Architecture of FPGAs

Michael Kleinschuster

Agenda

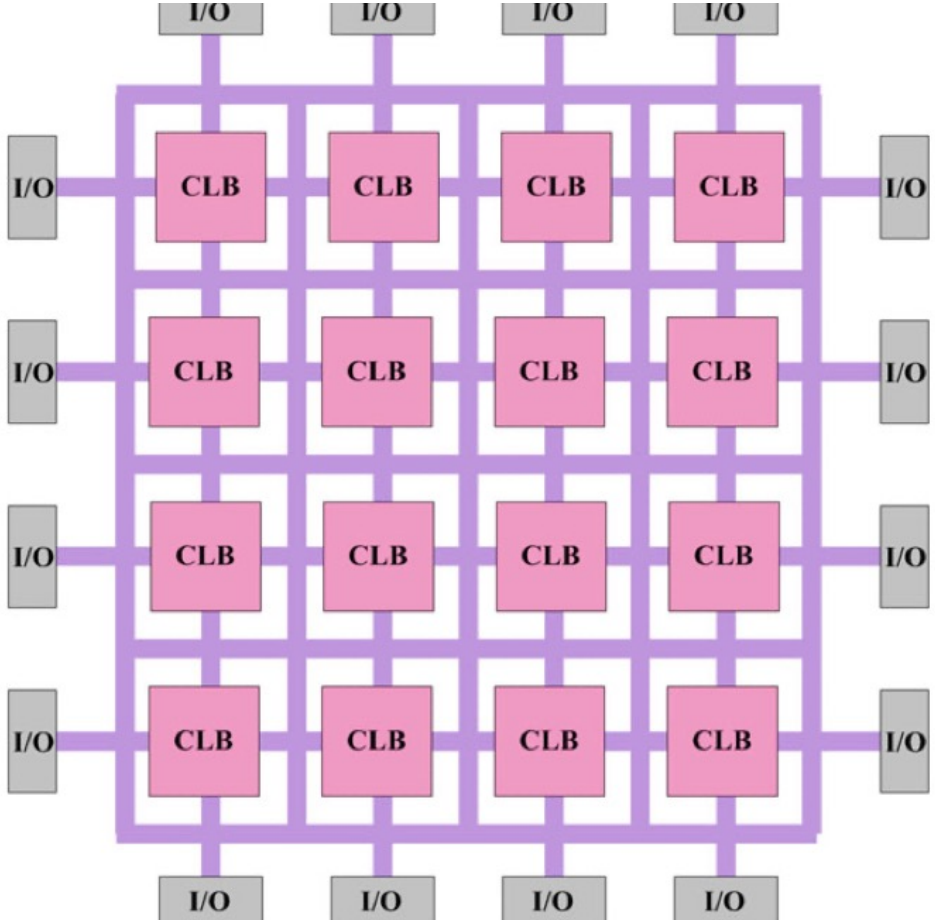
- What is a FPGA in general
- 3 fundamental elements of a FPGA
- Configurable Logic Block (CLB)
- Look-up table
- Interconnect Architecture
- Input / Output Blocks
- Modern FPGA architecture
  - CLB Architecture
  - Block RAMs
  - DSP
  - SERDES

What is a FPGA?

- Field programmable Gate Arrays
- Electrically programmed to implement digital circuits
- Simply convert arbitrary equation into a from of Boolean equation
- Implement Boolean equation as combinational and sequential logic
- Fast time to market
- Unchallenging future updates

3 fundamental building blocks

- Configurable Logic Block (CLB)
- Interconnect Architecture
- Input / Output Blocks

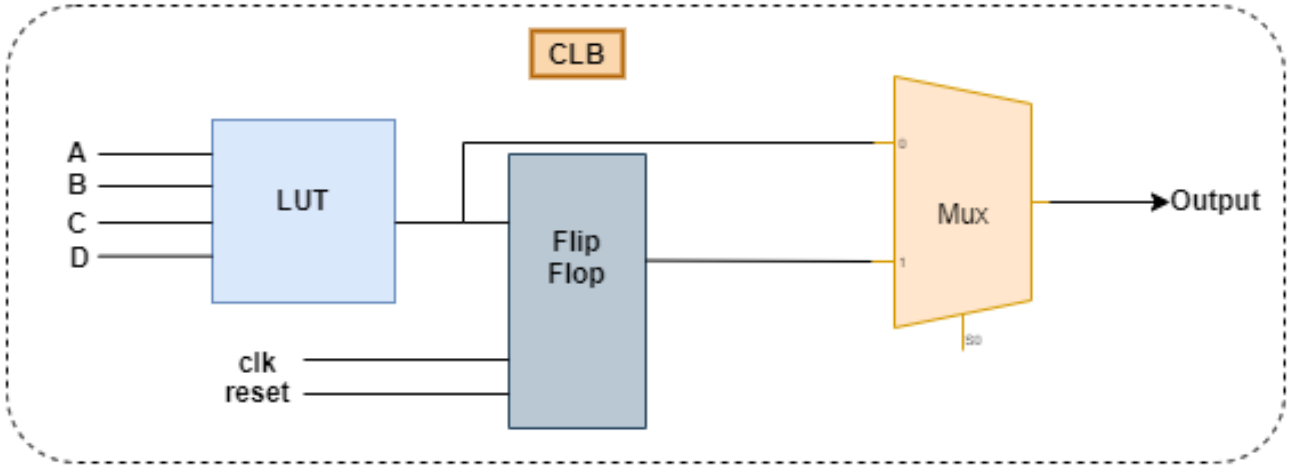


General overview of a FPGA [9, p. 9]

Configurable Logic Block (CLB)

- Provides the basic logic and storage functionality
- Execute complex logic functions
- Implement memory functions
- Synchornize code on the FPGA

- A CLB consist of three essential elemets:
  - LUTs
  - Multiplexer
  - Flipflop

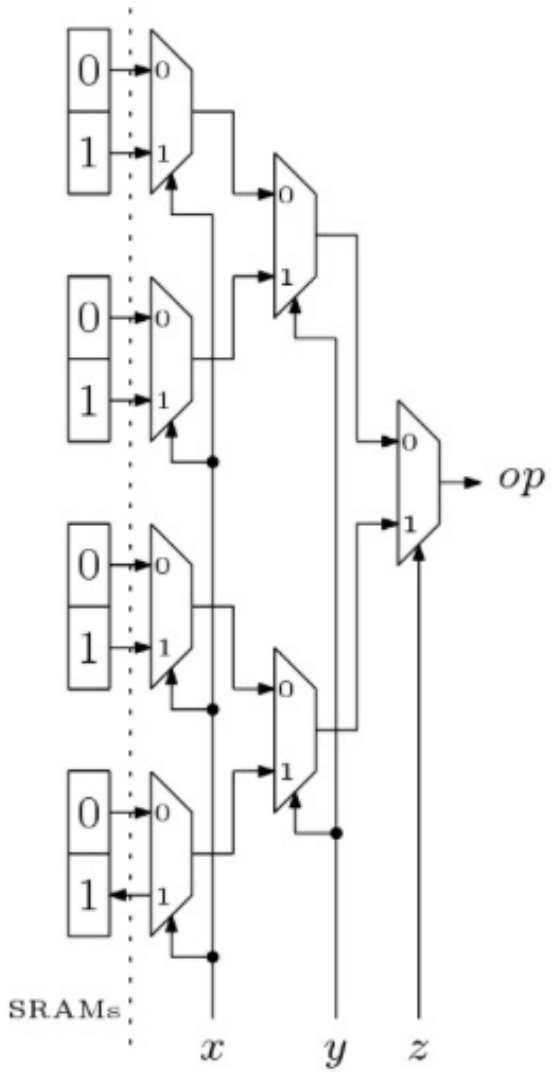


Structure of a CLB [1]

