Software and Computing R&D

Adam Lyon (Associate Division Head of Systems for Scientific Applications)

Inaugural Meeting of the ICAC

2019-03-14

Software & Computing Research and Development

Triggers (the why):

A. Requirements from experiments based on upcoming needs
B. Forward thinking to keep up with evolving computing landscape
C. Useful technologies that scientists adopt and needs support
D. Fruitful collaborations

Drivers (the what):

A. CMS in the HL-LHC era and DUNE
B. New computing architectures/accelerators and the Exascale High Performance Computing Era
C. Machine Intelligence's impact on HEP reconstruction and analysis
D. Specific funding calls (e.g. SciDAC from DOE-ASCR)

Guides (the how):

- Physics goals (of experiments and scientists)
- Software and Computing requirements from CMS and DUNE
- Community White Papers (HEP Software Foundation and IRIS-HEP)
- Goals of SciDAC and ECP
- Strive for common tools where possible and common principles for moving forward

There is overlap, of course

R&D Activities Overview - A broad program

- Physics and detector simulations with advanced architectures and techniques
- Accelerator Modeling on HPC
- Evolution of Infrastructure Frameworks (CMS, DUNE) and Root
- HPC, Advanced architectures/accelerators, multithreading
  - Containerization
  - HEP Data Analytics
  - Reconstruction
  - Spack & SpackDev [HPC compatible packaging]
- Machine Intelligence
- Data Acquisition
- Advanced networking (BigData Express)
- Workflow (HEPCloud)
- Astro (CCD/MKIDs)
- QIS now has its own program and I won't discuss, but some personnel comes from SCD (myself included)

Funding comes from many sources

- DOE-OHEP (CompHEP)
- USCMS Software and Computing (S&C) Operations Program
- SciDAC-4 [DOE-ASCR] $17.5M awarded total
  - 5 yr and 3 yr projects started in FY18
- Fermilab LDRD (Lab Directed R&D)
- Exascale Computing Project (ECP)
- HEP-CCE (Center for Computational Excellence)
  - Promote excellence in HPC and R&D
  - Enhance connection to ASCR
  - FNAL, ANL, BNL, LBNL
- Other experiment projects & Detector R&D (KA25)
  - e.g. CMS Outer Tracker, Mu2e TDAQ
- We supplement with SCD funds

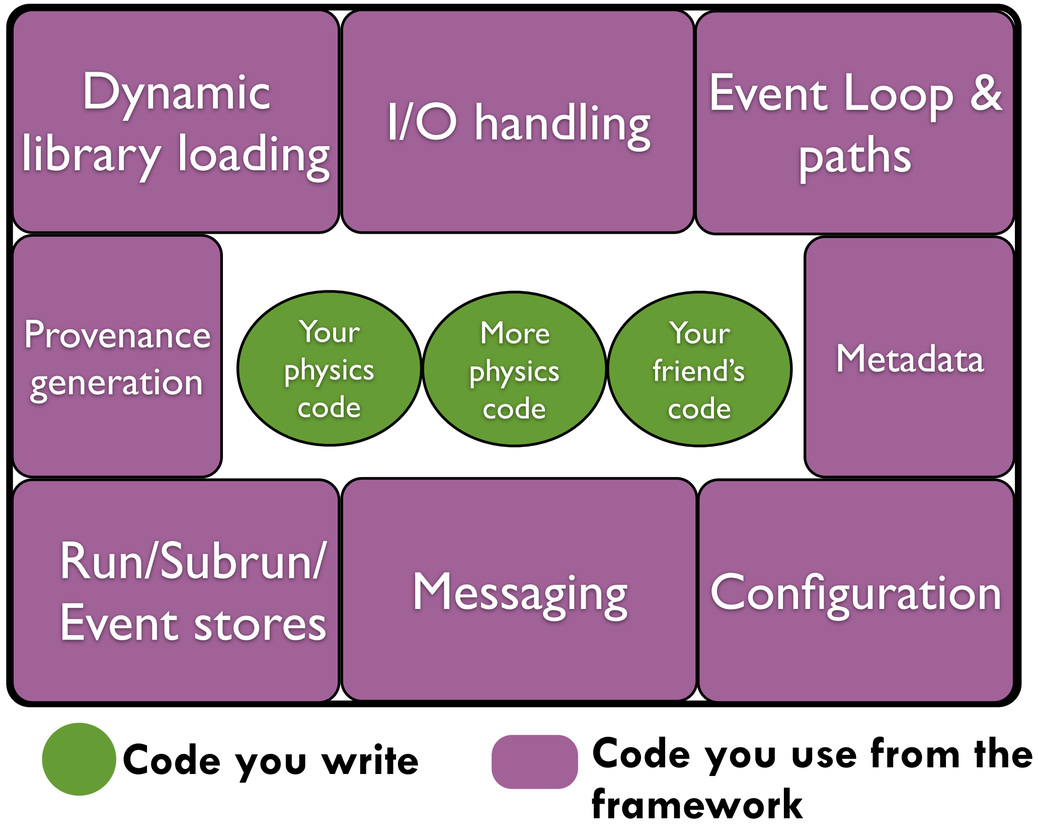
Personnel may be matrixed across projects

