

Modeling Post-Techmapping and Post-Clustering FPGA Circuit Depth

Joydip Das¹, Steven J.E. Wilton¹, Philip Leong², Wayne Luk³

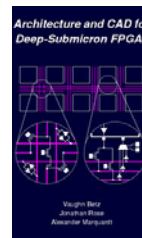
¹*The University of British Columbia,*
²*The Chinese University of Hong Kong,*
³*Imperial College London*

- *Funded by Altera and NSERC*

FPGA Architecture Design

Architectures are usually evaluated using experimental methods

- Using tools like VPR



Problems:

- Multi-dimensional optimization space – too much time
- Require CAD tools for each architecture
 - Or “tuning” of a generic tool like VPR
- No insight into what makes a good architecture

Can we supplement the experimental approach with analytical techniques?

[This talk](#): A model that makes it possible

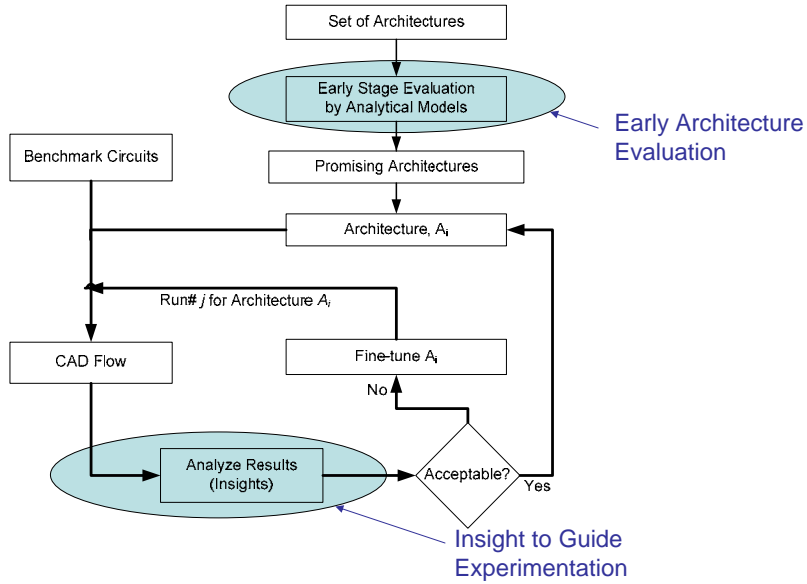
3

This Talk

1. Motivation: Speeding up architecture design
2. The model:
 - What makes a good model
 - What makes it hard
 - Overview
3. Details on Depth/Delay Model
4. Example of our Model's Application

4

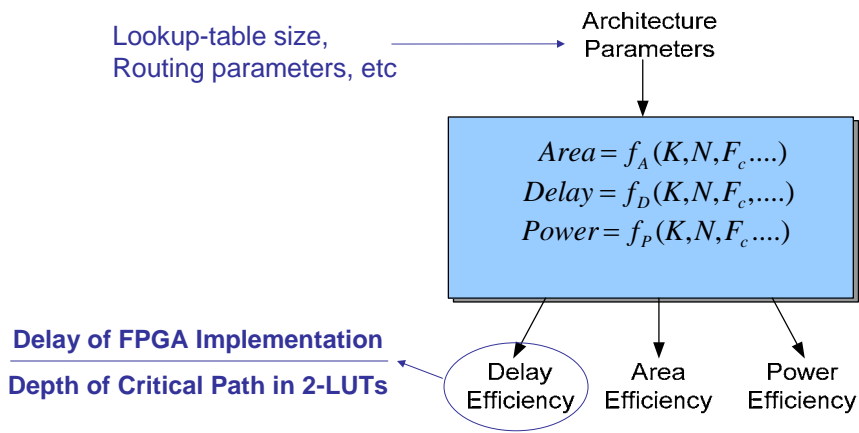
Accelerating FPGA Architecture Investigation



5

Analytical Model

The key is an analytical model that relates architecture parameters to efficiency of the FPGA:



6

Challenge: Capturing the “essence” of programmable logic
in a set of simple equations

7

What makes a good model?

1. Analytical Model
 - No curve-fitting or expensive experimental techniques
2. Balancing Complexity and Accuracy
 - Simpler equations provide significantly more insight into architectural tradeoffs
3. Architectural Relationships
 - Should be as independent of user circuit as possible

8

Estimation vs. Modeling

Estimation: What is the performance or density of a given user circuit on an FPGA?

