J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

PROGRAMMABLE LOGIC ARRAYS

Performance demand, design complexity, cost-effectiveness, and time-to-market have become the most crucial concerns for contemporary logic designers. With its user-programmability, field-programmable logic has emerged as an unparalleled solution to the cost and time-to-market challenges by providing short turnaround time of designs with low risk/cost, allowing easy design changes.

A *field-programmable logic device* (*FPLD*) refers to any logic device whose function is specified by the user after the device is manufactured; general-purpose commercially available FPLDs include Simple Programmable Logic Devices (*SPLD*), Complex Programmable Logic Devices (*CPLD*), and Field-Programmable Gate Arrays (*FPGA*).

An SPLD is capable of implementing hundreds of gates and is usually programmed by users at their sites (fields) using inexpensive programming hardware. For example, a Programmable Logic Array (*PLA*), introduced by Monolithic Memories (later acquired by Advanced Micro Devices, Inc.) and the first FPLD developed specifically for implementing logic functions, consists of two-level AND–OR planes that can be programmed by users to realize any sum-of-products logic function, subject to the size constraints of the device. See Fig. 1 for a realization of a 4-bit comparator using a small PLA with eight inputs, eight AND gates, one OR gate, and one output. Connections are made by programming the switches, indicated by solid circles, to be ON. The field-programmability provides SPLDs the attractive advantages of low start-up cost, low risk, short manufacturing time, and easy design changes.

The primary limitation of SPLDs lies in their low logic capacities (the gate count in a single chip), due to the restricted nature of the AND–OR plane; the size of the plane grows too quickly as the number of inputs increases. One feasible approach to substantially improving chip capacities based on SPLD architectures is to incorporate multiple two-level SPLDs on a single chip with programmable switching network interconnecting them; the resulting devices are known as CPLDs. Typically, a currently existing CPLD can provide a logic capacity equivalent up to tens of SPLDs; however, it is hard to obtain a higher-capacity logic device based on the SPLD architectures.

The need to build a programmable device with a very high capacity leads to the development of FPGAs which were first introduced by Xilinx in 1984 (1). An FPGA is a (re)programmable logic device that implements multilevel logic. It combines the scalable interconnection structure of a Mask-Programmable Gate Array (*MPGA*) and the programmability of an SPLD/CPLD; this combination results in a programmable device with high-capacity logic. For example, the usable gates in a current commercially available FPGA can be more than 100 K, and the gate count is expected to grow rapidly as new architectures are introduced. Like SPLDs/CPLDs, FPGAs are programmable by end users at their sites (fields), eliminating the time-consuming fabrication step, and thus result in lower prototyping cost and shorter manufacturing time. These advantages have made FPGAs very popular in a wide variety of applications such as system prototyping, small- and medium-volume production, logic emulation, device controllers, data communications and telecommunications systems, digital signal processing, reprogrammable computing, etc. As an important type of field-programmable devices that provides very high-capacity logic, FPGAs have been the fastest growing segment of the total logic market during the past decade.

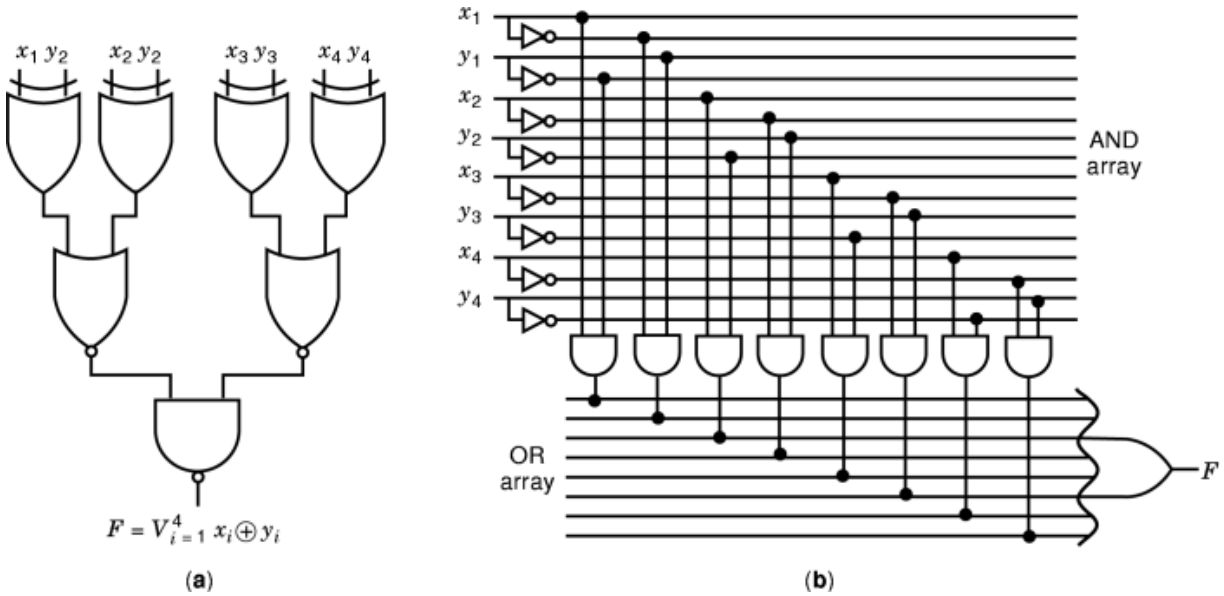

Fig. 1. Realization of an example circuit (4-bit comparator) using a PLA. (a) The example circuit. (b) An implementation of (a) using a *PLA* with eight inputs, eight AND gates, one OR gate, and one output.

A typical FPGA is composed of three major components: logic modules, routing resources, and input/output (I/O) modules (2). Fig. 2 depicts the conceptual FPGA model. In an FPGA, an array of logic modules is surrounded or overlapped by general routing resources bounded by I/O modules. The logic modules contain combinational and sequential circuits that implement logic functions. The routing resources comprise prefabricated wire segments and programmable switches. The interconnections between the logic modules and the I/O modules are user-programmable. A logic circuit is implemented in an FPGA by partitioning logic into individual logic modules and then interconnecting the modules by programming switches. A large circuit that cannot be accommodated into a single FPGA is divided into several parts; each part is realized by an FPGA and these FPGAs are then interconnected by a Field-Programmable Interconnect Component (*FPIC*) (see Fig. 3).

As a relatively new technology, high-capacity programmable logic devices such as CPLDs and FPGAs are still constantly evolving in their architectures, and virtually each CPLD/FPGA vendor (and even each generation of products from the same vendor) has its own unique architectural features. Nevertheless, each product can be distinguished from others by three key segments: programming technology, logic-module architecture, and routing architecture. In the rest of this article, we shall focus on the discussion of the popular high-capacity programmable logic, CPLDs and FPGAs, based on the three segments.

Programming Technologies

CPLDs/FPGAs use programmable switches to make connections, and a typical CPLD/FPGA may contain millions of switches. Therefore, the properties of the switches, such as manufacturing flow, physical size, ON resistance, OFF capacitance, power consumption, and manufacturing reliability greatly affect the CPLD/FPGA architectural decision. Specifically, the following properties for a switch are desirable (2,36):

- a simple manufacturing process

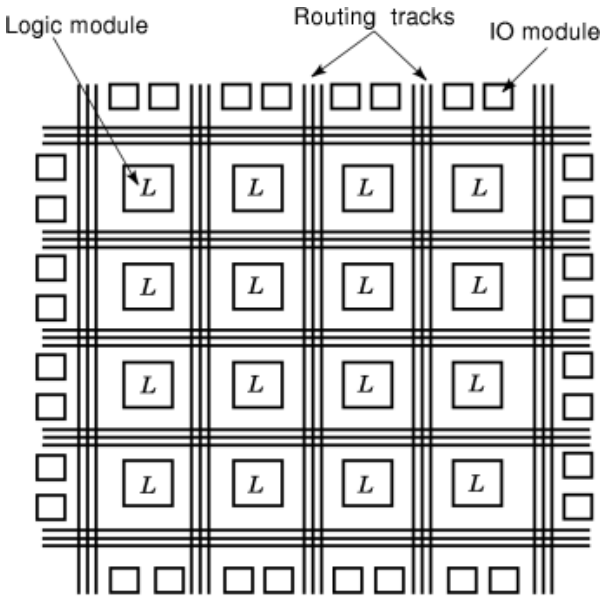

Fig. 2. A typical FPGA architecture with three major components: logic modules, routing resources, and I/O modules.

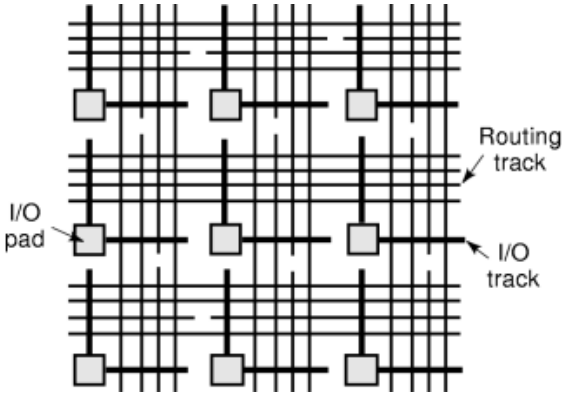

Fig. 3. An FPIC architecture.

- a small chip area
- a low ON resistance and a very high OFF resistance
- a low parasitic capacitance
- low power consumption
- high manufacturing reliability

In addition to the above, reprogrammability is also an attractive feature because it makes possible to reconfigure a CPLD/FPGA, reducing risk and allowing easy design changes.

