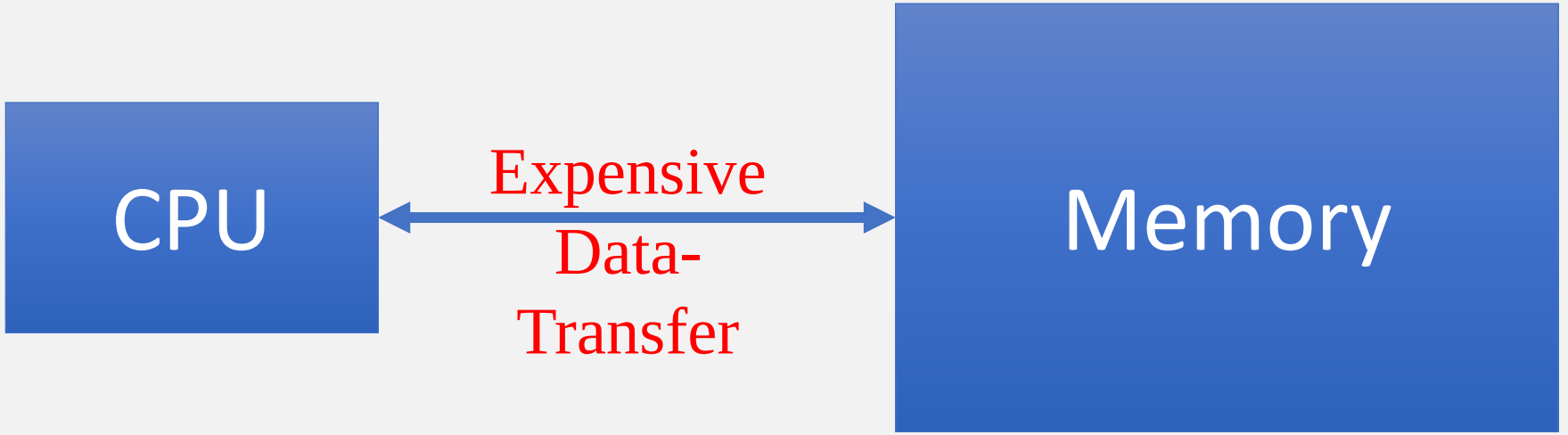
STREAM: Towards READ-based In-Memory Computing for Streaming based Data Processing

Muhammad Rashedul Haq Rashed∗, Sven Thijssen∗, Sumit Kumar Jha†, Fan Yao∗, and Rickard Ewetz∗

∗ECE, University of Central Florida, Orlando, USA

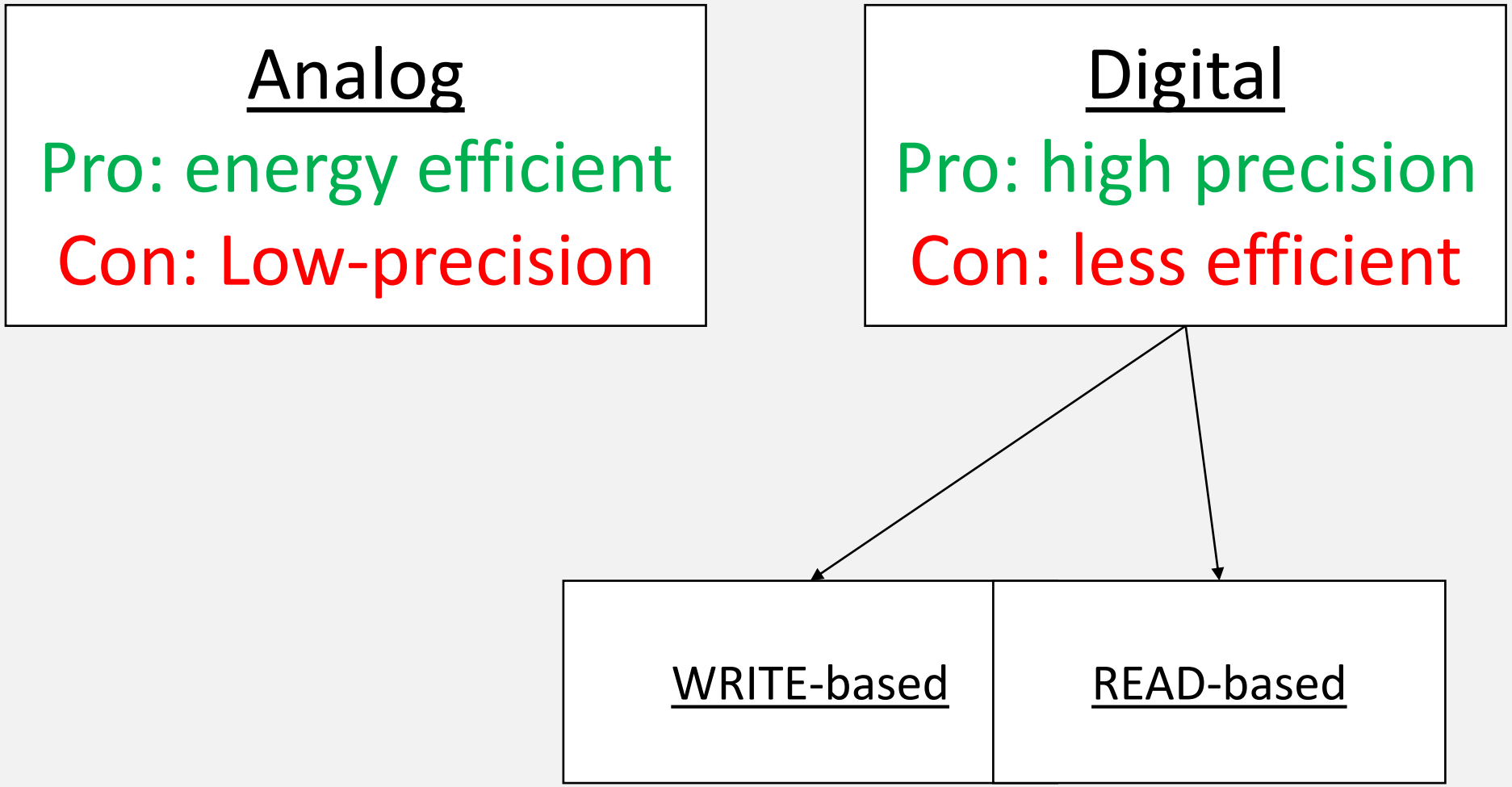
†CS, University of Texas at San Antonio, San Antonio, USA

