

Traineeships in Advanced Computing for High Energy Physics (TAC-HEP)

FPGA module training

Week-2: FPGA Architecture & its sub-components

Lecture-4: February 6th 2025



[Varun Sharma](#)
[University of Wisconsin – Madison, USA](#)



Content



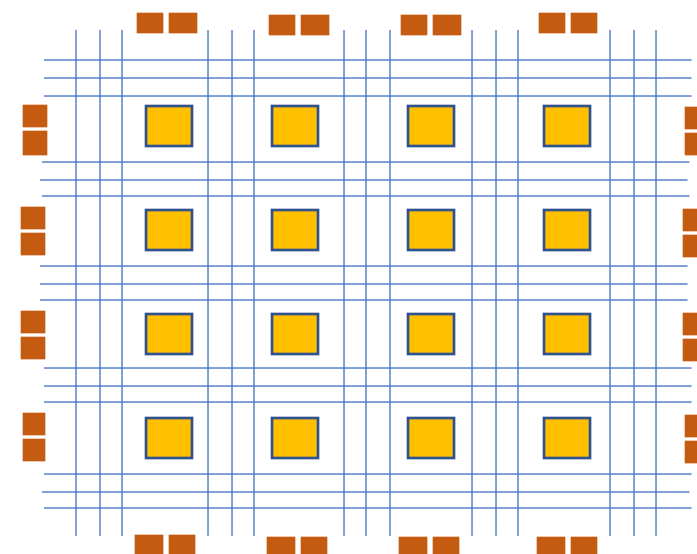
- FPGA and its architecture
- FPGA sub-components
- Basic of logic gates



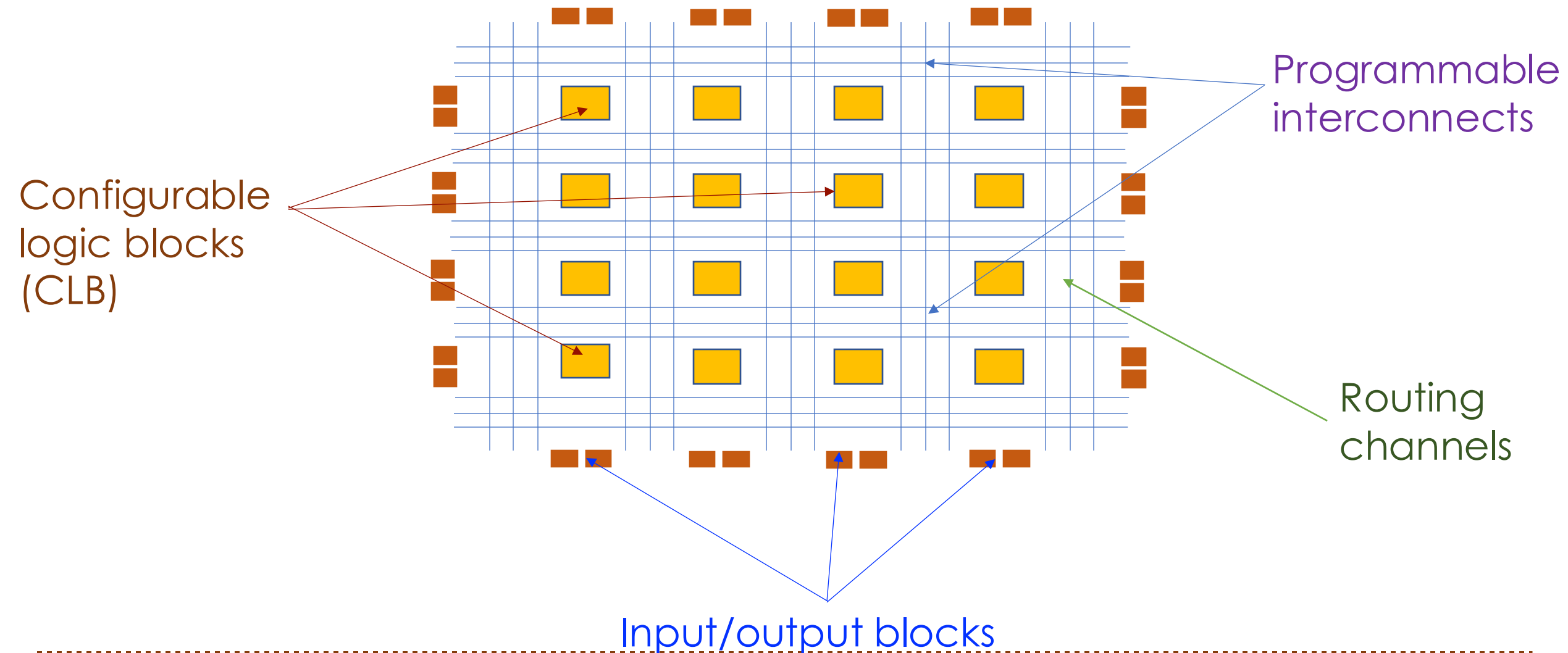
FPGAs:



- **FPGAs** are an array of programmable logic blocks and memory elements that are connected together using programmable interconnect
- **Over decades**, FPGAs have gone from small arrays of PL and interconnect to massive arrays of PL and interconnect with on-chip memories, custom data paths, high speed I/O, & microprocessor cores all co-located on the same chip



FPGA Architecture

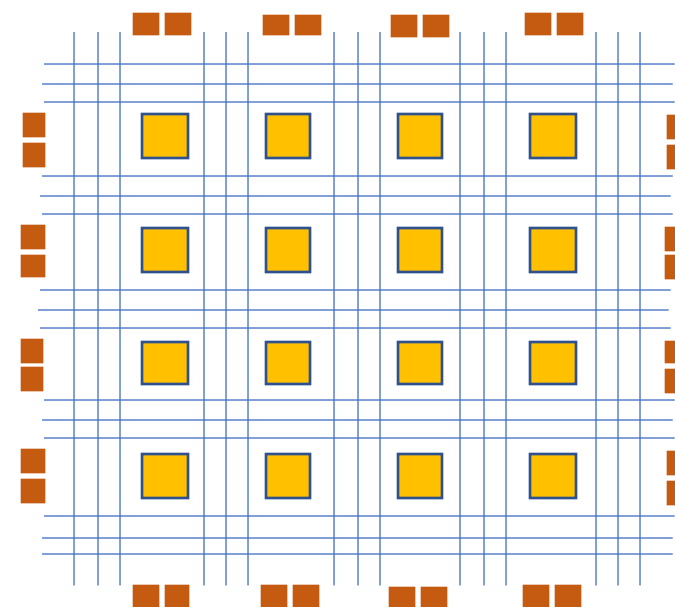


FPGA Architecture



The basic structure of an FPGA is composed of:

- Static Random Access Memory (SRAM):
- Configurable Logic Blocks (CLBs)
 - Look-up table (LUT)
 - Flip-Flop (FF)
 - Multiplexers
 - DSP Blocks
 - Block Memories (BRAM)
- Wires: These elements connect elements to one another
- Input/Output (I/O) pads: These physically available ports get data in and out of the FPGA
- Clocking Resources



SRAMs

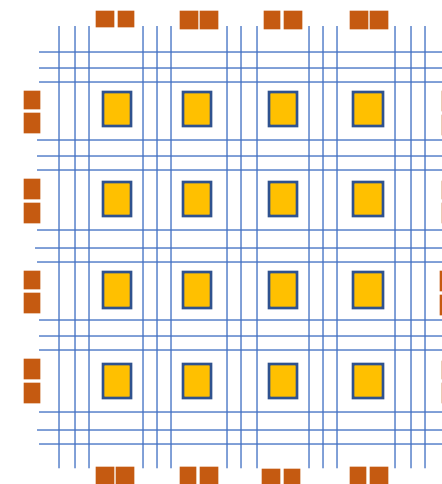


- These are memories that allows FPGAs to be configured and reconfigured so easily
- In SRAM-based FPGAs, configuration is stored in SRAM cells.
 - It determines how the logic blocks are connected, what function they perform, and how routing resources are configured
- They are inherently volatile.
 - Loses its data when power is turned off
 - Need to be configured every time they power on, usually done by loading configuration data from an external memory source

Configurable Logic Blocks



- CLBs are the heart of the FPGA, where logic functions are implemented
- They can be configured to perform a wide range of logical operations, from simple AND/OR gates to complex arithmetic functions
- **Components of CLB:**
 - Look-Up Tables (LUTs)
 - Flip-Flops
 - Multiplexers

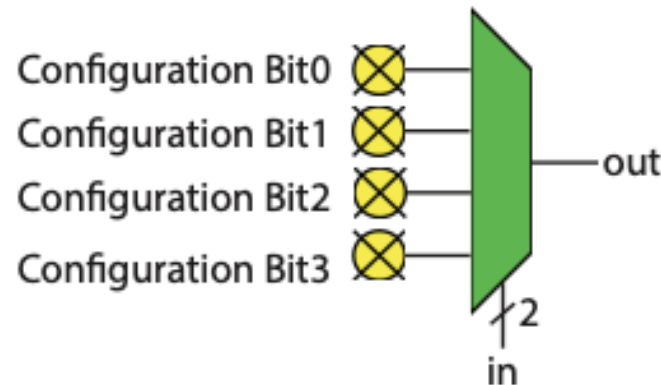


Look-up Tables (LUTs)



