OAK RIDGE National Laboratory

Streaming Readout and Data Reduction at ORNL

Joe Osborn

April 29, 2021

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

OAK RIDGE National Laboratory

Oak Ridge National Laboratory

- Oak Ridge is one of the large, multipurpose laboratories in the DOE laboratory family
- Recently at Oak Ridge, synergies between the physics and computing divisions have started to form
- Computing specialists in software, data reduction, and streaming readout discussing with physics division on possible collaboration looking towards EIC
- In this talk, I'll discuss some of the broad ranging data reduction work going on at the laboratory

Disclaimer

- I am speaking on behalf of many colleagues at ORNL! To name a few:
  - From physics : Ken Read, Jo Schambach, Friederike Bock *et al.*
  - From computing : Scott Klasky, Jason Wang, Norbert Podhorszki *et al.*

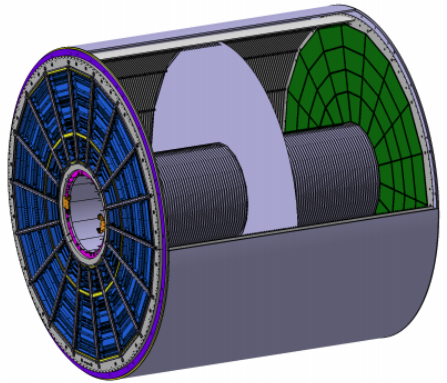
Physics Division

Overview

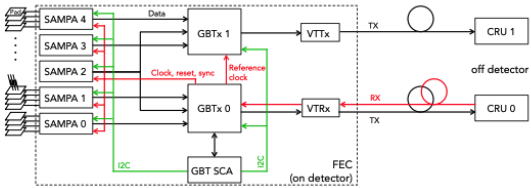
- Oak Ridge physics division has a strong background in electronics
- Recently published work on upgrade of ALICE Time Projection Chamber (TPC) for Large Hadron Collider (LHC) Run-3 and 4 (JINST 16 (2021) 03, P03022, arXiv: 2012.09518)
- Major contributor to ALICE inner tracking system (ITS)
- Current work on testing sPHENIX vertex detector readout and electronics

ALICE TPC

- ALICE TPC upgrade intended to handle 50 kHz Pb+Pb rate at LHC Runs 3-4
- Rare probes at low momentum $\rightarrow$ desire for continuous readout
- TPC readout time window vs nominal rate $\rightarrow$ continuous readout
- Expected data rates of $\sim$3 TB/s from $\sim$500k channels
  - Increase of data rate by $\mathcal{O}(100)$ compared to previous TPC
- Immense readout challenge requiring R&D for successful physics



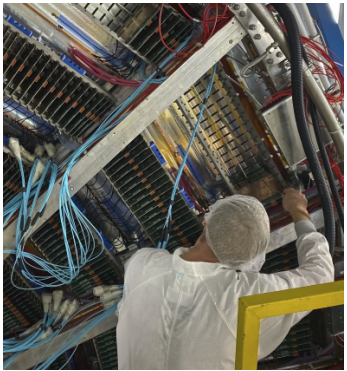
TPC Readout Scheme

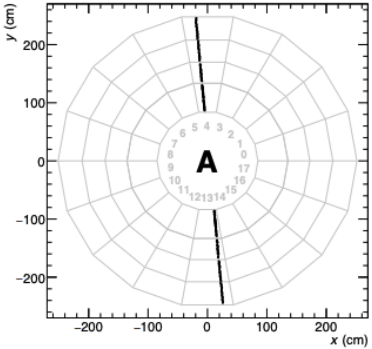


Front-end-card (FEC)

- TPC readout with ∼3000 front end cards, containing 5 SAMPA chips which perform signal shaping/ADC/DSP
  - Each SAMPA produces 1.6 Gbit/s data rate, leading to the ∼3 TB/s data rate of the detector as a whole
- 360 FPGA (CRU) cards receive the data and perform online processing
- RX link provides clock to the FEC, TX links transmit data

Detector Assembly

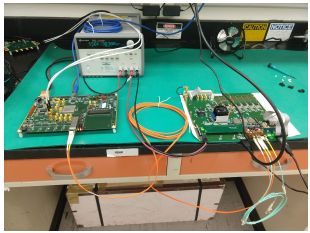


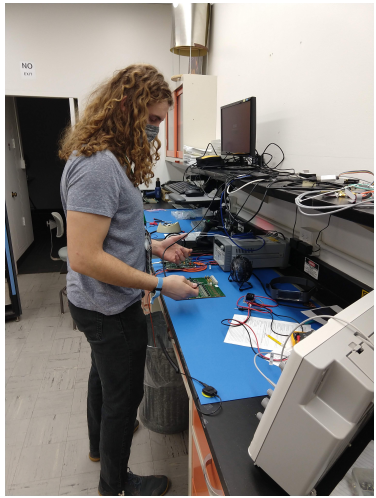


- Detector assembled and readout has been fully integrated into the experimental cavern
- Commissioning performed with x-rays, lasers, cosmics - shows good performance
- Continued testing and commissioning in preparation for LHC Run-3 expected

