Mixed precision multi-level approach for FFTs

University of Stuttgart
Cluster of Excellence in Simulation Technology

Author Name

Institute for Parallel and Distributed Systems
Universitätsstraße 38, D-70569 Stuttgart, Germany

Motivation

- FFT is used for a spectral approach to solve PDEs
- Medical image registration [3] involves large 3D FFTs
- Transformation of data is a bottleneck for large scale FFTs
- Modern hardware has native support for half precision arithmetics
- High precision not always needed, e.g. for preconditioning

$$\hat{u} = \mathcal{F}u \overset{1D}{\Leftrightarrow} \hat{u}_k = \sum_{l=0}^{N-1} u_l e^{-\frac{2\pi i}{N}kl}$$

$$\Delta u = g \overset{1D}{\Leftrightarrow} -k^2\hat{u}_k = \hat{g}_k$$

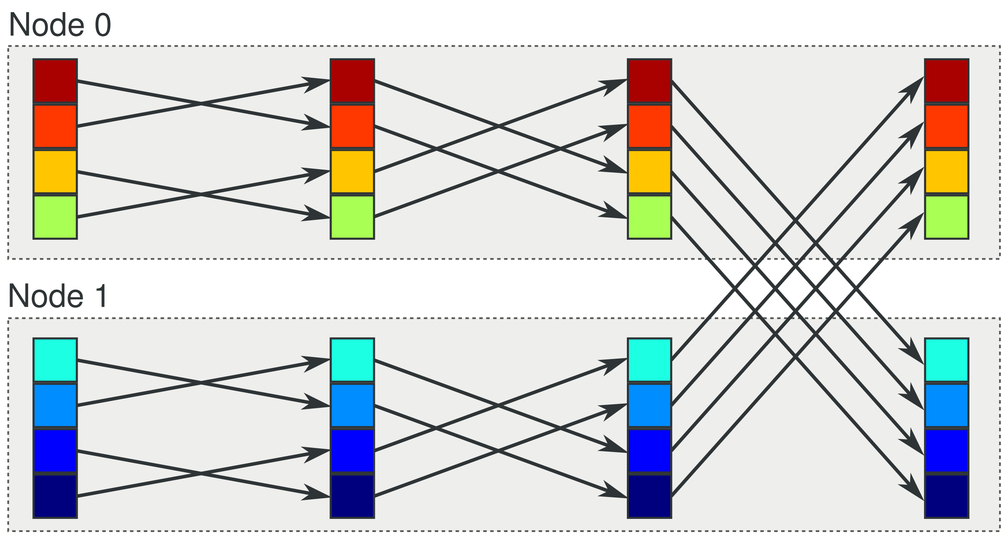

Figure 1: FFT with a divide and conquer scheme on two distributed nodes using the Cooley–Tukey algorithm.

- Lower accuracy reduces communication volume
- Higher computational throughput for lower accuracy

	NVIDIA Tesla P100	NVIDIA Tesla V100
half (FP16)	21.2 TFLOPS	31.3 TFLOPS
float (FP32)	10.6 TFLOPS	15.7 TFLOPS
double (FP64)	5.3 TFLOPS	7.8 TFLOPS

Table 1: Peak performance of NVIDIA Tesla accelerators [5]

Problems with low accuracy

Half precision (FP16) has very low dynamic range

	ϵ	min	max
half (FP16)	$2^{-10} \approx 10^{-3}$	$\approx 6.1 \times 10^{-5}$	65504
float (FP32)	$2^{-23} \approx 10^{-7}$	$\approx 1.2 \times 10^{-38}$	$\approx 3.4 \times 10^{38}$
double (FP64)	$2^{-52} \approx 10^{-16}$	$\approx 2.2 \times 10^{-308}$	$\approx 1.8 \times 10^{308}$

Table 2: Machine tolerance and range of floating-point values defined by IEEE754 [1].

Range issues

- FFT scales values by N (N^3 in 3D)
- Large FFTs map values outside of dynamic range

$$\mathcal{O}(1) \xrightarrow{\mathcal{F}} \mathcal{O}(N^3) \xrightarrow{N^{-3}} \mathcal{O}(1) \xrightarrow{\mathcal{F}^{-1}} \mathcal{O}(1)$$

$$\mathcal{O}(1) \xrightarrow{N^{-3/2}} \mathcal{O}(N^{-3/2}) \xrightarrow{\mathcal{F}} \mathcal{O}(N^{3/2}) \xrightarrow{N^{-3/2}} \mathcal{O}(1) \xrightarrow{\mathcal{F}^{-1}} \mathcal{O}(1)$$

$\Rightarrow$ pre- and post-scaling only allows for 512^3 values in half precision

Accuracy issues

- The RMS for the FFT is in $\mathcal{O}(\epsilon\sqrt{\log N})$ [2]
- Pure 3D FFTs in FP16 are not feasible

$$\mathcal{O}(1) \xrightarrow{\mathcal{F}} \mathcal{O}(\epsilon \log^{3/2} N) \xrightarrow{\mathcal{F}^{-1}} \mathcal{O}(\epsilon \log^{6/2} N)$$

