고급소프트웨어실습I 11주차 숙제

20141494

강동욱

실습 1-1

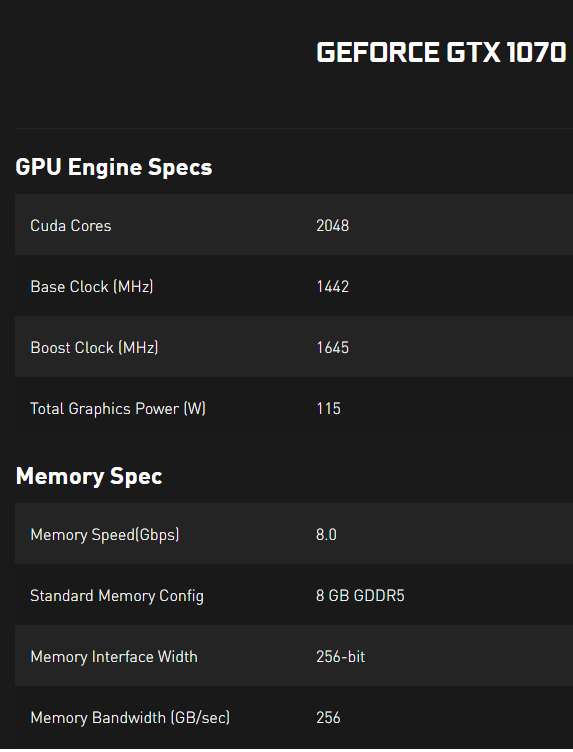
1. 자신이 사용하는 컴퓨터에 장착된 GPU의 기종

GEFORCE GTX 1070

1. 현재 설치된 CUDA 시스템의 Compute Capability

6.1

3. 현재 CUDA Compute Capability가 제공하는 각종 성능 및 스펙



실습 문제 1-2

(iii) 다음 이 CUDA 프로그램에 대하여 블럭 크기를 다양하게 변화 시켜가면서 시간을 측정한 후, 그 결과를 보고서에 테이블로 요약하라. 참고로 블럭의 크기는 와프의 크기인 32의 배수로 하고, 한 블럭이 가질 수 있는 크기에 어떠한 제한이 있는지 파악하기 위하여 실습 문제 1-1을 통하여 얻은 정보를 활용하라.

GTX 1070의 경우 Compute Capability이 6.1이므로 2.0이상의 Compute Capability에서 쓰레드 블럭 사이즈는 1024까지 가능하다.

1. 블럭 크기에 따른 수행 시간 변화가 있는지 확인하고 자신이 발견한

사항을 보고서에 기술하라.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Block size | 64 | 256 | 576 | 1024 |
| Compute time(ms) | 17.74 | 17.51 | 17.41 | 17.90 |

combine\_two\_arrays에서는 블록 사이즈에 대한 시간차이가 거의 없었다.

현재 커널 프로그램이 수행해주는 계산을 좀 더 (의미가 있고) 복잡하하여 한 쓰레드의 계산 시간을 길게 한 후, 블럭 크기에 따른 수행시간 변화를 분석하라.

일반적인 산술적계산으로는 계산시간의 차이를 보일 수 없었다.

실습1-3과 숙제1,2 에서의 수행시간을 보면 알 수 있드시 일반적으로(블록의 개수가 최대에 도달하기 전까지) 블록의 크기가 작을수록 수행시간이 짧다.

실습 문제 1-3

(ii) 다음 이 문제를 해결해주는 CUDA 커널 프로그램을 작성한 후 가급적 정확하게 GPU 수행 시간을 측정하라. 앞의 문제에서와 같이 쓰레드 블록의 크기를 변화시켜가면서 수행 시간 관점에서 CPU 방법과 비교 분석한 후, 그 결과를 보고서 기술하라.

N=1048575 대하여 CPU time은 평균 4172.6ms 이고 GPU time은 아래와 같다.

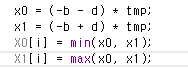
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Block size | 16 | 24 | 32 | 48 | 64 |
| Compute time(ms) | 121.8 | 183.0 | 270.3 | 365 | 395.08 |

하드웨어가 제공하는 쓰레드의 개수의 한계까지는 블록사이즈가 작을수록 더 많은 쓰레드 블록으로 태스크가 분할되어 더 빠른 계산이 이루어진다. CPU에서의 계산은 분산처리가 되지않아 블럭쓰레드가 프로세서의 개수 인것과 같은 계산이 이루어지기 때문에 이론적으로

GPU의 분산처리의 속도의 (프로세서수\*CPU에서의 계산속도)/(블록쓰레드의수 \*GPU에서의 계산속도) 배의 속도를 가진다. 실제로는 CPU에서의 계산속도가 훨씬 빠르기 때문에 수천배의 차이는 나지 않았다.

