Impact of the Blue Waters Fellowship Program

Steven I. Gordon
The Ohio State University
Columbus, Ohio
gordon.1@osu.edu

Scott Lathrop
University of Illinois
Urbana, IL
lathrop@illinois.edu

William Kramer
University of Illinois
Urbana, IL
wtkramer@illinois.edu

ABSTRACT
The Blue Waters Fellowship program supported by the National Science Foundation focused on supporting PhD candidates requiring access to high performance computing resources to advance their computational and data-enabled research. The program was designed to strengthen the workforce engaged in computational research. As the program developed, a number of modifications were made to improve the experience of the fellows and promote their success. We review the program, its evolution, and the impacts it had on the participants. We then discuss how the lessons learned from those efforts can be applied to future educational efforts.

Keywords
Computational science education

1. INTRODUCTION
The Blue Waters Graduate Fellowship program was launched in 2013 and continued through 2021 to provide graduate students across the nation with financial support, access to leading-edge petascale computing resources, and technical and scientific support to accelerate their computational and data analytics research. The program was supported with funding for the Blue Waters project [4] by the National Science Foundation (NSF) [13]. The Blue Waters Fellowship program was modeled upon the NSF graduate fellowship program [16] but with an emphasis on supporting PhD candidates needing access to high performance computing (HPC) resources to support their computational and data-enabled research.

The fellowship was designed to educate scientists and engineers in the use of computational modeling and data analytics applied to critical problems across multiple disciplines. The need for researchers and scholars skilled in the use of computational methods has been recognized in a number of national studies as well as in the ongoing Department of Energy Computational Science Graduate Fellowship program [20, 15].

The fellowship provided funding of up to $50,000 for one year to provide tuition and a stipend for each fellow. Each fellow received an allocation of 50,000 node hours (1.6 million core hours) on the Blue Waters supercomputer in support of sustained petascale computations. As the fellowship program matured, it became clear that a number of changes and additions to the original concept needed to be made in order to ensure the success of the fellows and their research projects.

The fellowship program has continued since 2021 and is now named the New Frontiers Fellowship Program [19], with a focus on national security research and funding from NSF. The program continues to evolve and improve based on our experiences.

In this paper, we describe the evolution of the program to provide the support the fellows needed. We then summarize the impacts of the fellowship program on the participating students and discuss the implications of the program for future efforts that aim to improve the pipeline of researchers that utilize high performance computing to advance their research.

2. GETTING STARTED
With the start of each new cadre of fellows, the fellows were invited to the National Center for Supercomputing Applications (NCSA) at the University of Illinois [6] to initiate their year-long fellowship. Their advisors were also invited to attend. This introductory meeting was beneficial for establishing a good working relationship among the fellows and the staff. The staff provided an overview of the NCSA organization, expectations of the fellows, and a technical overview of the computing environment. The fellows provided a summary of their research goals, computational plans, training needs, and anticipated challenges using the Blue Waters system.

The fellows were invited to the annual Blue Waters Symposium at the start of their fellowship to provide posters describing their proposed research and to meet the fellows from the previous cadre and undergraduate interns from the Blue Waters Student Internship program. Towards the end of their fellowship, the fellows were invited back to the next annual Symposium to present their research findings along with the other Blue Waters Investigators. Each fellow was also encouraged to submit a presentation to a conference within their own research domain, and, if needed, travel funding was provided in addition to their fellowship.

Prior to arriving at NCSA, the fellows completed a short survey about their HPC knowledge so that the staff could tailor start-up training on the use of the Blue Waters system. A similar approach was conducted at the annual Blue Waters Symposium in which fellows and other Blue Waters researchers were able to attend short workshops tailored to teaching participants about using the system effectively.

3. TECHNICAL SUPPORT
To make effective use of the computational resources, the fellows needed some training as well as technical and scientific support. The Blue Waters project assigned a primary point of contact (POC) to each fellow, similar to the support provided to the major NSF Petascale Computing Resource Allocations (PRAC) teams [14]. The POCs were professional staff with significant expertise using HPC systems and a diverse mix of science and engineering backgrounds. Many of the POCs had master's degrees or PhDs. We

DOI: https://doi.org/10.22369/issn.2153-4136/13/2/3

worked to match the technical and domain knowledge of the POCs with the research conducted by each fellow.

