Accelerating Scientific Discovery and Engineering Practice through Advanced, High Spectrum Computing and Data Analysis

Requirements for the NSF High-Performance Computing and Analysis Ecosystem

William Kramer

Partial List of Contributors (alphabetical):

Greg Bauer, Jerry Bernholc, Tom Bettge, Keith Bisset, Larry Di Girolamo, Thom Dunning, Steve Gottlieb, Tom Jordan, Al Kellie, Gerhard Klimeck, Phil Maechlin, Chris Malone, Mike Norman, Brian O'Shea, Jim Phillips, Nikolai Pogorelov, Patrick Reed, Vadim Roytershteyn, Klaus Schulten, Ed Seidel, Emad Tajkorshid, Warren Washington, Don Weubbles, Paul Woodward, P.K. Yeung

A INTRODUCTION

In just one short year of service, the current NSF Track-1 supercomputing system, NCSA's Blue Waters, has enabled *transformative* and *wide-ranging impacts* across a broad range of science and engineering disciplines. Blue Waters supports the country's leading research teams and is already transforming the scientific landscape in ways that are not possible on other systems nor by any other means. The system is delivering more computational and data analysis resources than all the other systems in the NSF portfolio combined, directly leading to many discoveries. A few examples include:

- unprecedented understanding of HIV and other biological systems,
- new predictions in earth sciences such as the never-seen-before detail of seismic hazard maps of the Los Angeles basin,
- new space science insights revealing breathtaking phenomena with the first fully kinetic, 3D hybrid turbulent simulation for understanding the impact of space weather on the earth's environment,
- new understanding of the formation of the universe and its constituents to hundreds to thousands of times the fidelity of past studies,
- the most complex study of the earth's hydrology ever attempted,
- deepened knowledge of the most fundamental particles in physics,
- expanded petascale to industrial OEM applications, and more.

As a balanced and integrated system with very high sustained computational performance, extraordinary analytical capabilities, very large memory, world-leading storage capacity and performance, leadership-level networking, and an advanced service architecture, the current Track-1 system is powering teams across all NSF directorates to do breakthrough science that would otherwise be impossible.

With the creation of its multiple-track cyberinfrastructure program, NSF established itself as a world leader in accelerating science and engineering discovery and fostering national competitiveness through computational and data analysis at the highest levels. While the current Track-1 system has several more years of grand challenge science potential, it is now time to establish a future computational and analytical requirements and practices for the next decade in the broader context of an integrative and comprehensive cyberinfrastructure for the academic and broader research community. As Paul Woodward, one NSF PRAC PIs stated, the "first Track-1 system available to the community has set in motion a significant rethinking by NSF investigators of what is possible and what is practical. It would be a very bad idea to nip this flowering of very large scale computation in the bud" by not having a path forward beyond the useful life of the first Track-1 system. The NSF Track-1 systems, a critical component to the nation's competitiveness and well-being, are similar in scale to DOE leadership systems but serve NSF's wider and deeper open research community in their pursuit of fundamental as well as programmatic research.

At the same time, the broader expectations of computing infrastructure have changed significantly since 2006, as the needs of science and engineering communities are diversifying, as large instruments

and high-fidelity simulations generate huge quantities of data that must be analyzed, and as multiple communities need to work together to address modern grand challenges with multiple computing challenges and modalities. Hence, the next High Spectrum HPCD system must be deeply embedded in a diverse ecosystem of instruments, data archives, smaller Track-2 systems, clouds, and digital services to support the diverse needs of the communities NSF serves.

The strategy provides both existing and new areas of computational science and engineering with a roadmap of enhanced resources that will allow research teams to again rethink "what is possible and what is practical" in advancing scientific discovery and the state of the art in engineering., science and engineering research drivers, and expected requirements; straw horse examples of system capabilities and characteristics; an evolution of service architectures to meet the evolving needs of science teams; and a proposed path forward along with a cost estimate.

B SCIENCE, ENGINEERING, AND RESEARCH DRIVERS AND EXPECTED REQUIREMENTS

The science, engineering and research drivers shared here came from multiple sources, including information presented and discussed at the NCSA Blue Waters Symposium in May 2014, interviews with over 25 NSF PRAC PIs during the spring and summer of 2014, and written summaries for S&E team goals and needs for the timeframe. Information and input also comes from other summary sources, such as community workshop reports, interactions with other projects, etc.

