Appendix A: IC Interfacing, System Design, and Failure Analysis

The invention of the transistor and the subsequent advent of integrated circuit (IC) technology is believed by many to be the start of the second industrial revolution. In this chapter we provide an overview of IC technology and interfacing. In addition, we look at the computer system as a whole and examine some general considerations in system design. In Section A.1 we provide an overview of IC technology. IC interfacing and system design considerations are examined in Section A.2. In Section A.2 we also discuss failure analysis in systems.

Section A.1: Overview of IC Technology

In this section we provide an overview of IC technology and discuss some developments in logic families.

The transistor was invented in 1947 by three scientists at Bell Laboratories. In the 1950s, transistors replaced vacuum tubes in many electronics systems, including computers. It was not until in 1959 that the first integrated circuit was successfully fabricated and tested by Jack Kilby of Texas Instruments. Prior to the invention of the IC, the use of transistors, along with other discrete components such as capacitors and resistors, was common in computer design. Early transistors were made of germanium, which was later abandoned in favor of silicon. This was due to the fact that the slightest rise in temperature resulted in massive current flows in germanium-based transistors. In semiconductor terms, it is because the band gap of germanium is much smaller than that of silicon, resulting in a massive flow of electrons from the valence band to the conduction band when the temperature rises even slightly. By the late 1960s and early 1970s, the use of the silicon-based IC was widespread in mainframes and minicomputers. Transistors and ICs were based on P-type materials. Due to the fact that the speed of electrons is much higher (about two and a half times) than the speed of the holes, N-type devices replaced P-type devices. By the mid-1970s, NPN and NMOS transistors had replaced the slower PNP and PMOS transistors in every sector of the electronics industry, including in the design of microprocessors and computers. Since the early 1980s, CMOS (complementary MOS) has become the dominant method of IC design. Next we provide an overview of differences between MOS and bipolar transistors.

MOS vs. bipolar transistors

There are two type of transistors: bipolar and MOS (metal-oxide semiconductor). Both have three leads. In bipolar transistors, the three leads are referred to as the *emitter*, *base*, and *collector*, while in MOS transistors they are named *source*, *gate*, and *drain*. In bipolar, the carrier flows from the emitter to the collector and the base is used as a flow controller. In MOS, the carrier flows from the source to the drain and the gate is used as a flow controller. See Figure A-1.

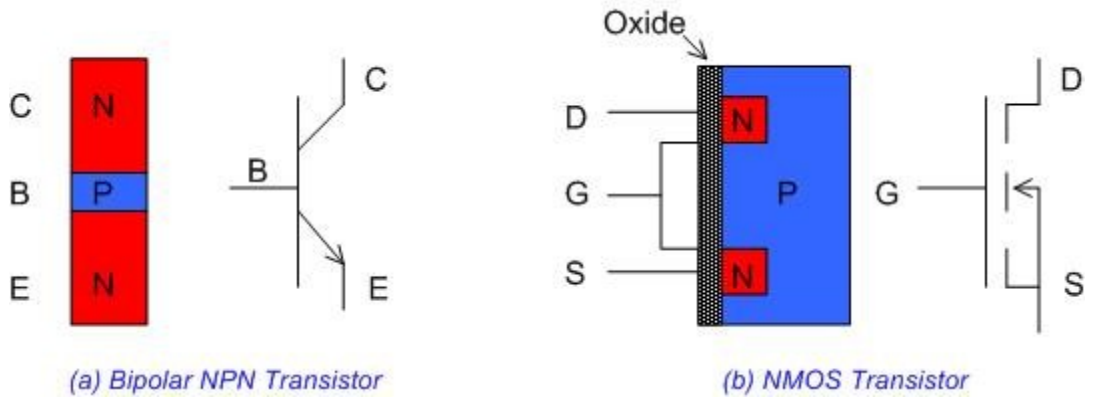


Figure A-1: Bipolar vs. MOS Transistors

In NPN-type bipolar transistors, the electron carrier leaving the emitter must overcome two voltage barriers before it reaches the collector. One is the N-P junction of the emitter-base and the other is the P-N junction of the base-collector. The voltage barrier of the base-collector is the most difficult one for the electrons to overcome (since it is reversed biased) and it causes the most power dissipation. This led to the design of the unipolar type transistor called *MOS*. In N-channel MOS transistors, the electrons leave the source reaching the drain without going through any voltage barrier. The absence of any voltage barrier in the path of the carrier is one reason why MOS dissipates much less power than bipolar transistors. The low power dissipation of MOS allows putting millions of transistors on a single IC chip. In today's million-transistor microprocessors and DRAM memory chips, the use of MOS technology is indispensable. Without the MOS transistor, the advent of desktop personal computers would not have been possible, at least not so soon. The use of bipolar transistors in both the mainframe and minicomputer of the 1960s and 1970s required expensive cooling systems and large rooms due to their bulkiness. MOS transistors do have one major drawback: They are slower than bipolar transistors. This is due partly to the gate capacitance of the MOS transistor. For MOS to be turned on, the input capacitor of the gate takes time to charge up to the turn-on (threshold) voltage, leading to a longer propagation delay.

Overview of logic families

Logic families are judged according to (1) speed, (2) power dissipation, (3) noise immunity, (4) input/output interface compatibility, and (5) cost. Desirable qualities are high speed, low power dissipation, and high noise immunity (since it prevents the occurrence of false logic signals during switching transition). In interfacing logic families, the more inputs that can be driven by a single output, the better. This means that high-driving-capability outputs are desired. This plus the fact that the input and output voltage levels of MOS and bipolar transistors are not compatible means that one must be concerned with the ability of one logic family driving the other one. In terms of the cost of a given logic family, it is high during the early years of its introduction and prices decline as production and use rise.

The case of inverters

As an example of logic gates, we look at a simple inverter. In a one-transistor inverter, while the transistor plays the role of a switch, R is the pull-up resistor. See Figure A-2.

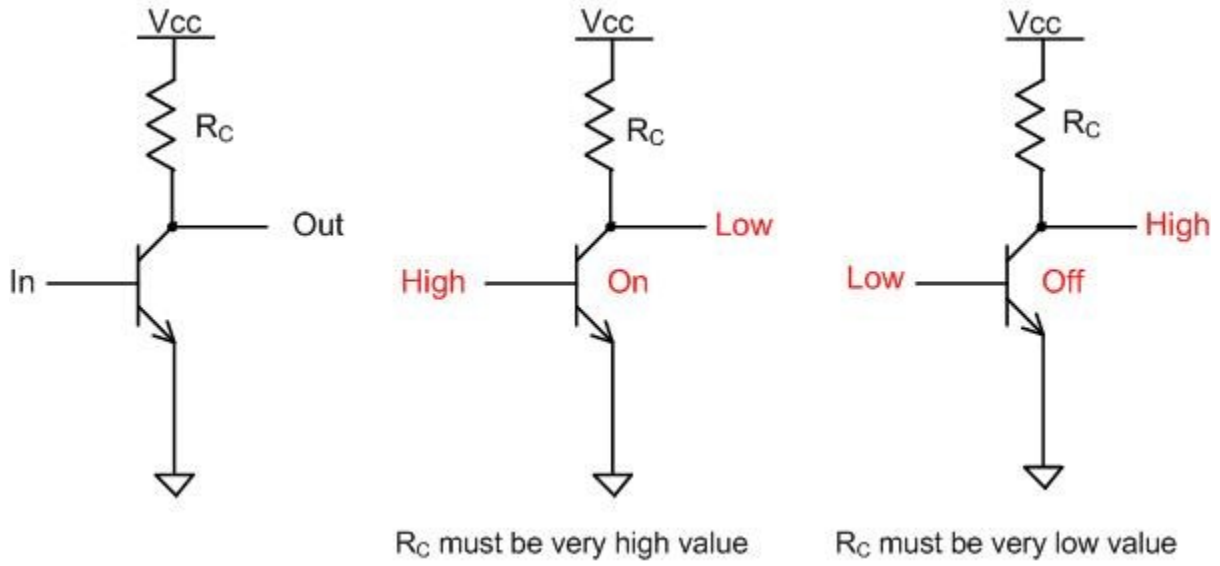


Figure A-2: One-Transistor Inverter with Pull-up Resistor

However, for this inverter to work effectively in digital circuits, the R value must be high when the transistor is "on" to limit the current flow from VCC to ground in order to have low power dissipation (P = VI, where V = 5 V). In other words, the lower the I, the lower the power dissipation. On the other hand, when the transistor is "off", R must be a small value to limit the voltage drop across R, thereby making sure that VOUT is close to VCC. These are opposing demands on the value of R. This is one reason that logic gate designers use active components (transistors) instead of passive components (resistors) to implement the pull-up resistor R.

