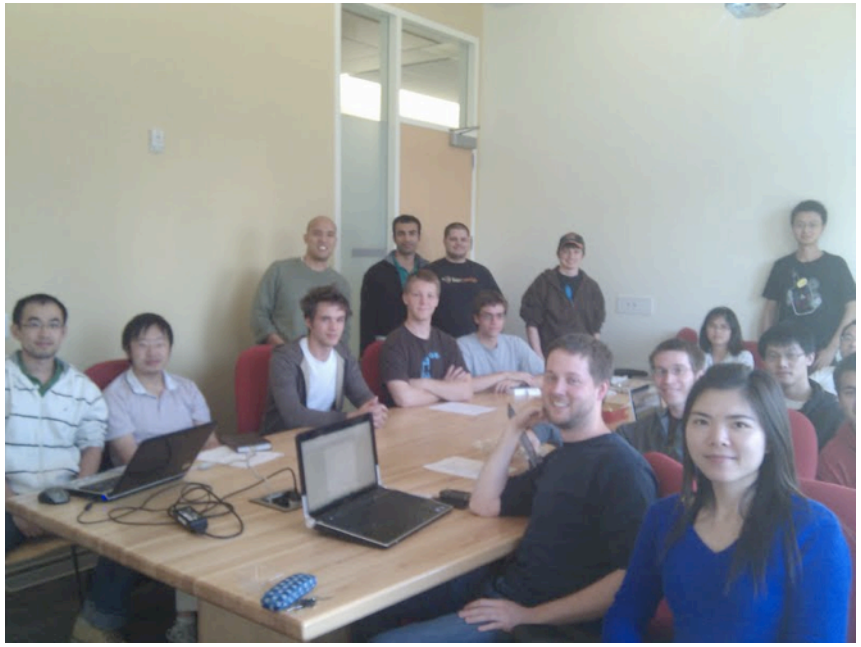
Sustainable Silicon: Energy-Efficient VLSI Interconnect Research



Collaborators: Department of Energy, Intel, Boeing, LSI, SRC, Air Force Research Laboratory, HP Labs, National Science Foundation

Patrick Chiang, Assistant Professor (and Students)
Oregon State University VLSI Research Group
pchiang@ece.oregonstate.edu

DOE, Aug 2011

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

What does exascale mean?

IBM: Roadrunner, Los Alamos (2009)

IBM Roadrunner (12000 PowerX CPUs 6000 AMD Opterons)	2009	2018
Power	2.35MW	10-20MW
Speed (petaflop = 10^{15} FLOP/s)	1.04	1000
Space	296 Racks 6000 sq. ft.	~2x
Memory	103 Terabytes	~1000x
Cost	$125M	????

Not just super-computers

	Attributes				
	Aggregate Compu-tational Rate	**Aggregate Memory Capacity**	**Aggregate Band-width**	**Volume**	**Power**
Exa Scale Data Center Capacity System relative to 2010 Peta Capacity System					
Single Job Speedup	1000X flops	Same	1000X	Same	Same
Job Replication	1000X flops	up to 1000X	1000X	Same	Same
Exa Scale Data Center Capability System relative to 2010 Peta Capability System					
Current in Real-Time	1000X flops, ops	Same	1000X	Same	Same
Scaled Current Apps	up to 1000X flops, ops	up to 1000X	up to 1000X	Same	Same
New Apps	up to 1000X flops, ops, mem accesses	up to 1000X - with more persistence	up to 1000X	Same	Same
Peta Scale Department System relative to 2010 Peta HPC System					
	Same	Same	Same	1/1000	1/1000
Tera Scale Embedded System relative to 2010 Peta HPC System					
	1/1000	1/1000	1/1000	1/1 million	1/1 million

EXASCALE

10^{18} FLOPS

~1MW

EMBEDDED

10^{12} FLOPS

~1W

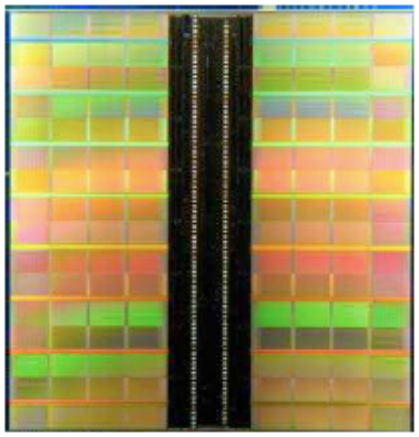
SENSOR

10^{9} OPS

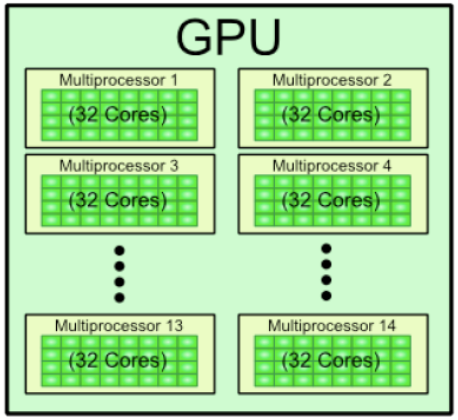
< 1mW

Source: DARPA Exascale Study, 2008

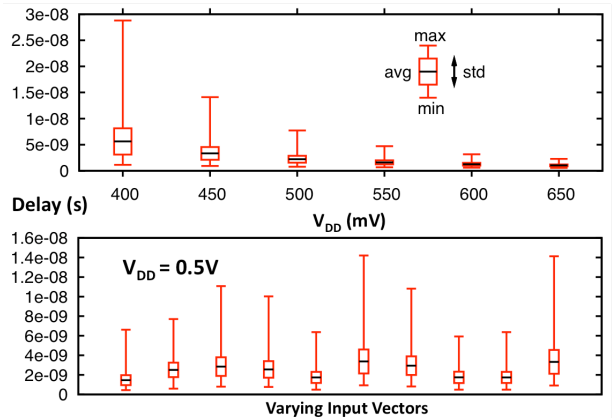
4 Major Challenges



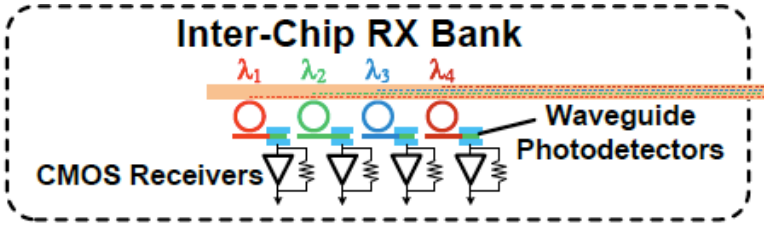
1) Memory and Storage
   - Capacity; Latency



2) Concurrency and Locality
   - f_{CLK} = 1GHz → Parallelism
   - Software / hardware



3) Resiliency Challenge
   - Low VDD
   - Process Variability



4) Energy and Power
   - Interconnect

Interconnect Energy Wall

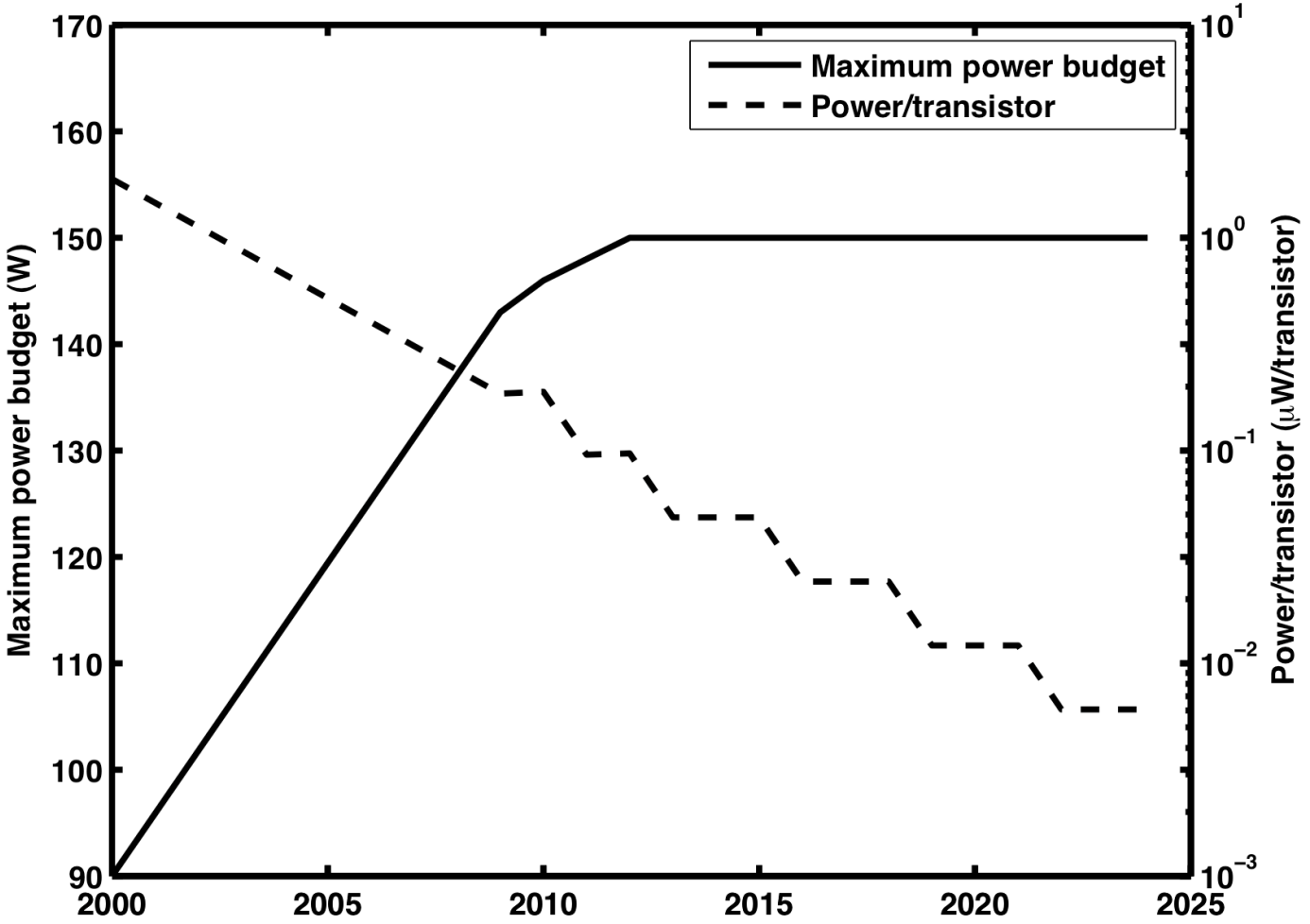
- Microprocessors: “at 0.13um approximately 51% of microprocessor power was consumed by interconnect, with a projection that without changes in design philosophy, in the next five years up to 80% of microprocessor power will be consumed by interconnect”, ITRS-07

- “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…”

