Efficient Generation of Energy and Performance Pareto Front for FPGA Designs

Sanmukh R. Kuppannagari
Viktor K. Prasanna
Ming Hsieh Department of Electrical Engineering,
University of Southern California.
Contact:{kuppanna, prasanna}@usc.edu

Analysis of trade-offs between energy efficiency and latency is essential to generate designs complying with a given set of constraints. Improvements in FPGA technologies offer a myriad choices for power and performance optimizations. Various algorithm intrinsic parameters also affect these objectives. The design space is compounded by the available choices. This requires efficient techniques to quickly explore the design space. Current techniques perform Gate/RTL level or functional level power modeling which are slow and hence not scalable. In this work we perform efficient design space exploration using a high level performance model. We develop a semi-automatic design framework to generate energy efficiency and latency trade-offs. The framework develops a performance model given a high level specification of a design with minimal user assistance. It then explores the entire design space to generate the dominating designs with respect to energy efficiency and latency metrics. We illustrate the framework using convolutional neural network which gained significance due to its application in deep learning. We simulate a few designs from the dominating set and show that the performance estimation for the dominating designs are close to the simulated results. We also show that our framework explores 6000 design points per minute on a commodity platform such as Dell workstation as opposed to state-of-the-art techniques which explore at 50 to 60 design points per minute.

Categories and Subject Descriptors

C.1.3 [**Other Architecture Styles**]: Adaptable architectures—*Design Space Exploration, Performance Modeling*
Keywords: High Level Performance Model; Design Space Exploration; Energy Efficiency; Design Framework; Convolutional Neural Networks
DOI: http://dx.doi.org/10.1145/2684746.2689133
Acknowledgements: NSF Grant Number: 1018801, DARPA Grant Number: HR0011-12-2-0023.