

CHAPTER 5 PLD Technology

PLDs are now commonly differentiated based on two common architectures: the CPLD and the FPGA. The CPLD has evolved directly from the SPLD which is based on two architectures, the PAL and the PLA as discussed in Book 1, Chapter 3. The major PLD vendors have manufactured a variety of devices that are marketed as being superior to, or having particular advantages over, their competitors. To appear unique, it is common for comparable structures to be identified differently by various manufacturers. Consider the two major players: Altera refers to their basic logic unit within an LAB (logic array block) as an LE (logic element) while Xilinx uses the term “Slice” when referring to similar entities residing within a CLB (configurable logic block). Adapting to the different terminology can be challenging.

Both CPLDs and FPGAs utilize similar if not the same development software. The overview presented here is intended to provide real substance to what is difficult to visualize abstractly, to satisfy the readers curiosity, and to provide basic concepts and terms that will be useful towards further self-study.

You will observe that many of the attributes of the SPLD carry forward into the structures of CPLDs and FPGAs. For example, the use of multiplexing as a means of selecting registered or combinatorial output format or to facilitate the feedback of an output back into the input array. Finally, it is most important to remember, based on the fact that CPLDs and FPGAs are in a state of evolution, that there is increasing overlap in the architectures.

To appreciate the differences in architecture of various advanced PLDs, a look at fabrication techniques at a simple transistor level is essential. In order to accomplish this a few basic concepts of semiconductor physics will be reviewed. This material is not essential to the programming of PLDs, but without it many of the points of comparison between the various available FPGAs and CPLDs will be largely meaningless.

1. Semiconductor Physics

A. Introduction

Today, the most commonly used semiconductor material is silicon. Silicon has four valence electrons which are shared to form tetravalent bonds in a very regular three dimensional pattern referred to as a crystal lattice. Pure silicon is a brittle glass-like material that generally does not occur naturally but is readily available, being derived from silica sand. It is typically manufactured with an impurity added, in the form of a cylindrical ingot that may be metres in length. The cylinder, which may have a diameter of up to approximately 300 mm, is sliced to produce many thin round discs (wafers). A single wafer can be processed as a unit so that it will contain thousands of identical integrated circuits (chips). Upon completion and after testing, the wafer is cut to yield the individual ICs that are then packaged and sold.

A silicon crystal lattice can be simplistically represented in two dimensions as shown in Figure 5-1 (a). At lower temperatures, pure (intrinsic) silicon is not a good conductor. However, at room temperature, higher energy electrons break free from their bonds creating thermally generated hole-electron pairs which are represented by “ $\circ e^+$ ” and “ $\bullet e^-$ ” in Figure 5-1 (b).

Hole-electron pairs that are thermally generated within the barrier region will be swept across, but this very small leakage current will be offset by an equivalent diffusion current, thus maintaining equilibrium. When an external bias voltage is applied across bulk P and N semiconductor material that has formed a junction, one of two things can happen.

Forward Biased Junction

If the applied voltage is positive at the P side (anode) and negative at the N side (cathode) the diode is said to be forward biased. If the applied voltage is greater than roughly 0.65 volts, the applied electric field will essentially neutralize the depletion region potential, and *diffusion* will take place freely across the junction. This is illustrated in Figure 5-4 (a).

