Enabling Quality-of-Service in Nanophotonic Network-on-Chip

PENNSTATE
1855

Jin Ouyang and Yuan Xie

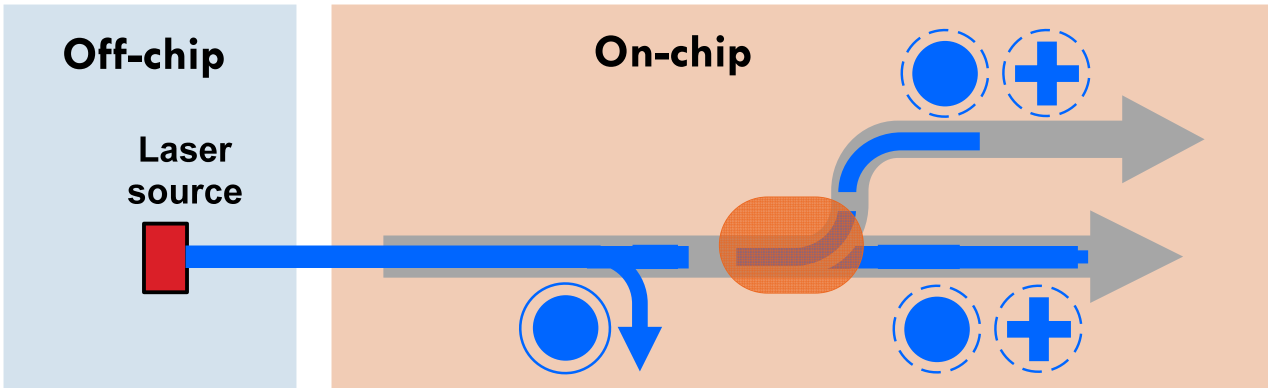
Microsystems Design Laboratory (MDL)
Dept. Computer Science & Engineering
Pennsylvania State University

Outline

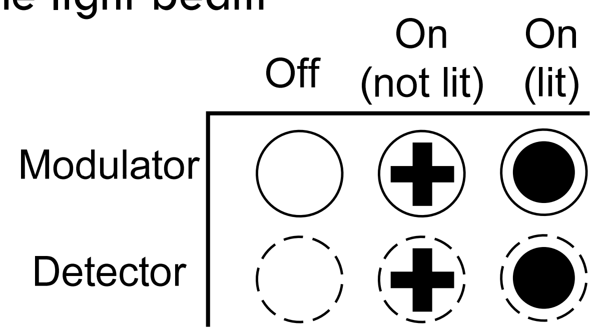
- Background and Motivation
- Baseline Optical NoC Architecture
- Quality-of-Service Support for Optical NoC
  - Principles of frame-based arbitration
  - Implementing frame-based scheduling *optically*
  - Optimization: Early frame-switching
- Experiment and Conclusion

On-chip optical interconnects

- **Components of on-chip optical interconnects:**



- **External laser source:** provides laser power.
- **On-chip waveguides:** carries and confines the light beam
- **Micro-rings:**
  - Out-of-resonance: let light pass by
  - In resonance: divert light from the waveguide
  - Modulation and detection (destructive)
- **Splitter:** broadcast



Motivation

- Various optical NoCs proposed, leveraging the power-efficiency of optical interconnects.
  - Direct networks: Shacham HOTI '07, NOCS'07; Gu CODES '08; Mo ISVLSI '10
  - Indirect networks: Gu DATE '09
  - Token-rings: Vantrease ISCA '08, MICRO '09; Pan MICRO '09; Zhang DAC '10
  - Bus: Kirman MICRO '06

- However, quality-of-service for optical on-chip network is absent.

- In this paper, we propose a simple and effective mechanism to provide QoS for optical NoC, leveraging optical frame-based scheduling.

Outline

- Background and Motivation
- Baseline Optical NoC Architecture
- Quality-of-Service Support for Optical NoC
  - Principles of frame-based arbitration
  - Implementing frame-based scheduling *optically*
  - Optimization: Early frame-switching
- Experiment and Conclusion

Baseline Optical NoC Architecture

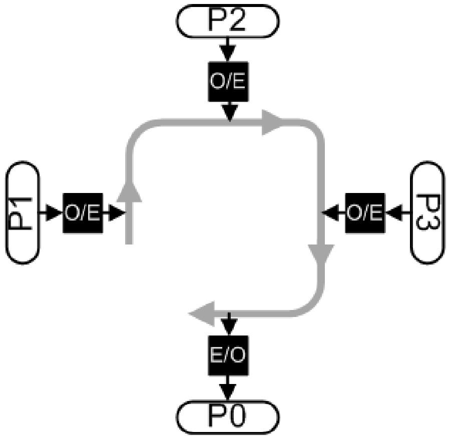
- Corona *†
  - Optical all-to-all crossbar, formed by multiple token-rings
  - For an N-node network, there are N token rings
- Features:
  - ✓ Efficient packet switching with low latency, area, and power overheads.
  - ✓ Efficient arbitration based on optical tokens
  - ✖ Poor fairness (upstream nodes always have higher priorities to obtain tokens).

* Vantrease, D. et al. "*Corona: System Implications of Emerging Nanophotonic Technology*", ISCA '08
† Vantrease, D. et al. "*Light speed arbitration and flow control for nanophotonic interconnects*", MICRO '09

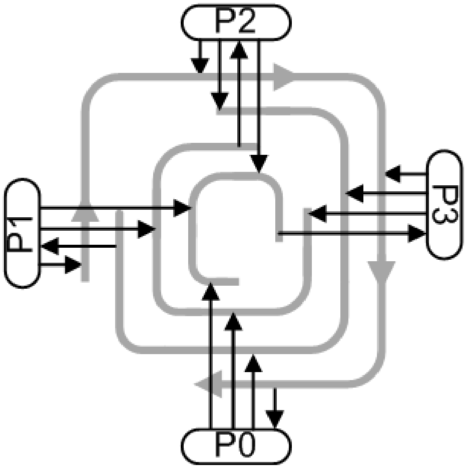
Multiple-Write Single-Read (MWSR) Rings

- Each token-ring has *multiple sources* but a *single destination*, called the *home node*.