Why In-memory Computing?

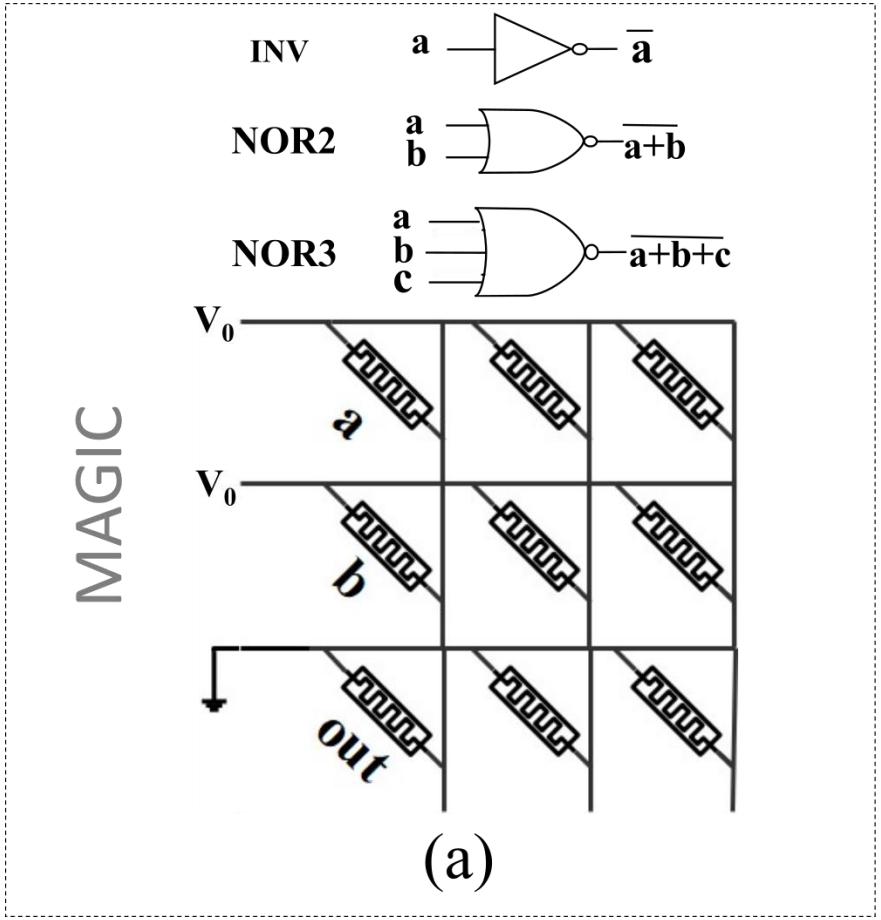


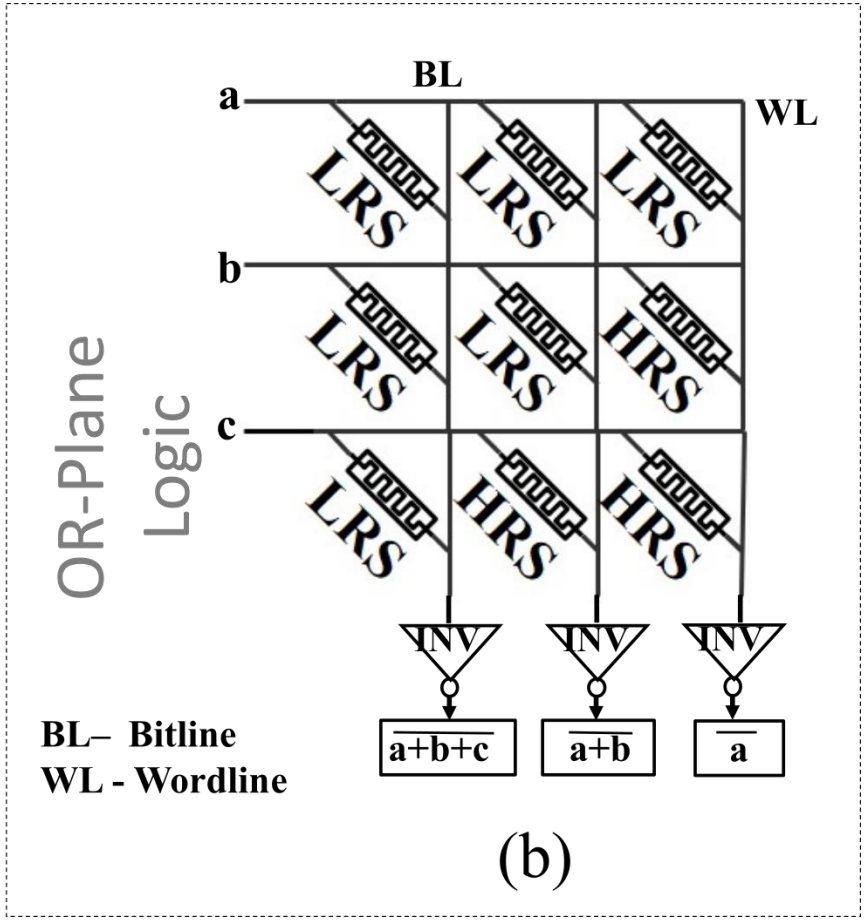
Why not Compute In-memory!