The case of a TTL inverter with totem pole output is shown in Figure A-3. In Figure A-3, Q3 plays the role of a pull-up resistor.

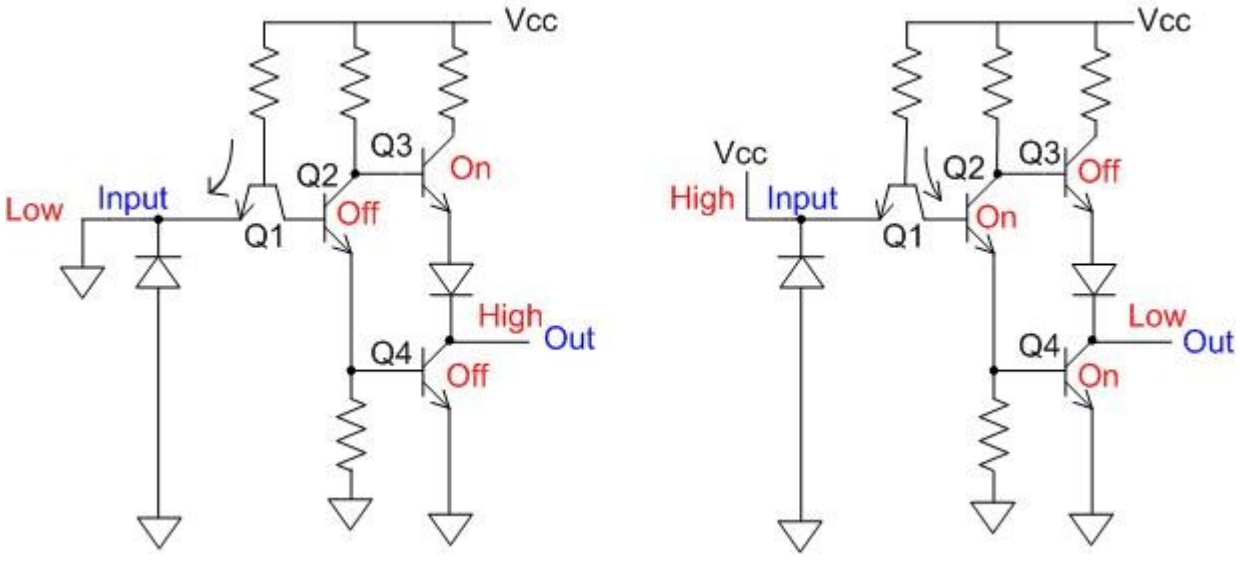


Figure A-3: TTL Inverter with Totem-Pole Output

CMOS inverter

In the case of CMOS-based logic gates, PMOS and NMOS are used to construct a CMOS (complementary MOS) inverter as shown in Figure A-4.

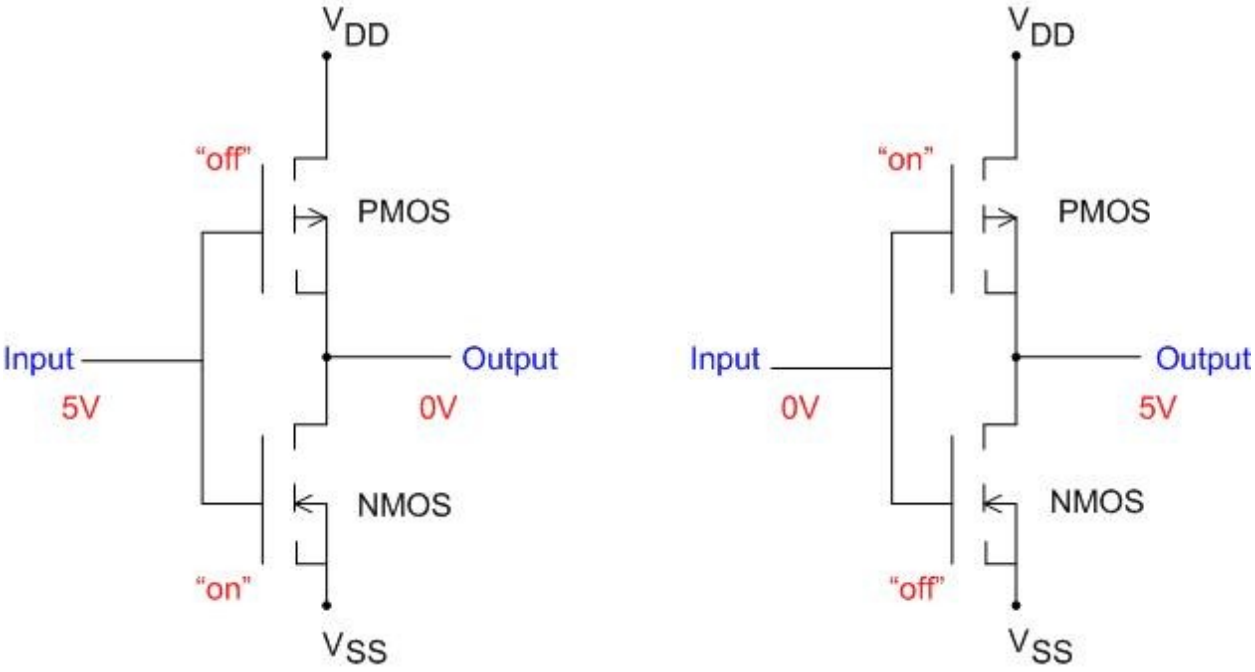


Figure A-4: CMOS Inverter

In CMOS inverters, when the PMOS transistor is off, it provides a very high impedance path, making leakage current almost zero (about 10 nA); when the PMOS is on, it provides a low resistance on the path of V_{DD} to load. Since the speed of the hole is slower than that of the electron, the PMOS transistor is wider to compensate for this disparity; therefore, PMOS transistors take more space than NMOS.

Input, output characteristics of some logic families

In 1968 the first logic family made of bipolar transistors was marketed. It was commonly referred to as the standard TTL (transistor-transistor logic) family. The first MOS-based logic family, the CD4000/74C series, was marketed in 1970. The addition of the Schottky diode to the base-collector of bipolar transistors in the early 1970s gave rise to the S family. The Schottky diode shortens the propagation delay of the TTL family by preventing the collector from going into what is called *deep saturation*. Table A-1 lists major characteristics of some logic families. In Table A-1, note that as the CMOS circuit's operating frequency rises, the power dissipation also increases. This is not the case for bipolar-based TTL.

Characteristic	STD TTL	LSTTL	ALSTTL	HCMOS
V_{CC}	5V	5V	5V	5V
V_{IH}	2.0V	2.0V	2.0V	3.15V
V_{IL}	0.8V	0.8V	0.8V	1.1V
V_{OH}	2.4V	2.7V	2.7V	3.7V
V_{OL}	0.4V	0.5V	0.4V	0.4V
I_{IL}	-1.6 mA	-0.36 mA	-0.2 mA	-1 µA
I_{IH}	40 µA	20 µA	20 µA	1 µA
I_{OL}	16 mA	8 mA	4 mA	4 mA
I_{OH}	-400 µA	-400 µA	-400 µA	4 mA
Propagation delay	10 ns	9.5 ns	4 ns	9 ns
Static power dissipation (f=0)	10 mW	2 mW	1 mW	0.0025 nW
Dynamic power dissipation at f = 100 kHz	10 mW	2 mW	1 mW	0.17 mW

Table A-1: Characteristics of Some Logic Families

History of logic families

Early logic families and microprocessors required both positive and negative power voltages. In the mid-1970s, 5V VCC became standard. For example, Intel's 4004, 8008, and 8080 all used negative and positive voltages for the power supply. In the late 1970s, advances in IC technology allowed combining the speed and drive of the S family with the lower power of LS to form a new logic family called FAST (Fairchild Advanced Schottky TTL). In 1985, AC/ACT (Advanced CMOS Technology), a much higher speed version of HCMOS, was introduced. With the introduction of FCT (Fast CMOS Technology) in 1986, at last the speed gap between CMOS and TTL was closed. Since FCT is the CMOS version of FAST, it has the low power consumption of CMOS but the speed is comparable with TTL. Table A-2 provides an overview of logic families up to FCT.

