

1.

0)

16 * 16 에서 32 * 32로 증가 시켰습니다.

1) cache block tiling

을 통하여 MNK 곱에서 K가 32를 넘지 않도록 조정하여 메모리 접근성을 높였습니다.

2) 1,2D tiling

을 통하여 메모리 타일성을 높였습니다.

2

MPI_Comm_rank: MPI 환경에서 현재 프로세스의 순위(rank)와 전체 프로세서의 수를 리턴합니다

cudaGetDeviceCount: 사용 가능한 gpu의 수 여기서는 엔비디아 지피유의 수를 얻습니다.

cudaGetDeviceProperties: 각 GPU의 속성을 얻습니다

cudaSetDevice: 특정 GPU를 활성화합니다.

cudaStreamCreate: 각 GPU에 대해 독립적인 스트림을 생성합니다.

cudaMalloc: mallocd의 쿠다버전으로 메모리할당

cudaMemcpyAsync: 호스트에서 디바이스 메모리를 복사합니다.

matmul_kernel<<<>>>: CUDA 커널을 시행 (여기선 matmul커널실행)

cudaStreamSynchronize: 모든 GPU의 작업이 완료될 때까지 대기합니다.

cudaFree: Freed의 cuda 버전할당된 메모리를 해제합니다.

cudaStreamDestroy: 생성된 스트림을 제거합니다.

3.

CUDA의 장점: 당연한 이야기일수 있지만NVIDIA GPU에 최적화되어 높은 성능을 제공합니다.

OpenCL의 장점: 오픈으로 여러 가지 최적화를 여러가지 기기에서 종합적으로 수행할 수 있습니다.

즉 엔비디아면 쿠다를 아니면 오픈씨엘을 쓰는것이 좋습니다.

0)

16 * 16 에서 32 * 32로 증가 시켰습니다.

단순히 gpu 점유율 증가만으로 20 % 정도의 성능향상을 발견하였습니다.

1) cache block tiling

을 통하여 MNK 곱에서 K가 32를 넘지 않도록 조정하여 메모리 접근성을 높였습니다.

여기서 블록 타일링을 추가하여 추가로 20%정도의 성능향상

2) 1,2D tiling

을 통하여 메모리 타일성을 높였습니다.

이 부분에서 현재 막혀 과제 진행이 잘 되지않습니다. 프로젝트까지는 이부분에 대하여 더 최적화를 잘하도록 해야할것입니다.

```
salloc: Relinquishing job allocation 287917
● shpc033@login0:~/hw6/hw6/matmul$ make && NODES=2 ./run.sh 65536 4096 4096 -n 10 -v
make: Nothing to be done for 'all'.
salloc: Granted job allocation 287933
(a05) Hello world, rank 0 out of 2
(a06) Hello world, rank 1 out of 2
Options:
  Problem size: M = 65536, N = 4096, K = 4096
  Number of iterations: 10
  Print matrix: off
  Validation: on

[rank 0] Initializing matrices...Done!
[rank 0] Number of devices: 4
[rank 1] Number of devices: 4
[rank 1] device 0: NVIDIA TITAN RTX
[rank 0] device 0: NVIDIA TITAN RTX
[rank 1] device 1: NVIDIA TITAN RTX
[rank 0] device 1: NVIDIA TITAN RTX
[rank 1] device 2: NVIDIA TITAN RTX
[rank 0] device 2: NVIDIA TITAN RTX
[rank 1] device 3: NVIDIA TITAN RTX
[rank 0] device 3: NVIDIA TITAN RTX
[rank 0] Calculating...(iter=0) 0.147322 sec
[rank 0] Calculating...(iter=1) 0.147323 sec
[rank 0] Calculating...(iter=2) 0.147285 sec
[rank 0] Calculating...(iter=3) 0.147299 sec
[rank 0] Calculating...(iter=4) 0.147387 sec
[rank 0] Calculating...(iter=5) 0.147085 sec
[rank 0] Calculating...(iter=6) 0.147462 sec
[rank 0] Calculating...(iter=7) 0.147466 sec
[rank 0] Calculating...(iter=8) 0.147456 sec
[rank 0] Calculating...(iter=9) 0.147825 sec
Validating...
Result: VALID
[rank 0] Avg. time: 0.147391 sec
[rank 0] Avg. throughput: 14919.674925 GFLOPS
salloc: Relinquishing job allocation 287933
○ shpc033@login0:~/hw6/hw6/matmul$
```

1.5만 기가 플롭스