Testimony of Dr. Georgia Tourassi
Director, National Center for Computational Sciences
Oak Ridge National Laboratory

Before the
Subcommittee on Energy
Committee on Science, Space, and Technology
U.S. House of Representatives

"Accelerating Discovery: the Future of Scientific Computing at the Department of Energy"
May 19, 2021

Chairman Bowman, Ranking Member Weber, and distinguished members of the Committee: Thank you for the opportunity to appear before you today. My name is Georgia Tourassi. I am Director of the National Center for Computational Sciences of the U.S. Department of Energy's Oak Ridge National Laboratory (ORNL) in Oak Ridge, Tennessee. I am a biomedical engineer and computational scientist by education and training. My research work lies at the intersection of high-performance computing (HPC), artificial intelligence (AI), and biomedicine. I am privileged to direct one of the Department of Energy's User Facilities, the Oak Ridge Leadership Computing Facility (OLCF), which hosts the nation's most powerful supercomputer for open science, Summit. It is an honor to provide this testimony on the transformative role of the Department of Energy's Leadership Computing User Facilities in spearheading a national high-performance computing and data infrastructure for accelerating research and development (R&D) giving the nation a significant advantage in the ever-changing scientific and economic landscape.

The views I offer in my statement are my own, shaped by my scientific experiences, my experience with inter-agency partnerships to advance respective missions, and my experience as Director of a DOE Leadership Computing Facility dedicated to open science.

INTRODUCTION

High-performance computing (or supercomputing) has been the cornerstone for research advancement, scientific innovation, competitive advantage, and economic prosperity[1]. The Department of Energy's Leadership Computing Facilities have a long history of enabling researchers to accelerate scientific discovery and deliver practical breakthroughs for some of the most computationally challenging problems. These research and development advances happen across many disciplines such as materials science, earth science, nuclear science, astrophysics, biology, and engineering to name a few. In addition, supercomputing has proven to be an effective ally to our society by helping us address critical and pressing challenges, such as climate change, national emergencies, and public health. For instance, supercomputers help researchers develop personalized medical treatments as well as better predict and manage the effects of natural disasters such as floods and earthquakes through the use of advanced computer simulations.[2,3] These discoveries help shape our understanding of the universe, bolster US economic competitiveness, and contribute to a better future.[4]

I will begin with a brief overview of activities at the Oak Ridge Leadership Computing Facility, exemplifying its lasting impact on scientific discovery, including its impact on global emergencies, most recently the COVID-19 pandemic. Then, I will discuss the need for sustainable R&D investments in supercomputing and data management strategies to ensure the nation's continued economic competitiveness and to help expand our nation's high-tech workforce. Lastly, I will discuss a vision of how DOE's national laboratories' leadership computing user facilities and computational

---

[1] US DOE Office of Science. High-performance Computing. https://www.energy.gov/science/initiatives/high-performance-computing. Accessed May 13, 2021.

[2] K.R. Milner, et al, Toward Physics-Based Nonergodic PSHA: A Prototype Fully Deterministic Seismic Hazard Model for Southern California," Bulletin of the Seismological Society of America (2021). doi:10.1785/0120200216.

[3] C.W. Tessum, et al. "InMAP: A model for air pollution interventions." PloS ONE 12, no. 4 (2017): e0176131. https://doi.org/10.1371/journal.pone.0176131

[4] Report. *The State of the DOE National Laboratories.* https://www.energy.gov/sites/default/files/2021/01/f82/DOE%20National%20Labs%20Report%20FINAL.pdf

workforce can support a thriving research and development ecosystem while promoting AI innovation, energy-efficiency, stakeholder engagement, and national security.

THE OAK RIDGE LEADERSHIP COMPUTING FACILITY (OLCF)

Collectively, the DOE laboratory complex occupies a distinctive position in the national science and technology innovation ecosystem. By fostering strong partnerships with industry and academia, the DOE national labs bring together one-of-a-kind and often first-of-a kind scientific instruments, powerful computing infrastructure, and computational and data science experts in multiple disciplines to push the boundaries of computing and solve some of the world's biggest challenges in science and engineering.

ORNL has been a global leader in high-performance computing for nearly 30 years. The Oak Ridge Leadership Computing Facility (OLCF) provides access to supercomputers of unprecedented capability to the scientific community on behalf of the US Department of Energy (DOE). These versatile systems are 10 to 100 times more powerful than most supercomputers available for research. Our users run a variety of codes and analyses, from traditional modeling and simulation to new and emerging classes of data science and artificial intelligence. But in every instance, our users tackle the most complex and difficult problems that cannot be easily solved using conventional computing resources.