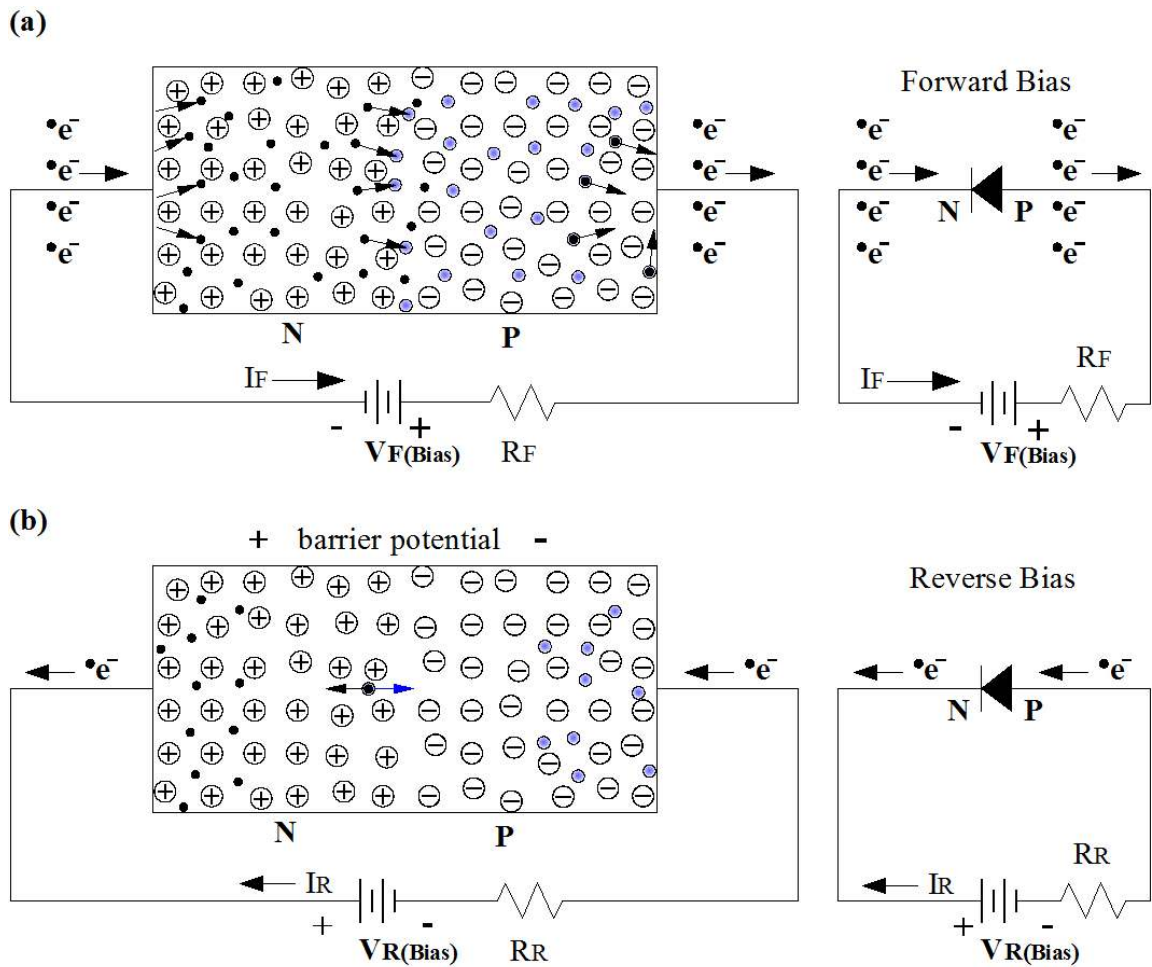


Figure 5-4 Biased PN Junction.

Biasing is the application of an applied voltage in either the forward or reverse direction. A resistor R is added to limit current. The diode symbol is an arrow pointing in the direction conventional current flows when forward biased. (a) Forward biasing counteracts the barrier potential effectively decreasing the depletion region to zero and carriers diffuse freely across the junction. If the relatively small semiconductor bulk resistance is ignored, the current can be approximated by $I_F = (V_F(\text{Bias}) - V_F) / R_F$, where V_F is the approximately constant forward voltage required to overcome the barrier potential. (b) A reverse bias increases the depletion region so that only thermally generated hole-electron pairs are available for conduction. This leakage current is typically in the order of nA and is often simply assumed to be zero. R_R can be ignored unless it is many M .

Substrate is the term used for a large area of silicon that is uniformly doped with the objective of providing an acceptable surface on which numerous devices can be integrated. In an N channel MOSFET (NMOS) the source and drain are N-type silicon located within either a P type substrate or a P type well that has been created within an N type substrate.

The identification of MOSFET type is associated with the substrate: an out arrow implies an N substrate or well, which means the channel will be P type (channel type is the opposite of the substrate or well). For an NMOS transistor, a conducting channel is formed when a *positive* voltage is applied to the gate with respect to the source or drain. This pulls majority carriers (electrons) into the P region immediately under the gate forming an inversion layer (the channel region is now effectively N type material).

