JLab Streaming Readout Activities

David Lawrence, JLab EPSCI group

Apr. 29, 2020 EIC YR SRO Meeting

The EPSCI Group at Jefferson Lab

Experimental **P**hysics **S**oftware and **C**omputing **I**nfrastructure

https://wiki.jlab.org/epsciwiki/index.php/Main_Page

- Group of 4-5 Physicists+Computer Scientists working on next generation software and computing technologies for ENP
- Under SciComp which is led by Graham Heyes (formally led Data Acquisition Group)
    - Carl Timmer and Vardan Gyurjyan, Physicists (former members of Data Acquisition Group)
    - Nathan Brei and Chris Larrieau, Computer Scientists
    - Thomas Britton and David Lawrence, Physicists
- SRO backend* is one of our major development projects

Mission Statement:

Identity, develop, implement, and maintain software and computing technologies in support of the Jefferson Lab Science Program.

**”backend” here means after the digitization and custom electronics*

Projects Driving JLab SRO

Experiment	Conditions	Event Rate	Data Rate	Comments
Moller	Production/ integrated mode	1920Hz	130MB/s	Can be handled with the traditional DAQ.
EIC	$L=10^{34}\ cm^{-2}s^{-1}$	450-550kHz (not including background noise rates)	20-25GB/s not included vertex tracker that will generate ~240GB/s	~10kHz/µb, track multiplicity = ~5 JLAB EIC detector design will have millions of channels. Only non-vertex detectors combined will have ~1M channels plus vertex detector: estimated 20-50M channels. In total ~1000 ROC's. Control nightmare (starting stopping a run). Streaming readout has less control requirements.
TIDIS		rTPC hit rates enormous (~800KHz/pad)	4GB/s	How to match up super Bigbyte detected electrons with rTPC detected spectator protons is a big question. Conventional triggered DAQ will be challenged.
SoLID	30 sector GEM		30GB/s	30 separate DAQ's each 1GB/s? How to combine GEM readout with other detectors? Handling GEM hits sharing adjacent sectors.
CLAS12	Phase 2	100KHz	5-7GB/s	

EIC Data Rate Considerations

- Initial signal processing on Front End FPGA
  - Higher intensity -> more pile up -> difficulty extracting hits
  - Complex algorithm on FE FPGA vs. downstream compute
    - Signal Processing on Front End reduces everything needed downstream (e.g. network)
    - FE is harder to scale up later. Downstream compute could include FPGA
- Final Detector Design
  - Details not finalized. (Vertex tracker?)
    - TOPsiDE
    - BeAST
    - JLEIC
    - sPHENIX
- Raw data vs. reconstructed vs. software filtered
  - Raw data O(~100Gbps)* - achievable with today's technology
  - Reconstructed (~1/10 of raw data) - requires large storage system to give time for calibrating
  - Software Triggered
    - Save raw data for filtered events (imperfect calibration needed)
    - Save only reconstructed values

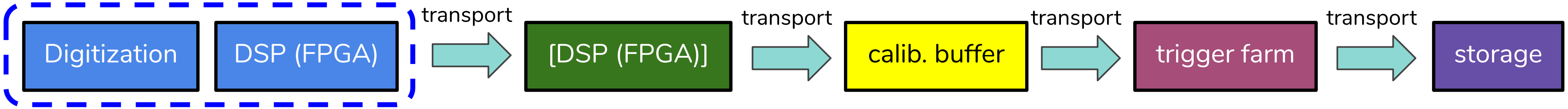
*See slide 30 of Jin Huang's talk https://indico.bnl.gov/event/5807/contributions/26937/attachments/21875/30184/EIC_DAQ_Streaming_Meeting.pdf