Physics and Detector Simulation

- *Generators and Geant*
  - Pythia
    - High energy collision generator
    - Steve Mrenna [SCD Scientist] is a main author
    - Event generator tuning on massive scale on HPC is part of SciDAC (see later)
  - Genie
    - Main Neutrino MC generator
    - Team adapts for Fermilab neutrino experiments
  - GeantV
    - Collaboration with CERN and others
    - Geant4 is the ubiquitous detector simulation toolkit…
    - GeantV is a re-architecture for GPUs, Vectorization, and Exascale
    - CMS is using alpha release
    - Beta release with ~x2 speed up is coming

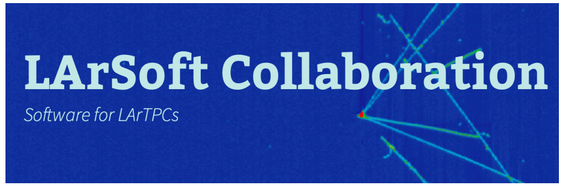
Infrastructure Frameworks (USCMS S&C & CompHEP)

- Benefits from Computing Professionals
- Enables advanced computing
- Important relationship between framework developers and experiment scientists
- ***CMSSW***
  - Multithreading pioneer and leader (in production)
  - Extensive project to upgrade algorithms done
  - Framework developers embedded in leading CMS software program
- ***art***
  - Fork of and diverged somewhat from CMSSW for muon and neutrino experiments
  - Special features for “non-collider physics” (e.g. redefinition of “event” for DUNE)
  - Driven by consensus of experiment stakeholders (no “special” versions for particular experiments - developers are not on experiments)



- Recently multithreaded capable (multiple events in flight)
- Shifting developers to LArSoft (next slide) [future art development only if necessary]

LArSoft



- LArTPC Toolkit atop *art* for DUNE (including protoDUNE), MicroBooNE, LArIAT, SBND, ICARUS
- Driven by steering committee with reps from SCD and experiment management
- Fermilab writes infrastructure (e.g. common data products, modules, and services, Geant4 interface)
- Experiments write algorithms
- Interfaces to external packages like WireCell (BNL) and Pandora

- Fermilab helping to make toolkit and algorithms multithreaded
  - Investigating advanced strategies like Kukkos and Raja, OpenMP SIMD, and OpenMP GPU offloading

- Event display needs work - engage collaborators

Infrastructure Framework R&D & Root

Moving frameworks ahead for the future…

- SCD working with experiments and stakeholders to agree on a unified framework for DUNE and HL-LHC to enable physics and analysis on a massive scale
- We welcome expanding stakeholders and developers beyond CMSSW/art
- Take advantage of future computing heterogeneity
- Take advantage of future I/O technology (e.g. object stores)

Root…

- Cross cutting application ubiquitous in HEP
- Hooks into current frameworks (especially C++ serialization and I/O)
- We have leadership in Root I/O, but need more effort for this important tool

Data Acquisition R&D

We develop(ed) DAQ for NOvA, MicroBooNE, single phase protoDune, SBND, mu2e, and member Dune DAQ consortium

artdaq – A Common DAQ toolkit atop *art*

- Front end adapters, routers, event builder, trigger modules, …
- Writes out same data format as *art* offline (with Root i/o) - significant advantages here and opportunity for common downstream tools
- Compatible with MPI style multiprocessing (though we've never exercised that feature)
- Significant development for protoDune, SBND, and mu2e

OTSDaq – An "off the shelf" DAQ system

- An end-to-end DAQ system based on a menu of hardware options (select by needs) and online & firmware libraries
- Initiated by a three year Fermilab LDRD
- Uses artdaq toolkit as well as CMS XDaq
- Used by CMS upgrade projects, test stands (e.g. LCLS II, CCD readout), and test beam experiments (on path to be an offering by Fermilab Test Beam Facility)
- Mu2e recently decided to use OTSDaq interfaces and run control system

Machine Intelligence R&D

- Recently formed *Machine Intelligence and Reconstruction* group to emphasize our expertise and work in this area

- Strong programs in adapting Machine Intelligence technology into Neutrino physics, CMS analyses and reconstruction, Cosmology and using advanced architectures such as FPGAs and GPUs

- Current LDRD: “Modeling Physical Systems with Deep Learning Algorithms”
  Extract cosmological parameters from large datasets with Deep Learning

- Past LDRD: “High Energy Physics Pattern Recognition with an Automata Processor”
  First use of automata processor for tracking