- Useful in CAD tools to predict long paths, congested regions, etc.

Modeling: *On average*, how does an architecture parameter affect the expected speed or density of an FPGA?

- Can we answer this independent of the user circuit?

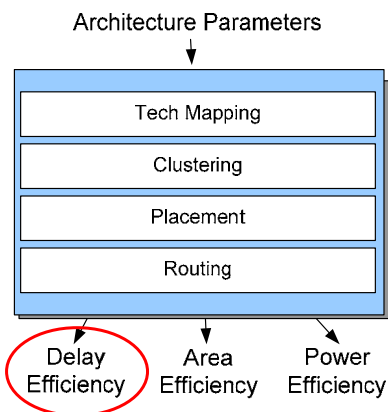
9

What makes it hard:

- Many parameters that interact in complex ways

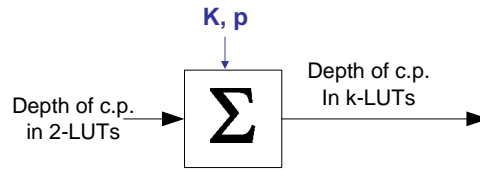
How we make it feasible:

- Break the model into *stages*, analogous to CAD flow



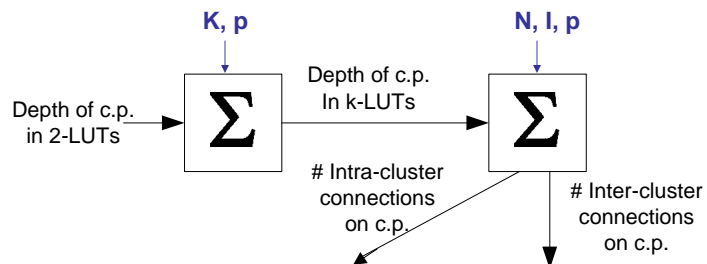
10

Breaking it up: Delay



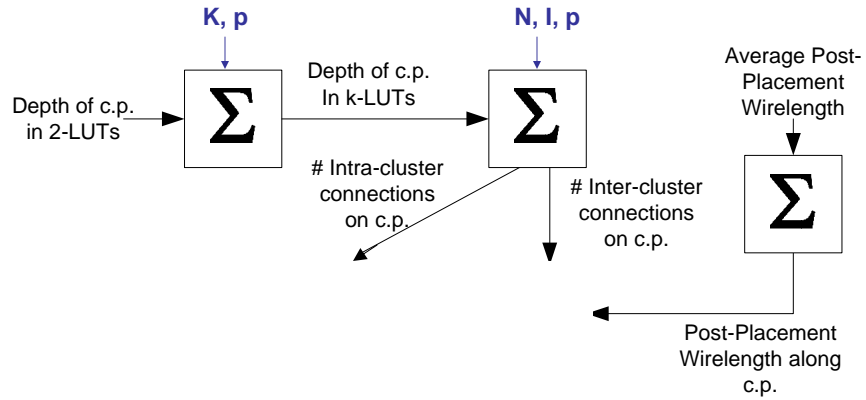
11

Breaking it up: Delay



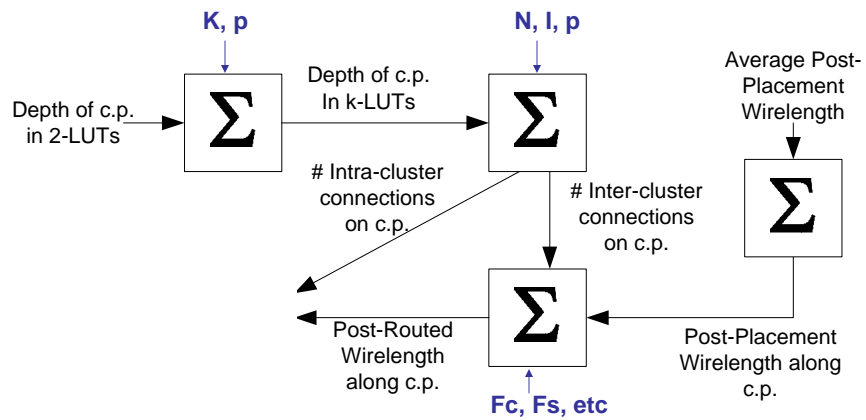
12

Breaking it up: Delay



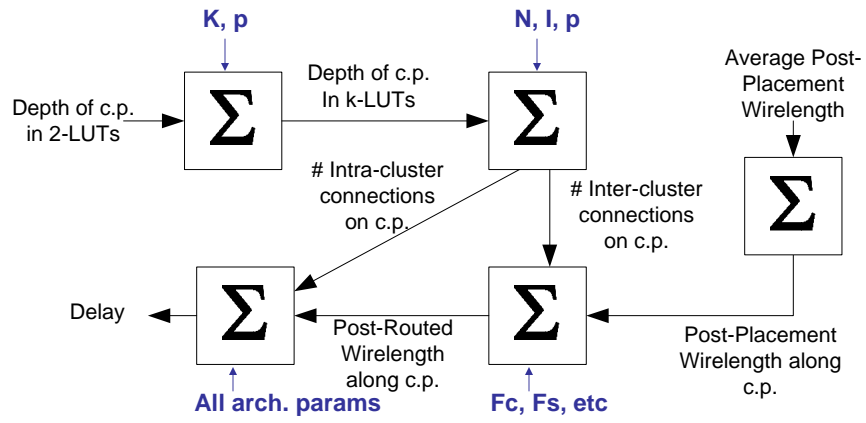
13

Breaking it up: Delay



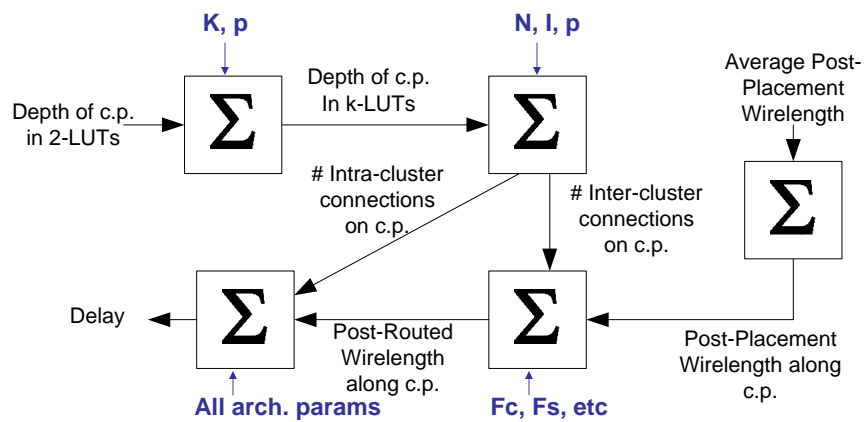
14

Breaking it up: Delay



15

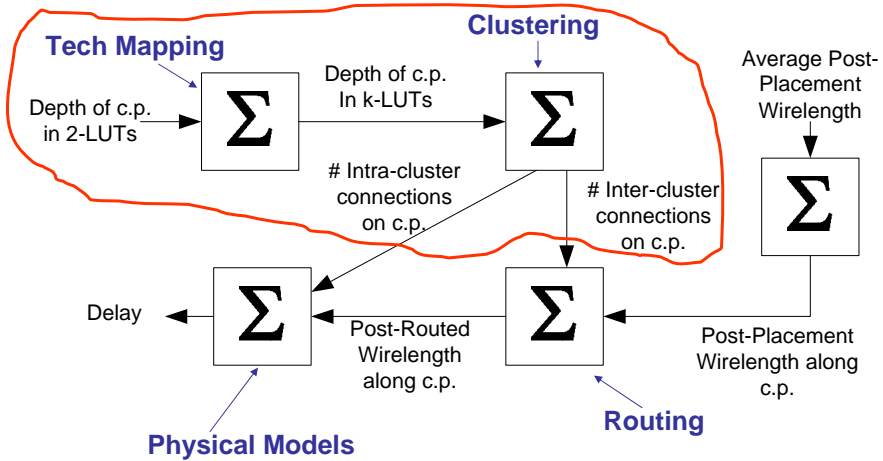
Breaking it up: Delay



Each part is simple, but together, they relate delay-efficiency to architectural parameters

