Hands-on Research and Training in High-Performance Data Sciences, Data Analytics, and Machine Learning for Emerging Environments

Kwai Wong, Stanimire Tomov, and Jack Dongarra
University of Tennessee, Knoxville
[kwong, stomov, dongarra]@utk.edu

ABSTRACT

This paper describes a hands-on Research Experiences for Computational Science, Engineering, and Mathematics (RECSEM) program in high-performance data sciences, data analytics, and machine learning on emerging computer architectures. RECSEM is a Research Experiences for Undergraduates (REU) site program supported by the USA National Science Foundation. This site program at the University of Tennessee (UTK) directs a group of ten undergraduate students to explore, as well as contribute to the emergent interdisciplinary computational science models and state-of-the-art high performance computing (HPC) techniques via a number of cohesive compute and data intensive applications in which numerical linear algebra is the fundamental building block.

The RECSEM program complements the growing importance of computational sciences in many advanced degree programs and provides scientific understanding and discovery to undergraduates with an intellectual focus on research projects using high HPC and aims to deliver a real-world research experience to the students by partnering with teams of scientists who are in the forefront of scientific computing research at the Innovative Computing Laboratory (ICL), and the Joint Institute for Computational Sciences (JICS) at UTK and Oak Ridge National Laboratory (ORNL). The program also receives collaborative support from universities in Hong Kong and Changsha, China.

The program focuses on scientific domains in engineering applications, image processing, machine learning, and numerical parallel solvers on supercomputers and emergent accelerator platforms, particularly their implementation on GPUs. The programs also enjoy close affiliations with researchers at the Oak Ridge National Laboratory. Because of these diverse topics of research areas and backgrounds of this project, in this paper we discuss the experiences and resolutions in managing and coordinating the program, delivering cohesive tutorial materials, directing mentorship of individual projects, lessons learned, and improvement over the course of the program, particularly from the perspectives of the mentors.

CCS Concepts

Applied Computing--Education; **Social and Professional Topics—Computing Educations—Computational Science and Engineering Education; Computing Methodologies—Parallel Computing Methodologies**

Keywords

Computational Science; Educational Outreach; Research Experiences for Undergraduates; Data analytics; Machine learning (ML), hands-on experiences and education; HPC

1. INTRODUCTION

Computational science is an emerging field of study that is truly interdisciplinary, involving researchers from mathematics, computer/information science, and many domain science areas. Computational modeling and simulation have become indispensable tools in nearly every field of science and engineering. The RECSEM predecessor, called CRUSE (2013-2016), and the current RECSEM (2017-2019) programs give students a synergetic set of knowledge and skills that are useful for them to perform scientific research in HPC. These programs aim to deliver a synergetic hands-on research experience to the students by combining the expertise at the Joint Institute for Computational Sciences (JICS) [3] and the Innovative Computing Laboratory (ICL) [4] at the University of Tennessee, focusing at HPC simulation in engineering applications, emergent schemes of numerical mathematics, and state-of-the-art numerical linear algebra software, and data intensive computing. ICL is leader in enabling technologies and software for scientific computing, developing and disseminating high-quality numerical libraries like LAPACK, ScaLAPACK, PLASMA, and MAGMA [5].

The RECSEM program focuses on scientific domains in engineering applications, images processing, machine learning, and parallel numerical solvers on HPC and emergent platforms. **Figure 1** shows the principle idea of the REU program. In general, the program starts with a two-week training session, introducing the students to the supercomputing environment and the common computational methods and tools to be used later. Each student is assigned a project complemented to his/her academics background and computing skill level and solves a computational modeling problem under the supervision of a team of mentors and advisors.

From 2013 to 2018, these programs have admitted a total of 92 students. Forty of them are international students from our four collaborating institutes from Hong Kong and three local students are supported under a separate REU grant from other colleges at UTK. The CRUSE and RECSEM programs have attracted students from 28 different colleges across the nation. Out of the 52 domestic students, 15 are women and 11 are African Americans (3 females). The students worked on a total of 55 different research projects with a total of 23 different lead advisors and 18 mentoring research staff and student associates at JICS and ICL. The program also enjoys tight collaborations with researchers at the Oak Ridge National Laboratory (ORNL). Given the scope of activities and size of students and staff in making this program a fruitful experience for the participants, we discuss the experiences and resolutions in managing and coordinating the program, directing mentorship of individual projects, and lessons learned via exemplary data science projects building on native linear algebra from ICL.

