Representation of Women in High-Performance Computing Conferences

Eitan Frachtenberg and Rhody Kaner

February 28, 2020

Representation of Women in High-Performance Computing Conferences

Eitan Frachtenberg
Department of Computer Science
Reed College
Portland, Oregon
eitan@reed.edu

Rhody Kaner
Department of Computer Science
Reed College
Portland, Oregon
rkaner25@gmail.com

***Abstract*—Women are severely underrepresented in the HPC workforce. Following the aphorism "if you can't measure it, you can't improve it," this paper aims to provide tangible and reproducible data on this gender gap. Specifically, this paper provides statistics on women's representation in HPC conferences, focusing mainly on authors of peer-reviewed papers, who serve as the keystone for future advances in the field.**

To this end, we analyzed participant data from nine HPC and HPC-related peer-reviewed conferences from a single year. In addition to examining gender distributions, we looked at other demographic factors, such as the authors' countries and sectors, and at post-publication statistics of the papers.

Our main finding is that women represent only about 10% of all HPC paper authors, with some geographical variations. Representation of women is particularly low among industry researchers and at higher experience levels.

TABLE I
HPC-RELATED CONFERENCES, INCLUDING START DATE, NUMBER OF PUBLISHED PAPERS, TOTAL NUMBER OF PUBLISHED AUTHORS, ACCEPTANCE RATE, AND COUNTRY CODE

Conference	Date	Papers	Authors	Acceptance	Country
CCGrid	2017-05-14	72	296	0.252	ES
IPDPS	2017-05-29	116	447	0.228	US
ISC	2017-06-18	22	99	0.333	DE
HPDC	2017-06-28	19	76	0.190	US
ICPP	2017-08-14	60	234	0.286	UK
EuroPar	2017-08-30	50	179	0.284	ES
SC	2017-11-14	61	325	0.187	US
HiPC	2017-12-18	41	168	0.223	IN
HPCC	2017-12-18	77	287	0.438	TH

I. Motivation

The gender gap in different fields of study is an active area of research [1]–[3], but some subfields remain poorly examined [4]. One example is High-Performance Computing (HPC), a large and vibrant research field with significant impact on the economy and the sciences, including computer science (CS). Despite the importance of the field, less than 17% of its workforce is estimated to be women, although the exact number is difficult to ascertain [3], [5]. The focus of this paper is to reduce some of this uncertainty by computing reproducible metrics, specifically on one important population of the HPC community, namely, HPC researchers. Although this paper cannot fully address the complex question of why women are so severely underrepresented in HPC, it attempts to quantify this underrepresentation, with two goals in mind: to faciliate understanding of its causes and to establish a baseline against which to measure the effects of future efforts and policies to increase gender diversity.

II. Methodology and Data

As in most subfields of CS, the primary channel for publishing research results in HPC is peer-reviewed conferences [6]–[9]. For this study, we collected extensive data from nine HPC-related conferences that took place in 2017 (Table I).

The largest of these conferences, SC, takes place in the US every November and attracts thousands of attendees. SC is also notable for its diversity initiatives, including an inclusivity chair, on-site childcare, a code of conduct, and diversity-related content. The other flagship HPC conference, ISC, takes place in Germany every summer and also has a diversity chair. The seven other conferences we investigate are not exclusive to HPC, but contain numerous HPC papers. None of them displayed diversity initiatives for 2017.

For each of these conferences, we downloaded all papers and gathered information about all authors, program committee (PC) members, and other roles. The most crucial piece of our data is gender information. Given the full name and affiliation of most researchers, we searched online for each person's pronouns or pictures to infer their *perceived* gender.

This approach has several limitations. First, gender identity is complex and non-binary, and yet as in many similar studies, we are confined to representing gender as a simple binary choice because of the available data [2], [10]–[14]. Second, not all researchers presented unambiguous Web presence or genders. For those 132 out of 4234 persons for whom we couldn't confidently infer gender from the Web or their name, we marked their gender as NA and omitted them from most analyses, which could introduce a small bias. Finally, this approach is labor-intensive, but our experiments suggest that it is more accurate and less biased than automated approaches based on name alone, especially for non-European names [15].