Data Transport

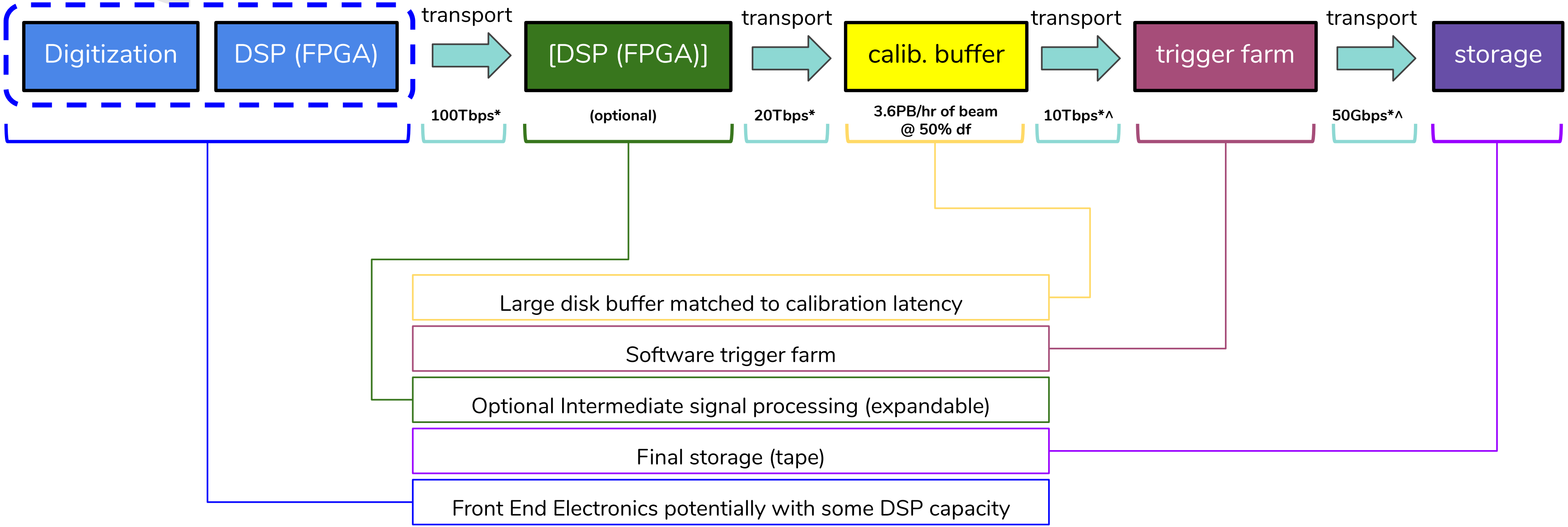
- Support of multiple technologies via transport layer
  - ADIOS from ORNL (testing currently underway)
    - MPI
    - RDMA
    - ZMQ
    - Sockets
  - Custom interface library
- Data formatting
  - Jan B., Markus D., Dmitry R.
    - Streaming Readout IV
      - https://agenda.infn.it/event/18179/contributions/89844/attachments/63450/76388/proto.pdf
  - Graham Heyes - (shhh...)
  - Serialization
    - Google Protocol Buffers / Flat Buffers (xMsg - JLab)
    - ProIO (ANL) (nothing done with this yet at JLab)
    - ROOT
    - Apache Arrow (xMsg, ...)

Workflow Management

- SRO system necessarily requires multiple processes distributed across several (many) compute nodes and devices
  - Some components are unique to dataflow diagram and required for system to operate
    - e.g. Event Recorder
  - Some components are farm-like and the system may still function if one or two disappear
- Workflow requirements
  - components be configured centrally
  - fault tolerant (redeploy components/services if they disappear)
  - performance monitoring and logging
    - alarm system
  - application elasticity (apply more/less resources as needed)
    - e.g. trigger farm uses portion of general purpose farm



Workflow Management



Digitization

DSP (FPGA)

transport

100Tbps*

[DSP (FPGA)]

(optional)

transport

20Tbps*

calib. buffer

3.6PB/hr of beam
@ 50% df

transport

10Tbps*^

trigger farm

transport

50Gbps*^

storage

Large disk buffer matched to calibration latency

Software trigger farm

Optional Intermediate signal processing (expandable)

Final storage (tape)