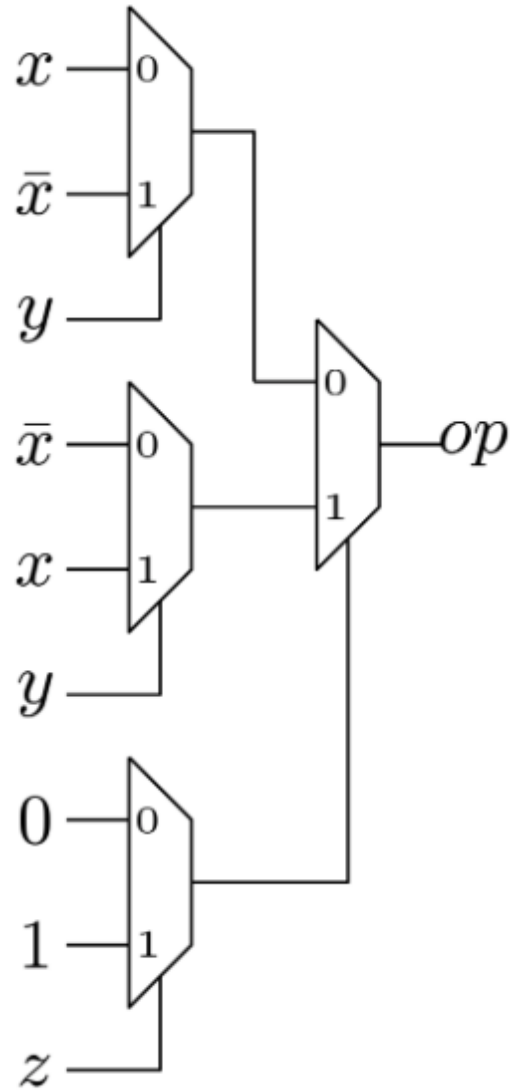
CLB types based on granularity

- Granularity: Defined by the smallest functional CLB

1. Fine Grained: Universal Gate like NAND or AND, OR or NOT
2. Middle Grained: Either Multiplexer based or RAM/ROM based
3. Coarse Grained: Floating point blocks or a processor as basic unit



LUT based CLB
(Programmed by SRAM) [4]



MUX based CLB [4]

CLB configuration

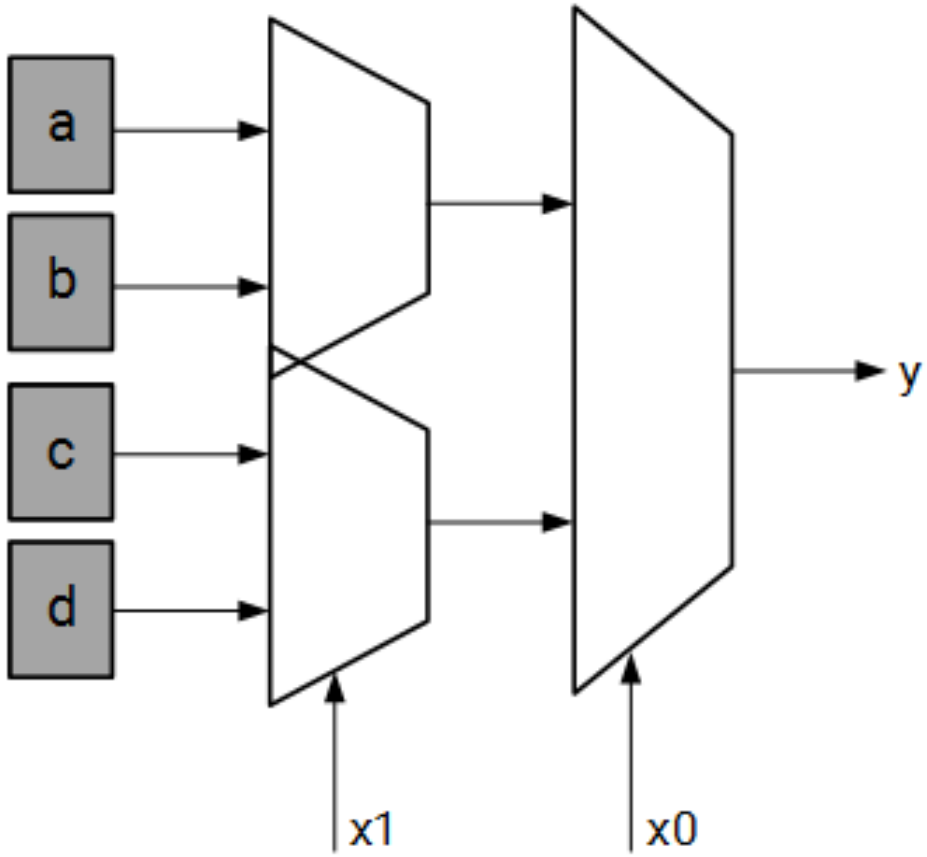
- Bitstream information is generated for the netlist
- Bitstream is programmed on the FPGA via the bitstream loader
- Bitstream contains information which SRAM bit on the FPGA programmed or not
- Routing information used to program SRAM bits of CBs and SWs

Look-up Table (LUT)

- Capable of implementing any logic function of N Boolean variables
- Predefined list of output for every combinations of inputs
- Fast way of retrieving a output
  - Possible results are stored and then referenced instead of doing a calculation

Functional Representation of a LUT

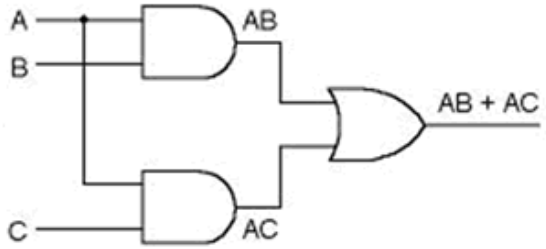
- For a N-input LUT the number of memory locations equals 2^N
  - Allows table to implement 2^{N^N} functions
- Collection of memory cells connected to set of MUXs
- Input bits select multiplexer for a desired output
- LUT used as function compute engine and a data storage element

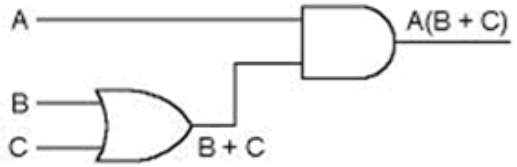


Basic LUT example [2]

Executing Boolean algebra in a LUT

Boolean equation: AB + AC



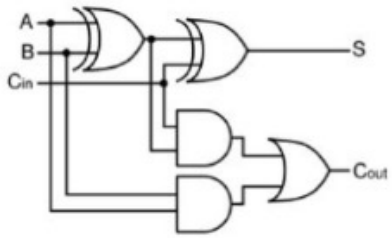


A	B	C	AB	AC	AB + AC
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	0
0	1	1	0	0	0
1	0	0	0	0	0
1	0	1	0	1	1
1	1	0	1	0	1
1	1	1	1	1	1

A	B	C	A	B + C	A(B + C)
0	0	0	0	0	0
0	0	1	0	1	0
0	1	0	0	1	0
0	1	1	0	1	0
1	0	0	1	0	0
1	0	1	1	1	1
1	1	0	1	1	1
1	1	1	1	1	1

Truth table for shown Boolean equation [4]

