An Intra-Chip Free-Space Optical Interconnect

Jing Xue, Alok Garg, Berkehan Ciftcioglu, Jianyun Hu, Shang Wang, Ioannis Savidis,
Manish Jain, Rebecca Berman, Peng Liu,
Michael Huang, Hui Wu, Eby Friedman, Gary Wicks, and Duncan Moore

Department of Electrical and Computer Engineering
The Institute of Optics
University of Rochester

Motivation

- Continued, uncompensated wire scaling degrades performance and signal integrity
- Optics has many fundamental advantages over metal wires and is a promising solution for interconnect
- Optics as a drop-in replacement for wires inadequate
  - Optical buffering or switching remains far from practical
  - Packet-switched network architecture requires repeated O/E and E/O conversions
  - Repeated conversions significantly diminish benefits of optical signaling (especially for intra-chip interconnect)

  $\Rightarrow$ Conventional packet-switched architecture ill-suited for on-chip optical interconnect

Challenges for On-chip Optical Interconnect

- Signaling chain:
  - Efficient Si E/O modulators challenging
    - Inherently poor non-linear optoelectronic properties of Si
    - Resonator designs also non-ideal: e.g., e-beam lithography, temperature stability, insertion loss
  - Off-chip laser (expensive, impractical to power gate)
- Propagation medium:
  - In-plane waveguides add to the challenge and loss
    - Floor-planning, losses due to crossing, turning, and distance
  - Bandwidth density challenge
    - Density of in-plane wave guide limited
    - WDM: more stringent spectral requirements for devices and higher insertion losses, more expensive laser sources

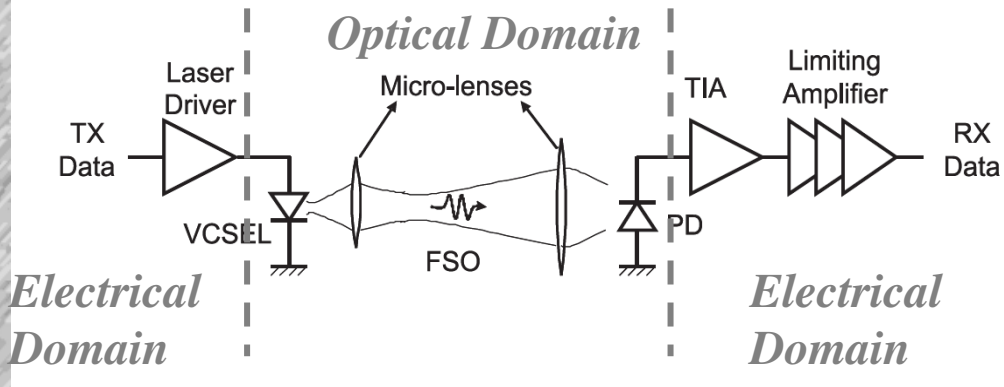
Free-Space Optical Interconnect: an Alternative

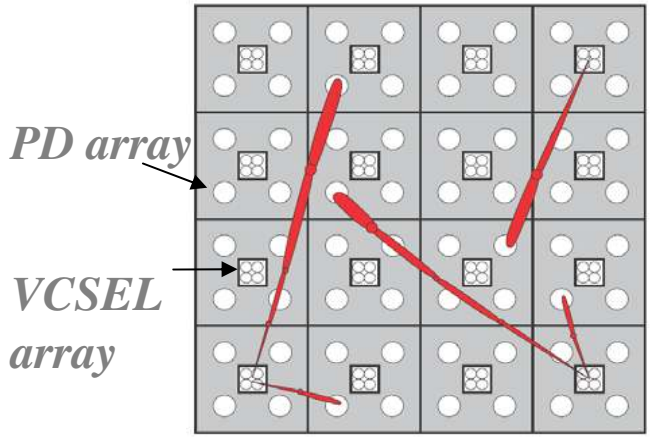
- Signaling
  - + Integrated VCSELs (Vertical Cavity Surface Emitting Laser) avoids the need for external laser and optical power distribution; fast, efficient photo detectors
  - – Disparate technology (e.g., GaAs)
- Propagation medium
  - + Free-space: low propagation delay, low loss and low dispersion
  - – Hindering heat dissipation
- Networking
  - + Direct communication: relay-free, low overhead, no network deadlock or the necessity to prevent it

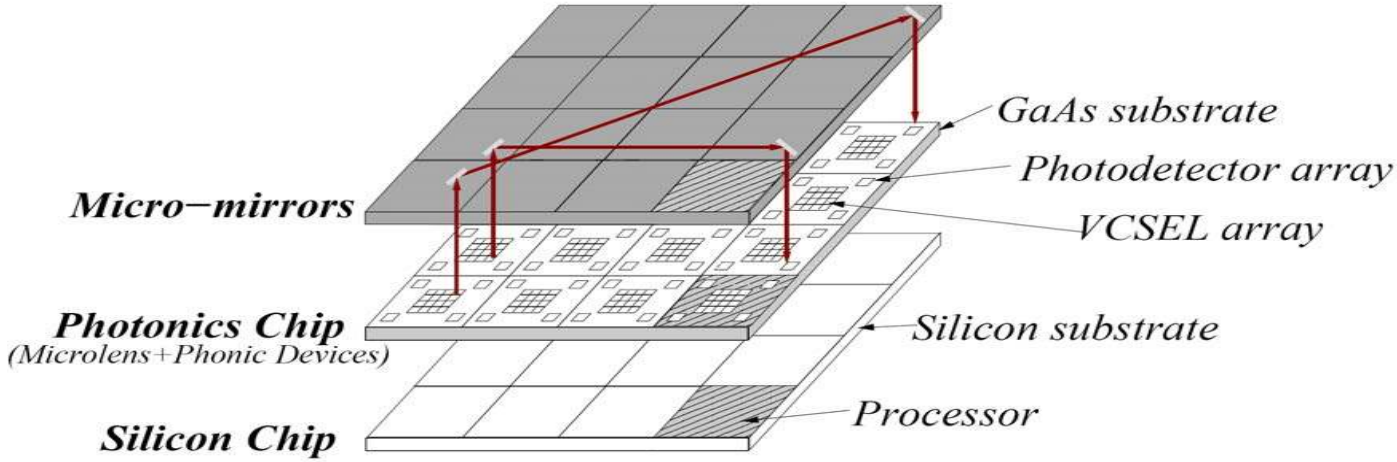
Outline

- System overview
- Interconnect architecture
- Optimization
- Evaluation
- Conclusion

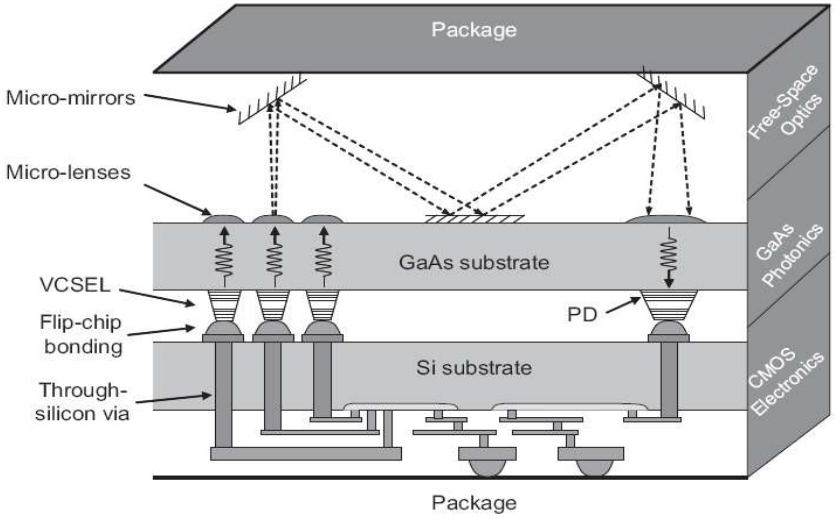
Optical Link and System Structure



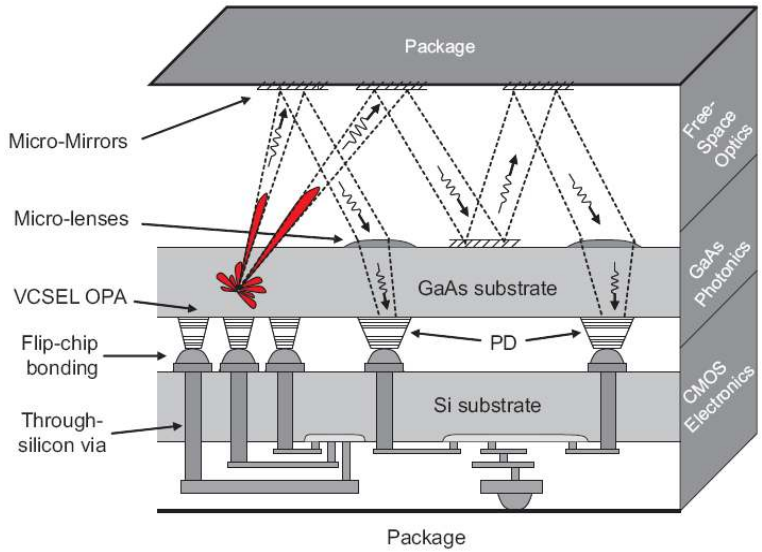




Chip Side View



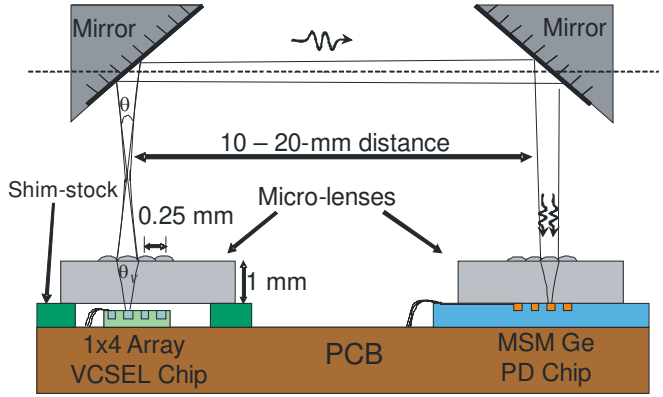
Side view (mirror-guided only)