Due to space and time limitations, it is not feasible to document the detailed science goals for each area, nor each S&E team's plans for achieving those goals. Instead, we discuss the six broad areas of future science and engineering (S&E) requirement trends and drivers that are common across many S&E communities in varying degrees as we move beyond petascale computing and analysis.

Before discussing the trends, however, we present several brief examples of future science team objectives. These examples illustrate directions seen in many more science communities, which will require high spectrum computational system and a more comprehensive and deeply integrated data anlytics. The details across all areas are in tables in the appendix.

- **Seismic Science**[a]**—**Seismic hazard analysis (SHA) is the scientific basis for many engineering and social applications for performance-based design, seismic retrofitting, resilience engineering, insurance-rate setting, disaster preparation, emergency response, and public education. All of these applications require a probabilistic form of seismic hazard analysis (PSHA) to express the deep uncertainties in the prediction of future seismic shaking. As currently applied, PSHA is largely empirical, based on parametric representations of fault rupture rates and ground motions that are adjusted to fit to the available data. The data are often very limited, especially for large-magnitude earthquakes. For example, no major earthquake (M > 7) has occurred on the San Andreas Fault in California during the post-1906 instrumental era. Consequently, the forecasting uncertainty of current PSHA models, such as the U.S. National Seismic Hazard Mapping Project (NSHMP), is very high. Reducing the uncertainty in PSHA through more accurate deterministic simulations of earthquakes at higher frequencies has many societal benefits, ranging from better safety designs to saving the costs associated with overdesign. With this in mind, the Southern California Earthquake Center is striving to extend the upper frequency limit of physics-based PSHA from 0.5 Hz, in current models, to 2 Hz in the 3-5 year timeframe and 5 Hz in the 5-10 year timeframe. This enables damage predictions to be made for most of the buildings in the Los Angeles areas (current predictions are limited to buildings of modest height). The long-term goal is to extend physics-based PSHA across the full bandwidth needed for seismic building codes; i.e.,

[a] Based on input from Tom Jordan and Phil Maechlin of SCEC.

up to 10 Hz, requiring machines of the capability of a next generation High Spectrum HPCD system.

- **Stellar Astrophysics Examples**[b]—Stellar explosions and heavy element synthesis provide insight into the creation of many fundamental elements needed for daily life, such as carbon and oxygen, which are created when stars explode. The current calculations for Hydrogen Ingestion Flashes in slowly exploding stars (e.g. Sakurai's Object) are challenging because of the need to simulate the entire central region of the star and over a very large number of dynamical time cycles. Simulations of just one sector of the star miss the global oscillation that develops, and a simulation for a shorter time would miss the enormous increase in the hydrogen ingestion rate at late times in the explosion. Simulating the nuclear reactions and heavy element synthesis in much greater detail and fidelity and simulating a larger radial volume of the star will show the boundary conditions for the event and the effects on the star's outer layers. Today, with the Blue Waters system, only a handful of such simulations in a year are feasible. Being able to investigate a wider variety of such events in stars of different masses and metal content is critical to better understanding of how heavy elements form.

  Raw computing power is necessary but not sufficient to investigate other important questions, such as stars in a galaxy, small particles in a proto-planetary disk, particles in a plasma, or photons or neutrinos in a non-opaque medium. Improved algorithms are needed, such as accurate methods up to six dimensions and time phase space fluids in order to make the simulations feasible in actual time. Today, science teams are using either simplified assumptions about the distribution function or a Monte Carlo approach by following representative particles in detail, as in simulations of the development of structure in the early universe. Treating such problems with a phase space fluid in 6-D will expand current understanding in important ways. Unfortunately, Blue Waters is only fast enough to host some of the initial developments of this new type of computation. It is not able to carry out real studies require next generation high-spectrum system.

