DGX-2 Based Optimization of the Application for Turbulent Combustion

Wen Minhua[1], Wang Shenpeng[1], Lin Linyin[2], Zhang Bin[2], Lin Xinhua[1+]
1. Center for High Performance Computing, Shanghai Jiao Tong University, Shanghai 200240, China
2. School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China
Nvidia Corporation

Background

Numerical simulation of turbulent combustion is a key tool for aeroengine design. Due to the need of high-precision model to Navier-Stokes Equation, numerical simulation of turbulent combustion requires huge amount of calculations, and the physicochemical models causes the flow field to be extremely complicated, making the load balancing a bottleneck for large-scale parallelization. We firstly designed this efficient DGX-2 based turbulent combustion numerical simulation system.

Method

The Navier-Stokes equation is a mathematical description of the conservation of mass, momentum, and energy under the assumption of a macroscopic continuous medium. Its general form is:

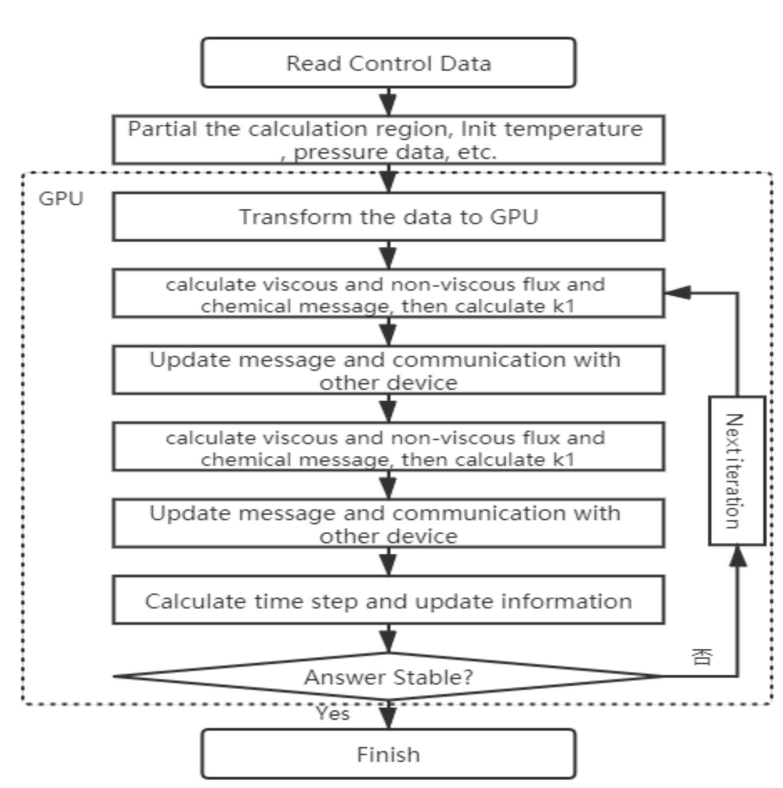
$$\frac{d\boldsymbol{u}}{dt} = \boldsymbol{f}(\boldsymbol{u}) + \boldsymbol{g}(\boldsymbol{u})$$

We use Semi-Implicit Runge-Kutta method on the numerical simulation of Navier-Stokes Equation, the final form we used is:

$$[\mathbf{I} - ha_1\mathbf{J}(\boldsymbol{u}^n)]\boldsymbol{k}_1 = h\{\boldsymbol{f}(\boldsymbol{u}^n) + \boldsymbol{g}(\boldsymbol{u}^n)\}$$
$$[\mathbf{I} - ha_2\mathbf{J}(\boldsymbol{u}^n)]\boldsymbol{k}_2 =$$
$$h\{\boldsymbol{f}(\boldsymbol{u}^n + b_{21}\boldsymbol{k}_1) + \boldsymbol{g}(\boldsymbol{u}^n + c_{21}\boldsymbol{k}_1)\}$$
$$\boldsymbol{u}^{n+1} = \boldsymbol{u}^n + \omega_1\boldsymbol{k}_1 + \omega_2\boldsymbol{k}_2$$

Pipeline

We use CPU only control the process of main process, before the code enter the iteration stage, all data will be transformed to the device memory. Then the whole iteration calculation will be only calculated in GPU, which avoids inefficient data transformation between device and host.



