

# Estimating Reliability and Throughput of Source-synchronous Wave-pipelined Interconnect

Paul Teehan, Mark Greenstreet, Guy Lemieux

University of British Columbia

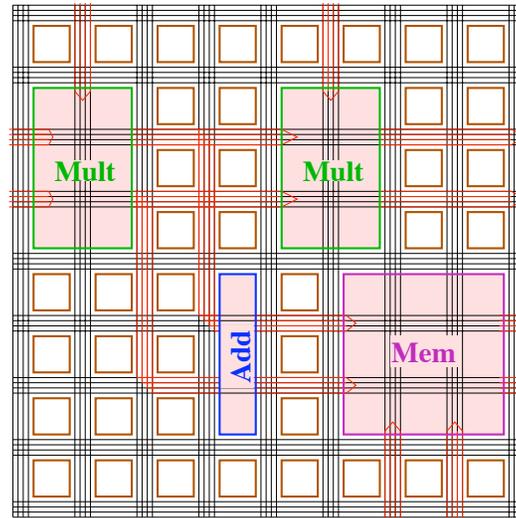
# Overview

---

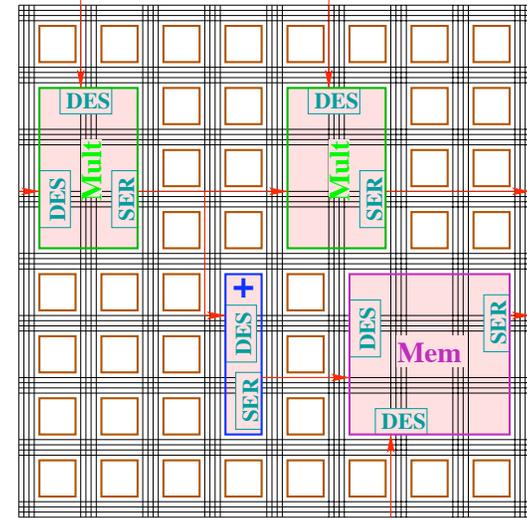
- NoCs motivate using highly optimized interconnect.
  - It's **easy** to make a link that has great performance in SPICE.
  - It's **harder** to get an acceptable bit-error-rate (BER).
- Real designs need BERs of  $10^{-20}$  or less.
  - Can't establish this with SPICE simulations.
  - Need to use statistical methods.
- We present a methodology for assessing the BERs of NoC interconnect.
  - Compare wave-pipelined and latch pipelined interconnect as a working example.

# Motivation

Bit-serial links for reconfigurable computing:



Today's FPGA

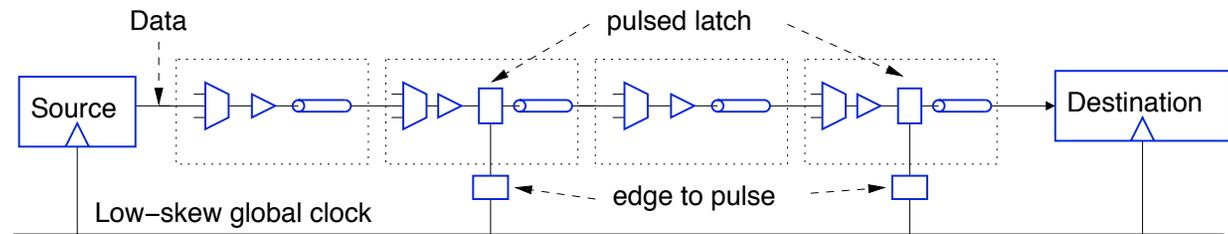


FPGA with serial interconnect

- Word-oriented blocks cause routing congestion for a traditional FPGA.
- FPGA clock periods are typically  $>100$  gate delays.
- Wires can be pipelined with  $\sim 10$  gate delay period.
- Bit-serial interconnect can alleviate routing congestion

# Bit-Serial Interconnects

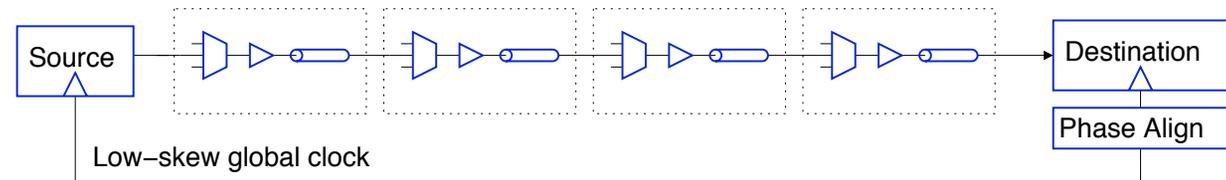
## Globally-Synchronous Interconnect:



- 😊 Standard, synchronous design.
- 😞 Needs high-speed, global clock that is unused by the rest of the system
- 😞 High-power consumption, even with clock gating.
- 😞 Complicates timing closure.

