Imperial College London

On-silicon Instrumentation

An approach to alleviate the variability problem

Peter Y. K. Cheung
Department of Electrical and Electronic Engineering

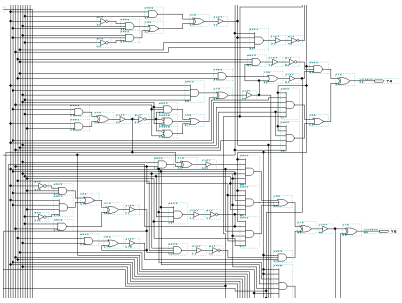
18th March 2014 – U. of York

…… How we started (in 2006) ……

- **Process variability – hot issue at the time**
- **The curses of FPGAs**
  - Used for ANY design, assume worst case in everything
- **The blessings of FPGAs**
  - Self-test is almost free (bitstream storage & time)
  - Ability to reconfigure

– **The opportunity**: **LATE BINDING**

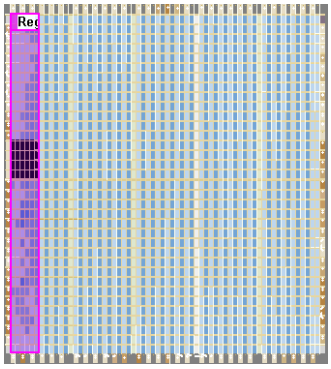
What is Conventional Binding?



logical view

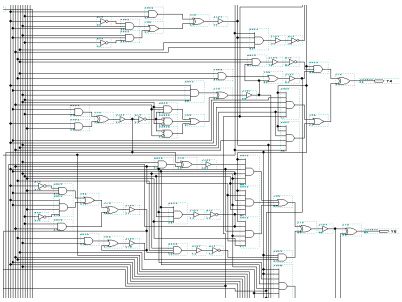
Current

Performed once for ALL chips at Place-and-Route



physical view

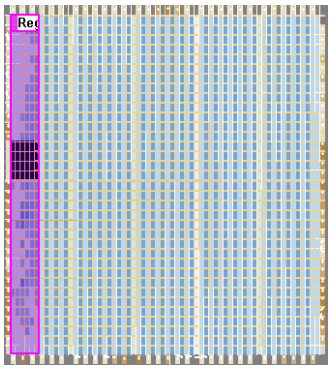
What is Late Binding?



logical view

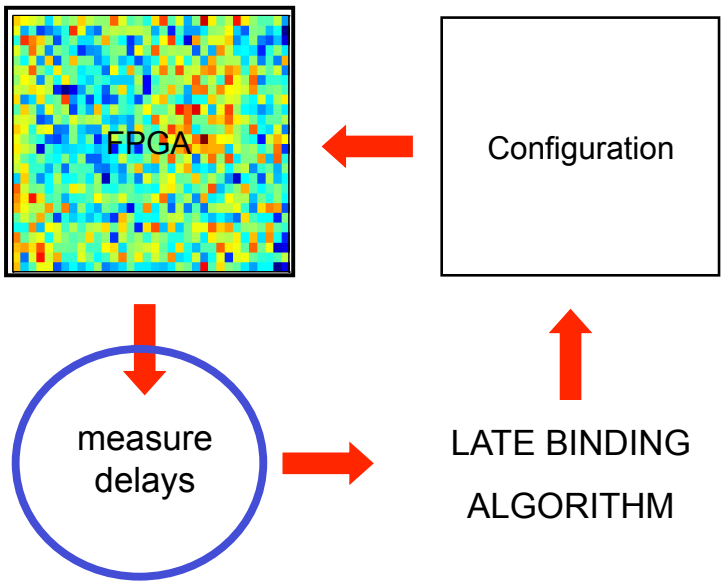
LATE BINDING

Performed part of this Mapping **AS LATE AS POSSIBLE** for **EACH** chip based on its **individual characteristics**

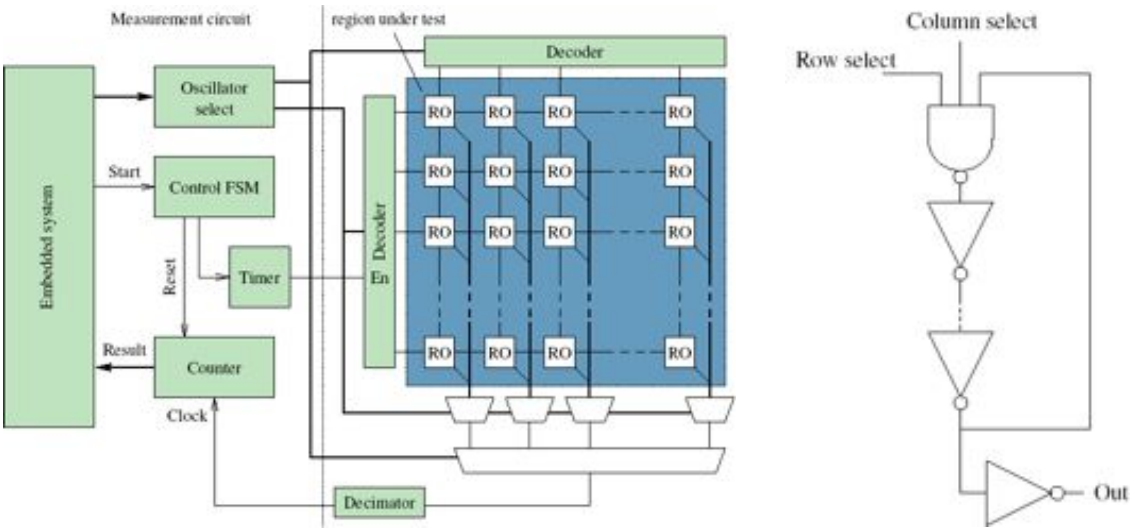


physical view

Late Binding



Instrument 1: Ring Oscillators



Application of Instrument 1: Investigate process variability in FPGAs

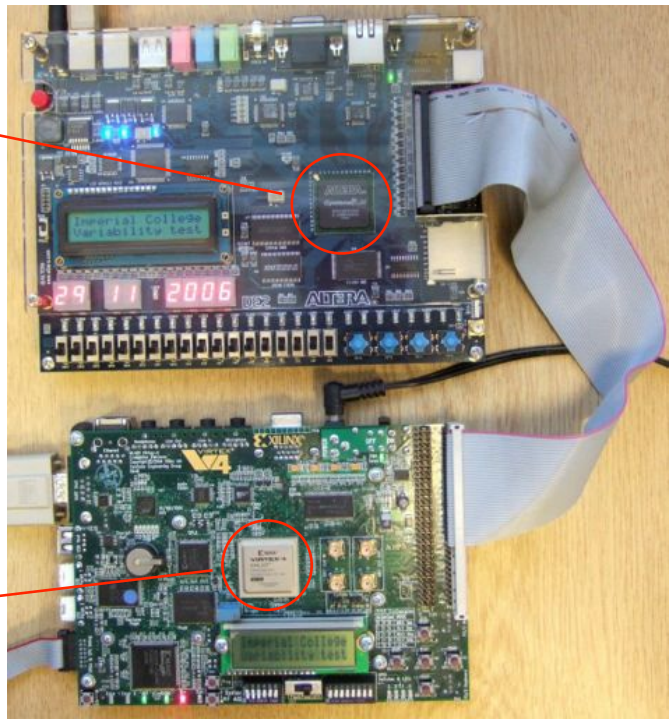
How bad is stochastic variation as compared with systematic variation for 90nm?

Xilinx – Altera interoperability!

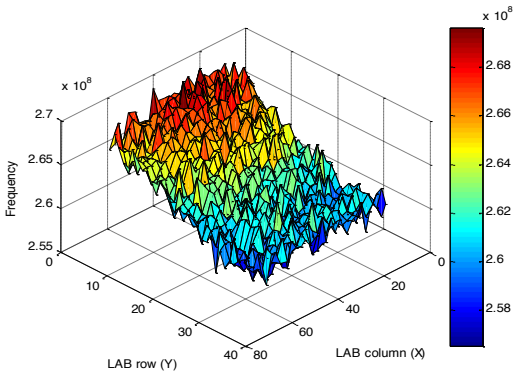
device under test:
Altera Cyclone II

measurement circuit:

Xilinx Virtex-4



Have we measured the right thing?



thermal effects & self heating?

