Message from the MICRO 2021 Program Co-Chairs

It is our great pleasure and honor to welcome you to the program of the 54th IEEE/ACM International Symposium on Microarchitecture (MICRO). The program showcases the continued growth and energy in our field with a record number of papers accepted. The program reflects the expanding interest of the community with roughly half the program covering important topics such as security and privacy, quantum computing and domain specific acceleration (e.g., machine learning, graph processing). The program also demonstrates the importance of synergy between architecture and software and also includes conventional paper topics that continue to enhance the state-of-the-art in CPUs, GPUs, and memory/storage hierarchy by improving any combination of performance, energy, and/or reliability. The program also features three keynotes by Michael Clark (AMD), Anastasia Ailamaki (EPFL), and Sean Lie (Cerebras) and a session with the finalists of the ACM Student Research Competition.

Timeline: ~13 weeks from Paper Submission to Author Notification

The timeline for MICRO 2021 was as follows. Abstracts were due on April 9 and full submissions were due on April 16. Authors were unable to submit a paper without having registered an abstract first. Because the review process is heavily dependent on proper review assignment, we spent ~2 weeks manually assigning papers to reviewers. The papers were released to the reviewers on April 29 with a requested review deadline of June 1 (~4 weeks). Roughly 80% of the reviews were finalized by the requested deadline and 98% finalized over the next five days. Before releasing the reviews to the authors, we performed a review quality check by assigning a "paper lead" to read through all the reviews and ensure they met our review quality standards (more details below). We released the reviews to the authors on June 8 and requested them to submit a revised manuscript (or simply provide a rebuttal) by June 21 (~2 weeks). We scheduled a two-week online discussion period between June 22 and July 6 before the virtual PC meeting to allow the reviewers to read the rebuttal/revision, discuss the paper, and reach online consensus to either online-accept, online-reject or discuss the paper at the PC meeting. During the online discussion period, for the first time, we introduced author-reviewer communication where reviewers directly interacted with authors to clarify concerns and questions raised during the discussion period, while also ensuring the double-blind review process (more details below). Finally, we held a 10-session Zoom-based virtual PC meeting over 2 days on July 8 and July 9 across nine different time zones. After the PC meeting, the submission outcome notifications were sent to the authors on July 14. For papers that underwent shepherding, the process concluded on September 1 and all shepherded papers were eventually included in the final program.

Overall Statistics: We received 430 full-paper submissions (on par with 446 at MICRO 2020). Of the 430 submissions, four were rejected before the review period started due to format violations and five were withdrawn during the revision/rebuttal period. From the remaining 421 submissions, the Program Committee selected 94 papers for inclusion in the program (23 of these papers were shepherded) – an acceptance rate of 22%. The Program Committee (PC) consisted of 122 members from around the globe who are experts on the broad range of topics within the field of computer architecture. Most PC members reviewed 14 or 15 papers. The PC was aided by a 181-member External Review Committee (ERC) where most ERC members reviewed 4 or 5 papers. We solicited 6 reviews per paper: 4 PC and 2 ERC. All papers had 6 reviews with the following exception: 16 papers had 7 reviews (to seek additional expert reviews) and 10 papers had 5 reviews (due to a missing review and seeking the sixth review was unnecessary because the existing 5 reviews had low scores). The PC and ERC combined provided a total of 2562 reviews, with an average of 6.1 reviews per paper. Antonia Zhai handled the review process for the six submissions that we were both conflicted on.

Submission and Review Process

During the first couple of weeks in our newly minted role as Program Co-Chairs for MICRO, after significant brainstorming, we identified the following opportunities to enhance the existing review process (a) improve review assignment, (b) ensure high review quality, (c) reduce randomness at the PC meeting, (d) gather author feedback on reviews received. We now provide a detailed description of our submission and review process.

Single Round Review: We investigated both the single and two-round review process based on the expected number of papers and the expected PC/ERC size. Our analysis revealed that a two-round review process would likely reduce the reviewer burden by 2-3 papers for the PC and 1-2 papers for the ERC. For the lower review burden, however, a two-round review process would potentially create unfairness issues where good papers are rejected in the first round. Furthermore, a two-round review process can also increase the overhead for (a) reviewers to manage their personal/professional schedules with multiple review deadlines and (b) Program Chairs to perform review assignments twice (which is already a lengthy process). Thus, based on this analysis, we adopted a single-round review process, which ensured maximizing fairness across all submissions while also providing better time management. In doing so, our review process provided a long review duration (4 weeks), sufficient time for the author response period (2 weeks), and enough turnaround time (2 weeks) between MICRO notifications and the next conference submission deadline (i.e., HPCA). The review load in the single-round process was kept low thanks to our large committee size (14-15 papers for the PC and 4-5 papers for the ERC).