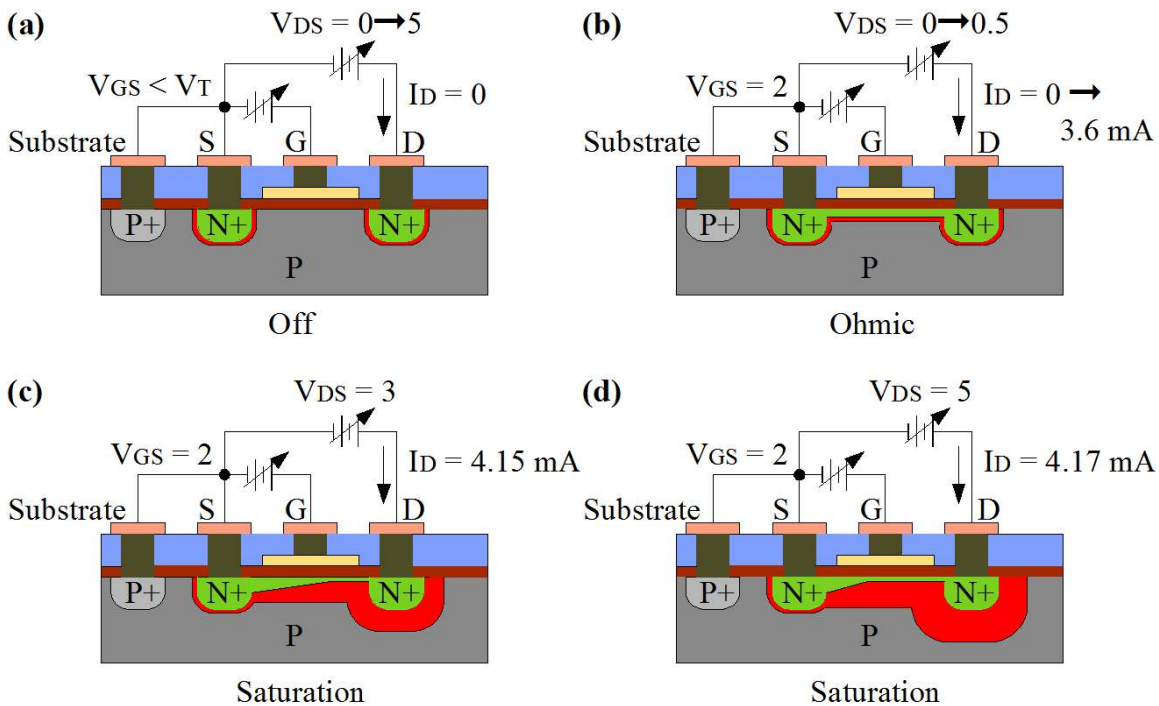


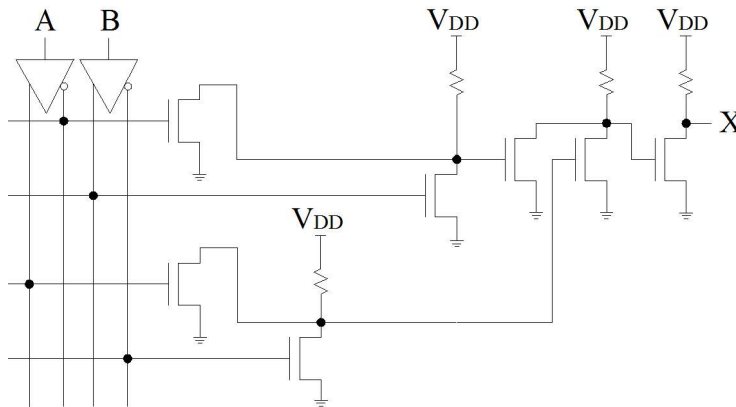
Figure 5-6 NMOS Transistor Biasing.

The biasing voltages represented above demonstrate NMOS behaviour over the full range from off to fully on, based on the characteristic curve (Figure 5-7). There is a region where operation is linear and amplification results but that is not the main focus relative to PLDs. (a) Without a modest positive V_{GS} voltage there is no conducting channel between the source and the drain so the drain current I_D is essentially zero (there is very little leakage current). (b) As V_{GS} exceeds the threshold voltage V_T , the gate induces a conducting channel in proportion to the difference. For small V_{DS} , I_D is proportional to V_{DS} within this ohmic region. (c), (d) As V_{DS} increases for a given V_{GS} , the drain voltage in relation to the substrate causes heavy reverse biasing of the PN junction formed by the two. This causes a depletion region to impinge increasingly upon the channel, restricting or "pinching it off". As a result substantial increases in V_{DS} cause virtually no increase in I_D and the NMOS is said to be in saturation. Saturation drain current values still depend directly on V_{GS} because V_{GS} controls the size of the channel. This is evident in Figure 5-7 for $V_{DS} = 5\text{ V}$ and $V_{GS} \approx 2.5\text{ V} \rightarrow 5\text{ V}$.

Referring to Figure 5-6, the NMOS substrate is typically tied to ground, or alternately to the source depending on the application. The manufacturer takes care of this. When the substrate is ground referenced an N type conducting channel will be induced by a positive voltage at the gate

Connecting a large number of NMOS in parallel in the wired-AND configuration results in a cumulative capacitance on the product line X with respect to ground. This slows switching times significantly, so a linear sense amplifier is often used to sense a small logic level change at X and produce a full logic output drive signal in response to it. This enhances performance, but at the expense of continuously wasted power. Given that each product line must have a sense amp, this can represent an unacceptable drain, for battery powered applications. Xilinx in their Coolrunner II CPLDs, advertizes that they have moved away from sense amps to “Real Digital” (CMOS based gating). The CMOS configuration as you will see shortly excels when it comes to low power because it replaces the upper resistor R with a PMOS that is off when the NMOS is on and vice versa. In this way it is able to produce the logical output voltages without any significant current flow. Of course it does require the complementary pair every time parallel or series MOS devices are required, which essentially doubles the number of devices.

