Project Proposal (Advanced Computer Architectures)

GPU-oriented estimation of stack size for OpenCL kernels

Stefano Cherubin

Politecnico di Milano

Size of the project: 1 student

Reference person: Stefano Cherubin <stefano.cherubin@polimi.it>

Abstract

You will work on a LLVM analysis pass to perform an estimation at machine-independent level of
what is likely to be allocated in registers, cache or global memory. This analysis will start
rely only on LLVM-IR source code and coarse-grained description of the target hardware.

Pre-requirements

- basic/intermediate knowledge of C++11 language and idioms

- basic Linux skills

- [preferred] previous experience with OpenCL

Involved Technologies and Frameworks

- LLVM

- C++11

- OpenCL

- GPGPU

Detailed description

Stack size is an important factor in the mapping decision when dealing with embedded heterogeneous architectures, where fast memory is a scarce resource. Trying to map a kernel onto a device with insufficient memory may lead to reduced performance or even failure to run the kernel. OpenCL kernels are often compiled just-in-time, starting from the source code or an intermediate machine-independent representation. Precise stack size information, however, is only available in machine-dependent code. We provide a method for computing the stack size with sufficient accuracy on machine-independent code, given knowledge of the target ABI and register file architecture. This method can be applied to make mapping and optimization decisions early, thus avoiding to compile multiple times the code for each possible accelerator in a complex embedded heterogeneous system.

The project will rely on a state-of-the-art analysis designed for x86_64 architectures and will provide an adaptation for GPGPU architectures.