Hierarchical Paper Topics: At submission time, the HotCRP submission form invites authors to select from a list of topics that apply to their submission. The reviewers are also presented a similar form to specify their expertise (or interest) on the list of topics. Program Chairs (and HotCRP's automated tool) perform paper assignments by identifying reviewers that have expertise (or interest) in the paper's topics. Such a paper assignment strategy can become challenging when the topics are overlapping, and the author selects many topics. Ideally, we desire the topics to be orthogonal to one another, so that authors can select the fewest number of topics that best represent the main contributions and insights of their submissions. In collaboration with Stefanos Kaxiras (Program Chair for HPCA 2022), we developed a hierarchical topics list. For example, our hierarchical topics list helped us distinguish between reviewers that had expertise on different computational platforms (e.g., accelerators, GPUs, CPUs, etc.), different application domains (e.g., cloud, graph processing, machine learning, etc.), and different parts of the memory hierarchy (e.g., cache, memory, storage, etc.), without requiring an exhaustive list consisting of the cross product of the different possible combinations. Authors were requested to limit their choices to at most four topics. Aside from a handful of papers, most authors adhered to this restriction, and we were able to easily group papers by topics when making review assignments.

Collecting Reviewer Expertise Information: Our Program Committee and External Review Committee consisted of 306 members. While we were aware of the research expertise of each reviewer, we wanted to gather data about their expertise and recent work using our hierarchical topics list. For example, using this approach would enable us to distinguish between accelerator experts in different domains such as graph processing, machine learning, or cloud computing. We invited each reviewer to specify their expertise using a Google Form of the hierarchical topics. The reviewers were instructed to only select topics where they considered themself an expert (or a sufficient expert). A reviewer could consider themselves an expert on a topic if the topic was their active area of research or they closely followed the topic and were familiar with the state-of-the-art on that topic. We used the Google form in lieu of the reviewers HotCRP "Topic Interests" profile information because we planned to make review assignments solely based on expertise (rather than

interest). The Google form also provided an opportunity for the reviewers to specify topics they preferred not to review. We had a near 100% response rate to our expertise survey suggesting that our approach for was well received.

Revision Letters for Review Continuity: We continued MICRO 2020 practice of giving authors the opportunity to submit a revision letter explaining how a submission that had been previously rejected from a previous venue was revised and improved. The intent of this revision letter was to provide some continuity in the review process and to also address cases in which comments regarding a lack of improvement are made during the review process. The deadline to submit this revision letter was one week after the paper submission deadline, which gave authors some time to construct a complete and concise revision letter. The authors had a choice about when and how to make the revision letter available to reviewers: (1) immediately available to all reviewers, (2) available upon explicit request (intended for past reviewers only), (3) available as part of the rebuttal, or (4) withheld from all reviewers. This process allowed authors to use their own judgment about whether and how to communicate to reviewers their effort in improving the paper, while keeping the reviewers unbiased with respect to the paper's submission history. The reviewers were instructed to use the revision letter strictly to understand how a submission improved over time.

Roughly 50% of the total submissions this year were resubmissions from past conference venues. Roughly 80% of these submissions included a revision letter with varying reviewer visibility settings. To measure reviewer continuity, we explicitly collected information from reviewers (as part of the review form) on whether they reviewed an earlier version of the paper at a past conference. Of the 426 papers reviewed, three papers had three past reviewers, 21 papers had two past reviewers, and 57 papers had one past reviewer. Since a large fraction of our reviewers have served as reviewers for past conferences, we expected these numbers to be higher. We also collected information from reviewers on the usefulness of the revision letter. The feedback from reviewers was that the revision letters were unnecessary for most papers and useful for a very small number of papers.

Review Assignment: We leveraged ISCA 2021 scripts that crawl public databases to populate collaborator information back into HotCRP. As part of the review assignment process, to ensure fairness, we also ensured that all paper submissions adhered to the format as specified in the submission guidelines. We relied on HotCRP's built-in format checking tool to detect format violations. We found four papers had gross format violations. These papers were rejected immediately. We wanted to ensure that paper format issues were addressed by us at review assignment time and did not come up during paper review/discussion time. This ensured that the time spent in making review assignments and performing reviews was efficiently utilized since format related issues did not impact paper outcome late in the review process.