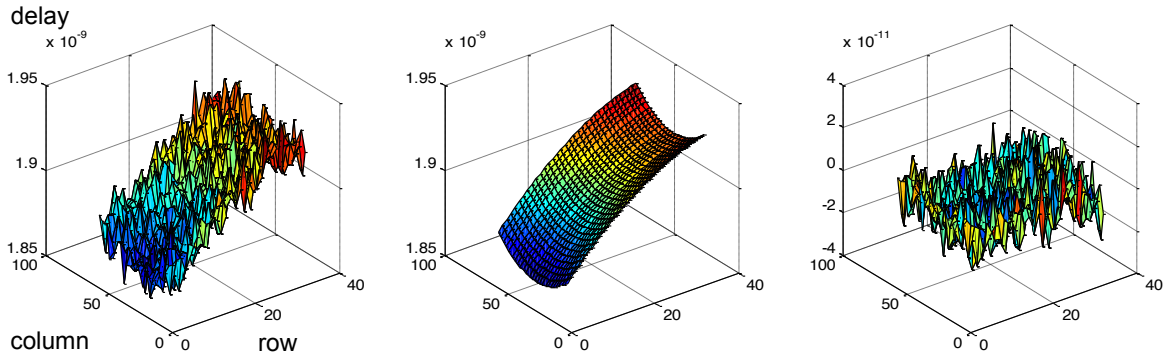
sensitivity to place and route?

measurement error?

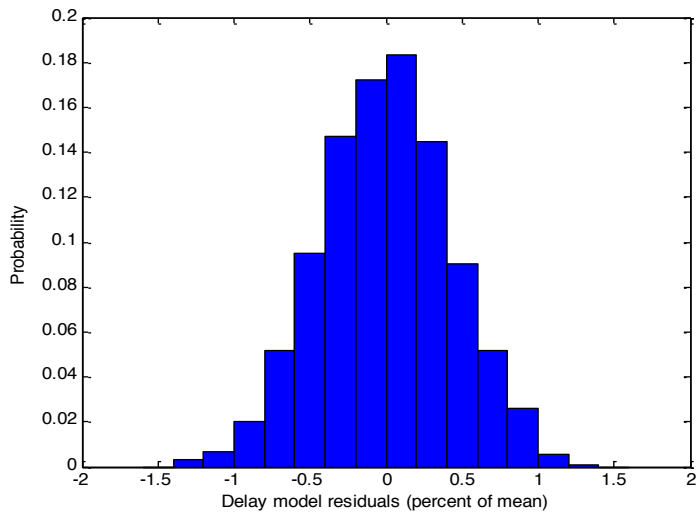
Error source	Error (3σ)
Noise	0.038%
Scan order	0.002%
Place and route	0.223%
LSB of counter	0.02% (max)

Modelling

measured loop delay = model of correlated variation + stochastic variation



$$a + b_x x + b_y y + c_x x^2 + c_y y^2 \qquad \epsilon(x, y)$$



Cyclone II FPGAs:

– 90nm technology

– EP2C35 part

– 18 devices

Stochastic 3σ variation per LUT = ±**3.54%**

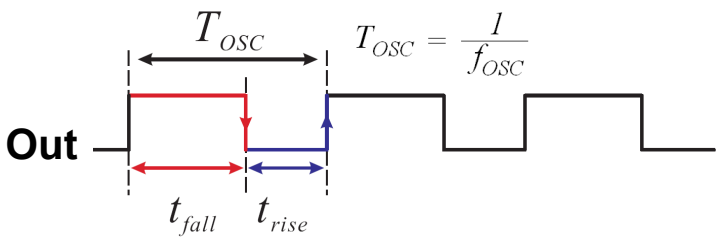
Correlated variation per LUT < **3.66%**

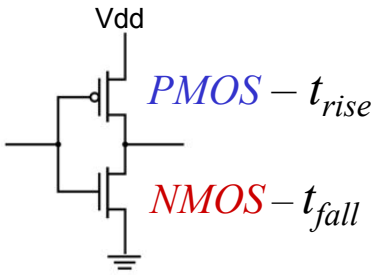
$$\mathrm{Var}(\epsilon) = N \times \mathrm{Var}(\text{one stage})$$

Sedcole & Cheung, "Within-die Delay Variability in 90nm FPGAs and Beyond", FPT 2006

Ring Oscillators is a BAD Instrument

- Easy to implement
- It gets Hot…
- Poor representation of circuit paths in real FPGA designs.
  "Combinatorial loops!?"
- Inaccurate – Only gives **average** delay between rising and falling transitions, NOT worst-case:



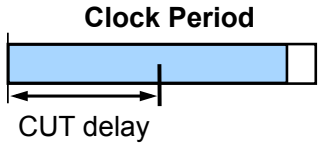


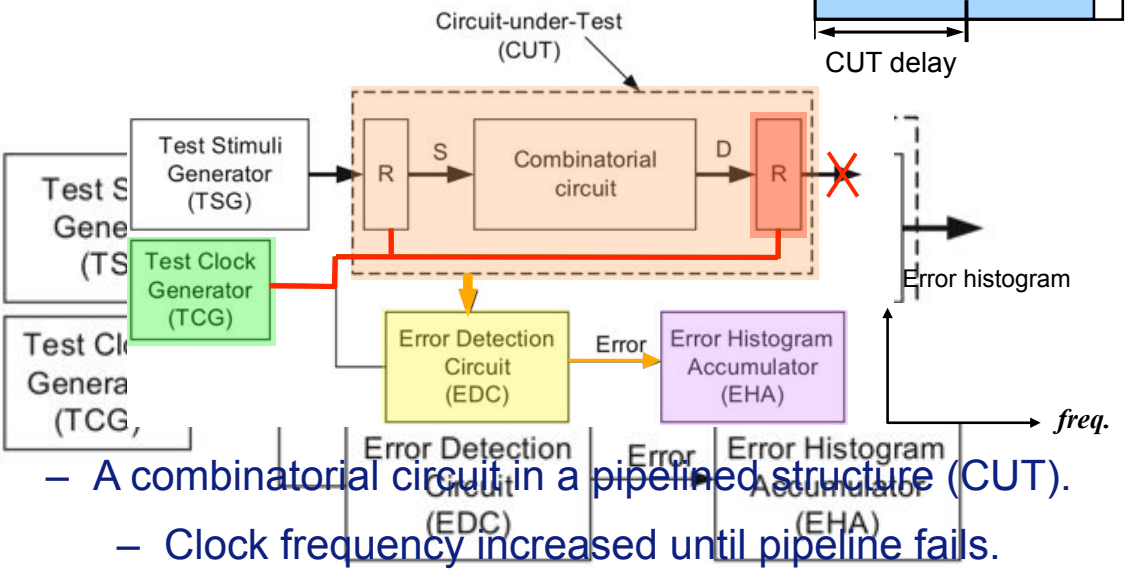
$$Path\ delay \approx \frac{T_{OSC}}{2} = \frac{t_{rise} + t_{fall}}{2}$$

$$t_{fall} \neq t_{rise}$$

Instrument 2: Failure Rate Detector

Failure Rate Detector (FRD) circuit

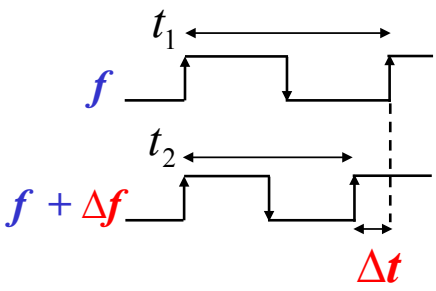


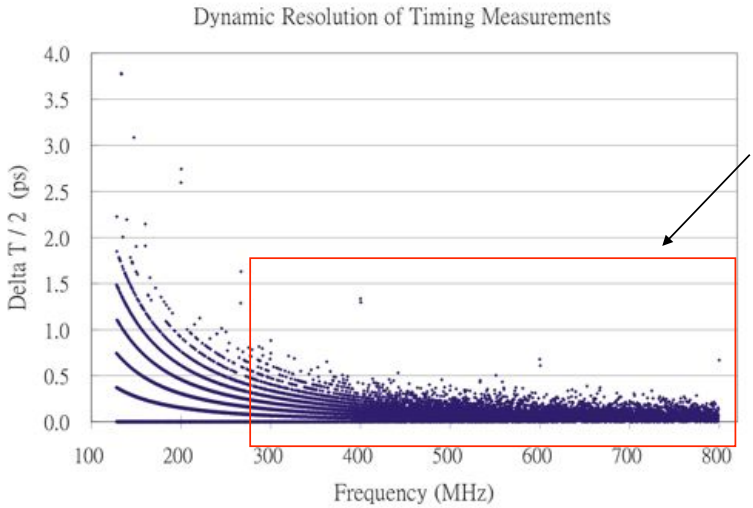


- A combinatorial circuit in a pipelined structure (CUT).
  - Clock frequency increased until pipeline fails.
- EDC detects the error and increment error count on the EHA.

