ePIC Streaming Computing Model

Markus Diefenthaler (Jefferson Lab) for the ePIC Collaboration

ePIC Streaming Computing Model

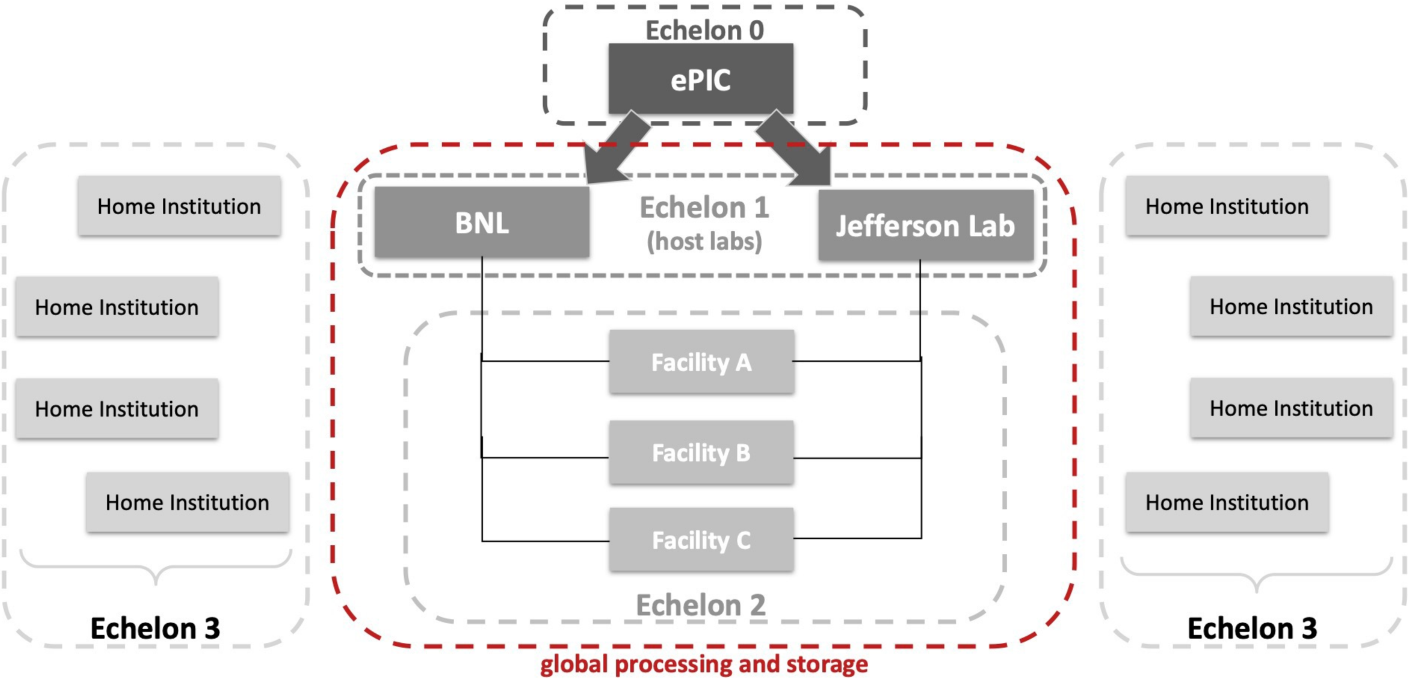
Four Tiers:

Echelon 0: ePIC Experiment

Echelon 1: Host Labs

Echelon 2: Global processing and data facilities, includes HPC and HTC resources.

Echelon 3: Home institute computing



ePIC Streaming Computing Model successfully reviewed by EIC Computing and Software Advisory Committee (ECSAC).

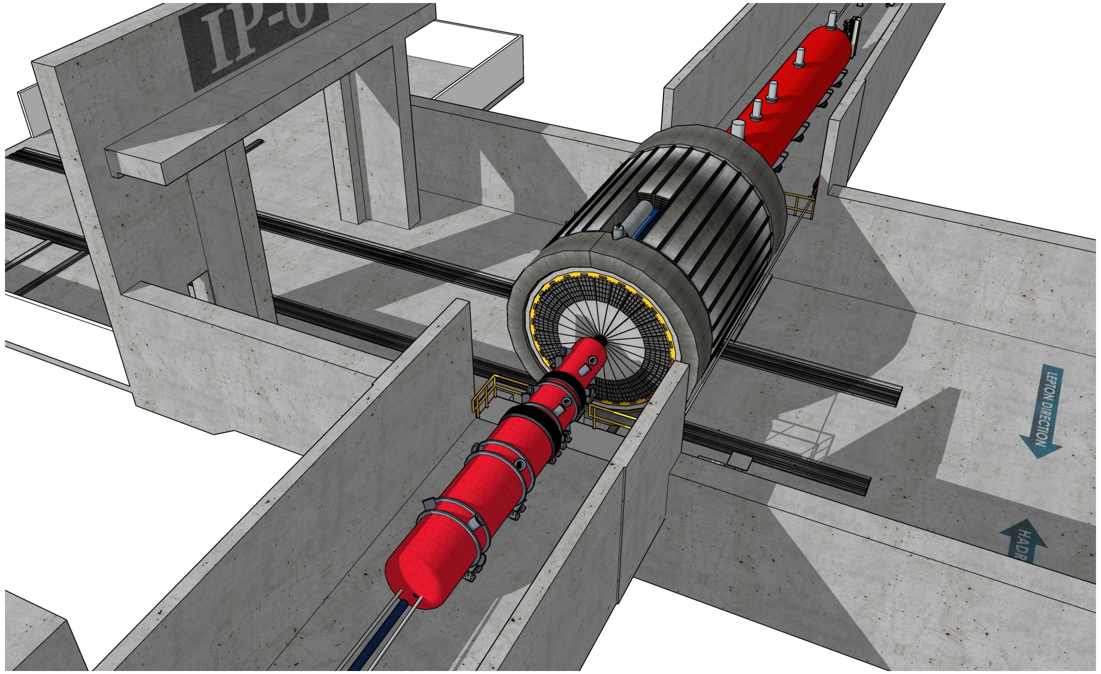
The structure of this presentation aligns with that of the report we prepared for the review.