- Starting involvement in Quantum ML

USCMS Software and Computing R&D

- USCMS and international CMS are making good progress in defining and executing a comprehensive R&D program for the HL-LHC era.
- Many areas and directions are part of the SCD portfolio and executed by or together with experts from SCD

  For example:
  - Address Heterogeneity challenge (be in a position to use any processor/accelerator made available)
    - Strategy is based on multi-threaded CMSSW, vectorized GeantV, PileUp pre-mixing, vectorized and re-designed reconstruction algorithms for advanced architectures
    - Foundation has been laid, future efforts needed in physics algorithm development - important to pair domain detector experts with core computing experts from HTC and HPC world

  …continued…

USCMS Software and Computing R&D (continued)

- Data Organization, Management and Access (DOMA)
  - Storage is cost driver for HL-LHC
  - CMS already demonstrated excellent data discipline through small and streamlined analysis data formats that are shared by the whole collaboration (single analysis working set)
  - Many R&D directions to control storage needs - Networking, Data Federations, Storage Technologies, Lossy Compression - Moving to Rucio by end of 2020, NANOAOD is being established as the newest smallest analysis data format
- Analysis
  - Novel strategies to optimize time-to-insight for very large analysis datasets - R&D in array programming
  - Delivery frameworks being investigated, for example Apache Spark, Striped LDRD

FNAL SCD the most important R&D partner on DOE side for USCMS, additional partners are IRIS-HEP (NSF), NESAP (co-development with NERSC for Perlmutter), Universities ➜ embedded in HSF and WLCG activities

Past LDRDs

- *Preparing HEP reconstruction and analysis software for exascale era computing*
  - Partnership with HDF5 Group
  - Starting point for component of a SciDAC project

- *Striped Data Server for Scalable Parallel Data Analysis*
  - Prototype No-SQL database server system for parallel data analysis
  - Cluster out of old hardware
  - Currently tested by multiple CMS analyses (dark matter search, Higgs measurements) and by DES for catalog processing
  - Using Jupyter as a user-facing interface

Workflow R&D

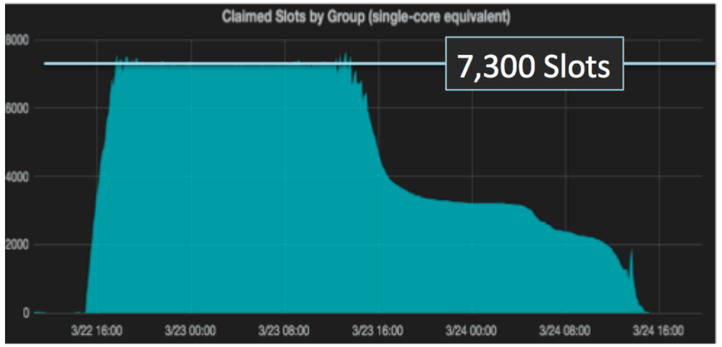
- Containerization (need to unify several efforts across the division)

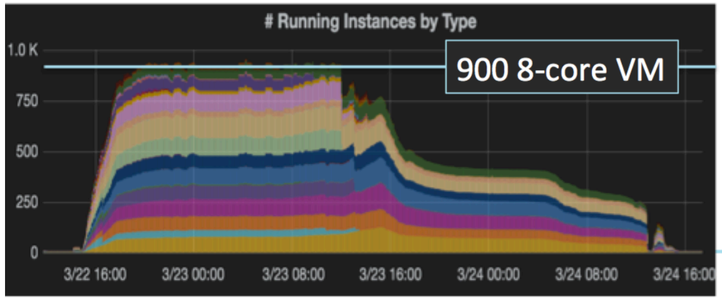
- Adoption of Rucio for the Fermilab Data Center
  - Rucio is an open-source project for managing community data developed by the ATLAS collaboration
  - Unify the CMS and Non-CMS Data management systems

- HEPCloud: An Elastic Hybrid HEP Facility using an Intelligent Decision Support System
  - Extend computing facility to provide access to disparate resources including commercial and community clouds, grid federations, local resources, and HPC centers
  - Novel Decision Engine (DE)
    - Makes decisions that aid in automatic provisioning of resources (heart of R&D)
  - BNL is contributing effort and will help with packaging HEPCloud as a product
  - Analytics from job failures is potential R&D topic
  - Strong endorsement from ASCR leadership
  - Go-live occurred this past Tuesday!!

