Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

P5 Science Drivers: Accelerator Experiments

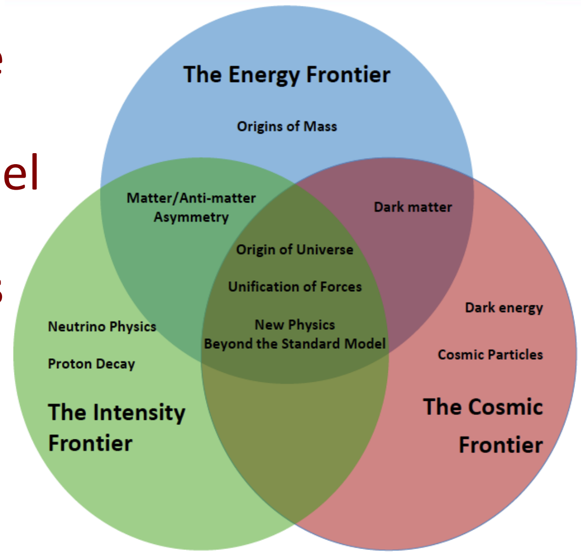
Panagiotis Spentzouris

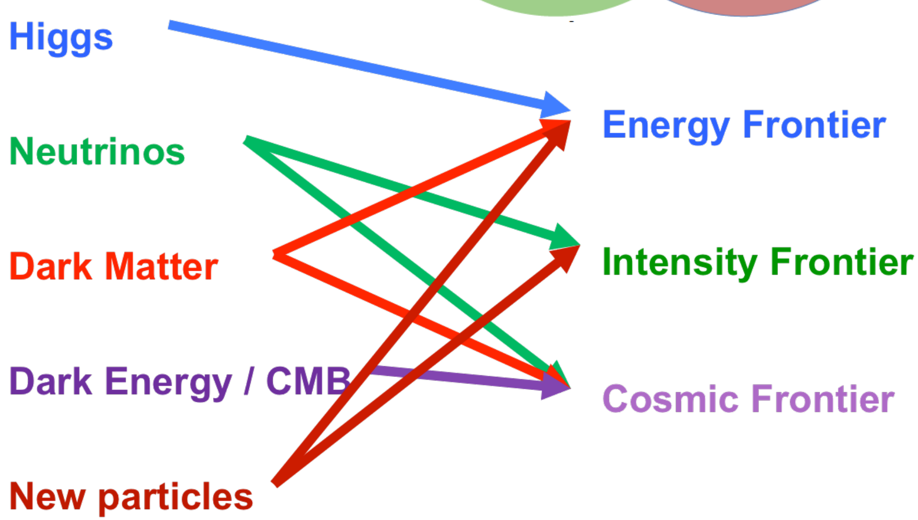
DOE SC Exascale Requirements Reviews: High Energy Physics