Product	Year Introduced	Speed (ns)	Static Supply Current (mA)	High/Low Family Drive (mA)
Std TTL	1968	40	30	-2/32
CD4K/74C	1970	70	0.3	-0.48/6.4
LS/S	1971	18	54	-15/24
HC/HCT	1977	25	0.08	-6/-6
FAST	1978	6.5	90	-15/64
AS	1980	6.2	90	-15/64
ALS	1980	10	27	-15/64
AC/ACT	1985	10	0.08	-24/24
FCT	1986	6.5	1.5	-15/64
Reprinted by permission of Electronic Design Magazine, c. 1991.				

Table A-2: Logic Family Overview

Recent advances in logic families

As the speed of high-performance microprocessors such as the 386 and 486 reached 25 MHz, it shortened the CPU's cycle time, leaving less time for the path delay. Designers normally allocate no more than 25% of a CPU's cycle time budget to path delay. Following this rule means that there must be a corresponding decline in the propagation delay of logic families used in the address and data path as the system frequency is increased. In recent years, many semiconductor manufacturers have responded to this need by providing logic families that have high speed, low noise, and high drive. Table A-3 provides the characteristics of high-performance logic families introduced in recent years.

Family	Year	Number Suppliers	Tech Base	I/O Level	Speed (ns)	Static Current	I_{OH}/I_{OL}
ACQ	1989	2	CMOS	CMOS/CMOS	6.0	80 µA	-24/24
ACTQ	1989	2	CMOS	TTL/CMOS	7.5	80 µA	-24/24
FCTx	1987	3	CMOS	TTL/CMOS	4.1–4.8	1.5 mA	-15/64
FCTxT	1990	2	CMOS	TTL/TTL	4.1–4.8	1.5 mA	-15/64
FASTr	1990	1	Bipolar	TTL/TTL	3.9	50 mA	-15/64
BCT	1987	2	BICMOS	TTL/TTL	5.5	10 mA	-15/64
Reprinted by permission of Electronic Design Magazine, c. 1991.							

Table A-3: Advanced Logic General Characteristics

ACQ/ACTQ are the second-generation advanced CMOS (ACMOS) with much lower noise. While ACQ has the CMOS input level, ACQT is equipped with TTL-level input. The FCTx and FCTx-T are second-generation FCT with much higher speed. The x in the FCTx and FCTx-T refers to various speed grades, such as A, B, and C, where the A designation means low speed and C means high speed. For designers who are well versed in using the FAST logic family, the use of FASTr is an ideal choice since it is faster than FAST, has higher driving capability (I_{OL}, I_{OH}), and produces much lower noise than FAST. At the time of this writing, next to ECL and gallium arsenide logic gates, FASTr is the fastest logic family in the market (with the 5V VCC), but the power consumption is high relative to other logic families, as shown in Table A-3. Since early 2000, a 3.3V VCC with higher speed and lower power consumption has become standard. The combining of high-speed bipolar TTL and the low power consumption of CMOS has given birth to what is called BICMOS. Although BICMOS seems to be the future trend in IC design, at this time it is expensive due to the extra steps required in BICMOS IC fabrication, but in some cases there is no other choice. For example, Intel's Pentium microprocessor, a BICMOS product, had to use high-speed bipolar transistors to speed up some of the internal functions in order to keep up with RISC processor performance. Table A-3 provides advanced logic characteristics. Table A-4 shows logic families used in systems with different speeds. The x is for the different speeds where A, B, and C are used for designation. A is the slowest one while C is the fastest one. The above data is for the 'LS244 buffer.

System Clock Speed (MHz)	Clock Period (ns)	Predominant Logic for Path
2 – 10	100 – 500	HC, LS
10 – 30	33 – 100	ALS, AS, FAST, FACT
30 – 66	15 – 33	FASTr, BCT, FCTA

Table A-4: Importance of Speed

Review Questions

1. State the main advantages of MOS and bipolar transistors.
2. True or false. In logic families, the higher the noise margin, the better.
3. True or false. Generally, high-speed logic consumes more power.
4. Power dissipation increases linearly with the increase in frequency in _____ (CMOS, TTL).
5. In a CMOS inverter, indicate which transistor is on when the input is high.
6. For system frequencies of 10–30 MHz, which logic families are used for the address and data path?

Section A.2: IC Interfacing and System Design Issues

There are several issues to be considered in designing a microprocessor-based system. They are IC fan-out, capacitance derating, ground bounce, V_{CC} bounce, crosstalk, transmission lines, power dissipation, and chip failure analysis. This section provides an overview of these design issues in order to provide a sampling of what is involved in high-performance system design.

IC fan-out

In IC interfacing, fan-out/fan-in is a major issue. How many inputs can an output signal drive? This question must be addressed for both logic "0" and logic "1" outputs. Fan-out for low and fan-out for high are as follows:

$$\text{Fan-out (of low)} = \frac{I_{OL}}{I_{IL}} \qquad \text{Fan-out (of high)} = \frac{I_{OH}}{I_{IH}}$$

Of the above two values the lower number is used to ensure the proper noise margin. Figure A-5 shows the sinking and sourcing of current when ICs are connected.

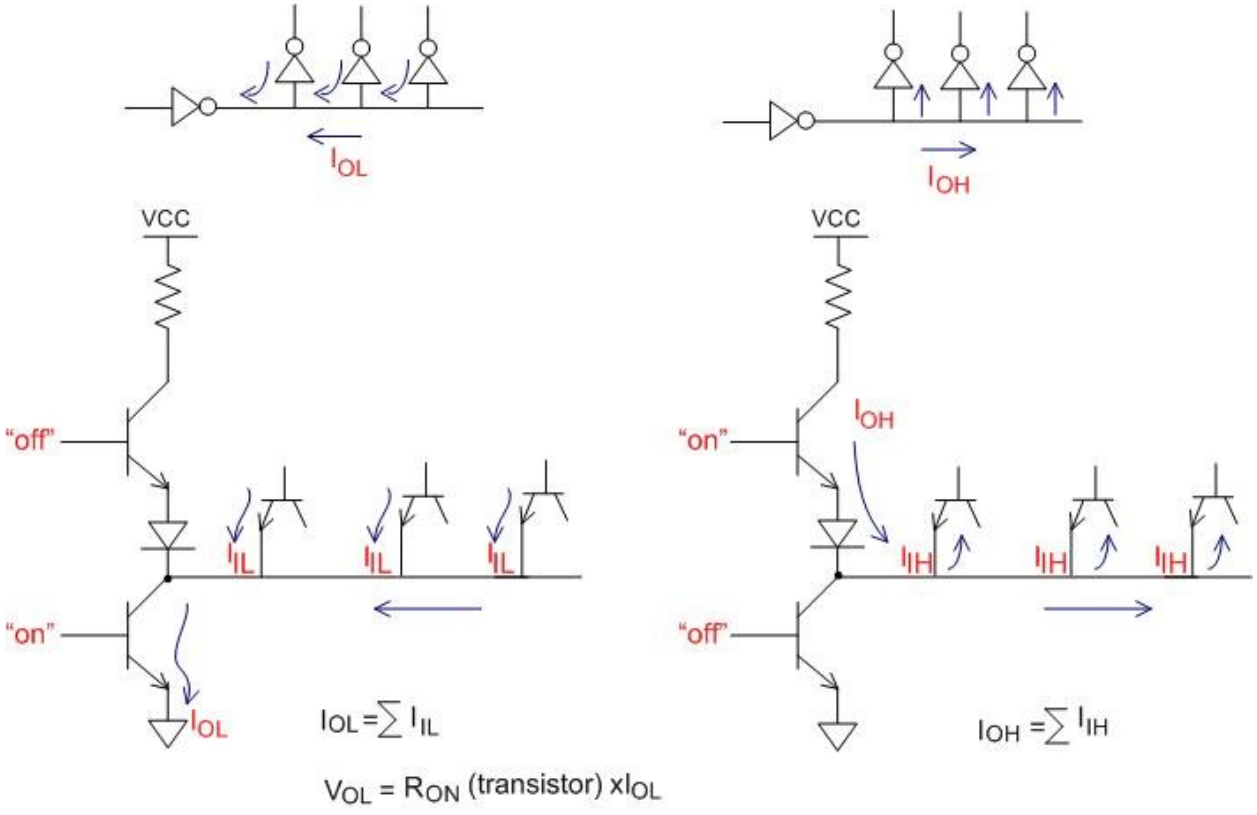


Figure A-5: Current Sinking and Sourcing in TTL

In Figure A-5, as the number of inputs connected to the output increases, I_{OL} rises, which causes VOL to rise. If this continues, the rise of V_{OL} makes the noise margin smaller, and this results in the occurrence of false logic due to the slightest noise.