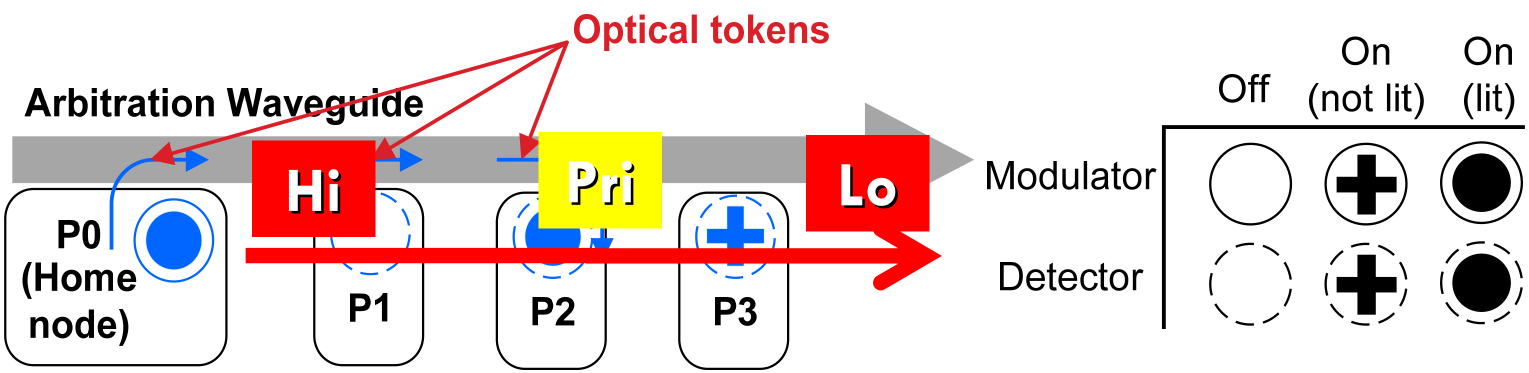
A single MWSR ring



Multiple MWSR rings for a 4-node network



- Basic optical token arbitration

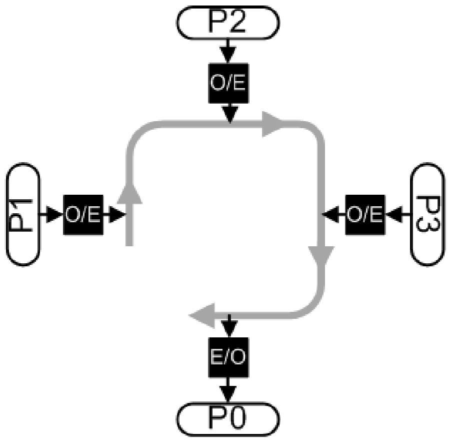


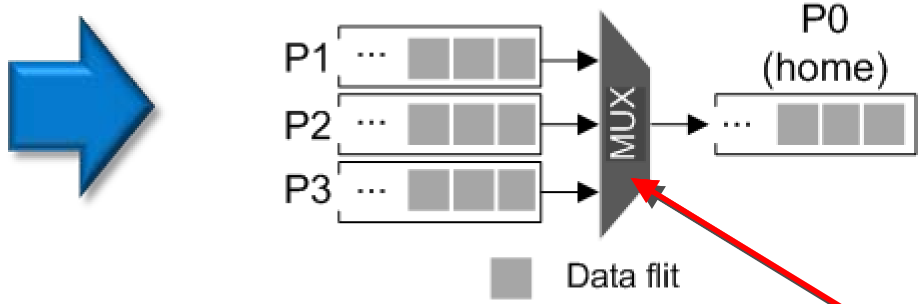
Outline

- Background and Motivation
- Baseline Optical NoC Architecture
- Quality-of-Service Support for Optical NoC
  - Principles of frame-based arbitration
  - Implementing frame-based scheduling *optically*
  - Optimization: Early frame-switching
- Experiment and Conclusion

Frame-Based Scheduling

- Frame-based scheduling
  - Simple and effective method to enable quality-of-service
  - Suitable for on-chip networks (GSF, ISCA ’08; PVC, MICRO ’09)
- Abstraction of a single MWSR ring





The multiplexer performs flit scheduling, dictating latency and bandwidth of each source.

With optical tokens, there is a strict ordering of priorities among P1, P2, P3 (P1 > P2 > P3).

Frame-based scheduling (con't)

- With frame-based scheduling, the queued flits are grouped into the so-called "frames".
  - A frame has a fixed size – F flits
  - A share assigned to each source node Pi – Ri flits

$$\sum_{Pi} Ri \leq F$$

- Example: (assuming $F=4$, $R1=1$, $R2=1$, $R3=2$)