June 10th, 2015

High Energy Physics Science Drivers

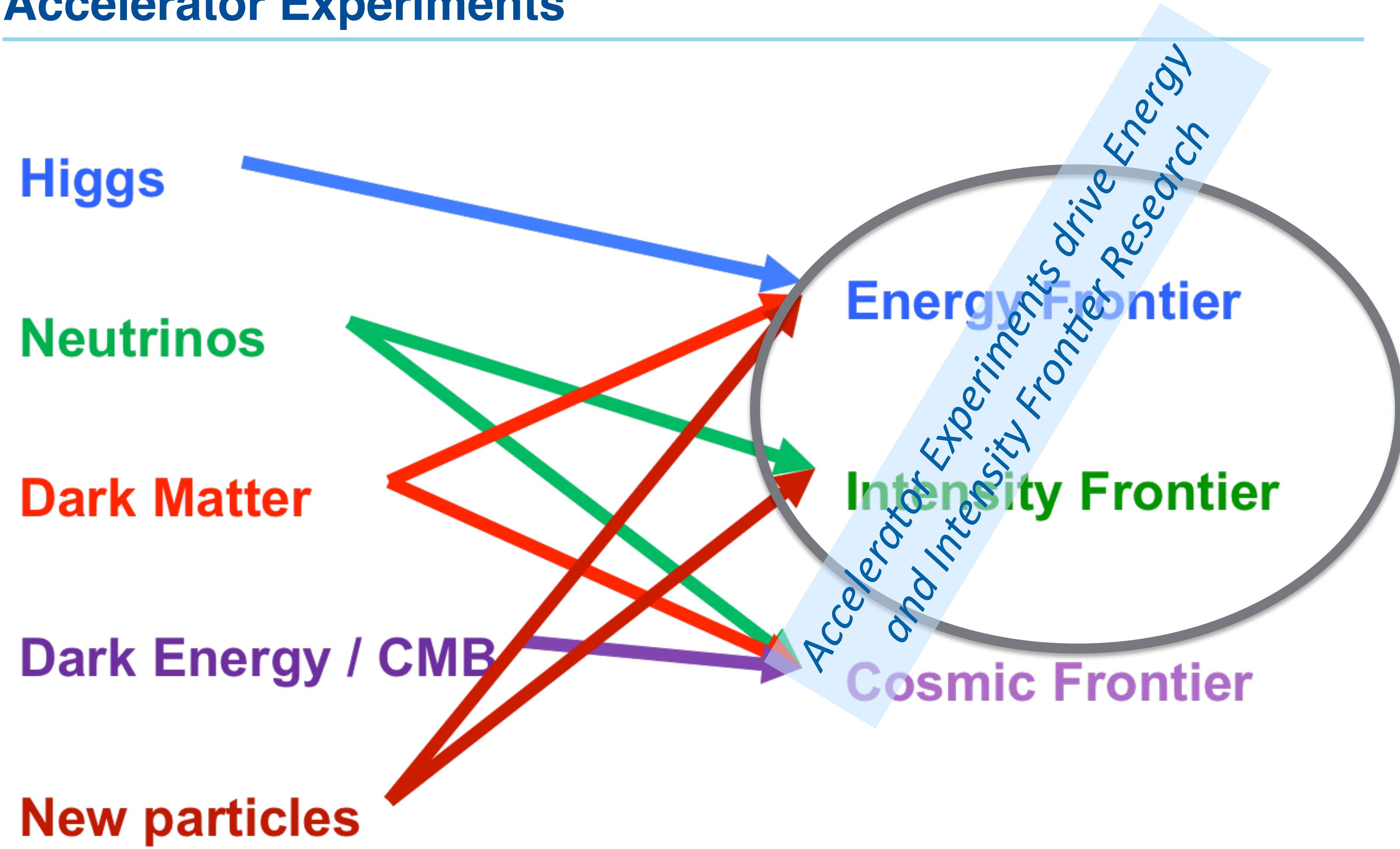
The 2014 Particle Physics Project Prioritization Panel report identified **5 Science Drivers Mapped to 3 Frontiers**





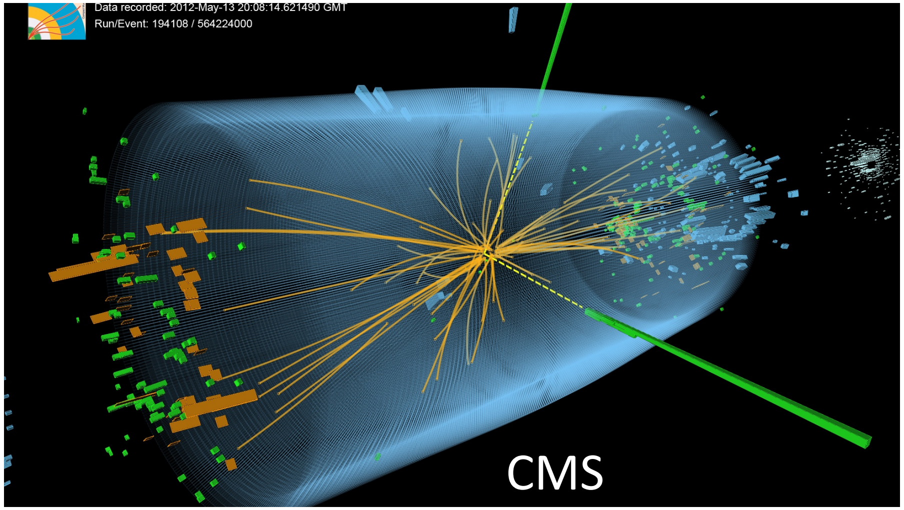
- At the **Energy Frontier**, high-energy particle beam collisions seek to uncover new phenomena
  - the origin of mass, the nature of dark matter, extra dimensions of space.
- At the **Intensity Frontier**, high-flux particle beams enable exploration of
  - **neutrino interactions**, to answer questions about the origins of the universe, matter-antimatter asymmetry, force unification.
  - **rare processes**, to open a doorway to realms to ultra-high energies, close to the unification scale
- At the **cosmic frontier** we seek to understand the nature of the contents of the universe: ordinary matter, dark mater and dark energy.