- A memory where **address signal are the inputs** and the **outputs** are stored in the **memory entries**
- A typical LUT is an **n-input truth table** stored in SRAM
- Instead of computing logic function in real-time, LUTs **store the output values for all possible input combinations**
- When inputs are applied, the LUT **retrieves the corresponding output from the memory**

a) Lookup Table (LUT)



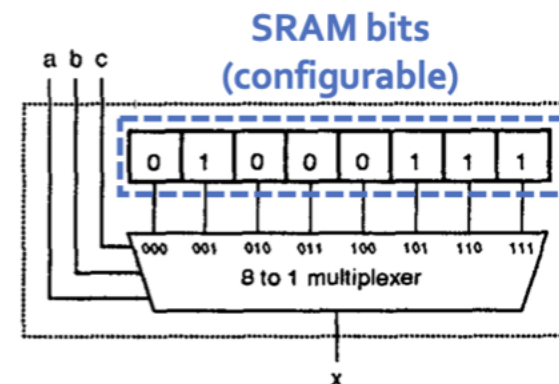
b)

in[1]	in[0]	out
0	0	0
0	1	0
1	0	0
1	1	1

out = in[1] & in[0]

a	b	c	$x = ab + \bar{b}c$
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Truth table



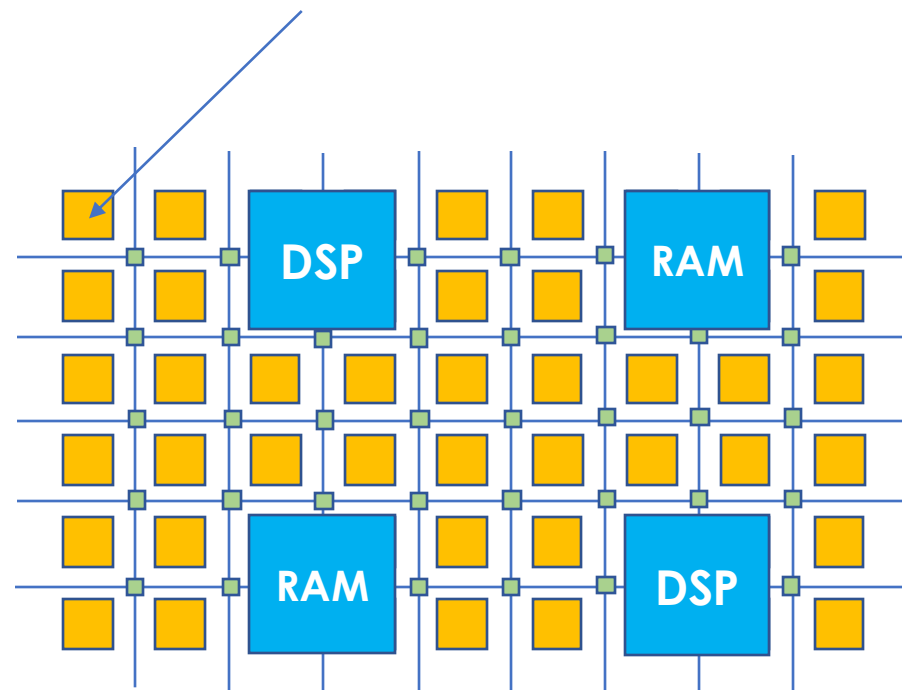
3-Input LUT

LUTs



- Capable of performing any arbitrary functions on small bitwidth inputs (N), generally $N \leq 7$
- Memory location accessed by LUTs: 2^N
- It can be used as both a function compute engine and a data storage element
- LUTs can be combined to create more complex logic circuits.

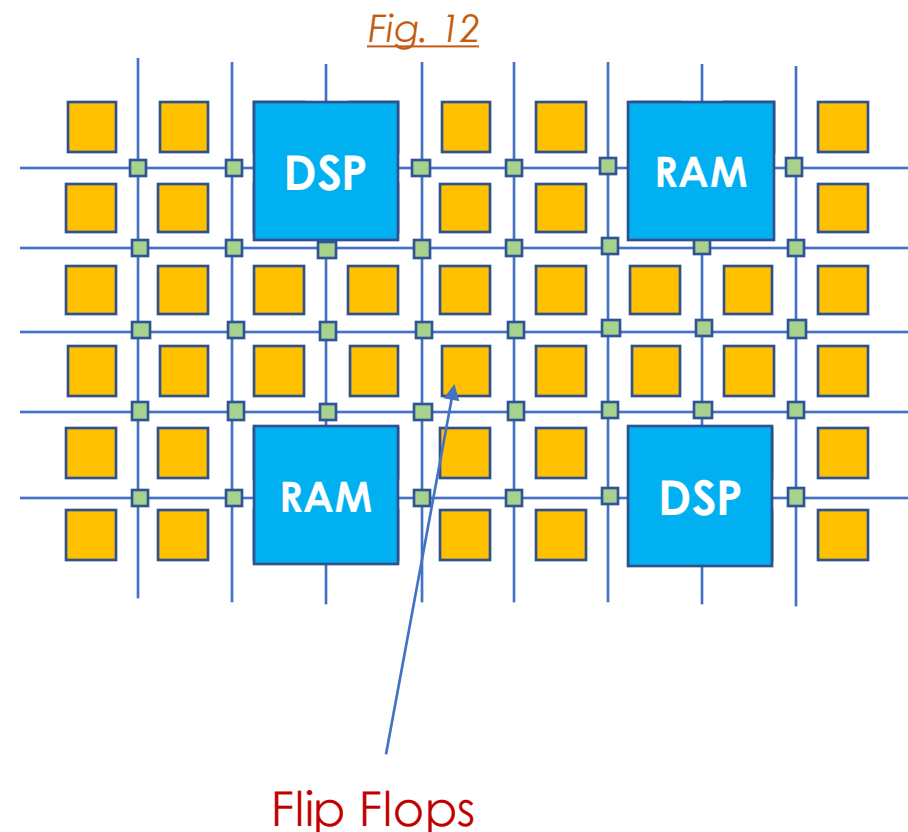
LUTs or logic cells:



Flip-Flops (FF)



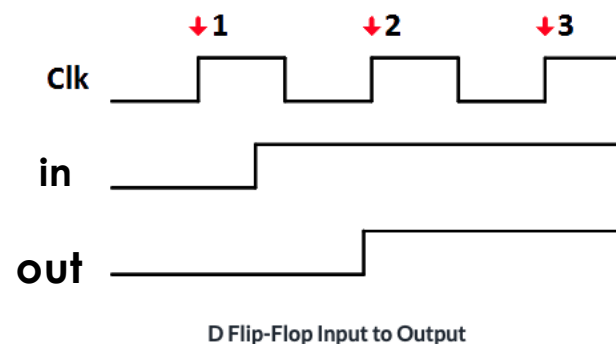
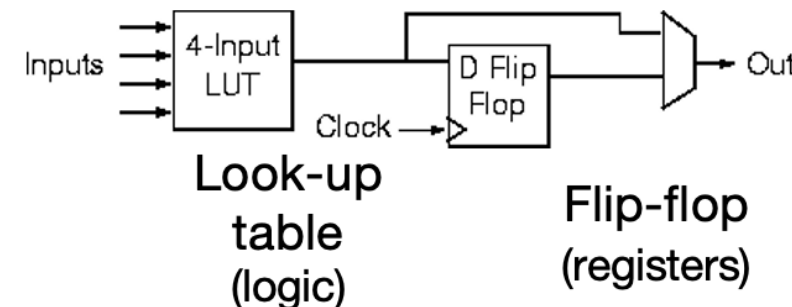
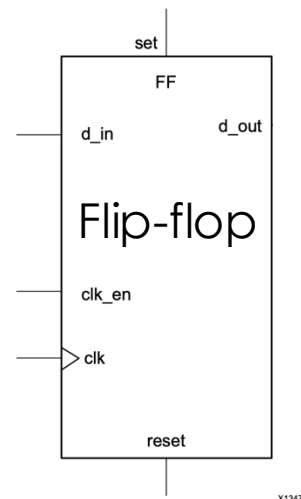
- Basic memory element, co-located with a LUT to assist in logic pipelining and data storage
- Its name comes from its ability to flip or flop between two stable states
- **Operation:** value at the data input port is latched and passed to the output on every pulse of the clock
- They can store data over time



FFs



- The clock is what allows a Flip-Flop to be used as a data storage element
- Any data storage elements are known as **sequential logic** or **registered logic**.
- Sequential logic operates on the *transitions* of a clock. Mostly on the rising edge (when the clock goes from 0 to 1).
- When a Flip-Flop sees a rising edge of the clock, it *registers* the data from the Input to the Output.
- CLBs typically contain several flip-flops that can be used to store intermediate results or outputs of logic functions