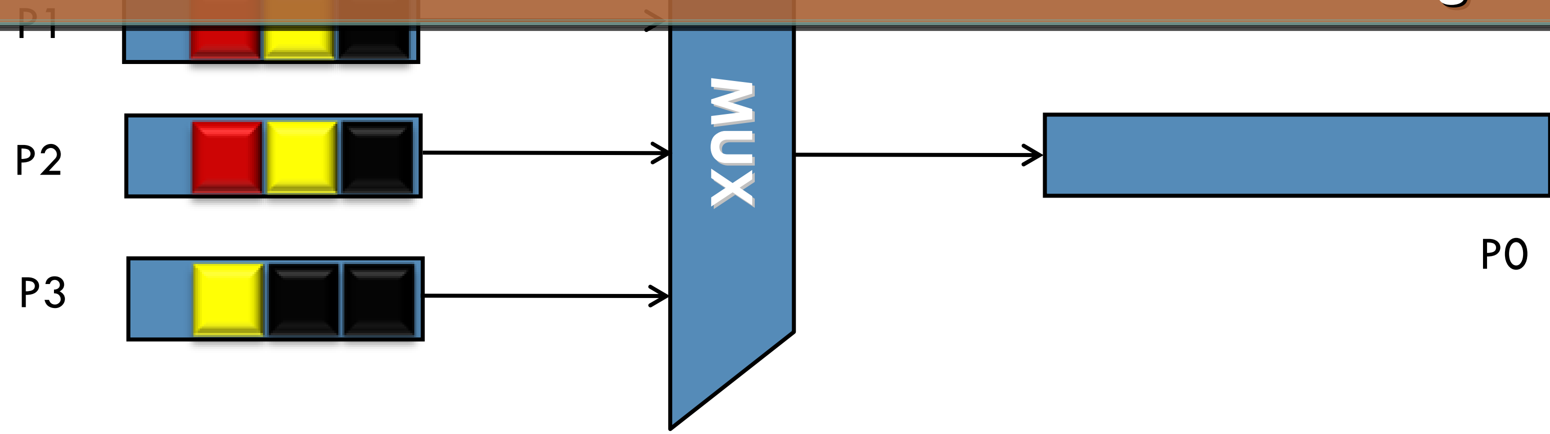
Frame-based scheduling (con't)

- **<u>Flits are serviced in the increasing order of frame number</u>**
  - Frame-based scheduling does not maintain any ordering inside a frame
  - However, with optical tokens, there is a strict ordering among flits from P1, P2, P3.

Bandwidth *Bi* received by node *Pi*:

$$Bi = Ri / F \times Bmax$$

where *Bmax* is the maximum bandwidth of the token-ring



How do we enforce frame-based scheduling in the optical token ring?

- **Two key steps:**

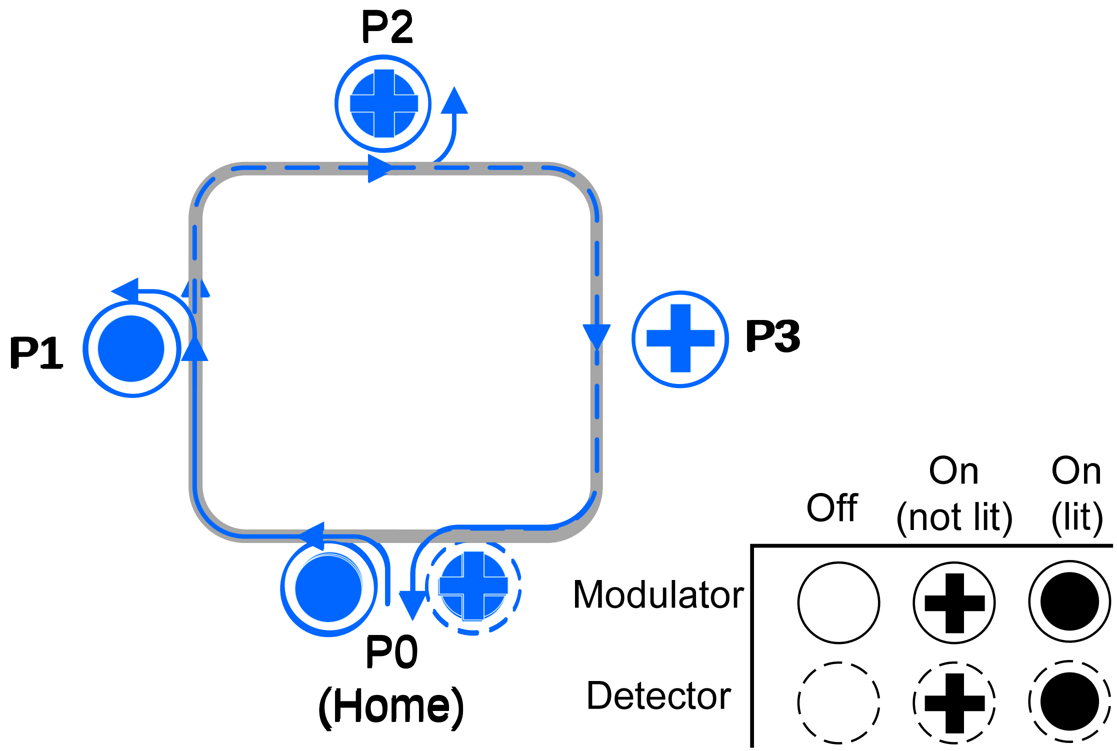
- 1. Mark the flits in the injection queue with different frame numbers (Easy)
  - Every source node maintains 2 local counters: IFi and Ci.
  - IFi: the frame number used to mark an incoming flit.
  - Ci: the remaining credits in frame IFi.

- 2. Request to send a flit only if frame-based ordering is not violated (Hard)
  - The oldest non-empty frame in the network (the *head frame*, or HF in short)
  - Request to send a flit only if it belongs to HF.
  - Every source node only knows *the head frame of itself*.
  - Need to synchronize HFs across the network.

