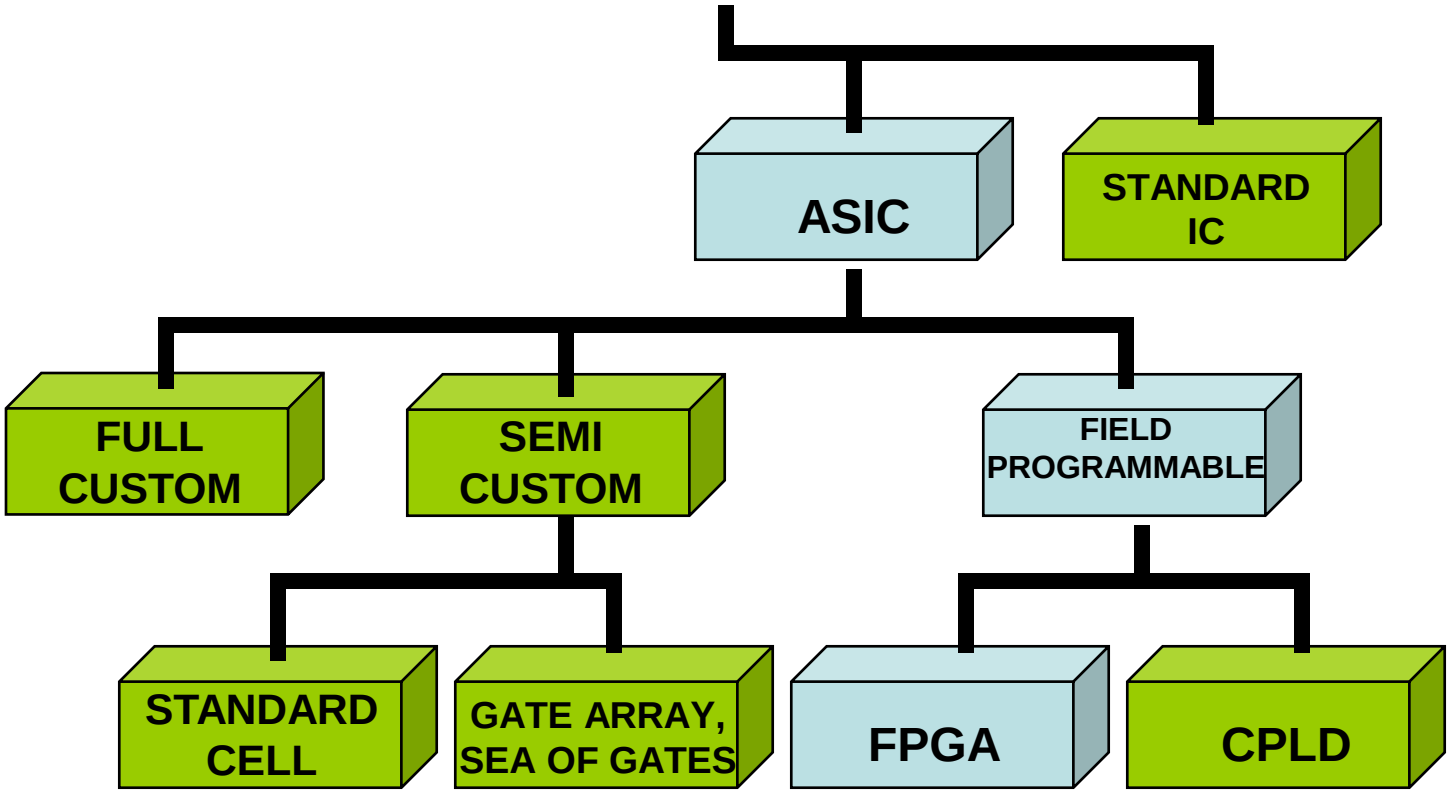
Programmable Logic Devices - FPGA

Parikshit Godbole
High Performance Computing -Technologies
C-DAC

2024 © Centre for Development of Advanced Computing

CMOS IC Design Alternatives



2024 © Centre for Development of Advanced Computing

World Of Programmable Logic

- A quiet revolution is taking place over the past few years, the density of the average programmable logic device has begun to skyrocket.
- The maximum number of gates in an FPGA is currently more than 20 Million (ASIC equivalent) and doubling every 18 months.
- Many types of programmable logic devices are available. The current range of offerings includes everything from small devices capable of implementing only a handful of logic equations to huge FPGAs that can hold an entire processor core (plus peripherals!).
- In addition to this incredible difference in size there is also much variation in architecture.

2024 © Centre for Development of Advanced Computing

World Of Programmable Logic

- Three phases of FPGA evolution called “Ages” by American computer scientist Stephen Trimberger
  - Age of Invention (1984-1991); FPGAs were smaller than the user application itself
  - Age of Expansion (1992-1999); FPGA vendors competing ASIC and EDA technology
  - Age of Accumulation (2000-2007); FPGAs became common components of digital systems

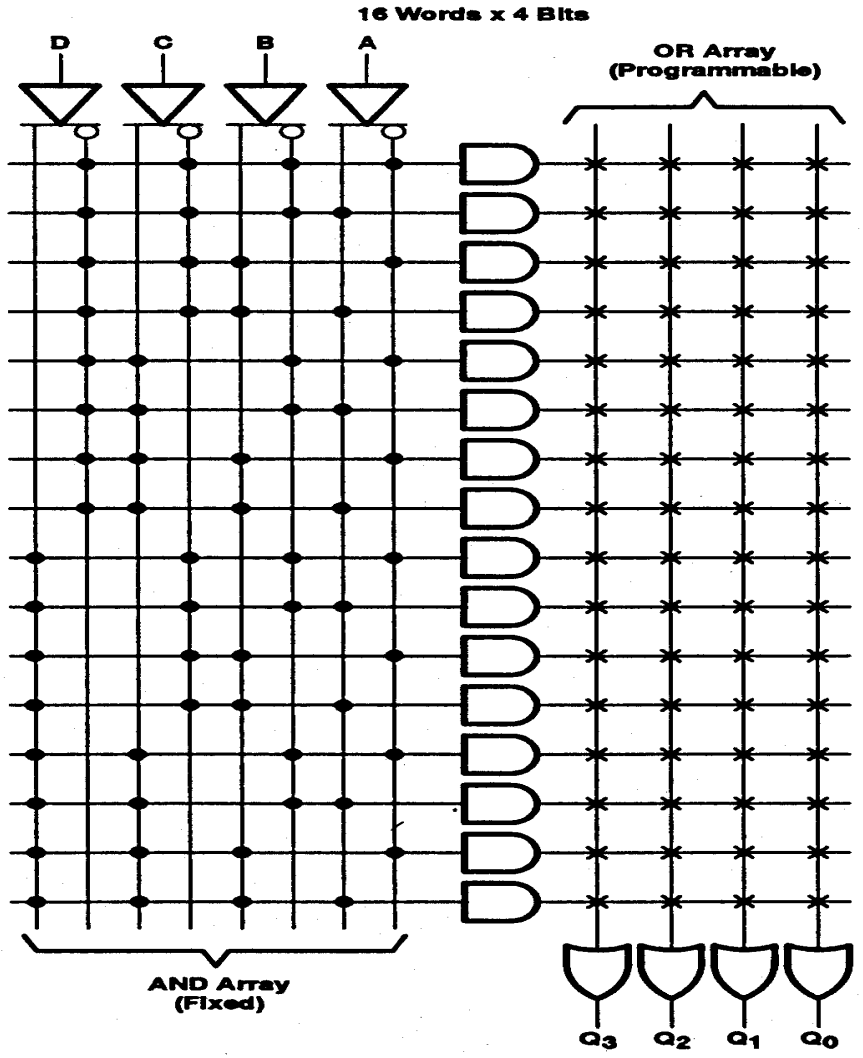
2024 © Centre for Development of Advanced Computing

PROM : The first PLD

- Basically available in 2 types viz. One time programmable (OTP) and field programmable.
- The field programmable had 2 types viz. EPROM based and EEPROM based.
- Its basic architecture is an input decoder configured from AND gates, combined with a programmable OR matrix on the outputs.
- This allows every output to be programmed individually from every possible input combination.
- As shown in Figure, with 4 inputs we have 16 combinations and with output width 4, each of the 16X4 bit words can be programmed individually.

2024 © Centre for Development of Advanced Computing

PROM



2024 © Centre for Development of Advanced Computing

Why PROM not used as a PLD

- Overall inefficient architecture for realizing logic circuits. (for example to implement O3 = D we require, sixteen 4-input AND gates and one 16-input OR gate).
- PROM have difficulty accommodating large number of input variables. (Since each input variable added, the size of fuse matrix doubles, additional address decoding logic is required)

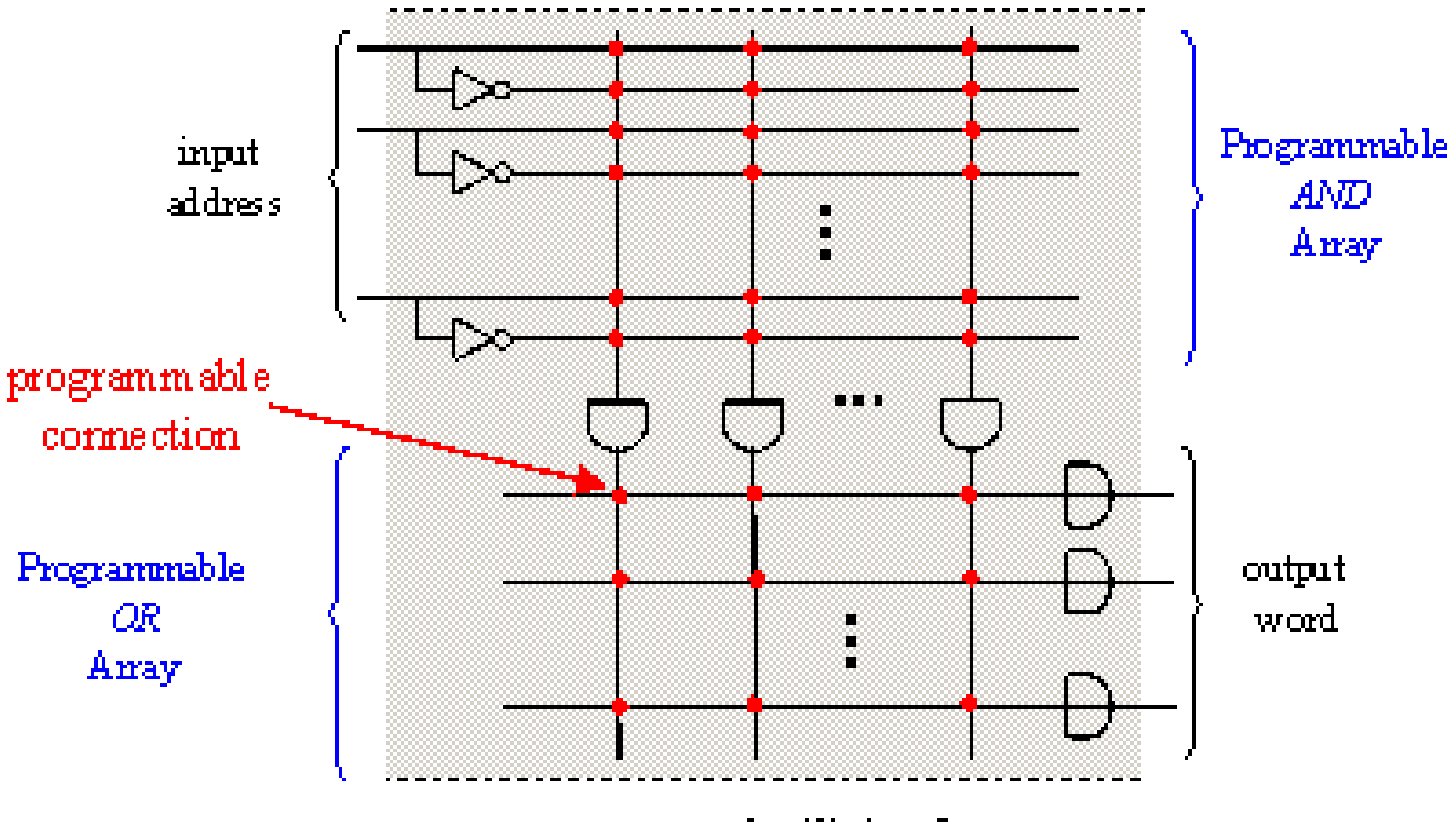
2024 © Centre for Development of Advanced Computing

PLA (Programmable Logic Array)

- First device specially for implementing logic circuits, introduced in the early 1970s by Philips.
- Consists of 2 level of logic gates : a programmable “wired” AND-plane followed by a programmable “wired” OR-plane.
- Designed to implement random logic expression in SOP form.

2024 © Centre for Development of Advanced Computing

PLA (Programmable Logic Array)



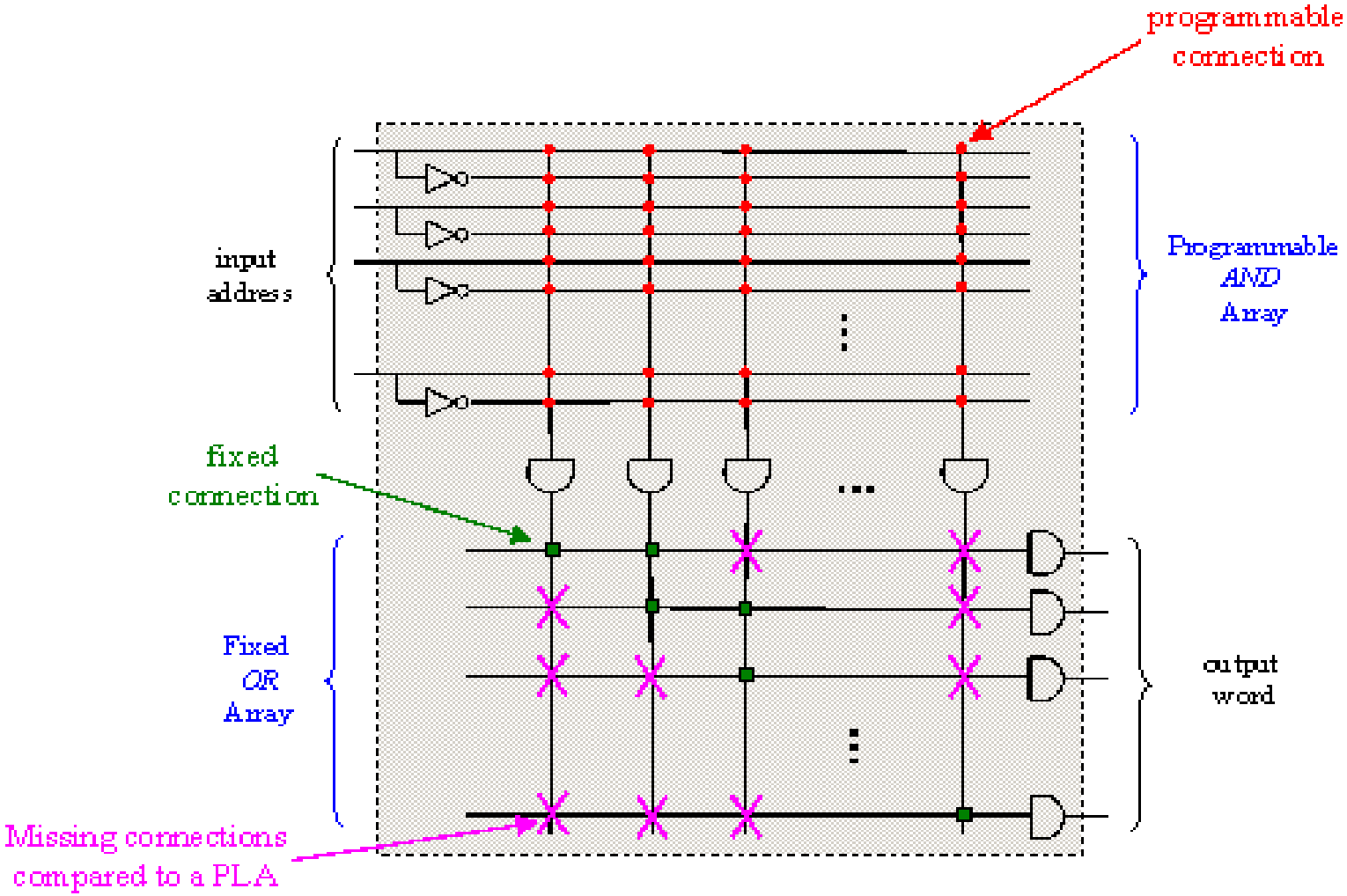
2024 © Centre for Development of Advanced Computing

PAL (Programmable Array Logic)

- Consists of 2 level of logic gates : a programmable “wired” AND-plane followed by a fixed OR-plane.
- Logic is represented in SOP form.
- The number of products in a SOP form will be limited to a fixed number.
- The number of variables in each product term limited by number of input pins.
- The number of independent functions limited by number of output pins.
- Features like Bi-directional inputs was added in advanced devices.

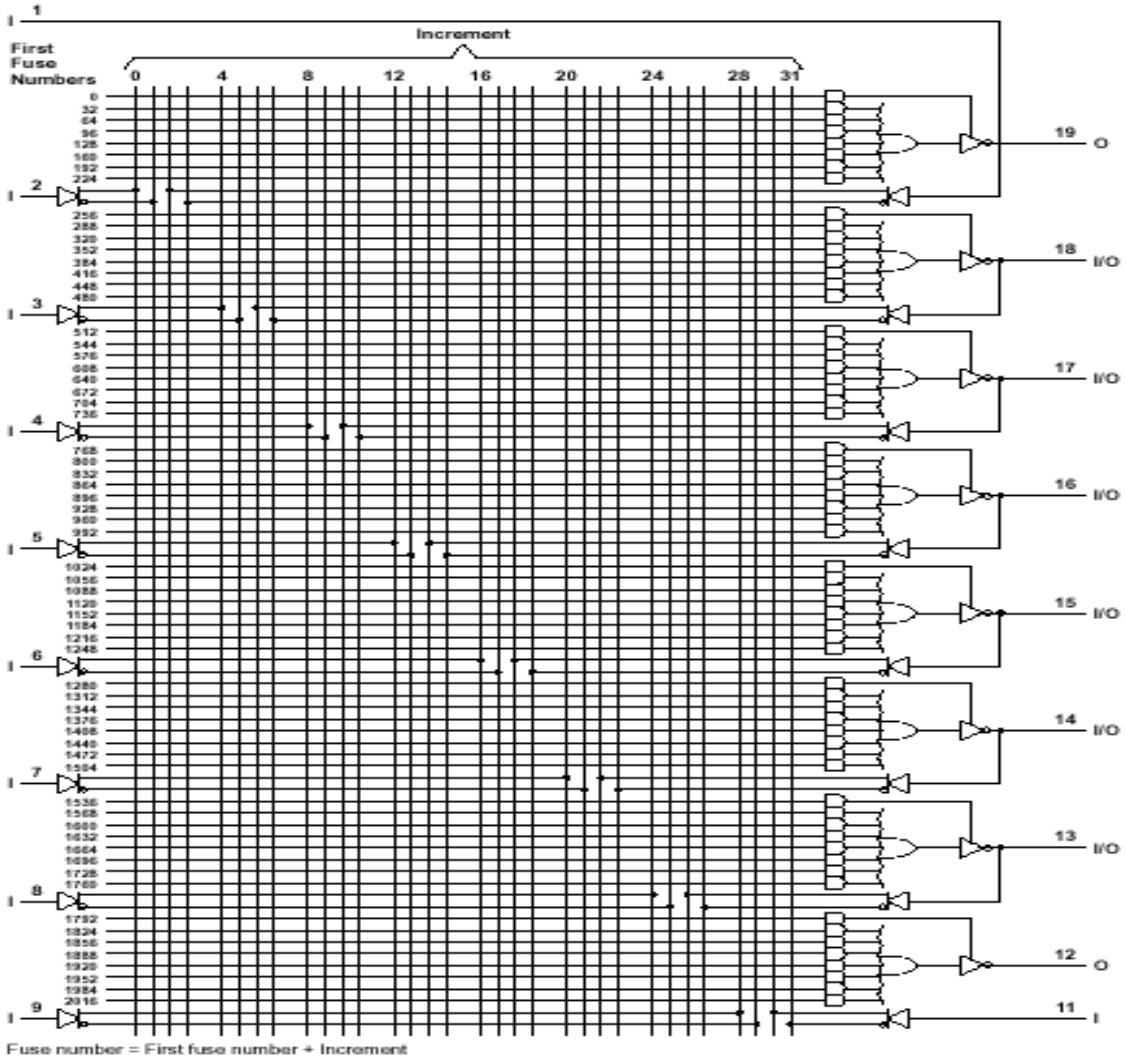
2024 © Centre for Development of Advanced Computing

PAL (Programmable Array Logic)



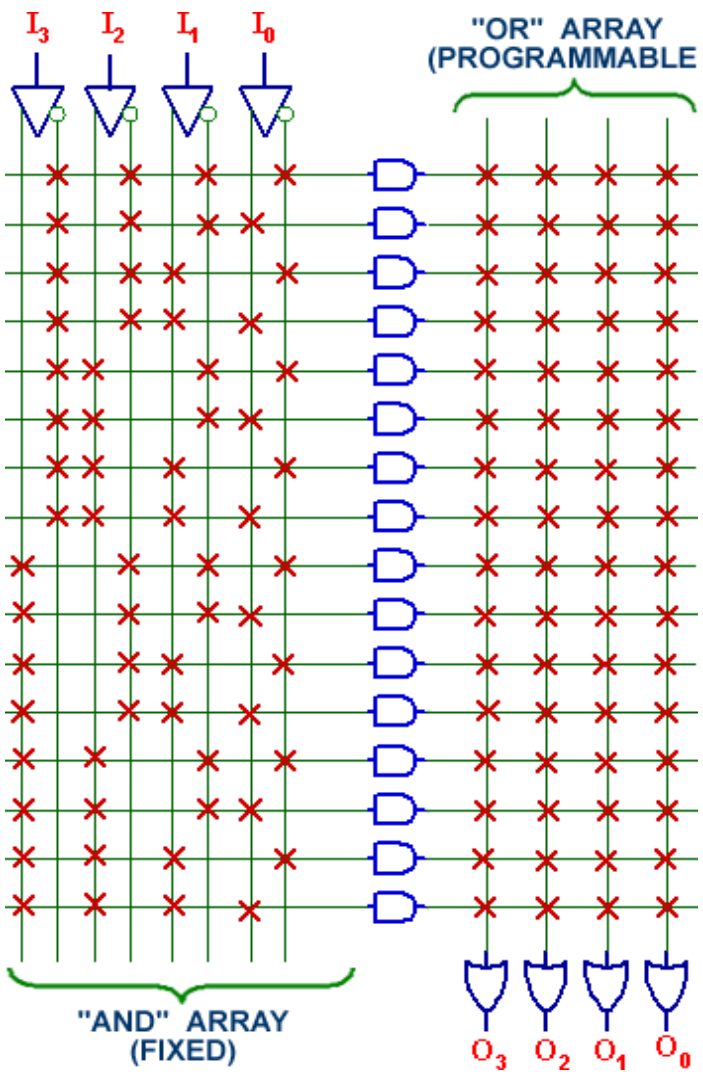
2024 © Centre for Development of Advanced Computing

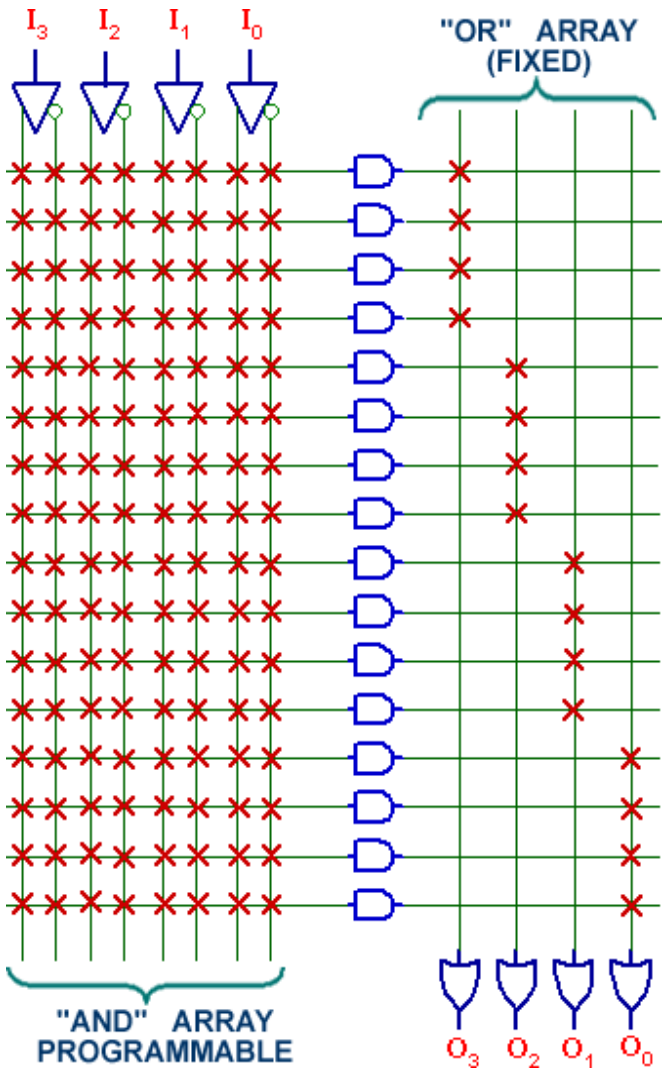
TIBPAL 16L8 Logic Diagram



2024 © Centre for Development of Advanced Computing

PROM Vs. PAL





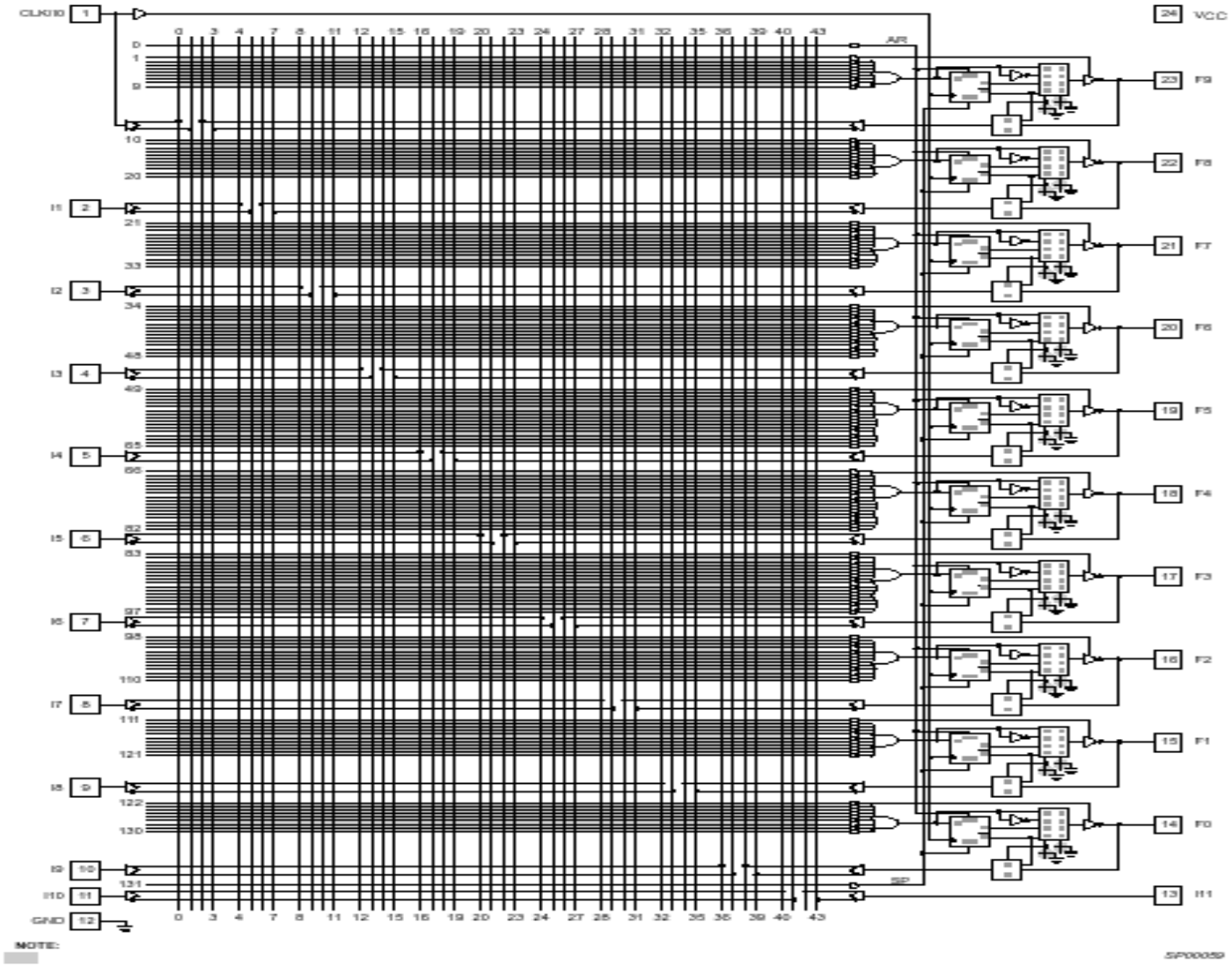
2024 © Centre for Development of Advanced Computing

PLD

- To implement sequential circuits & have flexibility at the IO, some manufactures like Xilinx added OMC (Output Macro Cell).
- The OR sop feeds an “Output Macro Cell (OMC)”, which can be individually configured as a dedicated input, a combinational output or a registered output with a internal feedback.

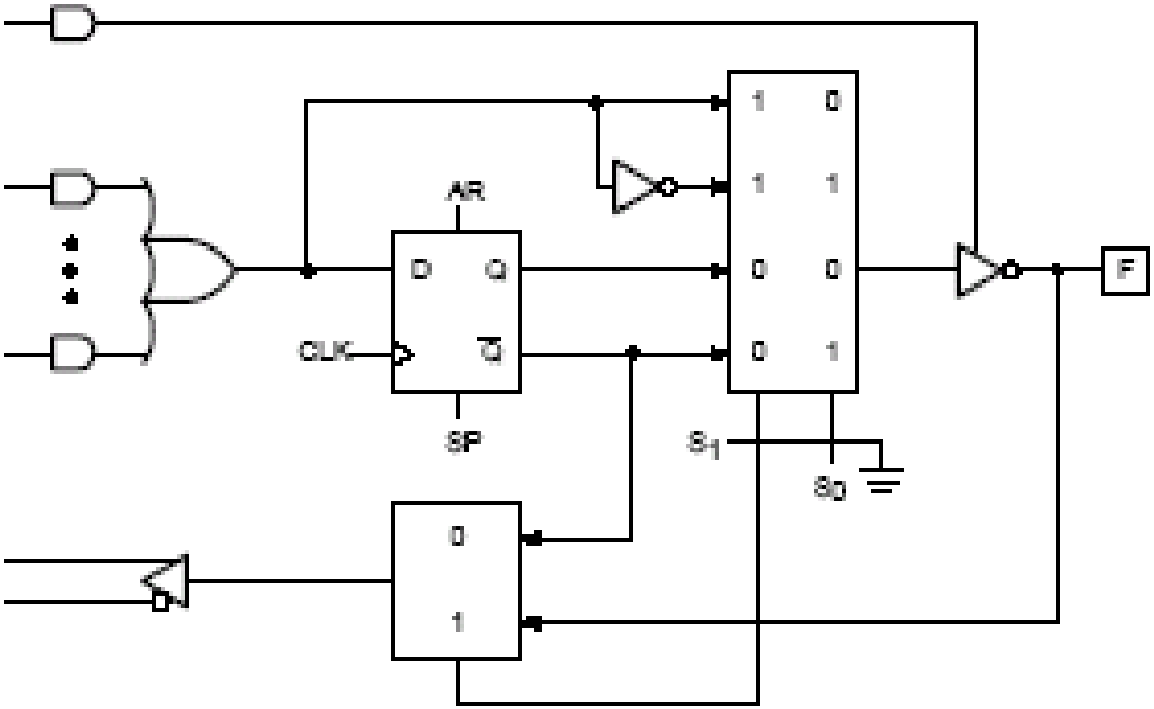
2024 © Centre for Development of Advanced Computing

XCR22LV10 Logic Diagram



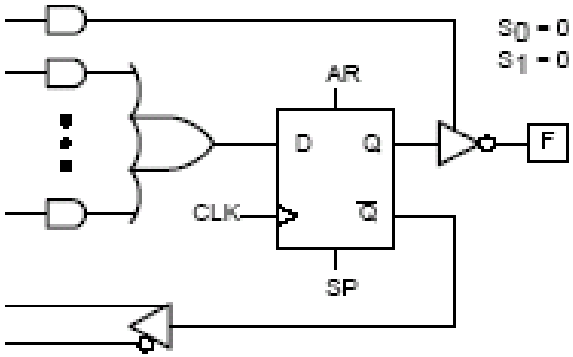
2024 © Centre for Development of Advanced Computing

Output Macro Cell Diagram

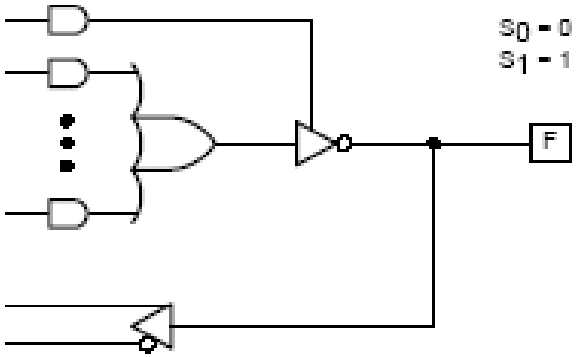


2024 © Centre for Development of Advanced Computing

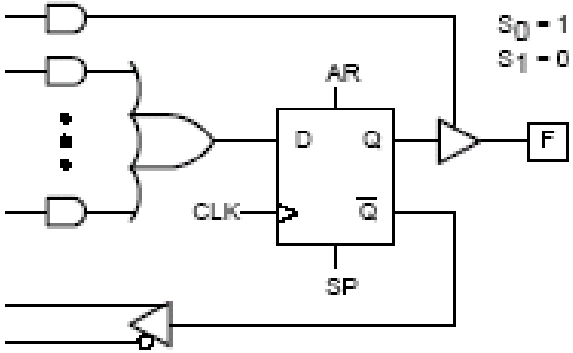
Output Macro Cell configuration



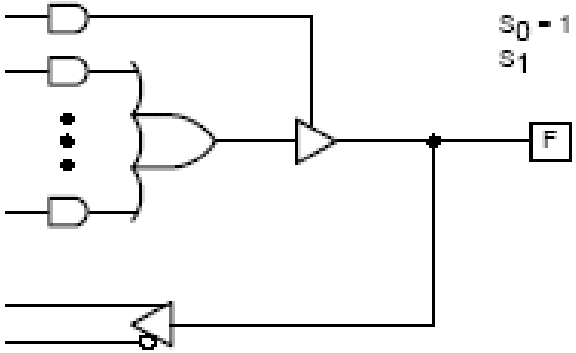
a. Registered/Active-LOW



c. Combinatorial/Active-LOW



b. Registered/Active-HIGH



d. Combinatorial/Active-HIGH

2024 © Centre for Development of Advanced Computing

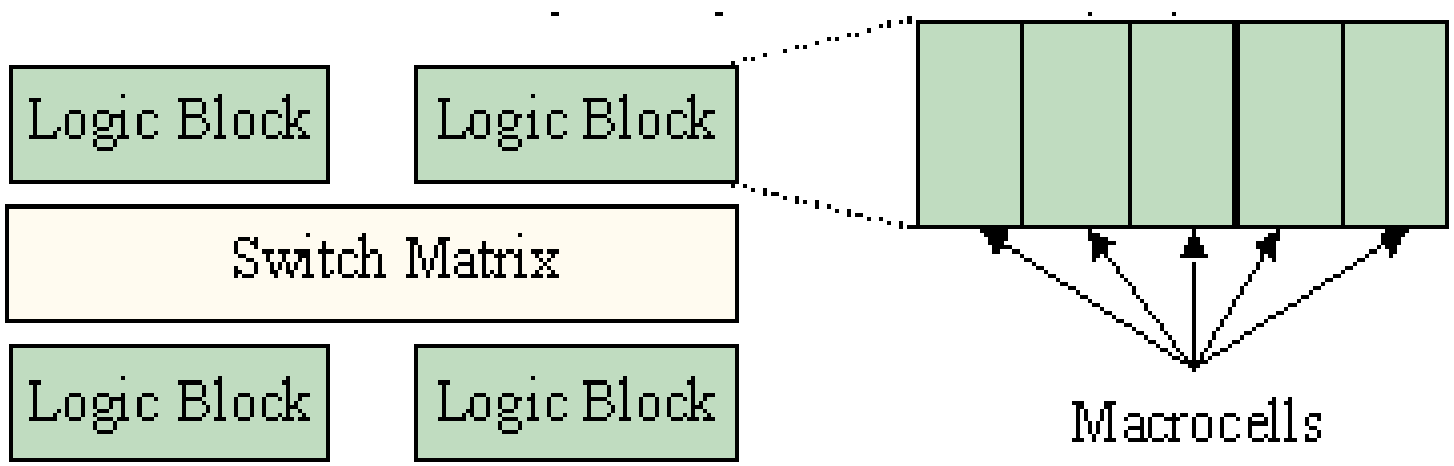
Evolution of PLDs

- As technology has advanced, it has become possible to produce devices with higher capacity.
- The structure of PALs grow too quickly in size as number of inputs is increased and thus its not possible to produce PAL devices with large capacity.
- The next generation PLD architectures then integrated multiple PALs onto a single chips, and they are called as CPLD.

2024 © Centre for Development of Advanced Computing

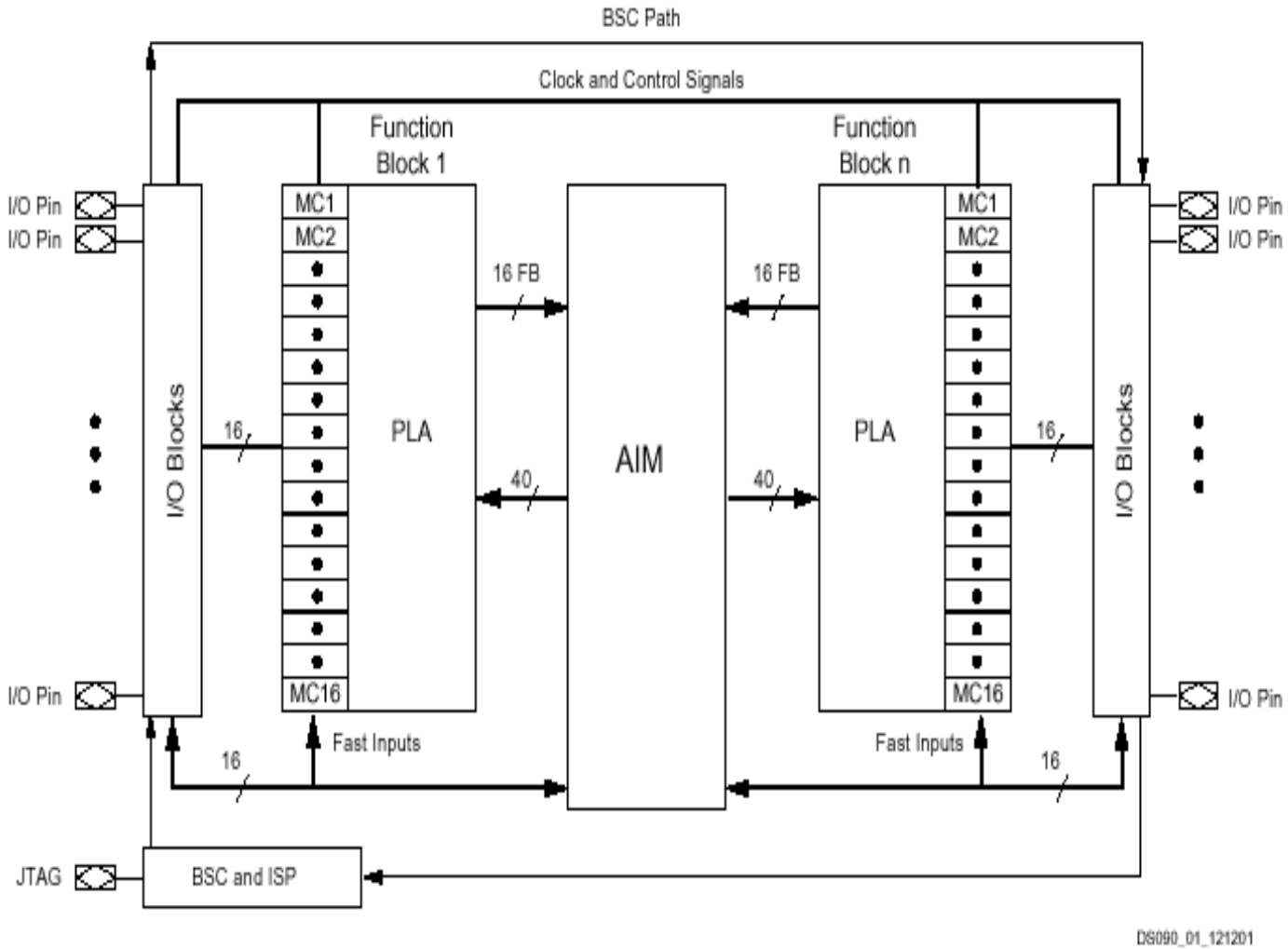
CPLD (Complex PLD)

- Consists of multiple PALs-like logic blocks interconnected via a programmable switch matrix.



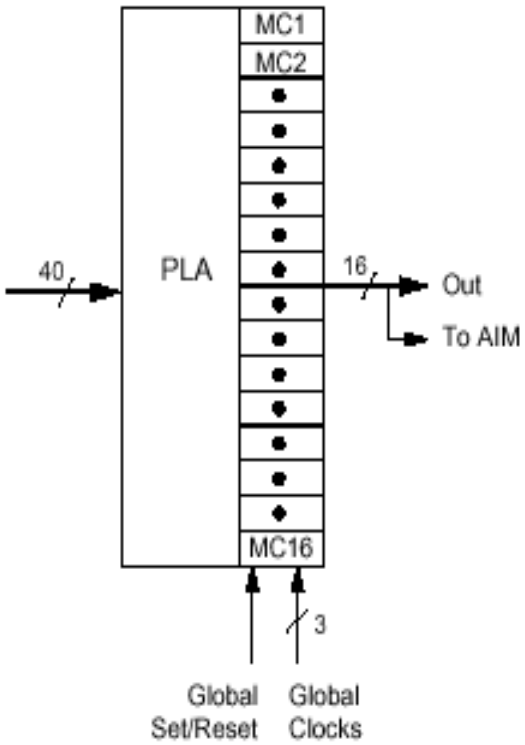
2024 © Centre for Development of Advanced Computing

A look at Xilinx CPLD



2024 © Centre for Development of Advanced Computing

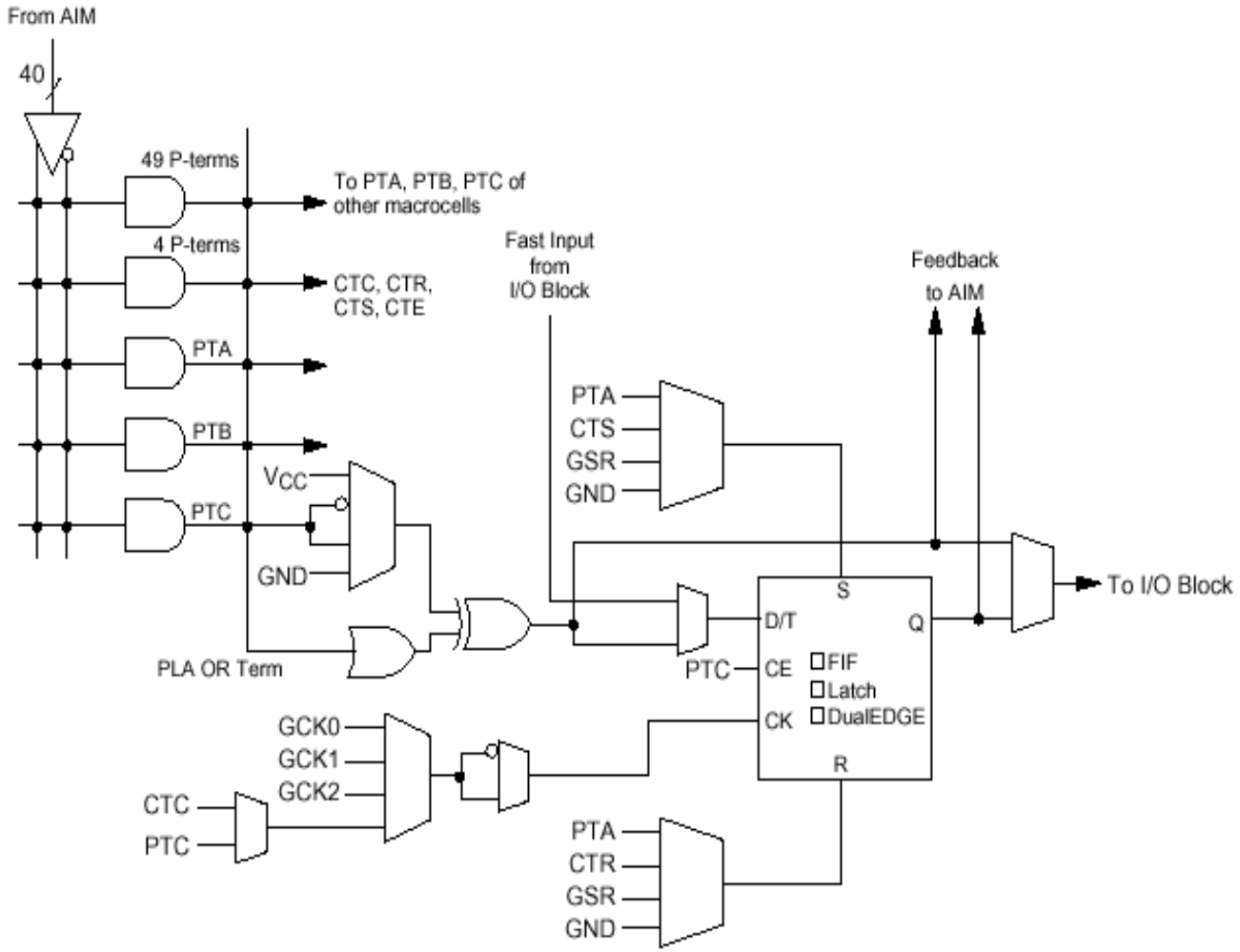
Functional Block



- Each Functional block has 16 macro cells.
- 56 product term PLA.

2024 © Centre for Development of Advanced Computing

Macrocell



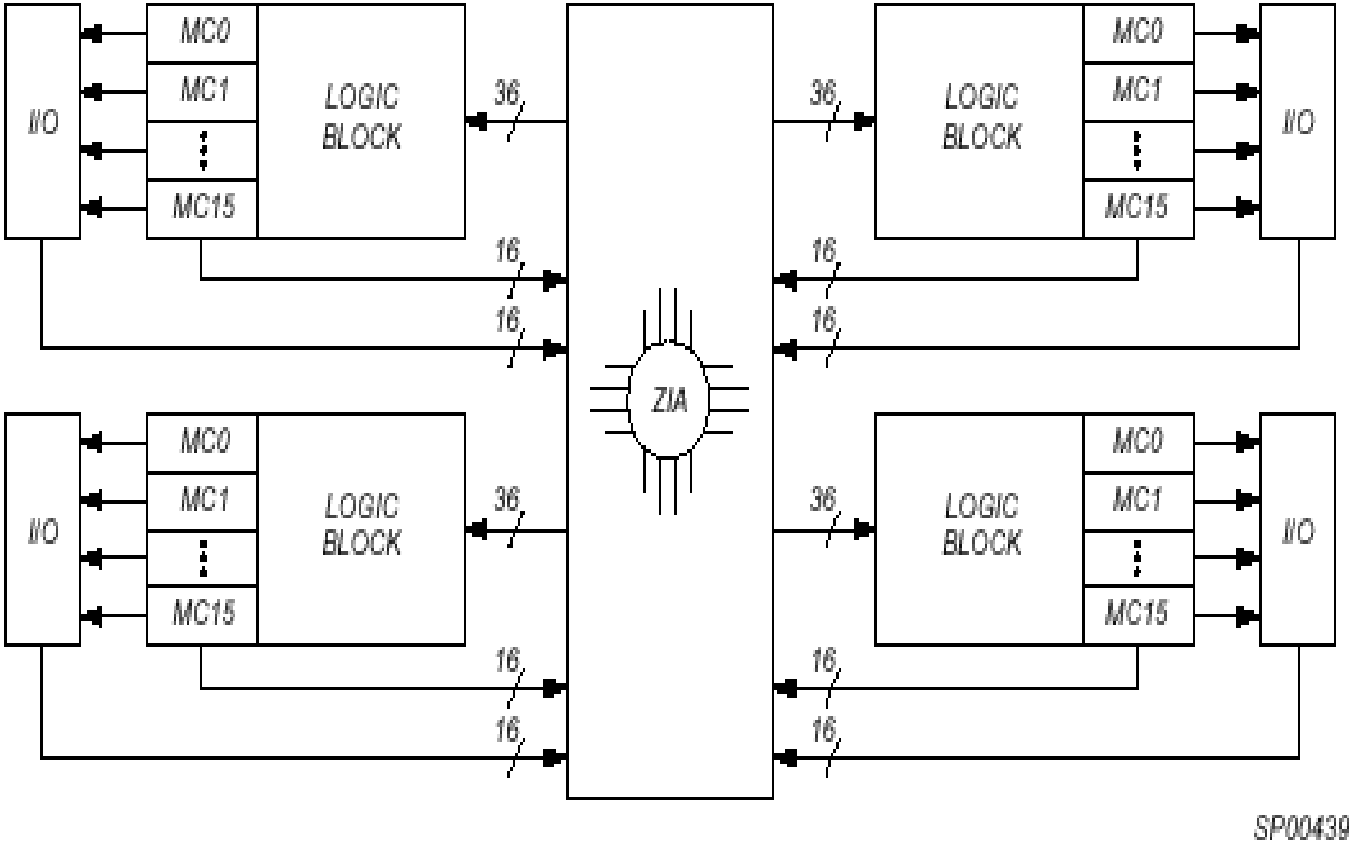
2024 © Centre for Development of Advanced Computing

Macrocell

- Users can develop SOP logic expressions that comprise up to 40 inputs and span 56 product terms within a single function.
- The macrocell can further combine the SOP expression into an XOR gate with another single p-term expression.
- At the high level, it is seen that the product terms (p-terms) the logic function can be pure combinatorial or registered, with the storage element operating selectable as a D or T flip-flop, or transparent latch.