# Bit-Serial Interconnects

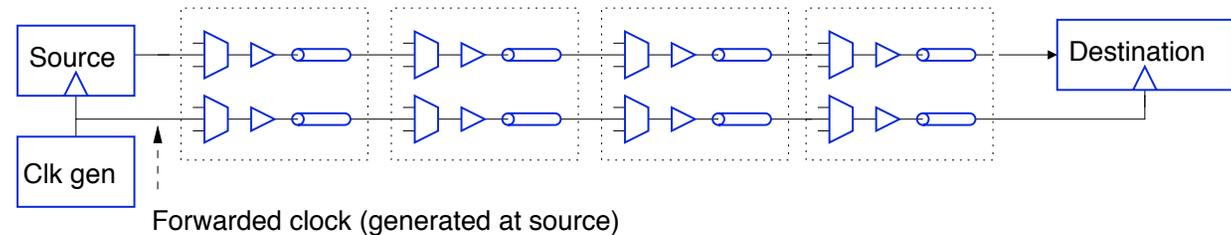
Wave Pipelined, No Latches:



- 😊 Minimal hardware.
- 😞 Requires close matching of **min** and **max** delays.
- 😞 Phase-alignment increases complexity and power consumption of receiver circuit.

# Bit-Serial Interconnects

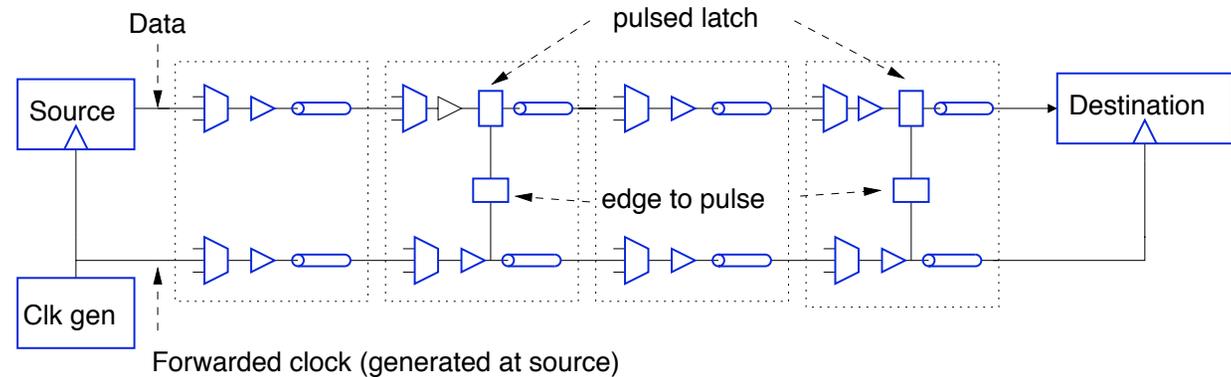
## Source-Synchronous, Wave Pipelined



- ☹ Strobe eliminates need for clock-phase recovery at receiver.
- ☹ Extra power and area for strobe.
- ☹ Need to match delays of data and strobe paths.
- ☹ Strobe pulses may be dropped due to jitter.

# Bit-Serial Interconnects

## Source-Synchronous, Latched, Wave Pipelined



- 😊 Periodic latching keeps data aligned with strobe.
- 😞 Strobe timing susceptible to jitter.
- 😞 Edge-to-pulse converters add circuitry and contribute to jitter between strobe and data.

# Timing uncertainty is the problem

---

- Source-synchronous is advantageous for on-chip, high-speed, serial communication.
  - No need for a global, high-speed clock.
  - Strobe only sent with actual data transfers – saves more power.
- Need to keep data and strobe aligned.
  - Requires analysis of delay variations in both paths.
- Worst-case timing analysis is overly pessimistic.
  - $6\sigma$  jitter at **every** stage for a particular data bit is extremely unlikely.
  - Need a statistical approach.

