Snowmass On the Mississippi CSS2013

Summary from the Computing Frontier Study Group

L.A.T. Bauerdick, S. Gottlieb,
for the Computing Frontier Group

Outline

✦ Introduction

  ★

✦ Computational Challenges

  ★

✦ Data Management Challenges

  ★

✦ Networking Challenges

  ★

✦ Technology Developments

  ★

✦ Software, Training, Careers

  ★

✦ Some Common Themes and Conclusions

Summary of the Summer Study

✦ Subgroups for "user needs"

  ★ Each subgroup to interacted with the corresponding physics frontiers to assess the computing needs

✦ Subgroups for "infrastructure"

  ★ The infrastructure groups project computing capabilities into the future and see how the user needs map onto the trends

✦ The main result is a written report from each of the subgroups, and a summary report

  ★ draft reports becoming available now, overall report until end of the month

  ★ heard about a DOE sponsored meeting in December on Scientific Computing & Simulations in High Energy Physics (building on results from Snowmass)

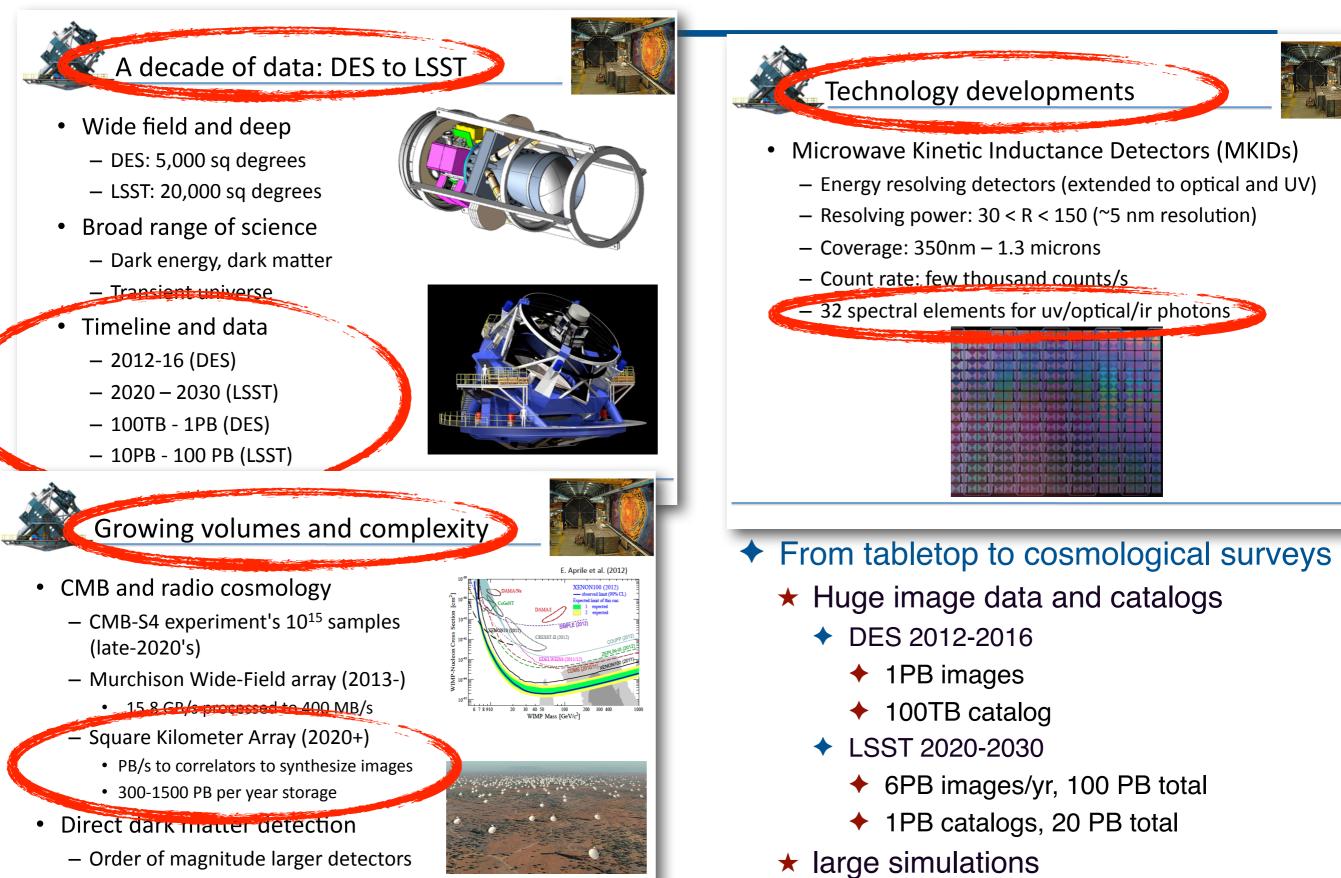# Subgroup Conveners

✦ Subgroups for "user needs"
  ✦ CpF E1 Cosmic Frontier
    ✦ Alex Szalay (Johns Hopkins), Andrew Connolly (U Washington)
  ✦ CpF E2 Energy Frontier
    ✦ Ian Fisk (Fermilab), Jim Shank (Boston University)
  ✦ CpF E3 Intensity Frontier
    ✦ Brian Rebel (Fermilab), Mayly Sanchez (Iowa State), Stephen Wolbers (Fermilab)
  ✦ CpF T1 Accelerator Science
    ✦ Estelle Cormier (Tech-X), Panagiotis Spentzouris (FNAL); Chan Joshi (UCLA)
  ✦ CpF T2 Astrophysics and Cosmology
    ✦ Salman Habib (Chicago), Anthony Mezzacappa (ORNL); George Fuller (UCSD)
  ✦ CpF T3 Lattice Field Theory
    ✦ Thomas Blum (UConn), Ruth Van de Water (FNAL); Don Holmgren (FNAL)
  ✦ CpF T4 Perturbative QCD
    ✦ Stefan Hoeche (SLAC), Laura Reina (FSU); Markus Wobisch (Louisiana Tech)

✦ Subgroups for "infrastructure"
  ✦ CpF I2 Distributed Computing and Facility Infrastructures
    ✦ Ken Bloom (U.Nebraska/Lincoln), Sudip Dosanjh (LBL), Richard Gerber (LBL)
  ✦ CpF I3 Networking
    ✦ Gregory Bell (LBNL), Michael Ernst (BNL)
  ✦ CpF I4 Software Development, Personnel, Training
    ✦ David Brown (LBL), Peter Elmer (Princeton U.); Ruth Pordes (Fermilab)
  ✦ CpF I5 Data Management and Storage
    ✦ Michelle Butler (NCSA), Richard Mount (SLAC); Mike Hildreth (Notre Dame U.)