HEPCloud

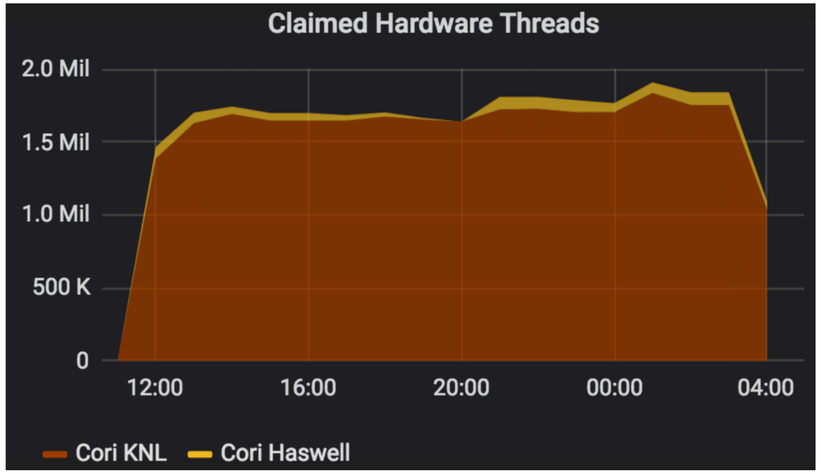
NOvA running on AWS (2016)



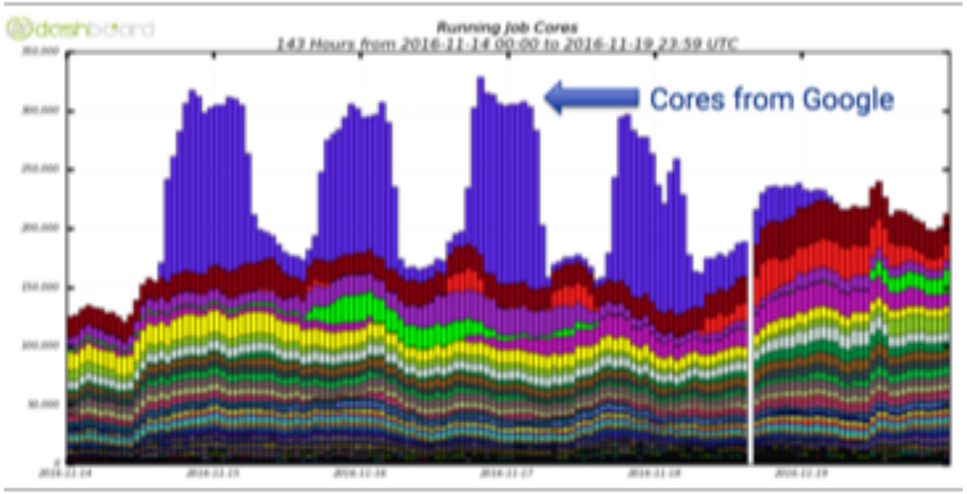


- Run ~7,300 jobs on 900 VMs for 21h out of 48 h for 203k h
- 10,000 jobs with 5 files per job
- ~6 types of 8-cores VMs in 8 overall Availability Zones in 3 AWS Regions

NOvA running at NERSC (2018)



CMS on Google Cloud (2016)



Google Cloud, HEPCloud and probing the nature of Nature

Monday, November 14, 2016

Posted by Paul Rossman, Google Cloud Developer Relations

Understanding the nature of the universe isn't a game for the resource-constrained. Today, we probe the very structure of matter using multi-billion dollar experimental machinery, hundreds of thousands of computing cores and exabytes of data storage. Together, the European Center for Nuclear Research (CERN) and partners such as Fermilab built the Large Hadron Collider (LHC), the world's largest particle collider, to recreate and observe the first moments of the universe.

Today, we're excited to announce that Google Cloud Platform (GCP) is now a supported provider for HEPCloud, a project launched in June 2015 by Fermilab's Scientific Computing Division to develop a virtual facility providing a common interface to local clusters, grids, high-performance computers and community and commercial clouds.