Front End Electronics potentially with some DSP capacity

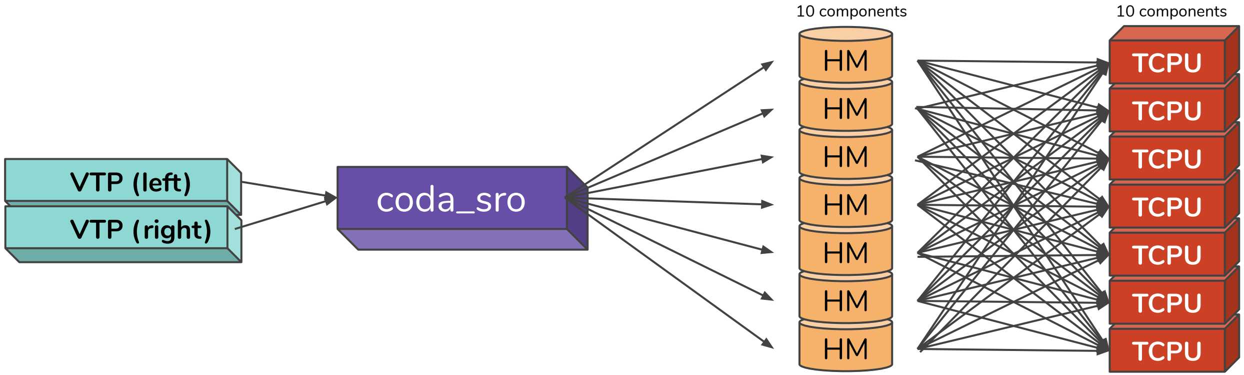
* See slide 2 of Jin Huang's talk https://indico.bnl.gov/event/6383/contributions/32820/attachments/25507/38249/EIC_DAQ.pdf