Exercise 5-3: Write the output X equation for the PAL type circuit shown below.
 (Soln. Pg. 196)



The previous illustrations although simplistic should be helpful in understanding the concept behind SPLD and CPLD fabrication. It is evident that the NMOS NOR/NOR converts very nicely to AND/OR while allowing very high input counts based on the parallel configuration of NMOS inverters. With resistors replaced by current sources, power dissipation is minimized. CPLDs are based on SPLDs and likewise have very high fan-in (number of inputs). In contrast the FPGA typically uses a 4-input LUT and cascades these to produce greater fan-in.

D. EEMOS

You may recall that the acronym EE (E^2) stands for *electrically erasable*. It has been mentioned from time to time that MOS transistors are used as the fuse type links in re-programmable PLDs. In the case of the PAL these links determine whether a particular input, i.e. $I_0, \bar{I}_0, I_1, \bar{I}_1, I_2$, etc., will contribute to a given product in a circuit such as Figure 5-11. The PLA also uses fuse type links to select the product terms that will contribute to the sum.

EEMOS cells (transistor pairs) are used for this purpose. This technology is non-volatile, which means that once the EEMOS is programmed its state is maintained indefinitely unless it is re-programmed.

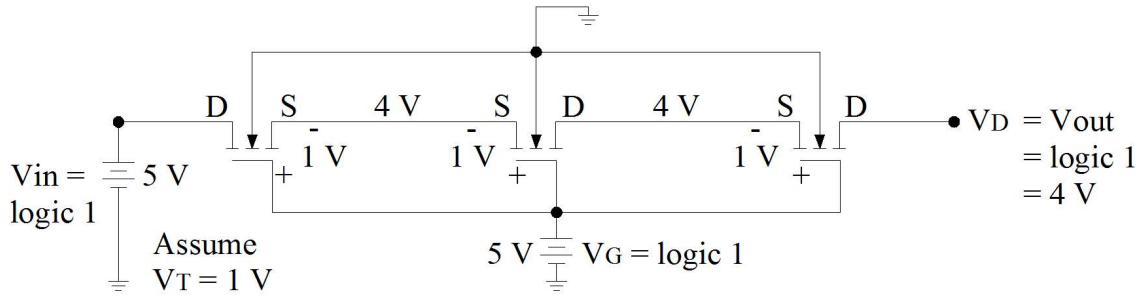


Figure 5-16 Multiple NMOS Pass Transistors.

When it is necessary to electronically interconnect many conducting wire segments in order to configure the logic in a PLD, it is only necessary to deal with a single degradation of a '1' level for NMOS or a '0' level for PMOS. As shown, assuming $V_T = 1\text{ V}$, consecutive NMOS will receive a 4 V input while having a 5 V, V_G , so all will meet the $V_{GS} = V_T$ requirement.

F. Transmission Gates

Since the PMOS passes '1's well and the NMOS passes '0's well, a parallel combination of the two, such that they are both on or off in unison, produces an ideal switch. Since NMOS and PMOS turn on with opposite gate voltages, an inverter must be added, as shown in Figure 5-17. To avoid extra clutter, the NMOS and PMOS are shown in their abbreviated form. The PMOS has its substrate tied to V_{DD} while the substrate of the NMOS is grounded. Assuming the required inverter is CMOS, this transmission gate structure is fabricated with a total of 4 transistors. Since the substrates are tied to V_{DD} or GND and not to the source, the source and drain are interchangeable. For this reason, they are not labelled in Figure 5-17. The transmission gate is an effective switch in digital applications as well as small signal analog where it is typically referred to as an analog switch. Of course, there is channel resistance associated with the switch, so its behaviour is most ideal when it is used in low current applications.