숙제 1

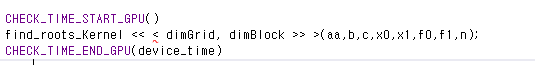
1. 첫 번째 두 배열 X0와 X1에는 자신이 구한 두 개의 실근 x0과 x1이 저장되어 있는데 반드시 x0 ≤ x1 조건을 만족시키도록 저장이 되어야 한다. 다음 두번째 두 배열 FX0와 FX1에는 각각 x0과 x1을 대응하는 이차 방정식에 대입하여 계산한 함수 값을 저장해주어야 한다. 물론 이론적으로는 모두 0 값이 계산되어야 하지만 수치 계산 시 발생하는 계산 오차로 인하여 정확히 0인 아닌 값이 나올 수 있음을 상기하라.



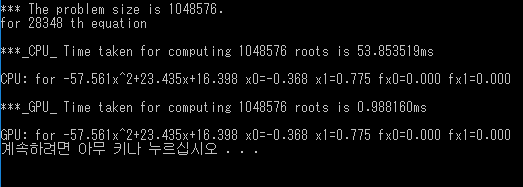
위와 같이 기존 코드를 수정하여x0<=x1 이 되도록 한다.

(ii) 이제 조교가 지정한 N EQUATIONS 값에 대하여 자신의 CUDA 프로그램이 가장 빠른 속도를 보이는 블럭의 크기를 실험적으로 결정한 후 그 값을 보고서에 기술하라.

(iii) 위의 C 함수와 자신의 CUDA 커널 프로그램의 속도를 가급적 정확히 측정하여 그 결과를 보고서에서 비교 분석하라.



커널에서의 계산시간만을 계산하기 위해 위와 같이 작성한다.



위와 같이 근을 확인하고 시간을 출력하였다. 임의의 식에 대해 fx0 fx1이 0임을 확인 할 수 있고, CPU로 계산한 값과 GPU로 계산한 값이 같음을 알 수 있다.

N=1048576에 대하여 CPU time은 평균적으로 13.84ms가 걸렸고,

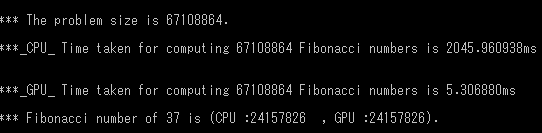
GPU time은 아래와 같다

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Block size | 16 | 64 | 256 | 576 | 1024 |
| Compute time(ms) | 0.706 | 0.407 | 0.406 | 0.407 | 0.434 |

간단한 계산이기에 계산시간이 1ms 이하로 매우 짧아 큰 차이가 나지 않았다. 다만 여전히 블록사이즈가 작을 때 더 빠른 것을 확인할 수 있으며, 블록사이즈가 너무 작을 경우 블록의 개수가 너무 많아지기 때문에 다시 계산시간이 길어짐을 볼 수 있다.

숙제 2

1. 먼저 본 수업 제공 코드에 기반을 둔 CPU 코드를 작성한 후 처리 속도를 가급적 신빙성 있게 측정하라. (당연히 Visual Studio에서 Release mode를사용할 것)



Release mode 에서의 5번의 측정결과 평균 2047.86ms가 기록되었다.

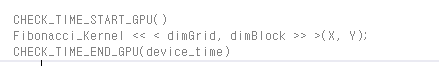
(ii) 다음 이에 대응하는 CUDA 프로그램를 작성한 후 다양한 크기의 블럭에

대하여 속도를 측정한 후, CPU 기반 코드에 비해 얼마나 성능이 향상이 되

는지 분석하라.

(iii) 위의 CPU 코드와 CUDA 코드로 실험한 내용을 자신의 분석 결과와 함께

보고서에 명확히 기술하라.



GPU 커널에서의 계산시간만을 측정한다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Block size | 16 | 64 | 256 | 576 | 1024 |
| Compute time(ms) | 10.03 | 4.712 | 4.733 | 5.157 | 5.308 |

이전의 결과와 같이, 특정 블록사이즈까지는 블록사이즈가 작을수록 계산속도가 빨라지다가 너무 작아질 경우 블록의 개수가 너무 많아져 느려지는 것을 확인 할 수 있다.

속도는 위에서 언급한 것과 같이 이론상으로는 cpu에서의 계산시간이 gpu에서의 계산시간의 (프로세서수\*CPU에서의 계산속도)/(블록쓰레드의수 \*GPU에서의 계산속도)의 배수만큼 길어지는데, 위의 계산의 경우 블록의 개수가 너무 많지 않을 때 대략 400배가량 빠른 것을 확인할 수 있다.