Userspace Networking in gem5

Johnson Umeike, Siddharth Agarwal*, Nikita Lazarev†, Mohammad Alian

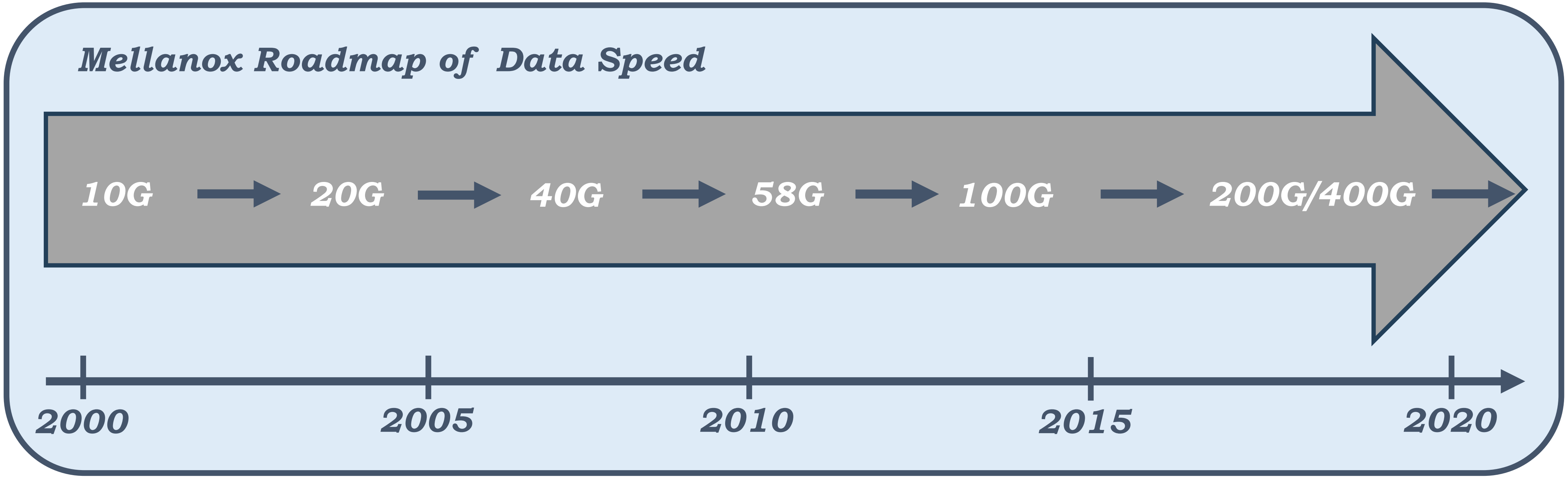
THE UNIVERSITY OF KANSAS

* UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

† MIT CSAIL

Network Bandwidth is Growing!

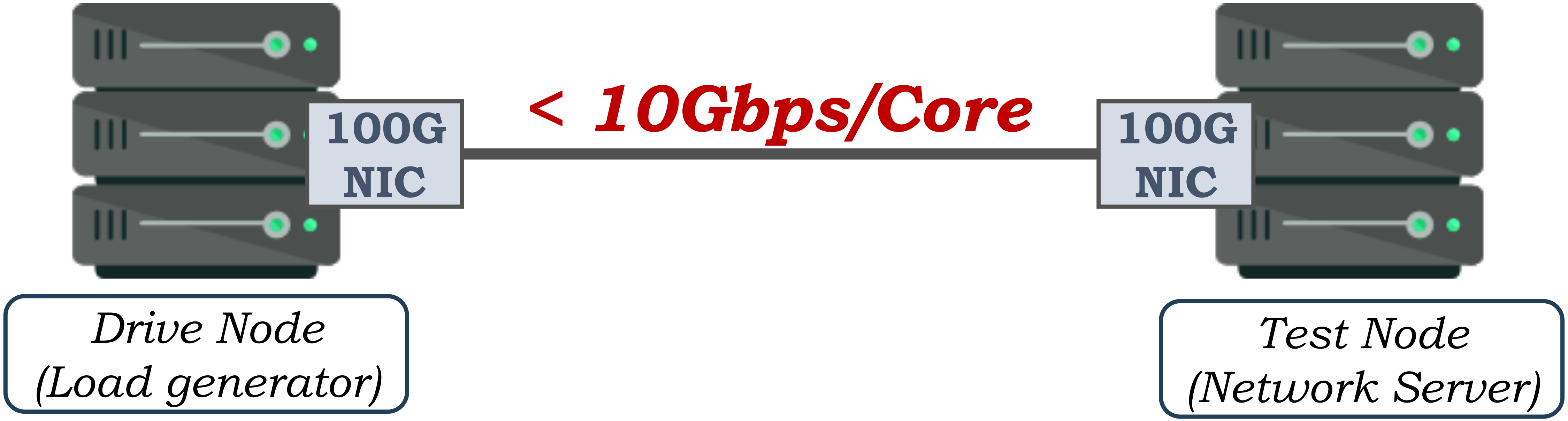
- Network bandwidth grows at **20% every year!**
- Tbps Ethernet is on the horizon.



There is a need to re-think the hardware and software stack of datacenter servers in the era to terabit per second networking!

Architecture Simulation

- Computer architects use simulators for early design explorations.



Limitations of Current Networked System Simulation in gem5

- gem5 uses kernel space networking
- gem5 lacks a hardware load generator model
- There is no benchmarking methodology for networked system evaluation in gem5

Contribution

In this work,

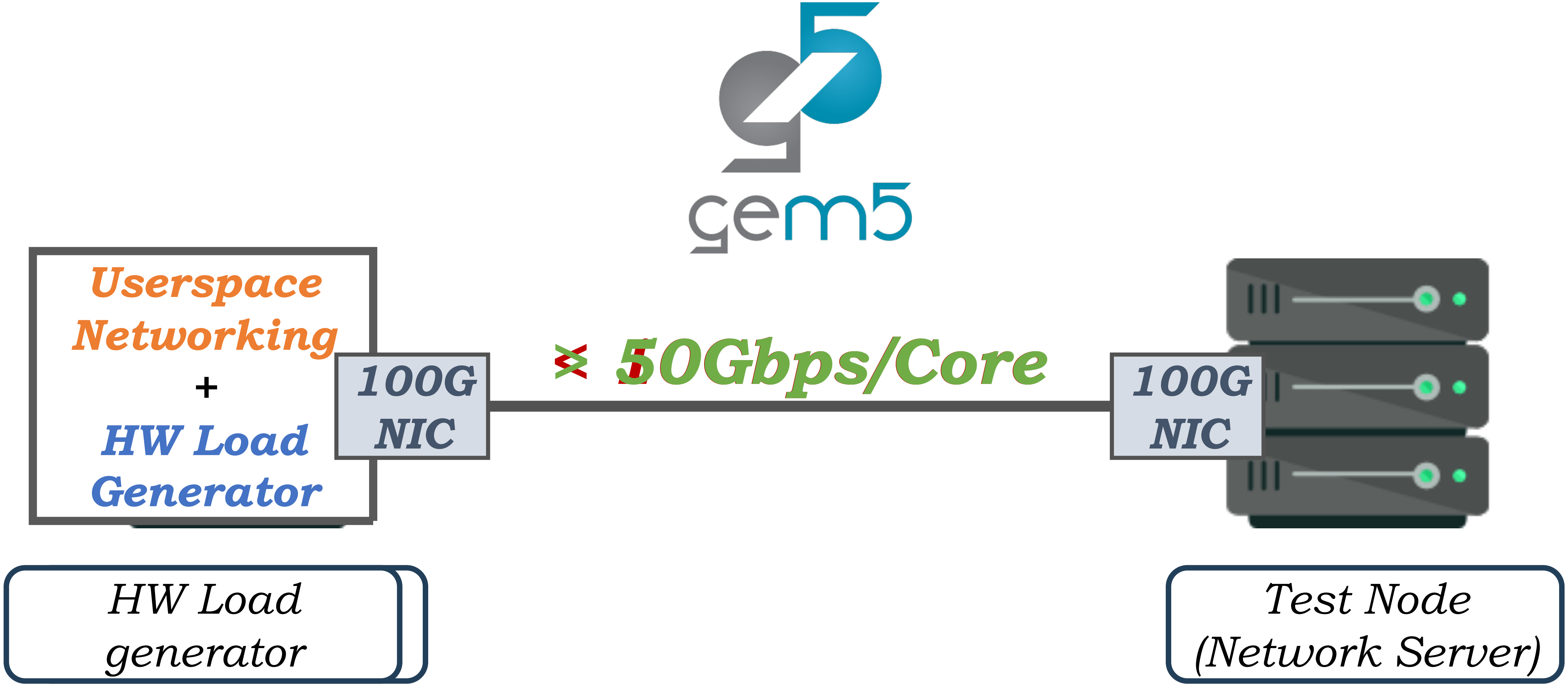
1 Extended gem5's to support **userspace networking**

2 Implemented a **hardware load generator** model

3 Streamlined **network subsystem evaluation** in gem5

- Introduced a suite of six **network-intensive applications**
- Enhancing gem5 statistics to report **causes of packet drops**

Result of Enabling Userspace Networking



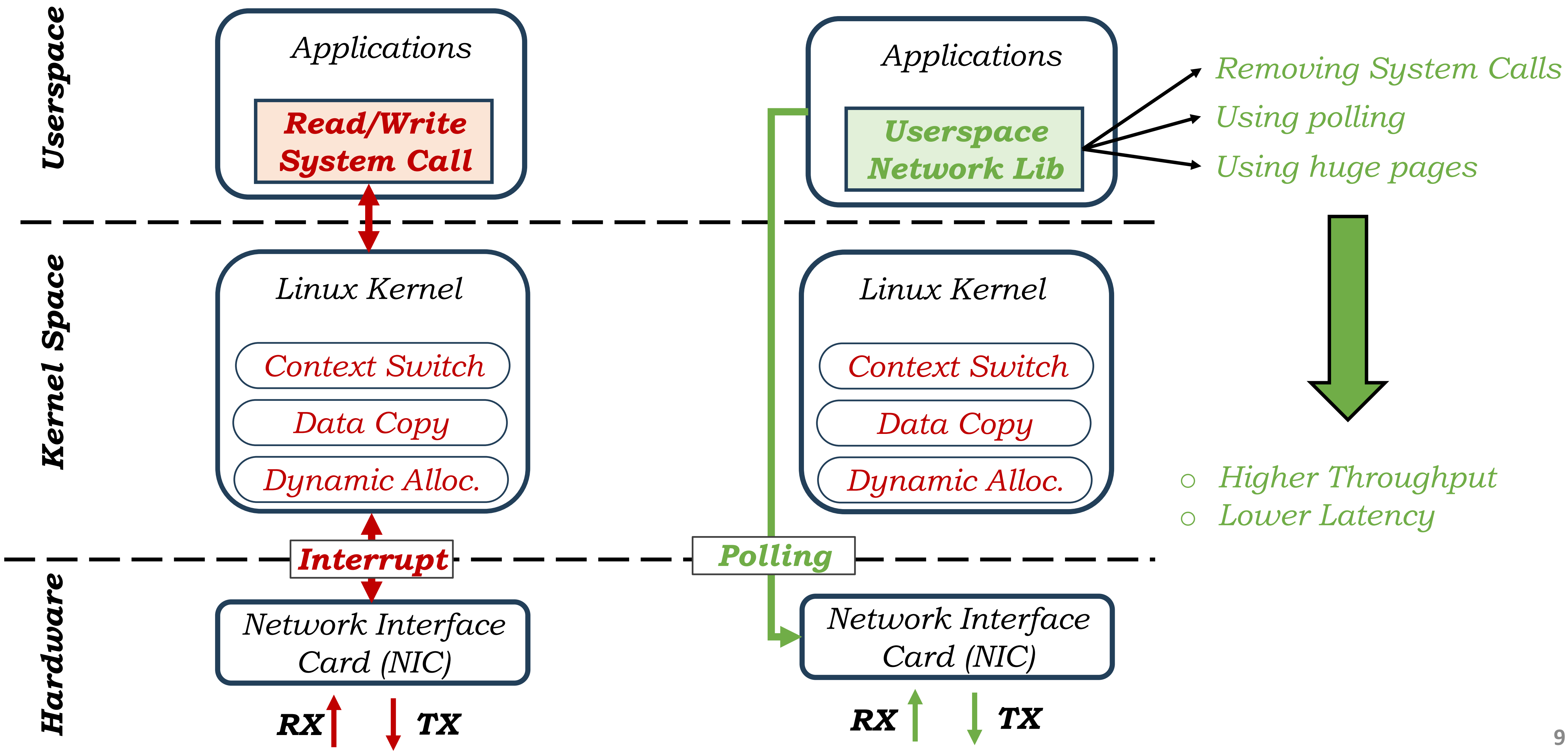
Outline of the Talk

- Background on userspace Networking
- Changes in gem5
  - Enabling userspace networking
  - Enabling NIC operation with a Poll Mode Driver
  - Enabling hardware load generator
- Network subsystem evaluation
  - Benchmarks
  - Causes of packet drop
- Experimental results
- Conclusion

