ECCE Computing

Cameron Dean
Los Alamos National Laboratory

08/04/2021

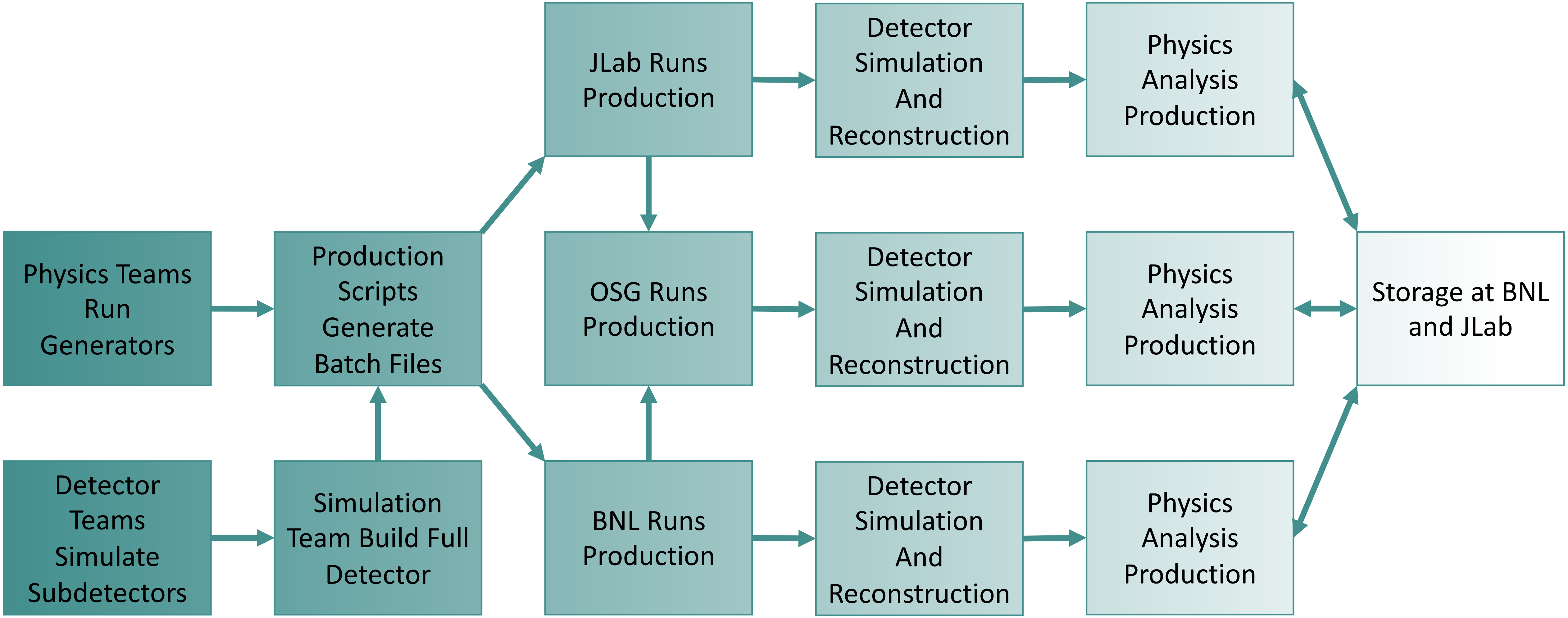
EICUG Summer Meeting

Overview

- ECCE computing is entangled in several areas of our consortium
  - Detector simulation
  - Data production and storage
  - User interaction
- AI/ML groups optimize design and physics potential
- 150M events already on disk

