Modeling Resource Utilization of a Large Data Acquisition System

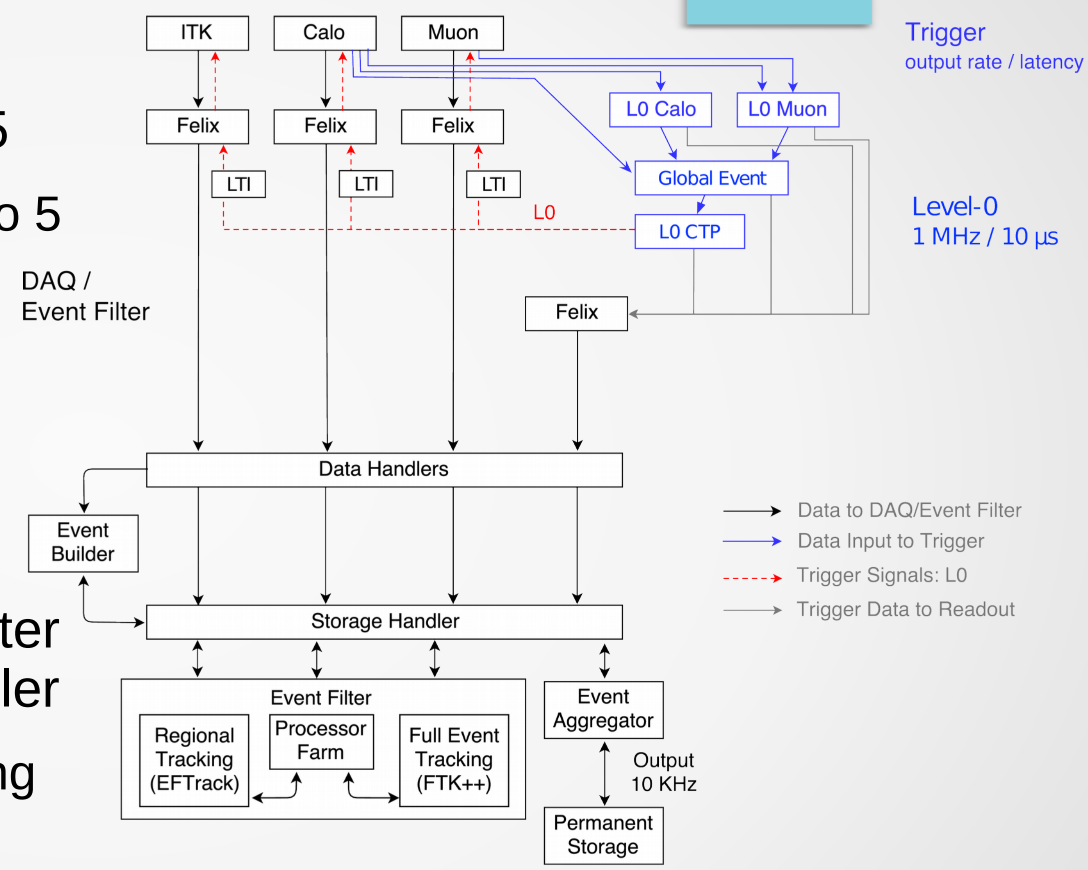
Alejandro Santos
CERN / Ruprecht-Karls-Universität Heidelberg
On behalf of the ATLAS Collaboration

Outline

- Introduction
- ATLAS TDAQ
- Simulation model
  - Implementation
  - Results
- Conclusion