The POCs were tasked with organizing monthly virtual meetings with each fellow to obtain updates on their progress and to address barriers to progress. The POC level of effort varied across the fellows depending on the students' background and goals. The interactions among the fellows and their POCs included getting software tools working, compiling and setting up job scripts, porting and compiling codes on the Blue Waters system, and optimizing their workflow to get the best throughput on the system. As a result of the selection process, the fellows already had a strong computational background and did not require significant introductory training or technical support. When the POC was unable to address a problem or challenge, the POCs were able to refer the issue to other University staff with the relevant expertise on algorithms, tools, and effective strategies for conducting computations on the Blue Waters system. Even when referring the request to other staff, the fellow's POC was responsible for managing the request to satisfactory conclusion.

Several of the fellows cited the support and training as key to their success. "One of my favorite components of the fellowship, and something that sets it apart from others, is the access to training seminars and workshops led by experts on topics across the high performance computing spectrum," said Lucas Ford, 2021–2022 Fellow. As Salme Cook, 2017–2018 Fellow, put it, "Having a designated person to provide mentoring and guidance about the Blue Waters system broke down social barriers that made the science progress faster than it otherwise would have. That to me was the "X" factor for the Blue Waters experience."

On average, the staff reported a low level of effort to support the fellows. Most of the fellows could be supported with about 1–2 hours of effort per month. There were a few cases in which four to 20 hours of assistance were provided in the first couple weeks during the initial port of their codes. Usually that initial support was the majority of staff time that was required.

The rationale behind this model of support is that the POC provides continuity of support and a holistic awareness of the research goals and challenges from the outset. The POCs viewed no question as inappropriate or too simple. As one staff member said, "the advantage isn't that the fellow is attached to someone that has an enormous amount of time to spend on them, it's that they're attached to someone who already knows their project, knows their background, and doesn't have to be brought up to speed if they have a quick question. The advantage to me is the already loaded background in the POC person, not the totality of their available time to spend."

4. EXTENDING THE EFFORTS

After the meeting with the first cadre of fellows, it became apparent that the initial meeting needed to be extended from one day to two days. Adding a second day allowed the fellows to spend more time with their POC and receive more in-depth advice for getting started on the Blue Waters System. With more time, the fellows were able to talk with other professional staff and make contact with University of Illinois researchers within their field of study.

It became apparent that 50,000 node hours were insufficient for many of the projects, and that one year of access to the system and POCs was not adequate for the fellows to bring their research projects to fruition. Fellows could propose to receive additional allocations on the Blue Waters system based on progress reports and ongoing plans. After the one-year fellowship, fellows could request additional allocations provided they were enrolled in an academic program and reporting significant research progress. For many of the fellows, the follow-on years were used to fine tune their codes and make production runs aimed at completing their PhD research.

5. BUILDING COMMUNITY

The fellows considered forming a community of practice to be very important. The fellows appreciated getting to know their fellow colleagues, the Blue Waters staff, and meeting other researchers attending the Blue Waters Symposium and other conferences. The quarterly meetings among all of the fellows and support staff added to the sense of community along with the welcoming attitude of the technical and research staff ready and willing to assist as needed. The fellows valued the ability to learn from one another and mentioned that some of these friendships lasted far beyond the timeframe of their fellowship.

Several fellows cited the community as critical to their success. "The Fellowship was instrumental in my career both because it gave me hands-on experience on a large-scale HPC system […] but also because the associated workshops gave me the courage and self-understanding to begin to perform significantly more independent research," said Rachael Mansbach, 2017–2018 Fellow. Micheline Soley, 2019–2020, Fellow, said, "Blue Waters was more than a supercomputer; it was a community and a family. […] Blue Waters was a transformative experience that I hope others will have in the future."

The annual Blue Waters Symposium [5] brought together principal investigators, researchers, students, and leaders in computational science and engineering. The participants shared challenges, solutions, and successes in large-scale heterogeneous computing. Along with presentations from the Blue Waters science teams, the symposium featured keynotes from innovative thinkers in science and provided opportunities to share and discuss specific topics of interest. The Blue Waters external evaluators highlighted the interdisciplinary nature of the Symposium as a key benefit to participating students, faculty, developers, and principal investigators.

The Symposium exposed the fellows to common challenges and technological solutions that span multiple disciplines. They were able to network with other researchers through formal sessions and informal gatherings. The ability to display their own posters and make presentations on their research gave them greater confidence in their activities as well as suggestions for improving their methods.

6. IMPACT

A total of 55 graduate students were selected through competitive national application processes. The fellows were selected from among 38 academic institutions across 27 states. Eleven of the institutions are in EPSCoR[1] jurisdictions and one institution is a Minority Serving Institution. Twenty-two of the fellows (40%) identified as female and/or a historically excluded race or ethnicity.