Multiplexer



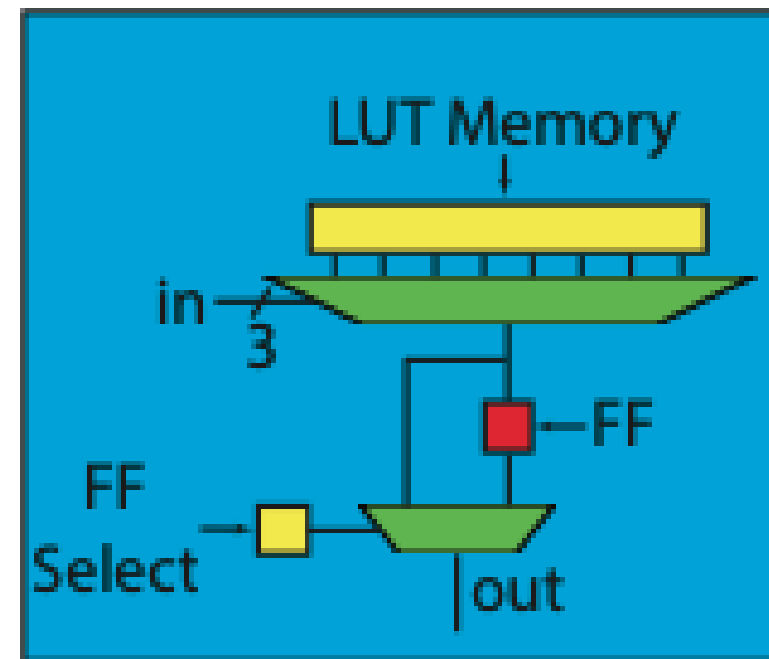
- Combinational logic circuits used to select one input from multiple inputs and pass it to the output based on a control signal
- Routing signals within the CLB
- FPGA interconnects rely on large multiplexers to route signals between logic blocks

Slice



- A small number of LUTs, FFs and multiplexers combined to make a more powerful programmable logic element
- Number & combinations varies per architecture

Slice



I/O Blocks



There are specialised blocks for I/O

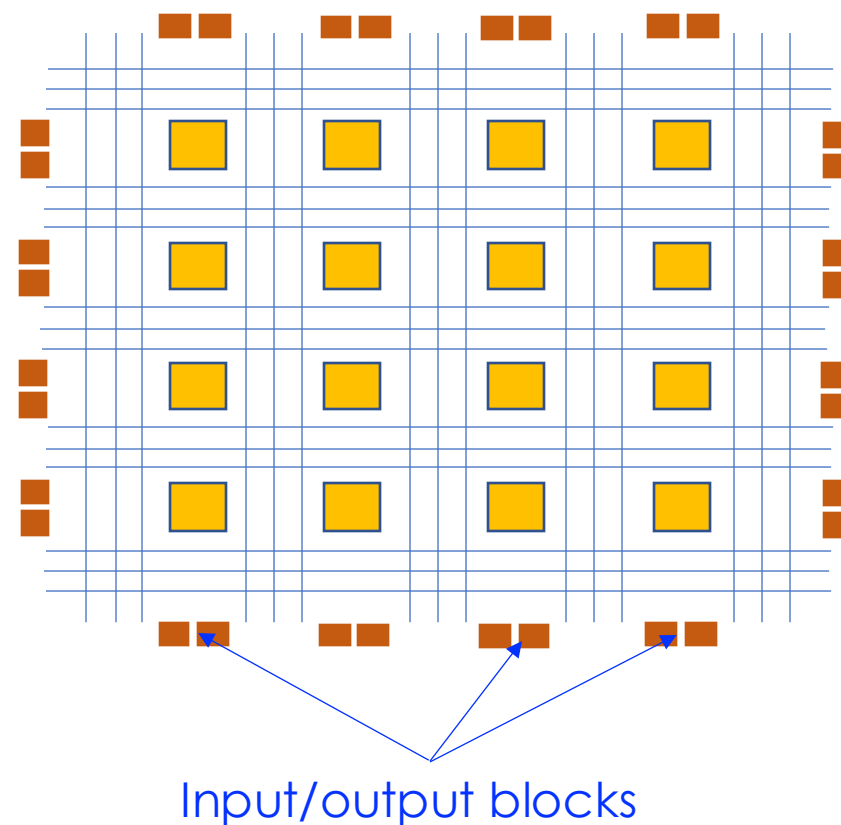
- Interface between FPGA & external devices
- Can be configured to handle a variety of voltage standards
- Making FPGAs popular in embedded systems and HEP triggers

Some of the I/O blocks are bidirectional – can be configured as input or outputs

Dynamically reconfigurable to accommodate changing requirements, such as switching b/w different voltage standards or signal types

Low power per Operation (relative to CPU/GPU)

Fig. 21



I/O Blocks



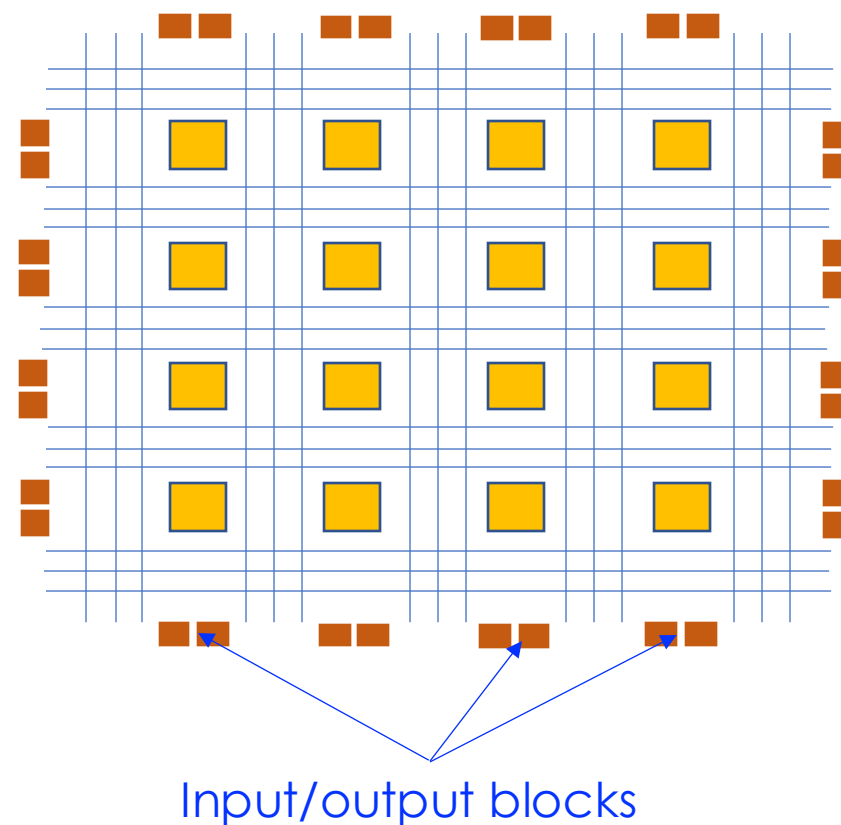
Clocking & Synchronization:

- Used for **clock input & output**, essential for synchronization the FPGA with other devices
- Support **clock domain crossing** & **buffering** to ensure stable signal timing across different clock frequencies

High speed transceivers

- with Tb/s total bandwidth PCIe
- (Multi) Gigabit Ethernet
- Infiniband

Fig. 21



Multi-Gigabit Opto-electronics

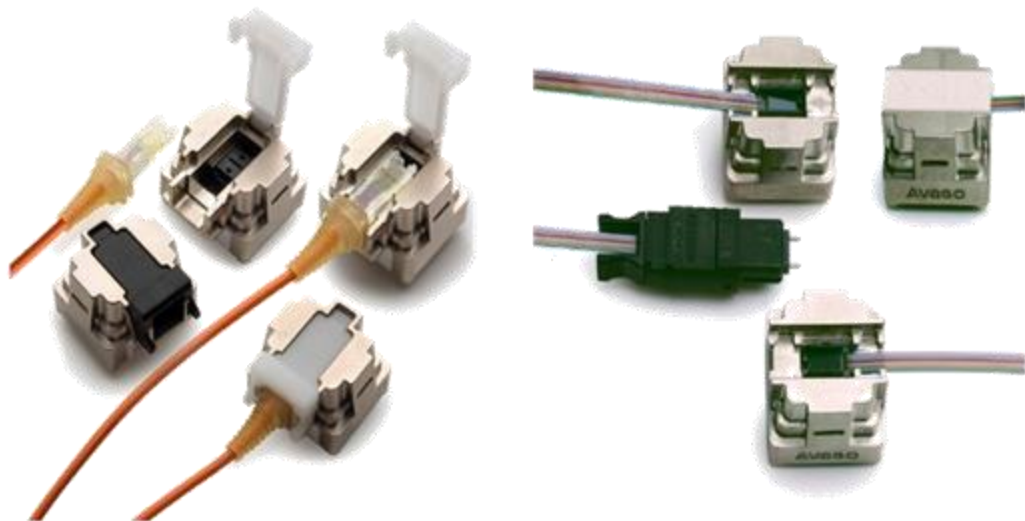


Figure 1. MiniPOD™ Transmitter and Receiver Modules with a) Round Cable and b) Flat Cable: shown with and without dust covers (White = Tx, Black = Rx).

