

# [CSEG483/CSE5483] 기초 GPU 프로그래밍

## HW 2: Matrix Multiplication

담당교수: 서강대학교 컴퓨터공학과 임 인 성

May 8, 2024

**제출 마감:** 5월 20일 (월) 오후 8시 정각 이전에 조교가 사이버 캠퍼스에 공지한 방식으로 제출

**참고:** LATE 없음

이번 숙제에서는 크기가 큰 두 행렬  $A$ 와  $B$ 를 곱하여  $C$ 를 계산하는 문제를 GPU를 사용하여 가속하는 문제에 대하여 실습하여 본다. 다음 7 가지 방법을 구현한 후, 세 가지 서로 다른 크기의 행렬에 대하여 계산 시간을 측정한 후 그 내용을 분석하라.

1. **[방법 1]** Host에서 작동하는 함수를 가급적 효율적으로 구현하라. (MM\_HOST)
2. **[방법 2]** Device에서 shared memory를 사용하지 않는 기본적인 함수를 가급적 효율적으로 구현하라. (MM\_DEVICE\_GM)
3. **[방법 3]** Device에서 shared memory를 사용하는 함수를 가급적 효율적으로 구현하라. (MM\_DEVICE\_SM)
4. **[방법 4]** Device에서 shared memory를 사용하고 “More-Work-per-Thread” 아이디어를 사용하여 최적화된 함수를 가급적 효율적으로 구현하라. (MM\_DEVICE\_SM\_MWPT)
5. **[방법 5]** Device에서 Tensor core를 기반으로 shared memory를 사용하지 않는 함수를 가급적 효율적으로 구현하라. (MM\_DEVICE\_TC\_GM)
6. **[방법 6]** Device에서 Tensor core를 기반으로 shared memory를 사용하여 최적화된 함수를 가급적 효율적으로 구현하라. (MM\_DEVICE\_TC\_SM)
7. **[방법 7]** Device에서 cuBlas에 기반을 둔 함수를 가급적 효율적으로 구현하라. (MM\_DEVICE\_CUBLAS)

**제출물:** 자신이 구현한 코드를 이름이 HW\_2\_학번인 디렉터리 아래의 Visual Studio 프로젝트를 생성한 후, zip으로 압축하여 제출할 것.

1. **자신이 작성한 코드:** Visual Studio 2022를 통하여 확인 할 수 있도록 위의 directory를 제출하되 .vs 파일 등 코드 수행에 불필요한 파일들은 반드시 제거한 후 제출할 것.
2. **프로그램 실행 결과:** 자신의 코드를 실행한 결과를 증빙할 수 있는 자료 (예를 들어, 콘솔 윈도우의 내용을 캡처한 영상)를 보고서에 포함할 것.
3. **보고서:** 자신의 실험 결과를 바탕으로 분석한 내용을 기술할 것.

**참고:**

1. 입력 행렬은 random number를 사용하여 생성한 후, 각 방법이 올바른 계산 결과를 산출하는지 반드시 확인하라.
2. 최소한 한 예는 자신의 컴퓨터에서 허용하는 가급적 크기가 큰 행렬을 사용할 것.
3. 각 방법의 실험 결과를 자신만의 방식으로 의견을 제시하라.