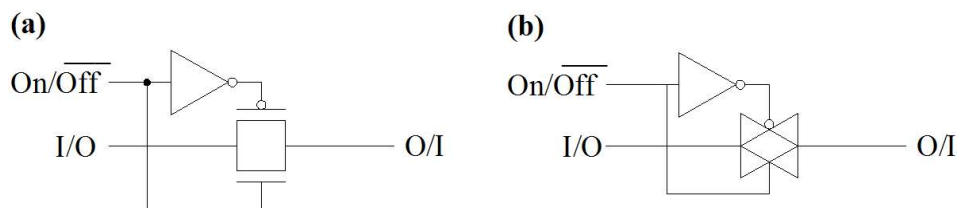


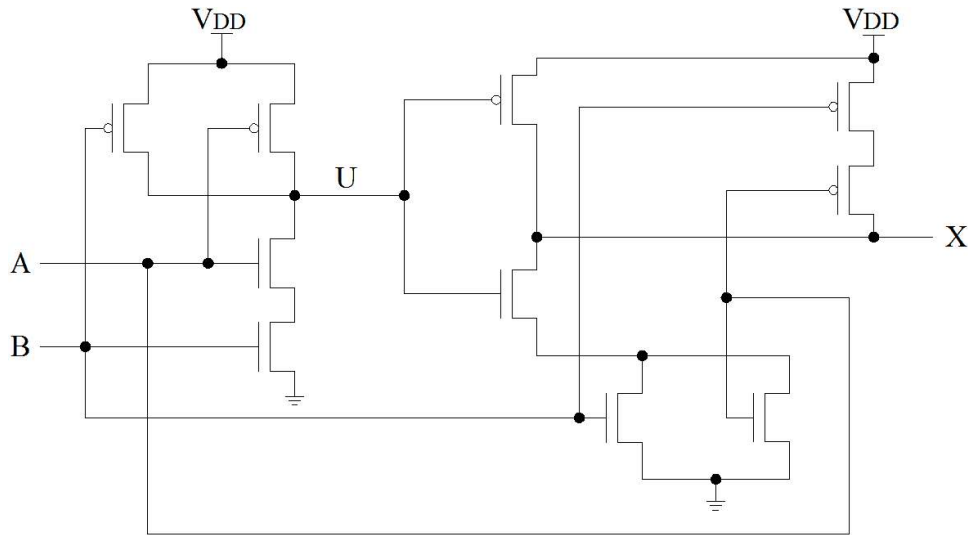
Figure 5-17 MOS Transmission Gate.

For digital or small signal analog, the transmission gate is capable of passing signals in either direction. Channel resistance is relatively low and operation is very symmetrical due to the pairing of PMOS and NMOS. (a) Typically, a CMOS inverter is used to provide the inverse V_G voltage required by the PMOS so that both the NMOS and PMOS are on together. (b) Alternate transmission gate symbol.

Exercise 5-5: Complete the truth table for the given circuit and determine its function. Write the minimal SOP expression for X. Why is it acceptable to hard-wire the outputs? What function is implemented by the hard-wired connection itself?

Figure 5-22 (b) uses the DeMorgan relationship that equates NAND/NOR (NAND/Negated input AND) to AND/AND. This configuration has potential for greater fan-in, i.e. if the two front end NANDs had five inputs each, then a ten input AND would result.

Exercise 5-6: Determine the function performed by the following gate. Hint: Examine the NMOS pull-down transistor configuration of the left half and right half sections of the circuit. The PMOS pull-up networks are simply complements of these. The left half is easy to recognize. For the right half remember series \Rightarrow AND, parallel \Rightarrow OR, and being a pull-down network, the result is active low (inverted). Use DeMorgan's theorem and you will recognize the final expression.



4. SPLDs

A very simple PAL type device is illustrated in Figure 5-23. Since it has three inputs, eight 3-input AND gates would be required to fully decode the input variables such as in a ROM. However, PAL devices do not include full input decoding because the number of product terms in a minimized truth table is never that large.

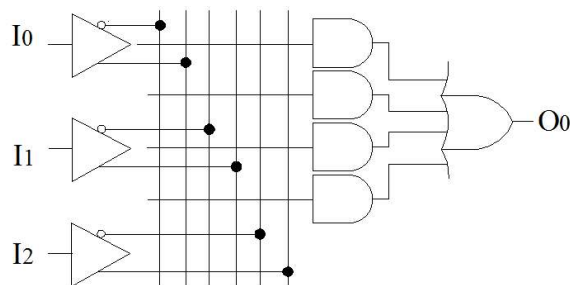


Figure 5-23 PAL Format.

A PAL consists of a programmable AND matrix followed by fixed OR. It typically has fewer than the theoretical maximum required product terms which in this example would be $2^3 - 1$.

