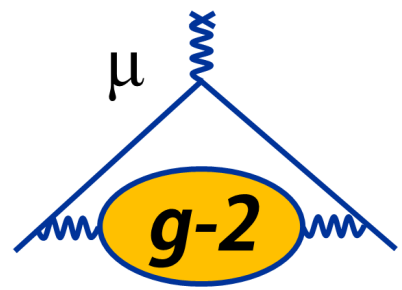
Data Acquisition with GPUs
The DAQ for the Muon g-2 Experiment at Fermilab

Wesley Gohn

University of Kentucky

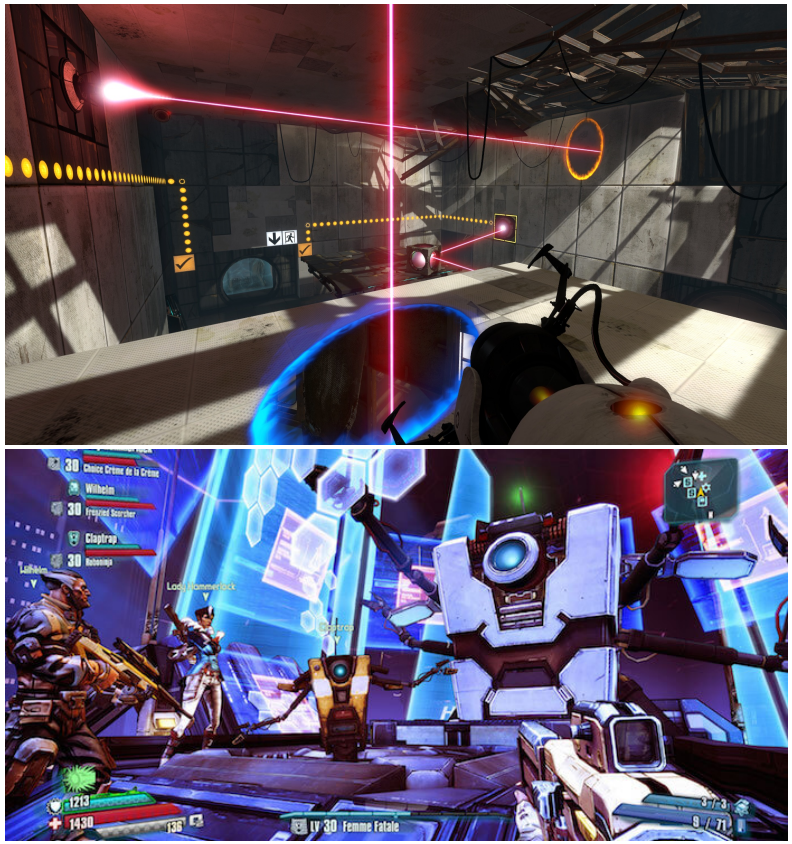
August 4, 2016



Outline

Why GPUs?

- Potential to dramatically improve performance at a reasonable cost by parallelizing data processing.
- Good for performing a simple algorithm many times, but less good for complex algorithms.
- Technology was developed initially for commercial applications, so it is well supported.
- Reasonably easy to learn programming environments are available.

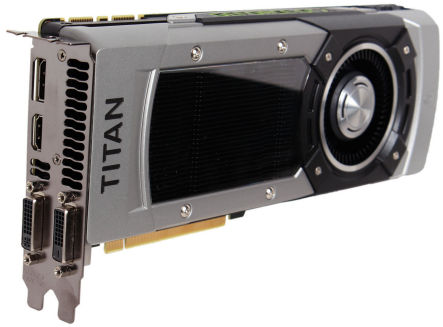


GPU Architectures

Scientific GPU
(i.e. Nvidia Tesla K40)

- 2880 CUDA cores at 740 MHz.
- 288 GB/s memory bandwidth
- 12 GB memory
- ECC memory protection.

Gaming GPU
(i.e. Black Titan)



- 2880 CUDA cores at 980 MHz.
- 336 GB/s memory bandwidth
- 6 GB memory
- No ECC memory protection.

Xeon Phi

- 68 cores / 272 threads at 1.4 GHz.
- 34 MB L2 Cache.
- PCIe version 2.0 limits data transfer bandwidth.

Other Hardware Considerations

The motherboard matters.

- Want a motherboard that supports PCIe version 3.0.
- Need a motherboard with a PLX chip, which allows it distribute the load over 40 lanes in order to operate all PCIe slots at 16x.
- We are using the ASUS X99-E WS/USB 3.1.

PCIe version 2.0 vs 3.0.

PCIe version	GPU	Host to device, Pageable	Host to device, Pinned
2.0	K20	3326.6 MB/s	5028.3 MB/s
3.0	K20	5628.6 MB/s	6003.6 MB/s
3.0	K40	6647.8 MB/s	10044.3 MB/s

GPU Programming

The choice: CUDA vs OpenCL?

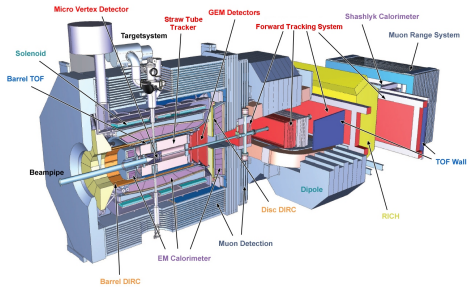
- Proprietary from Nvidia.
- Only works on Nvidia GPUs.
- More efficient code.
- Detailed documentation and references from Nvidia.

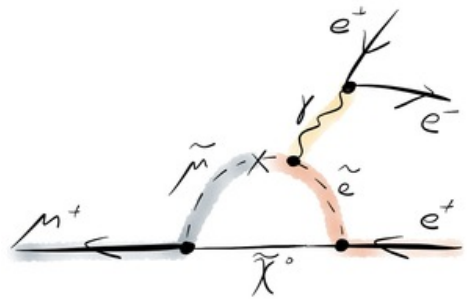
- Open Source.
- Works on many architectures, including Nvidia.
- Relies on community for support.

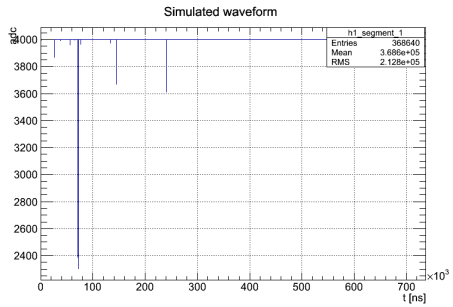
For our g-2 GPU programming, we are using CUDA.

GPUs in Particle Physics trigger and DAQ

- $\bar{P}$**ANDA at FAIR:** Online tracking of charged particle tracks (L. Bianchi *et al.*, J.Phys.Conf.Ser. 664 (2015) no.8, 082006)
- **NA62 at CERN:** Reconstructing ring-shaped hit patterns in a Čerenkov detector. (R. Ammendola *et al.*, arXiv:1606.04099 (2016))
- **Mu3e at PSI:** Track and vertex reconstruction, D. vom Bruch (2015)
- **LHCb at CERN:** Track pattern recognition S. Gallorini (2015)
- **Muon g-2 at Fermilab:** Deadtime-free analyzer of 700 μs waveforms – the rest of this talk.