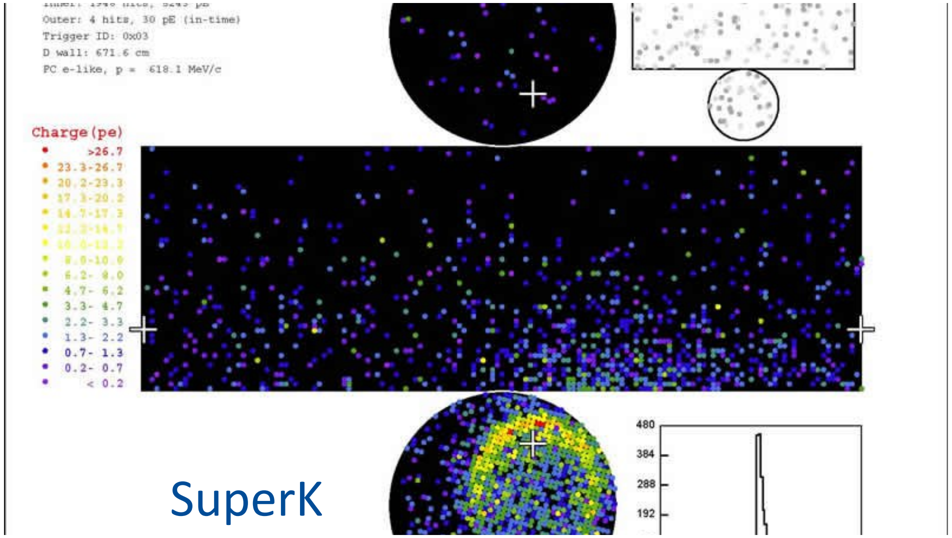
Accelerator Experiments



Where we are today

- **Discovery of the Higgs particle at the CERN LHC, responsible for electroweak symmetry breaking and the mass of elementary particles**
  - No physics beyond the “Standard Model” of HEP has been observed
- **Neutrinos oscillate, thus have mass**
  - No answers on mass hierarchy or symmetry properties
    - Potential explanation for matter anti-matter asymmetry observed in the universe!

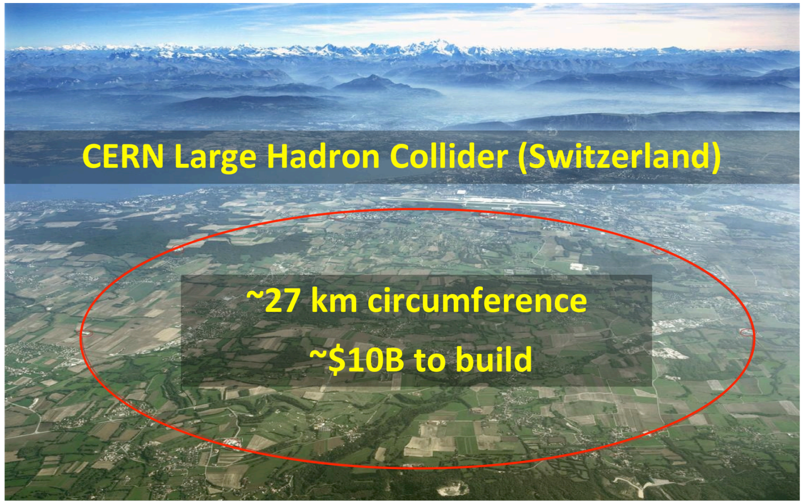




Fermilab

Next Steps (Energy Frontier)

- More **powerful** detectors & accelerators to facilitate discovery
  - **CMS & ATLAS** detectors at the **LHC** with higher energy and luminosity (major accelerator and detector upgrades)
  - a **new larger** hadron collider or a **dedicated** accelerator to study **Higgs properties**
- The US is playing a leading role in LHC upgrades and participates in designs for future machines
- **Computing evolution** essential for success
  - higher data rates, higher pile-up, will require utilization of new techniques and technologies
  - More powerful beams and complex detectors need higher fidelity numerical tools to design and optimize, to ensure program success!

