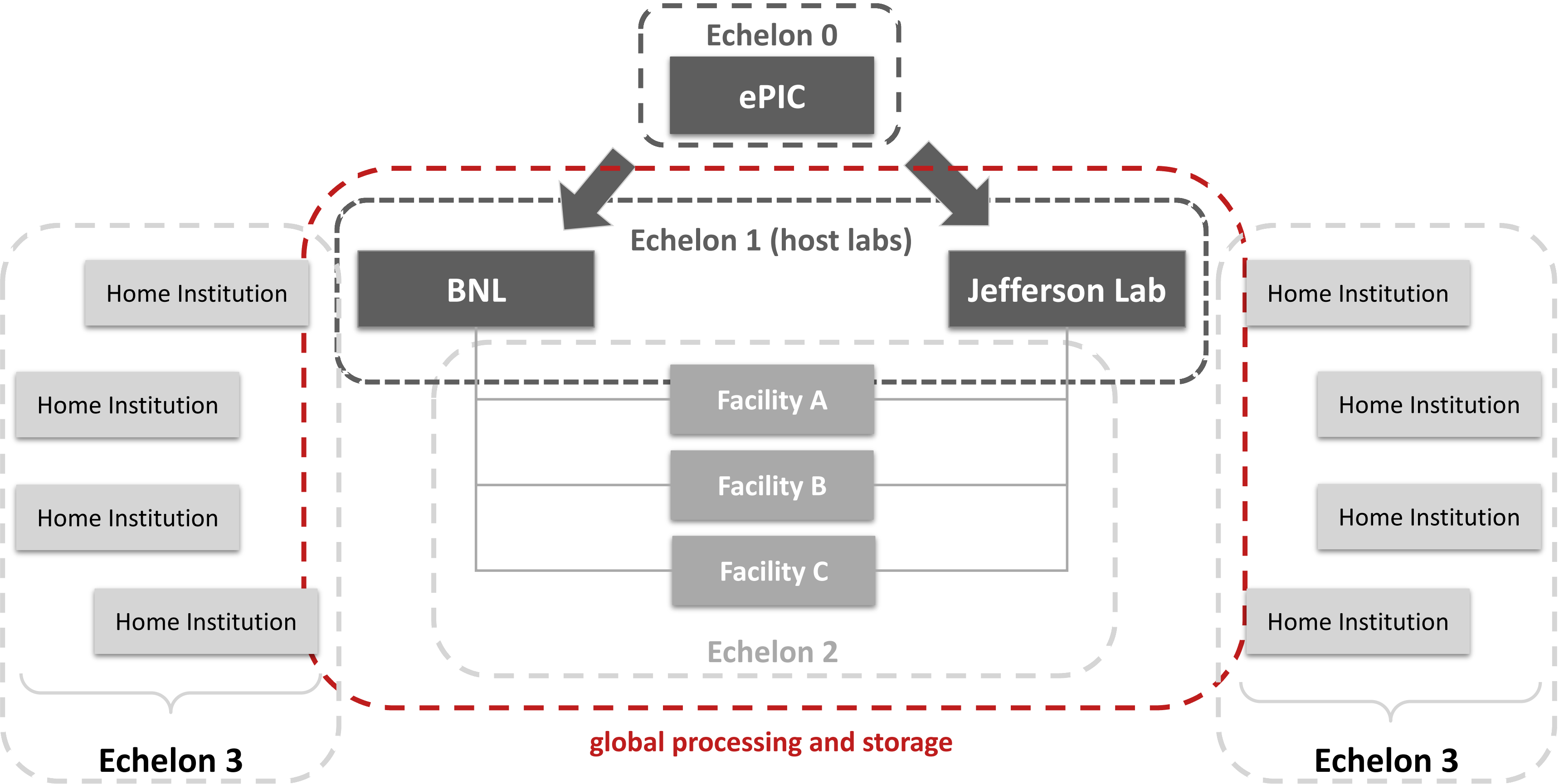
ePIC Streaming Computing Model and E0-E1 Data Flow

M. Diefenthaler (Jefferson Lab)

The ePIC Streaming Computing Model



Successfully reviewed in 2023–2025 and documented

The ePIC Streaming Computing Model
Version 2, Fall 2024

Marco Battaglieri[1], Wouter Deconinck[2], Markus Diefenthaler[3], Jin Huang[4], Sylvester Joosten[5], Dmitry Kalinkin[6], Jeffery Landgraf[4], David Lawrence[3] and Torre Wenaus[4]
for the ePIC Collaboration

[1]Istituto Nazionale di Fisica Nucleare - Sezione di Genova, Genova, Liguria, Italy.
[2]University of Manitoba, Winnipeg, Manitoba, Canada.
[3]Jefferson Lab, Newport News, VA, USA.
[4]Brookhaven National Laboratory, Upton, NY, USA.
[5]Argonne National Laboratory, Lemont, IL, USA.
[6]University of Kentucky, Lexington, KY, USA.

