Exploring Low Power Heterogeneous MPSoCs as an Option for High Performance Computing

William Simon – Professor David Atienza
Embedded Systems Laboratory
william.simon@epfl.ch

Introduction
Heterogeneous Multi-Processor Systems-on-Chip (MPSoCs) show promise in helping High Performance Computing (HPC) applications meet ever more stringent power, temperature and Quality-of-Service (QoS) constraints. This thesis utilizes an MPSoC to accelerate an HPC application, HEVC video encoding[1], and explore the possibilities of encoding multiple streams on limited hardware resources.

Theory/Method/Hypothesis

The interpolate, clip3, and SAD functions consume 98\% of encoding runtime for the HEVC application.

This thesis proposes to accelerate these functions via an accelerator/DMA (or IP) on the FPGA of a low power MPSoC.

Using an MPSoC and FPGA has three advantages, namely,
- reducing encoding time, making HEVC more real-time
- offloading computationally expensive functions to the FPGA
- reducing energy consumption in comparison to other hardware.

Also of interest is the ability to encode multiple video streams simultaneously, simulating a web streaming environment that may be encountered by companies such as Netflix.

The accelerator/DMA is designed with the following characteristics:
- Lightweight, filling less than 7\% of target board FPGA[2],
- Generic, supporting various levels of computational complexity,
- Implemented multiple times, allowing multiple video encodings.

The top level implementation of the hardware is displayed below.

Results

The above figures demonstrate the attained encoding acceleration in seconds at different CPU frequencies. A maximum speedup of 46x was achieved. In fact, the runtime on the low-power MPSoC approaches that of the same program run on a reference 3.2GHz Intel processor (seen in green above).

The above figure demonstrates that multiple encoding streams can utilize the same CPU with a ~10\% increase in computation time as computational complexity increases, validating the theory that offloading computation to the FPGA supports online, multi-stream encoding.

Conclusion/Perspectives
This thesis provides a strong motivation for FPGA acceleration of HEVC kernels. It can also be generalized to address the question of how to provide kernel acceleration to multiple processes utilizing limited hardware space. In future work we would like to perform this generalization to provide solutions for task mapping multiple processes to hardware accelerators.

Acknowledgments
I would like to thank Dr. Marina Zapater and Mr. Arman Iranfar for providing guidance and advice on this thesis, as well as Professor David Atienza for supervising this thesis.

References