# Energy-Efficient VLSI Interconnects

Patrick Chiang
Oregon State VLSI Research Group
Santa Fe, Oct. 13, 2010

# Fighting words

- What means more to you – performance or energy?
- Dark Silicon – more silicon than you have energy
  - Why is wider parallel and slow a problem?
- Multicore → lower CLK → deep parallelism
  - Low VDD at the extreme

# ITRS Roadmap 2007

- "at 0.13um approximately 51% of microprocessor power was consumed by interconnect, with a projection that without changes in design philosophy, in the next five years up to 80% of microprocessor power will be consumed by interconnect"

# Interconnect Energy Wall

- "The Energy and Power Challenge is the most pervasive of the four, and has its roots in the inability of the group to project any combination of currently mature technologies that will deliver sufficiently powerful systems in any class at the desired power levels."

- A key observation of the study is **that it may be easier to solve the power problem associated with base computation than it will be to reduce the problem of transporting data from one site to another** - on the same chip, between closely coupled chips in a common package, or between different racks on opposite sides of a large machine room…

- This is a "solved" problem if we are WILLING to deal with the implications

Source:
DARPA Exascale
Study, Sep. 2008

# Overview

- Motivation: Why high-speed links?

- Tutorial on channel loss
  - Channel loss limits:
    - Data Rate
    - Energy consumption
  - Optical interconnect
  - 3D Integration

- Energy-Efficiency
  - Assume channel is benign
  - 1mW/Gbps Off-Chip Serial Links
  - Fundamental limits to low-power on-chip links

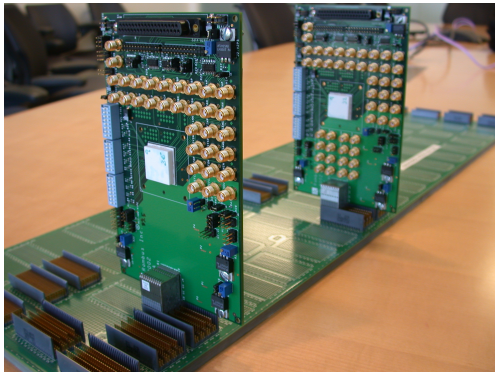
# Why serial links

- Wired communications from one chip to another
  - Memory->CPU (DDR3)
  - CPU->CPU (Intel Quickpath; AMD HyperTransport)
  - Router backplane (line-card to backplane)

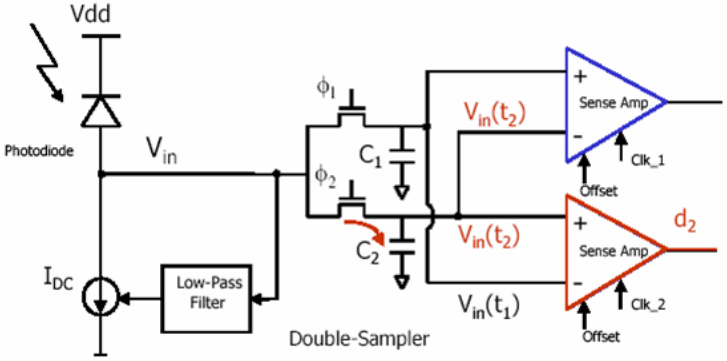
- Typically a major bottleneck for:
  - System performance
  - Power consumption
  - Chip complexity

- Why off-chip bandwidth scales slowly?
  - CPU transistor density increasing
  - Memory capacity increasing
  - Interconnect does not improve
  - # of pads does not increase

- Use Optics ??

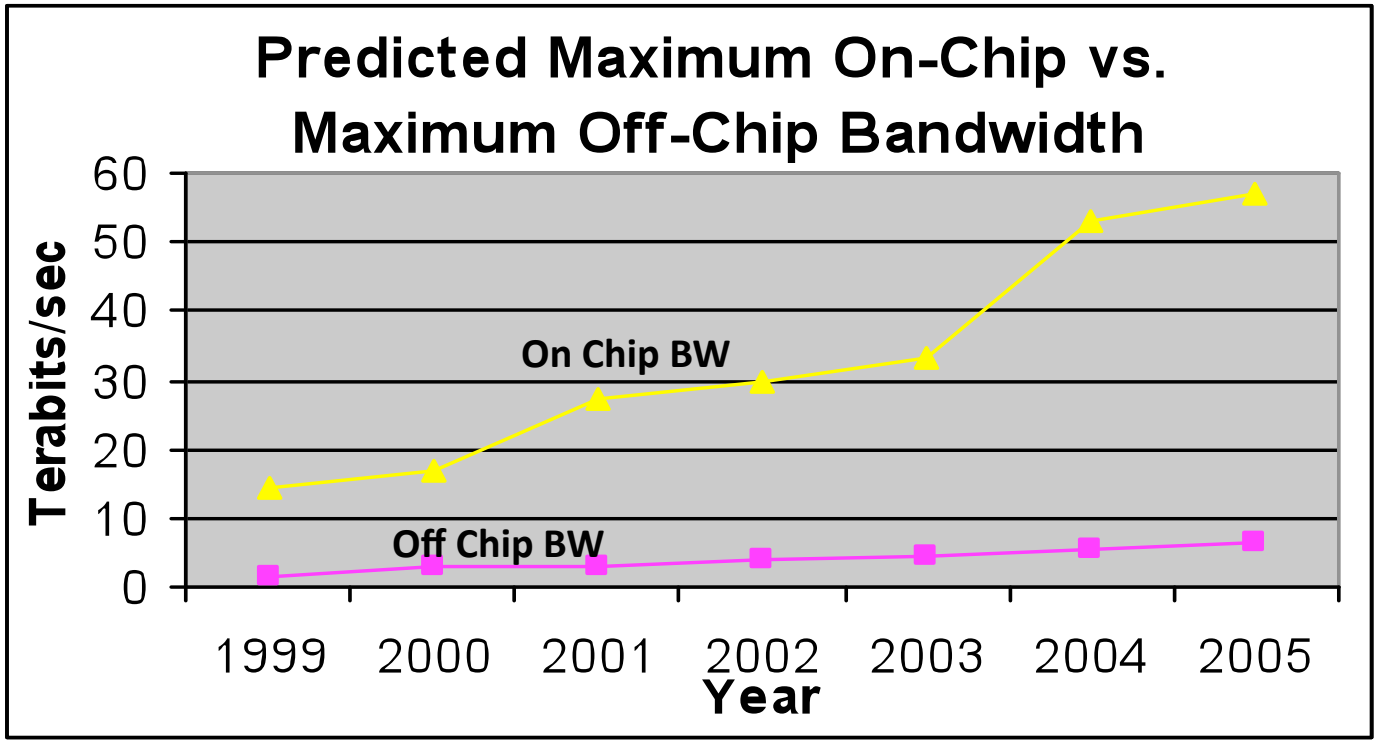


Router Backplanes




Optical Front-Ends

# I/O Bandwidth is Limiting Factor



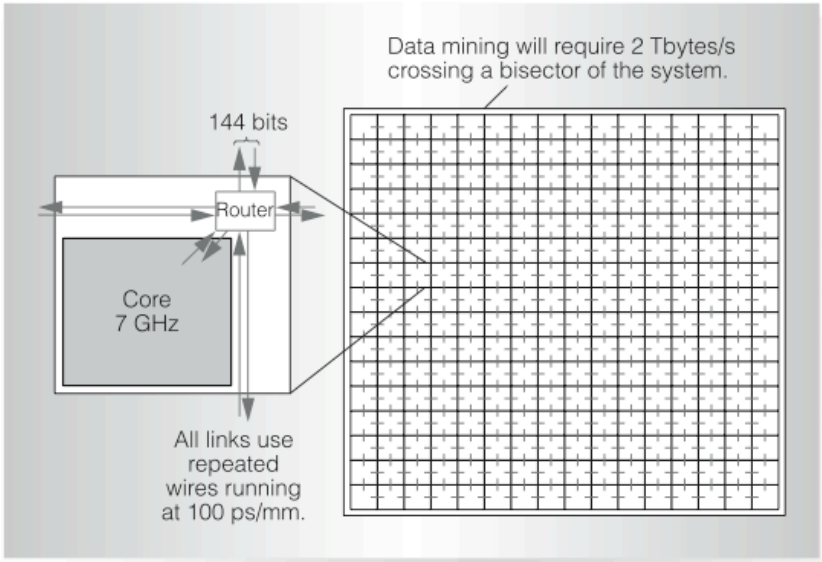

- GOAL:
  - high BW (Gbps / pad)
  - Energy-Efficient (mW/Gbps or pJ/bit)
  - Low-area (100's to 1000's on a die)

## The Energy Wall -- Computation is free; Communication is expensive

- Future processor: 256 Cores

- Total on-chip bandwidth
  - If $P_{WIRE}$=0.25mW/Gbps/mm, →150W

→ Need > 10x reduction



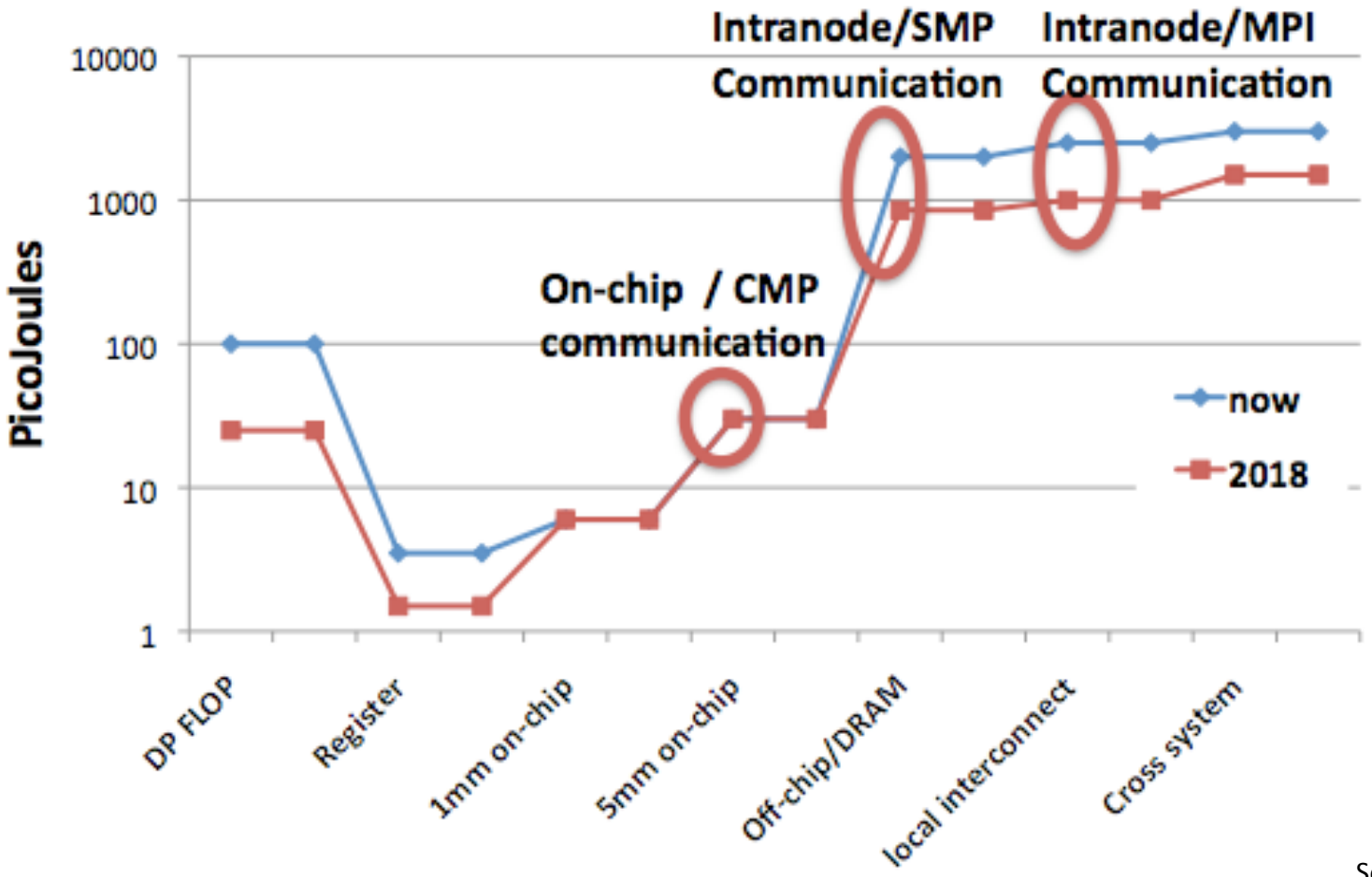

1. A CMP machine in 2015: 256 cores in a 16 × 16 grid.

Source:
IEEE Micro, Dec. 2007

- Total off-chip bandwidth
  - Predicted off-die bandwidth: 5.12Tbps
  - 512 Serial links @ 10Gbps (10mW/Gbps) = 50W

