**Programmable Interconnect**

- In addition to programmable cells, programmable ASICs must have programmable interconnect to connect cells together to form logic function.
- Structure and complexity of the interconnect is determined primarily by the programming technology and architecture of the basic cell.
- Interconnect is typically done on aluminum-based metal layers:
  - Resistance of approximately 50 mΩ/square
  - Line capacitance of approximately 0.2 pF/cm
- Early programmable ASICs had two metal interconnect layers, but current, high density parts may have three or more metal layers.

---

**Actel Programmable Interconnect**

- Actel interconnect is similar to a channeled gate array:
  - Horizontal routing channels between rows of logic modules
  - Vertical routing channels on top of cells
- Each channel has a fixed number of tracks each of which holds one wire.
- Wires in track are divided into segments of various lengths - segmented channel routing.
- Long vertical tracks (LVT) extend the entire height of the chip.
- Each logic module has connections to its inputs and outputs called stubs:
  - Input stubs extend vertically into routing channels above and below logic module.
  - Output stub extends vertically 2 channels up and 2 channels down.
- Wires are connected by antifuses.

---

**Figure 7.1** The interconnect architecture used in an Actel ACT family FPGA.
**Detail of ACT1 Channel Architecture**

- 22 horizontal tracks per channel for signal routing with 3 dedicated for VDD, GND, GCLK
- 8 vertical tracks per channel are available for inputs (4 from the LM above the channel, 4 from the LM below)
- Input stub
- 4 vertical tracks per LM for outputs — output stub
  - A vertical track extends across the two channels above the module and the two channels below
  - 1 long vertical track (spans the entire height of the chip)

---

**Elmore’s Constant**

- Approximation of waveform at node $i$:
  
  \[ V_i(t) = \frac{1}{R_{ki}} \int_{t_i}^{t_f} R_{o} C_i \, dt \]

  where $R_{ki}$ is the resistance of the path to $V_0$ shared by node $k$ and node $i$

  - Examples: $R_{24} = R_1$, $R_{22} = R_1 + R_2$, and $R_{31} = R_1$
  - If the switching points are assumed to be at the 0.35 and 0.65 points, the delay at node $i$ can be approximated by $\tau_{DI}$

---

**RC Delay in Antifuse Connections**

- A four-antifuse connection: $L_0$ is an output stub, $L_1$ and $L_3$ are horizontal tracks, $L_2$ is a long vertical track (LVT), and $L_4$ is an input stub.
- An RC-tree model: Each antifuse is modeled by a resistance and each interconnect segment is modeled by a capacitance.

---

**Routing Resources**

- ACT 1 interconnection architecture:
  - 22 horizontal tracks per channel for signal routing with 3 dedicated for VDD, GND, GCLK
  - 8 vertical tracks per LM are available for inputs (4 from the LM above the channel, 4 from the LM below)
  - Input stub
  - 4 vertical tracks per LM for outputs — output stub
  - A vertical track extends across the two channels above the module and the two channels below
  - 1 long vertical track (spans the entire height of the chip)
RC Delay in Antifuse Connections (cont’d)

- $R_n$ - resistance of antifuse, $C_n$ - capacitance of wire segment

\[
\tau_D = R_1 C_1 + R_2 C_2 + R_3 C_3 + R_4 C_4
\]

If all antifuse resistances are approximately equal and much larger than the resistance of the wire segment, then: $R_1 = R_2 = R_3 = R_4$, and:

\[
\tau_D = 4 R C_4 + 3 R C_3 + 2 R C_2 + R C_1
\]

- A connection with two antifuses will generate a $3RC$ time constant, a connection with three antifuses will generate a $6RC$ time constant, and a connection with 4 antifuses will generate a $10RC$ time constant.

- Interconnect delay grows quadratically ($\propto n^2$) as the number of antifuses $n$ increases.