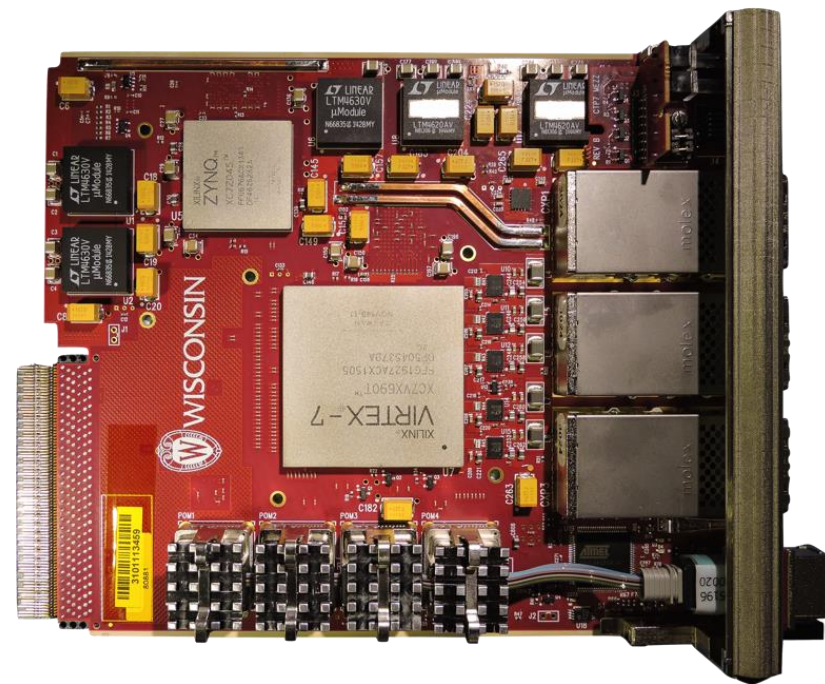
Key Product Parameters

The Avago Technologies MiniPOD™ modules operate at 850 nm and are compliant to the Multi-mode Fiber optical specs in clause 86 and relevant electrical specs in annex 86A of the IEEE 802.3ba specifications.

Parameter	Value	Units	Notes
Data rate per lane	10.3125	Gbps	As per 802.3ba: 100GBASE-SR10 and nPPI specifications
Number of operational lanes	12		100GbE operation utilizes the middle ten lanes (Rx and Tx) of the 12 physically defined lanes
Link Length	100	m	OM3, 2000 MHz·km 50 μm MMF
	150	m	OM4, 4700 MHz·km 50 μm MMF



Figure 2. MiniPOD™ Transmitter and Receiver flat ribbon cable modules in a tiled arrangement example.



Multi-gigabit-per-second serial links



HL-LHC

LHC
→
10 Gbps

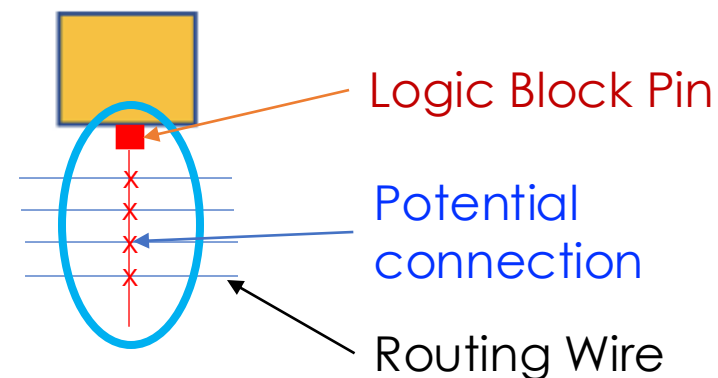
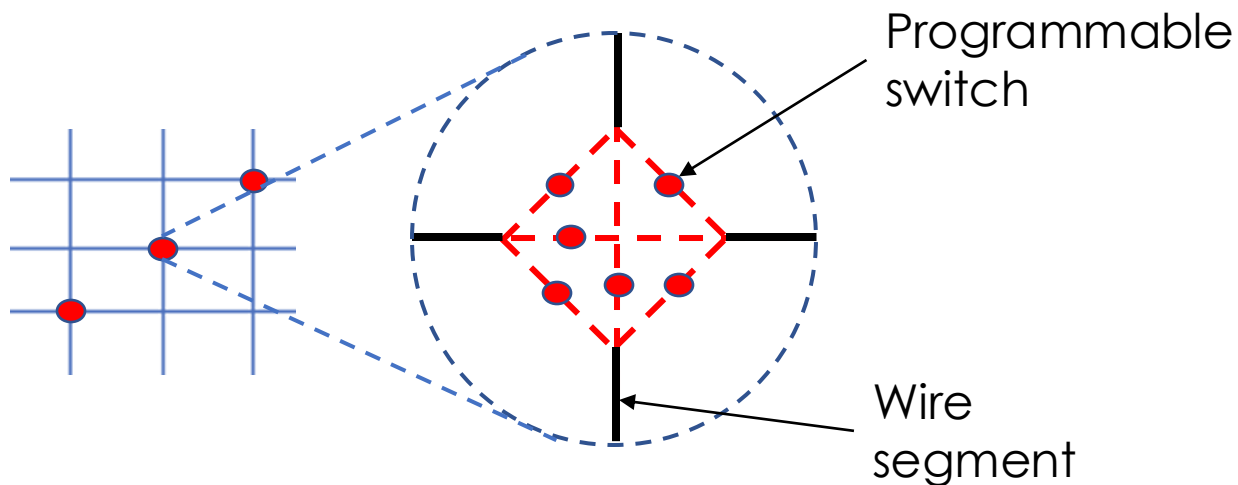
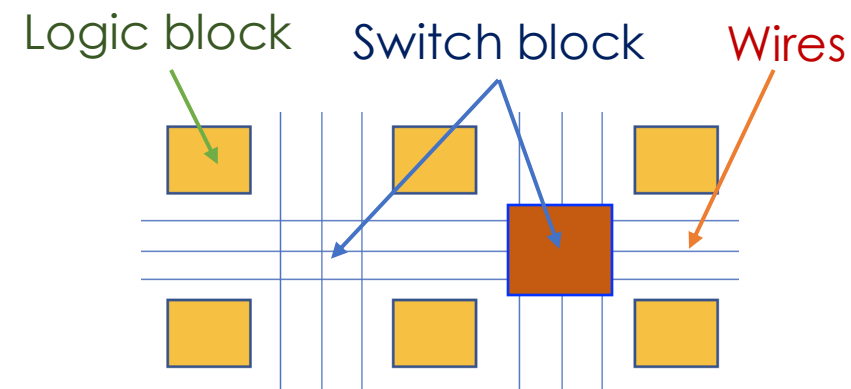
	Type	Max Performance ¹	Max Transceivers	Peak Bandwidth
Virtex UltraScale+	GTY	32.75	128	8,384 Gb/s
Kintex UltraScale+	GTH/GTY	16.3/32.75	44/32	3,268 Gb/s
Virtex UltraScale	GTH/GTY	16.3/30.5	60/60	5,616 Gb/s
Kintex UltraScale	GTH	16.3	64	2,086 Gb/s
Virtex-7	GTX/GTH/GTZ	12.5/13.1/28.05	56/96/16 ³	2,784 Gb/s
Kintex-7	GTX	12.5	32	800 Gb/s
Artix-7	GTP	6.6	16	211 Gb/s
Zynq UltraScale+	GTR/GTH/GTY	6.0/16.3/32.75	4/44/28	3,268 Gb/s
Zynq-7000	GTX	12.5	16	400 Gb/s
Spartan-6	GTP	3.2	8	51 Gb/s