From the beginning, the OLCF has been at the forefront of the rapid evolution in scientific computing. OLCF provided the **first teraflop system** (IBM Power3 Eagle) for open science and the science community's **first petaflop system** (Cray XT5 Jaguar). **Three OLCF systems ranked as the fastest in the world by the TOP500 list,** including OLCF's current **pre-exascale system**, Summit, an IBM system. Summit debuted in June 2018 and has a theoretical peak performance of 200 petaflops. This is over a million times the capability of Eagle and provides scientists with incredible computing power to solve critical problems in areas as diverse as clean energy, human health, advanced materials, climate, nuclear physics, and other research areas.

OLCF is at the brink of delivering **the first exascale system** in the U.S., the 1.5 exaflops Frontier supercomputer, expected this fall. Frontier will perform calculations up to 50 times faster than today's top supercomputers. Frontier will support research in solving problems in energy, medicine, and materials that were impossible as recently as 5 years ago. Exascale is the next level of computing performance and is needed to continue U.S. leadership in high-performance computing. DOE's Exascale Computing Project is developing applications across two dozen scientific and technical disciplines to run on Frontier on day one. These applications will offer new insights in quantum materials, chemistry, physics, additive manufacturing, fuels, fusion, fission, and more. Equally important, Frontier will mark an impressive trajectory for our facility, for DOE, and for the nation, demonstrating that in the past decade we improved computational power by a factor of 500 while suppressing growth of the carbon footprint which increased only by a factor of 4.

Summit is in high demand. Researchers annually submit proposals for groundbreaking computational challenges and are granted time on DOE leadership computers based on scientific merit and their need for and ability to use the full compute power and capacity of a leadership-class system. OLCF typically receives requests for **three to five times** more node hours than what the programs are able to award**.** Still, the facility preserves the ability to shift resources to address national emergencies, as you will hear later in my testimony. Last but not least, to train users for success on our HPC systems, we host workshops, conference calls, and training events. These events can help prepare new users for large-scale computing systems, provide useful tools to veteran users, and facilitate engagement both with the public and the user community.

ADVANCING SCIENCE VIA SUPERCOMPUTING

Using supercomputers, scientists have expanded the scale and scope of their research, solved complex problems in less time, and filled critical gaps in scientific knowledge. The following is a short list of recent examples of how scientists leveraged Summit and OLCF expert support to advance science.

Reaching new heights in weather forecasting[5]: Using Summit, a team of researchers achieved a computational first: a global simulation of the Earth's atmosphere at a 1-square-kilometer average grid-spacing for a full 4-month season. The milestone marks a big improvement in resolution which today operates at 9-square-kilometer grid-spacing for routine weather forecast operations. This improvement is a critical step in the effort to create multi-season atmospheric simulations at high resolution, pointing toward the future of weather forecasting—one powered by the emerging exascale supercomputers.

Gaining new insights into quantum materials[6]: Quantum materials—those that have correlated order at the subatomic level—have potential applications in electronic devices, quantum computers, and superconductors. For the first time, a team at ORNL used supercomputing and artificial intelligence (AI) to find patterns in neutron scattering data that can guide future neutron scattering experiments and lead to deeper understanding of the physics inside quantum materials.

Designing efficient and durable gas turbine jet engines[7]: General Electric (GE) designs gas turbine jet engines that are world-leading in efficiency, emissions, and durability. A key requirement to successfully design these engines is the ability to accurately predict the temperature and motion of fluids in the high-pressure turbines that power them. Internal turbine temperatures become so high during operation that turbine components will melt if they are not cooled properly. A GE team used the Summit supercomputer at ORNL to complete first-of-their-kind 3D flow simulations that provided breakthrough insights about these fluid flows and revealed how to better cool engine parts for more efficient and durable jet engines. The team performed the computationally intensive large simulations with unprecedented speed and exceptional detail that more closely matched results of actual engine tests.