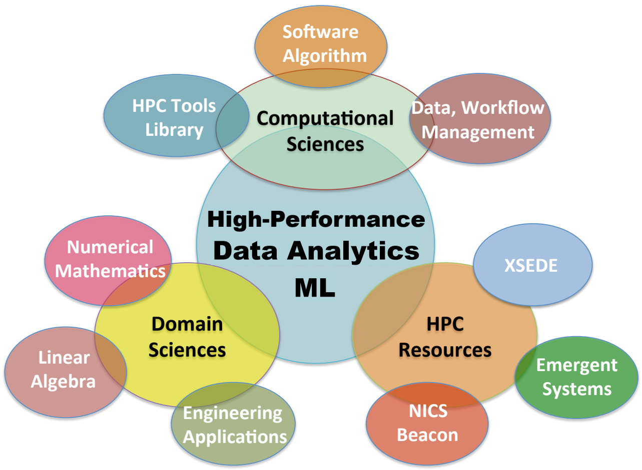

Figure 1: Design of the RECSEM Computational Science program

2. PROGRAM DESIGN and PLAN

The CSURE program started in 2013 and lasted for four years. The revised RECSEM program streamlines the operation and begins in 2107. These programs draw from the computational sciences experiences of JICS staff and the expertize of numerical linear algebra building on the HPC platform from ICL.

The principle goal was to promote the ability of undergraduate students to succeed in a research-oriented program in computational sciences. ***Hence the REU programs seek to mimic the pace and intensity of graduate-level or industrial-level projects with well-defined deliverable deadlines.*** The intention is to provide the participants a good knowledge of how a graduate project is organized and executed. In addition, its intellectual focus is not only to push for publishable research outcomes, but also to expose the students to research experiences through appropriate levels of motivations and accomplishments. These are major reasons we choose to do a ten-week long research program, giving students enough time to master the skills in accomplishing their research goals.

While the primary goal of these programs is to develop students' interest in pursuing research careers in computational sciences, we also provide strong professional development, post-program development opportunities, and social networking for the REU participants among themselves. Students are encouraged to continue their research activities at their home institutions afterward. There are several major tasks that the students are asked to follow. These tasks start with an informal in-class presentation, a midterm lecture presentation, an open poster presentation, and conclude with a final presentations and a final report in the last week. These tasks aim to gradually assist the students towards finishing their research goals in time. A detailed listing of the program is available at the program's webpage, www.jics.utk.edu/recsem-reu.

2.1 Schedule of the REU Program

To deliver such a diverse program, a well-planned step-by-step schedule for the entire summer is desirable to be in place by early December. Event items for the preparation period include logistical arrangements, program announcement and recruitment, selection of students, payroll registration, social activities, preparation of training materials, evaluation instrumentation, mentor selection and training, and most importantly identification of research projects and mentoring teams. Following that will be the ten-week summer program starting the first week of June and ending the first week of August. A typical daily schedule for the last three years can be found on the RECSEM webpage [1]. A typical timeline of the program is listed in the following table.

Table 1: Timeline of the REU program

Jan. - March	Student recruitment and research project identification
March	Student selection and research project selection
April - May	Prepare training materials, setup research plan, post detailed schedule
May	Mentor training, prepare reference materials, coordinate travel and logistics
First day	Goal statement, projects assignment, schedule, survey, social issues, Q&A,
1^{st} week	Training and hands-on workshop, meeting mentor, define and formulate research goals and plan of projects
2^{nd} week	Students finalize research plan with mentors, 1^{st} social gathering
2^{nd} – 4^{th} week	Preliminary study, in-class presentation
5^{th} week	Mid-term presentation, 2^{nd} social
6^{th} – 8^{th} week	Research and HPC implementation, final poster
9^{th} week	Prepare for final presentation, extending work
10^{th} week	Final presentation, project report, concluding
Last day	Survey, Q&A, retrospective movie, summary
August	Summarize results, follow up with students for possible extended work
September	Survey report, final report, project continuation
October	Final NSF yearly report submission

The last week of the program is reserved for reporting, presentations, surveys, and meetings with students. It is important to have a detailed check-out list for each student and a cordial discussion session with each student. The discussion session involves soliciting general impressions from each student, including upsides and downsides of the program, ideas for improvement, and future opportunities for project work and graduate school. These discussion sessions provide valuable insights to the advancement and improvement of the program.