HL-LHC
←
25 Gbps

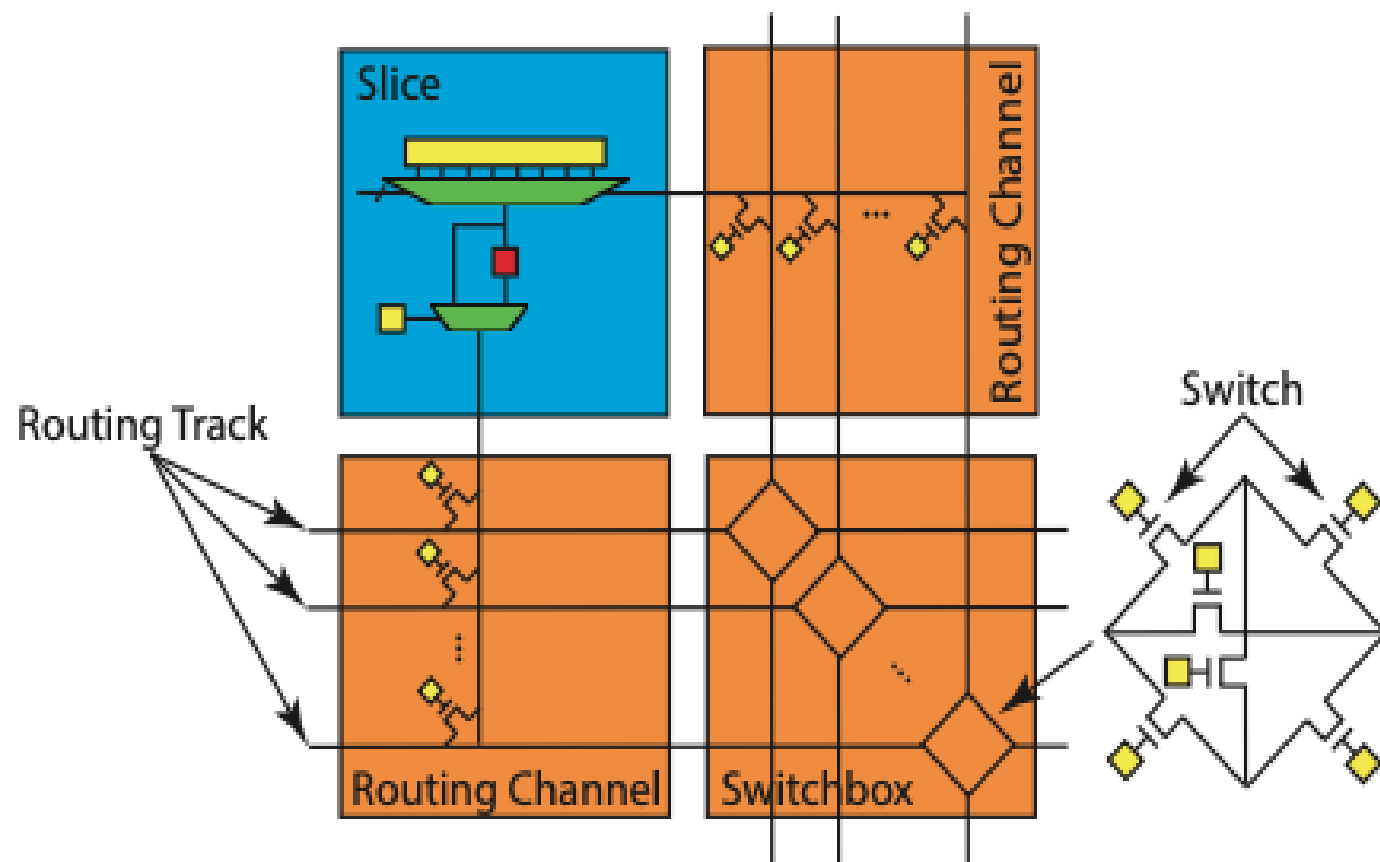
FPGA Components: Routing



- Between rows and columns of logic blocks are wiring channels
- These are programmable – a logic block pin can be connected to one of many wiring tracks through programmable switch
- Xilinx FPGA have dedicated switch block circuits for routing (flexible)
- Each wiring segment can be connected in one of many ways



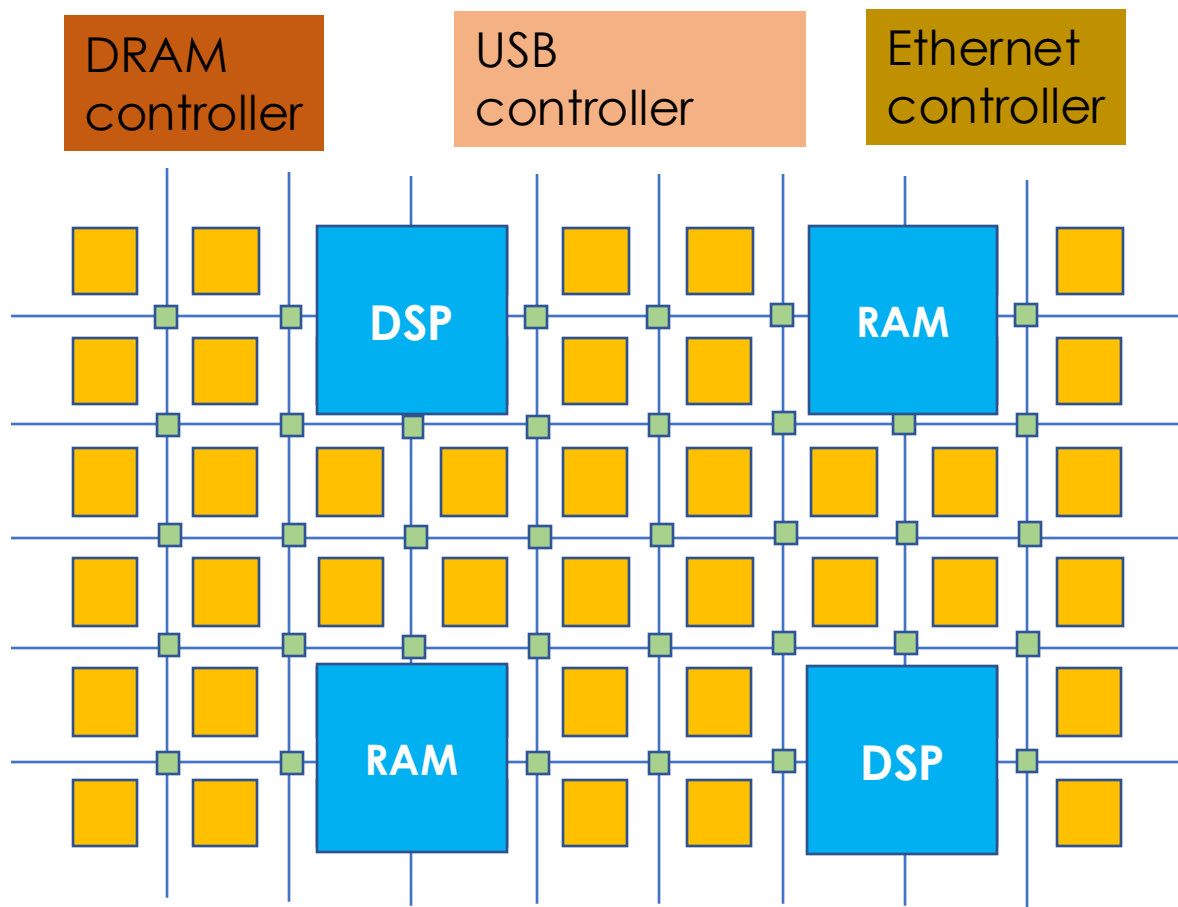
FPGA Components: Routing



- Simple slice with a LUT and a FF
- Slices are connected to one another using a routing channel & switchbox
- These two provide a programmable interconnect that provide the data movement b/w slices.
- The switchbox has many switches (typically implemented as pass transistors) that allow for arbitrary wiring configurations between the different routing tracks in the routing tracks adjacent to the switchbox

The main advantage and attraction of FPGA comes from the programmable interconnect – more so than the programmable logic.

FPGA Architecture



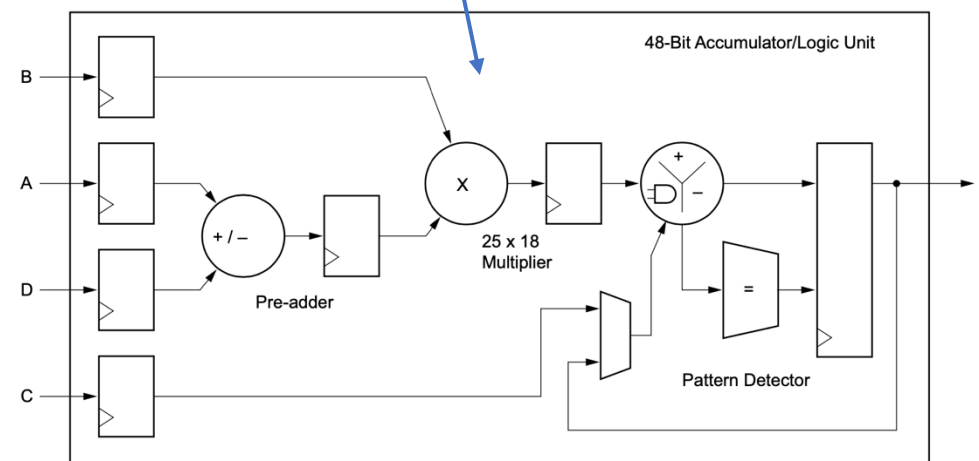
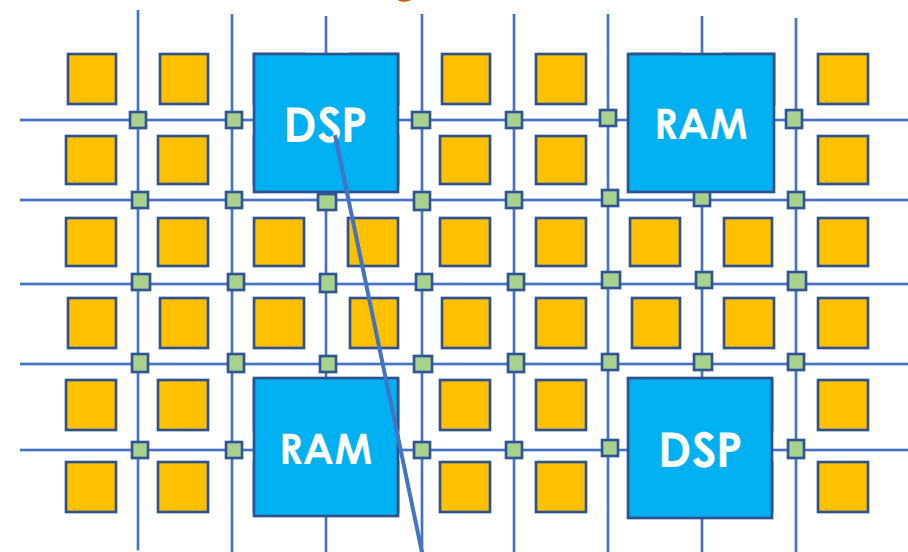
Digital Signal Processing (DSP) Blocks



