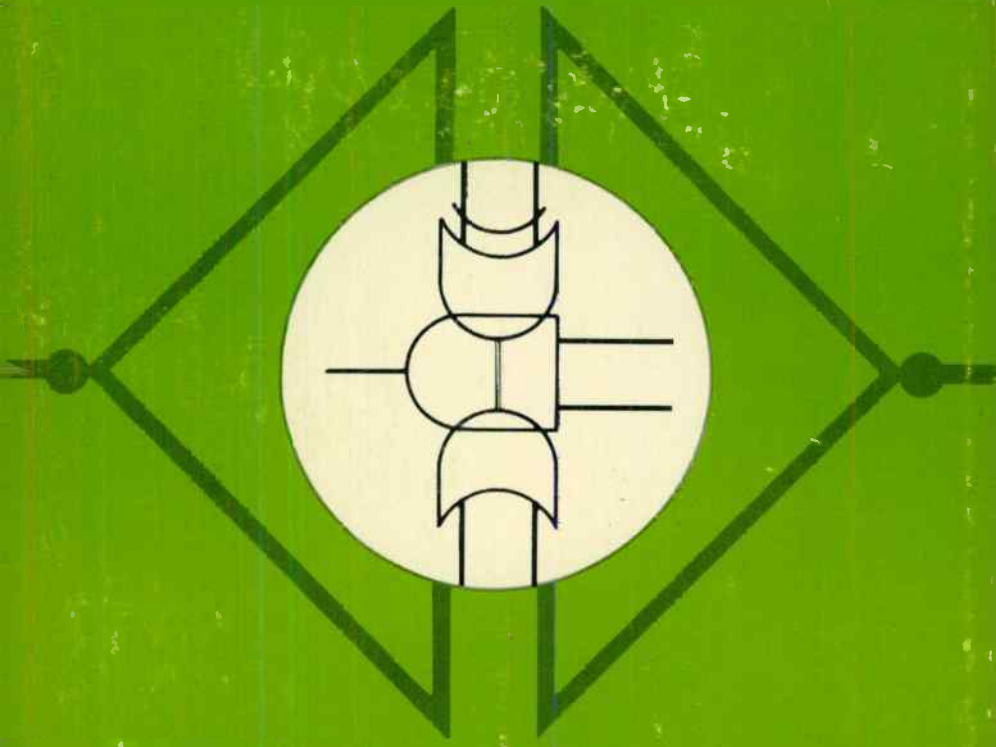
Howard W. Sams PUBLICATION 21129

Understanding CMOS INTEGRATED CIRCUITS

by
Roger Melen and Harry Garland



Understanding CMOS Integrated Circuits

by Roger Melen and Harry Garland

HOWARD W. SAMS & CO., INC.
THE BOBBS-MERRILL CO., INC.
INDIANAPOLIS · KANSAS CITY · NEW YORK

FIRST EDITION

FIRST PRINTING—1975

International Standard Book Number: 0-672-21129-7
Library of Congress Catalog Card Number: 74-33833

Preface

CMOS integrated circuits are amazingly versatile devices that can be used by anyone who designs or builds electronic circuits. In this book we describe what CMOS ICs are, how they work, and how they can be used in electronic circuit design. Many practical circuits, complete with parts values, are discussed in detail. These include a digital timer, a capacitance meter, an analog-to-digital converter, a frequency synthesizer, and an electronic wristwatch, to name a few.

Following an introduction to the use of CMOS ICs in Chapter 1, three chapters are devoted to the internal operation of CMOS ICs. Chapter 2 discusses the integrated components that make up CMOS ICs, and Chapter 3 discusses how these components are fabricated. Chapter 4 discusses integrated CMOS circuitry, while Chapter 5 develops a set of guidelines for using CMOS ICs. Over two dozen practical circuits using CMOS ICs are presented in Chapters 6 through 8.

The assistance of a number of people was indispensable in the preparation of this book. In particular we thank Bill Russell, Cathy Lincoln, Jane Fajardo, and Janis Birdwell. In the preface to a previous Sams book, *Understanding IC Operational Amplifiers,* we had the pleasure of thanking Margaret Hogan for her assistance. Margaret Hogan is now Margaret Garland, and we thank her once again.

ROGER MELEN
HARRY GARLAND

Contents

CHAPTER 5

CHAPTER 6

CHAPTER 7

CHAPTER 8

APPENDIX

1

Introduction to CMOS

CMOS digital integrated circuits are being used with increasing frequency in electronic circuit design, and with good reason. CMOS ICs are extremely versatile, easy to use, and have performance features not offered by any other family of digital integrated circuits.

CMOS ICs are so called because they are built with complementary *MOS* (metal-oxide-semiconductor) transistors. Being built in this way, CMOS ICs usually consume less power than other types of digital ICs and are able to operate over a much wider voltage range. Electronic wristwatches, automobile electronic systems, medical electronic instruments, television sets, and portable calculators are all examples of products that now use CMOS ICs.

In this chapter we will review a few of the building blocks that are basic to all digital IC families. These include gates, flip-flops, counters, and decoders. To give a design example, we will show how these standard components can be interconnected to make a CMOS digital timer.

LOGIC GATES

There are only two different proper signals that can be applied to any input of a digital IC. Likewise, there are only two different proper signals that can emerge at any output of a digital IC. One of these two signals is called "logic 1." The other is called "logic 0." For CMOS digital ICs a signal voltage greater than 70 percent of the IC power-supply voltage is logic 1. A signal voltage

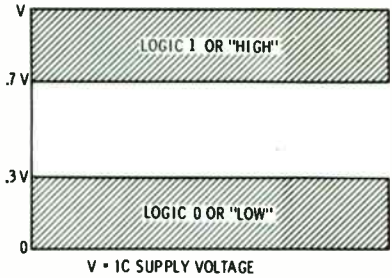


Fig. 1-1. The two logic levels for CMOS ICs.

less than 30 percent of the power-supply voltage is logic 0. A logic 1 signal is said to be "high" and a logic 0 signal "low." Fig. 1-1 illustrates the high and low logic levels for CMOS ICs.

The most basic circuit elements used to manipulate logic signals are the *logic gates*. Every logic gate has one or more inputs and one output. The logic signal that appears at the output of a gate is determined by the logic signals applied to the inputs of the gate. The relationship between the inputs and the output of any gate can be summarized in a table called a *truth table*. The basic logic gates are shown with their truth tables in Fig. 1-2.

The Inverter

The simplest logic gate is the *inverter*. The logic diagram of the inverter, together with its truth table, is shown in Fig. 1-2A. The truth table shows that the function of an inverter is to produce a logic 0 output when its input is logic 1, and to produce a logic 1 output when its input is logic 0. In other words, the output is low when the input is high, and high when the input is low. Since the output of an inverter is always the opposite of its input, the inverter is sometimes called a NOT gate.

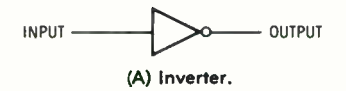
The AND Gate and the NAND Gate

The AND gate and its truth table are shown in Fig. 1-2B. The AND gate is so named because its output is high only when input A *and* input B are high. For any other combination of input signals, the output of an AND gate is low.

The output of the NAND or "NOT AND" gate is always the opposite of the output of an AND gate. In fact, the NAND gate symbol (Fig. 1-2C) is that of the AND gate symbol plus a small circle at the tip which indicates inversion of the output.

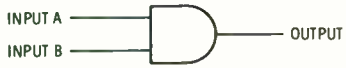
The OR Gate and the NOR Gate

The output of the OR gate is high when input A *or* input B is high. Otherwise the output is low. The OR gate is shown with its



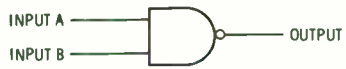
(A) Inverter.

INPUT	OUTPUT
1	0
0	1



(B) AND gate.

INPUT A	INPUT B	OUTPUT
0	0	0
1	0	0
0	1	0
1	1	1



(C) NAND gate.

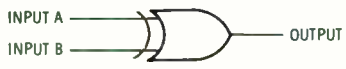
INPUT A	INPUT B	OUTPUT
0	0	1
1	0	1
0	1	1
1	1	0

(D) OR gate.

INPUT A	INPUT B	OUTPUT
0	0	0
1	0	1
0	1	1
1	1	1

(E) NOR gate.

INPUT A	INPUT B	OUTPUT
0	0	1
1	0	0
0	1	0
1	1	0



(F) Exclusive OR gate.

INPUT A	INPUT B	OUTPUT
0	0	0
1	0	1
0	1	1
1	1	0

(G) Exclusive NOR gate.

INPUT A	INPUT B	OUTPUT
0	0	1
1	0	0
0	1	0
1	1	1

Fig. 1-2. Logic gates and their truth tables.

truth table in Fig. 1-2D. The output of the NOR (NOT OR) gate, as shown in Fig. 1-2E, is always opposite that of the OR gate.

The Exclusive-OR Gate and the Exclusive-NOR Gate

The output of the exclusive-OR gate (Fig. 1-2F) is high when input A *or* input B is high but not when both inputs are high.

The output of the exclusive-NOR gate (Fig. 1-2G) is always opposite that of the exclusive-OR gate.

The NOR Gate and NAND Gate Used As Inverters

From the truth table for the NOR gate (Fig. 1-2E), we can see that if both inputs are low, the output is high, and if both inputs are high, the output is low. Thus, a NOR gate can be used as an inverter by simply connecting its inputs together. By the same reasoning, a NAND gate can also be used as an inverter by connecting its inputs together. Fig. 1-3 shows both a NAND gate and a NOR gate used as inverters.

(A) The NOR gate used as an inverter. (B) The NAND gate used as an inverter.

Fig. 1-3. Two-input gates used as inverters.

Multiple-Input Gates

Any of the gates described above (with the exception of the inverter) can have more than two inputs. The operation of multiple-input gates is completely analogous to that of two-input gates. The output of a multiple-input NAND gate, for example, is low only if all its inputs are high. Fig. 1-4A shows the diagram of a four-input NAND gate. The truth table of this gate is shown in Fig. 1-4B. It is interesting to note that any multiple-input gate can be built from two-input gates. Fig. 1-4C shows an equivalent four-input NAND gate built from three two-input gates.

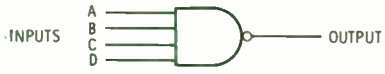
Logic-Gate Power Supply

Although the logic diagrams for logic gates do not normally show power-supply connections, all logic gates must be connected to a power supply in order to be functional. A single IC package can contain one or more logic gates, and the power-supply connection to the IC normally powers all the logic gates in the package. For most CMOS ICs, the power-supply voltage can be anywhere between 3 volts and 15 volts.

Buffers

Buffers are gates that are designed to be able to deliver more output current than standard gates. Buffers are particularly useful for driving external circuitry, such as LED (light-emitting diode) displays.

An easy way to increase the current-driving capability of CMOS gates is to connect the gates in parallel. An inverting buf-



(A) Logic diagram.

INPUTS				
A	B	C	D	OUTPUT
0	0	0	0	1
1	0	0	0	1
0	1	0	0	1
1	1	0	0	1
0	0	1	0	1
1	0	1	0	1
0	1	1	0	1
1	1	1	0	1
0	0	0	1	1
1	0	0	1	1
0	1	0	1	1
1	1	0	1	1
0	0	1	1	1
1	0	1	1	1
0	1	1	1	1
1	1	1	1	0

(B) Truth table.

(C) Four-input NAND gate built from two-input gates.

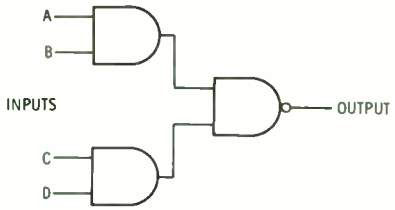


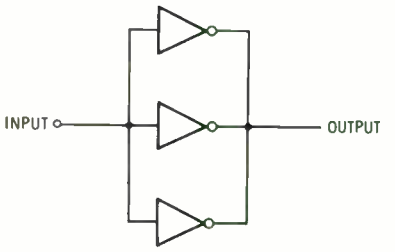
Fig. 1-4. The four-input NAND gate.

fer can be built, for example, just by connecting several inverters in parallel, as shown in Fig. 1-5.

A SQUARE-WAVE GENERATOR

In digital circuits it is often necessary to have a source of square waves. A practical square-wave generator circuit, using two CMOS inverters, is shown in Fig. 1-6. The operation of this cir-

Fig. 1-5. An inverting buffer with three times the output current of a single CMOS inverter.



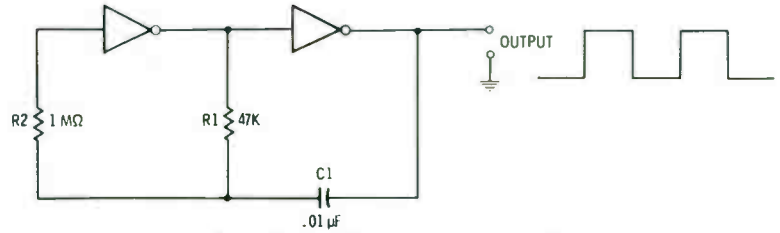


Fig. 1-6. A CMOS square-wave generator.

cuit is discussed in Chapter 6, and Chart 6-1 can be used to find the output frequency of the circuit for various values of R1 and C1. For the circuit values shown (R1 = 47K and C1 = 0.01 μF), the output frequency is approximately 1 kHz.

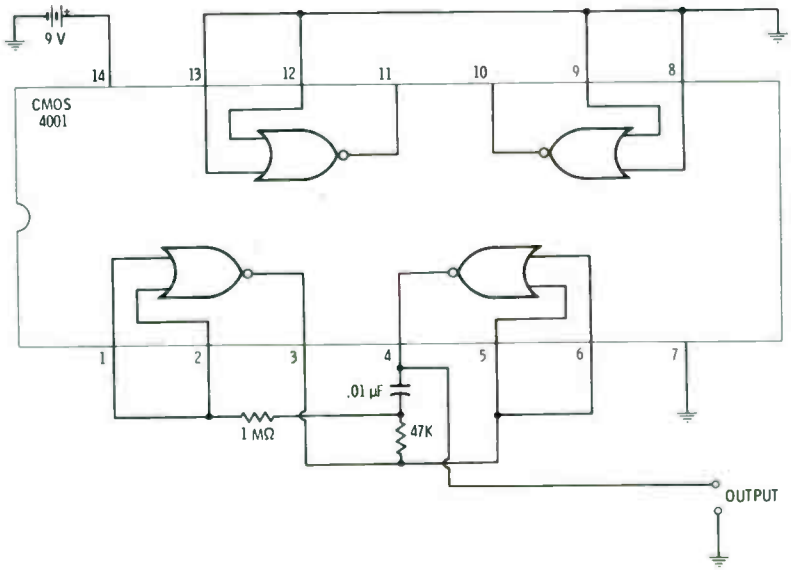


Fig. 1-7. A wiring diagram for the CMOS square-wave generator.

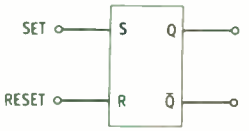
Fig. 1-7 illustrates how the square-wave generator can be built using a 4001 CMOS IC. The 4001 contains four two-input NOR gates in a 14-pin package. (This and other CMOS ICs are described in the Appendix.) Two of the NOR gates are used as inverters (as in Fig. 1-3). The other two NOR gates are not used. An important rule in using CMOS ICs is that all unused gate inputs must be connected either to the positive supply voltage or to ground. In this circuit the unused gate inputs are connected to ground. The circuit is powered by a single 9-volt battery.

FLIP-FLOPS

The flip-flop is the basic digital memory circuit. Three types of flip-flops are commonly used in digital systems: (1) the RS flip-flop, (2) the D flip-flop, and (3) the JK flip-flop.

The RS Flip-Flop

The RS flip-flop, or "reset-set" flip-flop, is shown in Fig. 1-8A. The RS flip-flop has two inputs, the set input and the reset input, and has two outputs, Q and $\overline{Q}$ (pronounced "Q bar"). Q and $\overline{Q}$ are said to be *complementary* outputs because when one output is high, the other output is low, and *vice versa.*



(A) Logic symbol.

INPUTS		OUTPUTS		
S	R	Q	$\overline{Q}$	
1	0	1	0	
0	1	0	1	
0	0	Q	$\overline{Q}$	(NO CHANGE)
1	1	—	—	(DISALLOWED STATE)

(B) Truth table.

Fig. 1-8. The RS flip-flop.

The remarkable feature of the RS flip-flop is its ability to remember, when both of its inputs are low, which of the two inputs was the last one to be high. If the set input was the last one to be high, then the Q output will be high (and the $\overline{Q}$ output will be low). If the reset input was the last one to be high, the Q output will be low (and the $\overline{Q}$ output high). If both inputs were high and simultaneously went low, the output state of the flip-flop would be unpredictable. For this reason the RS flip-flop is normally used in such a way that no more than one input can be high at any time. (Having both inputs high is said to be a *disallowed* state.) The operation of the RS flip-flop can be summarized in a truth table, as shown in Fig. 1-8B.

The RS flip-flop circuit can be built with two NOR gates, as shown in Fig. 1-9. To see how this circuit works, imagine that switch S1 is depressed. This applys a logic 1 signal to the set input and forces the Q output high and the $\overline{Q}$ output low. When S1 is released, Q will remain high, and $\overline{Q}$ will remain low. The flip-flop effectively remembers that S1 was the last switch to be depressed. Now suppose that S2 is depressed. This will force the Q output low and the $\overline{Q}$ output high. The flip-flop will remain in this state when S2 is released, and the outputs will not change unless S1 is depressed again.

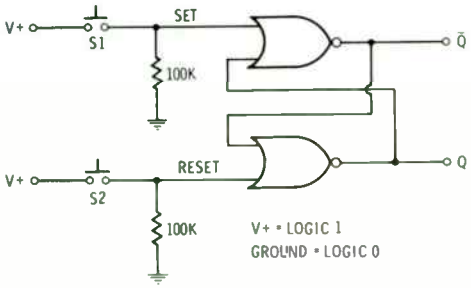


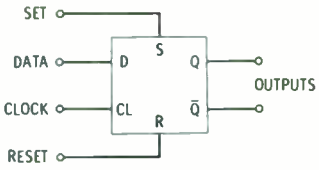
Fig. 1-9. RS flip-flop built with two NOR gates.

The D Flip-Flop

The logic diagram and truth table for the D flip-flop are shown in Fig. 1-10. The D flip-flop has four inputs: (1) the set input, (2) the reset input, (3) the data input, and (4) the clock input. The D flip-flop has two complementary outputs, Q and $\bar{Q}$.

The set and reset inputs of the D flip-flop work in precisely the same way as they do for the RS flip-flop. A logic 1 signal at the set input (with the reset input at logic 0) will force the Q output high and the $\bar{Q}$ output low. A logic 1 signal at the reset input (with the set input at logic 0) will force the Q output low and the $\bar{Q}$ output high. The D flip-flop will respond to the data input and clock input only when both the set and reset inputs are at logic 0.

With both the set and the reset inputs at logic 0, the D flip-flop has the powerful capability to sample the data input signal and to store its logic state at the Q output. The data input signal is sampled whenever the clock input signal goes from logic 0 to logic 1. Thus, whenever it is desired to sample and remember the state of the data input, all that is required is a positive transition (from



(A) Logic diagram.

INPUTS				OUTPUTS		
CL	D	S	R	Q	$\bar{Q}$	
↑	0	0	0	0	1	
↑	1	0	0	1	0	
↓	X	0	0	Q	$\bar{Q}$	(NO CHANGE)
X	X	1	0	1	0	
X	X	0	1	0	1	
X	X	1	1	—	—	(DISALLOWED)

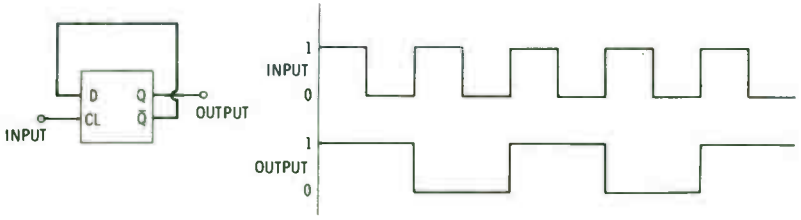
(X - DOESN'T MATTER)

(B) Truth table.

Fig. 1-10. The D flip-flop.

logic 0 to logic 1) of the clock input. This positive transition of the clock is indicated by an upward-pointing arrow in the truth table (Fig. 1-10B).

One very useful configuration of the D flip-flop is made by connecting the $\bar{Q}$ output to the data input, as shown in Fig. 1-11A. Since the $\bar{Q}$ output is always the complement of the Q output, the



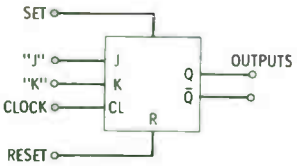
(A) Logic diagram. (B) Timing diagram.

Fig. 1-11. The D flip-flop connected as a toggle flip-flop.

flip-flop connected in this way will change output states, or "toggle," with every positive transition of the clock. The D flip-flop connected in this way is called a *toggle flip-flop*. If the clock input to this circuit is a square wave (as shown in Fig. 1-11B), the output of the circuit will be a square wave with a frequency exactly one half that of the input signal. The D flip-flop is often used in this way to divide the frequency of an input signal by a factor of two.

The JK Flip-Flop

The JK flip-flop, shown with its truth table in Fig. 1-12, is similar to the D flip-flop, but more versatile. The single D input is replaced in this device by a J input and a K input. An important feature of the JK flip-flop is that it can be used as a toggle flip-flop just by applying a logic 1 signal to both the J and the K inputs.



INPUTS					OUTPUTS		
CL	J	K	S	R	Q	$\bar{Q}$	
↑	1	1	0	0	$\bar{Q}$	Q	(TOGGLE)
↑	1	0	0	0	1	0	
↑	0	1	0	0	0	1	
↑	0	0	0	0	Q	$\bar{Q}$	(NO CHANGE)
↓	X	X	0	0	Q	$\bar{Q}$	(NO CHANGE)
X	X	X	1	0	1	0	
X	X	X	0	1	0	1	
X	X	X	1	1	—	—	(DISALLOWED)

(X - DOESN'T MATTER)

(A) Logic diagram. (B) Truth table.

Fig. 1-12. The JK flip-flop.

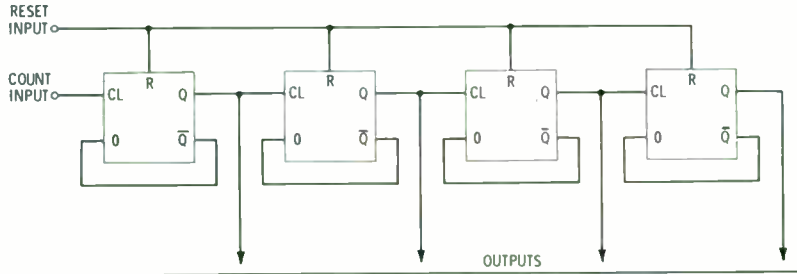
COUNTERS

The function of a digital IC counter is to keep count, at its outputs, of the number of logic pulses applied to its input. Since any output of the counter can only be at a logic 1 or a logic 0 level, an output code must be used to keep track of the input-pulse count. Different IC counters use different output codes. The two codes most commonly used in CMOS IC counters are the *binary code* and the *Johnson code*.

The Binary Counter

Fig. 1-13 shows a four-stage binary counter built with four D flip-flops. Each stage is connected as a toggle flip-flop, as in Fig. 1-11A. Prior to the beginning of a count, the outputs of the binary counter can all be reset to logic 0 by applying a logic 1 pulse to the reset input. When the reset signal returns to logic 0, the counter is ready to count.

The binary code is illustrated in the table of Fig. 1-13. After the first pulse arrives at the clock input of the counter, Q1 will be high and the other counter outputs will be low. After the second pulse



COUNT	Q1	Q2	Q3	Q4
0	0	0	0	0
1	1	0	0	0
2	0	1	0	0
3	1	1	0	0
4	0	0	1	0
5	1	0	1	0
6	0	1	1	0
7	1	1	1	0
8	0	0	0	1
9	1	0	0	1
10	0	1	0	1
11	1	1	0	1
12	0	0	1	1
13	1	0	1	1
14	0	1	1	1
15	1	1	1	1

Fig. 1-13. A four-stage binary counter and the binary code.

arrives, Q2 will be high, and the other outputs low. After the third pulse arrives, Q1 and Q2 will both be high, and the other outputs low. In this way, the binary counter continues to keep a count of the input pulses. A four-stage binary counter, as shown in Fig. 1-13, can count up to 15 input pulses. Additional binary stages can be cascaded for higher counts.

It is important to note that in the binary counter the input pulses are applied to the first flip-flop only, and that the output of each flip-flop is used to toggle the next flip-flop in the chain. Because there is a small delay in each flip-flop, the flip-flops cannot toggle at exactly the same time. When going from a count of seven to a count of eight, for example, first Q1 will change states, then Q2, Q3, and Q4 in rapid succession. Because the information "ripples" down the chain of flip-flops in this way, the binary counter is called a *ripple counter* or an *asynchronous counter*.

BCD Counters

Bcd (binary-coded decimal) counters use a four-place binary output code to code the decimal digits zero through nine. A two-digit bcd counter is shown in Fig. 1-14.

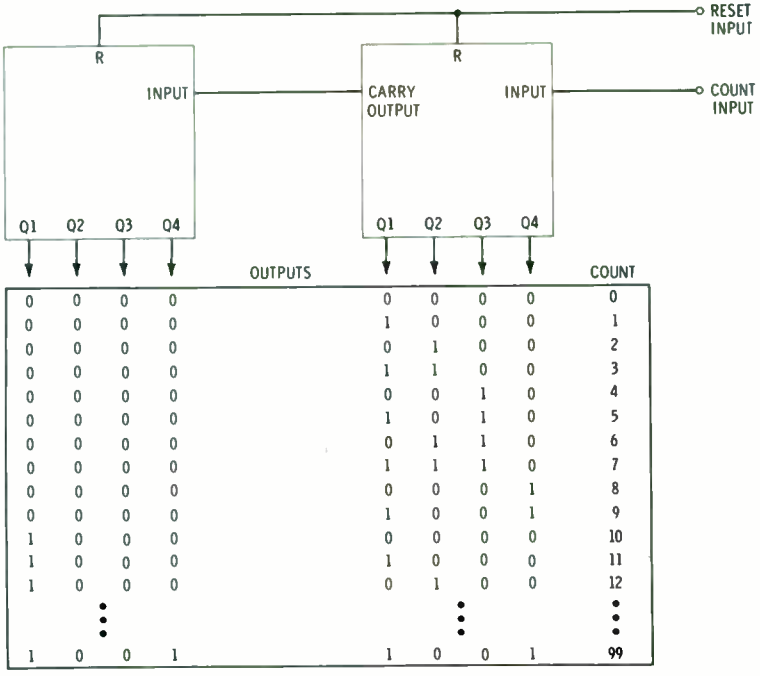
Each bcd counter is able to count up to the number nine (Q1 and Q4 high, Q2 and Q3 low). On the tenth input pulse the bcd counter generates a carry pulse in order to increment the next counter, while its own count returns to zero. Internally a bcd counter is just a four-stage binary counter with additional circuitry to reset itself on the tenth pulse and to generate a carry pulse.

Johnson Counters

The Johnson counter uses yet a different code to keep the input-pulse count. A 5-stage Johnson counter and a table of the Johnson code are shown in Fig. 1-15. An important feature of the Johnson counter is that the input signal is applied simultaneously to each of the counter flip-flops. Since information does not have to ripple down a chain of flip-flops, as it does in a binary counter, the Johnson counter is called a *synchronous counter*.

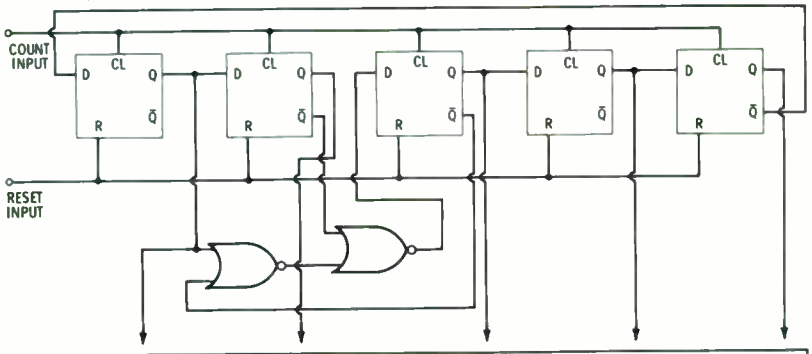
DECODERS

A *decoder* is a circuit that is able to convert a digital number, like the one that appears at the outputs of a counter, from one digital code to another. Decoders are most frequently used to convert a digital number into a form suitable for activating a numeric display. The two most common types of decoders are the one-of-n decoders and the seven-segment decoders.



0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	1
0	0	0	0	0	1	0	0	2
0	0	0	0	1	1	0	0	3
0	0	0	0	0	0	1	0	4
0	0	0	0	1	0	1	0	5
0	0	0	0	0	1	1	0	6
0	0	0	0	1	1	1	0	7
0	0	0	0	0	0	0	1	8
0	0	0	0	1	0	0	1	9
1	0	0	0	0	0	0	0	10
1	0	0	0	1	0	0	0	11
1	0	0	0	0	1	0	0	12
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
1	0	0	1	1	0	0	1	99

Fig. 1-14. A two-digit bcd counter.

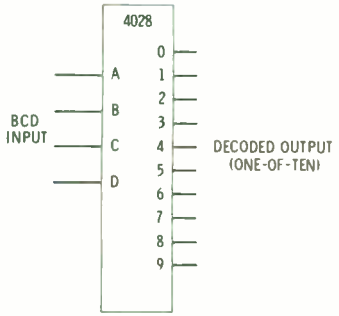


COUNT	Q1	Q2	Q3	Q4	Q5
0	0	0	0	0	0
1	1	0	0	0	0
2	1	1	0	0	0
3	1	1	1	0	0
4	1	1	1	1	0
5	1	1	1	1	1
6	0	1	1	1	1
7	0	0	1	1	1
8	0	0	0	1	1
9	0	0	0	0	1

Fig. 1-15. A five-stage Johnson counter and the Johnson code.

One-of-n Decoders

A one-of-n decoder has "n" outputs, one corresponding to each different possible combination of input signals. Each different set of input signals causes exactly one of these n outputs to go high, while the others remain low. The bcd-to-decimal decoder shown in Fig. 1-16A is an example of a one-of-ten decoder. From the truth table (Fig. 1-16B) it can be seen that each different bcd number indeed causes exactly one of the ten outputs to go high. One-of-n decoders are available both as separate ICs and as integral parts of counter ICs.



(A) Logic diagram.

INPUTS				OUTPUTS									
A	B	C	D	0	1	2	3	4	5	6	7	8	9
0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	0	0	0	0
1	1	0	0	0	0	0	1	0	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0	0	1	0	0	0	0
0	1	1	0	0	0	0	0	0	0	1	0	0	0
1	1	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	0	0	0	0	0	0	1	0
1	0	0	1	0	0	0	0	0	0	0	0	0	1

(B) Truth table.

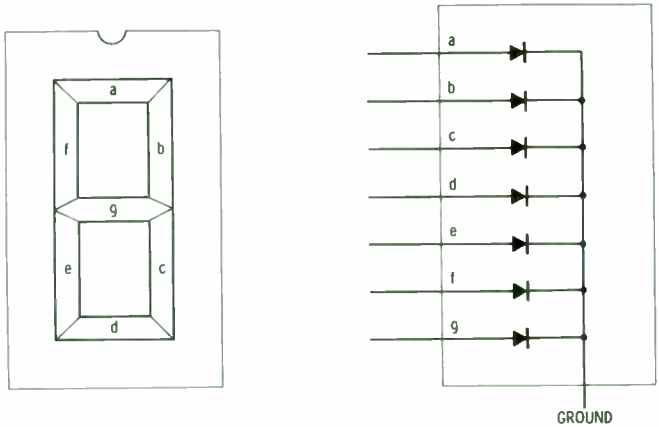
Fig. 1-16. The 4028 bcd-to-decimal decoder.

Seven-Segment Decoders

The most popular type of numeric display is made up of seven independent segments. As shown in Fig. 1-17A, each of the seven segments is assigned a letter, *a* through *g*. Any numeral, zero through nine, can be displayed by lighting the appropriate segments. The number "7" appears, for example, when segments a, b, and c are lit. The segments of a seven-segment display can be formed with light-emitting diodes, which glow when forward biased. The schematic diagram of a seven-segment LED display is shown in Fig. 1-17B.

The seven-segment decoder is used to convert a digital input number into a seven-segment code. The seven-segment code can then be used to turn on or off the individual segments of a display. The truth table for a bcd-to-seven-segment decoder is shown in Fig. 1-18.

Like the one-of-n decoders, seven-segment decoders are available either as separate ICs or as integral parts of counting ICs. The 4026 CMOS IC is an example of a 5-stage Johnson counter



(A) Each segment is designated by a letter. (B) Schematic diagram.

Fig. 1-17. A seven-segment LED display.

with integral seven-segment decoding. The logic diagram for this IC is shown in Fig. 1-19. In this diagram a small circle at the input of a logic gate means inversion of the input. Although this diagram may at first appear formidable, it is really not difficult to understand. On the upper left side of the diagram is a 5-stage Johnson counter identical to the one in Fig. 1-15. The Q and $\overline{Q}$ outputs of this counter are decoded into the seven-segment code by an array of NOR and NAND gates. Inverting buffers are placed in each out-

INPUTS				OUTPUTS							DISPLAY
A	B	C	D	a	b	c	d	e	f	g	
0		0	0	1	1	1	1	1	1	0	0
1	0	0	0	0	1	1	0	0	0	0	1
0	1	0	0	1	1	0	1	1	0	1	2
1	1	0	0	1	1	1	1	0	0	1	3
0	0	1	0	0	1	1	0	0	1	1	4
1	0	1	0	1	0	1	1	0	1	1	5
0	1	1	0	1	0	1	1	1	1	1	6
1	1	1	0	1	1	1	0	0	0	0	7
0	0	0	1	1	1	1	1	1	1	1	8
1	0	0	1	1	1	1	1	0	1	1	9

Fig. 1-18. The truth table for a bcd-to-seven-segment decoder.

