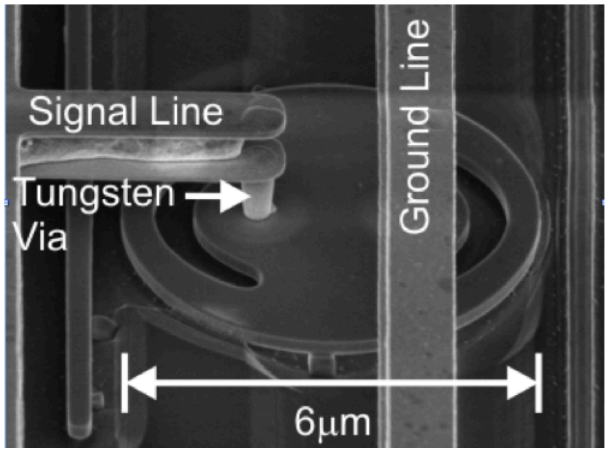
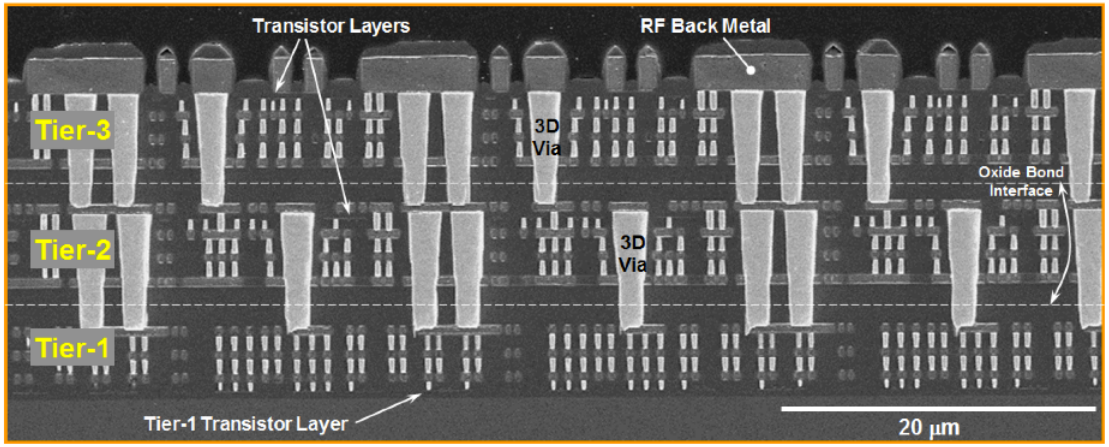
Electronic-Photonic Integration within Switches and Routers

Michael R. Watts

August 23, 2012

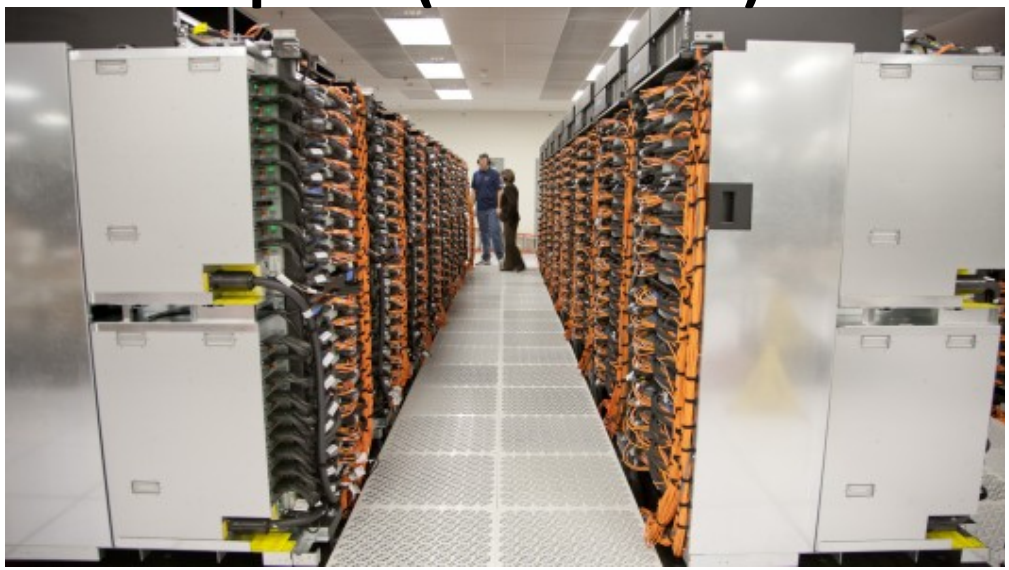
Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, MA

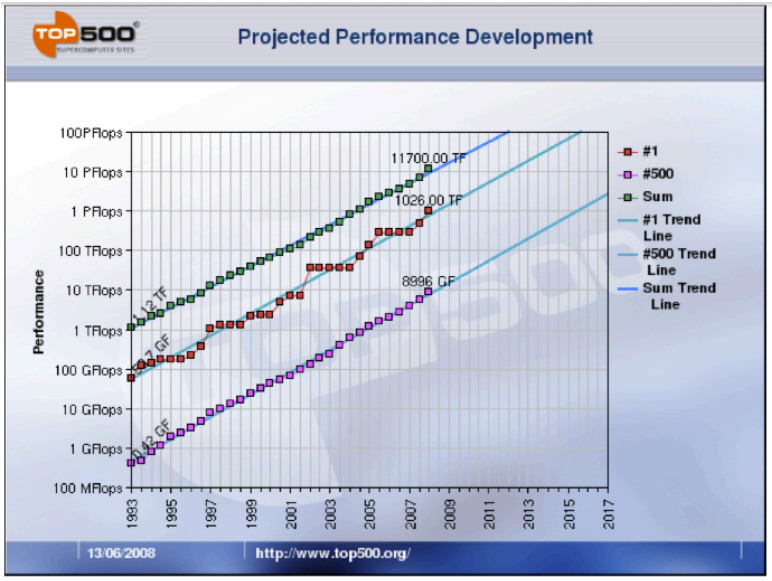




Supercomputer Scaling

IBM Sequoia (16 PFLOPs)





Data Center Power Consumption

- **Today:** #1, IBM Sequoia, 1.5M-cores, 16 PFLOPs, 8MW Future
- **Exascale:** 10^{18} FLOPs → 10^{18} Bytes/s Bandwidth/Port: 50× that required by Data Centers, and with no Moore's Law for Communications, expect future machines will be dominated by communications power consumption

Data Center Expansion and Power Consumption

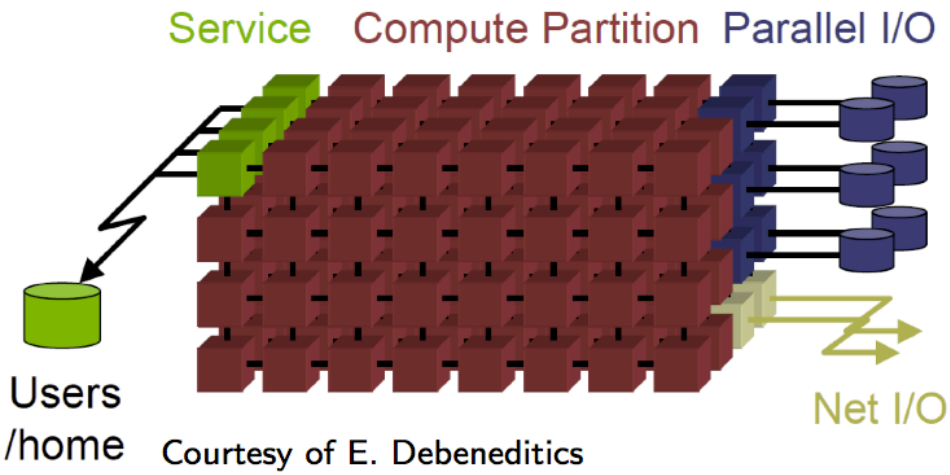
Microsoft's Chicago Data Center

200 containers
2500 servers per container

Data Center Power Consumption

- **Machine Scale:** Enormous machines cost ~$1B, powered by ~1-million servers. Google, Microsoft, Apple, Ebay, etc.
- **Power Consumption:** Consume ~2% of U.S. electricity (EPA), ~50MW each, today. Cost of power exceeds cost of server over three year server life.

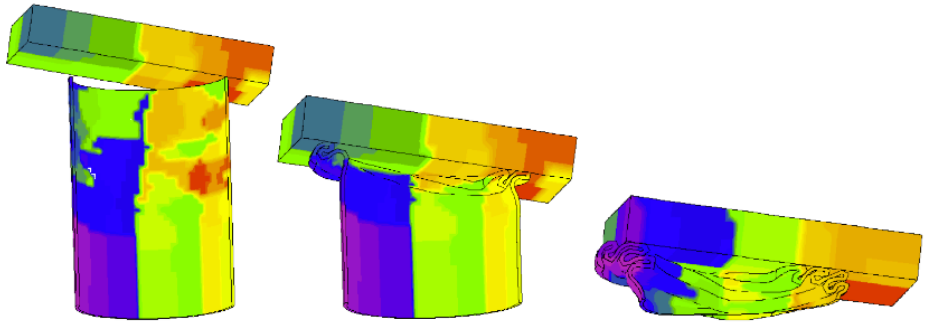
Traditional Electrical Networks



Courtesy of E. Debeneditics

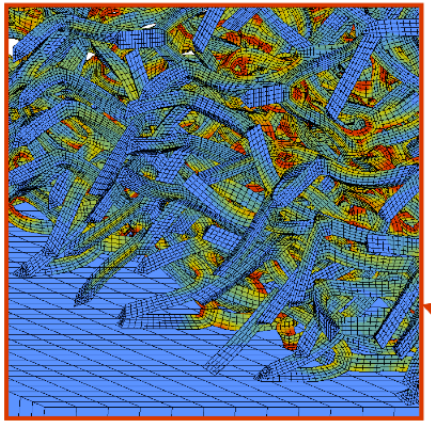
Finite Element Model (FEM) of a Crushed Can

Courtesy of Sandia CSRI



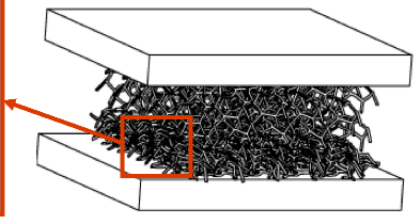
High-Order Filters

- **Data Patterns:** Irregular
- **Scale:** 100 × 100 × 100-nodes
- **Energy:** 100hops×10pJ/bit =1nJ/bit
- **Power:** 1nJ/bit × $8 \cdot 10^{18}$bits/s = 8GW
- **Wires:** 10^{11}, at \$0.5/m→\$50B (Too Many Hops!)

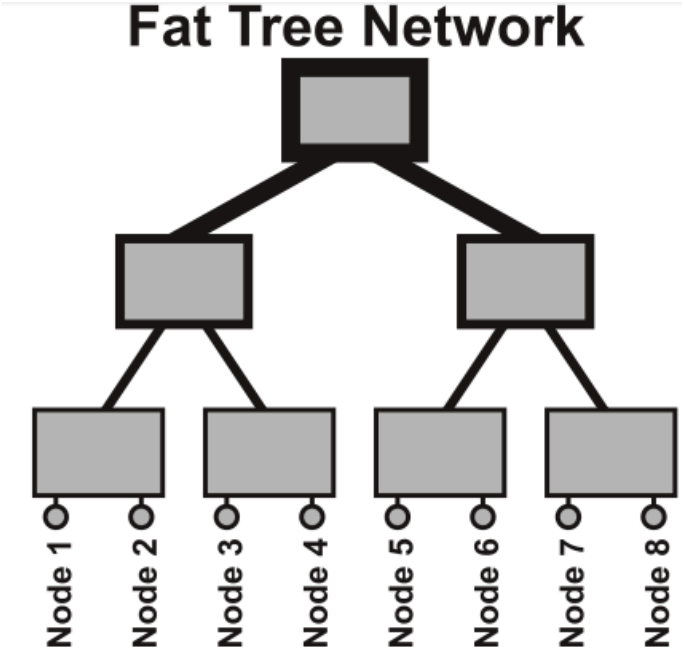


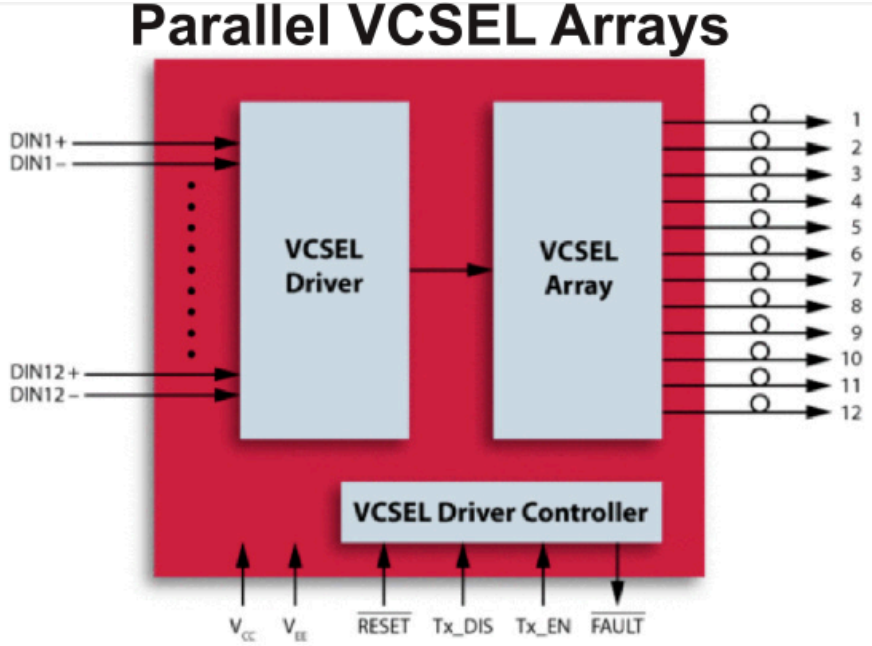
Finite Element Model (FEM) of Crushed Foam

Courtesy of Sandia CSRI



Optics: Enables Longer Reach and Higher Bandwidth

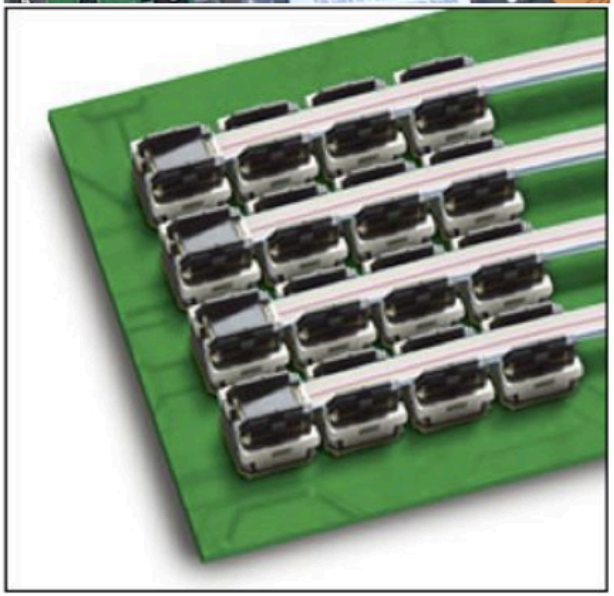




Fat Tree / Clos Networks: Fewer hops / greater distances

- **Scaling:** 2i + 1 stages→ n^{i+1} nodes / $(i + 1)n^i$, n × n switches
- **Exascale:** i = 3, 324 ≈ 10^6nodes, 7-hops→70pJ/bit
- **Power:** 70pJ ×8bits ×10^{18}bits → 560MW, better but . . .
- **Fibers:** 7 × 8 × 10^{18} bits → 5 × 10^9fibers/system
- **Reliability:** Have 5 × 10^9VCSELs, at 20-FIT → 100 fail/hour