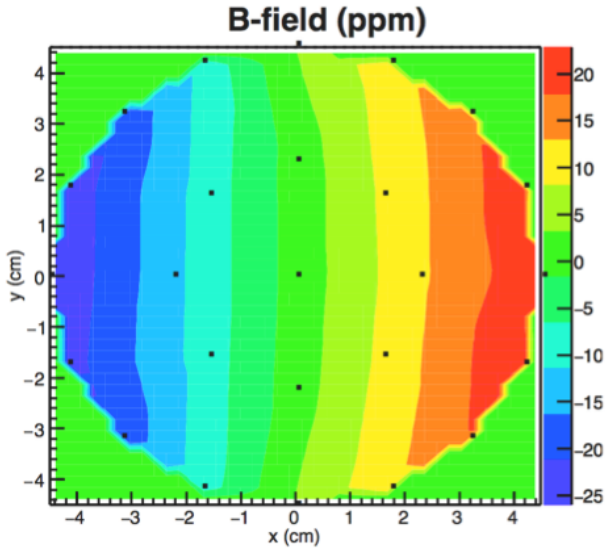






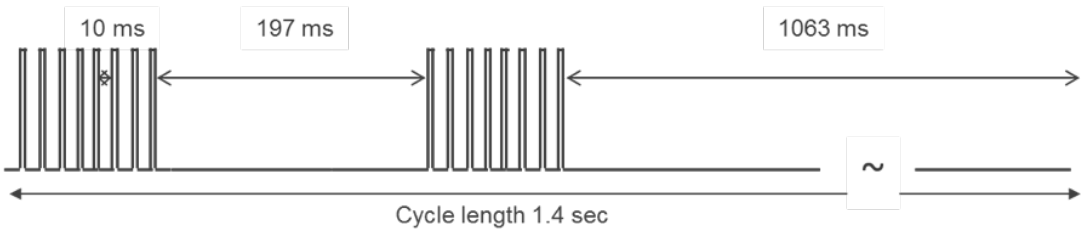
Muon g-2 Status

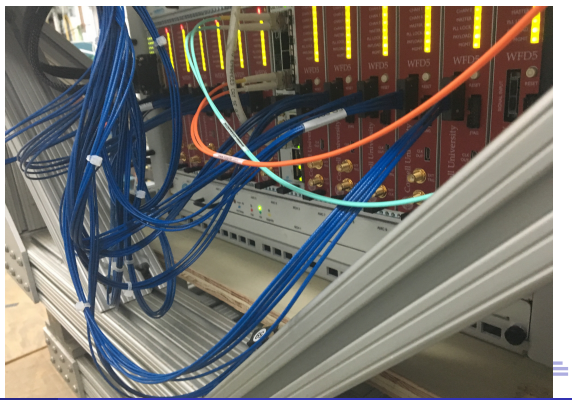
- We moved a 50' superconducting magnetic ring from BNL to FNAL in 2013 and fully installed it last year (see talk by V. Tischenko on details of muon storage).
- Measurement and stabilization of the magnetic field is nearly complete (see poster by B. Kiburg).
- Detector systems will be installed this year (see talk by J. Kaspar).
- Plan for data taking to begin in early 2017 (see talk by C. Polly tonight for more details).



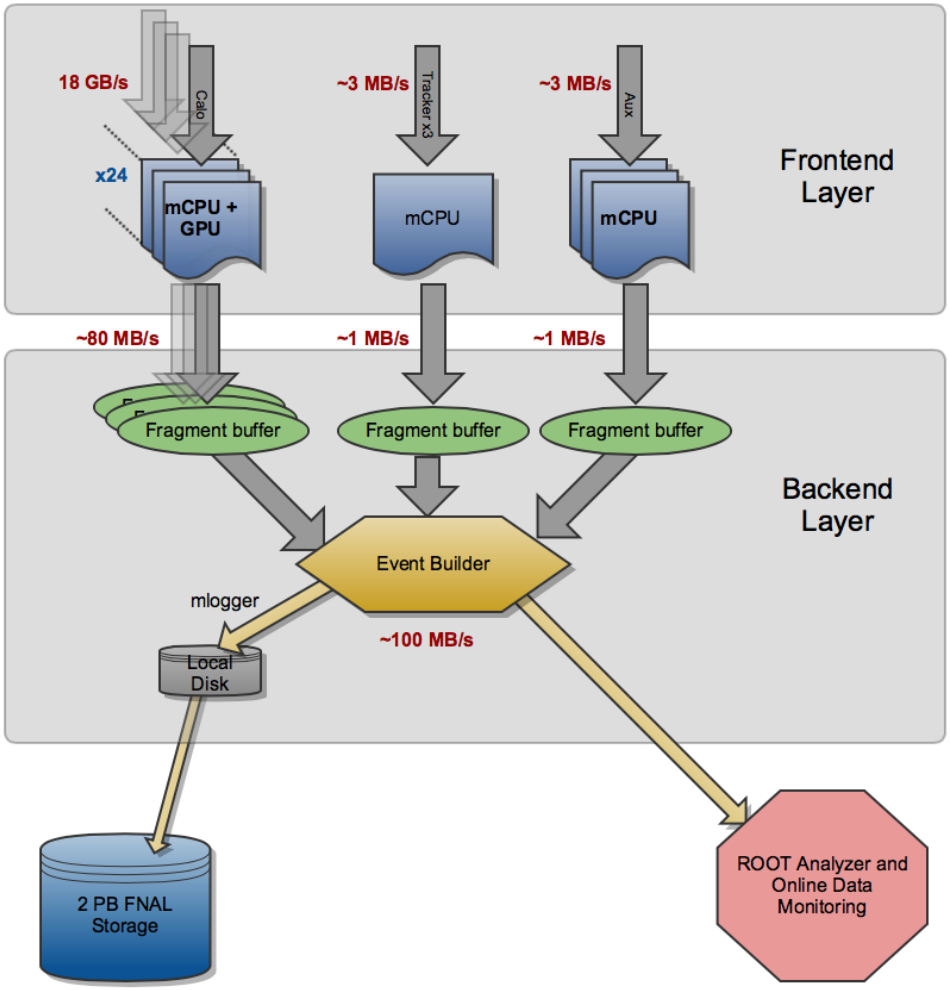
Experiment details and DAQ requirements

- Muons will be injected into the ring at an average rate of 12 Hz consisting of sequences of eight successive 700 μs fills with 11 ms fill separation.
- Positrons resulting from muon decays are detected using twenty-four electromagnetic calorimeters, each comprised of 9×6 arrays of PbF_2 crystals read out by SiPMs.
- The full 700 μs waveforms are read out by 1296 channels of custom μTCA 800 MHz, 12-bit, waveform digitizer (see poster by D. Sweigart for details).
- For a 12 Hz spill rate the time-averaged rate of raw ADC samples is 18.6 GB/s in total.