The programming technology of PLAs introduced by Monolithic Memories was based on bipolar fuses. However, most commercially available CPLDs/FPGAs are based on Complementary Metal Oxide Silicon (*CMOS*) technology which offers many attractive advantages such as low power dissipation, low risk, worldwide

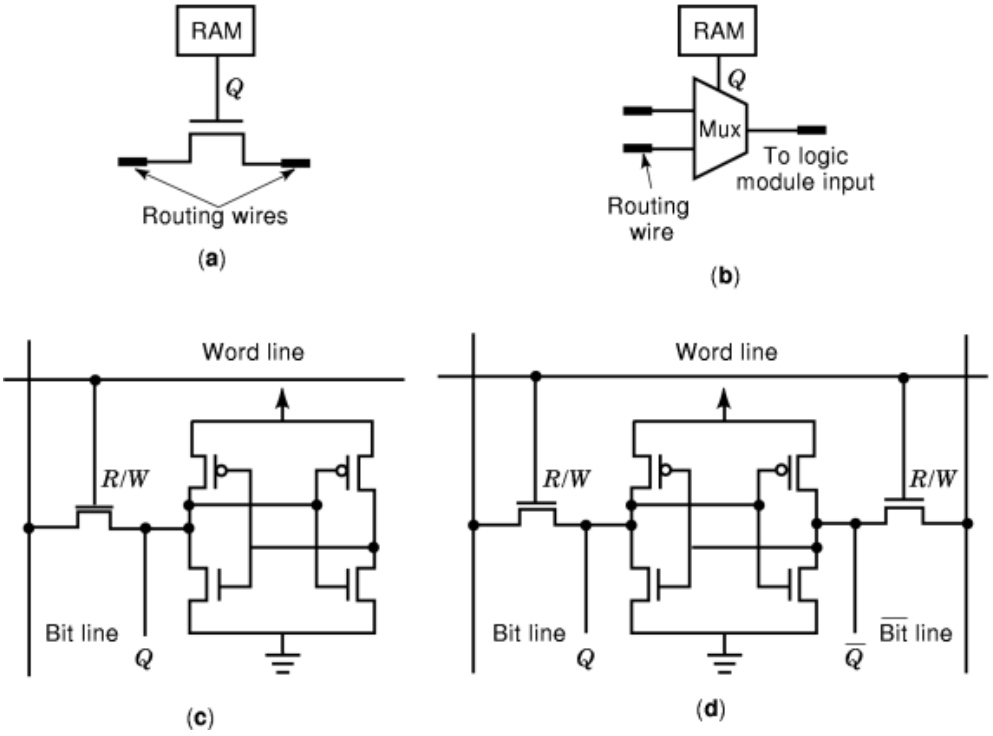

Fig. 4. The SRAM programming technology. (a) A pass-transistor switch. (b) A multiplexer switch. (c) A five-transistor SRAM cell. (d) A six-transistor SRAM cell.

research and development investment, etc. While programming technologies such as Erasable Programmable Read-Only Memory (*EPROM*) and Electrically Erasable PROM (*EEPROM*) are popular in SPLDs and CPLDs, Static Random Access Memory (*SRAM*) and antifuse are two major switch technologies for commercial FPGAs/FPICs and some CPLDs. Each technology has its strengths as well as weaknesses. The following subsections detail important properties of the programming technologies. Although examples are taken from commercial architectures, it is not our intent here to survey all existing structures.

SRAM Programming Technology. The SRAM programming technology employs SRAM cells to control pass transistors or multiplexers as shown in Fig. 4(a) and 4(b). For the pass-transistor element in Fig. 4(a), the state of the SRAM cell controls the ON and OFF of the transistor (switch): When ON, the pass transistor exhibits a relatively low resistance (<2 kΩ) between the two adjacent routing wires, and thus the switch is closed; when OFF, the switch is open and the transistor incurs a very high resistance between the two routing wires. For the multiplexer approach in Fig. 4(b), the states of the SRAM cells serve as select signals and control the choice of the multiplexer's inputs.

The SRAM cells shown in Fig. 4(c) and 4(d) are implemented using five and six transistors, respectively; four of them form two CMOS inverters interconnected to establish a bistable circuit element. In the case of five transistors [Fig. 4(c)], the remaining one transistor serves as Read/Write control component to load the cell and read back the programming; for the six-transistor SRAM, an additional transistor, on the other side of the cell, serves as another control component [see Fig. 4(d)]. The Q output of the SRAM cell is connected to and control a separate switch [could be a pass transistor or a multiplexer, as shown in Fig. 4(a) or 4(b)]. The five-transistor SRAM is more compact, whereas the six-transistor one has higher noise immunity and is

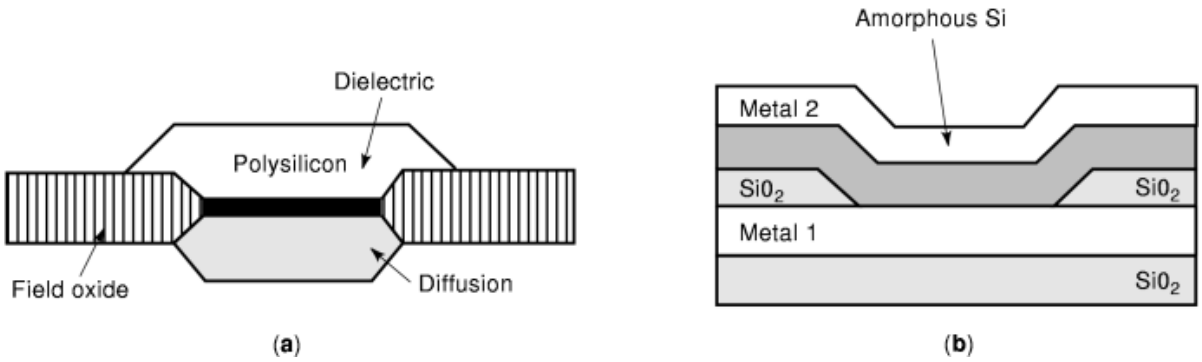

Fig. 5. The antifuse programming technology. (a) The ONO antifuse structure. (b) The amorphous antifuse structure.

easier to design. Due to its stability, the six-transistor SRAM gets advantages for deep-submicron process and low-power supply technologies (3).

SRAM programming technology has the following major advantages: simple manufacturing flow, fast in-circuit reprogrammability, and low power consumption. However, a major disadvantage of SRAM programming technology is its large physical size. For example, in Xilinx FPGAs (4), it takes at least one pass transistor to serve as a programmable switch, and five (e.g., XC4000, XC5200) or six (e.g., XC6200) transistors to implement an SRAM cell, as illustrated in Fig. 4(c) or 4(d).

One interesting property of SRAM is its volatility—a configuration is lost when its power is off. On one hand, this implies a system overhead that external permanent memory such as a Read-Only Memory (*ROM*), Programmable Read-Only Memory (*PROM*), or a disk is needed to store and provide programming configurations. On the other hand, volatility gives SRAM-based FPGAs better design security because a design configuration is lost when power is removed.

Antifuse Programming Technology. An antifuse is a two-terminal, one-time programmable circuit element with high resistance ($>$100 MΩ) between its terminals in the unprogrammed state and low resistance ($\leq$500 Ω) in the programmed state. Fig. 5(a) and 5(b) illustrate the structures of two antifuses, called a *PLICE* (Programmable Low-Impedance Circuit Element) and a ViaLink, used in Actel FPGAs (5,6) and QuickLogic FPGAs (7), respectively. The PLICE antifuse consists of an Oxygen–Nitrogen–Oxygen (ONO) dielectric layer (served as an insulator) sandwiched between a polysilicon and an $n+$ diffusion layers (served as conductors). Programming is performed by applying a higher-than-operating voltage (say, 18 V) across the antifuse's terminals, causing the dielectric breakdown and drastically reducing the device resistance. The ViaLink antifuse is composed of a sandwich of very high resistance layer ($>$1 GΩ) of programmable amorphous silicon between two mental layers. When a programming voltage is applied, a mental-to-mental link is formed by permanently converting the silicon to a low resistance state (30 Ω to 80 Ω).

Two major advantages of the antifuse are its small size and relatively low ON resistance and OFF capacitance; for example, the size of an amorphous antifuse is approximately 1 μm^2 in a 0.65 μm process (7). These advantages allow a much denser switch population and thus could alleviate the routing constraints imposed by the limited connectivity of routing resources. However, antifuse programming technology has the following major disadvantages: relatively complex manufacturing process and nonreprogrammability.

EPROM and EEPROM Programming Technology. An ultraviolet (UV)-erasable PROM (EPROM) is typically based on a floating gate structure as illustrated in Fig. 6. If sufficient charge is trapped on the floating gate by applying higher-than-operating voltages between the control gate (13 V to 14 V) and the drain of the transistor (12 V), the floating gate becomes charge negatively. The process increases the threshold voltage of the transistor to around 7 V, setting the transistor to the OFF state for all normal circuit voltages, maximally

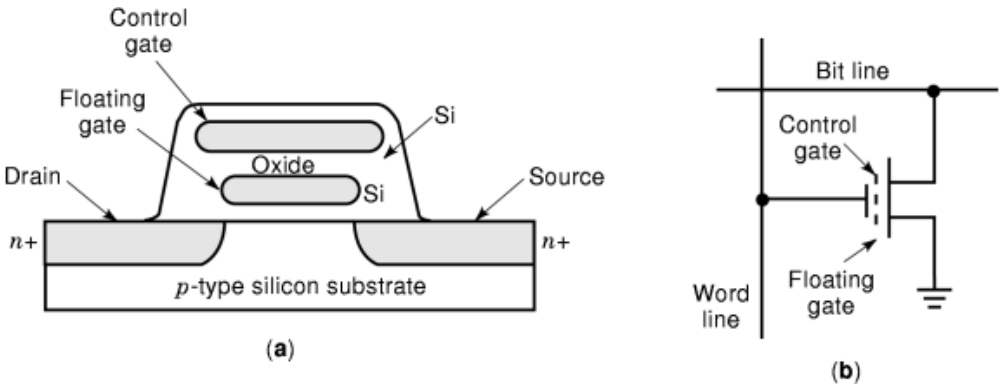

Fig. 6. The EPROM programming technology. (a) UV-erasable EPROM structure. (b) Circuit symbol for a floating-gate EPROM cell.

5 V to 6 V. The process can be reversed out-of-circuit by exposing the floating gate to UV light, giving the trapped charge sufficient energy to escape from the gate dielectric; this process reduces the threshold voltage and makes the transistor function normally.

The Electrically Erasable PROM (EEPROM) programming technology typically uses two transistors, one access and one programmed transistors, in a ROM cell. The programmed transistor performs the same function as the floating gate in an EPROM, with both charge and discharge being done electrically in-circuit, without UV light. The access transistor allows programming of a cell.

The major advantages of EPROM and EEPROM technologies are their reprogrammability and full testability before shipment (especially for EEPROM). Also, unlike the SRAM technology, EPROM or EEPROM requires no external permanent memory to program the chip at power-up. However, the EPROM/EEPROM technology suffers from some drawbacks such as relatively high ON-resistance, high static power consumption, and complicated manufacturing processes (8).

