إلى قارئ هذا الكتاب ، تحية طيبة وبعد ...

لقد أصبحنا نعيش في عالم يعج بالأبحاث والكتب والمعلومات، وأصبح العلم معياراً حقيقياً لتفاضل الأمم والدول والمؤسسات والأشخاص على حدٍّ سواء، وقد أمسى بدوره حلاً شبه وحيدٍ لأكثر مشاكل العالم حدة وخطورة، فالبيئة تبحث عن حلول، وصحة الإنسان تبحث عن حلول، والموارد التي تشكل حاجة أساسية للإنسان تبحث عن حلول كذلك، والطاقة والغذاء والماء جميعها تحديات يقف العلم في وجهها الآن ويحاول أن يجد الحلول لها. فأين نحن من هذا العلم ؟ وأين هو منا؟

نسعى في موقع عالم الإلكترون www.4electron.com لأن نوفر بين أيدي كل من حمل على عاتقه مسيرة درب تملؤه التحديات ما نستطيع من أدوات تساعده في هذا الدرب، من مواضيع علمية، ومراجع أجنبية بأحدث إصداراتها، وساحات لتبادل الآراء والأفكار العلمية والمرتبطة بحياتنا الهندسية، وشروحٍ لأهم برمجيات الحاسب التي تتداخل مع تطبيقات الحياة الأكاديمية والعملية، ولكننا نتوقع في نفس الوقت أن نجد بين الطلاب والمهندسين والباحثين من يسعى مثلنا لتحقيق النفع والفائدة للجميع، ويحلم أن يكون عضواً في مجتمعٍ يساهم بتحقيق بيئة خصبة للمواهب والإبداعات والتألق، فهل تحلم بذلك ؟

حاول أن تساهم بفكرة، بومضة من خواطر تفكيرك العلمي، بفائدة رأيتها في إحدى المواضيع العلمية، بجانب مضيء لمحته خلف ثنايا مفهوم هندسي ما. تأكد بأنك ستلتمس الفائدة في كل خطوة تخطوها، وترى غيرك يخطوها معك ...

أخي القارئ، نرجو أن يكون هذا الكتاب مقدمة لمشاركتك في عالمنا العلمي التعاوني، وسيكون موقعكم عالم الإلكترون ww.4electron.com بكل الإمكانيات المتوفرة لديه جاهزاً على الدوام لأن يحقق البيئة والواقع الذي يبحث عنه كل باحث أو طالب في علوم الهندسة، ويسعى فيه للإفادة كل ساعٍ ، فأهلاً وسهلاً بكم .

مع تحيات إدارة الموقع وفريق عمله

www.4electron.com

INTRODUCTION

Electronics is heavily relied on by most other areas of electrical engineering. While there is a considerable body of theory in communications, controls, etc., these areas ultimately use electronics to actually implement the functions.

Electronic circuits use electronic devices to perform functions on signals such as amplification, filtering, rectifying, switching, etc.. Electronics has been a major topic of study in Electrical Engineering for nearly a century. Early electronic circuits used devices such as spark gaps and point-contact crystal diodes to perform signal processing. Later on, vacuum tubes were invented which made electronic communications and control systems widely available. In the late 1940's, semiconductor devices such as diodes and transistors became available which created an electronics revolution. With these changes in technology, the study of electronics did not change significantly, only the devices changed. The circuits and the methods did not change substantially.

The study of electronics can be roughly divided into two areas, devices and circuits. The study of devices is concerned with physical processes such as electron flow while the study of circuits emphasizes using the devices in applications and signal processing functions. The study of electronic circuits is further subdivided into analog, or linear, and digital, or switching, electronics. This course focuses on digital electronic circuits.

By far, the greatest use of digital electronic circuits occurs in digital computers. Logic circuits are widely available from simple logic gates in small-scale integrated (SSI) circuits to very complex digital functions in very large-scale integrated (VLSI) circuits. In almost all digital circuits, transistors and diodes operate in two modes, on or off, carrying current or not carrying current; in essence, a switch. We will look at how digital logic circuits operate and what the terminal characteristics and manufacturer's specifications mean. We will then look at how to go beyond the logic circuits with interfaces both at the inputs and outputs.

We begin this course with a brief discussion of semiconductor materials and pn junctions. This material is neither rigorously developed nor complete. A rigorous study of semiconductor electronics is left for later. However, to effectively use semiconductor devices, it is necessary to have a basic understanding of how they work.

Because electronic devices are non-linear, we will look at their terminal characteristics and make circuit models of the devices that will allow us to use to linear circuit analysis techniques to analyze the circuits. We will then look at application of semiconductor devices in switching circuits including logic gates, interface circuits, and special applications. Many electronic systems involve both analog and digital circuits and during this course, we will look at some of these cases requiring a mixture of applications. The ultimate test of understanding the material of this course will be found in the design exercises.

Semiconductors

THE ELECTRON IN ELECTRIC FIELDS

If we were to take two parallel plates and connect a voltage source across them as shown in Figure 1, an electric field would be set up between the plates. Neglecting fringing around the edges, the electric field would be uniform everywhere between the plates. The electric field strength would be

$$E = V/d \qquad (1)$$

where V is the applied voltage and d is the distance between the plates. Thus, the electric field strength E has the units volts per centimeter and is a vector quantity going from a positive charge to a negative charge. Note: CGS units are normally used in semiconductor physics - centimeters, grams, seconds.

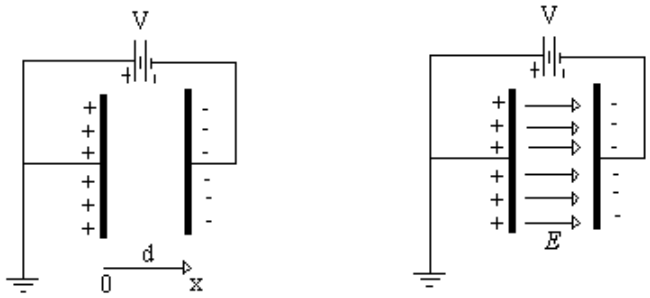


Figure 1. Parallel plate with a voltage source connected between the plates.

Now if a tiny person, let's call her Millie Micron, was able to carry an electron into the region between the plates and release it as shown in Figure 2, the electron would be attracted to the positive plate and repelled by the negative plate. The force on the electron would be

$$F_X = -qE_X \qquad (2)$$

where q is the electronic charge. The negative sign occurs because the electron is accelerated in the negative x direction, toward the positive plate. Of course, the electron would obey Newton's laws and the acceleration, a_X, would be a function of the mass of the electron, m, and the force exerted by the electric field,

$$F_X = -ma_X \qquad (3)$$

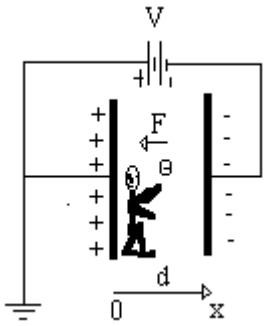


Figure 2. Millie releasing an electron within the electric field

As the electron accelerates, it gains kinetic energy. Just as with objects with mass in a gravitational field, the electron in an electric field has potential energy that can be converted to kinetic energy. The total energy then is

$$W = U + 1/2mv^2 \qquad (4)$$

where U is the potential energy and v is the velocity. The energy associated with a single electron is quite small compared to units we normally work with so we use the units electron volts defined as moving one electron across a potential difference of one volt.

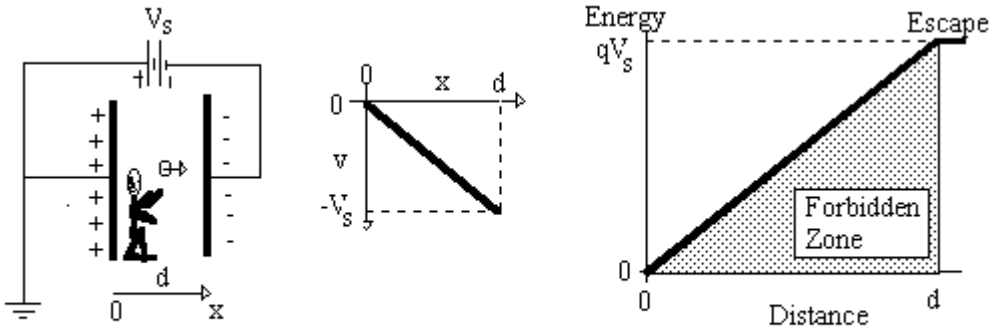
$$1 \text{ eV} = 1.602 \times 10^{-19} \text{ joules} \quad (1 \text{ joule} = 1 \text{ watt second})$$

The electronic charge is 1.602×10^{-19} coulombs (ampere-seconds).

In our example, if the voltage source is 5 volts and Millie released the electron at the negative plate, the electron would gain five electron volts of energy as it fell to the positive plate. At that point it would have zero potential energy. Thus, at the point of release, the electron had a potential energy of 5 eV. This 5 eV would be converted to kinetic energy by the time it arrived at the positive plate.

To look at this another way, let's look at a plot of the electric potential within the field. We will assume the positive plate is grounded and at zero potential. The negative plate is at negative five volts with the potential changing linearly in between as shown in Figure 3. In this example, let's assume Millie is standing on the positive plate and throws the electron toward the negative plate. If she throws it gently, it will start with only a small kinetic energy which is soon converted to potential energy as the electron goes against the electric field. When all the kinetic energy is converted to potential energy, the electron has zero velocity. The electric field accelerates the electron back toward the positive plate. The effect is that the electron falls back to Millie and she catches it. If she then throws it again, but this time a little harder, it will go further, but will again fall back. Say, this is a neat game isn't it?. This is similar to throwing a ball up a sloping roof and

catching it as it rolls back down. Just as with the roof, if Millie throws the electron hard enough, it will overcome the potential hill and escape. In this case, we must envision the negative plate as having lots of holes, like a wire screen, the electron can go through to escape; for example. This potential barrier concept will be used when we look at p-n junctions.



a. Millie throwing the electron.

b. Plot of potential versus distance

c. Potential energy barrier.

Figure 3. Illustration of the potential energy barrier.

ELECTRON EMISSION FROM THE METAL

Now, we can describe current flow between two metal plates. If a voltage is applied across the two metal conductors with just vacuum in between the two conductors, obviously no current flows between the two plates in the vacuum. There will be an electric field ($E = V_s/x$, where x is the distance between the two plates) dropped across two conductors. Charge builds up on each surface of the metal conductors that are facing each other. On one conductor, the cathode, electrons collect on the surface. On the other conductor, the anode, electrons are repelled from the surface, leaving the fixed metal ions at the surface. If a strong enough voltage (or a sufficiently high electric field) is applied, the electrons have a total energy equal to the vacuum energy, E_{vac}. The electrons at the surface of one metal conductor will be ripped out of the conductor and then move to the anode. And, current does flow across the vacuum!

This is called thermionic emission, a process by which tungsten filaments emit beams of electrons in cathode ray tubes for television, oscilloscope screens, and other instruments. The electric field required to rip the electrons out of the metal and into the vacuum is equal to $q\phi_m$, where ϕ_m is the work function of the metal and q is the charge on an electron. Different metals have different work functions. So, some metals work better than others as "electron guns".

When additional energy is added to the system, for example by heating the metal, the kinetic energy of the electrons in the metal is increased and the electric field required for

thermionic emission is decreased. Thus, there are some "electron guns" that are called cold cathode emitters and others called hot filament emitters, denoting the temperature of the metal from which the electrons are escaping. The tungsten filament in your television is a hot filament emitter, running at over 2500K. Cold cathode emitters are used when a high temperature filament is impractical in the system. The cold cathode emitters are being researched for thin film displays and other applications where the system can not handle a large thermal gradients.

CURRENT CONDUCTION IN METALS

Metal atoms have one or more very loosely bound valence electrons. These are the electrons in the outer most orbital or electron shell, in an s or d orbital. At normal temperatures, the valence electrons have enough thermal energy to be easily separated from the metal atom and move randomly throughout the material. The metal atom, then, becomes a positively charged ion. Figure 4 is a two-dimensional representation of the situation where the electrons are free and the metal ions are immobile. The material still has zero net charge as there are still as many electrons in the metal as there are positive charges on the metal ions. In its random motion, an electron occasionally collides with a metal ion, but with its thermal energy, it is not captured and rebounds at a random angle. The motion of electrons in metal described here is similar to the motion of molecules in gas. Thus, it is called the electron gas model.

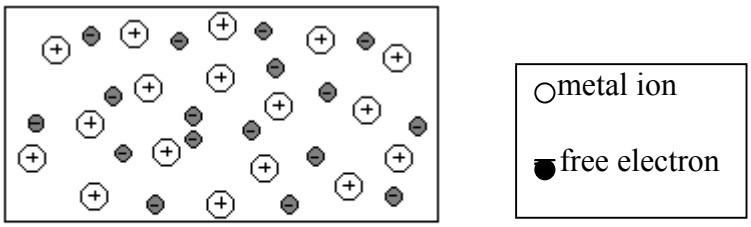


Figure 4. Metal with fixed ions and electrons in random motion.

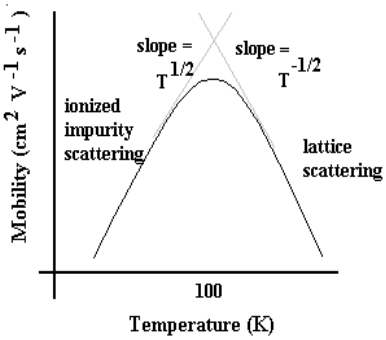
If we were to average the motion or velocity of the electrons in the metal in Figure 4, we would find zero net motion and zero average velocity. However, if we were to apply a voltage between the ends of the metal conductor, a field would be set up between the ends of the conductor and the electrons would be accelerated toward the positive end. Thus, there is a drift of electrons in the conductor toward the positive end and a current results. This current is called drift current. The average speed at which the electrons drift is called the drift velocity, v_d. It seems as though the electrons might continue to accelerate in the field and reach very high velocities, but instead, the electrons soon collide with a fixed ion and recoil in a random direction, to again be accelerated by the field. Therefore, the electron's drift velocity reaches a maximum at some electric field and does not increase any further with increasing electric field. This is called the saturation velocity. The drift velocity under low field conditions, when the drift velocity is well below the

saturation velocity, is a function of the electric field and the physical properties of the conductor,

$$v_d = \mu E \qquad (5)$$

where E is the electric field strength and μ is the mobility of the electron in that material.

Electron Mobility



Mobility is a physical constant that describes the ease in which an electron can move through a material. The mobility is a function of temperature as well as the electric field. Thus, the speed of the electrons in the metal changes with temperature. The exact relationship between the change in mobility as a function of temperature is dependent on a number of material properties including the number of grain boundaries and how pure the metal is.

At 0K, there is no kinetic energy. Therefore, the valence electrons are localized in a metal atom. As the temperature is increased from 0K, there are some free electrons. However, their thermal energy is low. They interact with the metal ions and undergo Coulombic scattering events. In coulombic scattering, they lose some of their energy to the metal ion so their mobility is limited by the number of scattering events.

A macroscopic example of a similar problem would be if you shot an iron marble (the electron) through a grid of magnets (the metal ions), where there is considerable space between each of the magnets. The magnets are regularly spaced and fixed in position. When you shoot the marble, your aim is great – the path is clear from one end of the grid to the other. However, if the marble is rolled slowly (low kinetic energy), it is likely that its motion will be perturbed by the magnetic field of the magnets. At some point in its travel, the magnetic attraction may be strong enough to divert the forward motion of the marble and the marble will collide with the magnet. As you increase the speed at which you send the marble through the grid (increase the kinetic energy/temperature), the farther the marble will go through the grid before it gets attracted to one of the magnets.

As the temperature is increased further, the kinetic energy of the free electrons increases and the scattering events do not result in significant energy transfer. So, the mobility of the electron increases. However, as temperature increases significantly (to ~100K), the kinetic energy of the metal ions becomes large enough that they are vibrating with sufficient movement to impede the movement of the electrons. The scattering events are called lattice scattering. Again, there is energy transfer and the free electrons' mobility is decreased because of the scattering events. Unlike Coulombic scattering, the effect of

lattice scattering increases as the temperature increases, further decreasing the mobility of the free electrons.

Continuing with the macroscopic example: Now modify the grid so that the magnets are mounted on springs. As the grid is moved (because both the electrons and the metal ions have kinetic energy), the magnets "jiggle" back and forth. This reduces the amount of open area between each magnet. As the movement of the grid increases (an increase in temperature), the magnets jiggle further into the open areas and jiggle faster back and forth. This increases the likelihood that the marble will hit one or more of the magnets during its travel, reducing the distance that the marble can move through the grid.

Current Density

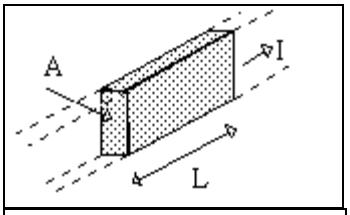


Figure 5. A section of a conductor with current I

Figure 5 shows a section of a conductor of length L and cross section A. Within this volume, there are N electrons. The number of free electrons, N, depends on the metal. If we apply a voltage, V, from end-to-end, the electrons will drift towards the positive end with a drift velocity of v_d. The time required for the electronics to traverse the length, L, at an average drift velocity, v_d is

$$t_t = \frac{L}{v_d} \qquad (6)$$

All N of the electrons pass through the end of the section in time, t_t. This electron drift constitutes a current I in amperes (coulombs/second).

$$I = \frac{qN}{t_t} \qquad (7)$$

Substituting the transit time, t_t, into equation (7)

$$I = \frac{qNv_d}{L} \longrightarrow J = \frac{I}{A} = \frac{qNv_d}{LA} \qquad (8)$$

where J is the current density which is defined as current over the cross sectional area.

The electron density is the number of electrons in a unit volume, n = N/LA. Thus

$$J = qnv_d \qquad (9)$$

Conductivity

We can write current density in terms of the electric field by substituting $v_d = \mu E$,

$$J = qn\mu E \qquad (10)$$

We know that voltage is the electric field, E, times the distance over which the electric field is dropped (V = EL). Thus, resistance, R, can be rewritten as follows:

$$R = \frac{V}{I} = \frac{EL}{qn\mu EA} = \frac{L}{qn\mu A} = \frac{\rho L}{A} \qquad (11)$$

In equation 11, we have related resistance of a sample to its shape and a parameter called resistivity, ρ. We define this parameter, ρ, as well as a second parameter, conductivity (σ), which is the inverse of resistivity, as:

$$\sigma = 1/\rho = qn\mu \qquad (12)$$

This result is an important equation. Conductivity is a function of the electronic charge, the (free) electron density, and the mobility of the electrons in the material. In metals, the free electron density is equal to the number of valence electrons times the density of atoms in the material that readily give up the valence electrons. Thus, mobility can be readily determined once the conductivity is found, which in turn is determined from the resistance measurement of a sample of the material. Also the physical dimensions of the metal are critical in determining its resistance. Resistance is a linearly proportional to the length of the metal and it is inversely proportional to its cross-sectional area.

Example:

Aluminum has a density of 2.7 g/cm^3, atomic weight of 27, and a resistivity of 3.44×10^{-6} Ω–cm. If Aluminum has three valence electrons, what is the mobility?

Avogadro's number is 6.02×10^{23} atom per mole. Thus, one mole of aluminum weighs 27 grams and has 6.02×10^{23} atoms. From this we get that one cm^3 of aluminum contains 0.1 moles or 6.02×10^{22} atoms. If each atom has three free electrons, there are 18.06×10^{22} free electrons per cubic meter. Thus we have

$$\sigma = qn\mu$$

or

$\sigma = (3.44 \times 10^{-6}\ \Omega\ cm)^{-1} = 1.602 \times 10^{-19}$ Coul x 18.06×10^{22} free electrons/ cm^3 x μ

$\mu = 10\ cm^2/Vs$

So, if we have 3 meters of Aluminum wire, the length of a normal oscilloscope cable, and there is 5 volts dropped across the cable, the time that it will take an electron at one end of the cable to make it all the way to the other end of the cable (called transit time, t_t) is 1800 seconds.

$$v_d = \mu E$$

$$t_t = L/(\mu * E) \qquad (13)$$

$$t_t = 300cm/(10cm^2/Vs * (5V/300cm)) = 1800\ s$$

Not too speedy is it? This is one of several reasons why we don't transfer data using individual electrons, except over very short distances.

Resistance as a Function of Temperature

Now, what happens to the resistance of a metal if we increase its temperature from room temperature (roughly 25°C) to, say, 125 °C, a commonly used maximum temperature of operation for certain devices and circuits? Look back at equation 11. Based upon our previous discussions, you know that the number of free electrons in the metal do not increase with increasing temperature. Of course, the length and cross-sectional area do increase slightly as the metal expands. For example, the length of an aluminum bar increases by 12 parts per million per degree C. Thus, a 1 meter long aluminum bar at room temperature will increase to a length of 1.012 meters.

However, the parameter that has most significant change with temperature is the mobility of the free electrons, μ. And, the mobility decreases as the temperature increases to 125°C. Thus, the resistance of the metal increases with increasing temperature. If you look at the specifications for metal- and carbon-based resistors, you will see that their resistance is not a constant as a function of temperature, but increases. This increase in resistance with temperature can be minimized over specific temperature ranges through a careful design of the materials used and the physical structure of the resistor. So, even the selection of resistors needs to considered when you design of circuit – the selection will depend on your application.

INTRINSIC SEMICONDUCTORS

Atoms in metals have relatively free valence electrons, which provide the carriers for conduction. Insulators are materials where there are virtually no free electrons, hence, no

conduction. Semiconductors are in between. The most common semiconductor material is silicon, with germanium, gallium arsenide and indium phosphide used for some special purposes. As you can see from the periodic table, both silicon and germanium are group IV materials and have four valence electrons, two in the outer most s shell and two electrons in the p orbital. Gallium arsenide is a III-V compound where gallium has three valence electrons (two in the s orbital and one in the p orbital) and arsenic has five (two in the s orbital and three in the p orbital). A maximum of two electrons occupy the s orbital and a maximum of six occupy the p orbital. The number of valence electrons is important because these materials form crystalline structures and each atom shares electrons with its neighbors in covalent bonds to, in effect, fill each atom's outer shell.

Because silicon is by far the most widely used semiconductor material, we will use it in our examples. A two-dimensional representation of the of the silicon structure is shown in Figure 6. Here the silicon atoms are arranged in regular rows and columns as in a crystal. Each atom shares four electrons with its four nearest neighbors. In the actual three-dimensional crystal, the structure is a tetrahedron and is called a diamond lattice. There are eight silicon atoms in each unit cell of the diamond lattice. A unit cell is the smallest volume that you can make that still has the same geometry of the overall crystal. In Si, the unit cell is a cube with dimensions of a = 0.545nm per side, where a is called a lattice constant.

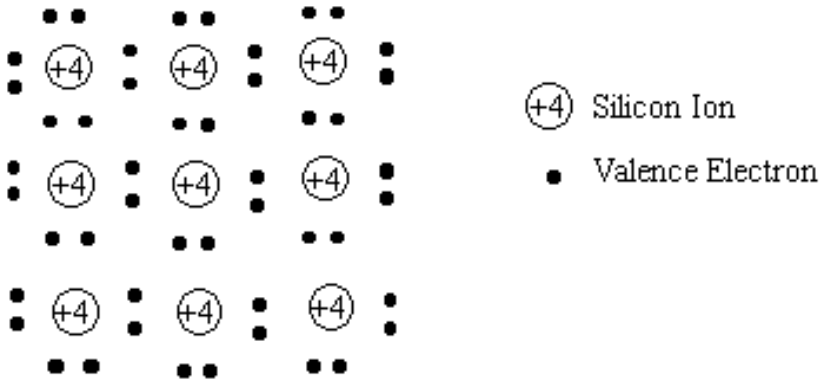


Figure 6. 2-D representation of a silicon crystal.

As shown in the drawing, each electron is part of a covalent bond and is rather tightly bound. However, due to thermal agitation, an occasional electron has enough energy to break free from the bond and become free to roam throughout the crystal as shown in Figure 7. The only place with a positive charge that can capture the free electron is another broken bond. Thus, these free electrons can contribute to electrical conduction.

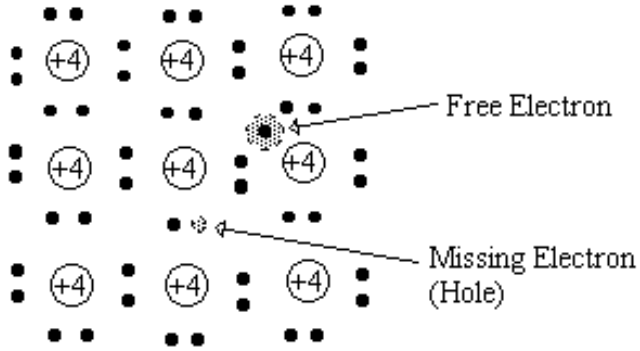
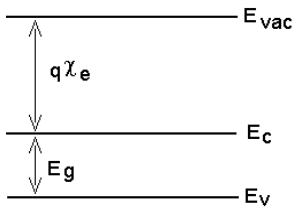


Figure 7. Silicon crystal with free electron and hole.



The free electrons are called conduction electrons. The conduction electrons reside in the conduction band, at an energy of E_C and above. The remaining electrons that are bound to the atoms, the valance electrons, reside in the valence band, at energies of E_V and below. The amount of energy required to free the electron is called the bandgap energy, E_g. The bandgap energy arises from the interaction of the atoms within the solid, the covalent bonds that results in the sharing of electrons between neighboring atoms. The energy required for thermionic emission of electrons from the semiconductor crystal, the emission of an electron from a semiconductor material into vacuum, is equal to $q\chi_e$, the charge on an electron times the electron affinity, χ_e. The graph to the right is called an energy diagram and shows these different energy levels.

The difference between semiconductors and non-conductors, or insulators, is the size of the energy gap between the valence bands, and the conductions bands. If the energy gap is very small, then the material will conduct electric current easily. If the band gap is large, then it will be a non-conductor. Semiconductors typically have a band gap of about 1 electon volt, while insulators have a band gap of several electron volts. This definition is somewhat vague and should only be used as a general guideline.

HOLE, THE MISSING ELECTRON

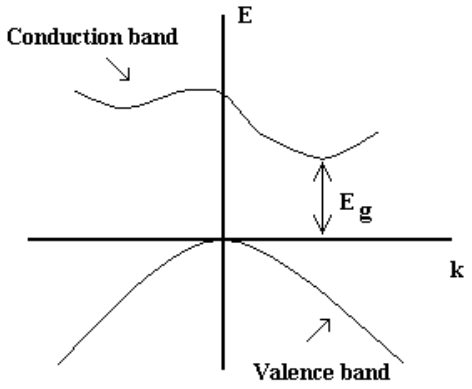
One interesting feature of the semiconductor is that the broken bond, called a hole, can also contribute to conduction. These holes reside at the top of the valence band, with an energy of E_V and are positively charged. Because the valence electrons and atoms have kinetic energy and the atoms are closely packed in a crystal, a valence electron can break its bond with a neighboring atom and complete the broken bond. Since there is now another broken bond on a different atom, a hole is now located on the neighboring atom. The results is that the hole has moved.

Just as the electron can be caused to drift in an electric field, so can the hole, although in the opposite direction because it has a positive charge. Thus, if an electric field is applied, the electrons move one direction and the holes the other. The result is that both contribute to current flow; the conventional current being in the direction of hole flow and opposite the direction of electron flow. This is unlike current flow in metals where only the free electrons contribute to current. The metal ions are fixed.

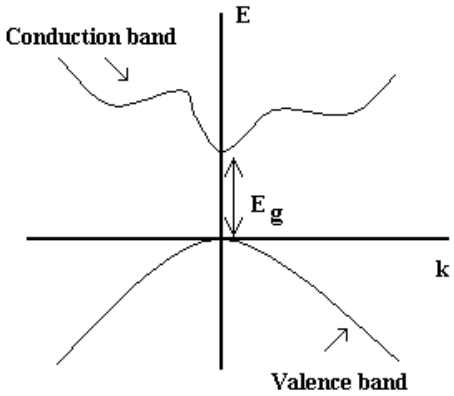
Electron-Hole Pair Generation and Recombination

At any time a hole and a free electron may come together and recombine, thus removing the conducting elements. In fact, recombination occurs naturally and is an on-going process where hole-electron pairs are continuously created and annihilated. When the hole and free electron recombine, the free electron needs to lose the excess energy that enabled it to break free of its silicon atom. This amount of energy is equal or greater than the bandgap of the semiconductor (E_g), the difference between E_C and E_V We use two quantum mechanical particles to describe how this energy is released. The energy may be released as phonons, a quantum of thermal energy or heat, or it may be released as photons, a quantum of light.

Phonons are usually released when the semi-conductor is an indirect semiconductor. Two examples of an indirect semiconductor are silicon and germanium. In these semiconductors, the location of the minimum energy point in the conduction band is not directly above the maximum energy point in the valence band, when plotting the energy of the conduction band and valence band as a function of momentum, k. Since momentum has to be conserved when the electron recombines with the hole, the difference in momentum between the electron and hole is given to another particle along with the excess energy. As a phonon has momentum and a photon does not, a phonon must also be emitted to conserve both energy and momentum.



Photons are usually released when the semi-conductor is a direct semiconductor. An example of a direct semiconductor is gallium arsenide. In these semiconductors, the location of the minimum energy point in the conduction band is directly above the maximum energy point in the valence band, when the conduction band and valence band energies are plotted as a function of momentum, k. The only particle that can be released when the electron and hole recombine that has energy, but no momentum



is a photon. Therefore, light is emitted by direct semiconductors, while heat is emitted by indirect semiconductors, during electron-hole pair recombination Thus, semiconductor laser diodes are composed of direct semiconductors such as gallium arsenide.

As mentioned, the generation of electrons and holes requires energy and when an conduction electron is created so to is a hole created. This energy can be thermal energy, heat, or the energy can come in the form of light, where the energy of the photon has to be equal to or greater than the bandgap of the semiconductor, E_g. This process is more apparent in some applications such as photo-generated current in photoconductors.

CARRIER LIFETIMES

In any semiconductor, hole-electron pairs are in a state of constant change. The total number of electrons and holes is a constant. However, individual electrons are sometimes found in the conduction band and other times are in the valence band. If a hole and an electron should meet, they will recombine - reducing the number of carriers. At the same time, the thermal energy of the crystal is exciting additional electrons to break the covalent bonds. At equilibrium, the number of pairs created is equal to the number that recombine. Thus, any free electron or any free hole will eventually recombine. Carrier lifetime, τ, is a measure of how long any individual electron or hole will remain free.

How would we measure this lifetime? Let us do a thought experiment.

Assume we have a block of silicon at room temperature. Let's attach a voltage source across it and measure the current as shown in Figure A below.

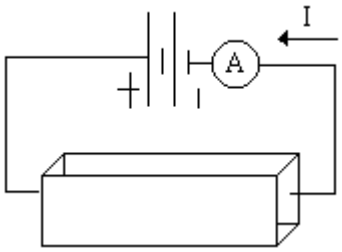


Figure A: A block of silicon with a voltage source

Now we can determine the resistance of the silicon block. From the resistance, we can calculate the electron and hole densities. Then we can determine the total current contribution by the electrons and the holes.

Now let's use a powerful strobe lamp and create a flash of light with an energy greater than or equal to the bandgap of the semiconductor, E_g. When this light falls on the silicon, the photons excite the electrons in the covalent bonds. Some of these electrons

will be excited into the conduction band and create hole-electron pairs in excess of the steady-state intrinsic level. These excess carriers are available for conduction so the current through the silicon block increases as the resistance of the silicon decreases. However, over time, the number of electron-hole pairs decreases because, the system will return to equilibrium without the continued light exposure (since it was only a pulse) providing the extra energy to free the electrons and create the holes. The excess electrons and holes will recombine and the electron and hole concentration returns to the equilibrium carrier concentration, n_i. Thus, the current will decrease in an exponential decay as shown in Figure B. We can measure carrier lifetime by measuring the rate at which the current decays. Carrier lifetime is one time-constant of the current pulse decay. ($\tau = 1/\alpha$) Note the similarity to the time constants for capacitive and inductive circuits.

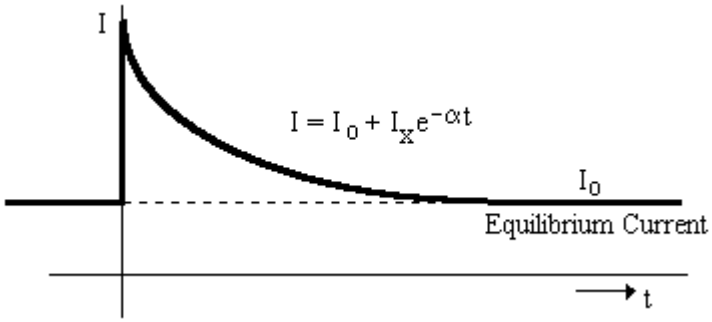


Figure B. Current following a pulse of light.

This experiment also leads us to the idea of photoconductors and photodetectors.

Photoconductors and Photodetectors

In a photoconductor, a voltage is dropped across the length of semiconductor. When a photon of light hits a covalent bond, it can transfer its energy to an electron by absorption, which can cause the covalent bond to disassociate and create a hole-electron pair. Because of the electric field across the semiconductor, the electron and hole move - in opposite directions. For as long as the photo-generated electron and hole exist, a current will flow in the photoconductor. When the electron and hole recombine, the current caused by the photogenerated carriers will cease. Of course, recombination occurs randomly recombining any hole with any electron, not necessarily the original pair. First of all, they are not monogamous. Secondly, they aren't even close to each other after a few femtoseconds, as they drift in opposite directions. In order for an electron to recombine with a hole to create a covalent bond between two Si atoms, they have to be within an atomic length of each other. In the case of Si, the atomic length is 0.235nm. As mentioned above, the average length of time that these photogenerated carriers will exist before they recombine is known as the carrier lifetime, τ.

The process of photon generation of electron-hole pairs can occur when the semiconductor is illuminated with photons that have an energy, $E_{photon} = h\nu = hc/\lambda$, greater than or equal to E_g, the bandgap of the semiconductor. h is Planck's constant and

is equal to 6.63×10^{-34}Js. ν is the frequency of the photon. λ is the wavelength of the photon. Any photons with energy less than E_g will pass through the semiconductor. I.e., the semiconductor is transparent at wavelengths less than its bandgap. Silicon, which has a bandgap of 1.12eV, is transparent to photons with a wavelength of less than 1.1μm, which is in the infrared wavelength region. Thus, Si absorbs visible light and looks grey. Gallium phosphide, on the other hand, has bandgap of 2.3eV. Its bandgap is equal to the energy of green light and, thus, absorbs green and blue light. Therefore, gallium phosphide appears to be orange and somewhat transparent as red and yellow light passes through the semiconductor without being absorbed.

Table 1 is a list of the bandgaps of several semiconductors. Included in this list is diamond. Generally, diamond is considered to be an insulator. As can be observed, its bandgap energy is considerably larger than that of the other semiconductors listed. However, in special situations, diamond can be a semiconductor.

The general trend for bandgaps is that the larger (and heavier) the atoms are, the smaller the bandgap of the semiconductor. Thus, Si (28amu) has a larger bandgap than Ge (73amu). The bandgap is dependent on a number of factors including the strength of the bond between atoms in the crystal and the spacing between each atom (or length of each bond). As the temperature of the crystal increases, the bond length increases and the strength of the bond decreases. So, the bandgap is a function of temperature of the semiconductor, decreasing with increasing temperature.. A general formula for the bandgap as a function of temperature for any semiconductor follows the form of:

$$Eg(T) = Eg(0K) - \frac{A\,T^2}{(B+T)}$$

Therefore, the lowest energy photon that the semiconductor can absorb changes with the temperature of the semiconductor. This change in bandgap energy as a function of temperature will also be important in understanding why the bias voltages of diodes and transistors vary with temperature.

Table 1 - Bandgap Energy

Semiconductor	Bandgap at 300K	Bandgap at 77K
Silicon	1.12eV	1.17eV
Gallium Arsenide	1.424eV	1.508eV
Germanium	0.661eV	0.733eV
Indium Phosphide	1.35eV	1.414eV
Diamond	5.46eV	5.48eV

Typical applications that rely on photogenerated carriers are in opto-isolators and opto-interrupters where the light source is turned on and off or where an object interrupts the light beam. The speed of detection and extinction is limited by the lifetimes of the carriers, typically nanoseconds to a few microseconds.

CARRIERS IN INTRINSIC SEMICONDUCTORS

In semiconductors, it is conventional to use the symbol n for the electron density, the number of electrons per cm^3, and p for the hole density or the number of holes per cm^3. In the intrinsic semiconductor, holes and electrons are created in pairs, n = p and the common symbol is usually designated n_i, where the subscript indicates intrinsic material, or pure silicon. n_i is also known as the intrinsic carrier concentration.

$$n_i^2 = np \qquad (14)$$

The number of thermally generated electron-hole pairs is a function of the semiconductor material and its temperature.

$$n_i = (N_c N_v)^{1/2} \exp(-E_g/2kT) \qquad (15)$$

where:

$$N_c = 2\left(\frac{2\pi m_e kT}{h^2}\right)^{3/2} \qquad (16)$$

$$N_v = 2\left(\frac{2\pi m_h kT}{h^2}\right)^{3/2} \qquad (17)$$

m_e and m_h are the effective mass of the electron and hole in the semiconductor, respectively. Strangely enough, these values can be very different from the mass of an electron in vacuum, $m_o = 9.1 \times 10^{-31}$ kg. The effective mass for electrons and holes for a number of semiconductors are given in Table 2.

Table 2

Semiconductor	Electron Effective Mass, m_e	Hole Effective Mass, m_h
Silicon	1.18 m_o	0.81 m_o
Germanium	0.55 m_o	0.36 m_o
Gallium Arsenide	0.063 m_o	0.48 m_o

Temperature Dependence of Hole-Electron Concentration

As the temperature increases, the electrons get more energetic and the probability that an electron is excited into a higher energy state, the conduction band, is increased. Thus, the number of electrons in the conduction band is increased and the number of hole-electron pairs is increased. Finally, the conductivity is increased. The equation describing n_i specifically for intrinsic silicon is

$$n_i = p_i = 4.8x10^{15} T^{\frac{3}{2}} e^{(\frac{-5797 E_g}{T})} / cm^3$$

T is the temperature in degrees K, and E_g = 1.12 eV, the energy gap in silicon.

This equation gives $n_i = 1.5x10^{10}$ per cm^3 at room temperature, 300 degrees K, or 27 degrees C. At 60 degrees C, $n_i = 10^{11}$ per cm^3

We can use this characteristic to build temperature sensitive devices to measure temperature or to compensate for temperature changes such as thermistors.

CURRENT IN INTRINSIC SEMICONDUCTORS

Conduction due to the electrons is

$$J_n = qn\mu_n E \qquad (18)$$

where the subscript n indicates electrons. Similarly, the hole conduction is

$$J_p = qp\mu_p E \qquad (19)$$

where this time, the subscript p indicates holes. As the different physical properties of the electron and hole are different, the mobilities and effective masses of the electrons and holes are not the same. Therefore, the hole current and electron current need to be calculated separately and then added together. The total current density is

$$J = J_n + J_p = qn\mu_n E + qp\mu_p E = q(n\mu_n + p\mu_p)E \qquad (20)$$

$$= qn_i(\mu_n + \mu_p)E \qquad (21)$$

Conductivity in intrinsic material is

$\sigma_i = qn_i(\mu_n + \mu_p)$ *[Be careful here, this equation applies only for intrinsic material*

where $n = p = n_i$ *]* (22)

It is of interest at this time to calculate the conductivity of intrinsic silicon and compare it to the conductivity of aluminum.

Example: Calculation of conductivity of intrinsic silicon.

Several properties of silicon at 300K (room temperature) are given in Table 3. From these properties we see that the intrinsic carrier density is 1.45×10^{10} carriers per cm^3, the mobilities of electrons and holes are 1500 and 475 cm^2/Vs, respectively. Conductivity is then

$$\sigma_i = 1.602 \times 10^{-19} \{coul\} \times 1.45 \times 10^{10} \{cm^{-3}\} \times (1500 + 475) \{cm^2/Vs\}$$

$$= 4.59 \times 10^{-6} (\Omega\ cm)^{-1}$$

TABLE 3
PROPERTIES OF SILICON

Atomic Number	14	μ_n at 300° K (cm^2/Vs)	1500
Atomic Weight	28.1	μ_p at 300° K (cm^2/Vs)	475
Density (gm/cm^3)	2.33	n_i at 300° K (cm^{-3})	1.45×10^{10}
Atoms/cm^3	5.0×10^{22}	D_n at 300° K (cm^2/s)	34
ρ_i at 300° K (Ωcm)	2.30×10^5	D_p at 300° K (cm^2/s)	13
Dielectric constant ε_r	11.7	Electron affinity χ_e	4.05eV

Compare this with the conductivity of aluminum at 2.9×10^5 $(\Omega\ cm)^{-1}$. Obviously, intrinsic silicon is not nearly as good a conductor as aluminum. We will see shortly, however, that adding impurities to silicon can change its conductivity by several orders of magnitude.

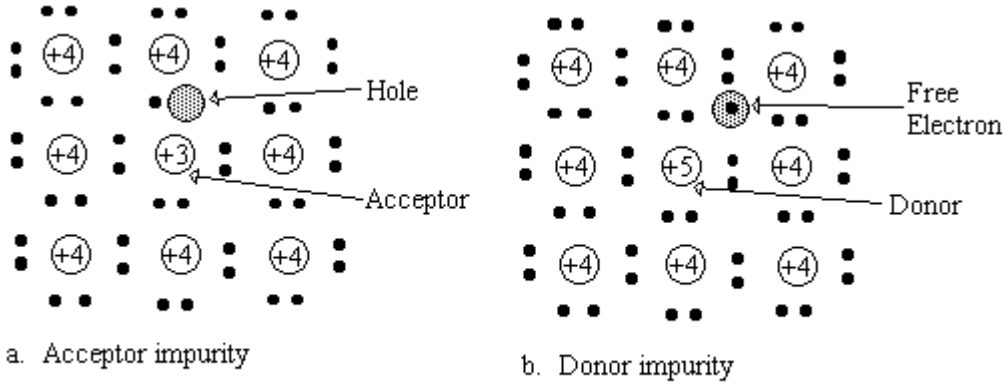
EXTRINSIC SEMICONDUCTORS

In order to increase the number of free electrons or holes in a semiconductor, impurities are introduced. There are two types of impurities used, donors and acceptors. In Si, donor atoms typically have 5 valence electrons. In the crystal structure, four are bound in covalent bonds with adjacent silicon atoms. The remaining valence electron is easily disassociated, thus donating an electron for conduction. Only about 5 to 50meV are required to remove these remaining valence electrons. Since the amount of thermal

energy available is equal to kT, where k is Boltzman's constant ($8.6x10^{-5}$eV/K) and T is temperature in Kelvin. At room temperature (~300K), the kinetic energy available to the system is ~25meV, which is greater than the average energy required to ionize an impurity in a semiconductor. Thus, essentially all of the extra electrons on the donor atoms are free conduction electrons at room temperature.

Acceptor atoms usually have 3 valence electrons, and in the crystal structure, leave a valence vacancy which can trap an electron, thereby creating a hole for conduction. The energy required for this is typically slightly more than what is required to ionize a donor - 10-100meV. However, most of the acceptors are ionized (created a hole) when the temperature of the semiconductor is above 100K. Pictorial examples are shown in Figure 8. A periodic table which identified the possible donor and acceptor atoms in IV semiconductors is below.

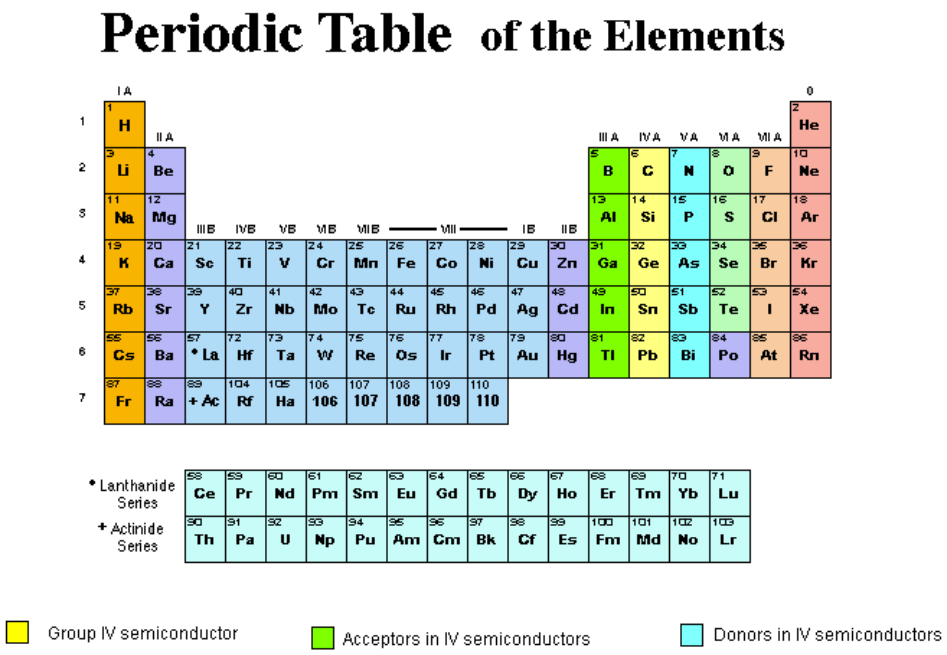
Silicon with an excess of free electrons is called n-type while silicon with an excess of holes is called p-type. Typical donors are pentavalent atoms: antimony, phosphorous, and arsenic. Typical acceptors are trivalent atoms: boron, gallium, and indium. In GaAs and other III-V semiconductors, column II atoms are acceptors and column VI atoms are donors. Column VI atoms in GaAs are donors if they occupy a Ga site and are an acceptor if they occupy an As site. Thus, Column IV atoms (e.g., Si and Ge) in III-V semiconductors are called amphoteric dopants, meaning that they can act as either a donor or an acceptor.



a. Acceptor impurity

b. Donor impurity

Figure 8. Silicon with impurities.



