Henry Moran
Santiago Segovia
Alex Yue

Project Proposal: Dense Linear Solver using LU Decomposition in CUDA

Solving systems of linear equations is a fundamental problem in scientific computing, often appearing in simulations, engineering applications, and data science workflows. LU decomposition is a classic technique for solving such systems, particularly when the coefficient matrix is dense and square. CUDA provides a powerful platform to accelerate numerical linear algebra routines.

Objectives:

- Implement a dense LU decomposition algorithm using CUDA C/C++
- Solve linear systems of the form Ax = b using the LU factorization
- Explore optimization techniques such as:
    - Tiling and shared memory usage
    - Coalesced memory access
    - Thread and block decomposition strategies

Stage 1: Implementation

- Develop a serial LU decomposition and forward/backward substitution in C++
- Create a CPU solver for Ax = b using LU

Stage 2: CUDA Parallelization

- Port LU decomposition to CUDA using naive row operations
- Implement forward and backward substitution kernels

Stage 3: Optimization

- Apply tiling to improve memory access patterns
- Use shared memory to minimize global memory access latency
- Investigate load balancing and occupancy tuning

Stage 4: Testing and Evaluation

- Benchmark speedup against CPU implementation for increasing matrix sizes (512x512, 1024x1024, …)

Goal:

- We want our algorithm to solve Ax = b accurately within a specific error tolerance
- Demonstrate performance gain over CPU implementation
- Can handle large matrices with minimal performance degradation