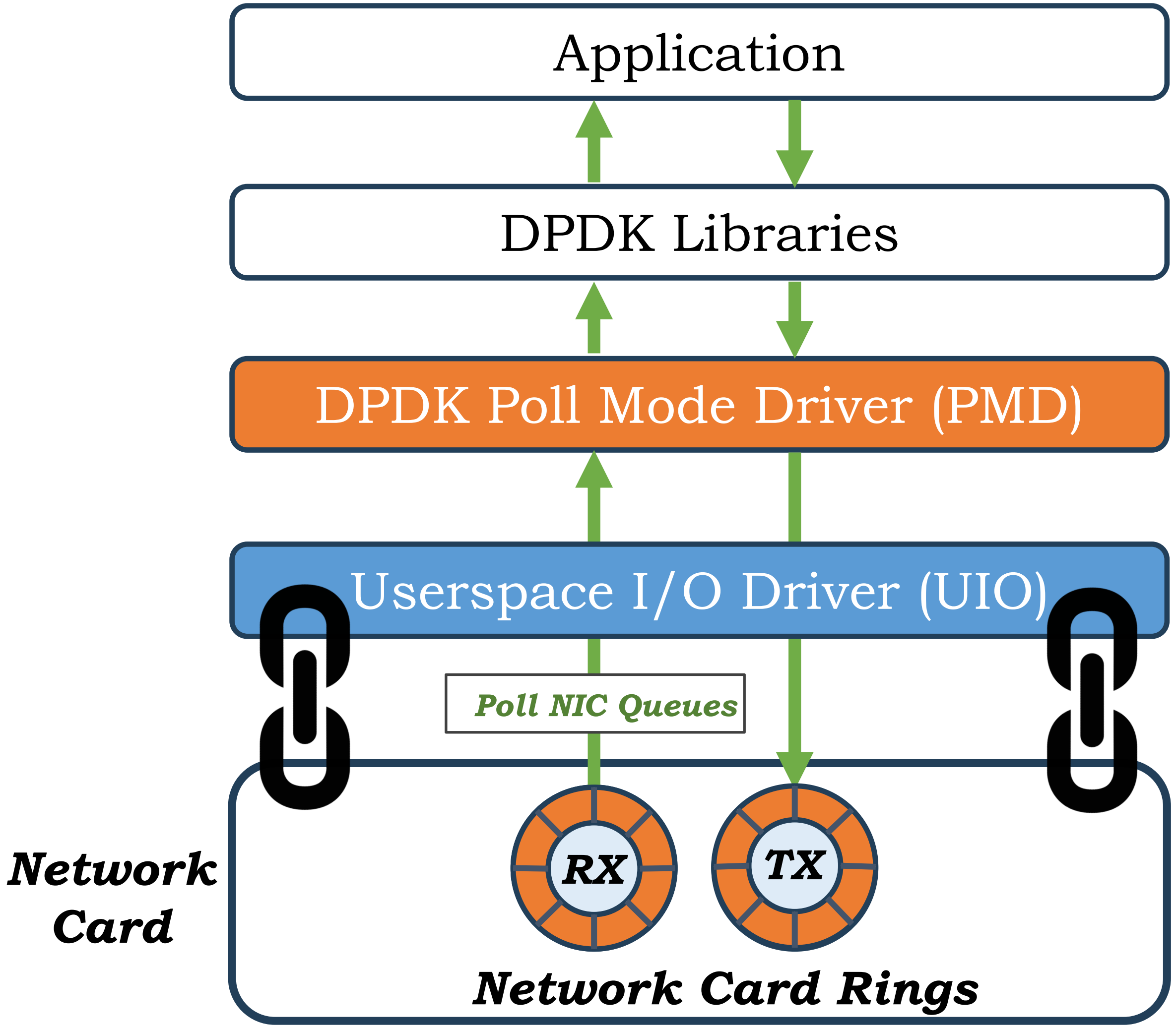
Outline of the Talk

- **Background on userspace Networking**
- Changes in gem5
  - Enabling userspace networking
  - Enabling NIC operation with Poll Mode Driver
  - Enabling the hardware load generator
- Network subsystem evaluation
  - Benchmarks
  - Causes of packet drop
- Experimental results
- Conclusion

Kernel Space vs. Userspace Networking



DPDK: The State-of-the-Art Userspace Networking Framework



Outline of the Talk

- Background on Userspace Networking
- **Changes in gem5**
  - **Enabling Userspace IO Driver**
  - Enabling NIC operation with Poll Mode Driver (PMD)
  - Enabling The hardware load generator
- Network subsystem evaluation
  - Benchmarks
  - Causes of packet drop
- Experimental results
- Conclusion

Enabling Userspace I/O Driver

- Userspace I/O Driver **allows direct access** to PCI device from userspace.
- Implementation is not complete in gem5

Offset	*bits[31:24]*	*bits[23:16]*	*bits[15:8]*	*bits[7:0]*
0x00	*Device ID*		*Vendor ID*	
0x04	*Status*		***Command***	

0x05

First 8 Bytes of PCI Configuration Space

Enabling Userspace I/O Driver

- Userspace I/O Driver **allows direct access** to PCI device from userspace.
- Implementation is not complete in gem5

Offset	*bits[31:24]*	*bits[23:16]*	*bits[15:8]*	*bits[7:0]*
0x00	*Device ID*		*Vendor ID*	
0x04	*Status*		***Command***	

0x05

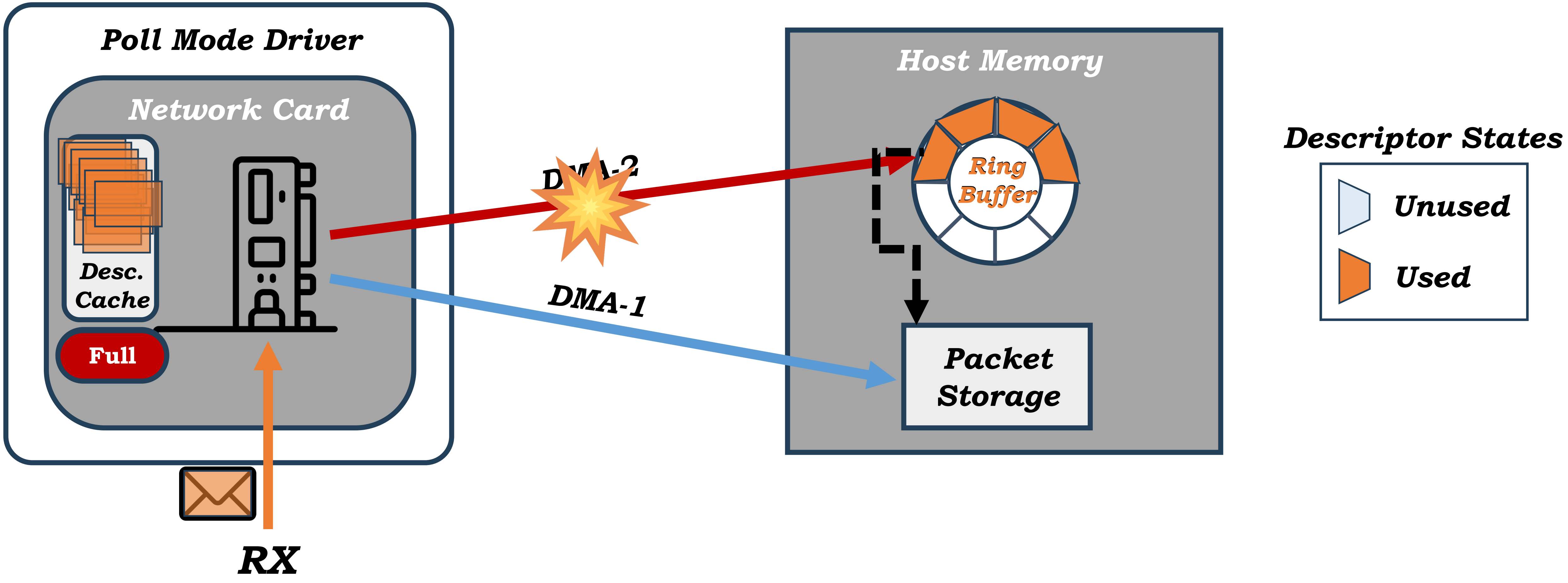
Bit<10>
Interrupt Disable

First 8 Bytes of PCI Configuration Space

Outline of the Talk

- Background on Userspace Networking
- **Changes in gem5**
  - Enabling Userspace IO Driver
  - **Enabling NIC operation with Poll Mode Driver (PMD)**
  - Enabling The hardware load generator
- Network subsystem evaluation
  - Benchmarks
  - Causes of packet drop
- Experimental results
- Conclusion

Enabling NIC Operation with a PMD



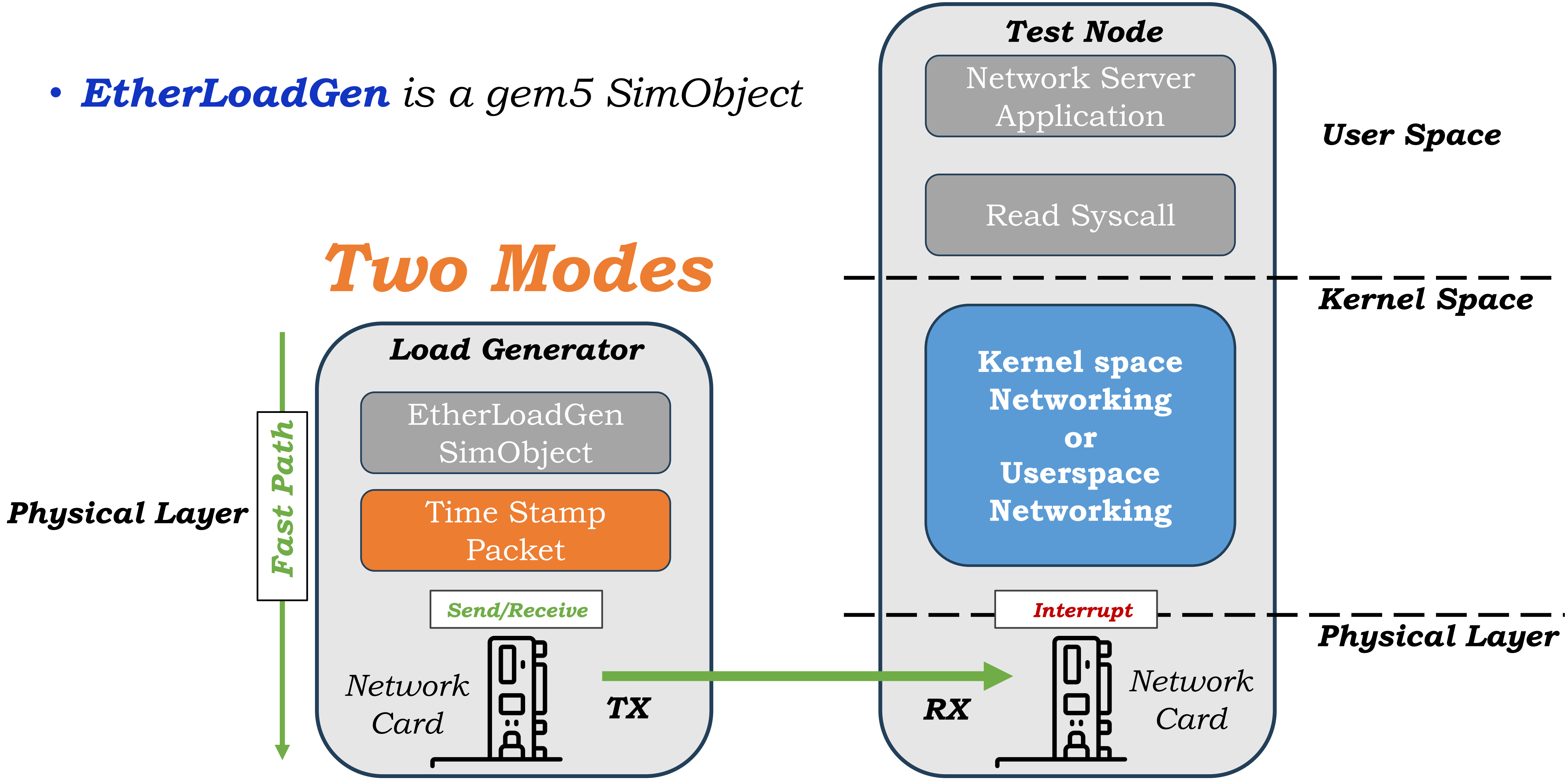
Outline of the Talk

- Background on Userspace Networking
- **Changes in gem5**
  - Enabling Userspace IO Driver
  - Enabling NIC operation with Poll Mode Driver (PMD)
  - **Enabling The hardware load generator**
- Network subsystem evaluation
  - Benchmarks
  - Causes of packet drop
- Experimental results
- Conclusion

