Joint Agency Assessment of the Responses to the Request for Information (RFI) on "Science Drivers Requiring Capable Exascale High Performance Computing"

The National Institutes of Health, the Department of Energy, and the National Science Foundation[1]

February 13, 2017

Executive Summary

This report provides a joint assessment by the National Institutes of Health (NIH), Department of Energy (DOE) and National Science Foundation (NSF) of the scientific and engineering research community's science drivers requiring capable exascale high performance computing (HPC), based on analysis of a total of 246 responses to a joint NIH-DOE-NSF Request for Information (joint RFI); an RFI to the DOE National Laboratories (DOE RFI); and an RFI to the NIH Intramural Research Principles, all conducted respectively in the fall and summer of 2015. Each of the three agencies analyzed the aggregate responses in the context of their missions and programmatic goals. This report comprises a joint collaborative synthesis of these respective analyses and summary conclusions. Individual agency perspectives are provided in Appendix 1. Text of the RFIs are provided in Appendix 2.

From the perspective of science opportunities, the responses describe a broad and exciting array of applications in physics, biological sciences, health sciences, materials sciences, geosciences, planetary sciences, chemical sciences, engineering and energy technology, mathematics, computer science, and information science. Together, the responses expand the range of potential applications and impacts from traditional large-scale computational areas such as cosmological simulations and climate modeling to emerging areas such as health science, biomedical modeling, and full-scale wind power plant simulation. Respondents across disciplines noted that advances in high performance computational and data science will considerably enhance our understanding and predictive capabilities of complex phenomena.

From the perspective of technology needs, the HPC applications described in the RFI responses spanned a broad range of computing approaches in three main application domains:

- Modeling and simulation: *Enabling vast improvements in spatial and temporal realism and associated predictive accuracy*;
- Data-intensive science: *Enabling analysis and visualization of multi-source and multi-scale data at unprecedented scales*; and
- On-demand and real-time computing*: Enabling real-time analysis of simulations, data-intensive experiments and streaming observations*.

The nature of the stated needs varied significantly even within disciplines and sub-disciplines, reflecting the strong linkage of specific computational requirements to specific project objectives. Responses

[1] This assessment report was developed by a multi-agency RFI team including Barbara Helland and Carolyn Lauzon (DOE), John Russell (DOE-AAAS), William Miller (NSF), and Peter Lyster and Susan Gregurick (NIH).

pointed to the need for a dynamic and agile HPC ecosystem to accommodate a diversity of requirements for HPC capabilities and resources across the spectrum of science and engineering research. Cited requirements include development of new HPC architectures that can handle a heterogeneous range of applications; new application and system software, and new algorithmic methods; as well as new efforts to validate the applications at the expanded scale.

Moreover, from the perspective of community needs, some fields are highly data driven and therefore require computing capabilities that vary significantly from traditionally numeric application-oriented high-end computers, while others require introduction of parallelism into existing algorithms and revising applications to be scalable to higher levels of machine performance. Overall, it was clear from the RFI responses that major efforts will be required to develop computational methods, software, and workflow technologies across many disciplines to take full advantage of significantly increased HPC performance.

Finally, many respondents pointed to a general lack of requisite familiarity, knowledge, and skills on the part of the scientific workforce in computational sciences, computer science, HPC technologies, and software development methods. A particular concern was a large knowledge gap between the domain scientists and experts in high-end parallel computing. These issues suggest a critical need for *developing the future HPC workforce* including interventions in educational curricula, training, and other workforce development, as well as advances in workflow systems, more accessible means of HPC availability and usage, and to maximize productive application of exascale computing in science, technology, engineering and mathematics.

I. Context of the Request for Information

The United States is a leader in the development and deployment of HPC systems. These computing systems are essential to U.S. scientific leadership, economic competitiveness, and national security. The performance of HPC systems is traditionally characterized by their capability to sustain rates of calculation measured in “floating point operations per second” or “flops”. Current advanced HPC systems can sustain calculations in the tens of petaflop range (petascale systems) while future HPC systems are currently being targeted to sustain exascale performance, or approximately 100 times the performance of current 10-petaflop systems. Scientists and engineers use the superior calculation power of HPC systems to perform research and make discoveries on some of the most complicated and challenging research problems.

HPC systems are complex and require an ecosystem of technology, tools, and expertise to be utilized productively. Development of capable exascale systems is anticipated to require a range of technical innovations, not only for the HPC systems themselves but for the whole HPC ecosystem comprising data, storage, communication, operational and analytical software, and accessibility resources and technologies, to achieve maximal end-to-end usability and performance of such systems and the associated computational environment.

In support of efforts to develop an interagency common understanding of the science drivers, requirements, applications, and future use-models that will be advanced by exascale resources, in the fall of 2015, NSF, DOE, and NIH published a Joint RFI under public notice NOT-GM-15-122[2]. In the summer of 2015, DOE issued a similar request for information to the DOE National Laboratories (DOE RFI); and similarly, in the winter of 2015, a request for information was issued to the NIH Intramural Research Principles community. The NIH Intramural RFI responses were included as part of the analysis of the joint RFI. Both RFIs solicited community input identifying scientific and engineering research that would benefit from a greatly enhanced next generation HPC ecosystem far beyond what can be done using current technologies and architectures. Appendix 2 provides the texts of the RFIs.

The collective response will inform NIH, DOE, and NSF planning to achieve objectives for advancing performance of the HPC ecosystem to support scientific research, and guide the research, engineering, and development process. It is likely that a range of advanced capabilities will need to be developed to respond to the varied computing needs across science disciplines. This report documents the results of the collective RFIs, provides a summary of the public response, and highlights key issues identified by the public that will inform both decision-makers and the planning process. Conclusions of this joint agency assessment are summarized in the final section; individual agency perspectives are provided in Appendix 1.

A. Description of Responses

The interagency joint RFI generated 113 responses including a group response from the NIH Intramural program[3] with an additional 133 responses from the DOE RFI to the DOE National Laboratories. Respondents included individuals and groups from academic and national laboratories, industry stakeholders and non-profit entities. Responses ranged from submissions from individual Principal Investigators (PIs) with research goals focused on a single sub-discipline to institutional responses submitted on behalf of multiple PIs conducting research in different research domains. Even with the domain diversity found within individual responses, an approximate breakdown of science categories was useful in gaining high-level insights into the responses.

The science domains represented by the responses are summarized below. Table 1 lists the primary research domains responding, sub-disciplines responding that fall into each domain, and a percentage of total responding by domain. To understand the domains represented in the Joint RFI and the DOE RFI, respectively, Chart 1 below presents the domain breakdown by each RFI.

[2] RFI link: http://grants.nih.gov/grants/guide/notice-files/NOT-GM-15-122.html and http://grants.nih.gov/grants/guide/notice-files/NOT-GM-15-123.html.

[3] The NIH Intramural research RFI resulted in a group submission comprising 27 responses from 9 NIH Institutes. This group submission was included in the set of Joint RFI responses to facilitate analysis.

Primary Domain	Discipline Responding	Percent of Total
Math, CS, Information Science	Applied Mathematics, Cybersecurity, Software Ecosystem, Data Science	6 %
Physics	Particle Physics, Space Physics, Plasma and Fusion Energy, Nuclear Physics, Fluid Dynamics	20 %
Chemical Sciences	Catalysis, Combustion and Energetics, Photo and electro chemistry, Heavy element chemistry, Quantum Chemistry,	5 %
Geo and Planetary Sciences	Solid Earth Science, Metrology, Seismology, Subsurface, Climate Science	10 %
Biological Sciences	Molecular and Cell Biology, Biophysics, Neuroscience, Omics, Systems Biology	23 %
Health Sciences	Precision Medicine, Clinical Medicine, Population Health	11 %
Materials Sciences	Condensed Matter, Superconducting Materials, Electronic Properties, Soft Matter and Polymer Physics, Materials Genome	14 %
Engineering and Energy Technology	Bioengineering, Chemical Engineering, Aerospace Engineering, Energy Storage, Electric Smart Grid, Renewable Energy Technologies, Vehicle and Combustion Engine Technologies	11 %

Table 1. *Science domains and sub-disciplines represented in the aggregated (joint and DOE) RFI responses. A variety of scientific communities responded to the RFI, with each identifying numerous scientific opportunities that would be enhanced by an exascale HPC platform. Areas of interest include, but are not limited to, astronomy, astrophysics, and cosmology; geosciences and climate sciences; materials and chemical sciences; wind energy technology, vehicles technology and safety; nuclear engineering; biology, biophysics and neuroscience; and population science and precision medicine.*

