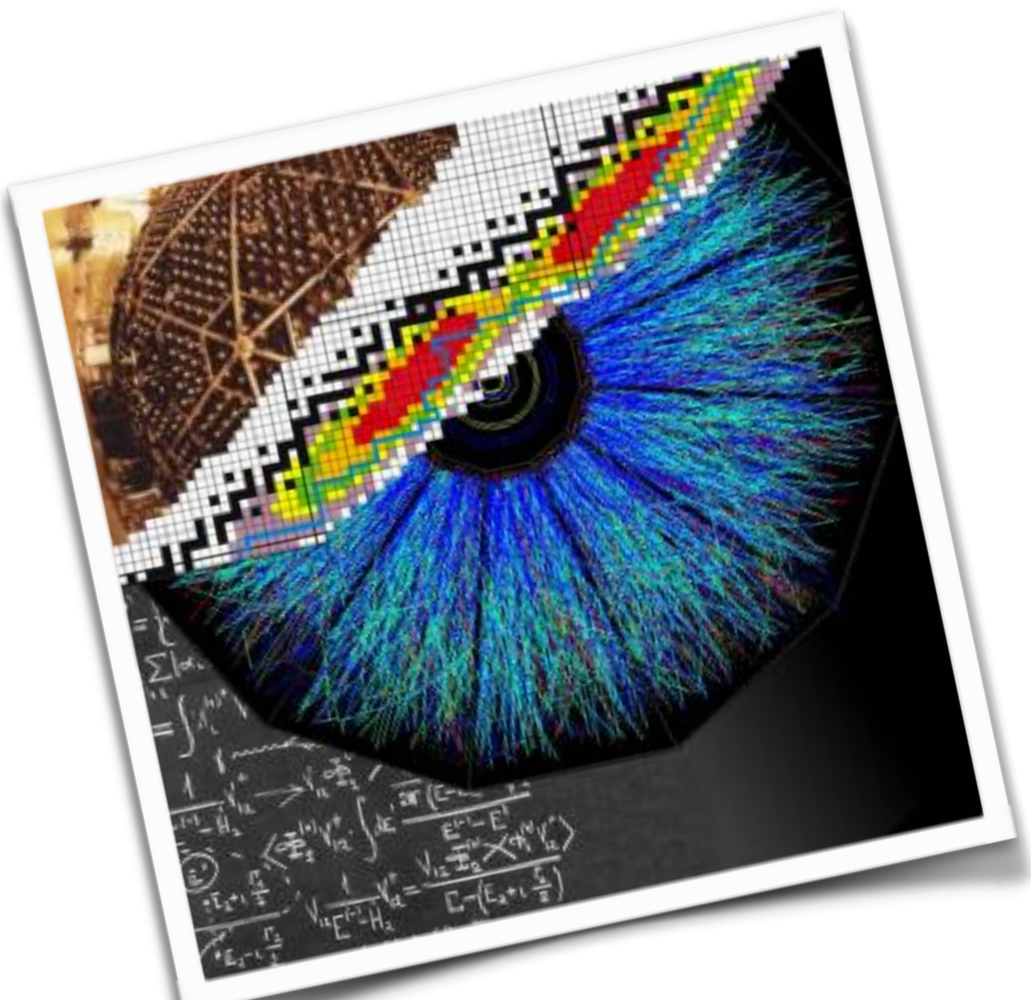
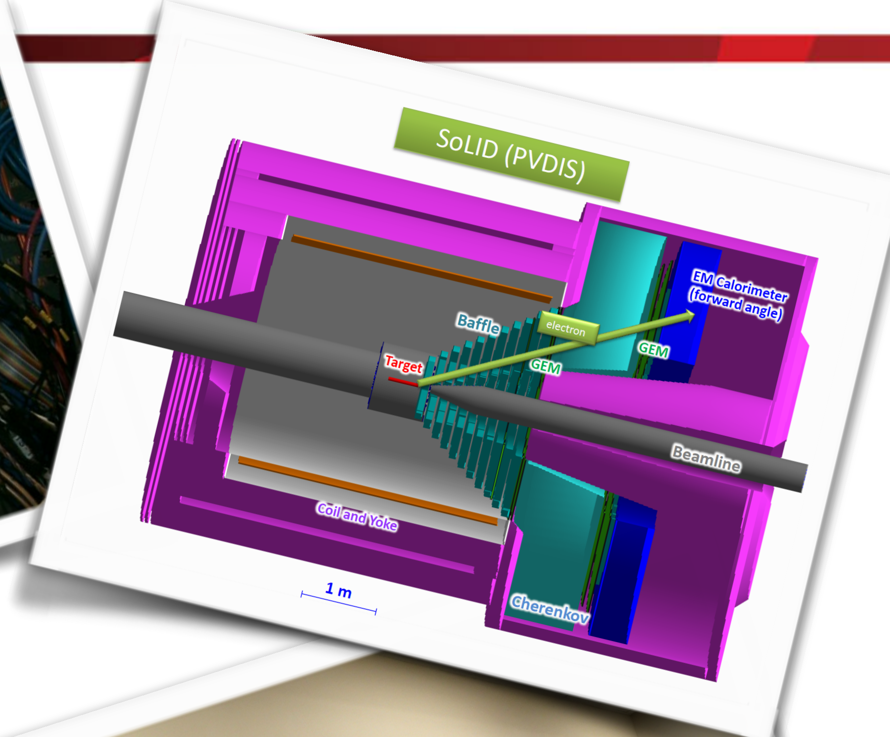
ASCR-NP : Experimental NP

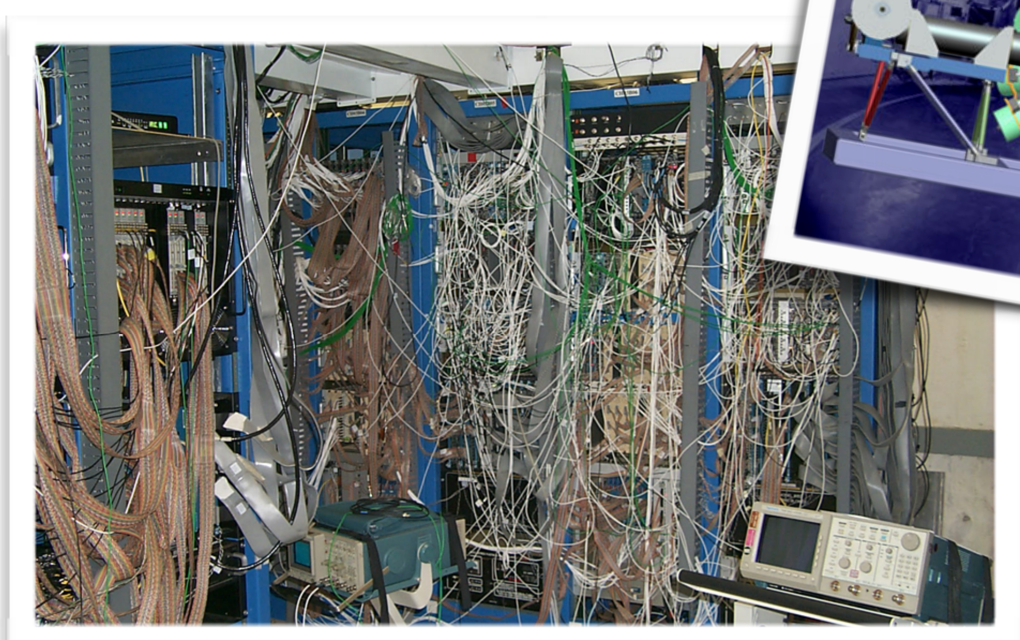
Graham Heyes - JLab, July 5th 2016

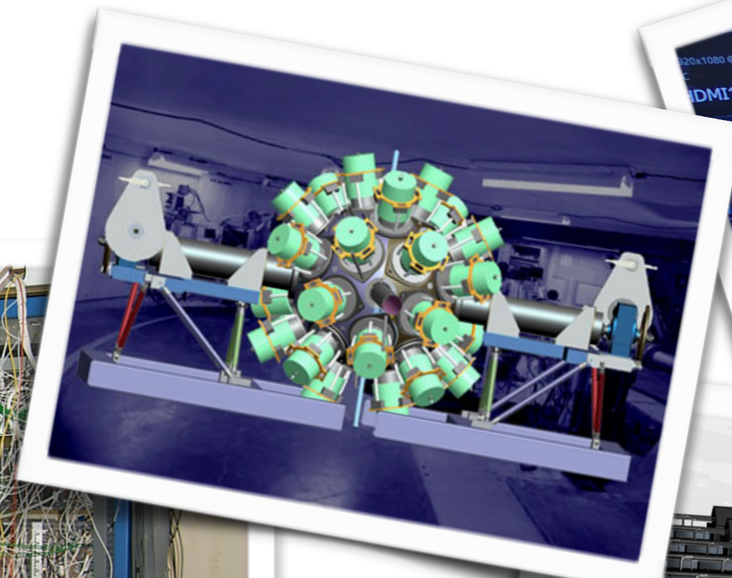


Introduction

- DAQ.
- Streaming
- Trends in technology.
- Other labs.
- Simulation and analysis.
- Opportunities for collaboration.
- Concluding remarks.

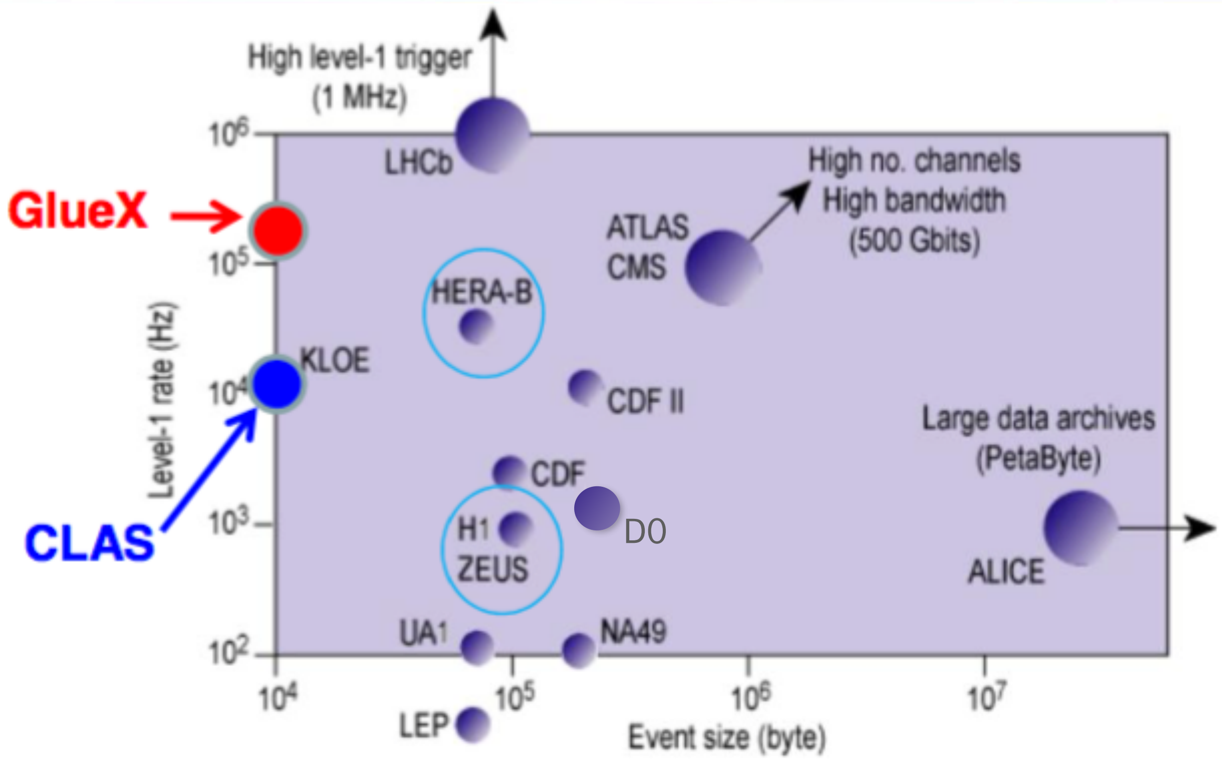






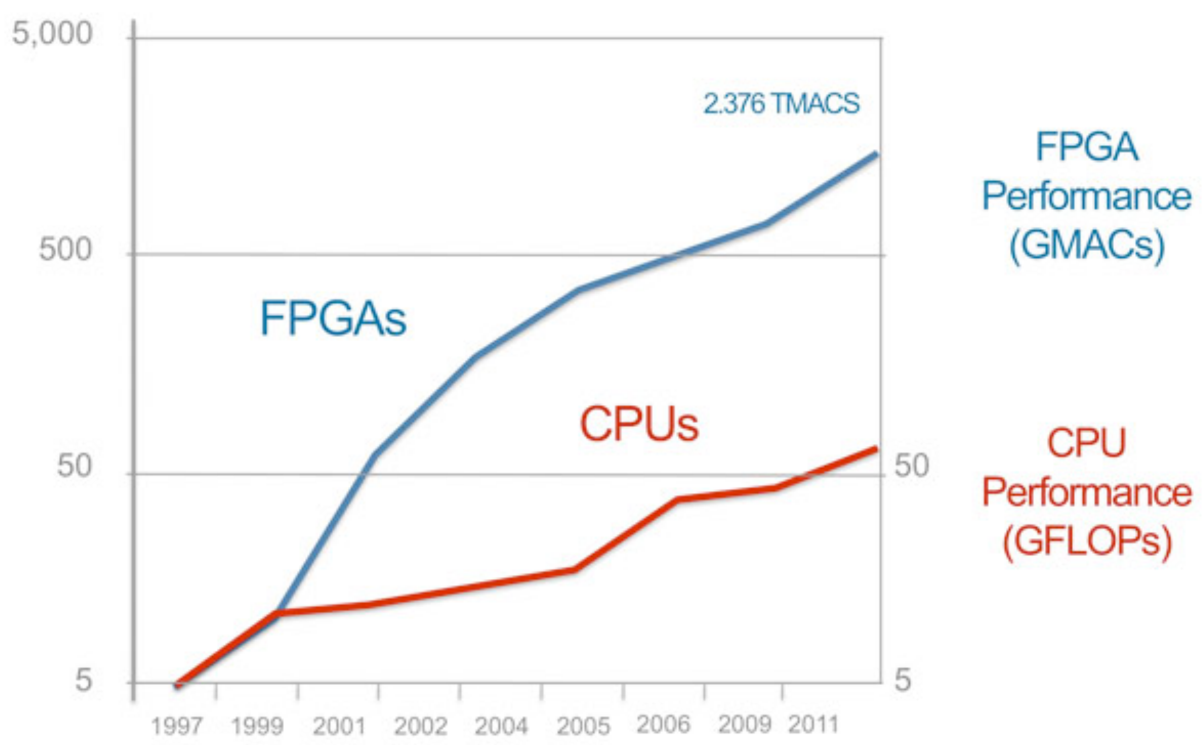
Trends in experiments

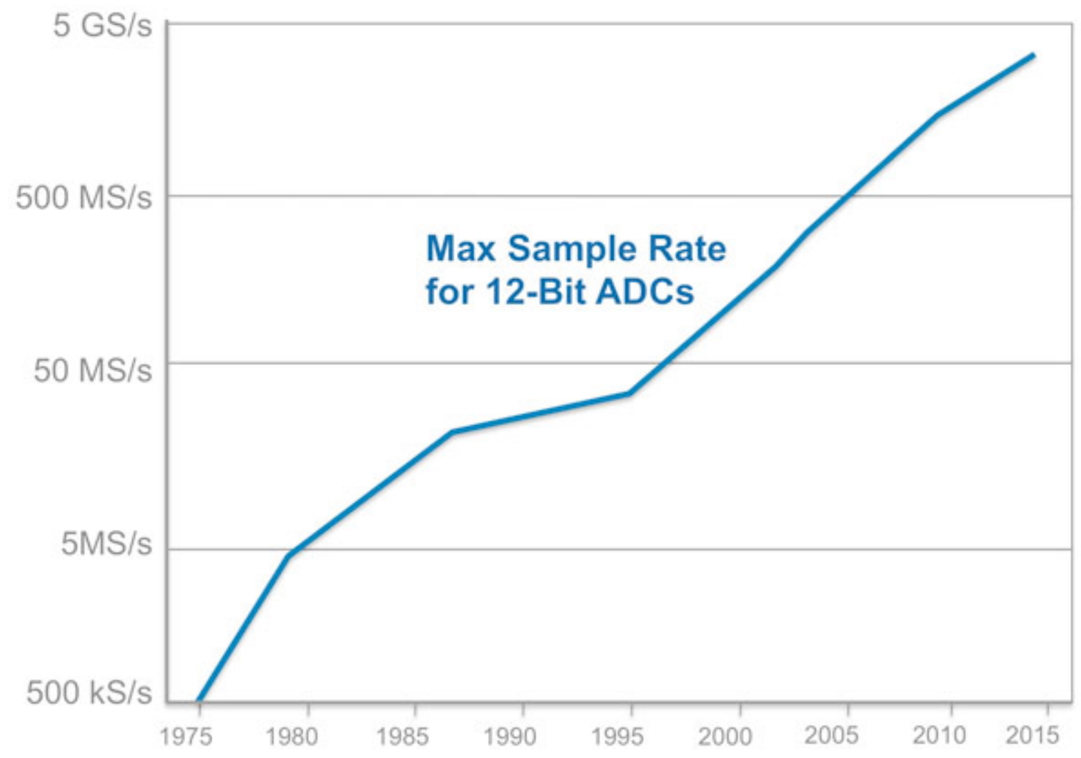
- Look at historical trigger and data rates.
- At JLab
  - mid 1990's CLAS, 2 kHz and 10-15 MB/s
  - mid 2000's - 20 kHz and 50 MB/s
  - mid 2010's
    - HPS, 50 kHz and 100 MB/s
    - GLUEX
      - 100 kHz, 300 MB/s to disk.
      - (Last run 35 kHz 700 MB/s)
- FRIB - odd assortment of experiments with varying rates
  - LZ Dark matter search 1400 MB/s
  - GRETA 4000 channel gamma detector with 120 MB/s per channel. (2025 timescale)
- RHIC PHENIX 5kHz 600 MB/s
- RHIC STAR - Max rate 2.1 GB/s average 1.6 GB/s
- Looking at the historical trends the highest trigger rate experiments increase rate by a factor of 10 every 10 years.



