SIMD Performance and Yield Optimization with Multi-granularity Redundancy

Daniel Epstein, Kevin Skadron
Department of Computer Science, University of Virginia

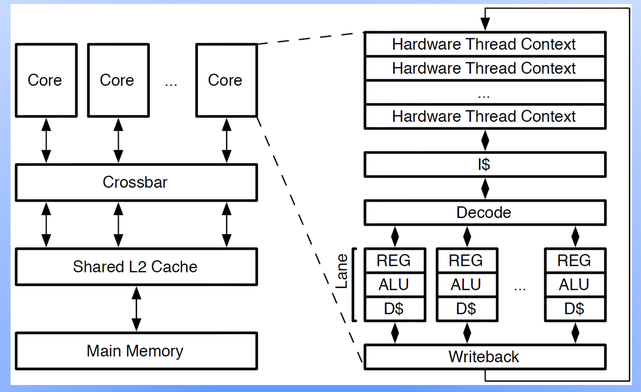
Brett Meyer
Department of Electrical and Computer Engineering, McGill University

The Problem

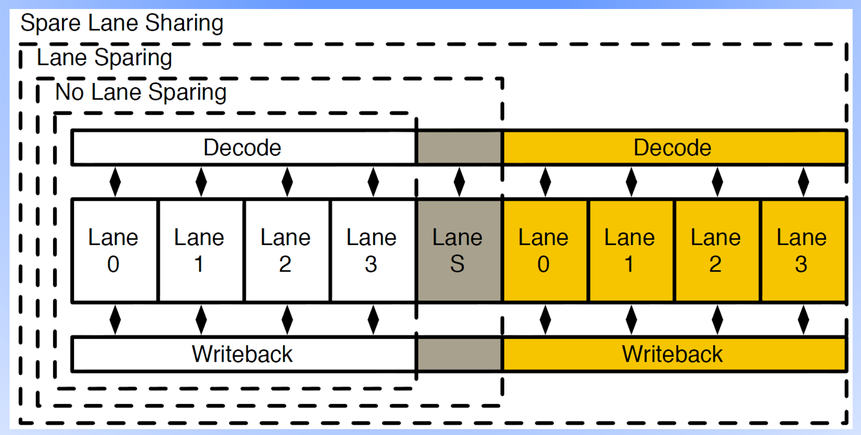
- Die yields are decreasing
- Sparing used to increase yield
- Sparing can be applied at multiple granularities
- What is "best" for a particular design?

SIMT

- Single-Instruction, Multiple-Thread



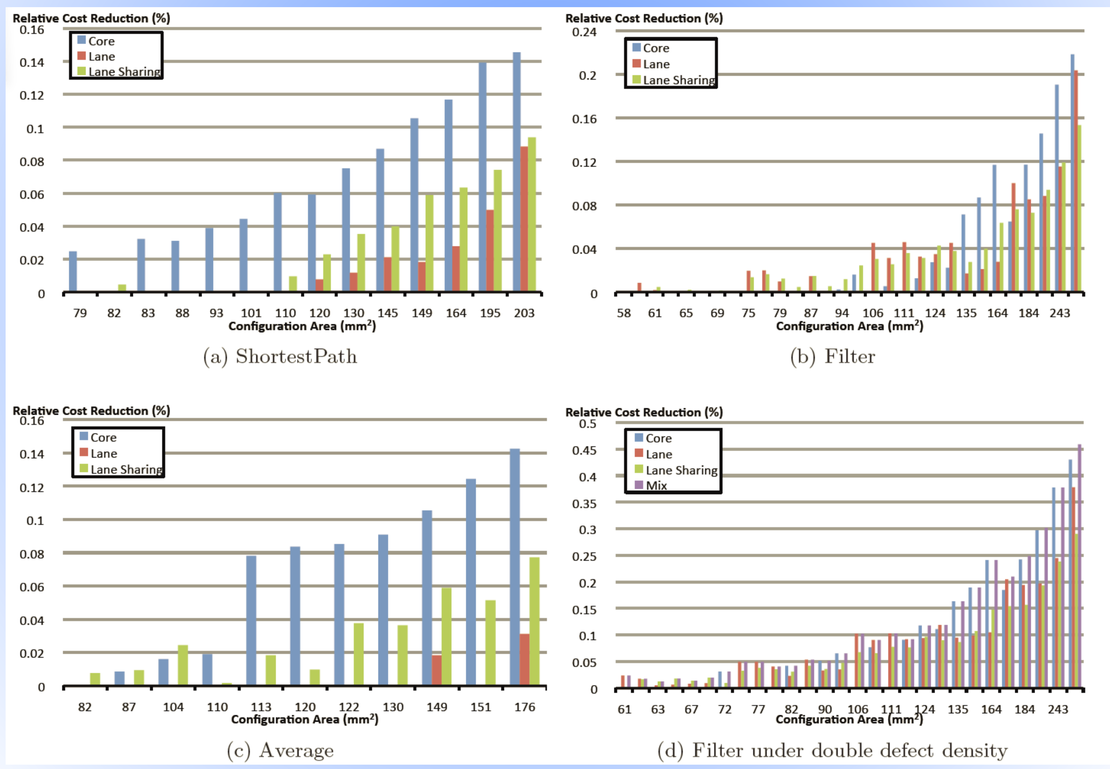
- Sparing becomes more beneficial as chip size increases
- Different sparing options: core, lane, shared lane



- As the number of cores increase, core sparing becomes more effective

Benchmarks

- MV5 (https://sites.google.com/site/mv5sim/)
- Optimize for a variety of benchmarks, or just one?
- Vary defect density?
- All up to designer



Simulated Annealing

- Need to search design space quickly
- Search a fraction of the design space
- Generate points on the cost-performance Pareto-optimal front

			% of Designs w/in X%		
Benchmark	% Expl	% Opt	0.1%	0.5%	1%
Average	1.7	91.6	91.6	91.6	91.6
FFT	30.7	98.0	98.0	98.0	98.0
Filter	2.2	85.1	85.2	97.4	99.9
HotSpot	1.6	78.6	78.8	79.3	84.8
KMeans	2.1	91.7	91.7	92.7	98.3
LU	2.3	72.2	80.8	81.7	98.3
MergeSort	1.7	97.8	97.8	98.5	98.5
ShortestPath	1.7	89.8	89.9	90.3	97.5
SVM	1.6	85.3	85.3	85.3	88.9

University of Virginia Engineering
McGill