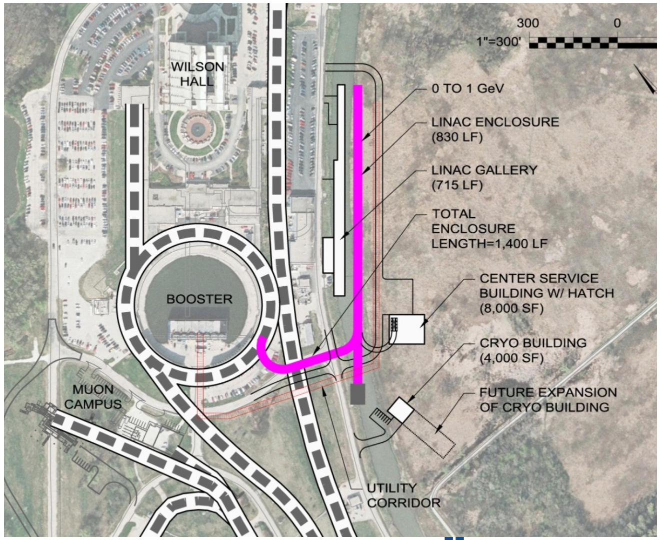


Courtesy S. Myers (IPAC 2012)


CMS detector

Next Steps (Intensity Frontier)

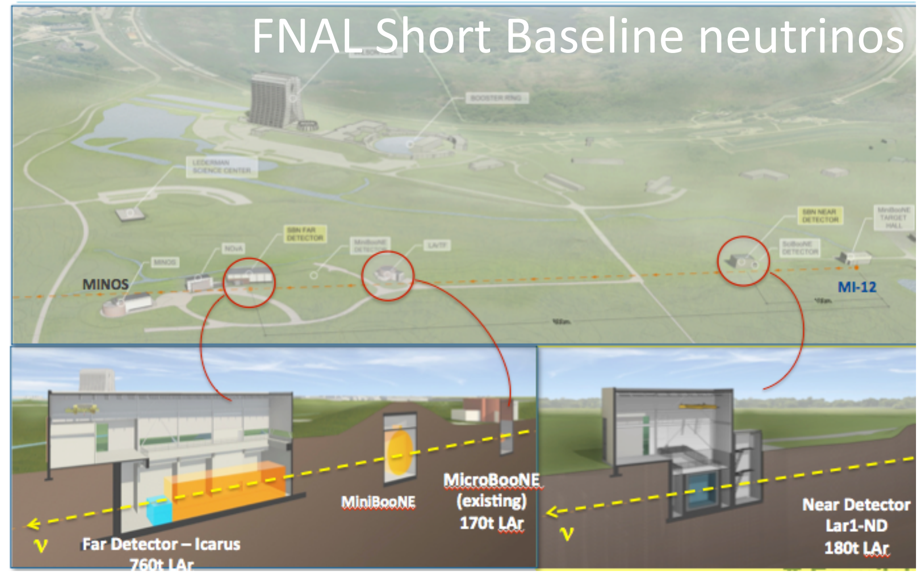
- A **high-intensity proton** accelerator for
  - **neutrino oscillation** experiments
    - Mass hierarchy, matter-antimatter asymmetry
  - rare process experiments
    - New particles and interactions
- Staged approach at Fermilab: major complex improvements support
  - Short and long baseline neutrino program
  - Rare decay program
- BELLE2 at KEK (Japan), LHCb (CERN): heavy quark and tau lepton decays
- **Computing evolution** essential for success
  - Many experiments with different timelines, need a fully supported computing ecosystem for data analysis
  - More intense beams and large, complex detectors need higher fidelity numerical tools to design and optimize, to ensure program success!