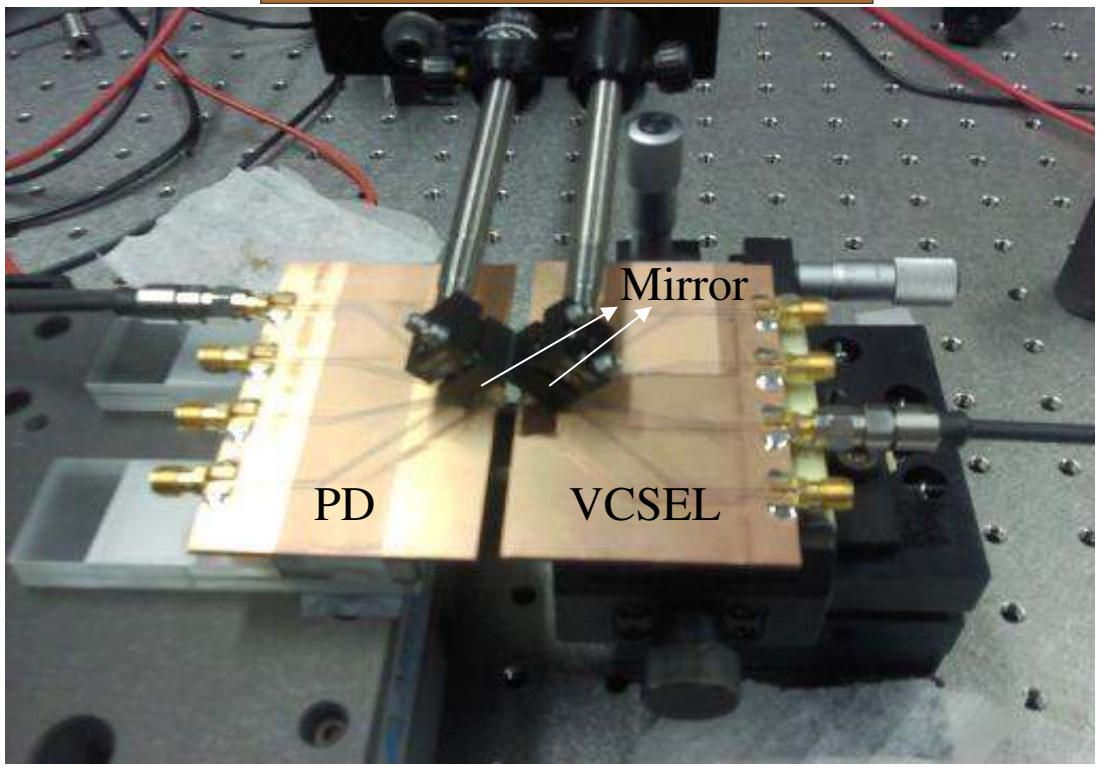


Side view (with phase array beam-forming)

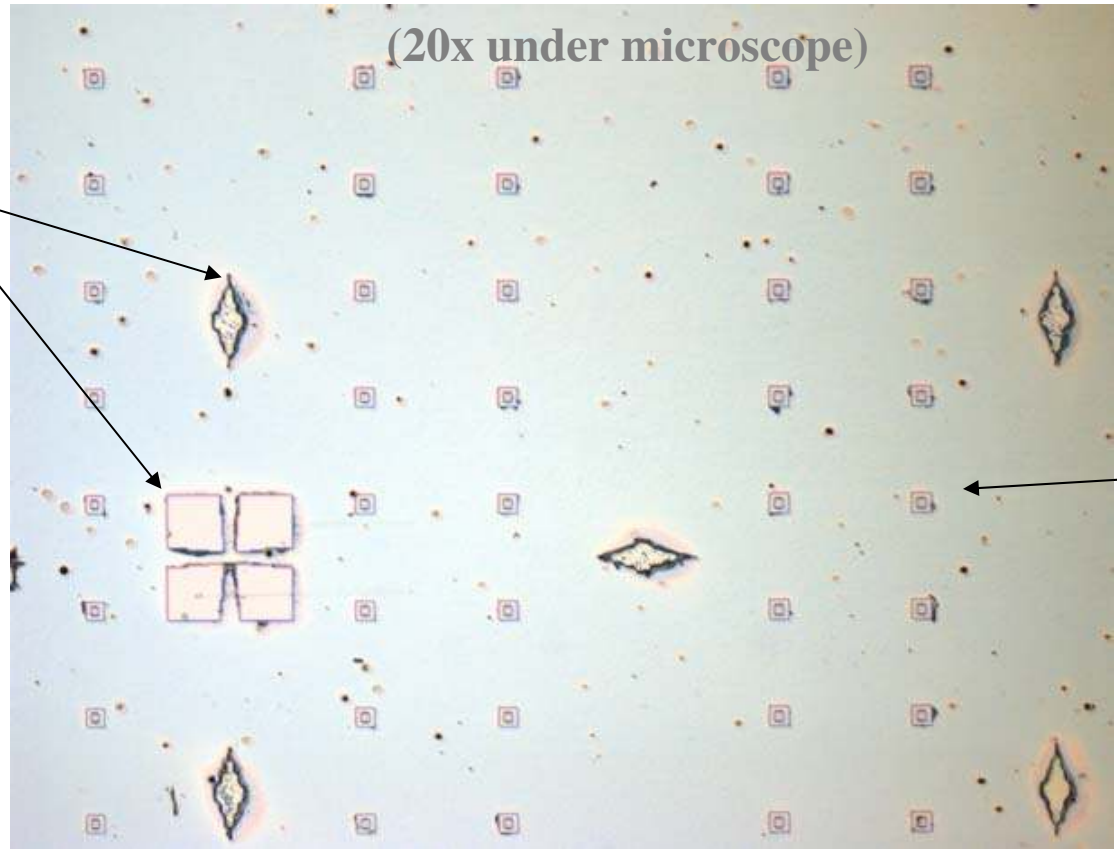
- Mostly current (commercially available) technology
  - Large VCSEL arrays, high-density (movable) micro mirrors, high-speed modulators and PDs
- Efficiency: integrated light source, free-space propagation, direct optical paths

Link Demo on Board Level



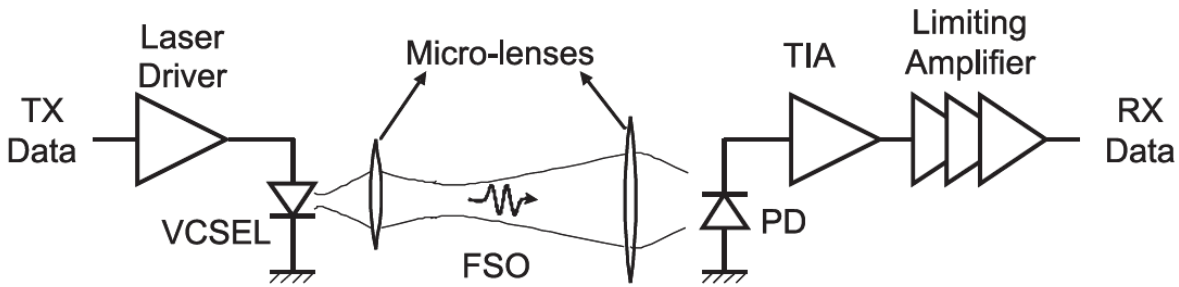


Prototype Custom-Made VCSEL Arrays



Photograph of VCSEL mesa structure

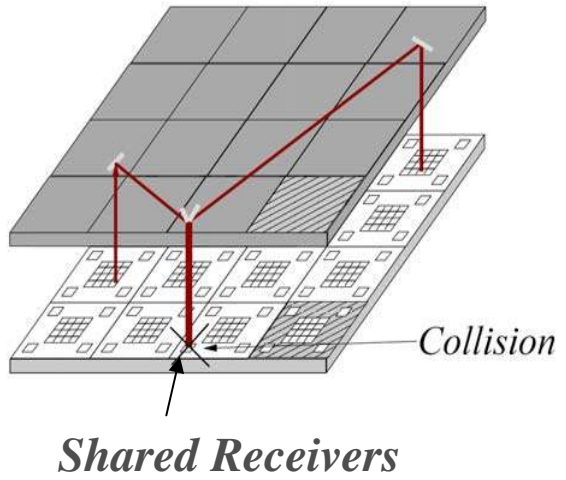
Efficient Optical Links



Free-Space Optics	
Transmission distance	2 cm
Optical wavelength	980 nm
Microlens aperture	90 μm @ transmitter, 190 μm @ receiver
Optical path loss	2.6 dB
Transmitter & Receiver	
VCSEL	aperture=5 μm, parasitic=235 Ω // 90 fF
	threshold=0.14 mA, extinction ratio=11:1
Laser driver	bandwidth=43 GHz
PD	responsivity=0.5 A/W, capacitance=100 fF
TIA & Limiting amp	bandwidth=36 GHz, gain=15000 V/A
Link	
Data rate	40 Gbps
Signal-to-noise ratio	7.5 dB
Bit-error-rate (BER)	10^{-10}
Cycle-to-cycle jitter	1.7 ps
Power Consumption	
Transmitter (active)	0.96 mW for VCSEL (0.48 mA@2V)
	6.3 mW for laser driver
Transmitter (standby)	0.43 mW
Receiver	4.2 mW

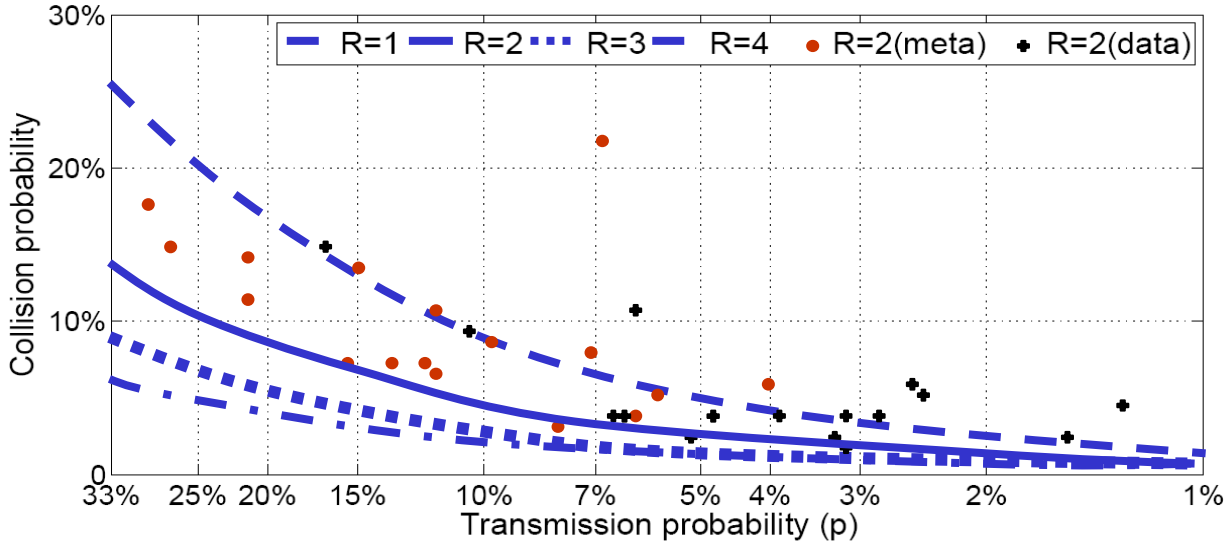
Network Design

- Allowing collisions: a central tradeoff
  - Avoid centralized arbitration
    - Improve scalability
    - Reduce arbitration latency for common case
    - Reduce the cost for arbitration circuitry
  - Same mechanism to handle errors
    - No extra support to handle collisions
    - Once collisions accepted can lower BER requirements (more engineering margins and/or energy optimization opportunities)
  - No significant over-provisioning necessary (later)
  - Simple structuring steps reduce collisions

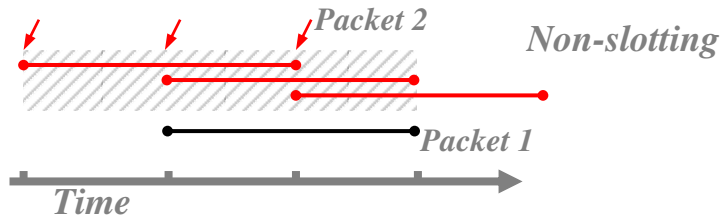


Structuring for Collision Reduction

- Multiple receivers $1-\left[\left(1-\frac{p}{N-1}\right)^{n}+\binom{n}{1}\frac{p}{N-1}\left(1-\frac{p}{N-1}\right)^{n-1}\right]^{R}$ $n=(N-1)/R$ Number of nodes sharing a receiver



