Transistor Switching and Sequential Circuits

BY

JOHN J. SPARKES

PERGAMON PRESS

OXFORD · LONDON · EDINBURGH · NEW YORK

TORONTO · SYDNEY · PARIS · BRAUNSCHWEIG

Pergamon Press Ltd., Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1
Pergamon Press (Scotland) Ltd., 2 & 3 Teviot Place, Edinburgh 1
Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford,
New York 10523
Pergamon of Canada Ltd., 207 Queen's Quay West, Toronto 1
Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street,
Rushcutters Bay, N.S.W. 2011, Australia
Pergamon Press S.A.R.L., 24 rue des Écoles, Paris 5e
Vieweg & Sohn GmbH, Burgplatz 1, Braunschweig

First edition 1969
Library of Congress Catalog Card No. 68-30842

Printed in Hungary

08 012981 1 (flexicover)
08 012982 X (hard cover)

Preface

NOWADAYS, in the field of digital circuitry, there is a great variety of elementary but challenging problems, which are suitable for students to tackle, and which introduce them to most of the basic ideas involved in the construction of computers, coders, pulse communication systems, instrumentation, automation and so on. This book is intended to help students, and others beginning in this field, to design, construct and interconnect digital or switching circuits and so begin to accumulate the experience which is still needed in large measure in order to construct the large and relatively complex digital systems manufactured today.

In order to keep the book short and yet sufficiently detailed to be useful I have selected for description what I believe to be the most generally useful and successful circuits and have described a reasonably straightforward design procedure for sequential circuits. Refinements of both are known to exist and are referred to at the appropriate places in the book. The circuits as they stand are, however, reliable and versatile and can be used to construct a wide variety of useful electronic equipment.

There are two parts to the book. The first is concerned with how to generate the kinds of waveforms needed in digital circuits, principally square waves, ramps and delays. The second half is concerned with gates and flip-flops and how to interconnect them in order to achieve a variety of logical or sequential functions. Several methods of achieving a particular logical function are described and contrasted. It is not necessary to read the first part of the book in order to understand the second.

The design of sequential circuits is dealt with in Chapters 6 and 7. The subject is first introduced in Chapter 6 by means of an example and the procedures used are then examined in more detail and in greater generality in Chapter 7. However, the reader may find it helpful to return to Chapter 6 again after reading Chapter 7 in order better to understand the procedures involved. In this way it is hoped that the rather numerous but quite simple detailed considerations which comprise sequential circuit design have been described in a reasonably condensed form.

J. J. S.

Principal Symbols Used

α_N, α_I	Normal, inverse common base current gains
β	Common emitter current gain ($\delta I_C/\delta I_B$ with V_{CE} constant)
β_S	On-demand current gain (see p. 18)
τ_C, τ_S	Collector time factor, saturation time factor
φ	Built-in potential difference in a *pn* junction
c_{tc}, c_{te}	Collector, emitter transition region capacitance
$f_T = \omega_T/2\pi$	Current gain–bandwidth product
k	Boltzmann's constant = $1{\cdot}38\times10^{-23}$ J/K°
q	Electronic charge = $1{\cdot}6\times10^{-19}$ C
r_e	Emitter junction slope resistance kT/qI_E
r_E	Linearized, large signal, junction resistance
t_d, t_f, t_r, t_s	Delay time, fall time, rise time, saturation time
A_V	Voltage gain
C_C	Large signal (linearized) collector capacitance
I_{CBO}, I_{BX}, I_{EX}, etc.	Cut-off currents (see pp. 9, 10)
$I_{C(ON)}$, $\hat{I}_{C(ON)}$	D.C., peak transient values of I_C of a conducting transistor
I_E, I_B, I_C	Emitter, base, collector currents (d.c.)
Q_B, Q_{BS}	Active base charge, saturation base charge
Q_{ON}, Q_{OFF}	Total charge required to turn on, turn off a transistor
Q_{VC}, Q_{VE}, Q_{VD}	Charge associated with reverse biased junctions—collector, emitter, both collector and emitter
T	Temperature in °K
V_E, V_B, V_C in circuits	The potential of the emitter, base, collector with respect to earth
V_E, V_C in Ebers–Moll equations	The emitter, collector junction voltages, (forward bias positive)
V_{XY}	The potential of X with respect to Y

CHAPTER 1

Properties of Transistors

THE depth of understanding of transistor action needed for circuit design varies very much with the type of circuit and kind of problem being considered. At its simplest the transistor can be thought of as just a current amplifying device, and for many circuit design problems this is quite adequate. But if, for example, it is necessary to calculate how much delay there is between a change in input current and the proportional change in output current a much more detailed understanding is needed. In this chapter the properties of transistors which are necessary for the thorough design of quite straightforward circuits are summarized for use later in the book.

A Brief Physical Description of Transistors

A transistor comprises two *pn* junctions placed close enough together in a single crystal of semiconductor for the currents through each to be influenced by the bias applied to the other.

The current from one junction to the other is carried by minority carriers which diffuse across the base region between the junctions if the region is homogeneous. If the base region has a *graded* density of impurities the carriers are helped across by an electric field built into the structure of the device.

The magnitude of the current depends upon the densities of the minority carriers at the two ends of the base region, and these

carrier densities are determined by the voltages applied to the junctions. With no bias applied the minority carrier density is everywhere its equilibrium value. Next to a *forward* biased junction (i.e. p region biased positively with respect to the n region) the density of minority carriers is increased, and next to a reverse biased junction the density is decreased. If in a *pnp* transistor, p_{n0} is the equilibrium hole density in the base region† and p_{n1} is the hole density next to the emitter when it is biased, it can be shown that[2]

$$p_{n1}/p_{n0} = e^{qV_E/kT} \tag{1.1}$$

where V_E is the voltage across the emitter junction and where q/kT is the electronic charge divided by Boltzmann's constant and the absolute temperature. When $T = 300°\text{K}$ $kT/q \approx 25$ mV.

Thus when the emitter is forward biased (V_E positive) the hole density is increased so that $p_{n1} > p_{n0}$. If V_E is negative $p_{n1} < p_{n0}$.

A similar equation with V_C replacing V_E applies to the densities next to the collector junction.

In Fig. 1.1 are diagrams of various operating conditions of a *pnp* transistor. It is assumed that the transistor has a homogeneous base region (so that there is no electric field and the minority carriers diffuse across the base region) and that therefore the *emitter–collector current through the base is proportional to the minority carrier gradient.*

In (a) the emitter junction is forward biased so that the hole density next to the emitter is high. The collector junction is reverse biased so the hole density next to the collector is low. Thus holes diffuse from emitter to collector. This is the normal *Active Region of Operation* of the transistor.

In (b) the emitter is reverse biased and the collector forward biased. Now the carrier density gradient is in the opposite direction and the holes flow from collector to emitter. This is the

† Letters p, n stand for hole, electron densities. Subscripts p, n indicate p-type, n-type regions occupied by the holes or electrons. Subscripts 0 or 1 indicate equilibrium or biased conditions.

Inverse Active Region of Operation. It is not often used but has some important properties. Its principal importance here is to emphasize that a transistor, unlike a vacuum tube, can operate either way round.

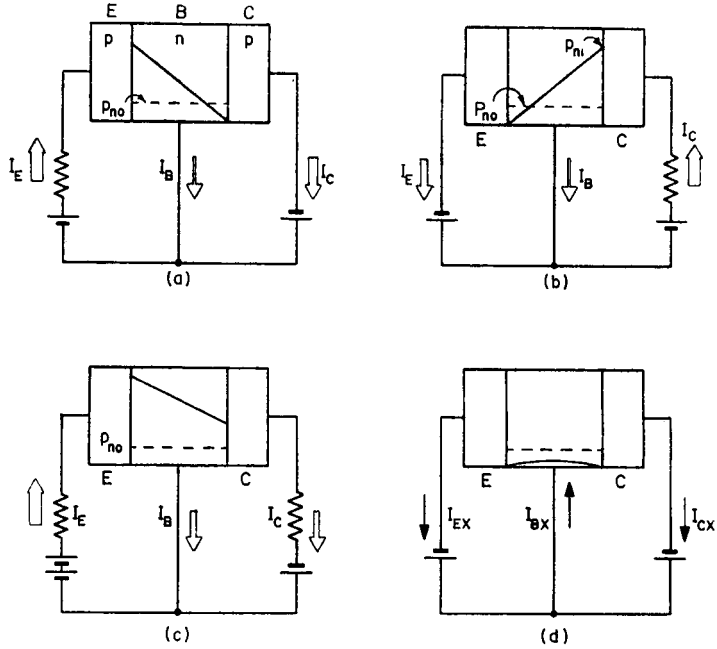


FIG. 1.1. Base region, minority carrier density distributions and the corresponding terminal currents (with relative magnitudes indicated) for (a) active region of operation; (b) inverse active region of operation; (c) saturation region of operation; (d) cut-off region of operation.

In (c) both junctions are forward biased, but since the emitter has a larger forward bias (the carrier density next to the emitter is larger than that next to the collector) the current flows down the gradient from emitter to collector. This is the *Saturation Region of Operation* and is very important in switching circuits.

Notice particularly that a forward bias voltage on both junctions *does not result in a forward current through both junctions*; the current through the collector junction is still a *reverse* current despite its forward bias voltage. The voltage applied to a junction only determines the minority carrier *density* near it; that is, it does not wholly determine the carrier gradient upon which the current depends. It takes two densities to make a density gradient, one is fixed by the emitter bias, the other by the collector.

In (d) both junctions are reverse biased and the transistor is in the *Cut-off Region of Operation*. The carrier density is everywhere reduced well below its equilibrium value, the gradients are small and if it were not for the thermal generation of carriers, negligible current would flow. In practice, however, the small currents which do flow as a result of thermal generation cannot always be neglected.

The d.c. Equations of a Transistor

The interdependence of the terminal currents and voltages of all the above four regions of operation can be summarized in three important equations.

In any isolated *pn* junction, it is a consequence of the dependence of minority carrier density on junction voltage that in the steady state the current I through the junction is related to the voltage V across it by the equation:

$$I = I_S(e^{qV/kT} - 1), \tag{1.2}$$

where I_S is a temperature sensitive parameter of the junction. (Actually it is the rate at which minority carriers are generated thermally within the device.)

In a transistor this diode equation for each junction is modified by the presence of the other *pn* junction. Thus considering collector and emitter currents separately,

$$I_C = I_{CB0}(e^{qV_C/kT} - 1) - \alpha_N I_E \tag{1.3}$$

and

$$I_E = I_{EB0}(e^{qV_E/kT} - 1) - \alpha_I I_C, \tag{1.4}$$

where V_E, V_C denote the junction bias voltages, where α_N is the proportion of the emitter current which reaches the collector, and α_I is the proportion of the collector current which reaches the emitter. I_{EB0} and I_{CB0} are the transistor parameters which correspond to I_S of the diodes. The subscripts are explained shortly.

It can also be shown[1, 2] that as a consequence of the interdependence of the two junctions, α_N and α_I are related by the equation

$$\alpha_N I_{EB0} = \alpha_I I_{CB0}. \tag{1.5}$$

These three equations are known as the Ebers–Moll[1] equations.

In order to preserve the symmetry and generality of these equations a rather special sign convention is used in which

(1) All currents flow into the transistor ($I_E + I_B + I_C = 0$).
(2) Forward bias voltages are positive; thus the subscript simply names the junction across which the voltage appears.
(3) α_N and α_I are positive numbers.
(4) I_{CB0} and I_{EB0} are *positive for pnp transistors* and *negative for npn transistors.*

It is convention (2) which can cause difficulty when the equations are related to practical circuits. "Forward bias" always means that the *p* region is positive with respect to the *n* region. This means that, for example, *for V_E to be positive in the equation, a positive voltage, with respect to the base, must be applied to the emitter of a pnp transistor, but a negative voltage, with respect to the base, must be applied to the emitter of an npn transistor.* That is, if V_E is positive V_{EB} is positive for a *pnp* transistor but negative for an *npn* one.

It is worth noting that these special conventions coincide with the more usual ones (in which currents and voltages are positive

if standard meters measure them to be so) for *pnp* transistors. That is, $V_E = V_{EB}$ and $V_C = V_{CB}$.†

Some Particular Solutions of the d.c. Transistor Equations

The Ebers–Moll equations are rather cumbersome for continual use so it is worth deriving from them some frequently needed results.

The Active Region of Operation

With the emitter forward biased and the collector reverse biased V_E is positive and V_C is negative. If V_C is more negative than $-0{\cdot}1$ volt, $\exp(qV_C/kT) \ll 1$ (since $q/kT \approx 39$ volts^{-1}). Thus V_E can be eliminated between eqns. (1.3) and (1.4) yielding

$$-I_C = \alpha_N I_E + I_{CB0}. \tag{1.6}$$

Or, since for these equations $I_E + I_C + I_B = 0$

$$I_C = \frac{\alpha_N}{1-\alpha_N} I_B - \frac{I_{CB0}}{1-\alpha_N}. \tag{1.7}$$

The more usual or "natural" sign conventions for currents are closer to the physical behaviour of the device under normal operating conditions. They specify that

(1) the emitter current flows into the transistor and divides to flow out through collector and base, so that $I_E = I_C + I_B$;
(2) I_{CB0} is the same sign as I_C.

Equations (1.6) and (1.7) can be written in terms of the natural conventions by changing the signs of I_C and I_B. Thus for any

†It is often simplest to perform *calculations* assuming the transistor being considered is a *pnp* one (even if it is not) and changing the signs at the end of the calculation, rather than try to get the conventions right in the calculations on *npn* transistors. This is true even though in nearly all the circuits considered later *npn* transistors are used.

transistor,

$$I_C = \alpha_N I_E + I_{CB0} \tag{1.8}$$

$$I_C = \frac{\alpha_N}{1-\alpha_N} I_B + \frac{I_{CB0}}{1-\alpha_N}. \tag{1.9}$$

In addition it is common practice to use a single parameter β, often called the common emitter current gain, instead of $\alpha_N/(1-\alpha_N)$.† Making this substitution in eqn. (1.9) yields

$$I_C = \beta I_B + (\beta+1) I_{CB0}. \tag{1.10}$$

Equations (1.8) and (1.10) are the forms of Active Region d.c. equations which will be used in this book. (Note that when $V_C = V_{CB} = 0$ eqn. (1.10) reduces to $I_C = \beta I_B$. Thus when the collector is shorted to the base $I_{CB0} = 0$ but I_C can still flow if $I_E \neq 0$.)

The Cut-off Region of Operation

The terminal currents when one junction is reverse biased and the other is either reverse biased, unbiased or only slightly forward biased are all classed as cut-off currents. To distinguish them they are given letter symbols involving three capital letter subscripts.

The first subscript indicates the terminal to which a meter to measure the current should be connected.

The second subscript indicates the terminal to which the other side of the meter should be connected, via a reverse biasing battery.

The third subscript indicates what is to be done with the third terminal. If it is left open-circuit the third subscript is 0. If the third terminal is shorted to the terminal indicated by the second subscript, the third subscript is *S*. If a resistor replaces the short,

† The British Standard symbol for this is h_{FEL}, but β is more commonly used.

R replaces S, and if a reverse biasing battery replaces the resistor the third subscript is an X.

Since $I_{CES} \equiv I_{CBS}$ both are called I_{CS}. Similarly $I_{EBS} \equiv I_{ECS} \equiv I_{ES}$. In addition when X is the third subscript the second subscript can be dropped so that with both junctions reverse biased the cut-off currents are I_{CX}, I_{EX}, I_{BX}.

It is possible to calculate the values of some of these cut-off currents in terms of I_{CB0} from the Ebers and Moll equations, as follows:

$$\left.\begin{aligned} I_{CX} &= \frac{I_{CB0}(1-\alpha_I)}{1-\alpha_I\alpha_N} & I_{EX} &= \frac{I_{EB0}(1-\alpha_N)}{1-\alpha_I\alpha_N}, \\ I_{CS} &= I_{CB0}/(1-\alpha_I\alpha_N) & I_{ES} &= I_{EB0}/(1-\alpha_I\alpha_N), \\ I_{CE0} &= I_{CB0}/(1-\alpha_N) & I_{EC0} &= I_{EB0}/(1-\alpha_I). \end{aligned}\right\} \quad (1.11)$$

The emitter–base voltage at which the above collector currents flow can similarly be calculated; thus, using the sign conventions of the Ebers–Moll equations:

$$\left.\begin{aligned} &\text{When} \quad I_C = -I_{CB0} \qquad V_E = \frac{kT}{q}\ln(1-\alpha_N). \\ &\text{When} \quad I_C = -I_{CS} \qquad V_E = 0. \\ &\text{When} \quad I_C = -I_{CE0} \qquad V_E = \frac{kT}{q}\ln\left(1+\frac{1-\alpha_I}{1-\alpha_N}\cdot\frac{\alpha_N}{\alpha_I}\right) \end{aligned}\right\} \quad (1.12)$$

Thus when $I_C = -I_{CB0}$ and the emitter is left floating the emitter junction acquires a reverse bias; but when $I_C = -I_{CE0}$ and the *base* is left floating the emitter junction acquires a forward bias. These results are illustrated in Fig. 1.2.

The Saturation Region of Operation

The Saturation Region of Operation is normally achieved in a circuit such as that of Fig. 1.3(a). If $I_B > I_C/\beta$, then by equation (1.10) the transistor must be bottomed and it is found that under these circumstances both junctions in the transistor are

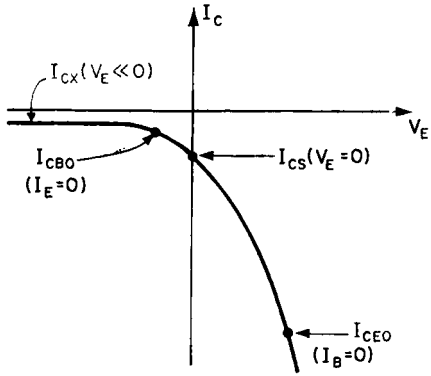


FIG. 1.2. The variation of collector current I_C with emitter junction voltage (forward bias positive), showing particularly the collector cut-off currents.

forward biased. The base charge distribution is as in Fig. 1.1(c) or as in Fig. 1.3(b); it is built up as follows.

As I_B is increased from zero the transistor remains in the Active Region of Operation until $I_C \approx E_{CC}/R_L$. I_C cannot then increase further but the base charge must continue to increase with I_B. The gradient of the charge at the collector cannot increase further so a distribution somewhat as shown is inevitable. Here the minority carrier density next to each junction is

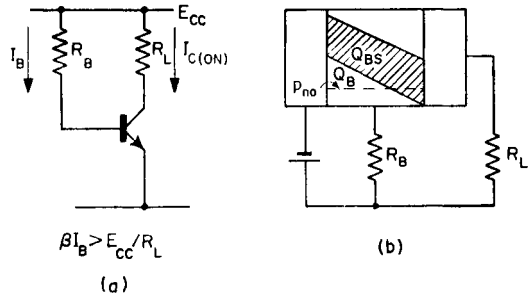


FIG. 1.3. A transistor driven into saturation (i.e. bottomed). (a) The circuit to achieve bottoming; (b) the base charge distribution.

greater than its equilibrium value so both junctions must be forward biased.

When the transistor is bottomed it is important to know the voltages between the terminals. If I_B, $I_C \gg I_{CB0}$, I_{EB0} then from eqns. (1.3), (1.4), (1.5), assuming negligible collector resistance,

$$V_{CE(\text{sat})} = \pm(V_E - V_C) = \frac{kT}{q} \ln \left\{ \frac{I_B + I_C(1-\alpha_I)}{\alpha_I(I_B - I_C(1-\alpha_N)/\alpha_N)} \right\}. \quad (1.13)$$

This is a much smaller voltage than that across either junction on its own. Typically with silicon transistors $|V_{BE}| \approx 0{\cdot}8$ V whilst $|V_{CE(\text{sat})}| \approx 0{\cdot}2$ V.

If the expression for $V_E - V_C$ in eqn. (1.13) is positive it means that the emitter junction has a larger forward bias than the collector. Thus the positive sign applies to *npn* transistors and the negative sign to *pnp* ones.

Breakdown Voltages and Zener Diodes

If the reverse bias voltage of a *pn* junction is increased sufficiently a voltage will be reached, usually called V_{BR}, at which the junction begins to break down and a large current begins to flow. Provided this current is prevented from becoming so large as to cause excessive heating ($V_{BR} \times$ current $<$ power rating) no damage results, but the onset of breakdown obviously limits the range of useful voltages which can be applied to the junctions.

The breakdown voltages of the collector and emitter junctions of transistors are normally stated by the manufacturers. The high doping near the emitters of diffused base transistors results in rather low emitter junction reverse breakdown voltages; as low as 0·5 V in amplifying transistors but from 2 to 6 V in transistors intended for switching circuits. The usual way to protect such transistors if larger voltages are unavoidable is to place a diode in series with the emitter (see p. 47).

In a Zener diode this breakdown phenomenon is used to provide a reference voltage.[2, 4] Beyond breakdown the voltage across

the diode varies only slightly as the current changes by orders of magnitude. At Zener voltages of about 6 V the slope resistance of the diode characteristic is minimal (normally a few ohms) and its temperature coefficient of voltage is near zero. Zener diodes are readily available in the range of about 4 V to 20 V. Zero temperature coefficients of breakdown voltage can be obtained at other than 6 V by cascading reverse biased Zener diodes of greater than 6 V breakdown, whose temperature coefficient is positive, with forward biased junctions, whose temperature coefficient is negative.

Transient Response of a Transistor

The Ebers–Moll equations refer to steady state conditions. If the bias conditions of a transistor change, the base charge of both minority and majority carriers must change too. It takes time to change these charges so in order to express the transient or high-frequency operation of a transistor it is necessary to set up time-dependent equations.[2]

Supposing a base current i_B is flowing into the base of a transistor it will in general have to perform two functions. It will first have to supply the recombination current for any charge already present and, second, if there is any current left over, it will cause the base majority carrier charge to change.

Thus

$$\boxed{i_B = \frac{q_B}{\beta\tau_C} + \frac{\mathrm{d}q_B}{\mathrm{d}t}\,;} \qquad (1.14)$$

where $1/\beta\tau_C$ is the expression used for the recombination rate of the base charge—for reasons which will shortly become clear. That is, $\beta\tau_C$ is the effective lifetime of minority carriers in the base region.

Ideally the collector current must be equal to the base region minority-carrier charge divided by the transit time, τ_t, whatever

the charge distribution (i.e. q_B/τ_t). In practice there is some recombination, so the constant relating i_C and base charge is not quite the transit time. It is called the *collector time factor* τ_C

and
$$\boxed{i_C = q_B/\tau_C.} \tag{1.15}$$

Equation (1.15) is the fundamental, charge control equation relating output current to base region *minority* carrier charge. Equation (1.14) states how much *majority* carrier charge accumulates in the base region as a result of a base current i_B. The fact that the same quantity q_B can be used in both equations is an expression of the fact that space charge neutrality exists and that the charge of minority carriers equals the excess charge of majority carriers.

Equations (1.14) and (1.15) imply that in the steady state if $I_B = 0$ then $I_C = 0$ too. But this is not the case, as already discussed. Consequently eqns. (1.14) and (1.15) should be corrected to:

$$i_B = \frac{q_B}{\beta\tau_C} + \frac{\mathrm{d}q_B}{\mathrm{d}t} - I_{CB0}, \tag{1.16}$$

$$i_C = q_B/\tau_C + I_{CB0}. \tag{1.17}$$

In the steady state
$$\mathrm{d}q_B/\mathrm{d}t = 0$$
so that
$$\beta = (I_C - I_{CB0})/(I_B + I_{CB0}).$$

This agrees with eqn. (1.10), which is why $\beta\tau_C$ was used for base region lifetime.

In practical circuits, however, cut-off currents normally significantly affect only the steady state conditions, so I_{CB0} will be neglected in any subsequent use of the charge control equations for transient calculations.

There is an additional effect which cannot, however, be neglected; it is the effect of the collector capacitance. When a transistor is switched on into a resistive load the collector–base voltage

decreases. This results in a narrowing of the collector transition region and a widening of the base region. This extra width of the base region requires an additional base charge, named q_{VC} (i.e. a voltage-dependent charge associated with the collector). The narrowing of the collector transition region also requires an equal charge to be drawn from the collector current and added to the collector region as shown in Fig. 1.4. Thus eqns. (1.14), (1.15) become

$$i_B = \frac{q_B}{\beta\tau_C} + \frac{\mathrm{d}q_B}{\mathrm{d}t} + \frac{\mathrm{d}q_{VC}}{\mathrm{d}t}, \tag{1.18}$$

$$i_C = \frac{q_B}{\tau_C} - \frac{\mathrm{d}q_{VC}}{\mathrm{d}t}. \tag{1.19}$$

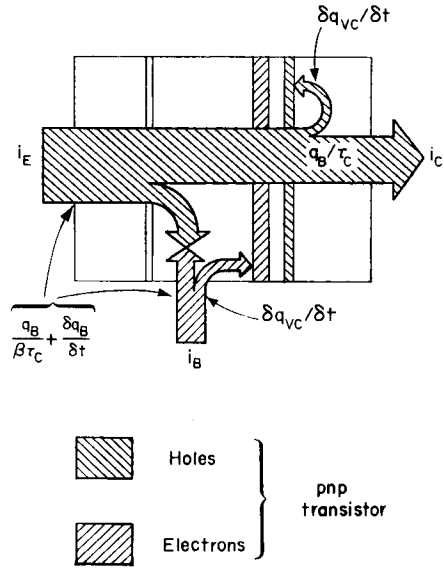


FIG. 1.4. Current flow through a *pnp* transistor at an instant during the turn-on transient. The base current, composed of electrons only, supplies electrons to build up the base charge ($\delta q_B/\delta t$), to charge up the collector capacitance ($\delta q_{VC}/\delta t$) and to allow recombination ($q_B/\beta\tau_0$). The steady state collector current (q_B/τ_0) composed only of holes is decreased by the hole current ($\delta q_{VC}/\delta t$) needed to charge up the collector side of the collector capacitance.

As shown in Fig. 1.5, q_{VC} is not linearly related to the collector–base voltage but for many purposes it is a good approximation to suppose that it is, and that

$$\frac{dq_{VC}}{dt} = -C_C \frac{dV_{CB}}{dt}, \tag{1.20}$$

where C_C is a "large signal" capacitance obtained by dividing the total change in q_{VC}, namely ΔQ_{VC}, by the total change in

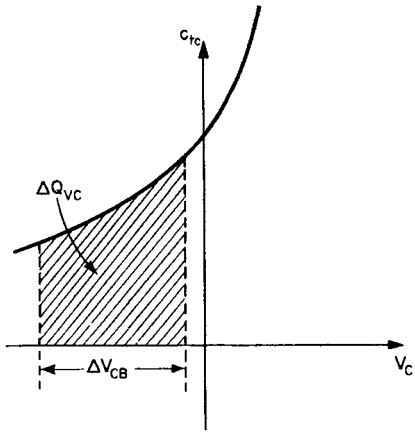


FIG. 1.5. The variation of the collector transition region capacitance c_{tc} with collector junction voltage (forward bias positive). The shaded area shows the total charge ΔQ_{VC} associated with a change of collector–base voltage ΔV_{CB}.

collector–base voltage $V_{CB1} - V_{CB2}$. Thus for a particular change ΔV_{CB} in collector junction voltage

$$C_C = |\Delta Q_{VC}/\Delta V_{CB}|.$$

Q_{VC} and hence C_C are best obtained graphically from the plot of the small signal collector capacitance c_{tc} versus V_C shown in Fig. 1.5.

The minus sign appears in eqn. (1.20) because q_{VC} increases as V_{CB} decreases.†

If it is assumed that C_C is a constant independent of voltage, then it follows that q_{VC} is proportional to q_B and eqns. (1.18), (1.19) can be linearized without too much error as

$$i_B = \frac{q_B}{\beta\tau_C} + \frac{\mathrm{d}q_B}{\mathrm{d}t}\left(1 + \frac{\Delta Q_{VC}}{\Delta Q_B}\right), \tag{1.21}$$

$$i_C = \frac{q_B}{\tau_C} - \frac{\Delta Q_{VC}}{\Delta Q_B}\frac{\mathrm{d}q_B}{\mathrm{d}t}, \tag{1.22}$$

where ΔQ_B is the total change in q_B during switching. This is the form in which these equations will be used subsequently.

Saturation Region

The above equations apply to the Active Region of Operation. When the transistor is driven into the Saturation Region of Operation both junctions acquire a small forward bias voltage and some extra charge Q_{BS} accumulates in the base region as shown in Fig. 1.3(b).

Here

$$Q_B = I_{C(\mathrm{ON})}\tau_C, \tag{1.23}$$

$$Q_{BS} = I_{BS}\tau_S, \tag{1.24}$$

where

$$I_{BS} = I_B - I_{C(\mathrm{ON})}/\beta. \tag{1.25}$$

Q_{BS} is called the saturation base charge, τ_S is the *saturation time factor*, and $1/\tau_S$ is the effective recombination rate of this saturation charge. (Normally $\tau_S < \beta\tau_C$.) I_{BS} is the *saturation base current*, the base current in excess of that needed to *just* bottom the transistor.

† The important point in regard to signs is that whatever the convention used for voltage, the signs of $\mathrm{d}q_B/\mathrm{d}t$ and $\mathrm{d}q_{VC}/\mathrm{d}t$ are normally the same. That is, if q_B increases, so does q_{VC}.

The variation of saturation base charge is evidently governed by the equation

$$i_B = \frac{Q_B}{\beta\tau_C} + \frac{q_{BS}}{\tau_S} + \frac{\mathrm{d}q_{BS}}{\mathrm{d}t} \tag{1.26}$$

in which, during saturation, Q_B is, by definition, a constant. This equation applies only when the transistor is saturated. Since both junctions remain forward biased in saturation q_{VC} can normally be neglected.

The On-demand Current Gain

When a transistor is being held bottomed by a constant base current, as in Fig. 1.3(a), and when the *collector* current may be subject to variation, then an additional parameter can be of importance.

In switching circuits it is common for the collector current of a bottomed transistor to change as gates connected to it are modified by other circuits. An extreme case is simulated in the circuit of Fig. 1.6(a) where the collector current can be reduced to zero by opening the switch. Suppose $I_C = 0$ with I_B flowing. The transistor is bottomed, $V_{CE} = V_{CE(\mathrm{sat})} \approx 0$, and the base charge distribution is as shown in Fig. 1.5(b). If now the switch is closed it is often found that I_B must exceed $I_{C(\mathrm{ON})}/\beta$ in order to ensure that V_{CE} remains at or less than the maximum permitted value of $V_{CE(\mathrm{sat})}$. Instead I_B must be at least I_C/β_S where β_S is called the "On-demand" current gain.

β_S is defined as the peak collector current $\hat{I}_{C(\mathrm{ON})}$ "immediately" available per unit base current, and in theory it differs from β for the following reason. When the switch is closed a reverse bias tends to be applied to the collector immediately. This lowers the base charge density at the collector to near zero. But there is as yet insufficient time for the base charge to change,

so the immediate result is that Q_{BS} of Fig. 1.6(b) redistributes itself to become Q_B of Fig. 1.6(c).

Hence, since $Q_{BS} = Q_B$

$$\underbrace{I_B = I_{BS} = Q_{BS}/\tau_S}_{\text{Fig. 1.6(b)}} = \underbrace{Q_B/\tau_S = \hat{I}_{C(ON)}\tau_C/\tau_S}_{\text{Fig. 1.6(c)}}. \quad (1.27)$$

Thus

$$\frac{\hat{I}_{C(ON)}}{I_B} \equiv \beta_S \approx \frac{\tau_S}{\tau_C}. \quad (1.28)$$

In practice with homogeneous base transistors β_S may be between β and about $\beta/3$; and the relation $\beta_S = \tau_S/\tau_C$ is only approximately obeyed. With diffused collector junction transistors it is

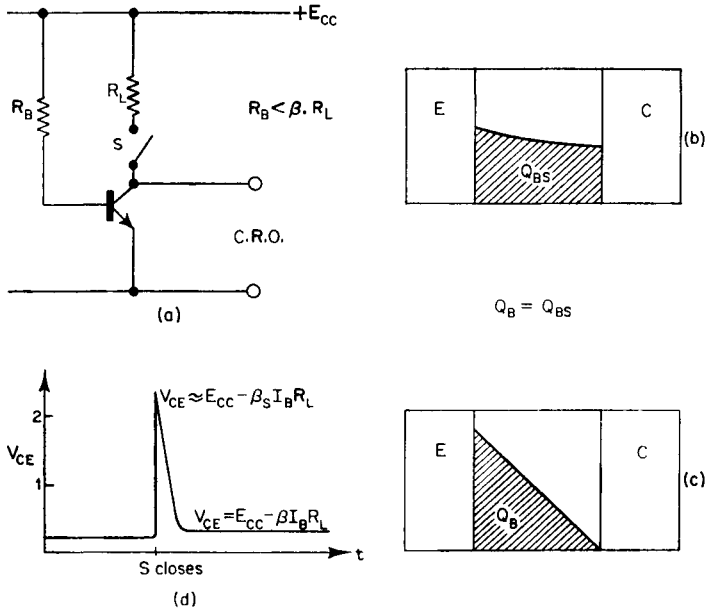


FIG. 1.6. On-demand current gain. (a) A circuit for displaying the effect; (b) the base charge distribution before S closes; (c) the base charge distribution shortly after S closes; (d) the pulse appearing on the CRO if $I_{C(ON)}/\beta = I_B < I_{C(ON)}/\beta_S$.

usual for β_S to be greater than β and is therefore not an important parameter.

Figure 1.6(d) shows the behaviour of the collector voltage when the switch is closed if $I_B < I_{C(ON)}/\beta_S = E_{CC}/R_L \beta_S$.

The Cut-off Region

In the Cut-off Region of Operation both junctions are reverse biased. Thus when a turn-on base current is applied in order to turn the transistor on again the junction capacitances must be charged up sufficiently to ensure a forward bias on the emitter junction before any minority carrier charge flows into the base region. During the charging process the reverse bias on both junctions is falling so that charge is supplied to both capacitances. The total charge is called Q_{VD} (subscript D for "delay"). It causes a delay between the application of a base current and the appearance of any collector current.

The Transient Switching Parameters of a Transistor

The five parameters introduced so far, namely, τ_C, Q_{VC}, Q_{VD}, τ_S and β_S, are not always quoted by transistor manufacturers; but if not, their near equivalents usually are. The important equivalent parameters of all except β_S, which has no reliable equivalent, are as follows.

First, as shown shortly, the above parameters determine quite simply the response times of a transistor when driven by a constant current source. Thus the response times can, conversely, be used to obtain the switching parameters. This is the most common method of describing transient response:

Second, the switching parameters are related to small signal parameters.

(a) $\tau_C \approx 1/\omega_T$, (1.29)

where $f_T = \omega_T/2\pi$ is the frequency at which the small signal-common emitter-current gain has fallen to 1. f_T should be measured at a low value of V_{CE}.

(b) $$Q_{VC} = Mc_{tc1}(|V_{CB1}| - |V_{CB2}|), \tag{1.30}$$

where $|V_{CB1}|$ and $|V_{CB2}|$ are the initial and final values of V_{CB} when the transistor is being turned on, c_{tc1} is the small signal transition region capacitance when $V_{CB} = V_{CB1}$ and M is a constant between 1 and 2. When $V_{CB1} - V_{CB2} \ll V_{CB1}$, $M \approx 1$, but when $V_{CB2} \approx 0$, M tends to 2. A value of about 1·7 for abrupt junctions or 1·3 for graded junctions is typical.

With high-speed transistors the capacitances of the encapsulation may be significant. These add to both Q_{VD} and Q_{VC}. The best way to obtain Q_{VC} is to take the area under a plot of c_{tc} versus V_{CB} between the limits applicable to the problem in hand as in Fig. 1.5.

(c) τ_S is related to f_T and to the corresponding frequency with the transistor operated in the inverse connection, namely $f_{TI} = \omega_{TI}/2\pi$.

$$\tau_S(1 - \alpha_I\alpha_N) \approx \frac{1}{\omega_T} + \frac{1}{\omega_{TI}}.$$

Information on ω_{TI} is, however, rarely available.

(d) Q_{VD} is determined largely by the emitter transition region capacitance, with a further contribution from c_{tc}. Again, the area under a plot of the small signal capacitances versus voltage is the best way to find Q_{VD}. Note, however, that V_{EB} usually passes through zero as the emitter is taken from reverse to forward bias.[3]

Response Times

If a square wave is applied to the input of the circuit shown in Fig. 1.7(a) (excluding C_B), the output current waveform is usually similar to that shown in Fig. 1.7(c). The turn-on time is divided

into two parts, the *delay time* t_d before the collector current starts to rise above its cut-off value (i.e. above zero approximately) and the *rise time* t_r which is the time taken for i_C to rise from zero

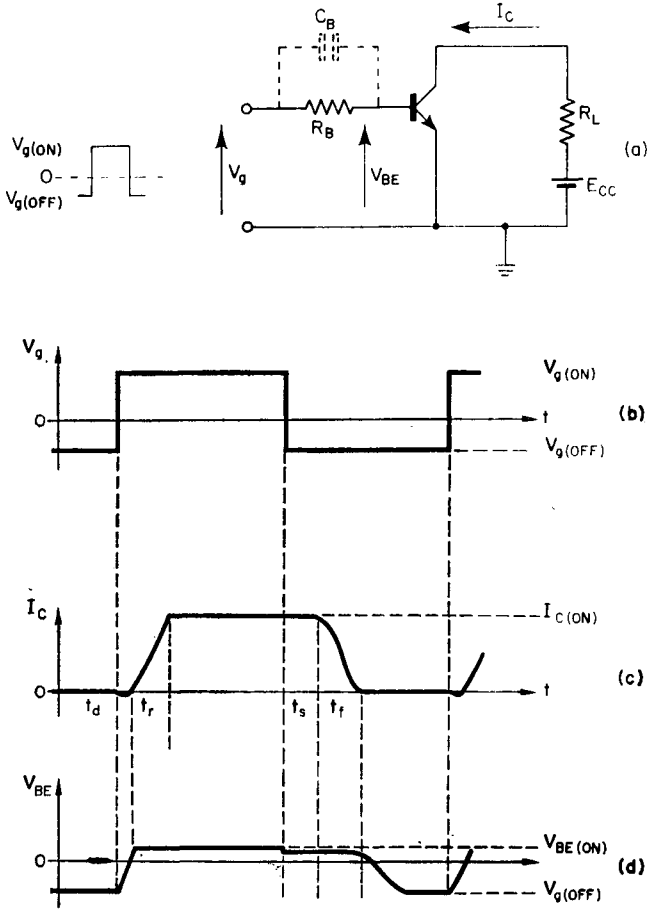


FIG. 1.7. The transient response of a transistor. (a) The basic circuit; (b) the input waveform; (c) the collector current waveform; (d) the base-emitter voltage waveform. (Typical values: $E_{CC}=10$ V $V_{g(ON)}=6$ V, $V_{g(OFF)}=-2$ V, $R_B=10$ kΩ, $R_L=1$ kΩ.)

to its final value (about E_{CC}/R_L in the circuit of Fig. 1.7(a), if the transistor is driven into saturation).

Similarly the turn-off time is divided into two parts. First the *saturation time*, t_s, during which the saturation charge q_{BS} is used up and the collector current does not significantly change. Second, the *fall time* t_f during which i_C falls from its maximum value to near zero as q_B is removed.

In the circuit of Fig. 1.7(a), the base–emitter voltage, V_{BE}, stays almost constant at $V_{BE(ON)}$ throughout t_r, t_f and t_s. Thus the base current $I_{B(ON)}$ during t_r is nearly constant, namely

$$I_{B(ON)} \approx \frac{V_{g(ON)} - V_{BE(ON)}}{R_B}. \tag{1.31}$$

During t_s and t_f, the constant base current is

$$I_{B(OFF)} \approx \frac{V_{g(OFF)} - V_{BE(ON)}}{R_B}, \tag{1.32}$$

where R_B includes the generator resistance and any extrinsic base resistance within the transistor and $V_{g(ON)}$, $V_{g(OFF)}$ are the open circuit voltages produced by the generator.

During the delay time t_d, V_{BE} increases from about $V_{g(OFF)}$ to $V_{BE(ON)}$, as in Fig. 1.7(d). The base current therefore varies with time. If the variation is assumed to be linear the average base current $\bar{\imath}_B$ is given by:

$$\bar{\imath}_B = \left(V_{g(ON)} - \frac{V_{g(OFF)} + V_{BE(ON)}}{2} \right) \frac{1}{R_B}. \tag{1.33}$$

Using these values of base current in eqs. (1.21) and (1.22) gives the following response time expressions:

$$t_d = Q_{VD}/\bar{\imath}_B, \tag{1.34}$$

$$t_r = \beta\tau_C \left(1 + \frac{Q_{VC}}{Q_B}\right) \ln \left(\frac{I_{B(ON)}}{I_{B(ON)} - I_{C(ON)}/\beta} \right), \tag{1.35}$$

$$t_f = \beta\tau_C \left(1 + \frac{Q_{VC}}{Q_B}\right) \ln \left(\frac{I_{B(OFF)} - I_{C(ON)}/\beta}{I_{B(OFF)}} \right), \tag{1.36}$$

provided

$$\frac{Q_{VC}}{Q_B}\frac{dq_B}{dt} \ll \frac{q_B}{\tau_C}$$

in eqn. (1.22) and

$$t_s = \tau_S \ln\left\{\frac{I_{B(OFF)} - I_{B(ON)}}{I_{B(OFF)} - I_{C(ON)}/\beta}\right\}. \tag{1.37}$$

These results are obtained by substituting the following boundary conditions in the appropriate charge control eqns. (1.21), (1.22), (1.26). Here $\Delta Q_{VC} = Q_{VC}$ and $\Delta Q_B = Q_B$. Also

Rise time When $t = 0$ $q_B = i_C = 0$,
when $t = t_r$ $q_B = \tau_C I_{C(ON)}$.

Saturation time When $t = 0$ $q_{BS} = \left(I_{B(ON)} - \frac{I_{C(ON)}}{\beta}\right)\tau_S$,
when $t = t_s$ $q_{BS} = 0$.

Fall time When $t = 0$ $q_B = \tau_C I_{C(ON)}$,
when $t = t_f$ $q_B = 0$

These rise and fall times are sometimes quoted for changes between 10% and 90% of the final value of collector current. The corresponding theoretical expressions are obtained using the appropriate boundary conditions.

Evidently eqns. (1.34) to (1.37) can be used either to calculate the response times from the transistor parameters, or the reverse. However, in using them to determine the switching parameters, two points should be noted.

First, eqns. (1.35) and (1.36) do not separate Q_B and Q_{VC}. To obtain values for each from measurements of t_r involves measurements of t_r at two widely different values of $I_{C(ON)}$, since Q_B is (almost) proportional to $I_{C(ON)}$ but Q_{VC} is independent of it.

Second, τ_C and τ_S are not completely independent of operating conditions so their values at one condition of $I_{C(ON)}$, $I_{B(ON)}$, etc., can only be used to *estimate* performance at other conditions.

It is often convenient to discuss transistor transient performance in terms of the *total* switching charge required.

If in Fig. 1.7(a) the input resistance is by-passed with the capacitor C_B shown dotted, the response times are greatly reduced. This is because at turn-on the capacitor discharges into the base region, rapidly injecting a charge Q_{IN} given by

$$Q_{IN} = (V_{g(ON)} - V_{BE(ON)})C_B, \tag{1.38}$$

whilst at turn-off the capacitance removes a charge of

$$Q_{OUT} = (V_{g(ON)} - V_{g(OFF)})C_B. \tag{1.39}$$

Notice that since the change of charge on a capacitance is proportional to the change of voltage across it, the voltage terms in eqns. (1.38), (1.39) are not the same. At turn-on the base voltage rises from $V_{g(OFF)}$ to $V_{BE(ON)}$ so that the voltage change across the capacitor is less than the driving step by this amount. At turn-off it is assumed that the transistor base voltage does not change (i.e. that the emitter remains forward biased until the collector current has dropped almost to zero).

The charge needed to take the transistor from the cut-off condition to the ON state (which may be the edge of the saturation region of operation) is

$$Q_{ON} = Q_{VD} + Q_{VC} + Q_B \tag{1.40}$$

where $$Q_B = I_{C(ON)}\tau_C.$$

After the turn-on transient the transistor will normally be held on by the base resistor R_B, and if the transistor is bottomed, the current through R_B will build up saturation charge, $Q_{BS} = I_{BS}\tau_S$, in the base region.

This charge will have to be removed at turn-off so that

$$Q_{OFF} = Q_{VC} + Q_B + Q_{BS}. \tag{1.41}$$

Notice that it is not necessary to inject Q_{BS} at turn-on, nor to remove Q_{VD} at turn-off, if we are only interested in the time taken for *collector current* to rise and fall between zero and $I_{C(ON)}$.

The optimum value of C_B is such that $Q_{IN} = Q_{ON}$ and $Q_{OUT} = Q_{OFF}$. In most circuits the capacitor is chosen so that for all possible conditions and variations of transistors

$$Q_{IN} \geqslant Q_{ON}$$

and

$$Q_{OUT} \geqslant Q_{OFF}.$$

This concludes the discussion of transistor transient responses; more details can be found in reference 2. Some problems appear at the end of the chapter.

Temperature Effects and Power Dissipation

Ratings

Ratings are statements by the manufacturer of the limiting conditions within which reliable operation of the device can be expected. They imply that if the transistor is operated at a voltage or power level which exceeds the voltage or power rating, then it is to be expected that the device will be damaged, or its performance decline, before the device has had a reasonable period of useful life.

Thus ratings are fixed, or decided upon, after extensive life tests have been undertaken and after it has been discovered under which conditions the device will not deteriorate significantly in a reasonable time.

In general it is true that the further the operating condition is from the rated value (on the conservative side) the longer the expected life of the device.

Ratings or Maximum Operating Conditions are usually stated for power dissipation, junction voltages and sometimes currents too. In addition a *thermal derating factor* is usually stated. The power dissipation is limited because too much dissipation produces too high a junction temperature. Evidently therefore, if the device is operated in a hot ambient or environment the amount of power which it is safe to dissipate is reduced. The thermal derating factor states by how much the power dissipated must be reduced per degree rise of ambient or case temperature.

The thermal derating factor is not to be confused with the *thermal resistance* θ. Thermal resistance is a characteristic of the

device. If P milliwatts of power are dissipated in the device, the resulting junction temperature rise T_r can be measured, and

$$\theta = T_r/P.$$

It may well be that the derating factor is the reciprocal of θ, but if this is so it means that it has been found from life tests that overall temperature rise is the sole cause of transistor failure when power is being dissipated. This may not be the case. For example, traces of volatile gases may be trapped in the transistor casing. If heat is supplied from the outside the gases will condense on the semiconductor elements within the casing. But if heat is generated in the device the gases will condense on the casing. Thus, where the heat comes from can make a difference.

If cooling fins are used, both thermal resistance and thermal derating factor are affected.

Thermal derating factors can be applied to other ratings (e.g. voltage) but this is not usual with transistors.

Temperature Effects

The most temperature sensitive parameters of a transistor are the cut-off currents I_{CB0} and I_{EB0}; both increase at a rate of about 10% per °C in germanium transistors, and at a rate of about 15% per °C in silicon transistors, when the devices are operated between about 0°C and 75°C. At higher temperatures the rate of increase falls off. (Germanium devices are not normally used at operating temperatures much above 75°C.)

Since both α_I and α_N also normally increase somewhat with temperature, it follows from eqns. (1.11) that the rate of increase of other cut-off currents may be even greater.

This variation of cut-off current also gives rise to a temperature variation of the forward bias voltage of, say, the emitter–base junction at a given emitter current. Eqn. (1.4) states that

$$I_E = I_{EB0}(e^{qV_E/kT} - 1) - \alpha_I I_C,$$

and if I_E is held constant, at some significant forward current, by an external circuit (so that exp. $qV_E/kT \gg 1$) and if we assume $\alpha_I I_C$ remains substantially constant too, then we can differentiate eqn. (1.4), and replace V_E by V_{EB}, to obtain

$$0 = \frac{dI_{EB0}}{dT} + I_{EB0}\left(\frac{dV_{EB}}{dT}\left(\frac{q}{kT}\right) - \frac{V_{EB}}{T}\cdot\frac{q}{kT}\right). \qquad (1.42)$$

But since $dI_{EB0}/dT \approx 0{\cdot}1\ I_{EB0}/°C$ in germanium transistors and since $V_{EB} \approx 0{\cdot}2$ V it follows that at $T=300°K$ (when $q/kT \approx 39\ V^{-1}$)

$$0{\cdot}074 = -\frac{dV_{EB}}{dT}\cdot\frac{q}{kT}$$

or

$$\frac{dV_{EB}}{dT} \approx -1{\cdot}9\ \text{mV/°C for Ge.}$$

In silicon transistors $V_{EB} \approx 0{\cdot}7$ V; since $dI_{EB0}/dT = 0{\cdot}15\ I_{EB0}/°C$,

$$\frac{dV_{EB}}{dT} \approx -1{\cdot}5\ \text{mV/°C.}$$

This smaller variation in silicon of both current and voltage is important in the design of level sensitive circuits which are not temperature dependent.

Thermal Resistance and Thermal Runaway

In circuit design the most serious possible consequence of the variation of I_{CB0} is "thermal runaway". This arises as follows When a transistor is dissipating power (approximately equal to $V_{CE}I_C$) the temperature of the transistor rises (by θ°C/mW) and causes I_C to rise—since part of I_C is a temperature sensitive cut-off current. But this causes a further increase in power dissipation and temperature, leading yet again to an increase in I_{CB0}. Whether this process is convergent or whether it "runs away"

depends on whether the rate of increase with temperature of the power being generated is less than, or exceeds, the reciprocal of the "thermal resistance", θ. Thus to avoid thermal runaway

$$\frac{\mathrm{d}(V_{CE}I_C)}{\mathrm{d}T} < \frac{1}{\theta}. \tag{1.43}$$

The worst condition is when the base is open circuited, or when the base circuit contains a high resistance, since then the base current is held constant. Now

$$I_C = \beta I_B + I_{CE0}.$$

Multiplying by V_{CE} and differentiating yields (if β is assumed to be constant),

$$\frac{\mathrm{d}(V_{CE}I_C)}{\mathrm{d}T} \approx V_{CE}\frac{\mathrm{d}I_{CE0}}{\mathrm{d}T} \approx V_{CE}\frac{\mathrm{d}I_{CB0}}{\mathrm{d}T}(\beta+1).$$

Thus there is a maximum collector–emitter voltage above which thermal runaway will occur, namely

$$V_{CE\,(\max)} = \frac{1}{\theta(\beta+1)\mathrm{d}I_{CB0}/\mathrm{d}T}. \tag{1.44}$$

The value of $\mathrm{d}I_{CB0}/\mathrm{d}T$ which appears in this expression is, of course, its value at the actual operating junction temperature of the device, and since $\mathrm{d}I_{CB0}/\mathrm{d}t$ increases with temperature this maximum permitted voltage falls as the operating temperature rises.

In practice both β and I_{CB0} increase with temperature so the maximum safe value of V_{CE} is less than that given in eqn. (1.44).

If the emitter current rather than the base current is held constant the term $(\beta+1)$ does not appear in the expression.

Thus, particularly in common emitter circuits, the collector voltage that it is safe to use may be limited by thermal runaway as well as by the normal voltage ratings.