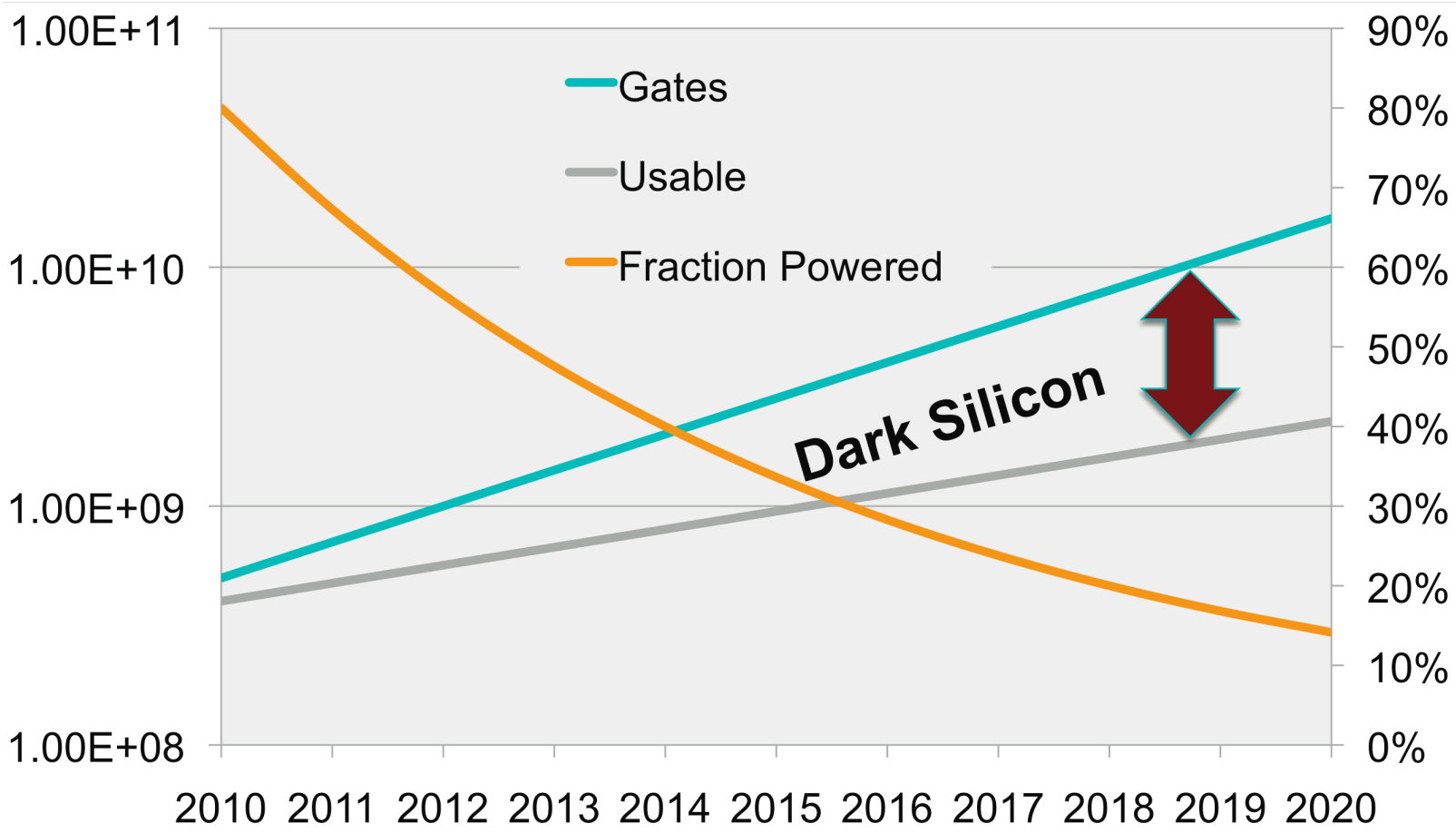
Source: DARPA Exascale Study, Sep. 2008

Power Density



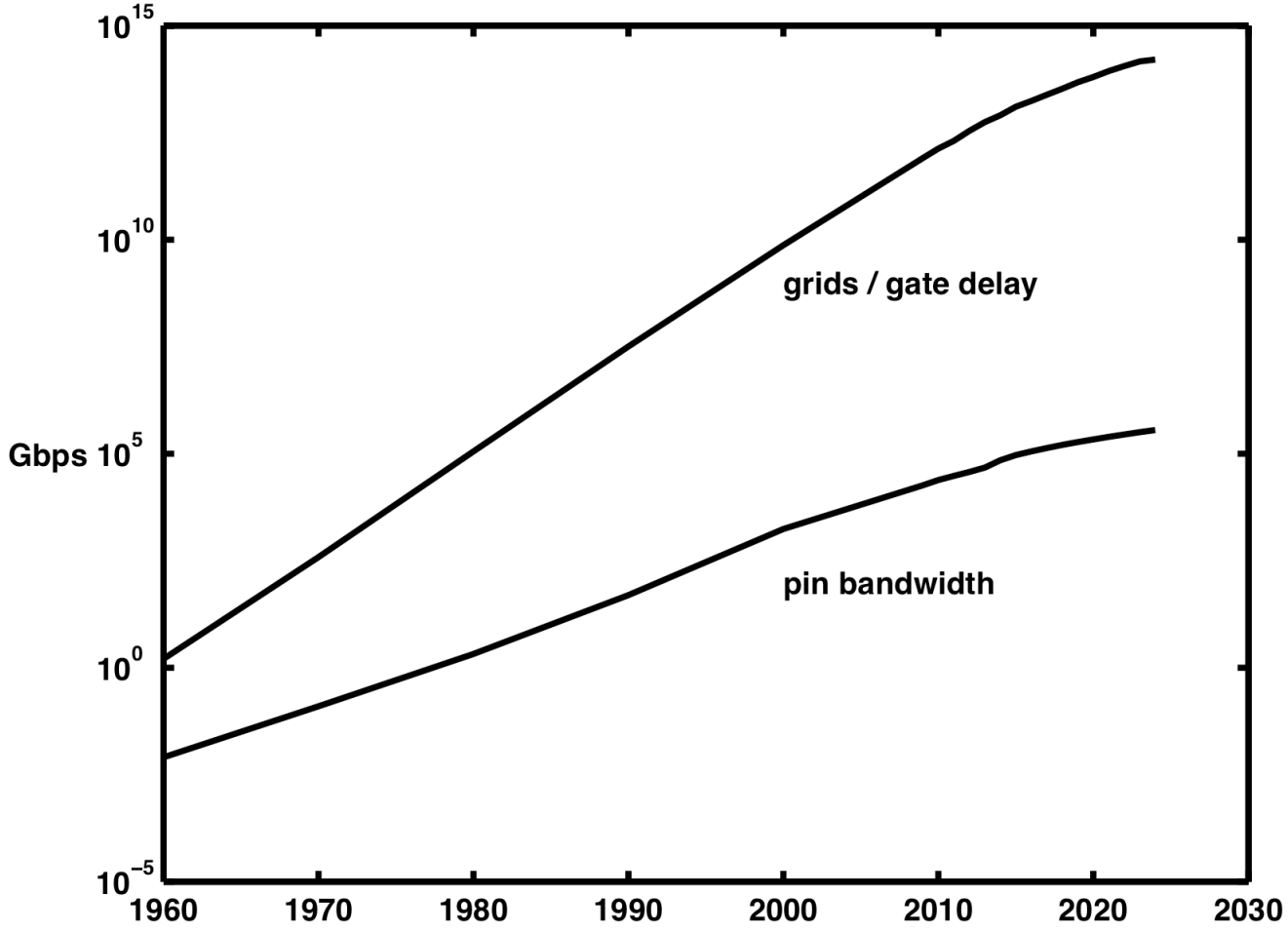
Source: ITRS 2010

Not enough power available

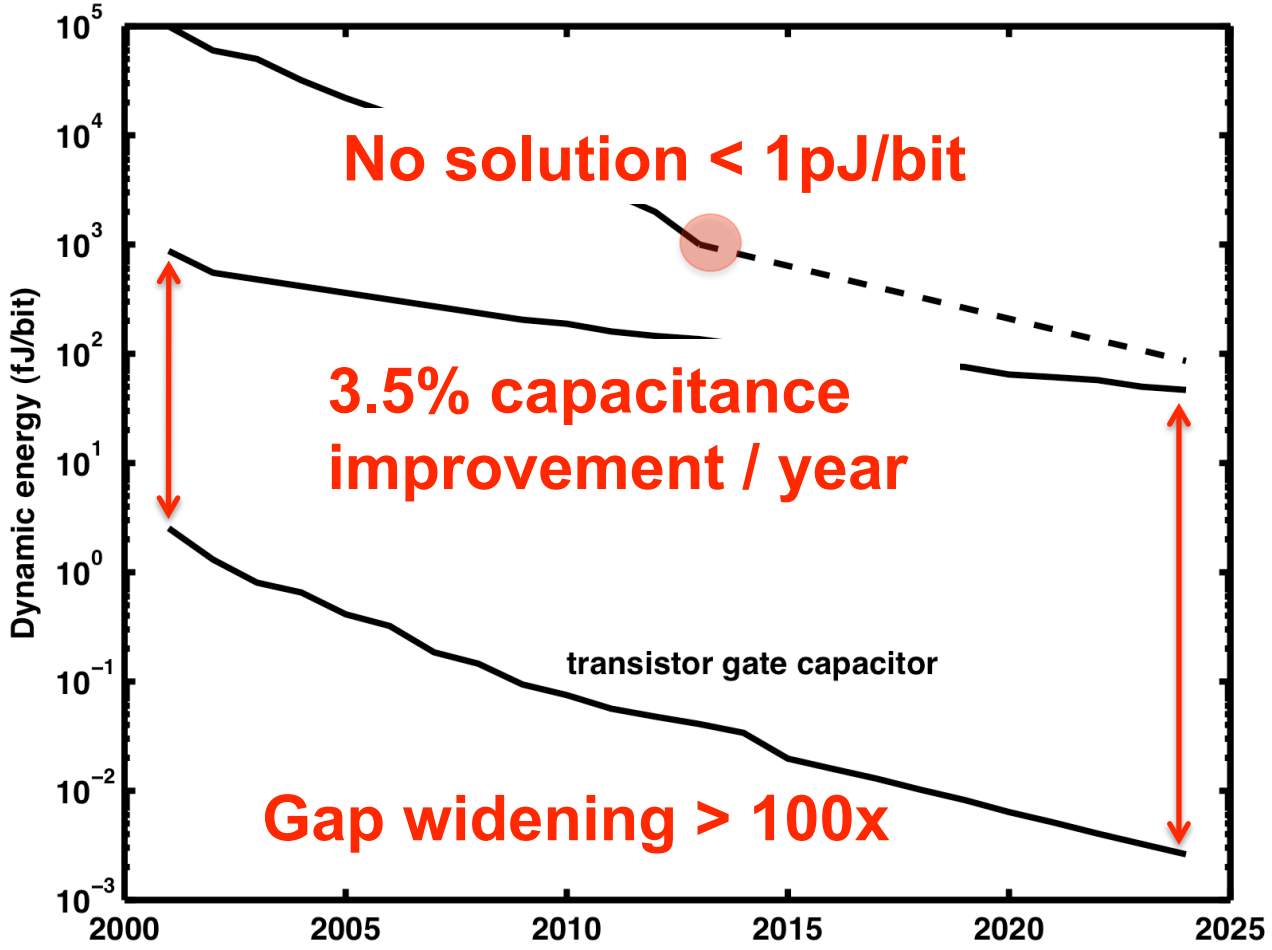


Source: Bill Dally, 2011

Need more bandwidth

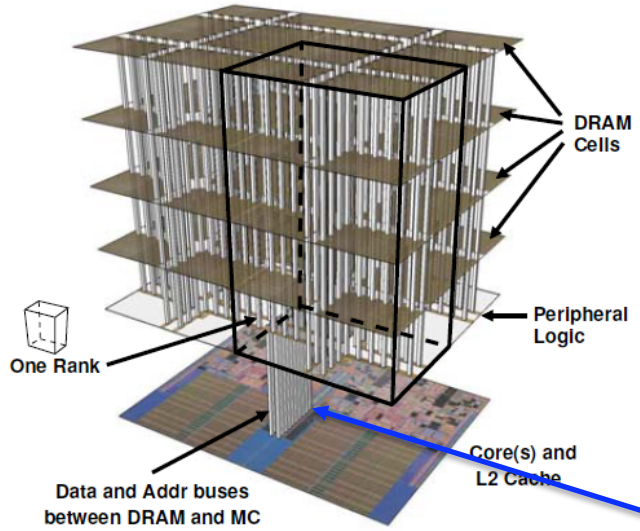


Interconnect Energies

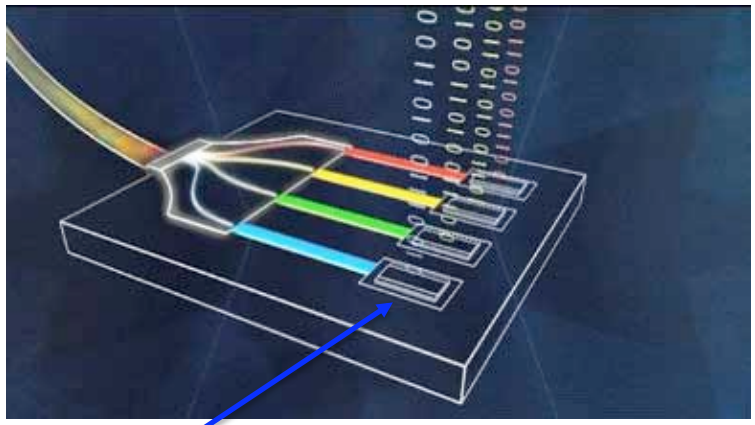


Alternatives

- 3D Stacking



- Silicon Photonics



Electrical transceivers

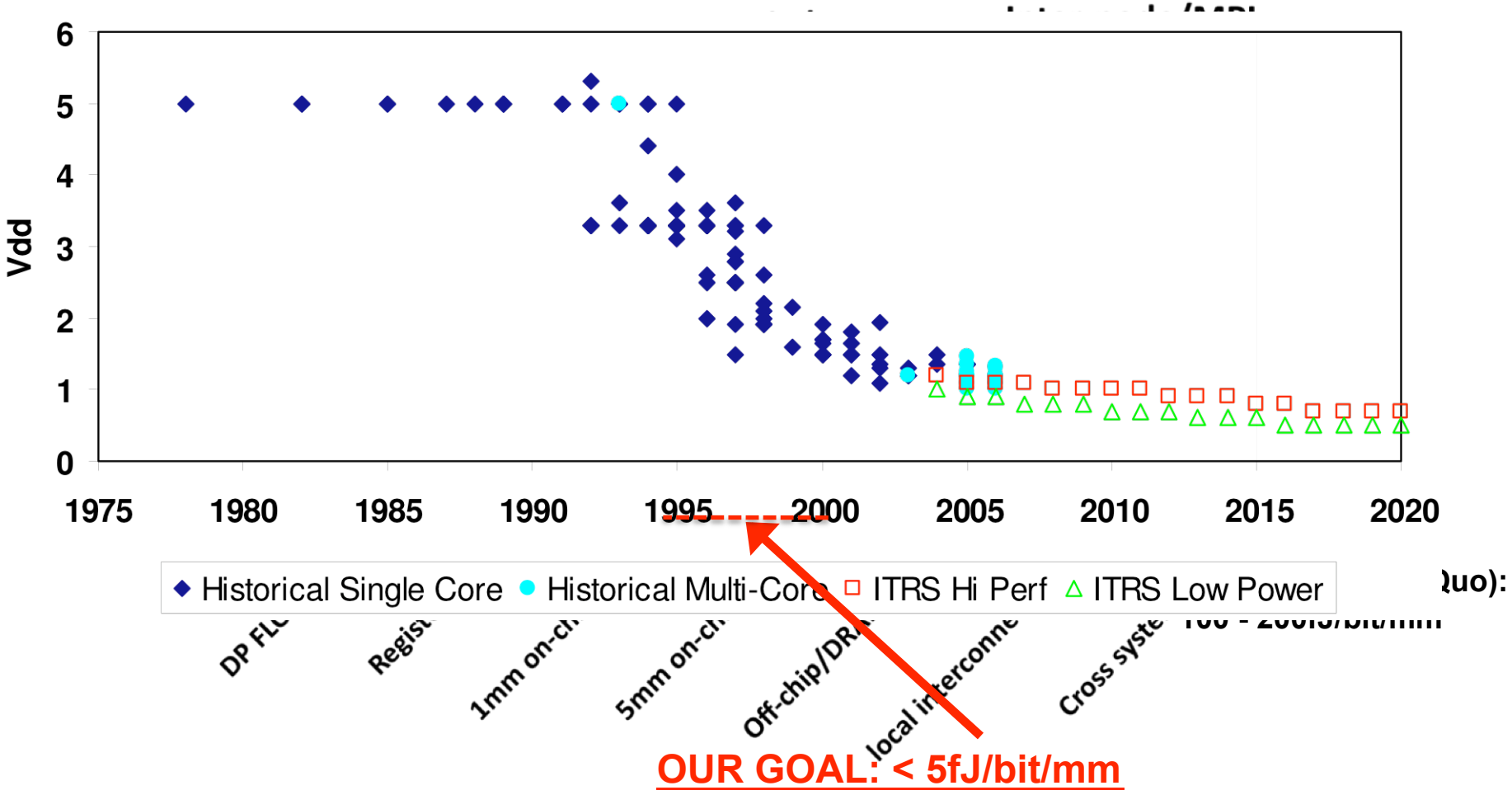
1. Off-chip bandwidth:
   - 3D: dense through-silicon vias
   - Photonics: wavelength division multiplexing

2. Off-chip energy:
   - 3D: make things closer
   - Photonics: use low-loss light

Source: G. Loh, 2008;
Intel 2010

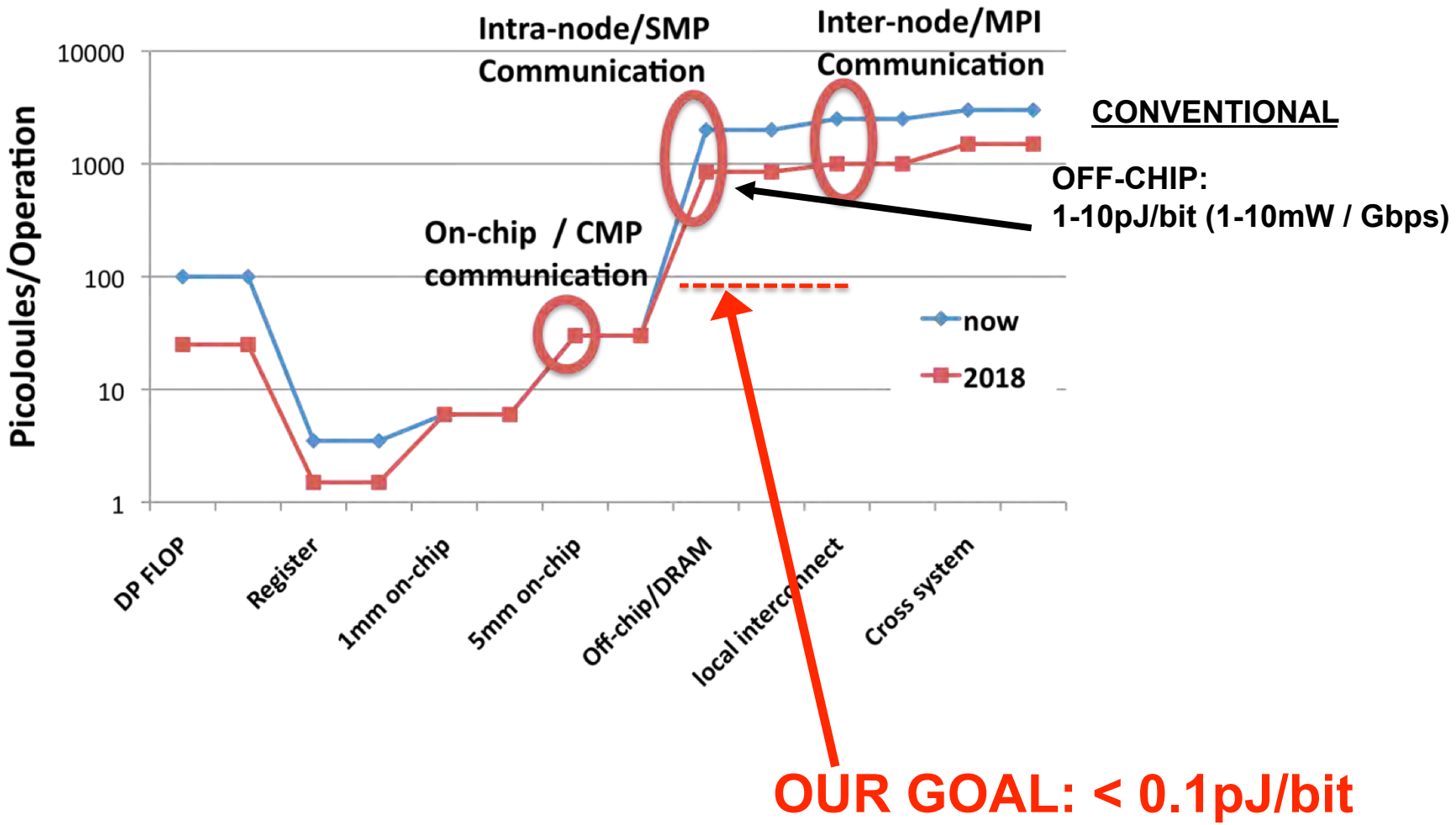
Problem 1: On-Chip Wires

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



[DOE, Exascale Workshop]

Problem 2: Off-chip I/O



Source: Exascale Roadmap Meeting, Dec. 2009

Low-Power Interconnects Overview

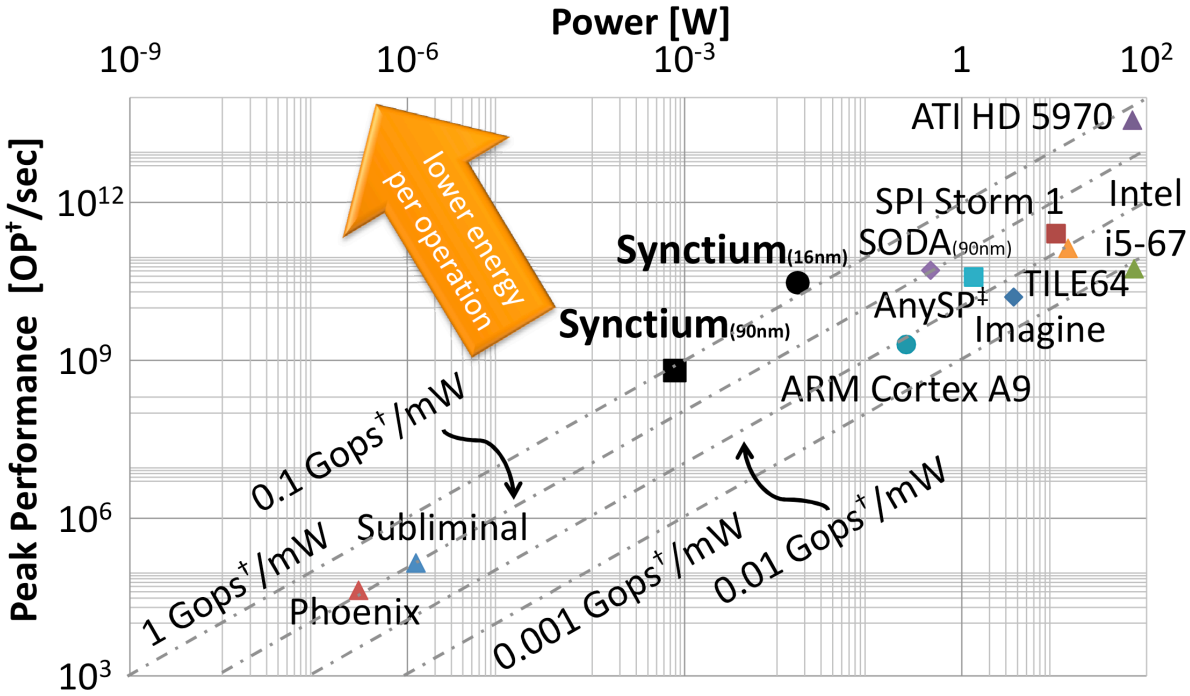
1. On-Chip I/O
   - Network-on-a-chip with reduced-swing interconnect
   - Fundamental limits to low-voltage swing

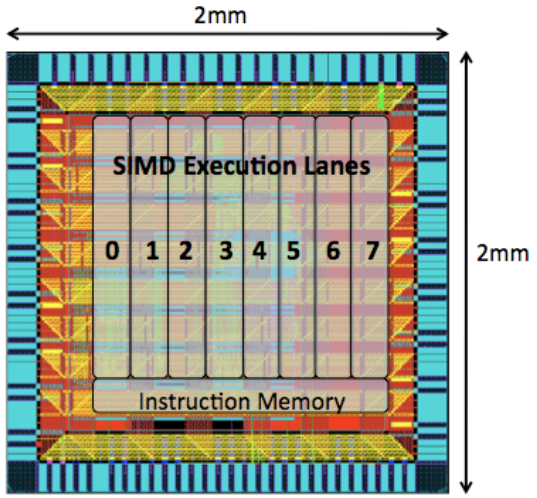
2. Off-Chip I/O
   - Sub-1mW/Gbps off-chip I/O

Near-Threshold Operation (V_{DD} ~ 0.4V)