Idea

- For smooth functions high modes decay fast
- Split data into a coarse grid for low modes in high accuracy and a fine grid in lower accuracy
- Use restriction $\mathcal{R}$ and prolongation $\mathcal{P}$, e.g. mean value of neighbors

$$u_c = \mathcal{R}_{f\to c}(u)$$
$$u_f = u - \mathcal{P}_{c\to f}(u_c)$$

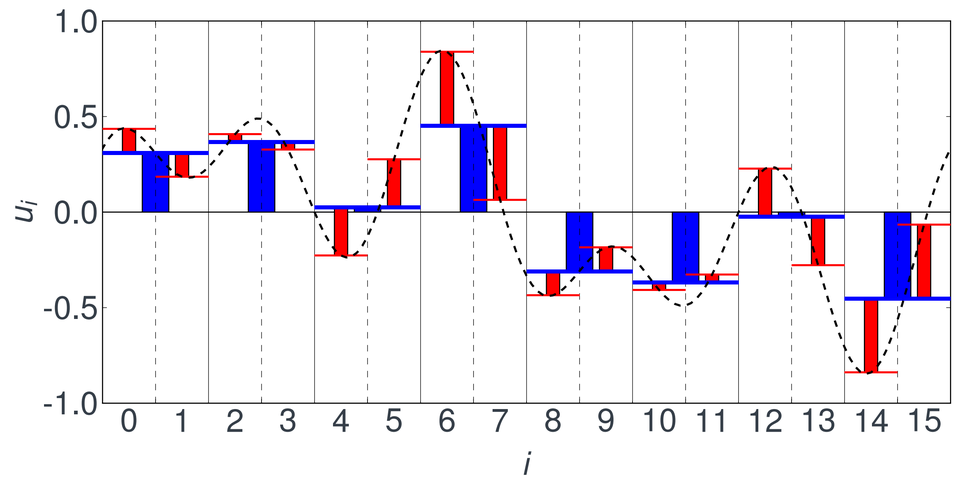

Figure 2: Exemplary function values with a sample rate of 16. The restricted values are shown in blue. The additive surpluses on the fine grid are red.

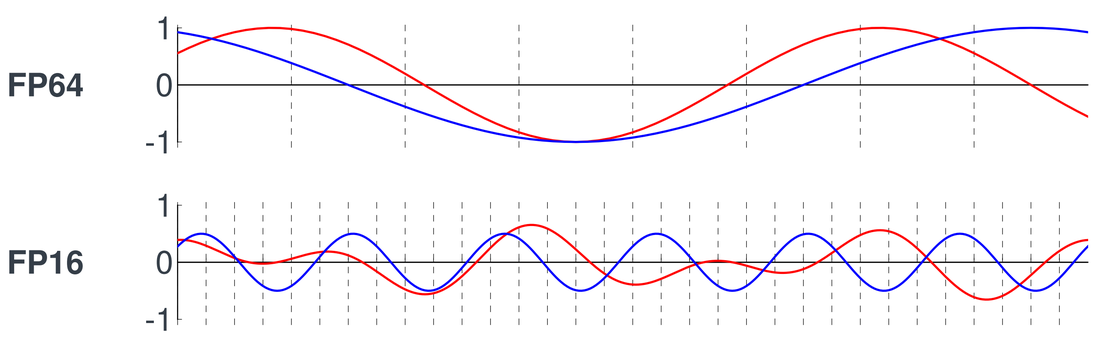

Figure 3: Low modes are represented on a coarse grid with high accuracy and high modes are represented on a fine grid with low accuracy.

Questions

- Effect of multi-level approach on derivatives
- Error analysis for mixed precision
- Performance for large scale FFTs

$$u = u_f^{FP16} + \mathcal{P}_{m\to f}(u_m^{FP32}) + \mathcal{P}_{c\to f}(u_c^{FP32})$$
$$\hat{u} \neq \hat{u}_f^{FP16} + \mathcal{P}_{m\to f}(\hat{u}_m^{FP32}) + \mathcal{P}_{c\to f}(\hat{u}_c^{FP32})$$
$$\Delta u \neq \Delta u_f^{FP16} + \mathcal{P}_{m\to f}(\Delta u_m^{FP32}) + \mathcal{P}_{c\to f}(\Delta u_c^{FP32})$$

References

[1] IEEE Standard for Floating-Point Arithmetic.
IEEE Std 754-2008, pages 1–70, Aug 2008.

[2] W Morven Gentleman and Gordon Sande.
Fast fourier transforms: for fun and profit.
In *Proceedings of the November 7-10, 1966, fall joint computer conference*, pages 563–578. ACM, 1966.

[3] Andreas Mang and George Biros.
A semi-lagrangian two-level preconditioned newton–krylov solver for constrained diffeomorphic image registration.
SIAM Journal on Scientific Computing, 39(6):B1064–B1101, 2017.

[4] NVIDIA Corporation.
CuFFT Library 9.0, September 2017.

[5] NVIDIA Corporation.
NVIDIA TESLA V100 Architecture, August 2017.