Computing Challenges
at the
Physics Frontiers

Cosmic Frontier

A decade of data: DES to LSST

- Wide field and deep
  - DES: 5,000 sq degrees
  - LSST: 20,000 sq degrees
- Broad range of science
  - Dark energy, dark matter
  - Transient universe
- Timeline and data
  - 2012-16 (DES)
  - 2020 – 2030 (LSST)
  - 100TB - 1PB (DES)
  - 10PB - 100 PB (LSST)

Growing volumes and complexity

- CMB and radio cosmology
  - CMB-S4 experiment's 10^{15} samples (late-2020's)
  - Murchison Wide-Field array (2013-)
    - 15.8 GB/s processed to 400 MB/s
  - Square Kilometer Array (2020+)
    - PB/s to correlators to synthesize images
    - 300-1500 PB per year storage
- Direct dark matter detection
  - Order of magnitude larger detectors
  - G2 experiments will grow to PB in size

E. Aprile et al. (2012)

Technology developments

- Microwave Kinetic Inductance Detectors (MKIDs)
  - Energy resolving detectors (extended to optical and UV)
  - Resolving power: 30 < R < 150 (~5 nm resolution)
  - Coverage: 350nm – 1.3 microns
  - Count rate: few thousand counts/s
  - 32 spectral elements for uv/optical/ir photons

✦ From tabletop to cosmological surveys

  ★ Huge image data and catalogs
    ✦ DES 2012-2016
      ✦ 1PB images
      ✦ 100TB catalog
    ✦ LSST 2020-2030
      ✦ 6PB images/yr, 100 PB total
      ✦ 1PB catalogs, 20 PB total
  ★ large simulations

Energy Frontier

✦ EF will go to very high trigger rates and more complicated events

  ★ we looked back 10 years to aid prediction prediction of the magnitude of changes expected from programs over 10 years

✦ programs suggested for EF all have the potential for another **factor of 10 in trigger and 10 in complexity**

  ★ Simulation and reconstruction might continue to scale with Moore's law as they did for LHC, but could just as easily increase much faster

  ★ LHC adds 25k processor cores and 34 PB a year —in 10 yrs at this rate (flat budget) the capacity would be up by 4x - 5x

✦ Need make better use of resources as the technology changes

	Tevatron	LHC
Trigger	50Hz	ATLAS 500Hz CMS 350Hz LHCb 2kHz
RAW Event Size	150kB	ATLAS 1.5MB CMS 0.5MB
RECO Event Size	150kB	ATLAS 2MB CMS 1MB
Reco Speed	1-2 seconds on CPU of the time	10 seconds on CPU of the time

Intensity Frontier

HEP Intensity Frontier Experiments

Outside US   Taking data   List from DOE: ⟹ There are MANY

US participation

Experiment	Location	Status	Description	#US	US participation
Belle II ★	KEK, Tsukuba, Japan	Physics run 2016	Heavy flavor physics, CP asymmetries, new matter states	10 Univ, 1 Lab	55
BES III ★	IHEP, Beijing, China	Running	Precision measurements charm, charmonium, tau; search for and study new states of hadronic matter	6 Univ	26
CAPTAIN ★	Los Alamos, NM, USA	R&D; Test run 2015	Cryogenic apparatus for precision tests of argon interactions with neutrinos	5 Univ, 1 Lab	20
Daya Bay	Dapeng Penisula, China	Running	Precise determination of θ_{13}	13 Univ, 2 Lab	76
Heavy Photon Search ★	Jefferson Lab, Newport News, VA, USA	Physics run 2015	Search for massive vector gauge bosons which may be evidence of dark matter or explain g-2 anomaly	8 Univ, 2 Lab	47
K0TO ★	J-PARC, Tokai , Japan	Running	Discover and measure $K_L \to \pi^0 \nu\nu$ to search for CP violation	3 Univ	12
LArIAT	Fermilab, Batavia, IL	R&D; Phase I 2013	LArTPC in a testbeam; develop particle ID & reconstruction	11 Univ, 3 Lab	38
LBNE	Fermilab, Batavia, IL & Homestake Mine, SD, USA	CD1 Dec 2012; First data 2023	Discover and characterize CP violation in the neutrino sector; comprehensive program to measure neutrino oscillations	48 Univ, 6 Lab	336
MicroBooNE	Fermilab, Batavia, IL, USA	Physics run 2014	Address MiniBooNE low energy excess; measure neutrino cross sections in LArTPC	15 Univ, 2 Lab	101
MINERvA	Fermilab, Batavia, IL, USA	Med. Energy Run 2013	Precise measurements of neutrino-nuclear effects and cross sections at 2-20 GeV	13 Univ, 1 Lab	48
MINOS+	Fermilab, Batavia, IL & Soudain Mine, MN, USA	NuMI start-up 2013	Search for sterile neutrinos, non-standard interactions and exotic phenomena	15 Univ, 3 Lab	53
Mu2e	Fermilab, Batavia, IL, USA	First data 2019	Charged lepton flavor violation search for $\mu N \to eN$	15 Univ, 4 Lab	106
Muon g-2	Fermilab, Batavia, IL, USA	First data 2016	Definitively measure muon anomalous magnetic moment	13 Univ, 3 Lab, 1 SBIR	75
NOvA	Fermilab, Batavia, IL & Ash River, MN, USA	Physics run 2014	Measure $\nu_\mu \to \nu_e$ and $\nu_\mu \to \nu_\mu$ oscillations; resolve the neutrino mass hierarchy; first information about value of δ_{cp} (with T2K)	18 Univ, 2 Lab	114
ORKA	Fermilab, Batavia, IL, USA	R&D; CD0 2017+	Precision measurement of $K^+ \to \pi^+ \nu\nu$ to search for new physics	6 Univ, 2 Lab	26
Super-K	Mozumi Mine, Gifu, Japan	Running	Long-baseline neutrino oscillation with T2K, nucleon decay, supernova neutrinos, atmospheric neutrinos	7 Univ	29
T2K	J-PARC, Tokai & Mozumi Mine, Gifu, Japan	Running; Linac upgrade 2014	Measure $\nu_\mu \to \nu_e$ and $\nu_\mu \to \nu_\mu$ oscillations; resolve the neutrino mass hierarchy; first information about value of δ_{cp} (with NOvA)	10 Univ	70
US-NA61 ★	CERN, Geneva, Switzerland	Target runs 2014-15	Measure hadron production cross sections crucial for neutrino beam flux estimations needed for NOvA, LBNE	4 Univ, 1 Lab	15
US Short-Baseline Reactor	Site(s) TBD	R&D; First data 2016	Short-baseline sterile neutrino oscillation search	6 Univ, 5 Lab	28

CSS2013; Intensity Frontier Intro; July 29, 2013, H.Weerts   * not explicitly surveyed

The Method

- We wanted to **qualitative survey the community of current and future experiments in the IF** in order to understand the computing needs but also the foreseen evolution of said needs.

- Computing liaisons and representatives for the LBNE, MicroBooNE, MINERvA, MINOS+, mu2e, g-2, NOvA, Daya Bay, IceCube, SNO+, SK, T2K, SEAQUEST collaborations all responded to the survey and provided input.

- This does not cover all experiments in all areas but we consider it a representative survey of the field.

- More input is of course welcome. Please **see/email/chat Brian Rebel and myself** over these days or the next few weeks. *We want to thank the people that took the time to give well thought answers this survey.*

Computing Model

- We found a **high degree of commonality among** the various experiments' **computing models** despite large differences in type of data analyzed, the scale of processing, or the specific workflows followed.

- The model is summarized as a traditional **event driven analysis and Monte Carlo simulation using centralized data storage** that are **distributed to independent analysis jobs running in parallel on grid computing clusters**. Peak usage can be 10x than planned usage.

