Lecture-10
BJT Switching Characteristics, Small Signal Model

BJT Switching Characteristics: The circuit in Fig.1(b) is a simple CE switch. The input voltage waveform v_s shown in the Fig.1(a) is used to control the state of the switch (between collector and emitter). For $t < T_1$, $v_s = -V_1$ and the emitter-base diode is reverse-biased. If we neglect the reverse-current components, since the collector-base diode is also reverse-biased, the BJT is cut-off and no current exists anywhere in the circuit. Consequently, $v_o = V_{CC}$, and with $i_C = 0$, this is an open switch. Actually, $i_C \approx I_{CO}$ and $v_o = V_{CC} - I_{CO}R_L$. However, with I_{CO} of the order of a nano-ampere and R_L of the order of kilohms, v_o differs from V_{CC} by only a few microvolts. Thus, for practical purposes, $v_o = V_{CC}$. The input voltage becomes V_2 for $T_1 < t < T_2$. The value of V_2 is selected to ensure that the BJT is at least at the edge of saturation. From Table-1 in LN-7, $v_{CE} = v_o = V_{CE(sat)} \leq 0.3\ V$ and $i_C = (V_{CC} - V_{CE(sat)})/R_L$; these values approximate the closed switch. Note that the current in the closed switch is determined by the external elements V_{CC} and R_L. For $V_{CC} \gg 0.3\ V$, $i_C = V_{CC}/R_L$. At $t = T_2$, the input wave form switches back to V_1, eventually causing the BJT to return to cutoff. Sketches of both v_o and i_C are depicted in Fig.1. The causes of the switching transients are described later in this section. The nature of the switching characteristics is readily discernible from the transfer characteristic, a graph of v_o versus v_s, for the circuit.

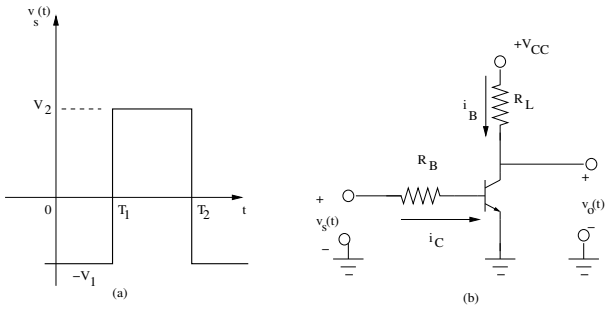

Figure 1: (a) Input Waveform applied to BJT (b) A simple BJT used as switch in CE configuration

BJT Switching Speed: Our description of the circuit in Fig. 1(a) and (b) at the beginning of this section focused on the ON and OFF states of the switch. We now consider the switching transients indicated in the waveforms in Fig.2. As seen in Fig.2, the current does not immediately respond to the input signal. Instead, there is a delay, and the time that elapses during this delay, together with the time required for the current to rise to 10 percent of it maximum (saturation) value, is

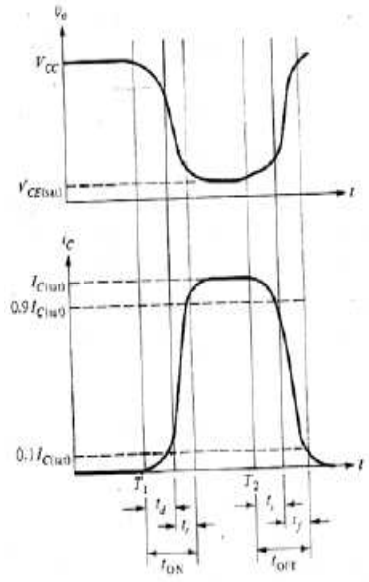

Figure 2: The waveforms for v_o and i_c displaying the rise time, fall time, delay and storage time during switching

