Reconfigurable Computing

FPGA Architecture

Architecture should speak of its time and place, but yearn for timelessness. – Frank Gehry

Philip Leong (philip.leong@sydney.edu.au)
School of Electrical and Information Engineering

http://phwl.org/talks

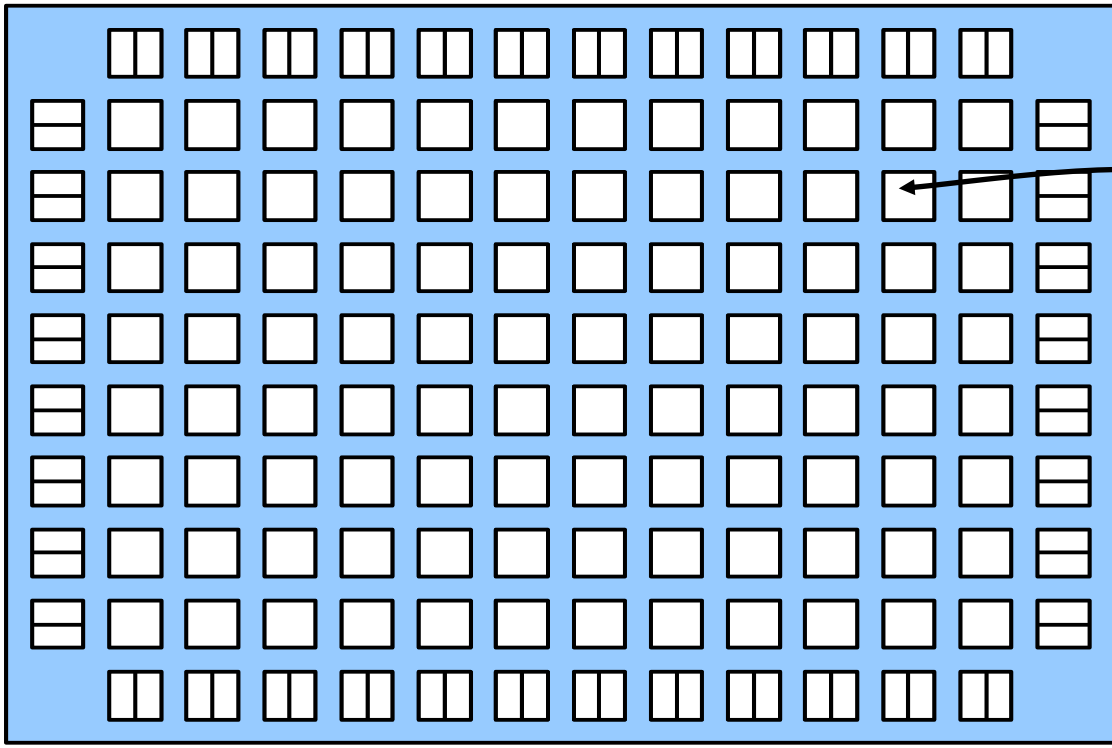
THE UNIVERSITY OF SYDNEY

Introduction

- Architecture of an “Island-style” FPGA and how CAD tools map to them
  - Homogeneous model – all logic elements are of a single type
  - FPGA only consists of BLEs and programmable routing
  - Commercial FPGAs have embedded blocks which differentiate products, choice of FPGA strongly influenced by availability of embedded blocks and IP cores
- Case study
  - Architectural Exploration

Island-style FPGAs

Island-style FPGA – Logic Block



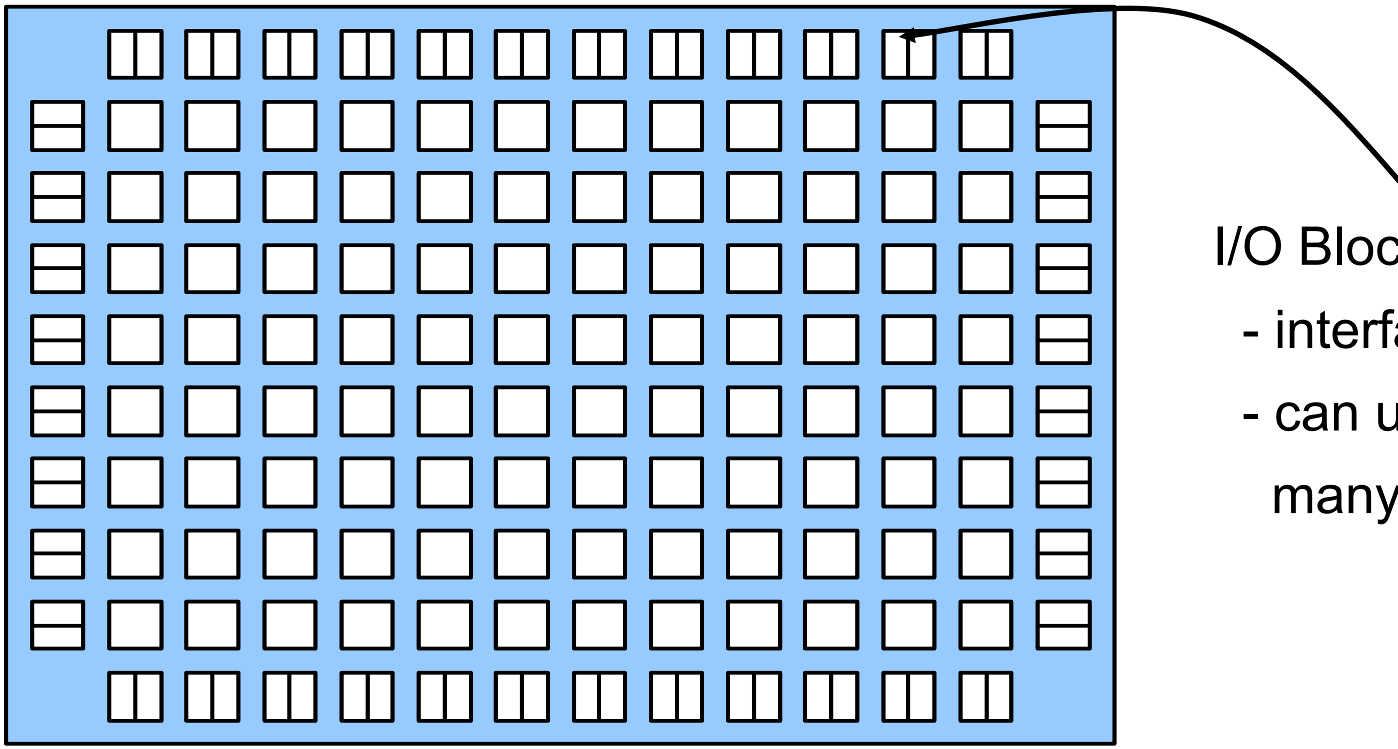
Logic Blocks

- used to implement logic
- lookup tables & flip-flops

Altera: LABs

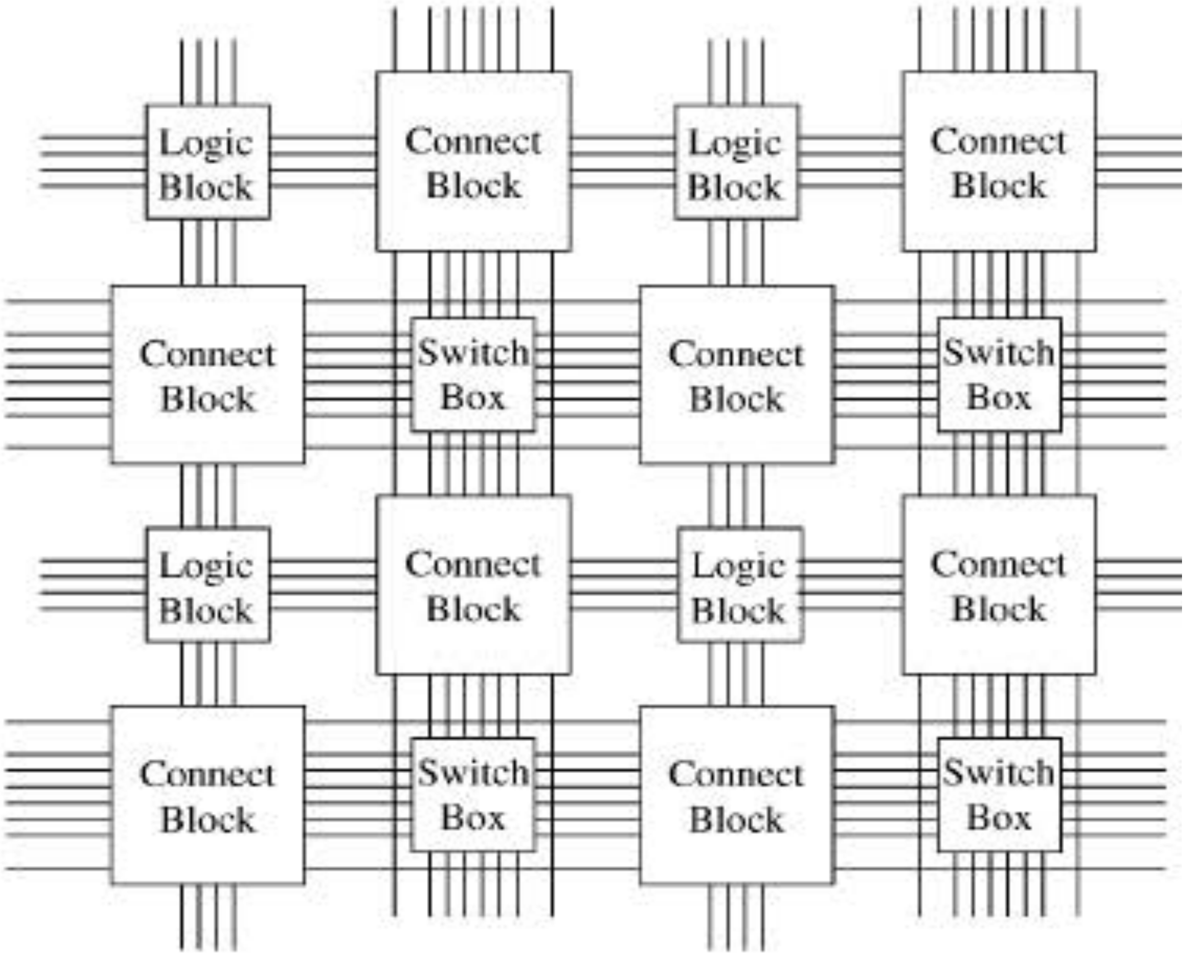
Xilinx: CLBs

Island-style FPGA – I/O Block



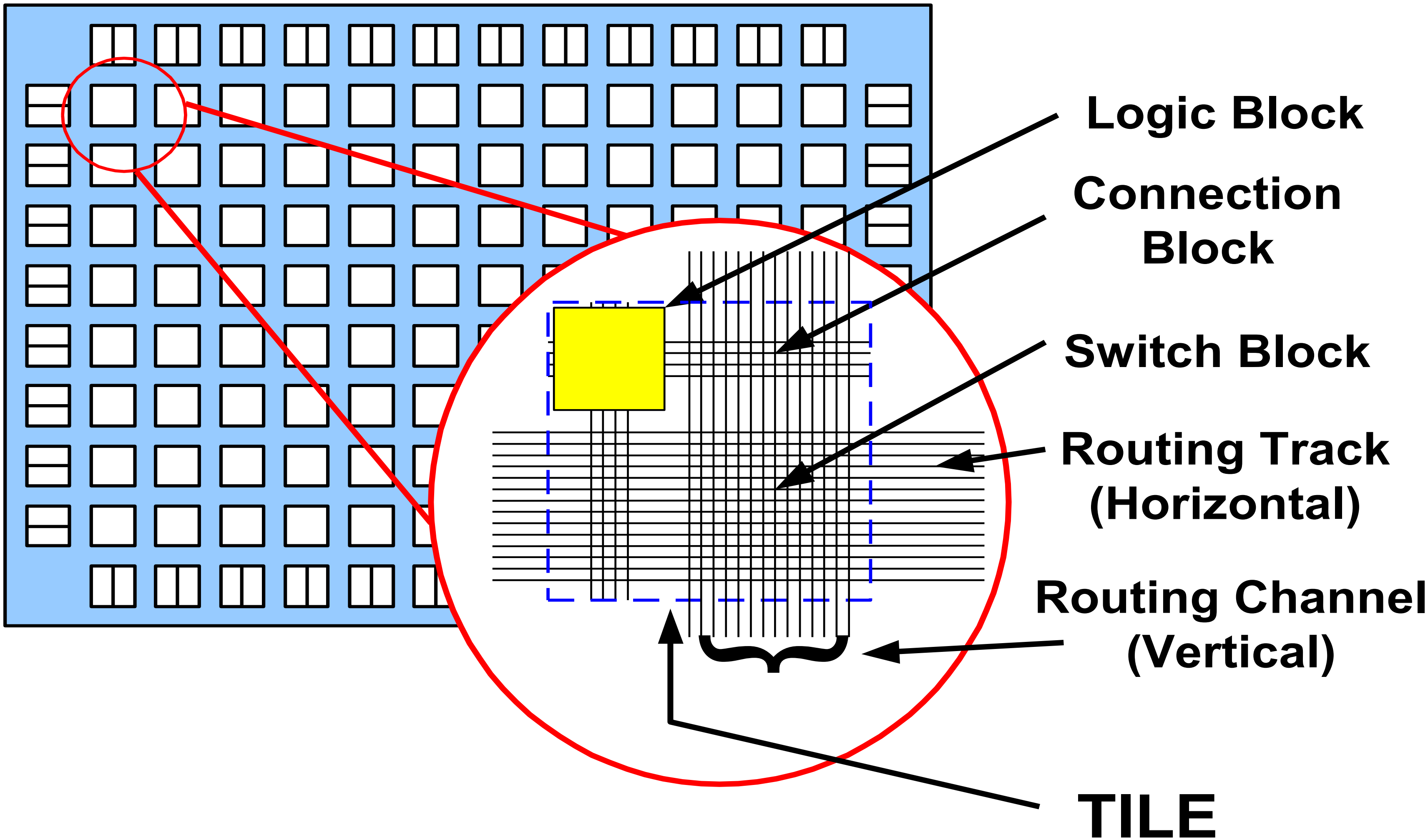
I/O Blocks
- interface off-chip
- can usually support many I/O Standards

FPGA Blocks



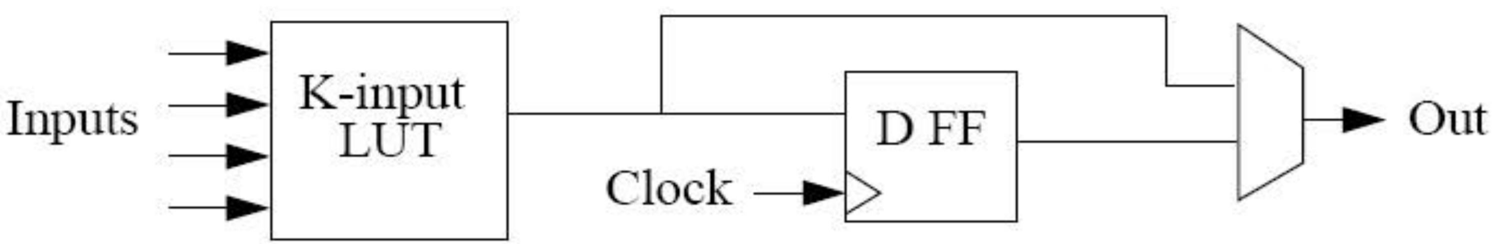
From Compton & Hauck, ACM Comput. Surv., V34, No. 2, June 2002

Tile


Logic Block
Connection Block
Switch Block
Routing Track (Horizontal)
Routing Channel (Vertical)
TILE