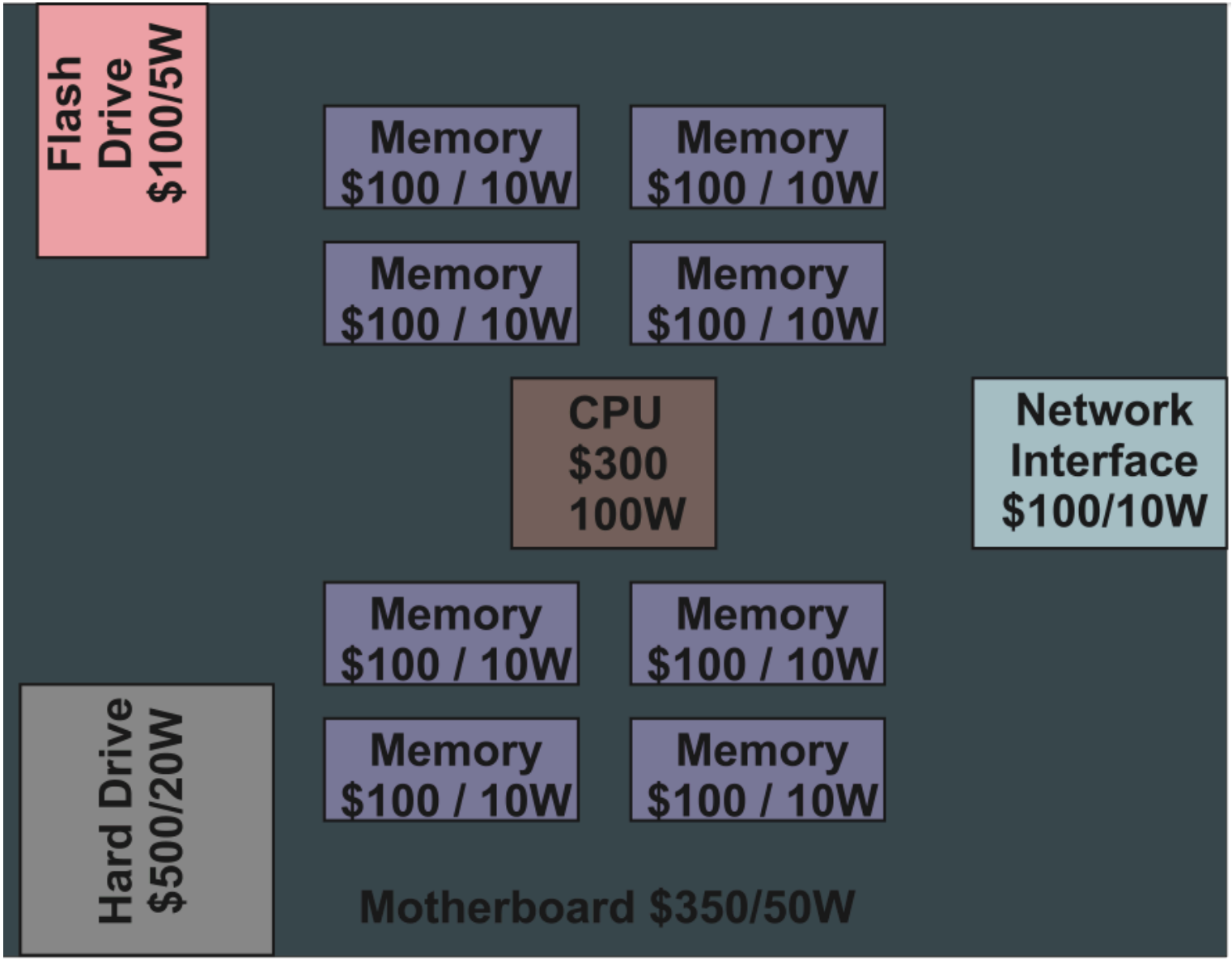
IBM's Latest Machine: Optics Everywhere



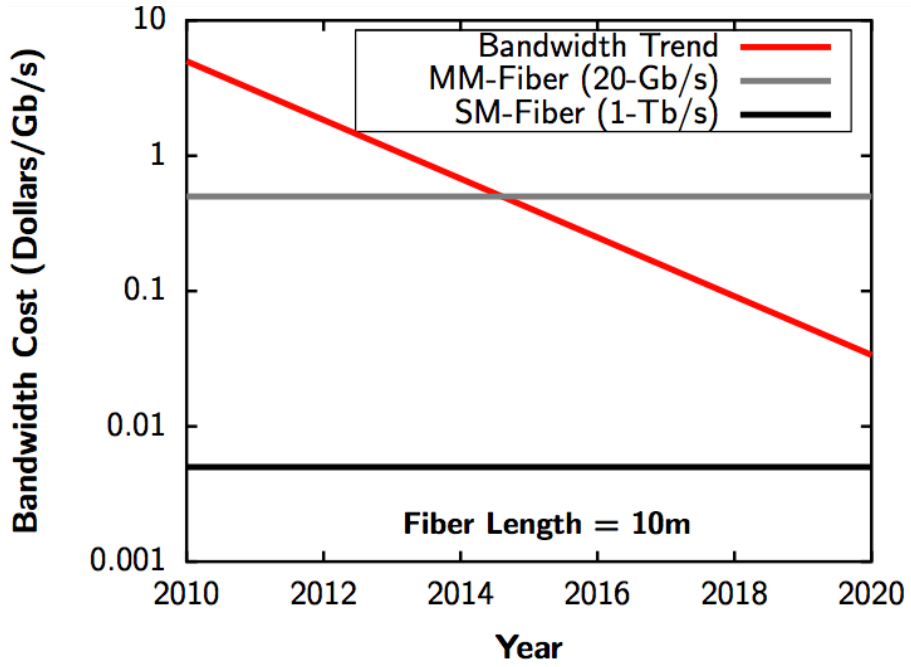
IBM Blue Waters

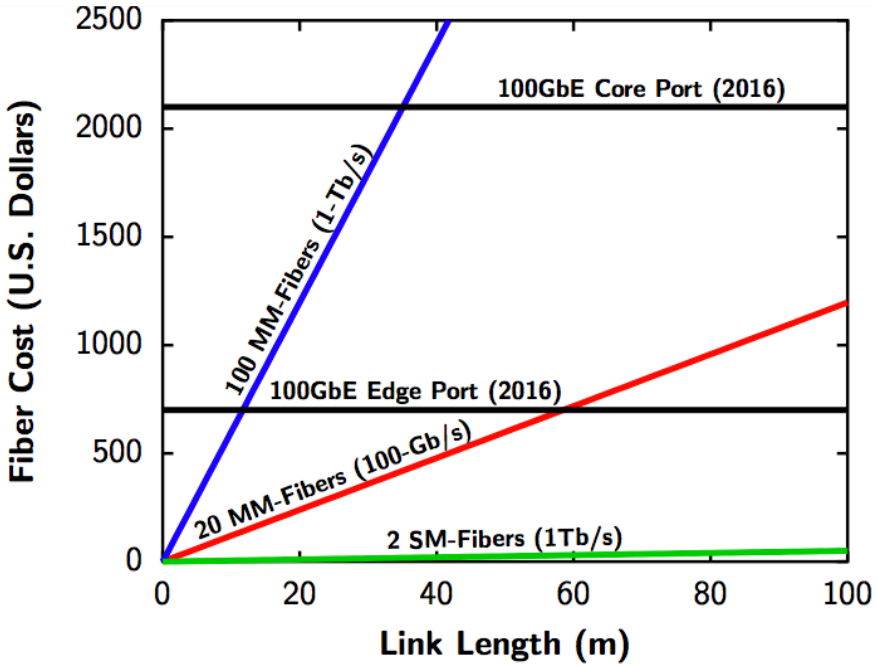
- **Drawer:** 16-Sockets, 8-Procs, 8-Switches Multi-Chip-Module
- **Processor:** 4, 8-core Power7 chips, 32-cores total, 1TFLOP/MCM
- **Memory:** 4GB/core, 128GB per Processor
- **Memory BW:** 512GB/s/MCM (4Tb/s
- **Network BW:** 192GB/s/MCM (1.5Tb/s)
  VCSELs: ~360/MCM

Performance Scales: Price Remains Constant



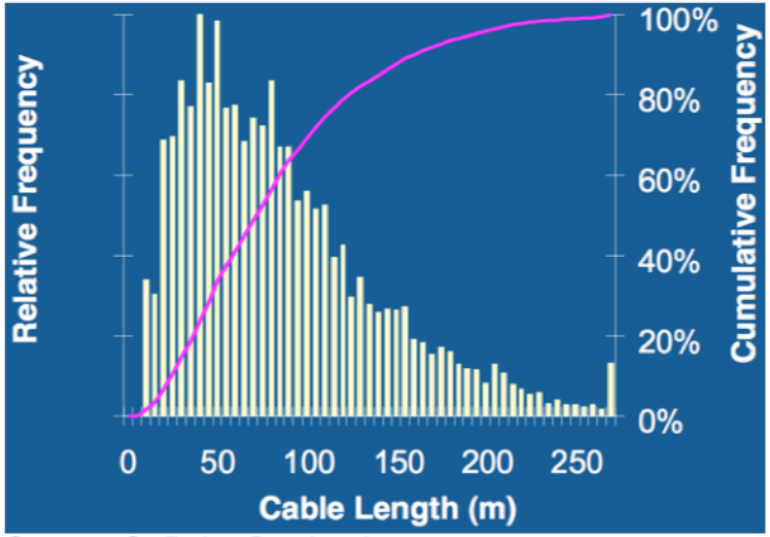
Fiber Cost Begins to Dominate Link Cost





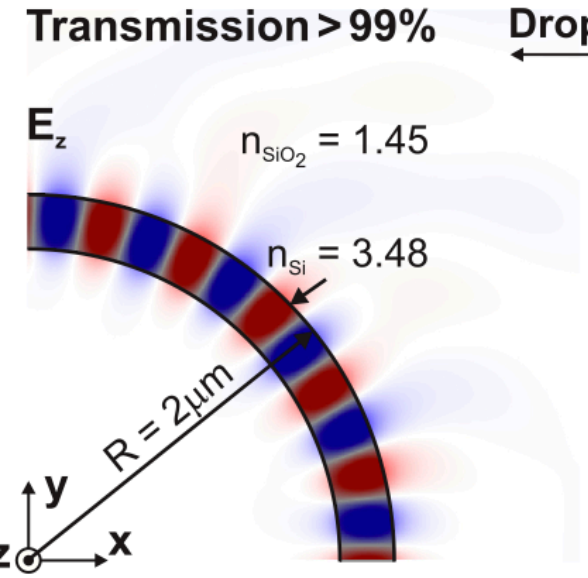
Parallel VCSEL Arrays

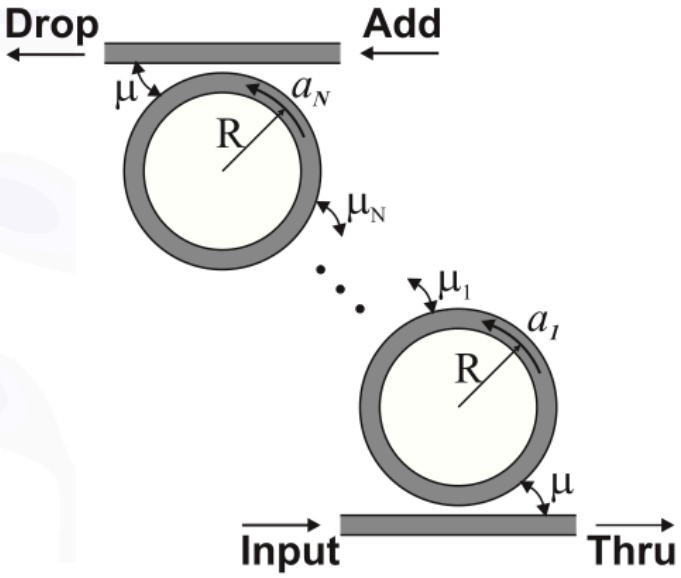
- **Parallel Links:** Required for VCSEL-based links above 10GbE
- **MM-Fiber:** Cost limits price scaling
- **Power Scaling:** Need constant port power with each generation

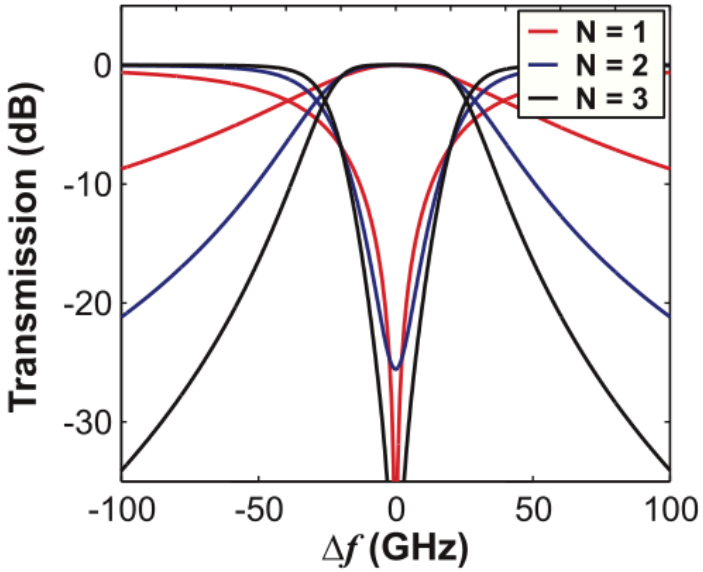


Source: S. Bois, Corning Inc.

Why Microphotonics?







$$\theta_c = \sin^{-1}\frac{n_1}{n_2}$$

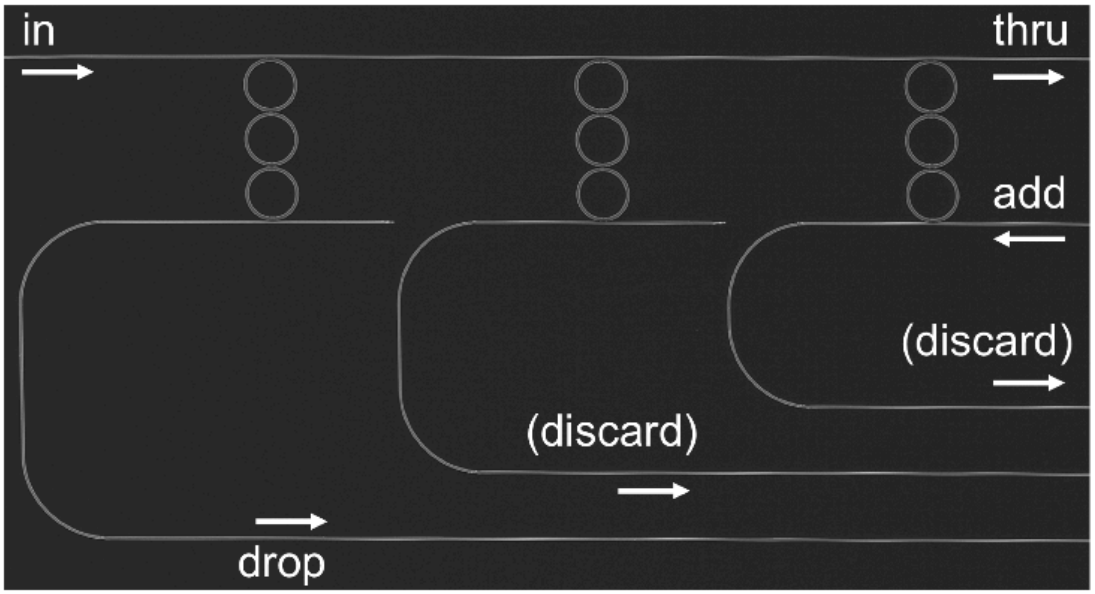
$$a_1 = \frac{-j\mu s_i}{j\Delta\omega_1 + \frac{1}{\tau} + \frac{\mu_1^2}{j\Delta\omega_2 + \frac{\mu_2^2}{j\Delta\omega_3 \cdots +}}}$$

N	Maximally Flat
2	$\mu_1^2 = 0.25\mu^4$
3	$\mu_1^2 = \mu_2^2 = 0.125\mu^4$

For Starters . . .

