A Petascale numerical library for multiscale phenomena simulations

D. Pekurovsky, San Diego Supercomputer Center/UCSD (*dmitry@sdsc.edu*)
D. Donzis, Department of Aerospace Engineering, Texas A&M University (*donzis@tamu.edu*)

Introduction

Phenomena with a wide range of scales are a grand challenge for theory, simulations and experiments. Simulations of such phenomena is a subject of study in diverse fields of science today. Scaling them to large node counts presents considerable challenge, however it is often essential for achieving required resolution, accuracy, complexity. Examples include:

- Cosmology/astrophysics: cosmological structure formation simulations.
  Challenge: Increasing resolution while retaining the essential physics
- Chemistry: all-atom molecular dynamics simulations of enzymes.
  Challenge: simulating a large number of atoms and having sufficiently high throughput.
- Engineering: compressible turbulence, high-speed mixing and MHD turbulence.
  Challenge: resolving enough details at a high Reynolds number

Need for a numerical library for multiscale phenomena simulations

Multiscale phenomena simulations often share the same or similar algorithms (pseudospectral/FFT/Chebyshev, compact schemes etc). Yet as of today there are no software packages that implement them in a flexible, unified framework, while providing scalability to O(10^5) cores and beyond. Existing open-source packages (P3DFFT, 2Decomp, PFFT) have limitations of particular transform types and data structures.
In addition, existing tools typically rely on blocking MPI calls. This is getting increasingly inadequate as communication mechanism for fast-approaching age of Exascale, especially for a communication intensive operations required in multiscale simulations, such as all-to-all exchanges. Innovative technologies/protocols are needed to take full advantage of modern HPC platforms.

Prototype Software

P3DFFT [1] is an open source library package known in the community. It is targeted towards 3D Fourier and related transforms, implementing real-to-complex FFT, Chebyshev (in one dimension, combined with 2D FFT), as well as empty transform in one dimension. Employing 2D domain decomposition, P3DFFT achieves good scaling on up to O(10^5) cores, subject to hardware support. It is built on top of FFTW/ESSL optimized libraries for 1D FFT. P3DFFT provides interfaces for Fortran and C. The project has been originally funded by an STCI EAGER grant. It is maintained at http://code.google.com/p/p3dfft

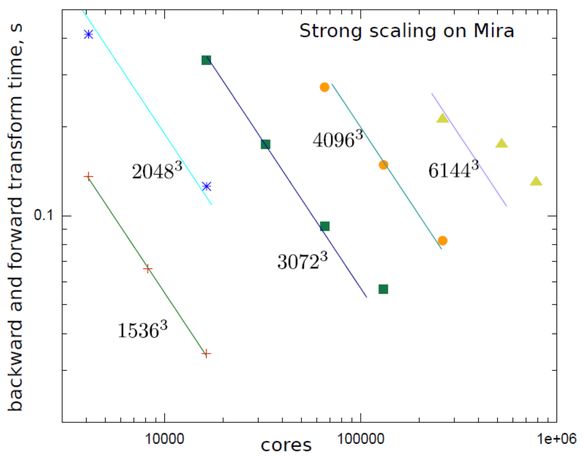

Figure 1. Scaling of P3DFFT on IBM BG/Q "Mira" at Argonne National Lab (A. Beresnyak).

P3DFFT infrastructure has been successfully applied in a number of codes. Below is a scaling plot of compressible turbulence code cDNS developed by co-PI Donzis.

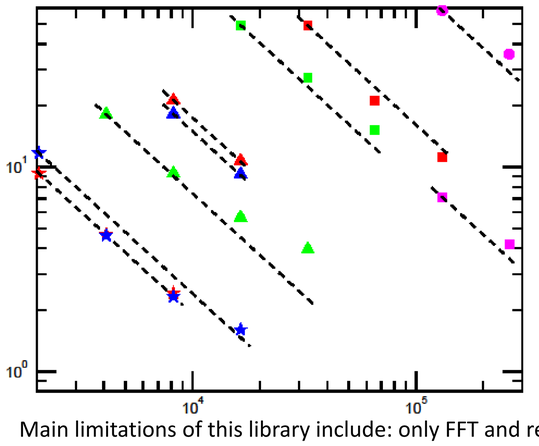

Figure 2. Scaling of cDNS, a DNS code for compressible turbulence that solves the fully compressible Navier-Stokes equations using tenth-order compact schemes in space in a 3D domain. Details of the code and performance using MPI and OpenMP have been described in [2]. Horizontal axis is the number of cores, vertical axis is execution time per step. The legend is as follows: Blue: BG/Q. Red: BG/P. Green: Kraken. Magenta: Titan. Lower to upper set of data points correspond to 512^3, 1024^3 and 2048^3.

Main limitations of this library include: only FFT and related transforms implemented; predefined data structure (X-pencil input, Z-pencil output); blocking MPI communication.

Proposed Solution

The library under development will attempt to fill the gap by providing software infrastructure that is performance oriented and scalable to high core counts, flexible in the interface and offering an extensive set of options:

- Compact finite-difference schemes as well as FFT/Chebyshev in 1D, 2D, 3D; other transforms?
- Arbitrary combination of transforms in all three dimensions
- Pruned input/output
- Multiple plans and multi-variable transforms within the same plan
- Flexible domain decomposition options
- User-friendly utilities: differentiation, halo exchanges, spectra, correlations
- Isolated transposes

Performance and Scalability

The library under development will provide overlap of communication with computation. A number of mechanisms are being explored for nonblocking all-to-all exchanges, including: MPI-3 (MPI_Ialltoall), MPI-2 (MPI_Put or MPI_Get), OpenSHMEM, CoArray Fortran. Communication will be implemented in a modular fashion, giving the user freedom to choose the most efficient mechanism depending on their platform of choice. In addition MPI/OpenMP framework will be implemented.