In designing the system, very often an output is connected to various kinds of inputs. See Examples A-1 and A-2.

Example A-1

Find how many unit loads (UL) can be driven by the output of the LS logic family.

Solution:

The unit load is defined as I_{IL} = 1.6 mA and I_{IH} = 40 µA. Table A-1 shows I_{OL} = 8 mA and I_{OH} = 400 µA for the LS family. Therefore, we have

fan-out (low) = I_{OL}/I_{IL} = 8 mA / 1.6 mA = 5

fan-out (high) = I_{OH}/I_{IH} = 400 µA / 40 µA = 10

This means that the fan-out is 5. In other words, the LS output must not be connected to more than 5 inputs with unit load characteristics.

Example A-2

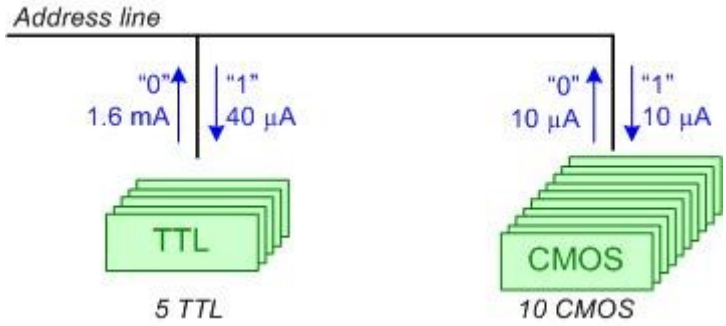
An address pin needs to drive 5 standard TTL loads in addition to 10 CMOS inputs of DRAM chips. Calculate the minimum current to drive these inputs for both logic "0" and "1".

Solution:

The standard load for TTL is I_{IH} = 40 µA and I_{IL} = 1.6 mA, and for CMOS, $I_{IL} = I_{IH}$ = 10 µA.

minimum current for "0" = total of all I_{IL} = 5 × 1.6 mA + 10 × 10 µA = 8.1 mA

minimum current for "1" = total of all I_{IH} = 5 × 40 µA + 10 × 10 µA = 300 µA



The total I_{IL} and I_{IH} requirement of all the loads on a given output must be less than the driver's maximum I_{OL} and I_{OH}. This is shown in Example A-3.

Example A-3

Assume that the microprocessor address pin in Example A-2 has specifications I_{OH} = 400 µA and I_{OL} = 2 mA. Do the input and output current needs match?

Solution:

For a high output state, there is no problem since $I_{OH} > I_{IH}$. However, the number of inputs exceeds the limit for I_{OL} since an I_{IL} of 8.1 mA is much larger than the maximum I_{OL} allowed by the microprocessor.

In cases such as Example A-3 where the receiver current requirements exceed the drivers' capability, we must use a buffer (booster), such as the 74xx245 and 74xx244. The 74xx245 is used for bidirectional and the 74xx244 for unidirectional signals. See current 74LS244 and 74LS245 characteristics in Table A-5.

Buffer	I_{OH} (mA)	I_{OL} (mA)	I_{IH} (µA)	I_{IL} (mA)
74LS244	3	12	20	0.2
74LS254	3	12	20	0.2
Note: V_{OL} = 0.4 V and V_{OH} = 2.4V are assumed.				

Table A-5: Electrical Specifications for Buffers

Capacitance derating

Next we study what is called capacitance derating and its impact in system design. A pin of an IC has an input capacitance of 5 to 7 pF. This means that a single output that drives many inputs sees a large capacitance load since the inputs are in parallel and therefore added together. Look at the following equations.

$$Q = CT \qquad \text{(A-1)}$$

$$Q / T = CV / T \qquad \text{(A-2)}$$

$$F = 1 / T \qquad \text{(A-3)}$$

$$I = CVF \qquad \text{(A-4)}$$

In Equation (A-4), I is the driving capability of the output pin, C is C_{IN} as seen by the output, and V is the voltage. The equation indicates that as the number of C_{IN} loads goes up, there must be a corresponding increase in I_O, the driving capability of the output. In other words, outputs with high values of I_{OL} and I_{OH} are desirable. Although there have been some logic families with I_{OL} = 64 mA and I_{OH} = 15 mA, their power consumption is high. Equation (A-4) indicates that if I = constant, as C goes up, F must come down, resulting in lower speed. The most widely accepted solution is the use of a large number of

drivers to reduce the load capacitance seen by a given output. Assume that we have a single address bus line driving 16 banks of 32-bit-wide memory. Each bank has 4 chips of 64K × 8 organization, which results in 16 × 4 = 64 memory chips, or 16 × 64K × 32 = 32M bytes of SRAM. Depending on how many 244s are used to drive the memory addresses, the delay due to the address path varies substantially. To understand this we examine three cases.

Case 1: Two 244 drivers

This option uses two 244 drivers, one for A0–A7 and one for A8–A15. An output of the 244 drives 16 banks of memory, each with 4 inputs. Assuming that each memory input has 5 pF capacitance, this results in a total of 4 ×16 × 5 = 320 pF capacitance load seen by the 244 output. However, the 244 output can handle no more than 50 pF. As a result, the delay due to this extra capacitance must be added to the address path delay. For each 50 to 100 pF of capacitance, an extra 3 ns delay is added to the address path delay. In our calculation, we use 3 ns for each 100 pF of capacitance. Figure A-6 shows driving memory inputs by two 244 chips. See Example A-4.

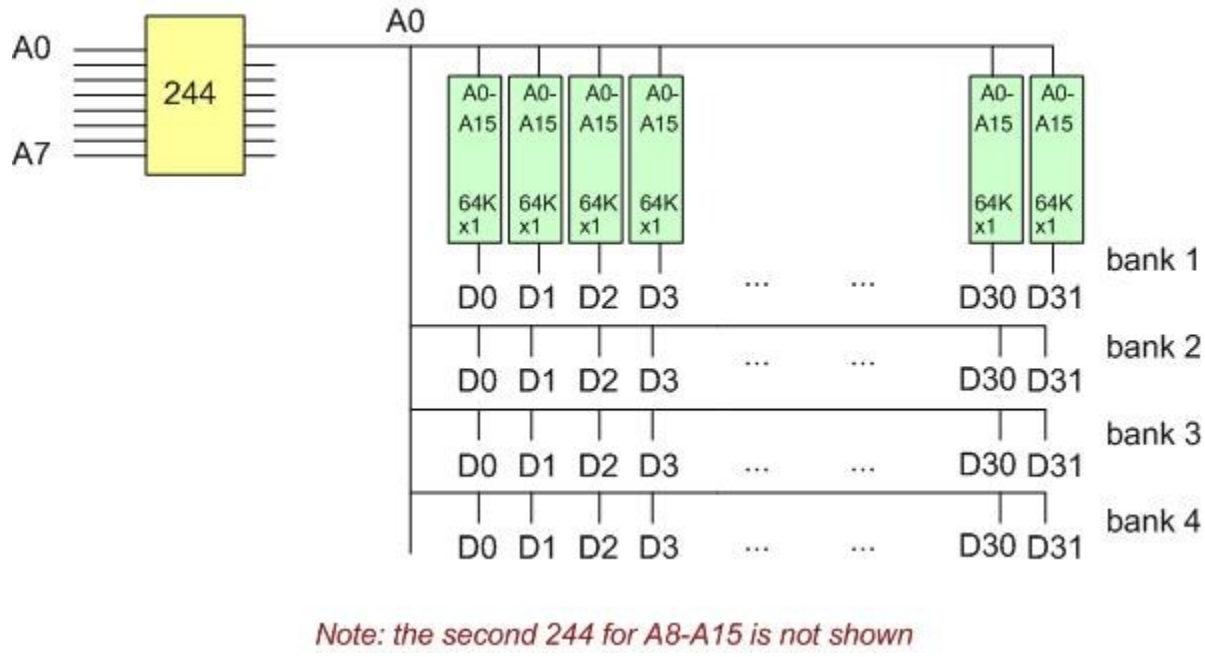


Figure A-6: Case 1, Two 244 Address Drivers

Example A-4

Calculate the following for Figure A-6, assuming a memory access time of 25 ns and a propagation delay of 10 ns for the 244.