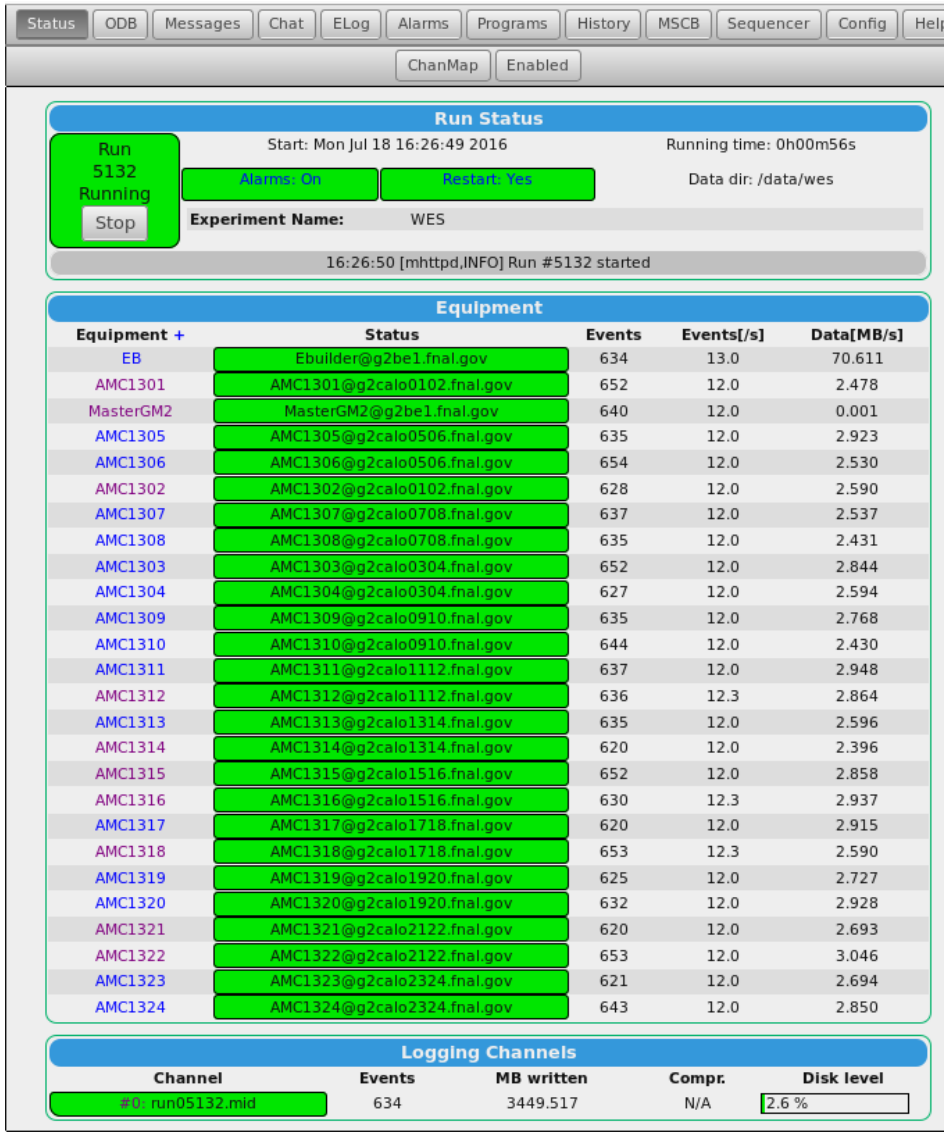
DAQ Schematic



- Layered array of commodity, networked processors
- FE layer for readout of digitizer (calo), MHTDCs (straws)
- BE layer for assembly of event fragments, storage
- Slow control layer for setting, monitoring of HVs, etc.
- Online analysis layer using *art*+JS for monitoring the integrity of raw data, physics data.

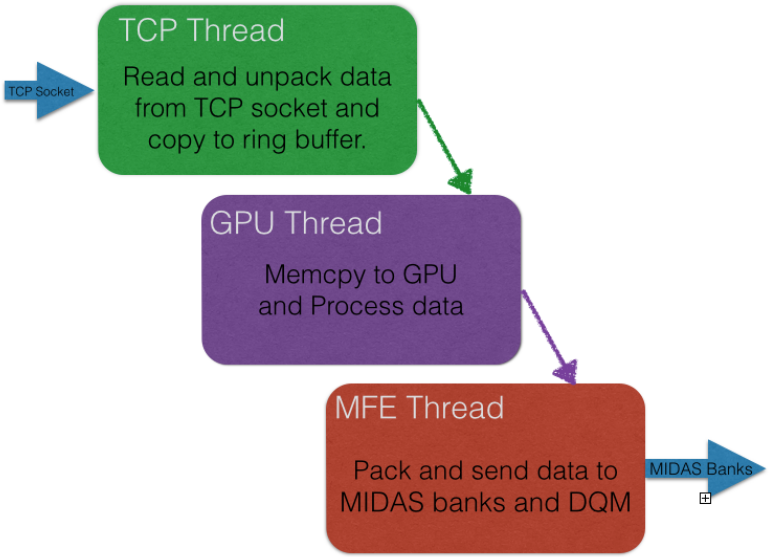
MIDAS

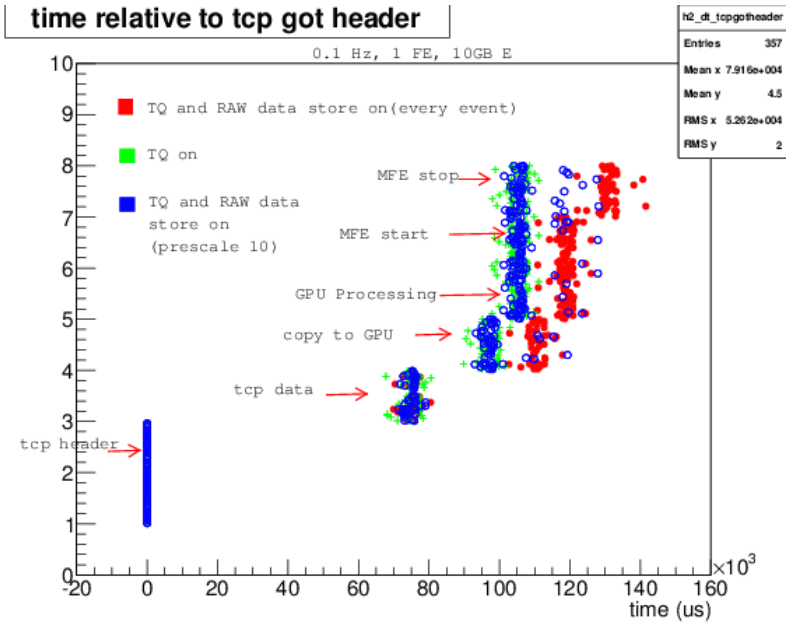
- MIDAS is a data acquisition software developed at PSI and also used extensively at TRIUMF.
- Includes web interface for easy control.
- Frontend acquisition code written in C/C++ with CUDA.
- Javascript based analyzer for online data monitoring via a web gui.
- Data will be written to tape as MIDAS datafiles.



Code structure

CPU multithreading with mutex locks





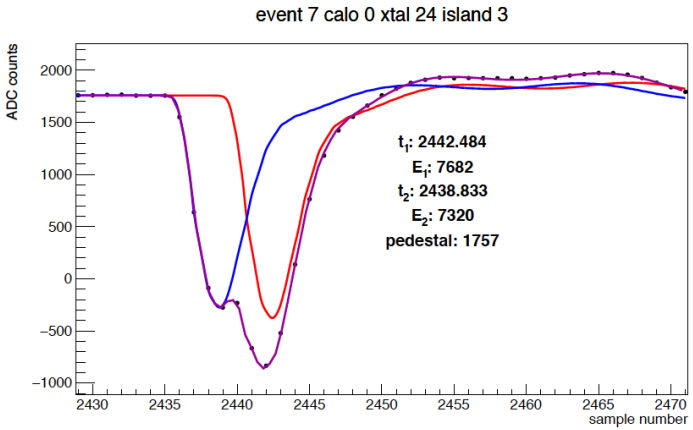
CUDA kernel routines: Data is copied to GPU memory in GPU thread, and then accessed by the following functions to identify and save islands, which are copied back to the computer memory and sent to the event builder.