2.2 Recruitment and student selection

The NSF Computer and Information Science Engineering (CISE) directorate has a joint recruitment program for REU students [2] but the program opts to do additional recruitment because of the diverse, interdisciplinary nature of the program. We rely on recruitment through emails and contacts with collaborative institutes of JICS and ORNL, particularly with an established outreach partner, Morehouse College in Atlanta. Many of the applicants are highly recommended students through the contacts of our collaborators.

Candidates considered for the program fill out an application form and write a short essay describing their background, interests in science, and their goal statements. This information is used to select students and then to assign them to work on the proposed core science domains, to ensure that the specific proposed projects are beneficial to the students and matched to their interests, background knowledge, and skills.

Student selection is not always a straightforward process because of the diverse, multidisciplinary nature of this program and the challenge in finding participants that match for the various research topics. A group of mentors meets to iterate over the applicants, ranking the students for their suitability to the program and the research topics. The deadline for applications is in the late February but generally moved back depending on the need for more applicants interested in specific research topics.

Participants are selected based on three major factors: the nature of their home colleges, their interests and background, and their letters of intent and references. Students from smaller schools with fewer research opportunities are preferred in order to expand the national research community. Rising senior students are preferred. GPA is a deciding factor only if two candidates have comparable qualifications. Over the course of CSURE, we have not seen that GPA is necessarily predictive of success with the program.

Two students are usually assigned to work on one research subject. Each pair starts off together but often splits up to work towards separate research aspects of the same topic at the midpoint of the program.

Acceptance letters will be sent out as soon as the first deadline is passed. Getting a written commitment from each accepted student is important. A second set of acceptances is always needed as there are always students declining to attend. Declination letters also need to be sent out in a timely manner; however, it’s wise to keep in contact with a few applicants in case of unexpected availability. There are cases that students withdraw late in April for sundry reasons.

2.3 International Students

Having a group of foreign students is an enlightening element to this program. The goal is to bring together students from different backgrounds in cultural thinking and education pathways, hoping to broaden the perspectives and understandings of ways of approaching solving problems. Hong Kong students have participated in the UTK summer research program under the direction of Dr. Kwai Wong over a decade. The international students are funded by their universities. Unlike the selection of the domestic students, these students are selected competitively primarily based on their academic achievements. The students are usually highly academically motivated and look for attending graduate school in the US, a fact confirmed by tracking their options after graduation. Pairing the domestic and the foreign students is done whenever a project permits. We had six pairs of students, so far and all of them worked out wonderfully, complementing each other in research efforts. Foreign students generally have better methodical skills while domestic students are more resourceful and investigative. Overall the domestic students are impressed by the mathematics and algorithmic training of others, while the foreign students gain tremendously in open minded ideas and team efforts. As the program goes by, the students mix extremely well and enjoy sharing thoughts as well as social activities beyond the academic ones, such as cooking, music, and playing video games together.

2.4 Logistic Support and Social Activities

There are complicating issues for the summer program. The program includes students from a foreign country. It has mentors from UTK and ORNL. The students will have access to supercomputers at NICS and XSEDE [14]. Conference and meeting rooms have to be arranged. Visitor badges to ORNL must be processed. The students each have separate travel plans. Housing must be arranged. Reservation of venues for the planned activities such as group photo sessions, lecture presentations, poster presentations, and social gatherings are done early to ensure availability. All of these issues require timely efforts, coordination, patience, and most importantly, a good supporting team for a successful program.

It is hard to foresee some of the logistic issues for a new program and usually takes two to three years to find good solutions for them. We recognize that getting the support from the school and the involvement of the research office do help tremendously.

Housing for students is the most urging logistical issue to be resolved and must be prepared early. The entire group should stay together in the arranged housing to help them to blend together socially. It helps to grow a solid bond among the students by organizing group activities and encouraging the students to create their own activities. We determine the housing in early February and proceed to place the students as soon as we have finalized the list.

Over the years, we have tried many options to minimize the cost but eventually settled in on-campus housing at a reduced rate negotiated with the help of our research office. In addition, a number of offices and meeting rooms are arranged to host the students in close proximity. Co-locating the entire group and student helpers in one or two rooms strongly enhances the cohesiveness of the program. There are also meeting rooms available nearby for private discussions.