# Outline

---

## Reliability and throughput estimation

- Motivation
- Timing uncertainty
  - Static vs. dynamic timing variation.
  - Inter-Symbol Interference (ISI)
  - Crosstalk
  - Power supply noise
- Statistical timing analysis for source-synchronous communication.
- Examples: analysis of bit-serial links in 65nm CMOS
- Conclusions and future work.

# Static vs. dynamic timing variation

---

- Static variation
  - Die-to-die, cross-chip, and device-to-device parameter variation.
  - Other variations that change slowly wrt. bit period
    - sub-GHz power supply noise
    - temperature variation
- Dynamic variation:
  - Crosstalk
  - ISI
  - Power-supply noise
  - Anything on a time scale of one to a few bits.

# Dynamic Uncertainty is the Problem

---

- Need to keep strobe and data aligned:
  - This is determined by timing variation that affects a single {strobe, data-bit} pair.
- Need to preserve pulses on strobe and data lines.
  - This is determined by timing variation between consecutive transitions on the strobe path or data path.
- Prior work:
  - Many proposals for on-chip, serial interconnect: [Ou 2004], [Zhang 2005a], [Dobkin 2007], [Joshi 2007].
  - [Zhang 2005b] applied statistical timing analysis to globally clocked, pipelined interconnect.
  - **Our contribution:** A systematic application of statistical timing methods to source-synchronous, on-chip interconnect.

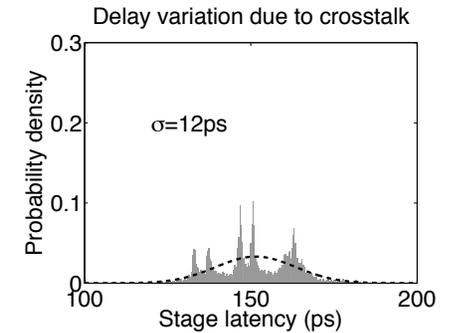
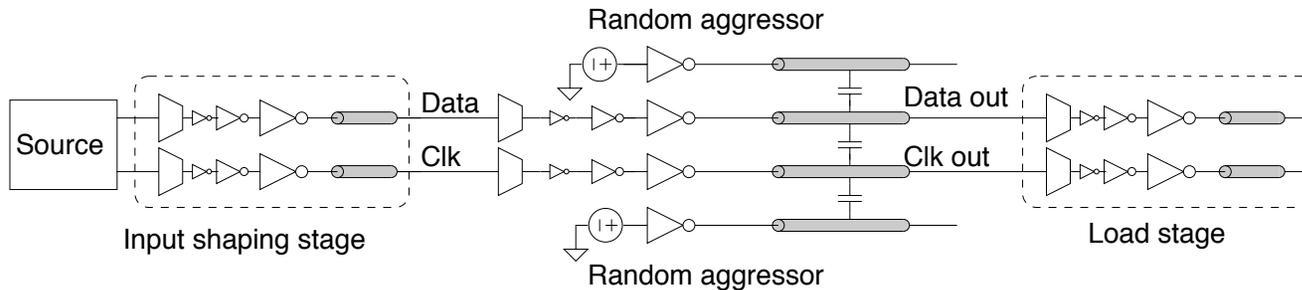
# Inter-Symbol Interference (ISI)