Abstract

This second version of the ePIC Streaming Computing Model Report provides a 2024 view of the computing model, updating the October 2023 report with new material including an early estimate of computing resource requirements; software developments supporting detector and physics studies, the integration of ML, and a robust production activity; the evolving plan for infrastructure, dataflows, and workflows from Echelon 0 to Echelon 1; and a more developed timeline of high-level milestones. This regularly updated report provides a common understanding within the ePIC Collaboration on the streaming computing model, and serves as input to ePIC Software & Computing reviews and to the EIC Resource Review Board. A later version will be submitted for publication to share our work and plans with the community. **New and substantially rewritten material in Version 2 is dark green. The present draft is preliminary and incomplete and is yet to be circulated in ePIC for review.**

https://doi.org/10.5281/zenodo.14675920

The ePIC Streaming Computing Model is **physics-driven**. It has been developed to **accelerate the pace of discovery** and **enhance scientific precision** through streaming readout, AI, and distributed computing.

What the ePIC Streaming Computing Model Enables

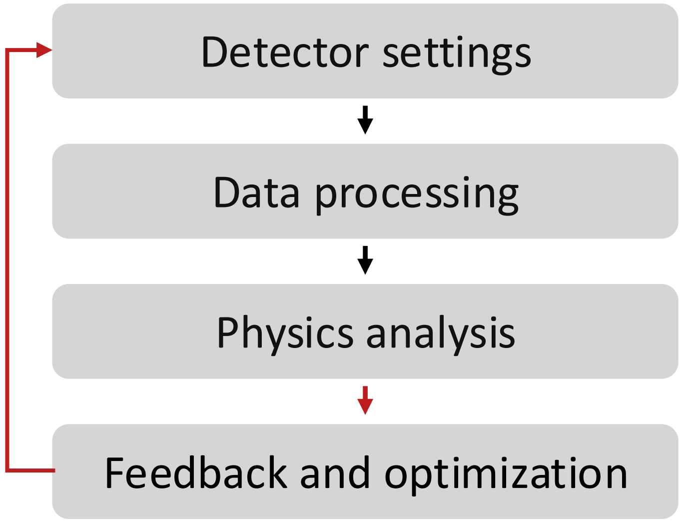
The ePIC Streaming Computing Model integrates streaming readout, AI, and distributed computing.

1. **Streaming Readout**
   Maximizes scientific reach through unbiased event selection and continuous measurement of backgrounds and detector noise.

2. **AI**
   Accelerates science through automated alignment, calibration, and validation, shortening the path to physics results and publications.

3. **Distributed Computing**
   Enables flexible access to dedicated and opportunistic computing resources where they are needed.

Closed-loop experimental control



Investing now in the ePIC Streaming Computing Model ensures that the collaboration will have the infrastructure, resources, and workflows needed to **commission and operate the experiment**.

Streaming DAQ and Computing Milestones

FY25	FY26	FY27	FY28	FY29	FY30	FY31	
PicoDAQ	MicroDAQ	MiniDAQ	Full DAQ-v-1	Production DAQ			DAQ
Streaming Orchestration			Streaming Challenges				
AI-Empowered Streaming Data Processing			Analysis Challenges				Computing
			Distributed Data Challenges				
AI-Driven Autonomous Calibration			AI-Driven Autonomous Alignment, Calibration, and Control				AI

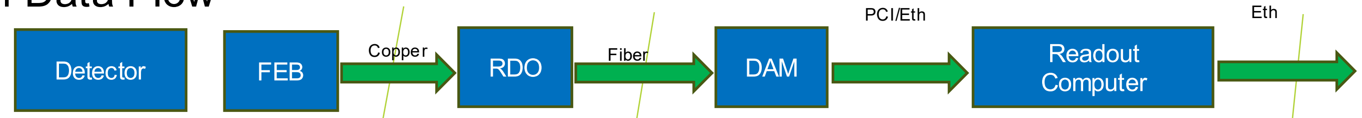
- **Compute-Detector Integration**:
  - Joint deliverables between DAQ and **computing** to develop an integrated system.
  - We define and test DAQ–computing interfaces to reduce integration risk and enable early system testing.
  - **FY28Q1 priorities**: Testbeds and prototypes for streaming orchestration and autonomous detector calibration for one detector system.

