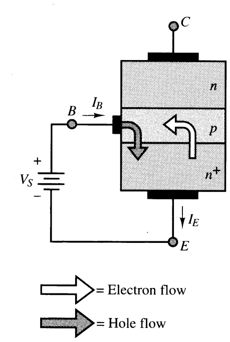
# Bipolar-Junction (BJT) transistors

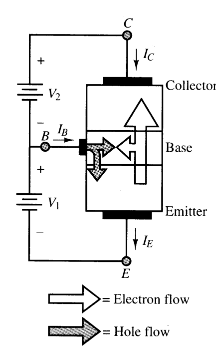
**References:**
**Barbow (Chapter 7), Hayes & Horowitz (pp 84-141), Rizzoni (Chapters 8 & 9)**

A bipolar junction transistor is formed by joining three sections of semiconductors with alternatively different dopings. The middle section (base) is narrow and one of the other two regions (emitter) is heavily doped. Two variants of BJT are possible: NPN and PNP.


We will focus on NPN BJTs. Operation of a PNP transistor is analogous to that of a NPN transistor except that the role of "majority" charge carries reversed. In NPN transistors, electron flow is dominant while PNP transistors rely mostly on the flow of "holes." Therefore, to zeroth order, NPN and PNP transistors behave similarly except the sign of current and voltages are reversed. *i.e.,* PNP = − NPN ! In practice, NPN transistors are much more popular than PNP transistors because electrons move faster in a semiconductor. As a results, a NPN transistor has a faster response time compared to a PNP transistor.



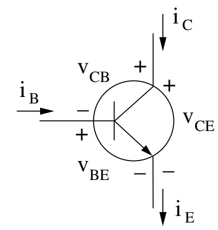

At the first glance, a BJT looks like 2 diodes placed back to back. Indeed this is the case if we apply voltage to only two of the three terminals, letting the third terminal float. This is also the way that we check if a transistor is working: use an ohm-meter to ensure both diodes are in working conditions. (One should also check the resistance between CE terminals and read a vary high resistance as one may have a burn through the base connecting collector and emitter.)




The behavior of the BJT is different, however, when voltage sources are attached to both BE and CE terminals. The BE junction acts like a diode. When this junction is forward biased, electrons flow from emitter to the base (and a small current of holes from base to emitter). The base region is narrow and when a voltage is applied between collector and emitter, most of the electrons that were flowing from emitter to base, cross the narrow base region and are collected at the collector region. So while the BC junction is reversed biased, a large current can flow through that region and BC junction does not act as a diode.

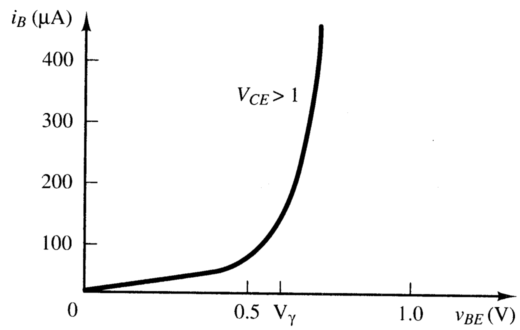
The amount of the current that crosses from emitter to collector region depends strongly on the voltage applied to the BE junction, $v_{BE}$. (It also depends weakly on voltage applied between collector and emitter, $v_{CE}$.) As such, small changes in $v_{BE}$ or $i_B$ controls a much larger collector current $i_C$. Note that the transistor does not generate $i_C$. It acts as a valve controlling the current that can flow through it. The source of current (and power) is the power supply that feeds the CE terminals.

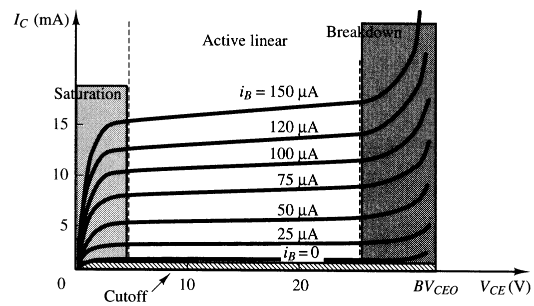
A BJT has three terminals. Six parameters; $i_C$, $i_B$, $i_E$, $v_{CE}$, $v_{BE}$, and $v_{CB}$; define the state of the transistor. However, because BJT has three terminals, KVL and KCL should hold for these terminals, *i.e.,*




$$i_E = i_C + i_B \qquad\qquad v_{BC} = v_{BE} - v_{CE}$$

Thus, only four of these 6 parameters are independent parameters. The relationship among these four parameters represents the "$iv$" characteristics of the BJT, usually shown as $i_B$ vs $v_{BE}$ and $i_C$ vs $v_{CE}$ graphs.




The above graphs show several characteristics of BJT. First, the BE junction acts likes a diode. Secondly, BJT has three main states: cut-off, active-linear, and saturation. A description of these regions are given below. Lastly, The transistor can be damaged if (1) a large positive voltage is applied across the CE junction (breakdown region), or (2) product of $i_C v_{CE}$ exceed power handling of the transistor, or (3) a large reverse voltage is applied between any two terminals.

Several "models" available for a BJT. These are typically divided into two general categories: "large-signal" models that apply to the entire range of values of current and voltages, and "small-signal" models that apply to AC signals with small amplitudes. "Low-frequency" and "high-frequency" models also exist (high-frequency models account for capacitance of each junction). Obviously, the simpler the model, the easier the circuit calculations are. More complex models describe the behavior of a BJT more accurately but analytical calculations become difficult. PSpice program uses a high-frequency, Eber-Mos large-signal model which is a quite accurate representation of BJT. For analytical calculations here, we will discuss a simple low-frequency, large-signal model (below) and a low-frequency, small-signal model in the context of BJT amplifiers later.