[S. Hanson, 2006]
MIT, MICHIGAN,
PURDUE, INTEL

Synctium: near-threshold parallel processor

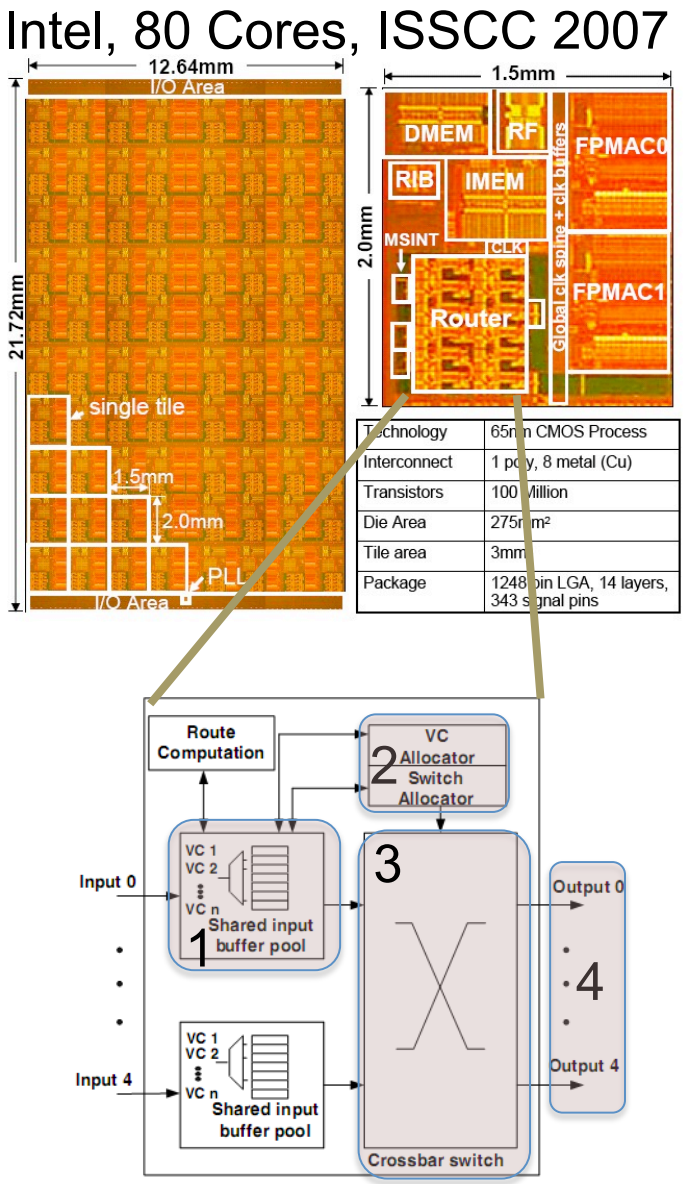




- **Throughput of SIMD; energy-efficiency of near-threshold operation**
- **Eight parallel lanes, at near-threshold (V_{DD}=0.5V)**
- **Variation Tolerance: Razor-like detection/recovery per lane**
  - **Lane weaving**
  - **Decoupled instruction queues**

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

(1) Energy-Efficient, On-Chip Links

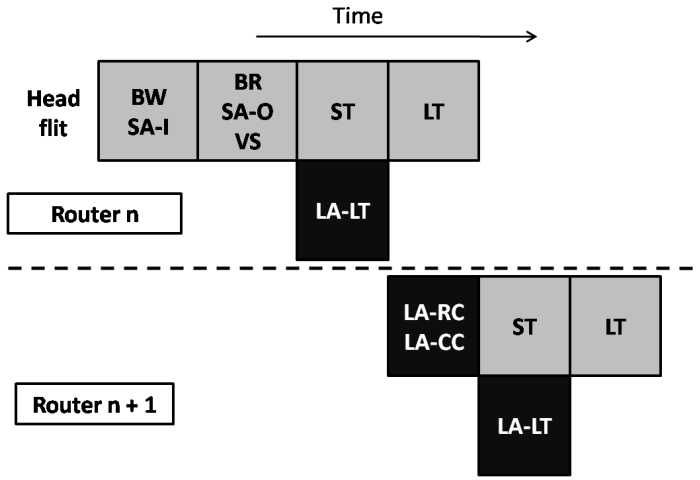


- Router Power:
  - (1) Buffering: 30%
  - (2) Arbitration: 10%
  - (3) XBAR: 30%
  - (4) LINKS: 30%

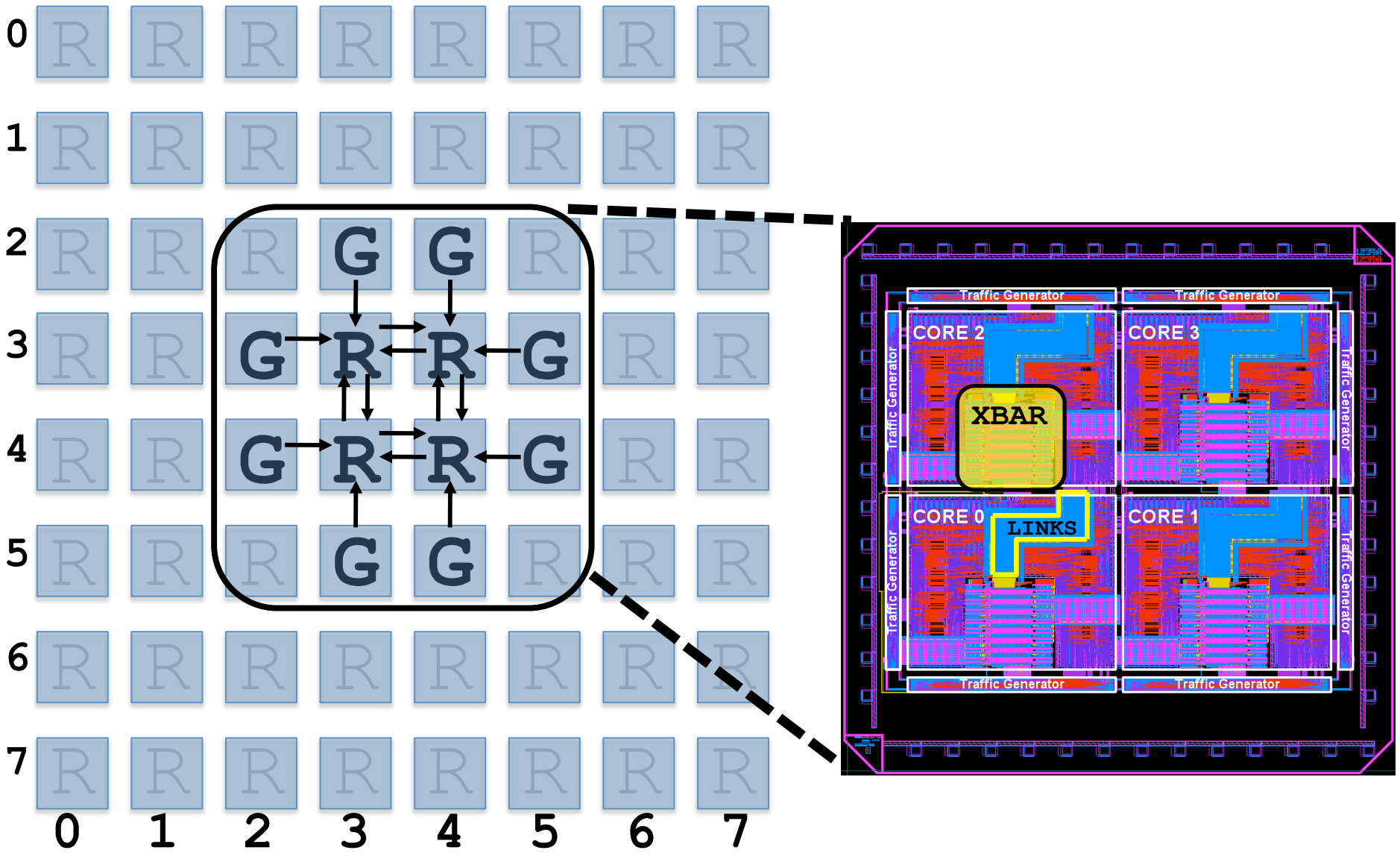
- Our Goal: low-power on-chip links
  - *Analog* low-voltage swing:
    - (3) XBARS
    - (4) LINK TRAVERSAL

Token Flow Control NoC (Li-Shiuan Peh, MIT)

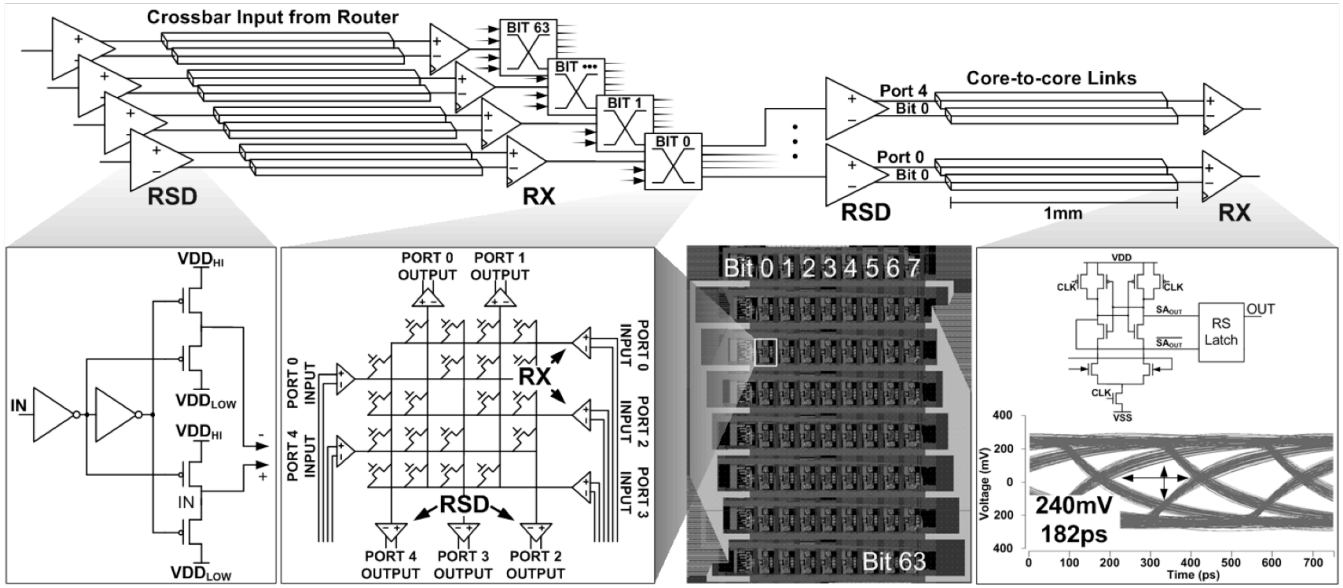
- **Conventional Router:**
  - Each hop requires 4 cycles
- **Proposed TFC Router:**
  - First hop requires 4 cycles
  - Following hops require 2 cycles

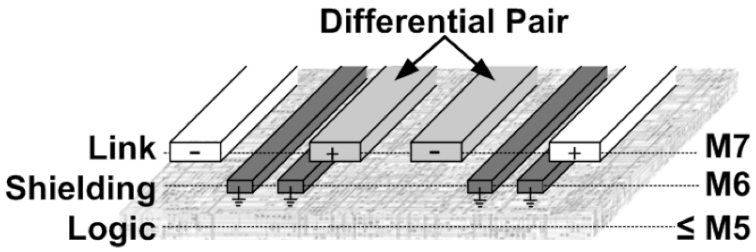


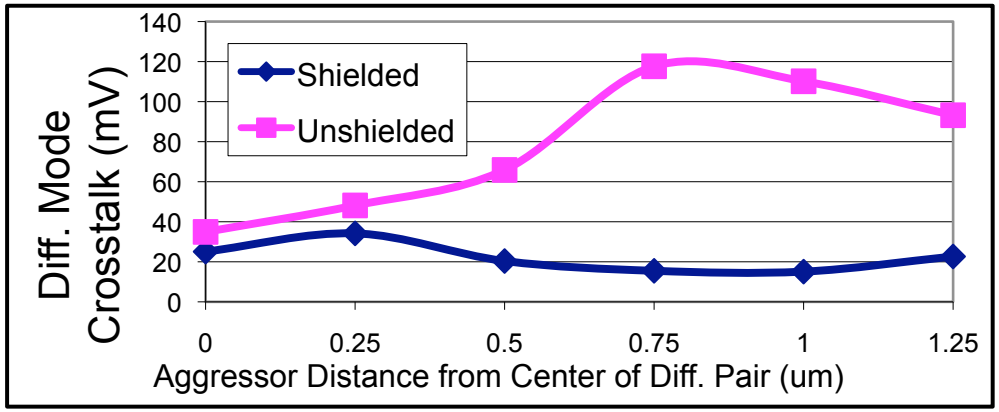
- **Tokens for advance allocation**
  - If little congestion, buffering is skipped
- **NoC power dominated by XBAR and LT**
  - TFC reduces buffer writes


