BECE406E
FPGA BASED SYSTEM DESIGN

DEEPA M
52829
deepa.m@vit.ac.in

Syllabus

Module:1 Programmable Logic Devices

Module:2 HDL Fundamentals

Module:3 Implementation of Arithmetic system

Module:4 FSM and memory modelling

Module:5 Synthesis and Timing Analysis

Module:6 SoC Design

Module:7 FPGA Applications

Reference

- Wayne Wolf, “FPGA-Based System Design, Pearson Education”, 2004.
- Michael D Ciletti, “Advanced Digital Design with the Verilog HDL”, Prentice Hall, 2017.

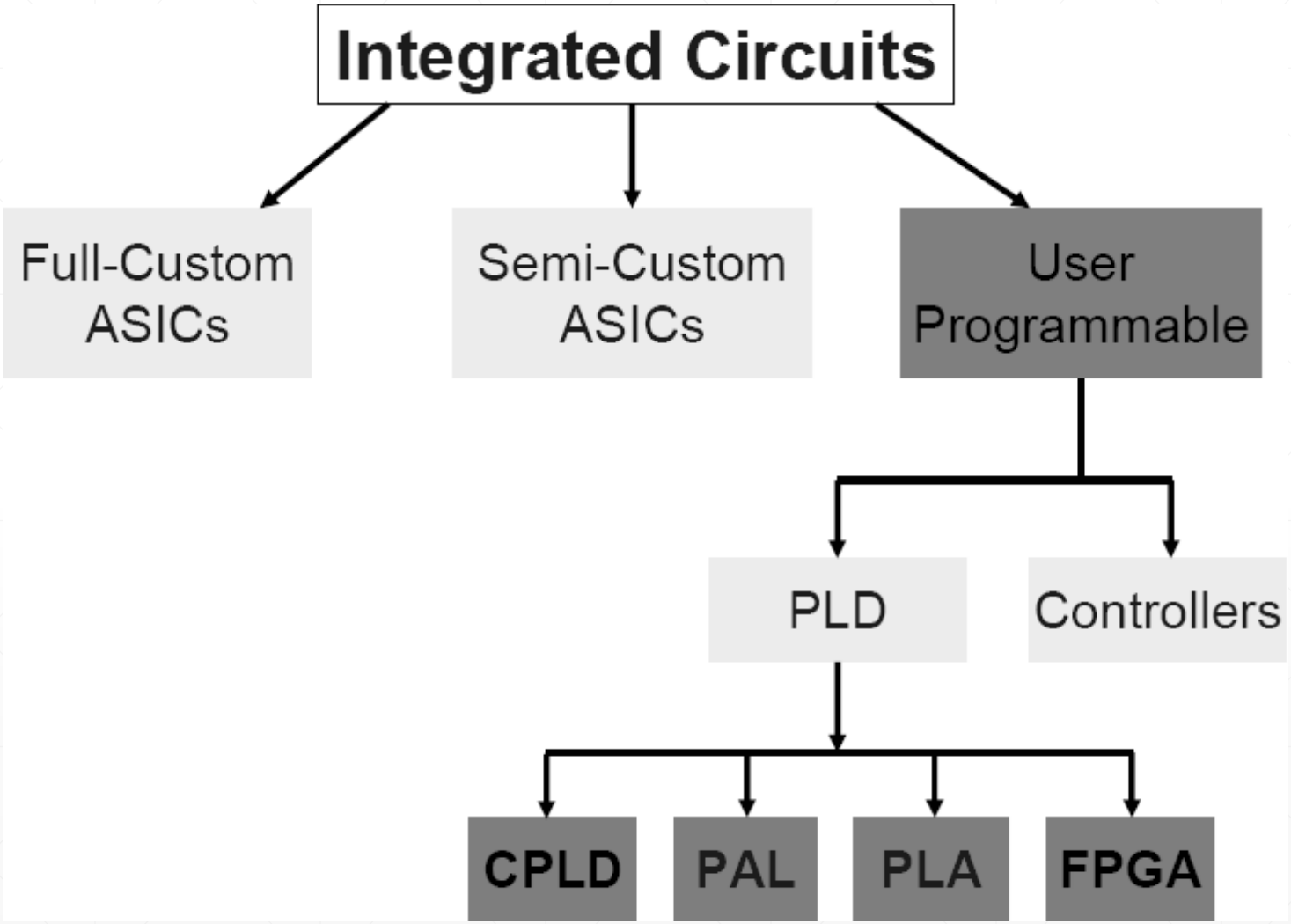
MODULE - 1

- Types of Programmable Logic Devices:
  - PLA, PAL, CPLD, FPGA
- FPGA Architecture
  - Programming Technologies
  - Chip I/O
  - Programmable Logic Blocks
  - Fabric and Architecture of FPGA

Why VLSI?

- Integration improves the design:
  - higher speed;
  - lower power;
  - physically smaller.
- Integration reduces manufacturing cost (almost) – no manual assembly.

Integrated Circuits



Full Custom ICs

- Can achieve very high transistor density (transistors per square micron)
- design time can be very long (multiple months).
- Involves the creation of a completely new chip, which consists of masks (for the photolithographic manufacturing process)
- Benefits - Excellent performance, small size, low power

Standard Cell

- Designer uses a library of standard cells
- an automatic place and route tool does the layout
- *Transistor density* and performance degradation depends on type of design being done.
- Design time can be much faster than full custom because layout is automatically generated.

Gate Array

- Designer uses a library of standard cells.
- The design is mapped onto an array of transistors which is already created on a wafer
- wafers with transistor arrays can be created ahead of time
- A *routing tool* creates the masks for the routing layers and "customizes" the pre-created gate array for the user's design
- Transistor density can be almost as good as standard cell.
- Design time advantages are the same as for standard cell.

Semi-custom ICs

- Flexible as portion of the IC is customized by the user
- Suitable for specific applications
- Gate array + standard cell
- Paves way for application specific ICs (ASIC)

Role of FPGA

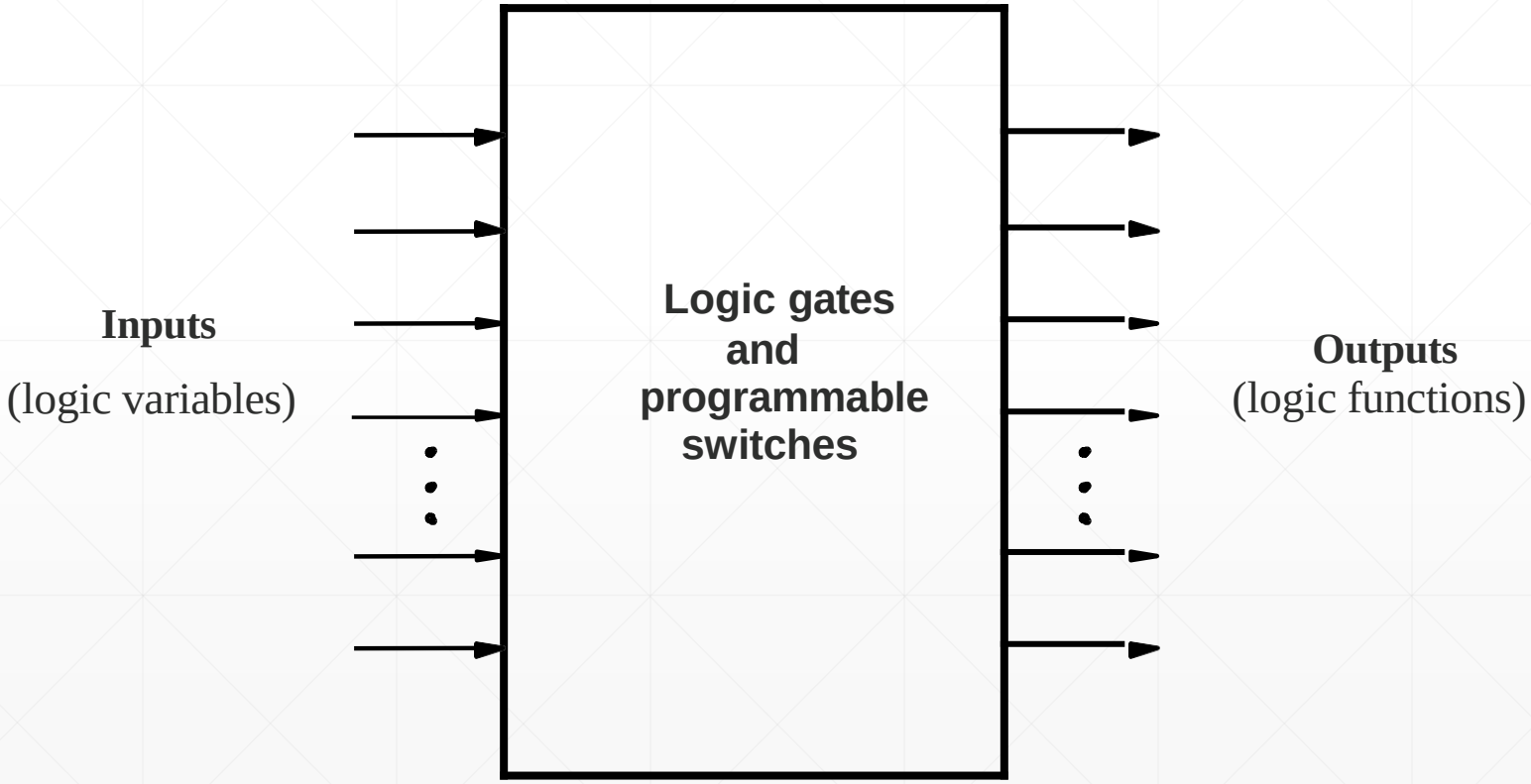
- Microprocessors used in variety of environments
  - Rely on software to implement functions
  - Generally slower and more power-hungry than custom chips
- When FPGAs?
  - Design economics
    - Shortest time to market
    - Lowest NRE cost
    - Highest unit cost
  - Make quick grab for market share
  - Same FPGA reused in several designs
  - Excellent prototyping, prototype to product --- much simple & easy to negotiate

Programmable logic devices

- Programmable Logic Device (PLD)
  - In market since 1970s.
  - An integrated circuit chip that can be configured by end user to implement different digital hardware.
  - Uses two-level logic structures to implement programmed logic.
  - First level – AND plane (fixed), Second level – OR plane (programmable)
  - Generally programmed by antifuses (large voltages to make connections)
  - Relatively useful for small logic functions.
  - For large logic functions, multi-level logic is essential.

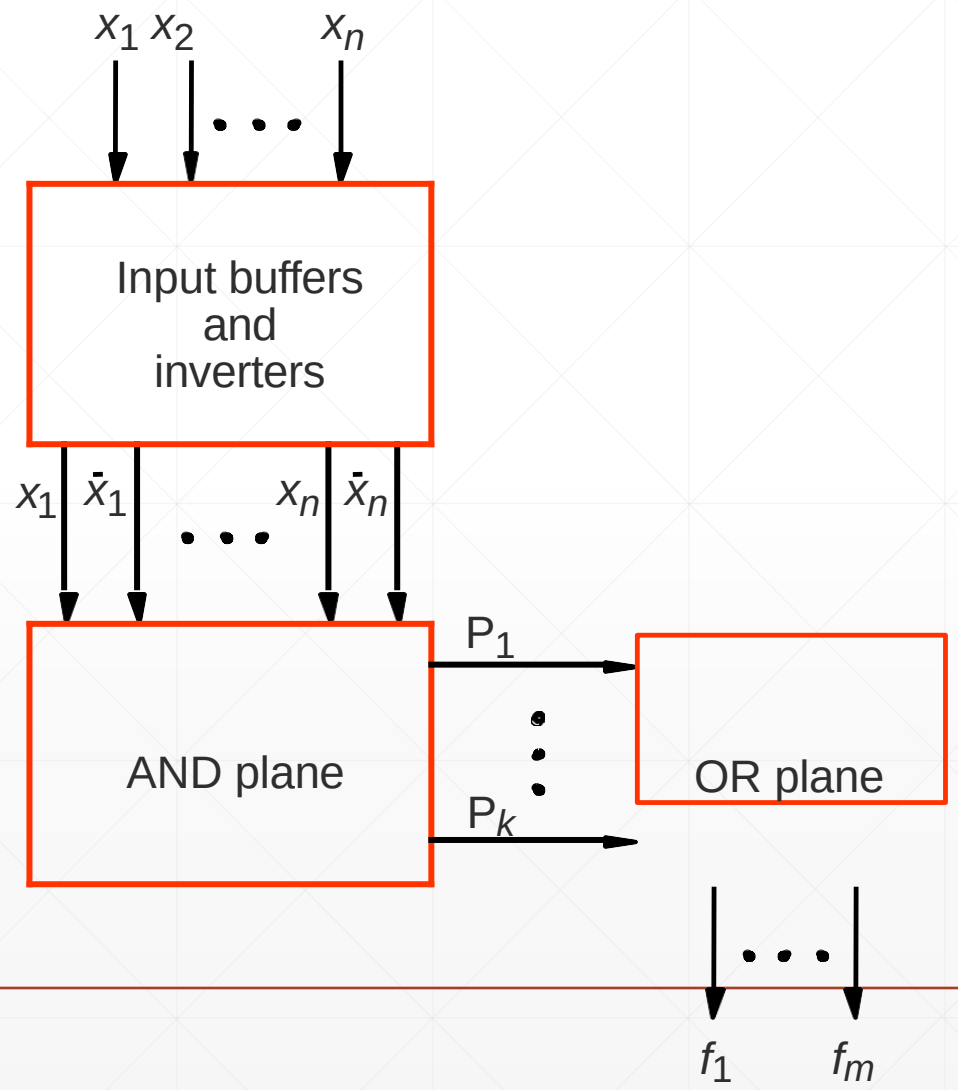
PLD

- PLD as a Black Box



Programmable Logic Array (PLA)

