Summary:
On July 29-30th, 2019 the Southern California Earthquake Center (SCEC) hosted 34 geoscientists, engineers, computer scientists, and software developers at the San Diego Supercomputer Center (SDSC) on the University of California, San Diego (UCSD) La Jolla campus to discuss the future of SCEC research computing activities. During the 2-day workshop, we discussed SCEC’s research computing needs, reviewed current SCEC computing projects, and summarized the current funding environment to identify future opportunities for SCEC research efforts. We also discussed new strategies for supporting SCEC research computing over the next five years. The workshop agenda, presentations, and participant list are available at: www.scec.org/workshops/2019/computing.

SCEC’s future computing needs will depend on its research and development goals. Over the next five years, if SCEC’s emphasis is on basic research, then computing needs will require maintenance and improvement of earth models and continuous development of advanced software. If SCEC plans to perform large-scale computations such as for earthquake rupture forecast (ERFs) model development such as UCERF-like projects, simulation-based probabilistic seismic hazard analysis (PSHA) projects, and/or for building code applications, then the organization will need significantly more computing resources to support the implementation, verification and validation required for these applications.

Motivation and Goals for Workshop:
SCEC researchers makes extensive use of research computing resources and tools including high-performance and cloud computing for observed and simulated data, community earth models, open-source scientific community software, and complex research computing software stacks. This workshop was motivated by the recognition that both (1) SCEC’s scientific research goals and (2) NSF/USGS/DOE open-science research computing environments are changing. The purpose was to bring together experts including geoscientists, engineers, computer scientists, and software developers to help SCEC develop a research computing plan that anticipates and accommodates these ongoing changes. The planning horizon targeted five years, so the discussions were relevant to the next SCEC proposal. The workshop conveners, Philip Maechling, Yifeng Cui, Tran Huynh, and Christine Goulet, defined four goals for the workshop:

1. Assess the current and future research computing needs of the SCEC community.
2. Prepare SCEC researchers for future NSF and DOE computing environments.
3. Design a research computing organization that will support future SCEC research.
4. Initiate future collaborative SCEC research computing projects.
Meeting Participants:
The participants included geoscientists, computer scientists, civil engineers, software developers, university faculty, research staff, post-docs, computer center staff, and government researchers. The SDSC conference room supported remote participation via webconference technology. A total of 34 people participated in the workshop, including those joining remotely (marked with a * below):

Chaitan Baru (SDSC), Greg Bauer (NCSA), Yehuda Ben-Zion (SCEC), Greg Beroza (Stanford), Scott Callaghan (SCEC), Yifeng Cui (SDSC), Rafael Ferreira Da Silva (USC/ISI), Christine Goulet (SCEC), *Ruth Harris (USGS), Elizabeth Hearn (Capstone Geophysics), Tran Huynh (SCEC), Thomas Jordan (USC), Christos Kyriakopoulos (UCR), Rich Loft (NCAR), Philip Maechling (SCEC), *Xiaofeng Meng (USC), Kevin Milner (SCEC), Mai Nguyen (SDSC), Mike Norman (SDSC), Kim Olsen (SDSU), Edric Pauk (SCEC), Dmitry Pekurovsky (SDSC), Andreas Plesch (Harvard), Doriam Restrepo (EAFIT), Zach Ross (Caltech), Daniel Roten (SDSU), William Savran (SCEC), Bruce Shaw (Columbia), *Fabio Silva (SCEC), *Mei-Hui Su (SCEC), *Ricardo Taborda (EAFIT), Frank Vernon (Scripps), Kyle Withers (USGS), Ellen Yu (Caltech)

Session Summaries:
The workshop sessions were organized around an important question related to SCEC Research Computing.

SESSION 1: What is SCEC Research Computing?
Christine Goulet presented an overview of (1) SCEC “special projects”—many of which have a computational emphasis and/or are synthesis activities that integrate SCEC research results into broad impact applications; and (2) SCEC’s supercomputer allocation development process which provides SCEC research with access to national open-science supercomputers. SCEC has developed a software eco-system of seismic hazard software including OpenSHA, UCVM, Broadband Platform, CyberShake, AWP-ODC, Hercules, RSQSim, and CSEP. In his following presentation, Philip Maechling described the challenges SCEC faces in supporting our research computing capabilities. He emphasized that software sustainability is a complex challenge, requiring long term strategies and planning. One strategy is to continue improving SCEC software development towards CIG and other open-source software organization best practices, to help preserve the research time invested in SCEC codes. Many codes will require improved software documentation, routine continuous testing, and a formal software approval process for new software releases.