ECCE overview talk with more information

Production Workflow



Detector Simulation & Reconstruction

Generator-level events produced in LUND format

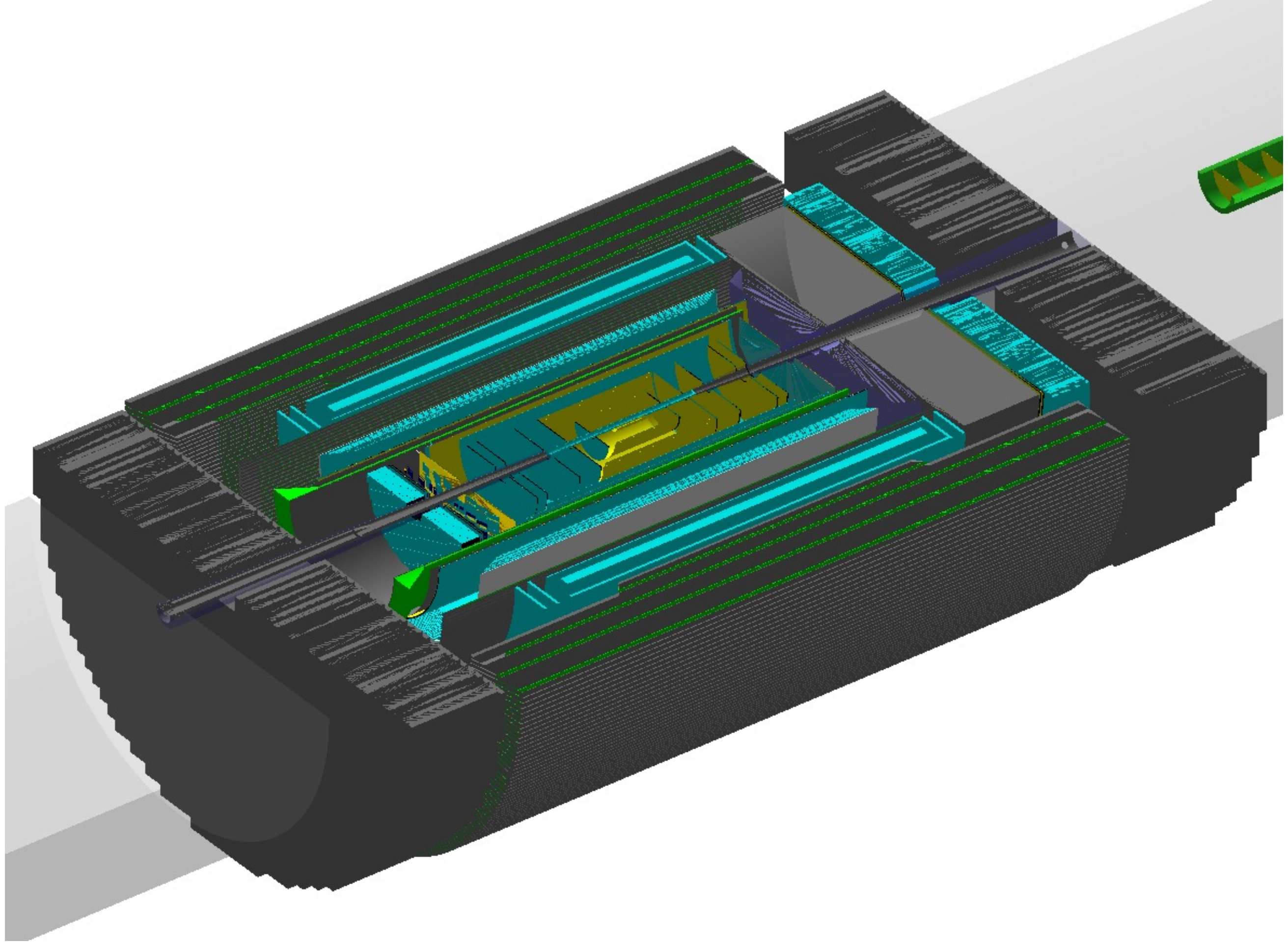
Fed through EIC-smear

Internal generator (pythia8)

Beam crossing and divergence

Detector response (plus digitization/PID/clustering)

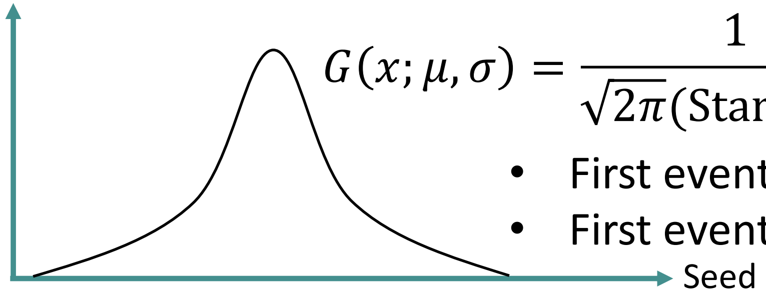
GenFit tracking



ECCE detector talks with more information:
Calorimetry, tracking, PID, far forward/backward

(Re-)Producing Physics Simulations

- Each collision sample (generator and beam conditions) produces a dataset file and associated ROOT files (direct analysis)
  - DST has truth info., hits, clusters, tracks, jet reco. and more
  - Typical file is 1k → 5k events
- Reproducibility is very important
  - Must fix: **software stack**, **detector simulation** and **seed**
- Seed solution: Use RooUnblindPar
  - Unique for each DST and reproducible anywhere with no user interaction!



$$G(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}(\text{Start}+1)}\exp\left(-\frac{(x-\text{nEvents})^2}{2(\text{Start}+1)^2}\right)$$

- First event and nEvents define width and mean
- First event, nEvents and file string (plus scale) selects seed from Gaussian