→ Need > 10x reduction (1mW/Gbps)

# Energy Barrier




Source:
Exascale Roadmap
Meeting, Dec. 2009

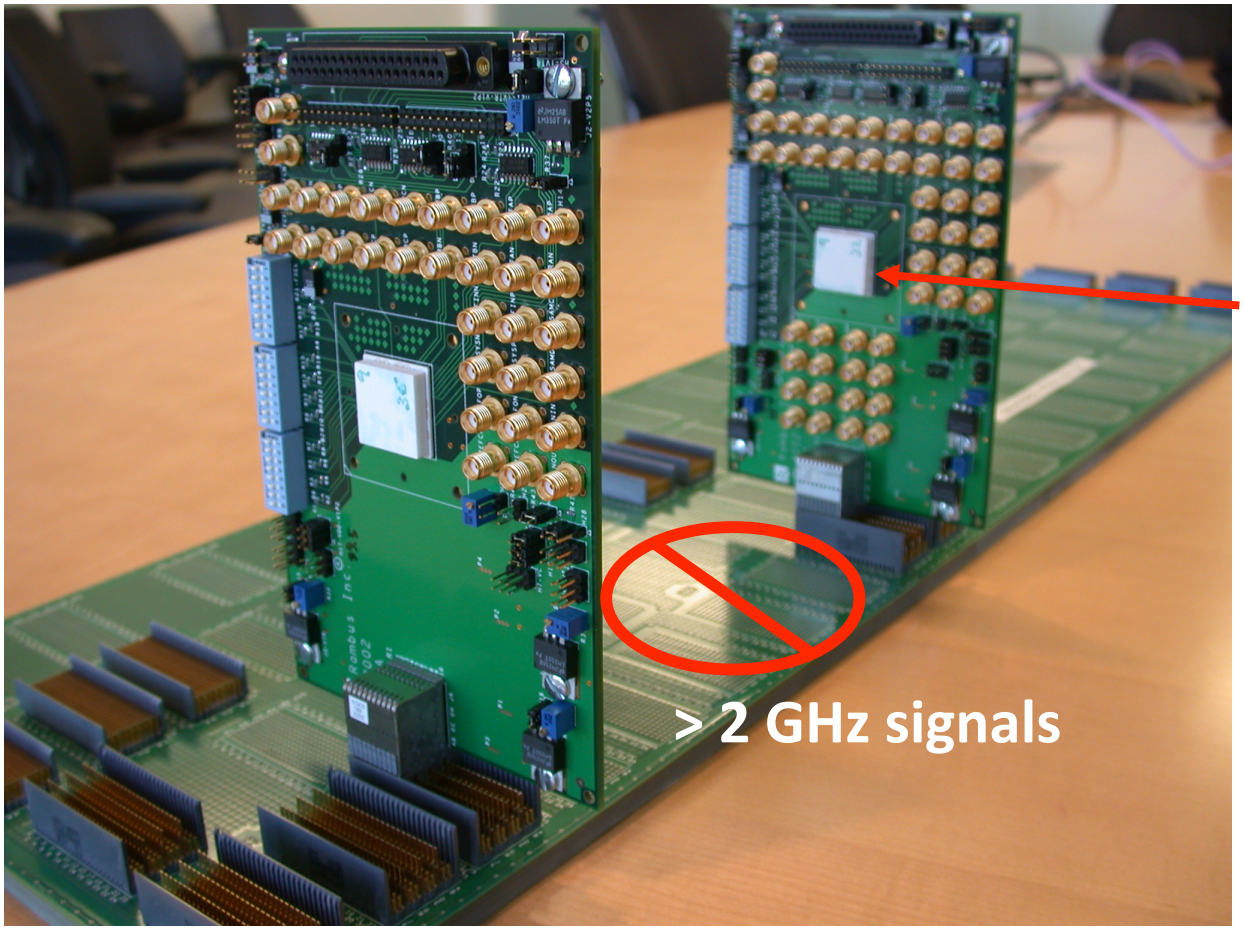
# Overview

- Motivation: Why high-speed links?

- Tutorial on channel loss
  - Channel loss limits:
    - Data Rate
    - Energy consumption
  - Optical interconnect
  - 3D Integration

- Energy-Efficiency
  - Assume channel is benign
  - 1mW/Gbps Off-Chip Serial Links
  - Fundamental limits to low-power on-chip links

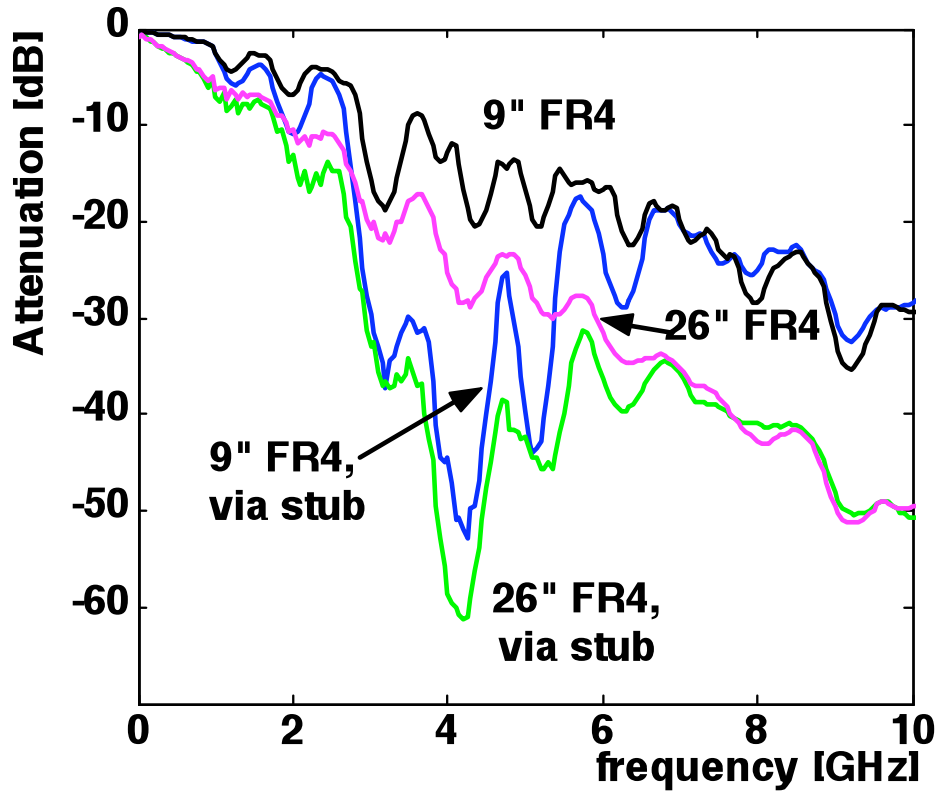
# What makes it more challenging




- Now, the bandwidth limit is in wires
- Take Home Note: tell vendor to throw kitchen sink at PCB

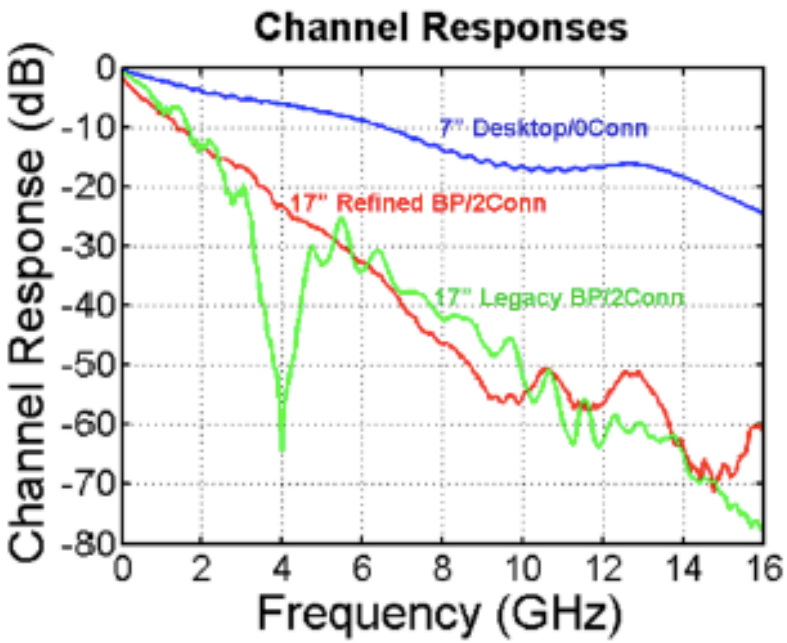
# Backplane channel

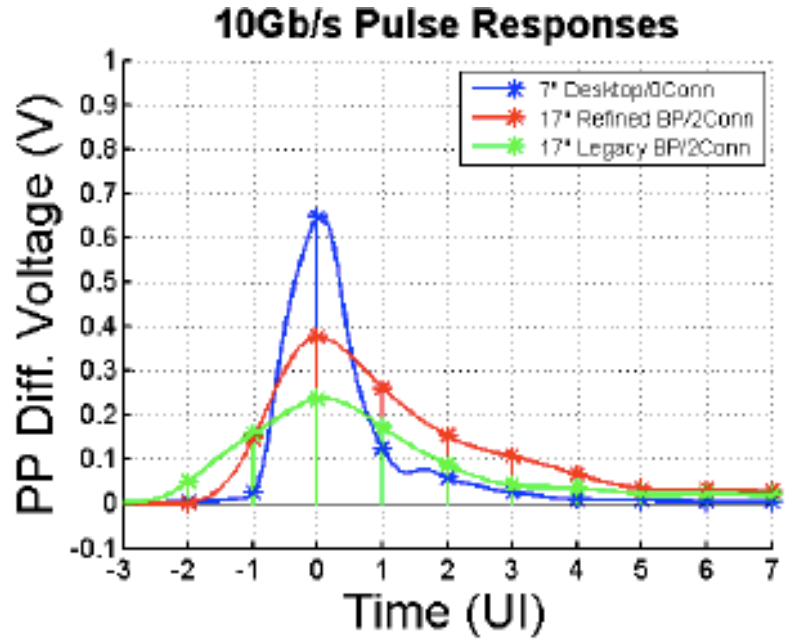
- Loss is variable
  - Same backplane
  - Different lengths
  - Different stubs
    - Top vs. Bot
- Attenuation is large
  - >30dB @ 3GHz
  - But is that bad?
- Required signal amplitude set by noise

