INNOVATIVE
COMPUTING LABORATORY

ECP

FFT

FIND OUT MORE AT

http://icl.utk.edu/fft/

FAST FOURIER TRANSFORM

Overview

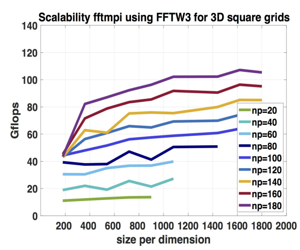
FFT is used in many domains—including molecular dynamics, spectrum estimation, fast convolution and correlation, signal modulation, wireless multimedia applications, etc. Distributed 3-D FFT is one of the most important kernels in molecular dynamics, and its performance can affect an application's scalability on larger machines. More than a dozen ECP applications use FFT in their codes.

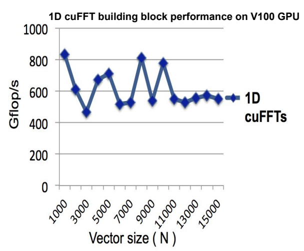
Objectives

The FFT-ECP project intends to provide a sustainable 2-D and 3-D FFT library for distributed-heterogeneous parallel systems as the one projected for the upcoming exascale computing systems. FFT-ECP will leverage established but ad hoc software tools that have traditionally been part of application codes, but not extracted as independent, supported libraries. These 3-D FFTs rely on third-party 1-D FFTs, either from FFTW or from vendor libraries.

The current state-of-the-art FFT libraries are not scalable on large heterogeneous machines with many nodes or even on one node with multiple high-performance GPUs (e.g., several NVIDIA V100 GPUs). Furthermore, these libraries require large FFTs in order to deliver acceptable performance on one GPU. Efforts to simply enhance classical and existing FFT packages with optimization tools and techniques—like autotuning and code generation —have so far not been able to provide the efficient, high-performance FFT library capable of harnessing the power of supercomputers with heterogeneous GPU-accelerated nodes. In particular, ECP applications that require FFT-based solvers might suffer from the lack of fast and scalable 3-D FFT routines for distributed heterogeneous parallel systems, which is the very type of system that will be used in upcoming exascale machines.
Thus, the main objective of the FFT-ECP project is to:

- Collect existing FFT capabilities recently made available from ECP application teams (LAMMPS/fftMPI and HACC/SWFFT);
- Assess gaps and make available as a sustainable math library;
- Explore opportunities to build 3-D FFT libraries on vendor 1-D and 2-D kernels, especially leveraging on-node concurrency from 2-D and batched 1-D formulations;
- Focus on capabilities for Exascale platforms;
- Emphasize leverage of vendor capabilities and addressing vendor deficiencies over creation of new and independent software stack.



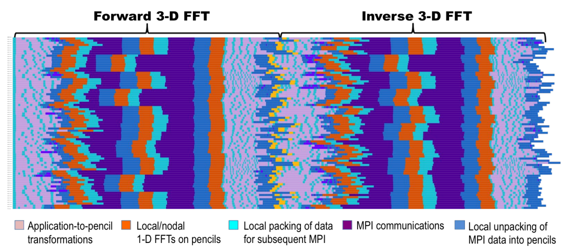


Scalability of FFTMPI (Left) on current CPU based systems. Codes have good weak and strong scalability, achieving up to 20 GFlop/s per node. GPU 1-D FFT building blocks can run up to about 800 GFlop/s on V100 GPUs (Right).

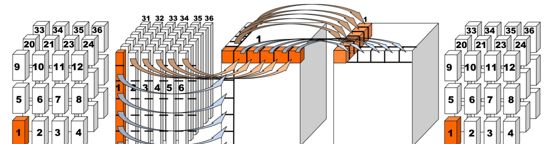
Key Challenges:

1. **Communication costs**: Today's machines have very complex memory hierarchies and thus data movement, data layout translation, and communication should be the main focus of any distributed FFT library that aims to improve the performance of any ECP application that relies on FFT.
2. **Application specifics**: ECP applications that require FFT-based solvers suffer from the lack of fast and scalable 3-D FFT routines for distributed-heterogeneous parallel systems as the ones projected for the upcoming exascale computing systems. Also, ECP applications may not be able to use existing FFT libraries without application-specific adjustments and tuning.
3. **Performance portability**: Finally, providing many application and hardware-specific versions along with their parameterizations and different optimization techniques will inevitably create a tuning challenge. The FFT-ECP software will be linked to our autotuning tools, which combined with our kernels designs and use of various state-of-the-art building blocks will provide performance portability, software interoperability, and sustainability.

A 3-D FFT EXECUTION TRACE WITH MAIN FFT COMPONENTS
(80 MPI processes on Intel Xeon E5-2650 v3 cluster, 1K × 1K × 1K grid)



FFT-ECP FRAMEWORK DESIGN WITH FLEXIBLE API
(Need flexible FFT API for application-specific input and output)



- Main objective is the design and implementation of a sustainable FFT library for exascale platforms;
- FFT-ECP goal is to help key ECP applications; to provide efficient and flexible FFT API to take applications-specific input and outputs;
- Approach is to: 1) Collect & leverage existing FFT capabilities to build sustainable FFT library (vs. the creation of new and independent software stack); 2) Optimize data movements and overlap computations with communications; 3) Autotuning.

Recent Progress: The FFT-ECP team completed an evaluation and design phase for FFTs targeting distributed accelerated systems. This included analysis of the current performance of FFT libraries and a design framework for the FFT-ECP project [1]. GPU-accelerated prototypes for FFTMPI and SWFFT are developed.

Next Steps:
1. Q1-Q2 FY19: Design and implementation phase.
2. Q3-Q4 FY19: Implementation optimization and features phase.

Deliverables:
FFT-ECP Milestone report ST-MS-10-1216
https://www.icl.utk.edu/publications/

Relation projects:
ECP LAMMPS, HACC, CoPA, Alpine, FFTX, SLATE, DTE, and MAGMA
(http://icl.cs.utk.edu/magma/)

Reference:

[1] S. Tomov, A. Haidar, D. Schultz, and J. Dongarra, "**Evaluation and Design of FFT for Distributed Accelerated Systems**," ECP WBS 2.3.3.09 Milestone Report, no. FFT-ECP ST-MS-10-1216: Innovative Computing Laboratory, University of Tennessee, October 2018.

ICL INNOVATIVE COMPUTING LABORATORY
THE UNIVERSITY OF TENNESSEE KNOXVILLE

SPONSORED BY

U.S. DEPARTMENT OF ENERGY Office of Science
ECP EXASCALE COMPUTING PROJECT
NNSA National Nuclear Security Administration

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-NA-0003525.