- **Life Sciences**—With current Track-1 computational resources, breakthroughs are being made in how viruses, in particular HIV, infect cells; how antibiotic and cancer drug resistance occurs; how to efficiently create biomass fuels; and how sunlight is converted to chemical energy in photosynthetic organelles. Many of these simulations require integration of complex experimental data, but at the same time, they create insights that are unachievable by experiment alone since multiple sets of experimental data need to be interpreted. The use of multiple sets of experimental data requires computation, and critical information must come from simulations. The simulations reveal all-atom structures and often entirely unknown physical and chemical mechanisms underlying cellular processes targeted for treatment through new drugs. Best-of-breed simulations today are covering 60 million to 200 million atoms that span one microsecond of simulated time, for example in the case of HIV. While impressive, this is only a step to the ultimate goal of full atomic scale simulations of an entire cell. An organelle such as a photosynthetic chromatophore is only about .1% of even a small cell, which, in the case of human cells, typically has about 100 billion atoms and 300,000 million independent proteins. Future High Spectrum HPCD systems can make the goal of simulating full cells a reality within 10 years or less. Full cell simulation would guide unprecedented treatments and have huge potential economic benefits as they enable highly controlled synthetic biology. Similarly, the examples of best-of-breed simulations today of HIV or drug resistance can lead to new and much more effective treatments, but require many simulated "drug trials" in order to reach effective solutions. In these and many other cases, future High Spectrum HPCD level systems are essential.

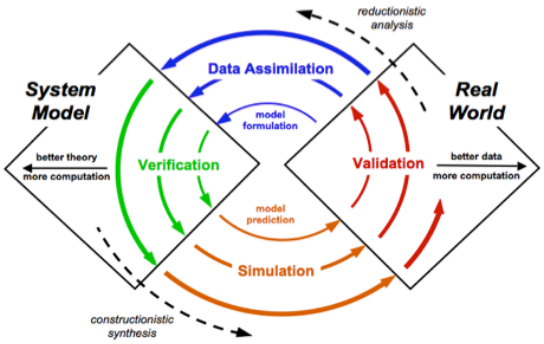
[b] Based on input from Paul Woodward of University of Minnesota

- **Multi-Messenger Astronomy/Physics—**The emerging field of multi-messenger astronomy provides an excellent example in which science needs are driving convergence of big compute and big data, as well as the need for a highly integrated environment of digital services connecting communities. A single astronomical event, such as a gamma-ray burst, generates transient electromagnetic, neutrino, and gravitational wave signals. Each is a separate "messenger" of a cosmic event, for which multiple billion-dollar class instruments are operating or are being developed by very different communities (e.g., DES and LSST in optical astronomy, ALMA in radio astronomy, LIGO in gravitational physics, and IceCube in neutrino, cosmic ray, and high-energy physics). Presently, separate, multi-petabyte data pipelines for compute and analysis are being developed, dedicated to each community's experimental instruments. These separate data pipelines often have lifetimes that are two or three times that of any High Spectrum HPCD or Track-2 resource. At the same time, a single 3D general relativistic simulation incorporating all the multi-physics and microphysics needed to compute optical, radio, neutrino, and gravitational wave signals of sufficient quality to be useful for identifying signal from noise, and/or for interpreting observations, from these systems simply cannot be done without a facility of the capability of a next generation High Spectrum HPCD system as proposed here. Hence, major High Spectrum HPCD resources, deeply embedded in an NCDI that incorporate big data services from multiple MREFC-class instruments, supported by new workflows, data analysis, and extraordinary simulation capabilities, all integrated together, will be needed to realize the science potential of all these investments. The future High Spectrum HPCD reinvestment program would be developed with all these needs—including simulation and modeling—as requirements.

The trends derived from the interviews and other sources may first be evident in *"best-of-breed"* implementations (aka breakthrough, hero, grand challenge calculations), and then the trends move into *"novel community"* practices as the best-of-breed methods, applications, and implementations are then adopted by other science teams to solve new problems, and eventually important methods become "*common community standard"* practices as they are widely used. The best-of-breed applications typically address a small number of a scientific discipline's very important yet previously computationally infeasible problems. Best-of-breed solutions are enabled by the largest computation and analysis systems available at the time—High Spectrum level HPCD systems plus the best, largest scale software. Novel community and common community standard practices follow best-of-breed implementations, applying their new developments to wider ranges of important problems.