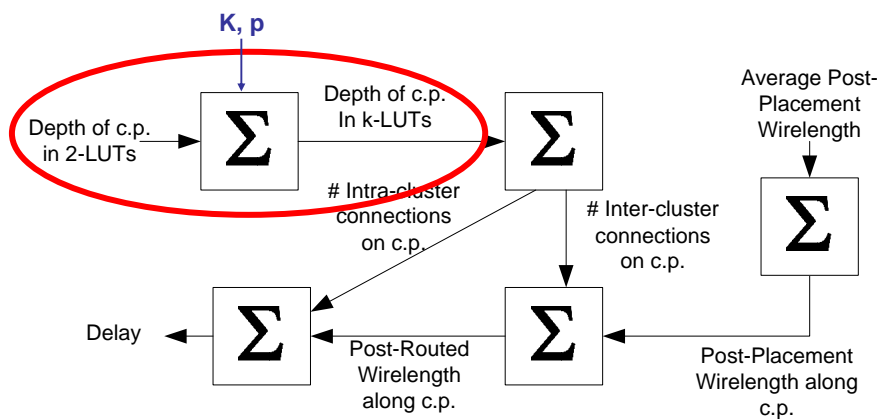
16

This Paper:



17

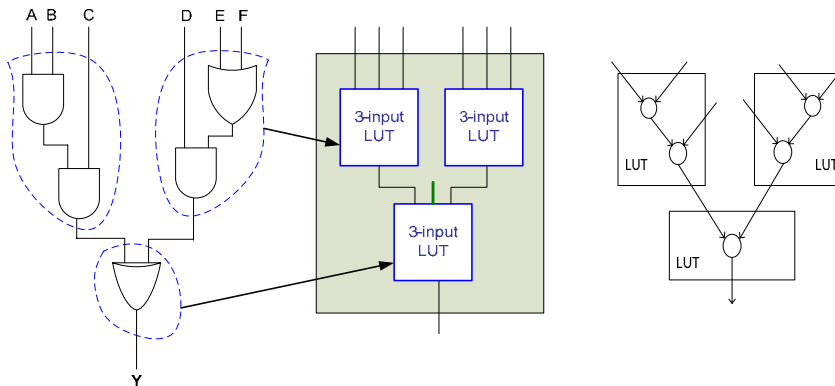
Technology Mapping Model



18

Review: Technology Mapping

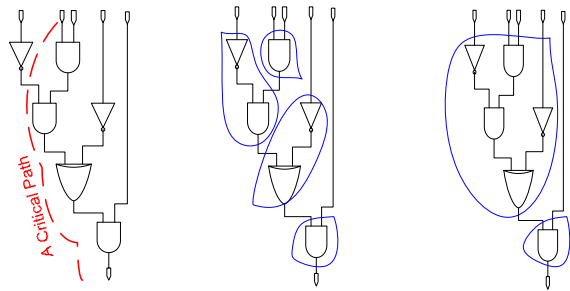
Map logic gates to lookup-tables:



Most algorithms give implementations of minimum depth

Modeling Technology Mapping:

Intuitively, Bigger LUT Size Means Smaller Depth



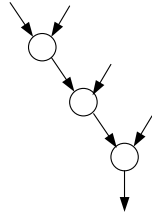
Circuit

2-LUT / Depth=4

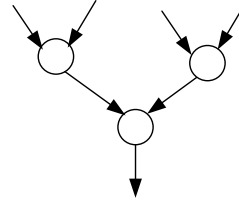
4-LUT / Depth=2

So, Larger LUT Size → Fewer Nets → Lower Depth

Mapping with $K=4$: Two Extremes



Depth = $(K - 1) = 3$
[Maximum Possible]



Depth = $\log_2(K) = 2$
[Minimum Possible]