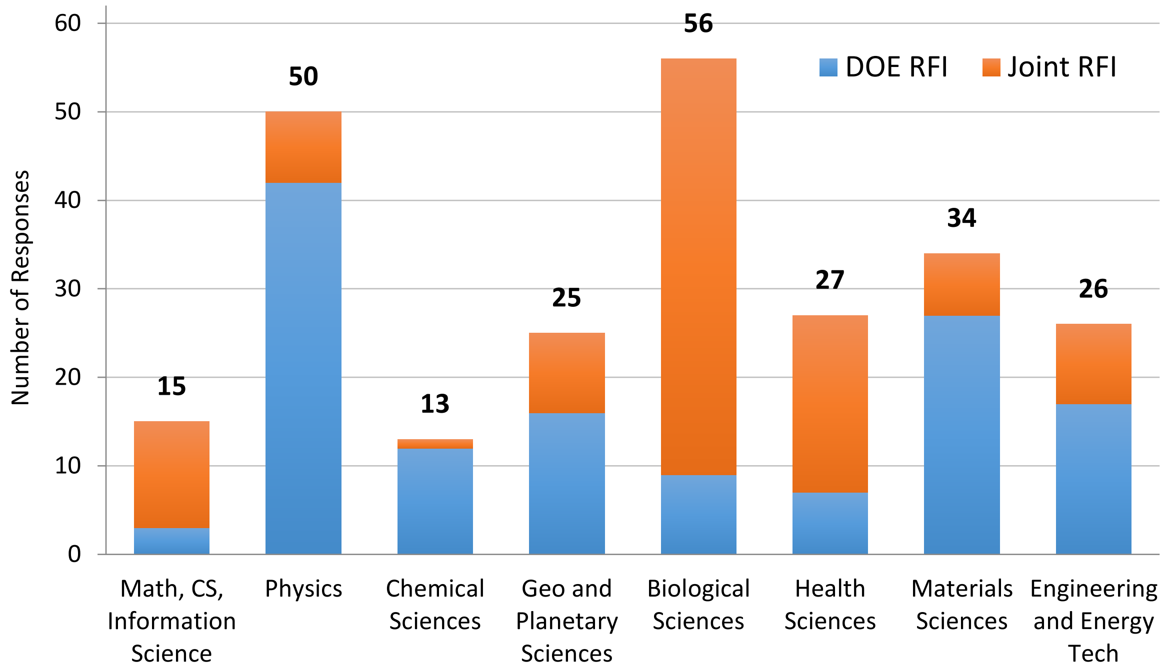

Chart 1*. Number of responses by science domain for the DOE RFI and the joint RFI.*

B. Gaps in Responses

Coverage was uneven across research disciplines, with physics and life sciences (biological and health sciences) representing the largest number of respondents. Physics disciplines were strongly represented in responses to the DOE RFI and are likely a reflection of this community's long history with HPC, as well as DOE's strong research and development (R&D) efforts in the physical sciences. The predominant representation from the health and biological sciences in the Joint RFI responses could be attributed to NIH issuing the Joint RFI but may also indicate a growing interest in the biomedical community to exploit new possibilities offered by advanced computing.

Although the two RFIs provided some complementary data, it is worth noting that several well-established computational fields, such as applied mathematics, computational chemistry, and combustion science, were not highly represented in the responses. Similarly, important emerging fields such as ecological, ocean, and polar sciences were largely absent from the responses. Industry represented only 4% of the responses (the Joint RFI was open to industry but did not target this sector) and, although a number of responses from different engineering fields were received, many engineering sub-fields such as aerospace engineering, advanced manufacturing, and vehicles technology design were likely underrepresented.[4] Additional opportunities for receiving community input and other outreach efforts by Federal agencies and stakeholders are necessary to ensure that agencies achieve a full and balanced understanding of the diversity of needs and scientific opportunities for an exascale ecosystem.

[4] The classified research community was out of scope for this RFI call and would likely include some areas in nuclear weapons programs, stockpile stewardship, national security or other classified areas in mathematics and engineering.

III. Analysis and Synthesis

A. Computing Approaches and Science Drivers

The RFI responses comprised a broad spectrum of needs for a transformative HPC ecosystem designed for scientific and engineering research. Three major categories of use cases were identified in the responses: modeling and simulation, data intensive science, and on-demand and real-time computing. Depending on the area of research, individual responses could belong to one or more of these categories. These categories are discussed further below.

1. Modeling and Simulation

Simulation science is a mainstay of HPC and is frequently used to test theories, explain and guide experiments, and study systems that are difficult or impossible to observe experimentally. Generally, RFI respondents in the modeling and simulation space described the need for greatly increased processing power to extend the length, time or parameter spaces of their current simulation efforts; capture new and critical physics in their models; and piece together multi-component multi-physics systems for whole integrated simulation efforts, as described below.

Extended Length, Time, and Parameter Spaces

RFI respondents identified the need for greater computational power to extend their current simulation efforts into longer timescales, greater lengths, and increased statistical sampling. Many respondents indicated a need for these extended regimes in order to capture critical physics. For example, to capture larger molecular complexes relevant to molecular biology, chemistry and materials science, researchers require simulation resources capable of simulating a significantly greater number of atoms than currently possible (Figure 1). Similarly, greater computational power will allow researchers to push the time duration of molecular dynamics simulations from nanoseconds into milliseconds. This time extension, by six orders of magnitude, is important for areas like protein folding or materials modeling.

A large segment of respondents needed large-scale resources to search vast parameter spaces in fine detail to discover new solutions or optimize current solutions for design problems. For example, materials genomics researchers need to simulate a greater number of configurations and elements to identify new materials with new properties. Drug discovery research will use parameter space searches to find new cures and identify the safest, most promising drug candidates. From industry, vehicles technology engineering will need large simulation resources for parameter space exploration to optimize vehicle design for durability, manufacturability, and safety. The connection of this domain of computing and new architectures is an area that needs further analysis.

Respondents also emphasized the need for greater statistical sampling for error reduction and improved uncertainty quantification (UQ). UQ is a technique that analyzes the accuracy of simulation predictions. For instance, efforts in nuclear physics to resolve the phase diagram of quark-gluon plasmas (QGP) using Monte Carlo techniques are hindered by the limitations of current computational systems to produce the needed statistical accuracy to capture realistic quark mass near the phase boundaries. QGP simulations are also crucial for interpretation and guidance of large-scale experiments, such as the Relativistic Heavy Ion Collider at Brookhaven National Laboratory. Respondents identified UQ as critical to connect simulated results to real-life decisions in areas as varied as combustion science for gas turbines, internal combustion engine design, predictive computational medicine and surgical diagnosis outcomes, and estimating sea level rise. UQ tended to be least cited in the biomedical and life science domains, possibly due to less awareness of this technique in those fields.

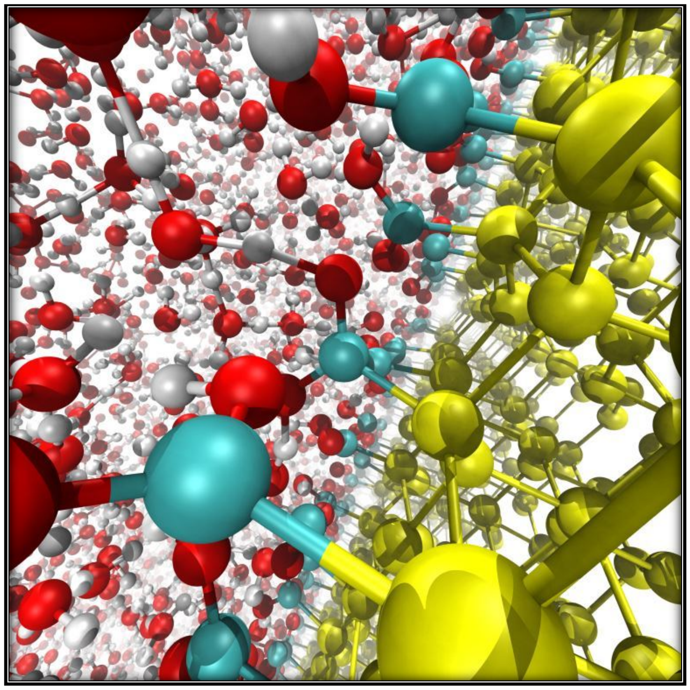

Figure 1. Si-COOH/water interface. *Exascale resources will allow unprecedented large-scale first-principle simulations on the atomic scale. With a greater number of atoms and longer time scales, exascale resources will substantially bolster scientists' ability to predict and engineer the properties and functions of materials. Image courtesy Giulia Galli and Marco Govoni (Argonne National Laboratory and University of Chicago).*

Finer Space Resolution and Increased Physical Accuracy

While many respondents expressed the need for large computational resources to extend their existing simulations, other respondents identified exascale computing as enabling finer resolution and increased physical accuracy of their simulations. In whole-cell modeling, for example, respondents identified exascale as enabling atomic-scale resolution of the full living cell. Capturing physical behavior at the atomic scale will dramatically improve our understanding of complex, cell-based biological processes with impacts in engineering, health and energy technologies. Like many respondents who fit into this category, the ability to incorporate smaller-scale physics into their simulations will require application codes with new mathematics, physics, and chemistry. Another example is in plasma and fusion science. In fusion plasmas, ions and electrons interact and together impact the behavior of the larger plasma system. Yet electron and ion scales are three orders of magnitude apart and computationally costly to couple and simulate together. Respondents are looking to exascale resources to couple physics on the electron scale with physics on the ion scale to more accurately match experimental results to models in order for simulations to accurately predict plasma behavior and control fusion energy devices. Other research domains where respondents expressed a need for greater fidelity and resolution in their models include combustion; nuclear reactor simulations; and geoscience, including geo-tectonics (e.g. Figure 2), climate modeling and weather prediction, earth system models, and environmental science (e.g., the study of watersheds).