The Hardware Load Generator Model

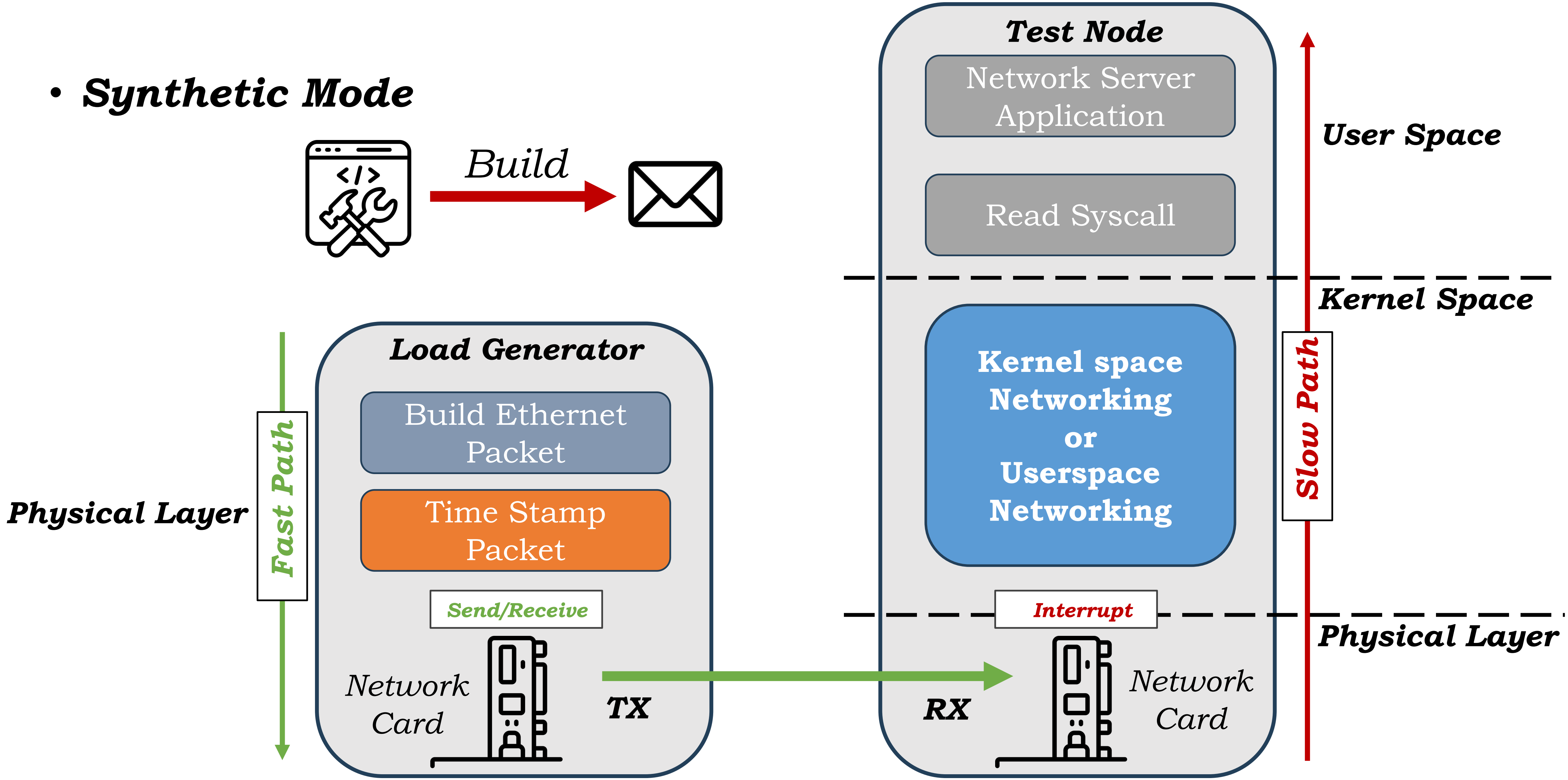
- ***EtherLoadGen*** *is a gem5 SimObject*

Two Modes



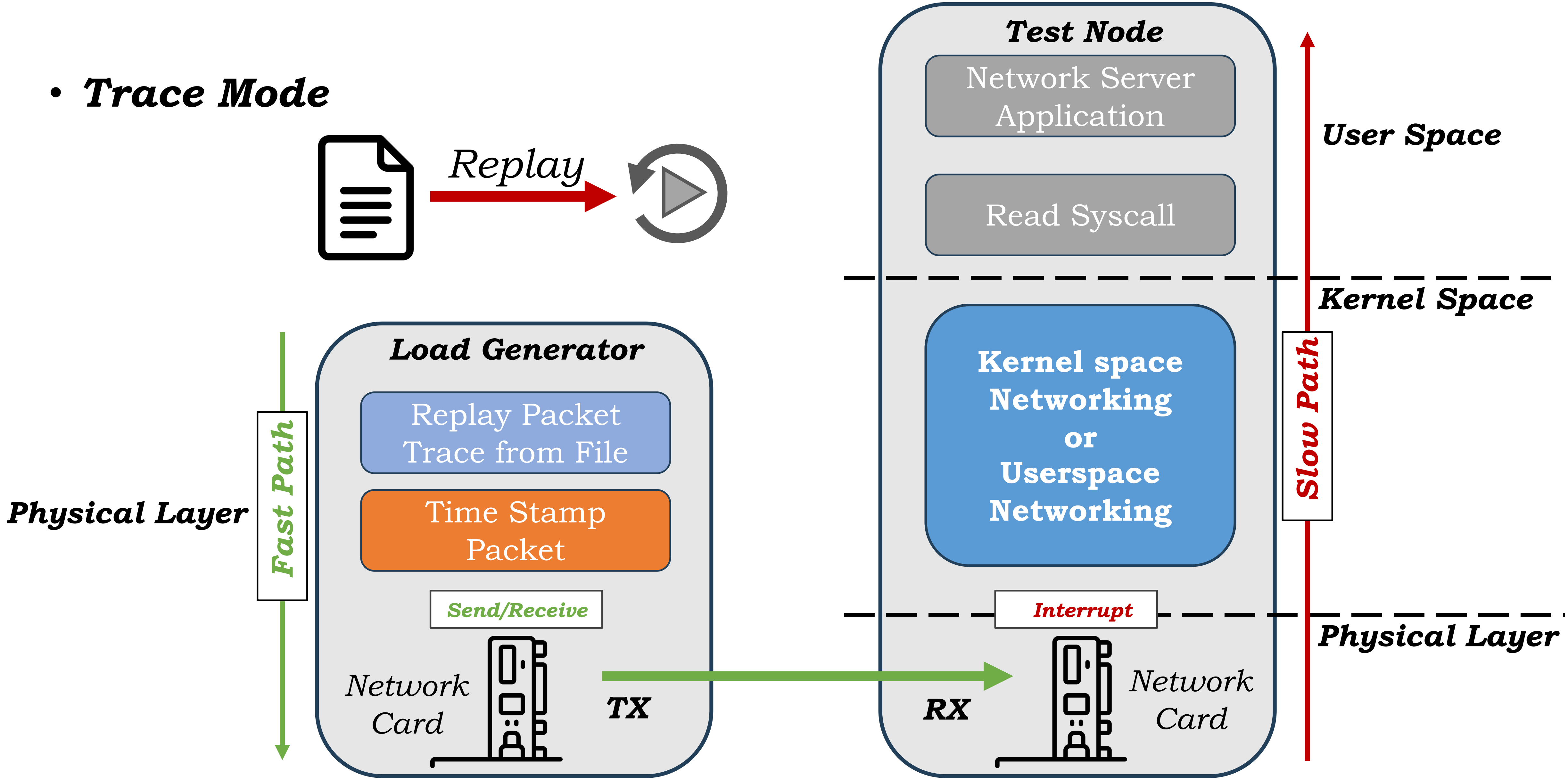
How LoadGen Works

- ***Synthetic Mode***



How LoadGen Works

- ***Trace Mode***



Outline

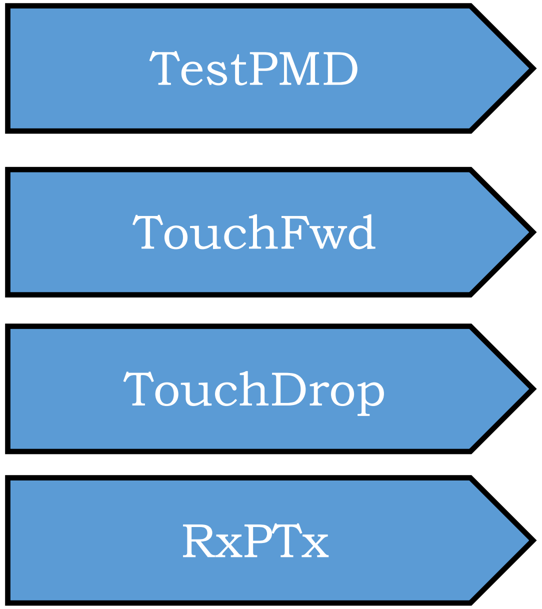
- Background on userspace networking
- Changes in gem5
  - Enabling userspace networking
  - Enabling Poll Mode Driver (PMD)
  - Enabling Hardware Load Generator
- **Network subsystem evaluation**
  - **Benchmarks**
  - Causes of packet drop
- Experimental results
- Conclusion

Benchmark Suite

- Network Applications to benchmark the end host performance in processing network packet.

Benchmark Suite

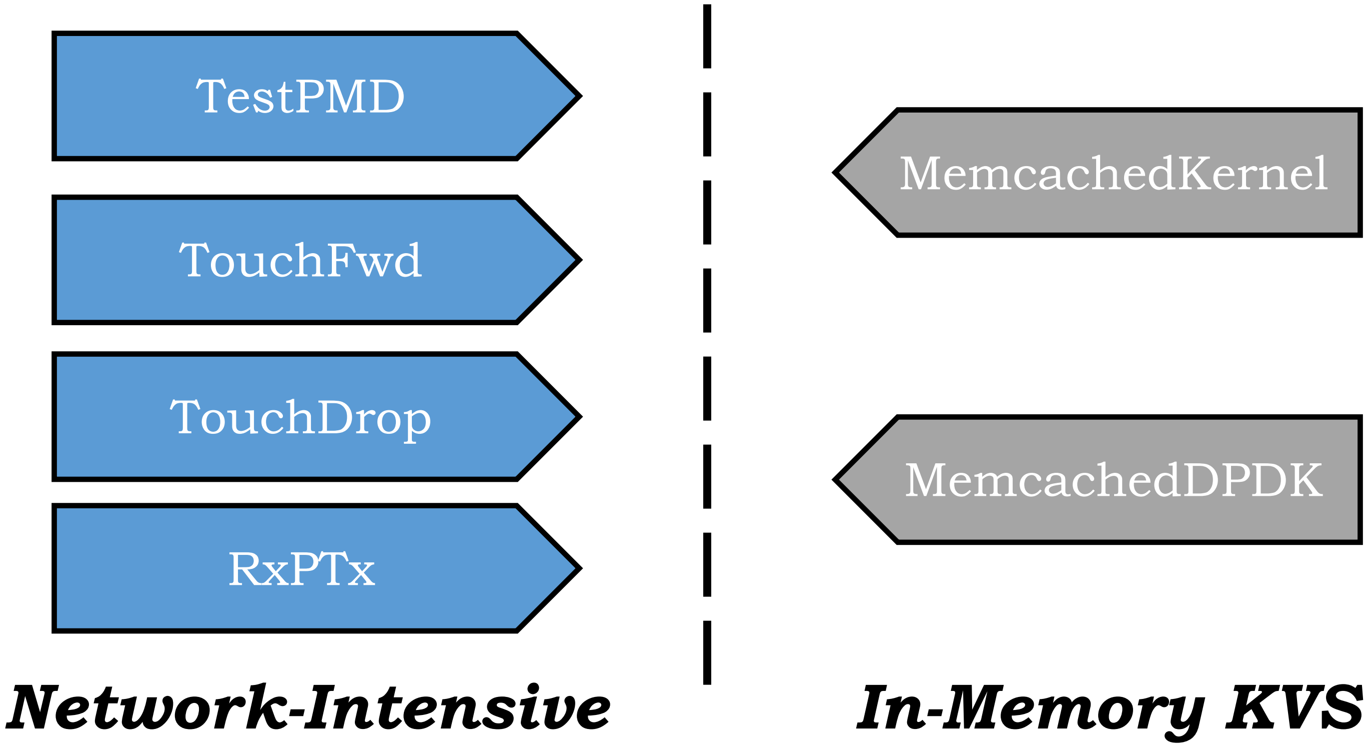
- Network Applications to benchmark the end host performance in processing network packet.
- **Four** network-intensive applications



Network-Intensive

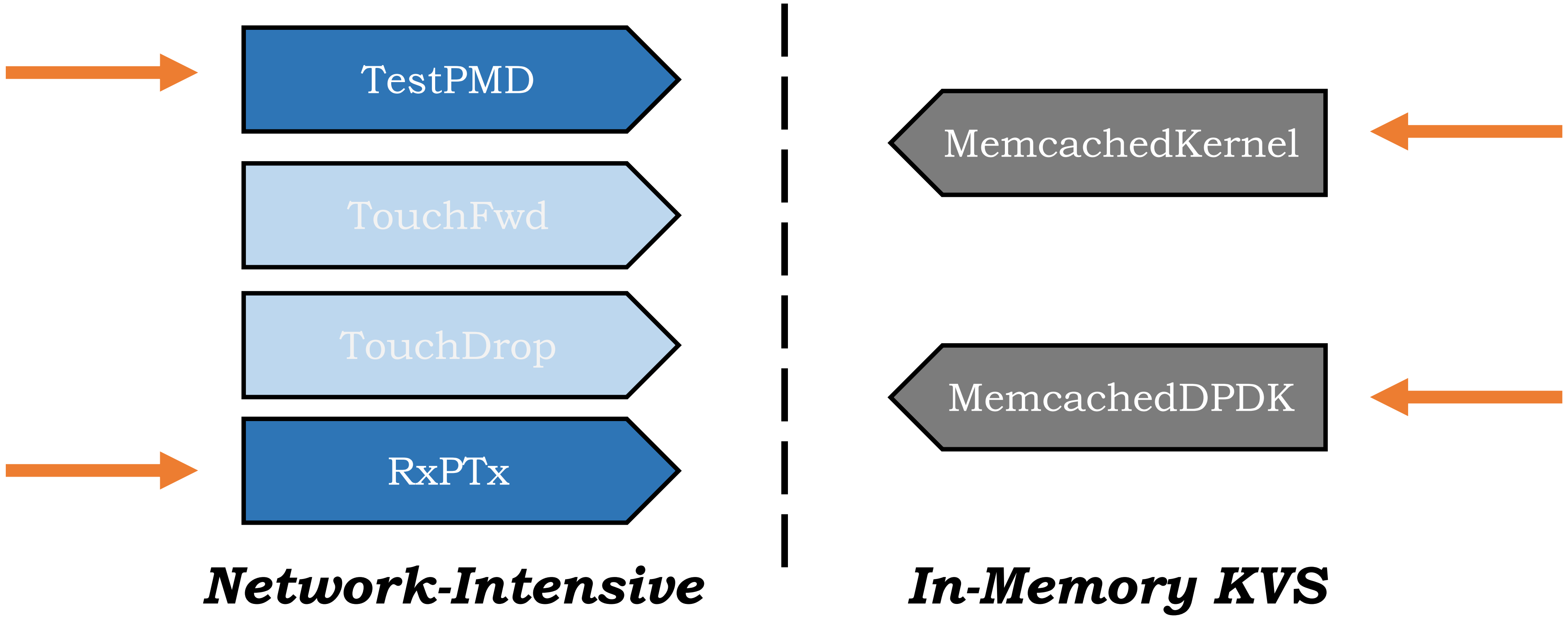
Benchmark Suite

- Network Applications to benchmark the end host performance in processing network packet.
- **Four** network-intensive applications
- **Two** in-memory key-value stores