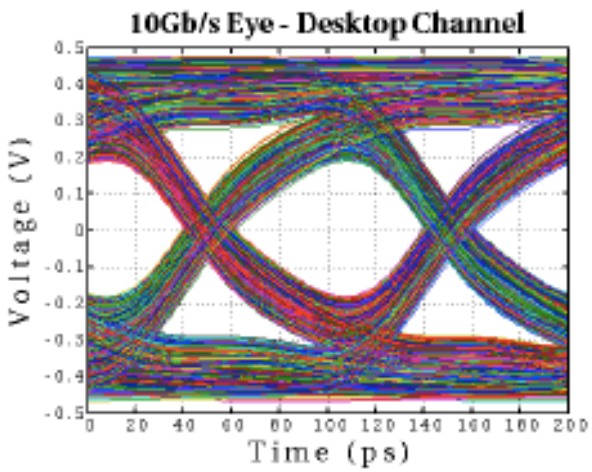


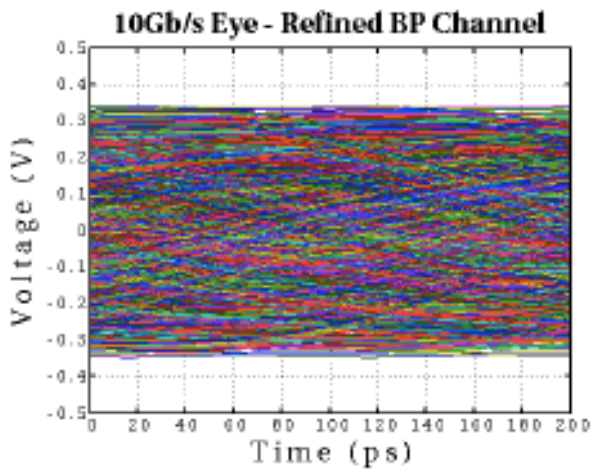
# Channel Responses

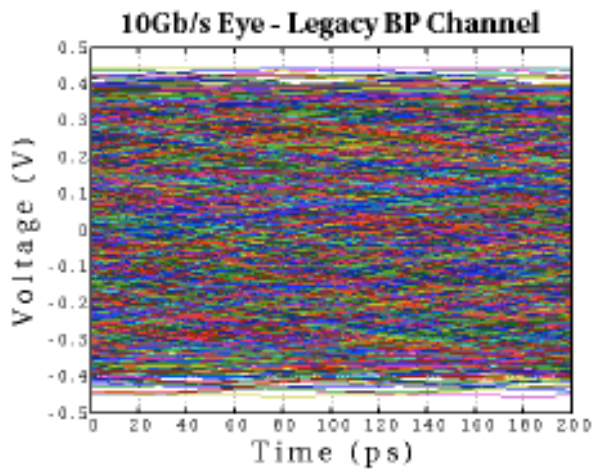
Source:
Palermo, Texas A&M



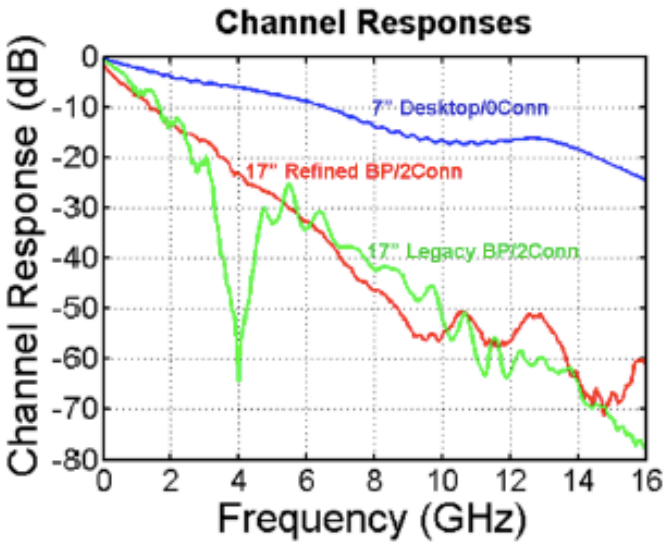



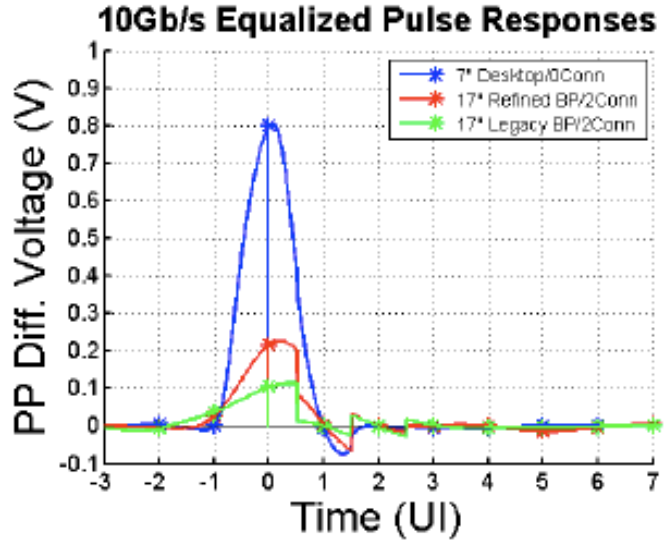

http://eecs.oregonstate.edu/research/vlsi

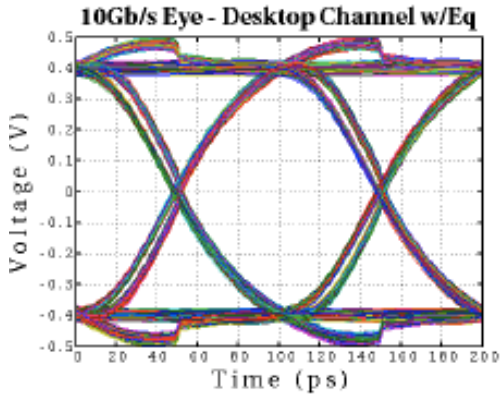
Patrick Chiang, Oregon State University

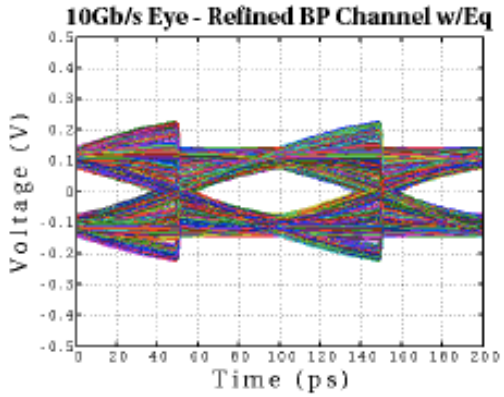
# 10Gbps Equalization

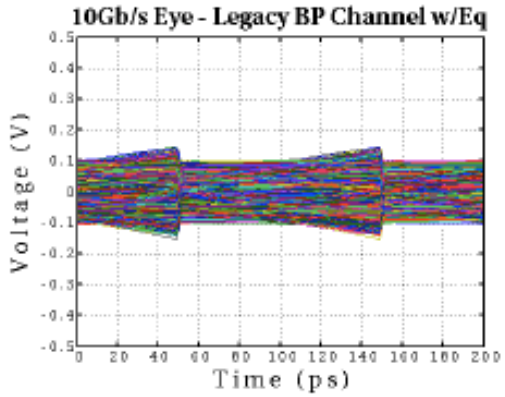
Source:
Palermo, Texas A&M



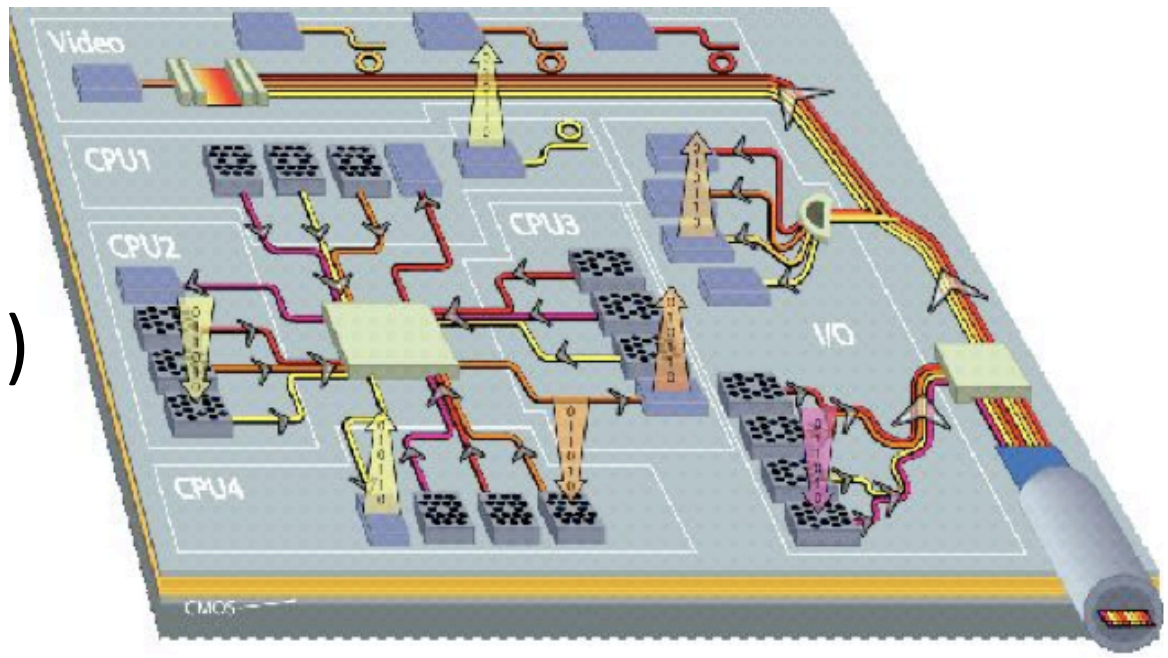




- NOTE: Tell vendor to throw kitchen sink at PCB

# Optics solves the channel

- Interconnect is now optical
  - On-chip wires are no longer RC
  - Off-chip backplane channels are no longer T-lines

- Multimode loss
  = 2.5dB (@ 850nm)



# Optics doesn't solve the electronics

TABLE II
TRANSCEIVER PERFORMANCE SUMMARY

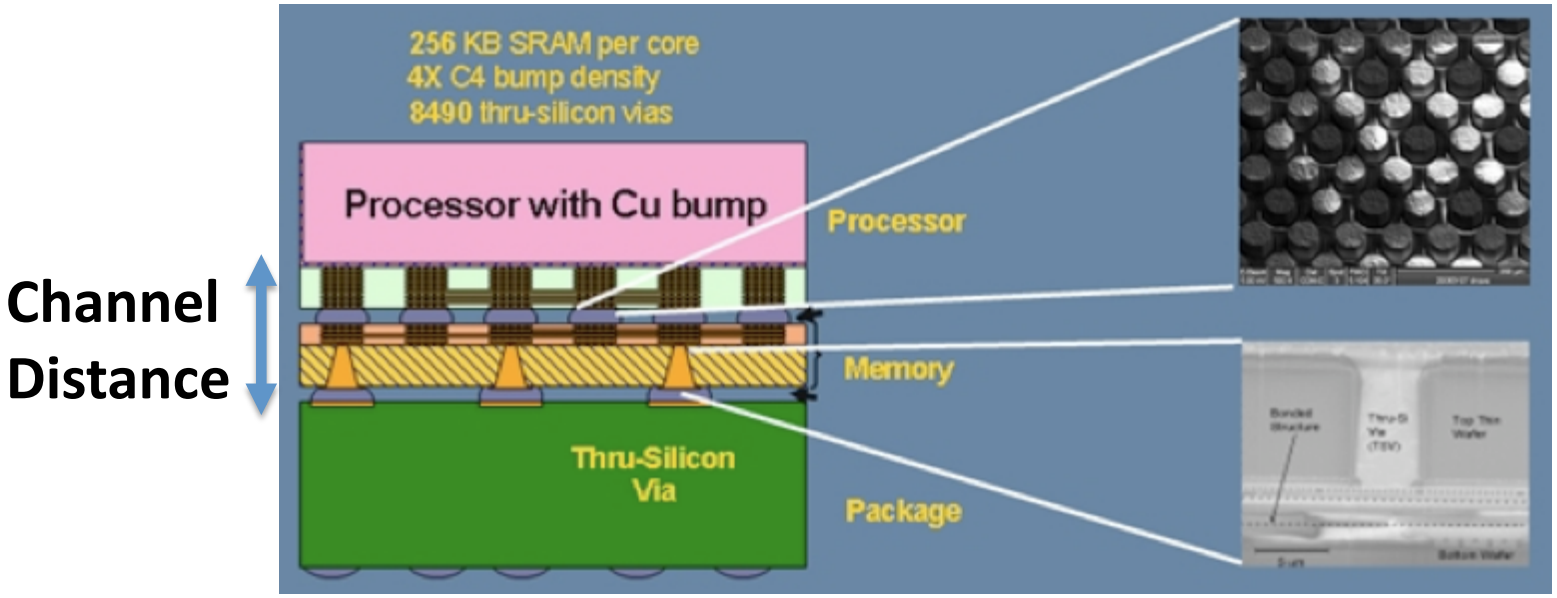
| | |
|---|---|
| **Technology** | **90nm CMOS** |
| **Supply Voltages** | **Vdd=1V, LVdd=2.8V** |
| **Data Rate** | **5-16Gb/s** |
| **ER** | **3dB** |
| **Average Optical Launch Power** | **3.1dBm** |
| **VCSEL $I_{avg}$** | **6.2mA** |
| **RX $C_{in}$** | **440fF** |
| **RX Sensitivity, BER=$10^{-10}$** | |
| **10Gb/s** | **-9.6dBm** |
| **16Gb/s** | **-5.4dBm** |
| **TX Clock Jitter @ 16Gb/s** | **$1.90ps_{rms}$, $14.0ps_{pp}$** |
| **RX Clock Jitter @ 16Gb/s** | **$1.90ps_{rms}$, $15.1ps_{pp}$** |
| **Area** | |
| TX Output | $0.017mm^2$ |
| TX PLL | $0.013mm^2$ |
| RX Front-End | $0.025mm^2$ |
| RX PLL/CDR | $0.050mm^2$ |
| **Total** | **$0.105mm^2$** |
| **Power Consumption @ 16Gb/s** | |
| TX Output | 48mW |
| TX PLL | 23mW |
| RX Front-End | 23mW |
| RX PLL/CDR | 35mW |
| **Total** | **129mW (8.1mW/Gb/s)** |