Periodic Table

These impurities are introduced into the silicon either during crystal growth or after a crystal of pure silicon has been grown and sliced into wafers. Silicon wafers are grown by melting a large quantity of noncrystalline silicon in a carbon crucible. The melting point of silicon is 1460^{o}C. To dope the silicon with impurities, an appropriate amount of donor or acceptor atoms are mixed into the melted silicon. A small single crystal of silicon, called a seed crystal, is then lowered slowly into the silicon melt, but pulled out before it completely melts. Because of silicon's high surface tension, some of the liquid silicon is pulled out with the seed crystal and begins to cool and solidify. The solidified silicon using the structure of the seed crystal as a template and, thus, grows as a single crystal. The seed crystal is continued to be pulled away from the silicon melt such that there is alway some liquid silicon wetting the solidified silicon. The solidified silicon is called a boule. Once the growth process is complete, the boule is sliced into wafers, which are then polished to remove the mechanical damage from the sawing process.

The impurities can also be selectively introduced into the silicon wafer via diffusion or ion implantation. Typically, the wafer is allowed to oxidize to create a surface film of silicon dioxide, SiO_2. Then some of the SiO_2 is etched away in places where n-type or p-

type regions are wanted. During a diffusion, the wafers are heated and exposed to high concentrations of the desired impurity. The impurity diffuses through the openings in the SiO_2 into the silicon displacing an occasional silicon atom in the crystal lattice. Upon cooling, the impurity atoms freeze in place, creating the n- or p-type regions. In ion implantation, a beam of high energy ions (ionized donor or acceptor atoms) bombard the surface of the silicon wafer. The SiO_2 prevents the ions from reaching the silicon surface. However, they penetrate the silicon where ever the SiO_2 has been removed. The silicon wafer is then heated to 700-1000^oC to give the ions enough energy to incorporate into the silicon lattice.

Only those impurities that occupy a site in the crystal lattice where a silicon atom would have been act as donors or acceptors and are called substitutional impurities. Those impurities that are in the silicon crystal but do not sit on a silicon site do not contribute a electron or hole to the semiconductor and are call interstitial impurities. Interstitial impurities, though not electrically active, reduce the mobility of the electrons and holes in the semiconductor as they will act as scattering sites as the electrons and holes move through the crystal

Carrier Concentrations

In intrinsic material, hole-electron pairs are thermally generated. These hole-electron pairs can also recombine. This process of generation-annihilation goes on continuously with the number of pairs at some equilibrium, where the number of electrons equals the number of holes. The average amount of time a hole or electron is free is called the carrier lifetime, τ. In n-type extrinsic material, there is no hole corresponding to the donated electron. There are still thermally generated hole-electron pairs, but the equilibrium is shifted so that there are fewer thermally generated carriers.

Within the extrinsic semiconductors, two processes take place which contribute to the carrier concentration. Hole-electron pairs are created by thermal agitation just as in the intrinsic case. In addition, the impurity atoms are virtually all ionized creating the corresponding carriers. Usually, the doping level is high enough that the extrinsically provided carriers predominate over the thermally generated carriers.

If both types of impurities are present simultaneously, they simply cancel each other out; an acceptor ion captures the excess electron provided by a donor ion. This is called compensation. Only the excess impurity concentration contributes to the conductivity. In most cases, one type of impurity predominates, creating an extrinsic material of the desired type. In perfectly compensated material, the concentration of electrons equals the concentration of holes and both are equal to the intrinsic carrier concentration, n_i. But, the mobility of the electrons and holes is lower than in intrinsic material because of the larger number of impurity ion to scatter with.

Because of the physics of hole-electron pair formation, the total number of free carriers in the semiconductor is controlled by the mass-action law, equation 16 which is $\mathbf{np = n_i^2}$. Thus, in an n-type material, the number of electrons is increased due to the donor ions, thereby causing the number of free holes to be reduced. In the n-type material, electrons are the majority carrier. Similarly in p-type material, holes are the majority carrier. Let N_D represent the concentration of donor ions and let N_A represent the acceptor concentration. The material must maintain charge neutrality so that the total number of positive charges will equal the total number of negative charges.

$$N_D + p = N_A + n \qquad (23)$$

Donors give up electrons so they become a positively charged ion. Ionized acceptors have a negative charge as they have captured an extra electron. In an uncompensated n-type material, $N_A = 0$ and because virtually all the donors will be ionized at most temperatures,

$$N_D + p = n \qquad (24)$$

Here the free electrons, n, have two sources; donors, N_D, and thermally created hole-electron pairs, p. In all but extremely lightly doped materials, the number of impurity atoms is a much greater in concentration than is the intrinsic electron concentration, n_i. In other words, $N_D >> n_i$. Thus, $n \cong N_D$. We can find the hole concentration from the mass action law ($np = n_i^2$)

$$p = n_i^2/n \cong n_i^2/N_D \qquad (25)$$

Likewise, when there is p-type semiconductor, the number of holes is determined by the number of acceptor atoms in the material. Thus, $p \cong N_A$. Again, we can find the electron concentration from the mass action law

$$n = n_i^2/p \cong n_i^2/N_A \qquad (26)$$

Now let's look at the conductivity of a doped n-type material.

Example: Conductivity of doped material

Assume silicon is doped with donors so that $N_D = 5 \times 10^{12}$ atoms/cm^3. Determine the conductivity of the material.

The electron density $n \approx N_D$ and we can find the hole density from the mass action law.

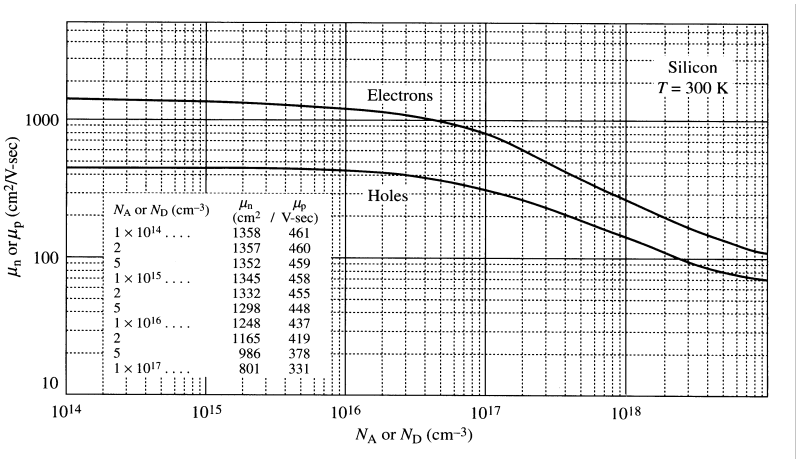
$$p = (1.45 \times 10^{10})^2/5 \times 10^{12} = 4.21 \times 10^7 \text{ /cm}^3$$

(Note that p = 4 x 10^7 are all thermally generated holes. There are an equal number of thermally generated electrons, but this number of thermally generated electrons is insignificant when compared to the number of donors that contributed electrons.)

The conductivity then is

$$\sigma = q(n\mu_n + p\mu_p)$$

However, the added donors do more than increase the number of electrons in the material, they also decrease the mobility of the electrons and holes. This is because they act as Coulombic (or ionic) scattering sites (see the figure in the discussion on electron mobility in metals). The donor atons also slightly distort the unit cell because they are not exactly same size as the silicon atom they replace and the bonds that they form with the adjacent silicon atoms are not the same length as the Si-Si bonds.



The figure above shows the relationship between mobility and doping levels for both n-type and p-type silicon. For the doping level in the example, the donor ions do not significantly change the mobility of the carriers. Therefore, the the mobility of the holes and electrons are equal to the value of intrinsic silicon. The lower values of mobility would need to be used when calculating the conductivity of more heavily doped material, 10^{15} cm^{-3} to 10^{18} cm^{-3}, which are typical doping levels for diodes and transistors.

Thus, for our example,

$$\sigma = 1.602 \times 10^{-19} (5 \times 10^{12} \times 1500 + 4.21 \times 10^{7} \times 475)$$

$$\sigma = 1.602 \times 10^{-19} (7.5 \times 10^{15} + 2.00 \times 10^{10})$$

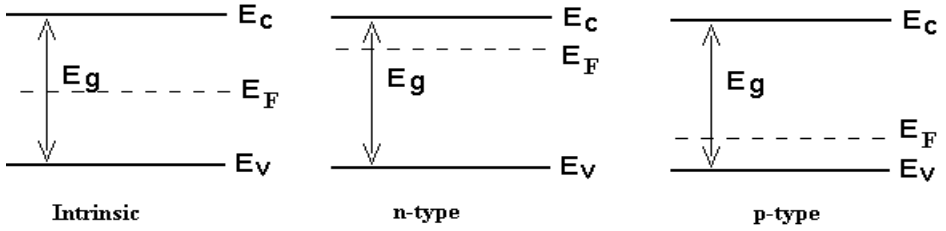
$$= 1.2 \times 10^{-3} (\Omega cm)^{-1}$$

This conductivity is almost 3 orders of magnitude greater than for intrinsic silicon.

Note that the electron density is five orders of magnitude greater than the hole density which means that the holes contribute very little to the conductivity. It can be seen from this example that if the doping level is significantly higher than the intrinsic electron/hole density, the majority carrier does most of the conducting in the material and that the conductivity is directly proportional to the doping level.

Fermi Level

When drawing a energy band diagram, we indicate what type the material is (n type or p type) by the placement of the Fermi level, E_F. The Fermi level is an imaginary energy level. When it is positioned halfway in between the conduction band and valence band, the material is intrinsically doped or evenly compensated. When the Fermi level is close to the conduction band, the material is n-type. The closer the Fermi level is to the conduction band, the more heavily doped is the material - the larger N_D is. When the Fermi level is close to the valence band, the material is p-type. The closer the Fermi level is to the valence band, the more heavily doped is the material - the larger N_A is.



No matter what the temperature of the semiconductor is, the Fermi level is halfway between the conduction and valence bands for an intrinsic semiconductor. However as the temperature increases, the Fermi level in a doped semiconductor moves towards the middle of the bandgap. The reason for this is that the intrinsic carrier concentration increases significantly with temperature. Thus, the extrinsic carrier concentration becomes less important as the temperature increases. Another way to express this is that as silicon becomes intrinsic as the temperature increases because $n_i > N_D$ or N_A. This is one reason why there is a maximum temperature of operation for most semiconductor devices.

CONDUCTION BY DIFFUSION

So far we have talked exclusively about conduction by drift of carriers in an electric field. However, conduction by diffusion is important in some cases. Diffusion result from the natural tendency of most particles move from regions of high concentration to regions of

lower concentration. That is, diffusion is driven by concentration gradients while drift is driven by electric fields. Everyone has experienced an example of diffusion where a person opens a bottle of perfume (or a package of fish), in a few minutes, a different person in another part of the room can smell it.

The reason this diffusion happens is illustrated in Figure 9. In this figure, we have simplified the situation by assuming only four possible directions of motion: up, down, right, and left. In random motion, one fourth of the particles will be going in each direction. Now the left side of the volume has one half the density of the right side. Thus, there are twice as many particles going left across the boundary as there are going right in any period of time. In a short time then, there will be more particles on the left than as shown in the diagram. After while, the two sides will have the same density and the diffusion current will stop. The random motions will average out on both sides.

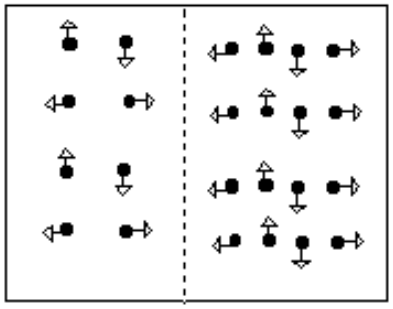


Figure 9. Illustration of diffusion

For n-type material, the diffusion current is given by

$$J_n = qD_n \frac{dn}{dx} \qquad (27)$$

where D_n is the diffusion constant. Note that if dn/dx has a positive slope (rises to the right), the concentration is higher to the right and electrons will diffuse to the lower concentration, to the left. Because the electrons have a negative charge, the current will be to the right, or in the positive x direction. Thus, if we look at diffusion current in a p-type material,

$$J_p = -qD_p \frac{dp}{dx} \qquad (28)$$

Here the negative sign reflects the fact that the holes will diffuse to the left causing current in the negative x direction.

The diffusion constant of a hole or electron is related to that carrier's mobility by the

following equation:

$$D/\mu = kT/q \qquad (29)$$

If we have both an electric field and a concentration gradient, the total current density is the sum of the drift current and the diffusion current densities.

$$J_n = q\mu_n nE + qD_n \frac{dn}{dx} \qquad \text{for electrons} \qquad (30)$$

and

$$J_p = q\mu_p pE - qD_p \frac{dp}{dx} \qquad \text{for holes} \qquad (31)$$

The total current density is

$$J = J_n + J_p$$
$$= q(n\mu_n + p\mu_p)E + q(D_n \frac{dn}{dx} - D_p \frac{dp}{dx}) \qquad (32)$$

(Drift) *(Diffusion)*

THE pn JUNCTION

We have now set the stage for looking at the p-n junction, the basis for most semiconductor devices. Figure 10 shows blocks of n-type and p-type material. In these drawings, only the donor and acceptor ions are represented along with the resulting free electrons and holes, the majority carriers in each material. The minority carriers, the holes in the n-type material and the electrons in the p-type materials, are not shown here to keep from cluttering up the diagram, especially since there are so few of them

.

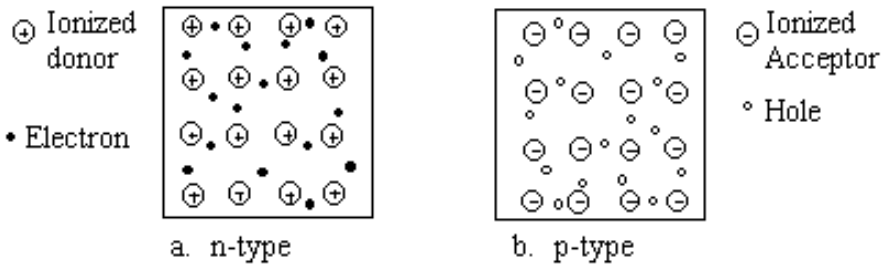


Figure 10. n-type and p-type blocks of material.

In each block in Figure 10, the positive and negative charges balance. Now if we bring those blocks together as shown in Figure 11, we still have charge neutrality, but we have

a concentration step dislocation. We have a high concentration of electrons to the left of the junction and a high concentration of holes to the right.

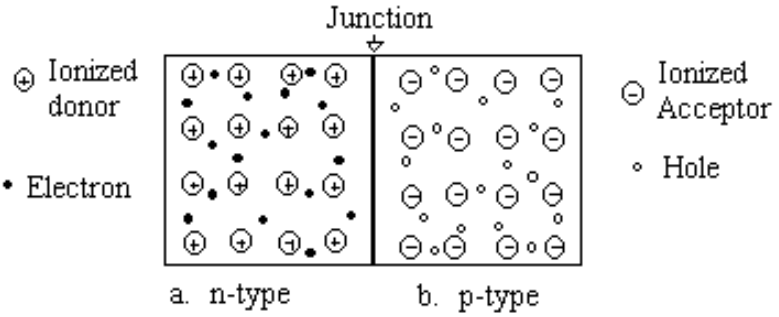


Figure 11. p-n junction at moment of contact.

The donor and acceptor ions are fixed by the crystal lattice, but the holes and electrons are free to move. These carriers diffuse to the regions of lower concentration until the device comes to equilibrium shown in Figure 12.

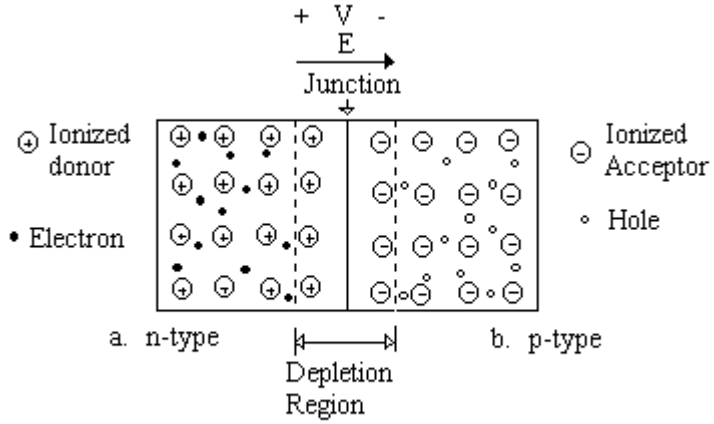


Figure 12. p-n junction at equillibrium

In Figure 12, some of the electrons diffused over into the p-type material and some of the holes diffused over into the n-type material. These become minority carriers and are not shown. In fact, outside the depletion region, most of the excess minority carriers recombine with majority carriers and thus, are annihilated. Obviously, the carriers closest to the junction were the first to migrate by diffusion. This migration left behind some charged ions in the area marked as the depletion region. These charged ions create an electric field with the positive side in the n-type material and negative side in the p-type material. This field is essentially a potential hill which repels further migration. The carriers left behind, beyond the depletion region, still have thermal energy and move randomly in the area. If an electron had enough energy to overcome the potential barrier, and was moving in the proper direction, it would pass across the junction. This electron

would leave another charged ion behind, increasing the potential barrier, making it harder for others to cross.

Now going back to the minority carriers. These are not shown in the diagram, but they too are distributed throughout their respective areas outside the depletion region. This distribution means that some are near the electric field at the edge of the depletion region, but instead of being repelled by it, they are attracted by it and are pulled across the junction. This act has the opposite effect on the barrier height, essentially neutralizing an ion, and lowering the barrier.

The result is a continuous motion of charges back and forth across the junction at equilibrium. Minority carriers diffuse to the junction, are swept across the depletion region. This causes the potential barrier to be lowered which allows majority carriers to diffuse against the barrier. The charge motion is random, with the total current balancing to zero. However, at any one instant in time, there is a measurable current flowing across the p-n junction, even with zero voltage applied. The currents are quite small compared to other currents we will be looking at later. However, these currents are important in the operation of many devices. For example, these currents contribute to the dark current of a zero biased photodetector.

It is often of interest to consider what happens if the doping levels are not the same on both sides of the junction. First, we note that charge neutrality in the device must be maintained and in fact is maintained both in the depletion region and in the non-depleted regions of the n-type and p-type materials. The depletion region is charge neutral, with the positive ions balancing the negative ions. Thus, if one of the regions has a higher concentration of doping ions, that part of the depletion region will be narrower than that in the region with lower doping. Similarly, if both regions had higher doping, the depletion would be narrower and the field gradient or field strength would be higher although the total potential difference would be the same. The depletion width, W, can be calculated using the following equation:

$$W = \left(2 \frac{\varepsilon_r \varepsilon_o}{q} \frac{N_A + N_D}{N_A N_D} (V_{bi} - V_a) \right)^{1/2} \qquad (33)$$

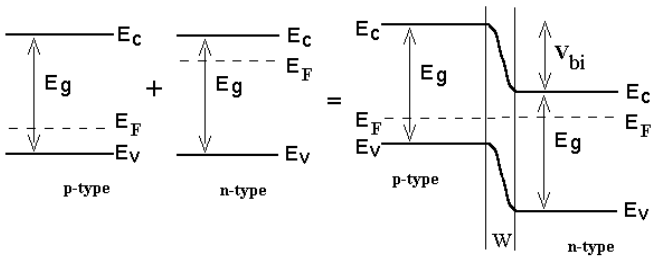
where ε_r is the relative dielectric constant which is between 10-12 for most semiconductors, ε_o is the permittivity of free space (8.85×10^{-14} F/cm), V_{bi} is the built-in voltage of the p-n junction, and V_a is any externally applied voltage across the p-n junction.

The built-in voltage of the p-n junction is equal to the electric field developed across the depletion width times the width of the depletion layer. It is equal to:

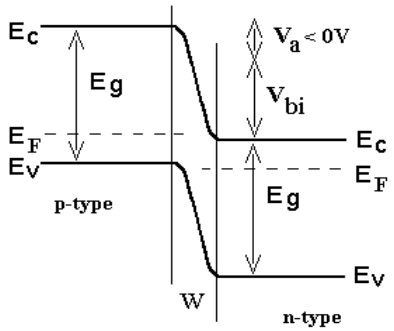
$$V_{bi} = \frac{kT}{q} \ln \left(\frac{N_A N_D}{n_i^2} \right)$$

(34)

One can see how the built-in voltage is developed when a p-type region is in contact with an n-type region by using the energy diagrams and the concept of continuity of the Fermi level. Continuity of the Fermi level means that, at zero bias, we redraw the energy diagrams of the p-type and n-type regions so that the Fermi levels of the two materials are at equal positions. Thus, the electrons in the n-type material need to gain enough energy, about V_{bi}, to overcome the energy barrier, the "hill" in the energy diagram, in order to move into the p-type material. And, the holes also need to gain energy to flow into the n-type material. [Increasing energy for a hole is negative energy.]



As can be seen from equation 33, as the voltage across the junction increases (i.e, the junction is reverse biased, $V_a < 0V$), the depletion region width increases as well. Furthermore, the Fermi level of the p-type material is no longer drawn at the same energy as the Fermi level as the n-type material and the energy barrier for the holes to flow into the n-type material and electrons to flow into the p-type material increases. Changes in



the width of the depletion region will result in charge transfers, similar to those in capacitors. This phenomenon give s rise to the concept of junction capacitance, C_j.

$$C_j = \varepsilon_r \varepsilon_o A/W \qquad (35)$$

where A is the cross-sectional area of the p-n junction.

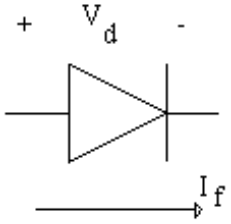
At higher doping levels, the depletion width is thinner and we get a higher capacitance at

the junction. In some cases, this higher capacitance can cause problems with device speeds when there are large voltage changes, but also the narrow depletion region means carriers crossing the region take less time (the transit time, t_t, is small) and that switching speeds can be increased if the change in voltage is low. In general, there is a speed/current trade-off in the design of these junctions. We will see the results of this trade-off later in this course. The design of semiconductor devices is a fascinating subject, but requires material beyond the scope of this course.

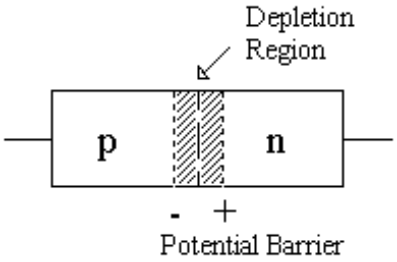
In forward bias, $V_a > 0V$, the depletion region decreases, the junction capacitance increases, and the electrons in the n-region need less excess energy to flow into the p-region/the holes in the p region needs less excess energy to flow into the n-region. From the discussion thus far, you would expect that when the externally applied voltage, V_a, is equal to the built-in voltage, V_{bi}, the junction capacitance becomes infinite. But, of course, this doesn't happen because the resistance of the junction becomes zero, which shorts the junction capacitor.

DIODES

A pn diode is the simplest junction semiconductor device. A simple representation of a a p-n junction is shown below. At the junction, a depletion region forms where all free carriers have gone away, leaving the fixed donor and acceptor ions which cause a static field. This static field just balances the tendency of the majority carriers on either side to diffuse into the depletion region.



Diode symbol



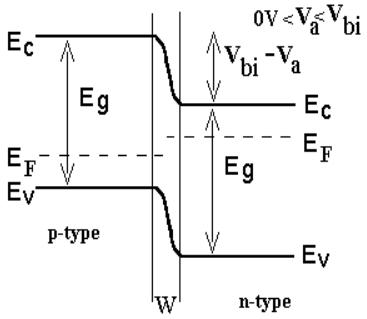
Diode representation

In the steady-state case, there are several processes going on all the time that result in zero net current flow. (1) Majority carriers, holes from the *p* region and electrons from the *n* region diffuse into the depletion region. A few are energetic enough to cross the potential barrier, V_{bi}, and result in a current flow. Balancing this flow, (2) minority carriers, holes in the *n* region and electrons in the *p* region diffuse into the depletion region. The field sweeps them across the region into their respective majority regions. (3) Hole-electron pairs are continuously created in the depletion region by thermal agitation. These carriers are swept by the field from the depletion region. (4) Recombination of hole-electron pairs balances the thermal generation to maintain charge neutrality.

The previous discussion has assumed no external connections. Let us now consider what

happens when a short circuiting wire is connected externally to the two ends of the device. Actually, the experiment is a big flop, nothing happens. A little thought will reveal to us that while there is contact potential difference between the two different semiconductor materials, there is also a contact potential difference between the wire and the semiconductor materials at both ends. It should come as no surprise that these contact potentials cancel each other; the sum of the contact potentials around the circuit is zero.

If a forward voltage is applied to the diode, positive to the *p* material and negative to the *n* material, the potential barrier is decreased - by the value of the applied voltage if we assume that there are no resistive voltage drops in the p and n regions. Then, majority carriers diffusing into the depletion region can more easily overcome the potential barrier. Because both holes and electrons are involved and have opposite charges, the result is a net current flow from left to right. In many cases, the doping levels will be unbalanced (either $N_D >> N_A$ or $N_A >> N_D$) such that either the hole current or electron current will dominate and the other will be ignored.



If a reverse voltage is applied to the diode, the additional voltage across the depletion region will widen it. There will also be a very small reverse current, called the reverse saturation current. This current arises from the minority carriers diffusing to the depletion region and being swept across. This current is no longer balanced by forward diffusion of majority carriers because the potential barrier is now much higher and is maintained by the external source.

$$I_s = qA\left(\frac{D_n}{L_n}\frac{n_i^2}{N_A} + \frac{D_p}{L_p}\frac{n_i^2}{N_D}\right) \qquad (36)$$

where D_n is the diffusion constant of minority electrons in the p region, L_n is the diffusion length of these minority electrons, D_p is the diffusion constant of minority holes in the n region, and L_p is the diffusion length of these minority holes. The minority carrier diffusion length is related to both the minority carrier diffusion constant and the lifetime of these minority carriers, τ.

$$L = (D\,\tau)^{1/2} \qquad (37)$$

It is the change in potential that is responsible for the difference in resistance of a diode when measured using a digital multimeter. Try this yourself, measure the resistance of a diode with the probes connected in one way. Then, measure the resistance again after switching the probes to the opposite terminals of the diode. One measurement will be very high, usually 1-10MΩ. The other resistance measurement will be much lower. The

reason for this is that the multimeter applies a very small voltage to the component under test. In one probe configuration, the diode is slightly reverse biased and a extremely small amount of current flows. In the other configuration, the diode is slightly forward biased and a much larger current flows.

If we are going to use diodes in circuits, we must have some way to relate the voltage to the current. We will not discuss the physical electronics here, but it can be shown theoretically that the current is an exponential function of the voltage

$$I = I_S \left(e^{\left(\frac{v_d}{v_T}\right)} - 1 \right) \tag{38}$$

where I_S is the reverse saturation current, v_d is the voltage across the diode in the forward direction shown earlier, and v_T is the thermal voltage

$$v_T = \frac{kT}{q} \tag{39}$$

where in this case, k is Boltzman's constant, T is the temperature in ^{O}K, and q is the electronic charge. We normally just use v_T in voltage form for simplicity. At room temperature, 300 ^{O}K, v_T = 25.9 mV. A plot of the diode equation is given in Figure 13. Note the change in scale for the forward and reverse currents.

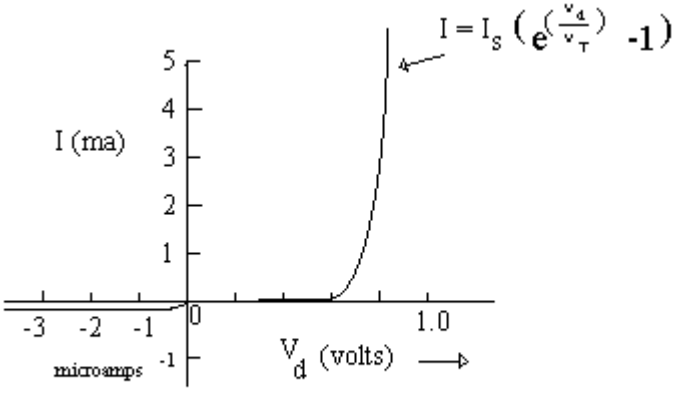


Figure 13. I-V characteristic for a silicon diode.

The current flows easily when the applied voltage is equal to the built-in voltage, a point known as the flat band condition - where the conduction band energies and the valence band energies of the p region are equal to that of the n region. At this applied voltage, the depletion width is equal to zero. The capacitance is now dominated to another capacitance, known as the diffusion capacitance. This is a capacitance associated with the excess minority carriers that are injected into the p and n regions. In order to switch the diode off, these excess minority carriers must be removed from the layers in which they have been injected. Thus, there is a transient current, t_{rr}, that flow as the applied voltage on a diode is switched from $V_a > V_{bi}$ to $V_a < 0V$ and the minority electrons that

were injected into the p region and the minority holes that were injected into the n region return to the n region and p region, respectively.

Non-Ideal Factors That Affect Terminal Characteristics

Because of many factors that arise during manufacturing, the measured characteristic is somewhat different. Nevertheless, for moderate current densities, the diode equation serves quite well. There are significant differences at low currents. One of the most significant is caused by surface imperfections that cause I_S to be much larger than predicted. Also, I_S is a function of temperature which can cause significant reverse leakage currents at high temperatures due to the strong temperature dependence of n_i and of the mobility of the minority carriers. At very high current densities, the resistance of the bulk semiconductor materials can have significant voltage drops that must be added to the junction characteristic to obtain the terminal characteristic. Lastly, as the reverse bias voltage is increased, at some point the p-n junction breaks down and a large amount of current flows. This reverse voltage is called the breakdown voltage, V_{BR}, and is a function of the doping levels, N_A and N_D. The closer the doping levels are to n_i, the larger V_{BR} is.

Similarly to the electron mobility in metals, the mobility of the electrons and holes in a semiconductor also have the same trend with temperature, initially increasing and then decreasing as temperature is increased. However, the relationship between resistance and temperature isn't quite as straight-forward due to the fact that the intrinsic carrier concentration, n_i, has an extremely strong dependence on temperature.

To further complicate the calculation of the resistance as a function of temperature, the bandgap of the semiconductor, E_g, is, also, a strong function of temperature. The general formula for the change in bandgap as a function of temperature is

$$E_g(T) = E_g(0K) - \frac{A\,T^2}{(B+T)} \qquad (40)$$

For Si, $E_g(0K)$ is 1.17eV, A is equal to 4.73×10^{-4} and B is equal to 636K. T, the temperature, is in degrees Kelvin. As can be seen from the formula, the bandgap of any semiconductor decreases with increasing temperature.

This complicates the calculation of n_i as the bandgap energy is a term in the equation (see eq. 15), with the result that n_i increases even faster with increasing temperature. Because n_i increases rapidly with temperature – faster than the mobility decreases with temperature, the resistance of a semiconductor decreases with increasing temperature.

All of the thermal changes in various material properties of silicon impact the performace of a silicon p-n junction. For example, the change in the turn-on voltage (V_{bi}) for silicon

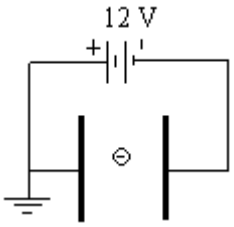
diodes is approximately –2mV/^{o}C for changes near room temperature. Other device parameters are similarly affected by a change in temperature. The resistance of a silicon diode decreases with temperature because both the series resistance (the sum of the resistance of the bulk p-type and n-type semiconductor layers) and the junction resistance (which is inversely related to the rate of thermal generation of carriers) decrease with increasing temperature. The junction capacitance decreases with increasing temperature because the materials become less extrinsically doped as the intrinsic carrier concentration increases with temperature. When designing circuits that need to work in all types of environments – not just in the lab, engineers need to evaluate their designs to make sure that they work over the expected temperature ranges.

All of these changes limit the maximum operating temperature of silicon devices. Generally, silicon devices are limited to well under 300^{o}C operation. Larger bandgap semiconductors are less affected by temperature and, thus, can operate to higher temperatures. SiC devices have been operated up to 600^{o}C. Diamond diodes are expected to be able to operate at far beyond this temperature. On the other hand, HgCdTe diodes, which have an extremely small bandgap, can not operate at room temperature, but must be operated at temperatures equal to that of liquid nitrogen (77K) or below.

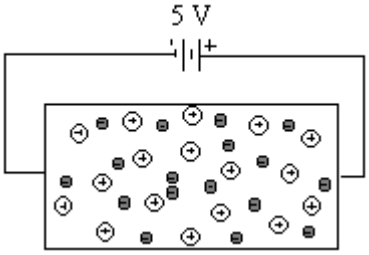
When you look at a data sheet, the operating temperature is specified. The manufacturer has guaranteed that the device or circuit will operate over this temperature range with the operating characteristics listed. Also, certain numerical assignments have been designated to identify the operating temperature range for some commonly used circuits. For example, digital logic circuits that begin with the numbers 74 are designed for commercial applications and generally can be operated from 0^{o}C to 70^{o}C. Whereas circuits that begin with the numbers 54 are designed to meet military specifications, the specified operation temperature range is –55^{o}C to 125^{o}C.

EXERCISES

1. An electron is released in the middle between two metal plates which are connected to the voltage source as shown.
   a. Draw an arrow beside the electron showing which way it will move.
   b. Draw an arrow showing the direction of current.
   c. If you stood on the grounded plate and threw the electron toward the other plate, how much energy would the electron initially have to have in order to reach the other plate?



2. An intrinsic silicon material is connected to a voltage source as shown. Draw arrows indicating the directions of motion for the holes and electrons. Which direction does current flow?

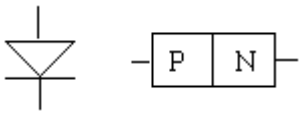


Holes Electrons Current

3. The doping level for an n-type silicon is $5x10^{15}$ donors/cm^3.
   a. What is the density of electrons? n = _____
   b. What is the density of holes? p = _____
   c. What is the conductivity? σ = _____

4. A pn junction has a reverse saturation current I_o = 1 μA. A forward voltage of 0.65 volts is applied. What is the current? I_F = _____
If a reverse voltage of 0.65 volts is applied, what is the current? I_R = _____.

5. The symbol for a diode is shown along with a pn junction.
   a. Draw the polarity that will cause forward current in both devices.
   b. Draw an arrow showing the direction of forward current for both devices


P
N

Problems

1. Determine the conductivity of copper by using resistance and wire diameter in wire tables in the "Electrical Engineering Design Compendium".

2. Determine the mobility of electrons in copper which has an atomic weight of 63.54, a density of 8.89×10^3 kg/m^3, and one valence electron.

3. Two parallel plates 10 cm apart have a 10 volt DC voltage source connected between them. If you dropped an electron from the negative plate, how long would it take to reach the positive plate? If you stood on the positive plate and threw an electron toward the negative plate, what initial velocity is required for it to reach the negative plate?

4. What fraction of atoms in intrinsic silicon have broken co-valent bonds at 300^{o}K?

5. You have a bar of intrinsic silicon that has a cross sectional area of 1 cm^2 and is 10 cm long.
   a. What is the resistance between the ends?
   b. What is the resistance of an aluminum bar of the same dimensions?
   c. What is the current density in both bars if a 10 volt source is connected at the ends?
   d. What is the average drift velocity in each?

6. What doping level is required to make n-type silicon have a conductivity of 1 $(\Omega cm)^{-1}$? What would the doping level be if the material was p-type?

7. Suppose we could make wire of n-type silicon doped at 10^{15}/cm^2. What would the diameter of a 10 cm long piece of silicon wire have to be to have the same resistance as the same length of #24 copper wire?

8. For the resistor in Problem 7, what temperature range can this resistor be used if the design specification for the Si resistor is +/- 5% of the room temperature value? Ignore the change in mobility as a function of temperature.

9. Using Matlab, plot the bandgap of Si as a function of temperature from 0^{o}C to 1200^{o}C.
   a. At what temperature is the bandgap of Si equal to the bandgap of Ge at 300K?
   b. If you replaced the curve with a line between the temperatures of 0^{o}C to 100^{o}C, what is the slope of the line (the change in bandgap as a function of temperature)?

10. What is the junction capacitance (C_j) and built-in voltage (V_{bi}) of a Si diode that is doped

a. with $N_A = N_D = 1x10^{15}$ cm^{-3}?
b. with $N_A = N_D = 1x10^{15}$ cm^{-3}?

11. Suppose that you have a piece of Si that is doped with arsenic to a concentration of $N_D = 1x10^{15}$ cm^{-3}, how large a volume is needed to make sure that there is one Ar atom in it? Assume that the volume is a cube, what is the length of each side in nanometers?

12. A bar of semiconductor can be used as a propagation delay line. These are used on the inputs and outputs of various digital gates to minimize possible timing hazards that can arise when the signals arrive at a gate at slightly different times. For example, you might want to have the output from an OR gate ANDed with the output of an NAND gate. However, it takes 15ns for the output to appear on a 74LS32 OR gate and it takes 22ns for the output to appear on a 74LS00 NAND gate.

a. What length of an n-type Si bar should be connected between the OR and AND gate so that the total propagation time (the time it takes for the output to appear on the OR gate plus the transit time of the carriers in the Si bar) equal the time that it takes for the output to appear on the NAND gate. Assume that the electric field across the semiconductor bar is 200V/cm and the mobility of the majority carrier is 1500 cm^2/Vs. Please write the solution in μms.
b. What is the voltage required to generate the 200V/cm across the n-type silicon?

CHAPTER 2

DIODE CIRCUITS

As discussed in the previous section, diodes are essentially one-way valves. They carry current in one direction, but block current in the other. In addition, in the forward conduction mode, the voltage must go above a threshold before significant conduction occurs. Diodes are used in a multitude of ways that utilize these characteristics.

The objective of this course is not to look at hole and electron flows, but to see how the diodes can be used in circuits and how to analyze those circuits. For example, consider the circuit below, Figure 1. We are generally not interested in the holes and electrons in the diode, but rather, the current in the resistor or the voltage across the resistor.

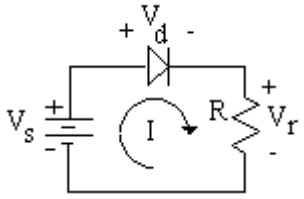


Figure 1. A simple diode circuit

NUMERICAL METHOD

The diode being a non-linear element means that the basic methods of circuit analysis learned in your circuits course cannot be used. Then, how does one go about analyzing the circuit? For example, we can write a loop equation

$$V_s = V_d + V_r$$

$$= V_d + I*R$$

Unfortunately, there are two unknowns in this equation, I and V_d. The relationship between voltage and current in the diode is given by

$$I = I_0(e^{(V_D/V_T)} - 1)$$

Where V_T is 25.9 mV at 300 degrees K, and I_O is the reverse saturation current. If we put the two equations together we get

$$V_S = V_d + RI_0(e^{(V_D/V_T)} - 1) \qquad (1)$$

This equation cannot be solved analytically. But it can be solved numerically. You can make a guess of the diode voltage and plug into the equation and see if you get a balance.

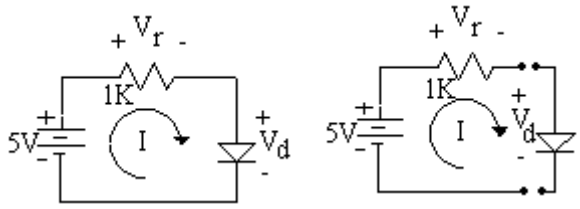
If not, try another guess. This process can be carried out several ways. The simplest is to use a calculator, while more complex methods would use a computer. One interesting approach is to use a math solving program on a personal computer such as MathCAD.

Literally all circuits could be solved this way, but that would be impractical, especially for more complex circuits. Secondly, the diode equation mentioned above only represents the hole-electron currents within the body of the diode. There is a considerable leakage on the surface of the semiconductor which makes the low current solutions in error.

GRAPHICAL METHOD

Instead of using the diode equation as the model, we can get quite satisfactory results by using simpler models, or by using graphical methods from the plotted V-I characteristic of the diode. Consider the circuit in Figure 2. The loop equation is

$$V_S = V_d + I*R \qquad (2)$$



a. b.

Figure 2. Diode circuit to be solved graphically

If we divide the circuit as shown in Figure 2b, the equation can be rewritten to solve for the voltage at the division.

$$V_d = 5 - I*R \qquad (3)$$

This equation can be solved for current

$$I = \frac{5 - V_d}{R} \qquad (4)$$

This equation represents the relationship between current and terminal voltage for the left-hand side of the circuit. The relationship between current and terminal voltage for the right-hand side is just the equation for diode.

$$I = I_0(e^{(V_D/V_T)} - 1) \qquad (5)$$

Each of these equations is just an I vs V_d equation and can be plotted on a graph. Equation 4 is a linear equation and its plot is called the load line. It can easily be plotted by selecting values for V_d and calculating I. For example, the two intercepts; $V_d = 0$ and I = 0 are convenient choices. The equation 5 is the diode I vs V characteristic for the diode. It is usually obtained from a curve tracer in the laboratory. Both of these equations are plotted on the same graph in Figure 3. The operating point for the circuit is where the two plots cross.

$$\frac{5 - V_d}{R} = I_0(e^{(V_D / V_T)} - 1) \qquad (6)$$

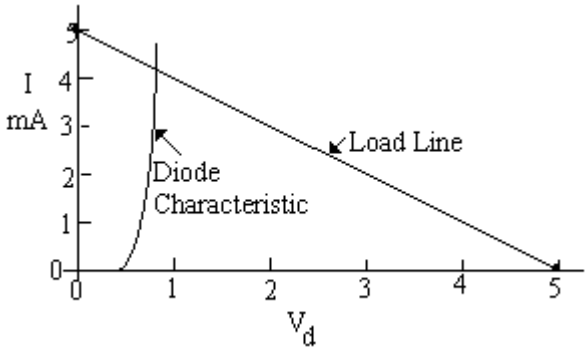


Figure 3. Plots for graphical solution

One problem with the graphical solution above is that the voltage scale is so large that precision in determining diode voltage is difficult. In many cases, the diode characteristic is plotted with a much lower maximum voltage than the voltage source in the circuit. In this case, the voltage intercept is off the scale and it is more difficult to plot the load line. This situation is easily handled by substituting in a fixed value of voltage in the load line equation and solving for the current. An example is shown in Figure 4. The maximum voltage on the diode characteristic plot is 2 Volts with the source voltage 5 Volts and a resistance of 1KΩ. Solving the load equation

$$I = \frac{5 - 2}{R} = 3mA$$

which gives us the current at $V_d = 2$ volts at the right hand end of the load line on the plot in Fig. 4.

Figure 4. Diode Characteristic and Graphical Solution

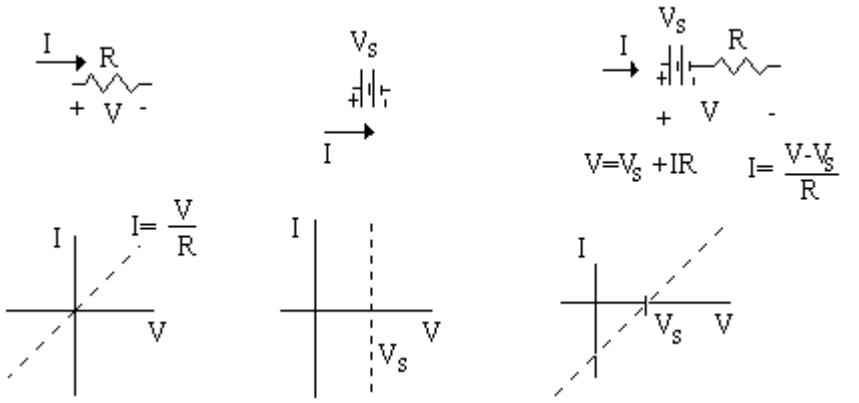
CIRCUIT MODELS

The graphical method presented in the previous section is one possible method of solution of circuits with diodes. For obvious reasons, this method will get very tedious and time consuming for more complex circuits. What we would like to do is to find a way to solve the circuits analytically; with equations. We do this by using circuit models which are combinations of ideal circuit elements that behave the same or almost the same as the actual device. We then solve the circuit using standard methods of circuit analysis.