Recently, a new nonvolatile, electrically erasable programming technology called *flash memory* was introduced. A flash memory element occupies a relatively small physical size (about one-third of the size of an EEPROM cell), offers higher reliability, and has faster device programming times than the EEPROM one. Due to the advantages, flash memory is becoming attractive on the market.

Summary of Programming Technologies. Table 1 compares the properties of SRAM, ONO antifuse, amorphous antifuse, EPROM, and EEPROM programming technologies based on a 0.65 μm process technology; it gives the name of the technology, the relative complexity of manufacturing process [measured in the number of extra processing steps required beyond standard CMOS (8)], the availability of reprogramming, the approximate relative size of the switch, the series resistance of an ON switch, the parasitic capacitance of an OFF switch, the relative degree of power consumption, and the volatility of the configuration.

Logic-Module Architecture

A logic circuit is implemented in a CPLD/FPGA by partitioning logic into individual logic modules and then interconnecting the modules by programming switches. A logic module has a fixed number of inputs and outputs. Certain set of functions can be implemented using a logic module, depending on the functionality of the module. The logic modules of a high-capacity commercial programmable device are typically based on lookup tables, multiplexers, transistor pairs, basic logic gates, or SPLDs, with the first four types being popular in FPGAs and the last in CPLDs. We detail lookup table- and multiplexer-based logic modules in the following.

Table 1. Comparison of Programming Technologies (+: Desirable; −: Not Desirable)

Programming Technology	SRAM	ONO Antifuse	Amorphous Antifuse	EPROM	EEPROM
Manufacturing complexity	+++	+	+	−	−−
Reprogrammable?	Yes in circuit	No	No	Yes out of circuit	Yes in circuit
Physical size (μm^2)	15–20	2	1	40–50	80–100
ON resistance (Ω)	600–800	100–500	30–80	1–4 K	1–4 K
OFF capacitance (fF)	10–50	3–5	1	10–50	10–50
Power consumption	++	+	+	−	−
Volatile?	Yes	No	No	No	No

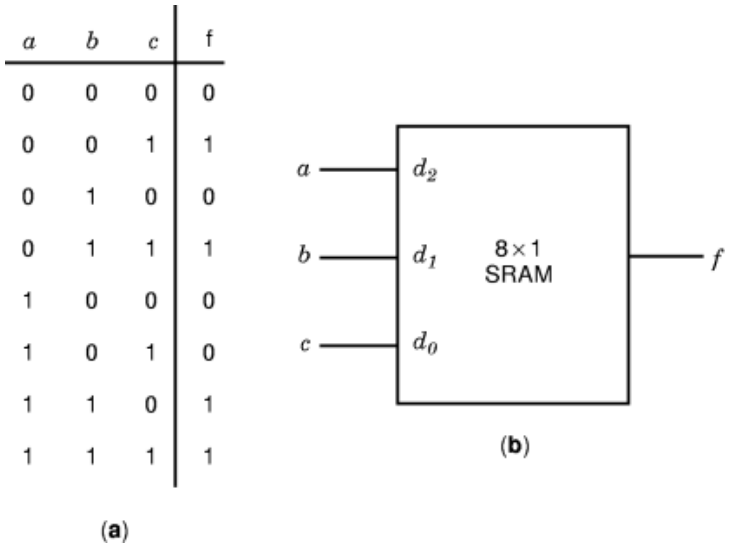

Fig. 7. Lookup table based logic. (a) Truth table for $f = ab + \bar{a}c$. (b) An 8×1 SRAM.

Due to their simplicity, we only brief those based on transistor pairs, basic gates, and SPLD at the end of this section.

Lookup Table-Based Logic Modules. A lookup table (*LUT*) based logic module is a segment of SRAM. The programming data defining the logic configuration are loaded into the SRAM at power-up. A K-input LUT is a $2^K \times 1$ SRAM with K address lines served as inputs and the 1-bit SRAM output as the LUT output. For example, if the function $f = ab + \bar{a}c$ needs to be implemented using a 3-input LUT, then the truth table shown in Fig. 7(a) is loaded into the LUT, and thus the $2^3 \times 1$ SRAM would have a 0 stored at address 000, a 1 at 001, etc., as given in the truth table.

The main advantage of LUTs lies in their high functionality—a K-input LUT can realize any function of up to K inputs and there are 2^{2^k} such functions. However, the LUTs are only feasible for small values of K because 2^K memory cells are required for a K-input LUT with only combinational logic; typically, the value of K is 3, 4, 5 or 6. Further, the value of K may be even smaller in order to implement multi-output or sequential logic.

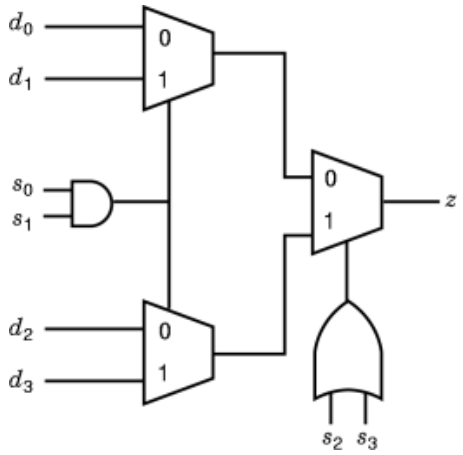

Fig. 8. A multiplexer-based logic module with data inputs d_0, d_1, d_2, and d_3, select inputs S_0, S_1, S_2, and S_3, and output z.

Multiplexer-Based Logic Modules. A multiplexer-based logic module is typically composed of a tree of 2-to-1 multiplexers. For example, the Actel ACT 3 C-Module (5) shown in Fig. 8 consists of three 2-to-1 multiplexers, implementing the following function:

$$z = \overline{(s_3 + s_2)}\,\overline{s_1 s_0} d_0 + \overline{(s_3 + s_2)} s_1 s_0 d_1 + (s_3 + s_2)\overline{s_1 s_0} d_2 + (s_3 + s_2) s_1 s_0 d_3$$

The inputs to the logic module are either data inputs d_0, d_1, d_2, d_3 or the multiplexer select inputs s_0, s_1, s_2, s_3. Therefore, the logic module can be used to implement a wide range of different functions of up to eight variables (specifically, 766 functions for the logic module); for example, the function $f = ab + \bar{a}c$ can be realized by setting $d_0 = \times$, $d_1 = \times$, $d_2 = c$, $d_3 = b$, $s_0 = a$, $s_1 = 1$, $s_2 = 1$, and $s_3 = \times$, where $\times$ means "don't care." Also, flip-flops can be incorporated into the multiplexer-based logic module to implement sequential logic.

Logic-Module Tradeoffs. Besides lookup tables and multiplexers, transistor pairs and basic gates such as 2-input NAND, EXOR, AND gates are used to form logic modules in several commercial FPGAs (9). Unlike lookup tables and multiplexers which can implement a versatile set of logic functions (but are substantially larger in size), the transistor pairs and basic gates can implement only a small set of functions (but are much smaller).

To capture the differences, logic modules are often classified by their granularity: coarse grain or fine grain (8). A coarse-grain logic module refers to one which can implement a large number of Boolean functions, whereas a fine-grain one would implement just a few. The choice of granularity lies in the tradeoffs in logic density, utilization, and performance (8). Typically, a coarse-grain logic module requires using a much smaller number of programmable switches and local interconnect, implying smaller area and delay. Therefore, coarse-grain programmable devices often have higher logic capacity and better timing performance. On the other hand, a fine-grain logic module can achieve a much higher logic utilization rate because it is easier to use small gates efficiently.

While FPGAs employ lookup tables, multiplexers, transistor pairs, and/or basic gates for their logic modules, most commercial CPLDs uses some type of SPLDs such as Programmable Array Logics (*PAL*s) or macrocells as their basic building modules (10,11). (A PAL comprises a programmable AND-plane feeding into

a fixed OR-plane, whereas the OR-plane in a PLA is also programmable). Often, the SPLD-based logic modules are grouped together, in a hierarchical manner, to form larger logic components.

Routing Architecture

The interconnect architecture of a CPLD/FPGA includes routing resources and their interconnection topology on the CPLD/FPGA. CPLD/FPGA routing resources consist of prefabricated wire segments and programmable switches; routing in a CPLD/FPGA is performed by programming the switches to connect the wire segments. There are four major interconnection topologies for commercial CPLDs/FPGAs: array-based, row-based, sea-of-gates, and hiearchical models.

An array-based FPGA consists of a two-dimensional array of logic modules which can be connected by general routing resources. [See Fig. 9(a).] As mentioned earlier, the logic modules house circuits that implement logic functions. The routing resources comprise horizontal and vertical routing tracks and user-programmable switches.

A row-based CPLD/FPGA consists of multiple rows of logic modules. [See Fig. 9(b).] Routing wires run in the channels between two adjacent rows and also inside the channels vertically. There are additional global vertical wires providing connections among different rows [not shown in Fig. 9(b)].

Unlike the array- and row-based architectures where there are routing channels separating logic modules, all logic modules in a sea-of-gates structure directly abut. [See Fig. 9(c).] The close proximity of logic modules allows direct connections between adjacent modules, significantly improving circuit performance. A routing network runs over the top of the modules, enabling efficient use of the available chip area.

The hierarchical structure is composed of logic modules connected by multiple levels of interconnect. [See Fig. 9(d).] At the first hierarchical level, several logic modules form a group, and local routing resources for the modules are provided. Together, the logic modules and their interconnect form a local component. Again, the second level of interconnect ties together a group of the components of the first level, with shared routing resources for the components. The construction for higher levels of components proceeds in the same manner.

Commercial Cpld/Fpga Examples

CPLD/FPGA architectures differ significantly from vendor to vendor. Due to the limitation of space, we shall give an overview for only the major programmable logic devices on the market today. Table 2 summarizes the architectural features of popular commercially available CPLDs/FPGAs.

Xilinx FPGAs/CPLDs. Xilinx introduced the first FPGA, the XC2000 series, in 1984 and now offers several more generations of FPGAs, including SRAM-based XC3000, XC4000, XC5200, and XC6200, and antifuse-based XC8100, and so forth. Besides the FPGA products, Xilinx also offers several families of CPLDs such as EPROM-based XC7300 and Flash-based XC9500. We focus on the SRAM-based XC4000 FPGAs because of their popularity on the market.

