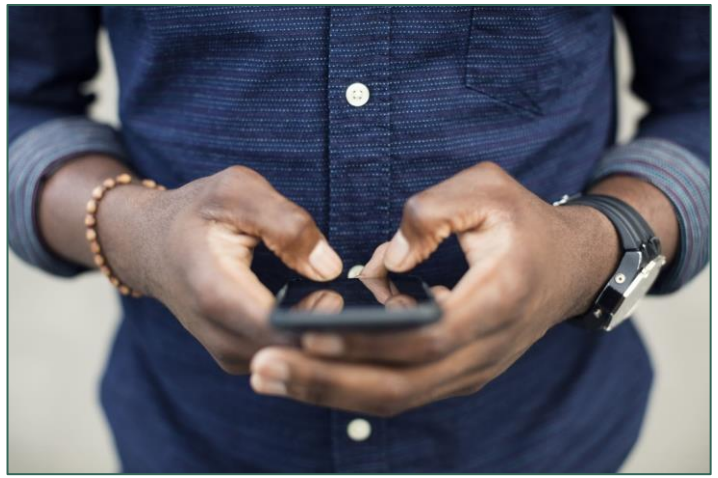
RELAX: A RECONFIGURABLE APPROXIMATE NETWORK ON CHIP

IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC-2021)
Singapore University of Technology and Design, Singapore
December 20-23, 2021

Richard Fenster and Sébastien Le Beux
(r_fenste@encs.concordia.ca, slebeux@encs.concordia.ca)

Department of Electrical and Computer Engineering
Concordia University
Montréal, Quebec
Canada