For reproducibility, all of the data and source code files for this paper can be found at https://github.com/eitanf/sysconf/.

III. Results

We organized our statistical observations into three perspectives: the conferences, the papers, and the researchers.

A. Conferences

We analyze the conferences in our set through the lens of the ratio of women in different roles (Fig. 1). In this section, we look at women across all conferences and papers in our set, and later we examine HPC papers only.

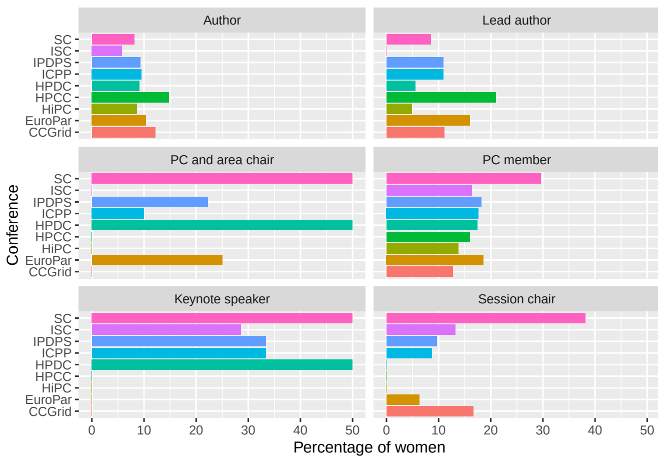

Fig. 1. Representation of women across six conference roles

1) Authors: We start by looking at the ratio of female authors and of first authors. Since these roles are peer-reviewed and ideally selected from the overall author population based on the merit of their papers alone, this sample population may be the most representative of the overall female population across publishing HPC researchers. The result, if indeed indicative of the overall population, is alarming. Only 10.06% of all 2236 authors are women, far lower than the already low 20–30% for the rest of CS [2], [16], [17]. It is particularly dismal for the HPC flagship conferences, with only 8.12% female authors in SC and 5.77% in ISC (excluding the few authors for which we have no gender information).

It's possible these results are biased. For example, survivor bias [18] or peer-review bias [19] would lead to a higher percentage of women among submitted papers than the percentage we observe across accepted papers. However, variations in the visibility of the author's gender do not support this hypothesis. For example, SC and ISC are the only double-blind conferences in our dataset, where the identity of authors are hidden from reviewers, and yet these conferences show the least female representation. And the ratio of female lead authors, where their gender could potentially be more prominent, is similar to other author positions.

2) Program Committee: Next, we examine gender representation among two of the most important elected roles of conference participants, the PC chairs and members. These roles have direct bearing on the technical content and author selection of a conference, and therefore have strong influence over the author population. Most conferences only have a handful of PC and area chairs, so it is difficult to draw statistical conclusions from this role. Nevertheless, it is instructive to note that four of the nine conferences elected no female chairs whatsoever.

Among the 1220 total PC members (with repeats), 18.54% are women, about twice the rate among authors. Even without the larger SC conference, the ratio is still 16.2%. Conceivably, women are not as poorly represented in this population because conferences often aspire to intentionally increase diversity, which can directly lead to selecting more women for this role. SC's example supports this explanation, with its explicit push for diversity. Then again, the other conference with a diversity chair, ISC, had a more average ratio of women in the PC. But even if these ratios are more representative of the overall female HPC population than the authors statistics, they are still far from equitable, and are ostensibly insufficient to increase the ratios among authors.

3) Visible Roles: Finally, we look at two more conference roles: keynote speakers and session chairs. These roles are only collateral to the main technical content of a conference. But they can be very visible to conference attendees, and as such, can represent the "face" of the conference [20]. A conference with low female visibility in these roles could conceivably further deter women from joining or staying in the conference or field. Like PC chairs, keynote speakers are few, but again we see four conferences with no women. Perhaps worse still, three conferences had no female session chairs whatsoever, out of a total of 45 session chairs. Only SC shows a ratio that is approaching equitability, again perhaps because of its explicit push for diversity and inclusivity.