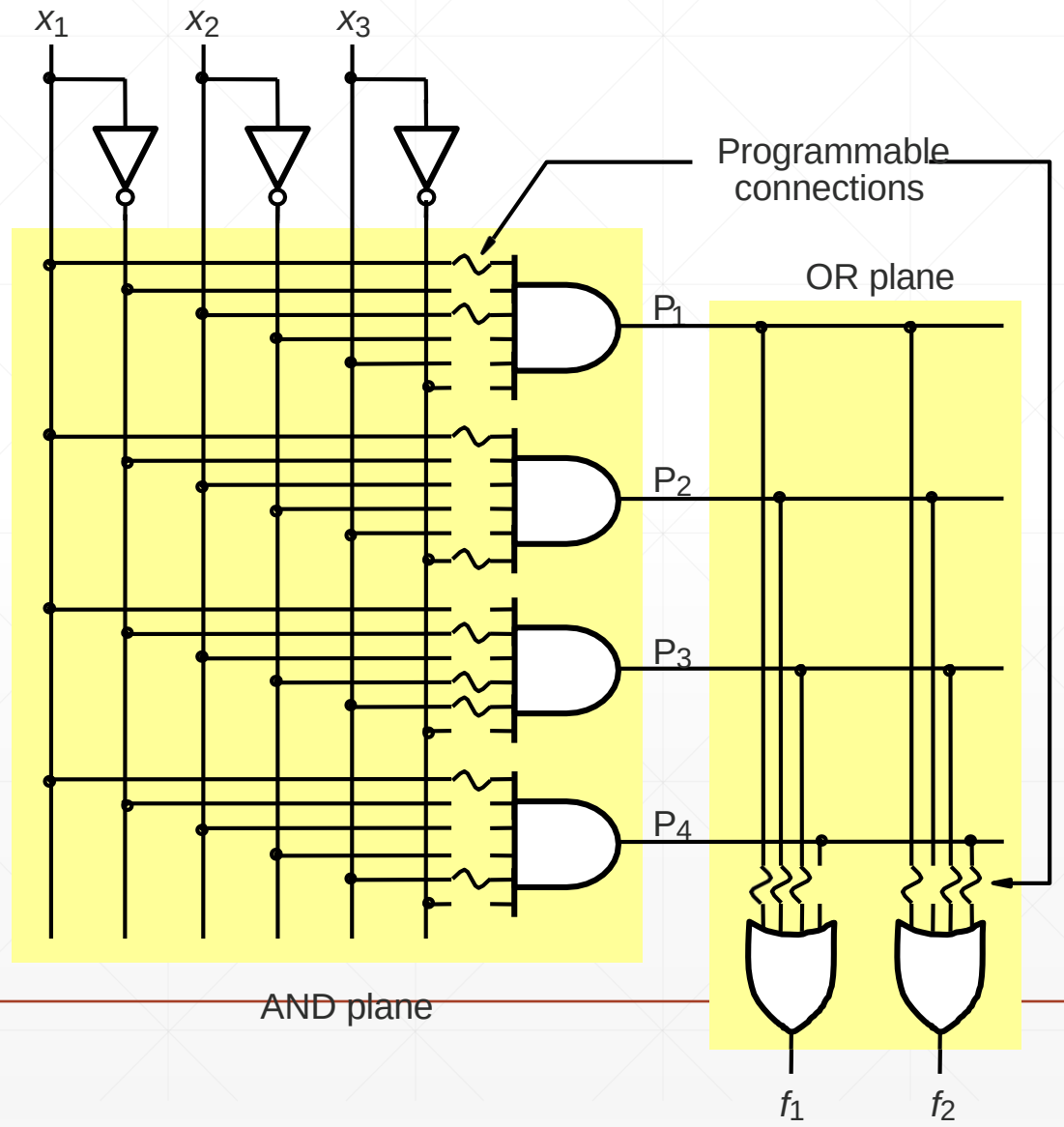
- Use to implement circuits in SOP form
- The connections in the AND plane are programmable
- The connections in the OR plane are programmable



Gate Level Version of PLA

$$f_1 = x_1x_2+x_1x_3'+x_1'x_2'x_3$$

$$f_2 = x_1x_2+x_1'x_2'x_3+x_1x_3$$

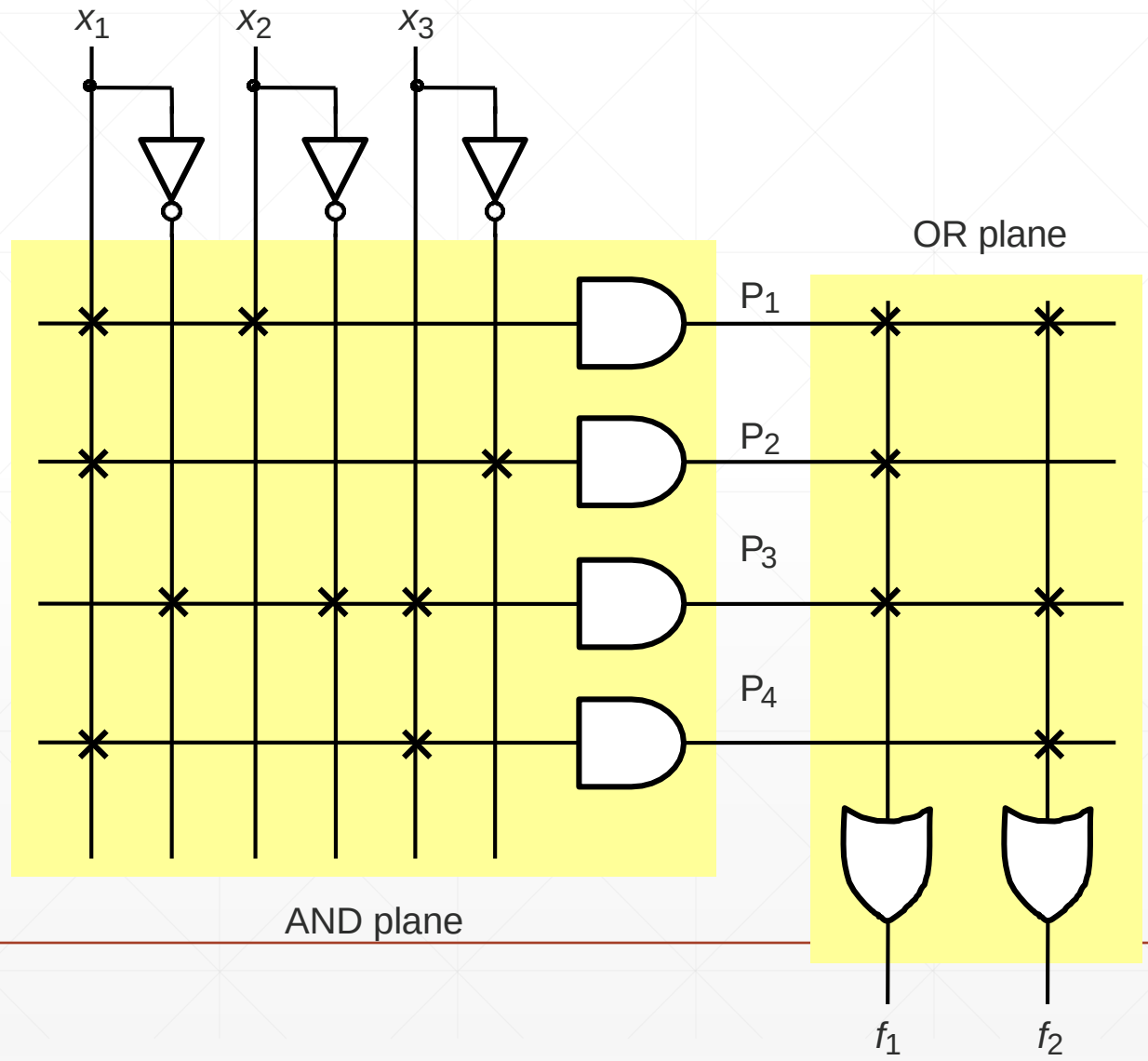


Customary Schematic of a PLA

$f_1 = x_1x_2+x_1x_3'+x_1'x_2'x_3$

$f_2 = x_1x_2+x_1'x_2'x_3+x_1x_3$

x marks the connections left in place after programming

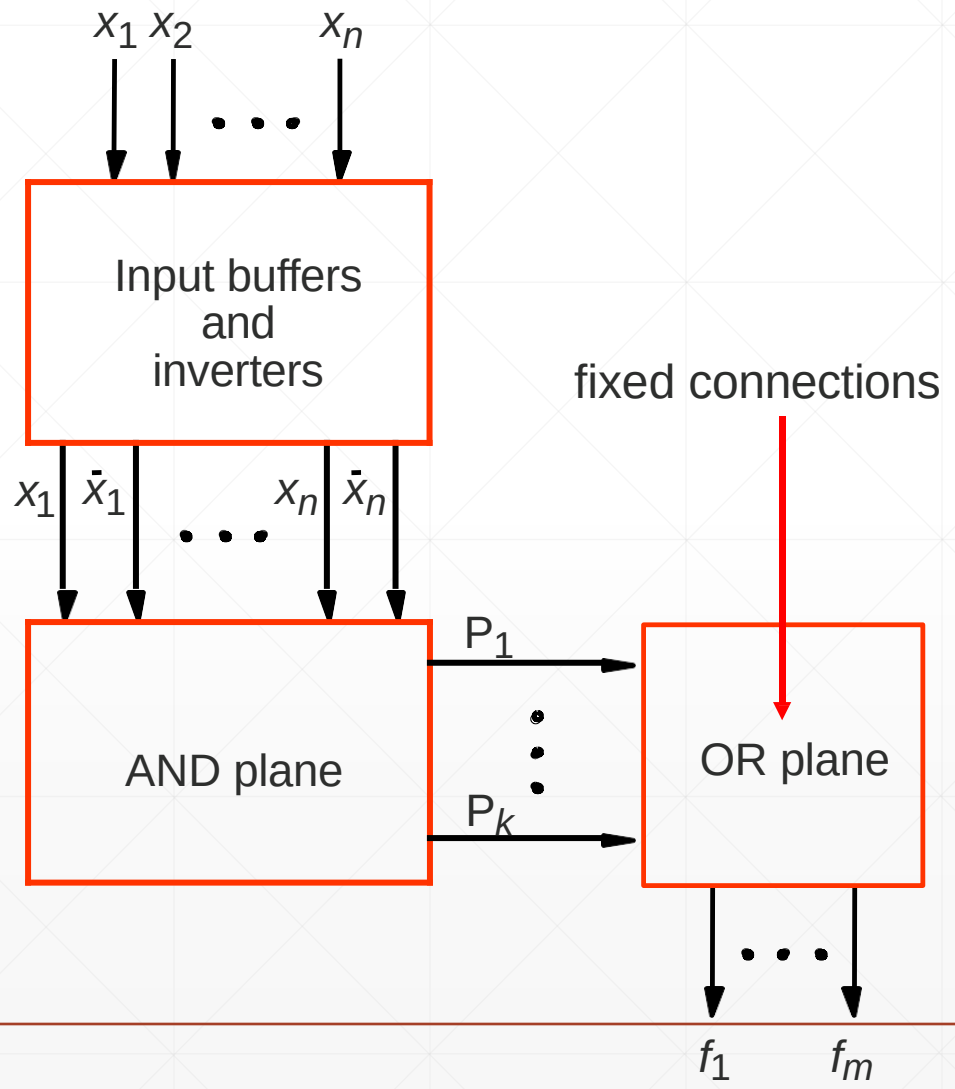


Limitations of PLAs

- Typical size is 16 inputs, 32 product terms, 8 outputs
  - Each AND gate has large fan-in - this limits the number of inputs that can be provided in a PLA
  - 16 inputs 2^{16} = possible input combinations; only 32 permitted (since 32 AND gates) in a typical PLA
  - 32 AND terms permitted large fan-in for OR gates as well
- This makes PLAs slower and slightly more expensive than some alternatives to be discussed shortly

Programmable Array Logic (PAL)