Finally, to complete this introductory chapter a table is included which presents the most important data for two quite

different types of switching transistor. The *npn* silicon planar one is a diffused epitaxial transistor based on the 2N914, though the registered specification for this type number contains a good deal more data than that shown in Table 1.1. The *pnp* transistor is an alloy junction type of which there are many versions made in the various countries of the world.

TABLE 1.1

Ratings and characteristics	*npn* planar silicon	*pnp* alloy Ge	*Units*
Absolute maximum ratings			
Operating junction temperature	200	90	°C
Power dissipation (ambient temp. = 25°C)	360	200	mW
Thermal derating factor[3] (in air)	2·1	3·1	mW/°C
Power dissipation (case temp. = 25°C)	1·2	—	W
V_{CB} ($I_E = 0$)	40	20	V
V_{CE} ($I_B = 0$)	15	—	V
($R_{BE} < 1$ kΩ)	—	20	V
V_{EB} ($I_C = 0$)	5	10	V
Electrical characteristics (at 25°C unless otherwise stated)			
d.c. Current gain, β (or h_{FEL}) ($I_C = 10$ mA, V_{CE} just not bottomed) typical	55	70	
d.c. Current gain, β (or h_{FEL}) ($I_C = 10$ mA, V_{CE} just not bottomed) limits	30 to 120	> 30	

Notes: (1) Only the magnitudes of d.c. voltages and currents are given. In general they are negative for the *pnp* transistor.

(2) The data have been adjusted slightly in places to apply to comparable operating conditions for the two transistors.

(3) Thermal resistance is the reciprocal of the thermal derating factor for each of these transistors.

Table 1.1 (cont.)

Ratings and characteristics	*npn* planar silicon	*pnp* alloy Ge	*Units*
$V_{CE\,(sat)}$ (I_C = 10 mA, I_B = 1 mA)	0·2 (<0·25)	0·1 (<0·15)	V
(I_C = 200 mA, I_B = 20 mA)	0·4 (<0·70)	0·2 (<0·3)	V
$V_{BE\,(sat)}$ (I_C = 10 mA, I_B = 1 mA)	0·7 to 0·8	<0·25	V
I_{CBO} (V_{CB} = 6 V)	—	1·0 (<2·0)	μA
(V_{CB} = 20 V)	4·0 (<20)	—	mμA
(V_{CB} = 6 V, T_{amb} = 80°C)	—	40 (<80)	μA
(V_{CB} = 20 V, T_{amb} = 150°C)	3(<15)	—	μA
f_T ($V_{CE} \approx$ 1 V, I_C = 10 mA)	~200	~10	MHz
τ_C (V_{CE} = 1 V, I_C = 10 mA)	0·8	15 (<30)	nsec
τ_S (I_C = 10 mA, I_B = 1·0 mA)	0·03	0·8 (<1·2)	μsec
β_S (I_C = 10 mA, $V_{CB} \approx$ 0 V)	—	30 (>20)	
c_{ob} (or c_{tc}) (V_{CB} = 6 V, I_E = 0)	5(<7)	8·5 (<12)	pf
c_{ib} (or c_{te}) (V_{EB} = 0·5 V, I_C = 0 V)	9	—	pf
r_{bx} (V_{CE} = 6 V, I_C = 0 V)	—	90 (<200)	Ω
$t_{ON} = t_{OFF}$ (I_C = 200 mA, $I_{B\,(ON)} = 2I_{B\,(OFF)}$ = 40 mA)	25(<40)	—	nsec

Problems

1.1. Verify eqns. (1.11), (1.12), (1.13), starting with the Ebers–Moll equations.

1.2. Plot curves of $V_{CE(sat)}$ versus I_B for constant I_C, and of $V_{CE(sat)}$ versus I_C for constant I_B when $\alpha_I = 0{\cdot}8$, $\alpha_N = 0{\cdot}98$, $I_{EB0} = 0{\cdot}81$ μA, $I_{CB0} = 1{\cdot}0$ μA.

1.3. If in Fig. 1.7(a) $R_B = 50$ kΩ, $R_L = 2$ kΩ, $E_{CC} = 10$ V and if $V_{g(ON)} = 8$ V and $V_{g(OFF)} = -2$ V calculate the delay time, the rise time, the fall time and the storage time using a transistor for which $\beta = 80$, $\tau_C = 0{\cdot}016$ μsec, $Q_{VD} = 50$ pC, $Q_{VC} = 80$ pC, $\tau_S = 1{\cdot}0$ μsec. (Assume $V_{BE(ON)} = V_{CE(sat)} = 0$ and that $I_{CB0} = 0$.)

(Ans.: 0·278 μsec, 1·27 μsec, 2·4 μsec, 0·668 μsec.)

1.4. Repeat the calculation of problem 1.3 for a silicon transistor for which $V_{BE(ON)}$ is 0·75 V ($V_{CE(sat)}$ remains almost zero).

(Ans.: 0·29, 1·45, 1·95, 0·53 μsec.)

1.5. (a) In the circuit of Fig. 1.7(a) the generator produces a square wave of 8 V amplitude returning to zero. R_B and C_B are adjusted for a square-wave output of 10 V amplitude. In the ON state the transistor is *just* saturated (i.e. $Q_{BS} = 0$).

With $R_L = 10$ kΩ, $C_B = 18{\cdot}7$ pF $\quad R_B = 320$ kΩ,

$R_L = 1$ kΩ, $C_B = 52{\cdot}5$ pF $\quad R_B = 32$ kΩ.

Calculate β, τ_C, Q_{VC} (assuming again that $V_{BE(ON)} = V_{CE(sat)} = 0$ and neglecting I_{CB0}).

(b) With R_L still equal to 1 kΩ, R_B is decreased to 16 kΩ (thus saturating the transistor). In order to obtain abrupt turn-off again C_B must be increased to 83·75 pF. What is the value of τ_S of the transistor? With R_L increased to 10 kΩ again (but R_B still 16 kΩ), what value should C_B have to obtain an abrupt turn-off again?

(Ans.: (a) 40, 0·030 μsec, 120 pC. (b) 1 μsec, 78·2 pF.)

1.6. In the circuit of Fig. 1.7(a) $E_{CC} = 10$ V, $R_B = 12$ kΩ, $R_L = 1$ kΩ, $V_{g(ON)} = +6$ V, $V_{g(OFF)} = -3$ V. At turn-off the storage time, t_s, is 0·6 μsec and the fall time, t_f, for the collector current to fall from its maximum value to zero is also 0·6 μsec. Calculate the rise time t_r, and the saturation time constant τ_S if the current gain, β, of the transistor is 50.

What is the minimum size of capacitor, C_B, placed in parallel with R_B which will ensure abrupt turn-off of the collector current? (Assume $V_{BE(ON)} = 0{\cdot}7$ V, $V_{CE(sat)} = 0$ V, $I_{CB0} = 0$.) (Ans.: 0·725 μsec, 1·53 μsec, 67·8 μμf.)

1.7. Calculate the charge Q_{VC} which must be supplied to the collector capacitance of a transistor when the applied d.c. voltage V_{CB} across it changes from $V_{CB1} = -10$ V to zero volts (the minus sign means reverse bias), given that when $V_C = -6$ V the small signal capacitance $c_{tc} = 10\,\mu\mu$F, and that $c_{tc} = \text{const}(-V_C - \phi)^{-\frac{1}{2}}$ where ϕ is the contact potential $= -0{\cdot}9$ V.

Derive a general expression for Q_{VC} before substituting numerical values. Plot a curve of Q_{VC} versus V_C.

(Ans.: 124 pico coulombs.)

1.8. Derive expressions for the rise time and fall time of output current of a transistor when the initial and final values of I_C are 10% and 90% of the fully bottomed value.

CHAPTER 2

Some Basic Circuits

WHEN a transistor is being used as a switch the type of behaviour required of it is much simpler than when it is used as a linear amplifier. It is expected to change rapidly between two states, usually between a cut-off condition, and a conducting state with the transistor bottomed. Consequently the problem in the design of switching circuits often reduces to ensuring that (a) a transistor which should be cut-off is held cut-off despite component tolerances and changes of temperature or of transistors; (b) a transistor which should be in a conducting state is similarly held there; (c) when a transistor changes state it does so sufficiently fast, despite tolerances of loading and drive, etc.

Once the details of individual circuits have been properly controlled, the subsequent problems of how to interconnect them to perform systematic functions can be tackled.

In this chapter the behaviour of some simple circuits will be considered; first the inverter, second the emitter follower, third the long-tailed pair.

npn devices will always be used where possible, primarily because this simplifies the signs in the circuit equations and avoids ambiguity in the explanation. (The sign convention used from now on is the "natural" one in which positive currents flow away from positive potentials.) *pnp* devices can be used instead but all polarities must be reversed.

The Inverter

A simple inverter is shown in Fig. 2.1(a). The transistor changes from the cut-off state (at which input voltage $V_{g(\mathrm{OFF})}$ is zero volts or less) to the bottomed state (at which the input voltage $V_{g(\mathrm{ON})}$ is about E_{CC}). The circuit is an inverter because V_{OUT}, the output voltage, is approximately zero when $V_g = E_{CC}$ and is approximately E_{CC} when $V_g = 0$.

Actually, of course, the output voltage changes by less than E_{CC}. $V_{\mathrm{OUT}} = E_{CC} - I_{C(\mathrm{OFF})} R_L$ when the transistor is cut-off, and $V_{\mathrm{OUT}} = V_{CE(\mathrm{sat})}$ when the transistor is bottomed. For best performance $I_{C(\mathrm{OFF})}$ and $V_{CE(\mathrm{sat})}$ should be as small as possible.

The Bottomed State

Transistor data provided by the manufacturers usually state the maximum value of $V_{CE(\mathrm{sat})}$ for given values of I_C and I_B. The minimum value of β is the ratio of these specified currents. Thus in order to achieve an output voltage no greater than $V_{CE(\mathrm{sat})}$ it is necessary that:

$$I_{B(\mathrm{ON})} \geqslant I_{C(\mathrm{ON})}/\beta_{\min}. \tag{2.1}$$

If the generator has the equivalent circuit shown in Fig. 2.1(a), then

$$I_{B(\mathrm{ON})} = (V_{g(\mathrm{ON})} - V_{BE(\mathrm{ON})})/(R_g + R_B). \tag{2.2}$$

The Cut-off State

If $I_{C(\mathrm{OFF})}$ is to be as small as possible the emitter–base junction should be reverse biased so that $I_{C(\mathrm{OFF})} = I_{CX}$. If V_{BE} is allowed to rise to zero volts or above, $I_{C(OFF)}$ will rise to I_{CS} or more, with the result that V_{OUT} may fall significantly

below E_{CC}. In particular, if, as is often the case, the input voltages $V_{g(OFF)}$ and $V_{g(ON)}$ are provided by the output of a similar transistor circuit, the minimum value of V_g will be $V_{CE(sat)}$ (instead of zero), so that the base–emitter voltage of the inverter will also be positive, and $I_{C(OFF)}$ will be greater than I_{CS}, and perhaps even greater than I_{CE0}. At high operating temperatures this is almost certainly too large when germanium transistors are

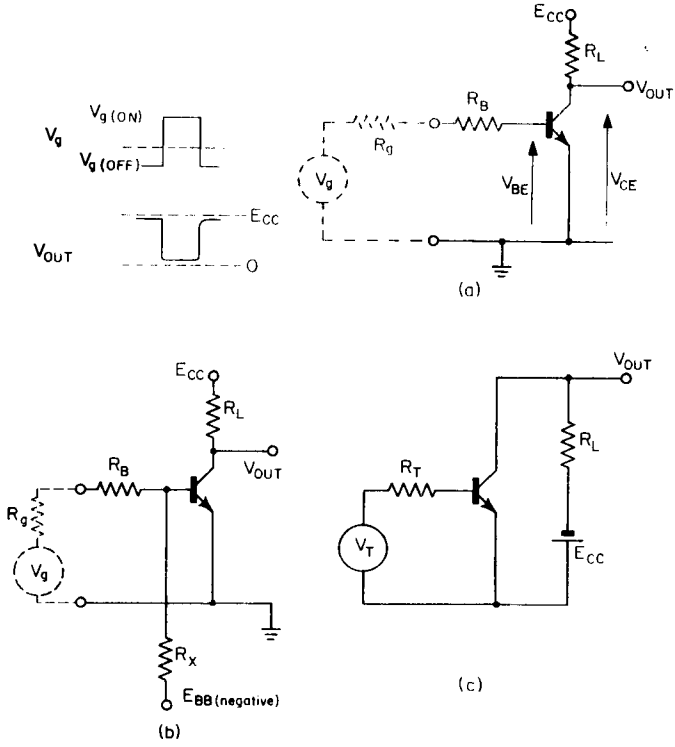


FIG. 2.1. The inverter. (a) The basic circuit coupled to a generator; (b) an inverter with base resistor to ensure a firm cut-off condition; (c) the Thévenin equivalent of (b) (Typical values: $R_L = 1$ kΩ, $R_B = 15$ kΩ, $V_{g(ON)} = E_{CC} = 6$ V, $V_{g(OFF)} = 0$, $R_X = 100$ kΩ.)

used. Consequently, the circuit of Fig. 2.1(a) is usually modified to that of Fig. 2.1(b) when germanium devices are used.

In Fig. 2.1(b) the condition for the cut-off state is determined by equating currents at the base node. Thus:

$$\frac{V_{g(\mathrm{OFF})}-V_{BE}}{R_B+R_g}+I_{BX}=\frac{V_{BE}-E_{BB}}{R_X}. \tag{2.3}$$

In a well-designed circuit using germanium transistors, R_B, E_{BB} and R_X will be chosen so that V_{BE} will be zero volts, or less, for the largest possible value of $V_{g(\mathrm{OFF})}$ and I_{BX}.

Using silicon transistors the collector cut-off currents are much smaller and they usually do not rise so rapidly with increasing V_{BE} (because α_N and α_I are usually significantly smaller at low currents than at high ones). Consequently it is usually possible to allow $I_{C(\mathrm{OFF})}$ to exceed I_{CS} or even I_{CE0}, in which case V_{BE} can become slightly positive and R_X can be removed, or taken to zero volts instead of to the negative supply, E_{BB}. In other words, Fig. 2.1(b) is needed when germanium transistors are used, but Fig. 2.1(a) is normally satisfactory with silicon transistors.

The Bottomed State Again

The equation for the saturated state of the inverter in Fig. 2.1(b) is evidently now

$$I_{B(\mathrm{ON})}=\frac{V_{g(\mathrm{ON})}-V_{BE(\mathrm{ON})}}{R_g+R_B}-\frac{V_{BE(\mathrm{ON})}-E_{BB}}{R_X}. \tag{2.4}$$

Strictly speaking, eqns. (2.2) or (2.4) can only be solved for $I_{B(\mathrm{ON})}$, say, given R_B, R_g, R_X, E_{BB}, by solving them simultaneously with the equation for the transistor input characteristic, namely $I_{B(\mathrm{ON})} \approx \mathrm{const} \times \exp(qV_{BE(\mathrm{ON})}/kT)$. Drawing a load line is one way of achieving this, but in practice it is normally

sufficiently accurate to assume that $V_{BE(ON)}$ is a constant of about 0·2 V for germanium transistors and about 0·7 V for silicon low level switching transistors.

The Transient Response

It is convenient to reduce the circuit of Fig. 2.1(b) to its Thévenin equivalent, shown in Fig. 2.1(c) where (using Thévenin's theorem)

$$V_T = \frac{V_g R_X + E_{BB}(R_g + R_B)}{R_X + R_g + R_B} \tag{2.5}$$

and

$$R_T = \frac{R_X(R_g + R_B)}{R_X + R_g + R_B}. \tag{2.6}$$

Using this circuit the transient response expressions derived in Chapter 1 can at once be applied.

The equivalent circuit containing V_g and R_g used for the generator in Fig. 2.1(a) may not be quite correct if the generator is another transistor inverter, for the following reason. When the driving transistor is cut-off, the output resistance of the generator is simply the load resistance R_L and $V_g \approx E_{CC}$. When the driving transistor is bottomed the output resistance is very low so long as no more collector current than β times its base current is demanded from it. In effect this means that when V_g is rising to $V_{g(ON)}$, $R_g = R_L$, but when V_g is falling to $V_{g(OFF)}$, $R_g \ll R_L$. These values also affect R_T and V_T.

The Inverter as a Chopper

It is often useful to reduce a varying signal of amplitude v_s (the output of some measuring instrument, for example) to a series of pulses of amplitude proportional to the signal. This is

called *chopping* or *pulse amplitude modulation* and can be achieved using the inverter circuit by replacing E_{CC} by $E_{CC}+v_s$.

When the transistor is conducting the output voltage is $V_{CE(sat)}$ and when the transistor is cut-off the output is $(E_{CC}+v_s-I_{C(OFF)}R_L)$. Thus, as for an inverter, the condition for a good chopper is that $I_{C(OFF)}$ and $V_{CE(sat)}$ should be as small as possible.

If the transistor is inverted as in Fig. 2.2, then the cut-off current flowing in R_L is I_{EX}, and is given by eq. (1.11):

$$I_{EX} = \frac{I_{EB0}(1-\alpha_N)}{1-\alpha_I\alpha_N}.$$

In a typical transistor in which $\alpha_N > \alpha_I$ it is usual to find that I_{EX} may be an order of magnitude less than I_{CX}.

When a transistor is bottomed at a particular value of collector current, $V_{CE(sat)}$ decreases as I_B is increased, as can be seen from eqn. (1.13). In the limit

$$V_{CE(sat)} = \frac{kT}{q} \ln \frac{1}{\alpha_I} \quad \text{when} \quad I_C = 0. \tag{2.7}$$

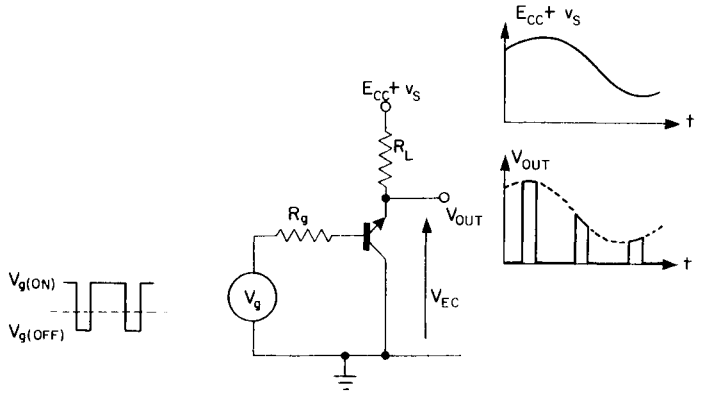


FIG. 2.2. The inverse connection of a transistor in a simple chopper.

If the transistor is inverted, so that the emitter is now performing the function of the collector, the emitter–collector saturation voltage $V_{EC(sat)}$ is given by

$$V_{EC(sat)} = \frac{kT}{q} \ln \left\{ \frac{I_B + I_E(1-\alpha_N)}{\alpha_N[I_B - I_E(1-\alpha_I)/\alpha_I]} \right\}. \qquad (2.8)$$

In the limit, when $I_E = 0$,

$$V_{EC(sat)} = \frac{kT}{q} \ln \frac{1}{\alpha_N},$$

and $V_{EC(sat)}$ is significantly less than $V_{CE(sat)}$ if $1 \approx \alpha_N > \alpha_I$.

Usually $V_{EC(sat)} < V_{CE(sat)}$ only if $I_B \geqslant I_{C(ON)}$, so that the transistor must be heavily overdriven to achieve the advantage of inverse operation in saturation. Thus theoretically, the transistor is a better chopper, both when cut-off and when bottomed, if it is operated in the inverse connection.

In practice symmetrical chopper circuits,[5] in which the signal can be either polarity, are normally used in instrumentation.

The Emitter Follower or Buffer

The emitter follower shown in Fig. 2.3(a) gives a voltage gain of slightly less than unity, but can provide a current gain of nearly β. In switching circuitry the emitter follower is usually

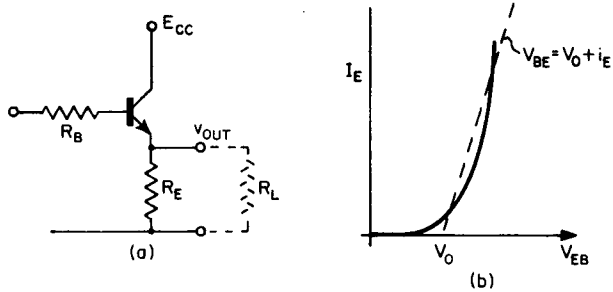


FIG. 2.3. (a) The emitter follower or buffer; (b) the transistor input characteristic and a linear approximation to it.

used as a *d.c. buffer* stage; that is, as a coupling element between a load which demands a large current, and a driving circuit which can only supply a small one.

If we assume that the emitter junction characteristic is represented by the equation $V_{BE} = V_0 + i_E r_E$, as illustrated in Fig. 2.3(b), then we can set up simple linear equations to describe the behaviour of the circuit, namely

$$i_B = [v_{IN} - (V_0 + i_E r_E) - v_{OUT}]/R_B, \tag{2.9}$$

$$i_E = v_{OUT}/R_E. \tag{2.10}$$

These relate the transistor currents to the circuit elements. i_B and i_E can now also be related through the charge control equations. Thus from eqns. (1.21) and (1.22)

$$i_B = \frac{q_B}{\beta\tau_C} + \left(1 + \frac{Q_{VC}}{Q_B}\right)\frac{\mathrm{d}q_B}{\mathrm{d}t}, \tag{2.11}$$

$$i_E = i_C + i_B = \frac{\beta+1}{\beta}\cdot\frac{q_B}{\tau_C} + \frac{\mathrm{d}q_B}{\mathrm{d}t}. \tag{2.12}$$

With a step voltage applied at the input, so that v_{IN} becomes a constant, V_{IN}, these equations can readily be solved. In particular, if R_B is not too small $\mathrm{d}q_B/\mathrm{d}t$ can usually be neglected with respect to q_B/τ_C in eqn. (2.12) (even though it cannot be neglected in eqn. (2.11)). Substituting q_B from eqn. (2.12) into (2.11) and solving yields

$$v_{OUT} = \frac{V_{IN} - V_0}{A}(1 - \mathrm{e}^{-At/B}), \tag{2.13}$$

where

$$A = 1 + \frac{1}{R_E}\left(r_E + \frac{R_B}{\beta+1}\right)$$

$$B = \frac{R_B}{R_E}\cdot\frac{\beta\tau_C}{\beta+1}\left(1 + \frac{Q_{VC}}{Q_B}\right).$$

Thus the output voltage rises exponentially to its steady state value of

$$V_{OUT} = \frac{(V_{IN} - V_0)R_E}{R_E + r_E + R_B(\beta + 1)} \tag{2.14}$$

with a time constant which is reduced by the presence of the emitter resistor R_E. That is,

$$\frac{B}{A} \approx \frac{\tau_C[1 + (Q_{VC}/Q_B)]}{1/\beta + (R_E + r_E)/R_B}. \tag{2.15}$$

If R_B, which includes the generator resistance, is very small, then the second order differential eqns. (2.9)–(2.12) must be solved without approximation.

If the output drives a load R_L, as shown dashed in Fig. 2.3(a), then R_E must be replaced by $R_E R_L/(R_E + R_L)$ in the above equation.

Since a load equal to the output resistance, R_{OUT}, of the circuit halves the voltage gain, it follows from eqn. (2.14) that:

$$\frac{1}{R_{OUT}} = \frac{1}{R_E} + \frac{1}{r_E + R_B/(\beta + 1)}. \tag{2.16}$$

Now r_E is approximately the slope of the emitter junction characteristic at the mid-value of the emitter current. For example, at $I_E = 5$ mA:

$$r_E \approx \frac{kT}{qI_E} = 5\,\Omega.$$

Thus for a circuit operating at 5 mA mean current and using a transistor with large β, $R_{OUT} \ll R_E$, and the circuit will drive large loads—provided they are directly coupled.

If the load is capacitively coupled, as in Fig. 2.4, the input voltage will normally now be a periodic one and distortion of the signal is likely to occur if $R_L \ll R_E$ and if significant voltage swings are involved. The reason for this is as follows.

Since, as shown, the output is now an alternating one, the current demanded by the load is also alternating. But the emitter

follower can only supply a large current in one direction, so that the output tends to clip if R_L is small and demands too much current in the other direction.

If the waveform is just about to be clipped the instantaneous total emitter current i_E will fall to zero at the negative peaks of the signal. At these instants the value of the instantaneous load current i_e must exactly balance the mean d.c. emitter current I_E since

$$i_E = i_e + I_E.$$

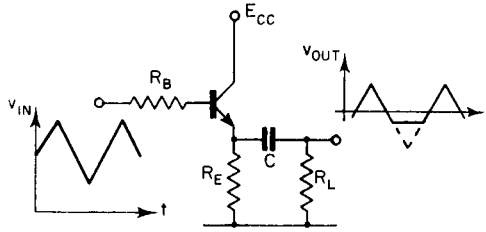


FIG. 2.4. An emitter-follower with capacitor-coupled load, showing the distortion that can result.

Thus if the peak negative excursion of the output voltage exceeds $I_E R_L$, the transistor will become cut-off and the waveform will be clipped.

Thus, if the peak value of v_{OUT} exceeds R_L/R_E times the mean value of emitter voltage, distortion will occur.

To avoid this difficulty push–pull emitter followers can be made using an *npn* and *pnp* transistor in which each transistor is cut off in turn whilst the other supplies the load current.[6]

An alternative circuit involving *npn* (or *pnp*) transistors only is shown as the output of H.L.T.T.L. gates on p. 108.

The Long-tailed Pair

The circuit of Fig. 2.5(a), often called the Long-tailed Pair, is a very important and versatile circuit. It operates in the following way.

The emitter resistor R_E, and the supply voltage E_{EE}, are intended to be sufficiently large to ensure that small changes of v_{IN} do not significantly affect the total current I_{EE} shared by the transistors.

Suppose the two transistors are identical. Then if $V_{BE1} > V_{BE2}$, I_{E1} will be greater than I_{E2}, and vice versa.

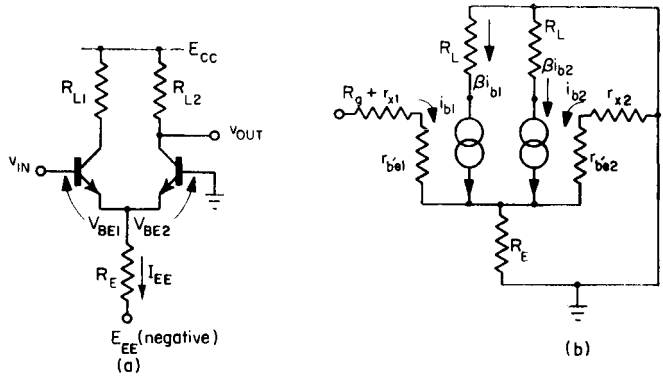


FIG. 2.5. The long-tailed pair. (a) The basic circuit; (b) its small-signal equivalent circuit. (Typical values: $E_{CC} = -E_{BB} = 10$ V, $R_E = R_{L1} = R_{L2} = 4{\cdot}7$ kΩ.)

If the emitter cut-off currents of the two transistors are the same, namely I_0, then, approximately,

$$I_{E1} = I_0(e^{qV_{EB1}/kT}) \tag{2.17}$$

and

$$I_{E2} = I_0(e^{qV_{EB2}/kT}),$$

so that

$$V_{IN} = V_{EB1} - V_{EB2} = \frac{kT}{q} \ln (I_{E1}/I_{E2}) \tag{2.18}$$

$$= \frac{kT}{q} \ln \frac{I_{EE} - I_{E2}}{I_{E2}}. \tag{2.19}$$

Thus the change in input voltage necessary to achieve a change of I_{E2} from 1% to 99% of I_{EE} is $9{\cdot}2\ kT/q$. In other words, a

change in V_{IN} of about 240 mV is required to switch the current from one transistor to the other. This is true whether the transistors are made from germanium or from silicon.

Within the range $V_{B1} = V_{B2} \pm 120$ mV the circuit behaves as an amplifier whose voltage gain rises to a maximum when the d.c. currents in the two transistors are equal. Using a simple, low-frequency, common-emitter small-signal equivalent circuit of a transistor, the equivalent circuit for the long-tailed pair is as shown in Fig. 2.5(b) (a more accurate transistor equivalent circuit yields only slightly improved results). If the current gains of the two transistors are the same it is a simple matter to show that if R_E is large the voltage gain A_V of the circuit is

$$A_V \approx \frac{R_L \cdot \beta}{R_g + r_{b'e1} + r_{b'e2} + r_{x2} + r_{x1}}, \tag{2.20}$$

where r_{x1} and r_{x2} are the extrinsic base resistances of the transistors.

Now $r_{b'e1}/\beta \approx r_{e1}$ the emitter junction resistance[2] of T_1 and $r_{b'e2}/\beta \approx r_{e2}$ the junction resistance of T_2, where, as before,

$$r_{e1} = \frac{kT}{qI_{E1}} \quad \text{and} \quad r_{e2} = \frac{kT}{qI_{E2}},$$

so that A_V becomes a maximum when $I_{E1} = I_{E2}$. (Linearity can be increased by including resistors in each emitter lead though the maximum gain is reduced.)

Later we shall be considering the long-tailed pair as part of a regenerative switch in which the output is fed back to the input.

If the base of T_2 is regarded as a second input it is evident that the circuit is a *differential amplifier* with an input dynamic range of about 200 mV. With equal load resistors on the transistors it also has a differential output.

The circuit can also be regarded as a *zero-crossing detector* or a *level detector*. If the base of T_2 is held at a constant voltage V_{B2} (between V_{CC} and V_{EE}) then the current will switch from one transistor to the other as the input passes through V_{B2}.

On subsequent pages, the various aspects of long-tailed pair behaviour are discussed in more detail, as follows:

As a clipping circuit or level detector on p. 71.

As a regenerative level detector (Schmitt trigger) on p. 73.

As a high-speed current switch on p. 111.

Problems

2.1. (a) An inverter as shown in Fig. 2.1(b) has the following parameter values: $E_{CC} = +6$ V, $R_L = 1$ kΩ, $V_{g(ON)} = +6$ V, $V_{g(OFF)} = +0{\cdot}2$ V, $R_B + R_g = 33$ kΩ, $E_{BB} = -6$ V.

If the circuit is designed so that $V_{BE(OFF)} = -1$ V, calculate the value of R_X (neglecting cut-off currents).

With this value of R_X what is the minimum value of β which will ensure the circuit is bottomed in the conducting state, assuming that $V_{BE(ON)} = 0{\cdot}7$ V and $V_{CE(sat)} = 0$?

(Ans.: 138 kΩ, 53·5.)

(b) If the minimum value of β of a batch of transistors is 25, redesign the circuit (i.e. calculate new values of $R_B + R_g$, R_X) so that again $V_{BE(OFF)} = -1$ V and so that all transistors are bottomed when in the conducting state.

(Ans.: 15·4 kΩ, 64 kΩ.)

2.2. Verify eqns. (2.13), (2.14), (2.15), given eqns. (2.11), (2.12) describing the transistor. Calculate the time for the output voltage of an emitter follower to reach 90% of $(V_{IN} - V_0)$ if $r_E = 5$ Ω, $R_B = 5$ kΩ, $R_E = 2$ kΩ and if $\tau_0 = 0{\cdot}02$ μsec, $Q_{VC} = 120$ pC, $\beta = 100$, and $V_{IN} - V_0 = 10$ V. (Ans.: 0·275 μsec.)

2.3. In a long-tailed pair containing two identical transistors what is the change in differential input voltage necessary to change the ratio of current flow in the two transistors from 3:1 to 1:3? Assume $kT/q = 25$ mV.

(Ans.: 55 mV.)

2.4. Derive eqn. (2.20).

CHAPTER 3

Waveform Generation and Wave Shaping

IN THIS chapter some of the more important methods of producing non-sinusoidal waveforms, such as square waves, or sawtooth waves, will be described. It is assumed that each circuit is designed to produce an ideal waveform, so that part of the analysis of each circuit involves assessing the causes of any departures from the ideal in the performance of the circuit.

Rectangular Waveform Generation

1. *The Astable Multivibrator*

The astable multivibrator is shown in Fig. 3.1(a). It has a stable condition with both transistors conducting and bottomed, and in which no oscillations occur. In these circumstances the capacitors are evidently inactive and can be ignored. From this diagram it can be seen that if each transistor is to be held bottomed, then $\beta I_{B(ON)} \geqslant I_{C(ON)}$, or, neglecting $V_{BE(ON)}$,

$$\frac{\beta_1 E_{BB}}{R_{B1}} \geqslant \frac{E_{CC}}{R_{L1}}, \quad \frac{\beta_2 E_{BB}}{R_{B2}} \geqslant \frac{E_{CC}}{R_{L2}}. \tag{3.1}$$

Now suppose the base supply E_{BB} is temporarily switched off, so that the base current of each transistor falls to near zero, and the output voltage rises to near E_{CC}. If now the base supply is switched on again both transistors will be turned on, the collec-

tor currents will tend to rise, and the collector voltages to fall. But with the capacitors connected, as in Fig. 3.1(a), these changing collector voltages will be coupled to the opposite base terminals, tending to drive the transistors off again. Actually one transistor will always turn on more rapidly than the other, with the consequence that it will be more successful in turning the other off, which allows it to turn on even more rapidly, and the

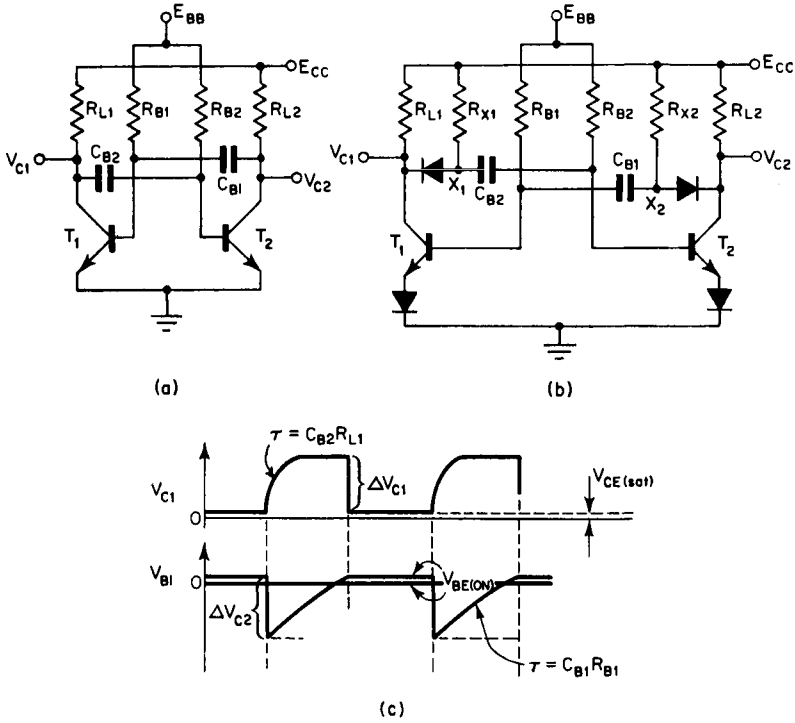


FIG. 3.1. The astable multivibrator. (a) The basic circuit; (b) the circuit modified to improve the rise time of the output voltage, and to protect the emitters from damage due to emitter junction breakdown; (c) the collector and base voltage waveforms of the basic circuit. (Typical values: $R_L = 1$ kΩ, $R_B = 22$ kΩ, $E_{CC} = E_{BB} = 6$ V in circuit (a). In (b), $R_L = R_X = 2{\cdot}2$ kΩ.)

circuit soon reaches a state with one transistor on and the other cut off.

But this state is only temporarily stable, since when T_2, say, is cut-off, the current through R_{B_2} will charge capacitor C_{B_2} and will eventually forward bias the emitter–base junction of T_2. Then T_2 will conduct, V_{C_2} (the collector voltage of T_2) will fall and T_1 will be turned off. Now the current through R_{B_1} charges up C_{B_1}, and, after a time, T_1 will conduct again. Thus the circuit oscillates, first one transistor is cut off, then the other.

Timing

To a first approximation, when one transistor, say T_2, is driven into saturation the whole of the *change* of its collector voltage ΔV_{C_2} ($\approx E_{CC}$) is transferred via C_{B_1} to the base of T_1. Thus V_{B_1} falls by ΔV_{C_2} to the voltage $V_{BE(ON)} - \Delta V_{C_2}$, as shown in Fig. 3.1(c). Thereafter V_{B_1} rises towards E_{BB} as the current through R_{B_1} charges C_{B_1}. Since the other plate of the capacitor is clamped to near zero volts by T_2 in the bottomed state ($V_{C_1} = V_{CE(sat)}$), the following equation applies,

$$C_{B1} \frac{dV_{B1}}{dt} = \frac{E_{BB} - V_{B1}}{R_{B1}}. \tag{3.2}$$

T_1 begins to conduct again when $V_{B_1} \approx V_{BE(ON)}$, so that if at $t = 0$, $V_{B_1} = V_{BE(ON)} - E_{CC}$, the time t_1 for which T_1 is cut-off becomes (solving eqn. (3.2)):

$$t_1 = C_{B1}R_{B1} \ln \left[1 + \frac{E_{CC}}{E_{BB} - V_{BE(ON)}}\right]. \tag{3.3}$$

A similar expression gives the period t_2 for which T_2 is cut-off. If $E_{BB} = E_{CC}$, and if $V_{BE(ON)} \ll E_{BB}$,

$$t_1 = C_{B1}R_{B1} \ln 2 \quad \text{and} \quad t_2 = C_{B2}R_{B2} \ln 2. \tag{3.4}$$

If the time during which switching occurs is negligible, then the frequency of oscillation is $f = 1/(t_1 + t_2)$ Hz.

The only transistor parameter in this expression is $V_{BE(ON)}$. In practice the magnitude of the base charge and the magnitude of the base cut-off current affect the timing too, as follows:

When T_1 is being turned on, T_2 will only become cut-off once a charge Q_{OFF} has been removed from its base. This means that the first negative-going part of the V_{C1} transient is used in discharging the base region of T_2 without any significant change in V_{B2}. The negative bias finally applied to the base of T_2 is therefore less than E_{CC} by Q_{OFF}/C_{B2}, and $V_{B2} = V_{BE(ON)} - (E_{CC} - Q_{OFF}/C_{B2})$ at $t = 0$. Whence

$$t_2 = C_{B2}R_{B2} \ln \left(1 + \frac{E_{CC} - Q_{OFF}/C_{B2}}{E_{BB} - V_{BE(ON)}}\right). \qquad (3.5)$$

Thus if C_{B2} is so small that $Q_{OFF}/C_{B2} = E_{CC}$, $t_2 = 0$ and no oscillations can occur. However, in practice, at the highest frequencies this expression is inaccurate because the time during which both transistors are in the Active Region of Operation becomes a significant part of the total cycle time and cannot be assumed negligible. But at this speed the waveform is near sinusoidal and does not concern us here.

Some diffused base transistors can withstand only 2 or 3 volts of reverse bias on their base–emitter junctions without the junction breaking down and possibly being damaged (see p. 12). Thus if E_{CC} is more than a few volts, such transistors might be damaged in the multivibrator during their cut-off periods. To prevent this, two methods are available.

Either a diode can be placed in series with each emitter, as in Fig. 3.1(b), so that if the emitter junction does break down the current is limited to the reverse current of the diode; or the diode can be placed in parallel with the junction so that a reverse bias greater than the forward conduction voltage of the diode cannot be applied to the emitter junction of the transistor. The first method allows the junction to break down, but because the current is limited no damage is caused. It is the better method

to use in this case since the timing of the multivibrator is only slightly affected. Both $V_{CE(sat)}$ and $V_{BE(ON)}$ are increased by the diode forward voltage, and the two effects almost balance.

At high temperatures the base current, I_{BX}, of the cut-off transistor, may not be negligible compared with the current through R_B.

Equation (3.2) becomes:

$$C_{B2}\frac{dV_{B2}}{dt} = \frac{E_{BB}-V_{B2}}{R_{B2}}+I_{BX} \tag{3.6}$$

whence

$$t_2 = C_{B2}R_{B2}\ln\left(1+\frac{E_{CC}-Q_{OFF}/C_{B2}}{E_{BB}-V_{BE(ON)}+I_{BX}R_{B2})}\right). \tag{3.7}$$

$V_{BE(ON)}$ decreases by 1·5 or 2·0 mV per degree centigrade (depending upon whether silicon or germanium transistors are being used) and this will also affect the timing lightly.

Waveform

The output waveform of the multivibrator is shown in Fig. 3.1(c). The relatively slow exponential rise of the output voltage appears at the collector of the transistor that has just been cut off. Since its collector current is almost zero, V_{C1} rises to E_{CC} exponentially with a time constant of $C_{B2}R_{L1}$.

Since the period for which T_1 is bottomed is about $C_{B2}R_{B2}$ it follows that this slowly rising edge must occupy a fraction of the waveform of at least R_{L1}/R_{B2}, so that with this multivibrator circuit the slowly rising turn-off edge will always be noticeable. This effect also limits the degree of timing asymmetry which the circuit can produce. If, for example, C_{B2} is made larger than C_{B1}, then T_2 will be cut-off for longer than T_1 (assuming $R_{B1} = R_{B2}$), and the rise of the collector voltage of T_1 will be

slower than the rise of V_{C2}. Thus with a timing asymmetry of only 20:1 or so, the output voltage of T_1 will not have risen to E_{CC} before the transistor is driven into the conducting state again. When this happens ΔV_{C1} (the change in V_{C1} at the switching transient) falls, the oscillation frequency increases, thus further accentuating the trouble. Thus with asymmetries of greater than about 10:1 the timing equations begin to be in error and above 20:1 or so the circuit may not oscillate.

It is possible to modify the circuit to achieve a greatly improved output waveform, but not to solve the timing problem just described. The improved circuit is shown in Fig. 3.1(b).

Whilst T_1 is being turned on, or is held conducting, its collector current flows through both R_{L1} and R_{X1}. When T_1 is turned off the voltage of X_1 rises with time constant $R_{X1}C_{B2}$, but the output voltage V_{C1} now rises rapidly since the diode becomes cut-off and disconnects the output from C_{B2}.

A similar addition of resistor and diode can be used to improve the output waveform of T_2.

The diode cut-off currents modify the timing in the same way as I_{BX}.

The fall time of the output voltage is much faster than the rise time, as shown in Fig. 3.1(c). The reason for this is that most of the turn-on of T_1 takes place with T_2 cut-off. But when T_2 is cut-off all the current through R_{L2}, nearly E_{CC}/R_{L2}, flows through C_{B1} into the base of T_1. Thus T_1 is a heavily overdriven transistor in which the base current is approximately equal to the final collector current. Thus a rise time approaching $\tau_C(1+Q_{VC}/Q_B)$ is to be expected. (This follows from eqn. (1.35) with $I_B \approx I_C$.)

The proportion of a period during which *both* transistors are in the Active Region of Operation is very small at slow oscillation frequencies, but increases as the frequency of oscillation rises. Thus, as already pointed out, the equations given become less accurate as the maximum frequency is approached.

Application

The multivibrator can, of course, produce square waves as well as rectangular waves simply by ensuring symmetry between the two halves of the circuit.

The frequency of oscillation can be easily controlled by E_{BB}; indeed with a signal added to E_{BB} a multivibrator is a very simple method of producing a frequency-modulated square wave.

In addition, it is possible to synchronize the multivibrator oscillations to another, higher frequency which is nearly a harmonic. If the signal is fed to the base of one transistor, via a resistor, then the exponentially rising base waveform will have the synchronizing signal superimposed upon it and the multivibrator will switch only when the signal is nearing its most positive excursion.

Alternatively, a small resistor can be placed in series with the emitter of one transistor, and the signal applied across this resistor. This time switching will take place near the most negative excursion of the synchronizing signal.

A Current-steered Multivibrator[7]

A circuit capable of producing rectangular waves with very sharp rising and falling edges, at mark-space ratios of between 1:1 and 500:1, and which will operate up to the highest speeds relative to the transistor capabilities, is shown in Fig. 3.2.

The ideal circuit is shown in Fig. 3.2(a) and a practical realization of it is shown in Fig. 3.2(b). E_{CC1} must be larger than E_{CC2} (in Fig. 3.2(a)) to prevent T_1 bottoming when it is conducting. In Fig. 3.2(b) this is achieved by the inclusion of R_X.

The currents I_1, I_2 are not quite constant in Fig. 3.2(b) since the emitter voltages vary during the oscillations by the same amount as the collector voltage of T_2.

The operation of the idealized circuit can be understood as follows.

As with the multivibrator the two transistors take it in turns to conduct or be cut-off, so that the total current I_1+I_2 (which is fixed) is switched between T_1 and T_2.

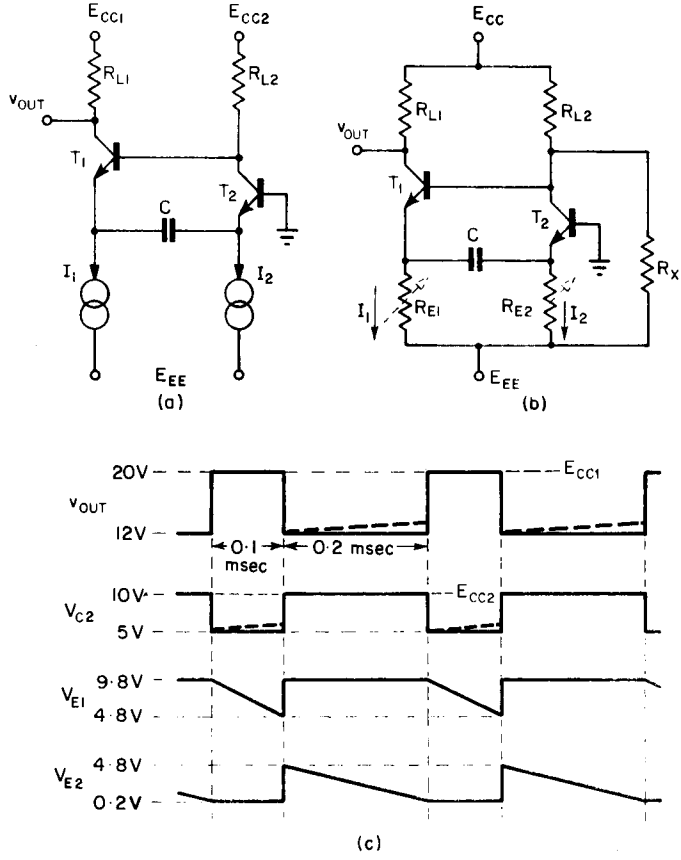


FIG. 3.2. A current steered multivibrator. (a) The ideal circuit; (b) a practical circuit; (c) the voltage waveforms at the emitters and collectors of the ideal circuit. (Typical values for (b): $E_{CC} = -E_{EE} = 10$ V, $R_L = 1$ kΩ, $R_X = 3{\cdot}9$ kΩ, $R_{E1} = 15$ kΩ, $R_{E2} = 10$ kΩ for about 1:1 ON–OFF ratio.)

When T_2 has just begun to conduct its collector voltage falls by $R_{L2}(I_1+I_2)$. Since the emitter voltage of T_2 cannot change (now that T_2 is conducting), the emitter voltage of T_1 cannot immediately change either (because of C), so the emitter junction of T_1 acquires a reverse bias of $R_{L2}(I_1+I_2)$. I_1 is now diverted to flow through C, and consequently lowers the emitter voltage of T_1 (at a rate of $\mathrm{d}V_{E1}/\mathrm{d}t = I_1/C$) until T_1 begins to conduct again. T_1 now turns on regeneratively (since the loop gain is greater than 1) and T_2 turns off. The output voltage of T_2 rises to E_{CC2} thus transferring a reverse bias of $R_{L2}(I_1+I_2)$ to the emitter of T_2 via T_1 (which is now conducting) and via C. The capacitor C is now discharged again from the other side, by I_2, until T_2 again conducts. I_1, I_2 and R_{L2} are chosen so that T_2 never bottoms.

The output is best taken from the collector of T_1 because this output is not part of the feedback loop in the oscillator but E_{CC1} must be greater than $E_{CC2}+R_{L1}(I_1+I_2)$. Also the output voltage can be varied by R_{L1} without affecting the frequency. Loading V_{C2} affects R_{L2} and therefore changes the circuit timing.

Timing

If we neglect the base charge of the OFF going transistors and any cut-off currents, the time t_1 for which T_1 remains cut-off is given by

$$t_1 = \frac{R_{L2}(I_1+I_2)C}{I_1} \tag{3.8}$$

and the time for which T_2 remains cut-off is

$$t_2 = \frac{R_{L2}(I_1+I_2)C}{I_2}, \tag{3.9}$$

so that the total cycle time is

$$t_1+t_2 = \frac{(I_1+I_2)^2 R_{L2} C}{I_1 I_2} = 4R_{L2}C \quad \text{if} \quad I_1 = I_2 . \tag{3.10}$$

If the turn-off base charge of each transistor is significant it reduces the magnitude of the reverse bias on the emitter junctions by Q_{OFF}/C, so that

$$t_1+t_2 = \frac{I_1+I_2}{I_1 I_2}[(I_1+I_2)R_{L2}C - Q_{OFF}]. \tag{3.11}$$

The timing of this circuit is much less dependent upon temperature than is that of the normal multivibrator, because firstly, the temperature-dependent transistor cut-off current which affects the timing is I_{EX}, the emitter cut-off current. But I_{EX} is very small—usually about an order of magnitude less than I_{EB0} (see eqn. (1.11)). Secondly, the variation of $V_{BE(ON)}$ with temperature only causes a second-order variation of timing arising out of differences between I_1 and I_2. If $I_1 = I_2$ the timing is not affected at all by changes in $V_{BE(ON)}$.

Similarly the timing is only slightly affected by changes in β. When T_2 is cut-off, V_{C2} is less than E_{CC2} because some base current flows into T_1, and this depends on β_1. Also β_2 determines the lower value of V_{C2} when T_2 is conducting.

The practical circuit of Fig. 3.2(b) affects the circuit behaviour as follows:

1. The effective collector supply voltage of T_2, E_T is (by Thévenin's theorem) $(E_{CC}R_X + E_{EE}R_{L2})/(R_X + R_{L2})$, and the effective load resistance R_T is $R_X R_{L2}/(R_X + R_{L2})$.

2. Since V_{E1} and V_{E2} fall during the timing periods, it follows that I_1 and I_2 must also fall, with the result that the lower output voltage level rises by an amount depending upon the magnitude of V_{C2}, V_{E1} and V_{E2} as compared with E_{EE}. The rising values

of V_{OUT} and V_{C2}, which occur in the practical circuit, are shown dashed on the ideal waveforms of Fig. 3.2(c).

Waveform

The rise and fall times of the waveform are very short because (a) the loop gain is large; (b) the transistors are never bottomed; (c) when T_2 is cut-off the capacitive load on V_{C2} is composed only of transistor junction capacitances (collector and emitter of T_2 and the collector of T_1) which are all small (even though the last is amplified a little by the Miller effect).

Application

As compared with the astable multivibrator, this circuit generates square waves with faster edges, of more stable repetition rate and with a much wider range of on–off time ratios (500 : 1 as compared with about 10 : 1). Its frequency can be varied, without altering this ratio by simply varying C, and its amplitude can be varied, without altering the frequency greatly, by varying R_{L1}.