B. Papers

1) Topic: Not all papers in our dataset relate directly to HPC. To limit our discussion only to HPC papers, as opposed to HPC conferences, we skimmed each paper and tagged it as "HPC" if its topic related directly to high performance hardware or software.[1] The resulting list of 178 papers is approximately a third of the complete list of 518 papers in our 9 conferences and can help us determine what the representation of women looks like among paper authors when we restrict ourselves to only HPC research.

Of the 809 authors of HPC-only papers with known gender, 10.14% are women, almost identical to the 10.06% in the overall conference author population. When we look specifically at lead authors, the picture is similar: of the 172 papers for which we know the first author's gender, 20 are women (11.63%), a statistically insignificant difference from the overall ratio ($\chi^2 = 0.149$, $p = 0.6995$).

These results suggest that gender representation among these HPC authors is in line with that of the conferences' representation overall. Subsequently, we will return to analyzing the complete set of papers throughout the rest of this study, for greater coverage.

[1]The determination of this tag was based on our experience in HPC research and is therefore subjective and error-prone.

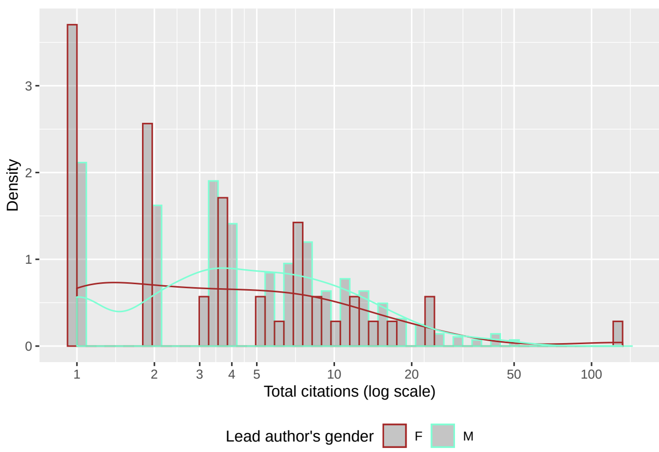

Fig. 2. Distribution of paper citations two years after publication

2) Reception: One way to measure a paper's influence is through its citations over time. We have allowed our data set to age to the point where all papers could be discovered and cited by other researchers. The density plot in Fig. 2 shows the distributions of numbers of citations of the papers 24 months after publication, broken up by the gender of the lead author.

Many of the 55 papers with a woman as lead author are concentrated on the left side of the chart, with fewer citations than the 434 papers authored by men. Women average more citations than men (6.95 vs. 6.08). However, there is one outlier at the long tail of the distribution. This paper, lead by a female author, garnered more than 100 citations, but does not appear to be strongly related to HPC [21]. When we exclude this paper, the mean citations for women's papers drops to 4.63, siginficantly lower than men's ($t = -1.72$, $p = 0.089$). There appears to be also a "ceiling" around 3 citations or so, with only 46% of female-led papers exceeding this threshold, compared to 65% for men ($\chi^2 = 6.4$, $p = 0.011$).

C. Researcher Demographics

There are a total of 3456 authors and PC members in our dataset, spanning numerous countries and institutions. We can estimate the distribution of these demographics by looking at email addresses or Google Scholar (GS) affiliations. Using simple regular expressions, we mapped most domains to an institution's country and sector, and marked the rest as NA.

1) Geography: Unsurprisingly, most of these researchers hail from the West, although China, Japan, and India are also prominent (Table II). The United States has the highest percentage of women of any country with more than 15 researchers in this list, although it is still far from gender equality. On the opposite end, Japan stands out among developed countries with particularly low female representation.

Another way to break down these numbers is by geographic region and conference role (Table III). This perspective confirms the stark lack of geographic diversity in the field, with a full half of the 1789 identified authors associated with US email addresses, and another 14.25% from Western Europe. It is encouraging to note, however, that Western reviewers are not over-represented compared to authors, as has been observed in journals in other fields [22]. For example, the percentage of PC members from the US or Western Europe are similar to the authors' (52.57% and 16.36%, respectively).

In terms of representation of women, none of the larger regions deviates much from the $\approx 10\%$ overall percentage. The regions with fewer than 25 authors or so exhibit more variance, due to the small denominators (for example, a single female author from Eastern Europe would have bumped the percentage from zero to $\approx 8\%$). These conclusions are limited, however, by the small sample sizes for most regions.