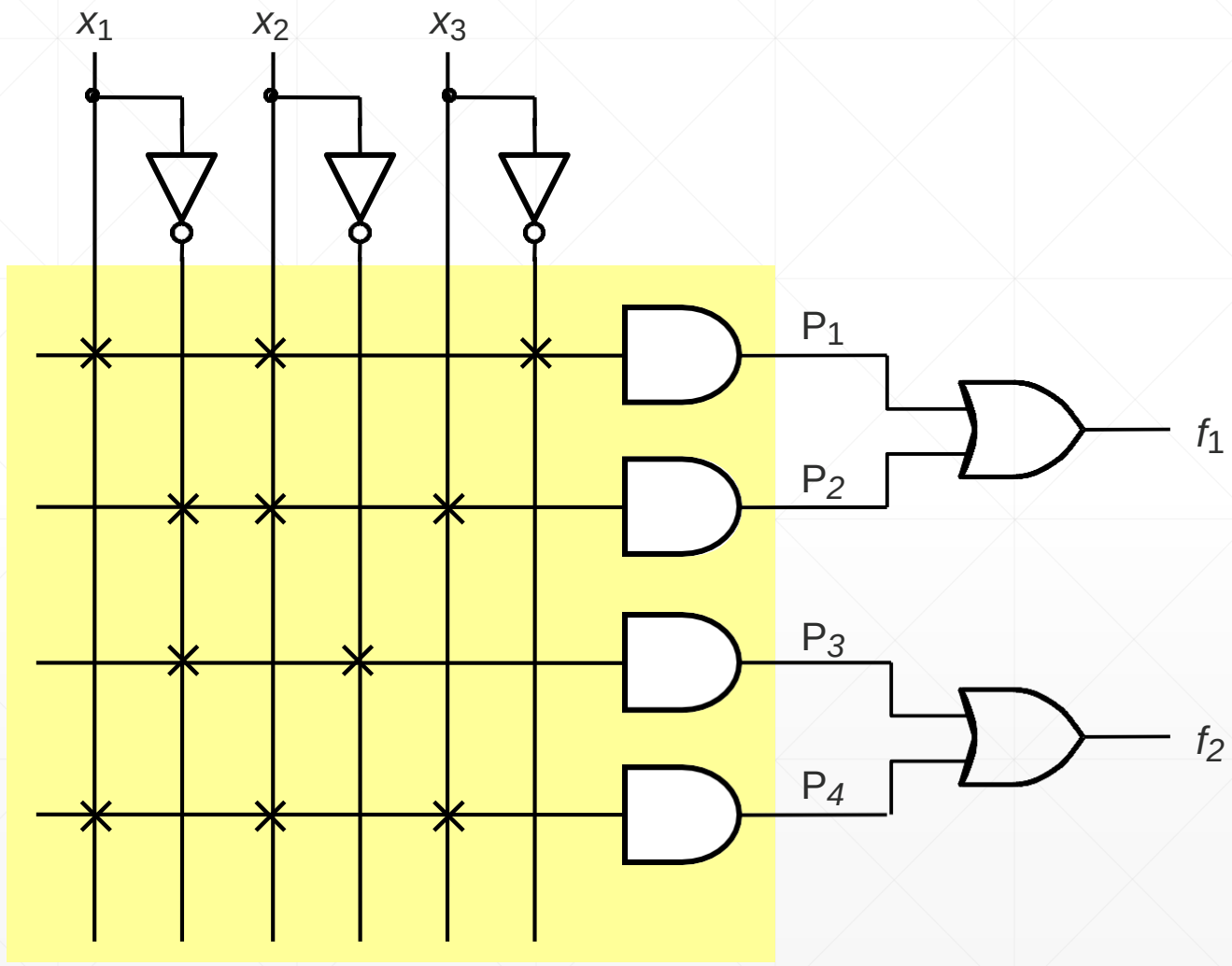
- Also used to implement circuits in SOP form
- The connections in the AND plane are programmable
- The connections in the OR plane are NOT programmable



Example Schematic of a PAL

$f_1 = x_1x_2x_3'+x_1'x_2x_3$

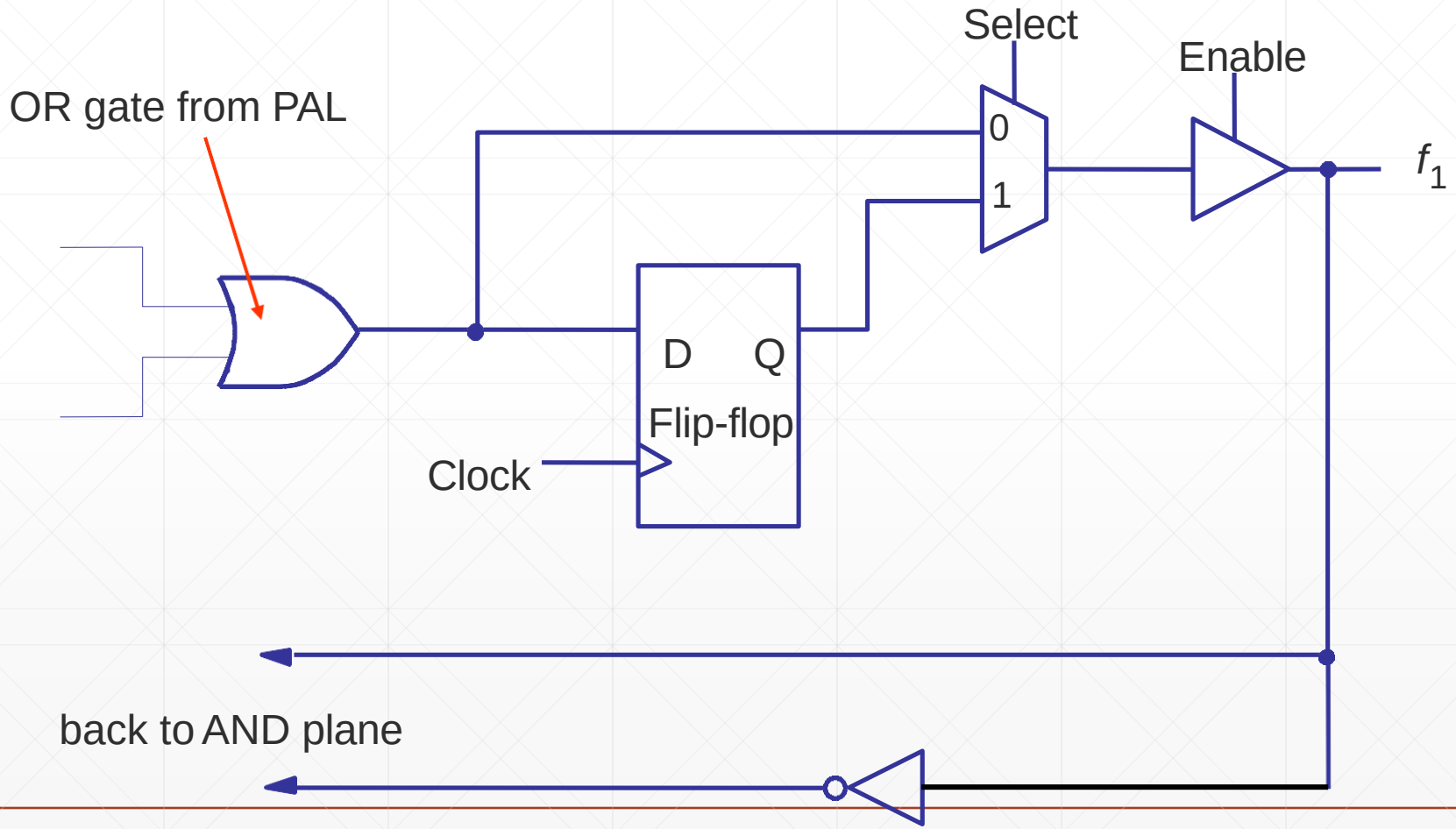
$f_2 = x_1'x_2'+x_1x_2x_3$



AND plane

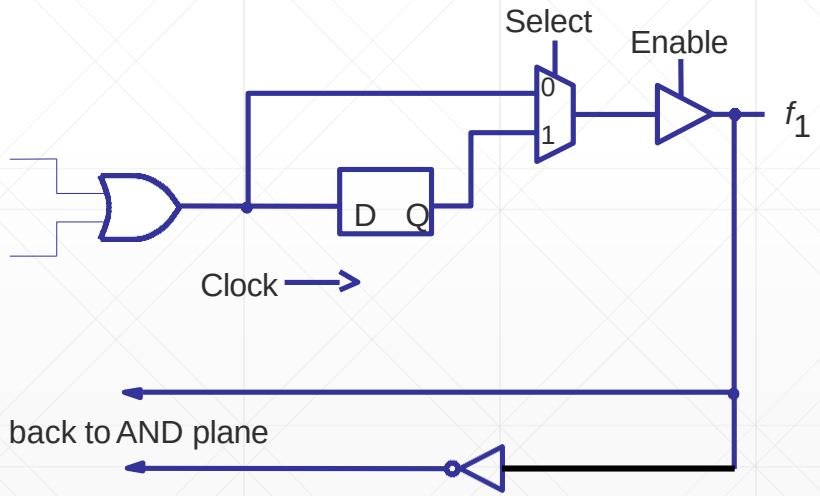
- Comparing PALs and PLAs
  - PALs have the same limitations as PLAs (small number of allowed AND terms) plus they have a fixed OR plane less flexibility than PLAs
  - PALs are simpler to manufacture, cheaper, and faster (better performance)
  - PALs also often have extra circuitry connected to the output of each OR gate
    - The OR gate plus this circuitry is called a *macrocell*

- Macrocell


Select
Enable
OR gate from PAL
0
1
f_1
D
Q
Flip-flop
Clock
back to AND plane

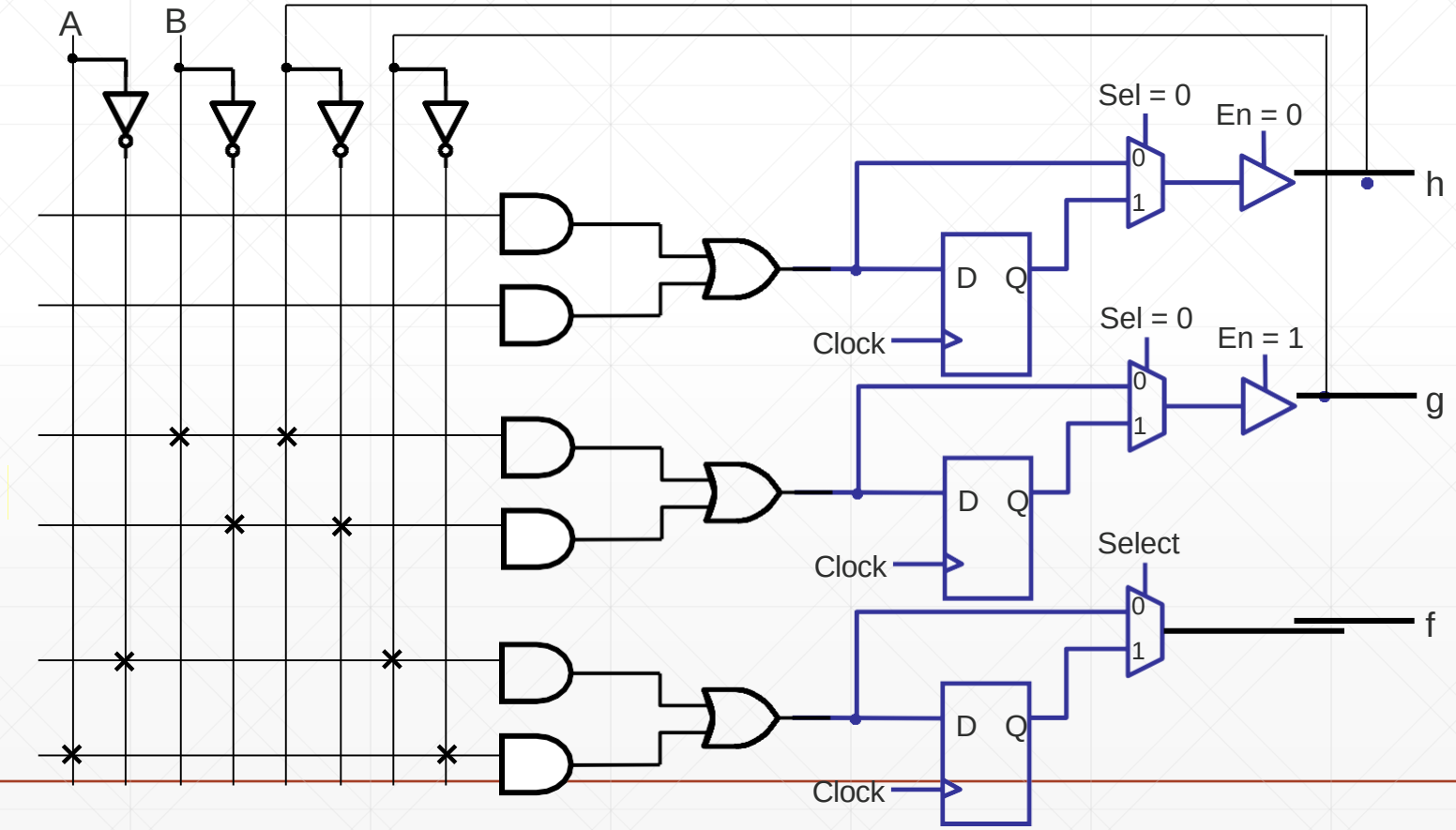
- Macrocell Functions

  - Enable = 0 can be used to allow the output pin for f_1 to be used as an additional input pin to the PAL.
  - Enable = 1, Select = 0 is normal for typical PAL operation.
  - Enable = Select = 1 allows the PAL to synchronize the output changes with a clock pulse.
  - The feedback to the AND plane provides for multi-level design

