

并行复习7 CUDA编程

CUDA编程

CUDA编程框架、线程结构、内存结构、优化方式（掩盖访存开销、利用 shared memory）

编程框架：host端 device端

两种优化方式

CUDA

- GPU：图形处理单元，CPU的协处理器
- CUDA：统一计算设备架构，建立在GPU基础之上的通用计算开发平台

CUDA编程框架

host主机端

指代CPU端的代码

device设备端

指代GPU端的代码

kernel

从host调用，在device端运行的函数

CUDA线程结构

线程

并行执行

线程块Block

GPU上的一组并行线程的集合

线程格Grid

线程块的集合

CUDA内存层次

- Device代码可以
 - 读/写每线程独占的寄存器
 - 读/写每kernel共享的全局内存
- Host代码可以
 - 在CPU主存和Device全局内存间传输数据

优化方式

掩盖访存延迟

SM的优化