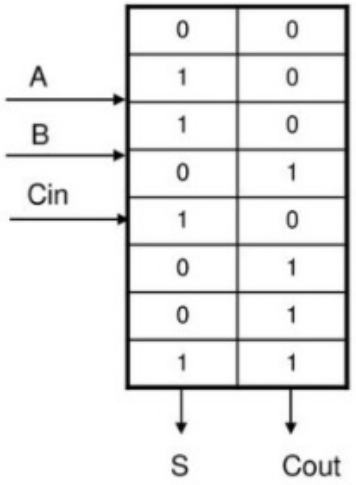
More complicated example



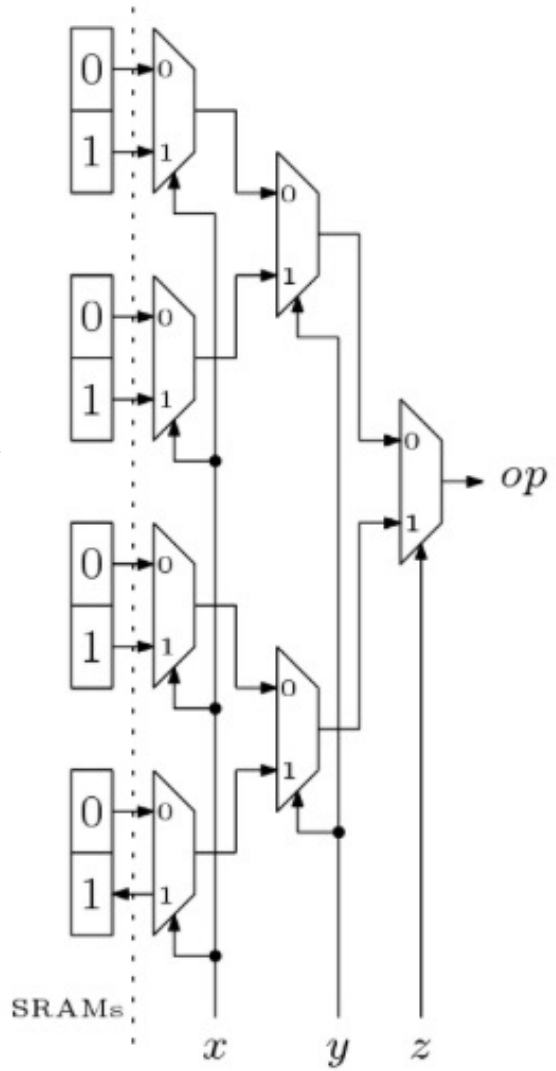
Truth Table

Inputs			Outputs	
A	B	Cin	S	Cout
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

3-input, 2-output LUT



Truth table and LUT for shown 3 Input and 2 Output Boolean Equation [4]

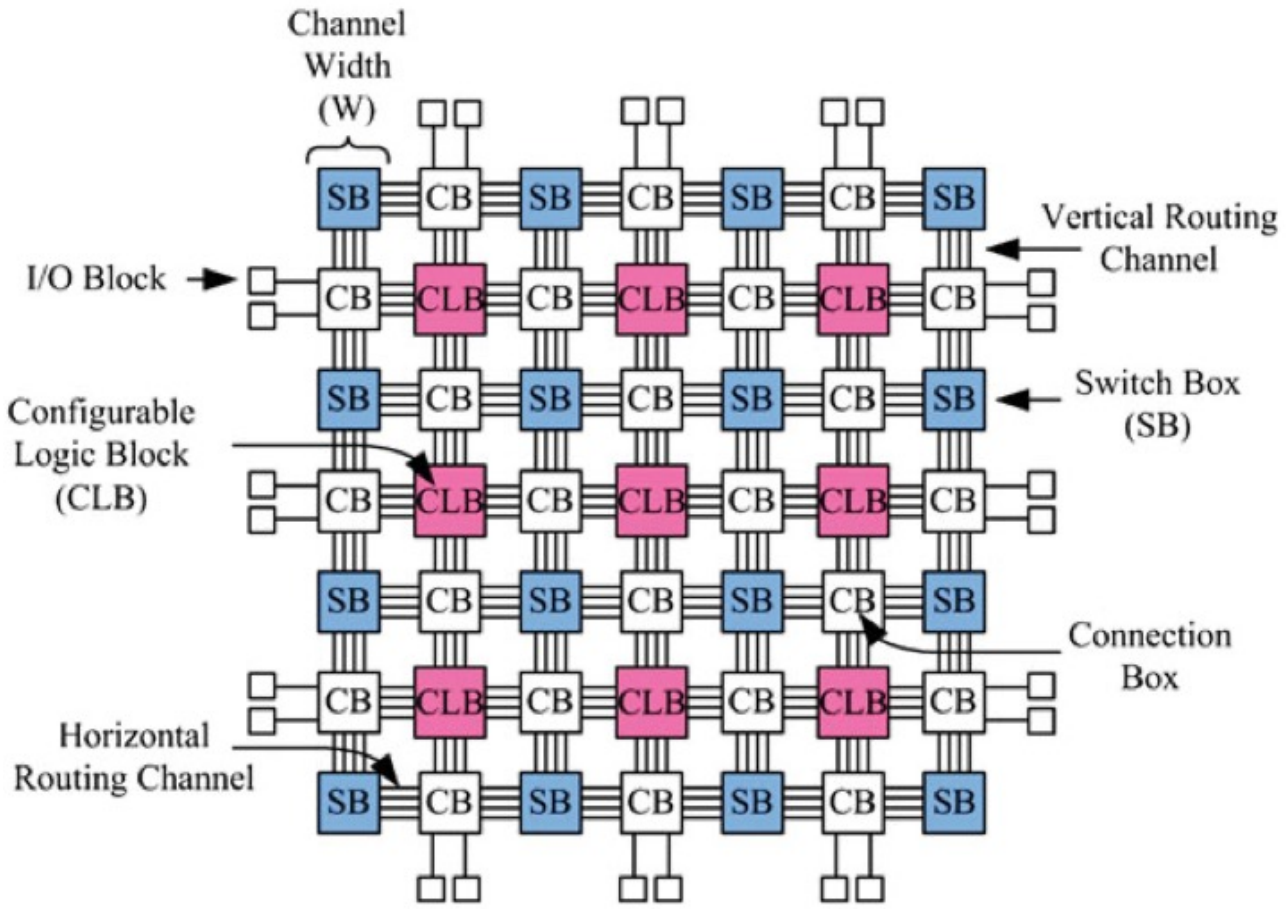


LUT configuration example [3]

Interconnect Architecture

- Interconnect of signal pathways between input and outputs of functional elements within the FPGA
- Interconnect also called Routing
- Routing interconnect consists of wires and programmable switches
- Routing interconnect must be very flexible to deal with wide variety of circuits
- Routing network consumes 80-90% of total area
- Distinguish between Island-Style and hierarchical routing architecture

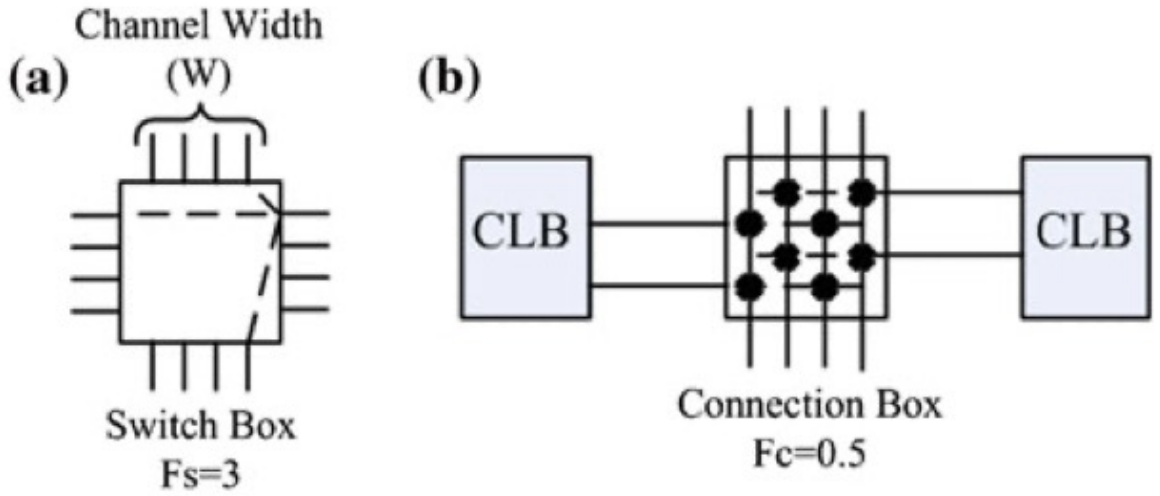
Island-Style Routing Architecture



Overview of a Island-Style Routing Architecture [9, p. 14]

