RP2: Written Research Assignment

Performance Optimization of Sequence Alignment Algorithms: The Role of GPU Acceleration

Sai Prasanthi Kurra

AIT 512-DL1(Spring 2025)

Part 1: Summaries of Five Selected Papers

This study conducts GPU acceleration used for high-performance sequence alignment by the library GASAL2 for large datasets like Next-Generation-Sequencing. This library uses optimised versions of the Smith-Waterman and Needleman-Wunsch algorithms to boost performance by parallelism, thread management, and memory coalescing. The speed improvement by GASAL2 is 40-80 times when compared to CPU systems. This research paper shows us the advantages of GPU acceleration in enhancing alignment output and correlates with my research interests.

This research paper presents the Database-Oriented Protein Alignment (DOPA), a GPU-optimized implementation of the Smith-Waterman algorithm, which is primarily designed for protein database searches. The authors introduce a study that leverages shared memory on the GPU to enhance the algorithm's performance by reducing memory access overhead. They employ various techniques to optimize database access patterns, which significantly boosts efficiency. The findings indicate that DOPA achieves a speed increase of 25-40 times compared to CPU-based methods while maintaining accuracy. This work highlights the benefits of GPU acceleration for database searches and underscores the importance of memory optimization in accelerating sequence alignment. This is particularly relevant to my research question as it pertains to GPU acceleration of the Smith-Waterman algorithm in terms of execution time and memory efficiency.

This research paper discusses the Frequency Distance Filtration (FDF) technique, which eliminates unlikely protein matches. A hybrid CPU-GPU strategy is aimed to speed up the Smith-Waterman algorithm through this technique. This is a fast strategy that decreases the workload on the GPU. Results show a 32 times performance increase when GPU is added. The algorithm aids in scalability and adaptability since it performs well with both short and long sequences. The accuracy was also maintained while processing speed increased, showing the advantages of hybrid GPU acceleration in improving the Smith-Waterman algorithm's effectiveness.

This research paper illustrates the performance of multi-GPU systems and their energy efficiency by using the Smith-Waterman algorithm. The authors give various frameworks that spread alignment tasks across several GPUs. This paper shows us the advantages of parallelism and workload management. Results include 186 times speed increase using GPU over CPU

approaches. This study also analyzes the energy efficiency of multi-GPU systems (lower per computation) even though there is a higher overall power consumption. The setups that were used to improve speed without additional energy costs are particularly relevant to my research question.

This paper explores a new, optimized version of the Needleman-Wunsch algorithm that employs Compute Unified Device Architecture (CUD) for parallel processing distributing tasks across GPU and MPI cores. This allows the distribution of work across multiple nodes in a supercomputer. The paper explains how the method was implemented and evaluates its performance. Dynamic workload allocation is utilized to prevent bottlenecks and optimize inter-node communication. The result shows that there are significant multi-fold speedups over CPU-based approaches, most particularly for larger datasets and multi-sequence alignments. This paper highlights the efficiency made by parallel processing and gives a detailed performance analysis, which directly correlates to my research question.

Part 2: Synthesis/Conclusion of the selected papers

The articles evaluate the performance of GPU acceleration and show that they greatly improve the efficiency of sequence alignment algorithms, especially Needman-Wunsch and Smith-Waterman. This happens by decreasing the execution times and optimizing memory usage. Multiple research cases noted in the articles illustrate the importance of using techniques like CPU-GPU collaborations and memory coalescing to achieve better alignments. The energy efficiency and scalability provided by multi-GPU systems are shown in these studies. GPUs provide speed improvements and can provide large dataset management. Future studies should aim to develop strategies to utilise GPU acceleration in various applications, including bioinformatics.

References
1. Ahmed, N., Levy, J., Ren, S., Mushtaq, H., Bertels, K., & Al-Ars, Z. (2019). GASAL2: a GPU accelerated sequence alignment library for high-throughput NGS data. BMC Bioinformatics, 20(1), 520–520. https://doi.org/10.1186/s12859-019-3086-9
2. Hasan, L., Kentie, M., & Al-Ars, Z. (2011). DOPA: GPU-based protein alignment using database and memory access optimizations. BMC Research Notes, 4(1), 261–261. https://doi.org/10.1186/1756-0500-4-261
3. Hung, C.-L., Lin, C.-Y., Hong, Y., Liu, Y., & Jiang, H. (2015). Accelerating Smith-Waterman Alignment for Protein Database Search Using Frequency Distance Filtration Scheme Based on CPU-GPU Collaborative System. International Journal of Genomics, 2015(2015), 1–12. https://doi.org/10.1155/2015/761063
4. Perez-Serrano, J., Sandes, E., Magalhaes Alves de Melo, A. C., & Ujaldon, M. (2018). DNA sequences alignment in multi-GPUs: acceleration and energy payoff. BMC Bioinformatics, 19(Suppl 14), 421–421. https://doi.org/10.1186/s12859-018-2389-6
5. Zwaka, L. (2024). *Parallel DNA Sequence Alignment on High-Performance Systems with CUDA and MPI.* https://doi.org/10.48550/arxiv.2412.21103