On the other hand, its frequency cannot be made dependent upon an input voltage (changing E_{EE} only alters the amplitude); neither output voltage is at zero volts (the common return of the d.c. supplies) and the lower output level does not remain constant.

Triggered Pulse Generators (Monostable Circuits)

Monostable circuits have one permanently stable operating state. The other state is like each state of the astable multivibrator; it is only temporarily stable, its duration being determined by some C-R or L-R timing circuit. To operate it the circuit is triggered by an external source into the temporarily stable state.

The Monostable Multivibrator

In the monostable multivibrator one transistor is biased to remain stable in the cut-off state whilst the other is held bottomed. But the circuit can be driven into a state in which the transistor states are interchanged. This state is "temporarily stable", just as each state in an astable multivibrator, and the circuit switches back to the first state, after a time determined by a C-R time constant. The circuit of the monostable multivibrator is shown by the heavy lines in Fig. 3.3. The lighter lines indicate triggering circuits which will be discussed shortly.

For T_2 to be stable when bottomed

$$\frac{E_{BB2}-V_{BE(ON)}}{R_{B2}} \geqslant \frac{E_{CC}}{\beta_{min}R_{L2}}. \tag{3.12}$$

For T_1 to be temporarily stable when bottomed

$$\frac{E_{CC}}{R_{L2}+R_{B1}} \geqslant \frac{E_{CC}}{\beta_{min}R_{L1}} + \frac{V_{BE(ON)}-E_{BB1}}{R_X}. \tag{3.13}$$

For T_1 to be stable when cut-off (with T_2 bottomed)

$$\frac{V_{BE(OFF)}-E_{BB1}}{R_X} \geqslant I_{BX} + \frac{V_{CE(sat)}-V_{BE(OFF)}}{R_{B1}}, \tag{3.14}$$

where $V_{BE(OFF)}$ must be zero volts or less with germanium transistors under worst case conditions, or less than a few tenths of a volt positive with silicon transistors.

The circuit is operated by driving T_2 into the cut-off state (or by driving T_1 on) by some external trigger source. Regeneration completes the transition to the temporarily stable state. But this is only stable for a short period, determined by the rate at which C_{B2} is charged by R_{B2}, after which the circuit returns to its permanently stable condition with T_2 bottomed. The period t_2 for which T_2 stays cut-off is dependent on the same parameters

as the period of the astable multivibrator. Thus approximately

$$t_2 = C_{B2}R_{B2} \ln \left[1+\frac{E_{CC}}{E_{BB1}-V_{BE(ON)}}\right]. \qquad (3.15)$$

More accurate expressions are given in eqns. (3.5) and (3.7).

The external trigger can be applied in various ways, three of which are shown in Fig. 3.3. Each trigger circuit consists of a

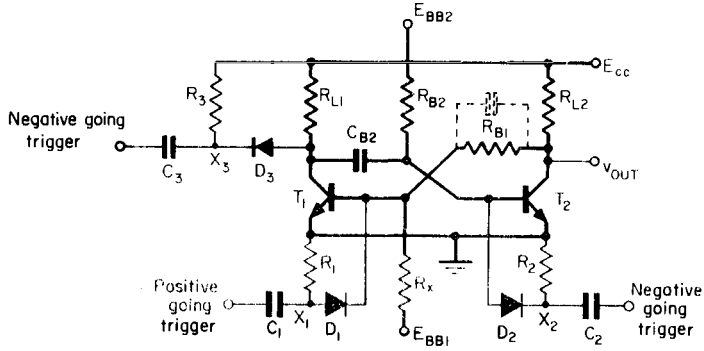


FIG. 3.3. The monostable multivibrator showing three possible trigger inputs. (Typical values: $R_L=1$ kΩ, $R_B=18$ kΩ, R_X (only needed for Ge transistors)$=180$ kΩ, R_1 or R_2 or $R_3 \approx 15$ kΩ.)

resistor, a capacitor and a diode, and is distinguished by a subscript 1, 2 or 3. R_1, C_1 and D_1 allow a positive going voltage step to reach the base of T_1, turning it on. R_2, C_2 and D_2 allow a negative going step to reach the base of T_2 turning it off. R_3, C_3 and D_3 allow a negative going step to reach the collector of T_1 and (since T_1 is normally cut-off) to pass via C_{B2} to the base of T_2, again turning T_2 off.

In each circuit the diode is included to prevent the wrong polarity edges reaching the circuit (a precaution which it is usually necessary to take) and to isolate the driving circuit from the multivibrator voltage transients.

Each circuit has its limitations. Circuit 1 turns on T_1 so that the trigger source "sees" D_1 and the emitter junction of T_1 as two forward-biased junctions in series. Thus the circuit loads the source severely. Circuit 2 is the most sensitive and loads the source very lightly (since T_2 turns off) but it affects the timing of the circuit (i.e. C_2 removes charge from the base of T_2 which has to be replaced by R_{B2} during the timing period). Circuit 3 does not affect the timing; it loads the trigger source moderately but is less sensitive than Circuit 2.

The resistors R_1, R_2 or R_3 are included to ensure that the potentials of points X_1, X_2 or X_3 return to their equilibrium values between the application of trigger pulses. Thus the interval between pulses should not be less than the time constant CR of the triggering circuit.

It is also possible to trigger the circuit by inserting a resistor in the emitter of T_2 so that a positive pulse turns T_2 off, but this reduces the loop gain and affects the timing since the voltage swing of the collector of T_2 is reduced.

The cross-coupling resistor R_{B1} can be shunted by a capacitor (shown dashed) to speed up the regenerative switching of the circuit. Its value should be about an order of magnitude greater than Q_{OFF}/E_{CC} so that only a small change in V_{OUT} will turn T_1 off. It will normally be significantly less than C_{B2}. Its inclusion will, however, slow up the rising edge of the output voltage.

The duration of the positive pulses can easily be controlled by varying E_{BB}. Thus by applying a clock pulse to the triggering input, and by adding a signal to E_{BB} a simple form of pulse width modulation can be produced.

The Emitter-coupled Monostable Circuit

The circuit of Fig. 3.4, like a long-tailed pair, has a common emitter resistance R_E serving as a coupling element between T_1 and T_2. The circuit is stable with T_2 conducting and T_1 cut-off. It

can be made to switch by applying a negative going step or pulse to the pulse input. This turns T_2 off, causing I_{EE} and the voltage of the emitters to fall, bringing T_1 into conduction (since V_{B1} is held by R_{B1}, R_X). The current that previously flowed through T_2 is then directed to T_1 causing V_{C1} to fall. At the completion of the switching process the base of T_2 will have acquired a reverse bias almost equal to the voltage swing of the collector of T_1,

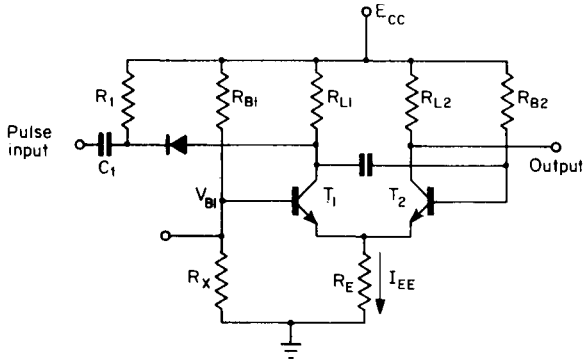


FIG. 3.4. An emitter-coupled monostable circuit. (Typical values: $R_L = R_E = 1$ kΩ, $E_{CC} = 10$ V, $R_{B2} = 22$ kΩ, $R_{B1} = R_X = 4{\cdot}7$ kΩ.)

approximately $I_{EE}R_{L1}$ if β of T_1 is large and if T_1 does not bottom. Thus immediately after the switching process the emitter junction of T_2 is cut-off and

$$V_{B2(\mathrm{OFF})} \approx V_{B1} - I_{EE}R_{L1} = V_{B1} - \left(\frac{V_{B1} - V_{BE(\mathrm{ON})}}{R_E}\right)R_{L1}$$

since T_1 starts to conduct when $V_{B2} \approx V_{B1}$.

The voltage of the base of T_2 now rises exponentially, with a time constant $R_{B2}C_{B2}$, towards E_{CC}, until T_2 begins to conduct again (which again occurs when $V_{B2} \approx V_{B1}$). Thus the time t_2, for which T_2 is cut off, is given by

$$t_2 = C_{B2}R_{B2} \ln\left(1 + \frac{(V_{B1} - V_{BE(\mathrm{ON})})R_{L1}}{(E_{CC} - V_{B1})R_E}\right). \qquad (3.16)$$

Evidently the period t_2 is a function of V_{B1}. Thus, for example, if a low frequency signal is applied to the base of T_1 and a high frequency pulse to the pulse input, a pulse width modulated signal can again be produced.

Apart from being a convenient source of simple pulse width modulated signals, the above two monostable circuits are a convenient way of delaying a negative going transient. If the initial transient is applied to the pulse input, T_2 will initially switch off. After the period t_2, T_2 will switch on again producing a negative going transient at the output. Thus the initial transient has been delayed by t_2. This kind of function is useful in asynchronous digital circuitry where circuit delays have to be introduced sometimes (see Chapter 7).

The Blocking Oscillator[9]

The blocking oscillator uses a transformer as the feedback coupling element, which because of its phase inversion can also replace one of the transistors in the multivibrator. The primary inductance of the transformer is used to obtain the timing of the temporarily stable operating state.

Many configurations of this circuit can be used of which perhaps the commonest is shown in Fig. 3.5(a) with triggering applied across a small resistor R_B in the base lead of the transistor. The circuit is stable with the transistor cut-off. A positive pulse applied to the base starts a current in the collector–emitter circuit. The collector current flowing in the transformer induces an emitter current which, if $N > 1$, is sufficient to bottom the transistor. At this point the equivalent circuit of Fig. 3.5(b) becomes valid.

Initially the current i_L through the primary inductance, L, is small. But since L is finite i_L grows whilst i_E begins to fall. Eventually i_E becomes too small to hold the transistor bottomed; i_C

ceases to rise, cutting off i_E, and the transistor turns off regeneratively.

If R_E is the effective resistance in the secondary circuit when the transistor is bottomed, the effective resistance seen by the collector circuit is N^2R_E and the initial collector current is $E_{CC}/(R_1+N^2R_E)$. As i_L grows the following equations apply:

$$i_C = \frac{E_{CC}}{R_1}(1-e^{-t/\tau})+\frac{E_{CC}}{R_1+N^2R_E}e^{-t/\tau},$$
$$i_E = \frac{NE_{CC}}{R_1+N^2R_E}e^{-t/\tau}, \tag{3.17}$$

where

$$\tau = L\left[\frac{1}{R_1}+\frac{1}{N^2R_E}\right].$$

The transistor cuts off when $\alpha i_E \approx i_C$.

If $t \ll \tau$, and if $\alpha \approx 1$

$$\frac{1}{t} = \frac{N^2R_E}{L}\left(\frac{1}{N-1}+\frac{R_1}{R_1+N^2R_E}\right).$$

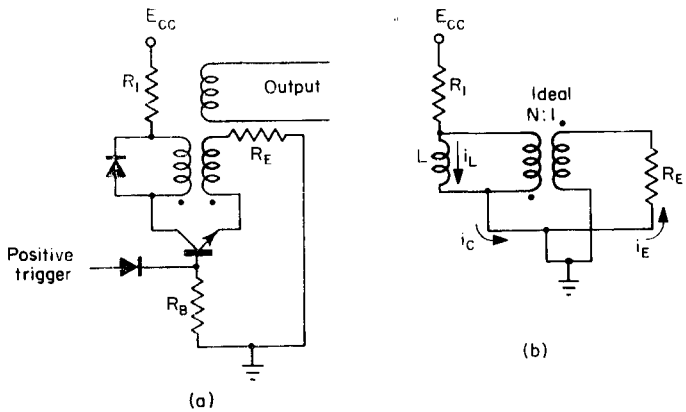


FIG. 3.5. The monostable blocking oscillator. (a) The basic circuit; (b) the equivalent circuit when the transistor is bottomed.

When the transistor cuts off the inductance of the transformer tends to ring and a large reverse voltage can develop at the collector. The inclusion of the damping diode prevents these effects.

The blocking oscillator can also be made astable, for example by shunting R_1 with a capacitor, but the special property of blocking oscillaters, that of producing large *current* pulses, is probably best achieved using the monostable form driven by another oscillator. This is because the timing of the two parts of the oscillation cycle are difficult to control owing to the influence of the transformer imperfections on the oscillator performance.

Sawtooth Waveform Generators

All sawtooth-wave generators integrate a constant voltage or current with respect to time and thus produce a linear ramp. This is subsequently quenched with a gating voltage so that the integration can begin again. The most convenient form of integrator is a capacitor fed by a constant current, and there are three important techniques for producing linear ramps in this way.

1. *Constant-current Source*

In Fig. 3.6(a) a common-base transistor circuit is used to produce a nearly constant current source. When switch S is opened the collector current flows into the capacitor charging it up from an initial value of $-E_{CC}$ towards zero.

Since I_E can be held almost constant at E_{EE}/R_E and $I_C = \alpha_N I_E$, the *variation* of the charging current depends upon how much α_N varies as V_{CB} changes.

The rate of rise of V_{OUT} is given by

$$C\frac{dv_{OUT}}{dt} = I_C.$$

This is the simplest of the three techniques and its linearity can be very good. But its output resistance is very high and any load resistance connected to the output will affect the linearity of the output according to the relation

$$C\frac{dv_{OUT}}{dt}+\frac{v_{OUT}}{R_L}=I_C. \tag{3.18}$$

This represents an exponential rise of output voltage (if I_C is constant) from $-E_{CC}$ towards zero with an initial rate of rise, dv_{OUT}/dt, of $(I_C-E_{CC}/R_L)/C$ and a final rate of rise of I_C/C. Thus for a linear rise $E_{CC}/R_L \ll I_C$.

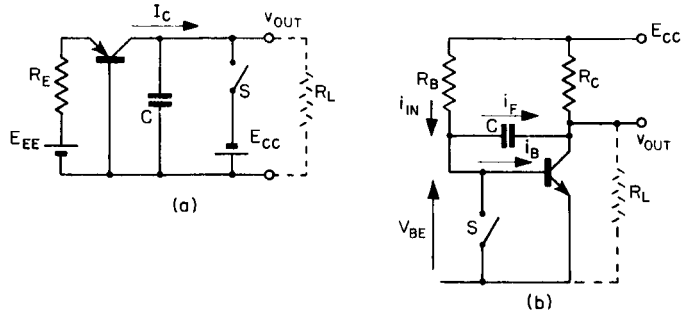


FIG. 3.6. Linear sweep generators: (a) using a transistor as a constant-current source; (b) the Miller integrator. (Typical values: $E_{EE} = |E_{CC}| = 10$ V; $R_E = R_B = 10$ kΩ, $R_C = 1$ kΩ.)

2. *The Miller Integrator*

The Miller integrator circuit uses feedback to keep the current flowing into the capacitor constant. Thus in Fig. 3.6(b) when switch S is opened V_{BE} will tend to move positively and will cause a larger negative excursion of V_{CE} (owing to the gain of the circuit). But this change in V_{CE} will be coupled back to base via C, tending to prevent the original change in V_{BE}. Thus V_{BE} and i_{IN} tend to remain constant. And provided $i_B \ll i_{IN}$, the current i_F into the capacitor stays almost constant too.

On the release of switch S, the base voltage rises rapidly to $V_{BE(ON)}$. This step is coupled to the output via C, so that before the output falls linearly (as just explained) a positive step appears at the output.

The equations describing this circuit, once the transistor is conducting, are evidently (equating currents at the base node):

$$\frac{E_{CC}-V_{BE}}{R_B} = i_B+i_F \tag{3.19}$$

and at the collector node:

$$\frac{E_{CC}-v_{OUT}}{R_C}+i_F = i_C+\frac{v_{OUT}}{R_L}, \tag{3.20}$$

where

$$i_F = C\frac{\mathrm{d}(V_{BE}-v_{OUT})}{\mathrm{d}t}. \tag{3.21}$$

At low frequencies $i_B \approx i_C/\beta$ and if it is assumed that V_{BE} is a constant and that at $t = 0$, $v_{OUT} = E_{CC}R_L/(R_L+R_C)$, eqns. (3.19) to (3.21) yield:

$$v_{OUT}\left(\frac{R_L+R_C}{R_L}\right) = (A+E_{CC})\mathrm{e}^{-t/B}-A, \tag{3.22}$$

where

$$A = E_{CC}\left(\frac{\beta R_C}{R_B}-1\right)-\frac{\beta V_{BE}R_C}{R_B}, \tag{3.23}$$

$$B = C\frac{R_CR_L}{R_C+R_L}(\beta+1). \tag{3.24}$$

If $t \ll B$, v_{OUT} falls linearly. Thus, putting $\beta/(\beta+1) = \alpha_N$,

$$v_{OUT} = \frac{E_{CC}R_L}{R_C+R_L}-\frac{(E_{CC}-V_{BE})\alpha_N t}{CR_B}. \tag{3.25}$$

When v_{OUT} has fallen to near zero volts the transistor will be bottomed and the (nearly) linear ramp output will cut off.

Putting $t = \infty$ into eqn. (3.22) it can be seen that in effect the circuit generates a negative aiming potential for the exponential variation of v_{OUT} equal to $-A$, or approximately $-\beta E_{CC}R_C/R_B$.

When switch S is closed the transistor is cut-off so the output voltage rises to its initial value with a time constant of $CR_CR_L/(R_C+R_L)$. This may not be much smaller than the fall time constant so that the quenching time may not be negligible.

The Bootstrap Sweep Generator

The Bootstrap Sweep Generator uses a different kind of feedback to ensure that the current which charges the timing capaci-

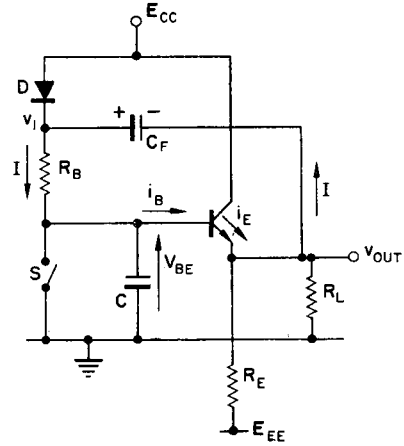


FIG. 3.7. The bootstrap linear sweep generator. (Typical values: R_B = 10 kΩ, R_E = 4·7 kΩ, $E_{CC} = -E_{EE}$ = 10 V, $C_F \approx 100 \times C$.)

tor C (Fig. 3.7) is held constant: when switch S is opened, capacitor C begins to be charged by the current I. The rising voltage across C raises v_{OUT} too, via the emitter-follower transistor. But v_{OUT} is coupled back to v_1 via C_F, which is a large capaci-

tor, so that v_1 also rises with v_{OUT}. Thus the voltage drop across R_B, which determines I, is held constant.

Evidently v_1 is initially nearly equal to E_{CC}, but as soon as v_{OUT} begins to rise diode D is cut-off and the whole of I is obtained through C_F.

The equations describing this circuit are as follows:

If to begin with it is assumed that C_F is sufficiently large for no voltage change across it to occur, then

$$v_1 - IR_B - V_{BE} = v_{OUT} \tag{3.27}$$

and

$$v_1 = E_{CC} - V_{D(ON)} + v_{OUT}, \tag{3.28}$$

where $V_{D(ON)}$ is the forward voltage drop across the diode when it is conducting.

Equating currents at the base and emitter nodes

$$C \frac{d}{dt}(V_{BE} + v_{OUT}) = I - i_B, \tag{3.29}$$

$$i_E = \frac{v_{OUT} - E_{EE}}{R_E} + I + \frac{v_{OUT}}{R_L}. \tag{3.30}$$

If V_{BE} is assumed constant and if $V_{BE} \approx V_{D(ON)}$, then eqns. (3.27) and (3.28) yield

$$I = (E_{CC} - 2V_{BE})/R_B. \tag{3.31}$$

If the rate of rise of v_{OUT} is not too rapid $i_E = i_B(\beta + 1)$ and eqns. (3.29), (3.30), (3.31) reduce to

$$v_{OUT} = A(1 - e^{-t/B}), \tag{3.32}$$

where, neglecting V_{BE} in eqn. (3.31),

$$A = \left(E_{CC} \frac{\beta}{R_B} - \frac{E_{EE}}{R_E}\right) \frac{R_L R_E}{R_L + R_E},$$

$$B = (\beta + 1)C \frac{R_L R_E}{R_L + R_E},$$

which are closely analogous to the equations for the Miller integrator.

A simpler understanding of the circuit can be gained by just considering eqn. (3.29) from which it follows, since $V_{BE} \approx$ const.

$$\frac{dv_{OUT}}{dt} = \frac{1}{C}(I - i_B). \tag{3.33}$$

As v_{OUT} rises the transistor draws more of the charging current away from the capacitor, thus decreasing the rate of increase of v_{OUT}.

In practice the charging current does not remain quite constant since the voltage across C_F changes. If t_r is the time for which S remains open, the change of voltage across C_F is It_r/C_F. If this change is much less than E_{CC}, then

$$I_{final} \approx \frac{E_{CC} - 2V_{BE}}{R_B + t_r/C_F}. \tag{3.34}$$

This value of I can be substituted in eqn. (3.33) to obtain the final value of dv_{OUT}/dt.

When switch S closes, the timing capacitor C discharges rapidly bringing v_{OUT} and v_1 with it until the diode becomes forward biased again. When this occurs v_{OUT} will not have reached its initial value owing to the change of voltage across C_F. Indeed the difference between the initial and quenched output voltages is simply It_r/C_F again, and the transistor will be temporarily cut-off. C_F then recharges primarily through R_E (the current in R_L being very small if $It_r/C_F \ll E_{CC}$). Thus the recharge time, until the transistor is again conducting, is approximately

$$t_{recharge} = \frac{It_r}{I_{EE}} \approx \frac{It_r R_E}{E_{EE}}. \tag{3.35}$$

This is not normally negligible compared with t_r. Thus the bootstrap generator usually includes two transistor switches,[10] one at S and one to quench the output.

High-speed Working

When the linear rise or fall of output voltage of a sweep generator is not sufficiently slow for the reactive effects in the transistor to be ignored, some account must be taken of them. In practice this simply involves substituting eqns. (1.21) and (1.22) for $i_B = i_C/\beta = i_E/(\beta+1)$ in eqns. (3.29) and (3.30). If dq_B/dt is still small compared with q_B/τ_C (but not small compared with $q_B/\beta\tau_C$) it is easy to show that in the bootstrap circuit the consequence is an increase in the effective value of C to

$$C_{(\text{effective})} = C+\tau_C\left(1+\frac{\Delta Q_{VC}}{\Delta Q_B}\right)\left(\frac{1}{R_E}+\frac{1}{R_L}\right)\frac{\beta}{\beta+1}. \quad (3.36)$$

At high rates of change of output voltage dq_B/dt cannot be neglected in eqn. (1.22), producing a second-order differential equation and greater non-linearity in the solution.

Example. If in Fig. 3.7 $E_{CC} = -E_{EE} = 10$ V, $R_B = 5$ kΩ, $R_E = 4$ kΩ, $C = 0{\cdot}2$ μF, $C_F = 5$ μF, $R_L = 2$ kΩ and in the transistor $\beta = 30$. Estimate the magnitude of v_{OUT} 0·5 msec after switch S opens. What is the effect of replacing the transistor by one with a very large β? (Assume $V_{BE} = V_{D(ON)} = 0{\cdot}7$V.)

(1) The initial rate of increase of v_{OUT} is given by eqns. (3.33) and (3.31).

Since, by eqn. (3.31), $I = 1{\cdot}72$ mA

$$\frac{dv_{OUT}}{dt} = 7{\cdot}95 \text{ V/msec} \quad \text{if} \quad \beta = 30.$$

(2) Thus the final output voltage would be about 4·0 V for $\beta = 30$ assuming no decrease in dv_{OUT}/dt.

(3) A current of 1·72 mA flowing through C_F for 0·5 msec causes a voltage change across it. Thus the final value of I is by eqn. (3.34)

$$I_{(\text{final})} = 1{\cdot}69 \text{ mA}.$$

(4) Using these approximate values of the final output voltage and charging current in eqn. (3.33) yields the final value of dv_{OUT}/dt.

$$\frac{dv_{OUT}}{dt} = 7{\cdot}29 \text{ V/msec.}$$

(5) The average rate of increase of output voltage is consequently about 7·6 V/msec and

$$v_{OUT} \approx 3{\cdot}8 \text{ volts after } 0{\cdot}5 \text{ msec.}$$

If $\beta = \infty$ the only non-linearity arises from the charging of C_F. Thus in this case

$$v_{OUT} = 4{\cdot}25 \text{ V.}$$

Level Detectors

Level detectors, which give a step function output whenever the input crosses and recrosses certain particular voltage levels, are widely used for the generation of digital waveforms. Such circuits are also sometimes called *zero crossing detectors* or clipping circuits.

The long-tailed pair has already been briefly considered as a clipping circuit, but its input voltage range of about 230 mV within which it is an amplifier is too wide for many clipping purposes. There are two ways of obtaining better performance. One is to increase the gain of the circuit so that the 230 mV range is reduced as low as required. The second is to include feedback in the circuit so that once the circuit has begun to switch it will complete the change over regeneratively. The first method can give more rapid response, the second is essential for slowly varying input waveforms.

1. *A High-gain Clipping Circuit*

The sensitivity of a long-tailed pair can be greatly improved by the addition of transistor T_3 as shown in Fig. 3.8. The extra d.c. level at the emitter of T_3, produced by the Zener diode — biased by R_D — is very important. V_Z and R_{L2} should be such that when the collector current through T_2 is $I_{EE}/2$, then T_3 is on the edge of conduction.

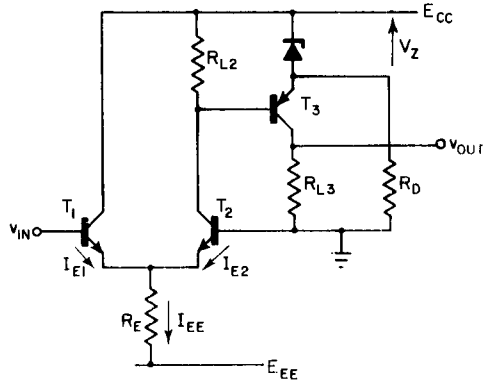


FIG. 3.8. A high-gain clipping circuit using a long-tailed pair. (Typical values: $E_{CC} = -E_{EE} = 10$ V, $R_E = R_{L2} = 4{\cdot}7$ kΩ, $R_{L3} = 1$ kΩ, $V_Z = 5$ V, $R_D = 2{\cdot}2$ kΩ.)

Suppose T_2 is being turned on and T_1 turned off by the falling input voltage v_{IN}. As explained on p. 43 T_1 will pass from the near cut-off state ($I_{E1} = 0{\cdot}005 I_{EE}$) to the half-on state when $I_{E1} = I_{E2}$ with an input voltage change of about 120 mV. During this change T_3 will remain cut-off. T_3 will, however, switch from the near cut-off state to full conduction when its base–emitter voltage V_{BE3} also changes by about 120 mV and when sufficient current is supplied to the base of T_3 to bottom it. Thus to switch T_3 fully the collector current of T_2 must increase beyond $I_{EE}/2$

by an amount ΔI_{C2} given by

$$\Delta I_{C2} \approx (\Delta V_Z + 0{\cdot}120)\left[\frac{1}{R_{L2}} + \frac{1}{r_{02}}\right] + \frac{E_{CC} - V_Z}{\beta R_{L3}}, \quad (3.37)$$

where r_{02} is the output resistance of T_2 (usually unimportant), where β is the current gain of T_3, and where ΔV_Z is the change in V_Z as the current through the Zener diode increases as T_3 conducts.

The change Δv_{IN} in the input voltage necessary to achieve this change in I_{C2} is given by

$$\Delta v_{IN} = \frac{kT}{q} \ln \frac{I_{E1} - \Delta I_{C2}}{I_{E2} + \Delta I_{C2}}, \quad (3.38)$$

where

$$I_{E1} = I_{E2} = I_{EE}/2.$$

The exact value of Δv_{IN} depends upon operating conditions, but there is normally no difficulty in reducing Δv_{IN} to 5 mV or less for a 5 V output voltage swing, giving a large signal voltage gain, A_V, of 1000 or more.

The rise time of the output voltage will normally be the time it takes for the input voltage to change by Δv_{IN} (having already changed by about 120 mV to bring I_{E1} to $I_{EE}/2$). If this change is sufficiently fast to involve a transient change in the base charge of T_3 which is not negligible compared with the last term in eqn. (3.37), then a larger base current must be supplied to T_3 to produce rapid switching. If $I_{B(ON)}$ is this larger base current supplied to T_3, the rise time of the output voltage can be found from eqn. (1.35). The value of Δv_{IN} to produce $I_{B(ON)}$ can be found from eqns. (3.37) and (3.40) by replacing $(E_{CC} - V_Z)/(\beta R_{L3})$ in eqn. (3.37) by $I_{B(ON)}$.

For example, in Fig. 3.8 let $R_{L3} = 500\ \Omega$, $E_{CC} = -E_{EE} = 10$ V, $R_{L2} = R_E = 1\ \text{k}\Omega$, and $V_Z = 5$ V. Assume that ΔV_Z, as T_3 turns on, is negligible.

If the turn-on is to be fast, with t_r of the order of $5\tau_C(1 + Q_{VC}/Q_B)$ (see p. 23), then $I_{B(ON)} \approx I_{C(ON)3}/5 \approx 2{\cdot}0$ mA.

Whence by eqn. (3.37) $\Delta I_{C2} = 2{\cdot}0 + 0{\cdot}120 = 2{\cdot}120$ mA. But $I_{EE} = 10$ mA, so

$$\Delta v_{IN} = \frac{kT}{q} \ln\left(\frac{7{\cdot}12}{2{\cdot}88}\right) \approx 23 \text{ mV}.$$

If base overdrive of T_3 were not needed, and a minimum β of T_3 of 30 can be assumed, $I_{B(ON)} = (E_{CC} - V_Z)/\beta R_{L3} = 0{\cdot}33$ mA. Then $\Delta I_{C2} \approx 0{\cdot}45$ mA and $\Delta v_{IN} \approx 4{\cdot}7$ mV. Thus, as is to be expected, the gain of the circuit falls at high switching speeds.

Two properly biased, cascaded stages of long-tailed pairs (one *npn* and the other *pnp*) will, of course, perform in much the same way as the circuit of Fig. 3.8.

2. *The Schmitt Trigger Circuit*

The Schmitt trigger circuit of Fig. 3.9 is again essentially a long-tailed pair (although the "tail" is not a very high resistance in this case) together with additional coupling from one transistor to the other, in the form of R_B—sometimes in parallel with C_B.

The circuit has two stable states, one with T_1 cut-off, the other with T_2 cut-off, and it switches rapidly from one state to the

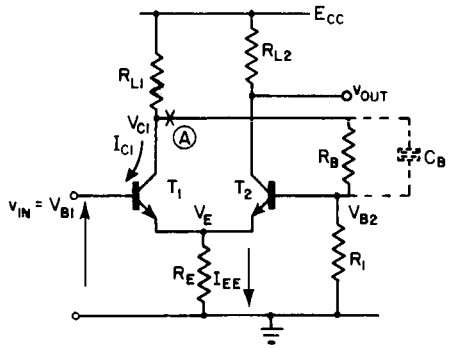


FIG 3.9. The Schmitt trigger circuit.

other when the input *voltage* V_{B1} reaches the appropriate trigger level. The larger the gain round the loop (forward via R_B and back again across R_E), the faster it switches.

The circuit can be designed so that the output transistor saturates or so that it does not. There are important differences between these two cases and they will be considered in turn. The input transistor can also be allowed to bottom but the difference in this case is less significant. C_B will not be included.

With non-saturating output transistor. The input voltages at which the circuit switches depend on the steady state values of V_{B2}. When T_2 is conducting, raising V_{B1} until it is about equal to V_{B2} turns T_1 on so that T_2 is switched off regeneratively. When T_2 is cut-off V_{B2} is lower than when it is ON, and lowering V_{B1} until it is again about equal to V_{B2} turns T_1 off again, so that T_2 comes on regeneratively.

Thus two parameters must be calculated in order to determine the triggering levels.

(1) The steady state values of V_{B2} when T_2 is (a) conducting and (b) cut-off.

(2) The difference between V_{B1} and V_{B2} when regeneration starts and the circuit rapidly changes state.

We will consider the second problem first.

Regeneration will only commence provided the loop gain A_V (determined by breaking the circuit at Ⓐ and by measuring the signal voltage ratio at the terminals exposed) is greater than 1. It was shown in Chapter 2 that A_V has an upper limit, of not greater than

$$A_V \approx \frac{R_{L1}}{r_{e1}+r_{e2}}, \tag{3.39}$$

where

$$r_{e1} = \frac{kT}{qI_{E1}} \quad \text{and} \quad r_{e2} = \frac{kT}{qI_{E2}}.$$

r_{e1} and r_{e2} are the emitter junction slope resistances.

Since $I_{E1}+I_{E2} = I_{EE}$ and I_{EE} is nearly a constant it follows that A_V has a maximum when $I_{E1} = I_{E2}$ and that $A_V = 0$ when either I_{E1} or $I_{E2} = 0$.

Suppose we design the circuit so that regeneration begins when the oncoming transistor takes $2\frac{1}{2}\%\ I_{EE}$. Then, as explained on p. 43,

$$V_{B1} - V_{B2} = \frac{kT}{q} \ln \left[\frac{1+19/20}{1-19/20}\right] \approx 100 \text{ mV}. \qquad (3.40)$$

This current change is sufficient to initiate regeneration but does not alter the steady state voltages significantly.

Thus when T_1 is being turned on

$$V_{B1(\text{ON})} = V_{B2(\text{ON})} - 100 \text{ mV} \qquad (3.41)$$

and as V_{B1} is being lowered T_1 turns off when

$$V_{B1(\text{OFF})} = V_{B2(\text{OFF})} + 100 \text{ mV}. \qquad (3.42)$$

In other words, the difference between $V_{B1(\text{ON})}$ and $V_{B1(\text{OFF})}$, called the *hysteresis*, is about 200 mV less than the difference between the two steady state values of V_{B2}.

The steady state values of V_{B2} are evidently the following, neglecting cut-off currents:

(a) when T_2 is conducting, then, by superposition,

$$V_{B2(\text{ON})} = \frac{R_1 E_{CC}}{R_1+R_B+R_{L1}} - I_B \cdot \frac{R_1(R_B+R_{L1})}{R_1+R_B+R_{L1}}, \qquad (3.43)$$

where $I_B = I_{EE}/(\beta+1)$ and I_{EE} depends on $V_{B2(\text{ON})}$,

(b) when T_1 is conducting

$$V_{B2(\text{OFF})} = \frac{(E_{CC} - I_{C1(\text{ON})}R_{L1})R_1}{R_1+R_B+R_{L1}}, \qquad (3.44)$$

where, just before T_1 turns off,

$$I_{C1(\text{ON})} \approx I_{EE} = (V_{B1(\text{OFF})} - V_{BE1(\text{ON})})/R_E. \qquad (3.45)$$

Combining eqns. (3.42), (3.44) and (3.45) leads to an explicit but cumbersome expression for $V_{B1(\text{OFF})}$, the voltage at which the circuit switches regeneratively. Thus for given values of

R_1, R_B, R_{L1} and R_E it is possible to calculate the voltage levels at which the circuit will switch, and hence the hysteresis.

The *design* problem is, of course, the converse, and the remaining factor of importance, the effect of current gain, will be considered through an example.

Example. *Design a Schmitt trigger to operate from a +10 V supply, to switch at a rising input voltage of 5±0·25 V and to switch a current I_{EE} of about 5 mA.*

We will only take transistor parameter variations (β and V_{BE}) into account in determining the tolerance of $V_{B1(ON)}$. Silicon transistors will be used.

(1) If the limits of variation of β in the transistors used are 25 to 200 then to ensure that $V_{B1(ON)}$ does not vary by more than 0·5 V the variation in $V_{B2(ON)}$ should be kept within about 0·4 V to allow for variations in V_{BE} in the two transistors. Thus eqn. (3.45) leads to the following inequality:

$$0{\cdot}4 \geqslant \frac{I_{EE}R_1(R_B+R_{L1})}{R_1+R_B+R_{L1}}\left(\frac{1}{26}-\frac{1}{201}\right)$$

or

$$\frac{R_1(R_B+R_{L1})}{R_1+R_B+R_{L1}} \leqslant 2{\cdot}4\ \text{k}\Omega.$$

Let

$$R_1(R_B+R_{L1})/(R_1+R_B+R_{L1}) = 2{\cdot}4\ \text{k}\Omega. \qquad (3.46)$$

(2) If the loop gain of the circuit is to exceed unity when I_{E1} or $I_{E2} = I_{EE}/40$, then by eqn. (3.39)

$$I_{EE}R_{L1} > 41\ kT/q$$

or

$$R_{L1} > 220\ \Omega \text{ since } kT/q \approx 25\ \text{mV}.$$

Let $R_{L1} = R_{L2} = 400\ \Omega$, giving a 2 V output.

(3) Dividing eqn. (3.43) by eqn. (3.46) and taking the mid-value of $1/(\beta+1)$, namely 47, yields:

$$\frac{E_{CC}}{R_B+R_{L1}} - \frac{I_{EE}}{47} = \frac{V_{B2(ON)}}{2{\cdot}4\ \text{k}\Omega} = 2{\cdot}1\ \text{mA},$$

whence $R_B + R_{L1} = 4{\cdot}5\ \text{k}\Omega$

and $R_1 = 5{\cdot}1\ \text{k}\Omega$ by eqn. (3.46).

(4) With $I_{EE} = 5$ mA when $V_{B1(\text{ON})} = 5$ V, then for silicon transistors in which $V_{BE(\text{ON})} \approx 0{\cdot}7$ V.

$$R_E = \frac{V_{B1(\text{ON})} - V_{BE(\text{ON})}}{I_{EE}} \approx 860\ \Omega.$$

(5) From eqns. (3.42), (3.44) and (3.45) the above values yield:

$$V_{B1(\text{OFF})} = 4{\cdot}5\ \text{V}$$

so that since $V_{B1(\text{ON})} = 5 \pm 0{\cdot}25$ V the hysteresis is approximately $0{\cdot}5 \pm 0{\cdot}25$ V.

The simplest way of altering the hysteresis is to vary the value of R_{L1}. The larger R_{L1} the larger the hysteresis, as shown by eqn. (3.44). If R_{L1} increases sufficiently to bottom T_1 eqn. (3.44) still applies but now $(E_{CC} - R_{L1}I_{C1}) = V_{E1} \approx V_{B1}$ and when V_{B1} is reduced I_{C1} *increases*. Indeed since by eqn. (3.44), $V_{B2(\text{OFF})}$ must always be less than $(E_{CC} - R_{L1}I_{C1})$ and since the circuit only switches when $V_{B1} \approx V_{B2}$ it follows that the circuit cannot switch until T_1 comes out of bottoming sufficiently to allow $(E_{CC} - I_{C1}R_{L1})$ to be greater than V_{B1}. Thus the hysteresis can be increased very considerably by increasing R_{L1} so that T_1 bottoms.

The penalty for allowing T_1 to saturate is that the input resistance of the circuit when T_1 is conducting drops from about βR_E to R_E because, of course, the current gain of the input transistor is lost when it is bottomed.

The Schmitt Trigger with Saturated Output Transistor

If the output transistor T_2 is driven into saturation, the principles of operation are unaltered, but the details change considerably.

With T_1 conducting eqns. (3.42), (3.44) and (3.45) again apply.

With T_2 conducting and bottomed, $V_{CE(sat)} = 0$ and $V_{B2} = V_{BE(ON)} + V_E$ so that equating currents into and out of the transistor

$$\frac{E_{CC} - V_E}{R_{L2}} + \frac{E_{CC} - V_E - V_{BE(ON)}}{R_{L1} + R_B} = \frac{V_E}{R_E} + \frac{V_E + V_{BE(ON)}}{R_1} \quad (3.47)$$

from which V_E can be found.

T_1 begins to conduct when V_{B1} rises to approximately V_{B2}, as discussed previously, and if T_2 was only just bottomed this would also be the voltage at which the circuit switches. Normally, however, T_2 is not only just bottomed and V_{B1} must rise further to start regeneration.

A better approximation assumes that β_2, the current gain of T_2, is large, and uses the fact that T_2 will cease to be bottomed when $I_{B2} = 0$. Thus we can write

$$V_{C1} = I_1(R_1 + R_B) = E_{CC} - R_{L1} I_{L1}$$

where

$$I_1 = \frac{V_E + V_{BE(ON)}}{R_1}$$

and where I_{L1} is the current in R_{L1} when T_2 is *just* bottomed. If β_1 is also large

$$I_{L1} = I_1 + \frac{V_E}{R_E} - \frac{E_{CC} - V_E}{R_{L2}}.$$

Now V_E can be found, and $V_{B1(ON)}$ is about 0·7 V larger, using silicon transistors.

If β_2 is finite the equations become very involved. To cause the circuit to switch it is necessary to turn on T_1 sufficiently to reduce I_{B2} until it is less than I_{C2}/β_2 (i.e. until T_2 is in the Active Region of Operation). The value of I_{C1} needed can be found as follows.

So long as T_2 remains bottomed its three terminals can (almost) be regarded as shorted together (by neglecting $V_{BE(sat)}$ and $V_{CE(sat)}$) so that as far as *changes* in current are concerned R_E, R_{L2} and R_1 are in parallel, as indicated in the equivalent circuit of Fig. 3.10(a). We will call their parallel resistance R_X.

Thus the incremental equivalent circuit reduces to that of Fig. 3.10(b).

Using the prefix Δ to denote changes (not necessarily small) in I_{B2}, I_{C1}, V_{C1} and V_E we obtain:

$$\begin{aligned} \Delta V_{C1} &= -R_{L1}(\Delta I_{C1} - \Delta I_{B2}), \\ \Delta V_E &= R_X(\Delta I_{C1} - \Delta I_{B2}) \end{aligned} \tag{3·48}$$

and $-\Delta I_{B2} R_B = \Delta V_{C1} - \Delta V_E$, so that

$$\Delta I_{C1} = \Delta I_{B2} \frac{R_B + R_X + R_{L1}}{R_X + R_{L1}}. \tag{3.49}$$

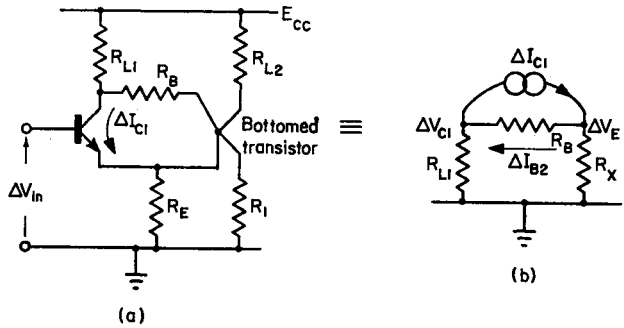


FIG. 3.10. The Schmitt trigger with transistor T_2 bottomed, together with its equivalent circuit (see text).

The required change in I_{B2} to drive T_2 into the active region is

$$\Delta I_{B2} = I_{B2(ON)} - I_{C2(ON)}/\beta_2. \tag{3.50}$$

Equations (3.49) and (3.50) are sufficient to determine the change in I_{C1} needed. The important quantity, however, is the value of $V_{B1(ON)}$ at which this current flows. This is

$$V_{B1(ON)} = (I_{C2} + I_{B2})R_E + \Delta V_E + V_{BE(ON)}. \tag{3.51}$$

Thus the parameters which determine the value of $V_{B1(ON)}$ when T_2 is saturated are quite different from those when T_2 is not saturated. These are inconvenient analytic equations so that *design* is most easily accomplished by choosing sensible values and calculating the consequences, as follows:

Example. *Design a Schmitt trigger circuit, with saturating output transistor, which will trigger to a rising input voltage at* $5 \pm 0{\cdot}25$ *V when* $E_{CC} = 10$ *V and* I_{EE} *is about 5 mA.*

After some brief trials to obtain approximate values, particularly of ΔV_E, the steps in the design are now

(1) Choose $V_E = 3{\cdot}7$ V, so that T_1 will turn on at 5 V, allowing for $V_{BE(ON)} = 0{\cdot}7$ V and ΔV_E in eqns. (3.48) and (3.51) = 0·6 V.

(2) Calculate $R_E = V_E/I_{EE} = 740\ \Omega$.

(3) Choose I_{C2} and I_{B2} so that $I_{C2}+I_{B2} = I_{EE}$ and so that $I_{C2}/I_{B2} < \beta_{\min}$.

Let $I_{C2} = 4{\cdot}7$ mA, $I_{B2} = 0{\cdot}3$ mA.

(4) Calculate $R_{L2} \approx (E_{CC}-V_E)/I_{C2} = 1{\cdot}34$ kΩ (assuming $V_{CE(sat)2} \approx 0$).

(5) Choose $I_1 = I_{B2}$. This choice is somewhat arbitrary. The consequences of alternative choices are easily explored. I_1 must not be so large as to destroy the loop gain, yet not too small to prevent T_2 being properly cut-off.

Thus choose $I_1 = 0{\cdot}3$ mA.

(6) Since T_2 is bottomed, $V_{B2} \approx V_E+0{\cdot}8$ V $\approx 4{\cdot}5$ V.
Allowing $V_{BE(ON)} = 0{\cdot}8$ V for a bottomed silicon transistor.

(7) Calculate $R_1 = V_{B2}/I_1 = 15$ kΩ.

(8) Choose R_{L1} to prevent T_1 bottoming but to provide ample loop gain.

Let $R_{L1} = 700\ \Omega$.

(9) Calculate R_B. $R_{L1}+R_B = (E_{CC}-V_{B2})/(I_1+I_{B2})$.

Hence $R_B = 8.5$ kΩ.

The range of values of V_{B1} at which T_1 turns on is now obtained as follows:

Since $I_{B2} = 0{\cdot}3$ mA, the change in I_{B2}, ΔI_{B2}, necessary to bring T_2 into the Active Region of operation is given by eqn. (3.50) for $25 < \beta < 200$.

$$\begin{aligned}\Delta I_{B2} &= 0{\cdot}3 - I_{C2}/25 = 0{\cdot}112 \text{ mA (min.)}\\ &= 0{\cdot}3 - I_{C2}/200 = 0{\cdot}276 \text{ mA (max.)}.\end{aligned}$$

Now $1/R_X = 1/R_{L2}+1/R_1+1/R_E$, so that $R_X = 460\ \Omega$, and by eqn. (3.49)

$$\Delta I_{C1} = 7{\cdot}65\ \Delta I_{B2}.$$

Hence by eqn. (3.48) ΔV_E is between 0·34 V and 0·84 V. A mean of 0·6 V was assumed initially.

From eqn. (3.51)

$$V_{B1(ON)(min)} = 4{\cdot}74\ \text{V},$$
$$V_{B1(ON)(max)} = 5{\cdot}24\ \text{V}.$$

These values are at the limits of the specification. The spread on $V_{B1(ON)}$ can be greatly reduced by increasing I_{B2} and the current through R_1 (i.e. by reducing R_B and R_1).

The value of $V_{B1(OFF)}$ is obtained from eqns. (3.42) and (3.44) and (3.45).

$$V_{B1(OFF)} = 4{\cdot}25\ \text{V}.$$

$V_{B1(ON)}$ can be reduced without greatly affecting $V_{B1(OFF)}$ by including an extra resistor in the emitter lead of T_2. This resistor also reduces the loop gain and therefore retards (or may even prevent) the regenerative switching. It must not therefore be too large.

The advantages of the saturated output circuit are:

(1) The output voltage is larger and is not dependent on the transistor current gain.

(2) It can be designed with smaller variation of the triggering levels, although it is perhaps more difficult to design.

In either circuit the inclusion of C_B, to improve the loop gain at the transitions, affects the trigger levels with varying input voltages.

Problems

3.1. In the multivibrator circuit of Fig. 3.1 (a) $R_L = 2{\cdot}2\ \text{k}\Omega$ and $R_B = 47\ \text{k}\Omega$ and $C_B = 0{\cdot}001\ \mu\text{F}$ and $E_{CC} = E_{BB} = 10$ V. Calculate the repetition frequency of the oscillation assuming:

(i) that $V_{BE(ON)} = 0$ and that $I_{BX} = 0$ and that $Q_{OFF} = 0$ (15·3 kc/s);
(ii) that $V_{BE(ON)} = 0{\cdot}7$ V but $I_{BX} = Q_{OFF} = 0$ (14·55 kc/s);
(iii) that $I_{BX} = 10\ \mu\text{A}$ but $V_{BE(ON)} = Q_{OFF} = 0$ (15·85 kc/s);
(iv) that $Q_{OFF} = 500$ picocoulombs but $V_{BE(ON)} = I_{BX} = 0$ (15·9 kc/s);
(v) that $V_{BE(ON)} = 0{\cdot}7$, $I_{BX} = Q_{OFF} = 0$ but that $E_{CC} = E_{BB} = 9$ V (14·46 kc/s).

3.2. To the circuit specified in question 3.1 add a resistor of 47 kΩ between base and collector of T_1 and another between base and collector of T_2. What effect does this have upon the frequency calculated in 3.1(i)?

(Ans.: 33 kc/s approx.)

3.3. In the Schmitt trigger circuit of Fig. 3.9, $E_{CC} = 12$ V and $E_{EE} = 0$, $R_{L1} = R_{L2} = 1\ \text{k}\Omega$, $R_B = 5\ \text{k}\Omega$, $R_E = 4\ \text{k}\Omega$ and $R_1 = 12\ \text{k}\Omega$. If the current gain β of each transistor is very large, what is V_{B1} at which T_1 (a) turns-on and (b) turns-off? (Assume $V_{BE(ON)}$ of each transistor is zero and that cut-off currents are negligible.)

(Ans.: 8 V, 6·86 V.)

If the current gains of the transistors were both 20 instead of ∞, what would be the values of V_{B1} at turn-on and turn-off?

(Ans.: 7·64 V, 6·91 V.)

3.4. R_{L2} in the circuit of question 3.3 is changed to 3 kΩ so that T_2 is bottomed when it is conducting. Determine the values of V_{B1} at which the circuit switches assuming that for each transistor $\beta = \infty$, and that $V_{BE(ON)} = 0{\cdot}7$ V.

(Ans.: 7.93 V, 6·96 V.)

3.5. Design a Schmitt trigger circuit with non-saturating transistors to switch at a rising input voltage of zero volts and a falling voltage of -1 V, when the supplies are ± 10 V. Design for a current in R_E of about 5 mA. Let β of each transistor be 50 and V_{BE} of each transistor be 0·7 V when it is fully conducting.

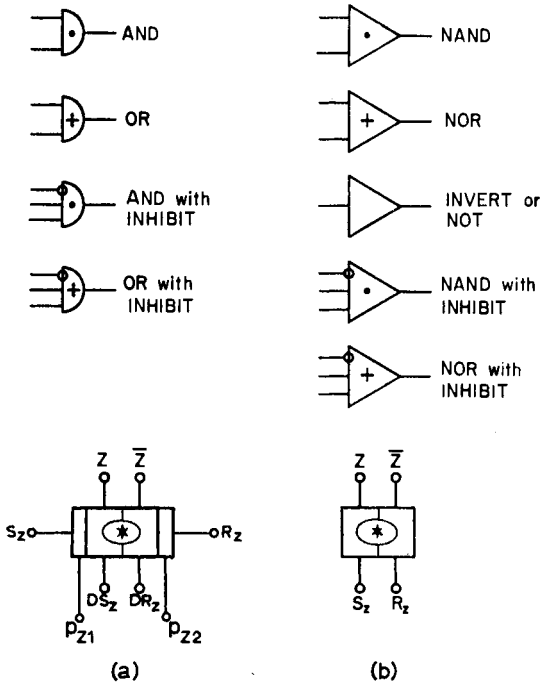
If β of each transistor falls to 25 what are the revised values of input voltage at which the circuit will switch?