MVTX at sPHENIX

- ORNL also leading readout testing and development for the sPHENIX MVTX
- Setting up readout chain test stand in lab
- Streaming readout will be utilized at sPHENIX as well, with silicon+TPC (see Martin Purschke's talk on Wednesday, Takao Sakaguchi and Yasser Morales talks this morning)





Pictures courtesy of Jo Schambach

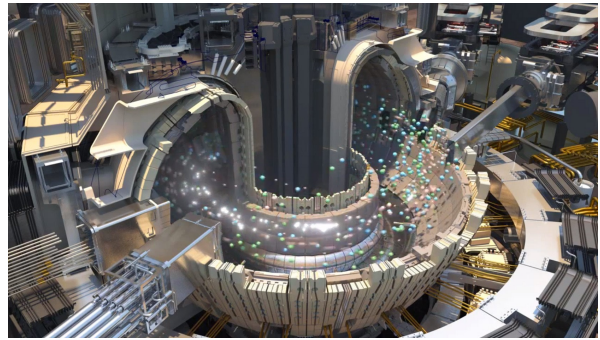
Computing Division

Overview

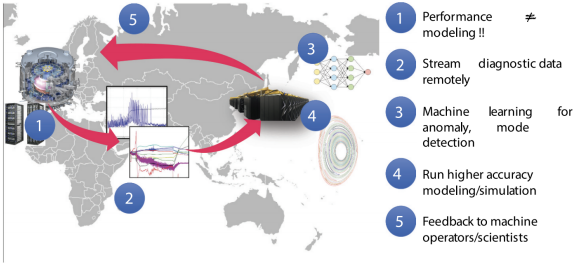
- ORNL computing is contributing to streaming readout and improved data reduction workflows in many areas
- Expertise available with Summit and CADES computing centers on data reduction
  - Summit (and Frontier, upcoming) are ORNL's flagship super computers - GPU focused
  - CADES is a compute center with cloud computing, storage, high speed data transfer nodes, and CPU/GPU nodes available for analysis
    - CADES used as tier 3 data analysis cluster for ALICE
- Additionally, dedicated contributions to improving data workflows at projects like ITER and the Square Kilometer Array (SKA), see :
  - Fusion Science and Technology, vol. 77, no. 2, pp. 98–108, 2021
  - SC '20: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, no. 2, pp. 1–12 (2020)

ITER and Fusion Experiments

- ITER is an international fusion project aiming to research fusion's applicability as a clean energy source
- Projected to produce approximately $\sim$1 PB per day
- Necessitates using large scale data movement and federated computing for data processing
- Additionally, near real time analysis to guide experimental operation strongly desired

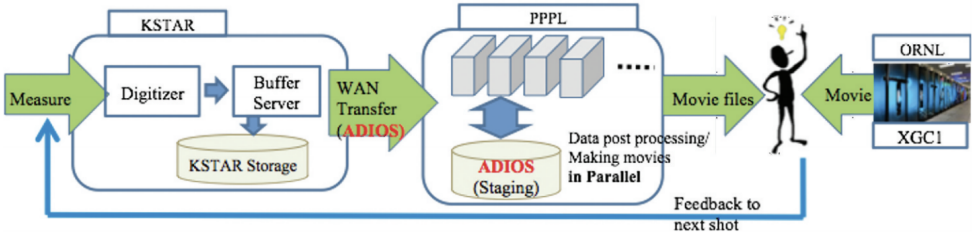


Example Workflow at Fusion Reactors



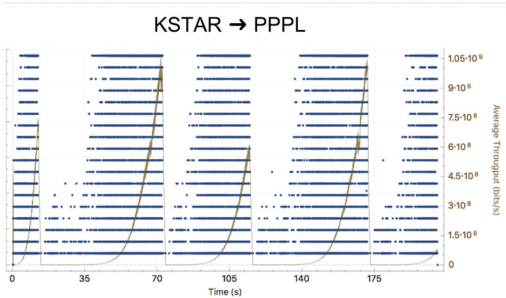
1. Compare performance locally to pre-run simulation
2. If performance does not match expectations from simulation, stream data to remote HPC
3. Use trained ML models to detect anomalies
4. Use ML information to run higher accuracy, more expensive simulations to understand discrepancy
5. Send diagnostic information back to scientists to guide next pre-run, simpler, simulation for performance diagnostics