We manually assigned papers to reviewers in HotCRP because it helped us immediately detect any possible paper conflicts using HotCRP's conflict hints. We assigned six reviews per paper rather than the common practice of assigning five reviews per paper. We did this to future proof discussions at PC meetings. For papers that have mixed opinions, using the majority-rule with six reviews per paper ensures a difference of at least two votes which provides sufficient signal when deciding paper outcome (five reviews provides a vote difference of only one). For example, five reviews per paper can result in a 3-2 vote (for or against acceptance) while six reviews per paper can result in a 4-2 vote (for or against acceptance). With 5 reviews per paper, a 3-2 vote always goes to a PC-wide vote, but with six reviews per paper, a PC-wide vote is only required for a 3-3 tie vote. Consequently, our hope was that six reviews per paper reduces the number of times we go to a PC-

wide vote, which can become a random process with a large PC size. *Six reviews per paper would ensure that the eventual paper outcome mostly rests in the hands of the paper reviewers*.

For our review assignments, we prioritized expertise by matching the hierarchical topics list of papers with that of reviewers. We also read abstracts and skimmed the related work. We also leveraged insights and tools written by past computer architecture conference chairs and developed a tool called MightyPC (with a GUI interface written in R). MightyPC interfaces with a local MongoDB database that consists of *all* publications by the MICRO 2021 PC and ERC members. MightyPC was also populated with *all* the MICRO 2021 paper submissions with the paper PDF and author, paper topic, and paper conflict information. MightyPC was also populated with assignments made by a machine-learning tool (called CSRA) that is actively being developed by Omer Anjum. Based on all these inputs, MightyPC recommends non-conflicted reviewers by (i) identifying reviewers whose papers were cited by the submission (ii) reviewers who identified themselves as experts on the paper topics (based on our review expertise survey) (iii) reviewer recommendations provided by CSRA.

MightyPC helped prune the reviewer search space from 306 reviewers to a dozen reviewers. From this pruned list, we manually selected reviewers by making the following choices. To reduce reviewer bias, we ensured that reviewers from the same institution (industry or academic) and/or those who co-authored a cited paper did not co-review a submission. When making assignments, we did our best effort to load balance among reviewers and to increase reviewer diversity for each paper. Where possible, we maximized expected expertise while minimizing the number of papers with more than two reviewers that were more junior, from industry, or were from the same geographic area (USA excluded). Review assignments were done directly in HotCRP which allowed us to load balance reviews simultaneously. The large number of reviewers provided us with sufficient experts to choose from and HotCRP allowed us to keep an eye out on the current load for each. While this did not eliminate load balancing, it helped bound the per-reviewer load and eventually simplified load balancing.

MightyPC enabled us to make review assignments in roughly 12 days. We were very pleased with our assignments, and we also received positive feedback from the reviewers. To facilitate paper review assignment by future conference program chairs, our team released the MightyPC toolkit on GitHub (https://github.com/TheNetAdmin/MightyPC). The toolkit features parsing DBLP and Microsoft Academic Graph database, maintaining a MongoDB for agile development and centralized data storage, parsing of reference lists from PDF files, reviewer suggestions based on PC/ERC papers cited by a submission, and finally generating PC meeting Zoom configurations and slides to show conflicts for each discussion.

Obtaining Quality Reviews: All reviewers were asked to acknowledge and abide by a set of ethics guidelines derived from ASPLOS 2020, ISCA 2020, MICRO 2020, and ISCA 2021. We required that the reviewers perform the reviews themselves, with the exception where academic reviewers requested involvement of their senior graduate students (on a select set of papers) to help train them in the research process. We requested these reviewers to reach out to us with the student's name to ensure there weren't any conflicts.

Overall, the review process went smoothly with roughly 80% of the reviews turned in by the requested review deadline and 98% were finalized five days after the review deadline. In these five days, we conducted a review quality check on the finalized reviews. This was adopted from ASPLOS 2020 where a "paper lead" was assigned to each paper. The job of the "paper lead" was to read ALL the reviews and check that reviews (a) were constructive, respectful, and of reasonable length (b) provided precise citations to related work and avoided comments such as "this is well known or has been done before" (c) did not dismiss the paper based on

concurrent work or non-peer reviewed publications (e.g., arXiv) that the authors could not have known of (d) did not dismiss a paper solely based on fit but instead based on technical merits. Based on feedback from the MICRO Steering Committee, we advised reviewers to err on the side of inclusion for papers that do not perfectly fit MICRO. We advised the "paper lead" to directly escalate fit concerns to us.

The "paper lead" filled out a Google form informing us of situations where the review quality standards were not met. If the paper lead felt comfortable, they were requested to also leave a comment in HotCRP requesting the reviewer to improve review quality. We observed that a small number of reviews were flagged during the review quality check and those flagged were updated in a timely fashion before releasing the reviews to the authors. Overall, we found that the review quality check helped us identify and resolve problematic reviews early which ensured that the authors could spend maximum time addressing the concerns raised in the reviews.