Types of In-memory Computing?



Digital in-memory computing

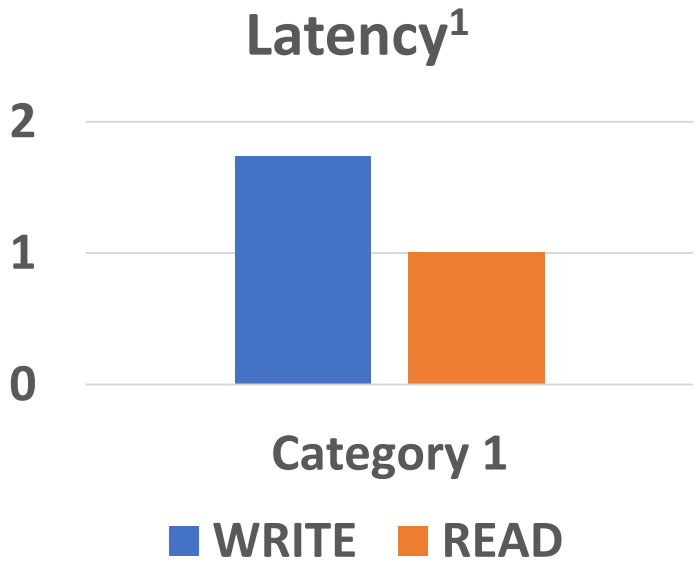




WRITE-based

READ-based

WRITE vs. READ Operation



NVM endurance

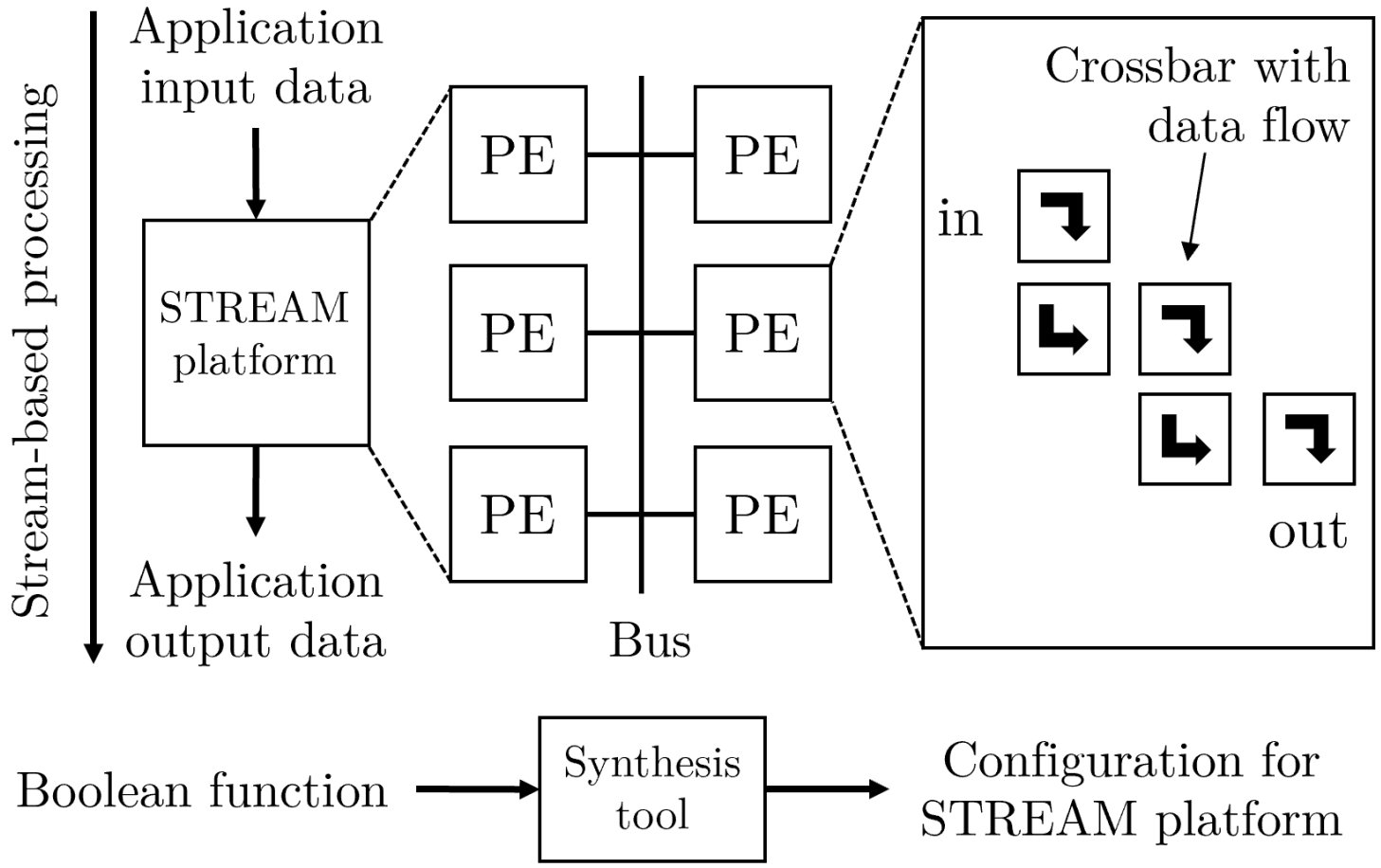
10^3 ~ 10^9

[1] L. Song, X. Qian, H. Li, and Y. Chen, “Pipelayer: A pipelined reram-based accelerator for