A model is what we use to represent an actual device. For example, we use a mathematical expression V=IR to represent the voltage across a resistor. This happens to be a very good model at low frequencies, below a few MHz. At higher frequencies, the inductive reactance becomes significant and the equation is no longer very accurate. Similarly, $V = j\omega LI$ and $I = j\omega CV$ are good models for inductors and capacitors at relatively low frequencies.

We also use models for voltage and current sources. For example, a typical voltage source is far from ideal. It usually has internal resistance and cannot supply unlimited current.

Now let us take a look at what these ideal circuit models represent. If we draw a plot of the relationship between current and voltage of a resistor, we get the plot in Figure 5a. Similarly if we plot the I-V characteristic of the ideal voltage source, we get the plot in Figure 5b. The I-V characteristic of the voltage source in series with the resistor is shown in Figure 5c.



a. Resistor *b. Voltage source* *c. Series combination*

Figure 5. Models for a Resistor and a Voltage Source

Circuit Models for Diodes.

Now let's look at possible circuit models we can use for a diode. Clearly, a single straight line would not make a very good model. However, a crude approximation can be developed using two straight lines as shown in Figure 6a, a model of an ideal diode. The two straight lines are used only for a specific region as shown. The ideal diode is considered an open circuit if the voltage across it is less than zero, and is a short circuit if the forward current is greater than zero. These two regions are called the cutoff and

Figure 6. Two Possible I-V Characteristic Models of a Diode

conducting regions or the reverse biased and forward biased regions respectively. We will carry this nomenclature forward with us through more advanced models. The semi-ideal model in Figure 6b is a better model of the actual diode curve and is used

extensively. In this model I = 0 for $V < V_\gamma$, and $V = V_\gamma$ for I>0, where V_γ is the threshold voltage.

Another way to look at these models is to translate them into circuit elements. In fact, this is the way we will use these models for conceptualizing circuit operation. The circuit elements representing these models are shown in Figure 7.

Figure 7. Circuit elements representing the models in Figure 6

SOLVING DIODE CIRCUITS USING DIODE CIRCUIT MODELS

As a first example, let us try the solution of a simple rectifier circuit shown in Figure 8. We will do this first with an ideal diode model and then with the semi-ideal model. Solution for this circuit may mean finding the current or the voltage across the resistor or diode.

Figure 8. Diode Rectifier Circuit

Solution Using The Ideal Diode Model

When attempting to solve this circuit assuming an ideal diode, the first thing you notice is that the diode has two regions of operation. How do you know which to use? The answer to that question comes more from experience than anything else. However, some quick mental analysis is often useful at arriving at a good starting point. If all else fails, simply guess and then check to see if the resulting solution is consistent with your original guess or assumption.

For this example, a reasonable guess is that the diode is forward biased and is therefore conducting. In this region, for the ideal diode, V=0 and I>0. Putting this model into the circuit, as shown below, the resulting current is

$$I = V/R = 2.5 \text{ mA}$$

Since I>0, our solution is consistent with the original assumption.

Figure 9. Rectifier Circuit With Diode Replaced by Ideal Model (Conducting)

Solution Using The Semi-Ideal Model

Now let us use the semi-ideal diode model and solve the rectifier circuit again. The model equations are:

V=0.70 for I>0 and

I=0 for V<0.70.

Again, let us assume that the diode is conducting, I>0. The circuit is redrawn below in Figure 10 with the diode replaced by the assumed model. Solving this circuit,

I= (5.0 - 0.70)/2K = 2.15 mA.

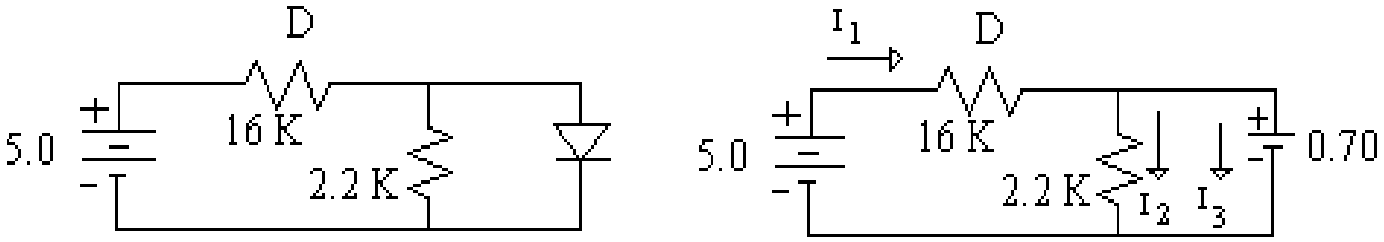
Again, since our solution is consistent with the circuit model, we can assume the solution is correct.

Figure 10. Rectifier Circuit With Diode Replaced by Semi-Ideal Model (Conducting)

We note that these two solutions are significantly different because we used significantly different models for the diode. The difference would have been smaller if the source voltage had been very large compared to V_γ.

Example 2, A More Complex Circuit

As a second example, let us look at the circuit below. We will use the semi-ideal diode model for this problem. At first glance it is not clear whether the diode is conducting or not. Thus, we are not given any guidance in selecting the appropriate model. Let us guess that the diode is conducting. In that case, the circuit is redrawn with the diode replaced by its model in the conducting region.



a. Original Circuit *b. Circuit With Conducting Model*

Figure 11. Diode Circuit for Example 2

We can solve this circuit by simply determining the currents in the branches.

I_1 = (5.0-.70)/16K = .269mA

I_2 = .70/2.2k .318mA

I_3 (The diode current) = .269-.318 = -.049 mA

The result implies that the current in the diode is in the reverse direction which is inconsistent with the diode model chosen. Our original guess was wrong.

Let's try this problem again with the model for the non-conducting or cutoff region. The circuit is shown below.

Figure 12. Circuit of Example 2 with Non-Conducting Model

Solving this circuit, we find that the voltage across the diode is:

V_d = 5.0x 2.2k/(2.2K + 16K) = 0.604 V (<0.70 V)

This result is consistent with our model and we can assume it is the correct result. What would have happened if we had chosen the non-conducting model first? We might not have tried the other model. This raises the question of whether both models will give answers that are consistent with the assumption. Actually, this can happen only in cases where there is feedback and the loop gain is greater than unity which can only happen with an amplifier in the circuit. In these cases, there are two stable states. A typical application is a flip-flop. These cases are easy to recognize.

As an exercise, modify the previous example and determine the value of R2 (originally the 2.2k resistor) that will ensure that the diode is conducting at 0.01mA.

Example 3, A Clipping Circuit

As a more advanced example, let us look at a clipping circuit shown in Figure 13. This is not a steady state case as before, but uses a time varying input signal. Let us assume the input signal is a sine wave. What is the output signal?

Figure 13. Example 3, Diode Clipping Circuit

We can do a simple calculation as we did before and discover that at the most negative peak, the diode is cutoff, and at the most positive peak, the diode is conducting.

At Vin = -5, Vo = -5 Volts.
At Vin = +5, Vo = 1.0 + 0.70 = 1.70 Volts

Obviously, some place in between, the diode switches states. One of the tricks is to find that transition point.

Let us start with the case of the diode non-conducting as shown below. It is easy to see that Vout = Vin as long as the diode is an open circuit. The diode will remain in this state as long as the diode voltage is less than 0.70 volts. As long as Vout (and Vin) is less than 1.70 Volts, the diode will be off. Thus

Vout = Vin for Vin < 1.70

Figure 14. Clipping Circuit With Diode Non-Conducting

For Vin > 1.70, The diode will be conducting. The circuit is shown below. For this case,

Vout = 1.70 for Vin > 1.70.

Figure 15. Clipping Circuit With Diode Conducting

If we combine the two regions of operation, and draw the waveforms generated by this circuit, we will get the results shown in Figure 16.

Actually, the diode begins to conduct when the forward voltage is in the neighborhood of 0.60 volts and fully turns on at about 0.70 volts. In between is some sort of smooth transition. The sharp transitions of the waveforms in Figure 16 would actually be rounded. For the purposes of this course, this transition is not very important.

Figure 16. Output Waveform of the Diode Clipping Circuit With A Sinusoidal Input Signal

Advanced Models

The models in Figure 6 are obviously only crude approximations to the actual curve. A more advanced model will take into consideration that the diode current is not zero below the knee of the curve. Looking just at that portion of the curve, it looks almost resistive, with a rather high resistance. In fact, this region is often modeled that way with a resistor on the order of 1 megohm. The fact of the matter is that a significant portion of the current in this region is due to surface leakage, (resistive). For the conducting portion of the curve, the curve is nearly vertical, but not quite. Again, this suggests another resistance, but in series with a voltage source. This resistance is mainly due to the body resistance of the bulk material. The result is a two-region linear model that is piece-wise continuous. The models and mathematical expressions are shown in Figure 17.

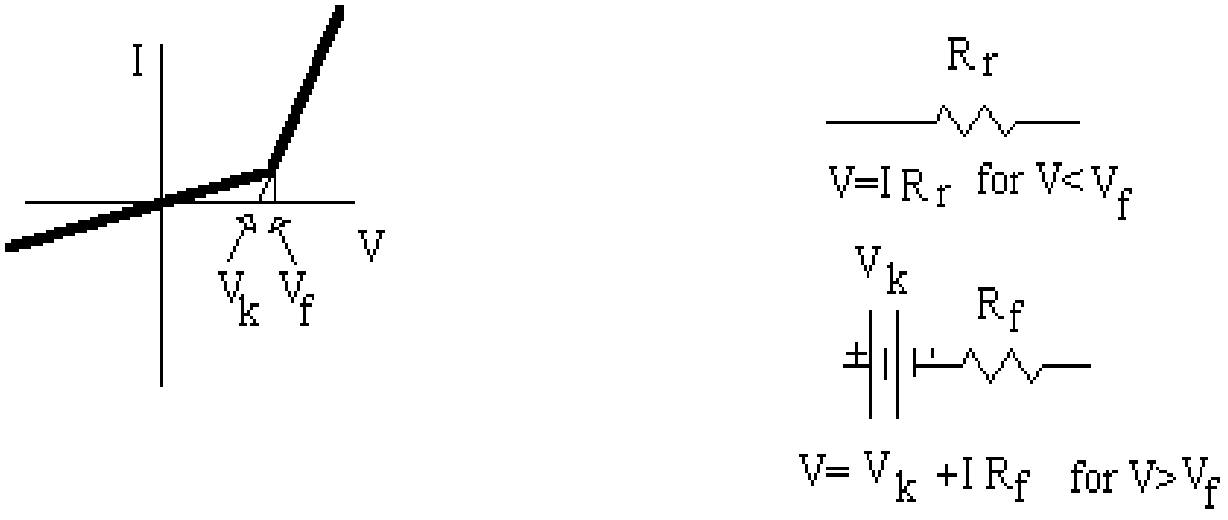


Figure 17. Two-Resistor Model of a Diode

The circuit model shown in Figure 17 is a much better representation than the earlier ones, but even this one differs greatly from the actual characteristic in the region of the knee. More sophisticated models would use three or more regions for the model.

Choosing an Appropriate Model

The model for the ideal diode is useful for many purposes that do not require great precision in circuit solution. The two-resistor model is rather complex and is sometimes used where relatively accurate analysis is required. However, this model is more difficult to use than the other two. As a compromise, we will use the semi-ideal model for most of the analysis in this course. It will give a reasonably good solution to digital switching circuits used while not adding unnecessarily to the circuit complexity.

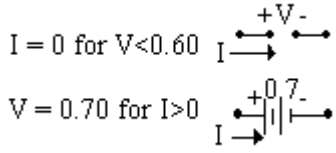
For silicon diodes, the value of V_γ is usually taken between 0.60 and 0.70 volts. There are physical justifications for the threshold voltage assumed, but the value used is often selected from observations of the typical voltage measured across the diode with normal operating current. For integrated circuit diodes, the cross sectional area of the diode body is kept rather small and the current density is often rather high. Thus, for integrated circuits, 0.70 volts is often used.

For discrete diodes, the current densities are often somewhat lower, allowing a lower value, usually 0.65 volts. In both integrated circuits and low-power discrete circuits, any reverse voltage applied during operation is quite low, allowing the reverse current to be neglected.

In power diodes, the current densities are often very high, approaching the upper device limit, and the reverse voltages also often approach the device limits. In these cases, the body resistance during forward conduction and the reverse leakage current during reverse biasing cannot be neglected. Thus, in this case, the two-resistor model is more appropriate and is often used.

Germanium diodes are usually used only for low power applications, and usually used as discrete devices as germanium is rarely used in IC's. In practice, V_γ is usually taken between 0.2 and 0.3 volts.

We will use a slight modification to the semi-ideal model for most of this course where we are talking about switching circuits. The model we will actually use is:



We will not operate where 0.60<V<0.70

This model makes some sense in that a close look at a typical diode curve reveals that the current is usually negligible with the voltage below 0.60 V. Once the current reaches an appreciable amount, the voltage across the diode has risen to about 0.70 V. In between, the I-V characteristic is a smooth curve. In digital switching circuits we find the diode will either be in the cutoff region or the forward conducting region with substantial current. We simply will not operate in the region in between.

DESIGN EXERCISE (Suggested by Dr. W. L. Cooley)

This design exercise is an example of an application that illustrates the use of diodes in circuits. This circuit utilizes the offset voltage of the diode. It assumes a semi-ideal characteristic: I=0 for $V<V_\gamma$ and $V= V_\gamma$ for I>0. In this case, $V_\gamma = 0.7$ volts.

The application is to make a non-linear voltmeter that will measure the available capacity of a 10 volt Ni-Cad battery. The Ni-Cad battery voltage at full charge is 10 volts. As the battery is discharged, the terminal voltage drops slowly down to about 9 volts. Below 9 volts, the battery capacity is nearly used up and further discharge will cause the terminal voltage to drop rapidly.

What we would like to have is a meter that will show full scale when the battery is 10 volts, and dropping linearly to 10% of full scale when the battery voltage drops to 9 volts. Below 9 volts, the meter reading drops linearly to zero at zero terminal voltage. Above 10 volts input, the meter reading is to be limited to full scale. This characteristic is shown graphically in Figure 18 below.

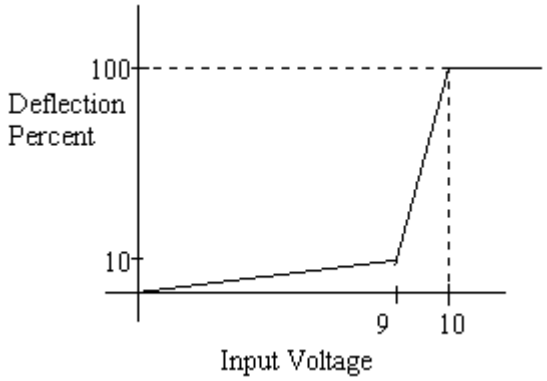


Figure 18. Graphical representation of desired meter characteristic.

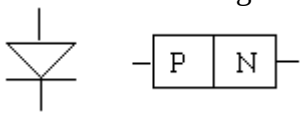
The meter to be used has 1Kohm resistance and a full scale deflection of 1 microamp.

Solution:

An example solution is given in the EE 56 Web site under Additional Information.

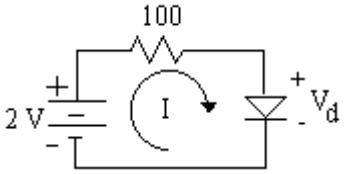
EXERCISES

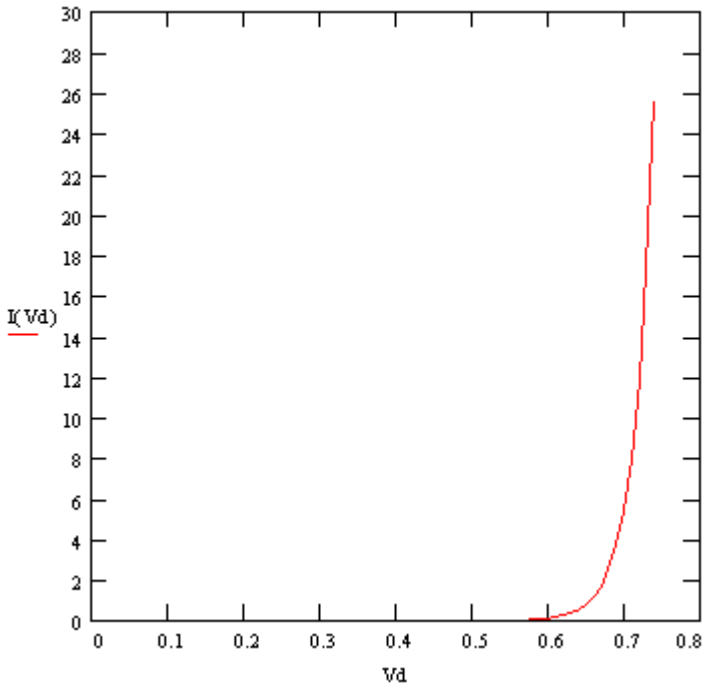
1. The symbol for a diode is shown along with a pn junction.
   a. Draw the polarity that will cause forward current in both devices.
   b. Draw an arrow showing the direction of forward current for both devices



2. Find the current in the following circuit. The diode I-V characteristic is given below.

I = _____





3. Solve again, if the voltage source is 3 volts. I = _____

4. Solve again, if the voltage source is 3 volts and the resistor is changed to 150 ohms.

I = _____

5. Solve Exercise #2 analytically using an ideal diode model.

I = _____

6. Solve Exercise #2 analytically, using the following model for the diode.
   $I_F = 0$ if $V_D < 0.60$ volts
   $V_D = 0.70$ volts, if $I_F > 0$

   I = _____

Problems

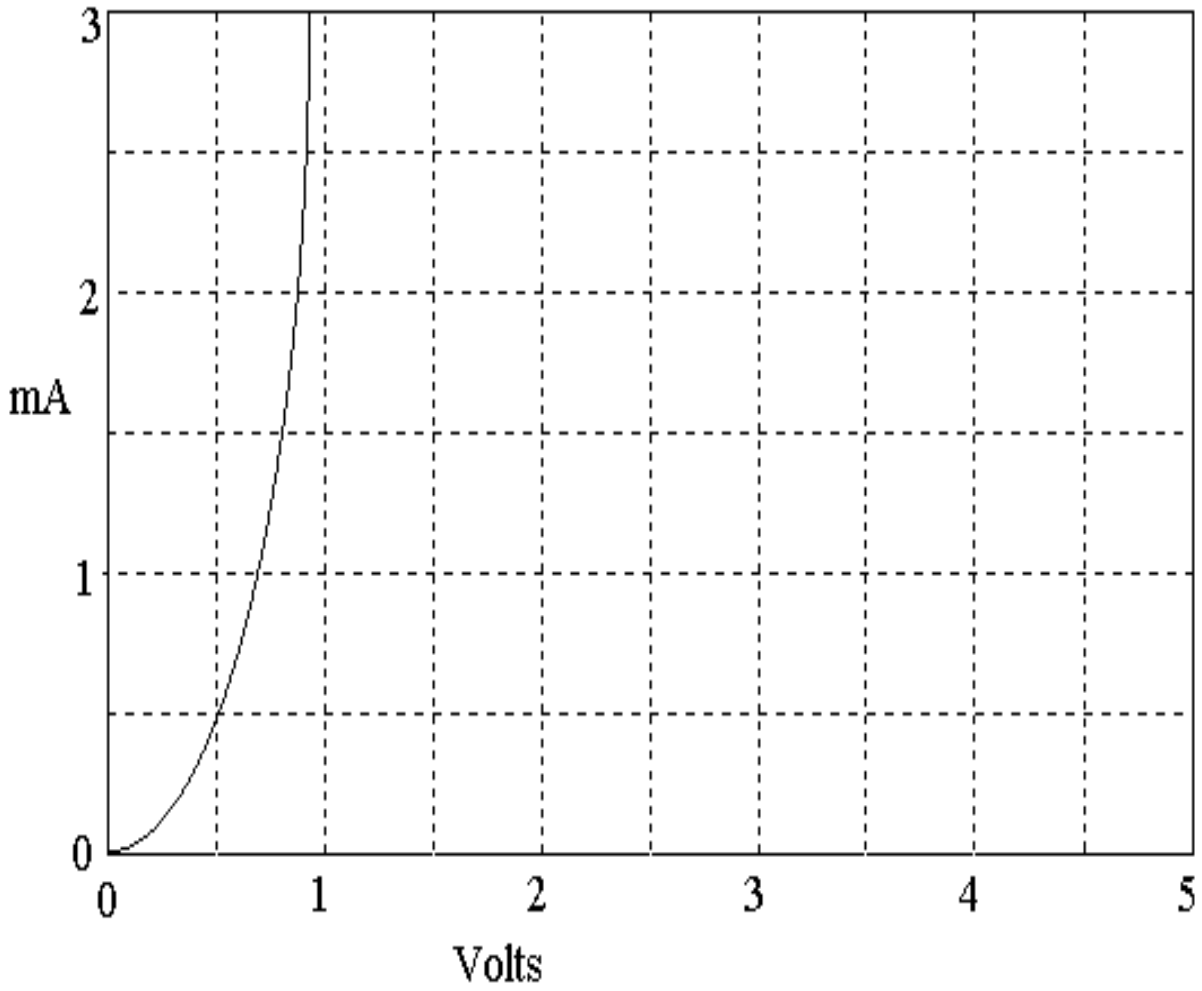
1. A diode has a reverse saturation current of 10 pA and V_T = 25.9 mV. What is the applied voltage to get a forward current of 100 mA?

2. A silicon diode was found to have a reverse saturation current of .001 microamp. What would the current be if a forward voltage of 0.50 volts was applied?

3. Use numerical methods to solve for the current in the circuit and the voltage across the 1NWV56 diode. Continue the iterative process until the error in the diode voltage is less than 0.1 mV. The reverse saturation current of the diode is 0.01 nA. The temperature is 27°C so that V_T = 25.9 mV. Use the equation below for the diode current.

$$I = I_0(e^{V_D/V_T}-1)$$

4. Solve the circuit above using graphical techniques. Assume the 1NWV56 diode has the characteristic given on the following page. Find the current in the circuit and the voltage across the diode.



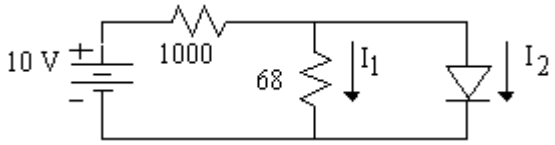
5. Repeat Problem 4, but with the voltage source at 10 volts and the resistor at 5 KΩ.

6. Solve the circuit below for currents I_1 and I_2. Demonstrate that your answer is consistent with the assumptions you made.

Diode Model:

Forward Conducting Region
V=0.70 volts for I>0

Non-Conducting Region
I=0 for V<0.70 volts



Circuit for Problems 6 and 7

7. Work Problem #6 again, but with a source voltage of 12 volts.

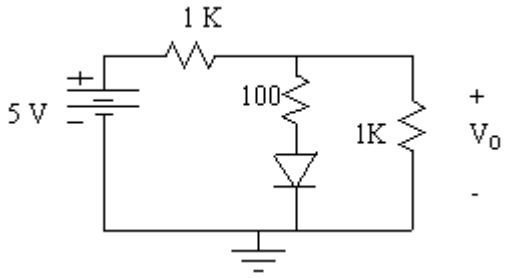
8. For the circuit given, draw the current waveform and the waveform for the voltages across the resistor and the diode. Use the diode model given in Problem #6. Carefully determine the breakpoints (abrupt changes in slope). Write the equations for each segment of each waveform showing the range of values of V_{in} for which the equations are valid.

9. Do a computer simulation of problem 8. Use the circuit models from problem 6. Show the waveforms.

10. Do a computer simulation for problem 8, but this time use the built-in model for a silicon diode. Show the waveforms.

11. Determine the output voltage for the circuit shown. The diode model is:

$V_D = 0.700$ for $I>0$ $\quad I = 0$ for $V_D < 0.700$



12. A device found in the Martian ruins has a model described by the following equations.

Equations:

i. $V_x = 2.5$ V for $I > 25$ mA

ii. $V_x = 100\ I$ for $-5.0 < V_x < 2.5$ V

iii. $I = -5.0$ mA for $V_x < -5$ V

a. Draw a circuit model for each region of operation.

b. Draw the I-V characteristic for the device

CHAPTER 8
MOSFET

Field effect transistors can be thought of as voltage controlled resistors for many applications. The physical construction of a typical enhancement MOSFET is shown in Figure 1. This figure represents an n-channel, enhancement mode device.

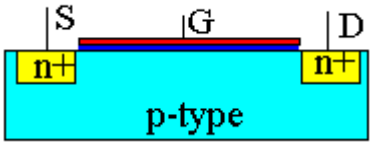


Figure 1. Construction of a MOSFET. (n-channel, enhancement),

This device is built on a substrate of p-type silicon. Two heavily doped n-type tubs are diffused in for the source (S) and the drain (D). The channel region is covered with a silicon-dioxide, SiO_2, insulating layer. A metal layer is deposited on top of the insulating layer to act as a gate (G). If we simply apply a voltage between the source and drain, one of the p-n junctions at either the source or drain would be reverse biased and the device would not conduct. The gate becomes the control element as discussed next.

A typical common-source circuit connection is shown in Figure 2. In this case, both the drain and the gate are made more positive than the source, which is grounded.

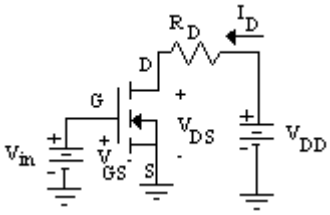


Figure 2. Typical biasing connection.

Ignoring the drain supply voltage, V_{DD}, for a moment, let us consider the gate supply only as shown in Figure 3. If the gate voltage is made positive enough, a number of electrons from within the p-type substrate will be attracted to the gate area. Because the gate is insulated from the substrate, these electrons will simply accumulate just below the gate, effectively changing that area to an n-type material; hence an n-channel.

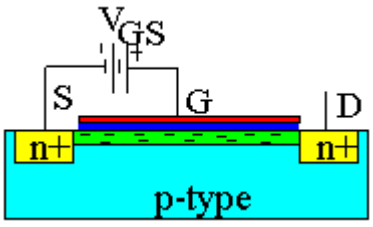


Figure 3. The MOSFET with a positive gate voltage.,

As shown in Figure 3, this locally enhanced n-type region between the source and drain allows conduction between the source and drain terminals. Note that if the gate voltage is made even more positive, more electrons accumulate, the electron density increases and the cross-sectional area increases, lowering the channel resistance. Also, there is a threshold voltage, V_T, below which not enough electrons will be attracted to turn the channel region into n-type material, and the channel will not conduct. This threshold voltage is a function of doping levels and thickness of the oxide layer.

With an insufficient voltage on the gate to establish the channel region as n-type, there can be no conduction between the source and drain. But when the gate voltage is above the threshold, the channel is established. Then, if we apply a voltage between the drain and the source,V_{DS}, we will get a current to flow. As long as we keep the drain-source voltage, low, the MOSFET acts just like a voltage controlled resistor. The resistance of the channel is controlled by the gate voltage. This region of operation is known as the *triode* region.,

Next, let's examine the voltage between the gate and the channel at three points along the channel as shown in Figure 4.

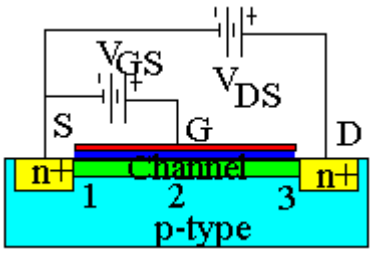


Figure 4. MOSFET under operation in the triode region.(V_{DS} small),

The voltage difference between the gate and the channel at point 1 is essentially the gate-source voltage, $V_{G1} = V_{GS}$. Likewise, the voltage between the gate and the channel at point 3 is the voltage between the gate and drain, and $V_{G3} = V_{GS} - V_{DS} = V_{GD}$. Note that $V_{GS} > V_{GD}$. Because the voltage between the gate and the channel at the drain end is less than the voltage between the gate and the channel at the source end, the channel is thinner and there is lower electron density in the channel and lower conductivity at the drain end. The voltage between the gate and middle of the channel is between the other two, $V_{G1} > V_{G2} > V_{G3}$.

Because the voltage between the gate and various points along the channel is not constant, the channel will not be a constant width. Figure 5 shows the channel in this condition. The channel-thinning effect becomes more pronounced as V_{DS} is increased. In fact, if we increase the drain voltage high enough, the voltage between the gate and the channel near the drain will fall below the threshold voltage V_T. One might think the current would stop, but instead the current simply remains constant as V_{DS} increases further. This effect is called PINCH-OFF. If the current were actually cutoff, the channel would be isolated from the drain, and the gate-channel voltage would increase back up to

the gate-source voltage, and conduction would start again. Pinchoff is simply an equilibrium point with constant current for a fixed gate-source voltage. Figure 6 shows the channel in pinchoff. Pinchoff begins when $V_{GD} < V_T$.

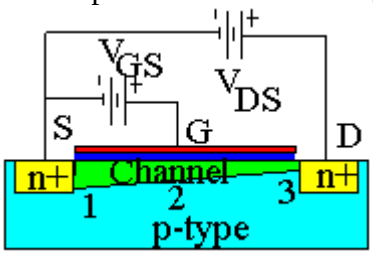


Figure 5. Channel in Triode

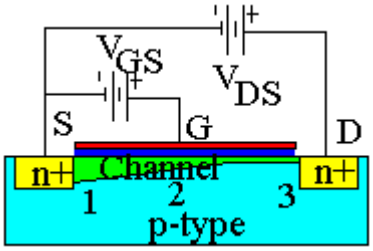


Figure 6. Channel in Pinchoff

The drain characteristic for a ZVN-3310, an enhancement-mode MOSFET is shown in Figure 7. The drain characteristic is a plot of drain current versus drain-source voltage at various gate-source voltages. If we held V_{GS} constant, we would get a single curve, but an infinite number are possible, so we select an appropriate set to represent the device. For any particular curve, it starts at $I_D = 0$ and $V_{DS} = 0$. As V_{DS} increases, the current increases almost linearly in the triode region. As V_{DS} increases further, channel thinning occurs and the curve does not rise as rapidly. The rate of rise decreases until there is constant current in pinchoff.

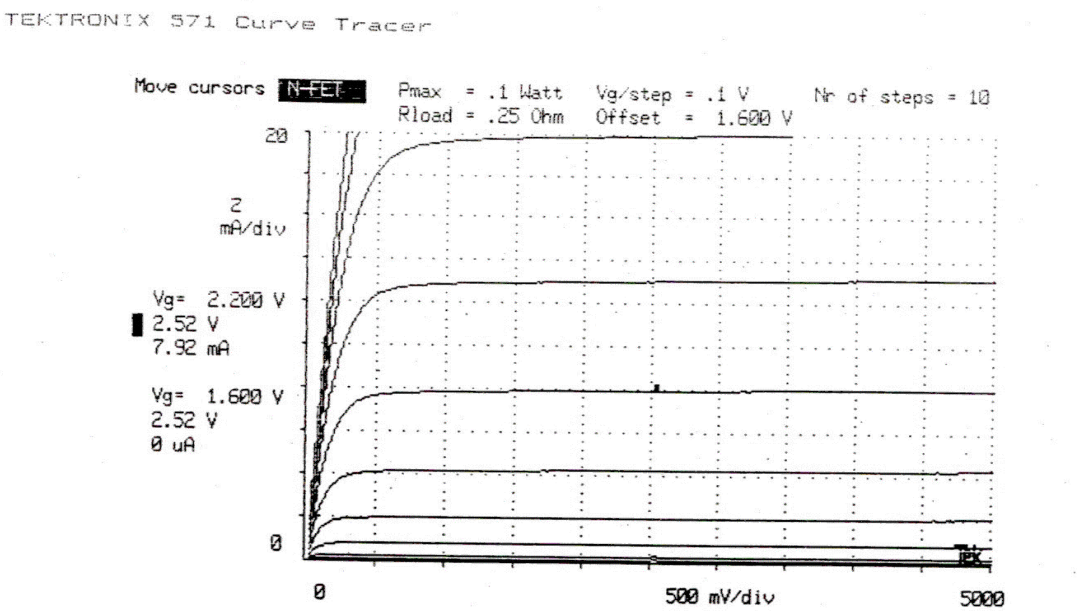


Figure 7. Drain Characteristic for a ZVN-3310, n-channel MOSFET`

MODELING EQUATIONS FOR THE MOSFET

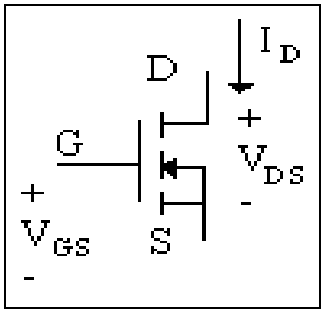


Figure 8. N-channel, enhancement MOSFET

There are three distinct regions of operation: cutoff where the gate voltage is below threshold and no current flows, triode where the current is nearly a linear function of drain voltage, and pinchoff where the current is constant for fixed gate voltage. Figure 8 shows the voltage polarity and current direction definitions. Positive drain current enters the drain and leaves by the source. Gate current is zero because it is insulated from the channel. The order of the subscripts In V_{GS} and V_{DS} indicate a voltage drop from gate-source and drain-source respectively.

CUTOFF REGION ($V_{GS} < V_T$)
In this case, The gate voltage is not high enough to create a conducting channel so there is no conduction. This case can be represented by a line on the x-axis of the characteristic curve with $I_D = 0$.

TRIODE REGION ($V_{GS} - V_T > V_{DS}$) [Could also be written $V_{GD} > V_T$],

As long as $V_{GS} - V_T > V_{DS}$, the voltage between the gate and every point in the channel is greater than the threshold voltage. Under these conditions, the approximate drain current is

$$I_D = k(2(V_{GS}-V_T)V_{DS}-V_{DS}^2) \qquad \text{Equation 1}$$

Where k is given by:

$$k = \frac{\mu\varepsilon W}{2t\mathrm{L}}$$

and:

μ= carrier mobility,
ε= dielectric constant,
W = channel width,
t = dielectric thickness,
L= channel length.,

If Equation 1 is compared to the characteristic curves, you will note that Eq. 1 is the equation for a parabola, as seen in the curves where the curves leave the origin. At $V_{GS} - V_T = V_{DS}$, Eq. 1 reaches its maximum value of $I_{dmax} = kV_{DS}^2$ and has a slope of zero. You will also note that, at that voltage, the voltage between the gate and the channel at the drain end is just equal to V_T; in other words, the threshold of pinchoff
PINCHOFF ($V_{GS} - V_T < V_{DS}$)

In this region, the voltage between the gate and some parts of the channel is less than the threshold voltage. The current will remain constant as V_{DS} increases.

$$I_D = k(V_{GS} - V_T)^2 \qquad \text{Equation 2}$$

Note that this current is just the same as the point where $V_{GS} - V_T = V_{DS}$ from Eq. 1.

Similar to the circuit models for all electronic devices, the model equations are approximations to the actual curves. For example, Equations 1 and 2 are plotted in Figure 9 for the case of V_{GS} = 4 V, V_T =2 V, and k=100 $\mu A/V^2$. Equation 1 is a parabola rising from the origin and curving back down. Equation 2 is simply a straight line with fixed current and not a function of the drain voltage. The two curves intersect at the peak of the parabola, the transition between triode and pinchoff operation. The parabola is valid only up to the intersection, and the straight line is valid only after the intersection, resulting in a composite similar to the actual characteristic curves in Figure 8.

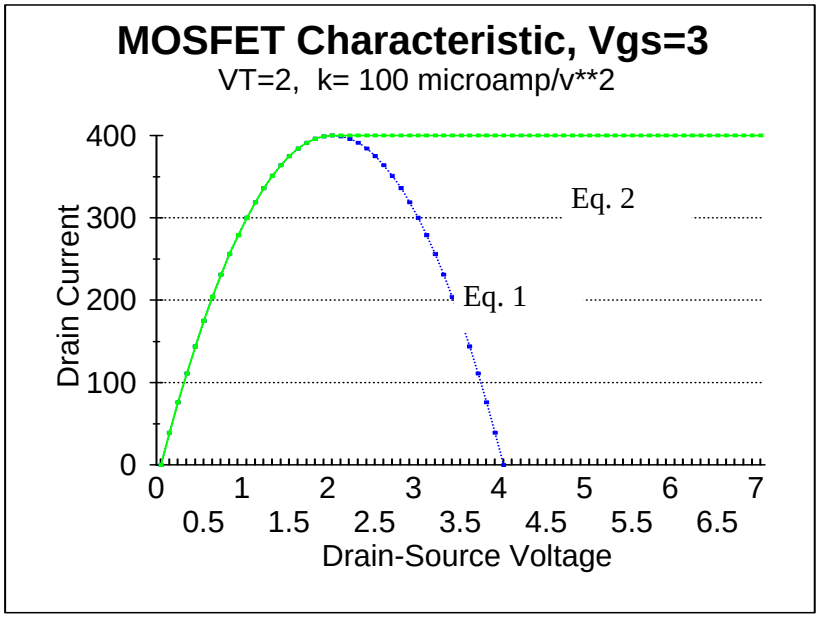


Figure 9. MOSFET Curve Using Modeling Equations.,

Figure 10 is a characteristic curve of an n-channel MOSFET using the models developed above. Also plotted on the graph is the boundary between the triode and pinchoff regions where $V_{GS} - V_T = V_{DS}$. On this graph, V_T = 2 V. This calculated characteristic curve is very much like actual characteristic curves taken on a curve tracer.

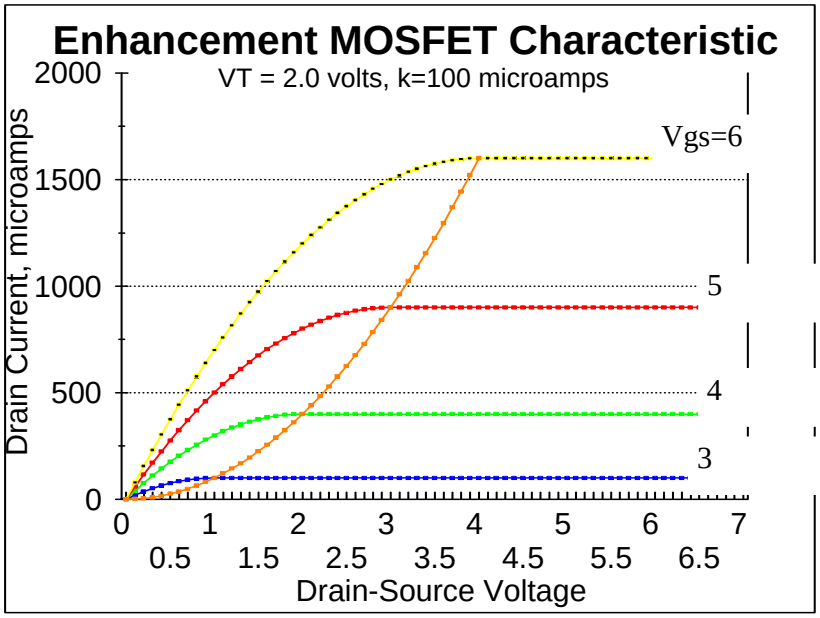


Figure 10. Drain characteristic from model equations for an n-channel, enhancement mode MOSFET with V_T = 2.0 volts and k = 100 $\mu A/V^2$.

CIRCUIT MODELS FOR THE MOSFET

We have described three distinct regions of operation: cutoff, pinchoff, and triode. Each region can be represented by the equations and used to solve circuits as shown in following sections, or each region can be represented by circuit elements and circuits solved using standard circuit analysis techniques.

Cutoff ($V_{GS} < V_T$), $I_D = 0$
In this case we can represent the MOSFET as shown in Figure 11. There is no current between the drain and source as the channel is open. Thus, we represent the channel as an open circuit. We also represent the gate terminal as an open circuit recognizing that the gate is insulated from both other terminals by an insulating oxide layer.

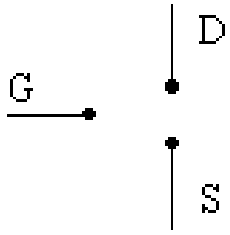


Figure 11. MOSFET Cutoff circuit model.

Pinchoff ($V_{GS} > V_T$ and $V_{GS}-V_T < V_{DS}$)

$$I_D = k(V_{GS} - V_T)^2$$

In this case, current is only a function of gate voltage and thus, the current in the channel can be represented by a dependent current source as shown in Figure 12.

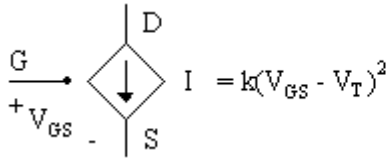


Figure 12. MOSFET Pinchoff model

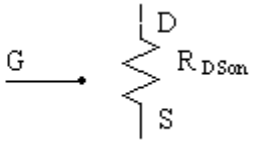


Figure 13. MOSFET Triode model

Triode ($V_{GS} > V_T$ and $V_{GS}-V_T > V_{DS}$)

$$I_D = k(2(V_{GS} - V_T)V_{DS} - V_{DS}^2)$$

If we assume the V_{DS}^2 term can be ignored, or $2(V_{GS} - V_T) >> V_{DS}$, then

$$I_D \approx 2k(V_{GS} - V_T)V_{DS}$$

and the term $2k(V_{GS} - V_T)$ has units of Siemens or Mhos, then the inverse is a resistance defined as R_{DSon}. In other words, we can represent the drain-source channel as a resistor as shown in Figure 13.

$$R_{DSon} = \frac{1}{2k(V_{GS} - V_T)}$$

SOLVING MOSFET CIRCUITS

To illustrate the use of models in solving MOSFET circuits, consider the circuit shown in Figure 14. In this circuit, V_{DD} = 6 volts and we will solve the circuit for input voltages of V_{in} = 0,3, and 6 volts. Solution of these circuits almost always involves finding the output voltage, which in this case is $V_O = V_{DS}$, the voltage between the drain and ground, or the current in the drain, I_D. This MOSFET is the same one whose calculated characteristic is shown in Figure 10. First, we will solve the circuit using the circuit models and then using the model equations.

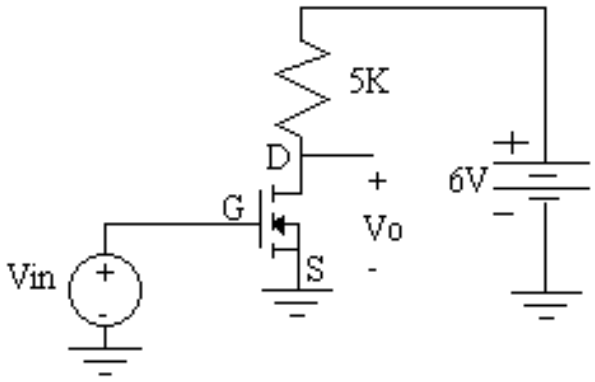


Figure 14. MOSFET inverter circuit. V_T = 2 volts and k = 100 microA/V^2.

SOLUTION USING CIRCUIT MODELS

V_{in} = 0 volts

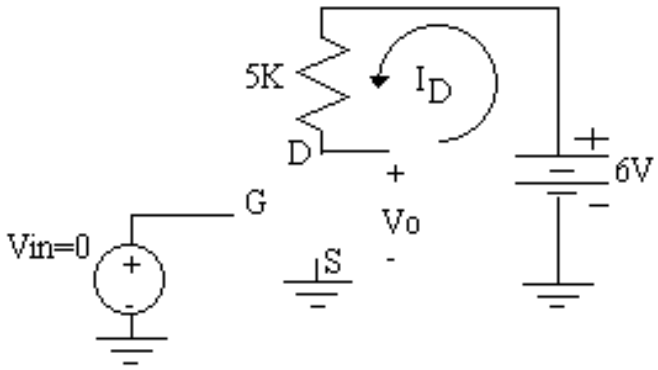


Figure 15. MOSFET inverter with cutoff model inserted.

In this case, $V_{in} = V_{GS} = 0$ volts. Because V_{GS} is below the threshold voltage, V_T, we will replace the MOSFET with its cutoff model as shown in Figure 15. From this circuit we see that $I_D = 0$ and because there will be no voltage drop across the 5K resistor, $V_O = V_{DD} = 6$ volts.

This solution is consistent with the cutoff assumption: $I_D = 0$ and $V_{GS} < V_T$.

$V_{in} = 3$ volts.

