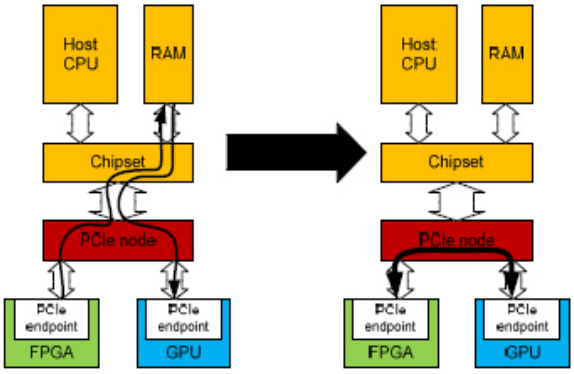
Hes·so
Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences
Western Switzerland



FPGA^2

Framework **P**CIe for **G**PU and **FPGA**

Project Summary

This project aimed at proposing a complete and easy-to-use framework enabling the integration of GPU and FPGA resources communicating through direct PCIe links in the same computational platform. These future heterogeneous HPC systems, taking the best of these two nowadays separate technologies, will allow the acceleration of computation in a wide diversity of different fields such as physics, finance, bio-informatics, high-definition video, etc.

The first step of this project concerned the understanding of a NVidia GPU memory access and mapping, with the hope of being able to directly access the internal memory of the GPU. A lot of trials in many directions were conducted, but concluded that the only directly available portion of memory accessible from the outside is the frame buffer. On the FPGA side, the development of a first prototype allowed to let two FPGAs communicate through PCIe, at a data rate close to 1.5Gb/s. A demonstrator consisting of a FPGA calculating a Mandelbrot curve and then displaying it on a screen by writing into the GPU frame buffer was also developed.

The final framework now consists in a complete FPGA design embedding everything needed to carry DMA transfers through PCIe in an optimized manner. The transfers have to be triggered by the host CPU, the only viable solution taking into account the black box nature of GPUs. These transfers allow of up to 200MB/s data rates, from FPGA to GPU, and slightly less in the opposite direction. This excellent data rate is clearly better than what can be done with a standard approach in which the data is first copied to the central memory before being sent to a device.

The hardware also supplies the ability to manage input and output FIFOs that can let a user application deal with streams of data. The user application being a module embedded into a simple interface, developing applications with this framework will be very easy. Additionally, the user application can raise interrupts in order to better synchronize with the host CPU.

On the software side, a FPGA2 driver has been developed and 4 examples software are available, showing all the capabilities of the system, and being potentially good starting points for new applications.

Valorisation

First of all, this project has been a unique opportunity to really investigate the GPU world, in terms of low-level communication and API, and the team has now a very good view on all the limitations and possibilities of PCIe exploitation of a Nvidia GPU. During the project, some industries have been approached in order to better understand their potential needs, mainly for video applications that could take advantage of the two worlds. We will now contact them with the results of this project, in order to try to build some CTI projects propositions. Academic contacts already showed great interest in this project, and will also receive the scientific report. The team also hopes that this project will help further projects of the ISYS institutes, and on the pedagogical point of view, the FPGA design will serve as a "fil rouge" for the master course about verification of digital systems. Finally, the scientific report, as well as all the source files, will be made available to the community by putting them online on the REDS website.

Contact / Prof. Yann Thoma (yann.thoma@heig-vd.ch)

This project has been carried out by the REDS Institute of the HEIG-VD

ISYS
heig-vd
Haute Ecole d'Ingénierie et de Gestion
du Canton de Vaud
REDS
Reconfigurable & Embedded
Digital Systems