**A Simple, Low-frequency, Large Signal Model for BJT:**

As the BE junction acts like a diode, a simple piece-wise linear model can be used :

**BE Junction ON:** $v_{BE} = v_\gamma$, and $i_B > 0$

**BE Junction OFF:** $v_{BE} < v_\gamma$, and $i_B = 0$

where $v_\gamma$ is the forward bias voltage ($v_\gamma \approx 0.7$ V for Si semiconductors).

When the BE junction is reversed-biased, transistor is OFF as no charge carriers enter the base and move to the collector. The voltage applied between collector and emitter has not effect. This region is called the **cut-off** region:

**Cut-Off:** $v_{BE} < v_\gamma, \qquad i_B = 0, \qquad i_C \approx i_E \approx 0$

Since the collector and emitter currents are very small for any $v_{CE}$, the effective resistance between collector and emitter is very large (100's of M$\Omega$) making the transistor behave as an open circuit in the cut-off region.

When the BE junction is forward-biased, transistor is ON. The behavior of the transistor, however, depends on how much voltage is applied between collector and emitter. If $v_{CE} > v_\gamma$, the BE junction is forward biased while BC junction is reversed-biased and transistor is in **active-linear** region. In this region, $i_C$ scales linearly with $i_B$ and transistor acts as an amplifier.

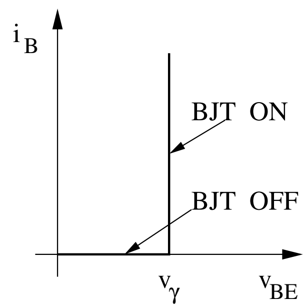
**Active-Linear:** $v_{BE} = v_\gamma, \qquad i_B > 0, \qquad \dfrac{i_C}{i_B} = \beta \approx constant, \qquad v_{CE} \geq v_\gamma$

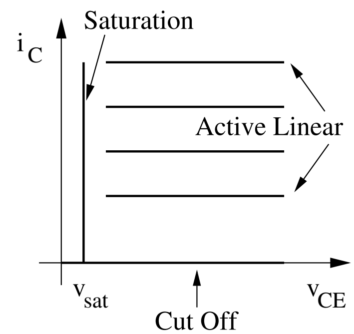
If $v_{CE} < v_\gamma$, both BE and BC junctions are forward biased. This region is called the **saturation** region. As $v_{CE}$ is small while $i_C$ can be substantial, the effective resistance between collector and emitter in saturation region is small and the BJT acts as a closed-circuit.

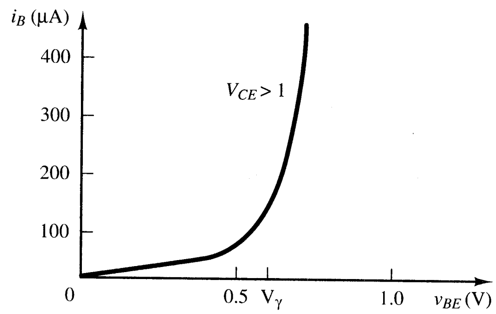
**Saturation:** $v_{BE} = v_\gamma, \qquad i_B > 0, \qquad \dfrac{i_C}{i_B} < \beta, \qquad v_{CE} \approx v_{sat}$

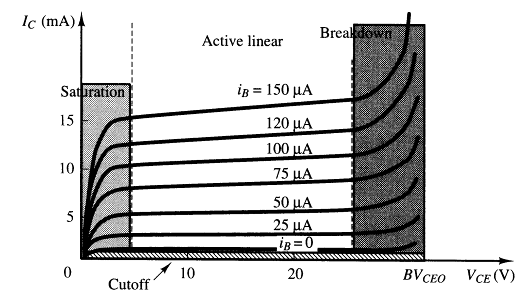
Our model specifies $v_{CE} \approx v_{sat}$, the saturation voltage. In reality in the saturation region $0 < v_{CE} < v_\gamma$. As we are mainly interested in the value of the collector current in this region, $v_{CE}$ is set to a value in the middle of its range in our simple model: $v_{CE} \approx v_{sat} \sim 0.5 v_\gamma$. Typically a value of $v_{sat} \approx 0.2 - 0.3$ V is used for Si semiconductors.

The above simple, large-signal model is shown below. A comparison of this simple model with the real BJT characteristics demonstrates the degree of approximation used.




**How to Solve BJT Circuits:**

The state of a BJT is not known before we solve the circuit, so we do not know which model to use: cut-off, active-linear, or saturation. To solve BJT circuits, we need assume that BJT is in a particular state, use BJT model for that state to solve the circuit and check the validity of our assumptions by checking the inequalities in the model for that state. A formal procedure will be:

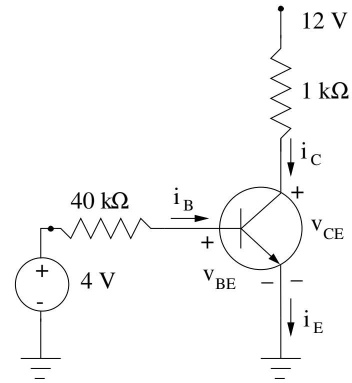
1) Write down a KVL including the BE junction (call it BE-KVL).

2) Write down a KVL including CE terminals (call it CE-KVL).

3) Assume BJT is in cut-off (this is the simplest case). Set $i_B = 0$. Calculate $v_{BE}$ from BE-KVL.
3a) If $v_{BE} < v_\gamma$, then BJT is in cut-off, $i_B = 0$ and $v_{BE}$ is what you just calculated. Set $i_C = i_E = 0$, and calculate $v_{CE}$ from CE-KVL. You are done.
3b) If $v_{BE} > v_\gamma$, then BJT is not in cut-off. Set $v_{BE} = v_\gamma$. Solve above KVL to find $i_B$. You should get $i_B > 0$.

