Concept study of using Alveo U50 hardware acceleration card in the RIBF DAQ system

Y. Ichinohe,[*1] H. Baba,[*1] S. Takeshige,[*1,*2] and T. Gunji[*3]

High-speed data processing in data acquisition (DAQ) systems using hardware accelerators such as graphic processing unit (GPU) and field-programmable gate array (FPGA) has garnered attention for the accommodation of the increasing intensity of accelerator beams. We have been investigating the possibility of implementing such hardware accelerator devices for the DAQ system at RIBF.

Alveo is a series of data center accelerator cards provided by AMD Xilinx. Each Alveo board contains an Ultrascale+ FPGA chip and supports PCI Express for installation in typical workstations or computing servers. For the concept study, we employed the Alveo U50 card, which has 8 GB HBM (high-bandwidth memory), and supports 100 Gbps connectivity with a QSFP28 port.

Using the Alveo U50 card, we implemented the particle identification (PID) algorithm based on the standard TOF-$B\rho$-ΔE method[1)] equivalent to that implemented in the typical offline analysis software `ANAROOT`. The original C++ source codes were ported into the register transfer level (RTL) design as per the following procedure. 1. The original C++ source codes were refactored to conform to the specification of high-level synthesis (HLS) toolkit implemented in Vitis Unified Software Platform (*e.g.*, dependency on `ROOT` is removed). 2. Compiler directives were manually added in the refactored codes to assist the HLS task in inducing efficient RTL codes. 3. The codes were converted into RTL codes using the HLS toolkit.

Figure 1 shows the schematic of the realized RTL dataflow of the algorithm. The inputs for the algorithm are the raw data streams of F3, F5 and F7 PPACs, F3 and F7 plastic scintillators (PL), and the F7 ion chamber. Further, the outputs are the streams of double-precision floating-point numbers corresponding to the inferred mass-to-charge ratio (A/Q) and atomic number (Z). Solid-lined rectangles indicate the individual tasks, each of which were independently implemented as an FPGA design. Arrows annotated with dotted-lined rectangles are the pipeline registers implemented as FIFO. The generated codes were compiled and linked with the board files such that it could be executed on the FPGA of the card. We confirmed that the results obtained with `ANAROOT` was successfully reproduced by the hardware implementation.

Table 1 shows the achieved performance compared to CPU. The overall PID process was completed in approximately 1000 clock cycles with the achieved clock frequency of approximately 220 MHz, resulting in the latency of approximately 4.3 μs. The achieved pipeline processing performance was the iteration interval of 5 clocks, resulting in throughput of approximately 44 MHz with the clock frequency of approximately 220 MHz. Although the latency for single event processing was longer than that achieved by CPU (Intel Core i9-10900X 3.70 GHz, single thread), the throughput was approximately 20 times higher than that of the CPU. Moreover, this performance was achieved without any manual tuning of the RTL design. Thus, this scheme has potential and we are exploring the possibility of further applications.

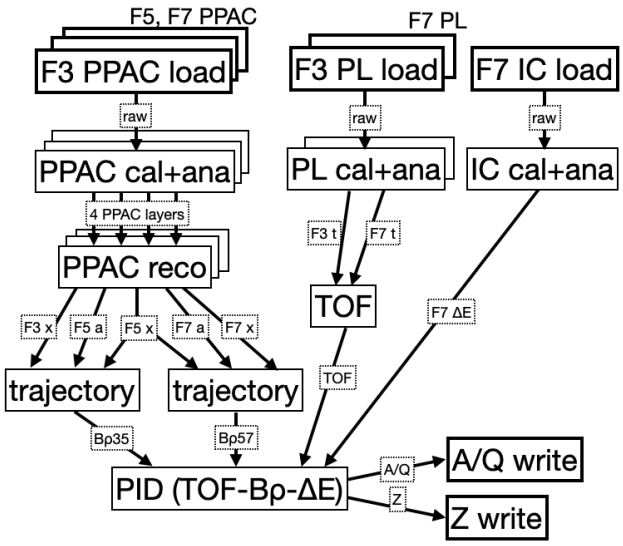

Fig. 1. Realized RTL dataflow of the standard TOF-$B\rho$-ΔE algorithm. *load* modules read the raw data from the HBM and generate raw data streams. *cal+ana* modules perform calibration and basic detector-level analysis. *reco* modules reconstruct the position and direction at each focal plane. *trajectory* modules solve the particle transfer to compute the magnetic rigidity $B\rho$. The *PID* module calculates A/Q and Z based on TOF-$B\rho$-ΔE method.

Table 1. Comparison of the achieved performance.

	Alveo U50	CPU*
Latency	∼4.3 μs	**∼450 ns**
Throughput	**∼44 MHz**	∼2.3 MHz

* Intel Core i9-10900X 3.70 GHz

Reference

1) N. Fukuda *et al.*, Nucl. Instrum. Methods Phys. Res. B **317**, 323 (2003).

*1 RIKEN Nishina Center
*2 Department of Physics, Rikkyo University
*3 Center for Nuclear Study, University of Tokyo

DOI:10.34448/RIKEN.APR.57-119