[1] EPSCoR — Established Program to Stimulate Competitive Research, https://beta.nsf.gov/funding/initiatives/epscor

We maintained communication with 96% of the fellows through ongoing communications and surveys to learn of their career progress and to capture their advice on how the fellowship program could be improved. We conducted a final survey of all fellows in March 2022. Eighteen of the fellows reported being in a postdoc position, twelve in a professional position, and eight with a faculty appointment. One of the faculty members reported pursuing tenure approval during 2022. One of the fellows in the 2021–2022 cadre accepted a postdoc position to begin upon completion of their fellowship. One fellow reported ending their academic studies before receiving their PhD. There were two fellows for whom we were not able to ascertain their status. The remainder of the fellows indicated they were working to complete their PhD within the next couple years.

The fellowship had major impacts on the pace and subject of the fellows' research. As George Slota, 2014–2015 Fellow, indicated, "The fellowship allowed me to focus on my research, which allowed me to finish my PhD much sooner and with a greater number of publications than expected; this enabled me to find a professor position immediately after graduating." The work of the fellows enabled their advisors to publish more papers and increase their HPC knowledge.

A total of 4,161,078 node-hours[2] were used by the 55 fellows. Twenty-four (48%) of the fellows used more than 50,000 node-hours either via supplemental allocations or by making use of reduced charge rates; twelve (24%) fellows used more than 100,000 node-hours. Fifteen of the fellows made little use of their Blue Waters allocations, primarily due to having access to other computing systems on their campus and national laboratories. For a petascale system like Blue Waters, this is a very small fraction of the system and could easily be scaled up to support many more fellows.

A number of the fellows mentioned pursuing different research goals and careers than they had originally intended. Based on the feedback from the fellows, they have been tremendously successful in leveraging their fellowships to advance their research, education, and careers. It would be a very rewarding experience to be able to continue to observe their endeavors over time, and to learn how they in turn impact their colleagues, students, and mentees. "The BW fellowship provided me with the opportunity to conduct independent research early on in my PhD. This helped prepare me for life as an independent researcher and greatly contributed to my ability to get my current position without doing a postdoc. I am very confident in saying that the BW fellowship was central to my development as an independent researcher and as a tenure-track faculty," said Jon Cameron Calhoun, 2014–2015 Fellow.

7. DISCUSSION

Given the success of the program, we believe there are opportunities to leverage the lessons learned to benefit the success of other graduate students pursuing research using computational science. We discuss the issues and possible approaches in the subsections below.

7.1 Scalability and Sustainability

The level of funding available for fellowship positions varied considerably each year. The program was very competitive, selecting between 4% and 10% of the applicants. It is clear that there are many more graduate students looking for fellowships than there are available positions. Due to the pent-up demand, the thematic focus of a fellowship can be as broad or narrow as needed to address the host institution's mission and goals. In the final years of the Blue Waters program, an emphasis was placed on applications from students pursuing geospatial intelligence research projects, and the selection process remained equally competitive.

The need and the value of supporting more fellowships is well documented [20, 15]. The nation would benefit greatly from supporting more fellowships. The Texas Advanced Computing Center (TACC) Frontera Computational Science Fellowships and the SC George Michael Fellowships support HPC fellowships [18, 2]. However, without significant funding, it is unclear how any single institution can realistically support a large number of fellowships. A rare exception is the Department of Energy Computational Science Graduate Fellowship (DOE CSGF) managed by the Krell Institute which focuses on Department of Energy mission oriented research [12].

7.2 Extending HPC Support

Replicating the Blue Waters Fellowship will be difficult given the level of resources needed. However, it may be possible to use parts of the fellowship model to provide assistance to graduate students pursuing research in science and engineering. The personnel time to support the fellows, as previously described, was mostly a light load on the staff. The staff generally supported one or two fellows at a time, along with supporting many other researchers conducting petascale class computational projects.

We believe this model of support is a viable option for other organizations. One possible option is to experiment with the POC idea through other means. Technical and scientific support is needed to complement access to leading-edge and emerging technologies and tools. Allocations of computing time at HPC and University computing centers could be accompanied by assignment of a POC for the initial year of a project allocation to ramp up the applications and codes to make effective use of the HPC system. This could be done via a special, additional application by the graduate students working on a project to receive this extra start-up assistance. It could be accompanied by either some mandatory training or confirmation of basic HPC and programming competencies to ensure the students are well prepared to initiate their computational research and have knowledge of whom to contact when they encounter barriers or need extra assistance.