Authors Were Given the Opportunity to Respond: Authors were able to see the reviews (and the review scores) on June 8 and were invited to submit a revised manuscript and/or a rebuttal by June 21. To enable the authors to judge the importance of the revision/rebuttal, we also released the scores for the importance of an author revision/rebuttal. The authors were given three options: (i) to not submit a rebuttal nor a revision (ii) to submit an 800-word rebuttal only (no revision). The rebuttal could either be a 1-page PDF of the rebuttal (single column) or plain text directly in the HotCRP form (but not both) (iii) submit a revised paper and a rebuttal. The paper revision had to comply with the same guidelines as the original paper with the changes highlighted. We recommended paper modifications to be less than 10% of the original submission. The goal of the revision option was to give the authors the opportunity to address small issues that could be fixed within the response period. In general, revisions provided a positive impact with post-rebuttal scores generally improving.

A small number of papers received late reviews, including those from additional reviews that we requested to increase overall reviewer expertise on the submission. In situations where the reviews were finalized after the rebuttal deadline, the authors were requested to provide a rebuttal as soon as possible (we allotted ~150 words per late review). Most of the rebuttals were provided within a couple days after forwarding the late reviews.

Author-reviewer Communication: For the first time, we enabled anonymous author-reviewer communication during the review process. Using HotCRP functionality, this process enabled reviewers to communicate with authors in real time. A natural question was when to enable the author-review communication period? There were three possibilities (i) review period (i.e., before author rebuttal/revision) (ii) during the rebuttal/revision period or (iii) during the online discussion period. The existing rebuttal/revision process already allows the authors to respond to individual questions/concerns raised by the reviewer. However, there exists no process that allows authors or reviewers to address questions/concerns raised by reviewers during the online discussion. Thus, we decided to enable author-reviewer communication during the second week of the online discussion period. This enabled reviewers to spend the first week discussing the paper and come up with suitable questions/comments to address concerns raised during the online discussion. We advised reviewers to ask the authors questions that were pivotal for making an online decision. Any reviewer could post a question that was related to (1) concerns that were not addressed in the author's response and/or revision and (2) content of the paper, rebuttal, and/or revision that needed further clarification. Of course, we requested that reviewer questions and author answers maintain the double-blind review process. In total 15 papers participated in author-review communication of which three papers reached consensus for online acceptance and three papers reached consensus for acceptance during review discussions at the PC meeting. Reviewers that participated in the author-reviewer communication had an overall positive

experience and felt that this process allowed them to make an informed decision on the eventual outcome of the paper.

Post-rebuttal Discussion with Optional Author-reviewer Communication: Papers were heavily discussed by the PC and ERC prior to the PC meeting (between June 22 and July 6) using the HotCRP review site. Each paper was assigned a PC member as a discussion lead to initiate and monitor the discussion – discussion leads were selected based on expertise and positivity, while balancing load across all PC members. All reviewers were requested to read through the rebuttal/revision, discuss the paper, acknowledge reading the rebuttal by providing post rebuttal comments and finally update their post-rebuttal overall merit scores based on the discussion. During this period, reviewers also had the option to communicate with authors anonymously to clarify questions/concerns raised during the online discussion (as discussed above).

Like MICRO 2020 and ISCA 2021, we opted for an anonymous discussion process where reviewer names were hidden before the PC meeting. The purpose of keeping much of the discussion anonymous was to encourage reviewers (regardless of their expertise level or seniority) to make complete discussion points. Furthermore, this approach reduced reviewer personalities from influencing the discussion. To hold the reviewers accountable for their discussion and to facilitate pre-meeting decisions, we revealed reviewer identities five days before the PC meeting. More than 5200 comments were posted during the discussion process. On average, this is ~12 comments per paper with reviewers exchanging more than 20 messages on 15% of the submissions.

To kickstart the discussion process, we grouped papers based on the 6-point overall merit score. We defined scores of 4 (or higher) as an accept vote and 3 (or lower) as a reject vote. We refer to reviewers who gave an accept vote as a supporter and those that gave a reject vote as an opponent. Papers were grouped into three tiers based on the number of supporters: #tier1, #tier2, and #tier3. Ideally, we should determine reviewer support for a paper based on the post-rebuttal overall merit score, but this was not available at the beginning of the discussion period. To address this, the paper tiers were continually updated as the reviewers updated their post-rebuttal scores. Using HotCRP's formatting feature, we color coded the three tiers similar to a traffic light: #tier1 was colored green (top 15% of submissions that were likely to be accepted) and comprised those papers that had four or more supporters; #tier3 was colored red (bottom 45% of submissions likely to be rejected) and comprised those that had five opponents; #tier2 were colored yellow (middle 40% of submissions requiring most discussion) and comprised those that had mixed overall merit scores. The coloring was intended to guide reviewers where discussion will be most effective. There is no value in spending significant time discussing 'green' papers that are likely to be accepted (unless there were critical issues being raised by an opponent). If a reviewer wishes to champion a paper in the 'red' category, they better know about it sooner than much later in the discussion process. Finally, we wanted the reviewers to easily identify the yellow papers that required the most discussion due to the mixed reviewer support.