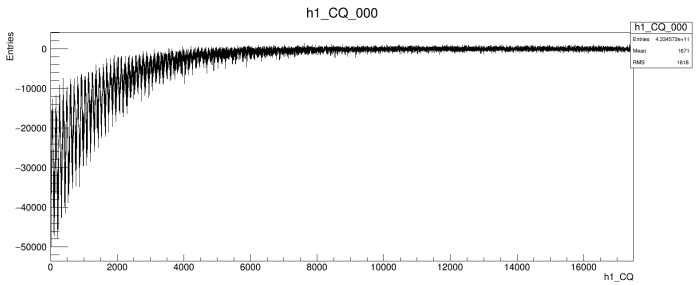
Function	Number of threads	Execution time (ms)
Compute pedestals as average of first 100 samples	54	0.1
Determine if threshold is passed for each sample	560k	1.7
Add pre-samples and post-samples to each island	560k	0.1
Check to see if any islands have merged	560k	0.2
Save an array of identified islands	560k	0.2
Sum all waveforms	560k	1.2
Decimate the sum for the Q-method	17.5k	0.3
Make a fill-by-fill sum of waveforms	30M	2.4

T and Q Methods

CUDA routines process data with two complimentary methods.

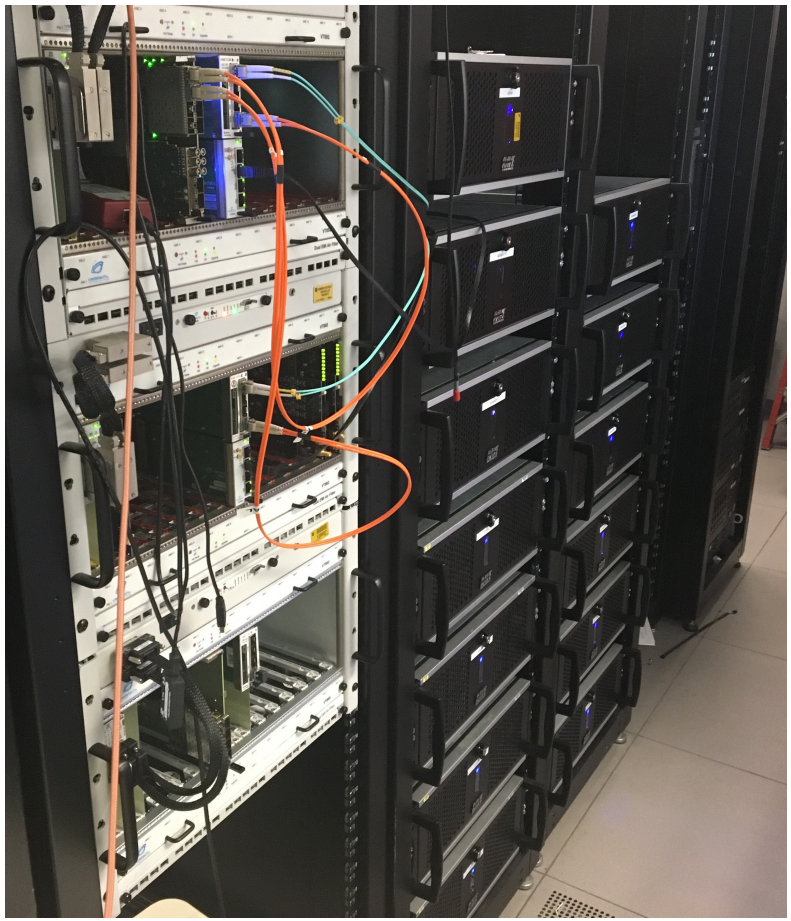
- T-method
  - Positron events in the calorimeter are individually identified, sorted and fit to obtain time and energy.
  - All events above an energy threshold are included.
  - $\vec{\omega}_a$ is determined from a fit to a pileup-subtracted histogram.
  - This was the method used in BNL E821.
- Q-method
  - Individual positron events are not identified.
  - Detector current is integrated as a proxy for event energy.
  - No pileup correction is necessary.





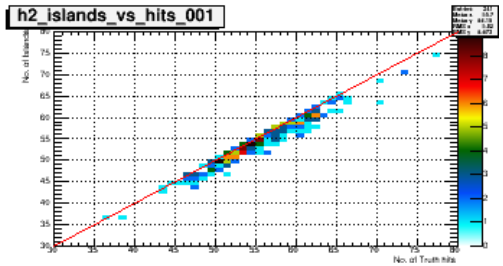
Infrastructure installation

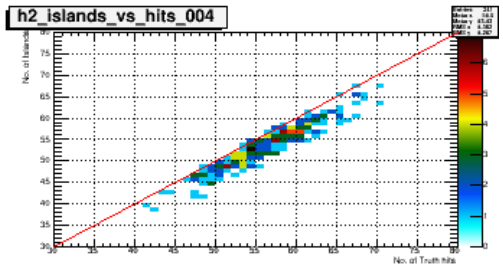
- GPU cluster containing 28 Nvidia Tesla K40 GPUs is up and running.
- Three of the four required backend computers are operational.
- We have 40 TB of RAID volume installed for short-term data storage.
- Fiber-optic 10 Gbe network has been installed.
- A Meinberg GPS unit is used to timestamp fills in the DAQ.

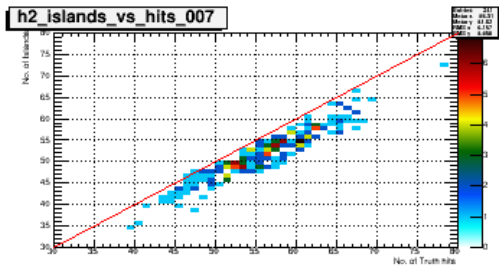


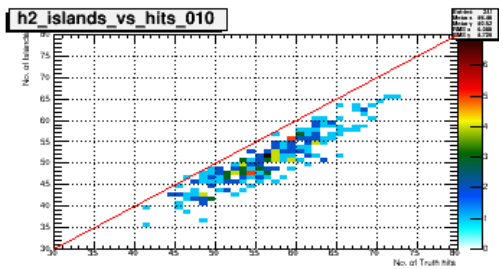
DAQ testing

- DAQ has been tested using a simulator that generates realistic data rates for all 24 calorimeters, which allows us to test our GPU pulse-finding algorithms.
- The full DAQ was used to read data from a full calorimeter at a SLAC test beam in June, 2016.
- Characterization of the 1296 WFDs required for the experiment is underway.
- Full commissioning with all detector systems will take place starting in October, 2016.