- CLBs blocks look like islands in a sea of routing interconnect
- CLBs are arranged on a 2D grid and are interconnected by a programmable routing network
- Horizontal and vertical tracks are connected via switch boxes
- Logic blocks are connected to routing network via connection boxes

Example of switch and connection box

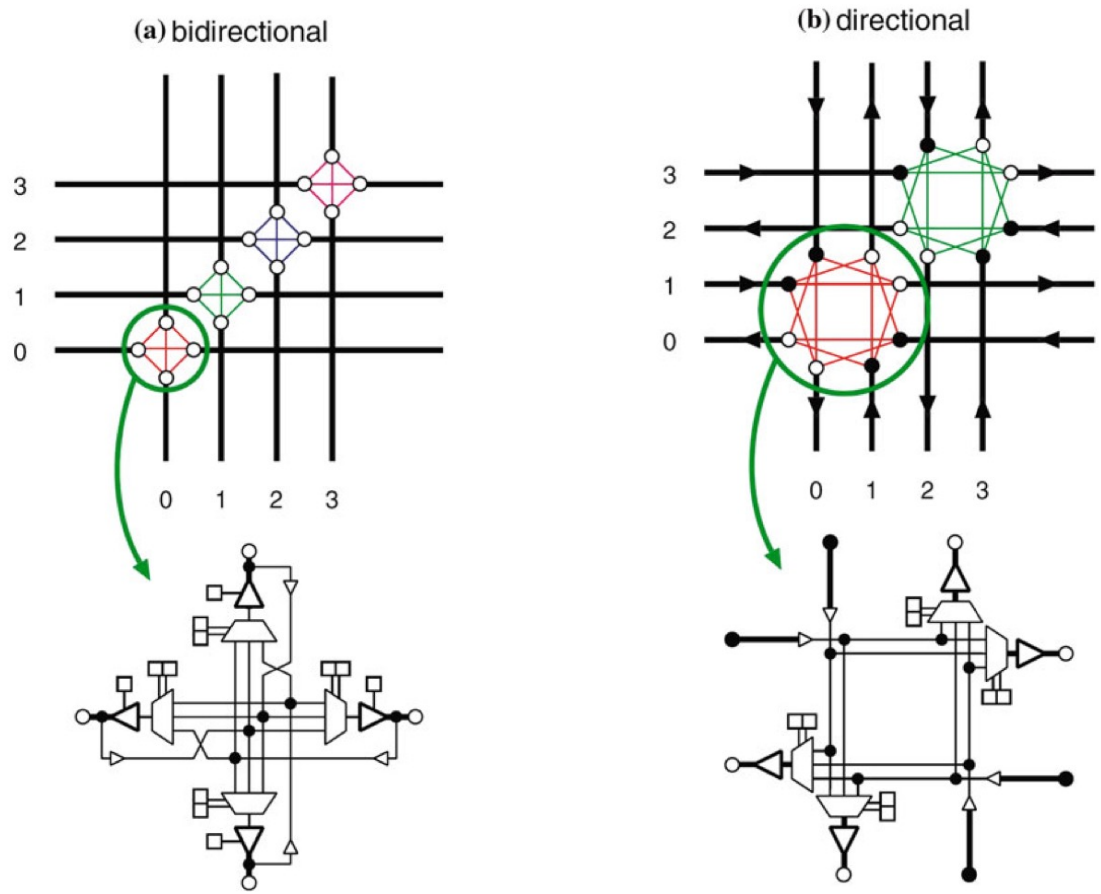


Switch Box and Connection Box Overview [9, p. 15]

- Flexibility of connection box (Fc):
  - Number of routing tracks of adjacent channel connected to the pin of a block
- Flexibility of switch box (Fs):
  - Total number of tracks with which every track entering in the SB

Bidirectional vs directional switch box

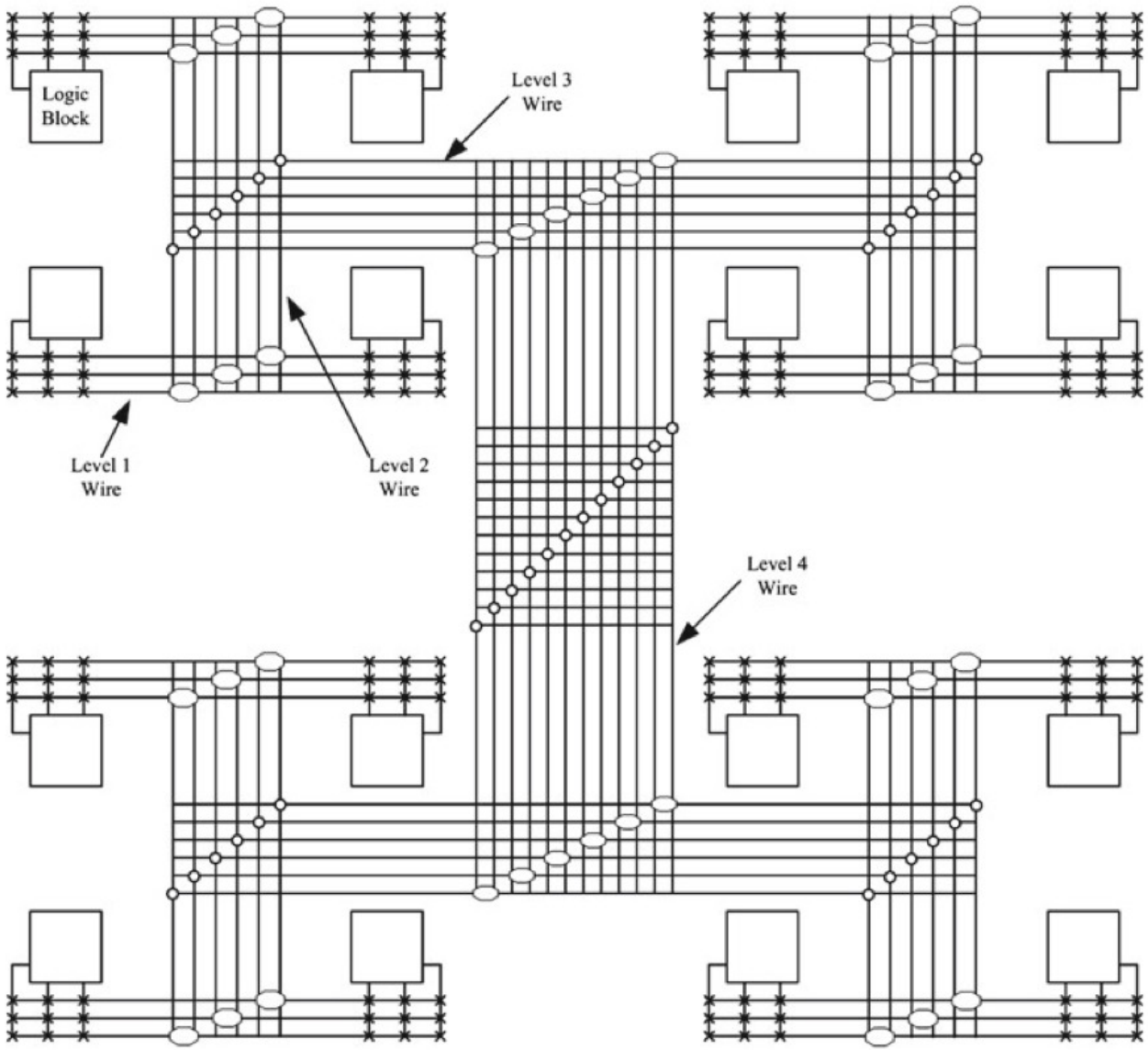
- Bidirectional wiring:
  - CLB output pins can connect to any track
- Directional wiring:
  - 25% improvement in area, 9% in delay



