NSF CISE AC Subcommittee on Growing and Diversifying the Domestic Graduate Pipeline

Final Report - June 1, 2021

Committee Co-chairs: Marie desJardins and Ran Libeskind-Hadas
Committee Members: Magda Balazinska, Susanne Hambrusch, Charles Isbell, Muriel Medard

At its December 2020 meeting, the CISE Advisory Committee discussed the issues of graduate enrollments and the diversity of the domestic graduate student population. An ad hoc CISE AC sub-committee was charged with continuing this discussion and, ultimately, identifying the most salient questions to be asked, as well as potential sources and methods for addressing these questions.

1. Introduction

The vitality of the pipeline of domestic students who pursue PhDs in computer science and related fields is an area of serious concern to the computing community with a concomitant impact on our national well-being and security. While undergraduate enrollments in computing have increased dramatically in the last decade, the number of domestic[1] students who chose to pursue doctoral degrees over that period has remained relatively flat [Hambrusch 2020]. The historical reliance on international PhD students may not be sustainable as an increasing number of outstanding international students choose to pursue their studies and careers in other countries. Moreover, some areas of computing research have national security implications and thus restrict funding to domestic students.

We have only a limited understanding of the factors that contribute to a student's decision to pursue graduate studies in computer science. Compelling undergraduate research experiences appear to have a positive impact [Tamer2021] while the very healthy job market for new college graduates in computing fields appears to be a force in the other direction. A recent study shows that 66% of students who aspired to a master's degree in their junior year and 37% of those aspiring to a PhD did not actually apply in their senior year. The most salient reported factors were wanting to get a job (87%), taking a break from school (53%), and worries about financial support (25%) [Wright2021].

Universities and industry are competing for a small pool of new PhDs. Approximately 60% of new computing PhDs take jobs in industry [ZwiBi19] and many academic departments are

[1] We use the term "domestic students" to refer to U.S. citizens and permanent residents of the U.S.

unable to fill faculty positions and retain their current faculty, particularly those in areas of greatest interest to industry.

The diversity of the computing PhD workforce is an area of particular concern. Although data is only incompletely reported, Black and Hispanic/Latinx students are underrepresented in PhD degrees by a factor of three to five, relative to their representation in the population as a whole [Straub2021; Taulbee]. Women represent roughly 25% of the computing PhD population and that level has remained relatively constant for nearly three decades [Straub2021]. These disparities contribute to the underproduction of computing professionals and result in narrower perspectives in computing research and development. Moreover, they inhibit an important vehicle for economic equity for traditionally underrepresented groups.

Finally, the baccalaureate origins of the majority of domestic PhD students are attributed to a relatively small number of institutions. A 2013 study [Hambrusch 2013] showed that approximately 50% of domestic PhD students come from only 54 institutions of baccalaureate origins, primarily from the most elite and selective colleges and universities. Moreover, between 2018 and 2021, 45% of NSF Graduate Research Fellowships from CISE went to students whose undergraduate degrees were from just ten schools. Many talented students are likely not receiving the guidance or undergraduate research opportunities that help pave the way for entry to PhD programs. Additionally, selection processes for graduate admissions and fellowships may have systemic biases that, ultimately, are detrimental to both individual students and to the health of the research community overall.

In Section 2, we formulate a set of questions that we believe should be addressed in order to help design and implement effective interventions. In Section 3, we offer some potential sources of data for addressing these questions. We recommend that a systematic investigation of the questions posed in Section 2 be conducted using sources including those in Section 3. In Section 4, we suggest some possible interventions whose viability and efficacy should be assessed in light of the proposed study.

2. Questions: Characterizing the Problem

Implicit in much of the discussion around graduate enrollments is the widespread understanding in the computing community that the pipeline is too small; that is, that there are simply not enough PhD students to fulfill the needs and demands generated by industry, academia, and others. In order to make concrete and meaningful recommendations, it will be necessary to understand this supply-demand gap more clearly.

There is abundant evidence that the PhD pipeline is substantially less diverse in terms of gender, race, and nationality than the undergraduate population of the country as a whole. We can readily understand these demographic characteristics at different stages of the pipeline: undergraduate majors, undergraduate degrees granted, graduate student populations, and PhD

recipients. What is less well understood are the primary influences on the demographics of the pipeline: what are the key factors in determining who attends graduate school? We also know relatively little about other important characteristics such as disability status, geographic distribution across the country, and the types of undergraduate institutions attended.

To that end, we pose the following questions. For each question, we recommend examining how the “answer” has changed over the last decade and trends for the near future.

