Scalable Fast Multipole Methods for Vortex Element Methods

Qi Hu *, Nail A. Gumerov †, Rio Yokota §, Lorena Barba ¶, Ramani Duraiswami‡
* † ‡University of Maryland Institute for Advanced Computer Studies (UMIACS)
* ‡Department of Computer Science, University of Maryland, College Park
† ‡Fantalgo LLC, Elkridge, MD
§ King Abdullah University of Science and Technology
¶ Mechanical Engineering Department, Boston University
*huqi,†gumerov,‡ramani@umiacs.umd.edu,§ Rio.Yokota@kaust.edu.sa,¶ labarba@bu.edu

We use a particle-based method to simulate incompressible flows, where the Fast Multipole Method (FMM) is used to accelerate the calculation of particle interactions. The most time-consuming kernels—the Biot-Savart equation and stretching term of the vorticity equation—are mathematically reformulated so that *only two* Laplace scalar potentials are used *instead of six*, while automatically ensuring divergence-free far-field computation. Based on this formulation, and on our previous work for a scalar heterogeneous FMM algorithm, we develop a new FMM-based vortex method capable of simulating general flows including turbulence on heterogeneous architectures. Our work for this poster focuses on the computation perspective and our implementation can perform one time step of the velocity+stretching for one billion particles on 32 nodes in 55.9 seconds, which yields 49.12 Tflop/s.

I. Problem Formulation

According to Helmholtz's laws the vortex elements move with the local velocity of the fluid. The velocity at locations $\mathbf{y}_j, j = 1, \ldots, M$ induced by vortex elements at locations $\mathbf{x}_i, i = 1, \ldots, N$ can be computed as

$$\mathbf{v}_j = \mathbf{v}(\mathbf{y}_j) = \sum_{i=1}^{N} \frac{\boldsymbol{\omega}_i \times (\mathbf{y}_j - \mathbf{x}_i)}{|\mathbf{y}_j - \mathbf{x}_i|^3} = \nabla \times \frac{\boldsymbol{\omega}_i}{|\mathbf{y}_j - \mathbf{x}_i|}. \quad (1)$$

This equation is derived from the definition of vorticity $\boldsymbol{\omega} = \nabla \times \mathbf{u}$ and the incompressibility condition $\nabla \cdot \mathbf{u}$, so calculating the velocity in this manner ensures the conservation of mass. For viscous flows, this singular kernel is smoothed by a Gaussian basis function for the vorticity field

$$\mathbf{v}_j = \sum_{i=1}^{N} \frac{\boldsymbol{\omega}_i \times (\mathbf{y}_j - \mathbf{x}_i)}{|\mathbf{y}_j - \mathbf{x}_i|^3} K_{\sigma_i}, \quad (2)$$

where the *cutoff function* $K(\mathbf{y}_j, \mathbf{x}_i)$ is defined by

$$K_\sigma(\mathbf{y}_j, \mathbf{x}_i) = \mathrm{erf}\left(\sqrt{\frac{r_{ij}^2}{2\sigma_i^2}}\right) - \sqrt{\frac{4}{\pi}} \sqrt{\frac{r_{ij}^2}{2\sigma_i^2}} \exp\left(-\frac{r_{ij}^2}{2\sigma_i^2}\right), \quad (3)$$

where $r_{ij} = |\mathbf{y}_j - \mathbf{x}_i|$.

Conservation of momentum is achieved by updating the properties of the vortex elements (position, strength, core size) according to the vorticity equation, obtained by taking the curl of the Navier-Stokes equation. This equation consists of the convection term, stretching term, and diffusion term. As mentioned earlier, convection is handled by moving the vortex elements, and diffusion is calculated by spreading the size of the Gaussian basis function at a rate that satisfies the analytical solution of the diffusion equation. As for the stretching term we insert Eq. 2 into

$$\frac{d\boldsymbol{\omega}_j}{dt} = \boldsymbol{\omega}_j \cdot \nabla \mathbf{v}_j, \quad (4)$$

which results in another N-body interaction. Note that this involves the contraction of the vorticity pseudovector with the velocity gradient tensor, which is computationally expensive.

II. Major Contributions

Direct evaluation of Eq. 1 or Eq. 4 on all $\mathbf{y}_j$ yields $O(N^2)$ cost, which can not scale to large size simulations in practice. However, since the Biot-Savart kernel is composed of dipole solutions of Laplace equation, we can use Fast Multipole Method (FMM) to approximate these sums to any precision ϵ at $O(N + M)$ cost [1], [2].