Benchmark Suite

- Network Applications to benchmark the end host performance in processing network packet.
- **Four** network-intensive applications
- **Two** in-memory key-value stores

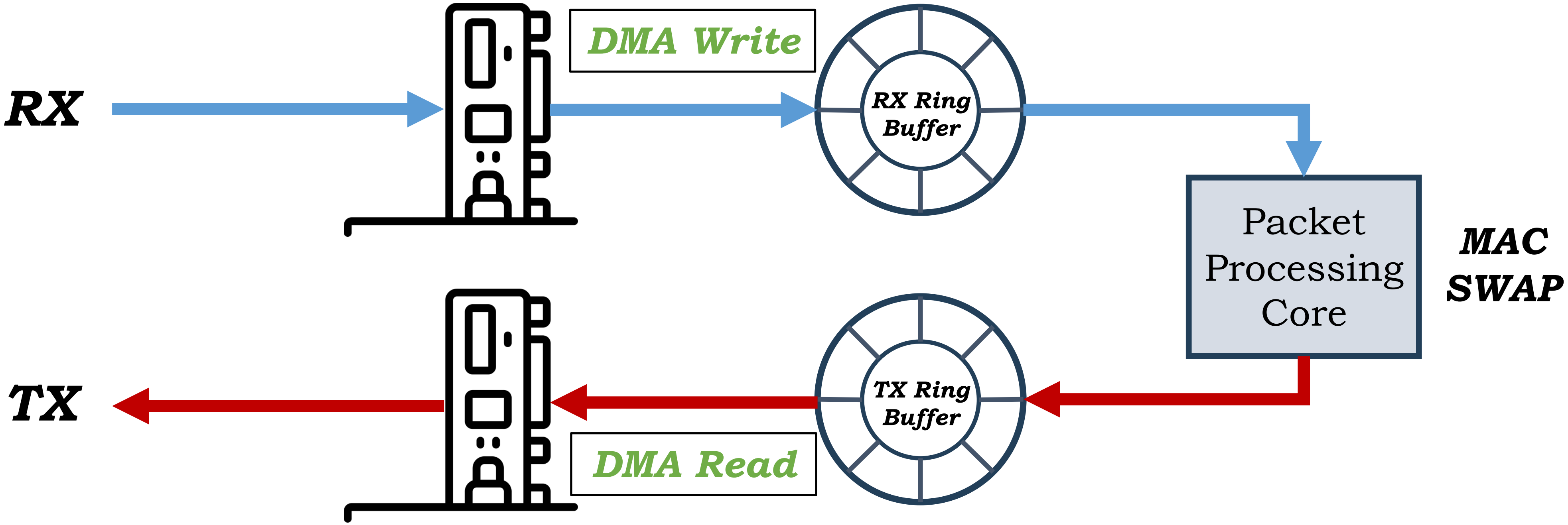


Please see our paper

- Other changes to gem5
- Benchmarks
- Changes to DPDK
- Analysis of Packet Drops
- Real-System vs. gem5 Results
- Bandwidth Sensitivity Analysis
- More interesting results
- And more . . .

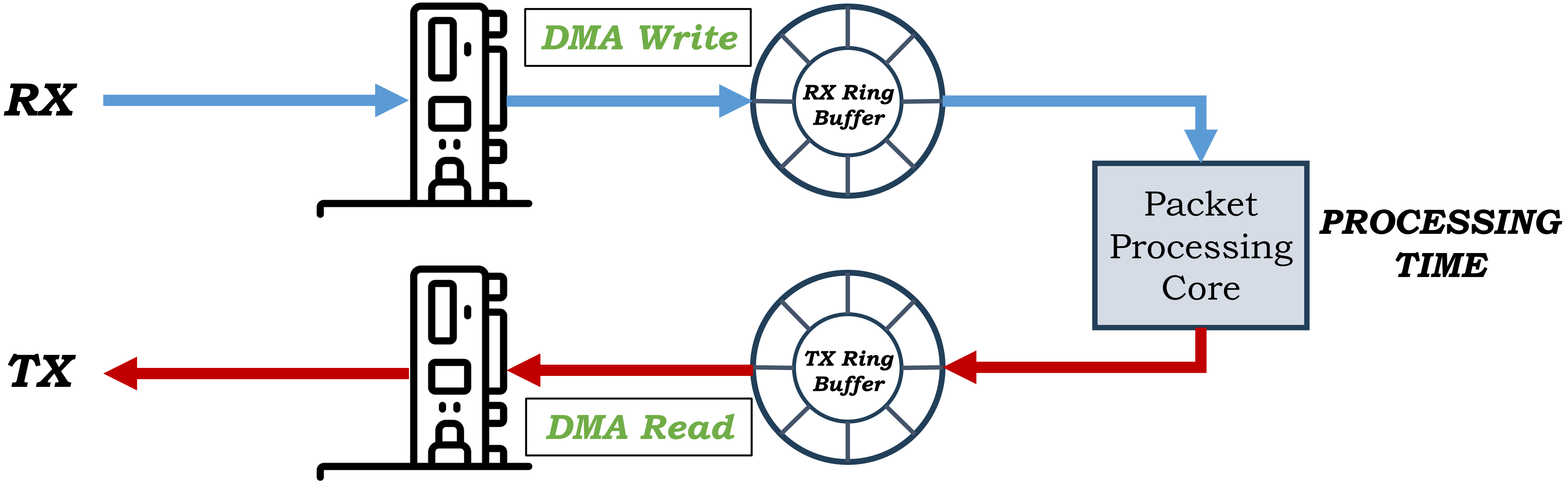
TestPMD

- Receives a Network packet, swaps MAC (if enabled), and Forwards it.
- Models shallow-network functionality



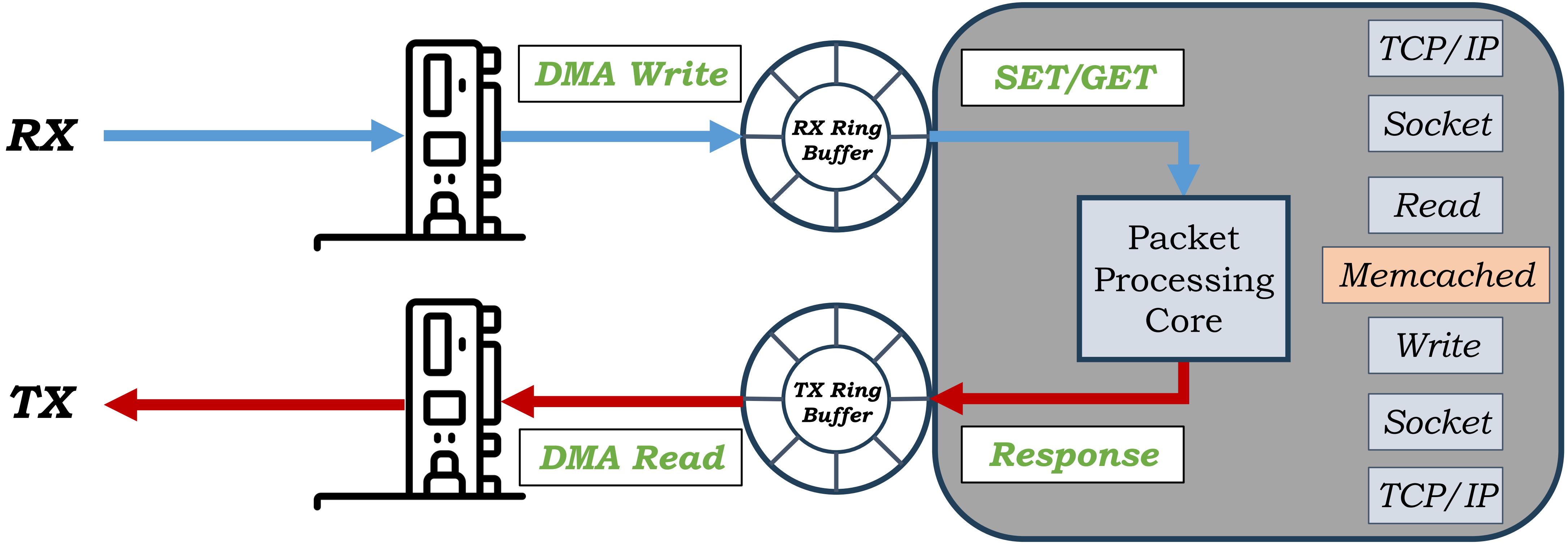
RxPTx

- Receives a Network packet, waits for ***processing*** time, and Forwards it.
- Models network functions with different core to DMA use distances



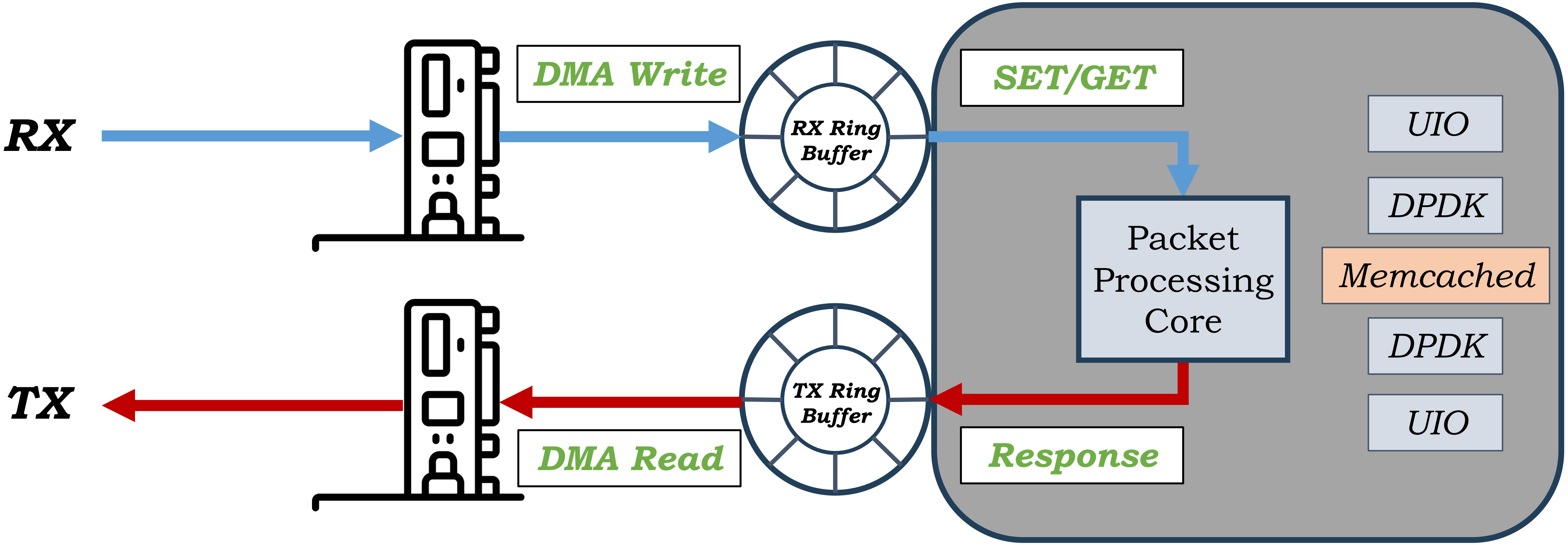
MemcachedKernel

- In-memory key-value store client implemented using the memcached library and Linux POSIX API
- RX/TX over the Linux Kernel Network Stack



MemcachedDPDK

- In-memory key-value store client implemented on top on DPDK to achieve higher throughput and lower latency
- RX/TX over DPDK