called the *delay time* t_d. The current waveform has a nonzero rise time t_r, which is the time required for the current to rise through the active region from 10 to 90 percent of $I_{C(sat)}$. The total *turn-on time* t_{ON} is the sum of the delay and rise time, $t_{ON} \equiv t_d + t_r$. When the input signal returns to its initial state at $t = T_2$, the current again fails to respond immediately. The interval which elapses between the transition of the input waveform and the time when i_C has dropped to 90 percent of I_{CS} or I_{C0} is called the *storage time* t_s. The storage interval is followed by the *fall time* t_f, which is the time required for i_C to fall from 90 to 10 percent of $I_{C(sat)}$. The *turn-off time* t_{OFF} is defined as the sum of the storage and fall times, $t_{OFF} \equiv t_s + t_f$. We shall consider now the physical reasons for the existence of each of these times. The exact calculations of these times are complex; Three factors contribute to the delay time: (1) when the driving signal is applied to the transistor input, a nonzero time is required to charge up the emitter-junction transition capacitance so that the transistor may be brought from cutoff to the active region; (2) even when the transistor has been brought to the point where minority carriers have begun to cross the emitter junction into the base, a time interval is required before these carriers can cross the base region to the collector junction and be recorded as collector current; and (3) some time is required for the collector current to rise to 10 percent of its maximum.

The *rise time* and the *fall time* are due to the fact that if a base-current step is used to saturate the transistor or return it from saturation to cutoff, the transistor collector current must traverse the active region. The collector current increases or decreases along an exponential curve whose time constant T_r. The failure of the transistor to respond to the trailing edge of the driving pulse for the time interval ts results from the fact that a transistor in saturation has excess minority carriers stored in the base. The transistor cannot respond until this excess charge has been removed. Consider that the transistor is in its saturation region and that at $t = T_2$ an input step is used to turn the transistor off, as in Fig.1. Since the turn-off process cannot begin until the abnormal carrier density has been removed, a relatively long storage delay time t_s may elapse before the transistor responds to the turn-off signal at the input. In an extreme case this storage-time delay may be several times the rise or fall time through the active region. It is clear that when transistor switches are to be used in an application where speed is at a premium, it is advantageous to reduce the storage time. A method for preventing a transistor from saturating, and thus eliminating storage time, is the use of a *Schottky diode* in conjunction with the BJT.

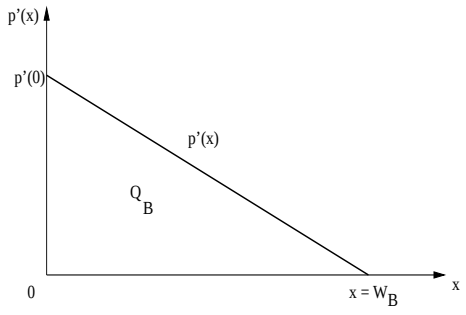

Figure 3: Distribution of excess minority carriers in the base of a *pnp* transistor

Stored Charge and Delay time t_d (or Transit Time t_{tr}): We have seen that BJT does not respond instantaneously to the fast changing signals. This is because BJT speed of response is limited mainly by the storage or diffusion capacitance, which accompanies the storage of minority carriers in the base. Let W_B be the width of the base $\ll$ diffusion length L_p or L_n of minority carriers so that distribution is linear in both active and saturation modes. Let the linear distribution of minority carriers is as shown in Fig.3, for the operation of active region. For *pnp* the hole current is the majority current, therefore excess hole current is related as

$$J_C = -qD_p\frac{dp'}{dx} \tag{1}$$

$$\Rightarrow I_C = -AqD_p\frac{dp'}{dx} \quad (2)$$

$$\Rightarrow \frac{dp'}{dx} = -\frac{p'(0)}{W_B} \quad (3)$$

$$\Rightarrow I_C = \frac{qAD_p p'(0)}{W_B} \quad (4)$$

The excess minority carrier charge stored in the base is given by

$$\begin{aligned} Q_B &= \text{concentration} \times \text{volume} \\ &= \text{area under the curve shown in Fig.2} \times qA \\ &= \left(\frac{p'(0)W_B}{2}\right) \times qA \end{aligned} \quad (5)$$

Now taking the ratio of Eqns.(5) and (4) we get,

$$\frac{Q_B}{I_C} = \frac{W_B^2}{2D_p} \equiv \tau_B \quad (6)$$

Let us find the relationship between 'τ_B' and approximate time it takes a hole to travel from emitter junction to the collector junction.

$$\frac{dx}{dt} = v(x) = \text{ velocity of the hole} \quad (7)$$

Let $t_d \equiv t_{tr}$ be the time for the hole to cross base of width W_B. Therefore, we have

$$t_{tr} = \int_0^{W_B} dt = \int_0^{W_B} dx/v(x) \quad (8)$$

The current due to holes is given by

$$\begin{aligned} I_p &= -qAD_p dp'/dx \\ &= qAD_p p'(0)/W_B \end{aligned} \quad (9)$$