Introduction

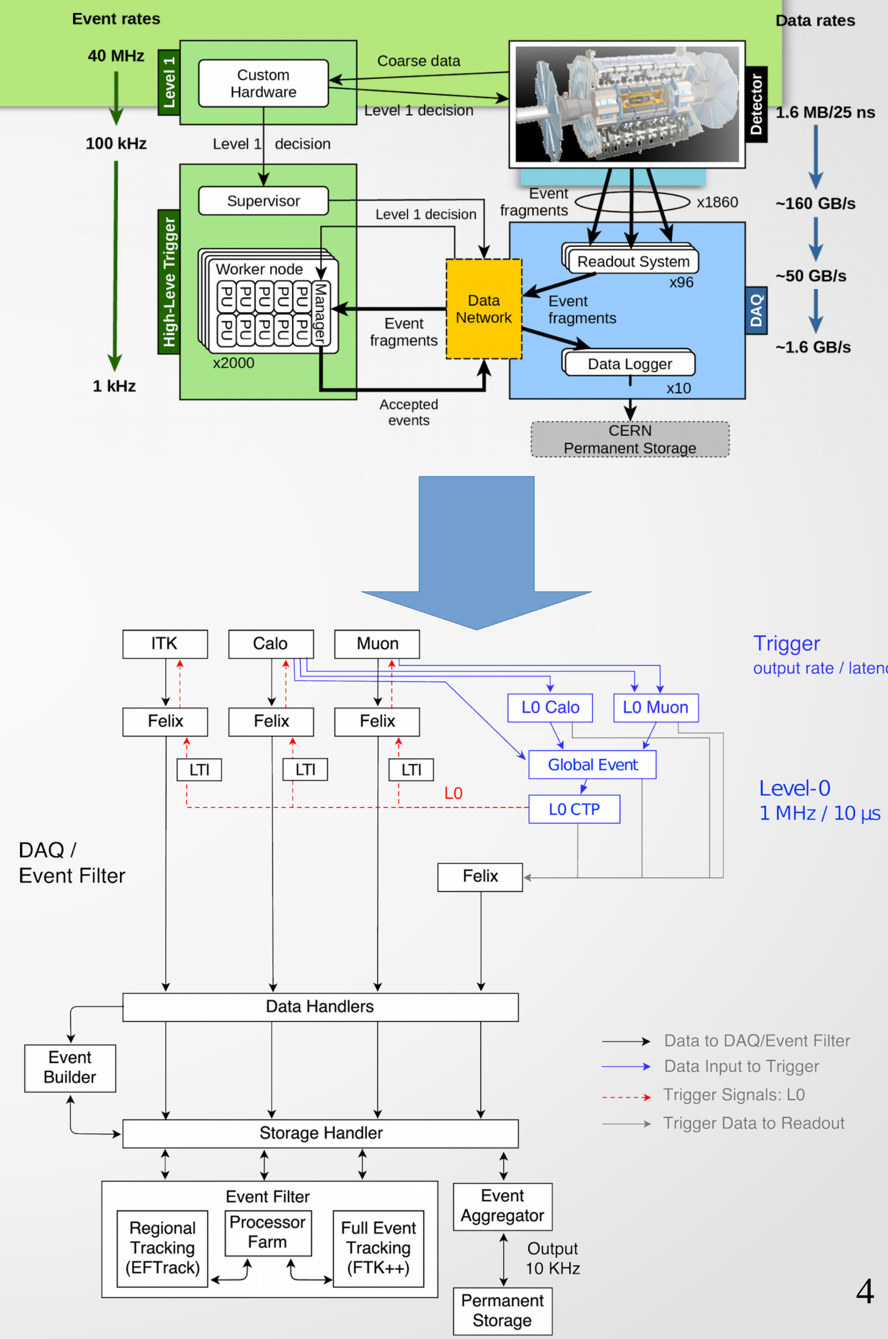
- ATLAS system after Phase-2 upgrade, scheduled for 2024-2025
  - DAQ challenges, e.g. 150 GB/s to 5 TB/s for input data rates
  - New physics requirements
- Great uncertainty on future technology and availability, very difficult to make predictions.
- Key component of DAQ system after Phase-2 upgrade is Storage Handler
  - Temporary buffer space decoupling data production from data processing.