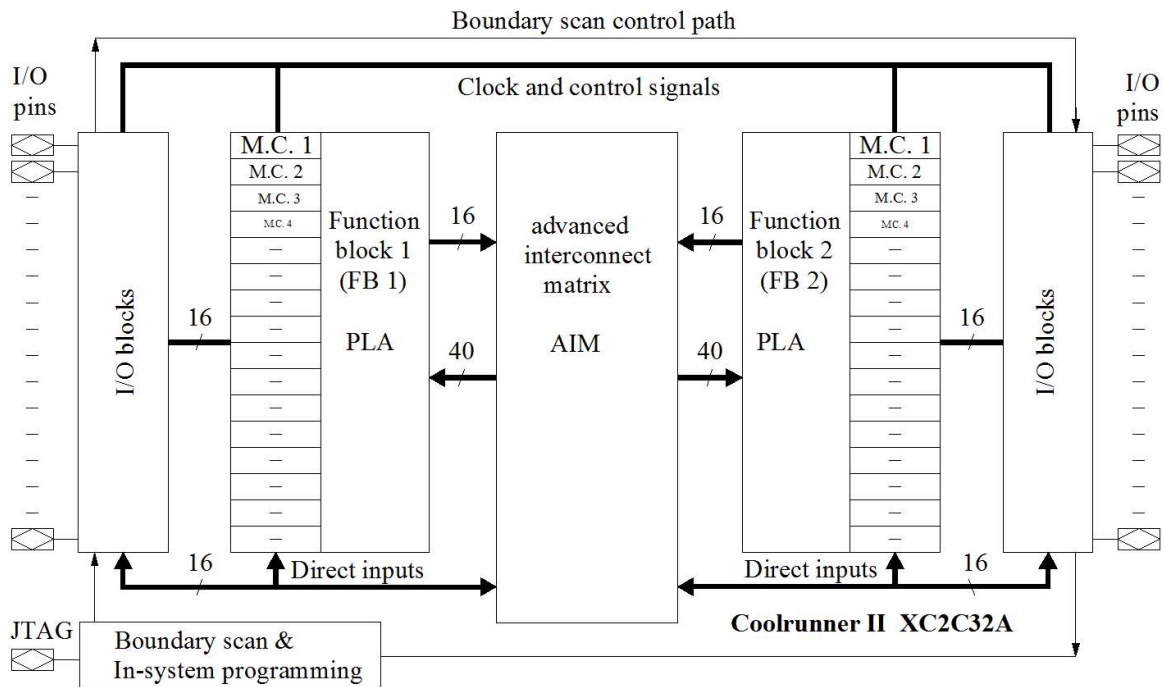


Figure 5-27 Xilinx Coolrunner II block Diagram.

The Coolrunner II XC2C32A is the smallest in the family, having only two function blocks. A function block is essentially 49 product term PLA having 16 OR gates with 56 programmable inputs each. Each OR gate is associated with a macrocell that is similar to that of the previous SPLD, allowing the typical output inversion, programmability, dual I/O functionality and so forth, as illustrated in Figure 5-28.

Figure 5-28 is a simplified representation of the basic function block, in this case FB₁. Referring to the hexagonal encircled numbers on the diagram that match the numbers below, some points of interest are:

1. There are 49 general purpose AND gates that span the 40 external input signals that are carried by the advanced interconnect matrix (AIM). Thus a product term can have up to 40 variables!
2. CTC, CTR, CTS, CTE are 4 AND gates dedicated to Control Terms (CT) used in the function block for: clocking (C), flip-flop asynchronous set (S) and reset (R), and output enable (E).
3. PTA is a single AND product term (PT) which provides alternate set and reset for the flip-flop.
4. PTB is a single AND product term that can be used for enabling tri-state outputs. Not shown, is the I/O block MUX which allows 10 different choices for enabling tri-state outputs.
5. PTC is a single AND product term that can be used to dynamically modify O/P polarity via XOR (8), or may be used to enable the output latch or provide a clock signal to the flip-flop.