KEY IDEA: Exploit PLL Measurement Resolution

$$\Delta t = t_1 - t_2 = f^{-1} - f^{-1}\left(1 + \frac{\Delta f}{f}\right)^{-1} \approx \frac{\Delta f}{f^2}$$





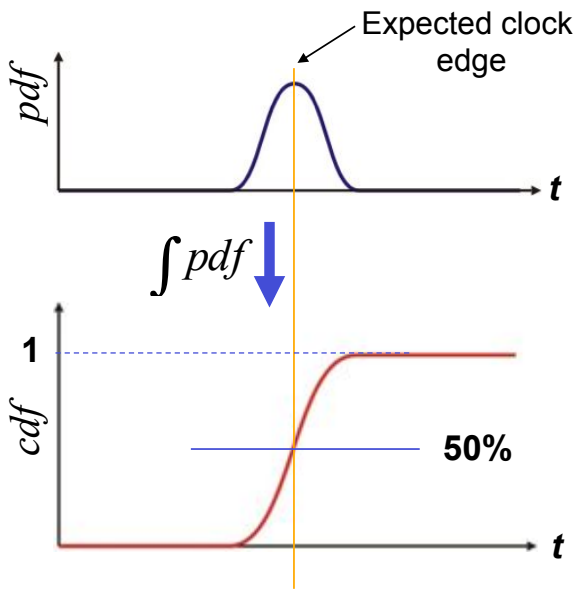
Worse-case timing resolution from 300 to 800MHz

= 1.33ps

Average timing resolution < 1ps

Assumptions

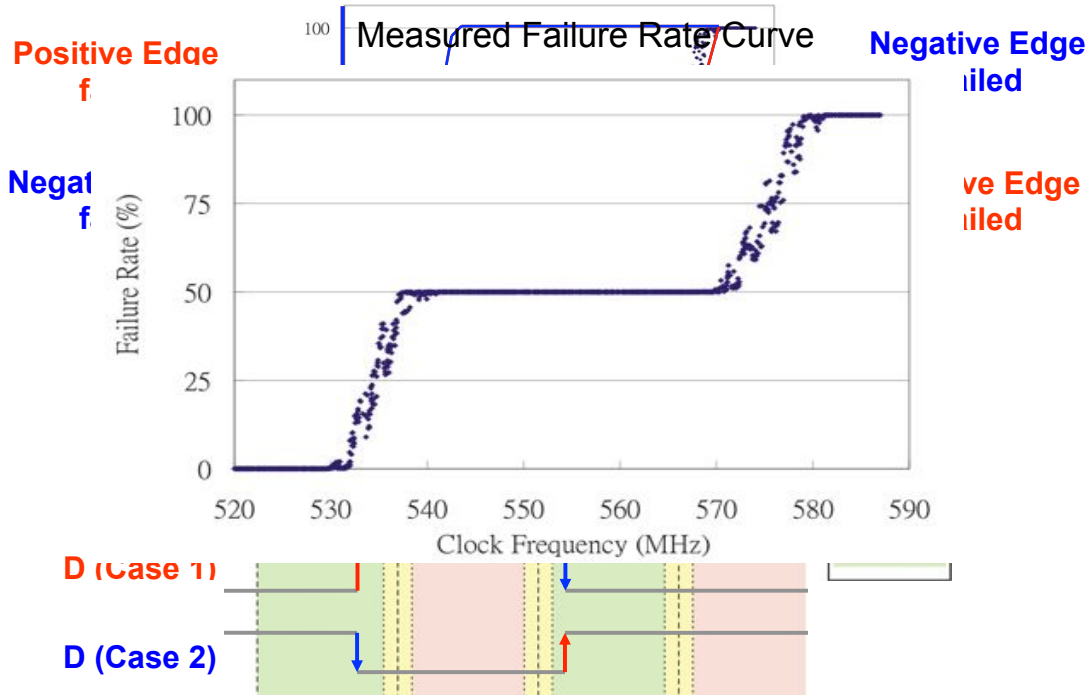
– Clock jitter is approximately Gaussian with symmetrical probability distribution.



– Given that the probability density function *(pdf)* of the clock jitter is symmetrical

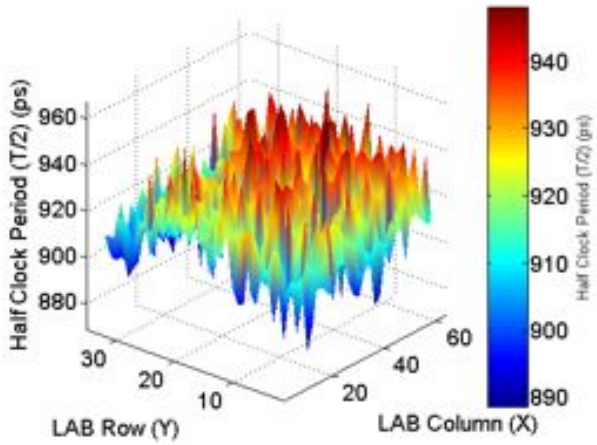
– The resultant cumulative distribution function *(cdf)* would have its 50% point centered at the expected position of the clock edge

Failure Rate Profile Explained



Application 1 (Instr 2): Better LUT Delay Map

FPGA Chip wide Delay map

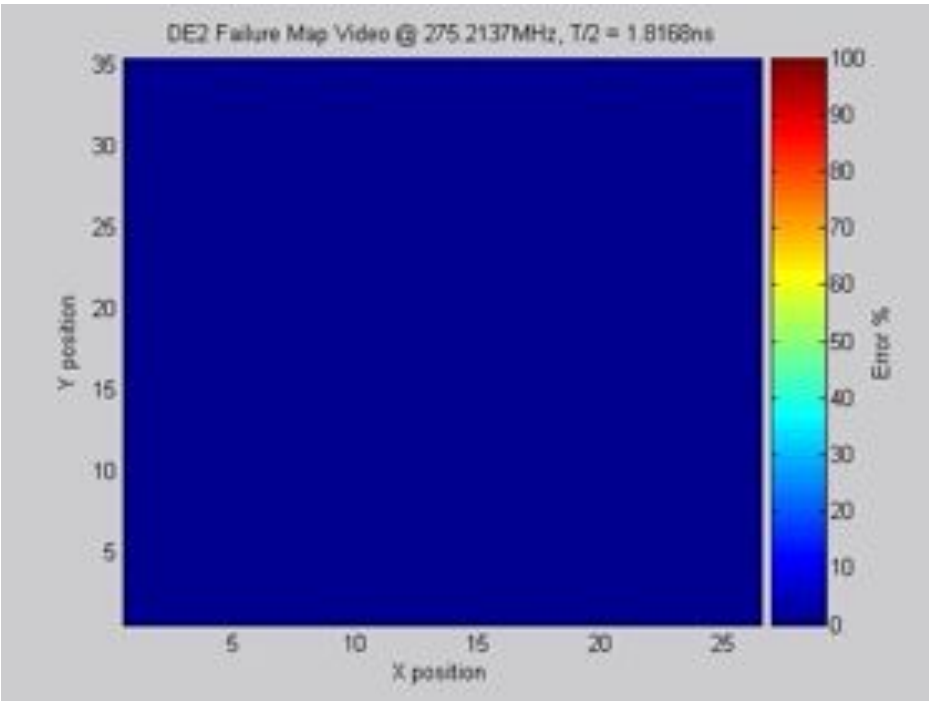


- Results obtained from Cyclone II using the measurement circuit
- CUT: minimum 2 LUTs as inverter.

