2022 IEEE 30th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM) | 978-1-6654-8332-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/FCCM53951.2022.9786139

Biological Activity Prediction of GPCR-targeting Ligands on Heterogeneous FPGA-based Accelerators

Ruiqi Chen[1], Yuhanxiao Ma[2], Shaodong Zheng[3, 4], Shizhen Huang[3], Chao Chen[4], Jun Yu[1], Kun Wang[1]
[1]Zhangjiang Fudan International Innovation Center, Fudan University, Shanghai, China
[2]Gallatin School of Individualized Study, New York University, New York, United States of America
[3]College of Physics and Information Engineering, Fuzhou University, Fuzhou, China
[4]VeriMake Innovation Lab, Nanjing, China

***Abstract*—In the drug discovery process, the biological activity value (BAV) of G Protein-Coupled Receptors (GPCRs) targeting ligands is a large consideration. Past BAV prediction on CPU consumes tremendous time and power, yet there is rarely any related acceleration research. Therefore, this paper proposes a series of heterogeneous FPGA-based accelerators for well-performing algorithms to predict GPCRs ligands BAV. Communication delay is reduced by compressing the sparse matrix and directly coupling accelerators on the system BUS. Computation is accelerated by the remapping during the weight storage. Experimental results show that our FPGA accelerator implemented on Xilinx XCZU7EV performs $54.5\times$ faster than CPU and $35.2\times$ more energy-efficient than GPU.**

I. Introduction

The biological activity value (BAV) of the G Protein-Coupled Receptors (GPCRs) targeting ligands is crucial to the drug development, and is the core of the Computer-Aided Drug Design (CADD). CADD usually adopts machine learning algorithms to generate BAV prediction models of the GPCRs ligands with numerous samples. However, existing methods [1] are time-consuming and almost no prior research focuses on its acceleration. Therefore, this paper presents a series of FPGA hardware accelerators for GPCRs ligands BAV prediction. High-performance algorithms including screening for lasso of extended-connectivity fingerprints and deep neural nets (SED), multitask regression learning with group lasso (MTR-GL), and multi-source transfer learning with graph neural network (MSTL-GNN) are flexibly deployed to FPGAs PL side with the High-Level Synthesis (HLS). As a result, the data processing speed is improved and the power consumption is significantly reduced compared with CPUs and GPUs.

II. The Accelerators Architecture

The HLS technology is used to quickly complete the SED, MTR-GL, and MSTL-GNN operations as the results solely depend on the weights on the PC side. However, to achieve lower communication latency and better access efficiency, it is necessary to compress the weight matrix and design the storage and access of the weights in a more focused way [2].

Molecular fingerprints are coded in binary, and are mostly composed of 1s and a majority of 0s. To achieve greater parallelization in BAV prediction, we stitch multiple molecular fingerprints into a matrix. Since the matrix is strongly sparse, we propose a method called binary compressed sparse row (BCSR), as shown in Fig. 1(a). It compresses the data structure of the compressed sparse row (CSR) before data transfer to reduce the communication time and improve the storage efficiency. Specifically, it retains the starting position of vertices or edges, and introduces matrices in the row and column directions to assist in identifying vertex and edge information. Moreover, to achieve greater parallelism, more weights from the BRAM are required. In order to streamline the calculation of these algorithms, the weights of modules are stored and read in a distributed manner. According to the equation shown in Fig. 1(a), the weights are obtained by pre-processing and then stored in the BRAM.

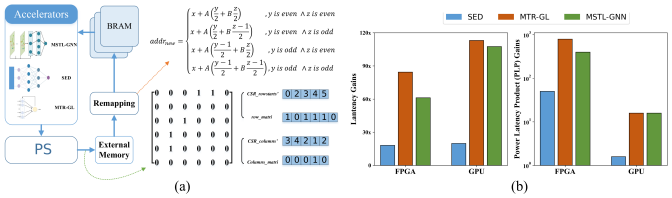

Fig. 1. Overview of the accelerators for predicting GPCRs ligands BAV. (a) The architecture of accelerators with BCSR format and weights remapping method. (b) Comparison of performance and PLP for BAV (Normalized over Intel i7-7700K CPU@4.2GHz data)

III. Conclusion

This paper presents a series of hardware accelerators for BAV prediction of GPCRs ligands, in which the users could select algorithms according to their practical needs. We are the first to deploy 3 well-performing BAV prediction algorithms on FPGA - SED, MTR-GL, and MSTL-GNN. And the system achieves significantly better performance, averagely $54.5\times$ than CPUs, and is more energy-efficient, averagely $35.2\times$ than GPUs, as shown in Fig. 1(b). Through address remapping and compression coding, we greatly improve the storage efficiency and reduce the communication latency.

References

1 Wu, J. *et al.*, "Homologous G protein-coupled receptors boost the modeling and interpretation of bioactivities of ligand molecules," *Journal of Chemical Information and Modeling*, vol. 60, no. 3, pp. 1865–1875, 2020.
2 Yu, Y. *et al.*, "Light-OPU: An FPGA-based overlay processor for lightweight convolutional neural networks," in *Proceedings of the 2020 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, pp. 122–132.

This work was supported by the young scientist project of MOE innovation platform.

978-1-6654-8332-2/22/$31.00 © 2022 IEEE