Project Report

Group Name: [Your Group Name]

March 28, 2025

1 Introduction

This report summarizes the performance results of various optimization tasks implemented in our project. For each task, we describe the methods used and present the corresponding speedups. Each section includes a performance figure illustrating the results.

2 Task 1: Loop Order Variants

In Task 1, we experimented with different loop orderings for matrix multiplication. The measured average speedups for each ordering are as follows:

Loop Order	mnk	mkn	kmn	nmk	nkm	knm
Speedup	1.00	4.91	4.81	1.01	0.88	1.01

Figure 1 illustrates the performance results.

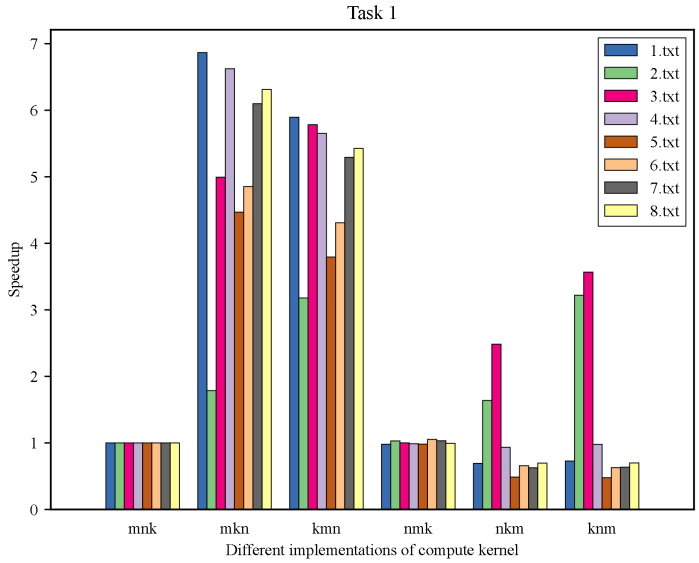

Figure 1: Performance results for different loop orderings in Task 1.

3 Task 2: Transposed Y and Blocking Techniques

In Task 2, we implemented several optimizations including:

- Transposing the matrix Y.
- Blocked matrix multiplication.
- Loop unrolling.

We experimented with different block sizes, loop orders, and unrolling factors. The best performance was achieved with the `t_mnk_lu4` variant, which attained an average speedup of 7.52. Other variants achieved the following speedups:

- `t_mnk`: 5.89,
- `mnkkmn_b32`: 4.63,

- `mnk_lu2`: 1.04,
- `t_mnk_b64_lu4`: 6.93,
- `knmknm_b16_lu2`: 4.33,
- `knmknm_b16_lu2` (alternative): 3.19.

Figure 2 shows the performance comparisons for Task 2.

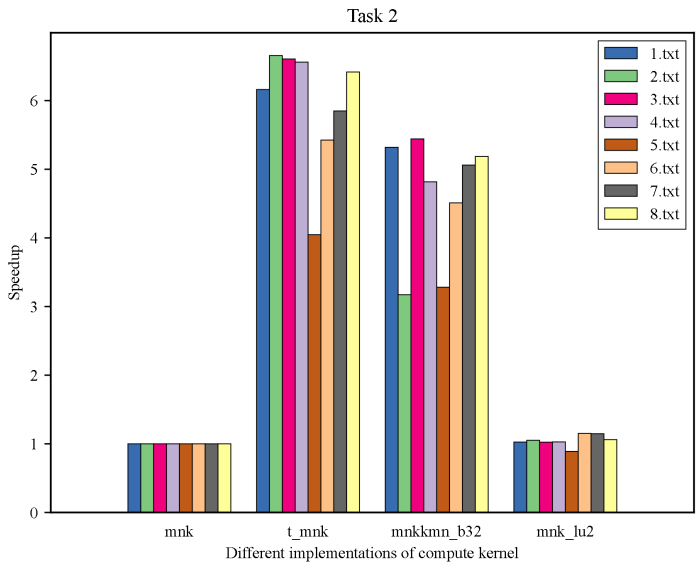

Figure 2: Performance results for different optimizations in Task 2.

4 Task 3: SIMD with 16-bit Data

Task 3 focused on accelerating the inner-product computation using SIMD with 16-bit data (using YP16 and X16). To prevent overflow, the multiplication results were first extended to 64-bit integers before accumulation. The speedups observed for various configurations were:

- `mnk`: 1.00,
- `simd`: 2.63,

- `o3`: 2.62,
- `simd-o3`: 39.61.

Figure 3 illustrates the performance results for Task 3.

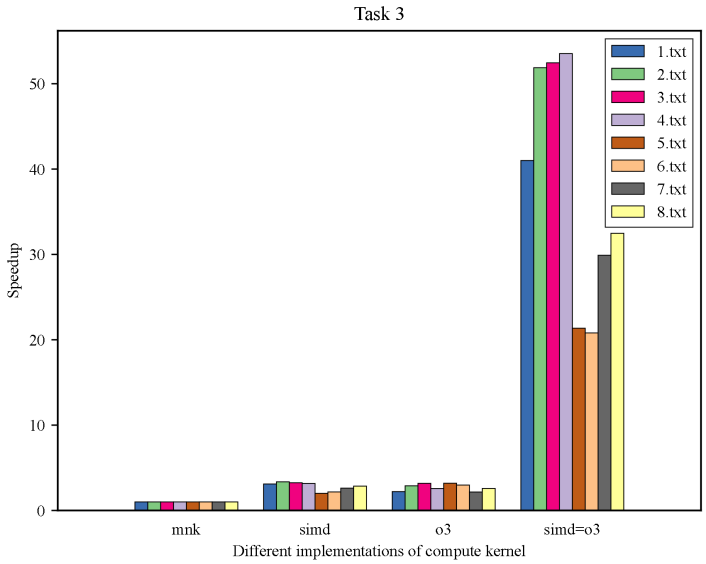

Figure 3: Performance results for SIMD optimizations in Task 3.