^ Assume 50% beam availability integrated out via calibration buffer

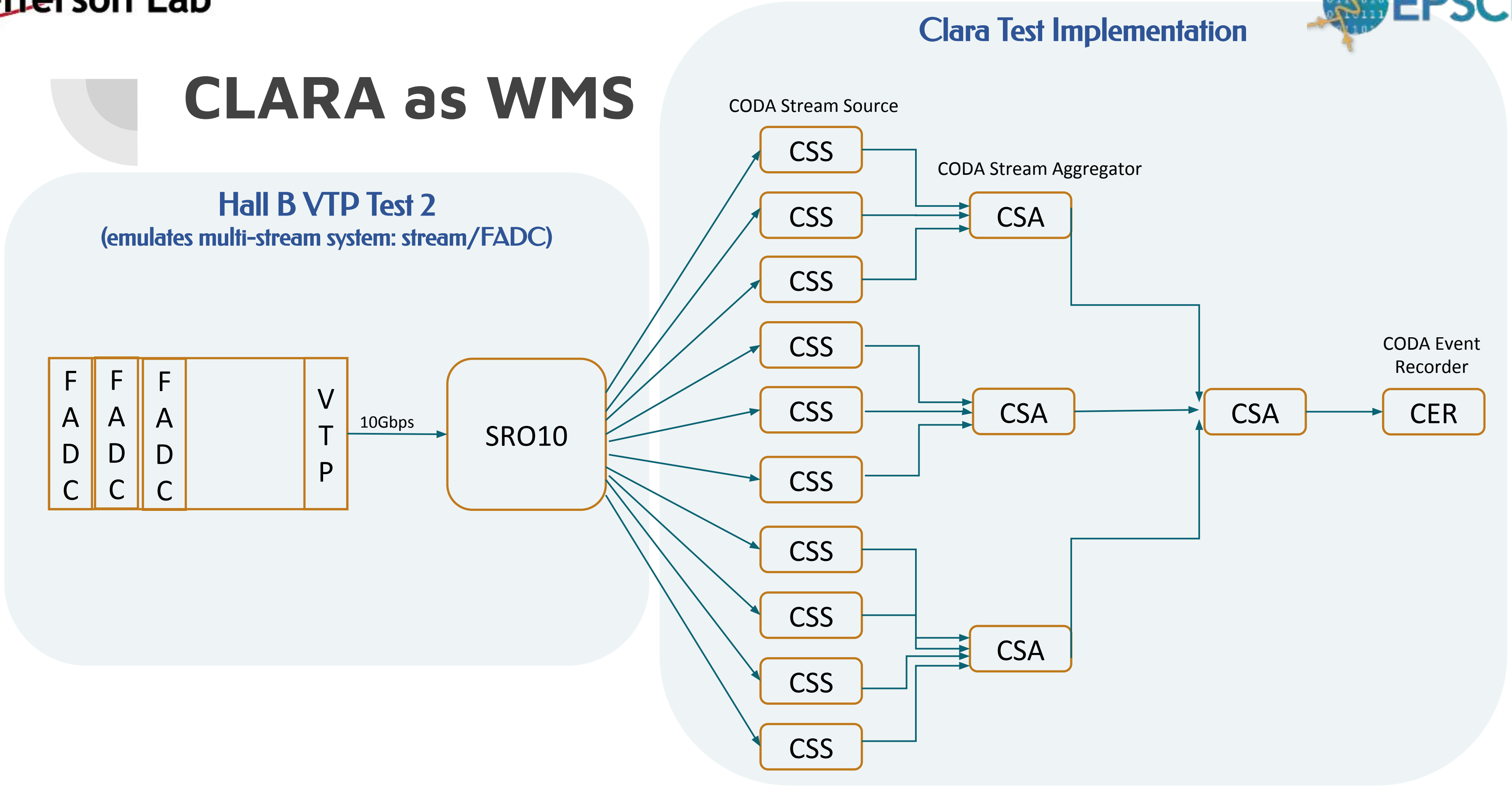
TriDAS Testing

see: https://agenda.infn.it/event/18179/contributions/89843/attachments/63451/76396/EIC-Stream_Readout-Camogli_20190524_chiarusi.pdf

- TriDAS system testing in Experimental Halls B and D at JLab
  - Existing Flash-ADC systems using VTP module with high speed VXS interconnects
  - Multiple testbeds currently available (Sergey B.)
  - Supports multi-node, multi-process and multi-threaded scaling options
    - integrated JANA2 for triggering
  - Only preliminary testing done so far at JLab
    - expect more stress testing over coming months
- Open issues
  - System designed for deep sea neutrino experiments (how well does it scale?)
  - Overall process management



CLARA as WMS

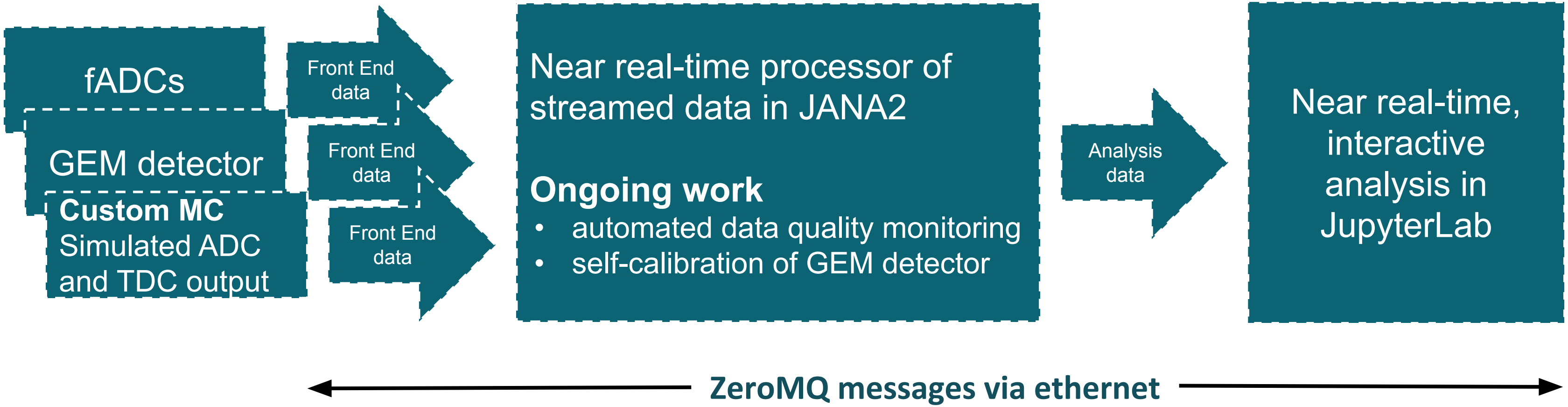

Clara Test Implementation
Hall B VTP Test 2
(emulates multi-stream system: stream/FADC)
FADC
FADC
FADC
VTP
10Gbps
SRO10
CODA Stream Source
CSS
CSS
CSS
CSS
CSS
CSS
CSS
CSS
CSS
CSS
CODA Stream Aggregator
CSA
CSA
CSA
CSA
CODA Event Recorder
CER