**If Transmit power=0,**
**→ 5mW/Gbps**

**Palermo, JSSC08**

# 3D solves the channel

- Memory->CPU interface is closer
  - Distances are shorter




- Electronics at the interfaces still required

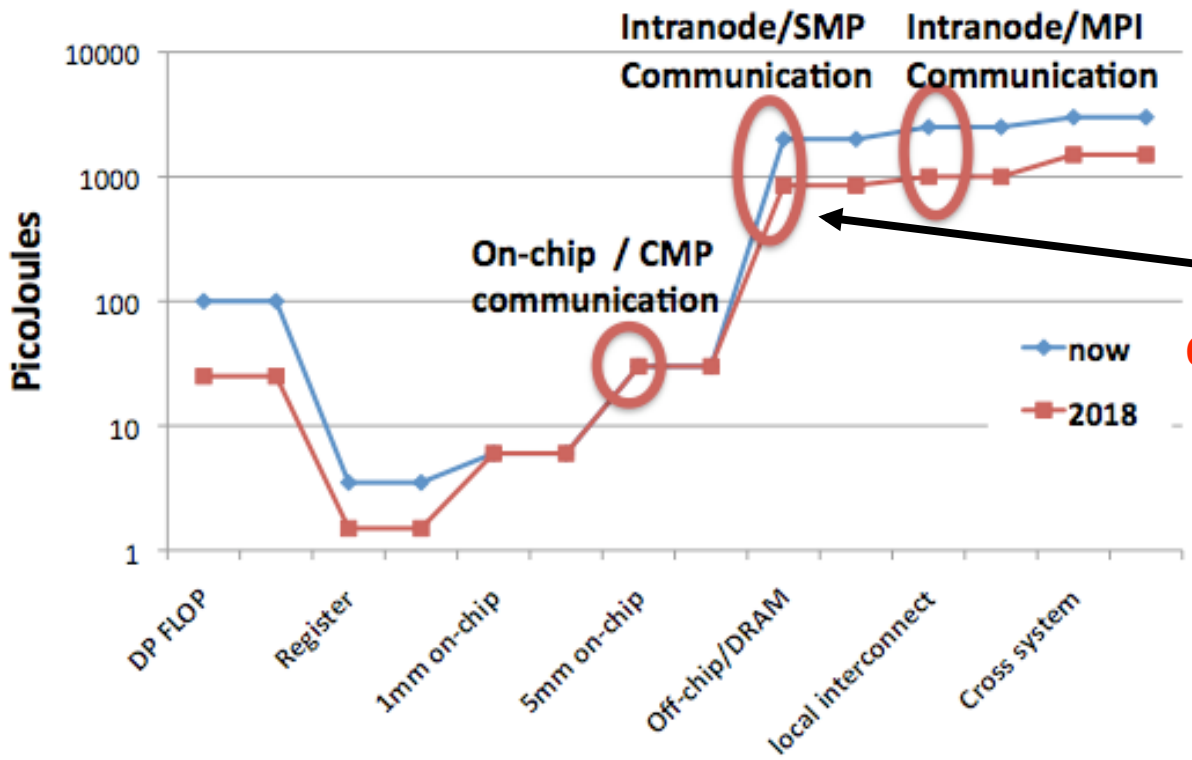
# Overview

- Motivation: Why high-speed links?

- Tutorial on channel loss
  - Channel loss limits:
    - Data Rate
    - Energy consumption
  - Optical interconnect
  - 3D Integration

- Energy-Efficiency
  - Assume channel is benign
  - < 1mW/Gbps Off-Chip Serial Links
  - Fundamental limits to low-swing on-chip links

# Problem 1: Off-chip I/O power is not scaling




**CONVENTIONAL**

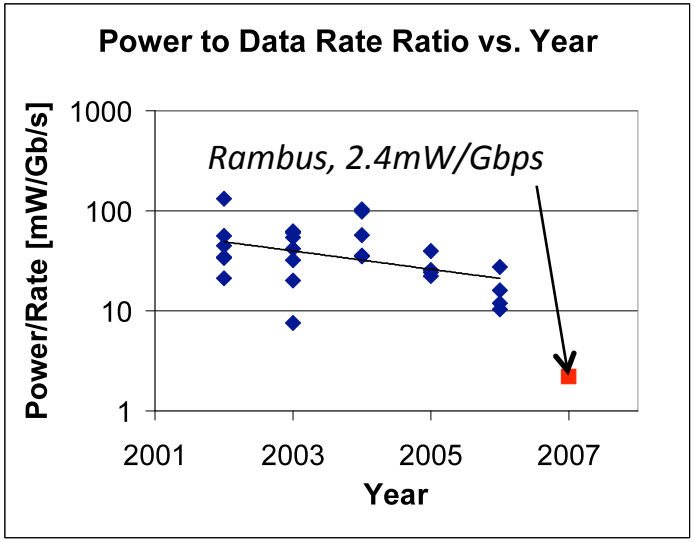
**OFF-CHIP:**
**1-10pJ/bit (1-10mW / Gbps)**

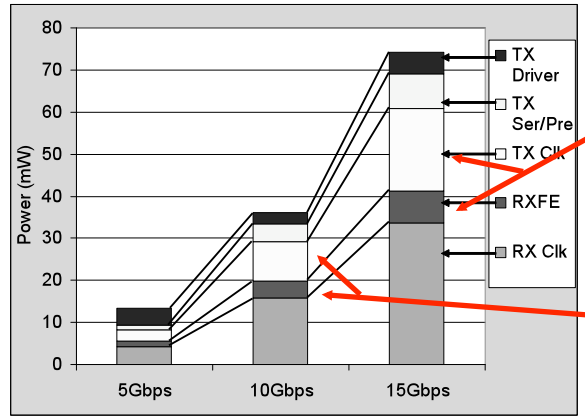
**OUR GOAL: < 0.1pJ/bit**

Source:
Exascale Roadmap
Meeting, Dec. 2009

# Parallelism: Reducing Clock Power Through Global Clock Distribution Optimization (with Intel)

- Goal: Reduce power < 1mW/Gbs



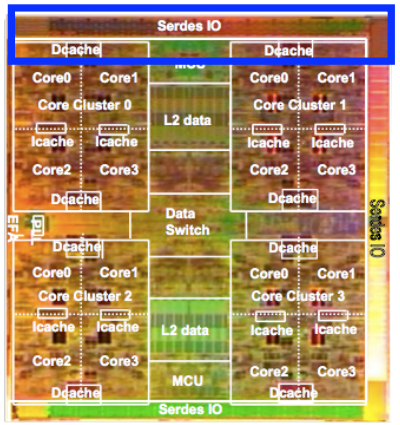

Intel (VLSI, 2007)

Clk/CDR is 71% of Total Power (15Gbs)

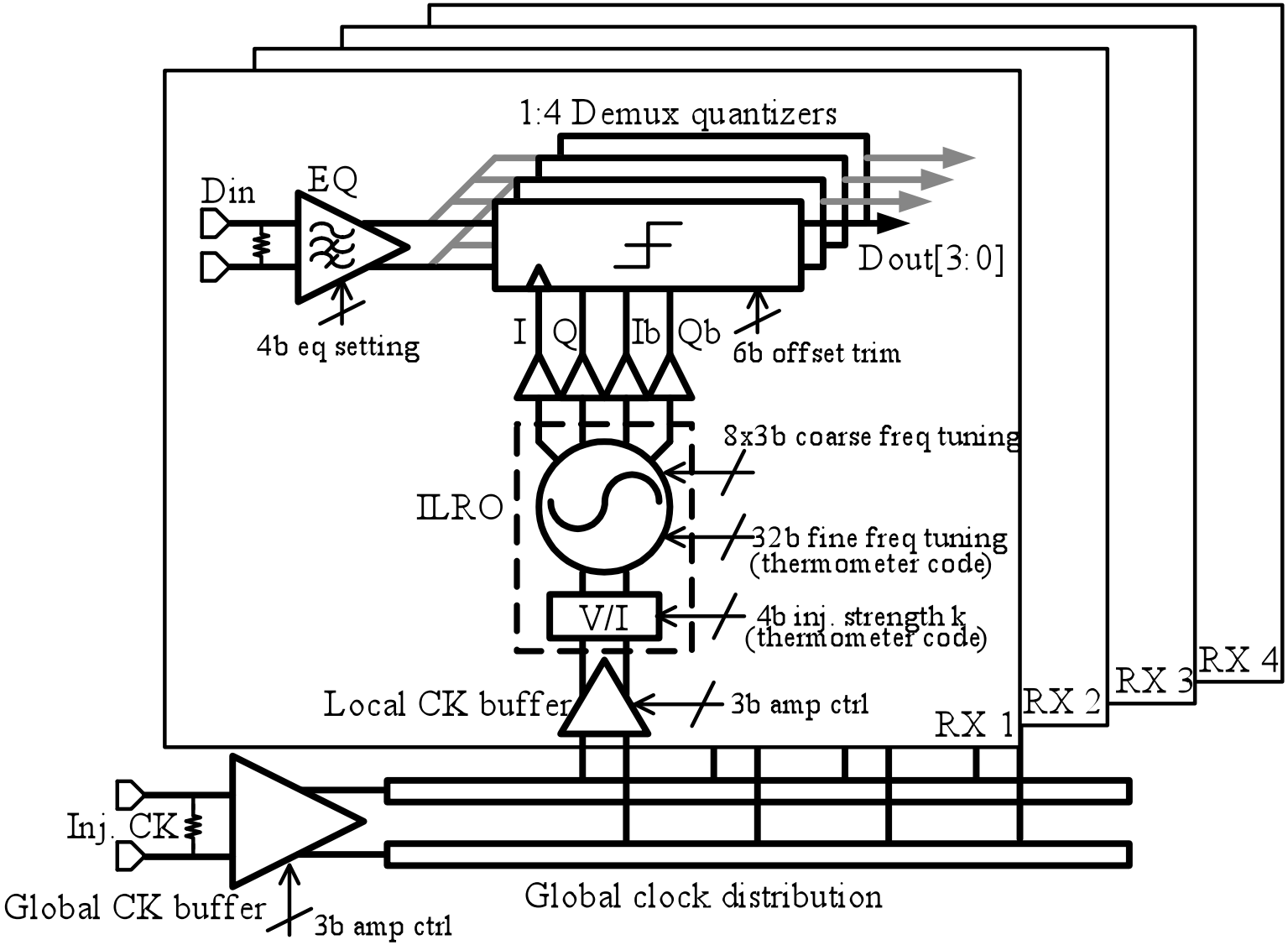
Clk/CDR is 54% of Total Power (10Gbs)

- Assume channel loss is **moderate**

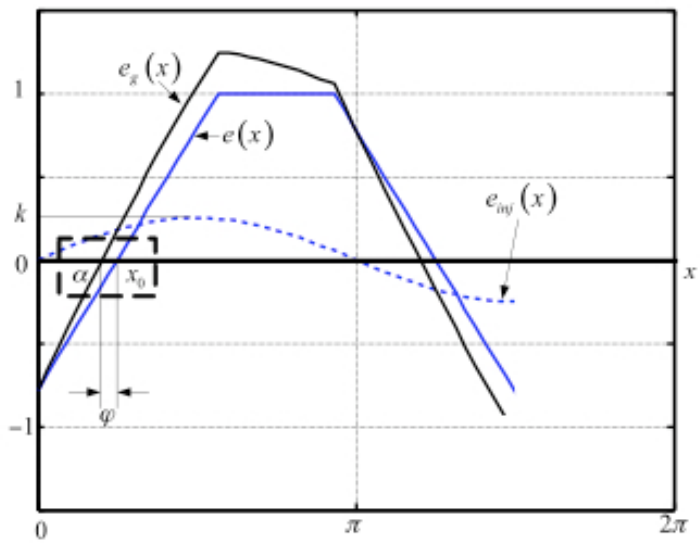
- Clock dominates power in serial links
- Is there a way to share clock power?