deep learning,” in *HPCA*, pp. 541–552, IEEE, 2017

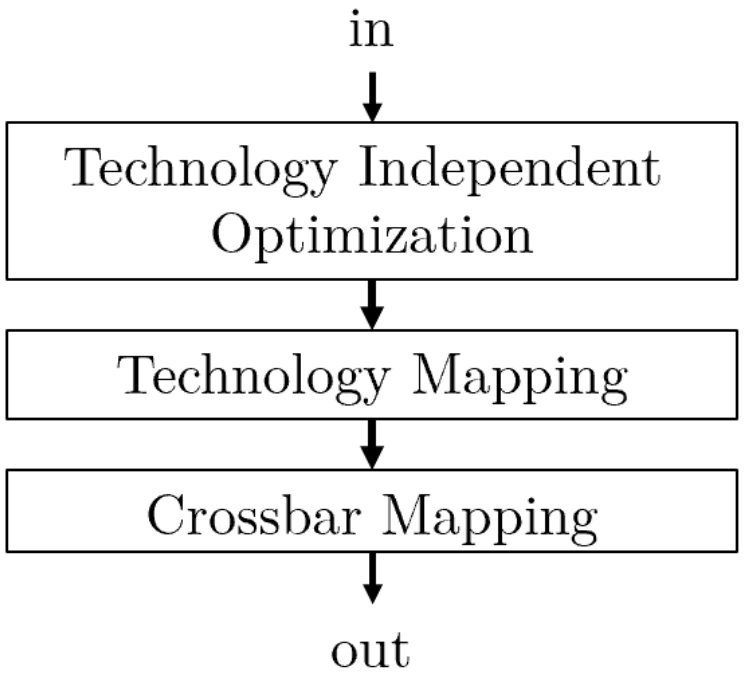
Motivation

Logic style	Work in	Initialization phase	Evaluation phase
Flow-based Comp.	[12]	WRITE	WRITE
Bitwise-In-Bulk	[11]	WRITE	WRITE
MAGIC	[5]	WRITE	WRITE
IMPLY	[6]	WRITE	WRITE
OR-plane logic	**(this work)**	WRITE	**READ**