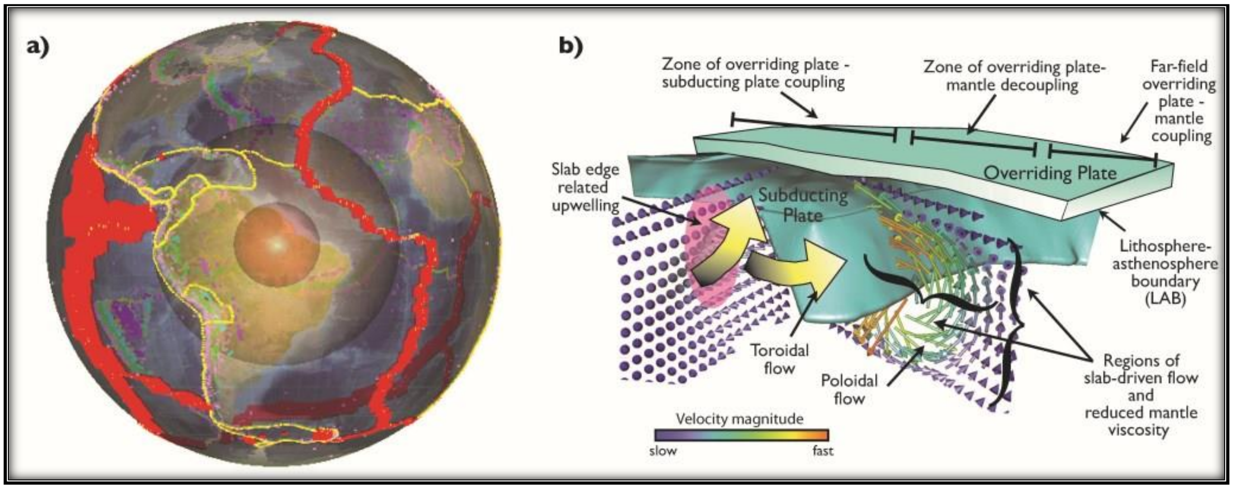

Figure 2. Multiscale global and regional tectonics. *Large-scale computational resources will allow geophysicists to model and simulate plate tectonics at higher-resolutions and increased physical accuracy. (a) Observational constraints on the nature of tectonic plates, overlain on the major structural boundaries within the Earth. (b) Results from a high-resolution 3D regional model of a single plate boundary show small-scale convection. Convection on Earth contains small-scale, highly non-linear processes shown in (b) embedded within the larger-scale, whole mantle convection that occurs in (a). Image courtesy of M. A. Jadamec (University of Houston); modified from Jadamec, J. Geodynamics v. 100:51-70, 2016.*

Systems Modeling

RFI respondents identified computing at the exascale as a regime that would allow connection between multi-component, multi-scale systems and enable unprecedented simulated experiments in areas like cosmology, climate, biology, engineering and health. In many of these fields, the behavior of the whole system cannot be understood by the study of individual parts, making full systems modeling and simulation critical for understanding and prediction. The computational expense of simulating multi-component systems drives the requirement for large computational resources. For example, respondents from systems biology pointed to the simulation of an ensemble of billions of cells, each with their own internal molecular systems, that together make up tissues. Understanding the underlying dynamics of cellular behavior, as a collective, would enhance our ability to, for example, simulate normal and abnormal processes in the human heart and other vital organs. In other areas, respondents were looking to incorporate new physical models to more accurately model the system of study. In wind farm simulations, for instance, the energy production from an individual wind turbine is dependent on the interactions with its neighbors. A predictive modeling capability for wind farms will not be achieved from modeling individual wind turbines; the full wind farm must be simulated (Figure

3). In addition, researchers in wind energy are looking to exascale resources to allow for the capture of atmospheric scale science (tens of kilometers) to model weather impacts on energy harvesting down to a single turbine rotor blade (meters) and couple the results to the full wind farm system model. Other areas where respondents identified exascale systems as putting full systems modeling in reach include wind tunnel experiments for aerospace design, asteroid deflection for planetary defense, whole-device modeling for fusion energy sciences, and coupled Earth-climate models.

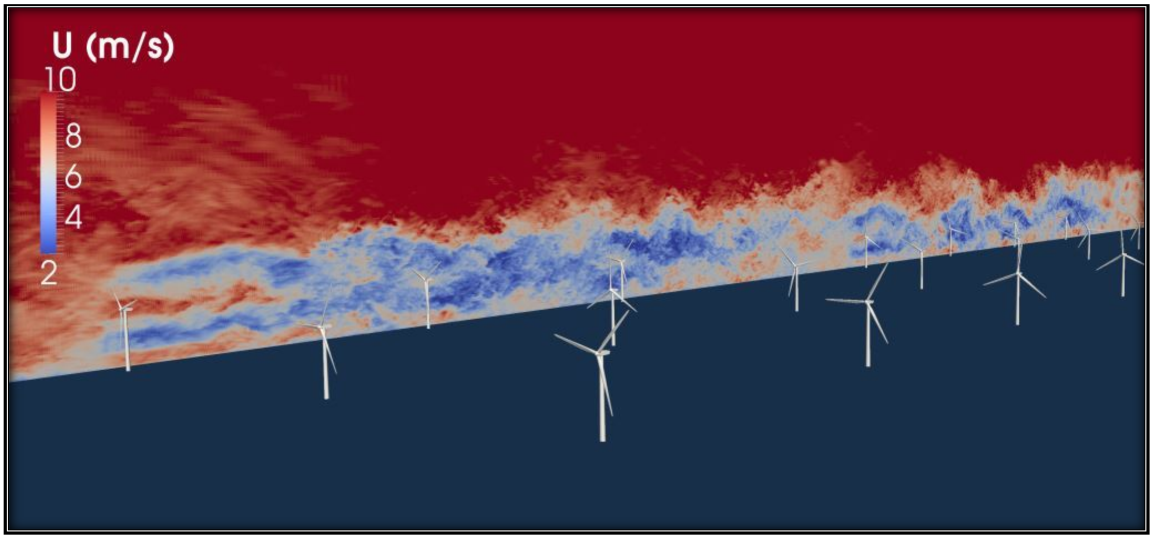

Figure 3. Systems level modeling of a wind farm*. Energy production from individual wind turbines depends on plant-level dynamics. Predictive, physics-based, high-fidelity modeling of wind farms at the full-plant level is critical to accurately predict and optimize wind farm performance. Exascale resources will allow for dramatic improvements in wind farm systems modeling and predictability and facilitate wide-scale development of cost-competitive wind energy facilities. Courtesy of Matt Churchfield, National Renewable Energy Laboratory, Golden CO.*

2. Data-Intensive Science

Respondents from many fields found that the processing of data, either through fast computation or through new experimental methods, frequently exceeds the capability of current analytical and archival processes. Responses to the RFI reflected the growing demand for greater capabilities in data analytics, visualization, and automation in data quality operations. Low-cost sensors and higher-intensity detectors are spurring data-intensive computing needs in fields including geosciences and environmental science, climate and weather studies, and urban network analysis. In many cases, respondents connected computationally-enabled fundamental research in these and other areas to policy and decision-making of societal relevance.

Machine learning is a growing field in data-intensive science, and some respondents expressed interest in data analytics for the purposes of machine learning to improve model parameters and guide simulations. For example, machine learning can facilitate parameter optimization in turbulence models for engineering design applications or integrating *in situ* machine learning into molecular dynamics simulations for materials or biological science.

Interestingly, a predominant number of biomedical researchers who responded to the RFI noted that large-scale biomedical data integration and advanced data analytics represent a significant bottleneck for advancing research in medicine. With the Precision Medicine Initiative accelerating research in genomics, whole-genome sequencing combined with variant discovery is likely to become a mainstream diagnostic tool. The desire to combine genome data and patient data that are collected over long periods of time will create an unprecedented amount of patient-centric data. The ability to infer correlations between patients, or within one individual over time, appears to be a grand challenge that will require exascale computing. A number of researchers noted that new algorithms, such as deep learning, have the ability to create actionable knowledge from these types of large and integrated datasets. The need to accomplish large-scale data assimilation and integration was also mentioned in many responses in other fields, particularly in the geosciences domain.