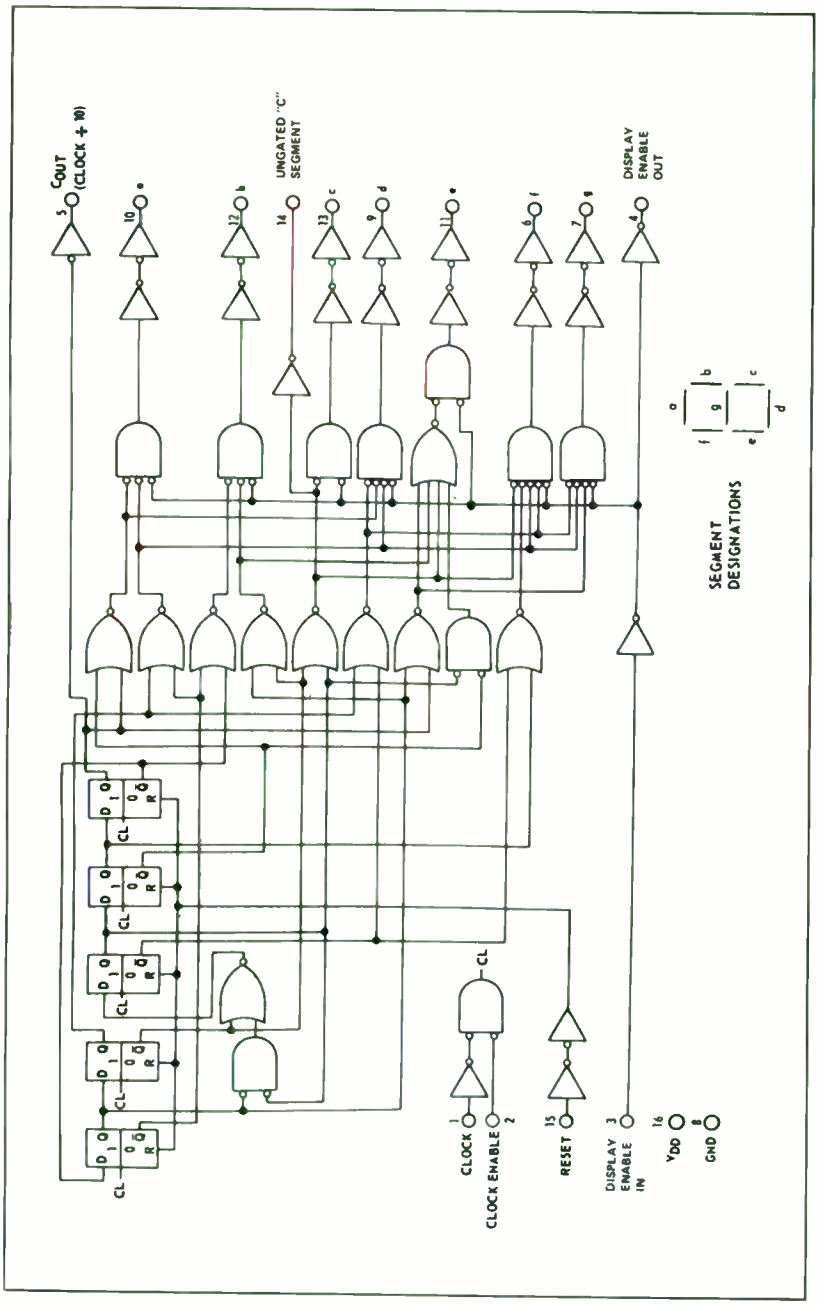


Fig. 1-19. Logic diagram for the 4026 CMOS IC.

put line to increase the current-driving capability of the outputs. When the IC is operated with a 9-volt power supply, each output can deliver a current of about 5 mA, which is sufficient to drive a small LED display.

In addition to a clock input, the 4026 has three control inputs. Normally the reset input is held at logic 0, the clock enable input at logic 0, and the display enable input at logic 1. A logic 1 signal to the reset input will reset the counter to zero, a logic 1 signal at the clock enable input will stop the counter, and a logic 0 signal at the display enable input will force all seven outputs low (turning off the display) while the counter continues to count. There are also three control outputs. The carry output goes from logic 0 to logic 1 when the counter goes from a count of nine to a count of zero. This positive transition of the carry output can be used to increment the next counter in a chain. The display enable output is simply a replica of the display enable input. The ungated "c"-segment output is the same as the "c"-segment output when the display enable input is high, but it is not disabled when the display enable input goes low. The ungated "c"-segment output is provided so that certain counting functions can be carried out while the display is disabled.

A CMOS DIGITAL TIMER

The digital timer circuit shown in Fig. 1-20 illustrates the practical application of several of the CMOS circuits we have discussed. The timer is built using one CMOS 4001 quad NOR gate and two CMOS 4026 counter/decoders. Two LED seven-segment displays (Monsanto MAN-3 displays or equivalent) are used for the readout.

NOR gates 1 and 2 are connected to form an RS flip-flop, like the one in Fig. 1-9; NOR gates 3 and 4 are connected to form a square-wave generator. This square-wave generator is similar to the one shown in Figs. 1-6 and 1-7, but with the distinction that one of the inputs to gate 3 is connected to the output of the RS flip-flop. When the output of the RS flip-flop is low, a square wave is produced at the output of gate 4. (The 500K pot is used to adjust the frequency of the square wave so that it is 1 Hz.) When the output of the RS flip-flop is high, however, the output of gate 3 is forced to logic 0 and a steady logic 1 appears at the output of gate 4.

Thus, the output of the RS flip-flop is used to determine whether or not a square wave is produced at the output of gate 4. When the start switch is depressed, the output of the flip-flop is forced to logic 0, and a square wave appears at the output of gate 4. When the stop switch is depressed, the output of the flip-flop is forced to

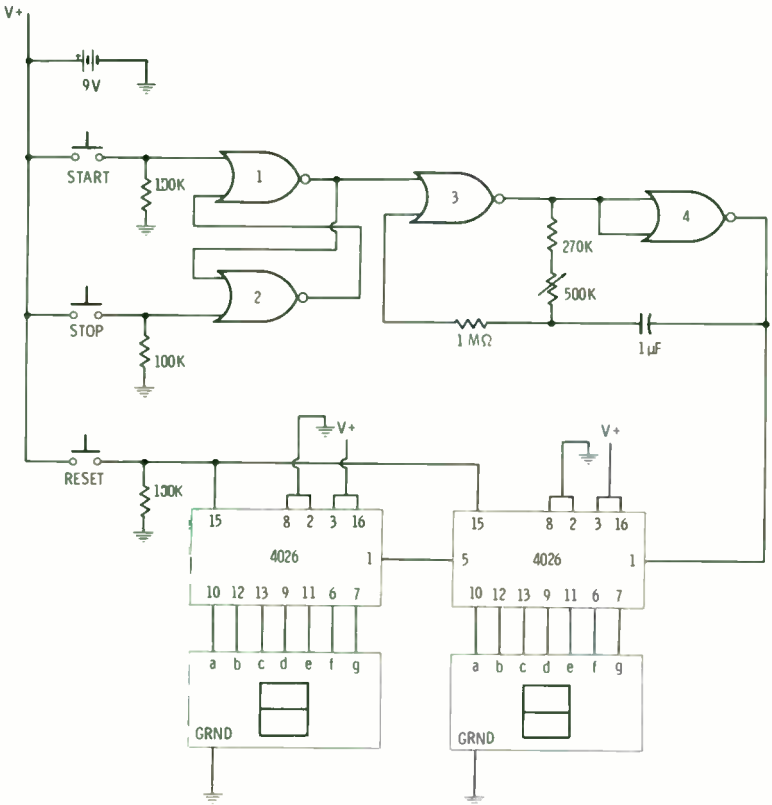


Fig. 1-20. A CMOS digital timer.

logic 1, the output of gate 3 goes to logic 0, and the output of gate 4 is fixed at logic 1.

The output of gate 4 is connected to the input of a 4026 counter/decoder. When there is a 1-Hz square wave at the output of gate 4, the counter advances once each second, on the positive transitions of the square wave.

The reset switch is used to reset the counters to zero. The start switch is used to start the timer, and the stop switch is used to stop the timer. With the two-digit display shown, the timer will measure elapsed time in 1-second increments up to a maximum of 99 seconds.

SUMMARY

We have discussed the basic CMOS digital integrated circuits and have shown how they can be used in practical circuit design.

We have defined the operation of ICs by means of truth tables, but we have not yet discussed the internal operation of the CMOS ICs. In order to be able to work with CMOS ICs most effectively, a clear understanding of their internal operation is indispensable. In the next three chapters we will discuss the operation of the integrated components that make up the CMOS IC, we will show how these components are fabricated, and we will describe the operation of the CMOS integrated circuitry.

2

CMOS IC Electronics

An elementary knowledge of CMOS IC electronics is essential to a thorough understanding of CMOS ICs. This chapter is an introduction to the fundamentals of CMOS IC electronics: semiconductor physics, CMOS devices, and CMOS parasitic devices.

SEMICONDUCTOR PHYSICS

A *semiconductor* is a crystalline structure with specific electrical properties. Silicon, gallium arsenide, and germanium are the most popular semiconductors in use today. All CMOS integrated circuits presently being built use silicon.

Fig. 2-1 shows a silicon ingot which has been grown for use in integrated circuits. These ingots are usually several inches in diameter and several feet long. The ingots are sliced, in the same way as bologna, into wafers about 0.3 mm thick. These silicon wafers are used as the starting material in the IC fabrication process.

Semiconductor Conduction

Metals provide a very simple illustration of the process of electrical conduction. When a voltage is applied across a metal, a "sea of electrons" drifts in response. Even though high currents may flow in a metal, the average velocity of the electrons is slow (typically less than a centimeter per second) because of the large number of electrons participating in the conduction. While the "sea of electrons" idea provides a clear picture of conduction in metals, it needs modification before it can be extended to semi-

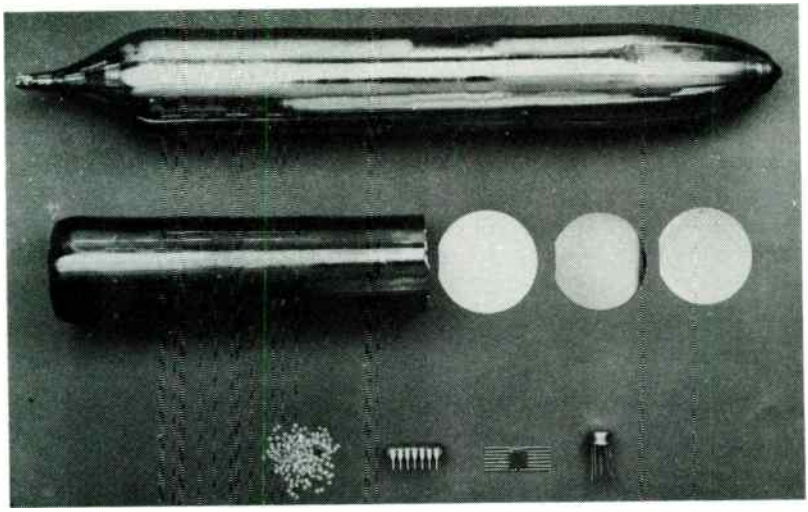


Fig. 2-1. The monolithic fabrication process from silicon ingot to packaged IC.

conductors. Just as in metals, though, electrons in semiconductors are the only type of charge that contributes to semiconductor conduction.

Electrons in semiconductors have certain specific energies. The allowed energies are grouped, and these groups of energy levels are called "energy bands," or just "bands." In nature, systems tend to exist in the lowest energy state. Consequently, electrons tend to occupy mostly the lower energy bands. The two highest energy bands containing electrons are called the *valence band* and the *conduction band* (shown in Fig. 2-2), and it is these bands that are important to conduction in semiconductors.

The conduction band is the higher of these two bands. It is only slightly filled. The electrons in this energy band act quite similarly to electrons in a metal during conduction. When a voltage is applied across a semiconductor, these electrons drift in response, similarly to the sea of electrons. The current identified with the flow of electrons in this band is called the *electron current,* and the *carriers,* or the type of charge said to be "carrying" the current, are electrons.

The valence band is just below the conduction band. The energy levels in this band are mostly filled. There are very few vacant energy levels available. It has been shown with quantum mechanics that these vacancies, or holes, behave very much like the electrons in the conduction band, except that they act as though they are positively charged. This can be seen more clearly by

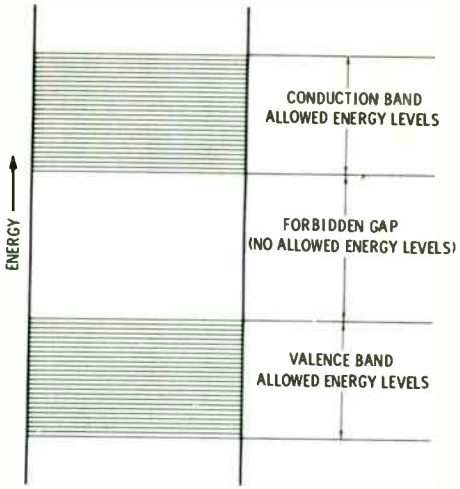


Fig. 2-2. Semiconductor energy levels.

looking at the conduction in the valence band, or the *hole conduction*. When a voltage is applied across a semiconductor, electrons in the valence band jump into nearby vacancies, leaving a vacancy at the energy level they came from. (The free electron and hole produced by this break are called a "hole-electron pair.") As seen in Fig. 2-3, the electrons are moving toward the right; but from another point of view, the vacancy, or hole, has moved toward the left. Thus, conduction in the valence band can be considered in terms of holes moving in a direction opposite that of the electrons in the conduction band.

In summary, when a voltage is placed across a semiconductor, there are two forms of current flow: current due to electrons in the conduction band, and current due to holes in the valence band. Unless otherwise noted, the term "electrons" will be used to de-

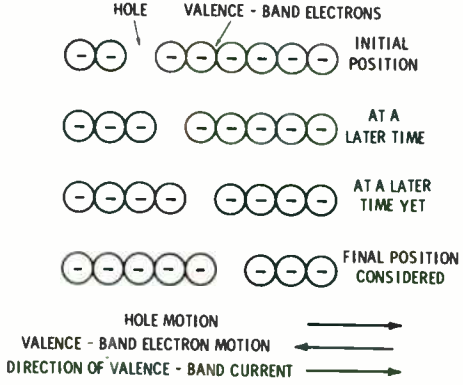


Fig. 2-3. Conceptual diagram of valence-band conduction.

scribe only the current carriers in the conduction band, and the conduction-band current will be referred to as "electron current." Similarly, "holes" will be used to describe current carriers in the valence band, and the valence-band current will be referred to as "hole current."

N-Type and P-Type Semiconductors

The concentration of electrons and holes in a semiconductor can be controlled. Two classes of impurity atoms are used to control the number of electrons and holes in semiconductors: *donors* and *acceptors*. Donor atoms have excess electrons which they are willing to "donate" when placed in a semiconductor. Acceptor atoms are the opposite of donors in that they "accept" electrons from the valence band when placed in a semiconductor. Most donors increase the number of electrons by one for every donor atom, and most acceptors increase the number of holes by one for every acceptor atom. By *selective doping*, or injecting impurity atoms into a semiconductor, it is possible to have a large number of holes in one area of a piece of semiconductor material and a large number of electrons in another. A large number of both holes and electrons cannot exist concurrently in the same area in a semiconductor, and if the number of holes in a certain area is increased tenfold by doping, the number of electrons in this area will be decreased tenfold. An area of semiconductor material with more electrons than holes is called *n-type*, and an area with more holes than electrons is called *p-type*.

Both n-type and p-type semiconductors can be fabricated on the same semiconductor wafer by selective doping. In fact, a p-type semiconductor can be changed to an n-type by *counterdoping* acceptor-doped material with a greater number of donor atoms. Counterdoping can likewise be used to change n-type material to p-type material.

Semiconductor Doping

Semiconductors are typically doped at two different times during IC fabrication. During the growth of the semiconductor crystals, a dopant is added to make the entire crystal n-type or p-type. After the crystal is sliced into wafers, it may be doped by a process called *diffusion*.

Diffusion is simply the migration of atoms from an area of high concentration to an area of low concentration. Many people have observed atomic diffusion when smelling perfume across the room from its source. The doping/diffusion process is carried out by passing a dopant-rich gas over the surface of a semiconductor wafer which has been heated to about 2000°F. The surface of the

semiconductor absorbs the dopant atoms, and they diffuse deep into the semiconductor.

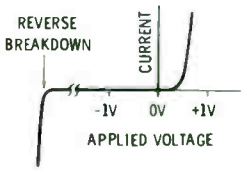
Another technique for doping semiconductors utilizes ion implantation. In this technique, a dopant-rich gas is ionized and accelerated toward the sliced semiconductor wafers, which are at room temperature. Since 100 kilovolts is typically used to accelerate the ions, the dopant atoms are implanted many atomic layers deep into the bombarded semiconductor wafer. The dose of implanted dopant ions can be very accurately measured and controlled by monitoring the ion current (the high-voltage power-supply current) as it is implanted. This doping technique can be combined with high-temperature diffusion to create deeper doped layers than is possible with practical values of accelerating voltage.

CMOS ICs are often fabricated with the high-temperature doping/diffusion process used to create the dose-insensitive high-conductivity *source/drain diffusion* (see Chapter 3). The ion-implantation process is used to determine the critical doping in the MOS transistor channels. (The channel doping will be shown to have great influence on the gate voltage required to turn the MOS transistors "on." Chapter 3 focuses on the details of these basic doping processes and their use in CMOS IC fabrication.)

THE DIODE

The diode is the basic element of CMOS integrated circuits. Diodes consist of an n-type doped semiconductor area adjacent to a p-type doped semiconductor area, as shown in Fig. 2-4A. The diode-characteristic curve (voltage vs current) and schematic symbol are also shown in Fig. 2-4. Typically, a diode is fabricated by counter-doping part of an n-type silicon wafer through a diffusion process so that it becomes a p-type.

(A) Semiconductor diode data.



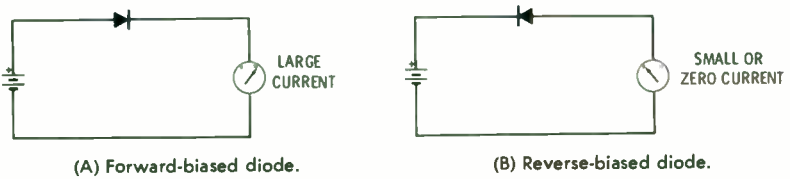
(B) V-I characteristic curve.

(C) Schematic symbol.

Fig. 2-4. Semiconductor diode data

A *diode* ideally passes current in only one direction, as shown in Fig. 2-5. In order to understand how this *diode action* occurs in a semiconductor, it is first necessary to understand the balance that goes on in semiconductors.

This balance exists between the two types of current that can exist in a semiconductor: drift current and diffusion current. *Drift current* is caused by the flow of carriers in an electric field. *Diffusion current* is caused by the migration, or diffusion, of carriers from an area of high carrier concentration to an area of lower concentration of the same type of carrier. Carrier diffusion is similar to the atomic diffusion of impurity atoms discussed previously. However, a major difference between these two types of diffusion is that carriers at room temperature diffuse readily, whereas doping atoms at the same temperature diffuse at a rate which, for all practical purposes, is zero. This is fortunate, since it is undesirable to have the doping atoms move after they are initially placed.



(A) Forward-biased diode. (B) Reverse-biased diode.

Fig. 2-5. Diagrammatic representation of diode action.

When there is a difference in hole concentration between two adjacent areas in a semiconductor, a diffusion current will flow from the area of high hole concentration to the area of low hole concentration. Since the holes are diffusing from an electrically neutral area, an electric field is created between the holes that have diffused and the negatively charged area they created in leaving. This electric field causes a drift current to flow, as shown in Fig. 2-6. At equilibrium, the diffusion current is equal and opposite to the drift current, and there is no net current flow.

The balance between drift and diffusion current is very important at the junction between the p-type and the n-type material in a semiconductor diode. On the p-type side of the junction, there are a large number of holes; while on the n-type side there are few. Thus, there is a balance between drift and diffusion currents for both the electrons and the holes. The amount of diffusion current that would flow, if it were not restrained by the electric field, would be several thousand amps for a typical diode. Yet in a silicon diode, this current is restrained by the electric field, which is only .6 volt. Note that electrons and holes are trying to diffuse in opposite directions across the junction; yet because they

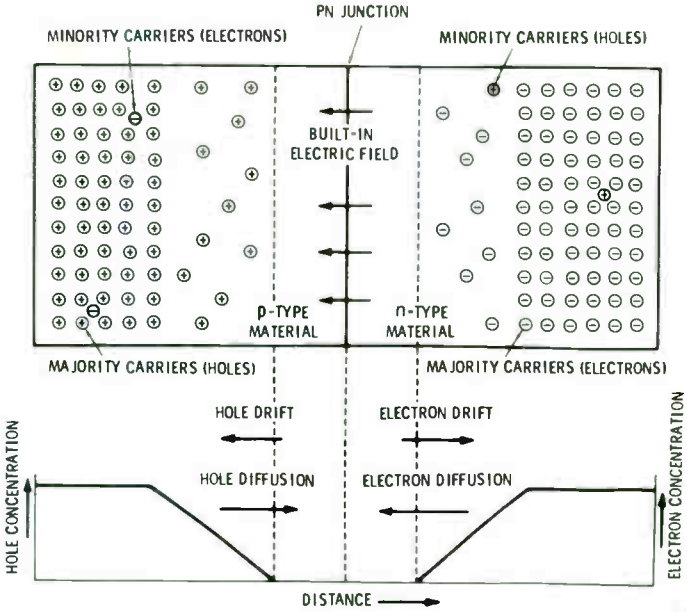
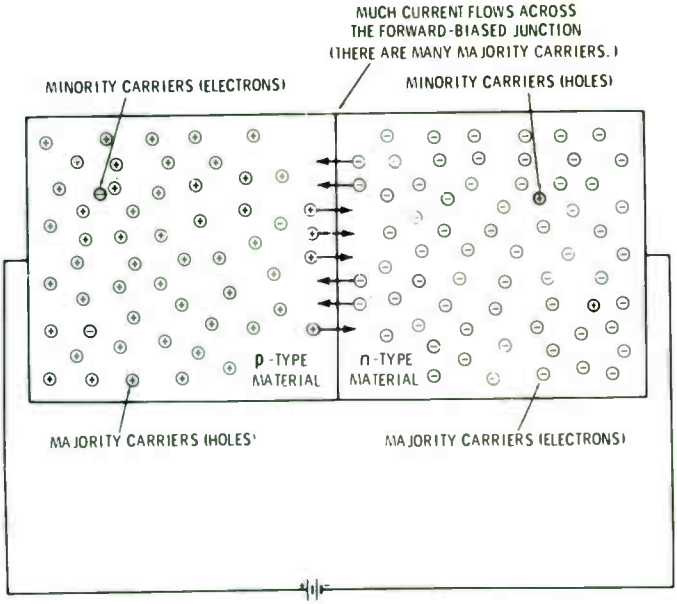


Fig. 2-6. Pictorial drawing of semiconductor diode. Plot of majority-carrier concentration.

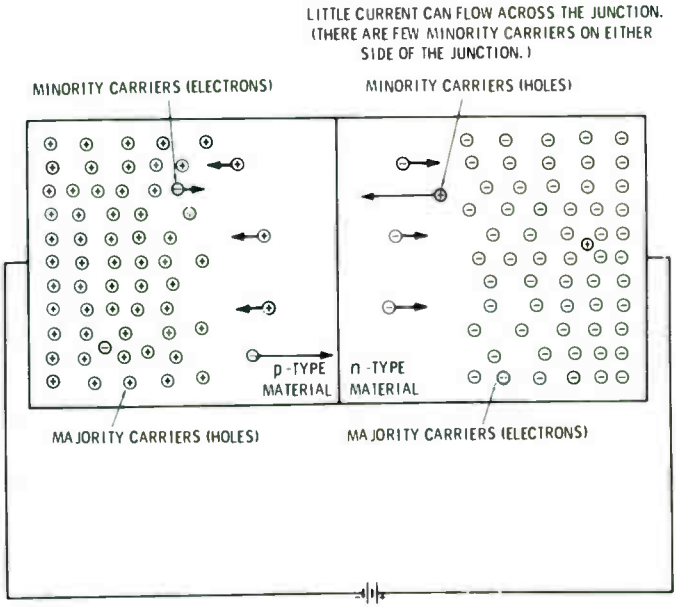
are oppositely charged, the same polarity of electric field restrains both types of carriers.

The diode action of a semiconductor junction may now be understood in terms of the drift/diffusion balance. In the conduction, or forward-biased direction (shown in Fig. 2-7A), the voltage applied from a circuit to the diode decreases the built-in electric field, releasing part of the potentially large diffusion current. For a silicon diode, several thousand amps of diffusion current would flow if .6 volt were applied across the junction. This is nearly impossible to do in practice because a very high voltage would have to be applied to the leads of the diode in order to overcome the resistive voltage drop in the semiconductor material as the current flowed to and from the junction. The diode would probably overheat long before this current could be achieved.

In the nonconducting, or reverse-biased, direction (shown in Fig. 2-7B), the applied voltage acts to upset the balance between drift and diffusion in favor of the drift current. The current flowing in this direction is limited by the diffusion of the minority carriers on either side of the junction. Upon diffusing to the junction, the minority carriers are swept across by the bias-aided electric field. These few minority carriers drifting across the junction constitute an imbalance between drift and diffusion currents at the



(A) Forward-biased diode.



(B) Reverse-biased diode.

Fig. 2-7. Pictorial representation of diode action.

junction. This reverse current is not dependent on reverse-bias voltage because it is limited by the rate of diffusion of minority carriers to the junction.

If the reverse-bias voltage is large, the diode will undergo *reverse breakdown*. Reverse breakdown is the phenomenon which causes a large increase in reverse current corresponding to a small increase in reverse bias above the breakdown, or *zener*, voltage of the diode (see Fig. 2-4B). (Zener diodes are ordinary pn junction diodes that have been designed and sorted for their reverse-breakdown characteristics.) When a diode breaks down, the reverse current must be limited to prevent excessive power dissipation by the diode. If the power is limited to the designed dissipation of the diode and package (typically ¼ to ½ watt), the reverse breakdown of the diode will be nondestructive. If the power is not limited, the diode will overheat and be permanently damaged.

Although it is often called "zener breakdown," reverse breakdown in a diode actually may be due to either zener breakdown or avalanche multiplication. Zener breakdown occurs in a diode when an area near the junction, which has been swept free of carriers, has a large enough electric field to "rip" electrons and holes from the atoms in that area. These electrons and holes are the carriers of the large reverse current resulting from zener breakdown.

With an electric field less than that required for zener breakdown, avalanche multiplication can occur. Avalanche multiplication is the effect that causes an increase in the reverse current. The increase is caused by the collision of the carriers constituting the reverse current with the crystal lattice near the junction. The carriers are accelerated by the large electric field near the junction. These collisions "knock loose" hole-electron pairs from the silicon atoms which, depending upon the location of the original collision, may be accelerated to cause more carrier-generating collisions. It is the electrons and holes knocked off by collisions that constitute carriers of the large reverse current in avalanche breakdown.

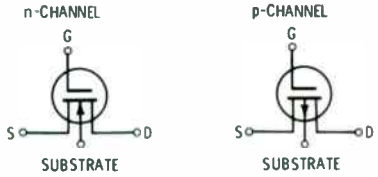
To summarize, silicon diodes conduct large amounts of current in the forward-biased direction when about 0.7 volt is applied to them. In the reverse-biased direction, only a little current flows, unless the bias is large enough to cause reverse breakdown. Reverse breakdown will not harm a diode as long as the power dissipated in the diode is limited to the designed level.

CMOS ICs contain many reverse-biased diodes as part of the IC chip. The diodes typically are used in the integrated-circuit design to form the source and drain of the MOS transistors.

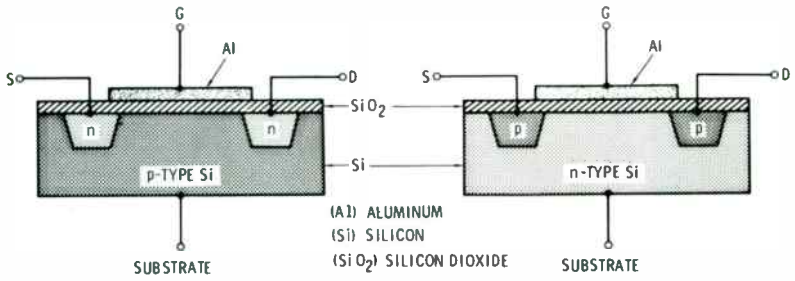
THE MOS TRANSISTOR

The MOS (metal-oxide-semiconductor) transistor is a semiconductor device that can be used to proportionally control large currents with small voltages. There are two basic types of MOS transistors: n-channel and p-channel. When fabricated together on the same IC chip, these two types of transistors constitute the basic building blocks of CMOS ICs. An understanding of the operation of these two types of MOS transistors will prove useful in understanding the CMOS circuit designs described in Chapter 4. The MOS transistors are quite different from the common npn and pnp bipolar transistors. The MOS transistor may be thought of as a voltage-controlled device, while the bipolar transistor may be thought of as a current-controlled device. The voltage-controlled nature of the MOS transistor will be discussed in detail in the remainder of this and the following section. This discussion of MOS transistors will be followed by a discussion of the bipolar transistor and other *parasitic devices* which are undesirable (but unavoidable) semiconductor devices found in CMOS ICs.

The schematic symbol and the semiconductor structure for each type of MOS transistor are shown in Fig. 2-8. The operation of these devices with a transconductance, g_m, of .01 mho is shown in Fig. 2-9 for both n-channel and p-channel transistors. It can be seen from Fig. 2-9 that the change in current flow into the drain is .01 times the change in input voltage.

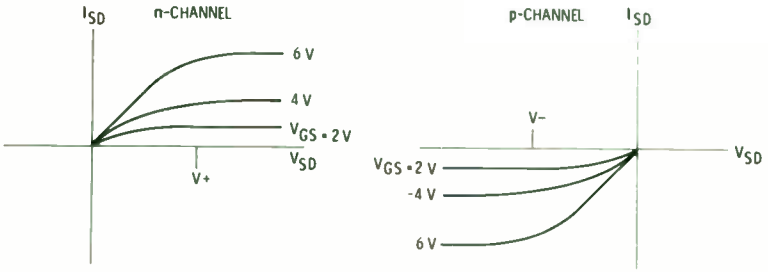


(A) Schematic symbols.

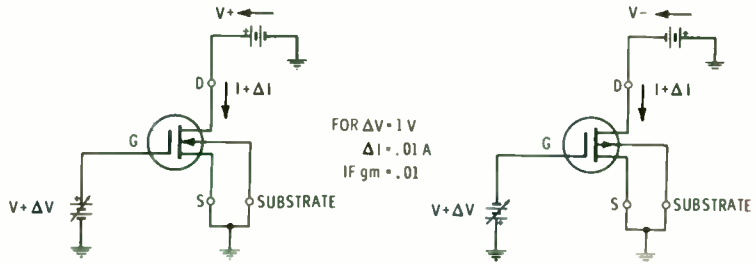


(B) Semiconductor cross section for n- and p-channel MOS transistors.

Fig. 2-8. MOS transistor symbols and cross sections.



(A) Source-drain voltage-current characteristics.



(B) Bias for normal operating condition of n- and p-channel MOS transistors.

Fig. 2-9. Characteristics and bias for MOS transistors.

Transconductance is defined as the change in source-drain current divided by the change in applied gate-source bias that caused it.

$$\text{Transconductance} = \Delta I/\Delta V$$

where,

ΔI is an incremental change in current,
ΔV is an incremental change in voltage.

The MOS transistor may be operated in one of two device modes (which will be discussed in more detail later): (1) the voltage-controlled resistor mode (sometimes called the *linear region*) and (2) the current-limited mode (sometimes called the *saturation region*). The current-limited mode is a special case of the voltage-controlled resistor mode. In this current-limited mode the voltage applied between the drain and the source is large, resulting in *saturation* or current limiting between these two terminals. This current limiting may be interpreted as the saturation of the hypothetical voltage-controlled resistor between the source and the drain. In actuality this current limiting is the result of all of the available charge from the source (for a particular value

of gate-source voltage) flowing at the maximum crystal-collision-limited velocity. Increases in source-drain bias above the saturation bias will not significantly increase the source-drain current.

For the current-limited mode of operation, the output current and, hence, the transconductance are independent of source-drain voltage. At lower values of source-drain voltage, the source-drain current and the transconductance are both directly proportional to source-drain voltage. The concept of transconductance (or "transfer-conductance") is less useful for the voltage-controlled resistance mode.

It is interesting to note that since the input impedance is very high (essentially that of a capacitance of a few picofarads), the input power required to control the output current is very small.

The MOS Transistor Fundamentals

Now that the basic circuit properties of the device are understood, the physical phenomena of the device in the semiconductor can be presented.

