Towards a Portable Drug Discovery Pipeline with SYCL 2020

Luigi Crisci, Majid Salimi Beni, Biagio Cosenza
University of Salerno
Fisciano, Italy

Nicolò Scipione, Davide Gadioli, Emanuele Vitali, Gianluca Palermo
Politecnico di Milano
Milan, Italy

Andrea Rosario Beccari
Dompé Farmaceutici S.p.A.
L'Aquila, Italy

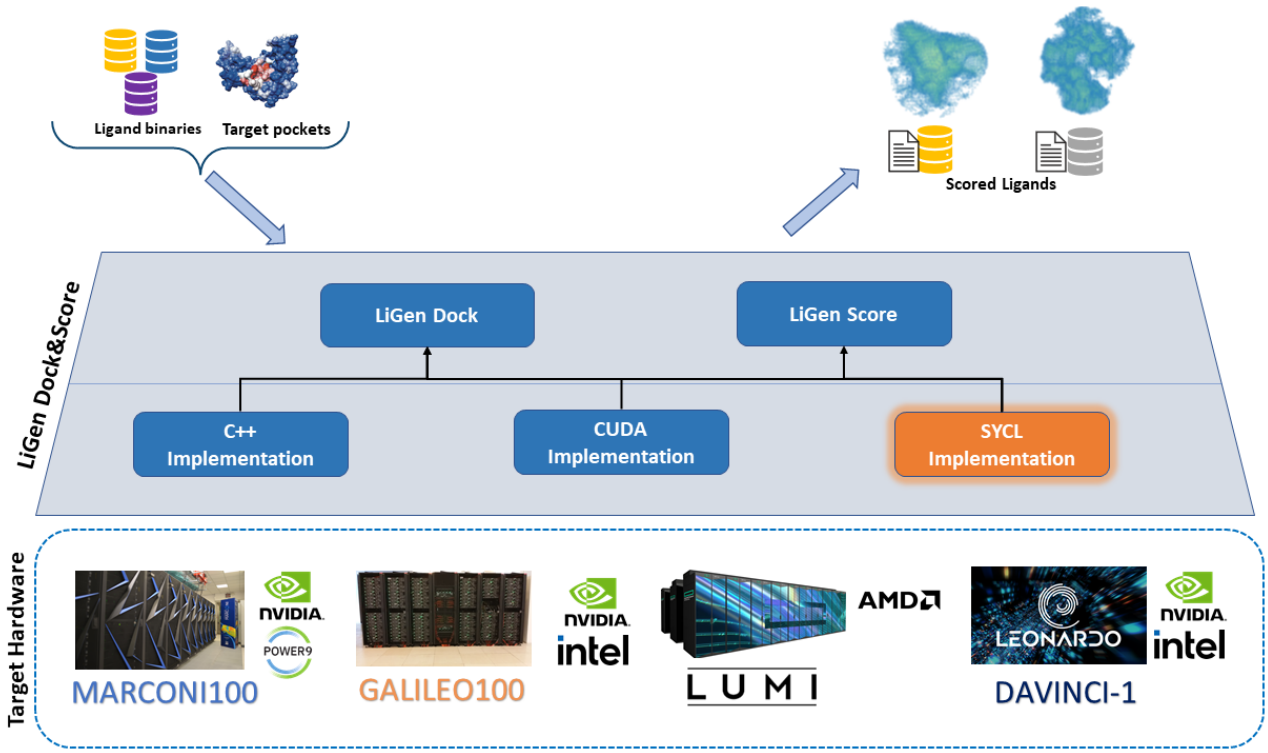

Figure 1: LiGen Dock&Score Architecture overview

ABSTRACT

The outcome of the drug discovery process is a molecule that has strong interaction with the target protein. Domain experts expect a beneficial effect from this interaction. The virtual screening is one of the early stages of the process and it aims at finding promising molecules to forward to later stages. We perform this task in-silico to evaluate a very large chemical library in a short time frame. This activity typically comprises two compute-intensive tasks: a docking function that predicts the displacement of atoms, and a scoring function, which estimates the interaction strength [6] Dompé Farmaceutici led the development of LiGen [1–3], a molecular docking platform targeting High-Performance Computing systems. LiGen has been used for the discovery of novel treatments in the fight against viral infections and multidrug-resistant bacteria [4]. The LiGen processing pipeline includes two main components, ligen-dock and ligen-score, originally developed in OpenACC, refactored to CUDA using non-portable target-specific optimizations [7].

