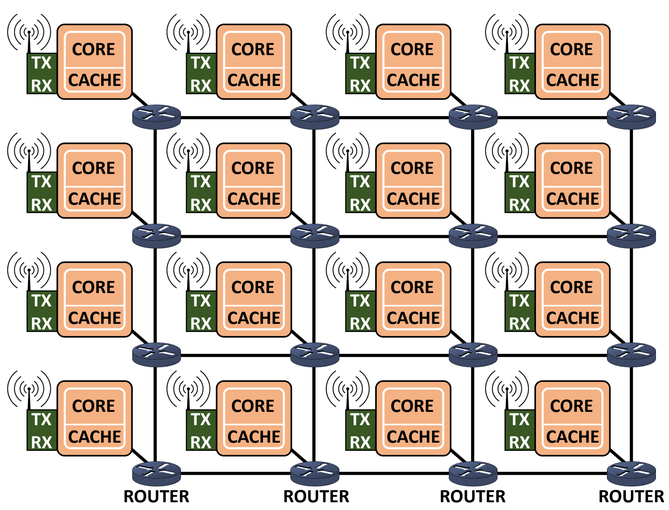
Millimeter Wave Wireless Network on Chip Using Deep Reinforcement Learning

Suraj Jog, Zikun Liu, Antonio Franques, Vimuth Fernando,
Haitham Hassanieh, Sergi Abadal, Josep Torrellas

Performance of Massive Multicores Bottlenecked by "Coherency Wall"



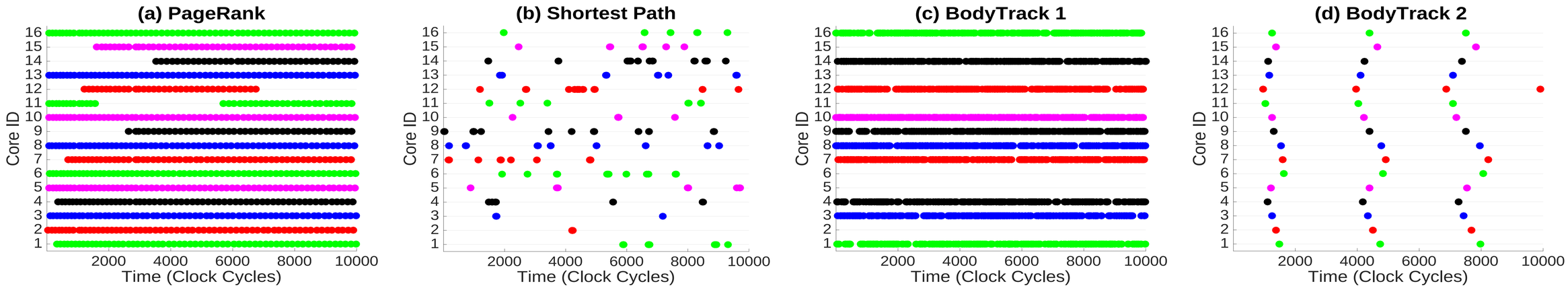
- Speedup gained by parallelism and multithreading is outweighed by the wired network's communication cost for keeping the caches coherent
- Augment wired interconnect with millimeter wave transceivers to form Wireless Network-on-Chip

Latency : Enables every core to reach every other core in just 1-hop

Broadcast : Local changes in the cache of a core can be instantaneously replicated at all other cores with single packet transmission

Traditional Medium Access Protocols Fall Short

Traffic patterns on wireless NoCs tend to be very dynamic and can change drastically across different cores, different time intervals and different applications.



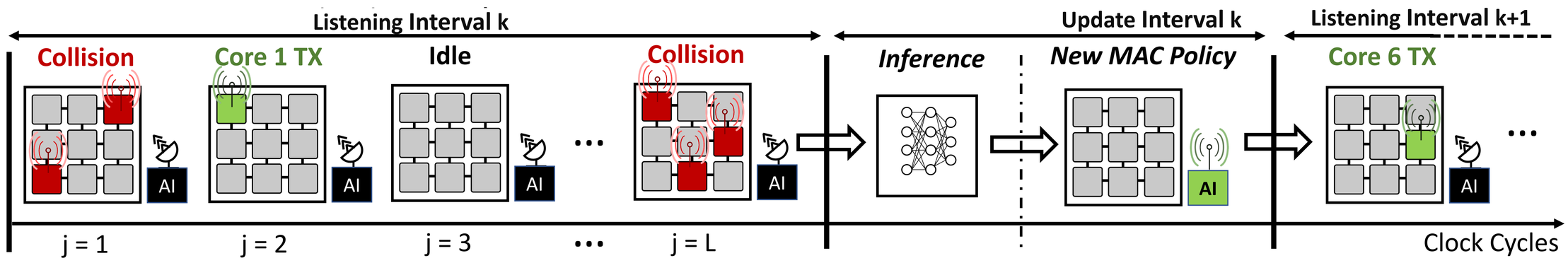
1. Building block functions like FFT, Graph Search, Sorting etc. repeatedly appear in applications producing unique periodic patterns
2. As number of cores increase, spatiotemporal correlations and structure in traffic increases