APPROXIMATE COMPUTING



Embedded Applications

Reduced power consumption

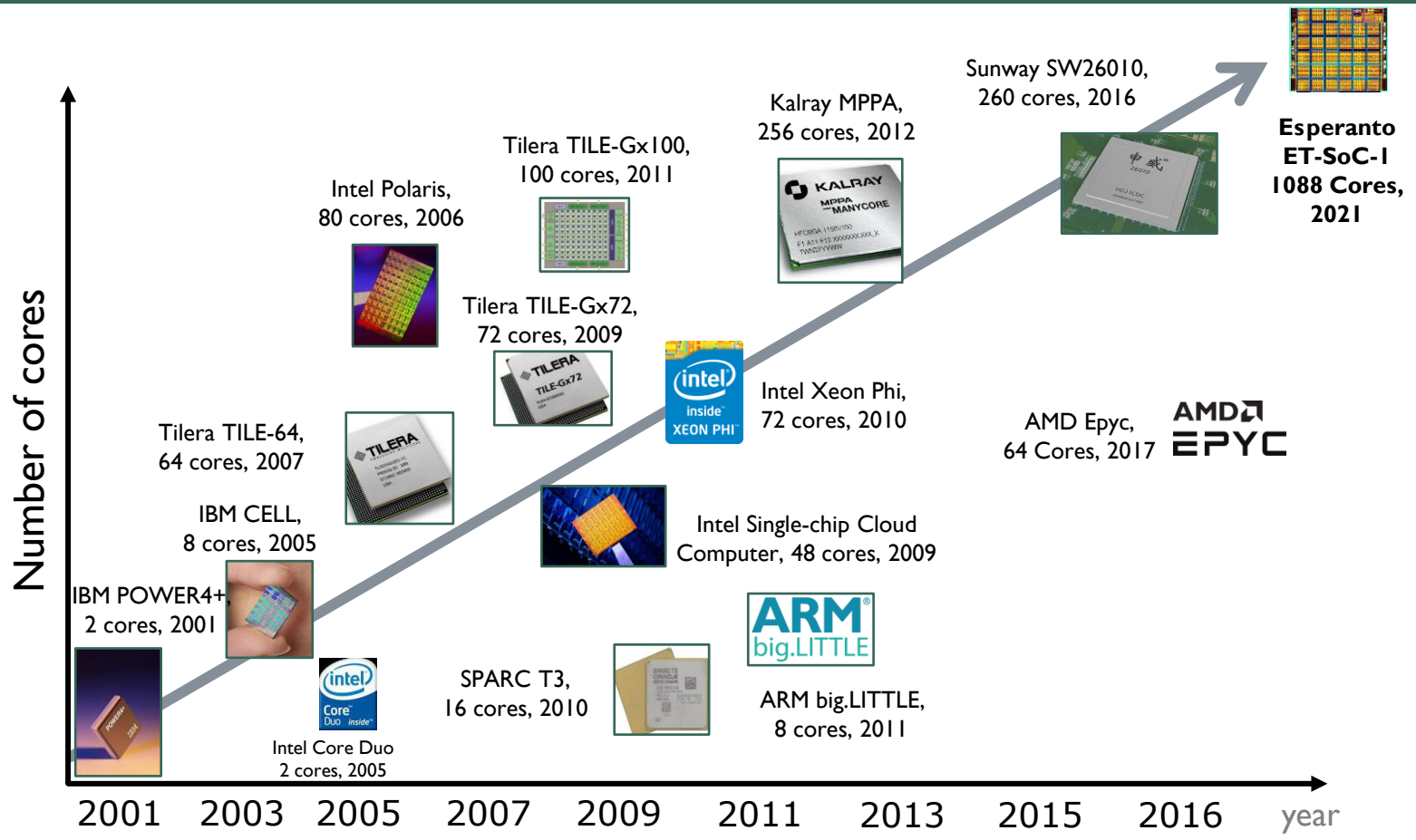
Machine Learning

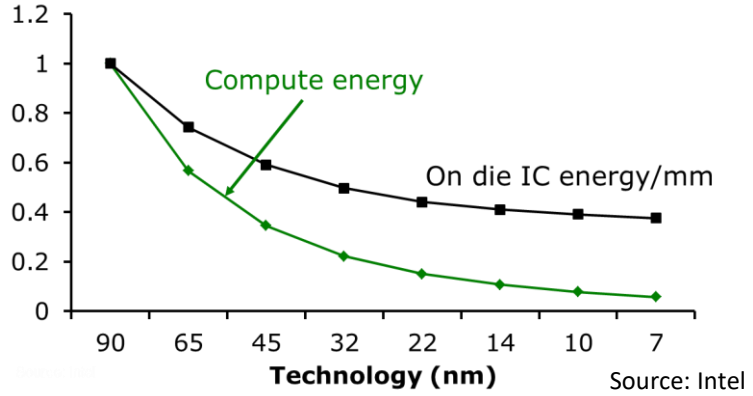
Increased throughput or reduced latency

Image Processing

Reduce storage requirements or processing time

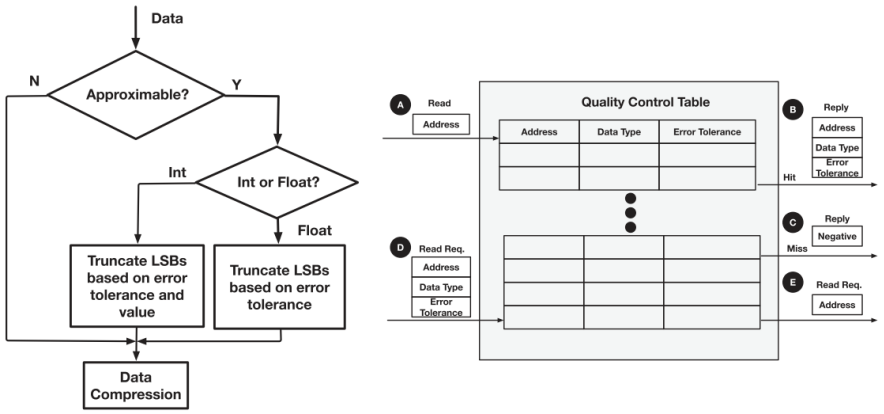