---

[5] Wedi, N. P., et al. A baseline for global weather and climate simulations at 1 km resolution. *Journal of Advances in Modeling Earth Systems*, 12 (2020), e2020MS002192. doi: 10.1029/2020MS002192
[6] Anjana M. Samarakoon et al., "Machine-Learning-Assisted Insight into Spin Ice $Dy_2Ti_2O_7$." *Nature Communications* 11 (2020): 892, doi:10.1038/s41467-020-14660-y.
[7] OLCF Website. *Sparks Fly in Marriage of GE's Genesis Code and the Summit Supercomputer*. https://www.olcf.ornl.gov/2020/08/10/sparks-fly-in-marriage-of-ges-genesis-code-and-the-summit-supercomputer/. Accessed May 13, 2021.

Deploying AI solutions for modernization of the national cancer surveillance program: Supporting efforts to reduce the cancer burden in the United States, the national Surveillance, Epidemiology, and End Results (SEER) Program manages the collection of curated data from population-based cancer registries on cancer incidence, prevalence, survival, and associated health statistics for the advancement of public health. As part of the National Cancer Institute (NCI), the SEER program is tasked with supporting cancer research to improve the understanding of patient care and outcomes in the "real world" beyond the clinical trial setting. By leveraging ORNL's pre-exascale computing technologies, a DOE-NCI team delivered state-of-the-art AI tools to help US state cancer registries improve the accuracy and efficiency of their operations in cancer incidence reporting. The AI tools make the dream of "real time" cancer incidence reporting a reality, while also enabling real-time eligibility assessment of cancer patients for clinical trials. This achievement to modernize the national cancer surveillance program exemplifies the benefits of a federally coordinated strategy to harness AI, high-performance computing, and sensitive health data assets for real-world application.

THE ROLE OF SUPERCOMPUTING IN FIGHTING THE COVID-19 PANDEMIC

In the past decade, supercomputing has emerged as a great ally not only to biology but also to biomedicine. This growing trend is even more apparent with life sciences being one of the leading application domains for artificial intelligence. Therefore, from the start of the COVID-19 pandemic, supercomputers, and by extension AI, have helped accelerate discovery, understand the virus, and inform management of the pandemic response.

In my role as director of the Oak Ridge Leadership Computing Facility, I have experienced firsthand our staff's herculean efforts to pivot fast and offer our world-leading computing resources and computational and data expertise in the fight against the coronavirus. Following are recent examples of how OLCF has leveraged its scientific tools and expertise to address challenges posed by the COVID-19 pandemic.

COVID-19 high-performance computing consortium: In March 2020, The OLCF and the Summit supercomputer joined forces with other U.S. federal agencies, industry, and academic leaders to provide access to the world's most powerful high-performance computing resources in support of COVID-19 research through the COVID-19 High-Performance Computing Consortium[8].

The Consortium is a unique private-public effort spearheaded by the White House Office of Science and Technology Policy, the U.S. Department of Energy and IBM to bring together federal government, industry, and academic leaders who are volunteering compute time and resources on their world-class machines. Since the beginning of the Consortium, OLCF has allocated 2,206,200 compute node hours across 21 most computationally demanding projects, supporting multi-disciplinary teams to improve our understanding of the virus's structure and biology in order to develop targeted therapies and vaccines, build better diagnostic and prognostic disease models, model how the disease may spread in communities as conditions change on the ground, as well as improve our understanding of how the virus may change as it spreads through the population.

Guiding vaccine development and therapeutic targeting: Research by a multi-disciplinary team leveraging OLCF's Summit led to a novel understanding of SARS-CoV-2 and a new method for studying disease on Summit. Since researchers first mapped the SARS-CoV-2 spike protein—the main infection machinery of the virus that causes the COVID-19 disease—scientists around the world embarked on a quest to understand the movement of the virus's spike protein, namely how it behaves and gains access to the human cell. A multi-institutional team built a first-of-its-kind AI-based workflow and ran it on the Summit supercomputer to simulate the spike protein in numerous environments, including within the SARS-CoV-2 viral envelope comprising 305 million atoms—the most comprehensive simulation of the virus performed to date. The team first optimized their molecular dynamics codes, which model the movements of atoms in time and space, on multiple smaller HPC systems. These optimizations prepared the team to run full-scale simulations leveraging the whole OLCF Summit supercomputer. The Summit runs have led to more detailed understanding of the virus

[8] The COVID-19 High-performance Computing Consortium Website. https://covid19-hpc-consortium.org/. Accessed May 13, 2021.

behavior but also its vulnerabilities, to guide vaccine development and better therapeutic targeting. Because one set of the Summit calculations generated a whopping 200 terabytes of data (equivalent to 62 million high resolution images), the team used AI to identify the intrinsic features from the simulations and break down the information to help them interpret what was happening. By layering the experimental data and the simulation data and combining it with their AI-based approach on Summit, the researchers were able to capture the virus and its mechanisms in unprecedented detail. The team received the prestigious ACM Gordon Bell Special Prize for HPC-Based COVID-19 Research at the 2020 International Conference for High Performance Computing, Networking, Storage and Analysis[9].