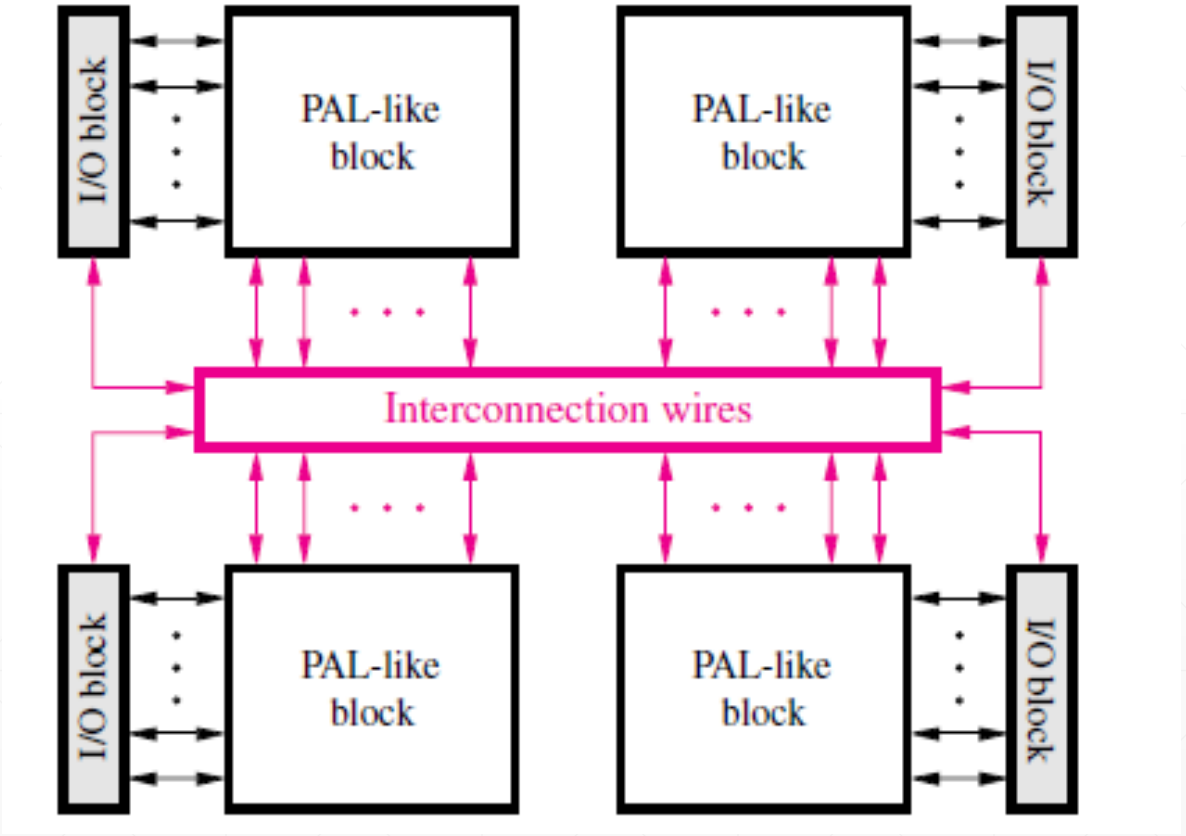


Multi-Level Design with PALs

- f = A'BC + A'B'C' + ABC' + AB'C = A'g + Ag'
  - where g = BC + B'C' and C = h below

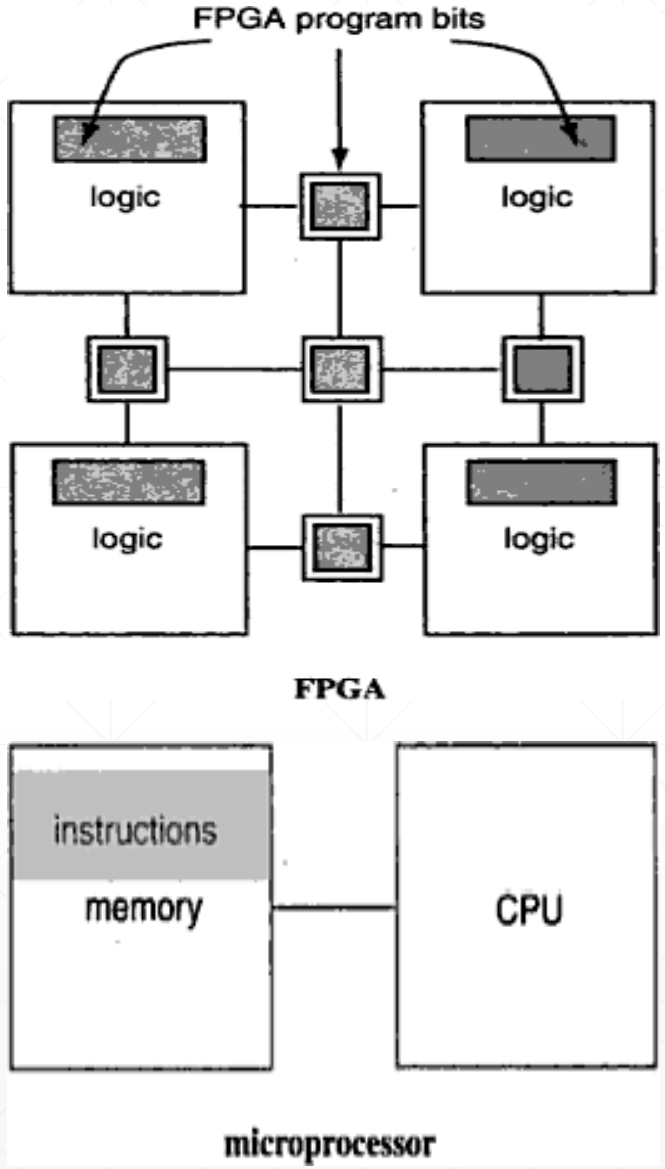


Complex Programmable Logic Devices (CPLDs)



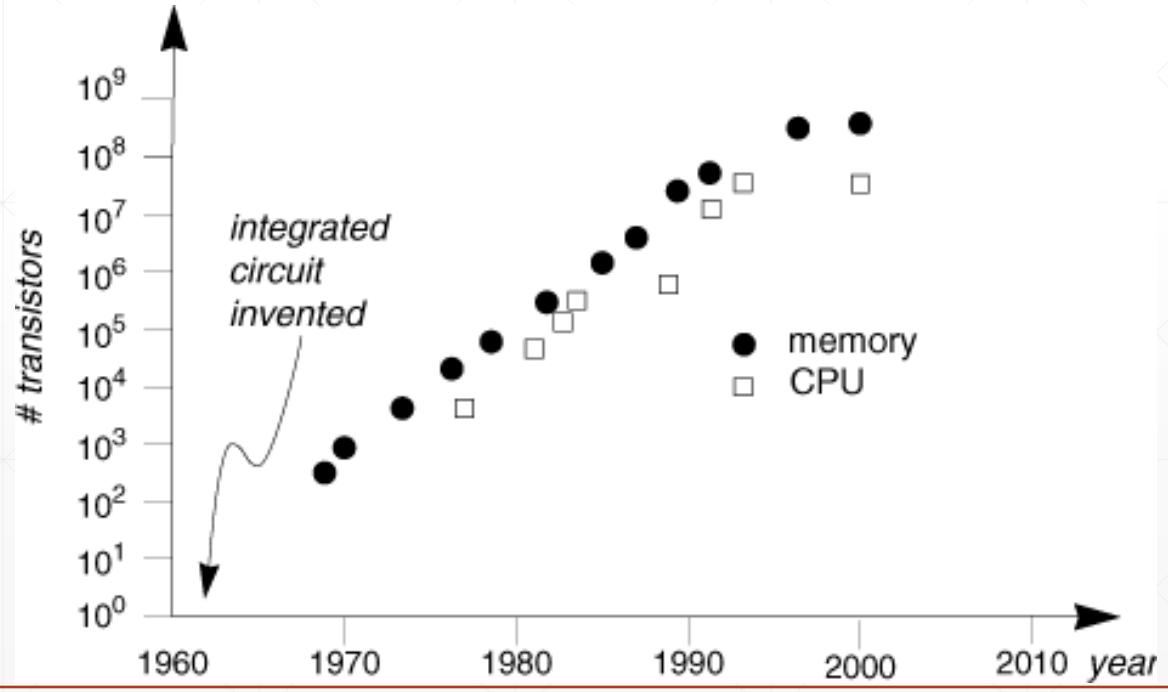
FPGA Programming

- FPGAs implement multi-level logic
- Need both programmable logic blocks and programmable interconnect
- Combination of logic and interconnect is fabric
- Microprocessor is a stored-program computer



Moore's Law

- Gordon Moore: co-founder of Intel.
- Predicted that number of transistors per chip would grow exponentially (double every 18 months).



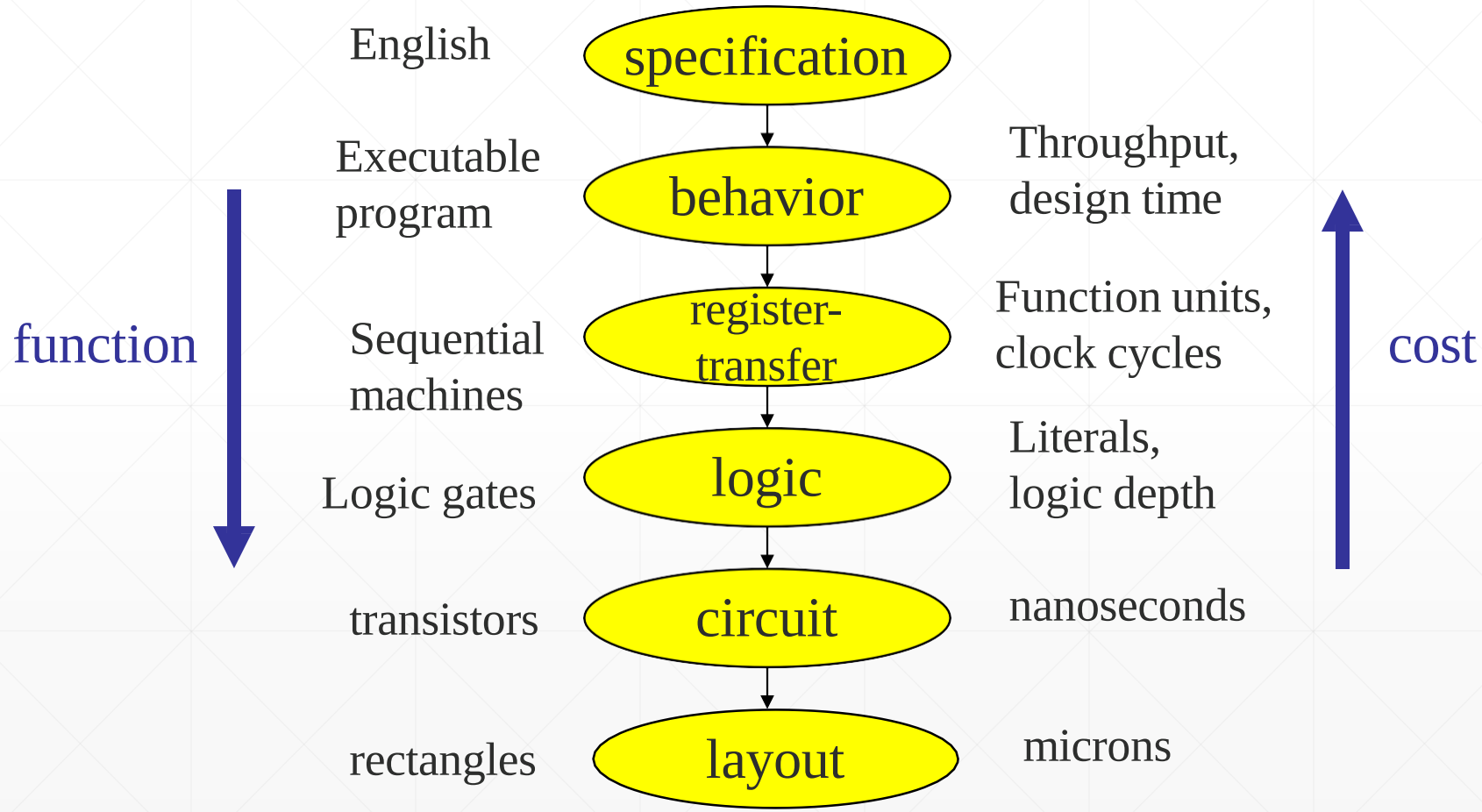
Goals and Techniques

- Performance
  - Logic run at a required rate
  - Throughput, Latency, Clock rate – performance measures
- Power/energy
- Design time
  - FPGAs are standard cells
  - Prototypes can be programmed quickly
- Design cost
  - FPGA tools less expensive than custom VLSI tools
- Manufacturing cost
  - Cost of replicating the system many times
  - FPGAs are relatively high cost than ASICs

Design Challenges

- Multiple levels of abstraction
- Power consumption
- Short design time

FPGA Abstractions



- Top-down design adds functional detail.
  - Create lower levels of abstraction from upper levels.
- Bottom-up design creates abstractions from low-level behavior.
  - More accurate delay information.
- Good design needs cycles of top-down approach followed by bottom-up efforts.