Summary of Channel Counts and Data Flow

Charge 1

Detector Group	Channels					Det Fiber Down	Det Fiber Up	RDO	Fiber Pair (DAQ)	DAM	Data Volume (RDO) (Gb/s)	Data Volume (To Tape) (Gb/s)
	MAPS	AC-LGAD	SiPM/PMT	MPGD	HRPPD/ MCP-PMT							
Tracking (MAPS)	16B					187	4976	323	323	7	15	15
Tracking (MPGD)				164k		640	2560	160	160	5	27	5
Calorimeters	500M		100k					522	522	17	70	17
PID (TOF)		6.1M				500	1364		1364	30	50	12
PID Cherenkov			318k		143k	1334	1334	1242	1334	33	1275	32
Far Forward		1.5M	10k					80	80	6	36	12
Far Backward	66M		3.4k					25	289	11	37	8
Lumi		128k	5.1k					41	41	4	264	8
Polarimetry	Independent Electronics, DAQ, & Controls from central detector but expected to build on same technologies											
TOTAL	**16.6B**	**7.7M**	**432k**	**164k**	**143k**	**2,661**	**10,234**	**2,393**	**4,113**	**113**	**1,774**	**109**

Summary of Data Flow



Aggregate	**2.0 Tb/sec**
Noise	1.6 Tb /sec
Signal from Physics + Background	400 Gb / sec

Aggregate	**2.0 Tb/sec**
Per RDO (Avg)	0.7 Gb/sec

Aggregate	**115 Gb/sec**
Collision Signal	62 Gb/sec
Synchrotron Rad	6 Gb/sec*
Electron Beam	4.5 Gb/sec
Hadron Beam	1.0 Gb/sec
Noise	41 Gb/sec

Scale of the system:

- **Electronics**
  - ~ 25 detector subsystems
  - ~ 5 Readout Technologies
  - ~ 2500 RDOs (on detector/in racks)
  - ~ 110 DAM boards (DAQ room)
  - GTU (with interface boards)
- **Maximum Data Volume**
  - ~ 2 Tb/sec digitized
  - ~ 115 Gb/sec recorded
- **Online Computing (Echelon 0)**
  - ~200 nodes (DAQ Room/SDCC)

* Synchrotron radiation caveats:

1. Rates are based upon hit rate for all ePIC detectors. In fact, data volumes depend upon specific detector hit (64 bits/hit assumed)
2. Highest Synchrotron radiation / electron beam gas will correspond to lower values for collision signal
3. Plan to analyze by component soon

E0-E1 Interface

E0 (Streaming DAQ)
- Data Aggregation
- Low-Level Filtering, e.g. zero suppression