The Completion Ring

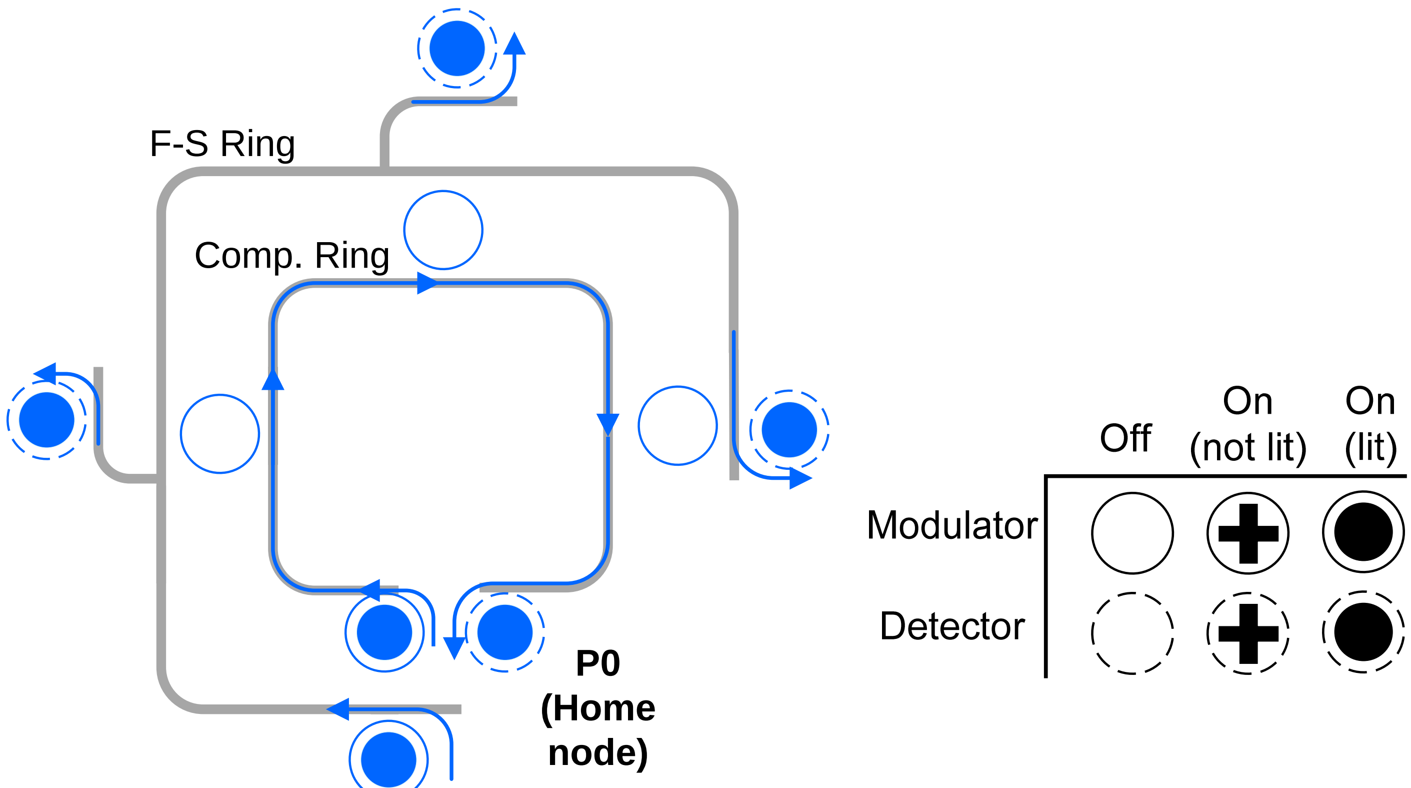
- A ring carrying a continuous laser sourced from the home node
  - Every source node has a micro-ring on this ring, originally all micro-rings are turned on.
  - The home node has a photo detector at the end of this ring.
  - When a source node finds its local HF is empty, it turns off its micro-ring.

1. The completion ring.
2. Initially, all micro-rings are on.
3. P1 and P3 finishe HF, turns of their micro-rings
4. P2 also finishes, turns off its micro-ring. Now the home node detects light from the completion ring.



The Frame-Switching Ring

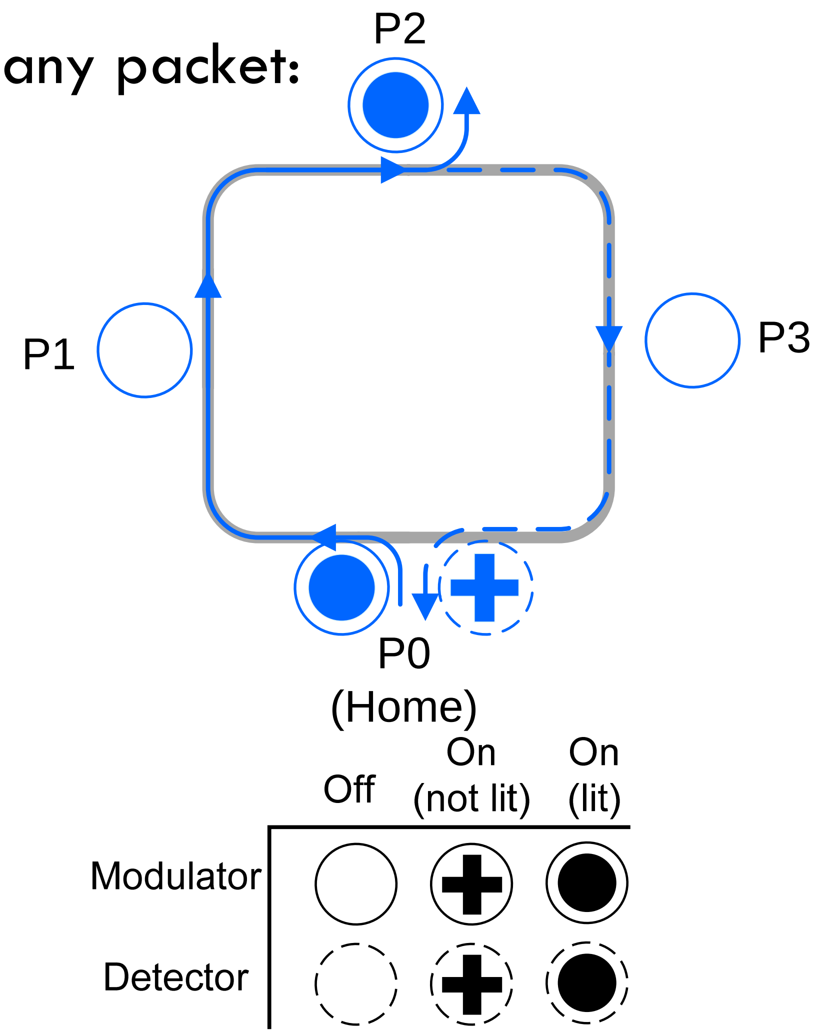
- A broadcast ring to signal incrementing the head frame number
- Once the home node detects light on the completion ring, it broadcasts frame-switching signal on this ring.



