The Future of Electrical I/O for Microprocessors

Frank O’Mahony
frank.omahony@intel.com
Intel Labs, Hillsboro, OR USA

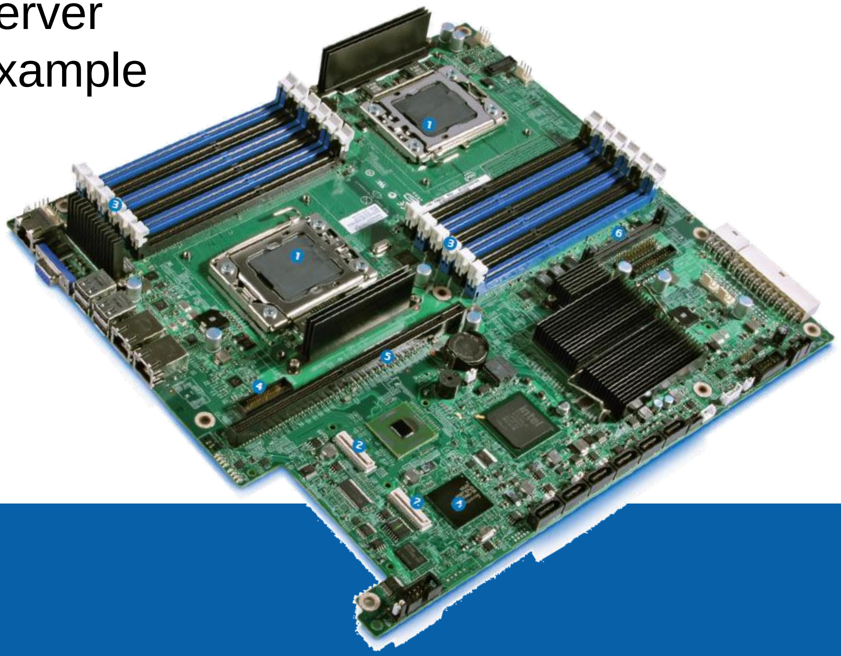
Outline

- 1TByte/s I/O: motivation and challenges
- Circuit Directions
- Channel Directions
- Tool Directions
- 470Gb/s Prototype

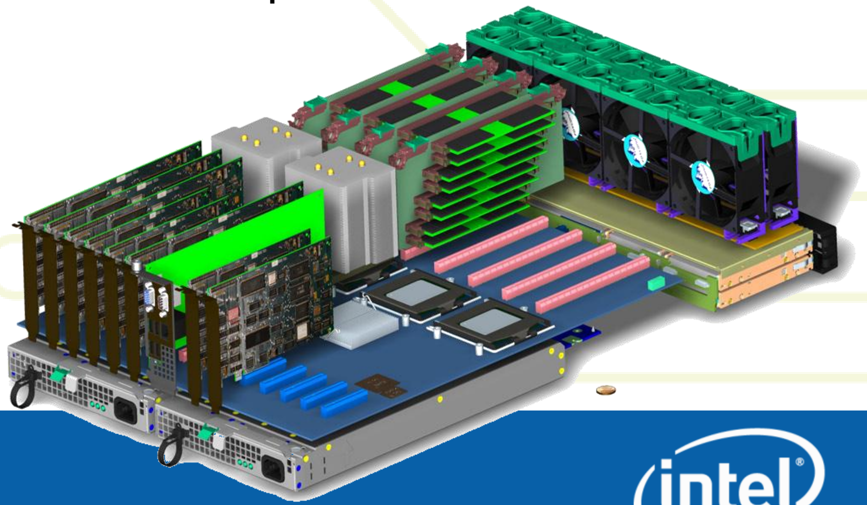
Microprocessor Bandwidth Needs

- As CPU core count increases, I/O bandwidth (BW) requirements will increase for all segments
- Current system bandwidth requirements (Y2010)
  - Client BW = ~50GB/s
  - Server BW = ~100GB/s
  - High-end Server BW = ~200GB/s

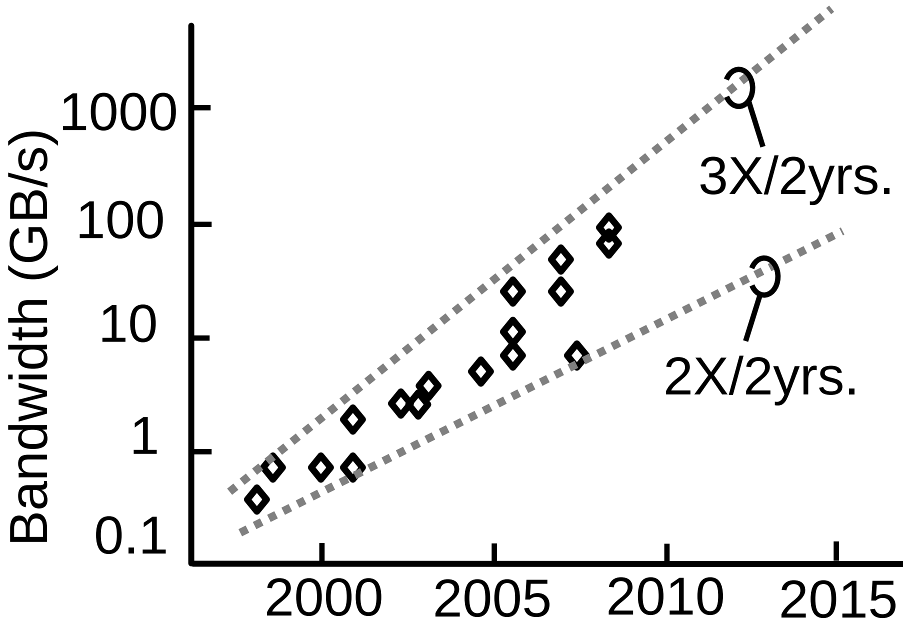
Server Example



High-End Server Example



Microprocessor Bandwidth Trends



<u>Bandwidth Drivers:</u>

CPU↔Memory
CPU↔CPU
CPU↔Peripheral
CPU↔I/O bridge

High-end microprocessors are expected to need ~<u>1TB/s</u> during coming decade

Microprocessor I/O Power

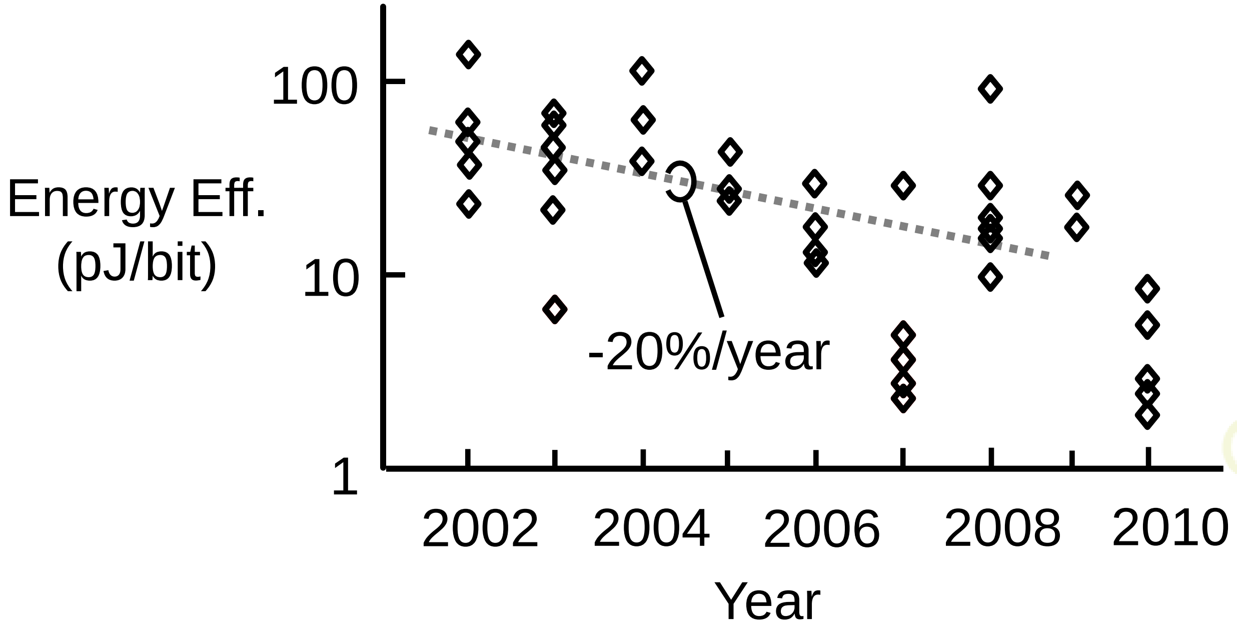
- Current system I/O power efficiency is 20-40pJ/bit

System	BW	I/O Pwr. Eff.	I/O Pwr
Client	~50GB/s	20pJ/bit	8W
Server	~100GB/s	20pJ/bit	16W
High-End Server	~200GB/s	20pJ/bit	32W

- If I/O power efficiency doesn't improve during the next decade, then:

1TB/s x 20pJ/bit = 160W

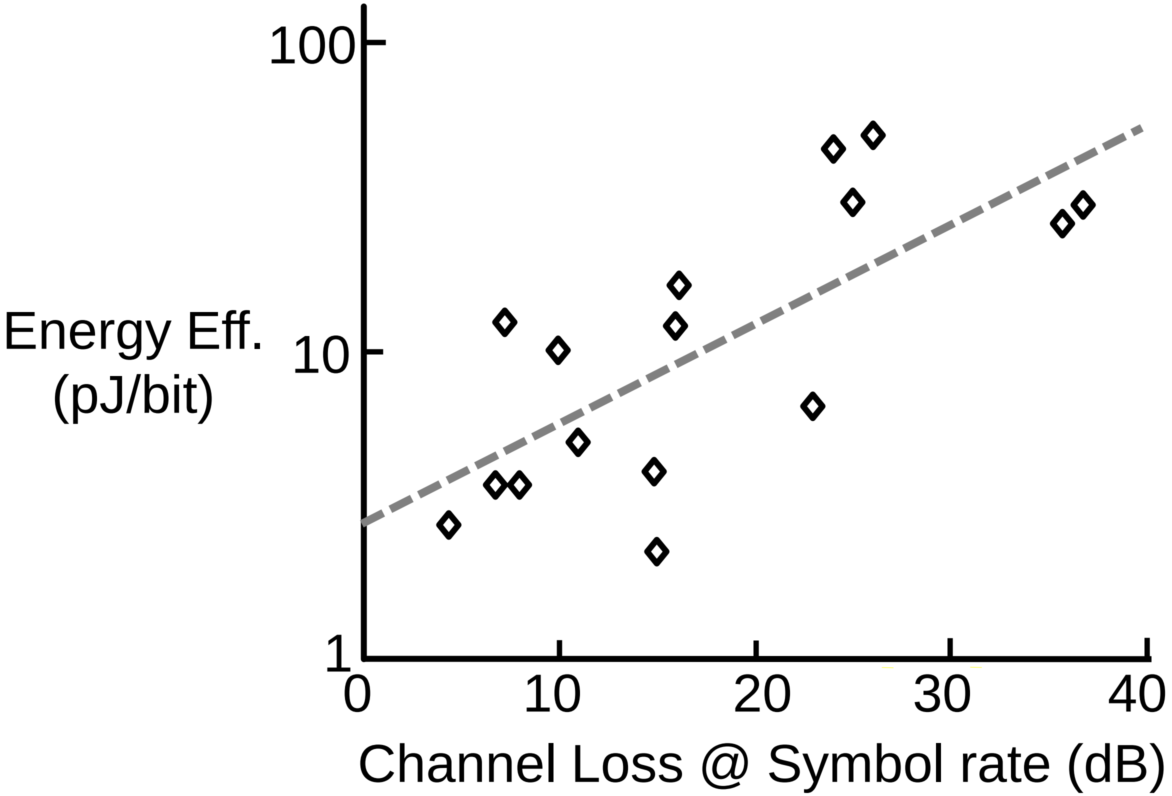
I/O Energy Efficiency Trends



Ref: R. Palmer, ISSCC '07

Issue: ~20% per year power reduction while bandwidth increasing 40-70% per year

Energy Efficiency and Channel Loss Tradeoff

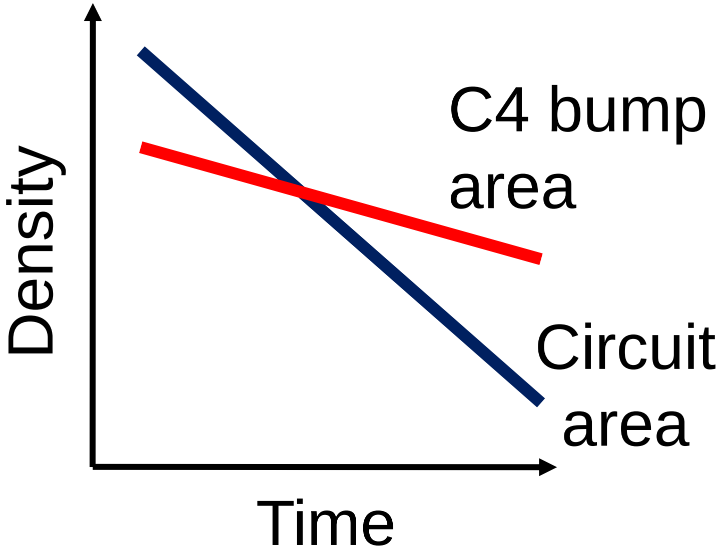


Based on transceivers reported 2006-2009 in 65-130nm CMOS

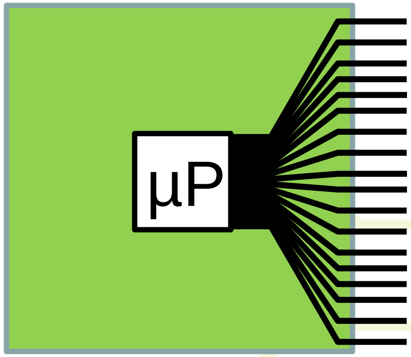
- Power efficiency is strongly correlated to channel loss
- Simply scaling per-pin BW will not meet power budget
- Low power interfaces should be “wider” not faster

Channel/Interconnect Density

- Conventional package/socket density does not scale with process
- “Width” of interfaces is limited by routing congestion



Flip-chip Package



C4 pitch << Pkg. pin pitch

Problem Statement Summary

- Bandwidth needs are quickly approaching 1TB/s

- Energy efficiency is not scaling as aggressively as bandwidth

- The channel limits our ability to increase per-pin data rate and/or increase the width of an interface

How Will Electrical I/O scale to 1TB/s?