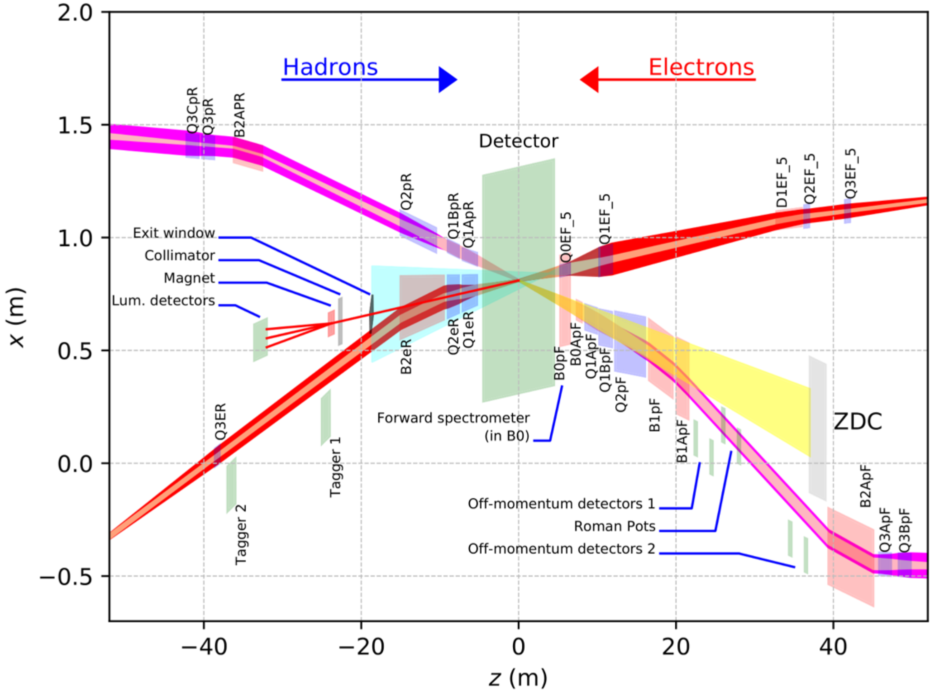
I provide references to various sections in the report for more detailed information on the topics.

Optimize Physics Reach

Integrated interaction and detector region (90 m)

Get ~100% acceptance for all final state particles, and measure them with good resolution. All particles count!





Compute-Detector Integration

Extend integrated interaction and detector region into detector readout (electronics), data acquisition, data processing and reconstruction, and physics analysis.

Compute-Detector Integration to Maximize Science

Broad ePIC Science Program:

- Plethora of observables, with less distinct topologies where every event is significant.
- High-precision measurements: Reducing systematic uncertainties of paramount importance.

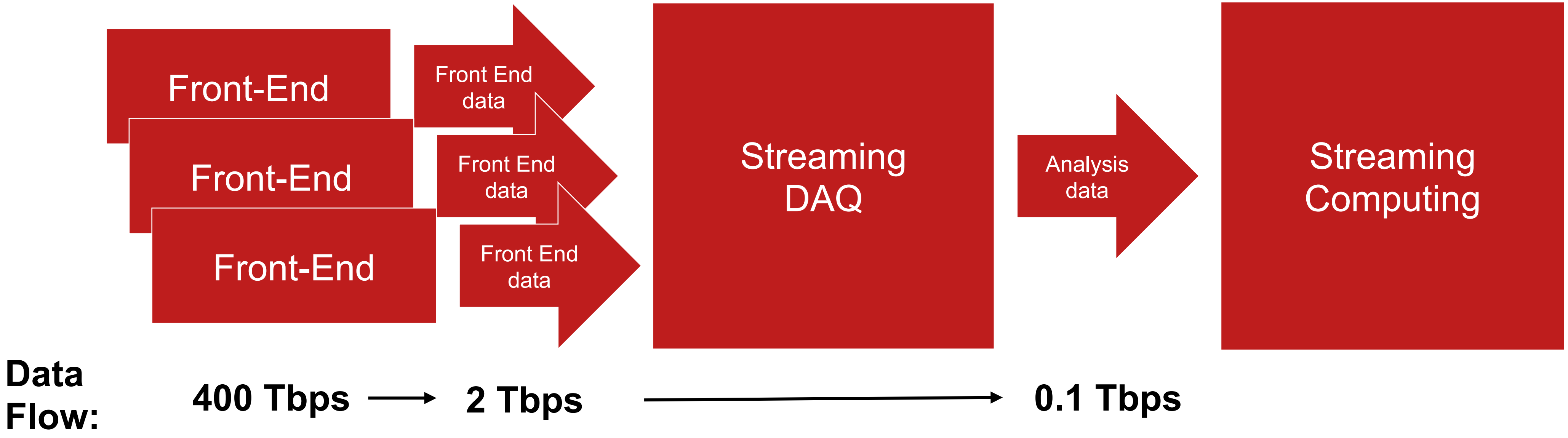
Streaming Readout Capability Due to Moderate Signal Rate:

- **Capture every collision signal**, including background.
- Event selection using all available detector data for **holistic reconstruction**:
  - **Eliminate trigger bias** and provide accurate estimation of uncertainties during event selection.
- Streaming background estimates ideal to **reduce background** and related systematic uncertainties.

	EIC	RHIC	LHC → HL-LHC
Collision species	$\vec{e}+\vec{p}, \vec{e}+A$	$\vec{p}+\vec{p}/A, A+A$	$p+p/A, A+A$
Top x-N C.M. energy	140 GeV	510 GeV	13 TeV
Peak x-N luminosity	10^{34} cm^{-2} s^{-1}	10^{32} cm^{-2} s^{-1}	$10^{34} \rightarrow 10^{35}$ cm^{-2} s^{-1}
x-N cross section	**50 μb**	**40 mb**	**80 mb**
Top collision rate	500 kHz	10 MHz	1-6 GHz
$dN_{ch}/d\eta$	0.1-Few	~3	~6
Charged particle rate	**4M N_{ch}/s**	**60M N_{ch}/s**	**30G+ N_{ch}/s**

Compute-Detector Integration to Accelerate Science

- **Problem** Data for physics analyses and the resulting publications available after O(1year) due to complexity of NP experiments (and their organization).
  - Alignment and calibration of detector as well as reconstruction and validation of events time-consuming.
- **Goal** **Rapid turnaround of 2-3 weeks for data for physics analyses**.
- **Solution** Compute-detector integration using:
  - AI for autonomous alignment and calibration as well as reconstruction and validation for rapid processing,
  - Streaming readout for continuous data flow of the full detector information,
  - Heterogeneous computing for acceleration.



Computing Use Cases

Interface between Streaming DAQ and Streaming Computing

- **Ongoing Discussion**: Where does the interface lie between “online” and “offline”?
- **Working Definition**: Point where data moves to archival storage.
- **Challenges**: Technical and sociological differences at this point:
  - **Pre archival storage**: Risk of permanently losing data in case of error or reduced live time.
  - **Post archival storage**: Less stringent requirements and latencies. More open environment.

