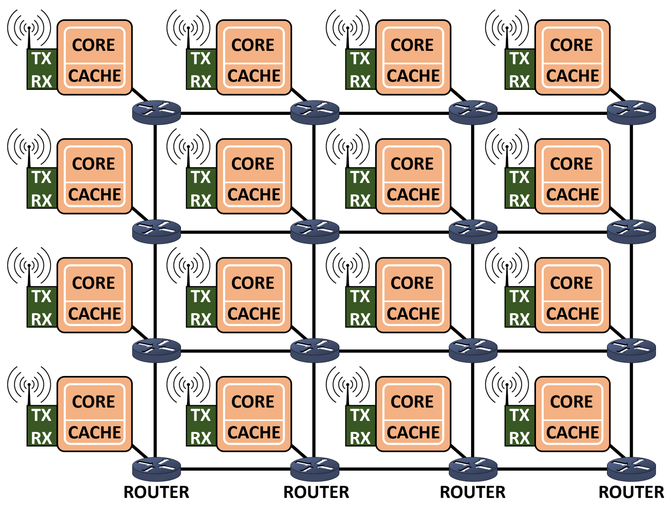
Millimeter Wave Wireless Network on Chip Using Deep Reinforcement Learning

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

ILLINOIS

Performance of Massive Network-on-Chip Multicores Bottlenecked by "Coherency Wall"

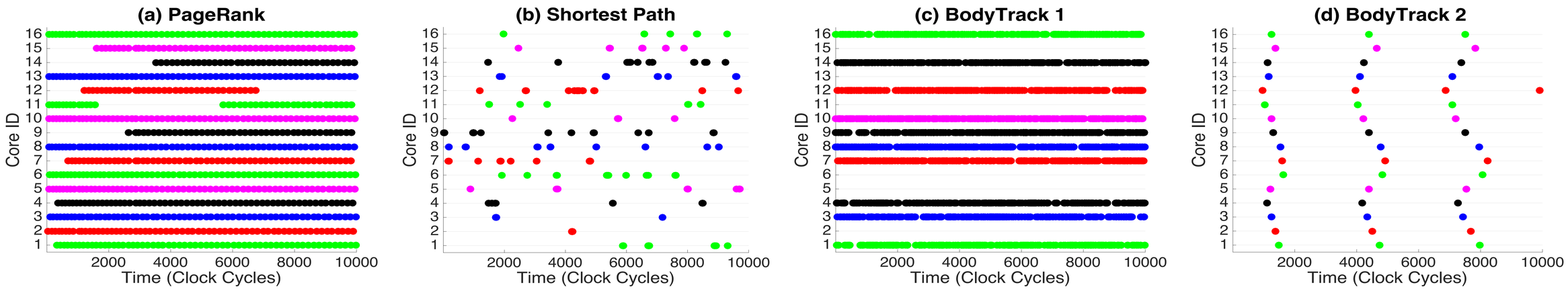


- Speedup gained by parallelism and multithreading is outweighed by the wired network's communication cost for keeping the caches coherent
- Augment wired interconnect with millimeter wave transceivers to form Wireless Network-on-Chip

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

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

Traffic patterns on wireless NoCs tend to be very dynamic and can change drastically across different cores, different time intervals and different applications. Hence, traditional medium access protocols fall short.



Can we learn new protocols that can adapt to accommodate drastically changing wireless traffic?