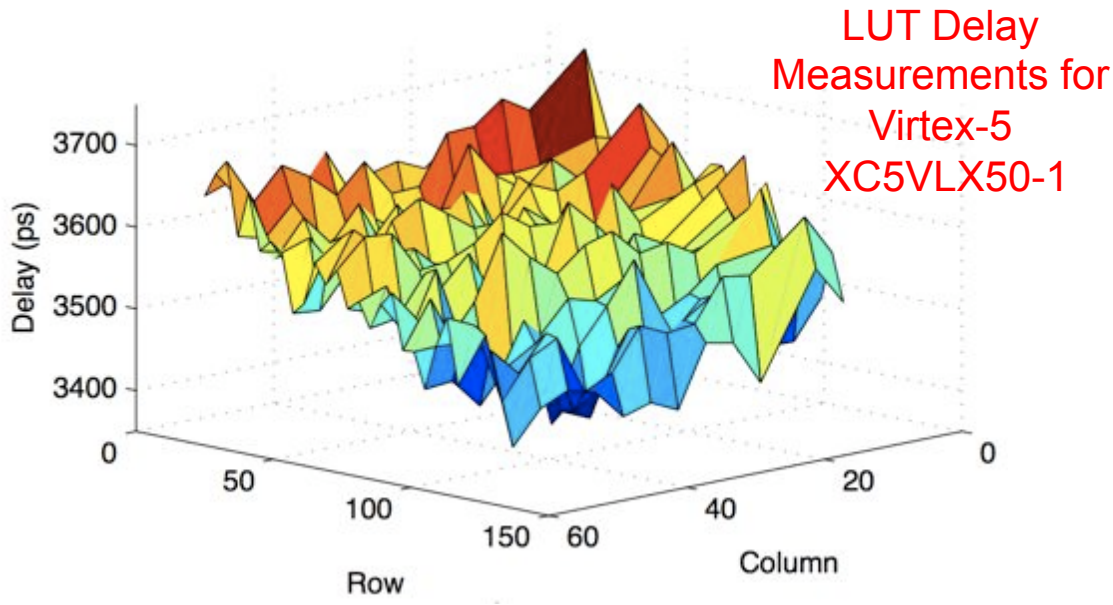
Wong & Cheung, "*Self-measurement of combinatorial circuit delays in FPGAs* ", ACM TRETS, (2) 2, pp. 1-22, 2009 & FPT 2008

Application 1: LUT Delay Map video

Videos showing how FPGA timing failure progressively as test clock frequency is increased

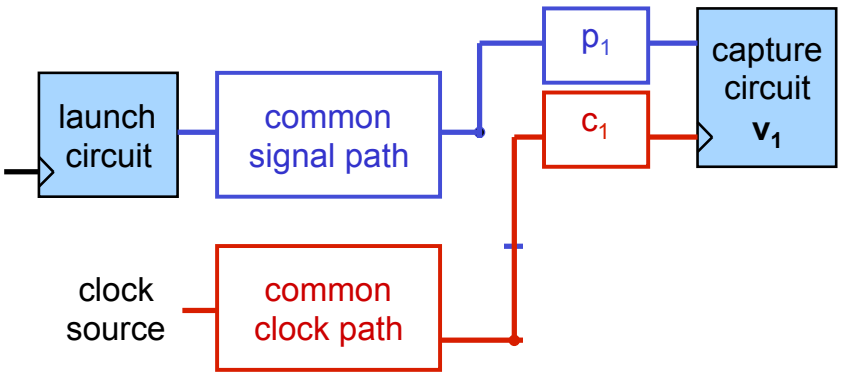


Application 2 (Instr 2) : Clock Delay Variabilities



How much variability comes from the **clock tree**?

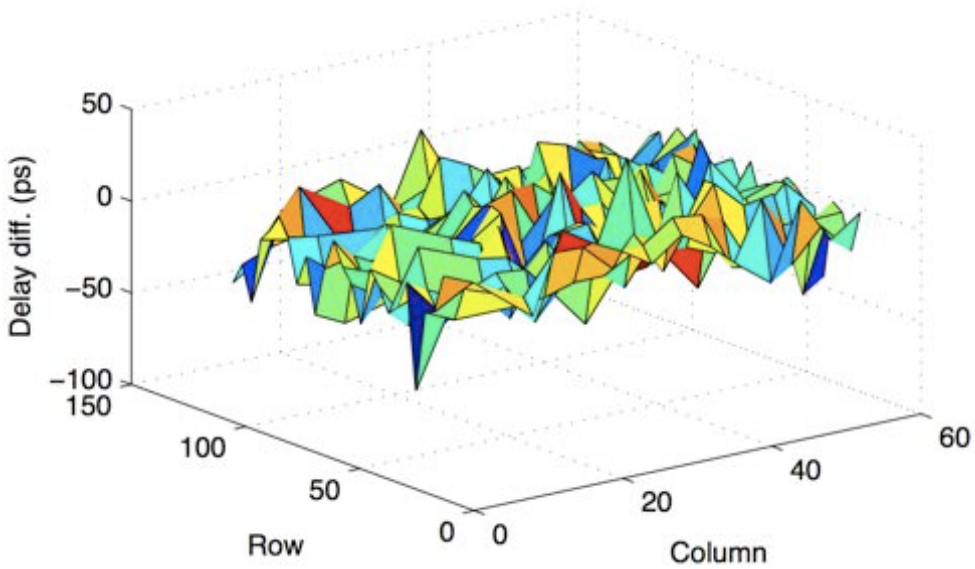
Differential delay measurement circuit



$$\text{Delay diff} = [t(p_1) - t(c_1)] - [t(p_2) - t(c_2)]$$

If p_1 is near p_2 (and c_1 near c_2) then spatially correlated variations cancel out

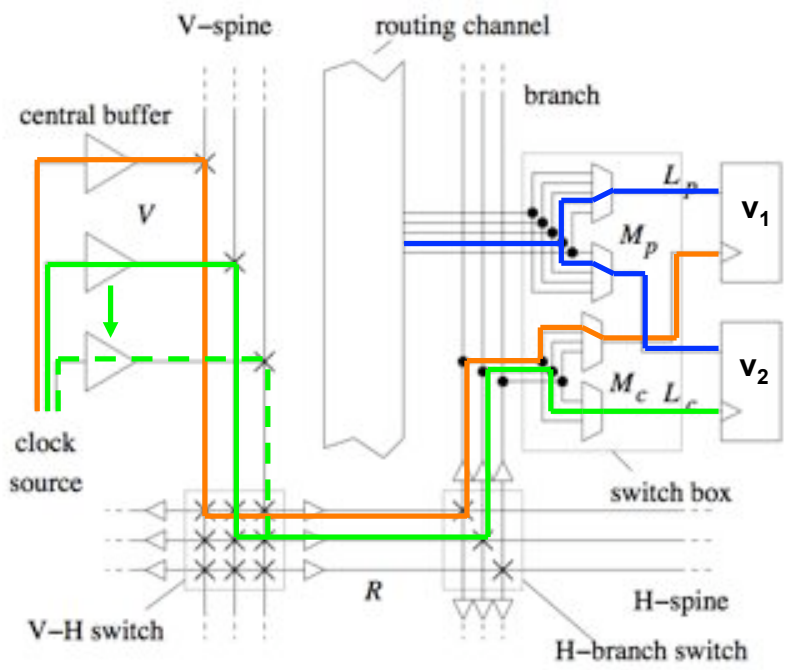
Differential delay measurement example



$$\text{Delay diff} = [t(p_1) - t(c_1)] - [t(p_2) - t(c_2)]$$

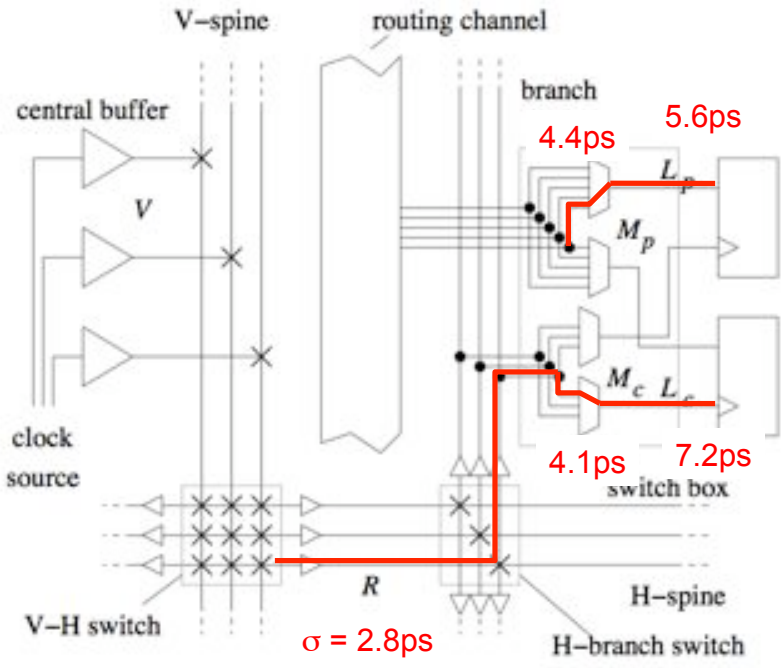
Components of signal path and clock tree

- Simplified lumped model of delays
- Components are isolated by making **incremental routing changes**
- **Variances** are calculated from the measured differences
- A **regression equation** of variances can be solved



Results

Solve linear regression equations to find standard deviations of delays:



Application 2 Results: How much clock skew?