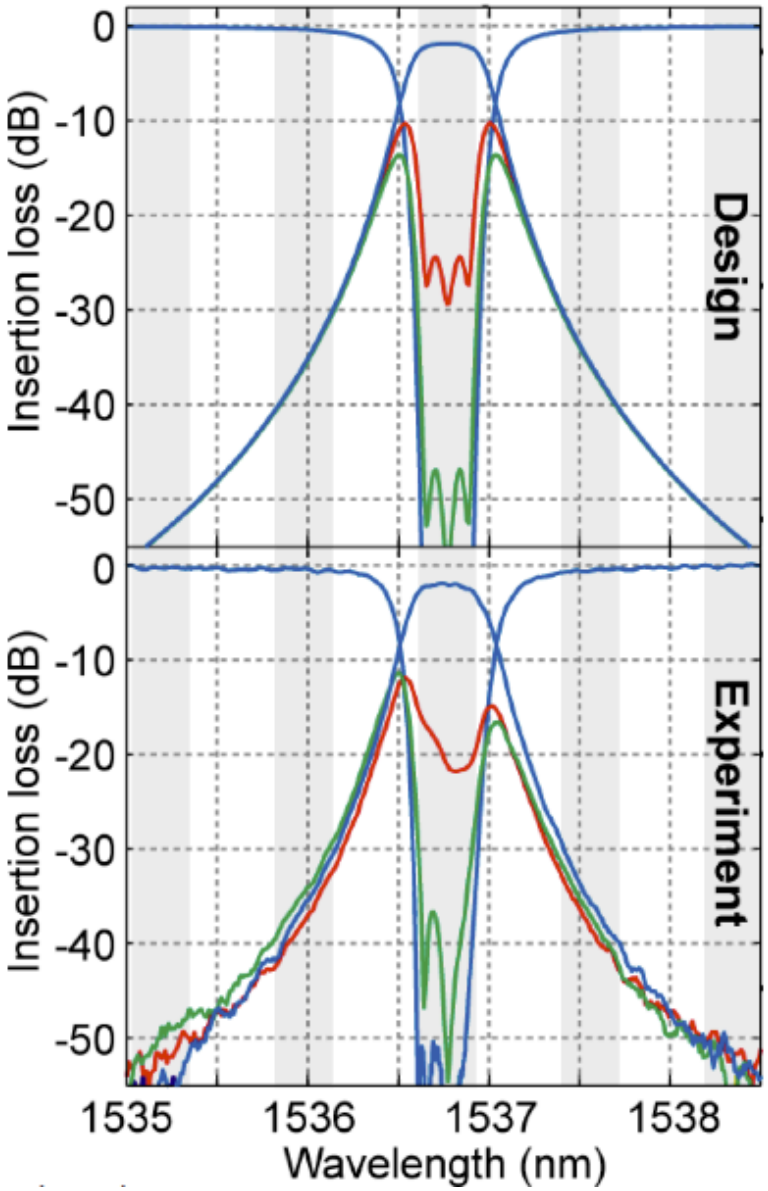
- ❑ **High Index Contrast:** Bends 2μm vs. 2cm for silica guides (10^8×)
- ❑ **Microrings:** Compact, unidirectional, huge Free-Spectral-Range
- ❑ **Filters:** High-order filters constructed with microwave filter tables
- ❑ **Bandwidth Scaling:** Wavelength Division Multiplexing (WDM)
- ❑ **Power:** Using single-mode optics & integrated CMOS ➔low power

WDM Communications

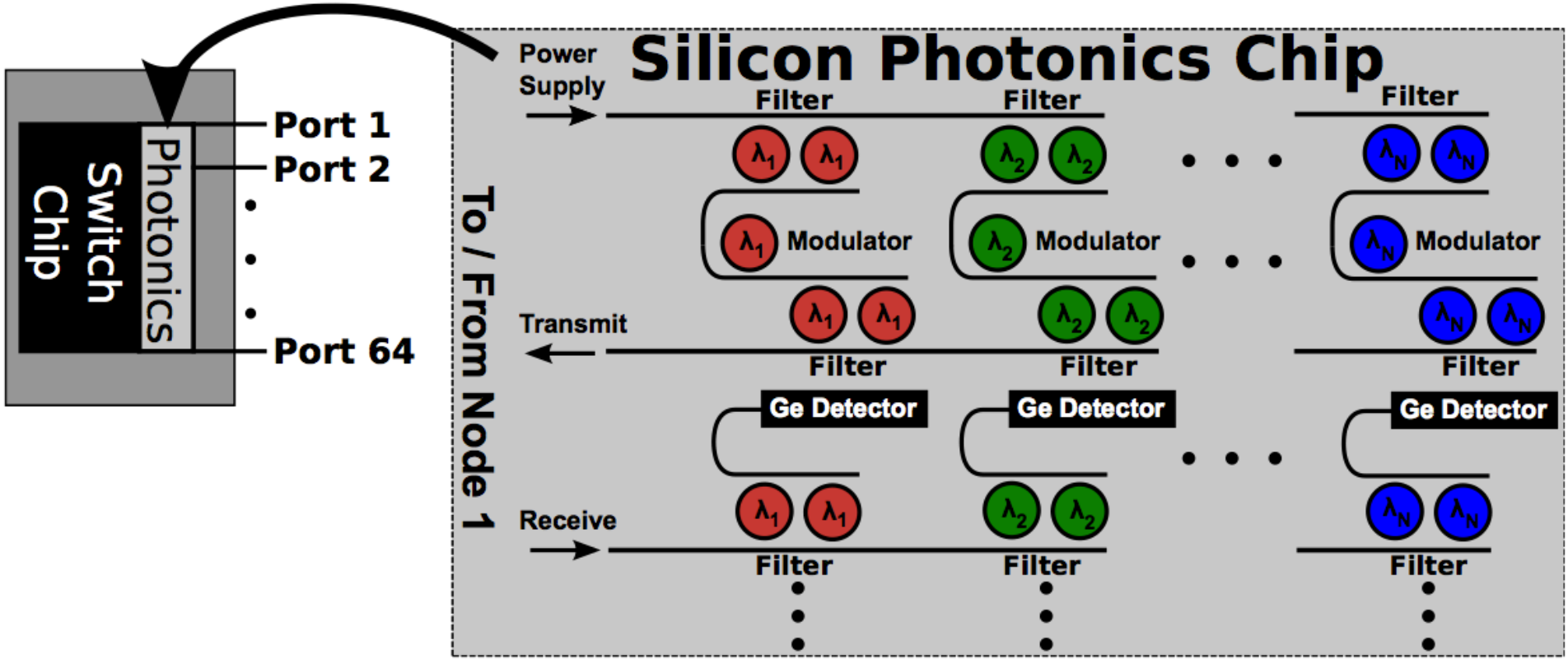


High-Order Filters

- **Predictable:** Using rigorous electromagnetic simulations → predict filter characteristics
- **High Quality:** Careful design/fab → exceptional performance



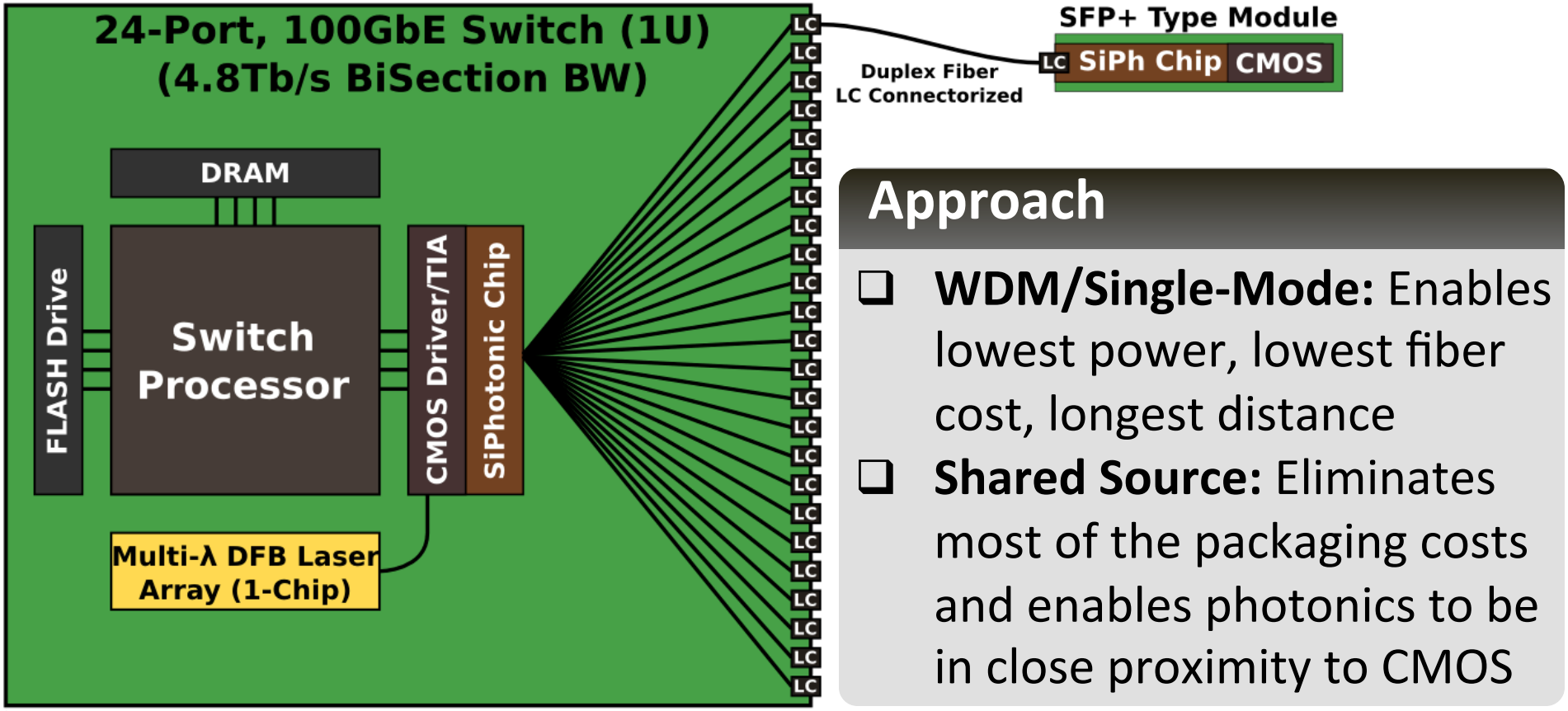
A Simple Architecture



Power Budget (1pJ/bit)

- **Photonic:** Modulation, filtering, optical power, and detection
- **Electronic:** Drivers, SerDes, amplifiers, and clock-recovery

OEO Approach



Approach

- ❑ **WDM/Single-Mode:** Enables lowest power, lowest fiber cost, longest distance
- ❑ **Shared Source:** Eliminates most of the packaging costs and enables photonics to be in close proximity to CMOS

Advantages

- ❑ **Cost:** Shared laser, reduced fiber, fewer parts, etc.
- ❑ **Port Density:** No longer limited by faceplate density, BW scalable
- ❑ **Power:** Photonic power consumption reduced by 3X

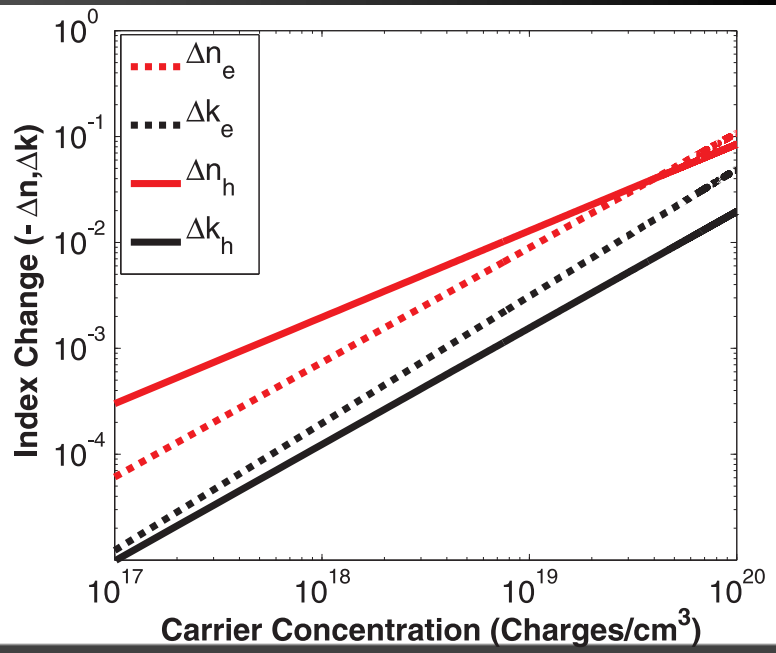
Free-Carrier Silicon Modulators

The Free-Carrier Plasma

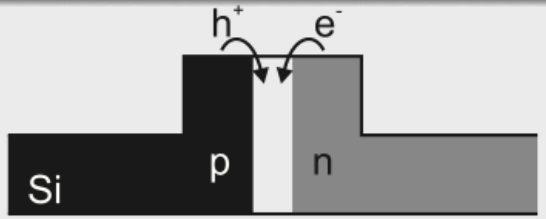
$$\mathbf{D} = \epsilon\mathbf{E} + \mathbf{\Delta P} = \left(\epsilon - \frac{Nq^2}{m^*\omega^2}\right)\mathbf{E}$$

$$n = \sqrt{\frac{\epsilon}{\epsilon_0}} \rightarrow \Delta n \approx \frac{\Delta\epsilon}{2}$$

$$\boxed{\Delta n = -\frac{Nq^2}{2m^*\omega^2}}$$



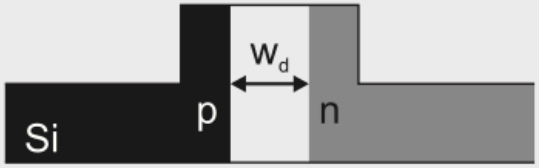
Injection: Drive charge into a *p-n* junction



Pros: Large effect → low voltage

Cons: Free-carrier lifetime ~1ns

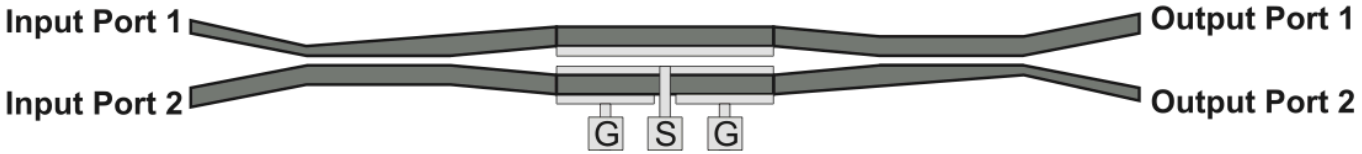
Depletion: Extract charge from a *p-n* junction