2024 © Centre for Development of Advanced Computing

Alternate Architecture



2024 © Centre for Development of Advanced Computing

FPGA Basics

- Field-Programmable Gate Arrays (FPGAs) provide the benefits of custom CMOS VLSI, while avoiding the initial cost, time delay, and inherent risk of a conventional masked gate array.
- FPGA consists of basically an array of logic blocks and routing channels.
- The FPGA has three major configurable elements: input/output blocks (IOB), configurable logic blocks (CLBs), and interconnects.

2024 © Centre for Development of Advanced Computing

FPGAs and CPLDs

- FPGAs have much smaller basic building block (e.g. 4-input LUT as compared to a 56 p-term PLA)
- FPGAs use more flexible and faster interconnects. Also number of interconnect tends to be much larger than a comparable CPLD
- FPGA structure is inherently scalable to realize very large logic matrices
- FPGAs tend to incorporate several features suitable for designing digital systems (more than just logic e.g. memories, clock buffers, DLLs etc)
- FPGAs offer true in-field programmability (fast, unlimited times)
- Fastest FPGAs, tend to be SLOWer than fastest CPLDs for small, fast logic design requirements
- FPGA design cycles tend to be more complex

2024 © Centre for Development of Advanced Computing

FPGA Basics

- The IOBs provide the interface between the package pins and internal signal lines.
- The CLBs provide the functional elements for constructing user's logic
- The programmable interconnect resources provide routing paths to connect the inputs and outputs of the CLBs and IOBs onto the appropriate networks.
- In terms of logic granularity, FPGAs offer finer granularity than CPLDs (But much courser granularity than ASIC libraries)

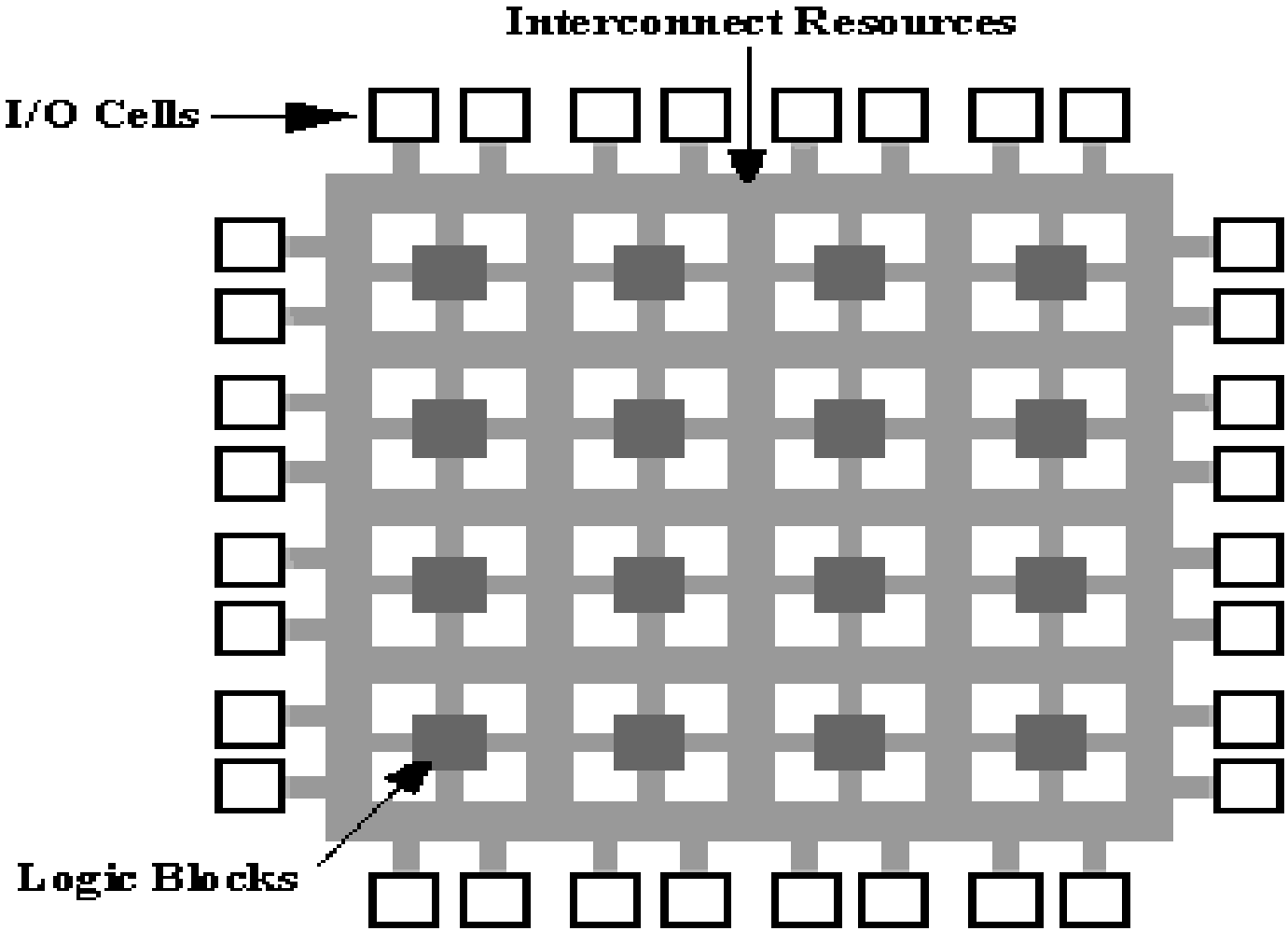
2024 © Centre for Development of Advanced Computing

Granularity of a PLD

- Granularity can be thought as a measure of how complex a basic building block in PLD is
  - A course grain structure will be less efficient towards silicon area utilization (e.g. CPLD macrocell) but will be easy to implement and provide predictable timing for a fixed set of routing resources
  - A fine grain structure will be more efficient towards area utilization, however it will require more local routing resources. Due to variable routing nature, predictable timing cannot be guaranteed for complex designs. (Note that for ASICs routing delays are much lesser due to fixed, metal interconnects) Also, these structures will demand more die resources if they are made truly re-configurable
  - Granularity therefore is always a tradeoff between **area utilization** Vs **speed** and **re-configurability** at smallest macrocell

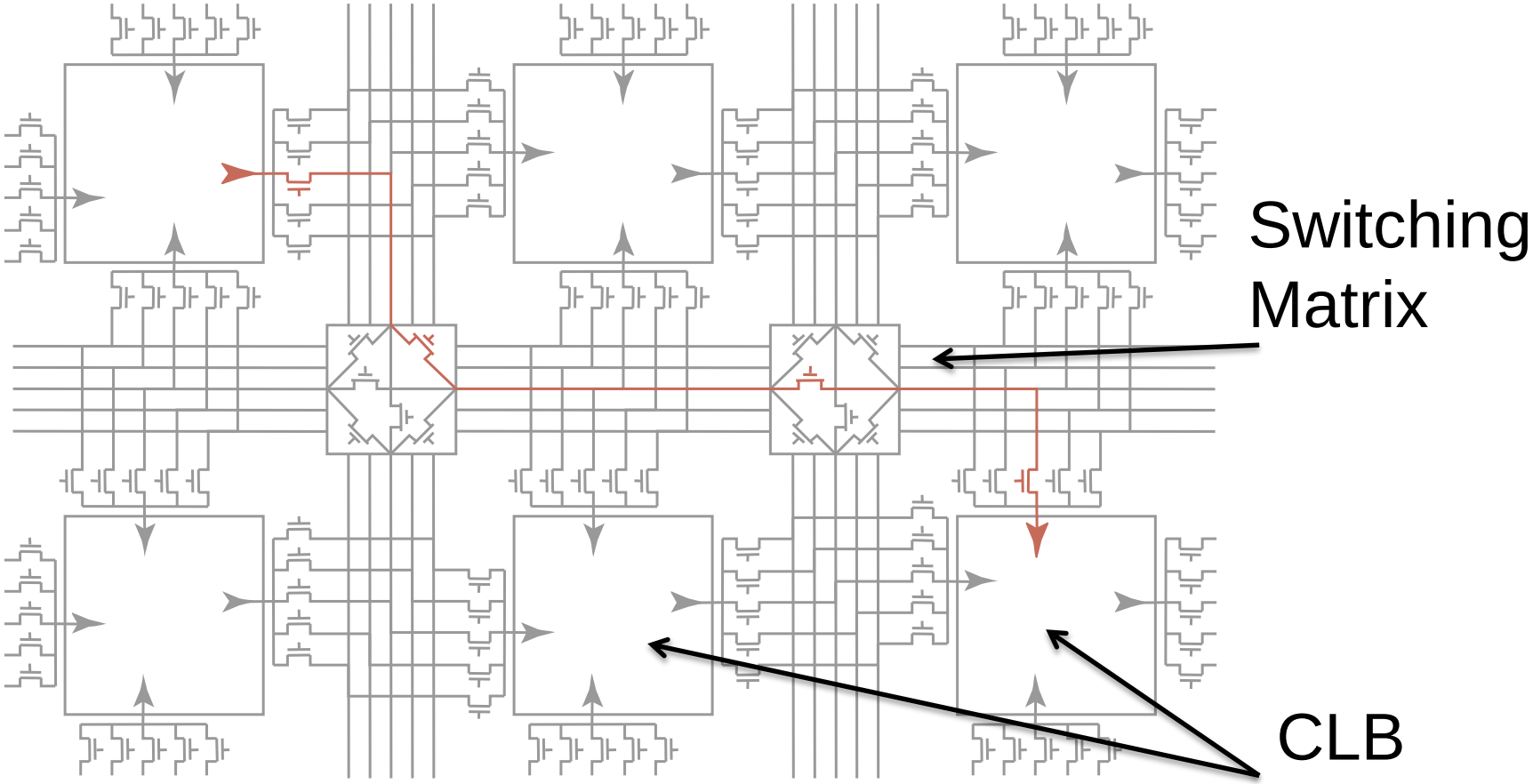
2024 © Centre for Development of Advanced Computing

FPGA Architecture



2024 © Centre for Development of Advanced Computing

CLB Interconnects



2024 © Centre for Development of Advanced Computing

FPGA Basics

- Figure depicts a FPGA with a two-dimensional array of logic blocks that can be interconnected by interconnect wires.
- All internal connections are composed of metal segments with programmable switching points to implement the desired routing.
- An abundance of different routing resources is provided to achieve efficient automated routing.

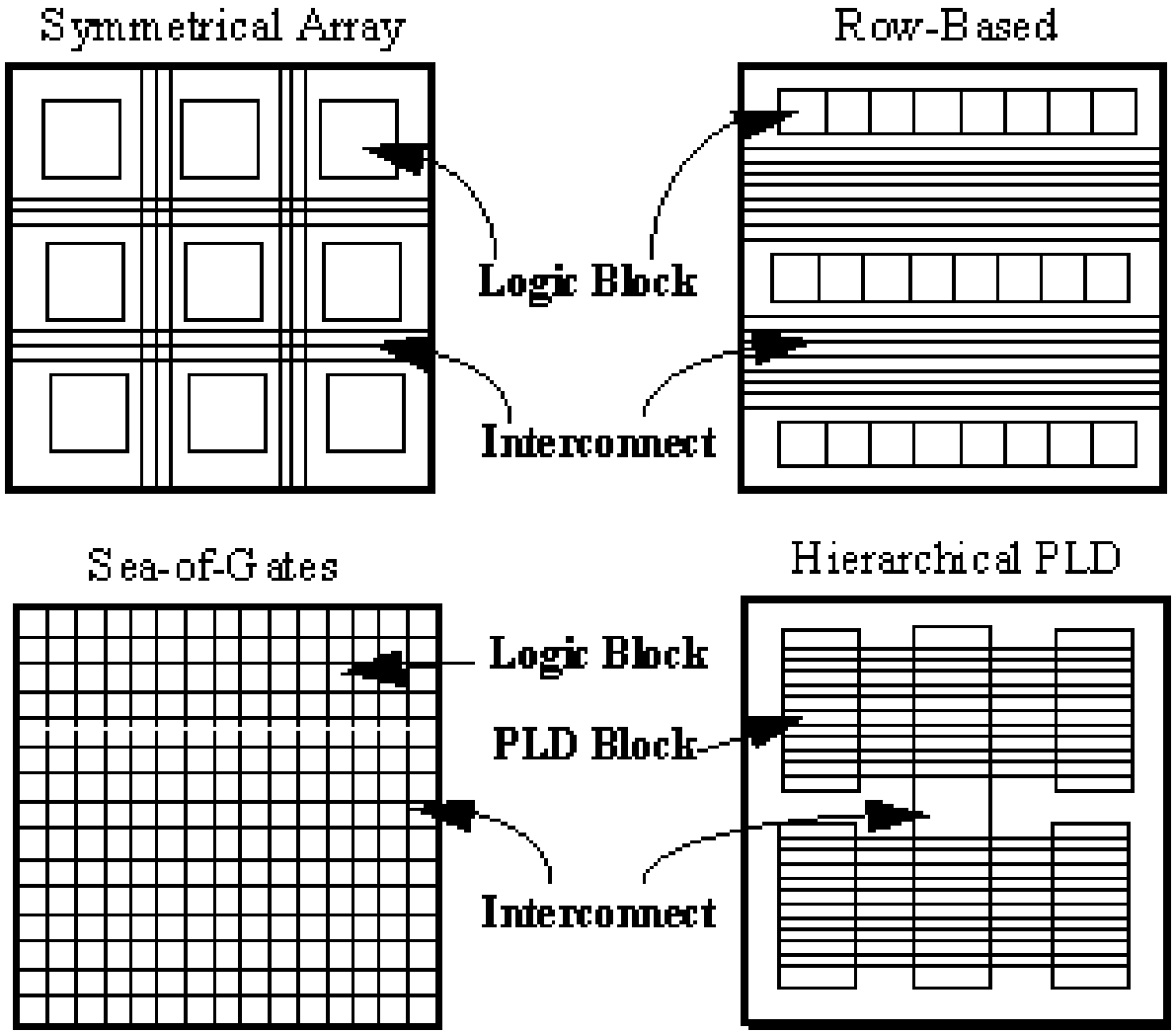
2024 © Centre for Development of Advanced Computing

FPGA Basics

- FPGAs are customized by loading configuration data into the internal memory cells.
- Based on placement of logic blocks FPGAs are classified into 4 categories viz. symmetrical array, row-based, hierarchical PLD, and sea-of-gates

2024 © Centre for Development of Advanced Computing

FPGA Basics



2024 © Centre for Development of Advanced Computing

FPGA configuration

- FPGA configuration technologies
  - Static RAM based, Antifuse based, FLASH based.
- Re-programmability, Speed, Density, Cost depends on the technology used in the FPGA.

2024 © Centre for Development of Advanced Computing

Static RAM Technology

- In the Static RAM FPGA, programmable connections are made using transistors, transmission gates, or multiplexers that are controlled by SRAM cells.
- The main feature of this technology is that it allows fast in-circuit reconfiguration.
- The FPGA can either actively read its configuration data out of external serial, or byte-parallel PROM or the configuration data can be written into the FPGA.

2024 © Centre for Development of Advanced Computing

SRAM Reconfiguration

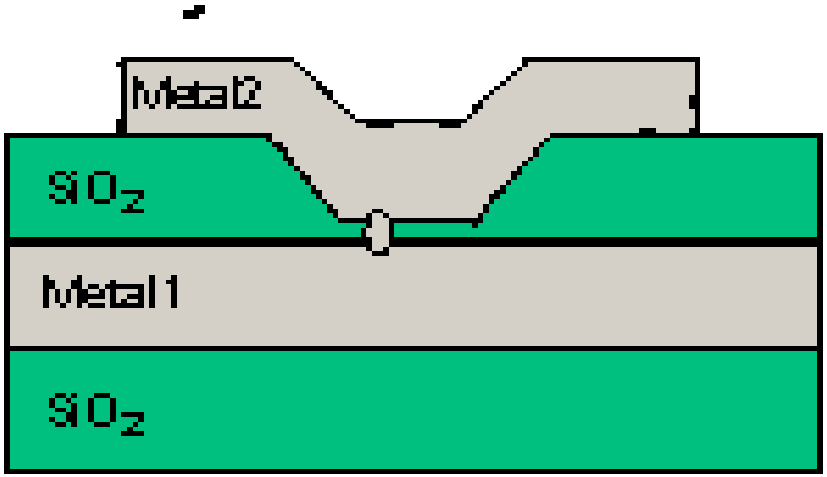
- Advantages
  - Design updates are easy, can be made to product already on field.
  - Reconfiguration is fast
  - Selective reconfiguration is possible
  - Simplifies hardware design and debugging. Reduces time to market.
  - IP protection using AES encryption

- Disadvantages
  - The programmability causes reduction in speed.
  - Less efficient utilization of Silicon; has cost implications

2024 © Centre for Development of Advanced Computing

Anti-Fuse Technology

- An anti-fuse resides in a high-impedance state; and can be programmed into low impedance or "fused" state.



metal-metal antifuse

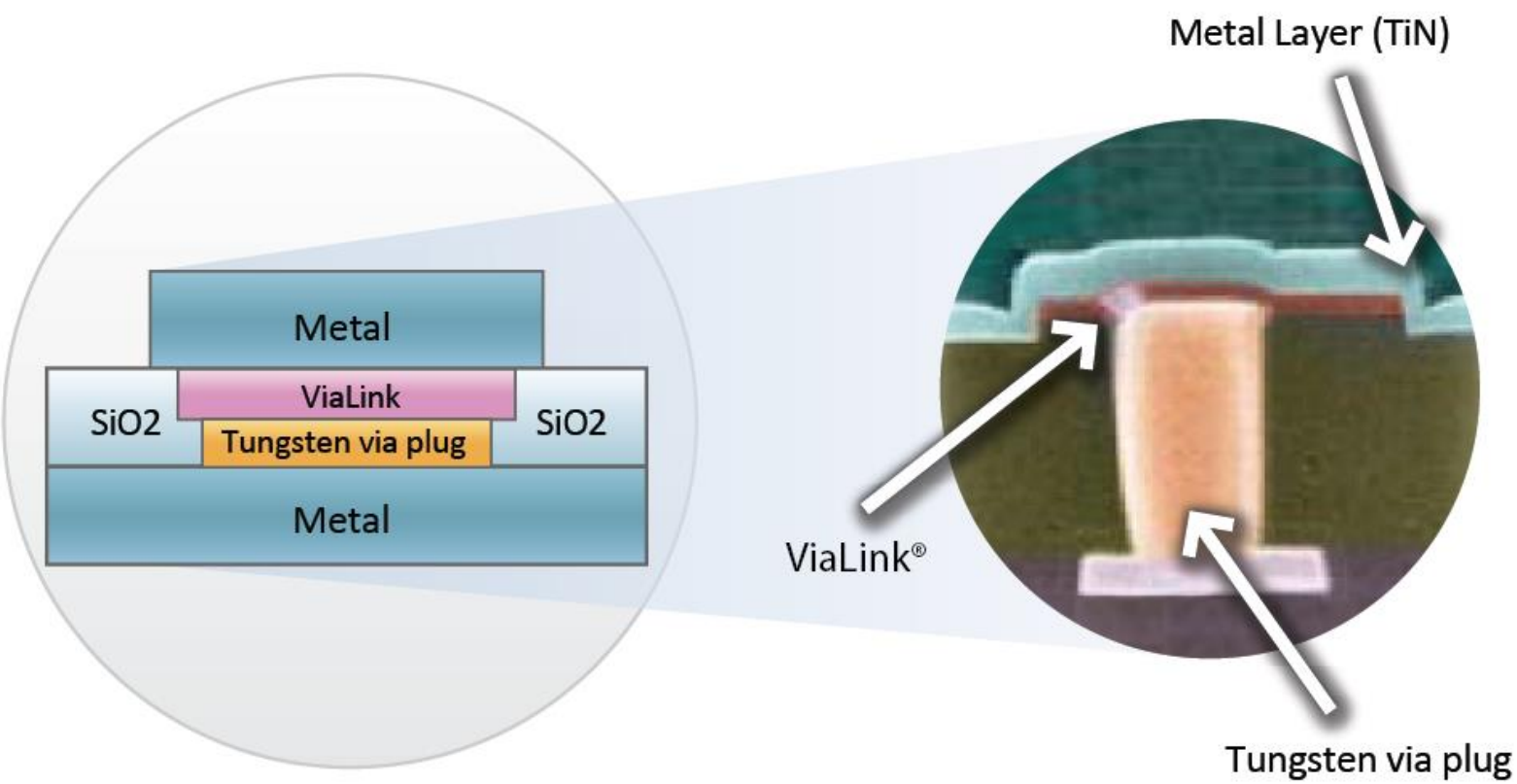
2024 © Centre for Development of Advanced Computing

Anti-Fuse Technology

- The link is created by melting the thin isolating dielectric between two metal layers.
- less expensive than the RAM technology, however, device becomes OTP
- Advantages
  - Faster than a programmable switch
  - Secure programming technology prevents reverse engineering and design theft
  - Retains configuration indefinitely

2024 © Centre for Development of Advanced Computing

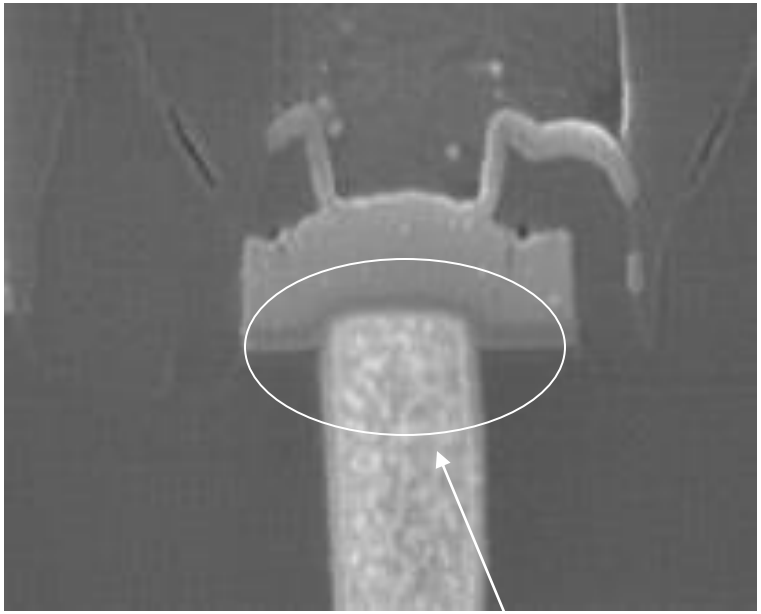
Anti-Fuse Technology



Curtsey: QuickLogic

2024 © Centre for Development of Advanced Computing

Antifuse snapshots



An unprogrammed Antifuse Cross Section

A programmed Antifuse Cross Section

2024 © Centre for Development of Advanced Computing

Interconnect

- Interconnect connects signal from one CLB to another CLB or to the I/O.
- Interconnects are distinguished by the relative length of their segments: single-length lines, double-length lines and Longlines.
- In addition global buffers drive fast, low-skew nets most often used for clocks or global control signals.
- “Active” interconnects dramatically reduce propagation delays of nets in newer devices
  - Individual R-C delays created by each programmable interconnect point (PiP) adds to the total delay of the net

2024 © Centre for Development of Advanced Computing

FPGA vendors

- Xilinx (Now with AMD since October 2020) (SRAM)
  - Virtex-II, Virtex-II Pro, Pro-X, V4 (SoC), V5, V6, 7-Series, Ultrascale, Ultrascale+, Versal
  - Spartan-3 to Spartan-7 (Low cost)
  - CoolRunner-II (CPLD)
- Altera (Now with Intel since June 2015)
  - Agilex F, I, M, D, Direct RF series and Stratix-V, 10
  - Arria-V and 10 (Low cost transceiver based)
  - Cyclone-IV, V and 10 (Low cost)
  - MAX-V and 10 (Low cost CPLD)
- Actel and Atmel (Now Microsemi)
  - ProASIC3 and Fusion mixed signal (Flash)
  - PolarFire
  - IGLOO/2 (Power Optimized)
- Lattice, QuickLogic (Antifuse), Achronix ...

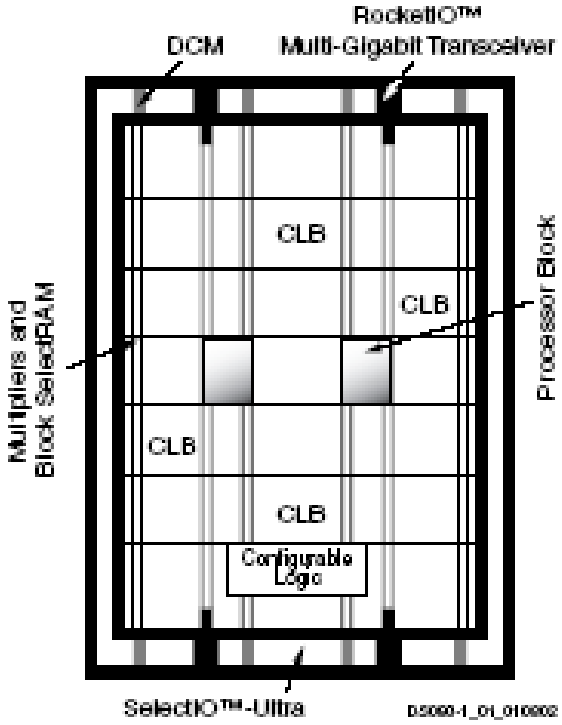
2024 © Centre for Development of Advanced Computing

Virtex-II, Pro and Pro-X devices

- 0.15 um, 8-layer metal process: 1.5V core and 3.3V aux supply
- 400+MHz system clock
- Support for 19 single ended and six differential standards
- Pro - Two Significant additions over Virtex-II:
  - On chip PowerPC processor
  - On chip serial transceiver and coding resource (RocketIO)
  - Improvements in speed grade
- Pro-X can scale upto 6 Gbps serial speeds

2024 © Centre for Development of Advanced Computing

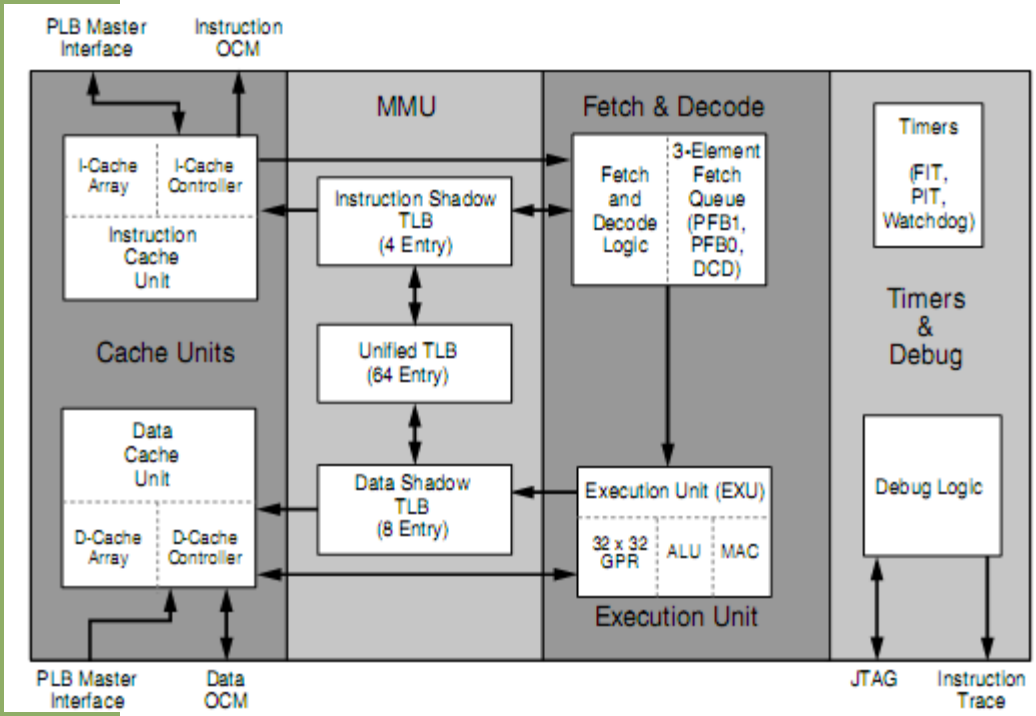
V2Pro



Device	RocketIO Transceiver Blocks	PowerPC Processor Blocks	Logic Cells(1)	CLB (1 = 4 slices = max 128 bits)		18 X 18 Bit Multiplier Blocks	Block SelectRAM+		DCMs	Maximum User I/O Pads
				Slices	Max Distr RAM (Kb)		18 Kb Blocks	Max Block RAM (Kb)		
XC2VP2	4	0	3,168	1,408	44	12	12	216	4	204
XC2VP4	4	1	6,768	3,008	94	28	28	504	4	348
XC2VP7	8	1	11,088	4,928	154	44	44	792	4	396
XC2VP20	8	2	20,880	9,280	290	88	88	1,584	8	564
XC2VP30	8	2	30,816	13,696	428	136	136	2,448	8	644
XC2VP40	0(2) or 12	2	43,632	19,392	606	192	192	3,456	8	804
XC2VP50	0(2) or 16	2	53,136	23,616	738	232	232	4,176	8	852
XC2VP70	16 or 20	2	74,448	33,088	1,034	328	328	5,904	8	996
XC2VP100	0(2) or 20	2	99,216	44,096	1,378	444	444	7,992	12	1,164
XC2VP125	0(2), 20, or 24	4	125,136	55,616	1,738	556	556	10,008	12	1,200

2024 © Centre for Development of Advanced Computing

PPC



- Available as hard macro (does not take FPGA logic)
- 300MHz operation
- On chip 16KB Icache and Dcache
- 5-stage pipeline for fetch/decode and execution
- MMU for virtual memory mgmt
- Connects to external logic over PLB (high performance) and OPB (low performance) busses. OPB access is thru a bus bridge
- OCM controllers allow connectivity to external memories
- Number of IPs to work with PPC are available from Xilinx and third party
- Debug interface is available over JTAG chain

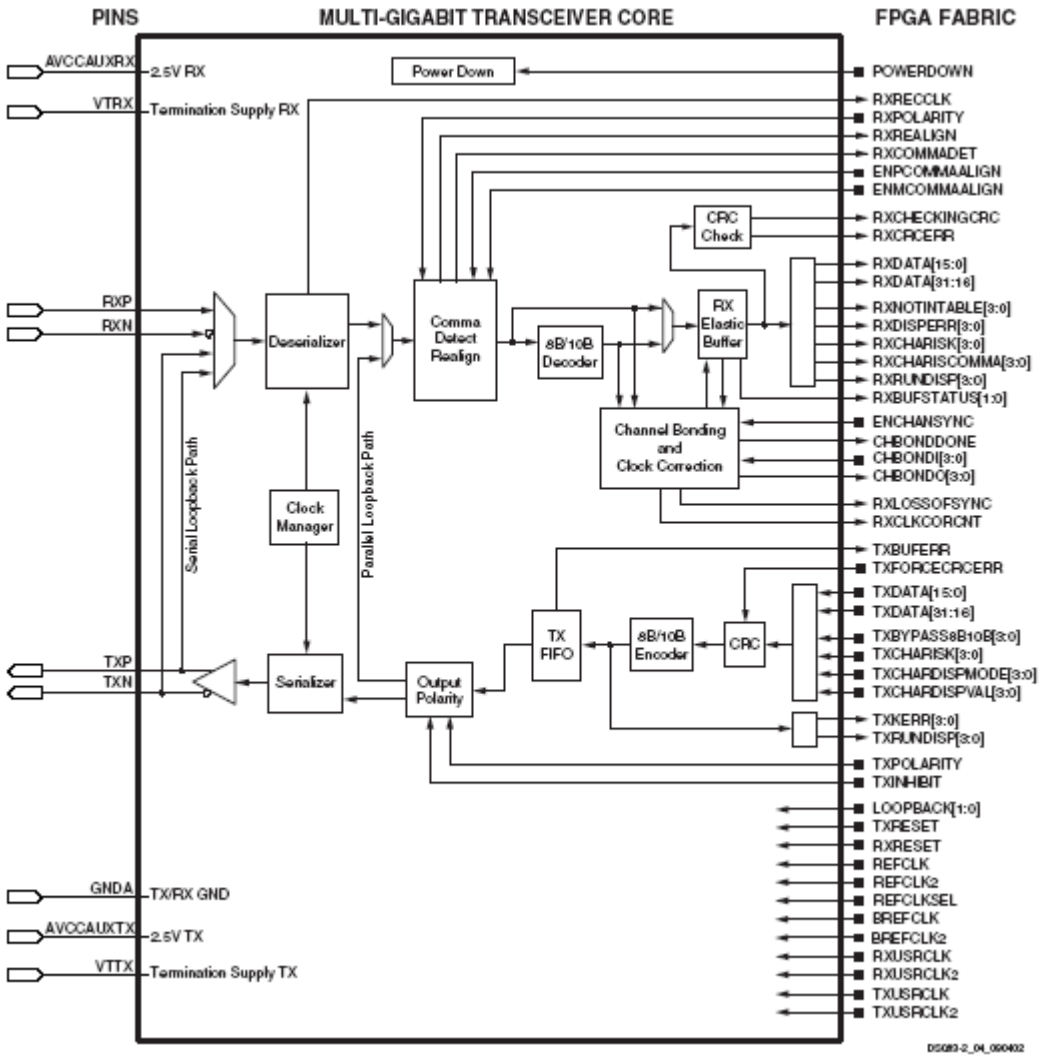
2024 © Centre for Development of Advanced Computing

Serial stds

- Emerging new high speed interconnect standards use multi-gigabit serial paths rather than parallel busses.
  - Gb ethernet (1Gb/10Gb)
  - Fiber Channel
  - Infiniband
  - PCI Express
- In addition serial interconnects are increasingly used as chip to chip interconnects (direct or across backplanes)
- RocketIO allows system designer to integrate Physical coding layer (PCS) and link layer within the FPGA

2024 © Centre for Development of Advanced Computing

RocketIO



Protocol	Channels (Lanes)	I/O Baud Rate (Gb/s)	Reference Clock Rate (MHz)
Fibre Channel	1	1.06	53
		2.12	106
		3.1875(1)	159.375
Gigabit Ethernet	1	1.25	62.5
10 Gbit Ethernet	4	3.125	156.25
Infiniband	1, 4, 12	2.5	125
Aurora	1, 2, 3, 4, ...	0.840 – 3.125	42.00 – 156.25
Custom Protocol	1, 2, 3, 4, ...	up to 3.125	up to 156.25

2024 © Centre for Development of Advanced Computing

Virtex-4 family

- 90nm, Vcore=1.2V
- LX, SX and FX series for logic, DSP and embedded processor applications
- Numerous enhancements over Virtex-2/Pro in many architectural resources, faster speed grades
  - Some of the enhancements are:
    - Faster PPC with optional APU
    - Programmable, Predictable delay on every I/P
    - Phase-matched clock divider (PMCD) along with DCM
    - On chip FIFOs using blockRAMs
    - On chip GMAC

2024 © Centre for Development of Advanced Computing

Virtex-5 family

- 65nm, Vcore=1.0V, triple oxide process (Toshiba/UMC)
- Available in following flavors
  - LX
  - LXT: Built in serial transceivers, Ethernet, PCI-e hard macros
    - Transceivers consume less power wrt V-4
  - SXT: DSP applications
- Changes to basic LUT: 5 I/P LUT
- Changes to CLB hierarchical routing
- BlockRAM capacity doubled
- ODELAY added alongside IDELAY for IOBs
- Analog PLL in DCM
- On chip monitoring

2024 © Centre for Development of Advanced Computing

Virtex-6 family

- 40nm, Vcore=1.0V (0.9 V available), Copper CMOS process
- Available in following flavors (with serial transceivers)
  - LXT: Ethernet(10/100/1000 Mb/s), PCI-e (Gen 2.0) hard macros
  - SXT: DSP applications (25x18 multiplier+adder+accumulator)
  - HXT: Highesh bandwidth serial transceivers
- Changes to basic LUT: 6 I/P LUT + Dual LUT5 (5 input LUT)
  - 64-bit (or two 32-bit) distributed LUT RAM
  - SRL32/dual SRL16 with registered o/p
- Mixed Mode Clock Managers (MMCM)
- Fine grained IO banking (up to 31 banks)
- AES-encrypted bitstream with 256-bit key

2024 © Centre for Development of Advanced Computing

Virtex-6 Architecture

- Widely used, mature architecture
  - 3000 → 4000 → 4000E → Virtex → Virtex-II→ Virtex-4 → Virtex-5 → Virtex-6 → Virtex-7
- Largest sales and market share for Virtex family
- Well defined roadmap (V2Pro, V2Pro-X, V4, V5, V6, 7-series, Ultrascale, Ultrascale+ ..)

2024 © Centre for Development of Advanced Computing

7-series Architecture

- 28nm, Vcore = 1.0V (0.9V low power)
- Three variants:
  - Artix-7 FPGAs: Lowest power and cost
  - Kintex-7 FPGAs: Industry's best price-performance
  - Virtex-7 FPGAs: Industry's highest system performance and capacity
  - EasyPath™-7 FPGAs: Fast, simple and risk-free solution for cost- reducing Virtex-7 FPGA designs
- 65% lower static power consumption
- 25 to 30% lower dynamic power consumption
- 30% lower I/O dynamic power consumption

2024 © Centre for Development of Advanced Computing

7-series Architecture

- Stacked Silicon Interconnect (SSI)
  - Multiple Super Logic Regions (SLRs) combined on a passive silicon interposer layer
  - More than 10000 routing lines across SLRs
  - Thousands of Super Long Lines (SLLs)
  - Virtex-7 HT devices support SSI

2024 © Centre for Development of Advanced Computing

Xilinx Family of devices

Current offering from Xilinx

45nm	28nm	20nm	16nm
SPARTAN^{6}	VIRTEX^{7}	VIRTEX$_{\text{UltraSCALE}}$	VIRTEX$_{\text{UltraSCALE+}}$
	KINTEX^{7}	KINTEX$_{\text{UltraSCALE}}$	KINTEX$_{\text{UltraSCALE+}}$
	ARTIX^{7}		
	SPARTAN^{7}		

2024 © Centre for Development of Advanced Computing

7-Series Family

MAXIMUM CAPABILITY	ARTIX-7 FAMILY	KINTEX-7 FAMILY	VIRTEX-7 FAMILY
Logic Cells	352K	407K	1,955K
Block RAM	12Mb	29Mb	65Mb
DSP Slices	700	1,540	3,960
Peak DSP Performance (symmetric FIR)	504 GMACS	1,848 GMACS	4,752 GMACS
Transceiver Count	4	16	80
Peak Transceiver Speed	3.75Gbps	10.3125Gbps	13.1Gbps+
Peak Serial Bandwidth (full duplex)	30Gbps	330Gbps	1,886Gbps
PCI Express® Interface	Gen1 x4	Gen2 x8	Gen3 x8
Memory Interface	800Mbps	2,133Mbps	2,133Mbps
I/O Pins	450	500	1,200
I/O Voltage	1.2V, 1.5V, 1.8V, 2.5V, 3.3V	1.2V, 1.35V, 1.5V, 1.8V, 2.5V, 3.3V	1.2V, 1.35V, 1.5V, 1.8V, 2.5V, 3.3V
Packaging Options	Low cost wire bond	Low cost lidless flip chip and High performance flip chip	Highest performance flip chip

2024 © Centre for Development of Advanced Computing

Specifications

- 28 nm, high-k metal gate (HKMG): 1.0V core
- Capability
  - 600MHz+ system clock (may not be achievable for complex designs)
  - 102K .. 1955K Logic cells
  - Upto 1200+ user I/Os
  - 1.4 Gbps data per I/O pin (using LVDS)
- SelectIO technology
  - Support for 19+ single ended and 10+ differential standards
  - Programmable slew rate and drive per pin
  - Digitally controlled impedance (DCI) for on-chip terminations
  - Support for major I/O busses: PCI, PCI-X
  - Support for DRAM, DDR-SDRAM, QDR-SRAM interfaces

2024 © Centre for Development of Advanced Computing

Specifications

- Logic
  - M X N matrix of ‘slices’ (device specific)
  - Slices supporting lookup tables, shift registers and registers/latches
  - Wide multiplexer support
  - Cascaded carry chains
- Clock management
  - Upto 24 clock management tiles each with 2 PLL based MMCMs
- Memory
  - SelectRAM block memory modules
  - Distributed memory supported
  - Dedicated on-chip FIFO logic

2024 © Centre for Development of Advanced Computing

Specifications

- Other resources
  - Dedicated multiplier blocks
  - Carry chain logic
- Interconnect
  - Large number of hierarchical routing resources
  - Active interconnect technology for predictable routing delays independent of fanout
- Configuration based on SRAM cells
  - AES for device IP protection
  - JTAG based configuration and readback
  - Support for master/slave modes, cascading of devices
- Wide packaging options

2024 © Centre for Development of Advanced Computing

UltraSCALE and + Family

	Kintex UltraScale FPGA	Kintex UltraScale+ FPGA	Virtex UltraScale FPGA	Virtex UltraScale+ FPGA	Zynq UltraScale+ MPSoC	Zynq UltraScale+ RFSoC
MPSoC Processing System					✓	✓
RF-ADC/DAC						✓
SD-FEC						✓
System Logic Cells (K)	318–1,451	356–1,143	783–5,541	862–3,780	103–1,143	678–930
Block Memory (Mb)	12.7–75.9	12.7–34.6	44.3–132.9	23.6–94.5	4.5–34.6	27.8–38.0
UltraRAM (Mb)		0–36		90–360	0–36	13.5–22.5
HBM DRAM (GB)				0–8		
DSP (Slices)	768–5,520	1,368–3,528	600–2,880	2,280–12,288	240–3,528	3,145–4,272
DSP Performance (GMAC/s)	8,180	6,287	4,268	21,897	6,287	7,613
Transceivers	12–64	16–76	36–120	32–128	0–72	8–16
Max. Transceiver Speed (Gb/s)	16.3	32.75	30.5	58.0	32.75	32.75
Max. Serial Bandwidth (full duplex) (Gb/s)	2,086	3,268	5,616	8,384	3,268	1,048
Memory Interface Performance (Mb/s)	2,400	2,666	2,400	2,666	2,666	2,666
I/O Pins	312–832	280–668	338–1,456	208–832	82–668	280–408

2024 © Centre for Development of Advanced Computing

Specifications

- 20nm and 16nm, Planar through FinFET technology
- Next generation routing, ASIC-like clocking, and enhanced logic blocks for a target of 90% utilization
- Capability
  - multi-hundred gigabit-per-second levels of system performance)
- GTH, GTY, GTM transceiver
  - 12.5Gbps (GTH) and 16.3Gbps (GTY) in Kintex family
  - 12.5Gbps (GTH) and 30.5Gbps (GTY) in Virtex family
  - 32.75Gbps (GTY), 58.0Gbps (GTM) in UltraSCALE+ family

2024 © Centre for Development of Advanced Computing

Specifications

- Memory
  - SelectRAM block memory modules
  - Distributed memory supported
  - Dedicated on-chip FIFO logic
  - DDR4 support of up to 2,666 Mb/s
  - UltraRAM (only in UltraSCALE+) provides massive on-chip memory for SRAM device integration
  - HBM DRAM (only in UltraSCALE+) provides up to 8 GB
- SLR (mostly in UltraSCALE+) provides easier and predictable clocking performance
- Static- and dynamic-power gating across a wide range of functional elements

2024 © Centre for Development of Advanced Computing

Versal

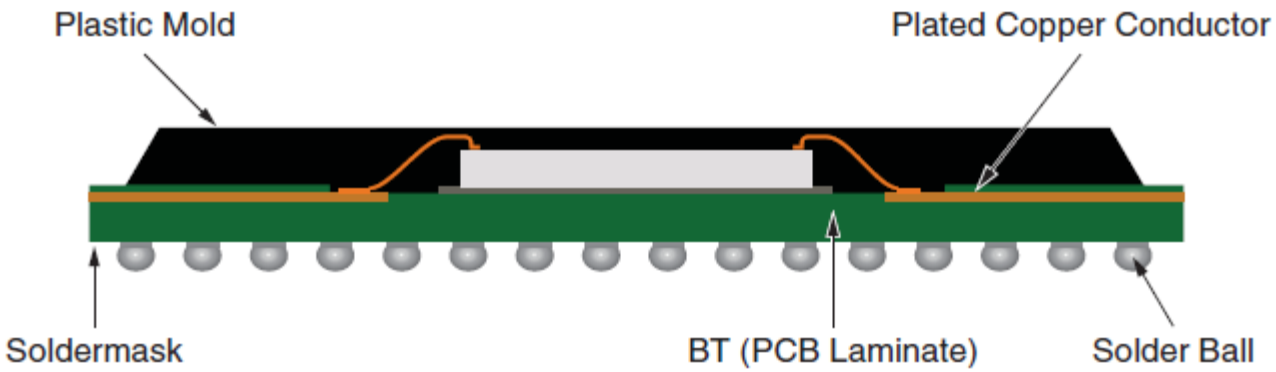
- 7nm, ACAP (Adaptive Compute Acceleration Platform)
- AI core and Prime series
- On chip NoC, AI cores
- Various memory types: Distributed LUTRAM, BRAM, Accelerator RAM, HBM, UltraRAM, DDR4 support
- Upto 112 Gbps transceiver, 58 Gbps mainstream device support
- PCIe x16, Gen4
- Multirate EMAC core
- Dual core ARM Cortex A72 (upto 1.7GHz) and dual core Arm Cortex R5 (Upto 750 MHz) Real Time processors

2024 © Centre for Development of Advanced Computing

Packaging

- 1.0mm BGAs, packaging either wire-bond (I/Os are limited) or Flip-chip (large number of I/Os supported for same die size)

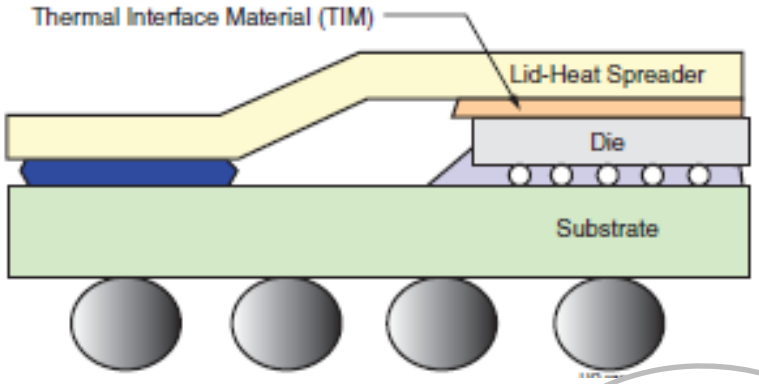
The substrate is made of a multi-layer BT (bismaleimide triazene) epoxy-based material

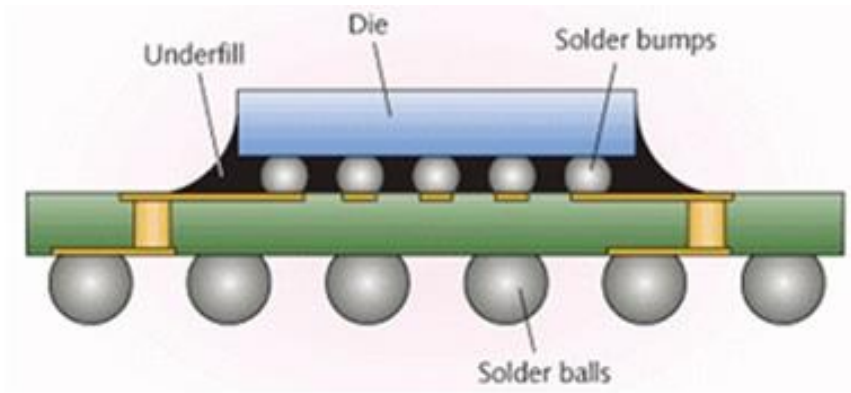


Cavity Up BGA package

2024 © Centre for Development of Advanced Computing

Flip Chip BGA package





2024 © Centre for Development of Advanced Computing

Architectural Components

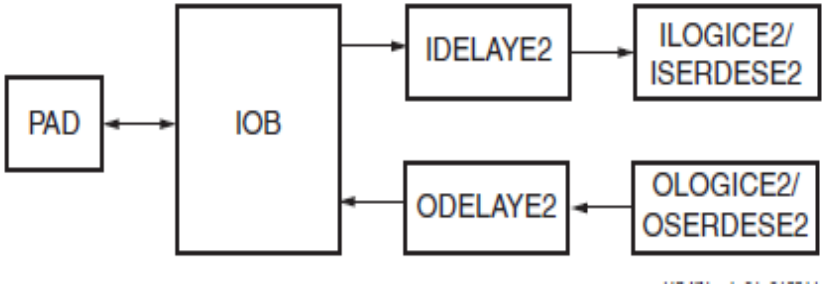
- Input/Output Blocks (IOBs)
  - SelectIO, Digitally controlled impedance (DCI)
- Configurable Logic Blocks (CLBs)
- Block Memories (SelectRAM) and dedicated multipliers
- Clocking resources
- Routing resources (local and global)
- Embedded logic resources
- Special DSP resources
- System Monitoring
- Configuration
- Testability/Access