From the structure of the transistor shown in Fig. 2-8, it becomes apparent that the basic device is comprised of two reverse-connected diodes shunted by a metal electrode closely spaced to, and insulated from, the semiconductor surface by an oxide layer. These are the layers of metal, oxide, and semiconductor from which the MOS transistor obtains its name. The properties of the MOS transistor with no gate bias applied are simply those of two reverse-connected diodes. No current flows between source and drain in this simple structure with no gate bias because one of the two junctions is reverse biased for either polarity of source-drain voltage. If bias voltage is applied to the gate electrode with respect to the substrate, it is possible to create an electric-field-induced region which interconnects the source and the drain. The n-channel transistor, for example, may be made to conduct between source and drain by applying a positive voltage to the gate electrode with respect to the p-type doped substrate. This positive bias on the gate repels the majority carriers (holes) away from the surface and attracts electrons from the MOS-transistor source region into the region beneath the gate electrode (as is shown in Fig. 2-10). The electrons beneath the gate electrode are then attracted to the drain region by a bias voltage, and the flow of source-drain current is established through a field-induced *channel.*

There are electromagnetic laws (Maxwell's laws) that must be obeyed during this process. All charges placed on the metal gate electrode must be balanced by equal and opposite charges in the semiconductor. Before the arrival of the electrons from the source

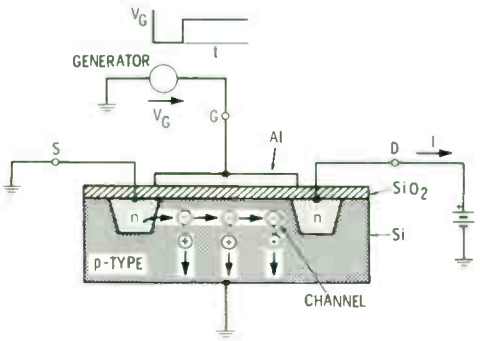


Fig. 2-10. Formation of conduction channel between the MOS transistor source and drain.

into the channel region, the positive charge is balanced by the negatively charged ionized dopant atoms of the p-type substrate (exposed by the departing holes) and by a few electrons which were native to the region beneath the electrode. As the attracted electrons from the source fill the channel region, these electrons balance the positive charge on the metal gate electrode. During this redistribution process, some of the ionized dopant atoms farther from the semiconductor surface are covered by holes as a result of the movement of electrons at the surface. Hence, the role of the dopant atoms in balancing the gate charge is sometimes superseded by the electrons from the source. As in many instances in nature, this action of the electrons filling in along the semiconductor surface has occurred because it represents a lower energy condition than that required for balancing the gate charge by ionized dopant atoms only. However, there is an equilibrium value of the channel electrons for a given gate voltage. That is the minimum energy condition, and the electromagnetic field laws require that the gate charge be balanced in part by channel electrons and in part by ionized dopant atoms. These channel electrons may be withdrawn from the channel region by the application of a bias voltage between the source and the drain terminals. Since the source and the drain are high-conductivity regions, virtually the entire voltage is dropped across the channel region. This results in the drift of the channel electrons into the drain and the flow of replacement electrons into the channel from the source.

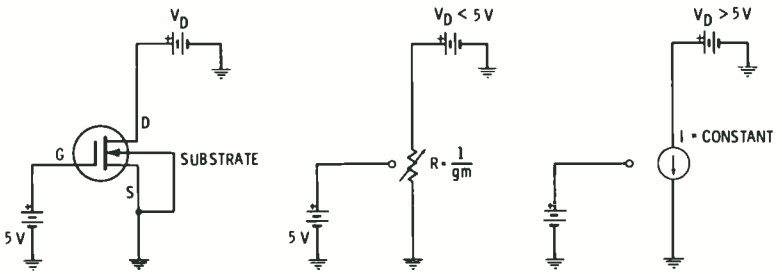
So far, the description of the MOS transistor has pertained to the operation of the device as a voltage-controlled resistor. This gate-voltage-controlled-resistor description of the device properties is accurate if (1) neither the drain/substrate nor the source/substrate junctions are forward biased during any time in the circuit operation and (2) the source-drain, drain-substrate and

source-substrate junction reverse-bias voltages are much smaller than the gate/substrate bias.

The first requirement of reverse-biased junctions is an obvious one from the device description, but it is a requirement that if violated can lead to "mysterious" circuit phenomena. Basically the junctions must be reverse (or zero) biased to make the conduction resulting from the gate-induced channel dominant. Forward biasing the source results in a drain current that is independent of the gate-source voltage. This is described in the next section on parasitic CMOS semiconductor devices.

The second requirement, that gate-source bias be much larger than the source and the drain bias, is necessary to ensure the voltage-controlled-resistor mode of operation frequently used in CMOS IC design. Large values of source-drain bias result in the source-drain resistance operating in the other MOS transistor mode, the current-limited mode of operation. This distinction between the voltage-controlled-resistor mode and the current-limited mode of operation can be more clearly understood by noting the differences in the current flow within the device for the two modes of operation.

For the present discussion, assume that the source of an n-channel MOS transistor is connected to the substrate, as shown in Fig. 2-11, and that the gate is biased at five volts with respect to the substrate. For values of drain-source voltages much less than five volts, the MOS transistor performs like the voltage-controlled resistor previously discussed. For a drain-source voltage much greater than five volts, the interaction of the drain-source voltage and the gate-substrate voltage significantly modifies the location and flow of channel current in the region immediately adjacent to the drain, as shown in Fig. 2-12. This figure shows the channel current as flowing at the semiconductor surface through the entire channel when the device is in the voltage-controlled-resistor

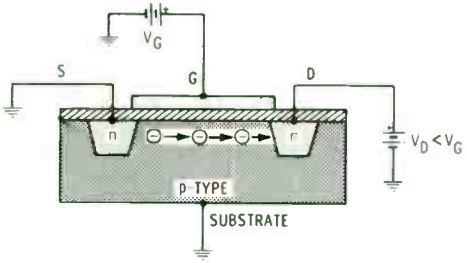


(A) Bias configuration.

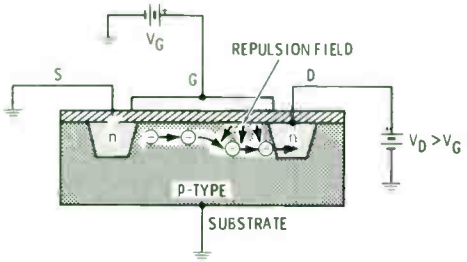
(B) Voltage-controlled-resistor equivalent circuit.

(C) Current-limited-resistor equivalent circuit.

Fig. 2-11. Grounded-source bias.



(A) Biased in voltage-controlled-resistor mode.



(B) Biased in current-limited mode.

Fig. 2-12. Flow of current from source to drain for an n-channel transistor.

mode. When the device is in the current-limited mode, the figure shows it as flowing along the surface through the greater part of the channel and beneath the surface in the region adjacent to the drain region. The subsurface flow of channel current near the drain and the current-limited source-drain property are the result of the electric field created by the gate-substrate and the drain-source voltages. The application of electrostatic field laws reveals that the net electric field at the surface near the drain beneath the gate is in opposite directions for the two operational modes of the MOS transistor.

Thus, the action of a large drain-source voltage on the surface electric field near the drain is to reverse the field direction by making the drain region more positive (for the n-channel example) than the channel and the gate electrode. The net effect of the field-direction reversal is to repel the channel carriers from the surface as they approach the drain. An exact analysis of the current-limiting phenomena is beyond the scope of this book. It is sufficient to say that increases in drain-source voltage for the MOS transistor operating in the current-limited mode result principally in increased depth of current flow in the drain region but do not result in a net current increase. This is because the

increases in drain-source/substrate voltage (the source is connected to the substrate for the purposes of this discussion) do not alter the magnitude of the electric field along the entire channel. (All of the voltage increase is dropped in a repulsion field at the surface near the drain.)

So far in this section, the basic operation of the MOS transistor has been discussed for the example of source connected to substrate ("grounded source"). The origin of the MOS transistor voltage-controlled-resistor mode has been shown to result from a gate-controlled electric field which induces a resistive channel at the semiconductor surface. The current-limited mode has been described as originating from the influence of the interaction of the drain and gate electric fields. This results in the source-drain current flowing along the semiconductor surface as it leaves the source, and also flowing in the semiconductor bulk as it reaches the drain. In the remainder of this section, the effects of voltages applied between the source and the substrate will be discussed. From the simplest description of the action of the MOS transistor, the source, the gate, and the drain all may be thought of as acting independently. The source injects charge into the channel region established by the gate bias, and the drain withdraws charge from the channel. As has been discussed in the beginning of this section, the interaction of the gate bias with the drain bias at the boundary between the gate and the drain results in current-limited mode of operation.

Similarly there is an interaction between the gate-substrate bias and the source-substrate bias. Mobile carriers flow into the channel from the source because they have a lower energy there. They essentially "roll" down (drift and diffuse) the potential hill resulting from the applied gate-substrate voltage. If the source is biased at some voltage with respect to the substrate, the height of the potential hill between the source and the channel is altered. The hill may be raised (aiding the supply of charge in the channel) by *forward biasing* the diode with respect to the substrate. As discussed previously in this section, forward biasing the source diode results in charge being injected in the substrate and is often undesirable. If the diode is *reverse biased* with respect to the substrate, the potential hill will be lowered. Reverse-biased source diodes are common in CMOS IC designs. Due to the lowering of the source-channel potential hill with the application of reverse bias to the source-substrate junction, less source-drain current will flow.

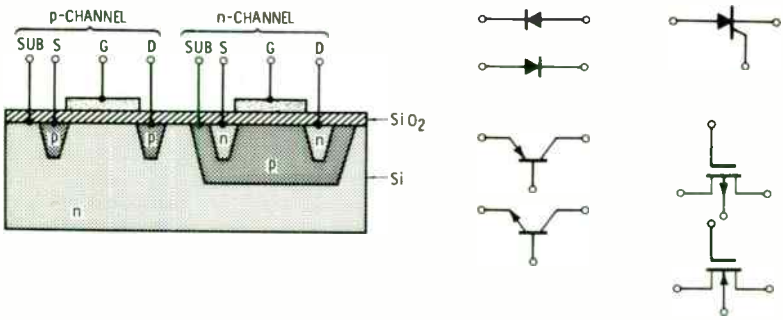
The voltage required to establish a gate-substrate MOS transistor channel between the source and the drain is called the *threshold voltage*. The idealized MOS transistor has no conduction be-

tween source and drain if the gate-substrate bias is less than the threshold voltage and if the MOS transistor conducts in the resistor or current-limited mode for values of gate-substrate bias greater than the threshold voltage.

If the source-substrate junction is reverse biased, the threshold voltage increases by a value slightly greater than the amount of source-substrate bias voltage. Historically the substrate of discrete MOS transistors has been regarded as a second "gate" (a back gate) terminal since the source-substrate bias influences the source-drain current. From a circuit point of view, the MOS transistor may be thought of as having source-drain conduction responding to gate-source bias modified by reverse bias on the source. In fact, the threshold voltage of the MOS transistor increases with increased source bias. This effect is referred to as the *back-gate-bias* threshold-voltage increase. More on the circuit properties of MOS transistors will be discussed in Chapter 4. The remainder of this chapter will cover the influence of undesired transistors, silicon-controlled rectifiers, and diodes that are often unavoidably fabricated at the same time as the n-channel and p-channel MOS transistors comprising the CMOS ICs.

CMOS PARASITIC DEVICES

As discussed in Chapter 3, the steps and sequence used to fabricate CMOS ICs are often optimized for good n- and p-channel MOS transistor performance at low manufacturing cost. Consequently, undesired *parasitic devices* are often fabricated on the same substrate as the CMOS transistors. As shown in Fig. 2-13, there are pnp and npn bipolar transistors in many CMOS ICs. It is useful to be aware of the influences of these components when



(A) CMOS semiconductor cross section.

(B) CMOS parasitic-device schematic symbols.

Fig. 2-13. CMOS cross section and parasitic symbols.

you are designing, maintaining, or debugging equipment with CMOS ICs. Understanding these influences will aid in understanding how small infractions of the manufacturers' recommended procedures may sometimes result in highly "unusual" IC performance.

Parasitic Bipolar Transistors

A bipolar transistor has an emitter, a base, and a collector which are similar in structure to the MOS transistor source, substrate, and drain. The principal differences between the two structures arise as a result of their intended application and bias conditions. In the bipolar transistor, for example, the base spacing between the collector and the emitter is typically small. An MOS transistor may be operated as a bipolar transistor, however, by connecting the source, substrate, and drain as the emitter, base, and collector (shown in Fig. 2-14). In operation, the source-substrate junction is forward biased, and the drain substrate junction is reverse biased. It is interesting to note that an MOS transistor may be operated as a bipolar transistor in a CMOS integrated circuit by forward biasing the source-substrate junction. In some CMOS ICs, *protection* diodes are attached between the gate inputs and the substrate. These diodes may be forward biased by establishing an input voltage greater than the positive supply voltage for the IC or below the negative supply voltage. Any of the drains in a CMOS IC circuit may collect the carriers injected by the forward-biased input protection diodes. Unusual circuit performance may result from the collection of the injected car-

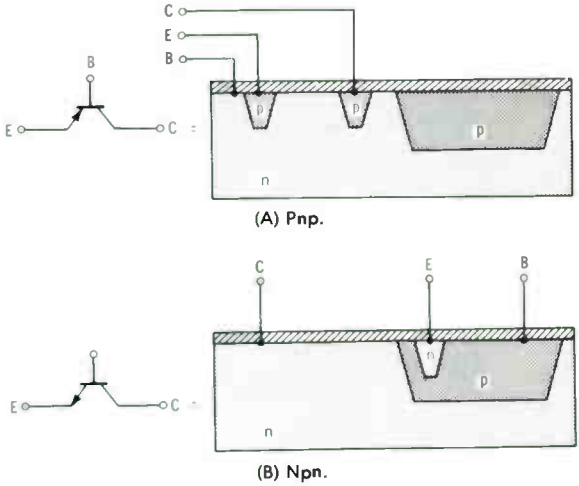
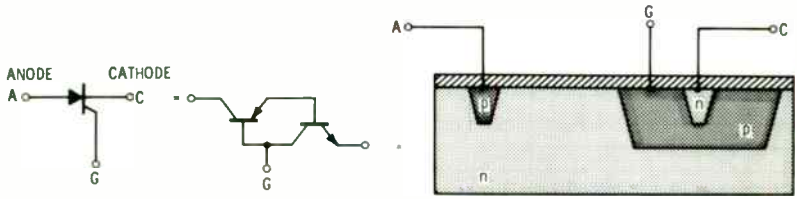


Fig. 2-14. Parasitic transistors and their realization in CMOS ICs.

riers. Digital signals, for example, passing through a CMOS IC that is experiencing substrate charge injection may be altered in a seemingly random fashion by the injected carriers.

Parasitic SCR's

As shown in Fig. 2-15, both npn and pnp parasitic transistor structures are present in CMOS ICs. When acting together, the npn and pnp transistors form a parasitic silicon-controlled rectifier (SCR). An SCR is a device that may pass either large forward currents or small leakage currents between the SCR *anode* and the *cathode*, depending on whether they are in the "on" or the "off" condition. In the on condition, the interconnection of npn and pnp transistors is such that each transistor supplies the neces-

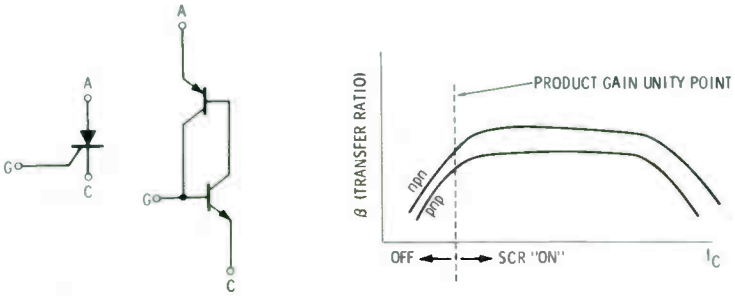


(A) Schematic symbol. (B) Equivalent circuit. (C) Semiconductor cross section.

Fig. 2-15. Parasitic SCR and its realization in CMOS ICs.

sary base current to the other transistor. In the off condition, neither transistor is conducting, which results in neither transistor supplying the other with the necessary base current for conduction. The small leakage currents that do exist are not sufficient to trigger the SCR because of the properties of bipolar-transistor current gain.

As shown in Fig. 2-16, the current gains of pnp and npn bipolar transistors may be constant for highly variable collector currents. However, for very small collector currents, second-order effects (such as recombination in the base-emitter depletion layer) result in decreased current gains. As indicated in the figure, the product of the npn and pnp transistor current gains must be greater than one in order to maintain the conducting, or on, state of the SCR. This requirement is the result of the interdependence of the npn and pnp transistors for base current. If the product of the npn and pnp current gains is less than one, neither transistor can adequately supply the base current needed by the other and the conduction of current between the SCR's anode and cathode will be reduced to leakage-current levels. Thus, the SCR is a device that continues to conduct after being triggered as long as the



(A) SCR equivalent npn/pnp transistor circuit.

(B) Equivalent circuit β versus IC curves.

Fig. 2-16. SCR equivalent transistor circuit.

conduction current is greater than the *sustaining current* (the current level for which the product of the current gains of the two devices exceeds one).

If triggered, the CMOS parasitic SCR can cause catastrophic failure in CMOS ICs. The action of the triggered SCR (as shown in Fig. 2-17) is to provide a low-impedance pathway between power supply and ground through the CMOS IC. CMOS ICs are not designed to withstand these current levels and may fail if the currents are not externally limited by a resistor or a regulated power supply.

Parasitic MOS Transistors

As shown in 2-13, there are parasitic MOS transistors as well as diodes, bipolar transistors, and SCRs. These transistors exist wherever metal interconnections cross oxide-covered source-diffused and drain-diffused regions. Typically, the covering-oxide

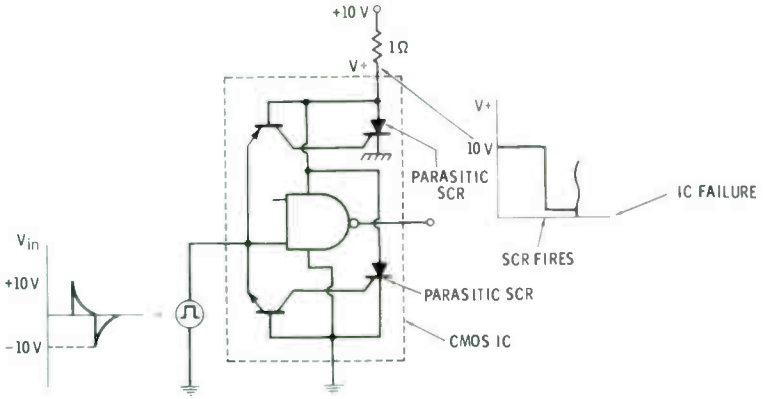


Fig. 2-17. CMOS IC destruction via parasitic SCR excitation.

layer is thick and the substrate regions surrounding these diffusions are intentionally heavily doped to provide *channel stop regions*. The thick oxide layering and the heavy doping raise the threshold of these MOS transistors above the maximum specified supply voltage and render these thick-oxide parasitic MOS transistors inactive.

Manufacturers of CMOS perform special procedures to minimize the effects of the IC parasitic devices already discussed. It is important, however, to be aware of them and their origin in order to detect errors made in designs using CMOS ICs. For example, applying waveforms to the inputs of CMOS ICs that exceed the CMOS IC positive or negative supply terminal voltage may result in circuit degradations due to parasitic bipolar and SCR devices.

The next chapter covers the fabrication techniques used to assemble CMOS ICs. The variations of techniques discussed result in significantly varied circuit performances and costs. An understanding of the basic elements of these structural variations makes possible the selection of the proper technology for a specific job.

3

CMOS Fabrication Technology

There are five basic fabrication techniques which are varied in sequence and degree to realize the five popular variations of CMOS fabrication. These fundamental CMOS wafer-fabrication techniques are:

1. Oxidation
2. Photolithography
3. Open-tube doping
4. Ion implantation
5. Deposition

There are five popular variations in the sequence of application of these basic steps. These variations have different effects on CMOS IC performance and have been given popular names. The most popular variations are:

1. Aluminum-gate CMOS
2. Silicon-gate CMOS
3. Oxide-isolation CMOS
4. Silicon-on-sapphire CMOS
5. Dielectric-isolation CMOS

The basic fabrication steps will be presented in the first part of this chapter, followed by a discussion of the fabrication variations and the influence of these variations on CMOS IC performance.

WAFER-FABRICATION TECHNIQUES

Oxidation

Oxidation is the fabrication step used to thermally grow layers of silicon dioxide (a very durable insulator) on silicon wafers. This growth process is accomplished by heating the silicon wafers to 2000°F (typically) in an oxygen-rich environment, as shown in Fig. 3-1. The oxygen and the silicon chemically bond to form silicon dioxide. The resulting oxide layers are used to protect the fabricated circuits from contaminants and are often used during IC fabrication as part of the photolithography process.

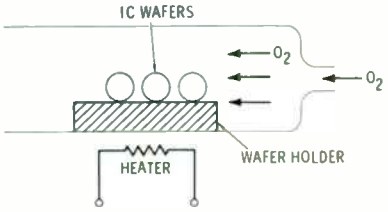
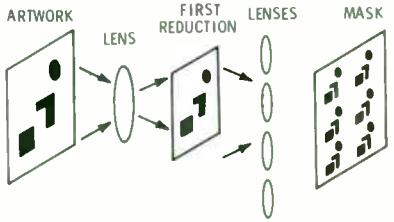


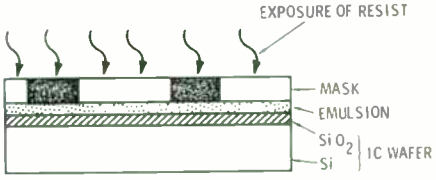
Fig. 3-1. Open-tube oxidation.

Photolithography

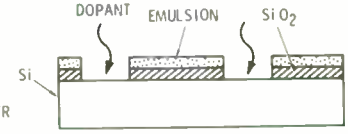
The photolithographic process for IC fabrication, shown in Fig. 3-2, originates with design drawings or taped artwork (typically four hundred times larger than the final IC). These design drawings specify the regions that are to be protected by an acid-resis-



(A) Photographic artwork reduction (200:1) and repetition (500 per mask).



(B) Photoresist emulsion exposure for IC wafer.



(C) Developed photoresist and etched SiO2 ready for doping Si wafer.

Fig. 3-2. Photolithography process.

tant emulsion. These design drawings are photographed, photoreduced and photorepeated so that hundreds of ICs may be fabricated simultaneously on one silicon wafer. The glass negatives, or *masks,* resulting from the photographic process are used in CMOS fabrication. The masks are placed in contact with the silicon wafers which have been oxidized and coated with a photosensitive emulsion. Exposure to ultraviolet light through the mask causes polymerization of the unprotected regions of the emulsion. After the unpolymerized emulsion has been chemically dissolved, the polymerized emulsion remains and protects the underlying regions during the chemical etching of the oxide layer.

After the emulsion, often called *photoresist,* is removed, the silicon dioxide layer remains in predetermined regions (because of the mask) on the wafer. This process of simultaneously etching selected windows in the oxide layer of hundreds of ICs on a wafer is made simple by the photolithographic process. The etched windows in the oxide layer allow open-tube doping to be selectively applied to the wafer.

Open-Tube Doping

CMOS IC fabrication requires that n- and p-type regions be selectively positioned in the silicon wafer. As shown in Fig. 3-3, this can be accomplished by heating the silicon wafers to 2000°F in a dopant-rich environment. Selective doping in the desired regions is accomplished by selectively etching windows (discussed in the previous section on photolithography) in a silicon dioxide layer. The doping is selective because the dopant *diffuses* into the silicon (through the etched windows) more easily than it does into oxide. Thus, during the doping process the silicon dioxide layer is used as a high-temperature mask bearing the image of the original photolithographic design drawings.

Thus, a high-temperature, dopant-rich environment allows the creation of the n- and p-type regions that are used as the sources and drains of CMOS transistors.

Typically, after a dose of the dopant is selectively diffused into the wafer, the dopant is driven in to create a deep junction. This *drive-in* process involves heating the wafers in an inert environment. The doping step is often called a *predeposition.* A predepo-

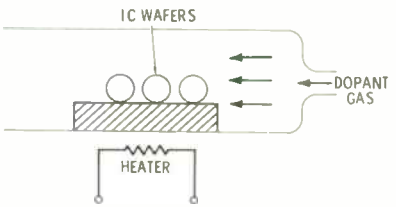


Fig. 3-3. Open-tube doping.

sition followed by a drive-in results in a controlled dose and a controlled junction depth in the CMOS transistor sources and drains.

Ion Implantation

The fabrication of CMOS transistors requires the generation of an n-type substrate in which p-channel transistors are embedded and also requires the generation of a p-type substrate region in which the n-channel transistors are embedded. *Ion implantation* is the technique used to realize the p-type islands in the n-type wafer which are used as the substrate of the n-channel transistors. Ion implantation is used in this step because the concentration of dopant in the p-well directly beneath the n-channel MOS transistor gate electrode plays a larger role in determining the n-channel transistor threshold voltage. Heavy doping of the well increases the threshold voltage and light doping decreases it. The key reason for using ion implantation in this critical doping task is its reproducibility of accurate concentrations. Ion implantation utilizes the acceleration of ionized dopant gas as a basis for implantation, as shown in Fig. 3-4. In operation, the ionized dopant is accelerated by a pair of widely spaced electrode plates (with typically 100,000 volts applied to them). The dopant gas is travelling at a very high velocity as it approaches the electrode plate to which it is attracted. Attached to this electrode is the n-type silicon wafer to be implanted. The dosage can be determined by monitoring the ion flow with a conventional ammeter; the ammeter measures the current supplied by the high-voltage supply. A magnet is used in the ion steam to separate the desired dopant from the other atoms. The deflection of the ion stream is determined by the mass of the atoms. Therefore, only

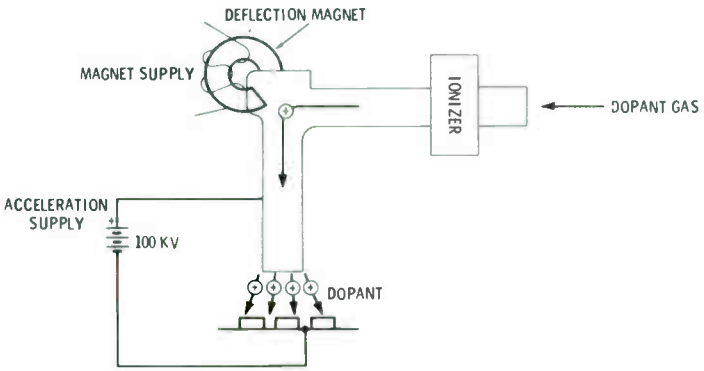


Fig. 3-4. Ion implantation.

the proper mass atoms—the desired atoms—are deflected the correct amount to collide with the silicon bulk in the following step. This *drive-in* step consists of heating the silicon wafers to roughly 2000°F for several hours. During this *drive-in* step the dopant-rich shallow layer is redistributed, thus creating the desired surface concentration and a doped region of greater depth. The surface concentration is an important factor in determining the MOS transistor threshold voltage, while the greater depth will be sufficient to completely contain the n-channel transistor sources and drains.

Like open-tube diffusion, ion implantation utilizes photolithography to selectively implant p-wells in the desired regions.

Deposition

Layers other than silicon dioxide may be *deposited* on the CMOS IC wafers during fabrication. Typically, the deposition process consists of evaporation in a vacuum chamber (for cleanliness) or of chemical vapor deposition in a high-temperature reaction chamber.

All CMOS ICs have a metal, usually aluminum, circuit interconnection layer. This layer is usually evaporated and photolithographically delineated as one of the final IC-fabrication steps.

Layers of silicon oxide may be chemically deposited at temperatures much less than the 2000°F used in thermal oxidation. This chemical deposition process uses the reaction of a silicon-rich gas and oxygen. This reaction is accelerated by the high temperature of the IC wafer and results in deposition on its surface. The resultant layer has less-desirable electrical properties than the high-temperature oxidized layers, but it can be used in applications for which high-temperature steps must be minimized.

Other layers such as silicon nitride are sometimes deposited on the wafers using chemical vapor deposition.

CMOS-FABRICATION PROCESSES

The remainder of this chapter discusses the uses of these basic steps in the five popular CMOS-fabrication processes.

Aluminum-Gate CMOS

The basic fabrication sequence for an aluminum-gate CMOS is shown in Fig. 3-5.

The first step in this sequence is to create the p-well substrate in the n-channel transistor locations. Typically, many n-channel transistors share a common p-well. These large, lightly doped p-type regions are created by using ion implantation masked by photoresist which has been photolithographically defined. This

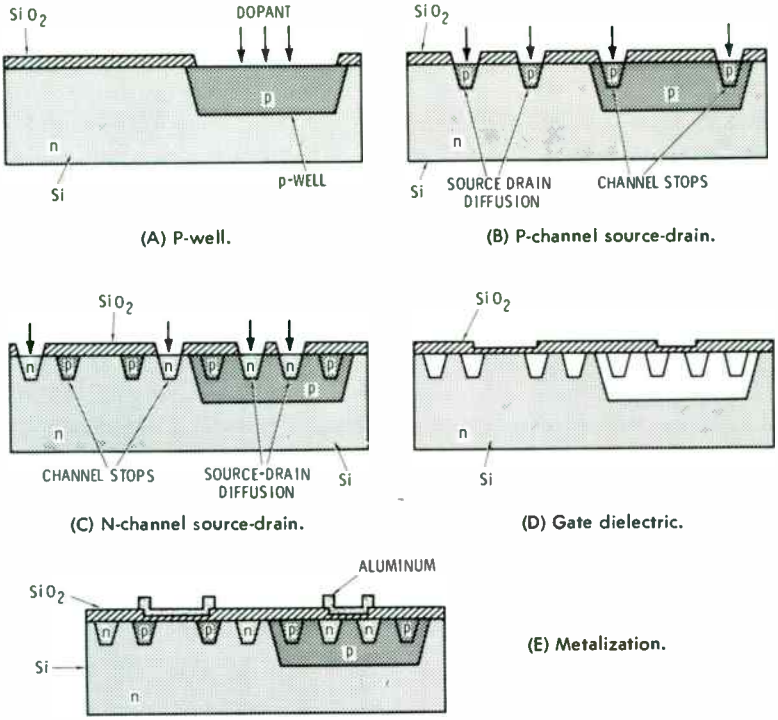


Fig. 3-5. Aluminum-gate CMOS fabrication. (Figures not to scale.)

light-dopant implant is redistributed by a drive-in step. This drive-in step increases the depth of the p-type region so that the region will completely contain the n-channel transistor sources and drains in the completed ICs.

The high-temperature p-well drive-in is formed in an oxygen environment, which results in the wafer surface being oxidized simultaneously with the redistribution of impurities. This silicon dioxide layer is fifteen times thicker than the MOS transistor gate electrodes. In the completed device this thick oxide remains everywhere on the IC surface except where there are junctions or transistor gate-electrode regions. The thick oxide, or field oxide, serves two functions: (1) to establish a minimum metal-to-silicon capacitance by spacing the metal-interconnection layer (final step) as far as is practical from the surface and (2) to minimize as much as is practical the electric fields at the silicon surface that are caused by potentials on these metal interconnections. (These fields, if not minimized, create active, undesired MOS transistors beneath every metal interconnection.)

The next two steps are to create the source and drain regions where n- and p-channel transistors are desired. Two open-tube diffusion steps are used, one for the n-type regions for the n-channel transistors and the other for the p-type regions for the p-channel transistors. These source-drain regions are doped by selectively etching the thick oxide in the desired regions.

The same doped regions used for the MOS transistor source and drain regions are also used as *channel stops*. Channel stops are heavily doped regions (of the same doping type as the substrate in which they are imbedded) used to prevent the formation of parasitic MOS channels beneath metal interconnection regions. Thus, the n-channel source and drain regions may be used as channel-stop diffusion in the n-type substrate to prevent parasitic p-channel transistors.

The next step is to etch the field oxide in the channel region between the source and the drain. The silicon surface is then reoxidized with a thin layer of silicon dioxide (Fig. 3-5D). This thin oxide serves as the MOS transistor gate insulator.

Next, small contact holes are etched through the field oxide. This step is followed by the deposition of an aluminum interconnection layer. This layer fills the contact holes and forms a metallurgical, low-resistance connection with the silicon. The aluminum layer is etched to form gate electrodes; interconnections between gate, source, and drain regions; and bond pads. These bond pads are comparatively large, square regions on the IC to which small wires are attached. These wires serve to interconnect the IC with the package that contains it.

The final step is to deposit a low-temperature (chemical-vapor deposit) silicon oxide layer on top of the IC structure to insulate and protect the IC from external contaminants and to prevent scratching of the IC surface during packaging. Windows are etched through this layer to the bond pads in order to ensure good contact of the package interconnection bond wires.

The final step in CMOS IC fabrication is to test all of the ICs on the wafer for electrical operation (typically only 30% are good at this point). The defective ICs are marked with ink. Next the ICs are scribed, like window glass, and separated by breaking the wafers along the scribe lines. (Lasers are sometimes used for the scribing operation.) The good ICs are selected, packaged, bonded (leads attached), and retested. At this point the fabrication is completed, and the devices are ready to be used.

Silicon-Gate CMOS

The basic fabrication sequence for a silicon-gate CMOS is shown in Fig. 3-6. This process is used to obtain ICs with less

coupling capacitance, lower voltage operation, and higher density interconnection than the ICs obtained with the aluminum-gate CMOS fabrication process.

The lower coupling capacitance results in high circuit speed and low power drain. Low-voltage operation for the silicon-gate CMOS is provided by the low electrochemical work function of this electrode-material type.

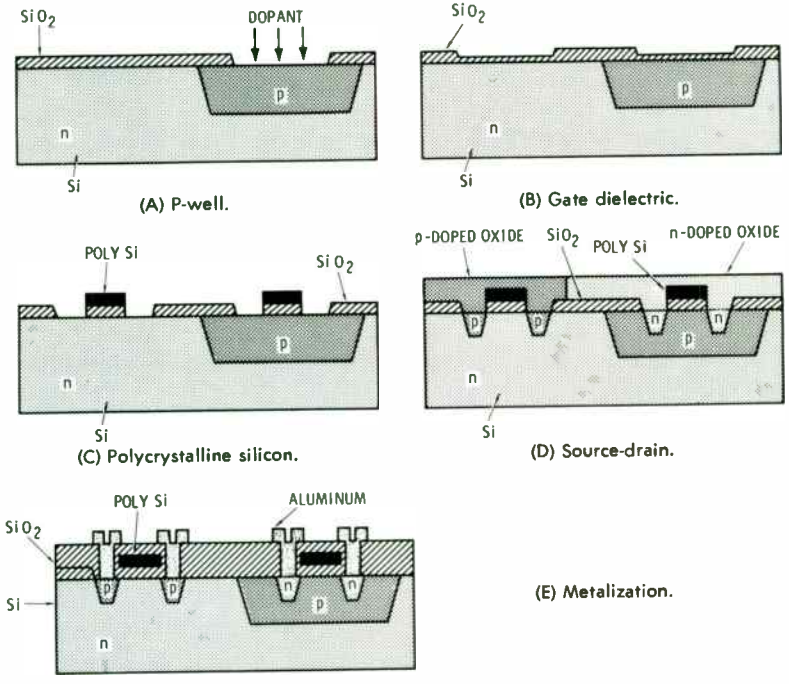


Fig. 3-6. Silicon-gate CMOS fabrication. (Figures not to scale.)

The field oxide and the p-well are fabricated in a manner identical to that previously described for the aluminum-gate CMOS process. Before the source and the drain regions are fabricated, however, the field oxide is etched away in regions where sources, drains, or gate electrodes are to be placed.