I would like to thank Congress for the CARES-Act funds OLCF received to augment Summit and help support the COVID-19 research community.

WHERE DO WE GO NEXT?

High-performance computing's impact on national competitiveness has long been embodied by the saying, "You must compute to compete." With advances in sensors and other data generation technologies, machine learning, and artificial intelligence, this saying can be amended to, "You must learn faster to compete." In today's world it is evident that the scientists, engineers, companies, and nations that learn the fastest are the most adaptive and most competitive. Those that can harness computation to analyze, identify patterns, and draw inferences from digital models and simulations, empirical measurements from large scientific instruments, or the wealth of data being collected from the connected sensors that have become ubiquitous will make discoveries and innovations at a faster pace than those that do not. It is imperative for the United States to expand and enhance the national research ecosystem and empower researchers with the computational tools so that they can unlock new insights hidden in data. We can accomplish this by taking the following steps.

[9] bioRXiv. "AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics." https://www.biorxiv.org/content/10.1101/2020.11.19.390187v1

First, we need to make the most of our Exascale systems: The upcoming exascale systems such as Frontier mark the beginning of a 5 – 7 year operations cycle. Researchers will be able to answer questions of national importance that simply can't be addressed with today's computing platforms. For example, users of Frontier will model the entire lifespan of a nuclear reactor to enhance efficiency and safety of operations. Biomedical scientists will use Frontier to improve health outcomes in new ways, from uncovering the underlying genetics of disease to extracting important information from complex patient data. Frontier will build on recent developments in science and technology to further integrate artificial intelligence with data analytics and modeling and simulation. These new capabilities will drastically reduce the time to discovery by automatically recognizing patterns in data and guiding simulations beyond the limits of traditional approaches. Overall, exascale systems will allow us to expand the application space and increase the impact of our investments. Furthermore, the exascale systems will offer training opportunities to grow a computationally-aware and computationally-savvy scientific workforce across academia, industry, national labs, and federal agencies.

Second, we need to continuously invest in new technologies and workforce development to maintain competitive advantage and ensure our global leadership: Ongoing investments in computing, data, and software technologies will provide the engine that drives an advanced research ecosystem that is energy-efficient, secure, and performant. Data infrastructure and software are needed to enable the flow of data between facilities. Scientific software must be maintained and extended to support these activities. Research in microelectronics will increase the power of computer processors to the limits of current transistor technology.

Furthermore, the energy demands of digital infrastructures and compute- and data-hungry AI algorithms pose additional challenges. AI and other computing activities are projected to use over half of the world's energy by 2040.[10,11] We cannot easily

[10] Greenpeace, "Clicking Clean: Who is winning the race to build a green internet?" Washington, DC, 2017.
[11] Semiconductor Industry Association and Semiconductor Research Corporation, "Rebooting the IT revolution: A call to action," September 2015.

predict how we can balance AI's energy demands with AI's demonstrated ability to guide management of our energy resources more efficiently. DOE's national laboratories are well positioned to lead hardware, software, and algorithmic innovations and deliver AI solutions that consume less energy, a challenge DOE is already working to address as part of the Exascale Computing Initiative. Energy-efficient AI is key to providing sustainable and affordable solutions that benefit our environment and our national and economic security.

Our societal advances in understanding the brain and its ability to process information in an energy-efficient manner will in turn influence how computing technologies will support society. While interest and advancements in the computing technologies of neural networks and neuromorphic computing were originally inspired by human neural capabilities, they have become broad fields of research in their own right. We anticipate that investments in AI methods and accelerated computing methods — from domain-specific accelerators and learning-and-inferencing at the network's edge to neuromorphic technologies and quantum accelerators — will directly enhance our national competitiveness.