2024 © Centre for Development of Advanced Computing

Input Output Block (selectIO)

2024 © Centre for Development of Advanced Computing

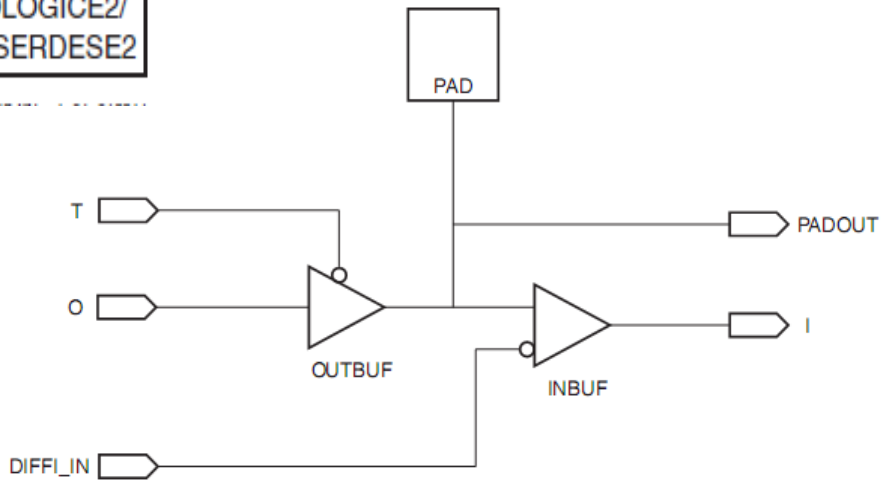
Basic Input/Output Block

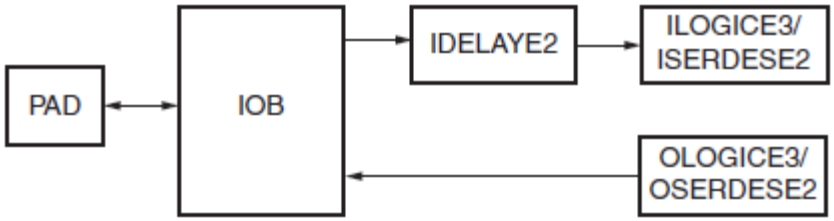


HP Bank

Ultrascale and + devices

Each tile contains two IOBs





HR Bank

Ultrascale devices only

2024 © Centre for Development of Advanced Computing

Supported features in I/O

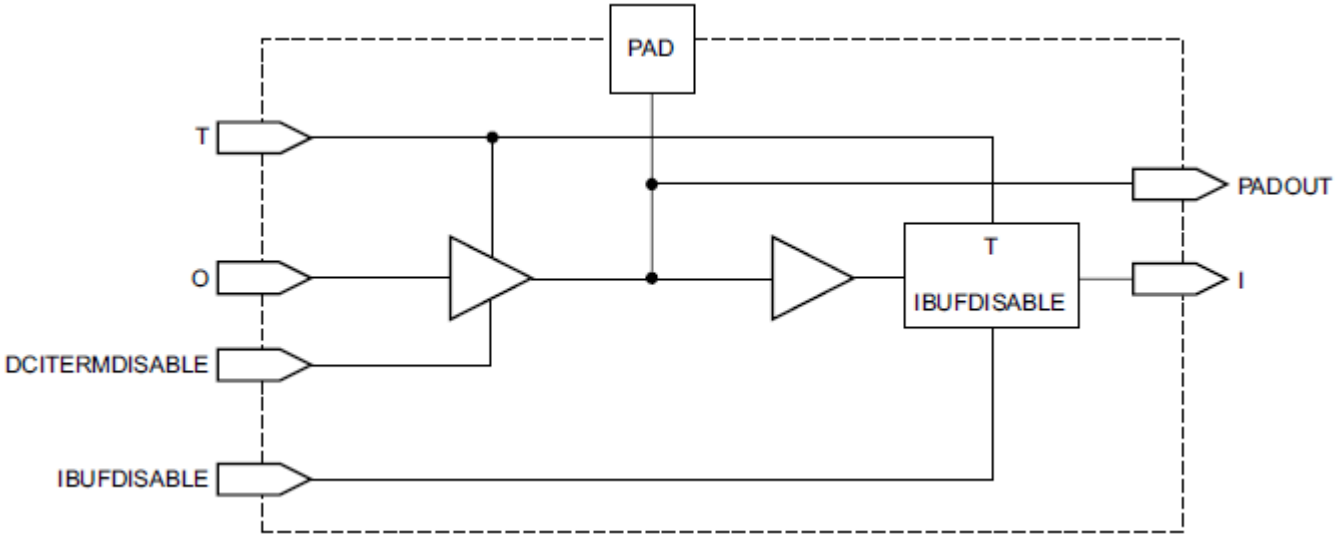
Feature	HP I/O Banks	HR I/O Banks
ODELAY	Supported	Supported
IDELAYCTRL	Supported	Supported
ISERDES	Supported	Supported
OSERDES	Supported	Supported
Transmitter pre-emphasis	Supported	Supported[3]
Receiver equalization	Supported	Supported
Receiver offset control	Supported	Not supported
Receiver V_{REF} scan	Supported	Not supported
MIPI D-PHY	Supported in Virtex UltraScale+ and Kintex UltraScale+ devices and Zynq UltraScale+ MPSoCs	Not supported

* High Density (HD) selectIO is only available in Zynq UltraScale + devices

2024 © Centre for Development of Advanced Computing

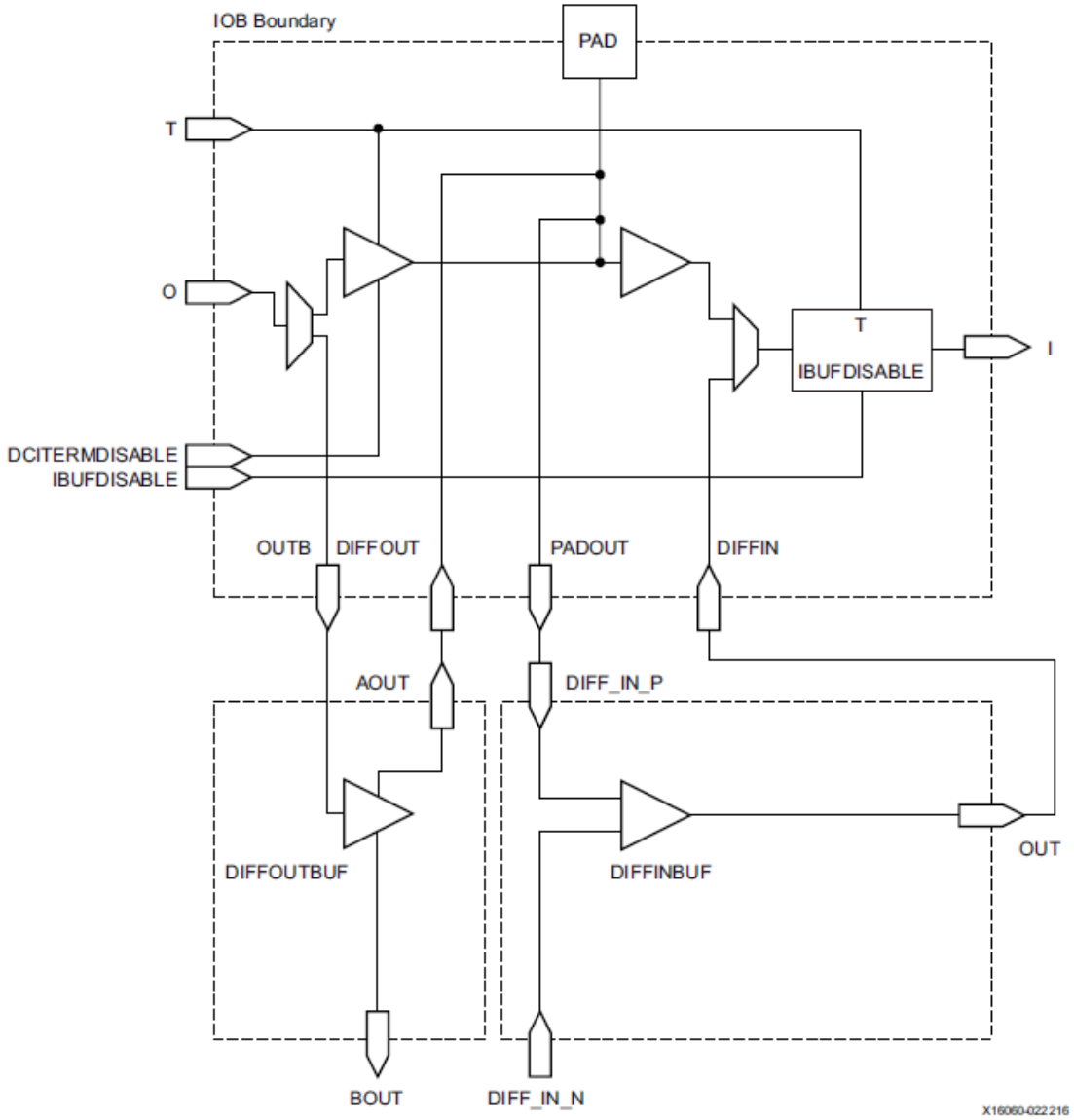
HP IOB

Single Ended (VRP pin used as general IO)



2024 © Centre for Development of Advanced Computing

HP IOB

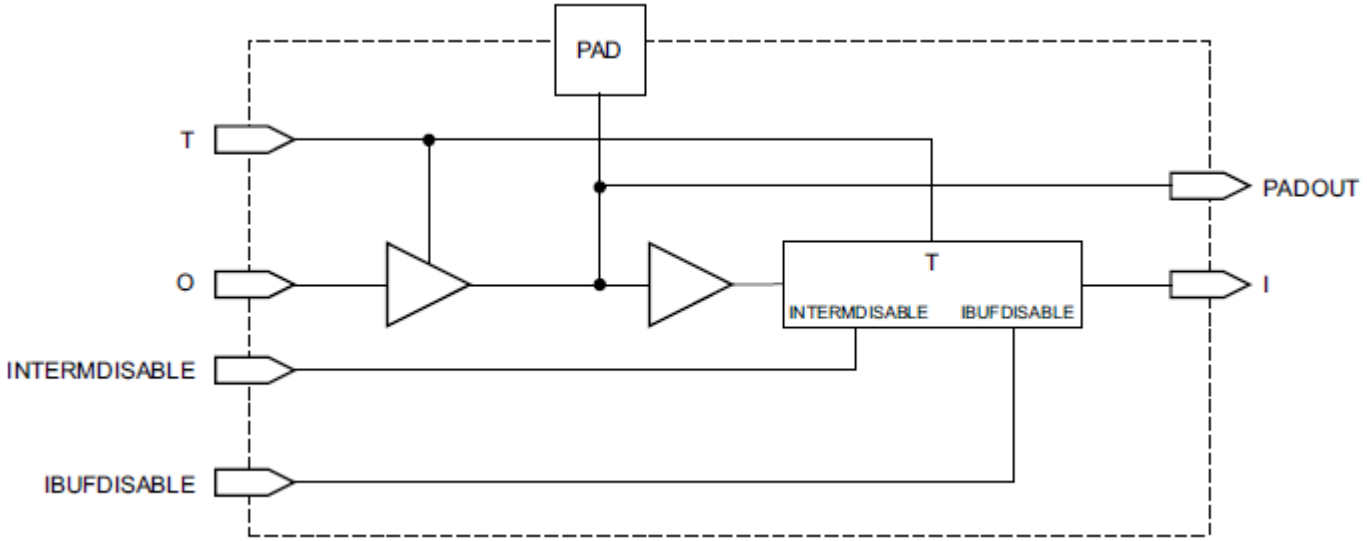


Standard

2024 © Centre for Development of Advanced Computing

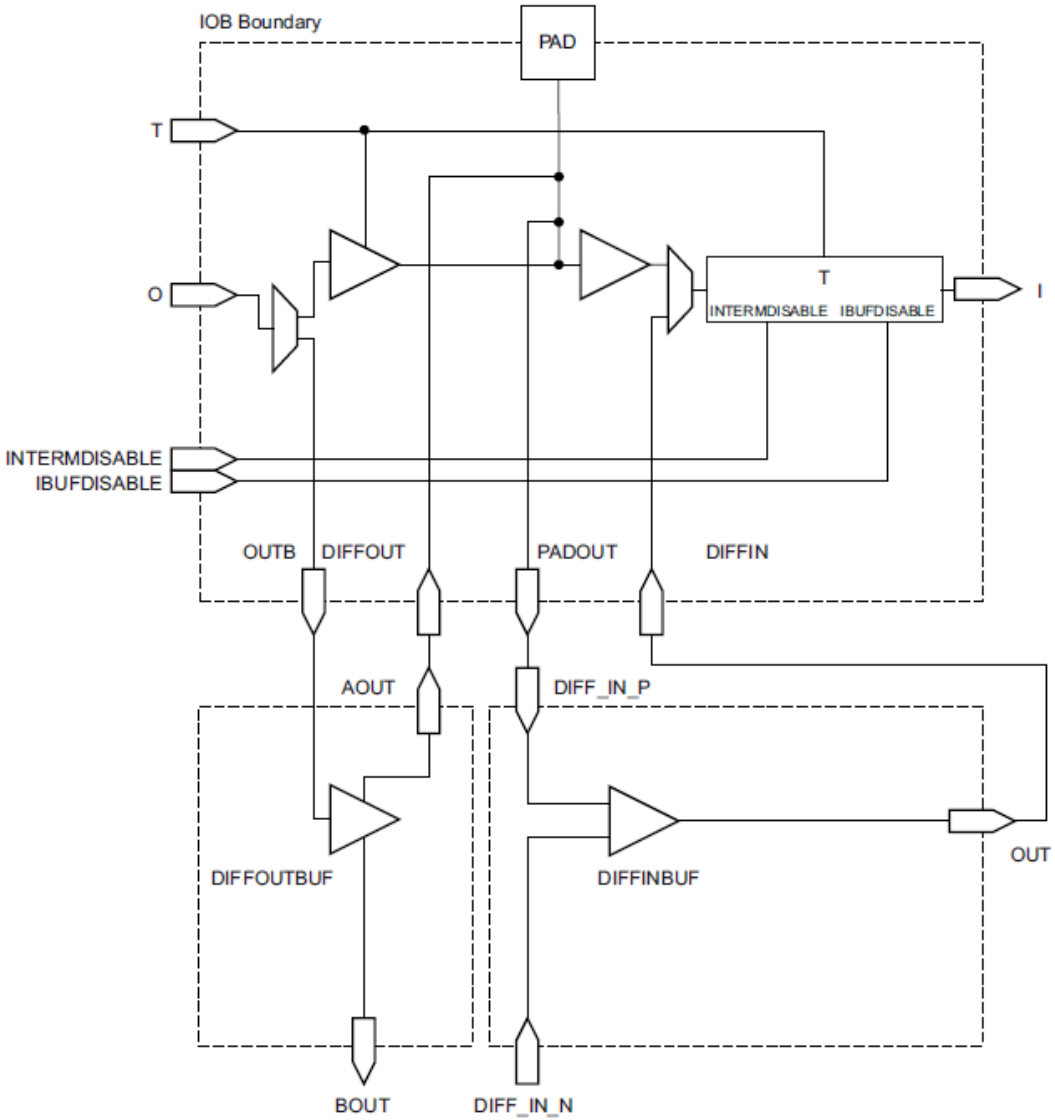
HR IOB

Single Ended (VRP pin used as general IO)



2024 © Centre for Development of Advanced Computing

HR IOB



Standard

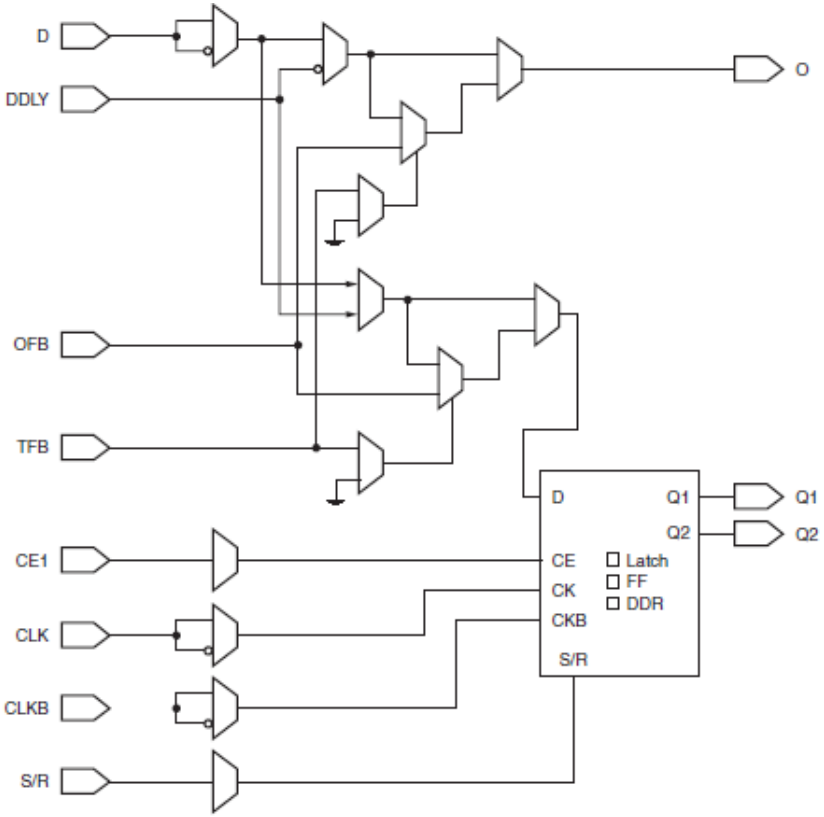
2024 © Centre for Development of Advanced Computing

IOB logic resources:

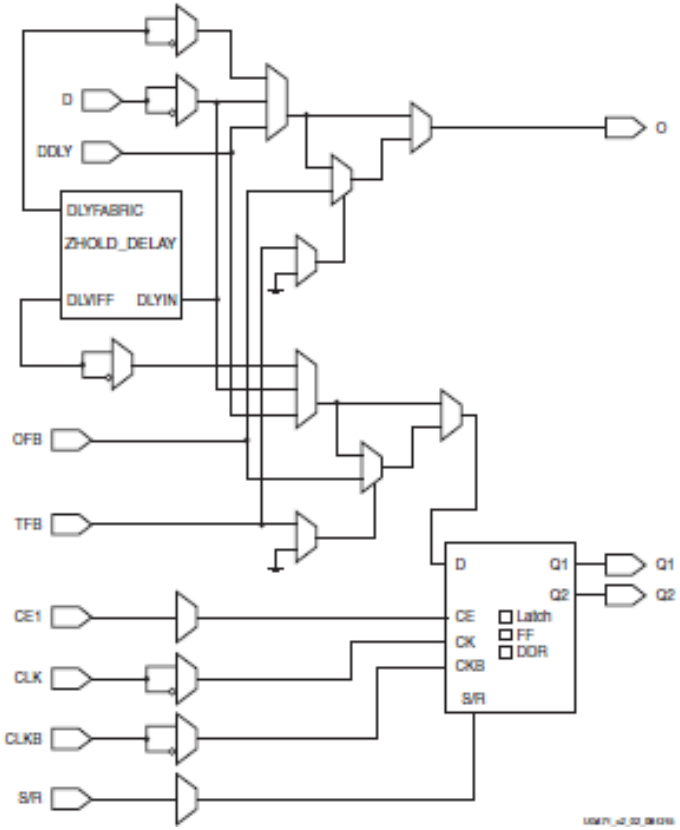
- Combinatorial input/output
- 3-state output control
- Registered input/output
- Registered 3-state output control
- Double-Data-Rate (DDR) input/output upto 2666Mb/s
- DDR output 3-state control
- IDELAYE3 and/or ODELAY3 along with IDELAYCTRL provides users control of an adjustable, fine-resolution delay element
- ISERDES3/OSERDES3

2024 © Centre for Development of Advanced Computing

ILOGIC



ILOGIC (HP IOB)



ILOGIC (HR IOB)

2024 © Centre for Development of Advanced Computing

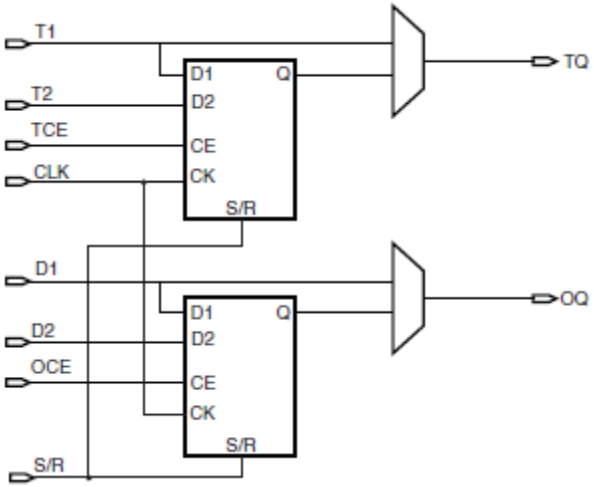
ILOGIC

- Input Path:
  - 52 IOBs per bank
  - Direct connection to ILOGIC/OLOGIC resources
  - Used to route the input signal directly to the internal logic or thru a FF/Latch at the input, or thru DDR input registers
  - Optional delay element at the input is useful to eliminate (or reduce) hold time at the cost of increased setup time at the inputs (IDELAYE3: more on this later)
  - Each input buffer is configurable to recognize logic ‘0’ and ‘1’ levels for either single ended (e.g. LVTTL), differential (e.g. LVDS) or user supplied V_{REF} levels (SelectIO: more on this later)
  - Input path can be configured for a number of termination options (DCI: more on this later)

2024 © Centre for Development of Advanced Computing

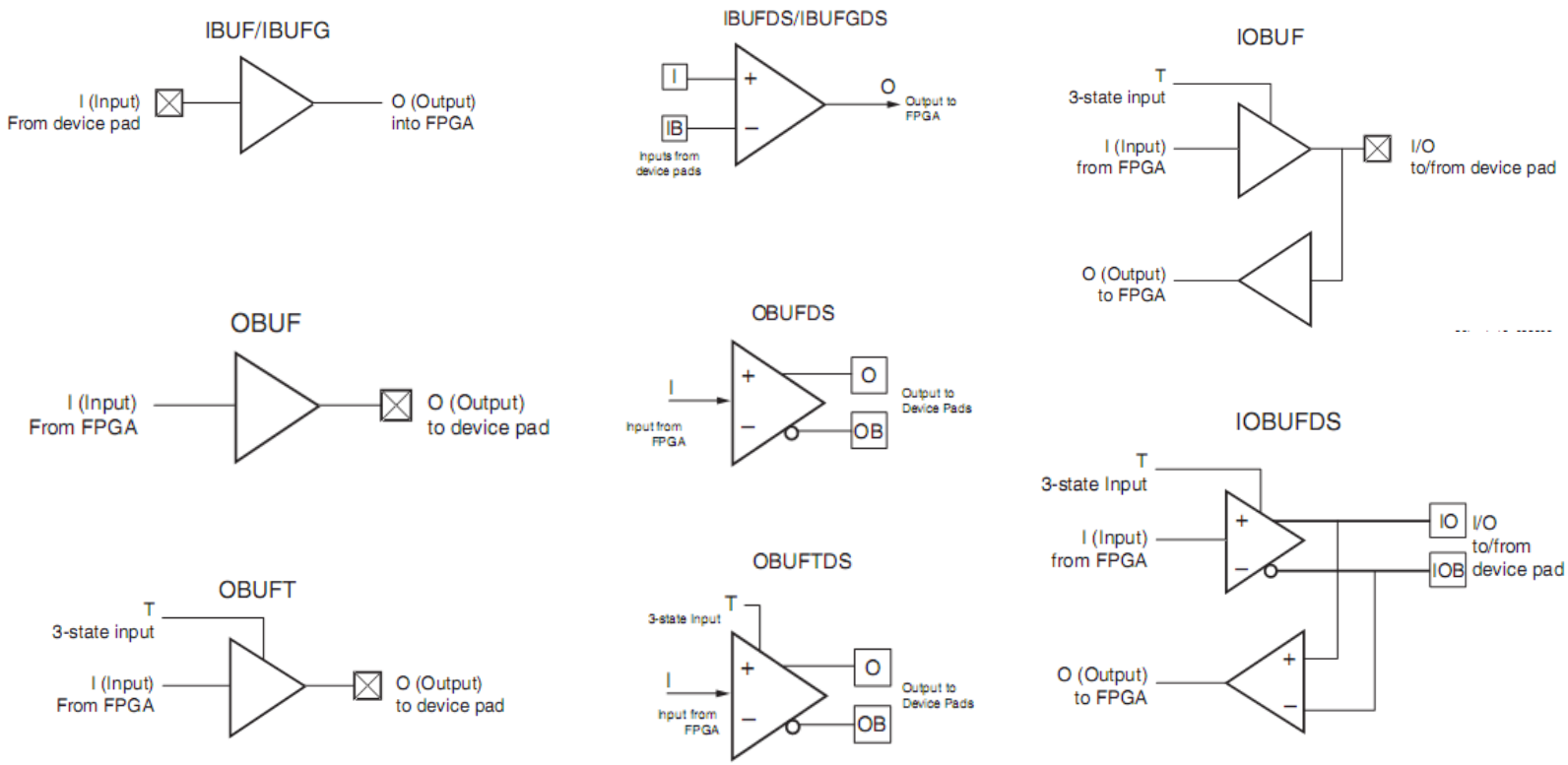
OLOGIC

- Output Path
  - Output from internal logic can be routed to the pin
    - Thru a FF or latch,
    - Thru a FF with tristate control,
    - Directly (Asynchronous/Combinatorial),
    - DDR capability for double BW per pin / clock: Pair of outputs given out on two clocks 180deg apart and combined via a mux (DDR o/p), optional tristate capability available



2024 © Centre for Development of Advanced Computing

Basic IOB



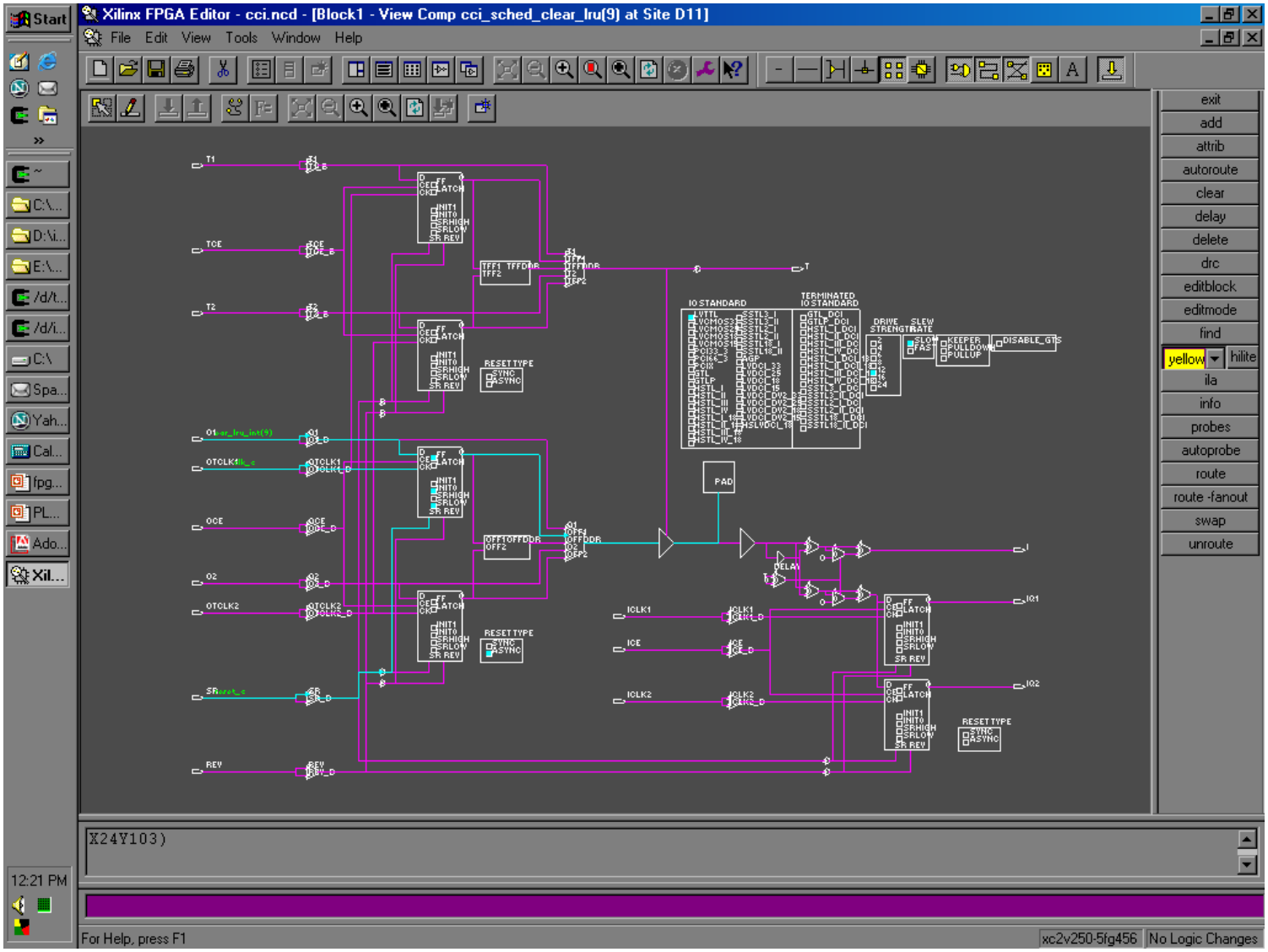
2024 © Centre for Development of Advanced Computing

Basic IOB

- FF attributes: INIT0/1: initial state, set/reset active low/high, set/reset asynch or sync
- Clock Enable: enable/disabled, polarity selectable
- Clock: polarity selectable
- All the configurable elements are set by setting the attributes in the HDL instantiation. This information is in turn passed to the device in the configuration phase
- Mux selection is also done in configuration

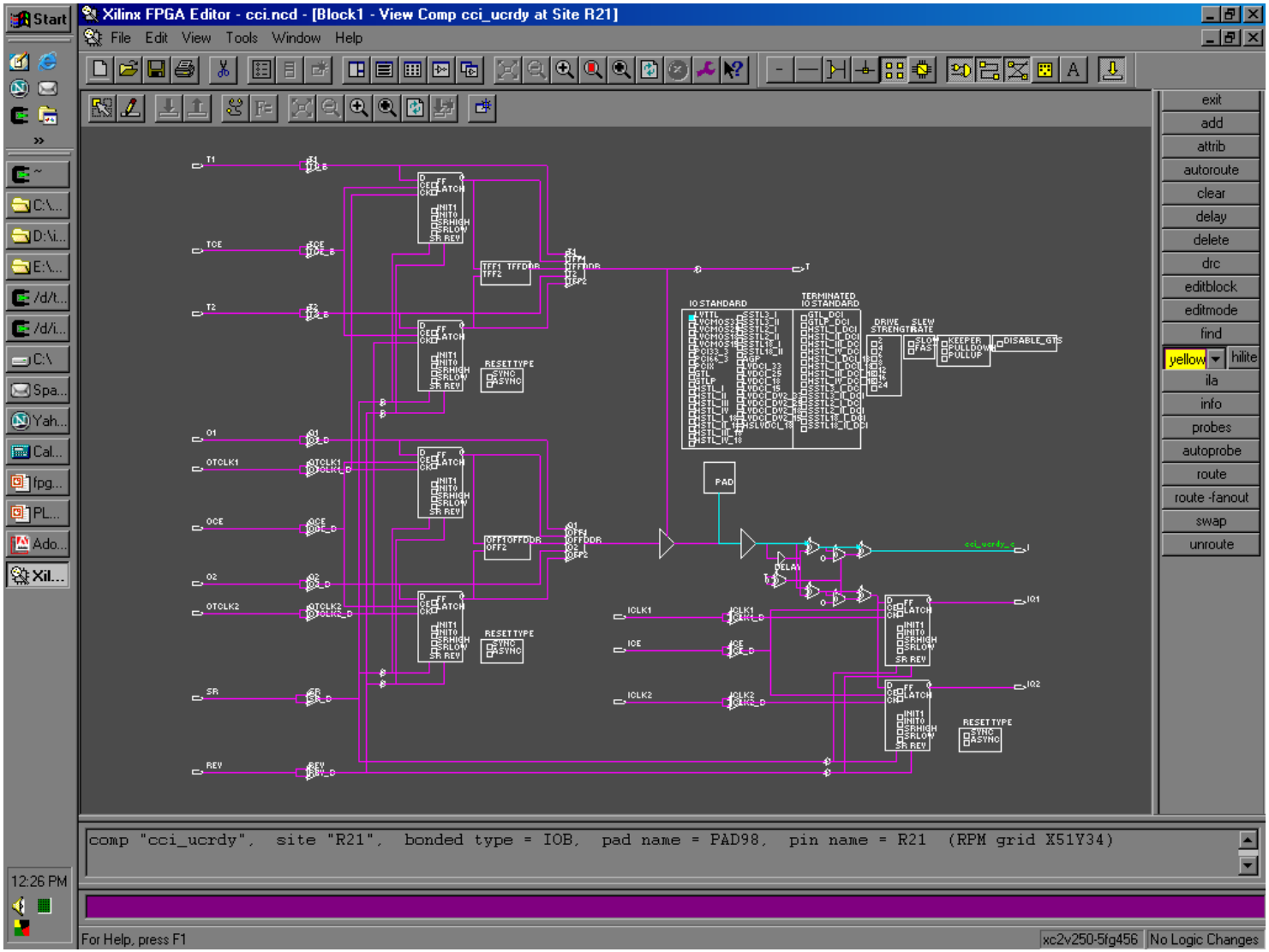
2024 © Centre for Development of Advanced Computing

IOB view: Registered O/P



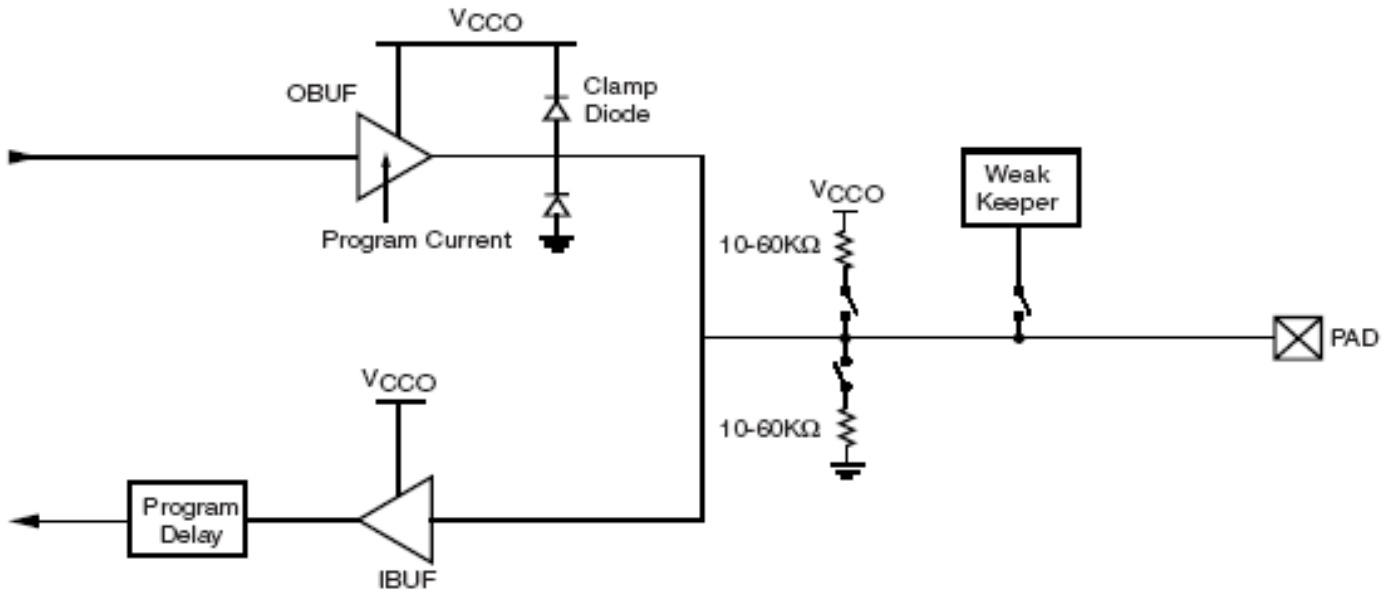
2024 © Centre for Development of Advanced Computing

IOB view: I/P



2024 © Centre for Development of Advanced Computing

IOB pin



- O/p path has tristate control, programmable slew rate and drive strength
- O/p swing controlled thru VCCO
- I/P path has programmable delay
- Pullup/Pulldowns, weak keeper
- Clamp diodes for overvoltage/ESD protection

2024 © Centre for Development of Advanced Computing

IOB pin

- Pull-up or pull-down within chip eliminates external resistors
- Weak keeper: “Remember”s the last voltage level and drives the pin appropriately (weakly).
  - Useful when multiple drivers are connected together and none of them is driving the line. In absence of weak keeper, voltage level tends to drift
  - Pull-ups or pull-downs may load the bus, increase power dissipation
- All pads have ESD and over-voltage protection (in the form of clamping diode)

2024 © Centre for Development of Advanced Computing

IOB features

- SelectIO capability addresses need of both single ended and differential standards with easy to instantiate I/O buffer macros in HDLs
- Digitally controlled impedance addresses need of terminations in a variety of configurations. Impedance is constant irrespective of process variation, temperature and supply voltage changes
- No external components are required for either of above except for reference values of V_{REF}, V_{TT} and resistors. This saves significant board area especially for designs using wide busses

2024 © Centre for Development of Advanced Computing

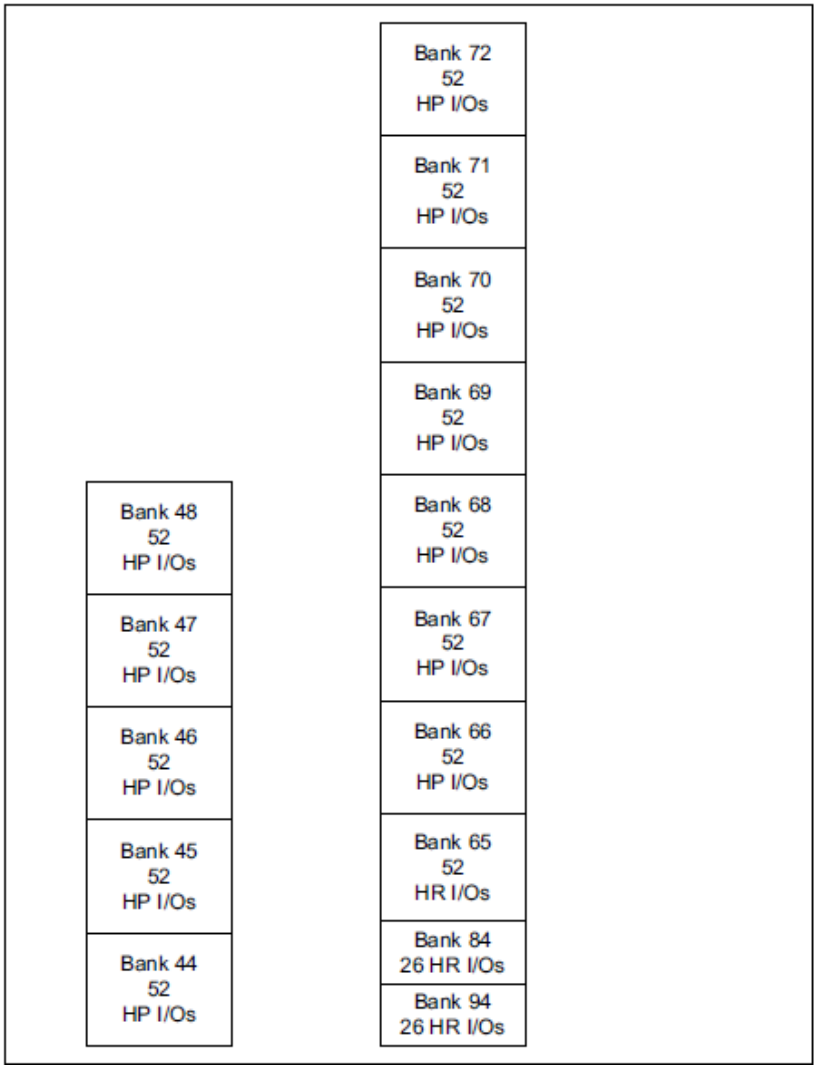
IOB drive strength

- Configurable drive strength for compensating line load
- Slew rate: slow or fast

Standard	HR I/O Bank Current Drive (mA)		HP I/O Bank Current Drive (mA)	
	Allowed Values	Default	Allowed Values	Default
LVCMOS12	4, 8, or 12	12	2, 4, 6, or 8	12(1)
LVCMOS15	4, 8, 12, or 16	12	2, 4, 6, 8, or 12	12
LVCMOS18	4, 8, 12, or 16	12	2, 4, 6, 8, or 12	12
LVCMOS25	4, 8, 12, or 16	12	N/A	N/A
LVCMOS33	4, 8, 12, or 16	12	N/A	N/A
LVTTL	4, 8, 12, or 16	12	N/A	N/A

2024 © Centre for Development of Advanced Computing

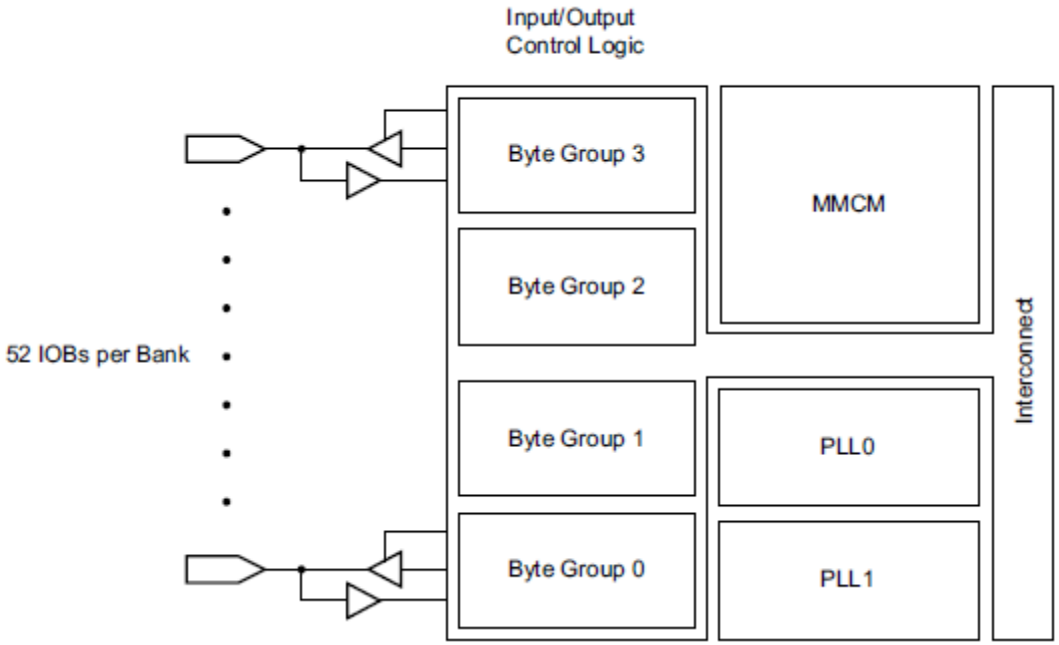
SelectIO and I/O banks



- I/Os are configured per-bank basis, For each bank, V_{REF} and V_{TT} needs to be supplied (and supply voltage Vcco)
- R_{REF} needs to be supplied for every bank if terminations are needed
- There are restrictions regarding mixing of standards within a bank

2024 © Centre for Development of Advanced Computing

SelectIO and I/O banks

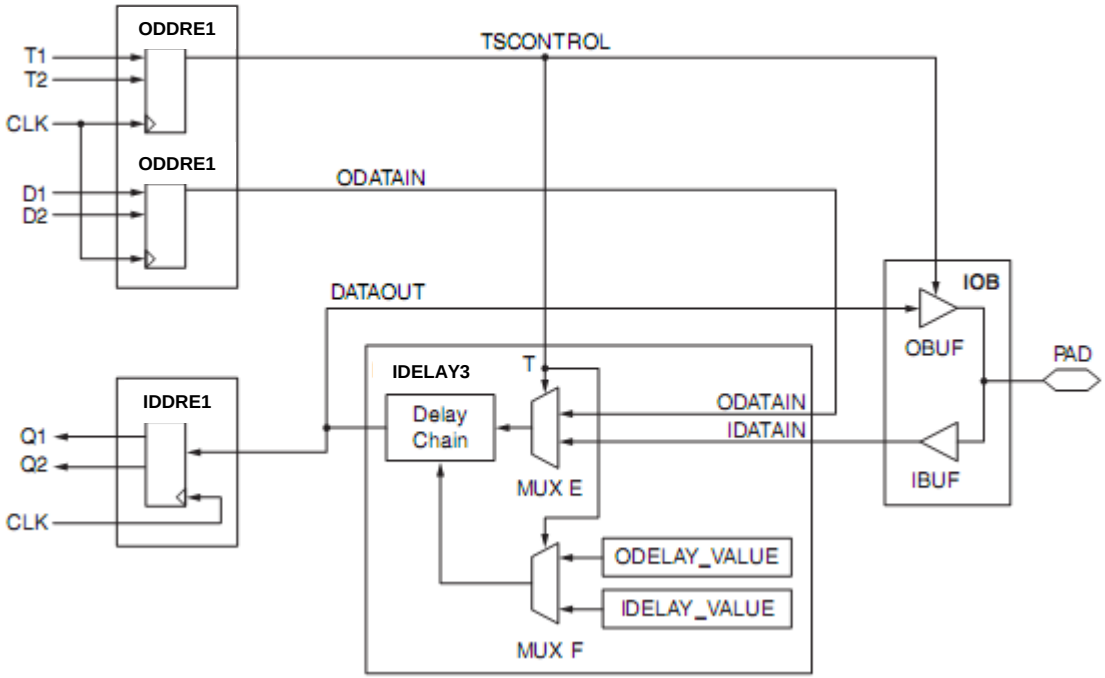


- 52 pins per bank
- Can be configured as 24 differential pairs identified as bit slice for single ended and master-slave bit slice for differential

2024 © Centre for Development of Advanced Computing

DDR path

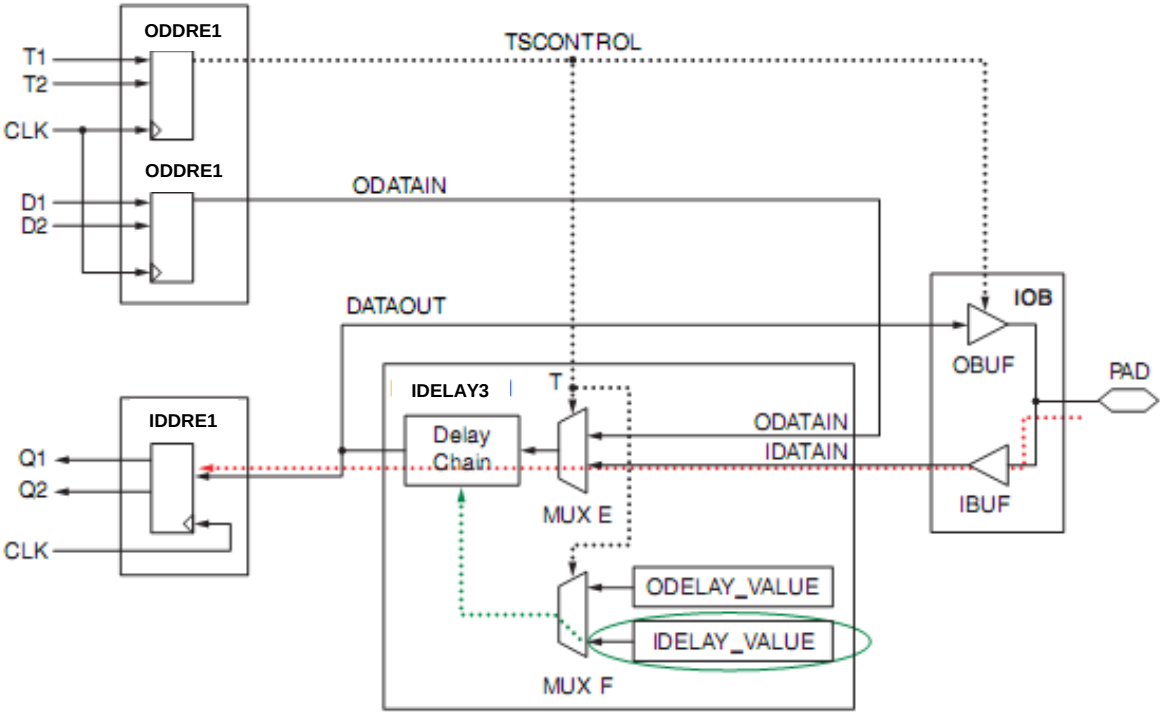
- D1 and D2 are derived from a 2X Wide data path @ 1/2 clock freq
- clock sync issues: FIFOs may be required
- Uses enhanced versions of IDDR and ODDR selectIO logic resources designated as IDDRE1 and ODDRE1



