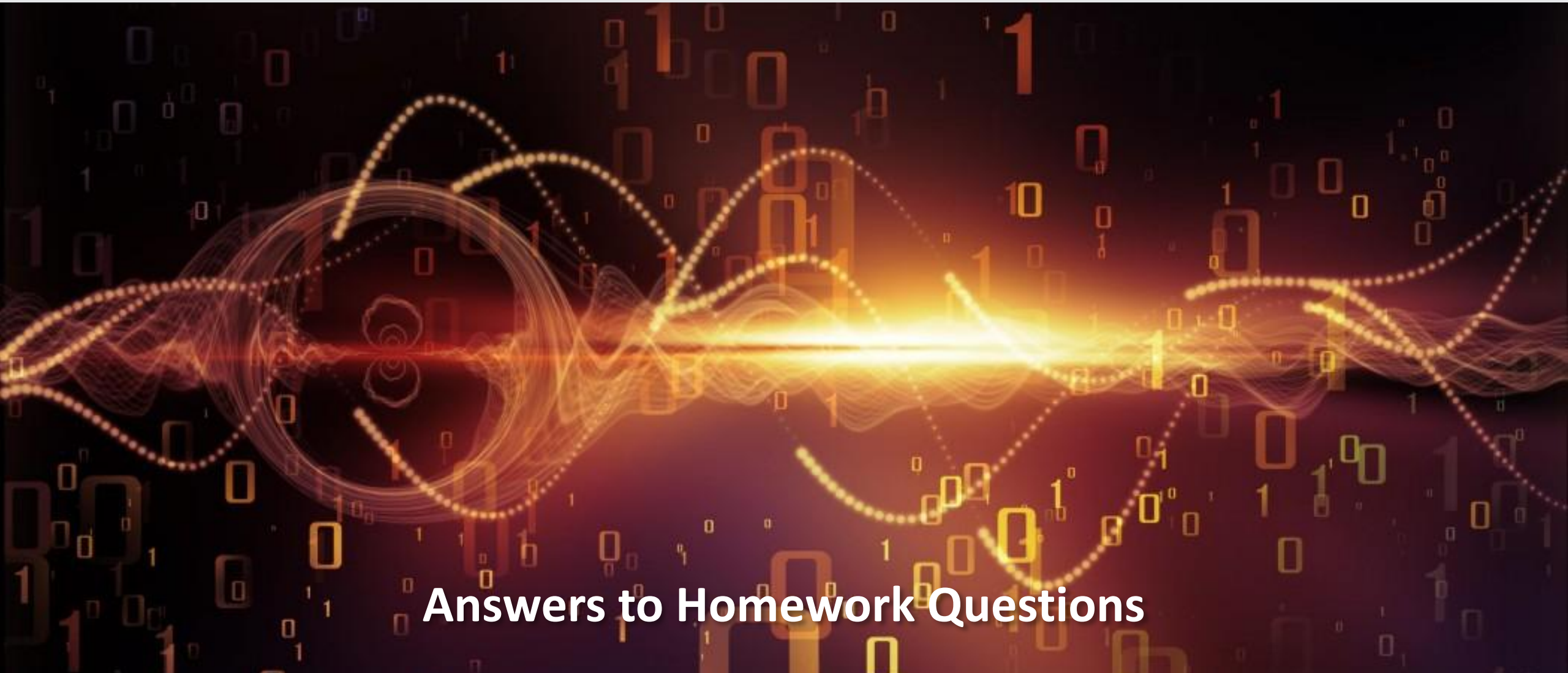


Computing & Software Review



Answers to Homework Questions

Questions

1. In Markus' talk, please replace slide 17, 18, and maybe 19 as well (?) with the actual needs in 2034, rather than needs scaled down by 5x or whatever the scaling factor is in these slides. Provide a rough estimate of the disk/tape split.
2. In Torre's talk slide 9, we understood that the ePIC s&c R&D is hurting for effort. Can you please provide us a rough estimate of committed effort at DOE lab, US university, and international. Both in FTE count and in how many people this FTE count is distributed across. We want here the FTE effort that you as s&c management consider committed to ePIC s&c given what you see is happening in terms of contributions to ePIC s&c R&D
3. If you were given 4 dedicated experts, make a prioritized list of the tasks/areas where you would employ them for your short/medium term needs (next 3 years)
4. Answer the same question as above, but for the more long term needs of the collaboration.
5. Explain Echelon 0 and 1 roles: how do you imagine the commissioning phase will analyze data and how the general and data quality monitoring will work during the production data taking (online and event based)? Issue: how can Echelon 1 actually be used before the construction is completed and operations have thus not yet started?
6. RHIC will possibly stop in 2025. What is the impact on ePIC, given that today ePIC does not have an operations budget and relies on synergies with other projects including RHIC?

Q1: Storage Resource Estimates

Actual needs in 2034

| Storage Estimates by Use Case [PB] | Echelon 1 | Echelon 2 |
|---------------------------------------|------------|------------|
| Streaming Data Storage and Monitoring | 71 | 35 |
| Alignment and Calibration | 1.8 | 1.8 |
| Prompt Reconstruction | 4.4 | - |
| First Full Reconstruction | 8.9 | 3.0 |
| Reprocessing | 9 | 9 |
| Simulation | 107 | 107 |
| Total estimate storage | 201 | 156 |

Echelon 1 sites arrive data, two copies
One copy (can and may be more) across
Echelon 2 sites for alignment, calibration,
and reconstruction use cases.

This table has been updated in the slides posted on Indico as well.

Q1: Computing Resource Estimates

Actual needs in 2034

| Processing by Use Case [cores] | Echelon 1 | Echelon 2 |
|---------------------------------------|----------------|----------------|
| Streaming Data Storage and Monitoring | - | - |
| Alignment and Calibration | 6,004 | 6,004 |