TOWARDS THE KILO-CORE PROCESSORS ERA



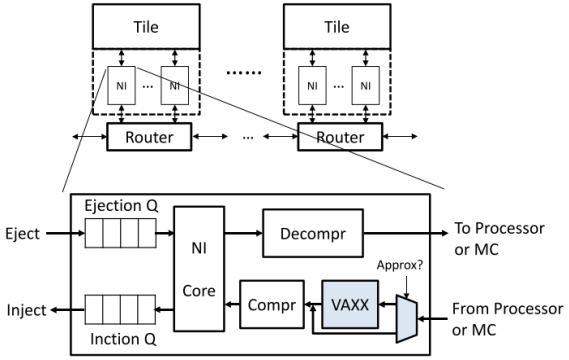


Challenge: it costs more energy to **move** data (e.g., 1pJ/bit/mm) than to **compute** with (e.g., 1 aJ/bit)

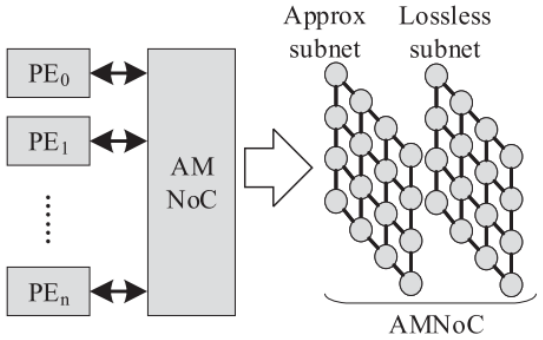
APPROXIMATE NETWORKS-ON-CHIP



Chen et al
IEEE TPDS, June 2020