Built in components for fast arithmetic operation optimized for DSP operations

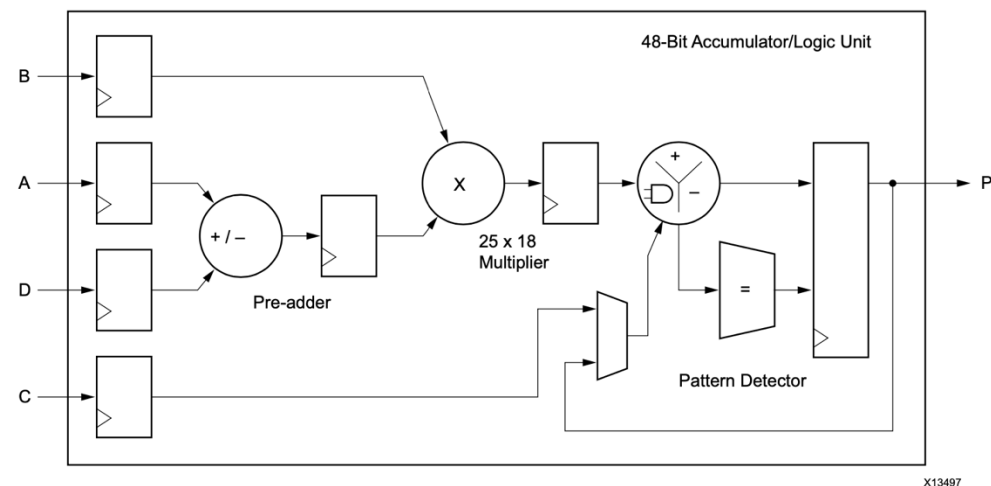
- Most complex computational block available in a FPGA
- Helps in accelerating signal processing algos
- Optimized for high performance **multiplication** and **accumulation**
 - Eg: $p = a \times (b + d) + c$
 - Foundation for many signal processing algorithms, like **filtering**, **transformations**, **convolutions**...
- Faster and more efficient than using LUTs for these types of operations
- Often most scarce in available resources

Fig. 14

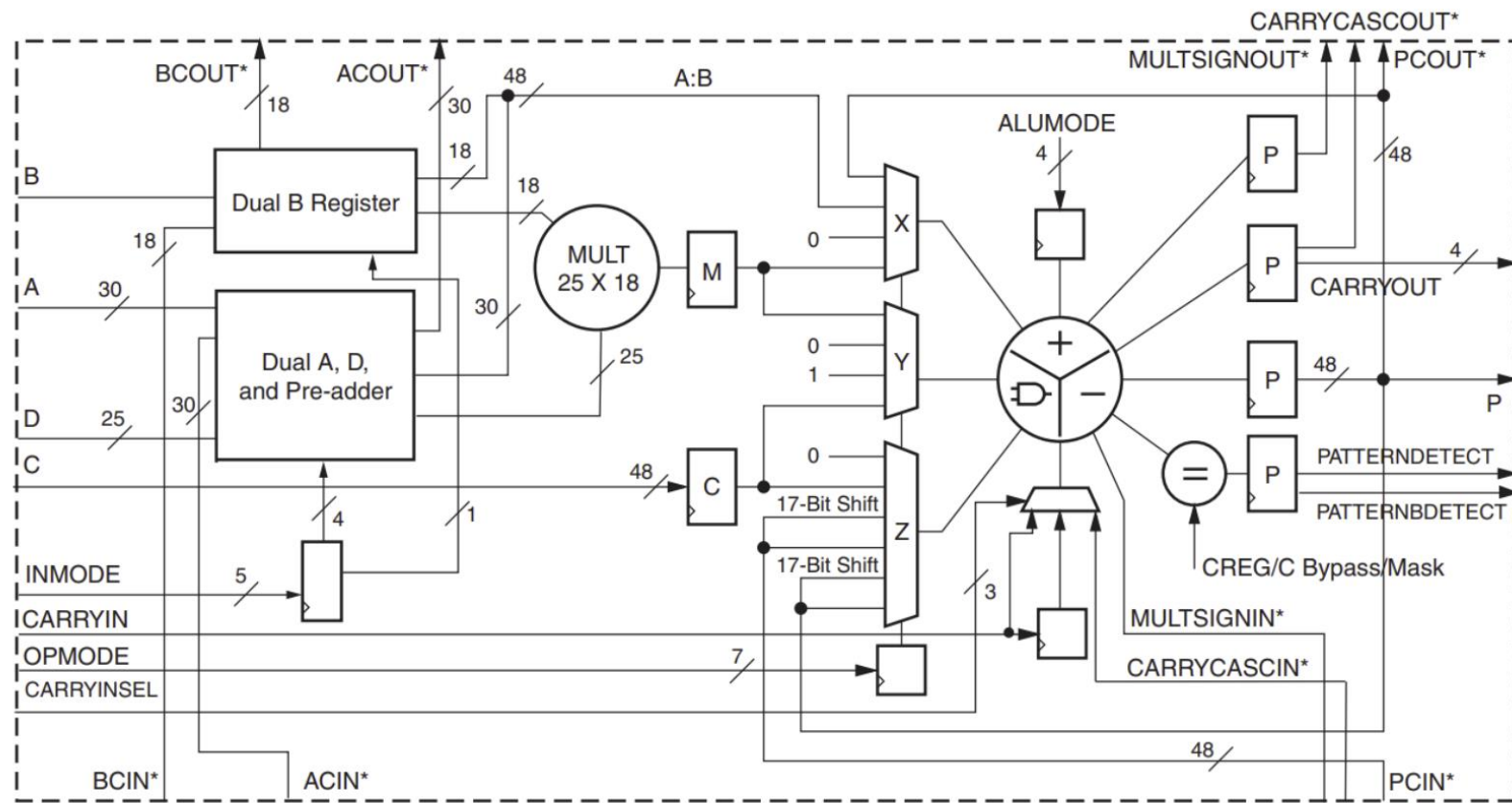




- Each DSP block typically contains a **multiply-accumulate unit**,
 - Allows a multiplication to be directly followed by an accumulation operation (multiplying two values and adding the result to an accumulator)
 - Highly efficient and minimizes the need for multiple steps in traditional processing
- These blocks support various data widths, typically ranging from **8 bits** to **64 bits** or more, and support both **signed** and **unsigned** arithmetic



Example: DSP48E1



*These signals are dedicated routing paths internal to the DSP48E1 column. They are not accessible via fabric routing resources.