The input voltage is clearly above the threshold, so the MOSFET must be conducting. Let us assume the device is in the pinchoff region. Then, we can replace the MOSFET by its pinchoff model as shown in Figure 16. The current in the dependent current source is

$$I_D = k(V_{GS} - V_T)^2 = 0.1\text{mA}/V^2(3-2)^2 = 0.1\text{ mA}$$

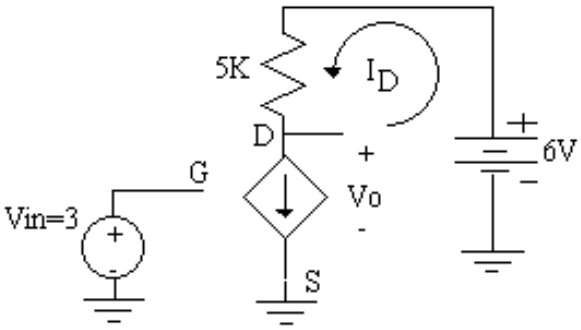


Figure 16. MOSFET inverter in pinchoff.

The output voltage is

$$V_O = 6 - 0.1\text{mA} * 5\text{ kOhms} = 6 - 0.5 = 5.5\text{ volts}$$

Now we must check to make sure we really are in pinchoff. $(V_{GS} - V_T) = 1$ volts which is less than $V_{DS} = V_O = 5.5$ volts. Therefore, we have met the conditions of pinchoff and our solution is valid.

$V_{in} = 6$ volts

Because we are now driving the gate to a higher voltage, let us assume the MOSFET is in triode. Then we can use the R_{DSon} model for the device in the circuit as shown in Figure 17.

$$R_{DSon} = \frac{1}{2k(V_{GS} - V_T)} = 1.25KOhm$$

Solving the circuit then,

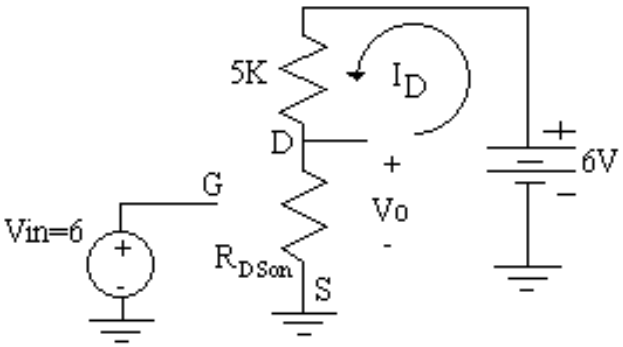


Figure 17. MOSFET inverter in triode

$$I_D = \frac{6}{5K + 1.25K} = 0.96mA$$

and

$V_O = 6 - I_D * 5K = 6\text{-}4.8 = 1.20$ volts

Now checking to see that we are truly in the triode region, $(V_{GS} - V_T) = 4$ volts which is greater than $V_{DS} = V_O = 1.20$ volts.

What if we had guessed the wrong region of operation? The cutoff region is easy because if $V_{in} < V_T$, then $(V_{GS} - V_T)$ is negative. But for instance we had assumed triode with $V_{in} = 3$ volts, then the circuit would be the same as in Figure 17 but with $R_{DSon} = 5$ K Ohms, and $V_O = 3.0$ volts which is greater than $(V_{GS} - V_T) = 1$ volt. In other words, our result would have been inconsistent with the assumption.

In the third case with $V_{in} = 6$ volts, If we had assume pinchoff, $I_D = 1.6$ mA and $V_o = -2.0$ volts, clearly an illegal result. (The dependent current source model is only valid for cases with $V_{DS} > (V_{GS} - V_T$.)

SOLUTION OF THE MOSFET INVERTER USING MODEL EQUATIONS

We can also solve the circuit using the model equations for the three regions. Using the model equations, we don't redraw the circuit for each case, but instead, we write an expression for Kirchoff's voltage law around the drain circuit in Figure 18.

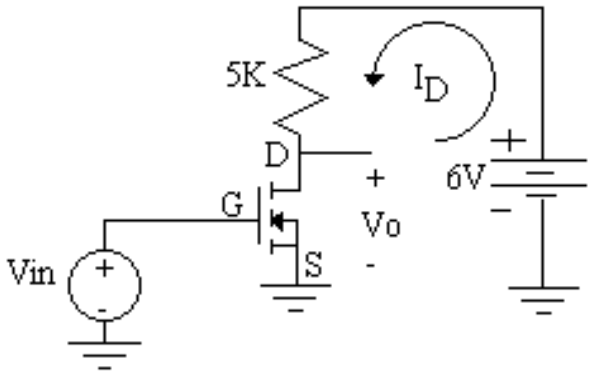


Figure 18. MOSFET inverter circuit

$6 = 5K * I_D + V_{DS}$

This equation has two unknowns, V_{DS} and I_D. We use the model equations to represent the relationship between the two. We will solve the circuit again for the same three cases using the model equations.

$V_{in} = 0$

This case solves the same way as with circuit models. Since we are in cutoff, $I_D = 0$, thus, $V_{DS} = V_O = 6$ volts.

$V_{in} = 3$ volts

Now we have to decide which region or model to use. If we assume pinchoff as in the previous method, then $I_D = 0.1mA/V^2(V_{GS} - V_T)^2 = 0.1(3-2)^2 = 0.1$ mA, and

$V_O = 6 - 0.1mA*5K = 5.5$ volts.

$V_{in} = 6$ volts

Assuming triode, $I_D = k(2(V_{GS} - V_T)V_{DS} - V_{DS}^2)$. Since this equation is a function of V_{DS}, we will have to use it to replace I_D in Kirchoff's equation for the circuit.

$6 = R*k(2(V_{GS}-V_T)V_{DS}-V_{DS}^2) + V_{DS}$

The only unknown is V_{DS}, but this equation is a quadratic. For our specific case,

$0.5V_{DS}^2 - 5V_{DS} + 6 = 0$

or $V_{DS} = 8.605$, or 1.394 volts

We get two possible solutions because our model equation is the equation for a parabola. However, only one solution fits our criteria for triode region, $(V_{GS} - V_T) > V_{DS}$ or $V_{DS} = 1.394$ volts.

But wait! This solution is not the same as the one we got using circuit models. In that case, we got $V_O = 1.2$ volts. This discrepancy is a result of approximation we used in developing the R_{DSon} circuit model where we assumed $V_{DS} << 2(V_{GS} - V_T) = 8$ volts. While we don't quite meet this criteria, the result is not real far off and usually good enough for this type of analysis. This get this error because we are into the curving portion of the parabola whereas the R_{DSon} model approximates the curve as a straight line. If this circuit was used to drive another inverter circuit as is often done in logic circuits, this "low" output is well below the threshold for the next MOSFET and the small error is immaterial.

Electronic systems are divided into two categories; analog and digital. This course deals primarily with digital or switching systems, and as a result, we will not pursue linear amplifiers. A large number of applications require the use of switches and most manufacturers of devices specify the parameters that are especially useful in designing switches. For example, The ME181 n-channel MOSFET in the set of specifications at the end of this book shows an R_{DSon} of 2.0 Ohms (maximum) with a forward current of 2 amps. This specification means that at 2 amps in the drain, there will be a maximum of 4 volts across the drain-source. At lower currents the voltage would be proportionally lower. At 100 mA, $V_{DS} = 0.2$ volts. Typically, this MOSFET would be used as a switch with the input voltage either 0 volts or very close to it, to cut the device off and the output voltage at the supply voltage, or $V_{in} = 10$ volts to cause the output voltage to be close to zero. The drain current would either be zero or fairly large and the output voltage near zero or at the supply voltage. These two states would be recognized as "on" or "off". In this kind of system, areas between the states are not allowed or are "illegal".

GRAPHICAL ANALYSIS OF THE MOSFET INVERTER

A third way to analyze the circuit is graphically on the characteristic curves. This method is sometimes useful, but not used extensively because characteristic curves are often not available.

If we write the Kirchoff equation for the drain loop,

$V_{DD} = I_D R + V_{DS}$

we can solve for I_D

$I_D = (V_{DD} - V_{DS})/R$

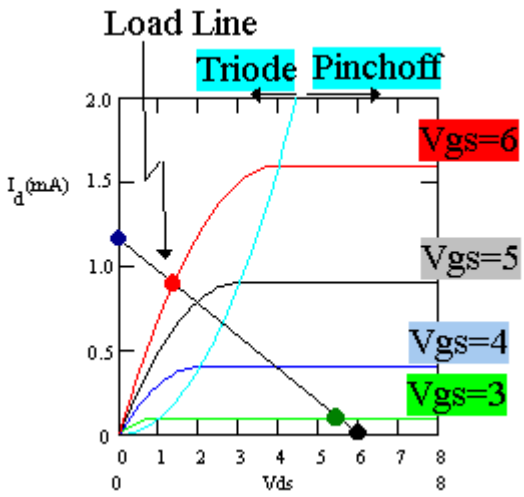


Figure 19. MOSFET characteristic curve with load line.

If we plot this equation, I_D vs V_{DS}, we get the load line as shown in Figure 19. It is easy to plot by finding the two points on the voltage and current axes; $I_D = V_{DD}/R$ when $V_{DS} = 0$, and $V_{DS} = V_{DD}$ when $I_D = 0$.

We can find the solutions for the three cases, $V_{GS} = 0$, 3, and 6 volts by finding the intercepts with these curves and the load line as shown by the three dots on the curves. The output voltages or $V_{DS} = 6$, 5.5, and 1.4, respectively for the three input voltages. These are very close to the calculated voltages. Obviously, it is a little difficult to get much precision with this method. But, the results are usually satifactory.

DESIGN EXAMPLE

We conclude our discussion of analysis of the MOSFET inverter circuit with an example of a logic level buffer circuit. The situation is shown in Figure 20. In this case, we have a TTL logic gate that we wish to use to drive a CMOS gate. Because of logic level incompatibilities, we cannot use the TTL gate to drive the CMOS gate directly. For example, the TTL low and high output voltages are 0.4 and 3.5 volts and the CMOS input requirements are <0.5 and >4.5 volts respectively. The current capabilities af the TTL gate are sinking 1.6mA when low and sourcing 0.4 mA when high. The current requirements of the CMOS gate are ±1μA (in when high and out when low). Clearly, the TTL gate current

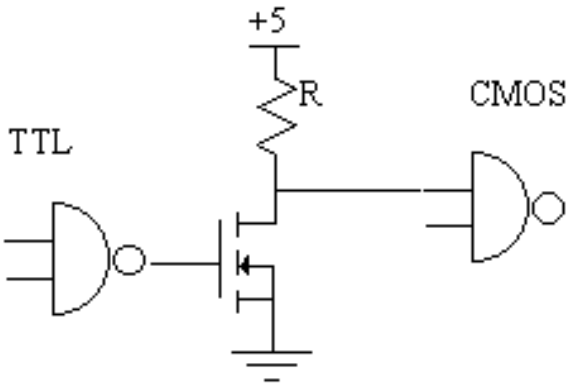


Figure 20. MOSFET logic level buffer circuit

capabilities are adequate, but the voltages are incompatible. For this design, assume we have an n-channel enhancement MOSFET available with the following specifications from the data sheet: V_T = 1.5 volts and R_{DSon} = 100 Ohms at V_{GS} = 5.0 volts and I_D = 30 mA . We will refer to the output of the TTL gate as the input of our buffer circuit and the input of the CMOS gate as the output of our buffer. Also assume we want our high output to be 4.8 volts and our low output to be 0.2 volts.

Input LOW, V_{in} = 0.4 volts

This input voltage is easily below the threshold for the MOSFET inverter, thus, the MOSFET is cutoff, I_D = 0. Then, V_O = 5.0 volts less the voltage drop across the pull-up resistor. Since we can allow 0.2 volts drop across the resistor and the current is 1 μA, the maximum resistor value is 200 K Ohms.

Input HIGH, V_{in} = 3.5 volts

This input is above the threshold voltage so the MOSFET will be conducting. Since we want the output voltage to be 0.2 volts or less, the device is in triode. We need to find the R_{DSon} for V_{DS} = 3.5 volts.

$$R_{DSon}(V_{GS} = 5.0) = 100 = \frac{1}{2k(5-1.5)}$$

$$R_{DSon}(V_{GS} = 3.5) = \frac{1}{2k(3.5-1.5)}$$

Solving, we find that

R_{DSon} (for V_{GS} = 3.5) = 175 Ohms

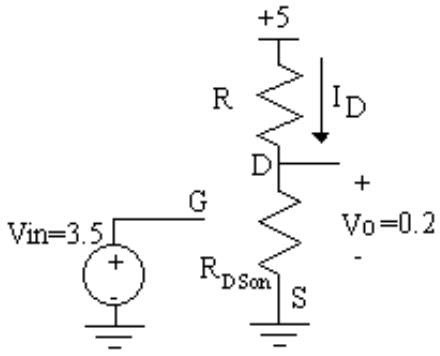


Figure 21. Buffer circuit model with V_{in} = 3.5 v

Drain current can be as much as 1.14 mA and still keep V_{DS} = 0.2 volts (175 Ohms * 1.14mA = 0.20 volts). Or, R can be as low as

$$R = \frac{5-0.2}{1.14mA} = 4.21KOhms$$

Thus, we can choose 4.21 K < R < 200 K Ohms. A convenient value is 10K Ohms. Using this value for the pull-up resistor, we find the low output voltage is 0.086 volts, well within our design objective.

OTHER FIELD EFFECT TRANSISTORS

So far we only discusssed the n-channel enhancement MOSFET, This type is by far the most prevalent, but, other types are sometimes useful. The first FETs were junction devices. Instead of an insulating layer between the gate and channel, a thin layer of n-

material is made for the channel and a p-layer made over that for the gate, a p-n junction. This junction was reverse biased to prevent conduction through the gate. As the reverse bias is increasde, the depletion region at the junction widens, narrowing the channel. Cutoff occurs when the gate is made sufficiently negative. These devices are rarely used today as MOSFETs are easier to make and perform better.

Or useful variation of MOSFETs is to make a lightly doped n-channel just under the gate area. Thus, the device will conduct even with zero volts on the gate. Conduction can be enhanced or decreased by making the gate more positive or negative respectively. If the gate is made sufficiently negative, all the free carriers will be driven from the channel and the device cutoff. This depletion of carriers leads to the name depletion-mode MOSFET.

P-CHANNEL MOSFET

P-channel devices can be made by switching the nand p regions in the device. A p-channel enhancement MOSFET requires the gate be made more negative than the source to bring it into conduction. It takes a little getting used to, but p-channel devices are very useful.

Like the n-channel devices, the p-channel MOSFET operates in three regions, cutoff, pinchoff, and triode, but the voltages and current directions are reversed. See Figure 22 for definitions. The model equations are:

Cutoff, $V_{SG} < V_T$

$I_D = 0$

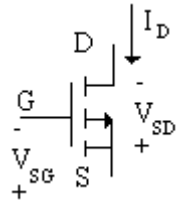


Figure 22. p-channel MOSFET

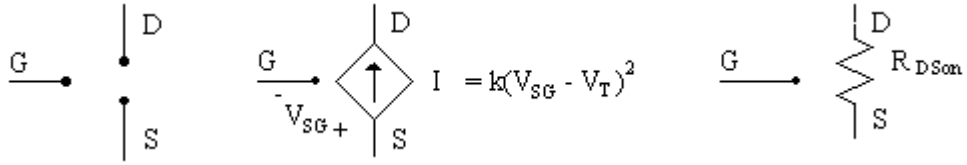
Pinchoff, $V_{SG} - V_T > V_{SD}$

$I_D = - k(V_{SG} - V_T)^2$

Triode, $V_{SG} - V_T < V_{SD}$

$I_D = -k(2(V_{SG} - V_T)V_{SD} - V_{SD}^2)$

We can also represent the p-channel MOSFET with the same circuit models as shown in Figure 23. Note the current reversal in the pinchoff model.



a. Cutoff *b. Pinchoff* *c. Triode*

Figure 23. Circuit models for the p-channel MOSFET

INVERTER CIRCUIT WITH A P-CHANNEL MOSFET

For an example, a p-channel MOSFET is used in an inverter circuit shown in Figure 24. In this circuit, the source is connected to the positive supply. Note that the drain current is negative so that actaul current goes from positive to negative. To avoid negative currents, we will use $I = -I_D$. The source-gate voltage is not the input voltage, but the voltage difference between the V_{DD} and the input $V_{SG} = V_{DD} - V_{in}$ and the output voltage is the voltage across the resistor or $V_O = V_{DD} - V_{SD}$.

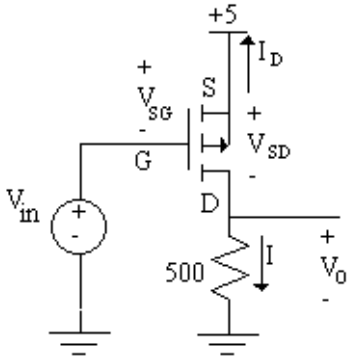


Figure 24. P-channel MOSFET inverter circuit.

For the purposes of this example, $V_T = 1$ volt, and $k = 5$ mA/V^2. We will analyze this circuit for input voltages of $V_{in} = 0$, 3, and 5 volts. We will start with $V_{in} = 5$ volts.

$V_{in} = 5$ volts

In this case, $V_{SG} = 0$ volts, below the threshold voltage, so the MOSFET is cutoff, $I_D = 0$, and $V_O = 0$ volts.

$V_{in} = 3$ volts.

Here, $V_{SG} = 2$ volts, and if we assume pinchoff, $I = -I_D = 5(2-1)^2 = 5$ mA. $V_O = 2.5$ volts. ($2.5 > (V_{SG} - V_T)$, so pinchoff is correct.)

$V_{in} = 0$ volts.

$V_{SG} = 5$ volts. If we assume triode operation, we can replace the MOSFET with $R_{DSon} = 25$ Ohms. Then $V_O = 5(500/525) = 4.76$ volts. Checking our assumptions, $V_{SD} = 0.24$ volts $< V_{SG} - V_T$), so triode is correct. Also, $2(V_{SG} - V_T) = 8 << V_{SD} = 0.24$, so the R_{DSon} approximation is valid.

SWITCHING SPEED OF MOSFETS

It would appear that the impedance of the circuit driving the gate is immaterial because we model the gate as an open circuit and not current flows. However, the source must charge or discharge the gate capacitance during switching and as such, the speed is limited by the R-C charging characteristic. There is capacitance between the gate and drain, channel, and the source. This capacitance is often lumped into a single "input capacitance" as is shown in the specifications, $C_{in} = 17$ pF, for the ME181 MOSFET in the appendix. If the Thevenin equivalent resistance of the source is 100 KOhm, it would take approximately 1 μsecond to switch the input signal from one value to another. This rate of change , dV/dt, for the gate is critical for CMOS logic circuits when running at high speeds as we shall see in the next chapter.

In addition to gate capacitance, rate at which the charges can be moved into and out of the channel by applying voltages to the gate also affect the rate at which the drain current and thus, the output voltage can change. Moving the charges takes time and is affected by the mobility of the carriers and the carrier lifetimes, both of which are a function of doping level in the substrate.

ANALYSIS OF A MOSFET INVERTER WITH AN ACTIVE PULL-UP

In processing MOS devices, making a resistor out of bulk material is an expensive process, both in terms of area and process steps. Therefore, it is appropriate to find other ways to develop logic circuits. One way is to use an active pull-up as shown in Figure 25. Analysis of this circuit follows a procedure similar to the circuit with the resistor pull-up. First, however, several observations must be made. The gate and the drain of Q_2 are connected together making $V_{GS2} = V_{DS2}$. Consequently, Q_2 always operates in the pinchoff region. (Assuming $V_T > 0$.) Because $V_{GS2} = V_{DS2} = V_{DD} - V_o$, then the equation of operation for Q_2 is:

$$I_{D2} = k_2(V_{DD} - V_o - V_{T2})^2 \qquad \text{Equation 6}$$

For Q_1, as in the previous example, $V_{GS1} = V_{in}$, and $V_{DS1} = V_o$. Also, with no load, the current in the two devices is the same, $I_{D1} = I_{D2}$. Now Q_1 may operate in cutoff, in the triode region, or in the pinchoff region.

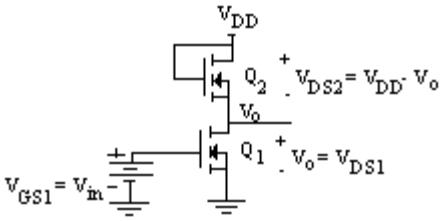


Figure 25. MOSFET inverter with an active pull-up.

Q_1 CUTOFF,
Cutoff, $V_{in} < V_{T1}$, means that the current will be zero and if Eq. 6 is solved for $I_D = 0$, then, $V_o = V_{DD} - V_{T2}$.

Q_1 IN PINCHOFF REGION $V_{in} > V_{T1}$, and $V_{in} - V_{T1} < V_o$

In this region, we start getting some current but the output voltage has not dropped very far. In this case the current in Q_1 is given by

$$I_{D1} = k_1(V_{in} - V_{T1})^2$$

but since this current must be equal to the current in Q_2

$$k_1(V_{in} - V_{T1})^2 = k_2(V_{DD} - V_o - V_{T2})^2$$

and

$$V_o = V_{DD} - V_{T2} - \left[\frac{k_1}{k_2}\right]^{\frac{1}{2}}(V_{in} - V_{T1}) \qquad \text{Equation 7}$$

Q_1 IN TRIODE REGION $V_{in} > V_{T1}$, and $V_{in} - V_{T1} > V_o$

In this case,

$$I_{D1} = k_1[2(V_{in} - V_{T1})V_o - V_o^2] = k_2(V_{DD} - V_o - V_{T2})^2$$

which again is a quadratic equation which can be solved for V_o.

$$V_o^2(k_1 + k_2) - V_o[2k_2(V_{DD} - V_{T2}) + 2k_1(V_{in} - V_{T1})] + k_2(V_{DD} - V_{T2})^2 = 0$$

Equation 8

Example:

Both Q_1 and Q_2 are made on the same piece of silicon at the same time, making the threshold voltage, $V_T = V_{T1} = V_{T2}$ as well as making ε, μ, and t the same for both devices. However, Q_2 is made much longer and narrower than Q_1, making k_2 much smaller than k_1.

For the circuit shown in Figure 25, let us assume $V_{DD} = 10V$, $k_1 = 1mA/V^2$, and $k_2 = 0.1mA/V^2$, and $V_T = 2V$. Plot V_o vs V_{in}.

Solution:

The solution must be taken in three parts: cutoff, pinchoff, and triode regions for Q_1.

Cutoff: $V_{in} < 2V$.

By the above discussion, $V_o = V_{DD} - V_T = 8V$,

Pinchoff: $V_{in} > 2$, and $V_{in} - 2 < V_O$,

To solve this region, we must solve Eq. 7 for every input voltage we wish to plot; or:

$$V_o = 10 - 2 - \sqrt{10}\,(V_{in} - 2)$$ (Note that this is a linear eq.)

For example, if $V_{in} = 3.0$ V, $V_o = 4.84V$. (Still in pinchoff.) Where does pinchoff end? It ends when $V_{in} - 2 = V_o$. Solving for this condition,

$$V_{in} - 2 = 10 - 2 - \sqrt{10}\,(V_{in} - 2)$$

$V_{in} = 3.922$ and $V_o = 1.922$ at the boundary between pinchoff and triode.

TRIODE REGION ($V_{in} - V_T > V_o$)

For all values of $V_{in} > 3.922$, Q_1 will be in the triode region, and Equation 8 must be solved.

$$V_o^2 (1.1) - V_o (2V_{in} - 2.4) + 6.4 = 0,$$

Because a solution to the quadratic equations will result in two answers, you must choose the one which conforms to the triode operation, $V_{in} - 2 > V_o$. A table of solutions is given below and the plot is shown in Figure 26.

TABLE 1
V_o vs V_{in} for the active pull-up inverter example.

Input voltage	<2	3	3.922	4	5	8	10
Output voltage	8.0	4.84	1.922	1.733	0.982	0.490	0.372

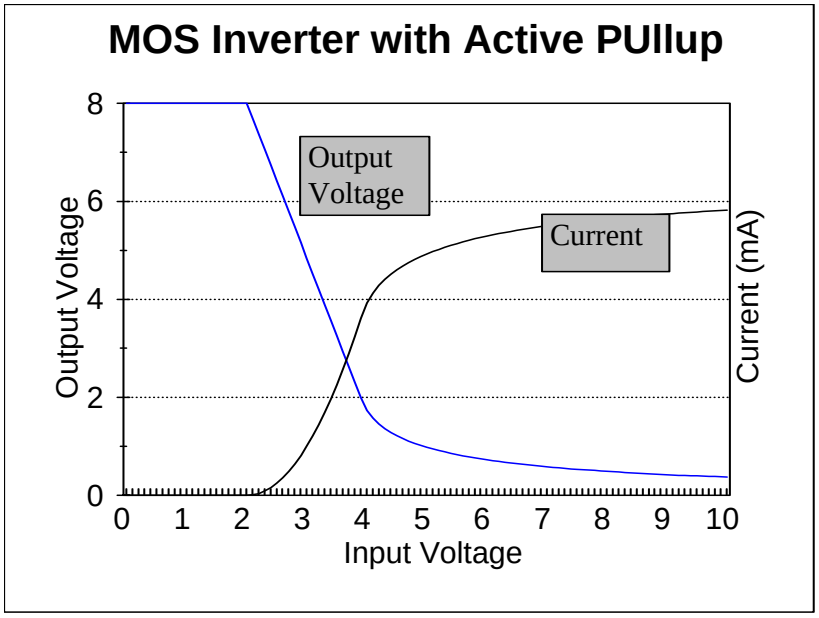


Figure 26. Voltage transfer characteristic for a MOSFET inverter with active pull-up. $V_T = 2.0$ volts, $k_1 = 1$, $k_2 = 0.1$ mA/V^2.

If we cascaded several of these inverters, we could show that the circuit will make a good logic level inverter. For example, if we connected the input of the first inverter to ground, the output would be 8.00 volts. If this was connected to the input of a second inverter, the output of the second would be 0.490 volts. Connecting this to a third, would provide 8.00 volts output, which would then produce another 0.49, etc.

1. $V_{in} = 0$ $V_o = 8.0$

2. V_{in} = 8.0 V_o = 0.49
3. V_{in} 0.49 V_o = 8.0, Which is back to a previous value.

NMOS LOGIC GATES

The circuit in Figure 25 can be made into a logic NOR gate by adding another MOSFET parallel to Q_1. A NAND gate can be made by adding another MOSFET in series with Q_1. These gates are shown in Figure 27 below.

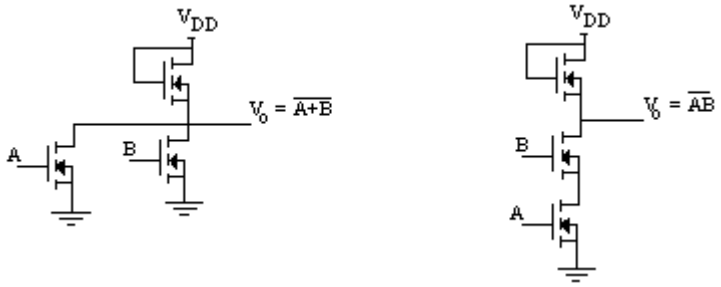


Figure 27. NMOS Logic Gates

NMOS DESIGN EXAMPLE

Design an NMOS buffer that will interface between a 0-5 volt signal and drive two 7400 TTL gates. For the sake of the example, we will assume the input signal comes from a switch as shown in Figure 28, below. The NMOS buffer will consist of an active pull-up inverter. Because the two transistors are fabricated on the same piece of silicon at the same time, all parameters of the devices are the same, except for the dimensions of the devices. Thus, you have control of the width and length of the channels. We can write the k for the devices as k'W/l where W and l are the width and the length, and k' is the rest of the parameters in the original k value. As the designer of integrated circuit devices, once the process has been fixed, you have control only over the dimensions of the devices. Let us assume that V_T = 2.0 volts and k' = 0.5 mA/V^2 are fixed by the fabrication process. Your job as a designer is then to specify the width/length ratio for the two transistors.

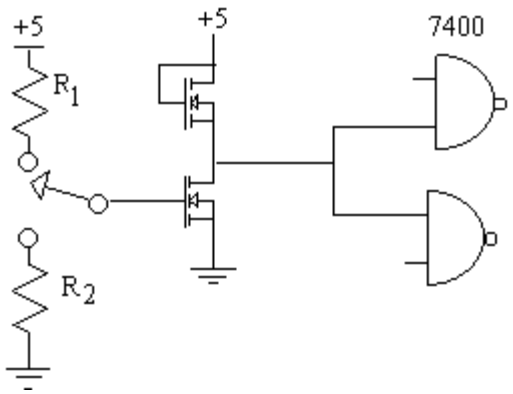


Figure 28. An NMOS circuit used to drive two 7400 gates

The 7400 gates have the following input specifications:

V_{inHmin} = 2.00, V_{inLmax} = 0.80, I_{inHmax} = 0.04 mA, I_{inLmax} = -1.6 mA

When the switch is in the up position, the input voltage to the MOSFET gate is 5.0 volts. (The gate current is zero, so there will be no voltage drop across the resistor.) When the switch is in the down position, the input voltage to the gate of the MOSFET is zero volts. As this circuit is an inverter, the output of the NMOS inverter is low and high respectively for the two cases. From these two cases, then, the load 7400 gates must recognize the low and the high.

Through the control of the transistor dimensions, we can control the high and low voltages at the output of the inverter. It would be a good idea to allow some noise margin at the input of the 7400 gates, so let us aim for V_{oL} = 0.40 volts, and V_{oH} = 2.4 volts, the same as the output of a 7400 gate. In this case, however, we only need to be able to drive 2 7400 gates rather than the 10 allowed as the fanout for the 7400.

When the switch is in the down position, the lower transistor is cutoff , but the upper transistor is always in pinchoff.

$-I_o = 0.080 \text{ mA} = k_2(V_{DD} - V_T - V_o)^2 = k_2(5\text{-}2\text{-}2.4)^2 = k_2 \times 0.36$

$k_2 = 0.222 \text{ mA/V}^2 = k'(W/l)_2$

$(W/l)_2 = 0.444$ for the upper transistor

When the switch is in the up position, the lower transistor is in triode mode and its current will be

$I_1 = I_2 + I_o = k_2(V_{DD} - V_T - V_o)^2 + 3.2 \text{ mA}$

$= 0.222(5 - 2.0 - 0.4)^2 + 3.2 \text{ mA} = 4.70 \text{ mA}$

Because the lower transistor is in triode,

$4.70\ mA = k_1(2(V_{in} - V_T)V_o - V_o^2) = k_1(2(5-2)0.4 - 0.16)$

$= k_1 \times 2.24$

$k_1 = 2.098\ mA/V^2$

$(W/l)_1 = 4.196$ for the lower transistor.

EXERCISES:

1. For an n-channel enhancement MOSFET, you must make the gate more _____ than the source in order to turn it on.

2. For a p-channel enhancement MOSFET, you must make the gate more _____ than the source in order to turn it on.

3. Draw the symbols for n-channel and p-channel MOSFETs. Draw arrows indicating normal, conducting current directions at each terminal.

4. A MOSFET Inverter with 2 Kohm pull-up resistor has an input voltage of 4 volts. The MOSFET has the following specifications:

$V_{DD} = 5$ $k = 2$ $V_T = 2.5$

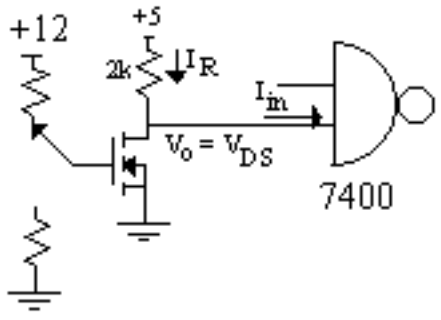
What is V_o?

5. A MOSFET with a 2 KOhm pull-up resistor is used as a buffer between a 12 volt switched signal and a 7400 gate as seen below.

$V_T = 2.5$ volts, $k = 0.5$ mA/V^2

What is the output voltage of the buffer, (the input voltage of the gate) when the switch is up? _____

Is this voltage sufficiently low to be recognized as a low? (<0.8volts)_____



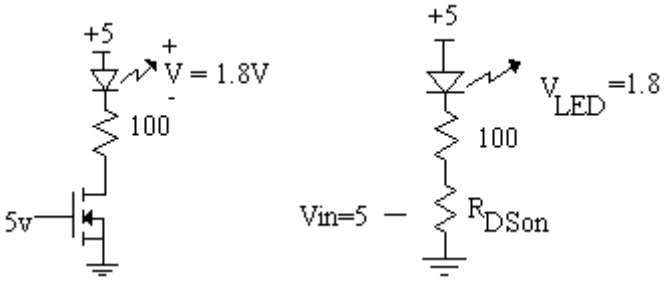
6. An n-channel MOSFET is used to drive an LED in the circuit shown. The device specifications are:

$1.0 < V_T < 2.0$, $R_{DSonmax} = 6$ Ohms @ $V_{GS} = 10$ volts, $I_D < 100$ mA

a. Using the linear approximation, what is the current?_____

b. Which limit on V_T gives the worst-case (minimum current) results?_____

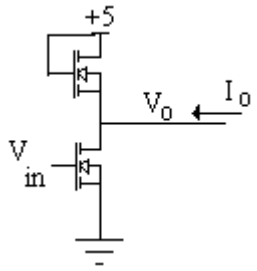
c. Does this application satisfy the linear approximation criterion?



7. An active pull-up inverter is shown below. What is the relationship between the gate-source and drain-source voltages for the pull-up device?

V_{GS} = _____

8. Design Exercise: The inverter below is used to drive five 7400 TTL gates. What k values are required for the two transistors? Assume V_T = 2.0 volts for both and the input voltage switches from 0-5 volts. For the purposes of this exercise, design for V_{oH} = 2.5 volts, V_{oL} = 0.5 volts, I_{oH} = -200 μA, I_{oL} = 8 mA.



9. Repeat the previous design example, except using a resistor pull-up in the inverter circuit. Determine the pull-up resistor value and the necessary k value for the MOSFET. The k value will be a function of the resistor value chosen.

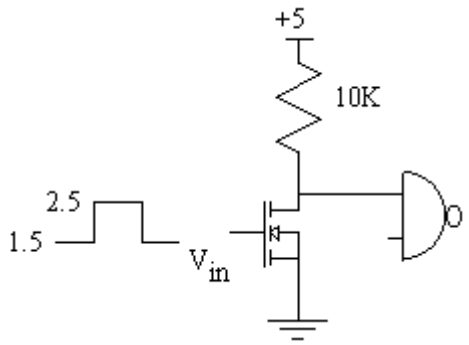
R = _____ k = _____

Problems

1. An n-channel enhancement MOSFET has V_T =1.8 Volts, and k=2.5 mA/V^2. This transistor is used in an inverter circuit with a 470 Ω pull-up resistor to a 5 Volt power supply. Calculate the output voltage with V_{in} = 1.8 - 5.0 volts. Plot the voltage transfer characteristic, V_o vs V_{in}. Find the input and output voltages where the MOSFET makes the transition between Pinchoff and Triode.

2. An inverter is made with two -n-channel MOSFET's. The bottom MOSFET has the parameters V_T =1.8 Volts, and k_1 =2.5 mA/V^2. The second is an active pull-up n-channel MOSFET with V_T = 1.8 volts, and k_2 = 1.0 mA/V^2. Do the same calculations and plot as for Problem 1. Find the input and output voltages where the MOSFET makes the transition between Pinchoff and Triode.

3. A MOSFET inverter is to be designed to convert a 1.5-2.5 volts square wave into a TTL compatible signal. The output of your inverter is to be able to drive one 7400 gate. The process parameters for making the NMOS transistor are such that V_T = 2.0 volts and k' = 0.2 mA/V^2, where k = k'W/L. In this equation, W is the width of the channel and L is the length of the channel. Your job is to determine the required W/L ratio. Assume V_{OLmax} = 0.6 volts for the inverter.



Problem 3 Circuit

4. The data sheet for an n-channel, enhancement MOSFET gives the following specifications:

V_T = 1.8 volts, R_{DSon} = 5 Ohms @V_{GS}=10 volts and I_D = 20 mA.

a. Determine k.

b. If this MOSFET is used at V_{GS} = 5 volts, what is the effective R_{DSon}? What is the maximum drain current such that $V_{DS} < 0.1\ (2(V_{GS} - V_T))$?

c. If this MOSFET is used at I_D = 10 mA and V_{GS} = 5 volts, how much error in V_{DS} occurs if you use the R_{DSon} from part b instead of the full triode model?

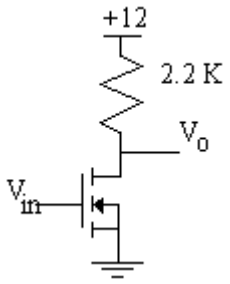
5. The process parameters for an NMOS process are:

$k' = 0.4\ mA/V^2$

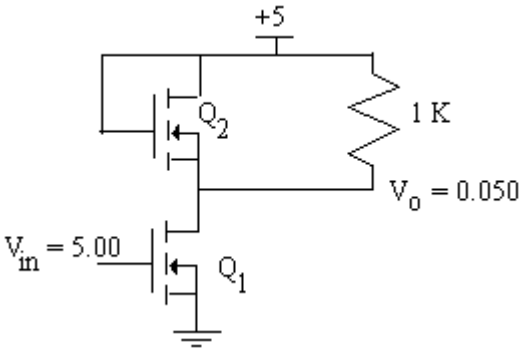
$V_T = 1.5$ volts

Determine the W/L ratios for both MOSFETs in an NMOS inverter circuit that drive 5 74LS00 gates. The required output voltages and currents of the NMOS inverter circuit are $V_{OH} = 3.0$ volts, $V_{OL} = 0.5$ volts, $I_{OH} = -0.1$ mA, and $I_{OL} = 2.0$ mA. The power supply is 5 volts, and the input voltages are 0.5 and 3.5 volts.

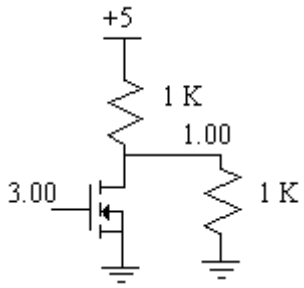
6. The input voltage for the following MOSFET inverter circuit is 2.5 volts. The MOSFET has the following parameters: $V_T = 1.5$ volts, $k = 4.0\ mA/V^2$. What is the output voltage?



7. An NMOS inverter is used to drive a resistive load as shown below. The upper transistor has $k_2 = 0.5\ mA/V^2$. When the input voltage is 5 volts, the output voltage of the inverter is 0.5 volts. What is k_1? V_T for both transistors is 1.00 volts.

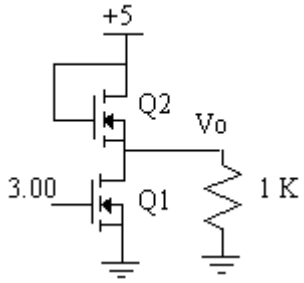


8. A MOSFET inverter with a resistor pull-up has a resistor load as shown. The input voltage is 3.00 volts and the output voltage is 1.00 volts. What is k for the MOSFET? $V_T = 1.00$ volts.



9. An NMOS circuit shown below has a resistive load. Find the output voltage when the input voltage is 3.00 volts.

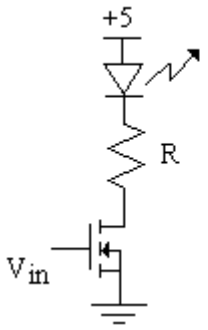
$k_1 = 1.00$ mA/V^2, $k_2 = 0.25$ mA/V^2, $V_T = 1.00$ volt for both devices.



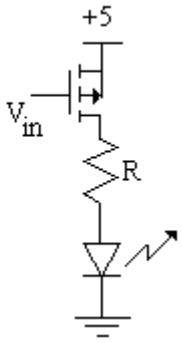
Design Problems

10. An n-channel MOSFET is being driven by a 7400 gate. The MOSFET is used to drive an LED at 25 mA when the LED has a forward voltage drop of 1.8 volts. The MOSFET has the parameters $V_T = 1.5$ volts and $k = 2.0$ mA/V^2. Find the required value of R using a 5 volt power supply.

11. A p-channel MOSFET is used to drive an LED as shown in the circuit below. The MOSFET has the following parameters: $V_T = 2.0$ volts and $k = 5$ mA/V^2. The LED is to have 20 mA and has a voltage drop of 2.0 volts at that current. Find the appropriate value of R if the input voltages are 0 and 5 volts.



Problem 10



Problem 11

CHAPTER 9

CMOS

CMOS, Complementary MOS, logic has become quite popular because of its low static power dissipation, wide noise margins, and its tolerance of a wide range of power supply voltages. We will begin the analysis of CMOS gates, by looking at an inverter as shown in Figure 1. Notice that the p-channel device is connected upside down with its source connected to the positive supply.

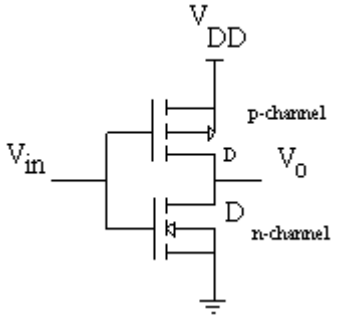


Figure 1. CMOS inverter circuit

The characteristic equations that we will use to model an n-channel MOSFET are:

(CUTOFF) I. $V_{GS} < V_T$,

$$I_D = 0$$

(TRIODE) II. $V_{GS} > V_T$, and $V_{DS} < V_{GS} - V_T$

$$I_D = k(2(V_{GS} - V_T)V_{DS} - V_{DS}^2) \quad (1)$$

(PINCH-OFF) III $V_{GS} > V_T$, and $V_{DS} > V_{GS} - V_T$

$$I_D = k(V_{GS} - V_T)^2 \quad (2)$$

A p-channel MOSFET works the same as the n-channel except that it has negative voltages and currents. The equations are the same, except that the subscripts are reversed for the relative voltages. Thus, in the three regions:

(CUTOFF) I. $V_{SG} < V_T$,

$$I_D = 0$$

(TRIODE) II. $V_{SG} > V_T$, and $V_{SD} < V_{SG} - V_T$

$$I_D = -k(2(V_{SG} - V_T)V_{SD} - V_{SD}^2) \quad (4)$$

(PINCH-OFF) III $V_{SG} > V_T$, and $V_{SD} > V_{SG} - V_T$

$$I_D = -k(V_{SG} - V_T)^2 \quad (5)$$

For this convention, V_T must still be used as the absolute value, or a positive number. Also, as noted by the negative sign, the current is leaving the drain rather than entering the drain as in the n-channel.

THE CMOS INVERTER

The CMOS inverter shown in Figure 1 is the basis for all CMOS logic. It is nearly the ideal logic device. Under normal operation, one transistor is on and the other is off, resulting in a low impedance path to either the ground or the positive supply and with no current flowing. For example, if V_{in} =0, then the n-channel is in cutoff, the p-channel is turned on, and $V_o = V_{DD}$. Similarly, if $V_{in} = V_{DD}$, then the p-channel is in cutoff, the n-channel is on, and V_o =0.

The CMOS inverter can be analyzed using the equations above. The two transistors are in series making the current the same in both, if there is no load. This no-load condition is met in any logic system which uses only CMOS gates since the gate current is very small (picoamperes). The equal current condition will not be met for any gate that has a significant load and the solution becomes more complex.

The voltage transfer characteristic is normally the first analysis done. In this case, we are interested in the output voltage as a function of the input voltatge. The two logic levels of interest are V_o= 0 volts and $V_o = V_{DD}$. Normally, the two transistors are made with $k_n = k_p = k$ and $V_{Tn} = V_{Tp} = V_T$. Also, we will assume that $V_{DD} > 2V_T$.

Let's rewrite Equations 1, 2, 4, and 5 in terms of the circuit parameters.

n-channel: $V_{GS} = V_{in}$ $V_{DS} = V_o$

TRIODE REGION,
$V_{GS} > V_T$, and $V_{GS} - V_T > V_{DS} \Rightarrow V_{in} > V_T$, and $V_{in} - V_T > V_o$

$$I_D = k(2(V_{in} - V_T)V_o - V_o^2) \qquad (6)$$

PINCH-OFF REGION,
$V_{GS} > V_T$, and $V_{GS} - V_T < V_{DS} \Rightarrow V_{in} > V_T$, and $V_{in} - V_T < V_o$

$$I_D = k(V_{in} - V_T)^2 \qquad (7)$$

p-channel: $V_{SG} = V_{DD} - V_{in}$ $V_{SD} = V_{DD} - V_o$

TRIODE REGION,
$V_{SG} > V_T$, and $V_{SG} - V_T > V_{SD} \Rightarrow V_{in} < V_{DD} - V_T$ and $V_{in} + V_T < V_o$

$$I_D = -k(2(V_{DD} - V_{in} - V_T)(V_{DD} - V_o) - (V_{DD} - V_o)^2) \qquad (8)$$

PINCH-OFF REGION,
$V_{SG} > V_T$, and $V_{SG} - V_T < V_{SD} \Rightarrow V_{in} < V_{DD} - V_T$ and $V_{in} + V_T > V_o$

$$I_D = -k(V_{DD} - V_{in} - V_T)^2 \qquad (9)$$

Now let us turn to calculating the voltage transfer characteristic. The characteristic is given in Figure 3 for arbitrary $V_{DD}>2V_T$. There are five cases or regions of the input voltage of interest. We will start with the input voltage low.