Bidirectional vs. directional switch box [5, p. 43]

Hierarchical Routing Architecture

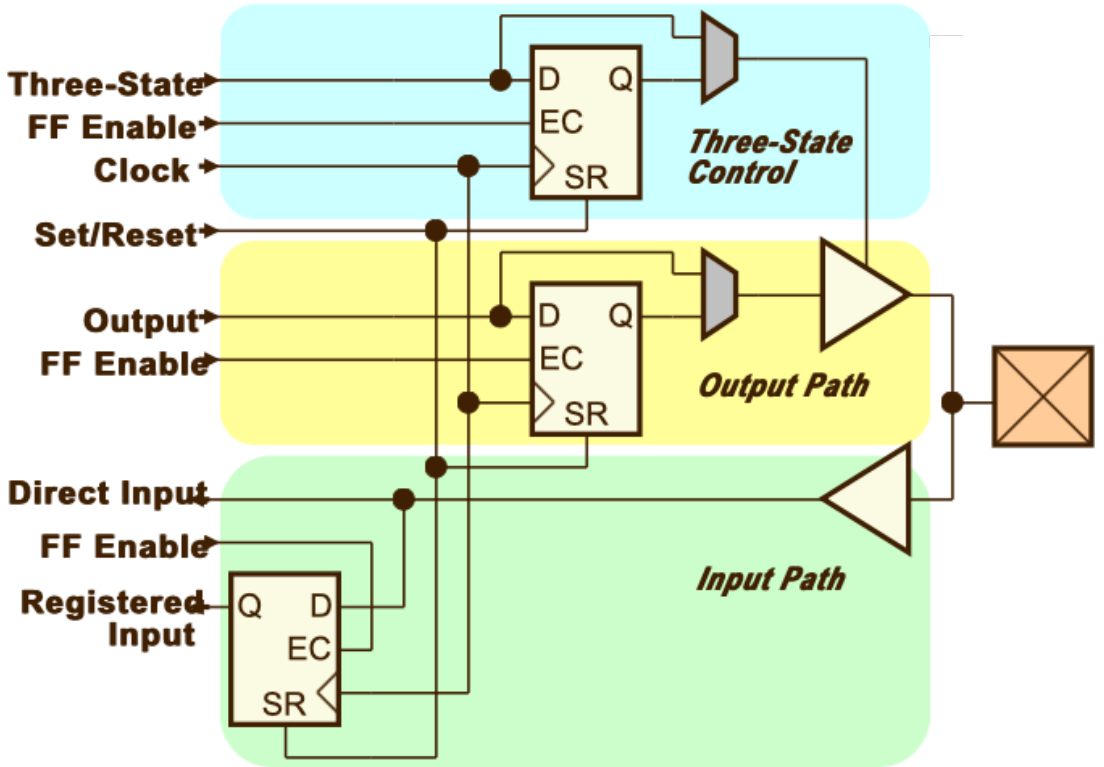
- Implying a hierarchy in placement and routing of connections between blocks
- Exploit locality by dividing FPGA logic blocks into separate groups/clusters
- Hierarchical architecture = tree-based architecture
  - Connections between same cluster are made by wire segments at the lowest level of hierarchy
- Blocks of different groups need traversal of one or more hierarchy levels



Hierarchical Routing Architecture [9, p. 22]

Input/Output Blocks

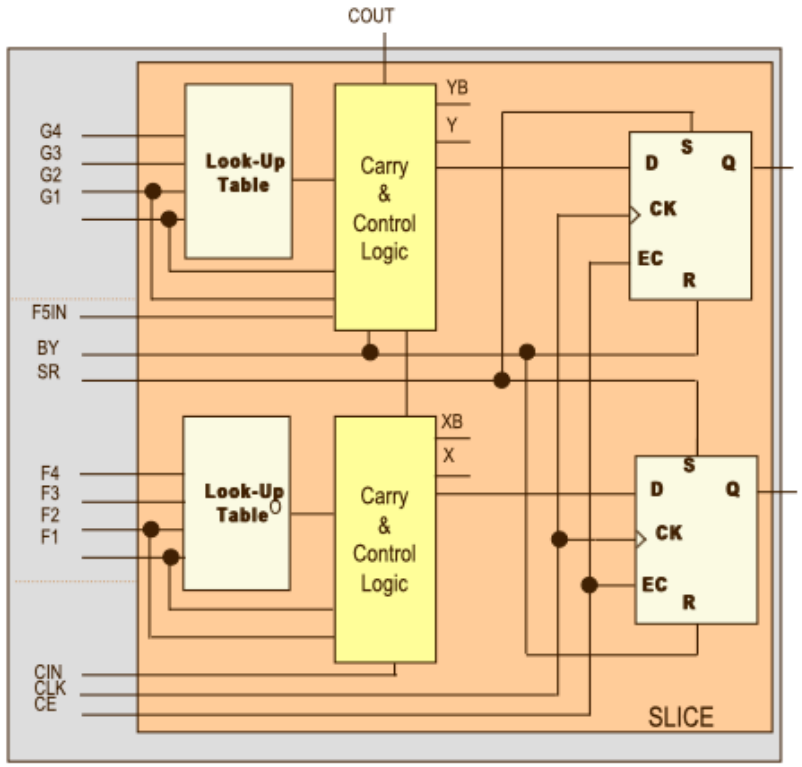
- Used to connect outside world to the FPGA
- Can be used for input and output signals (uni- or bi-directional I/O)
- Output can forced to Three-State (High impedance)
- Inputs and Outputs can be stored in D-Flip-Flop
  - For high-performance I/O
- Inputs can be delayed



General overview of a IO [6]