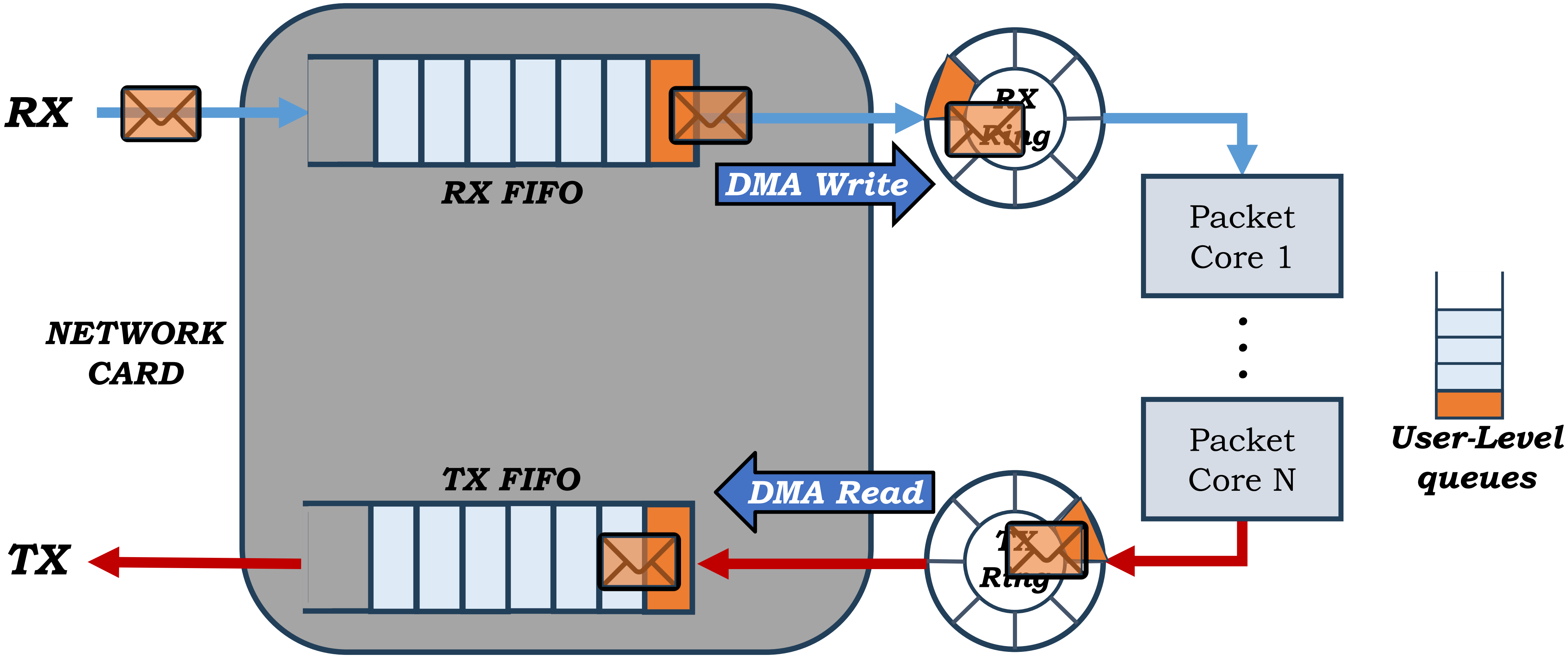


Outline

- Background on userspace networking
- Changes in gem5
  - Enabling userspace networking
  - Enabling Poll Mode Driver (PMD)
  - Enabling Hardware Load Generator
- **Network subsystem evaluation**
  - Benchmarks
  - **Causes of packet drop**
- Experimental results
- Conclusion