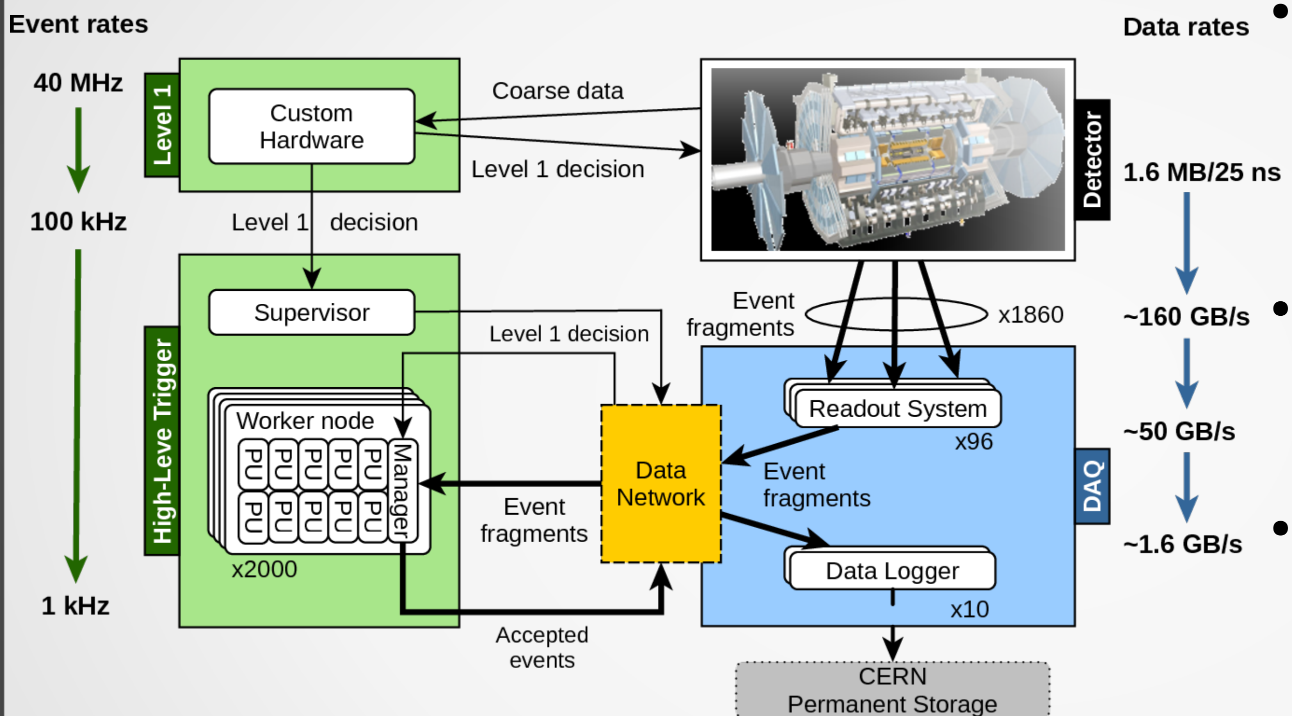
ATLAS Phase-2 TDAQ Baseline Architecture

Plan

- Model current ATLAS TDAQ system in simulation environment
- Build confidence in the model
  - Compare simulation results against real data
- Evolve the model toward the architecture for Phase-II upgrade
  - Simulation results for this model will help us to choose technology and design for the new architecture