Boyapati et al
ISCA, 2017



Wang et al
DATE, 2020

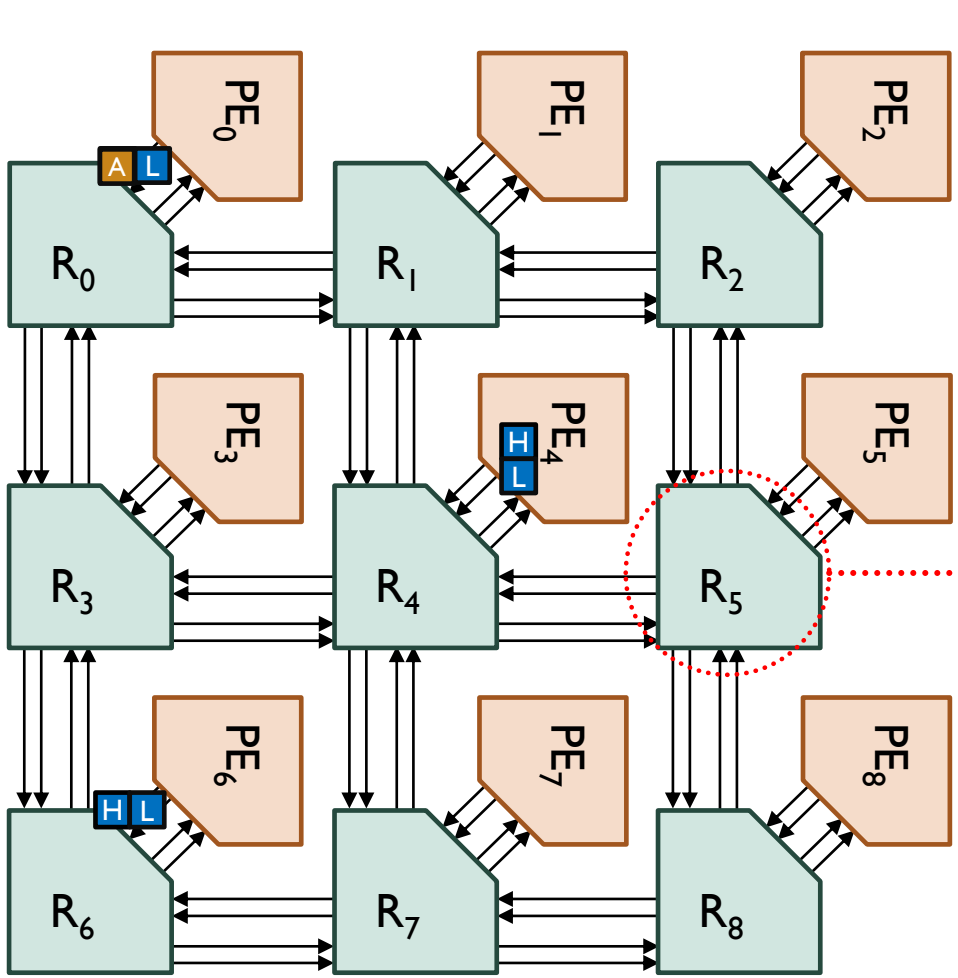
Name	Transmission Networks/Subnets	Dynamic Quality Control	Compression	Approx/Accurate concurrently
ACDC *(Chen et al)*	1	Yes	Yes	**No**
APPROX-NoC *(Boyapati et al)*	1	No	Yes	**No**
AMNoc *(Wang et al)*	2	No	No	**No**

