

GPU程序设计 2020-2021学年期末考试

整理: lzh

1. (20分)填空选择, 小题记不住, 但主要是考GPU计算的基本概念和CUDA网格模型
2. (10分)简述共享内存的模型, 共享内存冲突的解决办法和共享内存的两种使用方法
3. (10分)列出GPU中所有的内存种类及其特性
4. (15分)叙述统一内存(Unified Memory)的概念及其优缺点
5. (10分)结构体数组和数组结构体分别是什么, 它们在CUDA的内存访问中有什么不同
6. (15分)讨论不同情况下的同步及其实现方法
7. (20分)阅读一段GPU规约的代码, 回答如下问题
 - (1) (5分)for循环内为什么要 `__syncthreads()`
 - (2) (5分)13行-21行if语句的作用
 - (3) (5分)13行-21行if语句中为什么不需要 `__syncthreads()`
 - (4) (5分)代码中有哪些优化

```
1  __global__ void reduceInterleaved (volatile int *g_idata, int *g_odata, unsigned int n) {
2      unsigned int tid = threadIdx.x;
3      unsigned int idx = blockIdx.x * blockDim.x + threadIdx.x;
4      int *idata = g_idata + blockIdx.x * blockDim.x;
5
6      for (int stride = blockDim.x / 2; stride > 0; stride >>= 1) {
7          if (tid < stride) {
8              idata[tid] += idata[tid + stride];
9          }
10         __syncthreads();
11     }
12
13     if (tid < 32)
14     {
15         idata[tid] += idata[tid + 32];
16         idata[tid] += idata[tid + 16];
17         idata[tid] += idata[tid + 8];
18         idata[tid] += idata[tid + 4];
19         idata[tid] += idata[tid + 2];
20         idata[tid] += idata[tid + 1];
21     }
22
23     if (tid == 0) g_odata[blockIdx.x] = idata[0];
24 }
25 // 记得大概长这样, 细节有出入但是影响不大
```

注: 数院好不容易第一年开了GPU编程的课, 欢迎对高性能计算感兴趣的同学们选修