E1 (Streaming Computing)
- Data archiving
- Event building and filtering

Monitoring

low-level **Calibration** high-level

- Full reconstruction
- Simulation
- Analysis

ePIC Streaming Computing Model

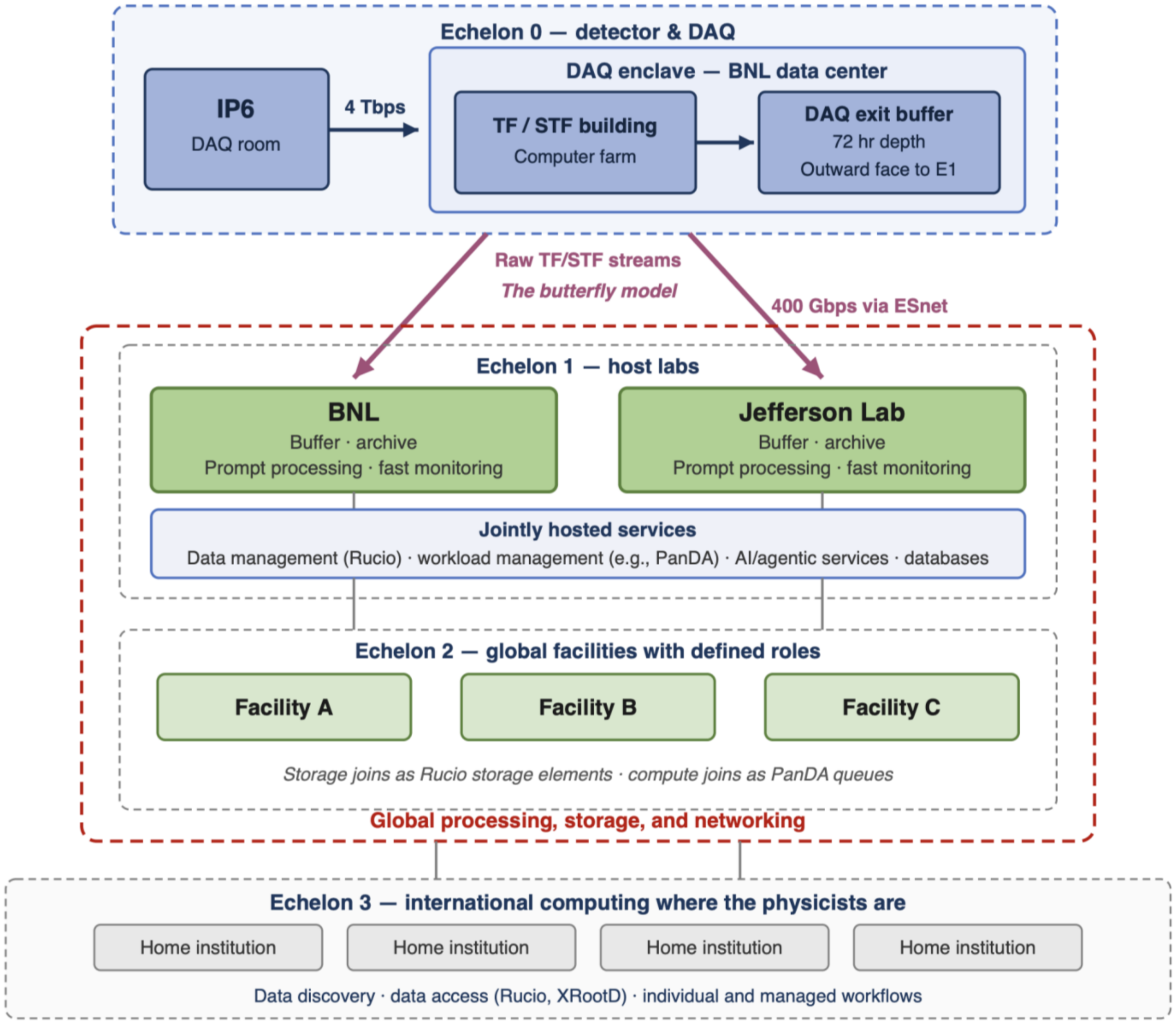
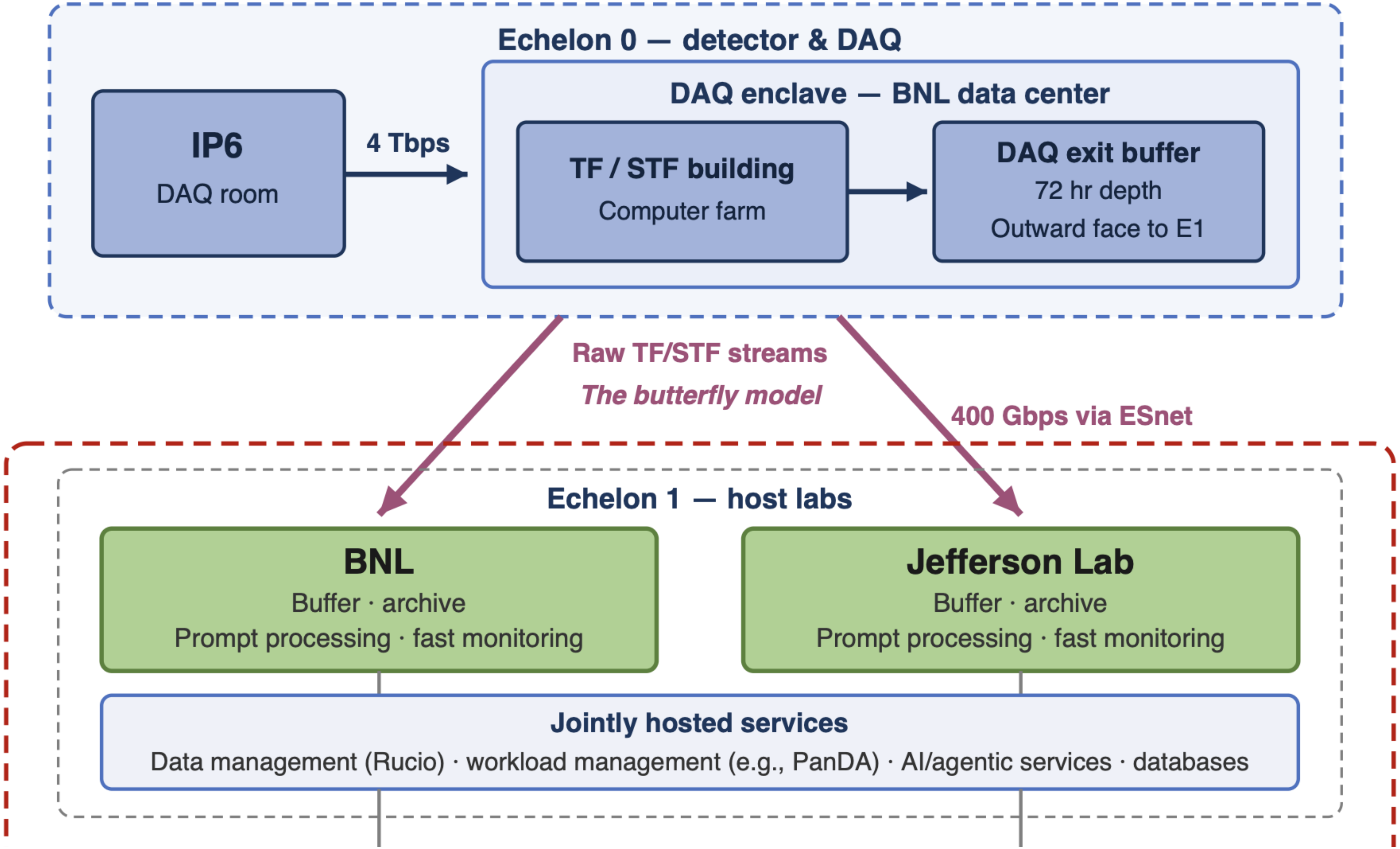
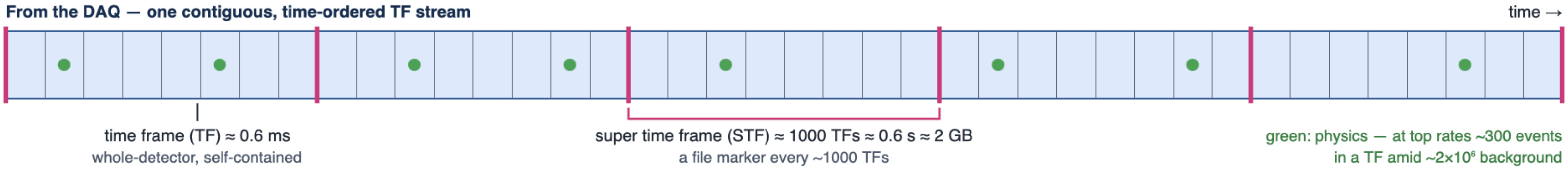