With the capacity ranging from 3 K to 85 K logic gates, the Xilinx XC4000-series FPGAs provide a wide range of selection. The XC4000-series FPGA adopts the symmetrical array-based architecture; each chip consists of a two-dimensional array of logic modules called configurable logic block (*CLB*s) and peripheral input/output blocks interconnected by horizontal and vertical routing channels.

XC4000 Logic Modules. The principal CLB structure is shown in Fig. 10. Each CLB of the XC4000 FPGA contains three lookup tables implementing combinational functions; two 4-input LUTs (implementing two functions labeled F' and G') feed a 3-input LUT which implements a function of F', G', and input $H1$. Setting the multiplexers properly, F' or H' can be connected to the combinational output X, and G' or H' can

Table 2. Summary of Major Commercially Available Programmable Devices

Company	Programmable Logic	Architecture	Logic-Module Type	Programming Technology	Example Products
Actel	FPGA	Row-based	Multiplexer	Antifuse	ACT 3
Altera	CPLD	Hierarchical-SPLD	SPLD block	EPROM	MAX 9000
				EEPROM	
		2-D array	Lookup table	SRAM	FLEX 10K
AMD/Vantis	CPLD	Hierarchical-SPLD	SPLD block	EEPROM	MACH 5
Aptix	FPIC	Symmetrical array	–	SRAM	FPIC/R
Cypress	CPLD	Hierarchical-SPLD	SPLD block	EPROM	
				Flash	Flash 370
Lattice	CPLD	Hierarchical-SPLD	SPLD block	EEPROM	ispLSI 6000
Lucent	FPGA	Symmetrical array	Lookup table	SRAM	ORCA 2
Motorola	FPGA	Hierarchical	Basic gates	SRAM	MPA1000
QuickLogic	FPGA	Symmetrical array	Multiplexer	Antifuse	pASIC2
Xilinx	FPGA	Symmetrical array	Lookup table	SRAM	XC4000
		Sea-of-gates	Multiplexer	SRAM	XC6200
		Sea-of-gates	Basic gates	Antifuse	XC8100
	CPLD	Hierarchical-SPLD	SPLD block	EPROM	XC7300
				Flash	XC9500

be assigned to the Y output. Therefore each CLB can implement any two functions of up to four variables, any single function of five variables, or selected functions of up to nine variables.

Each CLB also contains two edge-triggered D-type flip-flops for implementing sequential logic. Each of the D inputs is selected by one of the two 4-to-1 multiplexers which select from F', G', H', and input *DIN*. Thus the CLB can pass any of the combinational outputs to a flip-flop. If the combinational functions F', G', and H' are used as outputs and do not drive the flip-flops, *DIN* can be assigned to the flip-flops to store values from other sources.

XC4000 Routing Architecture. The routing channels in a Xilinx XC4000E-series FPGA consist of three major types of interconnect, distinguished by the relative length of their segments: single-length lines, double-length lines, and long lines. (See Fig. 11.) The single-length lines form a grid of horizontal and vertical lines that intersect at switch modules. The double-length lines consist of a grid of segments twice as long as the single-length lines. The long lines are a grid of segments that run the entire vertical or horizontal channel.

The horizontal and vertical single- and double-length lines can be interconnected using a programmable switch module shown in Fig. 12(a). There are six pass transistors per switch module interconnect point, used to establish connections between the lines. Several examples of connections are shown in Fig. 12(b). At point a, all four lines segments are electrically connected together by closing three transistors. At point b or c, two distinct signal nets pass through the interconnect point simultaneously by closing two nonadjacent transistors.

In addition to the wire segments mentioned above, there are also special interconnections for clocks and three-state buffers. Recently Xilinx offers the XC4000EX-series FPGAs which have several additional types of interconnect such as quad lines that span the length of four CLBs and direct interconnects that connects CLB outputs to adjacent neighbors.

Actel FPGAs. Actel now offers two major series of FPGAs, the Integrator and the Accelerator series, both with the row-based structure and the antifuse technology that provides one-time and nonvolatile programming. The Accelerator series is intended for higher-performance applications, while the Integrator offers a more attractive price. Due to the small size and low resistance/capacitance of the antifuse, the Actel FPGA

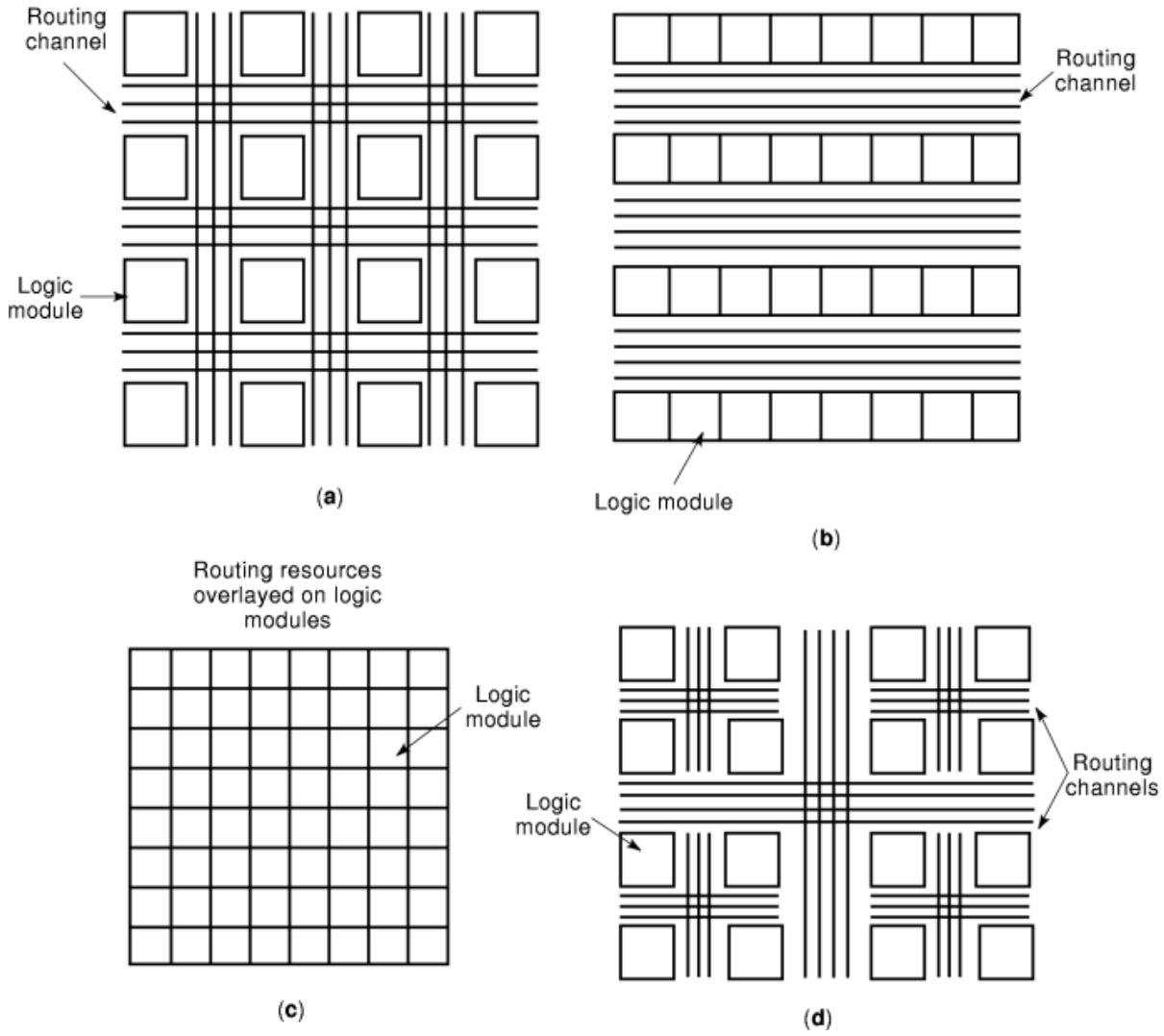

Fig. 9. The four major CPLD/FPGA architectures. (a) The array-based model. (b) The row-based model. (c) The sea-of-gates model. (d) The hierarchical model.

is compact and fast. We focus on the Accelerator ACT 3 FPGA in the following discussion. (See Fig. 13 for the basic architecture of an ACT 3 chip.)

ACT 3 Logic Module. The ACT 3 consists of two types of logic modules, C (combinational) modules and S (sequential) modules (see Fig. 14). The C module, as shown in Fig. 14(a) (and also in Fig. 8), is a multiplexer-based logic module that implements combinational logic functions. With the four data inputs d_0, d_1, d_2, d_3 and the four select inputs s_0, s_1, s_2, s_3, the C module can be used to implement 766 distinct functions of up to 8 variables. The S module uses a C-module front end with a dedicated edge-triggered D-type flip-flop on the output of the module, used for implementing sequential logic functions.

ACT 3 Routing Architecture. The Actel routing architecture consists of horizontal and vertical routing tracks. The routing tracks may either be of continuous length or broken into pieces called *segments*. Two