1. Co-design the interconnects and I/O circuitry to meet bandwidth, scalability and power efficiency demands

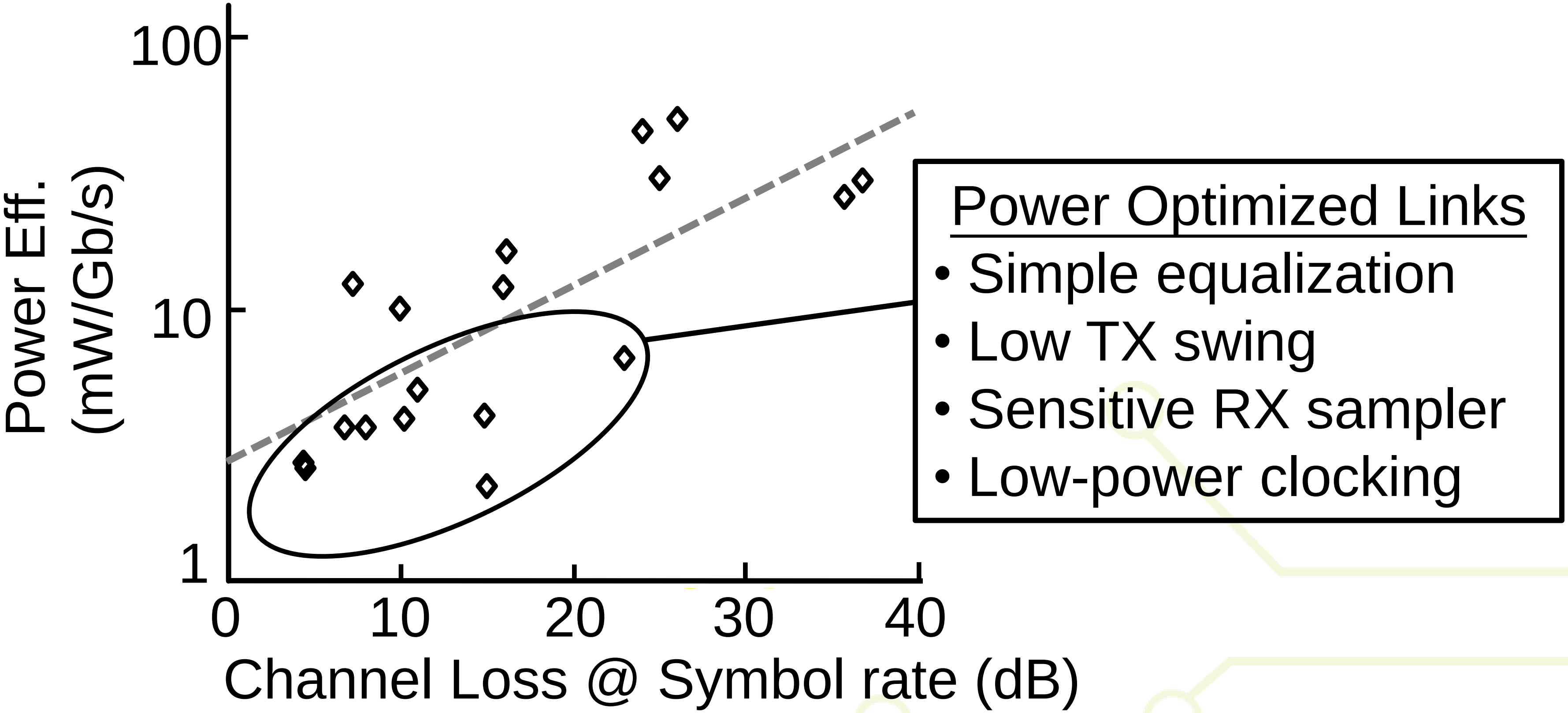
2. Scale the channel by transitioning to new channel configurations and materials

3. Use accurate, statistical link design tools to identify balanced architectures.

Outline

- 1TByte/s I/O: motivation and challenges
- Circuit Directions
- Channel Directions
- Tool Directions
- 470Gb/s Prototype

Low Active Power Techniques

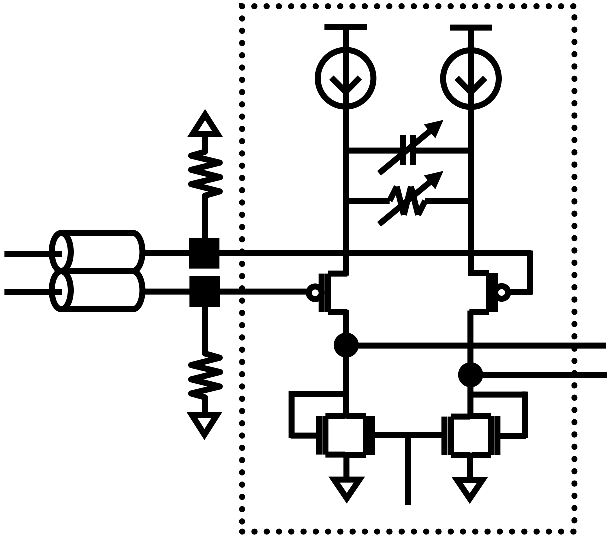


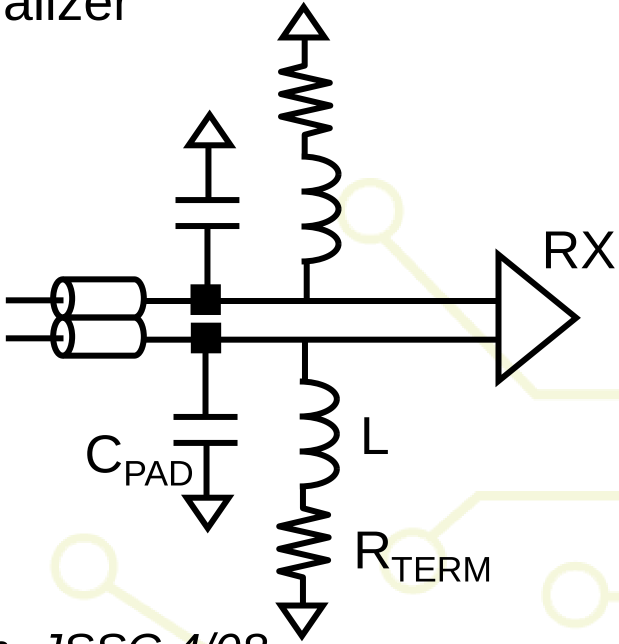
Power Optimized Links

- Simple equalization
- Low TX swing
- Sensitive RX sampler
- Low-power clocking

Minimize analog circuit complexity

- Lowest power links find ways to simplify equalization and clocking circuitry to reduce power
- Equalization examples:
  - Constrain equalization range by known channel characteristics
  - Continuous-time linear Rx equalizer

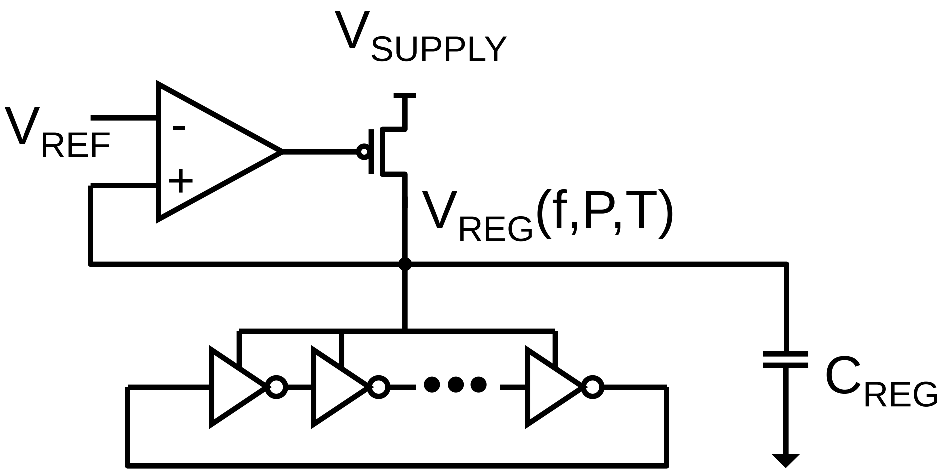




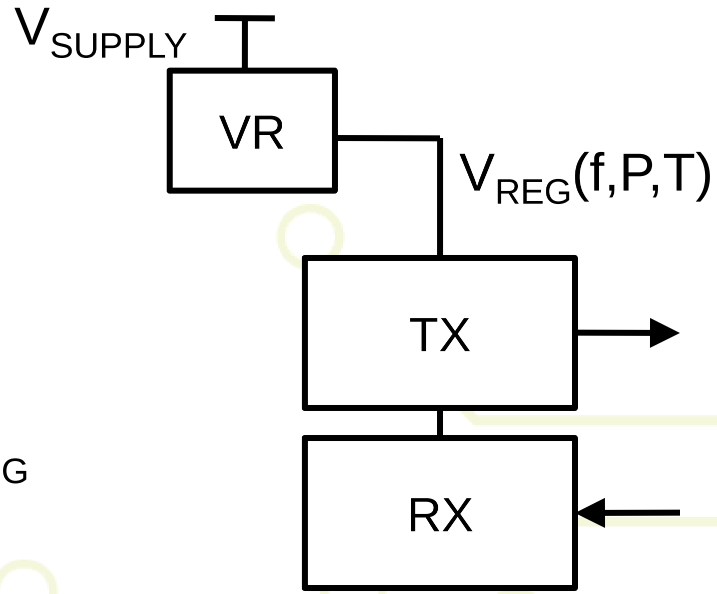
Ref: G. Balamurugan, JSSC 4/08

Power Management: Scalable supplies

- Adapt supply to frequency, process, temperature (f,P,T)
  - Digital: Power $\propto V_{SUPPLY}^2 \cdot f$
  - Analog: Power $\propto V_{SUPPLY} \cdot I_{bias}$
- Removes excess circuit BW and headroom