Trends in trigger and electronics

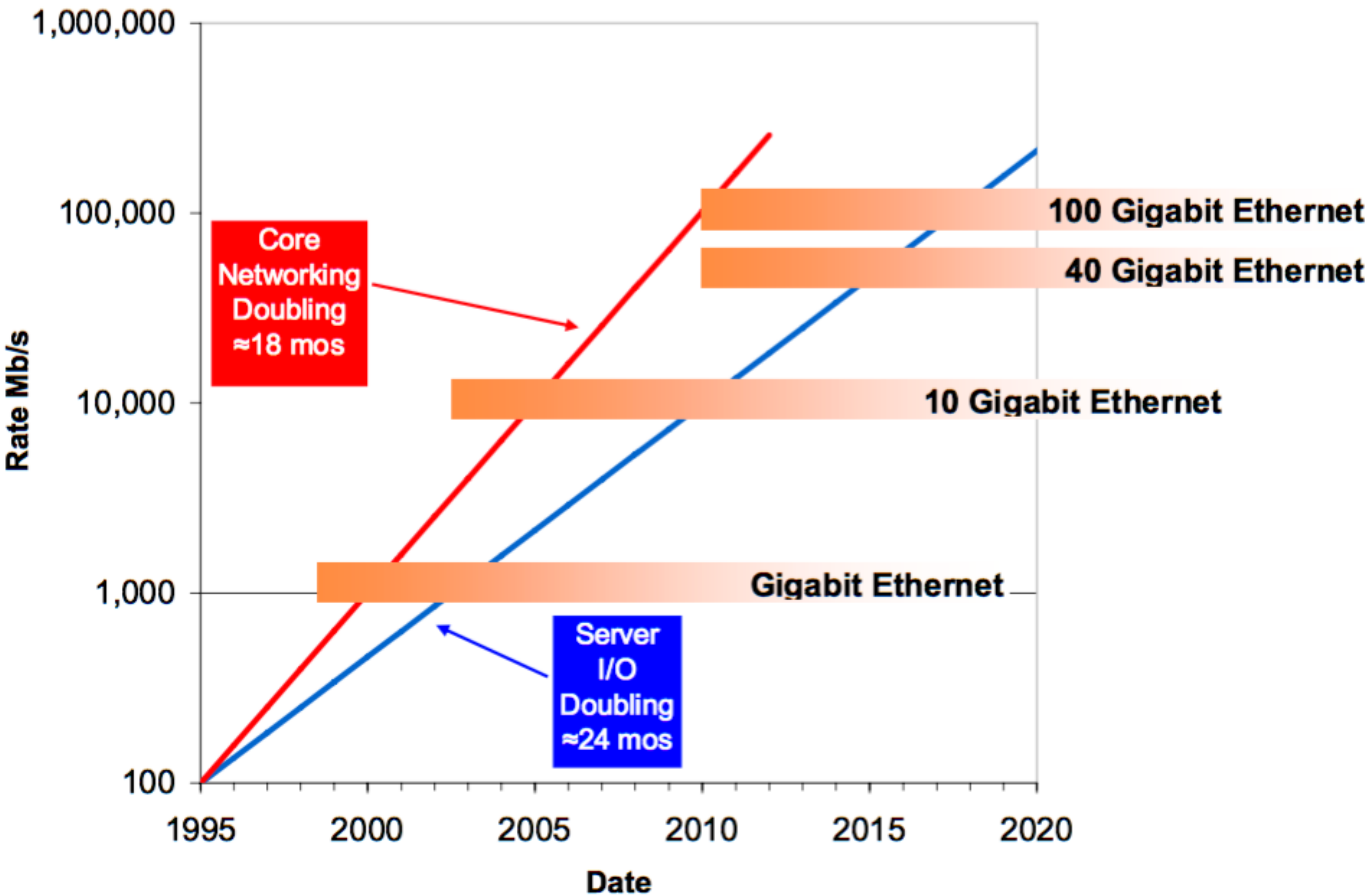
- FPGA performance is increasing faster than CPU performance. There is a delay between when technology is developed and when it becomes affordable for use in custom electronics. So there is room for growth over the next ten years.





- Current trend is to push some functionality currently performed in software running on embedded processors into firmware on custom electronics. This will probably continue.

Trends in data transport



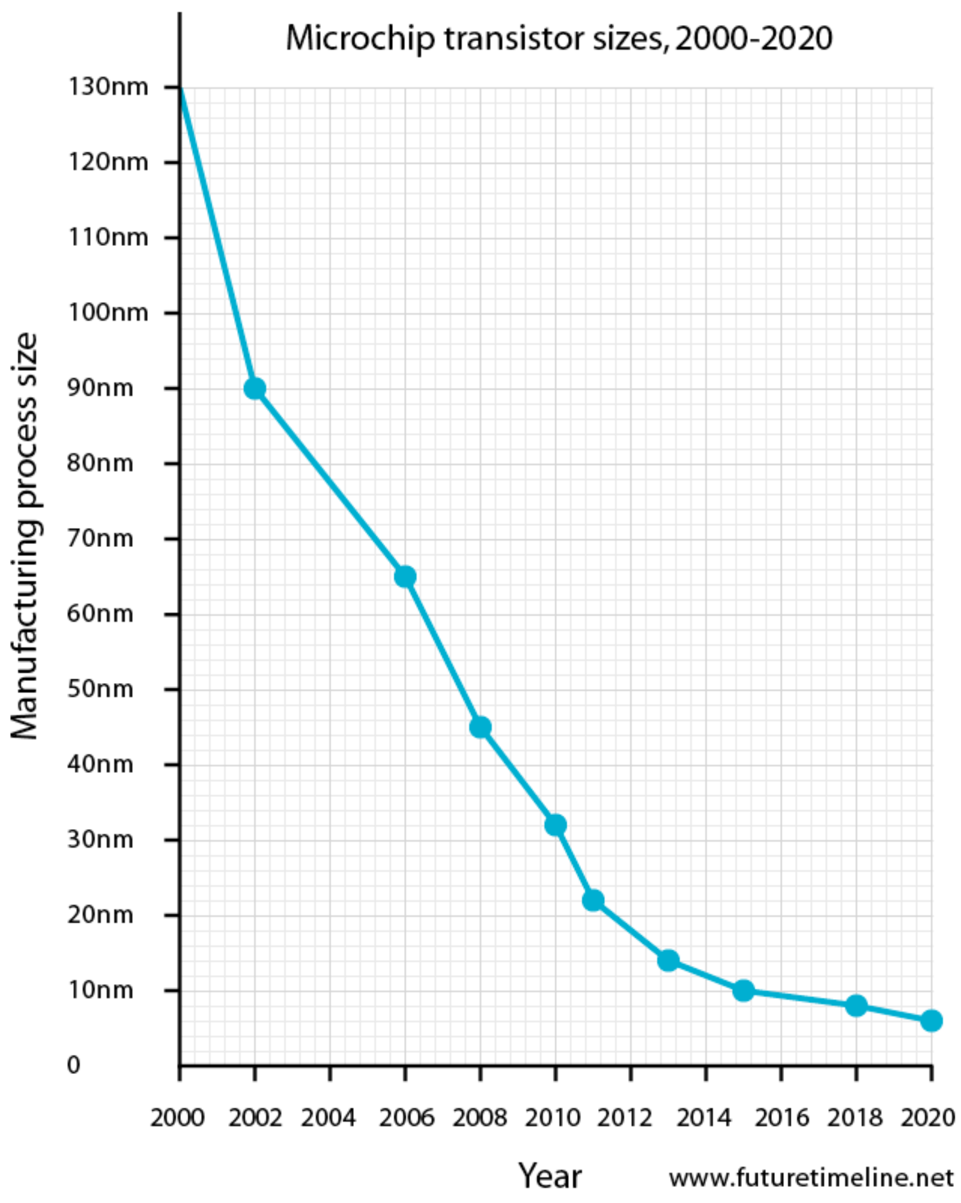
IEEE 802.3 Higher Speed Study Group - TUTORIAL

Challenges

- The precision of the science depends on statistics which leads to :
  - Development of detectors that can handle high rates.
  - Improvements in trigger electronics - faster so can trigger at high rates.
- Beam time is expensive so data mining or taking generic datasets shared between experiments is becoming popular.
  - Loosen triggers to store as much as possible.
- Some experiments are limited by event-pileup, overlapping signals from different events, hard to untangle in firmware.
- Often the limiting factor in DAQ design is available technology vs budget, a constraint shared by all experiments at the various facilities.
  - It is not surprising that trigger and data rates follow an exponential trend given the "Moore's law" type exponential trends that technologies have been following.
  - **What matters is not when a technology appears but when it becomes affordable**. It takes time for a technology to become affordable enough for someone to use it in DAQ.

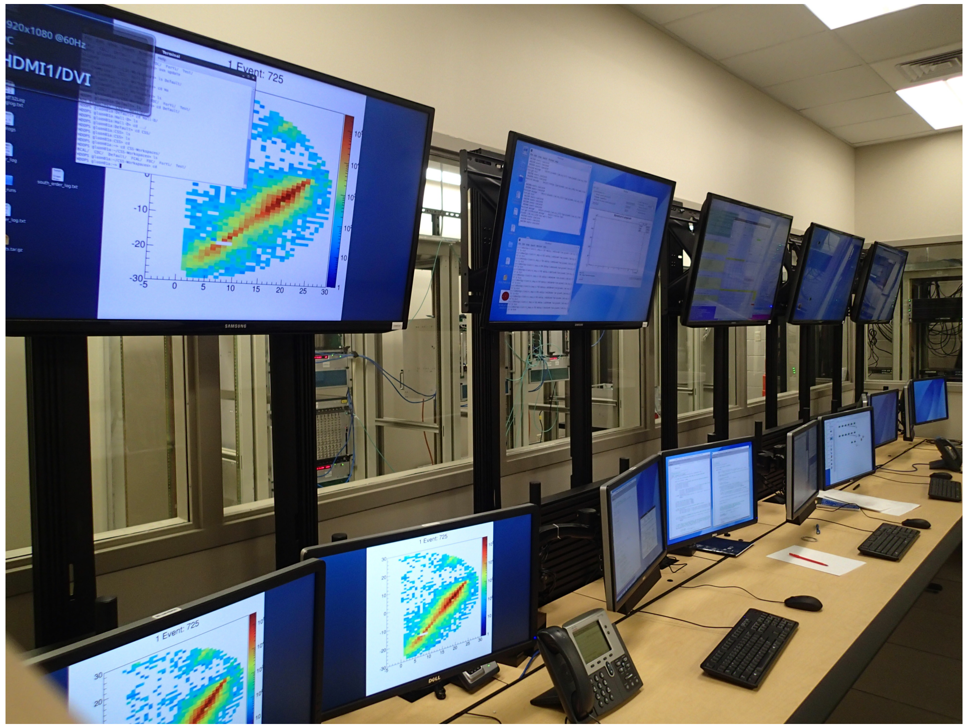
Challenges

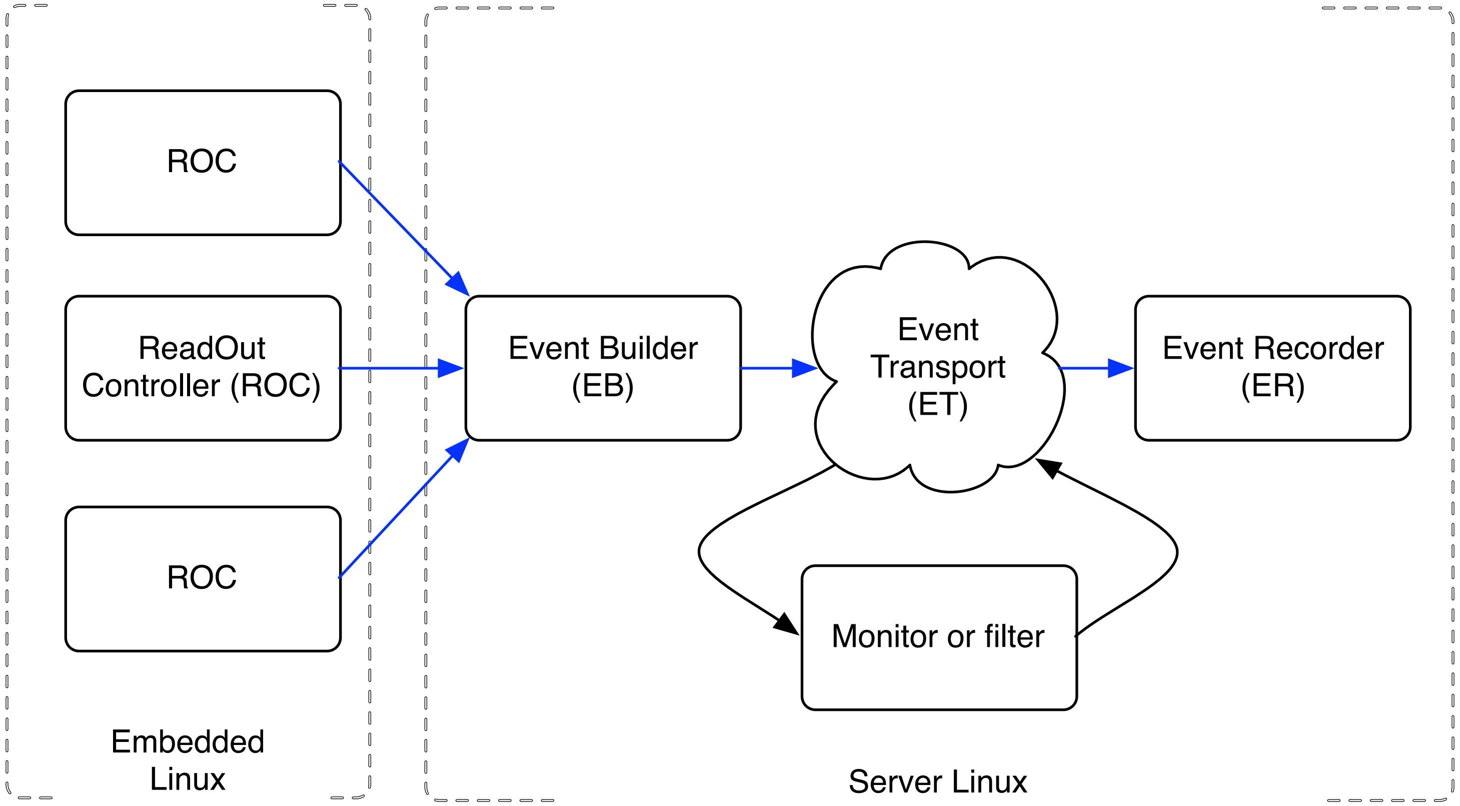
- Manufacturers are struggling shrink transistors.
  - How much further can Moore's law continue?
  - When does this trickle down affect the performance of other DAQ electronics?
- Use of mobile devices is driving tech in a direction that may not be helpful to NP DAQ, low power and compact rather than high performance.
- Are the rates for proposed experiments low because of low expectation?
  - Does the requirement of the experiment expand to take full advantage of the available technology?
  - If we come back in five years from now and look at experiments proposed for five years after that will we see a different picture than the one that we now see looking forward ten years? Probably yes.