Optimization: Early Frame-Switching

- The risk of under-utilizing network capacity:
  - An inactive source node can block the whole network
- For example, P2 does not generate any packet:
  - P1, P3 are starved

- Addressing this problem:
  - Force early frame-switching, even when the frame is not drained
  - Speculate that a source node will continue to be idle if it has been idle for L cycles.



Outline

- Background and Motivation
- Baseline Optical NoC Architecture
- Quality-of-Service Support for Optical NoC
  - Principles of frame-based arbitration
  - Implementing frame-based scheduling *optically*
  - Optimization: Early frame-switching
- Experiment and Conclusion

Experiment

- Baseline architecture
  - 64-node Corona
  - Token-slot arbitration
  - 10GHz, 256-bit wide optical channel

  → 20TB/s bi-section bandwidth
- QoS enhanced optical NoC
  - Implement frame-based scheduling
  - Default frame size F: 128 flits
  - Empirically set L to 2 cycles.
- Cycle-accurate simulator modeling both networks
  - 8X8 network
  - Synthetic traffic patterns
  - Collect results when network statistics are stable

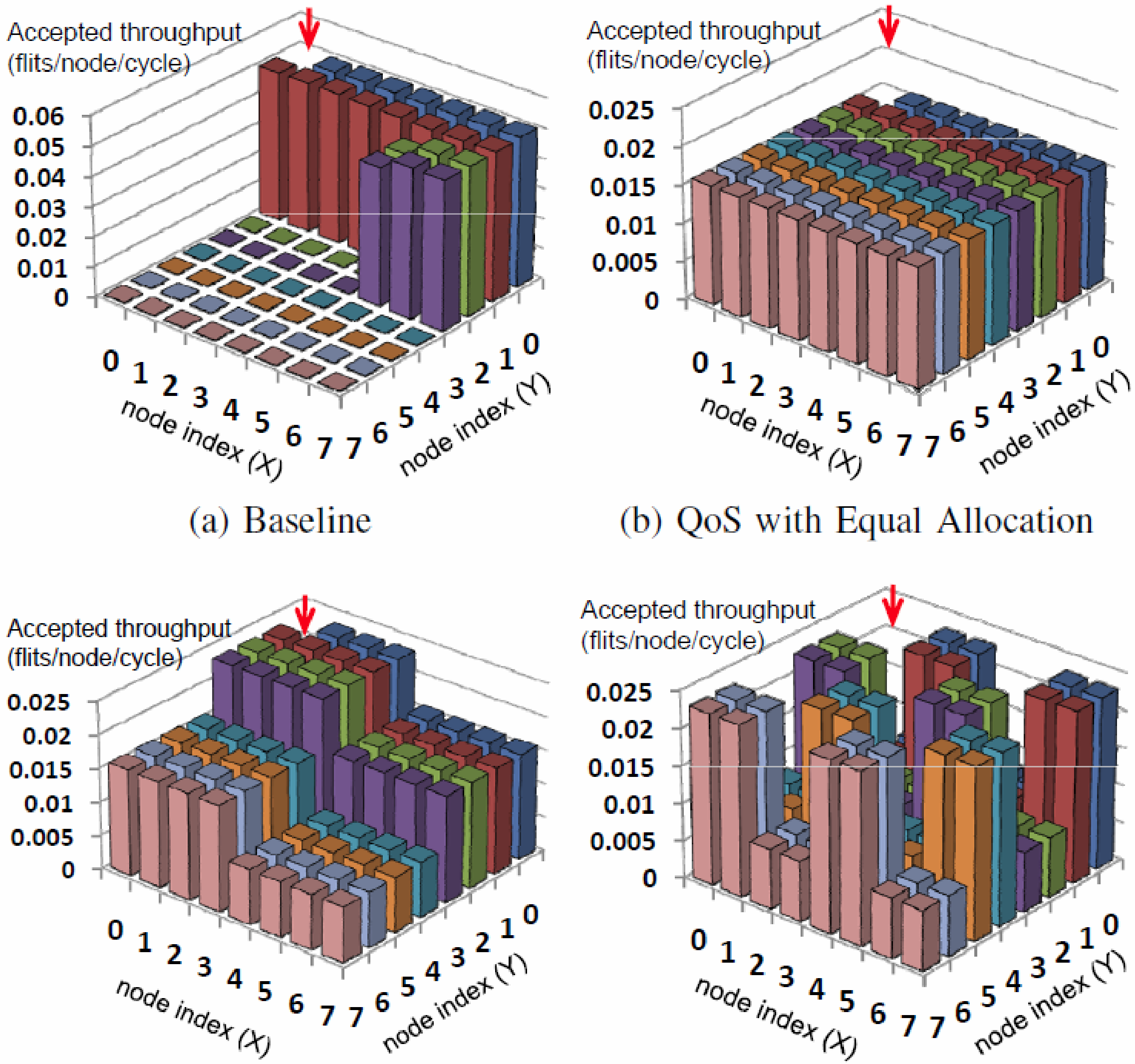
Hardware and Power Budget

Hardware and Power Budget table*

	Corona		QoS-Corona	
	Waveguides	Micro-rings	Waveguides	Micro-rings
Data MWSR rings	256	1024K	256	1024K
Arbitration ring	1	4K	1	4K
Comp. ring	0	0	1	4K
Frame-switch. ring	0	0	1	4K
<u>Total</u>	100%	100%	108%	108%
External laser	13.2W		14.1W	
Ring heating	26W		26.2W	
Ring modulation	50fJ		50fJ	

* Power consumptions are extracted from the paper by Zhang et al, DAC '10 for 22nm technology.