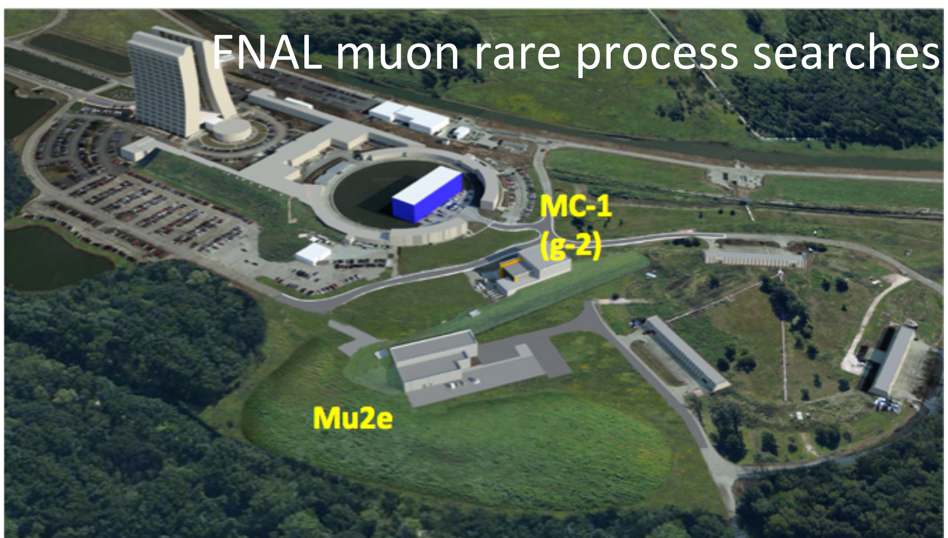


Fermilab

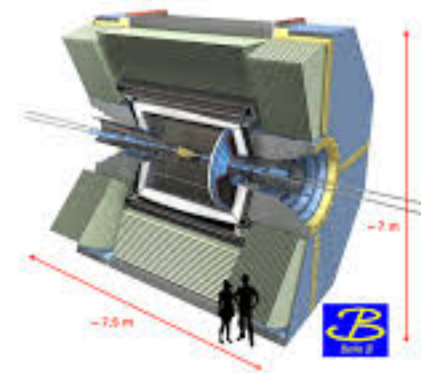
Next Steps (Intensity Frontier)

Many projects, different timelines,





KEK heavy flavor, tau new physics searches



2015-2020 (large, mid-size programs)

FNAL MI-LB neutrinos (IF)

- MINOS+, MINERvA, NOvA

FNAL Booster-SB neutrinos

- uBooNE, SBND, ICARUS (IF)

FNAL Recycler-muons (IF)

- g-2

KEK-heavy flavors (IF)

- Belle II

LHC beams: Run 2

- ATLAS, CMS (EF)
- LHCb (IF)

2020-2025

…

Recycler-muons (IF)

- Mu2e

LHC Run 3 (phase 1 upgrade)

…

2025-…

Long Baseline Neutrino Facility (LBNF)

- **DUNE (IF)**

LHC Run3 (HL-LHC)

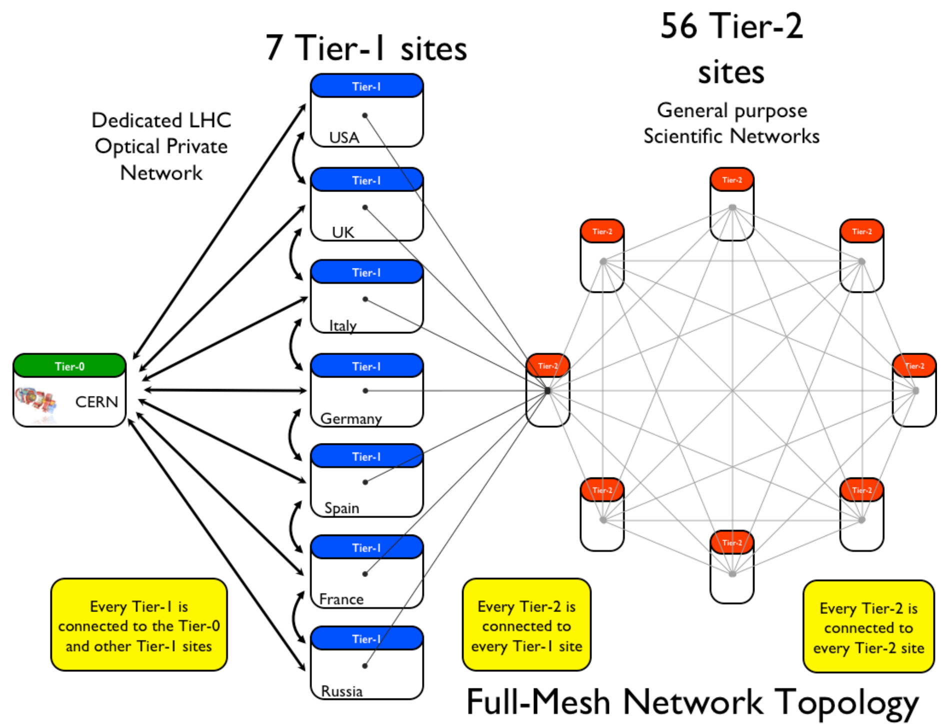
…

Computing paradigm

- A tiered architecture, with different reliability and availability requirements per tier, functioning as a coherent system
  - **Tier-0:** acquires, processes, archives, distributes raw data
  - **Tier-1:** receives subset of raw data & archives, provides compute resources for reconstruction & other processing, distributes data to Tier-2, receives and archives MC from Tier-2; user analysis
  - **Tier-2:** compute resources for user analysis, MC production; reconstruction and other processing
- Relies on GRID middleware and infrastructure, commodity compute resources, networking crucial for data and compute intensive workflows
- Resources either owned by the experiments (LHC) or deployed to cover the needs of specific program (e.g. IF experiments at FNAL)
  - Resource deployment objective to fully cover needs, peak demand offloading mostly within the HEP grid site ecosystem
  - Recently, some utilization of ASCR computing resources, for specific applications and with "special" workflow and workflow management