Fermilab

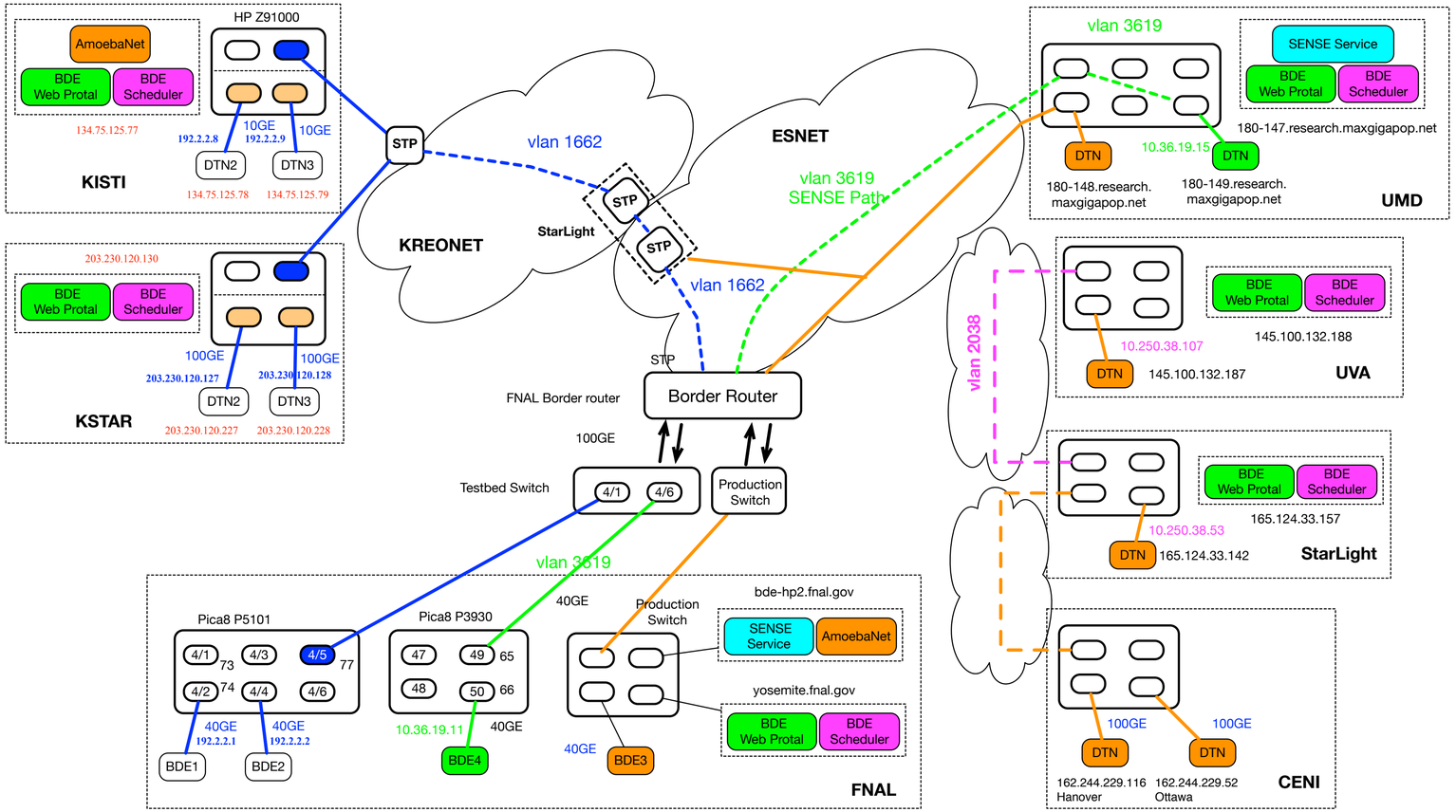
Networking for Science R&D - BigData Express

- Predictable, Schedulable, and High Performance Data Transfer

  - Support from DOE ASCR Network Research Program
  - A peer-to-peer, scalable, and extensible data transfer model
  - A visually appealing, easy-to-use web portal
  - A high-performance data transfer engine
  - A time-constraint-based scheduler
  - On-demand provisioning of end-to-end network paths with guaranteed QoS
  - Robust and flexible error handling
  - CILogon-based security

  - An improved version of Globus Online

BigData Express Supercomputing 2018 Demo



SciDAC-4 Programs

Scientific Discovery through Advanced Computing

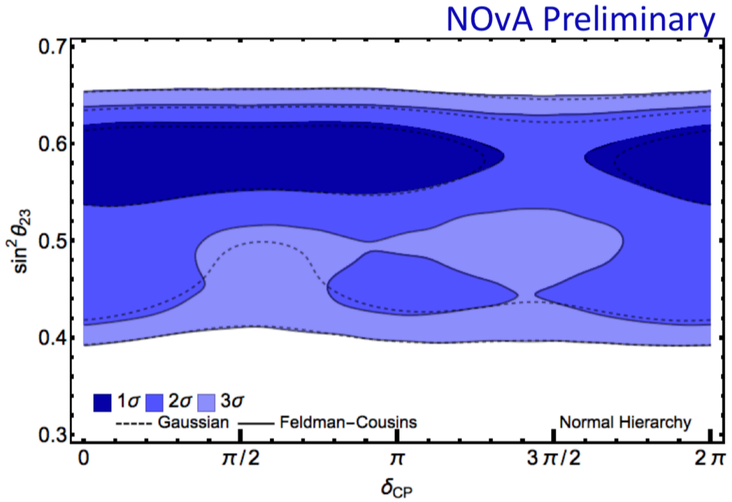
- Funded by DOE OHEP and Advanced Scientific Computing Research [ASCR] Program
- We have 3/5 HEP SciDAC programs and we participate in a 4th led by Fermilab Theory
- Have access to and take advantage of deep computing, HPC, and applied mathematics expertise at ASCR centers and institutes
- Joining this community and creating a presence in ASCR has been a long-term goal for us … successful with these programs…