Xilinx LCA Interconnect

- Xilinx LCA interconnect has a hierarchical architecture:
  - Vertical lines and horizontal lines run between CLBs
  - General-purpose interconnect joins switch boxes (also known as magic boxes or switching matrices)
  - Long lines run across the entire chip - can be used to form internal buses using the three-state buffers that are next to each CLB
  - Direct connections bypass the switch matrices and directly connect adjacent CLBs
  - Programmable Interconnect Points (PIPs) are programmable pass transistors the connect CLB inputs and outputs to the routing network
  - Bi-directional interconnect buffers (BIDI) restore the logic level and logic strength on long interconnect paths

Figure 7.5 Xilinx LCA interconnect. (a) The LCA architecture (notice the matrix element size is larger than a CLB). (b) A simplified representation of the interconnect resources. (c) A detailed view inside the switching matrix showing the pass-transistor arrangement. (d) The equivalent circuit for the connection between nets 6 and 20 using the matrix. (e) A view of the interconnect at a Programmable Interconnection Point (PIP). (f) and (g) The equivalent schematic of a PIP connection (h) The complete RC delay path.
Xilinx EPLD Interconnect

- Xilinx EPLD family uses an interconnect bus called a Universal Interconnection Module (UIM).
- UIM is a programmable AND array with constant delay from any input to any output.

![Figure 7.7](image1.png)  
**Figure 7.7** The Xilinx EPLD UIM (Universal Interconnection Module). (a) A simplified block diagram of the UIM. The UIM bus width, n, varies from 68 (XC7236) to 198 (XC73108). (b) The UIM is actually a large programmable AND array. (c) The parasitic capacitance of the EPROM cell.

Altera MAX 5000 and 7000 Interconnect

- Altera MAX 5000 and 7000 devices use a Programmable Interconnect Array (PIA).
- PIA is also a programmable AND array with constant delay from any input to any output.

![Figure 7.8](image2.png)  
**Figure 7.8** A simplified block diagram of the Altera MAX interconnect scheme. (a) The PIA (Programmable Interconnect Array) is deterministic - delay is independent of the path length. (b) Each LAB (Logic Array Block) contains a programmable AND array. (c) Interconnect timing within a LAB is also fixed.

Altera MAX 9000 Interconnect Architecture

- Altera MAX 9000 devices use long row and column wires (FastTracks) connected by switches.

![Figure 7.9](image3.png)  
**Figure 7.9** The Altera MAX 9000 interconnect scheme. (a) A 4 X 5 array of Logic Array Blocks (LABs), the same size as the EPM9400 chip. (b) A simplified block diagram of the interconnect architecture showing the connection of the FastTrack buses to a LAB.

Altera Flex

- Altera Flex devices also use FastTracks connected by switches, but the wiring is more dense (as are the logic modules).

![Figure 7.10](image4.png)  
**Figure 7.10** The Altera FLEX interconnect scheme. (a) The row and column FastTrack interconnect. (b) A simplified diagram of the interconnect architecture showing the connections between the FastTrack buses and a LAB.
Summary

- Antifuse FPGA architectures are dense and regular
- SRAM architectures contain nested structures of interconnect resources
- Complex PLD architectures use long interconnect lines but achieve deterministic routing

I/O Requirements

- I/O cells handle driving signals off chip
- Receiving and conditioning external inputs
- Supplying power and ground and
- Handling such things as electrostatic protection
- Different types of I/O requirements
  - DC output - driving a resistive load at DC or low frequency, LEDs, relays, small motors, etc.
  - AC output - driving a capacitive load with a high-speed logic signal off-chip, data or address bus, serial data line, etc.
  - DC input - reading the value of a sensor, switch, or another logic chip
  - AC input - reading the value of high-speed signals from another chip
  - Clock input - system or synchronous bus inputs
  - Power input - supplying power (and ground) to the I/O cells and logic core

Motor Control (Robotic Arm) Application

Figure 6.1 A robot arm. (a) Three small DC motors drive the arm. (b) Switches control each motor.

Can we replace the switches with an FPGA outputs and drive the motors directly?
CMOS Output Buffer

- CMOS output buffer has finite (non-zero) output resistance
- Data books specify typically A \( (V_{\text{olmax}}, I_{\text{olmax}}) \) and B \( (V_{\text{ohmin}}, I_{\text{ohmax}}) \)
  - Xilinx XC5200: A (0.4V, 8.0mA), B (4V, -0.8mA)
- Typical output currents that can be driven by a standard digital I/O pad are in the range of 50mA to 200mA

I/O Circuit for High Current Motor Control

Can we drive the motors by connecting several output buffers in parallel to reach a peak drive current of 0.5A?

Some FPGA vendors do specifically allow connecting adjacent output cells in parallel.

Problems?

