Response to Team Questions

PSAAP III Annual Review

September 30, 2022

Reproducibility

- We plan to use the reproducibility tools in their current form for managing data and experiments.

- If the current state of these tools becomes problematic, we will revisit them, and research and revise as appropriate. This could include using additional community tools such as Popper or others.

Fiesta

- We are not currently working on Fiesta as part of CUP-ECS, as we are using HIGRAD and small benchmarks for research on regular halo communication.

- We will keep Fiesta available to use for evaluating regular halo optimizations for which we cannot use HIGRAD, for example any optimizations that are Kokkos or C++-specific.

AMD and Cray Platforms

1. Access to and documentation for kernel/stream triggering libraries for use testing on AMD/HPE systems

2. Technical point of contact at AMD/HPE for discussion and troubleshooting

3. Additional information on network and other relevant system features (e.g., GPU/memory hierarchy) for forthcoming systems

4. Early access to relevant test systems

GPU Scaling and Unexpected Messages

- The abstractions, models, and optimizations we are developing are explicitly designed to address scaling to large numbers of GPUs.

- For example, the MPI_Pbuf_prepare primitive being examined for GPU partitioned communication was introduced specifically to prevent unexpected messages on GPUs.

Diversity and Inclusion

1. We are actively recruiting diverse students to the project, and our planned increased focus on undergraduate student recruitment will leverage campus diversity programs. This is a particularly promising avenue because of the diverse undergraduate populace at our institutions (e.g., UNM is a minority and Hispanic-serving institution).

2. We are recruiting our diverse undergraduated students into the accelerated Masters program so that they get exposure graduate education and laboratory internships.

3. We are participating in conferences and organizations that focus on increasing diversity and inclusion; for example Prof. Bienz is the faculty advisor for the UNM Women in Computing group. We are examining the possibility of offering a Learning Track workshop at the annual Grace Hopper conference next year.

4. PSAAP students attended the annual Grace Hopper conference this year, and we plan to continue encouraging and supporting active student participation in this and similar opportunities.

Network Emulation

- We are currently focusing on measuring existing systems and using that information to develop models for the performance and tradeoffs of potential new communication primitives.

- Network simulation and emulation systems provide an interesting opportunity to evaluate models and primitives on hardware beyond what is currently available that we will consider using when appropriate.

- This approach would also support the cross-validation of simulation and experimental results.

Technical Details and Evaluation

- Thorough technical specification and evaluation of our new abstractions, including partitioned collectives are planned for this award year.

- As part of this effort, we plan to publish papers that include each of the metrics in question.

- The availability of performance information on GPU triggered versions of these primitives will depend on the availability of robust vendor libraries supporting this functionality.