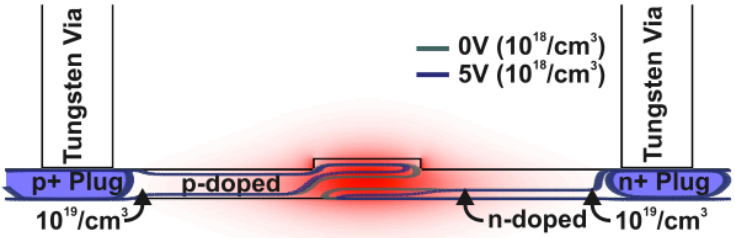


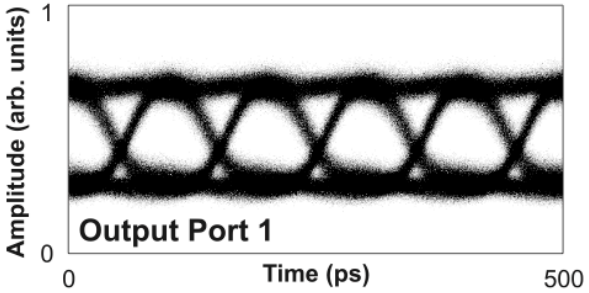
Pros: Limited by *RC* time constant

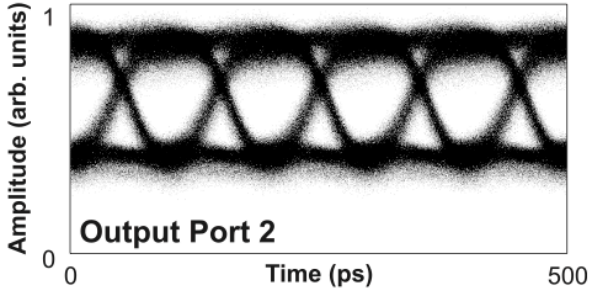
Cons: Smaller effect → higher voltage

Mach-Zehnder Modulators



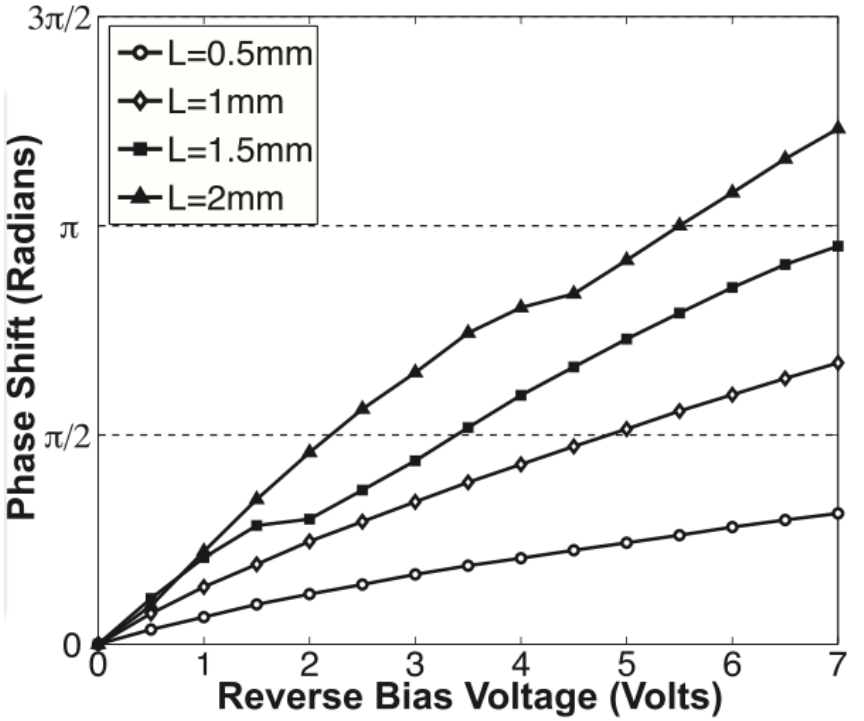






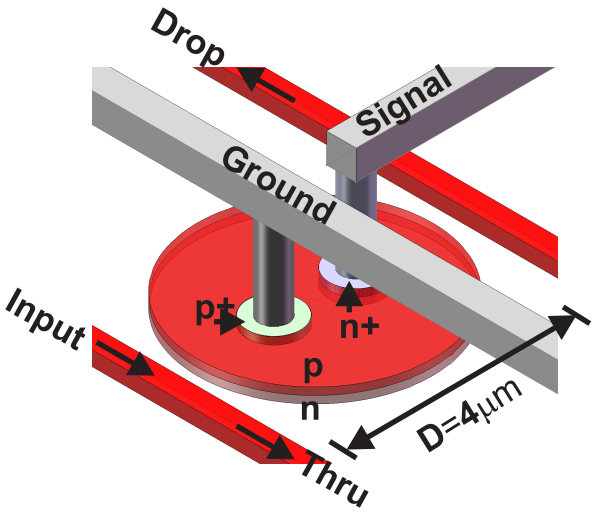
Maximum Overlap → Vertical Junction

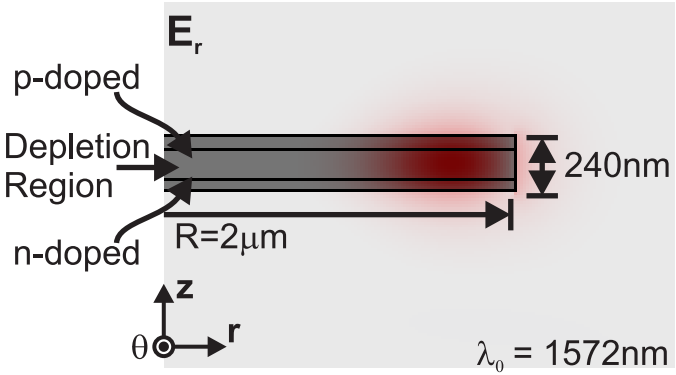
- **Effective Index:** $\Delta\bar{n} \propto \frac{nc\epsilon_0}{2}\int \Delta N|\mathbf{e}|^2 dA$
- **Phase Shift:** $\boxed{\Delta\phi \propto \Delta\bar{N}\frac{\Delta w_d L}{w}}$
- **Record** $V_\pi L$: $1V \cdot cm$
- **Power:** ~10pJ/bit, same as VCSELs

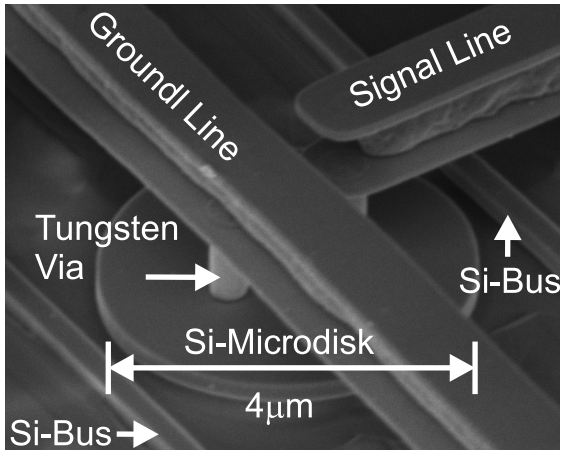


M. R. Watts et al., *IEEE JSTQE*, **16**, pp. 159-64 (2010)

Ultralow Power Modulators

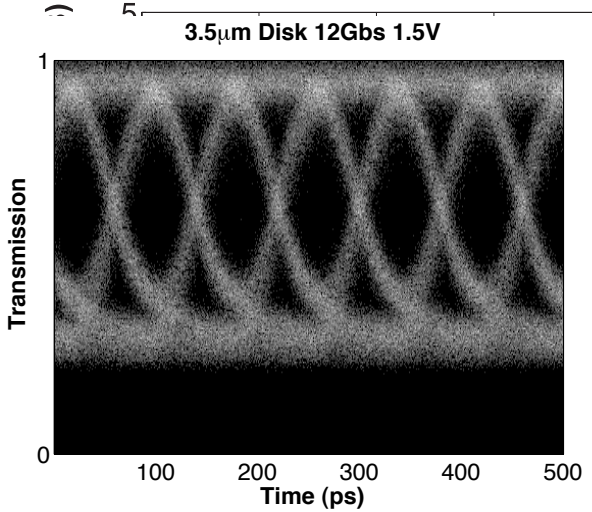


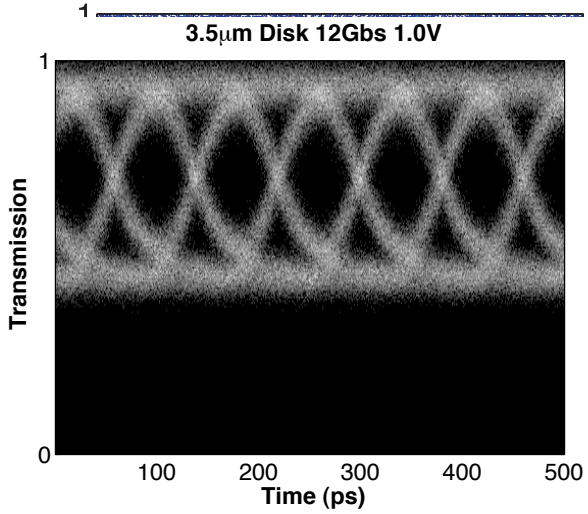




Modulator Power

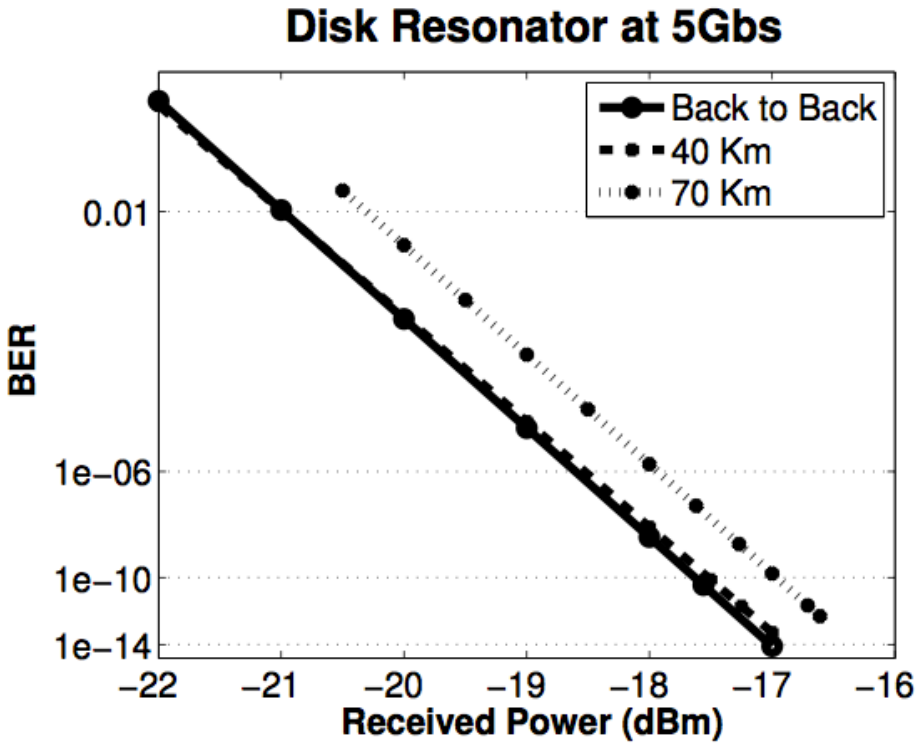
- Lowest Power, Lowest Voltage , 1V, and 3fJ/bit
- Directly CMOS Compatible
- >12Gb/s (expect 25Gb/s soon)

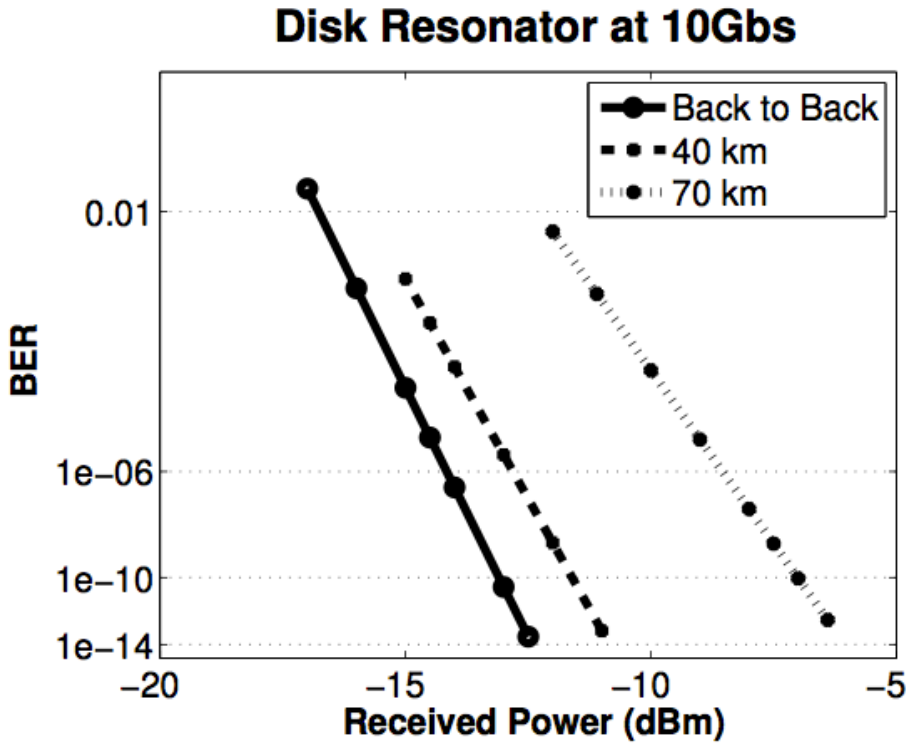




M. R. Watts, W. A. Zortman*, D. C. Trotter, R. W. Young, and A. L. Lentine, "Vertical Junction Silicon Microdisk Modulators and Switches," Optics Express, Vol. 19, pp. 21989-22003, October 2011.