Metadata logs

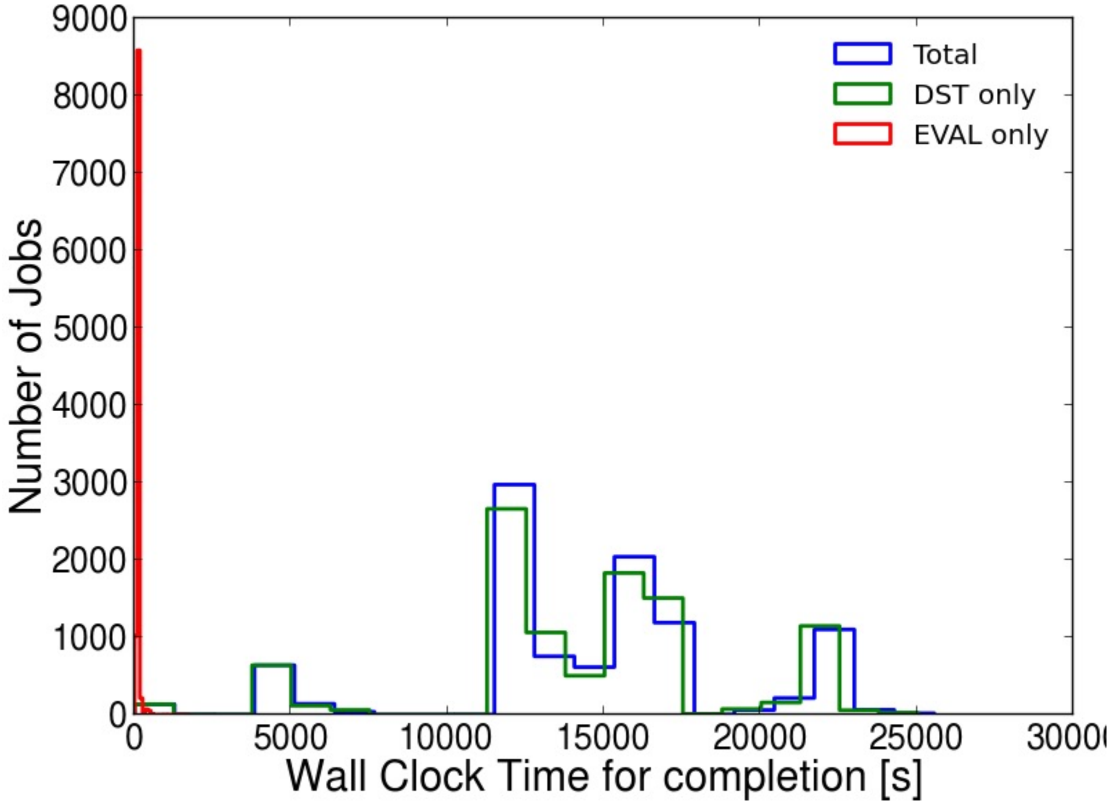
- Each production job automatically writes metadata to help debug issues

```
====== Your production details ======
Production started: 2021/07/25 17:10
Production site: BNL
Production Host: spool0680.sdcc.bnl.gov
ECCE build: prop.2
ECCE macros branch: production
ECCE macros hash: c131177
PWG: SIDIS
Generator: pythia6
Collision type: ep-10x100
Input file: /gpfs02/eic/DATA/YR_SIDIS/ep_10x100/ep_noradcor.10x100_run001.root
Output file: DST_SIDIS_pythia6_ep-10x100_000_0000000_05000.root
Output dir: /gpfs/mnt/gpfs02/eic/DATA/ECCE_Productions/MC/prop.2/c131177/SIDIS/pythia6/ep-10x100
Number of events: 5000
Skip: 0
=====================================
Seeds:
1322570549 (plus more)
md5sum:
01da8efd4555739dfa18fd96ee5b6a36
```

Fully and uniquely defines seed

Production Statistics and Notes

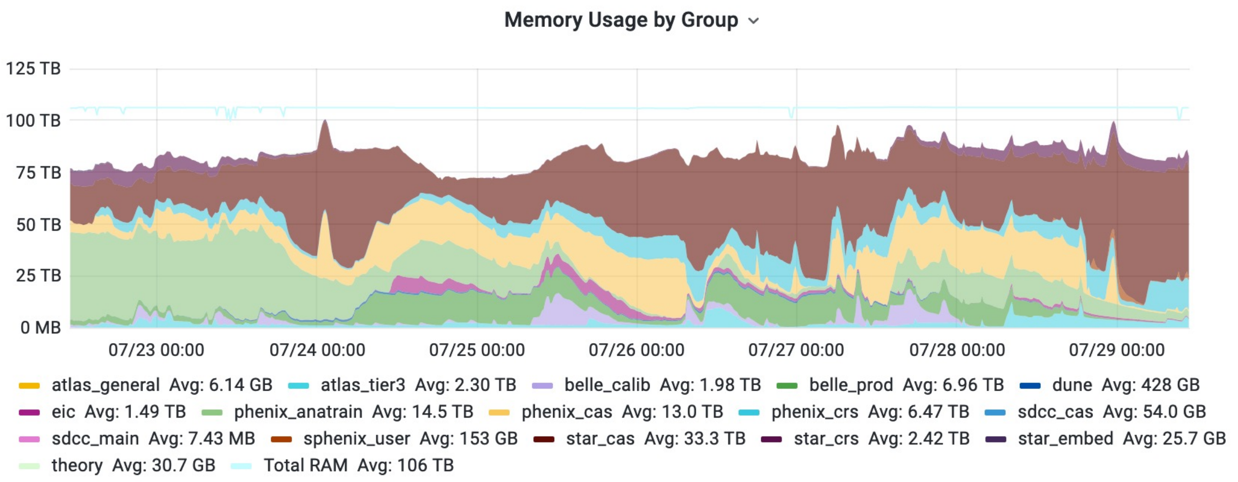
Planned simulation campaigns	2
Predicted events per campaign	120M - 160M
Typical event size	200kB
Typical event generation time	7s
Total storage per campaign*	30 TB
Typical job memory size	< 1.5 GB
Current events simulated in campaign one	101.3M