0
1
2
3
4
5
6
7
0 1 2 3 4 5 6 7
R
G
Traffic Generator
CORE 2
CORE 3
CORE 0
CORE 1
XBAR
LINKS

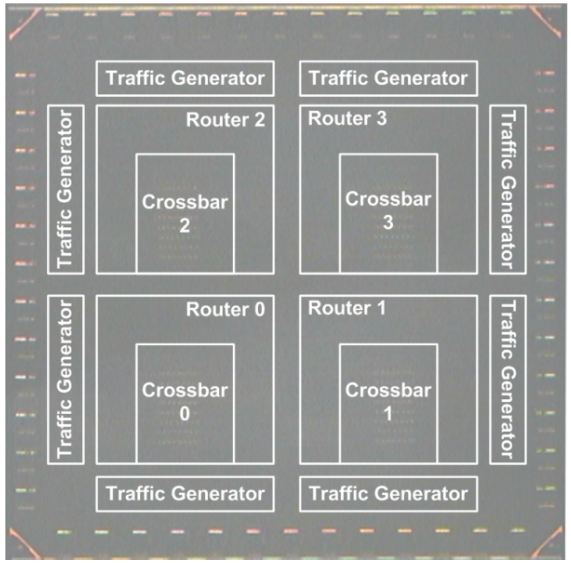
Low-Swing, Bitcell-Based Crossbar

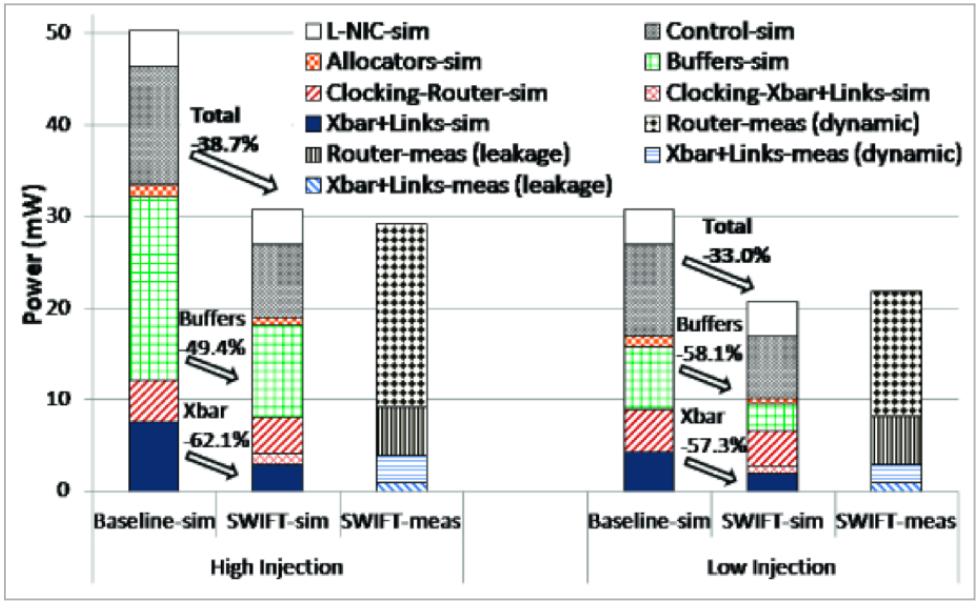






Measurement Summary





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

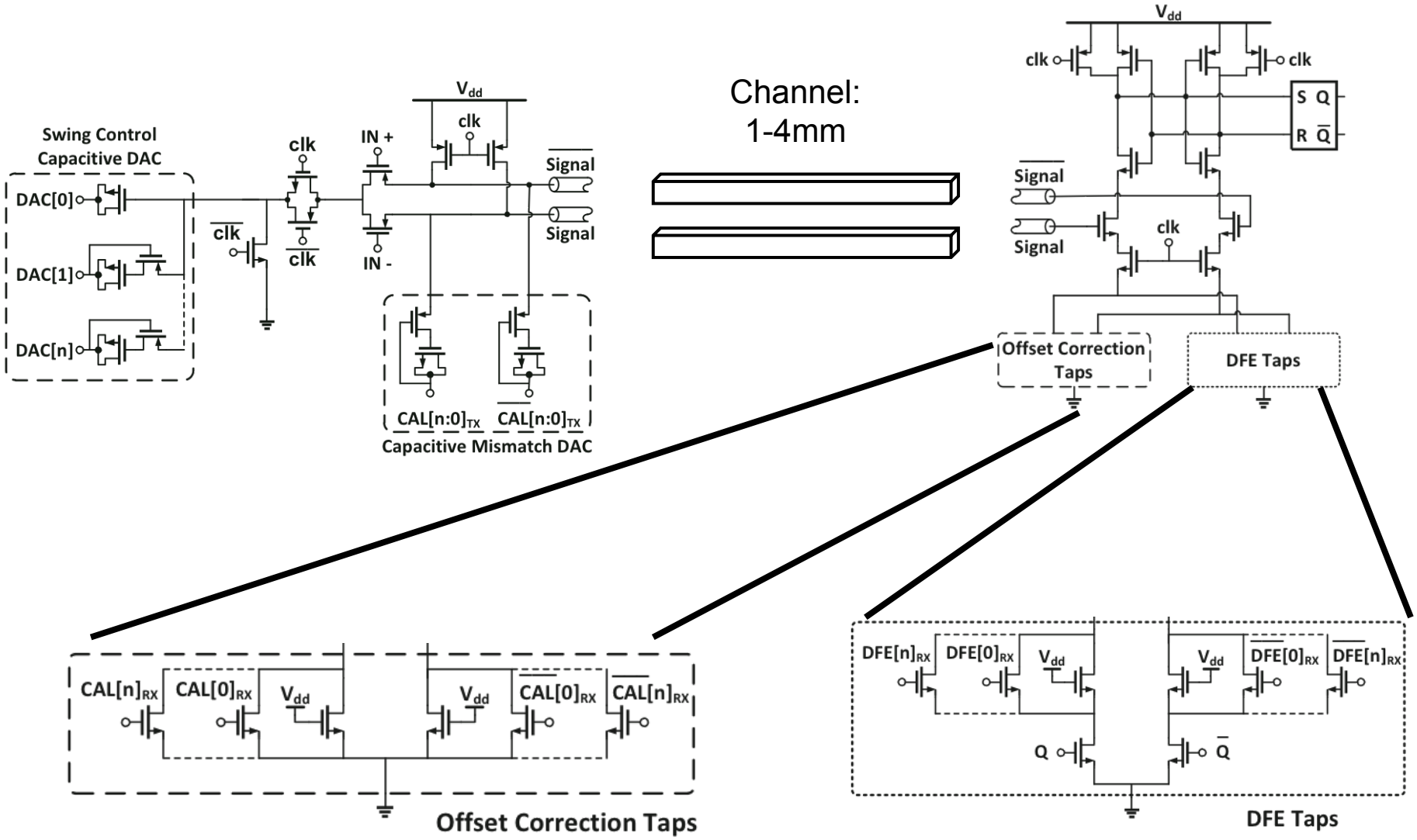
*Relative to a baseline synthesized VC NoC router

Next Goal: 1-5fJ/mm

	Conventional Full Swing	Schinkel (JSSCC '09)	Stojanovic (ISSCC '09)	Our TFC Router	New Goal
Wire Length	1mm	2mm	10mm	1mm	1mm-5mm
Supply	1.2V	1.2V	-	1.2V	0.5-1.0V
Transceiver Area	21um^2	TX:20um	2880um^2	23um^2	**20-30um^2**
Signal Swing	1.2V	120mV	200mV	250mV	50mV
Energy/Bit	305fJ	105fJ	356fJ	28-60fJ	**1-5fJ/mm**

- Determine: Fundamental limits to energy-efficient, on-chip link
- GOAL: 5-50x improvement in on-chip link energy
  - Energy scalability
  - Low area
  - Robust

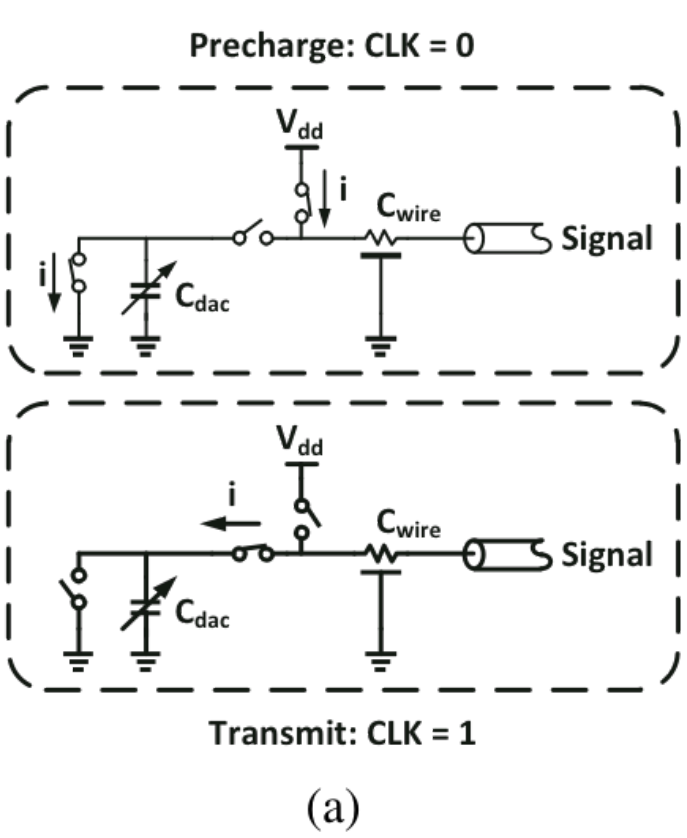
Energy-Scalable On-Chip TxRx



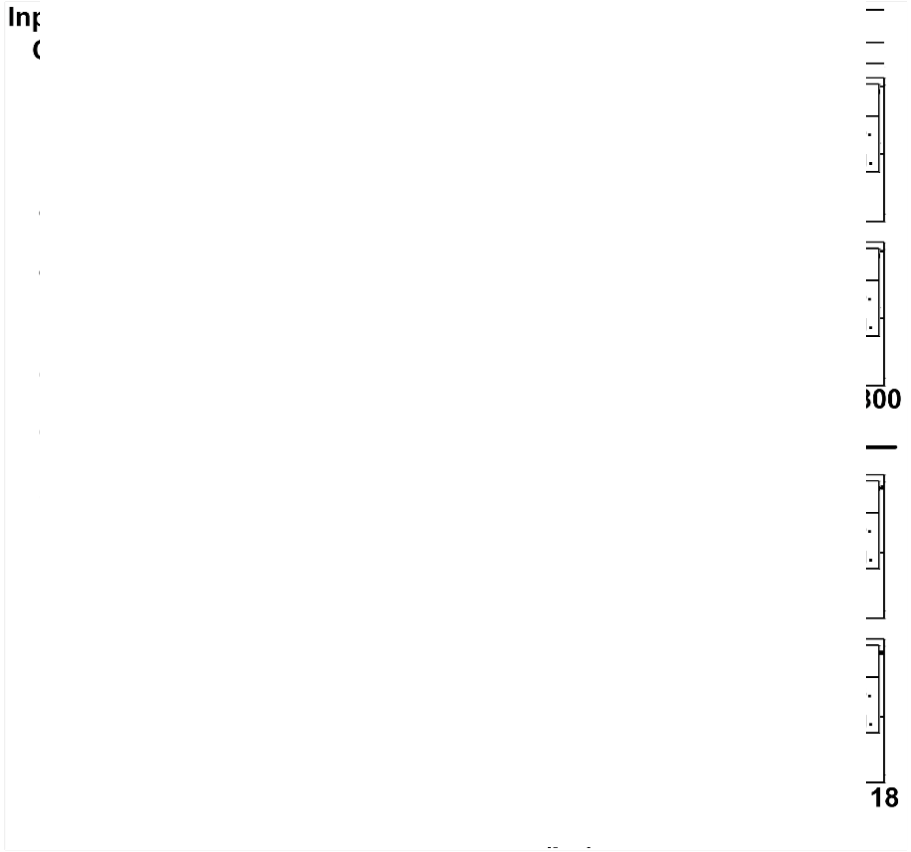
Digital Offset Cancellation

Decision Feedback Equalization

Dynamic Operation



(a)