Computing paradigm

- Model evolved from the LHC computing design (CMS description at http://cds.cern.ch/record/814248/files/note04_031.pdf)
  - programmatically supported, formally defined (MOUs)
- IF experiments follow variants
  - E.g. the Fermilab Facility provides both Tier-0 and Tier-1 function to some experiments, Tier-1 to most
    - using Fermilab provisioned and hosted resources, or remote resources through the Open Science Grid (OSG).
  - E.g. Belle2 has a Tier-0 at KEK and a Tier-1 at PNNL (see "Computing Requirements Report" at http://www.slac.stanford.edu/econf/C1307292/docs/submittedArxivFiles/1308.0672.pdf)

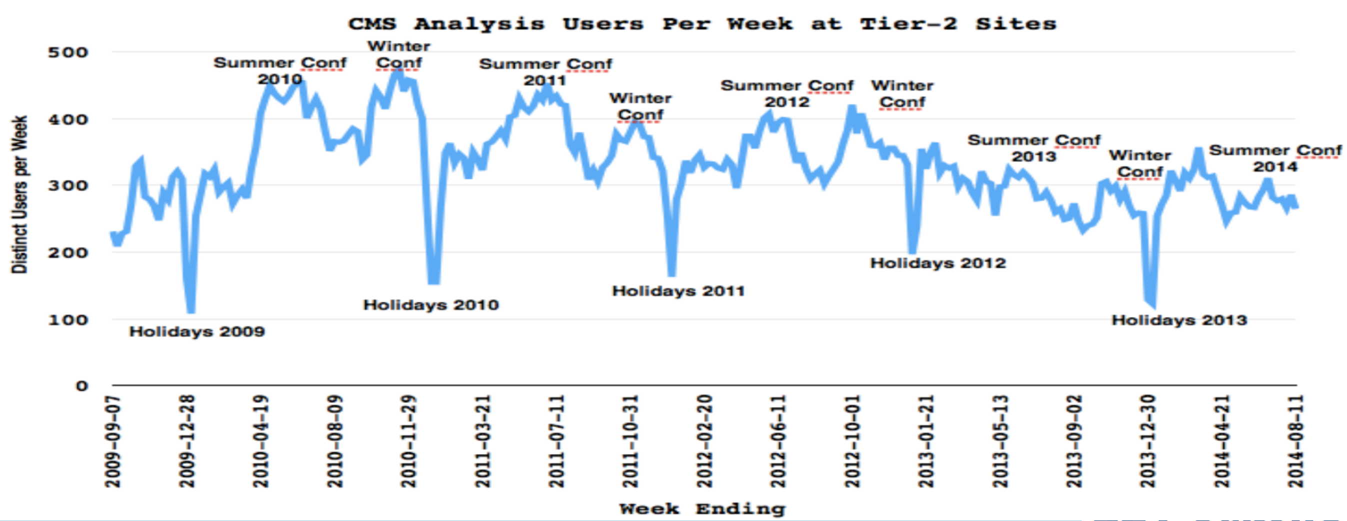


CMS grid infrastructure

Current size of computing infrastructure

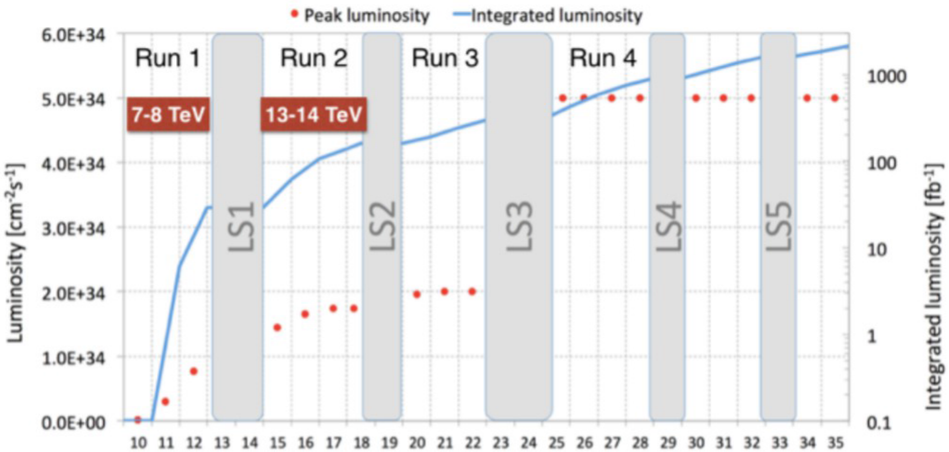
- World-wide CMS experiment: 100K cores, including Tier-1 and Tier-2 facilities
  – US CMS: 15K cores at Fermilab, 25K cores at Tier-2 and Tier-3 sites
- General purpose (non-CMS) at Fermilab: 12K cores
- Similar values for world-wide and US ATLAS, and for the ATLAS Tier-1 center (at BNL) and Tier-2 facilities