Totem-Pole Output

- Uses two n channel transistors as output drivers
- Advantage is that it has a higher output drive for a ‘1’ output
- Disadvantage is that output voltage will not be higher than VDD - V_Tn

AC Output

- AC outputs are often used to connect to a bi-directional bus - bus transceivers
- This functionality requires the capability for three-state (tri-state) outputs - ‘0’, ‘1’, and high-impedance or hi-z
- In addition to rise and fall times, bidirectional I/O pads have timing parameters related to the hi-z state (float time):
  - \( t_{\text{EZL}} \) - output hi-Z to ‘0’ time
  - \( t_{\text{EHL}} \) - output ‘0’ to hi-Z
  - \( t_{\text{EZH}} \) - output hi-Z to ‘1’
  - \( t_{\text{EHZ}} \) - output ‘1’ to hi-Z

Figure 6.2 (a) A CMOS complementary output buffer. (b) Pull-down transistor M2 sinks a current \( I_{\text{ol}} \) through a pull-up resistor \( R_1 \). (c) Pull-up transistor M1 sources current \( I_{\text{oh}} \) through a pull-down resistor \( R_2 \). (d) Output characteristics.

Figure 6.4 Output buffer characteristics. (a) A CMOS totem-pole output stage (b) Totem-pole output characteristics. (c) Clamp diodes. (d) The clamp diodes start to conduct if the output voltage exceeds the supply voltage bounds.
3 State Bus Example

Figure 6.5 A three-state bus. (a) Bus parasitic capacitance. (b) The output buffers in each chip. The ASIC CHIP1 contains a bus keeper, BK1.

3 State Bus Timing

Figure 6.6 Three-state bus timing for Figure 6.5.

1) CHIP2 drives BUSA.B1 high
2) CHIP2.OE goes low, floating the bus; the bus will stay high because we have a bus keeper
3) CHIP3.OE goes high, and the buffer drives a low

Characterizing AC Output Pads

Figure 6.7 (a) The test circuit for characterizing the ACT2 and ACT3 I/O delay parameters. (b) Output buffer propagation delays from the data input to PAD. (c) Three-state delay with D low. (d) Three-state delay with D high.

Supply (GND) Bounce

- Ground (also VDD) net has finite parasitic resistance and inductance
- Switching a load through a pull-down transistor causes a 2nd order response (ground bounce or ringing) on ground net
- Ground bounce can cause glitching on other logic signals

Figure 6.8 Supply bounce. (a) As the pull-down device M1 switches, it causes the GND net to bounce. (b) The supply bounce is dependent on the output state and the output load. (c) Ground bounce can cause other output buffers to generate a logic path. (d) Bounce can also cause errors on other inputs.
Transmission Lines

- Driving large capacitive loads at high speed gives rise to transmission line effects
- Transmission lines are defined by their characteristic impedance - determined by their physical characteristics
- Maximum energy transfer occurs when the source impedance matches the transmission line impedance
  \[ V_o = V_i \left( \frac{Z_o}{Z_o + R_o} \right) \]
- The time it takes the signal wave to propagate down the transmission line is called the time-of-flight \( t_f \)
- Typical time-of-flight for a PCB trace is on the order of 1 ns for every 15 cm of trace (about 1/2 the speed of light)
- When the signal wave is launched into the transmission line, it travels to the other end and is reflected back to the source
- Transmission line effects become important if the rise time of the driver is less than \( 2t_f \)

Terminating a Transmission Line

- Methods to terminate a transmission line:
  - Open circuit or capacitive termination - bus termination is the input capacitance of the receivers
  - Parallel resistive termination - requires substantial DC current - used in bipolar logic
  - Thévenin termination - reduces DC current on the drivers, but adds resistance across the source
  - Series termination - total series resistance (source and termination) equals the line impedance
  - Parallel termination - requires a third power supply
  - Parallel termination with series capacitance - eliminates DC current but introduces other problems
- Some high-speed busses actually use the reflection facilitate the data transmission (PCI bus)
- Other techniques include current-mode signaling or differential signals

Transmission Line Example

- Figure 6.9 Transmission lines. (a) A printed-circuit board (PCB) trace is a transmission line. (b) A driver launches an incident wave which is reflected at the end of the line. (c) A connection starts to look like a transmission line when the signal rise time is about equal to twice the delay.