3. On-Demand and Real-Time Computing

A significant number of responders pointed to the need for on-demand and real-time computing such as where experimental facilities may be linked to high performance computers via fast networks to enable real-time adjustments to active experiments. Several responders also noted that the ability to create real-time workflows to analyze and visualize high-volume data is an exciting computational challenge. Examples from responders included real-time, image-guided radiation therapy, real-time infectious disease modeling, streaming real-time urban systems data to predict movement of people and vehicles, and real-time power grid simulations. One example included new advances in cryo-electron microscopy (cryo-EM) which will allow structural biology researchers to determine the three-dimensional, atomic-scale structure of large macromolecules; however, achieving this goal depends on having the ability to process the terabyte-size datasets. Development of future advanced computing resources extending to the exascale level will consequently need to include an ability to support on-demand/real-time computing, with special demands on storage capacity, computational speed, and software applications, and handling unexpected workloads.

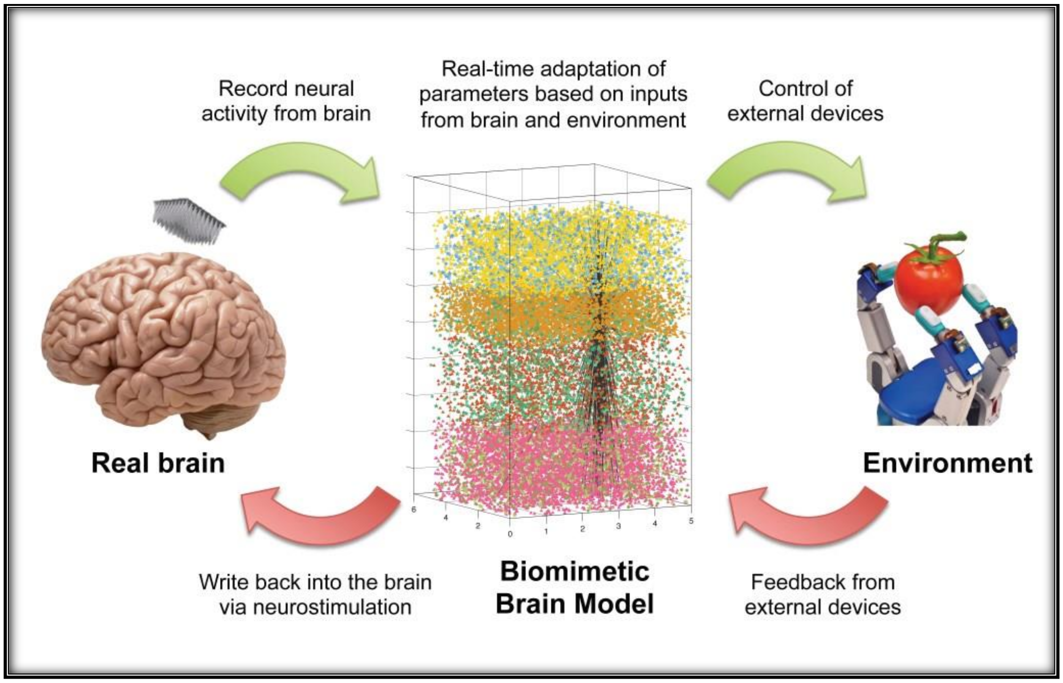

***Figure 4: Overview of neuroprosthetic data-driven model**. Coupling on-demand computing to neuroscience experiments could provide real-time experimental feedback and enable new bioengineering possibilities. Sensory information from the environment and recorded neural activity from the brain modulate the biomimetic brain model, which learns to control external devices, such as a robotic arm using biological learning rules; the external devices then affect the environment providing sensory feedback to the biomimetic model. To close the loop, neurostimulation could be fed back into the brain based on the biomimetic model's activity. This dynamic loop operates in real time. Courtesy of Dr. Salavador Dura-Bernal and Dr. Samuel Neymotin, SUNY Downstate Medical Center, Brooklyn NY.*

Several respondents identified real-time computing as enabling new types of research and engineering possibilities. Respondents noted new architecture and computing systems to collect experimental data from wearable or implanted devices, and the ability to process these data in real time would enhance scientific findings and predictability. Real-time processing coupled to neuroscience experiments could enable the bio-engineering of neuroprosthetics, i.e., prosthetics that decode and respond to brain signals (Figure 4). Another emerging area of research is the optimization of prosthetics and implantable devices that use computational simulations to pinpoint mechanical stresses *in situ*. This requires coupling of medical imaging with high-throughput computer simulations and rapid prototyping of devices that are then surgically implanted.

B. Challenges

1. Diversity of Science Applications

The readiness of different research fields to develop and utilize applications in a capable exascale computing environment appears to [or can] vary depending on prior experience with HPC capabilities. Responses from established High Performance Computational fields, such as physics and chemistry, specifically cited the need for sustained application performance levels 100 times greater than today's capabilities. A number of responders included detailed visions and architecture requirements based on current work with highly-developed, theory-based numerical modeling and simulation applications. Interestingly, simulations in the traditional HPC scientific fields were anticipated to grow in complexity and scale as well as enable comparisons against large experimental datasets.

Many other responders in fields with less established HPC usage also called for more computational capabilities, but were not as specific as to capacity or architecture requirements. Such responses were often associated with disciplines whose computational components are just now emerging, and that are being challenged by data inundation rather than numerics-intensive computational needs. Indeed, it was clear from such responses that the emergence of a plethora of experimental measurement and imaging techniques that extend to the nano, atomic, and quantal scales have made these disciplines vastly more data-intensive in current and anticipated science, with computing needs firmly oriented towards analytics capabilities integrated tightly with data management. A good portion of the data-intensive responses were focused on societal benefits of high-end computing, e.g., materials design for biomedical and energy applications, real-time patient-specific diagnostics and -omics, drug discovery, and biomedical imaging.

2. Architecture

The topic of computational architecture was a major area of interest in the RFI responses, with many communities providing different ideas on what an architecture might look like for exascale computing and how architectural decisions might impact their scientific research. There was broad agreement that the slowing or end of Moore's law and the hardware challenges of increased parallelism and power consumption would profoundly impact computational research. In the responses, discussion mainly focused on three topic areas: compute, memory, and data input/output (I/O). In the area of compute, some responses indicated that continued improvements in accelerators, such as graphical processing units (GPUs), could extend simulations with respect to length and timescale, as in models of biomolecules for drugs or for protein folding. Other respondents wrote of the difficulty and cost of having to adapt existing software code in order to take advantage of improvements in accelerators. With respect to memory, several responses argued that large memory architectures would be necessary for an exascale machine to be useful to them.

Respondents noted that the convergence of exascale simulations and exascale data analytics will require flexible systems with large memory, large I/O and advanced networking capabilities. Responses from data-driven communities, such as bioinformatics, observed that the speed of file I/O is not keeping pace with data production, and that new innovations such as nonvolatile memory are needed. Compression and data movement was seen as another challenge, as data will become increasingly conjoined with simulations in the near future. Overall, the responses indicated a variety of different architectural needs and proposed solutions. Respondents expressed a common concern for the increasing complexity of future architectures and the possibility of poor programmability of these systems.

3. Coherence of Big Data and Big Compute

An emerging feature of the scientific enterprise is the growing convergence of compute-intensive and data-intensive computing needs—which is an important consideration for the future exascale computing ecosystem. The RFI responses supported the idea that data production is and will continue to substantially increase in volume across experiments and fields such as synchrotron light sources, particle accelerators, ensemble simulations, genomics, medical imaging, and health science. Respondents in these communities stressed the need for an improved data storage ecosystem and improved networks. For some respondents, colocation of compute resources and data facilities is increasingly important due to the cost and latency of moving large data sets through existing networks. Other respondents, such as those from the climate or medical sciences, are looking to integrate and analyze data collected and stored in geographically dispersed databanks into single large-scale studies. In medical diagnostics, for example, respondents envision a future where large medical databases coupled with patient clinical data can inform patient-specific diagnostics and treatments (Figure 5). These respondents, rather than identifying a need for colocation, identified requirements for new networking and data management technologies that make the discoverability and usability of data independent of its geographical location. Across both types of use models, respondents indicated that I/O and file system considerations are limiting factors, where spinning mechanical disks are no longer keeping up with storage requirements, and new technologies such as Nonvolatile RAM are now being considered. In addition, new challenges are emerging for researchers in the areas of data provenance, security, privacy, and curation.