3.6. In the bootstrap sweep circuit of Fig. 3.7 $R_B = 3\ \text{k}\Omega$, $R_E = 2\ \text{k}\Omega$, $R_L = 1\ \text{k}\Omega$, $C_B = 0{\cdot}1\ \mu\text{F}$, $C_F = 5\ \mu\text{F}$, $E_{CC} = 12$ V, $E_{EE} = -6$ V.

Estimate the value of V_{OUT} 0·2 msec after the switch S has been opened. Assume $V_{BE} = 0{\cdot}7$ V and the current gain β of the transistor is 30. How long after switch S is closed again should be allowed for C_F to recharge before opening the switch again?

(Ans.: 6·25 V, 0·22 m sec. approx)

Graphic Symbols



Graphical symbols. Diagrams (a) and (b) show respectively pulsed and d.c. flip-flops. Logic inputs (S_Z, R_Z), direct set and reset inputs (DS_Z, DR_Z) and the pulse or trigger inputs (p_{Z1}, p_{Z2}) are also shown. Normally the asterisk in a flip-flop diagram will be replaced by the name of the type of flip-flop involved, namely *R–S*, *J–K*, *E*, *T*, *D*, *R–S–T*, etc.

Combinational & Sequential Circuits

THE remainder of this book is primarily concerned with circuits capable of performing logical or arithmetic functions, that is, circuits which can count pulses, circuits which detect coincidences or anti-coincidences between pulses, and so on. Such circuits have very wide uses in computers and telephone exchanges, in the automatic control of machinery and in the handling of data.

Circuits of this kind are assemblies of functional blocks, each of which is capable of performing some logical function. The problem in designing such circuits consists of two parts, the design of the functional blocks and the design of the appropriate interconnections. Nowadays many functional blocks are available in integrated circuit form, which leaves only the problem of interconnecting them to the system designer. However, it is true today, and will probably always be true, that the interconnections can be handled with confidence only if the circuits within the functional blocks are understood. Consequently both aspects of the problem are considered.

From time to time in the following pages comments are made as to which circuits can be easily constructed in integrated form. Here the type of integrated circuit being considered is the solid silicon circuit in which all the electrical components are formed by diffusing impurities into a silicon chip.[11] The kind of limitations which these circuits have are that, obviously, germanium components cannot be included, that silicon areas of differing carrier lifetimes cannot easily be made, that capacitors are expensive, and that inductors of above about 1 μH are almost impossible.

In this book there is only room to survey the possible types of functional blocks and their characteristics, and then to present a convenient method of designing or organizing their interconnections to form larger digital circuits. There are so many ways in which this organization can be performed for any particular circuit function that some method of choosing, say, the most economical circuit may be desirable. A rigorous method of *minimization*, as it is called, is, however, beyond the scope of this book, and is, in any case, still a matter for development. The method given is convenient and usually efficient but not always minimal. Some examples of minimization are included.

It is convenient to separate these logical circuits into two types, namely combinational and sequential circuits.

A combinational circuit is one whose output at any given time is a function only of its inputs at that time. Gates are typical examples.

A sequential circuit is one whose output at any given time is a function not only of inputs at that time, but also of previous inputs which have since been removed. A sequential circuit usually contains combinational circuits. Counters, control circuits, etc., are examples of sequential circuits. Since sequential circuits take note of inputs in the past, they must contain some form of *memory*; that is to say, the information imparted by the input must remain after the input is removed.

The problems of the design of both types of circuit are discussed in the next few chapters. But first the mathematical procedure used in the description of these circuits, namely Boolean algebra, will be briefly described. Only the most elementary aspects of the algebra are required in this book.

CHAPTER 4

Combinational Circuits

Boolean Algebra

Boolean Algebra is an algebra in which the variables can have one of only two possible values. The two values can stand for the truth or falsehood of a statement, the states of a switch (open or closed), the presence of one of two voltage levels, etc. The two values of a variable A, say, are usually written 1 or 0. Since A can have only two values, the *complement* of A, namely "not A" or $\bar{A}$, is also an important variable. If $A=1$, $\bar{A}=0$, or if $A=0$, $\bar{A}=1$.

In this book the variables refer to voltage levels and any Boolean equations we consider relate the voltage of one node or terminal to the voltages of other nodes. Thus $Z=A$ means that nodes Z and A are at the same voltage; both 6 V, say, or both zero volts. In Boolean notation, if we call one voltage level 1 and the other 0, then $Z=A$ means that if $Z=1$, $A=1$ and that if $Z=0$, $A=0$. Similarly $Z=\bar{A}$ means that when the voltage of Z is 6 V, the voltage of A is zero, and vice versa. Or in other words, if $Z=1$, $A=0$ and if $Z=0$, $A=1$.

There are three basic functions of primary importance in Boolean Algebra:

(a) *Complementation* (which has already been referred to). $\bar{A}$ is the *complement* of A, and vice versa. This is sometimes referred to as the NOT function.

(b) *Addition* An equation such as $Z = A+B$ means that $Z=1$

if either A OR B is equal to 1, or if both are. This is the basis of the "OR" function in logical circuits.

(c) *Multiplication.* An equation such as $Z=A.B$ means that Z is 1 only if both A AND B are 1. This is the basis of the AND function.

With these definitions it is now possible to proceed with the development of the algebra, beginning with Boolean arithmetic.

Rules of Boolean Arithmetic

$1+1 = 1$	$1.1 = 1$
$1+0 = 1$	$1.0 = 0$
$0+1 = 1$	$0.1 = 0$
$0+0 = 0$	$0.0 = 0$

With the exception of the first addition, these rules are identical with those of normal arithmetic. This close similarity is the justification for using addition and multiplication signs for what are actually quite different operations.

From these arithmetic rules we can begin the algebra, in which variables X, Y, Z can have one of only two values. The truth of these rules can be verified by simply substituting in turn the two possible values for each variable. Thus, below, in eqn. (1a) the two possible forms of the left-hand side are $0+1$ or $1+1$, both of which equal 1.

Rules of Boolean Algebra

(1a) $X+1 = 1$	(1b) $X.1 = X$
(2a) $X+X = X$	(2b) $X.X = X$
(3a) $X+0 = X$	(3b) $X.0 = 0$
(4a) $X+\overline{X} = 1$	(4b) $X.\overline{X} = 0$

In Boolean equations each occurrence of a variable, complemented or not, is called a *literal.* Thus in (4a) there are one variable, two literals and one value.

From these rules a number of useful theorems follow. (The full stop in the AND function will now be omitted.)

(5) $X+XY = X(1+Y) = X$ by (1a) and (1b).

(6) $X(X+Y) = XX+XY = X+XY = X$ by (2b) and (5).

(7) $X(\bar{X}+Y) = X\bar{X}+XY = XY$ by (4b) and (3a).

(8) $X+\bar{X}Y = X+XY+\bar{X}Y$ by (5).
$= X+Y(X+\bar{X})$
$= X+Y$ by (4a) and (1b).

(9) $(X+Y)(X+Z) = X(X+Y+Z)+YZ = X+YZ$ by (6).

(10) $(X+Y)(X+\bar{Y}) = X$ by (9) and (4b).

(11) $ZX+Z\bar{X}Y = ZX+ZY$ by (8).

(12) $XY+\bar{X}Z+YZ = XY+\bar{X}Z+YZ(X+\bar{X})$ by (4a).
$= Y(X+XZ)+\bar{X}(Z+YZ).$
$= XY+\bar{X}Z$ by (5).

(13) $(X+Z)(\bar{X}+Y) = XY+\bar{X}Z$ by (12).

The following two are known together as De Morgan's Theorem:

(14a) $\overline{X+Y+\ \ldots\ +Z} = \bar{X}\bar{Y}\ldots\bar{Z}.$

(14b) $\overline{XY\ldots Z} = \bar{X}+\bar{Y}+\ \ldots\ +\bar{Z}.$

These equations can be proved as follows: If all the variables X, Y, ..., Z are equal to 1 then both sides of theorem (14b) say, are clearly zero. If any one variable is changed to a zero, then both sides change to a value of 1. Any change in any other variable then makes no difference. Hence (14b) is always true. Similarly (14a) is always true.

In more general terms the theorem states that the complement of $f(A, B, C, \ldots,$ AND, OR) can be formed by

(i) changing all the AND's to OR's,
(ii) changing all the OR's to AND's,
(iii) changing all 1's to 0's,
(iv) changing all 0's to 1's,
(v) complementing each literal.

Thus if $$Z = A + A\bar{E} + FB(C + \bar{D}), \tag{4.1}$$
then $$\bar{Z} = \bar{A}(\bar{A} + E)(\bar{F} + \bar{B} + \bar{C}D). \tag{4.2}$$

With these theorems we can perform most of the operations needed for switching circuit design.

In this book the principal operations are those of simplification and rearrangement of rather complicated expressions. In principle this can always be done using the above theorems. In practice it is very difficult to achieve a minimal simplification without the use of some systematic procedure. The most general technique in use was worked out by McCluskey and refined by Quine[12] and can be used for any number of variables. Here we will usually consider only the case of up to four variables and make use of a matrix diagram known as the Karnaugh Map.[13] It will be seen that its use can be extended to deal with six variables by drawing three-dimensional maps, but not much further.

The Karnaugh Map

The method by which a Boolean expression can be simplified using a map is best explained by an example.

Suppose the output Z of a circuit is given by eqn. (4.3) (ignore the lower-case letters for the moment).

$$Z = \underset{a}{A(D + B)} + \underset{b}{\bar{B}D\bar{C}} + \underset{c}{A\bar{B}\bar{C}\bar{D}} + \underset{d}{\bar{A}\bar{B}CD} + \underset{e}{A\bar{B}\bar{D}}. \tag{4.3}$$

This expression can be simplified using the theorems just described or simply by expanding the expression in to its *canonical* form and regrouping. That is, each term is rewritten to contain all implied values of all the variables (e.g. since there are four variables, term (b) becomes $\bar{B}D\bar{C} = A\bar{B}D\bar{C} + \bar{A}\bar{B}D\bar{C}$). The resulting lengthy expression is then factorized.

If, however, the possible values of the variables are represented by the rows and columns of a map as in Fig. 4.1 the same process can be carried out, but is much simpler to accomplish. There

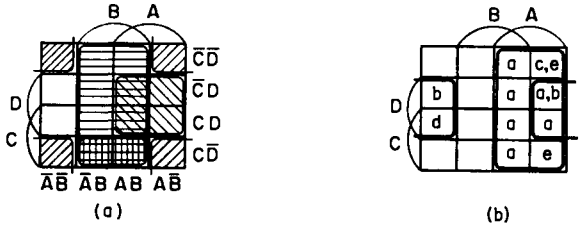


FIG. 4.1. The Karnaugh map. (a) Examples of element groupings or sets B, $B\bar{D}$, AD, $BC\bar{D}$; (b) the use of the map to simplify the Boolean expression of eqn. (4.3), namely $A\bar{B} + B(\bar{C} + \bar{A}) = \bar{A} + B\bar{C}$.

are sixteen ways in which the four two-valued variables can be combined, and each one is represented by an element in the matrix, or *map*. The columns and rows of the map are assigned in such a way as to ensure as tight a grouping of variables as possible. Thus the columns are labelled $\bar{A}\ \bar{B}$, $\bar{A}\ B$, $A\ B$, $A\ \bar{B}$, and the rows $\bar{C}\ \bar{D}$, $\bar{C}\ D$, $C\ D$, $C\ \bar{D}$. This same labelling is alternatively indicated by the brackets around the map. For example, the bottom right element is $A\bar{B}C\bar{D}$.

Elements can now be readily grouped in rectangular areas containing eight, four or two elements per group. Those marked in Fig. 4.1(a) are $\bar{B}\bar{D}$, AD, B, $BC\bar{D}$.

The terms of the Boolean eqn. (4.3) can now be plotted in a map, as shown in Fig. 4.1(b). The entries for each term in the

map are identified by the lower-case letters in the equation (for example, the term $A\bar{B}\bar{D}$ is indicated by e in the map).

Evidently these elements can be grouped in a much simpler expression, as indicated in the map. Thus

$$Z = A + D\bar{B}.$$

For problems containing more than four variables the grouping arrangements are not so simple to visualize although the procedure is evidently quite straightforward. If two more variables are added a three-dimensional map of $4\times4\times4$ elements is needed. Each level can be drawn separately, however, on the same plane, and with practice considerable skill in its use can be acquired.

Combinational Circuits

By representing the two states of binary logic or arithmetic by two voltages, say (+6 V and zero), it is possible to realize all the functions and the expressions hitherto discussed, and many more.

Suppose we have two voltage sources or circuit nodes A, B each of which can be either positive or zero. Either the positive polarity or the more negative one (i.e. zero voltage) may be regarded as the significant polarity. If the positive value is chosen, the system is called *positive logic*, the positive polarity is called the "1" state, and the other the "0" state. If the more negative polarity is chosen, the system is called *negative logic.*

If as is common the *complements* A and B are made available as two further sources or nodes they are called $\bar{A}$ and $\bar{B}$. (Thus if $A=1$, $\bar{A}=0$ and if $A=0$, $\bar{A}=1$.)

If the two voltage sources A, B are inputs to a combinational circuit, the output of the circuit may be one of many functions of the inputs. Two functions of two or more variables have already been mentioned, namely the AND and OR functions. These and

TABLE 4.1

Inputs		Outputs					
A	B	AND	OR	NOR	NAND	EX–OR	INH
0	0	0	0	1	1	0	0
0	1	0	1	0	1	1	1
1	0	0	1	0	1	1	0
1	1	1	1	0	0	0	0
		AB	$A+B$	$\overline{A}\overline{B}$	$\overline{AB}$	$A\bar{B}+\bar{A}B$	$\bar{A}B$

the other most commonly used functions of two variables are shown in the *truth table* of Table 4.1.

A truth table is one which tabulates the outcome of all the possible combinations of the inputs. In this case only two inputs A, B are considered, so there are four possible combinations of A, B and four rows to the table.

The NOR and NAND circuits are the complements of OR and AND. The Exclusive-OR function excludes from the OR function the simultaneous occurrence of $A=1$ and $B=1$. The output of the Inhibit gate is the same as input B except when $A=1$. Thus A inhibits B.

The extension of all these functions to gates with multiple inputs is clear, except in the case of the Exclusive OR gate. This gate can have only two inputs unless some further rules are laid down.

In combinational and sequential circuit diagrams it is usual to use special symbols to represent each logical function, and unfortunately a very wide variety of different symbols for the same functions have come into common use and have even been included in standards documents (see British Standard B.S. 530, supplement 5 and the American A.S.A. Y 32.14). The symbols used in this book are shown on page 84. The code used is quite

straight forward. The full stop stands for AND, the + sign stands for OR, both as in the Boolean equations, and the triangular shape indicates, as usual, an inverting amplifier restricted in this case to gain of unity, thus the triangle stands for the NOT function. The reasons for not using one of the sets of standards are as follows: in the British system there is no clear distinction between inputs and outputs, the inverting function is represented as something that happens after, or before, the signal has passed through the gate which is never the case, and the linking of AND gates and OR gates with threshold gates has no meaning in reality. In the American system the graphical codes have no symbolic meaning and are very difficult to remember.

The use of some of these functions, and their manipulation by Boolean Algebra, can be illustrated with the following simple example.

Example. *Suppose an output Z is required which is the following function of three inputs A, B, C, and their complements.*

$$Z = \bar{A}\bar{B} + B(\bar{C} + \bar{A}). \tag{4.4}$$

Realize this function using AND, OR, NOT, *gates,* NAND *gates and* NOR *gates.*

(a) *Using* AND, OR, NOT *gates*

The function can be realized in various ways using AND, OR and NOT gates. In Fig. 4.2(a) the function is realized as written in eqn. (4.4).

The presence of an inverter (NOT gate) on each input suggests an immediate simplification; that of realizing the complement of Z and then inverting this output, instead of inverting all the inputs.

Using de Morgan's theorem

$$\bar{Z} = (A + B)(\bar{B} + AC) \tag{4.5}$$

and the function Z can be realized as in Fig. 4.2(b).

However, eqn. (4.4) can be simplified as follows:

$$Z = \bar{A}(\bar{B}+B)+B\bar{C}$$
$$= \bar{A}+B\bar{C} \qquad (4.6)$$

and can therefore be realized in the simpler form of Fig. 4.2(c).

(b) NOR *and* NAND *gates*

It is always possible to express a Boolean equation in terms of two levels of NOR or NAND functions. The simplest procedure

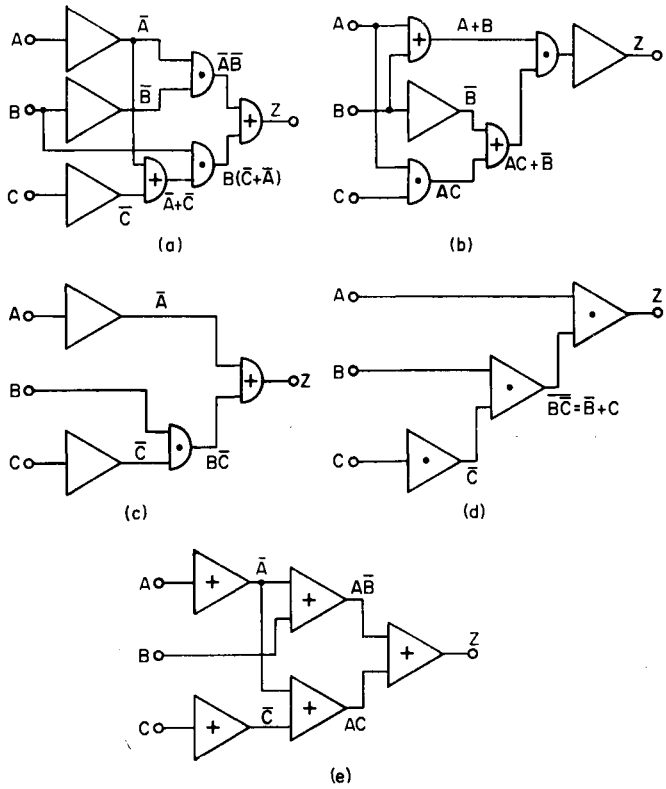


FIG. 4.2. Various combinational circuits all of which generate the function $Z = \bar{A}\bar{B}+B(\bar{C}+\bar{A})$. (a), (b), (c) Using AND, OR and NOT gates. (d) Using NAND gates. (e) Using NOR gates.

is as follows, but a more efficient method is considered in Chapter 8.

For NAND gates the Boolean expression is reduced to its simplest form containing the *sum of products*—using a Karnaugh map for example. Then each product term is the input equation for a NAND circuit, and the outputs of each of these NAND circuits form the inputs to a final NAND gate.

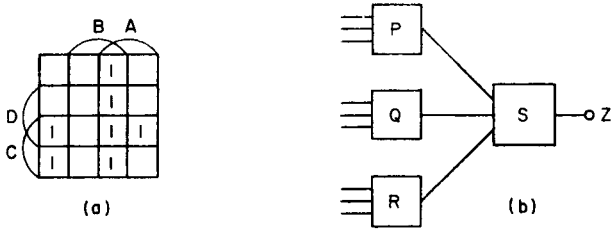


FIG. 4.3. (a) The Karnaugh map of eqn. (4.7). (b) The general arrangement of two-level NOR or NAND circuitry.

Consider the expression

$$Z = AB + C(D\bar{B} + \bar{A}\bar{B}\bar{D}).$$

The Karnaugh map of Fig. 4.3(a) shows a simpler grouping as

$$Z = AB + \bar{A}\bar{B}C + ACD. \tag{4.7}$$

Then with four NAND gates P, Q, R, S arranged as in Fig. 4.3(b), the inputs of P, Q, R will be AB, $\bar{A}\bar{B}C$ and ACD respectively. By de Morgan's theorem their outputs will consequently be $\bar{A}+\bar{B}$, $A+B+\bar{C}$, $\bar{A}+\bar{C}+\bar{D}$. These three outputs combined at the input to NAND gate S will yield the output Z above (by de Morgan's theorem again). If some of the inputs to the first bank of gates P, Q, R are complemented inputs, as in this case, preliminary inverters may be needed to to obtain them.

For NOR gates the Boolean expression is first reduced to its simplest form containing the *product of sums*. This is often easily done by grouping the *complement* of the expression as a sum of

products in a Karnaugh map. The required product of sums is then obtained using de Morgan's theorem. Thus using the example of eqn. (4.7) and its Karnaugh map of Fig. 4.3(a), the elements of the map *not* indicated by eqn. (4.7) can be grouped as

$$\bar{Z} = \bar{A}B + B\bar{C} + A\bar{B}\bar{D},$$

so that, by de Morgan's theorem,

$$Z = (A + \bar{B})(B + C)(\bar{A} + B + D).$$

The inputs to the first three NOR gates P, Q, R arranged as in Fig. 4.3(b) will be $A + \bar{B}$, $B + C$ and $\bar{A} + B + D$ respectively. Their outputs combined at the input to a fourth NOR gate, S, will give rise to the required function of Z, namely eqn. (4.7).

Returning to the simpler example of eqn. (4.6) in which

$$Z = \bar{A} + B\bar{C}$$

the NAND gate realization is now evidently that shown in Fig. 4.2(d) whilst the NOR gate realization is obtained by writing

$$Z = (\bar{A} + B)(\bar{A} + \bar{C})$$

(using theorem (9) or a Karnaugh map). Then Fig. 4.2(e) shows the obvious arrangement. (See problem 4.8.)

Encoders and Code Translators

A simple but important use of combinational circuits and of Karnaugh maps is in the construction of coders, as the following example illustrates.

The numbers 0 to 9 can be represented in a variety of logical combinations of four binary digits. Three of the more important are shown in Table 4.2.[14]

In the simple Binary code, if the digits A, B, C, D are given the weights 1, 2, 4, 8 respectively, then the code readily represents the decimal numbers. Thus, for example, 0110=6.

TABLE 4.2

	Simple binary code	Excess 3 code	Reflected Excess 3 code
	DCBA	*DCBA*	*DCBA*
0	0000	0011	0010
1	0001	0100	0110
2	0010	0101	0111
3	0011	0110	0101
4	0100	0111	0100
5	0101	1000	1100
6	0110	1001	1101
7	0111	1010	1111
8	1000	1011	1110
9	1001	1100	1010

The Excess 3 code is the same as the simple binary code but shifted by three decimal digits. In this code the "nines complement" (i.e. the number subtracted from nine, an important quantity in arithmetic processes) is formed simply by complementing all digits.

The Reflected Excess 3 code is one of many reflected or "Gray" codes in which consecutive numbers differ by only one digit. In this particular one 9 differs from zero by only one digit too. The nines complement is obtained by complementing only the *D* digit.

A frequent operation involves translating from one code to another. Suppose it is necessary to translate from Simple Binary to Reflected Excess 3 code. A map is drawn for each digit A' B' C' D' of the Reflected Excess 3 code in terms of the digits of the Simple Binary code as shown in Fig. 4.4. The entries in the maps can be derived from Table 4.2. For example, A of the Reflected Excess 3 Code, namely A', is "1" according to the following equation in terms of the Simple Binary code:

$$A' = \bar{A}\,B\,\bar{C}\,\bar{D} + AB\,\bar{C}\,\bar{D} + \bar{A}\,B\,C\,\bar{D} + A\,B\,C\,\bar{D}.$$

The blank spaces in the maps indicate binary code "words" that never occur.

The "1's" entered in the map have now to be grouped to form Boolean expressions, including any blanks which are convenient, but excluding the zero's.

$$A' = B$$
$$B' = \bar{A}\bar{C} + \bar{B}\bar{C} + ABC$$
$$C' = B + C + A\bar{D} + \bar{A}D$$
$$D' = D + CB + CA$$

If more than four variables are involved Karnaugh maps are inconvenient and more general methods have to be used.[15]

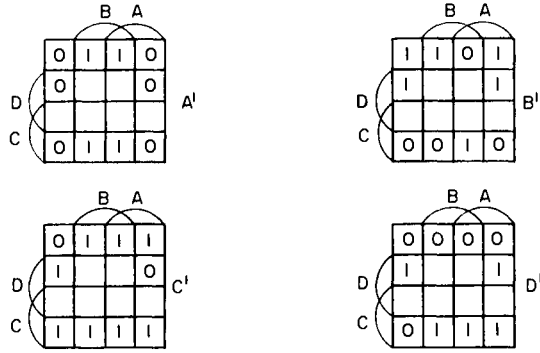


FIG. 4.4. The Karnaugh maps for the translation of simple binary code into Reflected Excess 3 code.

Practical Combinational Circuits

Table 4.1 defining the various logical functions (AND, OR, NAND, NOR, etc.) which combinational circuits can perform refers *either* to positive logic *or* to negative logic. If it is concerned with positive logic then all the 1's refer to the more positive of the two available voltage levels, and the 0's to the more negative. Except

where otherwise stated, *Positive logic will be used throughout this book* and all Boolean expressions describing circuit operations will be in terms of it.

However, it is important to realize that OR gates in one logic system are AND gates in the other, and that NOR gates in positive logic are also NAND gates in a negative logic system. This can be easily verified by inspection of Table 4.1. If the 0's are regarded as the significant values then OR becomes AND and vice versa, and NOR becomes NAND and vice versa. Thus OR and NOR are complements of AND and NAND functions respectively. The positive logic circuits described in the following pages are therefore also negative logic circuits of the complementary functions.

In addition to using only positive logic, *npn* transistors will be used, where possible and relevant. Of course, negative logic circuits can always be built according to the same circuit diagram but with all *npn* transistors replaced by *pnp* ones, all supplies polarities inverted and all diodes turned round.

Many methods of realizing these logical functions in combinational circuits have been proposed and are commercially available. The most important ones are briefly described below under the code names by which they are usually, though not always, described. In addition a brief analysis is given of why a circuit's performance may be inadequate for some purposes and why, therefore, a more complex circuit has been developed.

The performance differences between the various circuit designs are usually concerned with *speed of operation*, *fan-in* and *fan-out*, and *noise immunity*. The speed is characterized primarily by the rate at which the output, loaded "normally", especially by capacitance, rises (or falls) in response to a step input. The fan-out is the number of identical circuits the output can drive without serious deterioration of performance. (Evidently fan-out and speed of response are interrelated to some extent.) Fan-in is the number of identical inputs a circuit can accommodate without serious deterioration of performance. (This too, is often

associated with speed of response.) Finally noise immunity. Most circuits will give a spurious output if voltage fluctuations exist on the inputs, the earth line or on the supply lines (as a result, for example of pick up, or of the changing current demands of other circuits). The more resistant the circuit is to these fluctuations, the higher its noise immunity.

The most complex type of gate described, namely High-level Transistor–Transistor Logic, is designed to maximize all these parameters in one circuit.

Diode Logic (D.L.)

This is the simplest practical combinational circuit and is illustrated in Fig. 4.5. In the AND gate the most negative input holds the output down. Since each input can have only one of two voltages, the output will only become positive if *all* the inputs are at the *more positive* of the two available levels. The supply voltage must be greater than or equal to this level. In the OR gate *any one* positive input will produce a positive output.

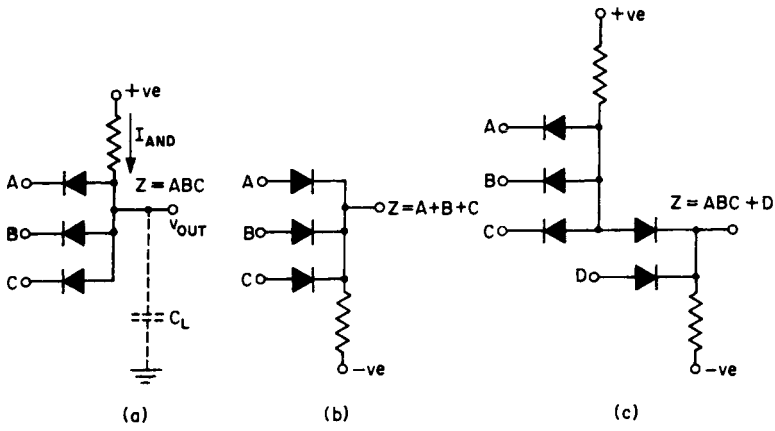


FIG. 4.5. Positive logic diode gates. (a) AND gate; (b) OR gate; (c) a two-stage AND/OR gate.

This circuit is usually designed to drive only one output (i.e. fan-out = 1) and to avoid loading its inputs the gate resistor is usually made as large as possible. But this means that the rising edge of the AND gate and the falling edge of the OR gate are slow if the output is capacitively loaded. For example, if C_L is the value of the capacitor loading the AND gate, and if I_{AND} is the AND gate current $C_L\, dV_{OUT}/dt = I_{AND}$. The smaller I_{AND} the slower the rise of the output voltage.

Two stages of diode logic can be constructed as shown in Fig. 4.5(c). When the output voltage is rising $C_L\, dV_{OUT}/dt = I_{AND} - I_{OR}$ and when it is falling $C_L\, dV_{OUT}/dt = -I_{OR}$. (A more detailed analysis of this kind of circuit appears in Appendix B.)

Resistor Transistor Logic (R.T.L.)

In this circuit (Fig. 4.6(a)) all the input resistors R_A, R_B, R_C are the same value and if any one input goes positive the resulting base current is sufficient to bottom the transistor and drive the output negative. The circuit is simply a multiple input inverter, or NOR gate. If more than one input goes positive the transistor becomes more heavily bottomed.

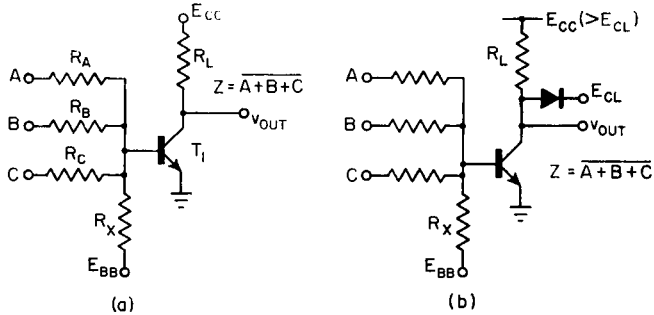


FIG. 4.6. (a) Transistor resistor logic NOR gate; (b) clamped R.T.L. NOR gate (positive logic). Typically $E_{BB} = -6$ V, $E_{CL} = +6$ V and $E_{CC} = +6$ V in (a) and $+12$ V in (b).

Normally each input A, B, C... will be the output of a similar circuit so that the more negative of the input voltage levels will be $V_{CE(sat)}$. When all inputs are at $V_{CE(sat)}$ transistor T_1 must be cut off, so that $V_{OUT} \approx E_{CC}$. But if $V_{CE(sat)}$ is not sufficiently small (and this depends upon the transistors involved) T_1 may be only partially cut off. For this reason with germanium transistors R_X is added to the circuit, to ensure that the NOR gate transistor is held cut-off when it should be (see p. 35). With silicon R_X is not necessary.

The gain of the transistor produces a much lower output resistance than in D.L. and therefore a larger fan-out.

A detailed analysis of this circuit appears in Appendix A where it is shown that the input resistors tend to draw current from each other causing a limit to be set on fan-in and fan-out.

The speed of this circuit is limited by the near-constant current drive provided by the input resistors. If these resistors are shunted by capacitors to inject the turn-on charge (and remove the turn-off charge) the speed of the circuit is significantly improved, but again the capacitors interfere with each other, and load the driving transistors.

Clamped R.T.L.

One of the consequences of attempting to obtain a large fan-out with R.T.L. is that the current drawn by all the circuits connected to the output of the NOR gate causes the more positive output voltage to fall significantly below E_{CC}. This can be avoided, and the fan-out increased, by increasing E_{CC} to above the logic level and by clamping the output to a second supply E_{CL} as in Fig. 4.6(b). With this arrangement an output current up to $(E_{CC}-E_{CL})/R_L$ can be drawn from the circuit without changing the output voltage significantly (see Appendix A).

Diode Transistor Logic (D.T.L.)

The loading of one input by the other inputs, which reduces the performance of R.T.L., can in principle be prevented by replacing the resistors with diodes, as in Fig. 4.7(a), giving a diode logic gate with inverted output. Unfortunately this circuit will not normally function properly since when all inputs A, B, C are at zero volts (or, more exactly, $V_{CE(sat)}$) the output of the

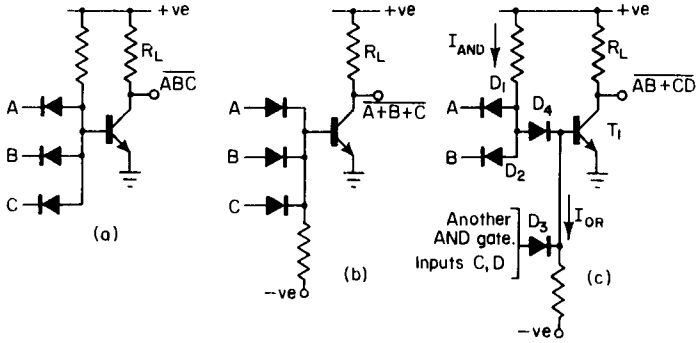


FIG. 4.7. Diode transistor logic. (a), (b) Unsatisfactory circuits. (c) A circuit with three stages of logic, AND/OR/NOT, which overcomes the inadequacies of (a) and (b).

diode gate is above zero by the magnitude of the forward bias voltage of the diodes, and this is too large to ensure that the following transistor is cut-off. (If the diodes were of germanium and the transistor of silicon the circuit would be satisfactory. However, mixed systems of this kind cannot be made in integrated form.)

The circuit of Fig. 4.7(a) is a NAND gate. The corresponding NOR gate shown in Fig. 4.7(b) will function correctly but the base current is unnecessarily large when any one input is driven positive by a similar, preceding stage. Thus the NAND gate is unsatisfactory when the transistor is cut-off, and the NOR gate is poor when the transistor is bottomed.

The cascade AND–OR–NOT arrangement of Fig. 4.7(c) neatly

solves both problems. When the AND gate inputs are positive the AND gate diodes are cut-off so the base current is I_{AND}–I_{OR}, and the transistor is not heavily bottomed. When, on the other hand, the AND gate inputs are at zero volts (or one of them is) both the AND gate and OR gate diodes are forward biased so that the base of the transistor is held at about zero volts, ensuring it is properly cut-off. Thus current is steered towards the transistor or away from it by the input voltages.

With germanium devices it is usual to use two (or more) diodes in series in place of each OR gate diode D_3, D_4 shown in Fig. 4.7(c). This ensures that the transistor will be cut-off despite diode tolerances, and provides greater noise immunity.

The transient response of this circuit depends upon I_{AND}, I_{OR} and the transistor properties (see eqn. (1.36)). The transistor is turned off by I_{OR}, and for rapid response I_{OR} should be large. The turn on current is I_{AND}–I_{OR} so I_{AND} must also be large. Thus fast switching involves reducing the fan-out of any driving circuit. (See Appendix B for a more detailed analysis.)

These gate currents can be reduced, somewhat fortuitously without reducing switching speed, by using "slow" diodes in the OR gate. If the OR gate diodes are slow to turn off, because of minority carrier storage, their reverse current will not fall immediately to zero when the forward bias is removed; instead quite a large transient current will flow. This current can be used to help turn off the transistor. Indeed if the charge stored in the diode is equal to Q_{OFF} of the transistor, no extra, transient turn-off current will be required from the OR gate and I_{OR} can be designed simply to maintain steady state conditions.

In order to achieve a rapid turn off of these OR gate diodes, the AND gate diodes must turn on rapidly when the input falls. Thus the AND gate diodes must be fast, even though the OR gate ones are slow. The need for diodes of different response times in one circuit makes the construction of this circuit very awkward in integrated form since silicon of two different lifetimes is needed.

Diode Resistor Transistor Logic (D.R.T.L.)

Some improvement in D.T.L. can be achieved by including a parallel R-C circuit in the base lead as shown in Fig. 4.8. Part of the difficulty in D.T.L. arises as follows. In order to turn on the transistor rapidly I_{AND} has to be large; this causes the transistor to be heavily bottomed in the steady state, necessitating a large current (I_{OR}) to turn it off again within a given time. The

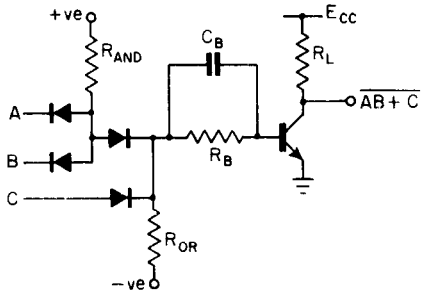


FIG. 4.8. Diode resistor-transistor logic. AND/OR/NOT gate.

capacitor C in Fig. 4.8 conveys the large transient current I_{AND} to the base of the transistor, but resistor R_B limits the steady state current to I_C/β(min). This allows a reduction in I_{OR} and therefore also in I_{AND}. (See Beaufoy[16] for an analysis of this circuit.)

Transistor–Transistor Logic (T.T.L.)

Each pair of back-to-back diodes which appears in diode transistor logic (Fig. 4.7(c)) can be combined in one transistor as shown in Fig. 4.9(a). When the emitters of all the input transistors are positive, their base current, I_B, flows through their forward biased collectors, bottoming T_3. When the emitter voltage of any one input transistor, T_1 say, falls to zero, I_B flows

through this emitter instead, reducing the base current of T_3 to near zero. Thus T_1 becomes heavily bottomed ($I_E \gg I_C$) and T_3 is cut-off with its base held firmly to near zero volts by T_1.

During the transient, when T_3 is being turned off, T_1 is momentarily in, or near, the Active Region of Operation ($V_{BC} \approx 0$) so that current readily flows from the base of T_3 through T_1 until the base charge of T_3 is exhausted—after which T_1 bottoms and T_3 is cut off. Thus the disposal of Q_{OFF} of T_3 is efficiently dealt with.

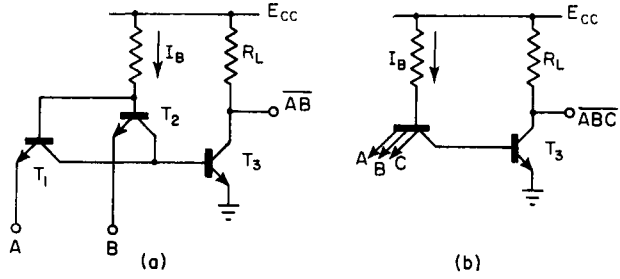


FIG. 4.9. (a) Transistor–transistor logic NAND gate. (b) Multiple emitter transistor logic NAND gate.

Multiple Emitter Transistor Logic (M.E.T.L.)

M.E.T.L. is a natural (but ingenious) development of T.T.L. in which all the input transistors are combined in one multiple-emitter transistor, as shown in Fig. 4.9(b). Any one emitter taken to zero volts will draw I_B away from T_3 cutting it off, as in T.T.L.

The *inverse* current gain of each transistor (whether of T.T.L. or M.E.T.L.) must be small so that when an emitter is positive, excessive current does not flow in the emitter of the transistor as a result of inverse transistor action.

The difficulty with both M.E.T.L. and T.T.L. arises primarily from its poor noise immunity. T_3 in either circuit is only

lightly cut off (the emitter junctions are not reverse biased) so that only a small noise pulse in the earth line or on the input is needed to cause a spurious pulse at the output.

The speed of response of the circuit is good. However, its output resistance is still too large to enable it to drive many similar circuits rapidly, especially if stray capacitance loading is severe. In other, words, at its normal, high response speed its fan-out is poor. This problem is solved as follows.

High-level Transistor–Transistor Logic (H.L.T.T.L.)

H.L.T.T.L. is a development of M.E.T.L. which removes its remaining disadvantages.

The relatively high resistance output of M.E.T.L. is buffered by a push–pull emitter follower circuit, which greatly increases the fan–out and the circuit's ability to drive capacitive loads. The diagrams of Fig. 4.10 show the two states of the circuit. It is assumed that $V_{CE(sat)}$ of each silicon transistor is not more

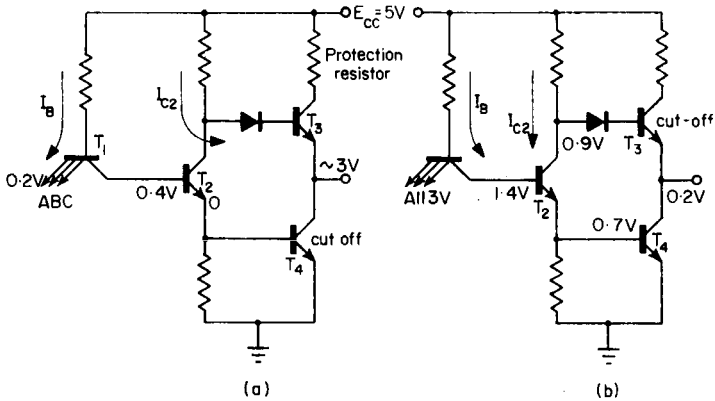


FIG. 4.10. High-level transistor–transistor logic NAND gate with push–pull output, showing voltage levels in the circuit. (a) With the output in the "1" state (positive); (b) with the output in the "0" state.

than 0·2 V. In Figure 4.10(b) it will be seen that a total of about 0·7 V exists across the diode and the emitter of T_3 in series. This is insufficient to cause significant current flow through two silicon *pn* junctions so that T_3 is cut-off.

Noise-voltage spikes in the earth line or at the input have to exceed about 600 mV before significant signals appear at the output, which is several hundred millivolts better than M.E.T.L.

This circuit completes the development at the time of writing of circuits in which logic functions are realized in terms of voltages at the inputs to transistors.

Direct Coupled Transistor Logic (D.C.T.L.)

The technique illustrated in Fig. 4.11 introduces the idea of combining logic functions at the *outputs* of inverting transistors instead of at their inputs. In the NOR gate, for example, when any one input goes positive the output drops to near zero (i.e. $V_{CE(sat)}$). This output is normally sufficiently low to cut off any following transistor, thus making directly coupled cascade circuits possible.

The simplicity of the circuit and the fact that it can be so easily manufactured in integrated form makes it very attractive. Unfor-

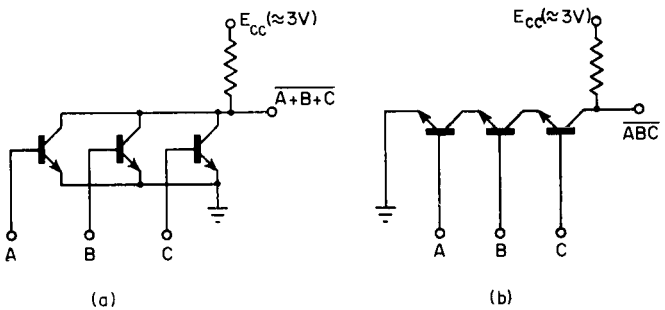


FIG. 4.11. Direct coupled transistor logic. (a) NOR gate, (b) NAND gate.

tunately, in practice, the emitter–base voltages of transistors cannot be made to sufficiently tight tolerances to prevent "current hogging" and the method is not satisfactory.

Current hogging arises as follows: If one transistor output is driving two or more other transistor inputs, the available base current will only be shared equally if the values of $V_{BE(ON)}$ of the load transistors are equal. In practice small differences in $V_{BE(ON)}$ occur, and owing to the exponential input characteristic of a transistor this results in large base current differences. The problem can be overcome by padding each base input with a series resistor. This produces R.C.T.L.

Resistor Coupled Transistor Logic (R.C.T.L.)

The current hogging of *D.C.T.L.* is prevented by the inclusion of a padding resistor in series with each base, as shown in Fig. 4.12. These resistors are kept as small as possible consistent

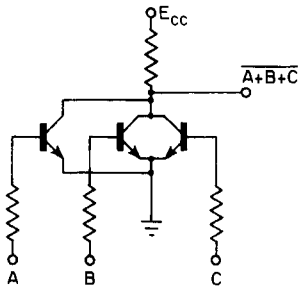


FIG. 4.12. Resistor coupled transistor logic NOR gate.

with the prevention of serious current hogging and this provides a heavy overdrive of base current at both turn-on and turn-off (using silicon transistors). It is a widely used technique, sometimes called Digital Computer Transistor Logic, D.C.T.L!

Current Steering or Emitter Coupled Transistor Logic (E.C.T.L.)

A technique known as *current steering* allows transistors to be switched at the ultimate speed of which they are capable. The penalty for this high speed of operation is the relatively large power dissipation of the circuits.

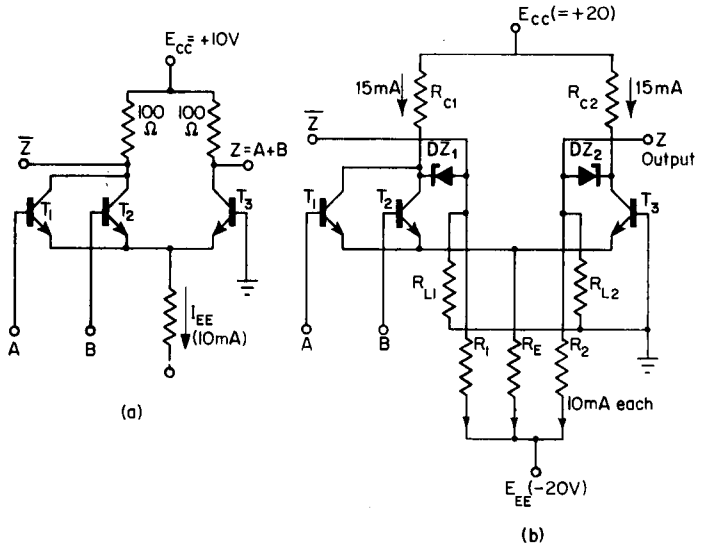


FIG. 4.13. Current steered, or emitter coupled transistor logic OR gate. (a) The basic circuit; (b) a practical circuit including level shifting Zener diodes. $R_{L1} = R_{L2} = 100\ \Omega$.

The basic circuit is a long-tailed pair, with extra input transistors added as shown in Fig. 4.13(a). In the example shown the output voltage change is only 1 V.

If, to begin with, all inputs A, B ... are at $-0{\cdot}5$ V, the current I_{EE} flows through T_3, lowering the output Z by 1 V, from 10 V to 9 V in this example. If any input goes to $+0{\cdot}5$ V, say input A, then I_{EE} will be diverted through T_1 (its base is now more posi-

tive than that of T_3) and the output Z will rise to $+10$ V. Thus the circuit is a positive-logic OR gate except that the output voltage levels are shifted by 9·5 V with respect to the input voltage levels.

The circuit has a complementary output so that it is also a NOR gate.

To overcome the difficulty caused by the change in voltage level between input and output, the circuit is modified as in Fig. 4.13(b).

The large supply voltages ensure that R_{C1}, R_{C2}, R_1, R_2, R_E all supply almost constant currents to the circuit. The currents through R_{C1} and R_{C2} are about 15 mA whilst the other three are each 10 mA in this illustration.

When T_3 is cut-off, all the 15 mA through R_{C2} flows through the Zener diode DZ_2. 10 mA flows away through R_2 and 5 mA through R_{L2}. Since $R_{L2}=100\,\Omega$, output Z rises to $+0{\cdot}5$ V.

When T_3 is conducting (i.e. both inputs A, B are at $-0{\cdot}5$ V) the 10 mA required by R_E flows through T_3. This means that only 5 mA now flows through DZ_2, but R_2 still draws 10 mA, which means that the current through R_{L2} must now be 5 mA again, but in the opposite direction, so that output Z falls to $-0{\cdot}5$ V. R_1, R_{L1} and DZ_1 need only be connected if a complementary output is required.

Very rapid switching times can be achieved with this circuit; rise times of only a few τ_C are possible. But this performance is only obtained at the cost of exceptionally high power dissipation in the circuit, about 1·5 W in this case. The high-speed switching occurs for the following reasons.

(1) The transistors remain in the Active Region of Operation throughout, and at a large reverse bias. (About equal to the Zener voltage of DZ_2.) This keeps τ_C and the collector capacitance small.

(2) The voltage swing is small and the current relatively large. This means that little charge is required to drive any stray cir-

cuit capacitances or the collector capacitance, and also that there is plenty of current to supply this charge rapidly.

(3) The current of 10 mA in high-speed, low-level transistors is a typical one for achieving minimal values of τ_C.

(4) Zener diodes exhibit no carrier storage effects in the breakdown region.

(5) The driving circuit which initiates the switching operation is a low impedance one. The importance of this can be understood as follows.

When, for example, T_3 is turning off and T_1 is turning on, the base charge Q_{OFF} of T_3 is transferred via the wire connecting the emitters to the base of T_1. If this is achieved in a switching time t_s, then a mean current of Q_{OFF}/t_s must flow between the base regions and the input circuit.

The current which the input circuit produces is approximately the magnitude of the driving voltage step, ΔV_{IN}, divided by the circuit resistance. If the source is the output of a similar circuit, the net circuit resistance is R_L+2r_x, where r_x is the extrinsic base resistance of each transistor. Thus

$$t_s \approx \frac{Q_{OFF}(R_L+2r_x)}{\Delta V_{IN}}.$$

If ΔV_{IN} is to be kept small, R_L, r_x must also be small.

If Q_{OFF} of one transistor does not equal Q_{ON} of the other, then overshoot or undershoot of the output waveforms will result.

In practice the inductance in the emitter wire through which the drive current has to flow is likely to limit switching speed or cause "ringing" in the output voltage.

This concludes the description of the various techniques for constructing AND, OR, NAND and NOR circuits. Finally, the construction of exclusive-OR and inhibit circuits must be briefly considered.

Exclusive OR *and Inhibit Gates*

Exclusive-OR gates can be constructed from combinations of NOR, NAND, AND and OR gates as indicated in Fig. 4.14. Note especially that since an AND gate can be made from three NOR gates, Fig. 4.14(a) can be modified to contain five NOR gates, which is

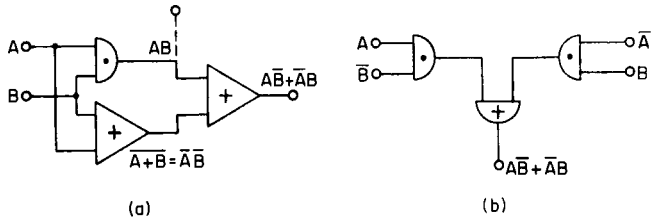


FIG. 4.14. Exclusive-OR gates. (a) With only uncomplemented inputs (also shown is the output AB needed in the half adder—see p. 198). (b) Using AND–OR gates.

particularly simple to construct using transistors. There are, however, interesting circuits which naturally generate the inhibit function, and which, when combined, form simple exclusive-OR circuits.

The two principal forms of inhibit gates are shown in Fig. 4.15. In the circuit of Fig. 4.15(a) the transistor is cut-off except when input B is positive and A is at zero. In Fig. 4.15(b) the output is at (near) zero volts unless input A is positive and B is at zero. Additional B inputs can be added, as shown dashed. In each case the A inputs inhibit the action of the B inputs.

The circuits of Fig. 4.15 use R.T.L. as the design method, but versions of the same circuit using other methods, are, of course, possible.