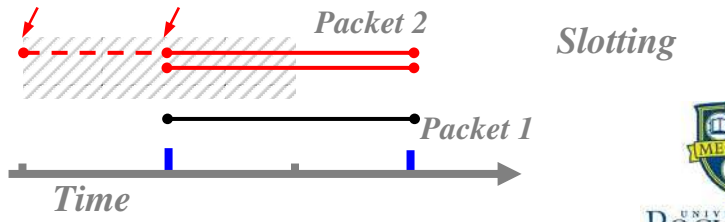
- Slotting and lane separation
  - Meta Packets
  - Data Packets



- Bandwidth allocation

$$\frac{C_1}{B_M}+\frac{C_2}{B_M^2}+\frac{C_3}{1-B_M}+\frac{C_4}{(1-B_M)^2}$$

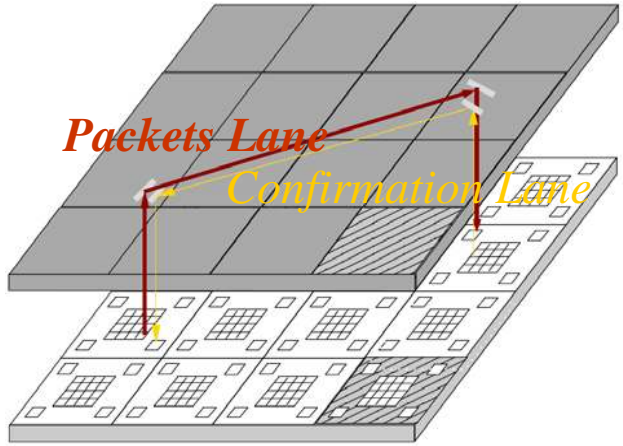
$$B_M = 0.285$$



Collision Handling

- Detection mechanism (at receiver)

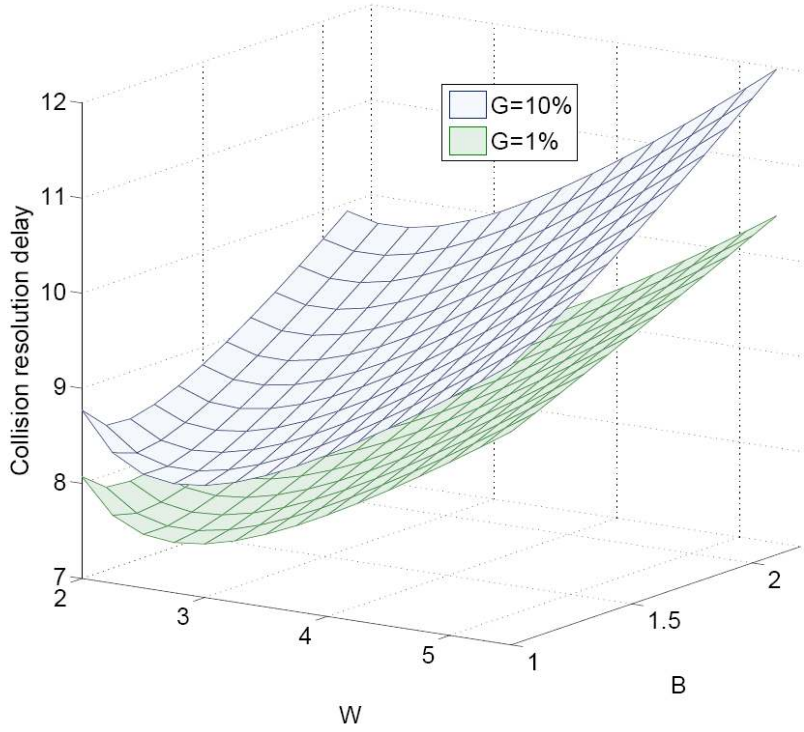
	PID	$\overline{\text{PID}}$
Node A	- - 1 -	- - 0 -
Node B	- - 0 -	- - 1 -
Received	- - 1 -	- - 1 -



- Notification/inference of collision at transmitter: confirmation
  - Dedicated VCSEL per lane
  - Collision free for confirmations
  - Allows coherence optimization

- Retransmission to guarantee eventual delivery
  - Exponential back-off: $W_r = W \times B^{r-1}$

    $W = 2.7$, $B = 1.1$ for minimal collision resolution delay



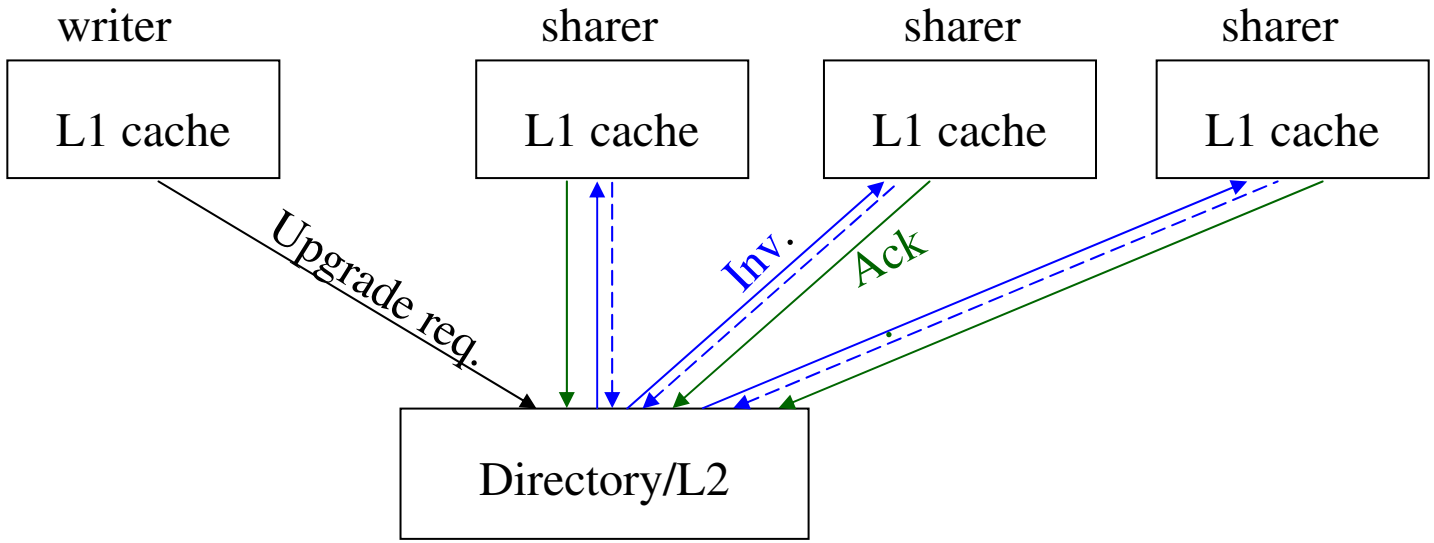
Optimizations: Leveraging Confirmation Signals

- Conveying timing information
  - Sometimes the whole point of communication is *timing*
  - E.g., releasing lock/barrier, acknowledging invalidation
  - Information content low (esp. when message is anticipated)
  - Inefficient use of bandwidth (~25% traffic for sync in 64-way CMP)

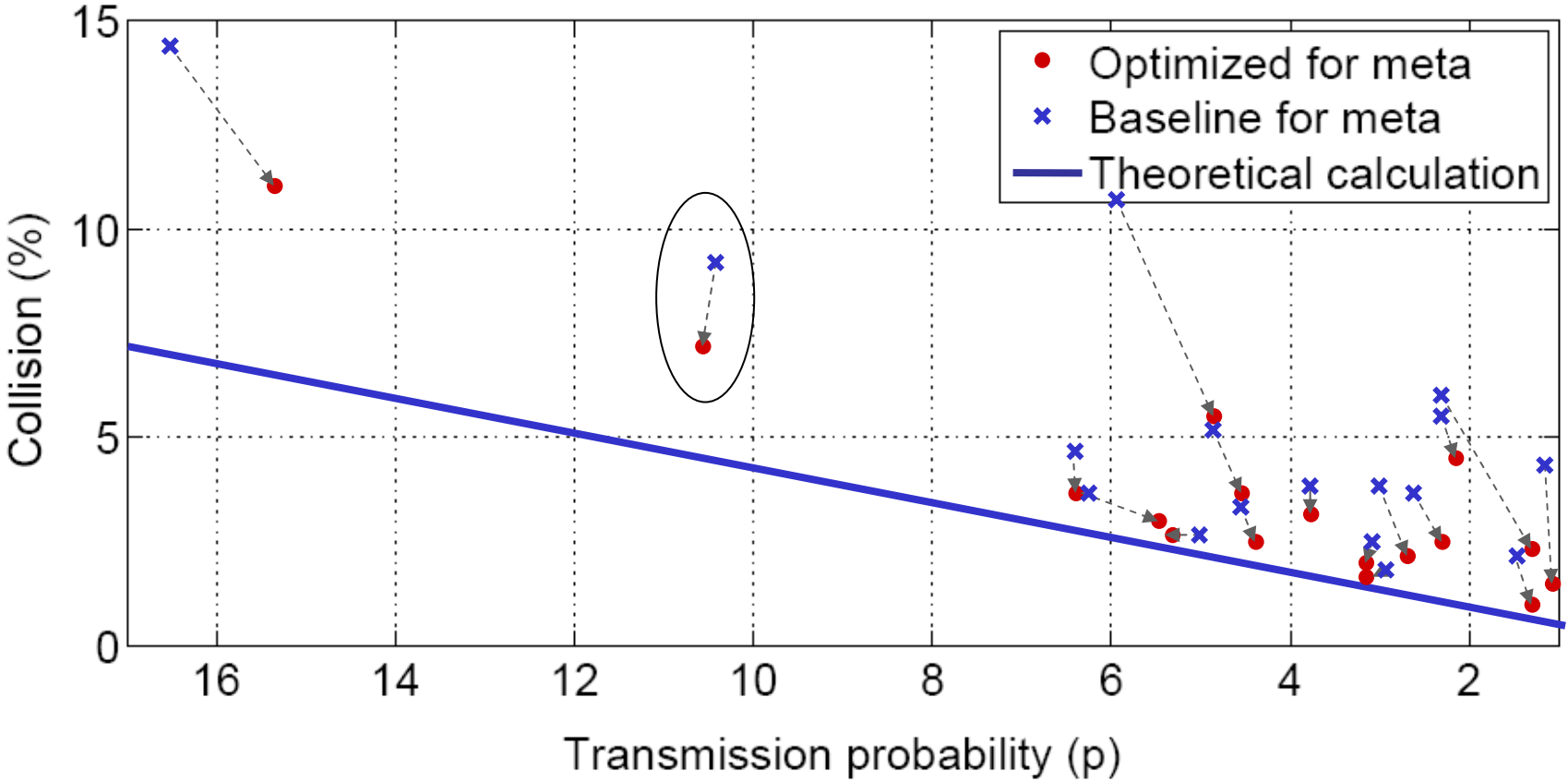
- Confirmation laser can provide the communication
  - Achieve even lower latencies than using full-blown packets (such communication is often latency sensitive)
  - Reduce traffic on regular channels and thus collision
  - Eliminate invalidation acknowledgement
  - Specialized boolean value communication

Eliminating Acknowledgements

- Acknowledgements needed for (global) write completion
  - For memory barriers, to ensure write atomicity, etc.
- Use confirmation as commitment
  - Only change: received invalidation is logically serialized before another visible transaction (same as some bus-based designs)
  - Avoid acks which are particularly prone to collisions