This current can also be written in terms of velocity of the carriers. Thus we have,

$$I_p = qp'(x)Av(x) \quad (10)$$

But equation for concentration gradient as shown in the Fig.2 is given by

$$p'(x) = p'(0)\left(1 - \frac{x}{W_B}\right) \quad (11)$$

Substituting Eqn.(11) in (10) we get,

$$I_p = qAv(x)p'(0)\left(1 - \frac{x}{W_B}\right) \tag{12}$$

Now equating Eqn.(9) and (10) we get,

$$v(x) = \frac{D_p}{\left(1 - \frac{x}{W_B}\right)} \tag{13}$$

Therefore,

$$t_{tr} = \int_0^{W_B} \frac{\left(1 - \frac{x}{W_B}\right)}{D_p} dx = \frac{W_B^2}{2D_p} \tag{14}$$

Thus $t_{tr} = \tau_B$.

Charge Control Relations: The general expression for the excess hole density charge stored in the base of *pnp* transistor, Q_B is given by

$$Q_B = qA \int_0^{W_B} p'(x)dx \tag{15}$$

where $p'(x)$ is the excess hole density at x. $p'(x) = p(x) - p_0$ that implies $dp'(x) = dp(x)$. From the continuity equation for the holes we have,

$$\begin{aligned} \frac{\partial p}{\partial t} &= \frac{p_0 - p}{\tau_p} - \frac{1}{q}\frac{\partial J_p}{\partial x} \\ \Rightarrow \frac{\partial p'}{\partial t} &= -\frac{p'(x)}{\tau_p} - \frac{1}{q}\frac{\partial J_p}{\partial x} \end{aligned} \tag{16}$$

Multiplying both sides by 'qA' and integrating from 0 to W_B with respect to x we get

$$\frac{\partial}{\partial t}\int_0^{W_B} p'(x)qA\ dx = -\int_0^{W_B} \frac{qAp'(x)}{\tau_p} dx - \int_0^{W_B} qA\partial J_p \tag{17}$$

The first integral in the above equation represents the rate of change of the stored charge, Q_B with time; the second integral is the ration of the stored charge to the life-time of holes; and the third integral becomes $i_C + i_E = -i_B(t)$. Thus equation can be written as,

$$\begin{aligned} \frac{dQ_B}{dt} &= -\frac{Q_B(t)}{\tau_p} - (-i_B(t)) \\ \Rightarrow i_B(t) &= \frac{Q_B(t)}{\tau_p} + \frac{dQ_B(t)}{dt} \end{aligned} \tag{18}$$

Turn-ON Time: The transistor switches from cut off to saturation by application of a step of base current. Let $i_B(t) = I_B$ for certain duration of time. Then the Eqn.(18) becomes,

$$I_B = \frac{Q_B(t)}{\tau_p} + \frac{dQ_B(t)}{dt} \tag{19}$$

Therefore solving this first-order differential equation, and using the initial condition at $t = 0$, $Q_B = 0$.

$$Q_B(t) = I_B\tau_p(1 - e^{-t/\tau_p}) \tag{20}$$

$$i_C(t) = \frac{Q_B(t)}{\tau_B} = \frac{I_B\tau_p}{\tau_B}(1 - e^{-t/\tau_p}) \tag{21}$$

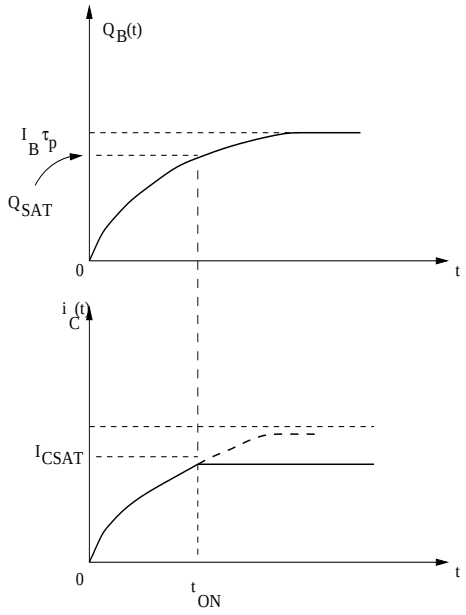

Figure 4: Sketches of $Q_B(t)$ and $i_C(t)$

As $t \to \infty$, implies $Q_B(\infty) = I_B\tau_p$ and $i_C(\infty) = \frac{I_B\tau_p}{\tau_B}$. We define Q_{SAT} is the value of Q_B at $i_C = I_{CSAT}$. Thus we have,

$$i_C(t) = \frac{I_B\tau_p}{\tau_B}(1 - e^{-t/\tau_p}) \text{ for } Q_B \leq Q_{SAT} \tag{22}$$