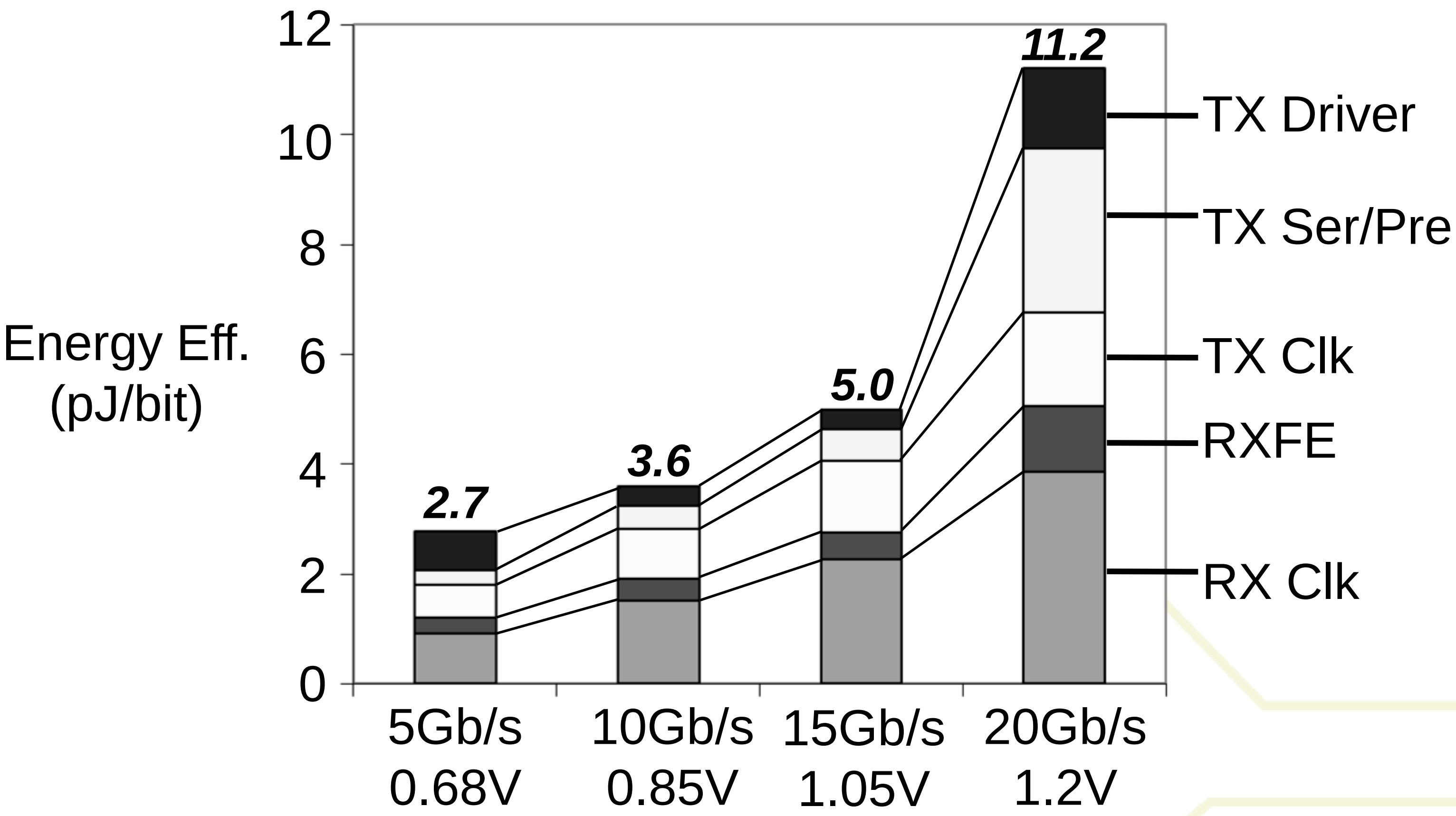


Regulated supply ring VCO



Data link with adaptive supply

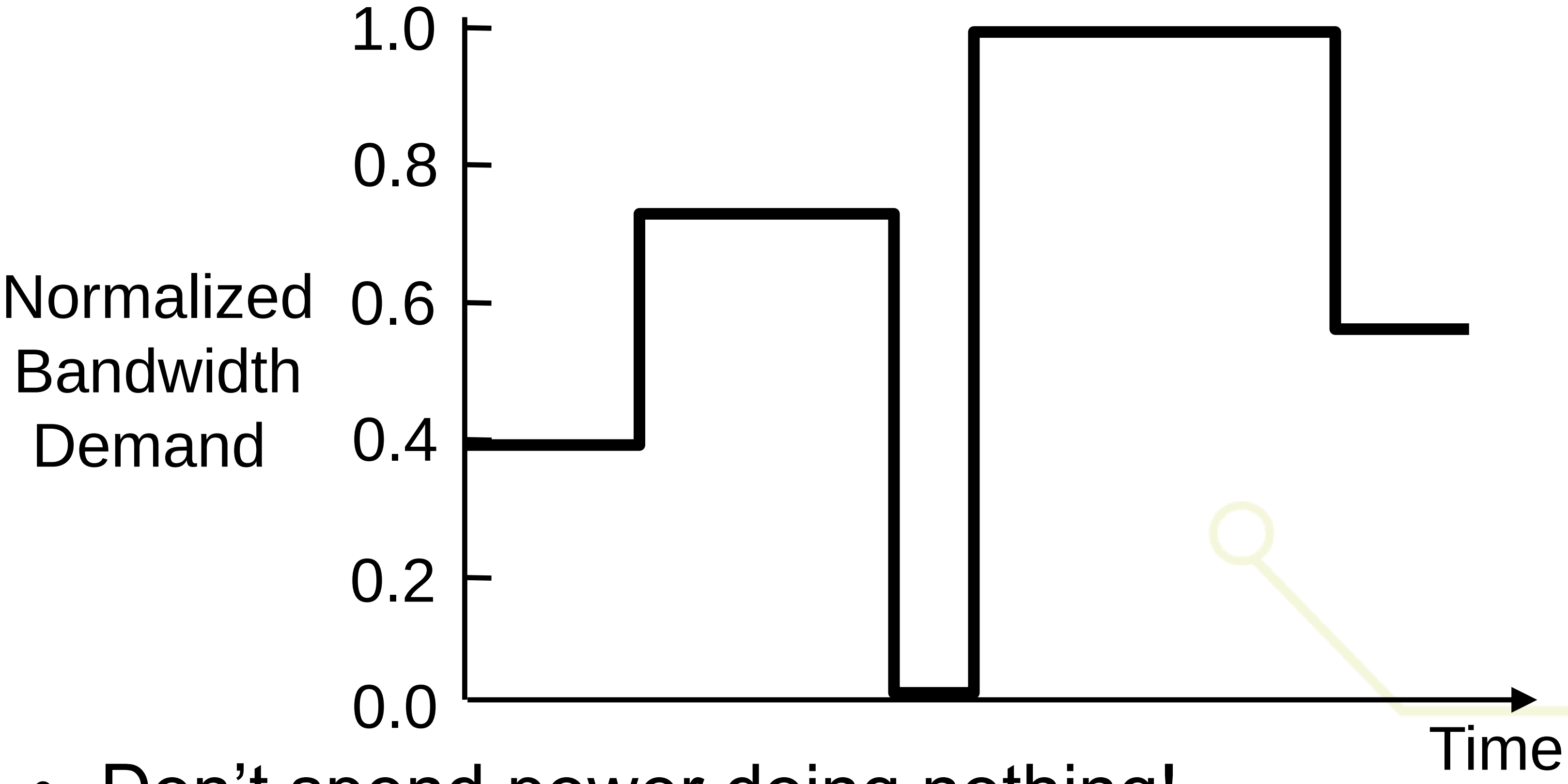
Power Management: Scalable supplies



- Power efficiency improves with adaptive supply/biasing

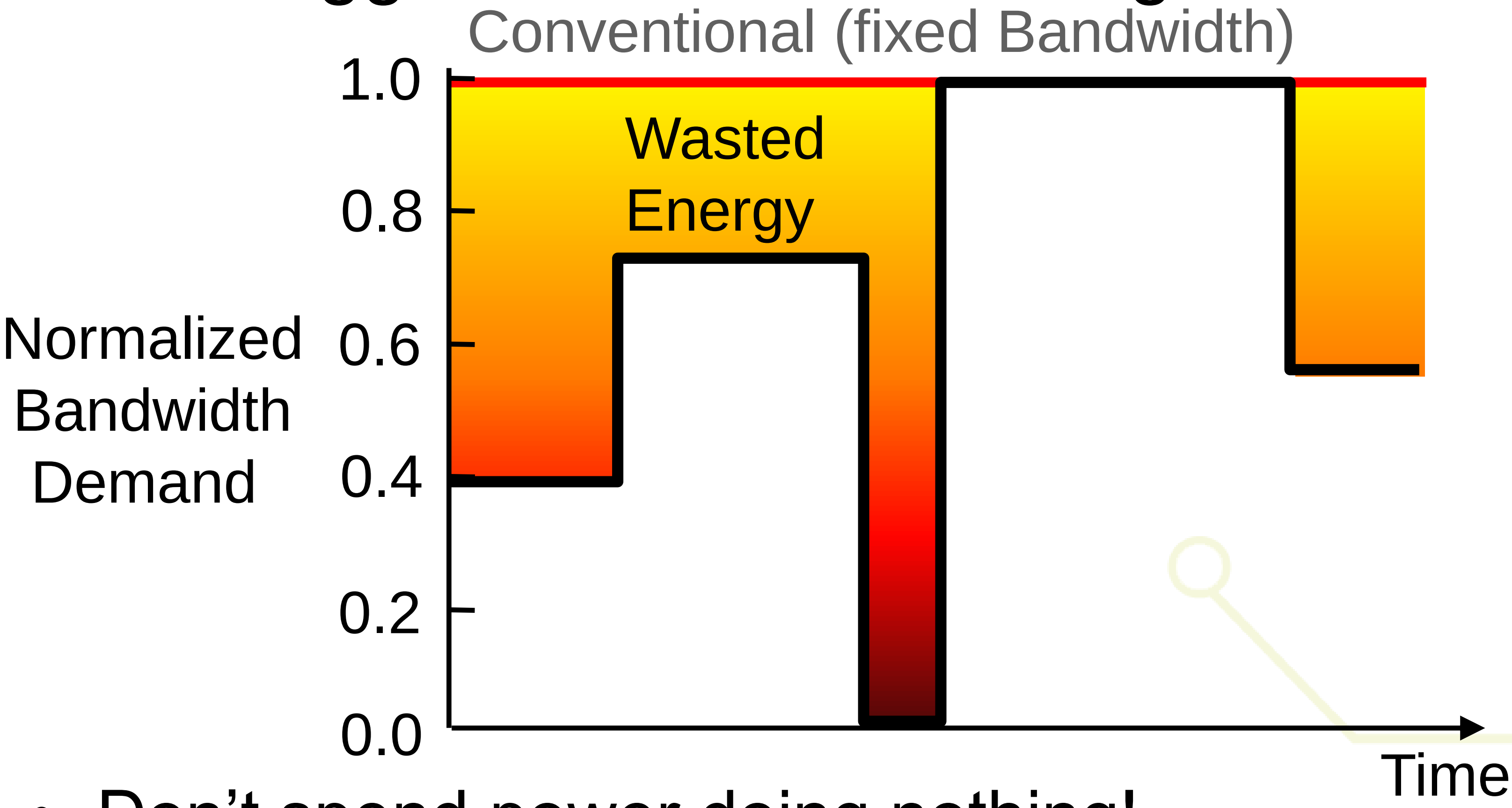
Refs: G. Balamurugan, JSSC 4/08 and B. Casper, ISSCC '06

Aggressive Power Management



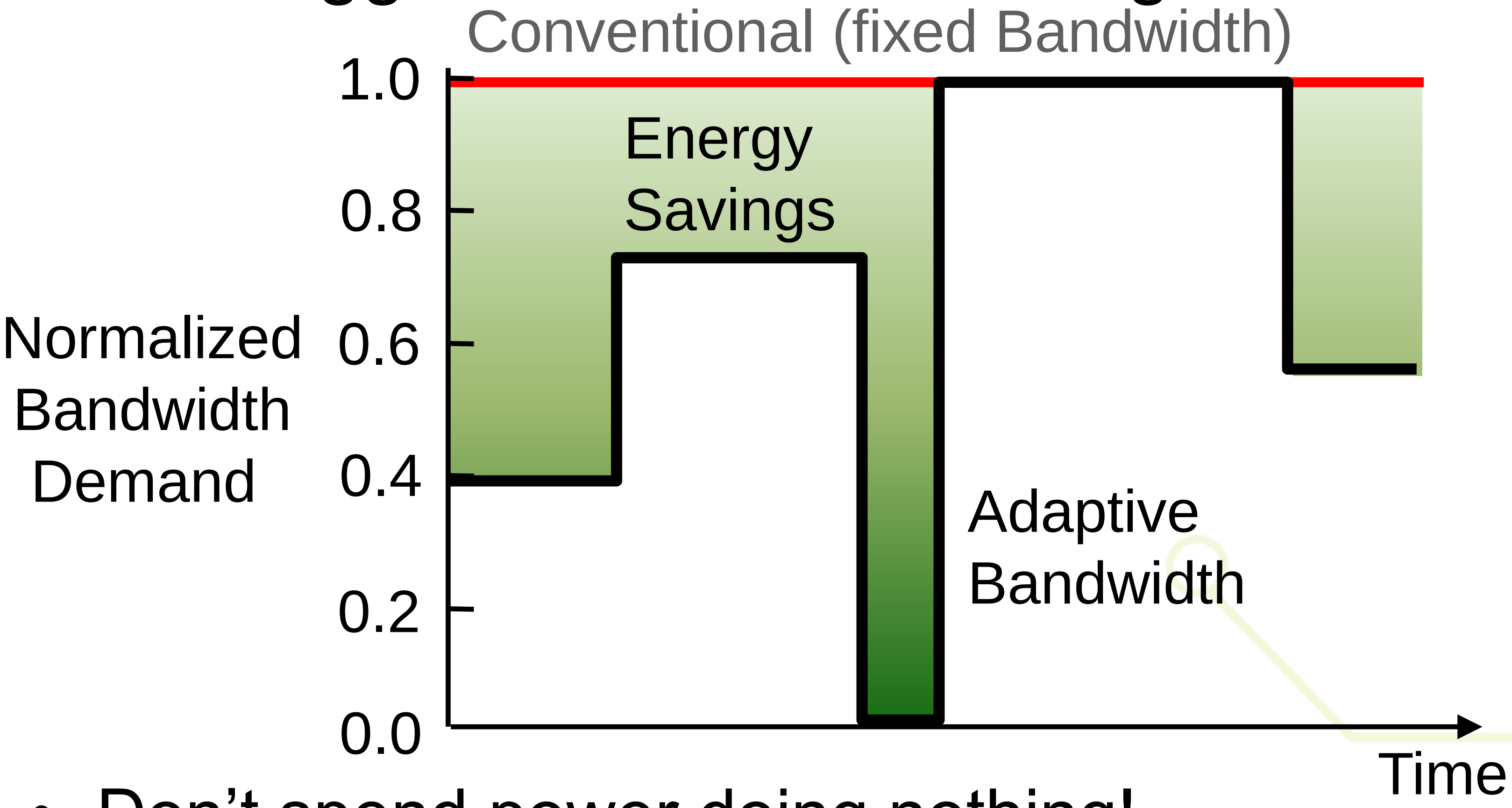
- Don’t spend power doing nothing!
- Rapidly adapt to bandwidth demand
  - – Requires fast, granular bandwidth adaptation

Aggressive Power Management



- Don’t spend power doing nothing!
- Rapidly adapt to bandwidth demand
  - Requires fast, granular bandwidth adaptation

Aggressive Power Management



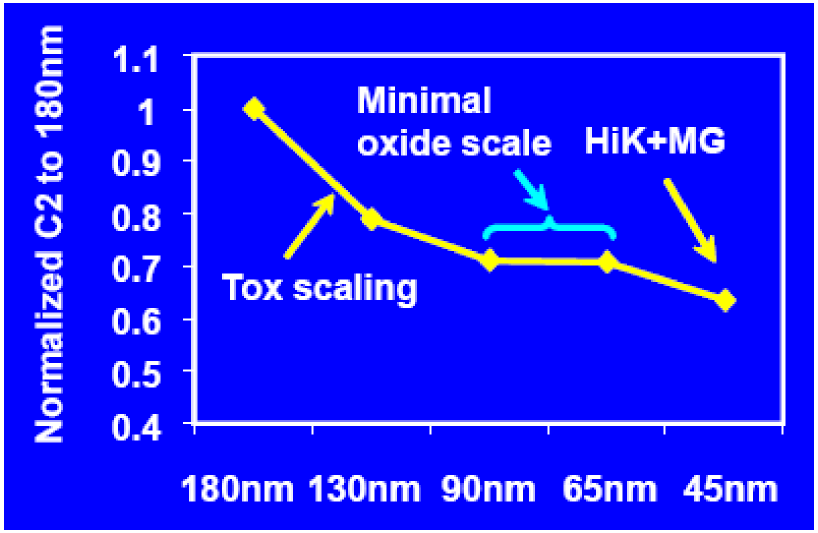
- Don’t spend power doing nothing!
- Rapidly adapt to bandwidth demand
  - – Requires fast, granular bandwidth adaptation

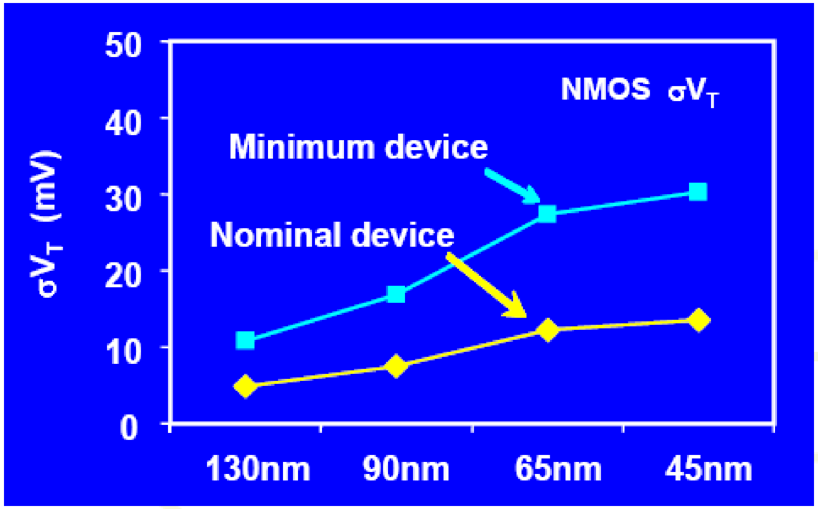
Device Variation in Scaled CMOS

- Device manufacturing tolerances are improving
- …but area scaling still causes higher variation
- Fundamental power/area to variation tradeoff is not acceptable

$$\sigma V_T = \frac{1}{\sqrt{2}}\left(\frac{c_2}{\sqrt{Weff \cdot Leff}}\right)$$

Need circuit architectures that fundamentally change this tradeoff.

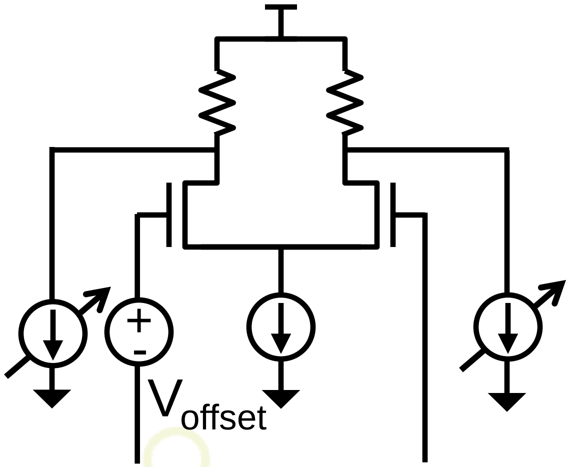




Ref: K. Kuhn, IEDM 2007

Mitigating Device Variation

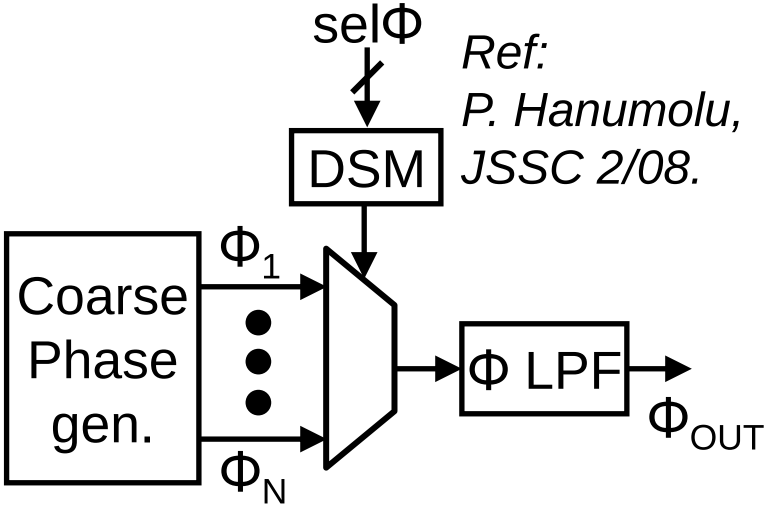
- Calibration greatly improves the power/variation tradeoff
  - Receiver offset calibration
  - Duty cycle correction
  - Adaptive equalizers
  - Clock recovery (or deskew)
- Simple calibration doesn’t alleviate all variation issues (e.g. PSRR)



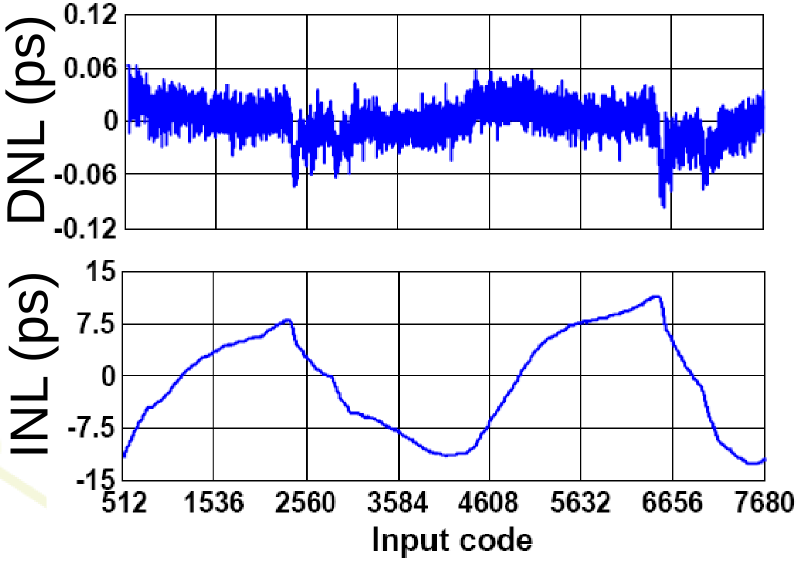
Circuit derivatives (gm, ro) are not calibrated by offset calibration → PSRR is not calibrated

Mitigating Device Variation

- Calibration greatly improves the power/variation tradeoff
  - Receiver offset calibration
  - Duty cycle correction
  - Adaptive equalizers
  - Clock recovery (or deskew)
- Simple calibration doesn’t alleviate all variation issues (e.g. PSRR)
- Possible solutions:
  - “Dynamic” calibration (e.g. auto-zero)
  - Redundancy/reconfigurability
  - Better “correct by design” circuits



Ref: P. Hanumolu, JSSC 2/08.