(a) delay due to capacitance derating on the address path

(b) the total address path delay for case 1

Solution:

(a) Of the 320 pF capacitance seen by the 244, only 50 pF is taken care of; the rest, which is 270 (320 – 50 = 270), causes a delay. Since there are 3 ns for each extra 100 pF, we have the following delay due to capacitance derating, (270/100) × 3 ns = 8.1 ns.
(b) Address path delay = 244 buffer propagation delay + capacitance derating delay + memory access time = 10 ns + 8.1 ns + 25 ns = 43.1 ns.

Case 2: Doubling the number of 244 buffers

Doubling the number of 244 buffers will reduce the address path delay. A single 244 drives only 8 banks, or a total of 32 inputs, since there are 4 inputs in each bank. As a result, a 244 output will see a capacitance load of 32 × 5 = 160 pF. In this case, we use only four 244 buffer chips, as shown in Figure A-7 and Example A-5.

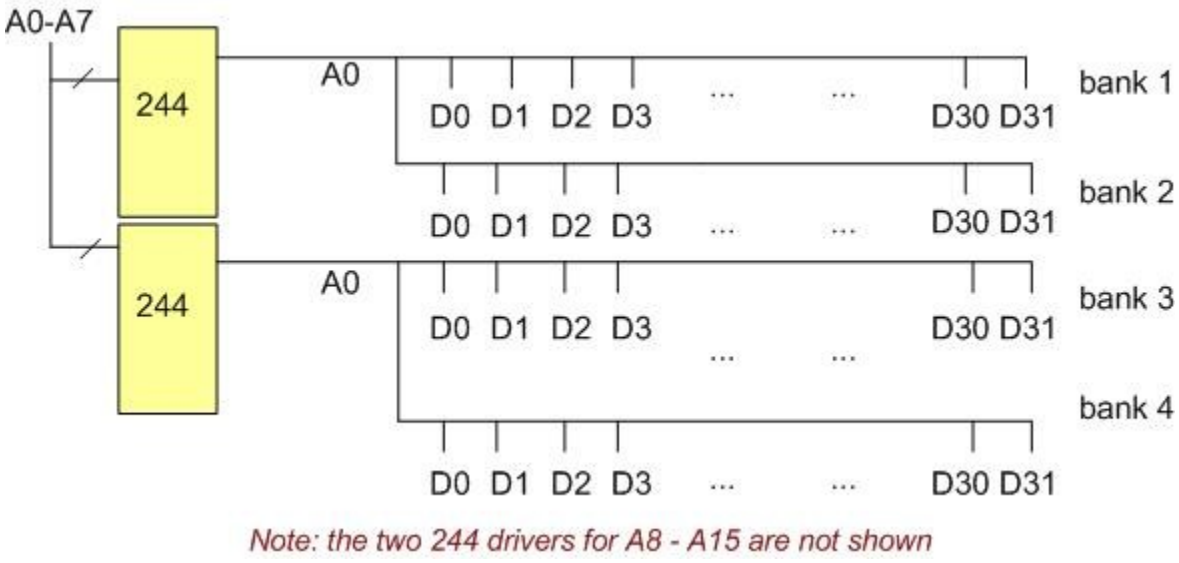


Figure A-7: Case 2, Four 244 Address Drivers

Example A-5

Calculate (a) delay due to capacitance derating on the address path, and (b) total address path delay for case 2. Assume a memory access time of 25 ns and a propagation delay of 10 ns for the 244.

Solution:

(a) Of the 160 pF capacitance seen by the 244, only 50 pF is taken care of; the rest, which is 110 pF, causes a delay. Since there are 3 ns for each extra 100 pF, we have (110/100) × 3 ns = 3.1 ns delay due to capacitance derating.

(b) The address path delay = 244 buffer propagation delay + capacitance derating delay + memory access time = 10 ns + 3.1 ns + 25 ns = 28.1 ns.

Case 3: Doubling again

In this case, we double the number of 244 buffers again, so that an output of the 244 drives four banks, each with 4 inputs. This results in a total capacitance load of 4 × 4 × 5 = 80 pF. Only 50 pF of it is taken care of by the 244, leaving 30 pF, causing a delay. See Figure A-8.

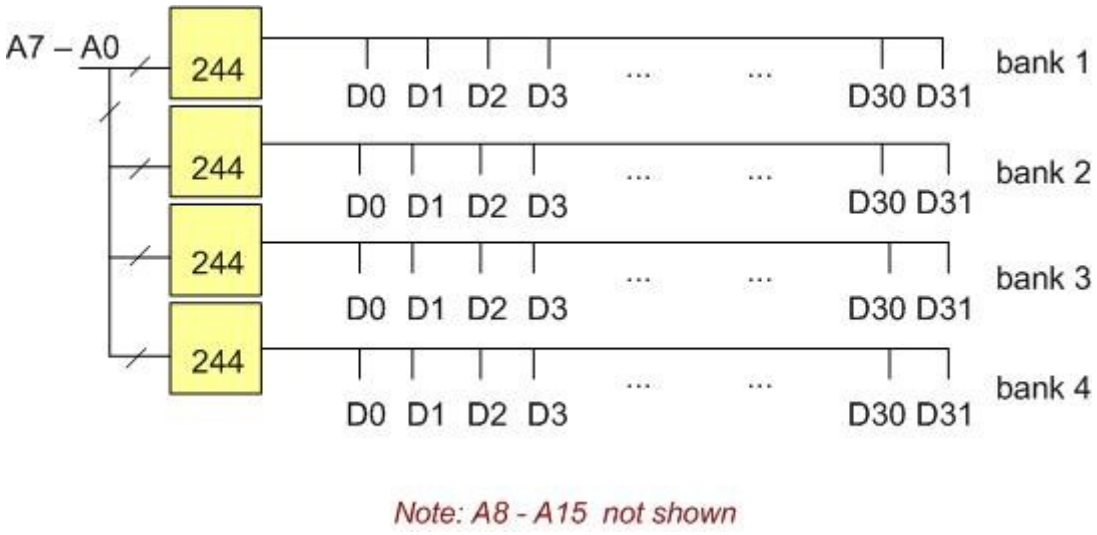


Figure A-8: Case 3, A Single 244 Address Driver for Each Bank

Examining cases 1 through 3 shows that for high-speed system design we must accept a higher cost due to extra parts and higher power consumption.

Power dissipation considerations

Power dissipation of a system is a major concern of system designers, especially for laptop and hand-held systems. Although power dissipation is a function of the total current consumption of all components of a system, the impact of V_{CC} is much more pronounced, as shown next. Earlier we showed in Equation (26-4) that I = CFV. Substituting this in equation P = VI yields the following:

$$F = VI = CFV^2 \qquad \text{(A-5)}$$

In Equation (A-5), the effects of frequency and VCC voltage should be noted. While the power dissipation goes up linearly with frequency, the impact of the power supply voltage is much more pronounced (squared). See Example A-6.

Example A-6

Prove that a 3.3 V system consumes 56% less power than a system with a 5 V power supply.

Solution:

Since P = VI, by substituting I = V/R, we have $P = V^2/R$. Assuming that R = 1, we have $P (3.3)^2 = 10.89$ W and $P = (5)^2 = 25$ W. This results in using 14.11 W less (25 – 10.89 = 14.11), which means a 56% power saving (14.11 W/25 W × 100 = 56%).