The key trends are:

1. Increased **integration with data sources and increased use of simulation data products** will continue to expand. There are several modes of this trend:
    a. Using data from multiple experiments and observations to set up the initial problem conditions for simulation (e.g. cosmology, biology, space science, climate, materials,



engineering). For example, in computational biology, interpretation of multiple experimental inputs requires computing as the only means to determine, for example, atomic-level models of very large macromolecular systems like the capsid of the AIDS virus that are consistent with all experimental data types.

b. Observation data assimilation and/or steering during simulations to extensive post analysis and validation (e.g. weather, climate, solar physics).

c. On the other hand, many traditional best-of-breed modeling and simulation teams are realizing that using High Spectrum HPCD systems for best-of-breed problems enables them to produce community data sets that are then useful for others to analyze in different ways. Examples on current High Spectrum HPCD systems include helioscience, satellite systems, particle physics, etc.

Figure 1: The Inference Cycle of System Science - As models become more complex and new data bring in more information, we require ever increasing computational resources - Courtesy of Tom Jordon of SCEC

d. Figure 1 – courtesy of Tom Jordan shows the computation and data cycle that is increasingly typical of high spectrum science and engineering.

2. The need to dramatically **increase fidelity**[c] in models and simulations to improve insights and address new problems. Fidelity increases tend to be domain specific, but lead to more accurate predictions as well as increases in the scope of the problems that can be simulated. The means used to increase fidelity include:

   a. *Increasing Use of Multi-scale and Multi-physics*. Increasing complexity and fidelity is driven by the increasing need for multi-scale and, in many cases, multi-physics applications, which are needed to accurately explore simulated phenomena.

   b. *Increasing Resolution*. Many areas require orders of magnitude increases in resolution to provide better insights. This is realized by finer grids, more elements or atoms, more particles, etc., and by increased resolution in observation. A key example is modeling turbulence of complex flows and chemistries.

   c. *Increasing complexity*. Increased understanding in physical models and simulation studies, combined with increased details in experiments and observations, drives the development of new, more complex models and simulations with more attributes, physical sub-processes, and higher degrees of precision. Examples include the use of full

[c] In this document *fidelity* means "accuracy in details" of the science problem.

cloud models rather than parameterizations, direct turbulence in fluid calculations, and complete treatment of fluids, magnetic fields, nuclear equations of state, radiation transport for multiple particle species in relativistic astrophysics.

d. *Increased number of "ensemble"[d] trials* provides statistical or other types of information for uncertainty quantification and probability analysis. Weather predictions may have up to 50 to 75 ensemble members for a single prediction. Materials, structures, bio-physics, climate, and astrophysics are a few other areas where the use of ensembles provides added value. Note this trend does not imply smaller scale runs but rather more simultaneous runs at scale (1,000s to 10,000s nodes).

3. The need for **longer simulated time periods**. Longer simulated time periods are often required to accurately simulate the system of interest. Sometimes long simulated time periods are the result of increases in fidelity but often they are required in their own right. As an example, striving to simulate hundreds of orbits of a binary black hole – neutron star system, or to tackle a 100 billion all-atom simulation to describe the behavior of a living cell, also means that the simulated time has to increase by three to five orders of magnitude without a corresponding increase in the time to solution. Another need for longer simulated time periods results from simulations of larger systems, which often require longer periods of time to stabilize. In many problems, the timescales of natural processes are longer than current simulations, e.g., in the magnetosphere, global effects can occur on scales of days whereas kinetic simulations can only simulate several hours currently.

4. Increased **number of problems to address**. As it becomes possible for new best-of-breed simulations (e.g., all-atom MD for 100 million atoms) to study complex systems (e.g., organelle, capsids, etc.), the solution of many other important problems also becomes possible, thereby elevating this level of simulation to "community standard practice." While the first 100 million all-atom simulations were completed in 2013, by 2020 there will be tens to hundreds of teams doing hundreds to thousands of 100 million atom simulations to solve outstanding problems in biology. Similar relationships between best-of-breed and community practice levels exist in aircraft and engine design, drug discovery, galaxy formation, weather and climate prediction, materials, chemistry, and many other fields.