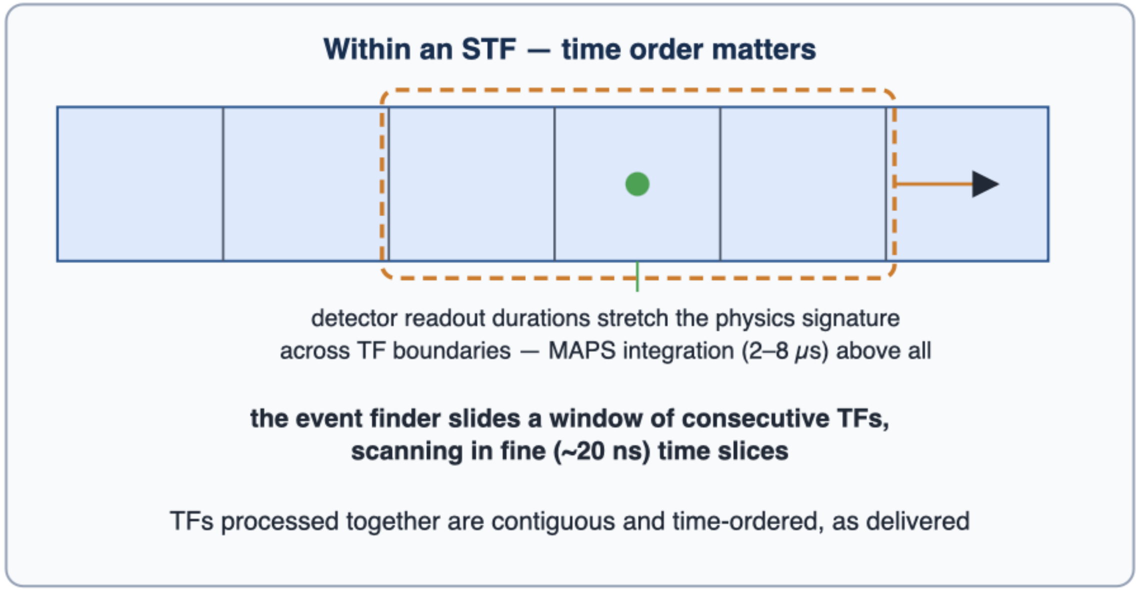
Diagram by Torre

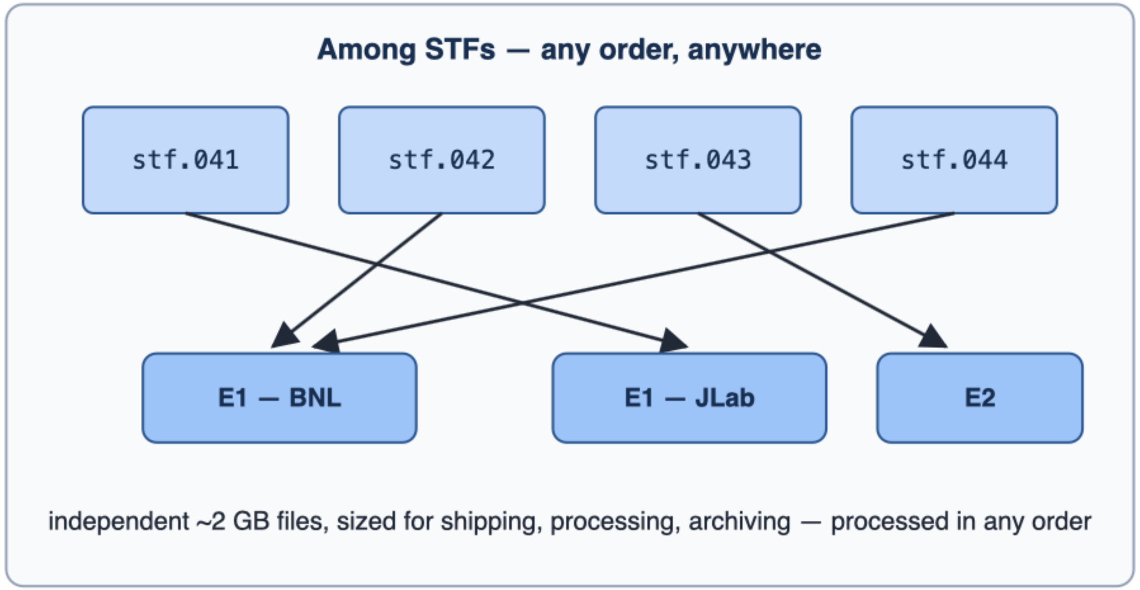
E0-E1 Interface



Timeframes and Supertimeframes







Next Steps and Open Questions

Next step

- Formalize the E0–E1 interface in the Streaming Computing Model report/publication.
- Discussed detailed technical notes if needed.
- **Target: September 2026**

Torre and his assistant summarized the recent discussions in a detailed note on The ePIC E0-E1 Interface.

Main open questions

- Raw-data and STF definitions and timeline
- Evolution of the state model and transition rules
- Latency requirements
- Calibration and conditions databases
- Information and control interfaces between E0 and E1
- AI readiness and AI-enabled control

Latency requirements: We are considering a 10-second latency for monitoring, but AI-assisted monitoring may require lower latency.