Therefore at $t = t_{ON}$ $i_C(t_{ON}) = I_{CSAT} = Q_{SAT}/\tau_B$

$$I_{CSAT} = \frac{I_B\tau_p}{\tau_B}(1 - e^{-t_{ON}/\tau_p}) = \frac{V_{CC}}{R_L} \tag{23}$$

Thus the expression for turn-ON time is given by

$$t_{ON} = \tau_p \ln\left(\frac{1}{1 - \left(\frac{V_{CC}}{I_B R_L}\right)\left(\frac{\tau_B}{\tau_p}\right)}\right) \tag{24}$$

Turn-OFF Time: To turn a transistor OFF, the excess stored charge in the base must be removed and the collector current must be made almost zero ($= I'_{C0}$ or I_{CEO}). This is done by making $i_B(t) = 0$. Thus Eqn.(18) becomes,

$$\frac{dQ_B}{dt} = -\frac{Q_B}{\tau_p} \tag{25}$$

The solution of the above first-order homogeneous differential equation is given by,

$$Q_B(t) = Q_B(0)e^{-t/\tau_p} \text{ for } t > 0 \tag{26}$$

The turn off time defined as the time required to reduce the collector current to almost zero, is made up of two increments: the time it takes Q_B to reach Q_{SAT}, is known as the *storage time* t_s and second, the time t_f, it takes collector current to reach zero, or more practically to a value of about $0.1I_{CSAT}$. The decrease of stored charge and current are shown in the Fig.5. At $t = t_s$ we have $Q_B(t_s) = Q_{SAT}$. For $t > t_s$, the transistor is in active region, so that i_C is given by

$$i_C(t) = Q_B(t)/\tau_B = (Q_{SAT}/\tau_B)e^{-t/\tau_p} = I_{CSAT}e^{-t/\tau_p} \tag{27}$$

The storage time t_s is found from,

$$Q_B(t_s) = Q_{SAT} = Q_B(0)e^{-t_s/\tau_p} \tag{28}$$

Solving for t_s, we obtain

$$t_s = \tau_p \ln\left(\frac{Q_B(0)}{Q_{SAT}}\right) \tag{29}$$

The BJT Small-Signal Model: The circuit shown in the Fig.6 is an elementary CE amplifier stage. The capacitor C_B (called a *blocking capacitor*) is used to isolate the dc bias from the signal source $v_s = V_{sm}\sin\omega t$ and its source resistance R_s. This capacitor acts as an open-circuit under the quiescent conditions (no input signal) because the reactance of a capacitor is infinite at zero frequency (dc). We assume that at the angular frequency of the signal, the reactance of C_B is sufficiently small compared with R_s that the series combination of these elements is R_s. Consequently, the effect of the capacitor on the signal transmitted from the source v_s to the amplifier input is negligible. The amplitude V_{sm} is chosen to provide the base current $i_b = I_{bm}\sin\omega t$. The total instantaneous base current i_B is the superposition

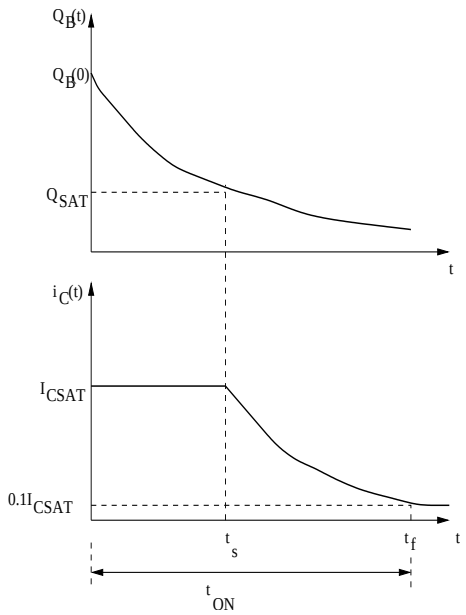

Figure 5: Sketches of $Q_B(t)$ and $i_C(t)$

of the dc bias level and signal current. Hence,

$$i_B = I_{BQ} + i_b = I_{BQ} + I_{bm} \sin \omega t \ A \tag{30}$$

As seen from the Fig.7, the effect of this signal causes both i_C and v_{CE} to vary (approximately) sinusoidally about their quiescent levels. These quantities are expressible as,

$$\begin{aligned} i_C &= I_{CQ} + i_b \\ &= I_{CQ} + I_{cm} \sin \omega t \ A \qquad (31) \\ v_{CE} &= V_{CEQ} + v_{ce} \\ &= V_{CEQ} + V_{cem} \sin \omega t \ V \qquad (32) \end{aligned}$$