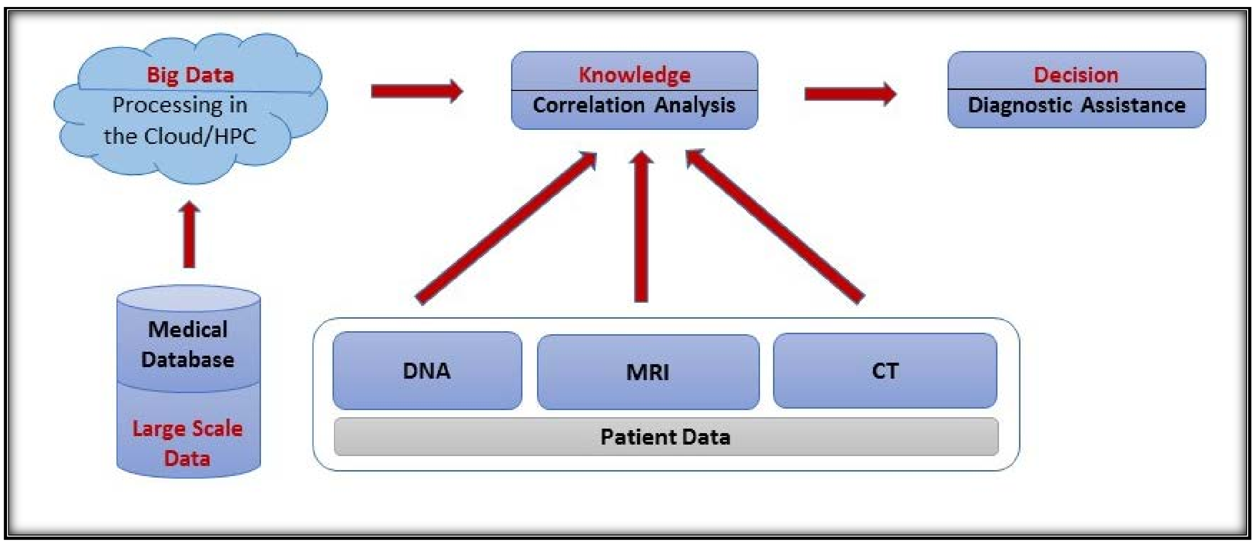

***Figure 5: Computer-Assisted Diagnostics: from Data to Knowledge to Decision**. Advanced computation and data technologies could provide opportunities for correlating single patient data with large-scale medical databases for computer-assisted diagnostic purposes. The above scheme represents an envisioning of an environment where medical personnel may consult computer-generated correlation statistics to guide diagnostics and treatment decisions. Courtesy of Dr. Frank Mueller, Department of Computer Science, North Carolina State University.*

4. Workflow and Visualization

Workflow was an important topic, well-represented in the RFI responses. However, the concept of what a workflow means in the context of exascale computing varied in the responses received, and can be grouped into three general categories. The first concept was that while today's applications and computational projects are typically developed and managed by individuals or small teams, the complexity of the exascale environment will require a more elaborate, labor-intensive workflow, i.e., a division of labor among domain, application, and computer scientists to efficiently accomplish a given effort. This divide-and-conquer approach implies a need for tools for domain scientists that enable high-level abstraction of desired computational problem design, while application and platform experts focus on platform-dependent code implementation and optimization.

A second view of workflow involved the anticipated need to introduce automation in analysis and visualization into the exascale environment to achieve efficient management of "ensemble simulations", i.e., multiple simulation runs to capture rare events and produce robust, statistically sound models. Ensemble simulations are an essential approach in many science domains from the geosciences, including climate modeling, weather, and earthquake prediction, to nanoscience, such as biomolecular models for drug discovery. The concern is that exascale high-throughput ensemble simulations will generate information beyond the human capacity to analyze, necessitating automated workflow capabilities and user interfaces to accomplish the computational goals. Some respondents recommended using machine learning for advanced data analysis and visualization.

Finally, a third concept of workflow focused on the anticipated tools, pathways and systems needed to accomplish rapid analysis of sensor and instrument data to adjust ongoing experiments, perform real-time analysis and predictive simulation, and make decisions in near real time. Examples include live fusion of data from geographically-distributed sensors, and fast turnaround adjustment of synchrotron light source experiments and telescope imaging via high-speed computation to focus on spatial features or time-sensitive events of interest. Exascale resources are anticipated to greatly increase the capability to achieve such rapid analysis and response pathways, and enable uncertainty quantification for real-time predictive models in areas such as earthquake research, weather and climate studies, and “smart city” environments. Advanced workflow and data infrastructure will be required in these rapid-response applications to deal with large-scale data flows to and from advanced computing resources.

All three ideas of workflow tackled the problems of how to use complicated hardware, exponentially increasing data, and how to separate the domain scientist focused on a particular problem from the minutiae of writing and executing code for an exascale machine.

5. Application Development

Many RFI responses emphasized that new theory and algorithms will be required to take advantage of capable exascale computing environments. A substantial number of respondents described the need to extend and improve existing HPC applications, to include fluid dynamics, continuum-mechanics, particle physics, and agent-based modeling. Additionally, respondents noted that the parameterization of physical phenomena will be important. Most of these responses describe the need for new methods and techniques to validate the models using experimental data. Because these dataset sizes are growing, large-scale resources will be needed to post-process and visualize results for human consumption. In addition, a large number of these responses described the need for optimized network and hardware configurations that are driven by the needs of increased dataflow in end-to-end applications. These needs will require substantive efforts to validate new science regimes and improve and parallelize existing applications, and develop new software solutions.

6. Security and Privacy

Data security, trustworthiness, and privacy are also a current important focus in the HPC ecosystem and will remain critical issues in the advance towards exascale computing. RFI responses highlighted precision medicine as a research area with special concerns for data security and privacy. In the near future, genomic data will be routinely mined for important health information from individuals seeking treatment for disease, and as such, it could be susceptible to intrusion or theft. Likewise, genomic data contributed anonymously by study participants to public databases may be accessed without authorization, a violation of privacy. It was noted that as the processing of personal data becomes ubiquitous in both the public and private domains, public confidence in the scientific enterprise may be increasingly linked to achieving trustworthiness in the cyber domain. Understanding the implications of security as computational overhead was discussed in the RFI responses, where bioinformaticists, in addition to articulating a desire for faster execution of software applications, mentioned a concern for computing which could be constrained by the requirements of data de-identification tools.

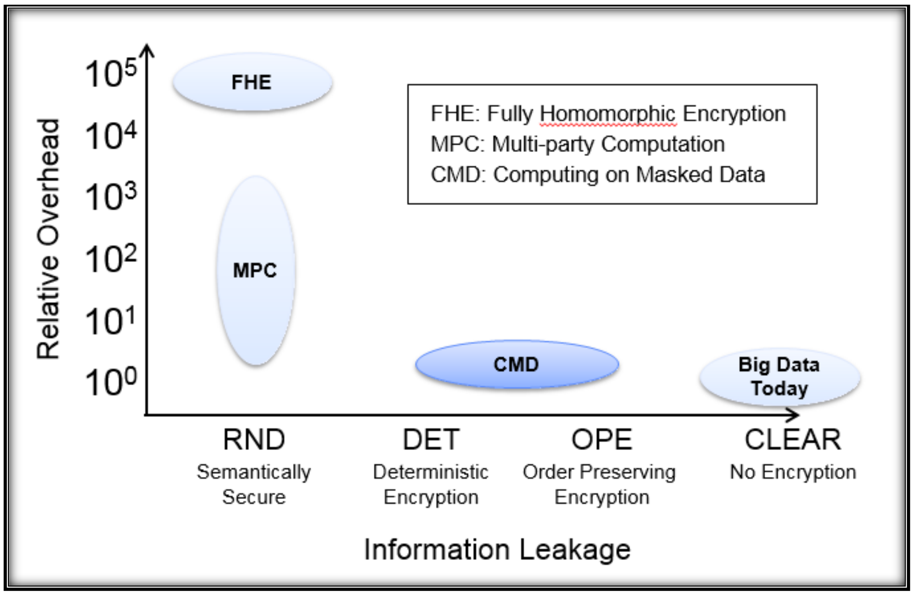

***Figure 6. Computational overhead of different cryptographic techniques**. Cryptographic protection adds significant computational cost. For example, the most effective protection schemes (FHE and MPC) require massive computation and communication overhead respectively. Yet medical and bioinformatics computations must add cybersecurity protections to protect privacy of patients and meet HIPAA regulations. The cryptographic overhead alone could push workstation codes into the exascale regime. Courtesy of Dr Patrick Dreher (Massachusetts Institute of Technology and North Carolina State University), Dr. Chris Hill (Massachusetts Institute of Technology), Dr. John Quackenbush (Harvard School of Public Health).*

The RFI responses also discussed the importance of exascale platforms for carrying out research in many areas associated with national security. For instance, greater computing power by 100x may put current encryption algorithms at risk for attack. In addition, some respondents noted that the computing and data requirements to support more sophisticated encryption regimes could put an extra burden on data storage, computation (Figure 6), and networks. Lightweight systems such as tablets, notebooks, cell phones, and smart electronics devices from the "Internet-of-Things" (IoT) may not have sufficient encryption strength to function in an exascale computing environment. Respondents noted that properly identifying hazards posed by adversaries possessing exascale platforms and effective controls to reduce residual risk will be important issues for consideration.

