SAND2017-5227R

A Unified Data-Driven Approach for Programming *In Situ* Analysis and Visualization: An Interim Report of Sandia Sub-Team Contributions

Janine C. Bennett, Philippe Pébaÿ, Hemanth Kolla, Giulio Borghesi
Sandia National Laboratories*, Livermore, CA

As we look ahead to next generation high performance computing platforms, the placement and movement of data is becoming the key-limiting factor on both performance and energy efficiency. Furthermore, the increased quantities of data the systems are capable of generating, in conjunction with the insufficient rate of improvements in the supporting I/O infrastructure, is forcing applications away from the off-line post-processing of data towards techniques based on *in situ* analysis and visualization. Together, these challenges are shaping how we will both design and develop effective, performant and energy-efficient software. In particular, the challenges highlight the need for data and data-centric operations to be fundamental in the reasoning about, and optimization of, scientific workflows on extreme-scale architectures.

The DOE ASCR-funded project titled, "A Unified Data-Driven Approach for Programming In Situ Analysis and Visualization" (UDDAP), seeks to understand the interplay between data-centric programming model requirements at extreme-scale and the overall impact of those requirements on the design, capabilities, flexibility, and implementation details for both applications and supporting in situ infrastructure. The project leverages the Legion programming model and runtime system [BTSA12], that was developed as part of the ExaCT co-design center. The team spans multiple institutions, including Los Alamos National Laboratory, Sandia National Laboratories, Stanford University, University of Utah, and Kitware. In this report, we briefly summarize research contributions from the Sandia sub-team of UDDAP project, directing the interested reader to relevant publications for details.

The central focus of the Sandia sub-team has been the design and development of a suite of data-driven statistical analysis kernels for *in situ* deployment in the Legion runtime system. The work leverages the team's previous advances in applied mathematics [PTKB16] and includes implementations of key aspects of the MPI-based parallel statistics framework described in [PTBM11] that the team had designed, implemented, and released as part of VTK [Kit10]. The framework supports both moment-based and quanta-based statistical analysis workflows comprising 4 disjoint operations: 1) Learn a model from observations, 2) Derive statistics from a model, 3) Assess observations with a model, and 4) Test a hypothesis.

In [PB15], we summarized our initial Legion porting work, noting the process was straightforward, thanks in part to the statistics package's initial design into four operations (as this had the benefit of cleanly separating computation from communication within the workflow). We found that our Legion implementation, based on *aggregation regions* as surrogates for bulk-synchronous collective operations, exhibited optimal on-node parallel scaling, thereby taking advantage of the multiplicity of cores on each node. In [PB16], we evaluated the overall per-

*Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

formance of the *in situ* system on a number of small-size test beds using a computational fluid dynamics (CFD) mini-application called MiniAero, cf. [FFL15]. We extended these early results in a more systematic study, the results of which were published in the workshop for High Performance Data Analysis and Visualization, co-located with IPDPS16 [PBH+16]. One of the primary benefits of the aggregation approach presented, was that only a small set of well contained code was required to connect the analysis to the main simulation application.

Although aggregation regions provide an elegant and simple solution for expressing collective communication needs, there are drawbacks that limit the approach's scalability: 1) it represents an all-to-one model of collective communication where all contributions to a collective are funneled to one-point (the aggregation region), and 2) the result is not immediately available in the tasks contributing to the collectives but rather to the parent task. Full-scale applications are based on main tasks that exist throughout all or most of the run, and the Legion's fork-join model can be seen as a "pinching" of this SPMD (Single Program, Multiple Data) parallelism. Nearly all *in situ* analyses require some form of global communication and most applications themselves might require frequent collectives, which can make the aggregation region approach incur noticeable overhead. The Legion team is developing a solution to address the root cause of the overheads that is based on automatic control replication [SLT+17]. In the mean-time, an alternative SPMD-style of Legion exists, which is seen as a practical way to achieve scalability until Legion's control replication work is finalzed.