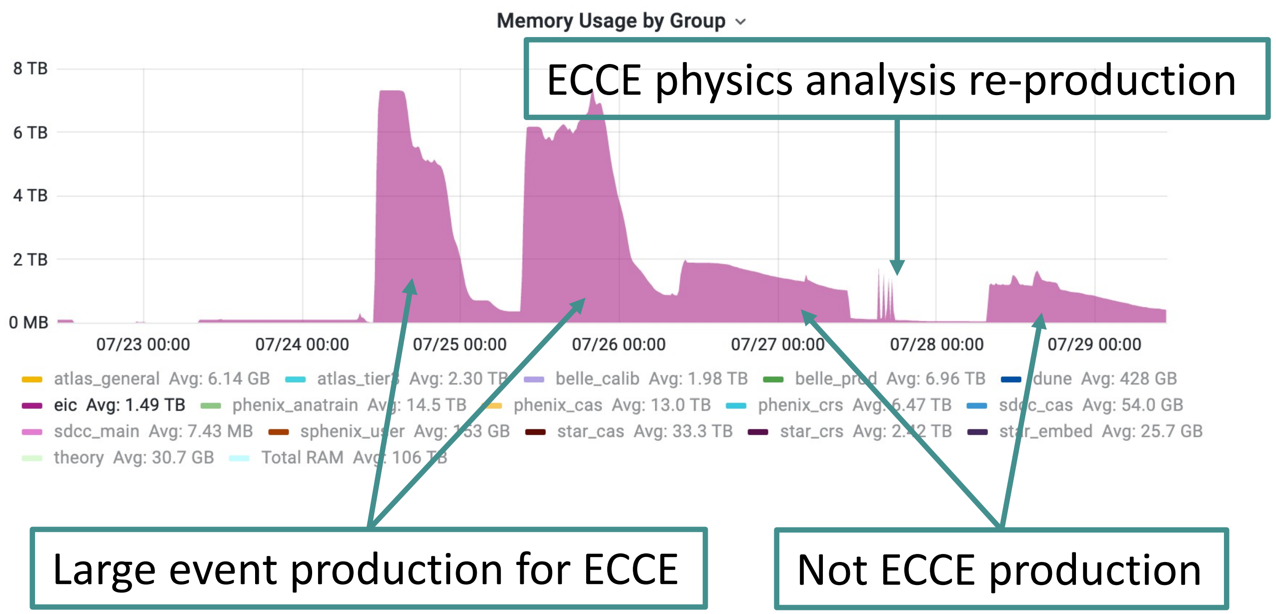


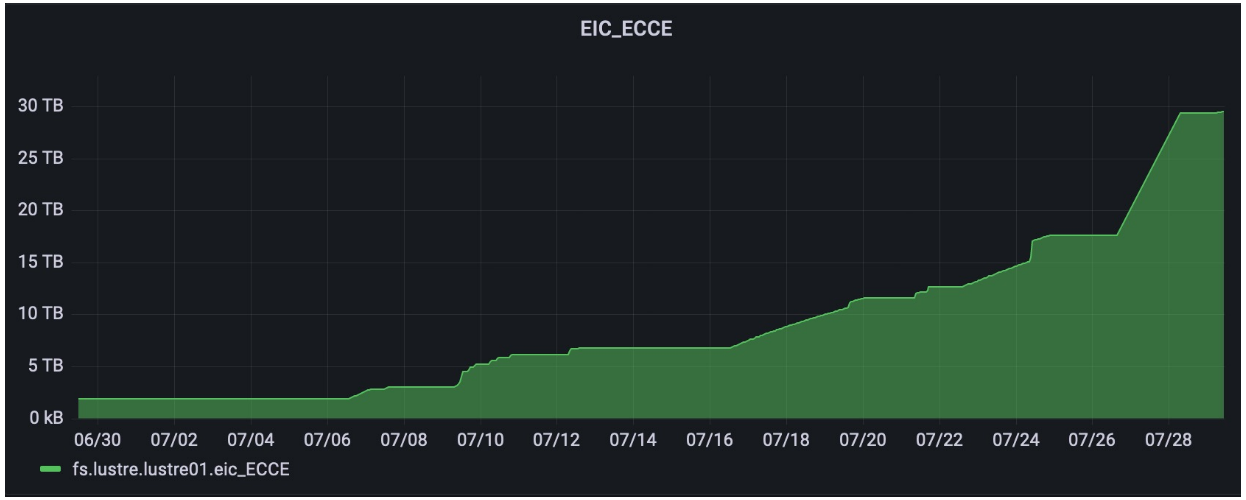
- > 90% job success rate (OSG > 98% success rate)
- Separated sim./reco. from physics production
- Automatic revisions of physics analysis production (compare updates, add features etc)
- To-do: add job monitoring/resubmission (in beta-testing)