Reviewers were instructed to reach online consensus by the end of the discussion period and mark the paper as #online-accept, #online-reject, or #discuss. Recall that each paper submission had 6 or more reviews. A #tier1(green) paper could be marked as #online-accept if it had at least four supporters (i.e., accept votes) and there was consensus among *all* the reviewers to accept (even those that did not give it an accept vote). If the reviewers felt that an online-accept paper could benefit from shepherding, they were also requested to mark the paper as #shepherd. If there was no consensus, the green paper would be marked as #discuss. A #tier3 (red) paper could be marked as #online-reject if it had at least four opponents (i.e., reject votes) and there was consensus among *all* the reviewers to reject (even those that did not give it a reject vote). If there was a

champion, the paper was marked as #discuss. For #tier2 (yellow) papers with three supporters (or two supporters with one being a champion), reviewers were encouraged to suggest additional reviewers (especially from the PC) to help make decisions. Since there was no consensus among reviewers, these papers were also marked as #discuss.

At the end of the discussion period, 281 papers from #tier3 were marked as #online-reject. The discussion leads for these papers were requested to provide a summary of the online discussion mentioning the reasons for rejection and areas for improvement. Of the remaining papers, 54 #tier1 papers were marked as #online-accept and 86 papers were marked as #discuss (28 from #tier1, 54 from #tier2, and 4 from #tier3). These 140 papers advanced to the virtual PC meeting and the discussion leads were requested to enter a summary in HotCRP on the paper contributions, strengths, weaknesses, and the key concerns raised during the online discussion.

Virtual PC Meeting: The PC meeting was held online due to the COVID-19 pandemic. We used Zoom as our conferencing software. Because we had an international PC committee, we organized a two-day PC meeting consisting of five sessions per day to accommodate 9 different time zones from California to Japan. Committee members were encouraged to attend all sessions that were reasonable in their own time zone. The paper discussion order was determined so that the time was reasonable for most PC reviewers. PC reviewers could see when a paper will be discussed by looking at the paper tags. The attendance of the sessions was excellent overall at 75-90%. To handle conflicts of interest, all PC members with a conflict were automatically moved to a breakout room prior to each discussed paper and all other PC members automatically moved into a separate discussion breakout room (to avoid the possibility of PC members joining the meeting in the middle of a discussion they were conflicted with).

To maximize time on papers that need discussion, it has now become the norm where online-accepted papers are no longer discussed at the PC meeting. However, we felt that bringing online accepted papers to the PC meeting would provide two benefits. First it enables the PC to calibrate against online accepted papers. Second, it ensures that all papers eventually accepted at MICRO are vetted by the PC. Thus, we needed to budget time for all 58 online accepted papers to be brought to the PC meeting.

In preparation for the PC meeting, the 140 papers were placed into discussion groups. One of the challenges is how to sort the papers and then where to draw the boundaries for the different discussion groups. Conventional strategies use some form of average overall merit (perhaps also considering reviewer generosity). This approach can significantly change the relative order of a paper based on a single bad score. Instead of relying on average merit scores, we sorted papers based on the number of supporters (i.e., by color/tier assigned during the online discussion period). We separated the #tier1 (green) papers which had four or more supporters into two groups: #groupO (online accept) and #groupA (#discuss). The #tier2 (yellow) papers were separated into two groups: #groupB (3 supporters) and #groupC (2 supporters). We also included in #groupC the #tier3 (red) papers that had a single supporter/champion. In total, there were 54, 28, 43, and 15 #groupO, #groupA, #groupB, and #groupC papers respectively. We planned to discuss all papers in #groupO, #groupA, and #groupB first and discuss #groupC papers on the second day only if there was time.

We discussed 140 papers in 10 sessions each consisting of ~6 online accepted papers and 6-8 discussion papers (total 12-14 papers per session). From a time-management perspective, the first 15 minutes of each session were allotted for online accepted papers (which also served as calibration for that session). The remainder of the session time was allotted to discussion papers with each paper receiving 10 minutes of discussion time and two minutes of voting time. Paper discussion order was randomized in each session with the only requirement

that #groupC papers appear on the second day (afternoon) and discussed after all #groupA and #groupB papers for that session were done.