JICS and ICL have research staff, students, and administrative staff. The JICS mentors include UT faculty, staff members in ORNL research groups, and joint faculty with appointments at both UTK and ORNL. The REU program has benefited tremendously from this infrastructure and staff support.

The program starts off with a campus walk and a group lunch on the first day. There are also two organized gatherings in the apartment complex. The students also participate in activities organized for undergraduate summer interns by the UTK office of research. Such activities include a tour of the Neyland football stadium and a few luncheon talks about graduate school application and scholarship information. Some highlight activities include a trip to the Great Smoky Mountain National Park or the Fall Creek Falls State Park, the Knoxville zoo and a tour to the Spallation Neutron Source facility at ORNL. These social activities help to bring the group together and improve morale.

Importantly, we have arranged a local student to serve as the lead of the group, helping the group to resolve some of the logistic issues in town.

2.5 Computing Resources

The JICS facility represents an investment by the state of Tennessee and features a state-of-the art auditorium, conference rooms, and suites for students and visiting staff. It also provides the access to different parallel computing platforms available at NICS and XSEDE [14]. The ICL has expertize in the fundamental building block of numerical libraries on HPC systems, with emphasis on GPUs. In particular, RECSEM uses the MAGMA libraries to build new data analytics and ML capabilities, e.g., MagmaDNN [6-13], as well as computational support for applications in various fields, as illustrated on Figure 2.

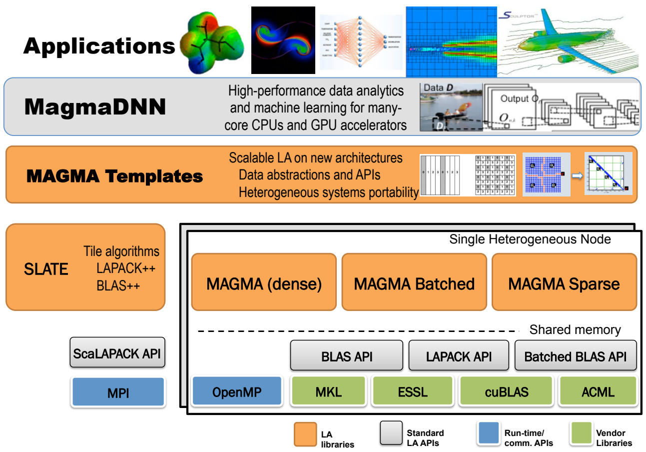

Figure 2: Software stack for high-performance data sciences in RECEM using linear algebra, data analytics and ML.

In the RECSEM program, we turn to XSEDE to support the computing need for the research projects. An educational allocation is obtained to access resources in PSC, SDSC, and TACC. Such arrangement has huge impact to the multi-discipline nature of the program that we organize, not just the variety of hardware platforms ranging from traditional core-based component to various types of accelerators, but also the availability of software and the interactive access for development and testing purposes. Matlab is openly available on XSEDE's *bridges* system, which helps tremendously. The GPU platform available on XSEDE's *comet* and *bridges* platforms provide excellent computing platforms fro data science projects. In addition, we have also arranged individual multicore workstations fitted with a low end P100 compatible GPU card used for code development. These workstations provide a good alternative to accessing supercomputing remotely.

3. RESEARCH WORK AND MENTORSHIP

This REU program addresses the growing importance of computational sciences in many advanced degree programs. The agenda of the program is organized around a synergistic set of ideas and practices that are common to many scientific domains. The focus of the projects leverages the multidisciplinary expertise of the staff in JICS, UTK, and ORNL.

In order to provide students with the most valuable and realistic experience in computational sciences we have identified several different areas of significant interest and expertise within out organization. A participant will select a scientific area in which he/she would like to be involved. Students are paired to work as a team together with their assigned scientific mentors and advisors.

One of the major theme in the RECSEM program is to deliver the fundamental concepts of numerical linear library which is the major building block of computational intensive and data driven sciences. We will provide exemplary data science projects using the home grown numerical libraries, LAPACK and Magma.

3.1 Stages of the Research Plan and Mentor Experiences

The schedule of the research program is organized into six progressing stages shown in Table 1.

