Particle identification on an FPGA accelerated compute platform for the LHCb Upgrade.

Christian Färber, Rainer Schwemmer, Jonathan Machen, Niko Neufeld,

***Abstract*—The current LHCb readout system will be upgraded in 2018 to a 'triggerless' readout of the entire detector at the LHC collision rate of 40 MHz. The corresponding bandwidth from the detector down to the foreseen dedicated computing farm (event filter farm), which acts as the trigger, has to be increased by a factor of almost 100 from currently 500 GBit/s up to 40 TBit/s. The event filter farm will pre-analyse the data and will select the events on an event by event basis. This will reduce the bandwidth down to a manageable size to write the interesting physics data to tape. The design of such a system is a challenging task, why different technologies are considered and have to be investigated for the different parts of the system. For the usage in the event building or in the event filter farm (trigger) an experimental FPGA accelerated computing platform is considered and therefore tested. FPGA compute accelerators are more and more used in standard servers like for Microsoft Bing search or Baidu search. The platform we use hosts a general CPU and an high performance FPGA linked via an high speed link. On the FPGA an accelerator is implemented. The used system is a two socket platform from Intel with a Xeon CPU and an FPGA. The CPU and the FPGA are connected via the point-to-point interconnect QPI, which is used to interconnect CPUs in industry standard server. The FPGA has cache-coherent memory access to the main memory of the server and can collaborate with the CPU. These cache-coherent architectures are better suited for real-time connections between FPGA and CPU as the usual PCIe FPGA accelerators. It is very likely that these platforms, which are built in general for high performance computing, are also very interesting for the High Energy Physics community. As First step it is tested to port the an existing LHCb RICH particle identification algorithm for the Cherenkov angle reconstruction to the experimental FPGA accelerated platform. We will compare the performance of this algorithm running on a Xeon CPU with the performance of the same algorithm, which is running on the Xeon-FPGA compute accelerator platform. This work is done in collaboration with Intel Corporation.**

***Index Terms*—Field programmable gate arrays, Triggering, Distributed computing.**

I. Introduction

THIS paper describes a study of porting an algorithm from the RICH reconstruction for particle identification on an FPGA accelerated compute platform for the LHCb Upgrade in 2018[1], in which the whole detector readout chain will be modified to make a detector readout of 40 MHz possible[2]. The accelerate compute platform uses a Intel(R) Xeon(R) CPU E5-2680 v2 at 2.80 GHz linked via QPI to a high performance Stratix V FPGA, on which an accelerator is implemented. The implementation of the RICH algorithm for the Cherenkov angle reconstruction[3] is presented and the results are discussed.

C. Färber, Rainer Schwemmer and Niko Neufeld are with the Experimental Physics Department, European Organization for Nuclear Research, Geneva, CERN CH-1211 Geneva 23 Switzerland e-mail: Christian.Faerber@cern.ch

J. Machen is with Intel Corporation.

II. The Intel Xeon/FPGA compute accelerator

The Intel Xeon/FPGA prototype is a two socket server machine. The first socket hosts an Intel(R) Xeon(R) E5-2680 v2 CPU and the second socket hosts an Altera Stratix V GX A7 FPGA. Both are connected via the high bandwidth and low latency interconnect QPI. The FPGA has 234720 ALMs, each contains inputs, several LUT-based resources and four registers to realize any function with up to six inputs. Furthermore, the FPGA has 940000 registers and 256 DSP blocks, which are crucial for any algorithm using floating point calculations and others.

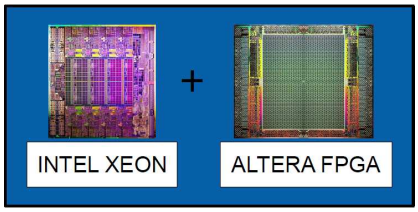

Fig. 1. Intel Xeon/FPGA prototype used for the tests. This is two socket server machine, one with Xeon CPU and the other with Altera FPGA.

The FPGA has cache-coherent access to the main memory of the machine. So far the programming of the algorithms running on the FPGA have to been written in Verilog, but OpenCL is also supported from now on. The power usage of the FPGA is low compared to GPUs and other computing accelerators.

III. FPGA-based accelerator for the RICH reconstruction

The RICH photon reconstruction is used for the particle identification and crucial for the LHCb physics program. In the current version the RICH part of the Trigger software takes so much time that it is not used for every proton-proton collision. A speed-up would be very appreciated for the LHCb Upgrade. For each proton-proton collision hundreds of particle tracks are measured, each of them creates several photons. To find the corresponding Cherenkov cone to every particle track the same calculation has to be done for every found photon. The reconstruction of the Cherenkov angle takes roughly 30% of the second High Level Trigger and is therefore a good candidate to test the possible acceleration with the Xeon/FPGA combination in the High Energy Physics community.