4) Assume that BJT is in active linear region. Let $i_E \approx i_C = \beta i_B$. Calculate $v_{CE}$ from CE-KVL.
4a) If $v_{CE} > v_\gamma$, then BJT is in active-linear region. You are done.
4b) If $v_{CE} < v_\gamma$, then BJT is not in active-linear region. It is in saturation. Let $v_{CE} = v_{sat}$ and compute $i_C$ from CE-KVL. You should find that $i_C < \beta i_B$. You are done.

**Example 1:** Compute the parameters of this circuit ($\beta = 100$).




Following the procedure above:

$$\text{BE-KVL:} \qquad 4 = 40 \times 10^3 i_B + v_{BE}$$
$$\text{CE-KVL:} \qquad 12 = 10^3 i_C + v_{CE},$$

Assume BJT is in cut-off. Set $i_B = 0$ in BE-KVL:

$$\text{BE-KVL:} \qquad 4 = 40 \times 10^3 i_B + v_{BE} \quad \rightarrow \quad v_{BE} = 4 > v_\gamma = 0.7 \text{ V}$$

So BJT is not in cut off and BJT is ON. Set $v_{BE} = 0.7\ V$ and use BE-KVL to find $i_B$.

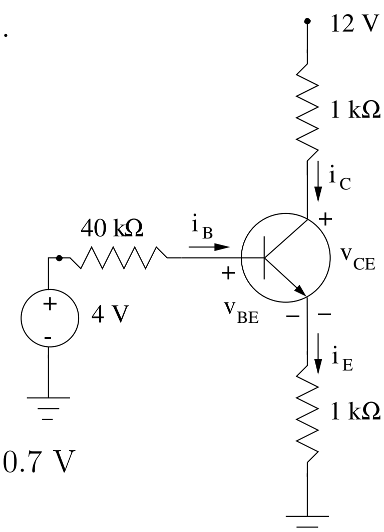
$$\text{BE-KVL:} \qquad 4 = 40 \times 10^3 i_B + v_{BE} \quad \rightarrow \quad i_B = \frac{4 - 0.7}{40,000} = 82.5\ \mu\text{A}$$

Assume BJT is in active linear, Find $i_C = \beta i_B$ and use CE-KVL to find $v_{CE}$:

$$i_C = \beta i_B = 100 i_B = 8.25 \text{ mA}$$
$$\text{CE-KVL:} \qquad 12 = 1,000 i_C + v_{CE}, \quad \rightarrow \quad v_{CE} = 12 - 8.25 = 3.75 \text{ V}$$

As $v_{CE} = 3.75 > v_\gamma$, the BJT is indeed in active-linear and we have: $v_{BE} = 0.7$ V, $i_B = 82.5\ \mu$A, $i_E \approx i_C = 8.25$ mA, and $v_{CE} = 3.75$ V.

**Example 2:** Compute the parameters of this circuit ($\beta = 100$).




Following the procedure above:

$$\text{BE-KVL:} \qquad 4 = 40 \times 10^3 i_B + v_{BE} + 10^3 i_E$$
$$\text{CE-KVL:} \qquad 12 = 1,000 i_C + v_{CE} + 1,000 i_E$$

Assume BJT is in cut-off.
Set $i_B = 0$ and $i_E = i_C = 0$ in BE-KVL:

$$\text{BE-KVL:} \qquad 4 = 40 \times 10^3 i_B + v_{BE} + 10^3 i_E \quad \rightarrow \quad v_{BE} = 4 > 0.7 \text{ V}$$

So BJT is not in cut off and $v_{BE} = 0.7$ V and $i_B > 0$. Here, we cannot find $i_B$ right away from BE-KVL as it also contains $i_E$.

Assume BJT is in active linear, $i_E \approx i_C = \beta i_B$:

$$\text{BE-KVL:} \quad 4 = 40 \times 10^3 i_B + v_{BE} + 10^3 \beta i_B$$
$$4 - 0.7 = (40 \times 10^3 + 10^3 \times 10^2) i_B$$
$$i_B = 24\ \mu\text{A} \quad \rightarrow \quad i_E \approx i_C = \beta i_B = 2.4\ \text{mA}$$
$$\text{CE-KVL:} \quad 12 = 1,000 i_C + v_{CE} + 1,000 i_E, \quad \rightarrow \quad v_{CE} = 12 - 4.8 = 7.2\ \text{V}$$

As $v_{CE} = 7.2 > v_\gamma$, the BJT is indeed in active-linear and we have: $v_{BE} = 0.7$ V, $i_B = 24\ \mu$A, $i_E \approx i_C = 2.4$ mA, and $v_{CE} = 7.2$ V.

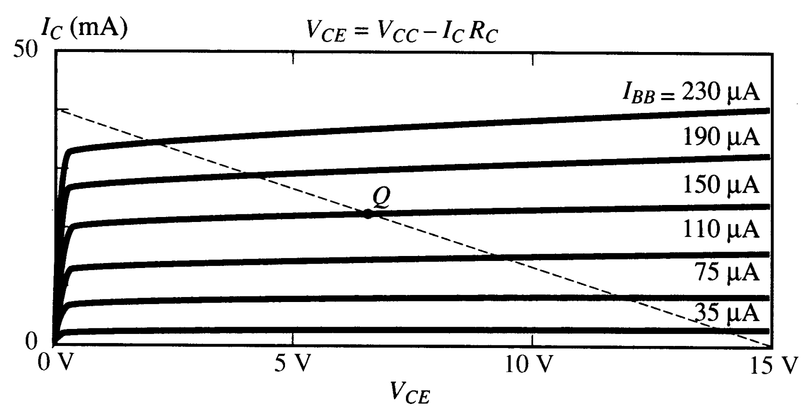
**Load line**

The operating point of a BJT can be found graphically using the concept of a load line. For BJTs, the load line is the relationship between $i_C$ and $v_{CE}$ that is imposed on BJT by the external circuit. For a given value of $i_B$, the $i_C v_{CE}$ characteristics curve of a BJT is the relationship between $i_C$ and $V_{CE}$ as is set by BJT internals. The intersection of the load line with the BJT characteristics represent a pair of $i_C$ and $v_{CE}$ values which satisfy both conditions and, therefore, is the operating point of the BJT (often called the Q point for Quiescent point)