1. What is the demand for computing PhDs from industry, academia, government, and others?
2. To what extent do PhD programs need more PhD students to conduct research and to serve as teaching assistants?
    a. How does the demand differ by type of institution, geographic location, departmental ranking, and other factors?
    b. How does the demand differ by subdiscipline?
3. What are the “sources” of PhD students?
    a. What fraction of students matriculate directly from undergraduate studies, master’s programs, and industry?
    b. What are the acceptance, yield, and retention rates in PhD programs and how do they vary by student demographics, type of baccalaureate institution from which the student graduated (e.g. selectivity, public/private, R1/master’s/PUI, HBCUs), and type of PhD program to which the student applies?
    c. To what extent do departments draw from their population of master’s students in filling their PhD programs?
4. What are the primary factors that contribute to an undergraduate deciding whether or not to pursue a PhD and how are these factors correlated with gender, race, ethnicity, familial context (e.g., parents’ level of education and country of origin), and geography (e.g., regional differences or urban/rural disparities)? Possible factors might include:
    a. Financial considerations: some students may have immediate financial needs and obligations; others may view immediate employment as a better financial decision; and others may take longer-term salary prospects more heavily into consideration
    b. Non-financial opportunity costs: the prospect of getting a job immediately after graduation from college mitigates risks incurred by waiting
    c. Availability of compelling undergraduate research opportunities
    d. Degree of awareness of what a PhD entails and the opportunities that a PhD provides
    e. Support and encouragement from parents, mentors, and others
5. What factors lead to success in a PhD program? Do these factors vary across demographic groups? Possible factors might include:
    a. High-quality mentoring
    b. High-quality advising
    c. Level and type of financial support
    d. Positive peer group environment

e. Positive and inclusive climate in the department
f. Types of qualifying and preliminary exams

3. Data: Assessing the Problem

There are a number of sources that may help address these questions.

1. NSF IPEDS, the Integrated Postsecondary Education Data System, provides data from surveys conducted by the National Center for Education Statistics.
2. The National Student Clearinghouse provides data on K-20 student outcomes.
3. The CRA Taulbee Survey provides data on computing enrollments and graduation at the undergraduate and graduate level including demographic information. The survey asks some questions that are not reported in the annual report, but are available in the full data download, which may be relevant to addressing the questions posed in the previous section. The Taulbee Report focuses on member institutions and may provide a limited view of non-research-focused undergraduate institutions.
4. HERI, the Higher Education Research Clearinghouse and Cooperative Institutional Research Program, provides data on undergraduates and has a large repository of reports that interpret the collected data.
5. The CERP Data Buddies project conducts surveys of undergraduates, graduate students, and faculty to assess past experiences, knowledge and confidence, sense of belonging, and many other factors that contribute to deciding whether or not to pursue a career path in computing research.
6. NSF Graduate Research Fellowships Program (GRFP) data is likely to shed light on the diversity of the recipients of that award and may help determine whether additional fellowships are warranted (and how to target those additional awards).
7. Although data on graduate admissions is not readily, widely, or consistently available, a few institutions do provide some public data on graduate admissions. For example, Purdue University provides a Data Dashboard that includes acceptance rates with some demographic information.

A number of published articles use these sources to address some of the issues enumerated in Section 1. Some examples include:

- J. Straub, "Analysis of the Changing Demographics of Computing Doctoral Degree Recipients", *ACM Inroads*, March 2021. This paper synthesizes a number of datasets to provide insights into the demographics of CS PhD students.
- S. Hambrusch *et al.*, "Exploring the Baccalaureate Origin of Domestic PhD Students in Computing Fields." 2013. This CRA whitepaper explores the origins of domestic PhD students.
- S. Hambrusch *et al.,* "Addressing the National Need for Increasing the Domestic PhD Yield in Computer Science", CRA Quadrennial Paper, November 2020.

- Burçin Tamer, "REU Participation Encourages Students to Pursue Graduate Degrees," *Computing Research News* 33(3): March 2021

4. Interventions: Addressing the Problem

Appendix A identifies a number of current "best practices" that are being pursued by institutions to grow and expand the pipeline. Here we outline a few key potential interventions that we believe will have the highest impact in the near term. The final one emphasizes the need for continued data collection and analysis.

Intervention 1: Increase Awareness and Outreach.

In many cases, students—especially members of underrepresented groups and students at smaller institutions—do not have adequate information about the PhD pathway. Information should be readily available for students to learn about PhD programs, how to decide where to apply, how to develop a strong application, what being a PhD student is like, and career pathways for doctoral graduates. CRA's Committee on Widening Participation already runs graduate cohort events that offer many of these materials, but NSF, CRA, and PhD institutions could partner to develop and disseminate valuable informational materials. Research-intensive universities could be encouraged to develop relationships with colleges and Primarily Undergraduate Institutions in their region, facilitating mentoring, research opportunities, and even pathways for graduate admission for undergraduates at those institutions. The REU participant survey may provide especially valuable data to ensure that student outreach meets the needs of potential PhD students. Useful resources in this area are available from Black in AI, AccessComputing, and CRA's CONQUER project.