Contact: wangshenpeng@sjtu.edu.cn, wenminhua@sjtu.edu.cn

Method & Result

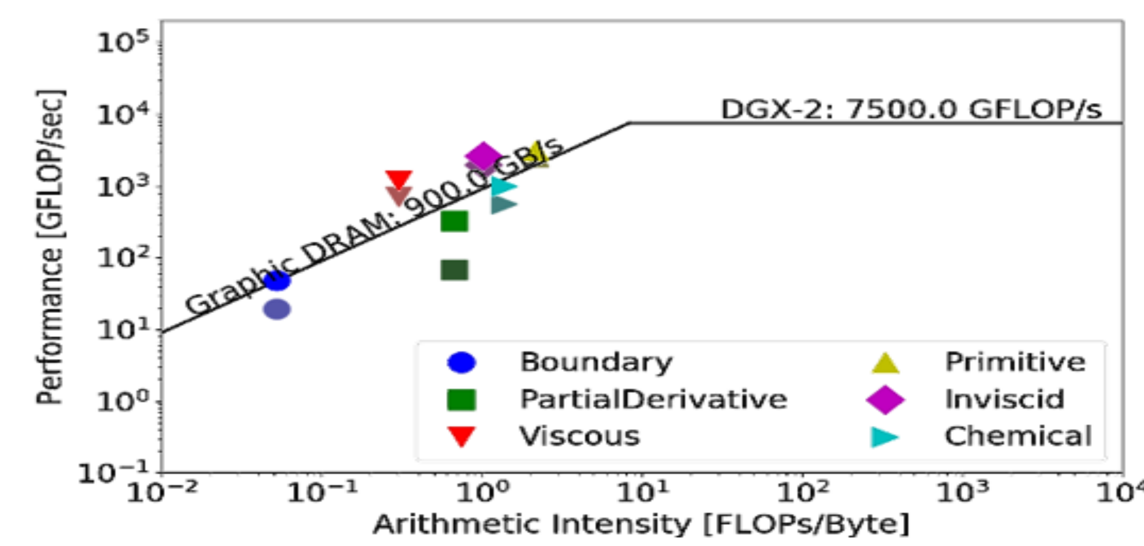


Fig.1 Roofline model on the original and optimized code on hot kernels

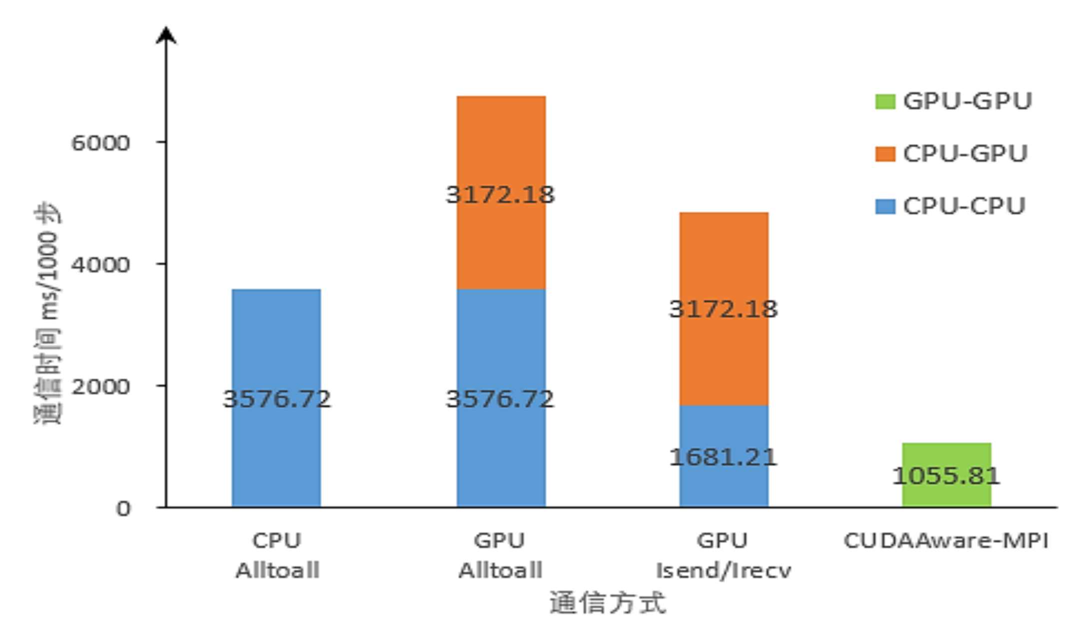


Fig.2 Communication Optimization

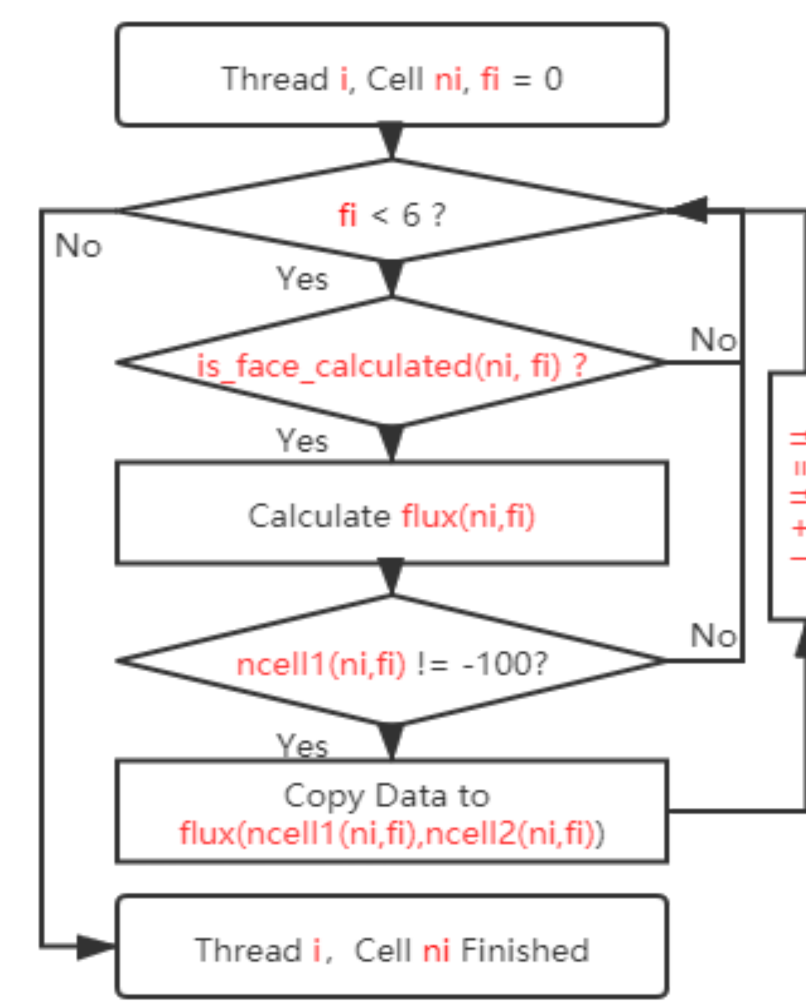


Fig.3 Flux Calculation to avoid data dependency

CPU parallelization has its limit due to the increasement of communication. By using DGX-2, only 8 V100 can exceed it.