Table 1: Timeline of the program

Stage	Week	Project targets
Training	1.5	Lectures, exercises, research skills
Science Study	1.5	Overview and set research plan
Formulation	1.0	Objectives and algorithm, **short talk**
Prototyping	2.0	Description, **midterm presentation**
Implementation	2.0	Results, **poster presentation**
Concluding	1.0	**Final presentation and report**

The program begins with a kick-off meeting to highlight the agenda of the program and introduce the team of researchers and staff working in the project. A set of tutorials containing a series of lecture materials and a clear calendar of schedule of work and activities is listed on the program website [1] and is available to the participants at start.

The first day is reserved for payroll paperwork, initial survey, introduction, exchanging email addresses, Q&A, introducing a local student team leader, and a campus walk. ***A list of safety reminders, health concerns, complaints, and emergency contact information is discussed in detail.*** During the program sessions, occasional health issues arise and absence and sick policy will be given.

The first stage of the program includes an in-depth introduction to the use of supercomputers, including programming languages and compiling procedures, batch queuing systems, and I/O tools. Training activities include classroom instruction, hands-on exercises, research and modeling design, and computational studies. Tutorials come with hands-on exercises that put them to work as teams. Recognizing there is an uneven level of expertise in computing, we always pair the team up to compensate for their knowledge in computer and domain sciences. The introductory sessions intersperse lectures with discussion questions, emphasis on group effort on problem solving, and hands-on exercises.

Research topics are assigned to students ahead of time; however, should they change their mind, they may do so in the first week. An important task of the first week is to give specific assignments to students to help them begin making progress on their research topic.

The second week of work includes an introduction to the domain science areas and the specific project content assigned to each team of students. This involves hour-long talks by the subject mentors. We avoid asking students to spend time on learning materials that they will not use. The rest of the week moves to scientific study with literature review, reading and discussing relevant articles, and hands-on practice with relevant computational methods and tools.

In later stages, students start to draw their research plan under the direction of their mentors. Every student will conclude the research plan and project goal in three weeks. Student progress toward their planned goals is evaluated frequently during the program. Mathematical formulation and algorithmic prototyping and testing are then followed. The last week will be reserved for concluding the project, presenting the final results, and finishing the final report.

3.2 Progress Oversight and deliverables

The program has five deliverables. These are designed to steer the students to finish their projects on time. The timeline of these deliverables is listed clearly on the webpage and emphasized in the first week of the program. ***The first deliverable is a short in-class summary talk*** of the research topic and the approach. ***The second deliverable is an open presentation*** of their research work and initial results. It is aimed to orient the students to focus on their works, help crystallize the approach, and make students aware of the project timeline.

The third deliverable is a public poster presentation, organized with other groups of REU students. The posters help students to organize their results. Students also have the opportunity to review other projects and potentially seek ideas to improve theirs.

The last two weeks of the program have the students working toward ***concluding their projects with a final presentation***. Each presentation lasts for 40 minutes and usually receiving a number of questions from their peers and attendants. Final presentations are great experiences for the participants and represent a concluding milestone for their research endeavors.

The last piece of work is a report. This is, in fact, a continuous process from the beginning, with students organizing their weekly summaries and articulating their results in detailed reports. Each student is encouraged to keep a weekly summary report. The final report will be a combined work that documents the student's progress and findings. ***Yet in fact, this turns out to be the most demanding part of the 10-week program. Hence, it is very important to keep reminding students throughout the program to work on documenting their efforts and results.***

3.3 Research Projects and Mentorship

The research topics available for the participants span a wide range of scientific and engineering domains. Each of the areas corresponds to significant capabilities at NICS or ORNL with active researchers and projects. All projects include hands-on experience and use of parallel computing in the various scientific and engineering domains of the program.

Research projects are selected based on the expertize of the core team of mentors and the backgrounds of applicants. Descriptions of previous research projects, from traditional domain sciences to cross-disciplinary data computation are listed in the following sections.

Data Analytics and Machine Learning

A common theme for all projects is the use of high-performance numerical libraries, data analytics, and machine learning. Several projects are specifically targeting the development of such capabilities. Examples are the development of MagmaDNN [6-13], a high-performance data analytics and deep neural networks (DNN) framework for manycore GPUs and CPUs. Students learn state-of-the-art algorithms and performance optimizations techniques for data analytics and machine learning, implement them in open source library, and also help other students use these capabilities for data-driven science projects. Projects have included the development of the MagmaDNN DNN framework [6], extensions with convolution algorithms [10], including Winograd [7], mixed-precision FFTs using the new FP16 Tensor Cores units on Nvidia GPUs [8, 9], parallelization and addition of new features [11], hyper-parameter optimization framework [12], and scalability improvements [13].