The equation of a load line for a BJT should include only $i_C$ and $v_{CE}$ (no other unknowns). This equation is usually found by writing a KVL around a loop containing $v_{CE}$. For the example above, we have (using $i_E \approx i_C$):

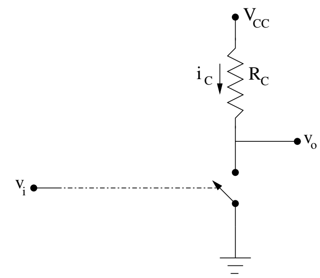
$$\text{KVL:} \quad 12 = 1,000 i_C + v_{CE} + 1,000 i_E \quad \rightarrow \quad 2,000 i_C + v_{CE} = 12$$

An example of a load line, $i_C v_{CE}$ characteristics of a BJT, and the Q-point is shown below.

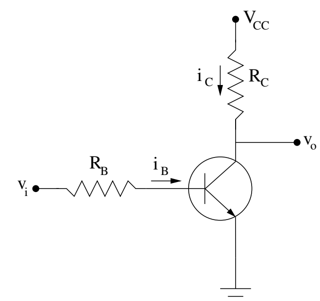


# BJT Switches and Logic Gates

The basic element of logic circuits is the transistor switch. A schematic of such a switch is shown. When the switch is open, $i_C = 0$ and $v_o = V_{CC}$. When the switch is closed, $v_o = 0$ and $i_C = V_{CC}/R_C$.




In an electronic circuit, mechanical switches are not used. The switching action is performed by a transistor with an input voltage switching the circuit, as is shown. When $v_i = 0$, BJT will be in cut-off, $i_C = 0$, and $v_o = V_{CC}$ (open switch). When $v_i$ is in "high" state, BJT can be in saturation with $v_o = v_{CE} = V_{sat} \approx 0.2$ V and $i_C = (V_{CC} - V_{sat})/R_C$ (closed switch). When $R_c$ is replaced with a load, this circuit can switch a load ON or OFF.




The above BJT circuit is also an "inverter" or a "NOT" logic gate. Let's assume that the "low" state is at 0.2 V and the "high" state is at 5 V and $V_{CC} = 5$ V. When the input voltage is "low" ($v_i = 0.2 < v_\gamma$), BJT will be in cut-off and $v_o = V_{CC} = 5$ V ("high" state). When input voltage is "high," with proper choice of $R_B$, BJT will be in saturation, and $v_o = v_{CE} = V_{sat} \approx 0.2$ V ("low" state).

### Resistor-Transistor Logic (RTL)

The inverter circuit discussed above is a member of RTL family of logic gates. Plot of $v_o$ as a function of $v_i$ is called the transfer characteristics of the gate. To find the transfer characteristics, we need to find $v_o$ for a range of $v_i$ values. THis plot will also help identify the values of $V_{IL}$ and $V_{IH}$.

When $v_i < v_\gamma$, BJT will be in cut-off, $i_C = 0$ and $v_o = V_{CC}$. Therefore, for input voltages below certain threshold (denoted by $V_{IL}$), the gate output is high. For our circuit, $V_{IL} = v_\gamma$.

When $v_i$ exceeds $v_\gamma$, BE junction will be forward biased and a current $i_B$ flows into BJT:

$$i_B = \frac{v_i - v_\gamma}{R_B}$$

As BE junction is forward biased, BJT can be either in saturation or active-linear. Let's assume BJT is is in saturation. In that case, $v_o = v_{CE} = V_{sat}$ and $i_C/i_B < \beta$. Then:

$$i_C = \frac{V_{CC} - V_{sat}}{R_C} \qquad \rightarrow \qquad i_B > \frac{i_C}{\beta} = \frac{V_{CC} - V_{sat}}{\beta R_C}$$

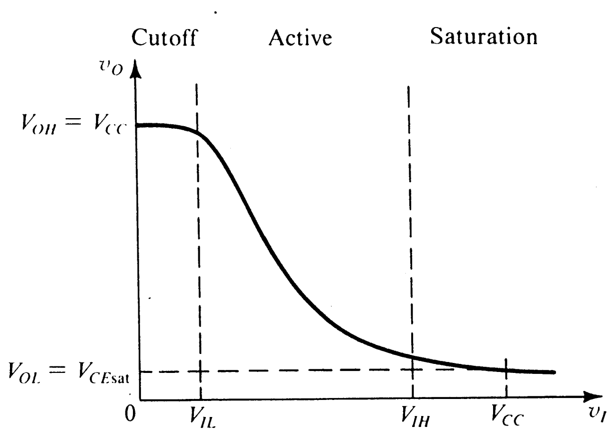
Therefore, BJT will be in saturation only if $i_B$ exceeds the value given by the formula above. This ouccrs when $v_i$ become large enough:

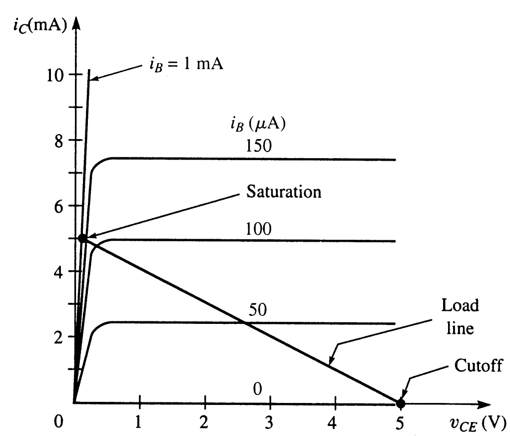
$$v_i = v_\gamma + R_B i_B > v_\gamma + R_B \times \frac{V_{CC} - V_{sat}}{\beta R_C} = V_{IH}$$

Therefore, for input voltages larger than the a certain value ($V_{IH}$) , the gate output is low.