Scope of Use Case Discussion:

- Discussion of use cases centers on “offline” computing.

Use Cases:

- Stored Data Streaming and Monitoring
- Alignment and Calibration
- Prompt Reconstruction
- First Full Reconstruction
- Reprocessing
- Simulation
- Analysis
- Modeling and Digital Twin

Computing Use Cases: Stored Data Stream, Monitoring, and Calibrations

Use Case: Stored Data Streaming and Monitoring

Sec. 4.2

- **Primary Duty**: Archive incoming raw data from DAQ.
- **Data Replication**: Geographically separate replicas at BNL and JLab.
- **Real-Time Workflows**: Calibration and prompt processing.
- **Monitoring**: Automated and user-interface based validation, alarming.

Use Case: Alignment and Calibration

Sec. 4.3

- **Requirement**: Prompt alignment and calibration for rapid turnaround of reconstruction.
- **Workflows**: As automated and autonomous as possible during operation.
- **Data Products**: Sent to a globally accessible conditions database.
- **Initial Alignment and Calibration**: Only Echelon 1.
- **Refinements**: Can proceed in multiple Echelons.

Computing Use Cases: Reconstruction

Unique Requirement of Streaming Based Processing:

- Reconstruction of physics events from time frames as produced from the DAQ.

Use Case: Prompt Reconstruction

Sec. 4.4

- **Real-Time**: Events reconstructed in near real-time.
- **Availability**: Crucial for quick raw data to production turnaround:
  - Required for monitoring, diagnostics, and quick-turnaround calibrations.
- **Limitations**: Confined to Echelon 1 due to strict latency requirements.

Use Case: First Full Reconstruction

Sec. 4.5

- **Resource Constraints**: Echelon 1 may not handle complete processing.
- **Shared Work**: Echelon 2 involved for remaining tasks.
- **Time-Scale**: Driven by calibrations. Maximum acceptable completion is 2-3 weeks.

Use Case: Reprocessing

Sec. 4.6

- **Forms**: Full reprocessing, re-reconstruction, regeneration of analysis object data.
- **Batch-Style**: Suited for batch processing.
- **Resources**: Utilizes Echelons 1-2 and other opportunistic resources.

Computing Use Cases: Simulation, Analysis, and Modeling

Use Case: Simulation

Sec. 4.7

- **Types**: Monte Carlo physics and detector simulation.
- **Volume**: At least one order of magnitude more simulated than real events.
- **Data Structure**: Similar to that of real data.
- **Workflow Strategy**: Mimic streaming data workflows to gain experience.

Use Case: Analysis

Sec. 4.8

- **Analysis Use Cases**: Broad science program with numerous observables.
- **Techniques**: Unfolding or joint theoretical-experimental analyses on the event level.
- **Resource Needs**: Varying Echelon requirements based on study complexity.

Use Case: Modeling and Digital Twin

Sec. 4.9

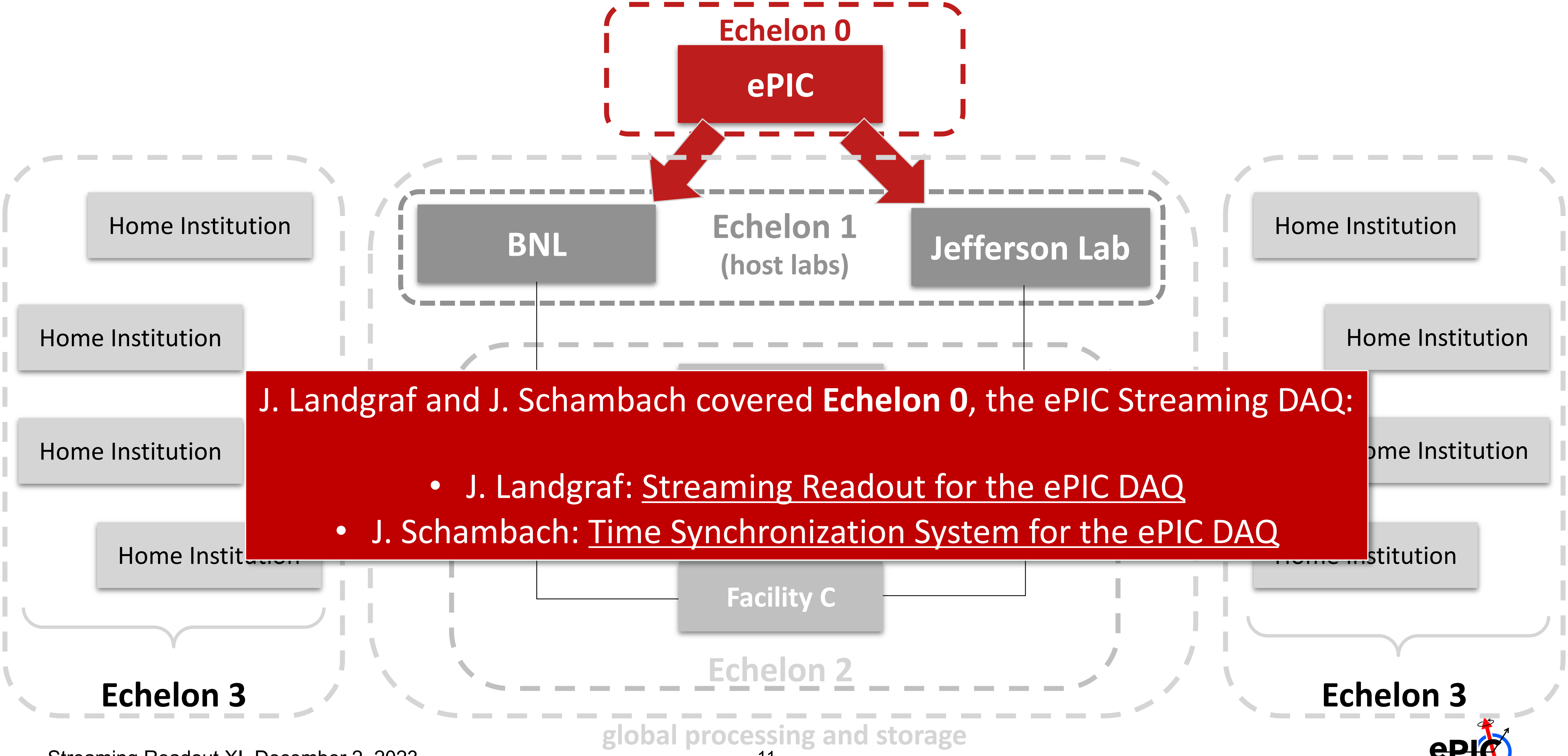
- **Modeling**: E.g., streaming data used for AI background modeling.
- **Digital Twin**: Complementary to detector simulations for real-time experimental control or optimization of experimental conditions.
- **Resources**: Echelon 1 and 2 primarily used for modeling workflows.