Computational Engineering Applications

Engineers have been using supercomputers to analyze and resolve many challenging problems for many years. Nowadays, computer simulation has become a mandatory step in the process of design and development for many industrial applications. Projects completed by the participants include climate and pollution transport simulations, biomechanics modeling, traffic flow computation, and power system evaluation.

Numerical Mathematics and data science projects

Numerical mathematics is the building block of computer simulation of every scientific application. A science problem can usually be modeled by a set of mathematical equations and then numerically solved on computers. The effectiveness of these solvers is often determined by the combination of the specific choice of numerical schemes and implementations, which is particularly true on HPC platform. A theme of the research projects is to develop efficient numerical schemes for equation-based and data-based applications, generally needed in a lot large-scale engineering simulations. Projects completed by the participants include continuous and discontinuous finite element formulations, machine learning algorithms for microscopy and brain signal problems, topological analysis of high-dimensional data, variational inequity problems, and stochastic programming modeling.

Linear Algebra Software Implementations

Linear algebra is the backbone of HPC. Many numerical simulations depend of efficient linear algebra libraries to scale on HPC platforms. Projects completed by the participants include parallel dense solve implementation on GPU and multi/many-core CPU processors, fMRI data analysis using Intel DAAL library, randomized SVD calculations, and building workflow and graphical user interface for specific applications.

Natural Sciences Applications

Computational chemistry, physics, and geography have big footprints on large scale supercomputers. Participants have completed a number of projects in quantum mechanics, molecular dynamics, neutron image reconstruction, and GIS modeling.

The lead mentors are designated persons committed to the program. Mentors are selected based on their availability and commitment. They are leading researchers in their domain science working at UTK and/or ORNL. The team of mentors defines the major element of success of the program. They are chosen early and are involved in the selection of students. The student research projects vary every year but fall in the scope of the major program subject areas. In general mentors meet with their students at least twice a week and are available for questions. Graduate students of the mentoring team are in general also available to provide constant guidance and direction to the students. Given the reality that travel for conferences, reviews, or other purposes makes it likely that mentors will be occasionally absent, having additional advisors is important to ensure steady progress. General oversight of the research progress by the program director is also recommended. Regular discussions between the program director and the mentoring team are also helpful.

4. ACOMPLISHMENTS, CHALLENGES, AND LESSONS LERNED

4.1 Survey

Evaluation of the program is centered on the toolkit distributed by the NSF CISE REU program as published by the University of North Carolina, Charlotte [2]. The evaluation provides the mentoring team with regular feedback for ongoing assessment of the program via in-person meetings along with formal mid-program and annual reports. Reports include evidence-based recommendations for program improvements in the form of clear actions items that program directors can apply directly to further program improvement. A final summative report examines and determines to what extent the program succeeded in meeting its stated goals.

Surveys for the students are performed at the start of the program and at the end of the program. We use the standalone A La Carte student survey from the CISE REU toolkit [2]. In order to evaluate the project's impact on participants, students are given pre- and post-evaluation surveys that assess their attitudes toward and interests in computational science, as well as their knowledge of computational science and its use in the domain focus area. The results of these surveys each year guide modifications to the project for future years. Surveys and summative evaluations are independently instrumented either professionally by a contract agency or a person that is familiar with the process. The REU program engaged Dr. Christian Halloy, a retired computational science leader to conduct the summative program evaluations. Dr. Halloy conducted pre- and post-participant surveys, a personal discussion with each participant, and provided a detailed final report. He also attended and critiqued the progress of the students' final lecture and poster presentations.

In summary, U.S. students' scores for survey constructs of self-efficacy, graduate school intentions, computing attitudes, help-seeking and coping, scientific leadership, and scientific identity were favorable at the start and end of the program (means above 4.00 on a 1 to 5-point scale for pre- and post-surveys). The largest improvement gain for REU participants after the 10-week period was found for the research skills and knowledge scale with a mean increase from 3.93 (SD = 0.48) to 4.38 (SD = 0.41). Overall, participants were satisfied with the program (M = 4.25, SD = 0.43) and their mentor (M = 4.12, SD = 0.63).