Basic Logic Element

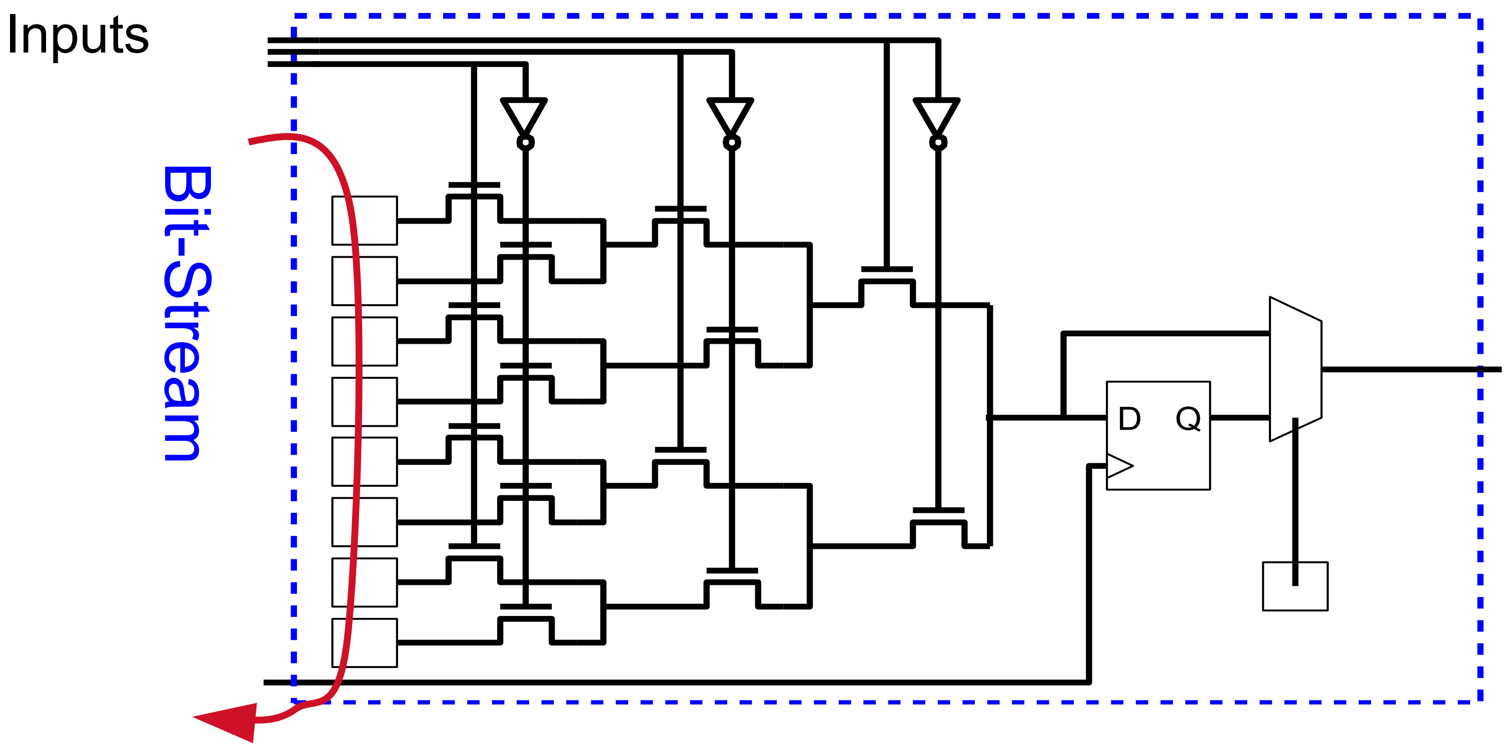
- Logic block and logic element mean the same thing
- The lookup table size is K
  - What are the consequences of this being too big/small?



(a) Basic logic element (BLE)

BLE Implementation

Basic Logic Gate: Lookup-Table



Function of each lookup table can be configured by shifting in bit-stream.

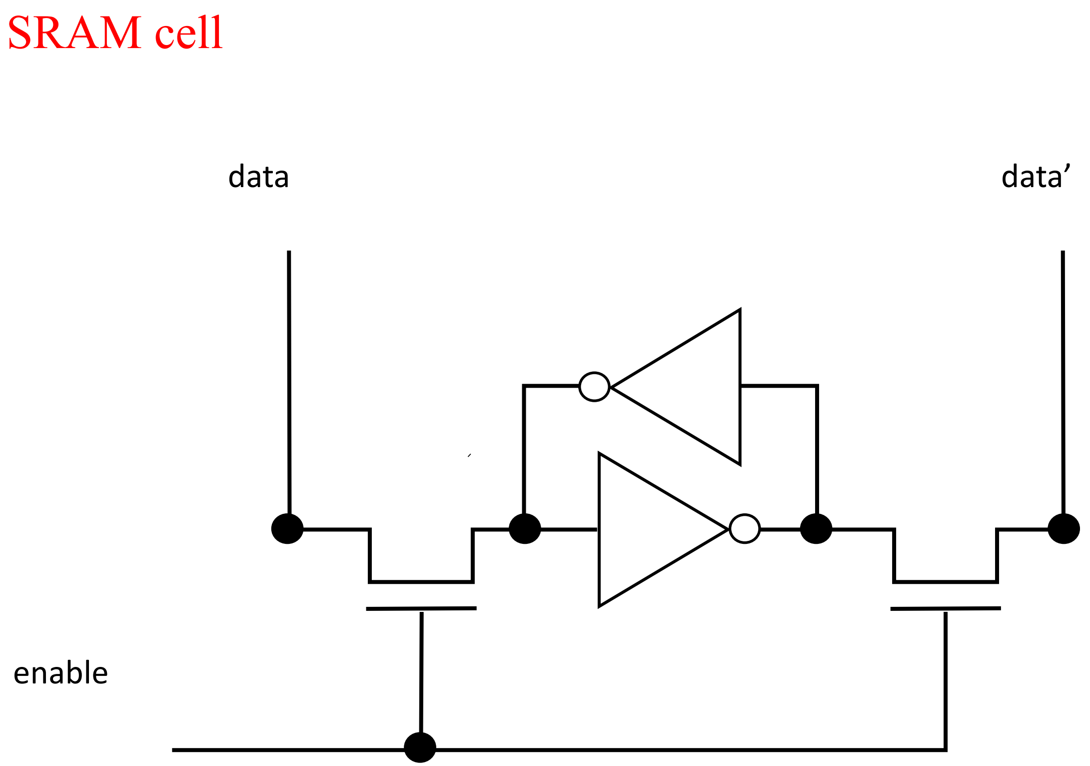
Exercise

- Show how we can implement A+B.C with the LUT in the previous slide
- How many of the following does a K-input LUT use?
  - SRAM cells
  - MUX pass transistors
  - MUX select buffers?

Exercise

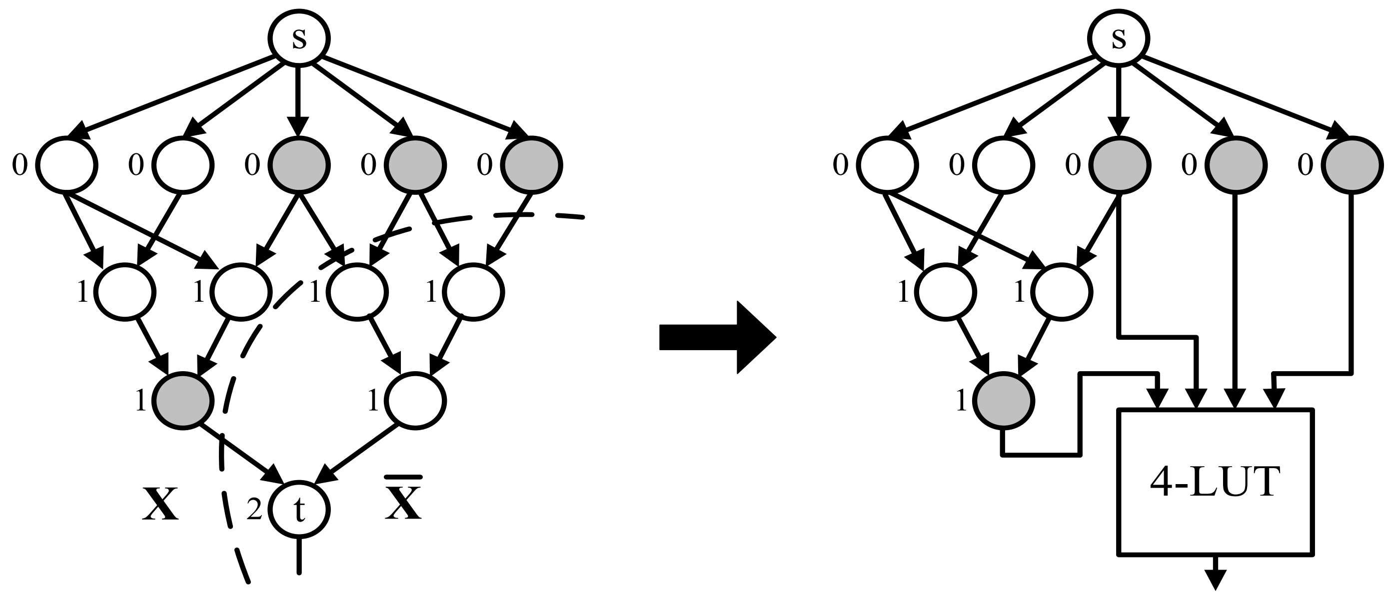
- Show how we can implement A+B.C with the LUT in the previous slide
- How many of the following does a K-input LUT use?
  - SRAM cells (2^K)
  - MUX pass transistors ($\sum_{i=1}^{k} 2^i$)
  - MUX select buffers? (K)

SRAM Cell



- How many transistors?

› Mapping gates to LUTs:

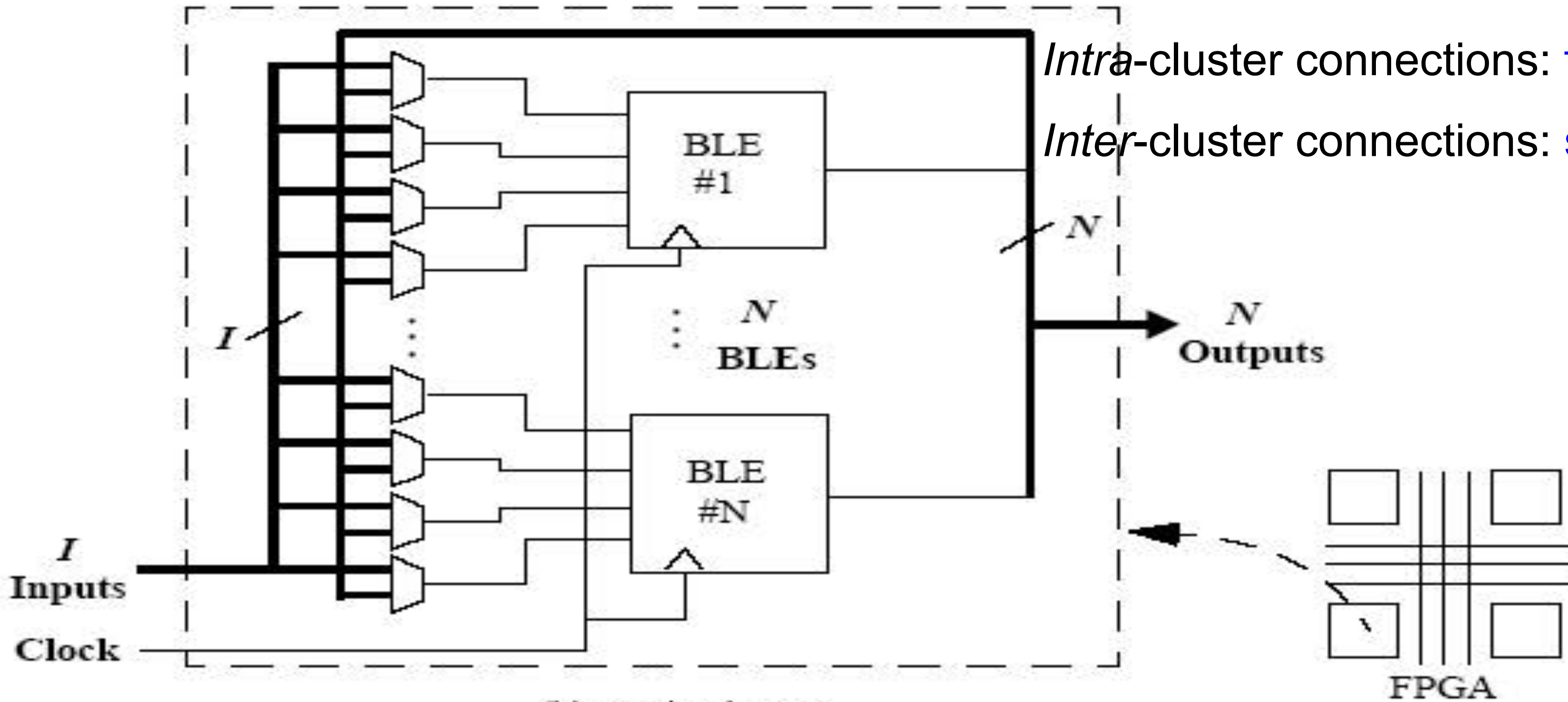


› Depth-optimal mapping

Logic Cluster

Intra-cluster connections: fast

Inter-cluster connections: slow

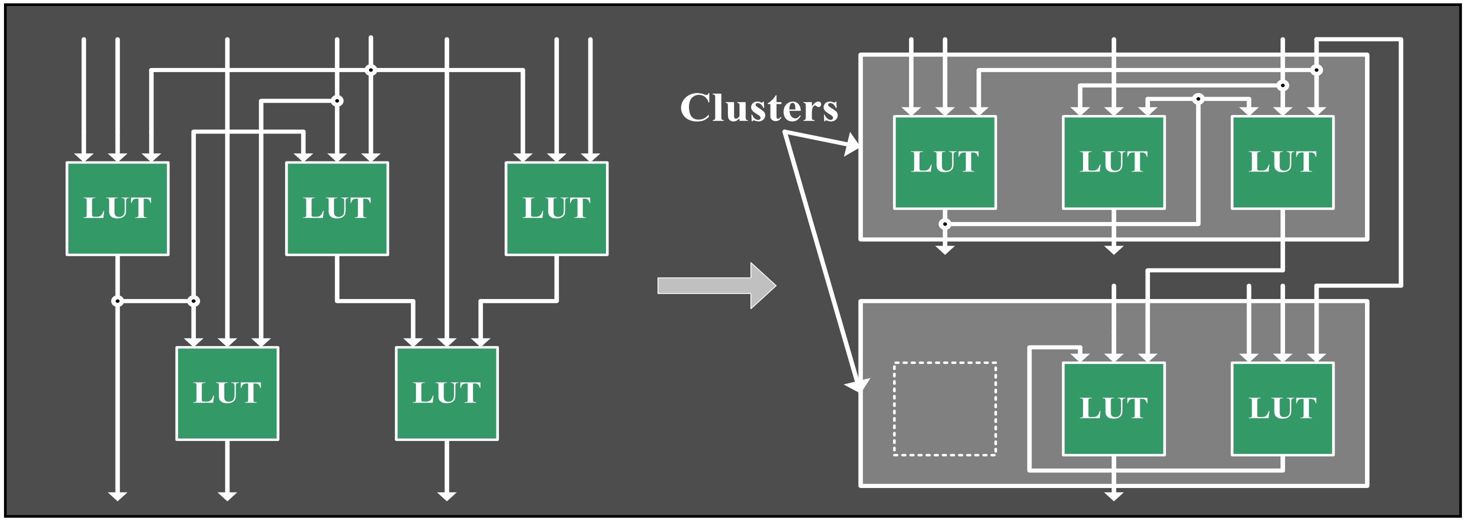


(b) Logic cluster

Cluster accepts I inputs and consists of N basic logic elements with multiplexed inputs