CASE 1. $V_{in} < V_T$,

If we start with the input at 0 volts, The n-channel transistor will be off and the p-channel will be on.

n-channel:

$V_{in} < V_T$

The n-channel is cutoff and therefore, $I_D = 0$

p-channel:,

$V_{in} < V_{DD} - V_T$, The p-channel is turned on and with no current because the n-channel is turned off. Equation 9 has no solution for $I_D = 0$ except $V_{DD}=V_{in}+ V_T$ which is contrary to the input conditions. Therefore, the p-channel is not in pinchoff. Equation 8 does has a solution with $I_D = 0$ at $V_o=V_{DD}$ with the p-channel in the triode region.

This case holds for the input voltage up to $V_{in}=V_T$ as long as $V_{DD}>2V_T$.

CASE 2. $V_T < V_{in} < \frac{V_{DD}}{2}$

We start into this case at the boundary V_{in} just begins to exceed V_T which makes the output voltage very close to V_{DD}, leaving the p-channel in the triode region and the n-channel just beginning to conduct in the pinchoff region. These conditions hold throughout this case.

n-channel: PINCH-OFF REGION,

$V_{in} > V_T$, and $V_{in} - V_T < V_o$

$$I_D = k(V_{in} - V_T)^2 \qquad (10)$$

p-channel: TRIODE REGION,

$V_{in} < V_{DD} - V_T$ and $V_{in} + V_T < V_o$

$$I_D = -k(2(V_{DD} - V_{in} - V_T)(V_{DD} - V_o) - (V_{DD} - V_o)^2) \qquad (11)$$

These conditions clearly exist for the input voltage just slightly greater than the threshold. Later, however, we will show that they exist as the input voltage nears half of the supply voltage, the upper end of the region.

Because the currents are equal,

$$k(V_{in} - V_T)^2 = k(2(V_{DD} - V_{in} - V_T)(V_{DD} - V_o) - (V_{DD} - V_o)^2) \qquad (12)$$

Collecting terms:

$$k(V_{DD} - V_O)^2 - (V_{DD} - V_O)[1k(V_{DD} - V_{in} - V_T)] + k(V_{in} - V_T)^2 = 0$$

This equation can be solved for V_o as a function of V_{in} and plotted in Figure 3. (Region 2)

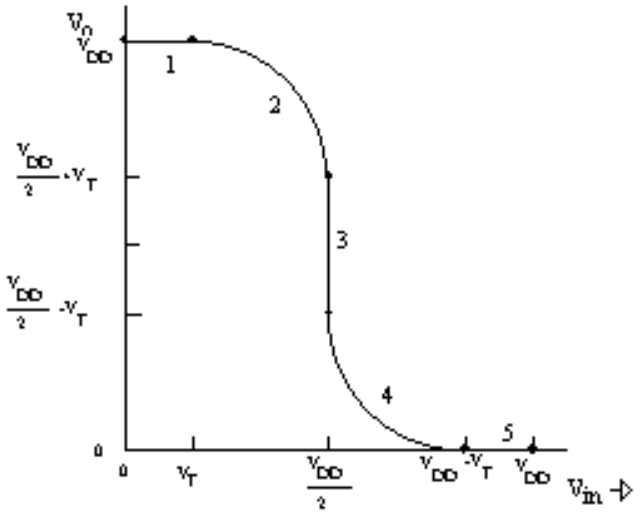


Figure 3. Voltage transfer characteristic for a CMOS inverter.
Case 1. n-channel cutoff, p-channel triode
Case2. n-channel pinchoff, p-channel triode
Case3. n-channel pinchoff, p-channel pinchoff
Case 4. n-channel triode, p-channel pinchoff
Case 5. n-channel triode, p-channel cutoff

When the input voltage reaches $V_{DD}/2$, the output voltage drops to $V_{DD}/2 + V_T$. At this point for the p-channel, $V_{SD}=V_{SG}-V_T$, the boundary between the triode and pinch-off regions. This result can also be shown by solving Equation 12 for V_o when the p-channel reaches the limit of the triode region, $V_{DD}-V_{in}-V_T=V_{DD}-V_o$ ($V_{in}=V_o-V_T$). Try it!

Case 3. Vin = $\frac{V_{DD}}{2}$

In this region, both the transistors are in their pinch-off regions.,

n-channel: pinch-off,

$V_{in} > V_T$, and $V_{in} - V_T < V_o$

$$I_D = k(V_{in} - V_T)^2 \qquad (13)$$

p-channel: pinch-off,

$V_{DD} - V_{in} > V_T$, and $V_{DD} - V_{in} - V_T < V_{DD} - V_o$ ($V_{in} + V_T > V_o$)

$$I_D = -k(V_{DD} - V_{in} - V_T)^2 \qquad (14)$$

Setting the two currents equal,

$$k(V_{in} - V_T)^2 = k(V_{DD} - V_{in} - V_T)^2 \quad (15),$$

or,

$$+/-(V_{in} - V_T) = +/-(V_{DD} - V_{in} - V_T) \quad (16),$$

This yields only a single value for V_{in},

$$V_{in} = \frac{V_{DD}}{2} \quad (17),$$

In this case, the output voltage can take on any value in the range

$$\frac{V_{DD}}{2} - V_T \le V_o \le \frac{V_{DD}}{2} + V_T \quad (18),$$

Case 4: $\frac{V_{DD}}{2} < V_{in} < V_{DD} - V_T$

This case is the reverse of case 2. The output voltage has fallen low enough that the n-channel is in the triode region while the p-channel continues in pinchoff.

n-channel: TRIODE REGION,

$V_{in} > V_T$, and $V_{in} - V_T > V_o$

$$I_D = k(2(V_{in} - V_T)V_o - V_o^2) \quad (19)$$

p-channel: PINCH-OFF REGION,

$V_{DD} - V_{in} > V_T$, and $V_{in} + V_T > V_o$

$$I_D = -k(V_{DD} - V_{in} - V_T)^2 \quad (20)$$

$$k(2(V_{in} - V_T)V_o - V_o^2) = k(V_{DD} - V_{in} - V_T)^2 \quad (21)$$

As in case 2, the two currents can be set equal to each other and the resulting quadratic solved for any input voltage within the range of this case. The end-points of the curve in Figure 3 can easily be found by plugging in the two extreme values of V_{in}, $V_{in} = 0.5V_{DD}$, or $V_{in} = V_{DD} - V_T$. The solutions are $V_O = 0.5V_{DD} - V_T$, or $V_O = 0$, respectively.

CASE V. $V_{in} > V_{DD} - V_T$,

In this case, the p-channel is off and the n-channel is in the triode region and $V_o = 0$ V.,

CURRENT DURING SWITCHING

When $V_T < V_{in} < V_{DD} - V_T$, both transistors are conducting. Outside of this region, one of the transistors is cutoff and the current is zero. This current occurs during the transistion and produces power dissipation in the transistors. Obviously, the power dissipation will be higher as the number of transitions during a period of time increases. At very low speeds, this power loss is insignificant, but can become quite high at several megahertz. What does this current look like?,

Because the currents in the two transistors are equal, it is easier to calculate the current for the transistor in the pinch-off region. Also, the circuit is symmetric so we only need

to calculate the current for one half of the transition. During the first half of the transition, the n-channel is in the pinch-off region. The current is given by

$$I_D = k(V_{in} - V_T)^2 \qquad (22)$$

The current during the transition is given in Figure 4. Notice that the current graph varies as a square function up to $V_{DD}/2$ and is symmetric about $V_{DD}/2$ for identical MOSFETs.

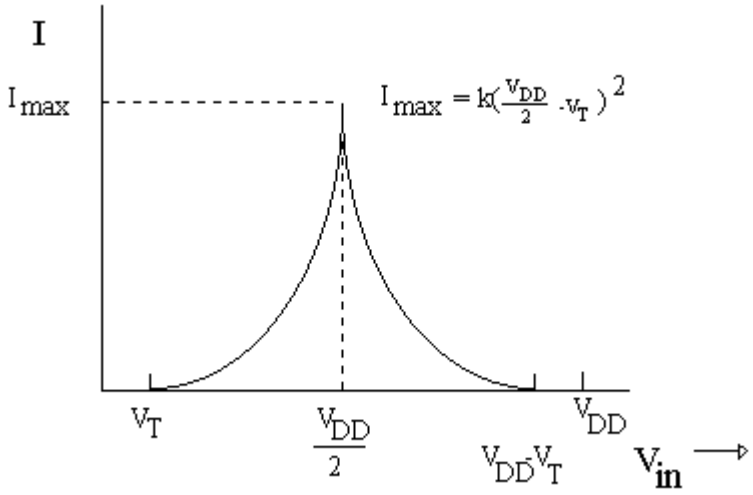


Figure 4. Current through a CMOS inverter as a function of V_{in}.

CMOS GATE TERMINAL SPECIFICATIONS,

From our discussion so far, we have only described the operation of the inverter and developed the voltage transfer characteristic. We have not yet developed terminal specifications for the circuit. It is clear that an input voltage at either rail voltage will cause one of the transistors to turn off and with no load, the output voltage will be at the other rail. That situation prevails even if the input voltage comes away from the rail by as much as V_T.

To go beyond this point or to operate with a load, involves a complex relationship between voltage and current. The CMOS manufacturers have taken a different route. If we take a look at the voltage transfer characteristic, we see that the slope of the curve changes drastically during the transition. The slope is the ratio of the change in output voltage over the change in input voltage. If the slope has a magnitude of less than one, the output voltage will change less than the change of the input voltage. Thus, theoretically, V_{inLmax} and V_{inHmin} are the points where the slope equals -1 as shown on Figure 5. The limits on the output voltage are selected to allow substantial noise margins.

If we look at the specification sheet of a 74C00, we see the results of this approach. With a 5 Volt supply, the noise margins are 1 Volt for both states. The output currents are

specified to allow a fanout of 10. Note that the input current maximum is 1 μA, some 200 times larger than the typical. At 5nA, the typical input current is negligible.

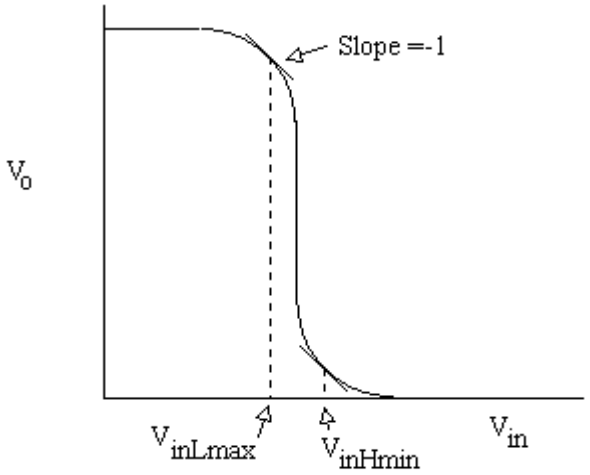


Figure 5. VTC for a CMOS inverter showing input voltage limits.

We observe that an output current of only 10 μA make this logic family difficult to use when interfacing with other logic families or other circuits. It is clear that one of these gates cannot drive any of the TTL gates. Manufacturers provide some help by developing special circuits that do have much higher current drive capabilities. Also, in an attempt to increase the speed of CMOS logic, advanced families of CMOS have been developed, HC, HCT, AC, and ACT. The H stands for High speed, the A stands for Advanced, and the T stands for circuits having input specifications compatible with TTL circuits.

CMOS LOGIC GATES,

MOS logic gates use a resistor or MOSFET which is awlays in the active region and conducts when the output voltage is low. The strength of the CMOS logic is that there is no current in either static state, only during the transistion. When we make logic gates from CMOS we must maintain this cutoff situation in both states. This constraint requires more complexity for CMOS gates, and hence, requires more silicon area. Two-input NAND and NOR gates are shown in Figures 6 and 7. In both cases, one pair of transistors is in series and the other pair is in parallel.

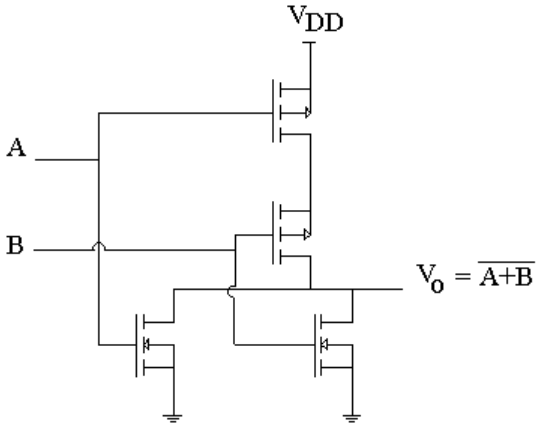


Figure 6. CMOS NOR Gate

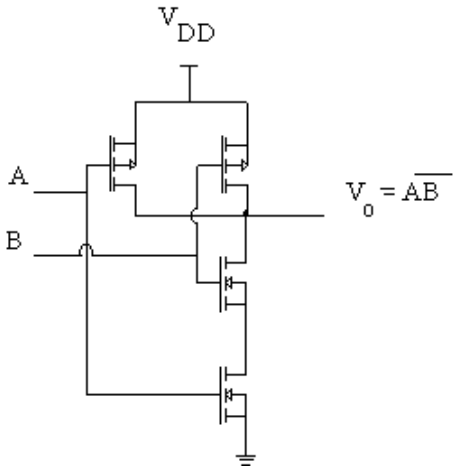


Figure 7. CMOS NAND Gate

NON-ZERO LOADS,

When non-CMOS loads are connected to CMOS logic gates, you can expect current to flow. Four cases are shown in Figure 8. In all cases, we will assume only logic levels are applied to the inputs of the gates. Thus, in each case, either the p-channel or n-channel transistors are cutoff. Two cases are trivial in that no current will flow (upper right and lower left). The other two cases reduce to the case discussed in the previous chapter with an n-channel MOSFET inverter with a resistor pullup or a p-channel MOSFET with a pull-down resistor. .

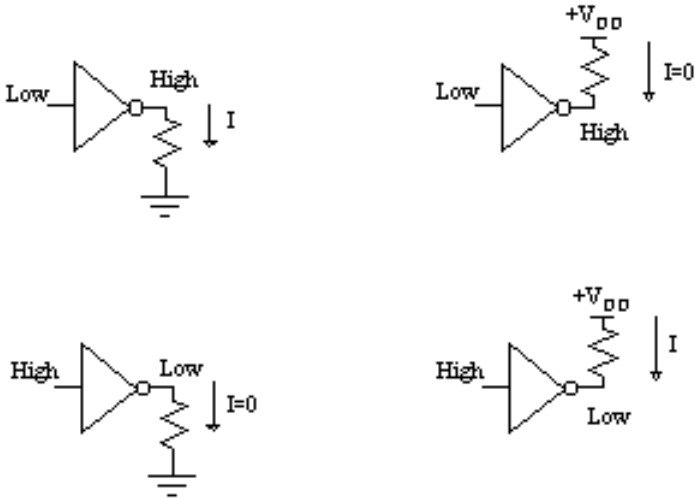


Figure 8. Four cases of resistor loads on CMOS gates.

CMOS DESIGN EXAMPLE

Interface Between CMOS and TTL

CMOS logic using a 5 volt power supply, normally switches between the rails; 0-5 volts. Design a CMOS inverter/interface that can drive one 7400 gate. This design requires the determination of a suitable value of k for the MOSFETs. Assume the MOSFETs are symmetric with $k_n = k_p$ and both have a threshold, $V_{Tn} = V_{Tp} = 2$ volts.

Requirements on the input for the 7400:

$V_{inLmax} = 0.8$ volts, $V_{inHmin} = 2.0$ volts, $I_{inLmax} = 1.6$ mA, $I_{inHmax} = -.040$ mA.

Design the transistors in the circuit below such that the noise margins are at least 0.4 volts.

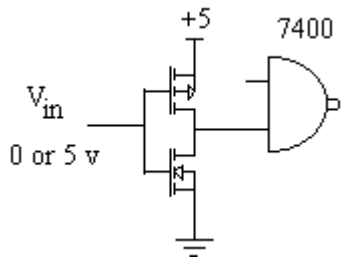


Figure 9. Circuit for design example

SOLUTION:

OBJECTIVE: Design a CMOS inverter/interface to drive one 7400 load.

PERFORMANCE REQUIREMENTS:

Must meet 7400 gate input specifications:

$V_{inLmax} = 0.8$ v, $V_{inHmin} = 2.0$ v, $I_{inLmax} = 1.6$ mA, $I_{inHmax} = -0.04$ mA

Noise margins must be at least 0.4 volts. ($V_{oL}<0.4$, $V_{oH}>2.4$)

CONSTRAINTS: The n-channel and p-channel FETs are to be symmetric; ie. $k_n = k_p$.

Threshold voltage = 2.0 volts

5 volt power supply

Interface driven by other 5 volt CMOS, High = 5 v, Low = 0 volts

DESIGN:

The design in this case consists of deciding on the value of k for the two transistors.

With the input voltage to the CMOS interface cirucit low, the output will be high, the n-channel will be off, and the p-channel will be on and supplying 40 microamps to the load gate. Depending on the output voltage, the p-channel could be either in triode of

pinchoff. However, because this current is much lower than for the output low case, and the k-value for both transistors is to be the same, the output-low case will be the determining case.

In the second case, the voltage at the input to the interface is 5 volts, the p-channel will be cutoff and the n-channel will be on. We will design the output voltage of the interface circuit to be at 0.4 volts. The n-channel will be in the triode region and must sink 1.6 mA.

$$I_n = 1.6 \text{ mA} = k(2(V_{in} - V_T)V_o - V_o^2) = k(2(5-2)0.4 - 0.16) = k \times 2.24$$

or

$$k = 0.714 \text{ mA/V}^2$$

We are now in the position to calculate the voltage when the output is high. If we assume the p-channel is in triode,

$$I_p = 0.04 \text{ mA} = k(2(V_{DD} - V_T - V_{in})(V_{DD} - V_o) - (V_{DD} - V_o)^2)$$
$$= 0.714(2(5-2-0)(5-V_o) - (5 - V_o)^2)$$

or

$$0.714(5 - V_o)^2 - 4.286(5 - V_o) + 0.04 = 0$$

Solving, $V_o = 4.991$. Clearly, the p-channel is in triode and the high output voltage easily exceeds the requirements.

POWER AND ENERGY DISSIPATION

During the transitions of a CMOS gate, L-H, and H-L, current flows through both transistors. This current results in a power dissipation in the devices.

$$p(t) = i(t)v(t)$$

Total power in the combined transistors is

$$p(t) = i(t)V_{DD}$$

To use a simple example, assume a symmetric CMOS inverter where

$$k_n = k_p = k$$

and

$$V_{Tn} = V_{Tp} = V_T$$

Also assume $V_{in}(t) = at$, a ramp function as shown in Figure 10. When $V_{in} = V_T$, the n-channel turns on, and when $V_{in} = V_{DD} - V_T$, the p-channel turns off. In between, both are conducting and power is being dissipated.

For the first half of the transition, the n-channel is in pinchoff, the current is

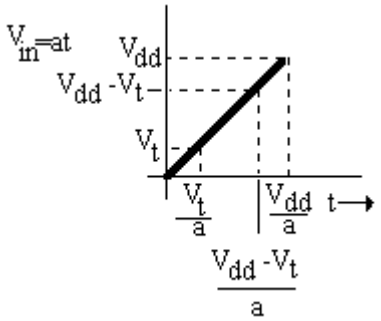


Figure 10. Input voltage ramp function

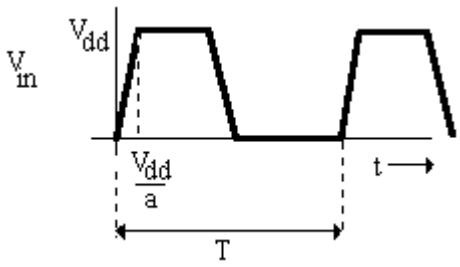


Figure 11. "Square" wave input

$$i(t) = k(V_{in}(t) - V_T)^2$$

and the power dissipation is

$$p(t) = i(t)V_{DD}$$
$$= V_{DD}\, k(V_{in}(t) - V_T)^2$$

Energy dissipated during first half transition is

$$E = \int p(t)dt$$

$$E = \int_{Vin=Vt}^{Vin=Vdd/2} kV_{DD}(at - V_T)^2\, dt$$

For first half transition, the total energy dissipated is

$$E = \frac{kV_{DD}}{3a}\left(\frac{V_{DD}}{2} - V_T\right)^3$$

We have assumed a symmetric CMOS gate so the energy dissipated during the total transition is twice the first half, or

$$E_T = \frac{2kV_{DD}}{3a}\left(\frac{V_{DD}}{2} - V_T\right)^3$$

We did the above calculations with a positive going transistion. The result would have been the same if the transistion were falling with the same slope.

Now let's look at the energy dissipated when a "square" wave shown in Figure 11 is used to drive the inverter. We assume the square wave has a rise and fall time with the slope a which we used in the calculation with a single transition. The frequency is

f = 1/T

Average power over any given time is the energy dissipated during each transition, divided by the time between transistions. There are two transitions per period, so the average power is twice the energy per transition divided by the period.

PAvg = E/T = Ef

$$P_{Avg} = \frac{4\,fkV_{DD}}{3a}\left(\frac{V_{DD}}{2} - V_T\right)^3$$

Note that the power increases with frequency, k, and V_{DD}.

Power decreases with a (slope of V_{in}) and as V_T approaches ½V_{DD}.

Example:

V_{DD} = 5 volts, k = 1 mA/V^2, V_T = 1 volt,

Frequency = 1 kHz,
a = 5V/ns (rail-rail in 1ns)

$$P_{Avg} = \frac{4x10^3 x1x5}{3x5x10^9}(2.5-1)^3$$

P_{Avg} = 2.08 μW

However, if the frequency goes to 100 MHz, (Period = 10 ns)

P_{Avg} = 208 mW

more than a typical IC can dissipate without getting too hot.

Capacitive Loads

The power dissipation problem is made worse with capacitive loads. The capacitor charges to V_{DD} when gate output is high, and, discharges to ground when gate output is low. The currents are shown in Figure 12, first as the output goes high, the capacitor charges toward the power supply, and then as the output goes low when the capacitor discharges to ground.

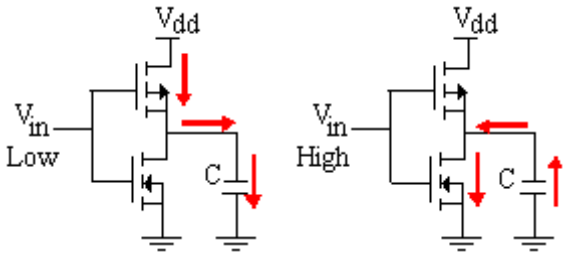


Figure 12. Capacitive load currents with rising and falling transitions.

With each cycle (L-H, H-L), a fixed charge or energy goes from power supply to ground.

$$E = \int V_{DD} i dt$$

The total charge is

$$Q = \int i dt$$

so the total energy dissipated in each cycle is

$$E = QV_{DD}$$

The charge transfered is $Q = CV_{DD}$, making the energy for each cycle

$$E = CV_{DD}^2$$

Average power dissipated due to the capacitive load with a square wave is

$$P_{Avg} = fCV_{DD}^2$$

Again, power increases with frequency.

Example:

Use the same gate as previous example with 20 pF load capacitance and calculate the power dissipated due to the capacitive load.

At 1 kHz

$$P_{Avg} = 10^3 x 20 x 10^{-12} x 5^2$$
$$= 0.5\ \mu W$$

If the frequency is raised to 100 MHz

$$P_{Avg} = 50\ mW$$

Independent of V_T and k.

In any system, the total power dissipation is the sum of power due to transition and load capacitance.

CMOS gates can be designed for reduced transitional power dissipation, but not capacitve load power loss. These losses must be controlled by limiting wiring and load input capacitance. It is interesting to note that CMOS gates inherently have a problem with input capacitance; the gate is essentially a capacitor. Input capacitance is limited by minimizing gate area.

EXERCISES

1. a. For the CMOS inverter shown, label the p-channel and n-channel devices.

b. Assume V_T=1.5 volts. If V_{in} = 1.0 volts, What are the states of the two FETs?

n-channel _____ p-channel _____

c. If V_{in} = 4 volts, what are the states of the two FETs?

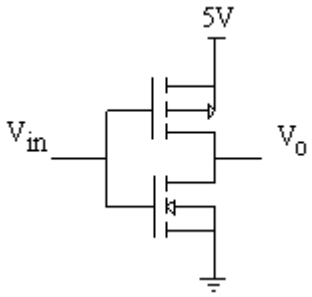
n-channel _____ p-channel _____

d. If V_{in} = 2 volts, what are the states of the two FETs?

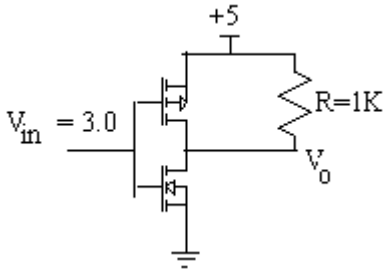
n-channel _____ p-channel _____

e. If V_{in} = 3 volts, what are the states of the two FETs?

n-channel _____ p-channel _____



2. Determine the output voltage and the currents in the two transistors for the following circuit. $k_n = k_p = 2$ mA/V^2, $V_{tn} = V_{tp} = 1.5$ volts.



3. Determine the total power dissipation of a CMOS gate if V_{DD} = 5 volts, C = 25 pF, and frequency = 300 MHz, with Rail-Rail input switching transition of 0.5 ns.

The CMOS transistors are symmetric with V_T = 1.5 volts, k = 0.5 mA/V^2

$P_{AVG\ Transistion}$ = _____ $P_{Capicitive}$ = _____ P_{Total} = _____

Problems

1. A CMOS inverter is to be designed such that it will drive 2 74LS00 gates. Because the largest load current occurs when the output of the inverter is low, this condition will drive the design. Determine the value of k_n such that $V_{oL} = 0.5$ volts and $I_{OL} = 0.8$ mA with two 74LS00 loads. Then assuming $k_p = k_n$, determine V_{oH} with 2 74LS00 loads ($I_{inH} = 0.02$ mA for each 74LS00). Assume $V_T = 1.8$ volts, power supplies of 5 volts, and the input signals to the inverter are 0 and 5 volts.

2. Using the k values from the first problem, determine the maximum current that occurs in the CMOS inverter during the switching transition. Make this calculation with no load.

3. Draw a compatibility chart showing how many loads of each kind each of the following gates can drive: 74LS00, CD4001, 74HC00, 74HCT00. Assume all are operating at 4.5 or 5 volts, whichever is appropriate.

4. A CMOS inverter gate is made with identical n and p-channel parameters, $V_T = 1.5$ volts and $k = 1$ mA/V^2. This gate is used to drive an LED at a curent of 20 mA with a voltage drop of 1.6 volts. The input voltages to the gate are 0 and 5 volts and the power supply voltage is 5 volts. What value of resistor should be used and what is the drain-source voltage across the drive transistor?

5. A CMOS inverter gate is made with identical n and p device parameters, $V_T = 1.5$ volts and $k = 5$ mA/V^2. This gate is used to drive an LED at a curent of 20 mA with a voltage drop of 1.6 volts. The input voltages to the gate are 0 and 5 volts and the power supply voltage is 5 volts. What value of resistor should be used and what is the drain-source voltage across the drive transistor?

6. A CMOS inverter gate is made with identical n and p-channel parameters, $V_T = 1.5$ volts and $k = 1$ mA/V^2. This gate has a capacitive load of 100 pF and is driven with a 0-5 volt square wave at 10 MHz. Calculate the power dissipation in the gate. The input square wave switches rail-rail in 10 ns.

7. A CMOS inverter gate is made with identical n and p-channel parameters, $V_T = 1.5$ volts and $k = 5$ mA/V^2. This gate has a capacitive load of 100 pF and is driven with a 0-5 volt square wave at 10 MHz. Calculate the power dissipation in the gate.

8. A CMOS inverter has a 5 volt power supply. The transistors are identical with $k = 5$ mA/V^2 and $V_T = 1.00$ volt. If the input voltage is 3.00 volts, find the output voltage.

9. A CMOS inverter has $V_T = 1.0$ volts. The data sheet specifies that $V_{oH} = 4.3$ volts with an output current of -4 mA, when $V_{in} = 0$ volts, and $V_{CC} = 5.0$ volts. Determine the value of k, assuming both the n-channel and p-channel transistors are the same.

10. A symmetrical CMOS inverter has $V_T = 1.0$ volts, and $k = 1.00$ mA/V^2. Determine the average power dissipation if the input voltage has a rail-rail transition time of 1 ns and a frequency of 100 MHz. The inverter has a 100 pF load, and $V_{CC} = 5.00$ volts.

11. A CMOS inverter with a 5 volt power supply is connected to an external load. With the input voltage at 5 volts, the output voltage is 1.00 volts. At this point, the inverter is sinking 2.00 mA from the external load. What is the value of k for the n-channel MOSFET? $V_T = 1.00$ volts.

12. A CMOS Hex inverter (6 inverters on the same IC chip) is being used in the laboratory. The transistors are identical with $k = 4$ mA/V^2 and $V_T = 0.5$ volts. Unfortunately, the technician hooks up the 5-volt power supply but allows the inputs to float and a static voltage buildup causes the inverters to be biased into the active region. What is the maximum power that might be dissipated in the IC chip?

CHAPTER 3

BIPOLAR JUNCTION TRANSISTORS

A bipolar junction transistor, BJT, is a single piece of silicon with two back-to-back P-N junctions. However, it cannot be made with two independent back-to-back diodes. BJTs can be made either as PNP or as NPN. The circuit symbols and representations of their configuration are given below. They have three regions and three terminals, emitter, base, and collector represented by E, B, and C respectively. The difference in the circuit symbols is the direction of the arrow. As we shall see shortly, the direction of the arrow indicates the direction of the current in the emitter when the transistor is conducting normally. As a way to remember which is which, a former student explained to me that NPN stands for "Not Pointing iN".

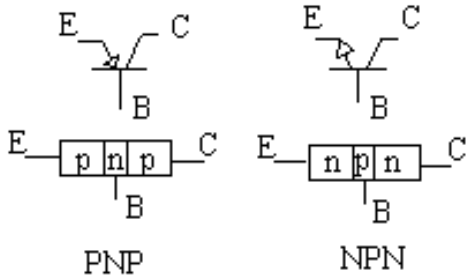


Figure 1. BJT symbols and representations

In digital electronics applications, we are interested in using the transistor as a switch and will concentrate on switching characteristics, rather than the linear properties. We will start with an overview of the operation of the transistor. Because most bipolar switching circuits use NPN transistors, we shall concentrate primarily on them.

There are four possible combinations, the base-collector junction may be either forward or reverse biased, and the base-emitter junction can also be biased either way. The four possibilities are shown in Figure 2 which also shows the operating region for each combination.

		B-E Junction: Reverse	B-E Junction: Forward
B-C Junction	Reverse	1 Cutoff	2 Forward Active
	Forward	4 Reverse Active	3 Saturation

Figure 2. The four operating conditions

The reverse active region is seldom used, but as we will see, it occurs in TTL gates. We begin our study of transistors by looking at each region of operation.

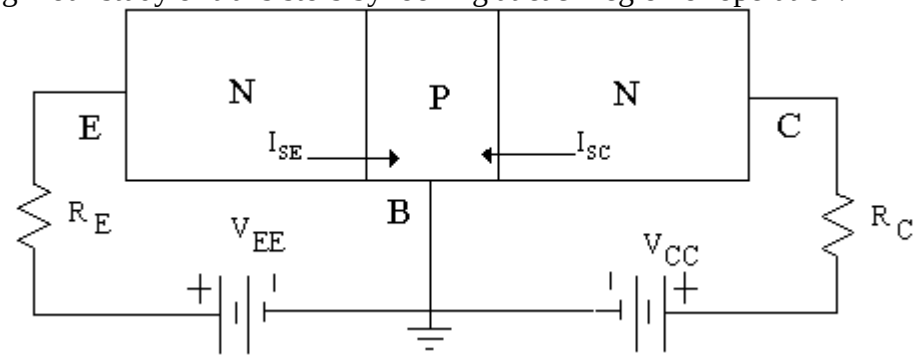


Figure 3. Common Base Configuration Showing Biasing Arrangement for Cutoff

CUTOFF REGION (Both junction reverse biased)

Let us start with the cutoff region, both junctions reverse biased as shown in Figure 3. With reverse biasing, we can assume that all currents are zero. We know that there are leakage currents associated with reverse biased junctions, but these currents are small and will be ignored.

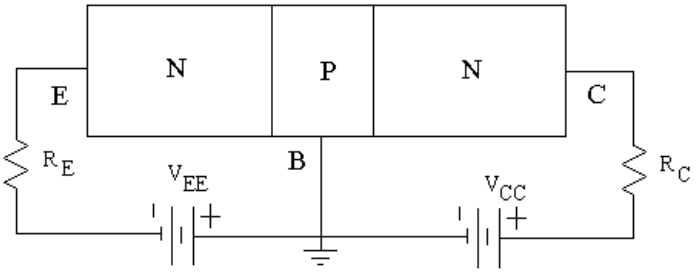


Figure 4. Common base configuration with biasing for the forward active region

FORWARD-ACTIVE REGION (BE junction forward biased, BC junction reverse biased)

The biasing condition for the forward active region of operation is shown in the Figure 4. The BE junction is forward biased and the BC junction is reverse biased. In this case, the forward bias of the BE junction will cause the injection of both holes and electrons across the junction. The holes are of little consequence because the doping levels are adjusted to minimize the hole current. The electrons are the carriers of interest. The electrons are injected into the base region where they are called the minority carrier even though they greatly outnumber the holes. They "pile up" at the BE junction. From there they diffuse across the base region due to the concentration gradient. Some are lost due to hole-electron recombination, but the majority reach the BC junction. At the BC junction, the electrons encounter a potential gradient (due to the depletion region) and are swept across the junction into the collector region.

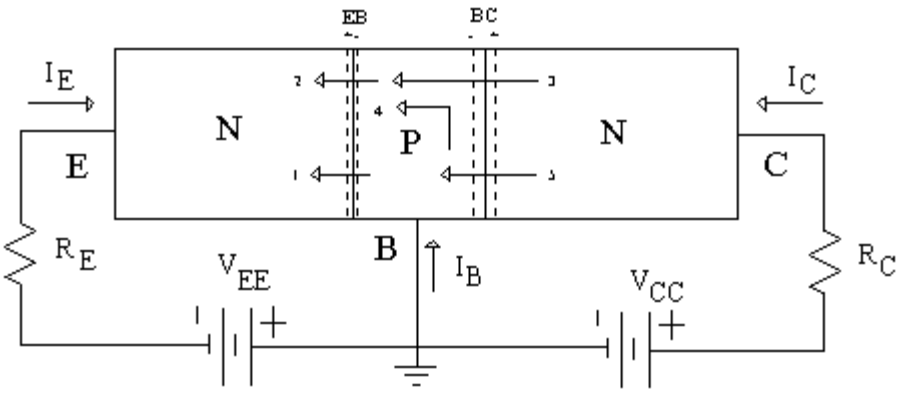


Figure 5. Common Base Configuration in forward Active region Showing Internal Currents

As shown in Figure 5, there are several components of current:

1. Holes injected from B-E. This is small and is ignored.
2. Electrons injected from base to emitter. This is $\approx -I_E$
3. Electrons that reach the collector. This is $-\alpha_F I_E$.
4. Recombination current. This is $-(1 - \alpha_F)I_E$.
5. Collector reverse saturation current I_{C0} , which we will usually neglect.

The term α_F is the forward current transfer ratio. This term refers to the fraction of electrons that reach the collector from those that are injected into the base region across the emitter junction. In most transistors, α_F is close to unity, typically 0.9-0.99. The difference between the current that is injected into the base region and that reaches the collector, becomes the base current. Thus, in most transistors, the base current is quite small, and the collector and emitter currents are close to the same magnitude.

The sum of currents entering the three terminals must equal zero,

$$I_B + I_C + I_E = 0 \qquad (1)$$

Also from current components 3 and 5 above,

$$I_C = -\alpha_F I_E + I_{C0} \qquad (2)$$

These two equations describe the currents in the common-base configuration in the forward-active region as shown in Fig. 5.

We can go back to the cutoff region from here by reducing the bias voltage between base and emitter. Since the BE junction is just a P-N junction, this current is essentially a diode current. Reducing this bias voltage to zero will reduce the emitter current to zero. In this case, $-I_B = I_C = I_{C0}$. We will hang on to this leakage current for a while yet before we completely drop it.

SATURATION REGION (Both junctions forward biased)

In the active region, the collector current is proportional to the emitter current (plus the leakage current, I_{C0}). This implies that the voltage bias across the base-collector junction is unimportant. But if we look again at Fig. 5, we see a resistor in series with the collector lead. If the current increases to a point that the voltage drop across the resistor plus the collector supply voltage begins to forward bias the collector junction, then holes will be injected into the collector region from the base. This hole current will counteract increases in electrons coming from the emitter, effectively limiting the transistor current. The base-collector voltage V_{BC} at which this limiting effect begins is at about V_{BC} =0.4 Volts, and becomes fully limiting at about 0.6 Volts. This region of operation is known as saturation. Please note that our terminology can get a little confusing. In this case, saturation refers to the circuit, not the transistor. The transistor could carry more current but the external circuit, the voltage source and resistor in the collector circuit, limits the current.

REVERSE ACTIVE REGION (BE junction reverse biased, BC junction forward biased)

In this case, the biasing arrangement is just the reverse of the forward active region. The collector junction is forward biased, while the emitter junction is reverse biased. The operation is just the same as the forward active region, except all voltage sources, and hence collector and emitter currents, are the reverse of the forward bias case. Another difference is that α_F is replaced by α_R . The equation corresponding to Eq. 2 is:

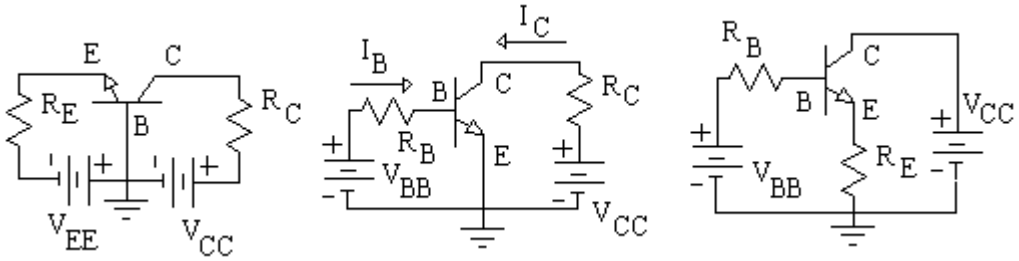
$$I_E = -\alpha_R I_C + I_{E0}$$

This configuration is rarely used because most transistors are doped selectively to give forward current transfer ratios very near unity, which automatically causes the reverse current transfer ratio to be very low.

NOTE
Please be aware that the above discussion and resulting equations are only approximations to the total operation of a transistor. A number of other phenomena are simultaneously occurring, but for typical applications, these phenomena cause only small errors in analysis. Such phenomena are: surface leakage currents, current due to holes injected from base into the emitter, hole-electron thermal generation, body resistance of the transistor, and avalanche multiplication. It is not our intention to study all effects in the transistor in this course, but to develop means to reduce the characteristics to simple, but appropriate models which will allow us to easily analyze transistor switching circuits. However, in the process of developing and using a model, an engineer must develop a feeling for when it is appropriate to use such a model and the limitations on the model.

COMMON BASE AMPLIFIER

Before we go on to the common emitter configuration, we should take a few moments to observe how a transistor can be used as an amplifier. Again referring to Fig. 5, assume that V_{EE} is increased by 1mV. This change in voltage would cause an increase in the current in the PN junction between the base and emitter. Most of this current would be reflected in an increase in the collector current. This change in collector current would manifest itself in a change in the voltage across the collector resistor, R_C. Now if the collector resistor is much larger than the emitter resistor, we would see a large voltage gain. We will not pursue this path further as this topic is the subject of another course.



a. Common Base *b. Common Emitter* *c. Common Collector*

Figure 6. Basic Configurations and Biasing for the NPN Transistors

COMMON EMITTER CONFIGURATION

The common-emitter configuration is much more prevalent than the common base configuration, especially in digital or switching circuits. This configuration is shown in Figure 6 along with the common base and common collector configurations for comparison. The name comes from the fact that the emitter terminal is held at ground or "common". If loop equations are written for both the input and the output circuits, the two currents of interest will be the collector current, I_C, and base current, I_B. Equation 2, above, is written in terms of the emitter and collector currents. By substituting Equation 1 into Equation 2, we can solve for the collector current in terms of the base current,

$$I_C = \frac{\alpha}{1-\alpha} I_B + \frac{1}{1-\alpha} I_{CO} \qquad (3)$$

Where we have dropped the subscript on α, since we are assuming forward active region. It is convenient to define a new variable,

$$\frac{\alpha}{1-\alpha} = \beta = h_{fe} \qquad (4)$$

Thus, Equation 3 becomes,

$$I_C = \beta I_B + \frac{1}{1-\alpha} I_{CO} \qquad (5)$$

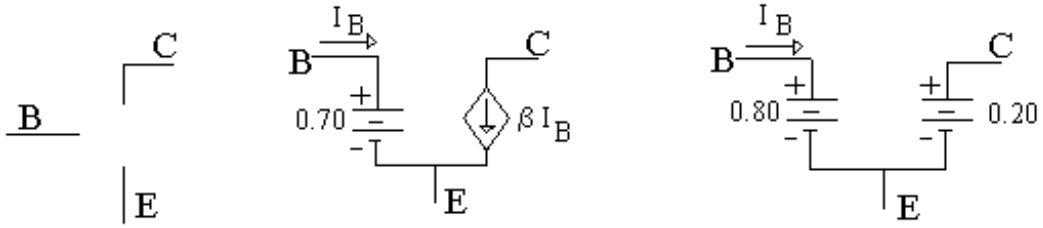
The terms β or h_{fe} relate a change in base current to a change in collector current, or current gain. If we ignore I_{CO}, then the dc current gain, h_{FE}, is the ratio of collector to base currents. β is used interchangeably for both ac and dc current gain. Many times we use β and h_{FE} interchangeably.

CIRCUIT MODELS FOR A COMMON EMITTER TRANSISTOR.

BASE-EMITTER CIRCUIT

When we go back and look at Figure 6, we can see that the relationship between base current and base-emitter voltage is controlled by the I-V characteristic of the base-emitter PN junction. While there is some effect caused by the collector-emitter voltage, these effects are quite small and for the purposes of this course, will be neglected. (The course in linear electronics does not neglect this effect, which is often expressed in terms of r_μ.) Thus, we can consider the base-emitter junction to operate just like a diode, and model it similarly. When talking about digital logic gates, we will assume that there will be no base current if the base-emitter voltage V_{BE} is less than $V_{BE\gamma} = 0.50$ Volts. We will also assume that if the base is driven heavily (into saturation), the base-emitter voltage V_{BE} is equal to $V_{BEsat} = 0.80$ Volts. Sometimes it is convenient to use an intermediate value of 0.7 volts when the transistor is in the active region. We will also use voltages of 50 mV higher for the ECL gates which are discussed later.

Figure 4 shows the circuit models we will use for the three regions of operation for the NPN transistor in the common emitter configuration. The base-emitter voltages are shown as discussed. Next we will. develop the collector models.



a. Cutoff
$V_{BE} < 0.50$
$(V_{BC} < 0.50)$

b. Active
$I_B > 0$
$V_{CE} > 0.20$

c. Saturation
$I_B > 0$
$I_C \leq \beta I_B$

Figure 7. Common Emitter Circuit Models for an NPN Transistor

COLLECTOR EMITTER CIRCUIT MODELS

The collector emitter circuit is more complex to model. Again we will talk about the three regions of operation. The collector characteristic is a plot of I_C vs V_{CE} for a few representative values of base current. Typical collector characteristics are given in Figures 8 and 9; the forward characteristic in Figure 8 and the reverse characteristic in Figure 9. Because the reverse active region is seldom used, we will not develop that model here.