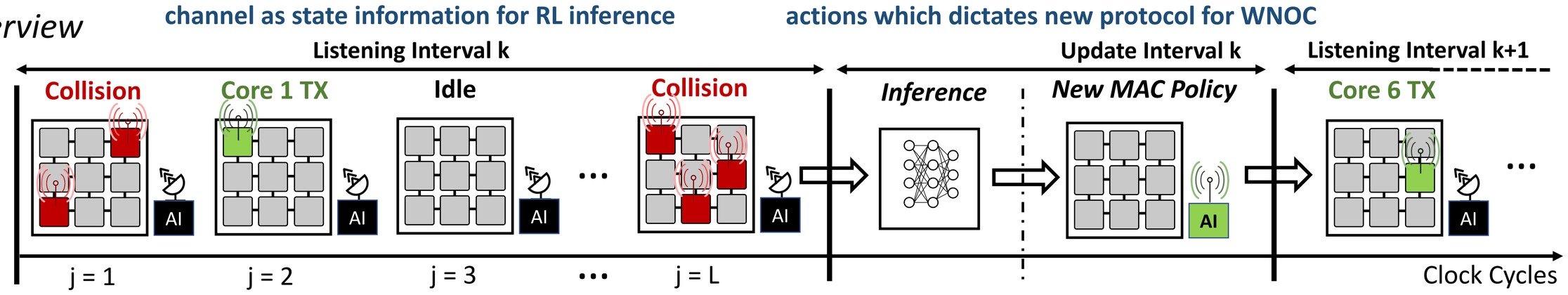
1. **Building blocks like FFT, Graph Search, Sorting etc. repeatedly appear in applications producing unique patterns**
2. **As number of cores increase, spatiotemporal correlations and structure in traffic increases**

NeuMAC: Bridging Networking, Architecture and AI

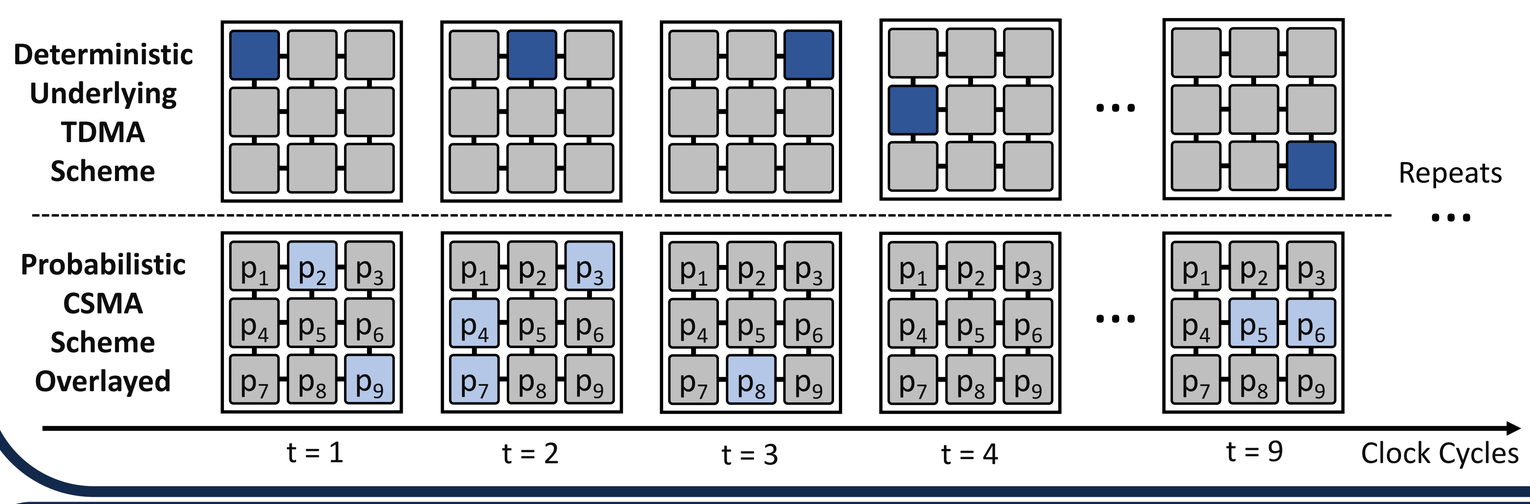
System Overview

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

NeuMAC agent periodically generates new actions which dictates new protocol for WNOC