THE STREAM FRAMEWORK



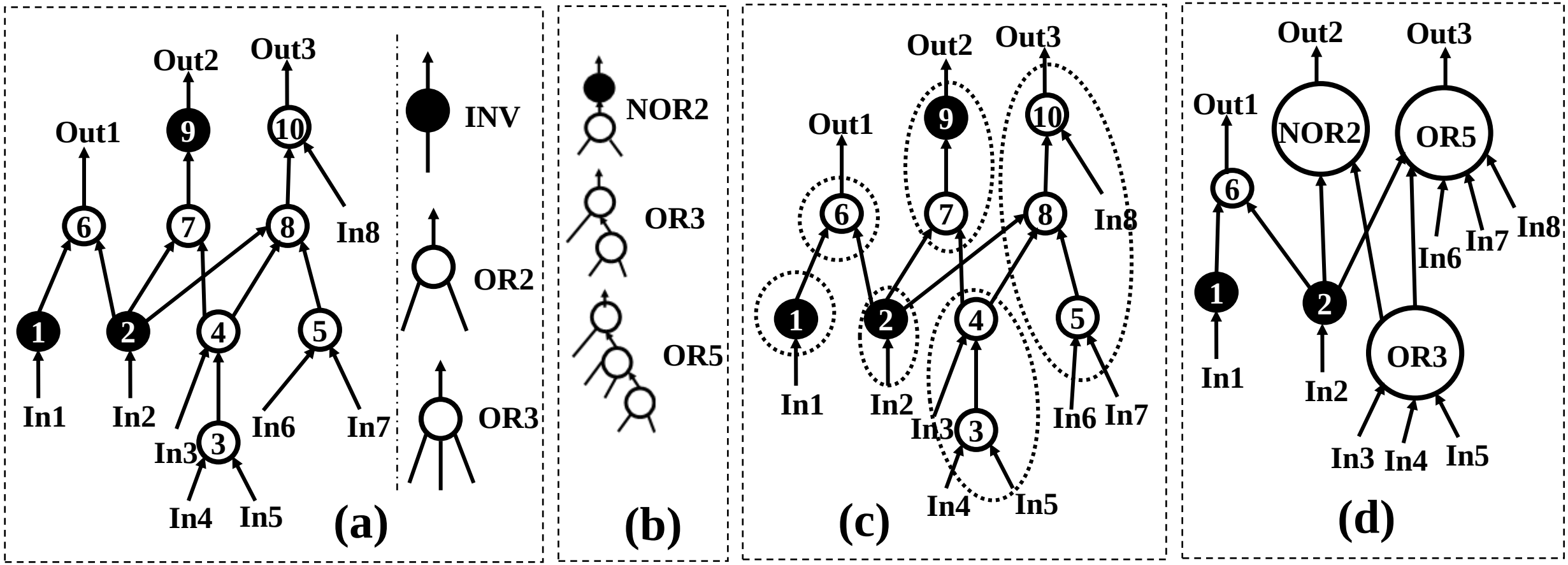
LOGIC SYNTHESIS FOR STREAM-BASED PEs



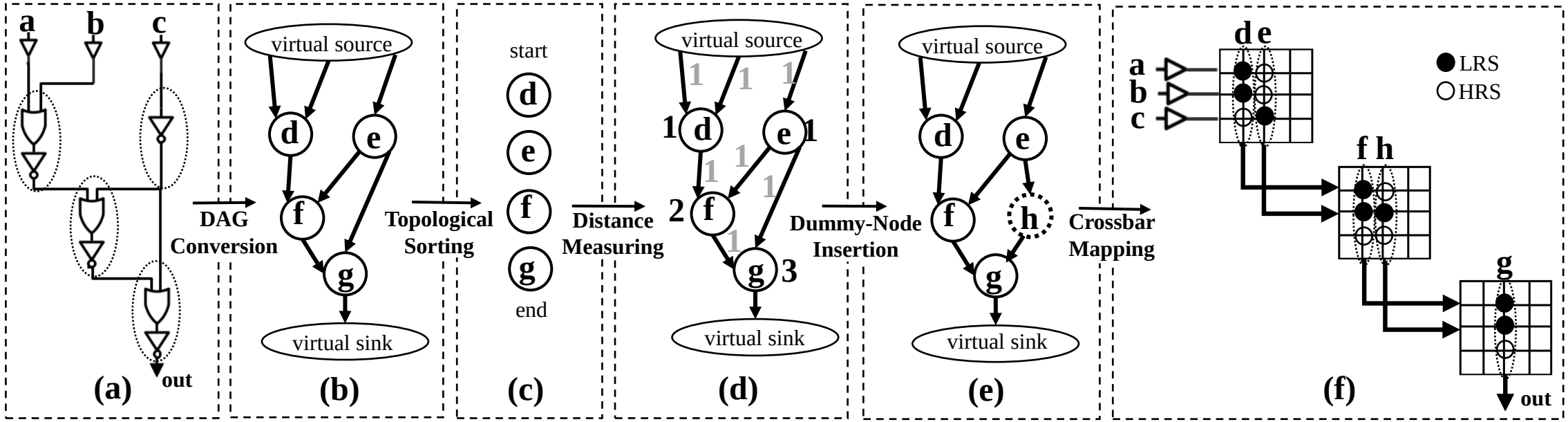
Netlist with low fan-in gates

Netlist with high fan-in in-memory compute kernels