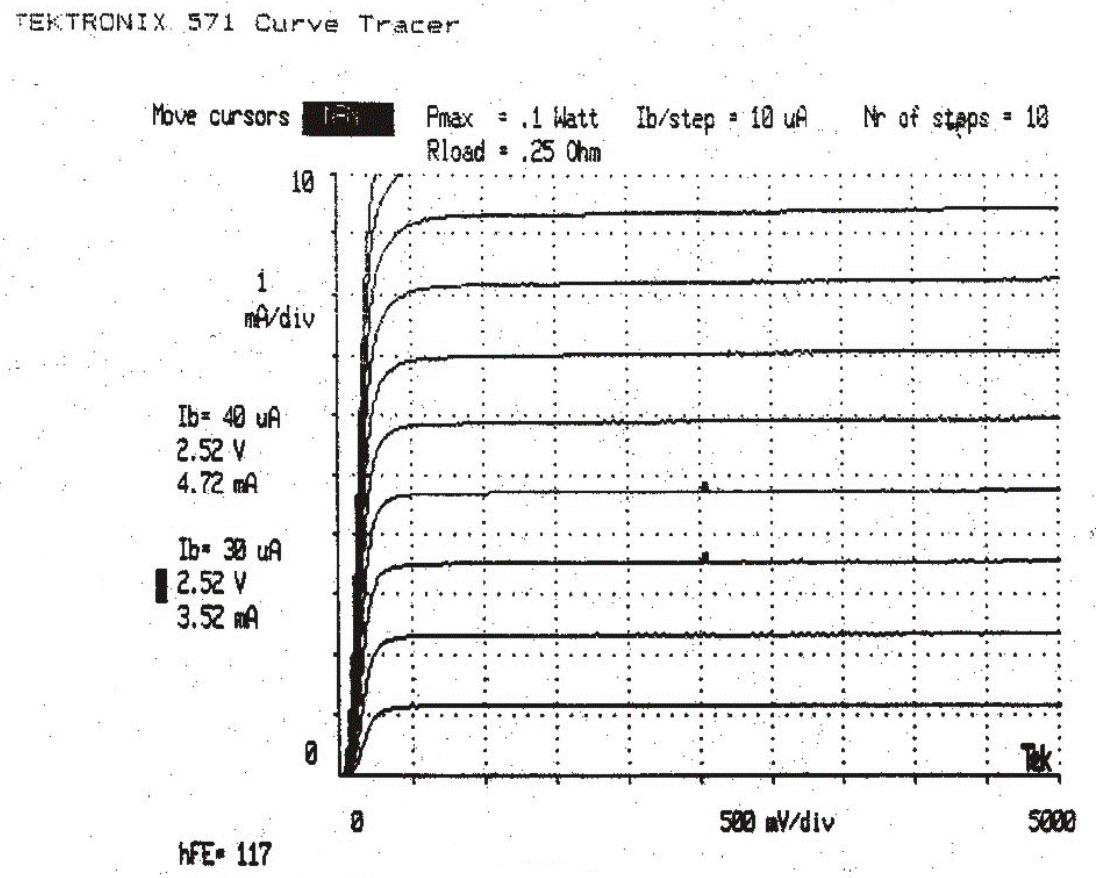


Figure 8. Forward collector characteristics for a 2N3903

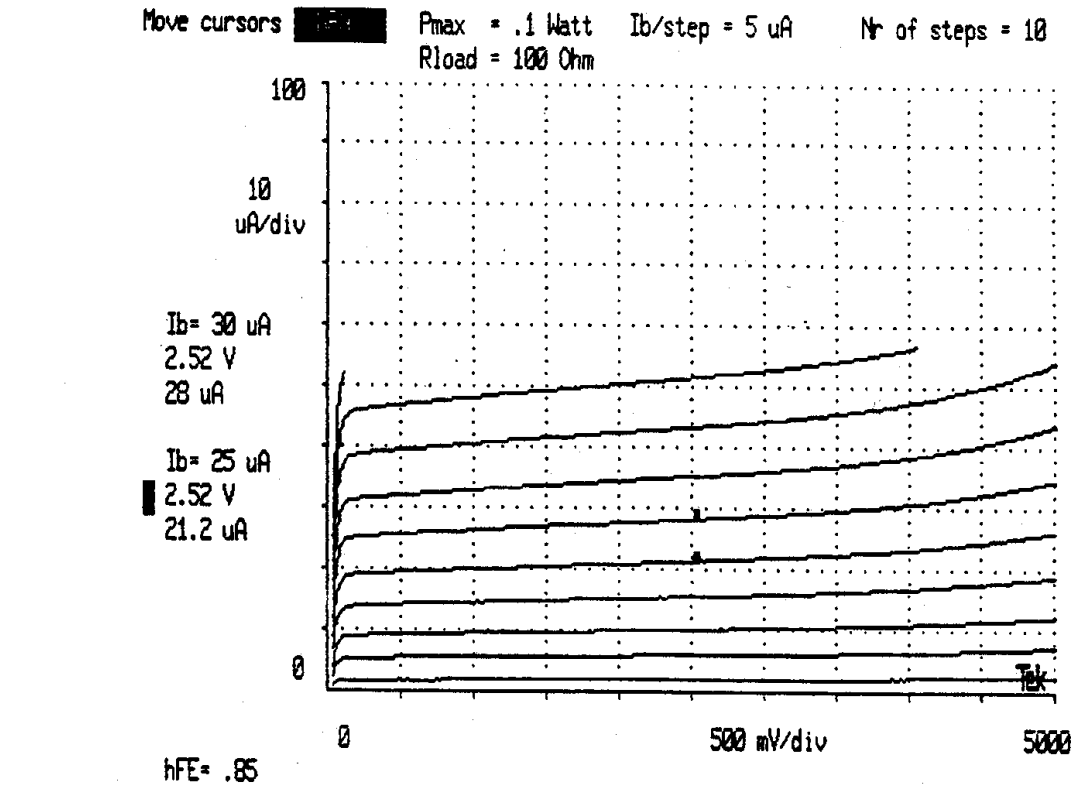


Figure 9. Reverse characteristics for a 2N3903

Cutoff

If the base-emitter voltage is less than cut-in ,$V_{BE\gamma}$, the base current is zero. From Equation 3, we see that the only current is the leakage current which we usually consider negligible. Thus, we can consider the transistor in cutoff to be an open circuit between the three terminals as shown in Figure 7a. The limit is, of course, that $V_{BE} < 0.50$ Volts.

Active Region

The discussion of the active and saturation regions rests on Equation 5 and the collector characteristic graph which is shown in Figure 8. The cutoff region is shown on the collector characteristic as the single line where the base current is zero; in this case the x-axis or $I_C = 0$. The active region is the large middle part of the family of curves where the collector current is essentially constant for a fixed base current. The saturation region is the left-hand portion of the curves where the curves are almost vertical.

The curves in the active region correspond closely to Equation 5. The current is only a function of base current, not a function of collector-emitter voltage, V_{CE} . In this case, a suitable model of the transistor is shown in Figure 7b. This figure shows the base-emitter model as a voltage source and the collector-emitter model as a current-controlled current source, βI_B . The defining limitations are $I_B > 0$ and $V_{CE} > 0.2$ V.

Again, you should be aware that this model is only an approximation to the real device. For instance, collector current does vary slightly with V_{CE} and the spacing between the curves is not exactly constant; β varies somewhat with collector current. The course in linear electronics more closely models the first of these effects with an additional parameter, r_o .

Saturation Region

The saturation region model is perhaps the poorest approximation to the real device. The saturation region is modeled simply as a fixed voltage of $V_{CE} = 0.2$ Volts. In this region of operation, the base circuit is modeled as $V_{BE} = 0.80$ Volts. This model is shown in Figure 4c. The limitations are that $I_B > 0$ and $I_C < \beta I_B$.

From the curves in Figure 8, there is a substantial portion of the curve between the straight-line portions in the linear region to the nearly vertical lines in the saturation region. Probably the best way to view this transition is that β is decreasing. A typical approach is to assume that when one is discussing saturation, the appropriate β is approximately 85% that for the linear region. In this course, we will consider this problem in two ways. First, when we are working with problems in the early part of the course where we are considering operation in all three regions, we will assume β to be the linear β for all regions. Here we will assume that saturation is reached abruptly. In the

second case, when we are considering only cutoff and saturation regions, we will use a saturation β only. We will simply ignore the linear region. The models we use in both cases are the same and should cause little difficulty.

CONCLUSION

The effect of using the models for the three regions can be seen in Figure 10. In this figure, the cutoff region is represented by the straight line on the voltage axis, indicating zero current. The linear region is represented by a series of evenly spaced horizontal lines. The saturation region is represented by the vertical line at V_{CE} =0.2 Volts. A comparison between Figures 8 and 10 will show that our models are only approximations to the actual characteristic. These approximations will, however, allow us to greatly simplify circuit analysis for the switching circuits in this course. These approximations will give us reasonably accurate solutions to these circuits. They will also be a great help to us when we attempt to understand circuit operation.

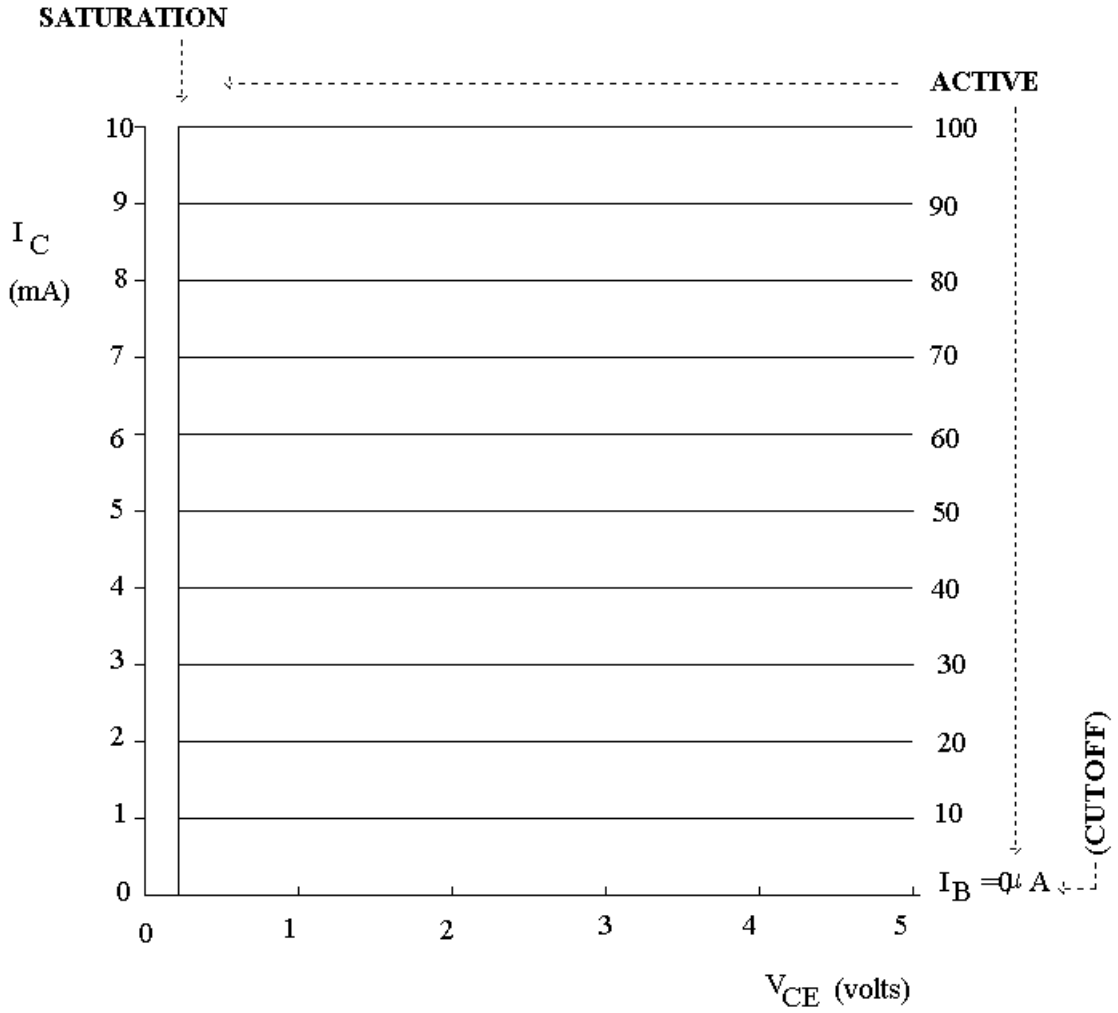


Figure 10. A composite characteristic represented by the circuit models.

THE TRANSISTOR AS AN INVERTER/SWITCH

The utility of a BJT in digital circuits is the ability of the transistor to block or conduct current with just a small control current. Thus, we are primarily interested in the cutoff and saturation modes of operation. In this section, we will discuss the transistor used both as a switch and as a linear inverter. We will use the circuit models of the transistor developed in the previous section.

To start the discussion, we will consider the circuit in Figure 11. This circuit is the classic switch. The operation of the transistor is controlled by the current in the base circuit. Thus, the input voltage controls the circuit. We will analyze the circuit for three cases: V_{in} =0, 5, and 10 Volts.

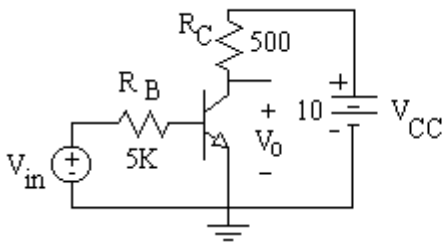
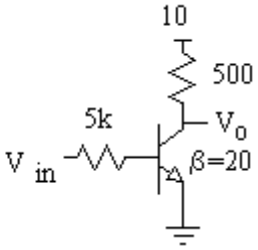


Figure 11. Transistor Switch.



Simplified Drawing.

$V_{in} = 0$

Because the base-emitter junction is a diode, this part of the circuit can be analyzed as we did earlier for diode circuits. With no source voltage to overcome the turn-on voltage of the diode, there will be no current flow. With no base current, the transistor is cutoff and there will be no collector current. See the circuit in Figure 12 where the transistor has been replaced with its cutoff model. With no current in the collector circuit, there will be zero voltage drop across the collector resistor. Hence the voltage between the collector and emitter will be

$$V_{CE} = V_O = V_{CC} = 10 \text{ Volts.} \qquad (1)$$

(Note the order of the subscripts, CE. A positive voltage for V_{CE} means that the collector is more positive than the emitter.)

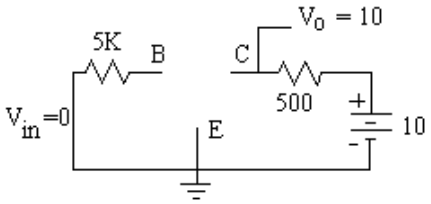


Figure 12. Circuit With V_{in} = 0. The transistor is replaced with its cutoff model.

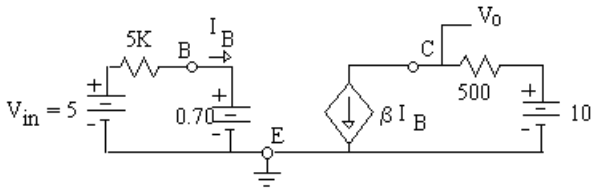


Figure 13. Circuit With V_{in} = 5 Volts

V_{in} = 5 Volts

An analysis of the base circuit in this case indicates that the base current is not zero. From Figure 13,

$$I_B = \frac{V_{in} - V_{BE}}{R_B} = \frac{5 - 0.7}{5K} = 0.86 \text{ mA} \qquad (2)$$

Since there is base current, there must be collector current. If we assume the transistor is in the active region, the active circuit model has replaced the transistor in Figure 13. We can solve this circuit for the collector current.

$$I_C = \beta I_B = 20 * 0.86 = 17.2 \text{ mA} \qquad (3)$$

The final information we would like to know is the output voltage, the voltage at the collector . We cannot get V_{CE} directly, we have to use the voltage drop across the collector resistor;

$$V_O = V_{CE} = V_{CC} - I_C R_C = 10 - 17.2 \text{ mA} * 0.500 \text{ K}\Omega = 1.4 \text{ V} \quad (4)$$

Since $V_{CE} > V_{CEsat}$ (=0.2V), this result is consistent with the assumption that the transistor is operating in the active region.

Now let us look at the final case.

V_{in} = 10 Volts

If we assume the transistor will be in the active region, the process will be the same as for the previous case. We can look at Figure 13 but with the input voltage at 10 volts.

$$I_B = \frac{V_{in} - V_{BE}}{R_B} = \frac{10 - 0.7}{5K} = 1.86 \text{ mA} \qquad (5)$$

Proceeding to the collector circuit, we will attempt to find the collector current as we did before;

$$I_C = \beta I_B = 20 * 1.86 = 37.2 \text{ mA} \qquad (6)$$

The output voltage is then,

$$V_O = V_{CC} - I\,R = 10 - 37.2 * .500 = -8.6 \text{ Volts ??} \quad (7)$$

This result says that the output voltage is negative. How can that possibly be? There is no source for the negative voltage; no negative power supply. The transistor is modeled as a current-controlled current source, but is not a current generator. It can only work within the limits of the power supplies. You will note on the collector characteristic curves that the collector voltage cannot go negative with a positive collector current. The obvious conclusion is that the transistor circuit has saturated, and the active region model is no longer valid.

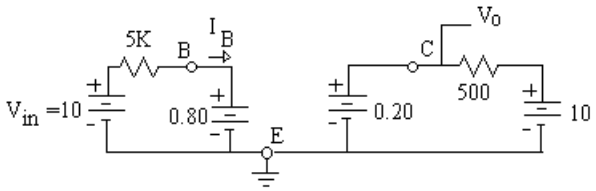


Figure 14. Circuit in Saturation

If we go to the saturation region model, the circuit is shown in Figure 14. Note that the only base circuit change is to change V_{BE} to 0.80 volts. This change makes the base current

$$I_B = \frac{10 - 0.80}{5K} = 1.84 \text{ mA} \quad (8)$$

We can also determine the collector current,

$$I_C = \frac{V_{CC} - V_{CEsat}}{R_C} = \frac{10 - 0.20}{0.5K} = 19.6 \text{ mA} \quad (9)$$

We already know that the output voltage is 0.2 Volts ($=V_{CEsat}$) since the transistor is in saturation. The only thing left is to verify that the model used is appropriate for the situation. (We already know the active region model is inappropriate, and that we are not in cutoff, $I_B > 0$.) To demonstrate that the saturation model is appropriate, we need only show the $I_{Csat} < \beta I_B$. Since $\beta I_B = 36.8$ mA > 19.6, the saturation model is appropriate.

SUMMARY

It is appropriate at this time to explicitly define the criteria for operation in each mode.

Cutoff: $V_{BE} < V_{BE\gamma}$ and $V_{BC} < V_{BC\gamma}$
If this condition occurs, the base current will be zero and the collector current will be zero. Of course, we are ignoring the leakage currents. We are also not considering the possibility of using the transistor backwards with collector and emitter reversed. Thus, we are assuming that the collector voltage is more positive than the base voltage. We will discuss the value of $V_{BE\gamma}$ later.

Active Region $I_B > 0$, $V_{BE} = 0.70$, $I_C = \beta I_B$, $V_{CE} > 0.2$ V.

Saturation Region: $I_B > 0$, $V_{BE} = 0.80$, $V_{CE} = 0.2$, $I_C < \beta I_B$.

The value of $V_{BE\gamma}$ we choose will have a significant bearing on how we view the transition between cutoff and the active region. For logic systems, we will use $V_{BE\gamma}$ = 0.50 Volts. This is the value the base-emitter voltage must be less than to guarantee the transistor is cutoff. Note that when the transistor is saturated, V_{BE} =0.80. For logic systems, any value in between is indeterminate. This inconsistency will not cause us any trouble in logic systems. However, this discrepancy will cause some difficulty when we are trying to find the transition between cutoff and active operation and between active and saturation operation.

For our next example, we will discuss a transistor as an inverting amplifier that varies continuously between cutoff and saturation.

INVERTING AMPLIFIER

To get a better feel for how a transistor works with real signals, we will discuss a transistor inverter with a sine wave input. The transistor will operate in all three modes. The circuit is shown in Figure 15. Temporarily, we will assume

$$V_{BE\gamma} = 0.75 = V_{BEact} = V_{BEsat}$$

In other words, there is no discontinuity from cutoff to the active region, and no discontinuity from the active region to saturation. Where do we start the analysis of such a circuit? First let us consider the possibilities. At any given instant of time, the transistor will either be cutoff, in active region, or saturated. From our previous example, we should note that if the input voltage is below a certain value, the transistor will be cutoff; above that value, the transistor will be in the active region with positive base current, and with the input voltage higher still, we get enough base current to saturate the transistor. Our job here is to determine which region or regions are utilized. First, let us look to see if the transistor is cutoff.

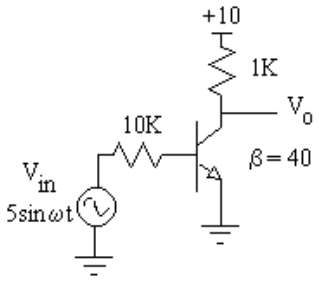


Figure 15. Transistor Inverting Amplifier

Model:
$V_{BE\gamma} = V_{BEactive} = V_{BEsat}$ =0.75

V_{CEsat} = 0.2

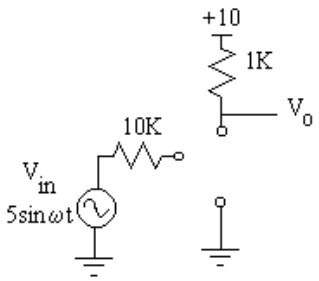
β = 40

Cutoff

In order for the transistor to conduct at all, the input voltage must be above $V_{BE\gamma}$. Conversely, if the input voltage is below $V_{BE\gamma}$, the transistor will be cutoff. Does that occur here? Yes, the input voltage may go as low as -5 Volts. Thus, for all the time the input voltage is below V_{BE} =0.75 Volts, the transistor is cutoff. The transistor is replaced by its cutoff model in Figure 16. What then is the output voltage during this time? Since there is zero collector current, zero current through the collector resistor, there is zero voltage drop across the resistor, and the output voltage is

$$V_O = V_{CC} = 10 \text{ Volts.} \qquad (10)$$

Figure 16 shows the circuit with the transistor replaced with its cutoff model and Figure 17 shows the input and output waveforms during the cutoff period.



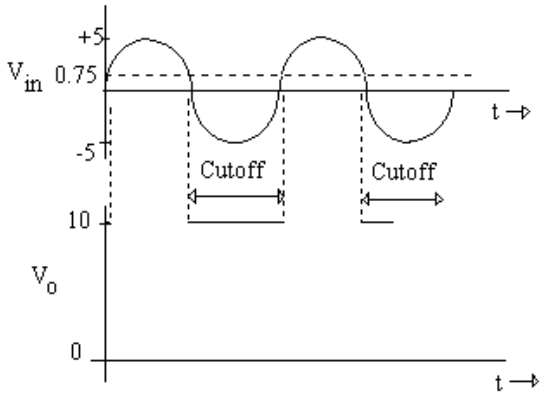


Figure 16. Circuit With Transistor Cutoff *Figure 17. V_O When Transistor is cutoff*

Active Region

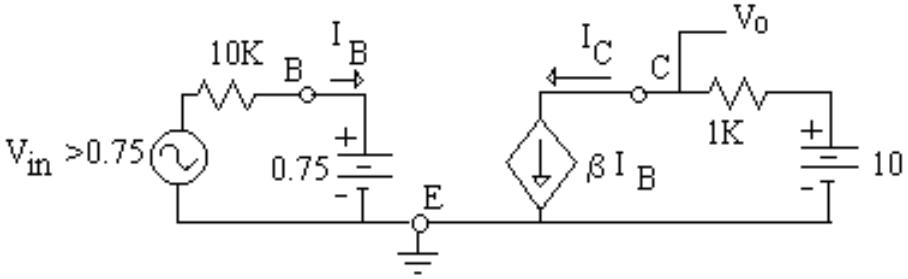


Figure 18. Transistor Inverter Circuit During Active Region

If the input voltage goes above 0.75 V, the transistor goes into the active region. If we use the active model for the transistor and place it into the circuit as shown in Figure 18, we can then analyze the circuit. The base current is

$$I_B = \frac{V_{in} - V_{BE}}{R_B} = \frac{5\sin\omega t - 0.75}{10K} \qquad (11)$$

Thus, the base current will have a sinusoidal component like the input voltage, only shifted. Note that this equation is valid only if $V_{in} > V_{BE\gamma}$. The next question, of course,

is what is the output voltage? Since the collector current is $\beta I_B = 40 I_B$, the output voltage is the supply voltage minus the drop across the collector resistor.

$$V_o = V_{CC} - I_C R_C = 10 - 40 \times 1K \times \frac{5\sin\omega t - 0.75}{10K} \qquad (12)$$

Again, the output voltage has a sinusoidal component. This equation is valid only for V_{in} >0.75 Volts, and also as long as the circuit is not saturated, $V_o = V_{CE}$ >0.2 v. Thus, the output voltage is sinusoidal only for a portion of the cycle. It should also be noted that if the input voltage = 0.75 Volts, the output is equal to 10 Volts, which is consistent with the cutoff case, no discontinuity.

We now need to determine if the circuit saturates.

Saturation

We note from our earlier example that the circuit will be in saturation if the collector-emitter voltage drops to 0.2 Volts. We could solve Equation 12 to find what input voltage would cause this to happen. However, we will take a more straightforward approach. Let us simply look at the saturation model for the circuit to determine what the input voltage must be in order to reach saturation.

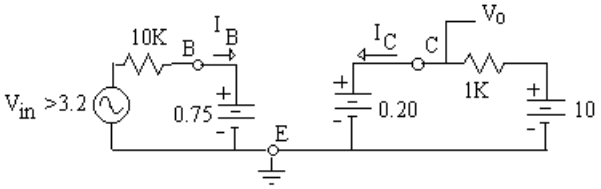


Figure 19. Transistor Inverter Circuit in Saturation

The saturation model for the circuit is shown in Figure 19. We now work backwards from the collector circuit. In Figure 19, the collector current is

$$I_{Csat} = \frac{V_{CC} - V_{CEsat}}{R_C} = \frac{10 - 0.20}{1K} = 9.8 \text{ mA} \qquad (13)$$

The minimum base current required to support this collector current is

$$I_{Bsat} = \frac{I_{Csat}}{\beta} = \frac{9.8}{40} = 0.245 \text{ ma} \qquad (14)$$

The minimum input voltage required to furnish this much base current is

$$V_{insatmin} = V_{BEsat} + I_{Bsat} * R_B = 0.75 + 0.245 \text{ mA} * 10 \text{ K}\Omega$$

$$= 3.2 \text{ Volts.} \qquad (15)$$

Since the input voltage rises above this value, the circuit does indeed saturate and will be in saturation whenever V_{in}>3.2 volts. In saturation, the output voltage is 0.2 Volts. Thus, Figure 20 shows the output voltage for the entire cycle.

Figure 20 shows that the circuit spends relatively little time in the active region. However, during this time the output voltage has a sinusoidal component and that sinusoidal part has a negative sign with respect to the input voltage, hence, the name of inverter. Actually this circuit is not really suitable for a linear signal inverter without substantial modifications. The linear inverting amplifier is discussed in more detail in the linear electronics course. It is useful to note, however, that the circuit we have analyzed does "square up" a sine wave, making the signal useful for digital switching.

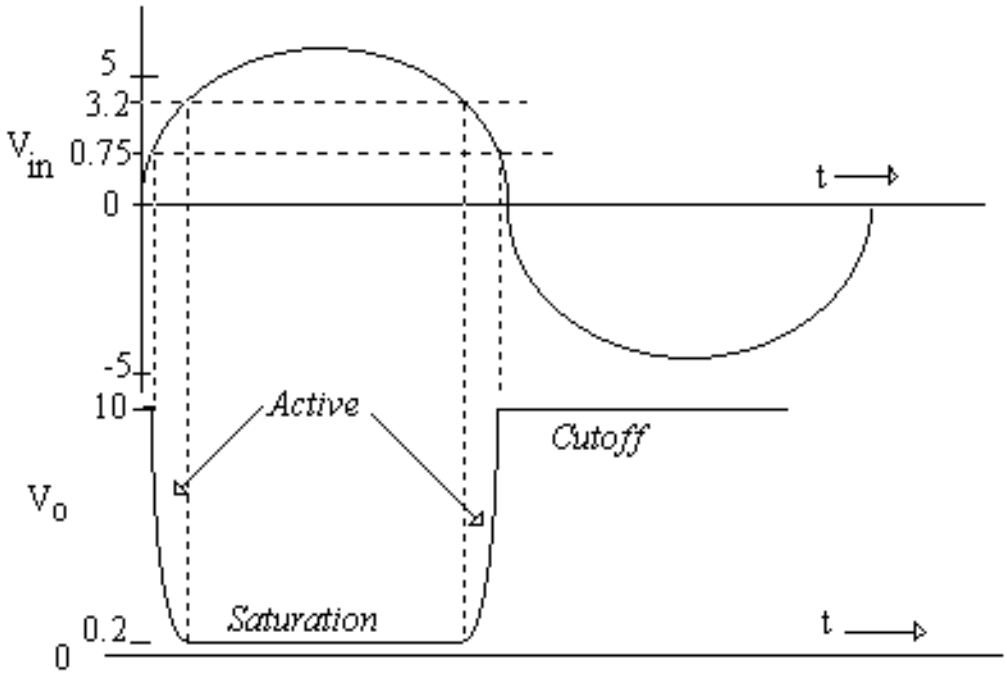


Figure 20. Waveforms for a complete cycle for the inverting amplifier circuit.

Voltage Transfer Characteristic

It is often useful to plot the relationship between the output voltage and the input voltage, called the voltage transfer characteristic or VTC. For the previous example, we already know the transistor is cutoff for $V_{in} < 0.75$ making $V_{OCutoff} = 10$ and the transistor is in the active region when $0.75 < V_{in} < 3.2$. We also know the transistor is in saturation when $V_{in} > 3.2$ making the $V_{Osat} = 0.2$. In active region:

$$V_{Oact} = 10 - 4(V_{in} - 0.75) = 13 - 4V_{in}$$

which is an equation for a straight line. The voltage transfer characteristic curve is shown in Figure 21.

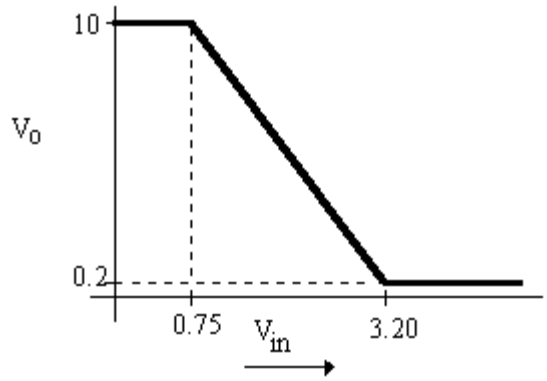


Figure 21. Voltage transfer characteristic for the transistor inverter

PNP Transistors

The PNP transistor works the same as the NPN, except that the carriers are holes instead of electrons, and all the currents and voltages are in the opposite direction.

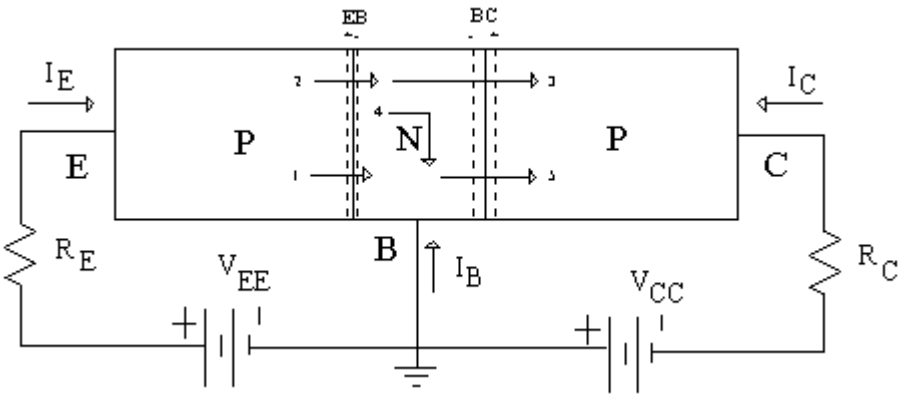


Figure 22. PNP transistor with internal currents.

Figure 22 is a drawing of a PNP transistor showing the internal and external currents. The internal currents correspond to the same currents as for the NPN transistor shown earlier. The currents of importance are numbered 2, 3, and 4, hole currents, injected from emitter to base(#2), the diffusion current that reaches the collector (#3), and the holes lost by recombination in the base region (#4). Currents #1, and #5, are parasitic and usually ignored.

Because the external currents are considered positive going into a terminal by convention,

I_E = current #2

I_C = - current #3 = $-\alpha I_E$

I_B = - current #4 = $-(1-\alpha)I_E$

The above equations relate the base and collector curents to the emitter currents as is usual for the common base configuration. However, because we usually use the transistors in the common emitter configuration, it is more convenient to reperesent the currents in terms of the base curents. We still get

$I_C = \beta I_B$

but now, I_B is a negative quantity. We also still have

$I_C + I_B + I_E = 0$

making

$I_E = I_B + I_C$

Now however, I_E will be a positive quantity, and I_B and I_C are negative quantities.

The currents in the PNP transistor are in the opposite direction as for the NPN transistor, making the voltage drops across the junctions in the opposite direction. Thus, for the PNP transistor operating in the active or saturated regions,V_{BE} and V_{CE} are negative values.

A typical common emitter configuration is given in Figure 23. From the circuit model,

$I_B = -1.3/5K = 0.26$ mA

The collector current is ($\beta = 10$)

$I_C = -0.26\beta = -2.6$ mA

and the voltage at the collector is

$V_O = -5 + 1K \times 2.6$ mA $= -2.4$ volts.

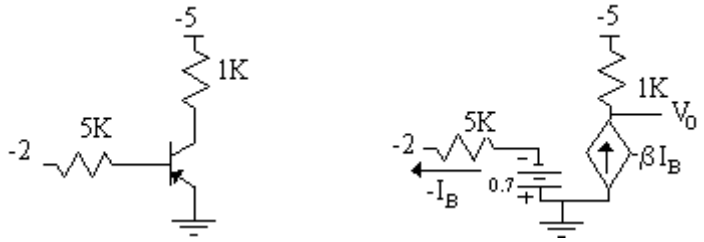


Figure 23. Typical common emitter application for a PNP transistor with its active model.

The difficulty in dealing with the PNP transistor is keeping track of the negative currents. Because most logic systems use positive voltages and because it is more intuitive to have currents going downhill instead of going uphill as in the above example, circuits can be drawn as shown in the application in Figure 24 as an output interface in a logic system. .

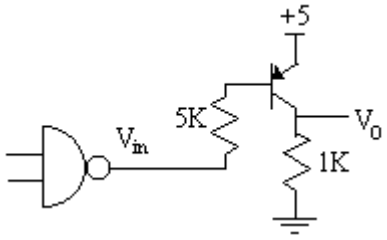


Figure 24. PNP application in a logic system interface

In this case, we will assume the logic gate output is either +5 volts, or 0 volts. (Later on, we will use more accurate data for the gate.) If the gate output is high at +5, the voltage across the base-emitter junction is zero volts, which will keep the transistor cutoff. The circuit with its cutoff model is shown in Figure 25. Note that we specify the emitter-base voltage as V_{EB}. In this case, $V_{EB} = 0$, because both terminals are connected to +5 volts.

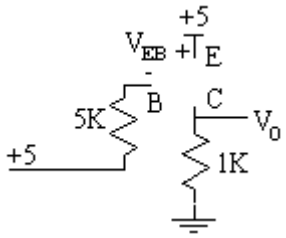


Figure 25. Cutoff model

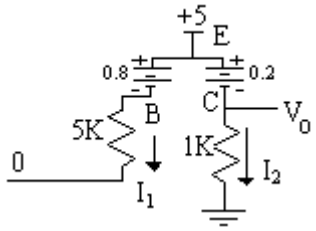


Figure 26. Saturation model

If the gate output is low, or zero volts, the emitter-base junction is turned on. We suspect the transistor is saturated, so we use the saturation model in Figure 26 for the analysis. It is convenient to use positive currents I_1 (= - I_B) and I_2 (= - I_C) instead of their negative counterparts. Note again, we use reverse polarities for the the two terminal voltages. We can calculate the two currents

$I_1 = (5\text{-}0.8)/5K = 0.84$ mA

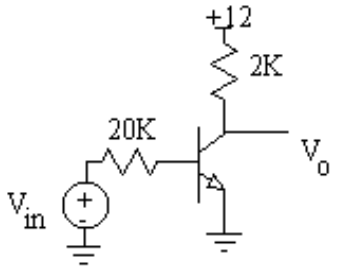
$I_2 = (5\text{-}0.2)/1K = 4.8$ mA

If $\beta = 10$, the transistor is obviously in saturation, and $V_O = 4.8$ volts. Notice that the voltage across the 1K load resistor is 4.8 volts.

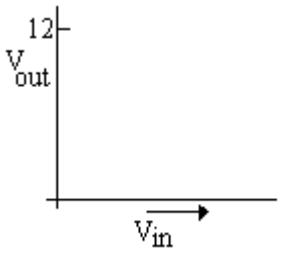
EXERCISES

1. For the following circuit, assume the following model for the transistor:

$V_{BE\gamma} = 0.5$ v, $V_{BEact} = 0.7$ v, $V_{BEsat} = 0.8$ v, $\beta = 30$, $V_{CEsat} = 0.2$ v.



Circuit Diagram



Voltage Transfer Characteristic

If $V_{in} = 5.0$ volts,

a. Draw arrows showing the actual current direction in each terminal of the transistor.

b. Determine three terminal currents: $I_B =$ _____

$I_C =$ _____

$I_E =$ _____

c. Determine V_o. $V_o =$ _____

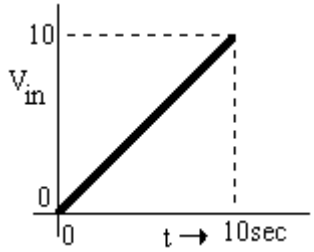
d. Find minimumV_{insat} and maximum $V_{incutoff}$. We can use these points to draw the voltage transfer characteristic, but because we are using a discontinuous transistor model, we cannot precisely predict the region in between. Simply draw a dotted line between the two points.

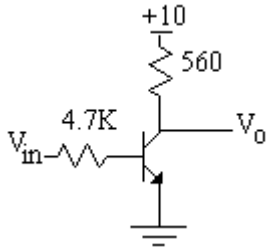
2. A ramp voltage is applied to the inverter circuit shown. A continuous model for the transistor is:

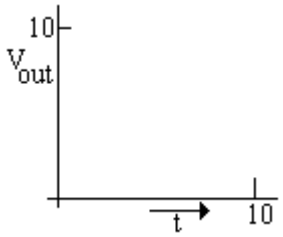
$V_{BE\gamma} = V_{BEact} = V_{BEsat} = 0.7$ volts

$V_{CEsat} = 0.2$ volts $\beta = 25$

Plot the output voltage waveform and determine the time it takes the output voltage to go from cutoff to saturation.







3. A PNP transistor circuit is shown below. The transistor model is:

$V_{BE\gamma} = V_{BEact} = V_{BEsat} = -0.70, \quad V_{CEsat} = -0.20, \quad \beta = 20$

If V_{in} = 5.0 volts,

a. Draw arrows showing the actual current direction in each terminal of the transistor.

b. Determine three terminal currents: I_B = _____

I_C = _____

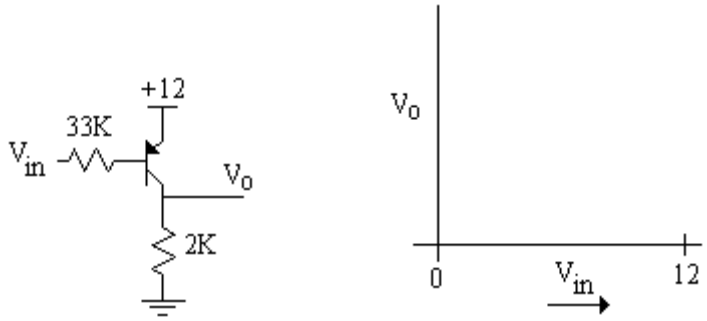
I_E = _____

c. Determine V_o. V_o = _____

d. Find V_{insat} and $V_{incutoff}$.

V_{insat} = _____ $V_{incutoff}$ = _____

e. Draw the voltage transfer characteristic.



DESIGN EXAMPLE: TRANSISTOR SWITCH

To illustrate the utility of a transistor as a switch, let's look at the following example. We have a signal from a 74HC00 CMOS logic gate that is supposed to turn on an LED when the output of the gate is a high. The specifications of interest are:

74HC00: V_{OHmin} = 3.84 volts when I_{OH} = -4.00 mA,
V_{OL} = 0 at I_{OL} = 0

LED: V_D= 1.80 volts when I_D =30 mA, I_{max}= 60 mA

Because we want the LED current to be about 30 mA, the logic gate clearly will not supply the necessary current. A BJT switch is a good candidate for this job.

A suggested circuit is shown below. In this case, when the output of the gate is high, we want the transistor to saturate with 30 mA through the diode. When the output of the gate is low, the transistor will be turned off. The high output case is shown.

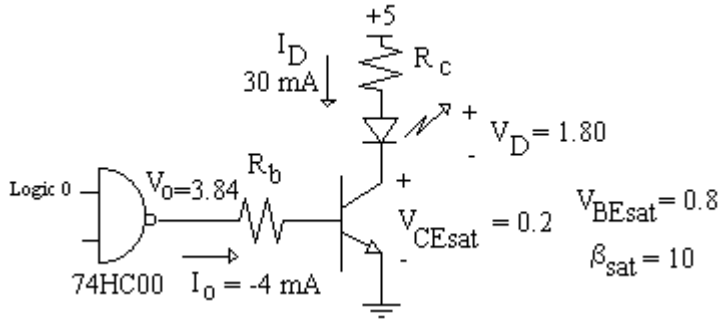


Figure 27. Circuit to be designed

First, let's look at the case when the gate output is low. In this case, the output voltage is zero and the transistor will be cutoff. There will also be no diode current.

When the gate output is high as shown in the circuit above, we want the transistor to be saturated. If the diode current is 30 mA, and the diode voltage is 1.80 volts, the voltage across the series (current limiting) resistor, R_C, is 3.0 volts. Thus, this resistor value must be 100 Ω. Assuming the transistor β_{sat} =10, the base current must be greater than 3 mA to guarantee saturation. In this case, the gate can source 4 mA when the output voltage drops as low as 3.84 volts. If the base-emitter voltage is 0.8 volts, the voltage across the resistor must be 3.04 volts. At a base current of 4 mA, the resistor value is 760 Ω.

Note that the base current in saturation needs only to be 3 mA or greater. Thus, the base resistor can be made larger, up to 1013 Ω. The standard 5% values between these limits are 820, 910, and 1000 Ω. Any of these values is satisfactory if we consider only nominal calculations. (If we must guarantee operation with 5% resistors, $800 < R_{NOM} < 964$.)

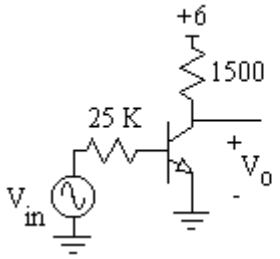
With this design, the LED will be off when the gate output is low and the LED will be on when the gate output is high.

Problems

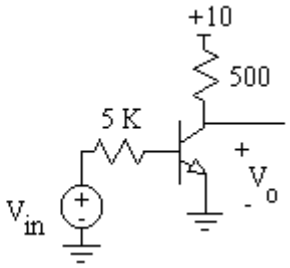
1. In the laboratory, take a set of transistor characteristic curves for a transistor in the forward active region. From these curves, develop the approximate, straight-line models for the three regions of operation. Present your results by drawing the circuit models with component values and by drawing the characteristic "curves" represented by these circuit models. Turn in the original collector characteristic curves, the circuit models, and the "curves" represented by these models.

2. For the circuit below, determine the values of the input voltages that will cause the transistor to be cutoff and saturated. Use the following model component values:

β=50, $V_{BE\gamma}$ =0.50, V_{BEsat} = 0.80, V_{CEsat} = 0.20.



Problem 2



Problems 3 and 4

For Problems 3 and 4, use the following model for the transistor:

β = 30, $V_{BE\gamma} = V_{BEact} = V_{BEsat}$ = 0.70, V_{CEsat} = 0.20

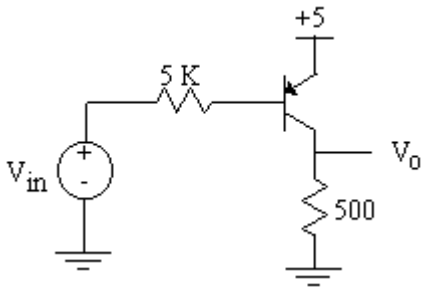
3. Draw the voltage transfer characteristic (V_o vs V_{in}) for the circuit given. Let V_{in} go high enough for the circuit to saturate.