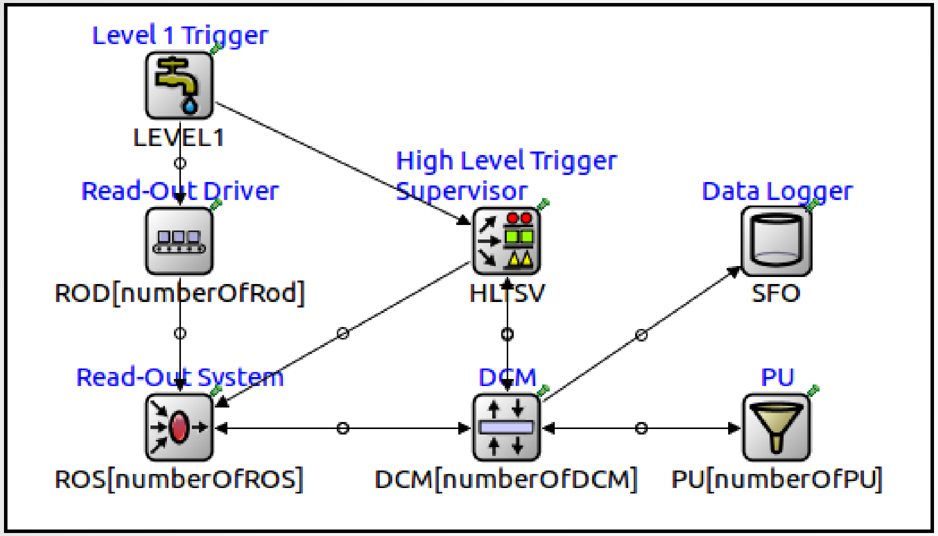
ATLAS TDAQ in Run 2



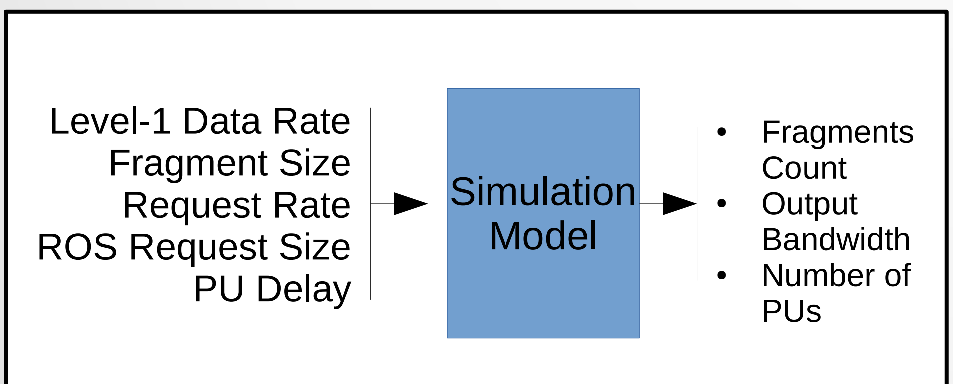
ATLAS Trigger and Data Acquisition (TDAQ) system

- Real-time heterogeneous computing system
  - Transport and filter ATLAS event data
- Events processed in parallel in a computer farm
- Readout System (ROS) stores and provides data during filtering
  - Similar role to Storage Handler after Phase-2 upgrade

Simulation Model (1/2)