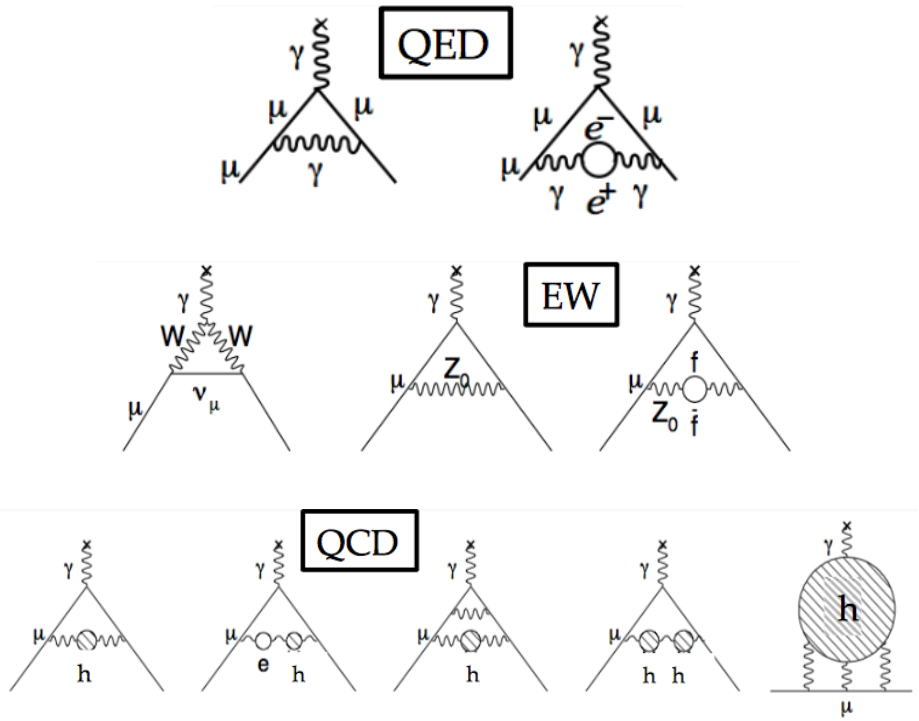


Conclusion

- The new muon g-2 experiment will run at Fermilab beginning in 2017 with the goal of reaching 20$\times$ the BNL statistics.
- A new state-of-the-art data acquisition system utilizing parallel data processing in a hybrid system of multi-core CPUs and GPUs is required to achieve the necessary data rates.
- The DAQ will acquire data from 1296 channels of custom μTCA waveform digitizers, as well as straw trackers and auxiliary detectors at a rate of 18 GB/s.
- The construction of the DAQ has been completed (with a few small pieces to be added), and commissioning is underway.
- First muons expected in March, 2017.

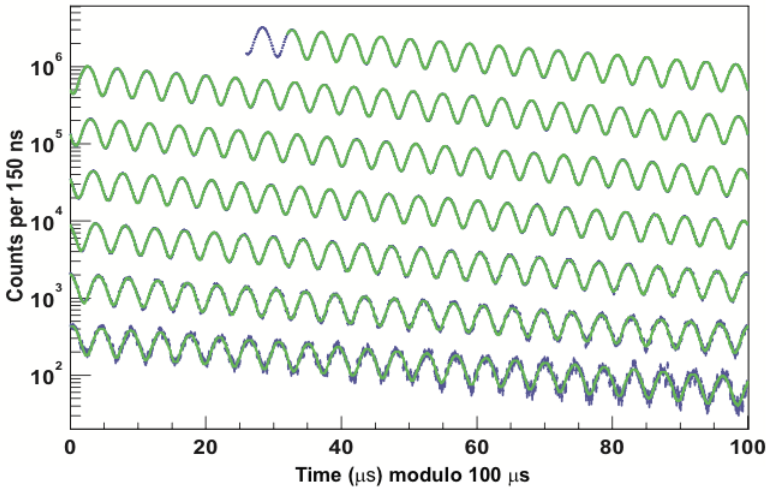
Physics of Muon g-2

- In the standard model, the muon is a spin 1/2 pointlike particle.
- It has a magnetic dipole moment of $\vec{\mu} = g\frac{q}{2m}\vec{s}$, with $g = 2$ for a pointlike particle (Dirac)
- Additional effects from QED, electroweak theory, and hadronic factors move the standard model prediction of g away from 2. It has become customary to measure this discrepancy, g-2.
- If a discrepancy with the standard model value is found, beyond standard model contributions to g-2 could come from SUSY, dark photons, or other new physics (NP).



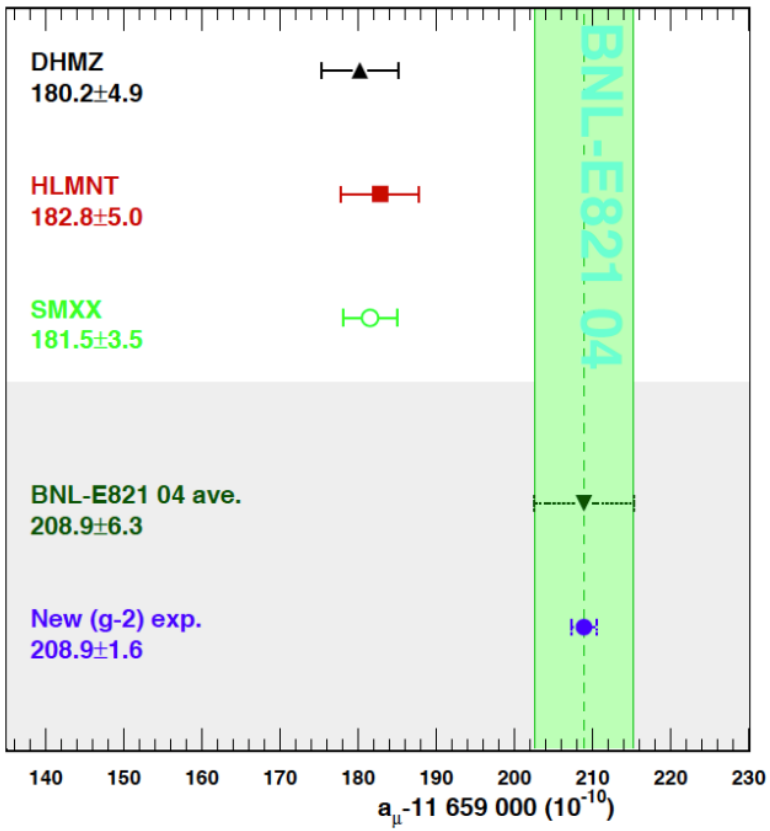
$$a_\mu = a_\mu^{QED} + a_\mu^{EW} + a_\mu^{QCD} + a_\mu^{NP}$$

Measurements of g-2



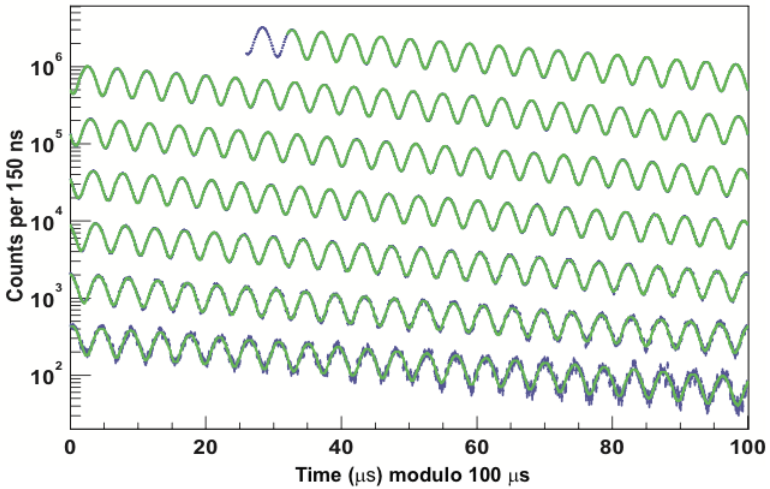
$$a_\mu \equiv \frac{g-2}{2}$$

$$\vec{\omega}_a = -\frac{Qe}{m}[a_\mu \vec{B} - (a_\mu - (\frac{mc}{p})^2)\frac{\vec{\beta}\times\vec{E}}{c}]$$



- BNL E821 measured g-2 to have a 3.3σ discrepancy from the standard model (2006).
- Fermilab E989 will measure 20 times the number of muons, reducing the uncertainty on this measurement by a factor of 4.
- Without theory improvements, discrepancy could reach $> 5\sigma$.