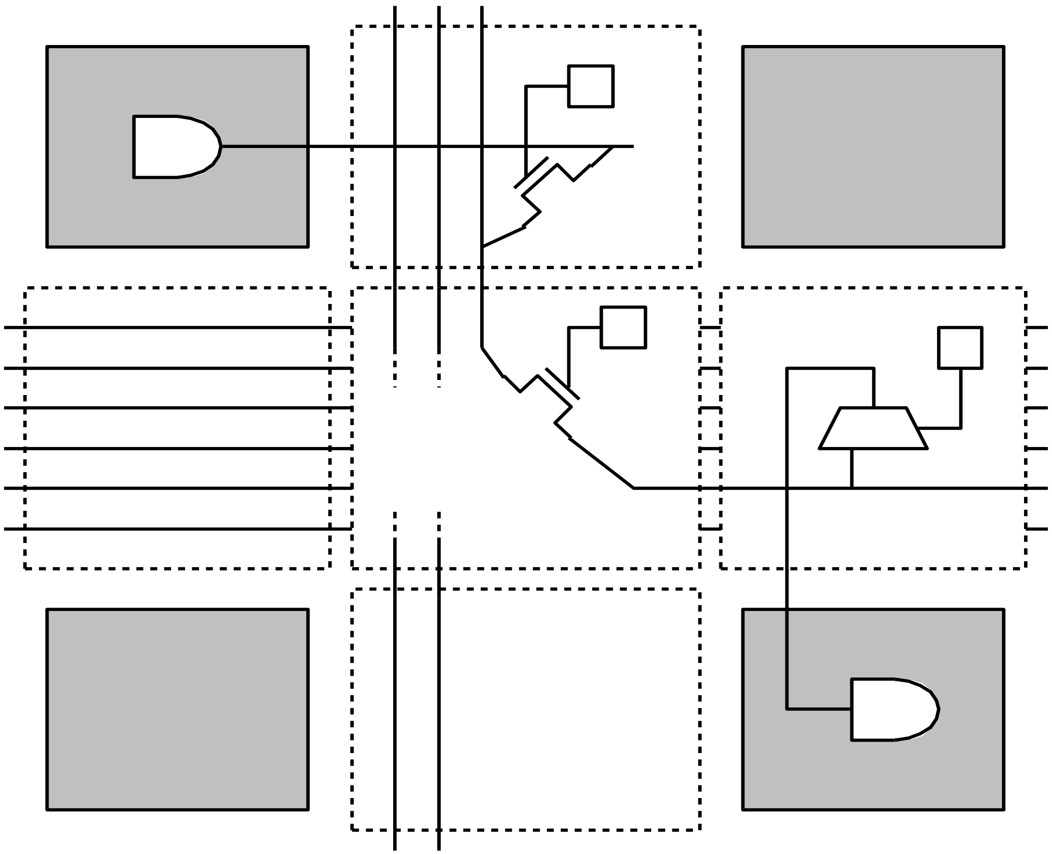
Clustering Algorithms

- FPGA logic blocks (LABs, CLB's) usually contain several LUTs:



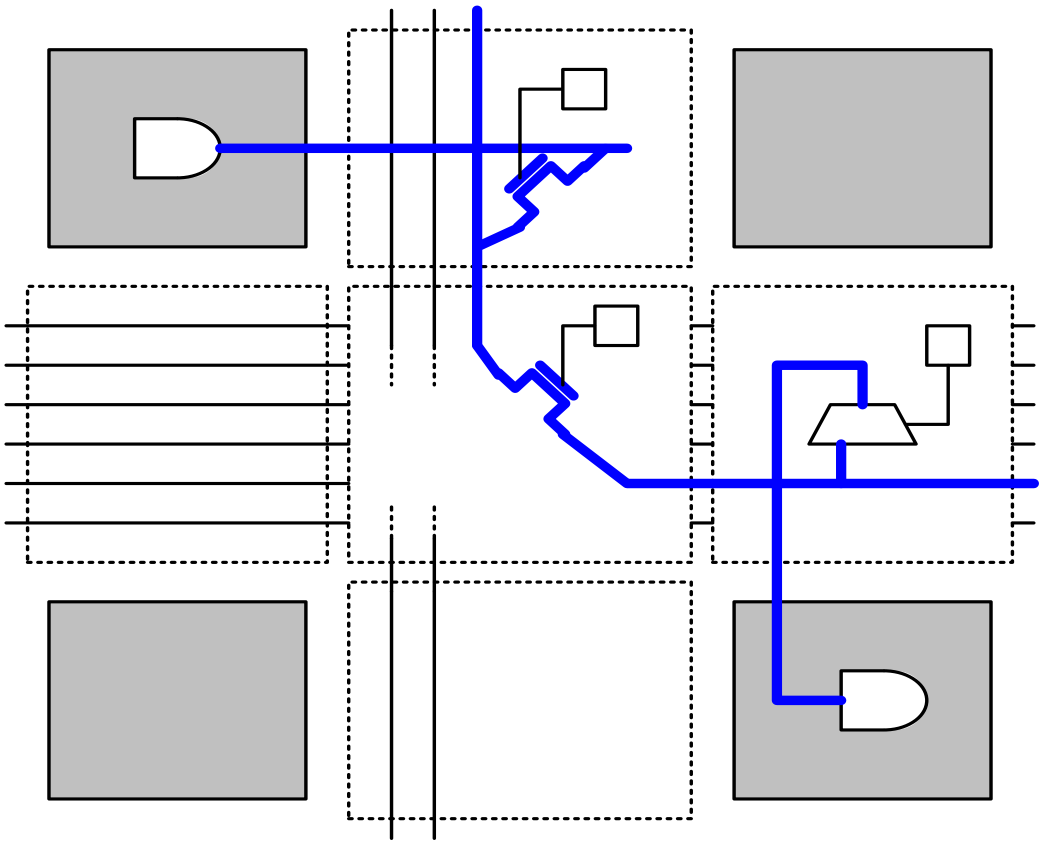
- Clustering groups LUTs into LAB-sized clusters
  - Idea: try to encapsulate as much activity inside each cluster as possible

Configurable Routing

- 

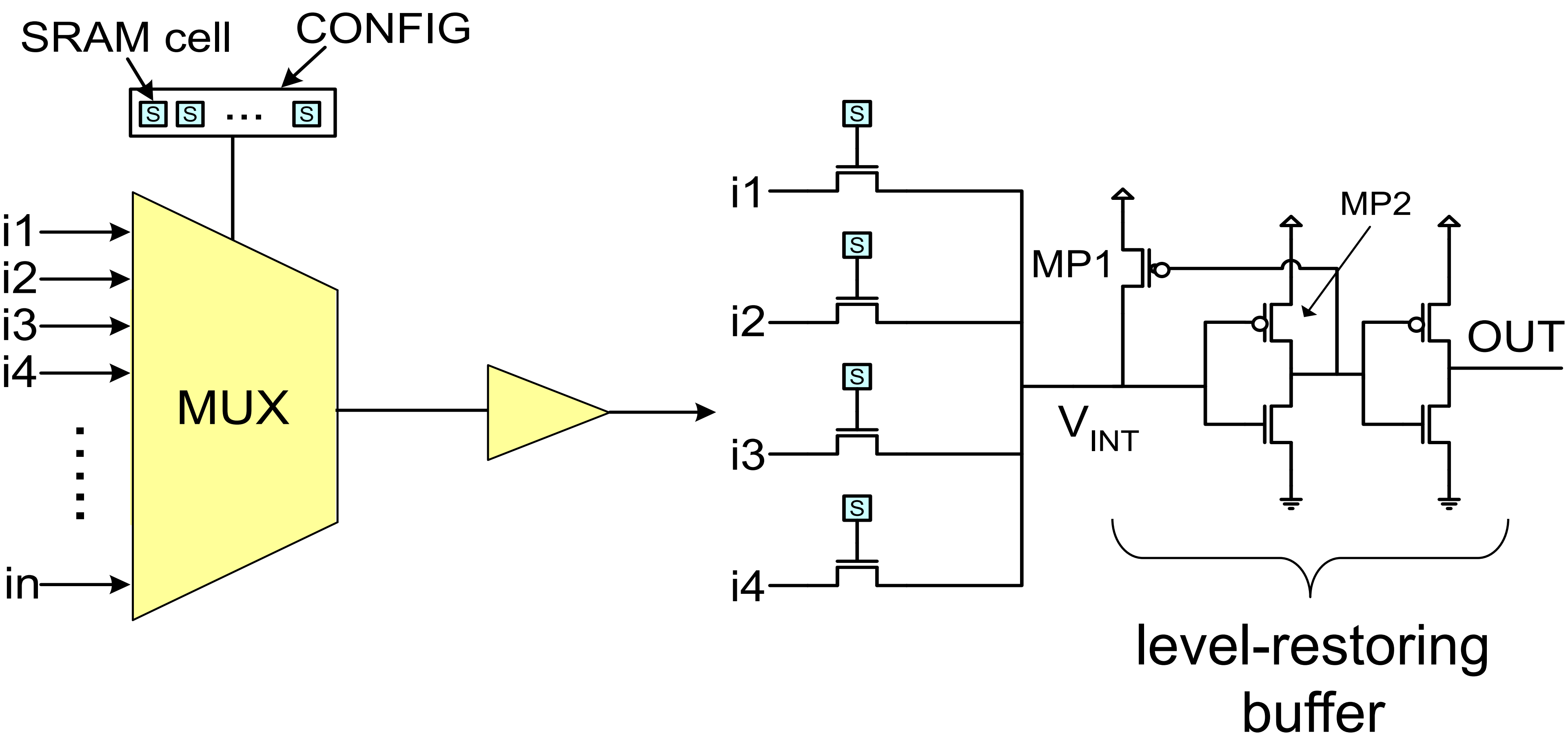
Connect Logic Blocks using Fixed Metal Tracks and Programmable Switches

Configurable Routing



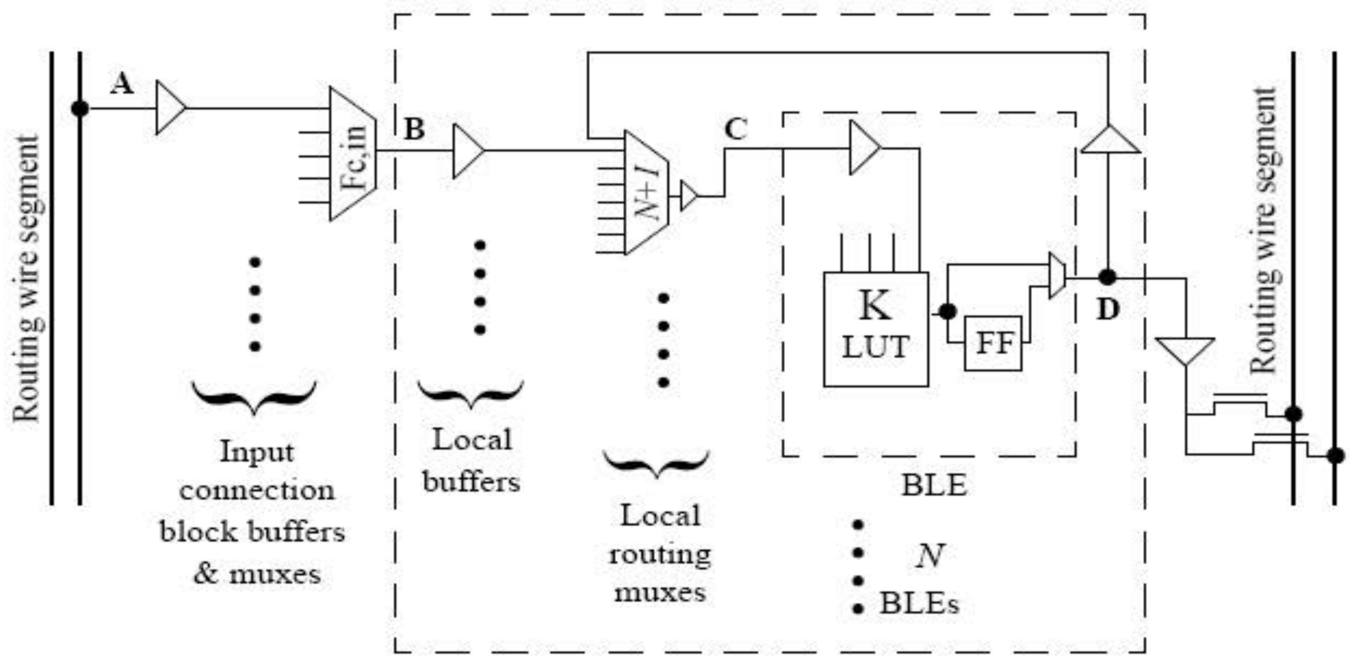
Connect Logic Blocks using Fixed Metal Tracks and Programmable Switches

Routing Switch



Logic Cluster Detail

› Fc,in is the number of input connections from routing to cluster



Case Study: Xilinx FPGAs

7 Series FPGA Overview

Source of slides that follow: Xilinx

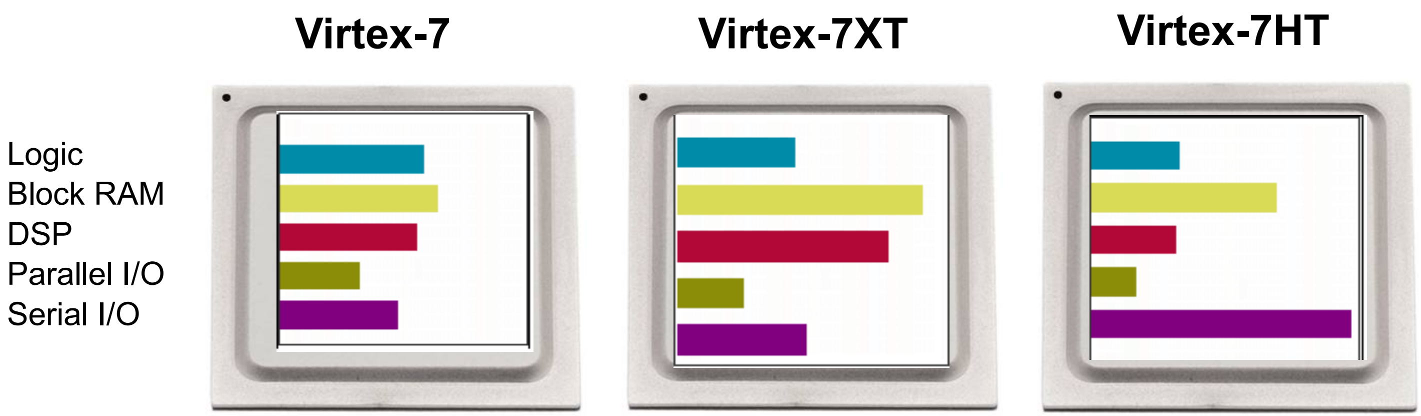
7 Series FPGA Families

Maximum Capability	ARTIX 7 — Lowest Power and Cost	KINTEX 7 — Industry's Best Price/Performance	VIRTEX 7 — Industry's Highest System Performance
Logic Cells	20K – 355K	70K – 480K	285K – 2,000K
Block RAM	12 Mb	34 Mb	65 Mb
DSP Slices	40 – 700	240 – 1,920	700 – 3,960
Peak DSP Perf.	504 GMACS	2,450 GMACs	5,053 GMACS
Transceivers	4	32	88
Transceiver Performance	3.75Gbps	6.6Gbps and 12.5Gbps	12.5Gbps, 13.1Gbps and 28Gbps
Memory Performance	1066Mbps	1866Mbps	1866Mbps
I/O Pins	450	500	1,200
I/O Voltages	3.3V and below	3.3V and below 1.8V and below	3.3V and below 1.8V and below

Source: Xilinx

Virtex-7 Devices

- The Virtex-7 family has several devices
  - Virtex-7: General logic
  - Virtex-7XT: Rich DSP and block RAM, higher serial bandwidth
  - Virtex-7HT: Highest serial bandwidth



Virtex-7	Virtex-7XT	Virtex-7HT
• High Logic Density • High-Speed Serial Connectivity	• High Logic Density • High-Speed Serial Connectivity • **Enhanced DSP**	• High Logic Density • **Ultra High-Speed Serial Connectivity**

Source: Xilinx

Architecture Alignment

- Common elements enable easy IP reuse for quick design portability across all 7 series families
  - Design scalability from low-cost to high-performance
  - Expanded eco-system support
  - Quickest TTM