*Undecided if first campaign raw data will be kept after second campaign

BNL Usage



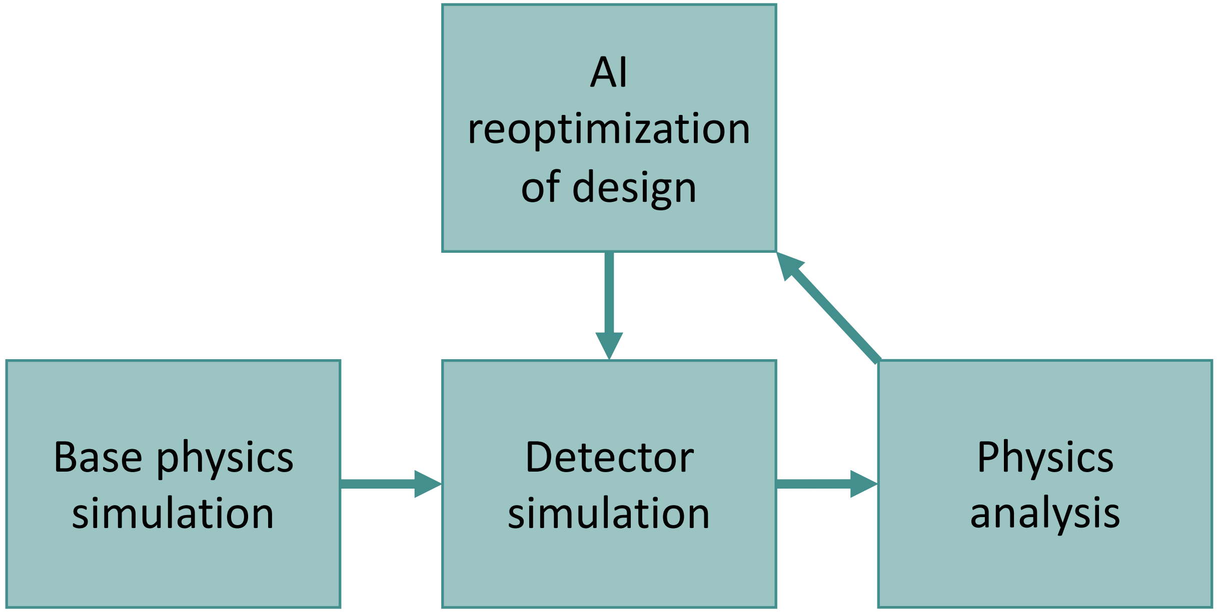




- EIC Condor use is small but noticeable
  - Note, no distinction between ATHENA, CORE and ECCE on condor
- S3 storage is creeping up for ECCE
- 150M events in storage so far

Co-design at ECCE

- What is co-design of particle detectors?
  - Machine learning to optimize layout, material and performance
- Detector design is not a 1D problem:
  - Physics goals met with reduced materials – reduced cost
  - New technology needed to reach targets – increased cost



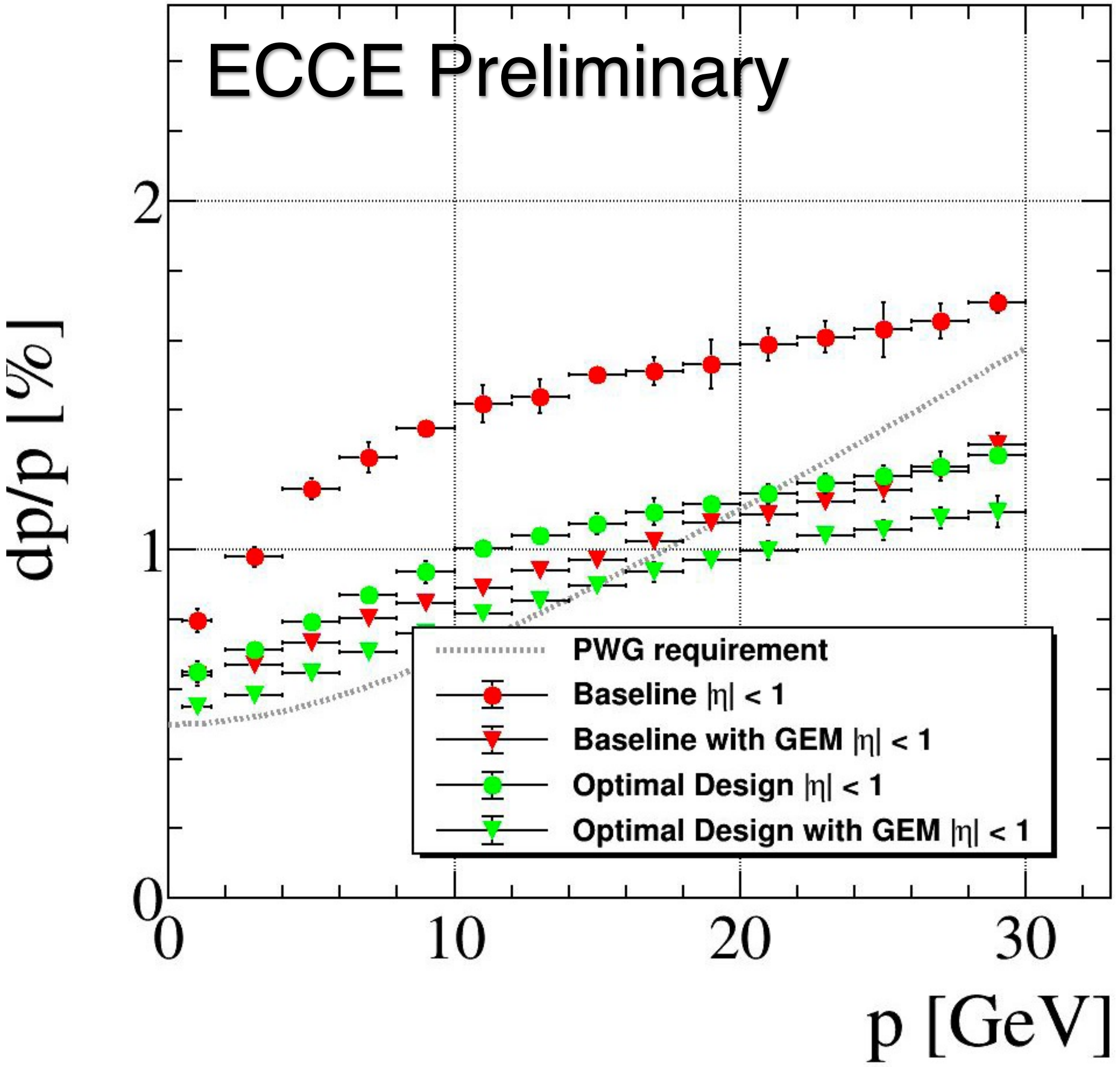
Tomorrows talk on co-design with more information