System architecture

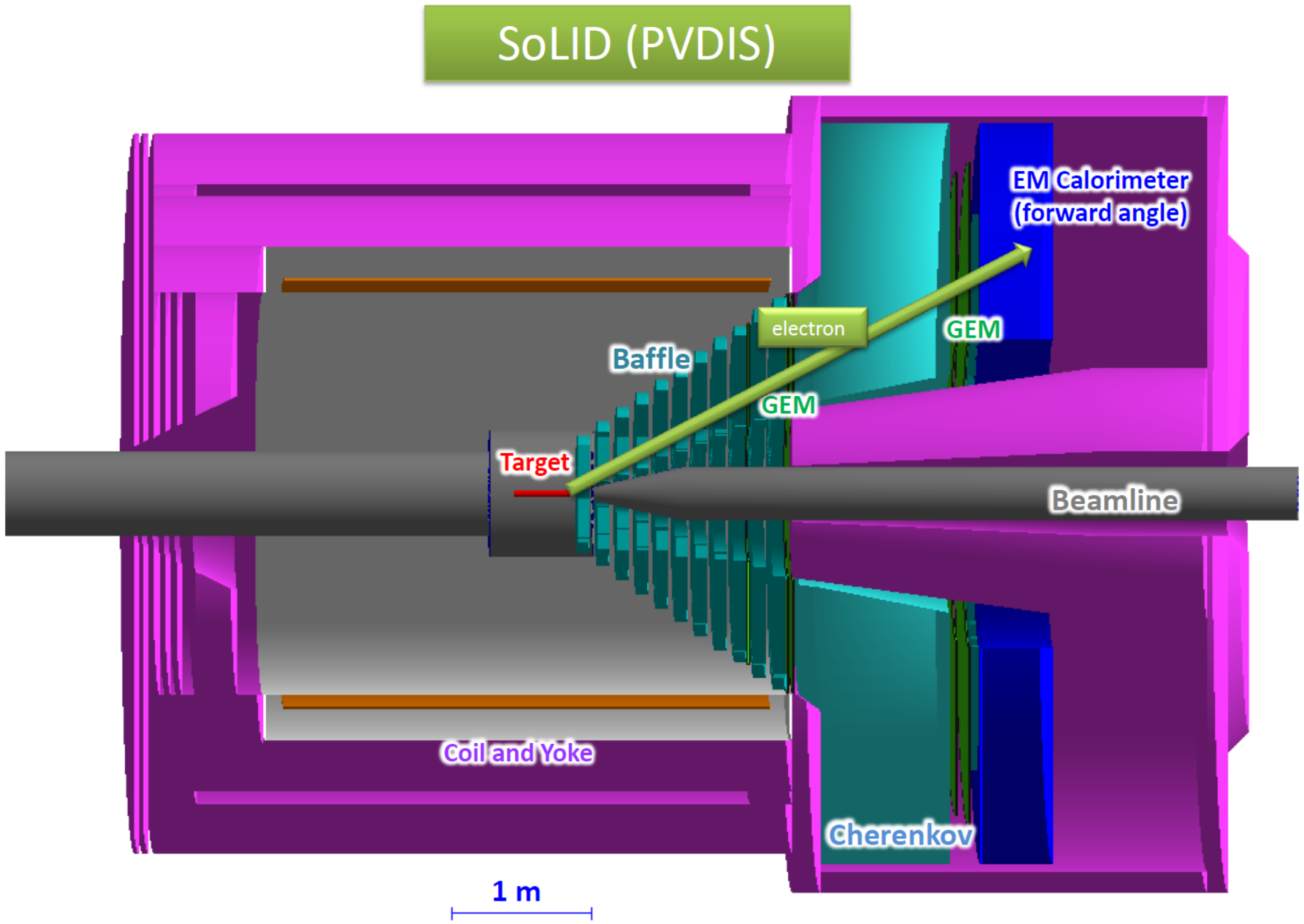
- DAQ architectures have not changed much in twenty years.
  - Signals are digitized by electronics in front end crates.
  - Trigger electronics generates trigger to initiate readout.
  - Data is transported to an event builder.
  - Built events are distributed for filtering, monitoring, display etc.
  - Event stream is stored to disk.
- Issues :
  - Single electronic trigger
  - Bottlenecks
  - Scalability
  - Stability





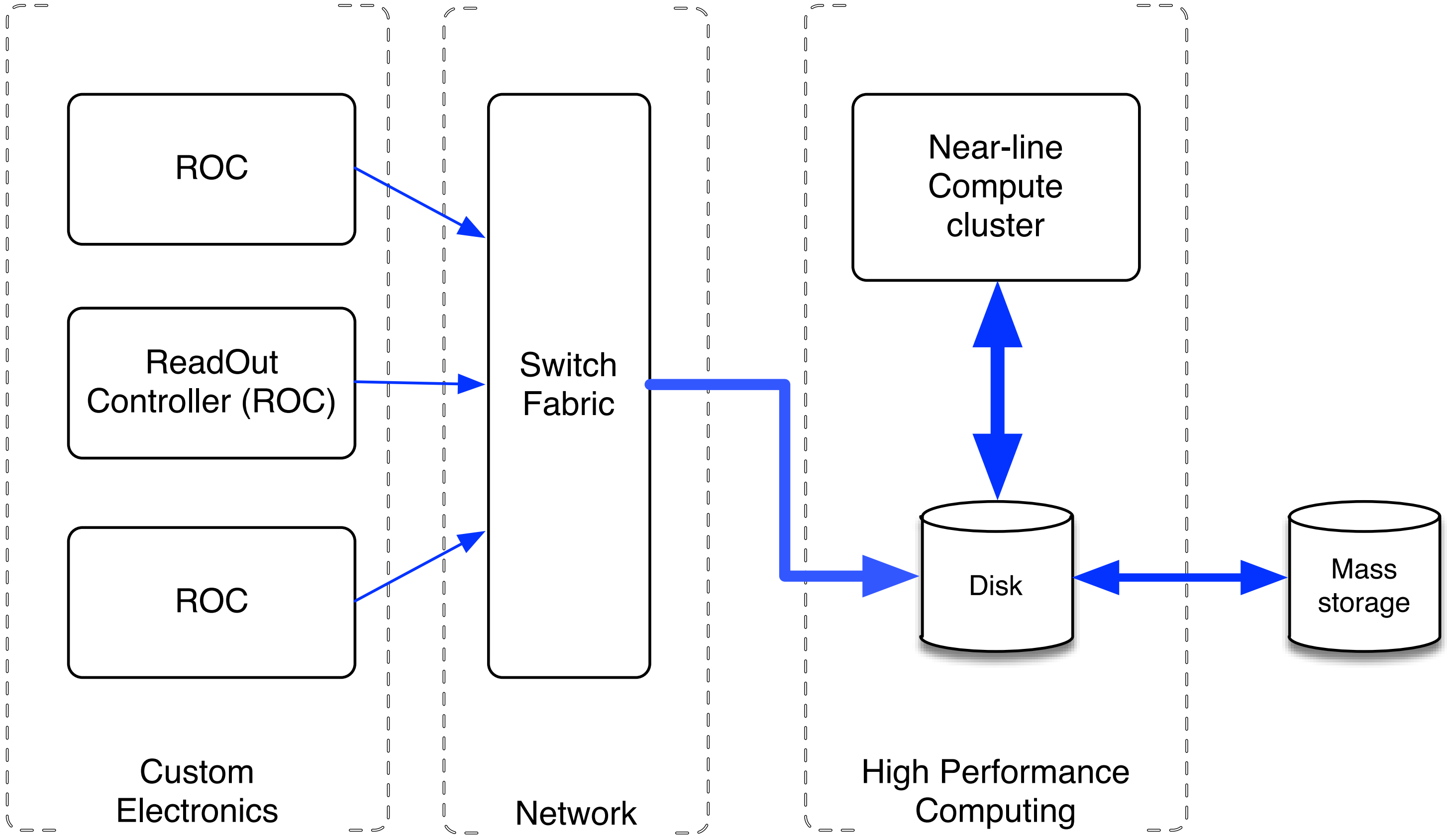
Future experiments, JLab - SoLID

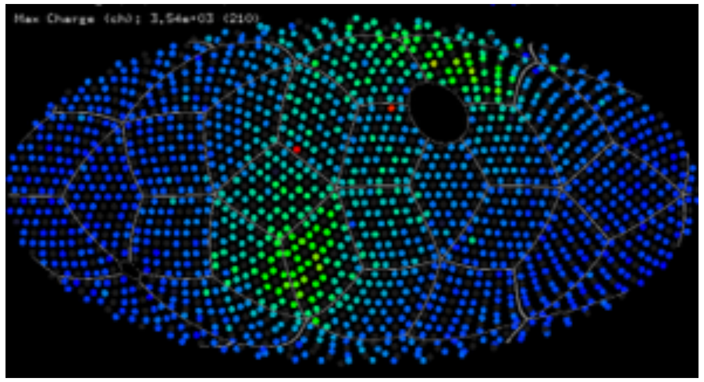
- SoLID is an experiment proposed for installation hall-A at JLab.
- The detector has two configurations. In the PVDIS configuration electrons are scattered of a fixed target at high luminosity.
- The detector is split into 30 sectors, the single track event topology allows 30 DAQ systems to be run in parallel at rates of 1 GByte/s each.

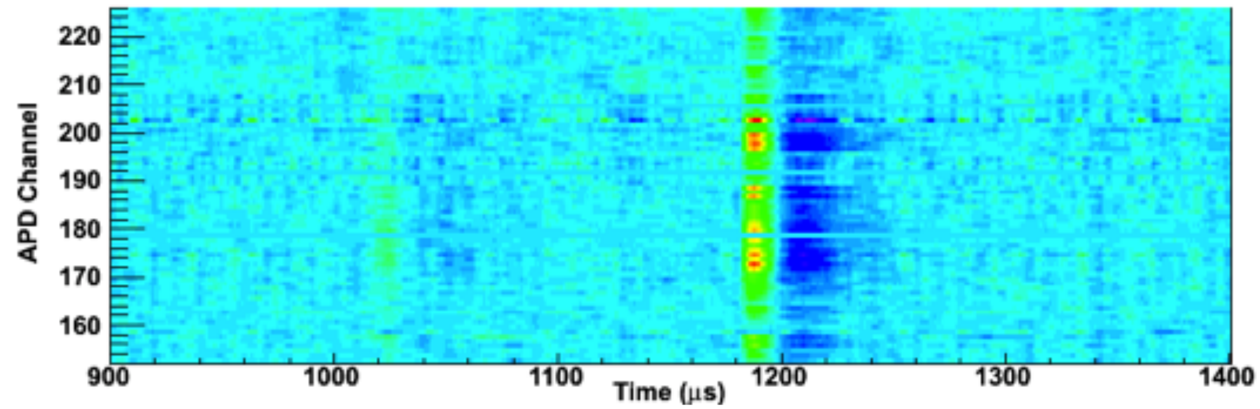


Alternative future solution

- Can’t escape some sort of crate to put the electronics in - MicroTCA ?
- Pipe the data through a network directly to temporary storage.
- High performance compute system processes the data online implementing a software trigger.
  – Several different triggers in parallel?
- Data surviving trigger or output from online processing migrates to long term storage freeing space for raw data.
- Much simpler architecture - more stable DAQ - but needs affordable versions of :
  – Reliable high performance network accessible storage.
  – High bandwidth network.
  – Terra scale computing.

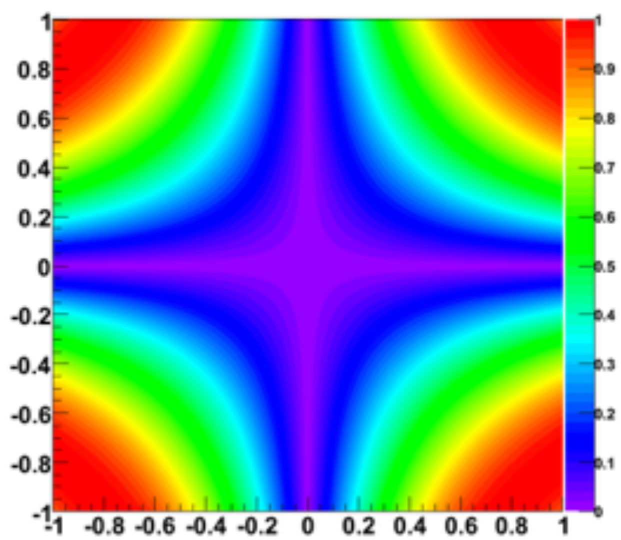


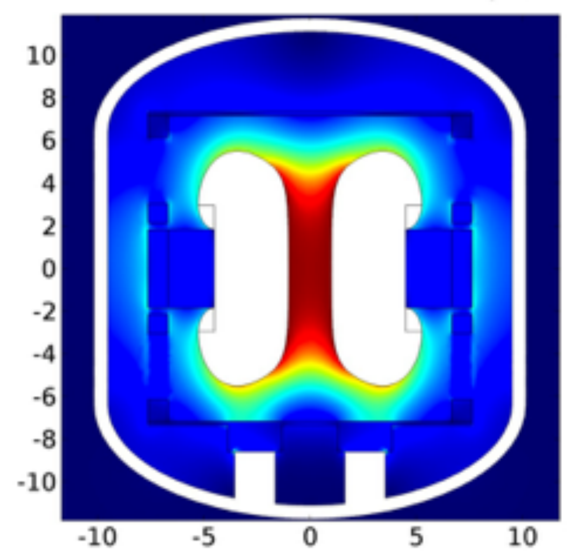


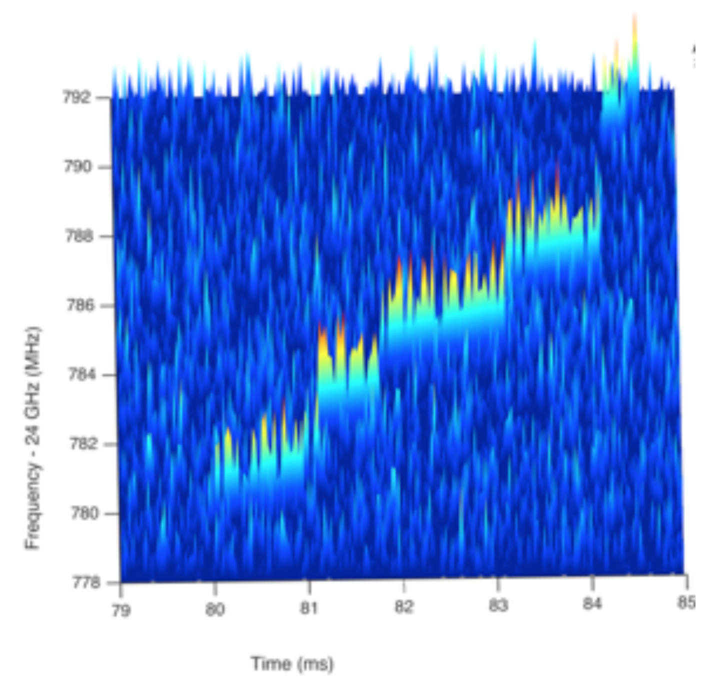


Experiments in Fundamental Symmetries and Neutrinos

Jason Detwiler, University of Washington
Exascale Requirements Review for Nuclear Physics
June 15, 2016