Comparing with the Laplace potential kernel $1/|\mathbf{y}_j - \mathbf{x}_i|$, the computation of Eq. 1 and Eq. 4 require much more operations. In Ref. [3], it is shown that the cost of the "velocity+stretching" kernels is about 6 times that of the "potential +force" kernels, while in [2] the "velocity" cost is shown to be around 2.5 times that of the "potential" calculation. In this work, our purpose is to develop an efficient FMM algorithm for "velocity+stretching" computation on heterogeneous clusters, using the efficient new formulation.

Firstly, given the incompressibility constraint

$$\nabla \cdot \mathbf{v} = 0, \quad (5)$$

we provide an efficient FMM translation method to calculate "velocity+stretching" at a cost of only two Laplace potential kernels by using Lamb-Helmholtz decomposition [4], [5]. This substantially speeds-up the overall vortex element method since this is the most expensive part among all FMM computations (see [6]).

Secondly, while the previous work [6] has demonstrated the efficiency of a heterogeneous FMM algorithm, this work displayed a large communication cost. We develop new data structures for the distributed algorithm which separate the computation and communication to avoid synchronization during GPU computations. The new data structures [7] build on the *local essential tree* (LET) [8], [9] concept but use a master-slave model and further have a novel parallel construction algorithm, in which the granularity is at the level of the spatial boxes (which allows finer parallelization than at the single-node level). Basically, each node divides its assigned domain into small spatial boxes via octrees and classifies each box into one of five categories in parallel. Based on the box type, each node determines the boxes that need to import and export data. This can be computed on the GPU at negligible cost and this algorithm can handle non-uniform distributions with irregular partition shapes and improve the load balancing. Together with our fast FMM octree data structures and interaction lists' constructions, we are able to reduce all data structure-related overhead substantially.

Thirdly, we analyze the local summations for the vortex method (Biot-Savart kernel and the vortex core stretching), which is the other part that dominates the cost of the FMM, to further improve the overall algorithm timing and accuracy. There are several transcendental functions involved in the sum of the cutoff functions (Eq. 3) which are expensive to compute. We develop an accurate and inexpensive approach to approximate them and demonstrate a large performance gain, which can be used in many other applications.

REFERENCES

[1] L. Greengard and V. Rokhlin, "A fast algorithm for particle simulations," *J. Comput. Phys.*, vol. 73, pp. 325–348, December 1987.

[2] Q. Hu, M. Syal, N. A. Gumerov, R. Duraiswami, and J. G. Leishman, "Toward improved aeromechanics simulations using recent advancements in scientific computing," in *Proceedings 67th Annual Forum of the American Helicopter Society*, May 3–5 2011.

[3] R. Yokota, T. Narumi, R. Sakamaki, S. Kameoka, S. Obi, and K. Yasuoka, "Fast multipole methods on a cluster of GPUs for the meshless simulation of turbulence," *Computer Physics Communications*, vol. 180, no. 11, pp. 2066–2078, 2009.

[4] N. A. Gumerov and R. Duraiswami, "Efficient FMM accelerated vortex methods in three dimensions via the Lamb-Helmholtz decompositions," *ArXiv e-prints*, January 2012.

[5] H. Lamb, *Hydrodynamics*. Cambridge, UK: Cambridge University Press, 1932.

[6] Q. Hu, N. A. Gumerov, and R. Duraiswami, "Scalable fast multipole methods on distributed heterogeneous architectures," in *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '11. New York, NY, USA: ACM, 2011, pp. 36:1–36:12.

[7] Q. Hu, N. A. Gumerov, and R. Duraiswami, "Scalable distributed fast multipole methods," in *Proceedings of the 14th International Conference on High Performance Computing and Communications (HPCC-2012)*, ser. HPCC '12. ACM, 2012, to appear.

[8] I. Lashuk, A. Chandramowlishwaran, H. Langston, T.-A. Nguyen, R. Sampath, A. Shringarpure, R. Vuduc, L. Ying, D. Zorin, and G. Biros, "A massively parallel adaptive fast-multipole method on heterogeneous architectures," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, ser. SC '09. New York, NY, USA: ACM, 2009, pp. 58:1–58:12.

[9] L. Ying, G. Biros, D. Zorin, and H. Langston, "A new parallel kernel-independent fast multipole method," in *Proceedings of the 2003 ACM/IEEE conference on Supercomputing*, ser. SC '03. New York, NY, USA: ACM, 2003, pp. 14–30.