INDRA-ASTRA: Seamless integration of DAQ and analysis using AI

LDRD goal

prototype components of streaming readout at NP experiments
→ integrated start to end system from detector read out through analysis
→ comprehensive view: no problems pushed into the interfaces

prototype (near) real-time analysis of NP data
→ inform design of new NP experiments



fADCs

GEM detector

Custom MC
Simulated ADC and TDC output

Front End data

Front End data

Front End data

Near real-time processor of streamed data in JANA2

Ongoing work
- automated data quality monitoring
- self-calibration of GEM detector

Analysis data

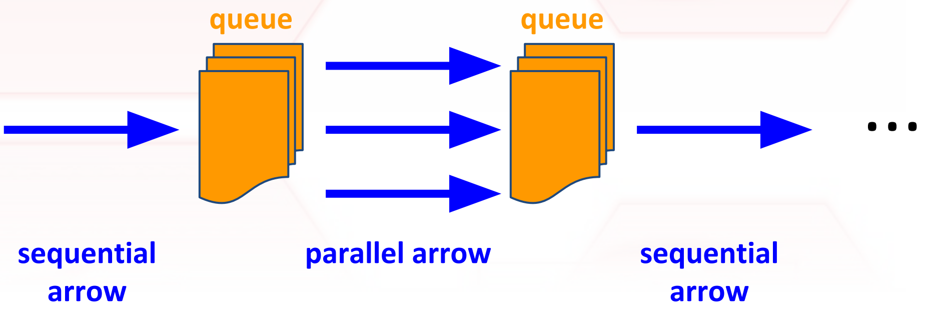
Near real-time, interactive analysis in JupyterLab

ZeroMQ messages via ethernet

Software Trigger Support with JANA2

https://jeffersonlab.github.io/JANA2/

- JANA is a multi-threaded event reconstruction framework
  - Original version in use for ~15years by GlueX
  - JANA2 mostly a rewrite that modernizes to newer C++ standards
  - Still under active development and very receptive to implement SRO support features
  - Available right now for production use
- Seamlessly integrates offline reconstruction software into Software Trigger and DQM systems
  - On-demand algorithm application naturally optimizes CPU usage in software trigger
  - Arrow-Queue design pattern couples sequential I/O with parallel algorithm execution in a “lockless” way
  - Support for parallel I/O
  - ZMQ source exists



My Opinions

1. **Is software event triggering and building necessary?**

   If Jin's numbers hold then, no.

   Should we do it? yes

   There is no technical reason for us not to and it will

   a. reduce the time it takes to start real physics analysis

   b. position us for potential detector upgrades without redesigning DAQ

2. **Could we drop raw data and keep only the reconstructed part?**

   Yes, but if the cost of keeping both is relatively small then don't. It mitigates risk considerably if you can keep the raw data as well. It does not necessarily need to be kept online and easily available (e.g. tape vault)

3. **Same question concerning the event selection?**

   Same answer.

4. **What sort of calibrations would be necessary?**

   It is conceivable to record raw data but only for events triggered based rough calibrations. Details on detector design and event topologies would be needed to answer this.

Summary

- Newly formed EPSCI group is now working to help develop SRO software for multiple future experiments
    - Close association with electronics group at JLab
    - Eager to contribute to EIC SRO efforts

- R & D
    - Data transport
    - Workflow Management
    - Microservices (CLARA)
    - Software triggering + DQM (JANA2)
    - VTP-based systems

- Other areas we can contribute
    - Data format
    - A.I. algorithm development and deployment (e.g. FPGA)

- Multiple talks scheduled for Streaming Readout VI workshop in May