In addition, resources are utilized in "burst" mode



2020-2025 Compute and Data needs

- Two new programs are coming online, DUNE (long baseline neutrinos) and High-Luminosity LHC (HL-LHC), while new physics search programs using rare muon decays, and heavy quarks and tau leptons (Mu2e, Belle2) are operating
- Increased precision, higher luminosity, increased event complexity push computing needs to ~10X-100X of HEP infrastructure currently deployed
  - Lower value assumes optimized algorithms on new computing architectures and new approaches



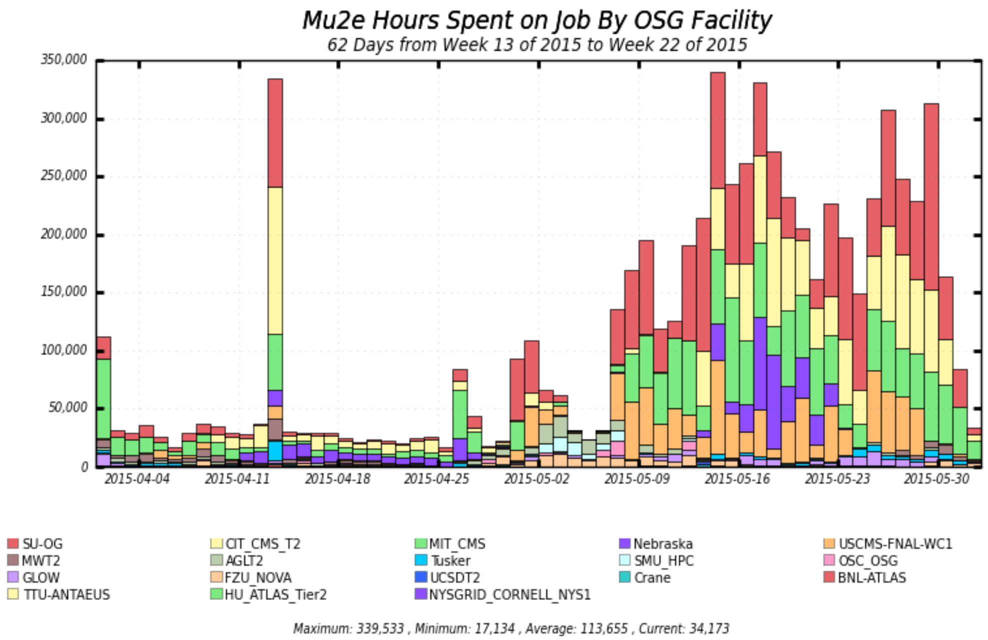
HL-LHC only will require from ~10X (anticipated optimization) to ~30X (no optimization) increase of compute capacity and ~10X to ~18X increase in storage (from 72PB/yr total today) from Run2

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Tape [PB]	2.8	2.8	2.8	2.8	19.24	54.43	103.55	153.89	204.64	255.39
Disk [PB]	4.00	4.00	5.00	8.00	27.98	79.17	115.68	153.10	190.82	228.55
CPU [kHepSPEC]	45.00	45.00	50.00	55.00	328.31	568.98	567.54	609.45	643.14	672.60

Belle2 computing requirements

2020-2025 Compute and Data needs

- Of course, new programs ramp up simulation campaigns much earlier than the beginning of data taking
  - Adding their own "cycles" of "burst usage" to those of operating experiments
- The need to provide "elasticity" in resource provisioning (to match "burst usage" patterns) becomes necessary