The silicon surface is oxidized, which leaves the thin silicon dioxide (SiO_2) layer to be used as the MOS transistor gate insulator. Next a layer of polycrystalline silicon (poly Si) is vapor-phase deposited on the wafer. This layer is then selectively etched, which leaves a polycrystalline stripe on the thin-oxide region. Next, the exposed thin oxide is etched by using the polycrystalline silicon stripe as a mask. (See Fig. 3-6C.)

The wafer is then coated with a chemical-vapor-deposition silicon oxide containing a p-type dopant. This layer is etched selectively; it is not etched on top of regions where p-channel transistors are to be fabricated. A similar layer, but one containing an n-type dopant, is then deposited (Fig. 3-6D). Next in one high-temperature step, the source and the drain regions of both n- and p-channel transistors are created. In addition, the polycrystalline silicon stripe between the sources and the drains is also doped into a low-resistance state. This conductive polycrystalline silicon electrode is the namesake of this technology. The low gate-drain capacitance of this structure results from the polycrystalline silicon stripe acting as a mask in the source-drain doping step and thus removing the gate-drain and gate-source overlap, which are required for ease of alignment of the gate mask in the aluminum-gate CMOS fabrication process. A fabrication process that uses the gate electrode as a mask for the source/drain positioning is called a *self-aligned* CMOS process. Note that aluminum cannot be used as the gate electrode in the self-aligned process because it has a melting point below the dopant drive-in temperatures.

The final assembly steps in the silicon-gate process are: to etch contact holes in the covering-oxide layers, to deposit an aluminum interconnection layer, to selectively etch the aluminum, and to package the ICs. (This last step is identical to the corresponding step discussed in the last section.)

Silicon-gate CMOS is needed where high-speed operation and/or low-voltage operation (less than 3.6-V supplies) are desired at low cost. This technology is typically not required for the assembly of simple logic functions, such as NOR/NAND gates (an aluminum-gate CMOS is typically used), because the on-chip circuitry is relatively simple for these devices and therefore does not benefit much from small (subpicofarad) on-chip capacitances or high-density interconnection. The large package and circuit capacitances (many picofarad) typically encountered overshadow the benefits obtained from their low-capacitance CMOS technology. However, in more-sophisticated circuits, such as counters and shift registers, the internal propagation delays play a more dominant role in determining speed. For these devices the silicon-gate CMOS has significant advantages. Moreover, the silicon gate provides a second layer of circuit interconnection which makes possible the construction of significantly more complex circuits in a given silicon surface area (more bits of storage in a shift register IC for example) than is possible with ICs using only one layer of interconnection.

The lower work function of the silicon gate provides lower-threshold-voltage MOS transistors and makes possible the opera-

tion of CMOS ICs with 1.2-volt supplies for miniature, low-power applications, such as electronic wristwatches. Not all silicon-gate CMOS ICs are capable of 1.2-volt operation, but they are almost universally capable of operation with lower supply voltages than similar aluminum-gate CMOS parts.

Some silicon-gate CMOS processes use the source-drain diffusions as channel stops also, and some do not. If supplies less than 5 volts are used, the thick-oxide parasitic MOS transistors will never be turned on. (These thick-oxide devices typically have a 7- to 10-volt threshold with no channel-stop diffusion.) If operation above 5 volts is specified, the silicon-area wasteful channel-stop diffusions must be used. In order to maintain high packing density, some manufacturers use a special ion-implantation doping step after the source-drain doping to create a lightly doped channel stop everywhere there are no transistors. The high packing density obtainable with silicon-gate technology would be compromised if it weren't for this implant step.

Oxide-Isolation CMOS

Oxide-isolation CMOS is an improved type of silicon-gate CMOS. The improvements claimed for this technology are higher speed, higher circuit density, and improved manufacturing as compared to basic silicon-gate CMOS. Some of the advantages of this technology are: the density, the ease with which the channel-stop regions may be implemented, and the flat topology of the IC surface just prior to metalization (which minimizes yield-reducing metal breakdown and cracks at the edges of windows in steep oxide cuts).

Although there are several *oxide*-isolation processes (such as the Isoplanar, Locos, and Oxim processes) which have become associated with specific semiconductor manufacturers' techniques, for the purposes of this discussion, these processes do have enough in common to describe them as if they were identical. A representative oxide-isolation process is shown in Fig. 3-7.

Fig. 3-7B shows the first variation from the silicon-gate CMOS process—the deposition of the highly inert silicon-nitride layer after the p-well is implanted. The channel-stop doping (an n-type diffusion everywhere the silicon nitride is not masked) is done next. The thin oxide beneath the silicon nitride is then etched back from the edges. In the following field-oxidation step, the edges of the field oxide (eventually the edges of the transistor source and drain) are determined by the edges of this undercut thin oxide. During the field-oxide growth, the silicon-nitride lip at the edge is bent back by the growing oxide layer, and the edge of the transistor region is spaced away from the channel-stop diffusion by

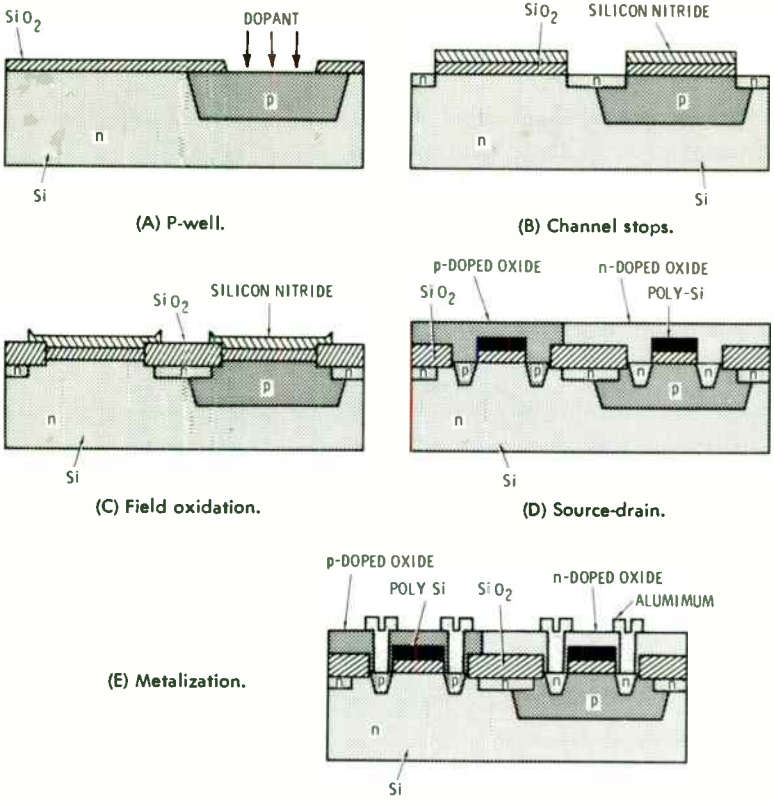


Fig. 3-7. Oxide-isolated CMOS fabrication. (Figures not to scale.)

the amount of thin-oxide undercutting. This spacing is necessary to prevent source-drain conduction of the n-channel transistor and to ensure high breakdown voltage of both n- and p-channel transistors.

Once the field oxide is grown, the nitride layer is etched away. What is left is a structure similar to the silicon-gate-process devices after the thin oxide layer has been grown. The principal differences in the structure are that the channel-stop doping is performed in a "self-aligned" doping step and that the field oxide is grown only in the regions in which it will remain in the completed device. This selective field growth results in a smaller step height between the top of the field oxide and the transistor thin-gate oxide. This oxidation converts silicon to silicon dioxide, resulting in a step in the silicon surface at the edge of the nitride layer. Thus, the selective consumption of silicon from the un-

protected surface being oxidized results in the "sunken" field oxide.

The assembly techniques after the silicon-nitride layer is removed (to reveal the thin oxide beneath) are virtually identical to the final silicon-gate CMOS assembly steps. They consist of silicon-gate deposition, delineation, and doping, followed by aluminum metalization, metal delineation, and packaging.

It is interesting to note that the transistors are built on "hills" of silicon formed by the "valley" created by the field-oxidation process. These oxide-isolated hills comprise the structure from which the devices obtain their name. Thicker field oxides may be used (for a given metalization-step coverage capability) for the oxide-isolated devices than may be used for the conventional silicon-gate devices discussed in the last section. Thus, circuits made with this process have less capacitance than those made with the silicon-gate process discussed previously.

One circuit feature offered by some manufacturers of oxide-isolation CMOS is *double-buffer* outputs, even on simple NOR, NAND, and inverter (NOT) circuits. Some manufacturers currently offer the oxide-isolated CMOS ICs in the simple functional blocks usually reserved for aluminum-gate CMOS and utilize the high packing density capability of this technology to incorporate an extra internal buffer cascaded with each output. These *double-buffered* output devices have more circuit gain, which simplifies logic-waveshape restoration. Double buffering aids in many circuit applications that do not have ideal square-wave inputs.

Silicon-on-Sapphire CMOS

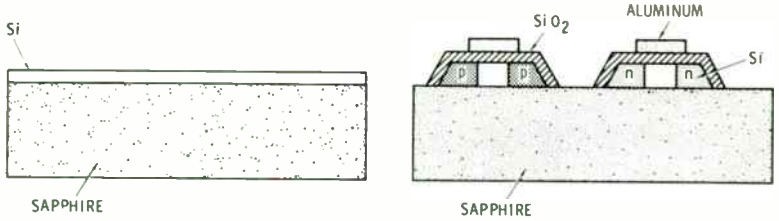
Silicon-on-sapphire CMOS represents a dramatic departure from the CMOS-fabrication techniques discussed thus far. The unique characteristic of this process is that the completed ICs have extremely low parasitic capacitances. This technology uses a special starting wafer which consists of a thin silicon layer deposited on an insulating sapphire substrate. The deposition of crystalline silicon is possible because the crystal-lattice spacing of sapphire (aluminum oxide) is nearly identical to that of silicon.

The basic fabrication sequence is outlined in Fig. 3-8. This sequence is similar in many ways to the previous assembly sequence, with the following two notable exceptions:

1. The source-drain doped regions extend down to the insulating substrate, which eliminates most of the junction capacitance associated with these regions.
2. The transistors are isolated by chemically etching away the silicon in regions that would otherwise be covered with field

oxide. Since an insulator is beneath the metal interconnection layer, the channel will stop diffusions.

Thus, the transistors on the completed devices resemble discrete ideal MOS p- and n-channel transistors attached to an insulating substrate. This technology offers fast on-chip circuit speed because of the low junction capacitances. However, the primary speed limitation of this CMOS technology is (as it is with most CMOS technologies) the output-stage speed. The 15-picofarad load capacitances that the output stage is typically required to drive are hundreds of times greater than on-chip capacitances. Thus, in order to benefit most effectively from the advantages of this technology, the IC must require high-speed circuitry on the IC chip and yet have only low-speed output waveforms. A decade-counter chip is an example of a circuit with these mixed speed requirements.



(A) SOS starting wafer.

(B) Fabricated IC cross section.

Fig. 3-8. Silicon-on-sapphire fabrication. (Figures not to scale.)

Dielectric Isolation

Dielectric isolation is a fabrication technique that incorporates many insulated substrate properties into a structure fabricated with a conventional silicon wafer as the starting material.

As shown in Fig. 3-9, the process begins by etching the silicon wafer and leaving transistor islands. The wafer is then oxidized and coated with a thick layer of polycrystalline silicon.

This wafer is mechanically lapped to remove all of the silicon wafer except for the transistor islands imbedded into the oxide and into the polycrystalline silicon coating. These islands of silicon fragments are embedded in the silicon oxide insulator which, with the polycrystalline silicon backing, serves as an insulated substrate and secures the devices in place during the remainder of processing and packaging. High-quality n- and p-channel transistors are fabricated in each of the *dielectrically isolated* fragments. Integrated circuits built this way have many of the speed and low-capacitance properties of the sapphire-insulated devices.

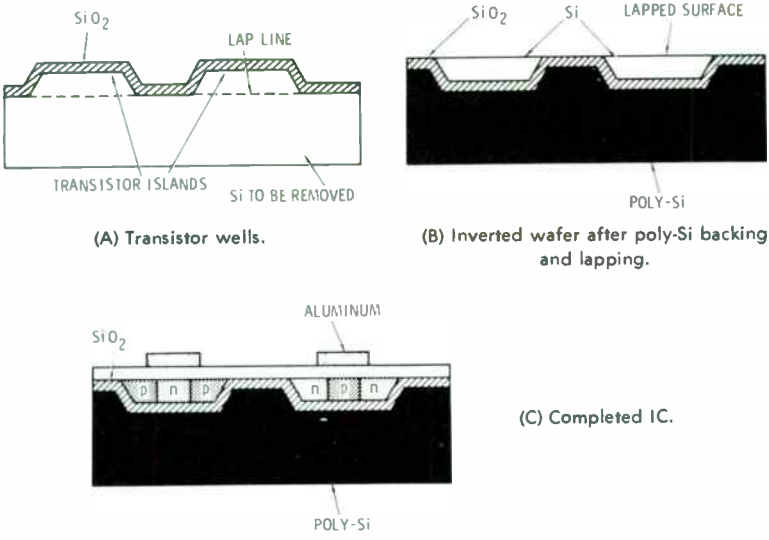


Fig. 3-9. Dielectric-isolated CMOS fabrication. (Figures not to scale.)

There are two additional desirable properties of the dielectrically isolated and insulated substrate devices. The first property is that of significant resistance to circuit malfunction during exposure to radiation (X-rays for example.) The small junction area associated with these technologies reduces the importance of the junction leakage currents which result from the incident radiation.

The second desirable property is that of complete electrical isolation of the transistor. This eliminates the troublesome parasitic SCRs associated with the junction-isolation techniques used in other technologies described in this chapter.

Many ICs are currently available fabricated with these insulated substrate technologies. Some are pin-for-pin replacements with devices of the other technologies. This allows for device-technology substitution during circuit troubleshooting. If, for example, it is thought that circuit operation is being disturbed by parasitic SCRs, an IC fabricated with an insulated-substrate technology may be substituted for comparison purposes.

In summary, this chapter covers the basic CMOS IC assembly techniques and their variations. In addition, the effects of these variations on circuit performance are discussed. The next chapter covers the design of the circuits used in these ICs and is based on the technology presented in this chapter, the device operation discussed in Chapter 2, and the desired ICs introduced in Chapter 1.

4

Monolithic CMOS Circuit Design

The preceding chapters have covered background topics that are useful for the understanding of CMOS IC circuitry. This chapter focuses on the circuitry integrated on the silicon IC die. The following chapter covers the topics associated with the interconnection of ICs with other components (such as resistors, capacitors, and other ICs).

CMOS IC COMPONENTS

Integrated-circuit design is limited by the components available as part of the IC structure. The basic components for CMOS ICs are:

1. Resistors
2. Capacitors
3. N-channel MOS transistors
4. P-channel MOS transistors
5. SCRs
6. Grounded-base pnp transistors
7. Grounded-base npn transistors
8. Grounded-collector npn transistors

These basic components are shown schematically in Fig. 4-1.

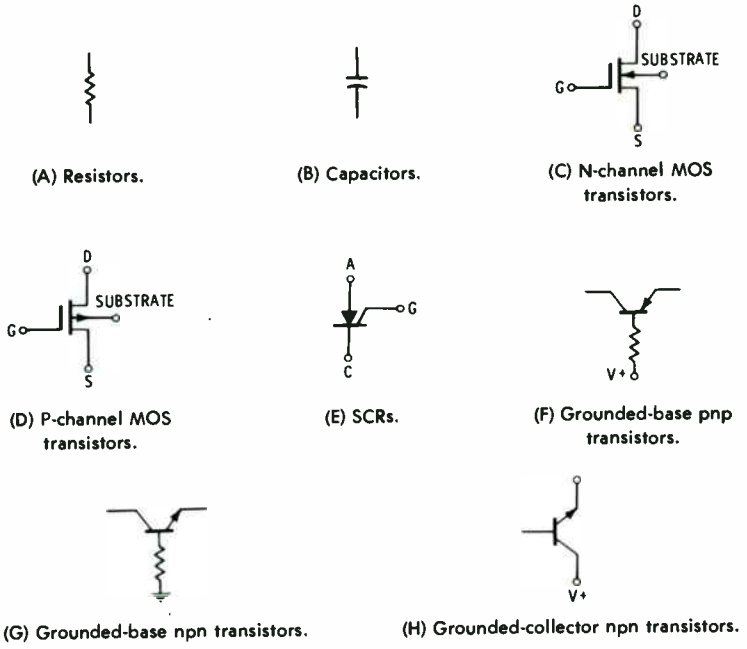
CMOS IC Resistors

The resistors available in the CMOS IC structure are:

1. N-channel transistor source-drain region resistors (typically 50 ohms per square)

2. P-channel transistor source-drain region resistors (typically 200 ohms per square)
3. P-well resistors (typically 4000 ohms per square)

The resistance of these resistors is a function of the doping parameters and the geometry of the resistor on the silicon surface. The rule of thumb for resistor design is that the resistance equals the ratio of resistor width to resistor length, multiplied by the



(A) Resistors. (B) Capacitors. (C) N-channel MOS transistors.

(D) P-channel MOS transistors. (E) SCRs. (F) Grounded-base pnp transistors.

(G) Grounded-base npn transistors. (H) Grounded-collector npn transistors.

Fig. 4-1. CMOS IC components.

ohms per square parameter of the region. A region 50 times longer than its width, for example, has a resistance 50 times the ohms per square of the region. (For a p-channel source-drain diffusion, this would be a resistance of 10,000 ohms.)

The resistors in CMOS IC designs must often be included in the IC circuit analysis not because they play a desired useful role, but because they present an often undesired voltage drop in long circuit bus lines.

CMOS IC Capacitors

The CMOS IC capacitors sometimes have a useful function and sometimes have a detrimental function in the performance of

CMOS circuits. These capacitances result from the reverse-biased diode capacitance of the MOS transistor source and drain region and from the capacitance between the metal interconnection lines and the silicon surface regions.

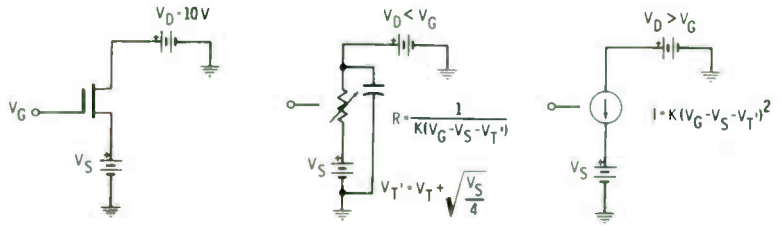
The useful function of these capacitors is the temporary storage of signals. These capacitors may, in some circuits, be charged or discharged to record a logic "one" or "zero."

The undesired function of these capacitors is the reduction of switching speed. They limit the circuit speed because they are charged and discharged through MOS transistors and resistors which limit the passage of current.

CMOS Transistors

The physical properties of the n- and p-channel transistors used in CMOS ICs are discussed in Chapters 2 and 3. For the purposes of this chapter, the basic equivalent circuit shown in Fig. 4-2 will be assumed. Note that a voltage greater than the threshold voltage, V_T, must be applied to the transistor (V_T is assumed to be negative one volt for the p-channel and positive one volt for the n-channel transistors) for source-drain current to flow. The transition from open circuit to R_o occurs gradually as the gate-source voltage is increased from the threshold voltage to the 10 volts shown. The capacitor shown is associated with the reverse-biased drain-region diode.

It is important to note that the 1-volt threshold is only applicable for the transistor with the source grounded to the substrate. For example, assume that the source is reverse biased to positive one volt with respect to the substrate for the n-channel transistor. Also assume that a voltage sufficient to just establish source-drain conduction is applied to the gate. This voltage will be greater than two volts because of the back-gate bias effect discussed in Chapter 2. (The gate-source threshold voltage has increased.) This increase in threshold voltage due to back-gate bias is approximately



(A) Circuit.

(B) Voltage-controlled-resistor equivalent circuit ($V_D < V_G$).

(C) Current-controlled equivalent circuit ($V_D > V_G$).

Fig. 4-2. CMOS.

$$V_{BG} = \sqrt{\frac{V_{source\text{-}substrate}}{4}}$$

Unfortunately, very few CMOS circuit designs utilize circuitry having the source connected to the substrate, which would eliminate back-gate bias as a design consideration. The digital gates and transmission gates discussed later in this chapter are frequently encountered examples of biased-source circuits.

CMOS SCRs and Bipolar Transistors

Rarely do the SCRs and transistors available in the CMOS process play a useful role. They are typically constrained by the IC structure so that one or more terminals are grounded or connected to the power supply, which renders them useless in circuit applications. These devices can have detrimental effects on the IC performance, as discussed in Chapter 2. These effects can be significant because of the noteworthy parameters of some of the devices. (The grounded-collector npn transistor typically has a common-emitter current gain of several hundred.)

CMOS IC DESIGN RULES

CMOS ICs are designed to provide economical high-switching-speed circuitry with low power drain. A set of simplified design rules are:

1. Use minimum-size transistors wherever possible to maximize the number of components and circuit functions available on the IC chip.
2. Use designs with no direct current path from the power supply to ground in order to minimize IC power drain.
3. Use large driver transistors, with low on-resistance, for IC output stages when switching IC outputs rapidly with the large circuit capacitances (15 picofarad) typically encountered off the chip.

The remainder of the chapter discusses specific circuit examples. These examples demonstrate how the CMOS components and design rules are used to realize a representative group of basic building-block circuits.

CMOS INVERTERS

The basic CMOS inverter used in the CD4007 IC listed in the appendix is shown schematically in Fig. 4-3.

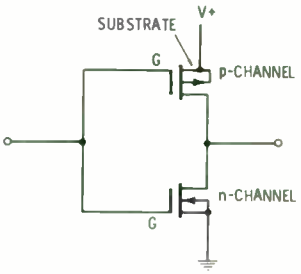


Fig. 4-3. CMOS inverter.

The operation of this circuit is relatively easy to visualize. When the input is 0 volts, the p-channel transistor gate-substrate bias (the p-channel, n-type substrate is connected to the +5-volt supply) is −5 volts and the transistor (having a −1-volt threshold) is turned on (low resistance). The n-channel transistor gate is biased to 0 volts and is turned off (high resistance). Thus, the output is connected to the 5-volt power supply through the p-channel transistor source-drain resistor.

When the input is +5 volts, the p-channel transistor is turned off and the n-channel transistor is turned on. This interconnects the output to the power-supply ground (negative) terminal.

Thus, the output terminal is connected to the power supply or to ground depending upon the input voltage. This circuit is called an inverter (also a NOT gate) because the output voltage is the logical complement of the input voltage.

The switching speed of this circuit between logic states is determined by the "on resistance" of the CMOS switches and by the capacitance of the IC and load circuitry. The switching time is given by the capacitor charging equation; that is,

$$t_{switching} = 2.2\, R_{on} C_{load}$$

where,

$t_{switching}$ is the switching time in seconds,
R_{on} is the on resistance in ohms,
C_{load} is the load capacitance in farads.

(The on resistance for typical off-chip driver transistors may be as low as a few hundred ohms.) This switching time for CMOS inverters can be as short as ten billionths of a second (10^{-8} seconds) for the fastest silicon-on-sapphire CMOS ICs. However, this basic inverter switching time typically limits the operation frequencies of common low-cost CMOS ICs to a few megahertz. (One megahertz is the same as 1,000,000 cycles per second).

The on resistance of the inverter switching transistor is dependent upon the supply voltage because these transistors are

biased by the supply voltages. Greater voltages decrease the on resistance and thereby increase the maximum switching speed. The highest CMOS switching speeds are obtained with the maximum-allowed supply voltage. The on resistance of the simple inverter switching transistors is shown as a function of supply voltage for typical CMOS ICs in Fig. 4-4. It can be seen in this figure that the on resistance is roughly inversely proportional to the supply voltage. (Doubling the voltage roughly halves the on resistance.)

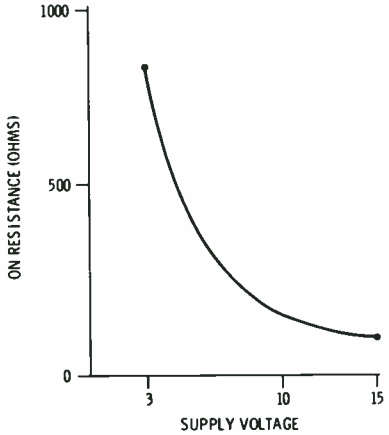


Fig. 4-4. CMOS transistor "on resistance."

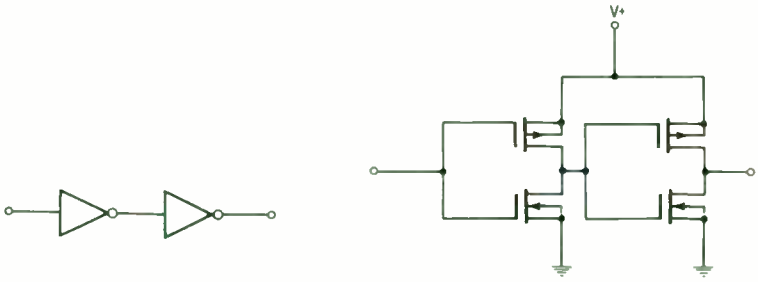
Inverter Fanout

An important parameter of digital ICs is *fan-out*, the number of ICs that may be driven from the output of the IC in question. The fan-out of CMOS ICs is very large due to the fact that CMOS ICs are driven to control their logical input states and therefore require virtually no continuous current at their inputs. Other types of digital ICs (such as transistor-transistor logic) may require that a few milliamperes be switched at the IC inputs, thus limiting the number of ICs which may be driven by a device with a specified current-handling capability. For CMOS ICs the practical fan-out is limited by the total capacitance presented by the inputs of the CMOS ICs being driven. Larger capacitances reduce the switching speed which may be obtained at the output of the driving CMOS IC.

Double-Buffered Inverters

Some IC inverters are double buffered, as shown in Fig. 4-5. This double buffering increases the gain of the inverter for input voltages between the logical-one threshold and the logical-zero threshold. For a range of voltages between these thresholds at the

input of a CMOS inverter, the output voltage is between logical-zero and logical-one voltage levels. Increasing the inverter gain narrows this range of voltages at the input and results in faster ICs. (During input transitions the output is turned "on" for a large fraction of the time if the inverter gain is large.) Increasing the inverter gain also results in ICs that are less sensitive to input-signal waveshape variations. (A triangular input waveshape will be converted to a rectangular waveshape by a high-gain double-buffered inverter.)



(A) Logic diagram. (B) Representative circuit diagram.

Fig. 4-5. Double-buffered output.

Inverter Noise Immunity

Undesired switching transients are sometimes superimposed on logical signals and power-supply voltages in digital circuits. These transients may have several causes: sending the logic signals through long multiconductor cables, providing poor ground connections, placing the logic circuitry near high-power circuits, etc. Noise immunity for logic circuits is often expressed in terms of the circuit input voltage required to cause a logic change at the circuit output.

As discussed in Chapter 1, the logical one and zero are defined in terms of voltages greater than 70% (logic 1) and less than 30% (logic 0) of the supply voltage. At some point between these defined logical transition points, the CMOS inverter output will change with changes in input voltage. The noise-voltage immunity is the voltage required to establish the input at this transition point, divided by the IC power-supply voltage. For the simple CMOS inverter circuit, the noise-voltage immunity is nominally 50% of the supply voltage.

Inverter Power Consumption

So far, little has been said about the power consumption of CMOS inverters. Since one inverter transistor is always on and

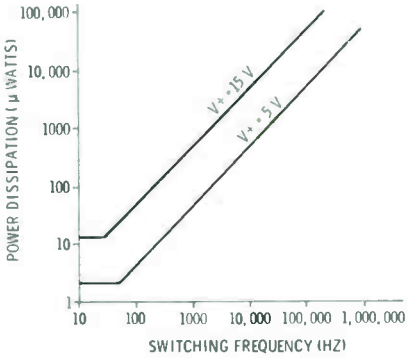


Fig. 4-6. CMOS inverter power consumption.

one always off, there is no direct path of current to ground. The charging of the load capacitors, however, requires energy. This energy is dissipated by the transistor switches which charge and discharge the load capacitors. The amount of power consumed is a combination of two things: the power consumed due to capacitor charging and discharging, and the power consumed in supplying a reverse-biased diode leakage current to the reverse-biased source-drain and p-well regions of the CMOS IC. While the leakage-current-related power is very small, it represents the principal source of power consumption in most CMOS ICs when no switching operations occur. The power dissipation of a typical CMOS inverter is shown in Fig. 4-6. The power dissipation shown in this figure may be expressed by:

$$\text{Power Dissipation} = I_1V + CV^2F$$

where,

I_1 is the leakage currents in amperes,
V is the supply voltage, in volts,
C is the load capacitance in farads,
F is the switching frequency in hertz.

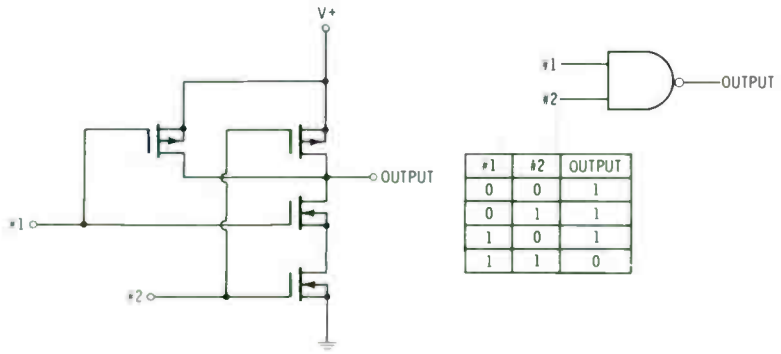
From this expression and the figure, we can see that the power dissipation is dependent upon the power-supply voltage. Large voltages result in greater dissipation even though there is no direct path to ground through the CMOS switches. At high switching speeds, the power increases four times for every doubling of the supply voltage.

From the standpoint of low power consumption, using the minimum supply voltage is desirable. (Typical CMOS ICs may be operated with supply voltages between 3 volts and 15 volts.) However, many applications require that higher voltage supplies be used to meet noise immunity and switching speed requirements.

The considerations for switching speed, fan-out, noise immunity, and power consumption of the simple CMOS inverted used in monolithic CMOS circuit design have been presented. Many of the considerations for the on-chip inverter design extend through all on-chip CMOS design and some off-chip CMOS circuits. The capacitances encountered in the on-chip monolithic CMOS circuits are often 100 times smaller than those encountered in off-chip capacitances. This allows much smaller transistors (using less silicon-surface area) to be used in internal circuitry than are used in the IC output stage to interface the IC with external circuitry.

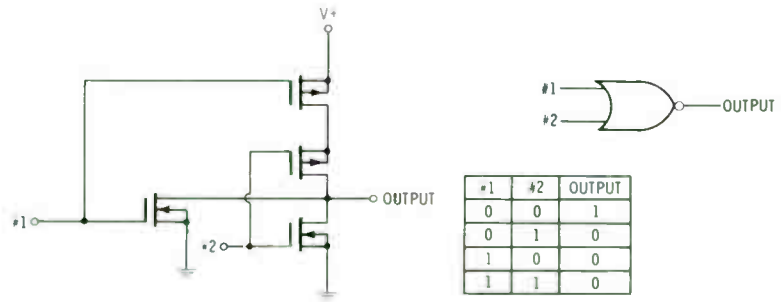
CMOS DIGITAL GATES

The CMOS inverter just discussed is the simplest member of the family of logic gates. The basic members of this family are NOT gates, AND gates, and OR gates. Many times these simple cir-



#1	#2	OUTPUT
0	0	1
0	1	1
1	0	1
1	1	0

(A) NAND gate.



#1	#2	OUTPUT
0	0	1
0	1	0
1	0	0
1	1	0

(B) NOR gate.

Fig. 4-7. CMOS two-input gates.

cuits provide combinations of these logic functions, such as NAND (NOT-AND) and NOR (NOT-OR) gates.

The basic logic functions of these gates are discussed in Chapter 1. Generally more than one input determines the logic output of the gates. The schematic circuit diagram and truth table for a two-input CMOS NAND gate and a two-input NOR gate are shown in Fig. 4-7. The operation of these circuits may be visualized by using the voltage-controlled-switch concept. The NAND gate output, for example, is equal to the supply unless both inputs are logic one, in which case the appropriate switches are activated, thus connecting the output to the ground terminal.

The principles relating to circuit performance that were discussed in the inverter section may be applied to the CMOS gates as well.

CMOS TRANSMISSION GATES

The CMOS transmission gate performs a floating-switch function not often used in bipolar-transistor logic circuits. The basic transmission-gate circuit is shown in Fig. 4-8.

The operation of this circuit is straightforward. The input is connected to the output via the double switch pair by the application of complementary bias to the transistor gates. (The p-channel transistor gate is grounded, and the n-channel transistor gate is connected to the supply.)

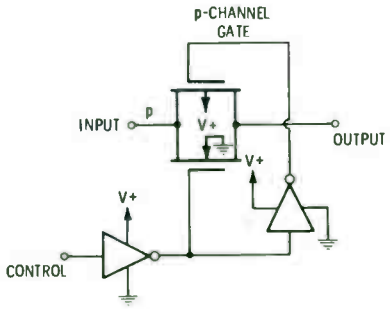


Fig. 4-8. CMOS transmission gate.

The transmission gate is an important example of a CMOS circuit in which the MOS transistor sources are not connected directly to the substrate. (The CMOS NAND gate is another.) As presented in Chapter 2 and reviewed briefly in the first part of this chapter, the floating source results in an increase in the source-drain resistance of that transistor for two reasons: (1) the gate-source bias is typically decreased by the amount of source-

substrate voltage and (2) the threshold voltage of the MOS transistors increases with an increase in the source-substrate voltage. (A four-volt bias on the source can increase the gate-source threshold voltage by one volt or can increase it to a composite threshold of two volts for a typical transistor.)

