Selecting Technical Papers for an Interdisciplinary Conference: The PASC Review Process *

Torsten Hoefler
Department of Computer Science
ETH Zürich, Switzerland
htor@inf.ethz.ch

ABSTRACT

We discuss the paper selection process of the ACM PASC16 conference. The conference spans multiple scientific fields used to very different publication cultures. We aim to combine the strengths of the conference and journal publication schemes in order to design an attractive high-quality publication venue for works in large-scale computational science. We use four non-standard key ideas (1) no pre-selected committee, (2) short revision process, (3) full double-blindness, and (4) suggested expert reviews to design a paper selection process for ACM PASC16. In this overview, we document observations of the process and provide data in an attempt to characterize the effectiveness of the used mechanisms. We hope that the ideas can be used beyond ACM PASC16.

Keywords

PASC conference series; paper selection in HPC; review processes

1. INTRODUCTION

Here, we describe the paper selection process of ACM PASC16, the Platform for Advanced Scientific Computing, a conference aiming to provide a venue for the discussion of applications of large-scale scientific computing. For the first time, this year's PASC will have a papers track with proceedings in the ACM Digital Library. The conference is designed to serve the community and advance the scientific standards in the field, thus, the process is completely open and we seek input from the community. In this paper, we define and evaluate the process in detail and provide some analysis of the innovative elements. We hope that the lessons learned can be adopted beyond the PASC16 conference.

PASC16 is the third conference in the PASC conference series and it is focused around the combination of computer science, applied mathematics, and various domain sciences such as Climate & Weather, Computer Science & Mathematics, Engineering, Life Sciences, Materials, Physics, Solid Earth. PASC16 also has a special track for Emerging Domains to capture newest developments. With regards to the sponsorship, we focused on the only common demoninator, high-performance computing (HPC) in its typical role as combining core computer science systems and applied mathematics research. Thus, PASC16 is sponsored by ACM's SIGHPC as one of the premier special interest groups in the HPC area. SIGHPC is the professional computer science organization sponsoring prestigeous conferences a such as the Supercomputing (SC) conference series.

The mix of various sciences makes PASC16 different from other conferences in their respective fields and requires a different paper selection process. It is especially different from conferences in computer science which are usually sponsored by ACM. PASC aims to be the venue to publish work spanning boundaries, for example, top-class science codes and methods for next-generation scalable HPC. All of the PASC areas have a strong link to computer science but many have a a stronger link to the respective domain science in terms of publishing culture. The different publishing models vary significantly between PASC's science fields. For example, computer science publishes in peer-reviewed, highly-selective conferences that are highly reputable while many other sciences publish in peer-reviewed journals. In fact, publishing in competitive conferences in CS is often higher regarded than journal publications while this is the opposite in other sciences where conferences are typically not selective and are seen mainly as networking events and not as a serious publishing venue. Bad voices could say that "computer science has yet to grow up to adopt a more thorough review model" or "journals are too slow in a fast moving sciences like computer science".

Thus, in PASC, we aim to combine the benefits of both publishing models, mainly the timeliness of conferences and thorough revision process of journals, into a hybrid publishing mode.

1.1 The conference publishing process

Today's top conferences often apply a multi-stage review process which is constantly evolving. Usually, authors submit a paper which is reviewed by multiple referees. These referees are drawn from a pre-determined set of people (the conference or track committee). Reviews are then made available to the authors who get the option to clarify misunderstandings in a rebuttal. However, authors are typically not allowed to add results to the paper or address any shortcomings in that phase because the rebuttal is only for the

*This article is not peer-reviewed, the author is the solely responsible for the content

DOI: 10.1145/1235

clarification of potential misunderstandings. At the end, the committee considers the papers together with the rebuttals in the final selection stage.

Our personal opinion is that rebuttals are largely ineffective for the authors. In fact, we observed that they more often aid in rejecting a paper than they do in accepting a borderline paper. Furthermore, authors often invest significant resources into drafting rebuttals (often in a very limited space) — this work is wasted in most of the cases as it does not improve the paper itself and can hardly be reused in other contexts.

Top-class conferences usually apply an "in-doubt reject" philosophy which reduces the false positive rate for low-quality accepted papers. This leads to an inefficiency in the community that papers are reviewed multiple times and a paper's review history is often lost. Furthermore, authors often wait for months for the next suitable deadline.