- 25x18 signed multiplier
- 48-bit add/subtract/accumulate
- 48-bit logic operations
- SIMD operations (12/24 bit)
- Pipeline registers for high speed

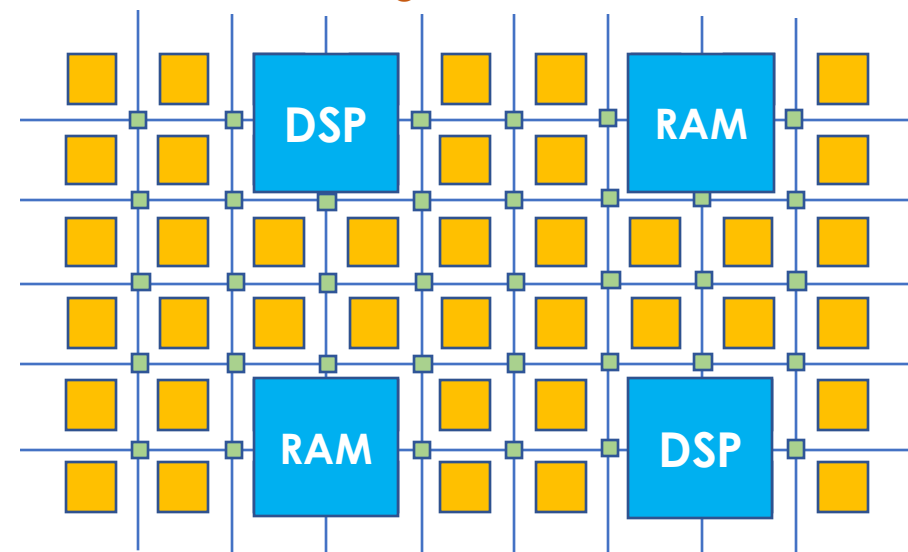
Storage elements: Block RAM



BRAMs are used for storing large amounts of data in a FPGA

- Embedded memory elements that can be used to provide high speed data storage & retrieval
- BRAM is a dual-port RAM module instantiated to provide on-chip storage for a relatively large set of data
 - can hold either 18kb or 36 kb
 - **Multiple Block RAMs** can be used in parallel to create larger memory arrays or buffers.
- Block RAMs are **dual-port**: two independent ports for **reading** and **writing** data simultaneously
 - Useful in applications like **FIFO buffers**, where one port handles writing data while the other port handles reading data

Fig. 16



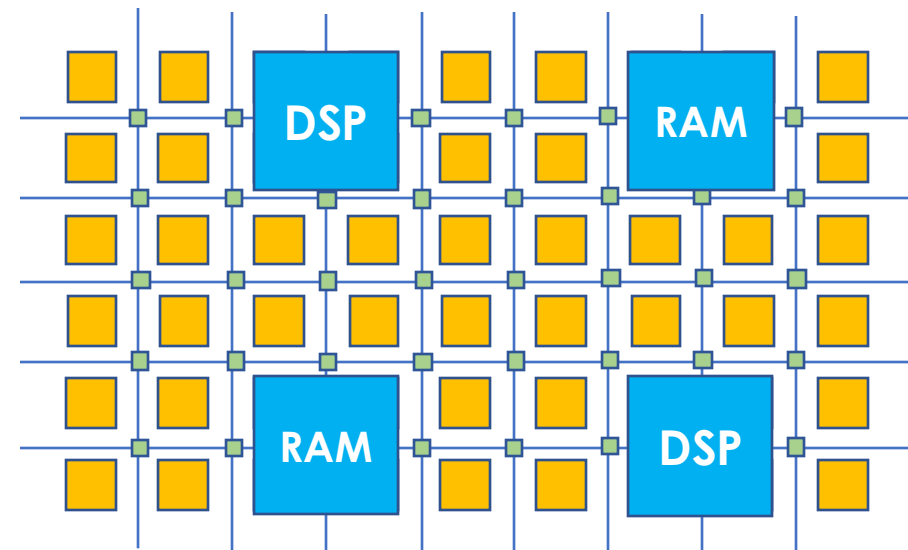
FPGA Components: Storage elements



LUTs as storage element:

- They can be used as 64-b memories due to its structural flexibility
- Commonly referred to as distributed memories
- Fastest kind of memory available on the FPGA device, because it can be instantiated in any part of the fabric that improves the performance of the implemented circuit
- Memories using BRAMs more efficient than using LUTs

Fig. 17





Xilinx Virtex Ultra Scale+ Product Table

COMPARE	Reset	XCVU3P	XCVU5P	XCVU7P	XCVU9P	XCVU11P
System Logic Cells (K)		862	1,314	1,724	2,586	2,835
DSP Slices		2,280	3,474	4,560	6,840	9,216
Memory (Mb)		115.3	168.2	230.6	345.9	341
GTY/GTM Transceivers (32.75/58 Gb/s)		40/0	80/0	80/0	120/0	96/0
I/O		520	832	832	832	624

Source: <https://www.xilinx.com/products/silicon-devices/fpga/virtex-ultrascale-plus.html#productTable>

Decide wisely which FPGA to use as per your needs

Example: Resource Utilization



Utilization Estimates

Summary

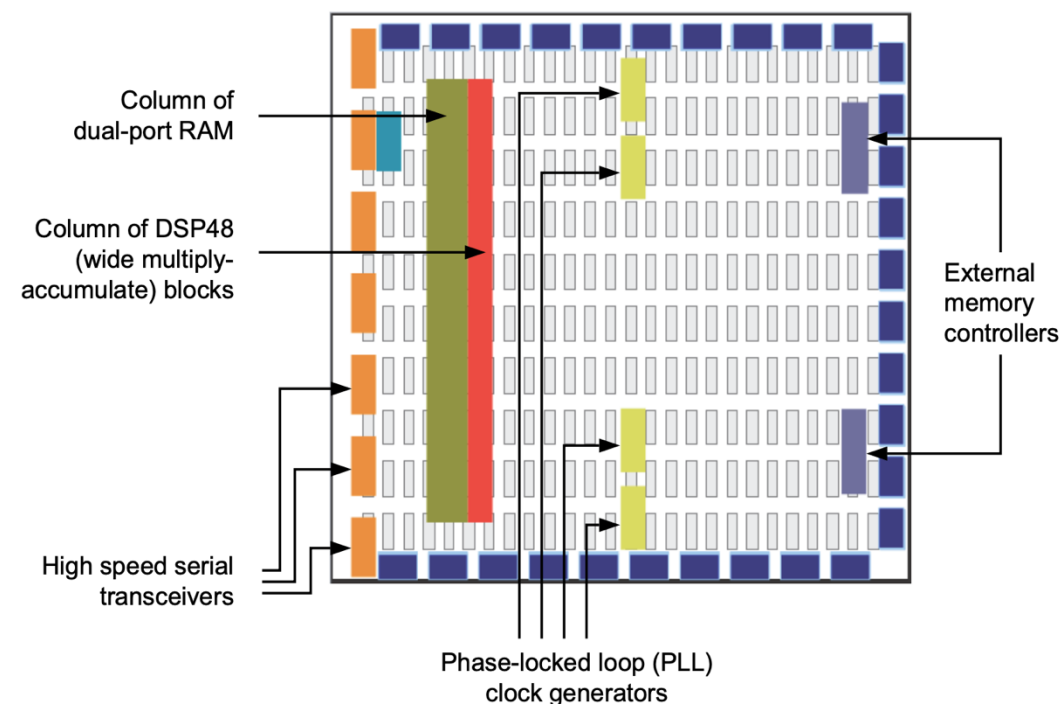
Name	BRAM_18K	DSP48E	FF	LUT	URAM
DSP	-	-	-	-	-
Expression	-	3	0	86	-
FIFO	-	-	-	-	-
Instance	-	-	-	-	-
Memory	0	-	64	6	0
Multiplexer	-	-	-	91	-
Register	-	-	111	-	-
Total	0	3	175	183	0
Available	650	600	202800	101400	0
Utilization (%)	0	~0	~0	~0	0

Clocking Resources



Clocking resources in FPGAs are essential components that manage **clock signals**, ensuring proper timing and synchronization across various logic blocks

- FPGAs include specialized **clock management circuits** to distribute and modify clock signals efficiently, enabling high-performance designs
- **Global Clock Networks**
 - Distribute clocks efficiently across the entire FPGA
- **Phase-locked loops (PLLs)** for driving the FPGA fabric at different clock rates
 - Adjusts the clock frequency by **multiplying** or **dividing** an input clock
 - Reduces **jitter** and maintains **clock stability**
 - Used in **frequency synthesis** and **clock recovery**



X13468

Fig. 9

Why are FPGAs fast



Fine-grained/resource parallelism

- Use the many resources to work on different parts of the problem simultaneously
- Allows us to achieve **low latency**

Most problems have at least some sequential aspect, limiting how low latency we can go

- But we can still take advantage of it with...