- **For large computing facilities** such a Fermilab, it is useful to **design a set of scalable solutions** corresponding to these patterns, with associated toolkits that would allow access and monitoring. Provisioning an experiment or changing a computing model would then correspond to adjusting the scales in the appropriate processing units.

- **Computing should be made transparent to the user**, such that non-experts can perform any reasonable portion of the data handling and simulation. Moreover, all experiments would like to see **computing become more distributed across sites**. Users without a home lab or large institution require **equal access to dedicated resources**.

★ Large diversity of experiments, but smaller-scale computing problems

★ Did survey of most experiments
  ✦ still live, <u>click here</u>

★ found reasonable convergence on a ~common computing model
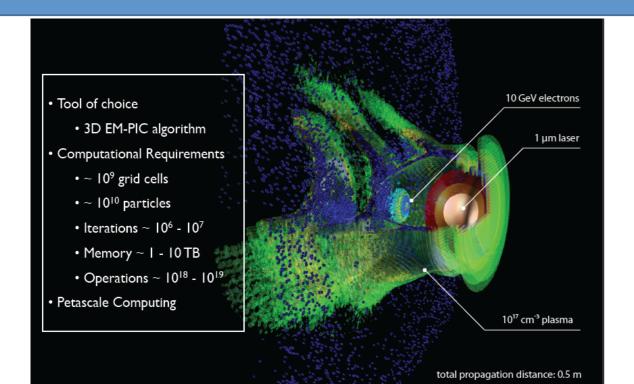
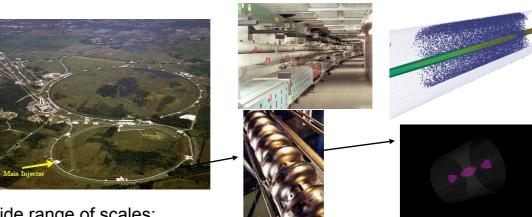# Accelerator Science

Accelerator Science Drivers

- New techniques and technologies
  - Optimize, evolve concepts, design accelerator facilities based on new concepts
- Maximize performance of "conventional" techniques and technologies
  - Optimize operational parameters, understand dynamics (manipulation and control of beams in full 6D phase space)
- Desirable outcome: achieve higher gradients for energy frontier applications, minimize losses for intensity frontier applications

✦ Considered community needs

- ★ beam loss characterization and control, even control room feedback
- ★ ability to produce end-to-end designs
- ★ better models, multi-physics
- ★ common interfaces, etc

Example: LWFA multi-scale physics

- Tool of choice
  - 3D EM-PIC algorithm
- Computational Requirements
  - ~ 10^9 grid cells
  - ~ 10^{10} particles
  - Iterations ~ 10^6 - 10^7
  - Memory ~ 1 - 10 TB
  - Operations ~ 10^{18} - 10^{19}
- Petascale Computing

10 GeV electrons

1 μm laser

10^{17} cm^{-3} plasma

total propagation distance: 0.5 m

Example: High-Intensity Proton Drivers

Main Injector

- Wide range of scales:
  - accelerator complex (10^3m) → EM wavelength (10^2-10 m) → component (10-1 m) → particle bunch (10^{-3} m)
  - Need to correctly model intensity dependent effects and the accelerator lattice elements (fields, apertures), to identify and mitigate potential problems due to instabilities that increase beam loss; thousands of elements, millions of revolutions
  - **Calculating 1e-5 losses at 1% requires modeling 1e9 particles, interacting with each-other and the structures around them at every step of the simulation**
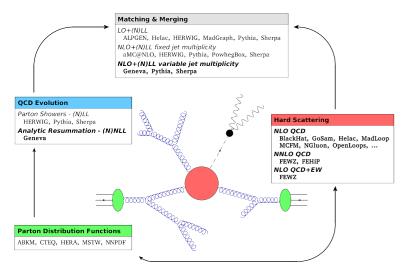
Lattice Field Theory

Summary bullet 1: Scientific goals

✦ The scientific impact of many future experimental measurements at the energy and intensity frontiers hinge on reliable Standard-Model predictions on the same time scale as the experiments and with commensurate uncertainties. Many of these predictions require nonperturbative hadronic matrix elements that can only be computed numerically with lattice-QCD. The U.S. lattice-QCD community is well-versed in the plans and needs of the experimental high-energy program over the next decade, and will continue to pursue the necessary supporting theoretical calculations. Some of the highest priorities are improving calculations of hadronic matrix elements involving quark-flavor-changing transitions which are needed to interpret rare kaon decay experiments, improving calculations of the quark masses m_c and m_b and the strong coupling α_S which contribute significant parametric uncertainties to Higgs branching fractions, calculating the nucleon axial form factor which is needed to improve determinations of neutrino-nucleon cross sections relevant experiments such as LBNE, calculating the light- and strange-quark contents of nucleon which are needed to make model predictions for the $\mu \rightarrow e$ conversion rate at the Mu2e experiment (as well as to interpret dark-matter detection experiments in which the dark-matter particle scatters off a nucleus), and calculating the hadronic light-by-light contribution to muon g – 2 which is needed to solidify and improve the Standard-Model prediction and interpret the upcoming measurement as a search for new physics.

Perturbative QCD

Broad impact of Perturbative QCD on collider physics

▷ interpreting LHC data requires accurate theoretical predictions

▷ complex SM backgrounds call for sophisticated calculational tools

▷ higher order QCD(+EW) corrections mandatory

✦ same importance of pQCD calculations in NLO and NNLO for EF experiment program

★ codes run on HPC

★ code libraries available for LHC at NERSC

This effort could greatly benefit from:

▷ unified environment for calculations/data exchange

▷ adequate computational means to provide accurate th[eory] pace and in a format useful to experimental analyses

▷ extensive computational resources to explore new tech[niques]

As pQCD component of the Computing Frontier we have set:

- **Short term goals**

  ▷ provide collider experiments with state-of-the-art theoretical predictions;

  ▷ make this process automated/fast/efficient;

  ▷ facilitate progress of new ideas and techniques for cutting-edge calculations (NLO with high multiplicity; NNLO).

- **Long term goals**

  ▷ take advantage of new large-scale computing facilities and existing computer-science knowledge;

  ▷ work in closer contact with computing community to benefit from pioneering new ideas (GPU, Intel Phi, programmable networks, …).

We have explored available options and provided some proofs of concept

U.S. Computational Infrastructure for HEP

U.S. HEP relies on a well-developed Compute and Data Infrastructure

HPC High Performance Computing
- often specialized processor architectures and interconnects

HTC High Throughput Computing
- distributed Grid or cloud of commodity compute nodes

Open Science Grid

National HPC Resources
- a planned cyber-infrastructure, based on "program needs"
- planned for, funded and built around **High Performance Computing** Centers to provide computing and storage resources
  - the NSF XD program,
    DOE Leadership-class facilities
- provide the "glue" across institutions (e.g. user accounting)
  - for XD program: the XSEDE project
- run an allocation process to satisfy computing need
  - example: XD X
  - DOE allocation

Computing Grids
- independently funded computing resources for science projects like LHC, HEP groups, labs, campuses
- Open Science Grid makes it into a grid infrastructure by forming a consortium of resource providers, science projects, campuses etc
- OSG services provide the "glue" to enable distributed **High Throughput Computing** across sites
- enable sharing of resources across stakeholders (VOs), and enable PIs to "opportunistically use" otherwise unused resources

Complementary Approaches!