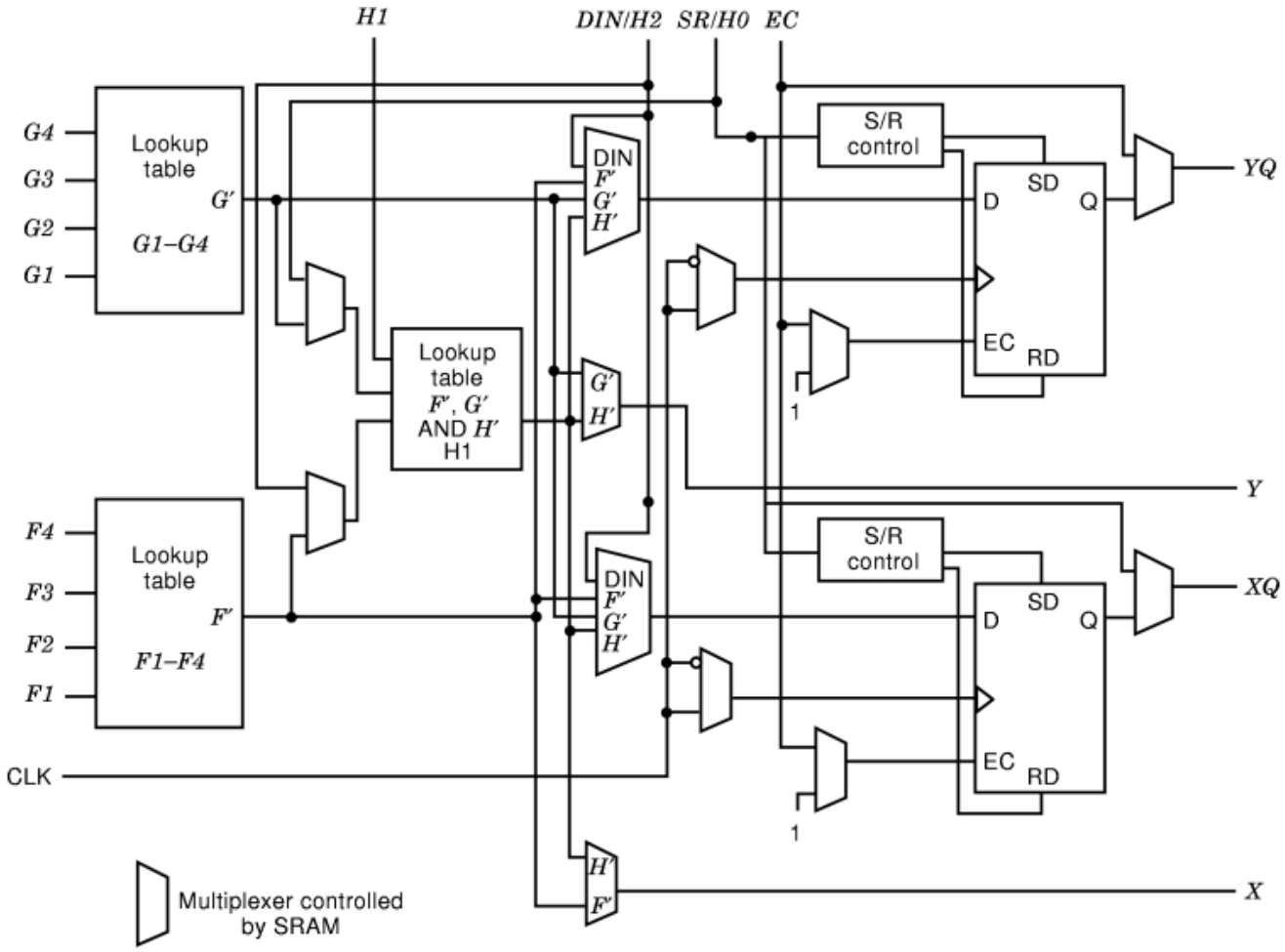

Fig. 10. The Xilinx XC4000 logic module. Each logic module contains two 4-input lookup tables, one 3-input lookup table, and two edge-triggered D-type flip-flops.

adjacent wire segments can be joined together by programming an antifuse switch to construct a longer connection. Many horizontal tracks lie in each channel between two adjacent rows of logic modules. The vertical tracks run through the logic modules and the horizontal channels. The segmented routing architecture leads to a better performance and routability tradeoff.

Altera CPLDs. Altera offers several types of CPLDs, including Classic, MAX, and FLEX series. The Classic- and MAX-series CPLDs are based on EPROM and/or EEPROM technologies and hierarchical-SPLD logic modules, whereas the FLEX-series CPLDs adopt the SRAM technology and lookup table-based logic modules. We focus on the FLEX-series CPLDs because of their state-of-the-art technology.

The FLEX programmable logic device, FLEX 8 K or FLEX 10 K, exhibits a combination of FPGA and CPLD technologies. Like CPLDs, the device consists of three levels of hierarchy and nonsegmented routing architectures; like FPGAs, its logic module is based on 4-input lookup tables. Figure 15 illustrates the FLEX 10 K architecture. The FLEX 10 K chip consists of a two-dimensional array of two types of logic modules, logic array blocks (*LAB*) and embedded array blocks (*EAB*s), interconnected by continuous row and column

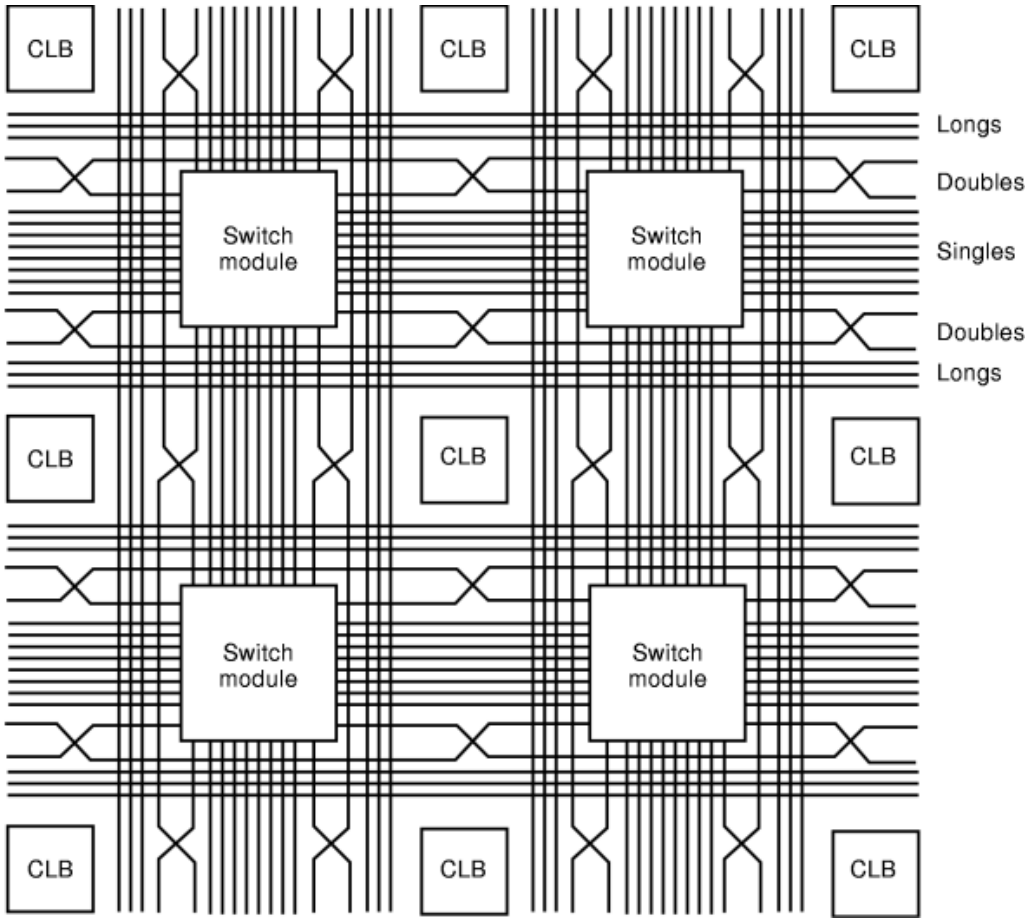

Fig. 11. The Xilinx XC4000E routing architecture. Each routing channel consists of three major types of interconnect: single-length lines, double-length lines, and long lines.

channels. Each LAB consists of eight logic elements that communicate through a local interconnect. Each logic element contains a 4-input lookup table, a D-type flip-flop, carry circuitry (for arithmetic operation), and cascade circuitry (for wide fan-in functions). Located on the left end of each row, each EAB provides 2 kbit of memory that is good for RAM and ROM functions. Also, the EAB can be configured to implement a more sophisticated function by using it as a large lookup table. The row and column routing channels consist of nonsegmented tracks that run the entire width and length of the device—similar to the long lines of Xilinx XC4000 FPGAs. Connections between horizontal and vertical routing channels are via active buffers and multiplexers. The homogeneous routing tracks make the FLEX CPLDs not only easier for design automation but also more predictable for delay estimation.

AMD/Vantis CPLDs. AMD/Vantis offers several series of CPLDs, including MACH 1, 2, ..., 5. The MACH-series CPLDs use the EEPROM programming technology and the hierarchical architecture, with their logic modules based on multiple PALs. In the following discussion, we focus on the most advanced MACH 5 series CPLDs.

The MACH 5 architecture consists of PAL modules connected by two levels of interconnect. (See Fig. 16.) Each group of four PAL modules has its own local routing resource called block interconnect. Together, the

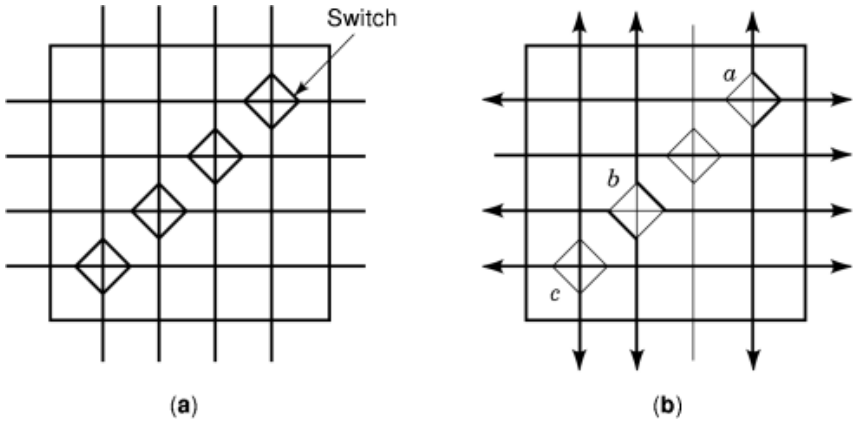

Fig. 12. The Xilinx XC4000 switch-module structure. (a) Six pass transistors per switch-module interconnect point. (b) Examples of connections.