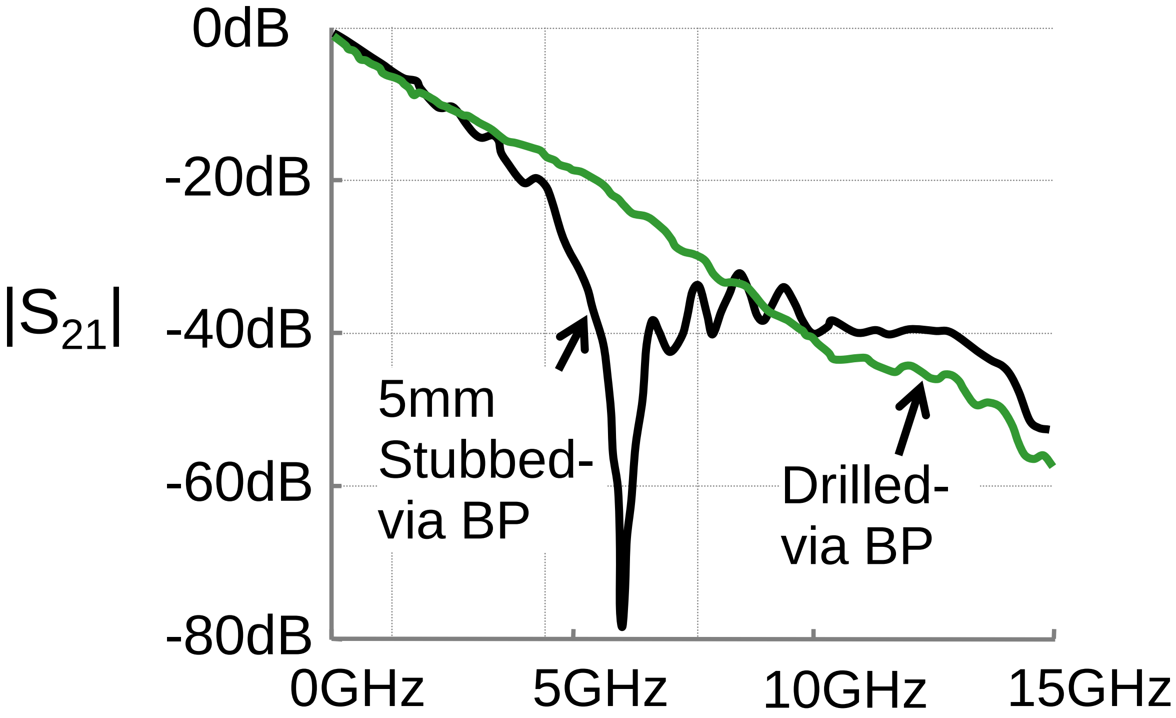
Outline

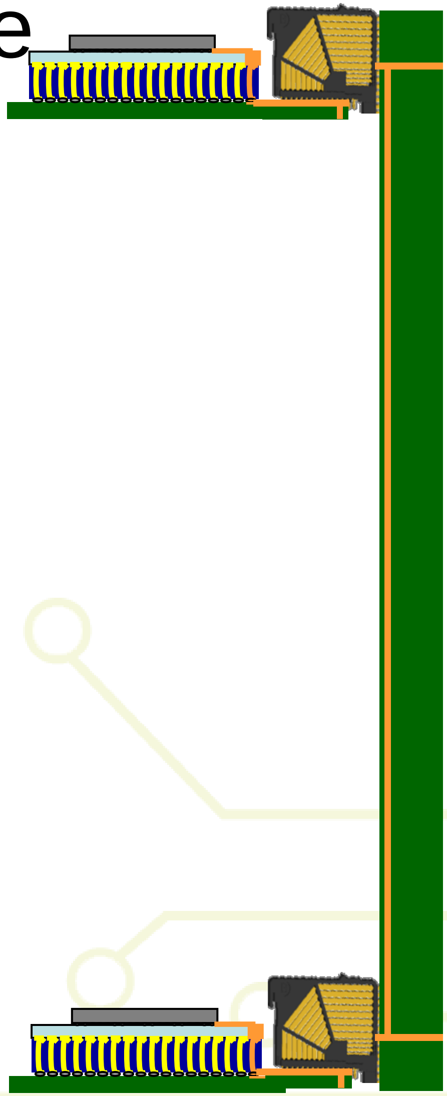
- 1TByte/s I/O: motivation and challenges
- Circuit Directions
- Channel Directions
- Tool Directions
- 470Gb/s Prototype

Channel scaling

- Circuit innovation alone will probably not be enough to reach the 1TB/s target → the channel needs to scale too!
- Better signal integrity: Improved electrical characteristics mean less power in clocking and equalization
- Higher density: More lanes allow each lane to operate at lower data rate → better power efficiency

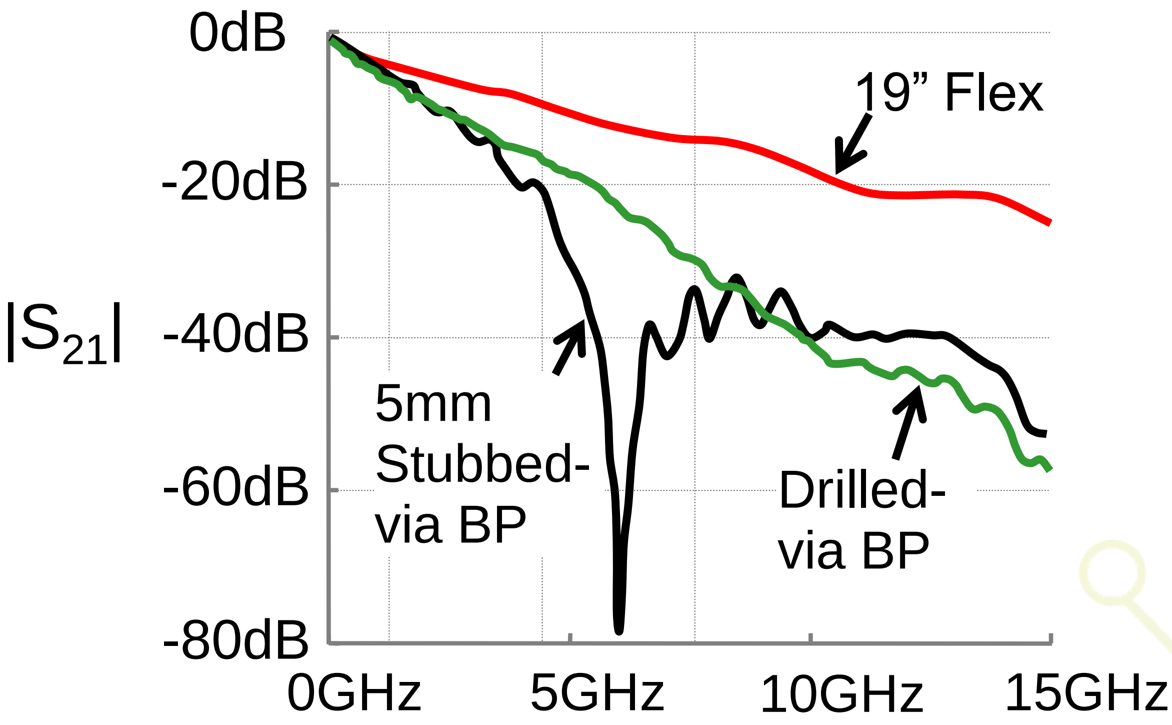
Channel vs. Equalization tradeoffs: Backplane example

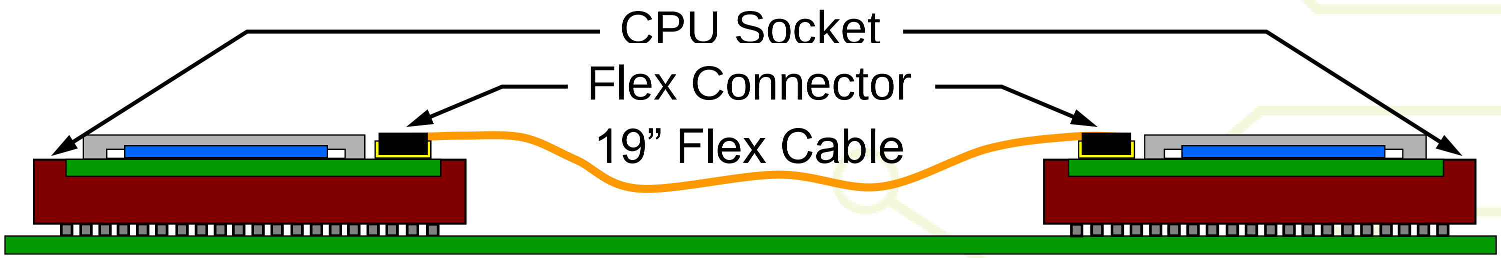




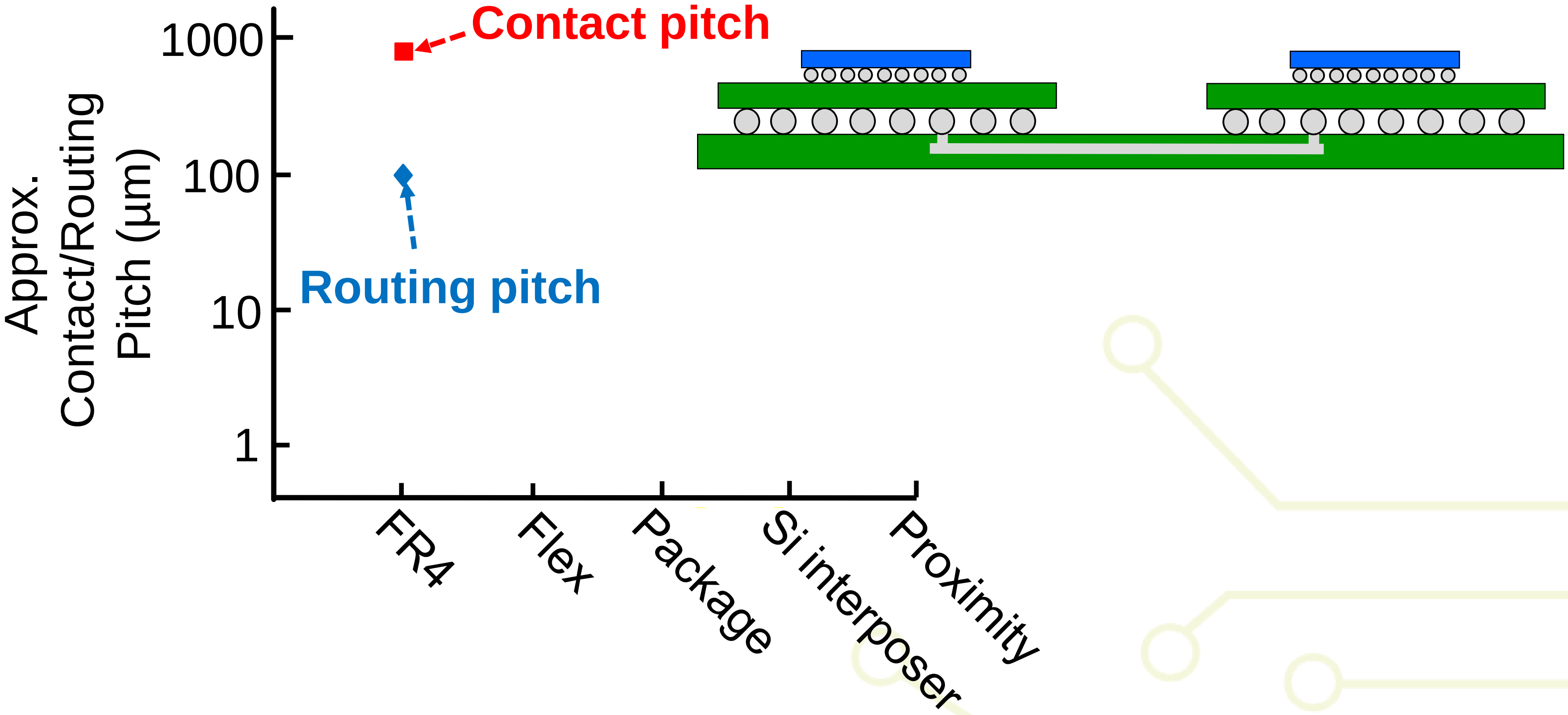
Ref: B. Casper, CICC '07.