- What is the minimum clock skew variation in a single clock region?
- Estimated σ = 12ps
- Similar to LUT delay variation (σ = 11ps)

Sedcole, Wong and Cheung, "Characterisation of FPGA Clock Variability", IEEE International Symposium on VLSI pp.322-328 (2008)

Problem with Instrument 2

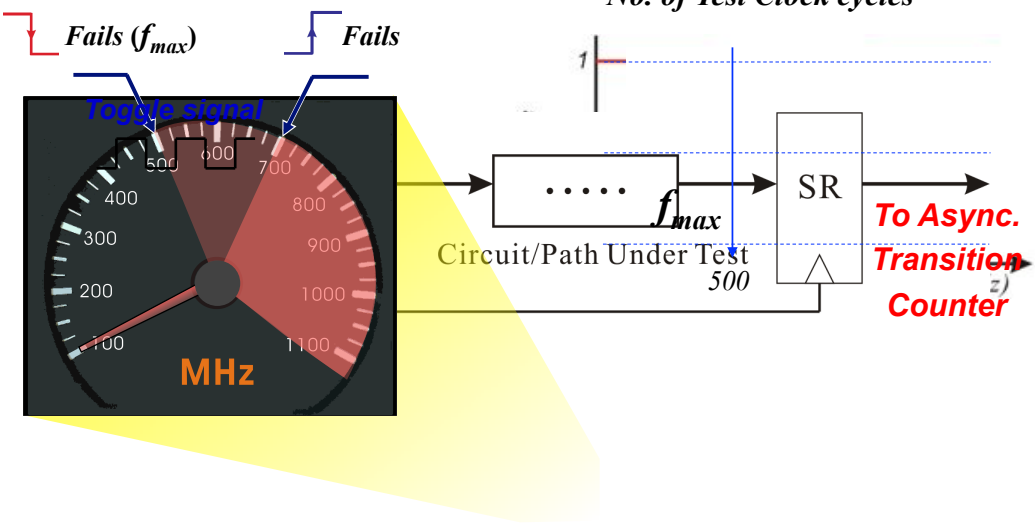
- Good resolutions
- Only works for combinational circuits
- Need to access both inputs and outputs of the capture registers

- Need a better method suitable for “black-box” approach

Instrument 3: Delay Measurement using Transition Probability

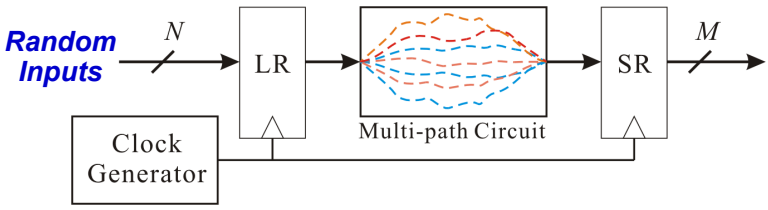
- No synchronous Error Detector needed
- Infer Timing Error by observing **Transitions Probability (TP)**
- The **TP Method**

$$TP = \frac{\textit{No. of Transitions}}{\textit{No. of Test Clock cycles}} \quad \textit{in a freq. step}$$

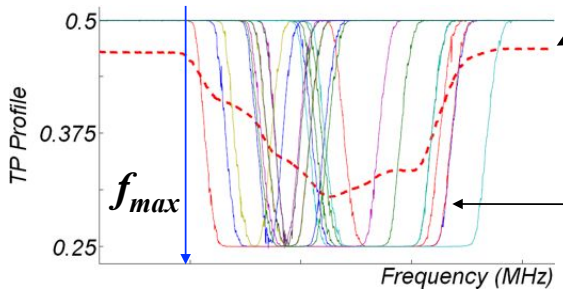


How about complex circuits?

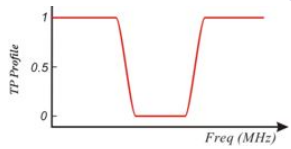
- Drive inputs with random patterns



TP @ 2nd output bit of a 4x4 fixed-point multiplier



No longer has the nice Characteristics of 1 path



But it is formed by a combination of them from each failing path

Wong & Cheung, "*Improved Delay Measurement method in FPGA based on Transition Probability*", ACM Symposium on FPGA 2011

Wong & Cheung, "A Timing Measurement Platform for Arbitrary Black-box circuit based on Transition Probability", TVLSI 2013

Accuracy and precision

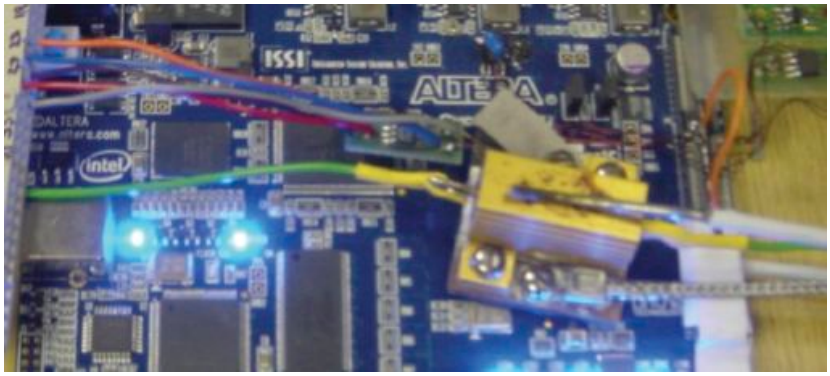
- Isolated single-path:
  - Resolution: ~1 to 2ps (depends on clock generator)
  - Measurement based on **nominal clock period** (centre of jitter distribution)
- Entire circuit (Multi-path):
  - Same as single-path.
  - Measurement based on **minimum clock period** (min. of jitter distribution)

Circuit-Under-Test (CUT)	Resource usage			Latency (cycles)	TP f_{max} (MHz)	Isolated path Reference (MHz)	TP Accuracy (Error %)
	LUT	Reg.	DSP*				
9x9 Embedded Multiplier	0	0	1	2	553.43	547.92	1.01
FP32 Multiplier	418	326	7	5	205.06	201.23	1.90
FP32 Adder	851	375	0	7	173.82	171.91	1.11
FP32 Divider	2233	2589	0	33	232.18	224.11	3.60
FP32 Square-root	762	1215	0	28	283.52	284.10	0.20
Butterworth IIR Filter	991	272	0	4	111.63	111.83	0.18
FIR Filter (15-taps)	361	196	30	3	164.67	165.58	0.55

Largest Design Tested

Application 1 (Instr. 3): Dealing with Delay variability due to ageing

- Degradation characterisation
  - Accelerated life test
  - Measure and model how logic slows down over time under stresses
    - Heat, voltage and different switching stresses

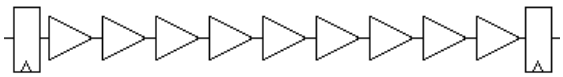


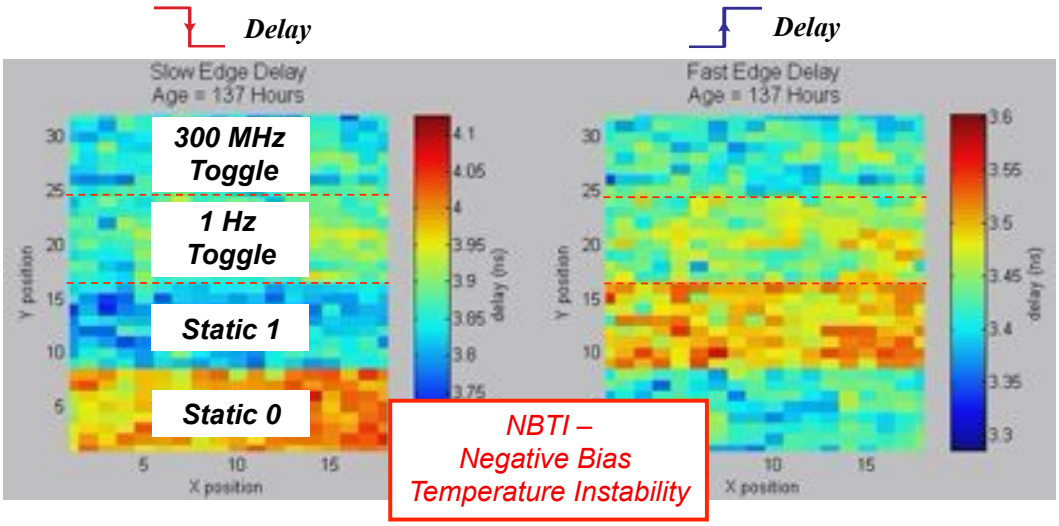
Stott, Wong & Cheung., "*Degradation in FPGAs: Measurement and Modelling*", ACM Symposium on FPGA 2010

Demo: 10 years worth of degradation in 17 seconds of video

- Cyclone III Accelerated life test
- with 4 types of input stress @ 125°C, 1.8v
- TP Test every hour @ 35°C, 1.2v (default voltage)

Path under test / stress:





What do the results tell us about degradations on FPGAs?