Long Distance Demonstrations





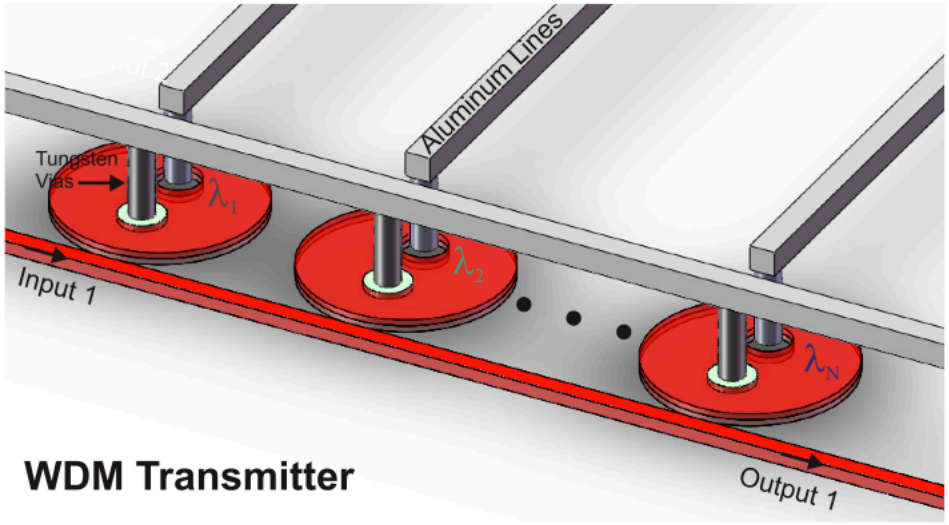
Advantages

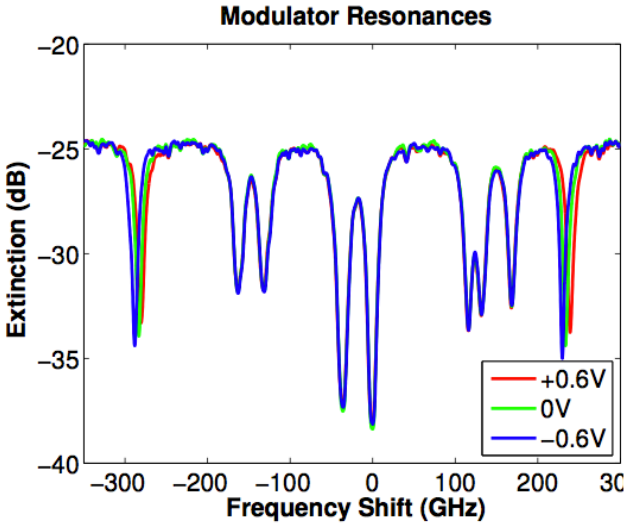
- **5Gb/s:** Power penalty of 0.5dB at 70km
- **10Gb/s:** Chirp and dispersive begin to appear limiting performance a bit, but not fundamentally so

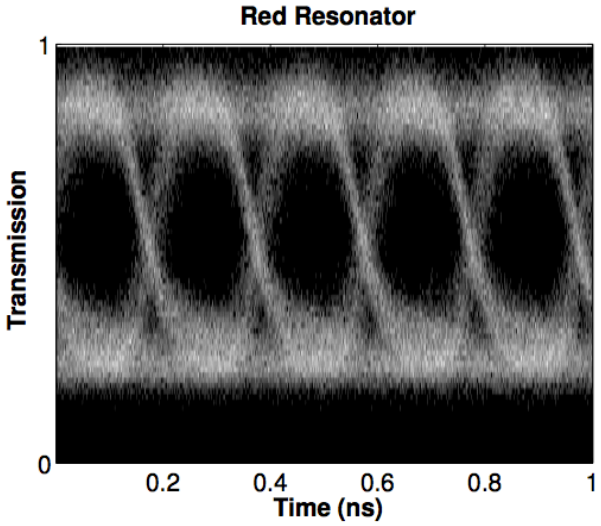
A Simple WDM Demonstration

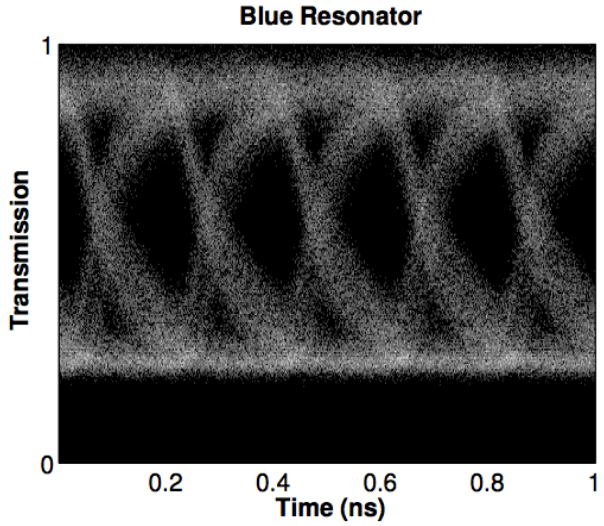
WDM Communications

- **WDM Transmit:** Successful demo across two channels
- **Implications:** WDM is now possible, but of course, now we need filtering

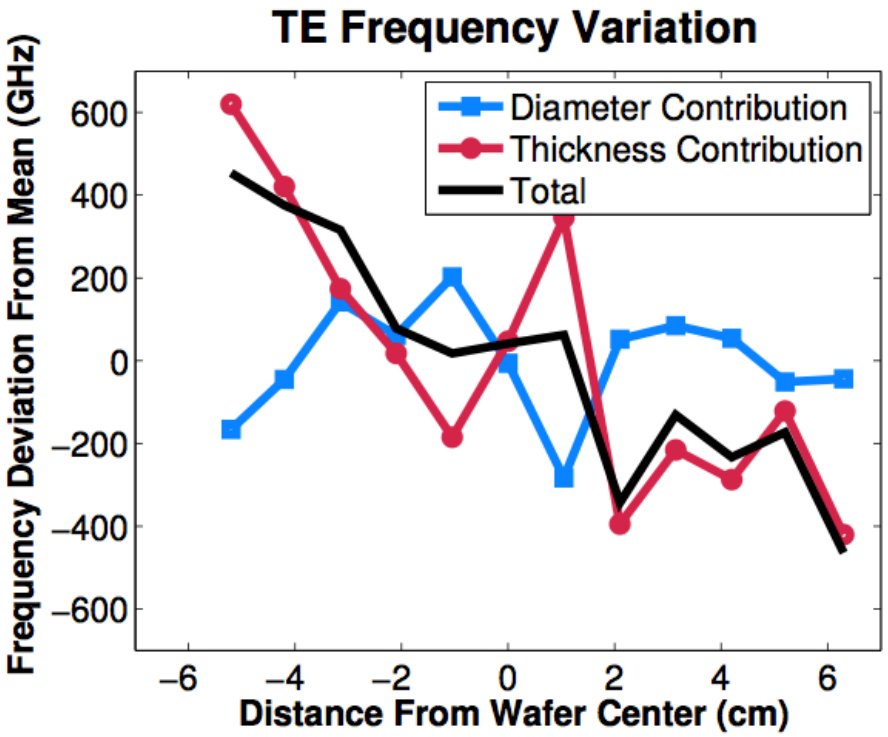


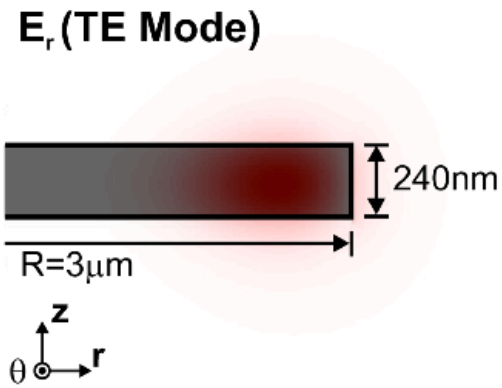


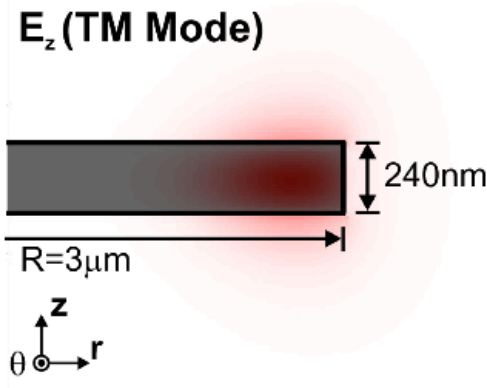




Uniformity: A Challenge for Fabrication





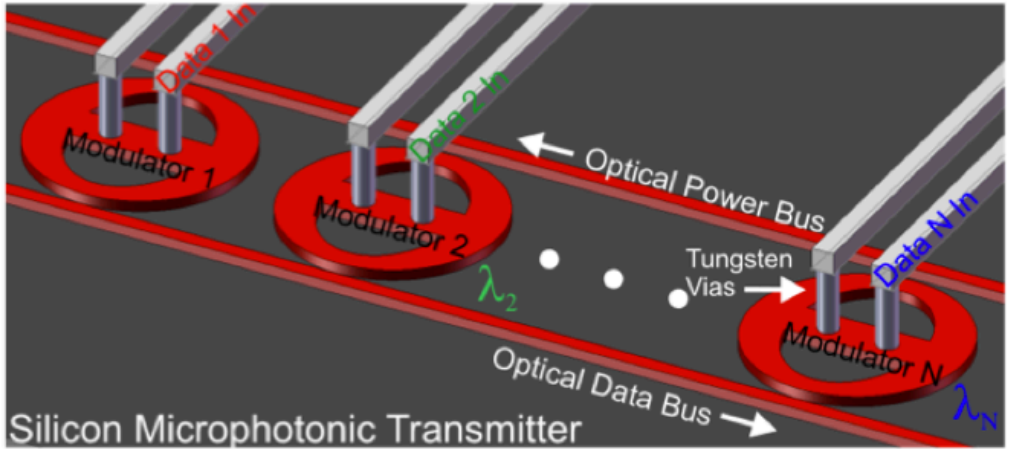


Uniformity

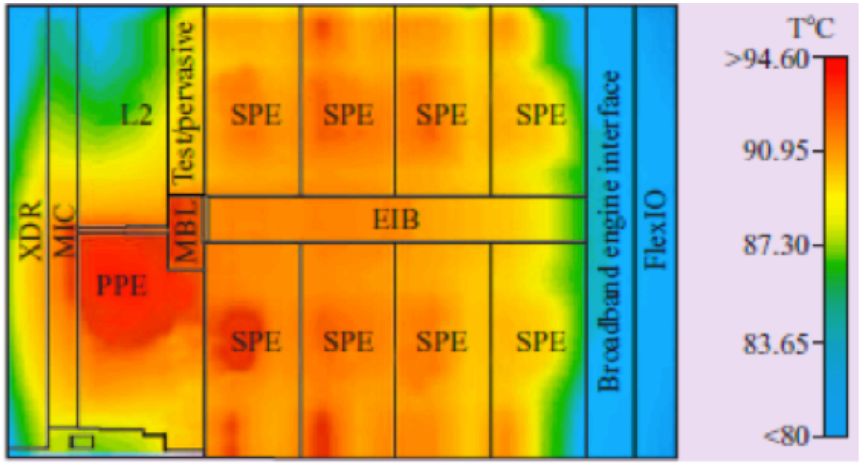
- **Global Variation (across a wafer):** Upto 1THz
- **Local Variation (within a chip):** Generally within 100GHz

Thermal Fluctuations

Microphotonic WDM Transmitter



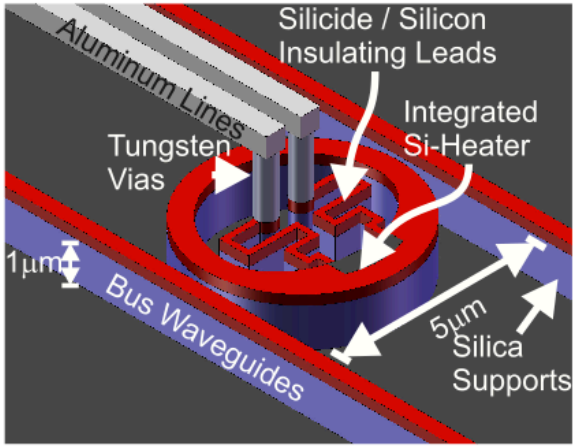
IBM Cell Thermal Map

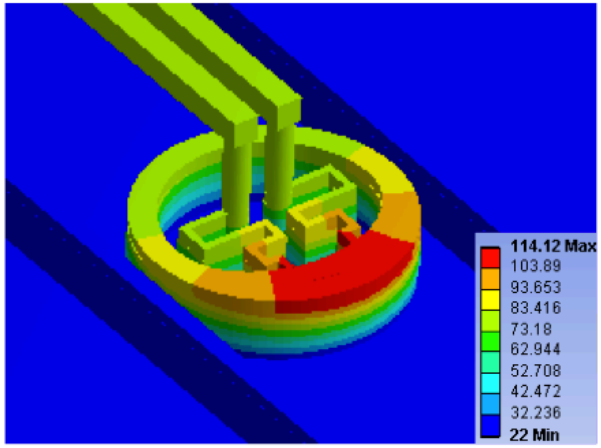


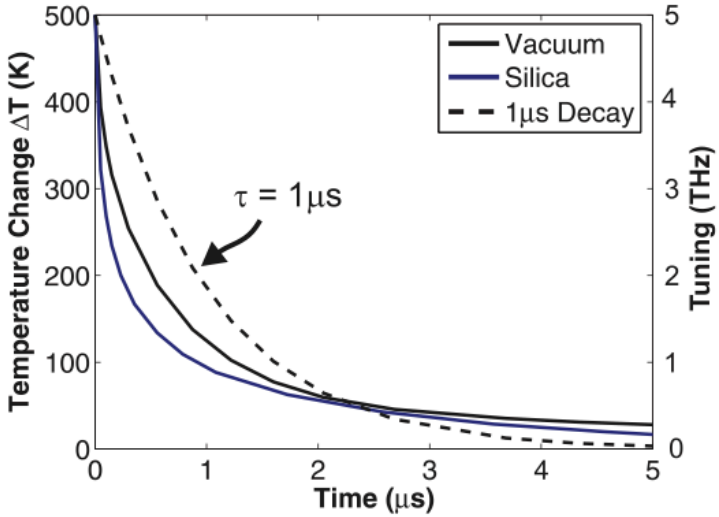
Achieving and Maintaining Resonance Position Challenging

- **Thermal Sensitivity:** Thermo-optic coefficient df/dT≈ 10GHz/K
- **Requirements:** Systems require a 0-to-50oC operating range
- **Challenge:** Can we overcome uniformity and thermal control with one integrated and electronic solution?