This increase in threshold with source voltage is called the back-gate bias effect and is related to the reason for using two transistors in place of one for the transmission-gate circuit function. The n-channel transistor in the transmission gate increases drastically in resistance as the transmission gate nears the supply voltage. In fact, for an n-channel transistor with a 1-volt threshold, the resistance will approach infinity (open circuit) as the source voltage approaches 12 volts with a 15-volt supply on the gate. Fortunately the p-channel operates in a complementary fashion. The substrate of the p-channel transistor is connected to the supply, and this transistor with its gate connected to ground has lowest source-drain resistance for input signals near the supply. (Of course, the p-channel transistor has large source-drain resistances for input signals near ground.) Thus, the complementary n-channel and p-channel pair combine to overcome the limitation imposed by floating sources. Since the transmission resistance is minimized for all amplitudes of input signal, the speed of charging capacitances connected to the output is maximized, as is the output voltage swing.

CMOS SHIFT REGISTERS

The shift register is a serial storage element. Logic signals enter at one end of the structure, and appear at the output located at the other end, at some later time that is determined by the shift clock rate applied to the register. There are two basic classes of shift register circuits: *dynamic shift registers* and *static shift registers*.

Dynamic Shift Registers

The two classes of shift register circuits differ in the minimum frequency of the shift clock. Dynamic shift registers must be clocked at rates greater than 1 kHz typically, while static shift registers may be clocked as slowly as desired. Dynamic circuits usually occupy much less silicon area than static circuits. This allows more shift register stages to be placed on an IC die (more information storage) at a lower cost per function.

The circuit of the single CMOS dynamic shift-register cell is shown in Fig. 4-9. This circuit is a combination of the simple CMOS inverter, CMOS transmission gate, and CMOS capacitor.

In operation the input signal is passed through the turned-on transmission gate and is applied to the CMOS storage capacitor. The input transmission gate is turned off, which leaves the inverted input signal applied to the input storage capacitor. The output transmission gate is turned on by the complement of the signal that turned the input gate off. The stored input signal is transferred to the output storage capacitor and is seen at the output of the second inverter. This circuit provides a delay of one clock period and stores one logic bit of information. In a typical shift-register IC, many of these cells are interconnected to provide many bits of storage and greater delay.

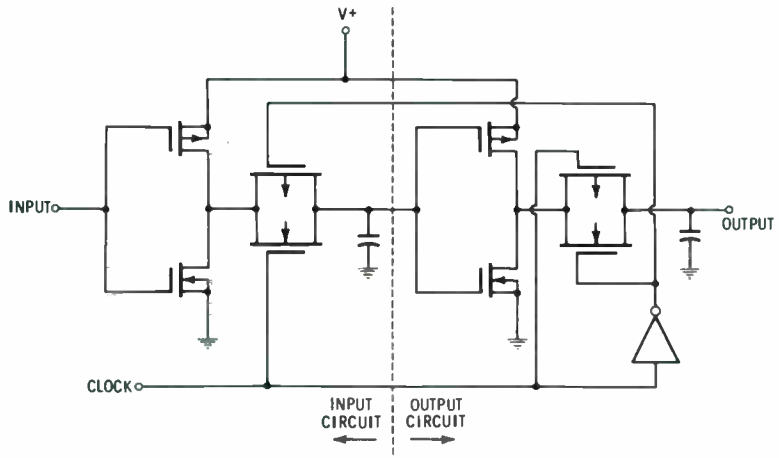


Fig. 4-9. CMOS dynamic shift-register stage (input and output buffers, one each per IC package, not shown.)

The minimum clock rate that may be used with dynamic shift registers is determined by the leakage current associated with the drains of the transmission gates connected to the storage capacitor. (Typically the reverse-biased capacitance of the drains provides the required capacitance, and no separate capacitor structure is used.) This leakage current discharges the storage capacitors, which can result in the destruction of the stored logic information. For example, at low-frequency clock rates, the capacitor may discharge below the inverter-logic-transition threshold, thereby disrupting the desired circuit operation. With each clock operation, however, the gain of the inverter restores the logic levels to their maximum (supply voltage or ground for logic one and zero, respectively) if the clock operation is performed before the storage-capacitor leakage discharges the capacitor below the logic-transition point (50% of the power-supply voltage).

Static Shift Registers

The static shift register has no minimum clock-rate restriction. These shift registers are used in applications where the clock rates are below the 1-kHz minimum dynamic shift-register clock rate.

A schematic diagram of a simple static shift register is shown in Fig. 4-10. It is interesting to note that this circuit is identical to the dynamic shift-register circuit, except that an additional inverter and transmission gate are connected to both the input and the output stages. This additional circuitry converts the input and output stages into reset-set flip-flops when information is being stored. It also switches the circuit and leaves the basic dynamic shift-register circuit operative during the time when the stored information is charged.

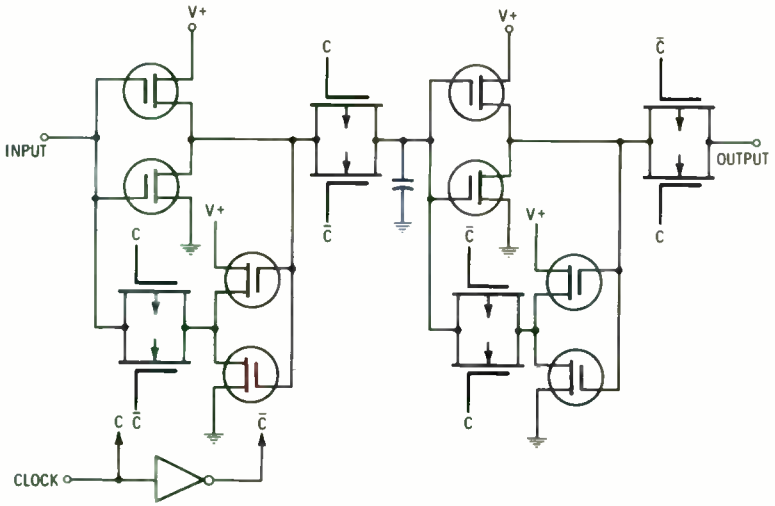


Fig. 4-10. CMOS static shift-register stage (input and output buffers, one each per IC package, not shown.)

The static shift-register ICs commercially available typically contain fewer bits of storage than dynamic shift registers of the same price (due to the increased silicon area required by the greater circuit complexity). Many of the static shift-register ICs offer parallel output terminals and input terminals from each storage bit. (Few dynamic shift-register ICs offer this.)

Thus, the design alternatives between dynamic and static shift registers are determined by clock frequency, cost, and input/output configuration.

CMOS MEMORIES

Memory ICs are used to store large quantities of digital information (1's and 0's). This information is typically stored in two-dimensional cubbyhole fashion. An individual storage point is addressed by entering row and column binary addresses to the IC. The information stored in the addressed cell may be *read* at the IC output or may be changed by *writing* the desired signal at the memory IC input.

Many memory ICs use only n- or p-channel transistors in order to obtain the largest number of storage points in a given IC at low cost. However, there are several CMOS memory ICs that take advantage of the extremely low power drain of CMOS when no memory points are being selected.

The simplified CMOS memory cell shown in Fig. 4-11 demonstrates this low-power operation. This circuit is a simplified version of the CMOS reset-set flip-flop described in Chapter 1. The

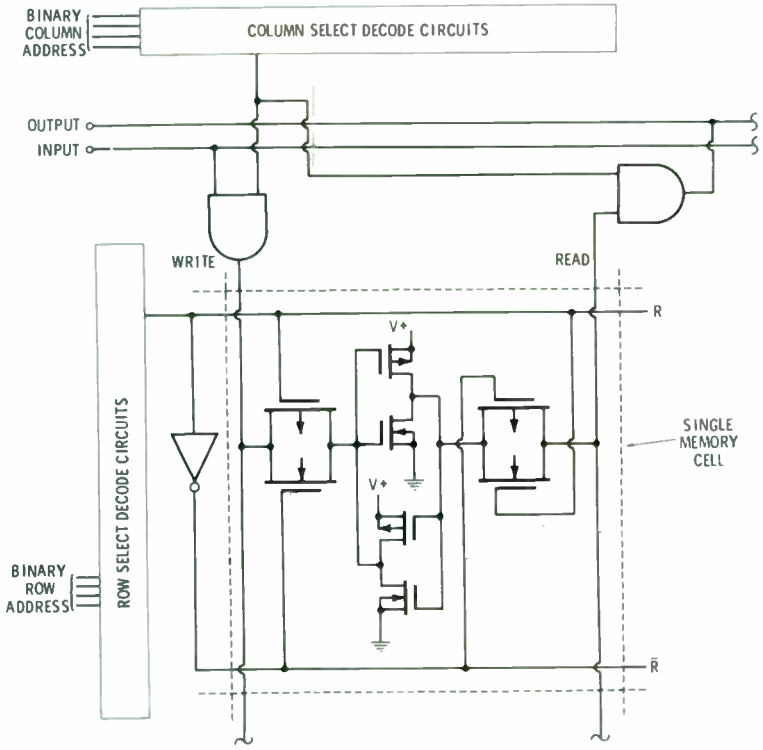


Fig. 4-11. CMOS memory.

cross-coupled structure is addressed to the bus lines through two single-transistor transmission gates discussed earlier in this chapter. The transmission gates of a given row in the square memory array are all turned on simultaneously. Thus, all the information stored in a given row is available on the read/write bus lines. The selected row is determined by the binary row address and the decoder (which converts the binary code to a single-row select command). The column address is decoded to select one of the read/write bus lines and to make it accessible to the IC output through a column-select transmission gate and inverter buffer.

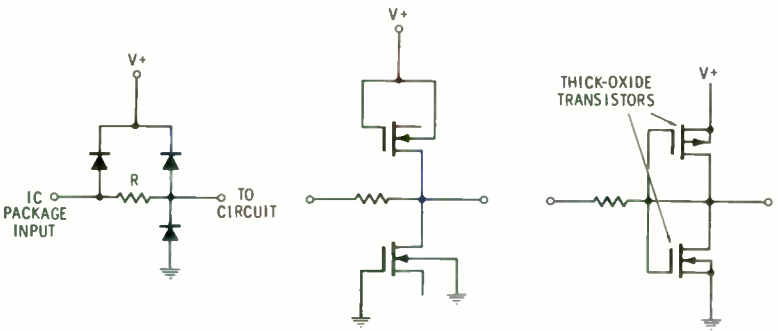
The low standby-power operation of this circuit is apparent after you note that no direct current paths exist from the power supply to ground. The principal sources of power consumption are reverse-biased diode leakage currents and charging and discharging circuit capacitances (during read/write and address-selection operations).

CMOS INPUT PROTECTION

Input-protection circuitry is commonly provided on the chip at all CMOS IC input points. This circuitry is designed to prevent static charge from destroying the MOS input transistors during handling and to prevent voltage transients from destroying the input transistor during operation.

This protection circuitry is provided because the breakdown voltage of the MOS transistor gate-to-semiconductor dielectric (gate silicon oxide) is typically 70 volts. The gate-insulator breakdown is usually catastrophic and results in the destruction of the associated MOS transistor.

Many CMOS ICs use the simple diode-clamp input circuit shown in Fig. 4-12A. These diodes are forward biased whenever



(A) Simple diode clamp. (B) Gate-controlled diodes. (C) Thick-oxide transistors.

Fig. 4-12. CMOS input-protection circuits.

the IC input signal is greater in voltage than the IC supply or less in voltage than the IC ground.

The gate-protection circuit shown in Fig. 4-12B contains MOS transistors connected as gate-controlled diodes. As we discussed in Chapter 2, semiconductor diodes are used because they may be "broken down" without harmful effects if they are kept within power dissipation ratings (i.e., if they do not excessively overheat). The breakdown voltage of these diodes is typically designed to be much less than 70 volts. This is done by using metal gate electrodes on the oxide covering the surface perimeter of the junction in order to assist the diode-breakdown process at the semiconductor surface.

Another protection circuit shown in Fig. 4-12C, uses thick-oxide parasitic MOS transistors. These transistors are fabricated by using the field oxide as a gate insulator. (These transistors typically have a turn-on threshold of 20 volts.) The gate and drain of the MOS transistors are interconnected and connected to the IC input, and the source of each transistor is connected to the substrate. Input overloads turn on the protection transistor which shunts current to ground.

All of these circuits use diodes that may stimulate the undesired parasitic devices discussed in Chapter 2. These undesired devices will not be activated if the CMOS ICs are operated within the system design rules discussed in Chapter 5. Some interface ICs, such as the 4049 and 4050, use only half of these input-protection circuits. This allows them to be operated properly with input voltage greater than the IC supply (but not with voltages less than ground).

5

CMOS Circuit and System Design

The first four chapters of this book cover reference material on the topics of digital electronics, semiconductor devices, fabrication, and CMOS IC design. The understanding of CMOS ICs presented in these chapters will prove valuable in the design and debugging of circuits and systems using CMOS ICs.

This chapter covers basic input circuit, output circuit, power supply, interfacing, and debugging rules for CMOS circuits and systems.

INPUT CIRCUIT RULES

The digital signals applied to CMOS IC input terminals must conform to a basic set of rules based on the following input parameters:

1. Input Amplitude
2. Frequency
3. Waveshape
4. Impedance
5. Noise Immunity

Input Amplitude

As shown in Fig. 5-1, the input voltage to the gate of a CMOS IC should not be greater than the positive power-supply voltage or less than the power-supply ground voltage. Exceeding these limits results in forward biasing of the IC input-protection di-

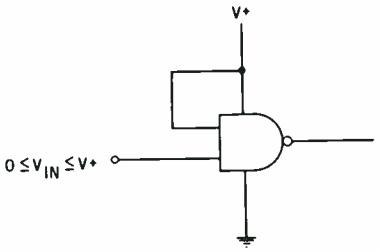


Fig. 5-1. Input circuit rules.

odes. And forward biasing of these diodes could result in the triggering of parasitic SCRs. Some manufacturers state that these input-voltage limits may be exceeded as long as the magnitude of the input current is less than 10 mA, but for many CMOS ICs this is a hazardous practice.

Also shown in Fig. 5-1 is another input circuit rule. All unused *inputs* must be connected either to ground or to supply (whichever is the inactive condition of the particular input). If these inputs are not properly terminated, it is virtually certain that the IC will not properly operate because of undesired signal pickup by the unterminated high-impedance input. (Many times, stray 60-Hz signals activate these unused input terminals).

Input Frequency

The maximum frequency of the input signal must be less than the maximum frequency of the IC being driven. (Typical maximum frequencies are 1 megahertz when a 5-volt IC power supply is used and 5 megahertz when a 15-volt power supply is used).

The minimum frequency is determined solely by the driving source, provided the waveshape of the driving waveform conforms to the design rules. Most CMOS ICs have no lower limit on operating frequency other than random IC failure. CMOS ICs are likely to operate for more than 50,000 hours. (A typical lifetime specification is that only one IC in one thousand fails during the first thousand hours.)

Input Waveshape

The digital input-signal waveshape can be important to proper circuit operation. For typical CMOS ICs, both n- and p-channel transistors are turned on if the input voltage exceeds one volt and is within one volt of the IC power-supply voltage. The usual CMOS circuit assumption is that the transition through this range of input voltage occurs quickly so that there is no significant direct path for current to flow from the IC power supply to ground. If the input waveshape is not rectangular, this assumption is not

accurate. The rise time and fall time of the input signals should be made less than 15 microseconds to minimize this shunt current to ground during switching. If input signals with rise and fall times of many milliseconds or seconds are used, the power dissipation rating of the CMOS IC being driven may be exceeded, or the particular IC circuit may not function due to temporary overlap of complementary timing waveforms.

Because of the dependence of CMOS power dissipation on input-signal waveshape, improperly designed CMOS oscillators can exhibit significant power consumption as a result of non-rectangular input waveshapes.

Input Impedance

The impedance associated with a CMOS IC input may be characterized as having a few picofarads (10^{-12} farads) capacitance shunted by a resistor of 100 megohms (or more) and having a leakage current of a few nanoamperes (10^{-9} amperes) (the reverse leakage current of the input-protection diodes).

The high, and nearly ideal, input impedance of CMOS ICs for the allowed amplitudes of input signals makes it very easy to interface this device with other circuits. CMOS ICs may be readily driven at moderate speeds from circuits having many megohms output impedance, for example, and many CMOS gates may be driven from a single circuit without performance degradation.

During input overloads, the impedance of CMOS ICs is that of a forward-biased diode, as shown in Fig. 5-2. If the input exceeds the power-supply voltage, the forward-biased diode connected to the supply passes the input current to the supply. If the input is more negative than the ground terminal, the input diode connected to ground is forward biased. These protection diodes act to clamp the input voltage to the bounds presented by the power supply and ground. Forward biasing these protection diodes should be avoided since the parasitic devices discussed previously in this chapter could be activated.

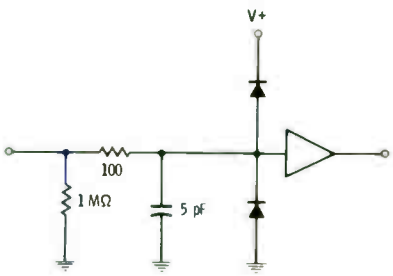


Fig. 5-2. Equivalent circuit of CMOS input stage.

Thus, the input impedance of properly operated CMOS ICs is typically quite high, while for overload signals it can be very low.

Input Noise Immunity

The transfer characteristic of a CMOS IC inverter (shown in Fig. 5-3) is the key factor in determining CMOS noise immunity. This plot of output voltage versus input voltage of the simple inverter demonstrates that the input signal has an effect on the output voltage only for a narrow range of input voltages around the *transfer point*. In the last chapter, double buffering was shown to reduce this active input range even further. The exact input voltage corresponding to the transfer point varies from IC to IC, depending upon the n- and p-channel transistor threshold voltages. These threshold voltages typically vary several tenths of a volt because they are sensitive to the exact quantity of dopant in the n-type starting-material substrate and in the p-well and also because they are sensitive to the quantity of residual ionized atoms in the gate oxide-insulating layer. For typical ICs this transition voltage is specified to occur between 30% and 70% of the IC power-supply voltage. Thus, for a 15-volt IC supply the transition point is between 4.5 volts and 10.5 volts.

The noise-voltage immunity of CMOS is often specified as being 30% of the supply voltage. For 15-volt supplies, an undesired input transient would have to create signals in the range between 4.5 to 10.5 volts to disrupt normal circuit operation. This high noise-voltage immunity is useful when circuits are interfaced by means of long cables which may have noise transients superimposed on the logic signals.

The noise-power immunity of CMOS ICs may be made large by using the technique shown in Fig. 5-4. The noise-power threshold

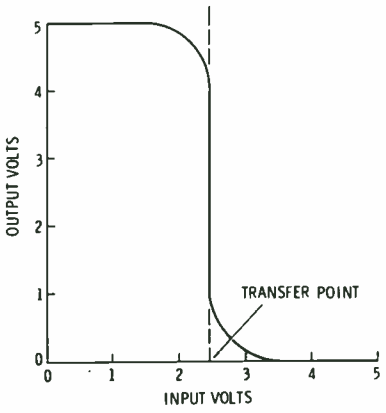


Fig. 5-3. CMOS transfer characteristic.

is usually defined in terms of the energy required to reach the transition point. The energy required to reach this point is determined primarily by the energy required to charge the capacitance of the external capacitor shown in the figure and by the capacitance associated with the IC input circuitry. This technique is useful for eliminating switching-transient noise introduced at the IC input by stray signals coupled to interconnection wires.

Another facet of noise immunity is the rejection of power-supply noise. Switching circuits can introduce a noise voltage on the power supply. If this voltage is coupled into the CMOS circuit, it will disrupt normal circuit operation. The rejection of noise voltage at this terminal is confused somewhat by the exact circuit affected, but the rejection is roughly of the same order as the input voltage rejection (30% of the supply voltage).

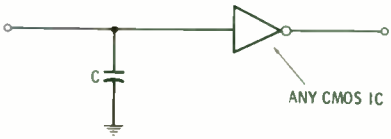


Fig. 5-4. Increased noise-power immunity with external input-loading capacitor.

Although the worst case of noise-voltage immunity is 30% of the supply, many specifications cite a nominal 50% of the supply voltage. Noise voltages that are greater than 1 volt and are a volt or more below the IC power supply, cause additional power consumption by the IC because both n- and p-channel transistors are turned on.

Voltage may be dropped across the on resistance of the p-channel transistor if current flows while the output is in the high state, and it may be dropped across the on resistance of the n-channel transistor if current flows while the output is in the low state. However, CMOS circuits driving CMOS circuits will exhibit nearly full supply-voltage swings at the IC output because of the very small current requirements of the CMOS IC input circuit between switching operations.

There are maximum current ratings associated with most CMOS outputs. Typically this maximum rating is about 1 mA except for some special high-current drivers which may provide up to 10 mA. Several CMOS ICs may be put in parallel to increase the current ratings, but the safe level is *less* than the sum of the current ratings because the currents passed by all outputs will not be identical. Exceeding the maximum current ratings slightly will usually not result in immediate destruction of the IC output but may cause premature failure after a number of hours of operation similar to the effect of minor current overloads on relays or

cables. Short-duration minor current overloads may be withstood many times without failure.

OUTPUT CIRCUIT RULES

As with CMOS inputs, the output amplitude should never be forced above the supply or below ground during normal circuit operation to prevent forward biasing the output circuit drain-substrate diodes and to prevent the stimulation of CMOS parasitic devices. Problems sometimes arise when driving reactive loads with CMOS. Switching the current through inductors, for instance, can result in transients which may cause a violation of this output rule.

Operating CMOS ICs at the maximum-allowed supply voltage minimizes the "on" resistance of the output-stage transistors and therefore minimizes the effects associated with this resistance.

Output Frequency

The maximum frequency at the output of a CMOS IC is determined by the total output load capacitance being driven and by the on resistance (see "CMOS Inverters," in Chapter 4) of the output stage transistor. Thus, the maximum frequency is determined by:

$$\text{Freq}_{max} = \frac{1}{2 \times 2.2 \times R_{on} \times C_{load}}$$

where,

R_{on} is the on resistance of the output stage transistor in ohms,
C_{load} is the output load capacitance in farads.

This maximum frequency is made large by minimizing the on resistance and the load capacitance. Two useful guidelines for obtaining maximum frequency operation at CMOS outputs are: (1) use the maximum supply voltage to minimize on resistance and (2) drive as few ICs as possible to minimize capacitance.

There are no special rules for low-frequency design of CMOS IC outputs.

Output Waveshape

The output waveshape of CMOS is determined by the input waveshape, the IC input-output transfer characteristics, the load capacitance, and the output transistor on resistance.

If the input waveshape is perfectly rectangular, the output waveshape of a CMOS IC is accurately characterized by the resistor-capacitor charging equation:

$$V_{out} = V_{supply}\ (1 - e^{-1/R_{on}C_{load}})$$

where,

V_{out} is the output expressed in volts,
V_{supply} is the total supply voltage,
e is the base of the natural logarithm (2.718),
R_{on} is the on resistance in ohms,
C_{load} is the output load capacitance in farads.

The simple single-buffered inverters, for example, often demonstrate significant interaction between input and output waveshapes because of transfer characteristics (Fig. 5-5).

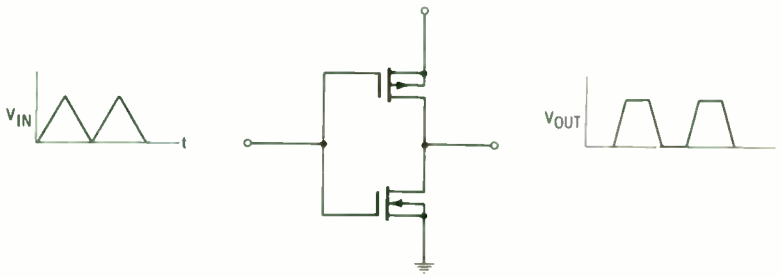


Fig. 5-5. Effect of inverter transfer gain on inverter input/output waveshape interaction.

Good practice dictates that the rise time should be less than the 15 microseconds required by the input circuits of most CMOS ICs.

Output Impedance

As discussed in Chapters 1 and 3, the output impedance of CMOS is predominately resistive. In the low state the output impedance is essentially the on resistance of the n-channel transistor, and in the high state it is the on resistance of the p-channel transistors. This on resistance is influenced by the IC power-supply voltage, shown in Fig. 5-6. Larger supply voltages result in greater gate-source transistor bias and lower on resistance.

POWER-SUPPLY RULES

Most CMOS ICs operate with any supply voltage between 3 volts and 15 volts. The power-supply rules are easily stated in terms of the power-supply parameters:

1. Supply Current
2. Supply Voltage

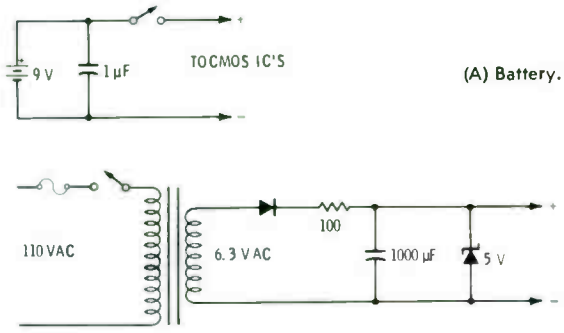


Fig. 5-6. CMOS power sources.

Supply Current

A typical CMOS IC circuit with some simple power supplies is shown in Fig. 5-6. The simple 9-V transistor-radio battery supply is useful for nearly all but the most complex and high-frequency CMOS circuits.

The current required from the supply depends upon the frequency of the circuit operation, the supply voltages, and the output load impedance.

Low-frequency operation reduces the current drain because the circuit capacitances shown in Fig. 5-7 are charged and discharged less frequently. Current flows into the CMOS ICs principally during switching of logic states within the ICs. The capacitor is placed in parallel with the transistor-radio battery to stabilize the supply voltage during these current transients (Fig. 5-6).

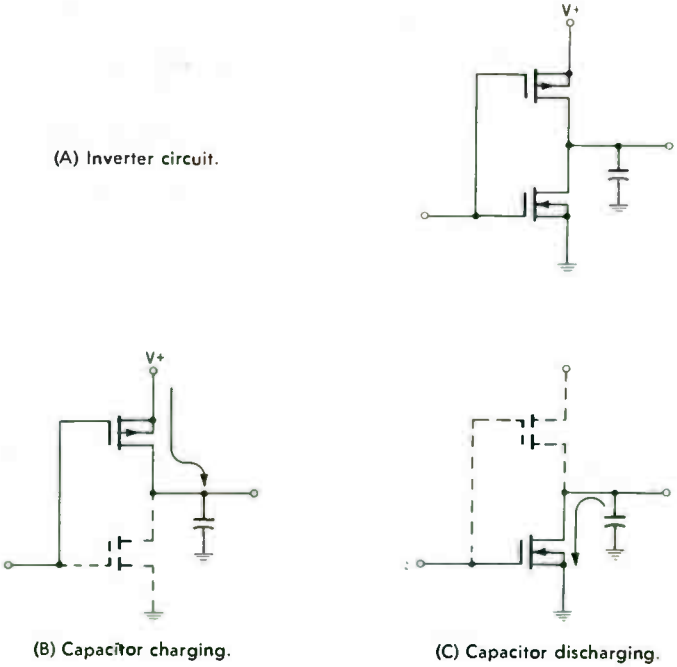
Low-voltage operation reduces the supply current requirements because less charge is placed on the circuit capacitances.

Thus, the reduced on resistance and increased output-circuit speed, which result from using the maximum supply voltages, come at a price of higher power-supply current drain and shorter battery life (if used).

The power consumption associated with the loads is slight if only CMOS ICs are driven. Often low-impedance circuits are driven at the output of the CMOS circuitry. The power consumption associated with driving this low-impedance circuitry may easily be much greater than the power consumption of all the CMOS circuitry used.

Supply Voltage

The supply voltage for popular CMOS ICs may range between 3 volts and 15 volts. Minimum supply-voltage operation (3 volts) is useful to minimize power consumption, while high-voltage op-



(A) Inverter circuit.

(B) Capacitor charging.

(C) Capacitor discharging.

Fig. 5-7. Inverter-switching power consumption.

eration (15 volts) is desirable to maximize the frequency of operation obtainable with CMOS and to minimize the effects of output-stage on resistance.

The CMOS circuitry is relatively insensitive to moderate power-supply ripple and switching-transient noise. Noise voltage or power-supply voltage fluctuations of a volt can typically be tolerated because of the high noise-voltage immunity of CMOS ICs.

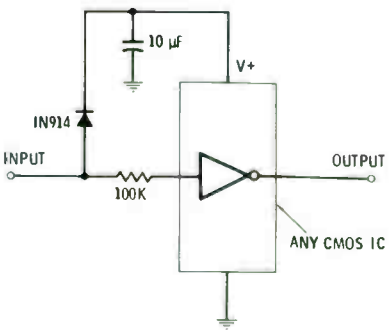


Fig. 5-8. Input-signal-powered operation of CMOS ICs.

A novel technique for obtaining a CMOS power supply is shown in Fig. 5-8. This circuit derives its power from the digital input signal, which may prove useful for powering remote circuits through long interconnection lines. (Only two wires are required for power and signal.) Another variation is the camera-control circuit which is powered by the circuit driven through its own output waveform; it is described in Chapter 8 (A Delayed-Release Timer).

INTERFACE RULES

Many systems use CMOS ICs mixed with other specialized IC types. For example, the high-speed circuitry in a system might be built by using TTL (transistor-transistor-logic) or ECL (emitter-coupled-logic), and the low-speed circuitry might be built by using CMOS. The interface rules provide a basis for designing circuitry that interfaces the drive requirements and output capabilities of CMOS with those of other specialized logic families. The interface rules for MOS, TTL, and ECL are given in this section as examples of interfacing techniques.

MOS Interface Rules

The output drive capabilities and input drive requirements of ICs fabricated with only n-channel or p-channel transistors are well suited for interfacing CMOS ICs. These MOS ICs may be driven directly from CMOS ICs and may drive CMOS ICs with no specific precautions other than observation of the manufacturer's recommended power-supply voltage amplitudes and polarity. This interface compatibility makes the intermixing of CMOS ICs and MOS ICs straightforward.

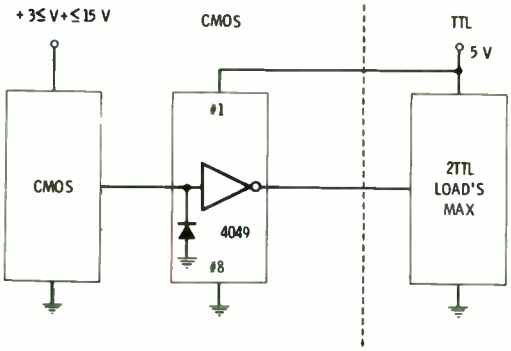
TTL Interface Rules

Three circuits that can be used for interfacing TTL and CMOS ICs are shown in Fig. 5-9. The 4049 or 4050 IC, as shown in Fig. 5-9A, is able to drive TTL directly. These ICs have special input-protection circuitry to allow the input voltage to exceed the IC supply voltage.

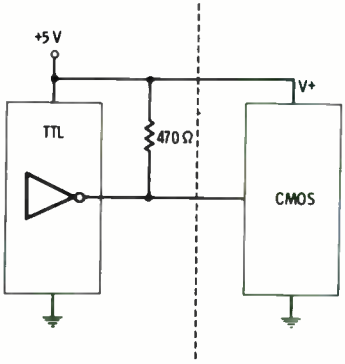
When a CMOS IC is driven from a TTL IC, the circuit of Fig. 5-9B can be used if the CMOS IC is operated from the same supply voltage as the TTL IC. Otherwise the circuit of Fig. 5-9C should be used.

ECL Interface Rules

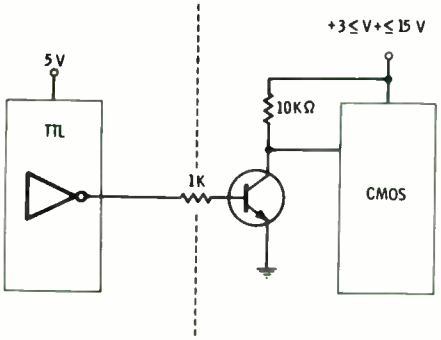
A simple interface to drive ECL ICs from CMOS ICs and to drive CMOS ICs from ECL ICs is shown in Fig. 5-10.



(A) CMOS (4049) as CMOS-to-TTL supply-voltage interface.

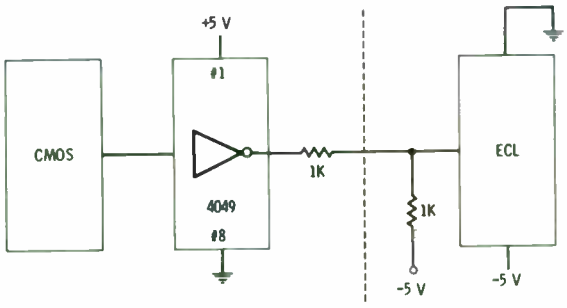


(B) Use of 470-ohm pull-up resistor for TTL-to-(5-V supply)-CMOS

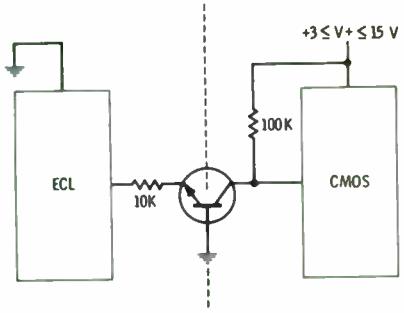


(C) CMOS and transistor for TTL-to-CMOS interface.

Fig. 5-9. CMOS and TTL interface circuitry.



(A) CMOS-to-ECL.



(B) ECL-to-CMOS.

Fig. 5-10. ECL and CMOS interface circuitry.

DEBUGGING RULES

Whenever a new circuit is built, there is a chance that it will not operate as intended when first tested. Here are seven debugging rules that can be used to help find the trouble in a CMOS IC circuit that works improperly.

1. Check all unused inputs to verify that they are connected either to ground or to the IC supply voltage.
2. Check the power supply to determine if the design voltage is present.
3. Check for input and output amplitude violations. (See first part of this chapter.)
4. If the power supply is intermittently shorted by a CMOS IC, check for possible sources of SCR firing.
5. Compare circuit rise and fall times with manufacturers' specification. (Typically no problems will be encountered for rise and fall times less than 5 microseconds.)

6. Verify that the CMOS ICs can provide the frequency of operation for the supply voltage used.
7. Touch-test the CMOS ICs for proper operating temperature. They should not be hot.

The following chapters give practical insight into the application of the circuit and system design rules discussed in this chapter.