Figure 4.16 shows two of the inhibit circuits of Fig. 4.15(a) combined to form an exclusive-OR circuit (again using R.T.L.). Since inputs A, B have to supply the emitter or collector currents

of T_1 or T_2, the logical gain (fan-out) is obtained by operating T_3 at a current which is an order of magnitude greater than that of the other two.

In the circuit of Fig. 4.16 the inverter T_3 can be replaced by an emitter follower to give an exclusive-NOR function ($AB+\bar{A}\bar{B}$) instead of the more usual exclusive-OR.

The inhibit circuit of Fig. 4.15(b) can be used to construct an alternative form of exclusive-NOR gate.

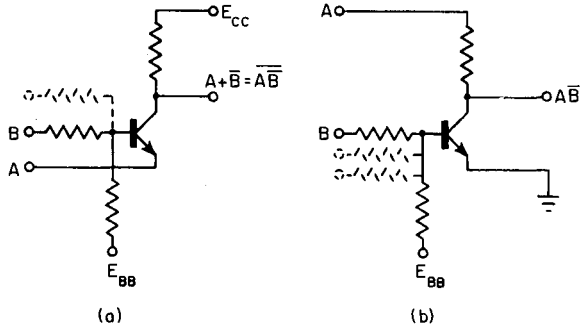


FIG. 4.15. Inhibit gates.

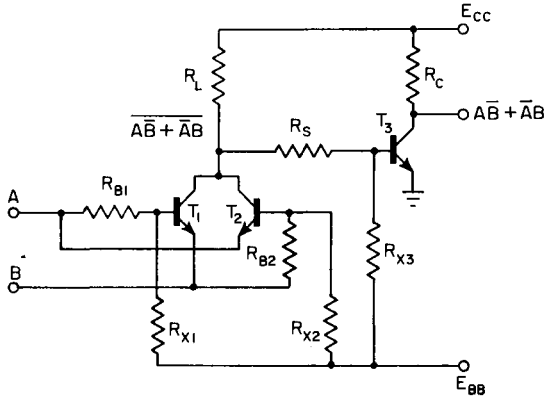


FIG. 4.16. Inhibit gates combined to produce an exclusive-OR gate.
$R_X \approx 3R_B \approx 20R_L \approx 20R_S \approx 200R_O$.

Problems

4.1. *Simplify* (using Karnaugh maps or otherwise)

(a) $A\bar{B}\bar{C}+A\bar{B}\bar{C}D+\bar{C}D+AB\bar{C}$. (Ans.: $A\bar{C}+\bar{C}D$.)
(b) $A\bar{B}C+\bar{A}\bar{C}D+A\bar{C}$. (Ans.: $\bar{C}D+A\bar{B}+A\bar{C}$.)
(c) $(A+B+CD)(\bar{A}+B)(\bar{A}+B+D)$. (Ans.: $B+\bar{A}DC$.)
(d) $(A+B+\bar{C})(B+\bar{C}+\bar{D})(A+D)$. (Ans.: $A\bar{D}+BD+D\bar{C}$.)

4.2. *Complement*

(a) $(A+B+\bar{C})(B+\bar{C}+\bar{D})(A+D)$. (Ans.: $\bar{A}\bar{B}C+\bar{B}CD+\bar{A}\bar{D}$.)
(b) $AB+\bar{C}+\bar{D}$. (Ans.: $(\bar{A}+\bar{B})CD$.)
(c) $AB+\bar{C}\bar{D}$. (Ans.: $(\bar{A}+\bar{B})(C+D)$.)
(d) $(\bar{A}B+\bar{C})D+\bar{E}F$. (Ans.: $(A+\bar{D}+\bar{B}C.)(E+F)$.)

4.3. Draw logic circuits to implement each of the above expressions,

(a) Using AND, OR, NOT circuits.
(b) Using NOR circuits.
(c) Using NAND circuits.

4.4. In the diode logic circuit of Fig. 4.5(c) the upper resistor = 56 kΩ and the lower = 150 kΩ, the supply voltages are +20 V and −10 V and the output is loaded by a 20 pf capacitor. If inputs B, C, D all remain at zero volts what are (a) the rise time, (b) the fall time of the output voltage when input A is driven abruptly from zero to +10 V and then, when a steady state has been reached, back to zero again?

Assume that there is no voltage drop across the conducting diodes and that no current flows through them when they are cut-off.

Repeat the calculation assuming 0·7 V drop across the diodes when they are conducting.

(Ans.: 15·25, 20·8, 15·9, 20·8 μsec.)

4.5. In the R.T.L. circuit of Fig. 4.6(a) $R_L = 1$ kΩ, $E_{CC} = +6$ V, the three base resistors are all 8 kΩ and $R_X = \infty$. If all the inputs are initially at zero volts, calculate the rise time and total turn-off time of the output *current* when one input rises abruptly to +6 V, and then, when a steady state has been reached, falls abruptly to zero again. (Assume $V_{BE(ON)} = 0{\cdot}7$ V, $\beta = 50$, $\tau_O = 0{\cdot}012$ μsec, $Q_{VO} = 80$ pc, $\tau_S = 0{\cdot}6$ μsec.) (Ans: 0.29 μsec, 0.95 μsec).

What are the turn-on and turn-off times if the fan-in is increased from three to six?

(Ans.: $t_r = 0{\cdot}612$ μsec, 0·42 μsec.)

If the transistor output is loaded by a 200 pf capacitor to earth, *estimate* the turn-on time (i.e. the time taken for the output *voltage* to fall to zero volts) assuming $V_{CE(sat)} = 0$ for both values of fan-in. If a much faster transistor is used in which $\beta = 50$, but τ_S, τ_C and Q_{VC} are negligible, what would be

the turn-on time for both values of fan-in and what would be the time constant of the turn-off edge?

4.6. Design the gates necessary to convert a simple binary code to an Excess-3 code (see p. 98).

(Ans.: $A'=\bar{A},\ B'=\bar{A}\bar{B}+AB,\ C'=A\bar{C}+B\bar{C}+\bar{A}\bar{B}C,\ D'=D+BC+AC$).

4.7. Design the gates necessary to convert a Reflected Excess 3 code to a simple binary code.

(Ans.: $A' = D(\bar{C}+AB+\bar{A}\bar{B})+\bar{D}(A\bar{B}+\bar{A}BC)$,

$B' = A,\ C' = AD+\bar{A}\bar{B},\ D'=\bar{A}BD$).

4.8. Prove the validity of the methods described on p. 96 for the constuction of NAND gate and NOR gate circuits from Boolean equations.

4.9. Design a circuit to compare two four-digit, binary coded numbers A, B. There should be three outputs to indicate respectively $A>B$, $A=B$, $A<B$.

4.10. Design a circuit with four inputs and five outputs, such that the outputs indicate the number of excited inputs: i. e. none, any one, any two, any three, all four.

CHAPTER 5

Bistable Elements for Sequential Circuits

SINCE sequential circuits must be capable of "remembering" previous inputs the simplest method of constructing them is by interconnecting circuits which possess memory. That is, by interconnecting circuits which will remain in a clearly distinguishable state after the inputs which set up that state have been removed.

The commonest memory device nowadays is the ferrite magnetic core which, because of its square hysteresis loop can be set in either of two quite distinct magnetic states, and it will remain there after the magnetizing field has been removed. Some special devices such as tunnel diodes or *pnpn* diodes or triodes are also bistable, but relays and transistors are simply switches which can be driven to one of two clearly distinguishable states (on or off) but which relapse to one preferred state, or dangle uncertainly, if the drive is removed.

In order to make a transistor into a bistable element the simplest method is to interconnect two of them, A and B, so that when A is cut-off B is held on and vice-versa. The usual circuit for achieving this is shown by the heavy lines in the diagrams of Fig. 5.1 and is called the Eccles–Jordan circuit, or the bistable multivibrator. It is evidently similar to the two previously considered multivibrators except that here both states of the circuit (T_1 on and T_2 off, or T_1 off and T_2 on) are permanently stable. (Previously only one state or neither state was permanently stable.)

The circuit can also be thought of as two R.T.L. NOR gates connected in a closed loop. The output of each NOR gate drives the input of the other.

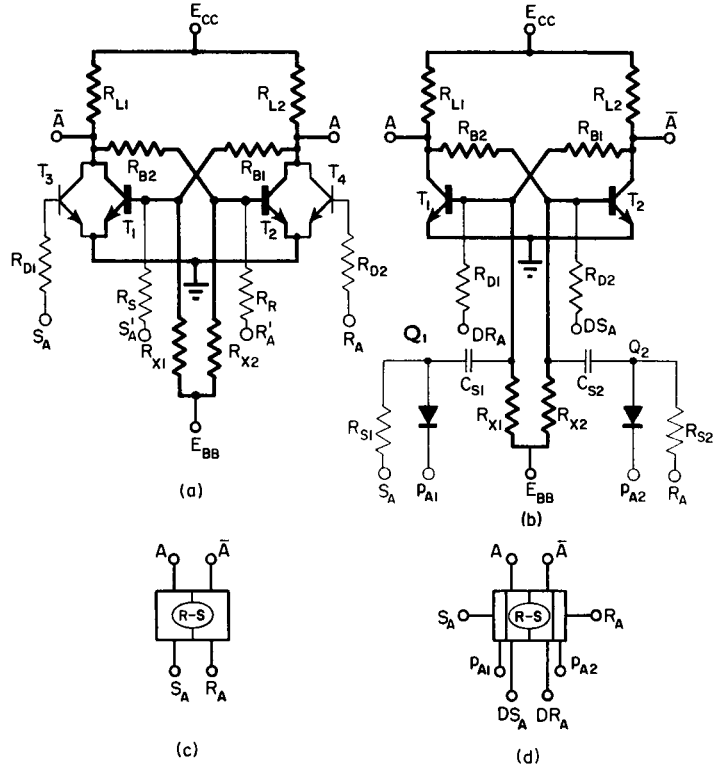


FIG. 5.1. (a) An Eccles–Jordan circuit (in heavy lines) with additional circuits connected to form a d.c. *R–S* flip-flop; (b) an Eccles–Jordan circuit (in heavy lines) with pulse steering circuits to form a pulsed *R–S* flip-flop, together with direct inputs DS_A, DR_A; (c) the symbol for a d.c. flip-flop; (d) the symbol for a pulsed flip-flop.

Triggered and Master–Slave Flip-flops

A bistable circuit is of no interest unless it can be driven from one stable state to another by means of external voltages. Of the many possible drive circuits three are illustrated in Fig 5.1. A circuit with such *logic* inputs is called a *flip-flop* and it is with flip-flops that this chapter is concerned.

Figure 5.1(a) shows a *d.c. flip-flop* and Fig. 5.1(b) shows a *pulsed flip-flop* sometimes called a triggered flip-flop.

Consider the d.c. flip-flop first. In Fig. 5.1(a) two types of logic input are shown. First, transistors T_3 and T_4 are added to form two R.C.T.L. NOR gates with T_1 and T_2 respectively. When the SET and RESET inputs (called *logic* inputs and labelled S_A and R_A) are at (or near) zero volts T_3 and T_4 are both cut-off and the Eccles–Jordan circuit is stable with either A or $\bar{A}$ positive. When S_A is driven positive T_1 is bottomed which forces output $\bar{A}$ to zero volts allowing output A to become positive. Thus, exciting S_A puts the flip-flop in the "1" state with A positive. Similarly exciting only the reset input R_A puts the circuit in the zero state with $\bar{A}$ positive. If *both* S_A and R_A are driven positive together, then both A and $\bar{A}$ go to zero; and when S_A and R_A themselves return to zero, the resulting state is in general indeterminate and the memory function of the circuit is lost. Thus in any sequential circuit care must be taken to ensure that S_A and R_A are not excited together. *A flip-flop which becomes indeterminate when both logic inputs are excited together is called an R–S flip-flop.*† The fact that the circuit becomes indeterminate when both inputs are driven *positive* identifies the circuit of Fig. 5.1a as a *positive logic*, *R–S*, d.c. flip-flop.

†Historically it is probably true that the name *R–S* flip-flop arose from the presence of SET and RESET inputs to the circuit. However, nowadays other types of flip-flops have SET and RESET inputs but do not respond to them in quite the same way as do *R–S* flip-flops. Consequently the name *R–S* has the functional meaning given above.

The second type of logic inputs shown in Fig. 5.1(a) and labelled S_A', R_A' form R.T.L. NOR gates with transistors T_1 and T_2. Their action is the same as the R.C.T.L. inputs and they comprise an alternative form of construction.

Figure 5.1(b) shows a *positive logic, pulsed, R–S flip-flop*. It has two logic inputs S_A, R_A as well as two pulse or trigger inputs p_{A1}, p_{A2}. The voltage levels applied to the pulse inputs are the same as for the logic inputs and the same as outputs A and $\bar{A}$, namely about E_{CC} (6 V, for example) and about zero volts. The pulse input circuits excite the flip-flop when the applied input voltage changes from positive to zero. The diode–capacitor–resistor arrangement which connects a pulse input to the base of a transistor is called a pulse *steering* circuit. Its operation can be understood as follows. (Ignore DR_A and DS_A for the moment.)

Suppose S_A and R_A are both at zero volts, then nodes Q_1 and Q_2 will also be at zero volts so that a negatively going voltage, from +6 V to zero (if $E_{CC} = 6$ V), applied to either p_{A1} or p_{A2} will be blocked by the diodes. But suppose S_A is driven to +6 V, (whilst p_{A1} is held at +6 V) the voltage of Q_1 will also rise (with time constant $C_{S1}R_{S1}$) to about 6 V. Now when p_{A1} is driven rapidly to zero volts the diode will conduct and Q_1 will follow. This negative going step will pass through capacitor C_{S1} to the base of T_1 and will therefore turn T_1 off. Thus when S_A is held at +6 V, output A goes to +6 V, output $\bar{A}$ goes to zero volts and the circuit switches to the "1" state. Similarly if R_A only is driven to +6 V, Q_2 will follow with time constant $C_{S2}R_{S2}$, and now if the pulse input p_{A2} is driven from +6 V to zero, T_2 will be turned off and the circuit switch to the "0" state.

Thus the logic inputs do not *cause* the circuit to switch, they predetermine the direction the circuit will tend to switch next time a negative-going voltage step (from E_{CC} to zero) is applied to the pulse inputs.

If *both* logic inputs are held positive, and a negative-going step is applied to both pulse inputs together, the next state will be

indeterminate since both transistors will be driven off momentarily and this is an unstable condition. This drive condition must therefore be avoided. The possibility of an indeterminate state when both logic inputs are driven positive identifies the circuit as a positive logic, pulsed, *R–S* flip-flop.

It is not, of course, necessary to drive both pulse inputs together. If both logic inputs are held at +6 V but the negative-going step is applied to only p_{A1}, T_1 only will be driven off and the next state will be clearly specified. This kind of pulse gating can be achieved by connecting additional diodes in parallel with the pulse input diodes. If any one of these is clamped to zero volts then a negative going step on p_{A1} (from 6 V to zero) will be blocked.

Thus two quite distinct methods of controlling the penetration of pulses to the transistor are available, namely *pulse steering* by applying logic voltages to S_A or R_A, or *pulse gating* via additional diodes at the pulse inputs.

In addition it is possible to add *direct* inputs to the bases of the transistor through DS_A, DR_A (in Fig. 5.1(b)) and so drive the circuit by voltage *levels* (as in the d.c. flip-flop) rather than by transients.

Evidently it is possible to predetermine the next state of a pulsed flip-flop in a wide variety of ways, and quite complex logic functions can be realized without additional circuitry. Indeed so many permutations are possible that some arbitrary limitations on the available variations are necessary initially. Accordingly in this book we shall be *primarily* concerned with the following techniques.

(1) The logical function, which is to determine the next state of the flip-flop, will normally be applied by the "logic" inputs S_A, R_A (i.e. pulse steering rather than pulse gating will be used).

(2) Pulse inputs p_{A1}, p_{A2} will normally be used as "clock-pulse" inputs and so will usually be connected together.

(3) The d.c. inputs to pulsed flip-flops called *Direct Set* and

Direct Reset (and labelled DS_A and DR_A in flip-flop A) will be regarded primarily as a means of *presetting* flip-flops to the "0" state before the beginning of a sequential operation.

However, examples of the use of separated pulse inputs appear on pp. 138, 143 and of the use of additional logic inputs appears on p. 201.

The graphical symbols to be used to represent the two types of flip-flop are shown in Fig. 5.1(c) and (d). The Set inputs, labelled S_A and DS_A, are always associated with putting the flip-flop into the "1" state, and inputs R_A, DR_A are associated with putting the circuit in the zero state.

It turns out that the circuit of Fig. 5.1(b) is satisfactory only for slow-speed operation, or for binary circuits, for reasons which will now be discussed. As we shall see, it is usual for higher-speed circuits to use more complicated steering circuitry or to construct negative logic flip-flops.

The Design of Flip-flop Circuits

Consider the d.c. flip-flop of Fig. 5.1(a) with R.T.L. inputs S'_A, R'_A. The two states are held stable by ensuring that the conducting transistor is bottomed and that the non-conducting one is sufficiently cut-off. Consider T_1. It must be possible to bottom T_1 by the current flowing through either R_{B1} or the Set resistor R_S. But when one is holding T_1 conducting the other draws current away from the base of T_1 to the collector of a bottomed transistor (either T_2, or a driving transistor in another circuit). Thus eqn. (3.13) becomes

$$\frac{E_{CC}-V_{BE(\mathrm{ON})}}{R_{L2}+R_{B1}} \geqslant \frac{E_{CC}-V_{CE(\mathrm{sat})}}{R_{L1}\beta_{\min}} + \frac{V_{BE(\mathrm{ON})}-E_{BB}}{R_{X1}} + \frac{V_{BE(\mathrm{ON})}-V_{CE(\mathrm{sat})}}{R_S}. \qquad (5.1)$$

Similarly, for T_1 to be cut-off, eqn. (3.14) must become

$$\frac{V_{BE(OFF)}-E_{BB}}{R_{X1}} \geqslant I_{BX}+(V_{CE(sat)}-V_{BE(OFF)})\left(\frac{1}{R_{B1}}+\frac{1}{R_S}\right). \quad (5.2)$$

And $V_{BE(OFF)}$ must be a suitable value to cut T_1 off adequately. With the R.C.T.L. inputs in Fig. 5.1(a) the terms containing R_S are omitted in eqns. (5.1), (5.2).

With the pulse steering circuits added (and the direct inputs renamed DR_A, DS_A) as in Fig. 5.1(b) a transient current may flow in R_{X1} and make the cut-off state more difficult to maintain. This current arises as follows.

When T_1 is cut-off and input S_A is taken from zero to about E_{CC} (the positive logic level), a reactive current charging C_{S1} must flow through R_{S1}, C_{S1} and R_{X1}. But this current raises the voltage of the base of T_1 and tends to take T_1 out of its cut-off state. The initial, and largest, value of this current is about E_{CC}/R_{S1}, so that for T_1 to be held cut-off in the circuit of Fig. 5.1(b)

$$\frac{V_{BE(OFF)}-E_{BB}}{R_{X1}} \geqslant I_{BX}+(V_{CE(sat)}-V_{BE(OFF)})\left(\frac{1}{R_{B1}}+\frac{1}{R_{D1}}\right)+\frac{E_{CC}}{R_{S1}}.$$

For fast switching R_{S1} must be fairly small necessitating a small value of R_{X1}, which in turn requires a small value of R_{B1} (by eqn. (5.1)) and the circuit soon becomes impracticable. (The diodes D_4, D_5 in the modified circuit discussed in Appendix B are included simply to divert this reactive current from R_{X1}.) Note that in the *binary* circuit, described in the next section, T_1 is never cut-off when S_A is taken positive, so this problem does not arise.

The problem of this reactive current can be removed by modifying the steering circuits as shown in Fig. 5.2 to produce a *negative* logic, *R–S* flip-flop. Notice that in Fig. 5.2 the capacitor and diode in each steering circuit have been interchanged as compared with Fig. 5.1(b). Now, since the base of T_1 (say) is always about zero volts, when input R_A is taken positive to about E_{CC} (the more positive logic level), a negative going edge (of

magnitude not greater than E_{CC}) applied to p_{A1} will be barred from reaching the base of T_2 by a reverse biased diode. On the other hand, if R_A is held at zero volts a negative going edge

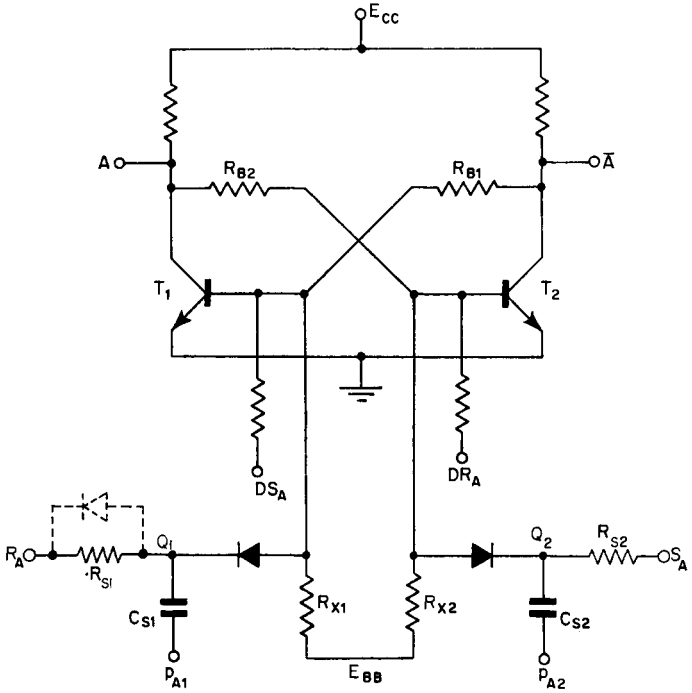


FIG. 5.2. An Eccles–Jordan circuit with pulse steering circuits added to produce a *negative* logic pulsed flip-flop. (Direct inputs also shown.)

applied to p_{A1} is transmitted by the diode to the base of T_1 turning the transistor off.

Thus when the logic input is at the *more negative* of the two logic levels a pulse is allowed to reach the transistor. If *both* logic inputs are at zero together, pulses can reach both transistors so that the resulting stable state may be indeterminate. Thus this circuit is a negative logic, *R–S*, flip-flop. (*R–S* because it has an

indeterminate state, negative logic because this indeterminate state occurs when both logic inputs are at the more negative level, zero volts in this case.) Notice that the *reset* input R_A is now connected to the transistor whose output is labelled A.

The reactive current involved in charging C_S through R_S now flows in the pulse generator, so for the cut-off condition the last term of eqn. (5.3) is replaced by the much smaller reverse current of a diode. The disadvantage of this circuit as compared with the positive logic one of Fig. 5.1(b) is that it is not so easy to gate the input trigger by just adding input diodes. Note too that the direct inputs DS_A, DR_A are still excited by *positive* voltage levels (i.e. with DS_A and DR_A both at zero, no change occurs), so that one type of input uses positive logic, whilst the other uses negative.

Thus, using *npn* transistors, pulsed *R–S* flip-flops using three basic types of pulse steering are available

(1) Positive logic with simple circuitry and pulse gating, but slow-speed working, except for binaries.

(2) Positive logic with more complex circuitry, pulse gating and high-speed working. (Various types of pulse steering are commercially available.)

(3) Negative logic without pulse gating but with simple circuitry and high-speed working.

The complements of these are, of course, available using *pnp* transistors.

Examples of all types are available commercially although the common need for high-speed operation tends to favour alternatives (2) and (3).

For simplicity of explanation this book will continue to discuss *positive logic* flip-flops which require *negative going trigger pulses* (i.e. which use *npn* transistors). It will be assumed that adequate pulse steering circuitry is being used, such as that described in Appendix B.

Finally, the reasons should be understood for always using a negative-going trigger to drive *npn* transistors off rather than a

positive one to drive them into the conducting state. The principal reason is that with the steering diodes reversed in Fig. 5.2 the pulse steering circuits could not work. With additional circuitry this difficulty could be overcome but the second reason would still apply, namely that the transistor when conducting offers a relatively low impedance and so loads the trigger source more severely.

Later we will be considering the use of gates to control the voltage levels reaching the logic inputs. In such cases, particularly when D.T.L. is being used, it is usual to remove the steering resistors R_S and replace them by diode gates, as for example in the circuits shown in Appendix B.

The Binary

In order to illustrate the points made so far in this chapter, the design of the simplest of all sequential circuits will be briefly discussed. This circuit is the *binary*. When driven by a square wave input its output is also a square wave but of only half the repetition frequency of the input.

The Binary Using Pulsed Flip-flops

Using pulsed flip-flops the circuit is constructed as shown in Fig. 5.3(a). The set input S_A is taken to the $\bar{A}$ output and the reset input R_A to the A output. This will ensure that the next triggering edge of the clock pulse waveform will always be steered to the base of the conducting transistor to turn it off. (In the physical circuit this means that the logic inputs of a positive-logic, pulsed flip-flop made of *npn* transistors are cross connected to the opposite outputs, whilst a negative-logic one has its logic inputs connected to the adjacent outputs.)

The circuit will only operate up to a clock pulse repetition frequency of the order of $1/C_S R_S$ because there must be sufficient

time between triggering input edges for the potentials of the steering circuits to approach their correct levels. Thus for rapid switching the delay introduced into the circuit by C_S, R_S must be small.

One way of decreasing the effective time constant $C_S R_S$ is to shunt R_S with another diode as shown dashed in Fig. 5.2. Thus when R_A falls to zero, Q_1 now follows much more rapidly in preparation for the next pulse applied to p_{A1}.

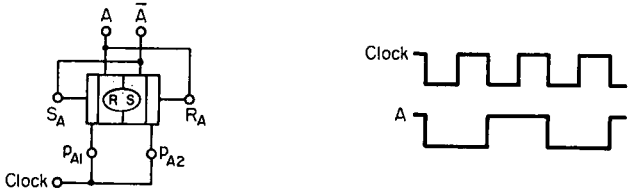


FIG. 5.3. A pulsed flip-flop used as a binary.

However, some delay in the circuit is essential. It is not possible to obtain binary action by connecting the Set and Reset inputs of a d.c. flip-flop back to the A and $\bar{A}$ outputs. This just sets up a half-on, half-off bias condition in both transistors. The delay introduced by C_S, R_S allows a new state at the output of the circuit to be set up before the input steering circuits are altered. Thus the change in the output only affects the *next* input pulse.

C_S and R_S also differentiate the input waveform so that the circuit responds to negative going edges in the input waveform rather than to particular voltage levels. In this it differs from circuits which use d.c. flip-flops.

The Binary Using d.c. Flip-flops

The differentiating and delaying functions of the capacitors C_S in the pulsed flip-flop are not available in the d.c. flip-flop, consequently the circuit must first be designed to respond to voltage

levels rather than to transients, and in addition the delaying function of the capacitors in the pulsed flip-flop must be obtained in some other way.

The delay is most readily achieved by means of a second "slave" flip-flop which receives the setting of the "master" flip-flop when the driving waveform, x, is in the "0" state, say. The setting of the slave is then used to control the master flip-flop when x returns to the "1" state again. Thus the levels, rather than the edges, of the input cause the switching, and the interval between level changes in the input waveform defines the delay. Two versions of this master–slave arrangement of a binary are shown in Fig. 5.4. (Which flip-flop is called the master and which the slave is quite arbitrary.)

Since positive logic d.c. flip-flops are being used, the input x must be inverted so that $\bar{x}$ becomes a positive voltage level which can be used to excite gates when $x = 0$. It can be seen that in both binaries each flip-flop drives the other via gates which are activated by either x or $\bar{x}$.

The equations for the input gates driving the d.c. flip-flops will be derived in Chapter 7. Here they are simply quoted without proof.

$$S_A = x\bar{B} \qquad S_B = \bar{x}A,$$
$$R_A = xB \qquad R_B = \bar{x}\bar{A}.$$

Using AND gates the circuit is constructed as in Fig. 5.4(b). Using NOR gates (which produce the AND function of the complements of the inputs) the circuit is constructed as in Fig. 5.4(a).

Thus binaries can be made from both d.c. and triggered flip-flops. The extra complexity of the d.c. binary brings as compensation a greater inherent switching speed since the only delays are the transistor switching times, and a greater ease of construction in integrated form since the capacitors of the pulsed flip-flops have been eliminated. For maximum speed, however, "speed-up"

capacitors may still be desirable in parallel with the base driving resistors. Alternatively two or three series diodes in place of the resistors are sometimes used.[17]

Modified Flip-flops

Flip-flops which do not respond to their logic inputs in the same way as *R–S* flip-flops can readily be constructed and were

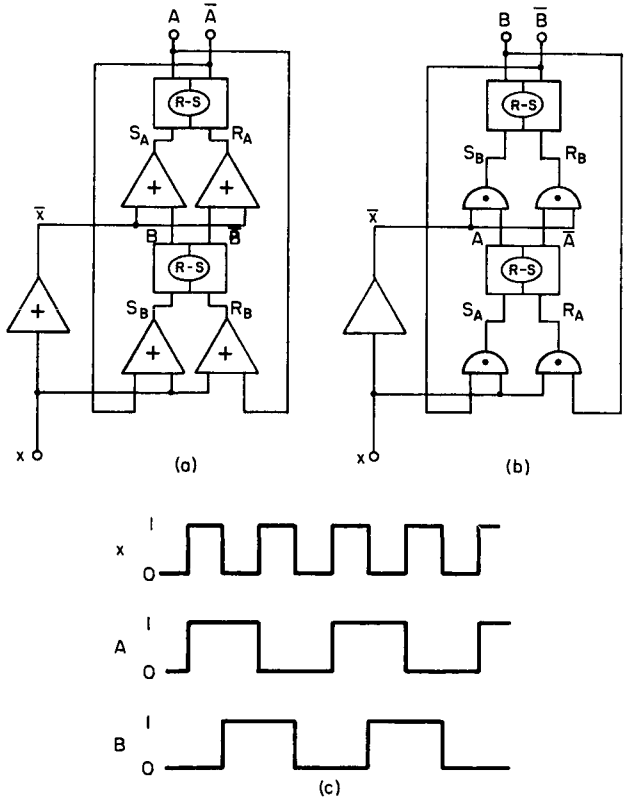


FIG. 5.4. The master–slave binary using two d.c. flip-flops. (a) Together with NOR gates; (b) together with AND gates and an inverter; (c) the circuit waveforms.

first identified by Phister.[18] In theory there are over 100 possible logical arrangements of input and output, but here we will only consider configurations which remove the indeterminate state of the *R–S* flip-flop which occurs when both inputs have been excited. To begin with only pulsed flip-flops will be discussed. Table 5.1 shows the logical function of five different types of pulsed flip-flop. Q^n is the state of the flip-flop after the nth trigger input (or clock pulse) and Q^{n+1} is its state after the $(n+1)$th trigger.

TABLE 5.1. SOME MODIFIED PULSED FLIP-FLOPS

Logic inputs		Q^{n+1}				
		Types of flip-flop				
Set	Reset	*R–S*	*S*	*R*	*J–K*	*E*
0	0	Q^n	Q^n	Q^n	Q^n	Q^n
0	1	0	0	0	0	0
1	0	1	1	1	1	1
1	1	?	1	0	$\overline{Q^n}$	Q^n

Where the *R–S* flip-flop is indeterminate (indicated by the question mark), the "*S*" flip-flop goes to the "1" state, the "*R*" flip-flop goes to the zero state, the "*J–K*" flip-flop *changes* state, and the "*E*" flip-flop stays where it was. For all other inputs the flip-flops are identical.

Two types of pulsed flip-flop with only *one* logic input are also of importance. The "*T*" or Trigger flip-flop is, in effect, a *J–K* flip-flop with its Set and Reset inputs connected together. The "*D*" or delay flip-flop merely adopts the state of its logic input when the trigger pulse arrives. Table 5.2 is the Truth Table for these two flip-flops, *Y* is the single logic input.

It is also possible to construct pulsed flip-flops with *more* than two logic inputs. One of importance is the *R–S–T* flip-flop which

TABLE 5.2

Input	Q^{n+1}	
Y	"T"	"D"
0	Q^n	0
1	$\overline{Q^n}$	1

has the Set and Reset inputs of the R–S flip-flop as well as the input of the "T" flip-flop.

The significance of these various pulsed flip-flops in the design of sequential circuits is discussed later (see p. 180).

With d.c. flip-flops the range of possibilities is much more limited than for pulsed flip-flops unless the flip-flops are taken in pairs in the master–slave arrangement described for the binary. This is partly because it is generally not possible to arrange for the output of a d.c. flip-flop to control its own inputs without danger of oscillation. This means that any of the modified pulsed flip-flops described above which include $\overline{Q^n}$ amongst its $(n+1)$th state cannot be realized in any straightforward way with single d.c. flip-flops. Thus the "J–K" and "T" flip-flop arrangements are excluded. They are, however, easy to construct as master–slave circuits and are readily available in this form (see Chapter 7).

Similarly the Delay (D) flip-flop can only provide a delay equal to the propagation delay of the flip-flop, unless it is made in either the pulsed or master–slave form, or unless a monostable multivibrator is used. Delays as short as the propagation delay of a d.c. flip-flop are often insufficient to avoid timing troubles.

Finally the E flip-flop cannot be made to work reliably, except in master–slave form, since the excitation of both its inputs gives rise to a critical race condition (to be explained later).

Thus the only types available in true d.c. form are the R–S, S and R flip-flops. Others have to make use of the master–slave arrangement, or some other delay mechanism.

The construction of all these various flip-flops is too large a subject to be discussed here in detail, but it is clear that it is always possible to construct them from an R–S flip-flop (d.c. or pulsed) together with combinational circuits on their logic inputs, as indicated in Fig. 5.5.

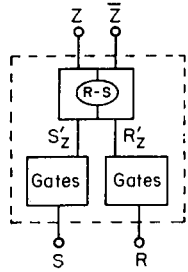


FIG. 5.5. A simple arrangement for constructing modified flip-flops.

The gate circuits shown in Fig. 5.5 have to be designed to obey the following logical equations:

$$\left.\begin{aligned} S' &= S\bar{R} \\ R' &= \bar{S}R \end{aligned}\right\} \text{“}E\text{”} \qquad \left.\begin{aligned} S' &= S\bar{Z} \\ R' &= RZ \end{aligned}\right\} \text{“}J\text{–}K\text{”}$$

$$\left.\begin{aligned} S' &= S \\ R' &= R\bar{S} \end{aligned}\right\} \text{“}S\text{”} \qquad \left.\begin{aligned} S' &= Y\bar{Z} \\ R' &= YZ \end{aligned}\right\} \text{“}T\text{”}$$

$$S' = Y \qquad R' = \bar{Y} \qquad \text{“}D\text{”}$$

Y is the single input formed by connecting S, R together. The inhibit gates of Fig. 4.15 can be used at the logic inputs of the E flip-flop.

If the complements of the inputs are available a J–K flip-flop requires only a NOR gate at each logic input.

A particularly simple pulsed D flip-flop is shown in Fig. 5.6 in which the steering circuits of positive and negative logic flip-flops are combined. But note that the d.c. designs of the two halves differ (see p. 124).

Mixed logic arrangements, along the lines of Fig. 5.5, can also be constructed; that is the central *R–S* flip-flop could be a negative logic one, whilst the logic inputs *S*, *R* to the modified flip-flop use positive logic. The use of these modified flip-flops in sequential circuits is discussed in Chapter 7.

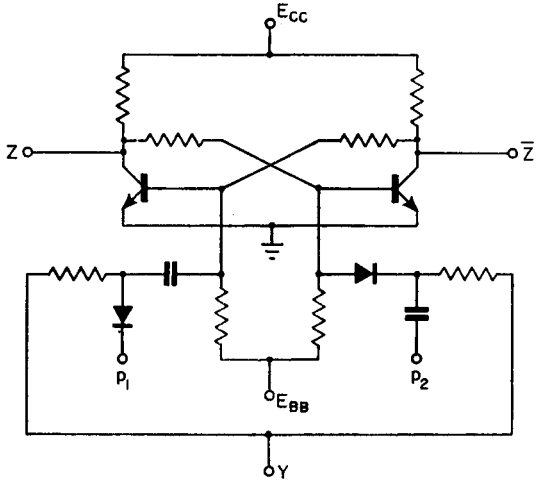


FIG. 5.6. A practical delay flip-flop with positive and negative pulse steering on the transistor bases.

CHAPTER 6

The Design of Sequential Circuits

Introduction

The design of sequential circuits consists primarily of determining what interconnections between combinational circuits, switches and flip-flops are sufficient to achieve some specific sequential function. As with all design problems there are many possible solutions and the problem is to choose the best in a particular application. We have already considered many different types of combinational circuits which differ as to speed, noise immunity, temperature range, cost, etc., and the specification of the required function determines, to some extent, which type of combinational circuit to work with. But even given the type of circuitry to be used there remain many alternative ways of interconnecting them. The purpose of this chapter is, therefore, to describe a systematic technique for arriving at reasonably economic solutions of simple design problems.

The subject is only complex in that there is a large number of possibilities to keep under control, each one being straightforward in itself. The procedures will therefore be introduced through a simple worked example and followed by a more generalized treatment with examples. For a more thorough and comprehensive treatment, the reader is referred to the original papers by Huffman, Mealy and Moore and to some of the larger textbooks on the subject.[18]

The first choice which has to be made in designing sequential

circuits is whether to use pulsed flip-flops, d.c. flip-flops, switches or simply feedback connections as the means by which memory is introduced into the circuit. The properties of each method will emerge as we proceed, but in general it is true that design is simplest using pulsed flip-flops and most complicated using d.c. feedback. We will therefore start with pulsed flip-flops.

It is assumed that the input to any sequential circuit comprises one or more binary signals, to be given the names x_1, x_2, x_3, etc., and that the output consists of one or more binary signals to be given the names Z_1, Z_2, etc. The steps in the process of design are as follows:

(1) Draw the Sequence Diagram (sometimes omitted).
(2) Draw the Sequence Chart or *State* Excitation Map.
(3) Remove redundancy, if any, and merge rows where possible, thus forming the Reduced Excitation Map.
(4) Assign the rows of the Reduced Excitation Map to "Secondary Circuits". This is usually called state assignment.
(5) Compile an *Input* Excitation Map.
(6) Derive minimal Boolean equations defining the inputs to the *Secondary* Circuits.
(7) Compile the Output or Z-map, and obtain Boolean equations for the required output.
(8) Implement the functions represented by the Boolean equations in hardware.

To give some meaning to each of these steps some simple but important examples will be described in some detail.

The simplest sequential circuit is, of course, the binary, which has already been described. It recognizes the sequence of two voltage levels, or changes of voltage, in the input waveform. Here, as an example, we will consider how to interconnect flip-flops, etc., in order to produce a *decade* counter. Counters to a different radix can be designed similarly.

Decade Counters using Triggered Flip-flops.

(a) *Sequence diagram.* In order to count up to ten it is evident that ten distinguishable states are needed so that the *sequence diagram* is particularly simple, as shown in Fig. 6.1(a). The states are numbered 0 to 9 corresponding to the count of the input, and the arrows show the transitions between states.

(b) *State assignment.* The ten states 0 to 9 can be represented by the combinations of not less than four binary digits, i.e. by *four flip-flops*, which we will call *A*, *B*, *C*, *D*. There are many codes which can be used to express decimal numbers in binary digits, three of which are shown on p. 98. With pulsed flip-flops the circuitry presents no design hazards for any kind of code. Accordingly, we will use the most common, namely, the simple binary code, in which the digits of the code have specific weights, 1, 2, 4, 8, and from which decimal numbers can readily be obtained as outputs, if required.

The State Assignment Map, in which states in the sequence are assigned to settings of the flip-flops, is now drawn, as in Fig. 6.1(b). Thus, for example, the seventh input pulse to be counted excites state 7 which appears in the map element *D*, *C*, *B*, *A* = 0, 1, 1, 1 (i.e. the decimal number 7).

(c) *State excitation map.* The state excitation map shown in Fig. 6.1(c) tabulates the *next* states of the circuit to be excited. That is the change from each state which must result from the arrival of the next pulse. In the case of counters this map is somewhat trivial since each decimal number is merely increased by one with respect to the map of Fig. 6.1(b), apart from state 9 which must return the circuit to state 0.

(d) *Input excitation map.* The changes recorded in Fig. 6.1(c) are now interpreted as drive conditions to the flip-flops *A*, *B*, *C*, *D* as shown in Fig. 6.1(d).

In this map the four possible entries (for flip-flops *A*, *B*, *C*, *D* respectively) have the following meanings:

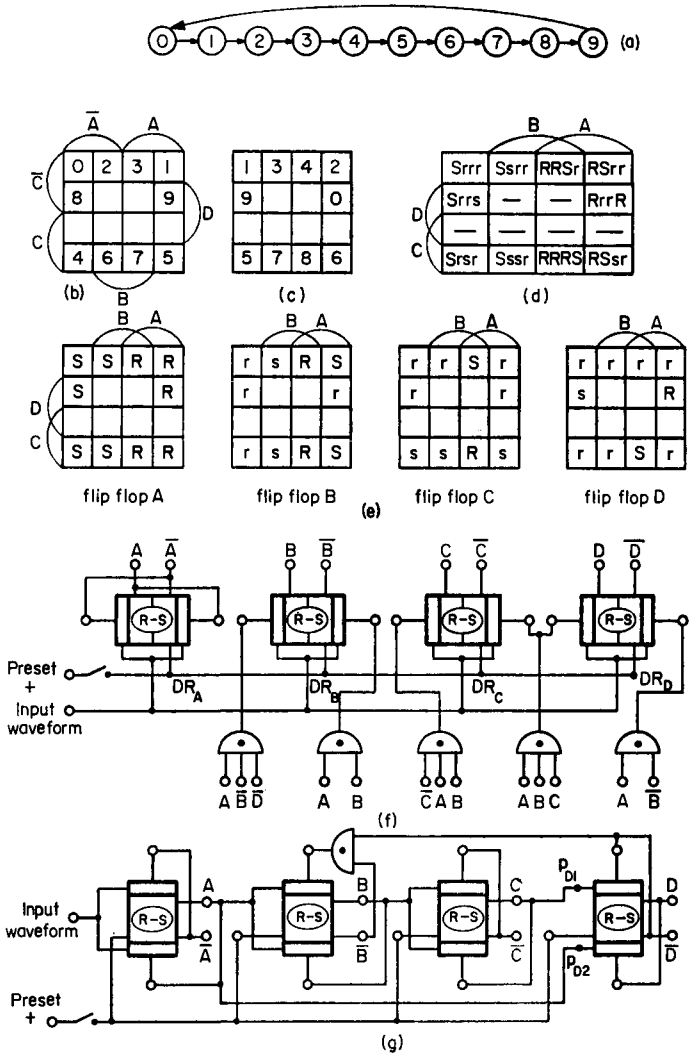


FIG. 6.1. The design of a decade counter. (a) The sequence diagram; (b) the state assignment map in which each state is assigned to a combination of settings of flip-flops *A*, *B*, *C*, *D*; (c) the state *excitation* map showing the next state to be excited for each prior state; (d) the *input*

If the state of the flip-flop *must be changed* the fact is recorded by a capital letter. Capital *S* means that the flip-flop must be *set*, an *R* means it must be *reset*.

If the state of the flip-flop must *not* change, then a choice of drive conditions is possible. One possibility is to excite *neither* input; the other possibility is to drive the flip-flop towards the state it already occupies. (If it is already in the "1" state, the set input can be excited, whilst if the circuit is already in the zero state the reset input can be excited). Since these drive alternatives are optional they are entered as lower-case letters, *s* or *r*.

Thus in Fig. 6.1(d) the first entry *Srrr* means that the Set input of flip-flop *A must* be excited, whilst the reset inputs of flip-flops *B, C, D can* be excited if it is convenient to do so. The entry *Srrr* is deduced from the fact that state 0 which is 0000, changes to state 1 which is 1000. Similarly, the change from state 7 to 8, for example, involves changing all flip-flops simultaneously.

The blank elements in the map apply to states that can never occur.

(e) *The Boolean equations.* For convenience the map of Fig. 6.1(d) is split into four separate maps, one for each flip-flop as shown in Fig. 6.1(e). The entries now specify under what conditions each flip-flop must Set, Reset or stay as it was. To ensure that the flip-flops behave as specified each flip-flop must be driven by gates whose Boolean equations contain the information held by these entries. Thus, grouping the capital *S*'s together with any convenient lower case *s*'s and blanks, and the *R*'s together with any convenient *r*'s and blanks for each flip-flop, the drive

---

excitation map in which the state excitation is interpreted as inputs to the flip-flops; (e) separate excitation maps for each flip-flop from which Boolean equations are derived; (f) the schematic circuit diagram of a synchronous decade counter using four flip-flops and five AND gates; (g) the schematic circuit diagram of an asynchronous decade counter, using four flip-flops and one AND gate.

conditions for each flip-flop become

$$\left.\begin{array}{llll} S_A = \bar{A} & S_B = A\bar{B}\bar{D}; & S_C = AB\bar{C} & S_D = ABC; \\ R_A = A & R_B = AB; & R_C = ABC & R_D = A\bar{B}. \end{array}\right\} \quad (6.1)$$

Circuit Realization

There are two principal ways of realizing this decade counter in practice.

(1) The first involves *synchronous working* in which all flip-flops which change state do so simultaneously.

The counter can be constructed by wiring flip-flop A as a binary, and by connecting a two- or three-input AND gate to each logic input of the other flip-flops, as shown in Fig. 6.1(f) (where the AND gate inputs obey eqns. (6.1) and are connected back to the appropriate outputs). The input pulse train, which is to be connected, is applied to the pulse inputs of all the flip-flops in parallel, as if it were a clock pulse, so that all are triggered together.

This is the fastest circuit arrangement, but it uses more components than are necessary.

(2) Counters using *asynchronous* working can also be made, in which some flip-flops are triggered by the outputs of others rather than by the input pulses.

It can be seen from the maps of Fig. 6.1(e) that apart from two exceptions each flip-flop *changes* when the previous one *resets*. The exceptions to this are: (a) B does not change as A resets when state 9 changes to 0; (b) D resets when 9 changes to 0 even though C does not reset at this transition.

Thus each flip-flop can be connected as a *binary* and each used to drive the next in line, provided the exceptions are dealt with separately, for example as follows:

(i) The Boolean equations for the control of flip-flop B are, from eqn. (6.1),

$$S_B = A\bar{B}\bar{D}, \quad R_B = AB. \quad (6.2)$$

But now, since B is being driven by the resetting of A, it follows that before B can change, A *must* be in the "1" state. Thus eqn. (6.2) can be reduced, without error, to

$$S_B = \bar{B}\bar{D}, \quad R_B = B. \tag{6.3}$$

(ii) By a similar argument, it follows from eqn. (6.1) that flip-flop D can be driven by the resetting of A, provided the logic inputs obey the eqns. $S_D = BC$, $R_D = \bar{B}$.

Alternatively, it can be seen that flip-flop D is reset on the first occasion that A resets following the setting of D. Thus, if pulse input p_{D2} of D is driven by the output of A, and pulse input p_{D1} by the output of C, and if D is connected as a binary, the required behaviour will be obtained.

The circuit for this last type of decade counter is shown in Fig. 6.1(g). It is slower than Fig. 6.1(f) since the circuits change in sequence instead of all together. Furthermore, its design depends upon recognizing a certain pattern of behaviour. However, any counter using the simple binary code should show significant economies in one of these serial forms.

The waveforms appearing at the outputs A, B, C, D of both these decade counters are shown in Fig. 6.2(a). (The slight delays between the instants at which the flip-flops switch in the asynchronous circuit have been ignored.) It can be seen that the counter consists of a binary followed by a scale-of-five counter.

It should perhaps be emphasized that the circuit shown in Fig. 6.1(g) applies to flip-flops which use positive logic and which are triggered by negative going edges, or vice versa. Positive logic flip-flops which are triggered by *positive* going edges (e.g those containing *pnp* transistors) will have their pulse inputs driven by the complemented outputs of the previous stages.

A number of logic systems available commercially use mixed logic. That is the flip-flop is in the "1" state when its primary output is positive (i.e. positive logic at the output) but its logic input is excited by a zero voltage input (i.e. negative logic at the

input). In such cases the AND gate in Fig. 6.1(g), giving $S_B = \bar{B}\bar{D}$, must become an AND gate giving a positive logic input of $S_B = BD$ to ensure the correct logical behaviour. Also in such systems the other Set inputs must be connected to uncomplemented outputs; the opposite of that shown in Fig. 6.1(g). This intermingling of significant logic levels can be very confusing initially, so it will not be discussed further in this introductory book.

By using different decimal codes it is possible to achieve decade (or other) counting using a variety of different circuits. A very common circuit uses the following property of a straight binary counter:

At some point in any straight binary counting sequence, all the flip-flops save the most significant are in the "1" state. At the next pulse the most significant flip-flop changes to the "1" state and all the others switch to "0". Thus, for example, 0111 becomes 1000 at count 8. Now in an asynchronous counter each flip-flop drives the next one to it as it changes to the "1" state, so that the most significant digit changes last (as the pulse ripples through the counter) although the whole sequence is very rapid. This being so the change of the most significant flip-flop can be used to change back again some of the earlier flip-flops, thus in effect omitting some of the numbers in the counting sequence. To turn a four-stage binary counter (normal count of 16) into a decade counter, six counts must be omitted. Consequently flip-flops B and C, whose weights are 2 and 4, must be switched back again as shown in Fig. 6.2(c).

For best operation separate pulse steering circuits for the forward ripple and the feedback are needed, so that the flip-flop can be set again very soon after it has been reset. The logic inputs of the second steering circuits are taken directly to the "1" level, namely E_{CC}, so that any negative going edge reaching its pulse input will cause the flip-flop to go to the "1" state. Thus if the pulse inputs of B and C are connected to $\bar{D}$, then whenever flip-flop D goes to the "1" state flip-flops B and C will also be

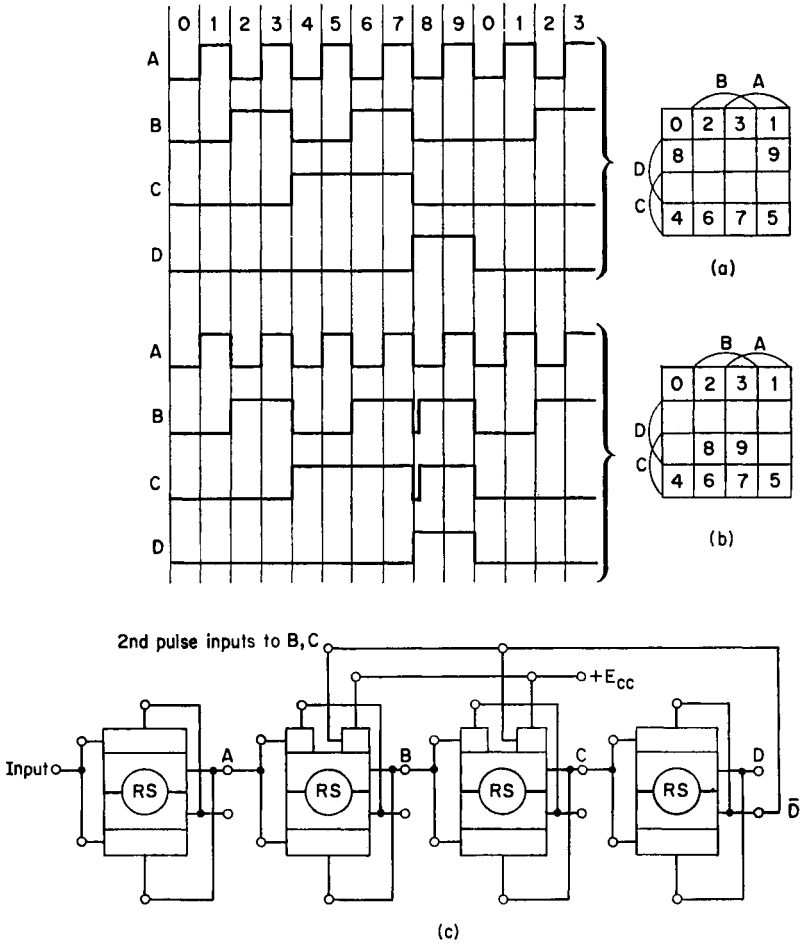


FIG. 6.2. (a) The output waveform of the circuits of Fig. 6.1; (b) the output waveform of the pulse-feedback counter shown schematically in (c). Flip-flops *B* and *C* each have two pulse steering circuits on one side.

driven to the "1" state. In other words after count 7 the state 0111 passes rapidly through 1000 to 1110, in effect omitting counts 8 to 13, as shown in Fig. 6.2(b).