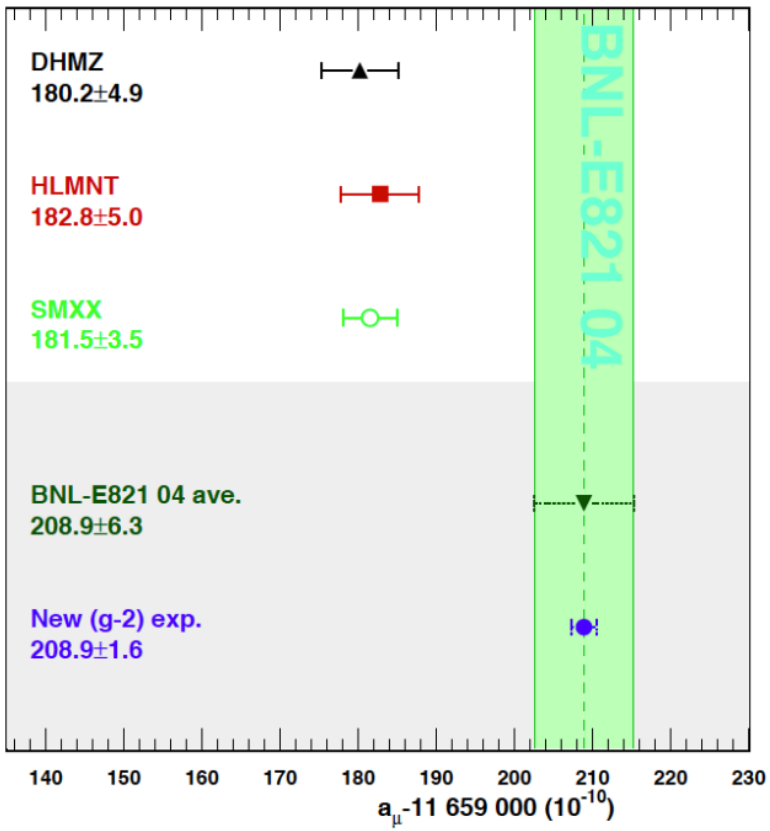
Measurements of g-2



$$a_\mu \equiv \frac{g-2}{2}$$

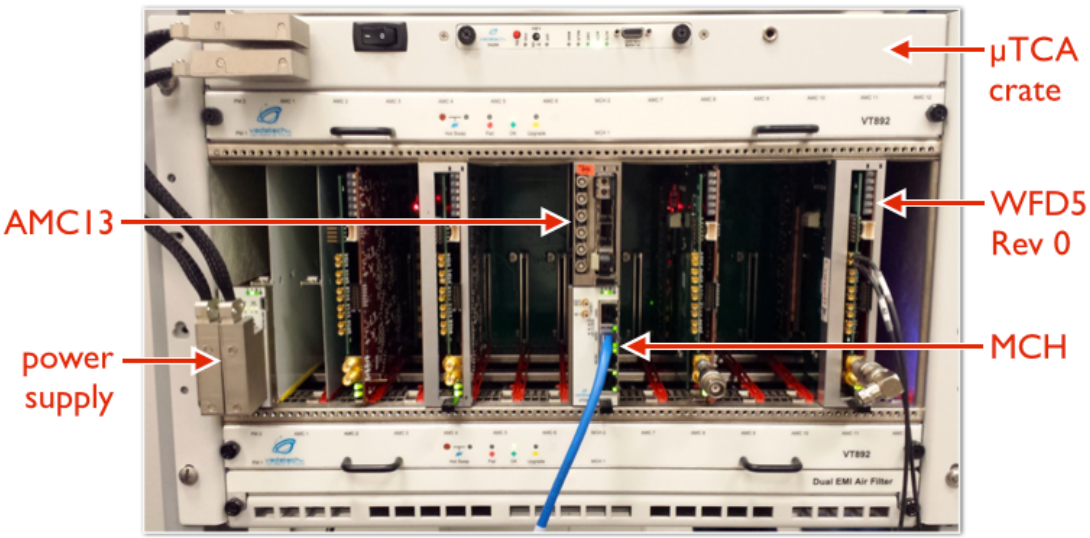
$$\vec{\omega}_a = -\frac{Qe}{m}[a_\mu \vec{B} - (a_\mu - (\frac{mc}{p})^2)\frac{\vec{\beta}\times\vec{E}}{c}]$$

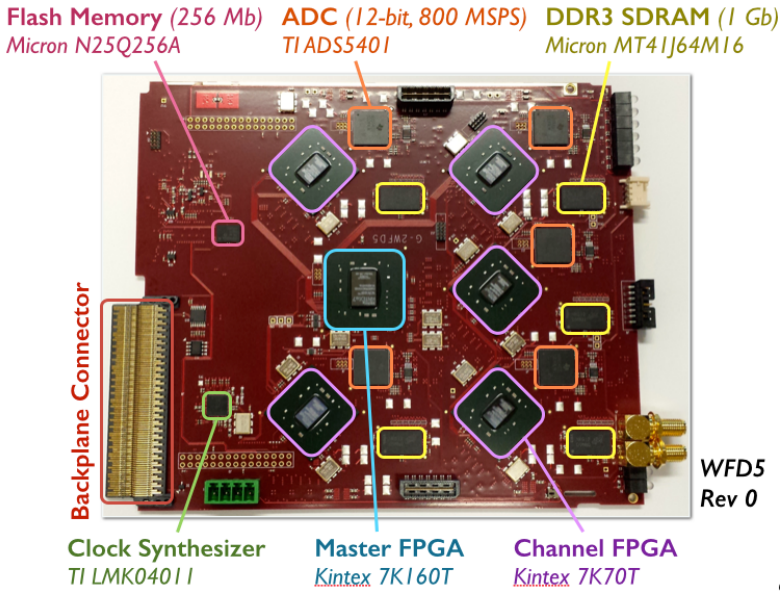
- BNL E821 measured g-2 to have a 3.3σ discrepancy from the standard model (2006).
- Fermilab E989 will measure 20 times the number of muons, reducing the uncertainty on this measurement by a factor of 4.
- Without theory improvements, discrepancy could reach $> 5\sigma$.



Detectors and Backend Electronics

- Measurement will utilize 24 calorimeters (each composed of 54 PbF_2 crystals read out by SiPMs), 3 straw trackers, and several auxiliary detectors.
- Each calorimeter will be readout by a custom WFD in a μTCA crate with an AMC13 control module controlled via IPBus.