6

Frequency Generators

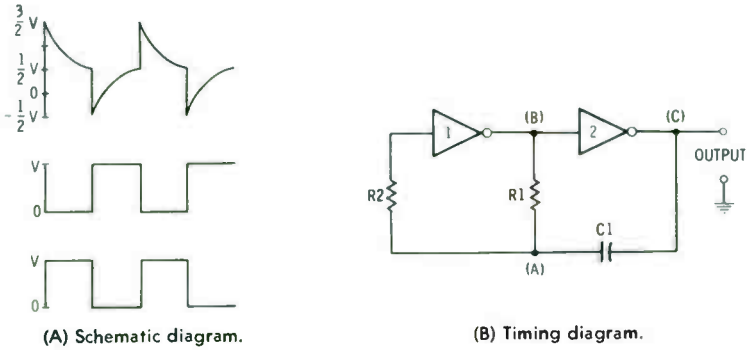
Frequency-generator circuits of one type or another are used in almost every digital IC system. Some frequency-generator circuits are designed to produce a waveform with a fixed frequency, while others are designed to produce a waveform with a frequency that depends on some external signal. Both types of frequency-generator circuits will be discussed in this chapter.

ASTABLE MULTIVIBRATORS

The function of an astable multivibrator is to generate a square wave. The basic CMOS astable multivibrator shown in Fig. 6-1A is identical to the square-wave generator circuit presented in Chapter 1.

To understand how this astable multivibrator circuit works, it is useful to assume that each of the two inverters in the circuit has an ideal transfer characteristic, like that shown in Fig. 6-2. An inverter with this ideal transfer characteristic has a transfer voltage equal to exactly one-half of its power-supply voltage. If the voltage at the input of the inverter is less than this transfer voltage, the voltage at the output of the inverter is equal to the supply voltage. If the input voltage is greater than this transfer voltage, the output voltage is equal to zero.

Assuming that the inverters have an ideal transfer characteristic, the operation of the astable can be seen from the timing diagram of Fig. 6-1B. This diagram shows the voltage at three points in the circuit as a function of time. Initially the voltage at the output (point C) is at +V. Since point C is at the output of an in-



(A) Schematic diagram. (B) Timing diagram.

Fig. 6-1. The basic CMOS astable multivibrator.

verter, the voltage at the input to this inverter (point B) must be less than ½ V. In fact, point B must be at 0 volts since it is at the output of an inverter which can have only a 0-volt or a +V-volt output. Since the output of this first inverter is at 0 volts, the voltage at point A must initially be greater than ½ V.

As the capacitor charges through resistor R1, the voltage at A will exponentially approach the voltage at B. But as soon as the voltage at A reaches the transfer voltage, the output of the first inverter (point B) will switch from 0 to +V, which causes the output voltage of the second inverter to switch from +V to 0. During this transient the voltage at A will decrease to $-(V/2.)$. The purpose of R2 is to prevent the voltage at point A from being limited by the protective-input circuitry of the first inverter. R2 should have a resistance at least twice that of R1.

Following this switching transient, the voltage at point A will again exponentially approach the voltage at point B as the capacitor charges. When the voltage at A again reaches the transfer voltage, point B will switch back to 0 volts, and point C will return to +V volts. During this transient the voltage at A will increase to 3/2 V. The voltage at A will again decrease exponentially, and as shown in Fig. 6-1B, the cycle will repeat.

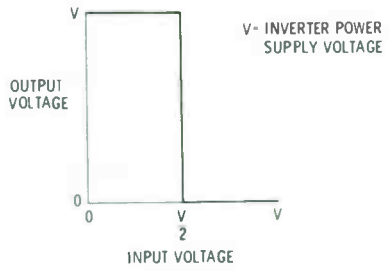
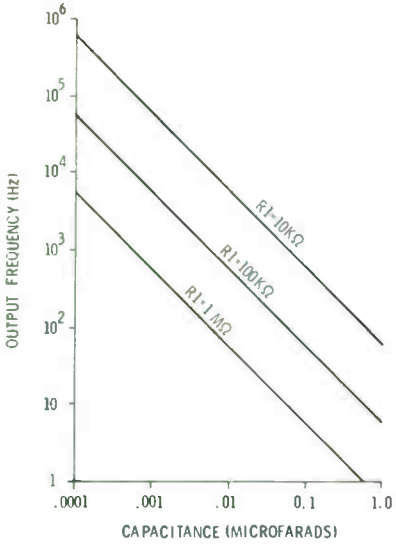


Fig. 6-2. Ideal inverter transfer characteristic.

The frequency of the square wave produced at the output of this circuit is dependent on the value of the resistor, R1, and of the capacitor, C1. Chart 6-1 shows the output frequency as a function of these values.

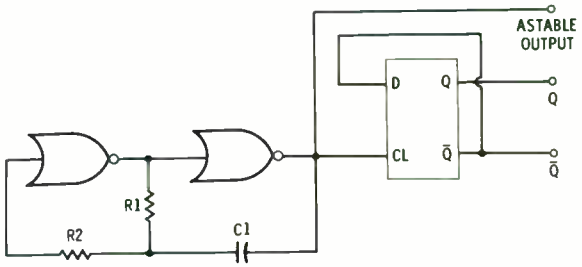
Chart 6-1. Astable Multivibrator Design Chart



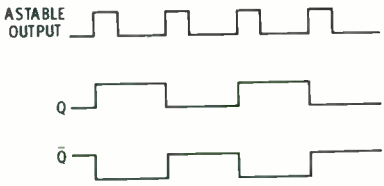
A Symmetric Square-Wave Generator

In the above analysis of the basic astable circuit, it was assumed that the inverter transfer voltages were 50% of the power-supply voltage. In fact, however, the transfer voltage of CMOS ICs can vary between 30 and 70% of the supply voltage. One consequence of this is that the output of the basic astable circuit might not be perfectly symmetric; that is, the length of time the output is at logic 1 for each cycle might not be exactly equal to the length of time the output is at logic 0.

If it is necessary to generate a symmetric square wave, the circuit of Fig. 6-3A can be used. In this circuit the output of the basic astable circuit is used to toggle a flip-flop. The flip-flop changes state on every positive transition of the astable output. As illustrated in Fig. 6-3B, the output square waves produced at the Q and $\bar{Q}$ outputs are symmetric even when the output of the astable is assymmetric. The frequency of these square waves is one-half the frequency of the astable output.



(A) Schematic diagram.



(B) Output waveforms.

Fig. 6-3. A symmetric square-wave generator.

The 4047 as an Astable Multivibrator

A single CMOS IC, the 4047, can perform the same function as the circuit of Fig. 6-3A. The wiring diagram of Fig. 6-4 shows how the 4047 can be connected to do this job. The astable output is available at pin 13 of the IC, and the Q and $\bar{Q}$ outputs are available at pins 10 and 11, respectively. The values of R1 and C1 required to obtain a particular frequency at the astable output can be determined from Table 6-1. The frequency of the square waves at pins 10 and 11 is one-half of the astable frequency.

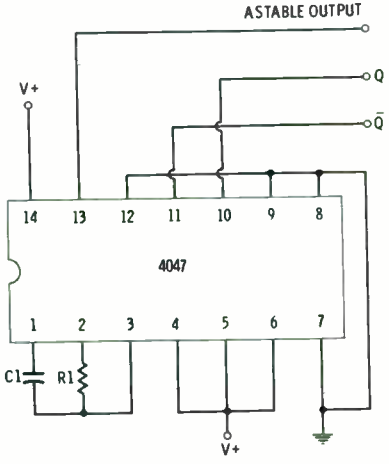


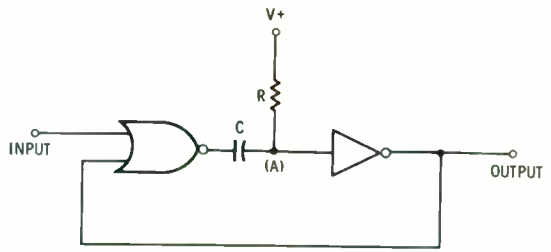
Fig. 6-4. Wiring diagram used to connect a 4047 IC as an astable multivibrator.

MONOSTABLE MULTIVIBRATORS

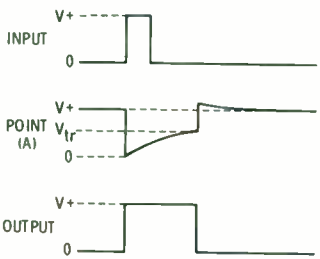
The function of the monostable multivibrator is to generate, upon command, a single output pulse of set duration. The basic CMOS monostable circuit is shown in Fig. 6-5A. This circuit is designed to produce an output pulse on positive transitions of the input signal.

The operation of the monostable circuit can be understood by referring to the timing diagram of Fig. 6-5B. Initially both the input and the output are shown at zero volts, and the voltage at point A (Fig. 6-5A) is at the positive supply voltage. When the input goes high, the output of the NOR gate is forced low, and point A goes to zero volts. Since point A is the input of the inverter, the output of the inverter goes high. The inverter output is fed back to one of the NOR gate inputs so that the output of the NOR gate remains low even after the input signal returns to logic 0. The capacitor now charges up through the resistor, and the voltage at point A rises exponentially. As soon as the voltage at point A reaches the transfer voltage of the inverter (V_{tr}), the monostable output returns to logic 0.

The duration of the pulse produced at the output of the monostable is dependent on three factors: (1) the value of resistor R, (2) the value of capacitor C, and (3) the transfer voltage of inverter V_{tr}. Assuming that the inverter has an ideal transfer char-



(A) Schematic diagram.



(B) Timing diagram.

Fig. 6-5. The basic monostable multivibrator.

acteristic, as shown in Fig. 6-2, the duration of the output pulse is given by the formula:

$$\text{Pulse width} = 0.7 \times R \times C$$

where,

Pulse width is in milliseconds,
R is in kilohms,
C is in microfarads.

A Monostable With Fast Recovery

One difficulty with the basic CMOS monostable circuit is that when its output returns to zero, the voltage at point A (Fig. 6-5A) increases to approximately 0.6 volt above the supply voltage (where it is limited by the input-protection diodes of the inverter). The voltage at point A, then, exponentially decreases to the supply voltage, as seen in Fig. 6-5B. While the voltage at A is decreasing to the supply voltage, the monostable circuit is said to be *recovering*. If the monostable is triggered while it is recovering, the output pulse duration will be shorter than it is when the monostable is fully recovered. This effect is particularly pronounced when the monostable is being operated from a low power-supply voltage.

The circuit of Fig. 6-6 demonstrates one way that a monostable can be designed to decrease the recovery time. In this circuit the power supply for the inverter of the monostable is not taken directly from V+, but rather is taken through a silicon diode from V+. Since the diode develops a voltage drop of approximately 0.6 volts, the voltage supplying the inverter is maintained at about 0.6 volts below V+. The advantage of this circuit is that when the output of the monostable returns to logic 0, the voltage at point A will return to V+ directly, without the overshoot and

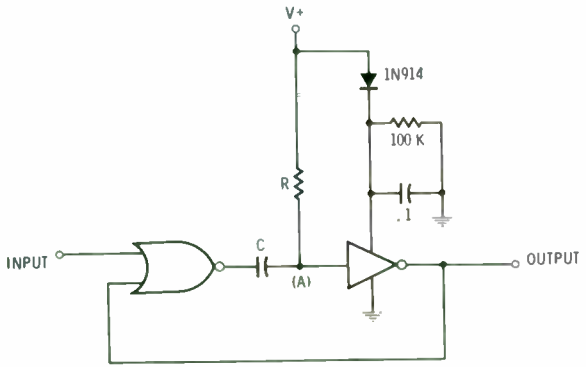


Fig. 6-6. A fast-recovery monostable multivibrator.

exponential recovery characteristic of the basic monostable circuit.

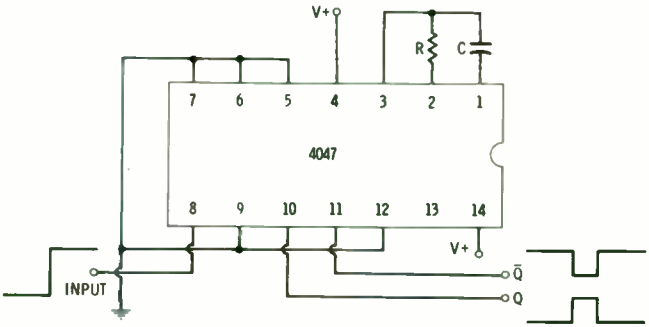
The 4047 As a Monostable Multivibrator

The same 4047 CMOS IC that was used as an astable multivibrator in Fig. 6-4 can also be used as a monostable. The 4047 monostable can be connected so that it is triggered on positive transitions or on negative transitions of the input. Both of these circuit connections are shown in Fig. 6-7. When triggered, the 4047 produces a positive and a negative pulse, at the Q and $\overline{Q}$ outputs, respectively. The duration of the output pulse is given by the formula:

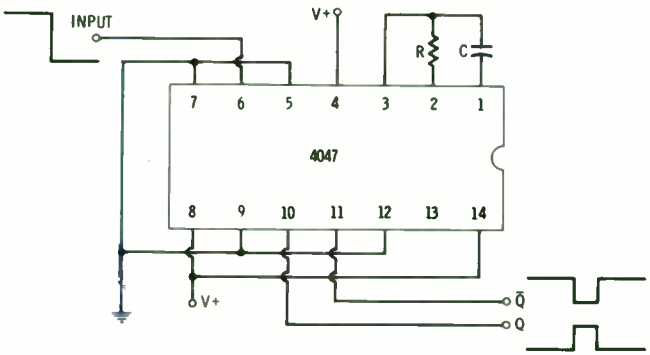
$$\text{Pulse width} = 2.5 \times R \times C$$

where,

Pulse width is in milliseconds,
R is in kilohms,
C is in microfarads.



(A) Triggered by positive-going input.



(B) Triggered by negative-going input.

Fig. 6-7. Wiring diagram used to connect the 4047 IC as a monostable multivibrator.

The value of the capacitor used in the 4047 monostable circuit should not be less than .001 μF, and the value of the resistor should not be less than 10K. The 4047 monostable pulse width is tabulated for several standard values of R and C in Table 6-1.

Table 6-1. 4047 Monostable Output Pulse Width

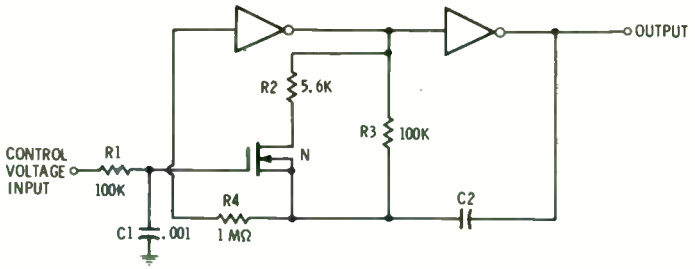
Output Pulse Width	R (ohms)	C (microfarads)
25 μsec	10K	.001
67 μsec	27K	.001
.12 msec	47K	.001
.25 msec	100K	.001
.67 msec	27K	.01
1.2 msec	47K	.01
2.5 msec	100K	.01
6.7 msec	270K	.01
12 msec	470K	.01
25 msec	100K	.1
67 msec	270K	.1
.12 sec	470K	.1
.25 sec	1 Meg	.1
.67 sec	2.7 Meg	.1
1.2 sec	4.7 Meg	.1

A VOLTAGE-CONTROLLED OSCILLATOR

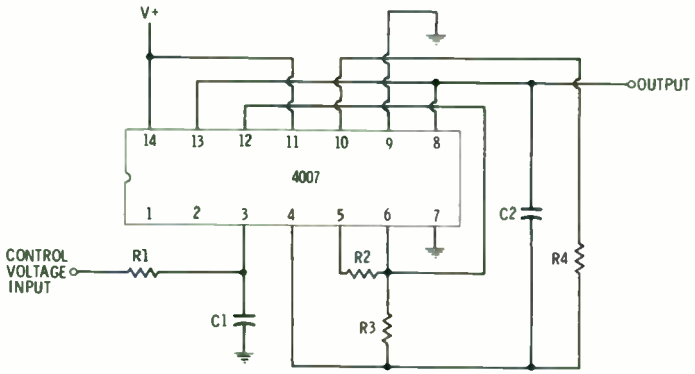
A voltage-controlled oscillator (vco) circuit is shown in Fig. 6-8A. This circuit is similar to the basic astable circuit; however, the output frequency of this circuit can be controlled by an input control voltage. The control voltage can range from zero up to the IC supply voltage. Once the control voltage exceeds the n-channel MOSFET threshold voltage (which is usually between 1 and 2 volts), the output frequency of the vco will increase as the control voltage is increased.

The low-frequency limit of the vco range is determined by the values of R3 and C2, and can be obtained from Chart 6-1. The high-frequency limit for the circuit shown is approximately ten times the low-frequency limit. A wider vco range can be achieved by reducing the value of R2. The purpose of R1 and C1 is to protect the n-channel MOSFET from possible damage due to static discharges or brief voltage surges that might be applied to the control-voltage input.

The vco circuit of Fig. 6-8A can be built by using a single CMOS IC. The wiring diagram for the circuit is shown in Fig. 6-8B. The



(A) Schematic diagram.



(B) Wiring diagram.

Fig. 6-8. A voltage-controlled oscillator.

CMOS 4007 IC was chosen because of the availability of an internal n-channel MOSFET at pins 3, 4, and 5.

A CRYSTAL OSCILLATOR

The astable multivibrator makes a very good general-purpose square-wave generator; but, when the utmost in frequency accuracy and stability is required, a crystal oscillator should be

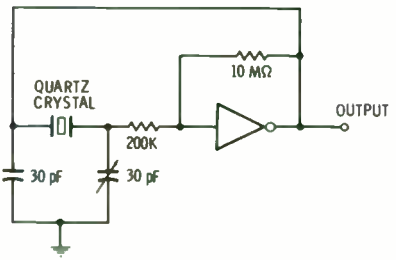


Fig. 6-9. A crystal-oscillator circuit.

used. A crystal-oscillator circuit using one CMOS inverter is shown in Fig. 6-9. Depending on the crystal selected, CMOS IC crystal oscillators can be designed with output frequencies anywhere from 10 kHz to 10MHz. If lower, very stable, accurate frequencies are required, a frequency divider can be used to divide the output frequency of a crystal oscillator.

FREQUENCY DIVIDERS

Frequency dividers are circuits designed to generate an output signal having a frequency equal to the frequency of the input signal divided by some integer. Digital frequency dividers are very precise circuits, are fairly easy to implement, and are ubiquitous in digital systems. Consider, for example, the CMOS digital wristwatch. The time base for the wristwatch is a 32,768-Hz crystal oscillator. The oscillator output is divided by 32 to generate a 1024-Hz signal for a voltage converter. The signal is further divided by 1024 to generate a 1-Hz signal for the seconds display. The 1-Hz signal is divided by 60 to derive minutes, and again by 60 to derive hours. The hours signal is further divided by 12 in a twelve-hour watch, or by 24 in a 24-hour watch. Frequency-divider circuits that divide by powers of two are particularly easy to implement, which is one reason why 32,768 Hz (a power of two) is the master frequency used in CMOS wristwatches.

Dividing by Powers of Two

The most basic frequency-divider circuit is the toggle flip-flop (or "divide-by-two"). The toggle flip-flop, as described in Chapter 1, generates an output frequency that is one-half of its input

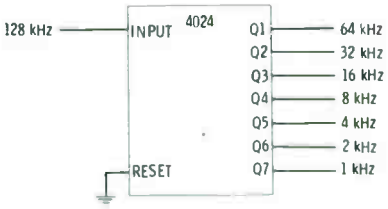


Fig. 6-10. A binary counter used to divide the frequency of an input signal by powers of two.

frequency. Several toggle flip-flops can be cascaded to divide by any desired power of two. The 4024 7-stage binary counter, for example, consists of seven cascaded toggle flip-flops. The 4024 can thus be used to divide an input frequency by 2, 4, 8, 16, 32, 64, or 128. The diagram of Fig. 6-10 illustrates the use of the 4024

as a frequency divider when the input signal has a frequency of 128 kHz.

Divide-by-N Circuits

The circuit of Fig. 6-11 is a handy circuit designed to divide the frequency of an input signal by any desired integer less than ten. The circuit uses just two CMOS ICs, a 4017 and a 4001. The 4017 is a 5-stage Johnson counter with integral one-of-ten decoding. The 4001 is a quad NOR gate.

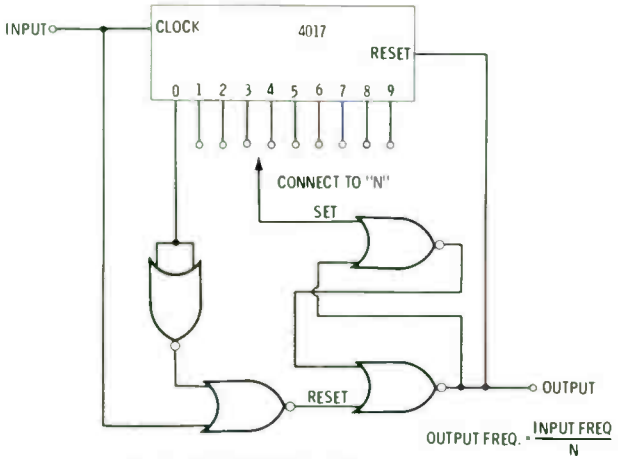


Fig. 6-11. A divide-by-n frequency counter.

The operation of this "divide-by-n" circuit is straightforward. The 4017 counter is incremented on every positive transition of the input signal and thus counts the number of input pulses. On the n*th* pulse a logic 1 signal appears at the n*th* output, which is connected to the set input of the RS flip-flop. This flip-flop is built from two NOR gates and is similar to the circuit of Fig. 1-9. When the flip-flop is set, the output of the RS flip-flop goes high, which resets the 4017 counter. Once the counter is reset and the input signal returns to logic 0, the RS flip-flop is reset and the count cycle repeats. Thus, a pulse appears at the output after every n input pulses, and the frequency of the input signal is effectively divided by n.

The circuit of Fig. 6-11 can be expanded to divide by an arbitrarily large n. The circuit of Fig. 6-12 illustrates how this is done. Additional counters must be cascaded for each additional digit of the number n, and additional gates must be added to detect when the count reaches n (to set the RS flip-flop) and when the counters are all reset to zero (to reset the RS flip-flop). The circuit of

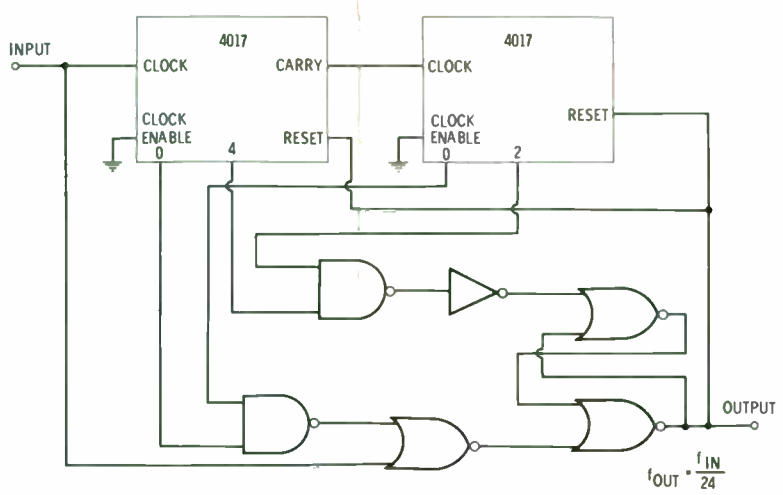


Fig. 6-12. A divide-by-24 frequency divider.

Fig. 6-12 is a divide-by-24 divider. If the appropriate outputs of the 4017s are used, this circuit can be modified to divide by any integer between 1 and 99.

Programmable-Frequency Dividers

A programmable divide-by-n frequency divider is designed so that the number n can be specified electronically. A circuit for a programmable divider is shown in Fig. 6-13. In this circuit the number n is specified as a binary-coded number at the data inputs, D1 through D4. With the 4522 CMOS IC in this circuit, n can be any integer less than 10. With the 4526 CMOS IC, n can be any integer less than 16.

Suppose you wish to use this circuit as a divide-by-7. All that is required is that the number seven be specified as a binary num-

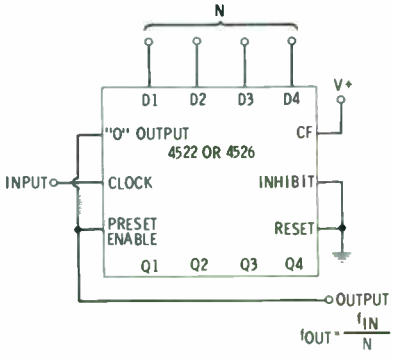


Fig. 6-13. A divide-by-n programmable divider.

ber at the data inputs. This is done by applying a logic 1 signal to inputs D1, D2, and D3, and by applying a logic 0 signal to D4. If an input signal with a frequency of 70 Hz were applied to this circuit, the output signal would have a frequency of exactly 10 Hz.

Before we explain how the circuit of Fig. 6-13 works, it is important to note that the 4522 IC is basically a bcd counter and that the 4526 IC is basically a binary counter. These counters are unusual, however, in that they count *down* rather than up. Thus, the binary-coded number at the counter outputs (Q1 through Q4) is *decreased* by one on every positive transition of the clock signal. Another important feature of these counters is that they are *presettable*. This means that the counter output may be *preset* at any time to the number presented to the data inputs. The counter is preset in this way whenever the preset enable input goes high.

Assume that an input signal with a frequency of f_{in} is applied to the input of the circuit of Fig. 6-13. The counter will count down on positive transitions of the input signal until the output count reaches zero. A count of zero is indicated by a logic 1 signal that appears at the 0 output. This output is connected to the preset enable input. So, as soon as the counter reaches zero, it is preset to the number n at the data inputs. With the next positive transition of the input, the counter again begins its count to zero. After every n*th* cycle of the input signal, a logic 1 pulse appears at the 0 output, and the cycle repeats. The output waveform of the programmable divider is taken from the 0 output and has a frequency of f_{in}/n.

The circuit of Fig. 6-13 can readily be expanded for larger values of n, as demonstrated by the circuit of Fig. 6-14. In this circuit three 4522 ICs are cascaded to allow frequency division by any number n between 1 and 999. N is coded as a three-digit bcd number at the data inputs of the three ICs. To assure that the preset enable is not activated until all three counters reach a count of zero, a CF (Cascade Feedback) input is provided on each counter. The 0 output of a counter cannot go high unless the CF input is high. Since the CF input of each counter is connected to the 0 output of the next most significant counter, the preset enable cannot be activated until each of the three counters reach zero.

The beauty of the programmable divider of Fig. 6-14 is that n can be programmed electronically. This circuit and programmable dividers similar to it are widely used in a number of important applications. One of these applications, frequency synthesis, is discussed later in this chapter.

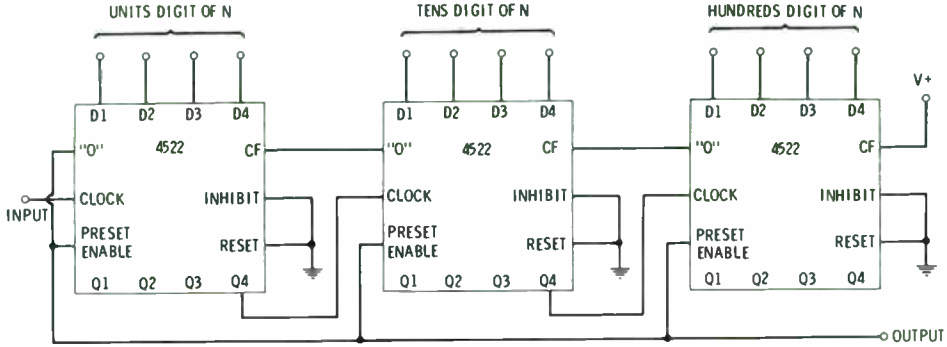


Fig. 6-14. A three-decade programmable divider.

A STAIRCASE GENERATOR

The circuit diagram of a staircase generator is shown in Fig. 6-15A. The circuit is built with two inverters, a 4024 7-stage binary counter, and a resistor summing network.

The two inverters in the circuit are connected as an astable. From Chart 6-1 it can be seen that the output frequency of the astable is approximately 100 kHz. The binary counter is incremented on negative transitions of the astable output. The seven outputs of the counter are summed by the resistor network to produce the staircase output waveform shown in Fig. 6-15B. The waveform starts at zero volts and rises in 128 steps, each step being 10 microseconds in duration. After 128 steps the output returns to zero and the cycle repeats. In Chapter 7 we will show how this staircase-generator circuit can be used in the implementation of an analog-to-digital converter.

PHASE-LOCKED LOOPS

Phase-locked loops are circuits that are designed to generate an output waveform having a fixed phase relationship with respect

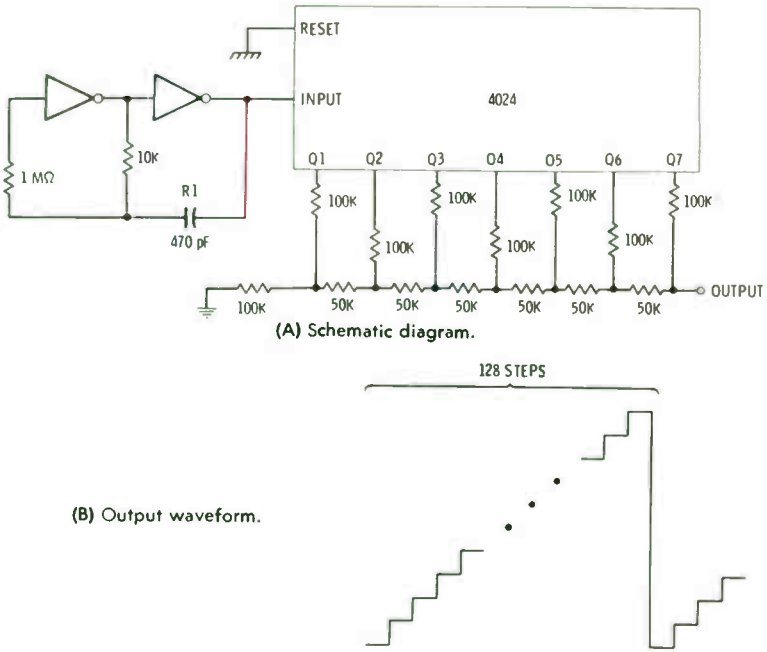


Fig. 6-15. A staircase generator.

to a reference input signal. To illustrate how a phase-locked loop works, suppose that you have a voltage-controlled oscillator (like the one of Fig. 6-8) and you want to match its frequency to the frequency of a reference signal. In addition, once the frequencies are matched, you want the two signals to be 90 degrees out of phase.

One way to match the frequency of the two signals is to display both of them on a two-channel oscilloscope. You then adjust the control voltage of the vco until the signals are stationary with respect to each other and are 90 degrees out of phase, as shown in Fig. 6-16. When you have achieved this condition, the two signals are said to be *phase locked.*

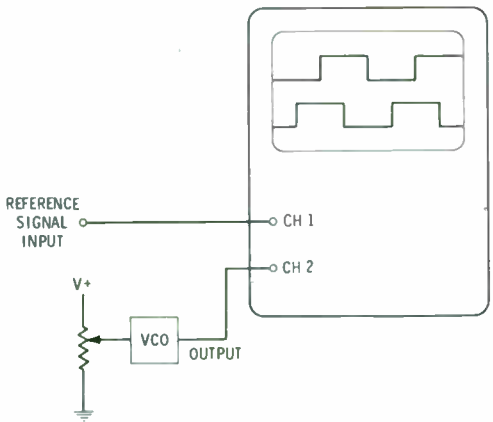
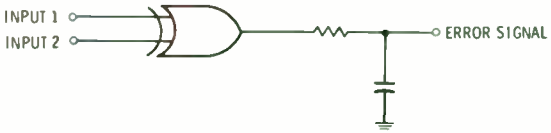


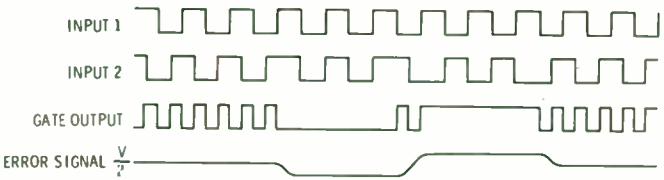
Fig. 6-16. A manually controlled phase-locked loop.

Now suppose that the frequency of the reference signal changes just slightly. One of the traces on the oscilloscope display will then begin to drift either to the left or to the right, and you will have to adjust the vco control voltage until the signals are again phase locked. In doing this, you are acting as part of a control loop, and you are doing manually what a phase-locked loop does automatically.

A phase-locked loop uses a *phase comparator* to measure the phase difference between a vco output signal and a reference signal. As shown in Fig. 6-17A, an exclusive-OR gate can be used as a phase comparator. The output of the gate is filtered by an RC filter to produce an *error signal.* As shown in Fig. 6-17B, as long as the reference signal and the vco output are 90 degrees out of phase, the error signal is equal to half of the supply voltage. If the two signals are more or less than 90 degrees out of phase, the error signal increases or decreases accordingly.



(A) Schematic diagram.



(B) Input and output waveforms.

Fig. 6-17. An exclusive-OR gate used as a phase comparator.

A phase-locked loop circuit is shown in Fig. 6-18. In this circuit the error signal is used as the control voltage of the vco. The signal generated by the vco is phase locked to the input signal. The circuit is designed so that if the output signal begins to drift with respect to the input signal, the resultant change in the error signal voltage will modify the vco frequency to reestablish a phase-locked condition.

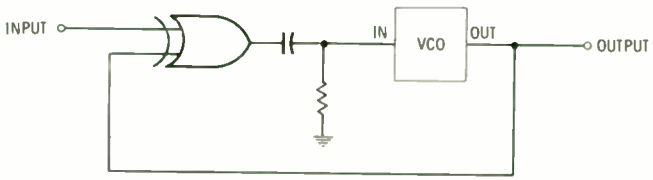


Fig. 6-18. A phased-locked loop circuit.

The 4046 Phase-Locked Loop

The CMOS 4046 phase-locked loop IC contains all the circuits required to build a phase-locked loop. A block diagram of the 4046 is shown in Fig. 6-19. The main components of this IC are a vco and two phase comparators. Either of the phase comparators can be used in the phase-locked loop design. Phase Comparator I is an exclusive-OR gate, while Phase Comparator II is a more sophisticated circuit. Phase Comparator I is less sensitive to noise on the input signal than Phase Comparator II is, but it will lock onto harmonics of the input waveform. Phase Comparator II is designed so that it will not lock onto harmonics of the input waveform.