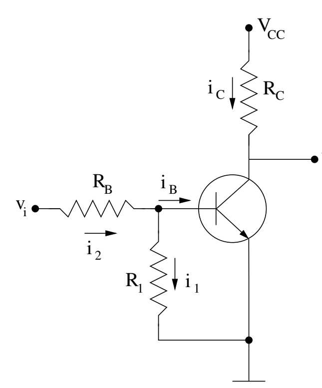
For $v_i$ values between these two limits, the BE junction is forward biased but the BJT is NOT in saturation, therefore, it is in active linear. In this case, the output voltage smoothly changes for its high value to its low value as is shown in the plot of transfer characteristics. This range of $v_i$ is a "forbidden" region and the gate would not work properly in this region.

This behavior can also seen in the plot of the BJT load line. For small values of $v_i$ ($i_B = 0$) BJT is in cut-off. As $v_i$ is increased, $i_B$ is increased and the operating point moves to the left and up on the load line and enters the active-linear region. When $i_B$ is raised above certain limit, the operating point enters the saturation region.




A major drawback of the this RTL inverter gate is the limited input range for the "low" signal ($V_{IL}$). Our analysis indicated that $V_{IL} = v_\gamma$, that is the gate input is low for voltages between 0 and $v_\gamma \approx 0.7$ V. For this analysis, we have been using a piecewise linear model for the BE junction diode. In reality, the BJT will come out of cut-off (BE junction will conduct) at smaller voltages (0.4–0.5 V). To resolve this shortcoming, one can add a resistor between the base and ground (or between base and a negative power supply) as is shown.

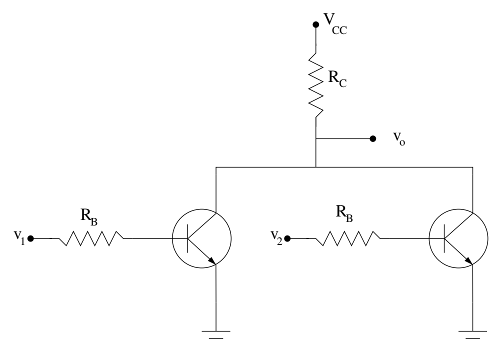



To see the impact of this resistor, note that $V_{IL}$ is the input voltage when BJT is just leaving the cut-off region. At this point, $v_{BE} = v_\gamma$, and $i_B$ is positive but very small (effectively zero). Noting that a voltage $v_{BE}$ has appeared across $R_1$, we have:

$$i_1 = \frac{v_{BE}}{R_1} \qquad\qquad i_2 = i_B + i_1 \approx i_1 = \frac{v_{BE}}{R_1}$$
$$V_{IL} = v_i = R_B i_2 + v_{BE} = v_{BE}\frac{R_B}{R_1} + v_{BE} = v_\gamma\left(1 + \frac{R_B}{R_1}\right)$$

This value should be compared with $V_{IL} = v_\gamma$ in the absence of resistor $R_1$. It can be seen that for $R_B = R_1$, $V_{IL}$ is raised from 0.7 to 1.4 V and for $R_B = 2R_1$, $V_{IL}$ is raised to 2.1 V. $R_1$ does not affect $V_{IH}$ as $i_B$ needed to put the BJT in saturation is typically several times larger than $i_1$.

### RTL NOR Gate




By combining two or more RTL inverters, one obtains the basic logic gate circuit of RTL family, a "NOR" gate, as is shown. More BJTs can be added for additional input signals. (You have seen in 20B that all higher level logic gates, *e.g.,* flip-flops, can be made by a combination of NOR gates or NAND gates.)

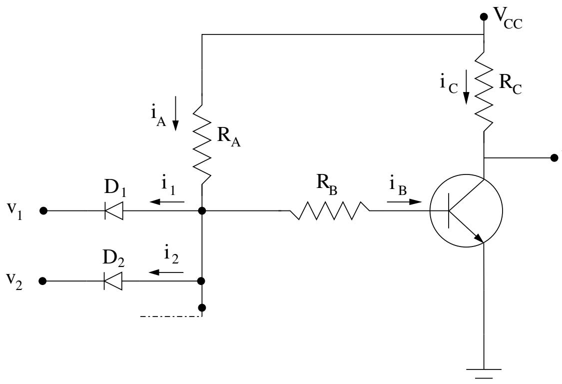
**Exercise:** Show that this ia NOR gate, *i.e.,* the gate output will be low as long as at least one of the inputs is high.

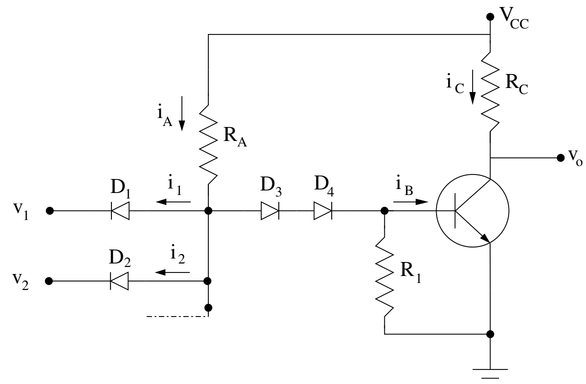
RTLs were the first digital logic circuits using transistors. They were replaced with other forms (DDT, TTL, and ECL) with the advent of integrated circuits. The major problem with these circuits are the use of large resistors that would take large space on an IC chip (in today's chip, resistor values are limited to about 20 kΩ and capacitance to about 100 pF).