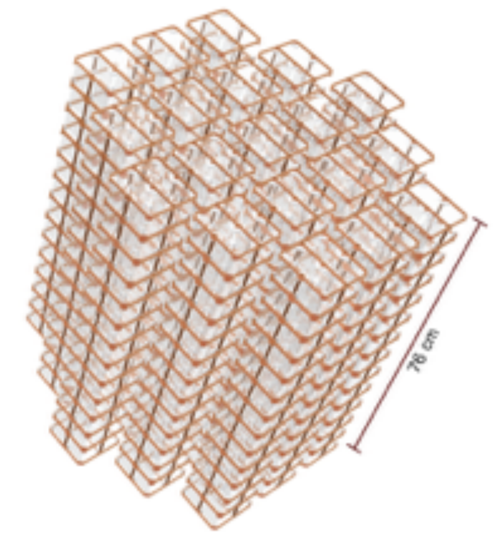




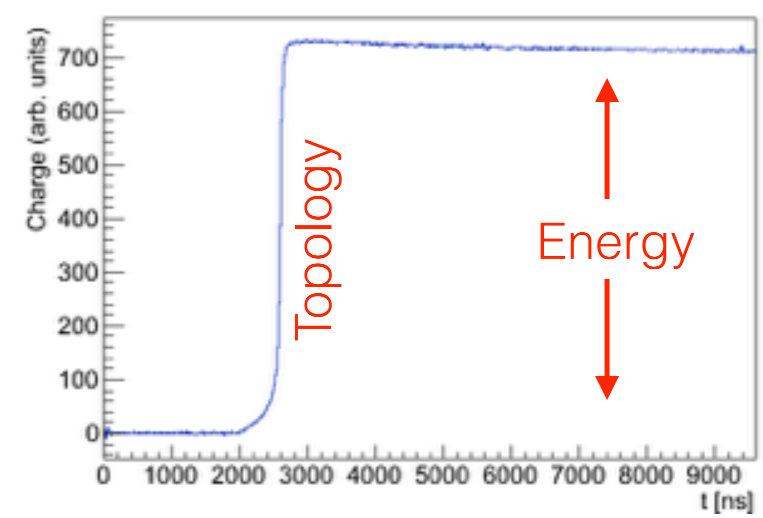
Neutrinoless Double-Beta Decay

- Current scale: 10's-100's of kg. 2015 NP LRP Rec II: ton(s) scale within the next decade
- Major technologies:
  - **Large crystal arrays** (CUORE, MAJORANA/GERDA): ionization / bolometer signals filtered for energy and pulse shape parameters. ~100 TB and hundreds of kCPU-hrs per year → ~3 PB/y, 3-10 MCPU-hrs/y (scales with volume).
  - **TPCs** (EXO, NEXT (SuperNEMO)): ionization and scintillation signals analyzed for energy and position reconstruction (some parallelization in-use) and other event topology info. 300 TB and ~1 MCPU-hr per year → 3 PB/y, 3-10 MCPU-hrs/y (scales with surface area).
  - **Large liquid scintillators** (SNO+, KamLAND-Zen): PMT signals analyzed for charge and time, used to reconstruct energy, position, and other parameters. ~100 TB and ~1 MCPU-hr per year (won't grow much)
- Many CPU-hours for simulations / detector modeling as well as signal processing

MAJORANA



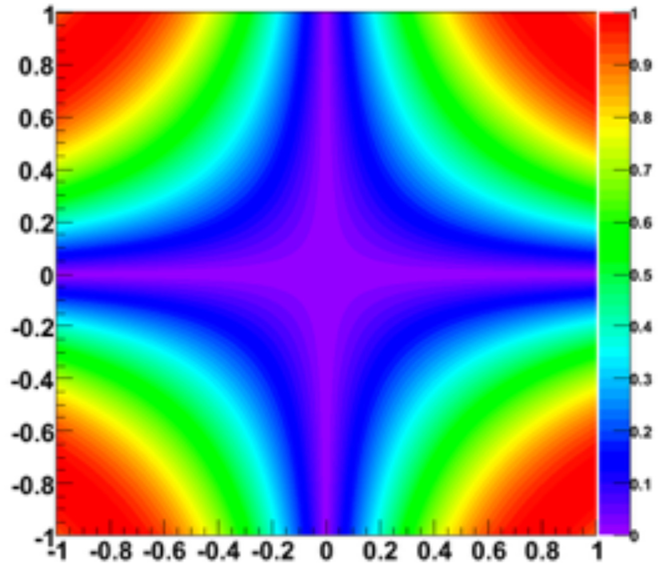
CUORE



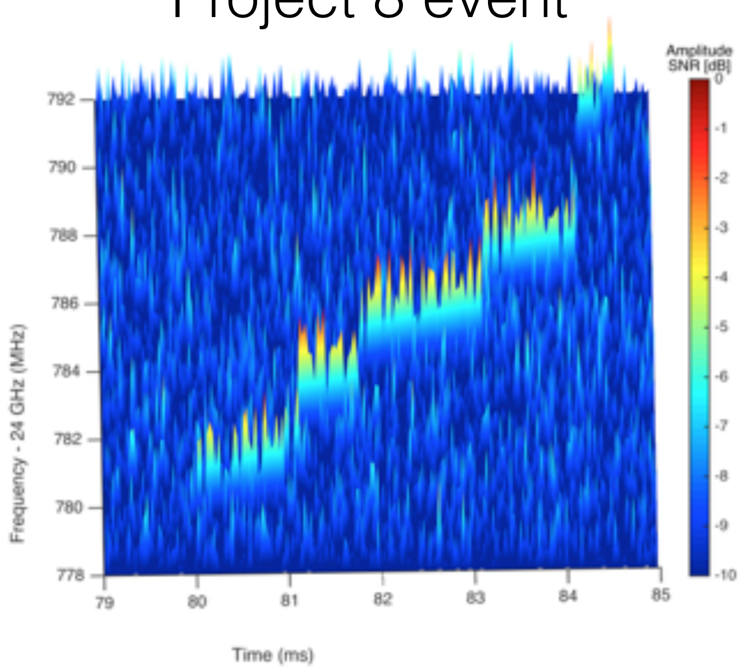
Kinematic Neutrino Mass Measurements

- **KATRIN**: MAC-E spectrometer (“dial and count”)
  - Data size is relatively small. Computing challenge: electron transport modeling.
  - 3D E&M, gas dynamics, MCMC techniques
  - Already using GPU techniques and parallel processing (field solver).
  - Modest resources required: TB of data, thousands of CPU-hr.
- **Project 8**: Cyclotron Radiation Emission Spectroscopy
  - RF time series recorded at 100 MB/s per receiver (~3 PB/yr)
  - Locate tracks and measure energy, pitch, other topology info (FFT, DBSCAN, Consensus Thresholding, KD-Trees, Hough Transforms…)
  - Current: 1 receiver, short runs: TB of data, hundreds of kCPU-hrs processing, little parallelism.
  - Future: 60 receivers, longer runs → ~200 PB/yr, millions of CPU-hrs. Data reduction and GPU methods under investigation.

KATRIN field

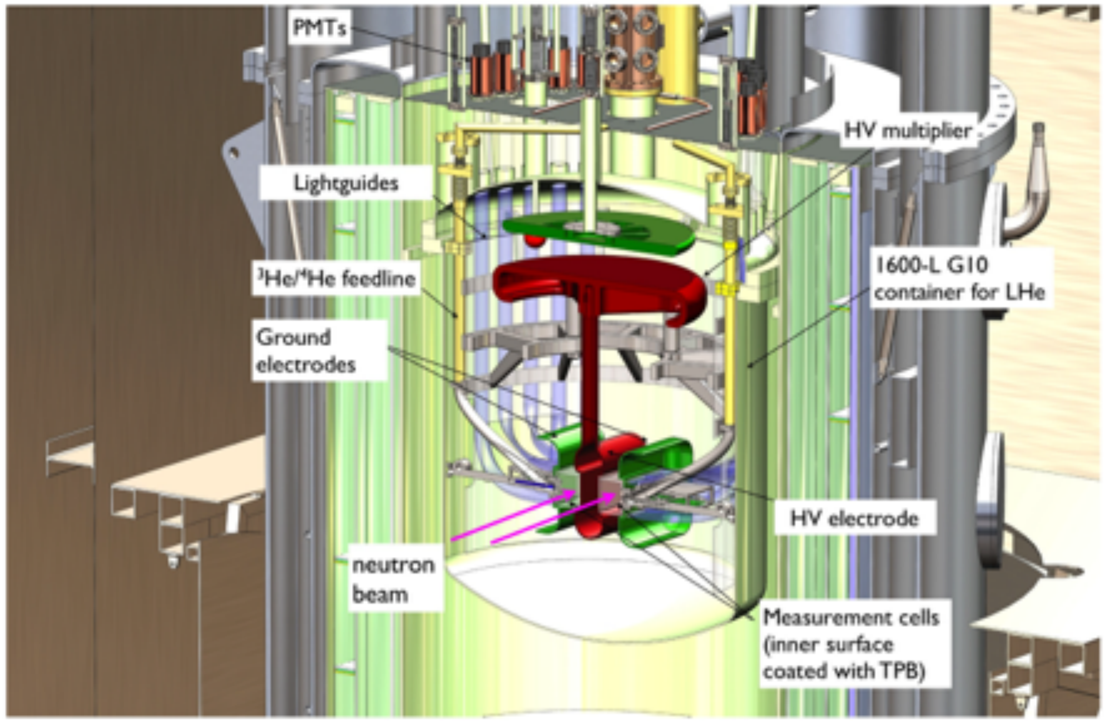


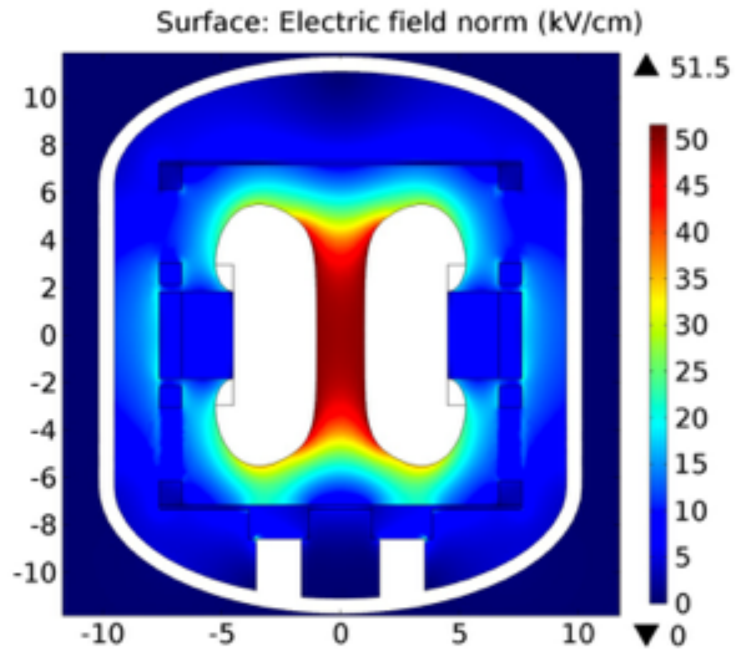
Project 8 event



Neutron EDM

- Hosted at SNS but most computing done on local clusters at collaborating institutions
- Data stream: SQUIDs and scintillators
- Detector response / background modeling: COMSOL (parallelized) for field solving, COMSOL or Geant4 for spin transport, Geant4 for background simulations
  - Many systematic studies required for ultimate sensitivity.
  - Currently limited by available memory
- Computation needs:
  - CPU: 0.1 → 100 MCPU-hr
  - Memory: 5 → 64 GB/node
  - Disk: 10 TB → 1 PB



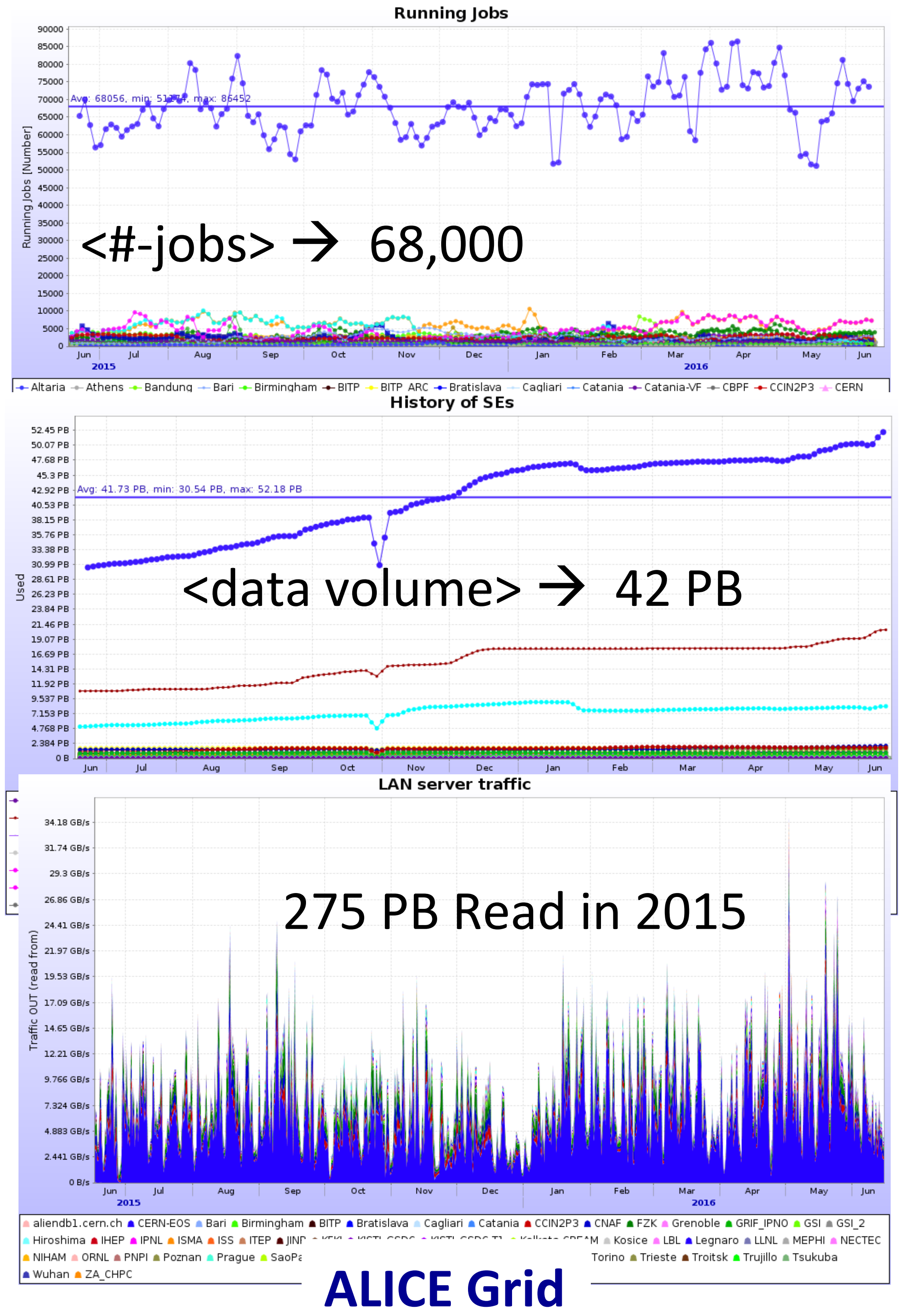


Large Data Sets – Needs at LHC

Jeff Porter (LBNL)
with ongoing input from Charles Maguire (Vanderbilt U.)