For those students who are selected for such a program, the assignment of a POC who becomes familiar with the project could help advance the progress of the research projects at minimal cost to the organization providing the HPC allocations. A side benefit might be that the codes being used would be more likely to be optimized to the host system, thus preserving scarce and expensive computing resources.

Providing ongoing access to leading-edge HPC resources is critical to application driven research and the ability to pursue complexity at scale. The level of computational resources used by the fellows was found to have negligible impact on the large-scale research projects on a system like Blue Waters. However, if there are limited resources on a campus computing system, education allocations are readily available on the NSF funded systems (e.g., ACCESS [1]).

[2] 133,154,496 core hour equivalents

7.3 Continuing to Build the Community

In our surveys of the fellows, there was a strong sentiment that even more community building should be pursued. We have seen the fellows transform from student to faculty/research member to major contributor within their field of science. It is useful to pair/group graduate fellows so they feel like they're part of a team and can support one another and answer each other's questions. The use of social media can bolster the connections among the fellows and their extended community of peers, advisors, POCs, etc. Regular scheduled meetings provide a useful forum for sharing progress, challenges, and potential solutions across diverse projects and further the process of building community.

Through the fellows program, we have seen that peer mentoring is very beneficial to all of the participants. We have seen that graduate students relate very well to undergraduate students as near-peer mentors, and these collaborations enhance their own confidence and leadership skills.

The knowledge and experience gained from a fellowship places the fellows in a unique position to advise faculty and their research teams about new techniques, tools, and methods. Their home institutions can leverage this knowledge to enhance local practices. A number of fellows shared their fellowship experiences with others within their research team, including students, advisors and researchers. This helped to raise the awareness of computational methods and improved the research teams' methods and approaches. Some of the fellows also conducted outreach sessions on their campuses through local presentations and informal discussions. They created communities of practice within their local community. The fellows were more attuned to sharing news of computational resources and opportunities that could benefit their colleagues.

7.4 Student Preparation

The fellows chosen for the Blue Waters fellowships had a strong research proposal and strong computational skills and experience. To ensure that more students are equally competitive, academic institutions need to place a greater emphasis on student preparation. However, the concepts and skills relevant to HPC research are still not a part of many scientific curricula even as there is a growing need for computational expertise in most disciplines to apply HPC within their research. The evolution of course curricula does not keep pace with the exponential evolution of technology. Workforce preparation remains a persistent challenge with each new generation of students and a rapidly evolving technological infrastructure.

There are many other activities academic institutions can pursue to prepare their faculty, staff, and students to improve workforce preparation. This begins with raising the basic awareness of computational thinking, parallel programming, and quantitative reasoning. Many campuses benefit from local Software Carpentry [17] and Data Carpentry [9] training sessions. The federally funded HPC centers offer numerous training sessions, workshops, hackathons, etc. that can be accessed virtually, at little or no cost to the students. NSF ACCESS [1] provides easy access to education allocations and start-up research projects on leading-edge HPC systems.

Course sharing between institutions [10] has been shown to be an effective strategy to bring courses to campuses that could not otherwise be offered. In addition to expanding access to more advanced topics, the faculty who are involved can gain confidence to teach the content on their own. HPC centers and professional societies [3, 11] can provide advice and guidance to assist campuses in bringing needed formal and informal courses to campuses.

Campus activities may also include working to host curriculum development workshops with their faculty to incorporate existing computational and data analytics modules into the mainstream educational system. Campuses can introduce informal computational science and data analytics training events offered by Software Carpentry and HPC centers at minimal cost. Establishing research and education partnerships with other institutions and industry will open new opportunities such as fostering internships and fellowships locally and through other organizations.

Every effort should be made to provide students in all disciplines with access to HPC resources with minimal barriers to access, learning, and achievement. Faculty and staff need to infuse new tools and methods for conducting computational methods. Workshops, internships, fellowships, and mentoring are essential to complementing formal coursework. Collectively, these efforts impact student retention and motivation to continue learning when they address authentic real-world problems and challenges.

Finally, a regular series of seminars and workshops will expand informal student learning opportunities. Upon returning from a conference or other similar event, students should be encouraged to share what they learned with fellow students and advisors to potentially incorporate new tools and methods into local practices.

8. CONCLUSIONS

There are many programs that provide in-depth preparation of talented individuals for advanced careers. However, the demand for such programs far exceeds the available resources and programs to give all students this level of engagement.