Before we move on to more modern gates, we consider two important characteristics of a digital gate.

## Diode-Transistor Logic (DTL)

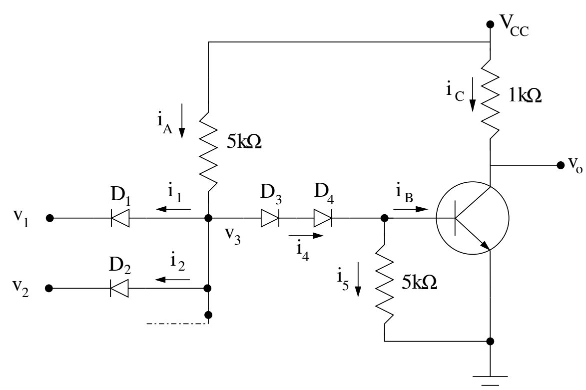
The basic gate of DTL logic circuits is a NAND gate which is constructed by a combination of a diode AND gate (analyzed in pages 67 and 68) and a BJT inverter gate as is shown below (left figure). Because $R_B$ is large, on ICs, this resistor is usually replaced with two diodes. The combination of the two diodes and the BE junction diode leads to a voltage of 2.1 V for the inverter to switch and a $V_{IL} = 1.4$ V for the NAND gate (Why?). Resistor $R_1$ is necessary because without this resistor, current $i_B$ will be too small and the voltage across $D_3$ and $D_4$ will not reach 0.7 V although they are both forward biased.




DTLs were very popular in ICs in 60s and early 70s but are replaced with Transistor-Transistor Logic (TTL) circuits. TTL are described later, but as TTLs are evolved from DTLs, some examples of DTL circuits are given below.

**Example:** Verify that the DTL circuit shown is a NAND gate. Assume that "low"state is 0.2 V, "high" state is 5 V, and BJT $\beta_{min} = 40$.




Case 1: $v_1 = v_2 = 0.2$ V It appears that the 5-V supply will forward bias $D_1$ and $D_2$. Assume $D_1$ and $D_2$ are forward biased: $v_{D1} = v_{D2} = v_\gamma = 0.7$ V and $i_1 > 0$, $i_2 > 0$. In this case:

$$v_3 = v_1 + v_{D1} = v_2 + v_{D2} = 0.2 + 0.7 = 0.9 \text{ V}$$

Voltage $v_3 = 0.9$ V is not sufficient to froward bias $D_3$ and $D_4$ as $v_3 = v_{D3} + v_{D4} + v_{BE}$ and we need at least 1.4 V to forward bias the two diodes. So both $D_3$ and $D_4$ are OFF and $i_4 = 0$. (Note that $D_3$ and $D_4$ can be forward biased without BE junction being forward biased as long as the current $i_4$ is small enough such that voltage drop across the 5 kΩ resistor parallel to BE junction is smaller than 0.7 V. In this case, $i_5 = i_4$ and $i_B = 0$.) Then:

$$i_1 + i_2 = i_A = \frac{5 - v_3}{5,000} = \frac{5 - 0.9}{5,000} = 0.82 \text{ mA}$$

And by symmetry, $i_1 = i_2 = 0.5 i_A = 0.41$ mA. Since both $i_1$ and $i_2$ are positive, our assumption of $D_1$ and $D_2$ being ON are justified. Since $i_4 = 0$, $i_B = 0$ and BJT will be in cut-off with $i_C = 0$ and $v_o = 5$ V.

So, in this case, $D_1$ and $D_2$ are ON, $D_3$ and $D_4$ are OFF, BJT is in cut-off, and $v_o = 5$ V.

Case 2: $v_1 = 0.2$ V, $v_2 = 5$ V    Following arguments of case 1, assume $D_1$ is ON. Again, $v_3 = 0.7 + 0.2 = 0.9$ V, and $D_3$ and $D_4$ will be OFF with $i_4 = 0$. We find that voltage across $D_2$ is $v_{D2} = v_3 - v_2 = 0.9 - 5 = -4.1$ V and, thus, $D_2$ will be OFF and $i_2 = 0$. Then:

$$i_1 = i_A = \frac{5 - v_3}{5,000} = \frac{5 - 0.9}{5,000} = 0.82 \text{ mA}$$

and since $i_1 > 0$, our assumption of $D_1$ ON is justified. Since $i_4 = 0$, $i_B = 0$ and BJT will be in cut-off with $i_C = 0$ and $v_o = 5$ V.

So, in this case, $D_1$ is ON, $D_2$ is OFF, $D_3$ and $D_4$ are OFF, BJT is in cut-off, and $v_o = 5$ V.

Case 3: $v_1 = 5$ V, $v_2 = 0.2$ V    Because of the symmetry in the circuit, this is exactly the same as case 2 with roles of $D_1$ and $D_2$ reversed.

So, in this case, $D_1$ is OFF, $D_2$ is ON, $D_3$ and $D_4$ are OFF, BJT is in cut-off, and $v_o = 5$ V.

