高性能计算程序设计基础 秋季 2020

提交格式说明

按照实验报告模板填写报告，需要提供源代码及代码描述至 https://easyhpc.net/course/121。实验报告模板使用 PDF 格式，命名方式为高性能计算程序设计_学号_姓名。如果有问题，请发邮件至 lidsh25@mail2.sysu.edu.cn，leong36@mail2.sysu.edu.cn 询问细节。

任务 1：

通过 CUDA 实现通用矩阵乘法（Lab1）的并行版本，CUDA Thread Block size 从 32 增加至 512，矩阵规模从 512 增加至 8192。

通用矩阵乘法（GEMM）通常定义为：

$$\mathrm{C} = \mathrm{AB}$$

$$C_{\mathrm{m,n}} = \sum_{n=1}^{N} A_{m,n} B_{n,k}$$

输入：M , N, K 三个整数（512 ~8192）

问题描述：随机生成 M*N 和 N*K 的两个矩阵 A,B,对这两个矩阵做乘法得到矩阵 C。

输出：A,B,C 三个矩阵以及矩阵计算的时间

任务 2：

将任务 1 改造成基于（OpenMP 或 MPI） + CUDA 的多层次并行矩阵乘法。矩阵被主进程切分成子矩阵分配给 OpenMP 或 MPI 并行线（进）程计算，并行进程调用任务 1 的 CUDA 版本矩阵乘法计算子矩阵，汇总并行进程的计算结果，并打印结果和运行时间，并行线程数：1，2，4，8。

任务 3：

通过 NVDIA 的矩阵计算函数库 CUBLAS 计算矩阵相乘，矩阵规模从 512 增加至 8192，并与任务 1 和任务 2 的矩阵乘法进行性能比较和分析，如果性能不如 CUBLAS，思考并文字描述可能的改进方法（参考《计算机体系结构-量化研究方法》第四章）。

CUBLAS 参考资料《CUBLAS_Library.pdf》，CUBLAS 矩阵乘法参考第 70 页内容。

CUBLAS 矩阵乘法例子，参考附件《matrixMulCUBLAS》