Eliminating Acknowledgements



Reduces 5.1% traffic but eliminates 31.5% of meta packet collisions

Invalidation acknowledgements systemically synchronized

Experimental Setup

Memory hierarchy	
L1 D cache (private)	8KB, 2-way, 32B line, 2 cycles, 2 ports, dual tags
L1 I cache (private)	32KB, 2-way, 64B line, 2 cycles
L2 cache (shared)	64KB slice/node, 64B line, 15 cycles, 2 ports
Dir request queue	64 entries
Memory channel	52.8GB/s bandwidth, memory latency 200 cycles
Number of channels	4 in 16-node system, 8 in 64-node system
Prefetch logic	Stream prefetcher
Network packet	Flit size: 72-bit, data packets: 5 flits, meta packet: 1 flit
Wire interconnect	4VCs, latency: router 4 cycles, link 1 cycle, buffer: 5x12 flits

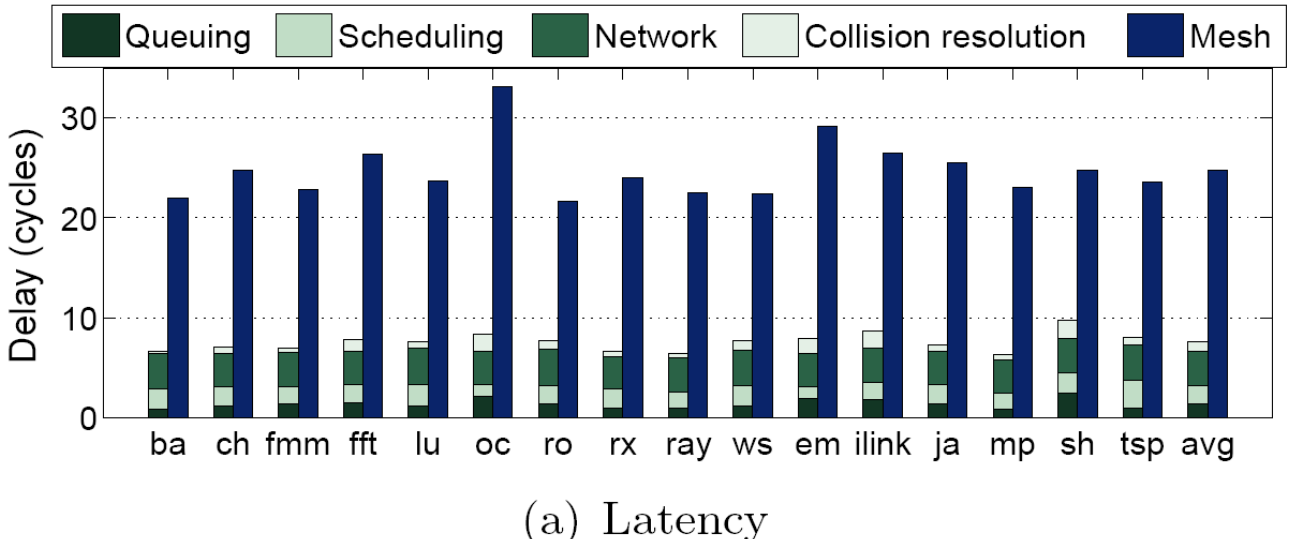
Processor core	
Fetch/Decode/Commit	4/4/4
ROB	64
Functional units	INT 1+1 mul/div, FP 2+1 mul/div
Issue Q/Reg.(int, fp)	(16, 16)/(64, 64)
LSQ(LQ, SQ)	32 (16, 16) 2 search ports
Branch predictor	Bimodal + Gshare
-Gshare	8K entries, 13bit history
-Bimodal/Meta/BTB	4K/8K/4K (4-way) entries
Br.mispred.penalty	At least 7 cycles
Process specifications	Feature size: 45nm, f_{clk}: 3.3GHz, V_{dd}:1V

Optical Interconnect (each node)	
VCSEL	40GHz, 12 bits per CPU cycle
Array	Dedicated (16-node), phase-array with 1 cycle setup delay (64-node)
Lane widths	6/3/1 bit(s) for data/meta/confirmation lane
Receivers	2 data (6b), 2 meta (3b), 1 for confirmation (1b)
Outgoing queue	8 packets each for data and meta lanes

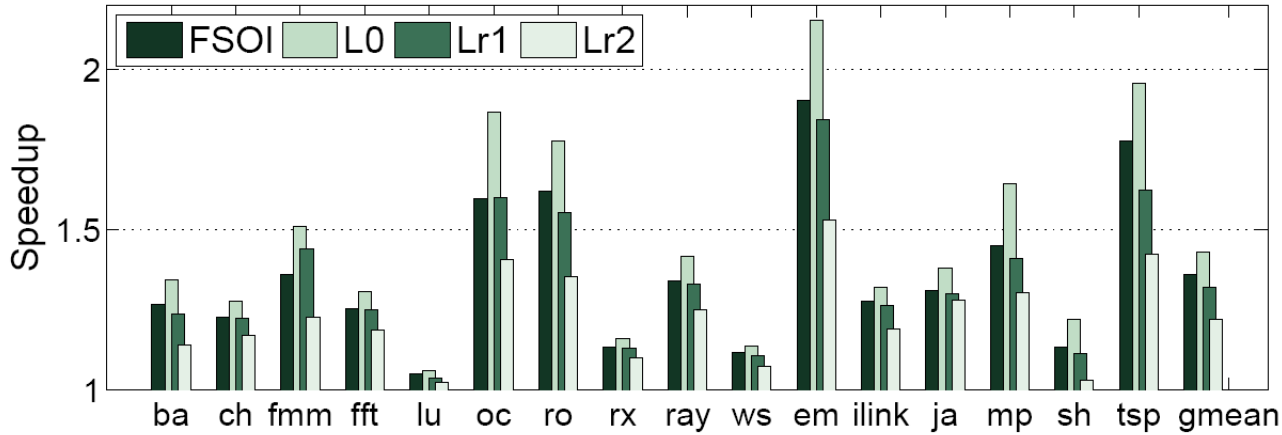
Applications: SPLASH 2 suite, electromagnetic solver (*em3d*), genetic linkage analysis (*ilink*), iterative PDE solver (*jacobi*), 3D particle simulator (*mp3d*), weather prediction (*shallow*), branch and bound based NP traveling salesman problem (*tsp*)

Performance – 16 Cores

- FSOI offers low latency
- Collisions do not add excessive latencies
- Speedup depends on code, but tracks L_0 (1.36 vs 1.43)
- Better than idealized single-cycle router mesh



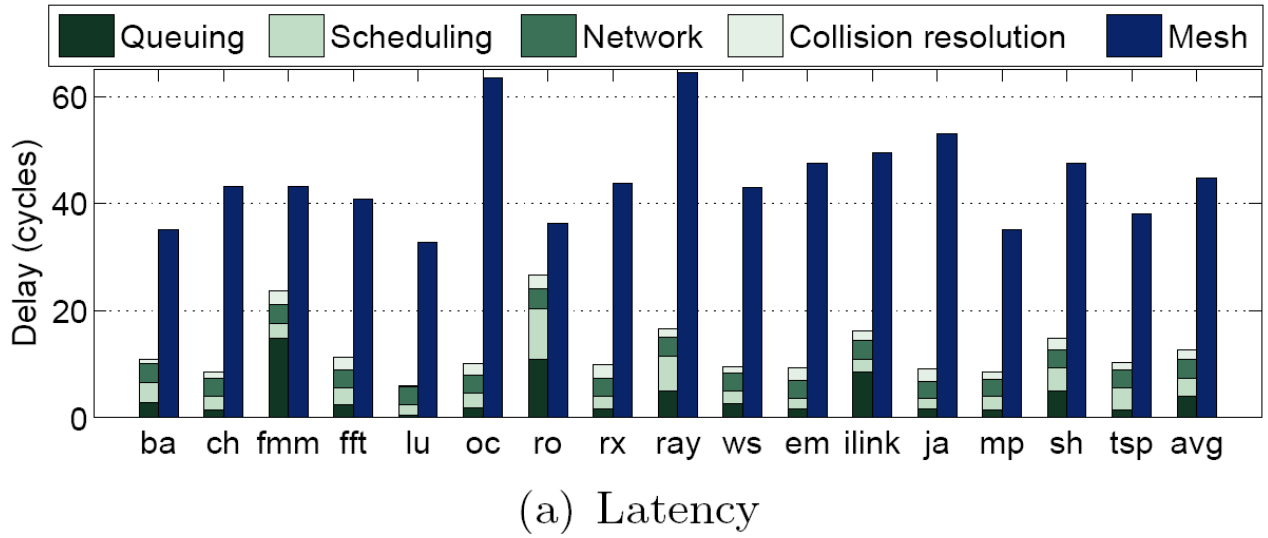
(a) Latency



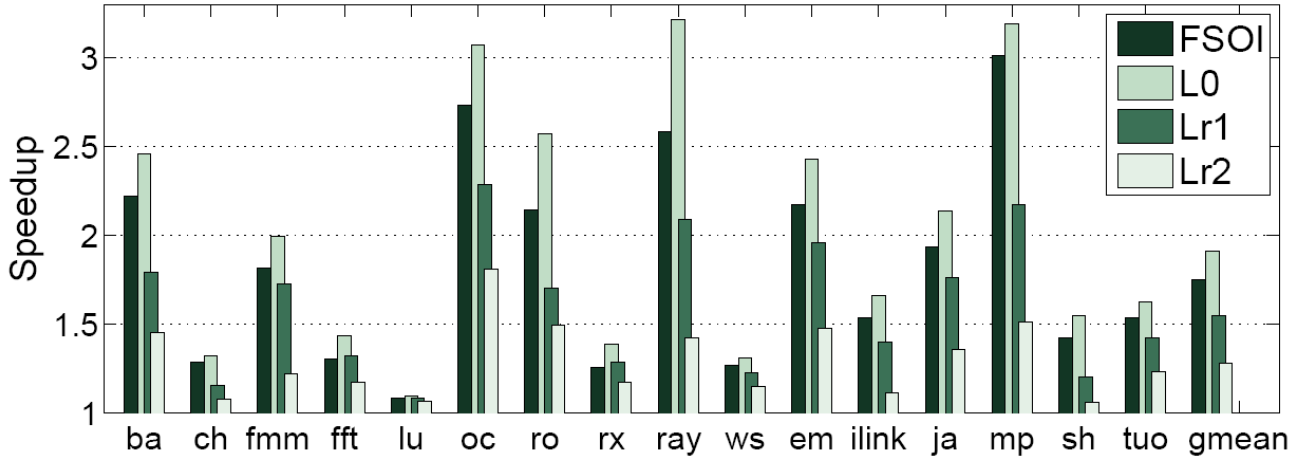
(b) Speedup

Performance – 64 Cores

- Latency does increase, but mostly due to source queuing
- Speedup continues to track that of L_0 (1.75 vs 1.90) and pulls further ahead of L_{r1}, L_{r2}