Gradual, no hard failures observed

Suggests NBTI dominant

- High frequency also severe (HCI)

LUTs and interconnect affected

- LUTs worst affected

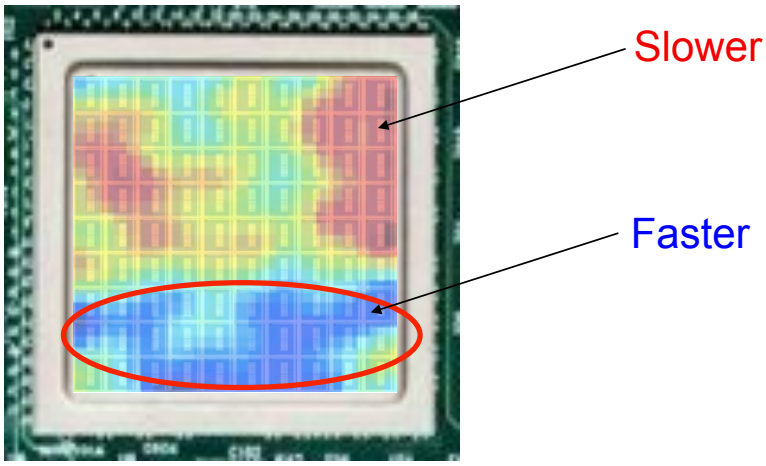
Promising for life-extension

- Non-uniform
- Gradual

Application 2 (Instr. 3): Variation-aware place-and-route

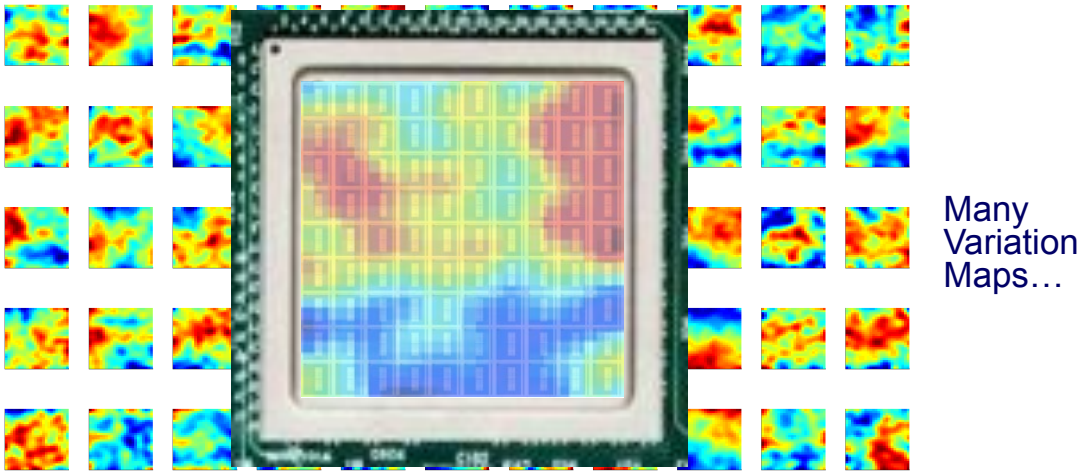
Idea:

- Measure chip-specific delay map (**Variation Map**)
- Place critical part of design into Fast Region (**Variation-aware Placement**)



However,…

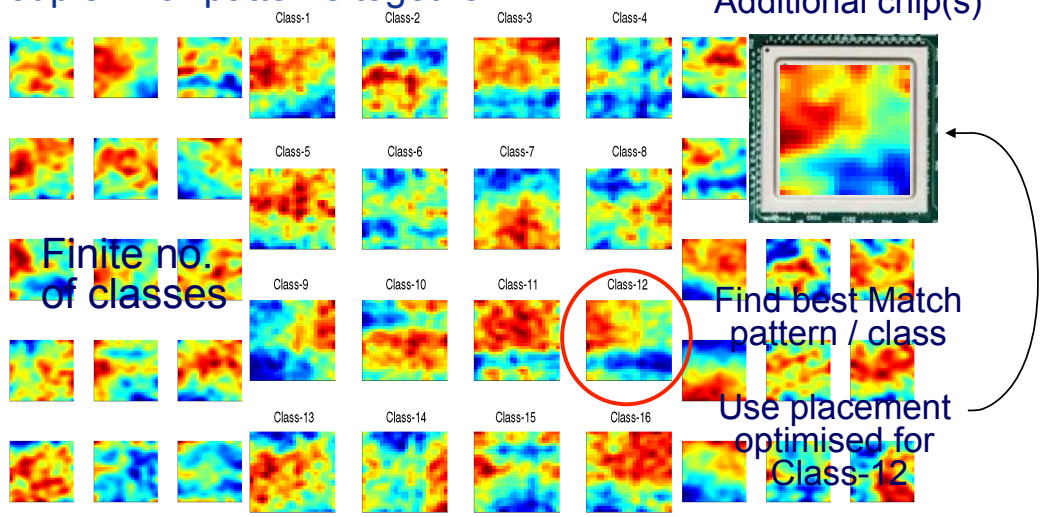
- Practical use of FPGAs involves large number of chips
- NOT just one specific chip…



- Each chip has unique Variation Map (and optimum placement)
- Very Time consuming: Variation-aware Placement for each chip

Solution

- **Pattern classification / clustering**
- Group similar patterns together



- Perform variation-aware placement for each class
- Reduce total run time, while retaining close-to-optimal placement

Goals of Variation-Aware

- Investigate how to use measured variation maps to improve timing performance
- With reasonable execution time overhead
- Integration into practical work flow for industry

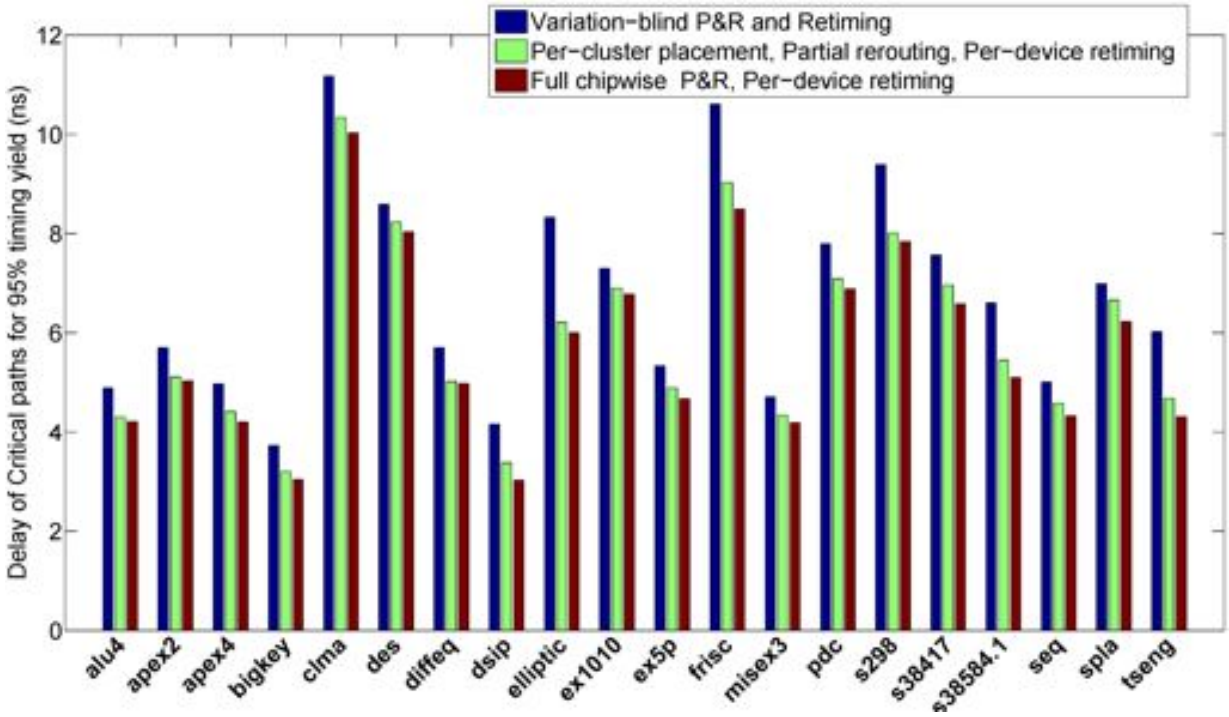
What we have achieved so far

- Two-stage variation-aware placement
- Variation-aware partial rerouting
- Variation-aware retiming