1.2 The journal publishing process

Journals are very different from conferences in that authors have the explicit option to revise a manuscript if it is deemed promising. Here, subject area editors are receiving the submissions and pick a set of competent reviewers. Multiple rounds of reviews and revisions can be used to establish a communication channel between authors and reviewers. Reviewers are often persistent in that process and help to ensure quality.

Many journals have a very long latency from the first submission to publication, often six months, sometimes years. Furthermore, there is no attached venue to present and discuss the results and the community relies largely on independent meetings, conferences, and other author interactions. We want to remark that there are journal publishing models that are attached to meetings, for example VLDB journal/VLDB [3], ACM TACO/HiPEAC, or the ACM TOPLAS invited track at ACM PLDI.

2. PASC PAPER SELECTION PROCESS

PASC16 aims to combine both models to achieve the timeliness and topical discussion of conferences as well as the thoroughness and efficiency of the journal process. We started from a conference process and modified the four key ideas:

(1) no pre-selected committee The paper selection is organized by chairs who appoint only area editors. This allows each area editor to select the appropriate experts in the field for each paper without being limited by the pre-established committee. Of course, this somewhat complicates the management of conflicts and requires more attention following up with reviewers that have not previously agreed to serve on the committee to review in a relatively short time. Yet, our experience suggests that the review thoroughness and confidence can be greatly improved over a standing committee.

(2) short revision process Allowing the authors to completely revise papers gives them the opportunity to improve the quality for the conference submission as well as for later submissions. Limiting the revision time guarantees a limited-time turn-around for each paper and enables synchronization of reviews to improve efficiency in the organization. Of course, a limited time may be too short for the necessary revision. However, in that case, authors can always resubmit for the following year or to different venues.

(3) full double-blindness A fully-blind review process where neither reviewers nor co-chairs know the author names significantly reduces author and institutional biases. In PASC16, only area editors knew the authors to properly detect and resolve conflict of interests when inviting reviewers. One negative point could be that double-blind reviewing is potentially less attractive for reviewers. Yet, since we were able to find the experts in the topic, nearly all of them immediately agreed to review papers that were close to their area, even without knowing the authors.

(4) suggested expert reviews We allowed reviewers in round 1 to propose additional expert reviewers for round 2 in order to get an even finer-grained resolution in expertise. We observed that the senior experts in the area often suggested rising junior experts that were not that well known. These were eager to write thorough and insightful reviews. A disadvantage of this model is that the authors, after receiving only round 1 reviews for their revision, could not take the additional review comments into account. This could create a perceived unfairness in the author's view but we believe that can be addressed with carefully considering the review rounds in the final selection process.

We now briefly discuss other choices made in the PASC16 review process.

2.1 Timeline

PASC16 was arranged around the following timeline:

Jan 15: abstract submission Having an early abstract submission allows the area chairs to line up expert reviewers to start the review process immediately when the final papers arrive.

Jan 22: full papers due Right after the deadline was closed, reviewers were invited.

Feb 26: first notification After the first reviewing round, only promising papers were invited for a revision. Here, we employed the strategy: "when in doubt: invite."

Mar 11: revisions due We required the revisions to mark-up all changes to simplify second round reviews. The authors had the option to upload a reply to the reviews.

Mar 31: final decisions In the final decision, we chose to accept only top papers by employing the conference strategy "in doubt: reject". For papers that were barely not making the cut, we encouraged the authors to resubmit their work.

The timeline was very compressed for PASC16. This was mainly due to delays in the sponsorship agreement and we remark that the overall process should really span 5-6 months.

2.2 Selection criteria

The selection criteria for PASC16 were focused on scientific quality. In the discussion, each area chair had to answer the question "What did I learn while reading the paper?". As a secondary aspect, we considered the relevance to PASC ("How many people would attend the talk?") as well as how well the work is presented ("Would I recommend my colleagues to read it?"). During the physical technical program committee (TPC) meeting, only the scientific quality played a role because all revised papers were clearly on topic and well presented.

However, during the discussion, the committee felt that there should be a session which does not focus exclusively on novelty but on solid software documentation, methods, and maybe even history (similar to the "State of the Practice" session at Supercomputing).