Co-design example: Tracking

- Simultaneously optimize all trackers
- Optimise for Kalman filter efficiency, DCA, p and ϕ resolutions
- Constraints:

1. Outer barrel radius < 51 cm
2. Vertex layer radius < 15 cm
3. Furthest disk position < 125 cm in z

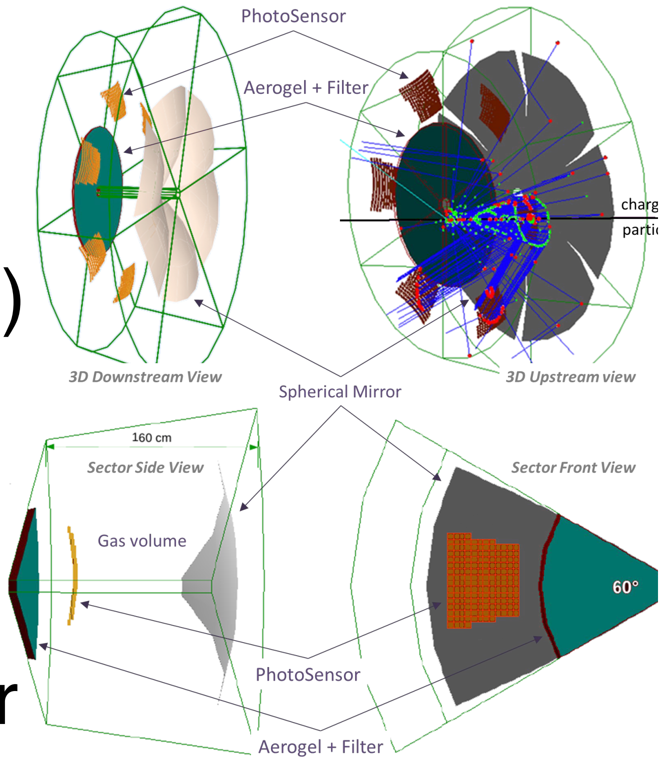
- Optimizing placement gives larger change than additional detectors



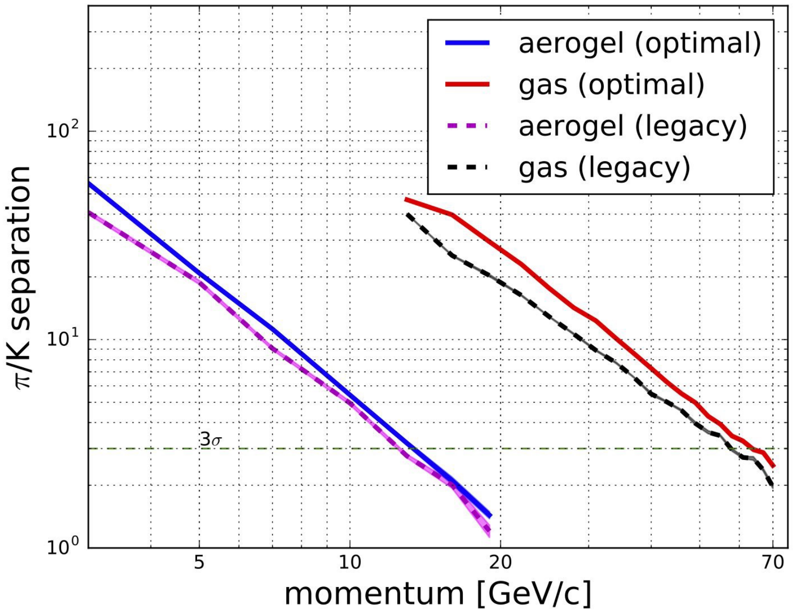
C. Fanelli and K. Suresh

Co-design example: PID

- Optimisation of mRICH:
1. Spherical Mirror radius
2. Mirror position (r, z)
3. PhotoSensor position (x, y, z)
4. Aerogel Refractive Index
5. Aerogel Thickness
- Optimised design reduces Cherenkov saturation at higher momenta



JINST 15 P05009 (2020)