Grids, HPC, Clouds

✦ Distributed High-Throughput (Grid) Computing
- ★ OSG provides 800M hours/year to EF and others
- ★ HTC workflows work well for IF and EF experiments
- ★ can implement improvements in evolutionary manner

✦ High Performance Computing
- ★ HPC is used and required by a number of projects
- ★ Adapting to ongoing and future architectural changes
  - ✦ diversity of complex nodes, memory/core, communication bottlenecks, multi-level memory hierarchy, power restrictions, --
- ★ requires new programming models -- rewrite large codes?

✦ Cloud Providers
- ★ Commercial clouds still too costly to replace dedicated resources
- ★ More existing Grid and HPC resources are moving to cloud interfaces
- ★ Cloud approach allows peak responses, dynamic allocation
- ★ Significant gaps and challenges exist in managing virtual environments, workflows, data, cyber-security, and other areas

Distributed and HPC Computing in HEP:

CMS + ATLAS in 2012:

~1.5 Billion Hours

National DOE & NSF HPC Centers 2013: HEP Only
~1.4 Billion Hours

Computational Needs from the Frontiers

Computational Needs: Cosmic Frontier

✦ Computational resources will have to grow to match the largely increasing associated data rates (makes for data intensive compute needs!)

   ★ In some cases require new computational models for distributed computing (including many-core systems)

   ★ Infrastructure for data analytics applicable to large and small scale experiments will need to grow over the next decade (with an emphasis on sustainable software, not just build-you-own)

      ✦ Data Archiving and Serving: data archives, databases, and facilities for post-analysis are becoming a pressing concern

      ✦ Archives now mostly used to "download", development of powerful, easy-to-use remote analysis tools

✦ Simulations

   ★ (cosmological, instrument) will play a critical role in evaluating and interpreting the capabilities of current and planned experiments

   ★ need for instrument simulation

Experimental Data	2013	2020	2030+
Storage	1PB	6PB	100-1500PB
Cores	10^3	70K	300+K
CPU hours	3×10^6 hrs	2×10^8 hrs	$\sim 10^9$ hrs
Simulations			
Storage	1-10 PB	10-100PB	>100PB - 1EB
Cores	0.1-1M	10-100M	>1G
CPU hours	200M	>20G	>100G

Computational Needs: Energy Frontier

✦ Driven by Trigger rate, event size, and reconstruction time
   ★ The programs suggested for energy frontier all have the potential for <u>another factor of 10 in trigger and 10 in complexity</u>
   ★ Simulation and reconstruction might continue to scale compatible with Moore's law as they did in the past, but much faster increase possible
   ★ <u>Moore's scaling requires full use of many-core architectures</u>
      ✦ EF has made transition to multi-core workflows, scaling to 16-32 per node
✦ New computing technologies allow to tap into more diverse resources
   ★ sharing, on-demand resource provisioning, opportunistic resources
      ✦ smooth out peaky computing needs, improve turn-around of "campaigns"
   ★ commercial clouds are not (yet) competitive, but the clouds are here
      ✦ e.g. use of HLT farms, use of Vodaphone cloud opportunistically for the LHC, etc
✦ More efficient code, high-performance low-power hardware (e.g.GPUs)
   ★ significant re-engineering effort has started, long-term program
✦ Possibly record more events than can be processed immediately
   ★ e.g. archive high-trigger-rate streams, and only selective reconstruction

Computational Needs: Intensity Frontier

✦ Computational demands of IF modest compared to those of EF
  ★ However the needs are NOT insignificant.
✦ Efficient use of available grid resources will have a huge impact
  ★ All efforts will benefit from dedicated transparent access to grid resources
    ✦ and this is a strategy for the Fermilab-based experiments and others
  ★ Grid approach will help w/ <u>peak CPU usage that can be 10x of the planned</u> average
✦ US participation in international IF efforts uses a combination of grid resources based mainly outside of the US and smaller local clusters
  ★ <u>It was widely noted that the lack of dedicated US resources has a detrimental impact on the science</u>.
  ★ Dedicated grid resources for the intensity frontier across experiments would have the largest impact on our international efforts
✦ Found a high degree of commonality among the various experiments' computing models at IF and EF, despite large differences in type of data analyzed, the scale of processing, or the specific workflows followed

Computational Needs: Field Theory LQCD and pQCD

✦ Lattice QCD simulations require parallel machines

  ★ Codes are floating point intensive, limited by memory bandwidth, network latency and bandwidth, implemented using message passing (MPI)

✦ Computationally intensive: generate gauge-field config. ensembles

  ★ <u>Capability Hardware</u>, jobs each use 10K to 100K+ processors

    ✦ USQCD uses DOE Leadership Computing Facilities (DOE ASCR funded)

    ✦ Also NERSC, LLNL, NSF XSEDE, Japan (RIKEN BNL), UK (UKQCD BlueGene/Q)

    ✦ allocations e.g. in 2013 290M CPUh at ANL, 140M CPUh at ORNL, among largest at LCFs

✦ "Measurements" repeat a calculation with all members of an ensemble

  ★ <u>Capacity Hardware</u> (now used for > 50% of Flops)

    ✦ USQCD has dedicated LQCD systems and support personnel at BNL, FNAL, JLab (including currently 812 GPUs)

✦ USQCD has submitted a follow-on 5 year proposal

  ★ 2.1 —> 12.5 billion CPUh capacity-class,
2.8 —> 17.8 billion CPUh capability-class in 2015..2019

✦ pQCD computational needs 100k-1M hours for dedicated measurements

  ★ for verifications of event generation libraries for NLO and NNLO

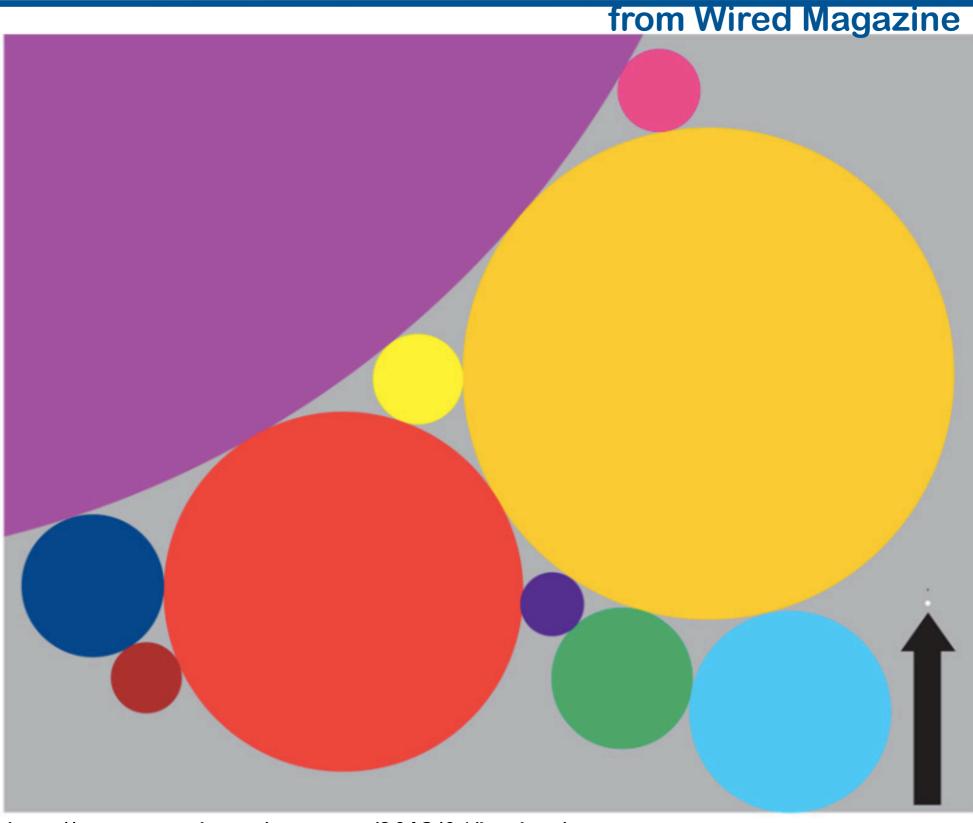 ★ otherwise computational needs contained in experiment requests