4. Let the input voltage be a 5 volt sine wave (5sinωt). Sketch the output waveform showing all breakpoints in the waveform and the input voltage at each breakpoint. Determine the time it takes the output waveform go from one "rail" to the other if the frequency is 60 Hz.

5. Use a computer simulation to verify the results from Problem #4. Use the library model for a 2N3903 with Beta = 30. Note that the results may not be quite the same because the PSPICE library model is different from the 3-region model we use in class and that model includes capacitance. Be sure to show on your plots the time it takes to make the transition from cutoff to saturation.

6. Find the input voltage required to cause the pnp transistor circuit to saturate.

PNP Transistor Model: $V_{EBsat} = 0.8$, $V_{ECsat} = 0.2$, $Beta_{sat} = 20$

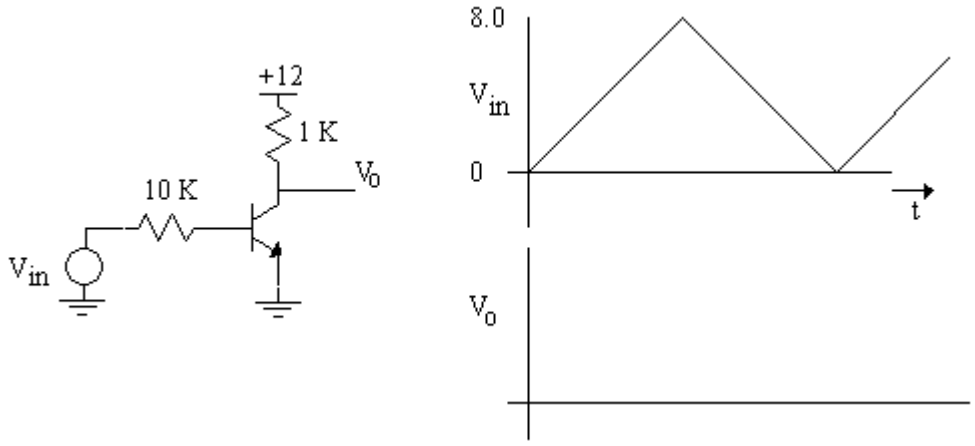


Problem 6

7. The input voltage waveform is shown. Draw the output voltage waveform showing the input and output voltages at each breakpoint.

The transistor has the following model parameters:

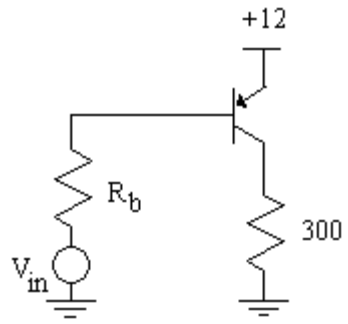
$V_{BEcutin} = V_{BEact} = V_{BEsat} = 0.7$ volts, $V_{CEsat} = 0.2$ volts, Beta = 25



8. A PNP transistor is used in the following circuit to drive a load as shown. Determine the value of the base resistor needed to guarantee saturation if the input voltage switches between +12 volts and 0.4 volts. Note that in one state, the pnp transistor will be cutoff and in the other state, the pnp transistor is to be saturated.

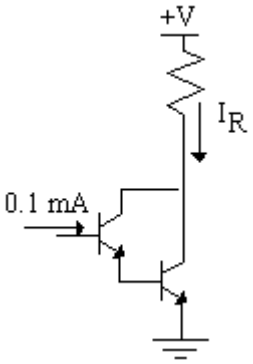
The pnp transistor has the following model parameters:

$V_{EBcutin} = 0.5$ volts, $V_{EBsat} = 0.8$ volts, $V_{ECsat} = 0.2$ volts, Beta = 20



Problem 8.

9. A Darlington connection is shown below. If 0.1 mA is going into the base of the first transistor, how much current can go through the resistor? Beta is 20 for each transistor. What is V_{CE} for the output transistor?



Chapter 4

Resistor-Transistor Logic

Resistor-transistor logic, RTL, is an old technology that is no longer commercially available. It is a useful starting place to study logic gates because it is a logical extension of the transistor inverter studied in the previous chapter, and because it is a useful circuit in itself.

A two-input gate is shown in Figure 1. As with most logic, the transistors operate only in saturation or cutoff. There are four possible combinations of logic inputs, L-L, L-H, H-L, and H-H for inputs A and B respectively. A low input will cutoff the associated transistor, and a high input will saturate it. Figure 2 shows these combinations where the transistors have been replaced by their corresponding models. If both transistors are cutoff, the output will be high, 3.0 Volts. If either transistor is saturated, the output will be 0.2 Volts, or low. The truth table for this gate is given in Figure 3 and shows that the gate performs a logic NOR function.

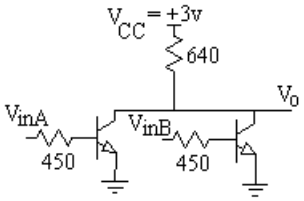
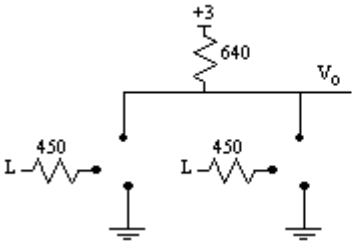


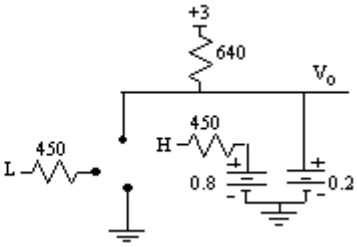
Figure 1. RTL gate

V_{inA}	V_{inB}	V_o
L	L	H
L	H	L
H	L	L
H	H	L

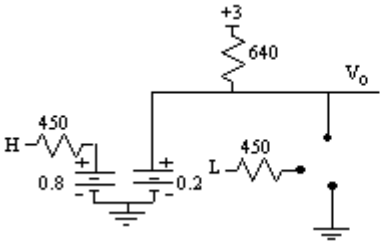
Figure 3. Truth table for RTL gate



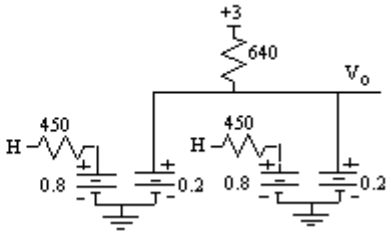
a. L-L inputs



b. L-H inputs



c. H-L inputs



d. H-H inputs

Figure 2. RTL gate with four combinations of inputs

Our job is to determine the eight terminal specifications, input and output voltages and currents when these terminals are both high and low as shown in Table 1. Definitions of these terms are given in Appendix A.

Table 1. The Eight Terminal Specifications to be Determined.

V_{inH}	V_{inL}	I_{inH}	I_{inL}
V_{oH}	V_{oL}	I_{oH}	I_{oL}

We begin the calculation of the terminal specification by noting that when a transistor is saturated, the collector-emitter voltage is 0.2 volts. The output terminal is connected directly to the collector terminal, thus, $V_{oL} = 0.2$ volts. Assume the gate on which this analysis is taking place is embedded in a logic system. Its inputs are driven by other identical logic gates. Then if the outputs of those driver gates are logic low, the inputs to the device under analysis are low. We will use this as a starting place for the calculations.

$V_{inL} = 0.2$, (LOW input to both transistors)

Under this condition, both transistors will be cutoff, making the output high. The circuit diagram with circuit models for the cutoff transistors is shown in Figure 4. The output voltage then is 3.0 Volts. We will call this a no-load condition when there is no load connected to the output.

$V_{oH} = 3.0$ Volts (No-Load)

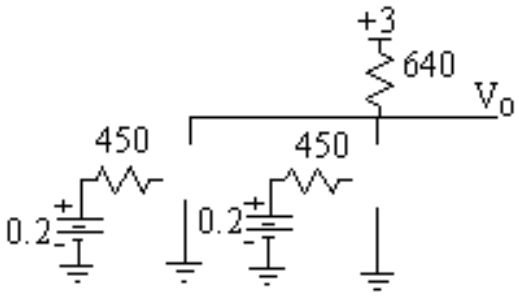


Figure 4. RTL gate with both inputs low. The transistors are replaced by the cutoff model.

Also note from Figure 4 that the input current is zero because the transistor is cutoff. Thus,

$$I_{inL} = 0$$

The question naturally arises about how high the input voltage may rise and still keep the transistor cut off. The input voltage must stay below $V_{BE\gamma}$ to guarantee the transistor is off, making

V_{inLmax} = 0.5 Volts.

The maximum value is given in this case because we need to know how high it may rise. This means that the input voltage may rise as high as 0.5 volts and still keep the transistors off, and will always be recognized by the gate as a "low".

INPUT HIGH

In this case we want the transistor to be saturated. Either transistor saturated will cause the output to go low, to 0.2 volts. This case is shown in Figure 5. We show the circuit with one of the transistors cutoff, although both saturated would produce the same result. We start by determining the minimum input current that will keep the transistor in saturation.

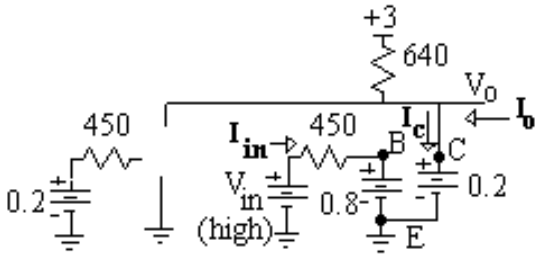


Figure 5. RTL gate with one input high.

$$I_{Csat} = \frac{3.0 - 0.2}{0.640k} = 4.375mA$$

The minimum base current to keep the transistor saturated is (assuming β=30)

$$I_{inH\min} = \frac{I_{Csat}}{\beta} = \frac{4.375}{30} = 0.146mA$$

The minimum value is of interest because this current is dependent on the input voltage. From this value, we can determine the minimum input voltage that will be guaranteed to be recognized as a "high".

V_{inHmin} = 0.8 + 0.146 mA x 0.450 KΩ = 0.866 Volts

Any input voltage greater than or equal to 0.866 volts will cause the transistor to be saturated, and thus be recognized as a "high". We previously found V_{inLmax} =0.5 volts. Any input voltage between 0.5 and 0.866 is an invalid logic level and the manufacturer assumes no responsibility if you provide an input voltage in that range. For the gate with no load, the relationships between input and output voltages are presented graphically in Figure 6. The noise margins NMH and NML are defined in Appendix A.

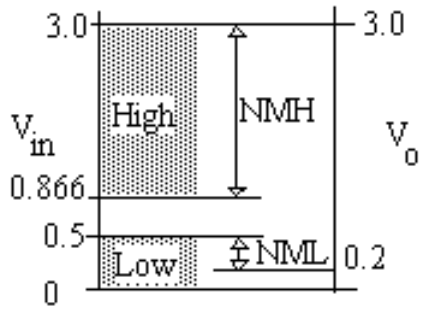


Figure 6. Graphical representation of input voltage ranges and output voltages for the unloaded RTL gate.

CALCULATIONS WHEN THE GATE IS USED TO DRIVE OTHER GATES

Up to this point we have done all calculations assuming no load on the gate. While convenient for visualization, it is not a very useful situation. Let us look at the loading of the gate as if it were embedded in a logic system driving identical gates as shown in Fig. 7.

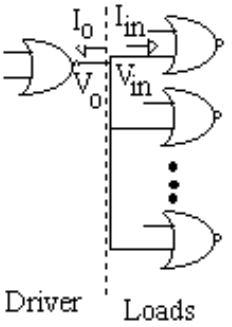


Figure 7. Driver gate being analyzed driving load gates.

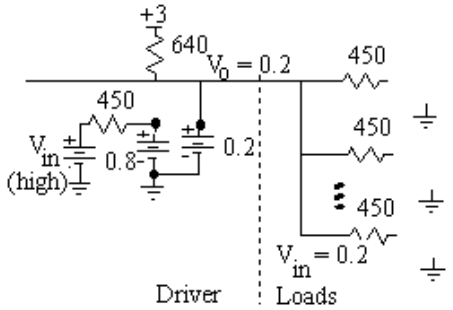


Figure 8. Driver with loads when driver output is low.

There are two cases, the driver output being low as shown in Figure 8, and the driver output being high as shown in Figure 9. When the output is low, the driver transistor is saturated and the output is 0.2 Volts. All the input transistors of the load gates are cutoff and no current flows. The output voltage remains at 0.2 volts, unchanged by the loads.

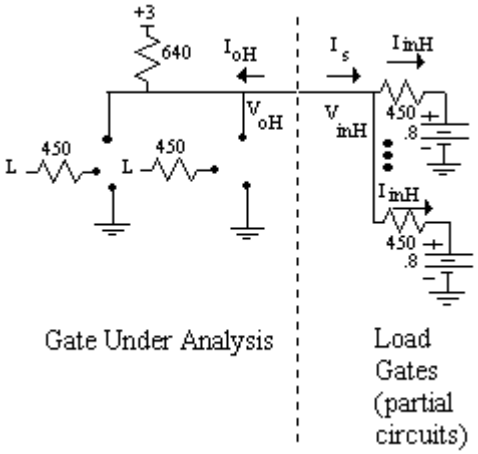


Figure 9. RTL gate with N loads when driver output is high.

However, when the output is high, the load gates require current, pulling the output voltage down. This situation is shown in Figure 9 with an unspecified number of loads. How many can you connect and still expect the gates to work right? This is called the maximum fanout. The output voltage is determined by the load current being SOURCED by the driver.

$V_{oH} = 3.0 - I_S \times 0.64\ k\Omega$

Where the current coming out of the driver gate is divided between the load gate inputs. Since the minimum voltage at the load gate inputs when high is 0.866 Volts, this is the lowest the output voltage of the driver gate may go. Thus,

$$I_{oH\max} = -\frac{3.0 - 0.866}{0.64k} = -3.334mA$$

The negative sign comes about because currents are always defined as positive when going into the terminal. Each load requires a minimum of 0.146 mA as input current, therefore, the maximum fanout is

$$MaxFanout \le \frac{3.334}{0.146} = 22.8$$

Thus, the maximum fanout is 22.

PRACTICAL FANOUT

The maximum fanout will not provide for any noise margin. Also, manufacturing tolerances will not allow such close calculations. Therefore, most manufacturers limit the allowed fanout to a much smaller number than found by nominal and typical calculations. Typically, the fanout would be limited to say 5 loads. In this case, the output voltage under maximum load would be substantially higher than calculated above. With five loads, the output current will be

$$I_{oH} = -\frac{3.0 - 0.8}{0.64 + \frac{0.45}{5}} = -3.014mA$$

and the output voltage will be

$V_{oH} = 3.0 - I_{oH} \times R_c = 3.0 - 3.014 \text{ mA} \times .64 \text{ k}\Omega = 1.071 \text{ Volts}$

Under this condition, the Noise Margin High is

NMH = 1.071 - 0.866 = 0.205 Volts

Note that in many cases the output voltage will be higher than the value just calculated and thus the noise margin larger. If we define the maximum allowed fanout to be five, then V_{oHmin} = 1.071 Volts.

CALCULATION OF I_{oL}

If we define the maximum allowed fanout to be five, then the minimum voltage ever seen by the input will be 1.071 Volts. The minimum current into the input will be

$$I_{inH\min} = \frac{1.071 - 0.8}{0.45k} = 0.602mA$$

The maximum collector current is β times this or 18.067 mA and still stay in saturation. Thus, the maximum output current when the output is low is this current minus the current down through the 640 Ω resistor.

$I_{oLmax} = 18.067 - 4.375 = 13.692 \text{ mA}$

We now have all the specifications for this gate. The results are shown in Table 2. We have assumed here that the fanout is limited to 5 loads. There is a certain amount of arbitrariness in the choice of V_{oH} and I_{oH}. Many other performance requirements could have been chosen. The values given are self consistent, however. First note that $V_{oHmin} > V_{inHmin}$ and $V_{oLmax} < V_{inLmax}$, as shown in Figure 10. The voltages are compatible.

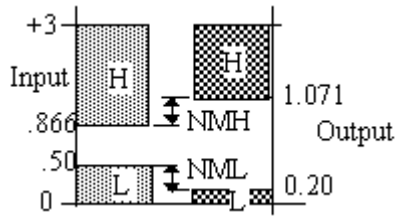


Figure 10. Voltage compatibility chart, with 5 loads

To determine fanout,

$$\frac{I_{oH}}{I_{inH}} = 5.007 \quad \text{and} \quad \frac{I_{oL}}{I_{inL}} = \infty$$

Therefore, the fanout is 5.

Table 2. Terminal specifications for the RTL Gate

V_{inLmax} = 0.500 Volts V_{oL} = 0.2 Volts

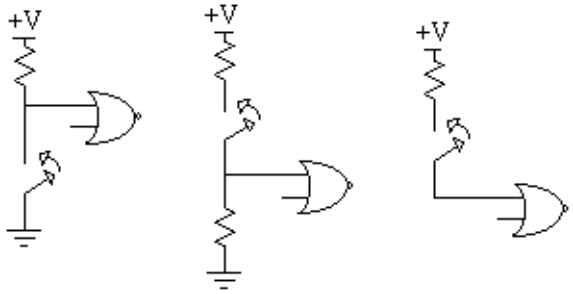
V_{inHmin} = 0.866 Volts V_{oHmin} = 1.071 Volts (N=5)

I_{inL} = 0 I_{oLmax} = 13.692 mA ($V_{in} > V_{oHmin}$)

I_{inHmin} = 0.602 mA I_{oHmax} = - 3.014 mA

INTERFACING TO RTL

Table 2 is a set of specifications that would be provided by the manufacturer of the RTL logic. From this information, you would be expected to be able to use the gates without exceeding any specification. Any gate embedded in a logic system using only the RTL gates analyzed, would be guaranteed to work within the specifications. The only limitation we must follow is limiting the fanout to five loads. However, logic is rarely used without interfacing to the outside world at some point. There must be some signals at some point that are generated outside the logic system that must be recognized as a logic one or a zero. Similarly, there is usually some point where a logic gate output must drive a device that is not another gate.



a. b. c.

Figure 11. Three possible configurations of a switch-to-logic gate interface.

Figure 11 shows three possible connection using a switch to generate the logic inputs to a gate. In Figure 11a the switch is used as a pull-down to make the input voltage a logic zero when the switch is closed. The resistor pulls the input voltage up to a logic one when the switch is open. Looking at the specifications in Table 2, we note that when the switch is closed, the input voltage is zero volts, well below the required V_{inLmax} of 0.5 volts. The gate input current is zero so the switch needs only to handle the current coming down through the resistor. When the switch is open, current will come down through the resistor into the gate input terminal. The specifications show the voltage must be above 0.866 volts and the current must be at least 0.602 mA. From these values, we can select a value of R_1. To provide a noise margin, we should certainly provide a voltage at least as high as V_{oHmin} = 1.071 volts. Let's shoot for 2.0 volts to provide even greater noise margin. We know this will provide greater than the minimum current, but we don't yet know how much. We now have to put on our engineer's hat and look inside the gate to finish the design. The parts of the circuit that pertain to this case are shown in Figure 12 with only the base-emitter voltage of the transistor shown. The transistor must be in saturation so we use 0.8 volts for V_{BEsat}.

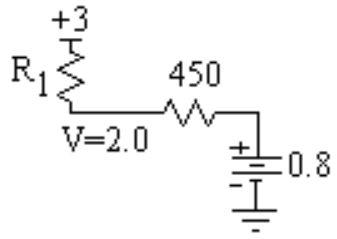


Figure 12. Circuit with switch open.

Solving this circuit, we can write a single node equation at the input node, V_{in} = 2.0 v,

$$(3.0-2.0)/R_1 = (2.0-0.8)/450$$

Solving, we get R_1 = 375 Ω. This is not a standard resistor value so we would select a nearby standard value and re-solve the circuit, this time to determine the input voltage. If we are satisfied with the result, then we are finished. If the input voltage found is not satisfactory, we would try with another value. Note that we do not have to get 2.0 volts, we simply chose that value as a place to start the design. We do, however, have to provide $V_{in} > V_{inHmin}$.

The design of the circuit in Figure 11b would follow a similar track. Notice that in this circuit the input voltage is high when the switch is closed; the opposite of the circuit in Figure 11a. Figure 11c is simpler than 11b, but it is not recommended because when the switch is open, the input lead of the gate is floating or disconnected and susceptible to noise which might cause random errors in the logic.

Figure 13 shows an example of an interface at the output of the gate. In this case, we want to drive an LED at 30 mA. Neither the high nor low output currents of the gate are sufficient, so we use a transistor as a current amplifier. Many other configurations are possible and some may be preferrable depending on the application. We will not go through the design of this circuit here. A similar circuit was designed at the end of the previous chapter.

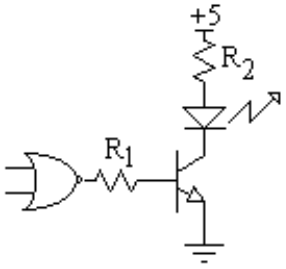


Figure 13. Example of an output interface circuit.

EXERCISE

Mountain State Electronics, Inc. has provided a data sheet for its line of RTL NOR gates:

V_{inLmax} = 0.45 volts V_{oLmax} = 0.3 volts

V_{inHmin} = 0.85 volts V_{oHmin} = 1.00 volts (@N=5)

I_{inL} = 0.0 mA I_{oLmax} = 10.00 mA (@ $V_{in} > V_{oHmin}$)

I_{inHmin} = 1.00 mA I_{oHmax} = -5.50mA (@$V_o = V_{oHmin}$)

1. Determine: Fanout = _____
   NMH = _____
   NML = _____

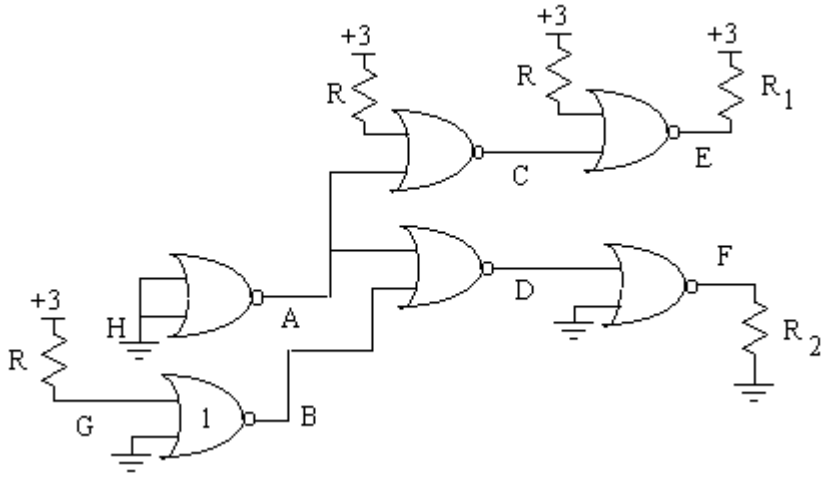
2. For the logic network below, write the logic level (H or L) at each node.

3. If you used a voltmeter to measure the voltage, what voltage would you expect to measure at each node? (Use <, > signs to describe the voltages. These voltages will be the thresholds you will use to determine if the network is operating properly.)
   V_A _____ V_C _____ V_E _____ V_G _____
   V_B _____ V_D _____ V_F _____ V_H _____

4. Beside each connecting wire, draw an arrow showing the direction of the actual current in the wire. Write I = 0 if no current flows.

5. In each wire labeled A-H, estimate the magnitude of the current.
   I_A _____ I_C _____ I_E _____ I_G _____
   I_B _____ I_D _____ I_F _____ I_H _____

6. Detemine the maximum allowable value of the pull-up resistor R at the input of gate #1. R_{max} = _____



In-Class Design Exercise:

Assume RTL Specifications are given by manufacturer:

Terminal Specifications:

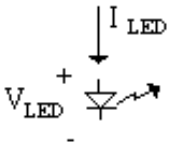
V_{inLmax} = 0.5 volts I_{inL} = 0

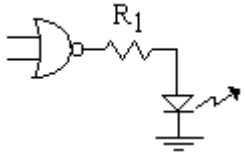
V_{inHmin} = 0.866 volts I_{inHmin} = 0.602 mA For max I_{oL}

V_{oL} = 0.2 I_{oLmax} = 13.692 mA @V_{inH}>V_{oHmin})

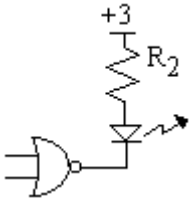
V_{oH} = 1.071 @Fanout = 5 I_{oHmax} = -3.014 mA

1. Drive an LED from an RTL gate.





Method 1



Method 2

LED Model:
V_{LED} = 1.8 for I > 0
I = 0 for V_{LED} < 1.8

Performance Requirement:
9 mA < I_{LEDon}

Constraints:
Use 5% standard value resistors, nominal calculations
3-volt power supply

Method 1 How does it work? (Draw the currents on the diagram)

When gate output is low LED is _____

When gate output is high LED is _____

How much current is available to turn on the LED? _____

R_1 = _____

Method 2 How does it work? (Draw the currents on the diagram)

When gate output is low LED is _____

When gate output is high LED is _____

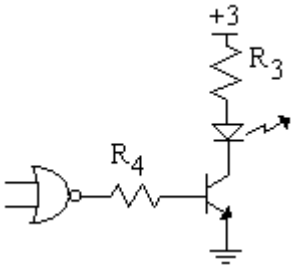
How much current is available to turn on the LED? _____

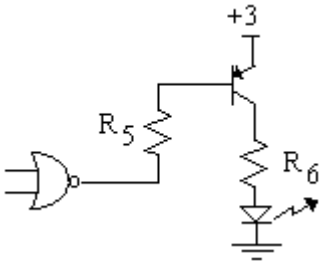
_____< R_2 < _____

2. Same problem, except the performance requirement:

20 mA < I_{LED} < 25 mA

Sssuggested Mthods:





Method 3 Method 4

Assume the following transistor specifications:
$V_{BE\gamma} = 0.5$, $V_{BEsat} = 0.8$, $V_{CEsat} = 0.2$, $\beta = 10$
(negative voltages for pnp)

Method 3 How does it work? (Draw the currents on the diagram)

When gate output is low, the transistor is _____, and LED is _____

When gate output is high, the transistor is _____, and LED is _____

How much current is available to saturate the transistor _____
to turn on the LED? _____

R_{3min} = _____, _____< R_4< _____

Designer's choice: R_4 = _____, then I_{Csat} = _____, I_{Bmin} = _____

R_{3max} = _____

Method 4 How does it work? (Draw the currents on the diagram)

When gate output is low, transistor is _____, and LED is _____

When gate output is high , transistor is _____, and LED is _____

How much base current is available to saturate the transistor _____
What is the maximum collector current is available to turn on the LED? _____

R_{5min} = _____, _____< R_6< _____

Designer's choice: R_6 = _____, then I_{Csat} = _____, I_{Bmin} = _____

R_{5max} = _____

Problems

Mountaineer Logic Inc. is developing a new line of RTL logic NOR gates. The gates are said to have a fanout of 5 and the following component values:

V_{CC} = 5.0 Volts, R_C = 470 Ohms, R_B = 1.5 K Ohms, and β_{min} = 20.

1. For the RTL gate above, present the following specifications in tabular form:

Noise margins NMH and NML, V_{oL}, V_{oH} (at N=5), V_{inLmax}, and V_{inHmin}.

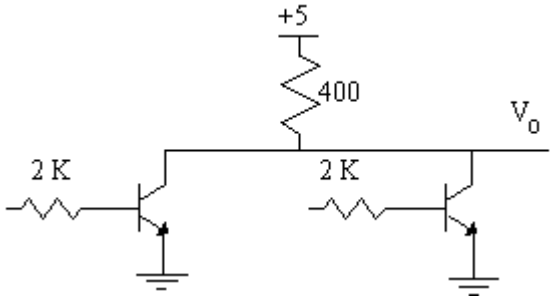
2. Draw the voltage transfer characteristic (V_o vs V_{in}) for the gate.

3. Determine the absolute maximum fanout with zero noise margin.

4. The RTL gate shown below has a specified "high" noise margin, NMH, of 1.00 volts. What is the specified fanout?

The transistors model parameters are:

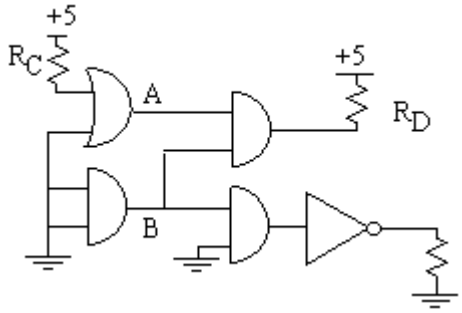
$V_{BEcutin}$ = 0.5 volts, V_{BEsat} = 0.8 volts, V_{CEsat} = 0.2 volts, Beta = 12



5. A logic family has the following terminal specifications:

V_{inLmax} = 0.6	V_{oLmax} = 0.4
V_{inHmin} = 1.8	V_{oHmin} = 2.4
I_{inL} = -1.0 mA	I_{oLMax} = 10 mA
I_{inH} = 1.0 mA	I_{oHmax} = -5 mA

This family of logic gates is used in the following logic system.



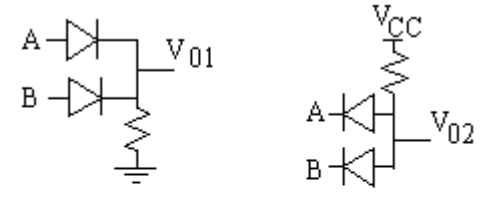
What voltage do you expect at nodes A and B?
What values of resistors R_C and R_D can be used? (Specify max or min.)

Chapter 5
DIODE TRANSISTOR LOGIC

Resistor-Transistor Logic was an early form of logic used in the 1950's and early 1960's. RTL was made from discrete transistors and resistors and manufactured on printed circuit boards with several gates per board. These boards were plugged into board sockets with wiring on the socket pins determining the system function. RTL was a big improvement over vacuum tube technology previously used, requiring less than one quarter the space and one tenth the power dissipation. RTL was superceded in the 1960's by Diode-Transistor Logic and then Transistor-Transistor Logic, DTL and TTL respectively. DTL was initially made with discrete transistors and resistors before being integrated onto silicon. One early form of DTL, used by IBM Corp in the 360 family of computers, was really a hybrid technology. Transistor and diode chips were glued to a ceramic substrate and aluminum resistor paste was deposited on the substrate to make resistors. Finally the ceramic base and components were hermetically sealed in an aluminum can. This family was used extensively in IBM products in the middle to late 1960's. While this family was not a true integrated circuit, it was very successful and was less expensive than true integrated circuits for several years. By the early 1970's integrated circuits became quite common and DTL gave way to TTL which was more appropriate to integrated circuit technology.

While DTL is no longer commercially used, we will discuss it because it is similar to and easier to understand than TTL, and because designers still find the configuration of value. First, however, we will discuss diode logic which is the front end of the DTL gate and performs the actual logic operation.

DIODE LOGIC



A	B	V_{01} = A+B
L	L	L
L	H	H
H	L	H
H	H	H

A	B	V_{02} = A•B
L	L	L
L	H	L
H	L	L
H	H	H

Figure 1. Diode logic and truth tables

Two diode logic configurations are shown in Figure 1. The truth tables show that the first circuit performes a logical OR and the second circuit performs a logical AND. One problem with these circuits is that there is a voltage level shift through the circuit. If several circuits are cascaded as shown in Figure 2, the output voltage for each stage is approximately one diode voltage drop further away from the rail. Logic "0" rises for the

AND gate and logic "1" drops for the OR gate. Only a limited number of series cascaded circuits can be used before the logic levels must be restored to the rails. Even more serious is the voltage degradation for the cascaded AND-OR function.

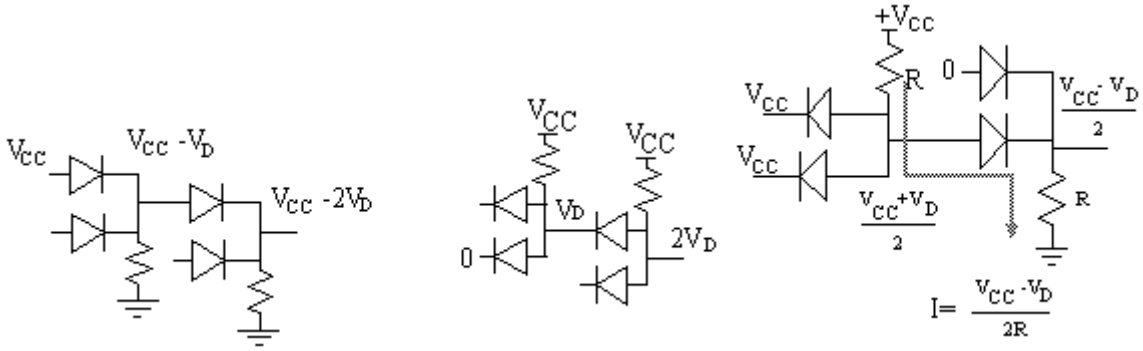


Figure 2. Cascaded diode logic showing level shifts.

The obvious way to restore the logic level voltages is to use a double transistor inverter. However, it does not take a genius to notice the a single inversion on the end of each gate will make a NAND or a NOR gate and the logic level is automatically restored for each gate. The NAND gate is used extensively. As you no doubt recall, any logic expression can be implemented using just NAND gates. Thus, other forms are not necessary.

DIODE TRANSISTOR LOGIC CIRCUITS

A typical DTL NAND gate is shown in Figure 3. Observe the diode AND function on the front end and the transistor NOT at the output end. The extra resistors and diodes are used to maintain appropriate currents, to maintain proper functioning, and to guarantee certain noise margins. We will completely analyze this circuit, but will be particularly interested in the "terminal" characteristics of the gate. Our ultimate goal is to understand the operation of the gate so that the manufacturer's specifications can be understood.

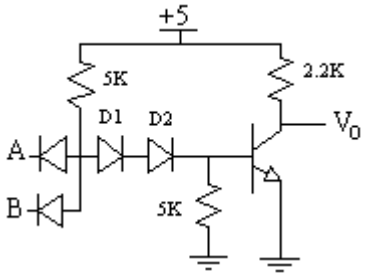


Figure 3. DTL Circuit

ANALYSIS OF THE DTL GATE

Analysis of the DTL gate is dependent on complete understanding of the currents within the gate under all logic conditions. First let us develop a generic understanding of the operation. *Of particular importance will be the direction of currents at the terminals of the gate*. As in most logic systems, the transistor will either be cutoff or saturated.

If all inputs are high, (+5v), no current will come out of the input diodes at the input and current will flow down through the first 5K resistor and through the diodes D1and D2 toward the base of the transistor. Some current will split off and go down through the lower 5K resistor to ground. However, most of the current will go into the base of the transistor causing it to saturate, pulling the output low, V_O =0.2 Volts. We will show this condition quantitatively shortly.

If one or more of the inputs to the gate are held low (0.2 V), then the current down through the 5K resistor will go out the input diode, away from the transistor base. Under this condition, the transistor will be cutoff and the ouput will be high with V_O =5 Volts.

ANALYSIS WITH INPUT LOW

Quantitatively, we will start with one or more inputs held low, at 0.2 Volts. From the logic function of the NAND gate, we know that the output is supposed to be high. Therefore, the transistor must be cutoff. To begin with, we will assume the two diodes D1 and D2 in series will also be cutoff. All the current coming down through the 5K resistor must all go out through the input diode, causing it to be on. The circuit with the models indicated is shown in Figure 4. From this circuit we can calculate all voltages and currents and prove (or disprove) our assumptions about the condition of each element.

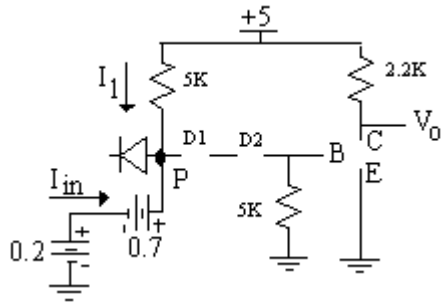


Figure 4. DTL gate model with input low

The voltage at point P is

$$V_P = 0.2 + 0.70 = 0.90 \text{ Volts.}$$

We need to show that this voltage is low enough that the two series diodes and the transistor will be cutoff. The argument is that if either diode carries current, then both must. Since 0.9 Volts is not enough across the pair to maintain conduction, then neither conducts. Given that the diodes are off, then the voltage at the base of the transistor is zero, and is also cutoff. We can verify that the input diode is conducting by observing that current I_1 is

$$I_1 = (5\text{-}0.9)/5K = 0.82 \text{ mA}$$

This current leaves through the input diode, hence it is on. The current entering the input terminal is

$$I_{in} = -I_1 = -0.82 \text{ mA}$$

The negative sign occurs because the input current is defined as going into the terminal.

We can now verify that the output voltage is 5 Volts because the transistor has been shown to be cutoff. Thus,

$$V_O = 5.0 \text{ Volts}$$

One other characteristic of the DTL gate that can be obtained at this point is the range of input voltages that will be recognized as a "low". From the logic function of the NAND gate, this can be translated into the question of how high the input voltage may rise and still keep the transistor cutoff.

The transistor will remain cutoff as long as the voltage at the base does not rise above 0.5 volts. At this voltage, there will be current down through the lower 5K resistor to ground. This current must come from the +5 supply down through the upper 5K resistor and the diodes, D1 and D2. Hence the diodes must be conducting. This current will be 0.1 mA. The voltage at point P will be

$$V_P = 0.5 + 0.7 + 0.7 = 1.9 \text{ Volts}$$

The current I_1 is

$$I_1 = (5-1.9)/5K = 0.62 \text{ mA}$$

The current going out through the input diode will be

$$I_{in} = -(0.62 - 0.1) = -0.52 \text{ mA}$$

indicating that the input diode is still conducting. Figure 5 shows the resulting circuit with the circuit models included. The maximum voltage at the input that is guaranteed to be recognized as a low is

$$V_{inLmax} = 1.9 - 0.7 = 1.2 \text{ Volts.}$$

This result is included in the table of terminal specification at the end of this chapter.

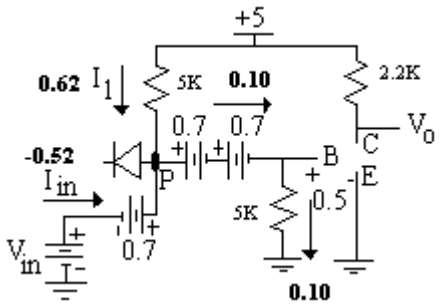


Figure 5. DTL circuit model with $V_{in} = V_{inLmax}$

ANALYSIS WITH ALL INPUTS HIGH

When all inputs are high, all current down through the upper 5K resistor will go toward the base of the transistor, causing it to saturate. The series diodes will obviously be conducting, and we will show that the input diodes are cutoff. Figure 6 shows the circuit with these models.

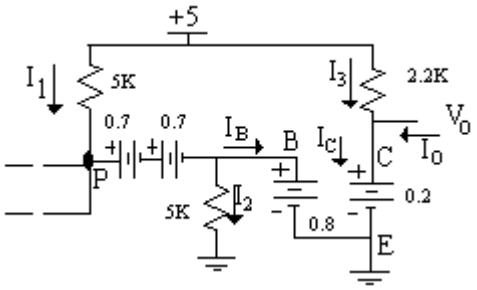


Figure 6. DTL circuit model with inputs high

The voltage at point P is

$$V_P = 0.8 + 0.7 + 0.7 = 2.2 \text{ volts}$$

Thus, I_1 is

$$I_1 = (5\text{-}2.2)/5K = 0.56 \text{ mA}$$

This current will go through the diodes toward the base of the transistor. Some of the current will go down through the lower 5K resistor with the rest going into the base of the transistor. With the transistor saturated, the current going down through the 5K resistor will be

$$I_2 = 0.8/5K = 0.16 \text{ mA}$$

The base current then is

$$I_B = 0.56 - 0.16 = 0.4 \text{ mA}$$

If the transistor is to be saturated, the maximum collector current is

$$I_{Cmax} = \beta I_B = 30 * 0.4 \text{ mA} = 12.00 \text{ mA}$$

at saturation, the current coming down through the 2.2K collector resistor is

$$I_3 = (5\text{-}0.2)/2.2K = 2.182 \text{ mA}$$

this current is much less than the maximum saturation current and we see that with no load, the transistor will, indeed, be in saturation. In fact, there is excess capacity in collector saturation current. This excess capacity can be used to **sink** external load current. This current is called I_o or load current. The maximum load current this gate can sink is

$$I_{oLmax} = 12.00 \text{ mA} - 2.182 \text{ mA} = 9.818 \text{ mA}$$

Note that this current is entering the terminal of the gate, hence, is positive.

CALCULATION OF V_{inHmin} and I_{inH}

We need to go back now and look at the input voltage in Figure 6 and determine the minimum allowable input voltage that will still be recognized as a high, V_{inHmin}. There are several ways we could define this value, but we will use a straightforward definition. That is, we will define the input to be high as long as no current flows through the input diodes. In other words, as long as the input diodes are cutoff. Thus,

$$V_{inHmin} = V_P - 0.60 = 2.20 - 0.60 = 1.6 \text{ Volts.}$$

Since the input diodes must remain cutoff, $I_{inH} = 0$.

CALCULATION OF FANOUT

If several load gates are connected to the output terminal of the gate we are looking at, we need to look at the current output drive capability compared to the input current requirements of the load gates. Because the input current is zero when high, an infinite number of load gates can be driven when high. However, the DTL gate requires current when the input is low. This situation is shown in Figure 7. We made that calculation earlier and found $I_{inLmax} = - 0.82$ mA. Note that this current is negative meaning that it is coming out of the input terminal. We also found the output of a gate can sink 9.818 mA when it is low. We can now calculate the fanout.

$$\text{Fanout} \leq \frac{I_{oLmax}}{I_{inL}} = \frac{9.818}{0.82} = 11.97$$

The maximum number of gates that can be driven as loads is 11.

Fanout = 11

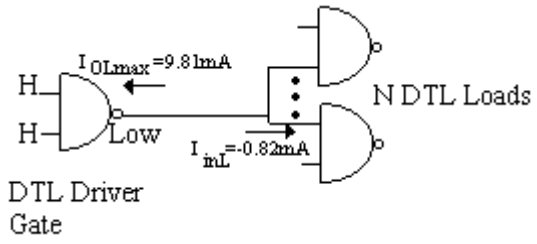


Figure 7. DTL driver gate with N identical DTL load gates

We also need to check voltage compatibility. That is, $V_{oH}>V_{inHmin}$ and $V_{oL}<V_{inLmax}$ which is true as shown in Figure 8 where V_{oH} is given at no load condition because I_{inH} = 0.

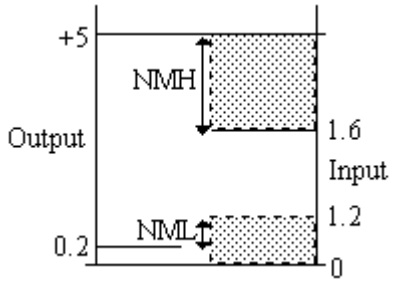


Figure 8. Comparison of output and input voltages for DTL gate.

CALCULATION OF V_{oH} AND I_{oH}

We now have calculated all terminal characteristics except for the output voltage and current when the output is high. Because I_{inH} = 0, with any number of loads there will be zero load curent. Thus, V_{oH} = 5.0 Volts. This specification is satisfactory if the only load is other gates. In many cases, however, we wish to use other types of loads and additional information is necessary.

When the output is high, the transistor is cutoff and any current coming out of the gate will come from the supply through the collector resistor. One way to define the output is when the output is allowed to drop as low as V_{inHmin} (although this is by no means the only way). Using this definition, we can calculate the output current as

$$I_{oH} = -(5-1.6)/2.2K = -1.545 \text{ mA}$$

Again, the negative sign indicates that current flows out.

We now have a complete set of specifications for the DTL gate as shown in Table 1.

Table 1. Terminal Specification for the DTL GATE

V_{inLmax} = 1.2 V	V_{oL} = 0.2
V_{inHmin} = 1.6 V	V_{oH} = 5 (at I_{oH} =0)
	= 1.6 (at I_{oH} = max)
I_{inL} = - 0.82 mA	I_{oLmax} = 9.818 mA
I_{inH} = 0	I_{oHmax} = -1.545 mA (at V_{oH} =V_{inHmin})

Fanout = 11

//
Exercise:
In the above discussion, V_{oH} was allowed to drop all the way to V_{inHmin}. Allowing it to drop this low, makes the high noise margin, NMH = 0. A more appropriate noise margin would be to make NMH equal to the NML. Calculate the I_{oH} under this condition.

(ans. -1.09 mA)

\\

OTHER LOGIC FUNCTIONS

While the NAND function can be used to implement any logic expression, it is often convenient to use other functions. Figure 9, 10, 11, and 12 show how four other function could be implemented. Figure 9 is the AND function, Figure 10 is the NOR fuction, Figure 11 is the OR function, and Figure 12 is the AND-OR-INVERT, AOI, a very useful function when implementing sum-of-products logic expressions.