7. Workforce Development and Usability

Learning and workforce development (LWD) appeared to be a nearly universal concern among respondents from all domains as a critical area of needed future growth. Many respondents indicated that a spectrum of workforce issues must be addressed in order to ensure the realization of the scientific benefits of future HPC systems. . Overall, ease of use was seen as a significant barrier to broadening participation. A number of respondents noted the present barriers to many domain researchers' use of advanced computing resources, given the complexity of the current HPC environment and applications. The anticipated increased complexity of exascale computing could be a significant barrier to expanding the community of users. There was a clear call for an expert workforce capable of developing and using HPC applications that use and maintain advanced computing frameworks. Respondents also observed that a cadre of experts, with hybrid skills in domain research and computer science knowledge, will be needed.

Some of the RFI respondents highlighted the challenges in the HPC workforce, including technical complexity of HPC platforms; the large private sector demand for computer scientists and software engineers; consequent retention difficulties in academic settings; and a compounding lack of academic incentives and/or stable career paths for scientists engaging in time-consuming, risky software development and maintenance. Respondents recommended solutions including training in HPC computing for students and researchers at all levels, and cross-training of graduate students in a domain science with computer scientists in order to learn software engineering.

A number of responses considered the alternatives of emphasizing the training of domain scientists in HPC and parallel computing versus embedding computer scientists within teams of domain scientists. Portals were suggested as one potential method of providing access to exascale computing information and resources to non-experts. As described earlier, other responses suggested additional abstraction layers to allow domain scientists to focus on problem design while letting computing experts focus on implementation.

IV. Conclusions

This report provides a joint assessment by the NIH, DOE and NSF of community input in response to Requests for Information (RFI) by NIH, DOE and NSF in the fall of 2015 to identify scientific and engineering research areas that would benefit from a greatly enhanced next-generation HPC ecosystem far beyond what can be done using current technologies and architectures.

Each of the three agencies analyzed the aggregate of responses in the context of their missions and programmatic goals; summaries of the perspectives of the individual agencies are provided in Appendix 1. Text of the RFIs are provided in Appendix 2. The main report represents a joint synthesis of those analyses and perspectives, as summarized in these conclusions.

Taken together, the entirety of responses comprise an exciting array of anticipated, potentially transformative scientific advancements and societal impacts that could be realized through the significant increase in application performance represented by capable exascale computing. The RFI responses underlined the importance to established computational fields of achieving major advances in application performance and scientific discovery, and additionally highlighted the rapid expansion of computational and data-intensive approaches in new domains. Indeed, it was clear that in many fields data-intensive research is rapidly emerging to complement modeling and simulation. The two being combined pressures the design of future HPC capabilities and resources. The responses also revealed that a rich collection of computational approaches and use models can be found even within a single discipline. The set of common computer science, mathematics, and applications development needs driven by cross-cutting computational use models could be a source for further interagency partnerships and collaborations.

Science drivers were highly varied both within and across disciplines, and were categorized into three broad application domains:

- *Modeling and simulation: Enabling vast improvements in spatial and temporal realism and associated predictive accuracy*, with emphasis on the transformative potential of extending simulation scales to match those of the natural phenomena under study, comprehensive exploration of the increasing parameter spaces of highly realistic models, increased understanding of simulation accuracy with uncertainty quantification, and multi-physics full systems modeling to more accurately simulate coupled systems and capture real-world complexity.

- *Data-intensive science: Enabling analysis and visualization of multi-source and multi-scale data at unprecedented scales,* particularly including comparison and validation of large-scale realistic models with large-scale observational and experimental data.

- *On-demand and real-time computing: Enabling real-time analysis of simulations, data-intensive experiments, and streaming observations*, to achieve computational steering of simulations and experiments, rapid event detection from observational data, and rapid societally-relevant decision-making from real-time, data-informed model predictions.

Across these application domains, the RFI responses also identified a number of challenges that must be overcome through sustained investment to achieve the full potential of capable exascale computing:

- *HPC architectures:* The strong dependence of HPC architectural requirements on scientific research objectives indicates that exploration and deployment of a broad spectrum of HPC capabilities, architectures and technologies must continue to be supported, along with efforts to achieve coherence in solutions for numerical-intensive and data-intensive computing.

- *Software:* Significant and sustained investments will be required in software development to incorporate new models and new science enabled by greatly increased computational power to adapt existing software applications to an exascale environment, and to accomplish validation and uncertainty quantification in increasingly complex multi-modal and multi-spatial analyses.

- *HPC Ecosystem:* A dynamic, agile and robust HPC ecosystem, in which advanced computing is coupled seamlessly with data and software infrastructure, workflow systems, and advanced networking, is critical to support the increasing demand for automation, data-intensive science, analysis and visualization, and real-time processing, with particular attention to data security and privacy.

- *Developing the future HPC workforce:* Achieving the full benefits of capable exascale computing will depend on comprehensive training of the next generation of computationally-savvy, multidisciplinary researchers; creating stable career pathways for a national cadre of computational technologists and HPC experts; fostering strong collaborations among domain scientists, application developers, and computing professionals; and facilitating access to HPC resources by a broad base of users across large-scale and long-tail science.

The assessment provided in this report is based on the set of RFI responses. Respondents included individuals and groups from universities, DOE national laboratories, industrial partners, and non-profit entities. While the responses cover a significant number of research disciplines, it was noted that a number of response gaps were identified in disciplinary coverage. Consequently, additional community input will be needed to ensure that agencies achieve a comprehensive understanding of the science and engineering research areas that would benefit from a capable exascale HPC ecosystem.

For over six decades, the United States has maintained leadership in advanced computing through continuous research, development, and deployment of computational and data facilities, capabilities, and resources comprising a rich HPC ecosystem. The three agencies issuing the RFI will use these responses, together with other inputs and analyses, to plan their investments in advanced computing and the HPC ecosystem to enable and accelerate future scientific and engineering research and discovery.

Appendix 1. Agency Perspectives on the RFI responses

National Institutes of Health (NIH)

Life science (basic biomedical and health sciences research) responses indicated a broad range of applications characterized by multivariate and multi-modal problems with data and phenomena often at multiple scales. The RFI responses in life sciences areas can be described in two categories:

1. Traditional HPC at exascale levels: Typical applications are Molecular Dynamics (MD) and many body-problem calculations that can scale to large number of processors, especially for some kinds of large-atom MD. Typical studies may be protein structure function studies, such as drug design in silico. Also, there are a number of areas in physiological research, such as whole cell, vascular, heart or other organ modeling applications that have legacy implementations in high end computers.
2. Big data at exascale: This includes image and data analysis, genome analysis, cell network analysis, and multivariate data integration. Many of these include correlative, association, maximum likelihood and network analysis algorithms. Quantitative analysis is challenging because the measurement of correlations in biology is very difficult, which subsequently makes it hard to develop and implement applications for data integration.

Future emphasis on data visualization post analysis, as some respondents have noted, will need to utilize new methods and analytics in order to make sense of the exabyte datasets from biological experiments or simulations. Similar to the situation for traditional HPC fields, biomedical and life science applications will eventually move to integrating real-time data assimilation and analytics with computation. The ability of the life science fields to utilize exascale computing will require extensive development of new and improved algorithms, attention to privacy and security, and training of a new workforce capable of both domain science and high end computing.

As a result of this initiative, a new three-year DOE/NIH pilot program is underway and exemplifies the interagency collaborations that will contribute to achieving capable exascale computing. Scientists at DOE National Laboratories and NIH's National Cancer Institute (NCI) will use "Collaboration of Argonne, Oak Ridge and Livermore" (CORAL) class computing to extend cancer science and ultimately clinical treatments[5] by:

- Developing new computational approaches to attack mutant proteins that result in certain types of aggressive cancers. This work may lead to potential new cancer therapies.
- Using large-scale computations to accelerate the development of patient-derived laboratory models of cancer. Together DOE and NCI will explore application of deep learning to combine and extract features for large-scale compound and cancer drug screening.
- Understanding the impact of existing therapies outside of clinical trials in real-world settings.