Methodology

- Hardware Description logic (HDL)
  - VHDL
  - VerilogHDL
- Synthesis Tool

Major FPGA Vendors

- SRAM-based FPGAs
  - **Xilinx, Inc**
  - **Altera Corp.**
  - Atmel
  - Lattice Semiconductor

Share 80% of the market

- Flash & Antifuse FPGAs
  - Actel Corp.
  - Quick logic Corp.

FPGA Vendors and Device families

- Xilinx
  - Spartan
  - Virtex
  - Kintex
  - Artix
- Altera
  - Stratix
  - Cyclone
  - MAX 3000/7000 CPLD
  - MAX-II

Xilinx Families

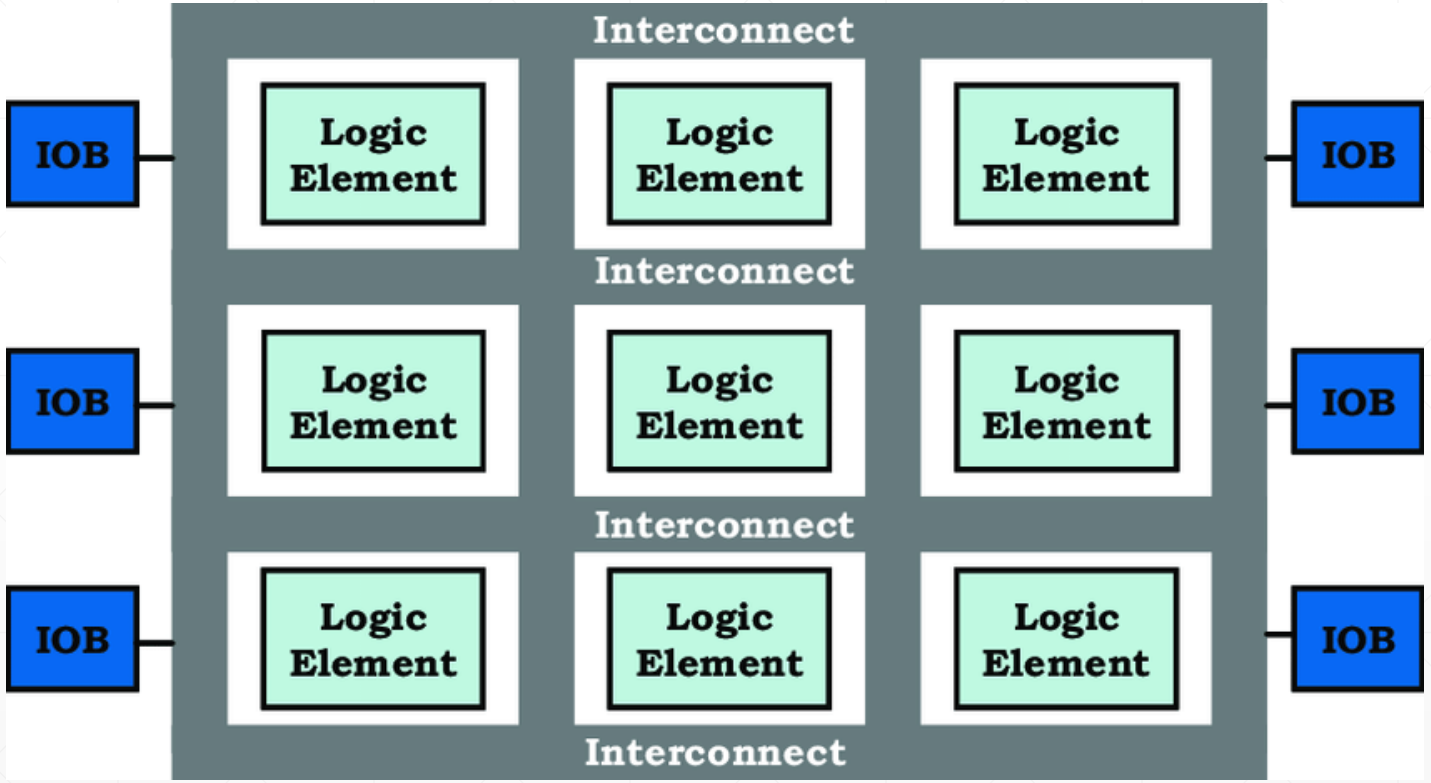
- **Old families**
  - XC3000, XC4000, XC5200
  - Old 0.5µm, 0.35µm and 0.25µm technology. Not recommended for modern designs.
- **High-performance families**
  - Virtex (0.22µm)
  - Virtex-E, Virtex-EM (0.18µm)
  - Virtex-II, Virtex-II PRO (0.13µm)
  - Virtex-4 (0.09µm)
- **Low Cost Family**
  - Spartan/XL – derived from XC4000
  - Spartan-II – derived from Virtex
  - Spartan-IIE – derived from Virtex-E
  - Spartan-3 – derived from Virtex-II

Altera Families

- **Old Families**
  - FLEX 10K, FLEX 6000, FLEX 8000
- **High-performance Families**
  - Mercury
  - Stratix, Stratix GX, Stratix II
  - APEX 20K , APEX II
  - Excalibur
- **Low Cost Family**
  - Cyclone, Cyclone II

FPGA FABRICS

- **3 major types of elements**
  - Combinational logic
  - Interconnect
  - I/O pins



FPGA Architectures

- **Combinational logic**
- The combinational logic is divided into relatively small units which may be known as logic elements (LEs) or combinational logic blocks (CLBs).
- The LE or CLB can usually form the function of several typical logic gates but it is still small compared to the typical combinational logic block found in a large design.

FPGA Architectures

- **Interconnects:**
- The interconnections are made between the logic elements using programmable interconnect.
- The interconnect may be logically organized into channels or other units.
- FPGAs typically offer several types of interconnect depending on the distance between the combinational logic blocks that are to be connected; clock signals are also provided with their own interconnection networks.

FPGA Architectures

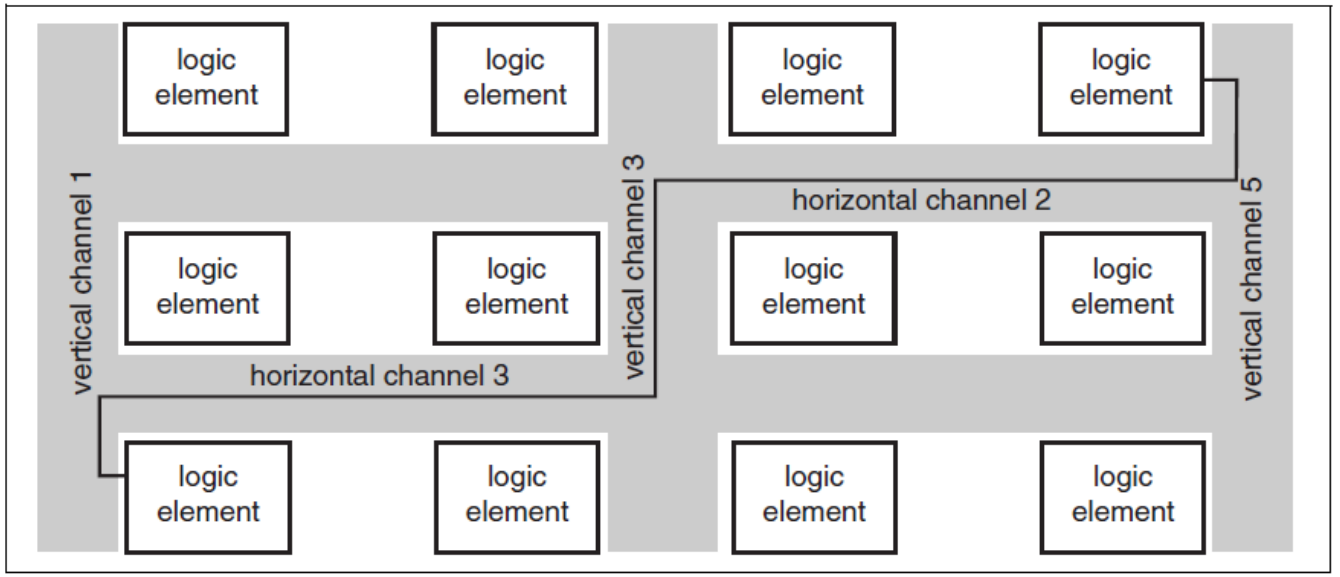
- **I/O pins:**
- The I/O pins may be referred to as I/O blocks (IOBs).
- They are generally programmable to be inputs or outputs and often provide other features such as low-power or high-speed connections.

FPGA Architectures

- **I/O pins:**
- The I/O pins may be referred to as I/O blocks (IOBs).
- They are generally programmable to be inputs or outputs and often provide other features such as low-power or high-speed connections.

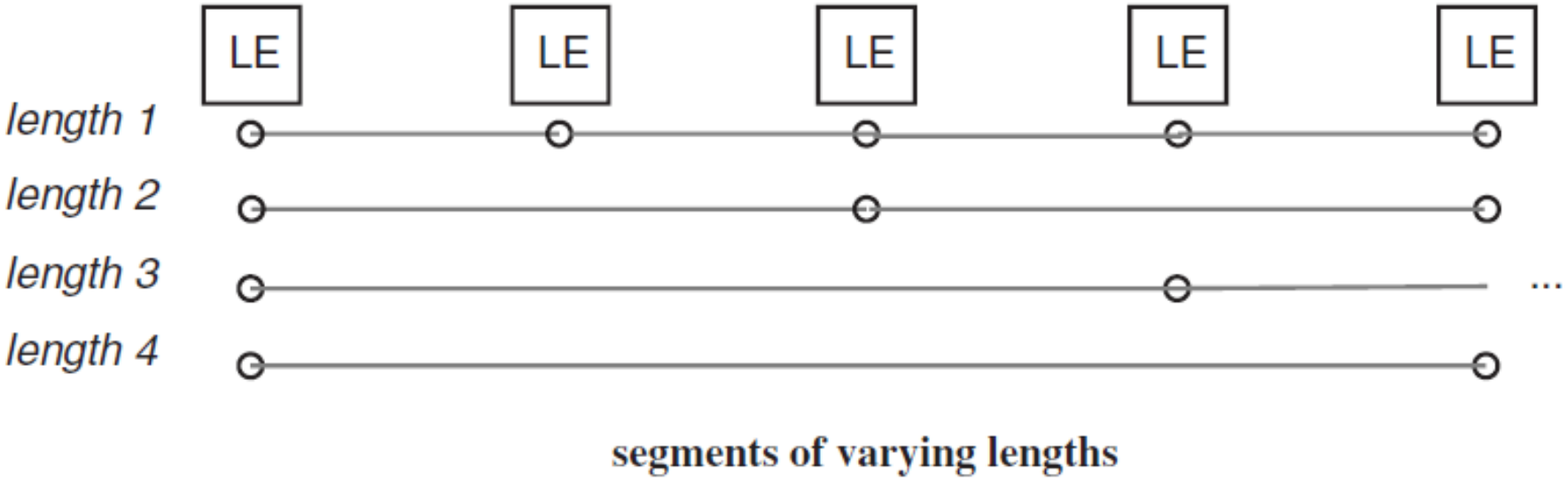
FPGA interconnect

- An FPGA designer must rely on pre-designed wiring, unlike a custom VLSI designer.
- LEs are arranged in some sort of 2D structure, so interconnect may require complex paths.
- Wires are typically organized in wiring channels or routing channels that run horizontally and vertically through the chip.
- Connections must be made between wires in order to carry a signal from one point to another.



Segmented wiring

- FPGA channels provide wires of a variety of lengths.

