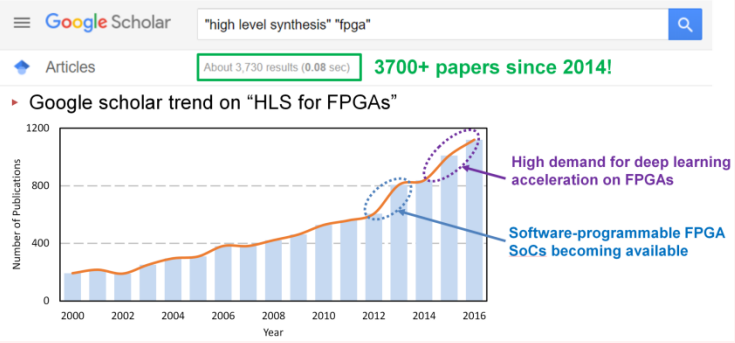
Rosetta: A Realistic HLS Benchmark Suite for FPGAs

Students: Yuan Zhou, Ritchie Zhao, Hanchen Jin
Faculty Advisor: Zhiru Zhang
School of Electrical and Computer Engineering, Cornell University

Motivation

Increasing Development of HLS for FPGAs



➔ Need Realistic HLS Benchmarks

- For testing a rich set of synthesis directives in modern HLS tools
- For evaluating new classes of HLS optimization techniques

Rosetta Overview

An open-source real-application based benchmark suite for software programmable FPGAs

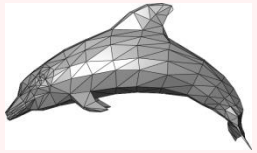
3**R** characteristics:

- **R**ealistic applications with optimizations and constraints, as it is complex enough for evaluating modern HLS tools and techniques
- **R**etargetable to different FPGA platforms and HLS tools, as it is specified in multiple programming languages
- **R**educible to sub-kernels for effective micro-benchmarking

Realistic Benchmarks

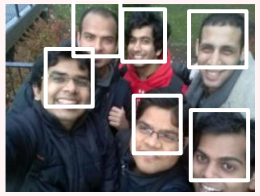
Six designs from machine learning and video processing domains

3D Rendering

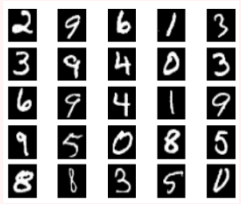


Optical Flow

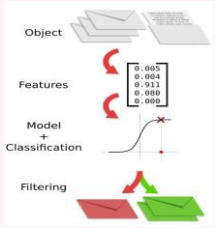
Face Detection



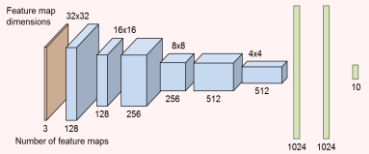
Digit Recognition



Spam Filtering



Binarized Neural Network



HLS Optimizations

To achieve application level constraints, Rosetta designs are customized using a variety of HLS optimization techniques:

- Datatype Customization: fixed-point types
- Compute Customization: parallel computing including loop unrolling, loop pipelining, and dataflow pipelining
- Memory Customization: customized memory hierarchy including data reuse and memory banking
- Communication Customization: proper data packing and careful design of data layout

Experimental Results

Embedded FPGA: Xilinx ZC706

Benchmark	#LUTs	#FFs	#BRAMs	#DSPs	Runtime (ms)	Throughput
3D Rendering	8893	12471	48	11	4.7	213 frames/s
Digit Recognition[1]	41238	26468	338	1	10.6	189k digits/s
Spam Filtering[2]	12678	22134	69	224	60.8	370k samples/s
Optical Flow	42878	61078	54	454	24.3	41.2 frames/s
Face Detection	62688	83804	121	79	33.0	30.3 frames/s
BNN[3]	46899	46760	102	4	4995.2	200 images/s

1: K=3, PAR_FACTOR=40. 2: Five epochs, PAR_FACTOR=32, VDWIDTH=64. 3: Eight convolvers, 1000 test images.

Cloud FPGA: AWS F1

Benchmark	#LUTs	#FFs	#BRAMs	#DSPs	Runtime (ms)	Throughput	Performance-cost Ratio
3D Rendering	6763	7916	36	11	4.4	227 frames/s	**496k frames/$**
Digit Recognition[1]	39971	33853	207	0	11.1	180k digits/s	**393M digits/$**
Spam Filtering[2]	7207	17434	90	224	25.1	728k samples/s	**1.6G samples/$**
Optical Flow	38094	63438	55	484	8.4	119 frames/s	**260k frames/$**
Face Detection	48217	54206	92	72	21.5	46.5 frames/s	**101k frames/$**

1: K=3, PAR_FACTOR=40. 2: Five epochs, PAR_FACTOR=32, VDWIDTH=512.

Conclusion & Future Work

Rosetta can serve as

- A realistic benchmark suite for HLS research community
- A useful design tutorial for HLS/FPGA users

We plan to further

- Continue to optimize current available applications
- Incorporate more applications from emerging domains

Find Rosetta here ➔ github.com/cornell-zhang/rosetta

Acknowledgements

This project was supported in part by a research gift from Xilinx. We thank Sumit Roy and Kumar Deepak for providing helpful feedback on the Rosetta designs.