| Prompt Reconstruction | 60,037 | - |
| First Full Reconstruction | 72,045 | 48,030 |
| Reprocessing | 144,089 | 216,134 |
| Simulation | 123,326 | 369,979 |
| Total estimate processing | 405,501 | 640,147 |

See prompt reconstruction.

Roughly 10% of data stream.

Must keep up with data taking; assume 2x headroom.

Repro includes simu as well as real data

Simply adding together the core counts is an overestimate. Reconstruction core hours used only part time.

This table has been updated in the slides posted on Indico as well.

Echelon 0: The raw data from the ePIC Streaming DAQ (Echelon 0) will be replicated across the host labs (Echelon 1). At the highest luminosity of $1e34$, the data stream from the ePIC Streaming DAQ is estimated at 100 Gbit/s. Consequently, Echelon 0 requires an outgoing network connection of at least 200 Gbit/s.

Echelon 1: Each Echelon 1 facility has similar requirements, as it will receive up to 100 Gbit/s of raw data and will share this data with Echelon 2. In addition, Echelon 1 will send a small amount of monitoring data, approximately 1 Gbit/s, back to Echelon 0. Echelon 1 will also receive calibration and analysis data from various Echelon 2 nodes at a comparable rate of about 1 Gbit/s.

Echelon 2: The network connection requirements for Echelon 2 facilities will depend on the proportion of raw data they intend to process. For the 10% of Echelon 1 scenario, a network connection of 20 Gbit/s would be required.

This slide is unchanged

Questions 2

In Torre's talk slide 9, we understood that the ePIC s&c R&D is hurting for effort. Can you please provide us a rough estimate of committed effort at DOE lab, US university, and international. Both in FTE count and in how many people this FTE count is distributed across. We want here the FTE effort that you as s&c management consider committed to ePIC s&c given what you see is happening in terms of contributions to ePIC s&c R&D

DOE Labs:

- 0 FTE are committed to ePIC S&C R&D (only best effort is possible)

US and International University: (counting postdoctoral and more senior)

- Total: 6.9 FTE, for a total of 20 people
- United States: 5.6 FTE, for a total of 16 people
- International: 1.3 FTE, for a total of 4 people

Questions 3 and 4

If you were given 4 dedicated experts, make a prioritized list of the tasks/areas where you would employ them for your short/medium term needs (next 3 years)

Answer the same question as above, but for the more long term needs of the collaboration.