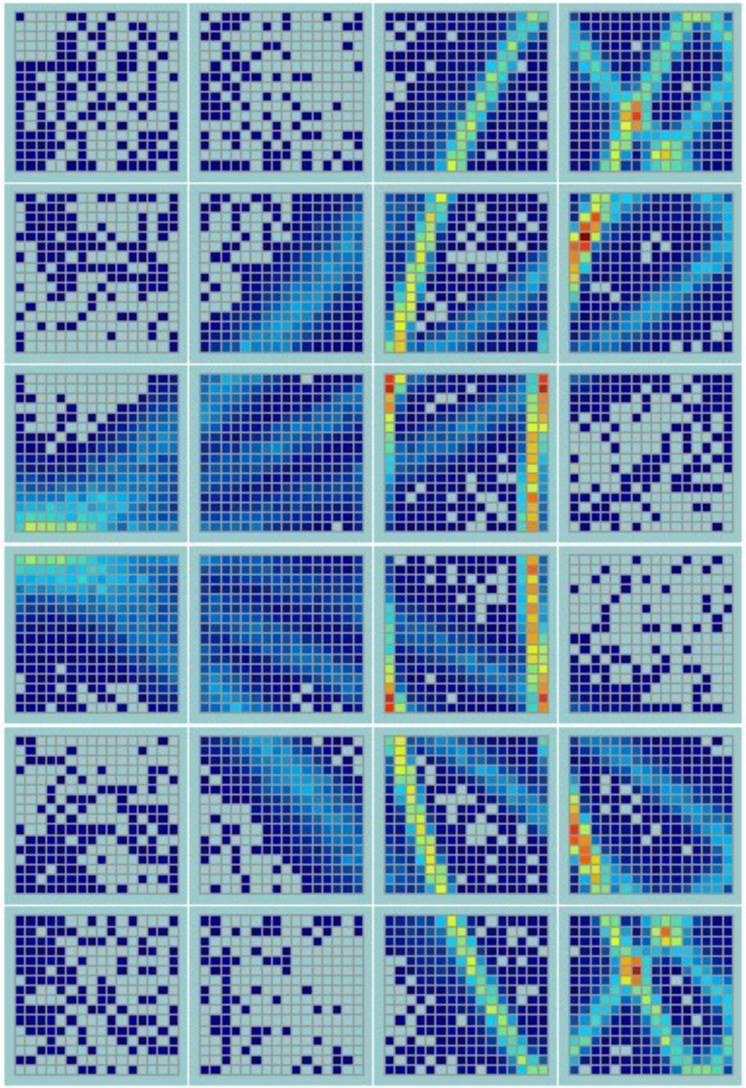
Co-design example: PID

- Current hpDIRC design: reused BaBar DIRC bars
  - Performs beyond required 6 GeV/c!
- Fast sim parametrization based on detailed G4 simulation
- Full reconstruction in Fun4All almost ready

- Next steps:

1. Cost/performance optimization
2. Validation of performance with magnetic field

- Optimisation performed using scikit-learn
- Right – $6\ \mathrm{GeV/c}\ \pi^{\pm}$ at 30^{o}

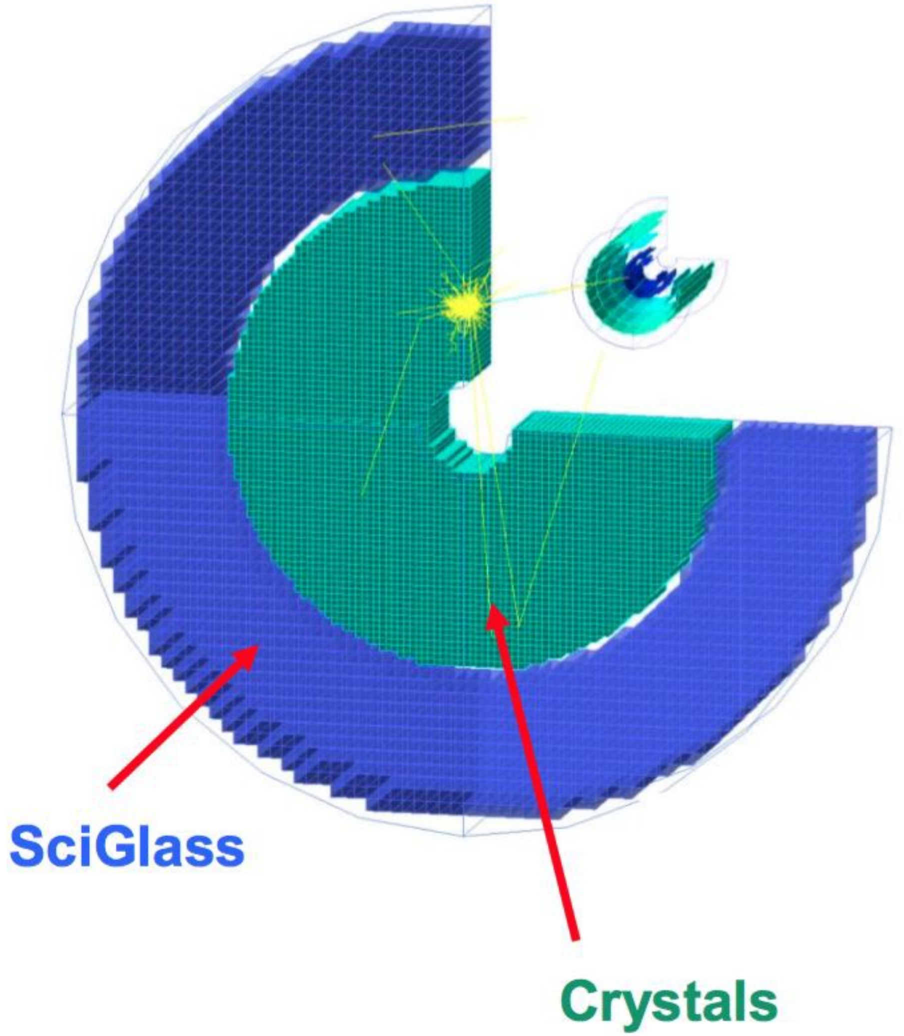


G. Kalicy, W. Phelps, A. Quiroga and N. Wickramaarachchi

Co-design example: Calorimetry

- Started co-design of electron-going endcap EMCal
- Maintain resolution while reducing the number of crystals?
- Optimises:

1. $PbWO_4$ crystal geometry
2. Inner/outer calo. radius
3. Densities
4. Efficiencies