Participants rated their mentors quite highly for all the indicators, the highest average score being for "approachable" (M = 4.89, StdDev = 0.31), while the lowest average score is seen for "accessible" (M = 4.56, StdDev = 0.50) which is nevertheless quite high per se.

In general, the following recommendations were provided as examples of practices the REU may consider to include or maintain to ensure continued and future success of the program.

1) Expand the evaluation to include feedback from additional stakeholder groups (i.e., faculty advisors/mentors and program administrators) in order to gain an additional understanding of the REU program.
2) If possible, create a system to follow the student participants over time to assess additional project impacts on a long-term basis. (e.g., graduate school attendance, career choice, presentation and publications, awards and honors, etc.).
3) Continue to integrate strategies that will enhance the experience across diverse backgrounds, considering that students in the program possess differing academic backgrounds and research preparation.
4) Carefully recruit faculty and graduate students who will be available throughout the duration of the program. Consider a back-up strategy to support students if a volunteering mentor is unavailable during parts of the program.
5) Continue to include and potentially increase hands-on instruction at the beginning of the program to engage and motivate participants.
6) Continue to provide opportunities for students and mentors to network at the start of the summer and throughout the research experience.

4.2 Accomplishments, Challenges, and Lessons

The success of the program counts on dedications and efforts of our mentors. We have instituted a total of 55 different projects with only a few of them sharing some levels of similarities. Selection and availability of mentors are constant subjects of concern even we enjoy having a large pool of volunteer scientists. As this REU program continues, we learn to streamline the dimension of projects and maintain the core subject areas the team of resident mentors and PIs are familiar with. Often the program director has to be prepare to spend over half of his time a day answering questions for the entire group.

Parallel computing to many participants has a steep learning curve, pairing students in their comfort knowledge backgrounds is essential to get a project done in time. In addition, to avoid duplication effort within a team, we often design a team project with multiple themes allowing every student has his/her research own contribution.

Human dynamics, emotion, frustration, and conflicts among students, however rare, are unavoidable issues. Listening, patience, caring, and professionalism are appropriate answers to most. After all, we put research experiences as the primary theme of the program. Having international students gives a good mix of cultural interaction, in fact, improves overall group dynamics.

4.3 Program Outcomes and Impacts

Over the last five years, we have instituted a multidisciplinary computational sciences REU program that encompasses 55 different projects, including a total of 92 students from 28 colleges. This program has established a continued relationship with undergraduate institutions such as Morehouse College in Atlanta, Maryville College near Knoxville, Centre College in Kentucky, and Slippery Rock University in PA. This is important in sustaining long-term viability of the REU program, which can continue to evolve and improve from listening the feedbacks and suggestions from our partner colleges. The outcomes of the students' research work included six sponsored conference presentations, three conference papers and a number of conference and journal papers to be submitted. A list of their reports is posted in the RECSEM website [3]. Close to 70% of the students have gone to or are applying for graduate schools. The program director has maintained yearly contacts with the participants. This is important to our sponsor. It helps to track the progress of the students and overall impact to the REU program.

5. CONCLUSIONS

This REU program intends to provide participants with an experience with a similar level of effort as in graduate school. The program provides students an exposure to research with high performance computing applied to a variety of scientific applications. In three summers, we have resolved many problems and met even more challenges. In particular, the following items summarize the highlights of the program:

1) A well-defined step-by-step timeline leading to the end of the program is in place in early December.
2) The participants are selected based on three major factors: the nature of their home college, their interests and background, and their letters of intent and references.
3) The project assignments are sent to students ahead of time.
4) Getting a written commitment from each enrollee is important.
5) A midterm preliminary presentation of the research topic and the approaches of the research, is very important.
6) Housing for students must be prepared in the early stage of the program. The entire group stays together in the arranged housing to get them to blend together socially.
7) A program director is important, with regular availability to the participants.
8) Co-locating all students and helpers in a multi-purpose lecturing room enhances the cohesiveness of the program.
9) A list of safety reminders, health concerns, and emergency contact information is discussed in detail in the first day.
10) An effective team of mentors represents a major element of success of the program. They are chosen early and are also involved in the selection of their students.
11) We have arranged a local student to serve as the site lead to the group, particularly for social activities.
12) The most demanding part of the 10-week experience is the final report. The program director should keep reminding participants and constantly check for progress.
13) A detailed checkout list for each student and a meeting with each student before the program ends are needed.
14) Surveys for the students are performed at the start of the program and at the end of the program.