Scale of LHC Operations

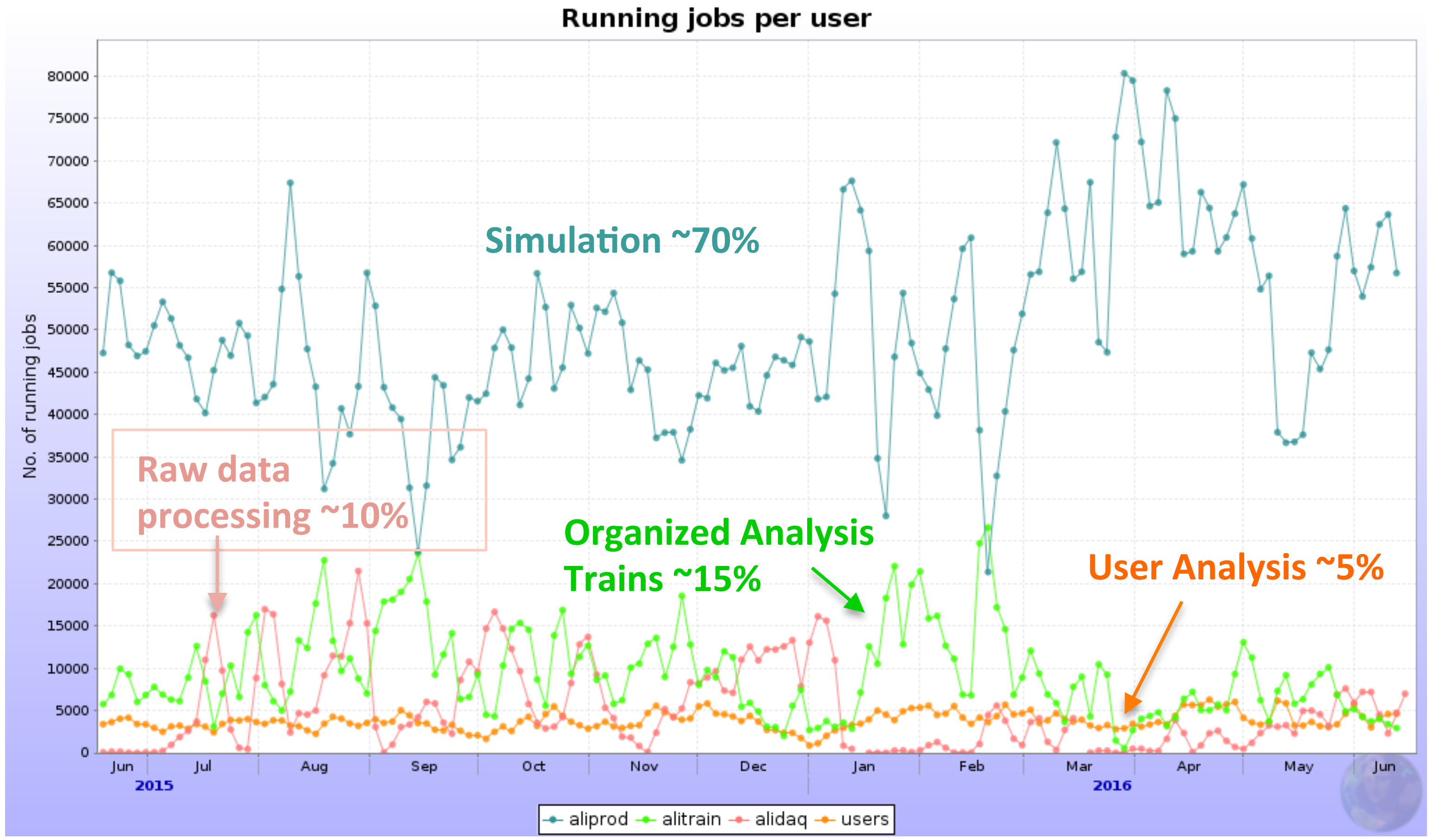
- **ALICE Distributed Processing**
  - 7 PB/year new raw data
  - 60,000+ concurrent jobs
  - 50 PB distributed data store
  - Process ~300 PB/year
  - ➢ ALICE-USA < 10%



- **CMS Heavy Ion Program**
  - US dominated,
    - primarily on NP Tier 2 & CERN Tier 0
  - 3000 concurrent jobs
  - 3+ PB Grid enabled storage
  - p+p data processing not included
    - common with HEP program

ALICE Offline Computing Tasks

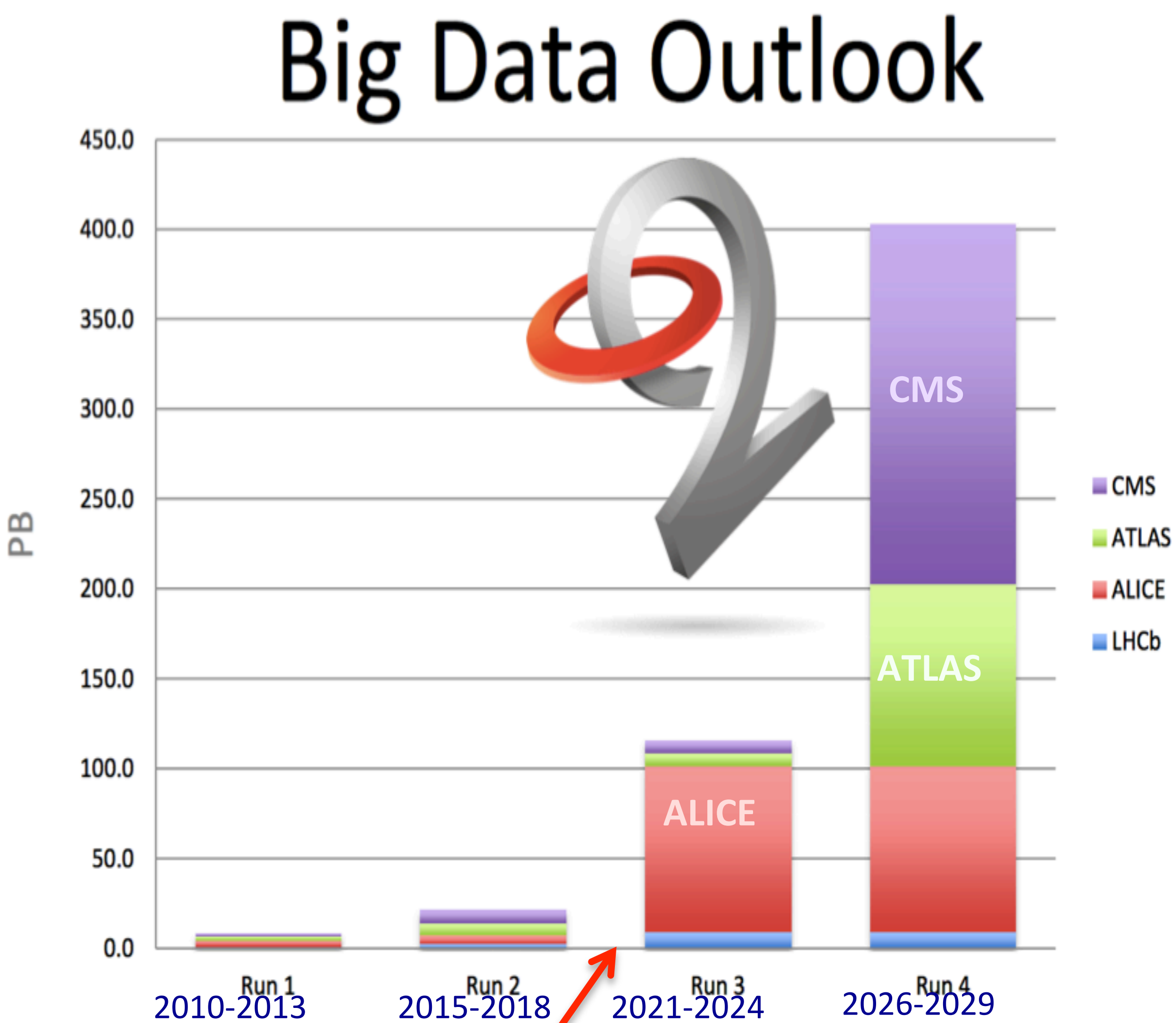
- **Raw Data Processing**
  - Calibration
  - Event Reconstruction
- **Simulation**
  - Event Generation
  - Detector Simulation
  - Digitization
  - Event Reconstruction
- **User Analysis**
  - AOD processing
  - Typically input-data intensive → low CPU efficiency
- **Organized Analysis → Analysis Trains**
  - AOD processing
  - Less I/O intensive → read once for many analyses
  - Adopted ~2+ years ago, now dominant AOD processing mode



ALICE Jobs Breakdown

LHC Running Schedule

- **Collider Running Schedule**
  - Run for 3+ years
  - Shutdown for 2 years
- **Run 1: 2010-2013 (early)**
  - ALICE ~7 PB Raw data
  - CMS HI
- **Run 2: 2015-2018**
  - estimate ~2-3x Run 1 both ALICE & CMS
- **Run 3: 2021-2024**
  - ALICE estimate is 100x Run 1
  - CMS (TBD)
- **Run 4: 2026-2029**
  - Official LHC High Luminosity Era

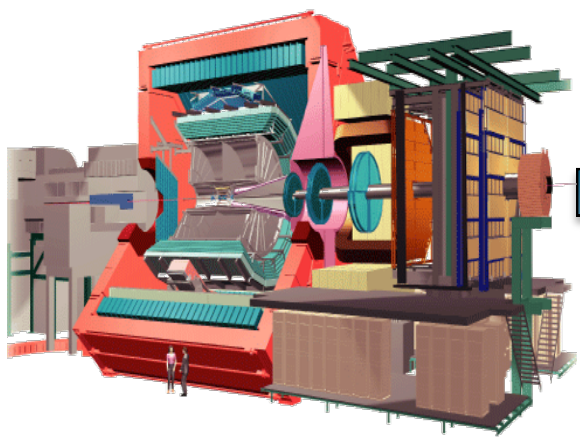


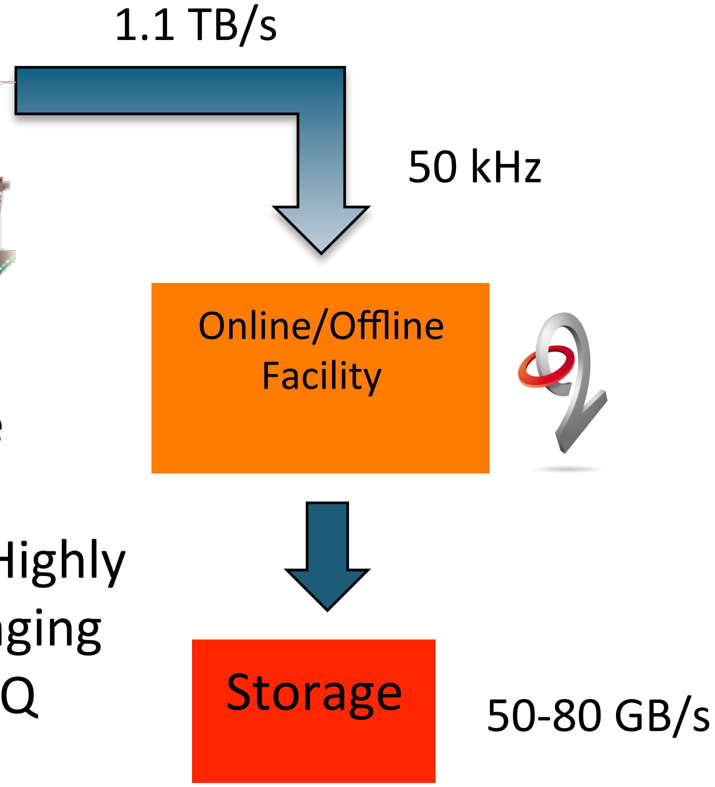
<u>Physics Driven Increase for Run 3</u>:
large statistics heavy-flavor & charmonium in minimum-bias data sample
(CMS HI may have similar goals)

Jeff Porter LBNL

ALICE O2 (Online-Offline) Project:
Offline quality reconstruction in Online for data reduction

ALICE O2 Project





- Data reduction by online reconstruction
  - ALFA Framework: Highly parallel with messaging service, e.g. ZeroMQ
  - <u>No event rejection</u>
  - Final volume ~10x increase

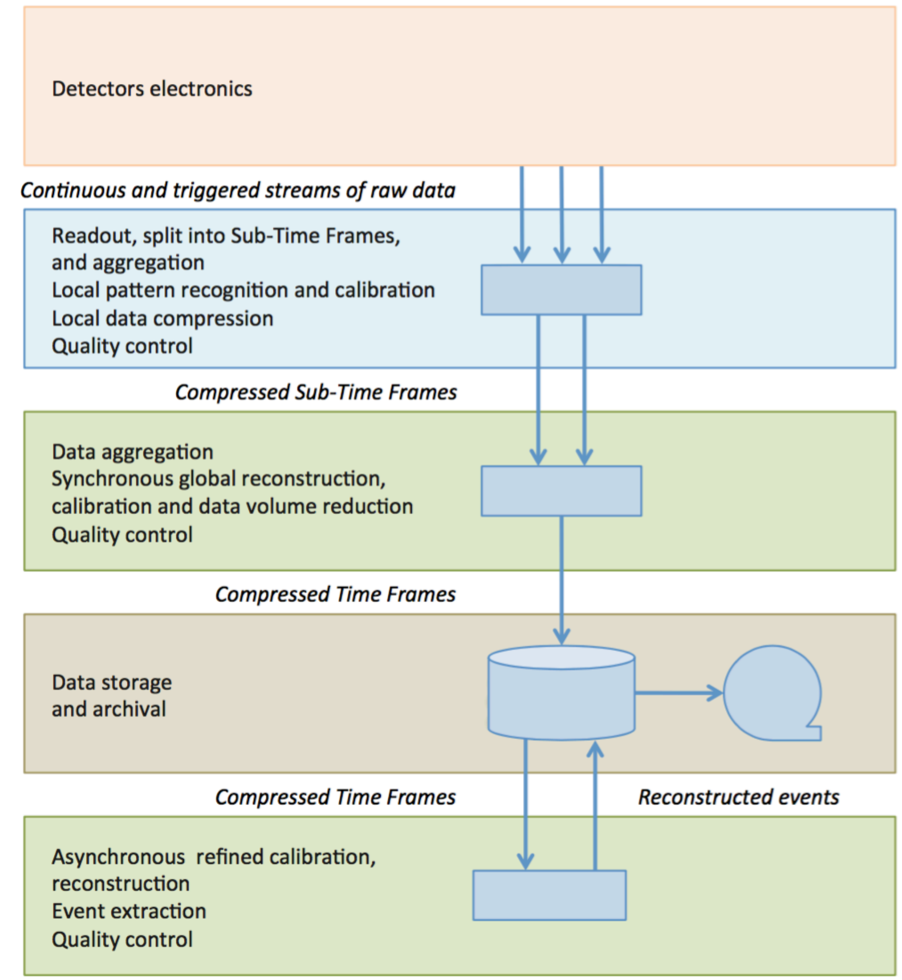
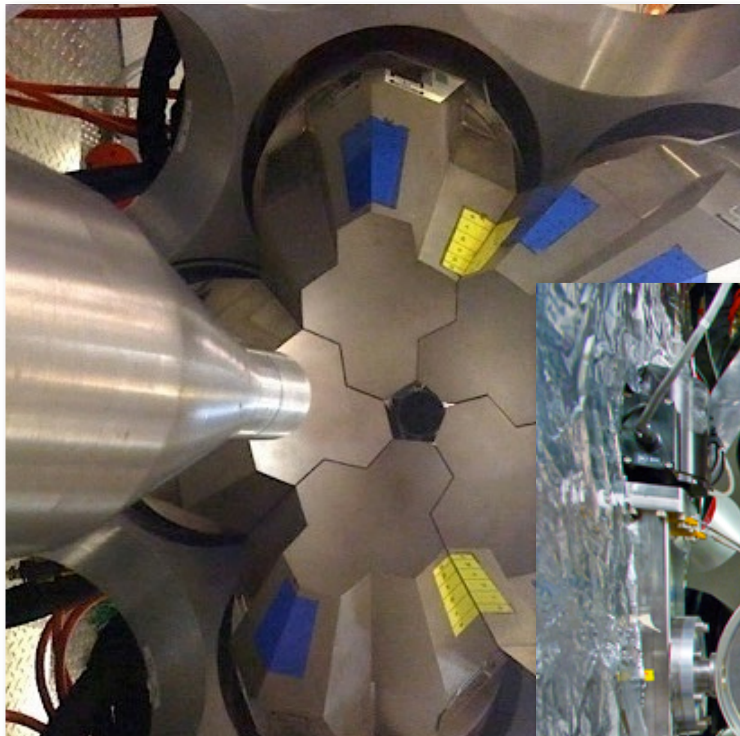
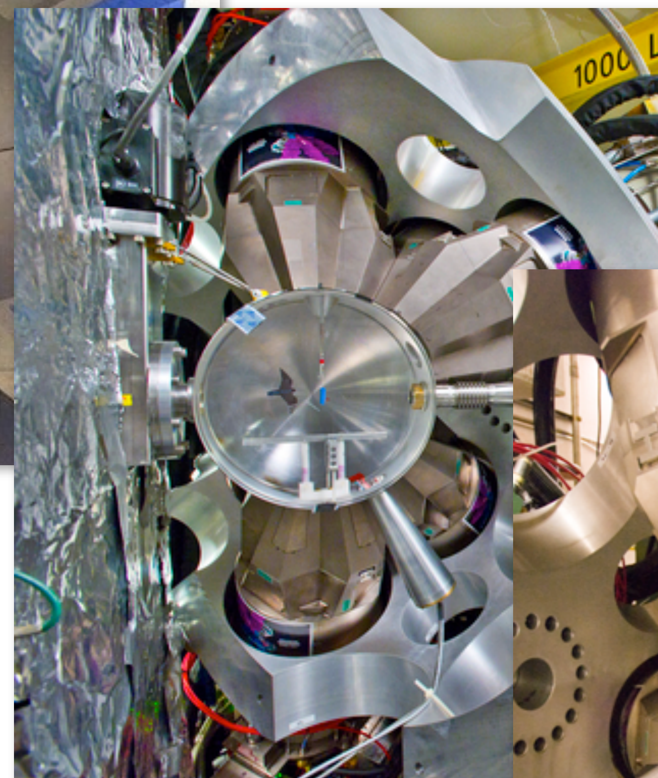


Figure 1.1: Functional flow of the O^2 computing system.