Integrated Heaters



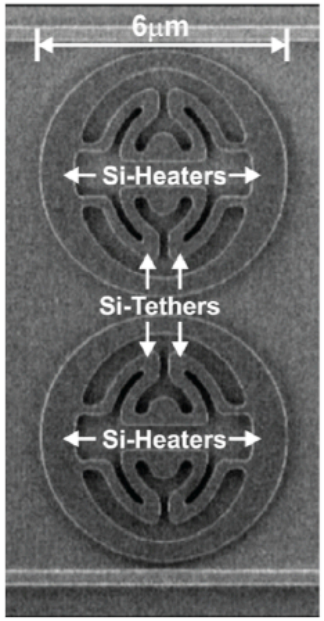


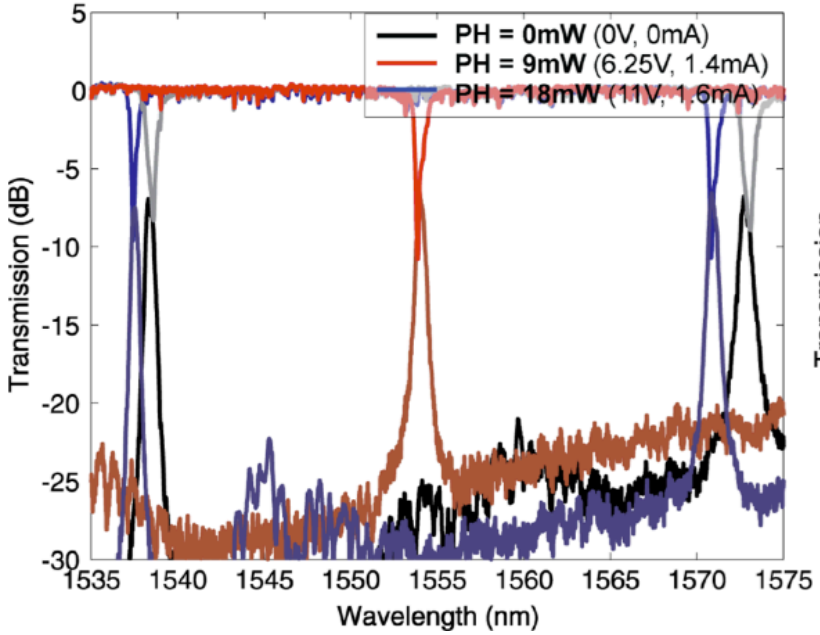


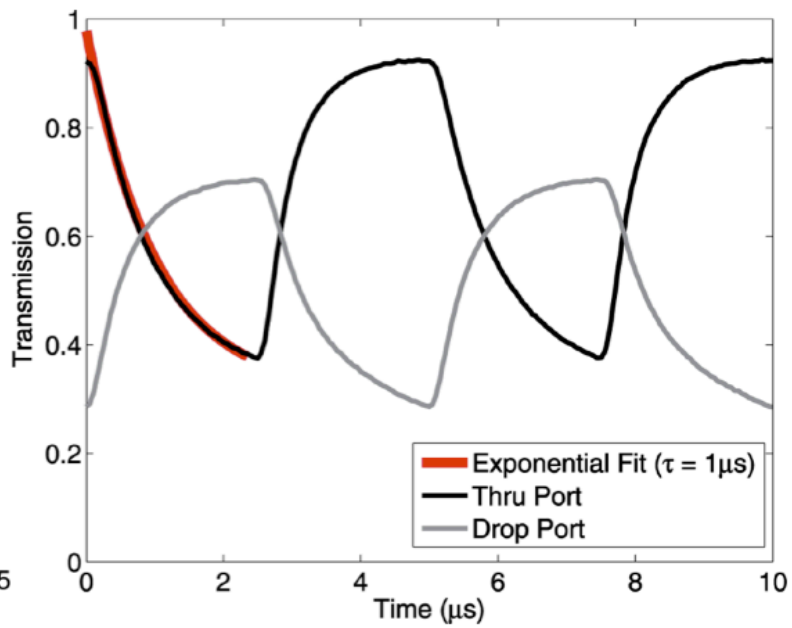
Consider Integrated Microheaters

- **Speed:** FEM indicates 1μs time constant
- **Power:** FEM indicates 5μW/GHz or 50μW/oC with potential to reduce this power by an order of magnitude

Tuning All the Way Across C-Band



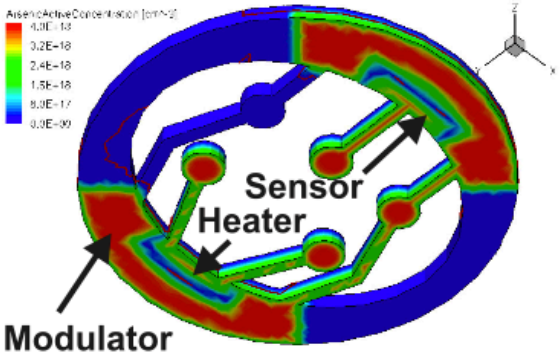


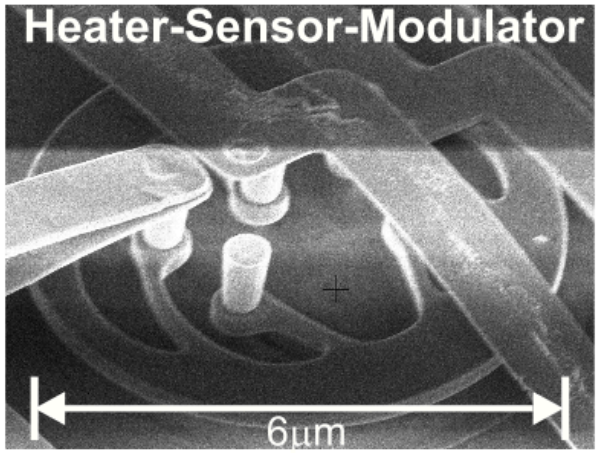


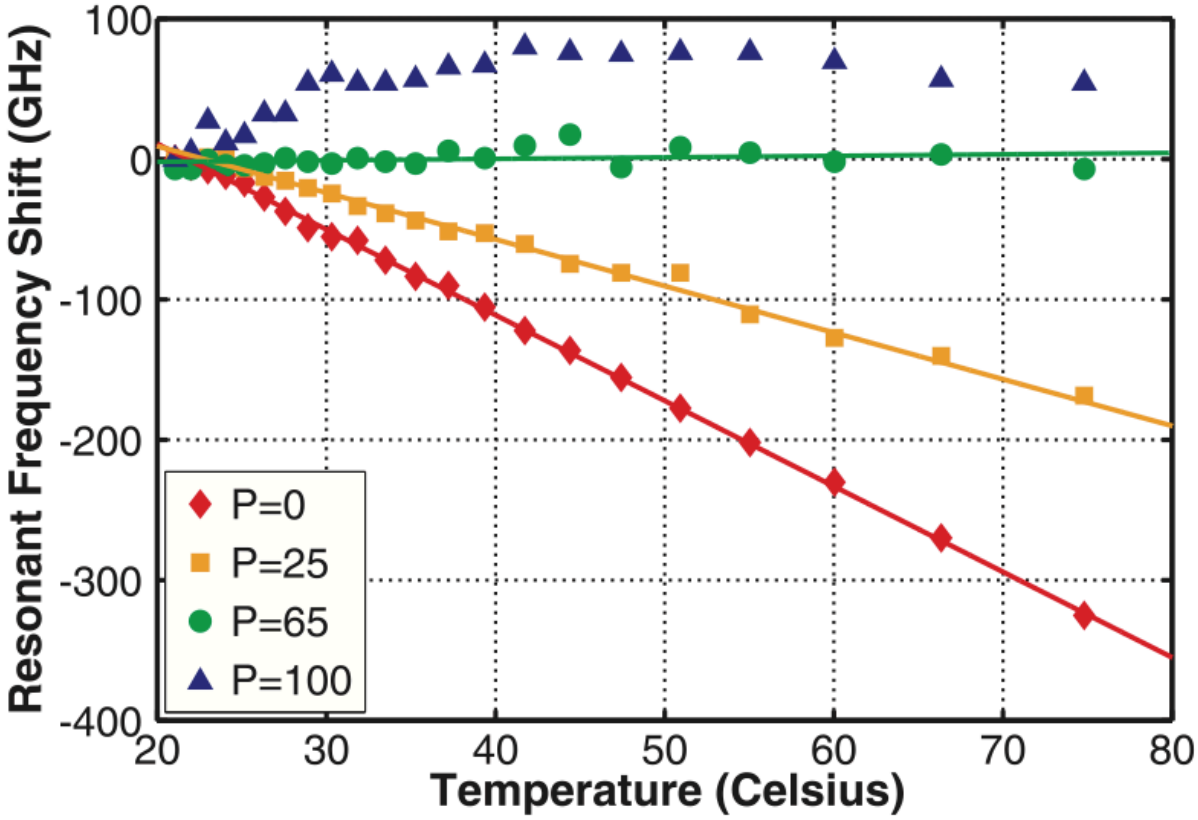
Results

- **Tuning Range:** Able to tune across C-Band
- **Speed:** Demonstrated 1μs time constant
- **Power:** Demonstrated 4.4μW/GHz or 44μW/K
- **Result:** Tuning okay, but what about control?

Integrated Heaters and Temperature Sensors



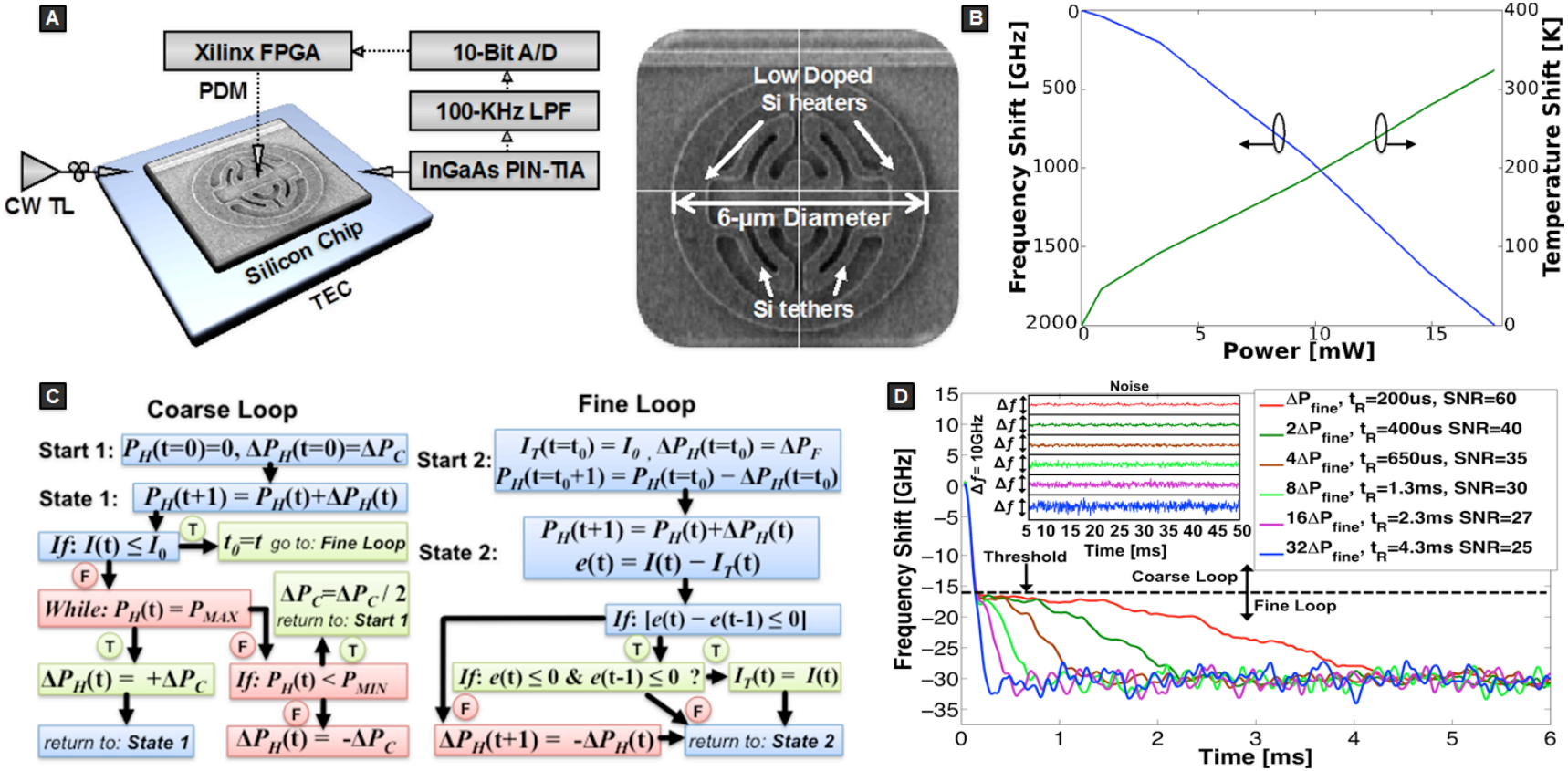




Demonstrated resonance stabilization over 55˚C

- **Sense Temperature:** Build-in diode temperature sensor
- **Lock Temperature:** Use control loop to lock temperature
- Challenge: Requires pre-characterization of ring frequencies

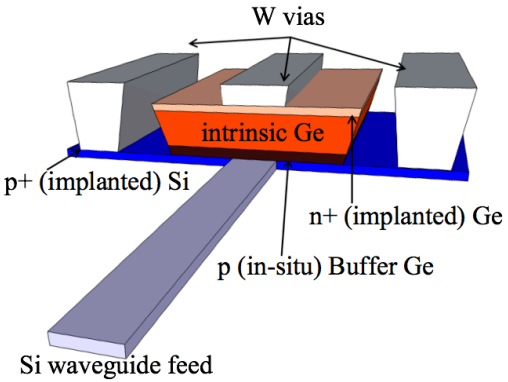
Wavelength Recovery and Control

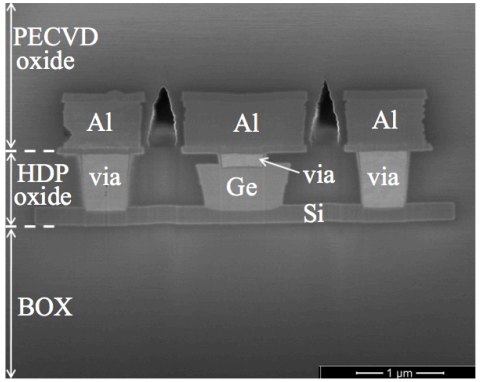


Wavelength Recovery (Amplitude Search Algorithm)

- **Coarse Outer Loop:** Rapidly search for dip in transmission
- **Fine Inner Loop:** Locks and holds onto the center resonance
- **Results:** In less than 1ms, we achieve lock, expect to reduce the transient to ~10μs