2) Sector: From authors' affiliations we can broadly categorize their sector as either "COM" for industry (8.7% of total), "EDU" for academia, (72.8%), and "GOV" for government and national labs (18.5%). Fig. 3 depicts the percentage of women in each sector, by role. It shows that among PC members, there is better female representation in government and academia than in industry. This result agrees with past studies that found relatively fewer female engineers in industry research positions [4], [23], [24].

TABLE II
TOP TEN COUNTRIES BY NUMBER OF RESEARCHERS

Country	% Women	Total
United States	15.44	1410
China	9.88	201
France	13.61	147
Germany	8.63	139
Spain	9.76	123
India	5.63	72
Switzerland	14.06	64
Japan	1.59	63
United Kingdom	9.62	52
Canada	6.82	44

TABLE III
REPRESENTATION OF WOMEN BY REGION AND ROLE

	Authors		PC members	
Region	% Women	Total	% Women	Total
Australia and New Zealand	8.33	24	0.00	14
Central America	100.00	1	NA	NA
Central Asia	0.00	1	NA	NA
Eastern Asia	11.44	201	2.90	69
Eastern Europe	0.00	12	11.76	17
Northern Africa	0.00	1	NA	NA
Northern America	9.88	931	24.47	523
Northern Europe	9.23	65	8.00	50
South America	8.33	36	27.27	11
South-Eastern Asia	5.00	20	0.00	4
Southern Asia	6.35	63	5.00	20
Southern Europe	8.49	106	13.75	80
Western Africa	50.00	2	NA	NA
Western Asia	27.27	22	12.50	24
Western Europe	9.02	255	16.35	159

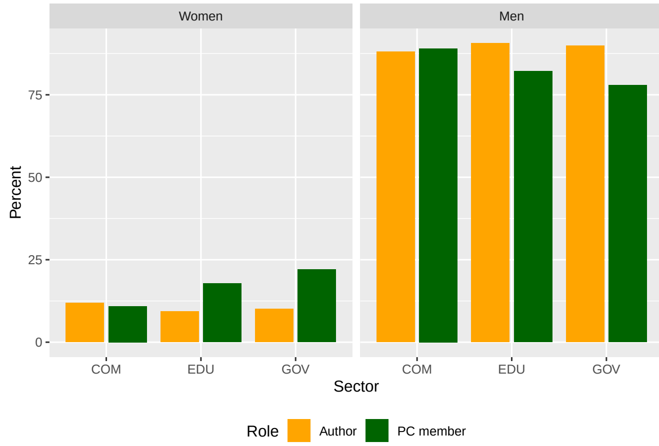

Fig. 3. Representation of women by sector and role

3) Experience: Another statistic we can evaluate from this dataset is the previous research experience of researchers. We can approximate this experience by identifying the unique GS profile of researchers whenever possible (in our data, 69.47% had a uniquely identifiable GS profile). The profile contains various bibliometric measures that can approximate the experience and research influence of a person. For example, Fig. 4 depicts the distribution of one such metric, the total number of previous publications (circa their conference's date).

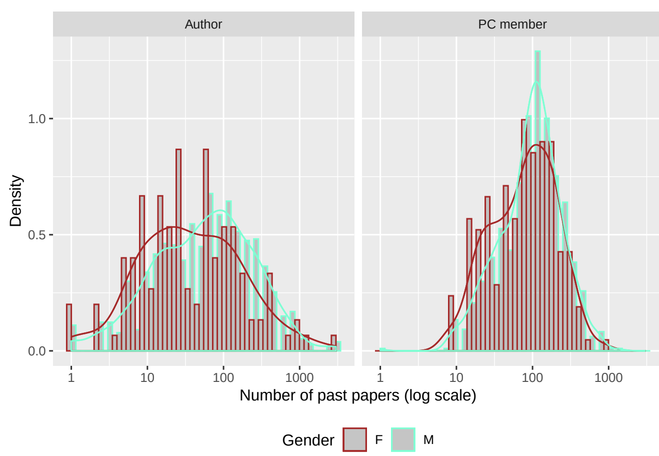

Fig. 4. Distribution of past publications by gender and role, as of 2017

As expected, PC members generally have more experience (publications) than authors, especially among women. This gap suggests that perhaps more of the female authors are novices, in relative terms. Another interesting observation is that the male authors' distribution "pulls to the right". In other words, there appear to be relatively more male authors in experienced or senior positions. This disparity may be related to the observations that women do not continue to senior research positions in the same rate as men [3], [4], [25], [26].