The small signal equivalent circuit of the BJT is shown in the Fig.8(a). The elements forming the equivalent circuits relate the *changes in voltages and currents about the operating point*. Each element in the model is a function of the quiescent voltages and currents established by the bias. As the input signal causes the changes, the equivalent circuit permits us to relate the output signal to the input signal. The *hybrid-π equivalent circuit* for CE connected BJT is shown the Fig.8(a). We can identify the elements in this model with coupled-diode representation of the

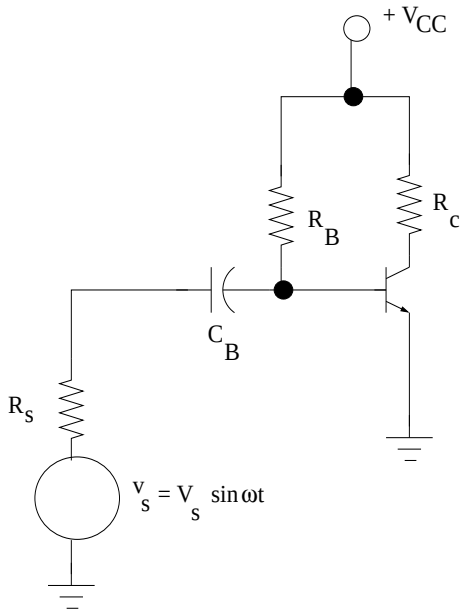

Figure 6: CE Amplifier

transistor. The forward-biased emitter-base junction is modeled by r_π and C_π where C_π is essentially the diffusion capacitance and r_π is the incremental resistance of the emitter-base diode.The capacitance C_μ is the depletion capacitance of the reverse-biased collector-base junction. The incremental resistance r_μ is shown by the dashed line. This resistance accounts for the feedback (base width modulation) between the input and output due to Early effect. r_μ is very large and usually neglected. Coupling between the junction is modeled by the controlled current source $g_m v_\pi$ and is proportional to the input current i_b. The output resistance r_o is the result of Early effect and equals the reciprocal of the slope of the dashed lines in the Fig.3 of LN-7. The resistance r_b is the *base spreading resistance* and accounts for the voltage drop in the path between the base contact and the active base region under emitter. Because of the larger cross-sectional area of the collector region, the *collector-spreading* resistance is in the order of 1 Ω and is usually neglected.

The Low Frequency Model: From the LN-9 we find the expressions for C_π (or C_D) and C_μ (or C_T) are dependent on the operating point values of the BJT voltages and currents. At typical quiescent levels, for both IC and low-power discrete transistors, values of C_π are in the order of tens of picofarads to one or two hindered picofarads. Values of C_μ are generally a few picofarads (1 to 5 pF). At low signal frequencies, the reactance of both capacitors are extremely high. At such frequencies the effects of C_π and C_μ can be neglected and consequently act as open circuit. This