Without the extra pulse steering circuits added to flip-flops *B* and *C* the pulse fed back from $\bar{D}$ must be delayed sufficiently to allow the normal steering circuits to recover. This can be done using simple CR circuits, but a monostable multivibrator gives a more reliable circuit.

The provision of multiple-pulse steering circuits at the inputs to flip-flops, as in the counter described above, evidently opens up another degree of freedom in sequential circuit design. Much more complex counting sequences can be realized in this way.[17] Again, the use of such circuits is beyond the scope of this book.

The only two methods for designing counters which can properly be regarded as general methods are, first, the synchronous counter; this will count to any number in any code and is potentially the fastest. The second is the pulse feedback method which can count to any number, but requires a code translater to obtain any but its own special code. The logic feedback method also described requires the presence of special code sequences for it to operate satisfactorily.

Decade Counters Using Level Sensitive Circuits

With d.c. flip-flops it is the voltage levels that are important, not the negative going edges of the input waveform. Thus, a series of ten pulses must be regarded as a sequence of twenty voltage levels. This rather large number of levels makes the maps and diagrams too big for convenient explanation. Consequently, rather than design a straightforward decade counter, a decade counter comprising a scale-of-five counter followed by a master–slave binary will be designed. This involves all the problems of a decade counter as well as introducing the design of the master–slave

binary referred to earlier (p. 128). (The straightforward decade counter is left as an exercise for the reader.)

Of course, a decade counter using four master-slave flip-flops (i.e. eight d.c. flip-flops) can be designed as for pulsed flip-flops. The counter to be described need only use six d.c. flip-flops, and the straight decade counter needs only five.

The Master-slave Binary Using d.c. Flip-flops

The master–slave binary must complete one cycle of operation in response to two pulses or four voltage levels, x, $\bar{x}$, x, $\bar{x}$. The circuit must consequently have four distinct stable states and four conditions in which it is being driven from one state to another. The sequence diagram of Fig. 6.3(a) again represents the stable states by circles with numbers in, and the transitions by arrows.

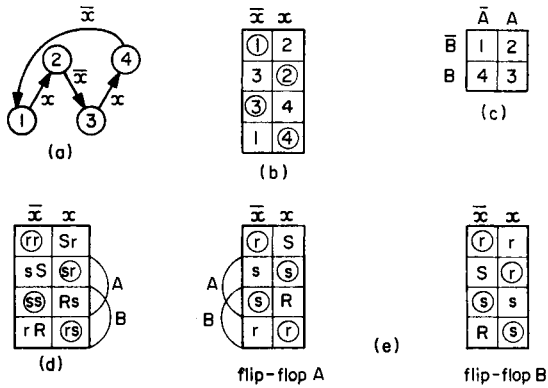


FIG. 6.3. The design of a master–slave binary. (a) The sequence diagram; (b) the sequence map or flow chart; (c) the state assignment map; (d) the input excitation map; (e) the separate input excitation maps, one for each flip-flop.

The information contained in the diagram can be put into the *sequence chart* (sometimes called the flow chart) of Fig. 6.3(b). This chart must be interpreted as follows: stable states are indicated by a circle in the chart. Thus state 1 is stable when the input is $\bar{x}$, but when the input changes to x, the circuit becomes unstable (i.e. it is going to change) and the number entered (2 in this case) is the next required stable state. The unstable state is normally very brief but it cannot be ignored.

With pulsed flip-flops all states were stable and changes of states were brought about by additional input pulses, quite separate from the logic voltage levels, so there were no unstable states mapped on Figs. 6.1 and 6.2. Here, however, the input levels both determine the input logic state and also initiate changes, so that a circuit can be in a particular state (i.e. with a particular pattern of voltages) but can either be stable in that condition, or unstable and about to change.

Since the change $\bar{x}$ to x is horizontal on the map, and since the *circuit state* has not changed until ② has been reached, we now regard a *row of the map* as representing a *state of the circuit* which may or may not be stable. Thus the first row of Fig. 6.3(b) represents a state of the circuit which is stable in column $\bar{x}$, but not stable in column x.

The four stable states of the circuit are now assigned to the four states of two flip-flops, or other two-state device. As we shall soon see, it is not necessary with level sensitive circuits to use flip-flops to indicate significant voltages in a circuit; switches or simply feedback paths may be used, and the term *secondary circuits* is used to embrace all three techniques.

They are called secondary circuits to contrast them with the input circuits, sometimes called primary circuits or inputs. Primary circuits carry the inputs, combinational circuits combine them logically, and secondary circuits are extra connections which have to be made to store or remember the various different states through which the circuit must pass. Thus, more correctly,

the four stable states of the circuit are assigned to the four states of two *secondaries A, B* as shown in the *state assignment map* of Fig. 6.3(c). For example, the row containing ④ is assigned to $\bar{A}B$. Notice that *the inputs to the circuit are used to identify the columns and that the states of the secondaries are associated with the rows.*

In this example we are at present using d.c. flip-flops as secondaries so that using the above assignments together with the sequence map we now draw the input excitation map of Fig. 6.3(d). The capital letters again imply necessary excitations and the lower-case ones optional excitations. The stable states, of course, always involve lower case letters. The input excitation map of Fig. 6.3(d) can now be split into two maps, one for each flip-flop, from which the excitation equations can be derived.

As before, the set input equations are found by grouping the capital S's together with any convenient lower case s's. The reset input equations are found by grouping the R's together with any convenient r's.

$$S_A = x\bar{B}; \qquad S_B = \bar{x}A;$$
$$R_A = xB; \qquad R_B = \bar{x}\bar{A}. \tag{6.4}$$

These are the equations of the master–slave binary described on p. 129 and in Fig. 5.4.

A Binary Using Switches or Feedback Paths

The difference between using switches or feedback paths rather than flip-flops as the secondary circuits A, B in sequential systems is that a switch has to be held in both the ON and the OFF state by an input voltage (or current) whereas a flip-flop will remain in either state after the input excitation has been removed.

In other words, a flip-flop is *bistable* and a switch is not. In practice, since there are only two input levels, it is sufficient to state that a switch will be ON unless it is driven OFF (or vice versa). There is no dangling, midway state.

Thus we can construct a binary using switches, instead of d.c. flip-flops, from the input excitation maps of Fig. 6.3(e) by grouping *all* the S entries together, *whether lower or upper case*, to form the Boolean equations for the switch excitation, since a switch will not stay set if the input is removed. Thus

$$Y_A = \bar{B}x + A\bar{x}, \tag{6.5}$$

$$Y_B = Bx + A\bar{x}. \tag{6.6}$$

Since there is only one input per switch, it is called Y rather than S or R, and since, now, the circuit always resets when the input is not excited it is not necessary to group the reset entries in the map.

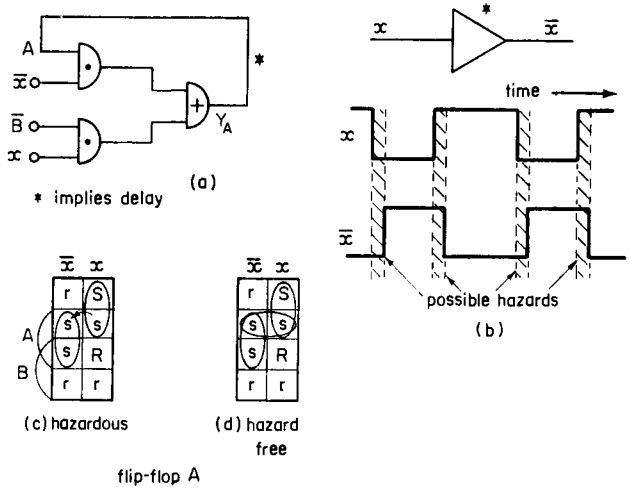


FIG. 6.4. Hazards in sequential circuits. (a) When an output is used to control its own input, delay in the feedback is usually necessary; (b) delay in the inversion of x causing possible hazards; (c) groupings which cause static hazards; (d) an extra group to avoid hazards.

Now although it is convenient to begin with to think of switches with inputs Y_A, Y_B, etc., and with outputs A, B, etc., it is not necessary to do so. The secondary circuit need only be a wire, or d.c. feedback path, with Y_A at one end and output A at the other, as in Fig. 6.4(a). Obviously, for such a direct connection $Y_A = A$.

Figure 6.4(a) is the implementation of eqn. (6.5). A similar circuit would implement eqn. (6.6), and the two together might be expected to constitute a master–slave binary circuit. In practice, however, such a circuit only works some of the time and is generally far too unreliable. This is because the instants at which transients occur in the voltage waveforms rarely occur precisely at the correct times. For example, as illustrated in Fig. 6.4(b) if there is a delay between $x = 1$ and $\bar{x} = 1$ there are brief periods when $x = \bar{x} = 0$ and when $x = \bar{x} = 1$, and during these periods it is probable that the circuit will not function correctly. Timing difficulties of this kind, which can lead to faulty operation of the circuit, are called *static hazards* and must be briefly discussed now since they affect the design of the master–slave binary circuit. We will consider them in more detail again later.

Static Hazards Affecting the Master–slave Binary Circuit

Consider eqn. (6.5), namely

$$Y_A = \bar{B}x + A\bar{x}. \tag{6.7}$$

Suppose that $\bar{B} = A = x = 1$ so that the circuit is in stable state ② (see Fig. 6.3(d)), and $Y_A = 1$ too. Now if x changes to $\bar{x}$, Y_A should remain in the "1" state, by eqn. (6.7). But if, temporarily, $x = \bar{x} = 0$, then for this period $Y_A = 0$. But if $Y_A = 0$, A changes to the "0" state also, so that when eventually $\bar{x} = 1$ it is too late ($A\bar{x} \neq 1$) and the output A has inadvertently changed permanently to the zero state. The possibility of this phenomenon is a *static hazard.*

This particular hazard can be very simply cured, however. Using theorem (13) on p. 89 we can rewrite eqn. (6.5) as

$$Y_A = (\bar{B}+\bar{x})\,(A+x) \tag{6.8}$$

which can be realized using two OR gates and an AND gate instead of the converse. Now the presence of the term $A\bar{B}$ prevents Y_A dropping out when $x=\bar{x}=0$. But, unfortunately, this new arrangement is hazardous when $x = \bar{x} = 1$.

Suppose that $A = 0$ and $B = x = 1$. The circuit is in stable state ④ with $Y_A = 0$. If $x = 1$ changes to $x=0$ (i.e. x changes to $\bar{x}$) Y_A should remain in the "0" state. However, if temporarily $x = \bar{x} = 1$ then temporarily $Y_A = 1$. But if $Y_A = 1$ then $A = 1$ too, so that when eventually $\bar{x}=0$ it is too late ($(A+x)\neq 0$) and the change in Y_A becomes permanent. Thus the circuit has again failed to function properly. This is another static hazard, which arises from the presence of $x\bar{x}$ in the expansion of eqn. (6.8).

Both these hazards can be removed by the same modification. Equation (6.7) contains too few terms, eqn. (6.8) too many. Using theorem (12) on p. 89 we can rewrite eqn. (6.7) as

$$Y_A = \bar{B}x+A\bar{x}+A\bar{B}. \tag{6.9}$$

This avoids both hazards, but includes an extra gate.

Hazards of this type can always be recognized in Karnaugh maps since they involve the transfer from one grouping of map elements to another *without a change of state*, as illustrated in Fig. 6.4(c).

Figure 6.4(c) is a copy of the entries in the input excitation map of Fig. 6.3(e) for flip-flop A, except that the stable states have not been ringed. Treating this as a Karnaugh map, we can group the s's and S's together, as shown by the oval rings or groupings in Fig. 6.4(c), to form the Boolean equation

$$Y_A = \bar{B}x+A\bar{x}$$

as before.

Referring back to Fig. 6.3(b), it can be seen that the two lower case s's on row 2 correspond to a change from stable state ②

to unstable state 3 (when flip-flop B is being set). And this transition is hazardous *because flip-flop A must not change state* as Y_A is held in the "1" state first by $\bar{B}x$ and then by $A\bar{x}$ (i.e. from one oval grouping to the other). The hazard is removed by bridging this hazardous transition with a third (redundant) grouping shown in Fig. 6.4(d), presenting the Boolean term $A\bar{B}$.

Thus the equations for the two secondaries, avoiding hazards from delays between x and $\bar{x}$, become

$$\left.\begin{aligned} Y_A &= \bar{B}x + A\bar{x} + A\bar{B}, \\ Y_B &= Bx + A\bar{x} + AB. \end{aligned}\right\} \tag{6.10}$$

These static hazards are the only hazards which normally affect a binary, although, in general, malfunctioning of the circuit can occur as a result of delays elsewhere in the circuit. Delays between changes in secondary outputs give rise to *critical race conditions*. Delays between changes in the inputs x and changes in the secondaries (which should be coincidental) can also cause trouble. These give rise to *race-hazards*, in particular to the *essential hazard* identified by Unger.[19] Both critical race conditions and race-hazards are discussed more fully in Chapter 7.

Equations (6.10) can be implemented using AND, OR, NOT circuits as shown in Fig. 6.5(a). As we shall see later, to avoid the essential hazard the changes in Y_A and Y_B must always follow the changes in x or $\bar{x}$. The points at which some delay is essential are marked by asterisks. Normally the delay in the cascaded AND–OR gates exceeds that of the inverter producing $\bar{x}$, thus ensuring proper operation.

The circuit shown in Fig. 6.5(b) is the master–slave binary circuit of Fig. 5.4(a) but with the transistors in the two flip-flops drawn as NOR gates. This circuit has four secondaries, A, $\bar{A}$, B, $\bar{B}$ with inputs Y_A, $Y_{\bar{A}}$, Y_B, $Y_{\bar{B}}$, though evidently two secondaries $\bar{A}$, $\bar{B}$ are closely dependent on the other two (or vice versa).

That this circuit obeys the same equations as Fig. 6.5(a) can be demonstrated as follows. The equation for the NOR gate whose

output is $Y_{\bar{B}}$ is evidently

$$Y_{\bar{B}} = \overline{B+\bar{x}A}.$$

Similarly

$$Y_B = \overline{\bar{B}+\bar{x}\bar{A}}.$$

But since $Y_{\bar{B}} = \bar{B}$ (the secondary circuit is a d.c. path),

$$Y_B = \overline{\overline{B+\bar{x}A}+\bar{x}\bar{A}}$$
$$= (B+\bar{x}A)(x+A),$$

or

$$Y_B = xB+AB+\bar{x}A,$$

in agreement with eqn. (6.10). A similar analysis gives the inputs for $Y_{\bar{A}}$ and Y_A.

This illustrates the point mentioned earlier that there are always many ways of realizing any Boolean equation.

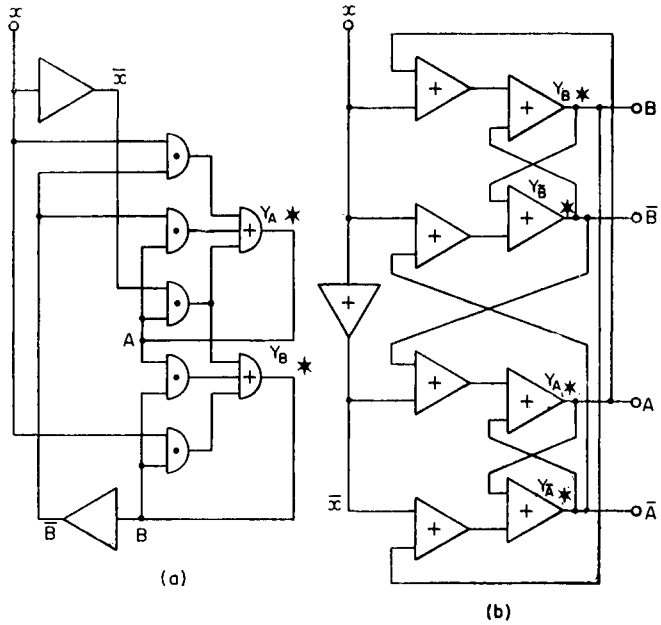


FIG. 6.5. Level sensitive binary circuits. (a) Using AND/OR logic (if all gates and the inverters are replaced by NAND gates the circuit is still correct); (b) the master–slave binary of Fig. 5.4 redrawn as a circuit involving only NOR gates.

To complete the level sensitive decade counter we must now design a scale-of-five counter, first using flip-flops as secondaries.

A Scale of Five Counter Using d.c. Flip-flops

The *sequence diagram* and the *sequence chart* are simply extended versions of those for the binary as shown in Fig. 6.6(a) and (b). There are ten secondary states, in this circuit (the same number as stable states) so four d.c. flip-flops will be needed as secondaries. We will call them *A*, *B*, *C*, *D*.

The secondary states must now be assigned to the settings of these flip-flops in such a way as to avoid *critical races*;† that is, each transition must, if possible, involve the change of only one flip-flop at a time—including the transition from ⑩ to ①. One such secondary assignment is shown in Fig. 6.6(c). Each state number is adjacent in the map to the state which follows it, so that only one secondary changes at a time. Another suitable assignment is the Reflected Excess 3 code given on p. 98.

The state excitation map now appears as in Fig. 6.6(d). In this problem, with four secondaries and one primary input, it is not possible to have a 4×4 Karnaugh map. The simplest method is to draw a separate map for each value of the primary input, x and $\bar{x}$, as shown.

These stable and unstable states are now interpreted as flip-flop input excitations. The four entries in each element refer to the four flip-flops in order. The capital letters refer to necessary excitations, the lower case to optional excitations. Thus in unstable state 10 (entry *srrR*) *A* must remain set, *B* and *C* remain reset and *D* must be reset.

Next, it is convenient to separate out the maps for each flip-flop as in Fig. 6.6(f) from which the Boolean equations for the

† See p. 162 for a full discussion of critical races.

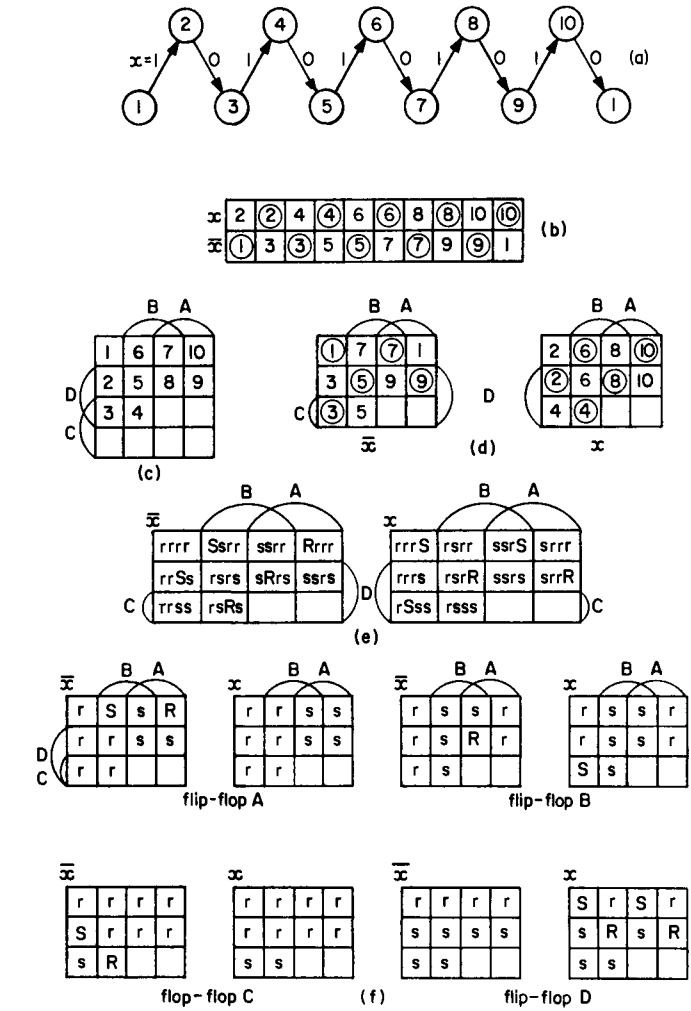


FIG. 6.6. Design of a scale of 5 counter. (a) Sequence diagram; (b) sequence chart; (c) secondary assignment—using stable state numbers to distinguish the states; (d) state excitation map; (e) input excitation map; (f) separate input excitation maps, one for each flip-flop.

flip-flop logic inputs can be obtained by grouping the capital letters with any convenient lower-case ones of the same sort (i.e. S's with some s's, R's with some r's). The equations are:

$$\begin{aligned} &S_A = \bar{x}B\bar{D} \quad S_B = xCD \quad S_C = \bar{x}\bar{A}\bar{B}D \quad S_D = x(\bar{A}\bar{B}+AB), \\ &R_A = \bar{x}\bar{B}D \quad R_B = \bar{x}AD \quad R_C = \bar{x}B \qquad R_D = x(A\bar{B}+\bar{A}B)\bar{C}. \end{aligned} \tag{6.11}$$

Thus the circuit can be realized with an AND gate on the input of each flip-flop except D. D requires two AND gates and an OR gate (or equivalent) on each logic input. (It is inevitable that one flip-flop changes state four times during the cycle, but it is not necessarily flip-flop D.)

No static hazards arise using flip-flops as secondaries so that finally the decade counter is constructed by cascading the master–slave binary with this scale of five counter.

Using d.c. Feedback Paths as Secondaries

The maps of Fig. 6.6(f) can also be used to derive the gates involved when d.c. feedback paths are used. In this case, all the set entries, lower or upper case, have to be grouped together. To avoid static hazards, redundant, overlapping groupings are needed to hold the secondaries in the "1" state when the *circuit* state is changing, but when the *secondary* state is not. Thus the excitation equations for the counter become

$$Y_A = \bar{x}B\bar{D}+\bar{x}AD+xA+AB.$$

(The last term covers the change ⑦, 8, ⑧.)

$$Y_B = B\bar{D}+xB+xC+B\bar{A}.$$

$$Y_C = \bar{x}\bar{A}\bar{B}D+xC+\bar{A}\bar{B}CD.$$

(The last term covers the change ③ to 4.)

$$Y_D = \bar{x}D+x(\bar{A}\bar{B}+AB)+C+\bar{A}\bar{B}D+ABD.$$

(The last two terms cover the changes ② to 3 and ⑧ to 9.)

Finally, for correct operation the changes in inputs x or $\bar{x}$ must always precede the changes in the secondaries.

CHAPTER 7

Generalized Sequential Circuit Design

THE previous chapter introduced some of the procedures which are helpful in organizing a circuit so that it will perform a particular sequential operation. In this chapter we shall generalize these procedures so that they can be used in more sophisticated applications than in the design of counters.

We shall not be involved in mathematical proofs of the validity of these procedures since, once stated, their purpose and validity is clear. The only difficulty associated with sequential circuit design is the handling of all the possible alternatives which are available and of organizing them into a useful array of equations. First, therefore, the various possibilities will be summarized and then the design method will be discussed in detail.

Digit Representation

In general it is not necessary to use voltage levels to represent the two values of a Boolean variable. Two frequencies, two phases of one frequency,[20] two states of a magnetic core, etc., can all be used. The logical design of all these types of circuits is the same, so we will continue to think primarily in terms of two voltage levels.

Secondary Circuits

In sequential circuits it is necessary for distinguishable stable states to exist in order to indicate that a particular sequence of inputs has reached the circuit. The circuits used to indicate these states, or stages in the sequence, are called *secondaries.* Several different types of secondaries can be used, namely:

(a) *Pulsed (or triggered) flip-flops.* These have pulse inputs as well as logic inputs. The pulse inputs incorporate differentiating circuits so that they are driven by steps in the pulse input waveform. The flip-flops only respond to their logic inputs when the appropriate change occurs at the pulse input.

(b) *Master–slave flip-flops.* The step in voltage which will trigger a pulsed flip-flop will also "trigger" a master–slave flip-flop, but since the master–slave circuit is a *level sensitive* flip-flop the step has to be regarded as a sequence of two voltage levels. Thus in effect the differentiating circuit of (a) is replaced by a simple sequence detector. Otherwise master–slave flip-flops are the equivalent of pulsed flip-flops and circuits built from either do not suffer from any form of race-hazards.

(c) *d.c. flip-flops.* These are level sensitive, bistable circuits with logic inputs only. They do not possess the pulse or clock waveform inputs of the previous types. Circuits built from them suffer from races but not from static hazards.

(d) *Switches or d.c. feedback paths.* These are the simplest form of secondary, but the design procedures using them are the most involved. Circuits built from them can suffer from all forms of race-hazards.

It is perhaps worth noting that the larger types of secondary, such as the master–slave flip-flop is itself a sequential circuit with internal secondaries of its own.

The main difference between pulsed or master–slave flip-flops on the one hand, and d.c. flip-flops, switches and feedback on the other, is the necessity to take into account the fact that the

logic inputs of the latter type of secondary also cause the circuit to change state. Thus the design of circuits using d.c. flip-flops or d.c. feedback is concerned with both *stable and unstable states.*

Synchronous and Asynchronous Working

When all the secondaries in a sequential circuit are driven by a clock waveform so that they can only change state at particular well-defined instants, the circuit is a synchronous one. When, however, some secondaries are driven by the changes of state in others (and are consequently a little delayed) the circuit is working asynchronously. Synchronous circuits always use pulsed or master–slave types of secondaries in which the logic is separated from the clocking waveform.

These then are the various different types of sequential circuit which can be built. Their efficient design requires techniques which go beyond those described in the last chapter. For example:

It is not always possible to see intuitively the minimum number of stable states or of secondary circuits needed to achieve a particular function.

It is not always possible to avoid hazards and critical race conditions at the state assignment stage of design.

The output requirements from sequential circuits are not always as simple as those in the previous examples, and the circuit has to be designed to meet them.

Each of these points will now be discussed in turn.

Sequential Circuit Design

(a) *Elimination of Redundant Stable States*

Any two states which are equivalent can be regarded as redundant and removed. For stable states to be equivalent their outputs and their stable and unstable states must be the same. Thus to eliminate stable states the sequence map should be drawn

together with the required outputs. An example is shown in Fig. 7.1. This is an arbitrary map for a sequential circuit with two inputs x_1, x_2 and two output terminals Z_1, Z_2. The map is constructed in a manner analogous to that used earlier for the counter, and we will return to this topic shortly. Here we are simply concerned with redundant stable states, and in this case states ② and ④ are obviously equivalent so that one row of the map or the other can be eliminated. However, the equivalence of states ② and ④ means that stable states ① and ③ are also equivalent. Thus the reduced map of Fig. 7.1(b) results.

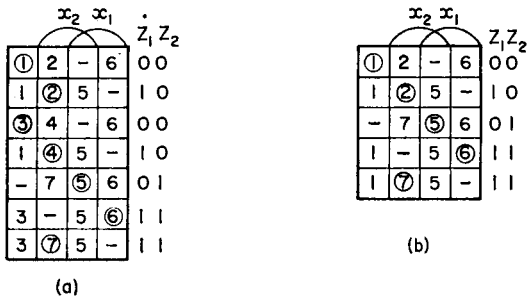


FIG. 7.1. Elimination of redundant stable states. (a) A sequence chart together with the required outputs from each condition; (b) with stable states ③ and ④ eliminated as they are redundant.

In more complex sequence charts less obvious equivalences can exist, and the method for finding the minimum number of stable states can become quite involved. We will not pursue the problem further here, however (see, for example, Marcus[18]).

(b) *Merging*

Merging is a procedure which reduces the number of rows (and secondaries) in a map but *does not* alter the number of stable states. Two rows can be merged if there are *no conflicting state numbers* in the two rows. Stable states merge with unstable ones,

of the same number, and become stable in the merged row. (Blanks in the map may be merged with any state number.)

Outputs from the circuit are not involved in the merging process. That is, two rows can be merged even if their outputs differ. This is because output values, which are determined primarily by the stable states, can be distinguished within one row by differences in the inputs.

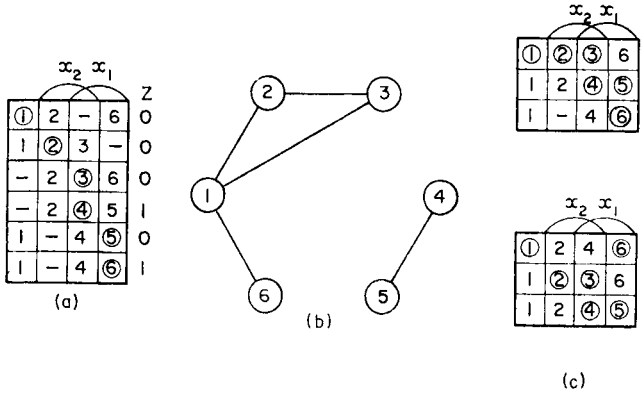


FIG. 7.2. Merging. (a) An example of a sequence diagram; (b) the merger diagram showing those rows (identified by their stable states) which can be merged (output, Z, values are ignored during merging); (c) two possible merged sequence charts.

These points should become clearer on consideration of the example of Fig. 7.2. Figure 7.2(a) shows the sequence chart, and (b) shows the *merger diagram*, on which the rows which can be merged are connected. Three or more rows can only be merged if each merges with every other. Thus ① ② ③ merge but ⑥ cannot merge with them. ⑥ and ① can merge only if ② and ③ are merged separately. Two merged three-row sequence diagrams are possible as shown in Fig. 7.2(c). Merger diagrams are only helpful in the simplest of cases. Intuitive judgement is often needed.

Race-hazards

Actual sequential circuits do not in general behave as ideally as we would like. The outputs of gates are delayed with respect to their inputs; changes in voltages at the outputs of secondaries may not be simultaneous even though their inputs are; inverters or buffers will cause delay, and at the highest speeds even the interconnecting wires may introduce significant delay. Unfortunately these various delays do not simply slow down the operation of the circuit, they may introduce functional errors. How these errors arise, and how they are to be cured, is what we must now consider.

In practice it is convenient to classify the various types of timing problem that can occur, since quite distinct solutions to the difficulties they cause are usually adopted, although in fact all types of timing problem are essentially the same. The classifications are these:

(a) Hazards due to timing differences between the outputs of secondaries are called *races*, and if they are likely to cause faulty operation they are called *critical races*.

(b) Hazards arising from delays between an input x and its complement $\bar{x}$ are called (i) *transient hazards* if they only cause faulty working during the period when $x = \bar{x} = 0$ or $x = \bar{x} = 1$, and (ii) *static hazards* if this delay results in a permanent error in the intended sequence of circuit states.

(c) Hazards arising from any other type of timing difficulty are grouped together under the general heading of race-hazards, a particularly important type being the so-called *essential hazard*.[19]

It is customary to arrange the logic so that all races are removed or rendered non-critical, to add extra circuitry to prevent static hazards causing trouble, and to introduce appropriate delays to handle the rest. Races, transient and static hazards and the others will now be discussed in turn.

Critical Races, Races and Cycles

During discussion of the design of the level sensitive counter in the previous chapter it was noted in passing that it is possible for the circuit to operate incorrectly if two or more secondaries are designed to change state simultaneously, because in practice one will always be faster than the other and an incorrect ordering may bring the circuit to the wrong stable state. If the circuit is designed so that such a phenomenon can occur it is said to possess a critical race condition.

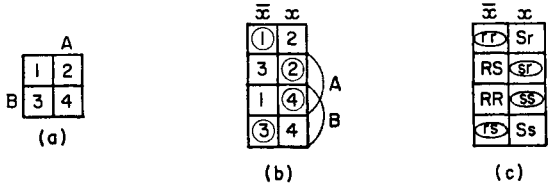


FIG. 7.3. Incorrect secondary assignment for a binary. (c) shows that in two unstable states *both* flip-flops have to be switched, causing a critical race condition.

A simple example of this can be produced by redesigning the level sensitive binary. Suppose the state assignment of Fig. 7.3(a) were adopted instead of that of Fig. 6.3(b). Proceeding with the design as before leads to the input excitation map of Fig. 7.3(c) which now contains a capital letter entry for both secondaries in two elements of the map, indicating that in two places during the sequence both secondaries must change simultaneously.

Thus in state ② an input change from x to $\bar{x}$ leads to A being reset and B being set. If both flip-flops change simultaneously, all well and good. But suppose one flip-flop responds more rapidly to a change in the input signal than the other. If flip-flop A changes first an intermediate state with both A and B reset will be reached which, with $x = 0$, is the stable state ①, so that no

further change will occur. The circuit will "stick" at ① and never change to stable state ③ as it should.

On the other hand, if B changes before A the intermediate state will be with both A and B in the "1" state. But this is not a stable state with input $\bar{x}$ (it is only stable with input x), and subsequently A will reset as required.

This *possibility* of incorrect operation resulting from unequal switching speeds of secondaries is the *critical race condition.*

Critical race conditions can be recognized as follows. If in a map the *columns* are identified by *input conditions*, then a critical race condition exists if

(1) there are two or more stable states in a column, and
(2) an unstable state involves the simultaneous switching of two or more secondary circuits.

Circuits made from pulsed flip-flops do not suffer from race conditions at all because the delay in the pulse steering circuits is made much greater than the switching time of any particular flip-flop, so that in effect all flip-flops change simultaneously.

Evidently the simplest way to avoid critical race conditions is to *assign states in such a way that only one secondary circuit at a time has to change state.* This means that consecutive states must be assigned to adjacent elements of the state assignment map, as in Fig. 6.3(b) but not as in Fig. 7.3(a).

But it is not always possible to assign the minimum number of secondaries so as to prevent two secondaries switching simultaneously. Under these circumstances secondaries should be assigned to produce either *non-critical races* or *cycles* or both. If these two techniques fail too, then additional secondaries must be used.

Non-critical races will not cause faulty operation. They occur when two secondary circuits have to switch simultaneously, but when there is no possibility of the wrong stable state being reached.

Cycles will not cause faulty operation either. They consist of sequences of unstable states which allow the required stable state to be reached without danger of sticking at the wrong one. They are indicated by arrows on a map.

Figure 7.4(a) shows two columns of a simple sequence map with only three stable states, namely $\bar{A}\bar{B}\bar{x}$, $\bar{A}B\bar{x}$ and ABx. The state $A\bar{B}$ is not needed here, but it can cause trouble if not taken care of; it can, on the other hand, be very useful. The map reveals two race conditions. In the first column (with input $\bar{x}$) there are two stable states, and both secondaries must be

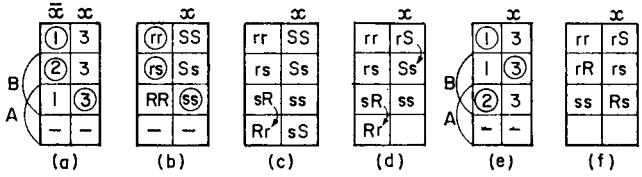


FIG. 7.4. *Races and cycles.* (a) A simple sequence chart; (b) the resulting input excitation map showing two races; (c) solution of the critical race condition by introducing a cycle and a non-critical race; (d) solution using two cycles; (e) and (f) alternative solution by reassignment of secondaries.

reset to get from unstable state 1 to stable state ①. (Stable state ② is reached from other parts of the map—not shown—involving other inputs.)

If A resets first, stable state ② will be reached. This is therefore a critical race condition.

However, in stable state ① when $\bar{x}$ changes to x, both secondaries should be set in order to reach stable state ③. This is not a critical race condition since there is no other stable state in the x column, although the blank state $A\bar{B}$ could be reached and made stable inadvertently by improper grouping.

A solution of both problems without reassigning the secondaries is shown in Fig. 7.4(c). The critical race has been removed

by introducing a *cycle*. Thus when both A and B have to be reset, B is reset first so that another unstable state $A\bar{B}\bar{x}$ is reached. Here A only resets giving $\bar{A}\bar{B}\bar{x}$ without danger of reaching the stable state $\bar{A}B\bar{x}$ on the way.

The uncertainty of the $xA\bar{B}$ state has been similarly removed in Fig. 7.4(c). The x column containing state ③ now shows a non-critical race condition. That is, even though in unstable state $\bar{A}\bar{B}x$ both A and B are set together, there is no possibility of reaching the wrong stable state. Figure 7.4(d) shows the same problem dealt with by introducing another cycle.

Figures 7.4 (e) and (f) show that in fact both problems in this example can in any case be solved by reassigning the secondary circuits. Now each unstable state is adjacent to the stable state to which it is changing.

When neither cycles nor non-critical races can be used to remove critical races, only two possible courses of action remain. Either to introduce delay into the circuit so that the secondaries always change in the proper order to avoid error, or else to introduce additional secondary circuits.

Figure 7.5(a) shows a merged sequence map containing two critical races (3 to ③ and 6 to ⑥) and a cycle. (Any reassignment of secondaries leaves a critical race somewhere in the map.) Consider first the timing problem. In moving from 3 to ③ it is necessary to reset A and set B. If B sets first, stable state ⑥ will be reached and no further change will occur. If A resets first, unstable state 6 will be reached. In this state both secondaries are excited, and whether ③ is reached depends upon whether B sets before A. Since B began to set when A first reset, B is almost certain to set first now and the required state will be reached. Thus by ensuring that A resets before B sets, the circuit will change from 3 to ③ correctly. Similarly, to get from 6 to ⑥ A must set before B. Thus correct operation can be achieved by ensuring that output B is always delayed with respect to A. (By a suitable capacitor or by inserting a monostable multivibrator.)

The second solution is to add another secondary, C. There are many ways of utilizing this additional secondary (see Marcus,[18] chapter 16). The simplest is to transfer one row (the top one, for example) to its corresponding place with $C = 1$ instead of $C = 0$, as shown in Fig. 7.5(c). The empty row in the map now becomes part of a cycle in each column, and the critical race is resolved.

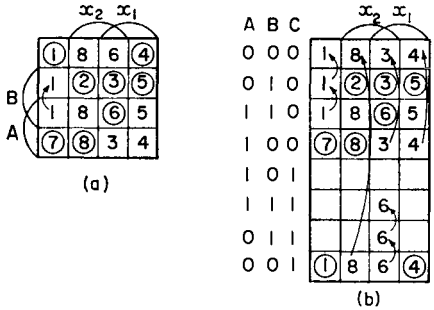


FIG. 7.5. The resolution of a critical race condition using an auxiliary secondary C.

Thus races which arise solely from the relative timings of secondaries can be dealt with by so assigning the secondaries that they never race one another, or by arranging things so that if they do race no trouble results (i.e. the right one always wins).

Transient and Static Hazards

Delays between an input x, say, and its complement $\bar{x}$ can cause a transient or static hazard.

Transient hazards give rise to the temporary appearance of unwanted voltage levels during the period of delay between $x = 1$ and $\bar{x} = 1$. They occur when the following kind of equation applies:

$$Y_A = B\bar{x} + Cx. \tag{7.1}$$

If $B = C = 1$, Y_A should be held in the "1" state continuously as x changes. But if temporarily $x = \bar{x} = 0$ (because $\bar{x}$ is delayed by

the inverter that generates it) then also $Y_A = 0$ temporarily. Y_A will reach its correct value when x does.

Static hazards are transient hazards which lead to a permanent, incorrect voltage level. They arise when d.c. feedback paths (or switches) are used as secondaries and a secondary's input is a function of its own output. Thus if

$$Y_B = B\bar{x} + Cx \tag{7.2}$$

a static hazard is likely to occur. In eqn. (7.2) if $B=C=1$ and x changes to $\bar{x}$ after a little delay, Y_B will fall to the "0" state during the period for which $x=\bar{x}=0$. But with d.c. feedback $Y_B = B$, so that by the time $\bar{x}=1$, $B\bar{x}=0$ and Y_B remains permanently in the "0" state—when it should be being held in the "1" state.

The cure for such hazards, when they occur in such simple circuits as those described by eqns. (7.1) and (7.2), is to add the extra gate according to theorem (12) on p. 89. Thus eqns. (7.1) and (7.2) become

$$Y_A,\ Y_B = B\bar{x} + Cx + BC. \tag{7.3}$$

Now when x changes to $\bar{x}$ the secondaries do not fall to the "0" state.

By considering all values of B, C and x, and all possible transitions of x, it is easy to show that the above transition (when $B=C=1$ and x changes to $\bar{x}$) is the only hazardous one. And, of course, faulty operation will only occur if this transition is indeed part of the intended sequence of states.

If eqn. (7.2) is written as a product-of-sums and constructed from two OR gates and an AND gate or three NOR gates the hazardous transition is a different one. Thus if

$$Y_B = (B+x)(C+\bar{x}) \tag{7.4}$$

and if $B=C=0$ and $x=1$, then $Y_B=0$. But when x changes to $\bar{x}$ there may be a brief period when $x=\bar{x}=1$. If there is then Y_B will change permanently to the "1" state although eqn. (7.4) is supposed to ensure that it remains in the "0" state. This again

is the only transition which causes a static hazard, and as before it can be cured by adding an extra term to the equation. Thus

$$Y_B = (B+x)\,(C+x)\,(B+C). \tag{7.5}$$

By multiplying this out term by term and omitting $x\bar{x}$ eqn. (7.3) can be formed again.

The difference between eqns. (7.3) and (7.5) in practice is that eqn. (7.3) leads directly to cascaded AND–OR circuits or to two cascaded NAND circuits, whereas eqn. (7.5) leads directly to OR–AND circuits or to NOR circuits. Evidently therefore the incidence of hazards depends to some extent on the intended form of circuit realization.

Thus the search for hazardous transitions involves, for sum-of-product equations, finding two states of the circuit (which are not contained in one Boolean product term) for which a particular secondary remains in the "1" state as the input changes, and between which a transition is intended to take place. The cure is to enclose the two states in an extra product term as in eqn. (7.3), so that when $x+\bar{x}=0$, $Y\neq 0$.

For product-of-sums equations the secondary must remain in the "0" state as the input changes. The cure is to group the two states in an extra sum term, so that when $x\bar{x}=1$, $Y\neq 1$.

Using Karnaugh maps this search is quite straightforward. If the map is arranged with columns corresponding to inputs and with rows corresponding to secondary outputs, then changes of input always involve horizontal transitions in the map—from stable to unstable states. If a particular secondary is not to change state, then its state must be expressed by lower case letters in the map: s's for the "1" state and r's for the "0" state. To ensure that a transition is not hazardous its beginning and end must be contained in one grouping or Boolean term. Examples of this method of finding hazards are to be found on pp. 175 and 229.

The reason why transient and static hazards do not occur when d.c. flip-flops are used as secondaries should now be clear. A

flip-flop does not need its set input held in the excited state to prevent it resetting, so the lower case entries in a Karnaugh map have no functional significance. They can be used or ignored according to convenience. Specifically, if eqn. (7.2) were the equation for the set input to flip-flop B, the output of B would not reset when $x=\bar{x}=0$.

Essential Hazards

Even when races and hazards have been appropriately dealt with as just described, timing uncertainties between inputs and secondaries can still cause trouble. To analyse this a detailed transition table can be compiled as suggested by Zissos.[21] The method is best explained by means of an illustration. Consider again the race-free, hazard-free excitation equations for the binary circuit discussed in Chapter 6, namely

$$\begin{aligned} Y_A &= A\bar{x}+\bar{B}x+A\bar{B}, \\ Y_B &= A\bar{x}+Bx+AB, \end{aligned} \tag{7.6}$$

for which the circuit is shown in Fig. 6.5(a).

All the changes in the inputs, gates and secondaries during one complete cycle of the binary are tabulated in Table 7.1, each row in the table representing one stable state.

The transitions between rows in the table which may be hazardous are those in which no change is intended, and for which the *excitation* involves changes in opposite directions. Such hazardous transitions are marked by an arrow. Thus, between rows 1 and 2 the output of gate $A\bar{x}$ could go wrong because at this transition A goes from 0 to 1 whilst $\bar{x}$ goes from 1 to 0. If the output of $A\bar{x}$ is to remain in the zero state it is clear that $\bar{x}$ must change before A. This ordering is also marked in the table. The outputs of gates Bx and AB also must stay in the zero state, but since B does not change state there is no danger of faulty operation.

TABLE 7.1

State	Inputs		Gates					Secondaries	
	x	$\bar{x}$	$x\bar{B}$	$A\bar{B}$	$A\bar{x}$	Bx	AB	A	B
①	0	1	0	0	0	0	0	0	0
					$\bar{x} \downarrow A$				
②	1	0	1	1	0	0	0	1	0
						$x \downarrow B$		$\bar{x} \downarrow \bar{B}$	
③	0	1	0	0	1	0	1	1	1
									$x \downarrow A$
④	1	0	0	0	0	1	0	0	1
			$x \downarrow \bar{B}$						
①	0	1	0	0	0	0	0	0	0
					Y_A		Y_B		

Similarly, working through the table, four other hazardous transitions are to be found. Of particular interest are the transitions in the secondaries. The state of each secondary is determined by the outputs of three gates, as indicated at the bottom of the table ($A = Y_A$, $B = Y_B$). Thus, for A to remain in the "1" state during the transition between states ② and ③, $A\bar{x}$ must change to the "1" state before the fastest of $x\bar{B}$ or $A\bar{B}$ changes to the zero state. Thus, as shown, $\bar{x}$ must change before $\bar{B}$. However, if the gate $A\bar{B}$ were not included, $\bar{x}$ would have to change before x, which is impossible, assuming, as is normal, that $\bar{x}$ is derived from x by means of an inverter. This is the static hazard we noted previously and its presence was the reason for including the gate $A\bar{B}$. Notice, however, that in the absence of the gate $A\bar{B}$, the condition for B to remain in the 1 state between states ③ and ④ is that x must change before $\bar{x}$. Only if long leads delay x but not $\bar{x}$, will x fail to change before $\bar{x}$. So normally the gate AB is superfluous. But without detailed timing information this was not easy to see. With the term AB included, the timing requirement is relaxed and it is only necessary for x to change before A, as shown in Table 7.1.

The conclusion from this table is, that changes in both the input and its complement must precede changes in the secondaries. The only transitions in the circuit where this could not occur are when $\bar{x}$ must precede A, or when $\bar{x}$ must precede $\bar{B}$. Both $\bar{x}$ and the secondaries are delayed with respect to x, so the requirement is that the inverter must be faster than two stages of gating.

More generally it must be concluded that there is an ordering of events which must be obeyed if a sequential circuit is to function correctly. The circumstances in which mistiming of this kind can cause trouble are known as *essential hazards* after Unger's investigations. Unger showed that an essential hazard can always be cured by the inclusion of sufficient delay in the appropriate part of the circuit.

Outputs and the Z-map

In any sequential circuit design the original sequence map should carry the information regarding the output required from each stable state. In the final, merged sequence map these stable states may have become intermingled so it is now necessary to draw another map—the Z-map—which displays the required outputs.

Consider the example of Fig. 7.2(a) reproduced in Fig. 7.6(a). One final merged map with secondary assignment and containing no critical races is shown in Fig. 7.6(b). (The maps of Fig. 7.2(c) both contain critical races, as we can now see.) The outputs shown in Fig. 7.6(a) for each stable state can now be marked on a new map, as shown in bold lettering in Fig. 7.6(c).

We now have to assign outputs to unstable states according to the following principle:

Each unstable state marks a transition from one stable state to another—possibly via a cycle—so to avoid transients in the output the unstable state must not introduce unnecessary changes

in the output (unless the transients are actually required to give outputs.) For example, if two stable states have the same output, so too must the unstable states between them.

Using the notation $Z_{(n)}$ for the output from stable state (n), and Z_n for the output from unstable state n, we can now consider each

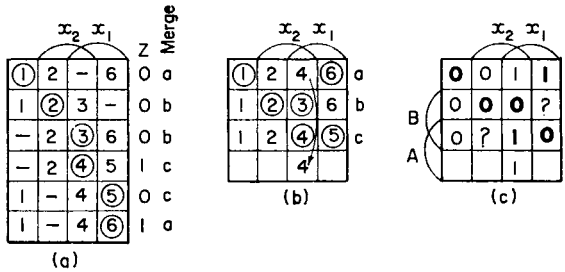


FIG. 7.6. The output and the Z-map. (a) A sequence chart; (b) the merged sequence chart; (c) the output to be obtained from each state of the circuit.

unstable state in Fig. 7.6(b), noting how it arises by inspecting Fig. 7.6(a).

Row 1: 2 comes from ①; $Z_{①}=Z_{②}=0$; so $Z_2=0$.
4 comes from ⑥; $Z_{④}=Z_{⑥}=1$; so $Z_4=1$.
Row 2: 1 comes from ②; $Z_{②}=Z_{①}=0$; so $Z_1=0$.
6 comes from ③; $Z_{③}=0$, $Z_{⑥}=1$; so Z_6 is optional.
Row 3: Z_1 must agree with Z_1 in row 2, but 1; comes from $Z_{⑤}$, so $Z_1=0$.
2 comes from ④; $Z_{④}=1$, $Z_{②}=0$; so Z_2 is optional.
Row 4: Z_4 must agree with Z_4 in row 1.

Thus the map of Fig. 7.6(c) can be drawn. The optional entries are resolved by finding which values give the simplest output groups. Thus with both put to zero and with the secondary assignment shown,

$$Z = x_1(\bar{B}+Ax_2). \tag{7.7}$$

Alternatively, with one optional entry put to "1" and the other to zero, a slightly simpler expression results, namely:

$$Z = Ax_2 + \bar{B}x_1. \tag{7.8}$$

The Master–Slave J–K Flip-flop

An important but simple illustration of some of the techniques described in this chapter is the design of the master–slave *J–K* flip-flop.

The problem is to design a circuit which is driven by a clock waveform, x, in a manner determined by two logic inputs R', S'. If neither logic input is excited the state of the flip-flop remains unchanged when the clock voltage changes; if R' only is excited the circuit goes to the zero state; if S' only is excited the "1" state results and if both S' and R' are excited together the circuit *changes* state.

The following sequence of events will be assumed. That when $\bar{x}$ changes to x a new stable state results but no change in output occurs. That when x changes to $\bar{x}$, the circuit again changes state, followed by a consequent change in output voltage and (since the inputs R', S', are driven by other circuits' outputs) by a change in logic input voltage. *Thus the change x to $\bar{x}$ always precedes the changes of output or of logic input. The change $\bar{x}$ to x only causes the internal state to change.*

Since the circuit has three inputs, a sequence map of eight columns is needed, together with a ninth to specify the output, Z, as shown in Fig. 7.7.

There seems no obvious reason why the circuit should not be stable with any input combination and with either output state. Thus sixteen possible stable states are first marked into the map as shown by the circled numbers in Fig. 7.7(a). It is next necessary to consider each stable state in turn and to mark the sequence of events which can follow.