Intervention 2: Encourage Recruitment of Diverse Students to PhD Programs.

The NSF should examine its graduate fellowship programs and explore ways to support a diverse group of students, including those from institutions with limited undergraduate research opportunities. Direct funding to students may incentivize graduate admissions committees to consider students whom they might otherwise overlook.

Additionally, support for students with financial need, such as relocation grants to enable students to move to a new locale to pursue graduate school, could be a low-cost but high-impact intervention. Undergraduate student debt is a critical challenge for many members of underrepresented groups and exploring ways to forgive or significantly reduce debt could have substantial impact.

While undergraduate research experiences are valuable in increasing students' likelihood of pursuing graduate studies, the trend in some admissions committees to expect undergraduate research publications significantly disadvantages many capable applicants. As long as the playing field is not level for accessing undergraduate research experiences, emphasizing

research as a precondition for acceptance will continue to exacerbate disparities. Pathways to doctoral studies should be available for students without research experience, students who majored in non-CISE areas, master's students, students with industry experience who are returning to graduate school, and veterans of the armed services. NSF can play a part by sponsoring workshops and working groups examining best practices for graduate admissions and for balancing experience with potential in the admissions process.

Departmental Broadening Participation (BP) statements should be required as part of grant proposals and those proposals should include evidence of success in those BP efforts.

Examining GRFP student statistics (disaggregated by gender, race, and institution type) may be instructive in understanding where additional investments could help to grow the pipeline. We also encourage NSF to incentivize departmental reporting of admissions data (applicants, accepted students, and matriculated students by gender, race, and undergraduate institution type) and retention data. NSF should encourage/request such reporting for all CISE departments and require regular reporting for departments who receive NSF funding.

Intervention 3: Expand Access to Early Research Experiences.

NSF should pursue multiple avenues to incentivize and support high-quality early research experiences for undergraduates. The REU survey may help to identify mechanisms for expanding the program and/or maximizing impact, including providing resources for faculty mentors. For example, a recent CRA-E whitepaper offers concrete strategies for departments to provide scalable undergraduate research programs [Alvarado et al. 2020]. Industry partnerships can also play a role in providing sponsorships for undergraduate research, as well as research-oriented internship experiences.

Intervention 4: Facilitate Collaborations with Industry to Support Graduate Study

Cooperation between industry, academia, and government in encouraging and supporting advanced graduate studies can benefit the entire research community. For example, the NSF has recently announced the CSGrad4US program to help support graduate studies for individuals who went to industry after college and are thus not eligible for the NSF GRFP. As another example, Google's exploreCSR program provides funding to students from traditionally underrepresented groups to pursue advanced graduate studies. We recommend exploring NSF-industry partnerships that would incentivize such programs while connecting them with the needs of graduate students and programs.

5. Recommendations and Next Steps

There is considerable evidence that the health of the domestic PhD pipeline in computing is at risk. International students are finding increasingly attractive options in their home countries while domestic undergraduate students are attracted by high-paying and compelling jobs in industry. Students from traditionally underrepresented groups are even less represented in

computing than in other STEM fields, and this disparity is exacerbated for all underrepresented groups at the PhD level. We make several recommendations for next steps:

1. Colleges and universities should:
    a. Seek to inform undergraduates about graduate school and careers in computing research.
    b. Provide high-quality undergraduate research opportunities at their own campuses and/or advising on research opportunities at other institutions; including through collaborations between research-intensive universities and proximate primarily undergraduate institutions.
2. Graduate programs should:
    a. Collect and publish aggregate data on their admissions and retention, including demographic information, towards the objective of this effort.
    b. Formulate and refine broadening participation statements supported by evidence of past success.
    c. Examine their graduate admissions processes to determine if they mitigate potential sources of bias and provide pathways for high-potential applicants from institutions that are not typically represented.
    d. Nurture supportive departmental cultures to enhance retention and well-being of their graduate students.
3. Industry should:
    a. Provide more research-oriented undergraduate internship opportunities.
    b. Provide more pathways for employees to pursue graduate education through fellowship programs and leaves of absence.
    c. Collaborate with academia to provide opportunities for graduate students and faculty to move between those environments including graduate programs that interleave time at the university with internships in industry.
4. The NSF should:
    a. Collect and publish innovative and effective practices used by various institutions to encourage their undergraduate students to pursue graduate studies and to admit, attract, and retain a diverse group of graduate students. Examples of a number of such practices are included in Appendix A, but we believe that a more comprehensive list would be valuable.
    b. Explore expansion of its graduate fellowship programs including offering more fellowships and developing selection processes that are mindful of students from underrepresented groups and from institutions that are not typically represented.
    c. Facilitate a comprehensive study that seeks to address the questions posed in Section 2 using data sources such as those enumerated in Section 3. That study may be most appropriately performed by another group such as the Computing Research Association or the National Academies.
    d. Use the findings from this comprehensive study to assess the viability and potential efficacy of the interventions suggested in Section 4.