Academic institutions need to pursue strategies to prepare more researchers able to effectively use HPC technologies and advanced algorithms to advance discovery and scholarship. There are HPC centers, professional societies, consortia, and other organizations committed to working with institutions [11, 7, 8] to address these needs with practical solutions, while also developing scalable and sustainable approaches that work for very diverse institutions and organizations. Persistent attention is needed to ensure that all students are provided with opportunities for engagement in these programs and activities.

Using some of the lessons learned from the Blue Waters Fellowship program can help to encourage both new formal teaching and learning efforts as well as informal programs that help build a cadre of STEM students pursuing computational science related research projects. The ongoing challenge is to raise the awareness of successful workforce development methods so that all academic institutions, and thus all students, have the potential to advance discovery using current and emerging technologies.

9. ACKNOWLEDGEMENTS

The Blue Waters Graduate Fellowship program is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) the State of Illinois, and as of December, 2019, the National Geospatial-Intelligence Agency. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

10. REFERENCES

[1] ACCESS. 2022. Home - Access. Retrieved November 2022 from https://access-ci.org/

[2] ACM, Inc. 2022. ACM-IEEE CS George Michael Memorial HPC Fellowships. Retrieved November 2022 from https://awards.acm.org/hpc-fellows

[3] ACM SIGHPC Education Chapter. Retrieved November 2022 from https://sighpceducation.acm.org/

[4] Board of Trustees of the University of Illinois. 2022. Blue Waters User Portal. Retrieved November 2022 from https://bluewaters.ncsa.illinois.edu/

[5] Board of Trustees of the University of Illinois. 2022. Blue Waters User Portal | Blue Waters Symposium. Retrieved November 2022 from https://bluewaters.ncsa.illinois.edu/blue-waters-symposium

[6] Board of Trustees of the University of Illinois. 2021. Home - NCSA. Retrieved November 2022 from https://www.ncsa.illinois.edu/

[7] Campus Champions. 2022. Home | Campus Champions. Retrieved November 2022 from https://campuschampions.cyberinfrastructure.org/

[8] Campus Research Computing Consortium (CaRCC). About – carcc.org. Retrieved November 2022 from https://carcc.org/about/

[9] Data Carpentry. Retrieved November 2022 from https://datacarpentry.org/

[10] Steven I. Gordon, James Demmel, Lizanne Destefano, and Lorna Rivera. 2015. Extending Access to HPC Skills Through a Blended Online Course. In *XSEDE '15: Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*. July 26, 2015, St. Louis, Missouri. Association for Computing Machinery, New York, NY, USA, 1–5. DOI: https://doi.org/10.1145/2792745.2792760

[11] IEEE. 2018. Technical Community on High Performance Computing - IEEE Computer Society Technical Consortium on High Performance Computing. Retrieved November 2022 from https://tc.computer.org/tchpc/

[12] Krell Institute. 2022. Front Page | Krell Institute. Retrieved November 2022 from https://www.krellinst.org/

[13] NSF - National Science Foundation. Retrieved November 2022 from https://www.nsf.gov/

[14] The National Science Foundation. Petascale Computing Resource Allocations (PRAC) (nsf17542) | NSF - National Science Foundation. Retrieved November 2022 from https://www.nsf.gov/pubs/2017/nsf17542/nsf17542.htm

[15] National Science Foundation Blue Ribbon Panel on Simulation-Based Engineering Science. 2006. Simulation – Based Engineering Science: Revolutionizing Engineering Science through Simulation. (May 2006). Retrieved November 2022 from https://www.nsf.gov/pubs/reports/sbes_final_report.pdf

[16] NSF Graduate Research Fellowships Program (GRFP). Retrieved November 2022 from https://www.nsfgrfp.org/

[17] Software Carpentry. Retrieved November 2022 from https://software-carpentry.org/

[18] Texas Advanced Computing Center, The University of Texas at Austin, Office of the Vice President for Research. 2020. Frontera Computational Science Fellowships. Retreived November 2022 from https://frontera-portal.tacc.utexas.edu/fellowship/

[19] University of Illinois Board of Trustees. 2022. New Frontiers Fellowships – New Frontiers Initiative. Retrieved November 2022 from https://newfrontiers.illinois.edu/opportunities/new-frontiers-fellowships-2023-2024/

[20] John Wells, Holly Bozeman, Brad Keller, Atsushi Miyaoka, Vasiliy Sergueev, Adrienne von Glatz, and Kerri Wills. 2021. U.S. Department of Energy: Computational Science Graduate Fellowship 1991–2021: A follow-up study of recipients and programmatic outcomes. (Dec. 2021). Retrieved November 2022 from https://www.krellinst.org/csgf/study