Scheduling Memory Access Optimization for HBM Based on CLOS

Shuang Xue* †, Huawei Liang* †, Qizhe Wu* † and Xi Jin* †
* State Key Laboratory of Particle Detection and Electronics, University of Science and Technology of China, Hefei, China
†Institute of Microelectronics, Department of Physics, University of Science and Technology of China, Hefei, China
Emails: {xs1159,lhw233,wqz1998}@mail.ustc.edu.cn, jinxi@ustc.edu.cn

Abstract*—With the recent release of FPGA boards based on High Bandwidth Memory (HBM), developers could employ unparalleled external memory bandwidth. HBM provides large-scale aggregated memory bandwidth by exposing multiple memory channels to the processing unit. This allows more memory-constrained applications to benefit from FPGA acceleration. However, it is difficult to take full advantage of available bandwidth: when an application requires multiple processing elements to access multiple HBM channels, the limited number of horizontal connections of the built in crossbar in HBM can result in a significant reduction in effective bandwidth for global addressing. To solve this problem, we propose HBM connection, which is a high-performance custom interconnection for FPGA HBM board. The high-performance custom switching network based on CLOS is introduced to replace the built in crossbar to optimize HBM access scheduling, and increase throughput of AXI bus hosts and switching components. The validity of HBM connection is proved by Xilinx VCU128 HBM board. Based on the breadth-first search BFS case study, we conducte an experimental exploration. The result shows that HBM connection improves the effective performance by 2.5X.

Keyword—HBM , CLOS , access schedule , FPGA , BFS

Shuang Xue received his B.S. degree in 2020 from University of Science and Technology of China, and he is currently a Master student in Department of Physics in University of Science and Technology of China, Anhui, China,under the supervision of Prof. Xi Jin. His current research work os mainly on FPGA-based Hardware Accelerator design.

Haiwen Liang received his B.S. degree in 2020 from University of Science and Technology of China, and he is currently a Ph.D. student in Department of Physics in University of Science and Technology of China, Anhui, China,under the supervision of Prof. Xi Jin. His current research work os mainly on FPGA-based Hardware Accelerator design.

Qizhe Wu received his B.S. degree in 2020 from Zhejiang University of Technology, China, and he is currently a Ph.D. student in Department of Physics in University of Science and Technology of China, Anhui, China,under the supervision of Prof. Xi Jin. His current research work os mainly on FPGA-based Hardware Accelerator design.

Xi Jin received his B.S. degree from University of Science and Technology of China, and he is currently an associate professor in Department of Physics in University of Science and Technology of China, Anhui, China,under the supervision of Prof. Xi Jin.His research interests include SOC design technology,VLSI design, computer-aided design methodologies for SoC system integration and FPGA-based Hardware structure design.