Use Case to Echelon Mapping

Echelon 0: ePIC Experiment
Echelon 1: Host Labs
Echelon 2: Global processing and data facilities
Echelon 3: Home institute computing

Use Case	Echelon 0	Echelon 1	Echelon 2	Echelon 3
Stored Data Streaming and Monitoring	✓	✓		
Alignment and Calibration		✓	✓	
Prompt Reconstruction		✓		
First Full Reconstruction		✓	✓	
Reprocessing		✓	✓	
Simulation		✓	✓	
Analysis		✓	✓	✓
Modeling and Digital Twin		✓	✓	

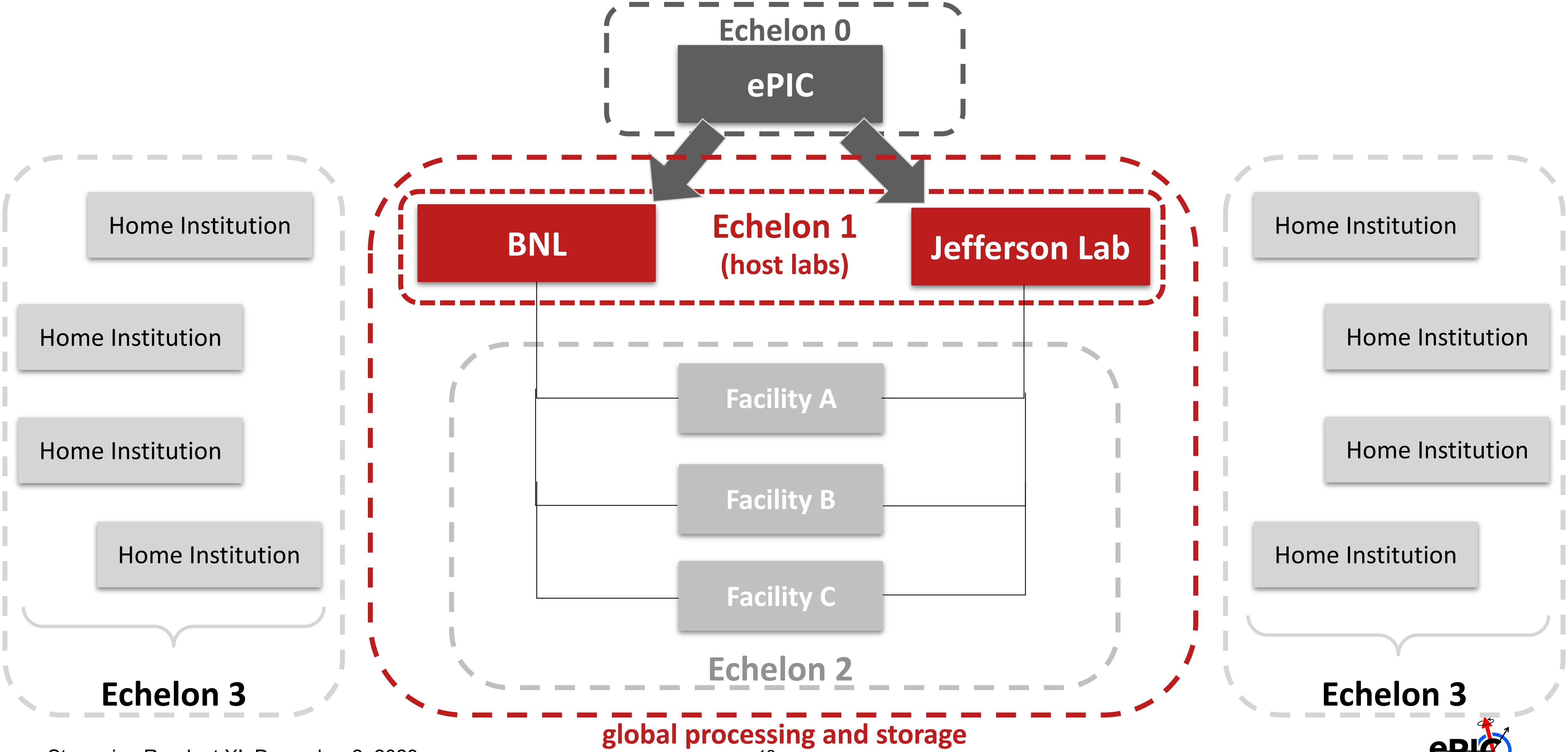
Echelon 0



J. Landgraf and J. Schambach covered **Echelon 0**, the ePIC Streaming DAQ:

- J. Landgraf: Streaming Readout for the ePIC DAQ
- J. Schambach: Time Synchronization System for the ePIC DAQ

Echelon 1


Echelon 0
ePIC
BNL
Echelon 1
(host labs)
Jefferson Lab
Home Institution
Home Institution
Home Institution
Home Institution
Echelon 3
Facility A
Facility B
Facility C
Echelon 2
Home Institution
Home Institution
Home Institution
Home Institution
Echelon 3
global processing and storage

Echelon 1 (and 2): Comment on Compute Resources

Status: Current estimates on computing resources:

Event size: 10MB frame size (1ms) based on our current detector readout design when running at peak luminosity and in standard operating conditions.

Number of Events expected to record and simulate per year: Assuming a 50% up-time for ½ year, we will record 15.5 billion frames in a year:

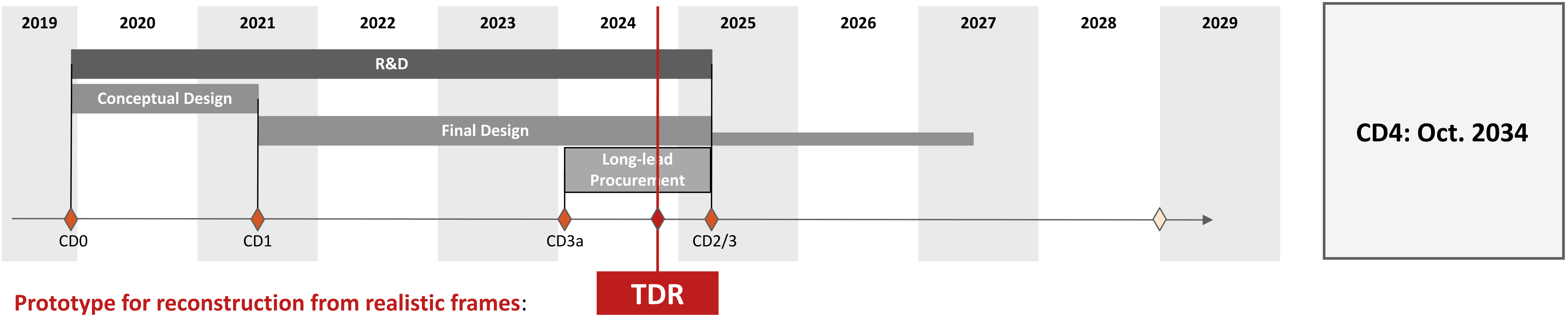
- The event rate at peak luminosity is 500kHz, which gives roughly 4×10^{12} events (60% background, 40% bunch crossing related):
- Lower at start of operations, where the luminosity will be lower (but relatively speaking background rate is expected to be higher).
- The actual physics events related to key EIC observables is only a very small fraction of the total physics bunch crossings. The expected number of DIS events / physics event of interest for one year of running at peak luminosity is ~ 10^{10}.
- This is the number that drives our simulation needs, and we expect to simulate 10x events for each event of interest, yielding $O(10^{11})$ simulated events. While considerable (~ 60k core years on today's hardware), this should be a realistic target in a decade.

Core-seconds for reconstruction and simulation (on a typical modern machine): Our current simulations of background embedded events take ~17s for simulation and ~ 2s for reconstruction, per event.
Unknown: How much this will change once changing to realistic frame to event reconstruction?

Goal: Reliable compute resource estimate on before TDR.

Reliable Compute Resource Estimates before TDR

Sec. 5.3.4



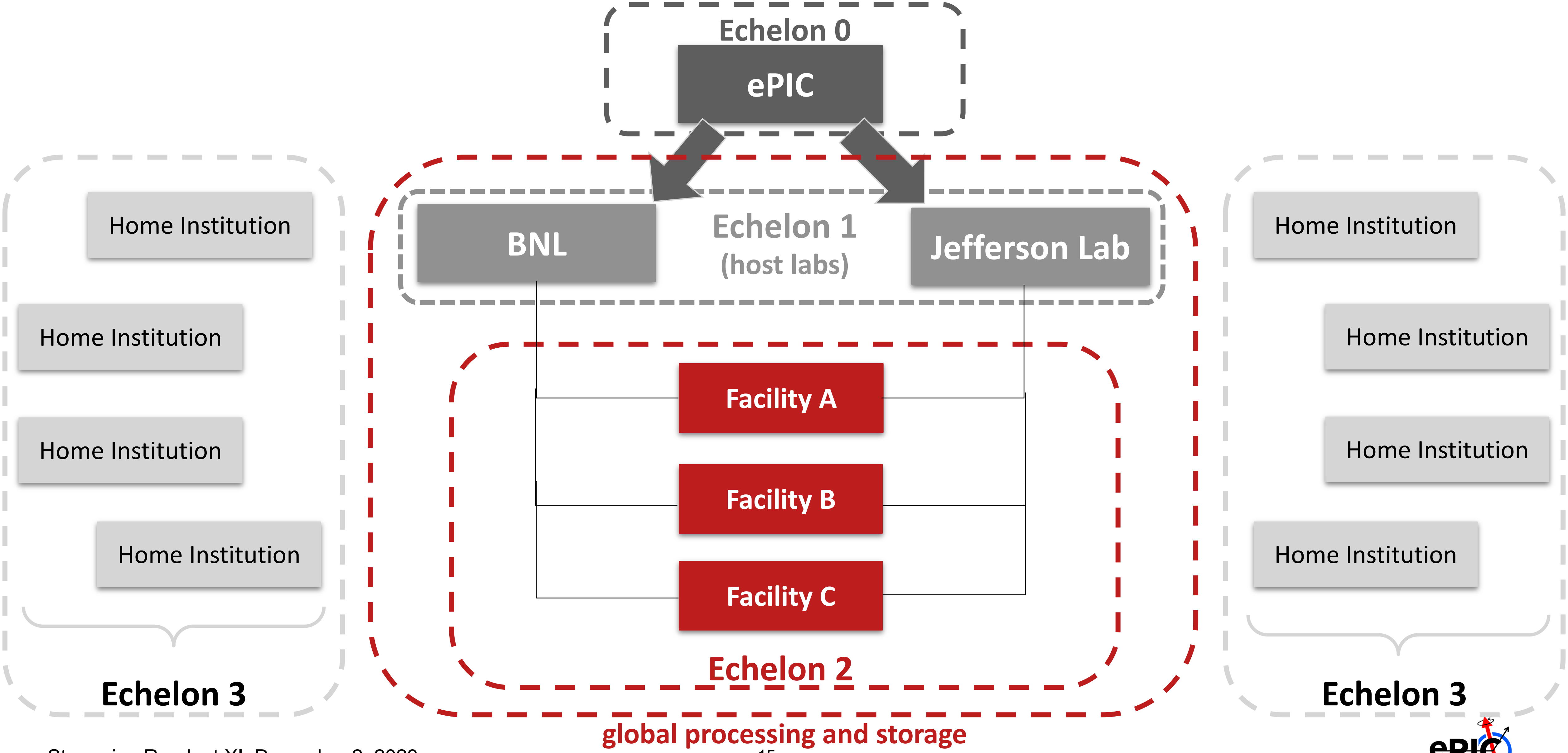
Prototype for reconstruction from realistic frames:

- **Scope limited to tracking**: Demonstrate we can correlate hits in a realistic time frame to the various events in the time window of the MAPS.

Framework Tests and Development:

  - **01/24** Integrate Jana2's built-in workflow for supporting frames in and events out in EICrecon.
- **Simulation**:
  - **12/23** Prepare simulation productions, using detailed information on FEEs for tracking detectors, utilizing the full, wide MAPS integration window for tracking purposes.
  - **01/24** Implement and utilize the frame-building infrastructure post-Geant4 and post-digitization.
- **Reconstruction Process**:
  - **02/2024** Adapt the reconstruction process to work with frames, making it frames-aware.
  - **03/2024** Demonstrate tracking from realistic frames
  - **04/2024** Deliver first estimate of reconstruction time from frames.

Echelon 2


Echelon 0
ePIC
BNL
Echelon 1
(host labs)
Jefferson Lab
Home Institution
Home Institution
Home Institution
Home Institution
Echelon 3
Facility A
Facility B
Facility C
Echelon 2
Home Institution
Home Institution
Home Institution
Home Institution
Echelon 3
global processing and storage

Echelon 2: Global ePIC Computing

Sec. 5.4

ePIC is an international collaboration and so is its computing:

- **Echelon 2** includes **global resources contributed by collaborating institutions**.
- Achieving scientific goals relies on effectively using Echelon 2's resources.
- Design of computing model aims for **effective integration and management**.