Images courtesy of David Sweigart

GPU Processing

- Data will be processed in an array of 24 GPUs (One GPU per calorimeter)
- Utilizing NVIDIA TESLA K40 GPUS
  - Peak double precision floating point performance: 1.43 Tflops
  - Peak single precision floating point performance: 4.29 Tflops
  - Memory bandwidth 288 GB/sec
  - Memory size (GDDR5): 12 GB
  - CUDA cores: 2880
- Data processing code is written using CUDA.

Results of bandwidth tests:

Frontend Machine	GPU	Host to device, Pageable	Host to device, Pinned
FE01	K20	3326.6 MB/s	5028.3 MB/s
RAVE01	K20	5628.6 MB/s	6003.6 MB/s
RAVE01	K40	6647.8 MB/s	10044.3 MB/s

Event building test

MIDAS experiment "UKY" — Wed May 29 20:08:06 2013 — Refr:60

Stop | ODB | Messages | Alarms | Programs | Config

RunLog | Logbook | Elog | Doc

Run #2071 | Running | Alarms: On | Restart: No | Data dir: /data/UKY/mid

Start: Wed May 29 20:07:16 2013 | Running time: 0h00m50s

Equipment	Status	Events	Events[/s]	Data[MB/s]
MagicBox	magic_box@mb	0	0.0	0.000
VMEcrate	(frontend stopped)	0	0.0	0.000
masterMT	(frontend stopped)	365	0.0	0.000
EB	Ebuilder@be	0	0.0	0.000
ATS9870	(frontend stopped)	0	0.0	0.000
EMC	(frontend stopped)	5	0.0	0.000
master	master@fe02	574	11.9	0.001
FakeCalo01	(frontend stopped)	0	0.0	0.000
FakeData01	FakeData01@fe01	564	12.0	2.058
FakeData02	FakeData02@fe01	578	12.0	2.057
FakeData03	FakeData03@fe01	555	12.0	2.053
FakeData04	FakeData04@fe01	566	12.0	2.059
FakeData05	FakeData05@fe01	575	11.7	2.002
FakeData06	FakeData06@fe01	551	12.0	2.058
FakeData07	FakeData07@fe01	564	12.0	2.059
FakeData08	FakeData08@fe01	576	12.0	2.059
FakeData09	FakeData09@fe01	551	11.6	1.999
FakeData10	FakeData10@fe01	563	12.0	2.059
FakeData11	FakeData11@fe01	573	12.0	2.059
FakeData12	FakeData12@fe01	551	11.9	2.052
FakeData13	FakeData13@fe01	561	12.0	2.058
FakeData14	FakeData14@fe01	571	12.0	2.057
FakeData15	FakeData15@fe01	547	12.0	2.060
FakeData16	FakeData16@fe01	558	12.0	2.059
FakeData17	FakeData17@fe01	570	12.0	2.057
FakeData18	FakeData18@fe01	544	12.0	2.057
FakeData19	FakeData19@fe01	555	12.0	2.057
FakeData20	FakeData20@fe01	567	11.6	1.997
FakeData21	FakeData21@fe01	578	12.0	2.059
FakeData22	FakeData22@fe01	555	12.0	2.060
FakeData23	FakeData23@fe01	567	12.0	2.059
FakeData24	FakeData24@fe01	578	12.0	2.059
FakeCaloNewQ01	(frontend stopped)	0	0.0	0.000
CaloSimulatorTCPIP01	(frontend stopped)	0	0.0	0.000
CaloReadoutTCPIP01	(frontend stopped)	0	0.0	0.000

Channel	Events	MB written	Compression	Disk level
#0: run02071.mid	579	99.383	N/A	63.9 %

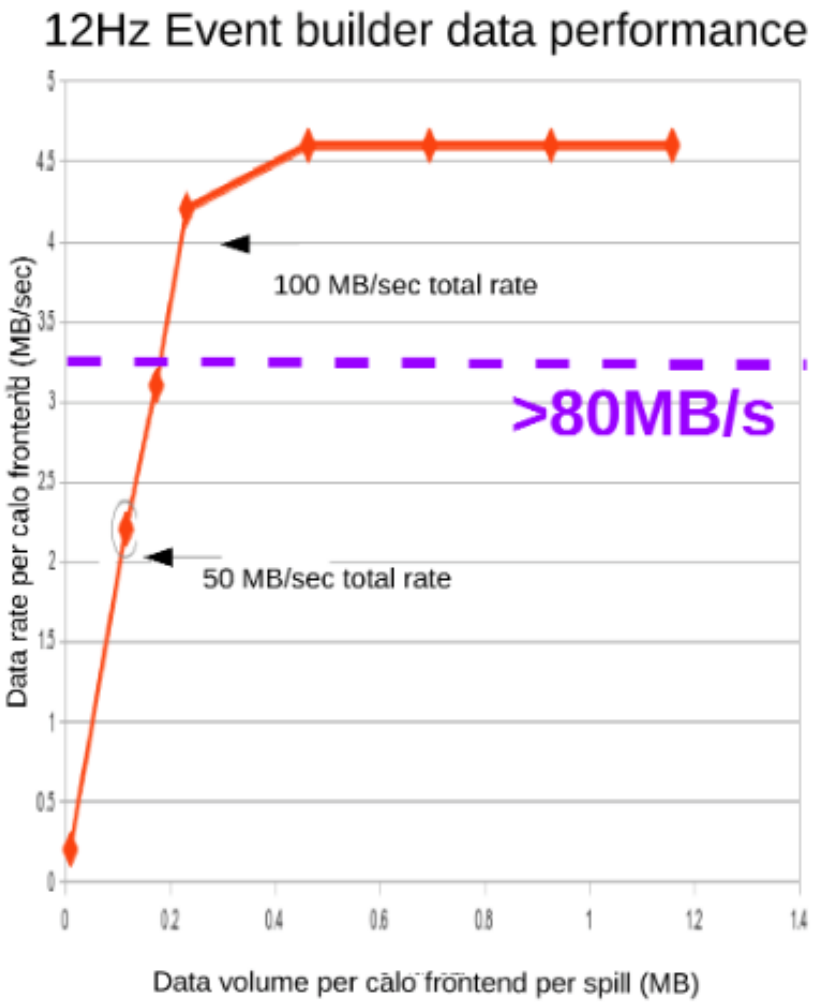
20:07:17 [mtransition,INFO] Run #2071 started