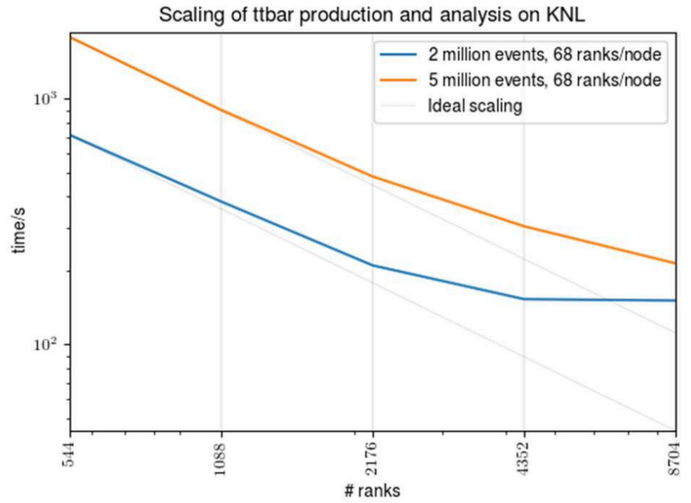
HEP Data Analytics on HPC

Fermilab Argonne NATIONAL LABORATORY Lawrence Berkeley National Laboratory University of CINCINNATI Colorado State University

Objective: Advance LHC and neutrino science by transforming data analysis applications, workflows, and data handling to effectively utilize resources available at HPC facilities

- 5 years starting in FY18
  - NOvA Neutrino/Antineutrino Analysis @ NERSC
    Most precise measurement of antineutrino oscillations 8x higher resolution; 50x faster to result than previous; billions of simultaneous multi-dimensional fits; Analysis with HDF5 I/O
  - HEPnOS: Fast event-store for HEP on HPC
    Object store distributed storage system
  - Automated & massively parallel event generation, analysis and tuning





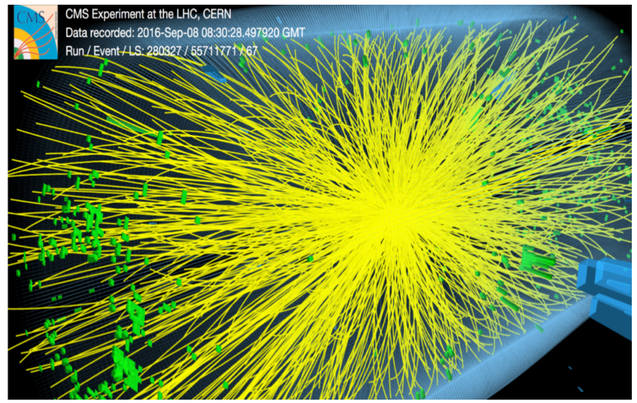
Fermilab

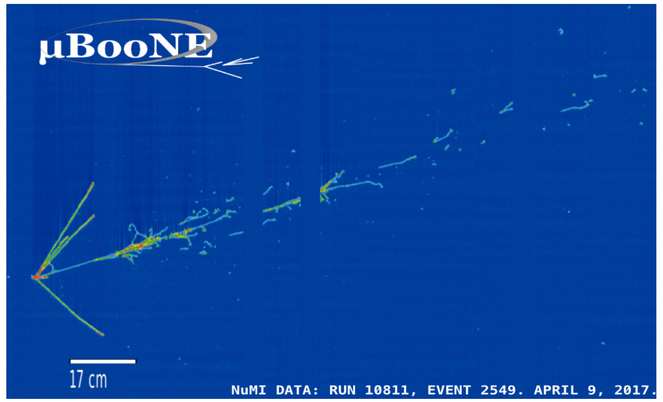
HEP Event Reconstruction with Cutting Edge Computing Architectures

Fermilab | UNIVERSITY OF OREGON

Objective: Accelerate HEP event reconstruction exploiting highly parallel computing architectures, focusing on CMS tracking and LArTPC Reconstruction

- 3 years starting in FY18
  - CMS Tracking prototype algorithm
    SIMD library and threading with TBB (simultaneous processing of multiple collision events),
    10x faster with one thread, 600x faster with >100 KNL threads without loss of physics
  - LArTPC Hit finding algorithm
    Replace MINUIT+Root Gaussian fit with local implementation of minimization,
    8x faster single threaded (no loss of physics), further 11x speedup with 20 KNL threads