Technology Mapping

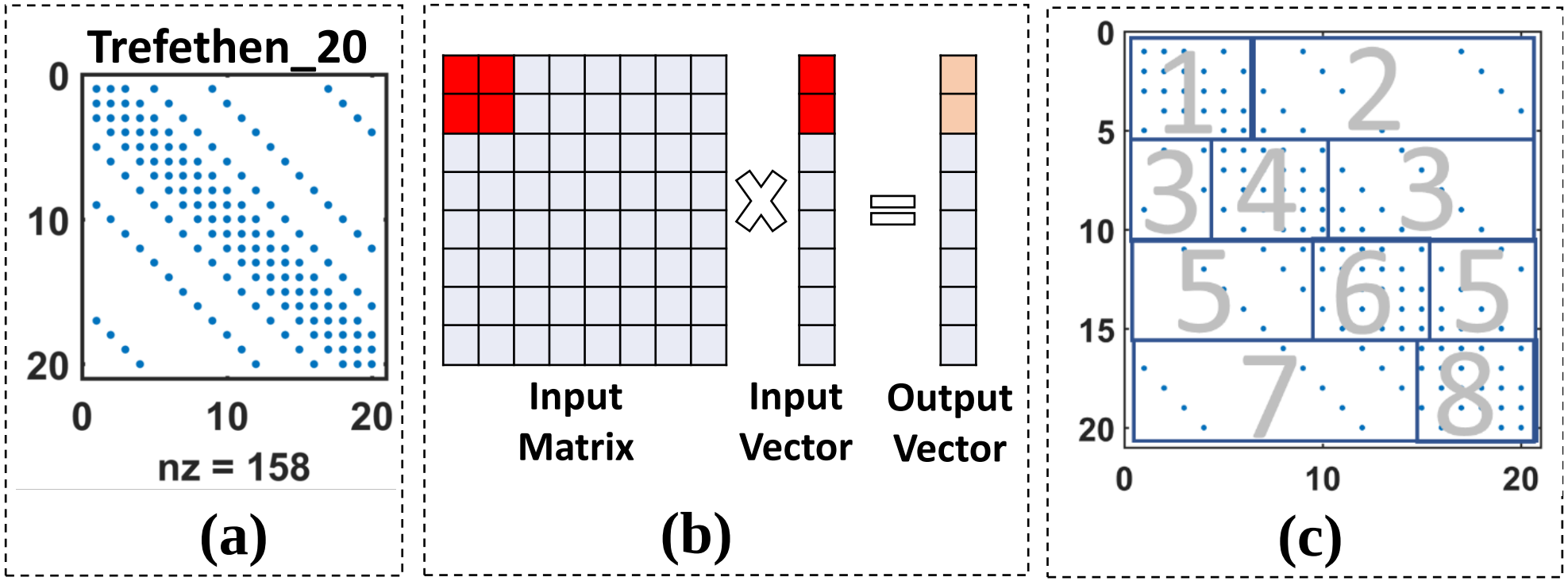


Crossbar Mapping

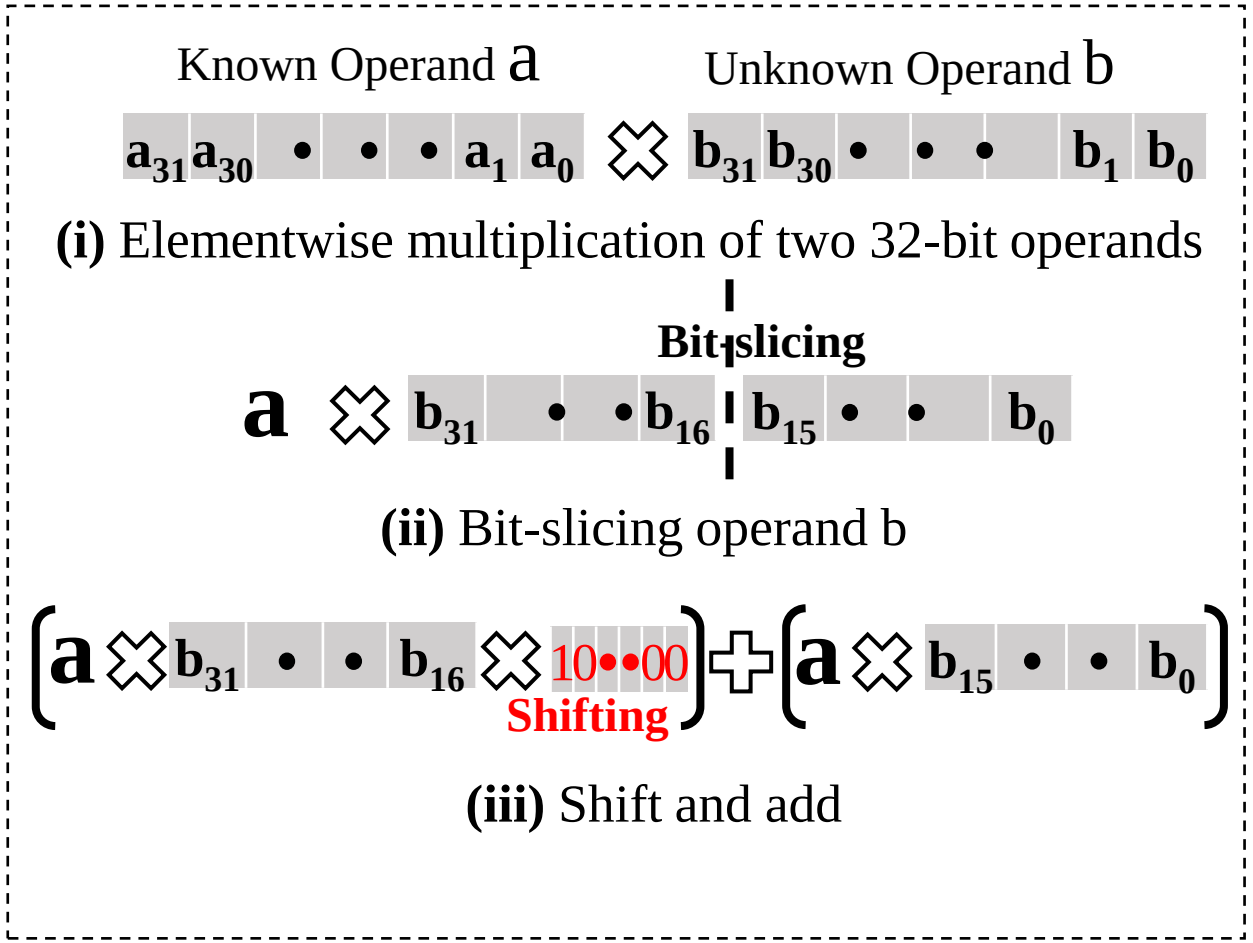


What if the logic is larger than the crossbar dimensions?