2024 © Centre for Development of Advanced Computing

DDR path

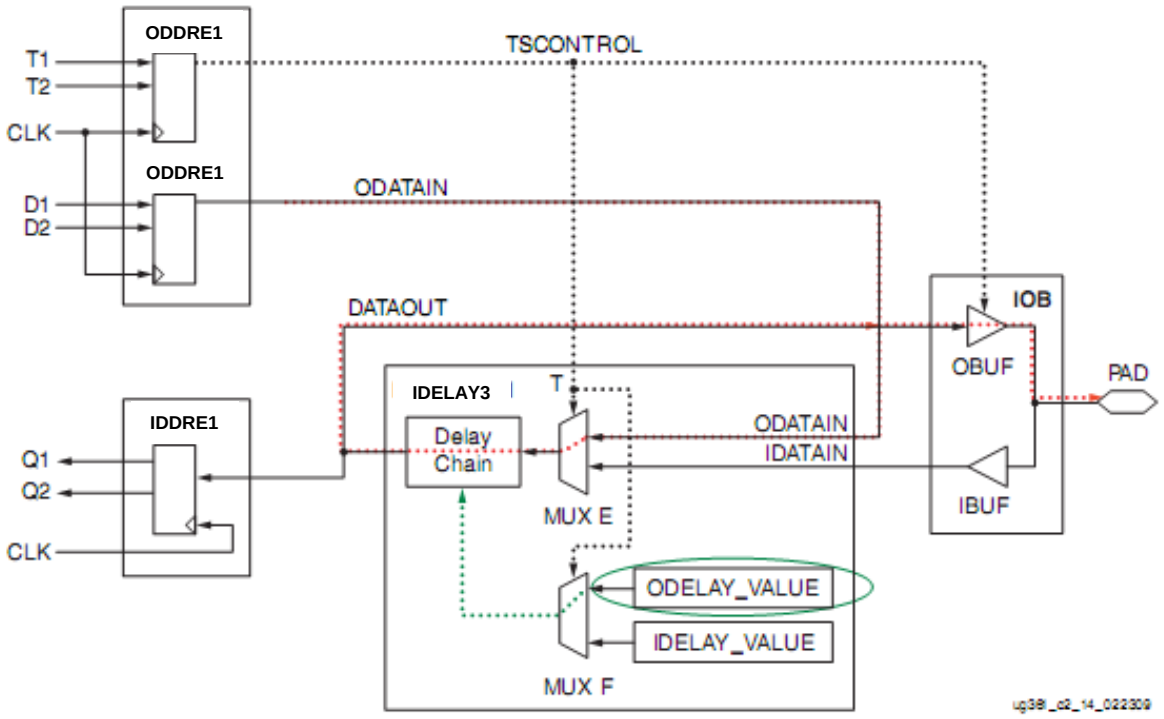
Tristate disabled



2024 © Centre for Development of Advanced Computing

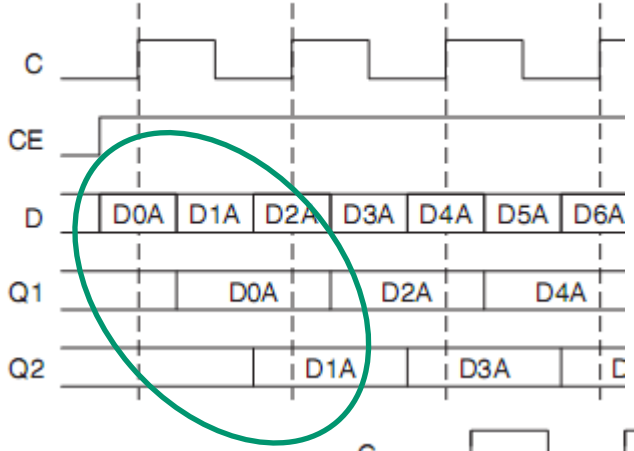
DDR path

Tristate enabled



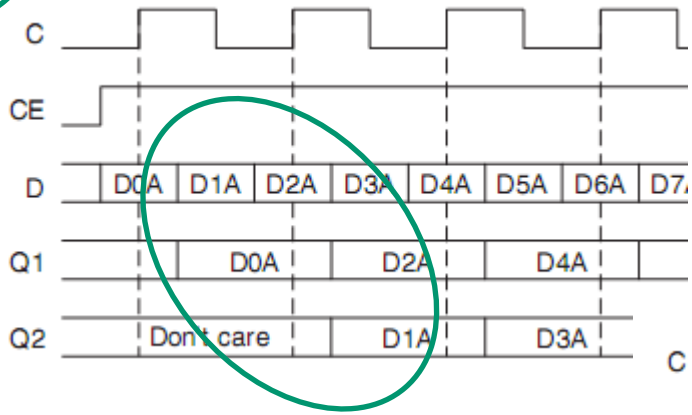
2024 © Centre for Development of Advanced Computing

DDR waveforms

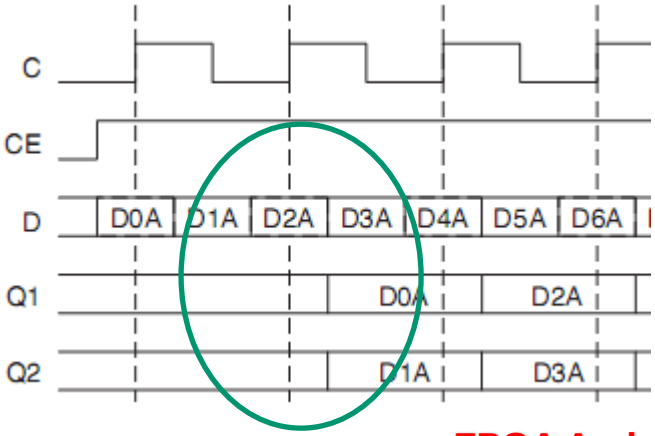


Input DDR Timing in OPPOSITE_EDGE Mode

Input path



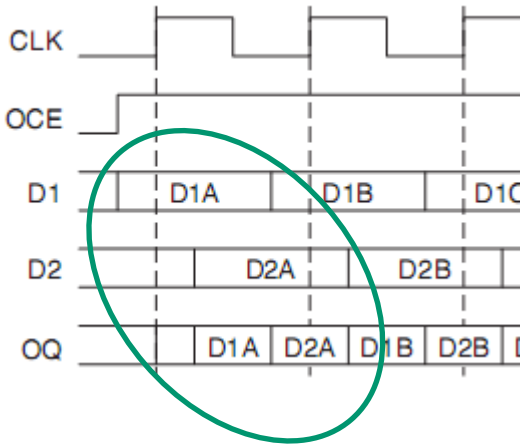
Input DDR Timing in SAME_EDGE Mode



Input DDR Timing in SAME_EDGE_PIPELINED Mode

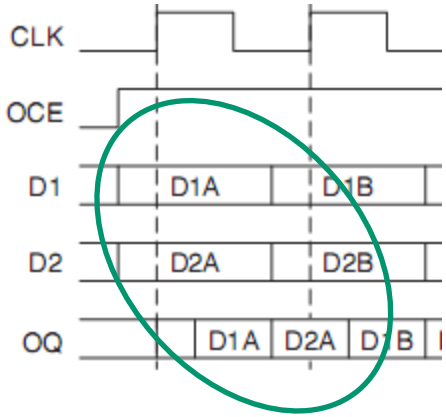
2024 © Centre for Development of Advanced Computing

DDR waveforms



Output DDR Timing in OPPOSITE_EDGE Mode

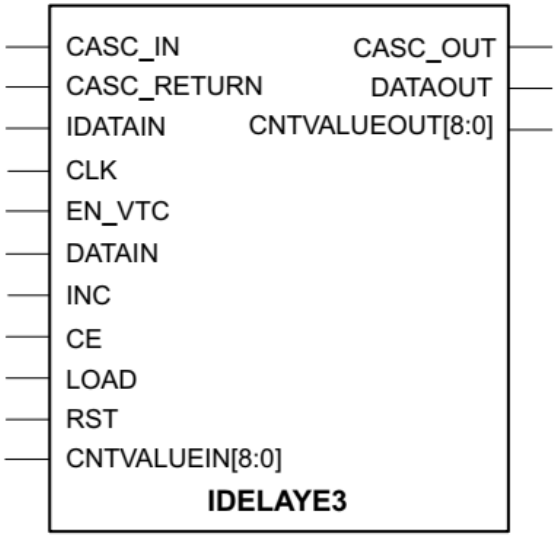
Output path



Output DDR Timing in SAME_EDGE Mode

2024 © Centre for Development of Advanced Computing

IDELAYE3

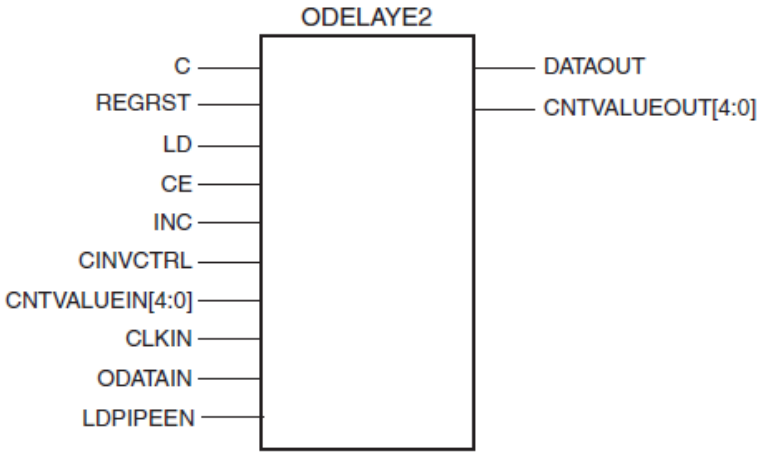


- It is a 512-tap, wraparound, delay element with a calibrated tap resolution of 2.5 – 15 ps
- It can be connected to ISERDES3
- It can also be accessed directly in FPGA logic

- Delay can be fixed or variable depending on implementation needs. Count mode also supported.
- Delay is continuously calibrated using IDELAYCTRL module to reduce effects of process, temperature and voltage variations

2024 © Centre for Development of Advanced Computing

ODELAYE3



It is a 512-tap, wraparound, delay element with a calibrated tap resolution of 2.5 – 15 ps

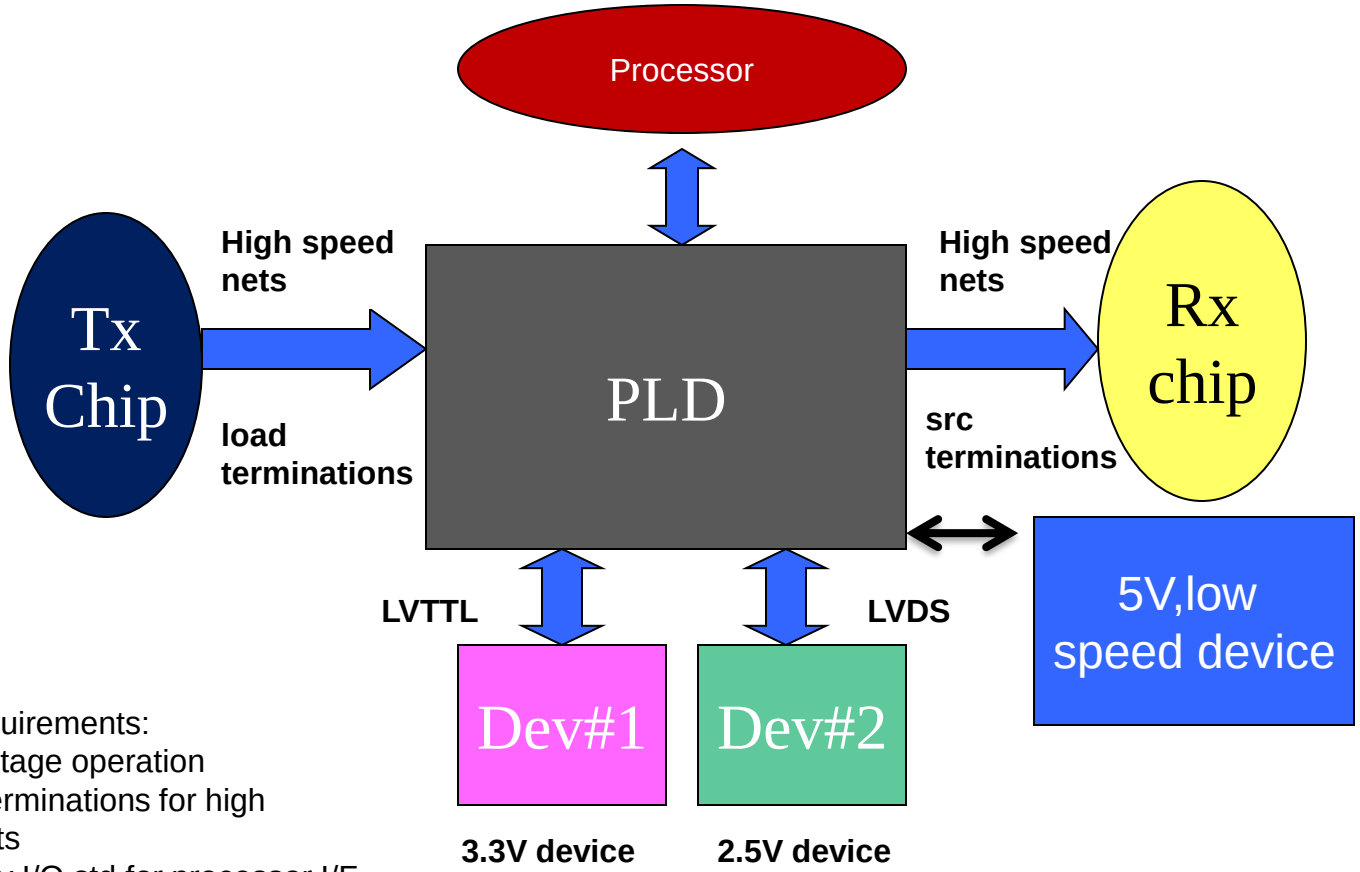
It can be connected to OSERDES3 only (HP IOB)

It can also be accessed directly in FPGA logic

- Delay can be fixed or variable depending on implementation needs
- Delay is continuously calibrated using IDELAYCTRL module to reduce effects of process, temperature and voltage variations

2024 © Centre for Development of Advanced Computing

IOB ecosystem



Diverse requirements:
- Mixed voltage operation
- Rx/Tx : terminations for high speed nets
- Proprietary I/O std for processor I/F
- Single Ended/ differential interfaces
- High speed serial transceivers

2024 © Centre for Development of Advanced Computing

Common I/O standards

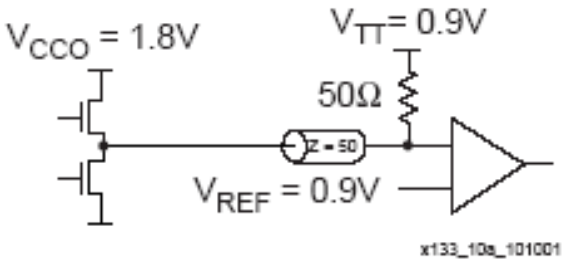
- LVTTL and LVCMOS: most common
  - ‘LV’ means ‘low voltage’ i.e. 3.3V going towards 2.5V
  - Except for output swing levels these two are similar
    - VIL(max)=0.8V, VOL(max)=0.4V; VIH(min)=2.0V, VOH(min)=2.4V ➔ min voltage swing ~2V
  - Motivation for going for lesser voltage swings is faster transitions (dV/dt) and lower power dissipation ($\propto V^2$)
  - These fail short for newer systems requiring high speed nets
  - Reduced voltage swings for CMOS 2.5, 1.8 and 1.5 stds
- Differential signaling
  - Limit output voltage swings so as to improve transition times and hence frequency
  - Greater immunity from electrical interference such as crosstalk because of differential nature of the signal
  - Matching transmission lines allows for larger distances
  - All above advantages have made LVDS a popular I/O std (3.5mA, 100E, ~350mv dV around ~1.2V)

2024 © Centre for Development of Advanced Computing

I/O stds

- Other I/O standards: These interfaces are often defined by industry leaders and used in their product offerings. Need of defining these comes from speed and/or power dissipation considerations
  - HSTL: High speed transceiver logic (IBM) (SSRAM memories)
  - SSTL: series stub TL (IBM) (DDR)
- Characterized by need of V_{REF} (ref voltage) and V_{TT} (termination voltage)

HSTL Class I (1.8V)



2024 © Centre for Development of Advanced Computing

2.5V IO standards

I/O Standard	V_{CCO}		V_{REF}	Termination Type	
	Output	Input	Input	Output	Input
LVDS_25	2.5	Note (1)	N/R	N/R	N/R
LVDSEXT_25			N/R	N/R	N/R
HT_25			N/R	N/R	N/R
RSDS_25(3)			N/R	N/R	N/R
BLVDS_25			N/R	N/R	N/R
LVPECL_25			N/R	N/R	N/R
SSTL2_I			1.25	N/R	N/R
SSTL2_II			1.25	N/R	N/R
DIFF_SSTL2_I			N/R	N/R	N/R
DIFF_SSTL2_II			N/R	N/R	N/R
LVCMOS25		2.5	N/R	N/R	N/R
LVDCI_25			N/R	Series	N/R
HSLVDCI_25			$V_{CCO}/2$	Series	N/R
LVDCI_DV2_25			N/R	Series	N/R
SSTL2_I_DCI			1.25	N/R	Split
SSTL2_II_DCI			1.25	Split	Split
SSTL2_II_T_DCI			1.25	N/R	Split
DIFF_SSTL2_I_DCI			N/R	N/R	Split
DIFF_SSTL2_II_DCI			N/R	Split	Split
DIFF_SSTL2_II_T_DCI			N/R	N/R	Split

2024 © Centre for Development of Advanced Computing

1.8V IO standards

I/O Standard	V_{CCO}		V_{REF}	Termination Type	
	Output	Input	Input	Output	Input
HSTL_III_18	1.8	Note (1)	1.08	N/R	N/R
HSTL_I_18			0.9	N/R	N/R
HSTL_II_18			0.9	N/R	N/R
DIFF_HSTL_I_18			N/R	N/R	N/R
DIFF_HSTL_II_18			N/R	N/R	N/R
SSTL18_I			0.9	N/R	N/R
SSTL18_II			0.9	N/R	N/R
DIFF_SSTL18_I			N/R	N/R	N/R
DIFF_SSTL18_II			N/R	N/R	N/R
LVCMOS18		1.8	N/R	N/R	N/R
LVDCI_18			N/R	Series	N/R
HSLVDCI_18			$V_{CCO}/2$	Series	N/R
LVDCI_DV2_18			N/R	Series	N/R
HSTL_III_DCI_18			1.08	N/R	Single
HSTL_I_DCI_18			0.9	N/R	Split
HSTL_II_DCI_18			0.9	Split	Split
HSTL_II_T_DCI_18			0.9	N/R	Split
DIFF_HSTL_I_DCI_18			N/R	N/R	Split
DIFF_HSTL_II_DCI_18			N/R	Split	Split
DIFF_HSTL_II_T_DCI_18			N/R	N/R	Split
SSTL18_I_DCI			0.9	N/R	Split
SSTL18_II_DCI			0.9	Split	Split
SSTL18_II_T_DCI			0.9	N/R	Split
DIFF_SSTL18_I_DCI			N/R	N/R	Split
DIFF_SSTL18_II_DCI			N/R	Split	Split
DIFF_SSTL18_II_T_DCI			N/R	N/R	Split

2024 © Centre for Development of Advanced Computing

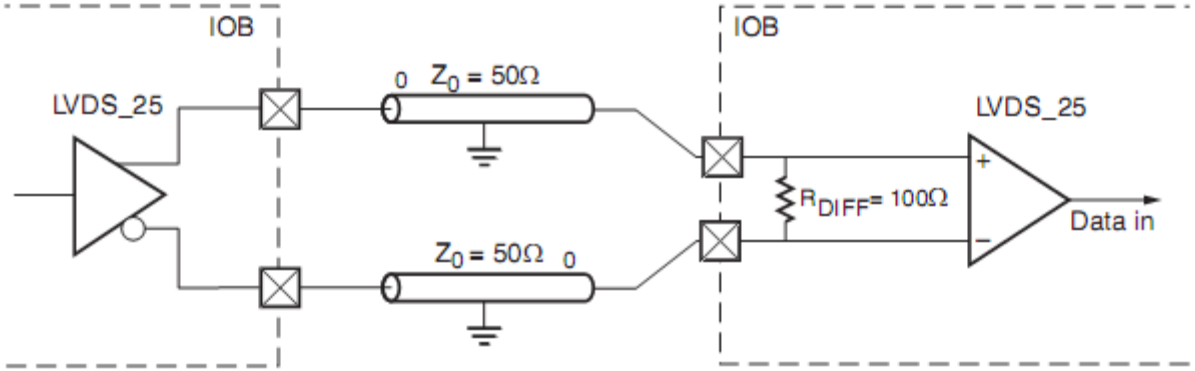
1.5V and 1.2V IO standards

I/O Standard	V_{CCO}		V_{REF}	Termination Type	
	Output	Input	Input	Output	Input
HSTL_III	1.5	Note (1)	0.9	N/R	N/R
HSTL_I			0.75	N/R	N/R
HSTL_II			0.75	N/R	N/R
DIFF_HSTL_I			N/R	N/R	N/R
DIFF_HSTL_II			N/R	N/R	N/R
LVCMOS15		1.5	N/R	N/R	N/R
LVDCI_15			N/R	Series	N/R
HSLVDCI_15			$V_{CCO}/2$	Series	N/R
LVDCI_DV2_15			N/R	Series	N/R
HSTL_III_DCI			0.9	N/R	Single
HSTL_I_DCI			0.75	N/R	Split
HSTL_II_DCI			0.75	Split	Split
HSTL_II_T_DCI			0.75	N/R	Split
DIFF_HSTL_I_DCI			N/R	N/R	Split
DIFF_HSTL_II_DCI			N/R	Split	Split
DIFF_HSTL_II_T_DCI			N/R	N/R	Split
SSTL15			0.75	N/R	N/R
SSTL15_DCI			0.75	Split	Split
SSTL15_T_DCI			0.75	N/R	Split
DIFF_SSTL15			N/R	N/R	N/R
DIFF_SSTL15_DCI			N/R	Split	Split
DIFF_SSTL15_T_DCI			N/R	N/R	Split
LVCMOS12	1.2	1.2	N/R	N/R	N/R
HSTL_I_12			0.6	N/R	N/R

2024 © Centre for Development of Advanced Computing

Need of termination

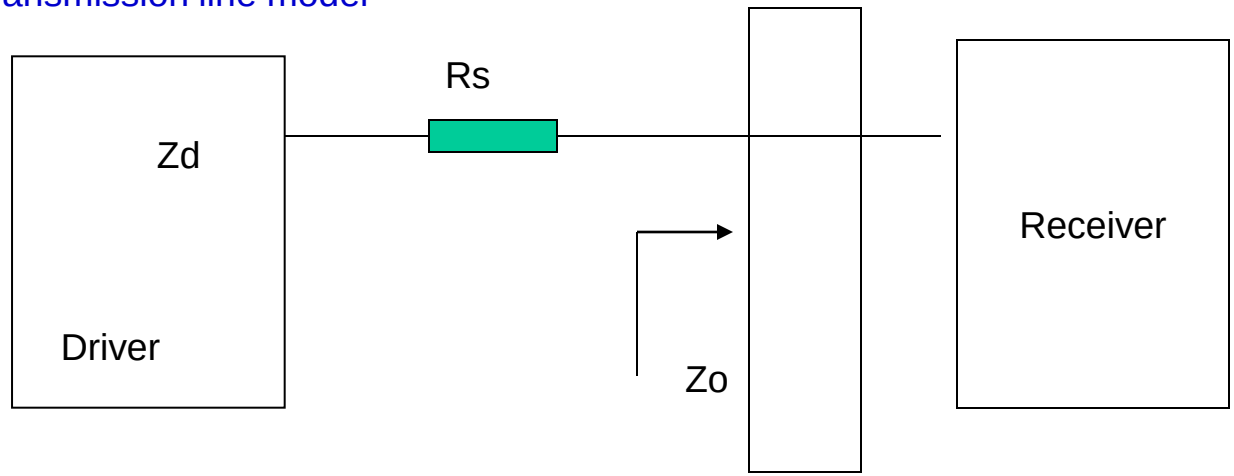
- Terminations are required
  - as defined by standards such as HSTL/SSTL
  - in conventional standards to meet signal integrity (SI) requirements at the I/O pin



2024 © Centre for Development of Advanced Computing

Case study: series terminations

Transmission line model



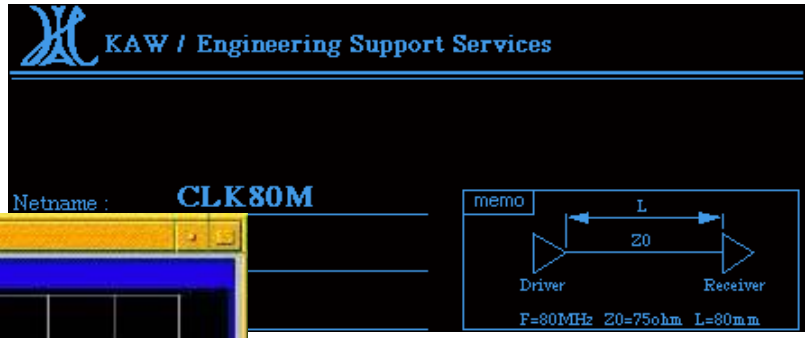
- Reflection coeff = (Zl – Z0)/(Zl + Z0). If nonzero, SI will suffer

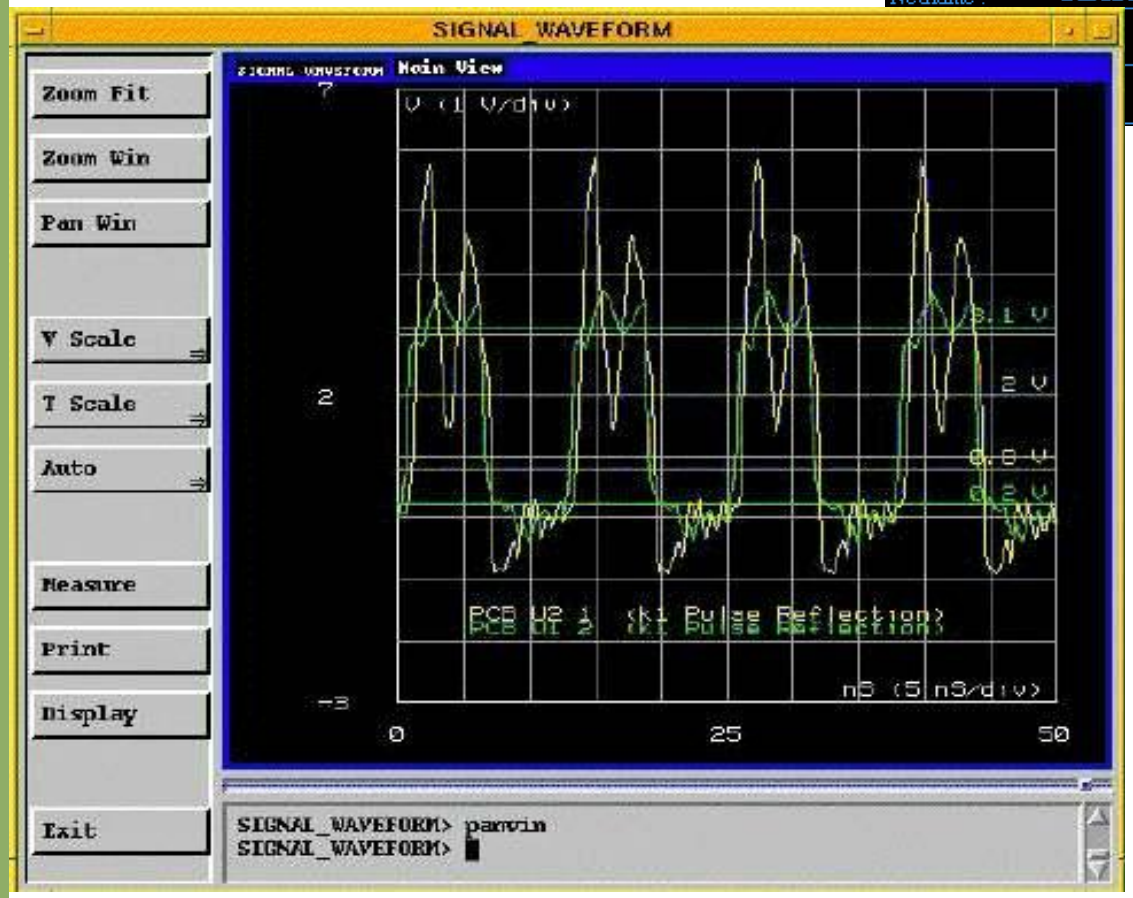
  where, Zl = Zd + Rs

  Zo is a function of PCB material and trace geometry, generally ~50 ohms
- Rs is generally in the same range (~30 ohms) since Zd is low.
- Series termination makes line slower because of R-C delays
- Parallel termination at the destination is faster, however draws more power.

2024 © Centre for Development of Advanced Computing

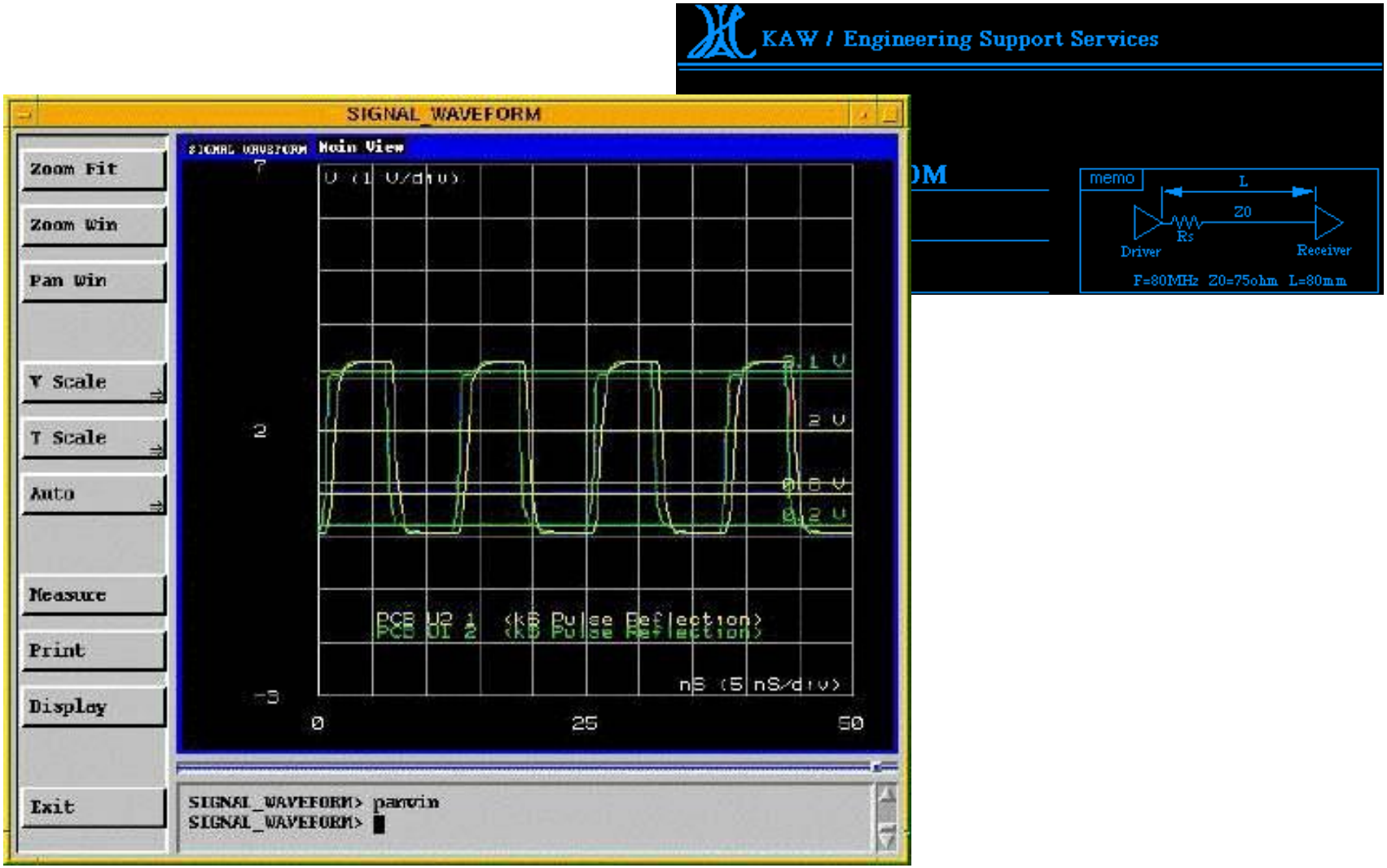
Simulation waveforms





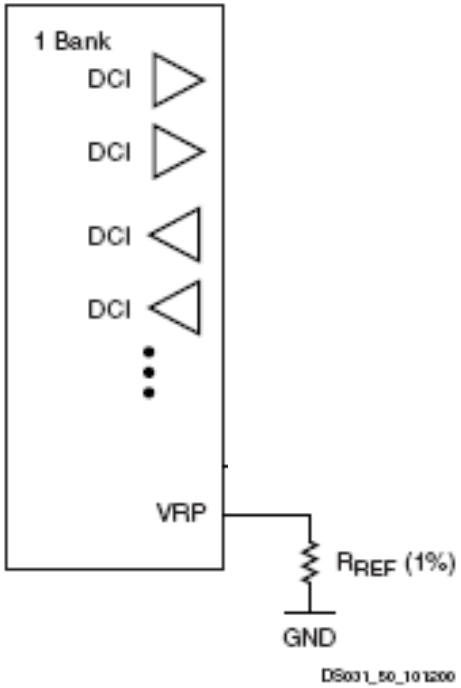
2024 © Centre for Development of Advanced Computing

Simulation waveforms

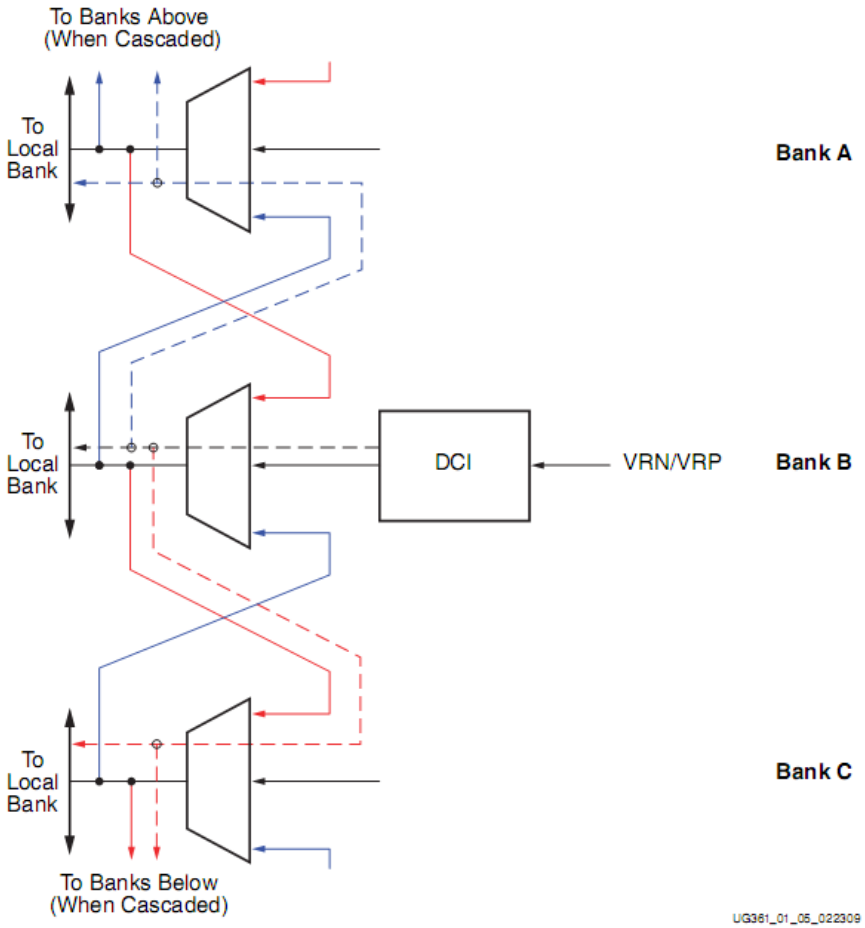


2024 © Centre for Development of Advanced Computing

DCI



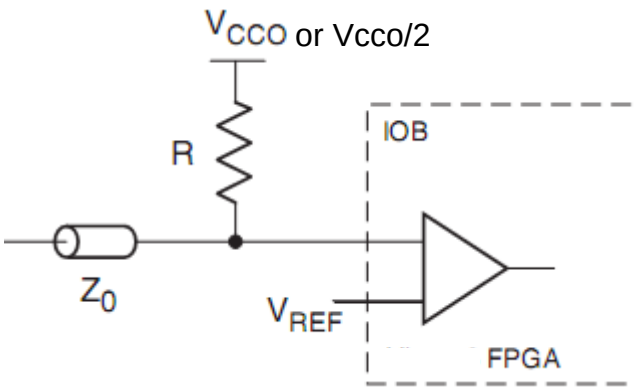
External resistor value must be 240 Ω



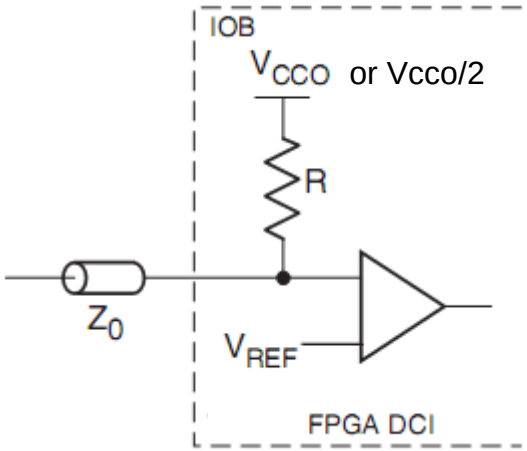
VRP pins on slave banks are available as user IOs in this case

2024 © Centre for Development of Advanced Computing

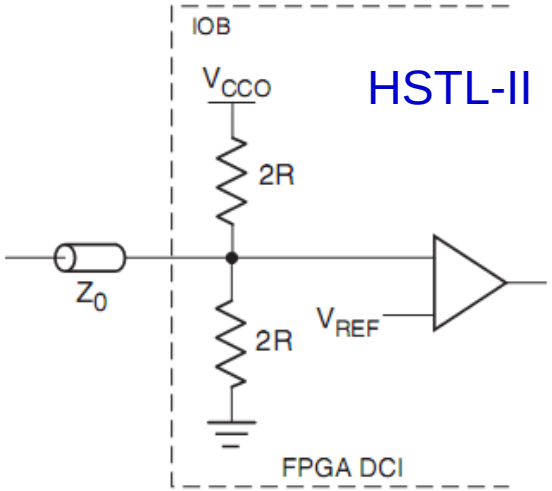
DCI



Without DCI



With DCI

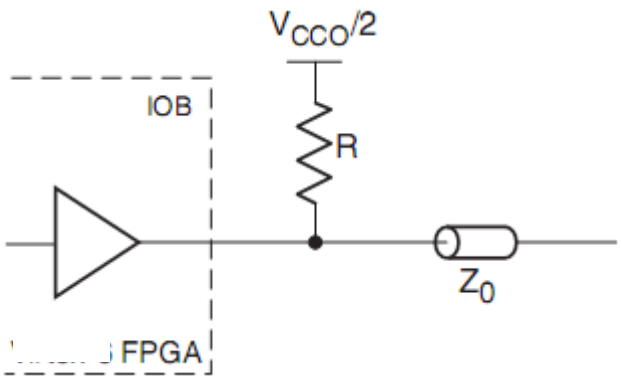


Guidelines:

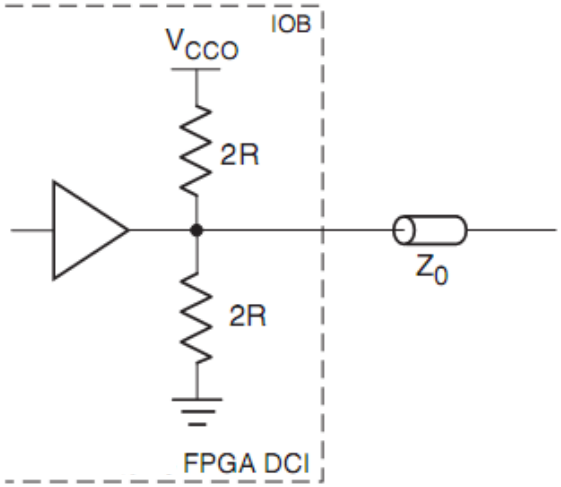
- Vcco must be connected to bank for a specific IO standard
- Correct DCI IO buffers should be used either by using IOSTANDARD attribute in constraint file (XDC) or instantiations in HDL code

2024 © Centre for Development of Advanced Computing

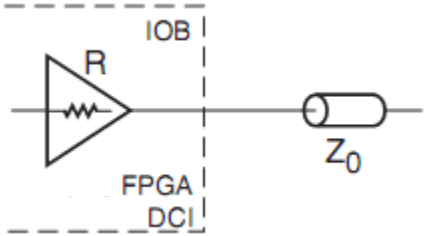
DCI

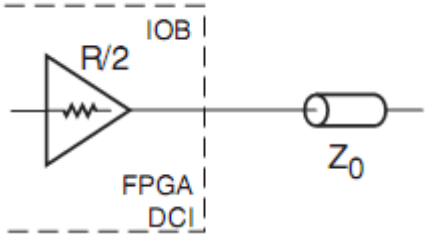


Without DCI



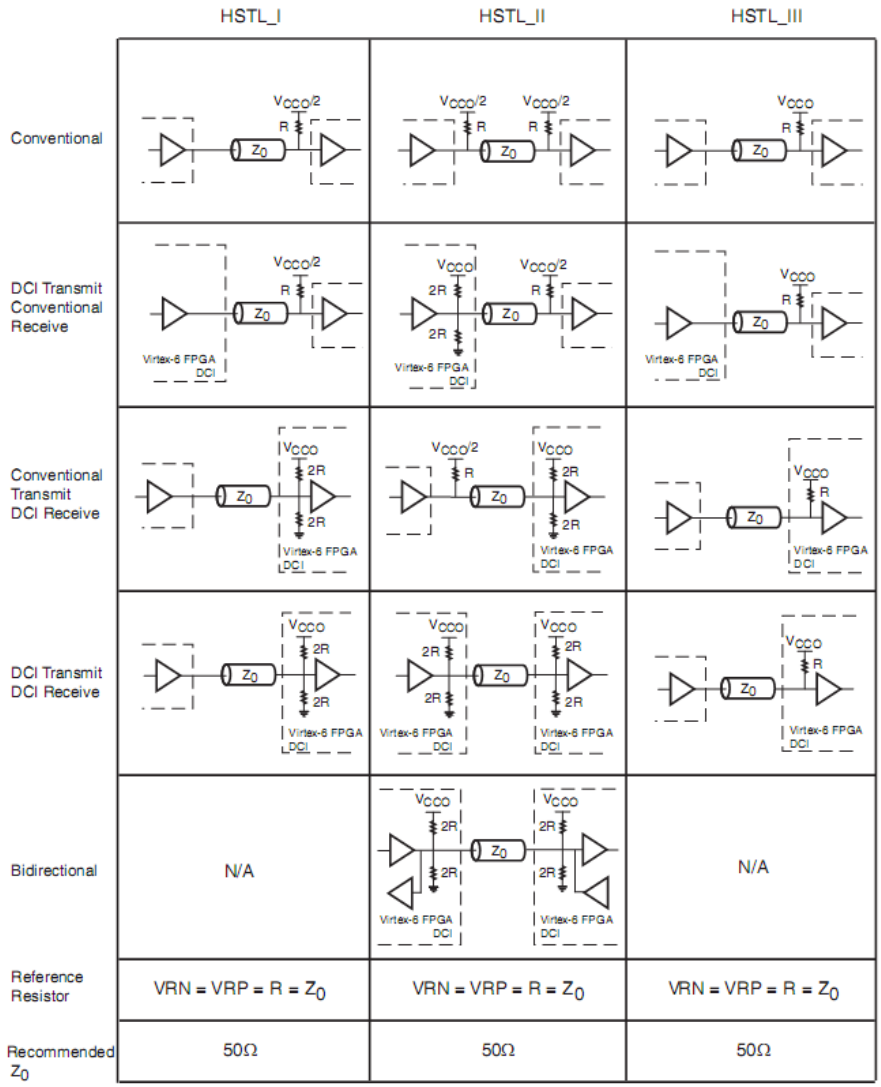
With DCI

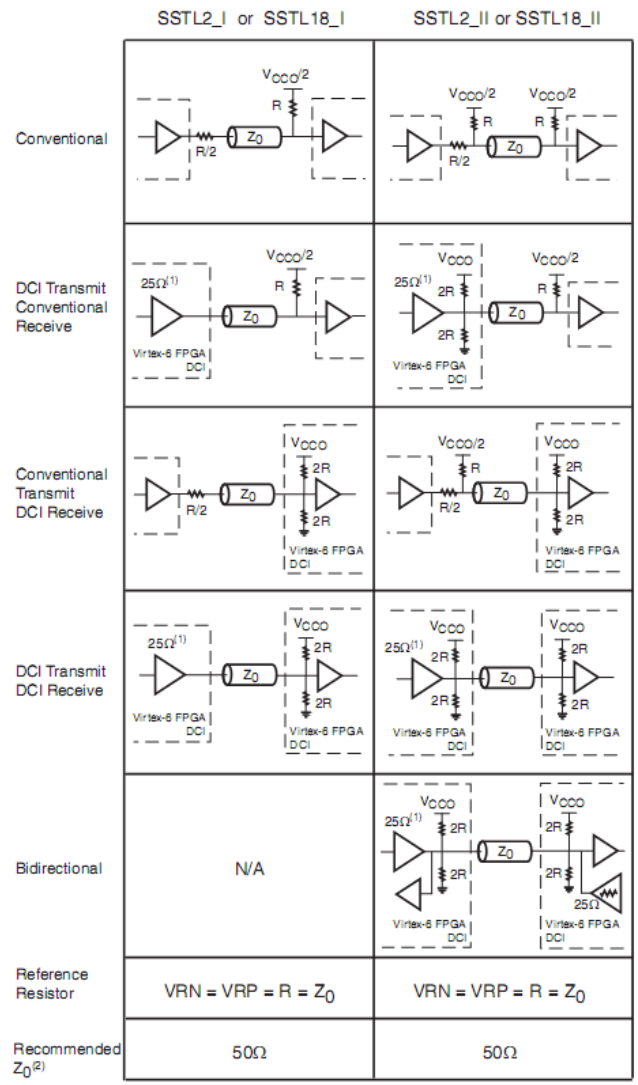




2024 © Centre for Development of Advanced Computing

DCI





2024 © Centre for Development of Advanced Computing

DCI

Rule book:

- Single termination, split termination, controlled impedance driver, and controlled impedance driver with half impedance can co-exist in the same bank provided they use same V_{cco} and V_{REF}
- No more than one DCI I/O standard using single termination type is allowed per bank
- No more than one DCI I/O standard using split termination type is allowed per bank
- Only available in HP bank

2024 © Centre for Development of Advanced Computing

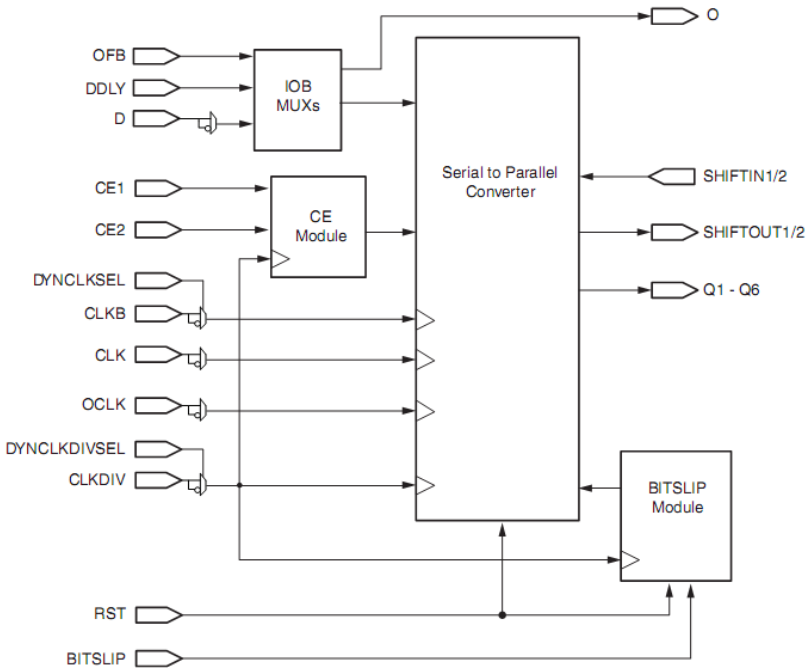
ISERDES3/OSERDES3

- This feature supports very high data rates and allow internal logic to run up to 10 times slower.
- Combined with LVDS IOB support designers can create very fast data path
- Together, Input serial-to-parallel converter (ISERDES3) and Output parallel-to-serial converter (OSERDES3) allow implementation of high speed source synchronous applications