Terminating a Transmission Line (cont.)

- Figure 6.10 Transmission line termination. (a) Open-circuit or capacitive termination. (b) Parallel resistive termination. (c) Thévenin termination. (d) Series termination at the source. (e) Parallel termination using a voltage bias. (f) Parallel termination with a series capacitor.
DC Input - Switch Bounce

- A pull-up or pull-down resistor is generally required on input buffers to keep input from floating to indeterminate logic levels.
- If the input is from a mechanical switch, the contacts may bounce, producing several transitions through the switching threshold.
- Some technique for debouncing mechanical switch inputs is usually necessary.

Debouncing Using Hysteresis

- A Schmitt-trigger inverter is used to prevent glitches.
- A typical FPGA input buffer with a hysteresis of 200mV centered around a threshold of 1.4 V.

Noise Margins - Another Representation

- Transfer characteristics of a CMOS inverter are shown.
- CMOS thresholds are represented as plug/socket clearance.

Noise Margins - Interfacing TTL and CMOS

- TTL and CMOS logic thresholds are compared.
- Raising V_{OH,min} solves the problem.
Noise Margins - Mixed Voltage Systems (e.g. 3.3V and 5V)

Figure 6.15 Mixed-voltage systems. (a) TTL levels. (b) Low-voltage CMOS levels. (c) A mixed-voltage ASIC. (d) A problem when connecting two chips with different supply voltages - caused by the input clamp diodes.

Metastability Example

Figure 6.16 Metastability. (a) Data coming from one system is an asynchronous input to another. (b) A flip-flop has a very narrow decision window bounded by the setup and hold times. If the data input changes inside this decision window, the output may be metastable - neither '1' or '0'.

Probability of Upset

• An upset is when a flip-flop output should have been a '0' and was a '1' or visa-versa
• Probability of upset is:

\[ P = T_0 e^{-t_c} \]

where \( t_0 \) is the resolution time and \( T_0 \) and \( t_c \) are constants of the flip-flop implementation

• Mean time between upsets (MTBU - similar to mean time between failures) is:

\[ MTBU = \frac{1}{e^{t_f}/t_{clock/data}} \]

where \( f_{clock} \) is the clock frequency and \( f_{data} \) is the data frequency

Probability of Upset Example

• Assume \( t_f = 5 \) ns, \( t_c = 0.1 \) ns, and \( T_0 = 0.1 \)s:

\[ P = 0.1e^{-0.1\times10^{-9}} = 2\times10^{-23} \]

• Assume \( f_{clock} = 100 \) MHz and \( f_{data} = 1 \) MHz:

\[ MTBU = \frac{5\times10^{-9}}{e^{5\times10^{-9}}/100\times10^{6} \times (6.3) = 5.2\times10^{9} \text{ sec} = 16 \text{ years} } \]

if we have a bus with 64 inputs, each using a flip-flop as above, the MTBU of the system is three months
Constants $\tau_c$, $T_0$

- $\tau_c$ – the inverse of the gain-bandwidth product of the sampler at the instant of sampling
  - may be determined by a small signal analysis of the sampler at the sampling instant or by measurement
  - we cannot change it
- $T_0$ (units of time) – function of process technology and the circuit design
  - may be different for sampling a positive or negative edge
  - usually only one value is given
  - may be determined by measurement and simulation
  - we cannot change it

MTBF as a Function of Resolution Time

Figure 6.17 Mean time between failures (MTBF) as a function of resolution time.

Clock Input

- Most FPGAs and PLDs provide a dedicated clock input(s)
- Clock input needs to be low latency $t_{\text{lat}}$, but also low skew $t_{\text{skew}}$
- Low skew is ensured by using a dedicated, balanced clock tree, but this tends to increase clock latency
- Example: Actel ACT1 FPGAs have a clock latency that can be as high as 15ns if the clock drives over 300 loads (flip-flops), but the skew is stated to be in the sub nanosecond range
- Large clock latency causes hold time restrictions on data inputs – data gets to the flip-flops faster than clock and must remain there until clock arrives

Clock Input Example

Figure 6.18 Clock input. (a) Timing model with values for Xilinx XC4005-6. (b) A simplified view of clock distribution. (c) Timing diagram. Xilinx eliminates the variable internal delay $t_{\text{lat}}$ by specifying a pin-to-pin setup time $t_{\text{PSUe}} = 2\text{ns}$.
Programmable Input Delay to Eliminate Hold Time on Data Inputs

- Pin-to-pin setup time: 2.2 ns
- Pin-to-pin hold time: 3.5 ns
- Programmable delay

Figure 6.19 Programmable input delay. (a) Pin-to-pin timing model with values from an XC4005-6. (b) Timing diagrams with and without programmable delay.