5. **Changing workflow methods**. Changes in computational methodologies impact the workflows of science teams. Frequently saving entire data sets becomes infeasible due to their large size. At this point, the use of *in-situ* visualization and analysis to reduce data movement and speed time to solution (e.g. severe weather, astrophysics) becomes necessary, just as it is for some forms of large-scale experiments. This trend also involves the integration of some high-throughput work to analyze and reduce large-scale simulation results (e.g. seismic, space science, materials genomics, and drug design):

    a. Support of data streaming pipelines for deadline-driven analysis for experimental and observational systems such as LSST, LIGO, and genomic sequencing. These workflow pipelines could be primary support for experimental projects or be backend expansion for projects that provision their primary resources within the projects. This area will require expanded integration of workflow and resource management methods, e.g. queued work (aka "batch"), topology-aware resource management, and OpenStack/real-time work management.

    b. Use of visualization to interpret and understand the simulation and analysis results, whether *in-situ* with simulations or after the simulation or analysis, is a critical part of at-scale workflows. Because there are multiple challenges in moving petascale data sets, in

[d] For this document *ensemble* means running the same application and basic problem but with different initial conditions and/or system parameters in order to obtain high-confidence results and provide new insights. It may also mean running the same problem with different applications. It does not mean running different problems.

many fields the simulation output and integrated observed data sets are becoming too big to move and must be analyzed in place on High Spectrum HPCD systems.

c. Malleable/elastic resource management for application load balancing and resiliency.
d. The scale and complexity of multi-stage research calculations require automation through workflows to support repeatability of the computational solution.
e. The use of data models (e.g. MapReduce, graph, NoSQL) programming methods, often combined with more traditional math model programming methods in a single application or workflow.

6. **Changing algorithmic methods**. S&E teams will substantially improve their algorithmic methods to reach new research goals over the next five to 10 years—not just to address new computer architectures, but also to improve the time to solution, independent of hardware changes, and to develop the algorithms needed for multi-physics and multi-scale simulations. This is a continuing re-engineering practice that is typically motivated by trying to use new technologies or trying to get better results in the same or less time.
    a. Examples in the past include adaptive mesh refinement and sparse linear algebra methods replacing dense LA methods. Going into the future, the majority of the S&E teams will have to change their algorithms to adjust to systems architectures that demand much more concurrency within and across nodes and much less I/O and communication bandwidth, and much less memory per core. Additionally, teams will upgrade to new algorithms and work methods to improve the quality and efficiency of their science output. Examples include Ultra Coarse Grain molecular dynamics, replacing Particle Mesh Ewald (PME) calculation with multi-level summation and higher order PME interpolation in all-atom simulations, replacing spectral methods with other methods in modeling fluids and climate/weather, using discrete event technology in space physics, or linear-scaling methods in density functional theory simulations of materials.
    b. Adaptive gridding and malleable/elastic processor management resource management for applications load balancing and resiliency. Improved load balancing is critical to overcoming both Amdahl's law limits and the increasing variation in system component performance, while resiliency is needed to address the number of single point failures in systems to millions to billions of discrete components.

Table 1 shows how these trends relate to the major science discipline areas.

Table 1: Trends Mapped to S&E Domains

Science Discipline	Workflow Changes	Data Integration and sharing	Increased Fidelity	Increased Simulated Time	Long-range investment program	Number of Simulation Problems	Algorithm re-engineering
Observational Astronomy	✔	✔	✔		✔		✔
Computational Astrophysics	✔	✔	✔	✔	✔	✔	✔
Atmospheric Science	✔	✔	✔	✔	✔	✔	✔
Computational and Analytical Biology	✔	✔	✔	✔	✔	✔	
Biophysics	✔	✔	✔	✔	✔	✔	✔
Chemistry	✔	✔	✔	✔		✔	✔
Climate	✔	✔	✔	✔	✔	✔	✔

Engineering	✔	✔	✔	✔		✔	✔
Geophysics	✔	✔	✔	✔	✔	✔	✔
Materials Science		✔	✔	✔	✔	✔	✔
Particle Physics (HE, NP, QCD)	✔	✔	✔	✔	✔	✔	✔
Social Science, GIS and, Economics	✔	✔	✔			✔	
Space-Based Earth Science		✔	✔			✔	
Space Physics	✔	✔	✔	✔	✔	✔	✔
Turbulence and Fluids		✔	✔	✔		✔	✔