Logic Fabric
LUT-6 CLB

Precise, Low Jitter Clocking
MMCMs

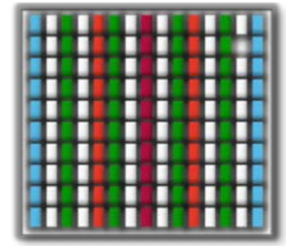
On-Chip Memory
36Kbit/18Kbit Block RAM

Enhanced Connectivity
PCIe® Interface Blocks

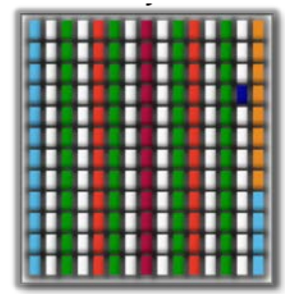
DSP Engines
DSP48E1 Slices

Hi-perf. Parallel I/O Connectivity
SelectIO™ Technology

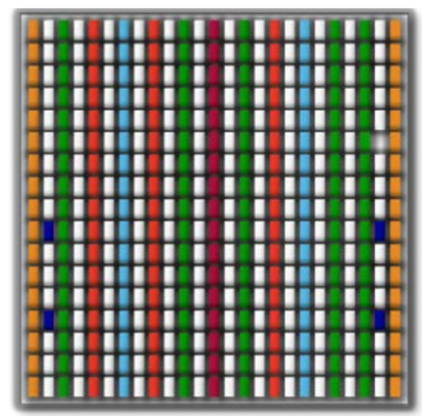
Hi-performance Serial I//O Connectivity
Transceiver Technology



Artix™-7 FPGA



Kintex™-7 FPGA

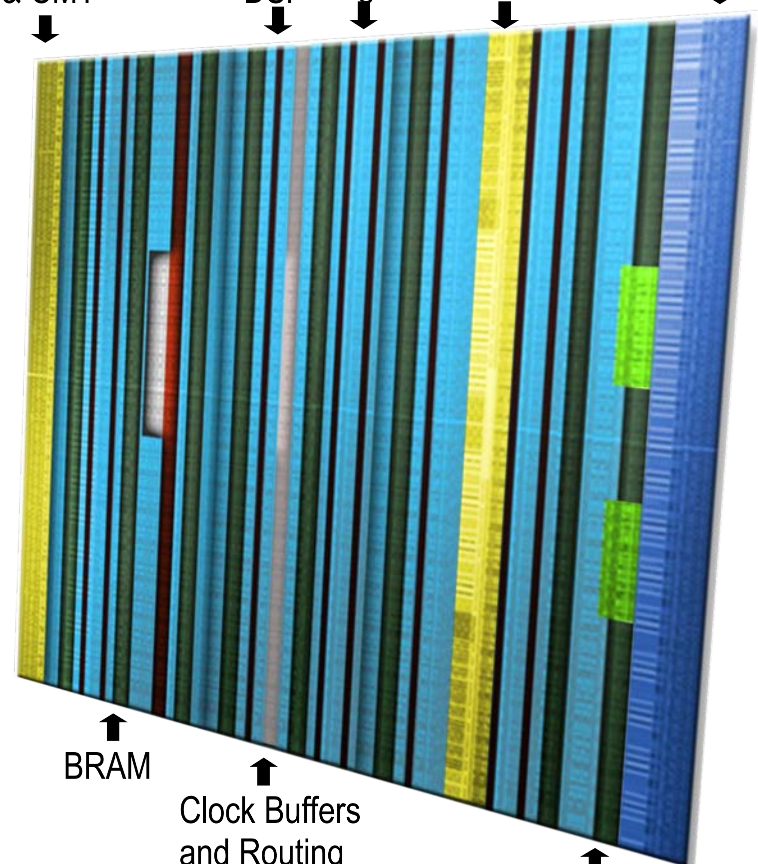


Virtex®-7 FPGA

Source: Xilinx

Fourth-Generation ASMBL Architecture

- Optimized FPGA feature mix for different families/members
  - FPGA comprises columns of different resources
    - Clocking, I/O, BRAM, DSP, HSSIO
- Enables the unified architecture between the different 7 series families
- Enables different resource ratios within the different devices

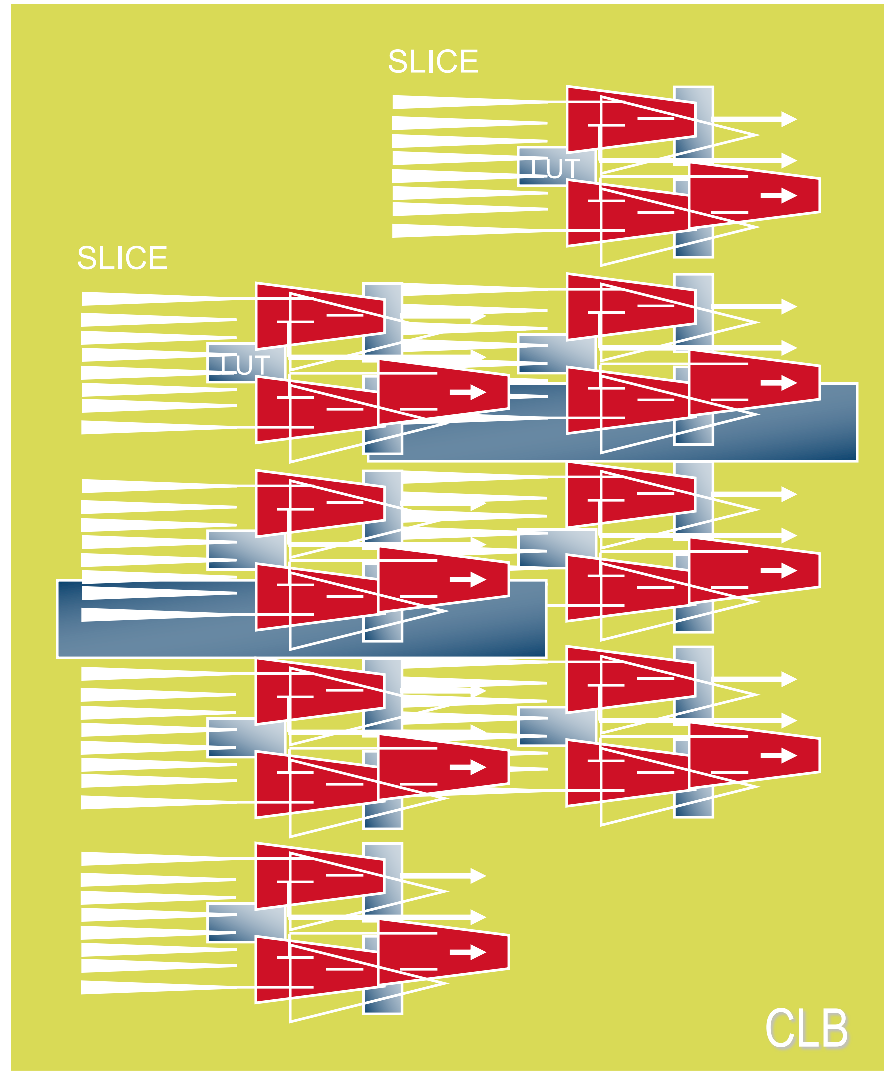


CMT=clock management tile, HSSIO=high speed serial I/O

Source: Xilinx

CLB Structure

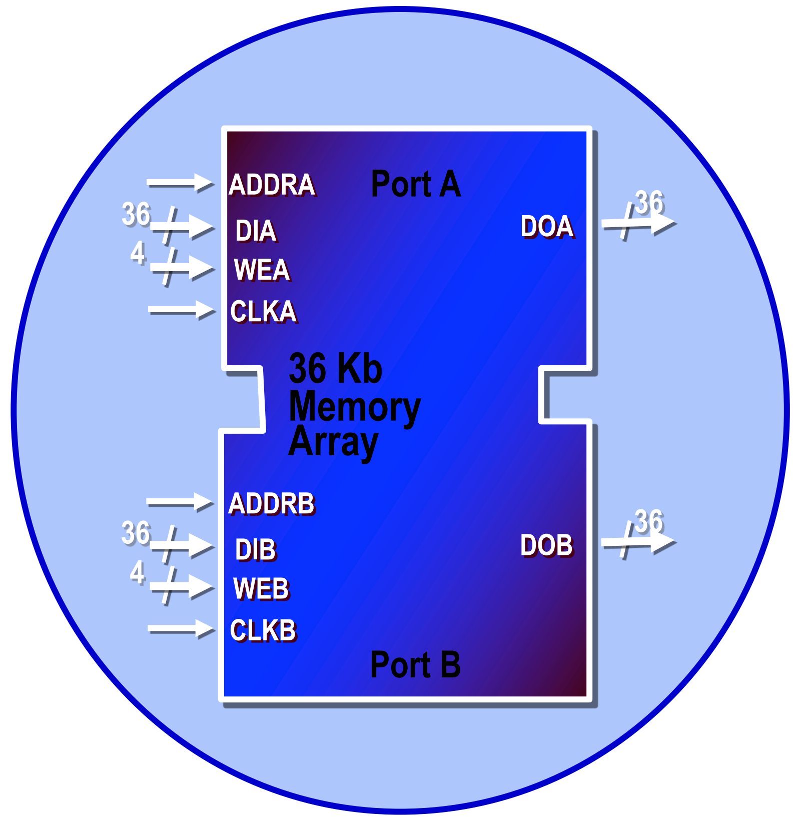
- **Two side-by-side slices per CLB**
  - Slice_M are memory-capable
  - Slice_L are logic and carry only
- **Four 6-input LUTs per slice**
  - Consistent with previous architectures
  - Single LUT in Slice_M can be a 32-bit shift register or 64 x 1 RAM
- **Two flip-flops per LUT**
  - Excellent for heavily pipelined designs
  -



Source: Xilinx

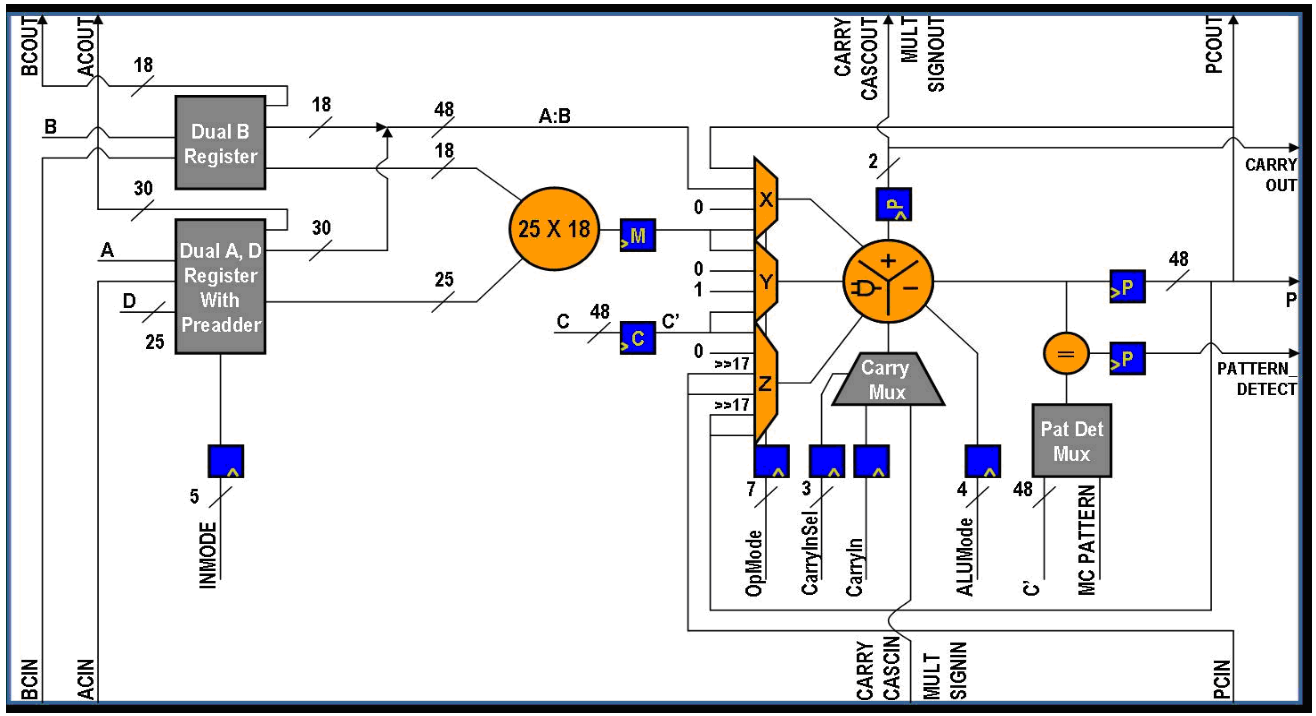
Block RAM

- **36K/18K block RAM**
  - All Xilinx 7 series FPGA families use same block RAM as Virtex-6 FPGAs
- **Configurations same as Virtex-6 FPGAs**
  - 32k x 1 to 512 x 72 in one 36K block
  - Simple dual-port and true dual-port configurations
  - Built-in FIFO logic
  - 64-bit error correction coding per 36K block
  - Adjacent blocks combine to 64K x 1 without extra logic



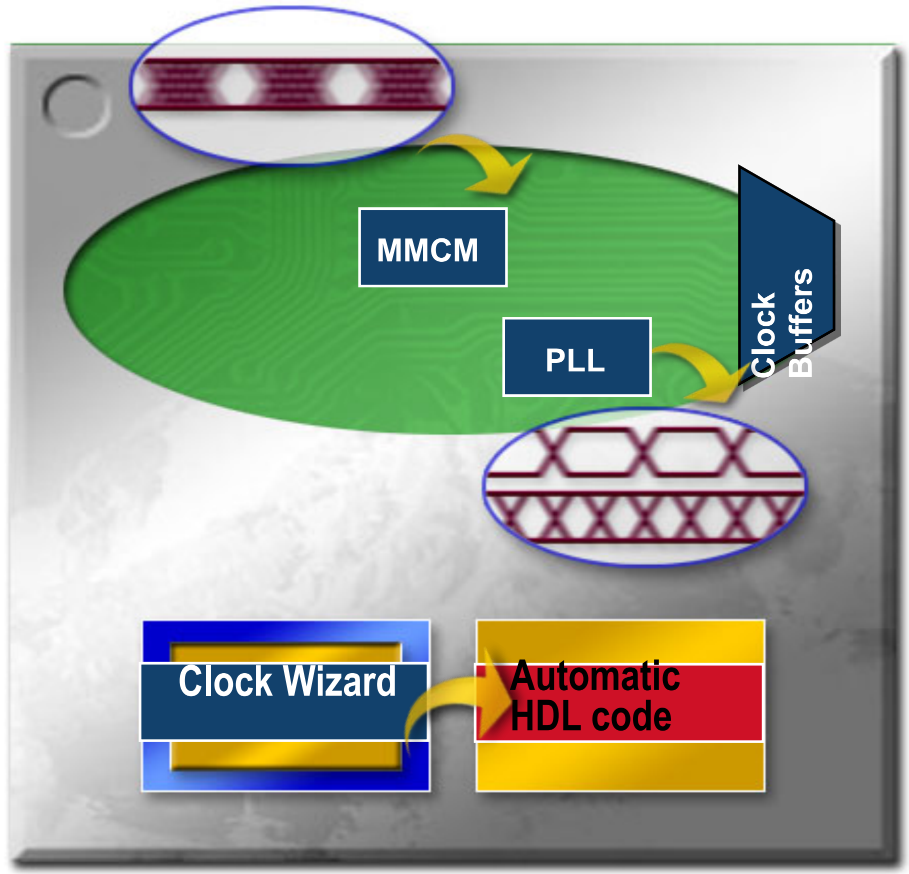
DSP Slice

- All 7 series FPGAs share the same DSP slice
  - 25x18 multiplier
  - 25-bit pre-adder
  - Flexible pipeline
  - Cascade in and out
  - Carry in and out
  - 96-bit MACC
  - SIMD support
  - 48-bit ALU
  - Pattern detect
  - 17-bit shifter
  - Dynamic operation (cycle by cycle)