Computational Needs: Accelerator Science

- ✦ Multi-physics modeling necessary to advance accelerator science
  - ★ Requires frameworks to support advanced workflows
  - ★ Efficient utilization of large computing resources, HPC in many cases
- ✦ Currently run production with 10k to 100k cores, 2013 allocation ~140M hrs
  - ★ needs are x10-100 for allocation/year (if all different accelerator options are pursued)
- ✦ Evolving technologies (Intel Phi) require R&D, could result in significant changes
  - ★ Advanced algorithmic research underway, will require continuing support
  - ★ Programmatic coordination necessary to efficiently utilize resources
  - ★ Opportunity for multi-scale, multi-physics modeling and "near-real-time" feedback, if could be used efficiently
  - ★ Intensity frontier machines would like "control room feedback" capabilities
- ✦ Current strategy is to abstract and parameterize data structures so that are portable and enable efficient flow of data to a large number of processing units in order to maintain performance.
  - ★ Already ported subset of solvers and PIC infrastructure on the GPU
  - ★ Evaluate current approach, develop workflow tools and frameworks
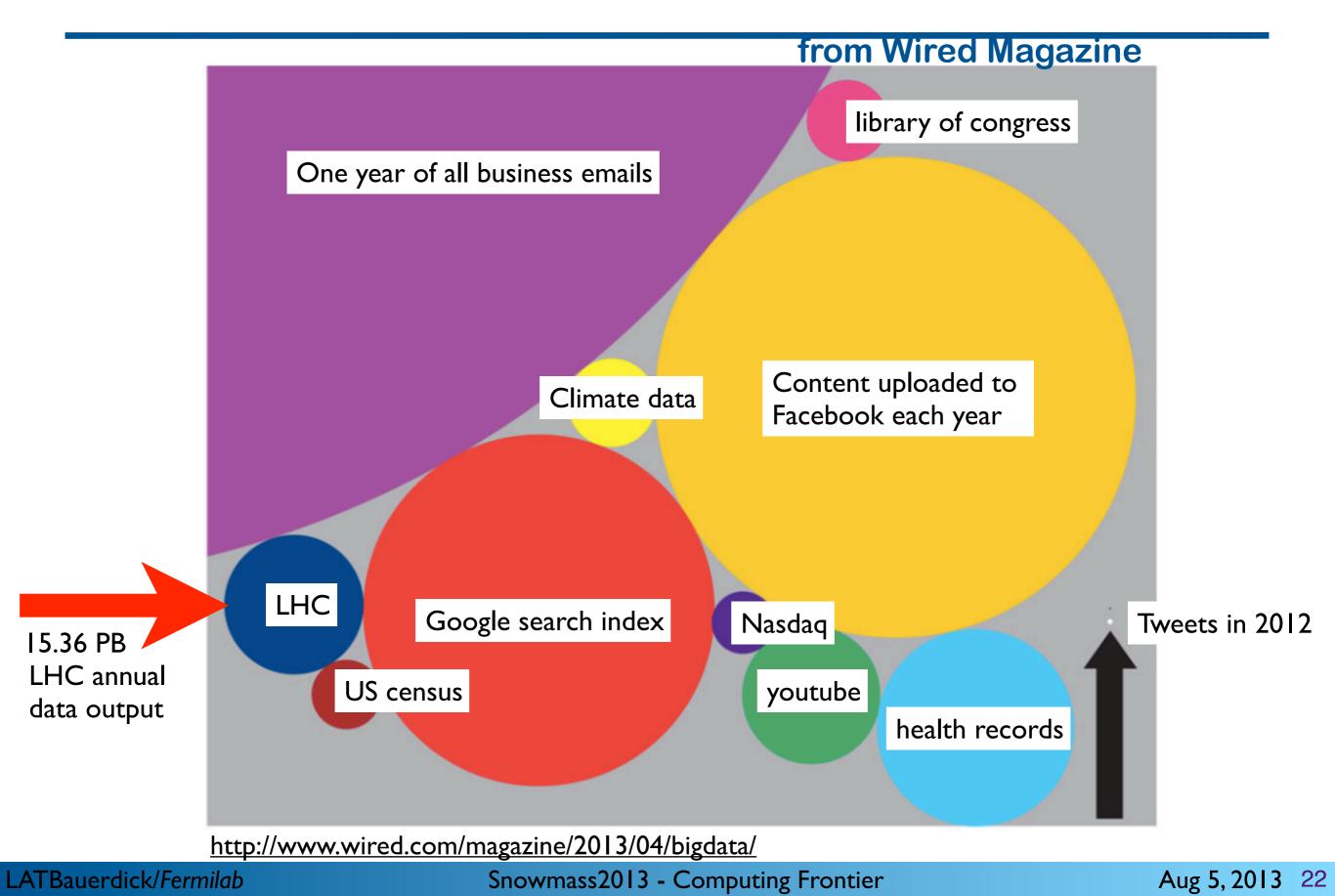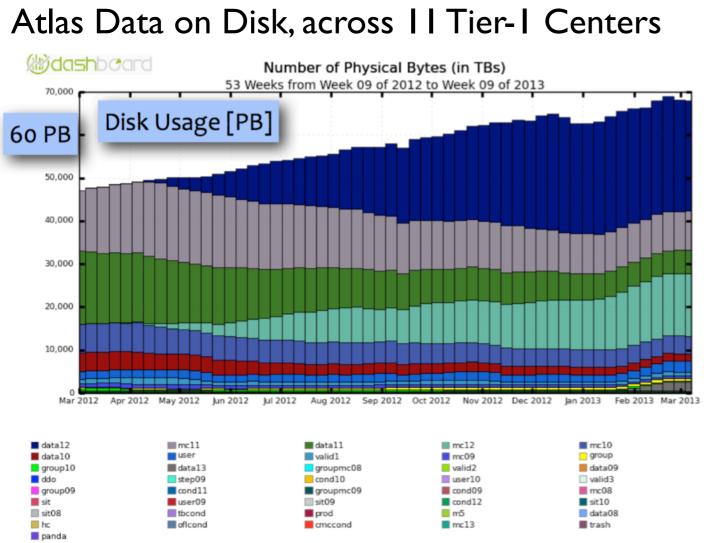 - ★ Investigate new algorithms and approaches