leads to the low frequency model shown in the Fig.8(b). We observe that,

$$v_\pi = r_\pi i_b \tag{33}$$

For $v_{ce} = 0$, no currents exists in r_o and,

$$i_c = g_m v_\pi = g_m r_\pi i_b \tag{34}$$

$$\Rightarrow \frac{i_c}{i_b} = g_m r_\pi \tag{35}$$

It is convenient to introduce,

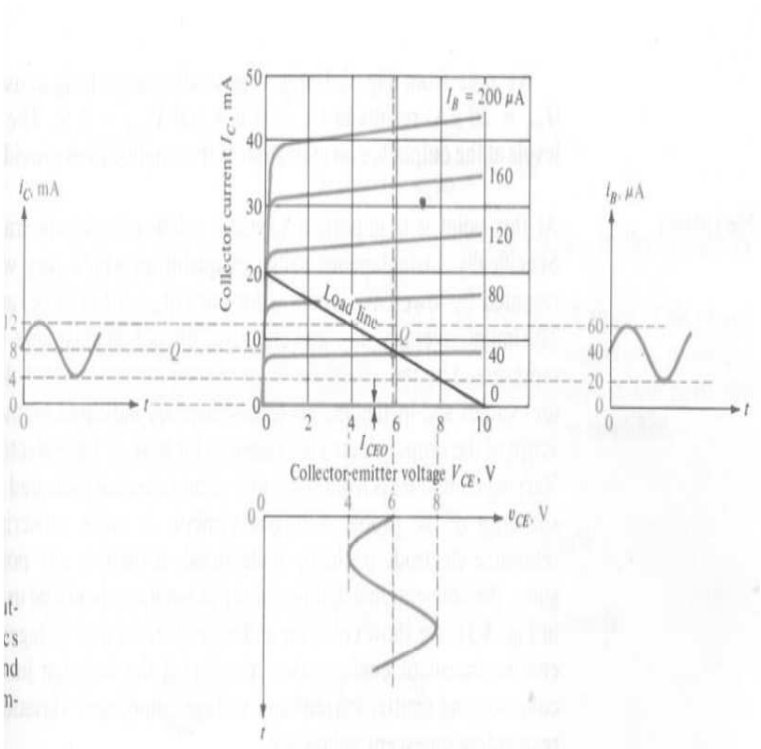

Figure 7: CE output characteristics showing load line and sinusoidal signal components

$$\begin{aligned} \beta_o &= \left.\frac{\Delta i_C}{\Delta i_B}\right| V_{CE} = \text{ const} = V_{CEQ} \\ &= \left.\frac{i_c}{i_b}\right| v_{ce} = 0 \end{aligned} \tag{36}$$

The parameter β_o is the incremental (ac) *common-emitter forward short-circuit current gain* and is evaluated at the operating point. The constant value of v_{CE} is indicative of incremental change in the quantity, and thus only $v_{ce} = 0$. (The condition $v_{ce} = 0$ and $i_c \neq 0$ represents a short circuit between collector and emitter relative to the *signal*. It does not, however, indicate a physical short-circuit connection between these terminals.) Thus we have,

$$\beta_o = g_m r_\pi \tag{37}$$

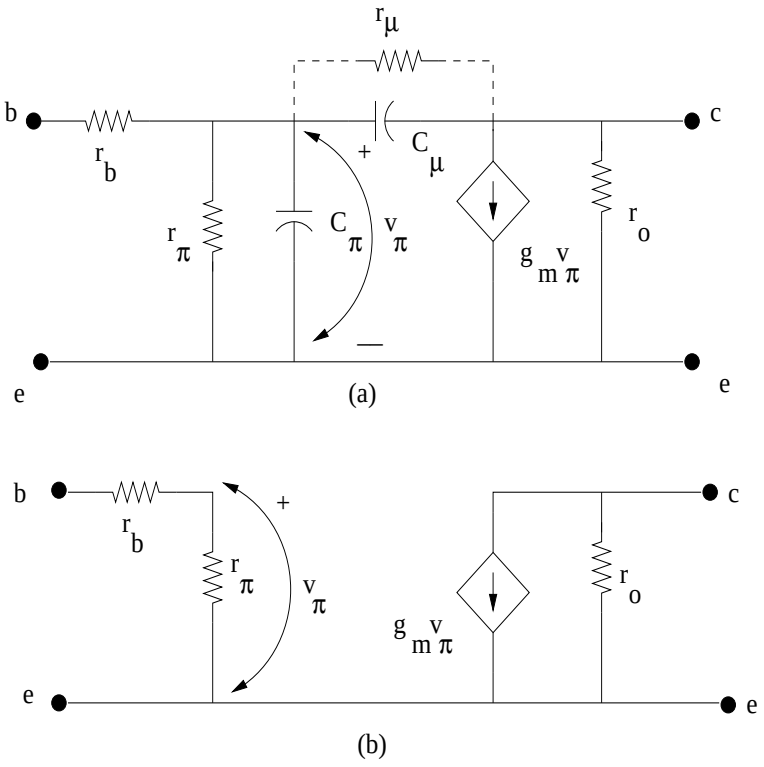

Figure 8: (a) Small Signal Hybrid-π equivalent circuit of BJT. (b) The Low frequency Hybrid-π