---

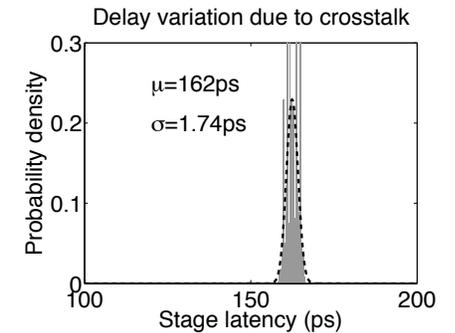
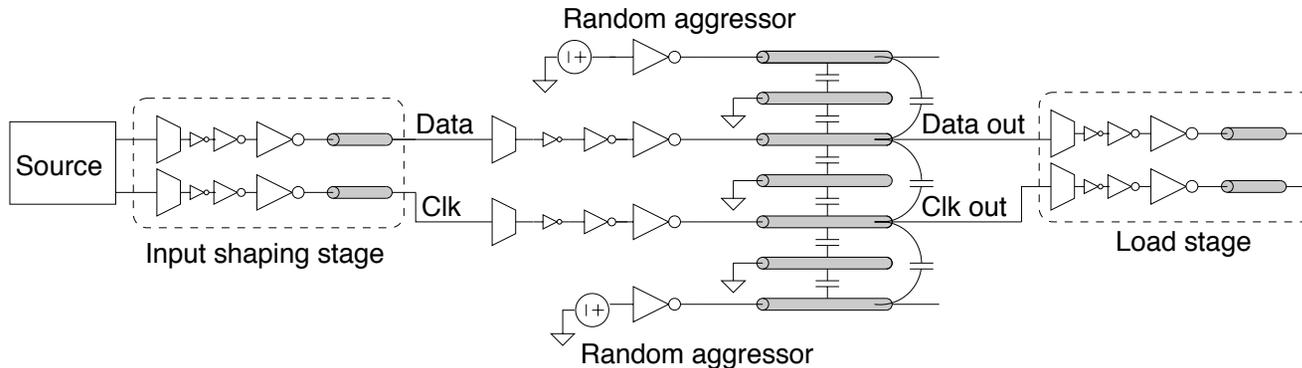
- Buffers propagate the trailing edge of a short pulse with less delay than that of a long pulse.
  - This is because the short pulse doesn't swing the wire all the way to the power supply rail.
  - Thus, the trailing edge of a short-pulse gets a head start.
  - If the minimum pulse width is  $> 10$  gate delays, then this is not a serious problem: all transitions make it very close to the rail.
- Wires propagate high-frequency signals faster than low-frequency ones.
  - RC delay dominates for low-frequency components of signal.
  - LC delay dominates for high-frequency components of signal.
  - In practice, ISI is not a serious problem for practical wire lengths with data rates  $< 10$  Gbps.
- Conclusion:
  - In 65nm, with bit-periods of 10-20 gate delays, ISI is not a serious problem.
  - Care needed for bit-periods  $< 10$  gate delays.
  - ISI will become more important for smaller processes.

# Crosstalk

## Interconnect without shielding



## Interconnect with shielding



Conclusion: Crosstalk effects are relatively small ( $\sim 0.1$  gate delay) if wires are shielded.

# Power Supply Noise

---

- High-frequency noise:
  - Impacts source synchronous interconnect.
  - Detailed models unavailable
    - includes clock network component
    - includes logic switching component
- Low-frequency noise:
  - Less critical for source synchronous interconnect.
  - Arises from ringing of off-chip inductance and on-chip capacitance.
- For our FPGA example:
  - Clock is relatively low frequency.
  - Serial transfer initiated after active clock edge.
  - Main source of noise is ongoing logic switching.

# Summary of Timing Uncertainty

---

- Inter-Symbol Interference (ISI)
  - Small impact for bit periods greater than 10 gate delays and 100ps.
  - Likely to become more significant in sub 65nm designs.
- Crosstalk
  - Shielding required for high-speed links.
- Power-supply noise
  - Main concern for serial interconnect.
  - Better models needed.

# Outline

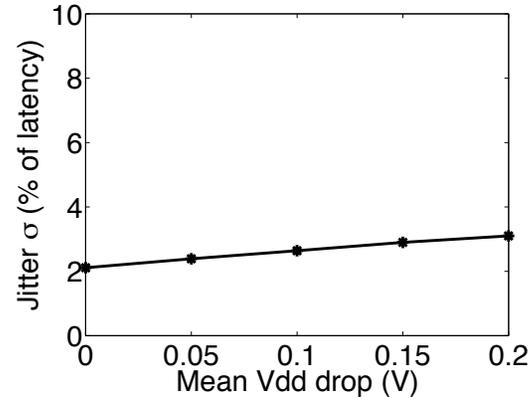
---

## Reliability and throughput estimation

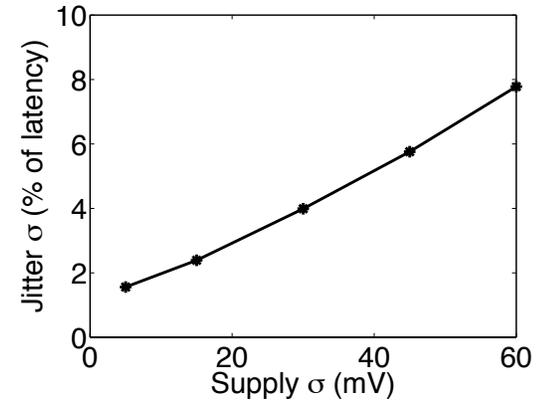
- Motivation
- Timing uncertainty
- Statistical timing analysis for source-synchronous communication.
  - Statistical modeling of timing variation
  - Identifying failure modes
  - Computing bit-error rates
- Examples: analysis of bit-serial links in 65nm CMOS
- Conclusions and future work.