The Big Data Frontier

http://www.wired.com/magazine/2013/04/bigdata/

The Big Data Frontier

from Wired Magazine

library of congress

One year of all business emails

Climate data

Content uploaded to Facebook each year

LHC

15.36 PB LHC annual data output

Google search index

US census

Nasdaq

youtube

health records

Tweets in 2012

http://www.wired.com/magazine/2013/04/bigdata/

Data Management: Energy Frontier

Atlas Data on Disk, across 11 Tier-1 Centers

Number of Physical Bytes (in TBs)
53 Weeks from Week 09 of 2012 to Week 09 of 2013

- **Current** LHC data sizes
  - ★ examples
    - ✦ Atlas: 70 PB on disk, world-wide
    - ✦ CMS 18.2 PB on tape, at Fermilab
- **Future** increases
  - ★ an estimate for 2021
    - ✦ ~130 PBytes detector data
    - ✦ ~350 PBytes simulated data
    - ✦ ~270 PBytes US "data library" (across Tier-1/2 centers)
    - ✦ CMS estimates for #events:
      - ✦ 30B data events, 46B simulated events

CMS Data in the Fermilab Tape Library

	Custodial [TB]		Non-Custodial [TB]		Total [TB]	
	Last week	Total	Last week	Total	Last week	Total
Data	143.39	9048.89	86.96	1854.74	230.35	10903.63
MC	0.69	6831.85	0.00	416.03	0.69	7247.88
Total	144.08	15880.73	86.96	2270.77	231.05	18151.51

Data Management: Energy Frontier

- ✦ Cost consideration (storage and processing) constrain the data rate
  - ★ make choices and set priorities about which type of events we can collect and what data analyses we can follow through on, based on how much computing and storage resources we can afford
  - ★ BTW, Science at EF lepton colliders is unlikely to be constrained by data management and storage issues
- ✦ LHC currently processes all data that's written to persistent storage
  - ★ thus, might not have enough resources to record all physics
- ✦ Possibility to stream large fraction of data for storage on tape only
    - ✦ for just a penny you can store almost 1,000 CMS raw events on tape!
  - ★ no further reconstruction, distribution unless physics case arose
    - ✦ exploits cost hierarchy for storage, also in view of slowed disk price decay
  - ★ progressively pair down the active dataset with understanding and time

Data Management: Energy Frontier

- ✦ Storage resources are expensive
  - ★ We cannot afford another factor of 100 increase in storage, so we need to find ways of being more efficient in the use of the space
- ✦ Distributed data management is successful, but manpower intensive
  - ★ need distribute and serve the data much more flexibly and dynamically
- ✦ New development: Remote data access over the network
  - ★ <u>less requirements on data locality</u> (enables more cost-effective clouds)
  - ★ allows flexible, dynamic data placement, federation of existing data stores
  - ★ possible centralization of cost-effective data stores and archival facilities
  - ★ such data federations already deployed as a first step, more work needed
- ✦ Data Management evolving to "content delivery networks"
  - ★ data management resources that deliver data on demand
  - ★ data to be cached and replicated intelligently
- ✦ This brings large demands on network connectivity
  - ★ a 10k core cluster (probably typical for 2020) would require 10Gb/s networking for organized processing, analysis would require 100Gb/s

Data Management:
Cosmic Frontier

✦ Continued growth in data from Cosmic Frontiers experiments

   ★ currently exceed 1 PB, 50 PB in 10 yrs, 400 PB per year in 10-20 yrs

✦ Simulation plays a vital role in understanding all aspects of astrophysics

   ★ is also needed for the design of observational programs and for their detailed technical elements.

✦ Already today, post-processing of simulation data presents a major data-intensive computing challenge, requiring data management, large-scale databases and tools for data analytics.

   ★ Some of today's pain relates to the much more ready availability of national resources for computation than those for data management and analysis: "we can easily generate many petabytes from simulations and <u>have [almost] no place to store them and analyze them</u>"

Data Management: Cosmic Frontier

✦ Sky Surveys

★ use of <u>innovative database technology</u> to make its data maximally useful to scientists.

★ DES takes data since 2012, 0.6 gigapixel camera, culminating in PB dataset

★ LSST's 3.2 gigapixel camera will produce 15 terabytes per night, building up to over 100 PB of images and 20 PB of catalog database during first ten yrs.

★ LSST develops a multi-petabyte scalable object catalog database that is capable of rapid response to complex queries. The data management needs are handling image catalogs and object catalogs

★ Baseline <u>LSST object catalog employs HEP's xrootd technology</u> in key role of providing a switchyard between MYSQL front ends and thousands of MYSQL backend servers

★ Not all LSST science will be possible using only the object catalog database. In particular, studies such as those for dark energy effects, of particular interest to the HEP community, are likely to require reprocessing of the LSST image data on HEP analysis facilities. The model for funding and executing these studies is not yet clear.

Data Management: Cosmic Frontier

✦ Arrays of radio telescopes

★ data-volume challenge <u>comparable with EF experiments</u>.

★ The most extreme example now being planned is the European-led Square Kilometre Array (SKA) project that expects to complete its Phase I system in 2020. SKA will feed petabytes/s to correlators that will synthesize images in real time, producing a reduced persistent dataset on the scale of <u>300 to 1500 petabytes per year</u>.

★ These volumes can only be realized if considerable evolution of computing and storage costs happens by the time SKA data flows. Although SKA currently has no US involvement, it presents a concretely planned example of the technologies and data-related challenges that will certainly be faced by US scientists involved in projects in the same timeframe.

✦ Todays example of the SKA concept is the Murchison Wide-Field array where a raw 15.8 gigabytes/s is processed to a produce a stored 400 MB/s

Data Management:
Intensity Frontier

✦ Largest data set: Lepton colliders in intensity frontier "factory mode"

  ★ also run up against the cost of storage, but the physics of lepton collisions is relatively clean and recording all events relevant to the targeted physics has proved possible in the past and is a realistic expectation for the future.

  ★ The Belle II TDR estimates a data rate to persistent storage of 0.4 to 1.8 gigabytes/s, which is comparable to LHC Run-2 rates but without the need to discard data with significant physics content.

✦ Most of the many other intensity frontier experiments do not individually challenge storage capabilities, but there is a recognition that data management (and workflow management) is often inefficient and burdensome.

  ★ Most experiments find it hard to escape from the comfort and constriction of limiting all their data-intensive work to a single site – normally Fermilab.