Starting at stable state ③, for example, $\bar{x}$ changes to x, but no logic inputs change. Thus the state moves, on the same row, to unstable state 7 and then settles to stable state ⑦ with no change

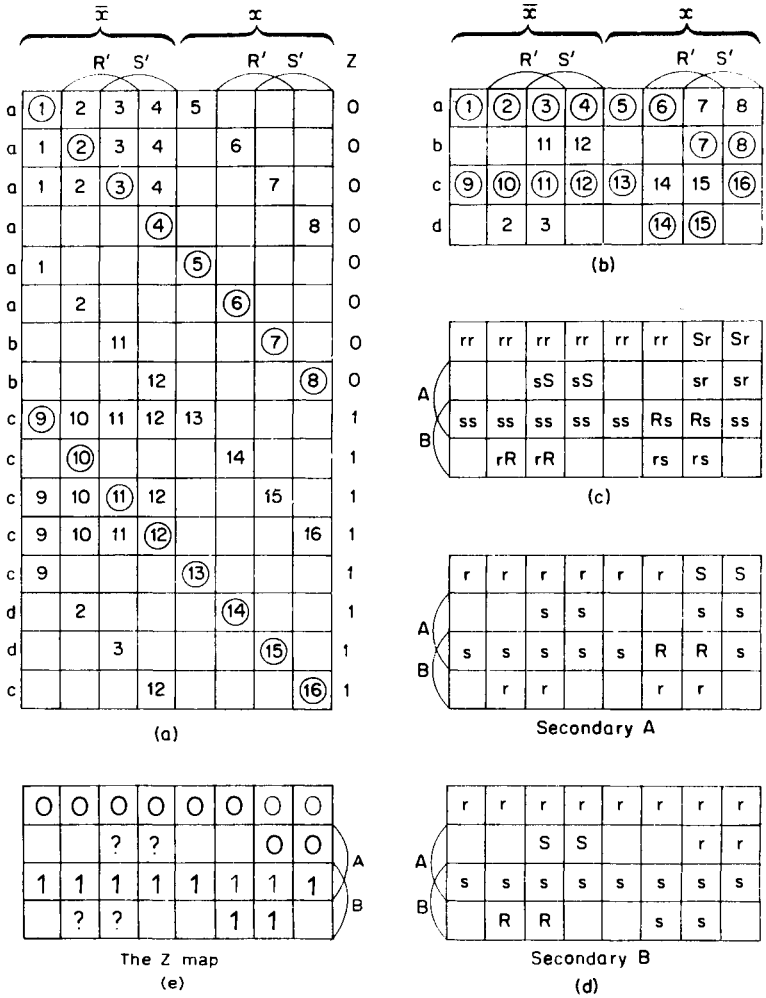


FIG. 7.7. The design of the master–slave J–K flip-flop.

in output. Now when x changes back to $\bar{x}$ the output Z has to change (because both S' and R' are excited in state ③). Thus the next stable state is ⑪, not 3. Finally *after* x has changed to $\bar{x}$, the logic inputs may change in readiness for the next clock wave, so the circuit may move to states ⑨, ⑩ or ⑫ via 9, 10, 12, without further change of output, depending on whether the next logic inputs are respectively $\bar{R}'\bar{S}'$, $R'\bar{S}'$ or $\bar{S}'R'$.

The complete statement of all possible sequences of events is shown in the map of Fig. 7.7(a).

There are no redundant states, but a good deal of merging is possible, and many possible merging patterns exist. The simplest appears to be that resulting in the map of Fig. 7.7(b) (in which the merged rows are indicated by letters beside Figs. 7.7(b) and 7.7(a)).

The assignment of secondaries A, B shown beside Fig. 7.7(c) results in no critical races, so the input excitation map of Fig. 7.7(c) is satisfactory. This can be separated into the two maps of Fig. 7.7(d), one for each secondary.

Finally, the Z-map is drawn as in Fig. 7.7(e). The stable states are indicated in bold type; the unstable states are written lightly. Two unstable states must produce "0" outputs to avoid transient outputs, two must produce "1"s and four can produce either. The simplest solution is, of course,

$$Z = B. \tag{7.9}$$

As regards the secondary excitation, three solutions can be taken from the maps of Fig. 7.7(d).

(a) *d.c. flip-flops* (upper-case letters with lower-case optional)

$$S_A = S'x\bar{B} \qquad S_B = \bar{x}A, \tag{7.10}$$
$$R_A = R'xB \qquad R_B = \bar{x}\bar{A}.$$

(b) *d.c. feedback paths* (all set entries, lower or upper case)

$$Y_A = S'x\bar{B} + A\bar{R}' + A\bar{x} + (A\bar{B}) = S'x\bar{B} + A.\overline{R'xB}, \tag{7.11}$$

$$Y_B = A\bar{x} + Bx + (AB). \tag{7.12}$$

The last terms in each equation are to prevent the static hazards arising respectively from ⑦ or ⑧ transferring to 11 or 12, or from ⑨ or ⑩, moving to 13 or 14.

(c) NOR *gates* (upper-case letters only in complemented pairs together with complement of pair)

$$\begin{aligned} Y_A &= \overline{\bar{A}+R'x\bar{B}}. & Y_B &= \overline{\bar{B}+\bar{x}\bar{A}}, \\ Y_{\bar{A}} &= \overline{A+x\bar{B}S'}. & Y_{\bar{B}} &= \overline{B+\bar{x}A} \end{aligned} \tag{7.13}$$

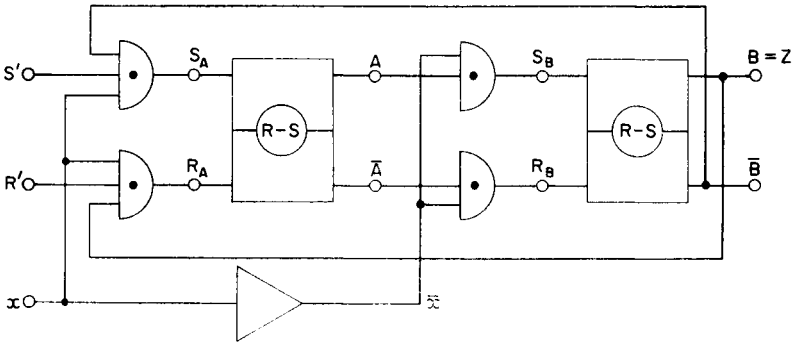
The three resulting circuits are shown in Fig. 7.8. Many variants of each can, of course, be found. The question as to which is the most economical circuit is not considered here. Note that circuit (b) functions properly with all gates drawn as NAND gates. A similar circuit using all NOR gates is, of course, possible (see problem 7.5).

Another example of sequential circuit design is given in Appendix C.

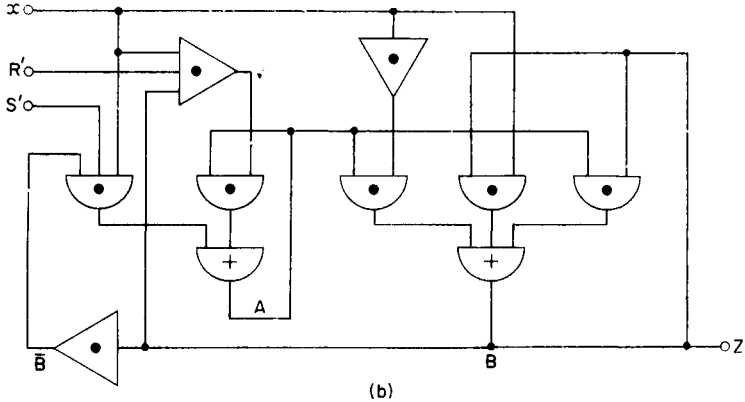
Minimization

The purpose of eliminating redundant rows in a sequence map, and of merging compatible ones, is, of course, to reduce the number of secondaries in a circuit to a minimum, so that hopefully the circuit will cost less to build. Now, as we have seen, secondaries can be triggered flip-flops, master–slave flip-flops, d.c. flip-flops, switches or simply d.c. feedback paths. The cost of each type of secondary depends, of course, on the method of construction, but in general it is evident that the secondaries just listed are in the order of decreasing cost. On the other hand, again as we have seen, the number and complexity of the combinational

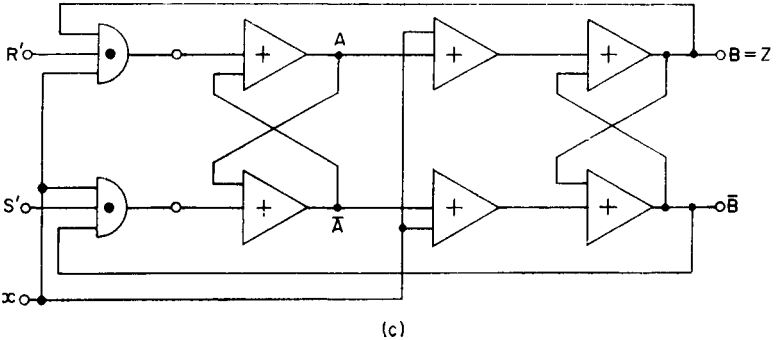
FIG. 7.8. Three versions of the level-sensitive *J–K* flip-flop. (a) A master–slave form using AND gates and two *R–S* flip-flops; (b) with minimal number of d.c. feedback paths as secondaries, using three inverters, five AND gates and two OR gates; (c) with redundant secondaries, $\bar{A}$ and $\bar{B}$, as complements of A and B, using six NOR gates and two AND gates.


S'
R'
x
S_A
R_A
R-S
A
$\bar{A}$
S_B
R_B
R-S
B = Z
$\bar{B}$
$\bar{x}$

(a)


x
R'
S'
A
$\bar{B}$
B
Z

(b)


R'
S'
x
A
$\bar{A}$
B = Z
$\bar{B}$

(c)

circuits associated with the secondaries tends to increase as the complexity of the secondaries decreases, so that economies in the type of secondary used may well result in an overall increase in cost. Whether or not this is the case can always be tested, however, by designing for each type of circuit and by counting the components.

There are more serious problems, however.

(a) By not minimizing the number of secondaries, the number and complexity of combinational circuits can sometimes be considerably reduced.

(b) At the secondary assignment stage in sequential circuit design, the rows of the merged sequence map are assigned to secondaries so as to avoid critical race conditions. It is evident, however, that in general several possible, race-free assignments exist and that they will lead to differing complexities of combinational circuitry.

(c) The avoidance of critical race conditions by ensuring that secondaries change state one at a time is a sufficient precaution: but it is not a necessary one. If two secondaries change state together but do not interact or affect another circuit immediately, then no harm is done (for example, in parallel shift registers; see Appendix C).

There is as yet no clear cut method of arriving at minimal circuits in which the number of secondaries is *not* minimized, which have the optimum state assignment and in which race conditions are allowed in a harmless manner. Indeed, at present, deliberate attempts to minimize a sequential circuit by using these extra freedoms usually results in extra complexity. The guide to circuit economy in general is still largely dependent upon the insight of experienced circuit designers.

Nevertheless, there are aspects of the design process which can be optimized in a useful and determinate manner, given certain premises, two of which will now be discussed. The first concerns the use of modified flip-flops instead of R–S flip-flops

in circuits whose logic has already been settled. The second concerns the minimal realization of combinational circuits using NAND or NOR gates given only uncomplemented inputs (A, B, C, etc., without in addition $\bar{A}$, $\bar{B}$, $\bar{C}$, etc.).

The Use of Modified Flip-flops

By using the modified flip-flops described in Chapter 5 as secondaries in sequential circuits, it is possible to simplify the associated combinational circuits. Even though the flip-flops themselves may be more complex, the economies in the gating arrangements may more than offset this extra cost. How to adapt the Karnaugh maps (to which sequential circuit designs can always lead) for use with modified flip-flops, will now be discussed.

When the input excitation maps obtained in the foregoing examples are separated so that there is one map for each flip-flop (e.g. Fig. 6.6(f)) each element of each map contains only one letter S, s, R or r (or blank) which means that either one logic input or the other or neither is excited. The presence of only one entry per element of the map ensures that the indeterminate state of the R–S flip-flop (with both inputs excited) cannot occur. But each of the modified flip-flops described on p. 131 will give a specific output with *both* inputs excited, so it is possible to modify the maps in accordance with the properties of each modified flip-flop and thus make use of these properties. In general the procedure only involves substituting a double-letter entry, SR, in place of some of the single-letter entries, implying that both logic inputs of the flip-flop should be excited together. The advantage of so doing is that the groupings of letters in the map, to form Boolean equations may be simplified. We will consider each type of modified flip-flop in turn, but it should be recalled that the J–K, T and E flip-flop in d.c. form require master–slave configuration.

1. *Flip-flops with Two Logic Inputs*

(a) *The J–K flip-flop.* The *J–K* flip-flop *changes* state (in either direction) when both inputs are excited together. This means that any upper-case letter (*S* or *R*) in the map can equally well be written as a double-letter, *RS*, entry implying the excitation of both inputs. The simplification which can be achieved depends upon the type of map involved.

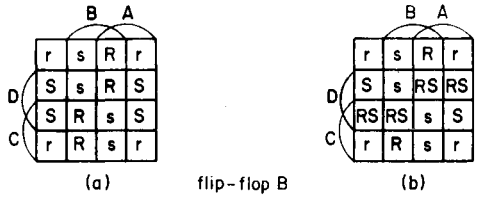


FIG. 7.9. An input excitation map. (a) For an *R–S* flip-flop; (b) modified for a *J–K* flip-flop.

Figure 7.9(a) shows a map to which the *J–K* flip-flop is especially suited, and eqns. (7.14) show the flip-flop excitation for this map when *R–S* flip-flops are used:

$$\begin{aligned} S_B &= \bar{B}D, \\ R_B &= \bar{A}BC + AB\bar{C}. \end{aligned} \tag{7.14}$$

The optimal modifications to this map for a *J–K* flip-flop are shown in Fig. 7.9(b) in which two *R*'s and two *S*'s have been changed to *RS*'s. The groupings of *S*'s and *R*'s are now simpler and lead to the simpler Boolean input excitation expressions of eqns. (7.15).

$$\begin{aligned} S_B &= D, \\ R_B &= \bar{A}C + A\bar{C}. \end{aligned} \tag{7.15}$$

For *J–K* flip-flops to lead to a simplification in the excitation equations, as compared with an *R–S* flip-flop, the capital letter

entries in the original map must be in horizontal and vertical arrays and intermingled with lower case entries. Without the lower case entries the *T* flip-flop achieves greater saving than the *J–K* flip-flop (see below, p. 182).

(b) *The E flip-flop.* The *E* flip-flop remains in its previous state if *neither* logic input is excited or if *both* are excited. Thus each *lower*-case letter in a map can be replaced by an *SR* entry without changing the function of the circuit.

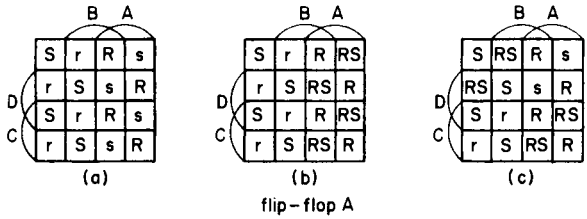


FIG. 7.10. An input excitation map. (a) For an *R–S* flip-flop; (b), (c) modified for an *E* flip-flop.

Simplification can often be achieved by changing some of those lower case entries in a map which are in any case included in the groupings for the set or reset input equation. The *E* flip-flop offers the greatest degree of simplification with a map of the kind shown in Fig. 7.10(a). The diagonal arrays of *S*, *R* entries which are characteristic of Exclusive OR logic also suggest the use of *E* flip-flops. The equations for an *R–S* flip-flop are

$$S_A = \bar{C}(B\bar{D}+BD) + C(\bar{B}D+B\bar{D}), \tag{7.16}$$

$$R_A = \bar{C}(B\bar{D}+\bar{B}D)+C(\bar{B}\bar{D}+BD). \tag{7.17}$$

With some lower-case entries rewritten as *RS*, as in Fig. 7.10(b), the excitation equations become

$$S_A = \bar{C}(B\bar{D}+BD)+C(\bar{B}D+B\bar{D}), \tag{7.18}$$

$$R_A = A, \tag{7.19}$$

or alternatively, as in Fig. 7.10(c), they become

$$S_A = \bar{A}\bar{C} + C(B\bar{D} + \bar{B}D),$$
$$R_A = AC + \bar{C}(B\bar{D} + \bar{B}D).$$

It should also be noted that the *E* flip-flop is its own inverse so that it can easily be used in negative, positive or mixed logic systems, and that it can both equalize and reduce the length of the set and reset Boolean equations.

(c) *The S flip-flop and the R flip-flop.* The set biased or reset biased flip-flops each offer half the advantage of the *J–K* flip-flop. The *S* flip-flop will set if both inputs are excited or if only the set input is excited. Thus any capital *S* entry in a map can be rewritten as an *SR* entry. Similarly, for the reset biased flip-flop any *R* entry can be rewritten as *RS*.

2. *Single-input Flip-flops*

In a similar manner it is possible to read the excitation map to determine the inputs to the "*T*" and "*D*" flip-flops.

The T flip-flop. The *T* flip-flop has only one logic input, and the circuit always changes state when this input is excited. Thus if *Y* is this single input any capital letter entry, either *R* or *S*, can be rewritten as a *Y* input to a *T* flip-flop.

Figure 7.11(a) shows the kind of circuit for which the *T* flip-

FIG. 7.11. An input excitation map. (a) For an *R–S* flip-flop; (b) for a *T* flip-flop.

flop has most to offer. Its revised form for T flip-flops is shown in Fig. 7.11(b).

$$\left.\begin{aligned} S_A &= \bar{A}B\bar{D}+\bar{A}BD \\ R_A &= A\bar{B}\bar{D}+ABD \end{aligned}\right\} \quad R\text{–}S \text{ flip-flop,} \tag{7.20}$$

$$Y = BD+\bar{B}\bar{D} \qquad T \text{ flip-flop.} \tag{7.21}$$

Notice that a T flip-flop can be constructed by connecting the two inputs of a J–K flip-flop together, though there are cheaper ways.

The Delay Flip-flop (D)

The D flip-flop has the same logical behaviour as an ideal switch (i.e. its output becomes a replica of its input) except that any output change involved is delayed until the voltage level of the clock pulse changes. Thus every set entry (S or s) in the map must be regarded as a necessary input to it, and be grouped to form the Boolean excitation equations for the single input, Y, to the flip-flop.

The D flip-flop is ideal for shift registers (see p. 195) and offers gating economies when the set entries are conveniently grouped and when the lower case ones would in any case be grouped with the upper case ones. Figure 7.12 shows an example of a map for which the D flip-flop is particularly suitable. The Boolean equation for the Reset input to the R–S flip-flop is simply eliminated.

$$\left.\begin{aligned} S_A &= \bar{A}B+\bar{C}D \\ R_A &= AC+A\bar{D} \end{aligned}\right\} \quad R\text{–}S, \tag{7.22}$$

$$Y = \bar{A}B+\bar{C}D. \tag{7.23}$$

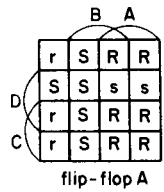


FIG. 7.12. An input excitation map suited for use with a D flip-flop.

Similar techniques can be deduced for interpreting the Karnaugh maps as input excitation to other types of modified flip-flops, such as the *R–S–T*, but these will not be investigated further here.

This section has shown that by selecting suitable modified flip-flops, economies in combinational logic circuits can be realized. In more complex systems than the examples shown here, the economies can be considerable. As to which are in practice the most generally applicable is not yet known. The penalty for choosing a modified flip-flop is that it may be more expensive. Nowadays, however, *J–K* flip-flops are being widely used and the price differences are becoming marginal.

Finally, it should be pointed out that any two-input flip-flop can always simply replace an *R–S* flip-flop, without changing the input logic. But the converse, obviously, is not normally true.

Map Factoring

When using NOR or NAND logic advantage can sometimes be taken of their inverting properties in order to reduce the number of gates involved in realizing a particular Boolean function.

Suppose inputs A, B and C are available and the following Boolean function is to be realized using NOR gates only:

$$Z = BC + \bar{B}\bar{A} = (B + \bar{A})(\bar{B} + C). \tag{7.24}$$

Using the method described on p. 97, five NOR gates will be required as shown in Fig. 7.13(a). A similar circuit using five NAND gates can also be constructed. In either case two gates are being used simply as inverters to complement A and B, and it might be asked whether a different arrangement could be devised which uses fewer gates whilst still using only *uncomplemented* inputs. In fact such an arrangement is possible, as shown in Fig. 7.13(b). One method of deriving this circuit is known as *map*

factoring, and will now be described. (For more detailed studies of this method see Maley and Earle.[18])

The procedure for design using NOR gates is not quite the same as that using NAND gates, but the similarity is so great as to be quite confusing when learnt together; consequently the techniques to be described will involve only NAND gates in the map

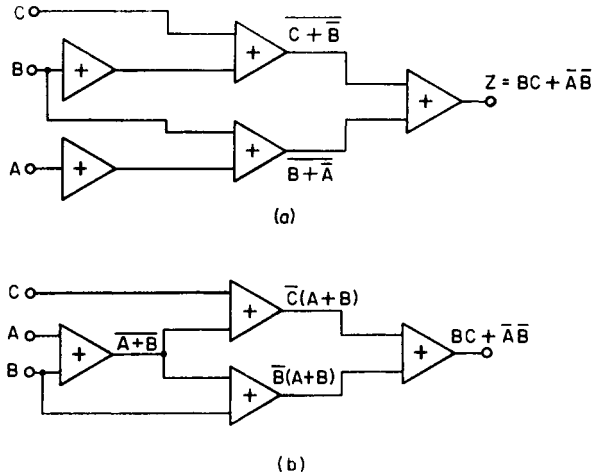


FIG. 7.13. (a) A simple method of realizing the function $Z = BC + \bar{A}\bar{B}$ using NOR gates; (b) a more efficient circuit.

factoring stage. The adaptation of the *same* technique to the design of NOR gate circuitry will then be explained.

The fundamental "element" of the minimized NAND gate circuit is a three stage circuit, an example of which is shown in Fig. 7.14(a). In this circuit the output is

$$Z = C\bar{A} + C\bar{B} = C(\bar{A} + \bar{B}) = C.\overline{AB} \qquad (7.25)$$

and is shown by the 1's in the map of Fig. 7.14(b). But this pattern of 1's can be regarded as the whole group C inhibited by the smaller grouping AB, as indicated in Fig. 7.14(b). The grouping C is shown dashed and the inhibiting function is shown by

crosses. This idea of expressing Boolean functions as larger groupings of uncomplemented inputs (C in this case), inhibited by smaller groupings of uncomplemented inputs, is at the heart of the map factoring method.

The circuit is then built up from this *inhibiting* interpretation by

(a) applying the *inhibiting* function (which must be in product form) to the first stage of the three stage circuit;

(b) applying the *inhibited* function (which must again be in product form) direct to the second stage;

(c) taking the required output from the third stage.

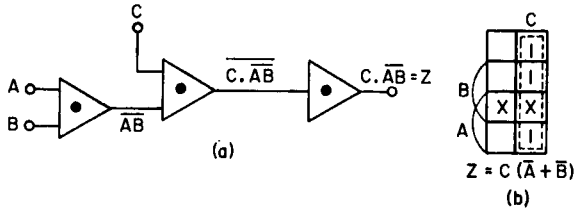


FIG. 7.14. The use of a NAND gate as an inhibitor in the realization of Boolean functions with NAND gates. Thus Z is C inhibited by AB.

In so simple a problem the third stage is serving only as an inverter. In more complex circuits it serves to combine several "factored" parts of the map to form the required overall output.

The technique therefore consists of factoring the required function into parts each of which can be realized by a function inhibited by another function, *both of which are, as far as possible,* AND *functions of uncomplemented inputs.*

Consider the example of Fig. 7.15. The map of Fig. 7.15(a) shows the Boolean function

$$Z = A\bar{B}\bar{C} + BD\bar{C} + \bar{A}CD + A\bar{B}\bar{D}.$$

Using the technique described on p. 96 this function would require nine NAND gates, four of which are used simply to complement inputs A, B, C and D.

This function can be "factored" as shown in Fig. 7.15(b), by picking out those parts of the function which can be synthesized as just described. Thus

$P = A$ inhibited by ACD and by B,

$Q = BD$ inhibited by ACD,

$R = CD$ inhibited by ACD.

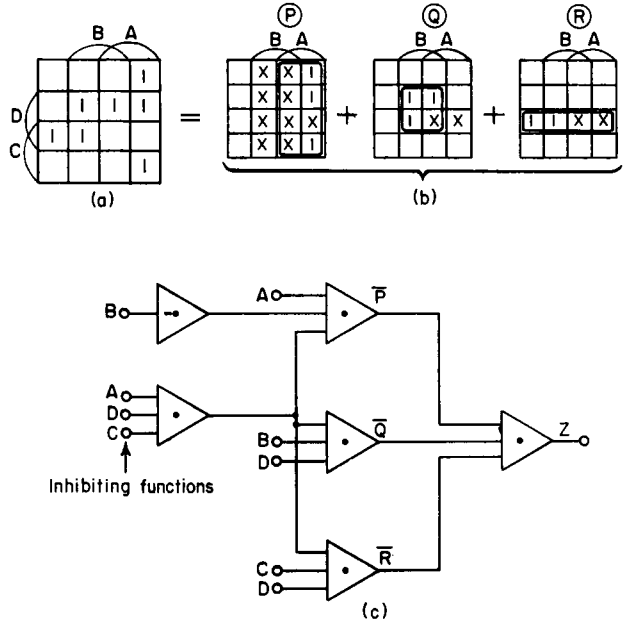


FIG. 7.15. Map factoring to realize a particular function Z using only NAND gates.

Both the inhibiting functions and the inhibited functions are AND functions of uncomplemented inputs. Note that in this case the inhibiting function ACD is common to each part of the circuit. This, of course, is exceptional, but it serves to illustrate the value of finding such functions.

It is now clear that the circuit can be built up as in Fig. 7.15(c) since the final (third stage) NAND gate forms the function $\overline{P\bar{Q}\bar{R}} = P+Q+R$. This uses only six NAND gates.

NOR *Gate Circuits*

Not all Boolean functions are as convenient to synthesize using NAND gates as this one is. This point can be illustrated by synthesizing the same function using NOR gates.

Since NOR gates are the complements of NAND gates it is evident that a NOR gate circuit with output Z can be formed by

(a) synthesizing the complement of the required output (i.e. $\bar{Z}$) using NAND gates and only *complemented* inputs;
(b) inverting all inputs and changing all NAND gates to NOR gates.

The output will now be Z.

Consider again the function expressed in the map of Fig. 7.15(a). The complement of Z is shown in the map of Fig. 7.16(a).

We now have to factor this map using only AND functions of *complemented* inputs, inhibited if necessary by further AND functions of complemented inputs. This is shown in Fig. 7.16(b).

$$P = \bar{D} \text{ inhibited by } A\bar{B},$$

$$Q = \bar{A}\bar{B}\bar{C},$$

$$R = ACD.$$

This time it is not possible to achieve the ideal functions. The inhibitor of P contains an uncomplemented input, whilst R contains three uncomplemented ones. The reason for this is, of course, that R contains the element $ABCD$, and so no function of complemented inputs can contain it, inhibited or not.

It is also worth noting that R could be regarded as CD inhibited by $\bar{A}$, or as AD inhibited by $\bar{C}$ or as AC inhibited by $\bar{D}$. But these are all the same in their use of gates as is ACD. This illustrates the point that efficient inhibiting functions involve multiple inputs to the inhibiting gates.

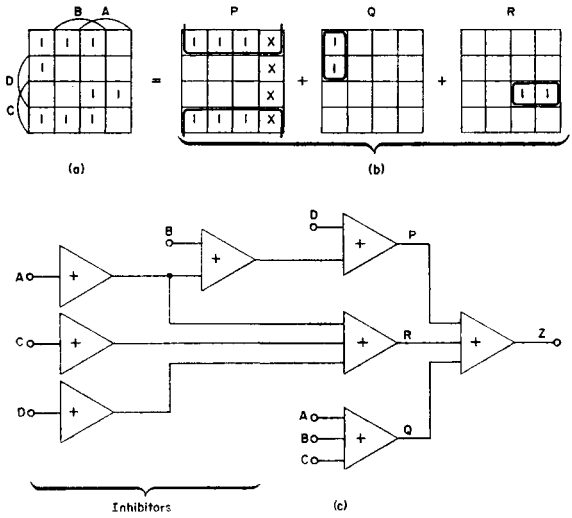


FIG. 7.16. Map factoring of $\bar{Z}$ to achieve the same function, Z, as in Fig. 7.15, but using only NOR gates (see text).

To implement this design using NOR gates, the circuit is now drawn *as if* NAND gates were involved; but the gates are labelled NOR gates and all the inputs are inverted so that they are now in uncomplemented form. This is illustrated in Fig. 7.16(c). This time the circuit requires eight gates, a saving of only one compared with the direct method of design.

The worst Boolean term from the point of view of NAND gate construction is $\bar{A}B\bar{C}\bar{D}$ on its own (i.e. adjacent to no other terms). The worst for NOR gate construction is $ABCD$ adjacent to no

other terms. In either case no reduction of gates below the number involved in the direct design method is possible.

Finally, it is left as an exercise to the reader to verify first that the circuit of Fig. 7.13(b) is an implementation of eqn. (7.24), and second to realize eqn. (7.24) using NAND rather than NOR gates.

Problems

7.1. Design a decade counter using Reflected Excess 3 code and (a) using pulsed R–S flip-flops, (b) using pulsed E flip-flops to remove fan-in's of 3 or more.

(Ans.: (a) $S_A = \bar{B}D+BC\bar{D}$, $S_B = AD$, $S_C = \bar{D}$, $S_D = \bar{A}\bar{B}$,
$R_A = \bar{B}\bar{D}+BD$, $R_B = A\bar{D}$, $R_C = \bar{A}BD$, $R_D = \bar{C}$.
(b) $S_A = \bar{A}C$, $S_B = AD$, $S_C = \bar{D}+A$, $S_D = \bar{A}\bar{B}$,
$R_A = \bar{B}\bar{D}+BD$, $R_B = A\bar{D}$, $R_C = BD$, $R_D = \bar{C}$.)

7.2. Design a scale of 3 counter using level sensitive circuits (i.e. to detect the sequence $\bar{x}$, x, $\bar{x}$, x, $\bar{x}$, x) in which the state of assignment of the three secondary circuits are (for states 0 to 5) 000, 001, 011, 010, 110, 100; (a) using d.c. flip-flops, (b) using d.c. feedback paths as secondaries.

(Ans.: (a) $S_A = \bar{x}B\bar{C}$, $S_B = \bar{x}C$, $S_C = x\bar{A}\bar{B}$,
$R_A = \bar{x}\bar{B}$, $R_B = xA$, $R_C = xB$.
(b) $Y_A = AB+xA+\bar{x}B\bar{C}$,
$Y_B = \bar{x}C+\bar{A}B+\bar{x}B$,
$Y_C = x\bar{A}\bar{B}+C\bar{B}+\bar{x}C$.)

7.3. Repeat problems 7.2 with the following assignment of secondaries; 100, 110, 111, 011, 001, 101.

(Ans.: (a) $S_A = x\bar{B}$, $S_B = x\bar{C}$, $S_C = \bar{x}B$,
$R_A = xBC$, $R_B = \bar{x}\bar{A}$, $R_C = \bar{x}A\bar{B}$.
(b) $Y_A = \bar{x}A+x\bar{B}+A\bar{C}+A\bar{B}$.
$Y_B = AB+xB+x\bar{C}$,
$Y_C = xC+\bar{x}B+\bar{A}+BC$.

Note: the last terms in Y_A, Y_C are to avoid static hazards.)

7.4. A sequential circuit using d.c. feedback path is constructed so that no race conditions exist in it. The Boolean equation for one secondary is

$$Y_A = xA+BC+\bar{x}\bar{C}.$$

Where might a static hazard occur and what should be added to cure it? (Draw Karnaugh map.)

(Ans.: $A\bar{B}\bar{C}\bar{x} \to A\bar{B}\bar{C}x$ or $AB\bar{C}\bar{x} \to AB\bar{C}x$. Add $A\bar{C}$.)

7.5. Using the maps of Fig. 7.7, design a master–slave J–K flip-flop using only nine NOR gates. Similarly modify the circuit of Fig. 7.8(b) so that it contains only nine NAND gates.

7.6. A sequence detector has inputs x_1, x_2, x_3 in random order and one at a time. Design a circuit to give an output when either the sequence x_1, x_2, x_3 or the sequence x_3, x_2, x_1 has occurred. (x_3, x_1 can be the end of one sequence and the beginning of another.) (See Appendix C.)

7.7. Design a level sensitive R–S flip-flop suitable for use in a clocked, synchronous system (a) in master–slave form (i.e. with two quite distinguishable flip-flops); (b) as an exercise in level sensitive sequential circuit design.

What is the smallest number of NOR gates needed to construct it (assuming in (b) that the complement of the output is not needed)?

(Ans.: In both cases two secondaries A, B are needed.

(a) Master–slave $S_A = Sx \qquad S_B = \bar{x}A,$

$R_A = Rx \qquad R_B = \bar{x}\bar{A}.$

(Eleven NOR gates needed.)

(b) $Y_A = Sx + A\bar{R} + A\bar{x} = (x+A)(S+A)(\bar{R}+\bar{x})$

$Y_B = Bx + AB + A\bar{x} = (x+A)(B+A)(B+\bar{x})$

(Nine NOR gates, including one to invert x).)

Note: As with a J–K flip-flop the circuit can be built with six NOR gates and two AND gates.

7.8. Repeat problem 7.7 for a delay flip-flop whose logic input is labelled y.

(Ans.: In both cases two secondaries A, B are needed.

(a) Master–slave $S_A = xy \quad S_B = \bar{x}A,$

$R_A = x\bar{y} \quad R_B = \bar{x}\bar{A}.$

(Ten NOR gates.)

(b) $Y_A = xy + \bar{x}A + Ay = (\bar{x}+y)(x+A)(A+y),$

$Y_B = \bar{x}A + xB + AB = (\bar{x}+B)(x+A)(A+B).$

(Eight NOR gates, including one to invert x).)

7.9. Design a circuit using only five NOR gates to realize the function

$$Z = \bar{A}\bar{B} + AD + BC.$$

CHAPTER 8

Arithmetic Operations

THE arithmetic operations to be discussed here are addition, subtraction, multiplication and division of binary coded numbers. Examples of binary and decimal arithmetic are shown below. Division is treated separately on p. 202.

Addition Sum = *X plus Y*†

	binary	decimal
X	11101	29
plus Y	10110	22
Carry	111	1
Sum	110011	51

Subtraction Difference = $X-Y = X$ *plus* $\bar{Y}$ *plus* 1

Either:

Borrow		22	
	X	11101	29
	$-Y$	10110	22
	Difference	00111	7
Or	*X*	11101	29
	$\bar{Y}$	01001	77
Carry *plus* 1		11 1	1
		(1)00111	(1)07

† In this chapter the symbol + is still reserved for the OR function. The arithmetical operation of addition is written out as *plus* each time it is needed.

Multiplication

X	11101	29
$\times Y$	10110	22
	11101	58
	11101	58
	11101	638
	1001111110	

The processes of addition and multiplication are both very similar to our customary decimal operations. In subtraction, however, our usual technique of "borrowing" is not as convenient as "carrying" during addition. So in practice in machine arithmetic it is usual to use the second alternative method in which the subtrahend Y is first complemented, then 1 is added, and the sum added to X. The 1 is introduced as an initial carry digit. The extra digit produced when $Y < X$ (shown in brackets) is rejected, or carried round as the initial carry digit.

The decimal equivalent is to obtain the "nines complement" of the subtrahend (i.e. replace each digit y by $9-y$) and add it to X *plus* 1. Again the extra digit (a "hundreds" unit in the example shown) is rejected to obtain the difference $X-Y$. (Some machines perform calculations using simple binary code for each *digit* of a decimal number. This is called Binary Coded Decimal coding or *BCD* code. When it is used, forming the nines complement in code becomes an important operation; see p. 98.)

Of course, if $Y > X$ a negative result is obtained by subtraction and in order to interpret such results it is necessary to define a system for expressing them. This is usually done as follows. Suppose we are working with five bit binary numbers, the method involves adding a sixth digit, and putting it first. This is the *sign digit*. When *writing* "signed" numbers the sign digit is identified with a bar over it, but in a machine the sign digit is always recognizable simply by being the first in the standard word length.

The sign digit is normally "0" for positive numbers and "1" for negative. The code used to express numbers -1 to -31 can easily be derived from the rules of subtraction.

Thus to obtain negative numbers, complement each digit and add 1 to the least significant digit.

Example:

	$+22 =$	$\bar{0}10110$
complement		$\bar{1}01001$
add "1"		$\quad\quad 1$
	$-22 =$	$\bar{1}01010$

The number can be read as $-32+10$. Thus, $29-22$ can now be written

29	$\bar{0}11101$
plus (-22)	$\bar{1}01010$
	$(1)\bar{0}00111 = 7$

But $22-29$ becomes

$22 =$	$\bar{0}10110$
plus (-29)	$\bar{1}00011$
	$\bar{1}11001$

This is a negative number whose magnitude is 1 *plus* its complement, namely 7 again.

Thus *the magnitude of a negative number is obtained by complementing each digit and adding 1*, so that $22-29 = -7$. Note that the sign digit is included in *all operations* just as if it were a number.

Using this scheme positive and negative numbers can be both added and subtracted using the same machine binary adder.

For all three arithmetic operations we now only need the ability to add, to shift and to complement. With flip-flops the complement of a digit is always available. The shifting and adding functions will now be described.

The Shift Register

A shift register is composed of a series of pulsed flip-flops (or master–slave pairs) so arranged that when a clock pulse is applied to them all simultaneously, the setting of each adopts the previous setting of its neighbour. This is achieved as shown in Fig. 8.1 using *R–S* flip-flops. (Using *D* flip-flops which are particularly suitable for this shifting function, only one logic lead to each previous flip-flop is needed.)

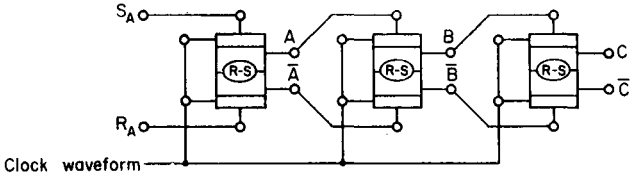


FIG. 8.1. A three-stage shift register, using triggered *R–S* flip-flops.

The set and reset logic inputs are connected to the two outputs of the preceding stage. If the outputs of the last flip-flop (*C* in this case) are connected back to the inputs of the first flip-flop then a *Ring Counter* is formed and any pattern set up in the register (by means of the preset inputs, not shown) is recirculated.

Shift registers can be made which shift in either direction according to a "forward" or "reverse" input voltage level, *X*. Suppose a shift register is constructed of delay flip-flops with inputs Y_A, Y_B, Y_C, etc. Consider the input to *B*. If Y_B is preceded by two stages of logic such that

$$Y_B = AX + C\bar{X},$$

then on the arrival of a clock pulse flip-flop *B* will adopt the setting of its left-hand or right-hand neighbour according to the value of *X*. Similar equations apply to the other stages of the register.

Adders

The full process of addition of two digits consists of two parts. First, the addition of the two numbers including a carry digit if there is one; second, the calculation of the presence or absence of a new carry digit.

Thus an "adder" will have three inputs, the digit inputs A and B (usually obtained from a pair of flip-flops, one from each of two shift registers) and (in serial adders) the carry input obtained from the output of a single element "carry" register or "carry flip-flop". It will also have two outputs, namely the "sum" and the "new value of the carry digit". All possible combinations of input and output are shown in Table 8.1.

TABLE 8.1. ADDITION OF DIGITS A, B

Inputs			Outputs	
A	B	Carry	Sum	New carry
0	0	0	0	0
1	0	0	1	0
0	1	0	1	0
1	1	0	0	1
0	0	1	0	0
1	0	1	1	1
0	1	1	1	1
1	1	1	0	1

(a) *Serial Adders*

The most straightforward way to realize this process of addition using circuit elements is shown in Fig. 8.2. Each row of Table 8.1 is represented by an AND gate, four of which combine to give the sum, and four its complement.

An inspection of the table also shows that the carry digit only *changes* in two circumstances. It sets for $AB\bar{C}$ and resets for $\bar{A}BC$. In Fig. 8.2 these are the set and reset inputs for a *Carry flip-flop* which holds the New carry digit ready for the next stage of the addition.

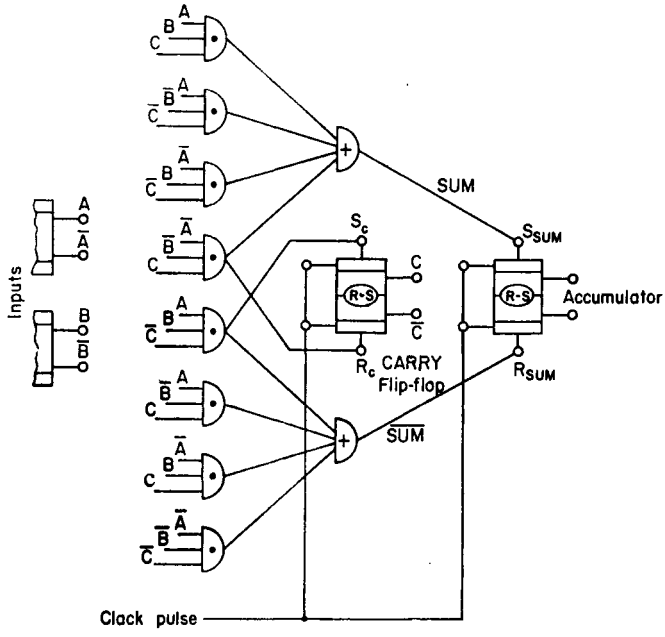


FIG. 8.2. A simple serial adder.

Now if the inputs to the adder, A, $\bar{A}$ and B, $\bar{B}$ are the outputs of two shift registers, and the outputs from the adder (SUM, $\overline{\text{SUM}}$) are the inputs to another register, then a clock pulse supplied to all the flip-flops will pass the numbers stored in the A, B registers into the adder and will accumulate their Sum in the output register, usually called *the accumulator*. The whole circuit is known as a *serial adder*.

This type of adder contains more circuit elements than are necessary. For example, $\overline{\text{SUM}}$ could be obtained by inverting the SUM. This would save 3 AND gates and an OR gate. The use of *D* flip-flops in the shift registers would achieve the same saving.

(b) *The Half Adder*

The usual method of addition involves the use of *half adders* and *full adders*. These are circuits whose function can best be explained by rewriting the operations listed in Table 8.1 in the form of Boolean equations.

$$\begin{aligned}\text{SUM} &= \bar{A}\bar{B}C + A\bar{B}\bar{C} + \bar{A}B\bar{C} + ABC \\ &= \bar{C}(A\bar{B} + \bar{A}B) + C(\bar{A}\bar{B} + AB).\end{aligned}$$

$$\begin{aligned}\text{New Carry} &= AB\bar{C} + A\bar{B}C + \bar{A}BC + ABC \\ &= AB + C(\bar{A}B + A\bar{B}).\end{aligned}$$

The factors in the second line of each equation show that the full addition process can be realized by two cascaded circuits, called half adders, each consisting of an AND gate and an Exclusive-OR gate, as shown in Fig. 8.3. (Figure 4.14(a) shows a type of half adder, the AND output being shown dotted.)

This full adder circuit can now replace the upper half of the simple serial adder. To drive *R–S* flip-flops in the accumulator and in the Carry flip-flop two inverters would also be needed to complement the SUM and CARRY outputs. For use with *D*

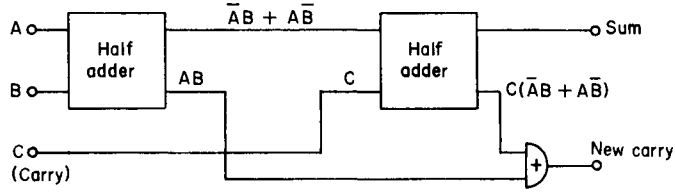


FIG. 8.3. The full adder, comprising two half adders and an OR gate.

flip-flops in both the registers and the carry flip-flop the circuit is sufficient on its own. This is also a serial adder.

If the output of the adder is not taken to the input of a third shift register, but is returned to the input of register A, the sum of A and B will replace A in the register as the clock waveform is applied. Thus register A is also the accumulator. This evidently allows the continued addition of a series of numbers placed successively in B.

If there are N flip-flops in the B register then one more than N clock pulses will be needed for each addition, and N plus M bistable elements will be needed in the accumulator, where M is the maximum number of additions which the adder is designed to handle before passing on the SUM and clearing.

The serial adder can be converted into a serial subtractor by replacing the B register outputs by their complements (i.e. replace B by $\bar{B}$ and vice versa on the AND gate inputs in Fig. 8.2), and by ensuring that the carry flip-flop is set to 1 before each subtraction, in order to add in the extra 1 as explained at the beginning of this chapter. In practice, this occurs automatically if $B < A$, since the rejected digit at the end of the word is left in the carry flip-flop ready for the next subtraction. Of course, for subtraction the accumulator must be the same length as register A; which is bound to be the case if A is also the accumulator!

(c) *The Parallel Adder (Subtractor)*

The serial adder is rather slow since one clock pulse is needed for each digit of the SUM. If the input registers are arranged as in Fig. 8.4 then since full adders are combinational circuits the SUM appears at the Sum outputs. One clock pulse then transfers all the digits to the accumulator—which may again be one of the input registers. Note that only a half adder is needed for the first pair of input digits.

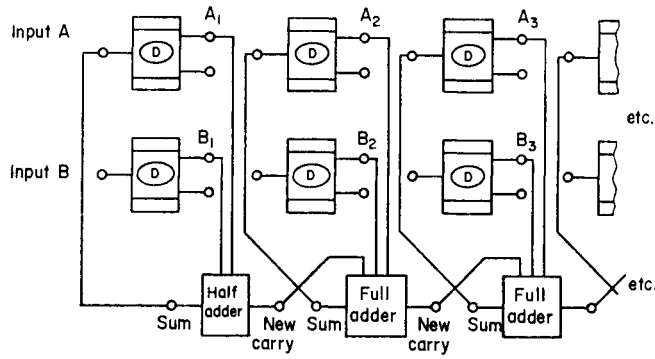


FIG. 8.4. Part of a parallel adder using pulsed flip-flops. (Pulse inputs not shown.) Register A is the accumulator.

Again the circuit can be used for subtraction, with appropriate minor modifications, in particular the provision of the extra digit.

Multiplication

The principle of machine multiplication is illustrated by the schematic diagram of a small multiplier in Fig. 8.5. The three bit number $A_2A_1A_0$ is to be multiplied by the three bit number $B_2B_1B_0$. The accumulator starts with all zeros and will finally contain the product.

In the accumulator of this multiplier it is convenient to use flip-flops (preferably delay type) which have two separate logic and pulse inputs per flip-flop, so that two independent functions can be performed with each flip-flop. One logic input of each D flip-flop is connected to the output of the flip-flop to the left of it, and all associated pulse inputs are driven in parallel by the train of *shift* pulses shown in Fig. 8.5(b). Thus the accumulator can behave as a shift register. The other logic inputs are used to accept the outputs of the adders and are triggered by the train

of *clock* pulses. Since both pulse trains are only effective on their negative-going edges, it is clear from Fig. 8.6(b) that the two functions, shift and add, occur alternately.

The multiplier circuit operates as follows. Initially the accumulator is at zero, and numbers to be multiplied are held in A and B.

If B_0 is in state 1, the clock pulse transfers the sum of the contents of A and of the accumulator S back into S. If B_0 is zero the contents of S are not altered. The shift pulse then shifts the contents of S and of B one place to the right.

These two operations are carried out alternately M times, where M is the number of bits which can be stored in the B register. Each pair of operations corresponds to one row of long multiplication except that each row is added to the previous one as

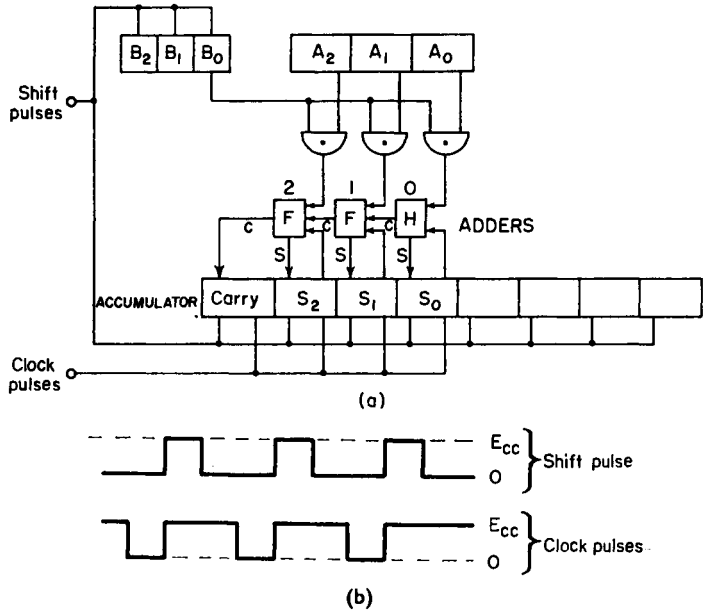


FIG. 8.5. A serial multiplier incorporating a parallel adder. (a) The schematic diagram; (b) the driving waveforms.

it is generated instead of adding all rows together at the end (as in the sum on p. 193).

Evidently A_0 and B_0 are the least significant digits in A and B. The "carry" is the most significant one in the final answer in the accumulator. If A can hold N digits, then S must be able to accumulate N plus M plus 1 digits.

Division

Just as multiplication is a process of alternate addition and shifting to the left, so division is a process of subtraction and shifting to the right. But there is one important difference. The sequence of operations in division normally depends upon the actual values of the numbers involved. Thus at each stage of long division, if the number to be subtracted is greater than the subtrahend, we do not subtract, we shift to the right and try again. In multiplication, on the other hand, the sequence is independent of the relative magnitudes of the two numbers.

In division, therefore, it is necessary to test for relative magnitudes; that is to subtract one number from the other, and if the result is negative to conclude that the subtraction should not have been done! Or, rather better, if the result is negative to add back in again the number that was just subtracted, and then shift the divisor one place to the right, and try again. The following example will illustrate the method.

Example: Divide $A = 135$ by $B = 27$.

Or in binary form divide $A = 10000111$ by $B = 11011$.

For subtraction we shall need 1 *plus* the complement of B, namely 00101.

We should also recall that when subtraction is performed by adding 1 *plus* the complement, a positive solution is indicated by a 1 being carried off the end of the accumulator (and disregarded). However, we can now use this rejected digit, to be called U, to indicate whether or not to add back in the number just

subtracted (i.e. if there is a rejected digit we do *not* add the divisor back in again).

The calculation proceeds as follows:

A		10000111	
subtract B		00101	add
No U		10101111,	$\therefore$ lst digit of result is 0
$\therefore$ add B		11011	add
$U = 1$	(1)	10000111	
$\therefore$ shift and subtract B		100101	add
$U = 1$	(1)	00011011	$\therefore$ 2nd digit of result is 1
$\therefore$ shift and subtract B		1100101	add
No U		11100101	$\therefore$ 3rd digit of result is 0
$\therefore$ add B		11011	add
$U = 1$	(1)	00011011	
$\therefore$ shift and subtract B		11100101	add
$U = 1$	(1)	00000000	$\therefore$ 4th digit of result is 1

There is no remainder so the result is $0101 = 5$.