Modern FPGA Architecture

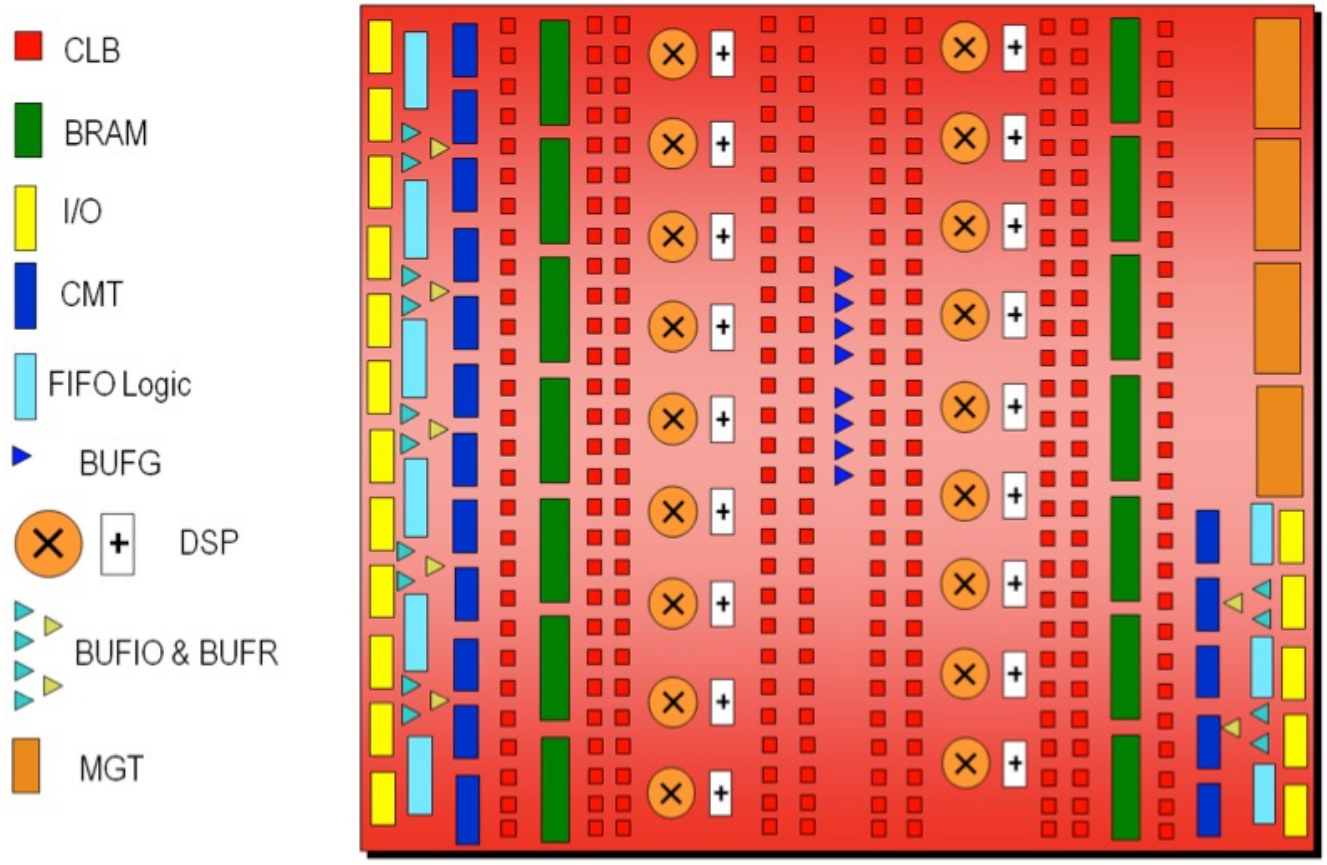
- Modern FPGAs emphasizes following points:
  - More LUTs in smaller area
  - Minimum power utilization
  - More flexible on-chip memory
- Putting Adders/Multipliers and DSP logic inside CLB
  - Reduce latency
  - Faster computation
  - Increasing throughput
- Implementing Fast Carry Logic
  - Fast generation of sum and carry signals
  - Increases performance and efficiency of adders, subtractors..



Slice with included Carry and Control Logic [6]

Architecture Overview of the Artix-7 FPGA

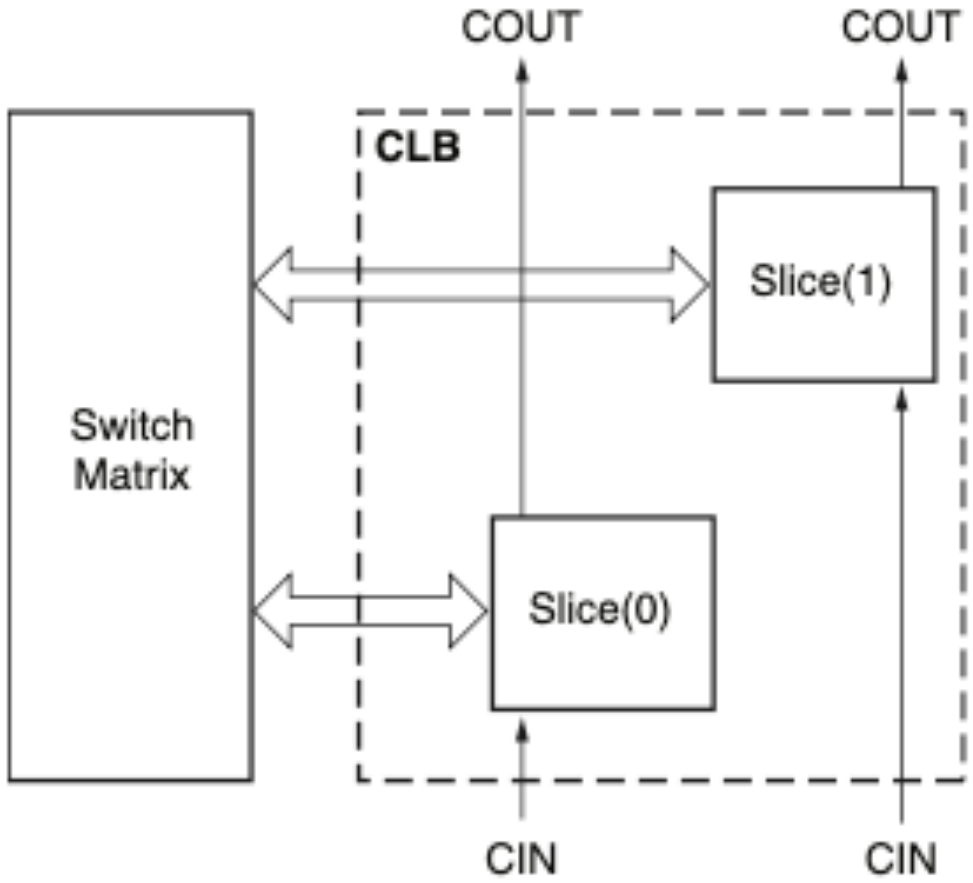
- Used on the ZYBO Z7 Development
- BRAMs
- DSPs
- IOBs



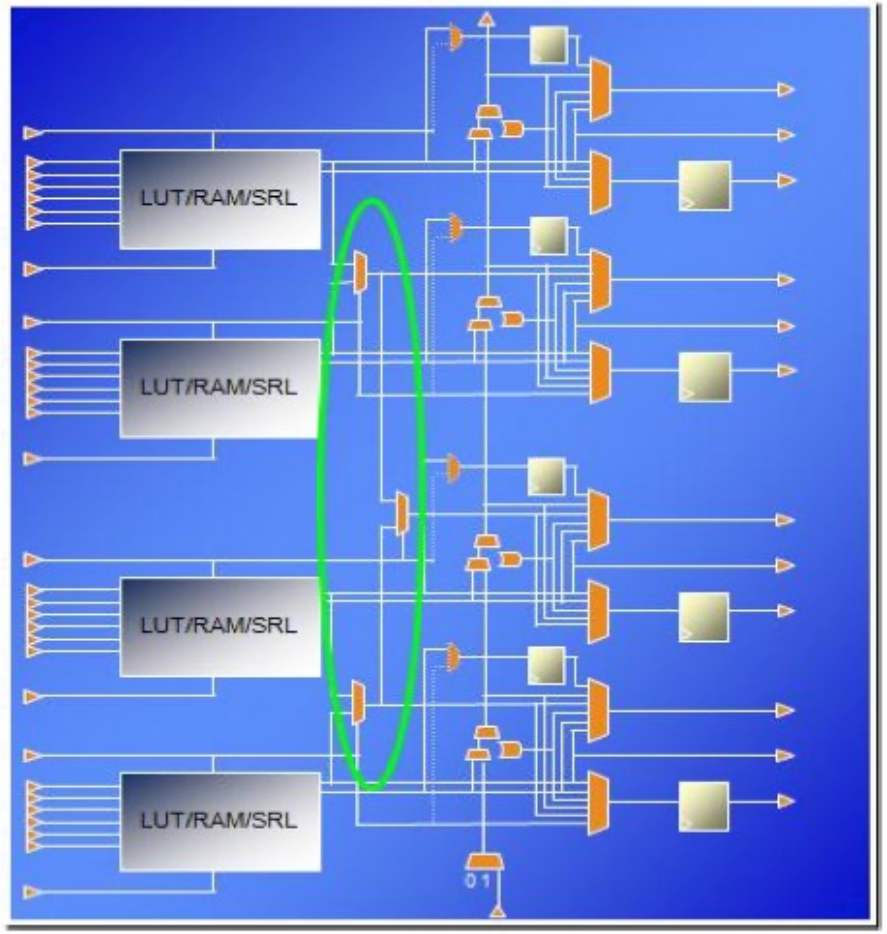
Overview of a Artix-7 FPGA [8, p. 9]