In addition to improved targeting of drug therapy for cancer patients, this pilot is designed to push the frontiers of HPC, for example, by integrating both data and scientists in new ways, potentially leading

[5] CORAL is a joint procurement activity among three of the DOE's National Laboratories launched in 2014 to build state-of-the- art high performance computing technologies that are essential for supporting U.S. national security and are key tools for technology advancement. This collaboration will procure leadership computing systems for the National Laboratories and is part of a National strategic computing effort that would align strategies and resources across the Federal enterprise.

to future innovations that shape the architectures for exascale platforms while transforming cancer treatment.

National Science Foundation (NSF)

NSF plays a central role in scientific discovery advances, the broader HPC ecosystem for scientific discovery, and workforce development. NSF is focused on increasing the coherence between technologies, modeling and simulation and data-analytic computing; "post-Moore's Law era" solutions; and increasing the capacity and capability of an enduring national HPC ecosystem. NSF conducted an assessment of the joint RFI responses from the viewpoint of its mission and these focus areas.[6]

The Foundation found that the RFI responses comprise an exciting range of scientific and engineering research opportunities across NSF-supported domains and disciplines. Transformative ideas were put forward in both well-established computational fields (e.g., physics, climate and geoscience) and emerging computational fields (e.g., neuroscience, -omics, and materials science).

NSF also found excellent alignment of stated community needs with recommendations of the recent National Academies of Science report on NSF leadership in computing and the HPC ecosystem,[7] particularly regarding the need to support the full range of science requirements for advanced computing – including both numerical and data-driven science; supporting development and maintenance of software infrastructure and applications; and supporting exploration of innovative, next-generation technologies.

Indeed, NSF's overriding conclusion from the responses is that a comprehensive, inclusive and holistic investment approach will be required to support the expanding scientific and engineering horizons in all research domains. Broad investments will continue to be necessary to explore and deploy a rich spectrum of HPC capabilities, architectures and technologies, and to advance the full HPC ecosystem of software, data, and information infrastructure, along with human capital. In particular, ambitious, highly diverse plans in numerical- and data-intensive science, as well as the rapidly increasing need, evidenced in the RFI responses, to inform large-scale models with comparing large-scale data will require investments in the convergence of architectures and technologies used for modeling and simulation and data integration, analysis, and analytics. Significant concomitant investments will be required in software, application development and portability, validation, and quantifying uncertainty in increasingly complex multi-modal and multi-spatial analyses. The development of a national cadre of computational technologists and HPC experts with stable career pathways; training the next generation of computationally-capable, multidisciplinary researchers; and facilitating access to HPC resources by an increasingly broad base of savvy users across large-scale and long-tail science will be critical.

Facilitated by this assessment of community needs, NSF has launched a Foundation-wide effort to begin to address future HPC challenges[8]. In FY 2016, NSF funded a series of projects through existing programs

---

[6] NSF Assessment of Responses to the Request for Information (RFI) on Science Drivers Requiring Capable Exascale High Performance Computing, http://nsf.gov/cise/nsci/NSFNSCIRFIassessmentfinal07-29-16.pdf.

[7] National Academies of Science, "Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020", https://www.nap.edu/catalog/21886/future-directions-for-nsf-advanced-computing-infrastructure-to-support-us-science-and-engineering-in-2017-2020.

[8] See NSCI @ NSF website, https://www.nsf.gov/cise/nsci/.

and special funding opportunities. The FY 2017 Budget Request for NSF includes more than $33 million to support pilot activities, community workshops, and core program investments to focus efforts on advancing the Nation's computational infrastructure for science and engineering research. These planned investments in FY 2017 are responsive to calls by RFI respondents for enabling fundamental research in computational techniques, foundational algorithms and technologies, architectures, methods and applications; for development and deployment of high-capability computing systems and shared research cyberinfrastructure; and for expanding opportunities for learning and workforce development across all participants in computational and data-intensive science. Existing and new NSF cross-agency initiatives, such Understanding the Brain (UtB), Innovations at the Nexus of Food, Energy and Water Systems (INFEWS), and the NSF Big Ideas,[9] are also incorporating NSF goals for advancing scientific discovery via a broadly capable HPC ecosystem and associated workforce.

In concert with the on-going HPC ecosystem efforts described above, in FY 2017, NSF has initiated planning for a refreshed vision and strategy for future investments in advanced cyberinfrastructure *writ large* over the next decade, as its current Foundation-wide effort, entitled "Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21),"[10] completes its final year. As a first step in this planning, and building on the present results of the RFI analysis, NSF has opened a new opportunity, entitled "NSF CI 2030", for the science, engineering, and cyberinfrastructure research communities to provide bold, forward-looking ideas on *Future Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research*.[11] This new effort will allow the Foundation to refine its understanding of intra- and inter-disciplinary computational and data science needs for advanced cyberinfrastructure resources and services, while also addressing identified gaps in the disciplinary coverage of the joint RFI responses noted in the main body of the present report. Such gaps include intense computational and data needs driven by activities such as ArcticDEM, a major new interagency coordinated effort to develop high-resolution maps of the Arctic using leading-edge computing capabilities like Blue Waters[12]. Finally, NSF is planning additional workshops and community engagements associated with recommendations made by the National Academies' report noted earlier.

The present RFI assessment, together with the additional community contributions cited above and alignments with agencies including the DOE and NIH, will inform the Foundation's HPC and cyberinfrastructure strategies and investments toward enabling future scientific and engineering research and discovery.

The Department of Energy (DOE)

DOE has the responsibility to execute a capable exascale computing program emphasizing sustained performance on relevant applications. Part of meeting this task requires DOE to understand the science drivers and applications that will be relying on exascale systems. Coupled with complementary efforts, DOE is using the RFI responses discussed here to inform plans to deliver the exascale computing

[9] See NSF Big Ideas, https://www.nsf.gov/about/congress/reports/nsf_big_ideas.pdf.
[10] See CIF21, https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504730.
[11] See NSF Dear Colleague Letter on NSF CI 2030 Request for information, https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031.
[12] See Arctic DEM first map release: https://www.nsf.gov/news/news_summ.jsp?cntn_id=189516

resources, supporting technology, human capital and networking infrastructure that will maximize science discovery and technological innovations in the exascale era.

The DOE Exascale Computing Project (ECP) carried out an in-depth analysis of the DOE RFI responses. The ECP is a DOE national laboratory led project to develop capable exascale hardware and software technologies and an application portfolio targeting high priority programmatic challenges in the DOE and the NSCI deployment agencies (NIH, NOAA, NASA, FBI and DHS). ECP goals include: (1) developing a broad set of modeling and simulation applications that meet the requirements of the scientific, engineering and nuclear security programs of the DOE; (2) developing a productive exascale capability in the U.S., including the required software and hardware technologies and preparing two or more facilities to house this capability; and (3) maximizing the benefits of HPC for U.S. economic competitiveness and scientific discovery.

The ECP chartered an Exascale Application Working Group (EAWG) with its initial membership consisting of subject matter experts from 16 DOE National Laboratories, with expansion planned in the near future including experts from academia and industry. The EAWG has helped the ECP to prioritize and reduce the DOE RFI responses to a more manageable set (57) that were then reduced to 33 based on peer review. Of the 33 responses, 30 were invited to submit full proposals for funding and support through the Exascale Computing Project. Criteria for the down selection process included identification of a key challenge across DOE science, energy, and national security missions; support from one or more DOE program offices as a strategic priority; a challenge requiring an exascale capability; acceptable quality and makeup of the project team; a thorough technical plan; and a manageable risk profile. An initial selection supported 15 application projects and 7 seed efforts. The supported projects deliver a broad coverage of strategic areas important to U.S. economy, security, and scientific leadership. Areas include high-efficiency combustion engine and gas turbine design; additive manufacturing; synchrotron light source-enabled analysis of protein and molecular structure; the design, acceleration and translation to cancer research; and a cosmological probe of the Standard Model of particle physics. This work on applications will be critical to the overall success of the exascale ecosystem.

In addition to in-depth reviews of the RFIs by the ECP team, DOE has completed a series of six workshops with DOE Office of Science partner offices to gather requirements from the scientific community for the exascale HPC ecosystem. Each workshop will produce a report identifying science drivers and requirements for exascale platforms. The requirements reviews, combined with knowledge gained from the RFI responses, will inform DOE HQ strategic planning for DOE HPC and networking facilities in the exascale era.

In response to the critical need for a science and technology workforce with advanced computing skills, the DOE began and continues to support the Computational Sciences Graduate Fellowship (CSGF). CSGF is a unique multidisciplinary program focused on developing future computational scientists by supporting fellows as they pursue substantive graduate work in both an application domain and in computer science/applied mathematics. Supporting the interdisciplinary mission of the program, fellows also participate in a practicum at a DOE laboratory in an area of research outside of the student's thesis dissertation. DOE CSGF alumni form a core group of computational science leaders that join industry, academia, and government labs and contribute to a workforce more fully prepared to use the power of exascale architectures.