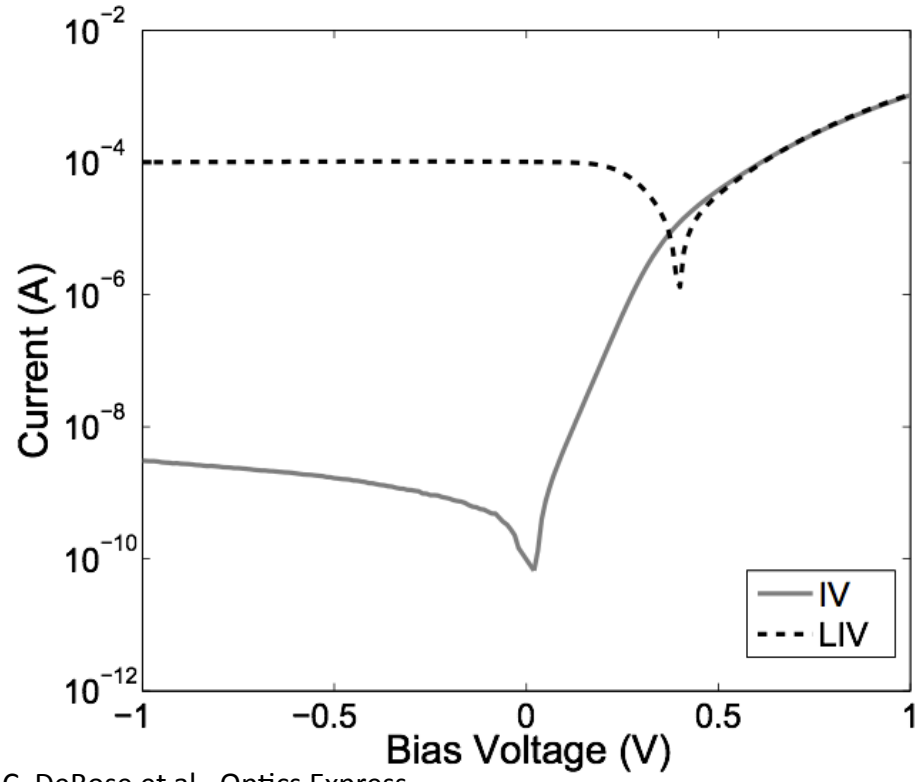
Ge Detectors

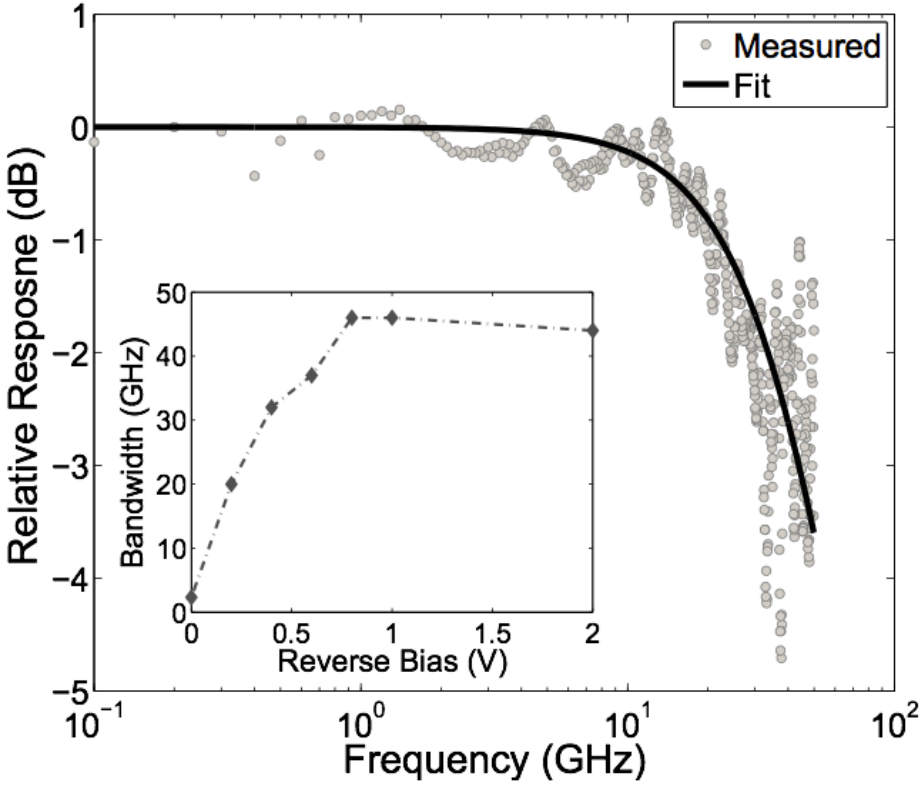




Detector Results

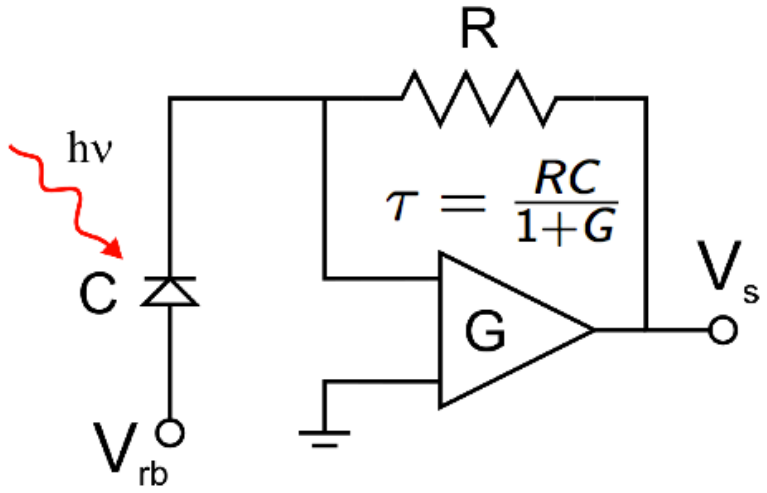
- **Bandwidth:** 45GHz
- **Responsivity:** ~1A/W
- **Dark Current:** ~30nA
- **Capacitance:** C ~ 1fF





C. DeRose et al., Optics Express

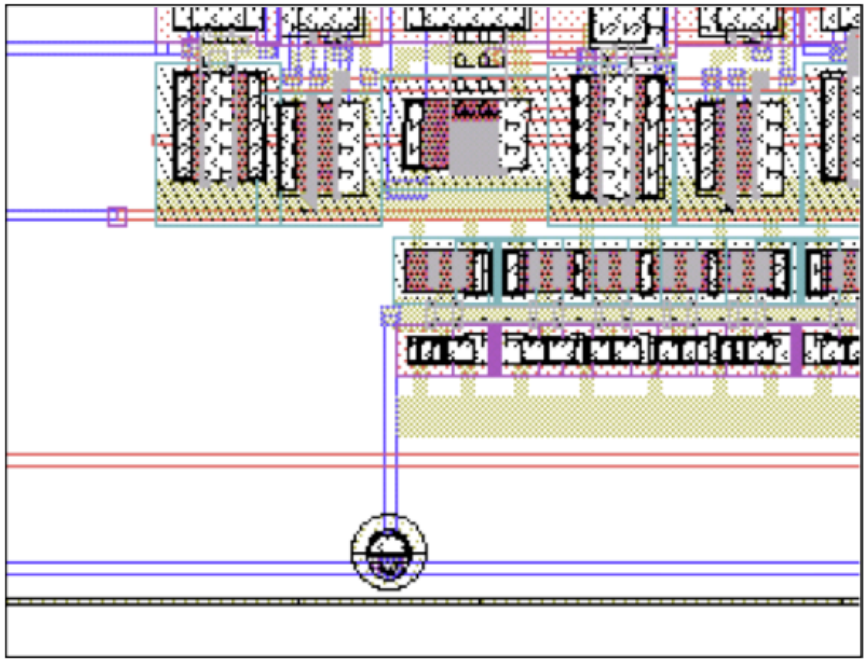
Capacitance Makes a Difference

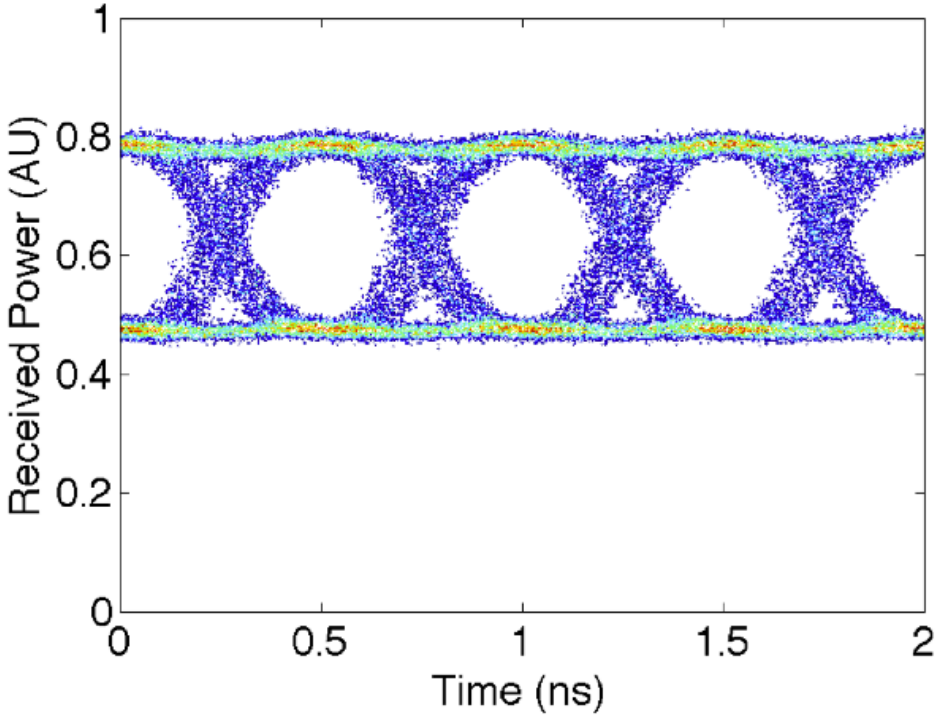


PIN Diodes: Receivers sensitivities $-18dBm$ @10Gb/s. Why?

- **Johnson:** $V_n = \sqrt{4RkT\Delta f} \rightarrow SNR = V_s/V_n = \frac{I_0\sqrt{R}}{\sqrt{4kT\Delta f}}$
- **Case 1:** C = 300fF. For $\tau = 15$ps, $R = 500$. $G = 9$, $SNR = 14$dB $\rightarrow 15\mu$W (-18dBm). Add 20dB loss $\rightarrow \boxed{150\frac{\text{fJ}}{\text{bit}}}$
- **Case 2:** $C = 1$fF$\rightarrow R = 150000$, $G = 9$, $SNR = 14$dB$\rightarrow 0.8\mu$W (-31dBm). Add 20dB loss $\rightarrow \boxed{8\frac{\text{fJ}}{\text{bit}}}$

Direct Integration with CMOS

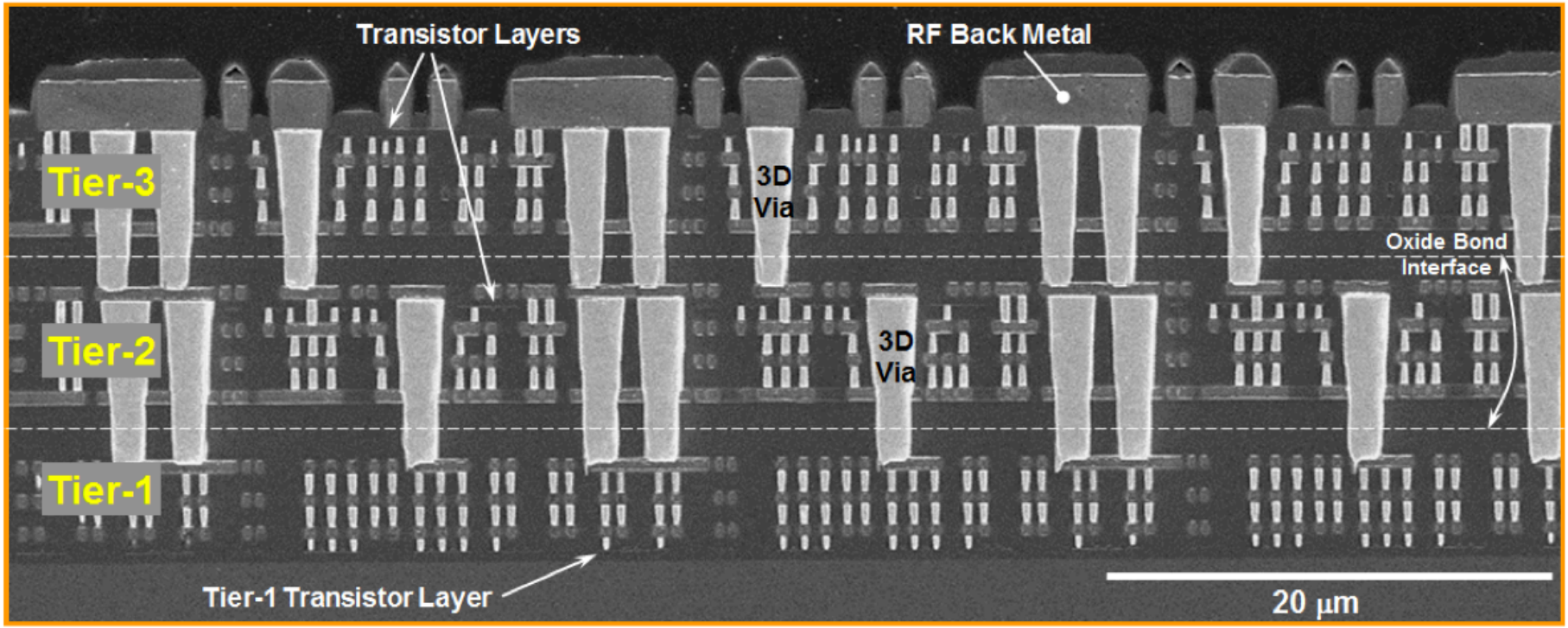




Direct Integration (using SOI for both gate and SiPhotonics)

- **Benefits:** Lowest capacitance, highest density, low cost
- **Challenges:** SOI Silicon layer is getting too thin in advanced CMOS so at 45nm and beyond direct integration is difficult

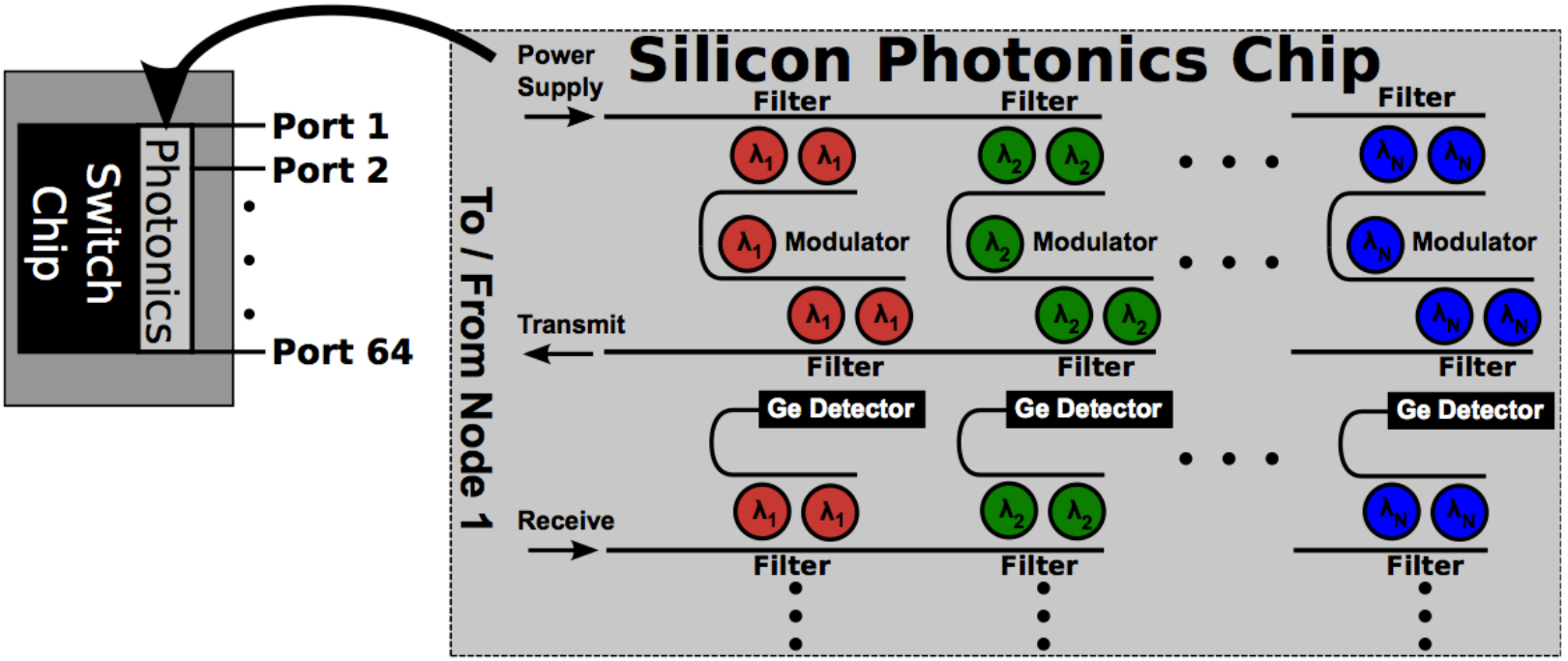
3D Integration with CMOS



3D Integration

- **Power:** Capacitance of 1fF → ultralow power
- **Density:** 1µm via enables densely interconnected photonics and massive bandwidth out of the chip

Power Consumption



Power Budget (~1pJ/bit, 0.5pJ/bit Optics, 0.5pJ/CMOS)

- **Uniformity:** 200GHz→ 200 µW → 1.4mW→ 140 fJ ring bit
- **Thermal Control:** 20°C→ 200 µW → 1.4mW→ 140 fJ ring/bit
- **Settled:** Modulation (~ 10 fJ), filtering (~ 0.3 pJ), optical bitbit power (< 10 fJ) and detection (< 70 fJ)→ ~ 0.4pJ