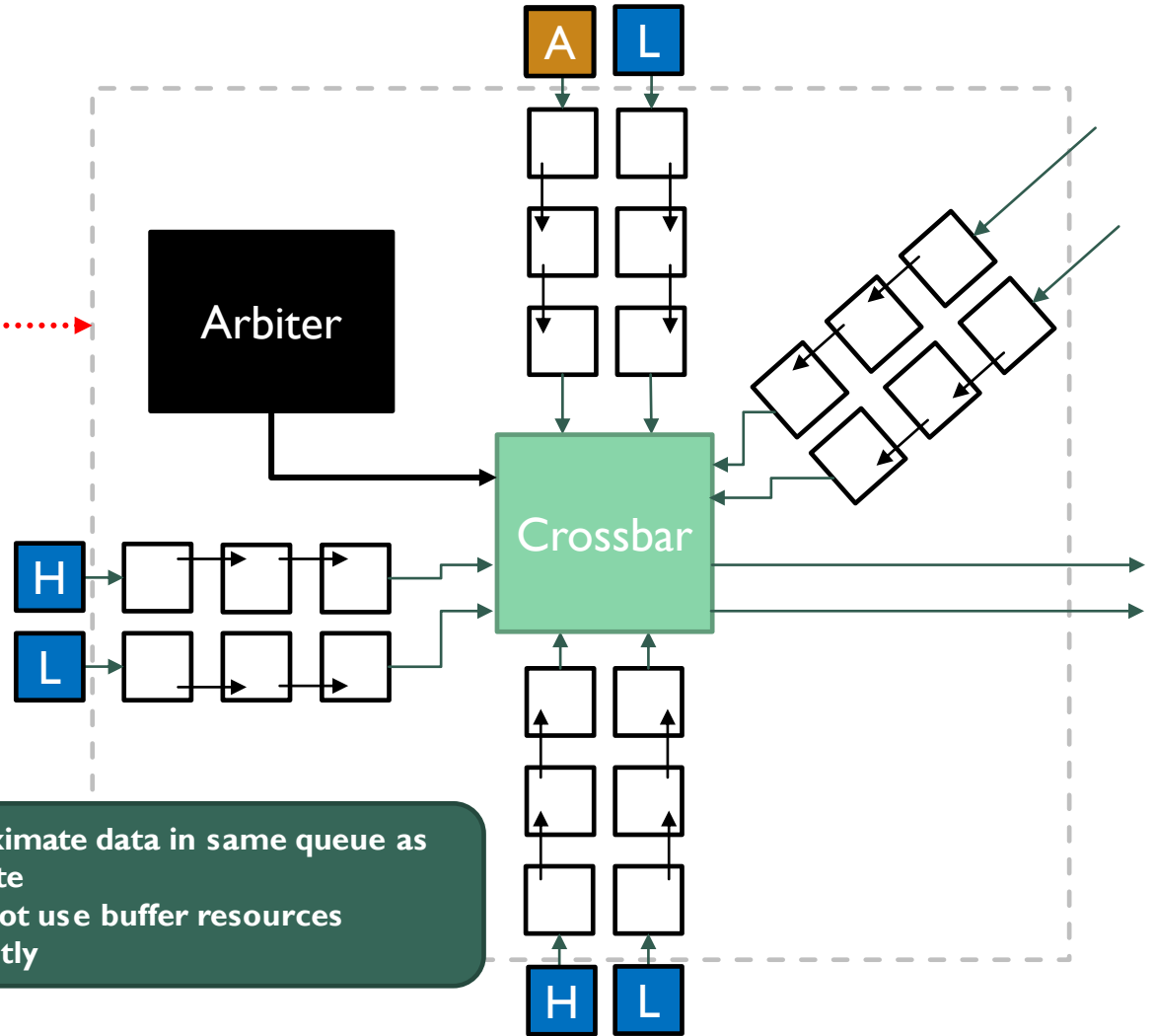
No approximation enabled NoC currently transmits accurate and approximate data concurrently using the same hardware resources!