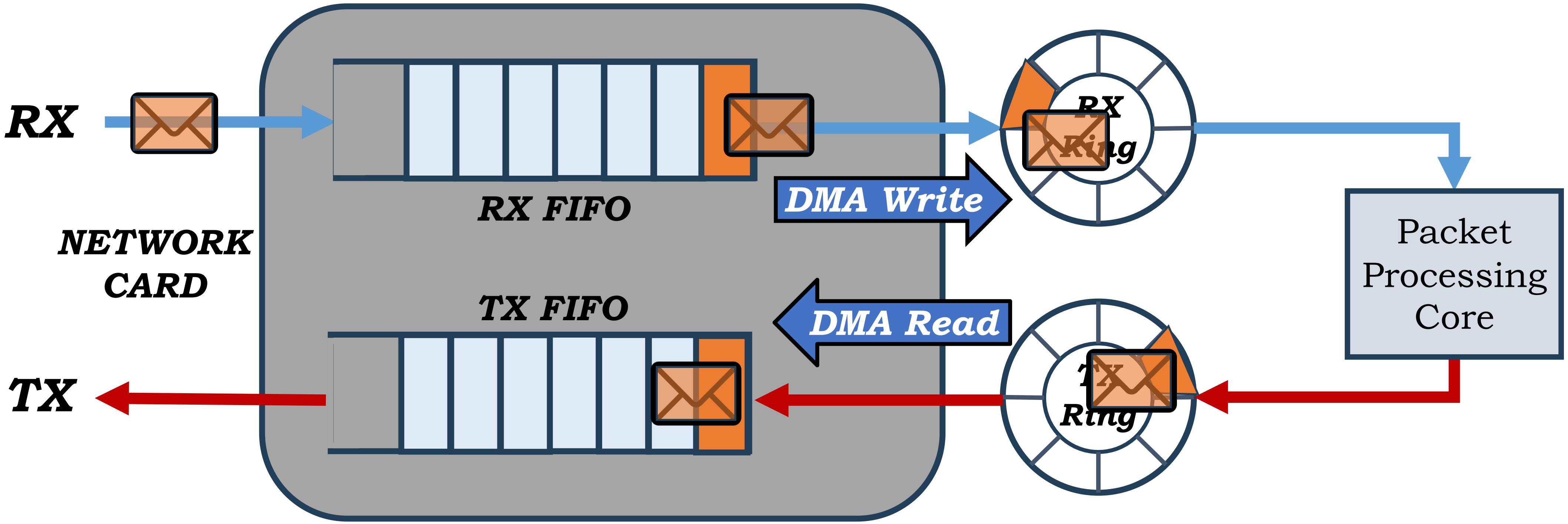
DPDK Modes: Pipeline

- Packet Life Cycle



DPDK Modes: Run-to-Completion

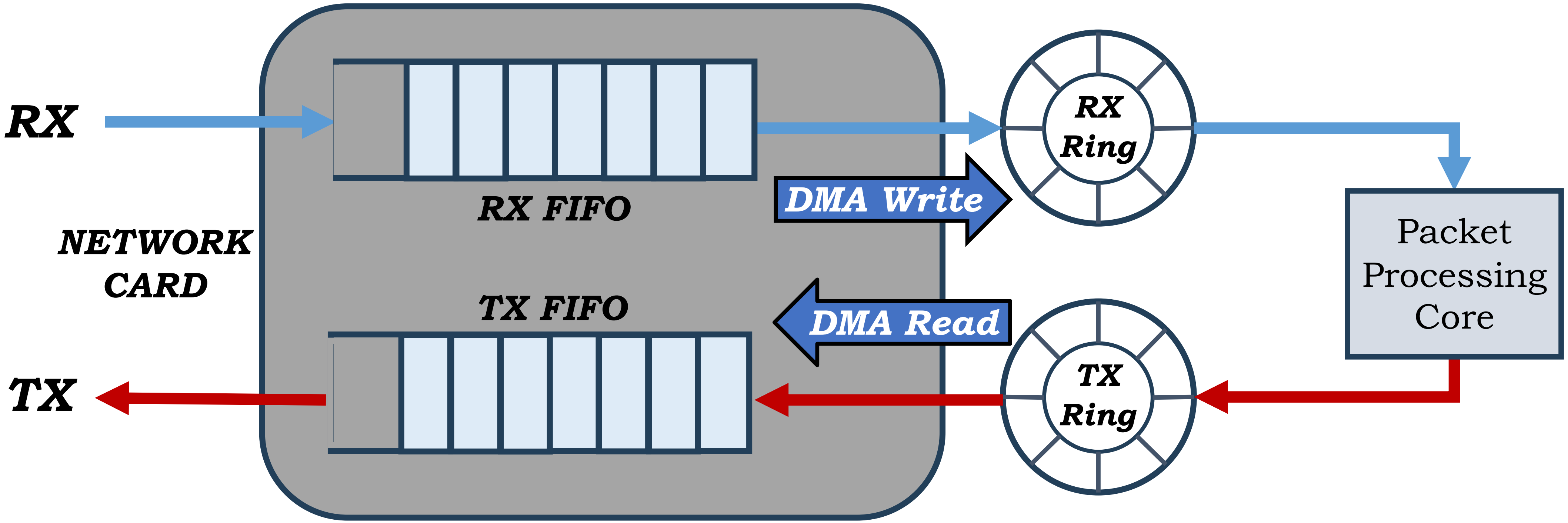
- Packet Life Cycle: Run-to-completion



We classify the causes of packet drops into DmaDrops, CoreDrops, and TxDrops

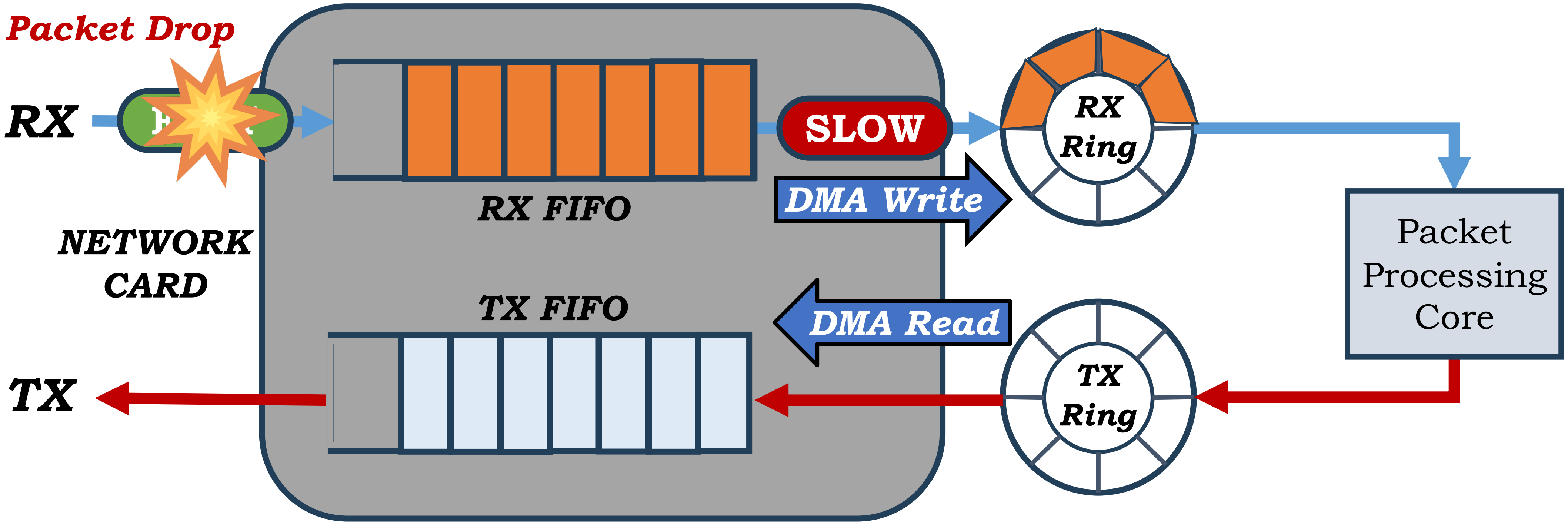
Causes of Packet Drops

- **DmaDrops**



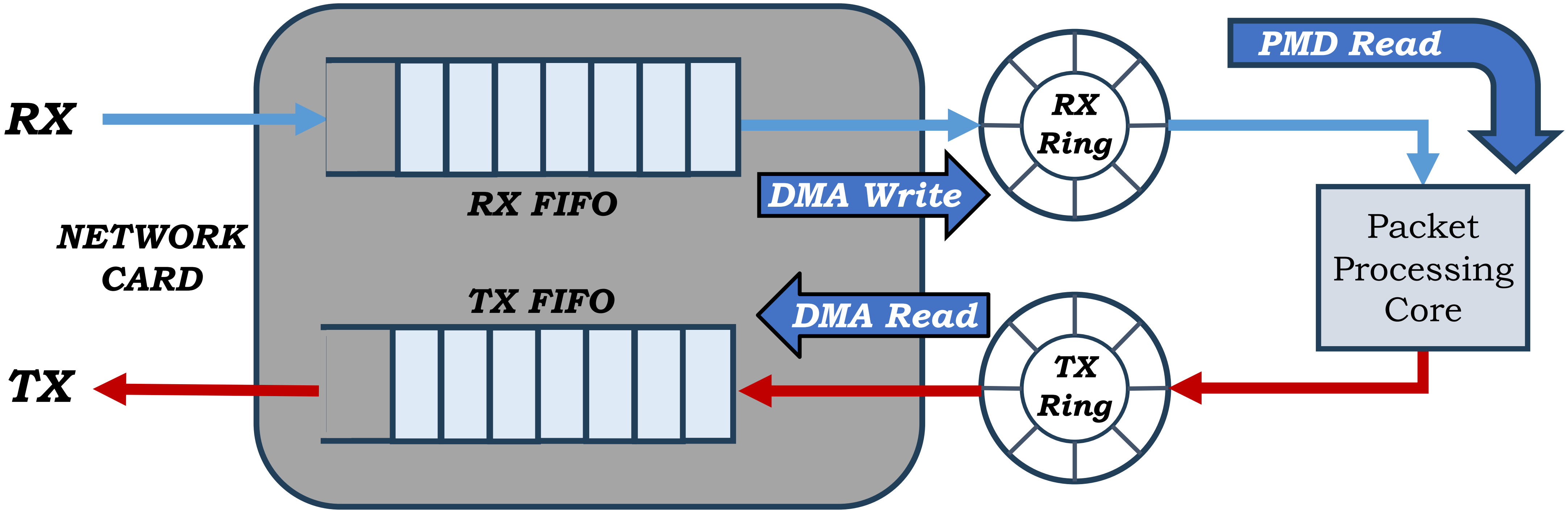
Causes of Packet Drops

- **DmaDrops**



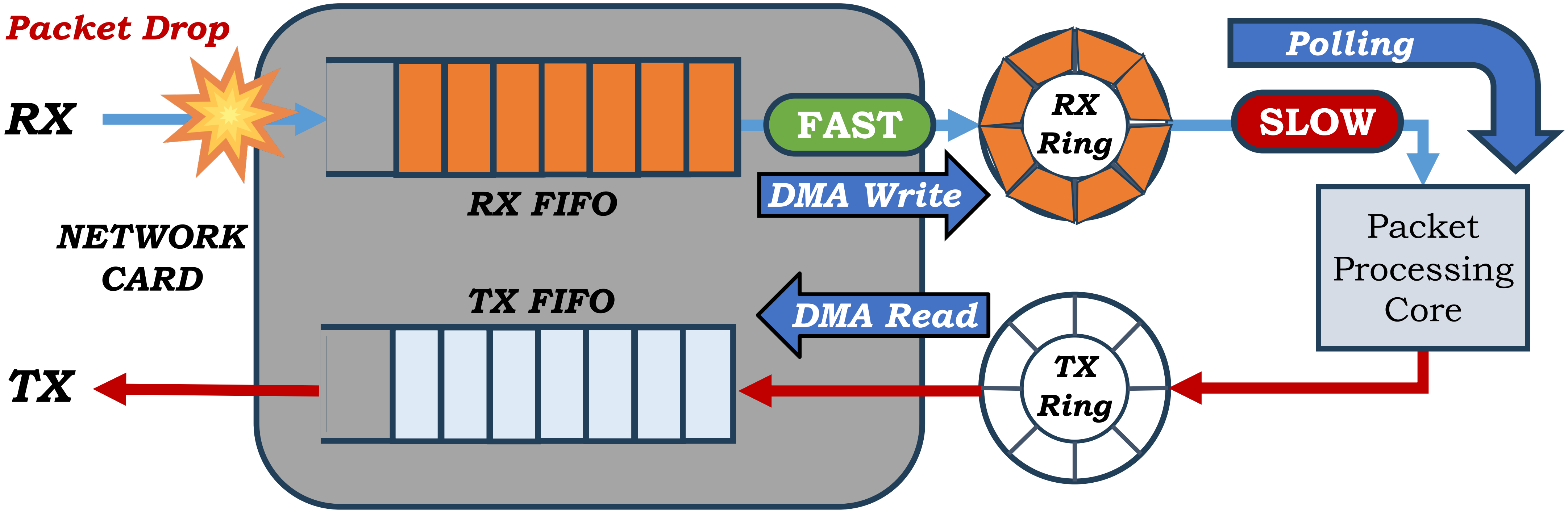
Causes of Packet Drops

- CoreDrops



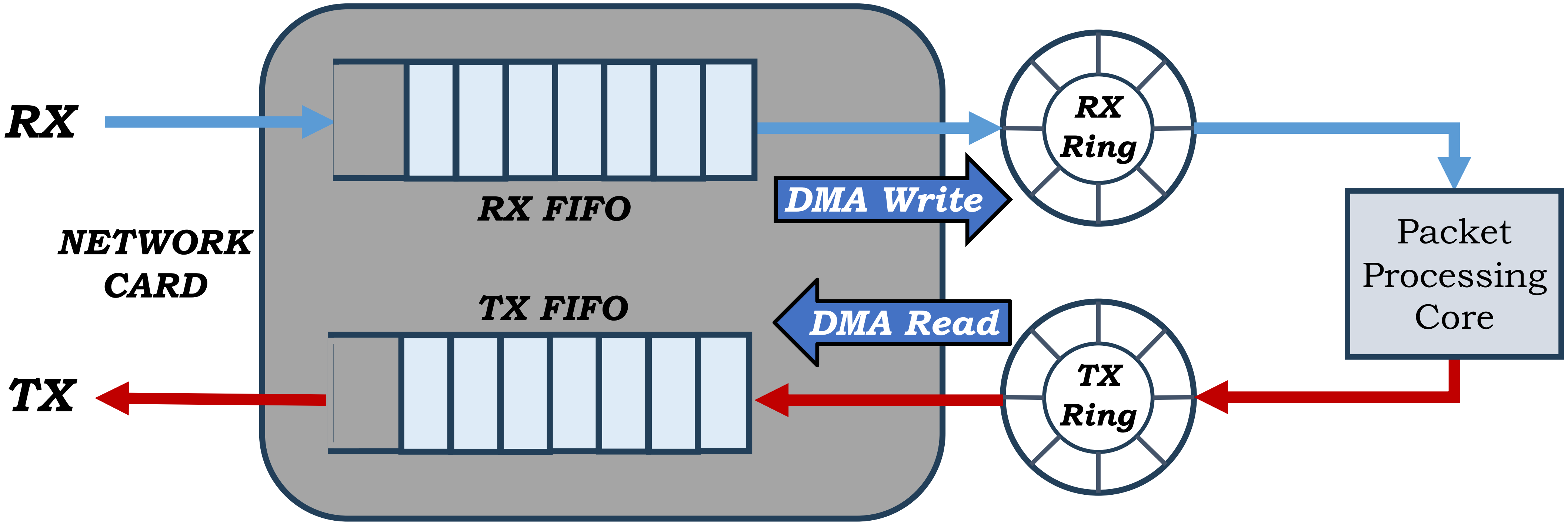
Causes of Packet Drops

- CoreDrops



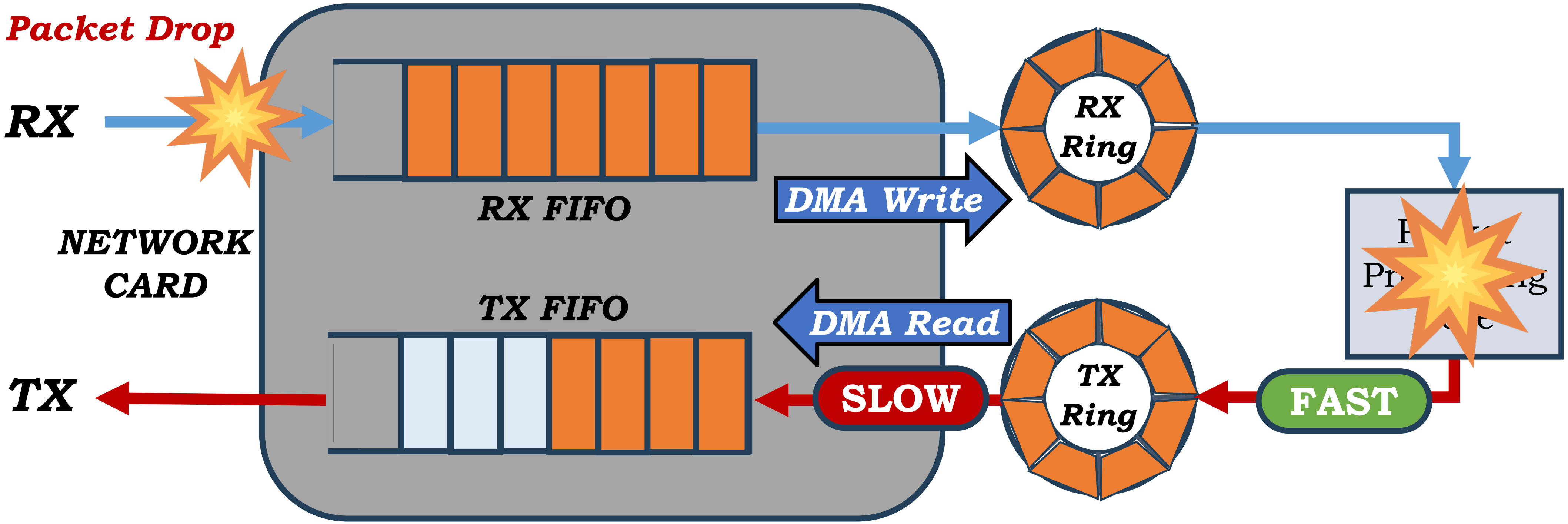
Causes of Packet Drops

- TxDrops



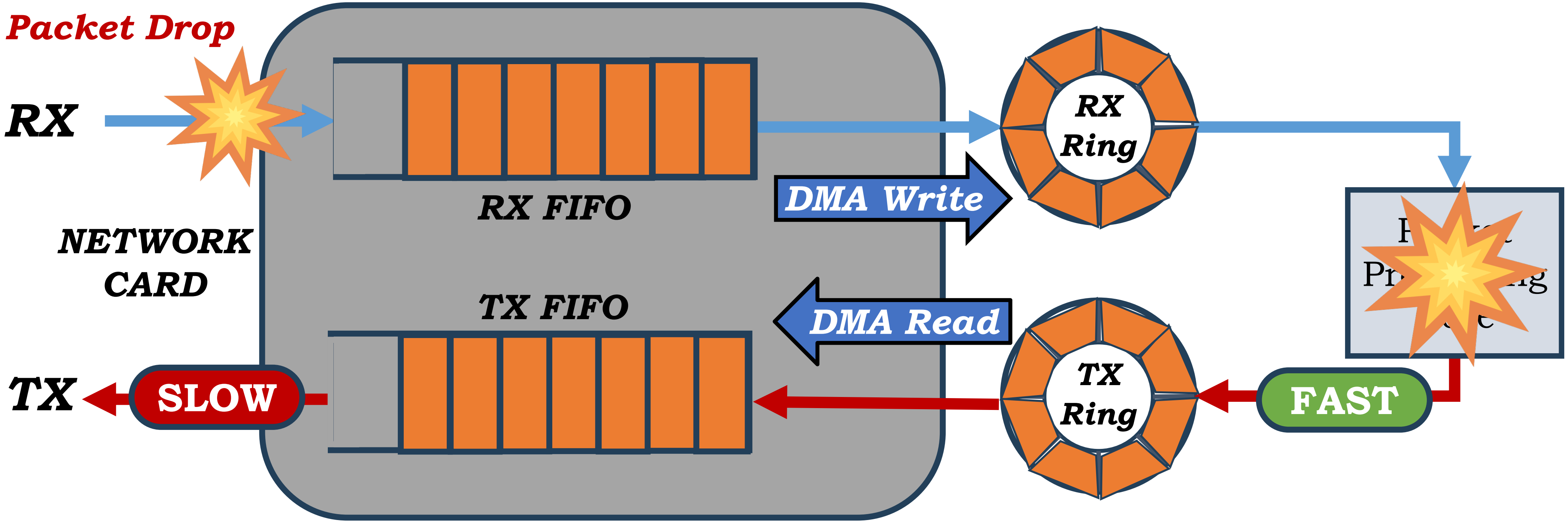
Causes of Packet Drops

- TxDrops



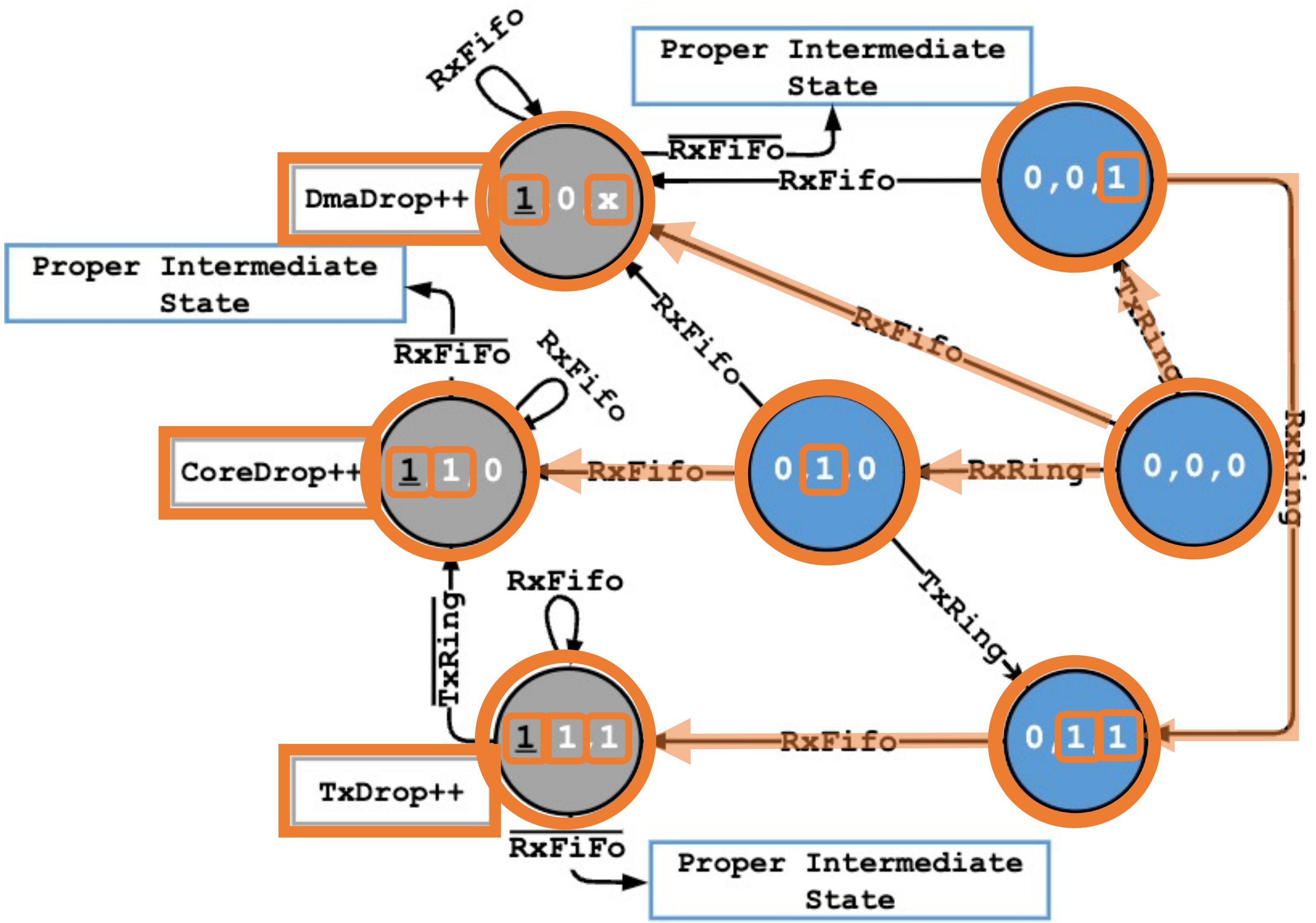
Causes of Packet Drops

- TxDrops

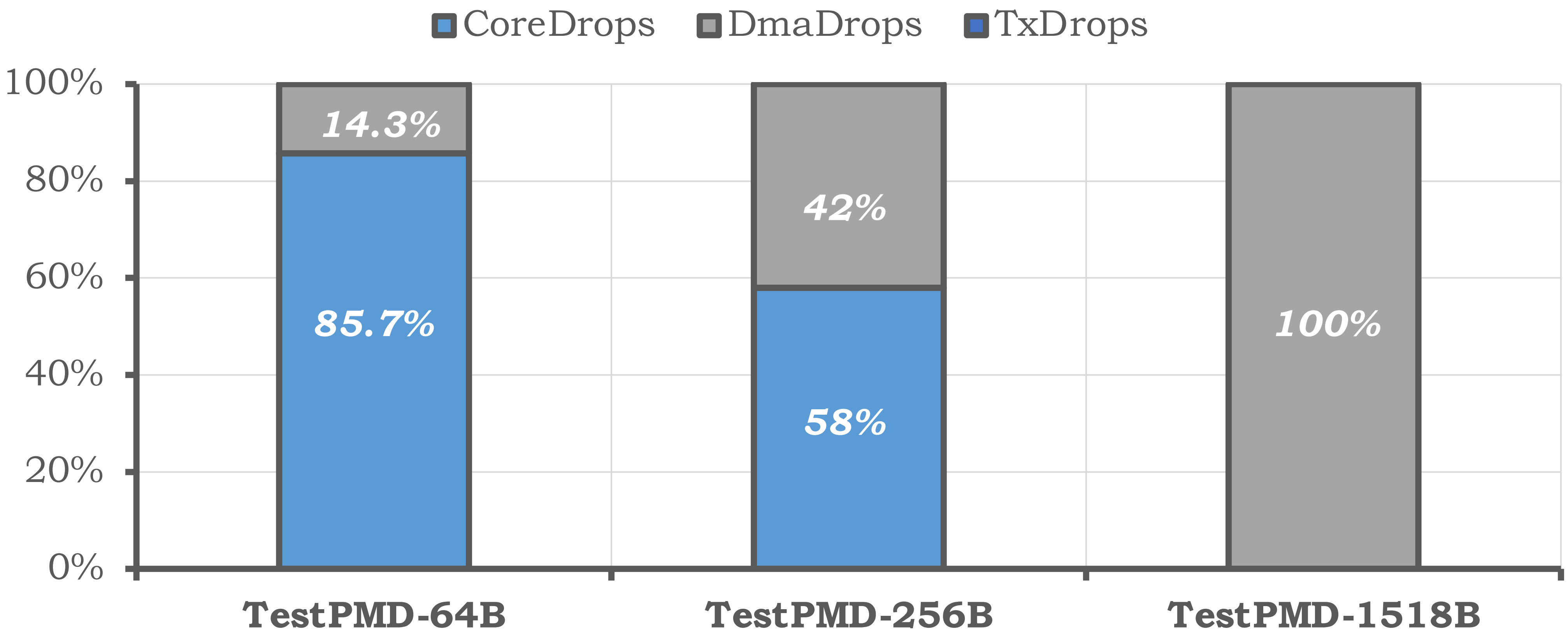


Causes of Packet Drops

- **FSM**



Causes of Packet Drops



Small packets are core bound, while large packets are DMA bound

Outline

- Background on userspace networking
- Changes in gem5
  - Enabling userspace networking
  - Enabling Poll Mode Driver (PMD)
  - Enabling Hardware Load Generator
- Network subsystem evaluation
  - Benchmarks
  - Causes of packet drop
- **Experimental results**
- Conclusion

Evaluation