C STRAW HORSE SYSTEM CAPABILITIES AND CHARACTERISTICS

The purpose of this section is to provide baseline expectations sustained performance for the computational and analytical capability and value of the systems that will be needed to address the science goals and trends discussed above. In order for S&E Teams to project potential science objectives, a series of system performance estimates were carried out by NCSA and vendor partners to estimate sustained performance of a Track-1 level system at 4 year intervals. The timing is chosen to provide a one-year overlap between each generation of High Spectrum HPCD system so the S&E community can continue to use stable, production Generation N system while the Generation N+1 system is being installed, tested, and accepted and put into early science phases. The systems are sized for an investment similar to the original NSF Track-1 program; this document does not assume they are in any particular location or facility, but rather just provides a feasible target for S&E Team planning. The details of the science goals associated with the system characteristics are below.

The alternatives presented here are developed based on vendor roadmaps. The estimates of performance are consistent with the NSF Track-1 program key metric for the high-spectrum follow-on HPCD systems are *sustained performance* for a wide range of science and engineering applications. Here sustained performance is defined as time to solution for real science, engineering, and research problems. The optimization target that represents sustained performance in a meaningful manner for evaluation is the Sustained Petascale Performance (SPP)[e] Metric developed as part of the current Track-1 acceptance process. Without getting into too much detail, the nearer term performance projections were based on real full application benchmarks benchmark results and extrapolations. Longer term performance projections were based on general technology trends and estimates. For simplicity and timeliness, the system configurations are not inclusive of all possible alternatives and architectures, and certainly there will be other estimates that are more or less optimistic.

[e] William Kramer, "How to Measure Useful, Sustained Performance." *ACM/IEEE SC11 Conference.* Seattle, WA: ACM/IEEE, November 12-18, 2011.

C.1 System Alternatives

C.1.1 2016/2017 Period

Table 2: Projected System Configuration and Sustained Performance for a $200M (FY 2014 dollars) purchase cost system in late 2016/early 2017

System Type	Peak (PF)	Example Nodes	Example Interconnect	Aggregate Memory (PB)	Estimated Running Power (MW)	BW SPP Estimate (PF)	On-line Storage Capacity (PB)
Reference: current Track-1	13.1	27,648	Cray Gemini	1.66	9-11	1.3	36
1. X86 General Purpose CPU (Based on Intel Skylake Processor[f])	~58	~19,200 100 racks	Cray Aries[g] or Intel Stormlake 1	3.5	9-10	7.8	200
3. Intel Many Core (Based on Intel's Knight Landing Processor[h])		26,500 136 racks	Cray Aries or Intel Stormlake1	2.3	8-10	6-10	200
4. X86 CPU with NVIDIA GPU (Based on Intel Skylake Processor with NVIDIA Pascal[i] GPU)		21,000 113 Racks	Cray Aries	1.2	9-11	7-11	200

In this timeframe, several new architectural features begin to appear in systems that together introduce more hierarchical levels for memory and storage. These features will enable the use of new methods in existing applications and enable new applications to make effective use of the systems. Examples are:

- The introduction of non-volatile RAM in nodes, which can be used to create very large memory (1-2 TB) nodes. Alternatively NVRAM can be used to accelerate I/O and checkpointing. Large memory nodes are expected to play a key role in enabling "Big Data" and data analytical applications.
- Advanced interconnects such as the Cray Aries Dragonfly will increase the performance of many applications as well as the I/O performance. For example, it is estimated that the

[f] The Skylake processor is expected to have six memory channels capable of supporting 16 GB (and later 32 GB) of DDR4 DIMM at 2.656 GHz, with a total of 12 DIMMS per node. The processors are estimated to have a peak performance of 1.1 TF/socket and a node has two sockets and a memory bandwidth of 16 Gbps. The Intel roadmap also discusses two additional DIMM slots that would be capable of holding ½ TB (and later 1 TB) of NVRAM.

[g] An Aries interconnect based on Dragonfly network topology provides three types of network links: "rank-1" links connecting all 16 of the Aries interconnect chips in a chassis via the backplane in an all-to-all fashion; "rank-2" links connecting all six chassis in the two-cabinet group via copper cables again in an all-to-all fashion; and "rank-3" links connecting all groups in the system via optical cables (not used in this system). Consequently, the interconnect provides very effective routing and low contention switching. Processors communicate with the Aries interconnect chips via PCIe Gen3 x16 links at 8 GT/sec for 16 GB/s of raw bandwidth per direction.