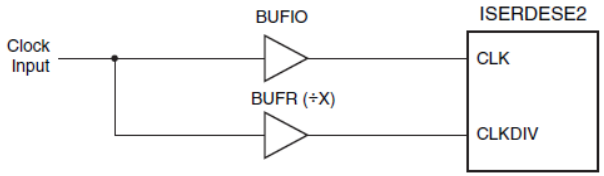
2024 © Centre for Development of Advanced Computing

ISERDES3/OSERDES3

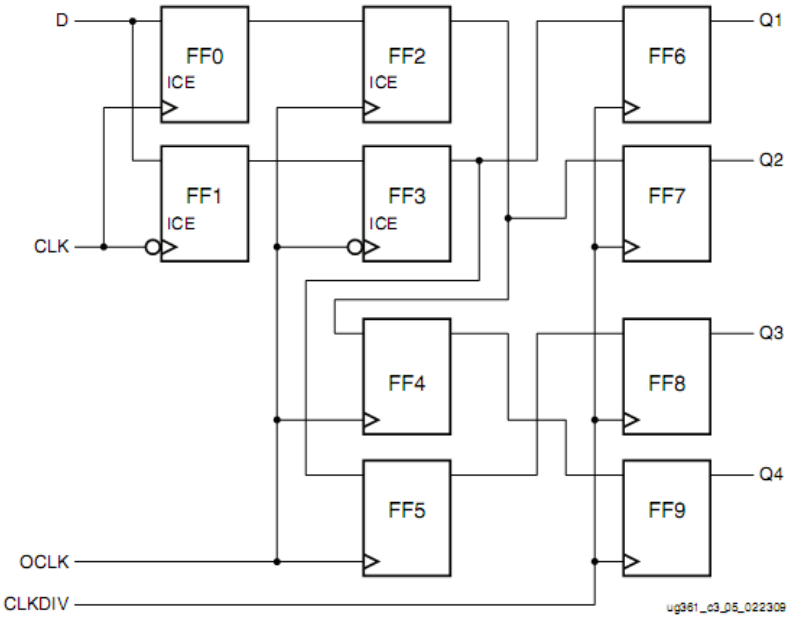
- Dedicated support for DDR3, QDR interfaces



ISERDES2



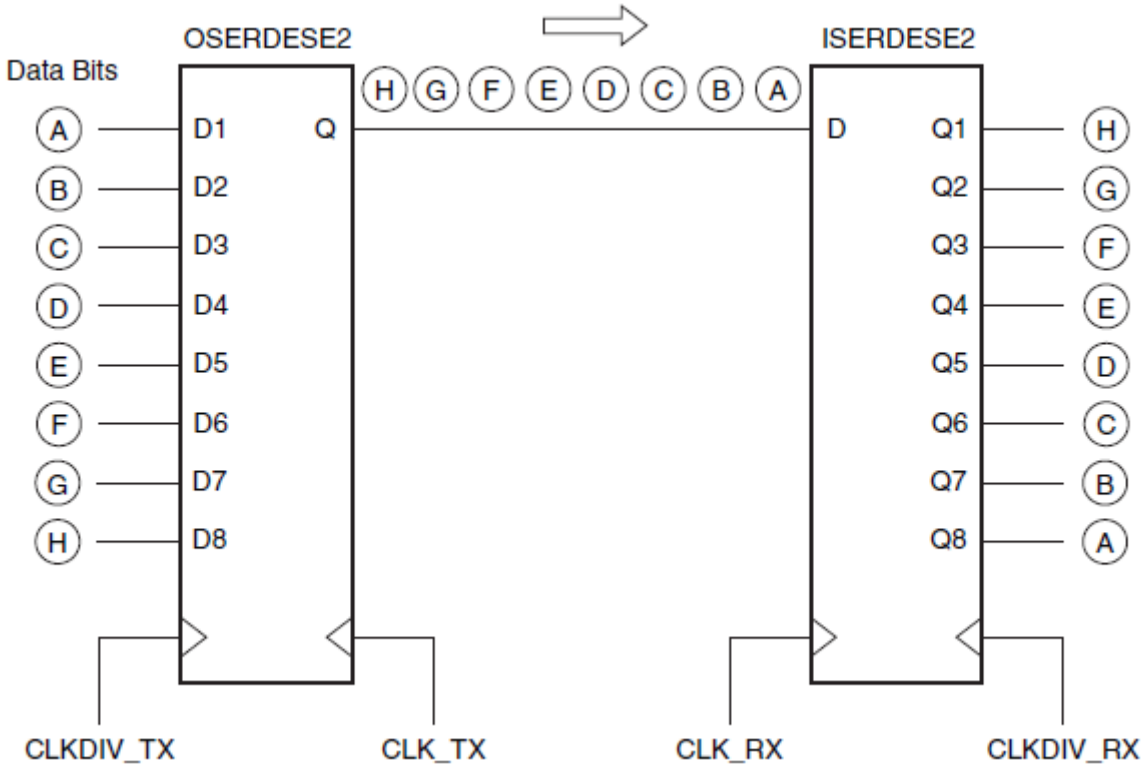
- Input pipeline



2024 © Centre for Development of Advanced Computing

ISERDES3/OSERDES3

- Dedicated support for DDR3, QDR interfaces

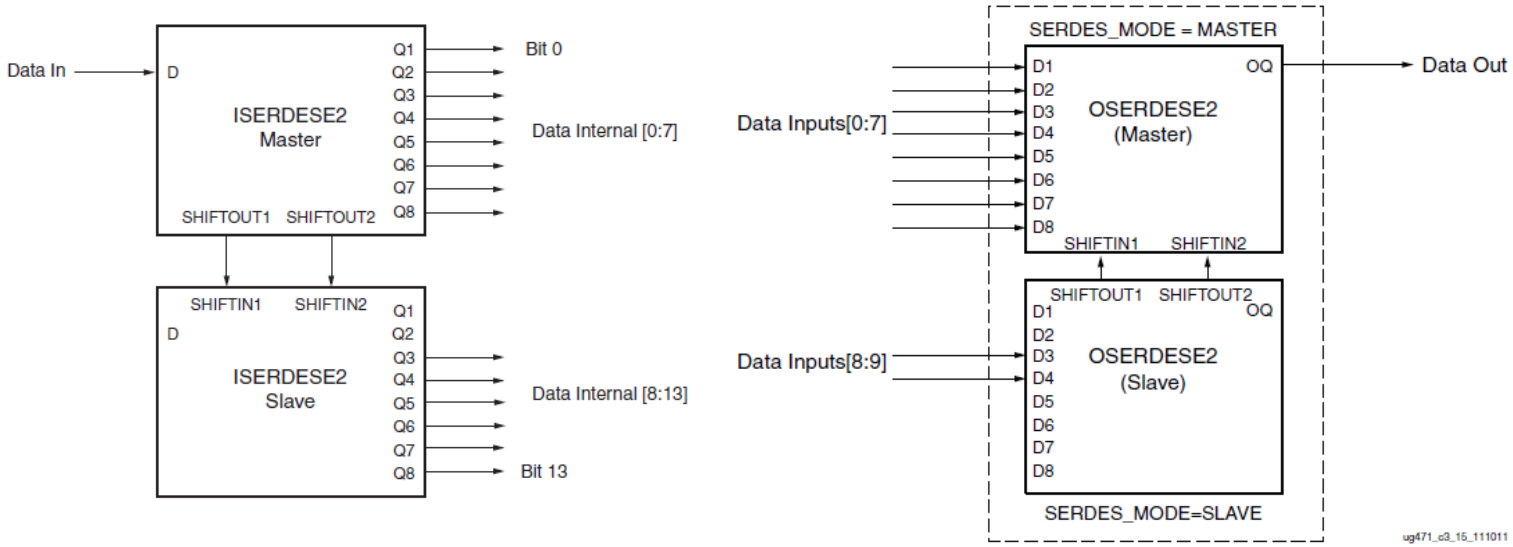


Note: Conceptual model same as ISERDES2/OSERDES2

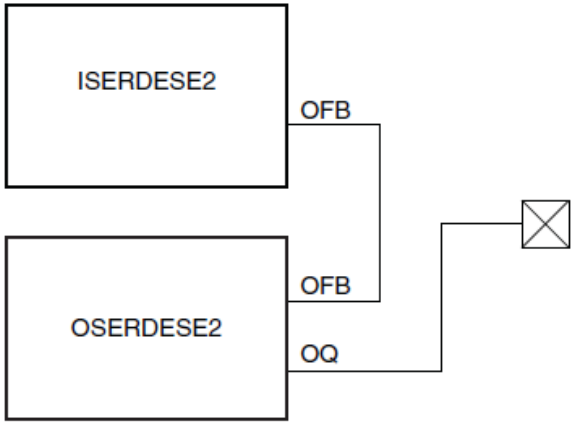
2024 © Centre for Development of Advanced Computing

ISERDES3/OSERDES3

- Expanding the width



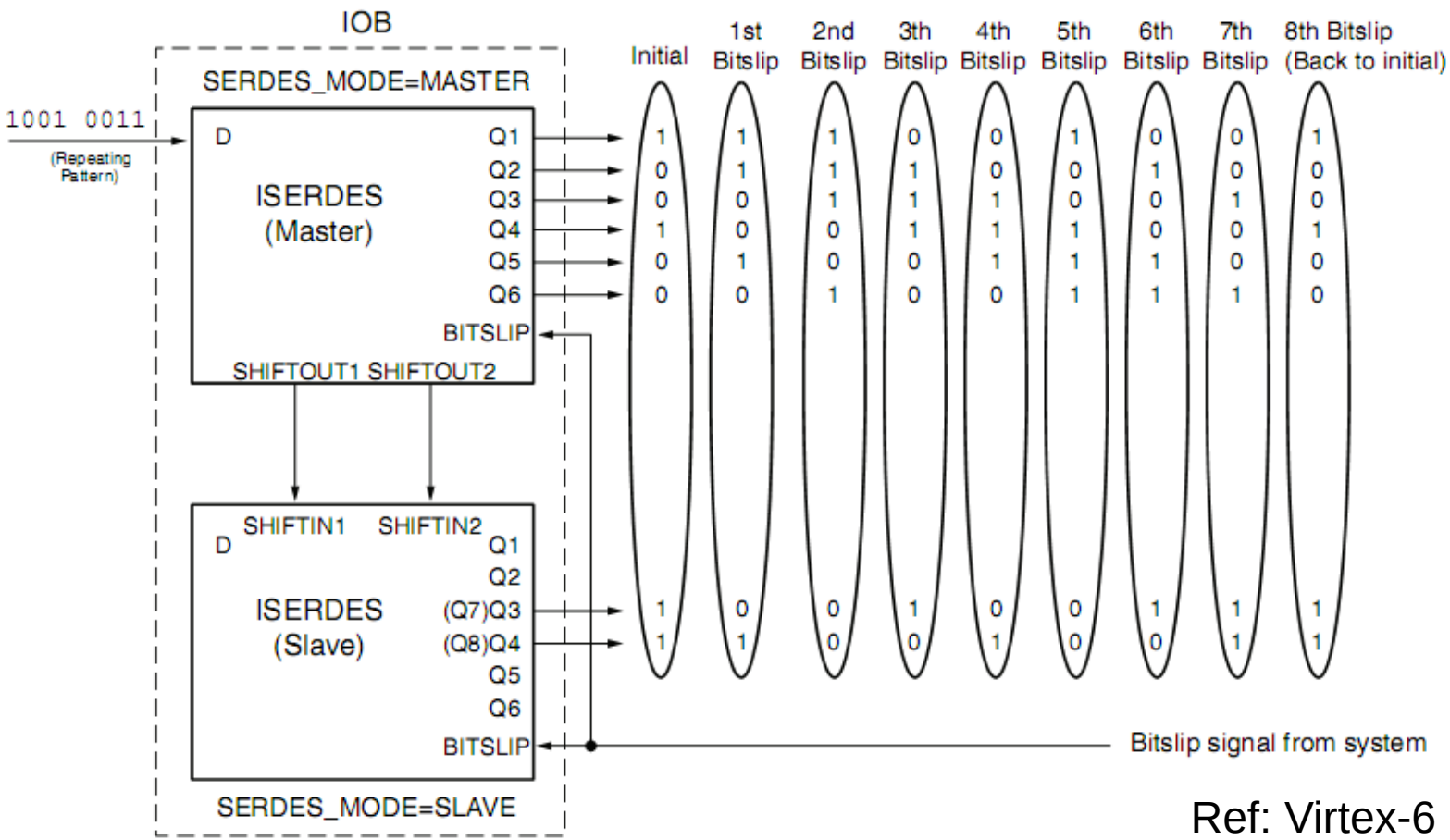
Note: Conceptual model same as ISERDES2/OSERDES2



2024 © Centre for Development of Advanced Computing

ISERDES3/OSERDES3

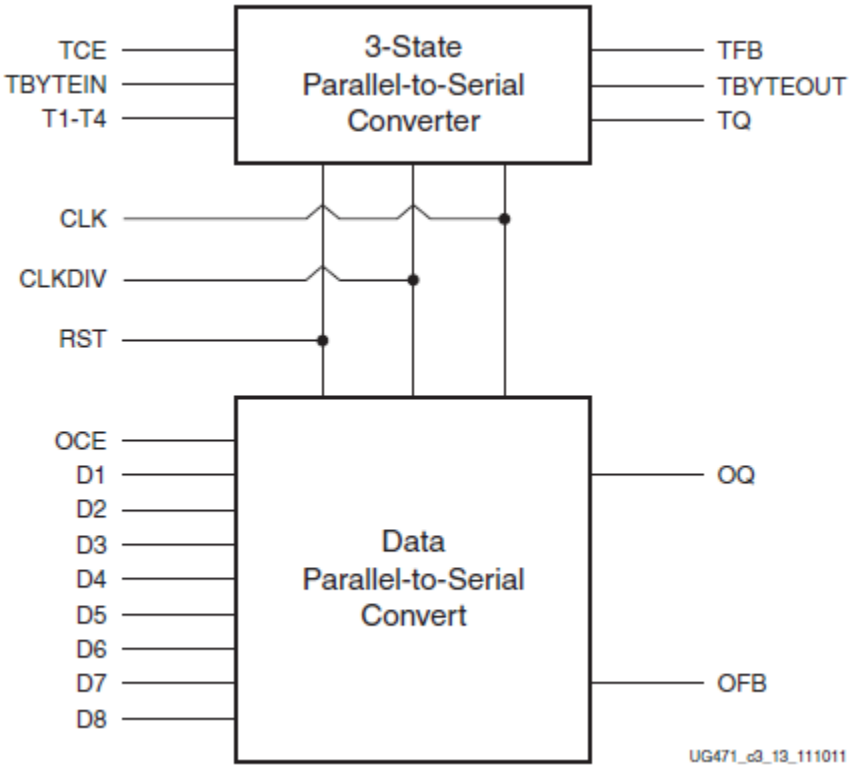
- Using training pattern and bitslip to align data



Ref: Virtex-6

2024 © Centre for Development of Advanced Computing

OSERDES3



2024 © Centre for Development of Advanced Computing

IOB usage

- Most commonly used BUFs are inferred by the synthesis tools
  - Alternatively, basic components are available in form of macros such as ‘IBUF’, ‘OBUF’ and ‘IOBUF’. IOB FFs and DDR macros are likewise available
  - User can use pre-defined macros such as ‘OBUFT_LVCMOS18_S_2’ or give attributes to the basic macros given above
- Number of other IOB resources can be used by instantiating macros, attributes or giving guidelines (constraints) to the backend tool during synthesis.

2024 © Centre for Development of Advanced Computing

Other design considerations related to I/O

- 3.3V, 5V tolerance considerations (latest devices are not 5V tolerant)
  - Important for Legacy device interfacing
- Simultaneous switching outputs (SSOs)
  - Limits number of useable pins for active I/Os
  - Related with ground bounce (Inadequate ground return paths)
  - New devices have better layouts
- Signal integrity and slew rate considerations
  - Important for any fast designs
  - Simulation tools allow pre/post layout validation
- Chip power and bypass considerations
  - Power supply quality is an important factor
  - Multiple inputs, few of which impose stringent requirements on input supply
- Hot-swap capability
- Power dissipation
  - Thermal dissipation within limits determined by system design

2024 © Centre for Development of Advanced Computing

Configurable Logic Block (CLB)

2024 © Centre for Development of Advanced Computing

Configurable Logic Block (CLB)

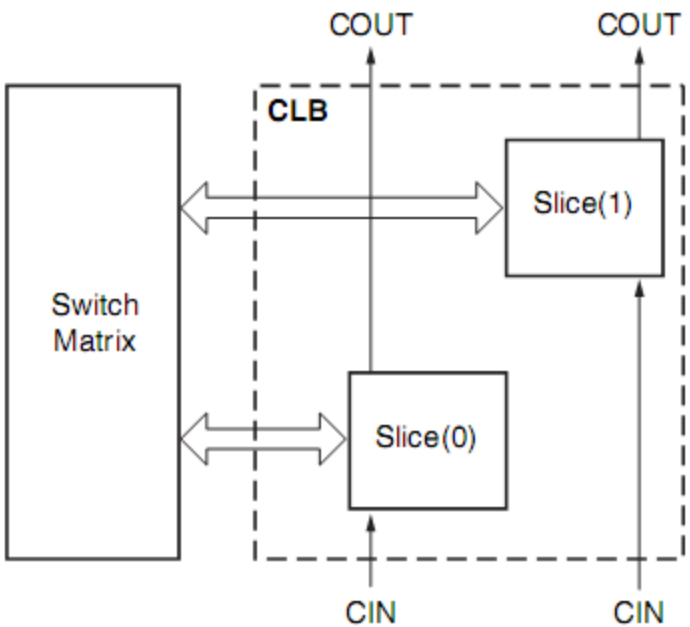
- UltraSCALE configurable logic blocks (CLBs) are arrays of groups of two ‘slices’ which are used to implement combinatorial and sequential functions
- Number of slices in a given device depends upon its density. There are both I/O rich and logic rich variants
- To the backend tools, any device appears as a resource of MxN slices (as far as logic design is considered). This makes design portable easily across devices
  - In case of ‘RPM’s, this may not be the case since a given RPM may be required to span across PxQ slices
- Individual CLBs are connected to the “switch matrix’ as well as fast interconnects to near neighbors

Table 2-1: **Logic Resources in One CLB Slice**

CLB Slice	LUTs	Flip-Flops	Arithmetic and Carry Chains	Wide Multiplexers	Distributed RAM	Shift Registers
SLICEL	8	16	1	F7, F8, F9	N/A	N/A
SLICEM	8	16	1	F7, F8, F9	512 bits	256 bits

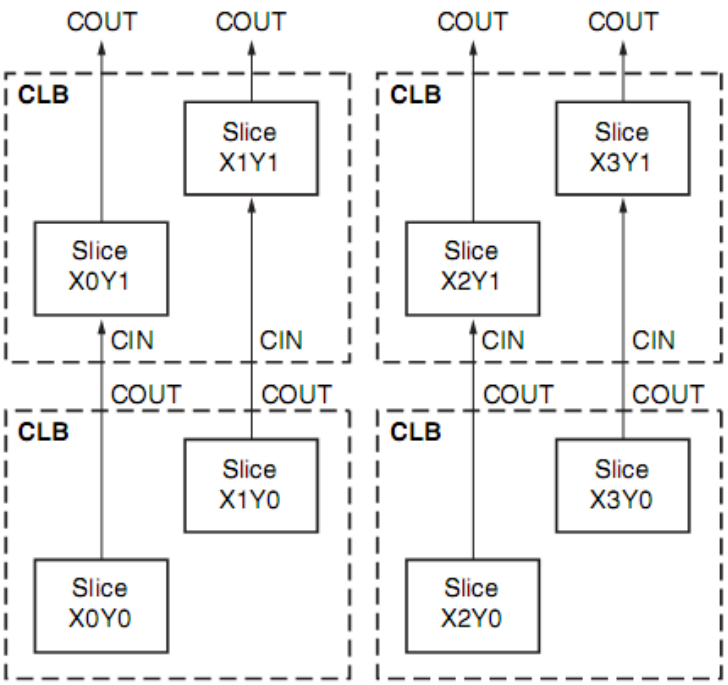
2024 © Centre for Development of Advanced Computing

CLB



Slices within CLB

CLB-Slice rows-columns



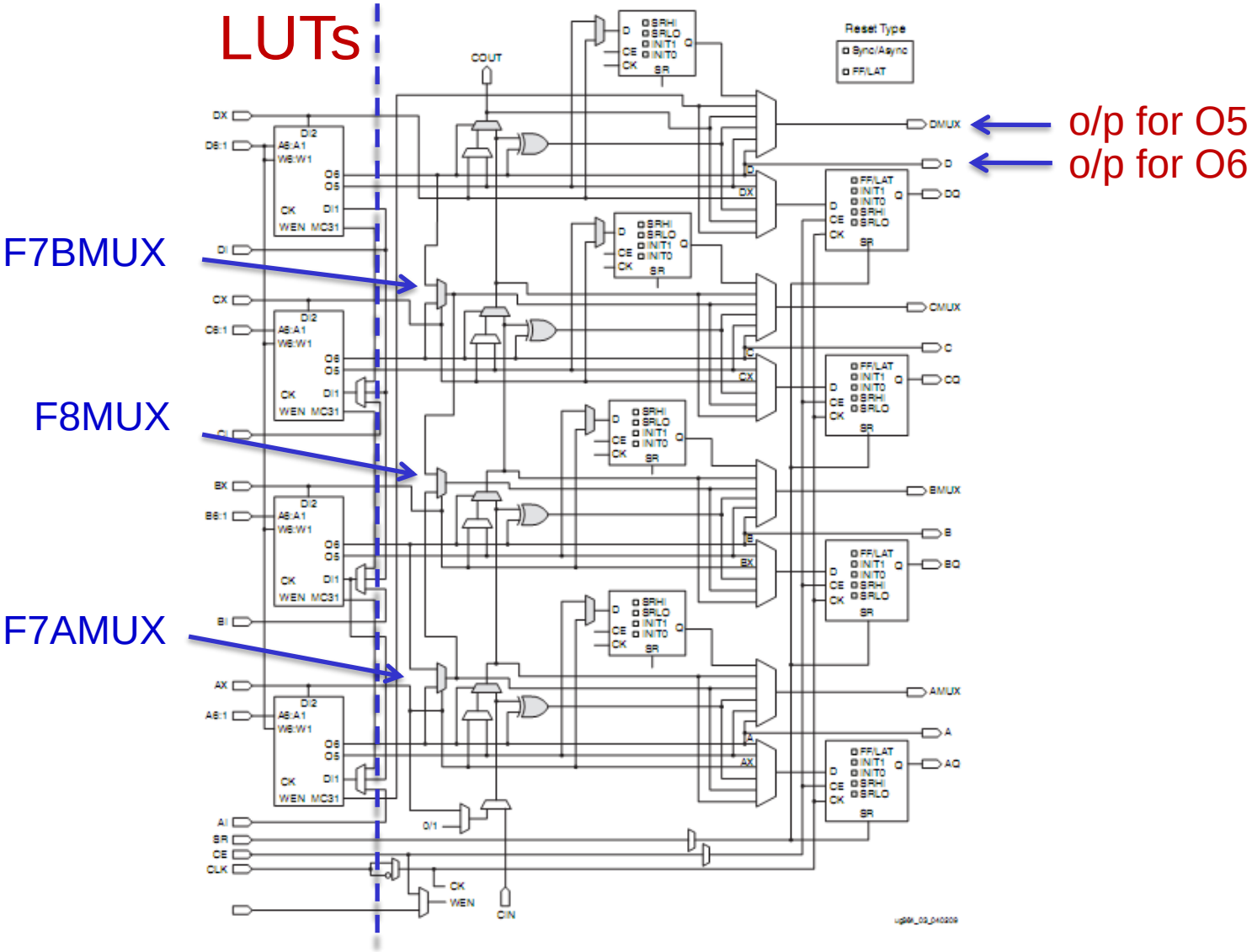
2024 © Centre for Development of Advanced Computing

Slice description

- Each slice (SLICEL) consists of the following:
  - Four 6-input function generators which can act as:
    - 6-Input LUT, dual 5-input LUT
    - 64x1 bit ROM
  - Eight storage elements configurable as FFs or Latch
  - Wide function MUXes to provide additional functionality
  - Cascading outputs for logic and arithmetic functions (Carry logic)
  - Some slices (SLICEM) support additional functions like distributed RAM, 32-bit shift register
    - 64,32-bit memory (RAM or dual port RAM)
    - 32-bit, or dual 16-bit variable tap shift register
  - Four MUXes (F7MUX, F8MUX and F9MUX)

2024 © Centre for Development of Advanced Computing

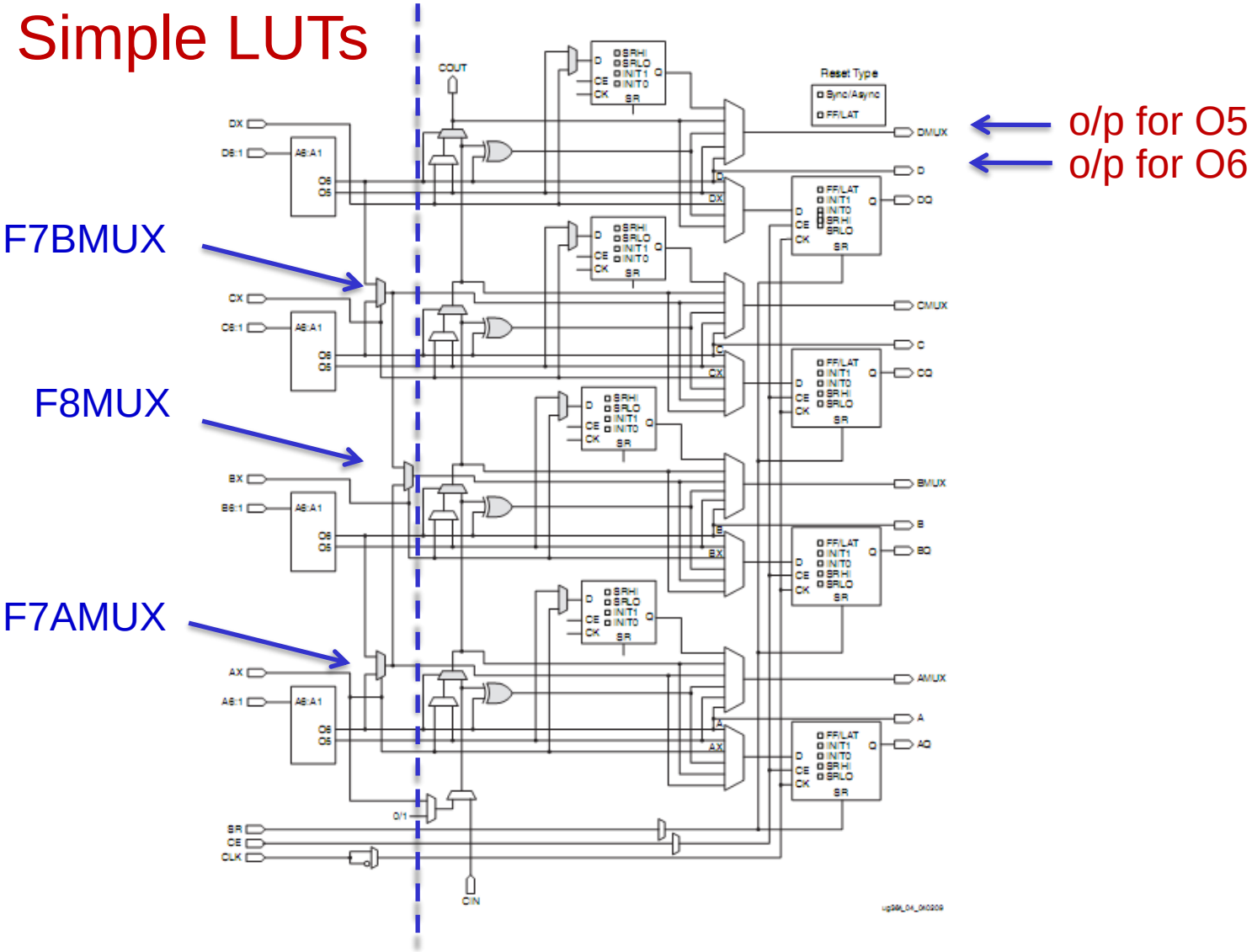
SLICEM



2024 © Centre for Development of Advanced Computing

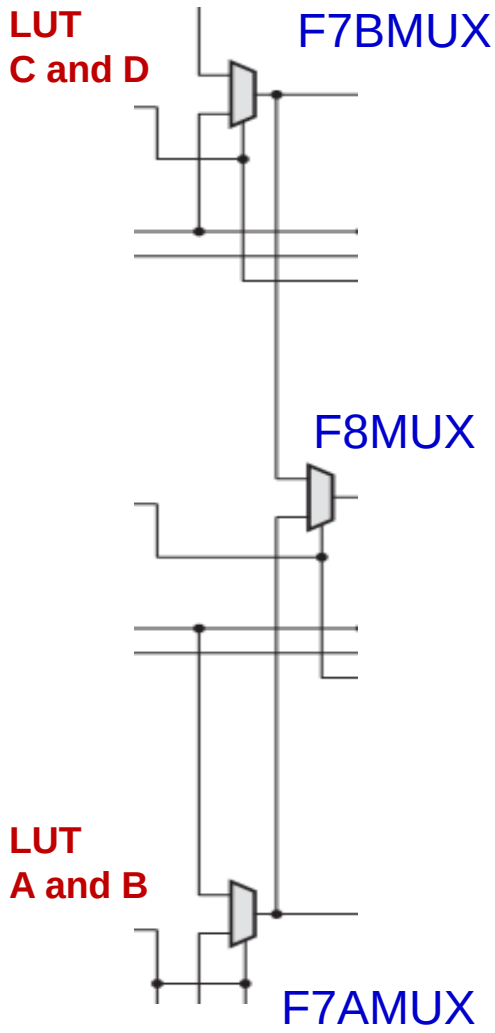
SLICEL

Simple LUTs



2024 © Centre for Development of Advanced Computing

Wide MUXes



- F7AMUX, F7BMUX and F8MUX are used to combine upto four LUTs to implement any function of seven or eight inputs in a slice
- F7AMUX and F7BMUX are used to generate seven input functions from LUTs A and B, or C and D
- F8MUX is used to combine all LUTs to generate eight input functions
- Functions with more than eight inputs can be implemented using multiple slices
- It is not possible to implement functions with more than eight inputs within a single CLB as there are no direct connections between slices
- Similarly F9 MUX combines nine input function. Not shown here.

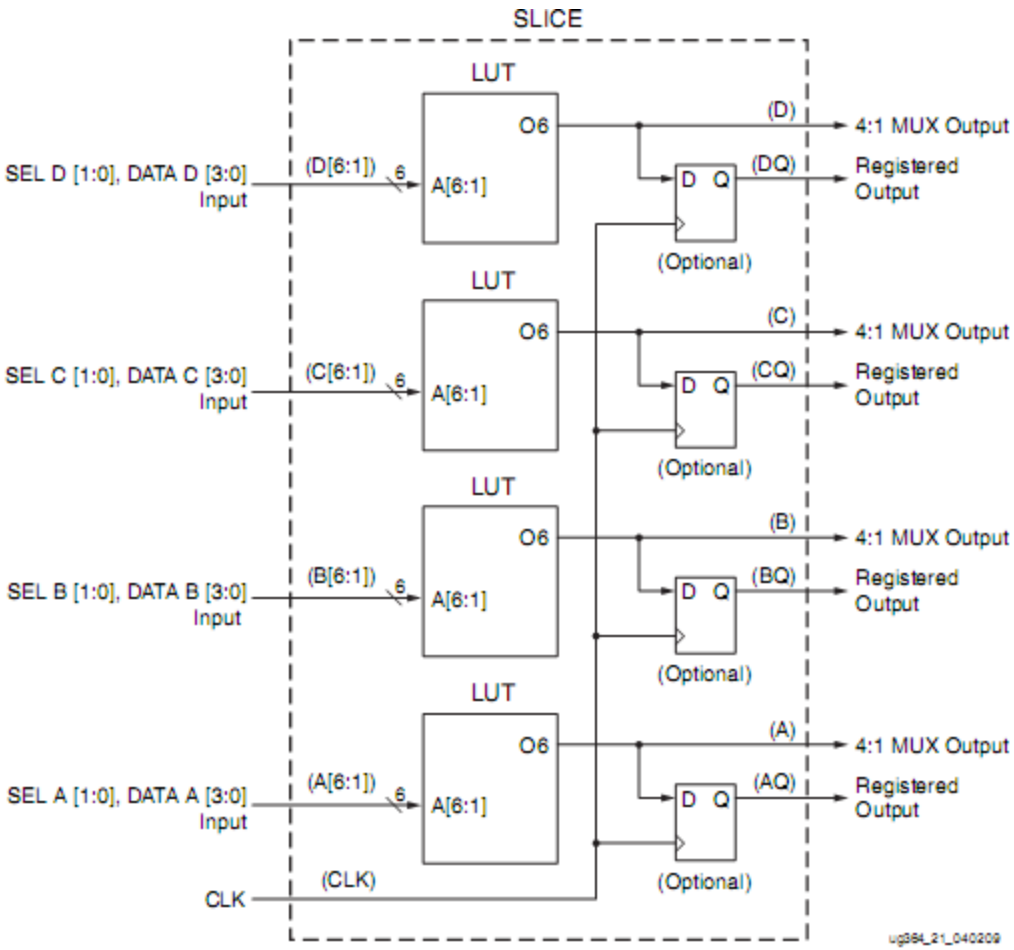
2024 © Centre for Development of Advanced Computing

CLB multiplexers

- MUXes with each CLB can be used for the following:
  - 4:1 MUX in one LUT
  - 8:1 MUX in two LUTs
  - 16:1 MUX in four LUTs
  - 32:1 MUX in eight LUTs

2024 © Centre for Development of Advanced Computing

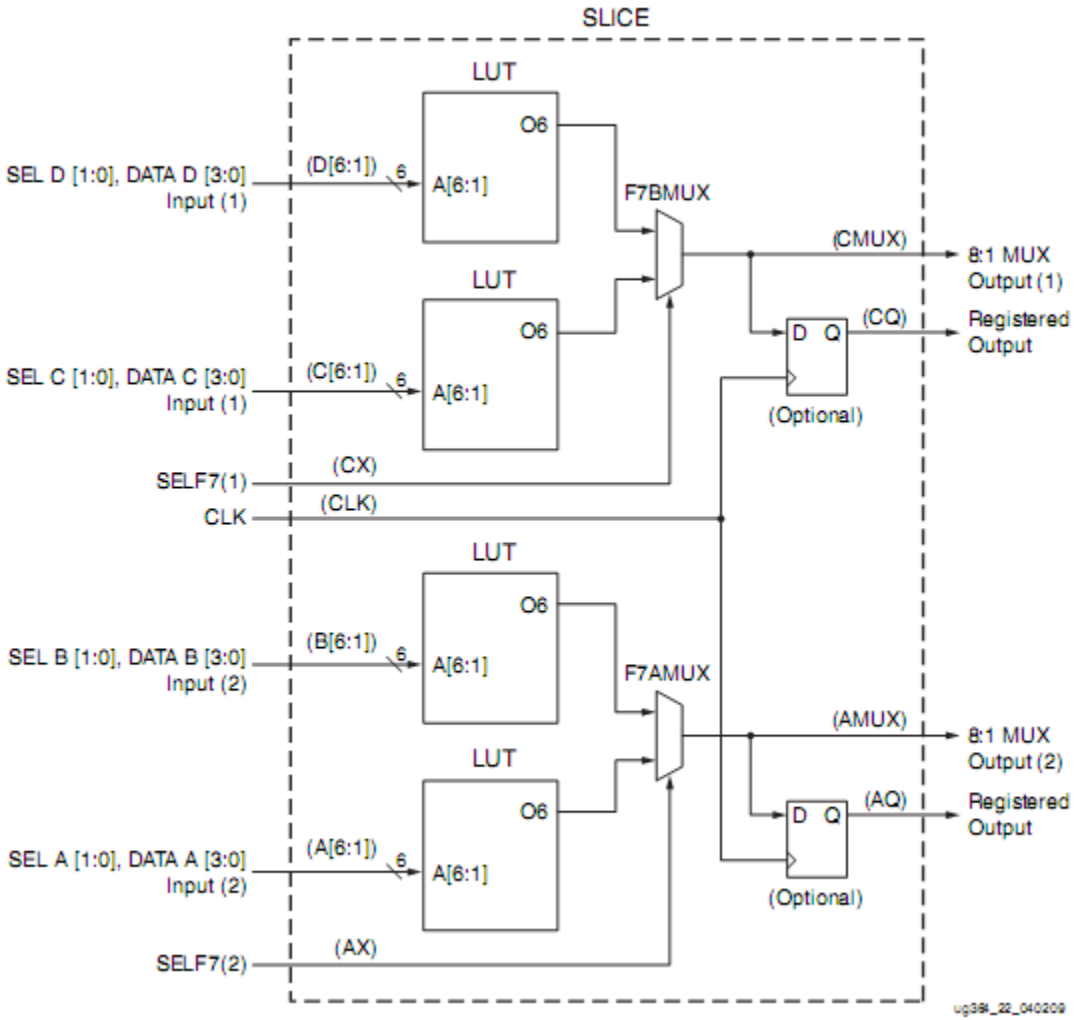
CLB MUXes



Four 4:1 MUX in a slice

2024 © Centre for Development of Advanced Computing

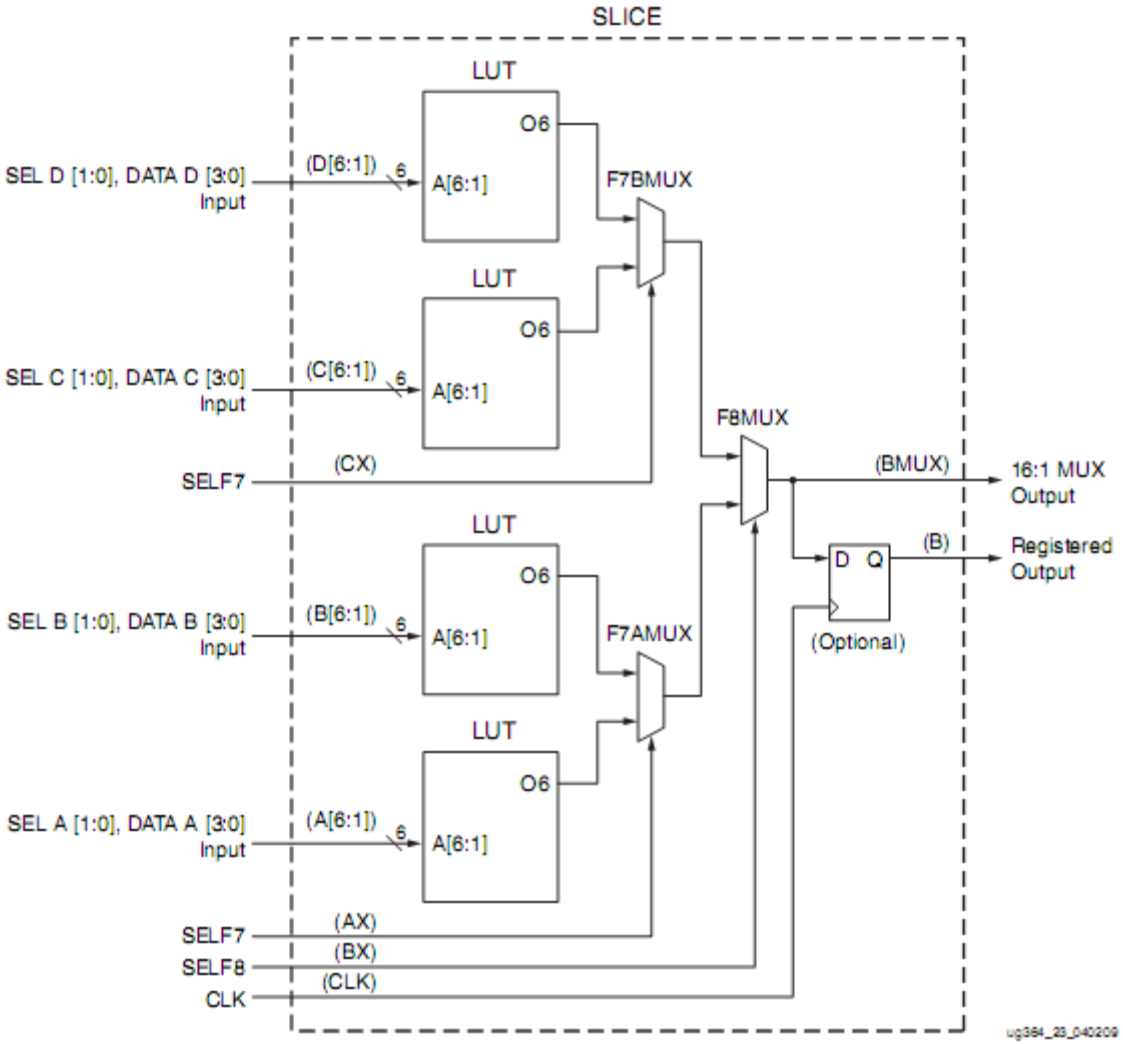
CLB MUXes



Two 8:1 MUX in a slice

2024 © Centre for Development of Advanced Computing

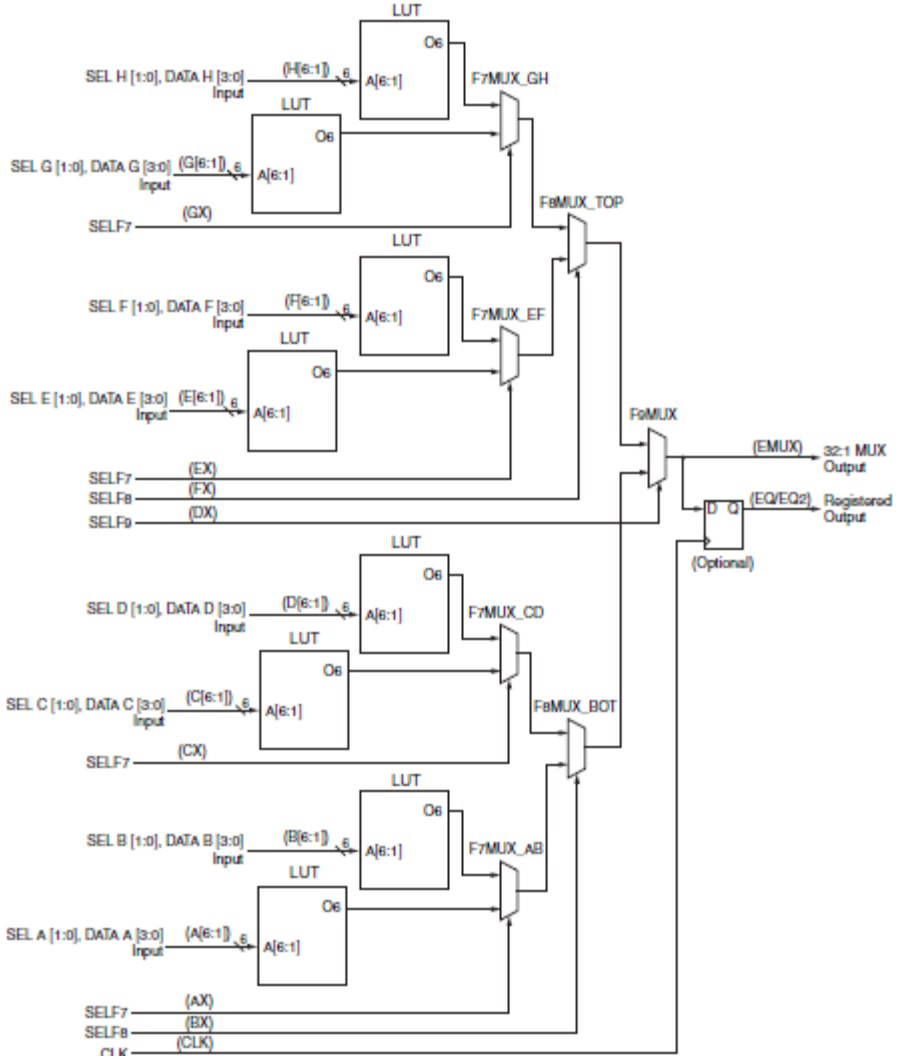
CLB MUXes



16:1 MUX in a slice

2024 © Centre for Development of Advanced Computing

CLB MUXes



32:1 MUX in a slice

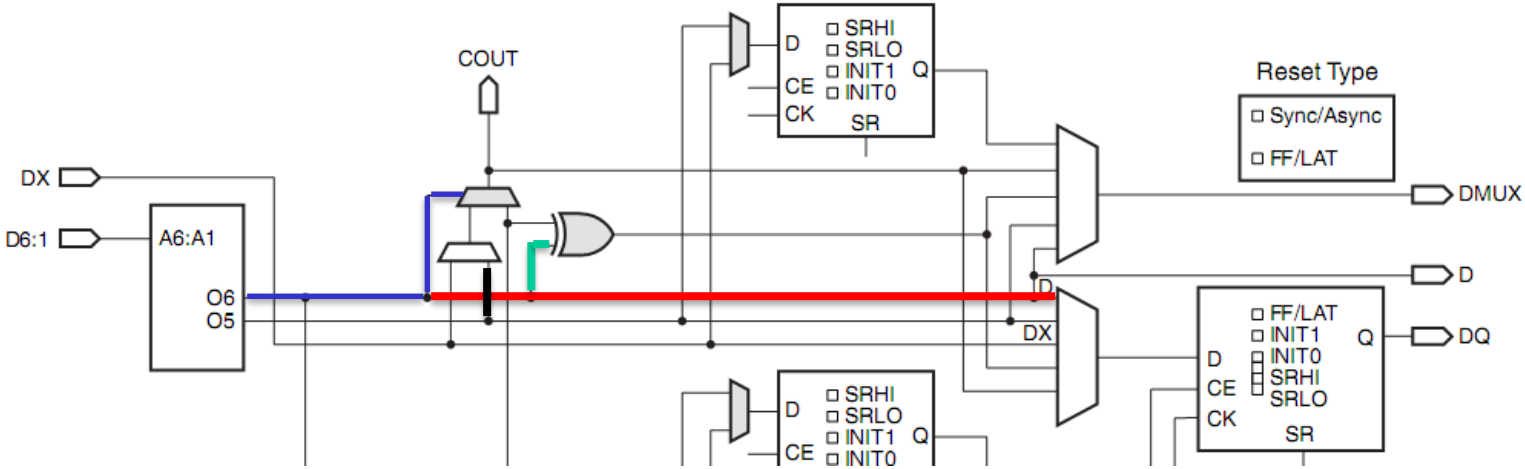
2024 © Centre for Development of Advanced Computing

LUT

- LUTs are 6-input lookup tables (A1-A6). These can be used to implement any 6-variable logic function. Basically, LUT acts as a PROM storing output values for each of the address selections. The values in ROM are loaded during FPGA configuration.
  - Advantage of ROM approach is its simplicity and predictable timing advantage
- Each LUT can be configured to function as two, 5-input LUTs where inputs are common. A6 is tied high by software. Both O5 and O6 are used in this case
- Propagation delay is independent of function implemented or whether LUT is used as single 6-input or dual 5-input LUT
- No direct connections between slices