International computing contributions are essential.

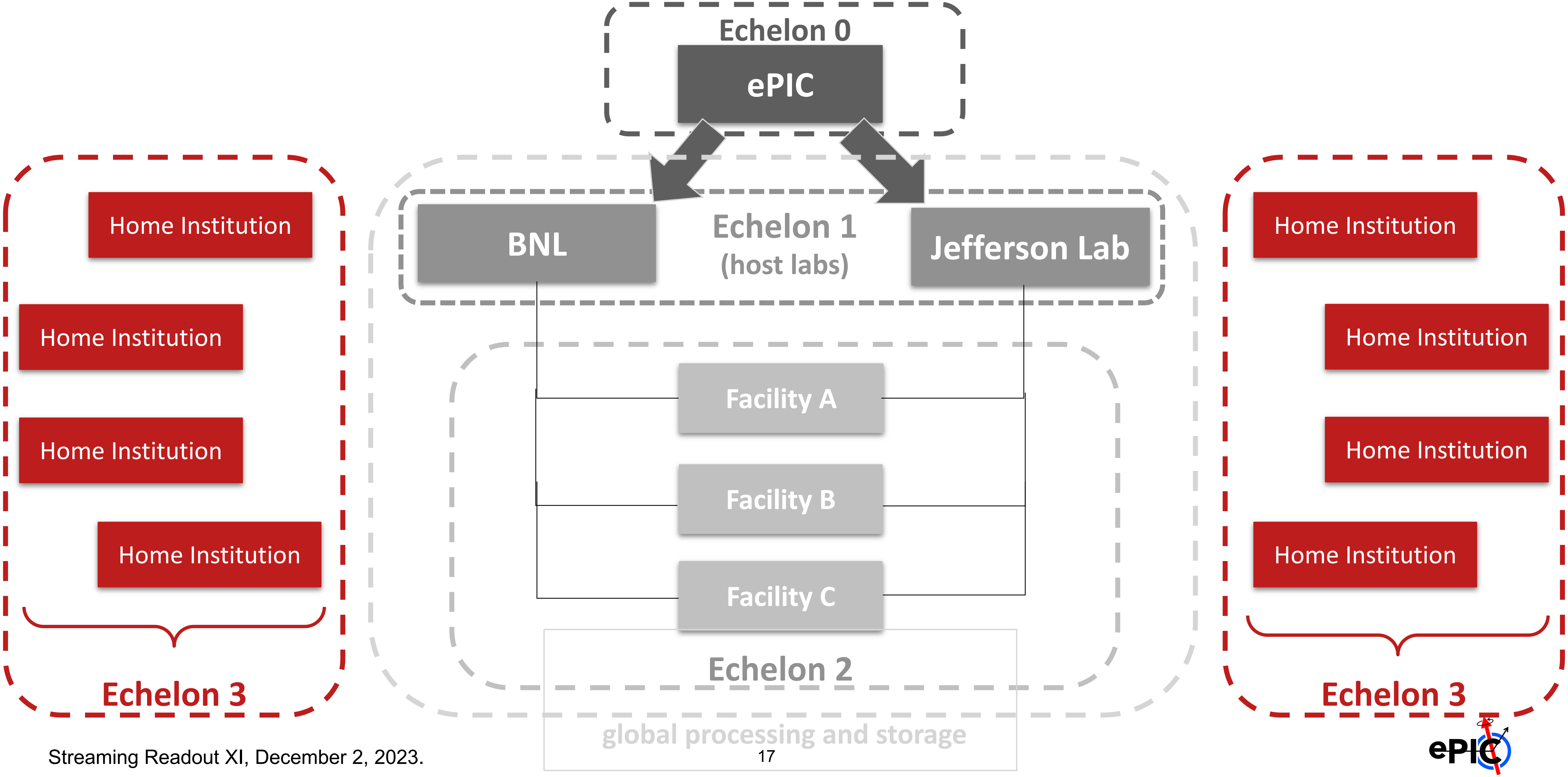
Lessons Learned from LHC

- Echelon 2 resources must have **MOUs specifying service requirements**.
- These MOUs **assure technical compatibility with the ePIC computing model**.
- **ePIC commits to facility integration** and robust testing/validation, includes monitoring and diagnostics.

Connectivity

- Echelon 2 sites connect equally to both Echelon 1 Host Labs.
- Connection is ultimately through ESnet network backbone.
- **Interconnected Mesh**:
  - Echelon 2 sites also interconnect based on their network environment.
  - Lessons from LHC show this interconnected mesh model is more effective than a hierarchical model.

Echelon 3



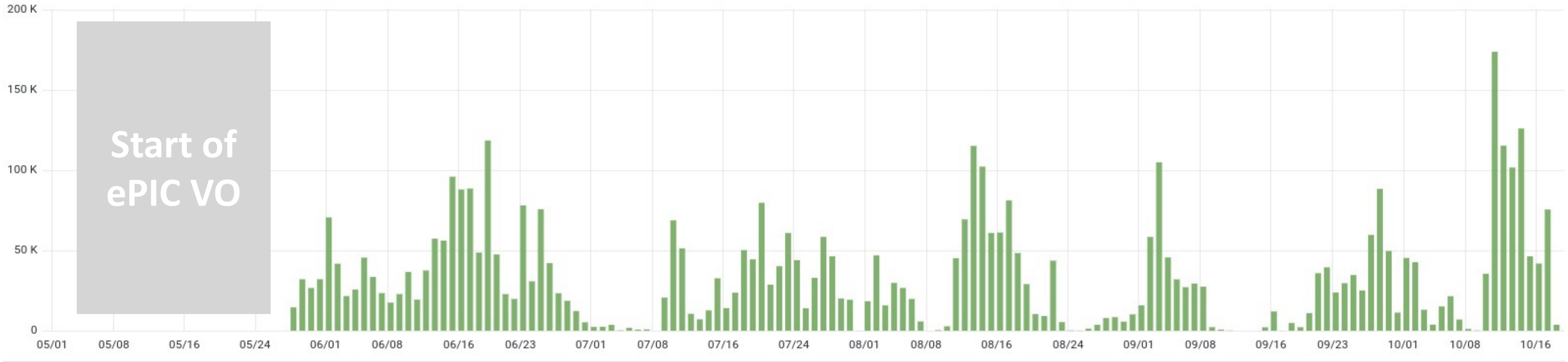
Echelon 0
ePIC
BNL
Echelon 1
(host labs)
Jefferson Lab
Facility A
Facility B
Facility C
Echelon 2
global processing and storage
Home Institution
Home Institution
Home Institution
Home Institution
Echelon 3
Home Institution
Home Institution
Home Institution
Home Institution
Echelon 3

Echelon 3: Home Institute Computing

- **Echelon 3**: Component in the computing model where collaborators directly interact with the computing system:
  - Users can access ePIC computing through various platforms like institutional clusters, work desktops, and personal laptops.
  - The role of Echelon 3 is to serve these diverse use cases.
- **Echelon 3 Resources**:
  - Echelon 3 resources are both global and local to the user, similar to Echelon 2.
  - The resources available in E3 are numerous, diverse, volatile, and often have restrictions on their use.
  - These resources are not intended to be managed as Collaboration resources.
- The collaboration will provide tools, interfaces, connection points, data access mechanisms, and support to make Echelon 3 resources effective for ePIC analysis.