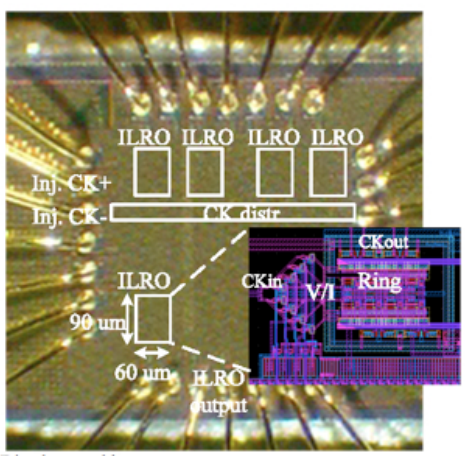
# Proposed Receiver Architecture

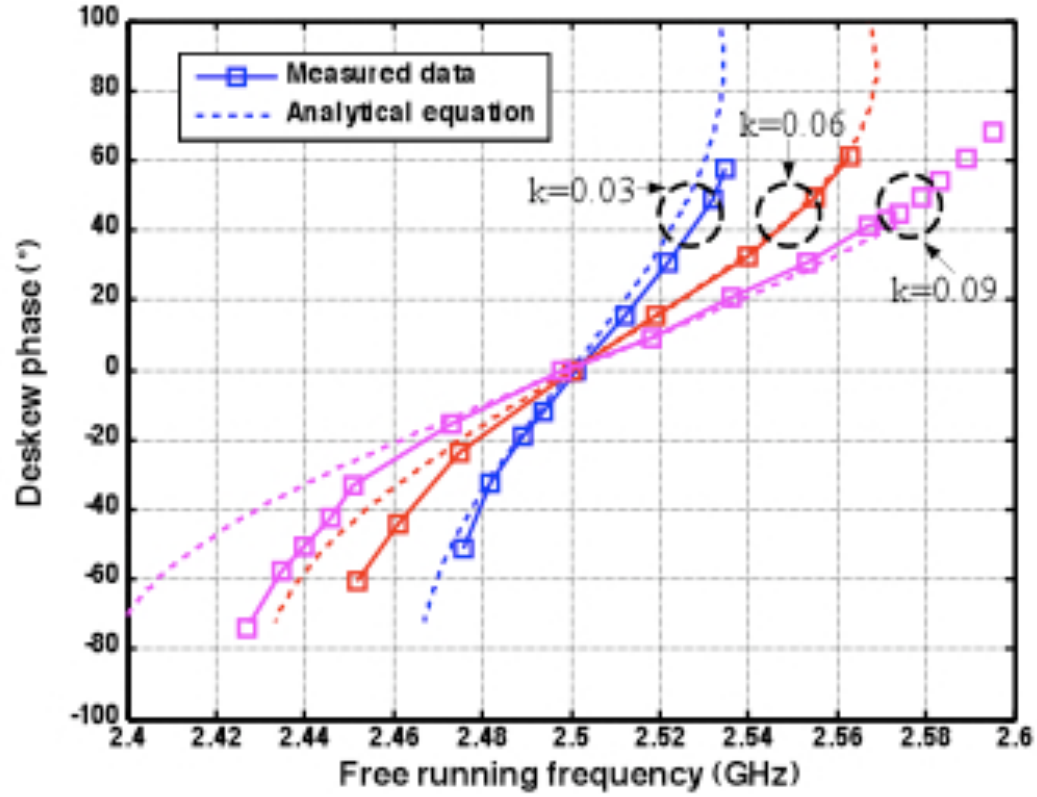


# ILRO: Extension of Adler's Equation



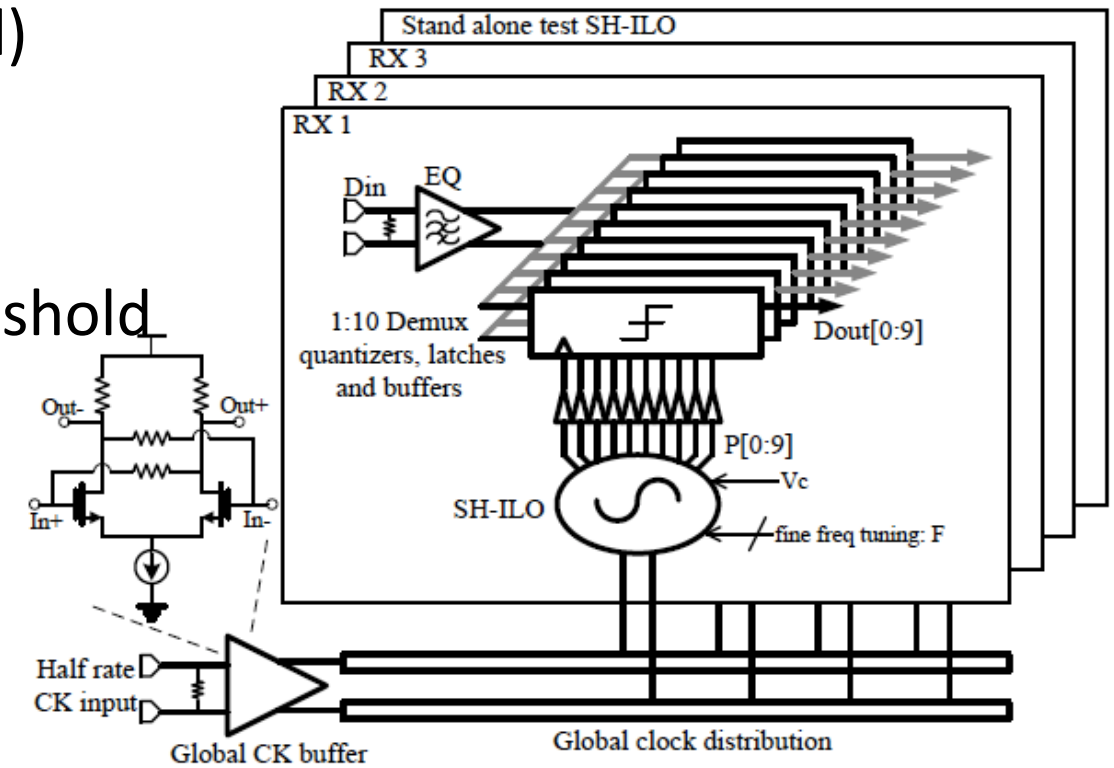

$$\omega_{SL} = \frac{k}{N\eta/\pi - k\cos\alpha}\,\frac{2\omega_0}{N\sin(2\pi/N)}$$



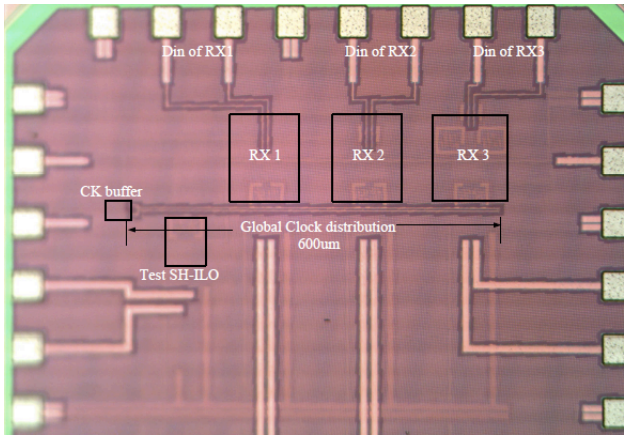



# Further: Near-Threshold, 0.13mW/Gbps, 8Gbps Serial Link Receiver

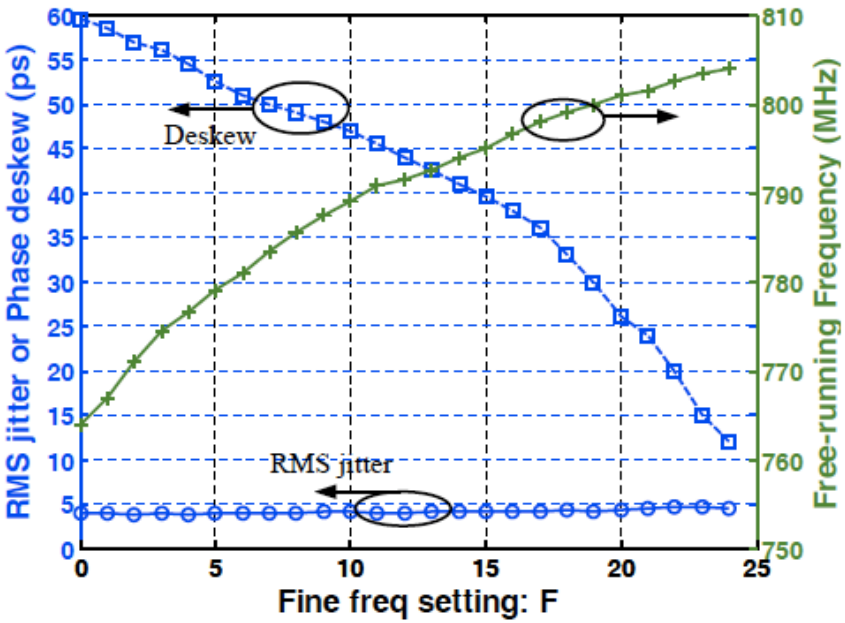
- Two innovations:
  - Exploits constrained deskew (Jaussi, Casper @ Intel)
  - "Kitchen Sink"
  - Operates in Near-Threshold (Vdd < 0.6V)

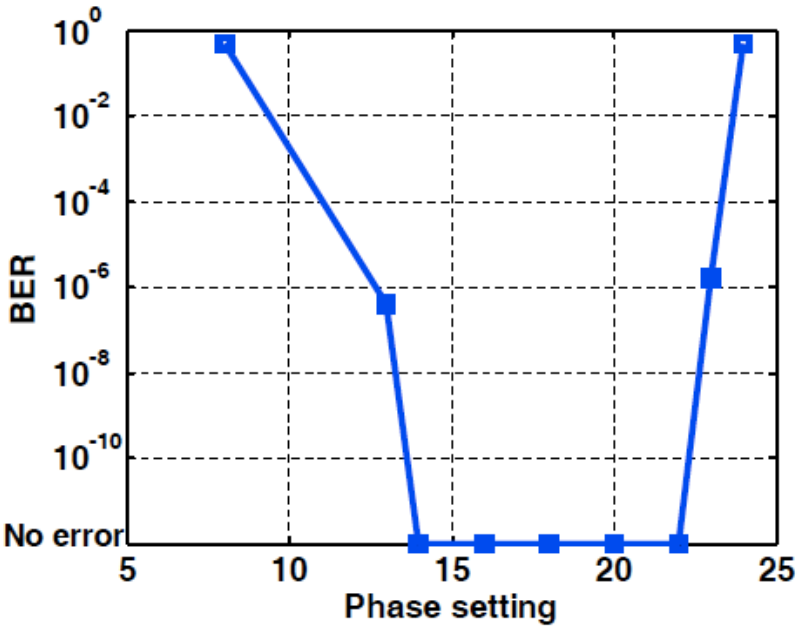


# Measured Deskew Range and BER




- Deskew range: 0.37UI (max)



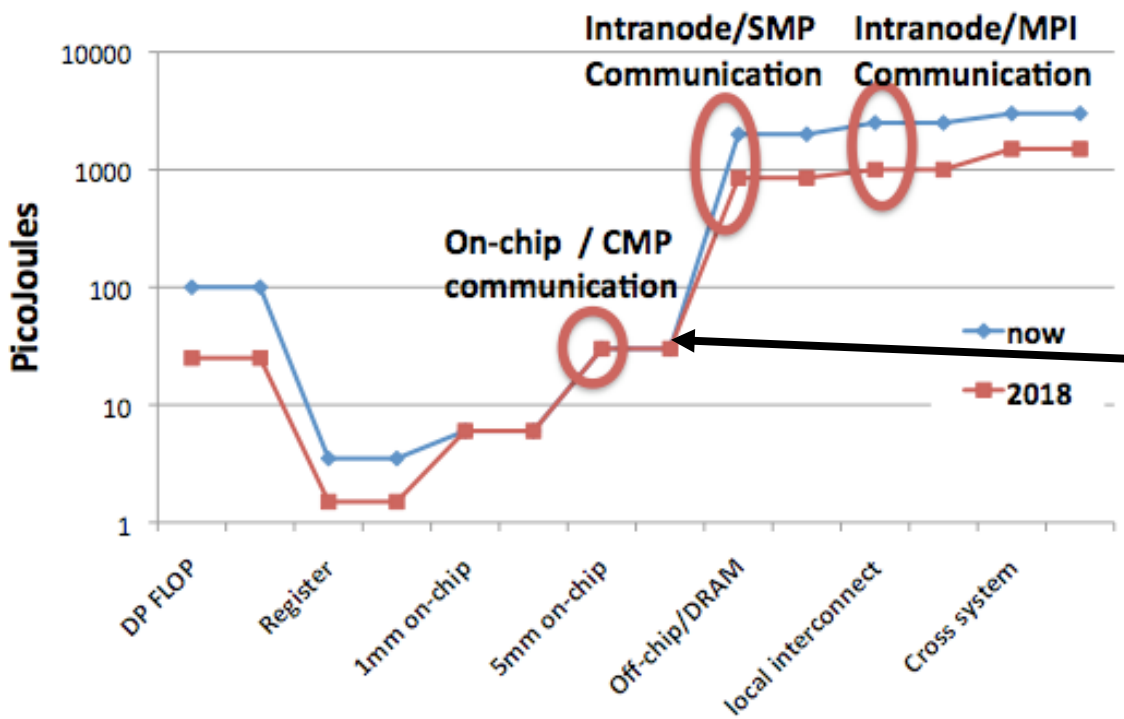



# Comparison Table

| | [6] | [7] | [8] | [9] | This work |
|---|---|---|---|---|---|
| Data rate | 6.25Gb/s | 27Gb/s | 7.2Gb/s | 7.4Gb/s | 8Gb/s |
| Architecture | Software CDR | Forwarded CK | Forwarded CK | Forwarded CK | Forwarded CK |
| Phase tuning method | PLL with PI | LC-ILO | Ring-ILO | ILO | SH-ILO |
| Deskew range | Full UI | Full UI | Full UI | Full UI | 0.37 UI |
| Supply voltage | 1V | 1.1V | 1V | 1V | 0.6V / 1V (global ck) |
| RX power | 8.22mW | 43mW | 4.3mW | 6.8mW | 1.04mW |
| Power efficiency | 1.31 mW/(Gb/s) | 1.6 mW/(Gb/s) | 0.6 mW/(Gb/s) | 0.92 mW/(Gb/s) | 0.13 mW/(Gb/s) |
| RX area | $0.153mm^2$ | $0.015mm^2$ | $0.017mm^2$ | $0.03mm^2$ | $0.014mm^2$ |
| Technology | 90nm | 45nm | 90nm | 65nm | 65nm |