Dynamic and static currents

There are two major types of currents flowing through an IC: dynamic and static. A dynamic current is a function of the frequency under which the component is working, as seen in Equation (A-4). This means that as the frequency goes up, the dynamic current and power dissipation go up. The static current, also called dc, is the current consumption of the component when it is inactive (not selected).

Power-down option

The popularity of laptops and tablets have led microprocessor designers to make an all-out effort to conserve battery power. Today processors have what is called *system management mode (SMM)*, which reduces energy consumption by turning off peripherals or the entire system when not in use. The SMM can put the entire system, including the monitor, into sleep mode during periods of inactivity, thereby reducing "power from 250 watts to less than 30 watts." The effects on the 3.3 V power supply alone translate into a power savings of up to 56% over systems with a 5 V power supply, as was shown in Example A-6.

Ground bounce

One of the major issues that designers of high-frequency systems must grapple with is *ground bounce*. Before we define ground bounce, we will discuss lead inductance of IC pins. There is a certain amount of capacitance, resistance, and inductance associated with each pin of the IC. The size of these elements varies depending on many factors such as length, area, and so on. Figure A-9 shows the lead inductance and capacitance of the 24 pins of a DIP IC.

Pin	Self-inductance	Capacitance
1	15.10 nH	1.86 pF
2	12.20 nH	1.70 pF
3	9.54 nH	1.29 pF
4	7.44 nH	0.95 pF
5	5.31 nH	0.61 pF
6	3.73 nH	0.43 pF
7	3.41 nH	0.43 pF
8	4.66 nH	0.61 pF
9	6.95 nH	0.95 pF
10	8.96 nH	1.29 pF
11	11.70 nH	1.70 pF
12	14.50 nH	1.86 pF
13	14.50 nH	1.86 pF
14	11.70 nH	1.70 pF
15	8.96 nH	1.29 pF
16	6.95 nH	0.95 pF
17	4.66 nH	0.61 pF
18	3.41 nH	0.43 pF
19	3.73 nH	0.43 pF
20	5.31 nH	0.61 pF
21	7.44 nH	0.95 pF
22	9.54 nH	1.29 pF
23	12.20 nH	1.70 pF
24	15.10 nH	1.86 pF

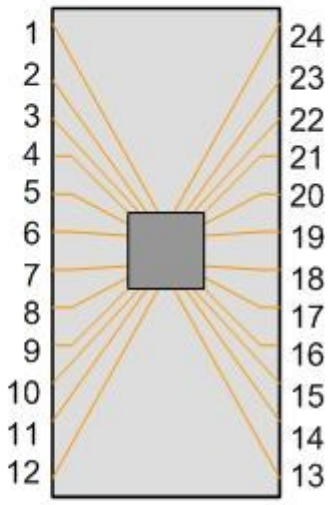


Figure A-9: Inductance and Capacitance of 24-pin DIP

The inductance of the pins is commonly referred to as *self-inductance* since there is also what is called *mutual inductance*, as we will show below. Of the three components of capacitance, resistance, and inductance, self-inductance is the one that causes the most problems in high-frequency system design since it can result in ground bounce. Ground bounce is caused when a large amount of current flows through the ground pin when multiple outputs change from high to low all at the same time. The voltage relation to the inductance of the ground lead follows:

$$V = L \, di / dt \qquad \text{(A-6)}$$

As we increase the system frequency, the rate of dynamic current, di/dt, is also increased, resulting in an increase in the inductance voltage L (di/dt) of the ground pin. Since the low state (ground) has a small noise margin, any extra voltage due to the inductance voltage can cause a false signal. To reduce the effect of ground bounce, the following steps must be taken where possible.

1. The V_{CC} and ground pins of the chip must be located in the middle rather than at the opposite ends of the IC chip (the 14-pin TTL logic IC uses pins 14 and 7 for ground and VCC). This is exactly what we see in high-performance logic gates such as Texas Instrument's advanced logic AC11000 and ACT11000 families. For example, the ACT11013 is a 14-pin DIP chip where pins 4 and 11 are used for the ground and VCC instead of 7 and 14 as in the TTL. We can also use surface mount technology such as the SOIC packages instead of DIP. Surface mount devices have much small size and shorter leads. The self-inductance of the leads is shown in Table A-6.

Pins	DIP (nH)	SOIC (nH)
1, 10, 11, 20	13.7	4.2
2, 9, 12, 19	11.1	3.8
3, 8, 13, 18	8.6	3.3
4, 7, 14, 17	6.0	2.9
5, 6, 15, 16	3.4	2.4
Courtesy of Texas Instruments		

Table A-6: 20-Pin DIP and SOIC Lead Inductance

2. Use logics with a minimum number of outputs. For example, a 4-output is preferable to an 8-output. This explains why many designers of high-performance systems avoid using memory chips or the drivers and buffers of 16- or 32-bit-wide outputs since all the outputs switching at the same time will cause a massive flow of current in the ground pin, and hence cause ground bounce (see Figure A-10).
3. Use as many pins for the ground and VCC as possible to reduce the lead length, since the self-inductance of a wire with length l and a cross section of B × C is:

$$L=0.002 \ln [2l / (B + C) + l / 2] \qquad \text{(A-7)}$$

As seen in Equation (A-7), the wire length, l, contributes more to self-inductance than does the cross section. This explains why all high-performance microprocessors and logic families use several pins for the V_{CC} and ground. For example, in the case of Intel's Pentium processor there are over 50 pins for the ground and another 50 pins for the V_{CC}.

The discussion of ground bounce is also applicable to V_{CC} when a large number of outputs changes from the low to high state and is referred to as V_{CC} bounce. However, the effect of V_{CC} bounce is not as severe as ground bounce since the high ("1") state has wider noise margin than the low ("0") state.

Filtering the transient currents using decoupling capacitors

In the TTL family, the change of the output from low to high can cause what is called *transient current*. In totem-pole output, when the output is low, Q4 is on and saturated, whereas Q3 is off. By changing the output from the low to high state, Q3 becomes on and Q4 becomes off. It is faster to turn a

transistor on than turn a transistor off. This means that there is a time that both transistors are on and drawing currents from the V_{CC}. The amount of current depends on the R_{ON} values of the two transistors, and that, in turn, depends on the internal parameters of the transistors. However, the net effect of this is a large amount of current in the form of a spike for the output current, as shown in Figure A-10.