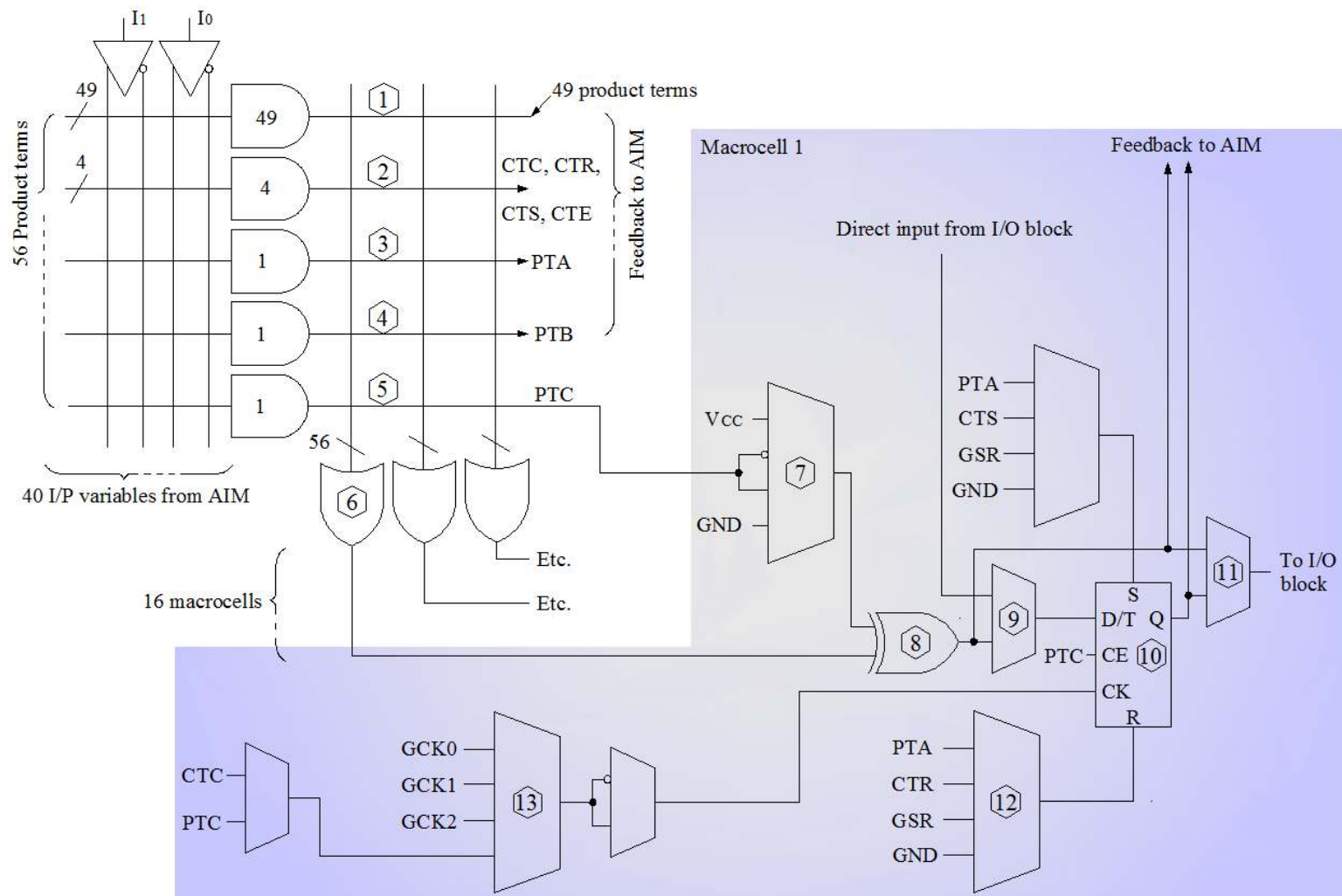


Figure 5-28 Xilinx Coolrunner II Functional Block 1.

The Coolrunner II XC2C32A has two functional blocks each having 16 macrocells. The PLA structure allows 49 AND inputs, $I_{48}..I_0$ to be selected for product terms and each of these product terms is also selectable in the OR matrix. Product terms are also fed back to the Advanced Interconnect Matrix (AIM) for reuse by any of the functional blocks. The macrocells allow a number of choices relative to registering outputs, selecting different clocks and using dual edge triggering, inverting signals, selecting D or T flip-flop format, setting and resetting the flip-flop, and so forth.

Logic blocks connect to the wire segments of the routing channels with buffered pass transistors or multiplexers in what is called a connection box (C) as shown in Figure 5-30. At every junction where wire segments would cross, there is a switch box (S) that extends the segment in the desired direction. A switch box consists of 6 pass transistors, which allows full interconnectivity. Each individual wire segment requires a switch box at every junction point. For clarity, the detail of Figure 5-30 illustrates only one switch box. The topology of the switch box that is illustrated is planar, meaning that segment-one connects only to other segment-one wires. Wire segments may be longer than the base length between adjacent switch boxes. Longer segments that bypass switch boxes will incorporate fewer pass transistor switches and therefore operate faster. Generally a mix of segment lengths improves performance without compromising routing efficiency significantly.