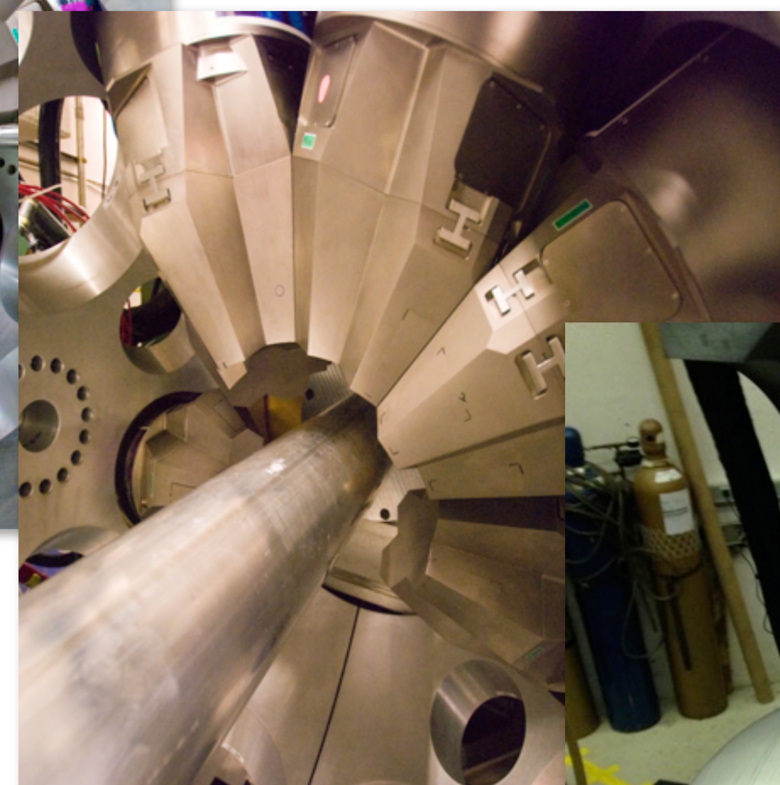
- **Project will produce a new flexible O2 framework designed for production purposes**
  - Will include capability for reconstruction of simulated data
  - Will not target event and detector simulations or ROOT-based user analysis
  - NOTE: Data volume is reduced by O2, not number of events → 100x increase

Jeff Porter LBNL

Streaming

Mario Cromaz, LBNL

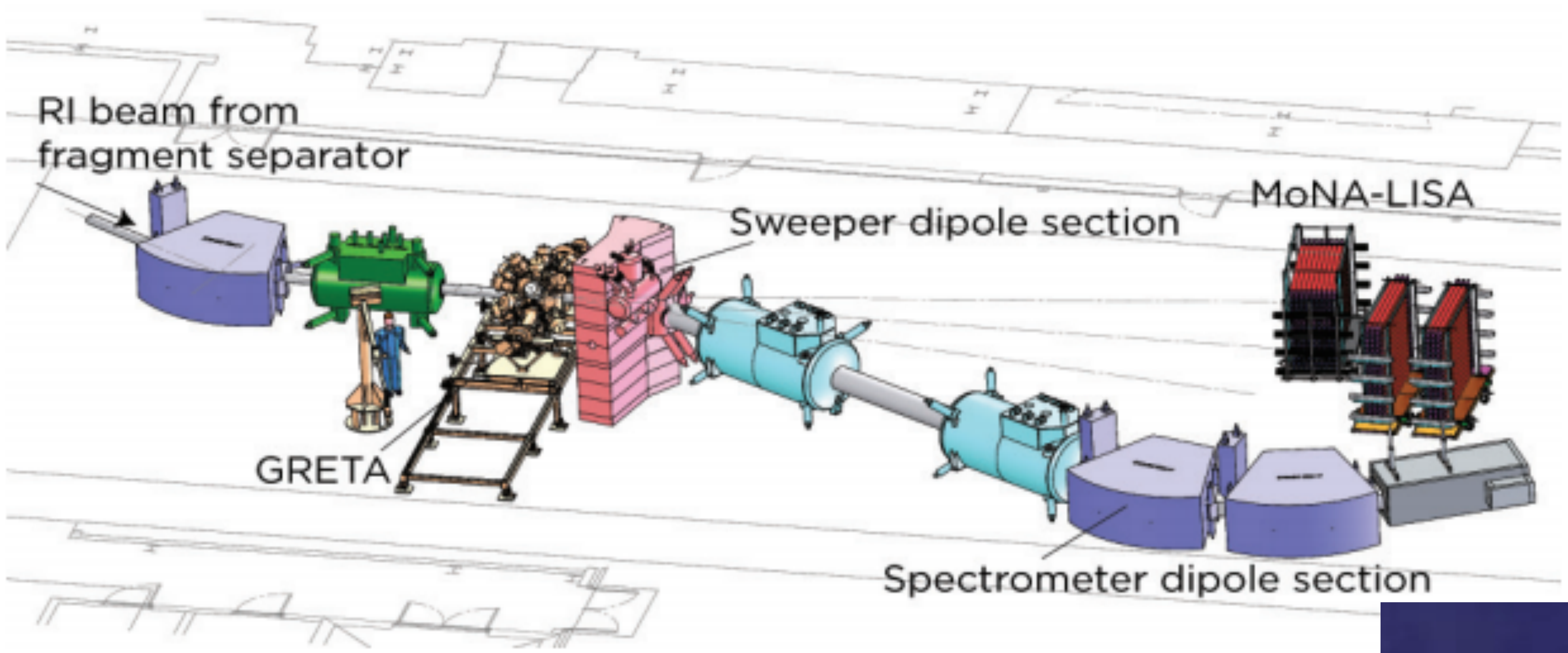


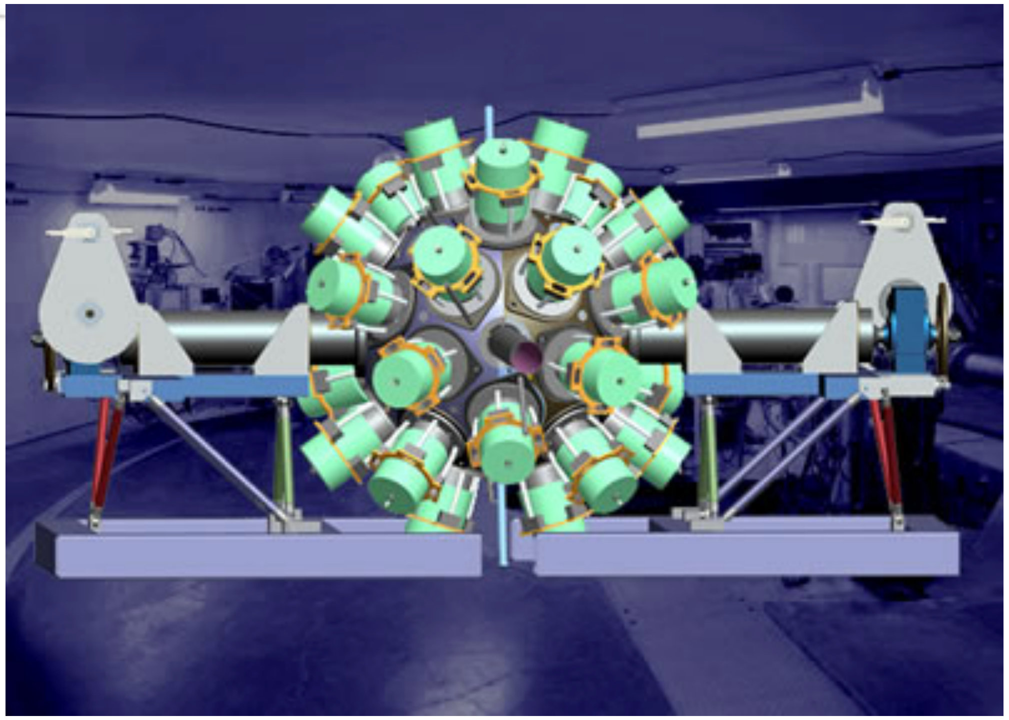

Exascale Requirements
Review for Nuclear Physics
Gaithersburg, 2016

Work supported under contract number DE-AC02-05CH11231.

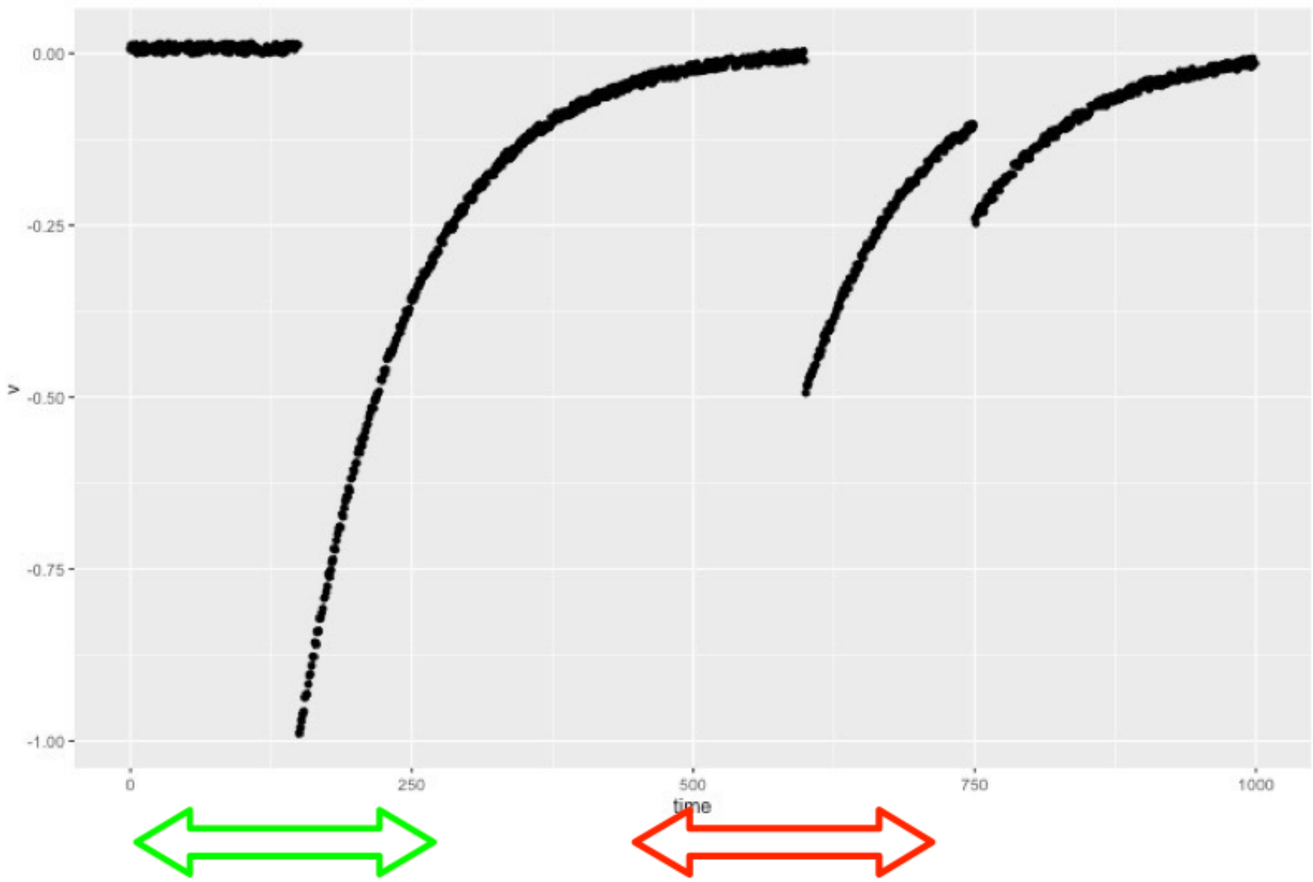
FRIB - GRETA

- Gamma ray spectrometer to be used at FRIB.
- Instrumented by 4000 x 100 MHz 16-bit ADCs.
- 2025 maximum I/O rate 100 MB/s per channel, 400 GB/s aggregate.

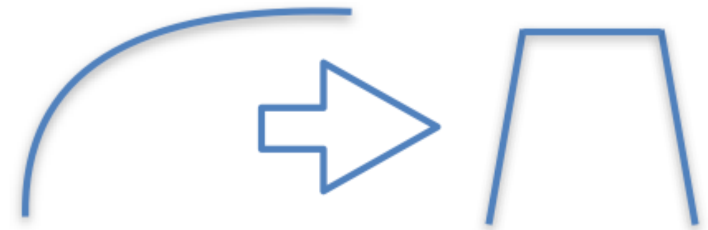




Energy Determination



(pileup)

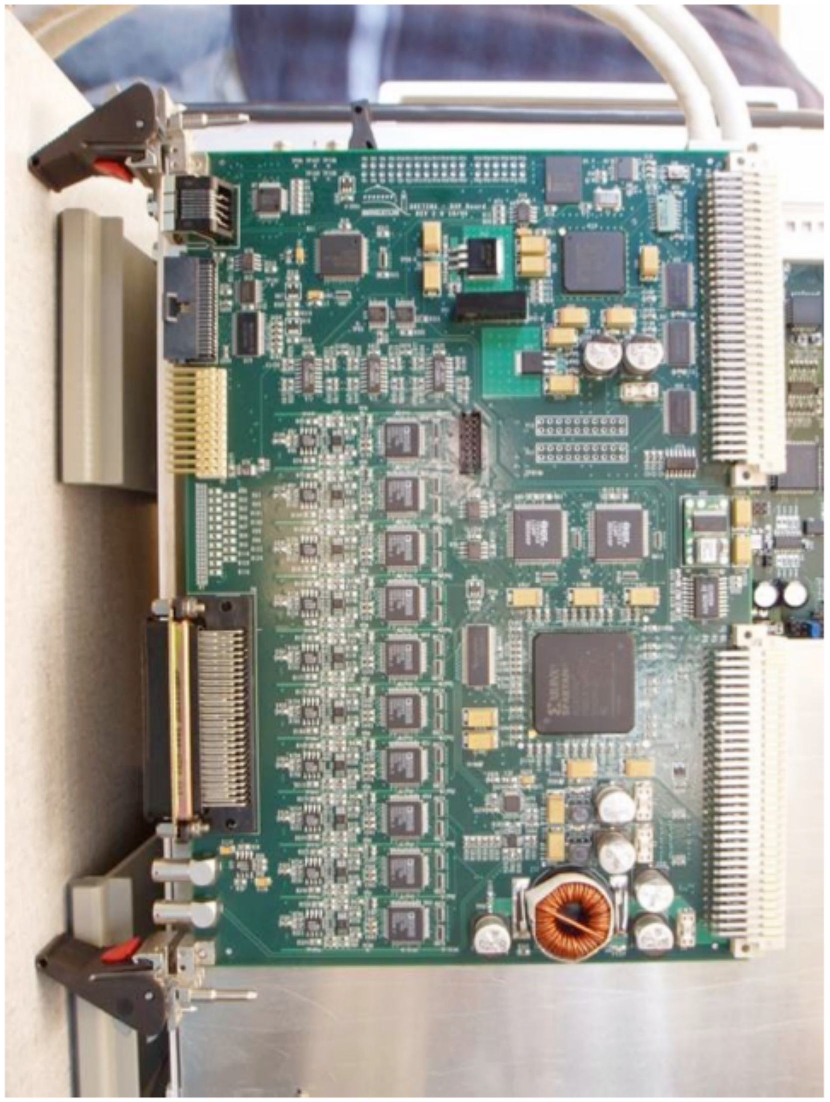


online trapezoidal filter implemented in FPGA memory pipeline (6μs flattop)