Simple Approach: Take the average

21

Technology Mapping Model

Depth after
Techmapping

$$\frac{d_k}{d_2} = \frac{2}{K - 1 - \gamma + \log_2(K - \gamma)}$$

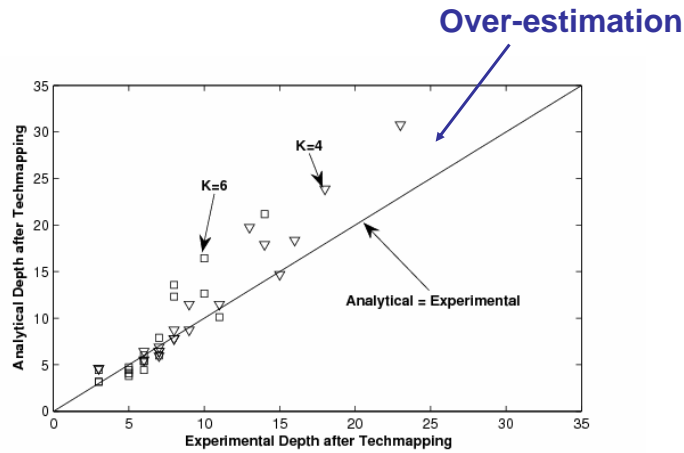
Depth before
Techmapping

LUT Size

Average Unused Inputs
in Each LUT

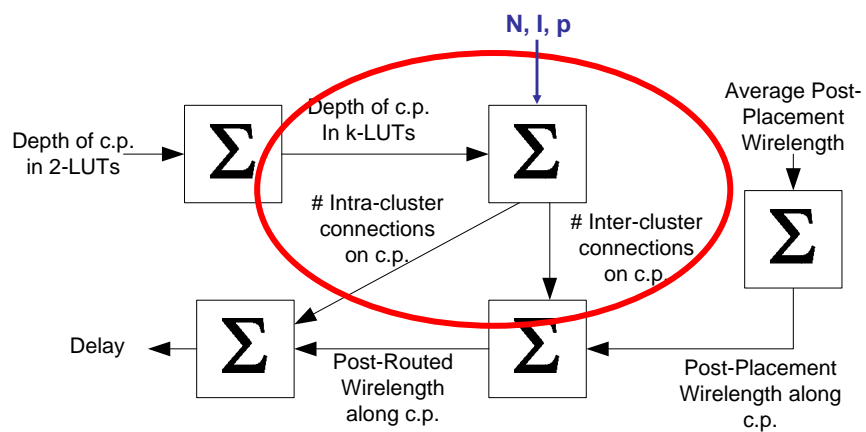
22

Techmapping Model : Validation



23

Clustering Model



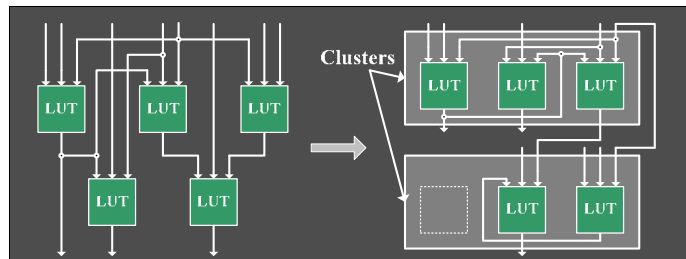
24

Review: Clustering / Packing

FPGA logic blocks usually contain several LUTs:

Altera: LABs

Xilinx: CLBs



Goal of Clustering Algorithms: Group LUTs into LAB-sized clusters

- Connections between LUTs within a cluster are fast

25

Clustering Model

Clustering does not eliminate nets:

- It just makes some nets local (intra-cluster) and some global (inter-cluster)

Intuitively: the larger the cluster size, the more nets are made local.

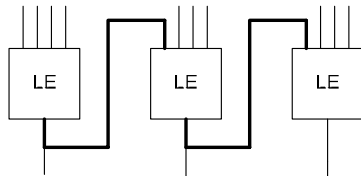
Goal: derive an equation for the proportion of nets along the critical path that are made local

26

Clustering Model

Sketch of derivation:

1. Some nets are made local "on purpose"



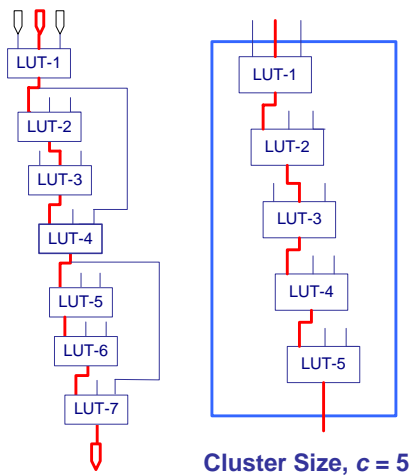
2. Some nets are made local "by chance"