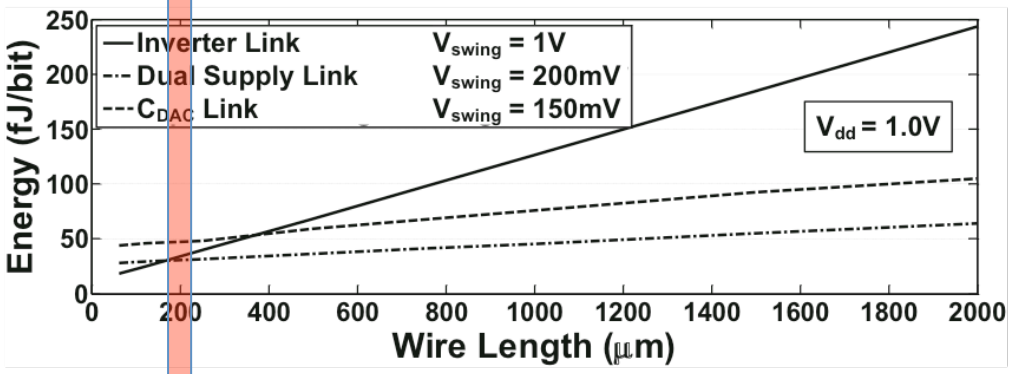


On-Chip Links – Measurement Results

Type	**Dual Supply**	**Capacitive FFE**		**Conventional Full-Swing**			**Capacitive-TX 10b RX Cal, No DFE**			**Dual Supply-TX, 2b RX Cal**							
										No DFE	3b DFE	No DFE	3b DFE	No DFE	3b DFE	No DFE	3b DFE
Reference	[4]	[2]	[5]	[5]	Simulated		This Work										
Technology	90nm	90nm	90nm	90nm	65nm		65nm										
Target Application	SoC	High Performance			SN	SoC	SN	SN	SoC	SN		SoC		SN		SoC	
Frequency (MHz)	100's	1000's			10's	100's	1's	10's	100's	10's		100's		10's		100's	
Wire Length (mm)	1	5	2	2	1		1			1		1		4		4	
Supply Voltage (V)	1.2	1.0	1.2	1.2	0.5	1.0	0.35	0.5	1.0	0.5		1.0		0.5		1.0	
Data Rate (bps)	300M	2.4G	9G	5G	55M	400M	5M	30M	622M	70M	72M	800M	805M	45M	50M	200M	200M
Signal Swing (mV)	250	100	120	1200	500	1000	40	75	230	250	250	250	250	250	250	275	275
E/b/mm (fJ)	64	48	52.5	210	30	126	8.4	10.9	136	6.5	6.6	37.6	38.4	4.1	4.0	20.2	20.2
Transceiver Area (μm^2)	23	730	N/A	N/A	INV 2.6	DFF 10.7	TX 122	RX 112	Total 234	TX 5.5		RX 35.1		Total 40.6			

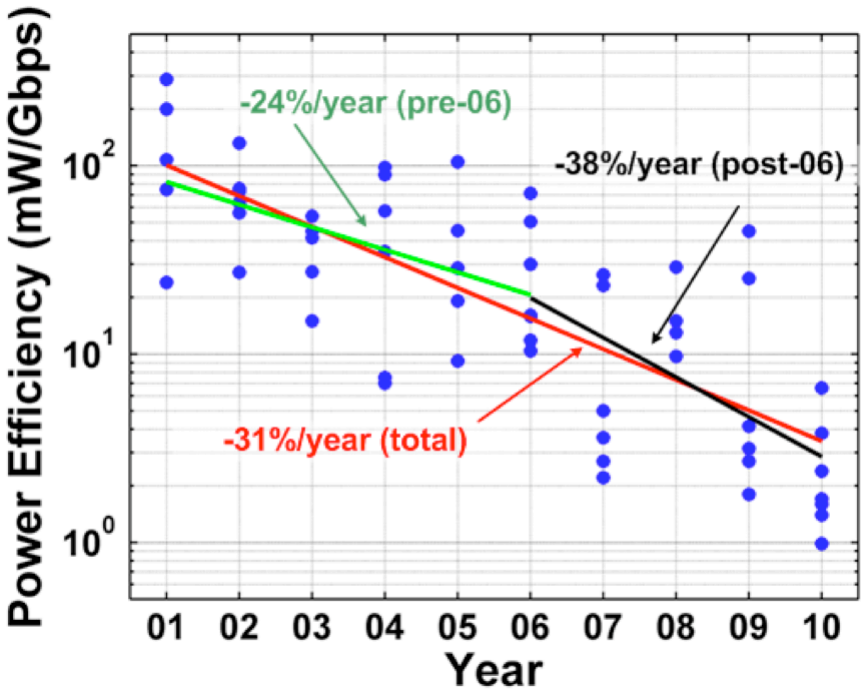
SoC - System-on-Chip, SN - Sensor Node/Network

Inverter Cross-Over Point



(a)

Off-Chip I/O Scaling Trends



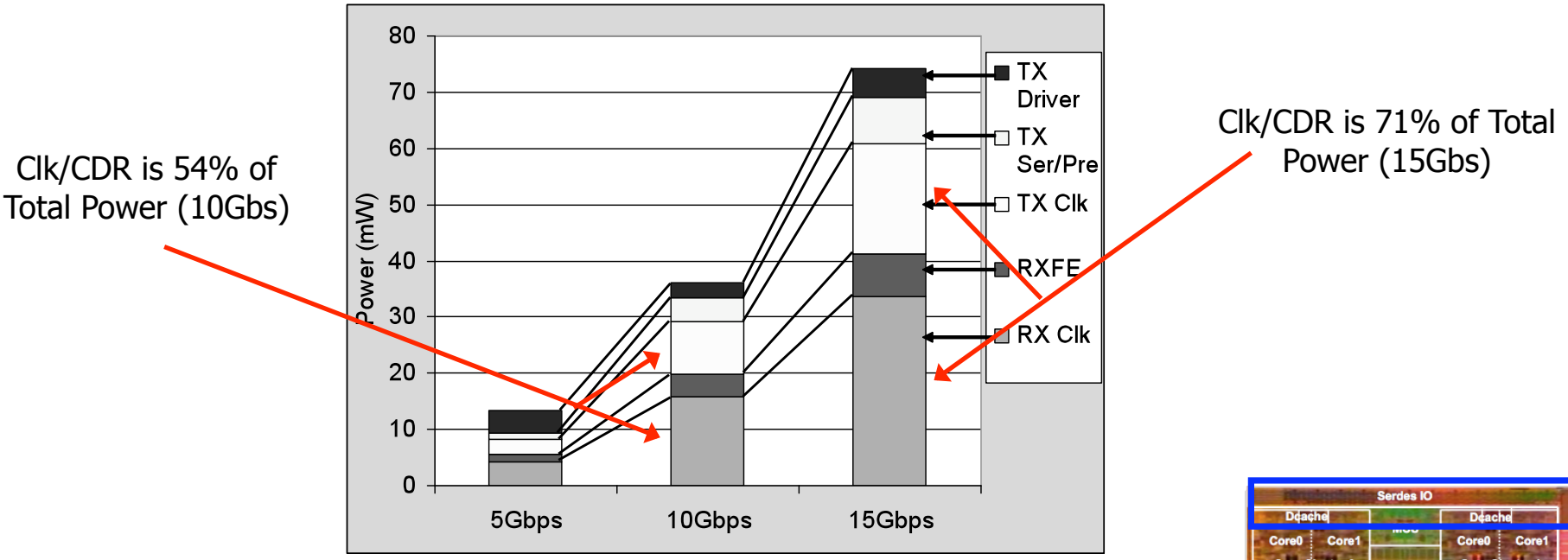
- I/O power efficiency is a function of:
  - data rate
  - process technology
  - channel loss
    → equalization complexity

- Project goal: < 1mW/Gbps over a scalable 5-10Gbps data rate

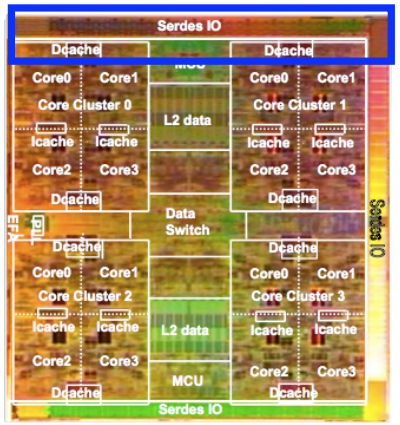
Source: S. Palermo, Texas A&M

(2) Off-Chip Links: Global Clock Distribution Optimization (with Intel)

- Clock dominates power in links

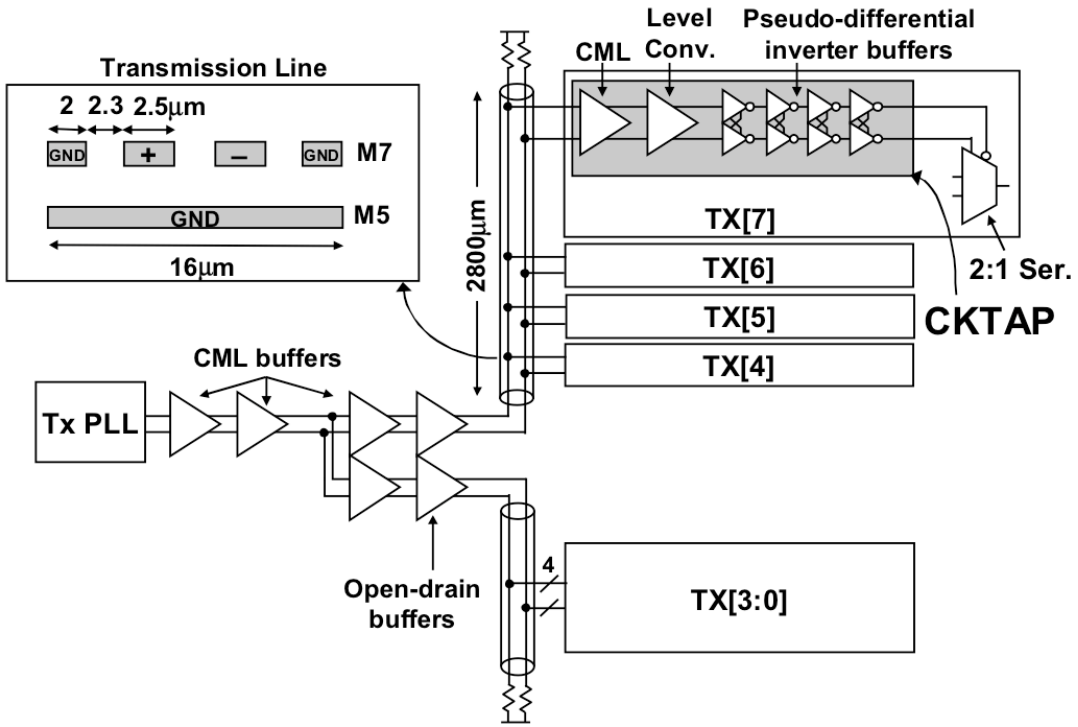


Intel (VLSI, 2007)



