Thesis Overview:

Viability Study of SYCL as a Unified Programming Model for Heterogeneous Systems Based on GPUs in Bioinformatics

Estudio de viabilidad de SYCL como modelo de programación unificado para sistemas heterogéneos basados en GPUs en bioinformática.

Manuel Costanzo
National University of La Plata, Argentina
PhD in Computer Science[1]
Enzo Rucci, Carlos García-Sánchez and Marcelo Naiouf
{erucci,mnaiouf@lidi.info.unlp.edu.ar}, garsanca@dacya.ucm.es

Abstract

Bioinformatics and Computational Biology are disciplines that have used GPUs for over two decades to accelerate data processing in computational applications, where CUDA has been the most utilized programming language. However, the exclusive use of CUDA presents a portability issue, as it is only compatible with NVIDIA GPUs and not with other heterogeneous architectures, such as AMD or Intel GPUs, or any other type of accelerator. To address this portability challenge, the Khronos Group recently introduced the SYCL standard, a multi-platform programming model that offers a high-level programming interface. This standard facilitates the development of portable applications that can efficiently leverage the capabilities of different hardware devices, such as NVIDIA, Intel, and AMD GPUs, without needing significant modifications to the source code. Therefore, this thesis set out as its general objective *to evaluate the feasibility of SYCL as a unified, portable, and efficient heterogeneous programming model for the design and development of computationally demanding applications on heterogeneous GPU-based systems, specifically in the field of bioinformatics*.

Initially, a detailed investigation was conducted about heterogeneous programming models, performance metrics, and bioinformatics concepts, in order to establish the theoretical foundations of this thesis. Then, the SW# suite was chosen as the case study, as it represents a clear example of a CUDA-based bioinformatics application for biological sequence alignment. Using the SYCLomatic tool, a complete migration of the CUDA code to SYCL was carried out, which involved modifying the generated code and solving runtime errors. Additionally, the functionality was verified, optimizations were applied, and the resulting SYCL code was standardized to be compatible with other SYCL implementations.

Subsequently, multiple experiments were conducted to evaluate the functionality and performance portability of the software migrated to SYCL. These experiments involved running the application on a wide variety of HPC platforms, including different CPUs and GPUs from various manufacturers. The results obtained demonstrated performance comparable to CUDA in most configurations, confirming the effectiveness of SYCL. Moreover, good performance portability across platforms was observed due to SYCL's ability to run on various hardware combinations. Furthermore, performance remained consistent when switching SYCL implementations. In conclusion, this study demonstrates that SYCL is a viable alternative as a unified, portable, and efficient programming model in the context of heterogeneous computing with GPUs for bioinformatics applications. The findings of this study lay the groundwork for the transition of legacy applications and the development of new solutions leveraging the capabilities of SYCL.

Contributions

- A study on the feasibility and efficiency of using SYCL for the development of bioinformatics applications in heterogeneous systems based on GPUs, with the purpose of identifying existing possibilities and methods for code creation, as well as the associated advantages and challenges in real development contexts. This is particularly relevant for developers in the field of bioinformatics (and potentially other areas) who are considering migrating an existing application to SYCL, or who are contemplating starting a new project from scratch using this technology. The study will help to identify possibilities and methods for code creation, as well as the associated advantages and challenges in real development contexts.
- The complete migration of the SW# suite, using automatic tools for migrating CUDA code to SYCL. SW# offers advanced features that allow computing alignments of both DNA and protein sequences;

[1] Full text available at *https://sedici.unlp.edu.ar/handle/10915/164928*

customizing the algorithm used according to its purpose, such as Smith-Waterman, Needleman-Wunsch, Semi-global, and Overlap; adjusting the scoring scheme (substitution matrix plus gap penalties); among others. Like the original version, the migrated version of SW# can combine the computing power of the CPU and GPU, although in this case, it is done through the same programming language. By allowing the configuration of the number of CPU threads and the GPU devices to be used, flexibility is provided for different hardware configurations. Finally, SW# can be used as an independent tool as well as a library, which facilitates its integration into broader bioinformatics workflows. For the benefit of the scientific community, the complete migration of SW# to SYCL is available in a public web repository.

- An exhaustive study on the functional and performance portability of SYCL in the field of bioinformatics. This analysis considers different variants of the ASB application, various implementations of SYCL, and their behavior on systems based on CPUs and GPUs from different manufacturers, as well as the combination of both in a hybrid environment. It is important to note that there are no previous studies that have used such a wide and diverse set of platforms in the existing literature. This study is especially useful for researchers and developers in the field of bioinformatics who seek to optimize their applications for more efficient processing. By providing an exhaustive analysis of the variations in SYCL's performance on different hardware configurations, this study contributes to determining whether SYCL represents a viable alternative as a unified programming model for heterogeneous computing. Additionally, this analysis also benefits the academic community, providing an educational resource for those interested in learning about performance portability across different computing environments and its practical application in bioinformatics.

Publications associated with this thesis

[1] M. Costanzo, E. Rucci, C. García-Sanchez, M. Naiouf, and M. Prieto-Matías, «Assessing opportunities of SYCL for biological sequence alignment on GPU-based systems», *J Supercomput*, feb. 2024, doi: 10.1007/s11227-024-05907-2.

[2] M. Costanzo, E. Rucci, C. García-Sánchez, M. Naiouf, and M. Prieto-Matías, «Migrating CUDA to oneAPI: A Smith-Waterman Case Study», en *Bioinformatics and Biomedical Engineering. IWBBIO 2022. Lecture Notes in Computer Science.*, I. Rojas, O. Valenzuela, F. Rojas, L. J. Herrera, y F. Ortuño, Eds., in Lecture Notes in Computer Science, vol. 13347. Cham: Springer International Publishing, jun. 2022, pp. 103-116. doi: 10.1007/978-3-031-07802-6_9.

[3] M. Costanzo, E. Rucci, C. García-Sánchez, M. Naiouf, and M. Prieto-Matías, «Comparing Performance and Portability Between CUDA and SYCL for Protein Database Search on NVIDIA, AMD, and Intel GPUs», in *2023 IEEE 35th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*, Porto Alegre, Brasil: IEEE, oct. 2023, pp. 141-148. doi: 10.1109/SBAC-PAD59825.2023.00023.

[4] M. Costanzo, E. Rucci, C. García-Sánchez, and M. Naiouf, «Brief performance portability analysis of a matrix multiplication kernel on multiple vendor GPUs», en *Short Papers of the 11th Conference on Cloud Computing Conference, Big Data & Emerging Topics (JCC-BD&ET 2023)*, jul. 2023, pp. 13-18. [Online]. Available at: http://sedici.unlp.edu.ar/handle/10915/155420

[5] M. Costanzo, E. Rucci, C. García-Sánchez, and M. Naiouf, «Early Experiences Migrating CUDA codes to oneAPI», in *Short papers of the 9th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2021)*, jul. 2021, pp. 14-18. [Online]. Available: http://sedici.unlp.edu.ar/handle/10915/125138

Citation: M. Costanzo. "*Thesis Overview: Viability Study of SYCL as a Unified Programming Model for Heterogeneous Systems Based on GPUs in Bioinformatics*". Journal of Computer Science & Technology, vol. 24, no. 2, pp. 193-194, 2024.
DOI: 10.24215/16666038.24.e18
Copyright: This article is distributed under the terms of the Creative Commons License CC-BY-NC-SA.