1. *How does our gem5 implementation compare with real-system?*

2. *What is the sensitivity of network performance to micro-architectural settings?*

3. *Does real-workload benefit from our userspace network stack in gem5?*

Evaluation

Parameters	gem5	Ampere Altra
Core Frequency	3GHz	3GHz
Superscalar	4 ways	4 ways
ROB/IQ/LQ/SQ	128/120/68/72	128/120/68/72
L1I, L1D	64KB, 64KB	64KB, 64KB
L2, L3	1MB, 4MB/Core	1MB, 4MB/Core
Direct Cache Access [1]	Enabled (default)	Unknown
Network Latency	200μs	200μs
Network Bandwidth	100Gbps	100Gbps
Network Interface Card	Intel i8254xGBE NIC	Nvidia ConnectX-6

[1] M. Alian, Y. Yuan, J. Zhang, R. Wang, M. Jung, and N. S. Kim, "Data direct I/O characterization for future I/O system exploration,"

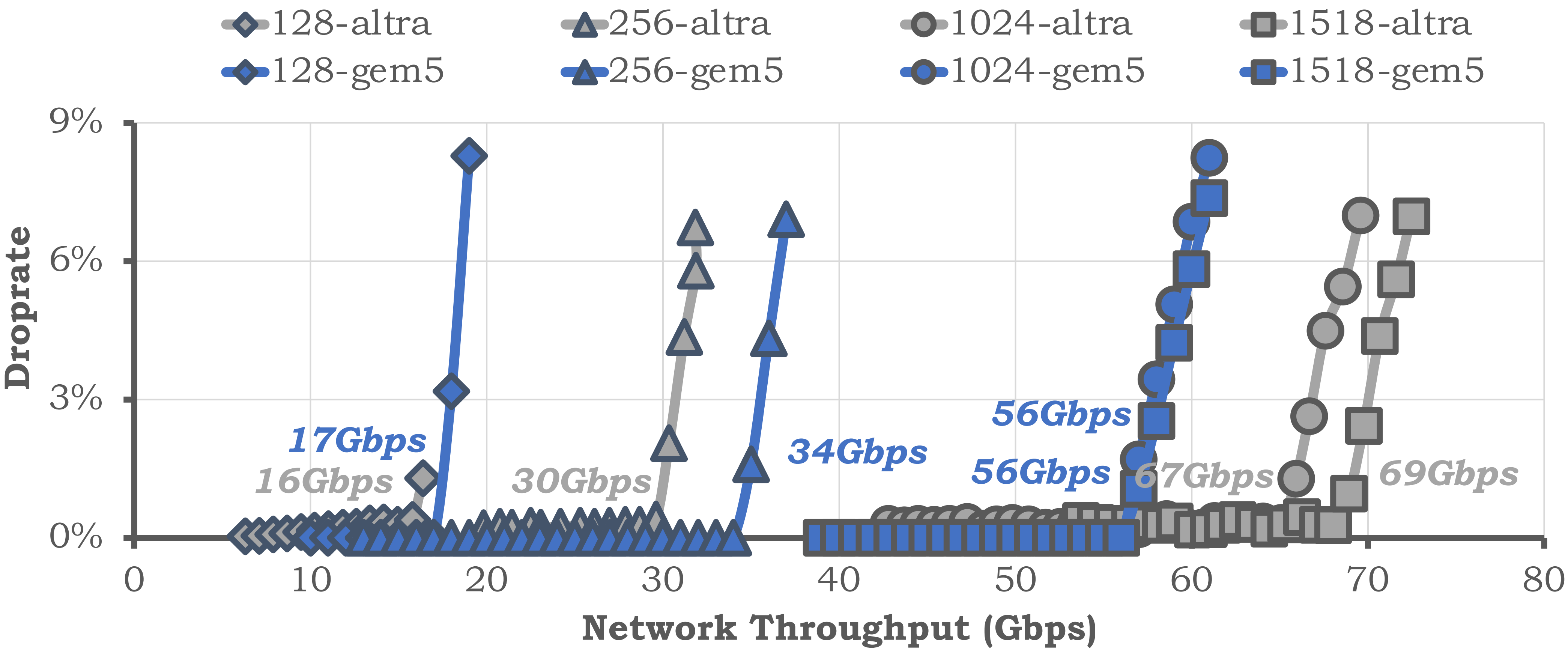
Evaluation

1 **How does our gem5 implementation compare with real-system?**

2 *What is the sensitivity of network performance to micro-architectural settings?*

3 *Does real-workload benefit from our userspace network stack in gem5?*

Real-System vs. gem5



gem5 networking follows a similar trend to that of real-system

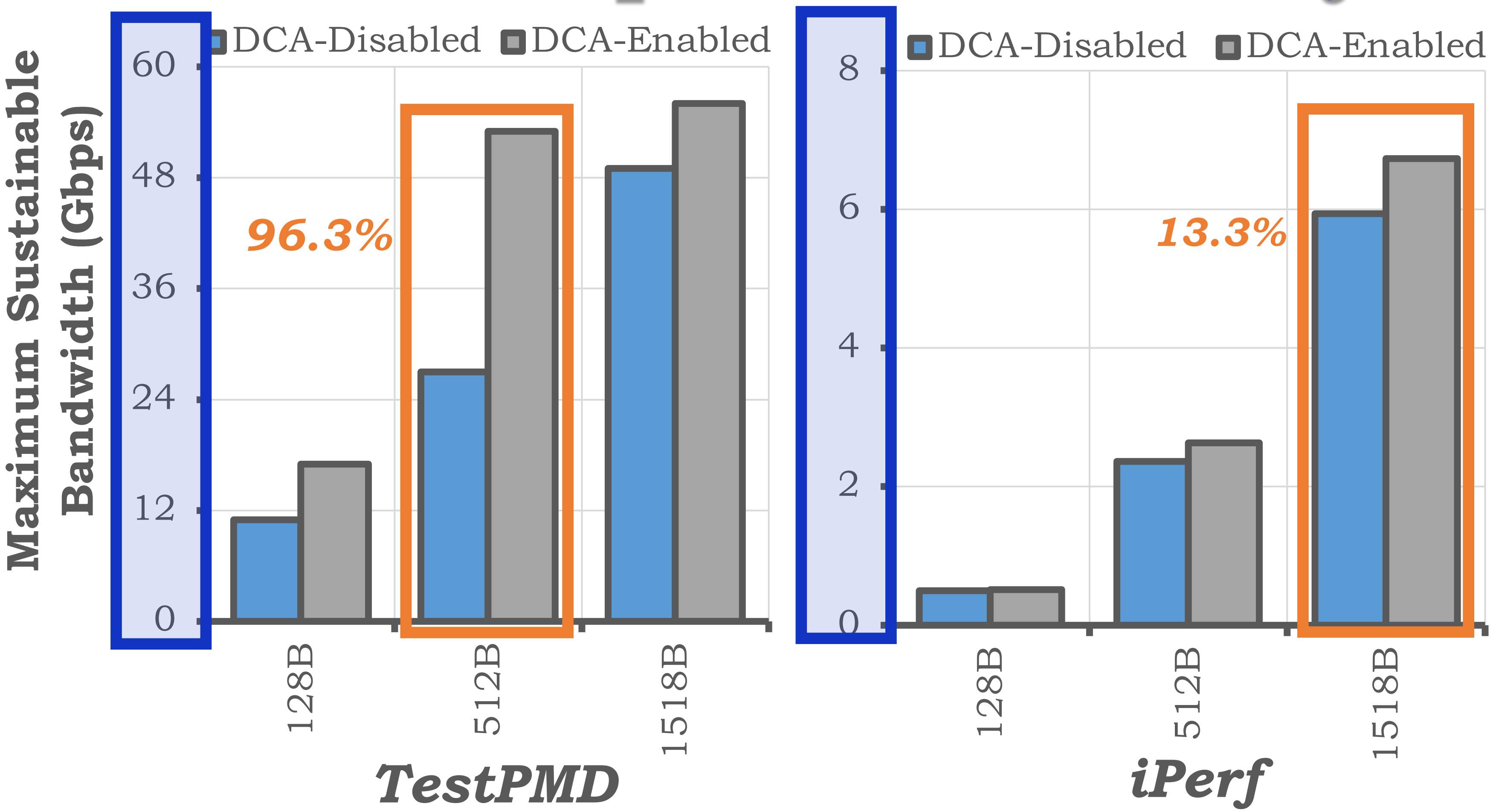
Evaluation

1. *How does our gem5 implementation compare with real-system?*

2. ***What is the sensitivity of network performance to micro-architectural settings?***