To confirm these observations, we stratified all our researchers into the following three groups, using the experience metric of H-index [27]: those with an H-index of less than 13, those with an H-index of 13 to 18, and the rest. We conveniently named these groups novices, mid-career, and experienced, although these breaks are arbitrary [27]. Indeed, Fig. 5 shows that fewer women than men reach the senior ranks of research, especially among authors (44.5% novice female authors compared to 36.5% men). This experience gap has also been observed in other engineering fields [4], [17].

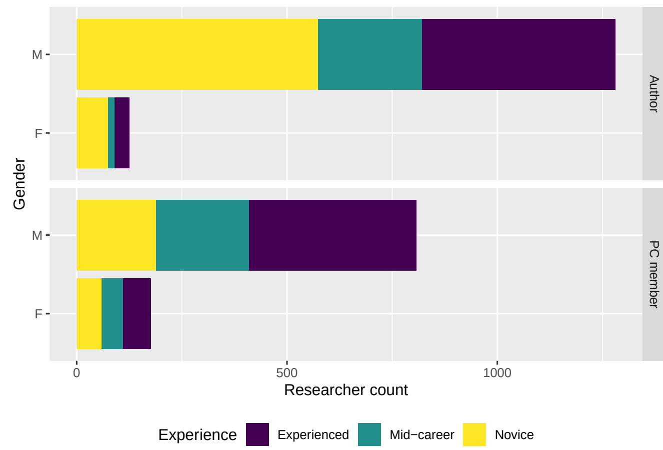

Fig. 5. Distribution of experience by gender and role

IV. Conclusion

This paper presented recent statistics on the representation of women in nine HPC conferences during 2017. The picture that emerges shows a field that is still far from achieving gender equality. Men in HPC comprise the vast majority of researchers, occupy most visible conference roles, and exhibit more research experience and higher citation counts than women.

Nevertheless, there are some indications that representation is improving. Both SC and ISC are actively measuring inclusivity metrics and have shown steady progress in female representation in 2018 and 2019. As the two conferences with a dedicated diversity chair and active push toward inclusion, they may serve as a benchmark to estimate the effect of diversity efforts. It is hard to generalize from these improvements given the short term since 2017, but they are encouraging. At the very least, these efforts are designed to make conferences more hospitable to all attendees, which may result in long-term improvement in gender representation. We plan to follow up at regular intervals to evaluate this hypothesis.

Other future work includes expanding this analysis to the full set of 56 conferences we have collected from all subfields of computer systems and to address more questions that emerge from the data, such as the differences in collaboration patterns between women and men.

Acknowledgments

We would like to thank Kelly Shaw for spearheading this research and providing detailed feedback along the way. We would also like to thank Josh Reiss for his part in the collection of gender data. This work was supported by a grant from the Social Justice Fund at Reed College.

References

[1] J. M. Cohoon and W. Aspray, "A critical review of the research on women's participation in postsecondary computing education. women and information technology," Cambridge Massachusetts, the MIT Press, 2006, pp. 137–180.

[2] J. M. Cohoon, S. Nigai, and J. Kaye, "Gender and computing conference papers," *Communications of the ACM*, vol. 54, no. 8, pp. 72–80, Aug. 2011.

[3] A. Frantzana, "Women's representation and experiences in the high performance computing community," PhD thesis, The University of Edinburgh, 2019.

[4] M. F. Fox, "Women, men, and engineering," in *Women, gender, and technology*, M. A. Fox, D. G. Johnson, and S. V. Rosser, Eds. University of Chicago Press Urbana, IL, 2006, pp. 47–59.

[5] T. Trader, "A celebration of women in HPC." https://www.hpcwire.com/2016/07/07/celebration-women-hpc/, july-2016.