Case 4: $v_1 = v_2 = 5$ V    Examining the circuit, it appears that the 5-V supply will NOT be able to forward bias $D_1$ and $D_2$. Assume $D_1$ and $D_2$ are OFF: $i_1 = i_2 = 0$, $v_{D1} < v_\gamma$ and $v_{D2} < v_\gamma$. On the other hand, it appears that $D_3$ and $D_4$ will be forward biased. Assume $D_3$ and $D_4$ are forward biased: $v_{D3} = v_{D4} = v_\gamma = 0.7$ V and $i_4 > 0$. Further, assume the BJT is not in cut-off $v_{BE} = v_\gamma = 0.7$ V and $i_B > 0$. In this case:

$$v_3 = v_{D3} + v_{D4} + v_{BE} = 0.7 + 0.7 + 0.7 = 2.1 \text{ V}$$

$$v_{D1} = v_3 - v_1 = 2.1 - 5 = -2.9 \text{ V} < v_\gamma \qquad v_{D2} = v_3 - v_2 = 2.1 - 5 = -2.9 \text{ V} < v_\gamma$$

Thus, our assumption of $D_1$ and $D_2$ being OFF are justified. Furthermore:

$$i_4 = i_A = \frac{5 - v_3}{5,000} = \frac{5 - 2.1}{5,000} = 0.58 \text{ mA}$$

$$i_5 = \frac{v_{BE}}{5,000} = \frac{0.7}{5,000} = 0.14 \text{ mA}$$

$$i_B = i_4 - i_5 = 0.58 - 0.14 = 0.44 \text{ mA}$$

and since $i_4 > 0$ our assumption of $D_3$ and $D_4$ being ON are justified and since $i_B > 0$ our assumption of BJT not in cut-off is justified.

We still do not know if BJT is in active-linear or saturation. Assume BJT is in saturation: $v_o = v_{CE} = V_{sat} = 0.2$ V and $i_C/i_B < \beta$. Then, assuming no gate is attached to the circuit, we have

$$i_C = \frac{5 - V_{sat}}{1,000} = \frac{5 - 0.2}{1,000} = 4.8 \text{ mA}$$

and since $i_C/i_B = 4.8/0.44 = 11 < \beta = 40$, our assumption of BJT in saturation is justified.

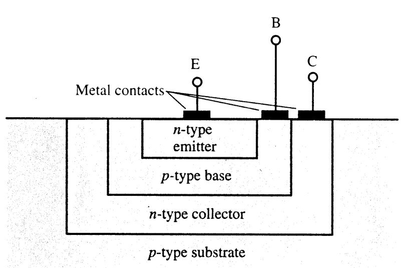
So, in this case, $D_1$ and $D_2$ are OFF, $D_3$ and $D_4$ are ON, BJT is in saturation and $v_o = 0.2$ V.

Overall, the output in "low" only if both inputs are "high", thus, this is a NAND gate.

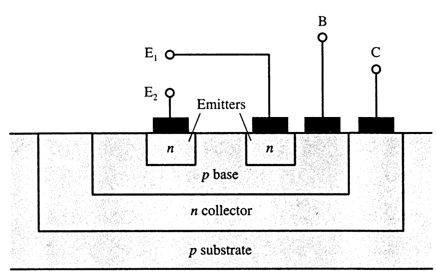
**Note:** It is interesting to note that at the input of this gate, the current actually flows out of the gate. In the example above, when both inputs were high $i_1 = i_2 = 0$, when both were low $i_1 = i_2 = 0.4$ mA, and when one input was low, *e.g.,* $v_1$ was low, $i_1 = 0.8$mA. The input current flowing in (or out of the gate in this case) has implications for the fan-out capability of logic gates as is shown in the example below.

**Transistor-Transistor Logic (TTL)**

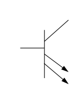
A simplified version of an IC-chip NPN transistor is shown. The device is fabricated on a p-type substrate (or body) in a vertical manner by embedding alternating layers of N and P-type semiconductors. By embedding more than one N-type emitter region, one can obtain a multiple-emitter NPN transistor as shown. The multiple-emitter NPN transistors can be used to replace the input diodes of a DTL NAND gate and arrive at a NAND gate entirely made of transistors, hence Transistor-Transistor Logic (TTL) gates.




(a) IC *npn* transistor



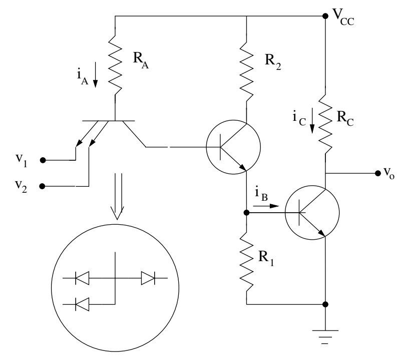

(a) two-emitter IC *npn* transistor



Circuit Symbol

A simple TTL gate is shown with the multiple-emitter BJT replacing the input diodes. This transistor operates in "reverse-active" mode, *i.e.,* like a NPN transistor in active-linear mode but with collector and emitter switched. Operationally, this BJT acts as two diodes back to back as shown in the circle at the bottom of the figure. As such the operation of this gate is essentially similar to the DTL NAND gate described above (note position of driver transistor and $D_4$ diode is switched).

Similar to DTL NAND gates, a typical TTL NAND gate has three stages: 1) Input stage (multi-emitter transistor), 2) driver stage, and 3) output stage. Modern TTL gates basically have the same configuration as is shown with the exception that the output stage is replaced with the "Totem-Pole" output stage to increase switching speed and gate fan-out. For a detailed description of TTL gate with "Totem-Pole" output stage, consult, Sedra and Smith (pages 1175 to 1180).

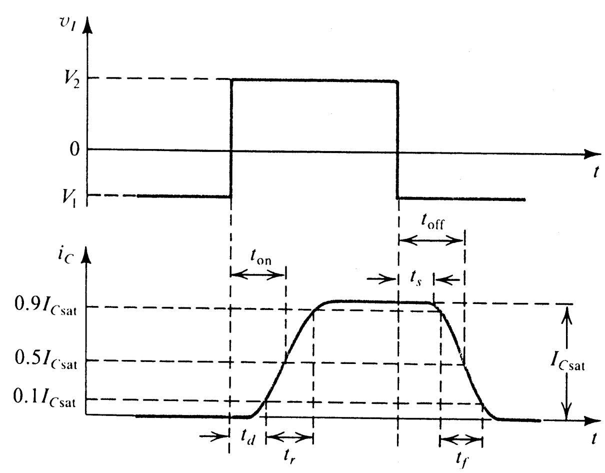



**Switching Time and Propagation Delay:**

Here we introduce two important considerations in designing logic gates.