Notice that the first subtraction is performed in the most significant position, and that subsequent ones are shifted one place to the right and that whenever B is added back in again it must be in the appropriate position.—Notice too that the result is indicated by the value of U at each stage *following a subtraction.* The value of U following an addition has no significance in the result. However, on all occasions that $U = 1$ the next operation is "shift and subtract B", and whenever $U = 0$ the next step is "add B". Thus the sequence of events is quite clearly defined.

In practice the subtraction is achieved, and the value of U obtained, by combinational circuits. Thus it is not necessary to store the result and add the divisor back in again if the result is negative. It is only necessary to inhibit the subtraction and shift instead. This evidently reduces the number of steps significantly.

REFERENCES

1. J. J. EBERS and J. L. MOLL, The large signal behaviour of junction transistors, *Proc. I.R.E.* **42**, 1761–72 (1954).
2. J. J. SPARKES, *Junction Transistors*, Pergamon Press, Oxford, 1966.
3. *Motorola Switching Transistor Handbook*, Motorola, 1966.
4. *Zener Diode Handbook*, International Rectifier Corporation, El Segundo, California, 1963. (A source of references.)
5. M. V. JOYCE and K. K. CLARKE, *Transistor Circuit Analysis*, Addison-Wesley, Reading, Mass., 1961.
6. J. MILLMAN and H. TAUB, *Pulse, Digital and Switching Waveforms*, McGraw-Hill, New York, 1965.
7. K. W. CATTERMOLE, *Transistor Circuits*, Heywood, London, 1959.
8. M. V. JOYCE and K. K. CLARKE, *op. cit.*, Chapter 11.
9. J. G. LINVILL and R. H. MATTSON, Junction transistor blocking oscillators, *Proc. I.R.E.* **43**, 1632–9 (Nov. 1955).
   J. MILLMAN and H. TAUB, *op. cit.*, Chapter 16.
10. K. P. P. NAMBIAR and A. R. BOOTHROYD, Junction transistor bootstrap linear sweep circuits, *Proc. I.E.E.*, Pt. B, 104, No. 15, 293–306 (1957).
11. *Integrated Circuits*, by Motorola Inc., McGraw-Hill, 1965.
    K. J. DEAN, *Integrated Electronics*, Chapman & Hall, 1967.
12. W. V. QUINE, The problem of simplifying truth tables, *Amer. Math. Monthly* **59**, No. 8, 521–31 (Oct. 1952).
    W. V. QUINE, A way to simplify truth functions, *Amer. Math. Monthly* **62**, No. 9, 627–31 (Nov. 1955).
    E. J. MCCLUSKEY, Minimisation of Boolean functions, *Bell System Tech. Journal* **35**, No. 6, 1417–44 (Nov. 1956).
13. M. KARNAUGH, The map method of synthesis of combinational logic circuits, *Trans. A.I.E.E. Comms. & Electronics* **72**, pt. 1, 593–9 (Nov. 1953).
14. L. NASHELSKY, *Digital Computer Theory* (Chapter 4), John Wiley, New York, 1966. (See also Marcus, Torng, etc., ref. 18.)
15. H. C. TORNG, *Logical Design of Switching Systems*, Addison-Wesley, Reading, Mass., 1964.
16. R. BEAUFOY, Transistor switching circuit design in terms of charge parameters, *Proc. I.E.E.*, Pt. B, vol. 106, suppl. 17, pp. 1085–91 (May 1959).

17. M. PHISTER, *Logical Design of Digital Computers*, John Wiley, New York 1958.
18. D. A. HUFFMAN, The synthesis of sequential circuits, *Journal of the Franklin Institute*, **257**, No. 3, pp. 161–90 (March 1954); No. 4, pp. 275–303 (April 1954).
G. H. MEALY, A method of synthesizing sequential circuits, *Bell System Tech. Journal*, **34**, No. 5, pp. 1045–79 (Sept. 1955).
E. F. MOORE, Gedanken-experiments on sequential machines, *Automata Studies*, pp. 129–53, Princeton University Press, Princeton, N. J., 1956.
M. PHISTER, *op. cit.*
M. P. MARCUS, *Switching Circuits for Engineers*, Prentice-Hall, Englewood Cliffs, 1962.
G. A. MALEY and J. EARLE, *The Logic Design of Transistor Digital Computers*, Prentice-Hall, Englewood Cliffs, 1963.
E. J. MCCLUSKEY and T. C. BARTEE, *A Survey of Switching Circuit Theory*, McGraw-Hill, New York, 1962.
E. J. MCCLUSKEY, *Introduction to the Theory of Switching Circuits*, McGraw-Hill, New York, 1965.
R. E. MILLER, *Switching Theory* (2 vols.), John Wiley, New York, 1965.
H. C. TORNG, *op. cit.*
19. S. H. UNGER, Hazards and delays in asynchronous sequential switching circuits, *I.R.E. Transactions*, Vol. CT-6, No. 1, 12–26 (Mar. 1959). (See also E. J. McCluskey, 1965.)
20. R. L. WIGINGTON, A new concept in computing, *Proc. I.R.E.* **47**, 516–23 (1959).
E. GOTO, The parametron, a digital computing element which utilises parametric oscillations, *Proc. I.R.E.* **47**, 1304–16 (Aug. 1959).
21. D. ZISSOS, and G. W. COPPERWHITE, *Logic Design Manual*, Pitman, London, 1968.

APPENDIX A

Design Consideration for R.T.L.

As AN example of the considerations needed to design logic circuits, a simple example will be discussed.

Consider two cascaded stages of R.T.L. as shown in Fig. A.1. Suppose each transistor has a fan-in of n and a fan-out of m as shown. What limitations are there on the magnitudes of n and m consistent with reliable operation?

For worst case design we need to know the range of values of $V_{CE(sat)}$ at the chosen operating current, the maximum value of $V_{BE(ON)}$, the minimum value of β and the maximum value of I_{BX}.

The equations describing the circuit can be chosen as follows:

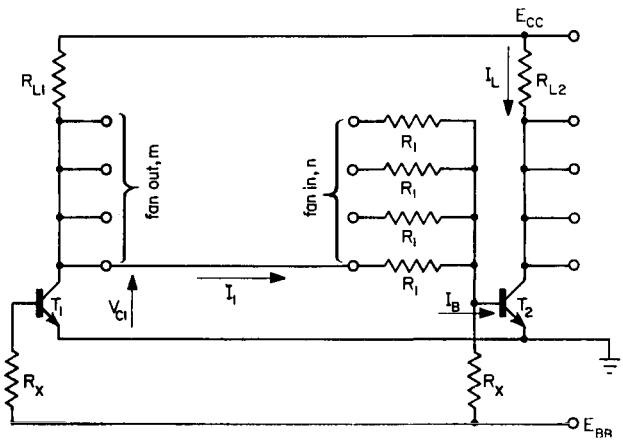


FIG. A.1. Two stages of R.T.L. NOR gates.

(a) *With T_1 off and I_{BX} negligible, T_2 must be bottomed*

$$\text{Smallest} \quad V_{C1} = E_{CC} - mI_1R_L. \tag{A.1}$$

$$\text{Smallest} \quad I_1 = \frac{V_{C1} - V_{B(\text{ON})\,(\max)}}{R_1}, \tag{A.2}$$

$$\text{whence} \quad V_{C1}\left(1 + \frac{mR_L}{R_1}\right) = E_{CC} + \frac{mR_L V_{BE(\text{ON})\,(\max)}}{R_1}. \tag{A.3}$$

To provide sufficient input current to bottom T_2:

$$I_1 = \frac{V_{C1} - V_{BE(\text{ON})\,(\max)}}{R_1} = I_B + \frac{V_{BE(\text{ON})\,(\max)} - E_{BB}}{R_X} + (n-1)\left\{\frac{V_{BE(\text{ON})\,(\max)} - V_{CE(\text{sat})\,(\min)}}{R_1}\right\} \tag{A.4}$$

where in the worst case $n-1$ input resistors obtain their input voltages from bottomed transistors. The last term implies that $n-1$ driving transistors all possess the minimum value of $V_{CE(\text{sat})}$. For large values of n this is very unlikely so that for a given confidence limit this term can be relaxed.

If I_B is to hold T_2 bottomed, then

$$I_B \geqslant \frac{1}{\beta_{\min}}\left[I_L + m\frac{(V_{BE(\text{ON})\,(\max)} - V_{CE(\text{sat})\,(\min)}}{R_1}\right]. \tag{A.5}$$

(b) *With T_1 bottomed and T_2 cut off*

$$\frac{V_{BE(\text{OFF})} - E_{BB}}{R_X} \geqslant I_{BX} + \frac{n}{R_1}(V_{CE(\text{sat})\,(\max)} - V_{BE(\text{OFF})}). \tag{A.6}$$

Again, within a specified confidence limit the last term in eqn. (A.6) can be relaxed owing to the improbability of finding n "worst" transistors connected in parallel.

The circuit has to be designed to ensure a low enough value of $V_{BE(\text{OFF})}$ so that under extremes of temperature, for example, T_2 will remain cut off.

These equations are sufficient to determine m and n for given limit parameters of the transistor where only slow speed operation is involved (i.e. response times ignored).

Example: *Silicon transistors*

Let $E_{CC} = -E_{BB} = 10$ V and suppose the circuit is designed for $I_L \approx 5$ mA (so that $R_L = 2$ kΩ) and for a base cut-off voltage $V_{BE(OFF)} \leqslant 0{\cdot}2$ V. Assume the transistor data to be:

$$\beta_{min} = 25 \text{ for } V_{CE} = V_{CE(sat)\,(max)},$$
$$V_{CE(sat)} = 0{\cdot}4 \text{ V (max)}, \quad 0{\cdot}2 \text{ V (min)},$$
$$V_{BE(ON)} = 0{\cdot}8 \text{ V (max)},$$
$$I_{BX} = 0.$$

To begin with, let $m = 5$. Then, taking the limiting equalities in eqns. (A.5), (A.6) and using the units volts, kΩ, mA, we obtain by eqn. (A.6)

$$\frac{10{\cdot}2}{R_X} = n\frac{0{\cdot}2}{R_1}. \tag{A.7}$$

From eqn. (A.5)

$$I_B = \left[4{\cdot}9 + 5\frac{(0{\cdot}8 - 0{\cdot}2)}{R_1}\right] \cdot \frac{1}{25}$$

or

$$I_B = 0{\cdot}196 + 3/25R_1. \tag{A.8}$$

From eqn. (A.3)

$$V_{C1} = \frac{10R_1 + 5 \times 2 \times 0{\cdot}8}{R_1 + 10}, \tag{A.9}$$

whence by substituting for V_{C1}, I_B and R_X eqn. (A.4) becomes:

$$\frac{10R_1 + 8}{R_1 + 10} - 0{\cdot}196R_1 = 0{\cdot}32 + 0{\cdot}812n. \tag{A.10}$$

Equation (A.10) can be plotted as shown in Fig. A.2, curve (a) and reveals a maximum fan-in of only 4 when R_1 is about 12 kΩ. Tolerances in component values would significantly reduce this fan-in.

Curve (b) of Fig. A.2 shows the improvement which can be achieved by putting $R_X = \infty$ and accepting $V_{CE(sat)\,(min)}$ of 0·4 V instead of 0·2 V. The fan-in rises to 7 for a fan-out of 5.

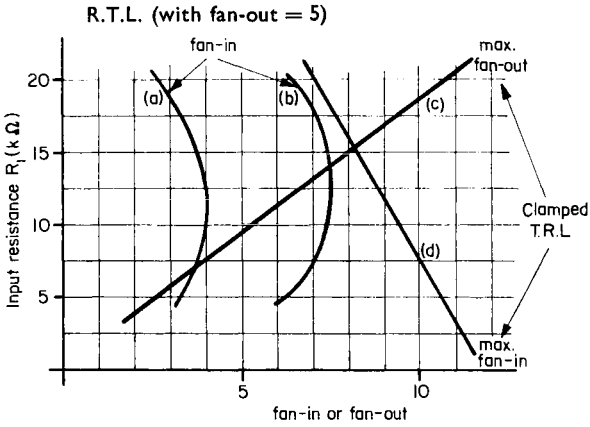


FIG. A.2. The effect of base resistance on R.T.L. (a) Variation of fan-in when $V_{BE(OFF)} = 0{\cdot}2$ and fan-out is 5; (b) variation of fan-in when $R_X = \infty$ and $V_{CE(sat)} = 0{\cdot}4$ V minimum, and fan-out = 5; (c), (d) with clamped collector voltage fan-in and fan-out are independent below their maximum values. The curves show maximum values. The optimum is near the intersection.

Similar curves can be obtained for different values of fan-out, m.

The problem with this circuit is that as R_1 is reduced in order to give a larger fan-in, the source becomes more heavily loaded thus reducing the advantage gained by making R_1 smaller.

This difficulty can be avoided using *clamped R.T.L.* as in Fig. 4.6(b). In this circuit there is a maximum fan-in and fan-out possible for each value of R_1. As R_1 increases the maximum fan-out increases too but the fan-out falls. The values of fan-in and fan-out for the case of

$$V_{BE(OFF)} \leqslant 0{\cdot}2 \text{ V} \quad \text{and} \quad V_{CE(sat)\,(min)} = 0{\cdot}2 \text{ V}$$

are plotted as two separate lines (c) and (d) in Fig. A.2. These apply to the same conditions as for curve A of R.T.L. At $R_1 =$ 15 kΩ both m and n are approximately doubled as compared with R.T.L.

APPENDIX B

An Analysis and Design of a D.T.L. R-S Flip-flop with the Two Levels of Input Gating

THE purpose of this Appendix is to describe the functions of each part of a flip-flop so that the appropriate value of each can be calculated. Both a positive logic and a negative logic flip-flop, using *npn* transistors, will be analysed since there are important differences.

The analysis is based on that described by Beaufoy[16] and is presented here to illustrate how the various parts of a digital circuit can be designed step by step.

1. The Positive Logic Flip-flop

An example of an *R–S* flip-flop is shown in Fig. B.1. The Eccles–Jordan part of the circuit is enclosed in the dashed line and includes two clamping diodes D_1 and D_2 which prevent the outputs A, $\bar{A}$ rising above about 6·2 V (in this example). The two levels of diode logic (which are provided on both sides) are easily identified. D_3 and C_S are pulse steering components whose function was explained on p. 121. The purpose of D_4 and D_5 will be described soon.

That the circuit uses positive logic can be recognized (with an *R–S* flip-flop) by finding which input condition does not lead to a determinate solution. In this case if both S_A and R_A are held positive (by the input gates) the application of a negative going trigger turns off *both* transistors, causing indeterminancy. *There-*

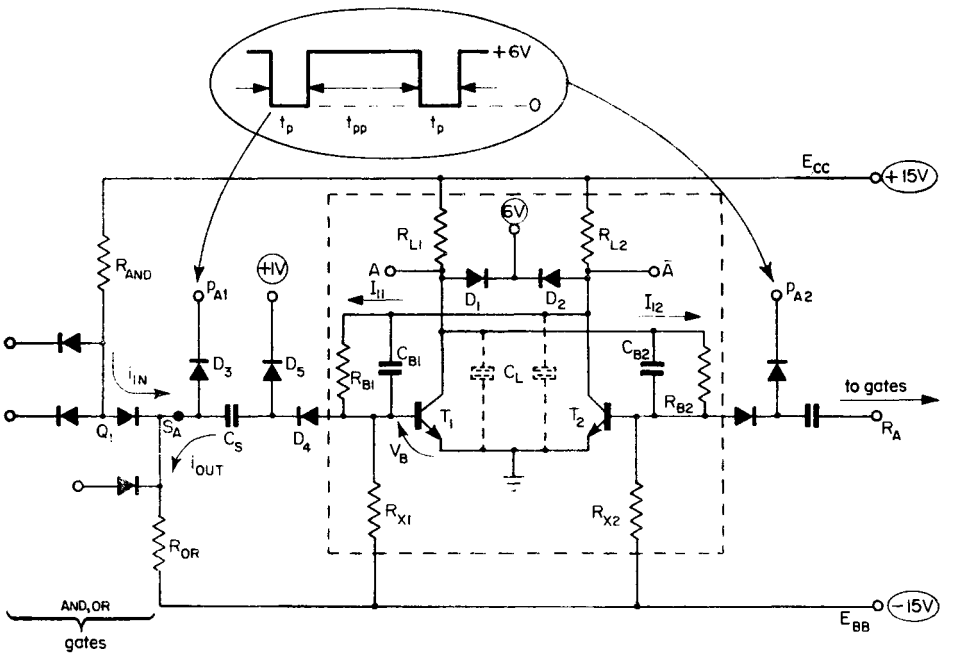


FIG. B.1. A practical circuit for a positive-logic, *R–S* pulsed flip-flop with two levels of logic on each logic input. (Negative-going trigger pulses.)

fore, the circuit is a positive-logic, *R–S* flip-flop. The "1" level is about +6 V. The "0" level about zero volts.

The circuit will be designed to operate at a clock pulse rate of 100 kc/s. That is $t_p = 1{\cdot}0$, μsec and $t_{pp} = 9$ μsec and the pulse waveform is a return-to-zero type, being normally at 6 V but falling to zero every 10 μsec, as shown in Fig. B.1.

The relevant transistor parameters will be introduced where appropriate. We will consider germanium transistors of quite low frequency since the problems are greatest in this case.

Since each parameter of the circuit is somewhat dependent upon every other a solution can only be obtained in closed form by setting up and solving simultaneous equations. Since many of these relations are inequalities such a method is not practicable. The design should begin therefore with those components which are least affected by other parts of the circuit.

We will begin with the bias resistors of the Eccles–Jordan circuit.

(a) *The Flip-flop Bias Resistors* R_{B1}, R_{B2}, R_{X1}, R_{X2}

When T_1 is cut off R_{X1} must hold the base of T_1 reverse biased despite (i) the cut-off current of T_1, (ii) the current in R_{B1}, (iii) the cut-off current of diode D_4.

$$\text{Thus } \frac{V_{B(\text{OFF})} - E_{BB}}{R_{X1}} \geqslant \frac{V_{CE(\text{sat})} - V_{B(\text{OFF})}}{R_{B1}} + I_{BX} + I_{D4}.$$

$V_{B(\text{OFF})}$ must be less than, say, $-0{\cdot}5$ V to provide sufficient protection against noise, which otherwise might momentarily forward bias the base of T_1. (With silicon transistors R_{X1} is not necessary and $V_{B(\text{OFF})}$ can be slightly positive.)

When T_1 is bottomed, the base current of T_1 must exceed $I_{C(\text{ON})}/\beta_{\min}$. Thus the current I_{11} in R_{B1} is given by

$$I_{11} = \frac{6{\cdot}2 - V_{B(\text{ON})}}{R_{B1}} \geqslant \frac{I_{C(\text{ON})}}{\beta_{\min}} + \frac{V_{B(\text{ON})} - E_{BB}}{R_{X1}}.$$

Since some of the collector current may be demanded when the transistor is already held bottomed, some of $I_{C(ON)}$ should be divided by $\beta_{S(min)}$, not by $\beta_{(min)}$ (see p. 18). The proportion can only be estimated at this stage.

Let $I_{BX(max)} = 50\ \mu A$,

$V_{CE(sat)\,(max)} = 0{\cdot}25$ V $\quad I_{D4(max)} = 50\ \mu$A,

$V_{B(ON)} = 0{\cdot}25$ V,

$\beta_{min} = 30 \qquad \beta_{S(min)} = 20.$

If the circuit is designed for a maximum value of I_C of 10 mA, and for $V_{B(OFF)} = -0{\cdot}5$ V then putting $\beta_{min} = 25$ as the average of β and β_S

$$R_{B2} = R_{B1} \approx 10\ \text{k}\Omega,$$

$$R_{X2} = R_{X1} \approx 82\ \text{k}\Omega.$$

(b) *The Input Capacitor C_S*

When the negative going edge of the clock pulse drives the set input S_A, negatively, the capacitor C_S must (a) remove the base charge of transistor T_1; (b) absorb the base current I_B (which would otherwise turn T_1 on again) until the circuit has switched and T_2 can hold T_1 off via the cross coupling circuit. Under the worst conditions, as we shall see, the collector voltage of T_2 may not begin to change until the end of the clock pulse period t_p. Therefore, if ΔV_S is the change in voltage across C_S as a result of the clock pulse

$$\Delta V_S C_S = Q_{OFF(max)} + I_B t_p.$$

Now $Q_{OFF} = I_{C(ON)}\,\tau_C + Q_{VC} + I_{BS}\,\tau_S$, so if the transistor data provided are

$\tau_{C(max)} = 0{\cdot}020\ \mu$sec

$\tau_{S(max)} = 0{\cdot}80\ \mu$sec,

$Q_{VC(max)} = 120$ pC (for $\Delta V_{CB} = 6$ V),

$\beta_{(max)} = 200,$

and with $I_{C(ON)} = 10$ mA and $I_B = 0{\cdot}4$ mA, $I_{BS(max)} = 0{\cdot}35$ mA.

Hence $Q_{OFF(max)} = 600$ pC

and $\Delta V_S C_S = 1000$ pC.

ΔV_S is nominally 6 V, but owing to diode D_5 (which clamps V_S at +1 V) and to loss in other diodes, the change in voltage across C_S is likely to be more nearly 4·5 V.

$\therefore C_S \approx 220$ pF.

(c) *The Gates*

(i) When the voltage of S_A is being driven negatively by the gates, all the OR gate inputs are at zero volts, so that all the OR gate diodes are cut-off. Thus the current i_{OUT} flowing away from C_S is given by

$$i_{OUT} = \frac{(\text{voltage of } S_A) - E_{BB}}{R_{OR}} + nI_D$$

if there are n OR gate diodes.

In the worst case $I_D = 0$, since the current i_{OUT} has to ensure that at the end of the period t_{pp}, S_A will be at or near zero volts; that is C_S must be discharged from its initial value, just after a clock pulse, to a final state with almost zero volts across it.

After the clock pulse, when the voltage at p_{A1} rises to 6 V again, D_3 cuts off so that S_A does not normally rise as far as +6 V, but to a smaller value depending on the transistor emitter cut-off capacitance, the capacitance of D_3, etc. We will assume that after the clock pulse S_A is at +3 V.

If $E_{BB} = -15$ V, i_{OUT} is almost constant. Taking its mean value

$$i_{OUT} = \frac{(15+3)+15}{2R_{OR}} = \frac{C_S \Delta V_S}{t_{pp}},$$

where ΔV_S is the change of voltage across C_S. If T_1 is conducting, V_{B1} is constant, so $\Delta V_S \approx 3$ V as just explained.

Thus $i_{OUT} \approx 75\ \mu$A and $R_{OR} \approx 220$ kΩ.

This current must be available from R_{B1} as C_S is discharged. Since in choosing R_{B1}, a value of $\beta < \beta_{\min}$ was used this current can be provided by R_B without danger of T_1 coming out of saturation.

(ii) When input S_A goes positive, the AND gate inputs are either already at +6 V or are going positive too, so that all the AND gate diodes are cut-off. Thus the current i_{IN} available to drive S_A positive is the current through R_{AND}, less that required by R_{OR}.

If V_{SA} is the varying voltage of point S_A

$$i_{IN} = \frac{E_{CC} - V_{SA}}{R_{AND}} - \frac{V_{SA} - E_{BB}}{R_{OR}}.$$

i_{IN} must remain positive as V_{SA} varies between 0 and 6 V, and it must be sufficient to charge C_S through 6 V in the interpulse period t_{pp}. (The initial value of $V_{SA} \approx 0$ in the worst case.)

During this transition the cut-off currents of all the AND gate diodes aid i_{IN}, whilst all but one of the OR gate diodes draw current from it. We will suppose that the net worst case effect is as if one OR gate diode remained.

A Thévenin equivalent of this gate circuit during positive transitions is a generator of voltage V_T in series with a resistor R_T where by superposition

$$V_T = \frac{E_{CC}R_{OR} + E_{BB}R_{AND} - I_D R_{OR} R_{AND}}{R_{OR} + R_{AND}},$$

$$R_T = \frac{R_{OR}R_{AND}}{R_{OR} + R_{AND}}$$

and the variation of V_{SA} is given by

$$V_{SA} = V_T\left(1 - e^{-t/R_T C_S}\right).$$

With

$E_{CC} = 15$ V, $E_{BB} = -15$ V, $I_D = 50$ μA, $R_{OR} = 220$ kΩ,

and $C_S = 220$ pF, a suitable value of R_{AND}, to ensure that V_{SA} reaches 6 V in time t_{pp} is

$$R_{AND} \approx 40 \text{ k}\Omega.$$

Thus i_{IN} is initially about 300 μA and falls to 120 μA as V_{SA} rises.

The current needed to hold down one AND gate is the current through R_{AND} (*not* reduced by the current through R_{OR} since all OR gate diodes might be cut off). It is therefore 0·38 mA.

The charging current i_{IN} flows through C_S and D_5. If D_4 and D_5 were not in the circuit i_{IN} would have to flow in R_{X1}, and R_{X1} would have to be reduced sufficiently to ensure that this extra current did not turn T_1 on during t_{pp}, (that is, by a factor of 5 or more). This large change would significantly affect R_{B1}, and therefore C_S and R_{AND}, too, so that the circuit performance would be seriously impaired. This is the reason for the presence of D_4 and D_5 in the circuit.

(d) *The Cross-coupling Capacitor* C_{B2}

C_{B2} has to ensure that T_2 is turned on adequately within the time t_p. Thus

$$\Delta(V_{C1}-V_{B2})C_{B2} = Q_{ON},$$

where $\Delta(V_{C1}-V_{B2})$ is the change of voltage across the capacitor C_2. If V_{B2} was initially reverse biased by 0·5 V, then $\Delta(V_{C1}-V_{B2})$ will be about 5 V taking into account $V_{BE(ON)}$, $V_{CE(sat)}$, etc.

Thus $$5C_2 = Q_{VD}+Q_B+Q_{VC}.$$

Using the values obtained in (b) and if

$$Q_{VD} = 50 \text{ pC} \quad \text{for} \quad V_{BE} = 0{\cdot}5 \text{ V},$$

$$C_2 = \frac{420}{5} = 84 \text{ pF}.$$

(e) *The Load Resistor* R_{L1}

When T_1 has been cut-off its collector voltage must rise to 6 V within the period t_p in order that the charge in C_{B2} will flow into the base of T_2 to turn it ON. Thus there must be sufficient

current available from R_{L1} to charge both C_2 and any stray capacitance, C_L, in 1 μsec, despite some loss in R_{B2}.

Deriving a Thévenin equivalent of the collector voltage of T_1

$$V_{C1} = V_T(1 - e^{-t/R_T(C_{B1}+C_L)}),$$

where $$V_T = E_{CC}R_B/(R_L+R_B),$$

$$R_T = R_BR_L/(R_L+R_B),$$

whence if $$C_L = 200 \text{ pF},$$

$$R_{L2} = R_{L1} \approx 5{\cdot}6 \text{ k}\Omega,$$

which means that $V_T = 9{\cdot}7$ V and $R_T = 3{\cdot}6$ kΩ.

Fan-out

With $R_{L1} = 5{\cdot}6$ kΩ, the current flowing in R_{L1} when T_1 is bottomed is evidently $15/5{\cdot}6 = 2{\cdot}7$ mA. But the circuit is designed for a collector current of 10 mA, so that 7·3 mA are available for driving gates.

The current per AND gate is 0·38 mA so that the theoretical fan-out is 18.

Tolerances

The above first stage calculation has been worked out taking little account of component tolerances. It should now be repeated adopting the most unfavourable component tolerances. Just what percentage tolerance for each component should be considered depends upon the system and its economics. The calculations are usually long and tedious though quite straightforward. Nowadays the task is usually performed by computer and is beyond the scope of this book. The result of allowing 10% tolerance on component values is usually a reduction of fan-out by a factor of at least 2, so that in practice a fan-out of 7 or 8 is to be expected.

Discussion

Since the AND gate current is determined by the rate at which C_S is charged it is evident that if the clock pulse rate were increased, the AND gate current would have to be increased and the fan-out reduced. Indeed, it is generally true that at high speeds (relative to the capability of the transistor used) fan-out and speed have to be traded to obtain an optimum system.

Higher-speed transistors (i.e. transistors with smaller Q_{ON} and Q_{OFF}) need smaller capacitances C_S, C_B so that the same current levels will achieve faster operation. Indeed C_B can often be omitted.

A better choice of supply voltages, say $+20$ and -10 V, would increase R_{AND} and improve the fan-out somewhat.

A more detailed consideration of some of the limiting conditions in the circuit can lead to some further improvements in design. For example, it is not necessary to turn on the full 10 mA in T_2 (part (d) above) in order to reach the new stable state. However, these details together with the problems of tolerancing will not be discussed further.

The Negative Logic R–S Flip-flop (using *npn* transistors)

The circuit is shown in Fig. B.2. As compared with Fig. B.1, D_4 and D_5 are omitted, C_S and D_3 are interchanged, and the gates are inverted.

There are three important differences of operation between this circuit and the positive logic circuit.

(a) In the positive logic circuit the fan-out was determined by how much current there was available from a bottomed transistor after the other functions it has to perform were satisfied. This was because a *bottomed* transistor held *down* the AND gate voltage. Now, however, an AND gate has to be held *up* by the flip-flop

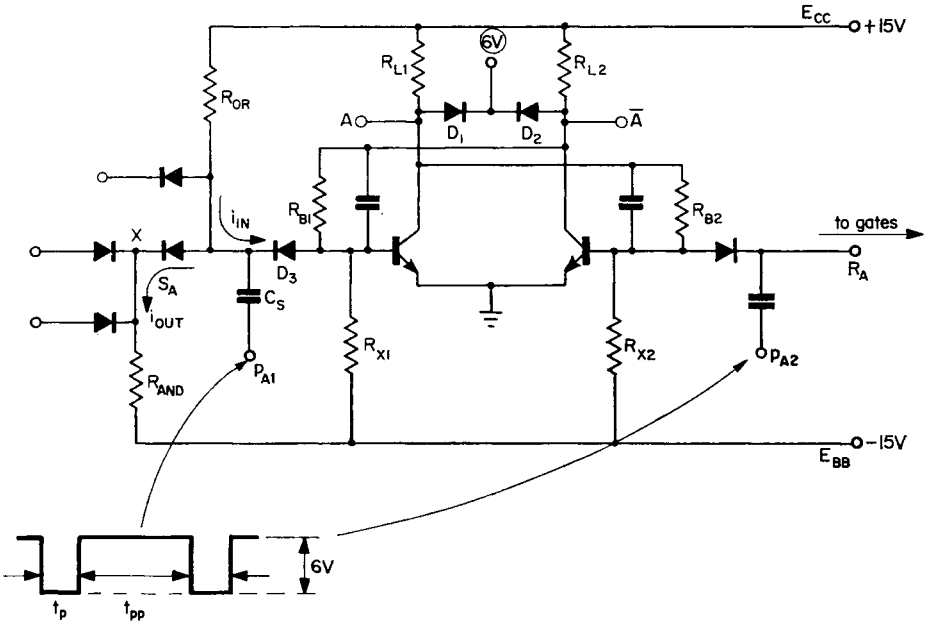


FIG. B.2. A practical circuit for a negative-logic *R–S* pulsed flip-flop with two levels of logic on each logic input. (Negative-going trigger pulses.)

output; consequently the fan-out is determined by how much current is available, from the collector load *resistor*, when the output transistor is *cut off*.

(b) Since the discharging and charging currents for C_S flow through the gates and the clock pulse generator, it is not necessary to include diodes D_4, D_5 to prevent its flowing in the transistor bias resistors R_X.

(c) In the positive logic flip-flop the capacitor C_S had to be capable of removing Q_{OFF} and of absorbing the base current which would otherwise flow into the transistor. Now it must in addition absorb the OR gate current.

The design now proceeds as follows:

The bias resistor R_B and R_X are the same as before and I_B is again 0·4 mA.

The value of C_S turns out to be 220 pF as before, since the increase due to I_{OR} is balanced by the fact that ΔV_S is now more nearly 6 V since the diodes D_5 and D_4 are removed.

The voltage of S_A just after the clock pulse will now be either about 0 or 6 V depending on whether T_1 is already bottomed or cut-off.

Thus R_{OR} is given by

$$i_{\text{IN}} \approx \frac{12}{R_{\text{OR}}} = \frac{C_s \Delta V_s}{t_{pp}} \approx 146\ \mu\text{A},$$

whence $R_{\text{OR}} = 82\ \text{k}\Omega$.

A suitable value of R_{AND} then becomes (using Thévenin again) $R_{\text{AND}} = 50\ \text{k}\Omega$.

Thus i_{OUT} falls from about 300 μA to about 120 μA as V_{SA} falls. The current necessary to hold an AND gate up is 21 V/ 50 kΩ = 0·42 mA.

The cross-coupling capacitors are the same as before.

The load resistors must now be as small as possible. When the transistors are bottomed all the collector current flows in

the load resistors. Thus

$$R_{L1} = R_{L2} = 1{\cdot}5\ \text{k}\Omega.$$

When the transistor is cut off and the collector voltage is rising the AND gate load must not be so great as to prevent the cross-coupling capacitor and the stray capacitance C_L from being charged up within 1 μsec. This is satisfied with a total AND gate load of about 4·7 kΩ giving a fan-out of about 10 (i.e. $10 \times R_{\text{AND}}$ in parallel). Again tolerance considerations are likely to reduce this by a factor of 2 or so.

Comparison between Positive and Negative Logic Circuits

The saving in diodes D_4 and D_5 in the negative logic circuit are paid for by a smaller fan-out and by an increase in power dissipation.

Because the load resistors in the negative logic circuit are so much less than those in the positive logic circuit, the dissipation in the circuit as a whole is much greater. Excluding the gates the dissipations of the two circuits are about 60 mW for the positive logic circuit and 210 mW for the other.

APPENDIX C

The Logical Design of a Sequence Detector

THIS Appendix is a worked example of the design of a sequential circuit using several different methods. The purpose is to throw the differences into clear relief and to compare the simplicity of the various techniques.

The problem is as follows. A sequential circuit has four inputs: x_1, x_2, x_3, x_4 and a two-level clock waveform P, $\bar{P}$, P, $\bar{P}$, etc. Only one x input is present at any one time. The x inputs occur in random sequence. They change immediately following a change from $\bar{P}$ to P in the clock waveform. The sequential circuit is to recognize (i.e. give an output) when the sequence x_2, x_3, x_4 appears in the input.

Method 1. Intuitive Design Using Shift Registers

The clock waveform is used to drive two shift registers, one of two stages, the other of only one stage. The input to the two-bit shift register is x_2, the input to the one-bit register is x_3, so that the sequence x_2, x_3, x_4 can be recognized by means of an AND gate between the outputs of the two registers and input x_4, as shown in Fig. C.1. The circuit operates by delaying x_2 by two bits and x_3 by one bit, so that the wanted sequence turns up as a coincidence.

The delay flip-flops can be either triggered flip-flops (see Fig. C.1) or level sensitive clocked flip-flops.

Consider particularly the master–slave flip-flop constructed from NOR gates (see problem 7.8). Each Delay flip-flop of the shift registers requires seven NOR gates. The final AND gate requires four more NOR gates (to invert x_2, x_3 (delayed) and x_4 and to combine them for the final output) and one more is needed to invert the clock waveform, making twenty-six NOR gates in all.

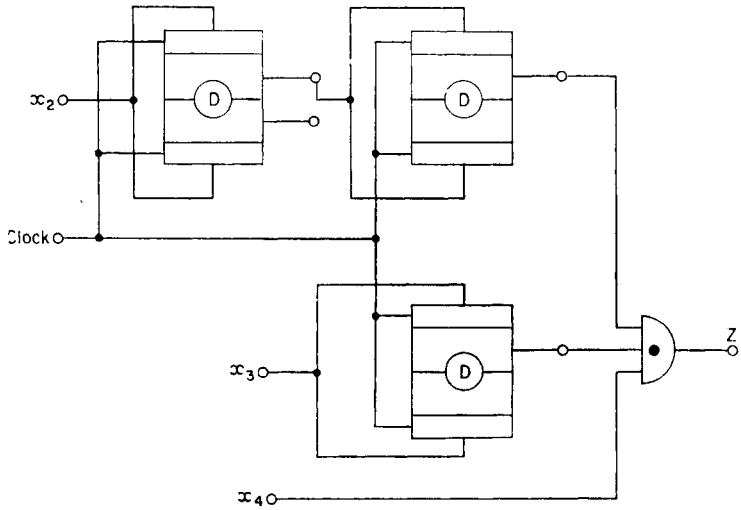
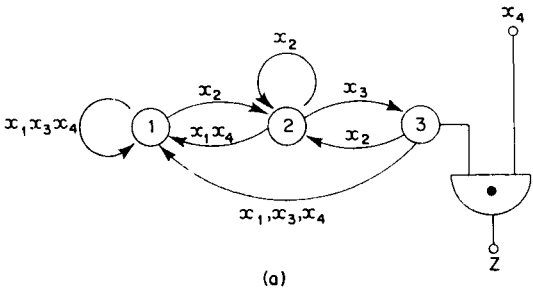


FIG. C.1. A sequence detector using delay flip-flops as shift register elements.

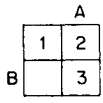
Method 2. Triggered or Master–Slave Flip-flops as Secondaries

Using triggered or pulsed flip-flops, the circuit can be designed according to the method described in Chapter 7.

The sequence diagram of Fig. C.2(a) shows three distinguishable states: (1) the resting state; (2) registering the arrival of x_2; (3) registering the arrival of x_2, x_3. Since these states are actually "next states", and are therefore predetermined during the clock period prior to their being set up, it is possible to obtain the required



(a)



(b)

	x_1	x_2	x_3	x_4	
1	1	2	1	1	$\bar{A}\bar{B}$
2	1	2	3	1	$A\bar{B}$
3	1	2	1	1	AB

(c)

x_1	x_2	x_3	x_4
rr	Sr	rr	rr
Rr	sr	sS	Rr
RR	sR	RR	RR

(d)

	x_1	x_2	x_3	x_4	
	r	S	r	r	$\bar{A}\bar{B}$
	R	s	s	R	$A\bar{B}$
	R	s	R	R	AB

flip flop A (e)

x_1	x_2	x_3	x_4
r	r	r	r
r	r	S	r
R	R	R	R

flip flop B

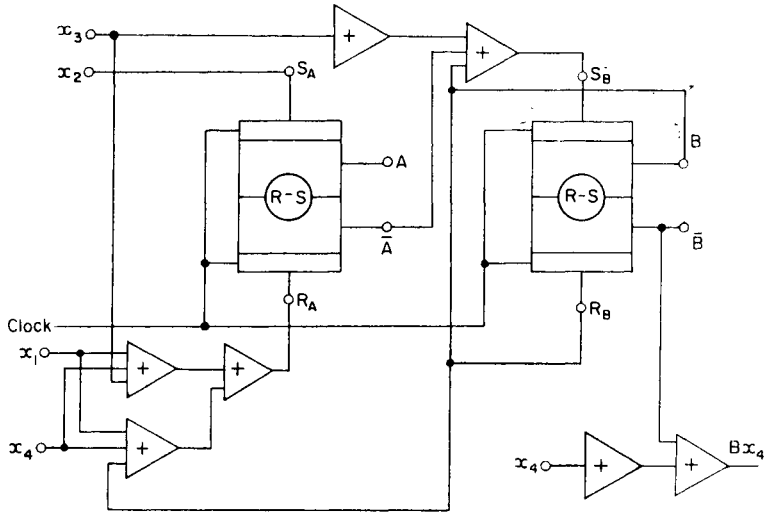


FIG. C.2. The design of the sequence detector using triggered or master–slave flip-flops as secondaries. (a) The sequence diagram; (b) the state assignment map; (c) the state excitation map; (d) the input excitation map; (e) separate maps for each flip-flop; (f) the resulting circuit using NOR gates and *R–S* flip-flops.

output for one clock period using an AND gate combining state 3 and input x_4, as shown. (It is also possible, but unnecessary, to distinguish a fourth state registering the arrival of sequence x_2, x_3, x_4.)

With the state assignment shown in Fig. C.2(b) the output will be given by $Z = Bx_4$. The state excitation becomes that shown in Fig. C.2(c), which leads to the input excitation maps of Fig. C.2(e). No races occur with triggered flip-flops so

$$\begin{aligned} S_A &= x_2 & S_B &= x_3A\bar{B}, \\ R_A &= x_1+x_4+Bx_3 & R_B &= B, \end{aligned} \tag{C.1}$$

(i.e. grouping the capital letters together with any convenient lower case ones of the same letter).

The circuit therefore requires only two R–S flip-flops as compared with three D flip-flops in method 1. (Indeed we may note in passing that method 1 is equivalent to choosing the state assignment $1 \equiv \bar{A}\bar{B}\bar{C}$, $2 \equiv A\bar{B}\bar{C}$, $3 \equiv \bar{A}BC$.)

Using clocked R–S flip-flops with nine NOR gates per flip-flop (see problem 7.7), five gates to implement eqns. (C.1) and two to generate the required output, as shown in Fig. C.2(f), involves a total of twenty-five NOR gates. (Actually one more can be saved by inverting the clock for both flip-flops with one inverter rather than one in each.)

The inputs to flip-flops are themselves gates, so it is to be expected that some further simplification of level sensitive circuits should be possible, as follows.

Method 3. Using d.c. Flip-flops or d.c. Feedback Paths

We now abandon the use of triggered flip-flops, or their master–slave equivalents, and regard the clock waveform as a fifth two-level input to the circuit. Consequently, we must now distinguish between stable and unstable states, each new state being set up immediately an input changes.

It is stated that the clock waveform changes from $\bar{P}$ to P shortly before an x input changes. Thus the kind of sequence of inputs to be expected is $x_1\ \bar{P}$, (x_1P), x_2P, $x_2\bar{P}$, (x_2P), x_3P, $x_3\bar{P}$, etc. Those shown bracketed are the brief input conditions occurring when the clock has changed state but the x input has not yet changed.

The behaviour of the circuit can now be mapped as in Fig. C.3(a). Beginning with stable state ①, say, the only possible change is $\bar{P}$ to P, producing unstable state 5 which then becomes stable at ⑤. Now the x input may change, leading via 6, 7 or 8 to ⑥, ⑦ or ⑧. When P changes back to $\bar{P}$, x's do not change and ⑤, ⑥, ⑦ or ⑧ lead back to ①, ②, ③ or ④.

When the circuit is in state ⑥ a change from x_2 to x_3 is the beginning of the required sequence and another stable state ⑨ must be set up. The remainder of the required sequence is specified in the bottom four rows of the map. The required output comes from stable state ⑫.

This sequence map can now be merged according to the lettering beside the map as shown in Fig. C.3(b). One race condition appears in column Px_2, but it is non-critical and can be resolved either by a cycle through ABx_2 or by leaving it as a race but specifying the excitation required in element ABx_2. The latter is slightly simpler in this case. The final input excitation maps for the two secondaries are shown in Fig. C.3(d) and lead to the following equations. The output is given by

$$Z = x_4BP. \tag{C.2}$$

(a) *For d.c. R–S flip-flops*

$$\left.\begin{aligned} S_A &= Px_2 & S_B &= \bar{P}x_3A \\ R_A &= x_1+x_4+\bar{P}x_2+Px_3B & R_B &= x_1+x_2+\bar{P}x_3\bar{A}+\bar{P}x_4 \end{aligned}\right\}. \tag{C.3}$$

No static hazards occur.

Using NOR gates only again, we can count two per flip-flop and thirteen to implement eqns. (C.3) as shown in Fig. C.3(e).

Generating the output $Z = BPx_4$ requires three more NOR gates, making twenty in all.

(b) *For d.c. feedback paths*

$$\left.\begin{aligned} Y_A &= Px_2 + \bar{P}x_3A + x_3A\bar{B} + (\bar{x}_1\bar{x}_4A\bar{B}P), \\ Y_B &= \bar{P}x_3A + Px_3B + (\bar{x}_1\bar{x}_2\bar{A}BP) + (x_3AB). \end{aligned}\right\} \quad \text{(C.4)}$$

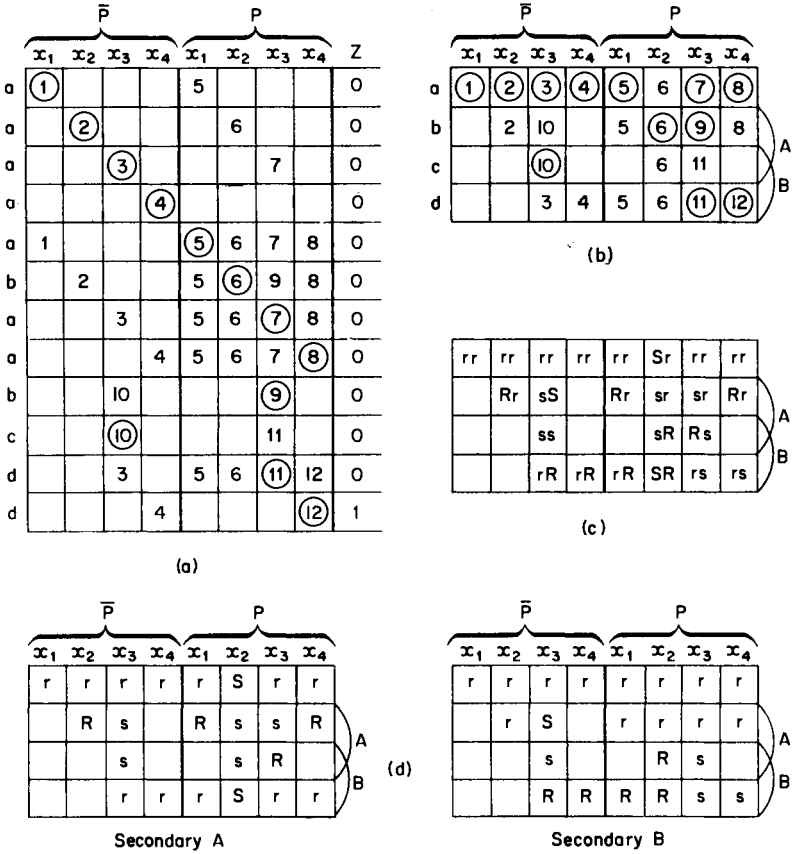


FIG. C.3. The design of the sequence detector using level sensitive circuits. (a) The sequence map (an x input changes shortly after the clock waveform has changed from $\bar{P}$ to P); (b) the merged state excitation map; (c) and (d) the input excitation maps.

The terms in brackets are to prevent static hazards, they are derived as follows. Static hazards can occur when the state of secondary should *not* change even when an input does change, and they arise because a hiatus between the two states of an input may exist. Thus in this case there may exist brief intervals of time when $P=\bar{P}=0$ or when one x input has gone off and no other x input has come on. These hazards can be recognized in Karnaugh maps, for which the Boolean groupings of inputs have been selected, by noting *those changes in state of the circuit which involve a secondary remaining in the "1" state and which involve transferring from one Boolean grouping to another.* Thus only lower case s's are involved.

In this problem circuit transitions which involve transitions from one lower case s to another are, for secondary A

(i) ⑨ to ⑩ via 10,
(ii) ⑥ to ⑨,
(iii) 6 to ⑥.

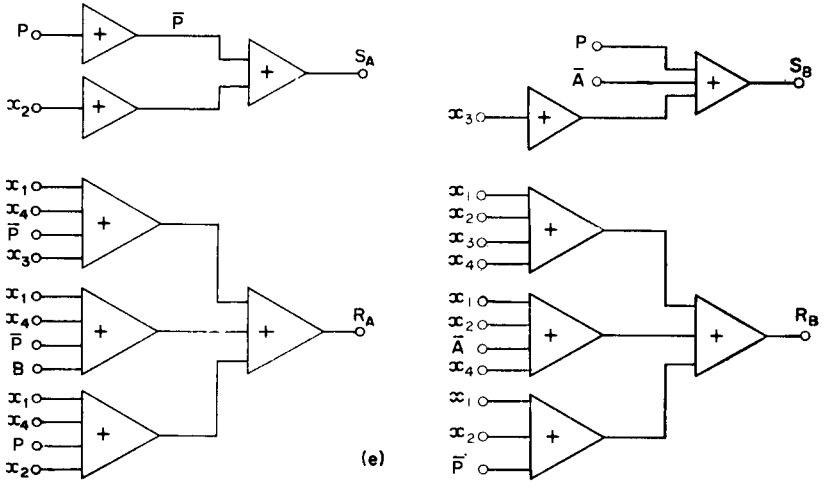


FIG. C.3. (e) the gates for the inputs to R–S flip-flops constructed from NOR gates.

These can be found by comparing maps 3.C(d) (for secondary A) and 3.C(b).

Of these transitions only ⑥ to ⑨ is hazardous. It involves leaving grouping Px_2 and entering $x_3A\bar{B}$. Evidently, if for a moment x_2 has dropped to zero and x_3 has not yet risen to the "1" state $Y_A = A$ will drop to zero, so that $x_3A\bar{B}$ will never rise as it should. The last term added to eqn. (C.4) for Y_A places this particular transition in one new grouping.

In the case of secondary B, the hazardous transition is ⑩ to 11. But note too that the grouping $(\bar{x}_1\bar{x}_2\bar{A}BP)$ includes the element $x_4P\bar{A}B$, in a hazard-free grouping. The grouping x_4PB is a simpler term which also includes element x_4PAB, but it leaves the transition ⑪ to ⑫ hazardous and is not therefore used.

Now realizing eqns. (C.4) in NOR gates, by map factoring, for example, involves manipulation of seven Boolean variables which is beyond any technique discussed in this book. Even to use the simple technique described in p. 96 is too difficult. We should put eqns. (C.4) in the form of the product-of-sums rather than the sum-of-products, but we cannot do this with seven variables. However, one technique, which in general is not so efficient, remains to us; to generate $\bar{A}$ and then invert.

Using de Morgan's theorem

$$\bar{Y}_A = (\bar{P}+\bar{x}_2)(P+\bar{x}_3+\bar{A})(\bar{x}_3+\bar{A}+B)(x_1+x_4+\bar{A}+B+\bar{P}).$$

This now requires nine NOR gates, one for each bracket, one to combine the brackets to produce $\bar{Y}_A = \bar{A}$, one to invert this output to produce A and one each to invert P, x_2 and x_3.

Similarly

$$\bar{Y}_B = (P+\bar{x}_3+\bar{A})(\bar{P}+\bar{x}_3+\bar{B})(\bar{x}_3+\bar{A}+\bar{B})(x_1+x_2+A+\bar{B}+\bar{P}).$$

This requires five more NOR gates. Three for the brackets not already generated in $\bar{Y}_A$, one to combine them and one to invert $\bar{B}$.

Finally, three more NOR gates are needed in order to give the output, $Z = x_4PB$ (one to invert x_4, one to generate the function $\bar{Z} = (\bar{x}_4 + \bar{P} + \bar{B})$ and one to invert Z).

Thus the total is now seventeen NOR gates (with a maximum fan-in of 5).

Thus by progressively simplifying the type of secondary used we have reduced the number of circuit elements involved, but have considerably increased the complexity of both the design procedures, and the interconnections between NOR gates. As to which is in the end the most economical is a question which can only be answered in the context of a particular production situation.

INDEX