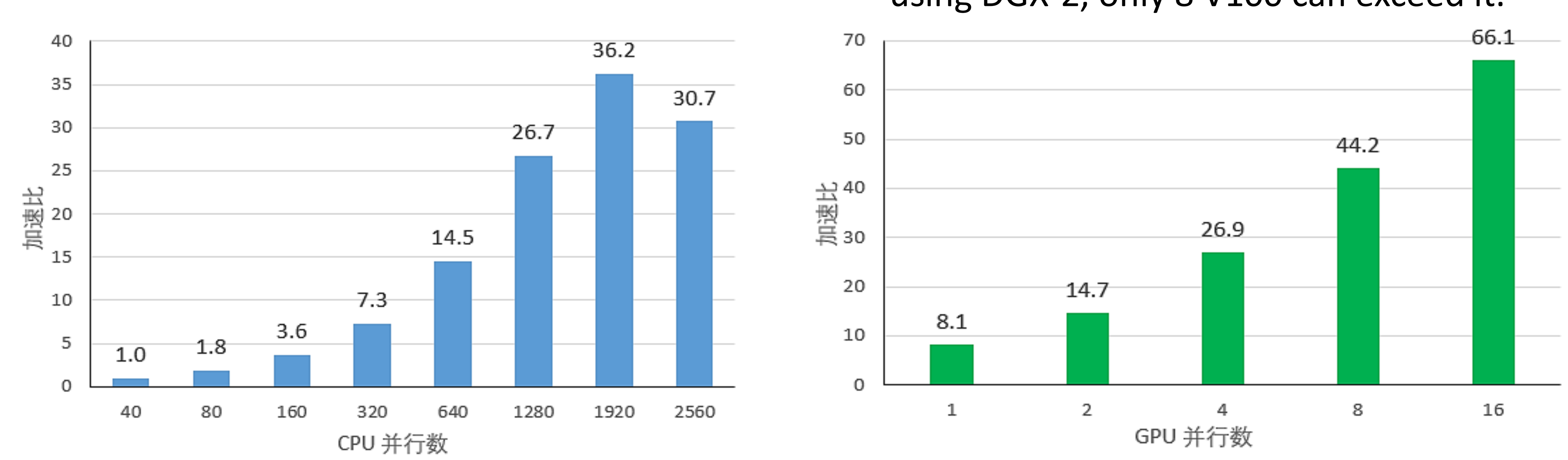


Fig.4 CPU and GPU Parallelization Analysis

Conclusion

We ported and optimized the numerical simulation of turbulent combustion to GPU. The performance of a single V100 GPU is 8.1x higher than that on dual-socket Intel Xeon 6248 CPU node with 40 cores. And the multi-GPU version on DGX-2 with 16 V100 GPUs achieves 66.1x speedup, which is higher than the best performance on CPU cluster.