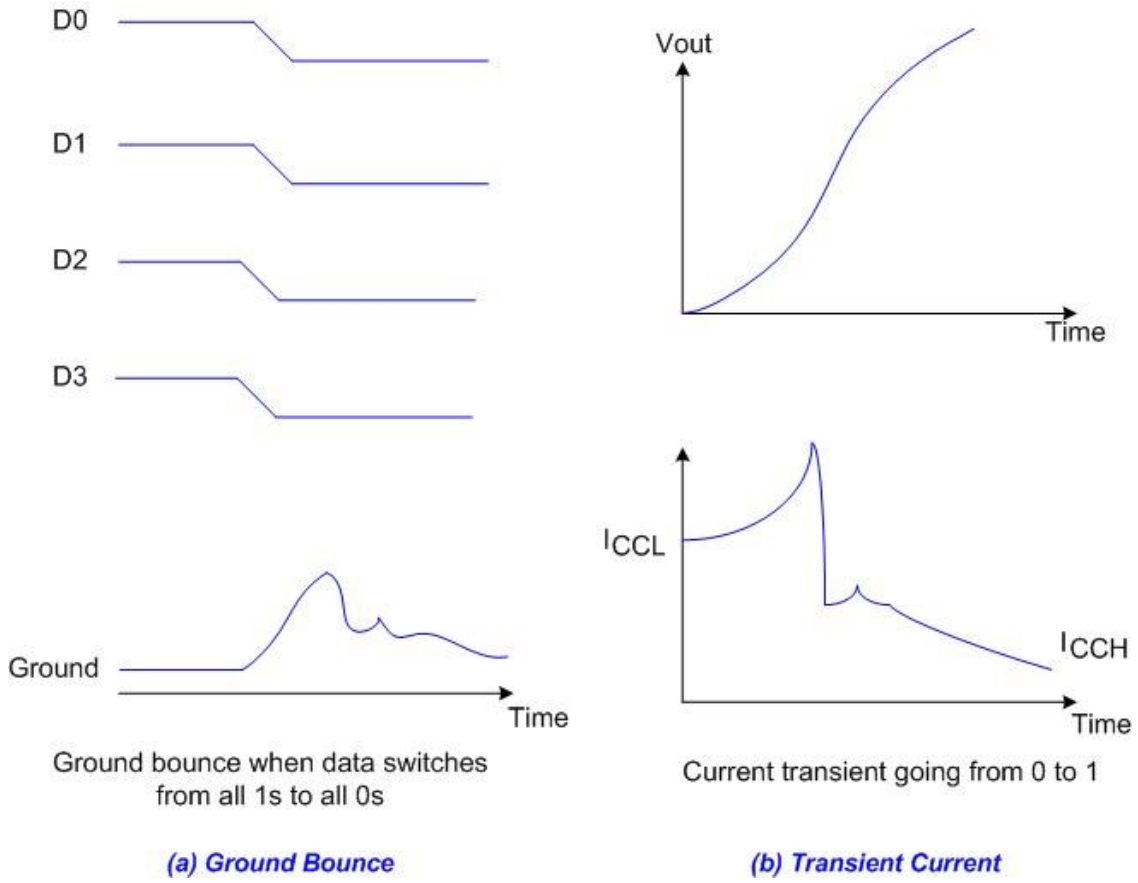


Figure A-10: (a) Ground Bounce (b) Transient Current

To filter the transient current, a 0.01 F or 0.1 F ceramic disk capacitor can be placed between the V_{CC} and ground for each TTL IC. However, the lead for this capacitor should be as small as possible since a long lead results in a large self-inductance and that results in a spike on the VCC line [V = L (di/dt)]. This is also called V_{CC} bounce. The ceramic capacitor for each IC is referred to as a decoupling capacitor. There is also a bulk decoupling capacitor, as described next.

Bulk decoupling capacitor

As many IC chips change state at the same time, the combined currents drawn from the board's V_{CC} power supply can be massive and cause a fluctuation of V_{CC} on the board where all the ICs are mounted. To eliminate this, a relatively large (relative to an IC decoupling capacitor) tantalum capacitor is placed between the V_{CC} and ground lines. The size and location of this tantalum capacitor vary depending on the number of ICs on the board and the amount of current drawn by each IC, but it is common to have a single 22 µF to 47 µF capacitor for each of the 16 devices, placed between the V_{CC} and ground lines. See Technical Notes TN0006 and TN4602 from Micron Technology.

Crosstalk

Crosstalk is due to mutual inductance. See Figure A-11.

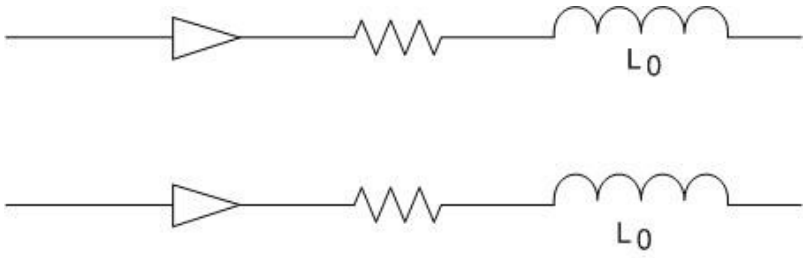


Figure A-11: Crosstalk (EMI)

Previously, we discussed self-inductance, which is inherent in a piece of conductor. *Mutual inductance* is caused by two electric lines running parallel to each other. It is calculated as follows:

$$M = 0.002l \times \ln (2l / d) - \ln (K - 1 + d/l - d/2l)^2 \quad \text{(A-8)}$$

where *l* is the length of two conductors running in parallel, and d is the distance between them, and the medium material placed in between affects K. Equation (A-8) indicates that the effect of crosstalk can be reduced by increasing the distance between the parallel or adjacent lines (in printed circuit boards, these will be traces). In many cases, such as printer and disk drive cables, there is a dedicated ground for each signal. Placing ground lines (traces) between signal lines reduces the effect of crosstalk. This method is used even in some ACT logic families where V_{CC} and GND pins are next to each other. Crosstalk is also called EMI (electromagnetic interference). This is in contrast to ESI (electrostatic interference), which is caused by capacitive coupling between two adjacent conductors.

Transmission line ringing

The square wave used in digital circuits is in reality made of a single fundamental pulse and many hArmonics of various amplitudes. When this signal travels on the line, not all the hArmonics respond the same way to the capacitance, inductance, and resistance of the line. This causes what is called *ringing*, which depends on the thickness and the length of the line driver, among other factors. To reduce the effect of ringing, the line drivers are terminated by putting a resistor at the end of the line. See Figure A-12.

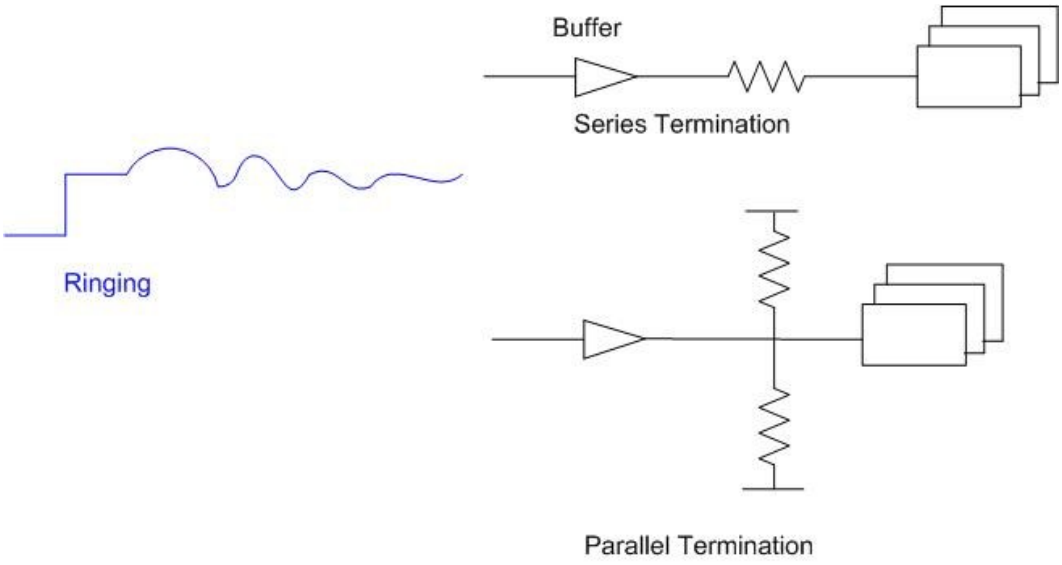


Figure A-12: Reducing Transmission Line Ringing

There are three major methods of line driver termination: parallel, serial, and Thevenin. In many systems resistors of 30–50 ohms are used to terminate the line. The parallel and Thevenin methods are used in cases where there is a need to match the impedance of the line with the load impedance. This requires a detailed analysis of the signal traces and load impedance, which is beyond the scope of this book. In high-frequency systems, wire traces on the printed circuit board (PCB) behave like transmission lines, causing ringing. The severity of this ringing depends on the speed and the logic family used. Table A-7 provides the length of the traces, beyond which the traces must be looked at as transmission lines.

Logic Family	Line Length (in.)
LS	25
S, AS	11
F, ACT	8
AS, ECL	6
FCT, FCTA	5
(Reprinted by permission of Integrated Device Technology, copyright IDT 1991)	

Table A-7: Line Length Beyond Which Traces Behave Like Transmission Lines

FIT and failure analysis

Chip manufacturers provide a parameter called *FIT (failure in time)* to measure the reliability for a single chip. The FIT of a single chip is the number of expected failures in a billion (10^9) hours of operation. If a chip has FIT of 300, then there will be 300 failures per billion device hours of operation. To reduce the

number of device failures, manufacturers use burn-in to eliminate the early failures before the product is shipped to the customer. This is commonly referred to as infant mortality since the failure rate starts high and eventually levels off to a constant level. See Figure A-13.