# Problem 2: On-Chip Wires Consume Energy

- On-chip wire power does not scale
  - Dominated by interconnect capacitance ($CV_{DD}^2$)



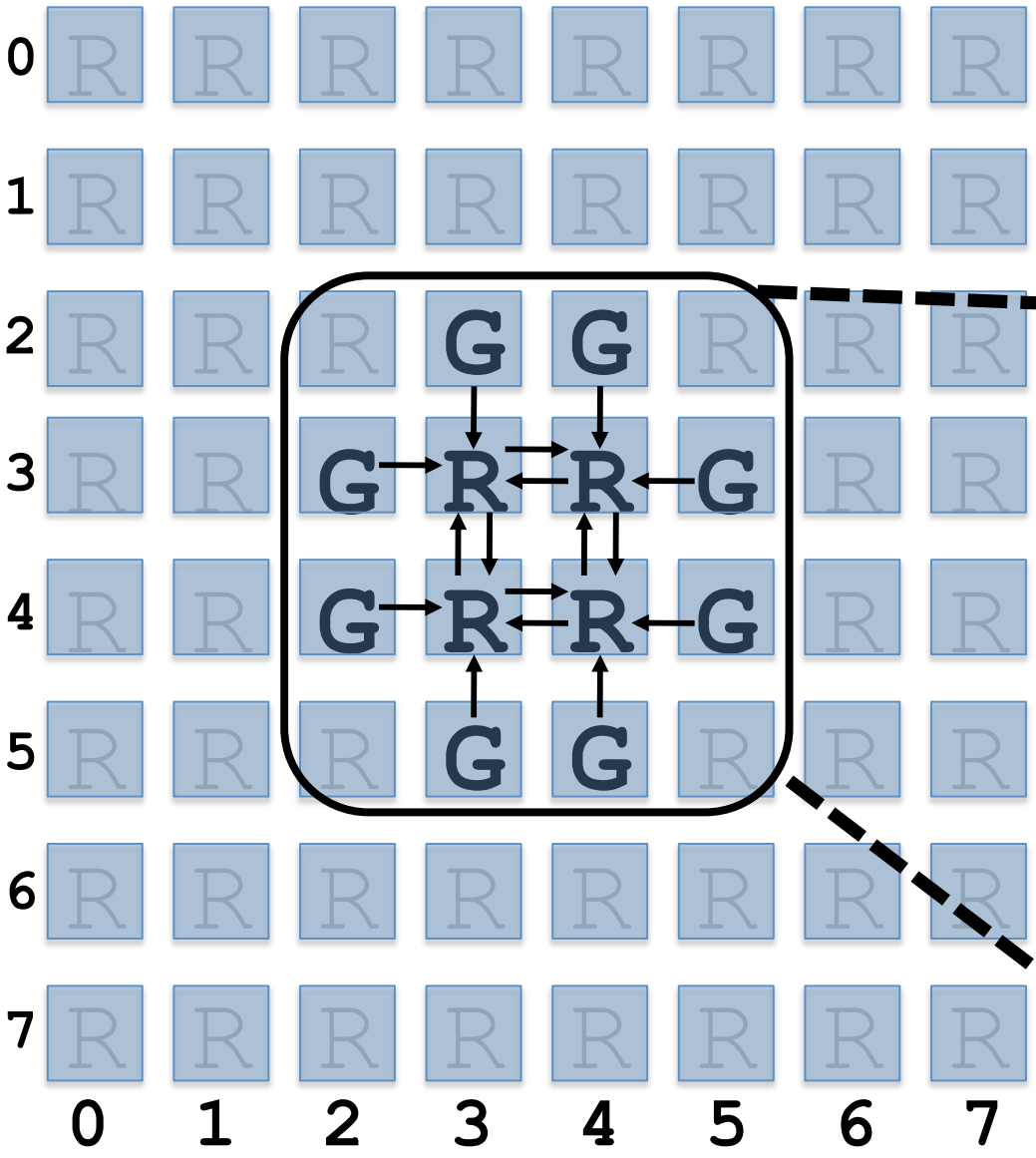

**ON-CHIP (Status Quo): 100 - 300fJ/bit/mm**

**OUR GOAL: < 5fJ/bit/mm**

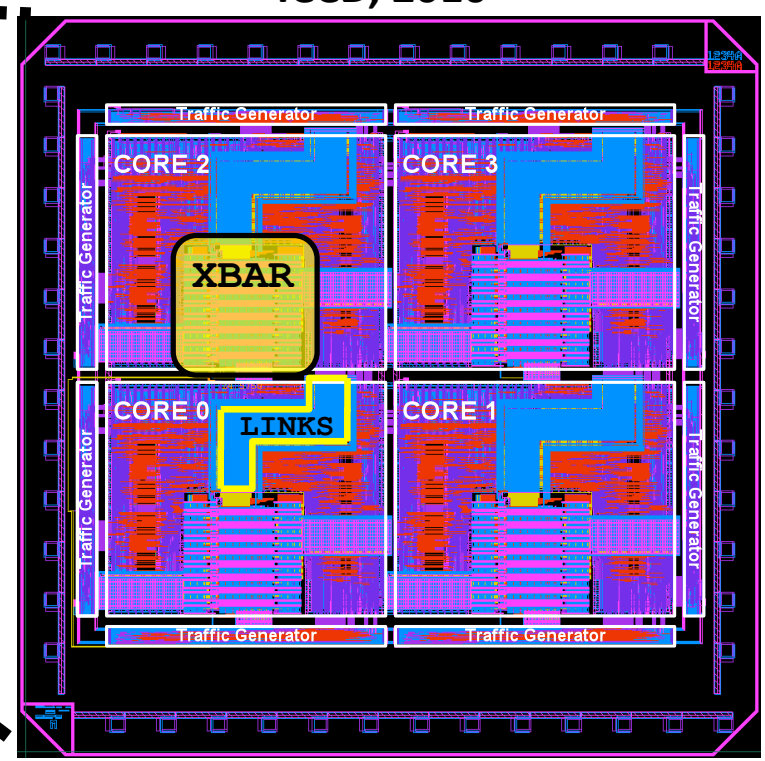
**NOTE: Sub/Near-Threshold doesn't help this problem!**

[DOE, Exascale Workshop]

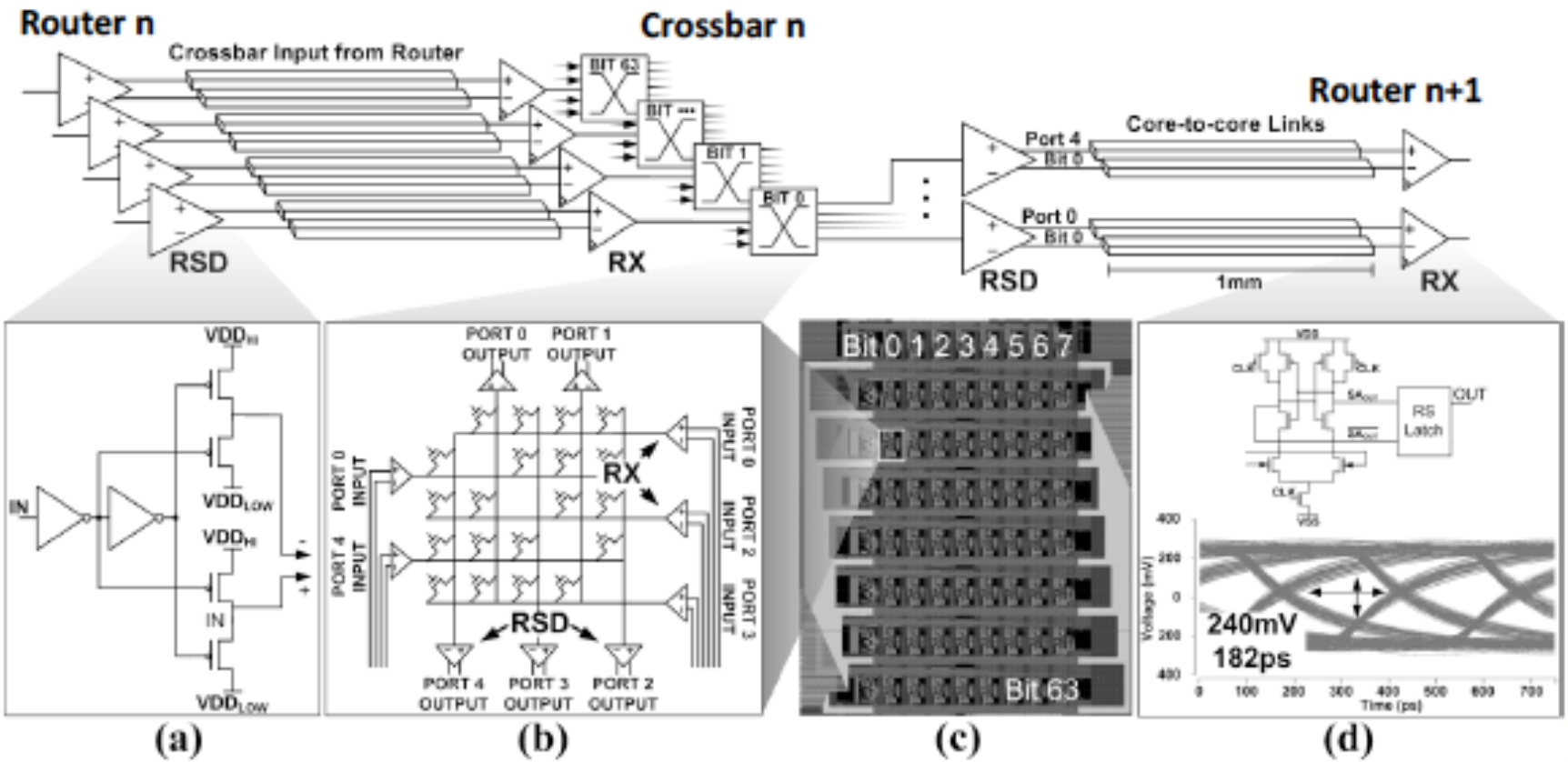
# Energy-Efficient On-Chip Links




**4-Core Network-on-a-Chip**
**(with Li-Shiuan Peh, MIT)**
**ICCD, 2010**



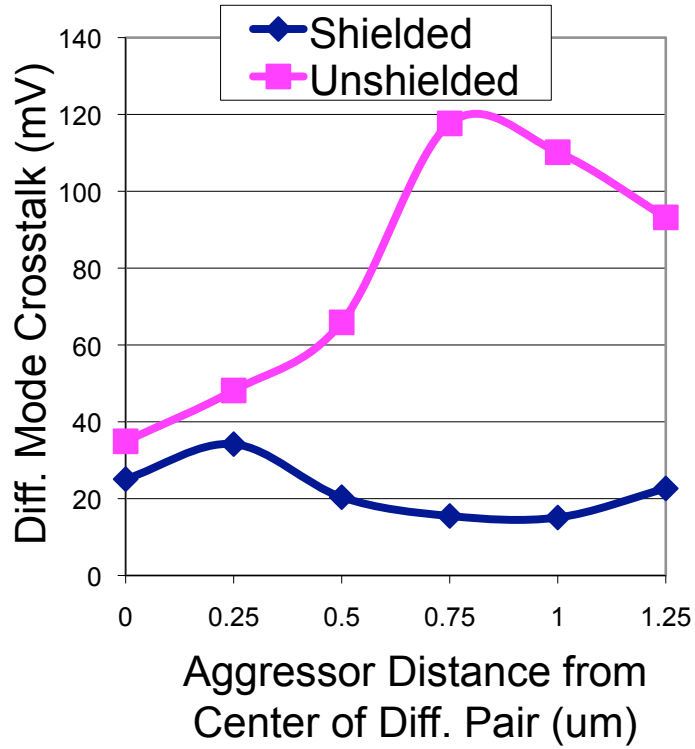
# Low-Swing Bitcell Based Crossbar

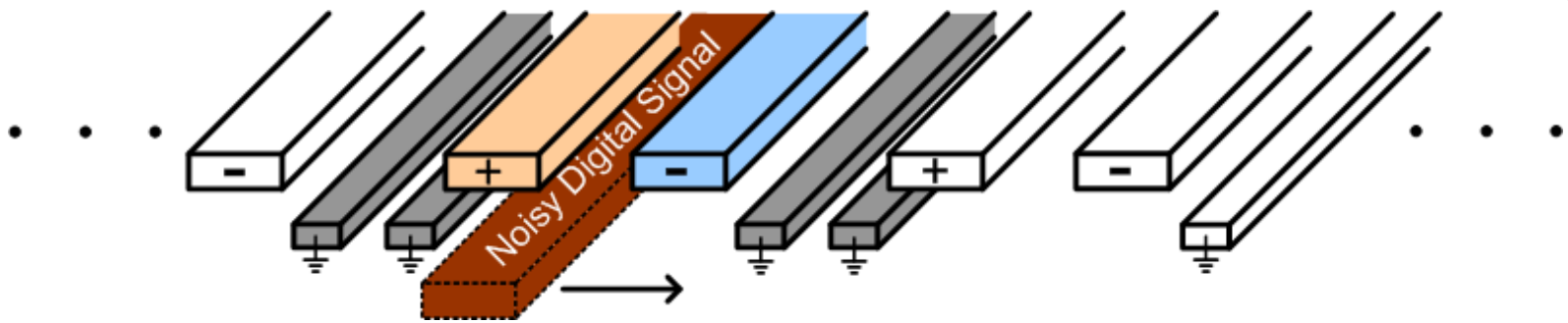



- Two supplies
  - Low-voltage swing = 200mV

# Low Swing Link Over Logic

- 1mm long, 64bit bus
  - Low swing diff. signaling (0.2V-0.4V)
- Routed over noisy digital logic
- Inter-pair differential shielding
  - Reduces differential mode noise
  - Less capacitance than full plane shielding