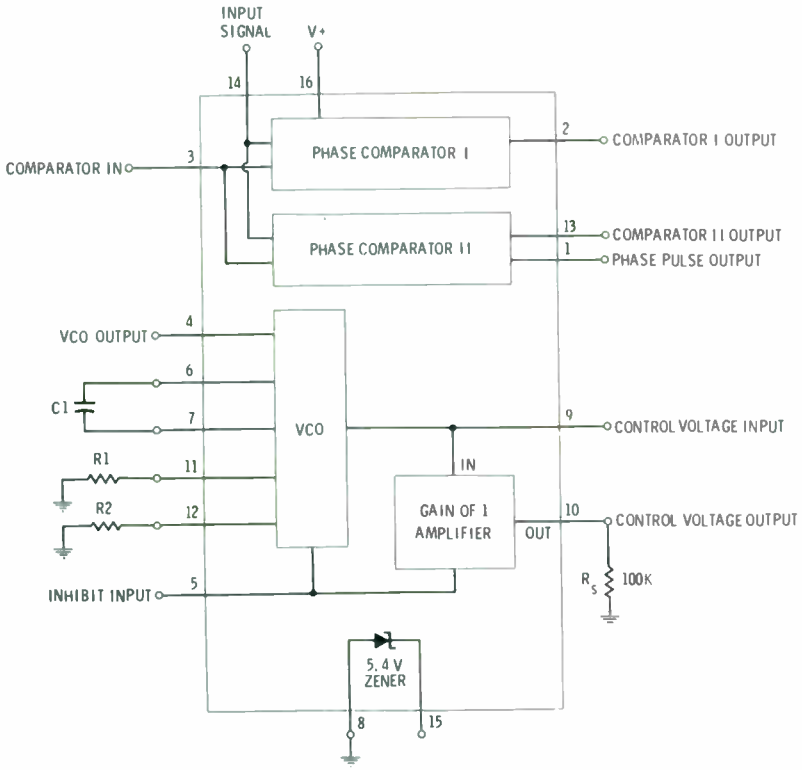
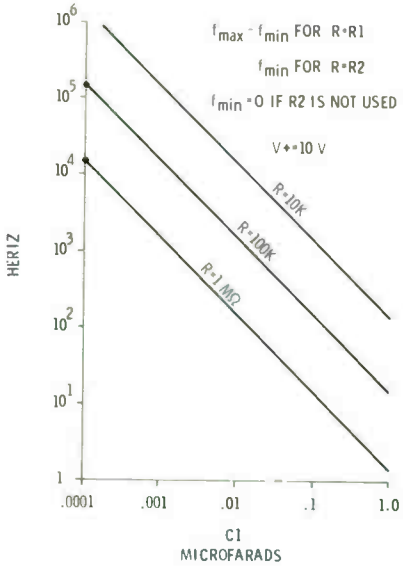


Fig. 6-19. A block diagram of a CMOS 4046 phase-locked-loop IC.

The vco of the 4046 IC has a minimum output frequency (f_{min}) when the control-voltage input is at zero volts and has a maximum output frequency (f_{max}) when the control-voltage input is at the positive supply voltage. The values of f_{max} and f_{min} are set by the external components, R1, R2, and C1, and can be found from Chart 6-2. With this chart the frequency range of the vco (f_{max}-f_{min}) can be found from the values of R1 and C1, and f_{min} can be found from the values of R2 and C1. If you want the frequency of the vco to be able to go all the way down to zero ($f_{min} = 0$), simply leave R2 out of the circuit.

The 4046 IC has several other features. A 5.4-volt zener diode is provided at pins 8 and 15, should you want to regulate the IC supply voltage. A gain-of-one amplifier is provided to give an output voltage equal to the vco control voltage; a 100K load resistor (R_s in Fig. 6-19) should be added if this output is used. A "phase pulse" output is provided by Phase Comparator II for detecting whether or not the phase-locked loop is locked onto the

Chart 6-2. A 4046 Phase-locked-Loop VCO Design Chart



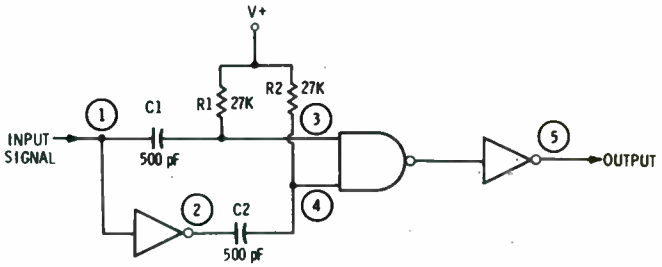
input frequency. Finally, an inhibit input is provided. The vco and gain-of-one amplifier are turned off when the inhibit input is high. Normally the inhibit input is held at logic 0.

FREQUENCY MULTIPLIERS

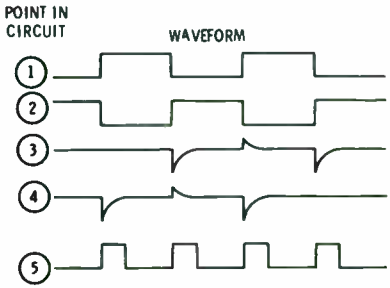
Frequency multipliers are circuits designed to generate an output signal having a frequency equal to an integral multiple of the input signal frequency. Digital frequency multipliers that multiply by powers of two can be built with just a few CMOS gates. More general multiply-by-n circuits can be built by using a CMOS phase-locked loop with a divide-by-n circuit in the feedback loop.

Multiplying by Powers of Two

A multiply-by-two circuit is shown in Fig. 6-20A. As seen from the timing diagram of Fig. 6-20B, the frequency of the output signal generated by this circuit is twice the frequency of the input signal. C1 and R1 are used to differentiate the input signal, and C2 and R2 are used to differentiate the complement of the input signal. The two differentiated signals are applied to the two inputs of the NAND gate. The input-protection circuitry of the NAND gate limits the positive excursions of the input signals to about 0.6 volt



(A) Schematic diagram.



(B) Timing diagram.

Fig. 6-20. A multiply-by-two frequency multiplier.

above the supply voltage. A negative excursion of either signal produces an output pulse.

Several multiply-by-two circuits can be cascaded to multiply by powers of two. When the circuits are cascaded, the output of

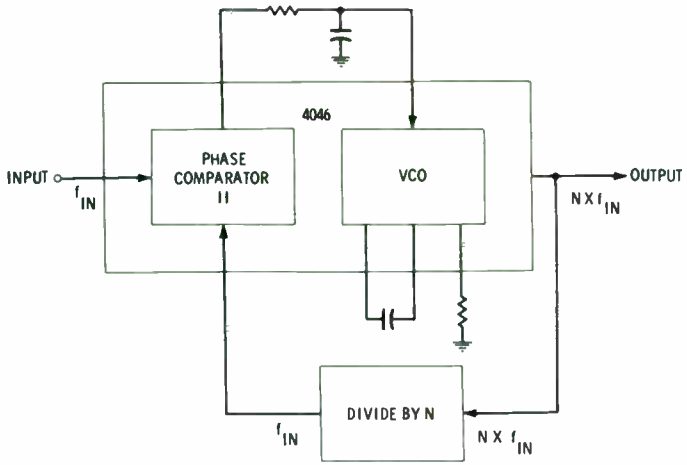


Fig. 6-21. A multiply-by-n frequency multiplier.

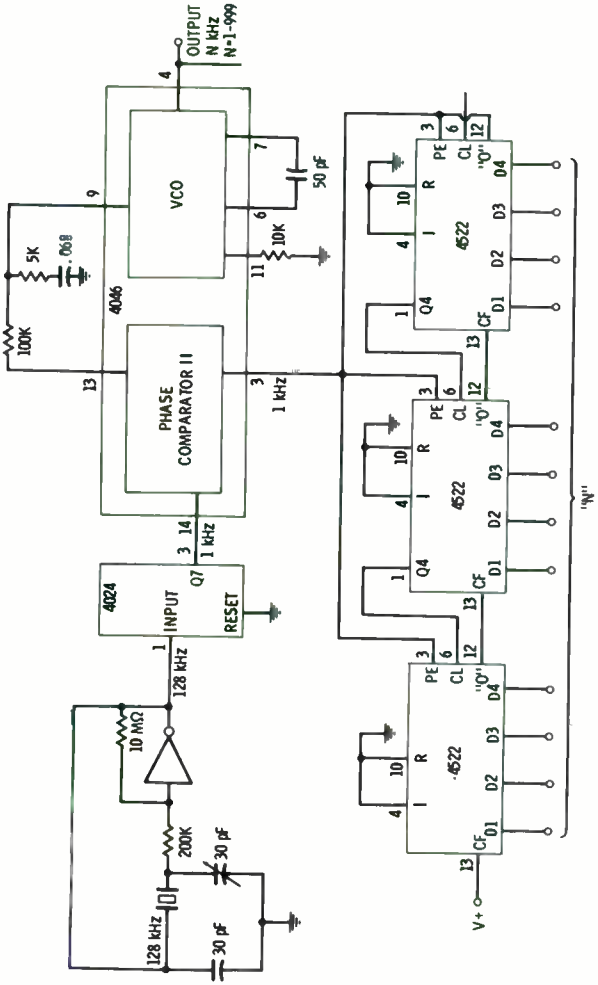


Fig. 6-22. A frequency synthesizer.

the n*th* stage will have a frequency of 2^n times the input frequency.

Multiply-by-N Circuits

Multiply-by-n circuits can be built by placing a divide-by-n circuit in the feedback path of a phase-locked loop, as shown in Fig. 6-21. In this phase-locked loop, the frequency of the vco is automatically adjusted until the two inputs to the phase comparator are exactly equal in frequency. This means that the output frequency of the divide-by-n circuit must be equal to the input frequency, f_{in}. Since the output frequency of the divide-by-n circuit is f_{in}, the input frequency to the divide-by-n must be equal to n times f_{in}. The output frequency of the multiply-by-n circuit is thus equal to n times f_{in}.

A FREQUENCY SYNTHESIZER

The circuit of Fig. 6-22 can be used to generate highly accurate and stable output frequencies that are digitally selectable. Three standard bcd thumbwheel switches can be used to set the output frequencies between 1 kHz and 999 kHz, in 1-kHz increments. This frequency synthesizer illustrates the application of several of the circuits discussed in this chapter.

The master frequency generator for this circuit is a 128-kHz crystal oscillator. A 4024 is used to divide the crystal-oscillator frequency by 128, which generates a 1-kHz square wave. This 1-kHz square wave is used as the input reference signal to a 4046 phase-locked loop.

The output frequencies are synthesized by a phase-locked-loop multiply-by-n circuit. A programmable divide-by-n circuit built with three 4522s is used in the feedback path. The output frequency of this frequency-synthesizer circuit is n kHz, where n is the number programmed at the bcd inputs of the 4522s.

7

Information-Processing Circuits

Electronic signals that carry information can be divided into two groups: analog and digital. *Analog signals* are signals that can vary continuously with time. *Digital signals,* on the other hand, can assume only one of two levels, logic 0 or logic 1. Integrated circuits have made a tremendous impact on the ease with which information-carrying signals can be processed. In this chapter we will discuss several practical CMOS IC circuits for processing both digital and analog signals and for converting from one type of signal to the other.

ANALOG-TO-DIGITAL CONVERSION

The process of converting an analog signal to a digital signal is called *digitization.* Once a signal has been digitized it can be processed by a digital computer or by other digital circuitry.

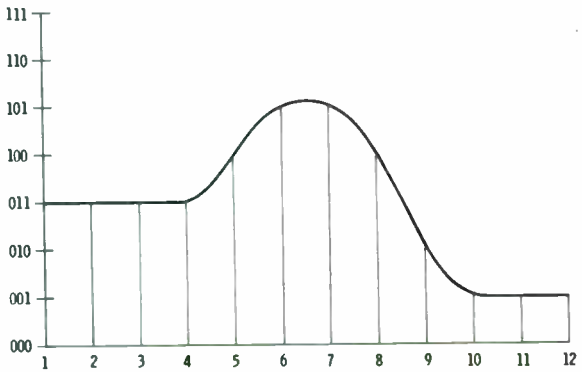
The process of digitization is illustrated in Fig. 7-1. In this illustration the analog signal shown is *sampled* at regular intervals, and the value of each sample is represented by a binary number. Since only three binary digits (bits) are used to represent each sample, there are only eight digitizing levels. Considerably finer resolution can be achieved when more than three bits are used to represent each sample.

A circuit designed for digitizing an analog signal is called an analog-to-digital (a/d) converter. A number of different types of circuits can be used to realize an a/d converter, one of which is illustrated by the seven-bit a/d converter circuit of Fig. 7-2. The three main sections of this circuit are: (1) a staircase generator

similar to the one of Fig. 6-15, (2) a 740 operational amplifier used as a voltage comparator, and (3) a bank of seven type-D flip-flops.

In order for this a/d converter to be effective, the analog input signal must be changing much more slowly than the sweep rate of the staircase generator. To see how the a/d converter works, it is important to note that as long as the staircase voltage is less than the voltage of the analog input signal, the output of the voltage comparator is logic 0. When the staircase voltage increases to the point where it first exceeds the voltage of the input signal, the output of the comparator goes to logic 1. This positive transition of the comparator output clocks all seven D flip-flops, thus storing in the flip-flops the output states of the 4024. These output states digitally represent the magnitude of one sample of the analog input signal.

As the staircase waveform goes through its cycle, the outputs of the D flip-flops are updated with every positive transition of the voltage comparator. As long as the analog input signal is changing slowly with respect to the staircase sweep rate, the set of 7-bit samples produced at the outputs of the D flip-flops accurately represent the magnitude of the analog signal as a function of time.



SAMPLE #	DIGITAL VALUE
1	011
2	011
3	011
4	011
5	100
6	101
7	101
8	100
9	010
10	001
11	001
12	001

Fig. 7-1. The digitization process.

DIGITAL-TO-ANALOG CONVERSION

A circuit used to reconstruct an analog signal from a digital signal is called a digital-to-analog (d/a) converter. The circuit diagram of a 7-bit d/a converter is shown in Fig. 7-3. This circuit is compatible with the a/d converter of Fig. 7-2.

The d/a converter is built with seven CMOS inverters and one 740 operational amplifier. The outputs of the inverters are summed

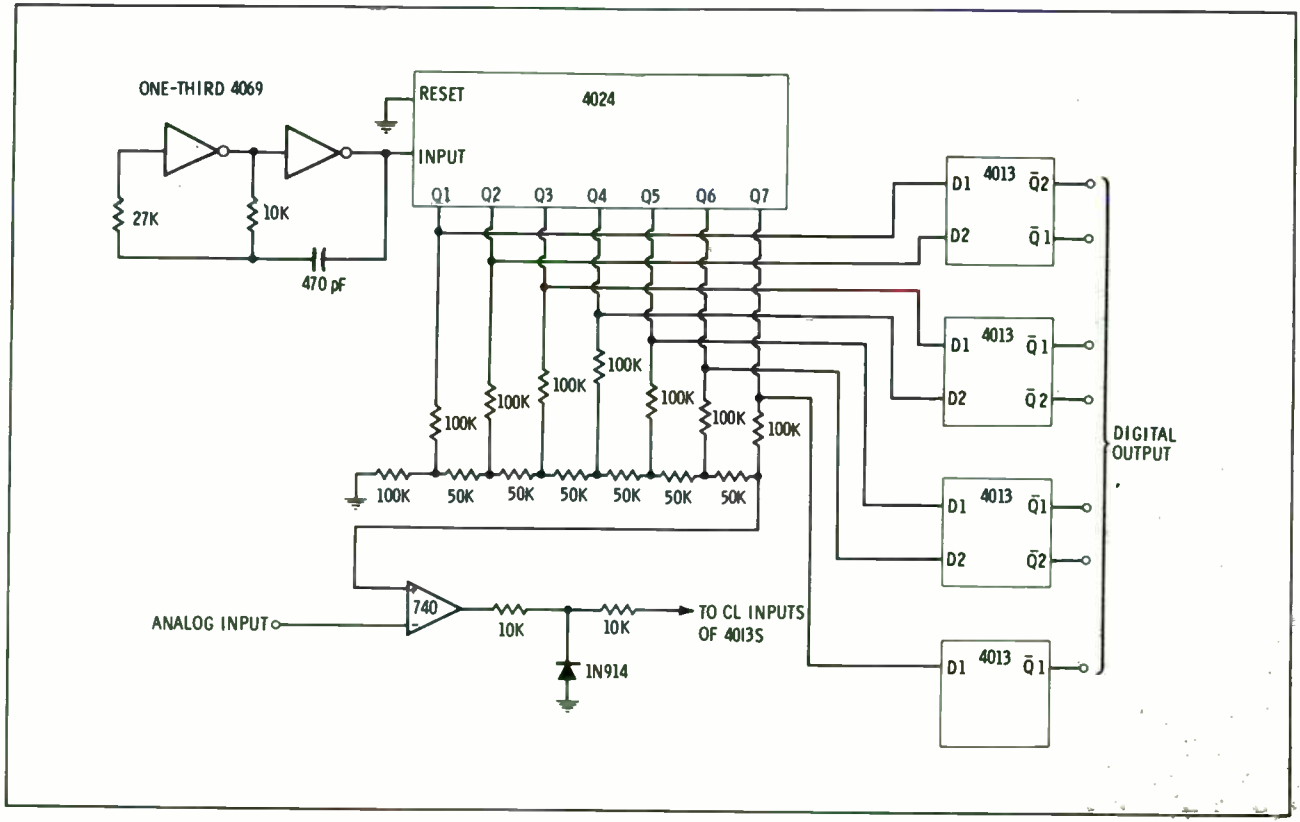


Fig. 7-2. A 7-bit a/d converter.

by a resistor network identical to the one used in the a/d converter. The operational amplifier is used as a voltage follower to buffer the output of the summing network. The analog signal produced at the output of the operational amplifier does not vary smoothly, but rather varies in small voltage steps. The original analog signal is reconstructed by smoothing the output of the operational amplifier with a low-pass filter. In the circuit of Fig. 7-3, a single RC low-pass filter is used to smooth the output.

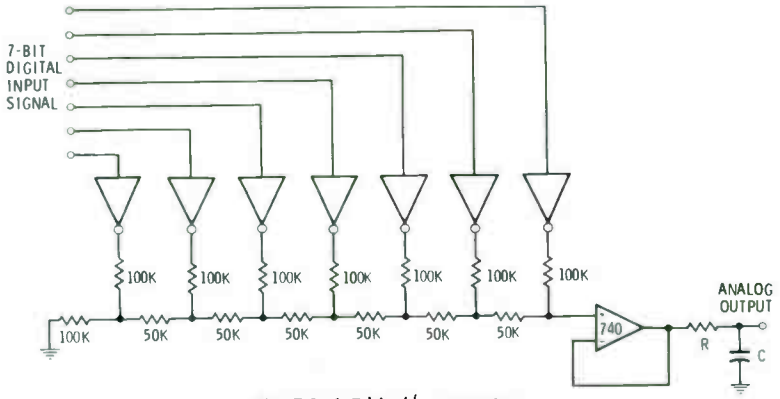


Fig. 7-3. A 7-bit d/a converter.

SAMPLE-AND-HOLD CIRCUITS

The function of a sample-and-hold circuit is to sample an analog signal at a particular time and to store that sample of the signal. Fig. 7-4 shows a sample-and-hold circuit built with a CMOS transmission gate, a storage capacitor, and a 740 operational amplifier. When the control signal to the transmission gate is high, the transmission gate is closed and the storage capacitor charges to the value of the analog input signal. When the control signal is low, the gate is open and the capacitor voltage appears at the output of the operational amplifier.

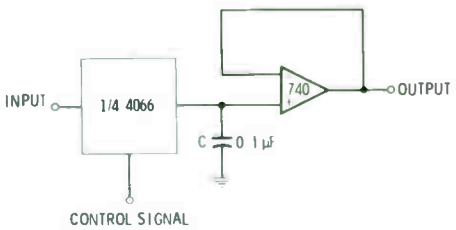


Fig. 7-4. A sample-and-hold circuit.

The capacitance of the capacitor, C, must be sufficiently small that the capacitor has time to charge to the value of the analog signal during the sampling interval. The capacitance must be sufficiently large to maintain its voltage to within required tolerances during the hold interval. The 0.1-μF capacitor shown in the circuit of Fig. 7-4 is a good choice for sample times greater than 100 microseconds and for hold times less than one second.

FREQUENCY MODULATION

A useful way to process an analog signal is to use it to modulate the frequency of some other signal. The time-varying amplitude of the analog signal is then coded as the time-varying frequency of the other signal. A frequency-modulated signal can be generated by a vco, as shown in Fig. 7-5. Here the vco of a 4046 CMOS IC is used to generate a digital signal that varies in frequency as the analog input signal varies in amplitude.

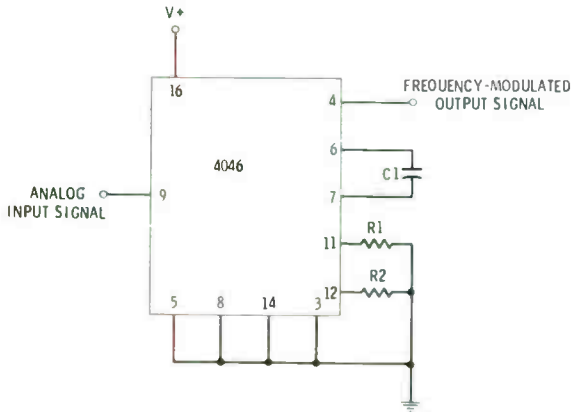


Fig. 7-5. A 4046 vco used for frequency modulation.

The frequency range of the vco is set by R1, R2, and C1. The minimum frequency of the vco (f_{min}) should be set several times higher than the highest-frequency component of the analog input signal. R2 and C1 determine the value of f_{min}, which can be found from Chart 6-2. The vco range (f_{max}-f_{min}) is set by R1 and C1, and can also be found from Chart 6-2.

FREQUENCY DEMODULATION

The process of reconstructing the original analog signal from a frequency-modulated signal is called frequency demodulation. A

phase-locked loop can be used for frequency demodulation, as shown in Fig. 7-6. Once the phase-locked loop is locked to the frequency of the input signal, the control voltage of the phase-locked loop (available at pin 10 of the 4046) is a replica of the analog signal originally used to modulate the input signal. The output at pin 10 of the 4046 IC is called the demodulator output.

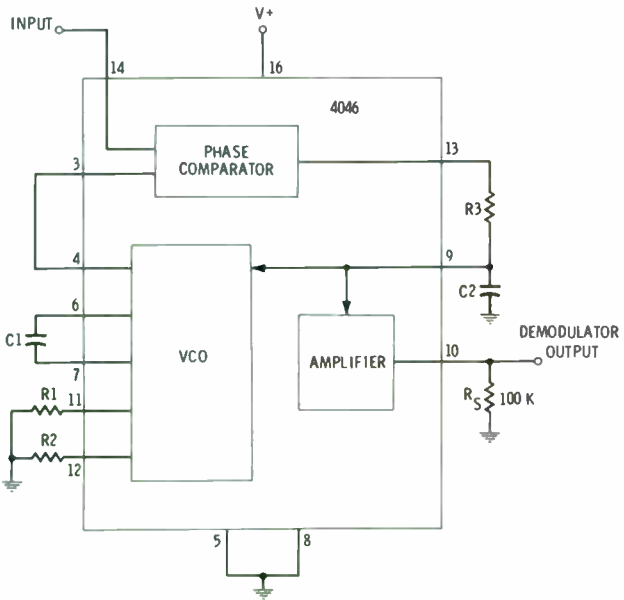


Fig. 7-6. A phased-locked loop used for frequency modulation.

The values of R1, R2, and C1 should be chosen (by referring to Chart 6-2) so that the vco has the same frequency range as the input signal. The values of R3 and C2 affect the stability of the phase-locked loop and affect the speed with which it responds to changes in the frequency of the input signal. The cutoff frequency of the low-pass filter formed by R3 and C2 is given by the formula,

$$f_{co} = \frac{1}{2\pi R3C2}$$

where,

f_{co} is in kilohertz,
2π is 6.283.
R3 is in kilohms,
C2 is in microfarads.

R3 and C2 should be chosen so that f_{co} is both higher than the highest-frequency component in the analog signal used to modulate the input signal and much lower than the minimum frequency of the vco (f_{min}).

MULTIPLEXING AND DEMULTIPLEXING

Multiplexing is a process used to transmit several different signals simultaneously over a single communications channel. The process of demultiplexing is used to reconstruct the original signals from a multiplexed signal.

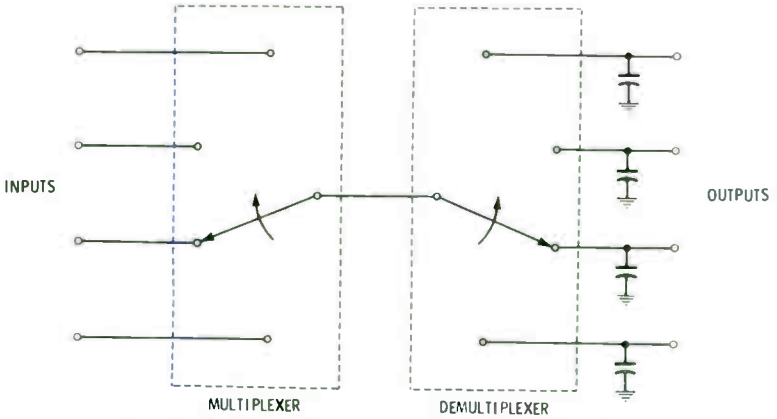
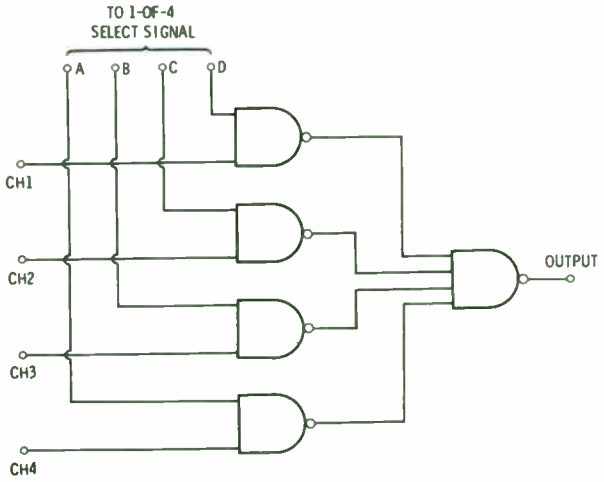


Fig. 7-7. The process of multiplexing and demultiplexing.

The processes of multiplexing and demultiplexing are illustrated in the diagram of Fig. 7-7. In this diagram a spinning rotary switch is used as a multiplexer, sampling each of the input signals in sequence. The demultiplexer is also a rotary switch, spinning in synchrony with the first switch. A small capacitor is used across each output to store the output samples. As long as the switches are spinning rapidly enough, the original input signals are accurately reproduced at the demultiplexer outputs. Of course, a spinning rotary switch is not very practical, but CMOS ICs can be effectively used to achieve the same result.

Digital Multiplexing and Demultiplexing

A four-channel digital multiplexer circuit is shown in Fig. 7-8A. The truth table for this multiplexer is given in Fig. 7-8B. A one-of-four select signal is used to sequentially sample each of the four input channels. When an input channel is sampled, its logic state appears at the output of the multiplexer.



(A) Logic diagram.

INPUT CHANNEL				SELECT INPUT				OUTPUT
1	2	3	4	A	B	C	D	
X	X	X	Q	1	0	0	0	Q
X	X	Q	X	0	1	0	0	Q
X	Q	X	X	0	0	1	0	Q
Q	X	X	X	0	0	0	1	Q

(B) Truth table.

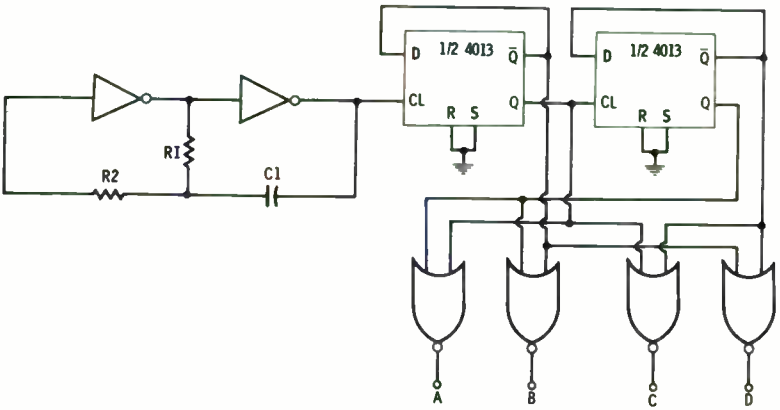
Fig. 7-8. A four-channel digital multiplexer.

Two circuits that can be used for generating the one-of-four select signal are shown in Fig. 7-9. The circuit of Fig. 7-9A uses an astable multivibrator to clock a two-stage binary counter. The outputs of the binary counter are decoded by a one-of-four decoder. The binary counter is built by using a 4013 CMOS IC, and the decoder is built by using a 4001 CMOS IC.

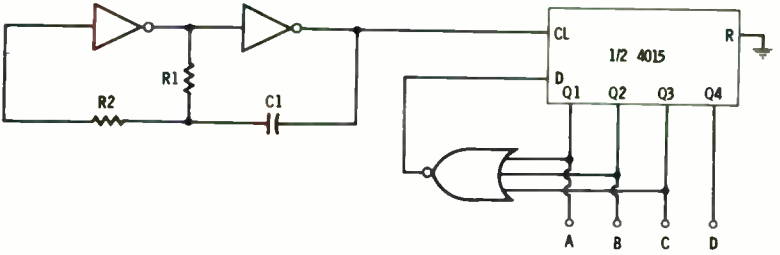
The circuit of Fig. 7-9B uses an astable multivibrator to clock a four-stage shift register. A three-input NOR gate is used to load the data into the shift register, loading a logic 1 pulse only when the first three outputs are logic 0. It is interesting to note that a four-stage shift register can also be used to generate a one-of-five select signal when used in conjunction with a four-input NOR gate. The four inputs of the NOR gate are connected to the four outputs of the shift register; the NOR gate output is used as the fifth select signal and as the data input to the shift register.

The circuit of a four-channel digital demultiplexer is shown in Fig. 7-10. A one-of-four select signal, generated by the circuit of Fig. 7-9B, is used to sequentially select each of the output chan-

nels. A D flip-flop can be used at each output to store the output sample of one channel while other channels are being selected. The connections required for using the D flip-flop are shown for channel 1 in Fig. 7-10.



(A) A binary counter with 1-of-4 decoding.



(B) A four-stage shift register.

Fig. 7-9. Circuits for generating a 1-of-4 select signal.

Analog Multiplexing and Demultiplexing

Since the input and the output of a CMOS transmission gate are interchangeable, the circuit of Fig. 7-11 can be used as either a multiplexer or a demultiplexer. The circuit works equally well with either analog or digital input signals. This multiplexer/demultiplexer circuit is built by using one 4016 CMOS IC. The operation of the circuit is analogous to the operation of the simple circuit of Fig. 7-7. The CMOS transmission gates are used as switches that are sequentially turned on or off by a one-of-four select signal. Either of the circuits of Fig. 7-9 can be used to generate the select signal.

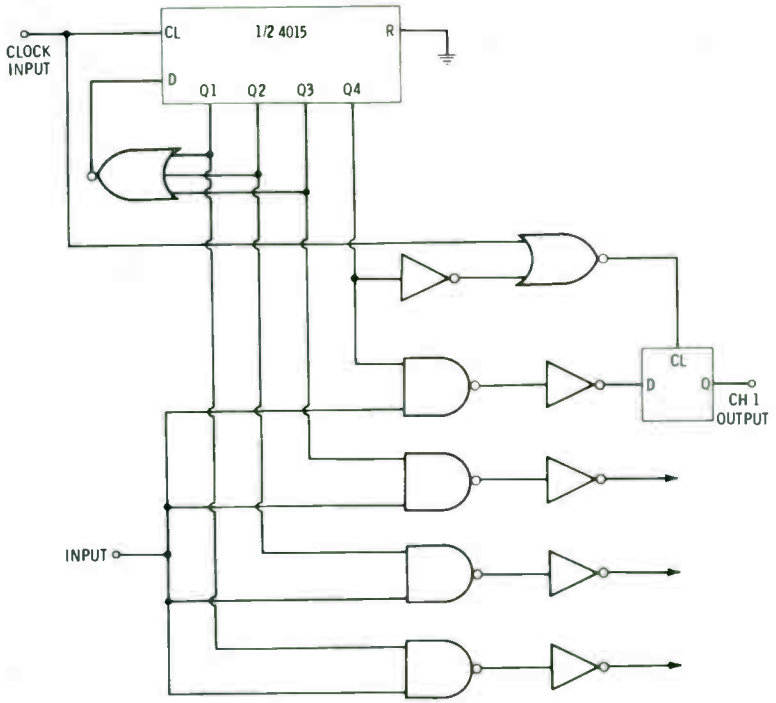


Fig. 7-10. A four-channel digital demultiplexer.

COMMUTATING FILTERS

A commutating filter is a digitally controlled analog bandpass filter. The operation of the commutating filter is illustrated in Fig. 7-12. In this illustration, the rotary switch is spinning with a commutating frequency of f_c. If the frequency of the input signal is equal to f_c, or to one of its harmonics, then each capacitor is always switched into the circuit during the same portion of the input waveform on each cycle of the rotary switch. Each capacitor then charges to the average value of the input voltage during its sample period. The resultant output waveform is a step approximation to the input waveform.