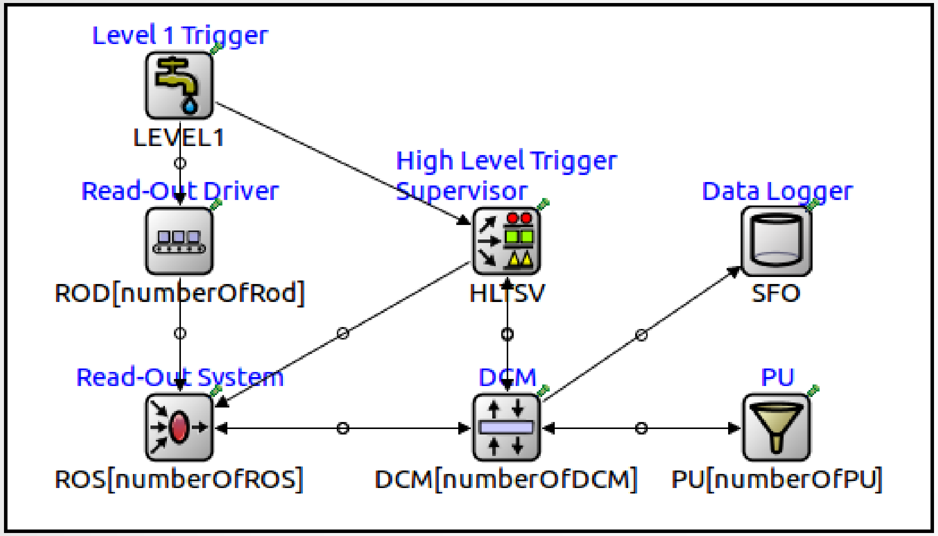
OMNeT++ Simulation Model



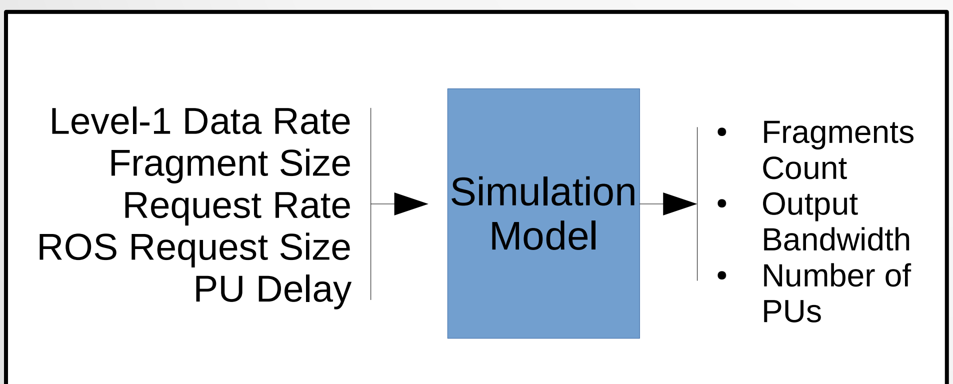
Simulation Overview

- Simulation model
  - Simplified version of TDAQ system
  - Network is assumed to be infinite and ideal, no packet loss
- Implemented in OMNeT++
  - Robust and user-friendly discrete event simulation framework
  - Configuration files are simple
  - Modular environment with a graphical and command line interface for running many simulations in parallel

Simulation Model (2/2)



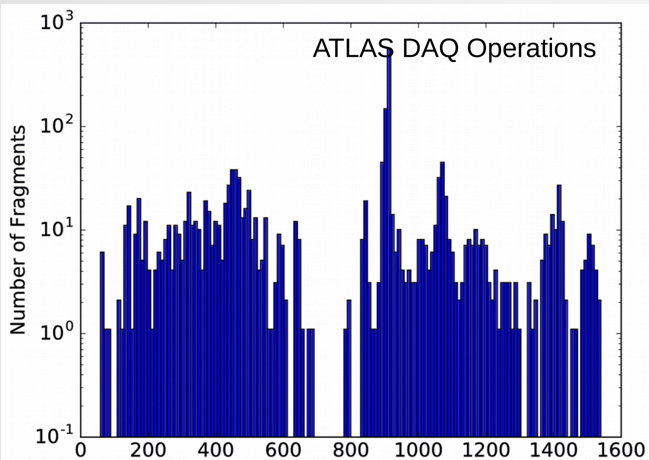
OMNeT++ Simulation Model

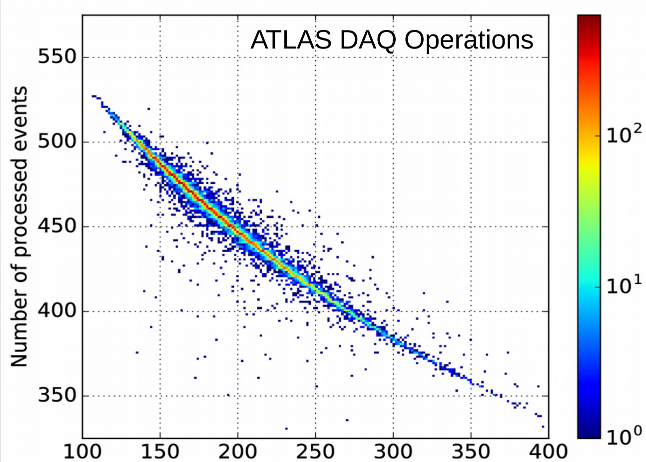


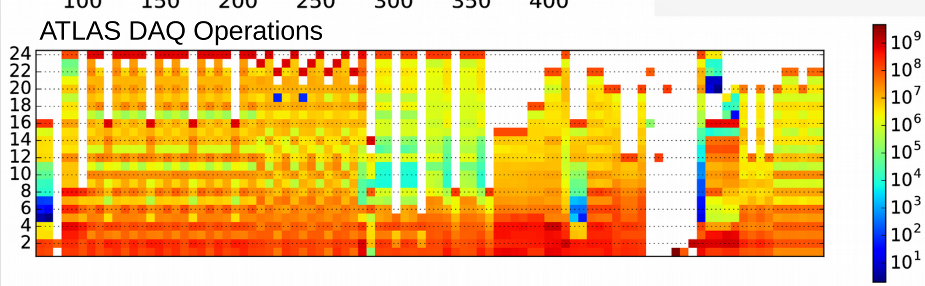
Simulation Overview

- Input values
  - Five minute averages of real operational data
- Output values
  - Number of Event fragments in the buffers of the ROS
  - Output bandwidth from the ROS to the PUs
  - Number of processing PUs