The group discussed how computing and testing requirements differ for “basic” research codes relative to “large-scale multidisciplinary” seismic data products that tend to have a broad impact. Basic research codes are used by a small group of researchers, with multiple versions, and are frequently modified, often without formal version control. Large-scale multidisciplinary codes are generally carefully managed by a scientific organization, have an authoritative version, and often have rigorous regression test suites. The need to iteratively re-test large multidisciplinary codes (at full-scale, as improvements are made, and with a forward-looking
plan for engagement with computing architecture changes) significantly increases the amount of computing time needed to develop these codes.

**SESSION 2: What research computing capabilities are needed to support SCEC research activities?**

SCEC Director Yehuda Ben-Zion described new research initiatives anticipated for SCEC for the next five years. He expressed that SCEC needs to identify and pursue compelling new earthquake research topics to maintain research support. SCEC research computing must support the new research activities, such as data processing from a planned dense fault zone observatory deployment, development of accurate representations of the top 350m of the Earth, merging of multi-resolution velocity models, and active monitoring of seismic transients. He also discussed new data management and machine learning based data analytics needs for SCEC community models, and for observational and simulation results.

These new research directions will call for new computing and data management technologies. Computer scientists in the group suggested that future SCEC computational research will need both high-performance computing (HPC) and data analytics, a computational research trend called Big Data and Exascale Computing (BDEC) convergence. SCEC’s current strength is in the more traditional parallel HPC technologies on shared clusters and will likely benefit in the future from expertise in data analytics technologies using virtual environments.

Furthermore, the Internet of Things (IoT) is likely to impact seismology in the future. Multiple examples were discussed including the development of inexpensive ground motion sensors installed by private entities, such as energy companies, and the use of fiber-optic cables as distributed ground motion sensors. As the number of sensors increases, it is likely that edge computing will be needed to perform data aggregation and data reduction by using advanced technologies such as wireless sensors to collect and process data at the source of the data rather than at a remote data center.

During the breakout session, Christine Goulet requested feedback on examples of key SCEC computing applications targets for the next five years: (1) M8, a large wave propagation simulation, (2) CyberShake, a large, heterogenous distributed workflow, (3) a mining seismic wavefield follow-up, a data intensive machine learning-based application, and (4) post-earthquake analysis, an on-demand computing application. These applications were discussed in order to capture the wide-variety of SCEC computing needs. SDSC Director, Mike Norman, described how the new SDSC system “Expanse” is designed to advance research that is increasingly dependent upon heterogeneous and distributed resources, including public cloud, support 99% of scientific computing, excepting on the largest “1%” jobs which are best suited for leadership class systems.

**SESSION 3a: What are the emerging technologies to consider in research computing?**

In this session, researchers discussed the computing requirements to advance their research. Greg Beroza presented machine learning methods developed at Stanford and emphasized the need for geoscience students to train in machine learning techniques. He described his group’s
main tools as Tensorflow with some current investigations into PyTorch. He emphasized the need to broaden the application of AI across SCEC in areas such as network seismology, LiDAR topography, satellite data (e.g. GNSS, InSAR), simulation/modeling, and aftershock forecasting. Rich Loft described National Center for Atmospheric Research’s (NCAR) computing, modeling software, machine learning, and data management capabilities. He presented guidelines for combining physics-based models and machine learning-based models. He also summarized the challenges atmosphere researcher face testing long term climate models, due to lack of historical observations, which has strong parallels to the challenges seismic hazard modelers face testing their own long term models, which also lack historical observations for validation purposes. Chaitan Baru presented SDSC and NSF data management initiatives including the Harnessing the Data Revolution (HDR). He suggested SCEC might explore NSF’s convergence research initiatives (due February 2020) and Open Knowledge Network (OKN). Greg Bauer described how the National Center for Supercomputing Applications (NCSA) supports geoscientific applications, including an ArcticDEM program and SCEC computing activities, using the Blue Waters system, and indicated NCSA’s interest in identifying other geoscientific computing applications with operational or reoccurring processing needs. The group identified SCEC’s Collaboratory for the Study of Earthquake Predictability (CSEP) earthquake forecast testing and CyberShake simulations in support of national seismic hazard map development as possible candidates for on-demand computational products. Mike Norman shared his views in scientific computing based on his experiences with DOE and NSF, involving the Exascale Computing Project (ECP) and other cyberinfrastructure developments. He noted that most DOE and NSF computing systems to come online in the next five years have already been planned and funded, and are in the pipeline. DOE’s current emphasis is on Exascale systems, whereas NSF’s current emphasis is on providing access to advanced computing for an increasingly diverse pool of academic researchers. He also emphasized that SCEC should participate in the leadership-class computing facility (LCCF) Phase 2 requirements gathering and perhaps encourage NSF to support domain-specific peta-apps-type software projects.