We used the following process in each session. For online accepted papers, the discussion lead provided a two-minute paper presentation providing the key contributions of the paper. No questions were allowed, and the paper was accepted into the program unless an objection was raised by a member of the PC. Only one online-accepted paper was objected to and after a short discussion the paper was accepted with shepherding. For the discussion papers, the discussion lead introduced the paper and summarized the main points made by external reviewers during the discussion period. Then, other PC reviewers chimed in and provided their perspectives on the paper. The discussion was then opened to the entire PC for further comments and questions. We then took a vote across all the reviewers of the paper. ERC vote and absentee PC vote was based on the post rebuttal overall merit score (pre-rebuttal if there was no post-rebuttal score). All papers discussed at the PC had at least six reviews and the paper outcome were decided by majority rule (i.e., vote difference of 2 or more). By design this strategy ensured that paper outcome mostly rested in the hands of the paper reviewers. This became important especially towards the end of the PC meeting where reviewers wished to champion papers to be brought for discussion at the PC meeting. A championed paper required a minimum of three supporters to move towards a PC wide vote. If this threshold could not be met due to absent PC members, the paper was not eligible for discussion. Again, we felt this was important as it prevented reviewer personalities from influencing paper outcome through a PC-wide vote and allowed paper outcome to mostly rest in the hands of those who read it.

When the reviewer vote was a tie, we went to a PC-wide vote. Before performing a PC-wide vote, the negative then positive reviewers were given an opportunity to summarize the reasons for rejecting/accepting the paper. This was to ensure that paper discussions ended on a positive note. The whole PC then voted based on the summary using the secret ballot mechanism in HotCRP. We modified the HotCRP secret ballot to also hide the vote tally counts until after the voting had finished. This was to prevent a PC member from being influenced by the direction the whole-PC vote was trending. We required a minimum 60% of the PC present to vote with the majority vote deciding the outcome (tie votes were considered an accept, though we had no tie vote).

Overall, 131 papers were discussed/presented at the PC meeting. The number of papers accepted in each category is as follows: 54/54 #groupO (online accepted), 24/28 #groupA (papers with four or more supporters), 14/43 #groupB (papers with exactly three supporters), and 2/15 #groupC (papers with less than or equal to two supporters). Only six of the #groupC papers were discussed, and the two #groupC papers accepted started with only one/two supporters prior to the PC meeting and became #groupA/#groupB papers (through continued online discussion) by the time the paper was discussed at the PC meeting, showcasing the importance of online discussion. Of the papers discussed, 26 papers went to a whole-PC vote of which only five were accepted. At the end of the virtual PC meeting, a record 94 papers were accepted of which 23 were shepherded. Discussion leads for all papers discussed at the PC meeting were requested to key-in a short summary of the paper discussion from the PC meeting.

Mitigating Review Process Side Channels: We took several precautionary measures to ensure that the review process had the least amount of information leakage. Because we performed all review assignments directly in HotCRP rather than offline, we wanted to ensure that submission information was not leaked until review assignments were finalized. Unfortunately, HotCRP did not have an approach to enable/disable paper visibility before the review period begins. To address this issue, we disabled all reviewer accounts for the two weeks when we were making assignments. During the PC meeting, we changed HotCRP visibility settings to allow

PC members access to only those papers that were planned to be discussed at the PC meeting (there was no reason to make online-rejected papers visible). After the PC meeting, we reset the HotCRP settings, so that PC/ERC members could only view papers they reviewed. We also ensured that PC/ERC members had no visibility of papers accepted in HotCRP (aside from the papers they reviewed, and the paper title information based on the accept list we sent out after the PC meeting). Additionally, author names were never revealed throughout the review process and were visible only when the final program was published online.

Shepherding: Shepherding was coordinated using author-review communication directly within HotCRP and remained double-blind. The shepherds were assigned during the PC meeting and shepherds were requested to provide a detailed PC meeting discussion summary and the required changes to be included in the final paper. Most shepherded papers went through a couple of iterations before receiving the green light. Going through the HotCRP comments for shepherded papers, we observed that the authors appreciated working with their shepherds and agreed that eventually they had a much stronger final camera-ready paper.

Best Paper Selection: We selected the best paper candidates by identifying 18 top papers based on the average post-rebuttal overall merit score. We requested the reviewers (both ERC and PC) of these papers to fill out a short Google form survey on whether they wished to nominate the paper for best paper. Ideally, the winner of the "Best Paper" award would demonstrate a novel idea/approach that expands the cumulative knowledge in the field and is likely to inspire future research, as opposed to providing improvements on a well-known approach/result (even if the study was conducted and reported extremely well). Thus, when voting, we asked the reviewers to consider (a) technical content, (b) novelty, (c) research significance and (d) potential for future impact. We received a 95% response rate from the reviewers. Based on the PC nominations, two papers clearly emerged as best paper candidates. However, we had a multi-way tie across four papers, from which we selected two papers by including ERC nominations and average overall merit score to resolve the tie.