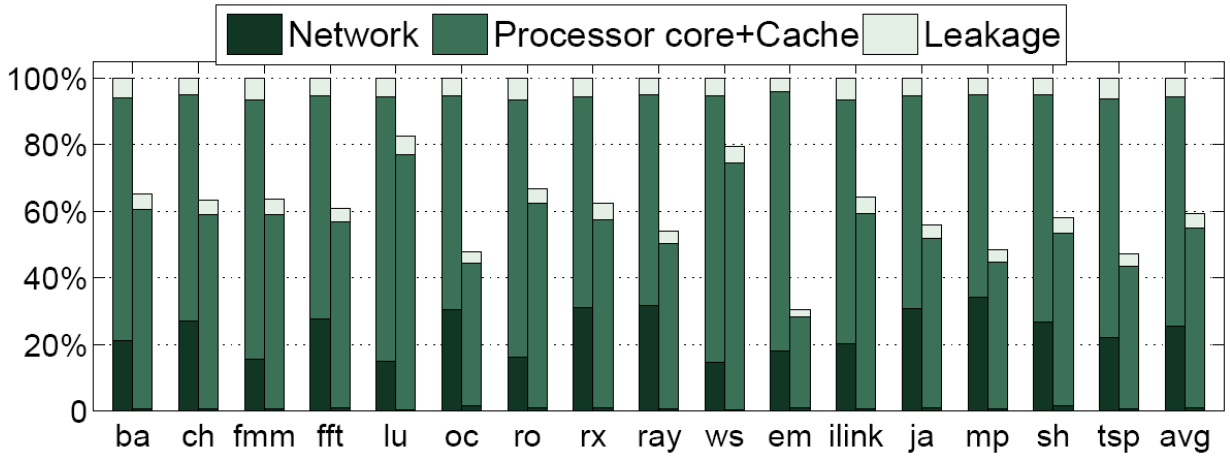
(a) Latency

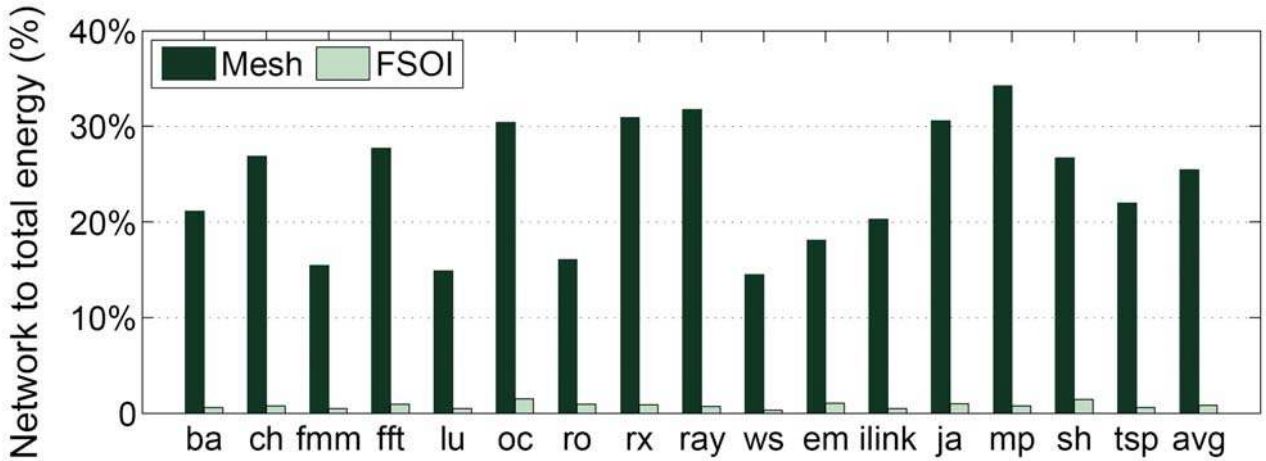


(b) Speedup

Energy Analysis

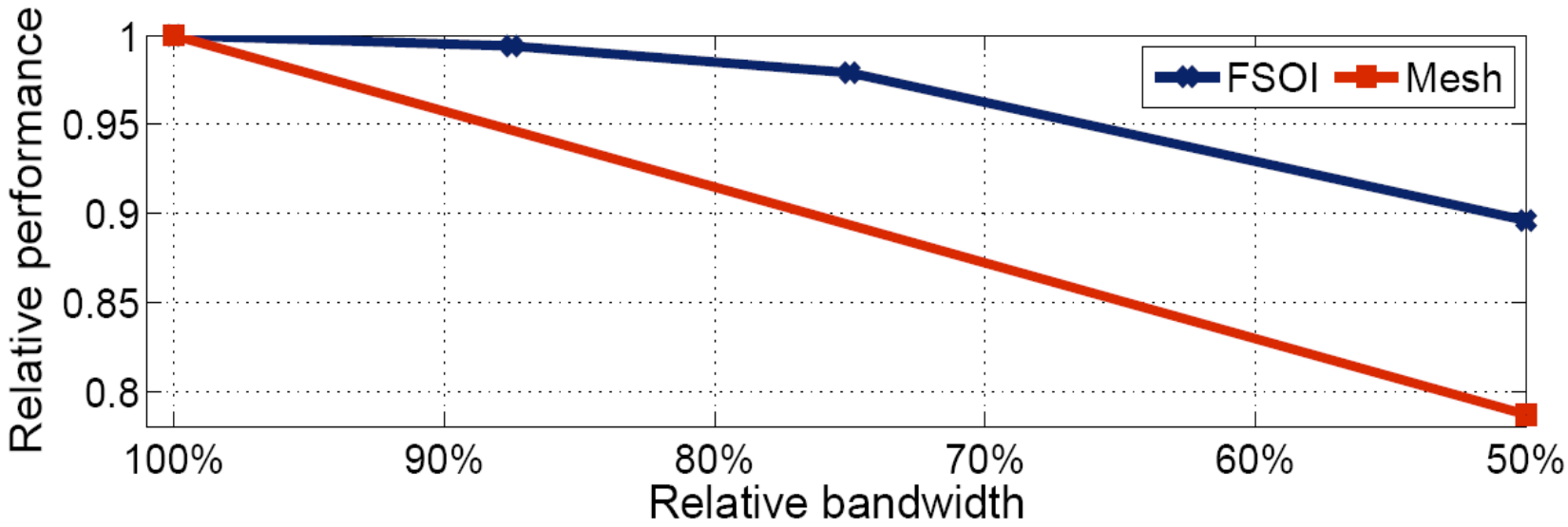
- 20x energy reduction in network
- Faster execution also reduces leakage and clock energy etc.
- 40.6% total energy savings
- 22% power savings (121W vs 156W)





Sensitivity Analysis

- Performance impact of progressive bandwidth reduction
  - Initial bandwidth comparable in both systems



- Allowing collisions ≠ requiring drastic over-provisioning

Other Details in Paper

- Using confirmation signal to provide specialized boolean value communication
- Spacing requests to ameliorating data packet collisions and its experiments analysis
- Improving collision resolution using info about requests
- Related work

Summary

- Proposed a scalable, fully-distributed free-space optical interconnect
  - Direct communication instead of packet relay: good performance
  - FSOI allows routing (virtual, on-demand) wires again: implementability
  - Integrating entire optical signal chain with efficient paths: excellent energy efficiency
- Allowing packet collision is a central tradeoff
  - Arbitration free and low overhead for contention-free traffic
  - Same mechanism to handle errors
  - No significant over-provisioning necessary
  - New opportunity for simple optimizations
- Technology readiness
  - Entire signaling chain is commercially available in large scale
  - 3D integration of disparate technologies common in small scale SoCs
  - Thermal issues may be avoided by piggybacking on other developments

Thanks!

Questions?

An Intra-Chip Free-Space Optical Interconnect*

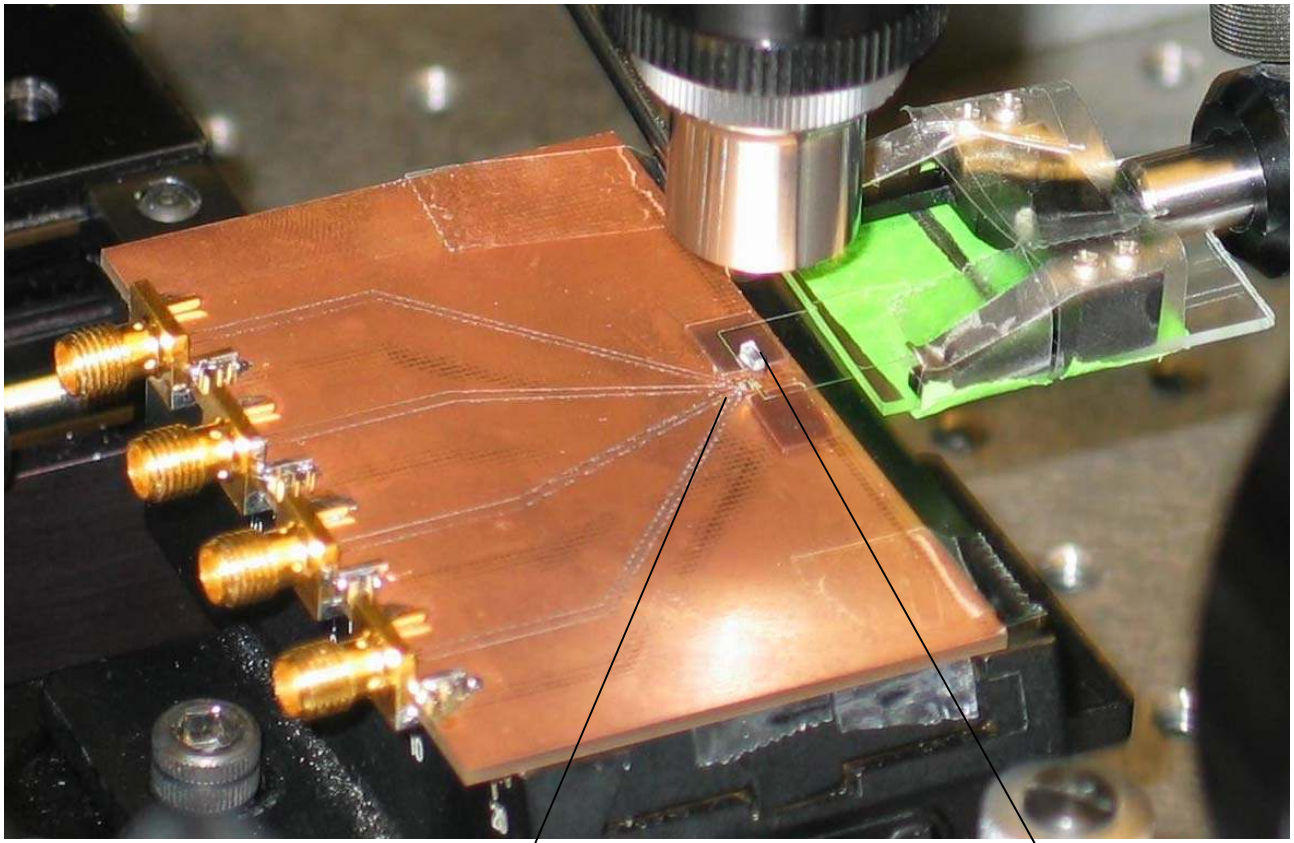
Backup Slides

Jing Xue

Dept. Electrical and Computer Engineering
University of Rochester

* To appear in Int'l Symp. on Computer Architecture, June 2010. Extended TR will be available online soon.