Short/Medium-Term (next 3 years)

- Establish a dedicated effort in collaboration with Electronics & DAQ to develop integrated DAQ-computing workflows, working towards a full streaming DAQ chain test
- Holistic full PID full reconstruction (lepton-hadron separation, lepton ID, hadron ID) implementation in the ePIC software stack utilizing the full capabilities of the integrated detector (PID, calo, tracking, etc.)
- Support AI/ML workflow integration in full simulation and reconstruction algorithms
- ACTS expert for track seeding, track fitting, vertex finding algorithm development, tuning, and evaluation

Long-Term (4+ years)

- Continued support for streaming DAQ workflows in collaboration with Electronics & DAQ
- Expert in fast simulations to reduce the computational cost of the simulation campaigns to interpret data
- Expert in hardware accelerators to develop collaboration expertise to speed up simulation and reconstruction and leverage HPC platforms
- Distributed computing expert to develop operations between Echelon-1/2 and support progressively scaled up challenges

Questions 5

1. Explain Echelon 0 and 1 roles: how do you imagine the commissioning phase will analyze data and how the general and data quality monitoring will work during the production data taking (online and event based)? Issue: how can Echelon 1 actually be used before the construction is completed and operations have thus not yet started?

Answer

The commissioning phase will exercise the same workflows planned for production data taking, to demonstrate readiness for data taking. Data monitoring will be accomplished by reconstructing a fraction of timeframes in Echelon 1 in near real-time, with the output made available to the experiment for data quality monitoring. Echelon 1 sites at both host labs will provide redundancy.

Echelon 1 at both labs will be required to support both the commissioning of the ePIC detector (with cosmics) and the EIC accelerator. We anticipate a phased development of Echelon 1, supported by pre-operations for the EIC, ramping up over time to the full EIC Phase-1 needs presented at this review.

Milestones Aligned with Streaming DAQ Development

Streaming DAQ Release Schedule:

PicoDAQ **FY26Q1**

- Readout test setups

MicroDAQ: **FY26Q4**

- Readout detector data in test stand using engineering articles

MiniDAQ: **FY28Q1**

- Readout detector data using full hardware and timing chain

Full DAQ-v1: **FY29Q2**

- Full functionality DAQ ready for full system integration & testing

Production DAQ: **FY31Q3**

- Ready for cosmics

Streaming Computing Milestones:

Start development of streaming orchestration, including workflow and workload management system tool.

Start streaming and processing streamed data between BNL, Jefferson, DRAC Canada, and other sites.

Support of test-beam measurements, using variety of electronics and DAQ setups:

- Digitization developments will allow detailed comparisons between simulations and test-beam data.
- Track progress of the alignment and calibration software developed for detector prototypes.
- Various JANA2 plugins for reading test-beam data required. Work started on an example.

↔ **Establish autonomous alignment and calibration workflows that allows for validation by experts.**

Analysis challenges exercising end-to-end workflows from (simulated) raw data.

↔ **Streaming challenges** exercising the streaming workflows from DAQ through offline reconstruction, and the Echelon 0 and Echelon 1 computing and connectivity.

Analysis challenges exercising autonomous alignment and calibrations.

↔ **Data challenges exercising scaling and capability tests** as distributed ePIC computing resources at substantial scale reach the floor, including exercising the functional roles of the Echelon tiers, particularly Echelon 2, the globally distributed resources essential to meeting computing requirements of ePIC.

Question 6

RHIC will possibly stop in 2025. What is the impact on ePIC, given that today ePIC does not have an operations budget and relies on synergies with other projects including RHIC?

Answer

Analysis of the RHIC experiments, the main driver for software and computing, will continue through 2030+. This will ensure that synergistic efforts for ePIC Software & Computing can continue.