- HPGe detectors have very high intrinsic energy resolution (< 0.1 %)
- require long integration times

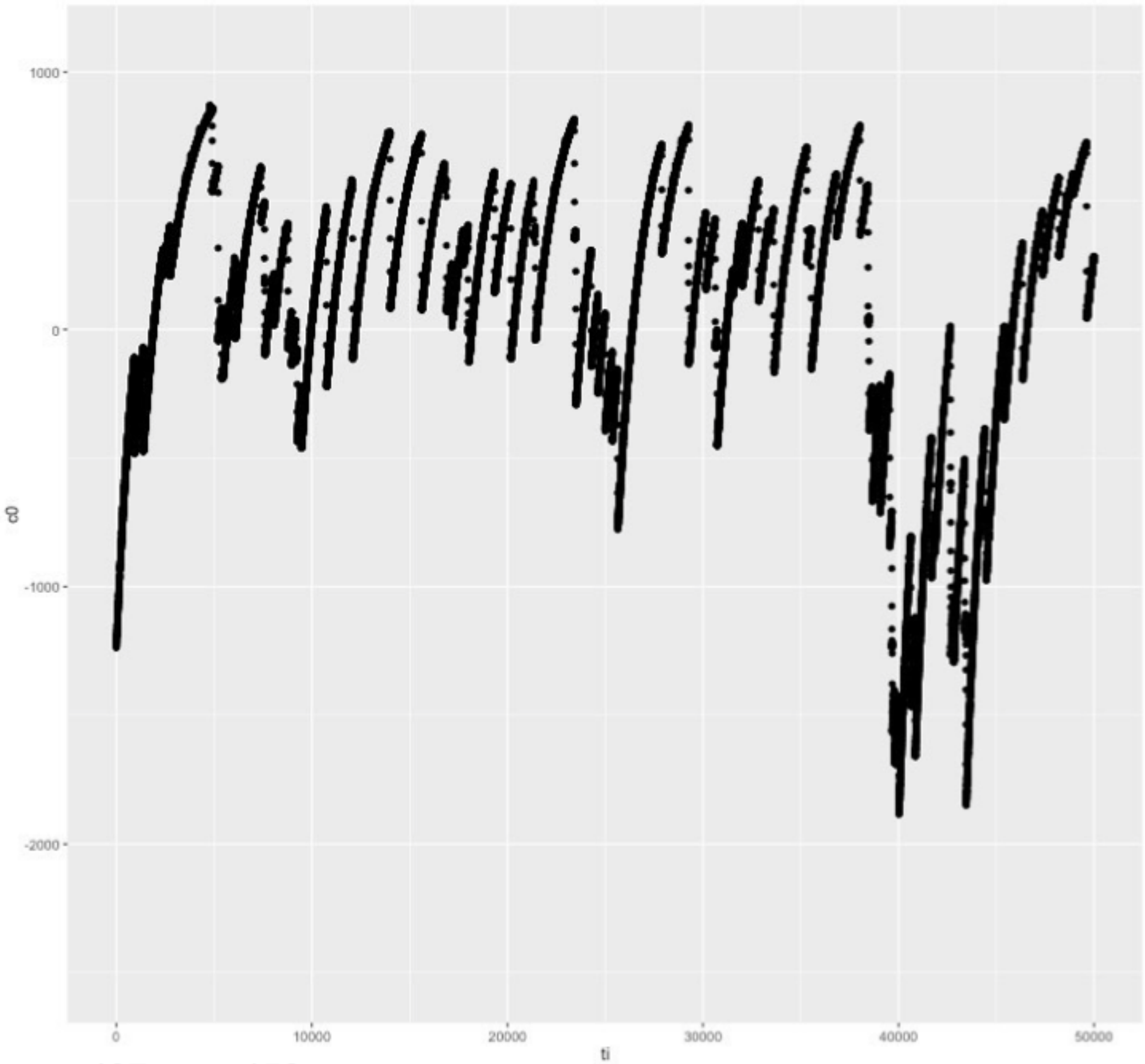
Todays Way

- Two timescales “baked in” to signal processing architecture
- Run the trapezoidal filter algorithm in an FPGA for energy - reject closely spaced signals (pile-up rejection)
- Generate a local trigger primitives from the pipeline - combine to form a global trigger
  - latch energies
  - windows off short sections of trace during charge collection
- Send trace segments to cluster for signal decomposition / tracking



GRETINA 10 ch digitizer board

Waveform taken at 50 kHz



^{137}Cs, ^{152}Eu sources; HPGe clover detector

- Most signals overlap
- Erratic baseline - accounting for electronic response and history is important
- Difficult to maintain 0.1% energy resolution for sizable fraction of events - throughput losses

? - 100 kHz, .. more ..

Streamers

- Faced with new requirements:
  - at the highest rates most of the waveform from the ADC is now necessary for accurate energy determination
  - the algorithms for extracting energy are progressing towards complex fits rather than simple filters
- **Sol'n:** Extract all waveforms in their entirety and perform real-time processing using high-performance computing resources (rather than FPGAs)
- *Can it be implemented? What resources are required?*

Detector simulations in NP: Scope of overview

Hot QCD / phases of nuclear matter

- ALICE
- sPHENIX
- STAR

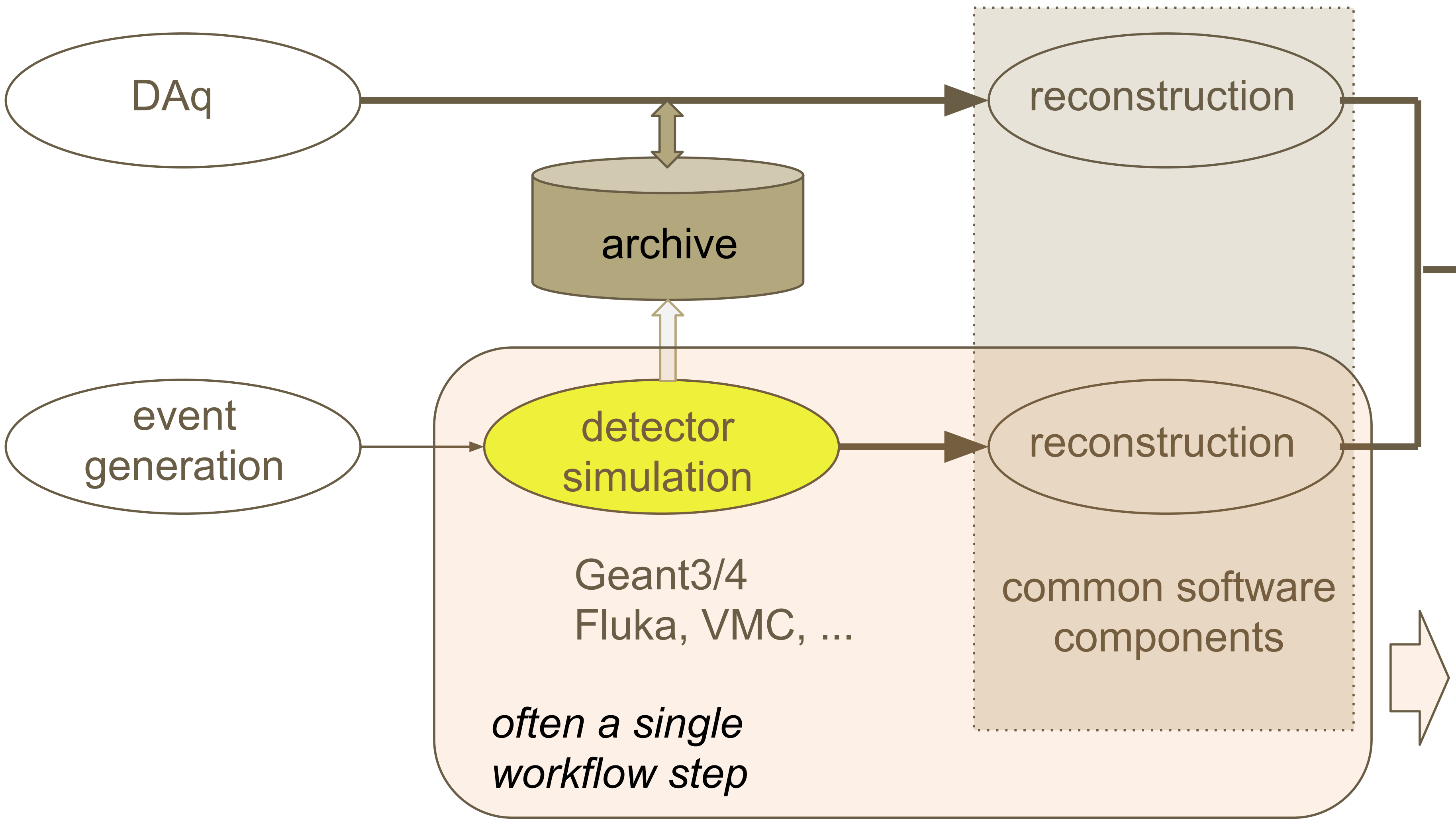
Nuclear structure / reactions

- GRETA
- FRIB spectrometers

Nucleon structure / cold QCD

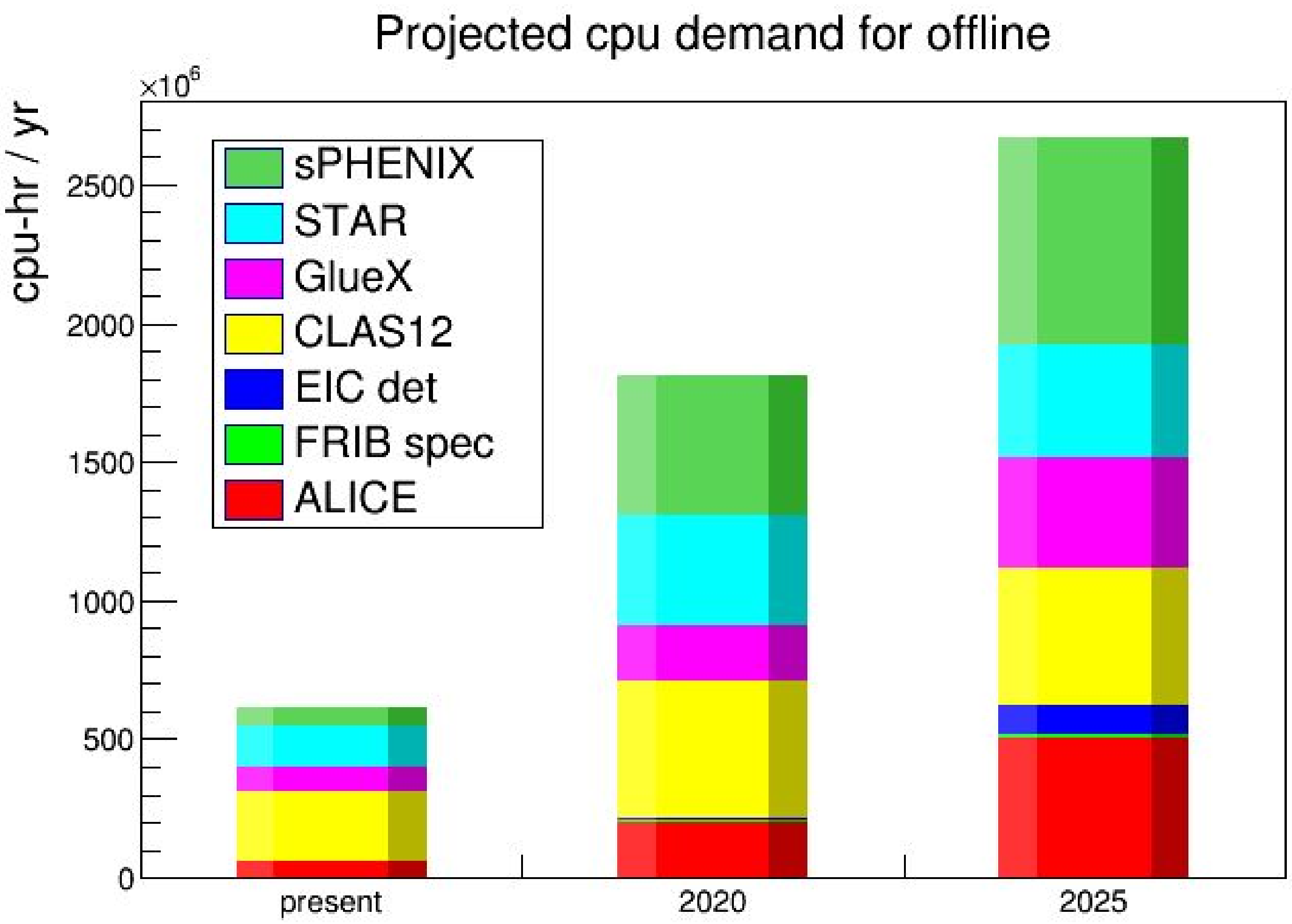
- CLAS12
- GlueX
- STAR
- sPHENIX
- EIC detectors

Detector simulation in NP context (expt)



- **detector optimization**
- **trigger efficiency**
- **resolution**
- **acceptance**
- **backgrounds**

Demand profile (cpu) for NP detector sims



Sims as fraction of total cpu:

ALICE	>50%
FRIB spec	not stated
EIC det	~50%
CLAS12	~75%
GlueX	>50%
STAR	not stated
sPHENIX	>90% now

R. Jones, Exascale requirements review for NP, Gaithersburg, June 16-19, 2016

Observations: general trends

- online / offline distinction is declining
  - driven mainly by **data transport and storage demands**, not cpu
  - growth demands **large special-purpose Tier-0 facilities** with cpu coupled to data

stands in sharp contrast with

- present model: mostly all-purpose homogenous compute resources
  - **common offline infrastructure** for event reconstruction, simulation, and even analysis
  - resource configuration is a **compromise between i/o, memory, cpu** throughput demands for these different tasks

Question: What would a **dedicated simulation resource** look like in 5-10 years?