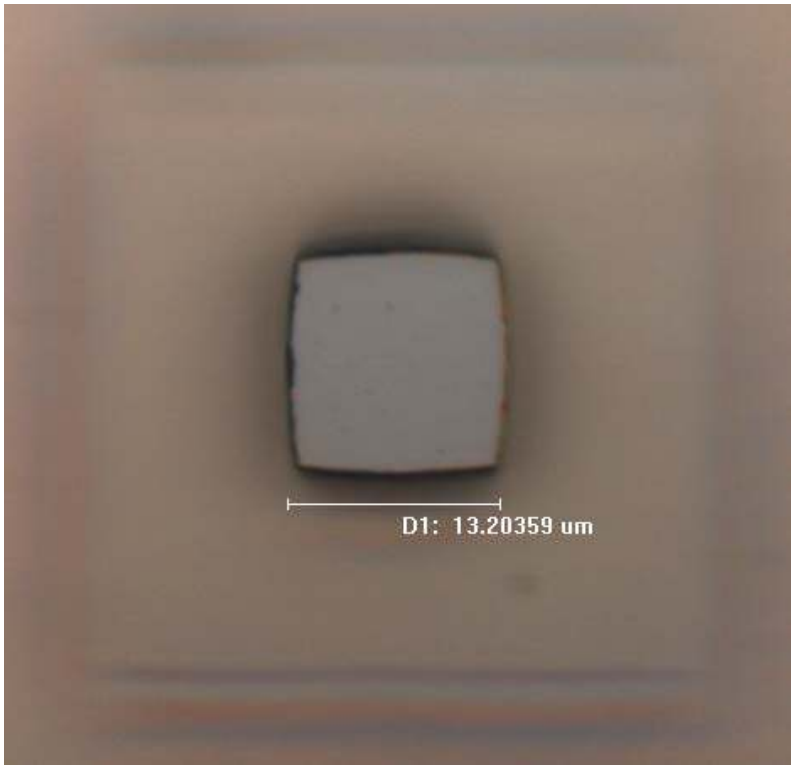
Readily Available Technology



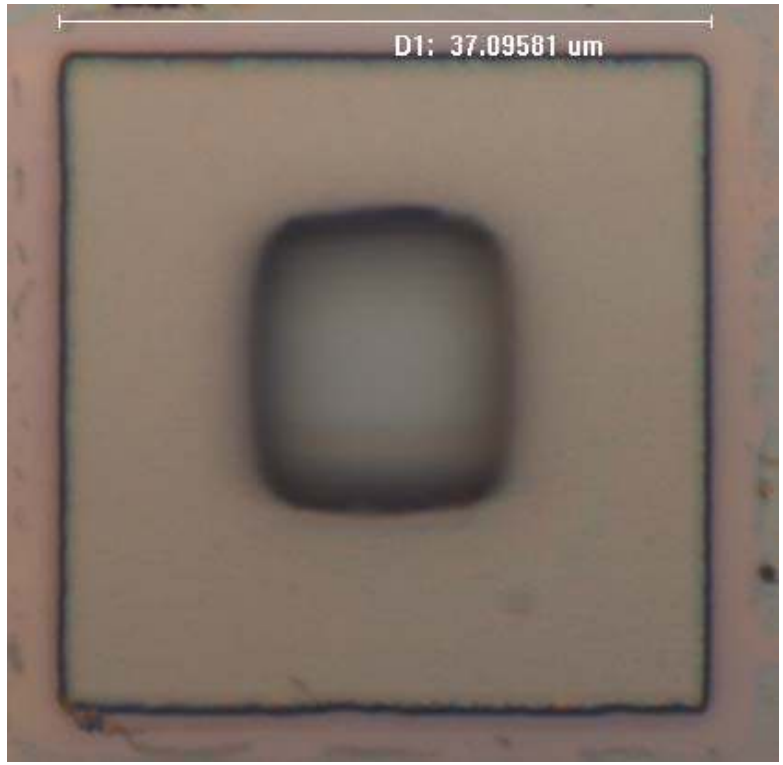
Commercial VCSELs

Commercial microlenses

Single VCSEL Structure (Under Microscope)

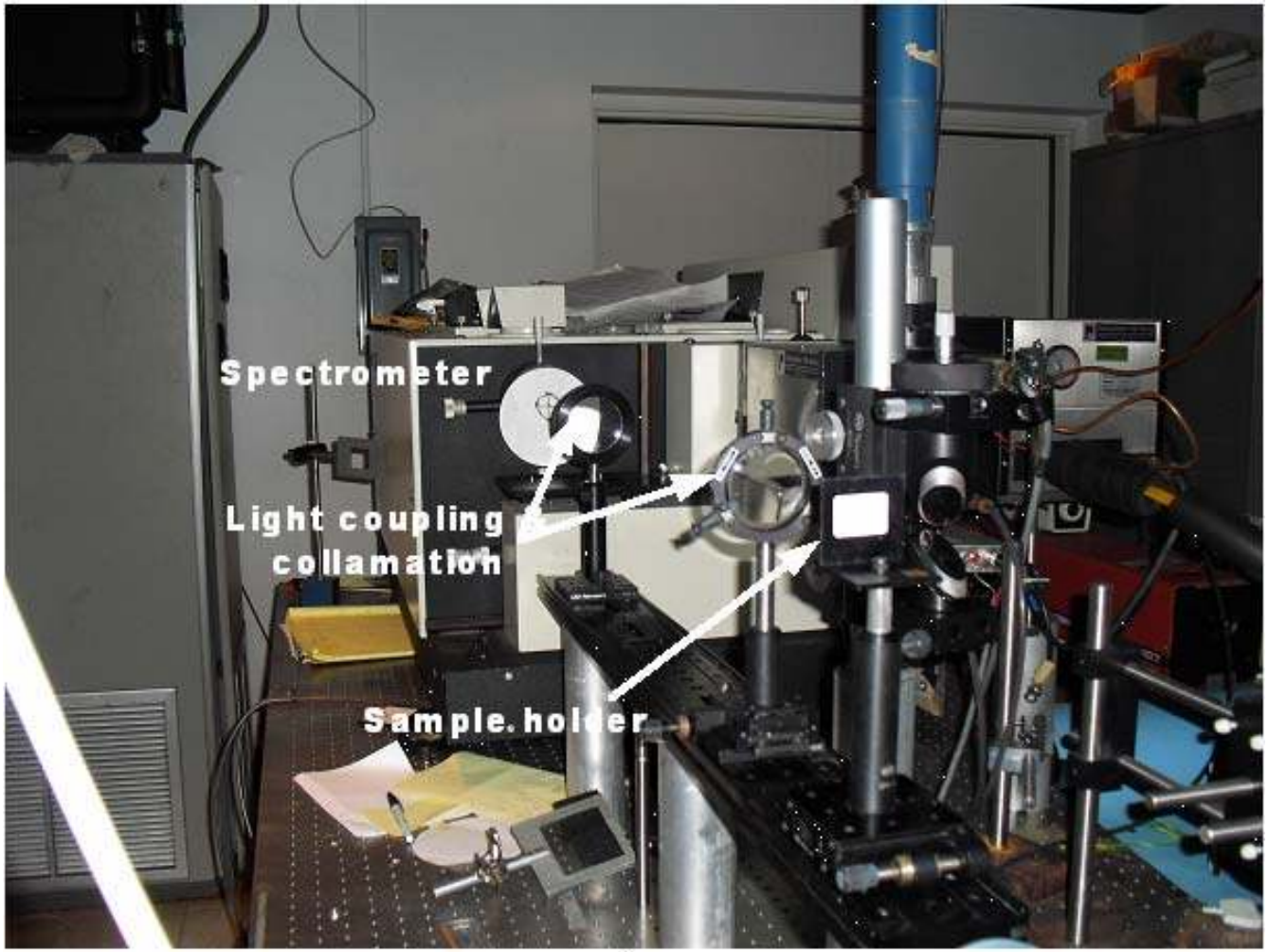


a) Top view of the etched mirrors

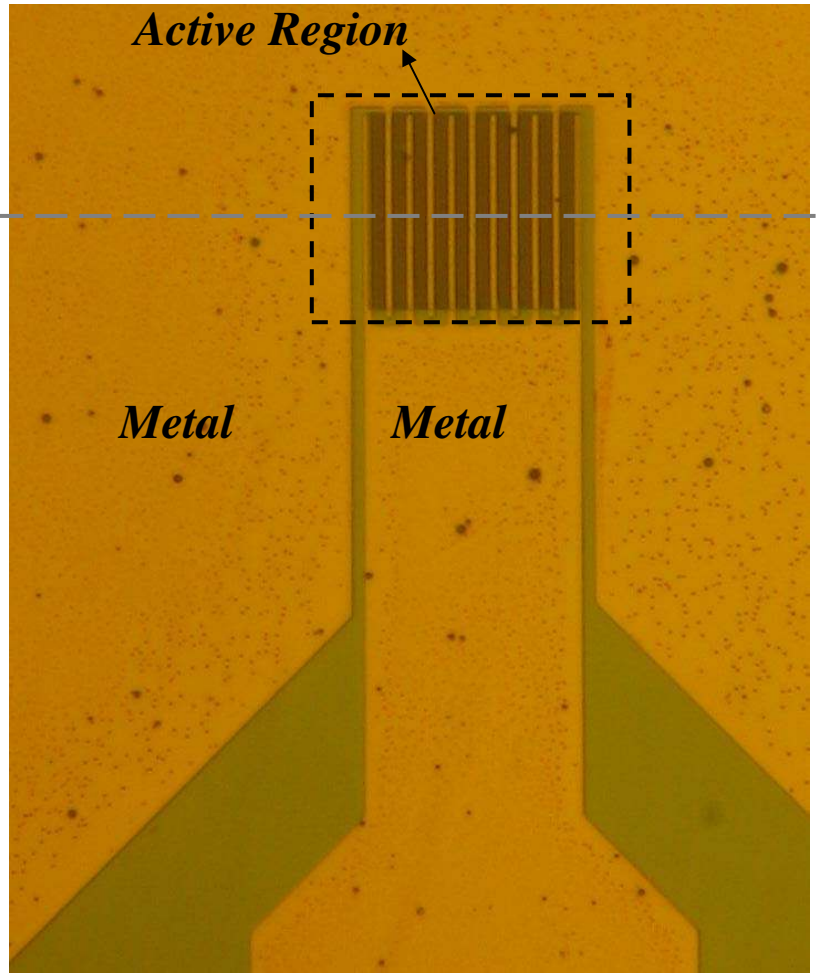


b) The p-contact region of the VCSEL, located below the mirrors shown in a)