CLB Architecture

- Arranged in columns
- Connection to Switch Matrix
- Slices are not directly connected
- Contains pair of slice
- Either two SLICEL or one SLICEL
  and
  one SLICEM



- CLB with Slices [7, p. 9]



- Detailed view on a slice [8, p. 15]

Types of slices

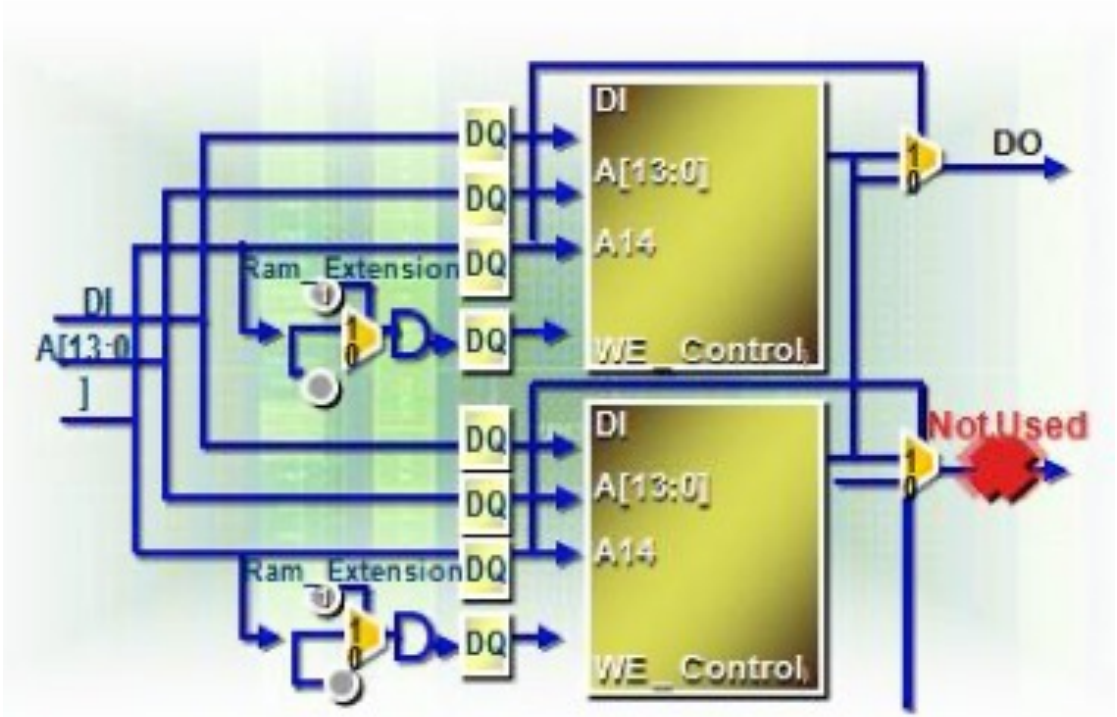
Distinguished between two types:

- SLICEM: Full slice
  - LUT is usable for logic and memory/Shift register lookup table
  - Contains wide range of multiplexers and carry chains

- SLICEL: Logic and arithmetic only
  - LUT only usalbe for logic, not for storage usage
  - Contains wide range of multiplexers and carry chains

BRAMs

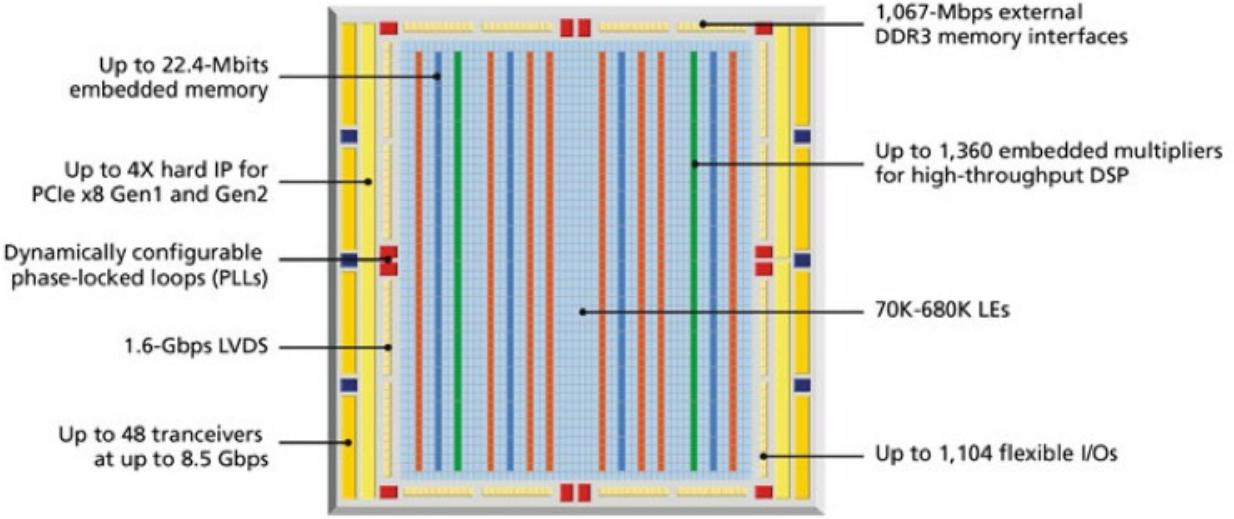
- Dedicated 36 kb block RAMs
  - Used to implement much larger memory arrays
- Cascade Group Size:
  - Casscading two vertically adjacent 32Kx1 to build on 64Kx1 RAM block
  - Saving resources, improving speed for large memories



Implementation of BRAM [8, p.42]

Digital signal processing block (DSP)

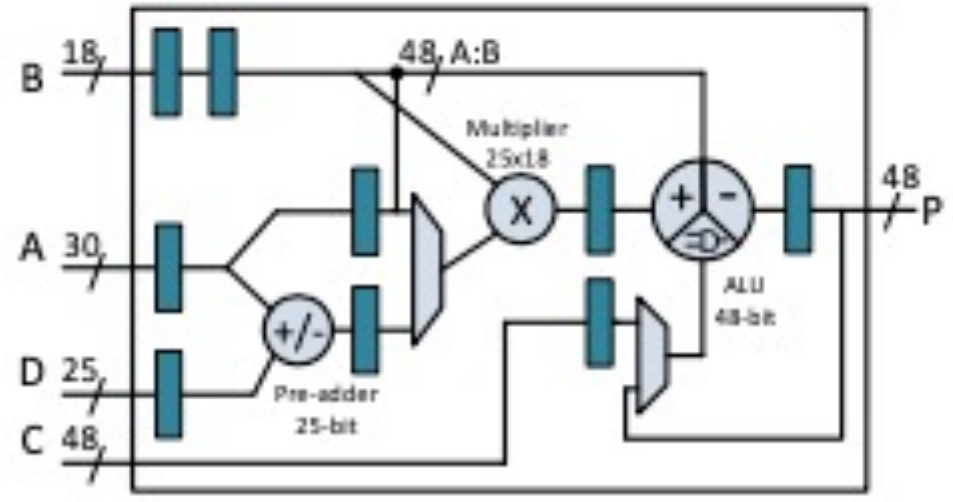
- Full-precision multipliers of different granularity types
- Placed in parallel to memory
- In Startix IV optimized for
  - signal processing applications
  - Finite Impulse Response (FIR)
  - Infinite Impulse Response (IIR),
  - Fast Fourier Transform functions (FFT) etc.
- DPSs can implement various functions:
  - Multiplication,
  - Multiply-add,
  - Logical shift