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

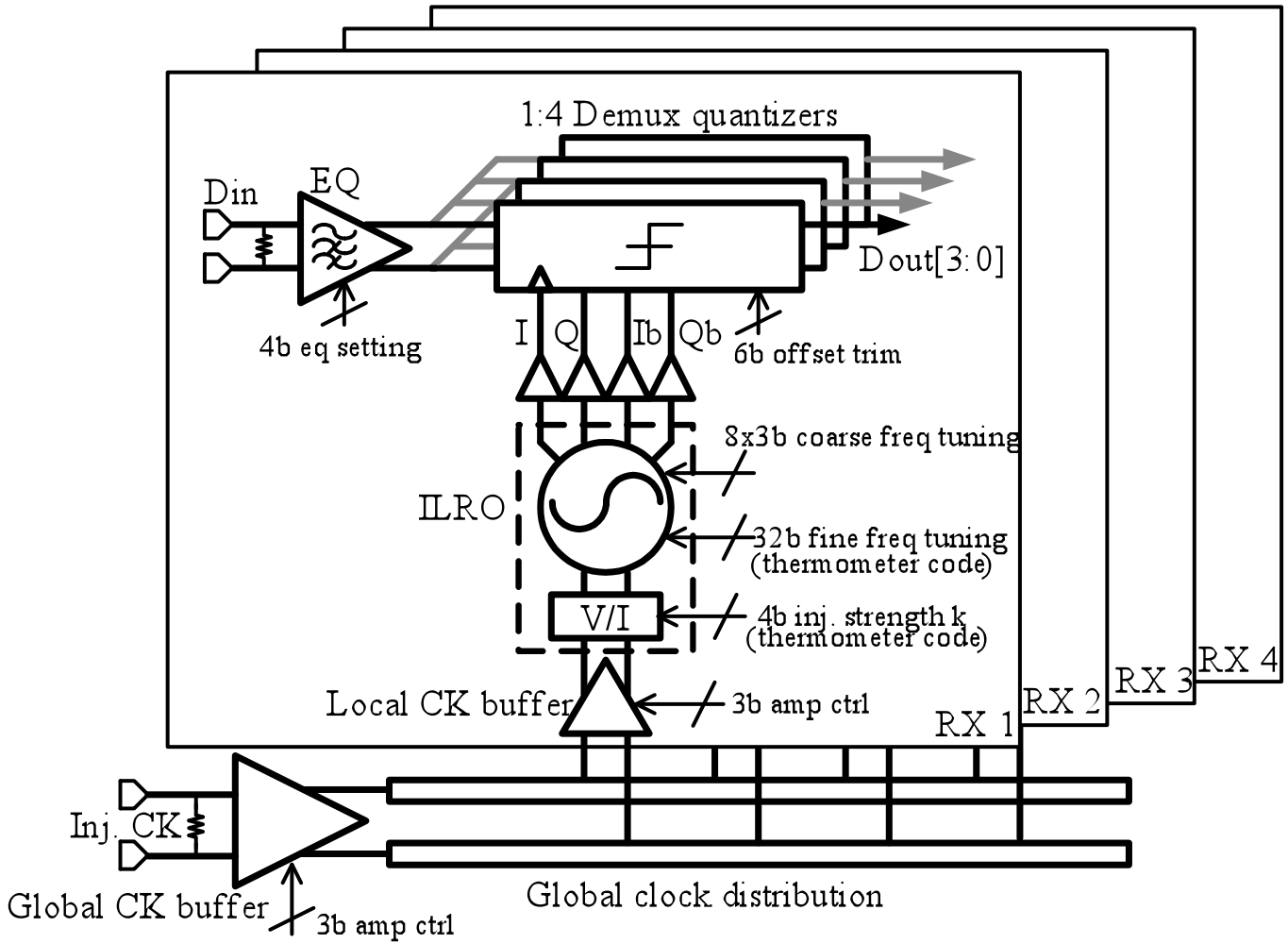
Conventional Clock generation and distribution



2) DESKEW: Significant energy in multi-phase generation

1) CLOCK DISTRIBUTION:
0.5W in the global distribution alone

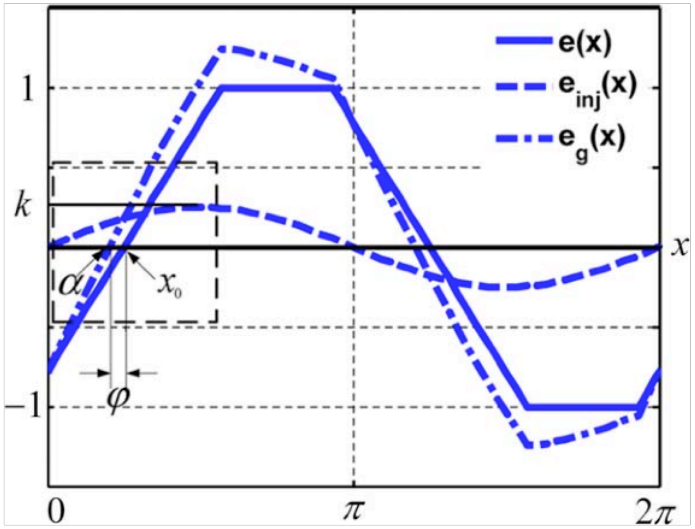
Chip #1: Injection-Locked Receiver Architecture

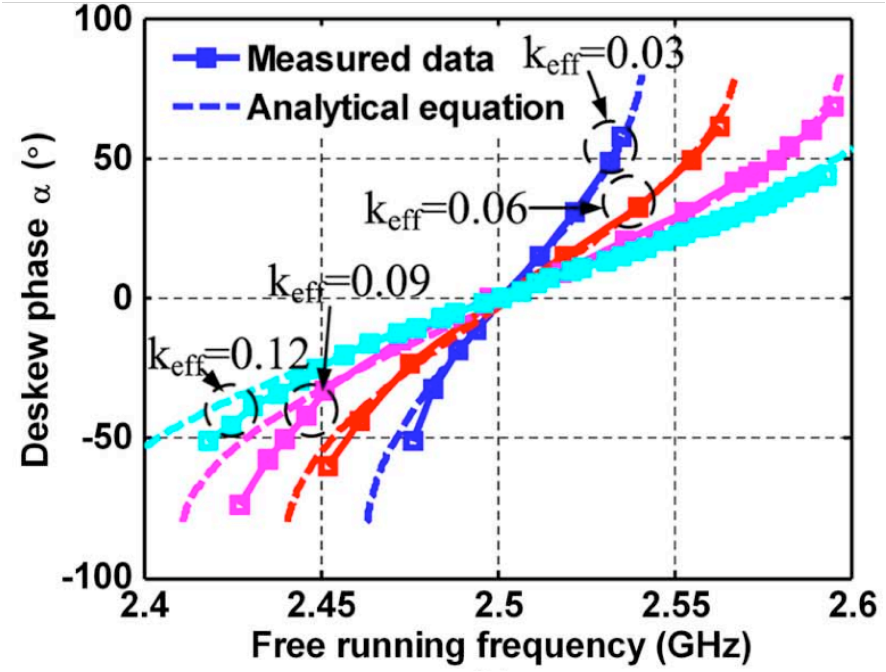


ILRO: Extension of Adler's Equation

$$\omega_{SL} = k/\frac{d\varphi}{d\omega} = k\frac{\omega_0}{2Q}$$

Adler's doesn't apply to ring oscillators!



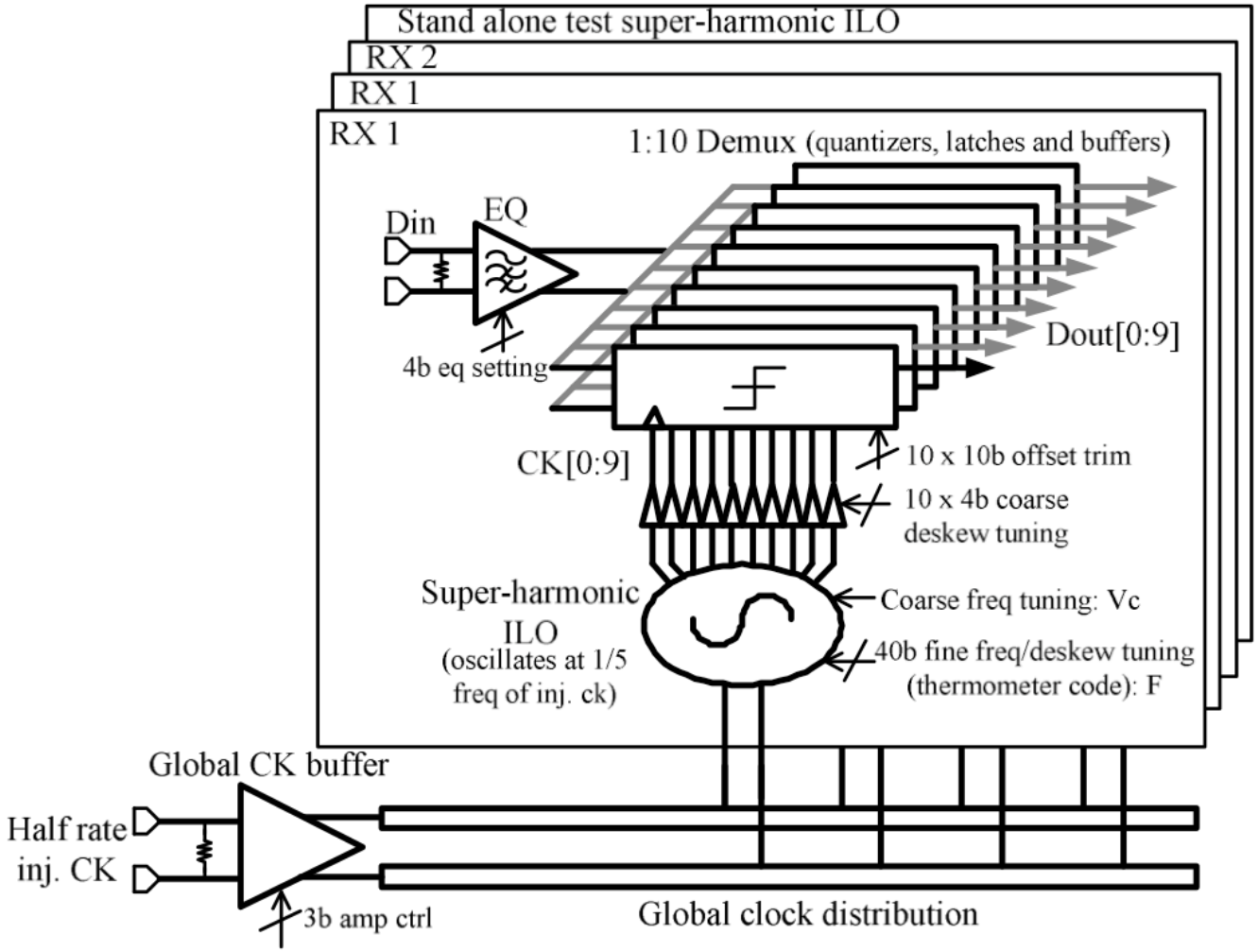


$$H(j\omega) = -\left(\frac{A_0}{1 + j\omega/\omega_{3\ \mathrm{dB}}}\right)^N$$

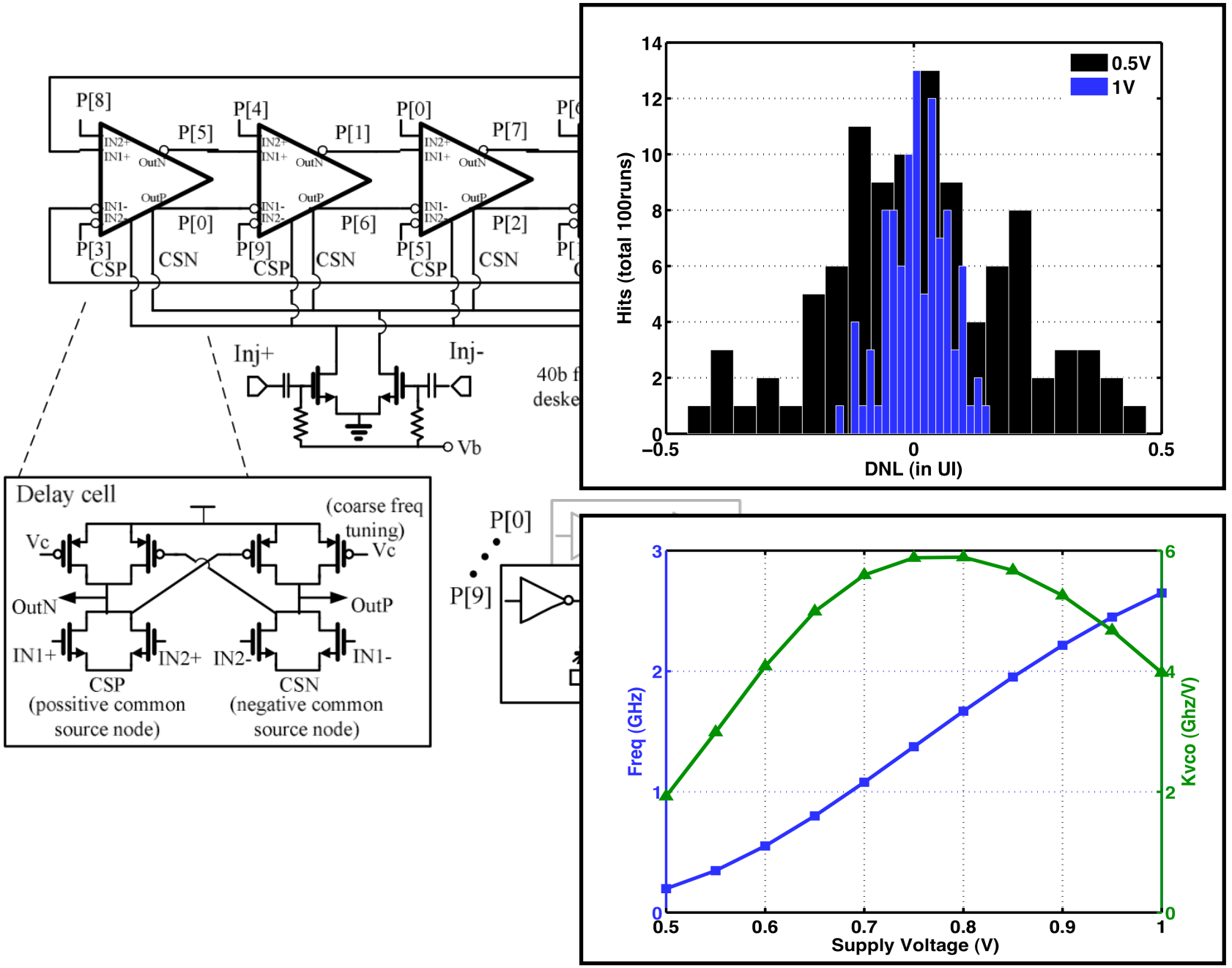
$$\omega_{SL} = \frac{k}{N\eta/\pi - k\cos\alpha}\frac{2\omega_0}{N\sin(2\pi/N)}$$