Opportunistic Resources

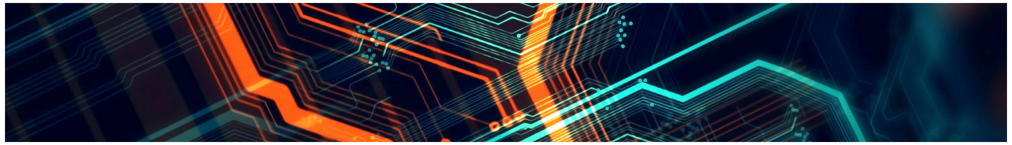
- **Use of Open Science Grid (OSG):**
  - Currently the most productive resource for ePIC. Expected to continue playing a role (simulation productions).
  - Stable concurrent core count of 5-10k.
    - Expect to grow substantially once we fold in Streaming DAQ simulations (and optical photon simulation).



- **Opportunistic Resources:**
  - Utilization of opportunistic resources like OSG is foreseen as ePIC science will likely be compute-limited.
  - WLCG evolving in also supporting non-LHC experiments.
- **Commercial Clouds:**
  - Actively used by other science communities.
  - Cost-effectiveness is promising, especially for fast-turnaround use cases.
  - ePIC is monitoring these developments for future consideration.

Special Resources

EIC SOFTWARE:
Statement of Principles



3 We will leverage heterogeneous computing:

- We will enable distributed workflows on the computing resources of the worldwide EIC community, leveraging not only HTC but also HPC systems.
- EIC software should be able to run on as many systems as possible, while supporting specific system characteristics, e.g., accelerators such as GPUs, where beneficial.

Special Resources:

- Includes non-x86 architectures like ARM, and accelerators like GPUs and TPUs.
- ePIC infrastructure aims to be flexible and extensible to support emerging architectures.
- ARM already supported. Will play important role due to its cost-effectiveness and ease of porting.
- FPGAs will be used for low-level data processing in Echelon 0.

Artificial Intelligence:

- Tracking the technologies for AI R&D and applications.
- Likely to exploit specialized accelerators like TPUs.

Large Supercomputers (LCFs): Developed by the DOE and NSF, often include special resources.

- Effectiveness for ePIC use will be evaluated on a case-by-case basis.

Streaming Data Processing

Traditional Workflow Characteristics in NP and HEP Experiments:

- Data is acquired in online workflows.
- Data is stored as large files in hierarchical storage.
- Offline workflows process the data, often with substantial latency.
- Batch queue-based resource provisioning is typical.
- Key features: discrete, coarse-grained processing units (files and datasets) and decoupling from real-time data acquisition.

ePIC Streaming Data Processing Characteristics

- Quasi-continuous flow of fine-grained data.
- Dynamic flexibility to match real-time data inflow.
- Prompt processing is crucial for data quality and detector integrity.
- Processing full data set quickly to minimize time for detector calibration and deliver analysis-ready data.

Challenging Characteristics of Streaming Data Processing:

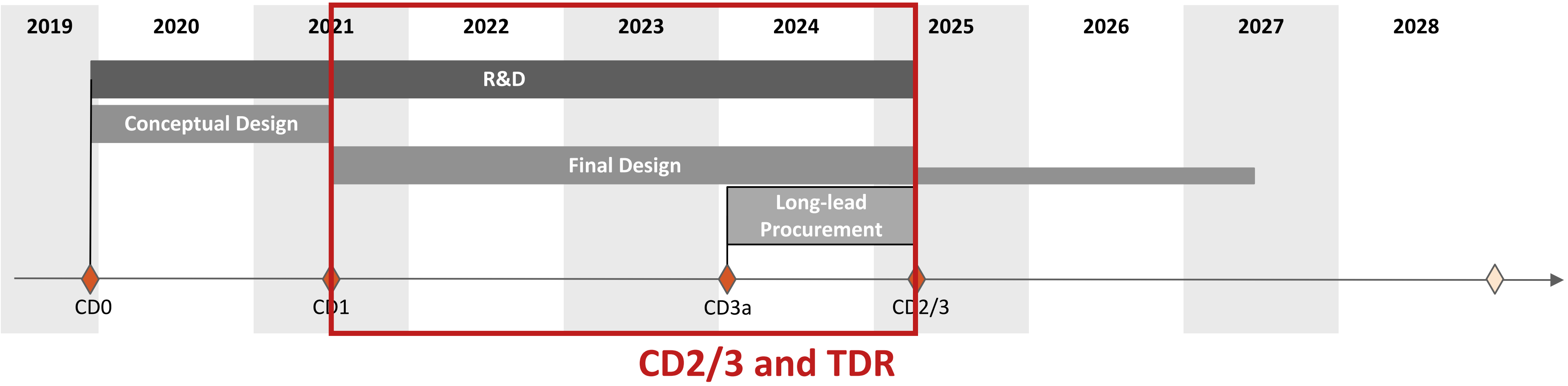
- **Time critical**, proceeding in near real time.
- **Data driven**, consuming a fine-grained and quasi-continuous data flow across parallel streams.
- **Adaptive and highly automated**, in being flexible and robust against dynamic changes in datataking patterns, resource availability and faults.
- **Inherently distributed** in its data sources and its processing resources.

Assumptions for Infrastructure:

- Existing batch-style processing likely to remain.
- Dynamic processing, e.g. Kubernetes, may displace the batch model.
- Design the system for both batch and dynamic processing to ensure resilience against technology evolution.
- Accommodate but effectively hide these underlying infrastructure characteristics.