6. ACKNOWLEDGMENTS

This work was conducted at the Joint Institute for Computational Sciences (JICS), sponsored by the National Science Foundation (NSF), through NSF REU Award #**1262937 and #1659502**, with additional Support from the University of Tennessee, Knoxville (UTK), and the National Institute for Computational Sciences (NICS). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Computational Resources are available through a XSEDE education allocation award **TG-ASC170031.**

7. REFERENCES

[1] RECSEM : www.jics.utk.edu/recsem-reu

[2] http://reu.uncc.edu/cise-reu-toolkit/results-cise-reu-toolkit

[3] Computational Science for Undergraduate Research Experience, 2013-17 internal reports: http://www.jics.utk.edu/csure-reu/csure13/projects, http://www.jics.utk.edu/cure-reu/csure-14/projects, http://www.jics.utk.edu/csure-reu/csure15/projects, http://www.jics.utk.edu/csure-reu/csure16/projects, http://www.jics.utk.edu/recsem-reu/recsem17/projects

[4] Innovative Computing Laboratory: http://icl.cs.utk.edu

[5] S. Tomov, J. Dongarra, M. Baboulin, “Towards dense linear algebra for hybrid GPU accelerated manycore systems”, Parallel Computing, vol. 36 (issues 5-6), June 2010.

[6] L. Ng, K. Wong, A. Haidar, S. Tomov, and J. Dongarra, “MagmaDNN – High-Performance Data Analytics for Manycore GPUs and CPUs”, 2017 Summer Research Experiences for Undergraduate (REU), Knoxville, TN, December, 2017. (Available at: http://icl.cs.utk.edu/projectsfiles/magma/pubs/71-MagmaDNN.pdf)

[7] L. Ng, S. Chen, A. Gessinger, D. Nichols, X. Cheng, A. Sorna, K. Wong, S. Tomov, A. Haidar, Ed D’Azevedo, and J. Dongarra, “MagmaDNN 0.2: High-Performance Data Analytics for Manycore GPUs and CPUs”, 2018 Summer Research Experiences for Undergraduate (REU), Knoxville, TN, January, 2019. (Available at: http://icl.eecs.utk.edu/projectsfiles/magma/pubs/MagmaDNN-v0.2.pdf)

[8] X. Cheng, A. Sorna , E. D’Azevedo, K. Wong, S. Tomov, "Accelerating 2D FFT: Exploit GPU Tensor Cores through Mixed-Precision," The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC'18), ACM Student Research Poster, Dallas, TX, November 11-16, 2018.

[9] A. Sorna, X. Cheng, E. D’Azevedo, K. Wong, S. Tomov, “Optimizing the Fast Fourier Transform Using Mixed Precision on Tensor Core Hardware”, IEEE 25th International Conference on High Performance Computing Workshops (HiPCW), December, 2018.

[10] S. Chen, A. Gessinger, and S. Tomov, “Design and Acceleration of Convolutional Neural Networks on Modern Architectures”, 2018 Summer Research Experiences for Undergraduate (REU), Knoxville, TN, 2018.

[11] D. Nichols, K. Wong, S. Tomov, L. Ng, S. Chen, and A. Gessinger, “MagmaDNN: Accelerated Deep Learning Using MAGMA”, PEARC’19 (submitted), July 28-August 1, Chicago, IL, 2019.

[12] F. Betancourt, K. Wong, E. Asemota, Q. Marshall, D. Nichols, and S. Tomov, “OpenDIEL: A Parallel Workflow Engine and Data Analytics Framework“, PEARC’19 (submitted), July 28-August 1, Chicago, IL, 2019.

[13] D. Nichols, F. Betancourt, S. Tomov, K. Wong, and J. Dongarra, “MagmaDNN: Towards High-Performance Data Analytics and Machine Learning for Data-Driven Scientific Computing”, ISC’19, Scalable Data Analytics in Scientific Computing (SDASC) Workshop (submitted), June 16-20, Frankfurt, Germany, 2019.

[14] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, Nancy Wilkins-Diehr, "XSEDE: Accelerating Scientific Discovery", Computing in Science & Engineering, vol.16, no. 5, pp. 62-74, Sept.-Oct. 2014, doi:10.1109/MCSE.2014.