The parameter $g_m = i_c/v_\pi$, called *transconductance*, reflects the incremental changes in i_C about the operating point produced by the incremental change in the emitter-base voltage. The voltage drop $i_b r_b$ is small so that changes in the base-emitter voltage can be assumed to appear across the junction . Quantitatively, g_m

is expressible as,

$$\begin{aligned} g_m &= \frac{\Delta i_C}{\Delta v_{BE}} \Bigg| \, V_{CE} = \text{ const} = V_{CEQ} \\ &= \frac{\partial i_C}{\partial v_{BE}} \Bigg| \, v_{ce} = 0 \end{aligned} \tag{38}$$

We know that $i_C = -\alpha_F i_E$ for either *npn* or *pnp* transistor and Eqn.(38) becomes,

$$g_m = -\alpha_F \frac{\partial i_E}{\partial v_{BE}} \Bigg| \, v_{ce} = 0 \tag{39}$$

We wish to relate g_m to the conductance of the emitter-base diode. The incremental conductance of the diode is given by, Eqn.(3) of LN-8, as

$$g_d = \frac{di_D}{dv_D} \tag{40}$$

where i_D and v_D are the forward current and voltage of the diode. For an *npn* transistor, v_{BE} forward-biases the emitter diode and $v_{BE} = v_D$; however i_E is in the opposite direction of i_D (from n to p) so that $i_E = -i_D$. Therefore, $\frac{\partial i_E}{\partial v_{BE}} = -\frac{di_D}{dv_D}$ and

$$g_m = \alpha_F g_d \tag{41}$$

Eqn.(41) remains valid for *pnp* transistor because forward-biasing the emitter junction makes $i_E = i_D$ and $v_{BE} = -v_D$. The emitter-diode conductance g_d is expressed in the Eqn.(5) of LN-8, with $\eta = 1$. Hence $g_d = -I_{EQ}/V_T$ for an *npn* transistor and $g_d = +I_{EQ}/V_T$ for an *pnp* device. For *npn* (*pnp*) transistor, I_{EQ} is negative (positive); thus g_d is positive in both the instances and can written as $g_d = |I_{EQ}|/V_T$, thus we obtain the following simple expression for the transconductance:

$$\begin{aligned} g_m &= \frac{\alpha_F |I_{EQ}|}{V_T} \\ &= \frac{|I_{CQ}|}{V_T} \end{aligned} \tag{42}$$

Figure of Merit: A measure of the quality of a high frequency transistor is its figure of merit f_T. As we shall see f_T is a measure of the ratio of g_m to the total capacitance of the transistor. By neglecting r_μ, r_b we will determine an expression for the short-circuit gain $\beta(j\omega)$, of the transistor. It is defined as the ratio of the current I_o in a short-circuit placed at the output, as shown the Fig.9. Because of the short circuit, C_μ is in parallel with C_π. By neglecting the current in C_μ compared

to $g_m v_\pi$, I_o is $g_m v_\pi$ and v_π is $I_i r_\pi/[1 + j\omega r_\pi(C_\pi + C_\mu)]$. The current gain is given by

$$\frac{I_o}{I_i} = \frac{\beta_o}{1 + j\omega r_\pi(C_\pi + C_\mu)} \tag{43}$$

At high frequencies, the magnitude of the imaginary part of the denominator of Eqn.(43) is much greater than unity, so that

$$|\beta(j\omega)| = \frac{g_m}{\omega(C_\pi + C_\mu)} \tag{44}$$

where $g_m = \beta_o/r_\pi$. The symbol, f_T is defined as the frequency at which the magnitude of the short-circuit current gain is unity, so that at $f = f_T$, $|\beta(j\omega)| = 1$ and

$$f_T = \frac{g_m}{2\pi(C_\pi + C_\mu)} \tag{45}$$

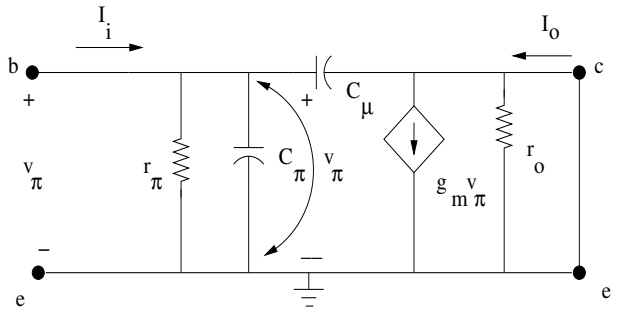

Figure 9: Circuit for calculating f_T of a transistor