[6] M. Franceschet, "The role of conference publications in CS," *Communications of the ACM*, vol. 53, no. 12, pp. 129–132, Dec. 2010.

[7] J. Freyne, L. Coyle, B. Smyth, and P. Cunningham, "Relative status of journal and conference publications in computer science," *Communications of the ACM*, vol. 53, no. 11, pp. 124–132, 2010.

[8] D. Patterson, L. Snyder, and J. Ullman, "Evaluating computer scientists and engineers for promotion and tenure," *Computing Research News*, September 1999.

[9] M. Y. Vardi, "Conferences vs. journals in computing research," *Communications of the ACM*, vol. 52, no. 5, pp. 5–5, 2009.

[10] Elsevier, "Gender in the global research landscape," 2017.

[11] C. W. Fox, C. S. Burns, A. D. Muncy, and J. A. Meyer, "Gender differences in patterns of authorship do not affect peer review outcomes at an ecology journal," *Functional Ecology*, vol. 30, no. 1, pp. 126–139, 2016.

[12] L. Holman, D. Stuart-Fox, and C. E. Hauser, "The gender gap in science: How long until women are equally represented?" *PLoS biology*, vol. 16, no. 4, p. e2004956, 2018.

[13] J. Huang, A. J. Gates, R. Sinatra, and A.-L. Barabasi, "Historical comparison of gender inequality in scientific careers across countries and disciplines," *arXiv preprint arXiv:1907.04103*, 2019.

[14] F. Karimi, C. Wagner, F. Lemmerich, M. Jadidi, and M. Strohmaier, "Inferring gender from names on the web: A comparative evaluation of gender detection methods," in *Proceedings of the 25th international conference companion on world wide web*, 2016, pp. 53–54.

[15] L. Santamaria and H. Mihaljevic, "Comparison and benchmark of name-to-gender inference services," *PeerJ Computer Science*, vol. 4, p. e156, 2018.

[16] L. L. Wang, G. Stanovsky, L. Weihs, and O. Etzioni, "Gender trends in computer science authorship." Jun-2019.

[17] S. Zweben and B. Bizot, "2017 CRA Taulbee survey," *Computing Research News*, vol. 30, no. 5, May 2018.

[18] M. Shermer, "How the survivor bias distorts reality," *Scientific American*, vol. 1, p. 106, 2014.

[19] C. J. Lee, C. R. Sugimoto, G. Zhang, and B. Cronin, "Bias in peer review," *Journal of the American Society for Information Science and Technology*, vol. 64, no. 1, pp. 2–17, 2013.

[20] J. R. Davenport, M. Fouesneau, E. Grand, A. Hagen, K. Poppenhaeger, and L. L. Watkins, "Studying gender in conference talks–data from the 223rd meeting of the american astronomical society," *arXiv preprint arXiv:1403.3091*, 2014.

[21] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, and L. Njilla, "Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability," in *Proceedings of the 17th iEEE/ACM international symposium on cluster, cloud and grid computing*, 2017, pp. 468–477.

[22] Clarivate Analytics, "Global state of peer review," Publons, 2018.

[23] G. Ghiasi, V. Larivire, and C. R. Sugimoto, "On the compliance of women engineers with a gendered scientific system," *PloS one*, vol. 10, no. 12, p. e0145931, 2015.

[24] Y. Tao, "Where do they do engineering? Gender differences in employment at the position level among engineering doctorate recipients," *Journal of Women and Minorities in Science and Engineering*, vol. 22, no. 1, 2016.

[25] S. Gerhard, F. M. Frank, and A. Kristen, "Undergraduate women in science and engineering: Effects of faculty, fields, and institutions over time," *Social Science Quarterly*, vol. 88, no. 5, pp. 1333–1356.

[26] M. C. Mattis, "Upstream and downstream in the engineering pipeline: What's blocking uS women from pursuing engineering careers," in *Women and minorities in science, technology, engineering and mathematics: Upping the numbers*, Edward Elgar Cheltenham, UK, 2007, pp. 334–362.

[27] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National academy of Sciences*, vol. 102, no. 46, pp. 16569–16572, 2005.