[h] Knights Landing processors are expected to contain up to 72 of these cores, with AVX, with double-precision peak performance exceeding 3 TFLOPs. Memory is expected to be Intel/Micron Hybrid Memory Cube (HMC) technology. See http://www.anandtech.com/show/8217/intels-knights-landing-coprocessor-detailed.

[i] NVIDIA's next-generation Pascal GPUs will be used in the final configuration of Nexus. Pascal-Solo GPUs will deliver at least 3 TF per GPU in double-precision floating-point arithmetic and will provide at least 12 GB of memory per GPU, with peak memory bandwidth of 750 GB/s.

communications-bound DNS Fluid Turbulence code running on Blue Waters will have a 10x improvement in time to solution on an Aries system with 1/10th the number of nodes. These advanced interconnects will also improve runtime consistency and reduce I/O congestion and will be important as data volumes continue to increase at a fast rate.

- The introduction of storage "burst buffers" will improve I/O performance. Many applications now experience significant I/O bottlenecks. Burst buffers can be used to improve time to solution for both output-based simulations and input-based analysis codes. They help address small I/O, less efficient access patterns, and rapid checkpointing, as a few examples.
- Introduction of additional tiers in the storage hierarchy along with improved storage management tools to automate many of the tasks S&E teams now have to do to move data.

C.1.2 2020/2021

Table 3: Projected System Configuration and Sustained Performance for a $200M (FY 2014 dollars) purchase cost system in late 2020/2021

System Type	Peak (PF)	Example Nodes	Example Interconnect	Aggregate Memory (PB)	Estimated Running Power (MW)	BW SPP Estimate (PF)	Online Storage Capacity (PB)
5. General Purpose CPU System[j]	200	~30,000 110 racks	Intel Stormlake 2 or other interconnect	~10	12-14	40-50	400
6. Accelerated, Many Core System[k]	500	~30,000	Intel Stormlake 2 or other interconnect	~4.0-10	8-10	40-50	400

C.1.3 2024/2025

Table 4: Projected System Configuration and Sustained Performance for a $200M (FY 2014 dollars) purchase cost system in late 2020/2021

System Type	Peak (PF)	Nodes	Interconnect	Aggregate Memory (PB)	Estimated Running Power (MW)	BW SPP Estimate (PF) BW 1.0 SPP = 1.3 PF	Online Storage Capacity (PB)
7. Accelerated, Many Core System[l]	1,200	~30,000	TBD	~15-30	15-20	100-200	1,000

C.2 Service Architecture

The service architecture for future High Spectrum HPCD systems will be more complex and diverse but will be expected to be as efficient as today and must support the largest best-of-breed challenges of the future. Specifically, the future High Spectrum HPCD service architecture will:

- Support a wider range of workflows from traditional, very large parallel jobs to many high throughput, but interconnected jobs. Resource management tools and methods have to greatly improve in order to satisfy this range of services while having sufficient utilization and responsiveness.

[j] The Intel x86 and ARM 64 processors will be competitive and approximately equivalent with different insertion timings. Hence, we do not distinguish which processor type is used for these estimates.

[k] This system is estimated using roadmaps for Intel Knights and NVIDIA augmented by general industry trends.

[l] This system is estimated based on general industry trends.

- A wider range of application frameworks, from tightly synchronized MPI codes to MapReduce style codes.
- A more varied set of I/O and data storage characteristics.
- Better integration with existing and future cyberinfrastructure capabilities.
- The ability to meet the deadline-based, data pipe requirements of large MREFC experimental facilities while servicing other workflows. This may include meeting high availability requirements so experimental data is not lost.
- Improved methods of data stewardship for both reliability and correctness. Also, a more robust set of data sharing and preservation services will be required.
- The ability to make it easier for leading research teams to gain access to the system and become productive. High Spectrum HPCD and Moderate Spectrum (aka Track 2) systems have to provide a lightweight but effective allocation process, while assuring the unique attributes of High Spectrum HPCD systems are used appropriately.
- Improved data management, cataloguing, and discovery services.
- Increased variety of programming and runtime environments.
- The ability to provide software-defined systems, software-defined networking, and software as a service will have to be features of future High Spectrum HPCD system architectures and software environments.