# Statistical modeling of timing variation

Focus on power supply noise.



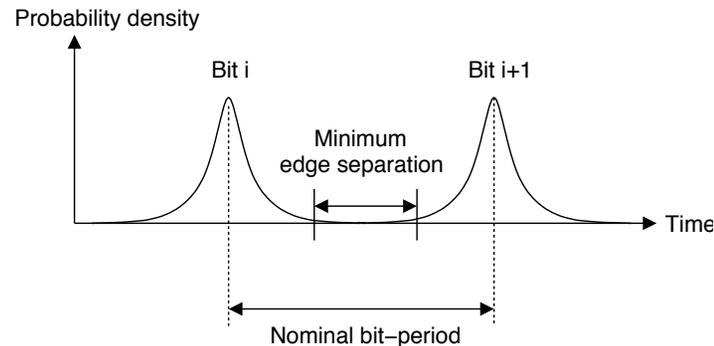
Response to nominal supply value,  $\mu$ , ( $\sigma = 15\text{mV}$ )



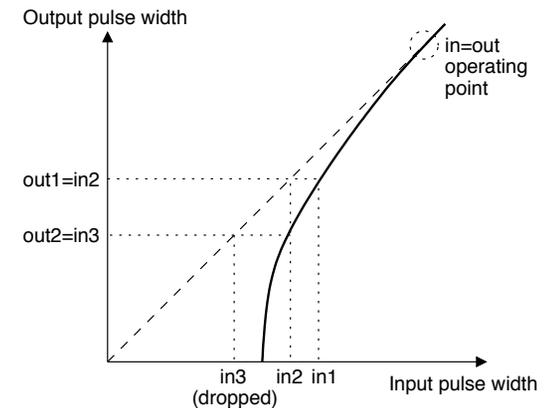
Response to transient noise,  $\sigma$ , ( $\mu = 0.95\text{V}$ )

- Residual jitter (at no noise) of slightly less than 2% due to HSPICE numerical noise.
- Jitter has slight sensitivity to  $V_{dd}$  drop.
- Jitter roughly proportional to transient noise.