Observations: concerns

Nearly all parallelism in experimental codes stops at *event-level:*

- gotten for free since the beginning,
- has scaled successfully for a long time, *but...*

This has fostered a certain reluctance to pursue parallelization in the offline,

- **hard to retro-fit** serial codes for significant speed-up (Amdahl's rule)
- new **ground-up designs, restrictive rules, significant effort** difficult to justify
- can be avoided by ***growing the per-core memory resources***

Event complexity is increasing

- present offline facilities: typically 2GB / core
- **pressure from detector simulation** is to increase this -- **double in the next 5 years?**
- ***this is going the wrong way!***

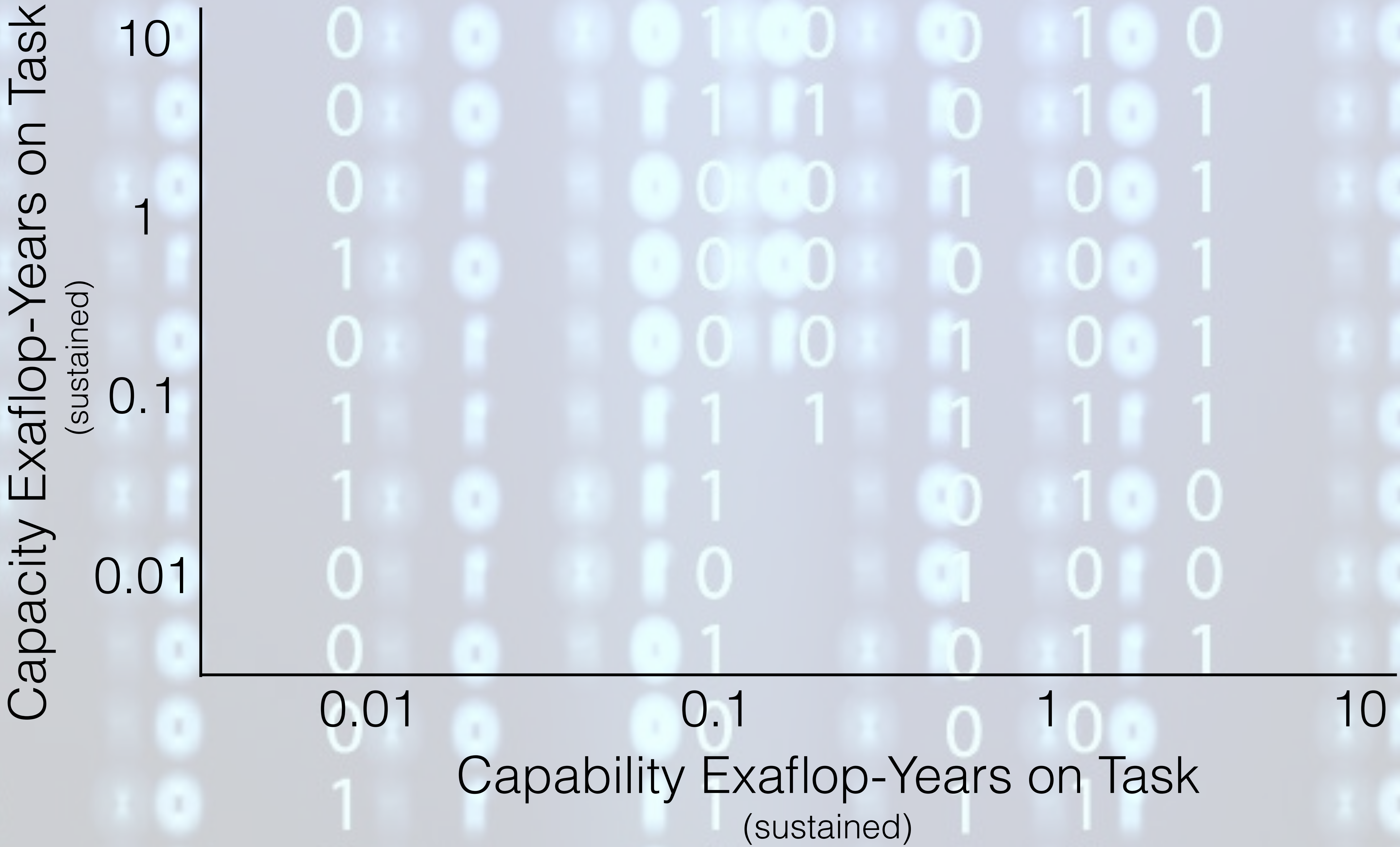
Observations: opportunities

- detector simulation parallelization
  - **Geant4**
    - recently added multi-threading (version 10)
    - still only *event-level parallelism*
    - incremental improvements may be possible at sub-event level
    - toolkit is incorporated into almost every NP experiment ⇒ *big impact*
  - **Geant5**
    - ground-up redesign for fine-grained parallelism
    - goal is full vectorization
    - *significant challenge / huge payoff* -- can benefit from HEP / NP collaboration
- shared virtual facility for NP simulation
  - OSG might serve as a prototype organization
  - might combine "leadership facility" and contributed resources

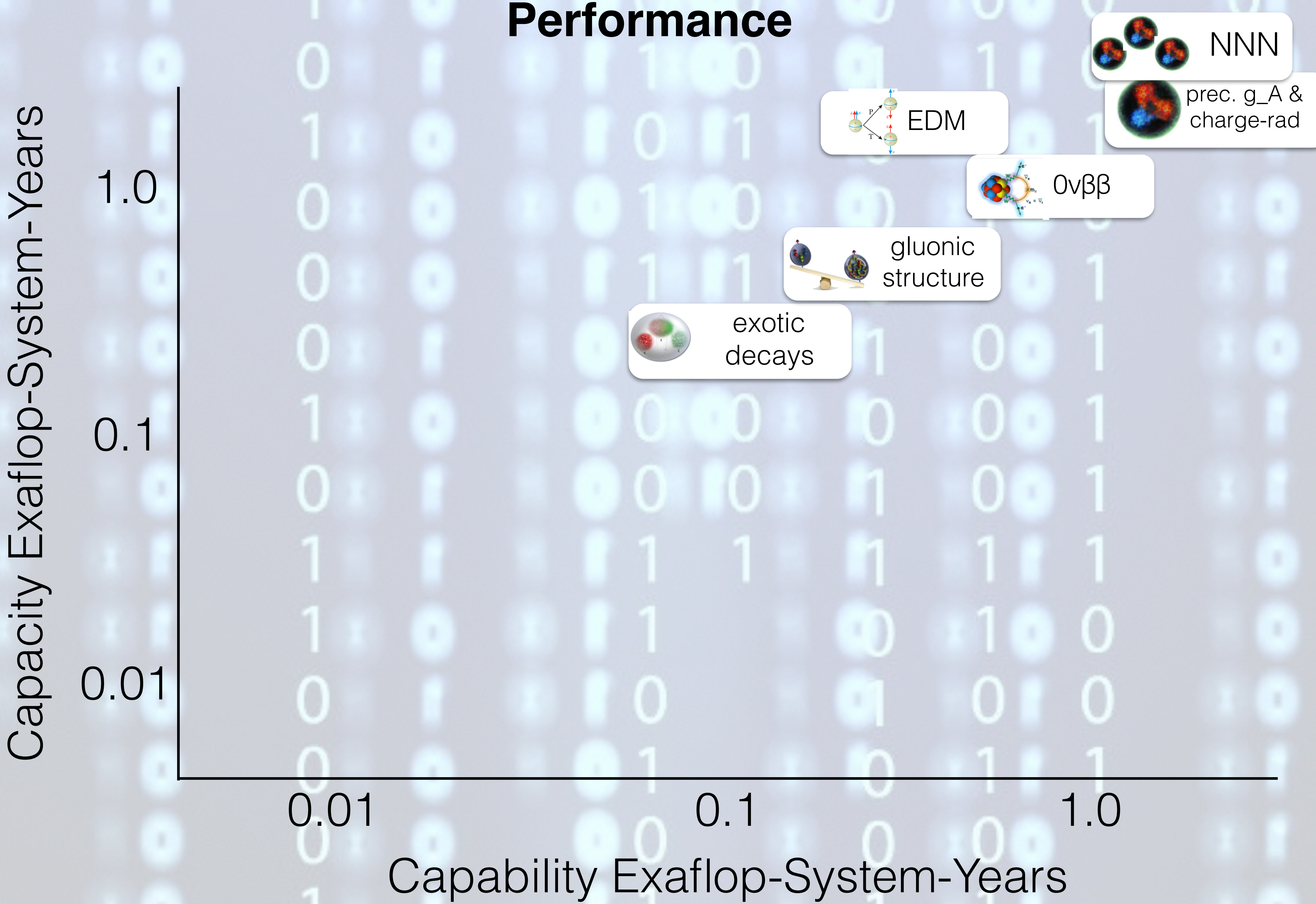
Opportunities for collaboration - DAQ

- Large projects like Exascale computing invariably lead to standards, software packages and hardware technologies that can be of use in the DAQ environment.
- DAQ and HPC face common problems that have been, or will be, solved on the HPC side. The solutions are not well known in the NP DAQ community.
  - DAQ has traditionally relied heavily on custom software.
  - It would be useful to collaborate to identify standards based solutions.
    - Monitoring and control, remote access, high performance storage, data transport, operating systems. streaming data, programming languages

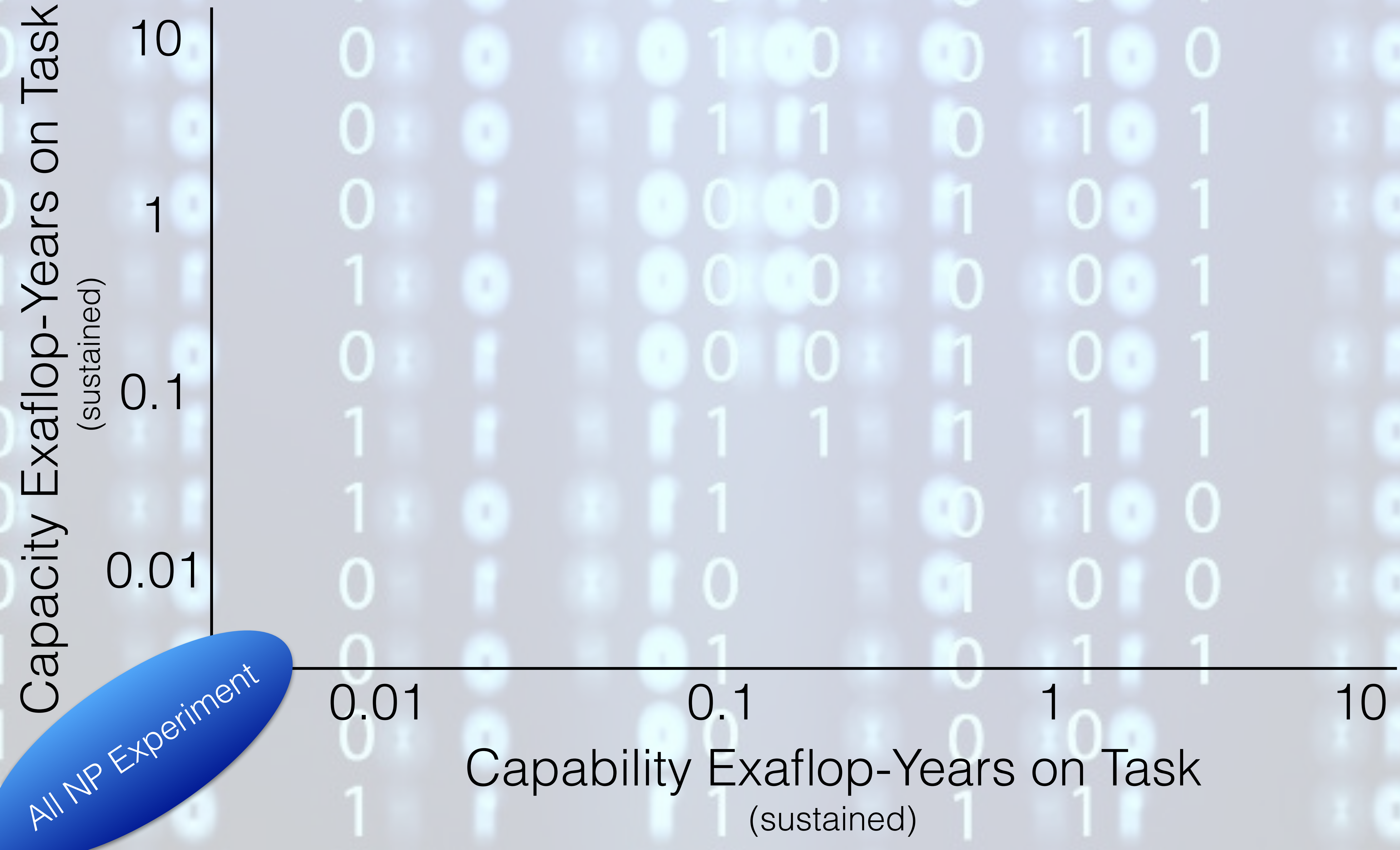
Performance



Performance



Performance



Opportunities for collaboration - Offline

- Detector simulation - GEANT is currently not optimized for massively parallel computing resources, certainly not for leadership class machines.
- Shared facility for NP simulation across the labs?
  - Up to 70% of an experiment's computing is simulation.
  - In principle simulation can run anywhere.
  - If simulation packages could make sufficiently efficient use of a leadership class resource then we could be the "small fish in the big sea" and benefit from currently unused resources.
- Centralized data archiving, outside ASCR scope but may be of interest across DOE science.

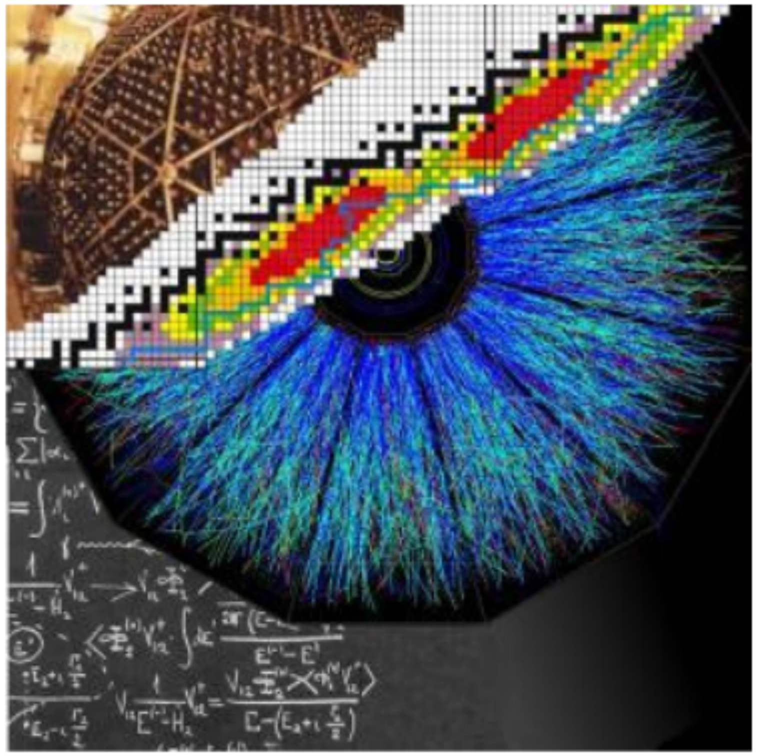
Concluding remarks

- At first glance comparing an extrapolation of current trends of experiment requirements out ten years to a similar extrapolation of technology indicates that ENP computing only gets easier with time.
- Caveats :
  - Technology trends in the next ten years show indications that simple extrapolation may be invalid. Good chance of disruptive technologies emerging.
  - Proposed requirements for new experiments may be based on perceptions of what will be possible that are artificially low.
  - Experiments are being proposed that are on the edge of the possible even with future technologies
  - Requirements are based on a workflow that may be non-optimal - there are other ways of doing things that are not accessible now but may be in ten years.
- All of the labs have recognized these issues and are gravitating towards a streaming solutions with HPC close to the experiment.
- Advances towards Exascale computing and other advanced computing projects will have a considerable impact on how NP DAQ and analysis are done.

NP Experiment and Data

Science Objectives

- Unfold the quark and gluon structure of hadrons and nuclei
- Realize a predictive model of nuclei and their role in the cosmos
- Tests of the particle-antiparticle nature of neutrinos and other fundamental symmetries and neutrino research that open new doors to physics beyond the Standard Model
- Study the properties and phases of quark and gluon matter in the high temperatures of the early universe and to explore the spin structure of the proton.



Composite image showing data aspects of the NP program

Methodology

Make optimum use of NP experimental facilities

- World-leading particle accelerators and detector facilities
- Reliable and safe facility operations
- Highly efficient detector and data acquisition systems
- Trigger defines the data,"real time" validation critical
- Online/offline data analyses are different, but gap is diminishing
- Simulated and observed data processed by same analysis chain
- Complex (distributed) workflows that may be adaptive and/or end-to-end optimized
- Plethora of ways to treat data by different experiments
- Some common building blocks used for modeling and analysis of data (Geant, ROOT, etc.)

Resource Requirements

People, time, infrastructure and (relevant) knowledge limited.

- Topical collaborative efforts between ASCR and NP researchers
- Realize streaming and real-time event processing
- Direct coupling of detectors with HPC resources
- Network infrastructure for improved data distribution and access (both online and offline)
- Best practices and high performance data management
- Incorporate modern ML techniques in data processing.
- Employing state-of-the-art visualization tools
- Evolutionary development for distributed workflows inclusive of resource diversity
- Optimized simulation tools fully utilizing HPC resources