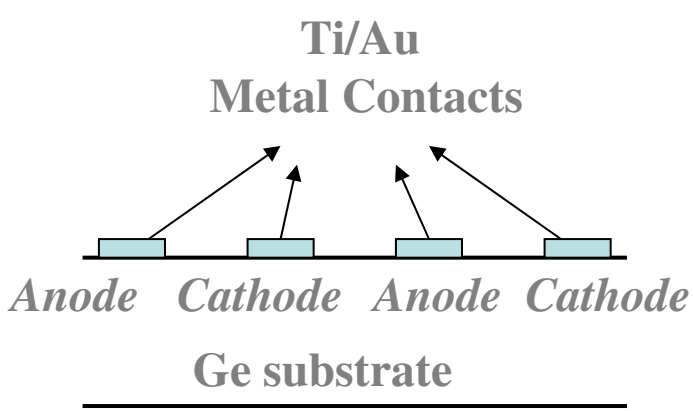
Spectrometer Setup



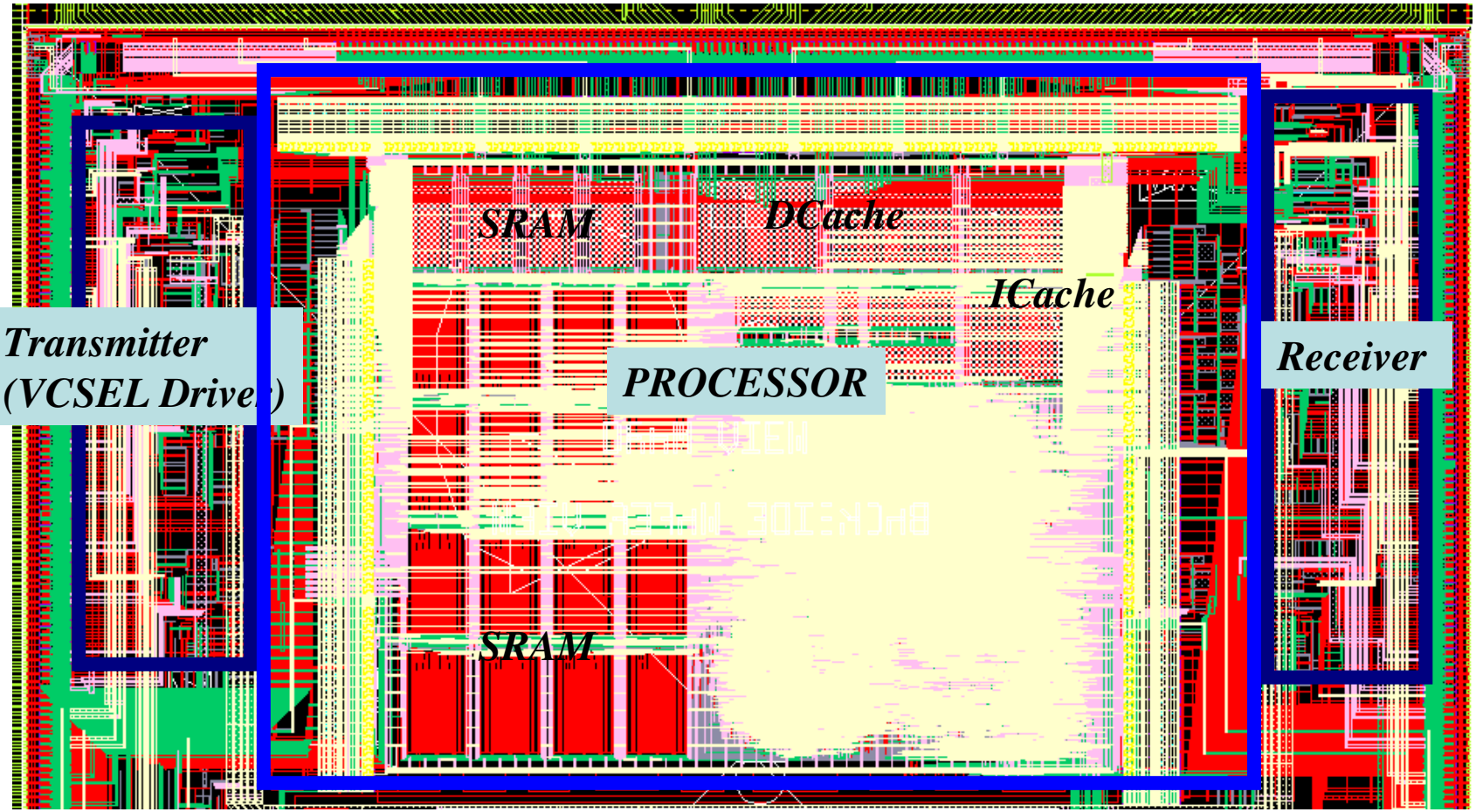
Germanium Photodetectors



Side view of Germanium Photodetector



3D Test Chip for System-Level Demo

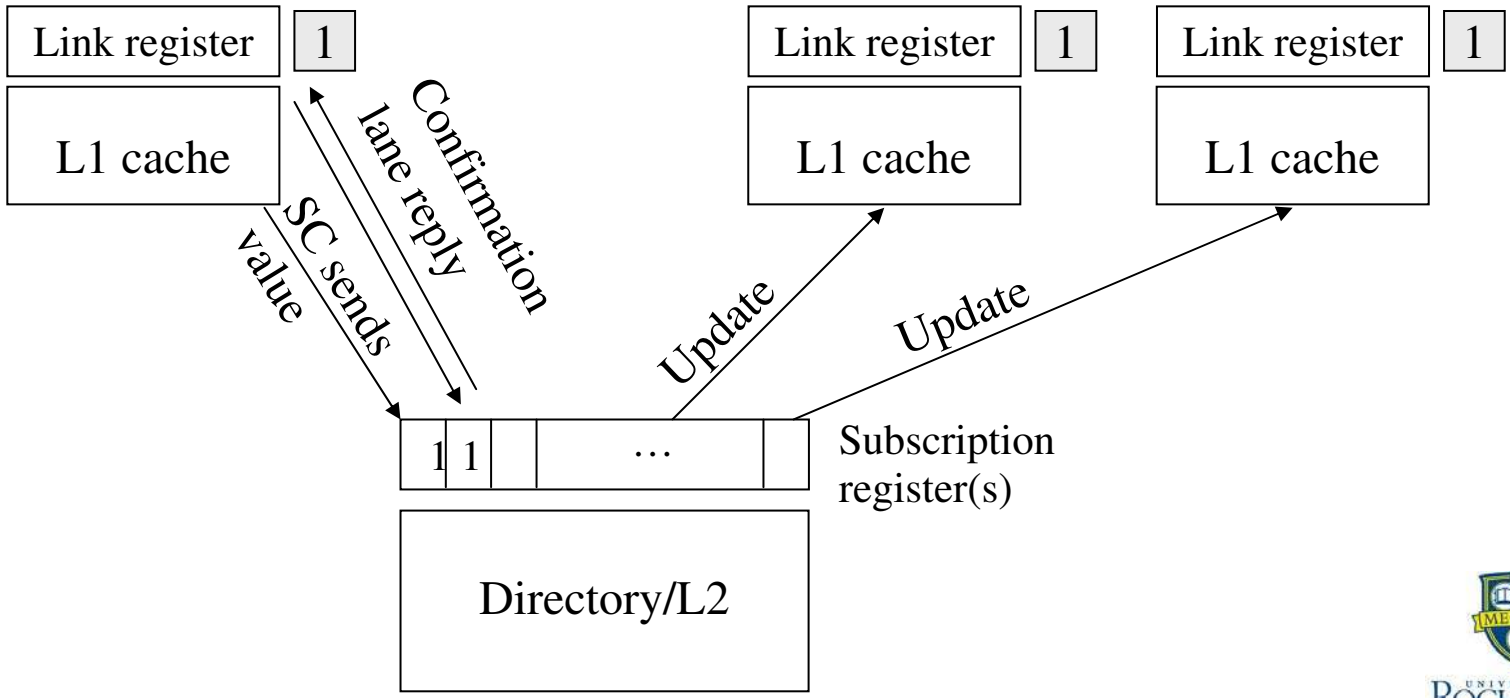


Specialized Boolean Value Communication

- Synchronization primitives:
  - Often boolean-based, unsuitable for inv.

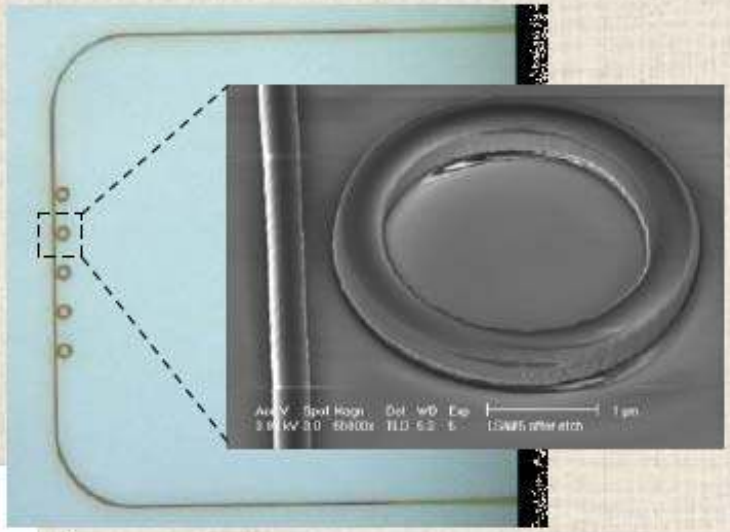
```
TEST:   LL      $1, 0($16)
        BNZ     $1, TEST
TAS:    BIS     $1, 1, $1
        SC      $1, 0($16)
        BZ      $1, TEST
```

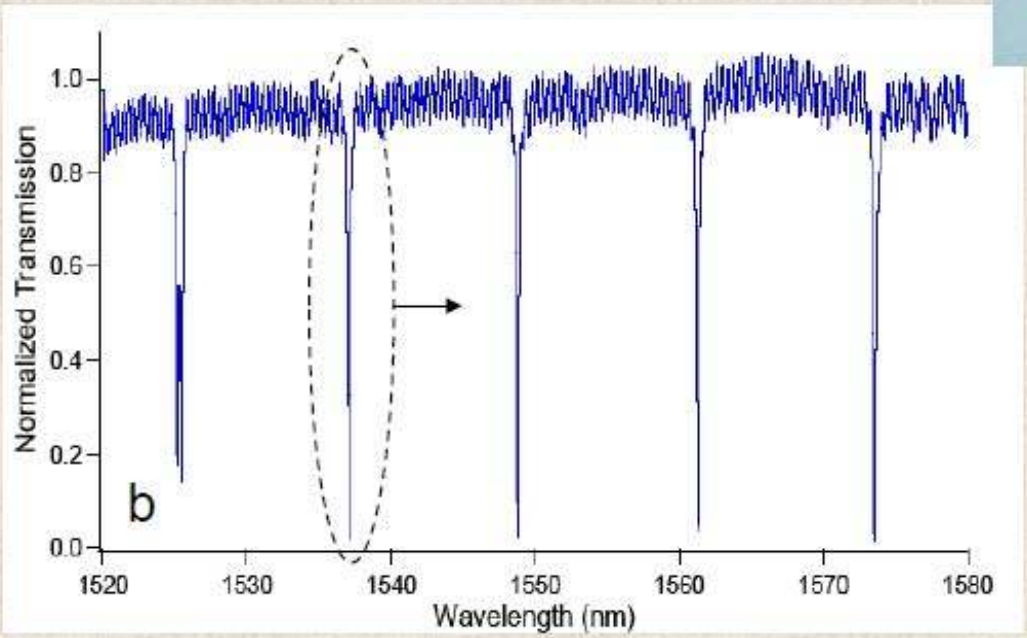
- Use confirmation laser to transparently:
  - Carry the values over confirmation lane (using pulse position)
  - Provide an update protocol (for LL/SC)

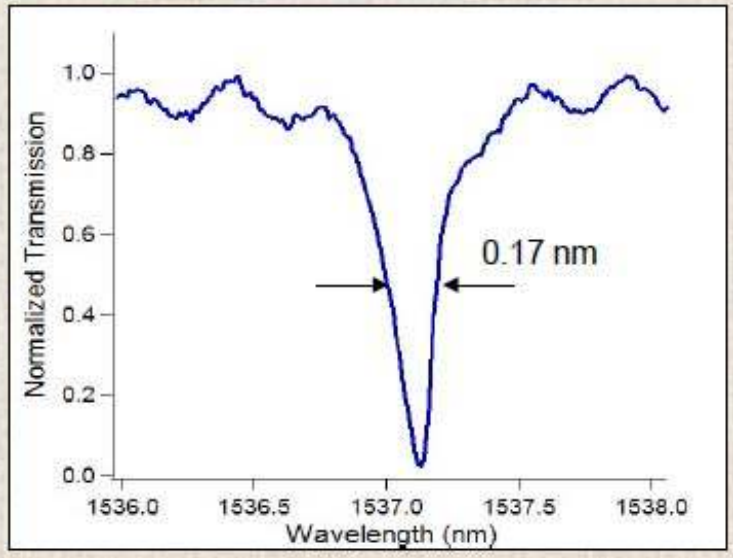


High-Q 1.5-μm-radius microring resonator

- Resonances from five cascaded microring resonators with slightly different radii ~ 1.5 μm
- High Q of 9,000 (BW ~ 20 GHz) and high extinction ratio of 16 dB.