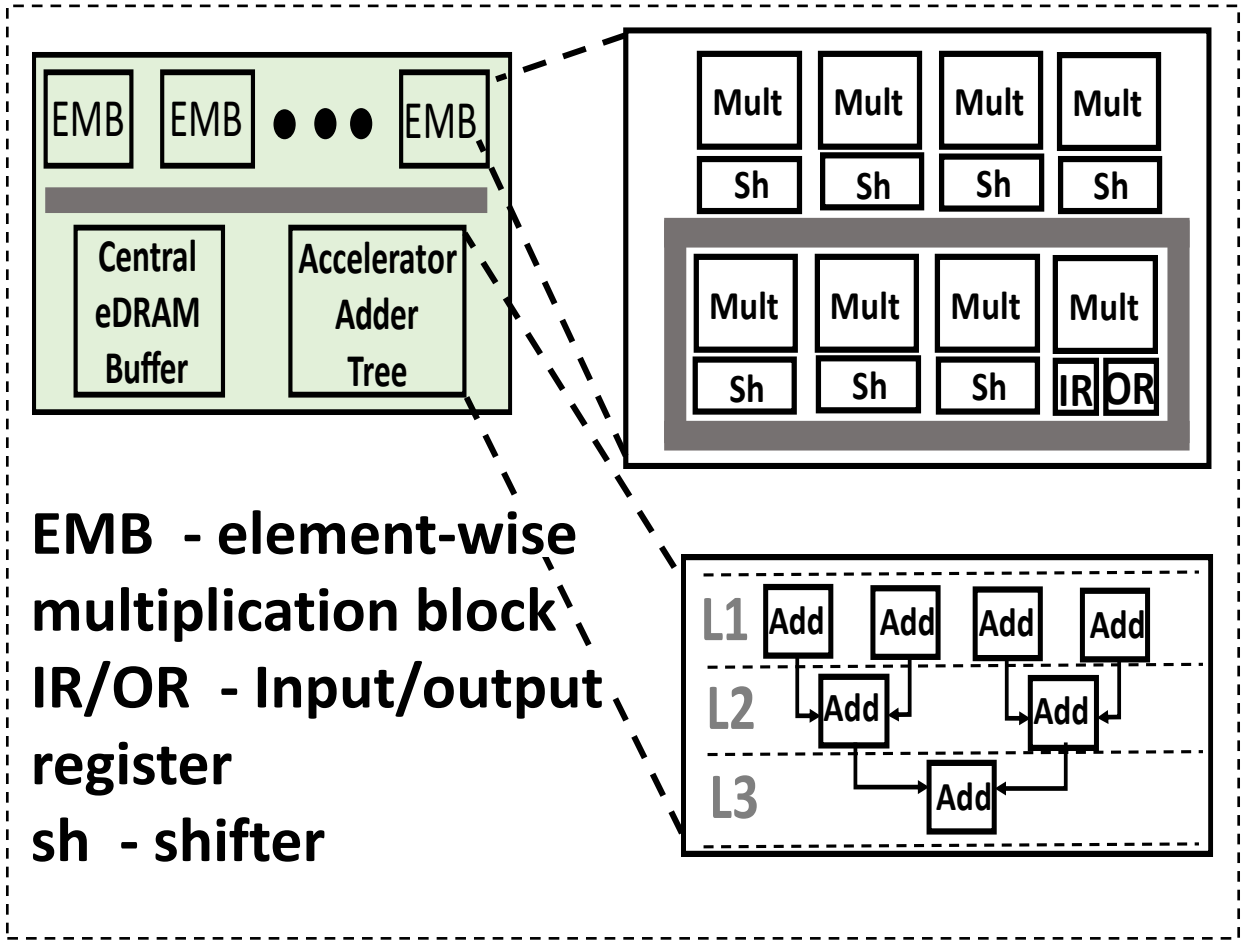
Spatial Partitioning



Bit-wise Partitioning



PE Architecture



Architectural overhead

TABLE III: Area-Power Cost of STREAM Components

Component	Parameter	Specs	Area	Power
Crossbar	dimension	128×128	25 μm^2	0.3 mW
Controller	# unit	1	400 μm^2	0.65 mW
Mult. (Total)	# crossbars	12	**0.005 mm^2**	**11.4 mW**
Shifter	# unit	1	60 μm^2	0.05 mW
IR	size	4 KB	4200 μm^2	2.48 mW
OR	size	512 B	1500 μm^2	0.46 mW
local bus	#wires	128	0.03 mm^2	2.33 mW
EMB (Total)	# Mult. # Shifter #IR/#OR	8 7 1/1	**0.077 mm^2**	**96.82 mW**
Adder Tree (Total)	# crossbars	198	**0.084 mm^2**	**188.1 mW**
eDRAM Buffer	size	128 KB	0.166 mm^2	41.4 mW
Bus	bandwidth	128-bits	15.7 mm^2	13 mW
PE (Total)	#EMBs #Adder tree #eDRAM Buffer	16 1 1	**17.18 mm^2**	**1791.67 mW**

Evaluation on ISCAS85 benchmarks

TABLE IV: Overview of ten ISCAS85 benchmarks.

Benchmark	Function	Inputs	Outputs
c432	Priority Decoder	36	7
c499	ECAT	41	32
c880	ALU and control	60	126
c1355	ECAT	41	32
c1908	ECAT	33	25
c2670	ALU and control	233	140
c3540	ALU and control	50	22
c5315	ALU and selector	178	123
c6288	16-bit multiplier	32	32
c7552	ALU and control	207	108

Evaluation on ISCAS85 benchmarks

TABLE V: Comparison of area, number of cycles, and power consumption for CONTRA and STREAM on ten benchmarks of the ISCAS85 benchmarks suite.