SESSION 3b: What new scientific or computing activity will benefit SCEC in the next 5 years?
In this session, participants presented 5-minute lightning talks with proposed research activities. There were 15 talks in this session, covering a range of suggestions. The proposed activities included: earthquake simulator research, development of new PSHA methods, post-earthquake processing including Epidemic-Type Aftershock Sequence (ETAS)-based forecast simulations and ground motion simulations, development of training data sets for seismological machine learning projects, scientific and software engineering improvements to SCEC research software, and other projects. Many of these lightning talk slides are posted on the workshop website.

SESSION 4: What new avenues will and should SCEC research computing take?
Yifeng Cui of SDSC discussed how future HPC changes are being driven by the end of clock speed increases and transistor density’s reaching its limit. HPC system developers are pursuing other methods to continue increasing computer performance. He summarized the top 10 exascale challenges, from energy efficiency, reformulating and co-design solution algorithm, to resilience and productivity, as identified by the Advanced Scientific Computing Advisory
Committee (ASCAC). He also described the increasing significance of data management needs for research computing groups, based on the increasing volume and value of scientific data, and the increasingly complex and heterogeneous computing needs (including HPC, workflows, data intensive computing, machine learning, and cloud computing). While the convergence of HPC, Big Data and deep learning are already happening, many of these computing specialties require very different software and computing infrastructure, and continuous integration and delivery of software stack. He indicated that U.S. is already behind Japan and China in supporting end-to-end earthquake simulations, and emphasized that a nation-wide infrastructure is needed to support computational seismology software research and development that accounts for computing time allocations which have an estimated value of about $26.5M federal investment annually. SCEC computing research needs are likely to increase by an order of magnitude in the next five years. He referred a white paper he and his team drafted in January 2018 for the necessity in creating an earthquake data and software center, in supporting extreme-scale architecture-aware numerical approaches and software development, including a small to middle-size hardware for operational earthquake forecasting and on-demand early warning computing needs.

Rafael Ferreira da Silva presented the Pegasus-WMS workflow concepts that allow researchers to define and automate multi-stage research calculations in shared computing environments that frequently change. He described how Pegasus workflow tools have contributed to the continuing advancement of the CyberShake platform, by automating this extended research calculation and distributing it across multiple supercomputers. SCEC’s CyberShake research and Pegasus-WMS development are a prime example of a collaboration between a domain research group and a cyberinfrastructure development group that can benefit both groups.

SESSION 5: How should SCEC develop opportunities and research computing initiatives?
In this session, Christine Goulet and Phil Maechling shared short presentations to engage the group in discussions. The group explored how SCEC needs both “basic” research computing, and “large-scale multidisciplinary” hazard computing. It was noted that SCEC “large-scale multidisciplinary” application development and evaluation is iterative, and as the scale and resolution of these applications increase, calculations must be repeated leading to large needs. The group recognized that SCEC is one of the few international research organizations that has the scientific breadth and depth needed to improve current “large-scale multidisciplinary” seismic hazard data products, because these products require a wide range of scientific inputs and expertise in velocity models, rupture models, wave propagation codes, and site-response models. A key point made in these discussions was regarding the future of initiatives and funding sources. It has been in SCEC’s tradition to develop projects and proposals driven by science needs, that brought in HPC centers as partners. However, as the funding scenery is changing, it was suggested that in some cases and some calls for proposals, the cyberinfrastructure aspect may be more prominent, in which case the HPC centers could take the lead and partner with SCEC for the application of the new sought-after technologies. The group discussed (1) whether it would be helpful to make this case through white papers that document SCEC’s research computing capabilities and broad impact and (2) strategies for approaching DOE, NSF, USGS, NASA, private entities or other potential funding organizations.
Conclusions:
The wrap-up session re-emphasized that earthquake system science research will have increasing needs for computing resources and services in future years. The required computing resources include computing time, data storage, scientific software, and model development, together with the services of high-performance computing experts, data management experts, machine learning experts, scientific software developers, and computer administrators. As an earthquake science research organization, it will be difficult for SCEC to develop internal expertise in all these fields, so SCEC should plan to develop partnerships and collaborations with computing organizations that have these specialized skills. Furthermore, the importance of research computing in modern earthquake science is likely to impact educational programs in the geosciences, adding research computing skills as essential training in earth science educational curriculums.