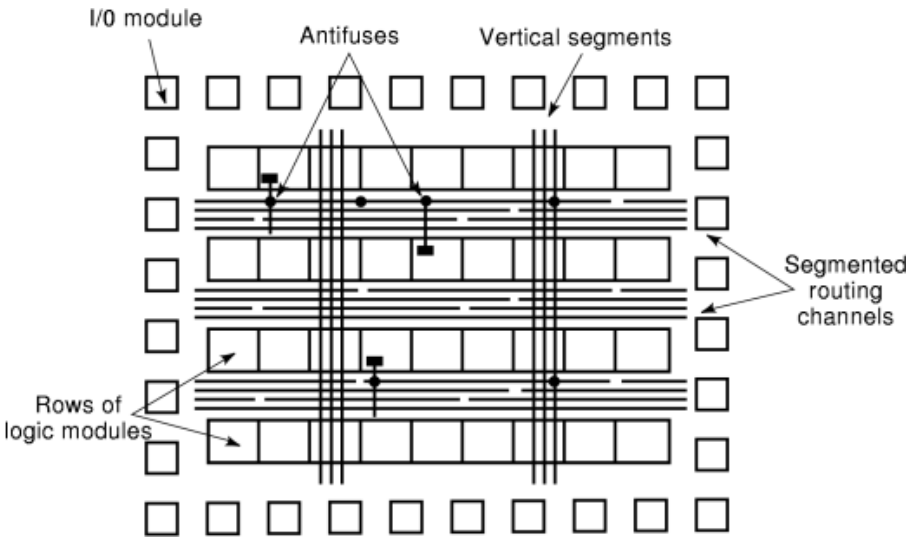

Fig. 13. The Actel ACT FPGA structure. An *ACT FPGA* consists of multiple rows of logic modules. Routing wires run in the channels between two adjacent rows and also inside the channel vertically.

four modules and their block interconnect form a supermodule. The second level of interconnect, the global interconnect, then ties all of the supermodules together.

The MACH 5 PAL block consists of a product-term array feeding into macrocells. The product-term array exhibits a sum-of-products architecture used in PAL devices. Comprising a D-type flip-flop, a control bus, and local routing resources, each macrocell can be configured for both combinational and sequential operations. A logic allocator is used to assign product terms to macrocells, and up to eight groups of four product terms can be fed to one macrocell. The logic allocator can also perform logic switching functions: as a design changes, the allocator will reassign logic to macrocells to retain pinout.

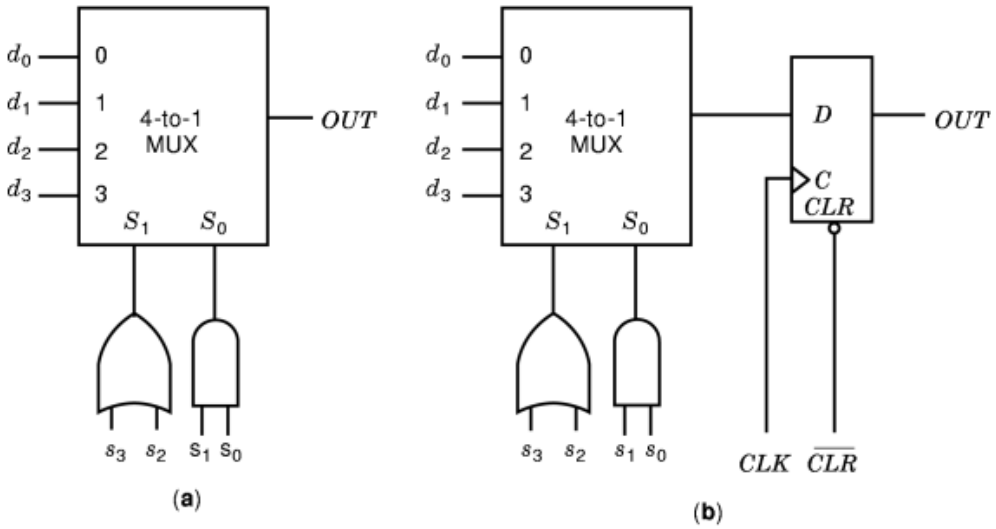

Fig. 14. The Actel ACT 3 logic module. (a) C (combinational) logic module; (b) S (sequential) logic module.

Design Process for Programmable Logic

This section briefly describes the CPLD/FPGA design flow; interested readers are referred to CAD for FPGA, (2,4), and data books from vendors (5,7,9,10,11,12,13) for more details. The CPLD/FPGA design cycle consists of three major stages described as follows:

(1) System Design The design process of CPLDs/FPGAs starts with system design, including specifying formal specifications of a system, drawing functional diagrams, and designing logic and circuit structures to realize the system. Typically, logic design uses schematic capture or hardware description language such as Verilog and VHDL. In system design, a designed circuit is usually optimized by a CAD tool and represented by Boolean expressions or circuit diagrams. Then, the optimized, technology-independent circuit is mapped to a set of logic modules of the target CPLD/FPGA; the mapping process is called *technology mapping*. Together, the process of logic optimization and technology mapping is called *logic synthesis*.

(2) Physical Design In the physical design, the optimized and mapped circuits are then converted into geometric patterns. The CPLD/FPGA physical design process consists of three major steps: partitioning, placement, and routing. A mapped circuit is partitioned into a set of subcircuits so that each subcircuit can fit into a logic module of the device. Then, the logic modules are placed in the device by a placement program. In the final step of the physical design, a router assigns wire segments and chooses programmable switches to establish the required connections among the logic modules.

(3) Customization Upon successful completion of routing, the output from the physical design process is fed to a programming unit, which customizes the programmable logic chip.

Note that, though not mentioned in the above, testing, simulation, and verification are integrated into many stages of the design process. While the tasks in the system design, except technology mapping, are common between programmable logic designs and traditional semicustom designs, those in technology mapping and the physical design are specific to the target technologies.

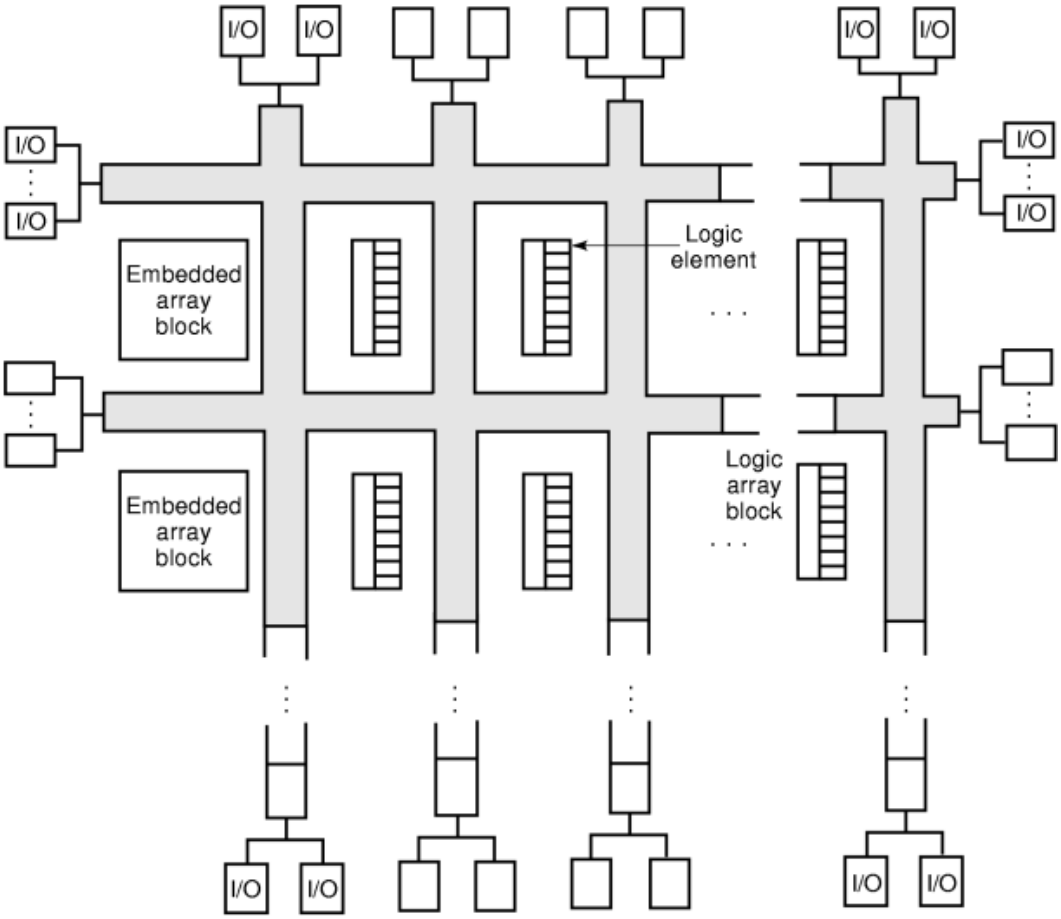

Fig. 15. The Altera FLEX 10K structure. A *FLEX 10K* chip consists of a two-dimensional array of two types of logic modules, logic array blocks and embedded array blocks, interconnected by continuous row and column channels.

Research in Cpld/Fpga Architectures

This section briefly describes important research results in the architectural design for CPLDs/FPGAs.

Most research in CPLD/FPGA architectures is based on the model shown in Fig. 17. In the model, a CPLD/FPGA consists of an array of logic modules which can be connected by routing resources. The routing resources comprise segments of wires and two kinds of modules, switch modules and connection modules, which contain user-programmable switches. An intersection of a horizontal and a vertical routing channels is referred to as a switch module; the switch module serves to connect wire segments, and this requires using programmable switches inside it. Connection modules are used to connect logic-module pins to wire segments; this is performed by programming switches inside connection modules.

Logic-Module Design.

Effect of Logic-Module Granularity on CPLD/FPGA Density and Performance. The effect of logic-module granularity on CPLD/FPGA density and performance has been studied extensively (14,15,16,17). The larger the granularity of a logic module, the smaller number and level of modules are required to implement a design. However, larger granularity of a logic module needs more circuitry in a module, implying more area is

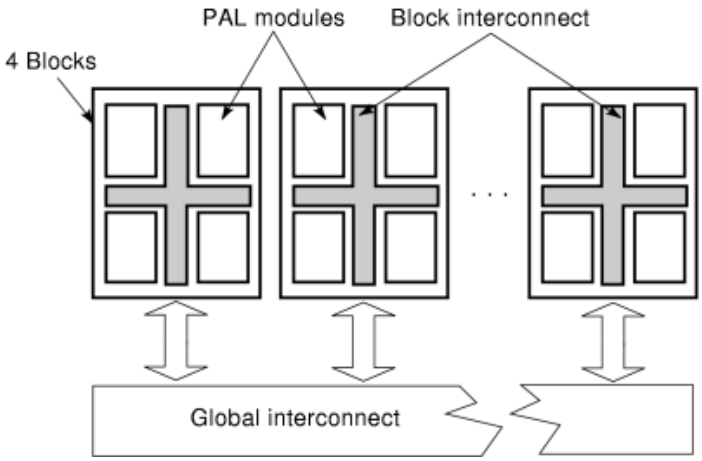

Fig. 16. The AMD/Vantis MACH structure. A MACH 5 *CPLD* is composed of *PAL* modules connected by two levels of interconnect.

