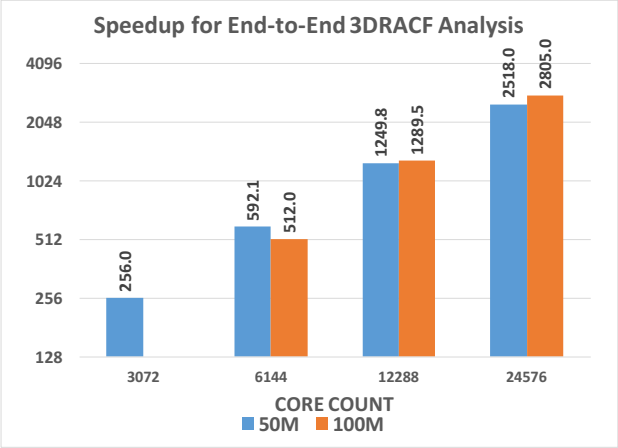
DAPPER Delivers First Ever At-scale Atom Probe Tomography (APT) Data Processing Capability



DAPPER enables first ever at-scale spatial correlation analysis

Achievement: Computation time of full 3D radial autocorrelation analysis of 100 million atoms APT data set reduced **from ~100 days on a single core to ~1 hour on 2048 cores (>2800X speedup).**

Significance and Impact: Reduces the turnaround time of an end-to-end APT correlation analysis on 100 million atoms by three orders of magnitude using 2048 MPI ranks on 1024 nodes (24 cores per node) of a Cray XC30. The software reported here equips material scientists for the first time with a high-speed scalable capability for efficient and timely analyses of massive APT data.

Research Details

- DAPPER (Data Analysis Parallel Package Maker) provides a collection of highly scalable core functionalities key to many kinds of analyses of multidimensional spatial data.
- PI: Sudip Seal (DCS Group, CSMD); Lead developer: Hao Lu (DCS Group, CSMD).
- Cutting-edge APT datasets have 100's of millions of atoms (10^2-10^3 times more than before). Existing APT analysis approaches unable to handle such massive data sets.

Facility: CNMS, NERSC

PI and affiliation: Sudip K. Seal, ORNL

Team: Sudip K. Seal, Hao Lu, Wei Guo and Jonathan Poplawsky

Publication(s):

- "*Spherical Region Queries on Multicore Architectures*", Lu, Seal, Guo and Poplawsky, Denver, USA, **IA^3, Supercomputing (2017).**
- "*Efficient, Parallel At-Scale Correlation Analysis for Atom Probe Tomography on Hybrid Architectures*", Lu, Seal, Guo and Poplawsky, **In Review (2017).**

Sponsor: LDRD (LOIS ID: 8449)

Publication Overview:

Spherical Region Queries on Multicore Architectures

In this short paper, we report the performance of multiple thread-parallel algorithms for spherical region queries on multicore architectures motivated by a challenging data analytics application in materials science. Performances of two tree-based algorithms and a naive algorithm are compared to identify the length scales at which these approaches perform optimally. The optimal algorithm is then used to scale the driving materials science application, which is shown to deliver over 17X speedup using 32 OpenMP threads on data sets containing many millions of atoms.

Efficient, Parallel At-Scale Correlation Analysis for Atom Probe Tomography on Hybrid Architectures

Atom probe tomography (APT) is a material probing technique that has undergone dramatic improvements in its capability to map individual atoms within a material sample resulting in data files with hundreds of millions of atoms. Understanding the nano-structural features hidden in these massive amounts of atomic data is a crucial analysis task for materials scientists. However, fast analysis capabilities for large APT workloads remains a critical bottleneck.

In this paper, we present the design, implementation and detailed performance evaluations of a parallel software capable of efficiently performing key but extremely time-consuming correlation analyses of massive high density APT data. We start by studying the performance of an existing shared memory software to motivate our design choices. Keeping realistic APT workloads in mind, we extend our shared memory implementation to hybrid architectures. We present detailed performance analyses, supported by empirical results on a Cray XC30 and a Cray XC40 architecture, of three different parallel implementations of our software. The usefulness of this software is demonstrated by reducing the turnaround time of an end-to-end APT correlation analysis on 100 million atoms by three orders of magnitude using 2048 MPI ranks on 1024 nodes (24 cores per node) of a Cray XC30. The software reported here equips material scientists for the first time with a high-speed scalable capability for efficient and timely analyses of massive APT data.