Improve channel signal integrity

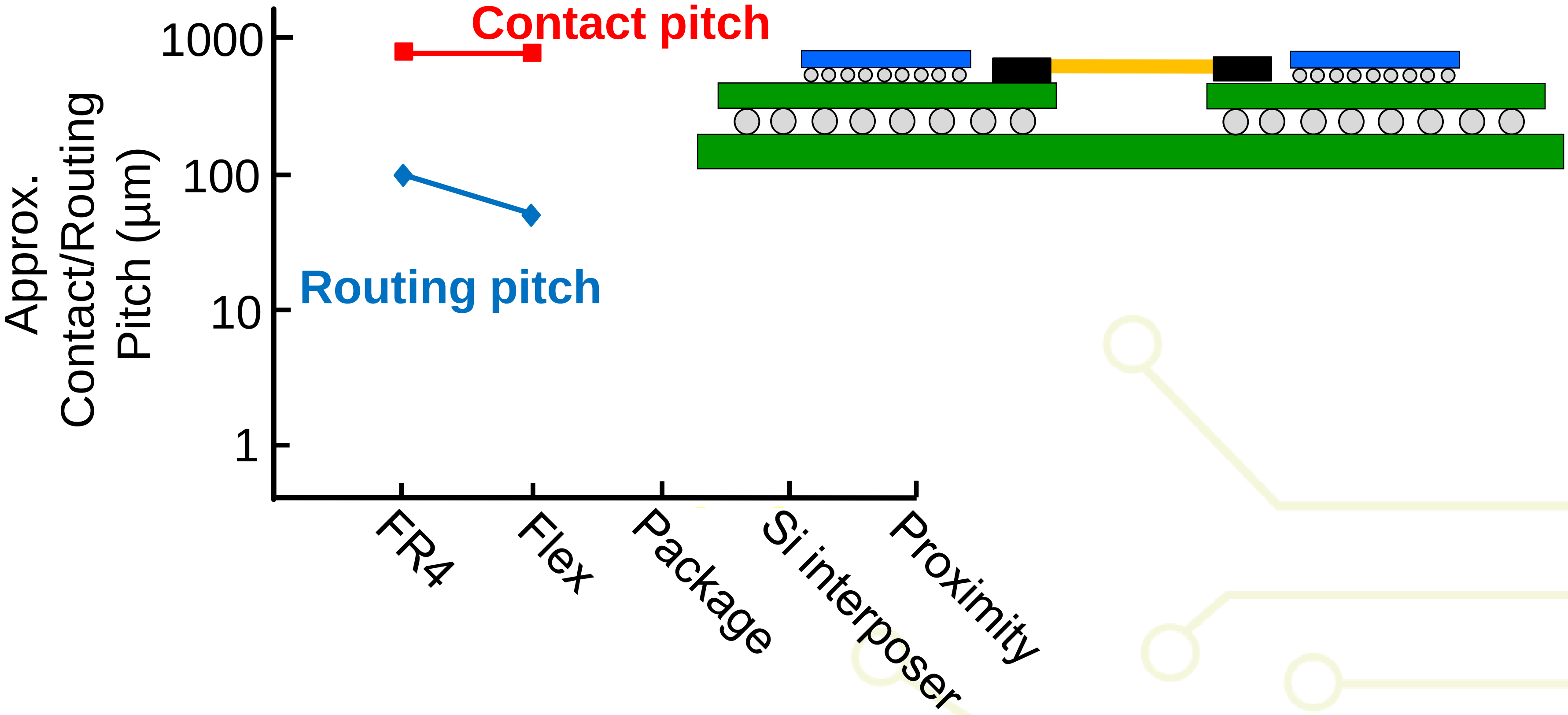




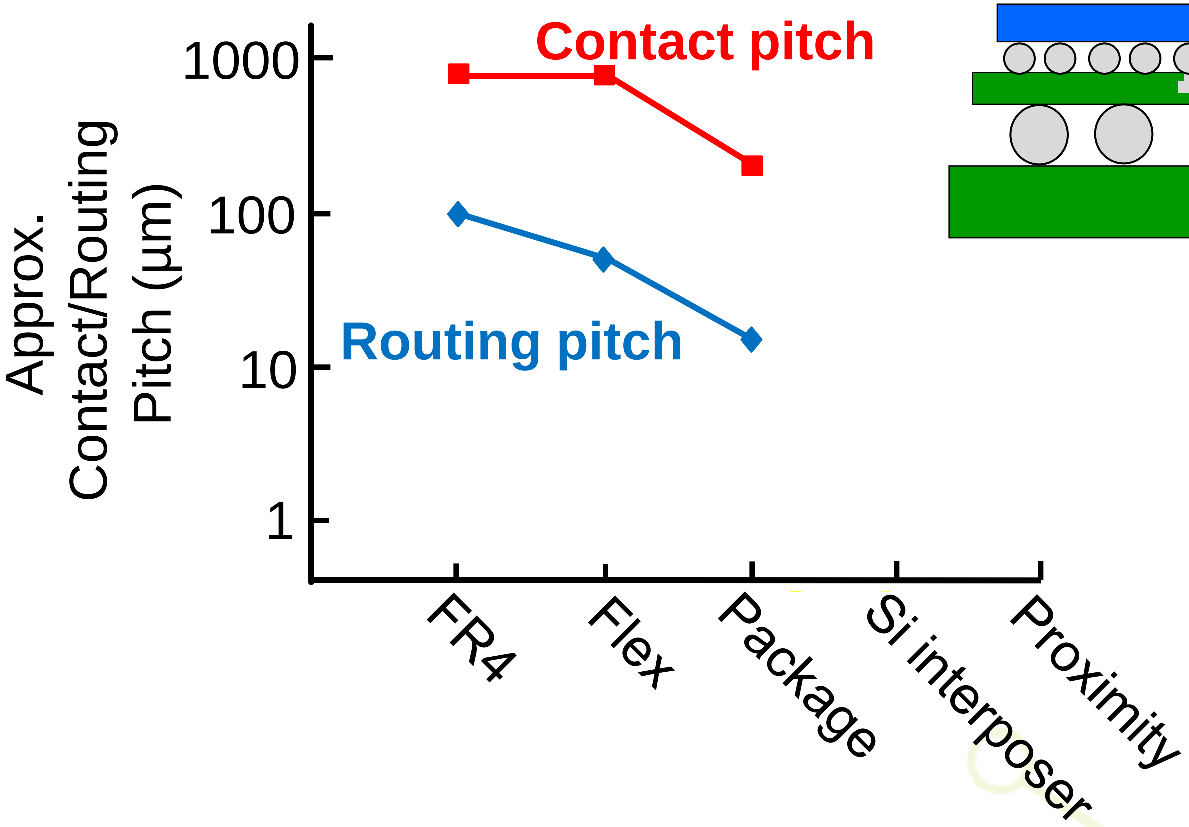
High density channels

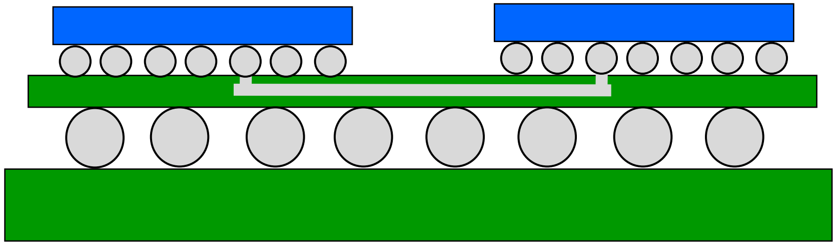


High density channels

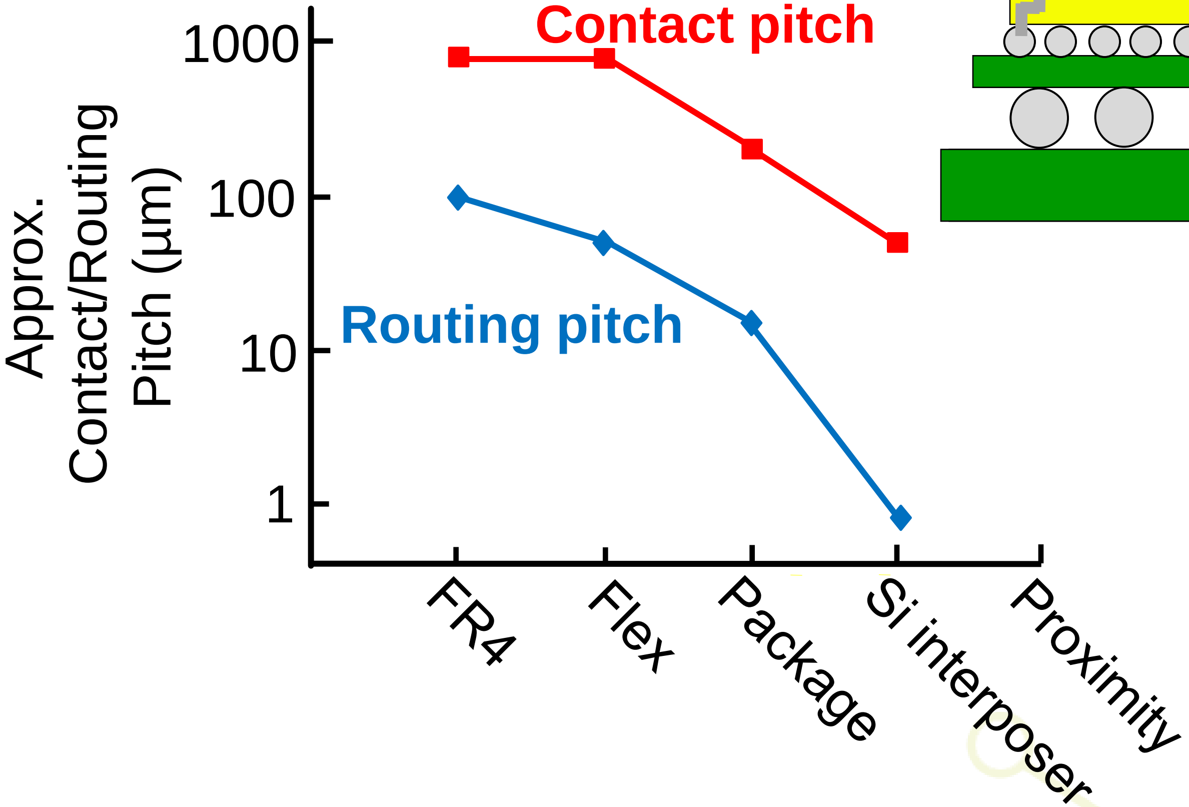


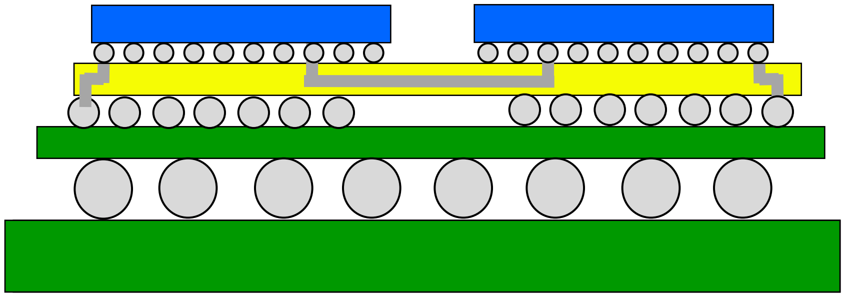
High density channels



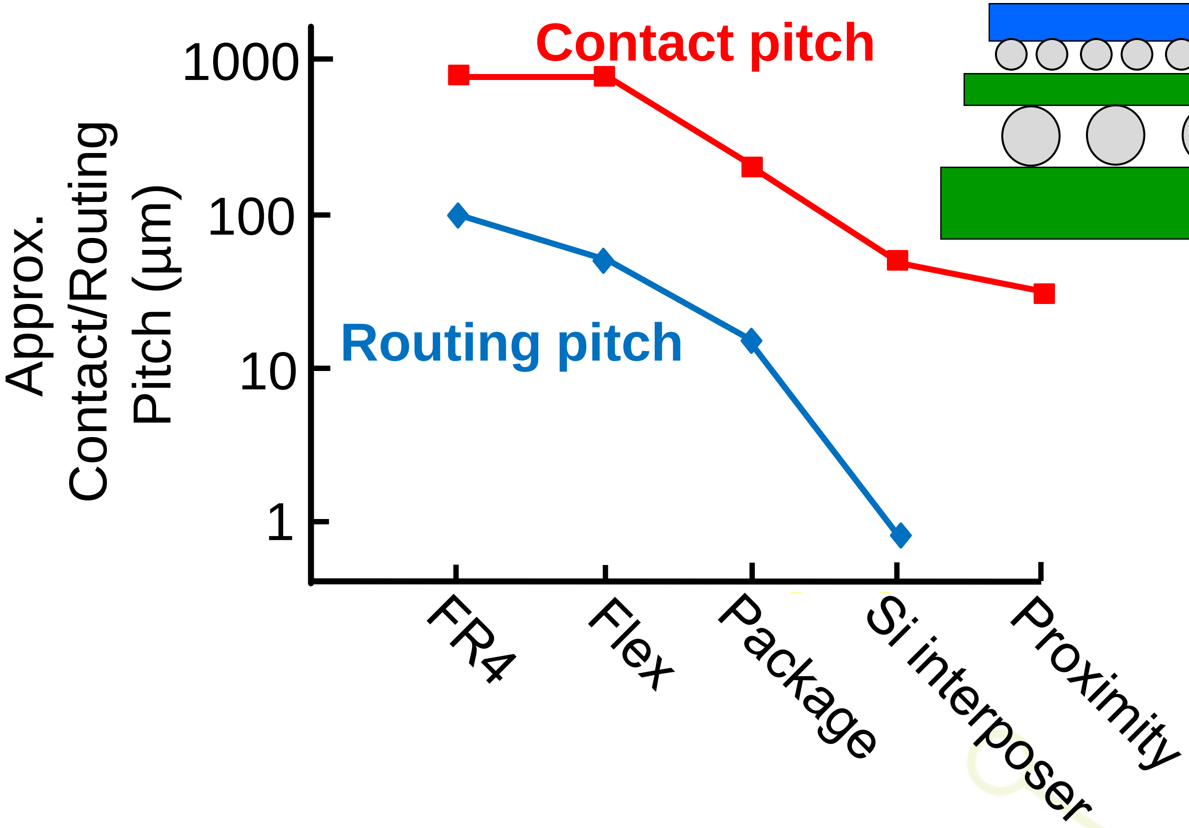


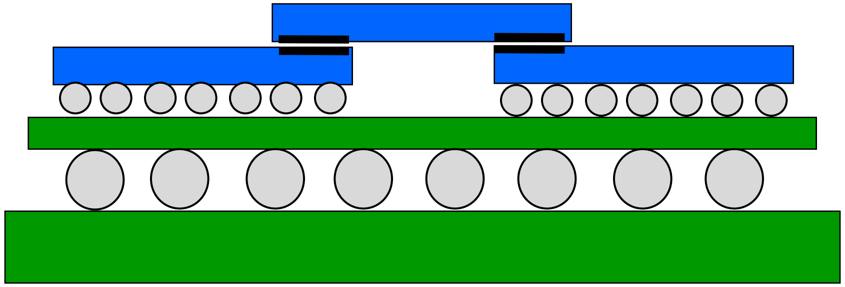
High density channels





High density channels

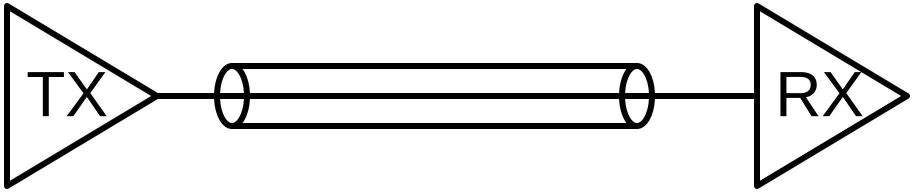




Outline

- 1TByte/s I/O: motivation and challenges
- Circuit Directions
- Channel Directions
- Tool Directions
- 470Gb/s Prototype

What is the “Right” Link Architecture?



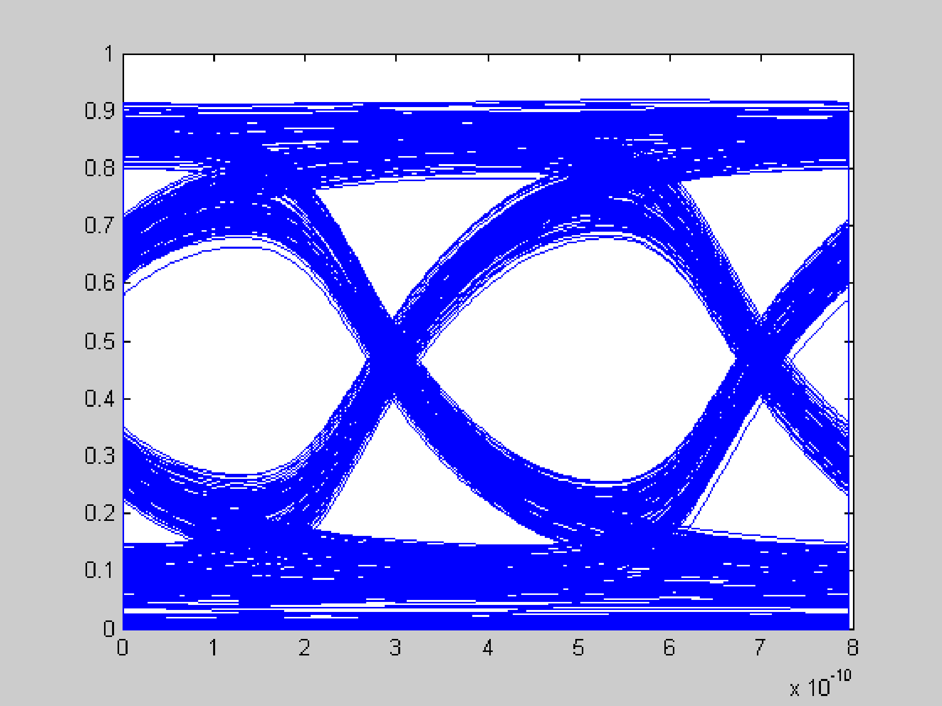
Clock Jitter?
Signal Swing?
Equalization?

ISI? Xtalk?
Modulation (PAM)?
Data Rate?
Interface width?

Clock Jitter?
Sensitivity?
Equalization?

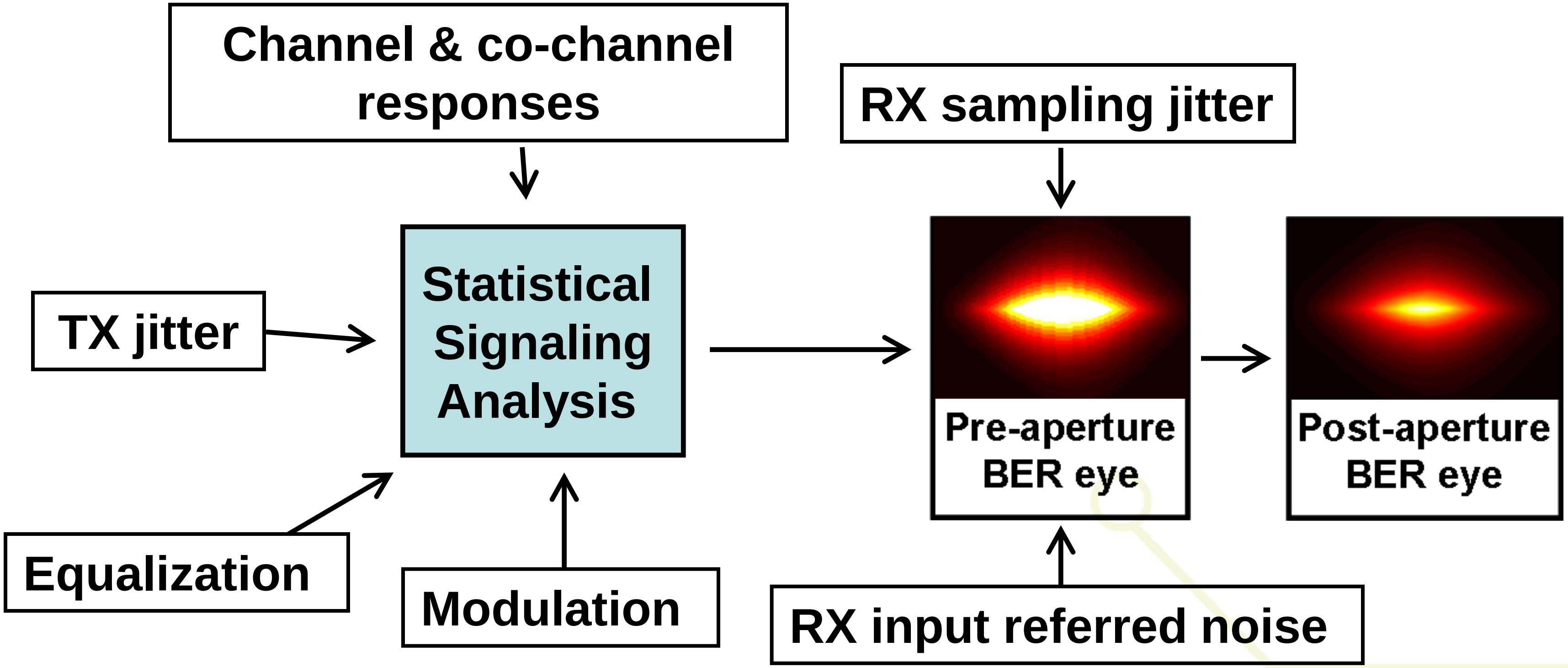
- Designers need the ability to quickly and accurately compare architecture options