Effect of Clock Latency on Registered Outputs

- Clock buffer
- IO pad
- Programmable delay

Figure 6.20 Registered output. (a) Timing model with values for an XC4005-6 programmed with the fast slew rate option. (b) Timing diagram.

Power Input

- All devices require inputs for VDD and Gnd during operation and programming voltage, VPP, during programming
- Larger devices with greater logic capacity require more power pins to supply the necessary power while maintaining a reasonable per-pin current limit
  - This reduces the number of signal pins possible for larger devices
- Some types of FPGAs (e.g. Xilinx) have their own power-on reset sequence to reset flip-flops, initialize and load SRAM, etc.

Power Dissipation

- General rule
  - plastic package can dissipate 1W
  - more expensive ceramic packages can dissipate about 2W
- Actel ACT 1 formula
  - Total chip power = 0.2 (N x F1) + 0.085 (M x F2) + 0.8 ( P x F3) mW
  - F1 = average logic module switching rate in MHz
  - F2 = average clock pin switching rate in MHz
  - F3 = average I/O switching rate in MHz
  - M = number of logic modules connected to the clock pin
  - N = number of logic modules used on the chip
  - P = number of I/O pairs used (input + output), with 50pF load
Power Dissipation (cont'd)

- An Example: Actel 1020B-2
  - Assumptions:
    - Clock is 20 MHz
    - 547 logic modules, each switches at an average speed of 5 MHz
    - 69 I/O modules, each switches at an average speed of 5 MHz
  - $P_{LU} = (0.2)(547)(5) = 547 \text{ mW}$
  - $P_D = (0.8)(69)(5) = 276 \text{ mW}$
  - $P_{CLK} = (0.085)(547)(0.2)(5) = 46.495 \text{ mW}$
  - $P_{MAX} = 869.5 \text{ mW}$
  - Max thermal resistance $\theta_{JA}$ is approximately $68^\circ\text{CW}^{-1}$ for VQFP (Very thin plastic Quad Flatpack)
  - Assuming worst-case industry conditions $T_A = 85^\circ\text{C}$
    - $T_f = 85 + 0.87 \times 68 = 144.16^\circ\text{C}$
    - Actel specifies $T_{JMAX} = 150^\circ\text{C}$

Example FPGA I/O Block

- Output features
  - Switch between totem-pole and complementary output
  - Include a passive pull-up or pull-down
  - Invert the 3-state control (OE)
  - Include a flip-flop, or latch, or a direct connection in the output path

- Input features
  - Configure the input buffer with TTL or CMOS thresholds
  - Include a flip-flop, or latch, or direct connection in the input path
  - Switch in a delay to eliminate an input hold time

Timing Model with I/O Block

Figure 6.22 The Xilinx XC4000 family input/output block (I/OB).
Example FPGA I/O Block (cont.)

Figure 6.23 A simplified block diagram of the Altera I/O Control Block (IOC) used in the MAX 5000 and MAX 7000 series.

Example FPGA I/O Block (cont.)

Figure 6.24 A simplified block diagram of the Altera I/O Element (IOE) used in the Flex 8000 and 10k series.

Summary

• Options available in I/O cells
  – different drive strengths, TTL compatibility, registered or direct inputs, registered or direct outputs, pull-up resistors, over-voltage protection, slew-rate control, boundary-scan test (JTAG)

• Important points to remember
  – outputs typically source or sink 5-10mA continuously into a DC load, and 50-200mA transiently into an AC load
  – input buffers can be CMOS (Tr. 2.5V) or TTL (1.4V)
  – input buffers normally have a small hysteresis (0.1-0.2V)
  – CMOS inputs must never be left floating
  – Clamp diodes are present on every pin
  – inputs and outputs can be registered or direct
  – I/O registers can be in the I/O cell or in the core
  – metastability is a problem when working with asynchronous inputs