Testing Workflow at KSTAR and PPPL



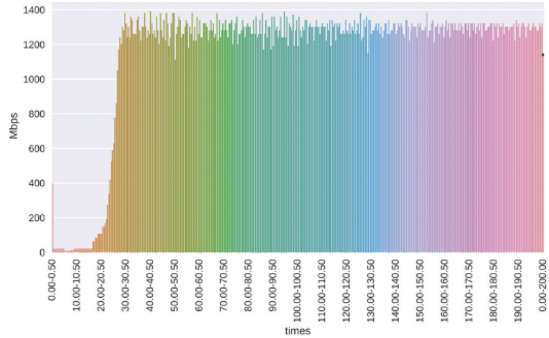
- ITER does not have data yet, so an example test is performed at KSTAR and Princeton Plasma Physics Laboratory (PPPL)
- Goal to demonstrate streaming data and compare to previously completed simulations

Testing Workflow at KSTAR and PPPL



- ITER does not have data yet, so an example test is performed at KSTAR and Princeton Plasma Physics Laboratory (PPPL)
- Goal to demonstrate streaming data and compare to previously completed simulations
- Previous throughput suffered from disruptions due to packet loss

Testing Workflow at KSTAR and PPPL

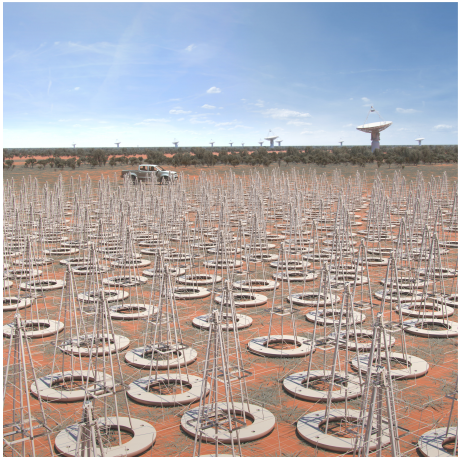


- ITER does not have data yet, so an example test is performed at KSTAR and Princeton Plasma Physics Laboratory (PPPL)
- Goal to demonstrate streaming data and compare to previously completed simulations

- New software achieves high, sustained data transfer throughput of $\sim$1.2 Gbps

Square Kilometre Array

- Square Kilometre Array (SKA) is a radio-astronomy experiment being built in Western Australia and South Africa
- Expected that the arrays will produce 5.2 Tb/s and 4.7 Tb/s data rates
- Will result in extremely large data sets that need to be calibrated, reduced, etc.

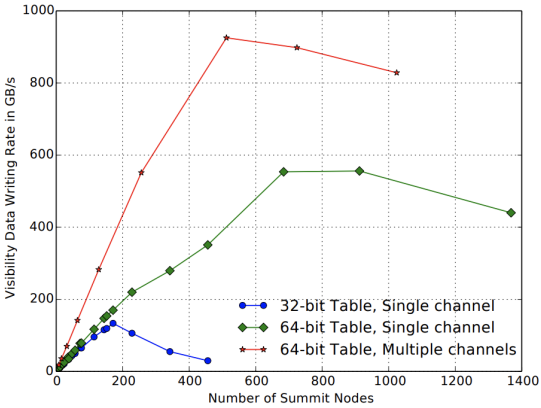


SKA Data Challenges

- SKA data rate will be 1-4 orders of magnitude larger than current telescopes, while SKA2 will be 1-2 orders of magnitude larger than SKA1
- Millions of files for a typical observation requires robust filesystem, metadata management, archiving, etc.
- SKA data must be simulated since the experiment is still being designed and built, so simulations were performed to mock the data rates

Telescope	Stations	Baseline (km)	Channels	Input rate (Gbps)	Deployed PFLOPS
MWA2	128	5	3,072	8	$\ll 0.1$
VLA	27	35	16,384	0.5	$\ll 0.1$
LOFAR	51	1,300	62,464	200	<0.1
ASKAP	32	6	16,384	23	0.2
SKA1-Low	512	65	65,536	4,700	125

Summit for SKA Processing



- Summit at ORNL was used to simulate data generation and processing at similar scales of SKA
- Table writing performance for several configurations shown, reaching peak of $\sim$0.9 TB/s
- Peak writing rate is configuration dependent - work ongoing

Conclusions

- ORNL is involved in several experiments world wide that are facing or will face data read out and reduction challenges
  - ALICE TPC at the Large Hadron Collider
  - sPHENIX MVTX at the Relativistic Heavy Ion Collider
  - ITER fusion experiment
  - SKA radio-astronomy experiment
- ALICE TPC commissioning at the LHC ongoing - detector already installed
- sPHENIX MVTX readout testing and development ongoing
- Test workflow with KSTAR and PPPL achieved high, sustained throughput
- SKA workflow demonstrated on Summit supercomputer at ORNL