Clocking Resources

- **Based on the established Virtex-6 FPGA clocking structure**
  - All 7 series FPGAs use the same unified architecture
- **Low-skew clock distribution**
  - Combination of paths for driving clock signals to and from different locations
- **Clock buffers**
  - High fanout buffers for connecting clock signals to the various routing resources
- **Clock regions**
  - Device divided into clock regions with dedicated resources
- **Clock management tile (CMT)**
  - One MMCM and one PLL per CMT
  - Up to 24 CMTs per device

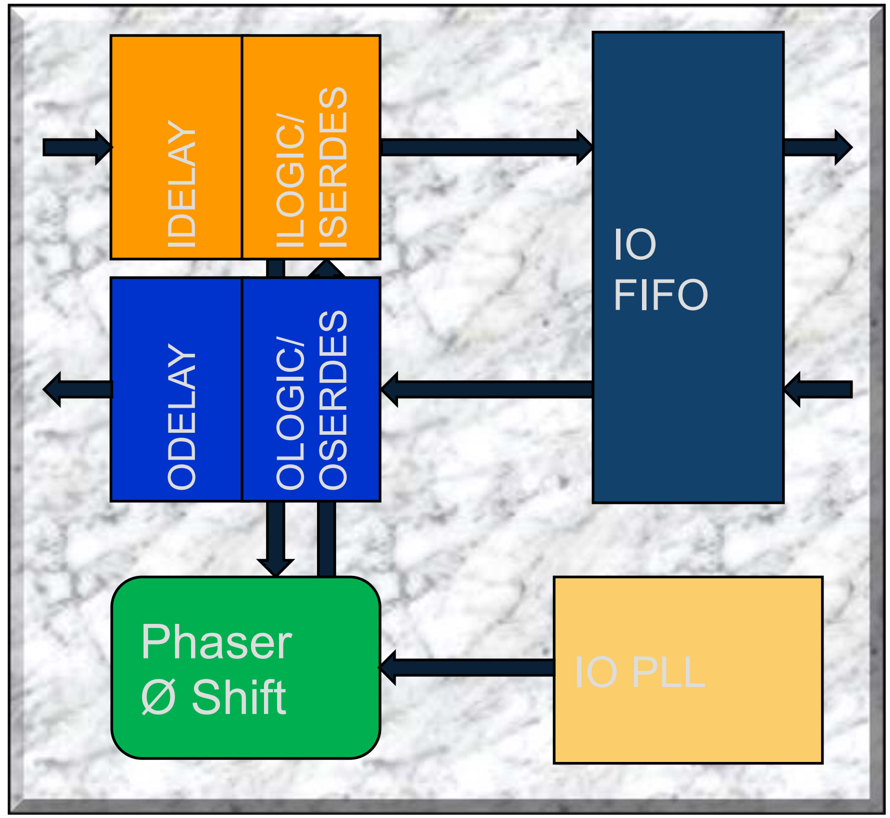


MMCM = mixed mode clock manager

Source: Xilinx

Input/Output Blocks

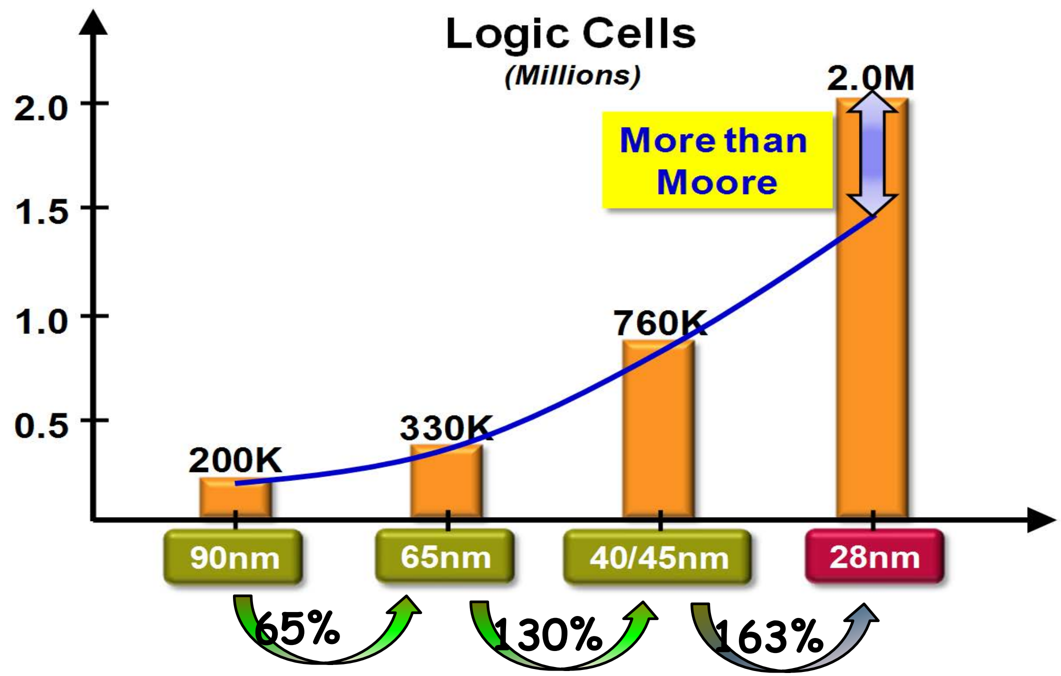
- Two distinct I/O types
  - High range: Supports standards up to 3.3V
  - High performance: Higher performance with more I/O delay capability
    - Supports I/O standards up to 1.8V
- Extension of logic layer functionality
  - Wider input/output SERDES
  - Addition of independent ODELAY
- New hardware blocks to address highest I/O performance
  - Phaser, IO FIFO, IO PLL



Source: Xilinx

Stacked Silicon Interconnect Technology

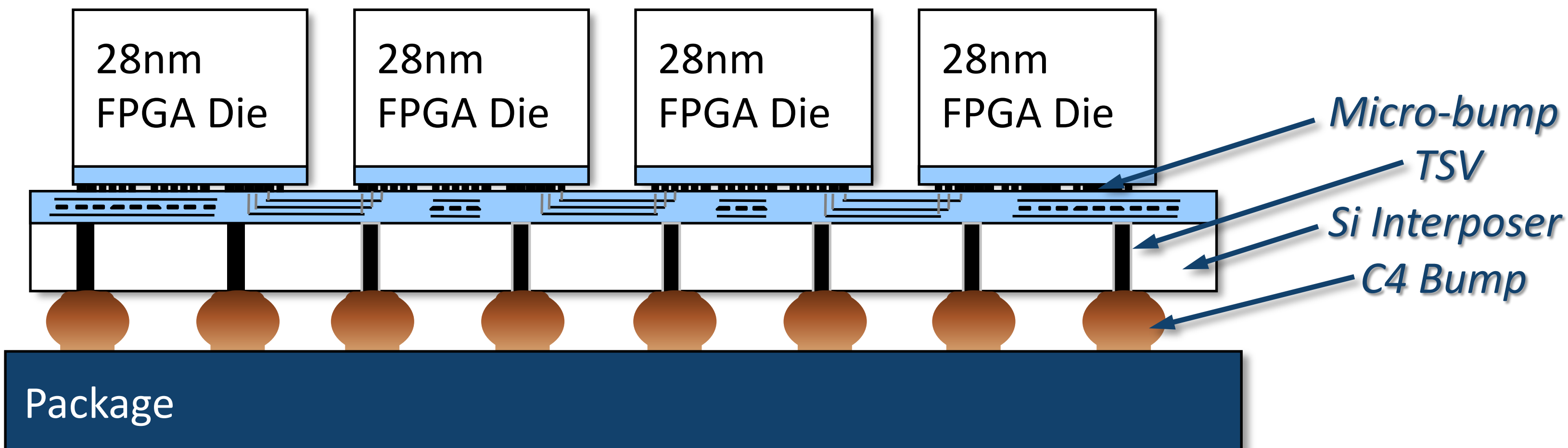
- **Largest Virtex-7 device is almost three times the size of the largest Virtex-6 device**
  - Growth is higher than Moore's Law dictates
- **Enabled by Stacked Silicon Interconnect (SSI) technology**
  - Multiple FPGA die on a silicon interposer
  - Each die is referred to as a Super Logic Region (SLR)
  - Vast quantity of interconnect between adjacent SLRs are provided by the interposer



Source: Xilinx

Stacked Silicon Implications

- Enables substantially larger devices
- Device is treated as a single monolithic device
  - Tool chains place and route complete device as if it was one die
- Minor design considerations around clocking and routing

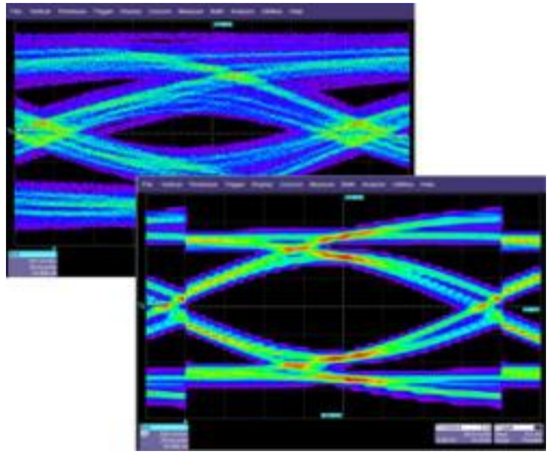


TSV=through silicon via, c4=controlled collapse chip connection

Source: Xilinx

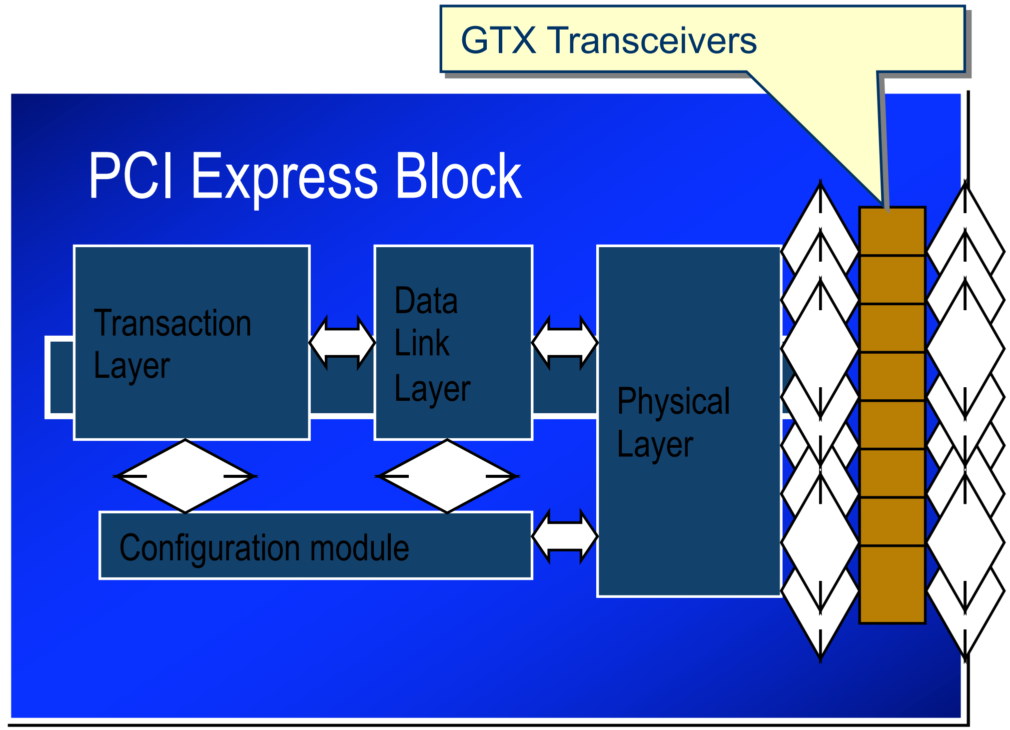
High-Speed Serial I/O Transceivers

- **Available in all families**
- **GTP transceivers – up to 3.75 Gbps**
  - Ultra high volume transceiver
  - Wire bond package capable
- **GTX transceivers – up to 12.5 Gbps**
  - Support for the most common 10 Gbps protocols
- **GTH transceivers – up to 13.1 Gbps**
  - Support for 10 Gbps protocols with high forward error correction overhead
- **GTZ transceivers – up to 28 Gbps**
  - Enables next generation 100–400Gbps system line cards



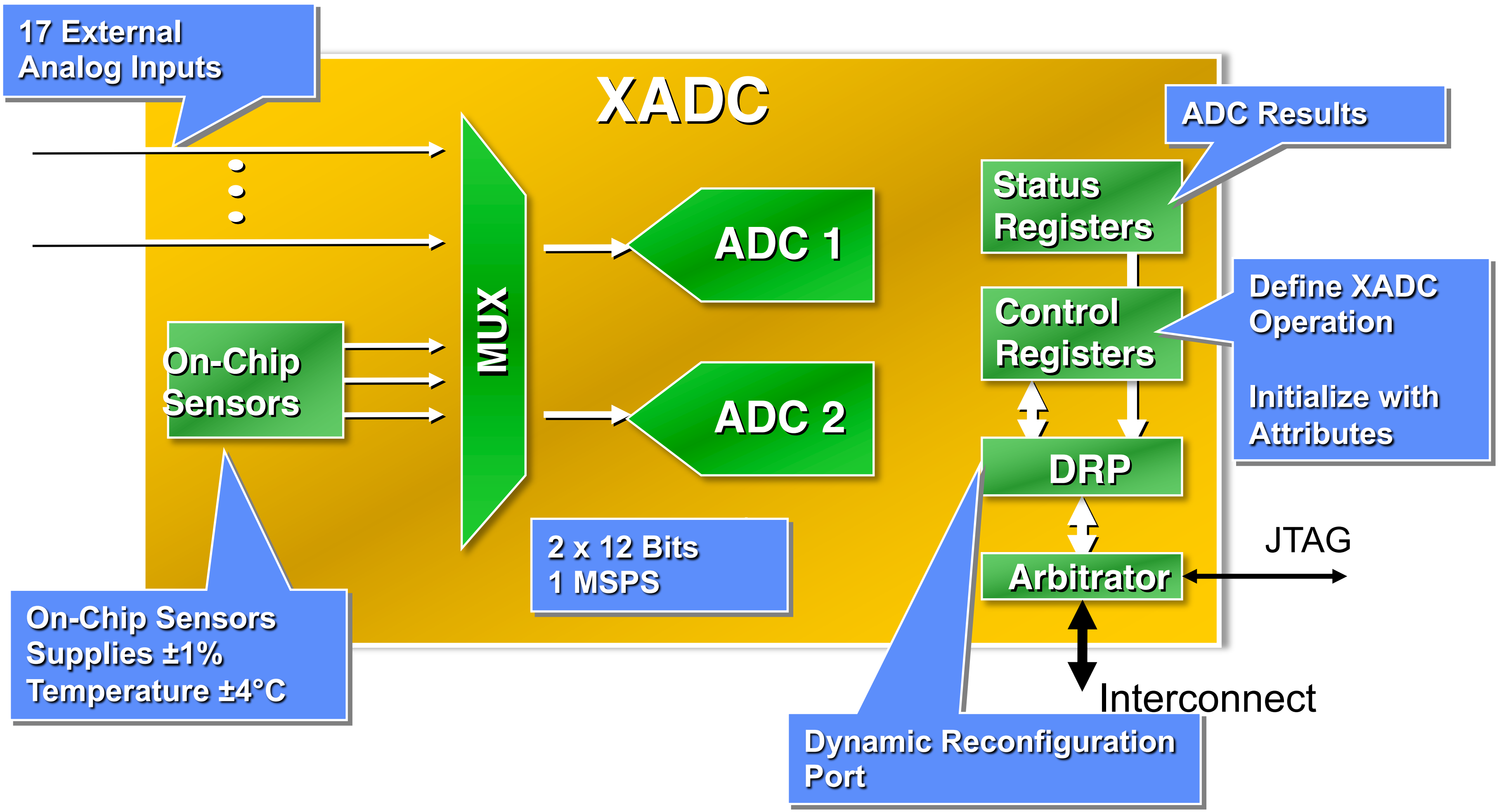
PCI Express

- **Features**
  - Compliant to PCIe Revision 2.1
  - Endpoint & root port
  - AXI user interface
  - <100 ms configuration*
  - FPGA configuration over PCI Express*
  - End-to-end CRC*
  - Advanced error reporting*
  - 100-MHz clocking
- **New wrappers**
  - Multi-function*
  - Single-root I/O virtualization*
- **Configurations**
  - Lane widths: x1-8
  - Data rates: Gen1 & Gen2 (2.5/5.0 Gbps)
  - Dependent on GT and fabric speed