---------- Event# 1 ----------

Evid:0001- Mask:0000- Serial:0- Time:0x512b8c40- Dsize:2881352/0x2bf748

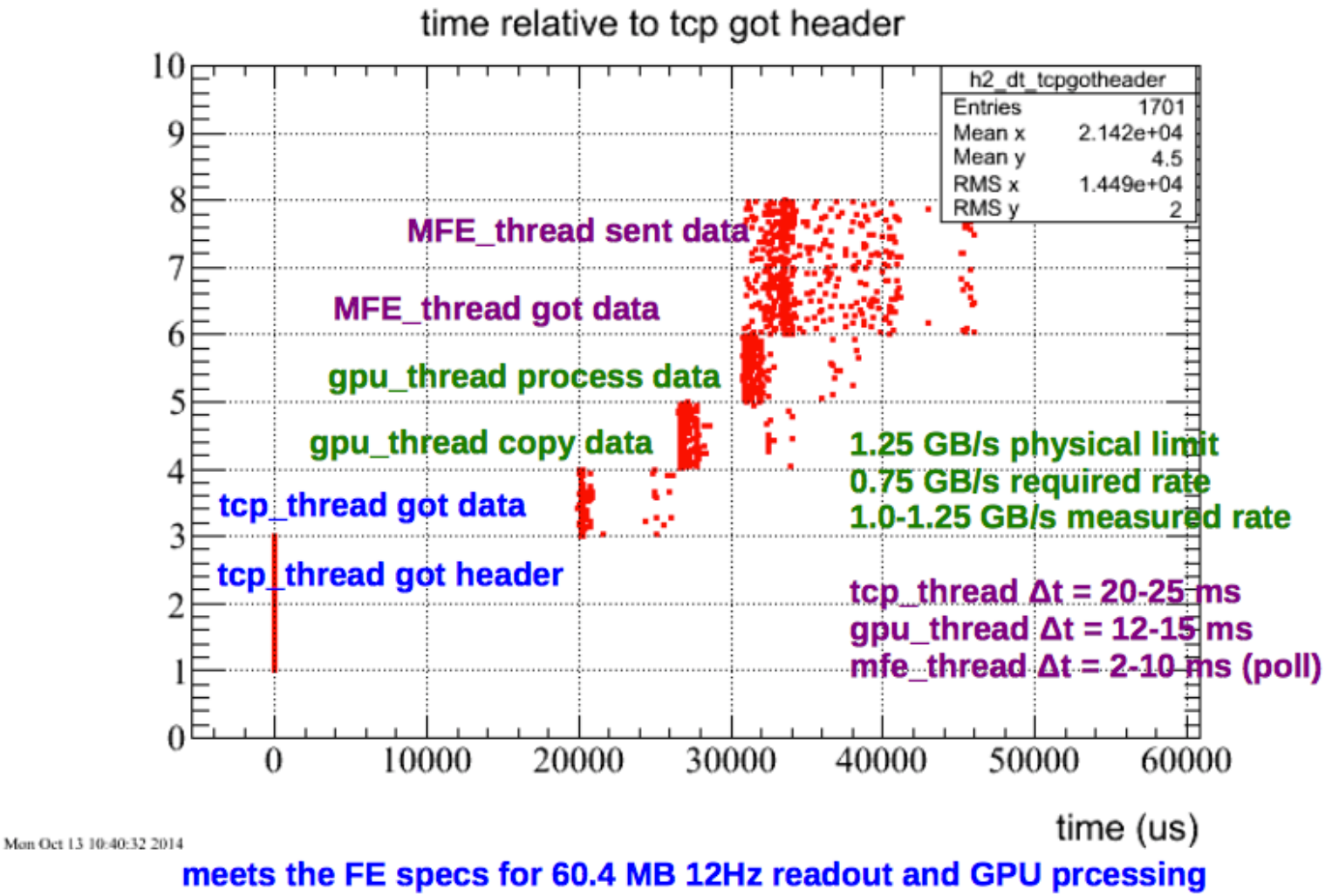
#banks:48 - Bank list:-FD01SR01FD02SR02FD03SR03FD04SR04FD05SR05FD06SR06FD07SR07FD08SR08FD09SR09FD10SR10FD11S

R11FD12SR12FD13SR13FD14SR14FD15SR15FD16SR16FD17SR17FD18SR18FD19SR19FD20SR20FD21SR21FD22SR22FD23SR23FD24SR24-



GPU Processing Time

Time is dominated by memcpy to GPU.



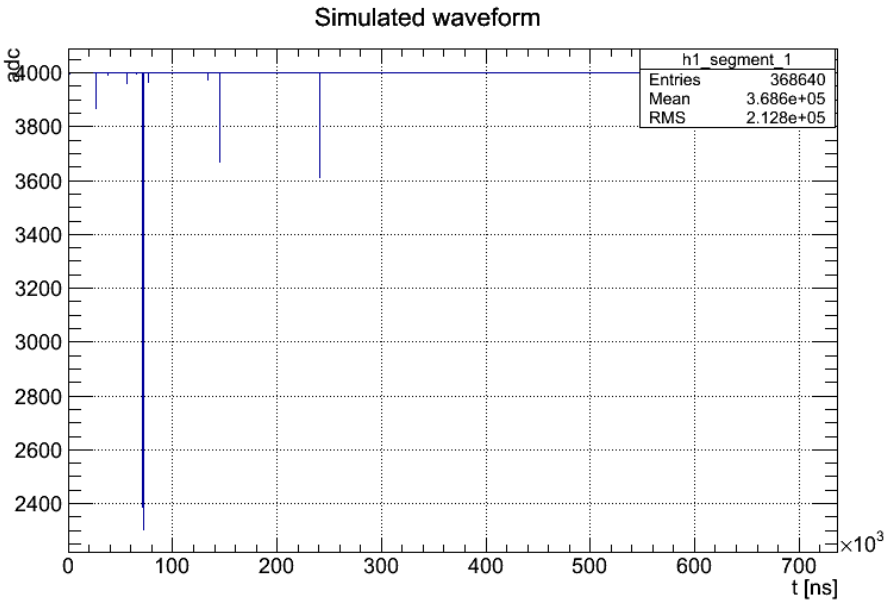
meets the FE specs for 60.4 MB 12Hz readout and GPU prcessing

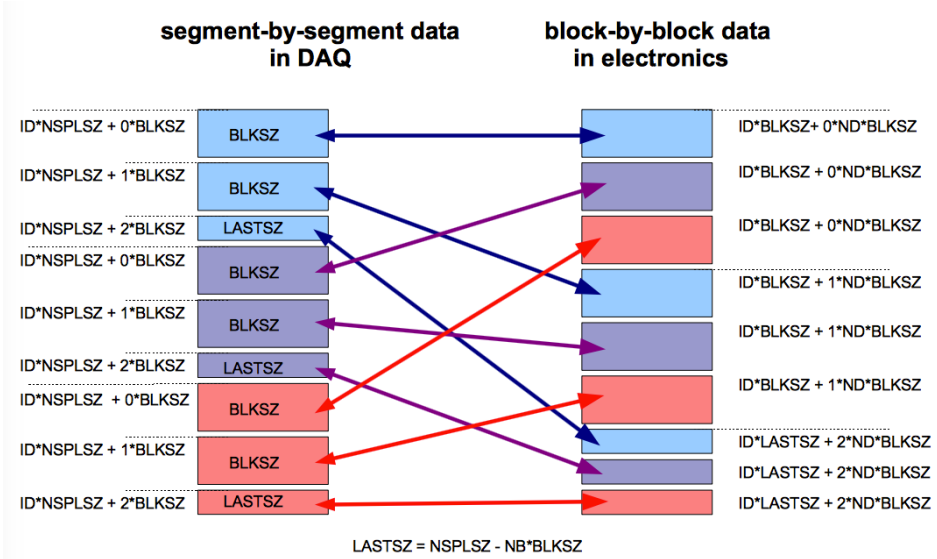
Test Stands

- DAQ was tested using test stands operating in parallel at Fermilab and U. of Kentucky
- Currently includes backend, frontend, gateway, and μTCA crate with WFD and AMC13

AMC13 Simulator

- Generates realistic waveforms and packs the data in the AMC13 data format.
- Allows us to exersize the DAQ without dependence on hardware.
- Plan to develop this into a tool that will recreate the full spectrum of DAQ input, which will be used for testing the complete data acquisition system.





Data unpacking in GPU