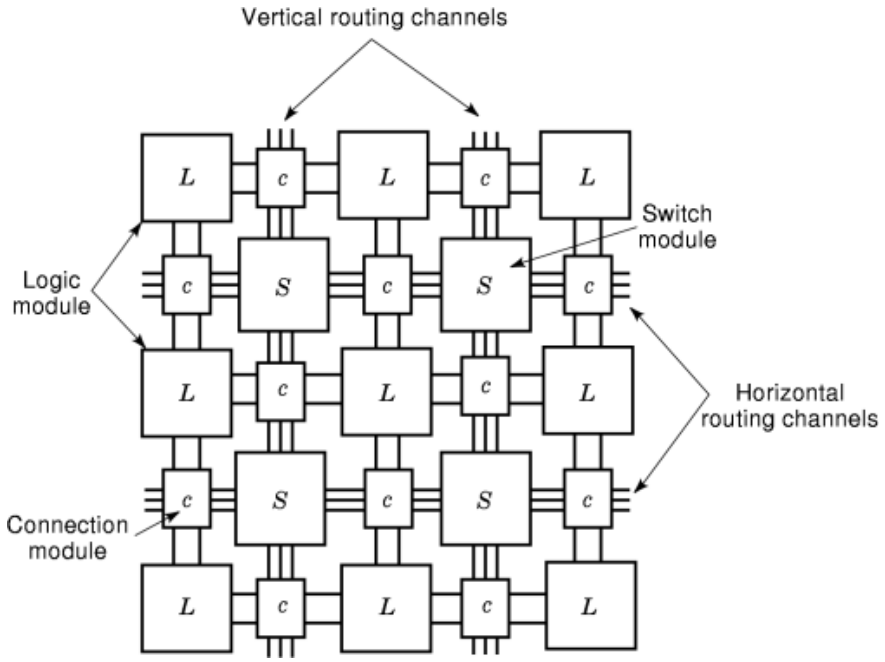

Fig. 17. The *FPGA* model. The FPGA model consists of an array of logic modules which can be connected by routing resources (switch modules, connection modules, and wire segments).

consumed and longer routing delay is incurred for a module. Hence, there are tradeoffs between logic-module granularity and area efficiency, and between the granularity and timing performance. The study by Kouloheris and El Gamal (15) shows that a 4-input, 1-output LUT gives best area efficiency of any K-input, M-output LUT-based logic module. Based on the study of (14,17), the effect of granularity on performance depends on the

RC delay in the programmable switch used: If the RC delay is relatively small, the best granularity is around $K = 3$ or 4; on the other hand, if the RC delay is larger, the best granularity is around $K = 6$ or 7.

Effect of Logic-Module Architecture on CPLD/FPGA Density and Performance. In addition to the study on the effect of a fixed-input LUT, other researchers have also investigated other logic-module architectures, including PLA-based logic modules (15), heterogeneous modules (18), hybrid modules (19), and universal logic modules (20,21).

Kouloheris and El Gammal explored the tradeoff between the area of a PLA-based CPLD/FPGA and its module granularity (15). They observed that a PLA-based architecture achieves the best area efficiency when a module with 8–10 inputs, 3–4 outputs, and 12–13 product terms, and the areas for the best PLA-based logic module and the 4-input 1-output LUT-based one are comparable when implemented by the same programming technology.

He and Rose noticed that 4-input LUTs make most efficient use of area, whereas more coarse-grain modules such as 5-, 6-, and 7-input LUTs are superior in terms of timing performance. Based on this observation, they proposed to use various sized LUTs within a CPLD/FPGA chip to provide a better tradeoff between density and performance (18). They investigated the effect of the combination of two differently sized LUTs, say a p- and q-input LUT combination [denoted by (p, q)], and suggest the combinations such as (5, 2), (4, 2), or (4, 3) to be better than a homogeneous 4-input LUT.

Kaviani and Brown presented a hybrid logic-module architecture based on a combination of LUTs and PALs/PLAs (19). The basis for this idea is that the strength of LUT-based architectures lies in their very high logic density, while the PLA/PAL-based ones usually achieve very high speed performance. Their hybrid architecture employs 4-input LUTs in the mix, since they have been found to be best for many applications. The PAL/PLA hybrid uses a programmable AND-plane connected to a half fixed and a half programmable OR-plane. Kaviani and Brown suggested allocating half of the logic-module area to 4-input LUTs and the other half to hybrid PLAs/PALs.

One other segment of research in logic modules focuses on designing the modules with high functionality, called *universal logic-module design.* The goal is to design a function $U(y_1, \ldots, y_m)$, with the minimum m, which can realize all the n-input functions $F(x_1, \ldots, x_n)$. The general approach is to specify the set of functions that can be covered by U by assigning y_i to 0, 1, x_j, or $\bar{x}_j$, permuting or complementing some inputs, or negating the output of U (10,23). Lin, Marek-Sadowska, and Gatlin (20) and Thakur and Wong (21) independently extended the above idea to the logic-module design for CPLD/FPGAs. While the techniques used in (20) are computationally intensive, the work by (21) gives a more efficient algebraic approach which is suited to larger designs.

Routing Resources. CPLD/FPGA routing resources consist of programmable switches and wire segments. Routing in CPLD/FPGA is performed by programming the switches to connect the wire segments. The programmable switches usually incur significantly high RC delay, and consume a large amount of area. Due to the area and delay constraints, the number of switches that can be placed in a switch module is usually limited, implying limited routability. Therefore, there is a basic tradeoff between routability and area/performance for switch-module architectures. Also, to achieve high performance, the number of switches used for routing a net is restricted. On the other hand, fewer switches would reduce routability. The tradeoff between routability and performance suggests the existence of an "optimal" segmentation architecture that needs to be explored.

Switch-Module Modeling. Zhu, Wong, and Chang gave the switch-module modeling as follows (22). A switch module is a square block with W terminals on each side. There are two types of switch modules, switch blocks and switch matrices (see Fig. 18). In a switch block [Fig. 18(a)], terminals on different sides may be connected by programmable switches, and connecting two terminals requires using exactly one switch. Switch blocks are used for connections among various types of interconnect in an FPGA (13). A switch matrix [Fig. 18(b)] consists of a grid of horizontal and vertical tracks. There are two types of switches in a switch matrix, crossing switches and separating switches. A crossing switch is used to connect two crossing tracks, and a separating switch can be used to make two terminals of a track electrically disconnected, and thus these two

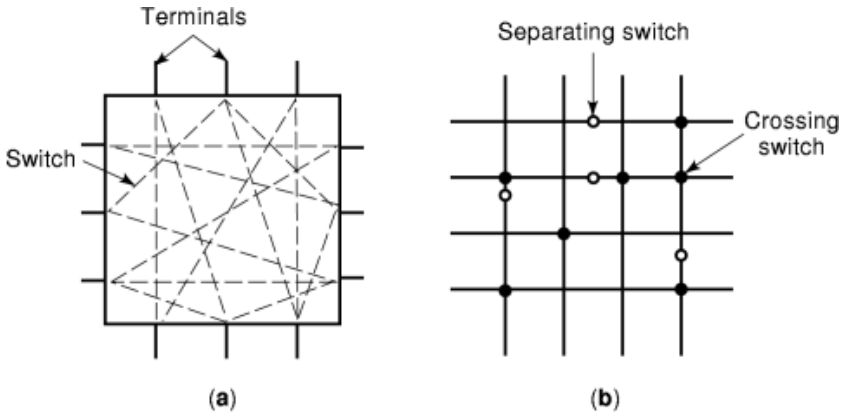

Fig. 18. Two types of switch modules. (a) A switch block; (b) A switch matrix.

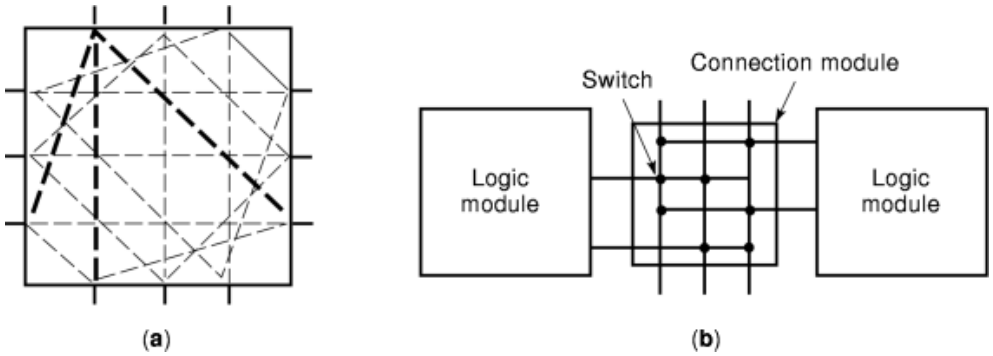

Fig. 19. Flexibility. (a) A switch module with $F_S = 3$; (b) A connection module with $F_C = 2$.

terminals can be used independently. In Fig. 18(b), the crossing switches are represented by solid circles and the separating switches by hollow ones. Switch matrices are used in various commercially available FPGAs (5,10,12,13), CPLDs (10,11), and FPICs (23).

Switch-Block Architectures. The effects of switch-block architectures on routing were first studied experimentally by Rose and Brown (24). The work in (24) defines the flexibility of a switch block, represented by F_S, as the number of programming switches between a terminal and others, and the flexibility of a connection module, denoted by F_C, as the number of tracks to which a logic-module pin can be connected; for example, the switch block in Fig. 19(a) has $F_S = 3$, and the connection module in Fig. 19(b) has $F_C = 2$. These definitions assume that all terminals and logic-module pins have the same degree of connectivity.

Rose and Brown investigated the effects of different switch-block and connection-module flexibilities on routing (24). The primary conclusions of the study in (24) are that for a high probability of 100% routing completion:

- F_C usually needs to be greater than one-half of the number of tracks in a routing channel.
- A relatively low F_S with as small as three is often sufficient.