The committee also consciously never disregarded an expert reviewer who objected strongly with the right arguments to build and maintain the reputation in the community. We understand that this may be slightly negative for the authors but we hope they understand the comments, improve their work and resubmit. Reject if in doubt is normal in the top-class conference model.

3. STATISTICS

In round 1, PASC16 received 44 submissions. The area chairs filtered incomplete and off-topic submissions in a basic review stage and moved 33 papers into the full review process. All papers received at least three reviews, six received a fourth review as tie breaker. After the first round of review comments, 23 of these papers papers were invited to submit a revision (all papers which were supported strongly by at least one reviewer). Most papers received three or more reviews in stage 2. Of these 23 invited, 12 papers were finally accepted after a second round of reviews. The final acceptance rate was 27% of the initial submissions. This acceptance rate is higher than top-class conferences such as Supercomputing because many papers improved significantly during the revision process.

The 105 reviews in stage 1 had an average length of 451 words. The 77 reviews in stage 2 had an average length of 153 words. Stage 2 reviews were generally shorter because the authors often succeeded to address all issues pointed out in stage 1.

3.1 Effectiveness of the revision

We will now briefly discuss how effective the revision was. We expect this to be close to a lower bound to the general effectiveness of revisions in a conference process because of the very limited time of two weeks. We believe a more appropriate revision time of 1 month would lead to even better results. We provided a free-form pdf upload for the revision itself as well as a second free form pdf upload for accompanying materials. All 23 papers invited to stage 2 submitted a revision as well as a reply to the reviews. Revisions ranged from addressing simple issues such as typesetting corrections up to a complete rewrite of the paper including numerous new experiments and an explanation exceeding the length of the actual submission.

We now compare the number of reviews for the papers invited to stage 2. Out of the 23 papers, 15 had three reviews in stage 1 and 8 had 4 reviews in stage 1. In stage 2, 4 had 2 revies, 11 had 3 reviews, 5 had 4 reviews, 2 had 5 reviews, and 1 had six reviews. The disparity in stage 2 resulted from the invitation of additional expert reviewers.

The system asked each reviewer to provide an absolute score in the range from 1 (strong reject) to 5 (strong accept). We observed that, on average, the scores from stage 1 to stage 2 increased but some reviewers reduced their scores dramatically. Yet, our results may not be statistically significant due to the small sample size, thue, we report the average difference for each paper. In total, 8 papers decreased their score by 1 point, 8 papers stayed unchanged, 6 papers increased their score by 1 point, and 1 paper increased its score by 2 points.

We now provide the complete data for the changes in all reviewers that reviewed the same paper in stage 1 and stage 2 and each of the categories (all ranging from 1 to 6):

type\change	-4	-3	-2	-1	0	+1	+2
overall	2	0	4	0	38	19	2
significance	1	2	1	7	45	6	4
soundness	1	0	1	10	47	10	4
originality	0	0	0	10	53	9	1
communication	0	1	2	3	41	13	3
expertise	0	0	0	5	47	10	1

We remark that the review system somewhat obscured the stage 1 reviews from the reviewers of stage 2 so that they could not easily find their own reviews and scores from stage 1. This was not intentional. We conclude, that the overall scores were significantly changed by the revision. Most improved slightly but some degraded significantly. The two -4 degradations were analyzed by the committee in detail but cannot be discussed here for confidentiality reasons.

3.2 Impact of expert reviewers

As mentioned before, we invited a number of expert reviewers proposed in stage 1 to the second stage. We explained to them that they joined late and review a revision and that their review time is very limited. We now present the final scores of all reviewers from stage 1 and stage 2 as two tuples of varying length, where a single digit denotes the score and "x" indicates that no second-round review was submitted. The tuple grows in the second round if additional reviewers submitted reviews. For example 135->12x4 means the three round 1 reviewers submitted scores 1, 3, and 5, respectively. The first two submitted scores 1 and 2, respectively, while the third reviewer did not submit a review but an additional suggested expert reviewer submitted a score of 4.