Quality-of-Service Results

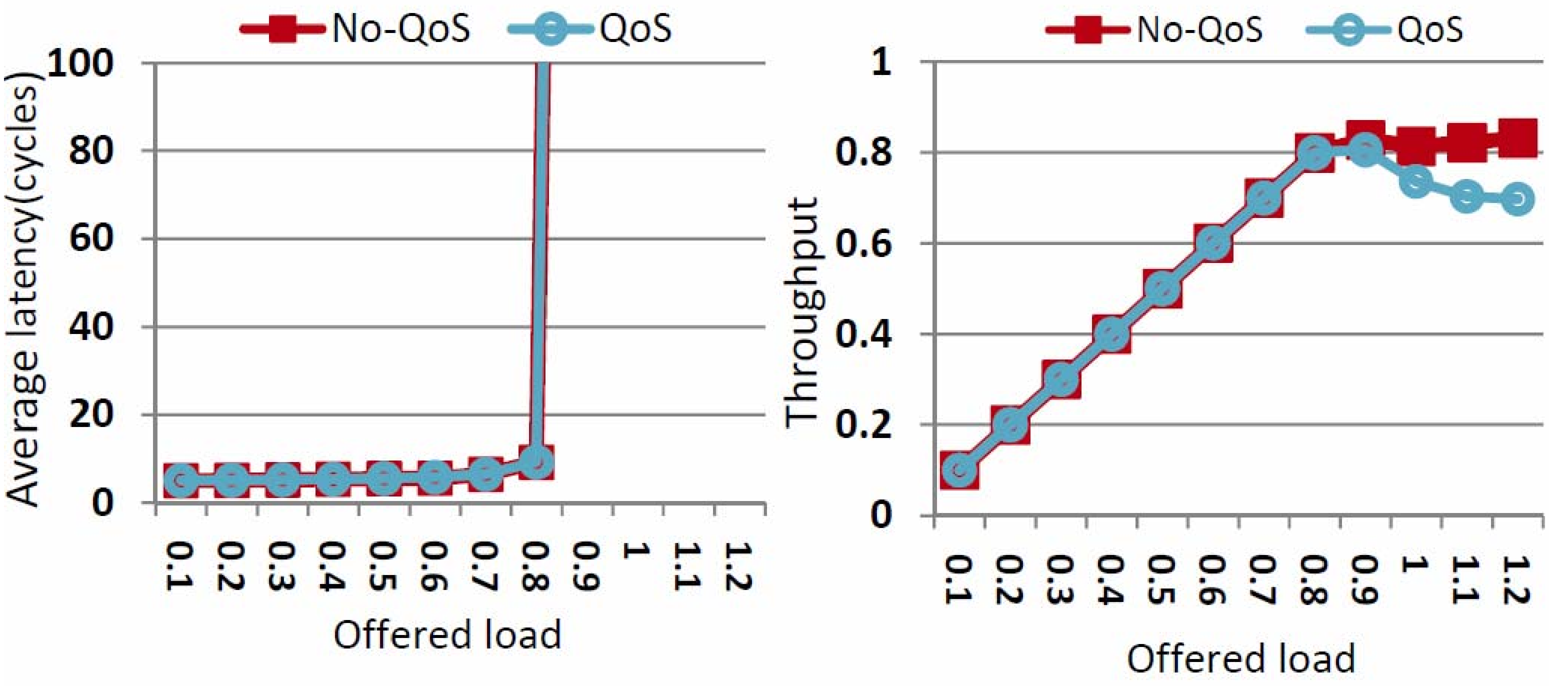


(a) Baseline (b) QoS with Equal Allocation

(c) Diff. allocation case 1 (d) Diff. allocation case 2

Performance overhead – Uniform

