The fast Fourier transform (FFT) is used in many domain applications—including molecular dynamics, spectrum estimation, fast convolution and correlation, signal modulation, and wireless multimedia applications. For example, distributed 3-D FFT is one of the most important kernels used in molecular dynamics computations, and its performance can affect an application’s scalability on larger machines. Similarly, the performance of the first principle calculations depends strongly on the performance of the FFT solver. Specifically, for the US Department of Energy (DOE), we found that more than a dozen Exascale Computing Project (ECP) applications use FFT in their codes. To address these needs, ICL released the Highly Efficient FFTs for Exascale (heFFTe) library. The heFFTe v2.3 library release features very good weak and strong scalability, and performance that is close to 90% of the roofline peak.

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.

S. Cayrols, J. Li, G. Boslca, S. Tomov, A. Ayala, and J. Dongarra
Lossy all--to--all exchange for accelerating parallel 3--D FFTs on hybrid architectures with GPUs
2022 IEEE International Conference on Cluster Computing (CLUSTER), 2022.

A. Ayala, S. Tomov, M. Stoyanov, and J. Dongarra
Scalability Issues in FFT Computation
Parallel Computing Technologies (PaCT 2021), Lecture Notes in Computer Science, vol 12942.

A. Ayala, S. Tomov, P. Luszczek, S. Cayrols, G. Raghvani, and J. Dongarra
Interim Report on Benchmarking FFT Libraries on High Performance Systems

D. Sharp, M. Stoyanov, S. Tomov, and J. Dongarra
A More Portable HeFFTe: Implementing a Fallback Algorithm for Scalable Fourier Transforms
HEFFTe Acceleration on GPUs
Local CPU kernels presented on Section 2 are typical on state-of-the-art parallel FFT libraries, HEFFTe provide new GPU kernels for these tasks achieving over 40x speedup.

Scalability
In the following figures, we present scalability curves of CPU and GPU versions of HEFFTe. Note that our GPU version has very good scalability and achieves (around 20 gigaflops/s per node), which is over 2x what FFTMPI and SWFFT typically achieve for these sizes. We use 24 MPis/node on each case, 1MPI/core for HEFFTe CPU and 4MPI/Volta100 for HEFFTe GPU.