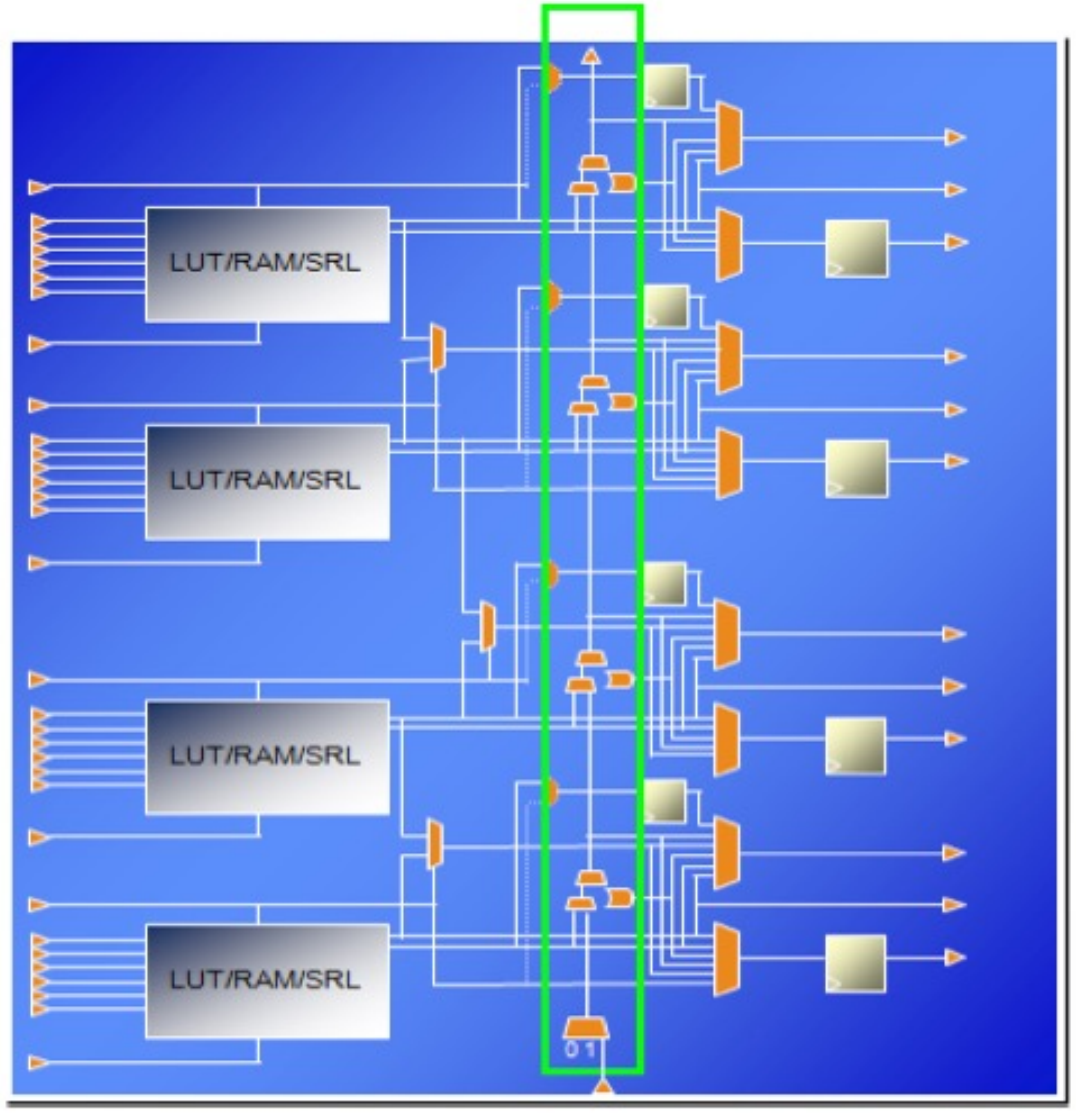
Overview of Startix IV [9, p. 42]

Representation of the DSP48E1

- 25-bit add/subtract unit on D
- Lower 25 bit from A as an input for the pre-adder
- Multiplier: 25 x 18 bit
- ALU block: usable for add/subtract/logic unit
- Sub blocks can be combined in various ways



Xilinx DSP48E1 structure [10, p. 2]



Detailed view of a slice [8, p.16]

Carry chains

- Implies fast addition and subtraction
  - Carry Out propagates vertically through the LUTs
  - Carry chain propagates from one slice to the slice above (same column)
- Can be used for adders and comparators

SEREDS

- SERDES = serializer/deserializer
  - Converts parallel data to serial data and vice-versa
- SERDES supports most commonly-used high-speed serial interface
- Each channel can be configured independently
  - Perform high-speed, full-duplex serial data transfers
  - Data rates from 270Mb/s to 3.2Gb/s


Any Questions?

Bibliography

- [1] The Ultimate Guide to FPGA Architecture. https://hardwarebee.com/ the-ultimate-guide-to-fpga-architecture//. Online; accessed 09 November, 2022.
- [2] Understanding FPGA architecture/ LUT. https://www.xilinx.com/htmldocs/xilinx2017_4/sdaccel_doc/yeo1504034293627.html//. Online; accessed 09 November, 2022.
- [3] FPGA: Basic Overview. https://digitalsystemdesign.in/fpga-basic-overview/ /. Online; accessed 09 November, 2022.
- [4] How to execute the Bolean Algebra in a Look-up Table. https://fpgabeginner.com/ how-to-do-executing-the-bolean-algebra-in-a-look-up-tab; Online; accessed 09 November, 2022.
- [5] G. Lemieux, E. Lee, M. Tom, and A. Yu, “Directional and single-driver wires in FPGA interconnect,” Proc. - 2004 IEEE Int. Conf. Field-Programmable Technol. FPT ’04, pp. 41–48, 2004, doi: 10.1109/fpt.2004.1393249.
- [6] FPGA Resources. https://people-ece.vse.gmu.edu/coursewebpages/ECE/ECE545/F17/viewgraphs/ECE545_lecture12_FPGA_Resources.pdf//. Online; accessed 09 November, 2022.

Bibliography

- [7] Xilinx, “7 Series FPGAs Configurable Logic Block UG474,” *Xilinx*, vol. UG474, pp. 1–74, 2016, [Online]. Available: https://www.xilinx.com/support/documentation/user_guides/ug474_7Series_CLB.pdf.

- [8] Z. Vivado, “7-Series Architecture Overview,” 2013, [Online]. Available: https://xilinx.eetrend.com/files-eetrend-xilinx/forum/201509/9204-20390-7_series_architecture_overview.pdf.

- [9] U. Farooq, Z. Marrakchi, and H. Mehrez, Tree-based heterogeneous FPGA architectures: Application specific exploration and optimization, vol. 9781461435. 2012.

- [10] B. Ronak and S. A. Fahmy, “Mapping for maximum performance on FPGA DSP blocks,” IEEE Trans. Comput. Des. Integr. Circuits Syst., vol. 35, no. 4, pp. 573–585, 2016, doi: 10.1109/TCAD.2015.2474363.