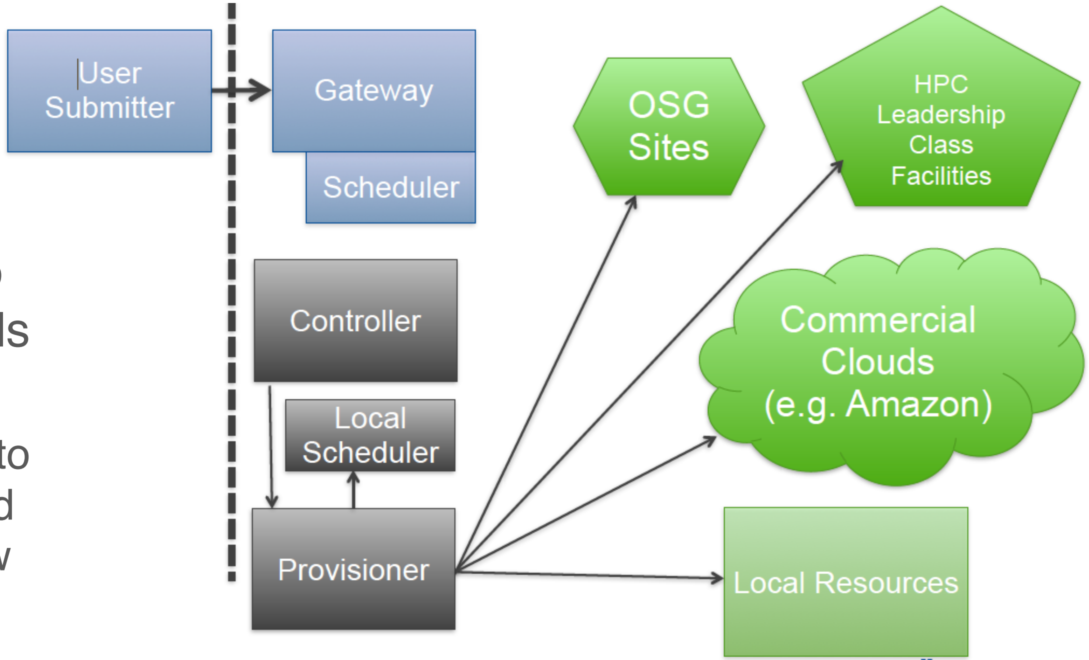
- Example: Mu2e experiment simulation campaign utilization of OSG resources through the Workflow management infrastructure and tools of the FNAL facility

Evolution of the computing paradigm

- Currently, Tier-0 and Tier-1 US HEP facilities support dedicated and shared resources, for data and compute intensive workflows
  - CPU, disk, hierarchical storage (including cache), tape, tape libraries; resources hosted at the facility or made available through OSG.
- Industry trend is to use Cloud services, either Infrastructure as a Service (IaaS), Platform as a Service (PaaS) or Software as a Service (SaaS)
  - Motivation includes high cost of provisioning and operating; need for redundancy or failover; ability to rapidly expand and contract resources; preference to purchase services (annual operating money) rather than make capital expenditures periodically; desire to pay only for the resources needed/used.
- Following this paradigm, US-HEP facilities could ***incorporate and manage*** "rental" resources, to achieve the "elasticity" that will satisfy demand peaks without overprovisioning local resources.
  - Options for obtaining access to such resources include experiment programmatic allocations at HPC facilities, access to HPC spare cycles, and commercial or academic cloud allocations
  - Model has to be studied for efficiency and cost-effectiveness

Example: the Fermilab HEPCloud Facility project

- HEPCloud Facility vision: a portal to a computing ecosystem of leadership and production class resources, either commercial or academic.
- Provides "complete solutions" to all users, with agreed upon levels of service
  - The Facility decides on routing to local or "rental" resources based on efficiency, cost, and workflow requirements
- Provide storage services appropriate to the system that the workflow is routed



Example: the Fermilab HEPCloud Facility project

- The goal is to integrate "rental" resources (such as Cloud and HPC resources) into the current Fermilab computing facility in a manner transparent to the user. Objectives include
  - A seamless user environment for all resource types, including necessary tools and infrastructure
  - The architecture, including network, needed to support required data rates.
  - The policies for efficiently using and prioritizing the use of different resources
  - The information security policies, procedures and monitoring.
- Partnership with resource providers to identify and perform necessary R&D essential for success
  - In the HPC case, will need to work both with ASCR research and facilities experts

Summary

- Over the next decade, two new major HEP accelerator based programs will be brought on-line: Dune and HL-LHC
- HEP will face significant computing challenges moving forward
  - Precision requirements lead to increased simulation needs
  - HL-LHC data volume and event complexity will push analysis needs and increase demand on central resources for analysis preparation (data reduction, ..)
- It will be advantageous for HEP to be able to utilize HPC resources outside “traditional” HPC workflows (accelerator, cosmology, LQCD), for simulation and analysis of experimental data.
- To achieve this goal we need to work in partnership with ASCR facilities and researchers to develop the infrastructure to
  - access HPC resources through the scientific workflows used by the experiments
  - move data in the HPC facility and distribute data out seamlessly to other facilities
  - optimize HEP software tools to take advantage of architectures that will be utilized in future HPC machines.