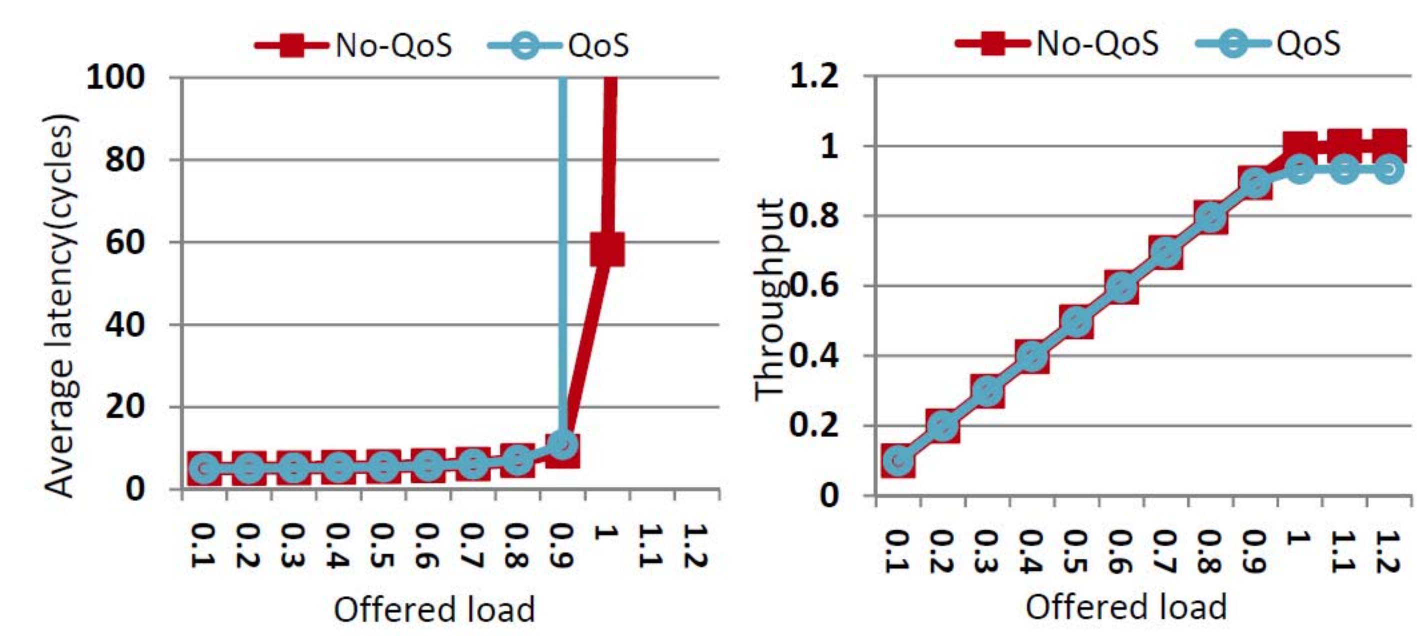
- Maximum throughput 17% lower than baseline



(a) Uniform

Performance overhead – Hotspot

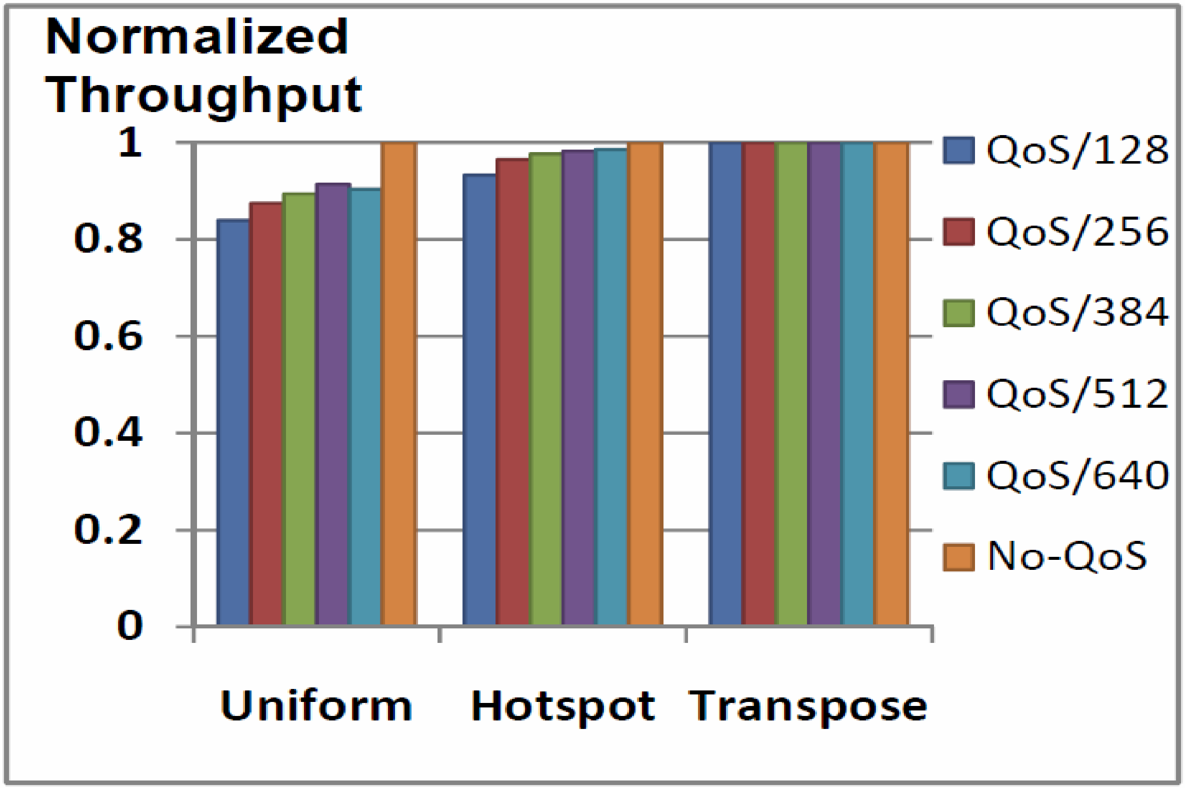
- Maximum throughput 7% lower than baseline