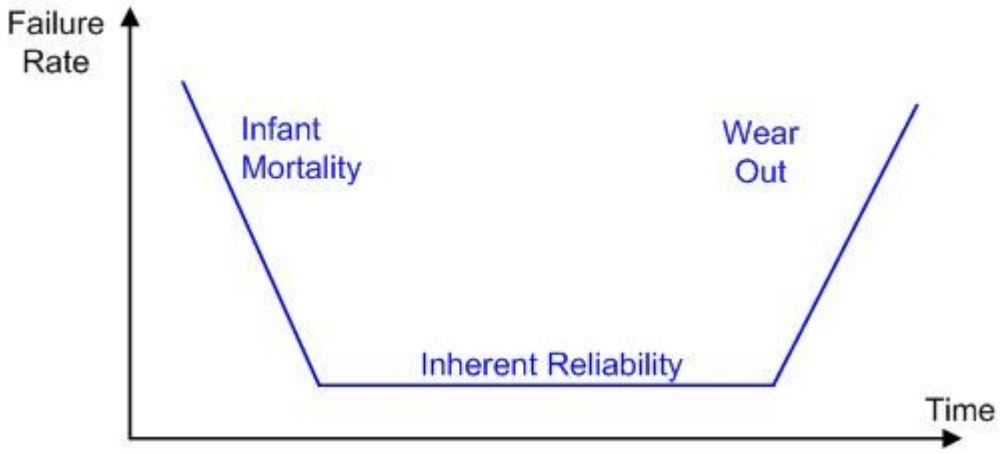


Figure A-13: Bathtub Failure Rate

Although we can eliminate the early failures using burn-in, we can never reduce the failure rate to zero due to wear out and other factors such as soft error. This is discussed next.

Soft error and hard error

In memory there are two kinds of errors that can cause a bit to change: soft error and hard error. If the cell bit gets stuck permanently in a "high" or "low" state, this is referred to as a *hard error*. Hard error is due to deterioration of the cell caused by wear-out (see Figure A-13). There is no remedy for hard error except to replace the defective RAM chip since the damage is permanent. The other kind of error, a *soft error*, alters the cell bit from 1 to 0 or from 0 to 1, even though the cell is perfectly fine (no hard error). Soft error is caused by alpha particle radiation and power surges. The sources of the alpha particles are the radiation in the air or the materials in the plastic package enclosing the RAM die. The occurrence of a soft error as a result of alpha particles ionizing the charges in a RAM cell is a greater source of concern since it is 5 times more likely to happen than a hard error. As the density of RAM chips increases and the size of the RAM cell goes down, the probability of a soft error for a given cell goes up, but the relation is not linear.

Mean time between failures (MTBF) for system

Reliability of system depends directly on two factors: a) the FIT (failures in time) value of a single part, and b) the number of parts in the system. We use these two factors to calculate what is called *MTBF (mean time between failures)*. The MTBF predicts the average time between the two consecutive failures. The MTBF for a single chip is calculated using the FIT as follows:

$$\text{MTBF} = 1{,}000{,}000{,}000 \text{ hours} / \text{FIT} \qquad \text{(A-9)}$$

To get the MTBF rate for the system, we must divide the single-chip MTBF by the number of chips in the system.

MTBF of system = MTBF of one chip / number of chips in system (A-10)

See Examples A-7 and A-8.

Example A-7

Assuming that the FIT for a single chip is 252, calculate the MTBF for:
(a) a single chip
(b) a system with 512 chips

Solution:

(a) The MTBF for a single chip is as follows: MTBF for 1 chip = 1,000,000,000 hr / 252 = 3,968,254 hr = 453 years
(b) The MTBF for 512 chips is = 453 years / 512 chips = 0.884 year = 323 days

Example A-8

Calculate the system MTBF for the system in Example A-7 if FIT = 745.

Solution:

MTBF for a single chip = 109 / 745 hrs. = 153 years. For the system it is 153 years / 512 = 109 days.

See Technical Notes TN-00-14 and TN-00-18 on the http://www.micron.com website.

http://www.micron.com/products/support/technical-notes

There is a paper called "Testing RAM for Embedded Systems" by Jack Ganssle and available from the following website:

http://www.ganssle.com/testingram.pdf

Also see the article "Thirteen feet of concrete won't shield your RAM from the perils of cosmic rays. What's the solution?" by Jack Ganssle in Dr. Dobb's Journal. It is available from the following website:

http://www.ddj.com/dept/debug/196800160

ECL and gallium arsenide (GaAs) chips

The use of L3 cache and EDC (Error Detection and Correction) in systems with speeds of 200 MHz and higher is adding to the data and address path delay. This is forcing designers to resort to using ECL and GaAs chips. Due to the fact that ECL chips have a very high power dissipation, they are not used in

low-cost x86 design. However, GaAs chips are showing up in high-speed x86 and RISC-based computers. This is especially the case for the GaAs EDC and cache controller chips. The mass of electrons in GaAs is lighter than in silicon, due to its quantum mechanics structure. As a result, the electrons in GaAs have a much higher speed. This means that GaAs chips can achieve a much higher speed than silicon. The power dissipation of the GaAs transistor is comparable to the silicon-based MOS transistor. Therefore, GaAs technology might appear to provide the ideal chip since it has the speed of ECL (it is even faster than ECL) and the power dissipation of CMOS. However, it has the following disadvantages.

1. Unlike silicon, of which there is a plentiful supply in nature in the form of sand, GaAs is a rare commodity, and therefore more expensive.
2. GaAs is a compound made of two elements, Ga and As, and therefore is unstable at high temperatures.
3. It is very brittle, making it impossible to have large wafers. As a consequence, at this time no more than 100,000 transistors can be placed on a single chip. Contrast this to the millions of transistors for silicon-based chips.
4. The GaAs yields are much lower than for silicon, making the cost per chip much more expensive than for silicon chips.

These problems make the building of an entire computer based on GaAs a visionary project, if not an impossible one. This was the case for the CRAY III supercomputer, which was based on GaAs, and the buses ran at speeds of multiple GHz; but the project was also several years behind and millions of dollars over budget, so it was eventually abandoned and the company went out of business.

Review Questions

1. What is the fan-out of the "0" state?
2. If the fan-out of "low" and "high" are 10 and 15, respectively, what is the fan-out?
3. If I_{OL} = 12 mA, I_{OH} = 3 mA for the driver, and IIL = 1.6 mA, IIH = 40 A for the load, find the fan-out.
4. Why do I_{IL} and I_{OH} have negative signs in many TTL books?
5. What are the 74xx244 and 74xx245 used for?
6. What is capacitive derating?
7. Ground bounce happens when the output makes a transition from _____ to _____.
8. Give one way to reduce ground bounce.
9. Transient current is due to transition of output from _____ to _____.
10. Why do high-speed logic gates using DIP packaging put the VCC and ground pins in the middle instead of the corners?
11. True or false. Soft error is permanent.
12. True or false. Hard error is permanent.
13. Alpha particle radiation causes _____ (soft, hard) errors.
14. FIT is in _____ (hours, months, years) of device operation.
15. What is the MTBF for 512 megabytes of memory if DRAM chips used are 16M × 8 with FIT = 252?
16. What is the MTBF for 512 megabytes of memory if DRAM chips used are 16M × 8 with FIT = 1000?

Answers to Review Questions

Section A.1

1. MOS is more power efficient, while bipolar is faster.
2. True
3. True
4. CMOS
5. NMOS
6. In the lower end, ALS, and in the higher end, FAST

Section A.2

1. It is the number of loads that the driver can support and it is calculated by I_{OL}/I_{IL}.
2. 10
3. I_{OL}/I_{IL} = 12 mA/1.6 mA = 7 and I_{OH}/I_{IH} = 3 mA/40 µA = 75. Fan-out is 7, a lower number.
4. The negative sign indicates that these currents are flowing out of the IC (conventional current flow).
5. They are used for the line driver: the 74xx244 for unidirectional and 74xx245 for bidirectional lines.
6. It is signal delay caused by excessive load capacitance.
7. High, low
8. Make the ground pin length as small and short as possible.
9. Low, high
10. To make the self-inductance of pins VCC and GND small in order to reduce the ground and VCC bounce
11. False
12. True
13. Soft
14. Hours
15. 453/32 = 14.1 years since we have 512M × 8/16M × 8 = 32 chips
16. 3.56 years (114.15 years for one DRAM divided by 32 chips)