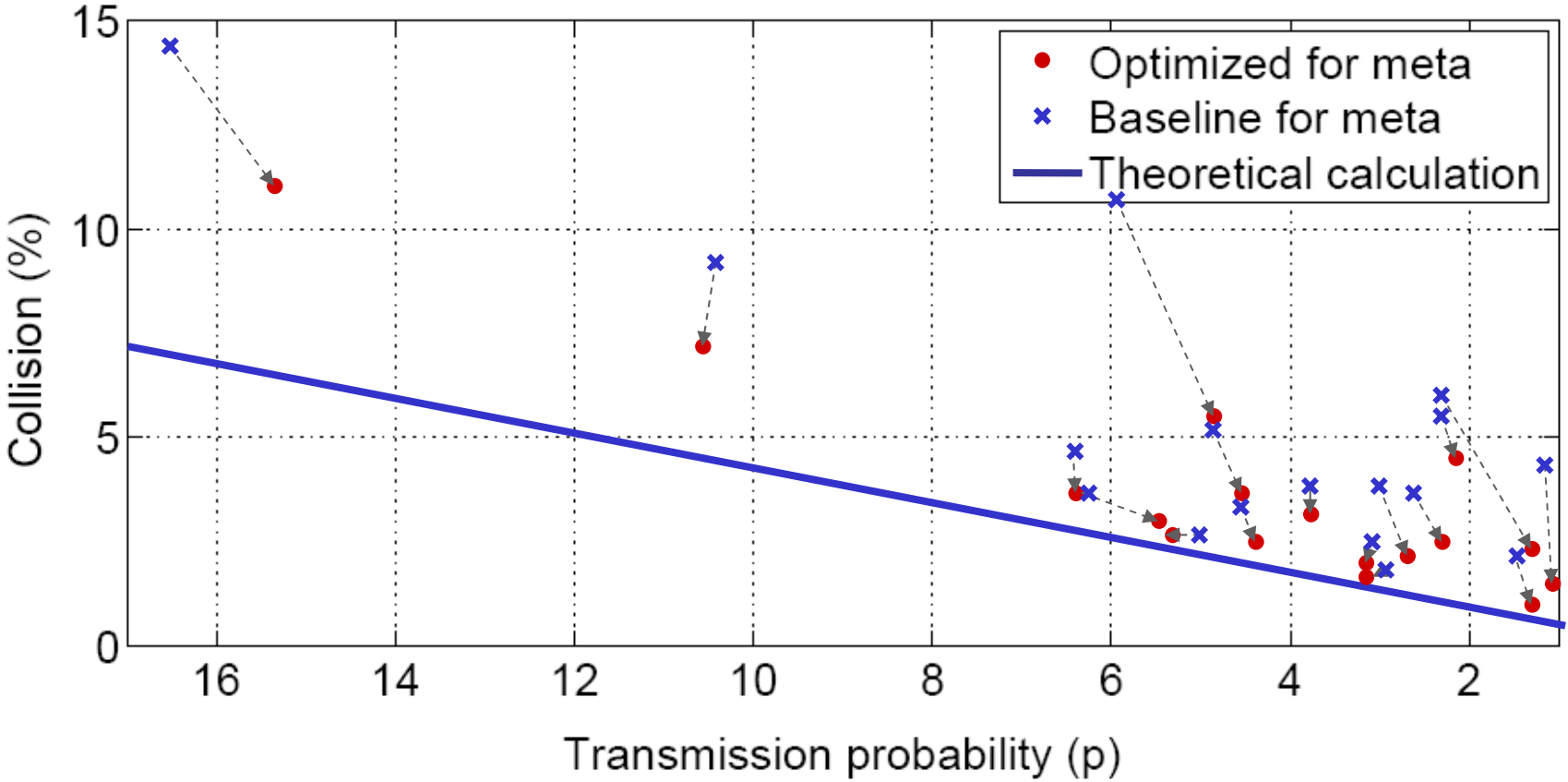


Q. Xu et al., *Opt. Express* **16**, 4309 (2008)

Recap of Main Tradeoffs

- Relay network
  - Relay contributes to energy cost and scalability challenges
  - Router complexity for performance also incurs costs
- Optical signaling can avoid relay using shared media
  - Off-chip light sources are expensive and power hungry
  - On-chip distribution and modulation chain (waveguide loss and insertion loss) reduce energy efficiency
  - WDM imposes stringent device constraints which pose challenges on fabrication
- FSOI avoids relay and minimizes loss in signaling chain
  - Requires 3D integration of disparate technologies
  - Makes air cooling very difficult

Eliminating Acknowledgements

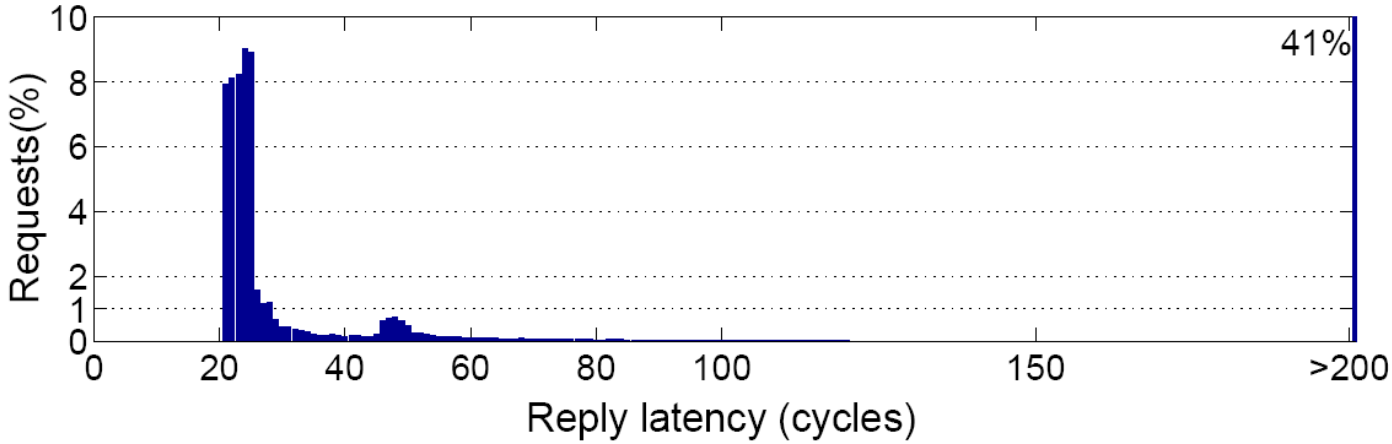


Reduces 5.1% traffic but eliminates 31.5% of meta packet collisions

Invalidation acknowledgements systemically synchronized

Ameliorating Data Packet Collisions

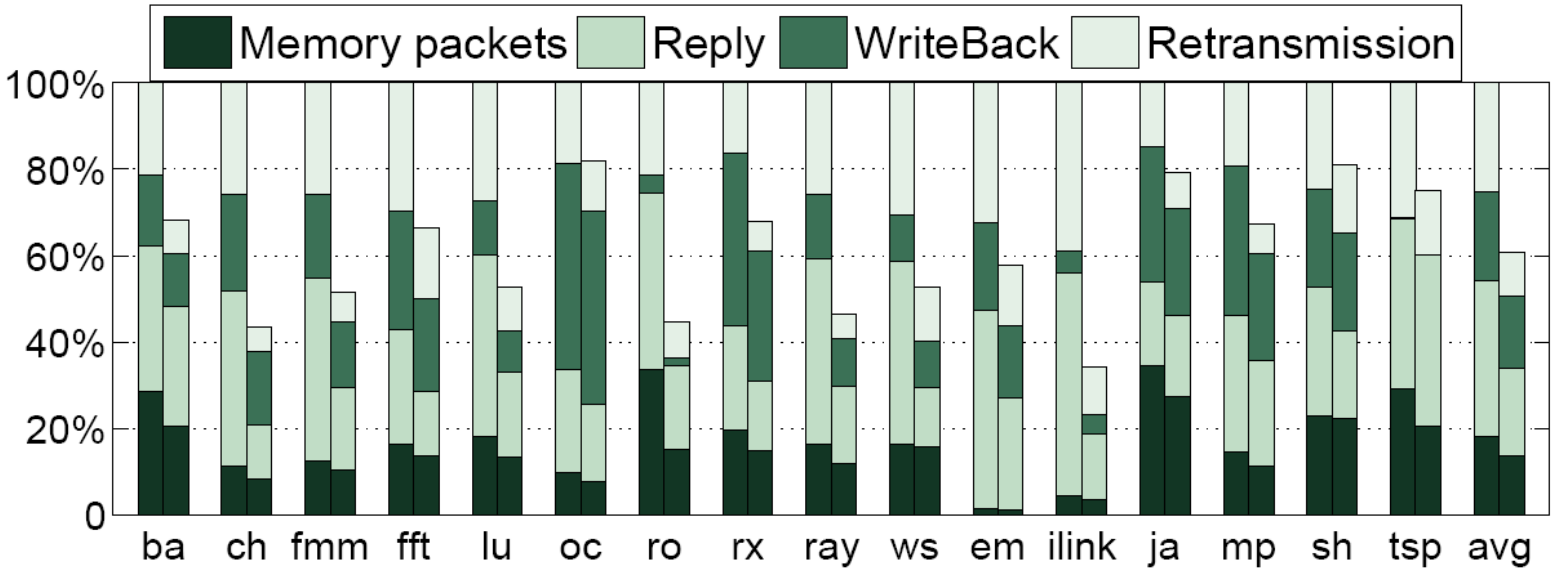
- Reduce the probability of data collision with spacing



- Improving collision resolution using info about requests
  - Collision-corrupted headers still reveal info about senders
  - Can notify one sender to immediately resend
  - Need not be correct, at most causing a collision for another node

Data Packet Collision Optimization

- Collision rate reduction: 38% of data collisions



- Collision resolution hint reduces delay by 30% (41→29 cycles)
- Performance impact depends on collision frequency
- Improves performance robustness

Related Work

- Buffer-less optical packet-switched network, Schacham and Bergman, *IEEE Micro* 2007
- Circuit-switched optical network, Schacham et al. NOC'07
- Bus or ring-based shared-medium optical interconnect
  - Ha and Pinkston *JPDC* 1997
  - HP Corona (Beausoleil *LEOS* 2008, Vantrease et al. ISCA'08)
  - Kirman et al. MICRO'06
- Free-space optics
  - Miller, *J. Sel. Top. in Quantum Elec*. 2007
  - Krishnamoorthy and Miller, *JPDC* 1997
  - Marchand et al. *JPDC* 1997
  - Walker et al. *Applied Optics* 1998