Pipeline parallelism

- Use the register pipeline to work on different data simultaneously
- Allows us to achieve **high throughput**



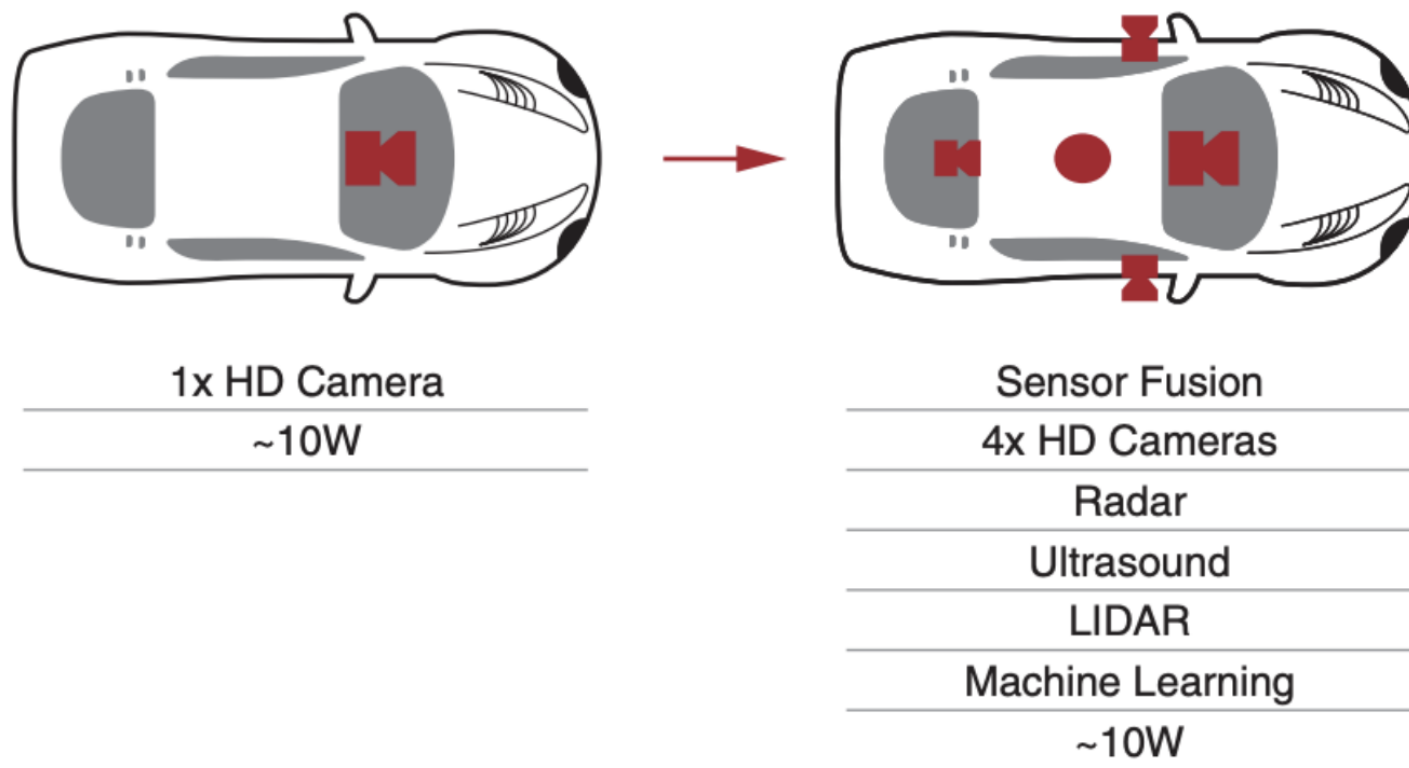
Fig. 22: Like a production line for data...

More Advanced Architectures



- Embedded FPGA System on Chip (SoC)
- High Bandwidth Memory (HBM) on Xilinx FPGA
 - A theoretical bandwidth up to 460 GB/s
- ACAP: Adaptive Compute Acceleration Platform
 - A fully software-programmable, heterogeneous compute platform that **combines Scalar Engines, Adaptable Engines, and Intelligent Engines** to achieve dramatic performance improvements of up to 20X over today's fastest FPGA implementations and over 100X over today's fastest CPU implementations—for Data Center, wired network, 5G wireless, and automotive driver assist applications.

ACAP Application



WP505_13_092818

Xilinx ACAP Devices enable sensor fusion in small power envelopes

Summary



- The information does not cover all the details about the FPGA architecture
- Rather to a concise report of some useful information needed to understand the HLS reports and successfully use and leverage the HLS directives, many of which very specifically target modern FPGA architectural features.



TAC-HEP 2025

Questions?



Xilinx FPGAs – Phase-1 choice: V7 690T

Xilinx Multi-Node Product Portfolio Offering

45nm

SPARTAN⁶

28nm

VIRTEX⁷

KINTEX⁷

Currently Deployed

20nm

VIRTEX⁷ UltraSCALE

KINTEX⁷ UltraSCALE

16nm

VIRTEX⁷ UltraSCALE+

KINTEX⁷ UltraSCALE+

Product Tables and Product Selection Guide



Cost-Optimized Portfolio

Spartan-7	Spartan-6
Artix-7	Zynq-7000

Product Tables and Product Selection Guide



7 Series

Spartan-7	Artix-7
Kintex-7	Virtex-7

Product Tables and Product Selection Guide



UltraScale

Kintex UltraScale	Virtex UltraScale
-------------------	-------------------

Product Tables and Product Selection Guide



UltraScale+

Kintex UltraScale+	Virtex UltraScale+
--------------------	--------------------

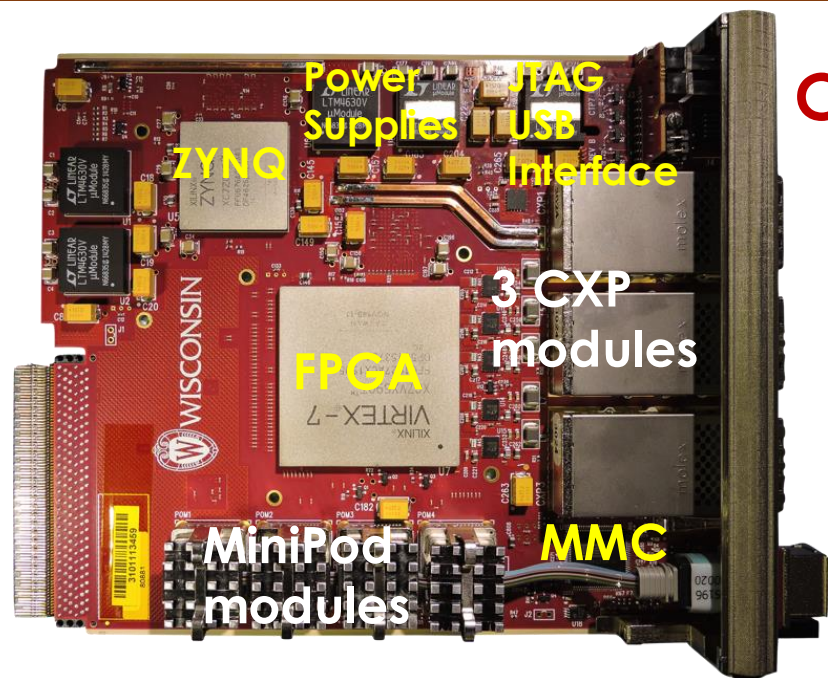
Decide wisely which FPGA to use as per your needs

	Spartan-7	Artix-7	Kintex-7	Virtex-7
Max Logic Cells (K)	102	215	478	1,955
Max Memory (Mb)	4.2	13	34	68
Max DSP Slices	160	740	1,920	3,600
Max Transceiver Speed (Gb/s)	--	6.6	12.5	28.05
Max I/O Pins	400	500	500	1,200

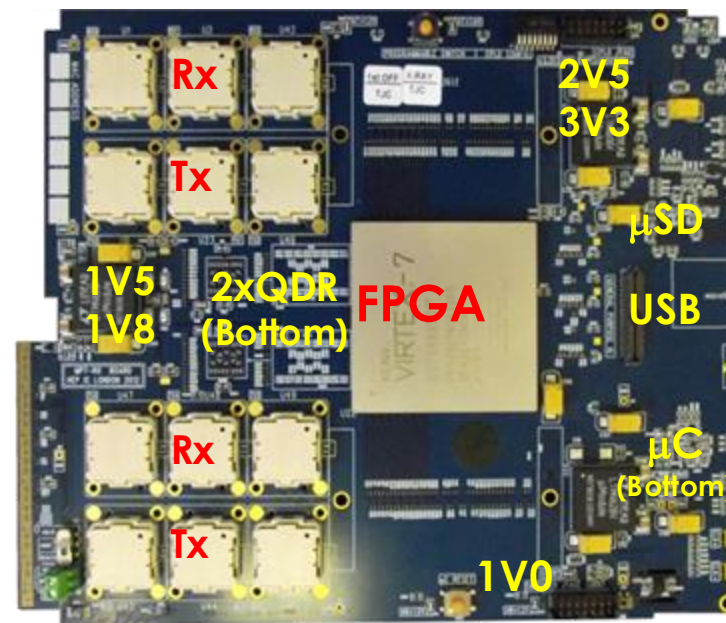
Speed grade:
maximum propagation delay for critical paths in the FPGA fabric or I/O operations

HL-LHC Prototypes

Trigger Processor Boards



CTP7



MP7

Calorimeter Trigger Processor(CTP7 - left), and Master Processor (MP7 - right)

- **CTP7 (Layer-1) - mTCA Single Virtex 7 FPGA, 67 optical inputs, 48 outputs, 12 RX/TX backplane**
 - Virtex 7 allows 10 Gb/s link speed on 3 CXP(36 TX & 36 RX) and 4 MiniPODs (31 RX & 12 TX)
 - ZYNQ processor running Xilinx PetaLinux for service tasks, including virtual JTAG cable
- **MP7 (Layer-2) - mTCA Single Virtex 7 FPGA, up to 72 input & output links**
 - Virtex 7 has 72 input and output links at 10 Gb/s
 - Dual 72 or 144MB QDR RAM clocked at 500 MHz



Workflow during FPGA development