References and Related Work

Christine Alvarado, Michael Hilton, Amy J. Ko, Lori Pollock, Kelly Shaw, Best Practice Report on Broadening Research Experiences for Undergraduates. CRA-E Report, 2020.

Susanne Hambrusch, Lori Pollock, Ran Libeskind-Hadas, Christine Alvarado, Addressing the National Need for Increasing the Domestic PhD Yield in Computer Science, CRA Quadrennial Paper, November 2020.

Susanne Hambrusch, Ran Libeskind-Hadas, Eric Aaron, Understanding the U.S. Domestic Computer Science Ph.D. Pipeline, *CACM* 58(8): July 2015.

Susanne Hambrusch and Ran Libeskind-Hadas, The PhD Pipeline, *IEEE Computer*, vol. 48, no. 5, pp. 76-79, May 2015.

Susanne Hambrusch, Ran Libeskind-Hadas, Fen Zhao, David Rabson, Amy Csizmar Dalal, Ed Fox, Charles Isbell, Valerie Taylor et. Exploring the Baccalaureate Origin of Domestic Ph.D. Students in Computing Fields. *Computing Research News* 25(1): 2013.

Tracy Camp, W. Richard Adrion, Elizabeth Bizot, Susan Davidson, Mary Hall, Susanne Hambrusch, Ellen Walker, Stuart Zweben. Generation CS: The growth of computer science. *Inroads* 8(2): 44-50, *Inroads* 8(3): 36-42, *Inroads* 8(4): 59-65 (2017). Report on CRA website.

Computing Research Association, Taulbee Survey.

Jeremy Straub, Analysis of the Changing Demographics of Computing Doctoral Degree Recipients at U.S. Universities and Implications of Change, *CACM Inroads* 12(1), March 2021.

Heather Wright, Senior Undergraduates Who Did Not Apply to Graduate School Waited Because of a Job, but Most Are Likely to Apply in the Future, Computing Research News 33(4), April 2021.

Appendix A: Ongoing Interventions

In this section, we review what several universities are currently doing to (1) increase the national pipeline of graduate students and (2) improve DEI in the graduate pipeline. This list is intended to be indicative of good practices and is not comprehensive.

These interventions are being implemented to address **growing the national pipeline**:

- Structured undergraduate research experiences, such as explicit research courses
- Regular information sessions on CS graduate school and its different flavors

- Systematic departmental processes for nominating students for the CRA Undergraduate Research Award
- Initiatives to send graduate students back to their undergraduate institutions to give talks about their experiences in graduate school and to foster relationships with nearby colleges and liberal arts colleges in general
- Summer research programs for high school students

These current initiatives are specifically designed to address **diversity, equity, and inclusion** in the graduate pipeline:

- Hiring of a graduate program recruitment and retention specialist, focusing on members of underrepresented groups
- Outreach and recruiting efforts for prospective graduate students, including a regular presence at the CMD-IT/ACM Richard Tapia Celebration of Diversity in Computing and at the Grace Hopper Celebration of Women in Computing
- Targeted outreach to high-potential undergraduates at HBCUs and other minority-serving institutions, with a particular focus on reaching students who may not be considering graduate school
- Admissions rubrics for initial applicant screening, designed to avoid prematurely filtering students from schools with limited research opportunities
- Admissions subcommittees that pay particular attention to students with unconventional backgrounds
- Panel sessions on graduate school involving multiple departments and held at forums such as SIGCSE and Tapia
- A pre-application review and mentoring program for prospective applicants; the applicants are matched with current students who coach them through the application process
- Application fee waiver program for qualified applicants
- Efforts to build personal relationships with prospective students identified during outreach
- Early arrival program for entering PhD students to establish community and begin research prior to first semester of graduate school

These interventions are related to **graduate student retention**:

- Increased mental health support:
    - Advising team
    - Workshops
    - Organized peer support
- Community-building initiatives
    - Affinity groups for URMs, women, LGBTQ+ students, and students with disabilities
    - Peer mentoring programs that include a mentor training component
- Career development workshops