Authors Were Given the Opportunity to Provide Feedback on Reviews Received: To gather some deeper insights into the review process, this year we collected author feedback for the reviews received on their submission(s). Reviewers were unaware that we were going to collect author feedback on reviews. This was to ensure that reviewers did not change their existing review practice based on knowledge of an upcoming survey. Reviewers were informed of our survey after the reviews were finalized and we entered the online discussion period. We conducted a five-question survey for each review an author received on their submission(s).

1) Will the Reviewer feedback improve the overall quality of your paper?

2) Based on the review content, do you feel that the Reviewer had an adequate understanding of your paper?

3) How was the tone of the review?

4) Were the concerns raised by the Reviewer reasonable?

5) Considering everything, how would you rate the overall quality of the review?

We requested the authors to reflect on each individual review and answer all five questions for each review. When filling out the survey, we requested the authors to focus on the review content of each individual reviewer (not the review scores). We requested only one feedback per paper submission, and asked authors to confer with their co-authors before responding to the survey. This was to avoid papers with large author lists from biasing the survey results. We informed the authors that the feedback had no influence on the paper outcome at MICRO-54 (or any future venue) and that the raw data collected will not be seen by anyone except

the MICRO-54 Program chairs. We expressed our plans to share aggregated feedback with the reviewers (after paper discussions were finalized) so that they can improve future reviews. We also expressed our plans to share summary information (anonymized) with future Program Chairs to improve review management.

The authors received the survey at the beginning of the reviewer discussion period with a survey response deadline shortly after the PC meeting ended. Out of the 421 papers, we received author feedback on 342 papers, a response rate of 80%. The author responses cover 2050 reviews provided by the PC and ERC. Due to the small number of reviews performed by the ERC, only PC reviewers were given choice to opt-in and receive a summary of the survey response. Most PC reviewers opted to receive a summary. A detailed summary of the survey responses will be provided at the business meeting and our blog article.

Lessons Learned and Opportunities for Enhancement

Need Conflict Database: Marking conflicts has become a long and tedious process. Authors often omit true conflicts accidentally while scrolling through hundreds of reviewer names in the paper submission form. Furthermore, HotCRP provides no mechanism to synchronize simultaneous conflict updates by multiple co-authors. Consequently, many paper submissions do not have all the true conflicts marked at the beginning of the review period. To address the problem of missing conflicts, we leveraged scripts written by past Program Chairs that use online databases (e.g., DBLP) to collect the co-author list of each PC/ERC member. Populating the co-author information into PC/ERC members HotCRP profile was a time-consuming process. This year, the profile updates needed to be done manually per reviewer (due to a HotCRP format issue). This process is clearly not scalable as the PC/ERC continues to grow. We need to invest in a centralized conflict database that can interface with HotCRP to reduce the author and Program Chair burden for managing conflicts. The conflict database would also aid Program Chairs in the difficult task of verifying author reported conflicts.

Sharing Review Assignment Overhead Beyond Program (Co-)Chairs: The review process is heavily dependent on a proper review assignment. We made all our paper assignments manually by selecting from the reviewer suggestions made by our tool chain. While our tools helped prune the reviewer search space, making the paper assignments by hand was a long and tedious process. It took us both 12 days to finish the review assignments for the 426 submissions. As the number of paper submissions continue to grow, two individuals making manual review assignments may no longer be tractable. Thus, it is perhaps wise to share the review assignment overhead across multiple individuals who can manage a topic or a track of the conference.

Revision Letters Beneficial Only if There is Reviewer Continuity: The positive impact from revision letters this year was most likely in the 3-5% range. This is because most reviewers reviewed a paper submission for the very first time. Thus, we did not find much value added for the extra effort spent by authors, reviewers, and program chairs to write, read, and manage the revision letters for this year. We believe that revision letters will be most effective if we can maintain some form of reviewer continuity across conferences. There is already significant opportunity for reviewer continuity, as there is heavy overlap in PC/ERC members across conferences. For example, MICRO this year had an 85% overlap in reviewers from the three most recent architecture conferences (ISCA, HPCA, MICRO). Providing reviewer continuity would require Program Chairs to coordinate and share reviewer information across conferences.