In this talk, we discuss the challenges of making the LiGen docking pipeline portable among different accelerators and GPUs by porting the original codebase from CUDA to SYCL. The code has been refactored by removing critical CUDA semantics with portable ones, and by exploiting several features from the SYCL 2020 standard [5], including sub-groups, group algorithms, and Unified Shared Memory. For comparison, we have developed two versions based on, respectively, accessor and USM-based memory accesses. Particular efforts have been spent on kernel tuning, in particular to optimize those kernels with high register pressure.

The final SYCL code base, comprising more than 20 SYCL kernels, has been evaluated on several architectures including NVIDIA V100, NVIDIA A100, AMD MI100 as well as Intel Xeon, and by using both HipSYCL and Intel DPC++ compiler. In terms of performance

IWOCL'22, May 10–12, 2022, Bristol, United Kingdom, United Kingdom

portability, the SYCL implementation achieves similar performance compared to the CUDA native version on NVIDIA V100 and AMD M100, with minimal modification needed.

CCS CONCEPTS

• **Applied computing → Chemistry**; • **Computing methodologies → Massively parallel algorithms**; **Parallel computing methodologies**.

KEYWORDS

drug discovery, docking, scoring, SYCL

ACM Reference Format:
Luigi Crisci, Majid Salimi Beni, Biagio Cosenza, Nicolò Scipione, Davide Gadioli, Emanuele Vitali, Gianluca Palermo, and Andrea Rosario Beccari. 2022. Towards a Portable Drug Discovery Pipeline with SYCL 2020. In *International Workshop on OpenCL (IWOCL'22), May 10–12, 2022, Bristol, United Kingdom, United Kingdom.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3529538.3529688

ACKNOWLEDGMENTS

LiGen has been funded by the LIGATE project, receiving funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956137. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Italy, Sweden, Austria, Czech Republic, Switzerland.

REFERENCES

[1] Claudia Beato, Andrea R. Beccari, Carlo Cavazzoni, Simone Lorenzi, and Gabriele Costantino. 2013. Use of Experimental Design To Optimize Docking Performance: The Case of LiGenDock, the Docking Module of Ligen, a New De Novo Design Program. *Journal of Chemical Information and Modeling* 53, 6 (2013), 1503–1517. https://doi.org/10.1021/ci400079k arXiv:https://doi.org/10.1021/ci400079k PMID: 23590204.

[2] Andrea R. Beccari, Carlo Cavazzoni, Claudia Beato, and Gabriele Costantino. 2013. LiGen: A High Performance Workflow for Chemistry Driven de Novo Design. *Journal of Chemical Information and Modeling* 53, 6 (2013), 1518–1527. https://doi.org/10.1021/ci400078g arXiv:https://doi.org/10.1021/ci400078g PMID: 23617275.

[3] Davide Gadioli, Gianluca Palermo, Stefano Cherubin, Emanuele Vitali, Giovanni Agosta, Candida Manelfi, Andrea R. Beccari, Carlo Cavazzoni, Nico Sanna, and Cristina Silvano. 2021. Tunable approximations to control time-to-solution in an HPC molecular docking Mini-App. *J. Supercomput.* 77, 1 (2021), 841–869. https://doi.org/10.1007/s11227-020-03295-x

[4] Davide Gadioli, Emanuele Vitali, Federico Ficarelli, Chiara Latini, Candida Manelfi, Carmine Talarico, Cristina Silvano, Carlo Cavazzoni, Gianluca Palermo, and Andrea Rosario Beccari. 2021. EXSCALATE: An extreme-scale in-silico virtual screening platform to evaluate 1 trillion compounds in 60 hours on 81 PFLOPS supercomputers. *arXiv preprint arXiv:2110.11644* (2021).

[5] Khronos SYCL Working Group. 2021. *SYCL 2020 Specification, revision 4.* Technical Report. Khronos Group.

[6] Nataraj S. Pagadala, Khajamohiddin Syed, and Jack Tuszynski. 2017. Software for Molecular Docking: a Review. *Biophysical Reviews* 9 (2017), 91–102.

[7] Emanuele Vitali, Davide Gadioli, Gianluca Palermo, Andrea Beccari, Carlo Cavazzoni, and Cristina Silvano. 2019. Exploiting OpenMP and OpenACC to accelerate a geometric approach to molecular docking in heterogeneous HPC nodes. *The Journal of Supercomputing* 75, 7 (2019), 3374–3396.