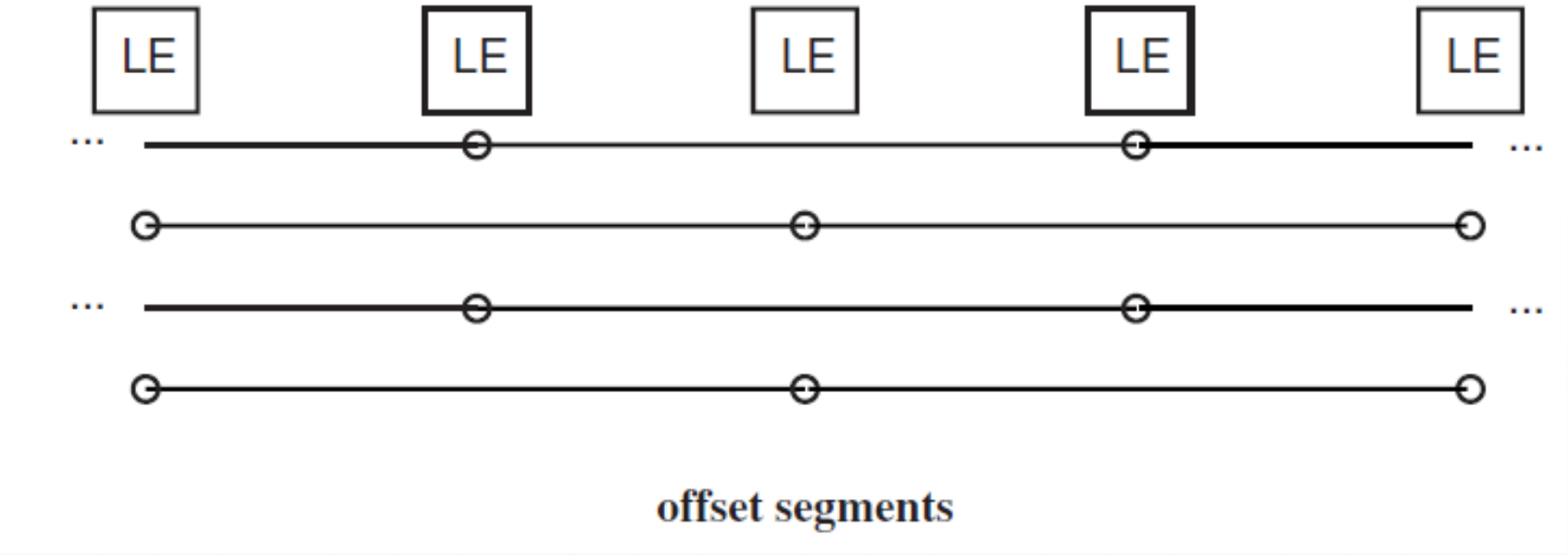


segments of varying lengths

- The figure shows connections of varying length as measured in units of LEs:
  - the top signal of length 1 goes to the next LE,
  - the second signal goes to the second LE, and so on

Offset wiring

- The alternative to segmented wiring is to make each wire length 1.
- The segment in a group need not all end at one point.



offset segments

FPGA configuration

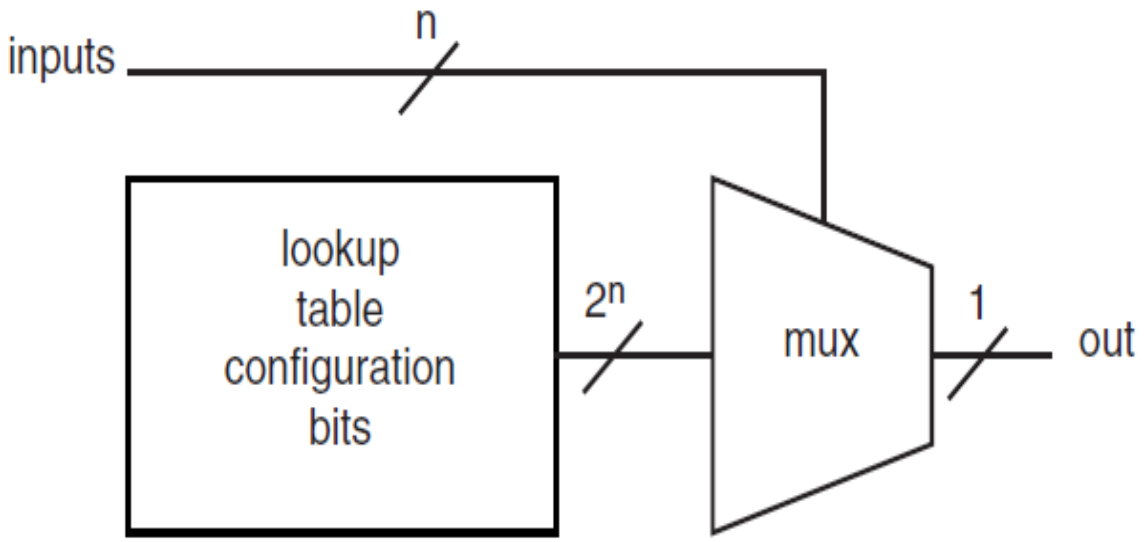
- 3 major circuit technologies for configuring an FPGA:
  - SRAM
  - Antifuse
  - Flash
- Characteristics of interest to the system designer:
  - How much logic can I fit into this FPGA?
  - How many I/O pins does it have?
  - How fast does it run?

SRAM-Based FPGAs

- Characteristics of SRAM-based FPGAs
  - SRAM-based FPGAs hold their configurations in static memory
  - The output of the memory cell is directly connected to another circuit and the state of the memory cell continuously controls the circuit being configured.
- Advantages:
  - FPGA can be easily reprogrammed
  - dynamically reconfigurable systems
  - fabricated with standard VLSI processes
- Disadvantages: Power & Security
- Each combinational logic element requires many programming bits.
- Each programmable interconnection point requires its own bit.

Logic Elements

- The basic method used to build a combinational logic block (CLB)— also called a logic element or LE—in an SRAM-based FPGA is the **lookup table (LUT)**.
- LUT is an SRAM that is used to implement a truth table.
- Each address in the SRAM represents a combination of inputs to the logic element.
- The value stored at that address represents the value of the function for that input combination.

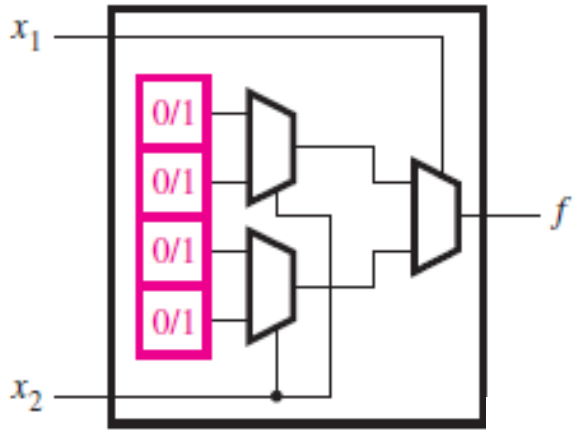


LUT

Logic Elements

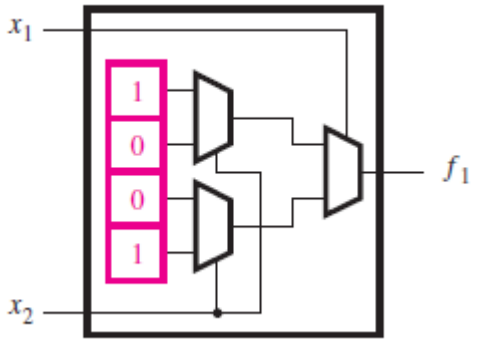
- An n-input function requires an SRAM with 2^n locations.
- A basic SRAM is not clocked --- as its inputs change, its output changes after some delay.
- A typical logic element has four inputs.
- The delay through the lookup table is independent of the bits stored in the SRAM, so the delay through the logic element is the same for all functions.
- For example, a lookup table based LE will exhibit the same delay for a 4-input XOR and a 4-input NAND.