Third, we need to make additional investments to support a national data infrastructure: The explosive growth of AI is driven by the convergence of big data, massive computational power, and novel algorithms. There is a pressing need to consider expanding federal investments in managing our national data assets in a secure environment that can provide the compute resources and scalable data analytics capabilities. The Department of Energy national laboratories are uniquely positioned to lead this effort for the nation, given the Department's long history of serving as the steward of large data infrastructures and of the nation's nuclear security enterprise. With their leading role in high-performance computing and their extensive data science and AI capabilities, the national laboratories could serve as a neutral entity, an honest broker for democratizing AI, while providing meaningful and responsible access to sensitive data assets and compute resources. This would further drive the need to educate and train a new workforce necessary to support these activities. Such an investment is critical to support the continuum of scientific discovery for effective domain-specific application of the nation's high performance computing and data assets. The COVID-19 pandemic demonstrated the value and importance of having established interagency programs and data integration ahead of this unanticipated crisis, and the utility of high-performance computing and artificial intelligence for rapid, complex, real-world data analyses.[12]

CONCLUDING REMARKS

The Department of Energy's national laboratories are a remarkable asset for the nation. Over the past 75 years, they have consistently provided the science and technology needed to address national problems. As a DOE national laboratory and Federally Funded Research and Development Center, ORNL is equipped with exceptional computational and experimental resources to support the Department's mission needs and extend to the nation's research community. Researchers at Oak Ridge and other national laboratories have a wealth of experience in delivering, operating, and using supercomputers to solve the most complex problems of our times. With the steadily increasing power of today's supercomputers and the massive data sets that are becoming available in a variety of areas, we are also in a position to build AI solutions that help us accumulate, analyze, and automate the delivery of functional knowledge in many application domains.

Our nation faces a formidable set of challenges: ensuring national security in a changing world; increasing the availability of clean, reliable, and affordable energy while protecting the environment and addressing climate change; improving human health; and enhancing U.S. competitiveness in the global economy by fostering scientific leadership and encouraging innovation. Leadership computing is expected to offer solutions to many of these challenges, but the importance of building a sustainable and integrated compute and data infrastructure cannot be underestimated.

A cohesive national plan for scientific computing, supporting emerging technologies as well as advancing AI, is imperative to secure the nation's economic

[12] Ramoni, Rachel, et al. "COVID-19 Insights Partnership: Leveraging Big Data from the Department of Veterans Affairs and Supercomputers at the Department of Energy under the Public Health Authority." Journal of the American Medical Informatics Association (2021).

competitiveness and well-being. At the same time, we as a nation have the unique opportunity to create a well-defined, federally coordinated roadmap for integrating our leadership computing facilities with national experimental facilities and federal data assets to deliver benefits across private and public sectors. The DOE Leadership Computing Facilities are uniquely equipped and positioned not only to make substantial contributions to address these opportunities and challenges, but also to support the execution of a national research computing ecosystem.

Thank you again for the opportunity to testify. I welcome your questions on this important topic.

Georgia Tourassi, Ph.D.
Director, National Center for Computational Sciences
Oak Ridge National Laboratory

tourassig@ornl.gov
865-576-4829

Dr. Georgia Tourassi is the Director of the National Center for Computational Sciences at the Oak Ridge National Laboratory (ORNL). Concurrently, she holds appointments as an Adjunct Professor of Radiology at Duke University and as a joint UT-ORNL Professor of the Bredesen Center Data Science Program at the University of Tennessee at Knoxville.

Her scholarly work includes 13 US patents and innovation disclosures and more than 260 peer-reviewed journal articles, conference proceedings articles, editorials, and book chapters. She is elected Fellow of the American Institute of Medical and Biological Engineering (AIMBE), the American Association of Medical Physicists (AAPM), the International Society for Optics and Photonics (SPIE), and the American Association for the Advancement of Science (AAAS).

Her research interests include high performance computing and artificial intelligence in biomedicine. For her leadership in the Joint Design of Advanced Computing Solutions for Cancer initiative, she received the DOE Secretary's Appreciation Award in 2016. In 2017, she received the ORNL Director's Award for Outstanding Individual Accomplishment in Science and Technology and the UT-Battelle Distinguished Researcher Award. In 2020, Dr. Tourassi received the DOE's Secretary Honors Award for her contributions to the COVID 19 Insights Partnership Team and to the COVID 19 HPC Resource Team.

Tourassi holds a B.S. in Physics from Aristotle University of Thessaloniki, Greece, and a Ph.D. in Biomedical Engineering from Duke University.