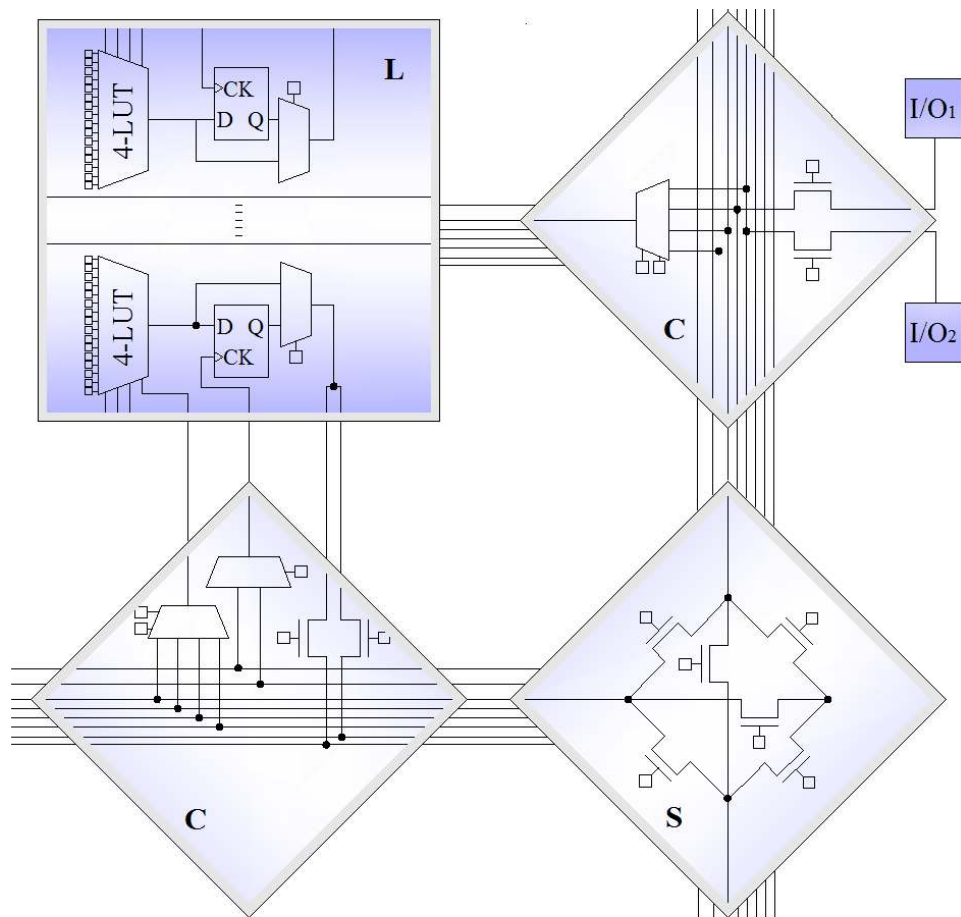
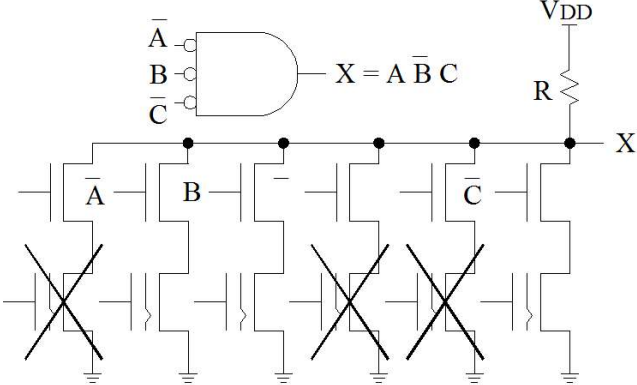


Figure 5-30 FPGA Architecture.

The logic block must interconnect with other blocks as well as with input and output. Multiplexers and pass transistors serve as switches selecting the desired components or wire segments. Volatile SRAM cells provide the bit values for the programming (small boxes in diagram). Connection boxes carry the signals to and from the logic block while switch boxes provide the vertical and horizontal wire segment interconnections that achieve the routing.

The logic block of a typical FPGA contains a number of LUTs, some simple gating, multiplexers for selective routing, and registers (flip-flops). The configuration may be optimized for certain common operations such as multi-bit addition. The SOP expressions that can be implemented are typically no more complex than what a 4 or 5 variable truth table would imply.

Pg. 172 Exercise 5-4: Place an “X” through the EEPROM transistors in Figure 5-13 (a) that must be programmed “off” in order to implement the result shown in Figure 5-13 (b).

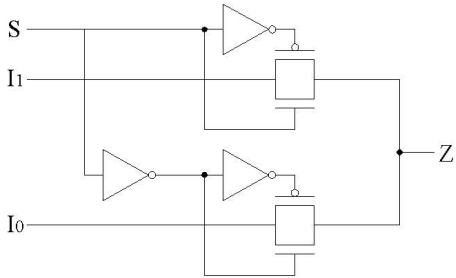


Pg. 176 Exercise 5-5: Complete the truth table for the given circuit and determine its function. Write the minimal SOP expression for X. Why is it acceptable to hard-wire the outputs? What function is implemented by the hard-wired connection itself?

S	I ₁	I ₀	Z
0	0	0	0
0	0	1	1
0	1	1	1
0	1	0	0
1	1	0	1
1	1	1	1
1	0	1	0
1	0	0	0

Only one of the two parallel switches can be on at a time in this wired-OR configuration.

$$Z = \bar{S} I_0 + S I_1$$



Pg. 182 Exercise 5-6: Determine the function performed by the following gate.