Simulation Data Input





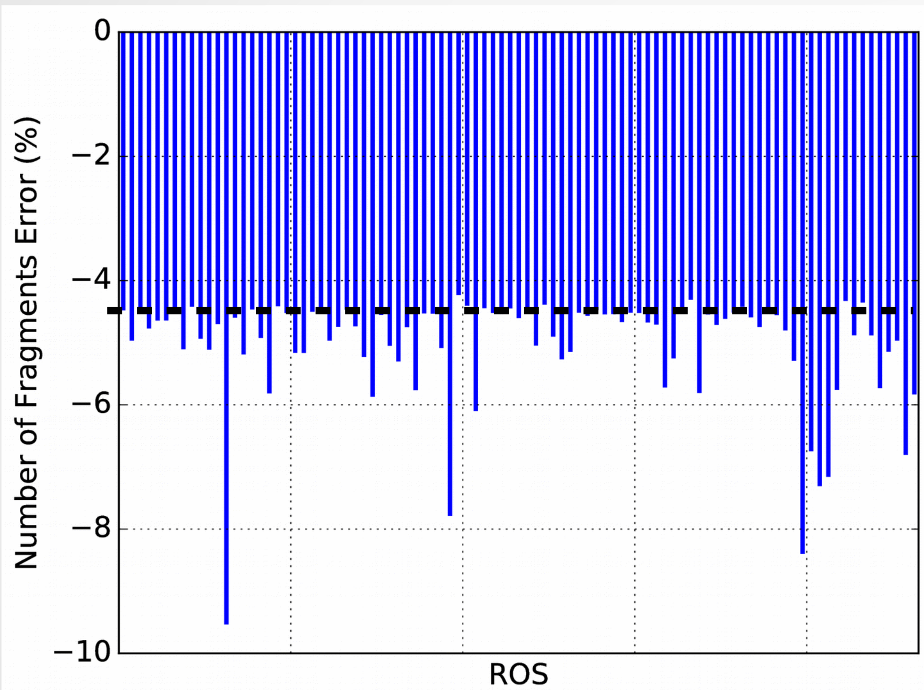


- Configuration and model data for the simulation is obtained from ATLAS databases
  - Records historical operational state
- Input values to simulation are average values over five minutes of real data.
  - Some values are already averages, need to re-normalize values

Simulation performance

- Simulations are run independently
- Each simulation is run for 60 simulated seconds in ~6hs (factor of 360) and takes ~2 GB of RAM.
  - Give enough time for simulation to warm-up
- And run on 4 independent machines, Intel Xeon E5645 2.4 GHz with 24 GB of RAM