We present the full data in the following:

```
555->5x54; 435->3451; 445->4433; 5455->54554;
333->33x; 245->24x; 3455->445445; 335->135;
3343->434; 1554->155x; 454->554; 4434->55454;
5455->x455; 445->4455; 5444->5555; 423->3214;
254->354; 453->554; 232->x42; 455->5x52; 243->3x5;
4555->5x5x5; 553->xx555.
```

During the TPC discussion, we kept track of the impact of expert reviews. At the end, 2 expert reviews led to accepting a paper which may not have been accepted without them, 19 did not change the final decision, and 2 expert reviews led to rejecting two papers which may not have been rejected without them. We discussed the two reject cases at length but decided that the standing of the reviewer in

the respective field warrants the decision and that the paper will benefit from a longer revision taking into account the additional comments and resubmission. The committee communicated this to the authors.

In addition to influencing the decisions, the expert reviewers usually wrote rather long and insightful reviews that benefit the authors in their revision for the final version or a later resubmission.

3.3 The physical TPC meeting

The PASC16 TPC meeting was conducted as a single-day face-to-face meeting of all area chairs. The challenge was again the diversity in the areas and thus paper topics. The main purpose of this meeting was to ensure a consistent and qualitiatively high selection process. We proceeded in two rounds: in round 1, we discussed each paper for approximately six minutes (some much longer). Here, the responsible area chair had to present the set of reviewers and their relation to the topic/qualification. Then, the area chair summarized the papers contributions and the reviewer's opinions for the others (implicitly answering the questions posed in Section 2.2). During the discussion, we established an approximate ranking (partial order) of the papers, weakly following Douceur's proposal [1]. In the second round, we re-discussed the borderline papers and decided where precisely to draw the line of acceptance.

We did not explicitly consider acceptance rates and accepted all papers where there was no doubt of excellence. We discussed the borderline cases but applied the conference-typical "reject if in doubt" philosophy. For the borderline cases, we made sure to include a summary of the committee discussion in the feedback for the authors. The committee points out that all revised papers in round two were good and the quality of selected papers was outstanding. It seems that, with a longer revision process, all papers could be accepted. Yet, in the PASC16 selection model, they would need to be resubmitted by the authors.

We decided to assign shepherds to five of the accepted papers. Shepherds guarantee that revisions that the committee deemed essential are applied to the final papers. Shepherds are drawn from the committee and area chairs have the final say on acceptance of a manuscript.

3.4 Handling conflicts

Handling fully double-blind submissions is can be somewhat cumbersome because it needs an additional entity to arbitrate conflicts of interest. Thus, we decided to keep co-chairs as well as all reviewers double-blind while the area chairs had access to the author's names in order to select reviewers. We asked the area chairs to ignore the names in the selection process though.

For the TPC meeting, we appointed an additional person, who was not directly involved in the selection solely to manage conflicts. This was necessary as the co-chairs and area chairs did not know the author names of the different areas. Conflicts were handled by sending all author names to all chairs who then marked conflicts.

In one case, the committee decided to use the conflict manager as an oracle to ask a binary question like "is person X a co-author". We felt that this was necessary because a review pointed out that the paper included contents that should be credited to person X. We note that all content was correctly credited in the revised version.

4. DISCUSSION

How to organize conference committees and paper selection is a much discussed topic in top-class computer science conferences. Mogul and Anderson provide a nice summary of issues for organizing computer systems conferences [5]. Among the highly-discussed issues is double-blindness and the related problems. This is largely unresolved in the general community and different conferences apply different rules regarding double-blindness [7]. Other committee reports (e.g., [6, 8]) document the established rules and experiments in the different communities. There were even workshops on how to organize meetings [4]. Yet, all these conferences are focused solely on computer science. Here, we document our experiences with a more interdisciplinary community.

We now briefly summarize issues that the committee identified during the paper selection process:

Performance reporting The different communities have very different styles of reporting performance results. There was no overarching standard and result discussion ranged from log-log plots of application runtime, which did not provide insights into the performance of the code, to detailed statistically sound bounds (roofline) analyses. We suggest that the conference should provide guidelines to the submitters to ensure basic interpretability of the results, cf. [2].

Face-to-face meeting The PASC16 face-to-face meeting only included the area chairs. We believe that it would be better to also invite all reviewers to the meeting (if possible).

Timing The time-frame for most steps (revision, reviews) was too short and should be increased.

Areas By design of the conference, everything is related to HPC/computer science. So the computer science & mathematics area received many more submissions than other areas even though some submissions clearly fit into other areas.