2-Layer Protocol Design

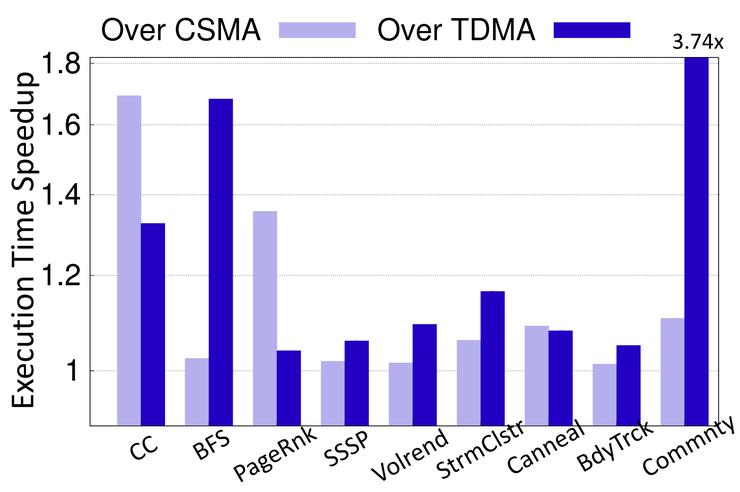


- NeuMAC assigns contention probabilities p_i to cores for opportunistic channel capture
- Can gracefully shift from TDMA to CSMA scheme while supporting all intermediate protocols, since $p_i = 0$ for all i, emulates pure TDMA and $p_i > 0$ mimics CSMA with varying degrees of aggressiveness
- Allows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

Results

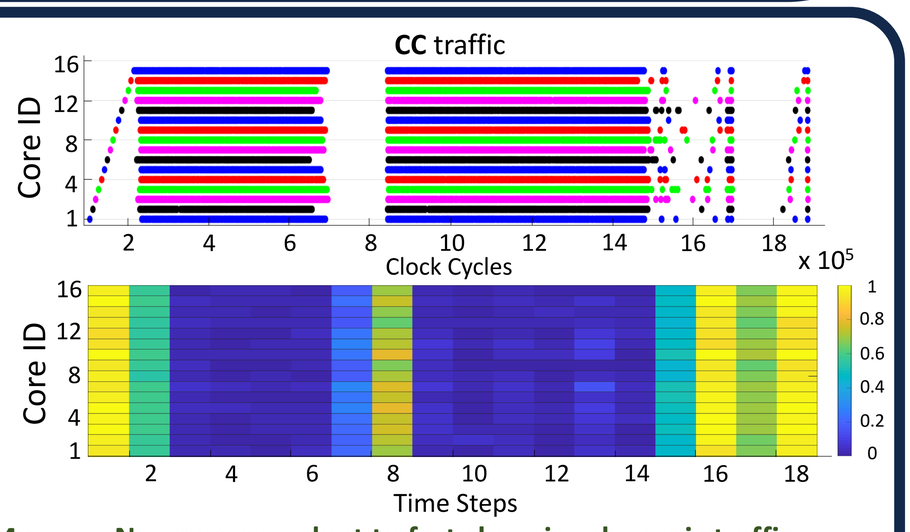
Apps	NeuMAC	Inf. Capacity Channel	% Achieved
CC	1.96x	2.06x	95%
BFS	6.53x	6.56x	99.5%
Pagerank	1.07x	1.11x	96.4%
SSSP	2.24x	2.25x	99.5%
Volrend	1.32x	1.33x	99.2%
StrmClstr	9.70x	9.77x	99.28%
Canneal	1.14x	1.15x	99.13%
Bodytrack	1.37x	1.38x	99.3%
Commnty	3.77x	3.82x	98.6%

(i) Application Speedup over Purely Wired NoC



Neumac improves execution time by up to 1.69x – 3.74x

(ii) Application Speedup over Baselines



Neumac can adapt to fast changing dynamic traffic

(iii) Qualitative Results: What did NeuMAC learn?