Our most recent work, summarized in [PBKB17], addresses some current limitations of the aggregation-region approach at-scale, by introducing an alternative model for collectives that aligns exactly with the SPMD-Legion philosophy. Our solution removes the all-to-one collective funneling limitation and, moreover, our collective result will be available within each of the SPMD sub-tasks launching the collectives. This model is "MPI-like" and relies on a binary-tree algorithm that has a $\log_2(N)$ scaling on the number of SPMD elements. In addition to detailing our new collectives library, [PBKB17] describes an updated implementation of the statistics kernels based on this new capability. In the updated statistics kernels, control is passed down from the top-level task to the SPMD sub-tasks, which have access to the global statistical model once it has been properly Derived. The Assess phase can then happen inside these sub-tasks since the global model will be available locally to each of them. The report concludes with a performance study of the kernels deployed *in situ* with S3D serving as the full-scale application driver. S3D is a massively parallel DNS solver developed at Sandia National Laboratories, cf. [HSSC05], for which a SPMD-Legion implementation already exists and is routinely run on DOE's largest production platforms. The performance studies reveal optimal scaling, both weak and strong, across the considered example space and the S3D team plans to include the *in situ* workflow within upcoming science campaigns.

Now that the Sandia team has a functional *in situ* workflow that can be deployed at-scale, our work can be directly leveraged by the other sub-teams from the UDDAP project. Notably, 1) our collectives library can be used by these teams to support their analyses' collective communication requirements, 2) the variant of Legion-S3D that supports the *in situ* workflow can be extended to support their analysis kernels, and 3) the statistics implementations can serve as a template example for how an analysis kernel can interact with each of the aforementioned codes. For the remainder of the project, the primary focus of the Sandia team will be to work with the other sub-teams to explore the effects of increasingly complex *in situ* workflows comrpising different types of analyses.

References

[BTSA12] M Bauer, S Treichler, E Slaughter, and A Aiken. Legion: expressing locality and independence with logical regions. In *SC '12: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–11, 2012.

[FFL15] K. J. Franko, T. C. Fisher, and P. Lin. CFD for next generation hardware: Experiences with proxy applications. In *Proc. 22rd AIAA Computational Fluid Dynamics Conference*, Dallas, TX, U.S.A., June 2015.

[HSSC05] Evatt R Hawkes, Ramanan Sankaran, James C Sutherland, and Jacqueline H Chen. Direct numerical simulation of turbulent combustion: fundamental insights towards predictive models. *Journal of Physics: Conference Series*, 16(1):65, 2005.

[Kit10] Inc. Kitware. *The VTK User's Guide, version 5.4*. Kitware, Inc., 2010.

[PB15] P. P. Pébaÿ and J. Bennett. An asynchronous many-task implementation of *in situ* statistical analysis using Legion. Sandia Report SAND2015-10345, Sandia National Laboratories, November 2015.

[PB16] P. P. Pébaÿ and J. Bennett. Scalability of asynchronous many-task *In Situ* statistical analysis on experimental clusters: Interim report. Sandia Report SAND2016-1487, Sandia National Laboratories, February 2016.

[PBH+16] P. P. Pébaÿ, J. Bennett, D. Hollmann, S. Treichler, P. McCormick, C. Sweeney, H. Kolla, and A. Aiken. Towards asynchronous many-task *in situ* data analysis using Legion. In *Proc. 30th IEEE International Parallel & Distributed Processing Symposium, High Performance Data Analysis and Visualization Workshop*, Chicago, IL, U.S.A., May 2016.

[PBKB17] P. P. Pébaÿ, J. C. Bennett, H. Kolla, and G. Borghesi. Scalability of several asynchronous many-task models for *In Situ* statistical analysis. Sandia Report SAND2017-5220, Sandia National Laboratories, May 2017.

[PTBM11] P. P. Pébaÿ, D. Thompson, J. Bennett, and A. Mascarenhas. Design and performance of a scalable, parallel statistics toolkit. In *Proc. 25th IEEE International Parallel & Distributed Processing Symposium, 12th International Workshop on Parallel and Distributed Scientific and Engineering Computing*, Anchorage, AK, U.S.A., May 2011.

[PTKB16] P. P. Pébaÿ, T. B. Terriberry, H. Kolla, and J. Bennett. Numerically stable, scalable formulas for parallel and online computation of higher-order multivariate central moments with arbitrary weights. *Computational Statistics*, 31(4):1305–1325, 2016.

[SLT+17] E. Slaughter, W. Lee, S. Treichler, W. Zhang, M. Bauer, G. Shipman, P. McCormick, and A. Aiken. Control replication: Compiling implicit parallelism to efficient SPMD with logical regions. 2017. In Submission.