# Identifying failure modes



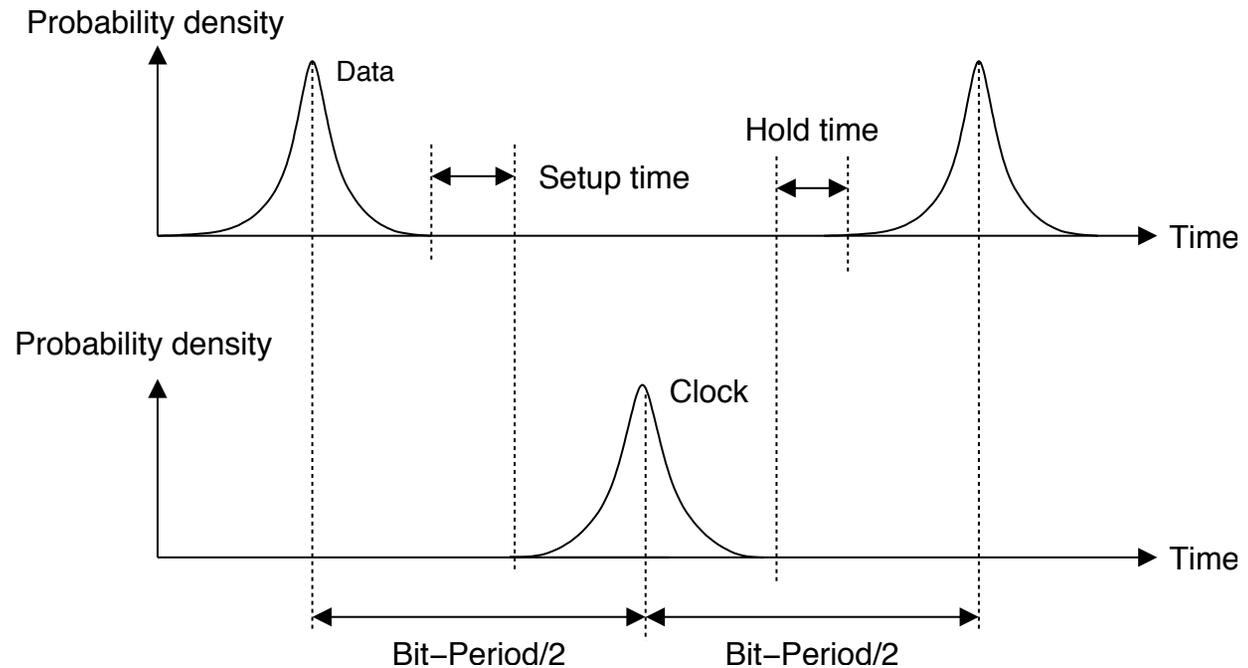
Edge Separation



Pulse-width transfer function.

- Loss of strobe pulses
  - Minimum pulse width is point where ISI becomes dominant and leads to loss of the pulse.

# Identifying failure modes



- Set-up or hold failures
  - Set-up failure: data arrives too late relative to the strobe.
  - Hold failure: data arrives too early relative to the strobe.

# Computing bit-error rates

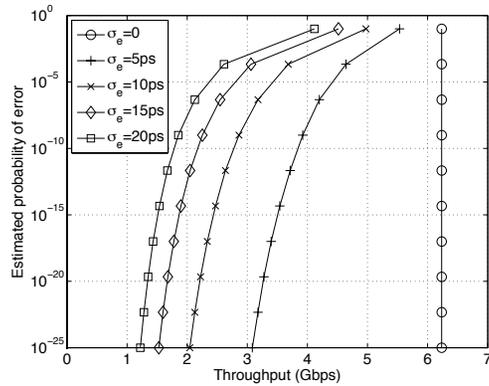
---

- Determine failure limits (HSPICE)
- Estimate per-hop jitter statistics (HSPICE)
- Compute total link jitter statistics
  - Assume independent jitter at each hop (because we don't have better models):

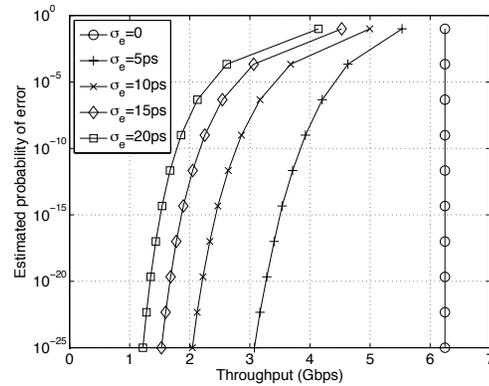
$$\sigma_{total} = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}$$

- Determine failure probability

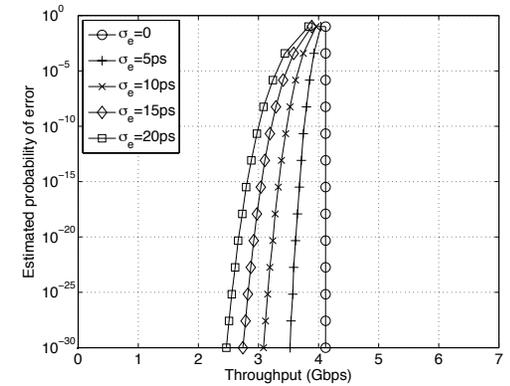
# Results



Wave-pipelined, no latches



Wave-pipelined, latch every 5<sup>th</sup> stage



Globally synchronous, latch pipelined

# Summary

---

## ● Conclusions

- Timing uncertainty must be accounted for in NoC high-bandwidth interconnect.
- Statistical methods necessary to validate BERs required for practical interconnect.
- Not all pipelining methods degrade equally.

## ● Future work

- Model other signaling methods.
- Better power supply noise models:
  - account for link latencies greater than one clock period
  - account for spatial correlations
- Apply to applications other than FPGAs.

# Summary

---

- Conclusions
  - Timing uncertainty must be accounted for in NoC high-bandwidth interconnect.
  - Statistical methods necessary to validate BERs required for practical interconnect.
  - Not all pipelining methods degrade equally.
- Future work
  - Model other signaling methods.
  - Better power supply noise models:
    - account for link latencies greater than one clock period
    - account for spatial correlations
  - Apply to applications other than FPGAs.
- Questions?