***New features in 7 series**

XADC: Dual 12-Bit 1-MSPS ADCs

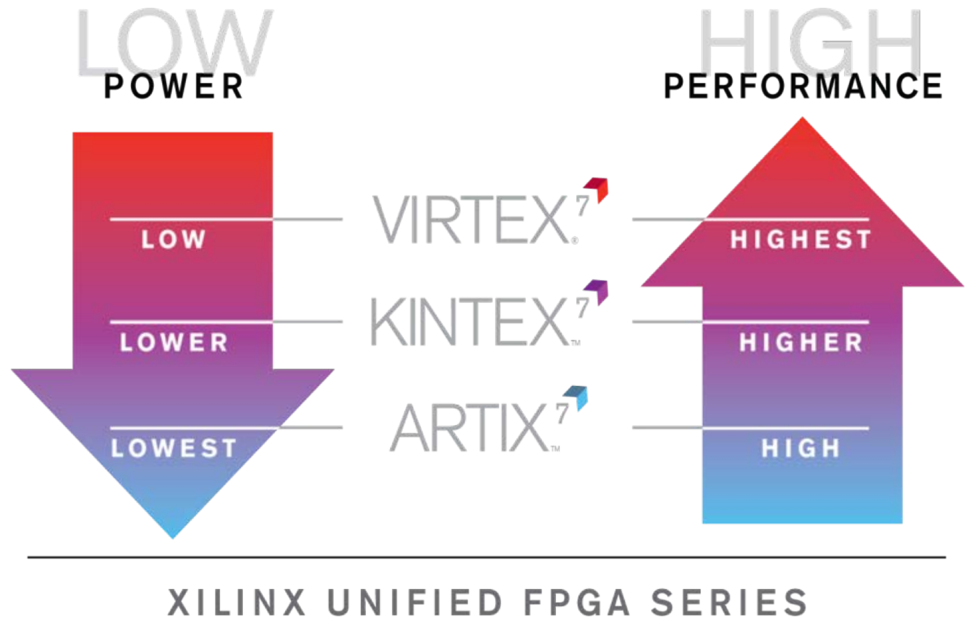


Source: Xilinx

THE UNIVERSITY OF SYDNEY

Cost, Power, and Performance

- **The different families in the 7 series provide solutions to address the different price/performance/power requirements of the FPGA market**
  - Artix-7 family: Lowest price and power for high volume and consumer applications
    - Battery powered devices, automotive, commercial digital cameras
  - Kintex-7 family: Best price/performance
    - Wireless and wired communication, medical, broadcast
  - Virtex-7 family: Highest performance and capacity
    - High-end wired communication, test and measurement, advanced RADAR, high performance computing



Source: Xilinx

I/O Composition

- Each 7 series I/O bank contains one type of I/O
  - High (voltage) Range (HR)
  - High Performance (HP)
- Different devices have different mixtures of I/O banks

I/O Types	Artix-7 Family	Kintex-7 Family	Virtex-7 Family	Virtex-7 XT/HT Family
High Range	All	Most	Some	
High Performance		Some	Most	All

Source: Xilinx

Multi-Gigabit Transceiver

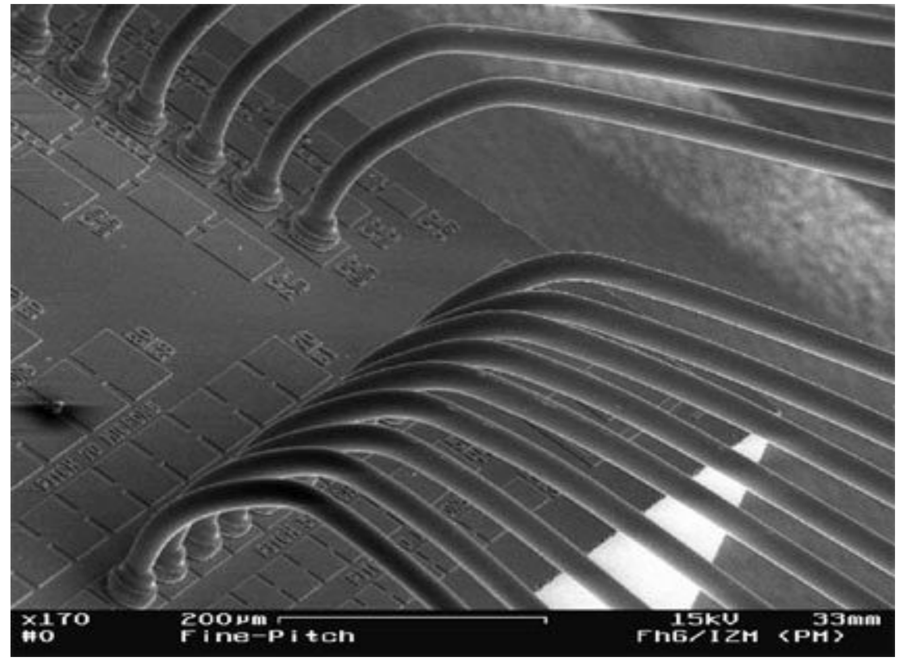
- **Different families have different MGT devices**
  - Artix-7 family: GTP
  - Kintex-7/Virtex-7 family: GTX
  - Virtex-7 XT family: Mixture of GTX and GTH
  - Virtex-7 HT family: Mixture of GTH and GTZ

Speed Grade	Artix GTP		Kintex GTX			Virtex GTX		Virtex GTH		Virtex GTZ	
	min	max	min	max	max (FF)	min	max	min	max	min	max
1LC/I	0.612	3.125	0.612	5.0	6.6	0.612	6.6	0.612	10.3125	N/A	N/A
1C/I	0.612	3.125	0.612	5.0	6.6	0.612	6.6	0.612	10.3125	TBD	TBD
2C/I	0.612	3.75	0.612	6.6	10.3125	0.612	10.3125	0.612	13.1	28.05	28.05
3C	N/A	N/A	0.612	6.6	12.5	0.612	12.5	0.612	13.1	28.05	28.05

Source: Xilinx

Packaging – Artix-7 Family

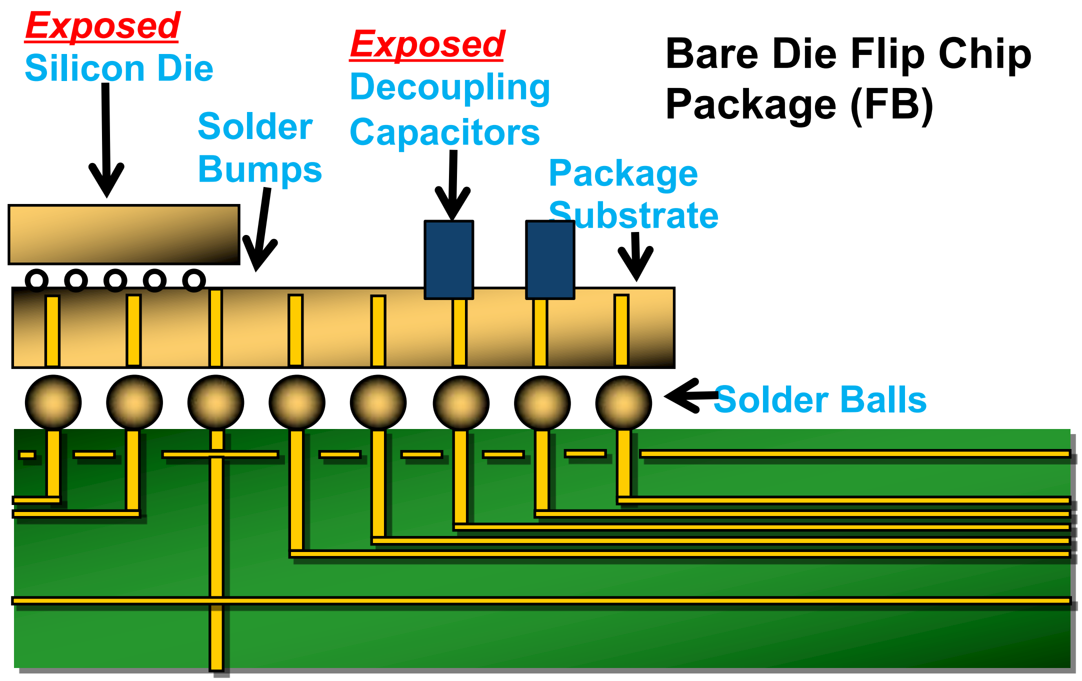
- Ultra low-cost wire bond technology
- Small form factor
- Fourth generation sparse chevron pin pattern
- Speeds up to 1.066 Gbps for parallel I/O
- Speeds up to 3.75 Gbps for MGT



Source: Xilinx

Packaging – Kintex-7 Family

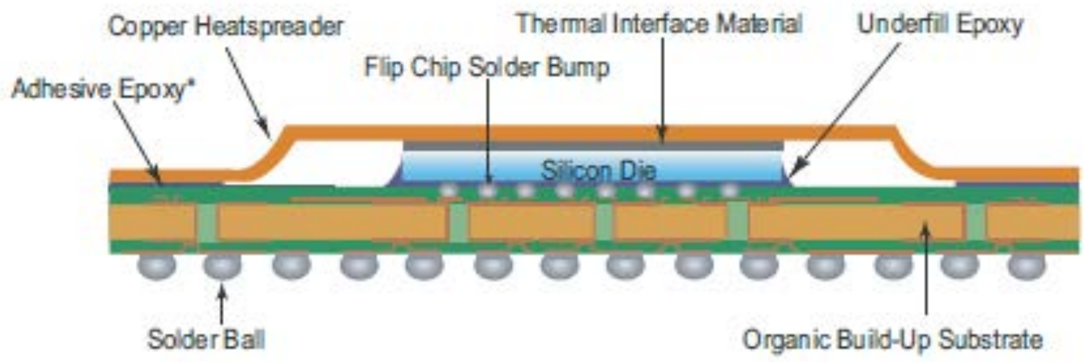
- Kintex-7 devices are available in two different packages
  - Low cost bare die flip chip (FB) and conventional flip chip (FF)
  - Small form factor packaging available
- Fourth generation sparse chevron pin pattern
- Speeds up to 2.133 Gbps for parallel I/O
- Speeds up to 12.5 Gbps for MGT in FF package, and 6.6 Gbps in FB package
- FB package has discrete substrate decoupling capacitors for MGT power supplies



Source: Xilinx

Packaging – Virtex-7 Family

- High performance flip chip (FF) package
- Fourth generation sparse chevron pin pattern
- Speeds up to 2.133 Gbps for parallel I/O
- Speeds up to 28.05 Gbps for MGT
- Discrete substrate decoupling capacitors:
  - MGT power supplies
  - Block RAM power supplies
  - I/O pre-driver power supplies



Source: Xilinx

THE UNIVERSITY OF SYDNEY

Summary

- Hard blocks needed for performance, power and low area
- Different types of FPGAs have different features to address FPGA market
  - Artix-7 family: Lowest price and power
  - Kintex-7 family: Best price/performance
  - Virtex-7 family: Highest performance/capacity

Source: Xilinx

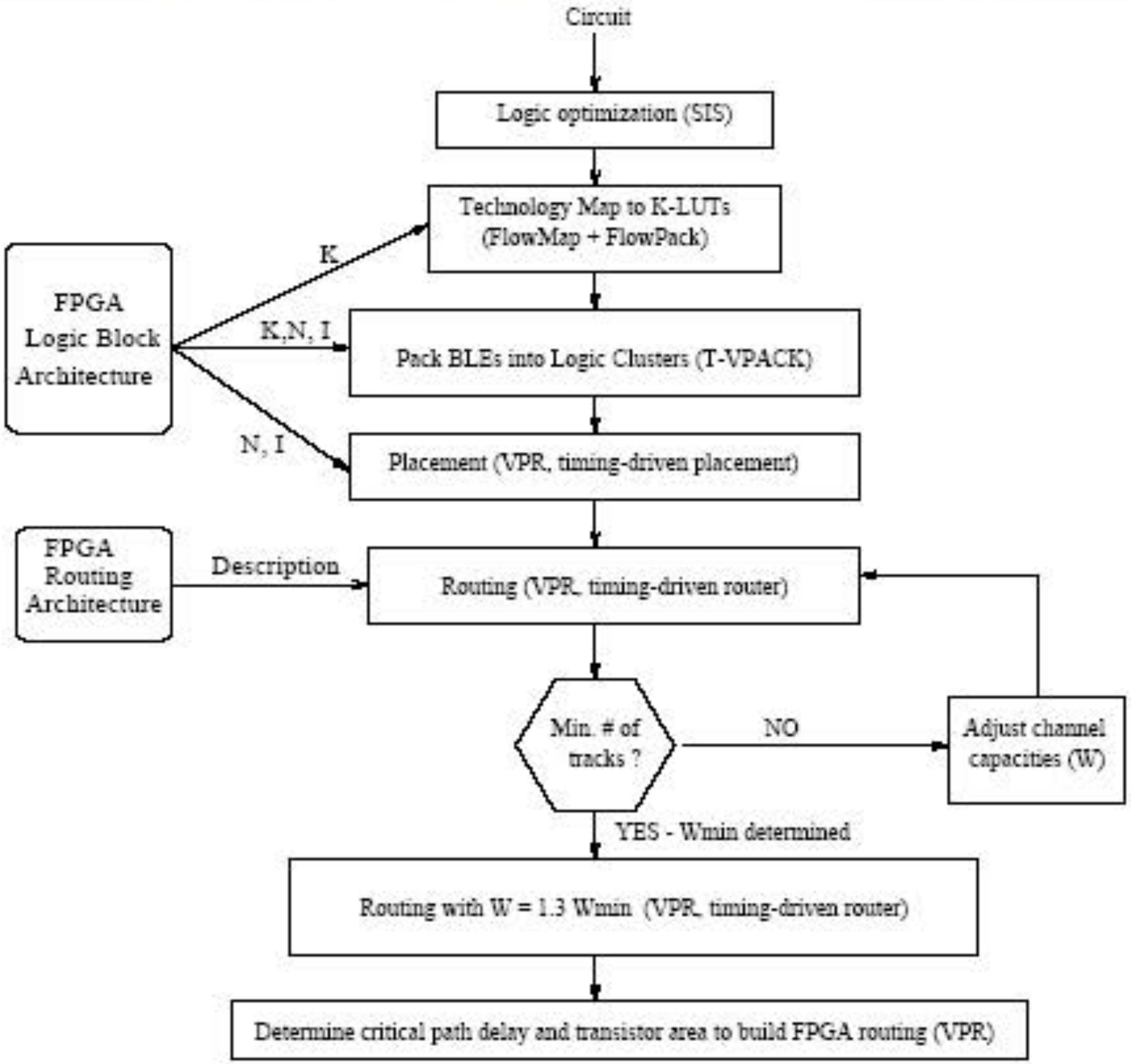
Architectural Evaluation

THE UNIVERSITY OF SYDNEY

Architectural Evaluation

- What value of I should we choose so 98% of LUTs in a cluster can be used?
- What is the effect of K and N on area and delay?
- These questions are circuit-specific so they involve an interaction of CAD tools with the architecture

Architectural Evaluation [1]



From [1]

Delay

- Spice simulations used to characterize cluster and routing delays
- Timing model in VPR updated

MCNC Benchmarks [1]