V. Berdnikov, M. Bondi, C. Fanelli, Y. Furletova, T. Horn, I. Larin, D. Romanov, R. Trotta

Conclusions

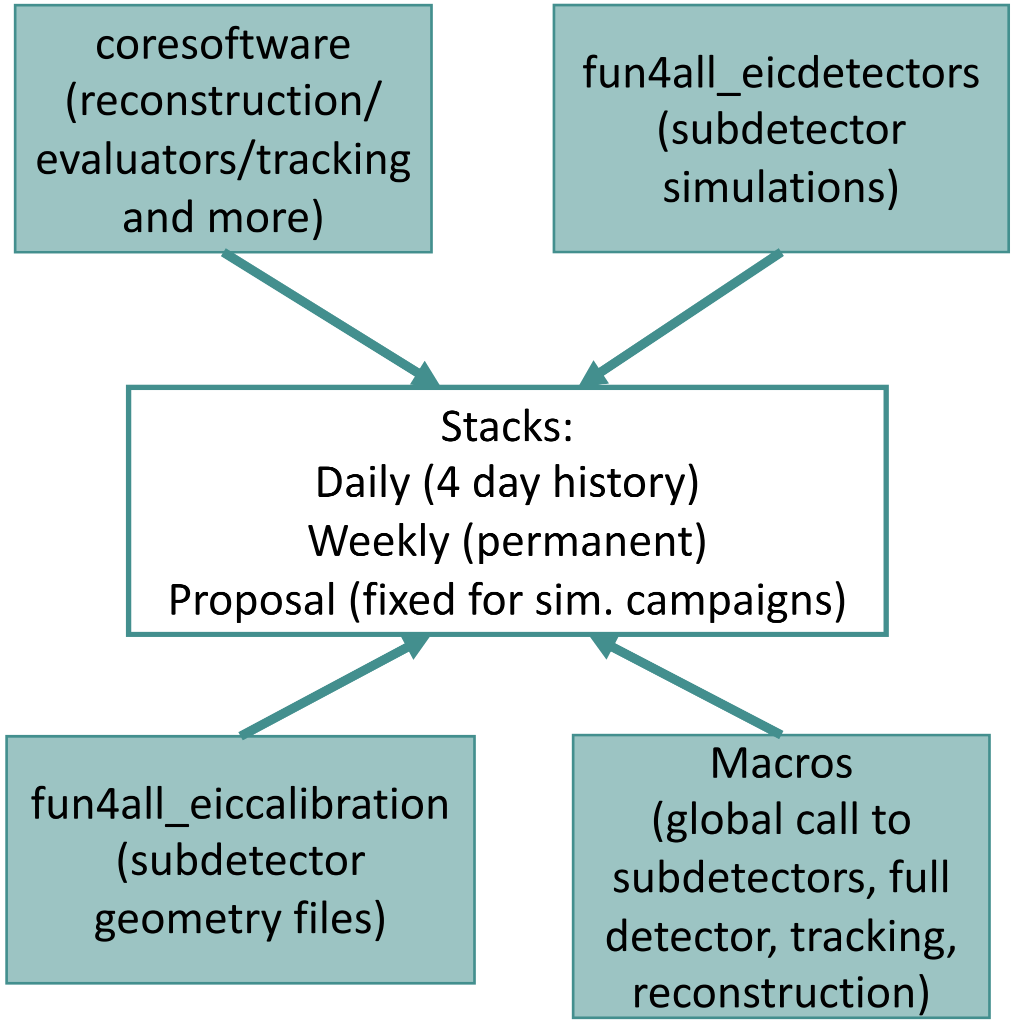
- ECCE members have risen to the challenge of the detector charge
- Realistic detector simulations are in place
- 150M already in storage
- Aim to store > 300M events by end of August
- AI working group is looking to the future with continual optimization
- UI streamlined – Data can be studied < 20 minutes after joining ECCE

Thank you

Backup

Code Production

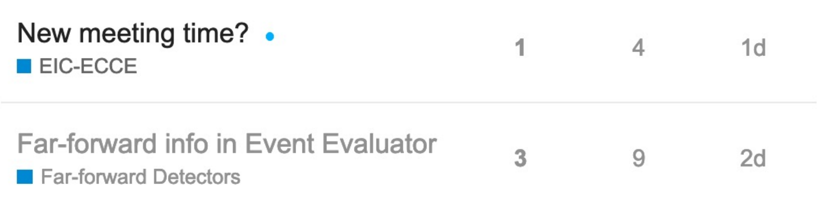
- All users are welcome to submit packages or updates
- Analysis code is written in C++
  - Other languages are used for more specific tasks (e.g. python and bash for productions)
- Code-conventions established by Chris Pinkenburg
- All code must compile with clang, have no unused objects, cpp-check must have no serious issues, valgrind used to find memory leaks



ECCE software talk with more information

User Work Methods

- Many users across the world, not everyone joins with active BNL membership
  - Not feasible to get everyone accounts in proposal timescale
- Singularity (and VirtualBox) is used to distribute daily software stacks (and simulations)
- S3 (BNL) and xrootd (JLab) protocols used to distribute data
  - minIO client and read-access keys are distributed from ECCE stack
- Users are encouraged to use low-volume nTuples over DSTs
  - Keeps bandwidth to a minimum
  - Many physics plots can be made without large data processing



We use mattermost and discourse for quick communication. We can manually authenticate, no need to have a BNL account