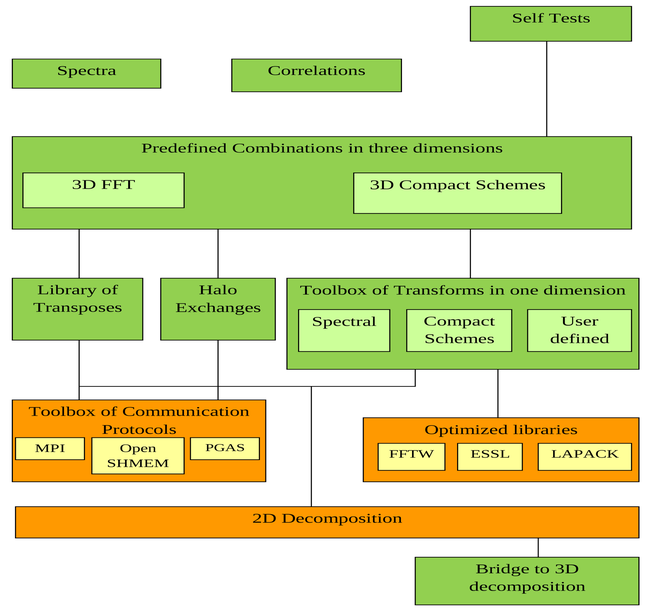

Figure 3. Structure of proposed software package. It will be built on top of well-optimized libraries such as FFTW, ESSL, LAPACK, and will be based on 2D domain decomposition. It will be built in a modular fashion providing flexibility to the user.

Modules

1. Spectral transforms
   a. FFT/Chebyshev in 1D, 2D, 3D
   b. Pruned input/output
2. Differentiation
   a. Spectral space
      i. Divergence, Laplacian, curl
      ii. Opportunities to overlap communication with computation on multiple vector components
   b. Physical space
      i. Explicit
      ii. Compact

$$f'_j = \sum_{k=-l}^{r} a_k f_{j+k} + O(h^n)$$
$$\sum_{k=-L}^{R} b_k f'_{j+k} = \sum_{k=-l}^{r} a_k f_{j+k} + O(h^n)$$

Includes the standard three-point stencil 6th order and five-point stencil 10th order compact schemes

3. Communication
   a. Isolated transposes
   b. Halo exchanges

Sustainability and Portability

Package will be available as open source through a commonly used interface such as SVN as well. Project website will be maintained including a wiki, an issue tracker, and mailing list. Code development will emphasize portability by employing modular components with underlying choice of mechanisms. We welcome participation of the growing user community in further code development and testing as one long-term sustainability strategy.

User engagement

In 2014 P3DFFT has been **downloaded** approximately 2000 times and **cited** 67 times. Through the regularly updated **project wiki** and the user mailing list the users keep up to date with the developments. Users can contribute to project by suggesting or directly implementing other transform types or library features, as well as providing scaling and performance feedback, using **self-testing and timing utilities** provided within the suite. Talks at HPC and domain science conferences (such as APS Fluids 2014) and informal contacts with colleagues are also used as a means of outreach. New features will be explored and tested with the help of 5 collaborator groups we have identified, across several scientific fields.
P3DFFT has been designated as an **XSEDE community software** and as such has been installed, tested and documented across XSEDE sites. It will also be integrated into a turbulence XSEDE Science Gateway under development. The budget includes partial support for one graduate (at Texas A&M) and one undergraduate (at UCSD) students who are gaining valuable experience in the HPC field and open source software.

Progress to date and ongoing work

1. Multivariable version of P3DFFT has been developed. Significant speedups (up to x2) are observed when transforming several variables (such as vector components) at once.
2. P3DFFT has been ported to a number of new architectures including IBM BG/Q. C/Fortran interface has been restored.
3. Work on redesigning the application interface to include many additional types of transforms and data layout options is under way.
4. A trial version employing overlapping communication with computation has been developed. Its performance is being studied on a number of architectures.
5. A new post-processing framework for the turbulence simulation codes is being developed.

Conclusion

A powerful and flexible numerical library and a suite of programs is being developed to facilitate simulations of multiscale phenomena at large scale. The demand for such software is great, in multiple disciplines. In many cases significant breakthroughs are expected through research enabled by this tool.

References

[1] D. Pekurovsky, *P3DFFT: a framework for parallel computations of Fourier transforms in three dimensions*, SIAM Journal on Scientific Computing 2012, Vol. 34, No. 4, pp. C192-C209.
[2] Jagannathan, S., Donzis, D.A. 2012 *Massively parallel direct numerical simulations of forced compressible turbulence: a hybrid MPI/OpenMP approach*. XSEDE'12 conference, Chicago IL, July 16-20, 2012.
[3] Subramoni, H., Kandalla, K., Jose, J., Tomko, K., Schulz, K., Pekurovsky, D., & Panda, D. K. (2014). *Designing Topology-Aware Communication Schedules for Alltoall Operations in Large InfiniBand Clusters*. 43rd International Conference on Parallel Processing (ICPP14).
[4] Donzis D.A., Aditya K., Yeung P.K. & Sreenivasan K.R. (2014). *The turbulent Schmidt number*. Journal of Fluid Engineering. 136 060912.
[5] Donzis D.A., Gibbon J.D. Gupta A., Kerr R.M., Pandit R. & Vincenzi D. (2013). *Vorticity moments in four numerical simulations of the 3D Navier-Stokes equations*. Journal of Fluid Mechanics. 732 316.