  ★ The statement "all international efforts would benefit from an LHC-like model" was made and should probably be interpreted as a need for LHC-like data management functions at a much lower cost and complexity than the current LHC system

Data Management:
The Challenge of Data Longevity

- ✦ What to do with 100s of PB of data over 10s of years?
  - ★ Irreplaceable resource, should be preserved, somehow, for the future
  - ★ also requirement for open access to experiment data
- ✦ With respect to data preservation and open availability, the LHC experiments are actively developing appropriate policies
  - ★ The intensity frontier community does not have a plan yet, but recognizes that the issue exists across the frontiers
  - ★ techniques exist for data preservation and open availability, once policies have been decided, but requires funding to do it
  - ★ first U.S. projects for EF, also linking to Biology, Astrophysics, Digital Curation
- ✦ For the Cosmic Frontier, images, and tabular object catalogs of sky surveys and other image-based astronomy are readily intelligible by other scientists and even the general public

Technology Challenges

Computations and Data

- ✦ Federated data is well-matched to distributed high-throughput computing on grids and clouds
  - ★ but requires networking and last-mile problems to be solved (e.g. science DMZs etc)
  - ★ The example of the tiered computing used by the LHC experiments and the distributed data handling system is a good basis for developing the model for the IF
- ✦ Enabling HPC access to large data set
  - ★ use cases increasingly require data intensive computing also at HPC
  - ★ current systems like at Argonne and Oakridge can be configured for data-centric analysis, but they are not specifically designed for this task, nor will next generation systems
  - ★ ALCF, NERSC, and OLCF are arguing for the Virtual Data Facility (VDF) concept as their preferred mode for data storage/analysis
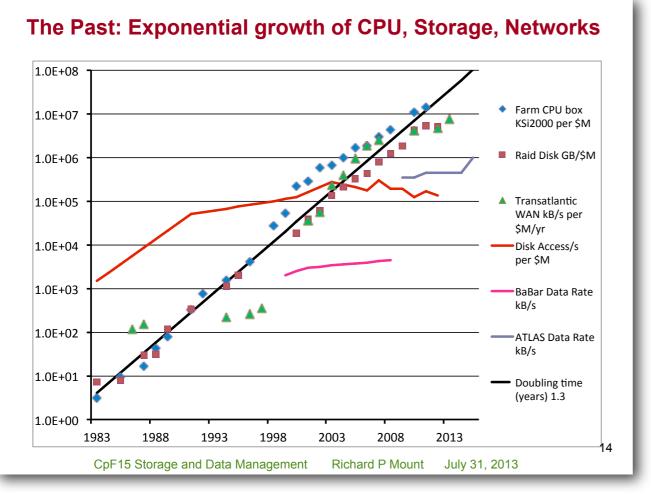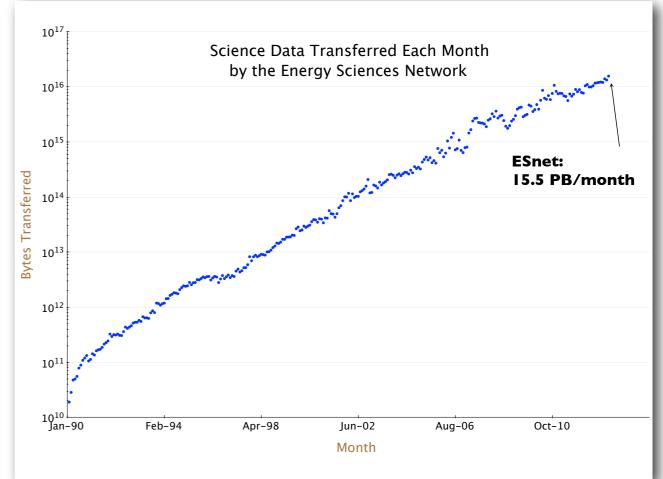
Networking

✦ All Frontiers depend on reliable, high-bandwidth, feature-rich networks

✦ Most HEP-related data is transported by National Research and Education Networks (NRENs), supplemented by infrastructures dedicated to specific projects (like USLHCnet)

★ NRENs differ from commercial network providers; they are optimized for transporting massive data flows from large-scale scientific collaborations.

★ NRENs offer advanced capabilities (such as multi-domain bandwidth guarantees) that are not generally available commercially.

★ HEP was a pioneer in exploiting international research networks, other science disciplines are making a similar transition now.

★ NRENs will be challenged as a result, and must be adequately motivated, resourced, and engaged

✦ HEP's objectives through 2020 require basic and applied network research

★ Need support and growth for research in areas relevant to networking science

★ Recent investments have been too small, using up dividends from prior research

★ <u>Translate research results into operational practices</u> is critical, but poorly funded

✦ No fundamental technical barriers to transport 10x more traffic within 4 years

Networking

✦ Networks should be treated like a resources and need to be managed
- ★ R&D to develop cost-efficient architectures, to manage complexity
- ★ exploit programmability and other emerging network paradigms
- ★ assure that networks and applications become more tightly integrated

✦ Networks today provide reliable and very high throughput

✦ Collaborations should design workflows around this fact
- ★ The large gap between peak and average transfer rates must be closed.
- ★ Campuses must deploy high performance Local Area Network Infrastructure

✦ The cost for networks will continue to go down
- ★ Whether the overall cost of networking remains stable over the next decade depends on the declining cost-curve for optical components, as well as the price of trans-Atlantic capacity
- ★ but there is a broad market for both

✦ In general, it's much cheaper to transport data than to store it

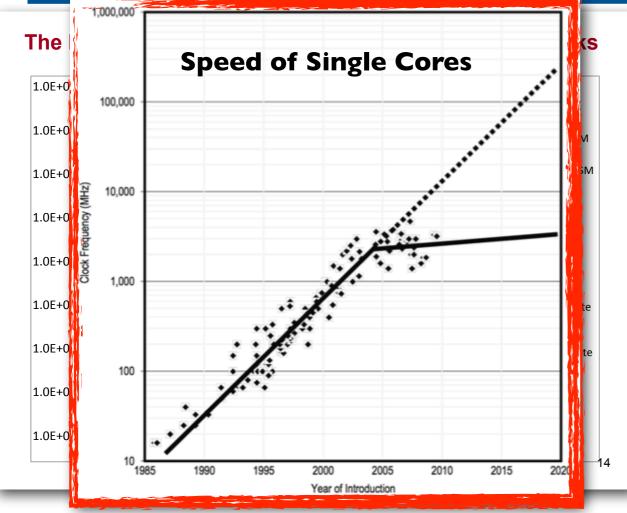# Technology Challenges after Decades of Exponential Growth

The Past: Exponential growth of CPU, Storage, Networks

Legend:
- Farm CPU box KSi2000 per $M
- Raid Disk GB/$M
- Transatlantic WAN kB/s per $M/yr
- Disk Access/s per $M
- BaBar Data Rate kB/s
- ATLAS Data Rate kB/s
- Doubling time (years) 1.3

CpF15 Storage and Data Management   Richard P Mount   July 31, 2013

Science Data Transferred Each Month by the Energy Sciences Network

ESnet: 15.5 PB/month

- ✦ Major shift in the nature of processors
  - ★ single sequential applications has roughly stalled due to limits on power consumption
- ✦ We have been living in a temporary period of "multicore", but even this cannot last due to power constraints
- ✦ Rotating disk will suffer marked slowdown capacity/cost.
  - ★ Computing models must attempt to optimize roles of tape, rotating disk, solid-state storage, networking and CPU

Technology Challenges after Decades of Exponential Growth

Speed of Single Cores

[Graph: Clock Frequency (MHz) vs. Year of Introduction, 1985–2020, showing single core speed growth leveling off after ~2005]

Disks – from Per Brashers/DDN

SLAC

- The area of a "bit" in current products is close to the limit where what is written will remain magnetically stable.

