Automatic Hardware Pragma Insertion in High-Level Synthesis: A Non-Linear Programming Approach

Stéphane Pouget, *University of California, Los Angeles*
Louis-Noël Pouchet, *Colorada State University*
Jason Cong, *University of California, Los Angeles*
Contact: pouget@cs.ucla.edu

High-Level Synthesis enables the rapid prototyping of hardware accelerators, by combining a high-level description of the functional behavior of a kernel with a set of micro-architecture optimizations as inputs. Such pragmas may describe the pipelining and replication of units, or even higher-level transformations for HLS such as automatic data caching using the AMD/Xilinx Merlin compiler. Selecting the best combination of pragmas, even within a restricted set, remains particularly challenging and the typical state-of-practice uses design-space exploration to navigate this space. But due to the highly irregular performance distribution of pragma configurations, typical DSE approaches are either extremely time consuming, or operating on a severely restricted search space.

In this work we propose a framework to automatically insert HLS pragmas in regular loop-based programs, supporting pipelining, unit replication (coarse- and fine-grain), and data caching. We develop a simple analytical performance and resource model as a function of the input program properties and pragmas inserted. We prove this model provides a lower bound on the actual performance for any possible configuration. We then encode this model as a Non-Linear Program, by making the pragma configuration as unknowns of the system, which is computed optimally by solving this NLP. This approach can also be used during DSE, to quickly prune points with a (possibly partial) pragma configuration, employing this latency lower bound property. We extensively evaluate our end-to-end fully implemented system, showing it can effectively manipulate spaces of billions of designs in seconds to minutes for the kernels evaluated.

Keywords: HLS; FPGA; Non-Linear Programming; Program Optimization

DOI: https://doi.org/10.1145/3626202.3637593