Problem statement: High performance batched LU factorization of simultaneous systems:

Implement an optimized batched LU factorization (with partial pivoting) solver for dense (and sparse) matrices that can efficiently handle hundreds to thousands of small-to-medium linear systems (50×50 to 500×500) simultaneously on GPU architectures. The solution should demonstrate:

- Effective use of GPU parallelism with CUDA
- Efficient batched processing strategies
- Performance competitive with standard libraries (e.g., cuSOLVER, MAGMA)
- Scalability across different matrix sizes and number of matrices

Key challenges:
- Efficient memory management for batched operations
- Thread-level parallelism vs. batch-level parallelism tradeoffs
- Handling pivot sequences across different matrices
- Minimizing thread divergence
- Optimizing for different matrix sizes in same kernel

SIEMENS

Phases

Phase 1: Baseline Implementation

- Implement sequential CPU version with partial pivoting
- Create OpenMP parallel version for multi-core CPUs

Phase 2: CUDA Implementation

- Develop naive CUDA version using:
  - One thread per matrix approach
  - Simple block/grid strategies
- Implement batched cuBLAS/cuSOLVER/MAGMA reference version
- Profile initial implementation with NVIDIA Nsight Compute and Nsight Systems

Phase 3: GPU Optimization

- Implement efficient parallelization strategies
- Optimize for different matrix size regimes:

SIEMENS

Phases

Phase 4: Final Tuning & Benchmarking

- Fine-tune implementation for performance
- Compare against state-of-the-art libraries
- Analyze weak/strong scaling characteristics

Assessment metrics

Correctness (30%): Passes all test cases
Performance (40%): Comparison with cuSOLVER/Magma's performance
Analysis (20%): Insightful optimization documentation
Code Quality (10%): Readability, Kernel organization, and documentation

Optional extensions

Mixed precisions
Sparse matrix support

SIEMENS

Relevant resources

https://en.wikipedia.org/wiki/LU_decomposition
https://ieeexplore.ieee.org/document/10074796
https://docs.nvidia.com/cuda/cusolver/index.html
https://developer.nvidia.com/magma

SIEMENS