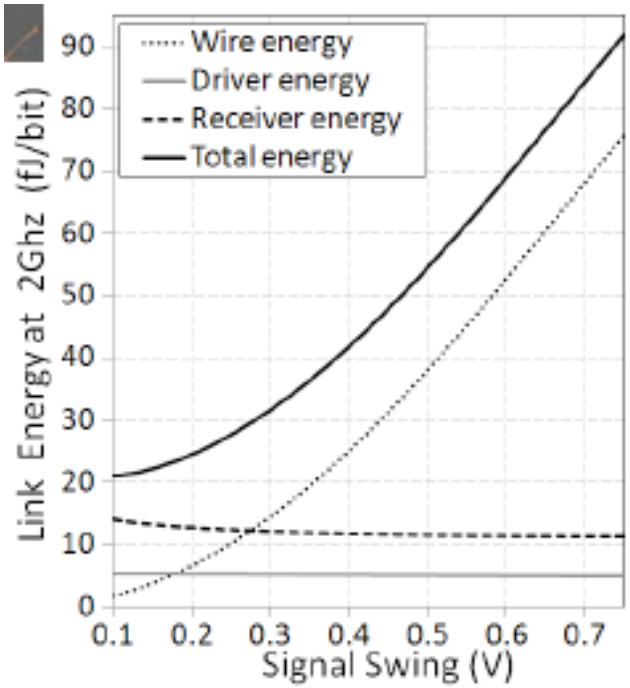




Post-layout Simulation of 1mm Parallel Coupled Digital Aggressor

# Chip Simulations

TABLE I: LINK SUMMARY

| | [4] | [5] | Conventional | This work |
|---|---|---|---|---|
| Technology | 90nm | 90nm | 90nm | 90nm |
| Data rate | 9Gbps | 4Gbps | 1Gbps | 2Gbps |
| Length | 2mm | 10mm | 1mm | 1mm |
| TX Size | C=7.29μm$^2$ $W_P$=1.6μm $W_N$=4μm | 1120μm$^2$ | 6.35μm$^2$ (unrepeated) | **7.8μm$^2$** Wp=2.4μ |
| RX Size | N/A | 1760μm$^2$ | 14.8μm$^2$ | **15.2μm$^2$** |
| Swing | 120mV | ~200mV | 1.2V | 250mV |
| Energy/bit | 105fJ | 356fJ | 305fJ | **28fJ** |



# Performance Summary

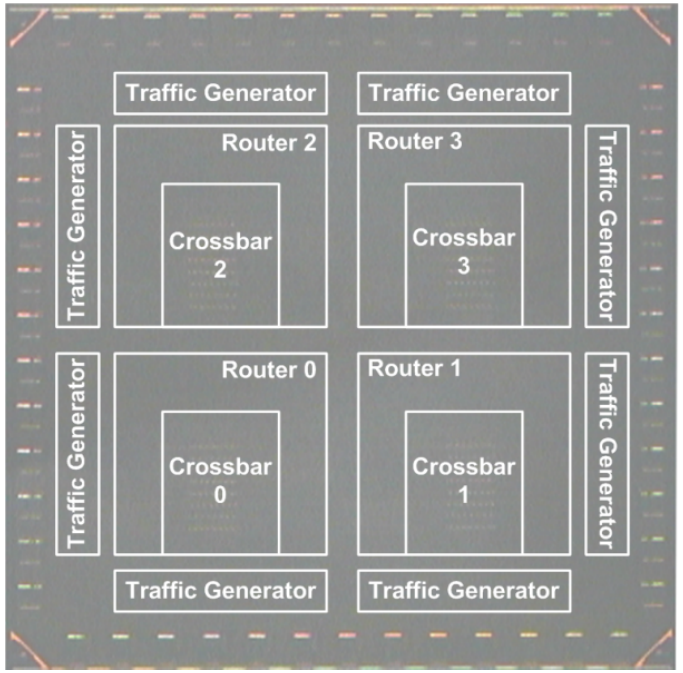
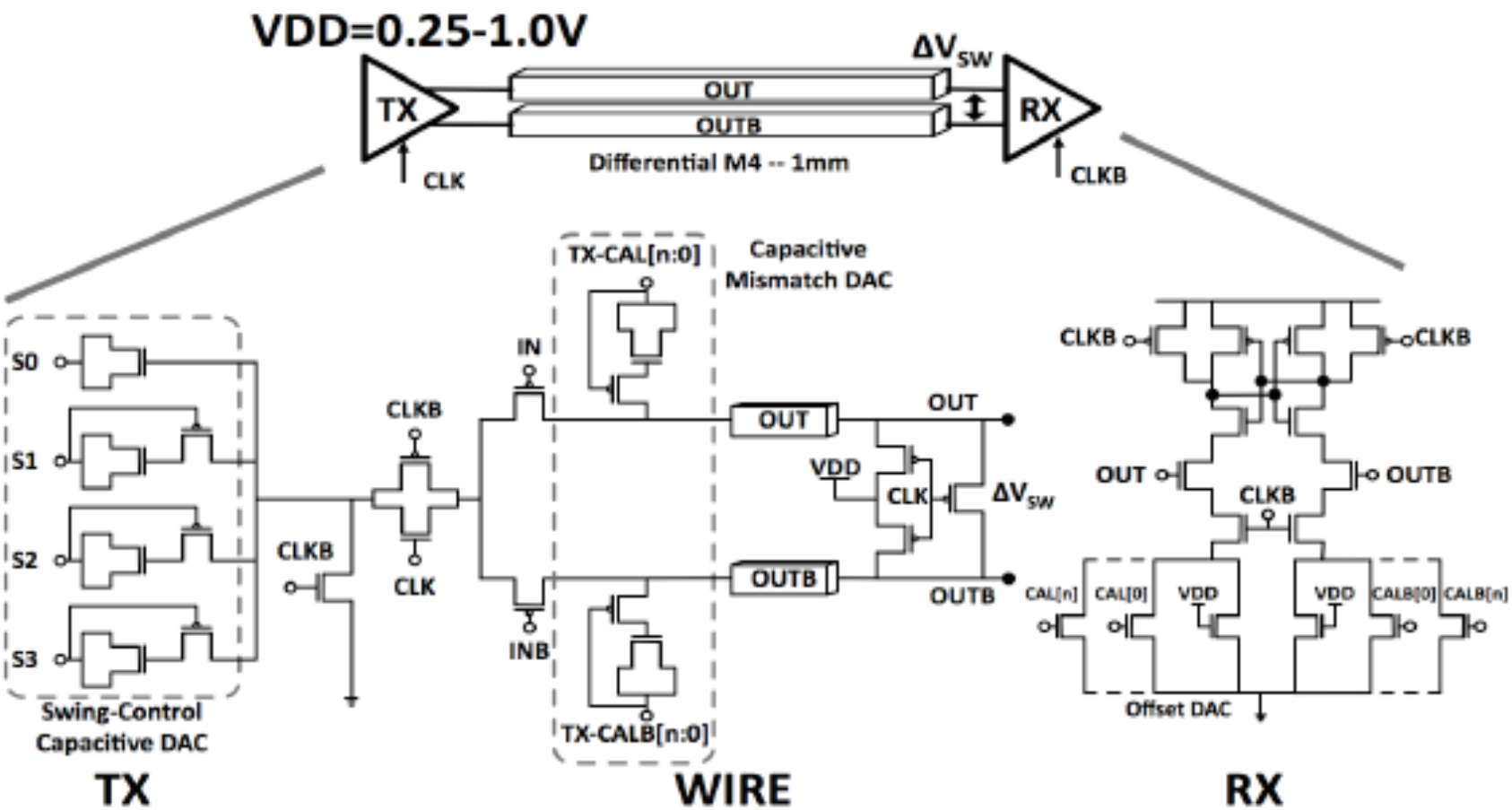



TABLE II: TEST CHIP SUMMARY

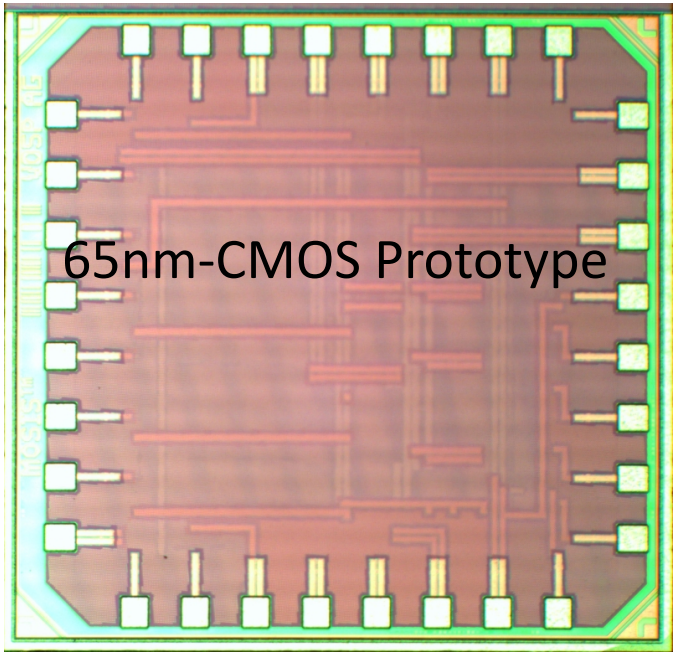
| | |
|---|---|
| Technology | 8-Layer, 90nm CMOS |
| Supply Voltage | 1.2V |
| Chip Size | $4mm^2$ |
| Transistors – Chip | 688k |
| Frequency | 400MHz (low injection)<br>225MHz (high injection) |
| Measured Energy/bit (1mm) (signal + clock) | 64fJ/bit |
| Network Latency Reduction* | 39% |
| Power Improvement* | 38% Network Total<br>53% Data Path |
| Throughput Improvement* | 15% |