OUR CONTRIBUTION

- **Key idea:** A reconfigurable approximate data enabled NoC
  - Two modes
    1. Accurate Only: All resources dedicated to transmitting accurate data
    2. Mixed: Hardware resources are partitioned to concurrently transmit accurate and approximate data
- Implementation:
  - RTL Description
  - Area, latency, power

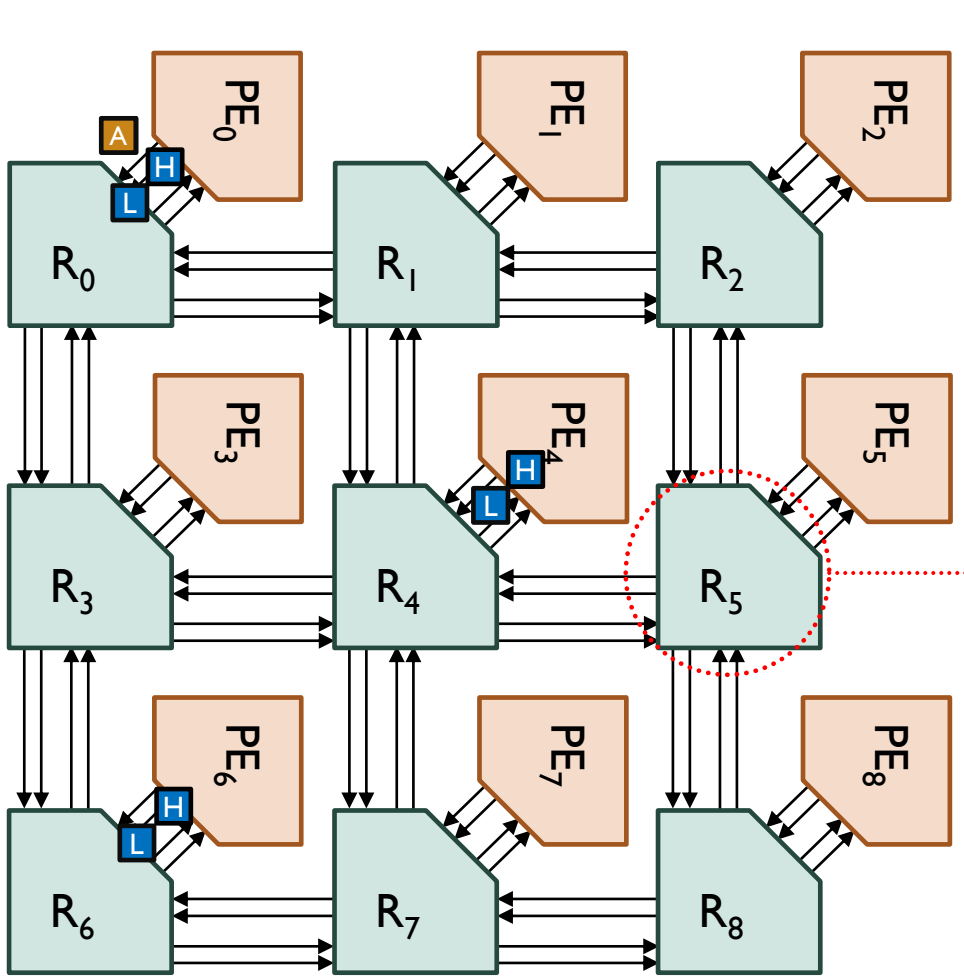
APPROXIMATE AND ACCURATE TRANSMISSION USING SAME RESOURCES

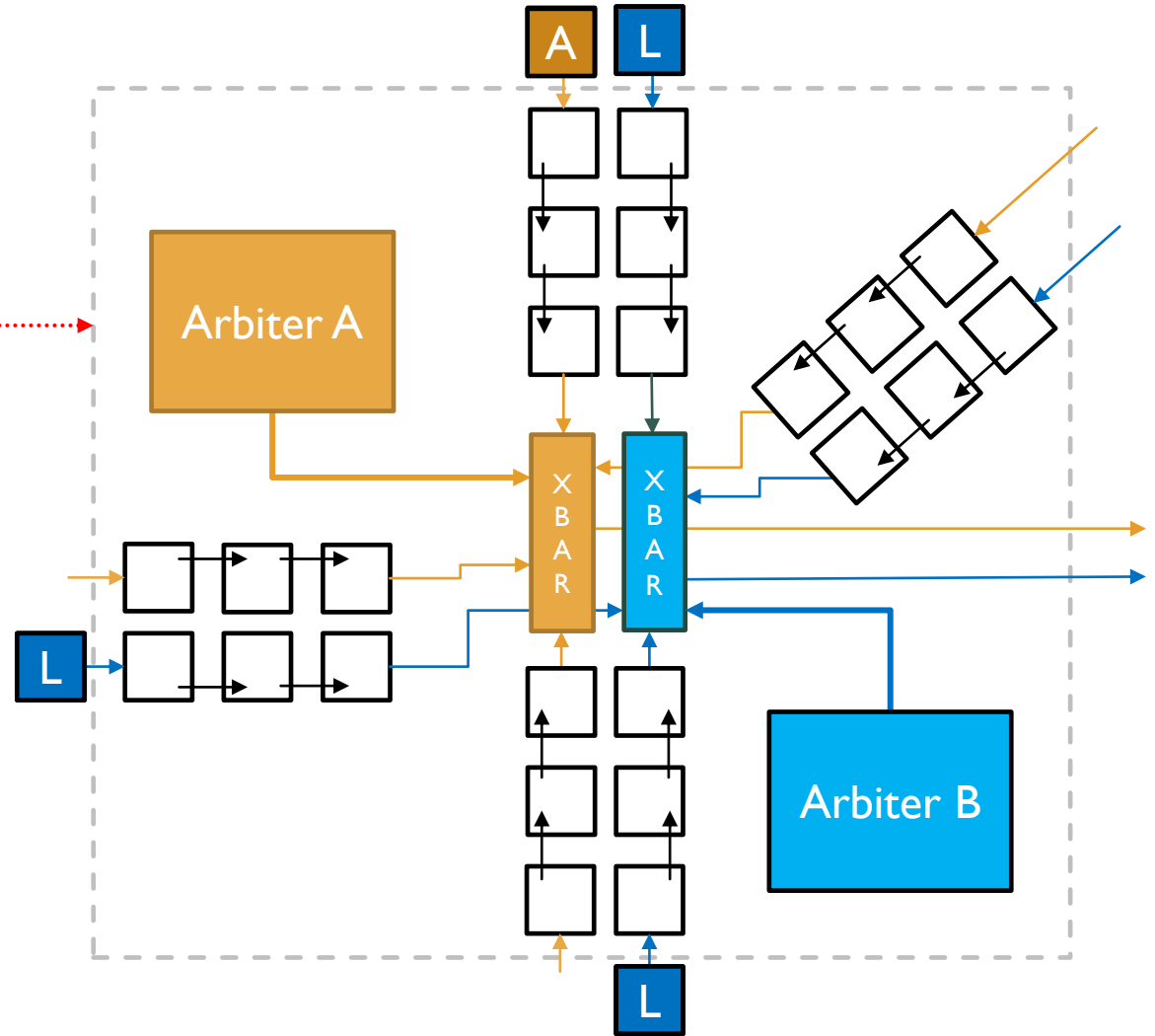




- Approximate data in same queue as accurate
- Does not use buffer resources efficiently

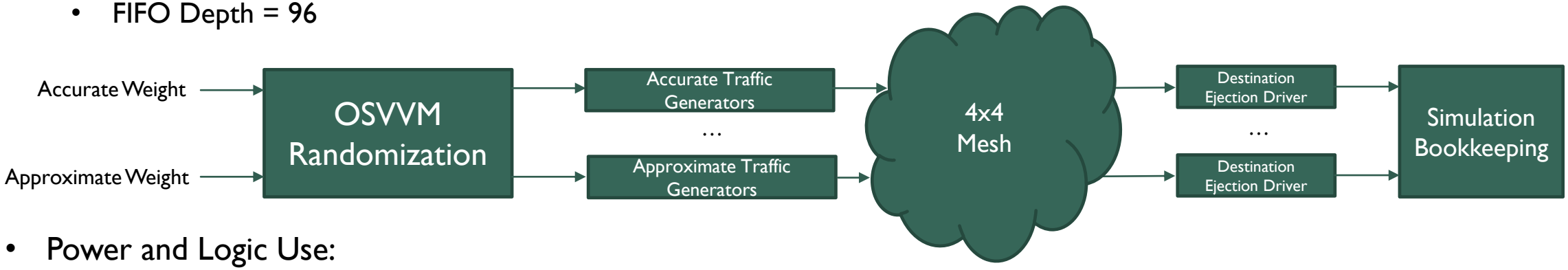
RELAX MIXED MODE





TESTING AND CHARACTERIZATION METHODOLOGY

- Generic RTL Implementation:
  - Packet Size 2N = 32
  - Written in VHDL-2008
  - X-Size, Y-Size = 4
  - Simulation Length = 262144
  - FIFO Depth = 96



- Power and Logic Use:
  - Synthesis for Xilinx Virtex Ultrascale XCVU080FFV1760
  - Assumed frequency of 100MHz
- Latency and Throughput
  - Simulated with Mentor Graphics Modelsim 10.7c
  - Round-robin, XY routing
  - Weighted-average distribution of generated accurate and approximate traffic

LOGIC UTILIZATION OF A SINGLE ROUTER (FPGA SYNTHESIS)

Name	Baseline		RELAX		Increased Cost	
	LUTs (x 10^3)	FFs (x 10^3)	LUTs (x 10^3)	FFs (x 10^3)	LUTs (%)	FFs (%)
RX	1.32	3.90	1.54	4.67	16.7	19.7
TX	1.32	3.90	1.54	4.67	16.7	19.7
Crossbar	0.112	0.006	0.162	0.013	44.6	117
Injection Port	1.32	3.90	1.71	4.69	29.5	20.2
Ejection Port	1.33	3.90	1.90	4.69	42.9	20.2
Routing Logic	10.7	31.2	12.5	37.4	16.8	19.8
Router (Combined)	**13.3**	**39.0**	**16.1**	**46.8**	**21.1**	**20.0**