2024 © Centre for Development of Advanced Computing

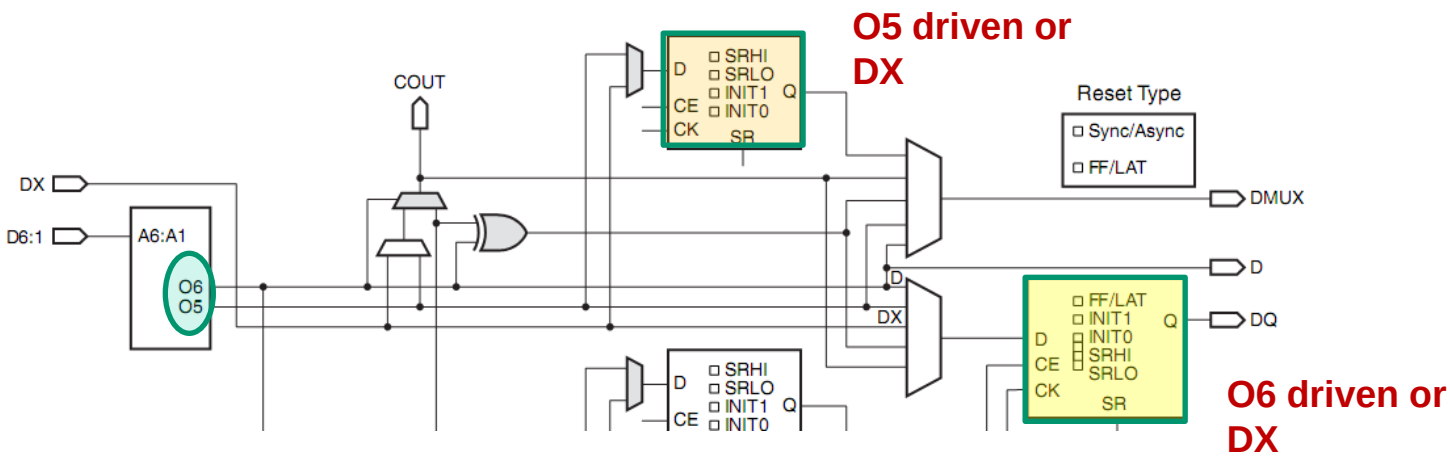
LUT o/p path



- O5, O6 exit slice via
  - Either A,B,C,D for O6 or AMUX, BMUX, CMUX, DMUX for O5
  - O6 to XOR dedicated gate
  - O5 to carry chain
  - O6 to carry chain MUX select line
  - D input of register
  - O6 to F7AMUX/F7BMUX

2024 © Centre for Development of Advanced Computing

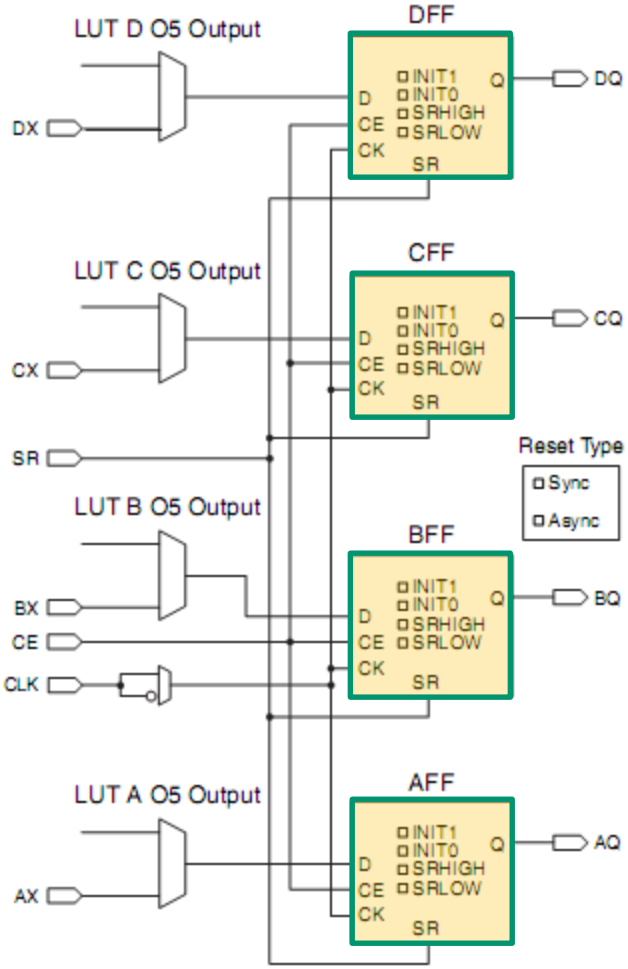
Slice storage



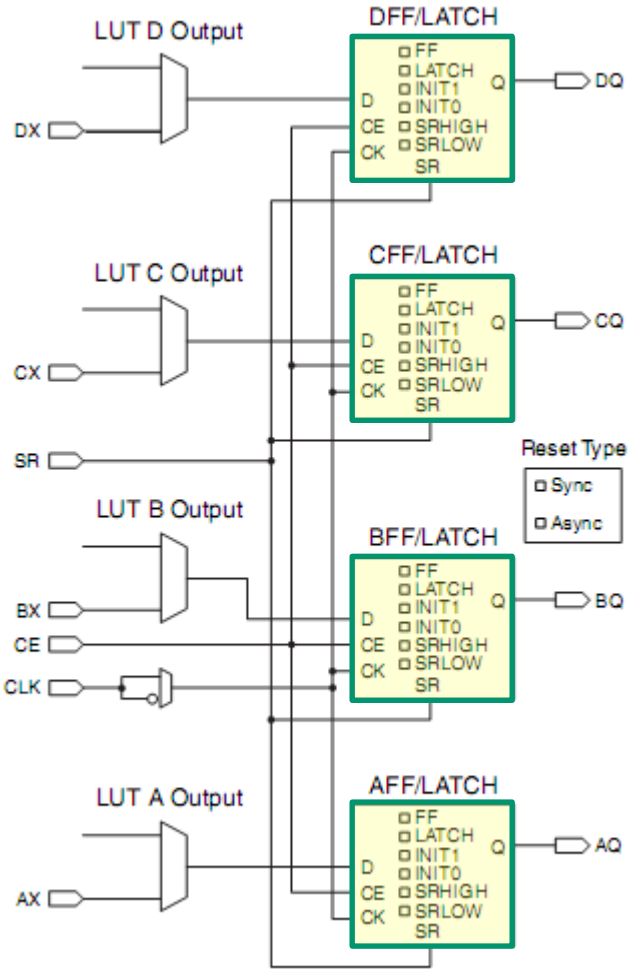
- Two sets of 4 storage elements
- O5 driven elements are edge triggered only
- O6 driven elements can be FFs or Latch
- Input : local (A-D) or remote (AX-DX)
- FF mode has very useful CE function
- Has set and reset inputs (sync or asynch)
- In addition, has a global set/reset capability using a dedicated chip resource
- Programmable polarity for CE,CLK

2024 © Centre for Development of Advanced Computing

Slice storage



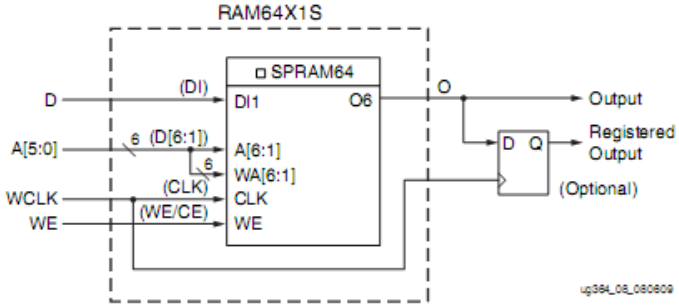
O5 driven or by DX



O6 driven or by DX

2024 © Centre for Development of Advanced Computing

Slice as memory



- CLBs with SLICEM can be used as memory. Such memory is called as ‘distributed memory’
- Each LUT is 64X1 bit sync RAM. This can be configured as ROM, SPRAM, dual port, quad port RAM. In case of dual port memory, density gets halved.
- Memory is synch write, sync/async read
- Multiple slices used to implement wider (e.g. 16X8) and deeper (e.g. 128X1) memory blocks.
- Distributed memory blocks serve as extremely fast (but small) memory elements. To be used judicially in design since precious LUT resources are consumed

2024 © Centre for Development of Advanced Computing

Slice memory configurations

- SP: Common address for async read and sync write
- DP: One port sync write/async read and other for async read
- SDP: One port sync write, other async read
- Quad port: One port sync write/async read and three additional ports for async reads

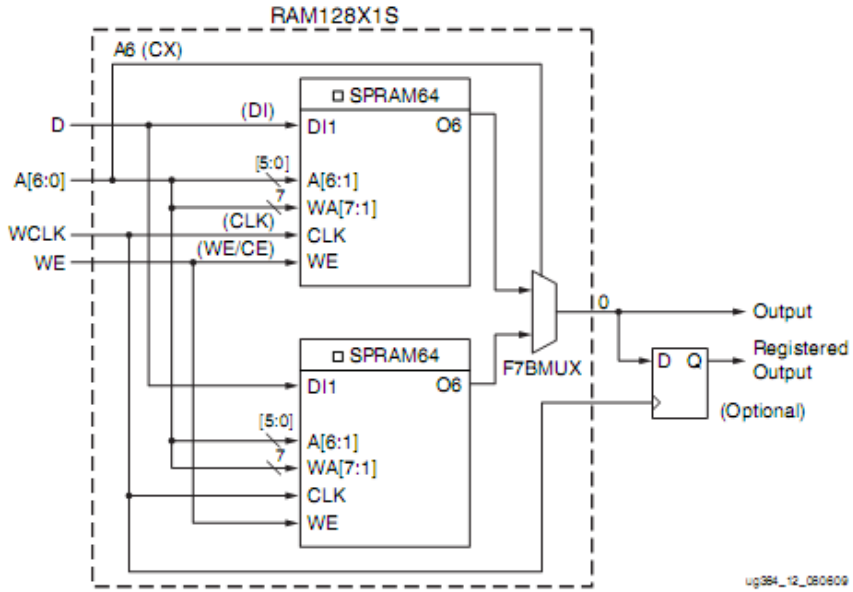
Note: If sync read is required, designer can use slice FFs at output side with additional latency getting introduced.

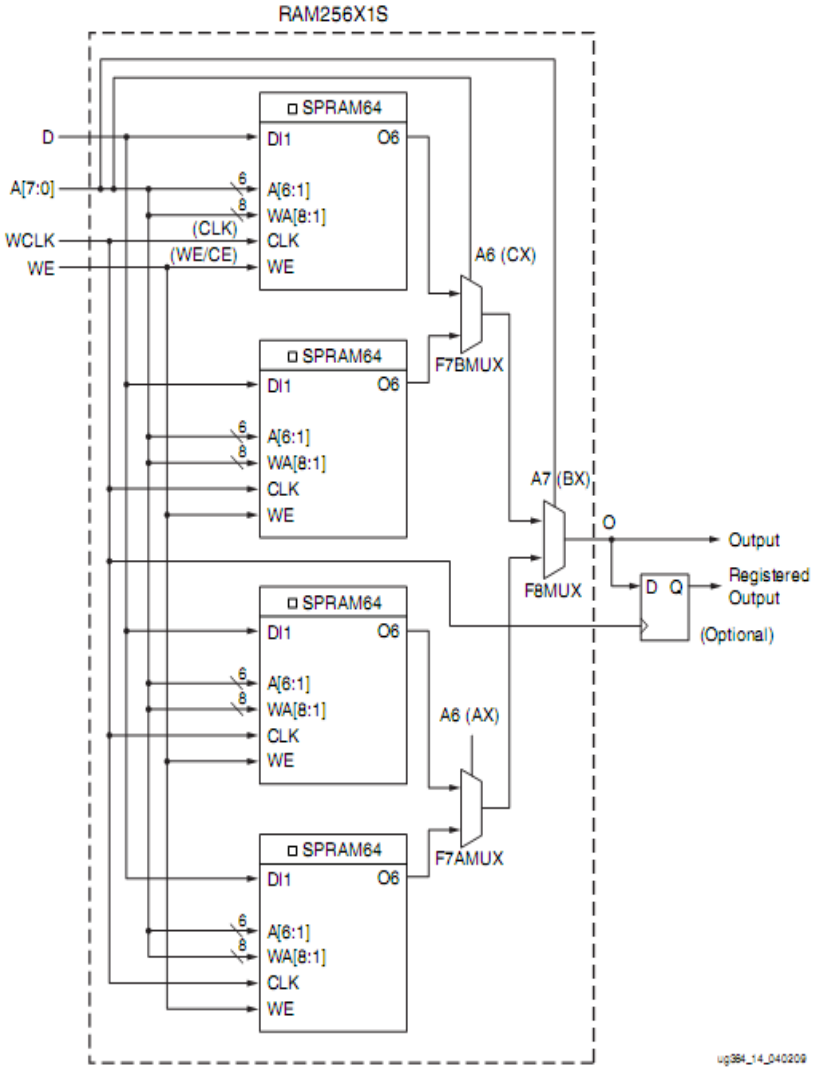
- ROM: Each LUT in SLICEM/L can be used as ROM. Contents of ROM are loaded during configuration

ROM	Number of LUTs
64 x 1	1
128 x 1	2
256 x 1	4
512 x 1	8

2024 © Centre for Development of Advanced Computing

Slice memory (SP)

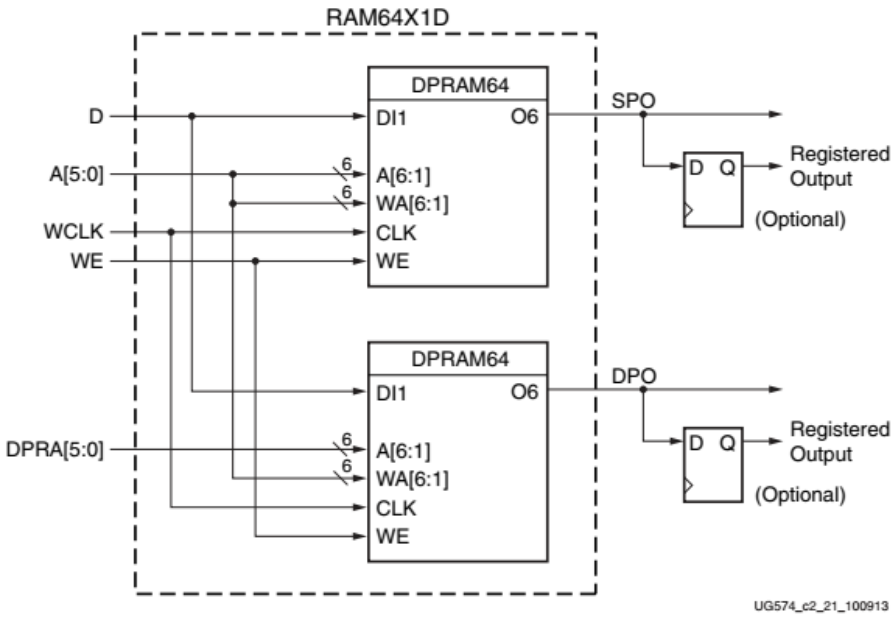




A[6:0] is read address, A6 connected to output F7BMUX
WA[7:1] is write address common to read address
WE: write enable
Memory o/p can be sync/asynch

2024 © Centre for Development of Advanced Computing

Slice memory (DP)



DPRA[5:0]: dual port read address

Dual port memory: two ports for Read only.

Write done using A[5:0],D and WE

DP functionality achieved using duplicate resources in the form of 64X1RAM.

Read/write port

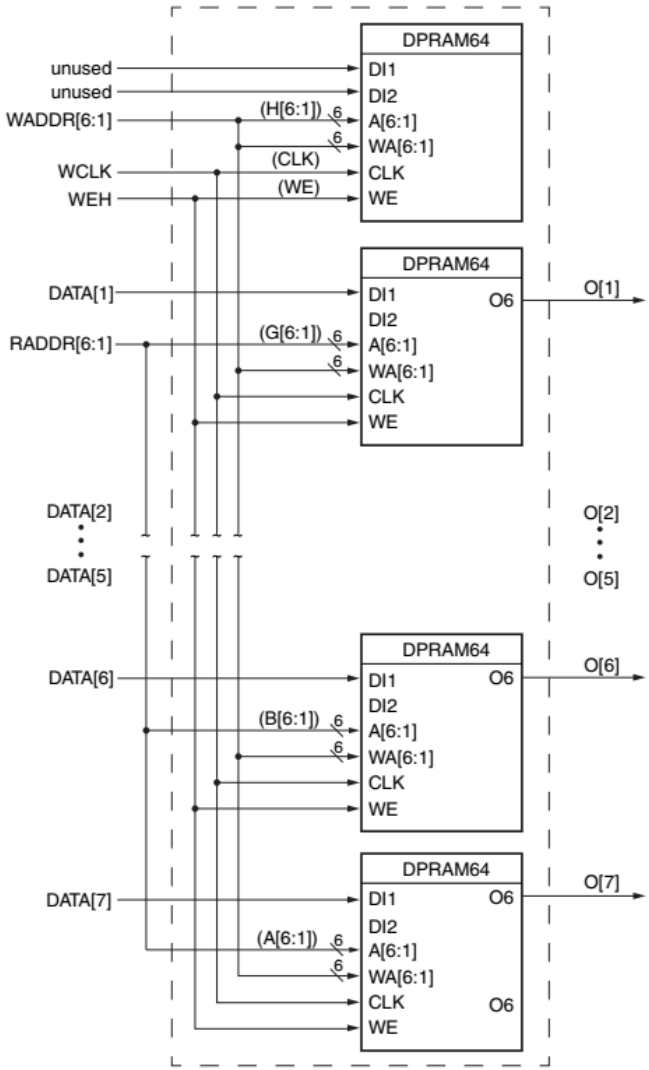
A[5:0], WE, WCLK, D, SPO

Read port

DPRA[3:0],DPO

2024 © Centre for Development of Advanced Computing

Slice memory (SDP)



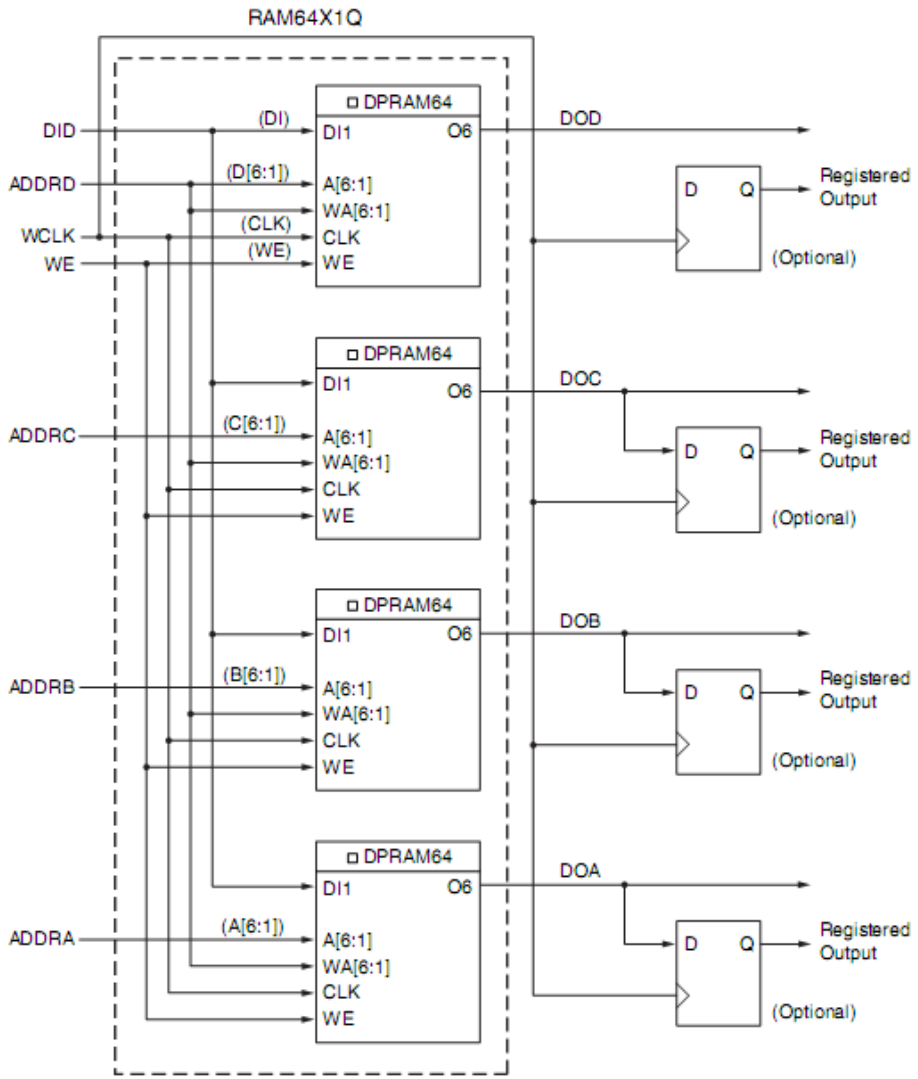
RADDR[6:1]: simple dual port read address

SDP memory: one port for Read only.

Write done using WADDR[6:1],D and WE

2024 © Centre for Development of Advanced Computing

Slice memory (Quad port)



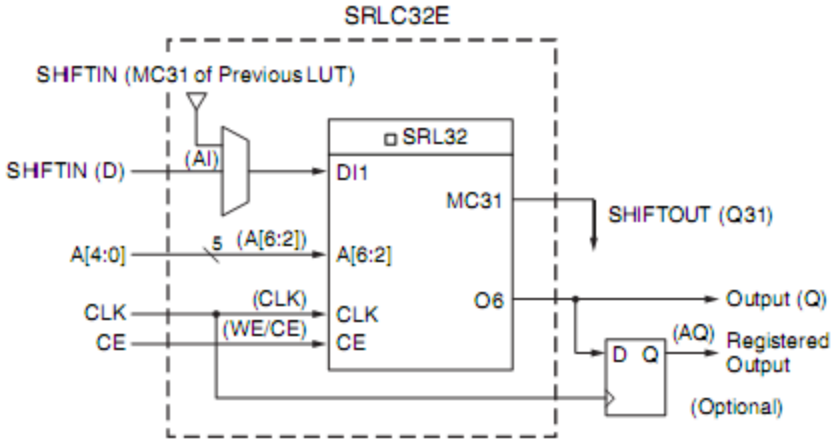
2024 © Centre for Development of Advanced Computing

Shift registers

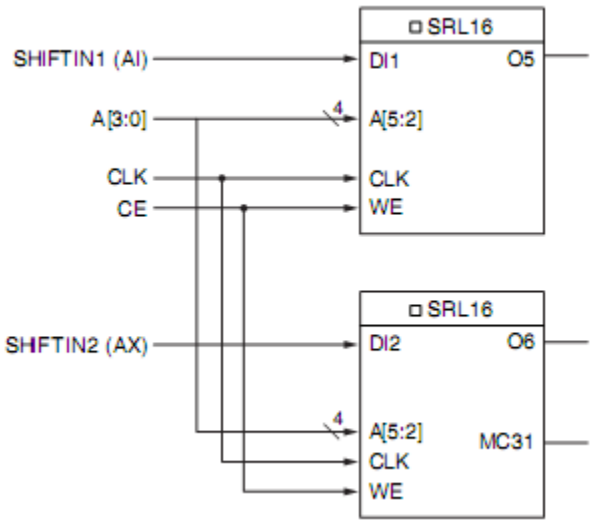
- Shift registers are a powerful design resource to implement innovative implementations for
  - Dividers
  - Psuedo counters for FIFOs
  - Psuedo random bit generators
  - DSP related algorithms
- SLICEM LUTs can be configured as 32-bit shift registers. Dedicated routing resources are available to realize bigger SRs with a given SLICEM

2024 © Centre for Development of Advanced Computing

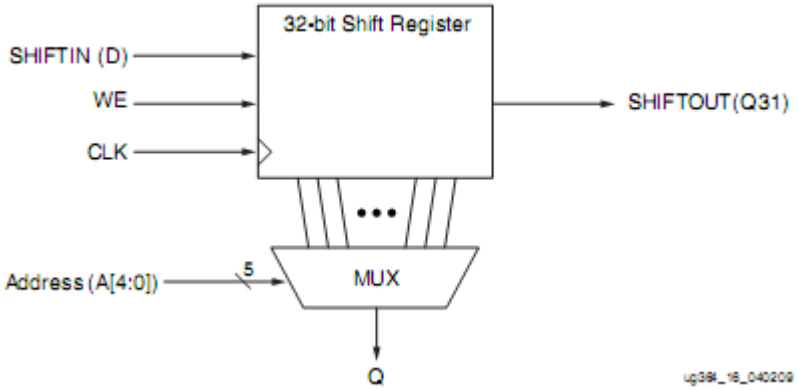
Shift regs



Single 32-bit SR (one LUT)



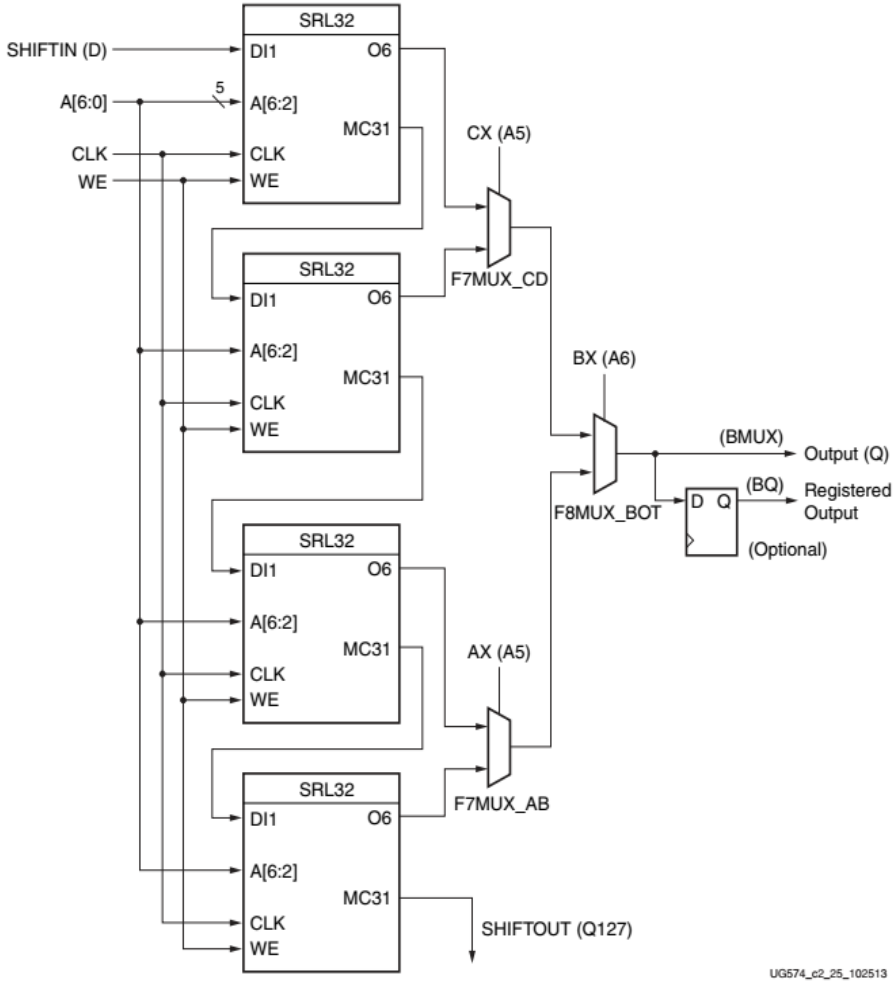
Dual 16-bit SR (one LUT)



- Synch write, async or sync read
- SHIFTOUT and SHIFTIN allow for cascading
- Address used to take output at desired shiftout tap

2024 © Centre for Development of Advanced Computing

Cascaded shift registers



- 128-bit SR can be built using single SLICEM
- SR longer than 128-bit require more slices. Direct connection to other slices is however not available

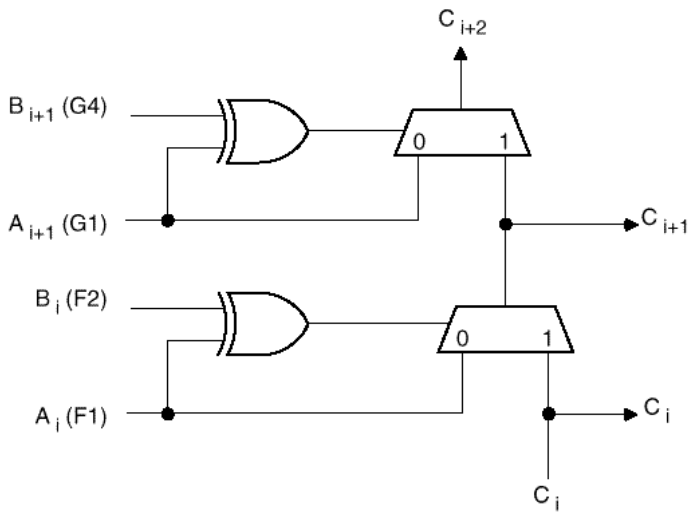
2024 © Centre for Development of Advanced Computing

CLB: arithmetic carry chains

- For addition and subtraction functions. A slice can be used to implement a 4-bit full adder
  - Sum = A exor B
  - CY = A.B
- Carry chains run upwards using Cin and Cout resources
  - Carry chain is required to ‘ripple’ thru many levels of logic when large vectors are used. They often define the critical path for arithmetic operations and need to be fast
- Carry chains can also be used to implement wide logic functions

2024 © Centre for Development of Advanced Computing

Carry Chain example (add)

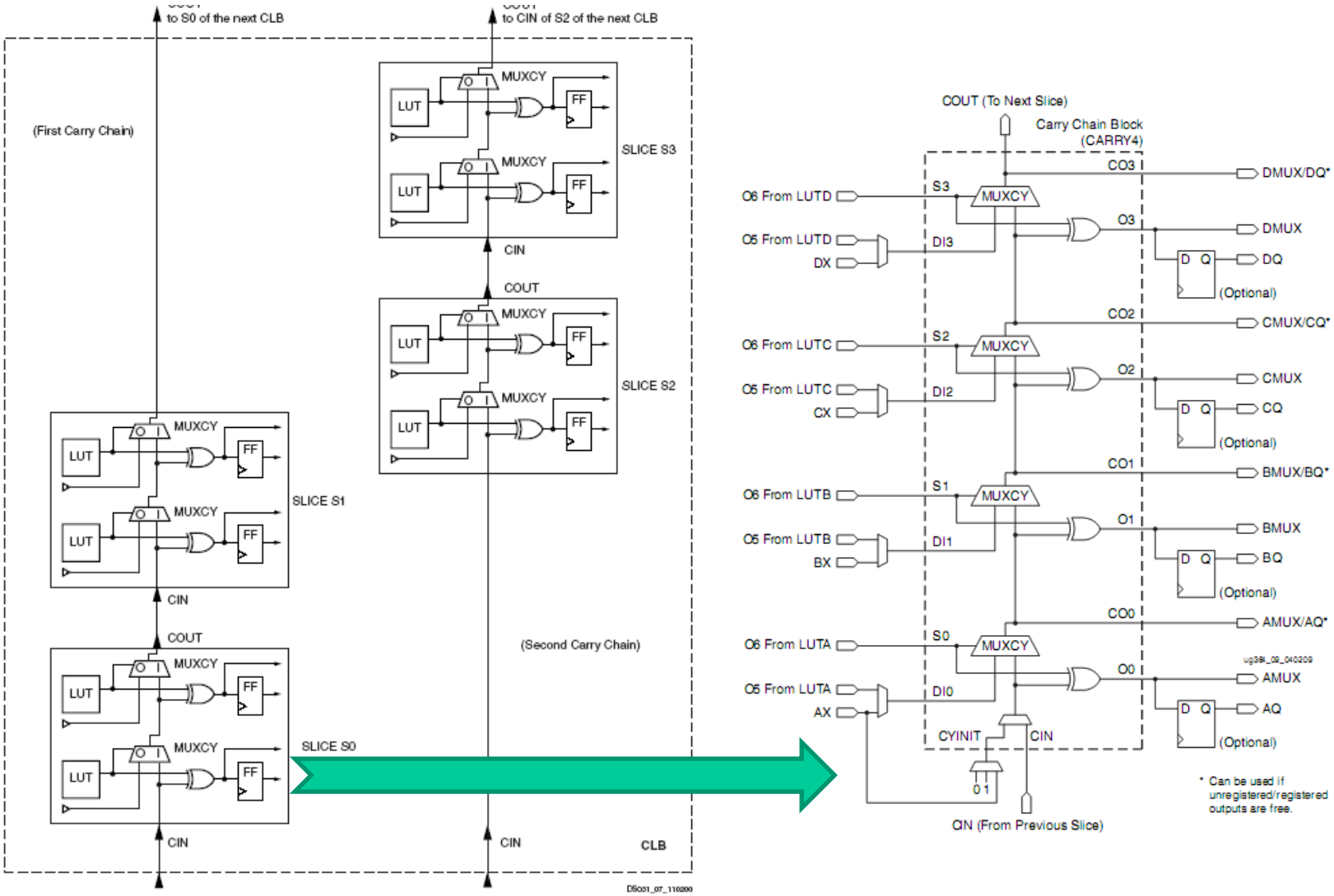


A	B	O/P	Ci+1	action
0	0	0	0	block
0	1	1	Ci	propagate
1	0	1	Ci	propagate
1	1	0	1	generate

If SUM(A,B) generates local carry, it needs to be propagated.
In worst case, carry will propagate thru the entire bit vector C0..Cn

2024 © Centre for Development of Advanced Computing

Carry Chains

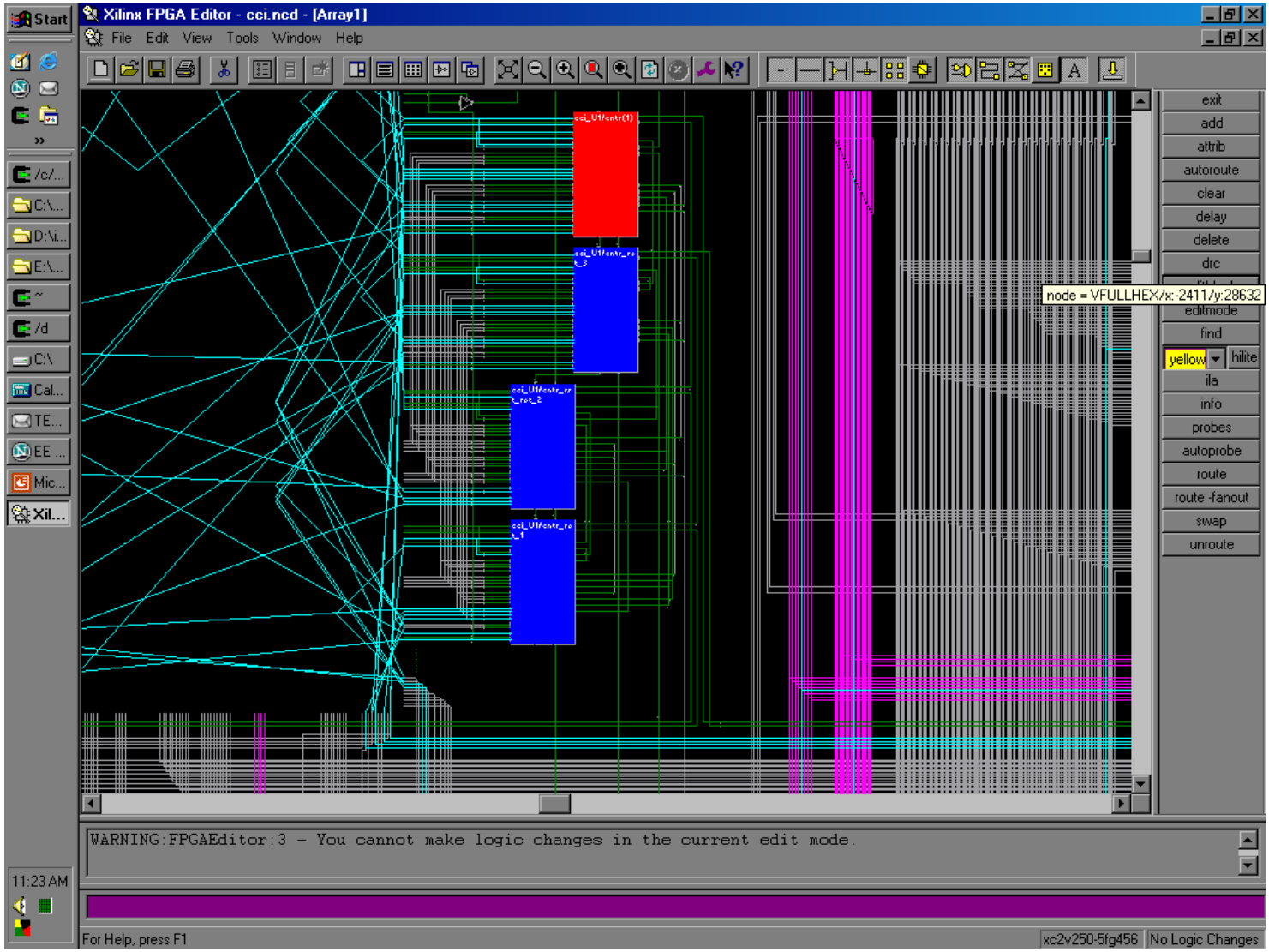


2024 © Centre for Development of Advanced Computing

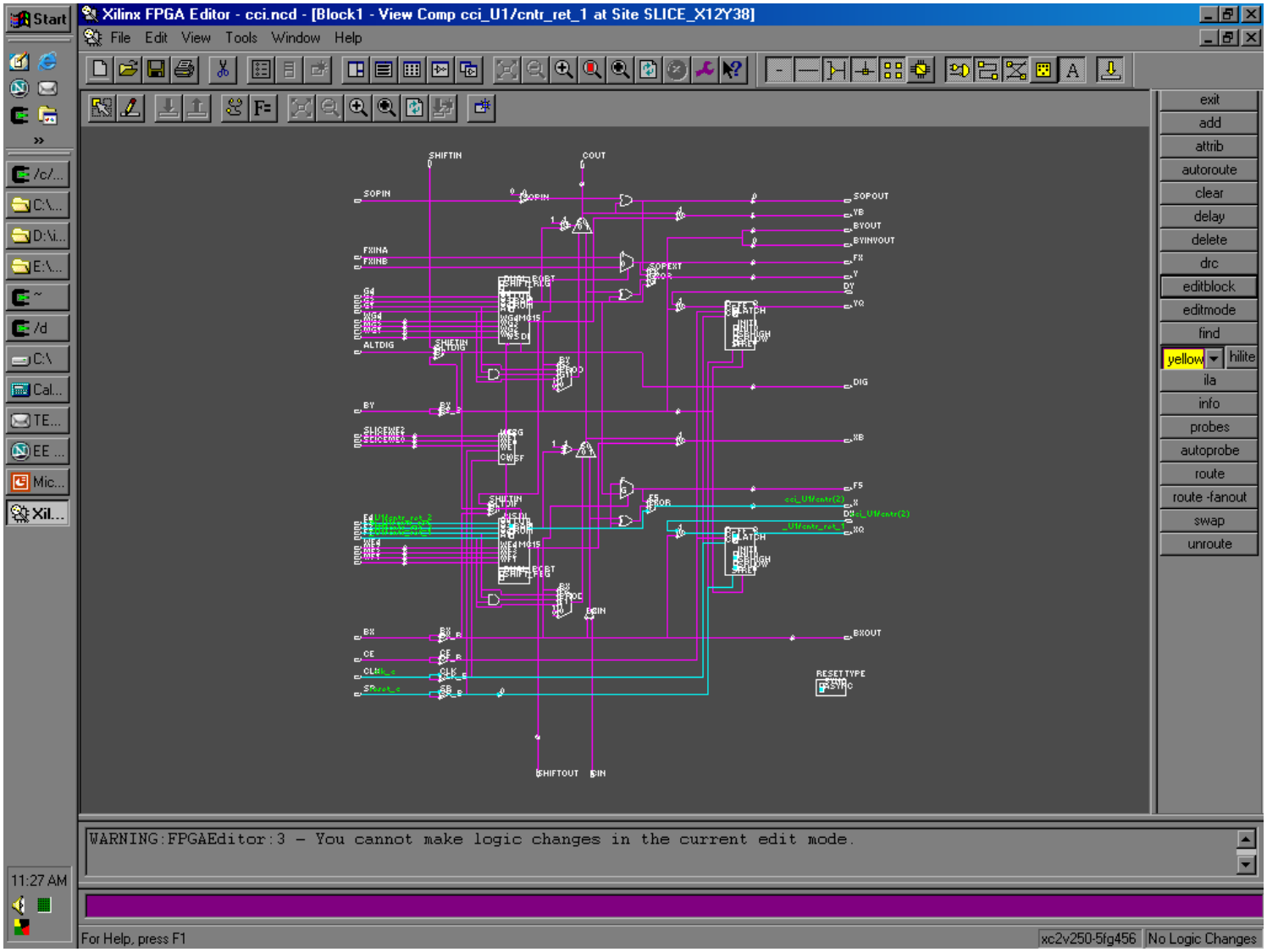
CLB

- Why should I care if architecture details are hidden by synthesis and backend tools?
  - For simple designs this is true. But when designs are stretched to limits (area and/or timing), hand optimizations considering device architecture may become necessary.
  - Optimal designs result when architectural aspects are considered during RTL stage itself
  - Macro designs require low level directives to be passed to backend tools
  - Innovative designs are possible using raw resources CLBs offer. E.g. FIFOs using SRs, delay lines using CLBs etc

2024 © Centre for Development of Advanced Computing



2024 © Centre for Development of Advanced Computing



2024 © Centre for Development of Advanced Computing

BlockRAM and DSP block

2024 © Centre for Development of Advanced Computing

Need of lumped memory

- Memory resources are almost always needed in every design for tasks such as:
  - FIFOs: For clock synchronization (elastic buffers) and as message passing blocks between data path processing elements (with or without bus width conversion)
  - Digital Filter designs in DSP (storing coeffs)
  - Instruction and data memories in microcontroller and processor cores
    - Local, fast cache for systems using external memories
- Memory realized using F/Fs are prohibitively resource intensive
- Memories can be realized using LUTs also. However this tends to be spread out in multiple CLB columns and there are timing issues for large width/ depth
- Dual port memories make FIFO implementations more efficient

  - Which is efficient utilisation: Bungalow(FF) or Flat(BRAM) ?

2024 © Centre for Development of Advanced Computing

BlockRAM

- Dedicated 36-bit wide and 1K deep blocks available within the chip. Can be used as dual independent 18Kbit RAMs
- Two 36K BRAMs can be combined to form 64kX1 w/o any logic.
- Basic unit is true dual port synchronous memory which can be cascaded to support deeper/wider configurations
- Basic unit can be configured to support 1-bit .. 72-bit widths
- A number of read/write modes in dual port mode are supported.
- Using memory INIT attributes, each element can be initialized to desired value (useful for storing coefficients in filter design and initializing instruction memory of processors after power on)
- In dual port mode, unequal data widths for two ports are supported (e.g. 32-bit input path converted to 8-bit output data path)
- Total number of blockRAM elements as well as their location depends upon device selected.
- ECC supported

2024 © Centre for Development of Advanced Computing

BlockRAM

Table 1-8: **Port Aspect Ratio for RAMB18E1 (in TDP Mode)**

Port Data Width	Port Address Width	Depth	ADDR Bus	DI Bus DO Bus	DIP Bus DOP Bus
1	14	16,384	[13:0]	[0]	NA
2	13	8,192	[13:1]	[1:0]	NA
4	12	4,096	[13:2]	[3:0]	NA
9	11	2,048	[13:3]	[7:0]	[0]
18	10	1,024	[13:4]	[15:0]	[1:0]

Table 1-9: **Port Aspect Ratio for RAMB18E1 (in SDP Mode)**

Port Data Width	Alternate Port Width	Port Address Width	Depth	ADDR Bus	DI Bus DO Bus	DIP Bus DOP Bus
32	1	14	16,384	[13:0]	[0]	NA
32	2	13	8,192	[13:1]	[1:0]	NA
32	4	12	4,096	[13:2]	[3:0]	NA
36	9	11	2,048	[13:3]	[7:0]	[0]
36	18	10	1,024	[13:4]	[15:0]	[1:0]
36	36	9	512	[13:5]	[31:0]	[3:0]

2024 © Centre for Development of Advanced Computing

BlockRAM

Table 1-10: **Port Aspect Ratio for RAMB36E1 (in TDP Mode)**

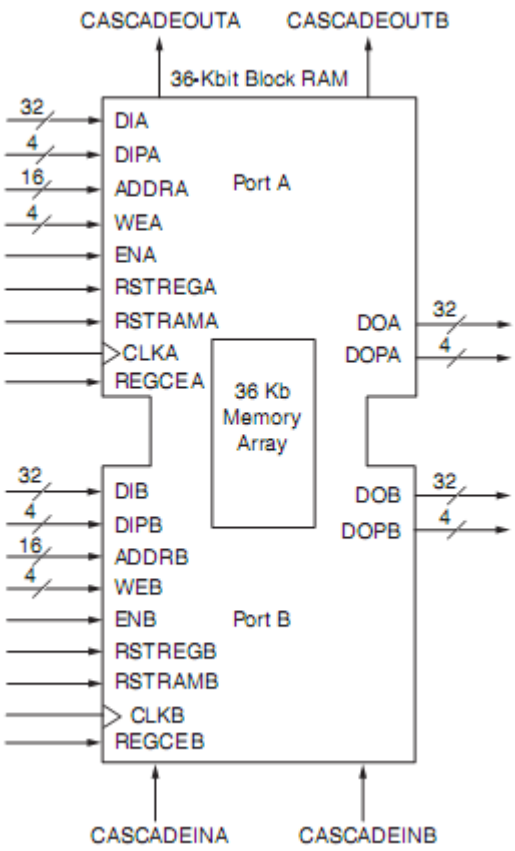
Port Data Width	Port Address Width	Depth	ADDR Bus	DI Bus DO Bus	DIP Bus DOP Bus
1	15	32,768	[14:0]	[0]	NA
2	14	16,384	[14:1]	[1:0]	NA
4	13	8,192	[14:2]	[3:0]	NA
9	12	4,096	[14:3]	[7:0]	[0]
18	11	2,048	[14:4]	[15:0]	[1:0]
36	10	1,024	[14:5]	[31:0]	[3:0]
1 (Cascade)	16	65536	[15:0]	[0]	NA

Table 1-11: **Port Aspect Ratio for RAMB36E1 (in SDP Mode)**

Port Data Width(1)	Alternate Port Width	Port Address Width	Depth	ADDR Bus	DI Bus DO Bus	DIP Bus DOP Bus
64	1	15	32,768	[14:0]	[0]	NA
64	2	14	16,384	[14:1]	[1:0]	NA
64	4	13	8,192	[14:2]	[3:0]	NA
72	9	12	4,096	[14:3]	[7:0]	[0]
72	18	11	2,048	[14:4]	[15:0]	[1:0]
72	36	10	1,024	[14:5]	[31:0]	[3:0]
72	72	9	512	[14:6]	[63:0]	[7:0]

2024 © Centre for Development of Advanced Computing

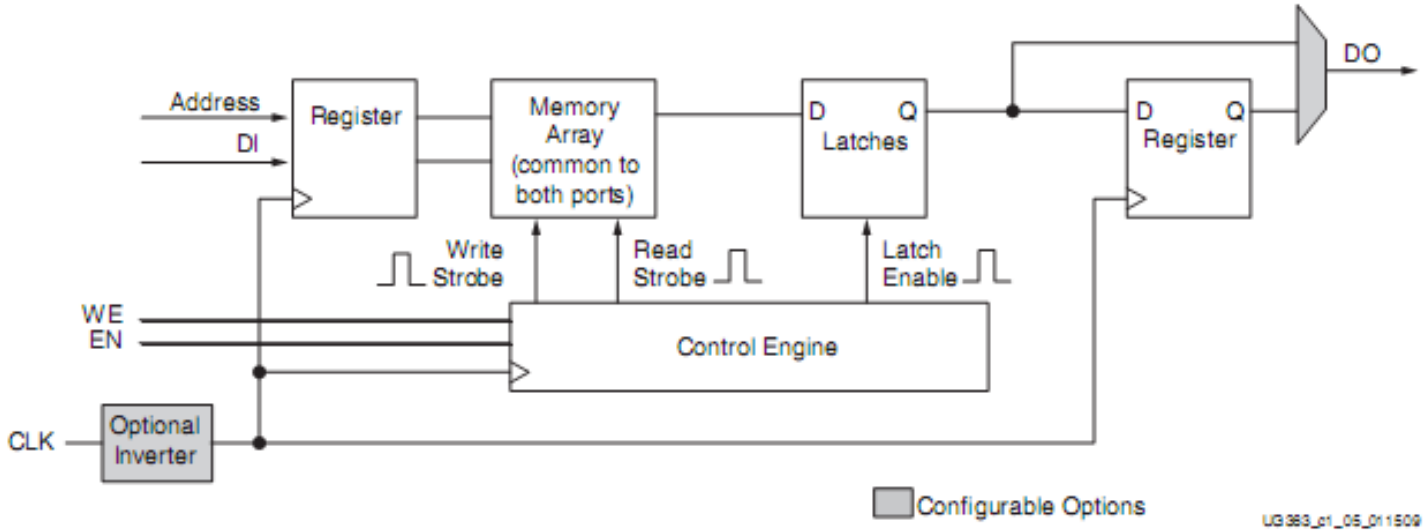
BlockRAM port map (dual port)



- Independent clock domains for port A and B
- Dix,DIPx: Data and parity input (32+4 bits in 32-bit config)
- ADDRA,ADDRB: memory address for port A and B
- WEA, WEB: Write enable
- ENA,ENB: memory enable
- RSTREG[A|B]: Sync-set/reset of DOx/DOPx registers
- RSTRAM[A|B]: Sync-set/reset of output data latches
- CLKA,CLKB: clocks
- DOx,DOPx: Data outputs
- Width of ADDR and DO depends upon the memory organization (width/depth)