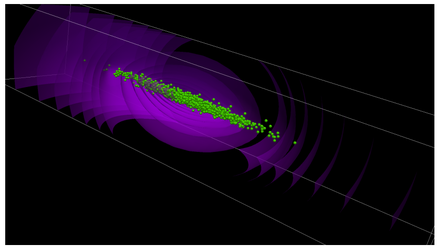




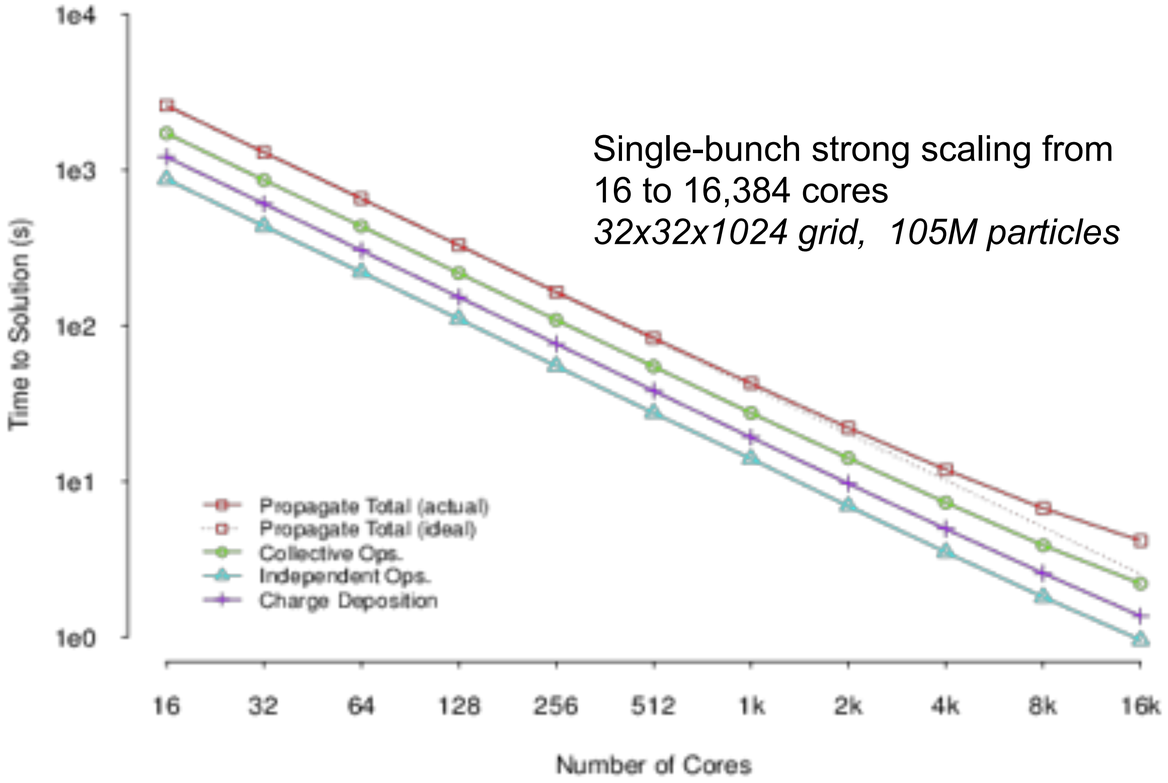
Accelerator Modeling

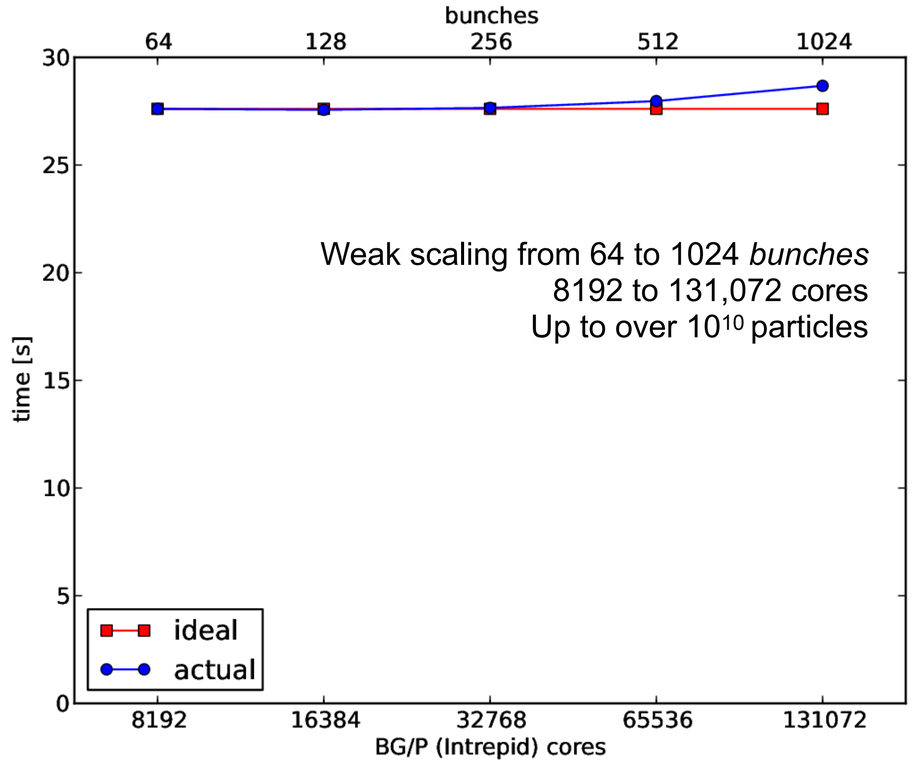
ComPASS4
High Performance Computing for Accelerator Design and Optimization

Synergia beam dynamics framework



- 5 year project started in FY18
- Collaboration with Fermilab, Argonne, LBL, and UCLA
- Optimizing for HPC (GPU, KNL)
- A native supercomputer application





Fermilab

… and there's more!