Circuit	# of 4-Input BLEs	# of Nets
alu4	1522	1536
apex2	1878	1916
apex4	1262	1271
bigkey	1707	1936
clma	8383	8445
des	1591	1847
diffeq	1497	1561
dsip	1370	1599
elliptic	3604	3735
ex1010	4598	4608
ex5p	1064	1072
frisc	3556	3576
misex3	1397	1411
pdc	4575	4591
s298	1931	1935
s38417	6406	6435
s38584.1	6447	6485
seq	1750	1791
spla	3690	3706
tseng	1047	1099
display_chip	1794	2419
img_calc	10141	10180
img_interp	2727	2769
input_chip	807	841
peak_chip	809	840
scale125_chip	2632	2654
scale2_chip	1189	1202
warping	1353	1394

I required for 98% Utilization [1]

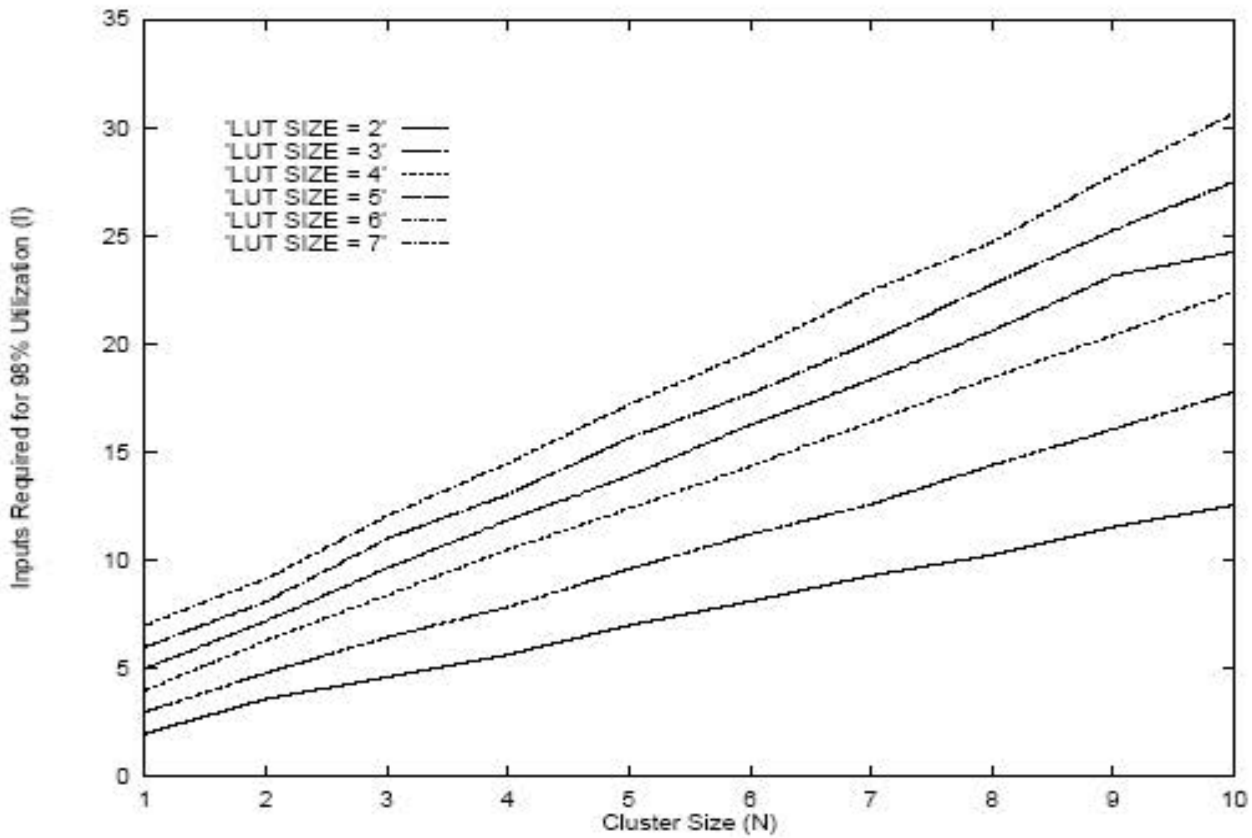
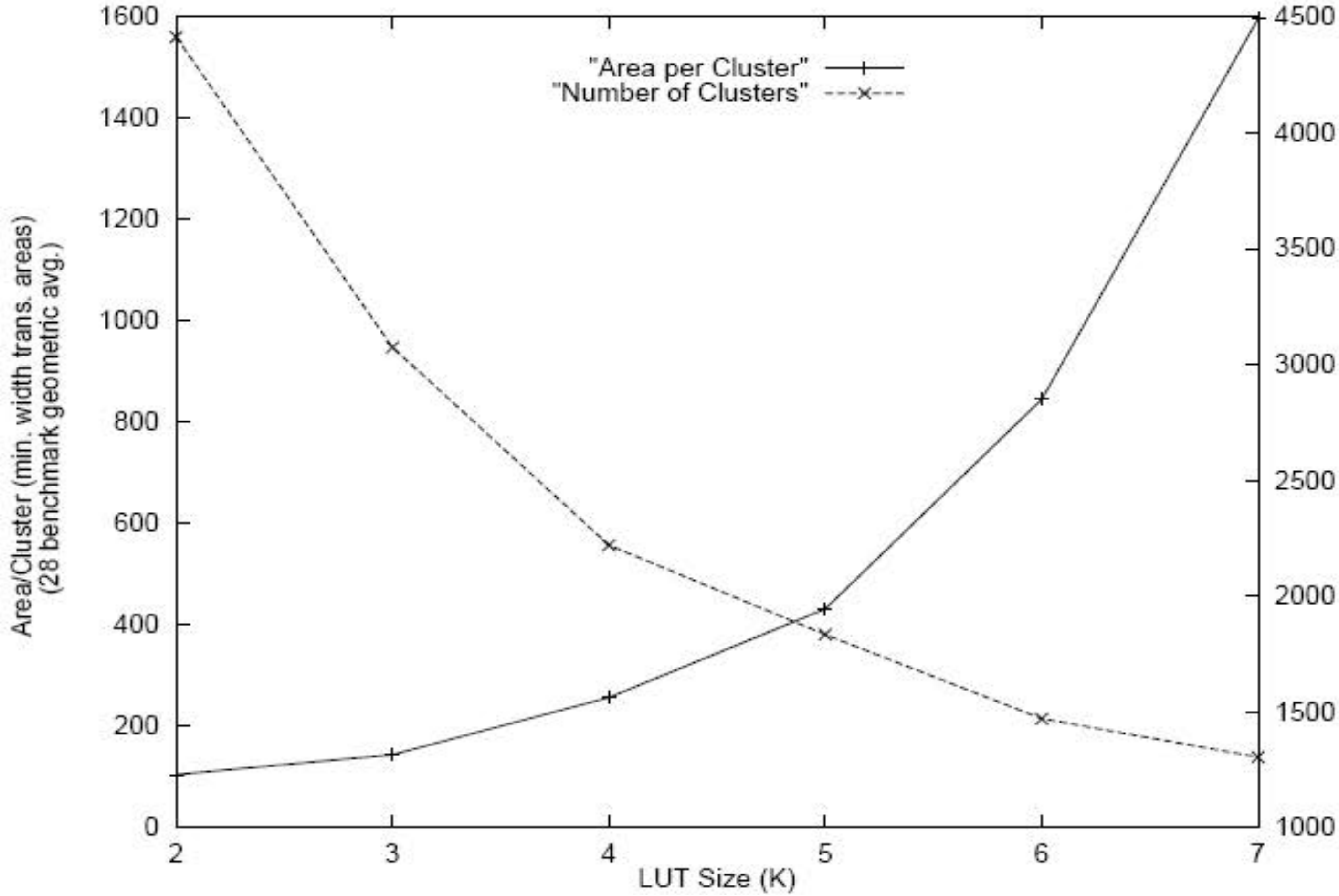


Fig. 6. Number of Inputs Required for 98% Logic Block Utilization

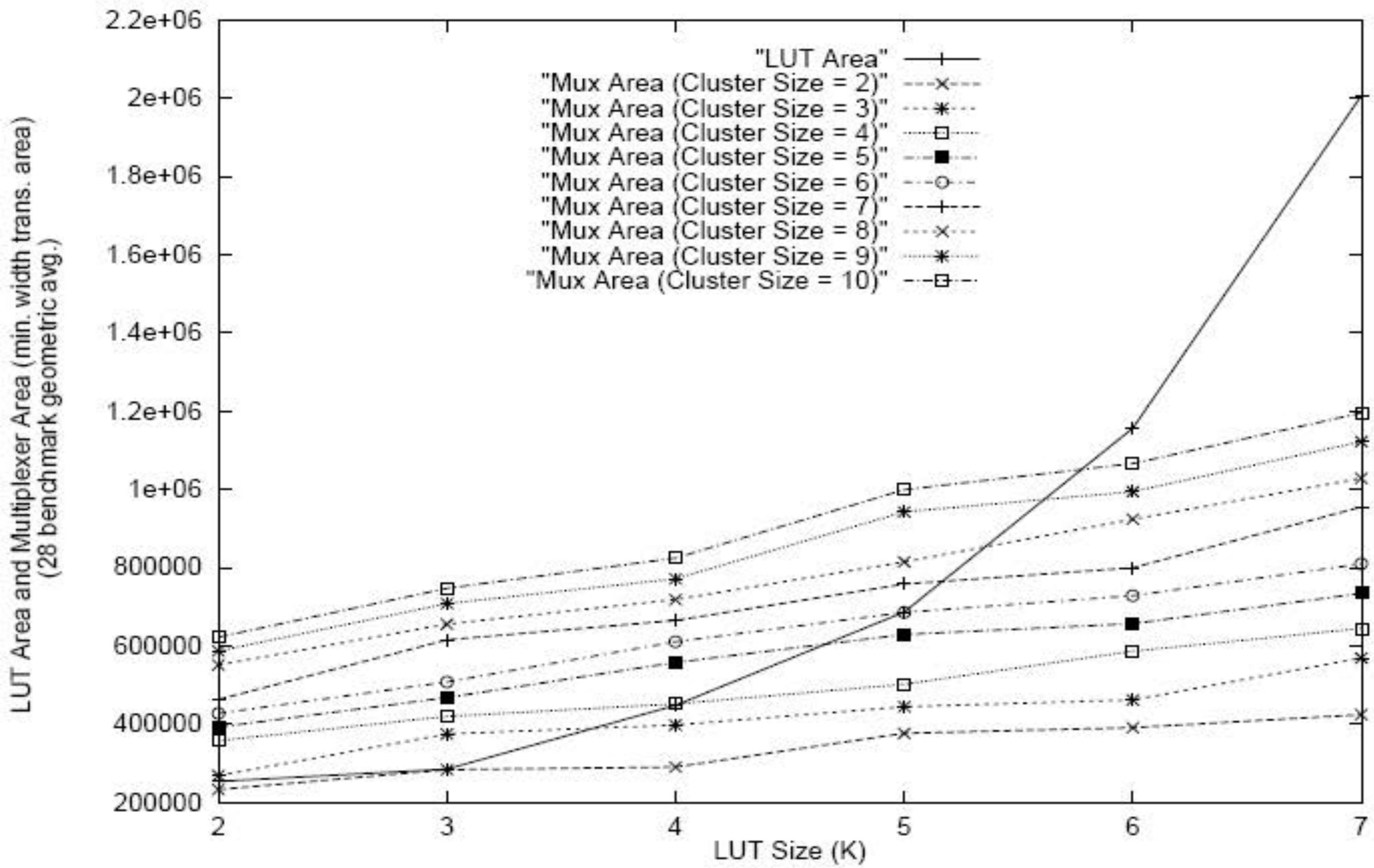
Empirical relationship: $I = \frac{K}{2}(N+1)$

Area vs N and K [1]



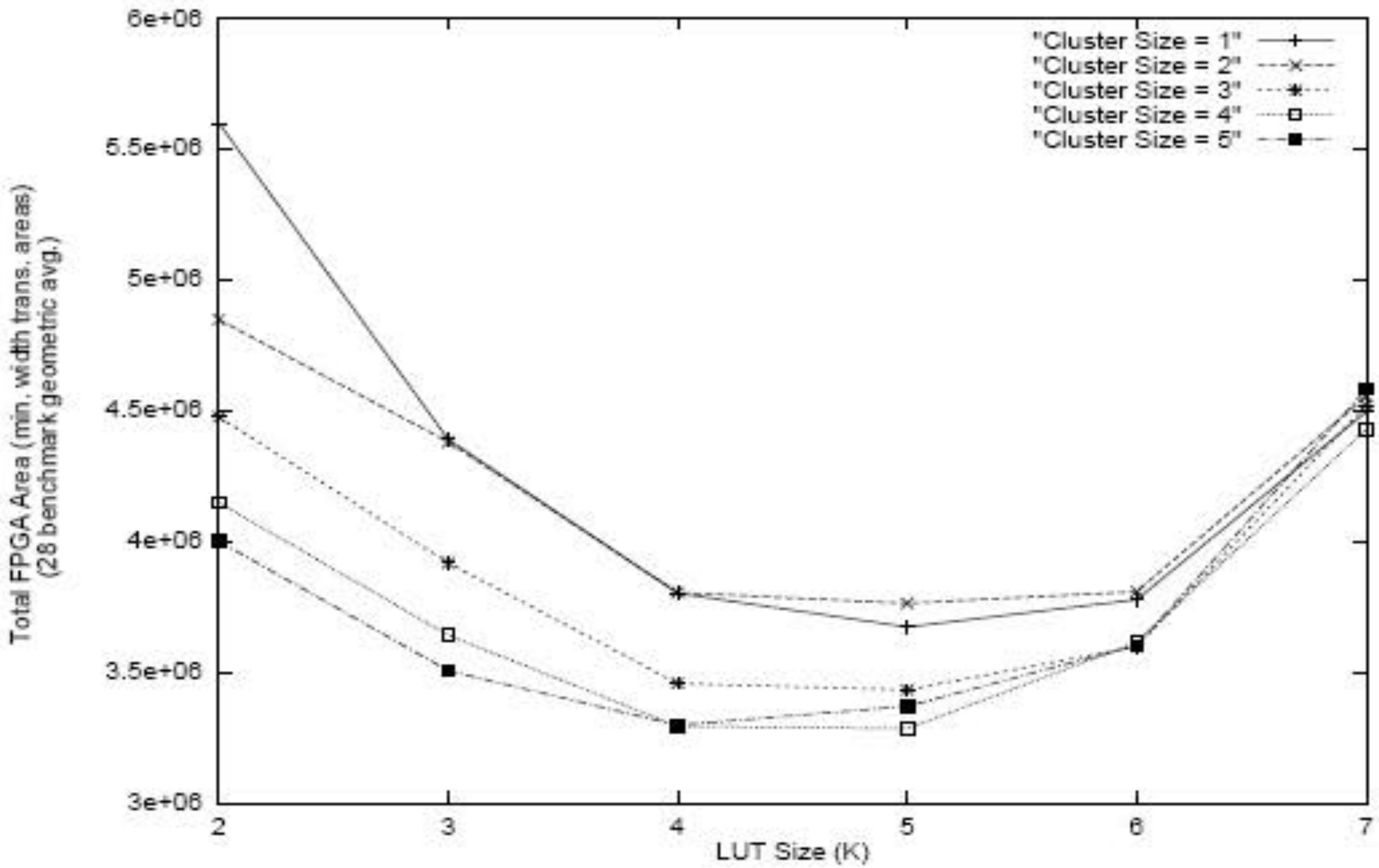
Number of clusters reduced as K increases but area increased

Breakdown of area usage [1]

