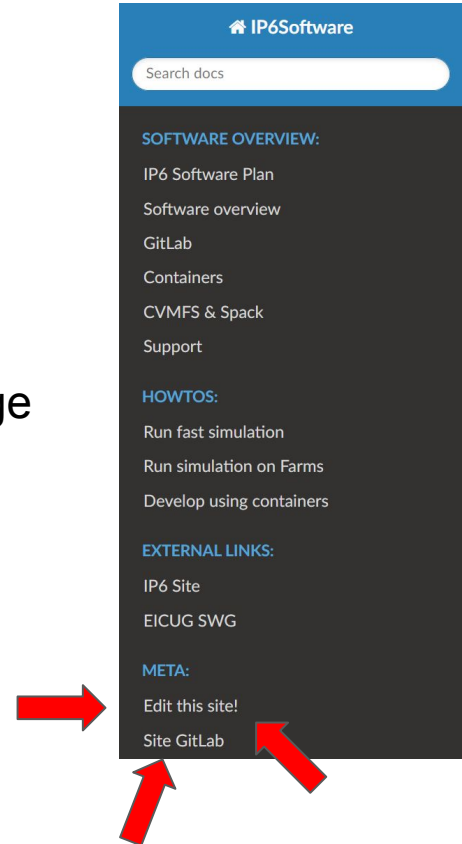4. How to work with CI
5. Generators/mc-actions

Tutorials for each of the topics to be organized!

What we need/missing nere?

Contributing

- This site is open to everyone.

- Users are encouraged to contribute to this documentation!

- The pages are automatically updated on master branch change

- [IP6 SWG documentation GitLab repo](#)

- [How to edit website instruction](#)

If you would like to add/edit the information on this site - you are welcome to do this!

Support

- **Slack chat => #ip6-software-wg**

- Mailing list eic-ip6-software-l (not actively used yet)

- GitLab issue tracking (everything internally is planned and distributed through issues and agile/project boards). Gitlab repos directory

- Reference Detector issue tracking

- Should we create a central issue tracker where anyone can leave anything and it will be processed and correctly addressed by SWG?

EIC@IP6 Computing

Raw estimation of needed resources

Requirements

- Up to 250M events of full simulation/reconstruction
- Based on Andrea's experience with COMPASS, this should take <40M core hours (very conservative upper limit).
- Data size: I feel that ~500TB should be more than plenty.

Estimation assumptions (our needs):

1. Isolated benchmark figures for each subsystem (and probably at least 2 options/subsystem early on)

2. Full simulation/reconstruction results for each serious subsystem candidate

3. Physics validation results with at least one fully integrated detector candidate

~10 key physics processes x 1M events x 2 tries —> 20M

~6 subsystem WGs x 20 x 10k events —> 1.2M

~25M events from generators, and probably < 250M events with full detector simulation.

It would be good to start determining which plots we exactly want for the proposal and then work backwards.

JLab and BNL commuting group meetings

Current resources:

1. At least 2000 & 2000 job slots from JLab and BNL => 4000 on OSG
    (shared among all collaborations)
2. Workflow for institutions to add their resources on OSG for collaborations
3. Hard to quickly provide large dedicated space for every collaboration.
    (EIC quotas for disk space for BNL and JLab)
4. Both labs looking at Rucio, but it won't appear soon and we have to push for it
5. A committee to control fair share of the resources between collaboration is to be formed
6. Labs are working on federated resources (single login, shared resources) but it won't appear soon and will be applied incrementally

Computing action items and discussion

1. **Assign a person to be responsible/convener for computing@EIC@IP6**
2. Prepare official request for Rucio (will be used as a leverage for both labs)
3. Provide detailed request on storage requirements.
   How much storage we need that is:
   a. Volatile (not shared)
   b. Archived
   c. Replicated across labs
   d. Globally accessible
4. Need to better understand our workflows to estimate this, more accurate CPU estimation, etc.
   a. Most CPU will go towards running large-scale detector simulations. These need to be centrally coordinated.
5. Define storage policy for large files (in particular from full simulation), as full Rucio rollout unlikely within the detector proposal timeline.
6. Question other institutions/universities of possible available resources now

Current CI resources and path forward

CI workflow centered around [eicweb](#):

1. Fully featured GitLab install that we own/control
2. Backed by dedicated build cluster (~500 threads, 100TB storage)
3. Used for all automatic builds (containers, etc), and small scale benchmarks
4. For large scale benchmarks we need to identify a path towards integration with compute resources at BNL and JLab, and HPC resources at other institutions (ANL, INFN, LBL, ...)
5. Data storage and policy will likely be biggest hurdle.
   a. May need to standardize around S3 to leverage storage available at BNL (~1PB)
   b. Could also consider workflows that include Globus

Working groups interaction

PWG ⟺ Software WG ⟺ DWG interaction

Our vision so far (very open to discussion):

1. Every group delegates 1 person to work with software groups

2. For PWG imminent TODO:
   a. What benchmarks/plots/studies should we have for report
   b. Workflow to integrate analysis in CI pipeline

3. For DWG imminent TODO:
   a. Work on detector implementation details
   b. How-to implement/modify a detector in DD4Hep

4. For all - gather information
   a. What are their plans and vision of now
   b. If additional resources/help needed for that