2024 © Centre for Development of Advanced Computing

BlockRAM optional o/p register



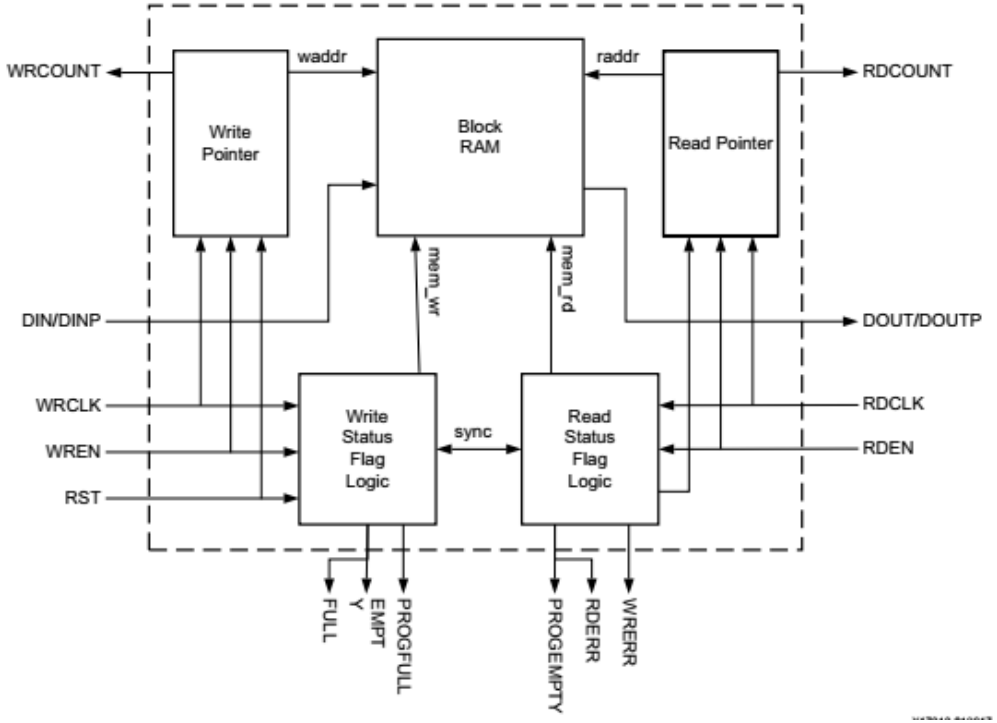
2024 © Centre for Development of Advanced Computing

BlockRAM usage

- Some tools can infer blockRAM from behavioral description of memory
- Recommended way is to use desired macro or IP from the vivado toolchain
- Since blockRAMs are distributed, take into account routing delays
- Consider using distributed RAM for small memories or when fast access is required

2024 © Centre for Development of Advanced Computing

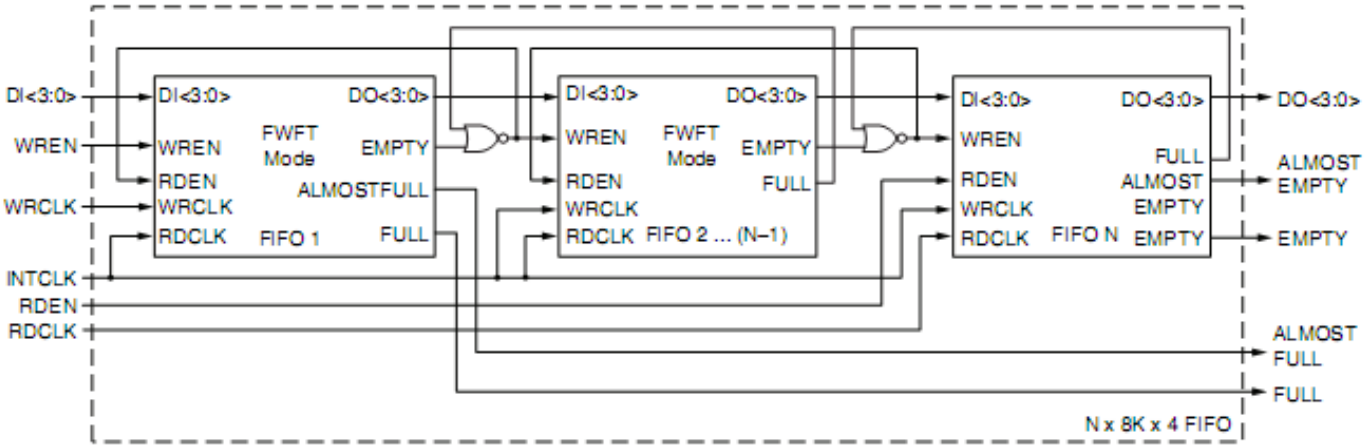
BlockRAM usage – Built in FIFO



- Does not use FPGA logic (CLBs) to realise FIFO
- Sync/Async FIFOs can be generated
- No metastability issues when used as Async FIFO
- Two operating modes: Standard and FWFT

2024 © Centre for Development of Advanced Computing

Cascading FIFOs to increase depth

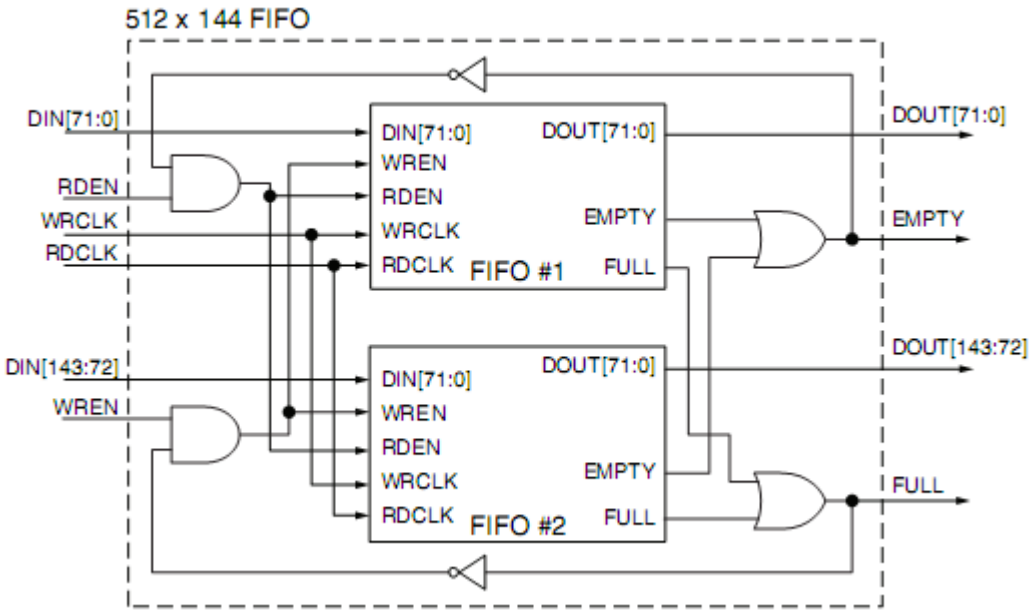


Ref: 7-series architecture

- N can be 2 or more; N= 2, only two FIFOs required
- N-1 FIFOs will be in FWFT mode
- NOR gates implemented with CLBs
- ALMOST_EMPTY threshold is set in the Nth FIFO; ALMOST_FULL threshold is set in 1st FIFO

2024 © Centre for Development of Advanced Computing

Cascading FIFOs to increase width



Ref: 7-series architecture

- AND/OR gates implemented with CLBs
- All FIFO FULL signals must be Ored together
- All FIFO EMPTY signals must be Ored together
- Max frequency limited by logic gate feedback path

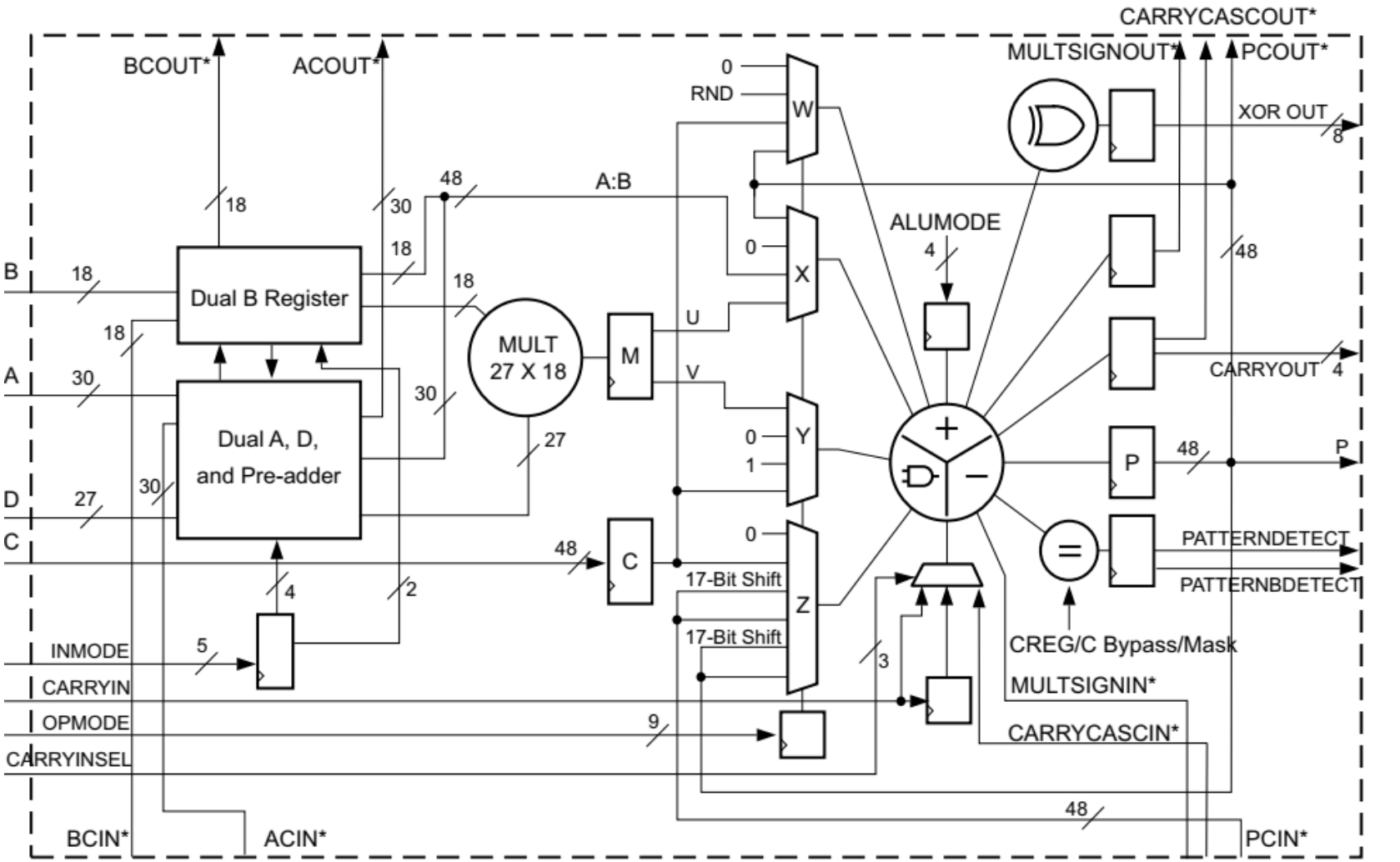
2024 © Centre for Development of Advanced Computing

DSP block

- DSP is characterised with a specific set of operations like multiply, multiply accumulate (MACC), multiply add, three-input add, barrel shift, wide-bus multiplexing, magnitude comparator, bit-wise logic functions, pattern detect, and wide counter
- Multiply operation is resource intensive
- Multipliers used along with blockRAM for storing coefficients are useful in digital filter designs for DSP applications
- DSP48E2 has 27x18 bit multiplier block
- Can be instantiated as a macro or inferred automatically by synthesis tools
- Some of the applications that DSP48 slice can support are Fast Fourier Transforms (FFTs), Systolic and Multirate FIR filters, CIC filters, floating point computation (multiply, add/sub, divide), counters, and large bus multiplexers

2024 © Centre for Development of Advanced Computing

DSP48E2 slice details

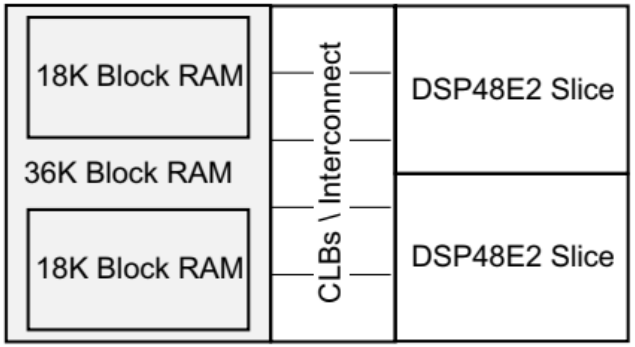


*These signals are dedicated routing paths internal to the DSP48E2 column. They are not accessible via general-purpose routing resources.

X16752-042617

2024 © Centre for Development of Advanced Computing

DSP48E2 slice arrangement



- Two DSP48E2 slices with dedicated interconnects form a tile. Tiles stack vertically.
- Height of tile matches that of BRAM. It also spans 5 CLBs. Each split BRAM (18KBit) can serve each DSP slice horizontally

	Max Cascade	Total
Kintex UltraScale		
KU025	72	1,152
KU035	120	1,700
KU040	120	1,920
KU060	120	2,760
KU085	120(1)	4,100
KU095	192	768
KU115	120	5,520

Virtex UltraScale		
VU065	120	600
VU080	192	672
VU095	192	768
VU125	120	1,200
VU160	120(2)	1,560
VU190	120	1,800
VU440	120	2,880

2024 © Centre for Development of Advanced Computing

MMCM and clocking

2024 © Centre for Development of Advanced Computing

Clock management

- Modern designs almost always require complex clocking schemes, multi-clock domains, tight margins at the I/Os and dynamic clocking schemes for power efficient designs
- Some type of on-chip clock management and distribution network is therefore essential
- Mixed Mode Clock Manager (MMCM) along with special buffers and MUXes are a resource useful for clock mgmt in Ultrascale devices.

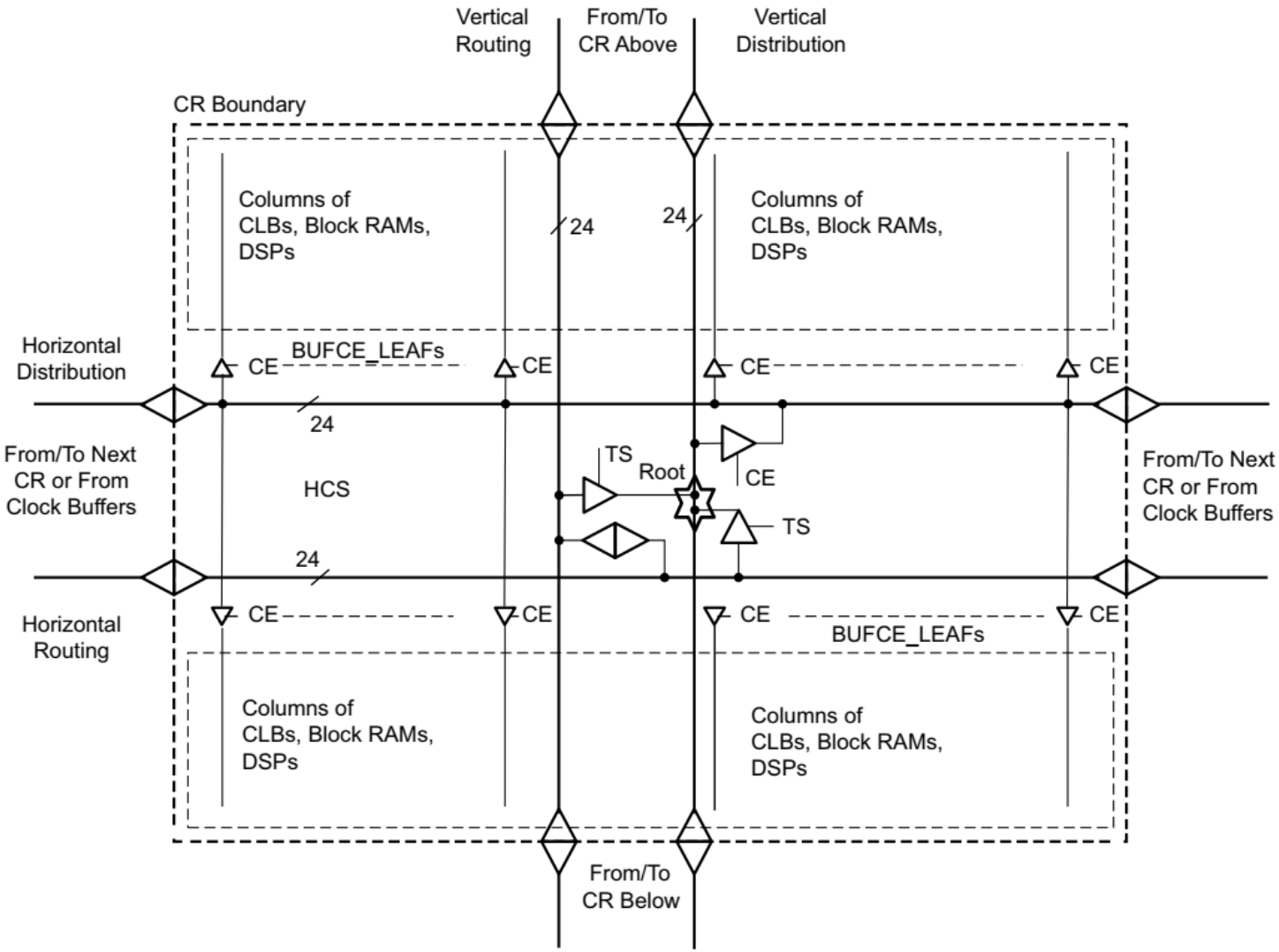
2024 © Centre for Development of Advanced Computing

Basics of clock mgmt in Virtex-6

- Devices are divided into clock regions or CR (60 CLBs high, 24 DSP slices, 12 BRAMs); 1 to 24 regions depending on size of device
- 24 horizontal/vertical routing and distribution clocks in each CR
- These clocks are distributed only by global clock buffers which can also be used as a clock enable circuit, or a glitch-free multiplexer to select between two clock sources. Clock buffers are often driven by Clock Management Tile (CMT)
- Each bank has 4 clock capable pin-pair inputs

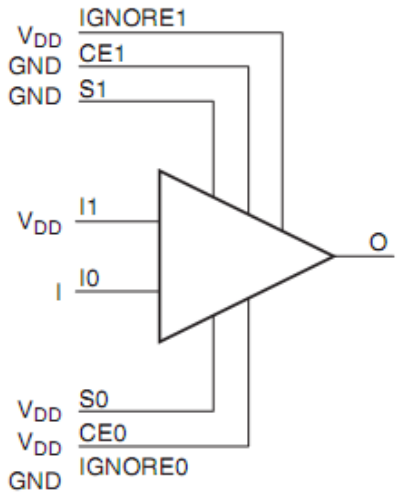
2024 © Centre for Development of Advanced Computing

Clock mgmt

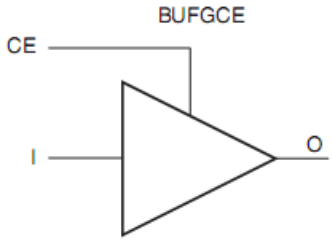


2024 © Centre for Development of Advanced Computing

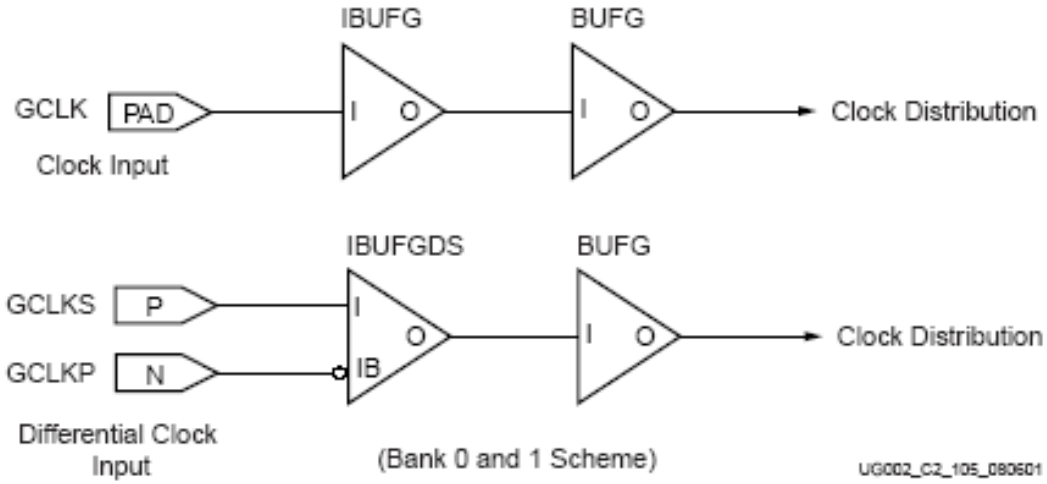
Clock mgmt



Variation of this primitive used



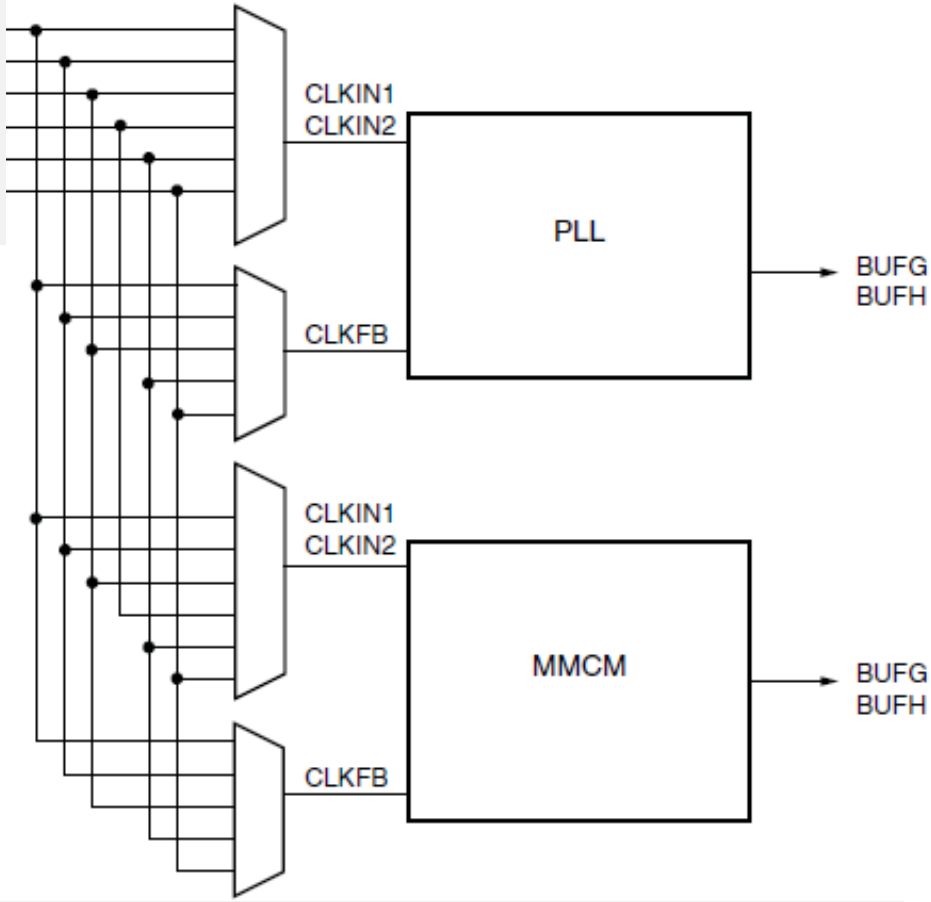
CE is CE0 pin of BUFGCTRL



2024 © Centre for Development of Advanced Computing

CMT with MMCM

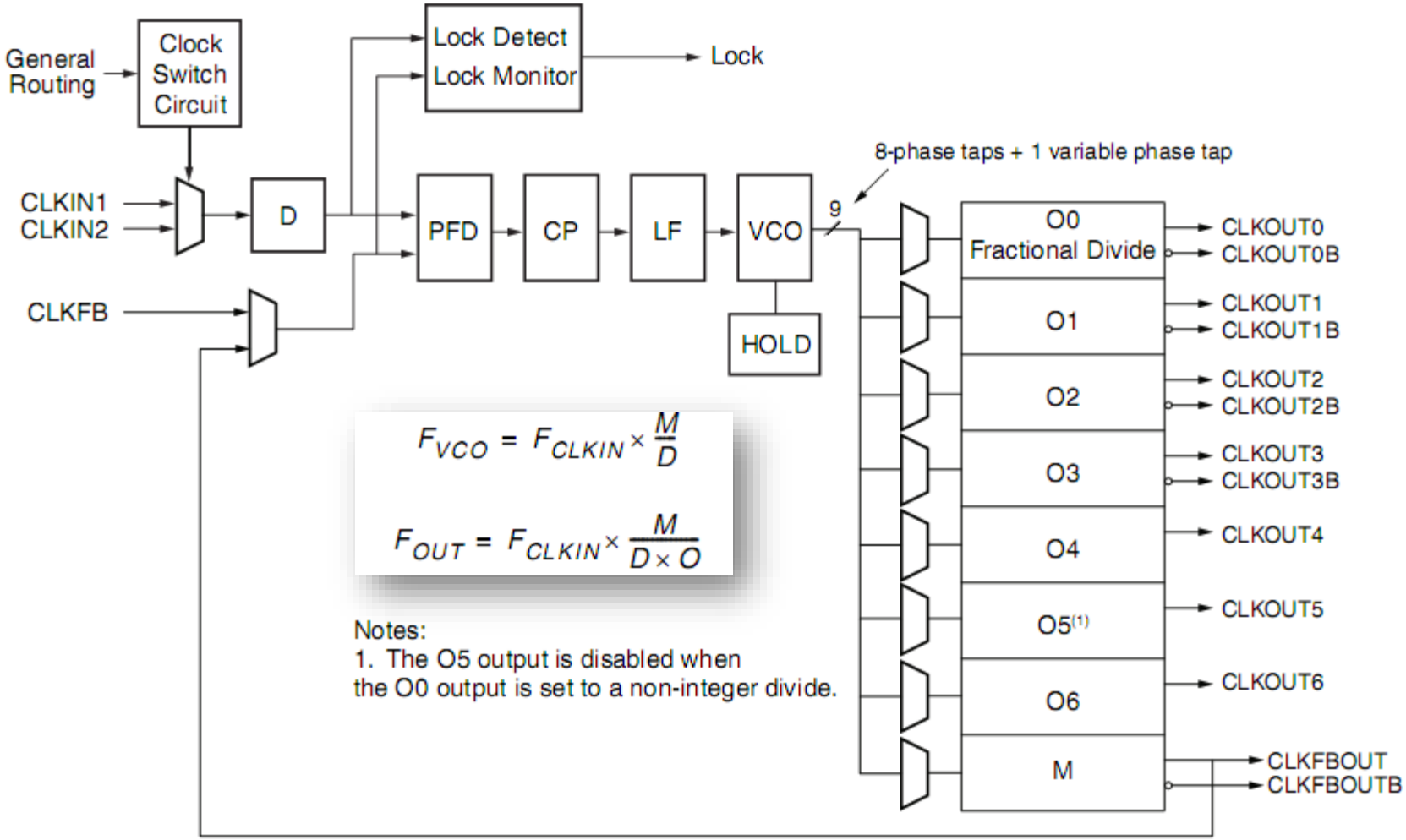
Various clock inputs



One CMT contains 2 PLL and 1 MMCM

2024 © Centre for Development of Advanced Computing

MMCM



$$F_{VCO} = F_{CLKIN} \times \frac{M}{D}$$

$$F_{OUT} = F_{CLKIN} \times \frac{M}{D \times O}$$

Notes:
1. The O5 output is disabled when the O0 output is set to a non-integer divide.

2024 © Centre for Development of Advanced Computing

PLL

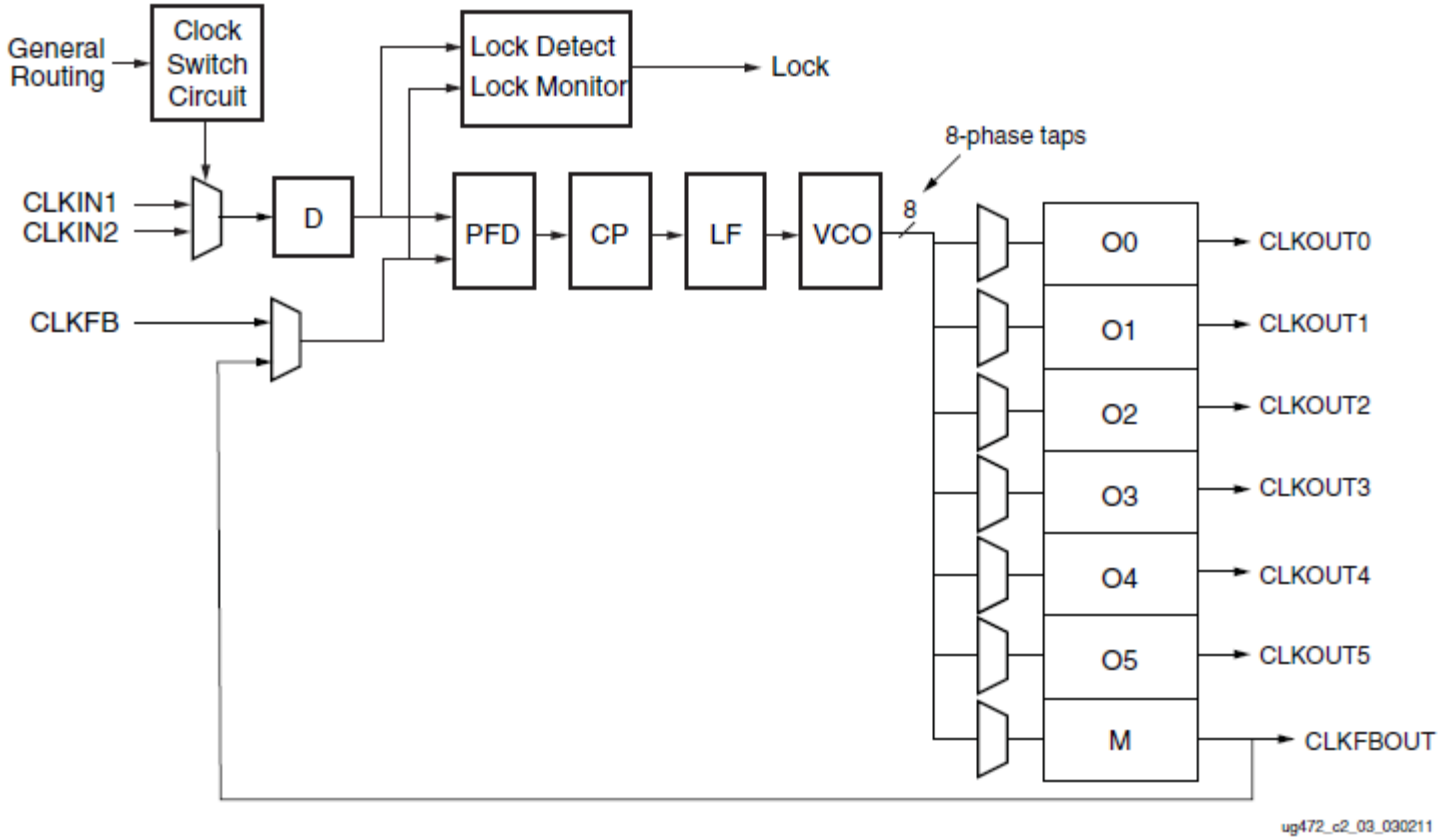


Figure 3-3: **Detailed PLL Block Diagram**

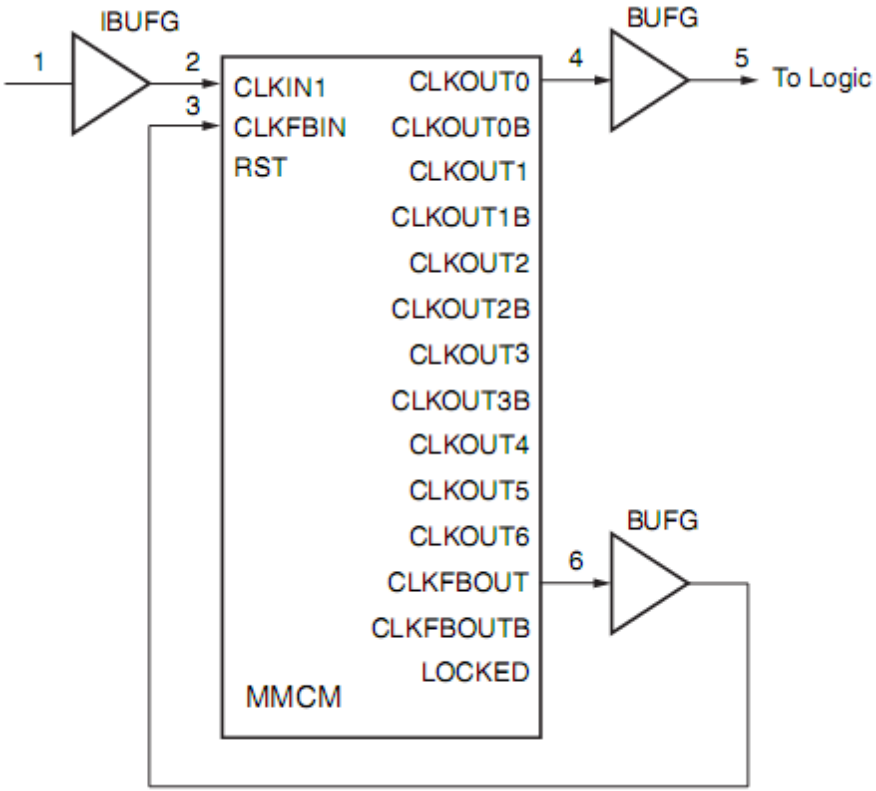
2024 © Centre for Development of Advanced Computing

MMCM

- MMCM & PLL serve as a frequency synthesizer for a wide range of frequencies, a jitter filter for either external or internal clocks, and deskew clocks

2024 © Centre for Development of Advanced Computing

Deskewing (Internal)

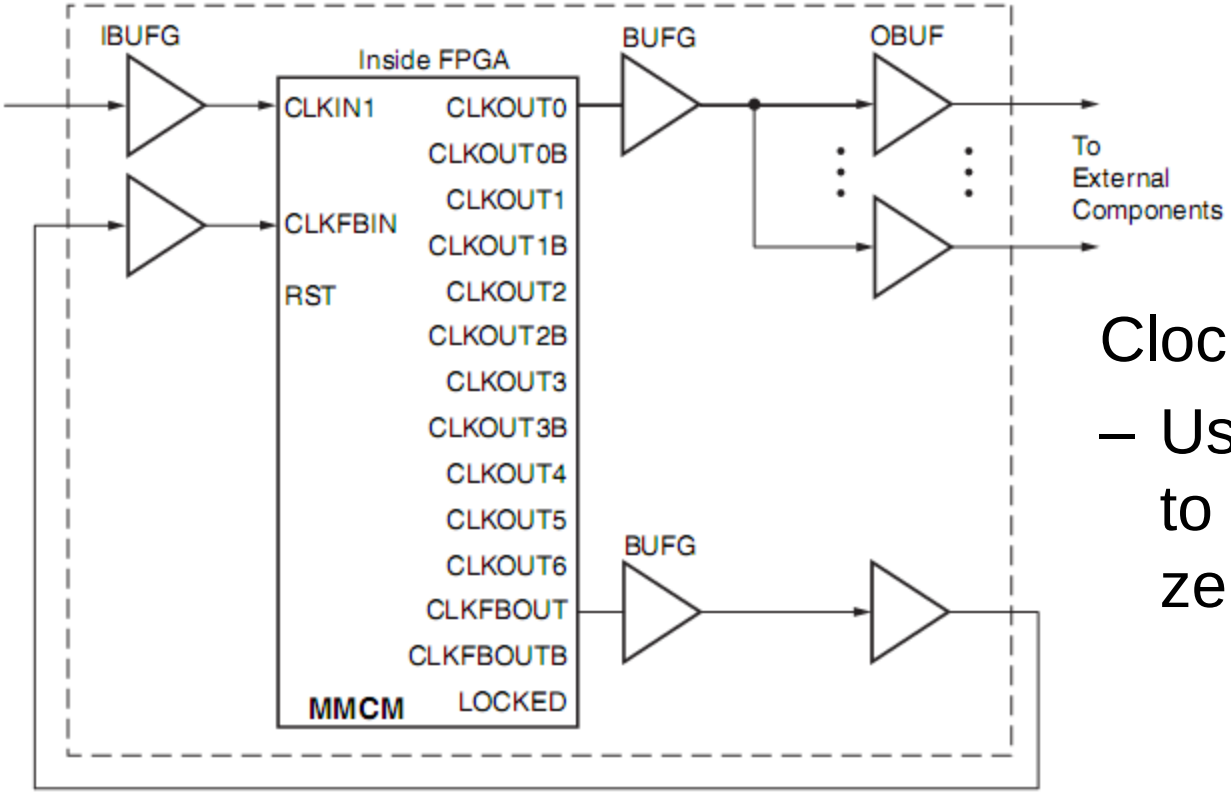


Clock De-skew

– Used frequently to compensate for delays of clock distribution buffer (BUFG)

2024 © Centre for Development of Advanced Computing

Deskewing (External)



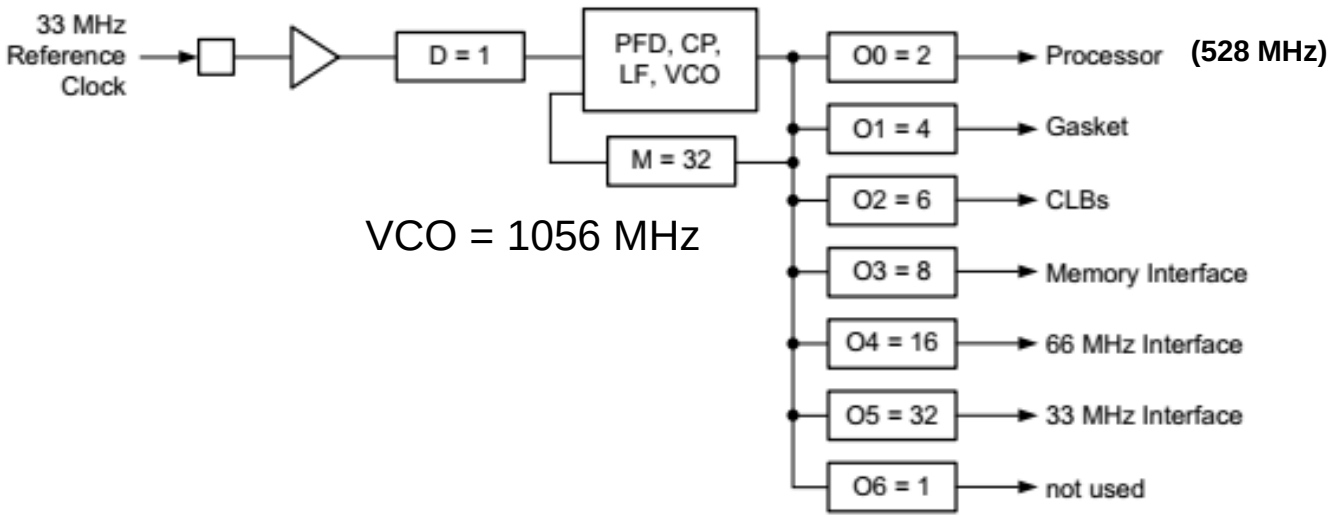
Clock De-skew

– Used frequently to implement zero delay buffer

2024 © Centre for Development of Advanced Computing

Frequency synthesis

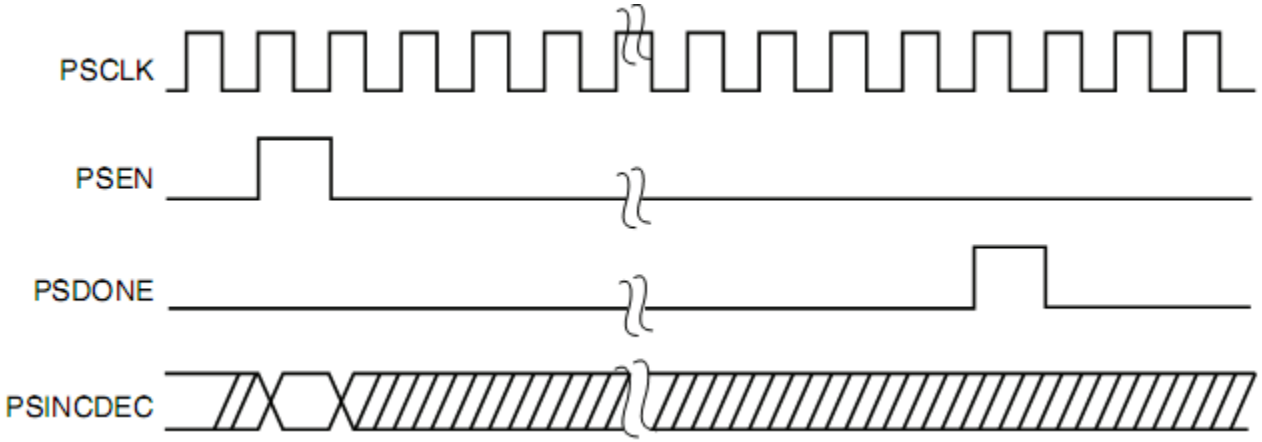
- Frequency synthesis with both integer and fractional factors can be done



2024 © Centre for Development of Advanced Computing

Phase shifting

- Phase shifting
  - Static phase shifting is available in steps of 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°
  - In addition, interpolated fine phase shifting can be done as fixed value (1/56 VCO period) during configuration or dynamically after configuration



2024 © Centre for Development of Advanced Computing

MMCM limitations

- There are numerous operational limits on MMCM, such as max frequency, range of locking, M/OxD limits, etc
- Number of MMCMs < other clock mgmt resources. Coupled with location constraints this may represent difficulties for some clocking schemes
- MMCM can only drive BUFGs in respective half and the feedback can also be taken only from respective half

2024 © Centre for Development of Advanced Computing

Routing Matrix

2024 © Centre for Development of Advanced Computing

Routing resources

- Routing delays account for a major component in timing budget in FPGAs
  - This is because of the configurable nature of the switching matrix with plenty of active ‘switching points’. Each of these add to the delays (metal interconnect is not a major contributor)
- Fine granular nature of modern FPGAs demand huge routing resources at local (to implement a macro or module) and global (for interconnecting nets with huge fanouts or between big modules in a hierarchical design) levels

2024 © Centre for Development of Advanced Computing

Routing resources

- Active interconnect for more predictable routing delays independent of locations e.g. IOB/CLB, and fanout. This allows for a unified timing model to be used throughout the chip
- Routing resources (local and global) are connected to ‘switch matrix’ which is a common block for elements such as CLBs, IOBs, BlockRAM
- Over the years, FPGA routing architectures have evolved, offering fast, hierarchical, predictable routing resources for automated routing tools
  - Guidelines from user help a lot
- Power for interconnect is derived from V_{CCAUX}

2024 © Centre for Development of Advanced Computing

Routing strategies

- Routing resources are not visible to the front end designer and controlling them at HDL level is difficult
- However, good design practices allow for backend tools to optimally place & route the design
- Routing inefficiencies can be interpreted by looking at the PAR results, timing reports etc
- Directives given to the backend tools in the form of constraints help in optimal routing strategies
- Tools such as FPGA editor (backend) allow for manipulation of routing at micro level which may be required to be done for critical nets or for creation of hand optimized macros

2024 © Centre for Development of Advanced Computing

Configuration

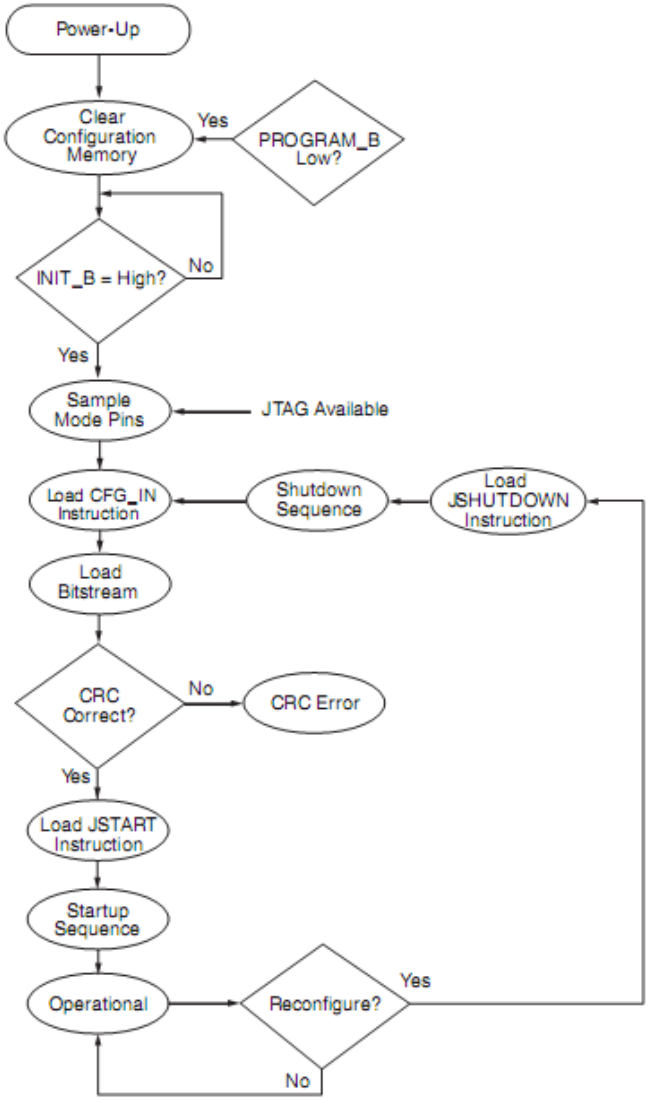
2024 © Centre for Development of Advanced Computing

Configuration

- Devices are configured at the time of power-on using a bit stream (serial or parallel)
- A number of modes determine “master” or “slave” mode of operation, serial/parallel bit stream etc
- Multiple devices can be daisy chained together
- Partial reconfiguration is supported using selectMAP feature
- To protect IP, this bit stream can be encrypted using AES algorithm with ‘key’ stored on the device (and protected using a battery)
- Configuration can be done using dedicated pins or thru JTAG pins
- Readback feature can be used to read contents of all the registers, distributed memories and blockRAM memories. This feature when used with Integrated Logic Analyzer (ILA) core, can be used as a powerful debugging tool
- Configuration data can be saved in serial/parallel PROMs or can come from elsewhere (e.g. a microcontroller reading from its data memory and configuring the device)

2024 © Centre for Development of Advanced Computing

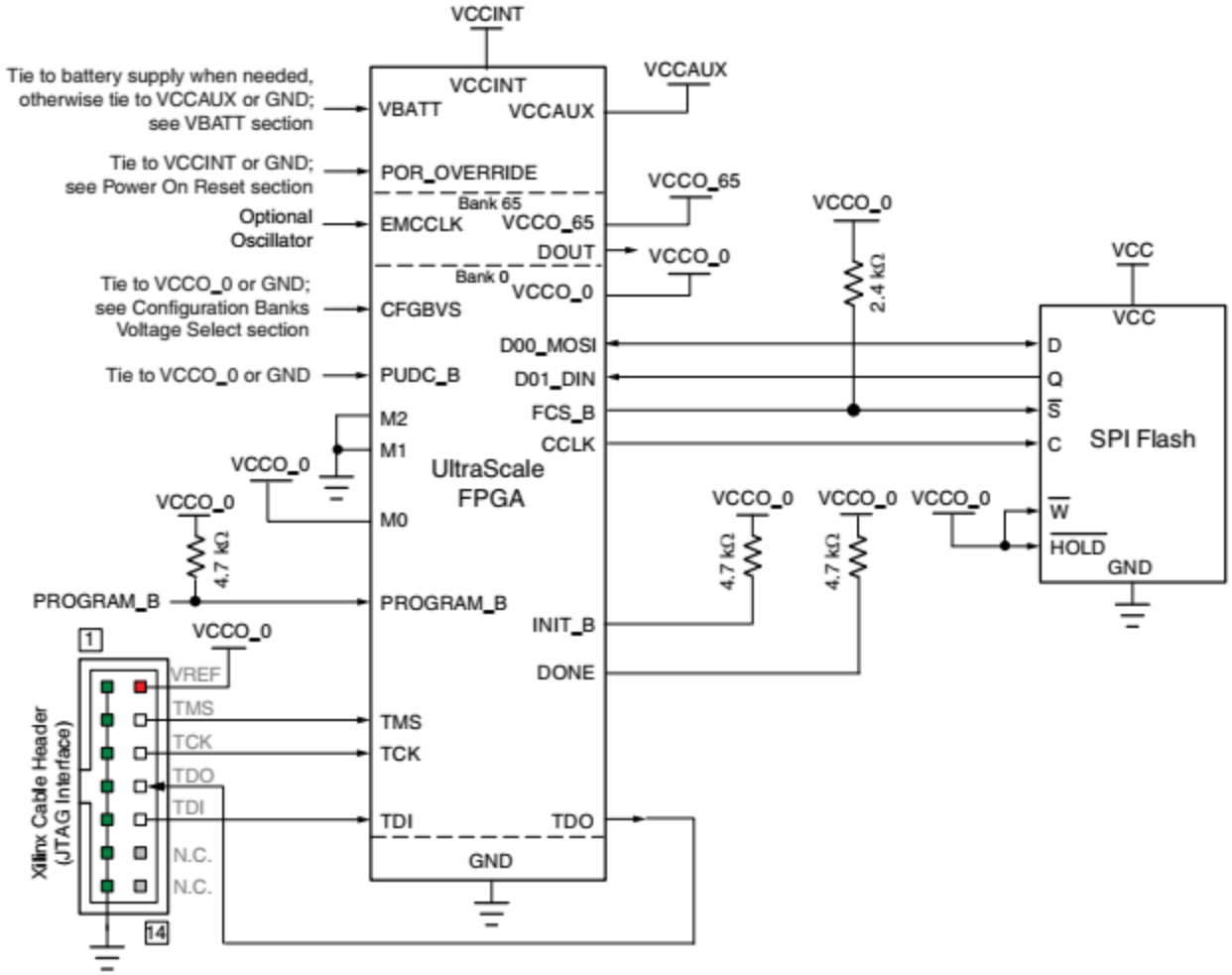
Configuration



- FPGA clears its config memory after input power stabilizes
- It then checks for delayed configuration option (INIT_B=HIGH)
- Depending upon the mode selection, configuration bit stream is loaded into device
- Bitstream is protected by CRC
- Startup sequence is configurable, consists of driving the outputs and enabling the F/F and RAM state transitions
- Configuration can be re-initiated any time after power-on