Which Markets Make Sense

Designation	Speed	Wavelength	Media	Technology	Distance
40GBASE-LR4	40 Gbps	1310 nm	Singlemode	2 Fibers Using WDM	10 km
40GBASE-SR4	40 Gbps	850 nm	LO 50 µm Multimode (OM4)	8 Fibers (Tx & Rx) @ 10 Gbps	125 m
40GBASE-CR4	40 Gbps	NA	Parallel Coaxial Copper (Twinax) Cabling	8 Pairs 4 Tx & 4 Rx @ 10 Gbps	7 m
100GBASE-ER4	100 Gbps	1310 nm	Singlemode	2 Fibers Using WDM	40 km
100GBASE-LR4	100 Gbps	1310 nm	Singlemode	2 Fibers Using WDM	10 km
100GBASE-SR10	100 Gbps	850 nm	LO 50 µm Multimode (OM4)	20 Fibers (Tx & Rx) @ 10Gbps	125 m
100GBASE-CR10	100 Gbps	NA	Parallel Coaxial Copper (Twinax) Cabling	20 Pairs 10 Tx & 10 Rx @ 10 Gbps	7 m

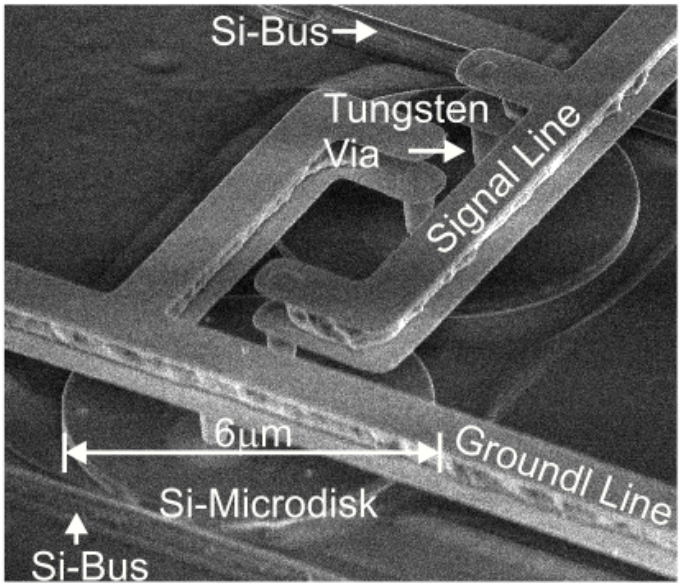
IEEE802.3ba Draft Standard for 40 and 100 GbE

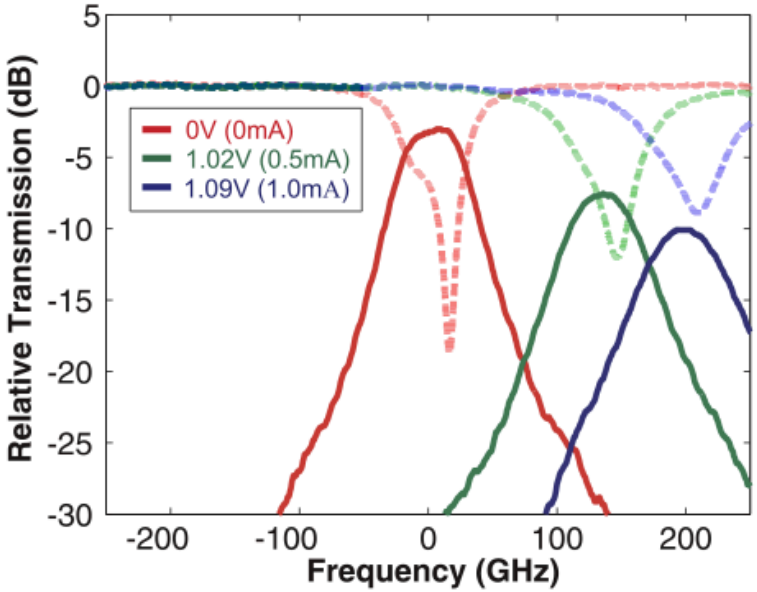
Standards Addressed Readily Addressed by SiPhotonics

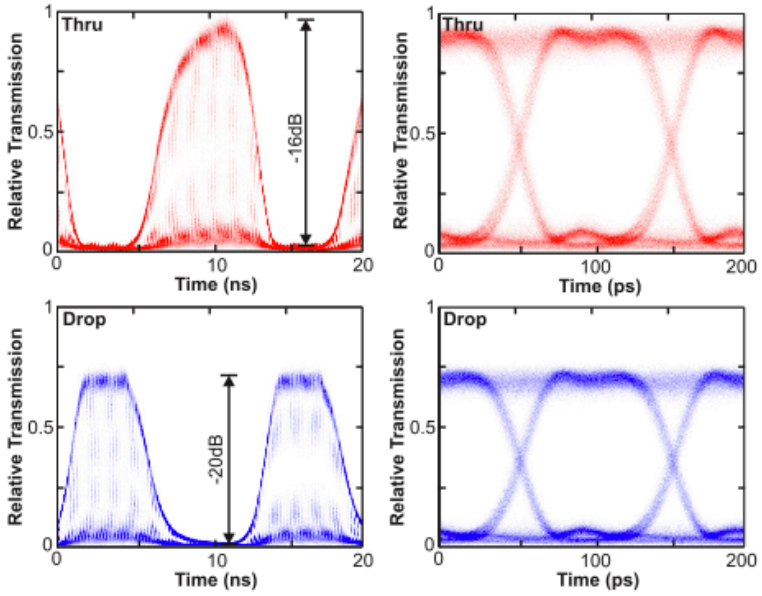
Impact: Likely 100GbE LR/ER standards will be first addressed

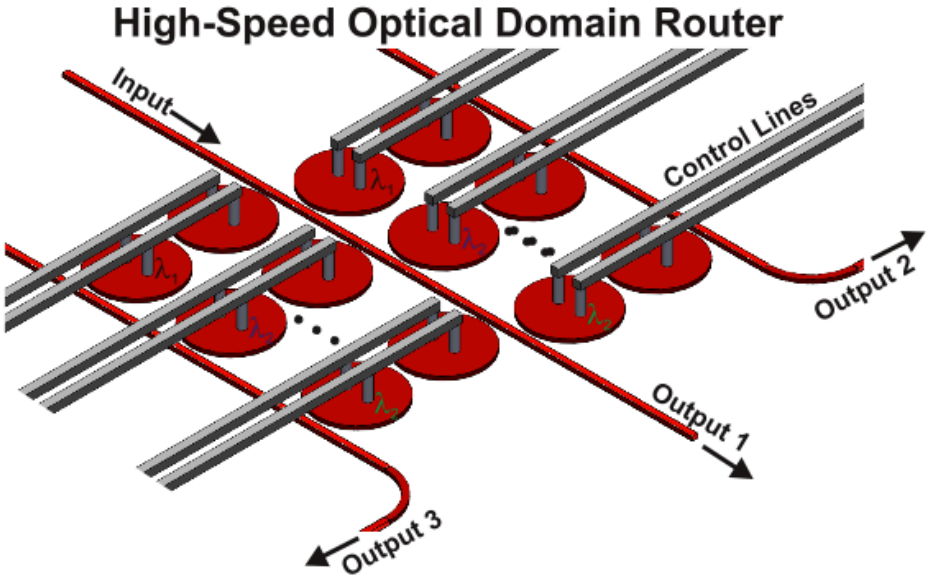
Future: Impact is likely to be greatest at higher rates, 400GbE, 1TbE

High-Speed Optical Domain Switching and Routing

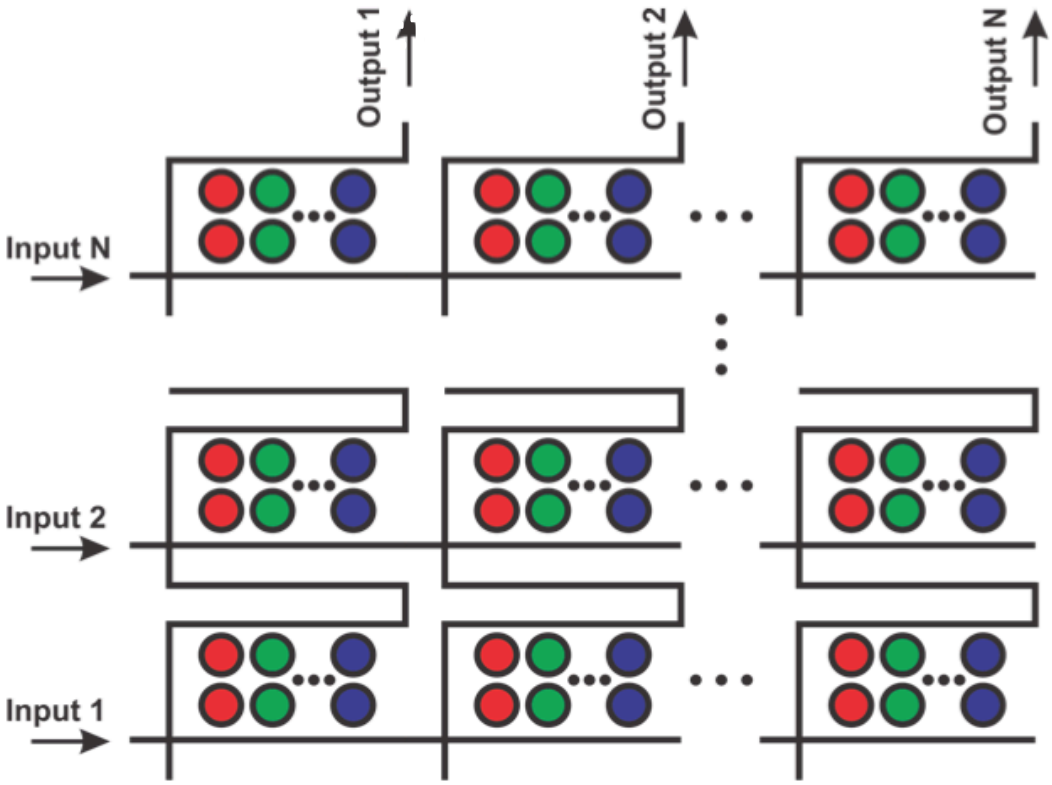


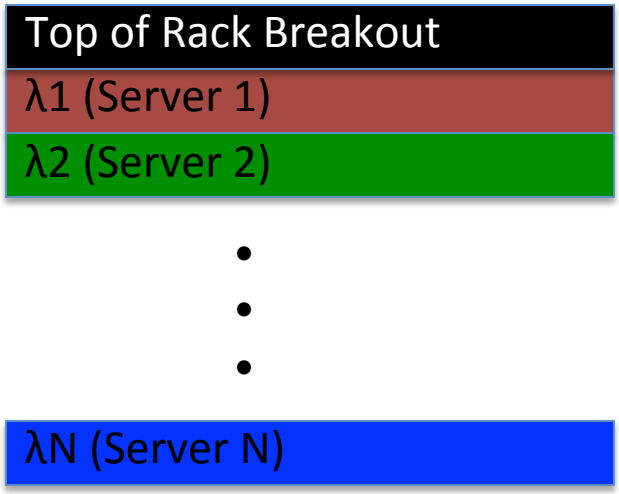






Architecture Optical Domain Switching

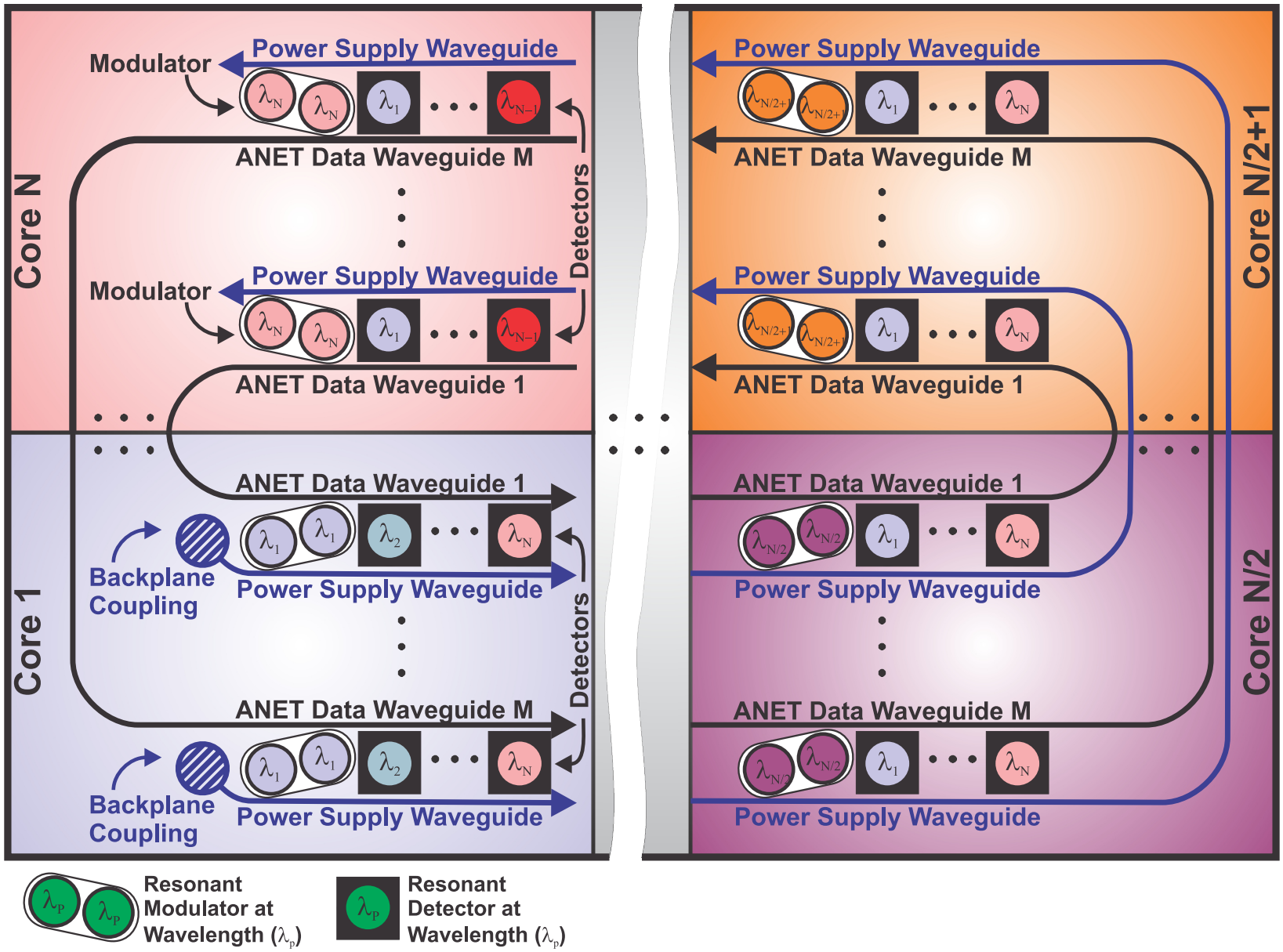




Benefits of Optical Domain Switching

- **Power Consumption:** No O-E-O conversions
- **Switch Radix:** Can achieve a switch radix ~1000 or more
- **Challenges:** Generally, optical domain switching is circuit based – achieving packet based switching would require a lot of effort

All-to-All WDM Networks



Summary

Silicon Photonics

- **Rapidly Maturing:** All devices work, we really just need to get to system demonstrations
- **Question:** What do traditional interconnect folks want to see for a demonstration to be convinced?

Market Need

- **What:** 100GbE and beyond
- **When:** A guess would be 2015 / 2016
- **What Market:** Likely HPC first

Acknowledgements

- **Funding:** DARPA-MTO (S. Raman and J. Shah), Sandia National Labs, APIC Corporation
- **Contributors:** Zortman, Trotter, Timurdogn, Biberman, Coolbaugh