Empirical Approach



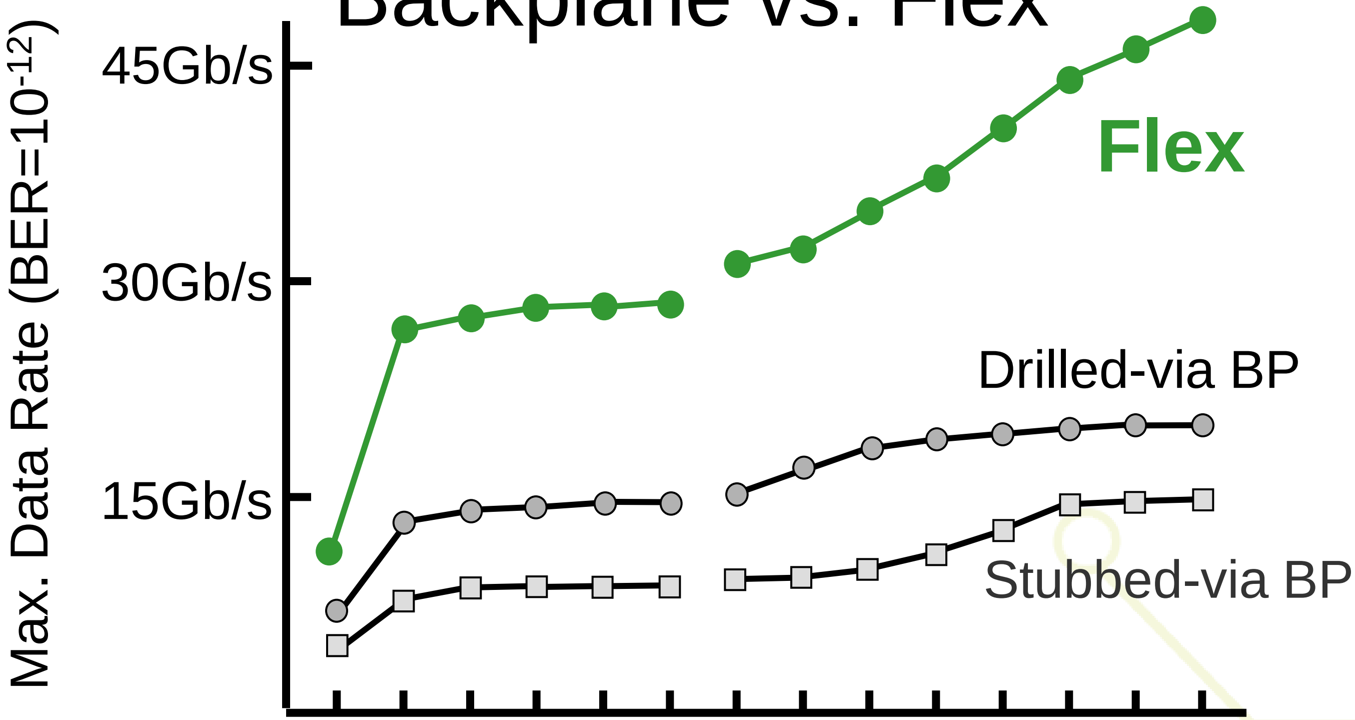
- Simulate system with random data
- This doesn't provide adequate accuracy (BER<10^{-12})

Full System Statistical Analysis



- Specify high-level architecture and block characteristics
- Enables fast evaluation of link sensitivities

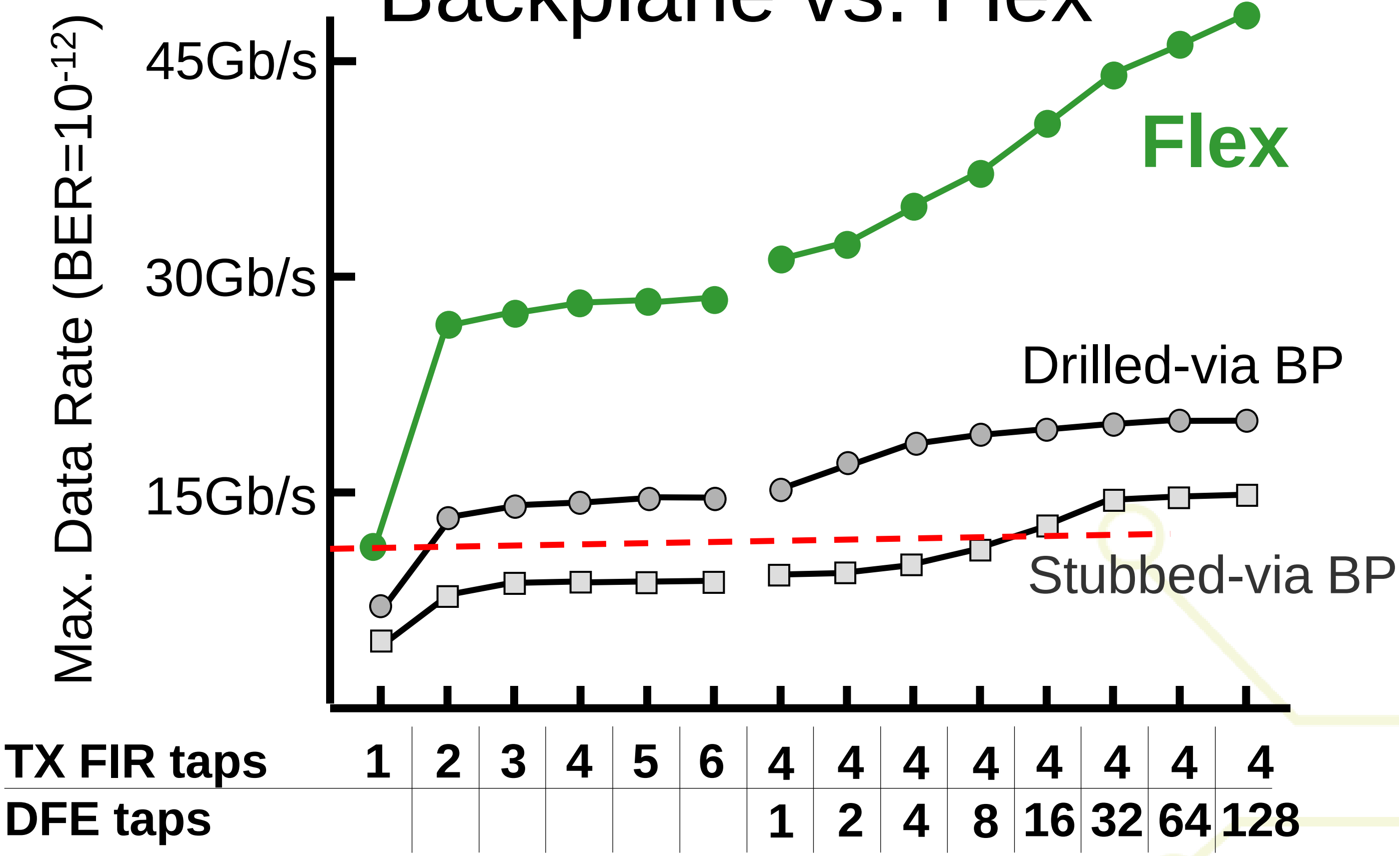
Maximum Data Rate Comparison: Backplane vs. Flex



TX FIR taps	**1**	**2**	**3**	**4**	**5**	**6**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**
DFE taps							**1**	**2**	**4**	**8**	**16**	**32**	**64**	**128**

- Statistical system analysis provides designers with real performance tradeoffs and “brick walls”

Maximum Data Rate Comparison: Backplane vs. Flex



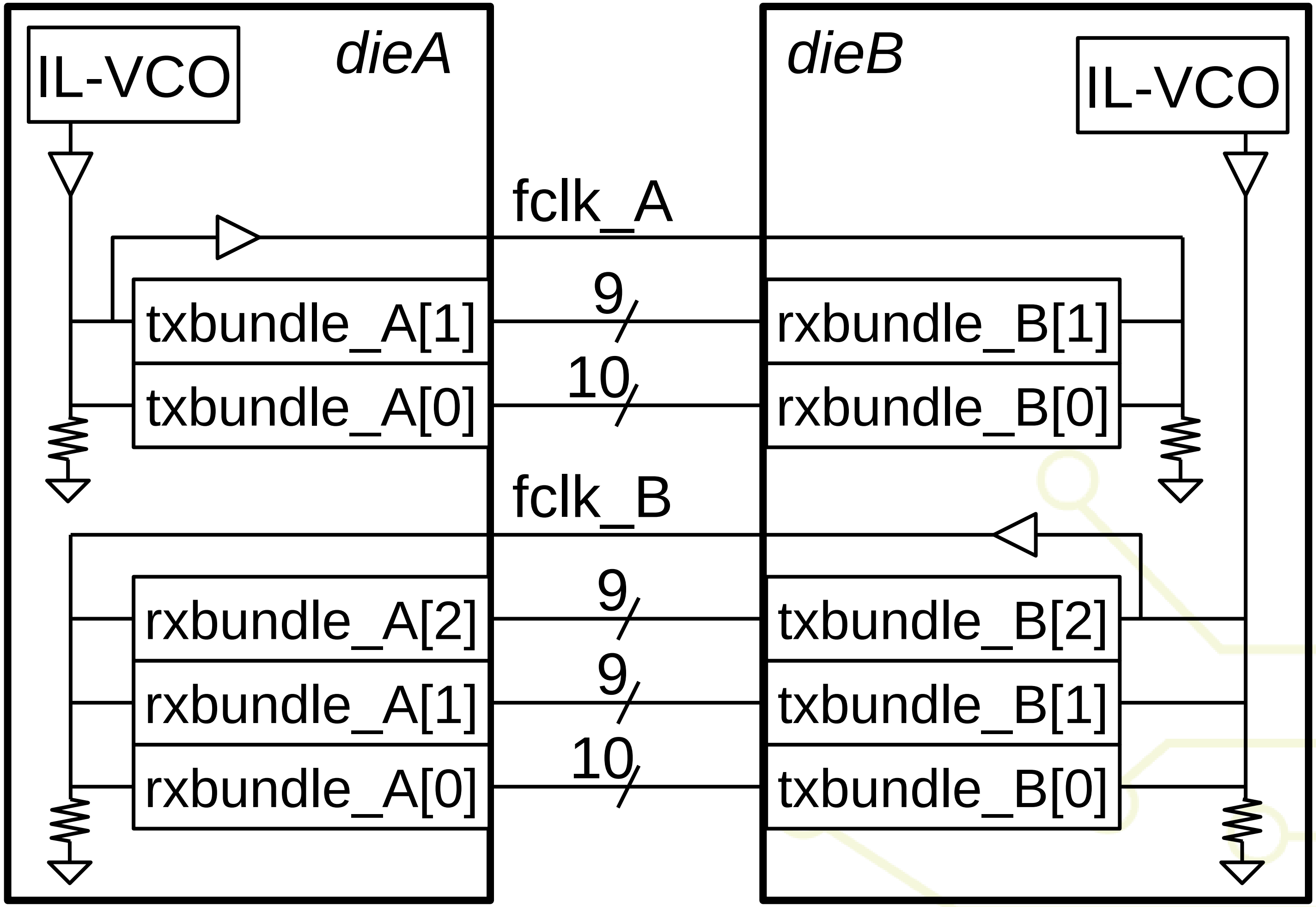
TX FIR taps	**1**	**2**	**3**	**4**	**5**	**6**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**
DFE taps							**1**	**2**	**4**	**8**	**16**	**32**	**64**	**128**

- Statistical system analysis provides designers with real performance tradeoffs and “brick walls”

Outline

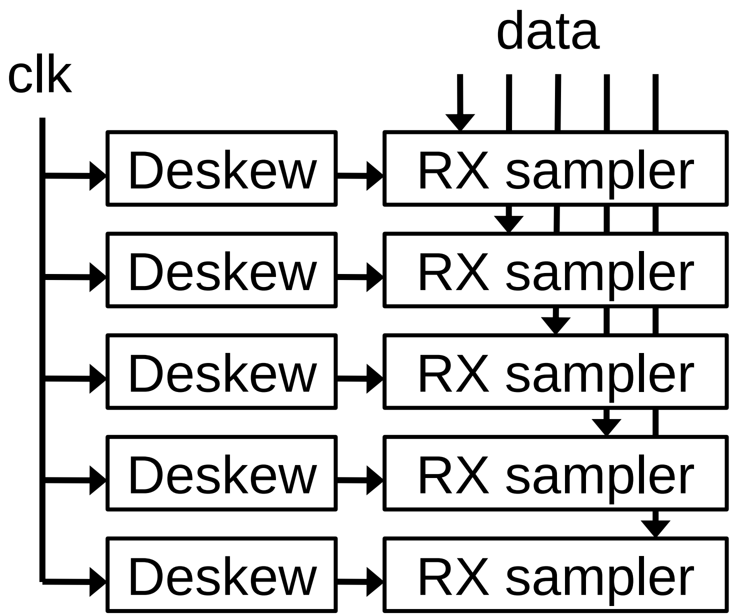
- 1TByte/s I/O: motivation and challenges
- Circuit Directions
- Channel Directions
- Tool Directions
- 470Gb/s Prototype

47x10Gb/s, 1.4pJ/bit Interface (45nm CMOS)