A. Implementation of Cherenkov Angle reconstruction

The algorithm was implemented in a long 753 clock cycle long pipeline, written in Verilog. The design uses 100% of the

Stratix V FPGA. In Table I the FPGA resource usage after the synthesis optimization is shown. The fast QPI interface uses already 30% of the FPGA ALMs, and after the optimization 88% of all the ALMs are used for the whole design. The DSP blocks are used to implement all soft-core floating point calculation blocks and these don't limit the design, like the registers. Only 50% of the Stratix V registeres are used for the design. The pipeline is running with 200 MHz, which makes a calculation for a single photon within 5 ns possible, if the pipeline is completely filled.

TABLE I
FPGA RESOURCE USAGE FOR CHERENKOV ANGLE RECONSTRUCTION

FPGA Resource Type	FPGA Resource used [%]	For QPI used [%]
ALMs	88	30
DSPs	67	0
Registers	48	5

B. Measurement

The runtime for the Cherenkov angle reconstruction was measured on an Intel(R) Xeon(R) E5-2680 v2 using a single thread. The code is partially vectorized. The time measurement for the FPGA version was started after the photon data were written into the main memory and ended after the results are written back to the main memory from the FPGA. For the time measurements the time stamp counter of the CPU was used with the library rdtsc, which was validated with a second different time measurement method. The time was measured for different number of photons from 1 photon up to 2097152 photons.

C. Results

The results are shown in Figure 2 for small number of photons the Xeon CPU is faster due to the large latency of the photon pipeline on the FPGA. The break even is reached for roughly 200 photons. The time for the CPU version rises linearly and the time for the FPGA version stays constant until 10000 photons, for more photons also the time for the FPGA version rises linearly. The time ratio between CPU and FPGA is a factor 20 up to 35. These acceleration factors are very promising. The measured fill stand of the photon pipeline is 50% so the bottleneck of the prototype system is the bandwidth between Xeon CPU and FPGA. In theory with a higher bandwidth an acceleration factor of 64 would be possible with the realized photon pipeline.

IV. CONCLUSION

The LHCb experiment will upgrade its whole detector in 2018 to a flexible 40 MHz readout chain. The raw-data bandwidth will be increases up to 40 TBit/s which the Event Filter Farm has to filter for the interesting events. The result is that the runtime for all the different High Level Trigger algorithms have to be reduced by quite a factor. Different technologies are considered to accelerate these algorithms like GPUs, FPGAs and other computing accelerators.

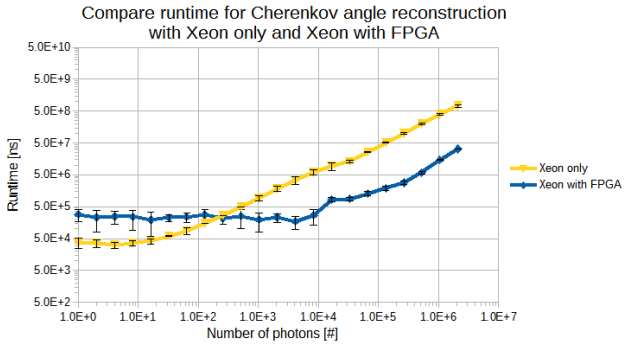

Fig. 2. Time for reconstructing the Cherenkov angle on a CPU compared to the runtime on an FPGA.

Therefore, a study was done to investigate the possible usage of the new Xeon/FPGA compute accelerator. The first High Energy Physics algorithm tested was the Cherenkov angle reconstruction for the particle identification, for which an encouraging acceleration of a factor 35 was achieved with the Xeon/FPGA versus the Xeon CPU alone. The acceleration is limited by the bandwidth to the FPGA, the developed photon pipeline on the Stratix V could run a factor 64 faster than the Xeon CPU alone. These results are very encouraging to use this platform in a High Energy Physics Trigger, especially with the new coming Arria 10 Xeon/FPGAs in a single package and faster interconnect between CPU and FPGA.

Also other algorithms like decompression and re-formatting of packed binary data from the detector for the Event Building are very promising and will be tested in the near future.

ACKNOWLEDGMENT

The authors would like to thank Intel Corporation for the collaboration and support in the High Throughput Computing Collaboration, which made this work possible. Furthermore, the author would like to thank Dr. Manuel Schiller from the Experimental Physics Department at the European Organization for Nuclear Research in Geneva for fruitful discussions about algorithms to calculate cubic-roots.

REFERENCES

[1] LHCb Collaboration, *Framework TDR for the LHCb Upgrade : Technical Design Report*.CERN, Geneva, 2012.
[2] LHCb Collaboration, *LHCb Trigger and Online Upgrade Technical Design Report*.CERN, Geneva, 2014.
[3] R. Forty and O. Schneider, *RICH pattern recognition*.CERN, Geneva, 1998.