NeuMAC: Unifying Networking, Architecture and AI

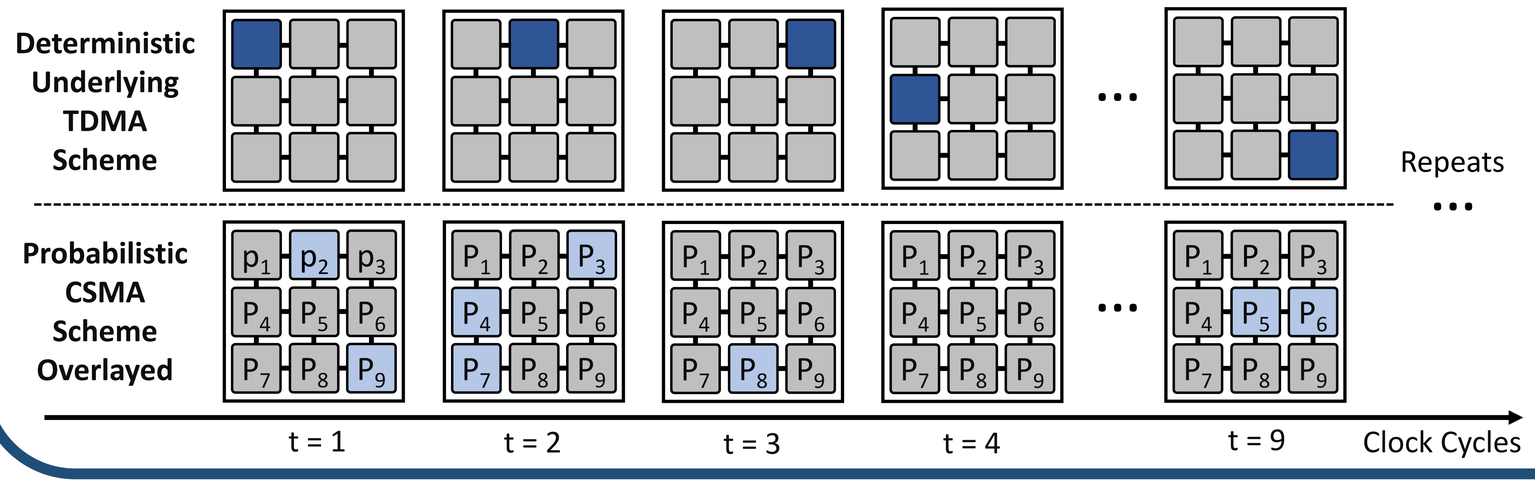
System Overview

NeuMAC's AI agent collects traffic statistics over wireless channel as state information for RL inference

NeuMAC's agent periodically generates new actions which dictate new MAC protocol for WNOC



2-Layer Protocol Design

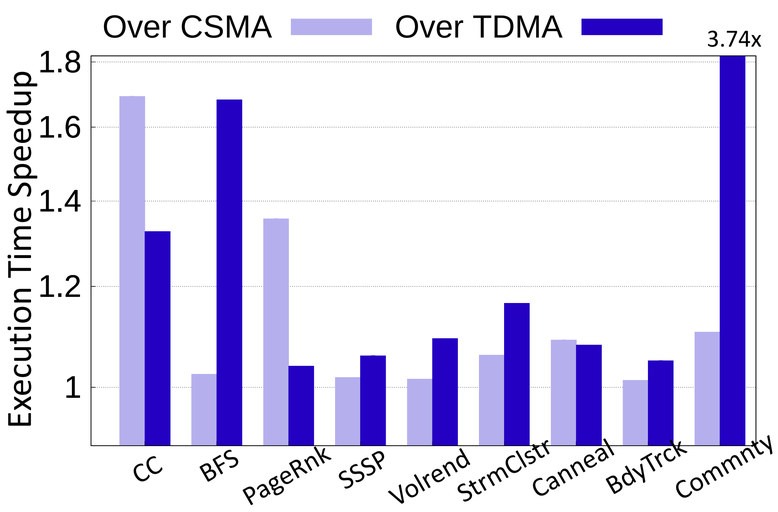


- NeuMAC assigns contention probabilities p_i to cores for opportunistic channel capture
- Can gracefully shift from TDMA to CSMA scheme while supporting all intermediate protocols, since $p_i = 0$ for all i, emulates pure TDMA and $p_i > 0$ mimics CSMA with varying degrees of aggressiveness
- Allows fine grained control to each core's action so as to generate highly optimized protocols++

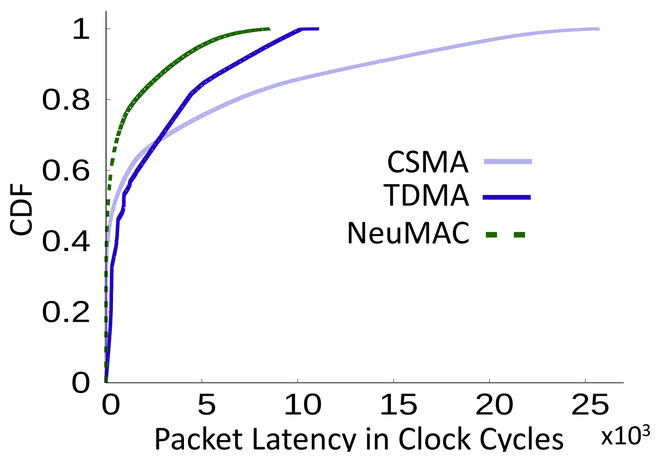
Preliminary Results

Apps	NeuMAC	Inf. Cap. Channel	% Achieved
CC	1.96x	2.06x	95%
BFS	6.53x	6.56x	99.5%
Pagerank	1.07x	1.11x	96.4%
SSSP	2.24x	2.25x	99.5%
Volrend	1.32x	1.33x	99.2%
StrmClstr	9.70x	9.77x	99.28%
Canneal	1.14x	1.15x	99.13%
Bodytrack	1.37x	1.38x	99.3%
Commnty	3.77x	3.82x	98.6%

Application Speedup over Purely Wired Network



Application Speedup over Baselines



Packet Latency CDF