- These are not nets that are specifically targeted by the cluster algorithm

Work out proportion of each and combine them

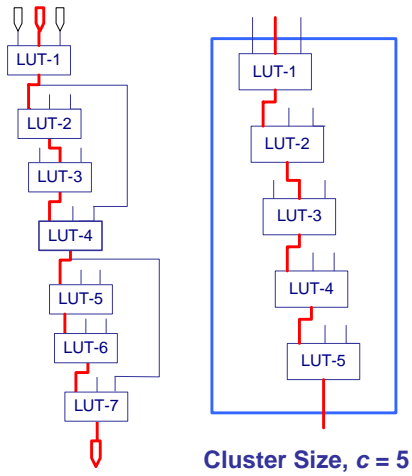
27

Connections on Critical Path – *Primary Goal*



28

Connections on Critical Path – *Primary Goal*

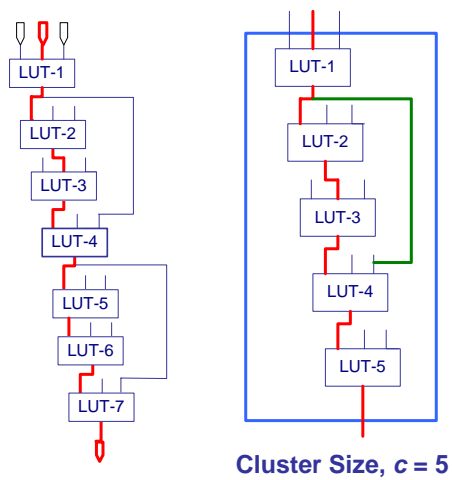


Cluster Size = c

Absorbed / Local = (c - 1)

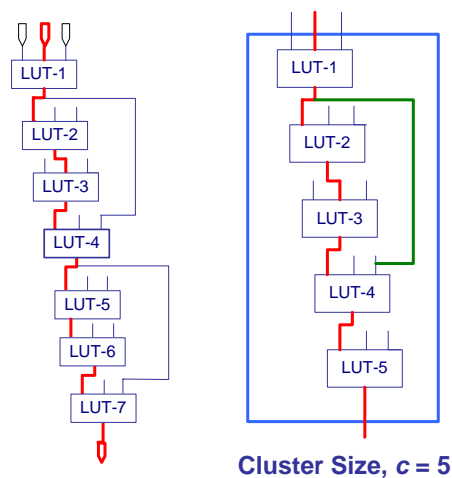
29

Connections Absorbed: *Not on Critical Path – by chance*



30

Connections Absorbed: *Not on Critical Path – by chance*



Absorbed by chance:

$$\frac{c}{n_k} [c(K - \gamma) - c + 1]$$

Details of derivation can be found in our paper

31

Clustering Model:

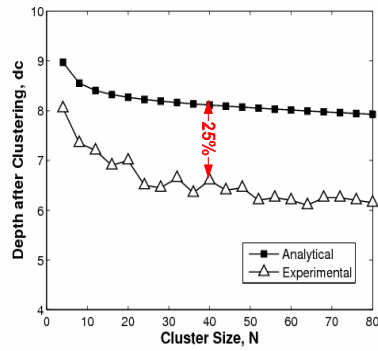
Which leads to

$$\frac{d_c}{d_k} = \left[\frac{(c - 1) + \frac{c}{n_k} [c(K - \gamma) - c + 1]}{c(K - \gamma)} \right], \text{ where } c = \frac{n_k}{n_c}$$

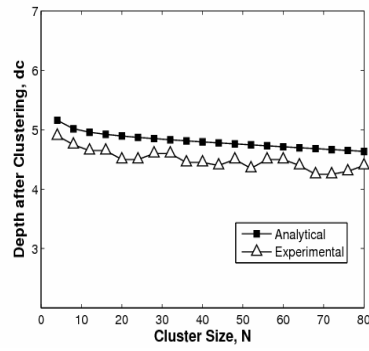
Details of derivation can be found in our paper

32

Clustering Model : Validation



LUT Size, K=4

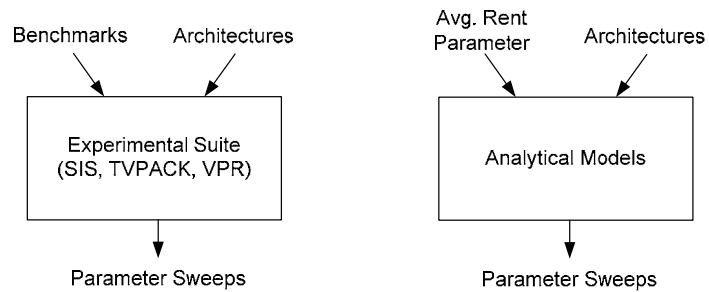


LUT Size, K=6

Is our Model Actually Useful?