Engineering/Software track The conference steering should consider a track that accepts papers that are not necessarily original research contributions but of huge value to the community. This year, we had to reject some excellent submissions fitting this category due to limited novelty. This track could be similar to SC's "State of the Practice".

Conflict handling The conflict handling can be further improved by allowing authors to submit a list of conflicts. This will enable a fully double-blind process where even the area chairs are not aware of the author's names or institutions.

Chair load The main tradeoff of the new approach is that much of the burden is now shifted to the area chairs. Each chair needs to understand every paper at a high level to make an assignment. This requires very reliable chairs and may also pose scalability problems with if more submissions are received.

5. CONCLUSIONS

The authors feel that the overall selection process was successful in that all four key ideas, (1) no pre-selected committee, (2) short revision process, (3) full double-blindness, and (4) suggested expert reviews, benefitted the rigor of the selection and hopefully the author's experience. Especially (1) proved to be beneficial when we discussed the reviewers expertise during the TPC meeting. We had very strong re-

viewers, for example, if the paper used new method Y, we were often able to recruit the inventor of Y, or the authors of similar software packages in the field. This would hardly be possible with a pre-selected committee. Revisions were perceived as very successful and influential, most of the time raised scores, author's efforts were remarkable.

After the final selection was finished, the committee checked how many papers focused on scalable systems/software — in fact all but one were emphasizing very large-scale systems, and for the one large-scale is not relevant!

Overall, the process was tough; good papers were rejected, we chose to only accept if there was no doubt about contribution and correctness (we have some controversial accepts but that's the nature of science :-)). We believe that PASC16's key principles can be applied in the context of different conferences and we will work to improve the paper selection process in the future.

Acknowledgments

I want to specially thank my papers co-chair David Keyes from KAUST. This process would not have been possible without his help and tireless effort to find great reviewers for the many computer science and mathematics papers. Similar thanks belongs to the area chairs (alphabetically) George Biros (University of Texas), Omar Ghattas (University of Texas), George Lake (University of Zurich), Mark van Schilfgaarde (King's College London), Jeroen Tromp (Princeton), Michael Wehner (LBNL), and Ioannis Xenarios (Swiss Institute of Bioinformatics) with a special thanks to George Lake who was very helpful when balancing across tracks during the TPC meeting. Tim Robinson and Tatjana Ruefli from CSCS/ETH Zurich did a wonderful job helping with managing the submission system, keeping reviewers, area chairs, and co-chairs on time, organizing the physical TPC meeting, and managing the overall publication process. Their unique contributions to the conference cannot be emphasized enough. Furthermore, we thank Olaf Schenk for helping with the selection of area chairs and Thomas Schulthess for various input related to the conference. We also thank all involved reviewers, who will be listed separately. Furthermore, we want to thank all authors who submitted for their efforts and support.

6. REFERENCES

[1] John R. Douceur. Paper rating vs. paper ranking. *SIGOPS Oper. Syst. Rev.*, 43(2):117–121, April 2009.

[2] Torsten Hoefler and Roberto Belli. Scientific benchmarking of parallel computing systems: Twelve ways to tell the masses when reporting performance results. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '15, pages 73:1–73:12, New York, NY, USA, 2015. ACM.

[3] H. V. Jagadish. The conference reviewing crisis and a proposed solution. *SIGMOD Rec.*, 37(3):40–45, September 2008.

[4] Jeffrey C. Mogul. Workshop on Organizing Workshops, Conferences, and Symposia for Computer Systems, 2008. http://www.usenix.org/event/wowcs08/, San Francisco, CA.

[5] Jeffrey C. Mogul and Tom Anderson. Open issues in organizing computer systems conferences. *SIGCOMM Comput. Commun. Rev.*, 38(3):93–102, July 2008.

[6] Konstantina Papagiannaki and Luigi Rizzo. The ACM SIGCOMM 2009 Technical Program Committee Process. *SIGCOMM Comput. Commun. Rev.*, 39(3):43–48, June 2009.

[7] Henning Schulzrinne. Double-blind reviewing: More placebo than miracle cure? *SIGCOMM Comput. Commun. Rev.*, 39(2):56–59, March 2009.

[8] Peter Sewell. Popl 2014 program chair's report. *SIGPLAN Not.*, 49(4):10–26, July 2014.