Bundled Architecture

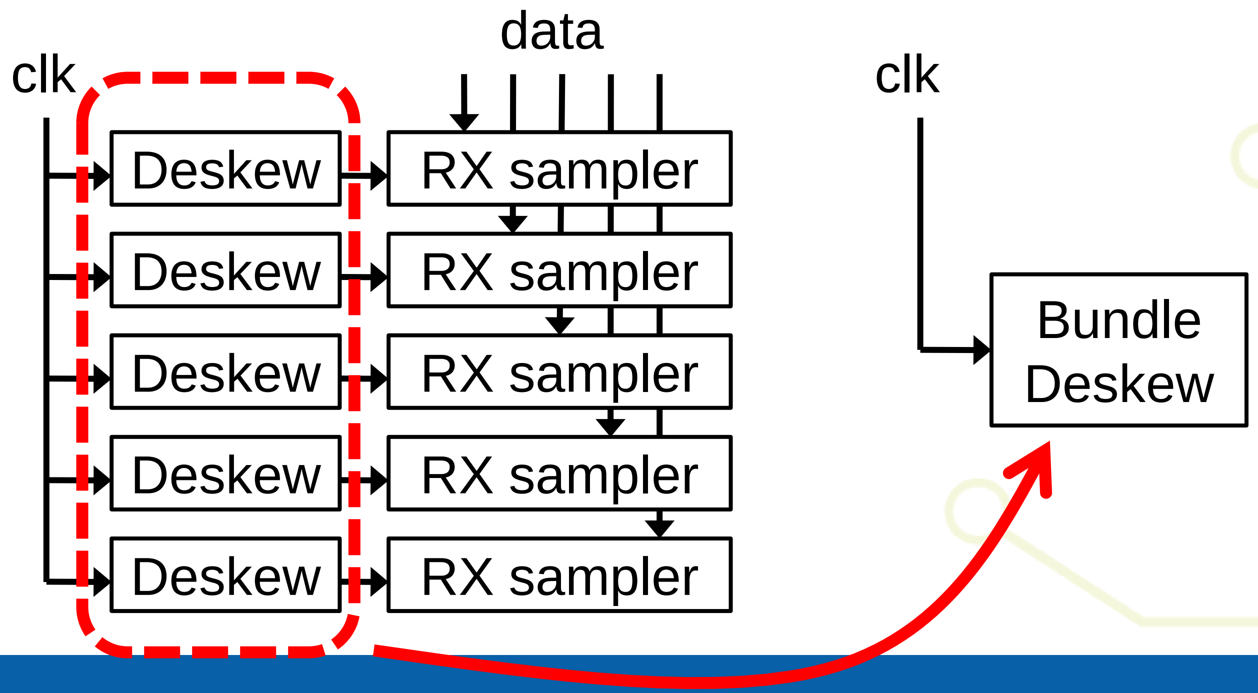
Conventional:
Independent clocking



Bundled Architecture

- **Clocking innovation → Bundle clocking**

Conventional:
Independent clocking

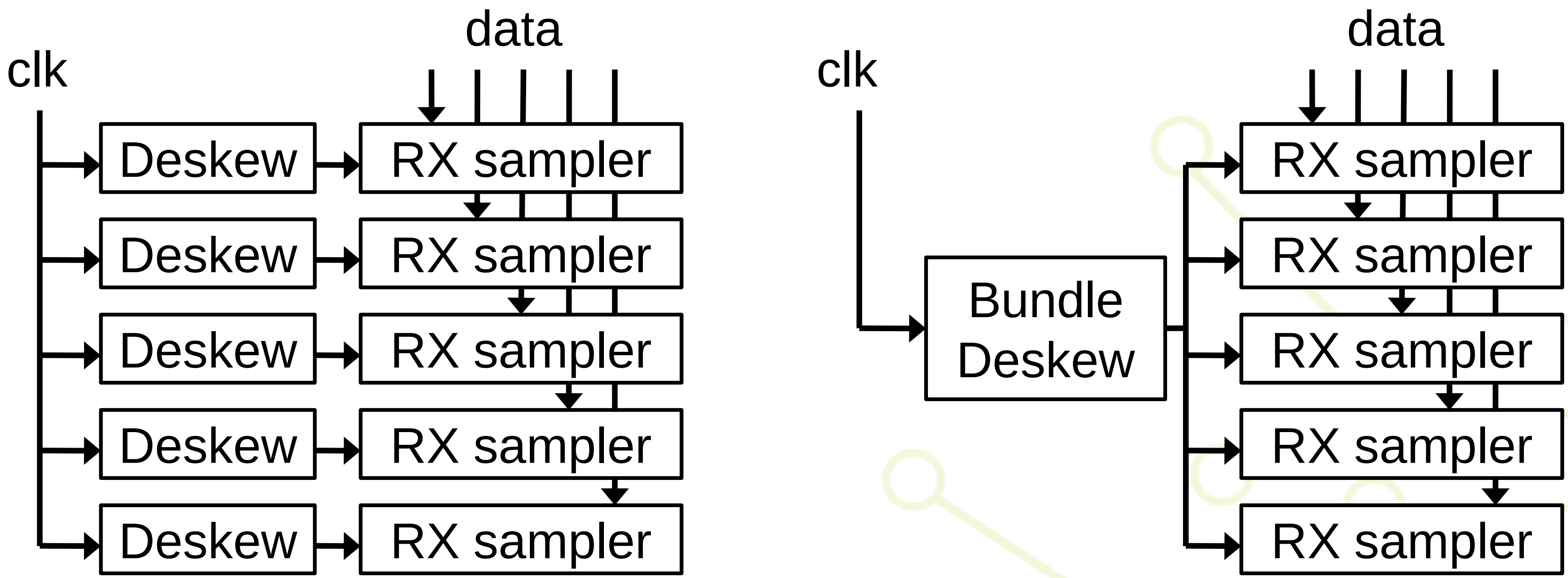


Bundled Architecture

- **Clocking innovation → Bundle clocking**

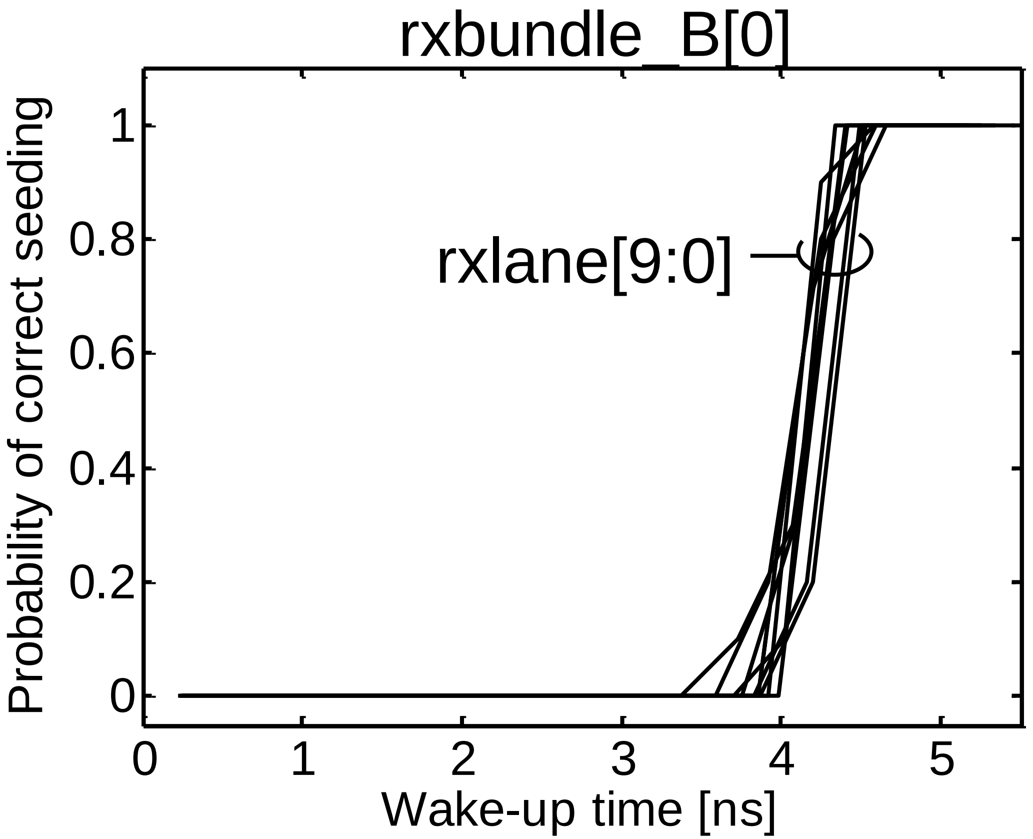
Conventional:
Independent clocking

Optimized:
Bundle clocking



Bundled clocking reduces I/O power

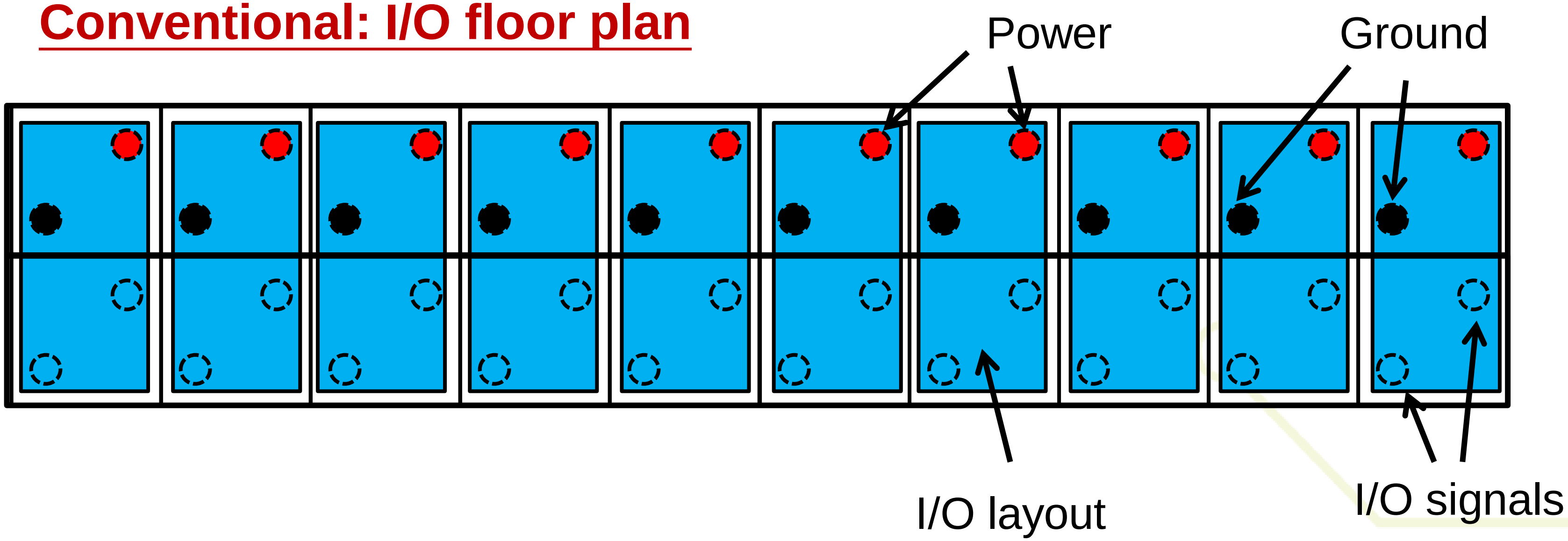
Fast RX Power States



- RX bundle power reduced by 93% in standby
- All RX lanes return to reliable operation in <5ns

Silicon Area Compression

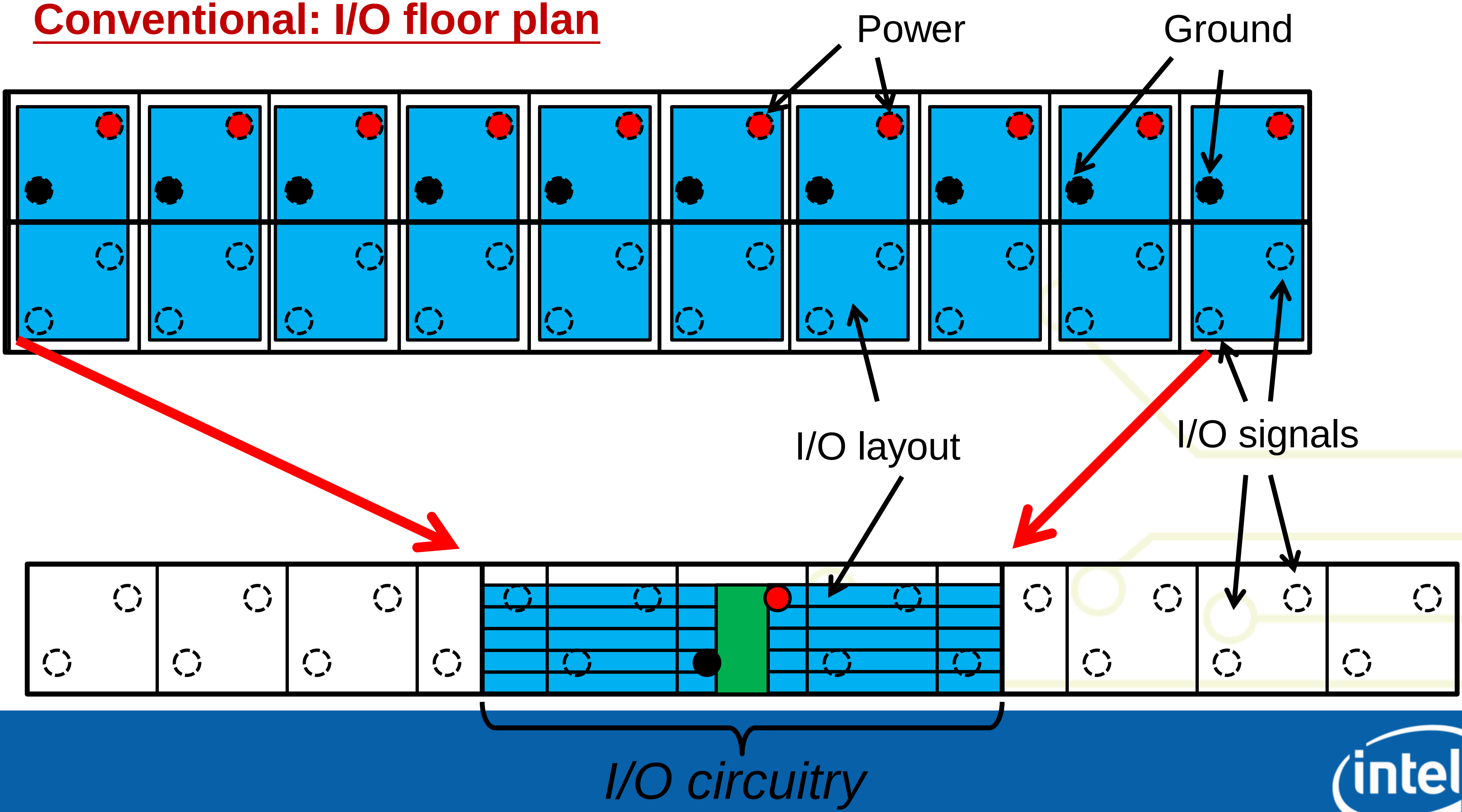
Conventional: I/O floor plan



Silicon Area Compression

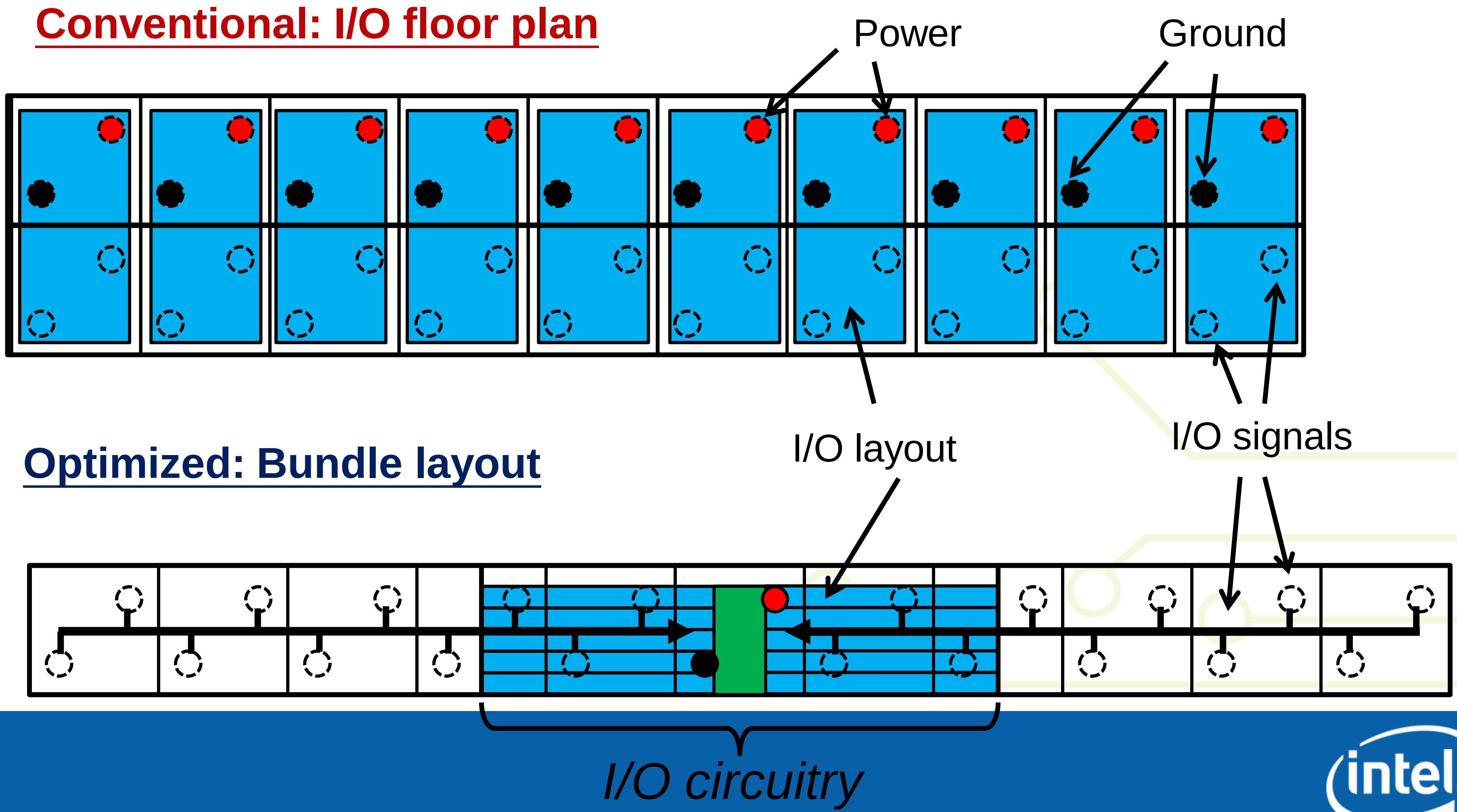
- Floor plan optimization→ minimize I/O area