Milestones

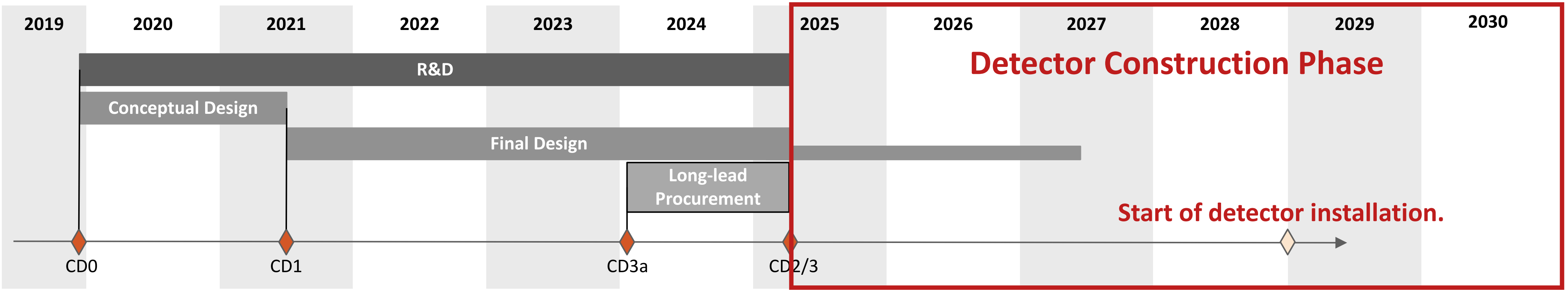
Sec. 9.2.1



Milestones Prior to CD2/3 and TDR

- Software and simulation readiness for TDR preparation (and subsequent phases of the CD process).
- Provide for each use case detailed estimates on the compute resources; update the networking and storage estimates according to format of streaming data format that is currently being defined.

Milestones

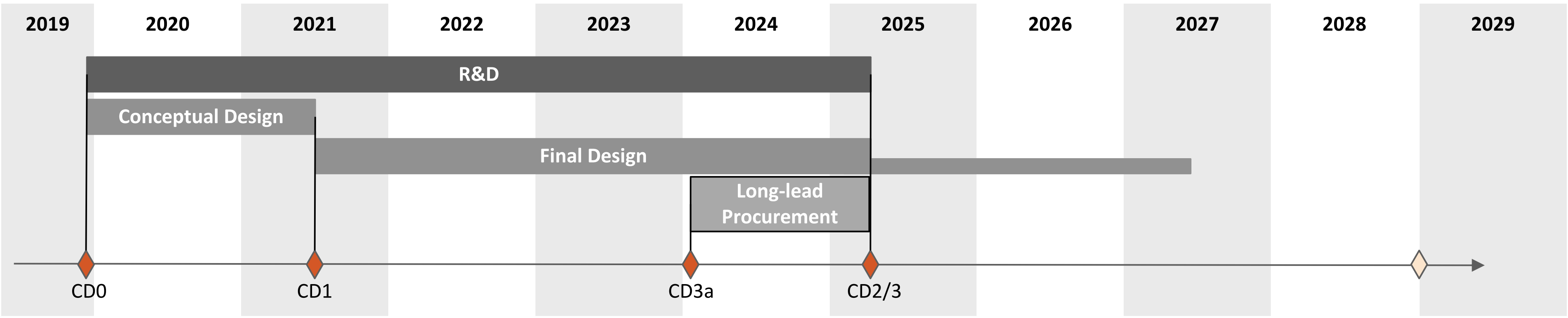


Milestones During Detector Construction Phase

- Provisioning DAQ and software sufficient for test beams, which can serve as small scale real-world testbeds for the developing DAQ and software.
- Streaming challenges exercising the streaming workflows from DAQ through offline reconstruction, and the Echelon 0 and Echelon 1 computing and connectivity.
- 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 ePIC's computing requirements.
- Analysis challenges exercising autonomous alignment and calibrations.
- Analysis challenges exercising end-to-end workflows from (simulated) raw data to exercising the analysis model.

Milestones

Sec. 9.2.3



CD4: Oct. 2034

Milestones During Detector Commissioning

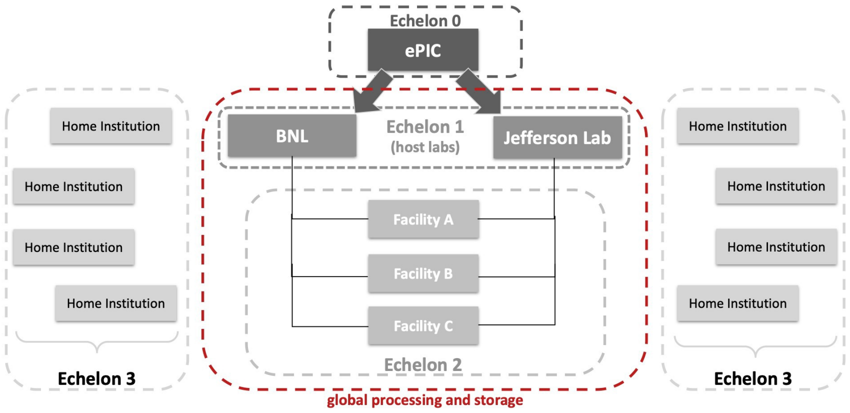
- This phase has unique expectations and requirements compared to steady-state operation:
  - Utilization of semi-triggered data-taking modes.
  - Initial calibrations.
  - Gradual extension of first pass processing from Echelon 1 to Echelon 2.
- Careful planning of software & computing efforts and leveraging experience from data and analysis challenges during the detector construction phase essential.

Milestones during Early Datataking Phase

- Simpler and more conservative approaches will be used during initial data-taking phase.
- The ePIC Streaming Computing Model will be gradually deployed and validated.

Summary

- **Streaming Readout of the ePIC Detector to maximize and accelerate science:**
  - ePIC aims for **rapid turnaround of 2-3 weeks for data for physics analyses**.
  - Timeline driven by calibrations.
- **Four tiers of the ePIC Streaming Computing Model computing fabric:**
  - **Echelon 0**: ePIC experiment and its streaming readout.
  - **Echelon 1**: Crucial and innovative partnership between host labs.
  - **Echelon 2**: Essential global contributions.
  - **Echelon 3**: Full support of the analysis community.



- ePIC supports the ECSJI formation to oversee and coordinate the complex computing fabric of the EIC.
- **High level milestones** ensures that the agile development process is continuously confronted with real world exercising of the software and the developing realization of the computing model:
  - Priority always given to meeting near-term needs.
  - Longer range timeline progressively exercising the streaming computing model to deliver for the needs of the CD process, for specific applications, e.g. test beams, for scaling and capability challenges, and ultimately for the phases of datataking.