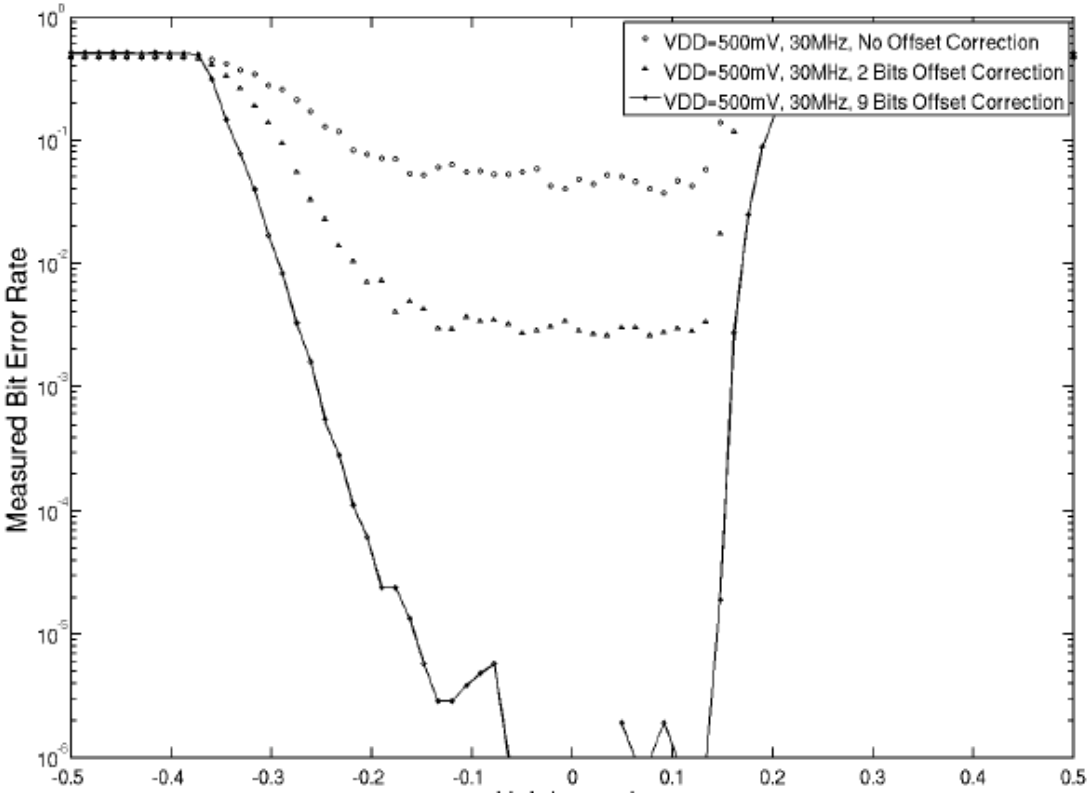
*Relative to a baseline synthesized VC NoC router

# Fundamental Limits to On-Chip Links

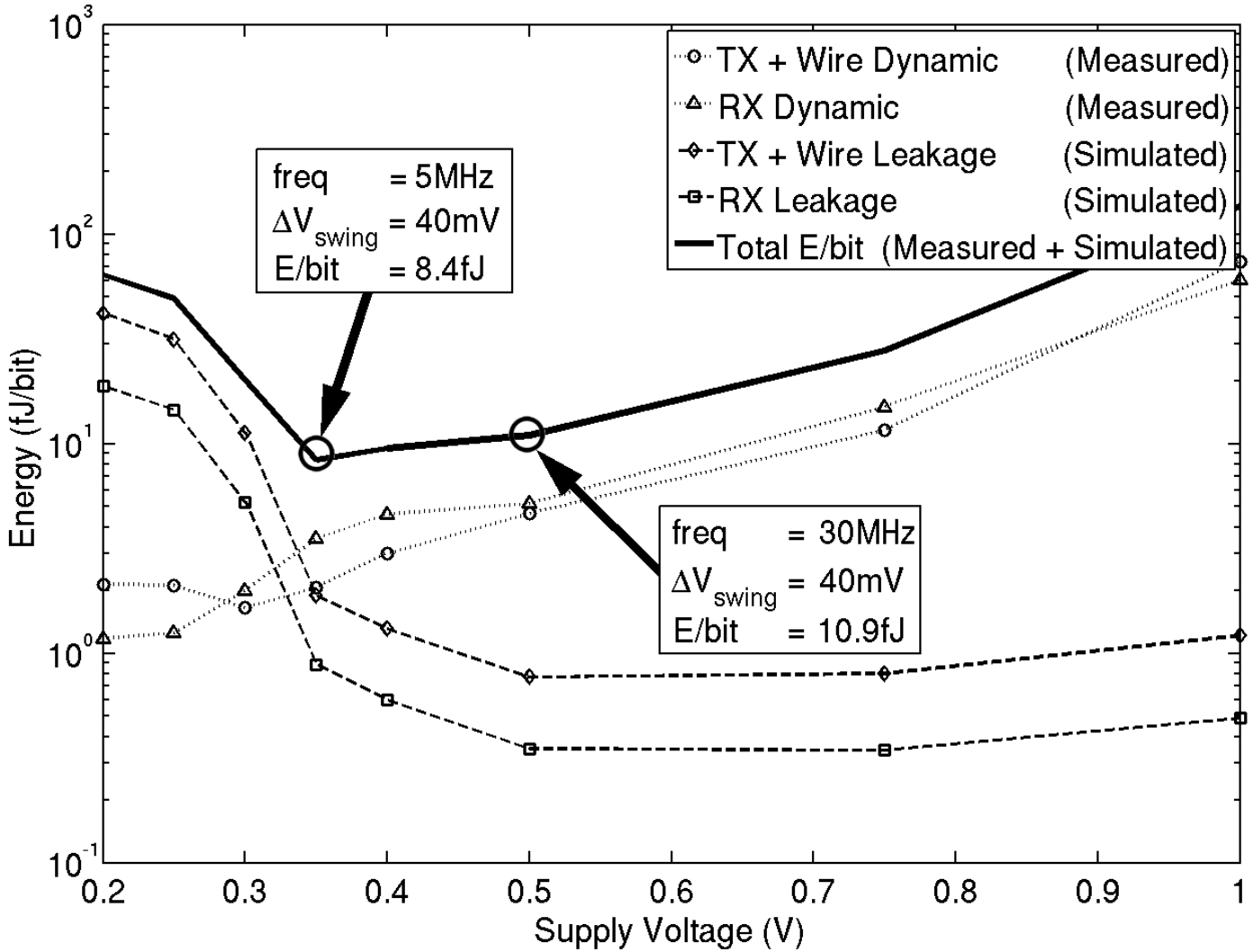


# Bit Error Rate



# Measured Energy/b/mm



# On-Chip Link Comparison

| | Conventional Full Swing | Schinkel (JSSCC '09) | Stojanovic (ISSCC '09) | 1st Chip | 2nd Chip |
|---|---|---|---|---|---|
| Wire Length | 1mm | 2mm | 10mm | 1mm | 1mm-5mm |
| Supply | 1.2V | 1.2V | - | 1.2V | 0.2-1.0V |
| Transceiver Area | 21um$^2$ | TX:20um | 2880um$^2$ | 23um$^2$ | **20-30um$^2$** |
| Signal Swing | 1.2V | 120mV | 200mV | 250mV | 34mV |
| Energy/Bit/mm | 305fJ | 105fJ | 356fJ | 28-60fJ | **8fJ/b/mm** |

- Approaching the fundamental limits to energy-efficient, on-chip links
- Lowest energy/b/mm of on-chip link
  - Energy scalability
  - Low area
  - Robust

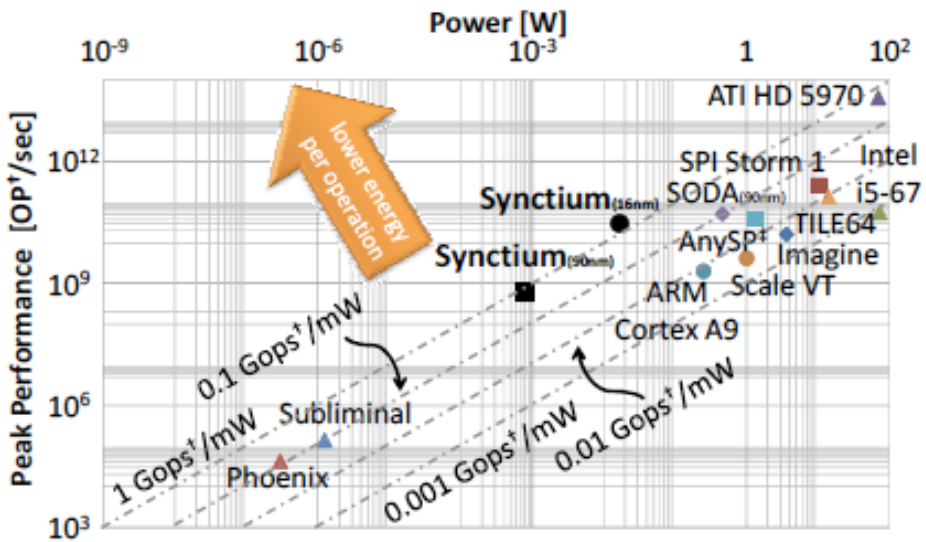
# Co-Design

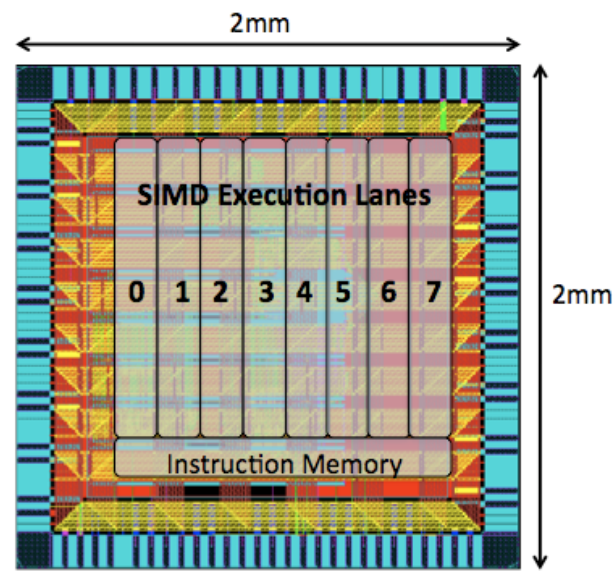
- Take Home: tell vendor to throw kitchen sink at PCB

- Energy can be saved by:
  - OFF: Turning things off when not being used
  - ON: Going parallel (low Vdd)

- Software needs to predict:
  - OFF: Things are not being utilized, turn things OFF
    - Coarse, fine grain clock/power gating
  - ON: Things are utilized, go wide parallel
    - Lower Vdd dynamically, coarse/fine grain DVFS

# Multicore: near-threshold parallel processor – Synctium in 45nm-CMOS



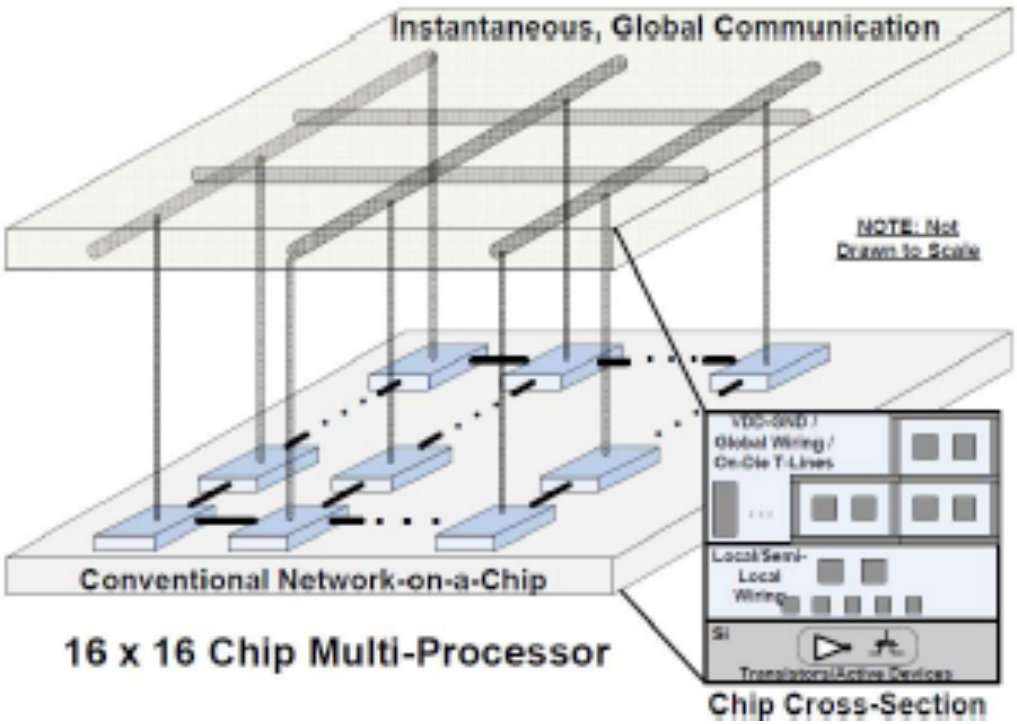




- **Ten parallel lanes, running in near-threshold operation (Vdd=0.4V-1.0)**

- **Incorporates Razor-like detection/recovery in every lane (variation tolerance)**
- **Throughput of SIMD; energy-efficiency of near-threshold operation**
- **GOAL: 1 GOPs / 1mW power; (Eight parallel 16b Multiply/Adds)**

E. Krimer, R. Pawlowski, M. Erez, P. Chiang, "Synctium: a Near-Threshold Stream Processor for Energy-Constrained Parallel Applications", IEEE Computer Architecture Letters, 2010.

# Heterogeneous Interconnect




Micro, 09.