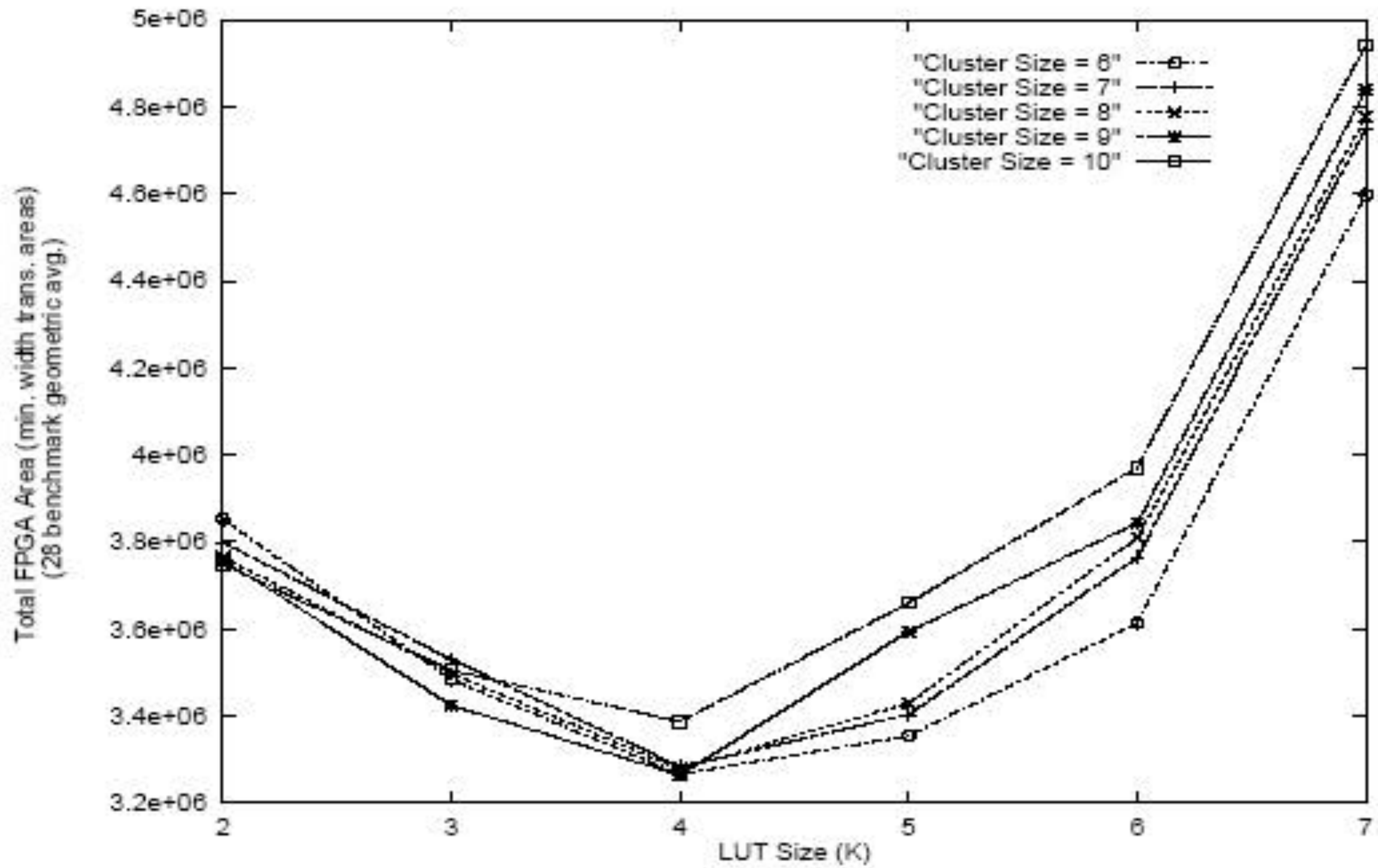


- Intra-cluster mux area significant for large cluster size

Area vs N and K [1]



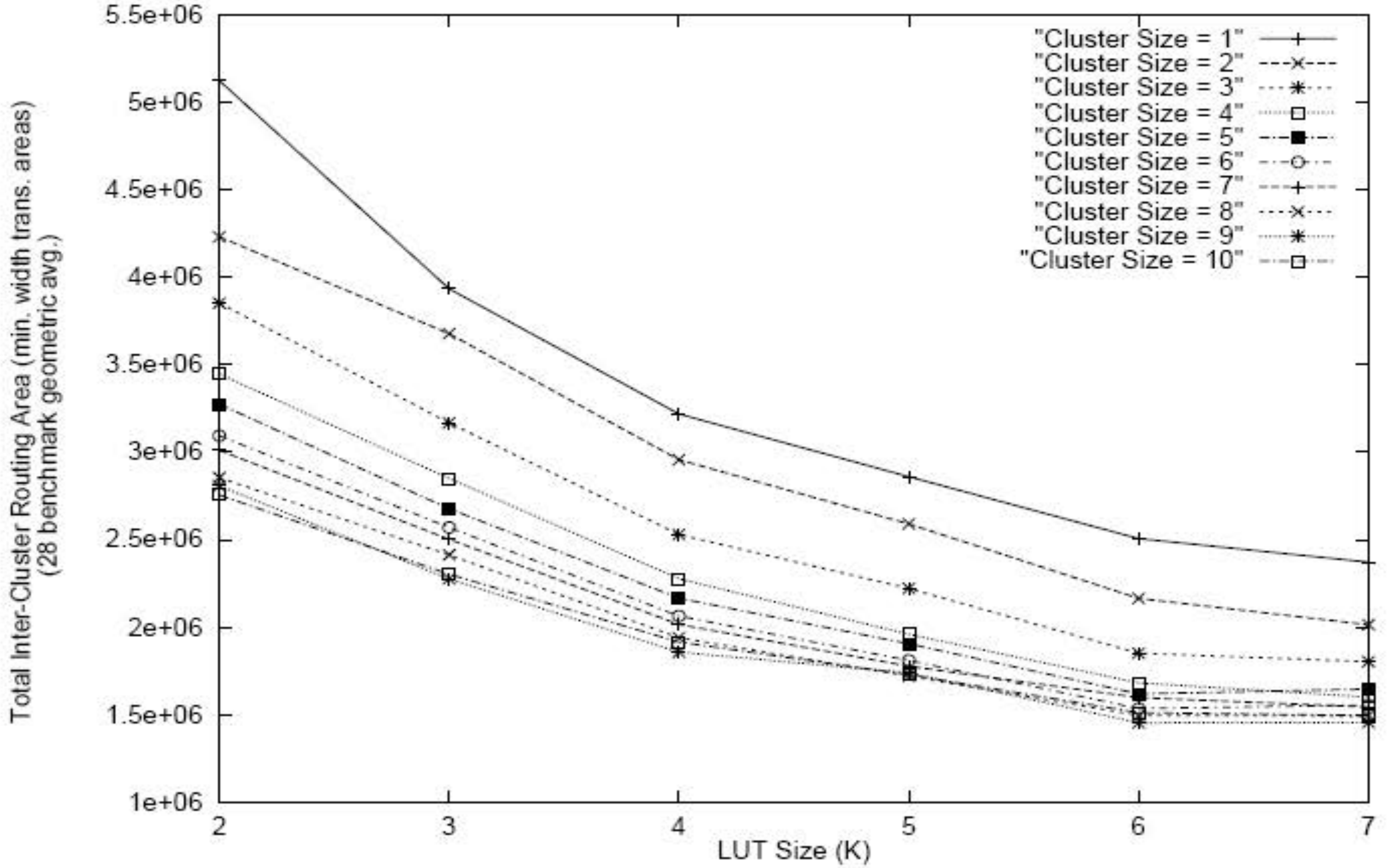
Area vs N and K [1]



Area vs N and K

- Area measure is geometric average of total area of all benchmarks
- LUTs of size 4-5 are most area efficient
- Reduction in area as cluster size increased from 1 to 3 for all LUT sizes. For N>4, little impact on total area

Inter-cluster routing area [1]

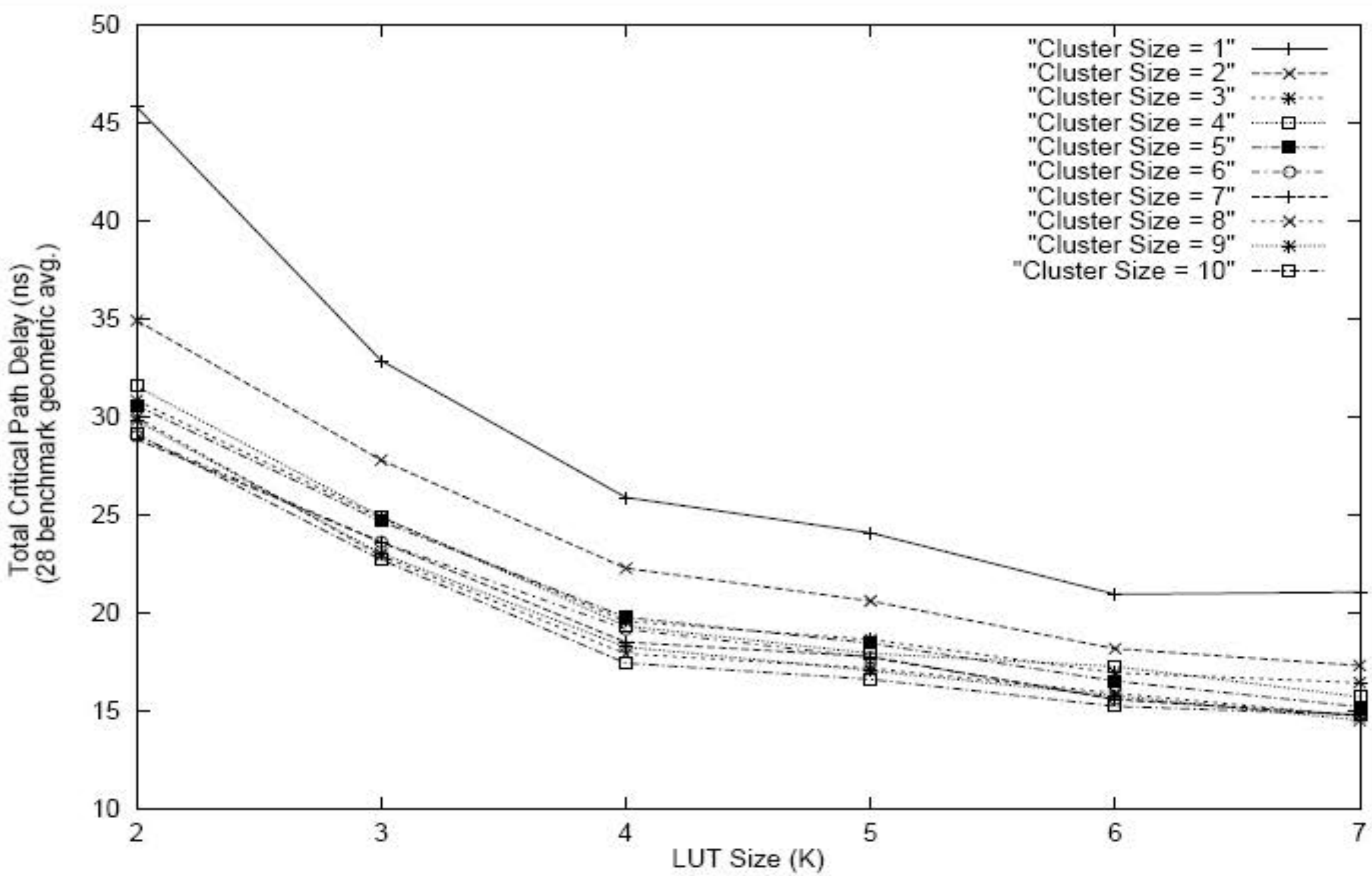


› Decreases linearly with K

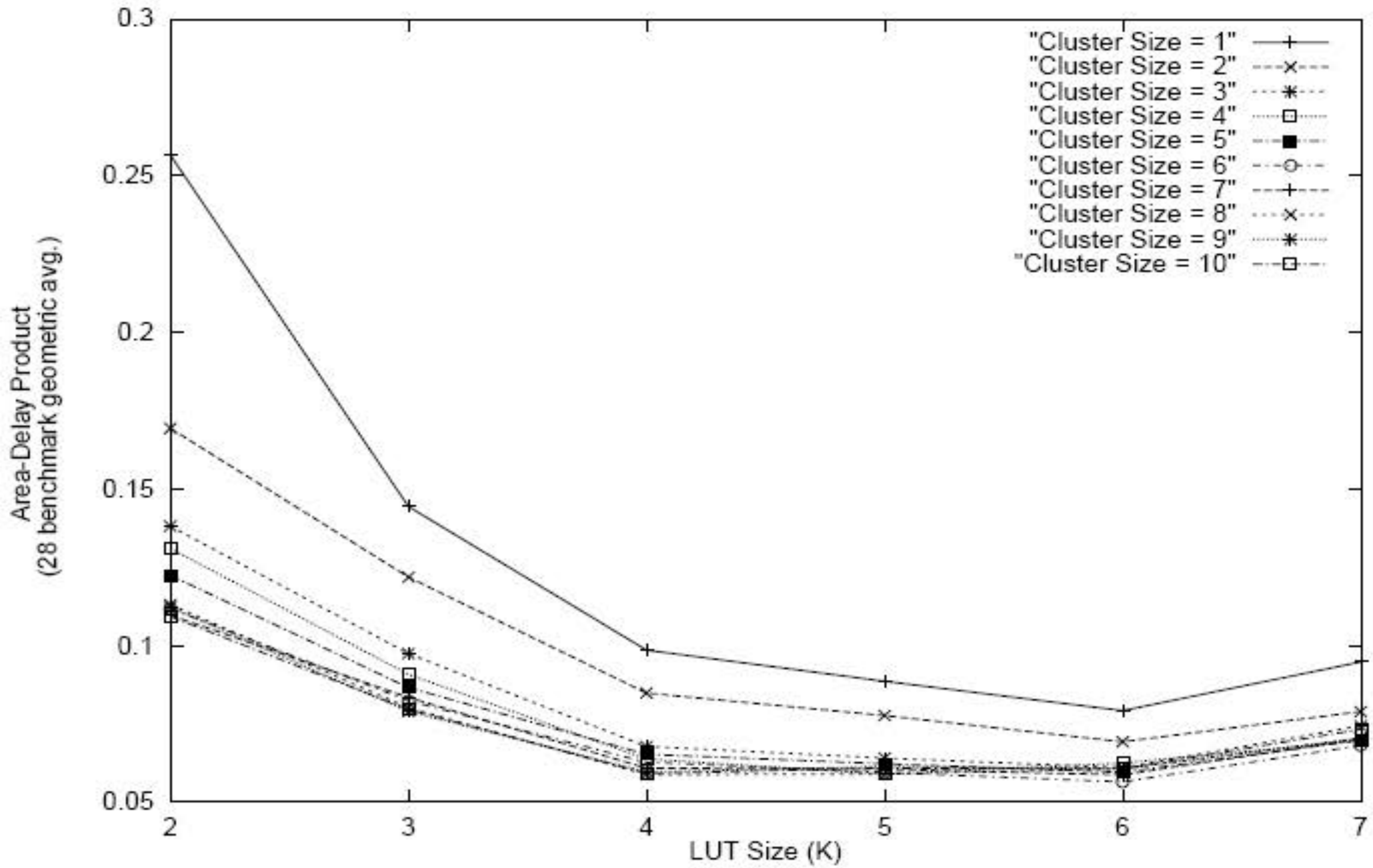
Delay vs Cluster Size

- As LUT and cluster size increases
  - delay through a cluster increases
  - number of LUTs and clusters in series on critical path decreases

Delay vs Cluster Size [1]



Area-delay product [1]



LUTs size 4-6 and cluster size 3 to 10 best

Summary

- Introduced island style FPGA architecture
- [1] describes a methodology for evaluating the impact of different architectural parameters on area, delay, area*delay
- Remember that the results are also a function of IC technology, CAD tools and benchmarks
- The VPR tool allowed **exploration** of different architectural choices without actually building the FPGAs
  - Common theme in this course

THE UNIVERSITY OF SYDNEY

References

[1] Elias Ahmed, Jonathan Rose: The effect of LUT and cluster size on deep-submicron FPGA performance and density. IEEE Trans. VLSI Syst. 12(3): 288-298 (2004)

Review Exercises

- What is K and N for the Virtex-7 and Stratix-V architectures?
- How many LUTs does it take to implement a full adder?
- How is I related to K and N?
- How does K&N affect LUT area?
- Would changing the benchmarks change the results of this study?