Guan, Wong, Constantinides & Cheung, “A two-stage variation-aware placement method for fpgas exploiting variation maps classication”, FPL 2012

Guan, Wong, Constantinides & Cheung, “A Variation-adaptive Retiming Method Exploiting Reconfigurability”, FPL 2013

Results Combined all Optimisation Methods



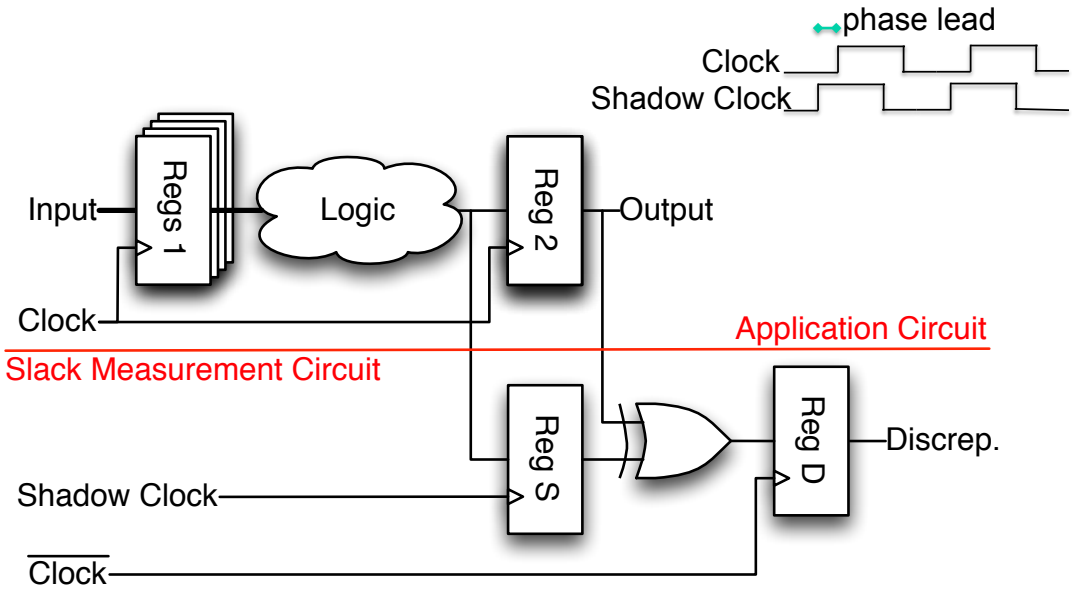
Where have we got to?

Instrument	Applications
1. Ring Oscillator	• Stochastic vs Systematic Variation
2. Timing Error Detection	• LUT delay map characterisation • Clock skew measurement
3. Transition probability	• Degradation characterisation • Variation-aware P&R and re-timing

- Our instruments so far operate OFF-LINE
- Need another method to perform delay measurement under normal operational condition

4. Online Slack Measurement	• **Online Health Monitoring** • **Dynamic voltage/frequency scaling**

Instrument 4: Online Slack Measurement (OSM)



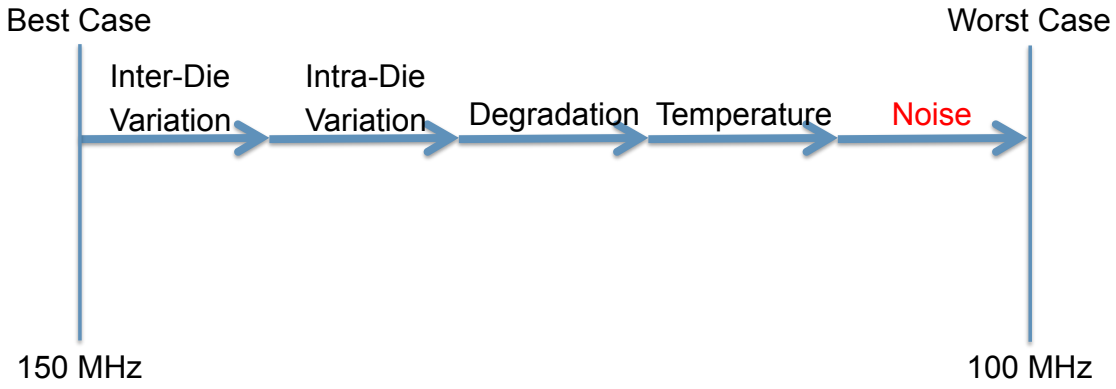
Levine Stott, Constantinides, **& Cheung**, "Online Measurement of Timing in Circuits: for Health Monitoring & Dynamic Voltage and Frequency Scaling", FCCM 2012

Applications (Instr. 4): Health monitoring & Dynamic Voltage/Frequency Scaling

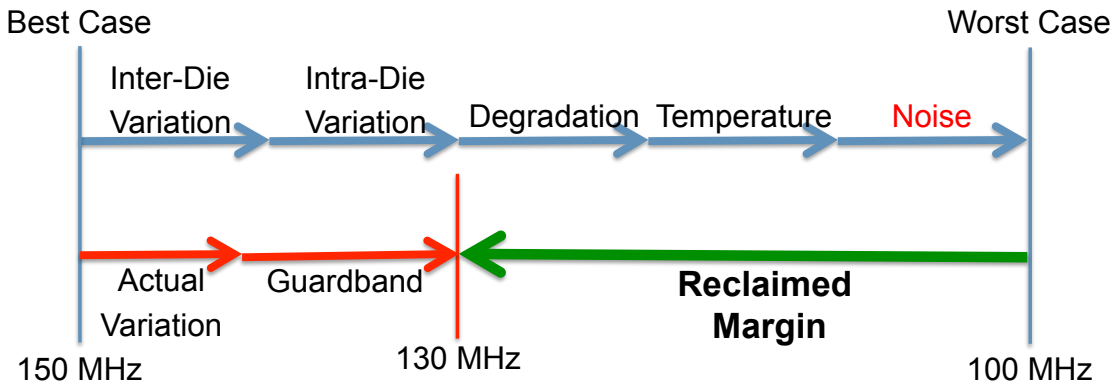
- Measure the actual timing slack in the circuit while it is working normally using Online Slack Measurement (OSM) technique
- Use timing slack to reduce the timing margin in order to:
  - Reduce power, or
  - Increase throughput, or
  - A combination of the two

Levine, Stott **& Cheung**, "Dynamic Voltage & Frequency Scaling with Online Slack Measurement", ACM FPGA Symposium, 2014

Timing Safety Margins



Reduced Timing Margin



Dynamic Scaling

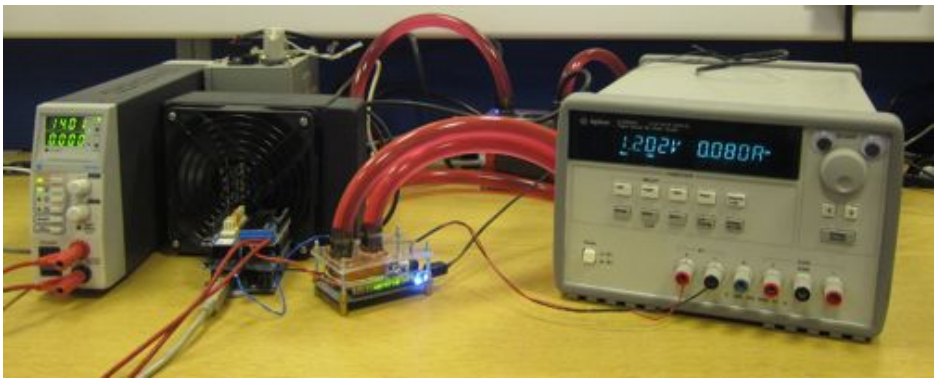
- Dynamic Voltage Scaling (DVS):
  - Scale the voltage
  - Frequency is constrained
- Dynamic Frequency Scaling (DFS):
  - Scale the frequency
  - Voltage is constrained
- Dynamic Voltage & Frequency Scaling (DVFS):
  - Scale both the voltage and frequency
  - Power is constrained