- New technologies to make the "bits" more stable are on the horizon:
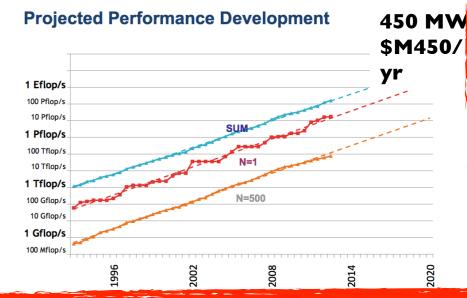 - "Shingled Recording" Not easily re-writable

  Disk Platter

  - Heat Assisted Magnetic Recording (HAMR)
  - [Laid-out-in-advance] Bit Patterned Recording
- None of these looks good for the near future.

- ✦ Major shift in the nature of processors
  - ★ single sequential applications has roughly stalled due to limits on power consumption
- ✦ We have been living in a temporary period of "multicore", but even this cannot last due to power constraints
- ✦ Rotating disk will suffer marked slowdown capacity/cost.
  - ★ Computing models must attempt to optimize roles of tape, rotating disk, solid-state storage, networking and CPU

NERSC Revolution in Energy Efficiency Needed

Even though energy efficiency is increasing, today's top supercomputer (N=1) uses ~9 MW or roughly $9M/year to operate. Even if we could build a working exaflop computer today, it would use about 450 MW and cost $450M/year to pay for power.

Projected Performance Development

450 MW $M450/yr

[Graph: Projected Performance Development, flop/s vs. year 1996–2020, showing SUM, N=1, N=500 curves]

Addressing Technology Challenges

✦ Exploit multi-threading and GPU environments, limit energy consumption

   ★ Advances in adapting key software tools to will be beneficial across frontiers.

   ★ Writing efficient codes is likely to become more difficult as we move to more exotic processors like GPUs or the Xeon Phi. It is not clear that one can abstract the details of the hardware in such a way that a single code can be written for both those targets

   ★ Applies also to large scale simulations in the cosmic frontier and lattice gauge theory are probably fairly similar

✦ Keep storage cost low: compute instead of store

   ★ Many of the components required to support virtual data already exist in the data and workflow management software of the largest experiments. The rigorous provenance recording required to support the virtual data concept would also benefit data preservation.

   ★ Computing model implementations should be flexible enough to adapt to a wide range of relative costs of the key elements of HEP computing. In preparing for Run 3, the LHC program should seriously consider virtual data as a way to accommodate scenarios where storage for derived and simulated data becomes relatively very costly

Software and Training

Software and Training

- ✦ Three main themes or goals, and a number of recommendations
  - ★ maximize the scientific productivity in an era of reduced resources
    - ✦ <u>use software development strategies and staffing models that will result in products that are generally useful for the **wider HEP community**</u>
  - ★ evolving technology especially with respect to computer processors
    - ✦ <u>develop, evolve software that will perform with efficiency in future computing systems</u>
  - ★ increasingly complex software environments and computing systems
    - ✦ <u>insure that developers and users have the training needed to create, maintain, use</u>
- ✦ Some of the recommendations
  - ★ Significant investments in software to adapt to the evolution of computing processors: R&D into techniques, and as reengineering "upgrades"
  - ★ Allow flexible, reliable funding of software experts to facilitate transfer of software and sharing of technical expertise between projects
  - ★ Facilitate code sharing: open-source licensing, publicly-readable repositories
  - ★ Include software i/s, frameworks, and detector-related applications early in project reviews, integrate software professionals with scientists

All Frontiers Agree:
Need for Training and Career Paths

✦ Encourage training, as a continuing activity

✦ Use certification to document expertise and encourage learning new skills

  ★ Use mentors to spread scientific software development standards

  ★ Involve computing professionals in training of scientific domain experts

  ★ Use online media to share training

  ★ Use workbooks and wikis as evolving, interactive software documentation

✦ Provide young scientists with opportunities to learn computing and software skills that are marketable for non-academic jobs

✦ Training and career paths (including tenure stream) for researchers who work at the forefront of computation techniques and science is critical

Conclusions

Conclusions on Computations

✦ Challenging resource needs require efficient and flexible use of all resources

   ★ HEP needs both Distributed High-Throughput computing (experiment program) and High-Performance computing (mostly theory/simulation/modeling)

   ★ emerging experiment programs might consider a mix to fulfill demands

   ★ programs to fund these resources need to continue

   ★ sharing and opportunistic use help address resource needs, from all tiers of computing, eventually including commercial

   ★ more need for data intensive computing, including at HPC, for data analytics, combining simulations and observational data etc.

✦ To stay on the Moore's law curve, need to proactively make full/better use of advanced architectures

   ★ with the need for more parallelization the complexity of software and systems continues to increase: frameworks, workload management, physics code

   ★ important needs for developing and maintaining expertise across field, including re-engineering of frameworks, libraries and physics codes

✦ Unless corrective action is taken we could be frozen out of cost effective computing solutions on a time scale of 10 years.

   ★ There is a large code base to re-engineer

   ★ We currently do not have enough people trained to do it

Conclusions on Data

✦ The growth in data drives need for continued R&D investment in data management, data access methods, networking

  ★ Continued evolution will be needed in order to take advantage of new network capabilities, ensure efficiency and robustness of the global data federations, and contain the level of effort needed for operations

✦ Networks can be relied on to serve as foundation of data intensive distributed computing

  ★ emerging network capabilities and data access technologies improve our ability to use resources independent of location, over the network

  ★ enables a large spectrum of provisioning resources: dedicated facilities, universities, opportunistic use, commercial clouds, leadership-class HPC,...

  ★ treat networks as resource, include in computing models, etc

  ★ significant challenges with data management and access, projecting solutions will be based on content delivery approaches, dynamic data placement, remote data access

✦ Have to learn to do more with less. This requires being more flexible and perhaps tolerating higher levels of risk

Computing is at a great starting point for moving into the HEP future

- ✦ We have established and well-working computing models
  - ★ the different frontiers are at some level separate in terms of facilities
  - ★ <u>but we are identifying many commonalities in terms problems and approaches</u>
  - ★ by coming together we are mapping out a good way to go forward
- ✦ HEP success has always been tied to advances in computing
  - ★ like LHC computing being enabled by networks and the Grid
- ✦ Push for using new technologies and approaches that are transformative
  - ★ like parallelization and multi-core, virtualization, GPUs etc
  - ★ for sure we'll see things over the coming years we have not yet thought of
- ✦ Industry caught up to us, and in cases surpassed us
  - ★ might reassure us that with hard work computing won't be a road block
  - ★ look at how industry can help
- ✦ HEP still drives areas of the technology, and certainly the collaborative space
  - ★ distributed computing requires collaboration and partnerships, between sites, science communities, with computer scientists, between funding agencies etc
- ✦ We have a lot to bring to the table working with other communities
  - ★ and are learning a lot in doing so!
- ✦ More commonality and community planning is needed for future computing systems

Thanks!

to all sub-conveners and observers, to all participants in and contributors to the Computing Frontier summer study, and to the organizers of an exciting and inspiring Snowmass meeting here in Minneapolis