(b) Hotspot

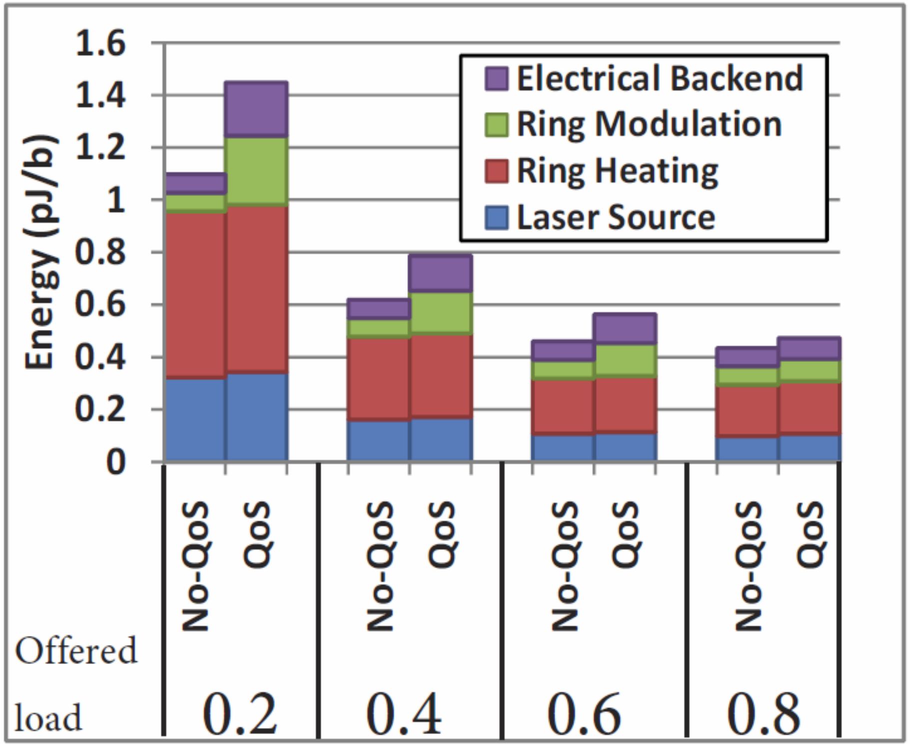
Performance overhead against frame size

- Frame size increases → Throughput increases
- When F=512 flits, 10% and 2% overheads for uniform and hotspot respectively.



Power overhead

- Power overhead mainly comes from dynamic power
  - Due to frame-switching overheads
- When load increases, dynamic overhead decreases
  - Frame-switching overhead is amortized
  - 32% for a rate of 0.2, 8% for a rate of 0.8



Conclusion

- A QoS-enabled optical NoC with frame-based scheduling is presented.
- Effective in providing strong bandwidth allocation
- Low performance overhead
  - 17% with 128-flit frame
  - 10% with 512-flit frame
- Power overhead depends on loads
  - Favors heavily loaded network
  - Future work: adaptively adjust early frame-switching frequency to further reduce power overhead, especially for low loads.

Thank You!
Questions?

Notations

- Recap:
  - F: frame size in flits
  - Ri: the share of a frame assigned to node Pi.

- New notations
  - HF: the current *lowest number frame* in the network.
  - IFi: the frame being used by node Pi to inject flits.
  - Ci: the remaining share of node Pi in IFi.
  - L: the node idle time to trigger *early frame switching*.

Implementing Frame-Based Scheduling in Optical NoC

- Each source node Pi maintains two counters
  - ***IFi***: Current injection frame number
  - ***Ci***: Remaining credits in the current injection frame
- Each source node Pi runs 2 procedures:
  - *1. Group flits into frames:*

    On the generation of a flit:

    If $Ci = 0$ then

    $IFi = IFi + 1$;

    $Ci = Ri$;

    end if;

    Mark the flit with IFi; $Ci = Ci - 1$;

    Put the flit in the injection queue
  - *2. Injection admission:*

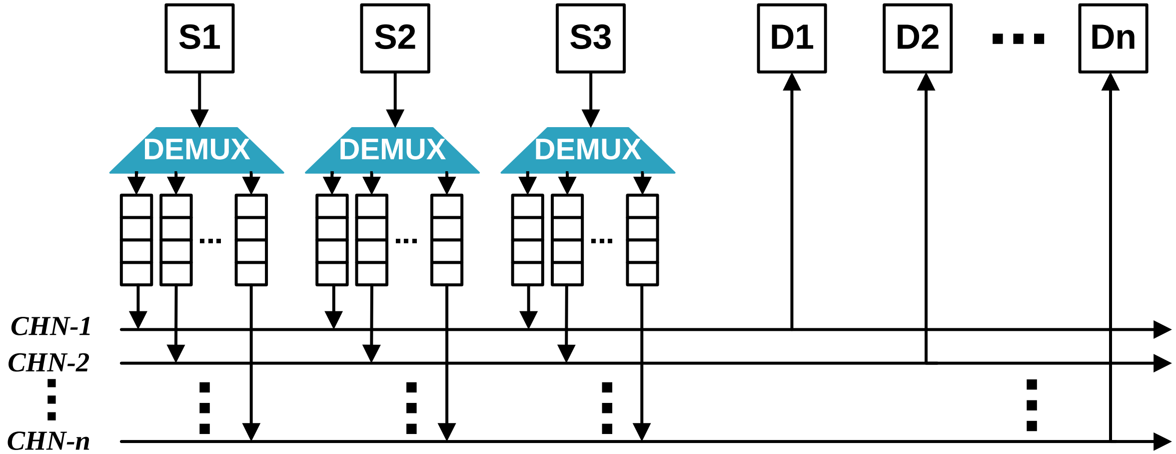
    If *there is a flit at the head of the injection queue*

    && *the flit's frame number is equal to HF* then

    Request for optical tokens in order to send this flit.

Put it together – Multiple token rings

- The discussion so far only considers a single token ring.
- For multiple token rings, virtual output queue (VOQ) is used
  - The source node has multiple injection queues, each for a different destination.



- Each ring runs an independent copy of frame-based scheduling algorithm.
  - No interaction among channels.