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：

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+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并行线程计算，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》