Circuit for a two-input LUT

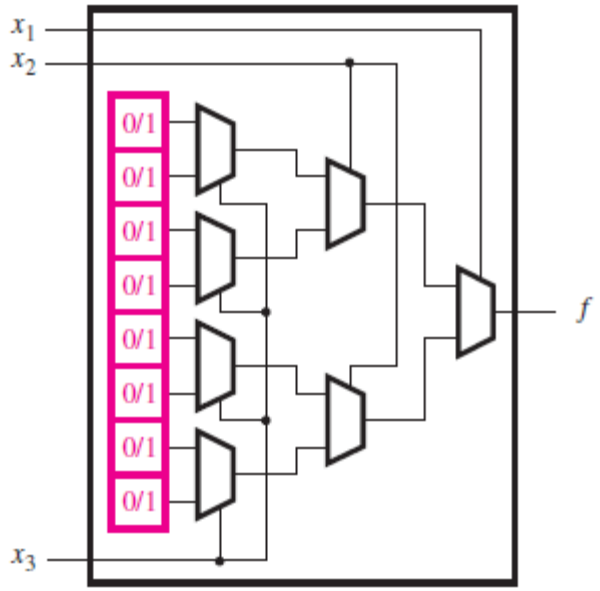


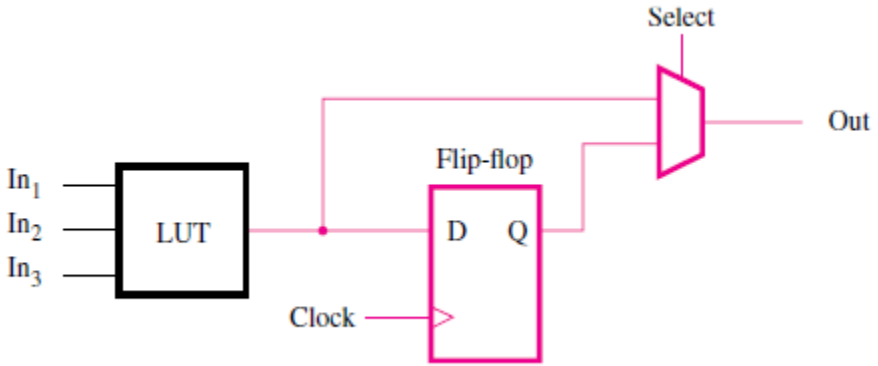
x_1	x_2	f_1
0	0	1
0	1	0
1	0	0
1	1	1

(b) $f_1 = \bar{x}_1\bar{x}_2 + x_1x_2$

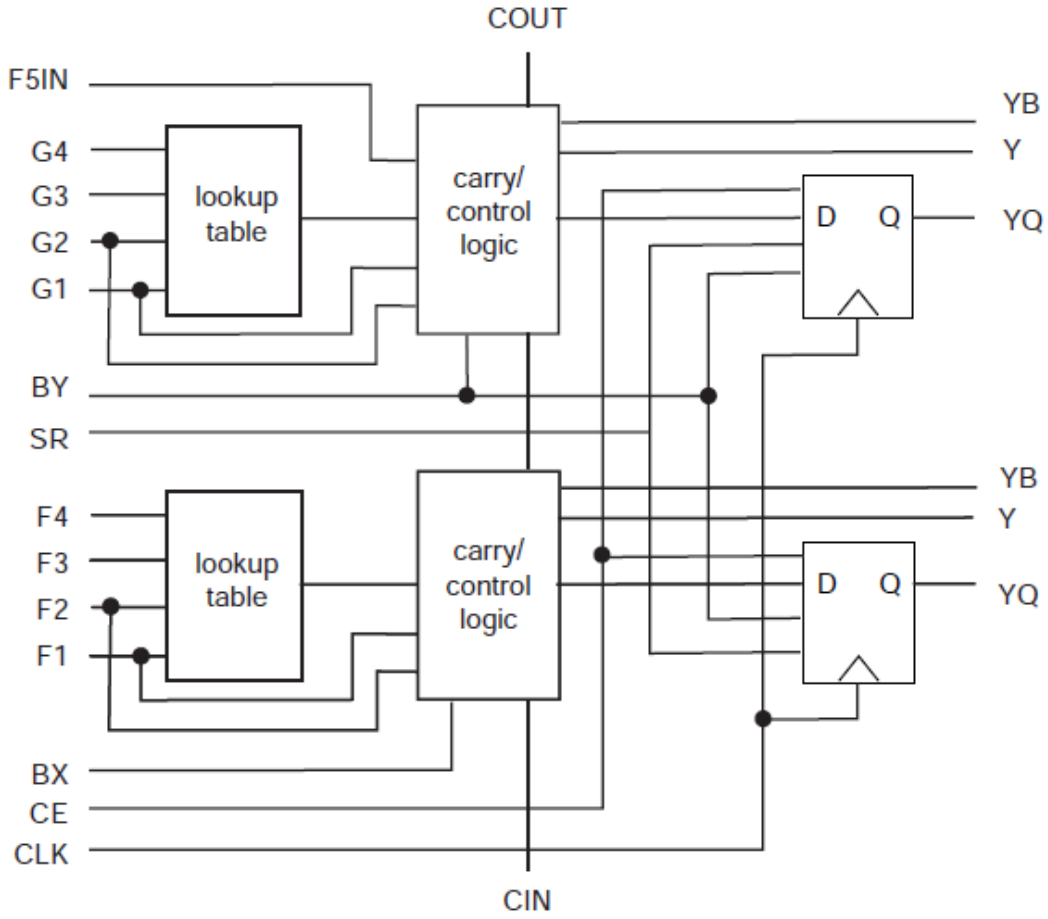


A three-input LUT





Xilinx Spartan-II CLB

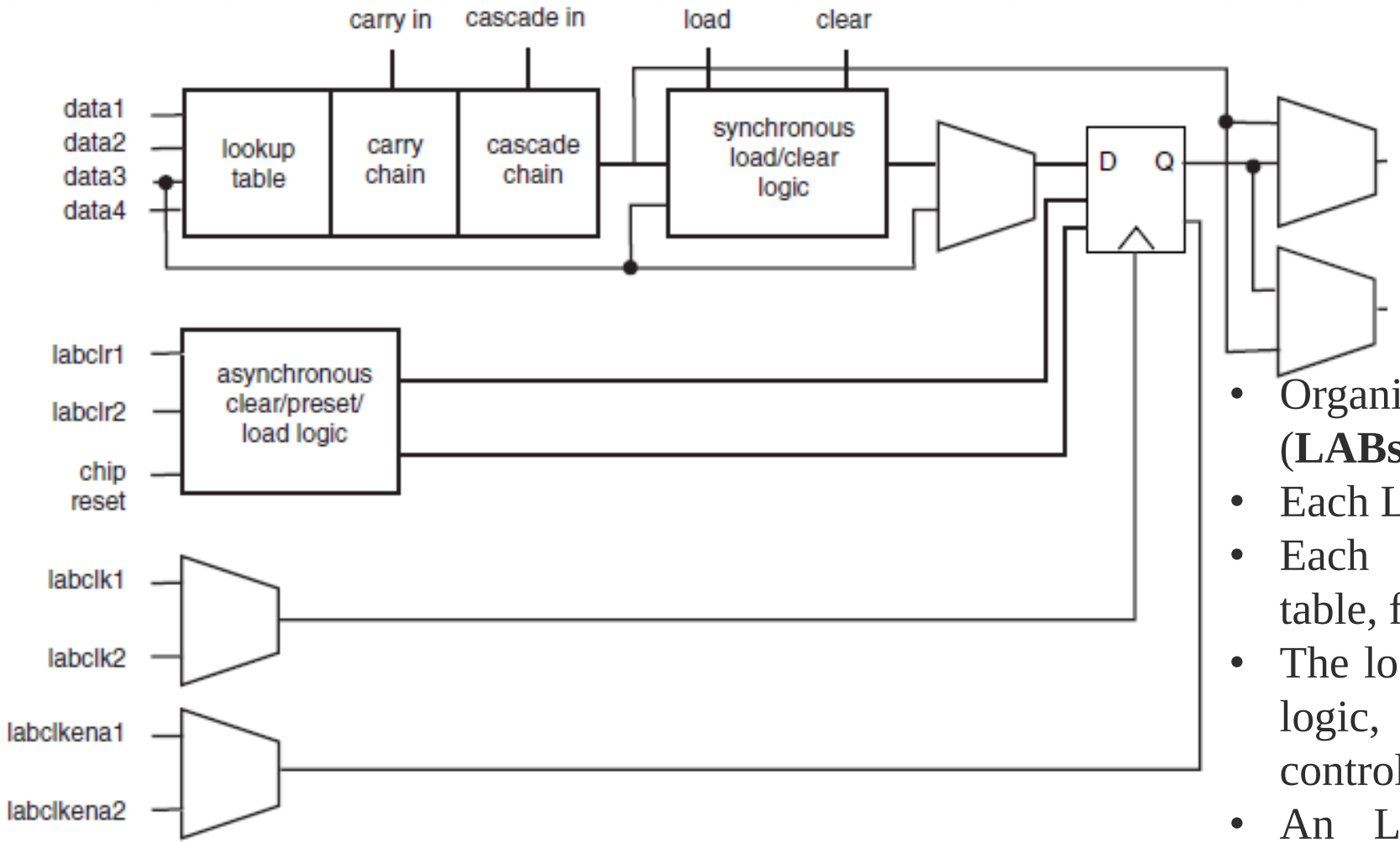


- Consists of two identical slices, with each slice containing a LUT, some carry logic, and registers
- A slice includes two logic cells (LCs).
- The foundation of a logic cell is the pair of four-bit lookup tables.
- Their inputs are F1-F4 and G1-G4.
- Each lookup table can also be used as a 16-bit synchronous RAM or as a 16-bit shift register.

Xilinx Spartan-II CLB

- Each slice also contains carry logic for each LUT so that additions can be performed.
- A carry in to the slice enters the CIN input, goes through the two bits of carry chain, and out through COUT.
- Each slice includes a multiplexer that is used to combine the results of the two function generators in a slice.
- Another multiplexer combines the outputs of the multiplexers in the two slices, generating a result for the entire CLB.
- The registers can be configured either as D-type flip-flops or as latches.
- Each register has clock and clock enable signals.
- Each CLB also contains two three-state drivers (known as BUFTs) that can be used to drive on-chip busses.

Altera APEX II logic elements



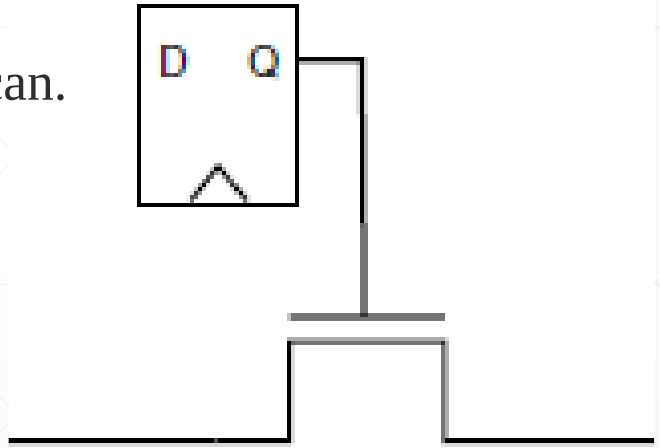
- Organized into **logic array blocks** (**LABs**).
- Each LAB includes 10 logic elements.
- Each logic element contains a lookup table, flip-flop, *etc*.
- The logic elements in an LAB share some logic, such as a carry chain and some control signal generation.
- An LE can be operated in normal, arithmetic, or counter mode.

Logic Elements

- An n-input function requires an SRAM with 2^n locations.
- A basic SRAM is not clocked --- as its inputs change, its output changes after some delay.
- A typical logic element has four inputs.
- The delay through the lookup table is independent of the bits stored in the SRAM, so the delay through the logic element is the same for all functions.
- For example, a lookup table based LE will exhibit the same delay for a 4-input XOR and a 4-input NAND.

Interconnection Networks

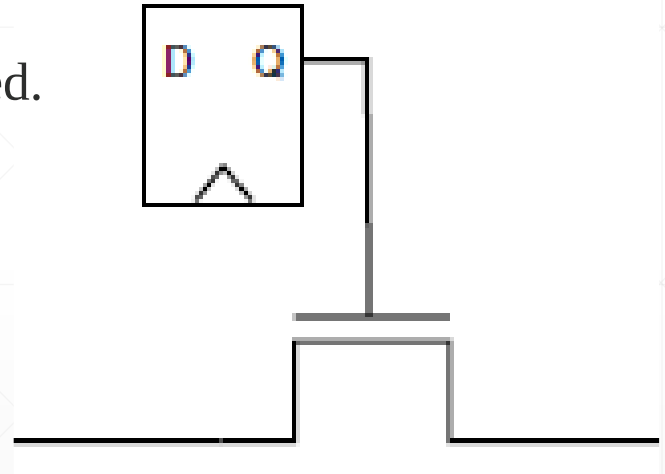
- An SRAM-based FPGA uses SRAM to hold the information used to program the interconnect.
- As a result, the interconnect can be reconfigured, just as the logic elements can.
- Figure shows a simple version of an interconnection point,
  - often known as a connection box.



- A programmable connection between two wires is made by a CMOS transistor (a pass transistor).
- The pass transistor's gate is controlled by a static memory program bit (shown here as a D register).

Interconnection Networks

- When the pass transistor’s gate is high, the transistor conducts and connects the two wires.
- When the gate is low, the transistor is off and the two wires are not connected.
- The transistor also conducts bidirectionally.
- Pass transistor is relatively slow.
- Alternative circuits (mostly unidirectional) provide higher performance at the cost of additional chip area.



Types of programmable interconnect

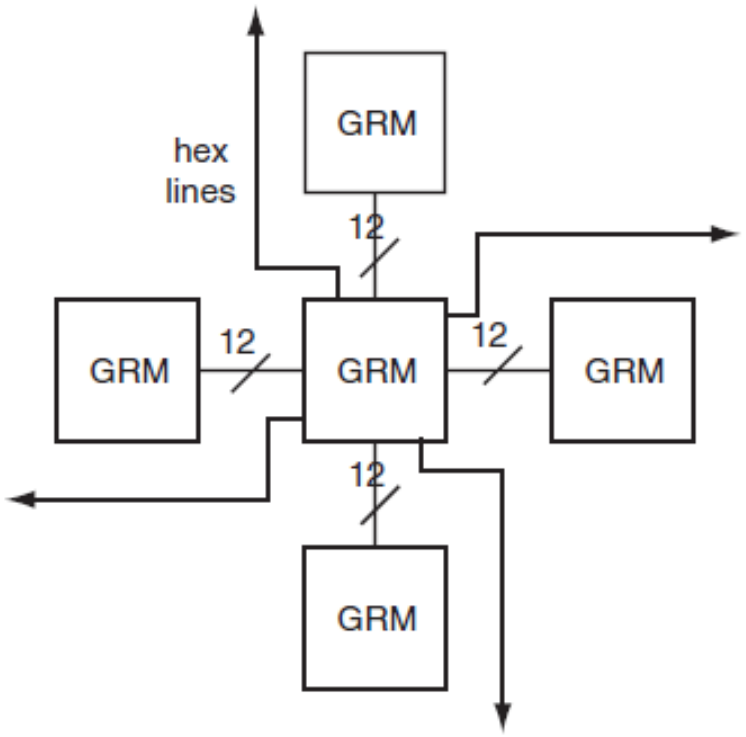
- An FPGA requires a large number of programmable wires in order to take advantage of the logic in the LEs.
- Wiring is often organized into different categories depending on its structure and intended use.
- **Short wires** connect only local LEs.
  - These wires don’t take up much space and they introduce less delay.
  - The carry chains through the LEs are one example of short interconnect.
- **Global wires**
  - designed for long-distance communications with high-speed highways
  - they have fewer connection points than local connections
  - repeaters to reduce the effects of delay
- **Special wires** may be dedicated to distribute clocks or other register control signals.

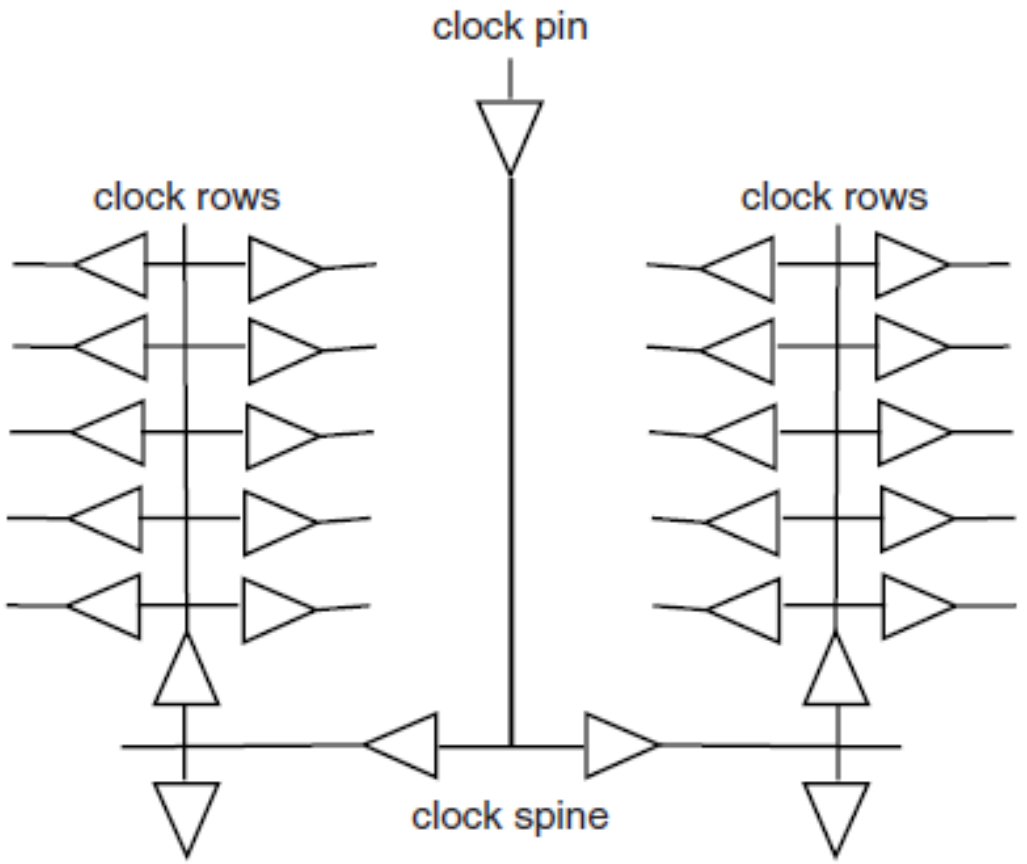
The Xilinx Spartan-II interconnect system

- The **Spartan-II** includes several types of interconnect:
  - local
  - general purpose
  - I/O
  - global
  - clock
- **Local interconnect** :
  - It connects the LUTs, flip-flops, and general purpose interconnect.
  - It also provides internal CLB feedback.
  - Finally, it includes some direct paths for high-speed connections between horizontally adjacent CLBs.
  - These paths can be used for arithmetic, shift registers, or other functions that need structured layout and short connections.

