Research Report: Measurement-based Adaptive Task Partitioning

Ting Li (SN:2017124076)

June 7, 2019

1 Difficulties in Balanced Division of Scientific Computing Tasks for Heterogeneous Platforms

1. GPU is sensitive to the scale of the problem, and its performance efficiency varies greatly with the size of the problem. (Higher efficiency on large-scale issues)

2. GPU performance is constrained by external factors with wide range of performance fluctuations

3. In heterogeneous collaborative computing, the processing power of each CPU core varies greatly.

2 Method Proposed

In the paper, the author proposed a measurement-based adaptive task partitioning method. The flow chart is shown bellow:

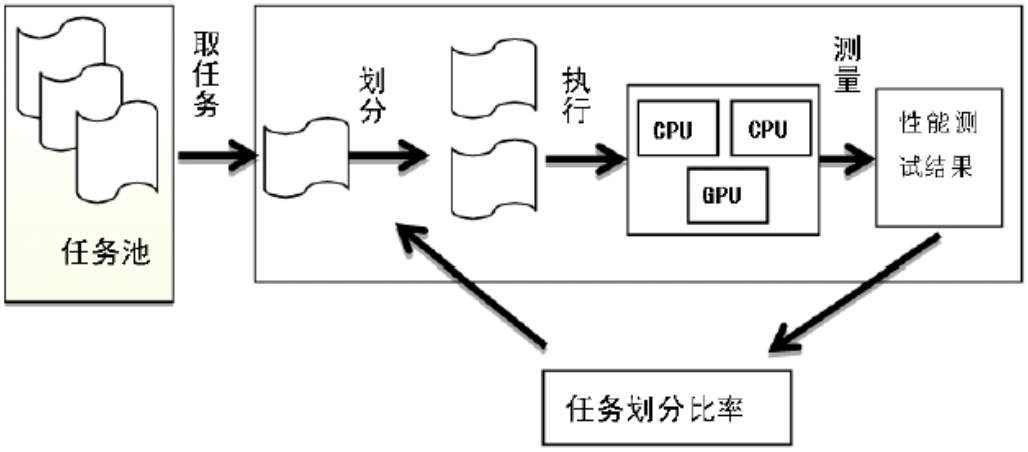

Figure 1: Flow chart of measurement-based adaptive task partitioning

The main steps are:

1. Initialize 'Task Partition Ratio'.

2. Cyclically get tasks from the tasks pool and divide them into GPU and CPU part according to the current 'Task Partition Ratio' each time.

3. After partitioning, the two parts are executed on GPU and CPU respectively.

4. Perform performance testing after execution, and update the 'Task Partition Ratio' by combining the results with the task load.

5. Repeat 2 to 4 steps until all tasks are completed.

3 Thoughts and Next Step

The detailed steps and formulas are not shown here, and there are very specific and detail descriptions in the paper, as well as a practical experiment. The author use this model to deal with two problems: DGEMM and LINPACK.

The next step, I will try to follow the author's steps and realize this model to seal with some examples like General Matrix Multiplication.

4 References

[1]王锋. 面向千万亿次CPU-GPU异构系统的编程模型与性能优化关键技术研究[D].国防科学技术大学,2013.