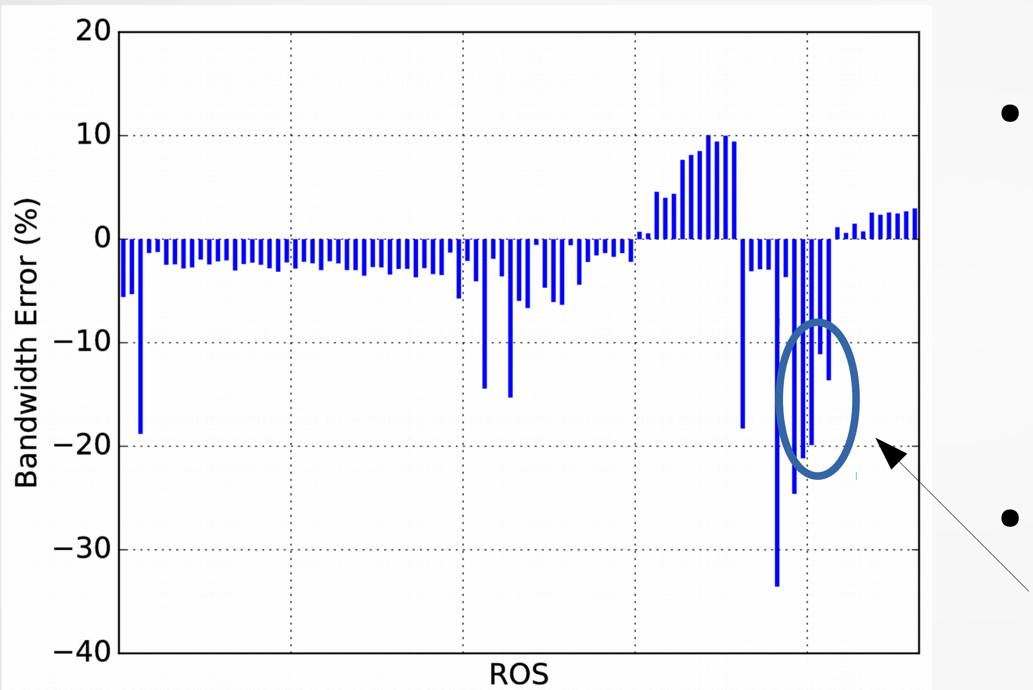
ROS Occupancy



Difference (in %) between simulation and real data for the number of fragments in ROS buffers. Each bin in the plot represents a separate ROS computer, there are ~100 ROS computers

- Common ~4% shift between simulation and real data
  - Model only includes processing latency
  - Network latency and software latency are not added to the model
  - This gives an additional ~10ms latency

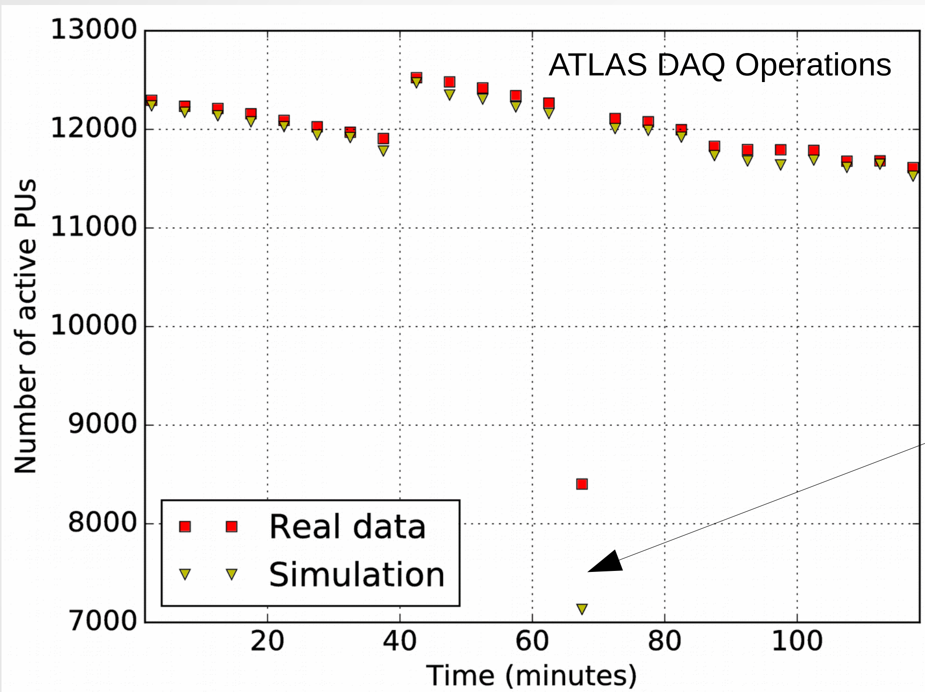
ROS Bandwidth



Difference (in %) between real data and simulation results for ROS output bandwidth. Each bin in the plot represents a separate ROS computer, there are ~100 ROS computers

- Most results are within 5% below real data
- Largest outlier is a ROS with a very small fragment size. Small changes in absolute values will have a large impact in error percentage
- TCP retransmissions and network protocol overheads are not modeled, and explain further differences