- Big data analyses with Apache Spark using CMS & Neutrino data

- Application and library packaging and development environment management with Spack/SpackDev
  - Spack popular at HPC centers
  - We want to replace a "vintage" Fermilab built and maintained system (UPS)
  - SpackDev is a development environment management system using Spack

- CCD readout/DAQ and MKIDS R&D
- CMS Fast timing

- Cosmology/LSST analysis (ComsoSIS)

Summary and Challenges

Fermilab SCD is engaged in a very broad R&D program to meet the needs of experiments and the opportunities of advanced diverse computing in the future

Challenges:

- Our personnel are spread thin and the funding is complicated
  - Keeping coherency is non-trivial. Many funding sources creates a tangled web
- Difficult to make room for new opportunities and funding calls
- We have successfully handed off R&D from project to project, but not easy
  - e.g. HDF5 LDRD to SciDAC
  - e.g. OTSDaq
  - We will need to integrate the SciDAC program results into future R&D
- We need to build on our success with follow-on R&D funding

BACKUP BACKUP BACKUP BACKUP BACKUP BACKUP

Scientific Computing Thrusts

Facility (Stu's talk)

Scientific Operations and Workflows (Liz's operations talk)

Development, Integration and Research (this talk)

- Provide **common** solutions to experiments in the areas of…
  - Data Acquisition
  - Simulation Tools (Generators and Geant)
  - Software Infrastructure (Frameworks and Toolkits)
- Areas of expertise
  - Large scale programming
  - C++
  - Physics/Detector simulations
  - DAQ engineering
  - Algorithms including Machine Intelligence

Hit Finding in LArTPC

Scientific Achievement

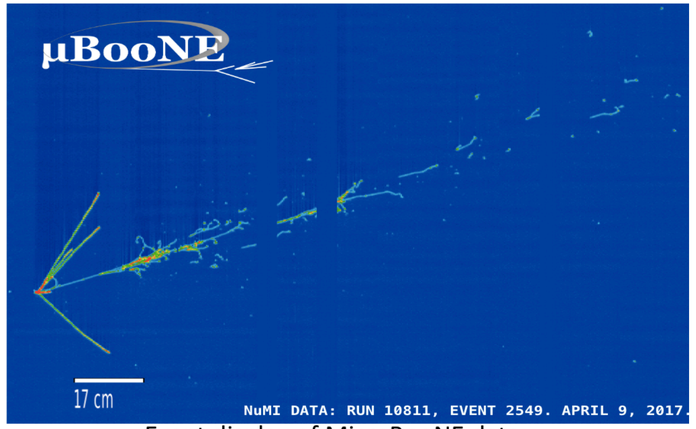
Identified hit finding algorithm as feasibility study in the area of LArTPC reconstruction and produced standalone implementation to test speedups from SIMD and thread parallelism

Significance and Impact

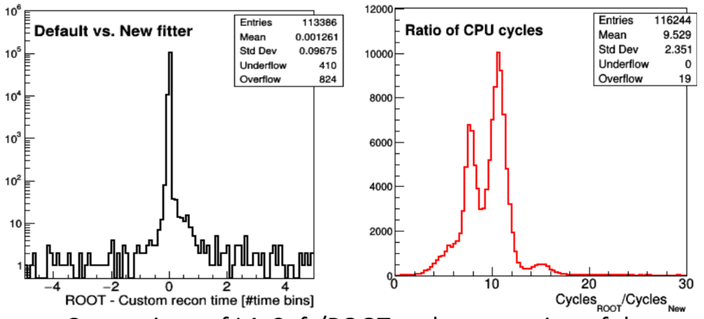
Reconstruction in LArTPC neutrino experiments is challenging, takes O(100) s/event in MicroBooNE, so speedups are critical for future detectors (DUNE)

Research Details

- Signal from charged particles produces Gaussian pulses, hit finding is the process of identifying such pulses and determining their properties (peak position and width)
- Currently a non negligible contributor to timing, takes ~15% of the MicroBooNE reconstruction workflow
- Suitable for parallel processing: wires are independent
- Standalone implementation where Minuit+ROOT-based Gaussian fit is replaced with a local implementation of Levenberg-Marquardt minimization
- Comparing with the LArSoft version of the algorithm, results nearly identical in terms of hit properties but ~8x faster already without parallel processing techniques



Event display of MicroBooNE data



Comparison of LArSoft/ROOT and new version of the algorithm: output (left) and CPU cicles (right)

Fermilab