$$e_g(x) = e_{inj}(x) + e(x) = k\sin x + k_f(x - x_0)$$

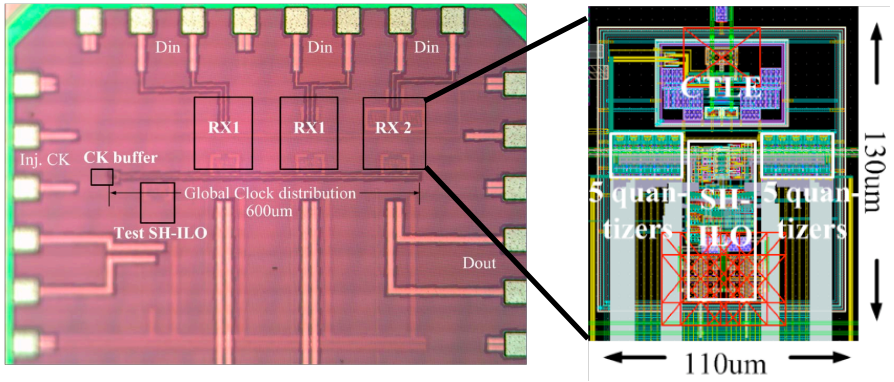
Chip #2: Near-Threshold, 0.15mW/Gbps, 8Gbps Serial Link Receiver



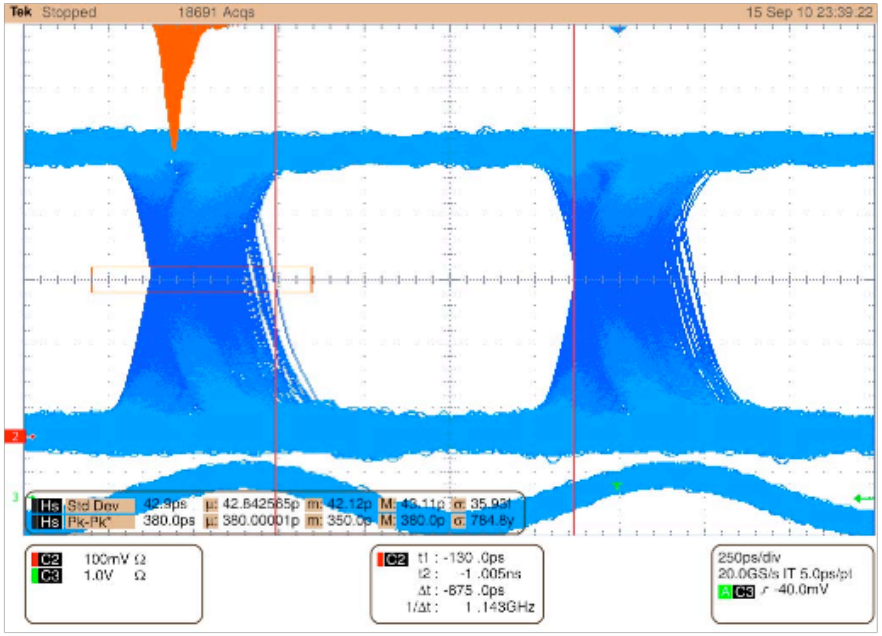
- Sub-harmonic Injection-Locking
- Operates at V_{DD} ~ 0.6V


P[8]
P[5]
P[4]
P[1]
P[0]
P[7]
P[3]
P[0]
P[9]
P[6]
P[5]
P[2]
CSP
CSN
IN2+
IN1+
IN1-
IN2-
OutN
OutP
Inj+
Inj-
Vb
40b f
deske
Delay cell
(coarse freq tuning)
Vc
OutN
OutP
IN1+
IN2+
IN2-
IN1-
CSP
(possitive common source node)
CSN
(negative common source node)
P[0]
P[9]
Hits (total 100runs)
0.5V
1V
DNL (in UI)
Freq (GHz)
Kvco (Ghz/V)
Supply Voltage (V)

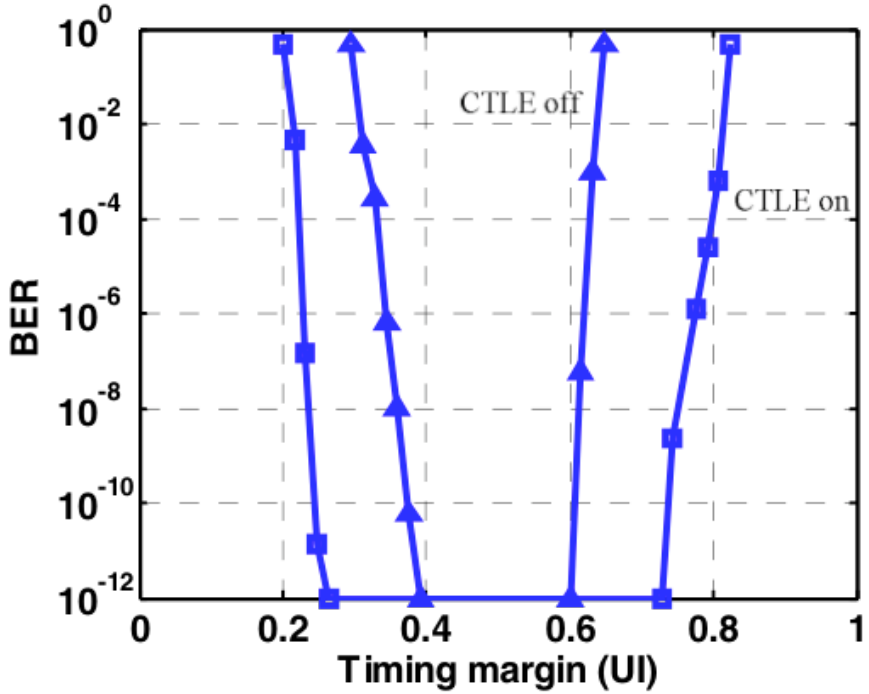
Measured Deskew Range and BER



$$\omega_{SL} = \eta \cdot \alpha_N \cdot \frac{2\omega_0}{N \sin(2\pi / N)} V_{inj}$$



(a)



(b)

Comparison Table

TABLE II. COMPARISION WITH PREVIOUS WORKS

	[6]	[4]	[5]	This work (RX1 and RX2)	
Data rate	6.25Gb/s	7.2Gb/s	7.4Gb/s	8Gb/s	
Architecture	Software CDR	Forwarded CK	Forwarded CK	Forwarded CK	
Phase deskew method	PLL with PI	Ring-ILO	ILO	Super-harmonic ILO	
Technology	90nm CMOS	90nm CMOS	65nm CMOS	65nm CMOS	
RX power	8.22mW	4.3mW	6.8mW	1.3mW	1.98mW
Power efficiency	1.31 mW/Gb/s	0.6 mW/Gb/s	0.92 mW/Gb/s	0.163 mW/Gb/s	0.25 mW/Gb/s
RX area	0.153mm^2	0.017mm^2	0.03mm^2	0.014mm^2	0.018mm^2

accepted, CICC-2011

Take-Away Points

- Lower energy silicon is possible
  - Aggressive interconnect circuits show:
    - Off-Chip: 5x improvements
    - On-Chip: 50x improvements
- Reliability at low-Vdd is issue
  - Explore in-situ adaptation to self-heal autonomously
- Magic bullets do NOT exist
  - Lower energy --> lower performance
  - Dynamically adapt the entire system
  - Requires co-design interaction between software, architecture, and underlying silicon

Outreach

- Publications
  - **K. Hu, T. Jiang, S. Palermo, and P. Chiang, "Low-Power 8Gb/s Near-Threshold Serial Link Receivers Using Super-Harmonic Injection Locking in 65nm CMOS", Accepted, IEEE Custom Integrated Circuits Conference, 2011.**
  - **R. Bai, J. Wang, L. Xia, F. Zhang, Z. Yang, W. Hu, P. Chiang, "Sinusoidal Clock Sampling for Multi-Gigahertz ADCs", accepted, IEEE Transactions on Circuits and System-I, 2011.**
  - **Lingli Xia, Jingguang Wang, Will Beattie, Jacob Postman, and Patrick Yin Chiang, "Sub-2ps, Static Phase Error Calibration Technique Incorporating Measurement Uncertainty Cancellation for Multi-Gigahertz Time-Interleaved T/H Circuits", accepted, IEEE Transactions on Circuits and System-I, 2011.**
  - **K. Hu, L. Wu, and P. Chiang, "A Comparative Study of 20-Gb/s NRZ and Duobinary Signaling Using Statistical Analysis", accepted, IEEE Transactions on VLSI Systems, May 2011.**
  - **T. Jiang, W. Liu, C. Zhong, F. Zhong, P. Chiang, "Single-Channel, 1.25-GS/s, 6-bit, Loop-Unrolled Asynchronous SAR-ADC in 40nm-CMOS", IEEE Custom Integrated Circuits Conference, Sep. 2010.**
  - **J. Postman and P. Chiang, "Energy-Efficient Transceiver Circuits for Short-Range On-chip Interconnects", Accepted, IEEE Custom Integrated Circuits Conference, 2011.**
  - **B. Goska, J. Postman, M. Erez, P. Chiang, "Hardware/software co-design for energy-efficient parallel computing," accepted, Department of Energy SciDAC Conference, July 2011.**
  - **Joseph Crop, Robert Pawlowski, Nariman Moezzi-Madani, Jarrod Jackson and Patrick Chiang, "Design Automation Methodology for Improving the Variability of Synthesized Digital Circuits Operating in the Sub/Near-Threshold Regime," accepted, Workshop on Low-Power System on Chip (SoC), 2nd Green Computing Conference, 2011.**
  - **T. Krishna, J. Postman, L.- S. Peh, P. Chiang, "SWIFT: A SWing-reduced Interconnect For a Token-based Network-on-Chip in 90nm CMOS", International Conference on Computer Design (ICCD), Amsterdam, Netherlands, October 2010.**
  - **E. Krimer, R. Pawlowski, M. Erez, P. Chiang, "Synctium: a Near-Threshold Stream Processor for Energy-Constrained Parallel Applications", IEEE Computer Architecture Letters, Jan/June 2010.**
- Talks
  - **P. Chiang, "Hardware/software co-design for energy-efficient parallel computing," accepted, Department of Energy SciDAC Conference, July 2011.**
  - **P. Chiang, Carnegie-Mellon, May 2011.**
  - **P. Chiang, Princeton, May 2011.**
  - **P. Chiang, UC-Davis, Feb 2011.**
  - **P. Chiang, Stanford, Feb 2011.**
  - **P. Chiang, Intel, Mar 2011.**
  - **P. Chiang, UC-San Diego, Jan 2011.**
  - **P. Chiang, Broadcom, Dec 2010.**
  - **P. Chiang, Illinois, Oct 2010.**
  - **P. Chiang, Michigan, Oct 2010.**
  - **P. Chiang, USC, Oct 2010.**