Example of Model's Application

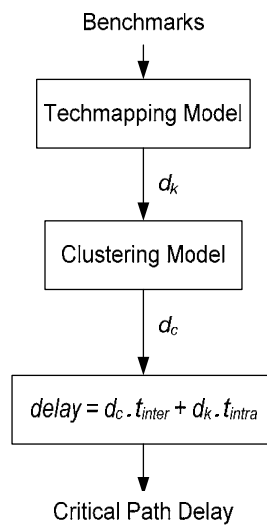
We considered two flows:



The “shape” of the results is more important than the actual values

35

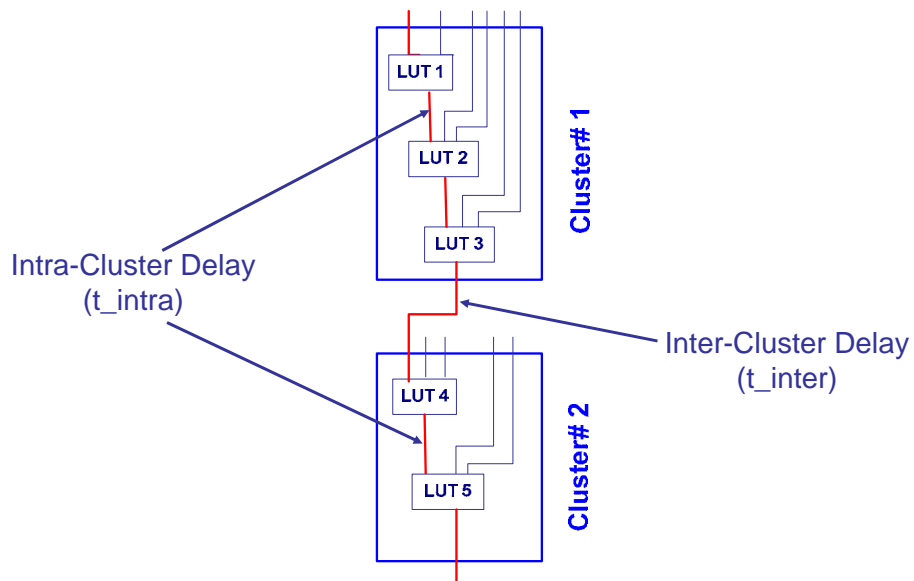
Critical Path Delay from Analytical Model



Analytical Flow

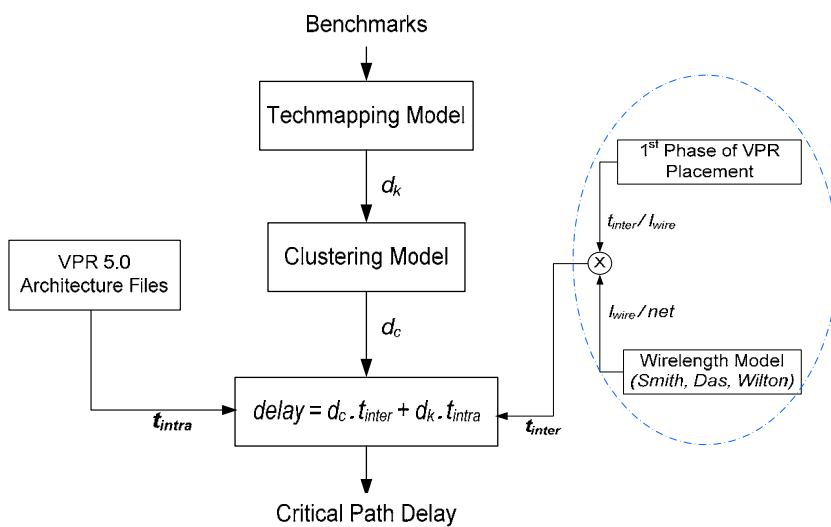
36

Intra-Cluster & Inter-Cluster Delay:



37

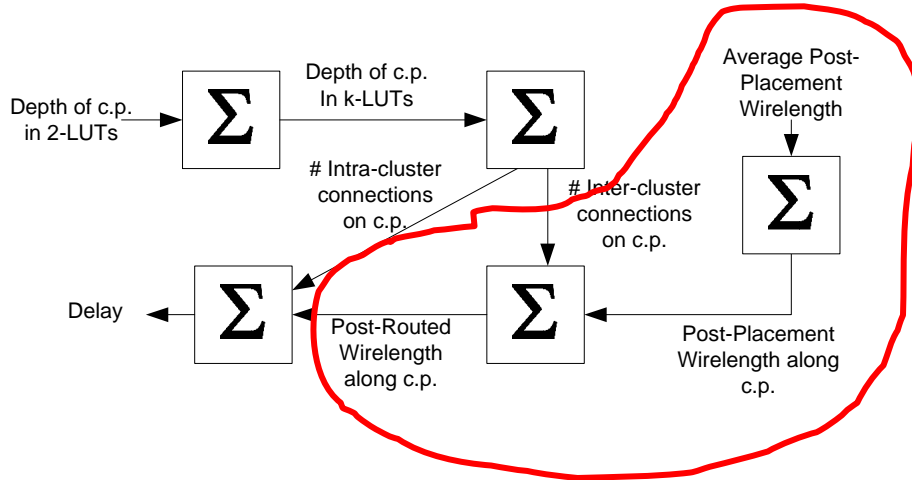
Critical Path Delay from Analytical Model



Analytical Flow

38

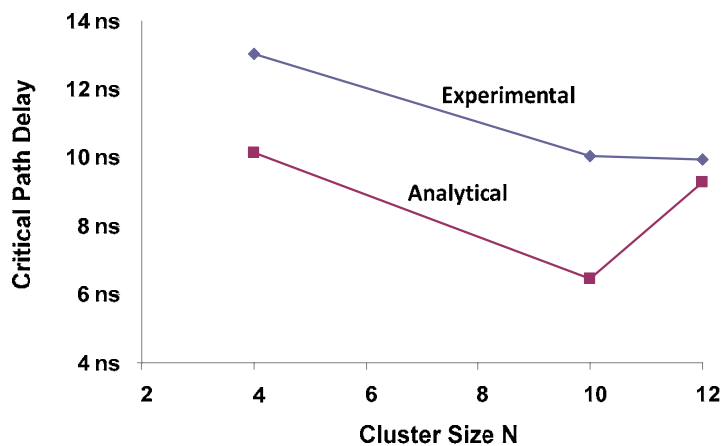
Important caveat: We do not yet have a model for delay routing



For now, we use experimental results for this part

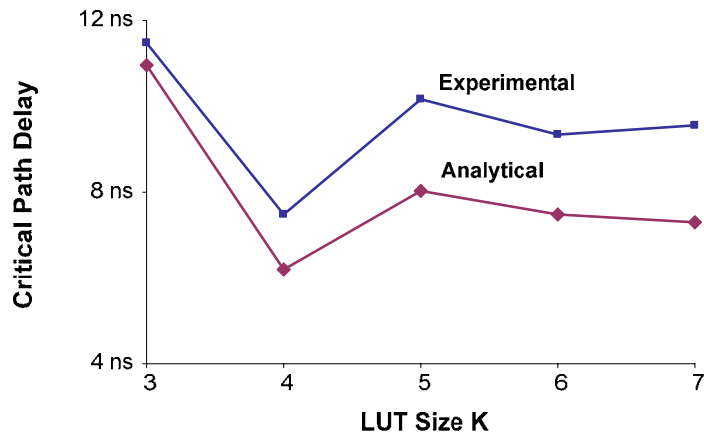
39

Overall Results: Delay



40

Overall Results: Delay



Same Conclusion in both cases: K=4.

41

Summary

Key Result: It is possible to describe an FPGA architectures using a set of simple equations

This Talk:

- Analytical model for techmapped & clustered depth
- Example of model's application for early stage architecture evaluation

Ongoing Works:

- Analytical model for post-routing delay
- Investigation of "Discrete Effects"
- Analytical model for the whole design flow

42