The Xilinx Spartan-II interconnect system

- The general-purpose routing network provides the bulk of the routing resources.
- This network includes several types of interconnect:
  - A **general routing matrix (GRM)** is a switch matrix used to connect horizontal and vertical routing channels as well as the connections between the CLBs and the routing channels.
  - There are **24 single-length lines to connect each GRM to the four nearest GRMs**, to the left, right, above, and below.
  - **Hex lines route GRM signals to the GRMs six blocks away**. Hex lines provide longer interconnect. The hex lines include buffers to drive the longer wires. There are **96 hex lines, one third bidirectional and the rest unidirectional**.
  - **12 longlines provide interconnect spanning the entire chip**, both vertically and horizontally.

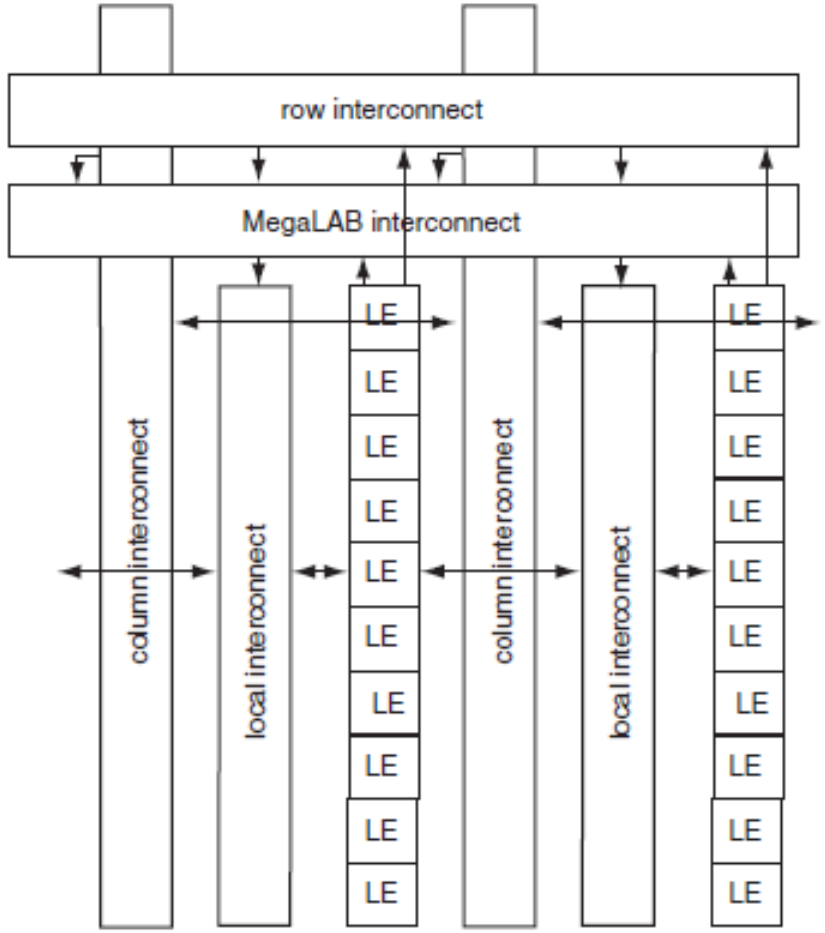




The clock distribution network is buffered to provide low delay and low skew

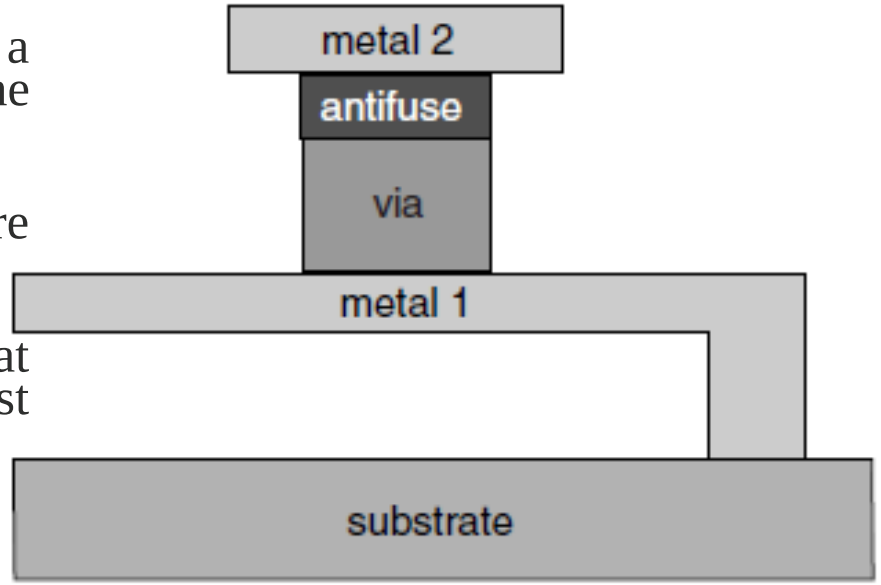
The Altera APEX II interconnect system

- The APEX II uses horizontal and vertical interconnect channels to interconnect the logic elements and the chip pins.
- A row line can be driven directly by an LE, I/O element, or embedded memory in that row.
- A column line can also drive a row line; columns can be used to connect wires in two rows.
- Some dedicated signals with buffers are provided for high-fanout signals such as clocks.
- Column I/O pins can directly drive these interconnect lines.
- Each line traverses two MegaLAB structures, driving the four MegaLABs in the top row and the four MegaLABs in the bottom row of the chip.



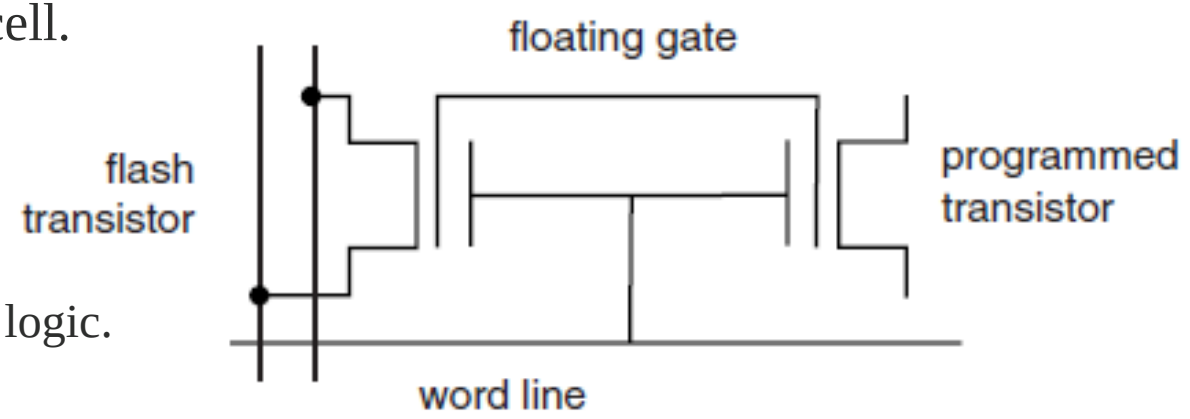
Antifuses

- An antifuse, as shown in Figure, is fabricated as a normally open disconnection.
- When a programming voltage is applied across the antifuse, it makes a connection between the metal line above it and the via to the metal line below.
- An antifuse has a resistance on the order of 100 Ω, which is more resistance than a standard via.
- The antifuse has several advantages over a fuse, a major one being that most connections in an FPGA should be open, so the antifuse leaves most programming points in the proper state.
- An antifuse is programmed by putting a voltage across it.
- Each antifuse must be programmed separately.
- The FPGA must include circuitry that allows each antifuse to be separately addressed and the programming voltage applied.



Flash Configuration

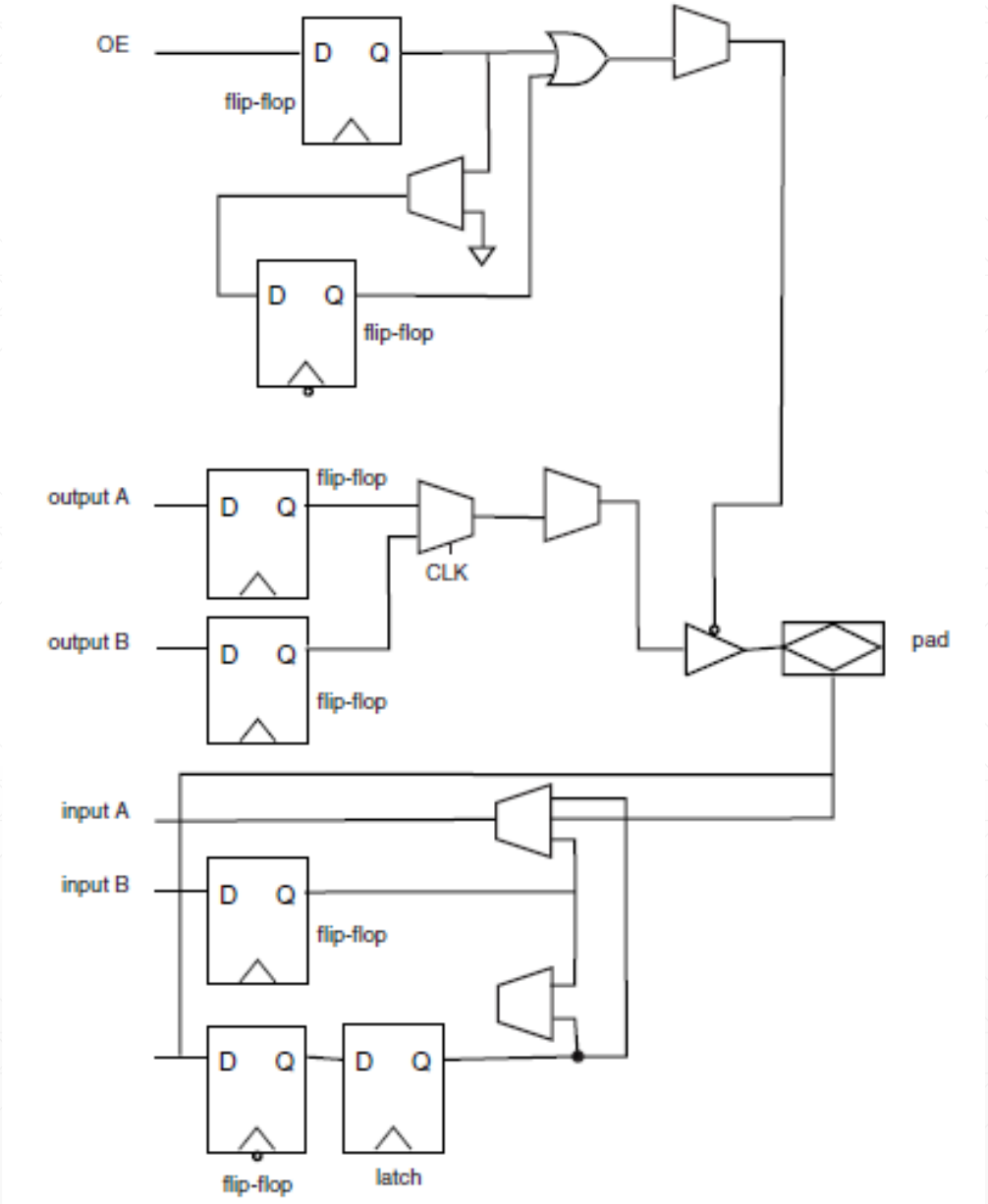
- Flash memory is a high-quality programmable read-only memory.
- Flash uses a floating gate structure in which a low-leakage capacitor holds a voltage that controls a transistor gate.
- This memory cell can be used to control programming transistors.
- Figure shows the schematic of a flash-programmed cell.
- The memory cell controls two transistors.
- One is the programmable connection point.
  - used for interconnect electrical nodes in interconnect or logic.
- The other allows read-write access to the cell.



Chip I/O

- The I/O pins on a chip connect it to the outside world. The I/O pins on any chip perform some basic functions:
- Input pins provide electrostatic discharge (ESD) protection.
- Output pins provide buffers with sufficient drive to produce adequate signals on the pins.
- Three-state pins include logic to switch between input and output modes.
- The pins on an FPGA must be programmable to accommodate the requirements of the configured logic.
- A standard FPGA pin can be configured as either an input, output, or three-state pin.

Altera APEX-II I/O pin



The Xilinx Spartan-II 2.5V I/O pin