The Joint RFI activity provided input from communities outside of DOE's research mission but important to the Federal enterprise and for delivering on DOE's capable exascale objective -- in particular the health sciences community. The large response from this community suggests the future user communities of HPC systems will be more diverse than they are today. These new user communities will benefit from HPC resources and expertise at the National Laboratories and also spur new research challenges for DOE. Currently, DOE is partnering with NIH in a pilot program supporting the Cancer Moonshot.. This pilot program has not only identified key areas of NIH research that can be advanced by DOE expertise in mathematics and computer science, but also areas where clinical and health science goals bring new research challenges in machine learning and data science to the DOE community.

Taken together, these efforts will guide DOE R&D, inform DOE public-private partnerships, and guide interagency collaboration needed to enable a capable exascale HPC ecosystem. DOE looks forward to continued partnerships with NSF, NIH, and the current and future user community to deliver a capable exascale HPC system.

Appendix 2: Texts of the Request for Information

A. Text of Joint RFI from NIH, DOE and NSF

Note: This RFI is available at http://grants.nih.gov/grants/guide/notice-files/NOT-GM-15-122.html. A deadline extension was announced at http://grants.nih.gov/grants/guide/notice-files/NOT-GM-15-123.html.

Request for Information (RFI) on Science Drivers Requiring Capable Exascale High Performance Computing

---

Notice Number: NOT-GM-15-122

Key Dates
Release Date: September 15, 2015

Related Announcements
NOT-GM-15-123

Issued by
National Institute of General Medical Sciences (NIGMS)

Purpose

This is a multi-agency request for information to identify scientific research topics and applications that need High Performance Computing (HPC) capabilities that extend 100 times beyond today's performance on scientific applications. Currently, computational modeling, simulation, as well as data assimilation and data analytics are used by an increasing number of researchers to answer more complex multispatial, multiphysics scientific questions with more realism. As the scientific discovery horizon expands and as advances in high performance computing become central to scientific workflows, sustained petascale application performance will be insufficient to meet these needs. In addition, HPC is expanding from traditional numerically oriented computation to also include large-scale analytics (e.g., for Bayesian approaches in model refinement, large-scale image analysis, machine learning, decision support, and quantifying uncertainty in multimodal and multi spatial analyses). Architectures and technologies used for modeling and simulation currently differ from those used for data integration and analytics, but are increasingly converging. The extreme computing ecosystem must therefore accommodate this broad spectrum of growing data science activities.

Background

The White House Executive Order, July 29, 2015, establishes the National Strategic Computing Initiative (NSCI) as a whole-of-government effort designed to create a cohesive, multi-agency strategic vision and Federal investment strategy, executed in collaboration with industry and academia, to maximize the benefits of HPC for the United States. The Department of Energy (DOE), the National Science Foundation (NSF), and the Department of Defense (DOD) are the lead agencies for this effort to support a significantly advanced HPC ecosystem within the next decade. One of the objectives of the initiative is to deliver "capable Exascale" computing capability that delivers 100 times today's application performance.

This is a request for information from NSF, DOE, and NIH for community input identifying scientific research that would benefit from a greatly enhanced new generation of HPC far beyond what can be done using current technologies and architectures. This information will be used to assist agencies to construct a roadmap, build an exascale-capable ecosystem required to support scientific research, and inform the research, engineering and development process. It is likely that a range of advanced capabilities will need to be developed to respond to the varied computing needs across science disciplines.

We seek responses for applications in subfields of life (e.g., biological, social, health and biomedical) sciences, mathematical and physical sciences, geosciences, energy science, and engineering research. We seek not only traditional areas of numerical intensity such as simulations of nuclear physics, biomolecular physics, weather and climate modeling, and materials science, but also any areas that rely on deriving fundamental understanding from large scale analytics that would require a 100-fold increase over today's application performance.

Information Requested

With respect to your field of expertise in traditional and non-traditional research areas in applications of HPC, agencies request your input/feedback. Your comments can include but are not limited to the following areas of concern:

- The specific scientific and research challenges that would need the projected 100-fold increase in application performance over what is possible today.
- The potential impact of the research to the scientific community, national economy, and society.
- The specific limitations/barriers of existing HPC systems must overcome to perform studies in this area. Your comment can also include the level of performance on current architectures, and the projected increase in performance that is needed from future architectures.
- Any related research areas you foresee that would benefit from this level of augmented computational capability. Identification of any barriers in addition to computational capability that impact the proposed research can also be considered.
- Important computational and technical parameters of the problem as you expect them to be in 10 years (2025). In addition to any specialized or unique computational capabilities that are required and/or need to be scaled up for addressing this scientific problem, e.g., in the areas of computing architectures, systems software and hardware, software applications, algorithm development, communications, and networking.
- Alternative models of deployment and resource accessibility arising out of exascale computing. Improvements in scientific workflow as well as particular requirements that may be needed by specific domains.
- Capabilities needed by the end-to-end system, including data requirements such as data analytics and visualization tools, shared data capabilities, and data services which includes databases, portals and data transfer tools/nodes.
- Foundational issues that need to be addressed such as training, workforce development or collaborative environments.

- Other areas of relevance for the Agencies to consider.

Submitting a Response

All responses must be submitted to NIGMS_exascale@nigms.nih.gov by October 16, 2015. All comments must be submitted via E-mail as text or as an attached electronic document. Microsoft Word documents are preferred. Please try and limit your response to two pages total.

This RFI is for planning purposes only and should not be construed as a solicitation for applications or an obligation on the part of the government. The government will not pay for the preparation of any information submitted or for the government's use of that information.

The agencies will use the information submitted in response to this RFI at their discretion and will not provide comments to any responder's submission. Responses to the RFI may be reflected in future funding opportunity announcements. The information provided will be analyzed, may appear in reports, and may be shared publicly on agency websites. Respondents are advised that the government is under no obligation to acknowledge receipt of the information or provide feedback to respondents with respect to any information submitted. No proprietary, classified, confidential, or sensitive information should be included in your response. The government reserves the right to use any non-proprietary technical information in any resultant solicitation(s), policies or procedures.

Inquiries

Please direct all inquiries to:

Susan Gregurick, Ph.D.
National Institute of General Medical Sciences (NIGMS)
National Institutes of Health
Telephone: 301-451-6446
Email: susan.gregurick@nigms.nih.gov

William L. Miller, Ph.D.
Office of Advanced Cyberinfrastructure
Directorate for Computer and Information Science and Engineering
National Science Foundation
Telephone: 703-292-7886
Email: WLMiller@nsf.gov

Barbara Helland
Advanced Scientific Computing Research (ASCR)
Department of Energy
Telephone: 301-903-9958
Email: barbara.helland@science.doe.gov

B. Text of the DOE RFI to the DOE National Laboratories

Template for Additional Exascale Applications

Please limit to 2 pages. No proprietary, classified, business confidential, or sensitive information should be included

Title/Name of Your Application: Provide a title that uniquely identifies your project or code.

Name, Institutional affiliation, and e-mail address: for any follow up questions or clarifications.

Overview Description and Impact: A half-page description of the specific scientific and research challenges that need the projected 100-fold increase in application performance. Clearly state the scientific question(s) that could be answered and potential impact(s) of the research to the scientific community, national economy, and society.

System Requirements: Identify the specific limitations of existing HPC systems that must be overcome to perform the planned studies in this area. If available, discuss the level of performance achieved on current architectures in terms of figures of merit for your application, and the projected increase in performance that is needed from future architectures to meet your science goals. Describe any specific hardware and software requirements for the system.

- **Code and Tools**: Describe the code and/or tool suite that you have now or will need to address these research objectives. Include details about existing or anticipated language(s), libraries, I/O and any special runtime requirements. Indicate the systems it currently runs on if a code exists. For existing codes, indicate the source of support (e.g. agency and program) for the development of these codes and tools.

- **Models and Algorithms:** Describe any new mathematical models and algorithms that will be needed to reach your scientific objectives.

- **End-to-End Requirements:** Describe any capabilities needed by the end-to-end system, including data requirements such as data analytics and visualization tools, shared data capabilities, data services, databases, portals and data transfer tools/nodes.

Related Research: Identify any related research areas in your domain of expertise that you foresee that would benefit from this level of augmented computational capability. Identify any barriers in addition to computational capability (e.g. availability of experimental data) that impact the proposed research.

10-Year Problem Target: Describe the computational and technical parameters of example problems in this area as you expect them to be in ten years (2025). Indicate key performance parameters specific to this area (e.g. simulated years per day, number of particles, etc.).

Other Considerations/Issues: Describe any other foundational considerations or issues, such as requirements for training or collaborative environments.