Consider the inverter gate with an input voltage close to zero (and/or negative). In this case, the BJT is in cut-off, $i_C = 0$ and the output of the gate is high. Suppose a "high" voltage is applied instantaneously to the gate at some point. We expect BJT to enter saturation with $i_C = I_{Csat}$ and output to drop to the "low"state. However, this does not occur instantaneously.

When the BJT is in cut-off, BE junction is reversed biased. When a forward voltage is applied to the BE junction, it takes some time for the BE junction transition capacitance to charge up. Time is also required for minority carries to diffuse across the base and enter the collector. This results in the delay time $t_d$, which is of the order of a nanosecond for a typical BJT.



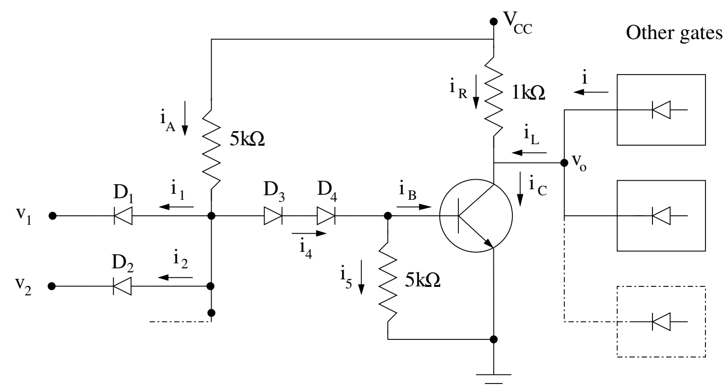

Before BJT can enter saturation, it should traverse the active-linear region. The rise time, $t_r$ (on the order of 1-10 ns) account for this transition. The time that takes for the gate to switch "ON" is represented by $t_{on}$.

Suppose that the input voltage to gate is then reduced instantaneously to low state. BJT will leave saturation region and go to cut-off. Again, this not occur instantaneously. When a BJT is in saturation, both BE and BC junctions are forward biased and conducting. As such, an excess minority charge is stored in the base. For the transistor to leave saturation and enter active-linear (BC junction to become reversed biased), this excess charge must be removed. The time required for the removal of excess charge determines the storage time, $t_s$ (order of 100 ns). Then, transistor traverses the active-linear region before entering cut-off. This account for the fall time $t_f$ (1-10 ns). The total time it takes for the gate to switch "OFF' is represented by $t_{off}$. As can be seen, BJT switching is mainly set by the storage time, $t_s$.

Propagation delays introduced by transistor switching time are important constraints in designing faster chips. Gate designs try to minimize propagation delays as much as possible.

**Fan-out:** All digital logic circuits are constructed with cross-coupling of several basic gates (such as NOR or NAND). As such, a basic gate may be attached to several other gates. The maximum number of gates that can be attached to a digital gate is called "fan-out." Obviously, one would like to have large fan-out.

**Example:** Find the fan-out of this NAND DTL gate. Assume that "low"state is 0.2 V, "high" state is 5 V, and BJT $\beta = 40$.




The circuit is the same DTL NAND gate of previous example and we can use results from previous example here. "$N$" other NAND gates are attached to the output of this gate. Fan-out is the maximum value of $N$. Since we want to make sure that our gate operates properly under all conditions, we should consider the worst case, when all of the second stage gates have maximum currents.

For a NAND DTL gate, the maximum current $i$ occurs when all of the inputs are high with exception of one input. We found this value to be 0.82 mA (Cases 2 & 3 in the previous example). Therefore, the worst case is when the input of all second stage gates are low (for the first stage, $v_o = 0.2$ V) and each draw a current 0.82 mA (a total of $i_L = N \times 0.82$ mA is drawn from the first stage gate).

Considering the first stage gate, we had found that $v_o = 0.2$ V only for Case 4. For that case, we found $i_B = 0.44$ mA. Then:

$$i_R = \frac{5 - V_{sat}}{1,000} = \frac{5 - 0.2}{1,000} = 4.8 \text{ mA}$$
$$i_C = i_R + 0.82N = 4.8 + 0.82N$$

The first stage gate operates properly as long as the BJT is in saturation, *i.e.,*

$$i_C < \beta i_B \quad \rightarrow \quad 4.8 + 0.82N < 40 \times 0.44 \quad \rightarrow \quad N < 13.7$$

As the fan-out should be integer, the fan-out for this gate is 13.

Fan-out of DTL gates can be greatly increased by a small modification. Fan-out can be increased by increasing the base current of the BJT. $i_B$ is, however, limited by the current $i_A$ (and $i_4$). Reducing the value of $R_A$ in the AND diode part of the circuit would increase $i_B$ but it also increases the input current demand and, in fact, the reduces the fan-out capability.

**Exercise:** Find the value of $R_A$ which leads to the maximum fan-out. Assume that "low"state is 0.2 V, "high" state is 5 V, and BJT $\beta = 40$. Also, $R_C = 1$ k$\Omega$ and $R_A = R_1$ (See circuit on page 84 for notation).

A simple solution which keeps current $i_A$ small but increases $i_B$ drastically is to replace diode $D_3$ with a BJT as is shown. As can be seen, the DTL NAND gate is now made of 3 stages: 1) input stage (diodes), 2) driver stage (first BJT) and 3) output stage (2nd BJT). In this case, value of $R_A$ can be raised substantially.