Highest relative increase is when injecting or ejecting data

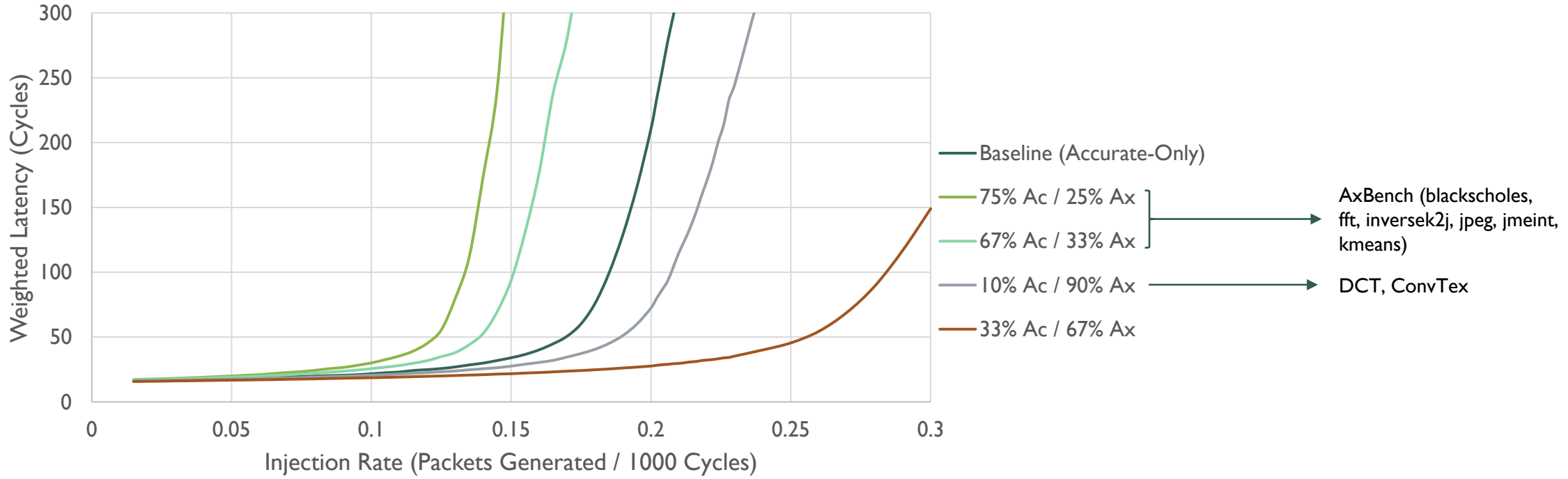
Overall implementation cost is roughly 20% more

POWER CONSUMPTION OF A SINGLE ROUTER (FPGA SYNTHESIS)

Channel Width (N)	**Static Power** (mW)		**Dynamic** (mW)		**Cost** (µJ/packet)		
	Baseline	RELAX	Baseline	RELAX	Baseline	RELAX Mixed Ac.	RELAX Mixed Ax.
8	910	911	97	119	10.1	10.3	5.2
16	910	911	151	169	10.6	10.8	5.4
32	912	913	237	258	11.5	11.7	5.9
64	915	915	425	416	13.4	13.3	6.7
128	919	920	755	780	16.7	17.0	8.5

NETWORK LATENCY

- Baseline is accurate-only mode
- All other results in mixed mode, transmitting Ac and Ax with weighted distributions
- Ac packets take 2 cycles, Ax packets take 1 cycle



RELAX: A RECONFIGURABLE APPROXIMATE NETWORK ON CHIP

- RELAX is the first network on chip that supports **concurrent** accurate and **approximate** data transmission by reconfiguring the same network resources
- Can reduce latency up to 44.2% from baseline with minimal (< 4%) power increase
- Future Work
  - Greater reconfigurability via profiling for different applications and traffic distributions
- Open source
  - Sources and more information available at available at: https://github.com/rf3nster/relax-noc

THANK YOU!

ANY QUESTIONS?