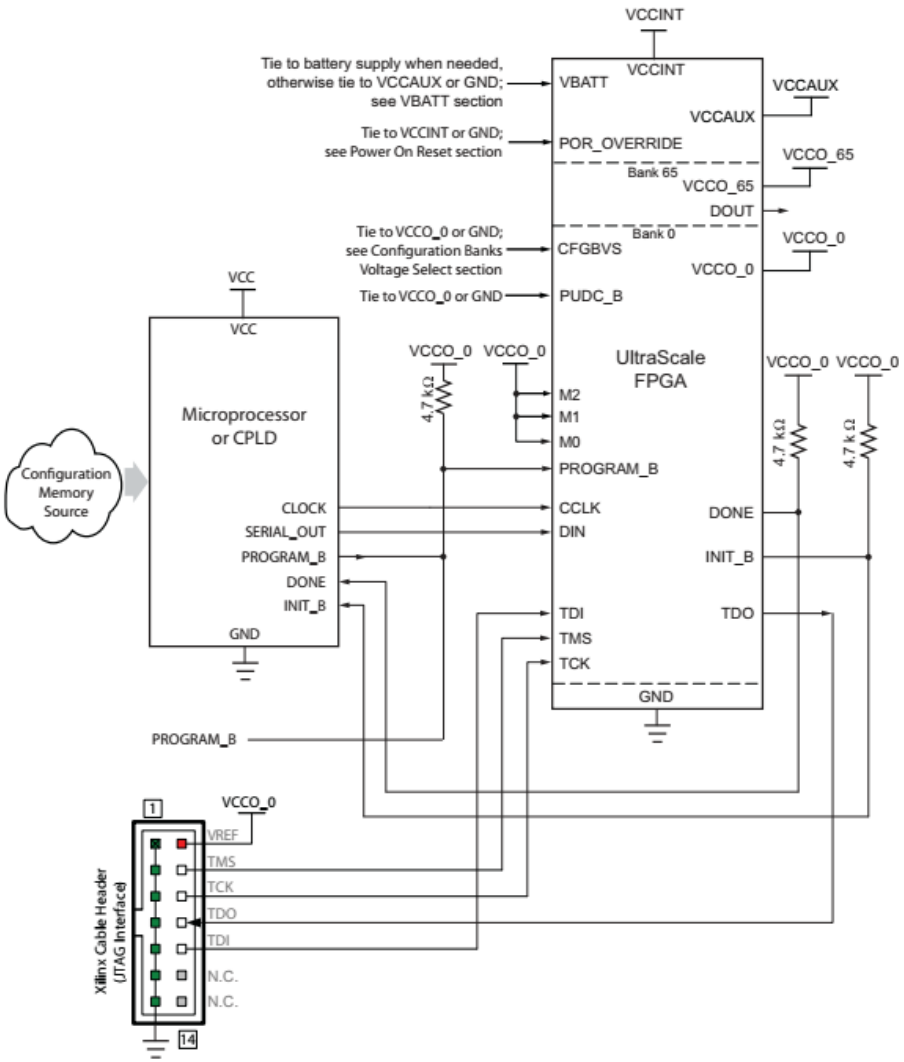
2024 © Centre for Development of Advanced Computing

Configuration: Master SPI



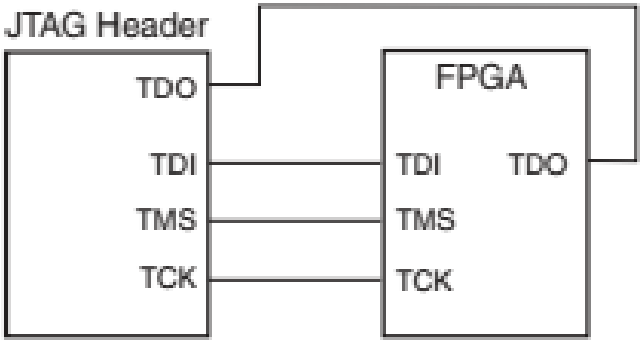
2024 © Centre for Development of Advanced Computing

Configuration: slave serial

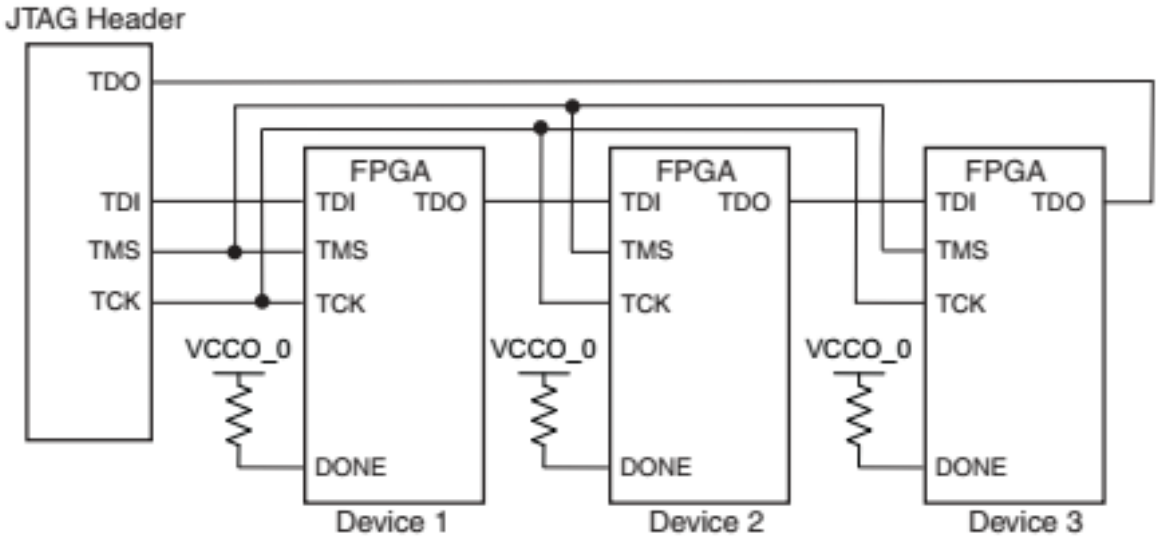


2024 © Centre for Development of Advanced Computing

Configuration: JTAG read/write



Single device JTAG programming



Daisy chain JTAG programming

2024 © Centre for Development of Advanced Computing

Configuration: JTAG read/write

- JTAG defines a standard way of accessing device data using test access port (TAP) and four wire interface (TDI, TDO, TMS and TCK)
- Multiple devices supporting JTAG can be daisy chained. Each device has a unique ID and location in the chain.
- Standard JTAG access control software can be used for access and control of the device. Xilinx supplied “Impact” tool supports JTAG functionality.
- Readback is done using diff set of commands and during device operation
- Apart from these, there are few other configuration modes using parallel interfacing or SelectMAP modes. These modes are high performance modes.
- Fast configuration is critical in certain implementations

  e.g. PCIe card

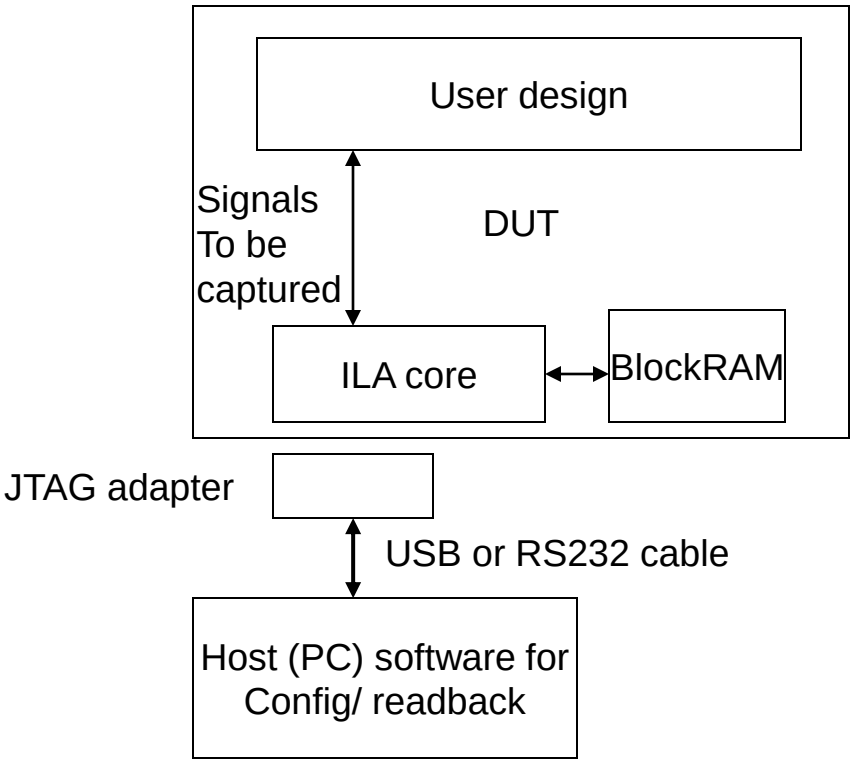
2024 © Centre for Development of Advanced Computing

Readback

- Can be done using selectMAP or JTAG interfaces
- All memory and storage elements can be accessed
- In addition, ILA core allows an arbitrary number of channels to be captured and displayed in a logic-analyzer like manner. ILA core is a MUX-like design which captures data based on a predefined trigger and stores it to blockRAM.
- Note that ILA allows for functional verification since timing may be affected by its inclusion in the design

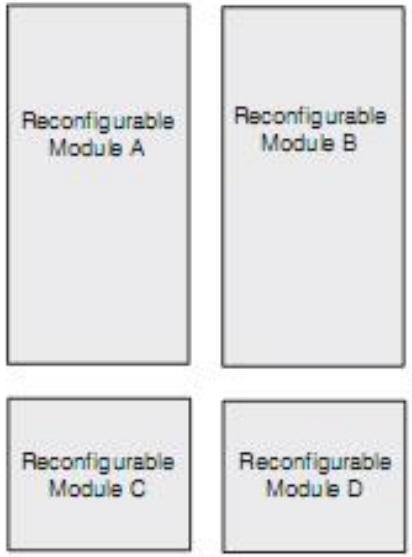
2024 © Centre for Development of Advanced Computing

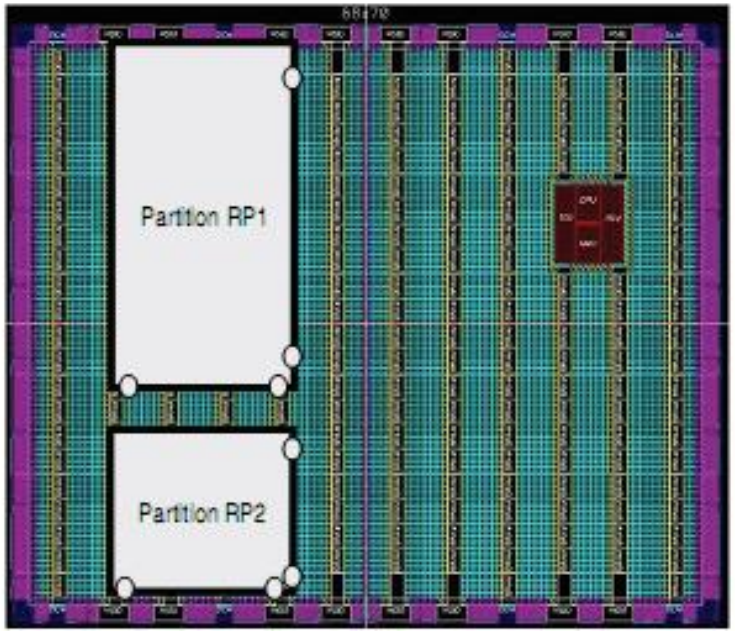
ILA



2024 © Centre for Development of Advanced Computing

Partial Reconfiguration





2024 © Centre for Development of Advanced Computing

Xilinx Ultrascale

Key features

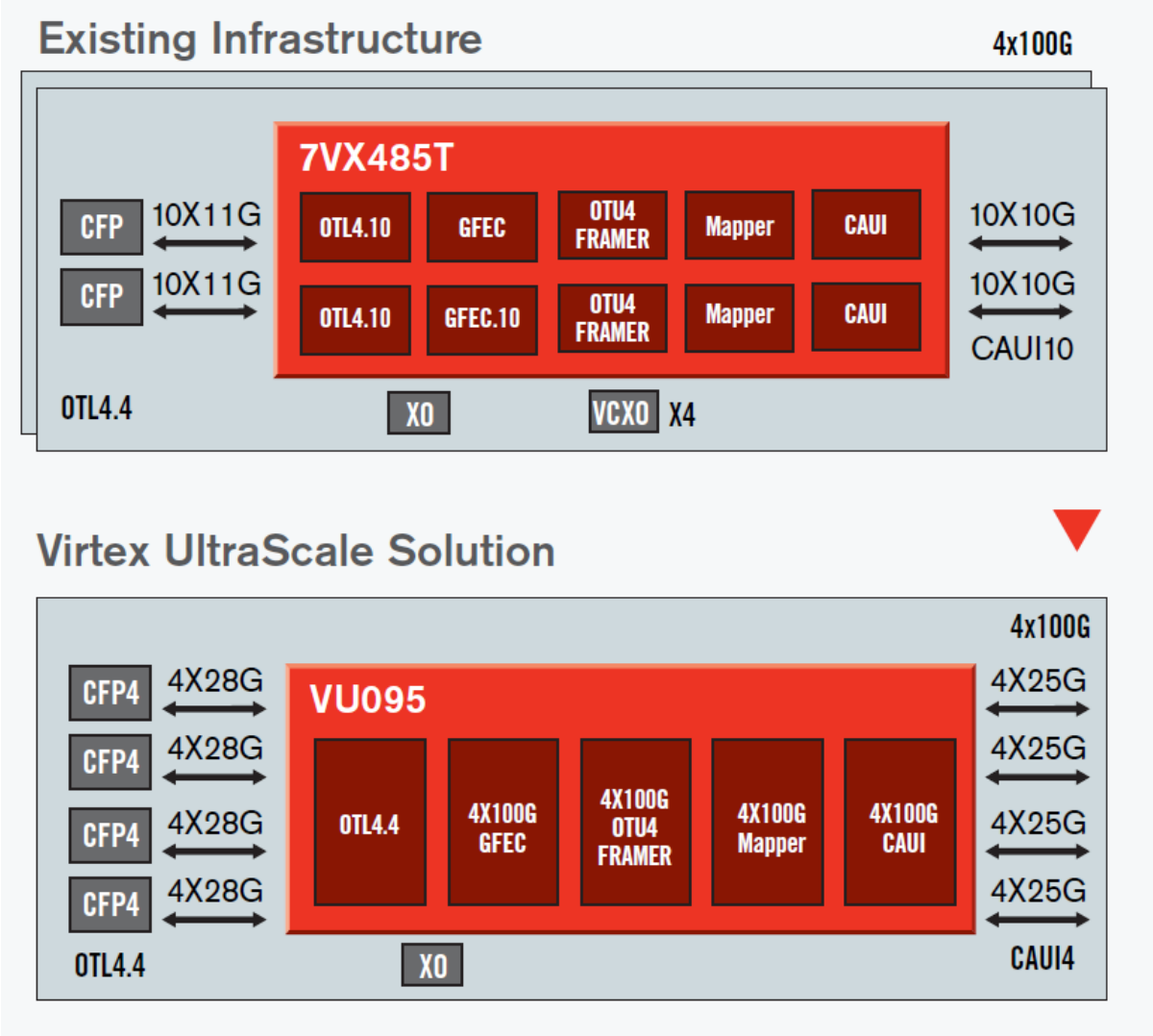
- 20 nm SoC process TSMC
- Upto 5.5M system LCs
- Integrated 100G MAC and 150G Interlaken core
- Enhanced serial transceiver bandwidths
- Integrated PCIe cores
- Next-Generation security to protect IP

Advantage

- >40% performance-per-watt w.r.t Virtex-7
- Higher densities
- Saves 60K-100K LC; 80% Dynamic power Vs. soft core
- 30.5G chip-to-chip; 16.3G and 28G backplane
- Gen3 x8 for 50G BW
- AES-GCM decryption, RSA and DPA

2024 © Centre for Development of Advanced Computing

100G options



2024 © Centre for Development of Advanced Computing

Xilinx Ultrascale+

Key features

- 16 nm LP FinFET+ TSMC process
- Upto 3.6M system LCs
- Upto 8GB integrated HBM
- Integrated 100G MAC and 150G Interlaken core
- Enhanced serial transceiver bandwidths
- Integrated PCIe cores with CC CCIX ports
- Next-Generation security to protect IP

Advantage

- >2x performance-per-watt w.r.t 7-series
- Higher densities
- Highest DRAM BW
- Saves 60K-100K LC; 90% Dynamic power Vs. soft core
- 32.75G chip-to-chip; 16G and 28G backplane
- Gen3 x16; Gen4 x8 for 100G BW
- AES-GCM decryption, RSA-2048 and DPA

2024 © Centre for Development of Advanced Computing

Alternate Architectures

2024 © Centre for Development of Advanced Computing

Stratix-V Family (Altera)

- 2-Dimensional row/column based FPGA architecture
- SRAM based configuration
- 28 nm TSMC process
- Core voltage - 0.85V or 0.9V
- ~236K – 952K Logic elements
- RAM blocks, DSP blocks
- Support for numerous single ended/ differential stds with OCT
- 1.6Gbps LVDS I/O
- 28.05Gbps Transceivers

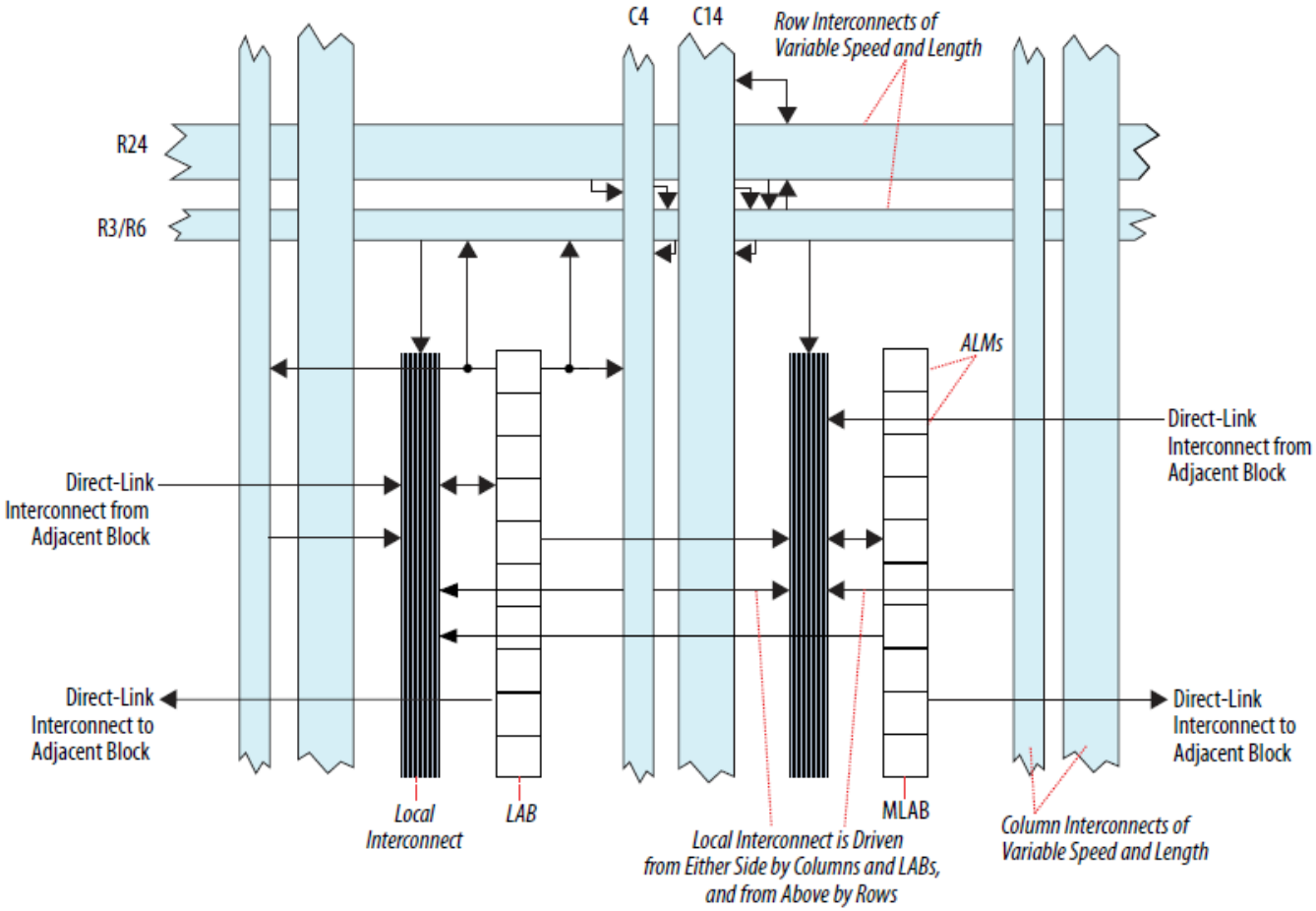
2024 © Centre for Development of Advanced Computing

Blocks

- Logic Array Blocks (LABs and MLABs) consisting of eight Adaptive Logic Modules (ALMs)
- M20K (dual port) memory blocks: 20Kbit
- MLAB: 640-bit
- DSP blocks capable of 27x27 multiplication
- 800MHz fabric clock
- IO Element (IOE) per pin, supports multiple stds, support for fast memory interfaces (DDR/QDR)

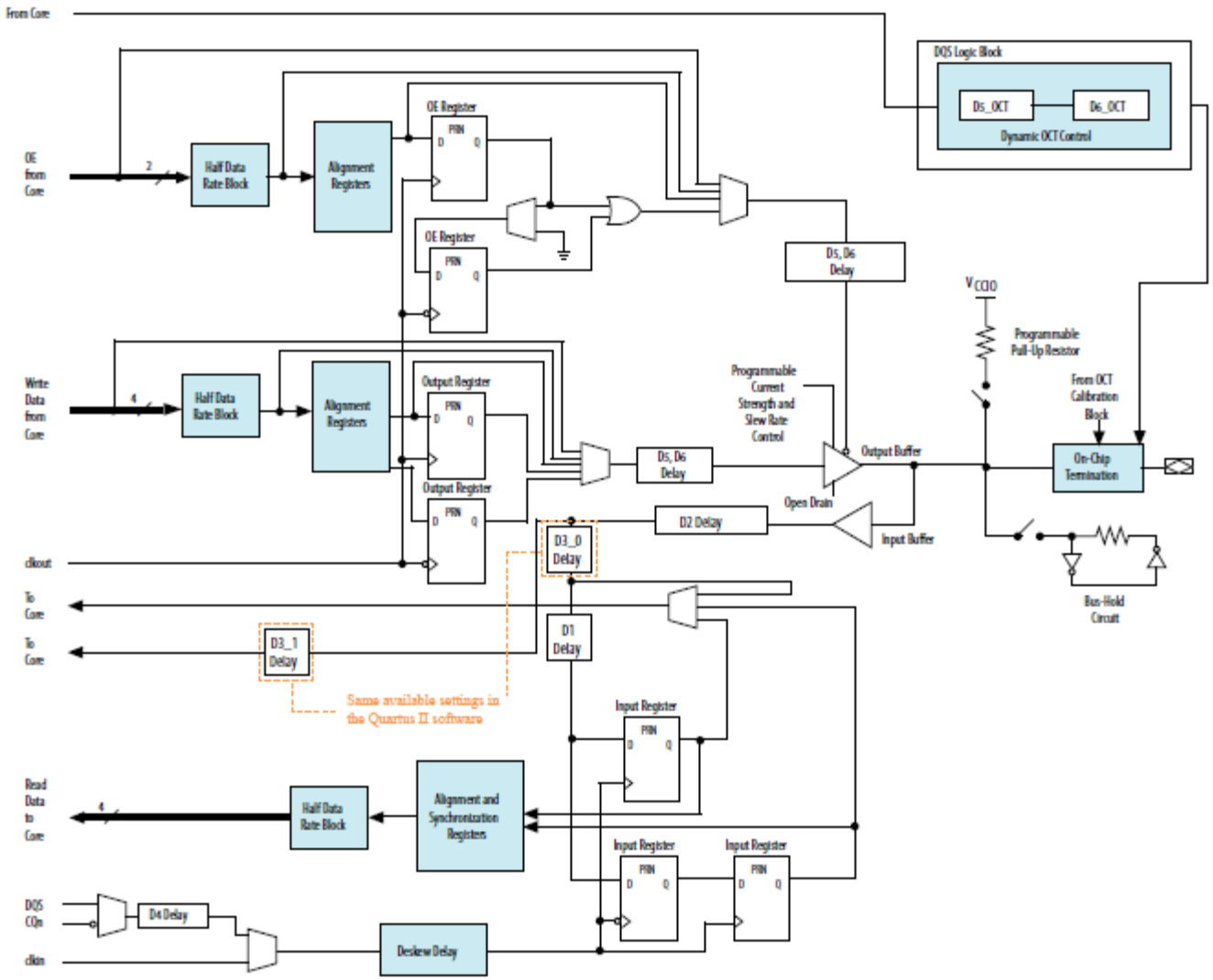
2024 © Centre for Development of Advanced Computing

Overall Architecture



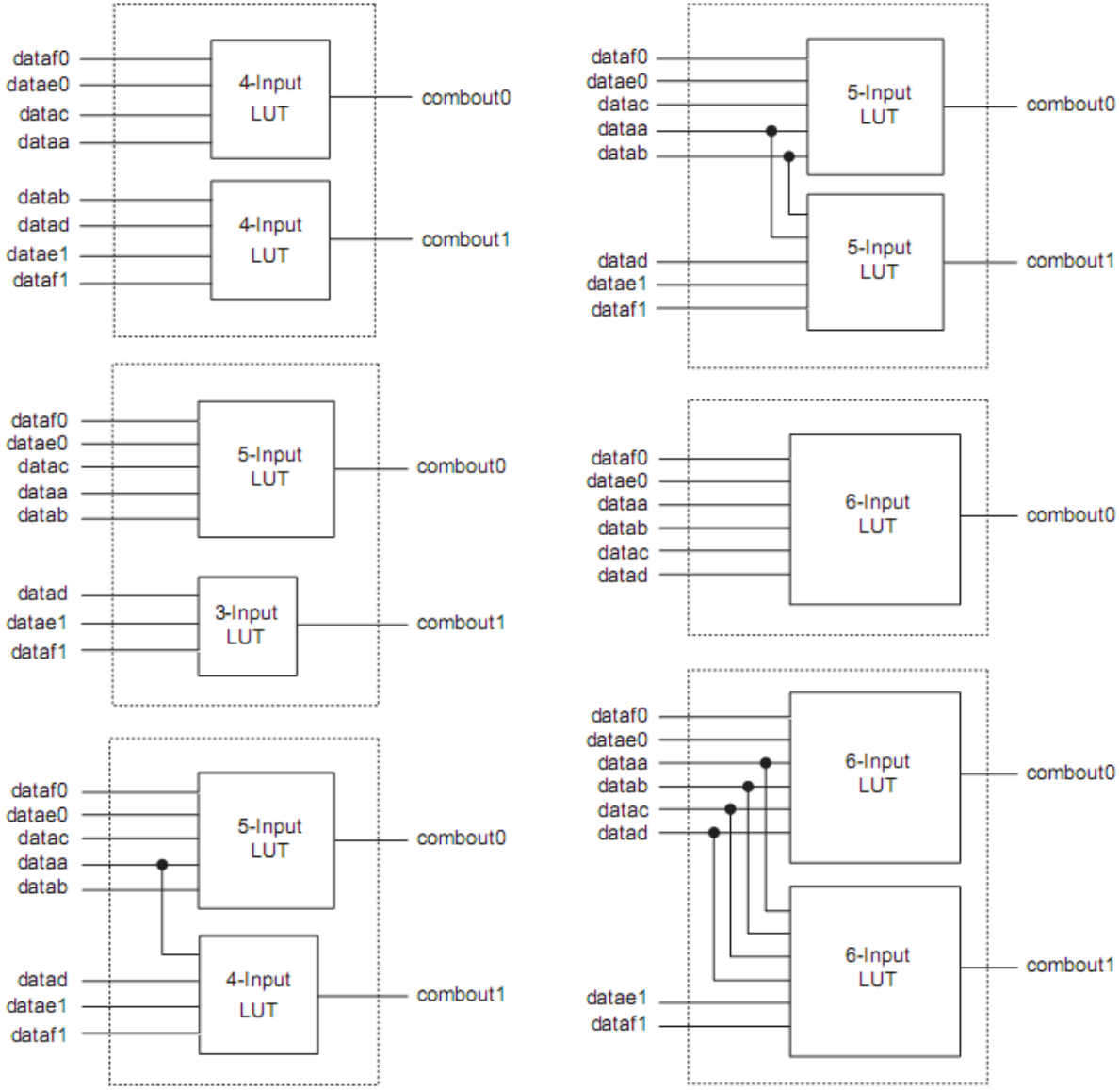
2024 © Centre for Development of Advanced Computing

IOE



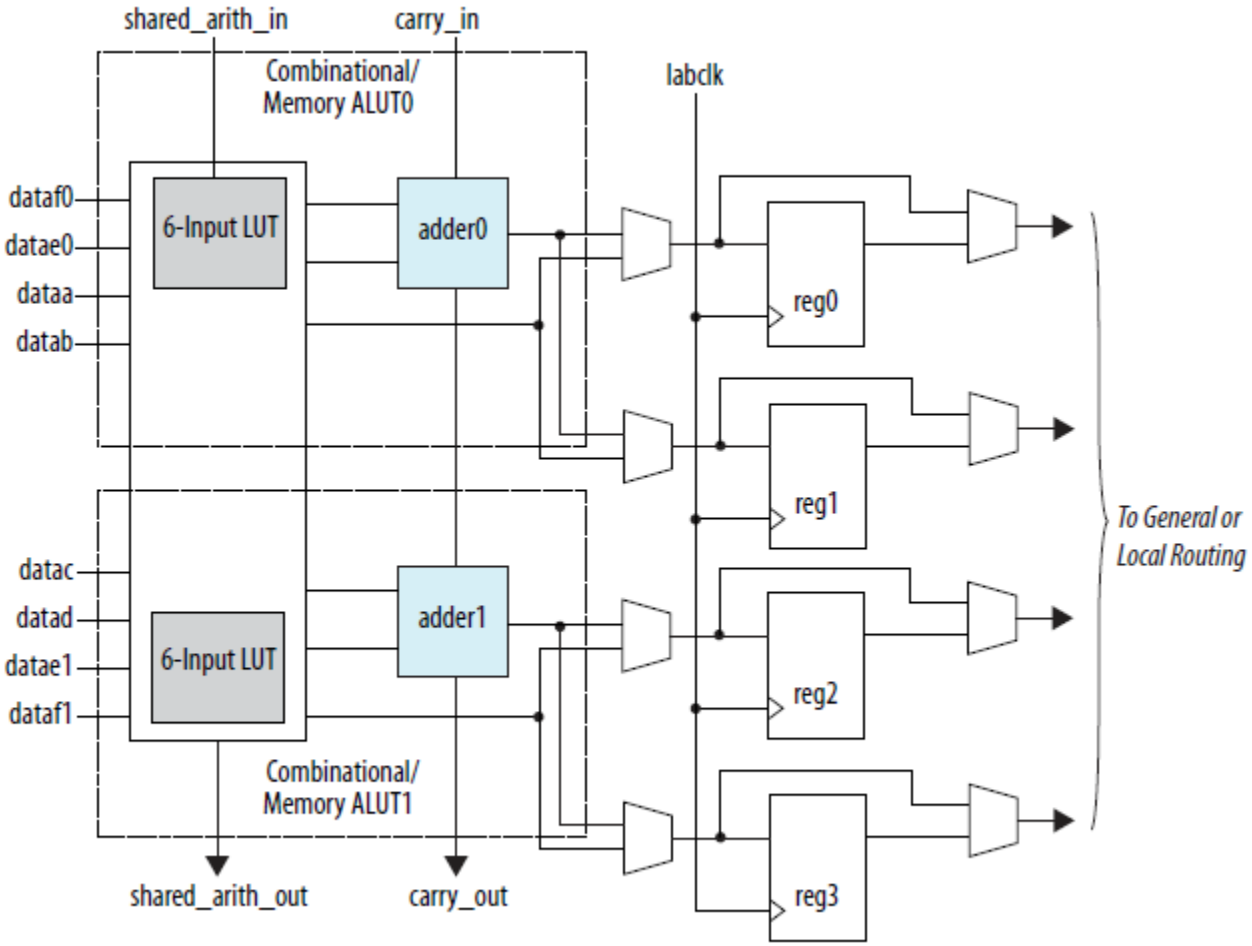
2024 © Centre for Development of Advanced Computing

Stratix-V LUT combinations



2024 © Centre for Development of Advanced Computing

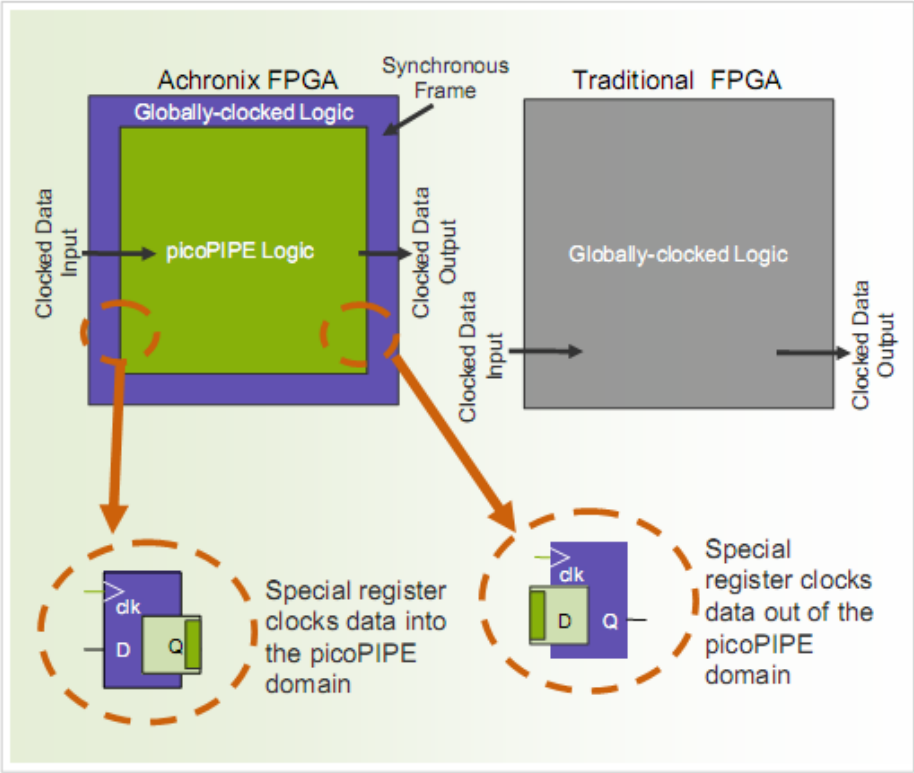
ALM



2024 © Centre for Development of Advanced Computing

Achronix Speedster family

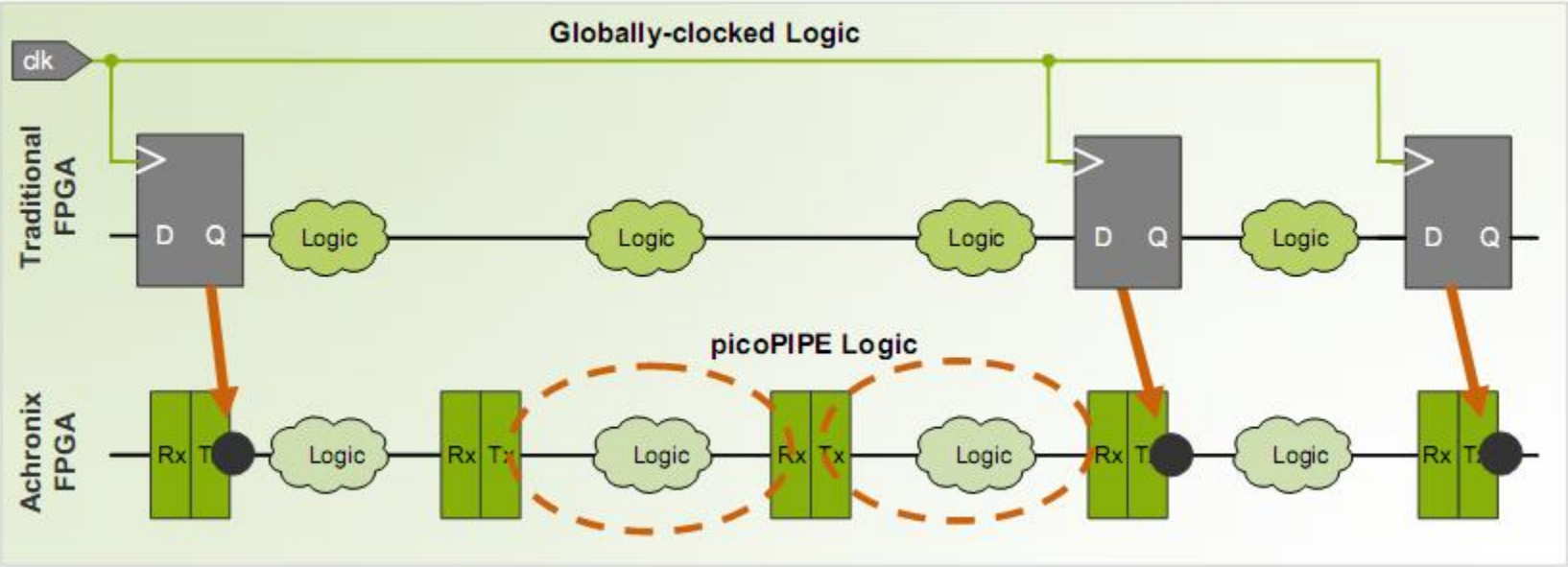
- 65nm, Vcore=1.0V (0.9 V available), TSMC process
- Standard 4-input LUT (RLB)
- Core contains picoPIPE technology used for both logic and routing similar to globally synchronous locally asynchronous
- DDR3/DDR2 controller
- 18x18 bit multiplier
- 18Kbit BRAM (1.5GHz)
- 10.3Gbps SERDES



2024 © Centre for Development of Advanced Computing

PicoPIPE concept

- Features
  - Zero design modification to migrate to Achronix picoPIPE
  - Converts combinational logic and routing to picoPIPE paths



2024 © Centre for Development of Advanced Computing

Factory Programmed FPGA

- Xilinx – Easypath
- Altera – Hardcopy
- The solution is around same FPGA family that designer is targetting. Alternative to ASIC.
- Customer-specific
- Lower volume cost – Suitable for low to medium volumes (1000-100000)
- Available in 12 weeks.
- Need to send implementation files only.

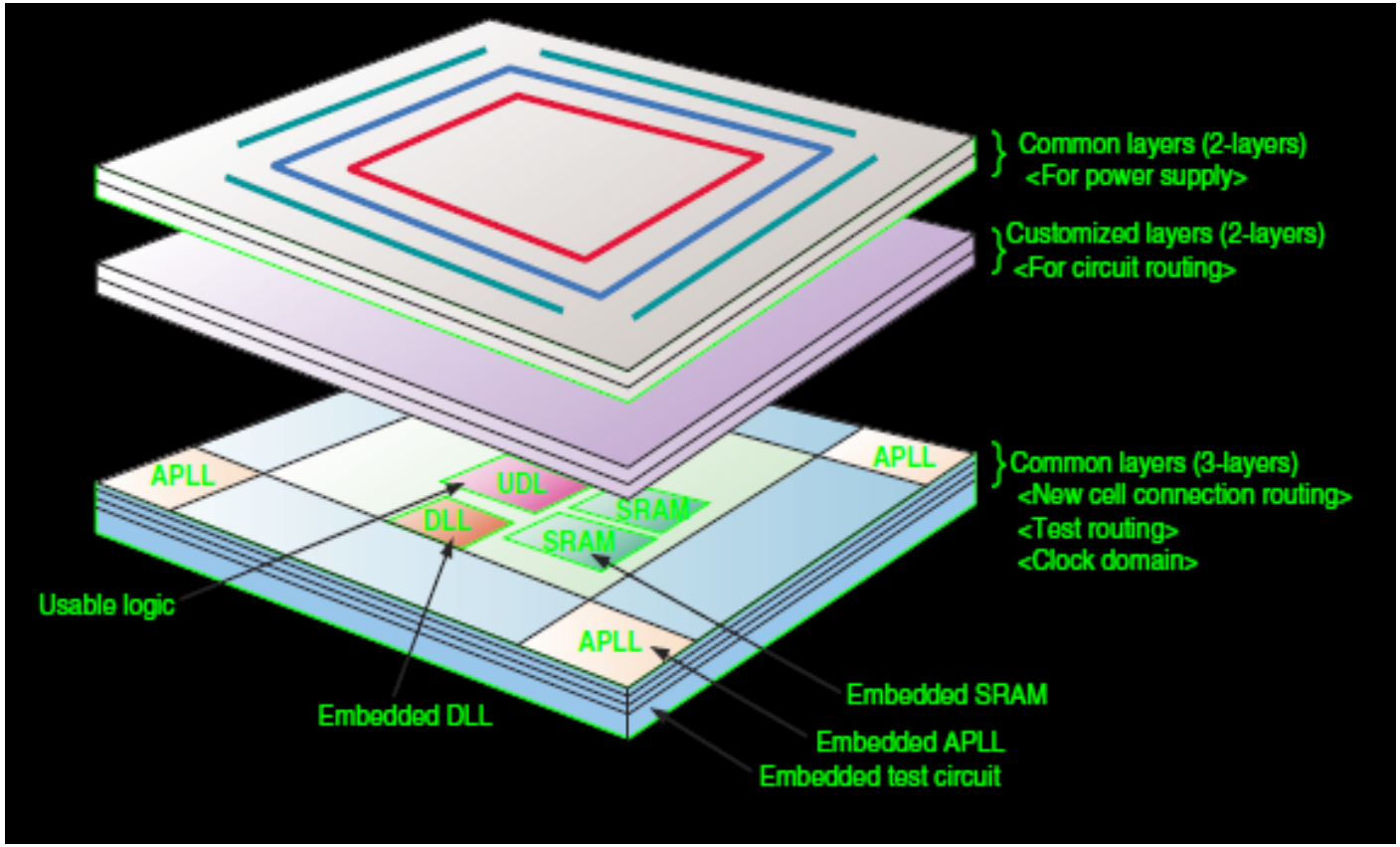
2024 © Centre for Development of Advanced Computing

NEC ISSP Structured ASIC

- Structured ASIC: “Instant Silicon Solution Platform”
- Pre-fabricated templates (Masters) with resources and IPs with two metal layers kept for user level customization
- Two lines: ISSP1 (150nm) and ISSP90 (90 nm)
  - STD and HSI variants
- User selects a particular template keeping in mind resource requirements (IOs, Logic, memory, high speed interfaces, pre-fabricated IPs etc)
- Design flow similar to classic FPGA/ASIC design cycles

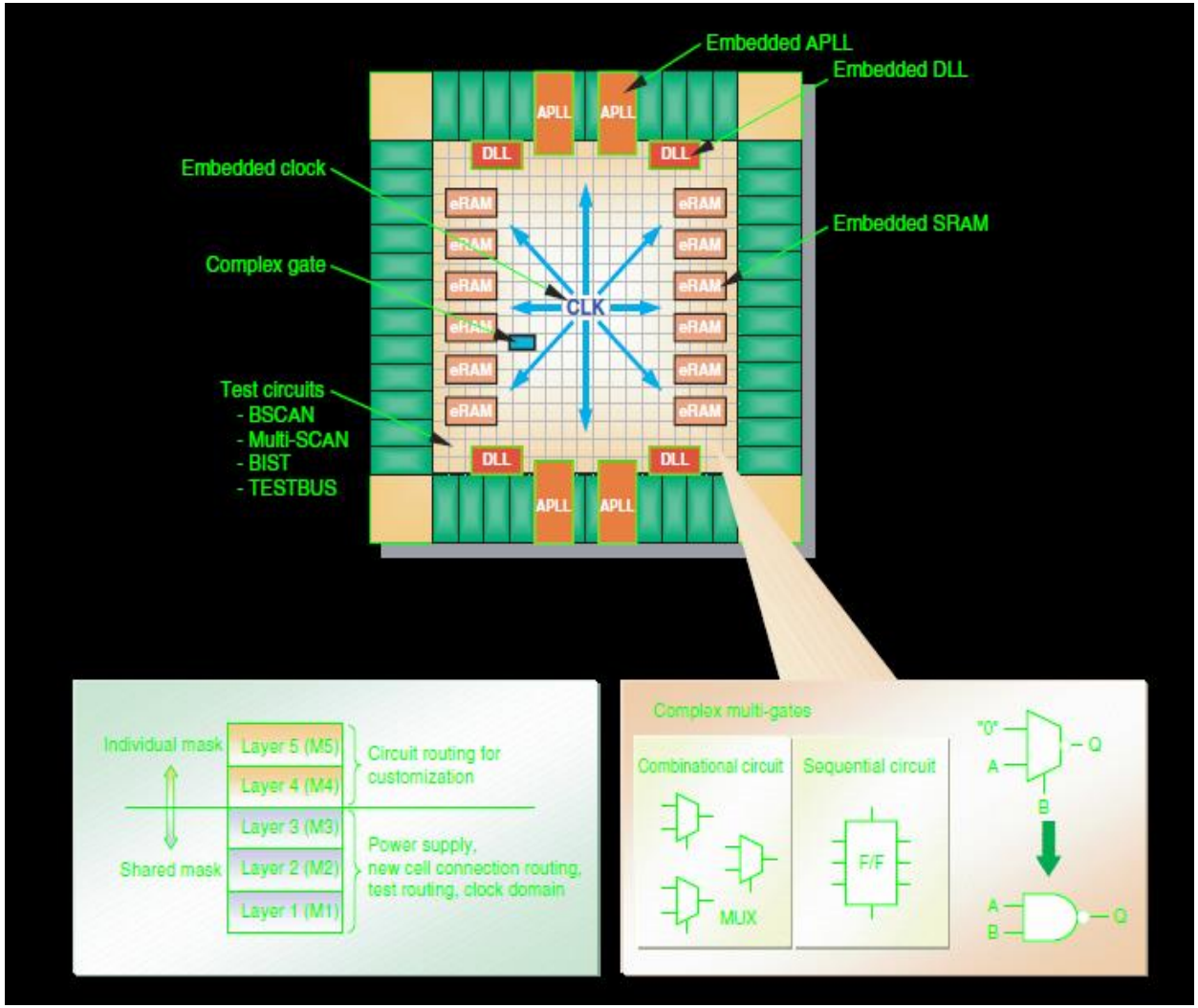
2024 © Centre for Development of Advanced Computing

NEC ISSP



2024 © Centre for Development of Advanced Computing

NEC ISSP



2024 © Centre for Development of Advanced Computing

- Advantages
  - Speed and area advantage of ASIC
  - FPGA turnaround times (foundry fab takes 2 wks typical)
  - Less NRE than ASIC
  - Number of IPs available from third party

2024 © Centre for Development of Advanced Computing

Technology Comparison

- Generally speaking
  - 100:33:1 ratio between the number of gates in a given area for standard cell ASIC's, structured ASIC's, and FPGA's, respectively
  - 100:75:15 ratio for performance (based on clock frequency)
  - 1:3:12 ratio for power

2024 © Centre for Development of Advanced Computing

Summary

- FPGAs architecture continue to evolve. Building blocks in today's FPGA offer a rich set of powerful, flexible resources to the developer
- Synthesis and backend tools are constantly evolving so as to hide platform specific details from the HDL developer
- Performance or area constraints require developers to be made aware of the underlying architecture so as to make optimal use of resources
- Logic and physical design cycles go hand in hand. Similar to the functional verification, many iterative cycles may be required in design flow before the design objectives are satisfied in the hardware
- FPGA design cycles are approaching ASIC design cycles for complex system designs. It is expected that FPGAs will continue proliferating into areas previously reserved for ASIC or other Custom chips
- New hybrid architectures such as structured ASIC may offer best of the FPGA/ASIC alternatives in near future

2024 © Centre for Development of Advanced Computing

Thank you

2024 © Centre for Development of Advanced Computing