PC Meeting Necessary for Only a Small Fraction of Papers: We felt that the MICRO 2021 program was mostly decided by the end of the online discussion period. Going into the PC meeting, 82 papers met the required threshold for #online-accept, but only 54 were marked as #online-accept. The PC meeting provided a

forum for the paper reviewers to discuss the 28 papers in real time and 24 out of the 28 papers were accepted based on reviewer discussion alone. This was largely by design to avoid randomness associated with whole PC votes where reviewer personalities can potentially influence many people have *not* read the paper. While the PC had the opportunity to ask questions and participate in the discussions for all the papers, of the 131 papers discussed at the PC meeting, only 26 papers went to a whole PC vote. The remainder of the papers were primarily decided amongst the reviewers themselves. Spending a total 20 hours of 122 PC members for 26 papers seems highly inefficient. The saving grace this year is the virtual PC meeting where PC members were able to be involved in paper discussions where they had interest/expertise while also tending to their personal/professional obligations. Alternate approaches where paper discussions are organized by topics and PC members are requested to attend virtual sessions where they have the expertise (or interest) can be more time efficient. This is likely the way of the future as the PC size continues to increase to match the growing number of paper submissions.

Virtual PC Meetings: Based on our experience this year, virtual PC meetings offer abundant benefits compared with traditional in-person meetings. Virtual PC meetings enable global attendance without the overhead of traveling away from professional and personal commitments. Virtual PC meetings also enable networking to some extent while PC members share conflict rooms. To enable networking, we also accommodated a social session each day on Gather Town where PC members could freely roam around having informal conversations with other PC members. While it is not as great as in-person conversations, it was the next best thing and we received positive feedback from the attendees. While virtual PC meetings are great, there are still some downsides that need to be accommodated.

- Differences in time zones make it difficult for the entire PC to get together simultaneously. While getting the entire PC together is not strictly necessary, it is important that all PC reviewers be able to get together. This must be accounted for when making review assignments. We tried to maximize geographic diversity when making review assignments, and that made paper scheduling difficult for us since most papers had one PC reviewer who was required to attend the PC meeting during their late-night hours. Accounting for time zones when doing paper assignments will make scheduling paper discussions at the PC meeting easier.
- Managing conflicts at a virtual PC meeting is much harder than at a physical PC meeting. It is easy to "send somebody out" from a physical PC meeting, but much more difficult during a virtual PC meeting. While Zoom scripts are supposed to seamlessly move PC members into conflict breakout rooms, in practice it turned out to be a manual job. On the first day, we quickly realized that Zoom scripts were not automatically moving PC members between the different rooms. There are hiccups when PC members dial into the meeting using different devices simultaneously which confuses the scripts. As such, we paused at the beginning of each paper discussion waiting for a student volunteer to manually move conflicted PC members. This additional time (which was longer for papers with large conflict lists) was something we had not accounted for in our time management.

Acknowledgements: We are grateful to the collaborative efforts of many people across multiple time zones who helped compose the strong MICRO 2021 program. We would like to thank:

- Graduate students Theodore Michailidis and Zixuan Wang, for all their hard work in testing out HotCRP, managing and uploading conflicts, developing the MightyPC tool, helping to organize the paper discussion schedules for the PC meeting, and organizing the Zoom management scripts; Haolan Liu, Yun Joon Soh, and Hengyu Zhao for managing and taking notes during the virtual PC meeting;

- Dimitris Gizopoulos, the General Chair, for his constant support and advice all along, it has been a joy to get to know you both professionally and personally;
- Emery Berger, Mattan Erez, Lizy John, Christos Kozyrakis, Moinuddin Qureshi, and Jun Yang, for their advice and information on the review process;
- Scott Mahlke and the Steering Committee members for their guidance and support as we experimented with new additions to the review process;
- Mattan Erez, Jun Yang, and Stefano Kaxiras for feedback in developing the hierarchical topic list;
- Omer Anjum, Jinjun Xiong, and Wen-mei Hwu for providing paper assignments using their tool;
- Eddie Kohler for his help with bug fixes on HotCRP.com;
- Saugata Ghose, the web chair, for prompt actions on website updates, along with Stefano Di Carlo for their management of online activities for the virtual conference;
- George Papadimitriou and Mengjia Yan, the publication chairs, for collecting the final paper versions and assembling the proceedings;
- Lisa Wu Wills and Dionisios Pnevmatikatos, Workshop and Tutorial Co-Chairs, for assembling an excellent lineup of workshops and tutorials;
- Samira Khan and Gennady Pekhimenko for organizing Artifact Evaluation (AE) in conjunction with MICRO. Artifact Evaluation promotes reproducibility of experimental results and encourages code and data sharing to help the community quickly validate and compare alternative approaches;
- Our three keynote speakers Anastasia Ailamaki (EPFL), Michael Clark (AMD), and Sean Lie (Cerebras) for spending their precious time to share their thoughts with the MICRO community;
- The 122 PC members and 181 ERC members for their time and dedication performing reviews and attending the two-day PC meeting;
- All the authors for their submissions to MICRO 2021, regardless of the outcome. We hope you found the reviews useful, and a huge thank you for participating in the feedback on the reviews received;

Aamer Jaleel and Jishen Zhao

Program Co-Chairs, MICRO-54