	CONTRA [19]			STREAM		
Benchmark	Area (μm^2)	Latency (μs)	Power (W)	Area (μm^2)	Latency (μs)	Power (W)
c432	601	39.18	2.35	13222	0.64	0.35
c499	601	68.33	4.10	17429	0.73	0.41
c880	601	64.26	3.85	17429	0.85	0.47
c1355	601	68.38	4.10	14424	0.59	0.33
c1908	601	74.74	4.48	16227	0.79	0.43
c2670	601	104.81	6.28	28848	0.88	0.54
c3540	601	181.89	10.90	28247	1.35	0.74
c5315	601	245.80	14.73	37863	0.97	0.62
c6288	601	401	24.04	105175	3.31	2.00
c7552	601	356	21.46	59499	1.5	0.96
Norm. avg.	**0.018**	1.000	1.000	1.000	**0.0072**	**0.071**

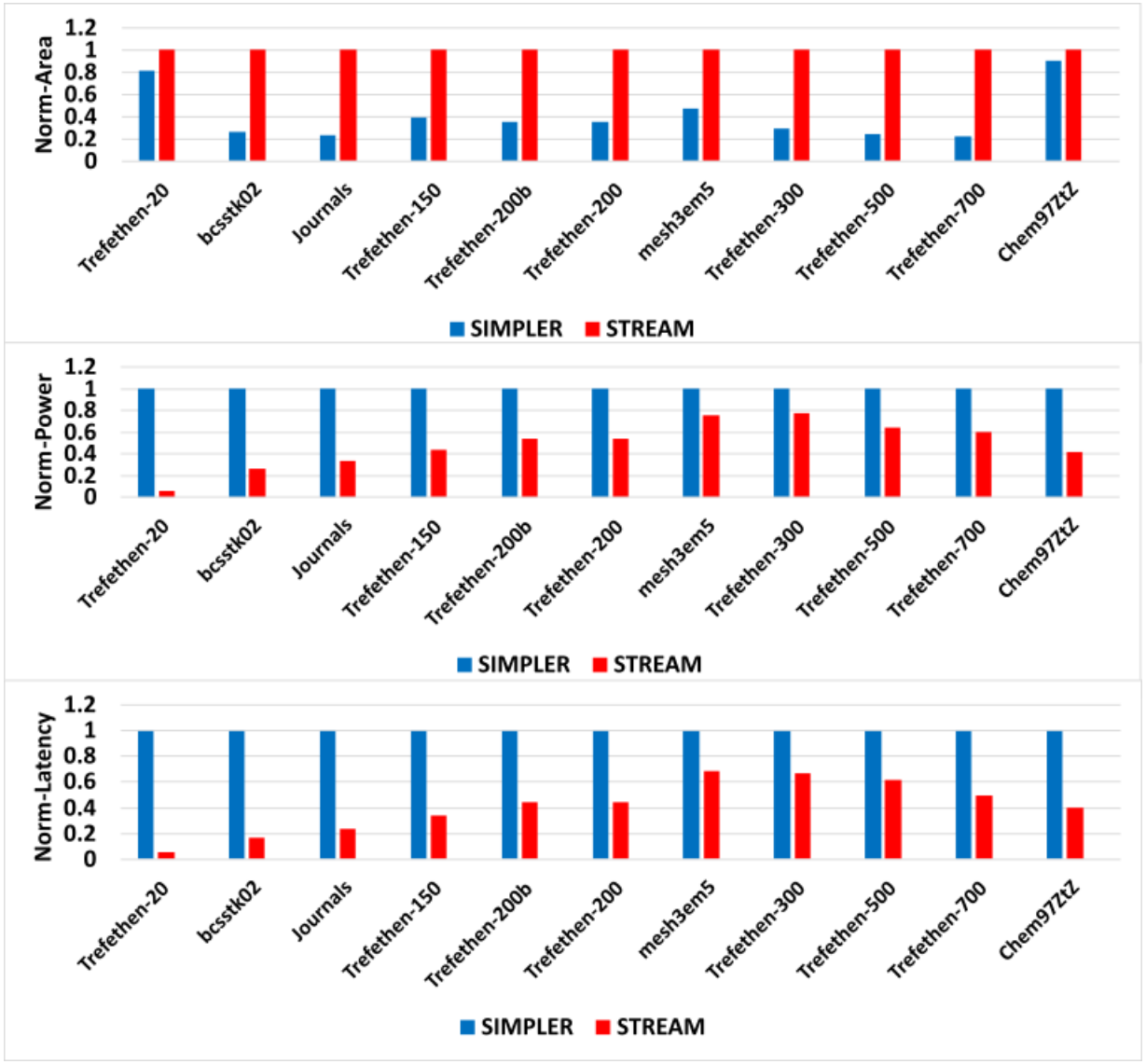
- **Latency improved by 139X**
- **Power consumption improved by 14X**
- **Area usage increases by 56X**

Evaluation on SuitSparse Matrix Applications

TABLE VI: Overview of eleven matrices of the SuitSparse Matrix Collection in terms of application type, matrix dimensions, and number of non-zero elements.

Applications	Systems	Matrix Dimensions	#Non-zeros
Trefethen-20	Combinatorial	20×20	158
mesh3em5	Structural	289×289	1377
Trefethen-150	Combinatorial	150×150	2040
Trefethen-200b	Combinatorial	199×199	2873
Trefethen-200	Combinatorial	200×200	2890
bcsstk02	Structural	66×66	4356
Trefethen-300	Combinatorial	300×300	4678
Chem97ZtZ	Statistical/Mathematical	2541×2541	7361
Trefethen-500	Combinatorial	500×500	8478
Journals	Undirected Weighted Graph	124×124	12068
Trefethen-700	Combinatorial	700×700	12654

Evaluation on SuitSparse Matrix Applications



- **2.2 X larger area**
- **2.0 X lower power**
- **2.4 X smaller latency**

Summary

Thank You

Contact us:

rashed09@knights.ucf.edu