Experiments

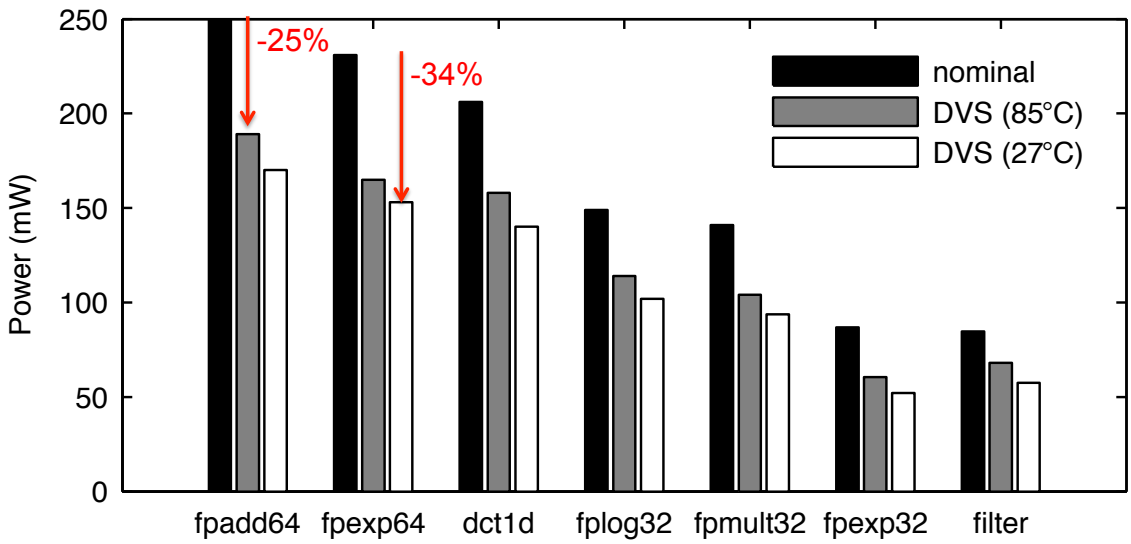
- A variety of functional benchmarks from FloPoCo and Spiral
- Contain memory and DSP
- LUTs: 1.1k – 5.4k, Regs: 0.9k-5.1k
- Automatically instrumented for online slack measurement
- Overheads:
  - 1.1% increase in LUTs
  - 2.5% increase in Regs
  - 1.8% decrease in model fmax

Experiment Rig

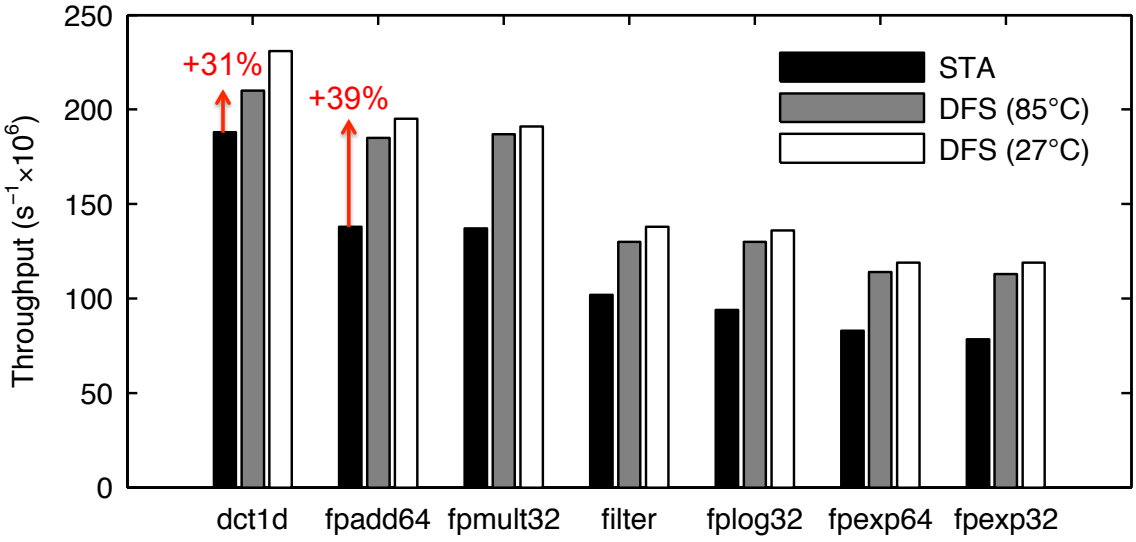
- Altera Cyclone IV FPGA (Tersaic DE0-nano)
- Temperature controlled package
- PSU supplies core voltage and provides power measurement



Dynamic Voltage Scaling Results

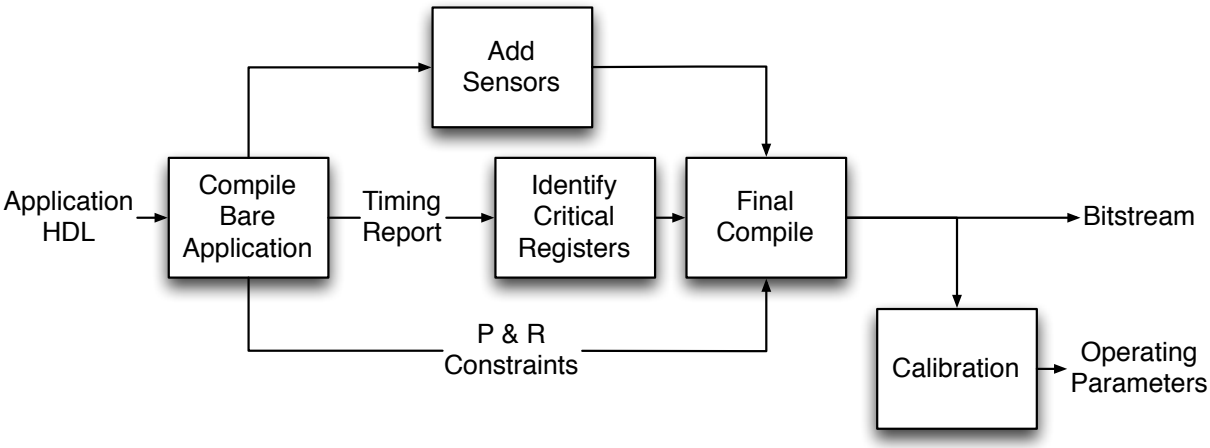


Dynamic Frequency Scaling Results



Automation Tools – no hands!

- Tools to automatically insert TP delay (TPD) and online slack measurement (OSM) circuitry
- Fully compatible with vendor's compilers
- Requires little to no manual intervention



Summary

Instrument	Applications
1. Ring Oscillator	• Stochastic vs Systematic Variation
2. Timing Error Detection	• LUT delay map characterisation • Clock skew measurement
3. Transition probability	• Degradation characterisation • Variation-aware P&R and re-timing
4. Online Slack Measurement	• Online Health Monitoring • Dynamic voltage/frequency scaling

Conclusions

- **Variability: this problem is here to stay. What are our response?**
  - "… Just give up …"
  - "….. yield will become zero …"
  - "….. semiconductor industry will always solve the problem .."
- **On-silicon instrumentation**
  - Will become increasingly important
  - When coupled with reconfigurability, open up new possibilities
- **VLSI chips: no need to treat them all the same (clones)**

Reconfigurability ~~may~~ is be the answer to the variability and reliability challenge

Acknowledgement

- Thanks to EPSRC for support of these grants:
  - Variation-Adaptive Design in FPGAs
  - PLATFORM: Custom Computing for Advanced Digital Systems
  - PLATFORM: Field-Programmable Logic for Custom Computing
  - PROGRAMME: PRiME (Power-efficient, Reliable, Many-core Embedded systems)
- Xilinx and Altera
- My students/RAs working/worked with me on this topic:


Secole


Wong


Stott


Guan


Levine


Davis

Advertisement

EPSRC funded **CENTRE FOR DOCTORAL TRAINING** (CDT)

In

HIGH-PERFORMANCE EMBEDDED AND DISTRIBUTED SYSTEMS (HiPEDS)

Department of EEE and Department of Computing

Imperial College London

50+ new PhD positions

from October 2014 until 2020