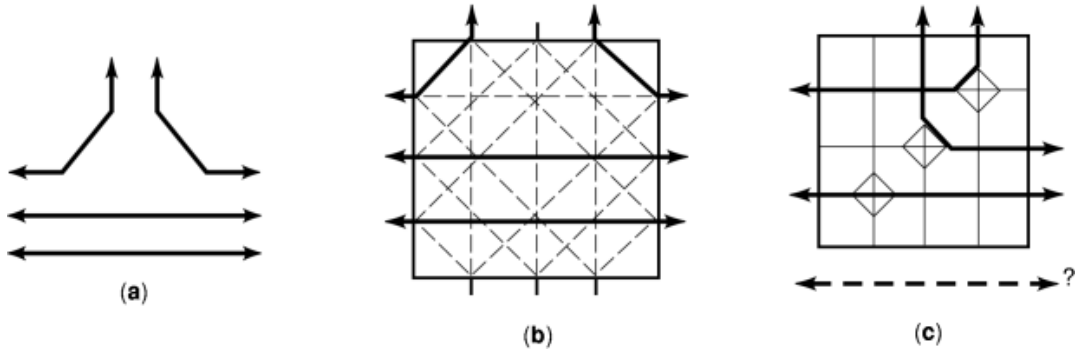

Fig. 20. Routing on two types of switch blocks of the same size and both with 18 switches. (a) A routing instance with four nets; (b) A symmetrical, universal switch block on which the routing instance of (a) is routable; (c) A Xilinx XC4000-type switch block on which the routing instance of (a) is not routable. Notice that one straight net cannot be routed in the figure shown.

These conclusions lead to the architectural choice: $F_S = 3$ combined with a high F_C, which gives the best area and routability tradeoff. A theoretical study of flexibility and routability was later presented based on a stochastic model (25), which confirms the experimental results in (24).

Chang, Wong, and Wong further investigated the various topologies of switch blocks of $F_S = 3$ and presented a class of universal switch blocks (26). A switch module M with W terminals on each side is said to be *universal* if every set of nets satisfying the dimensional constraint (i.e., the number of nets on each side of M is at most W) is simultaneously routable through M. They showed that $6W$ switches and $F_S = 3$ are sufficient to construct a "cheapest" universal switch block. They also proved that each of the universal switch blocks can accommodate significantly more routing instances than the Xilinx XC4000-type one of the same size. Figure 20(b) illustrates a routing on a symmetrical, universal switch block for the instance of (a); the same routing instance is not feasible for the XC4000-type switch block [see Fig. 20(c)]. The work also provides a theoretical insight into the important observation by Rose and Brown (24) that $F_S = 3$ is often sufficient to provide high routability.

Switch-Matrix Architectures. The switch-matrix architecture was first modeled and investigated by Zhu, Wong, and Chang (22). They studied the feasibility conditions of switch matrices and presented a heuristic guided by stochastic information for switch-matrix design. Chang, Wong, and Wong later proposed a network-flow-based approach for switch-matrix design (27); this approach leads to switch matrices with substantially higher routability. Wu and Chang studied the universality of switch *matrices* (34). They showed that there exist no universal switch matrices; however, they presented a class of "cheapest" *quasi-universal* switch matrices which have almost the same routing capacity as universal switch *blocks*.

Sun, Wang, Wong, and Liu (28) considered a variation of the switch-matrix architecture on which "separating switches" are placed outside a switch matrix. According to their study, this class of routing architectures has the advantage of using a smaller number of switches (and thus smaller signal delays) at the expense of consuming a larger number of routing tracks per channel, compared with the devices with switch blocks. Based on a stochastic model, they proved that 100% routing completion in that architecture can be achieved with a high probability using a reasonable track count in each channel.

Wiring Segmentation. The segmentation design problem is to arrange segmented tracks in the routing channels to optimize routability and performance simultaneously, preferably comparable to the case in mask programmable gate arrays.

El Gamal et al. showed that with appropriate arrangement of segment lengths, the channel width needed for routing completion can be close to that for gate arrays (29,30). Zhu and Wong presented an algorithm for the

channel segmentation problem based on a stochastic analysis (31). Given an arbitrary distribution of nets and routing requirements, they computed the number of segmented tracks of various types. Pedram, Nobandegani, and Preas later extended the stochastic approach to an analytical model for designing effective segmented channel architectures (32).

For the segmentation design for a symmetrical array-based programmable logic device, Zhu, Wong, and Chang observed that it can be done in two phases: channel segmentation design followed by switch-module design (22). In the first phase, a channel segmentation is constructed for each horizontal and vertical channel independently. In the second phase, a switch module is constructed at each channel intersection. Based on the similar idea, Mak and Wong later employed a decomposition procedure and showed in details how the two-phase approach can be done (33).

Chang, Lin, and Wong recently proposed a graph matching-based algorithm for channel- and array-based segmentation design (35); the algorithm substantially improves the routability of an FPGA chip. More significantly, the graph matching-based algorithm can readily be extended to higher-dimensional segmentation design, which is crucial to the design of large-scale CPLDs/FPGAs.

BIBLIOGRAPHY

1. W. Carter *et al.* A user programmable reconfigurable gate array, *Proc. 1986 IEEE Custom Integr. Circuits Conf.*, 1986, pp. 233–235.
2. S. D. Brown *et al. Field-Programmable Gate Arrays*, Norwell, MA: Kluwer, 1992.
3. J. Oldfield R. Dorf *Field Programmable Gate Arrays*, New York: Wiley, 1995.
4. S. Trimberger (ed.) *Field-Programmable Gate Array Technology*, Norwell, MA: Kluwer, 1994.
5. Actel Corp., *FPGA Data Book and Design Guide*, 1996.
6. E. Hamdy *et al.* Dielectric-based antifuse for logic and memory ICs, *IEEE Int. Electron Devices Meet. Tech. Dig.*, Piscataway, NJ: 1988, pp. 786–789.
7. QuickLogic Corp., *QuickLogic 1996/97*, Santa Clara, 1996.
8. J. Rose A. El Gamal A. Sangiovanni-Vincentelli Architecture of field-programmable gate arrays, *Proc. IEEE*, **81**: 1013–1029, 1993.
9. Concurrent Logic, *CFA6006 Field-Programmable Gate Arrays Data Sheet*, Sunnyvale, CA, 1991.
10. Altera Corp., *FLEX 10K Handbook*, 1996.
11. AMD/Vantis Inc., *The Mach 5 Family*, 1996.
12. Lucent Technologies Inc., *ORCA Series Field-Programmable Gate Arrays*, 1996.
13. Xilinx Inc., *The Programmable Logic Data Book*, 1996.
14. J. Kouloheris A. El Gamal FPGA performance vs. cell granularity, *Proc. IEEE Custom Integr. Circuits Conf.*, 1991, pp. 6.2.1–6.2.4.
15. J. Kouloheris A. El Gamal FPGA area versus cell granularity—Lookup tables and PLA cells, *Proc. ACM First Int. Workshop Field Programmable Gate Arrays*, 1992, pp. 9–14.
16. J. Rose *et al.* Architecture of programmable gate arrays: The effect of logic block functionality on area efficiency, *IEEE J. Solid-State Circuits*, **25** (5): 1217–1225, 1990.
17. S. Singh *et al.* The effect of logic block architecture on FPGA performance, *IEEE J. Solid State Circuits*, **27** (3): 281–287, 1992.
18. J. He J. Rose Advantages of heterogeneous logic block architectures for FPGAs, *Proc. IEEE Custom Integr. Circuits Conf.*, 1993, pp. 7.4.1–7.4.5.
19. A. Kaviani S. D. Brown Hybrid FPGA architecture, *Proc. ACM Int. Symp. Field Programmable Gate Arrays*, 1996, pp. 3–9.
20. C. Lin M. Marek-Sadowska D. Gatlin Universal logic gate for FPGA design, *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, San Jose, CA, Nov. 1994, pp. 164–168.
21. S. Thakur D. F. Wong On designing ULM-based FPGA logic modules, *Proc. ACM Int. Symp. Field Programmable Gate Arrays*, Monterey, CA, 1995, pp. 3–9.

22. K. Zhu D. F. Wong Y.-W. Chang Switch module design with application to two-dimensional segmentation design, *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Santa Clara, 1993, pp. 481–486.
23. Aptix Inc., *Programmable Interconnect System Data Book*, Nov. 1993.
24. J. Rose S. D. Brown Flexibility of interconnection structures for field-programmable gate arrays, *IEEE J. Solid State Circuits*, **26** (3): 277–282, 1991.
25. S. D. Brown J. Rose Z. G. Vranesic A stochastic model to predict the routability of field-programmable gate arrays, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, **12**: 1827–1838, 1993.
26. Y.-W. Chang D. F. Wong C. K. Wong Universal switch modules for FPGA design, *ACM Trans. Des. Automation of Electronic Syst.*, **1** (1): 80–101, 1996.
27. Y.-W. Chang D. F. Wong C. K. Wong Design and analysis of FPGA/FPIC switch modules, *Proc. IEEE Int. Conf. Comput. Des., VLSI Comput. and Processors*, Austin, TX, 1995, pp. 394–401.
28. Y. Sun *et al.* Routing for symmetric FPGA's and FPIC's, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, **16**: 20–31, 1997.
29. A. El Gamal *et al.* An architecture for electrically configurable gate arrays, *IEEE J. Solid-State Circuits*, **24** (2): 394–398, 1989.
30. A. El Gamal J. Greene V. Roychowdhury Segmented channel routing is nearly as efficient as channel routing (and just as hard), *Advanced Research in VLSI*, Univ. of California, Santa Cruz, 1991, pp. 193–211.
31. K. Zhu D. F. Wong On channel segmentation design for row-based FPGAs, *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Santa Clara, CA, 1992, pp. 26–29.
32. M. Pedram B. Nobandegani B. Preas Design and analysis of segmented routing channels for row-based FPGA's, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, **13**: 1470–1479, 1994.
33. W.-K. Mak D. F. Wong Channel segmentation design for symmetrical FPGAs, *Proc. IEEE Int. Conf. Comput. Des., VLSI Comput. and Processors*, 1997.
34. G.-M. Wu Y.-W. Chang Switch-matrix architecture and routing for FPDs, *Proc. ACM. Int. Symp. Physical Design*, Monterey, CA, 1998.
35. Y.-W. Chang J.-M. Liu D. F. Wong Graph matching-based algorithms for FPGA segmentation design, *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, San Jose, CA, 1998.
36. S. D. Brown J. Rose FPGA and CPLD architectures: A tutorial, *IEEE Des. & Test. of Comput.*, **13** (1): 42–57, 1996.

YAO-WEN CHANG
National Chiao Tung University
D. F. WONG
University of Texas at Austin
C. K. WONG
The Chinese University of Hong Kong