Conventional: I/O floor plan

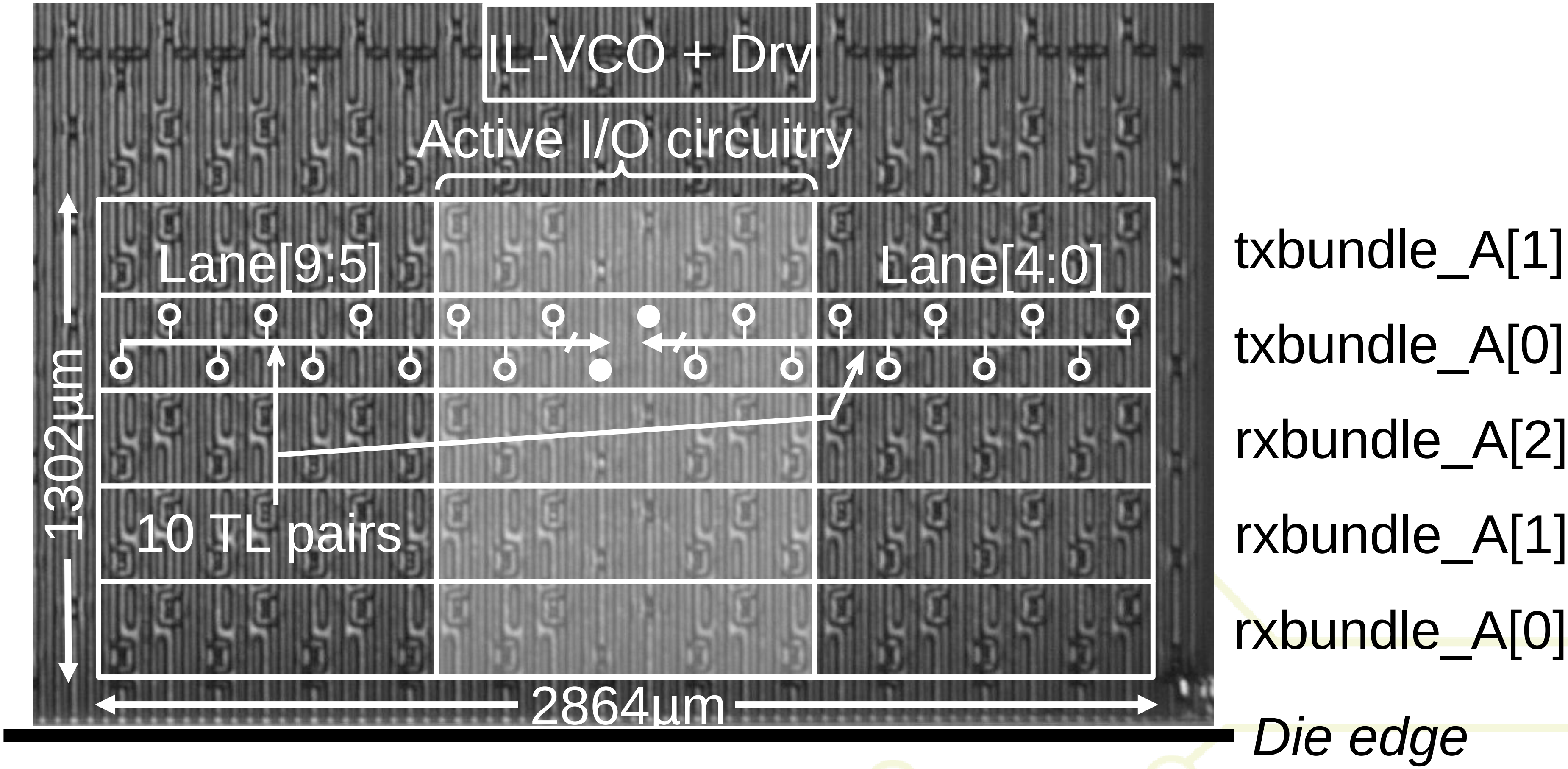


Silicon Area Compression

- Floor plan optimization→ minimize I/O area

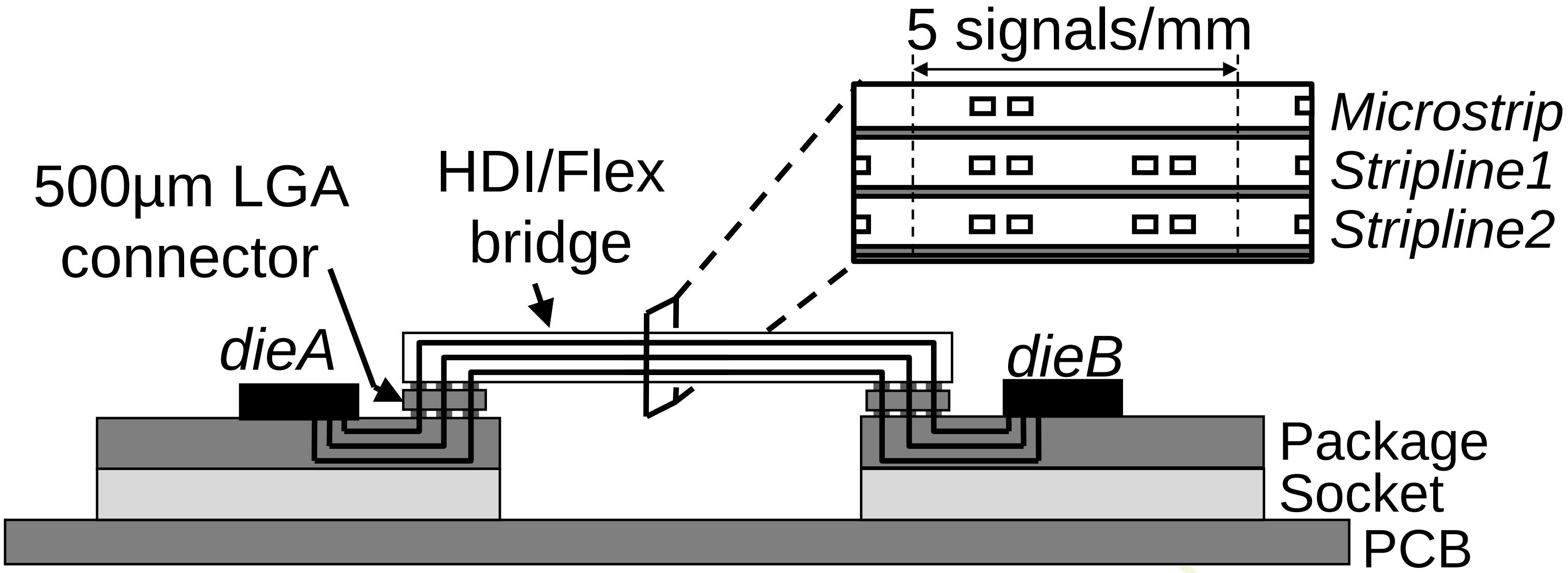


Interface Floorplan



- Active circuit area is reduced with TL routing.

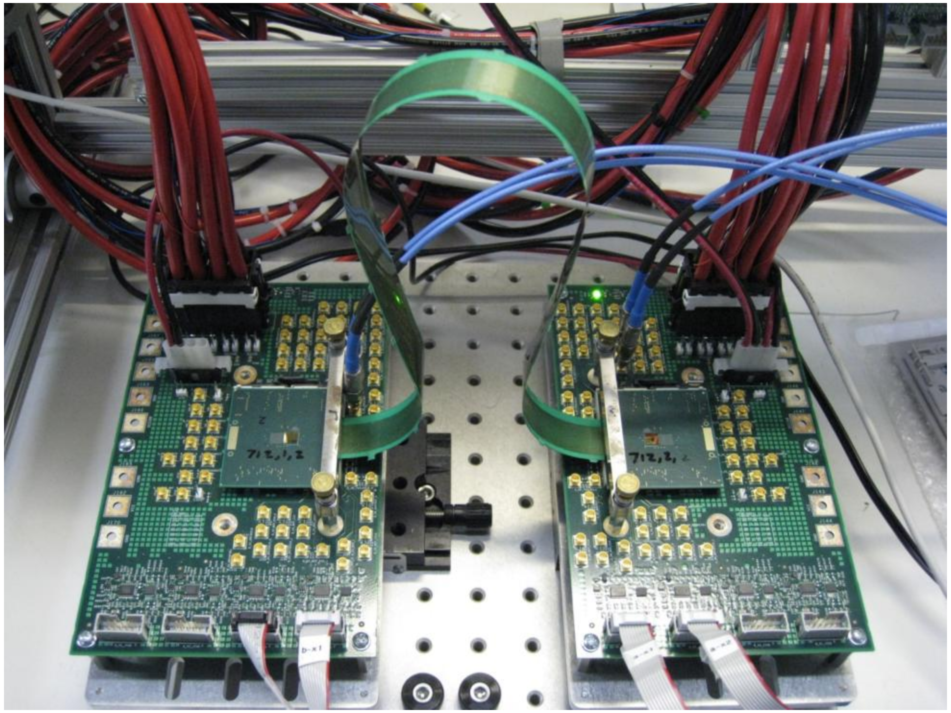
Interface Configuration



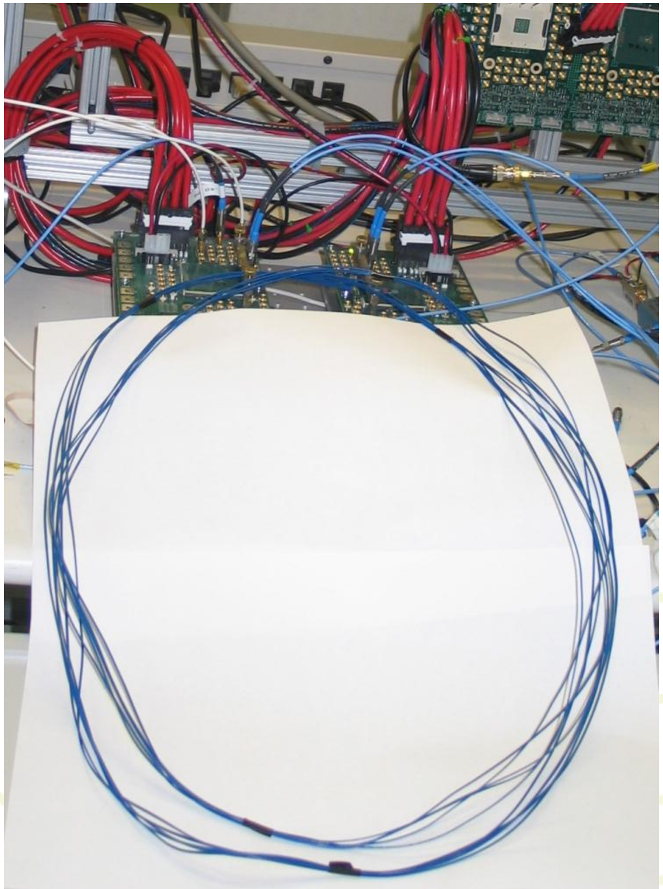
- Within-bundle lanes matched to <100µm
  - Dense LGA connector minimizes breakout area
  - Bundles share the same routing layer
  - 2X density on stripline layers due to reduced Xtalk

Silicon and Interconnect Prototypes

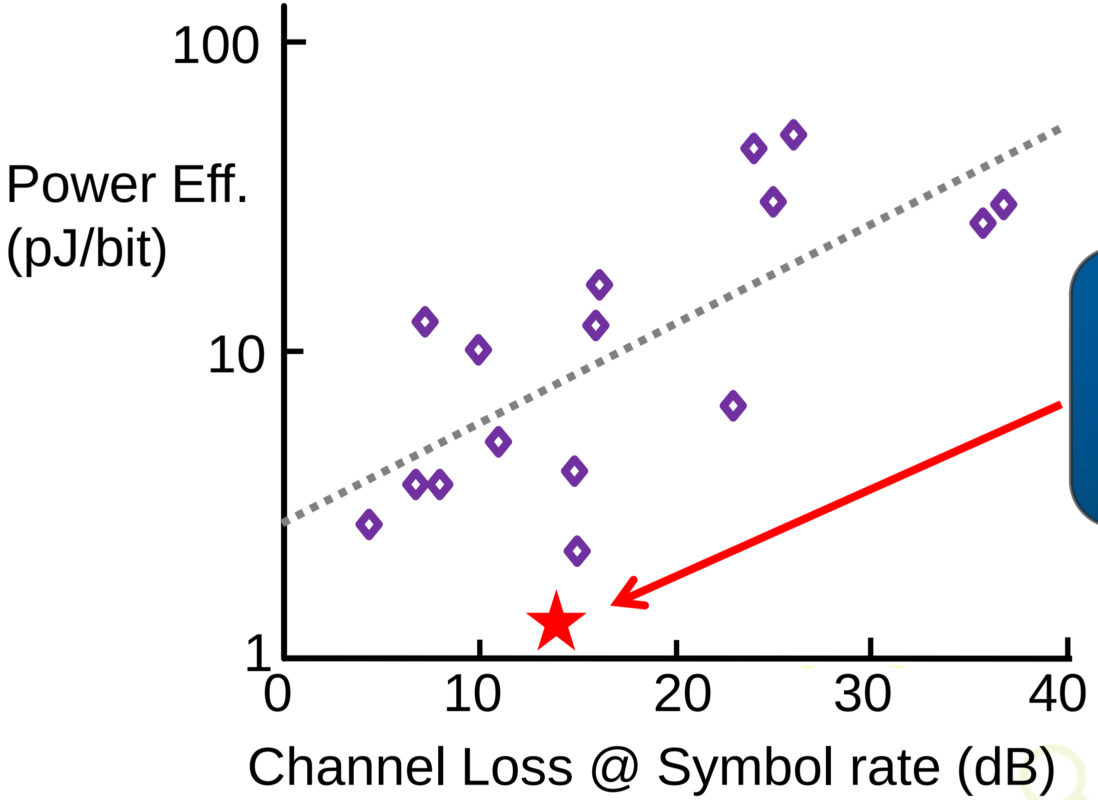
0.5m flex interconnect



3m twinax cable



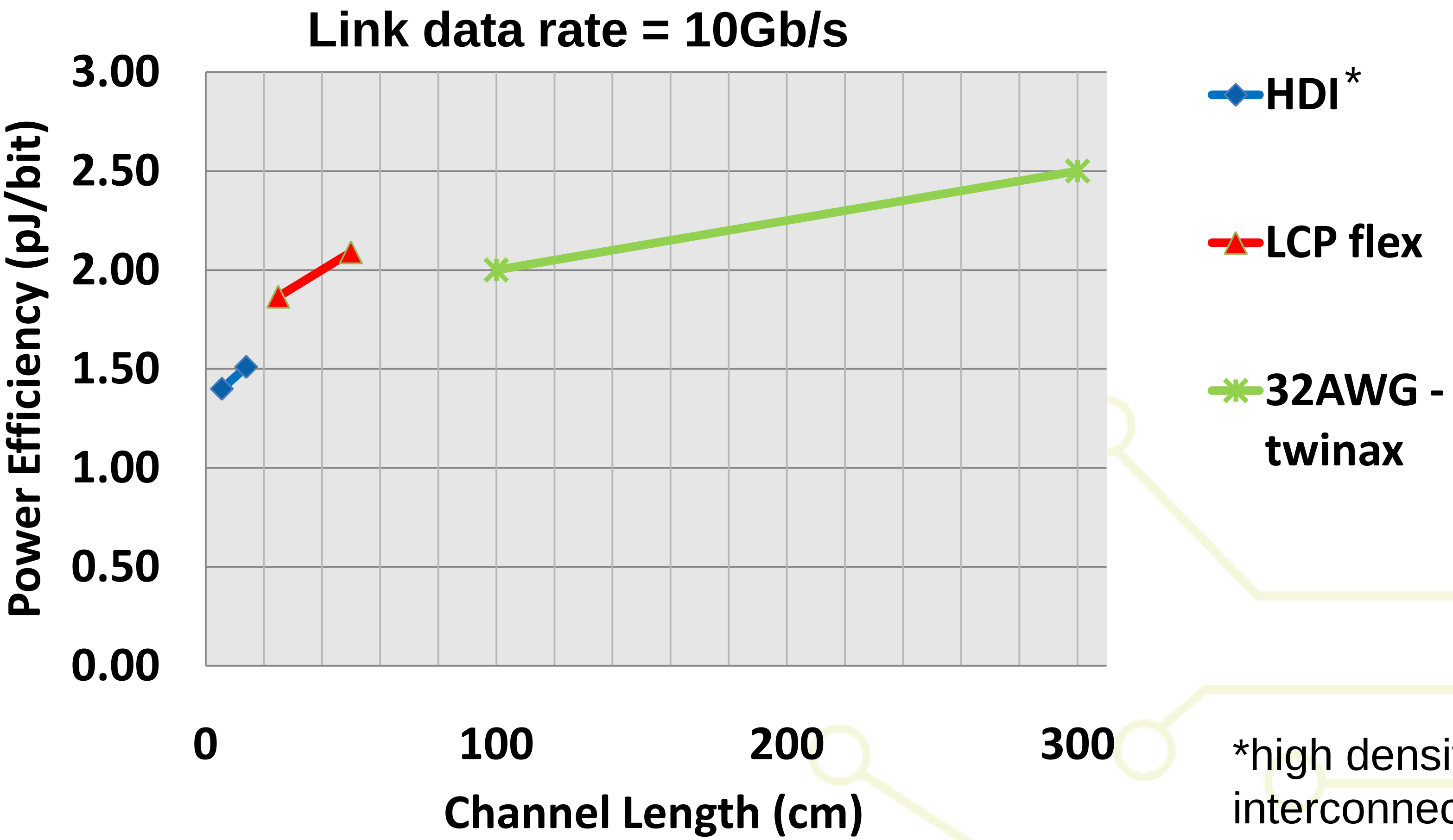
Electrical Interconnect Scaling Challenges



This work: 45nm CMOS Prototype @10Gb/s data rate

(Based on transceivers reported 2006-2009 in 65-130nm CMOS)

I/O Power Efficiency Measurements



*high density interconnect (HDI)

Summary

- Bandwidth needs are quickly approaching 1TB/s
- Extending electrical I/O to 1TB/s requires balance between power, data rate, density and cost
- Evaluate alternate channel configurations and materials
- Recent results indicate that electrical will be up to the task for “in-box” I/O