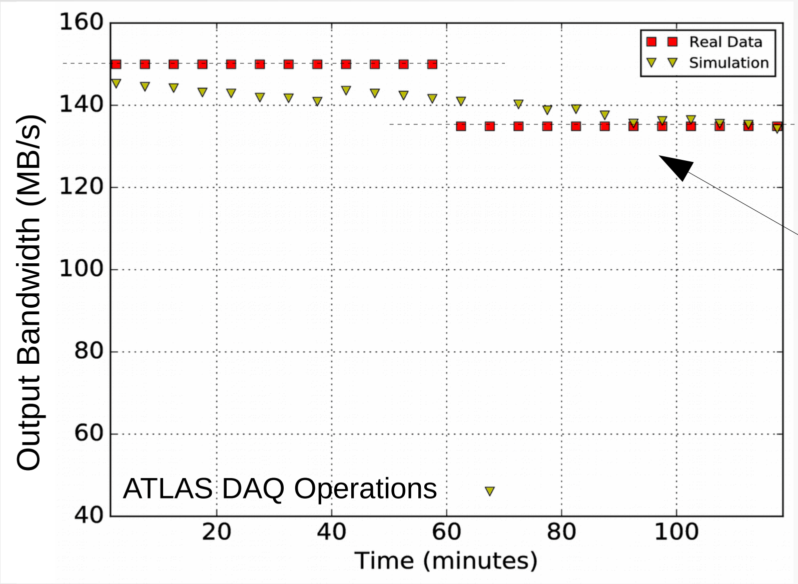
Simulation stability (1/2)

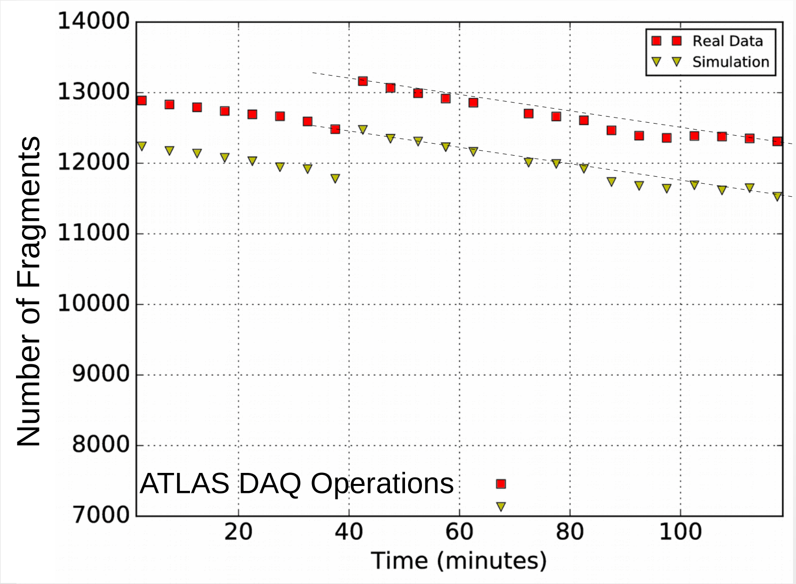


Each point in the plot represents one simulation. There are 24 simulations covering 2 hours of data.

- Analyze model result over large time window (2 hours)
- Good, stable agreement
- Outlier at minute ~70
  - Data-acquisition stopped due to external factors
  - Model assumes constant conditions
  - Conditions changed and simulation was not able to keep up with change

Simulation stability (2/2)





- Overall good agreement to within 5%
- ROS real data for bandwidth is stored as the average for one hour, data does not have better resolution
- Constant difference
  - Systematic difference → missing elements in the model

Conclusion

- A simulation model has been developed for studying the behavior of the current ATLAS TDAQ system
- Results produced by this model are in good agreement with the real information recorded during the second ATLAS run period
- Simulation results can be further improved by adding accurate simulation of the TDAQ system and network latencies to the model
- Upgrade of the TDAQ system architecture is planned for the years 2024-2025:
  - The simulation model can be used as the basis to studying the behavior of the future candidate architectures for the new TDAQ system