If the frequency of the input signal is not equal to f_c, or to one of its harmonics, each capacitor is switched into the circuit during different portions of the input waveform. As a result, the output of the filter is greatly diminished in amplitude. The commutating filter thus passes frequencies near f_c and its harmonics while greatly attenuating other input frequencies. The more capacitors

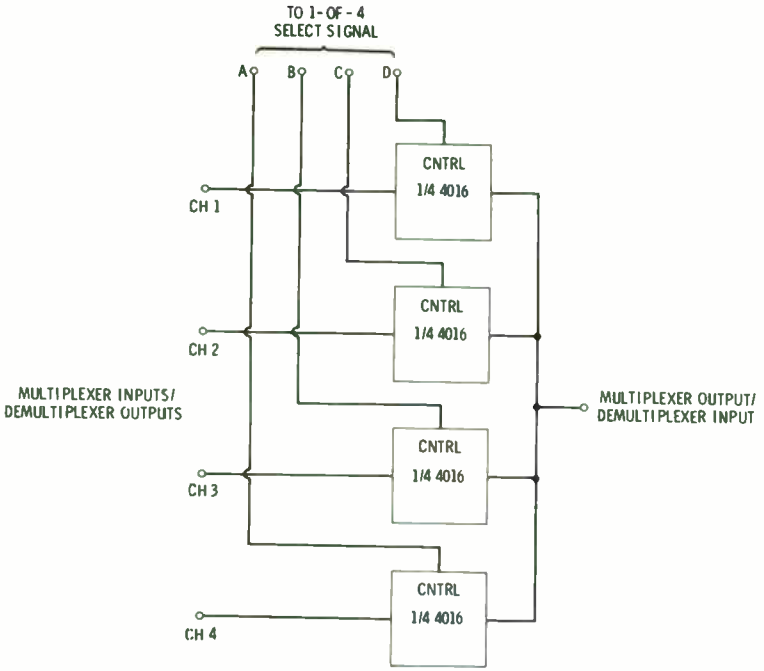


Fig. 7-11. A analog multiplexer/demultiplexer.

and switch positions used with the commutating filter, the more ideal its performance.

A practical commutating filter can be built by replacing the spinning rotary switch with an analog multiplexer. The circuit diagram of a commutating filter built with an eight-channel analog multiplexer is shown in Fig. 7-13. The 4022 IC (a four-stage Johnson counter with integral one-of-eight decoding) is used to

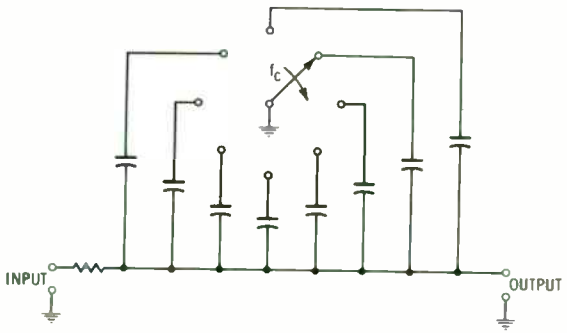


Fig. 7-12. The operation of a commutating filter.

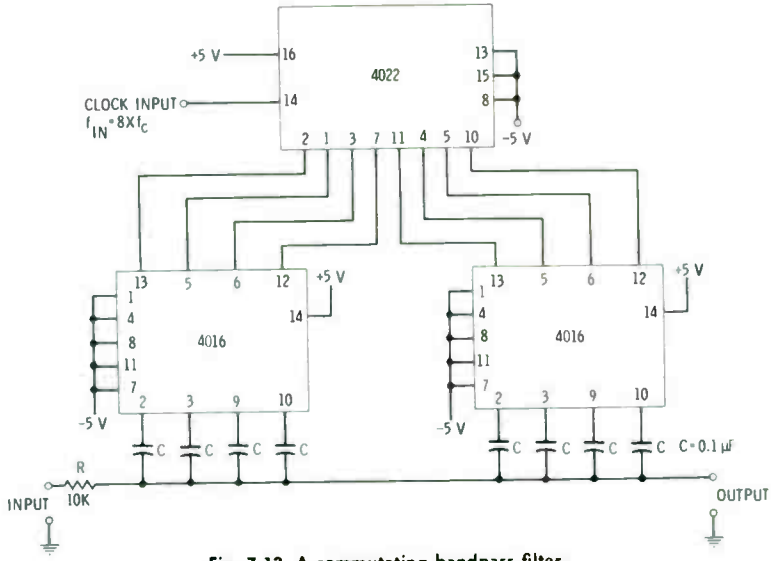


Fig. 7-13. A commutating bandpass filter.

generate a one-of-eight select signal, and two 4016 ICs are used to build the multiplexer itself. The clock input frequency, f_{in}, is set at eight times the desired commutating frequency.

MICROPROCESSORS

One approach to the design of information-processing systems is to wire together a set of ICs to perform a particular function. An alternative approach is to program a *microprocessor* to perform the function. The microprocessor is programmed with a set of instructions coded as a sequence of logic 1's and 0's in a memory IC. In the first approach the system performance is specified by *hardware* (the ICs and their interconnections). In the second approach the system performance is specified by *software* (the instructions stored in the IC memory).

A microprocessor is in many ways like a miniature computer built with just a few CMOS, MOS, or bipolar ICs. A microprocessor is usually built with one central processing unit (CPU) IC and one or more memory ICs. Two types of memory ICs can be used. A *read-only memory* (ROM) is a permanent memory which, once programmed, cannot be altered. The information stored in a *random-access memory* (RAM), on the other hand, can be changed at will. The 4061 CMOS IC shown in the appendix is an example of a 256-bit RAM.

Microprocessors are now sufficiently inexpensive that more and more information-processing systems are designed with software rather than with hardware. The great advantage of microprocessors is that with just a few standard microprocessor ICs, a great number of different capabilities can be realized by simply changing the instructions stored in an IC memory.

8

CMOS Systems

The inherent capabilities of CMOS ICs become clearly evident when CMOS ICs are employed in practical electronic system design. The CMOS IC circuits presented in this chapter illustrate the great versatility of CMOS devices in system design applications.

A CODE-PRACTICE RADIO TRANSMITTER

Fig. 8-1 shows the circuit diagram of a code-practice radio transmitter designed to transmit on the a-m broadcast band. The transmitter is a very low-powered one, in compliance with FCC regulations, but it is capable of broadcasting Morse code across a room to a standard a-m receiver.

NOR gates 1 and 2 are used in an audio-frequency astable multivibrator circuit. NOR gates 3 and 4 are used in a radio-frequency astable multivibrator circuit. The rf carrier produced by the second astable is tone-modulated by the output of the first astable. The frequency of the rf carrier is set by R5 and C2, while the frequency of the audio tone is set by R3 and C1.

A DELAYED-RELEASE TIMER

The circuit of Fig. 8-2 is designed to be used as a delayed-release timer for electrically driven cameras (the Polaroid SX-70, for example). The unit is connected to the remote-release contacts of the camera. After SW1 is depressed, there is approximately a

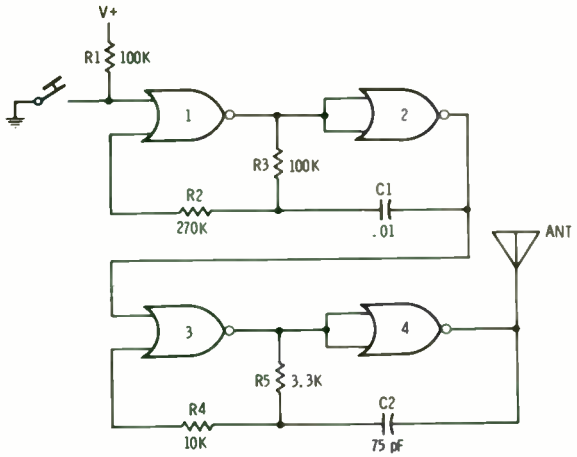


Fig. 8-1. A code-practice radio transmitter.

7-second delay before the camera is tripped, which gives the photographer a chance to put himself in the picture.

This circuit illustrates how the power for a CMOS circuit can be "stolen" from another circuit. The power for this circuit is taken from the camera remote-release contacts through R4 and D1, and is stored in C3. NOR gates 1 and 2 are used in a monostable circuit, as are gates 3 and 4. When SW1 is depressed, a

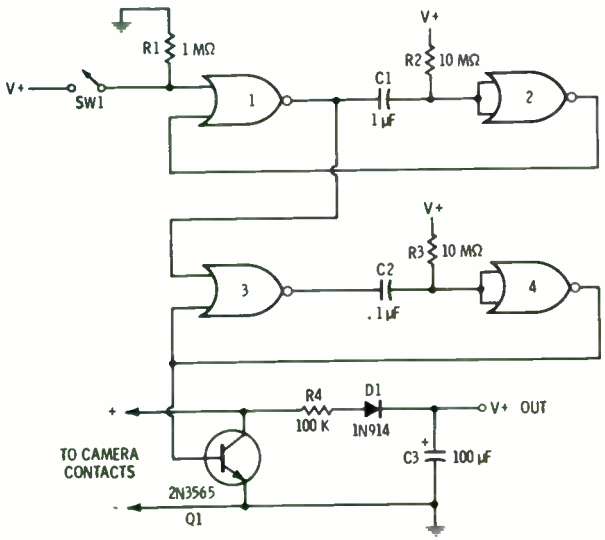


Fig. 8-2. A delayed-release camera timer.

negative-going pulse, approximately 7 seconds in duration, is generated at the output of NOR gate 1. The end of this pulse triggers the monostable formed by gates 3 and 4. The pulse generated by this monostable turns on Q1, which trips the camera.

A CAPACITANCE METER

Fig. 8-3 shows the circuit diagram of a capacitance meter that can be built with one 4001 CMOS IC. NOR gates 1 and 2 are used in an astable circuit, and NOR gates 3 and 4 are used in a monostable circuit. The value of the capacitor being measured, C_x, determines the output-pulse length of the monostable.

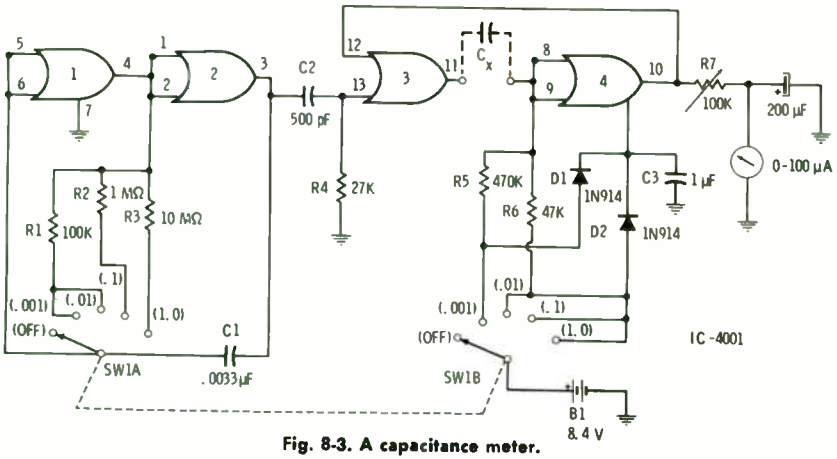


Fig. 8-3. A capacitance meter.

The monostable is repetitively triggered by the astable at a rate determined by the setting of the range switch, SW1. To speed the recovery time of the monostable, diodes D1 and D2 are used to decrease the IC supply voltage to approximately .6 volt below the battery voltage, as shown in the monostable circuit of Fig. 6-6. Since the output-pulse length of the monostable is proportional to C_x, the reading on the 100-μA meter is proportional to this capacitance. Potentiometer R7 is used to calibrate the meter to read capacitance directly, times a scale factor determined by the setting of the range switch.

A MULTICHANNEL ADAPTER FOR OSCILLOSCOPES

The circuit of Fig. 8-4 can be used to display up to four different signals simultaneously on the screen of a single-channel oscilloscope. Four potentiometers are provided to position each of the

four traces on the screen. The operation of the circuit is similar to that of the analog multiplexer discussed in Chapter 7.

A circuit similar to the one of Fig. 7-9B is used to generate a 1-of-4 select signal. Eight transmission gates, in two 4016 packages, are activated a pair at a time. Each pair of transmission gates will switch one input channel and one position-control voltage to

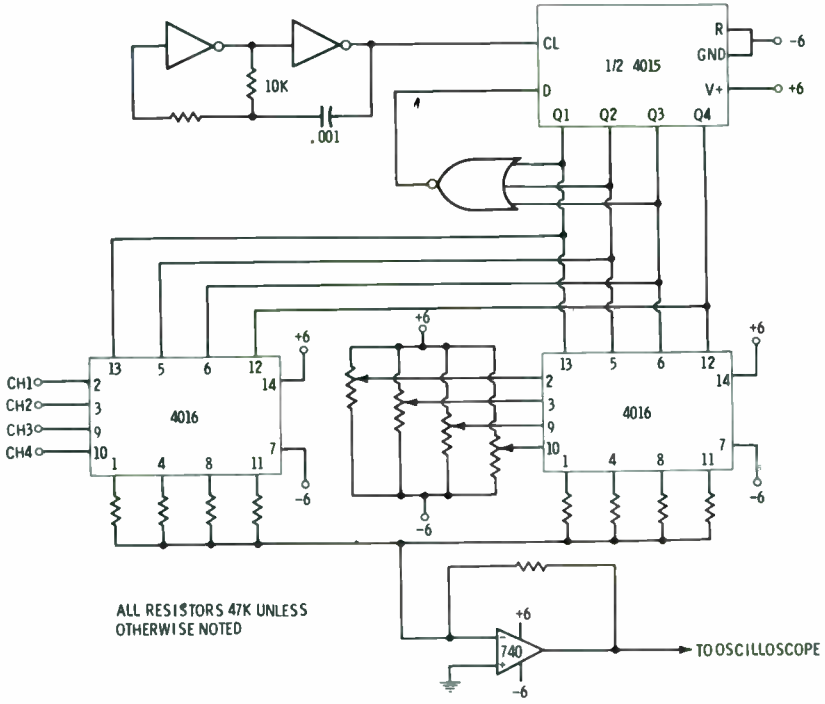


Fig. 8-4. A multichannel adapter for oscilloscopes.

the input of a 740 operational amplifier. Each of the four input channels and position-control voltages are scanned in rapid succession. The output of the operational amplifier is displayed on the oscilloscope. As long as the sweep rate of the oscilloscope is slow by comparison to the scan rate of the 1-of-4 select signal, all four channels appear simultaneously on the oscilloscope screen.

A CMOS ELECTRONIC WRISTWATCH

The power capability of CMOS ICs is dramatically demonstrated by the CMOS electronic wristwatch. A block diagram of a CMOS electronic-wristwatch circuit is shown in Fig. 8-5. This wristwatch circuit is built by using two silicon-gate CMOS ICs,

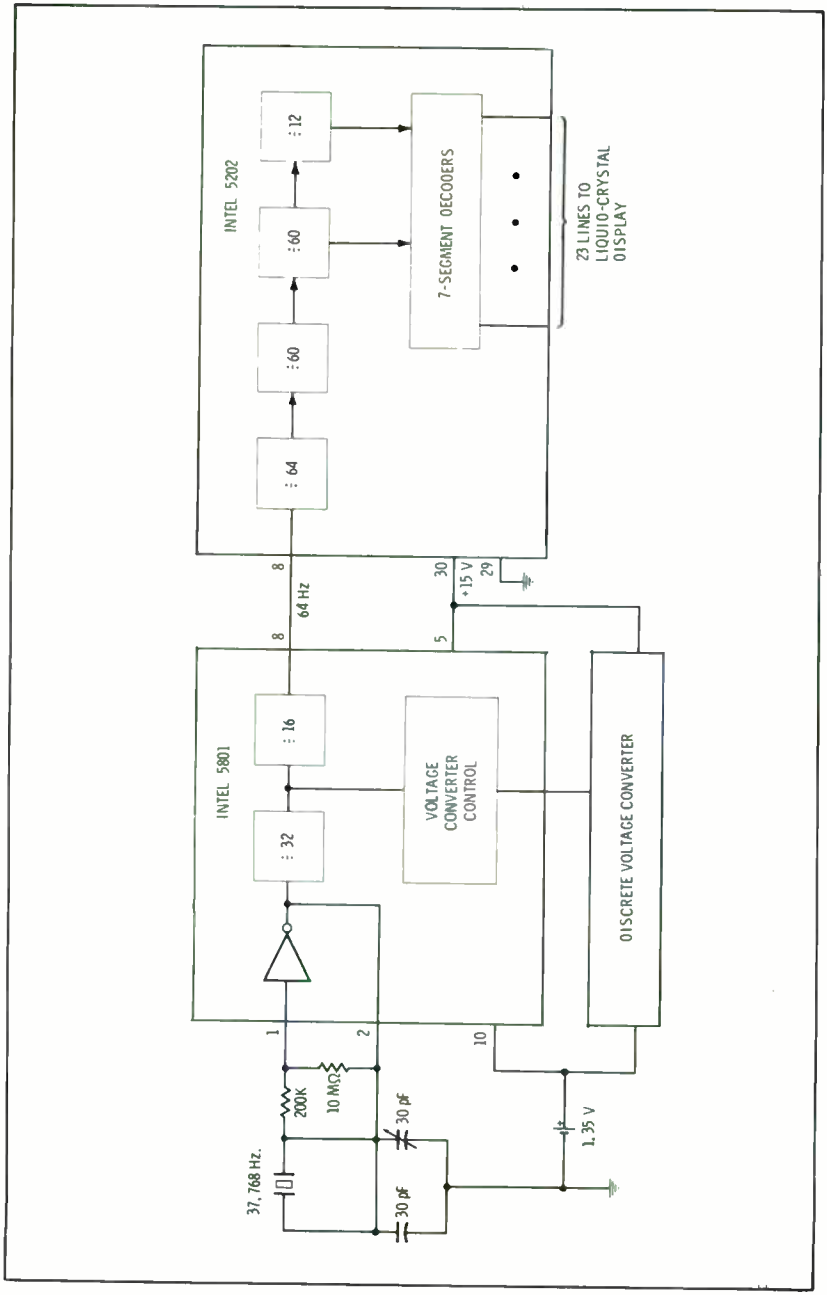


Fig. 8-5. Block diagram of electronic wristwatch built with two silicon-gate CMOS ICs.

an Intel 5801 and a 5202. Silicon-gate CMOS construction is used to achieve low-voltage operation, low power consumption, and high circuit density, as discussed in Chapter 3.

The time base for the wristwatch circuit is a 32,768-Hz quartz-crystal oscillator circuit like the circuit of Fig. 6-9. The CMOS inverter for the crystal oscillator is contained within the 5801 IC. The 5801 divides the output frequency of the crystal oscillator first by 32 and then by 16. The 1024-Hz output of the divide-by-32 circuit is used by a separate voltage-converter circuit. The voltage-converter circuit generates a 15-volt signal for the liquid-crystal digital display. The 64-Hz output of the divide-by-16 circuit is used as the reference input frequency to the 5202.

A divide-by-64 circuit in the 5202 generates a 1-Hz signal from the 64-Hz input signal. One divide-by-60 counter generates a "minutes" signal from the 1-Hz signal, and another divide-by-60 counter generates an "hours" signal from the minutes signal. A divide-by-12 counter cycles the hours-count every 12 hours.

Courtesy Microma Inc.

Fig. 8-6. A CMOS wristwatch.

The outputs of the counters are decoded by a set of seven-segment decoders. The outputs of these decoders are used to drive the wristwatch digital display. (A photograph of an electronic wristwatch using CMOS ICs is shown in Fig. 8-6.) Because of the low power consumption of the CMOS ICs used in this circuit, a tiny 1.35-volt battery will power the wristwatch for over a year.

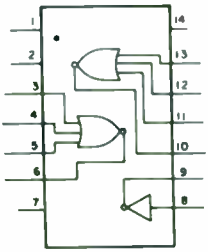
APPENDIX

Pin Diagrams of Selected CMOS

The CMOS ICs in this appendix are listed by their generic number. The generic number is often prefixed to indicate the manufacturer. A 4001 CMOS IC from RCA, for example, is labeled as a CD4001, while the same IC from Motorola is labeled as an MC14001. In addition, these numbers are often suffixed to indicate package style. All pin diagrams in this appendix are courtesy of RCA Corp.

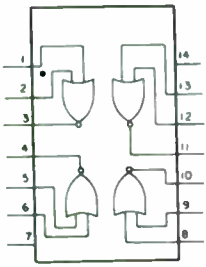
Only the most popular CMOS ICs are listed in this appendix. Individual IC manufacturers should be contacted for comprehensive listings of current CMOS ICs and for detailed IC specification sheets.

NOTE: 14-pin ICs: V+ to pin 14; ground (or V−) to pin 7.
16-pin ICs: V+ to pin 16; ground (or V−) to pin 8.



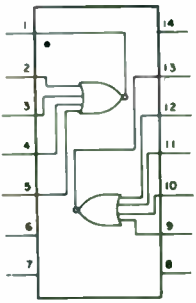
4000

Two 3-input NOR gates and one inverter.



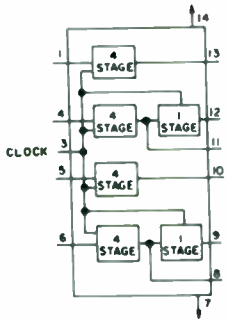
4001

Four 2-input NOR gates



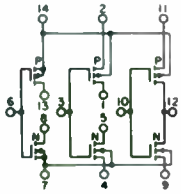
4002

Two 4-input NOR gates



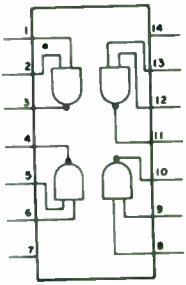
4006

18-stage shift register



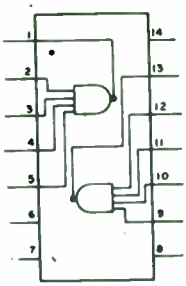
4007

Two CMOS transistor pairs and one inverter



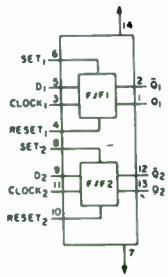
4011

Four 2-input NAND gates



4012

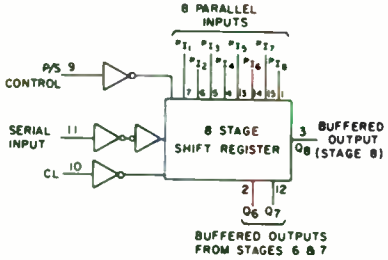
Two 4-input NAND gates



4013

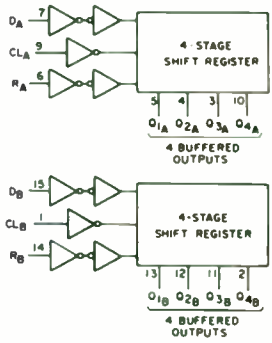
Two D flip-flops

NOTE: 14-pin ICs: V+ to pin 14; ground (or V−) to pin 7.
16-pin ICs: V+ to pin 16; ground (or V−) to pin 8.



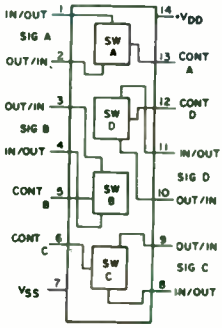
4014

8-stage shift register
(Synchronous inputs)



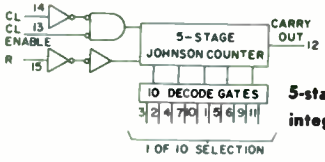
4015

Two 4-stage shift registers



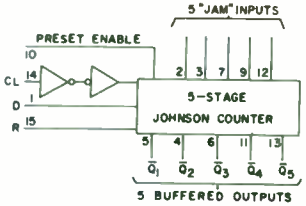
4016

Four transmission gates



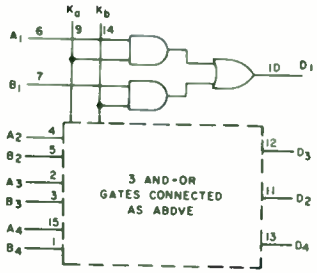
4017

5-stage Johnson counter with integral 1-of-10 decoding



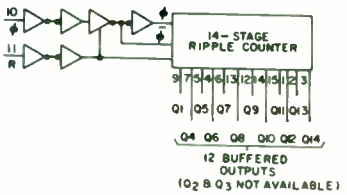
4018

Presettable 5-stage Johnson counter



4019

Four AND-OR gates



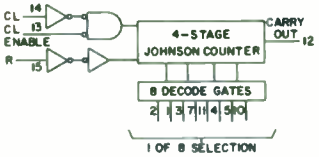
4020

14-stage binary counter

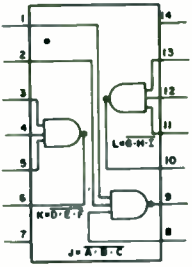
4021 See 4014

8-stage shift register
(Asynchronous inputs)

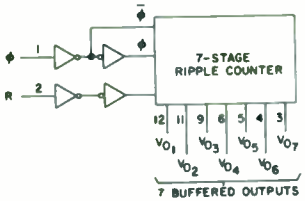
NOTE: 14-pin ICs: V+ to pin 14; ground (or V−) to pin 7.
16-pin ICs: V+ to pin 16; ground (or V−) to pin 8.



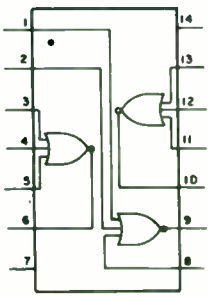
4022
4-stage Johnson counter with integral 1-of-8 decoding



4023
Three 3-input NAND gates

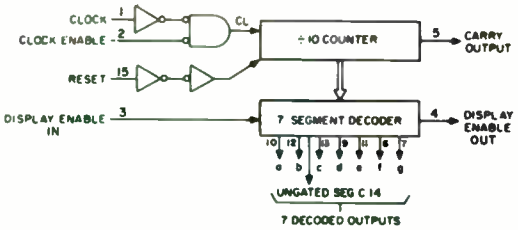


4024
7-stage binary counter



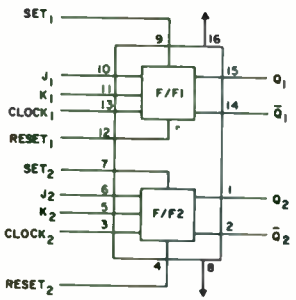
4025
Three 3-input NOR gates

NOTE: 14-pin ICs: V+ to pin 14; ground (or V−) to pin 7.
16-pin ICs: V+ to pin 16; ground (or V−) to pin 8.



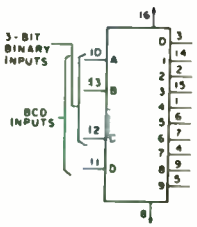
4026

5-stage Johnson counter with integral seven-segment decoding



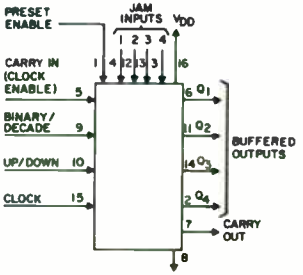
4027

Two JK flip-flops



4028

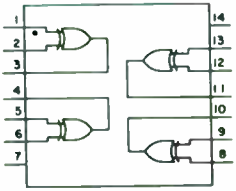
bcd-to-decimal decoder



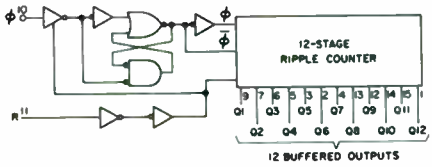
4029

Presettable up/down counter

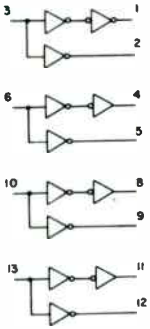
NOTE: 14-pin ICs: V+ to pin 14; ground (or V−) to pin 7.
16-pin ICs: V+ to pin 16; ground (or V−) to pin 8.



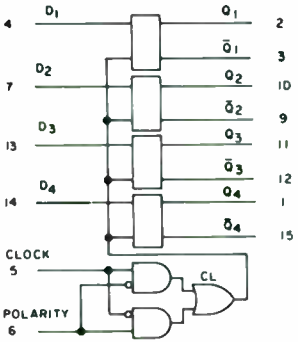
4030
Four exclusive-OR gates



4040
12-stage binary counter



4041
Four inverting/noninverting buffers



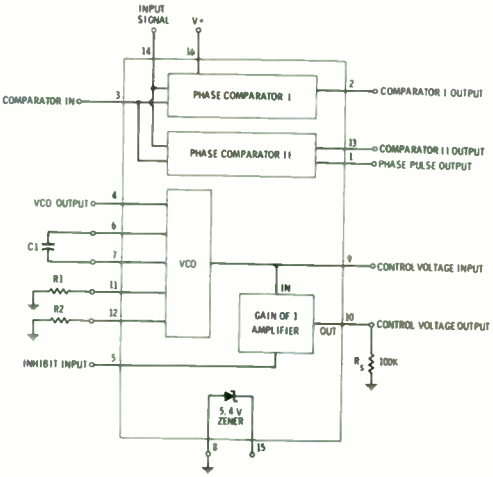
4042
Four clocked D flip-flops

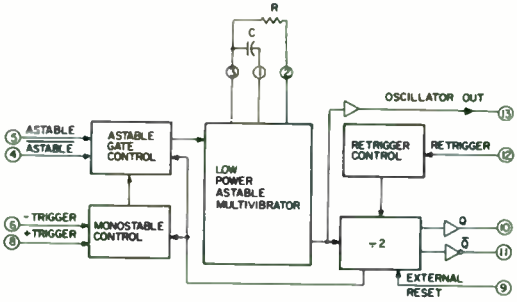
NOTE: 14-pin ICs: V+ to pin 14; ground (or V−) to pin 7.
16-pin ICs: V+ to pin 16; ground (or V−) to pin 8.

See text
Fig. 6-19

4046

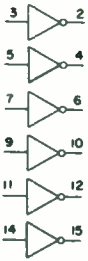
Phase-locked loop





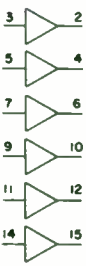
4047

Monostable/astable multivibrator



4049

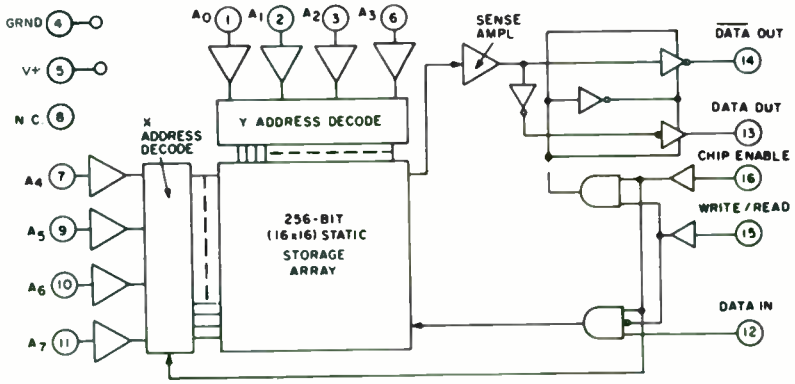
Six inverting buffers
(replaces type 4009)



4050

Six noninverting buffers
(replaces type 4010)

NOTE: 14-pin ICs: V+ to pin 14; ground (or V−) to pin 7.
16-pin ICs: V+ to pin 16; ground (or V−) to pin 8.



4061

256-bit random-access memory (RAM)

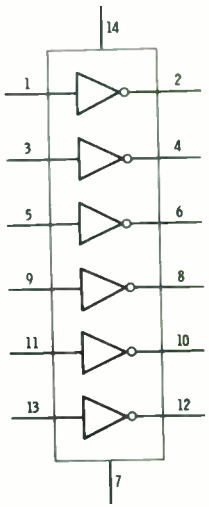
Note exception to power-supply pin convention

See 4016

4066

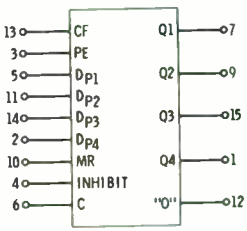
Four transmission gates

(lower on resistance than 4016)



4069

Six inverters



4522

Presettable bcd counter

See 4522

4526

Presettable binary counter

Index

Understanding
CMOS INTEGRATED CIRCUITS

by Roger Melen and Harry Garland

Complementary metal-oxide-semiconductor (abbreviated CMOS) integrated circuits are monolithic integrated circuits containing p- and n-channel enhancement-type MOS field-effect transistors. Available since 1968, they are now being fabricated by a rapidly growing list of manufacturers and have received wide-spread publicity because of their use in electronic calculators and digital wristwatches.

The authors take a solid building-block approach to their subject. The presentation is thorough, easy to absorb, and devoid of jargon. The many illustrations are meaningful and amplify the clear text.

The first chapter introduces basic digital ICs (integrated circuits)—logic gates, flip-flops, and decoders. These are then used in practical application, the design of a CMOS digital timer.

A "refresh" discussion in Chapter 2 of semiconductor physics progresses to the diode, the MOS transistor, and CMOS parasitic devices. Chapter 3, CMOS fabrication technology, includes the five popular variations as well as the basic fabrication techniques. The monolithic CMOS design and CMOS design rules of Chapters 4 and 5 encompass CMOS transmission gates and memories, as well as interfacing and debugging. Frequency generators, Chapter 6, covers multivibrators, oscillators, phase-locked loops, and divide-by-n circuits. Chapter 7 discusses analog-to-digital and digital-to-analog conversion, multiplexing, and microprocessors.

Then the authors put it all together in the last chapter on CMOS systems—a code-practice radio transmitter, a delayed-release timer (for cameras), a capacitance meter, an oscilloscope multichannel adapter, and a CMOS electronic wristwatch.

All the CMOS ICs used as examples throughout the book are currently available in the marketplace. The appendix has the pin diagrams of 33 selected CMOS ICs.

Although the reader is assumed to have a basic grasp of electronics, business-oriented persons interested in new technology will find this book just as enlightening as electronics people "in the business."

ABOUT THE AUTHORS

Roger Melen has a B.S.E.E. degree from Chico State College, Chico, California, and M.S.E.E. and Ph.D. degrees from Stanford University, Stanford, California. He is currently associate director of the Stanford University Integrated Circuits Laboratory and a consultant to several major electronics manufacturers. Dr. Melen writes technical articles for both popular magazines and professional journals.

Harry Garland has a B.A. degree in mathematics from Kalamazoo College, Kalamazoo, Michigan, and a Ph.D. degree in biophysics from Stanford University, Stanford, California. He holds an advanced-class amateur radio license and a first-class commercial radiotelephone license. Dr. Garland is currently assistant chairman of the Electrical Engineering Department at Stanford University.

HOWARD W. SAMS & CO., INC.
THE BOBBS-MERRILL CO., INC.

21129
$4.95 (In Canada $5.95)