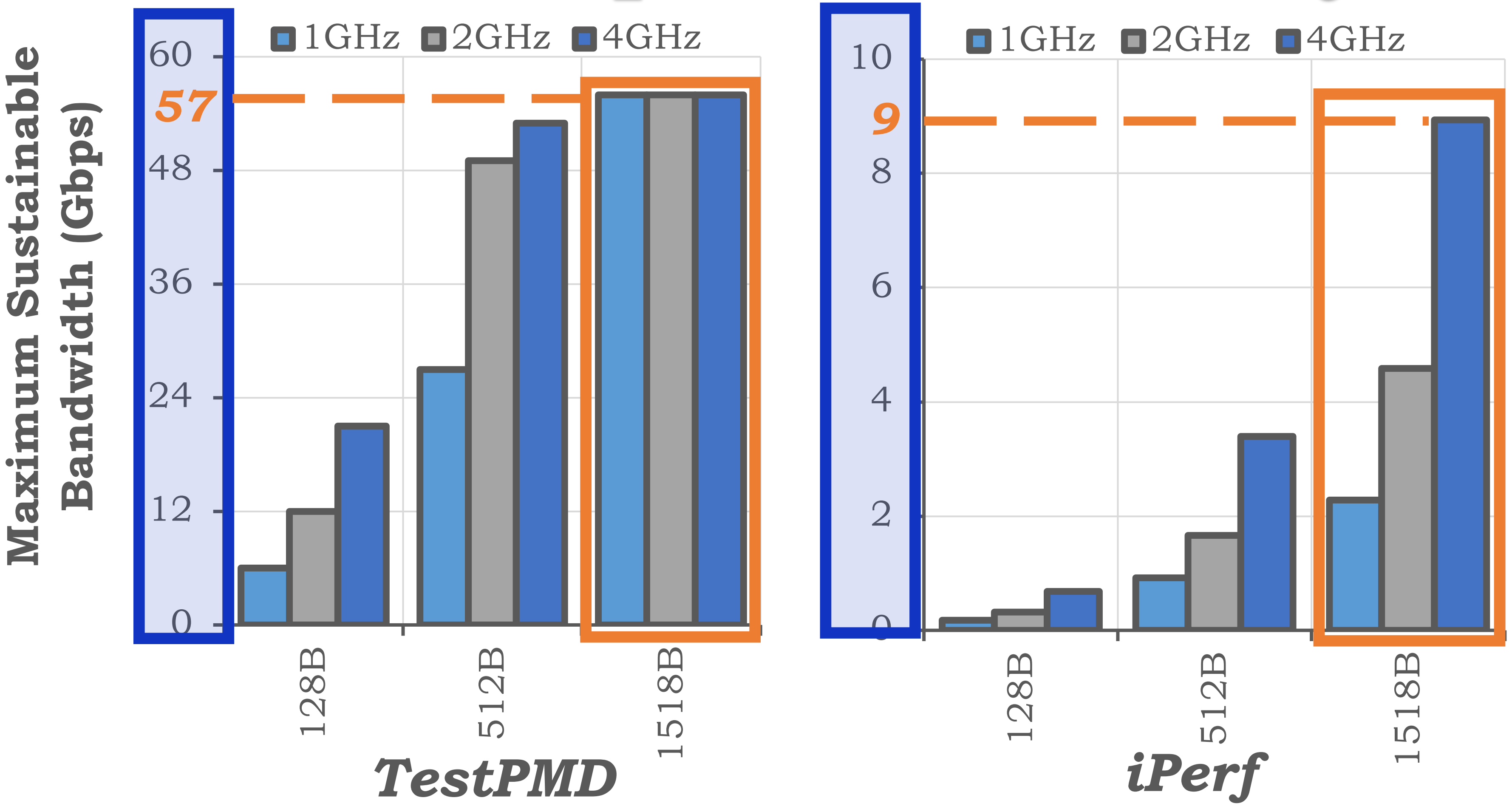
3. *Does real-workload benefit from our userspace network stack in gem5?*

User vs. Kernel Space Sensitivity to DCA



Userspace networking application are more sensitive to DCA than kernel space networking

User vs. Kernel Space Sensitivity to Freq.



Userspace networking application show a different sensitivity to CPU Frequency compared to kernel network application

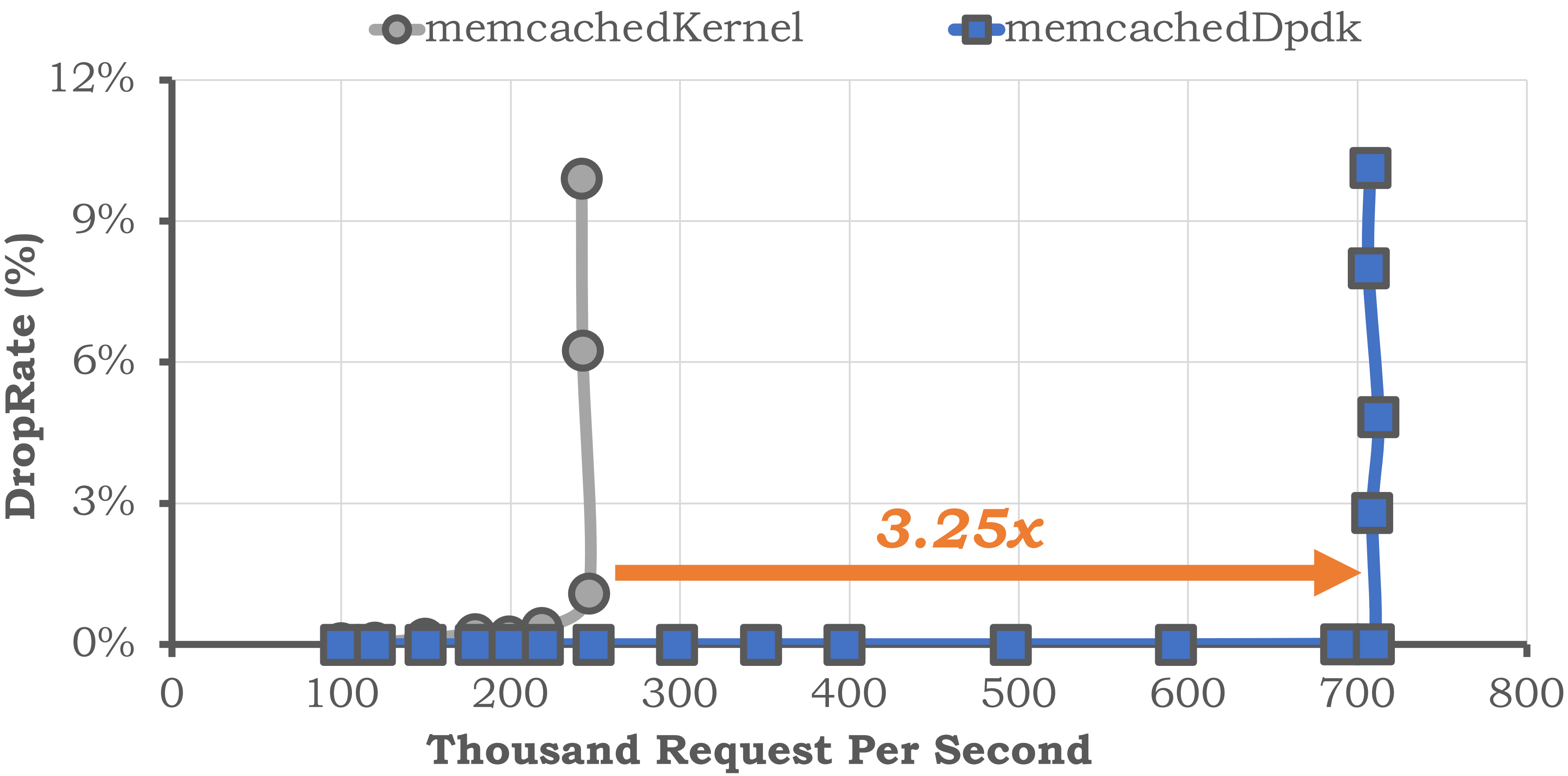
Evaluation

1 *How does our gem5 implementation compare with real-system?*

2 *What is the sensitivity of network performance to micro-architectural settings?*

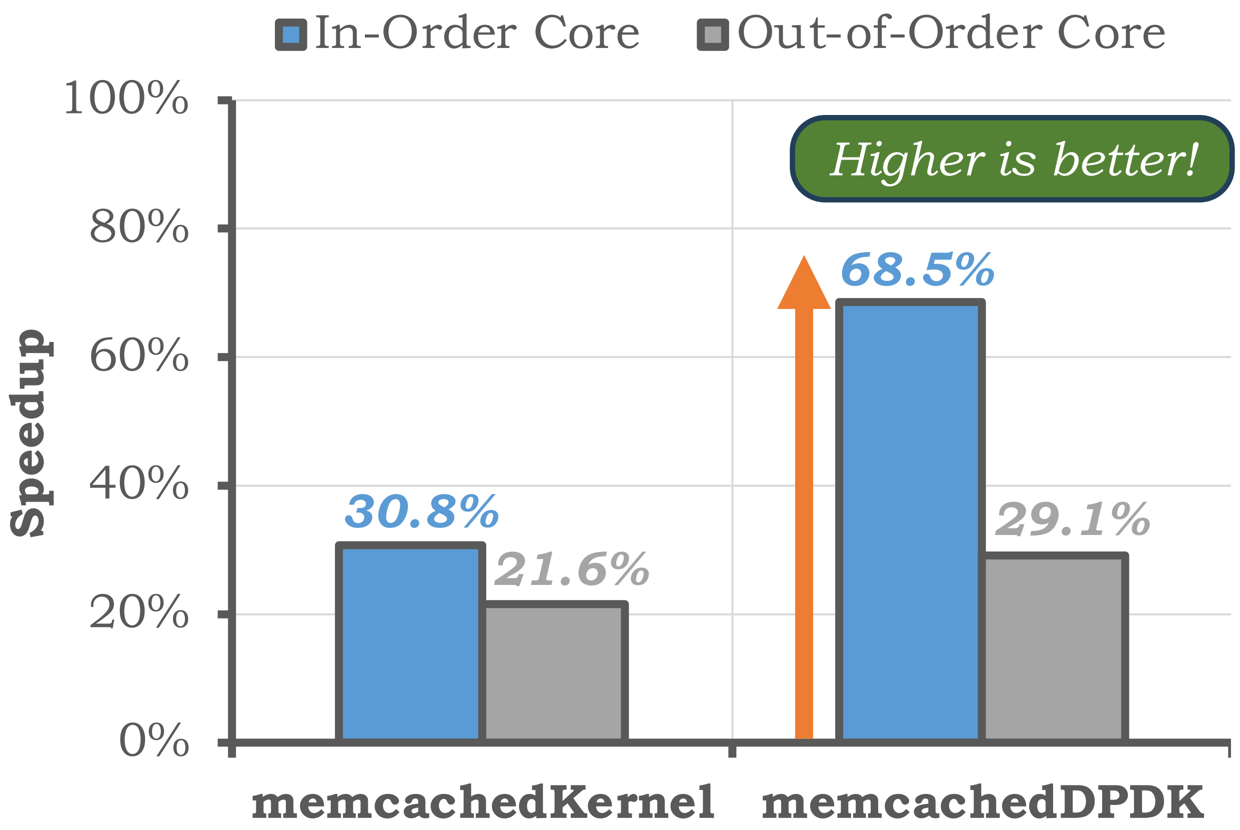
3 ***Does real-workload benefit from our userspace network stack in gem5?***

Benchmarking with Real Workloads



memcachedDPDK sustains 3.25x higher throughput than MemcachedDPDK

EtherLoadGen Speedup



EtherLoadGen speeds up simulation by ~68.5% running memcachedDPDK

Contributions

http://arg.ku.edu

1. Enabled gem5 to run userspace network applications
2. Included a hardware load generator model to gem5
3. Introduced a suite of 6 network intensive benchmarks
4. Enhanced gem5 to provide insights into packet processing bottlenecks
5. Performed sensitivity analysis of network intensive applications to microarchitectural settings

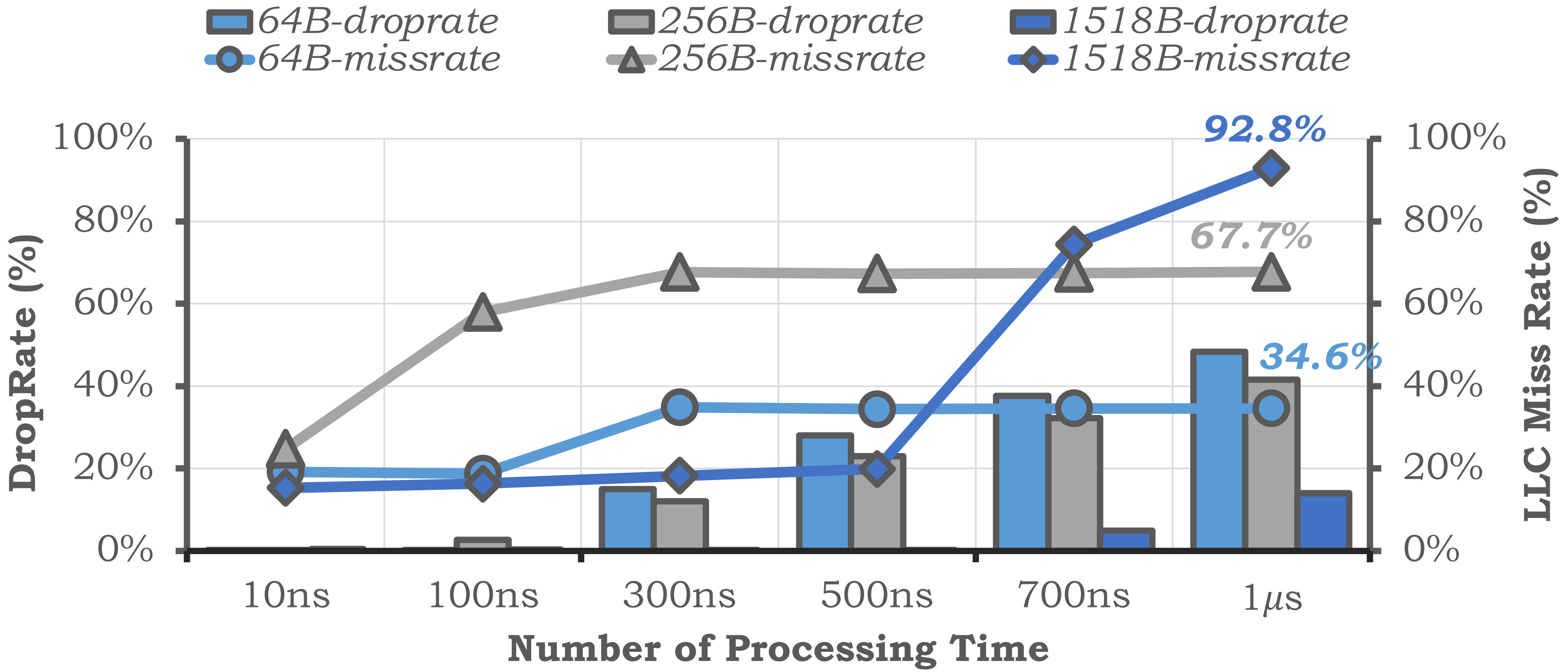
Thank You!

This work was supported in part by ACE, one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

Electrical Engineering and Computer Science

KU THE UNIVERSITY OF KANSAS

Backup Slides



As the DMA to core use distance increase, DMA leak is inevitable, leading to a higher likelihood for packet drop