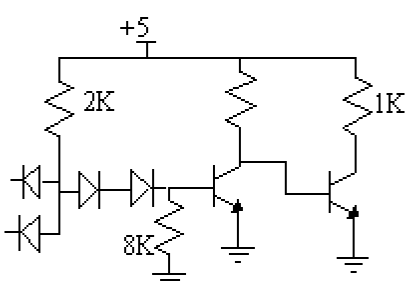


Figure 9. DTL AND gate

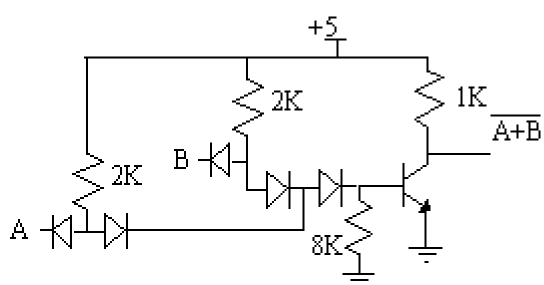


Figure 10. DTL NOR gate

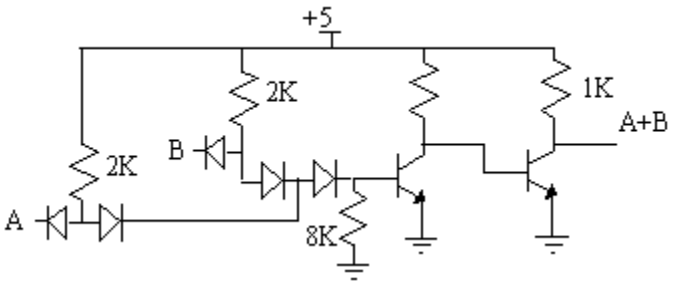


Figure 11. DTL OR gate

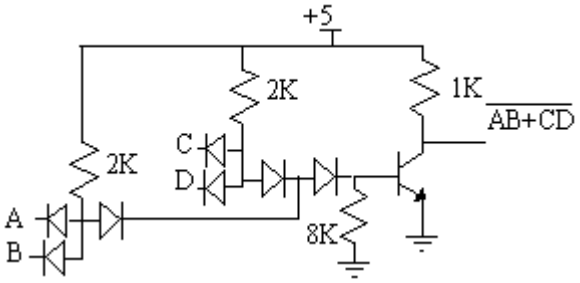
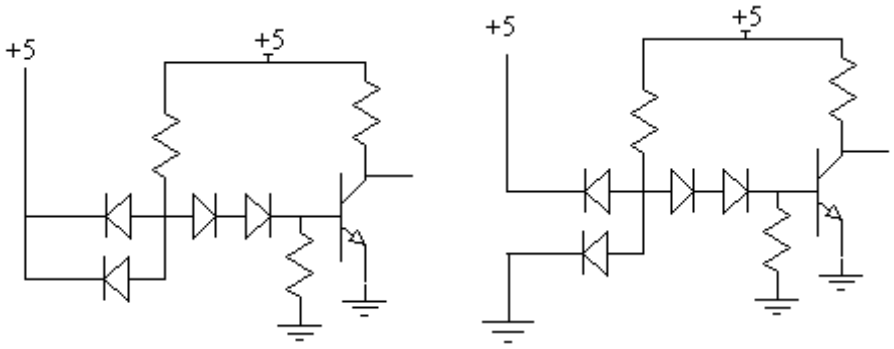


Figure 12 DTL AND-OR-INVERT gate

EXERCISES

1. For the following two DTL gate circuits as shown:
    a. Write the logic levels (H or L) at both inputs and the output.
    b. What is the state of the transistor?
    c. Draw arrows showing the direction of all currents in the circuit.



For the following exercises, the manufacturer of DTL gates gives the following specifications:

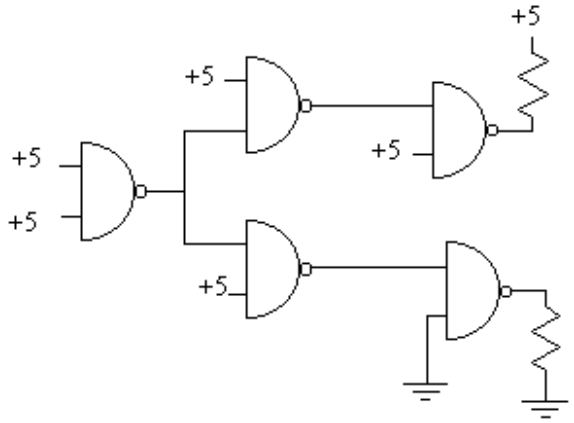
V_{inLmax} = 1.00 Volts V_{oLmax} = 0.30 volts

V_{inHmin} = 2.00 volts V_{oHmin} = 2.75 volts

I_{inLmax} = -1.00 mA I_{oLmax} = 10.00 mA

I_{inH} = 0.00 I_{oHmax} = -5.00 mA (@V_o = 2.75 v)

3. A logic system shown below is made up of DTL gates.
    1. Write the logic level (H orL) beside each input and output for each gate.
    2. Draw arrows beside each connecting wire, including inputs and outputs, showing the direction of current in the wire. If zero, write I=0.



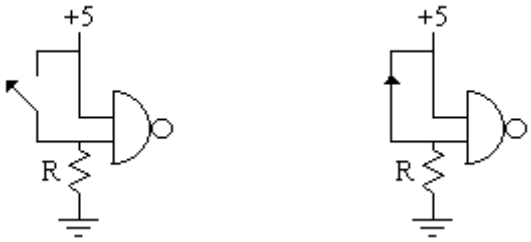
4. On the diagram, give the magnitude of all currents except the two final outputs.

5. Determine the minimum value of each of the load resistors if the gates are to be operated within specifications and logic levels are preserved.

$R_{pullup\ min}$ = _____ $R_{pulldown\ min}$ = _____

6. The circuit below is used to interface a switch to the DTL gate specified above. It is assumed that with the switch in one position, the gate sees a valid logic low, and when the switch is in the other position, the gate sees a valid logic high.

a. Indicate on the drawings, which logic level the gate input sees.
b. Draw arrows showing currents when the switch is open and when the switch is closed.



7. For the input interface circuit above, determine the maximum value of R that can be used and still have the gate operate within specifications. What determines the minimum value of R?

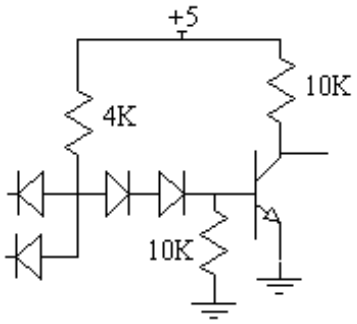
R_{max} = _____

R_{min} is determined by:

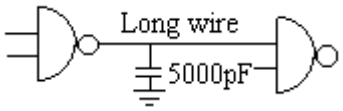
Problems

1. Your employer, Mountaineer Logic, Inc., is developing a new line of discrete DTL gates shown below. Write up a specification sheet for the device. (Use nominal calculations only.) Define the high level output current at $V_O = V_{inHmin}+1.0$ volts. Use a tabular format similar to the DTL data sheet given in this chapter. Assume the following diode and transistor models.
$V_{Dcutin} = 0.6$, $V_{Don} = 0.7$, $V_{BEcutin} = 0.5$, $V_{BEsat} = 0.8$, Beta = 20

2. The Quality Control Department of MLI just called you about the DTL gate above. It seems they connected it to a network with a single load gate. The output rise time seemed much too long. Upon questioning, you discover that they used a very long wire to connect between the driver and the load. You estimate that the capacitance of the wire to be 5000 pF as shown in the figure below. Estimate how long it will take to charge the capacitor so the output of the gate to rise from a "low" to V_{inHmin}. Because you are a professional, present only the circuit you used, the results, and the assumptions you made (including circuit models). Do not show details of the calculations.



DTL Gate



Circuit with 5000pF capacitive load

3. The manufacturer's specifications for a DTL gate are:

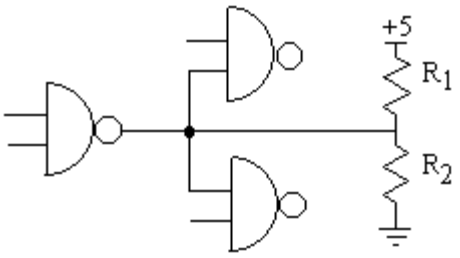
V_{inLmax} = 1.00 volts V_{oLmax} = 0.30 volts

V_{inHmin} = 2.00 volts V_{oHmin} = 2.75 volts

I_{inLmax} = -1.00 mA I_{oLmax} = 10.00 mA

I_{inH} = 0.00 I_{oHmax} = -5.00 mA (@V_O = 2.75 v)

This gate is used to drive two identical gates, plus a resistive load as shown. Determine the allowed range for R_1 if R_2 = 2200 Ohms.



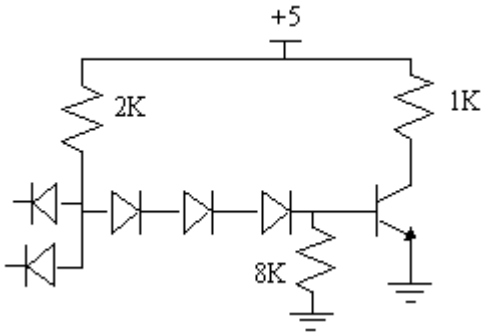
4. A DTL gate has an additional diode added as seen in the circuit below. Determine V_{inLmax}, V_{inHmin}, I_{inL}, I_{inH}, and I_{oLmax}.

Diode Models:

$V_{D\gamma} = 0.60$, $V_{Don} = 0.70$ volts

Transistor Models:

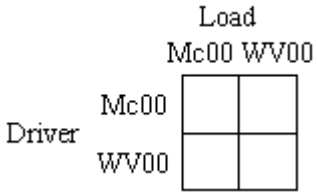
$V_{BE\gamma} = 0.50$, $V_{BEsat} = 0.80$, $V_{CEsat} = 0.20$, Beta = 20



6. Two logic families are being used in a logic system. Their terminal specifications are given below. Fill in the compatibility chart.

	Mc00	WV00
V_{inLmax}	1.00	1.50
V_{inHmin}	2.00	3.00
I_{inL}	- 0.500	- 1.00 mA
I_{inH}	1.00	0.050 mA
V_{oLmax}	0.50	0.75

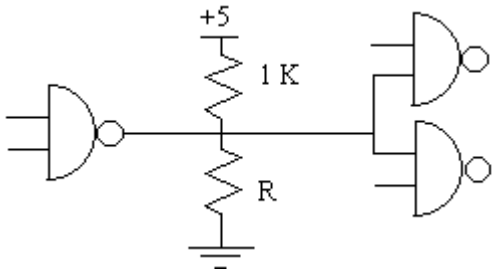
V_{oHmin}	2.50	4.00
I_{oLmax}	5.50	8.00 mA
I_{oHmax}	-7.00	-2.50 mA



7. In the following logic circuit, all the gates are Mc00 gates whose specifications are given below. Determine the range of resistance allowed for the resistor R such that all gates are working within their specifications. Note that the load gates must see a legitimate logic low and legitimate logic high.

V_{inLmax} =1.00 V_{inHmin} =2.00 I_{inL} = - 0.500 I_{inH} = 1.00

V_{oLmax} = 0.50 V_{oHmin} = 2.50 I_{oLmax} = 5.50 I_{oHmax} = -7.00



Chapter 6

TRANSISTOR-TRANSISTOR LOGIC

The evolution from DTL to TTL can be seen by observing the placement of p-n junctions. For example, the diode D2 from Figure 2 in the chapter on DTL can be replaced by a transistor whose collector is pulled up to the power supply; transistor Q_2 in Figure 1 below. The p-n junction of D2 is replaced by the BE junction of Q2 and with the current gain of the transistor, the current going into the base of Q3 is greatly increased, increasing the fanout.

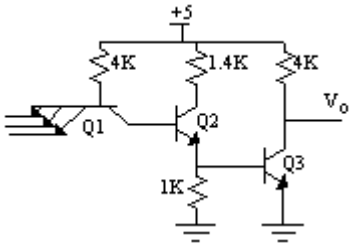


Figure 1. TTL Gate.

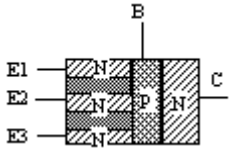


Figure 2. Configuration of Q_1 as a 3-emitter transistor.

The input diodes and D1 are replaced by the multi-emitter NPN transistor, Q1, in Fig. 1 and represented by the drawing in Figure 2. Later on, we will make additional modifications to this curcuit to improve its performance further.

The analysis of this circuit follows very much the same path as the analysis of the DTL gate. For the most part, we will consider the input transistor, Q1, to act just like two diodes. The transistor Q2, however, will operate in all three regions. The treatment of the output voltages and currents will be treated the same as the DTL gate and Q3 will either be cutoff or saturated, corresponding to an output high and an output low, respectively.

ANALYSIS WITH ONE OR MORE INPUTS LOW

With an input low, Q3 should be cutoff. We will assume Q2 is cutoff and then check our assumption. If Q2 is cutoff, then there can be no current coming out of the collector of Q1, hence its base-collector junction can be modeled as an open circuit. The base-emitter junction of Q1 will be conducting. The circuit with these models substituted for the transistors is shown in Figure 3. Note the similarity to the DTL circuit under the same conditions. The two unused inputs are assumed to be high, and are thus, modeled as open. From this case, we can see that V_{oH} = 5 volts with no load, and

$$I_{inL} = -I_1 = -(5-0.9)/4K = -1.025 \text{ mA}$$

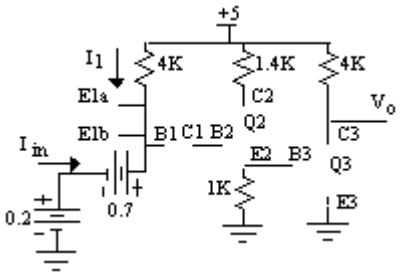


Figure 3. TTL circuit model with one input low.

We turn now to finding V_{InLmax}. We will use the criterion that V_{in} will be considered as a low as long as Q3 is kept cutoff. If the base voltage for Q3 can be raised to 0.5 Volts without turning it on, then there will be 0.5 mA current in the 1KΩ resistor. This current can only come from Q2, which means it must be conducting. Even assuming all this 0.5 mA comes through the collector of Q2, the voltage drop across the 1.4 KΩ resistor will be 0.7 Volts, not enough to cause the transistor to saturate. Thus, the active model for Q2 is appropriate as shown in Figure 4.

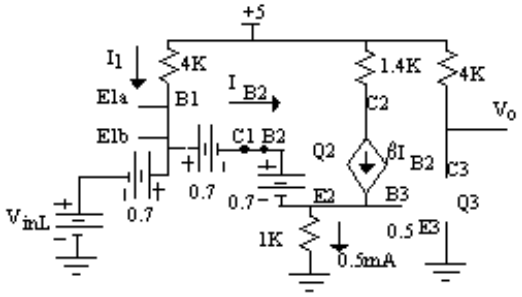


Figure 4. TTL circuit model to determine V_{inLmax}.

If we assume that β=30, the base current in Q2 is

$$I_{B2} = \frac{0.5mA}{\beta+1} = \frac{0.5}{31} = 0.016mA$$

Because this current is coming out of the collector of Q1, the base- collector junction of Q1 is on, and is modeled as a diode in Figure 4.

The voltage at B1, the base of Q1, is

$$V_{B1} = 0.5 + 0.7 + 0.7 = 1.9 \text{ Volts}$$

The current coming down through the 4 KΩ resistor, I_1, is

$$I_1 = \frac{5.0 - 1.9}{4K} = 0.775mA$$

This is considerably more than is going into the base of Q2, therefore, the input BE junction of Q1 will also still be conducting. The maximum voltage at the input is

$$V_{inLmax} = 1.9 - 0.7 = 1.2 \text{ Volts}$$

CALCULATIONS WITH INPUT HIGH

The circuit model for the TTL gate with all inputs high is shown in Figure 5. Both Q2 and Q3 are modeled as saturated, an assumption that must be verified. With the inputs high, Q1 is modeled as two diodes with the B-E diodes cutoff, and B-C diode conducting.

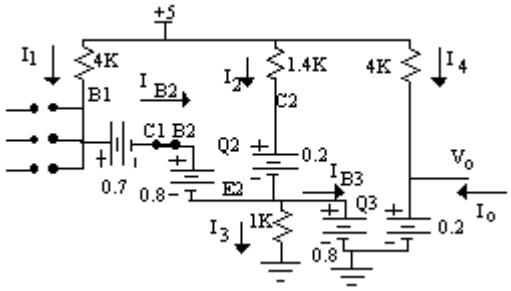


Figure 5. TTL gate circuit model with all inputs high.

The voltage at the base of Q1 is

$V_{B1} = 0.8 + 0.8 + 0.7 = 2.3$ Volts.

The current down through the 4 KΩ resistor, I_1 is

$$I_1 = \frac{5.0 - 2.3}{4K} = 0.675mA$$

All this current goes into the base of Q2.

$I_{B2} = 0.675$ mA

If Q2 is saturated, voltage at its collector terminal is

$V_{C2} = 0.8 + 0.2 = 1.0$ Volts

And the collector current is

$$I_{C2} = I_2 = \frac{5.0 - 1.0}{1.4K} = 2.857mA$$

Clearly, if $\beta = 30$, $\beta I_{B2} > I_{C2}$, and, therefore, Q2 is saturated.

The current coming out of the emitter of Q2 is the sum of the base and collector currents. Part of this current will go down through the 1 KΩ resistor to ground and the rest will enter the base of Q3.

$I_{B3} = I_{B2} + I_{C2} - I_3 = 0.675 + 2.857 - 0.8 = 2.732$ mA

The maximum collector current that Q_3 can carry and still be in saturation is βI_{B3} =81.96 mA, assuming β=30. The maximum current the gate can sink when the output is low

$$I_{oLmax} = I_{Csatmax} - I_4 = 81.96 - 1.2 = 80.76 \text{ mA}$$

Now let's turn our attention back to the input and determine V_{inHmin} and I_{inH} . We will define the input voltage to be high as long as no current goes out the input terminal. Thus, all we have to do is keep the input voltage high enough so that the B-E p-n junction of Q1 does not turn on. Thus,

$$V_{inHmin} = 2.3 - 0.6 = 1.7 \text{ Volts}$$

CALCULATION OF I_{inH}

With the input voltage at a high, say 5 volts, the transistor Q1 will be operating in the reverse active mode. The B-E junction is reverse biased, and the B-C junction is forward biased with a base current of 0.675 mA. If there were significant curent gain, you would expect to see a large current going into the input. However, the reverse β is typically on the order of 0.02. Thus,

$$I_{inH} = \beta_R * I = 0.02 * 0.675 = 0.0135 \text{ mA}$$

This current would add to the current going into the base of Q2, but is ignored because it is quite small and because β_R is made as small as possible and this input current is a maximum and cannot be counted on.

THE TOTEM POLE OUTPUT STAGE

One of the problems with the TTL gate circuit we have been analyzing is that the pull-up resistor on the output transistor will prevent rapid charging of any wiring capacitance on the output. One way to improve the rise time is to reduce the resistance value as is often done, but this also increases the power dissipation when the output is low.

If we look at the circuit, we observe that when the transistor is saturated, it presents a very low effective resistance to ground. The problem arises when the output is high and the pull-up resistor is too large. Ideally we would like to have a very low resistance pull-up when the output is high, but a very high pull-up resistance when the output is low. In this way, we could get quick charging and very low power dissipation. The totem-pole output stage for TTL, shown in Figure 6, does just that.

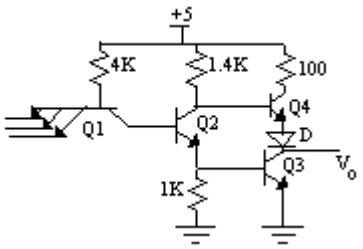


Figure 6. TTL gate with totem-pole output.

This circuit operates just like the original circuit except that Q_4 is on when the output is high and off when the output is low. We need to verify this operation.

OUTPUT LOW

Figure 7 shows the TTL circuit with all inputs high and the output low. The models for the transistors are shown as before, except diode D and transistor Q_4 are added and shown as cutoff.

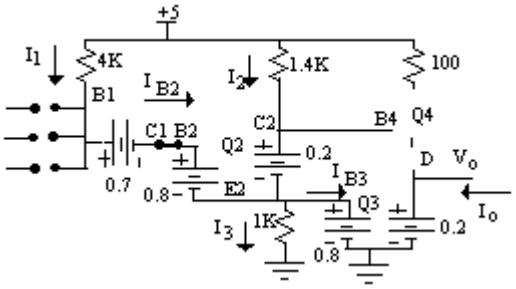


Figure 7. TTL gate with totem-pole output circuit model with inputs high.

The analysis of this circuit proceeds exactly the same as before. The currents, I_1, I_2, I_3, and I_{B3} are the same as before. With the diode and Q_4 not conducting, I_{oLmax} is now the same as I_{C4max} , 81.96 mA. We only need to show that the diode D and transistor Q_4 are indeed off.

The voltage at the bottom of the diode is 0.2 Volts and the voltage at the base of Q_4 equal to the voltage at the collector of Q_2; V_{C2} = (0.2 + 0.8) = 1.0 Volts. Thus, the voltage across the B-E junction of Q_4 plus the diode is 0.8 Volts. If one conducts, the other must also. To take both out of cutoff would require at least 0.5 + 0.6 = 1.1 Volts. Thus, both are off.

OUTPUT HIGH

This condition occurs when one or more inputs are low. The circuit is shown in Figure 8 with the appropriate models used for the transistors and the diode. In this case, Q2 and Q3 are both cutoff while Q4 and the diode are conducting. We have to assume here that there is some load and that the output current is not zero.

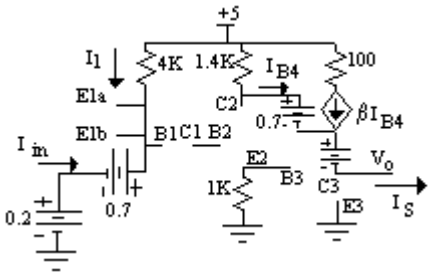


Figure 8. TTL totem-pole circuit model with output high.

The current coming out the output terminal I_S (=-I_o) is the sum of the currents coming down through the base and the collector. Thus,

$I_S = I_{B4} + \beta I_{B4}$

Because each TTL load represents 13 μA, if we assume there are 10 loads, then I_S = 130 μA. The base current is

$$I_{B4} = \frac{130\mu A}{1+\beta} = 4.2\mu A$$

where we have assumed a β of 30. Then taking the path down through the 1.4 KΩ resistor to the output. the output voltage is

$V_o = 5.0 - 4.2\mu A * 1.4K - 0.7 - 0.7 = 3.6$ Volts

The voltage drop across the 1.4 KΩ resistor is neglegible. Of course as the current increases, the output voltage will drop further.

TERMINAL SPECIFICATIONS OF THE TTL GATE

We are now ready to make the table showing the terminal specifications for the TTL gate. These are shown in Table 1.

Table 1. Terminal Specifications For TTL

V_{inLmax} = 1.2 V	V_{oL} = 0.2
V_{inHmin} = 1.7	V_{oH} = 3.4 (@I_o = -130 μA)
I_{inL} = -1.025 mA	I_{oLmax} = 81.96 mA
I_{inH} = 13 μA	I_{oH} = undetermined

TTL DATA

Recommended Operating Conditions

	74	74H	74L	74LS	74S	Units
$V_{CC(min)}$ — $V_{CC(max)}$	4.75-5.25	4.75-5.25	4.75-5.25	4.75-5.25	4.75-5.25	V
I_{OH}	-400	-500	-200	-400	-1000	μA
I_{OL}	16	20	3.6	8	20	mA
Operating Free-Air Temperature Range	0 - 70	0 - 70	0 - 70	0 - 70	0 - 70	°C

Electrical Characteristics Over Recommended Operating Temperature Range

	Condition	74	74H	74L	74LS	74S	Units
$V_{IH(min)}$		2.00	2.00	2.00	2.00	2.00	V
$V_{IL(max)}$		0.80	0.80	0.70	0.80	0.80	V
$V_{OH(min)}$	@ I_{OH} = max	2.40	2.40	2.40	2.70	2.70	V
$V_{OL(max)}$	@ I_{OL} = max	0.40	0.40	0.40	0.5	0.5	V
$I_{IH(max)}$		40	50	10	20	50	μA
$I_{IL(max)}$		-1.60	-2.00	-0.18	-0.40	-2.00	mA
I_{OS}	*	-20 - -55	-40 - -100	-3 - -15	-20 - -100	-40 - -100	mA

* Not more than one output should be shorted at a time and for H, LS, and S series, duration of short should not exceed one second.

Figure 9. Data for ‘00, ‘04, ‘10, and ‘30 NAND gates for several TTL families (Abstracted from Texas Instruments TTL Data Book.)

MANUFACTURER'S DATA SHEETS

The terminal specifications of several TTL families are shown in Figure 9. You will note the values given for various voltages and currents are quite different from those we calculated. This difference comes from the fact that manufacturing tolerances and variations cannot be closely controlled, hence, the specifications given by the manufacturers are much more conservative than our calculations which were based on nominal values. Also note that the limits are usually given as a maximum or a minimum, depending on which limit is normally used in design. For example, I_{inLmax} is given as -1.6 mA for the 74xx series. What this means is that as a designer, your driver must be able to sink as much as 1.6 mA when the input to the gate is pulled low.

You will note that the TTL gate is rather loosely specified. The question invariably arises as to how one reads the data sheets or designs with this data. Figure 10 shows the allowed operating regions for a 7400, 2-input NAND gate. The best description of these operating regions is probably given by the following examples.

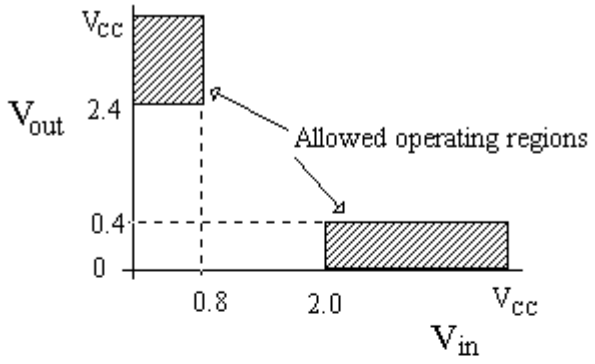
Example 1: If the input voltage is between 0.00 and 0.08 volts, the output voltage will be below V_{CC} and above 2.4 volts as shown in Figure 10a.

Example 2: If the input voltage is between 2.00 and V_{CC}, the input current will be between 0 and 40 μA as shown in Figure 10b.

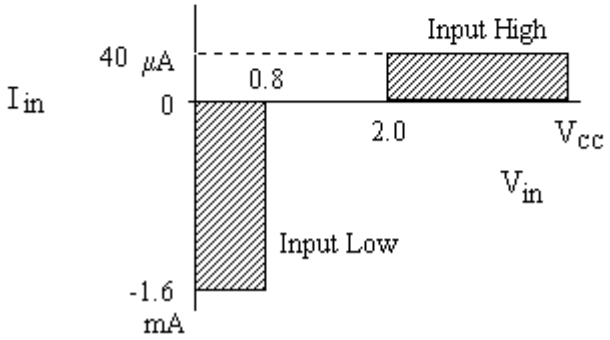
You will note in the above examples that there is no mathematical relationship between one variable and another. There is simply not enough data to develop one and the variability of the manufacturing process prohibits the manufacturer from providing one.

In a design setting, you must stay within the limits provided by the manufacturer. For example, if you wanted to connect a resistor from the output of a 7400 gate and ground, what would be the limits allowed on the resistance value? Figure 10c provides part of the answer. If we assume that we must operate within the shaded region which represents a "High" level output, we should not allow the output to drop below 2.4 volts with 400 μA coming out of the gate. The minimum value would be 6 KΩ. The upper limit is, of course, infinite; an open circuit ($V_o = V_{CC}$, $I_o = 0$).

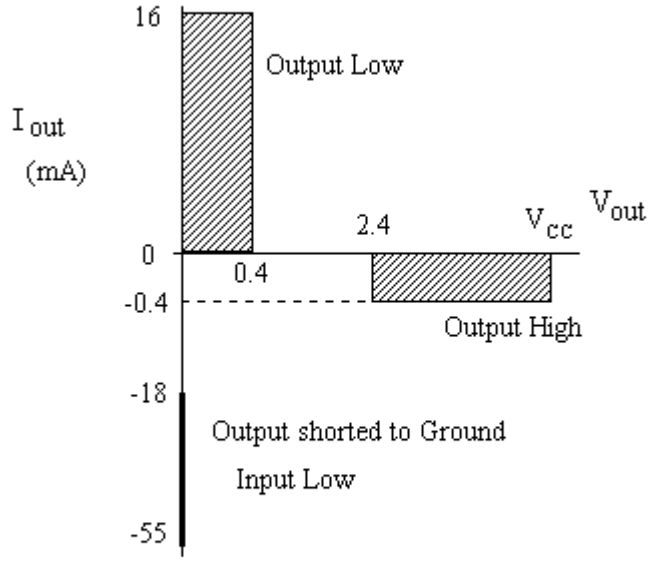
There are a few data points provide by the specifications that are not within the limits of normal operation. For example, if the input voltage drops below zero, it is allowed to drop to -1.5 volts where you may expect as much as 12 mA coming out of a 7400 gate input. Most of this current comes from an input clamp diode which has not been shown on our drawings.



a. Allowed operating region, V_{out} vs V_{in}, for a 7400 gate.



b. Allowed operating region, I_{in} vs V_{in}



c. Allowed operating region, I_{out} vs V_{out}

Figure 10. Allowed operating regions for a 7400 gate.

Another non-standard data point of interest is I_{oS}, the short-circuit output current. This is the current you get from the gate output if the output is shorted to ground when the output of the gate would otherwise be high. In this case, both minimum and maximum values are given, -18 to -55 mA. If you went to the laboratory and actually performed this deed, you could expect a current somewhere in this range. How does this affect the designer? For example, a designer might be tempted to connect the output of the 7400 gate directly to the base of an NPN transistor whose emitter is grounded. In this case, the "high" output voltage is clamped at 0.80 V by the BE junction. What current can you expect into the base of the transistor when the gate output goes "high"? This condition is tricky and perhaps open to some debate, but the conservative designer must recognize that the operation is between the short circuit case and the case where V_{oHmin}=2.4V when I_{oH} =-400μA. The conservative designer would conclude that the current might be as low as 400 μA and as high as 55 mA; the worst cases. It is possible to go back to the circuit of the gate with 0.8 volts at the output terminal and calculate the current. However, this analysis would be for nominal values only and not provide definitive limits on the current.

Note the notation at the bottom of the specification table in Figure 9. This notation discusses the limit on the amount of time a short circuit is allowed to be connected to the output of some gates. This time limit is based on the amount of time it takes the internal components of the integrated circuit to heat up to its maximum allowed value. While connecting a transistor base to the output is not exactly a short circuit, it is outside the allowed operating region and probably should have the same time limits as the short circuit.

TERMINAL CHARACTERISTICS

During the previous discussions on TTL, we were looking at circuit operation and developing an understanding of how the terminal specifications were arrived at. Let us now take a broader look at these characteristics.

First, the input currents are quite high when the input is low, requiring the driver to **sink** a lot of current. When the input is high, the input current into the gate is quite low. Thus, any circuit which is supposed to drive the input to a TTL gate must concentrate on sinking current, and only needs to source a little current when the driver output voltage is high.

Second, the output strength of the TTL gate matches the strength requirements at the input. An example is given in Figure 11. The TTL gate can sink a large current when its output is low, but can only source a small current when the output is high. Thus, if the TTL gate is expected to drive a circuit that is not another TTL circuit, you must exercise care when designing the interface. The load circuit must not require large input currents when its input is high, but may use larger currents when the input voltage is low.

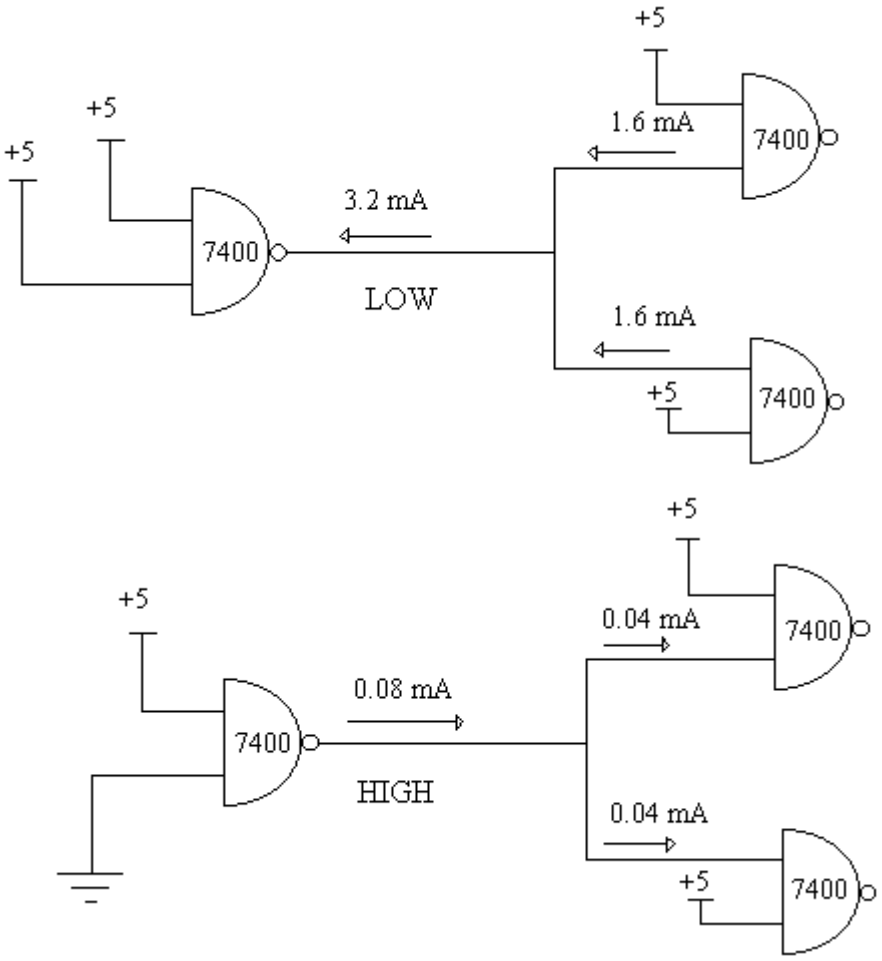


Figure 11. When driving other TTL gates as loads, a 7400 gate must be able to sink more current than it needs to source.

These requirements must be kept in mind when designing interfaces with the TTL gate at both the input and the output. Examples of interfacing with TTL gates are shown in Figure 12 and 13. See data books for more complete data.

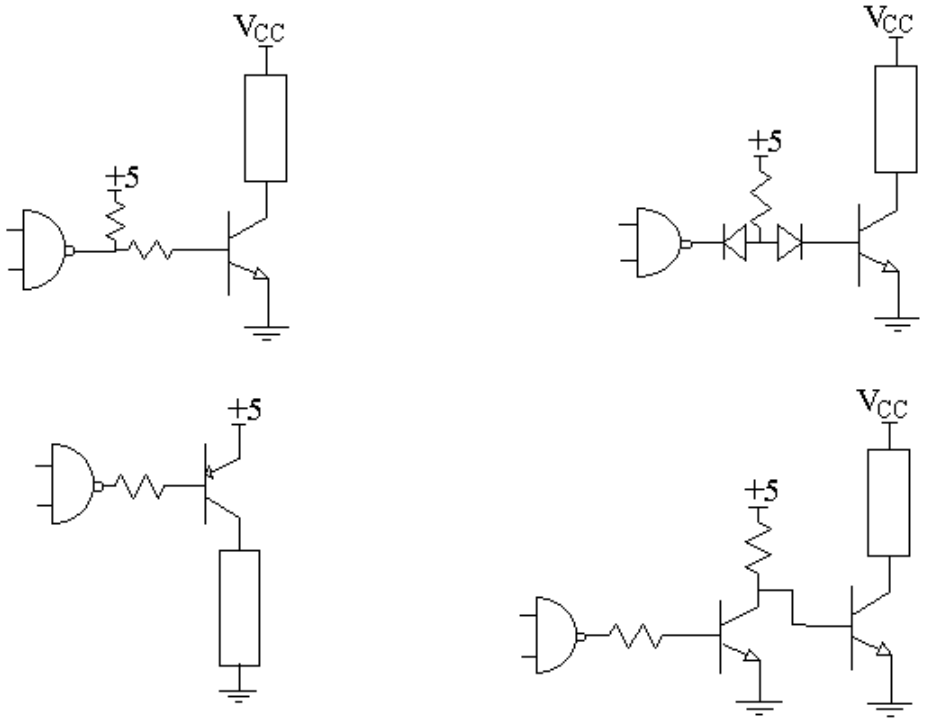


Figure 12. Several ways to drive loads from TTL gates.

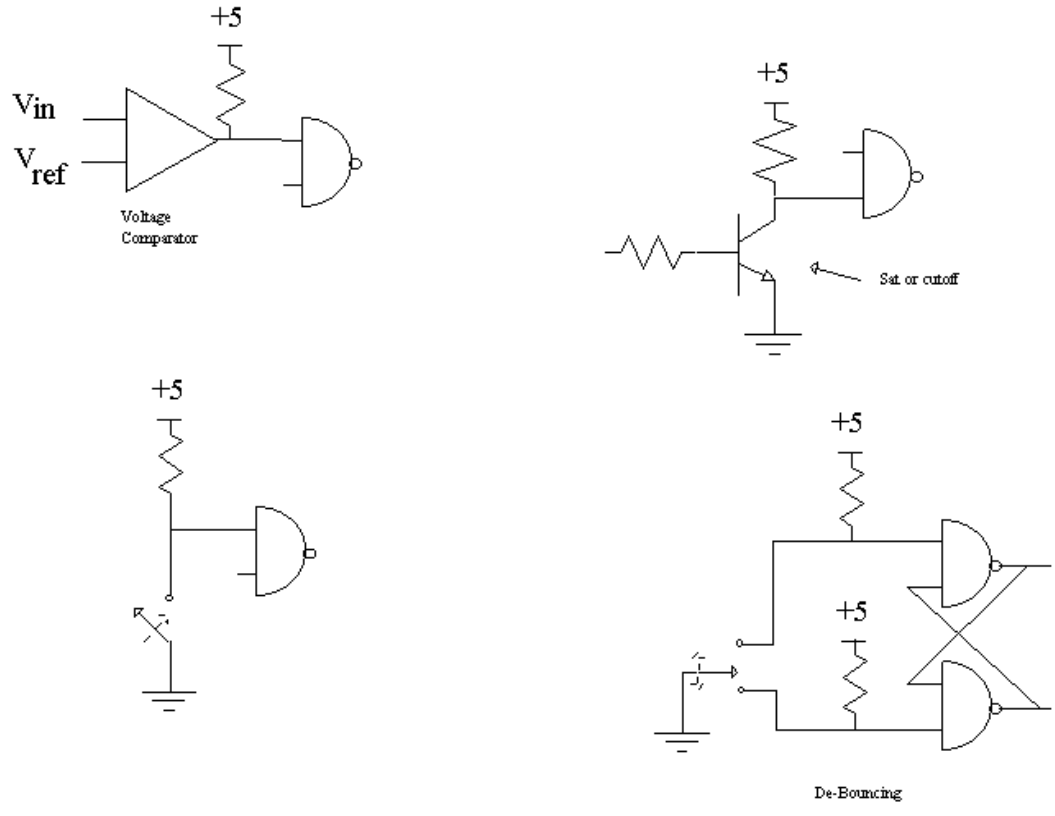


Figure 13. Several interfaces to drive TTL gates.

WIRED-AND CONNECTION

Because the active pull-up or totem-pole output of the TTL gate always has one transistor cutoff and the other turned on, you cannot connect two outputs together. If one is trying to pull the output high, and the other is trying to pull it low, you will have a very low impedance path to ground and very large currents.

For the same reason, the output must not be connected to any voltage source or to ground through a low impedance path. In one state or the other, there would be a low impedance path and large currents.

OPEN COLLECTOR GATES

In order to overcome the limitations created by the totem pole output circuit, some gates are manufactured with the output collector left open. One example is the 7405, a quad 2-input NAND gate with open collector outputs. If you connect a resistor as the pull-up, you can use this resistor to source current when the output is high and/or you can wire-AND the collectors together.

TTL FAMILIES

As the designers of TTL gates became more sophisticated, they developed modifications which would provide special characteristics. The original series of TTL was designated as 74XX, where the XX is replaced by logic function (00 is a quadruple 2-input NAND, 04 is a hex inverter, etc.) The 74LXX series is a low power family. 74HXX is a high speed family. 74SXX is a family based on Schottky diodes and transistors. 74LSXX is a family of low power Schottky. A 54xXX is also provided as a companion family to the 74xXX families. The 54... families are identical to the 74... families, except for operating temperature range and tolerance on power supply voltage.

Each family has different characteristics, but the same logic functions. The L family is low power, but is much slower than the standard family. The H family is high speed, but also has higher power dissipation. The Schottky families are quite fast without increasing the power dissipation. More recent advances in TTL family have given us several other versions. For example, 74F, 74AS, and 74ALS, for Fast, Advanced Schottky, and Advanced Low-power Schottky. The AS family is the fastest, with a propagation delay of less than 5 ns. Table 2 shows the propagation delays and power supply current for each type of gate. The power supply current, I_{CC}, is the average for a 50% duty cycle with the output spending half its time low and half the time at a high.

In addition, these different families use slightly different circuit configurations. A little study of the circuits will reveal the same operations.

Table 2. Propagation delays and power supply current for TTL families
Data abstracted from Texas Instruments TTL Data Books

Gate	t_{PLH} (ns)			t_{PHL} (ns)			I_{CC} (mA)
	min	typ	max	min	typ	max	typical
7400		11	22		7	15	2.00
74L00		35	60		31	60	0.20
74H00		5.9	10		6.2	10	4.50
74LS00		9	20		10	20	0.40
74S00	2	3	4.5	2	3	5	3.75
74ALS00	3		11	2		8	1.00
74AS00	1		5	1		4	6.20

SCHOTTKY TTL

A Schottky PN junction is made up of a semiconductor and a metal. This kind of junction has two characteristics: low turn-on voltage and low junction capacitance.

When a Schottky junction is used in place of or in parallel to the Base-Collector junction of a transistor, the transistor is faster because of the lower junction capacitance and because the transistor cannot go so deep into saturation. Because the turn-on voltage for the BC junction is lower, V_{CEsat} is higher.

Schottky TTL is thus faster than standard TTL and the terminal voltages are slightly different. See the data sheet.

TRI-STATE OUTPUT

The totem-pole output of a TTL gate provides additional speed at lower power for the gate than a simple pull-up resistor. The cost, however, is that the gate outputs cannot be connected in parallel. This problem is serious when you need to make a bus structure such as a data bus, where several gates need to put data onto the bus at different times. The outputs can be OR'd or AND'ed using appropriate gates, but this solution is less than satisfactory and slows down the operation.

A better solution is the TRI-STATE output as shown in Figure 14. The added input allows normal operation of the gate when the "enable" input is high. Both output transistors are cutoff when "enable" is low.

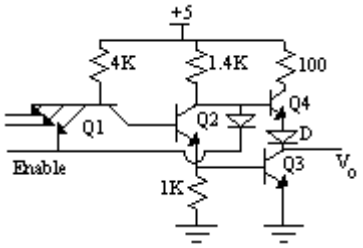


Figure 14. TTL gate with Tri-State output.

OTHER LOGIC FUNCTIONS IN TTL

TTL integrated circuits have been manufactured that perform over one hundred different functions. Many are small-scale integrated circuits that perform simple logic functions in addition to the basic NAND as well as flip-flops. Others are medium-scale integrated circuits that perform more sophisticated functions such as encoders, decoders, and registers. The more complex functions are implemented in TTL using combinations of the basic logic gates. It is of interest to look at how some of the other logic funcitons are implemented. Figures 15, 16, and 17 show a NOR gate, an AOI gate (AND-OR-INVERT), and an AND gate.

In each case, the input and output circuits are the same as for the NAND gate. Extra parts are added to perform the desired function. For example, the NOR gate has extra Q_1 and Q_2 added in parallel to the existing transistors. The extra emitters in Q_1 are left out. These extra emitters perform the AND function in the NAND gate. The OR is performed by the Q_2s in paarallel. The final inversion is performed by the output stage by Q_3. Note the similarity to the AOI gate in which the AND function is intact with the extra emitters in the Q_1s.

The AND gate in Figure 16 has an extra internal inversion. The actual inversion is done with Q_5. The other additional transistor, Q_6 is used simply to guarantee Q_2 is turned off when Q_5 is turned on. This extra transistor is necessary to maintain voltage levels consistent with the NAND gate.

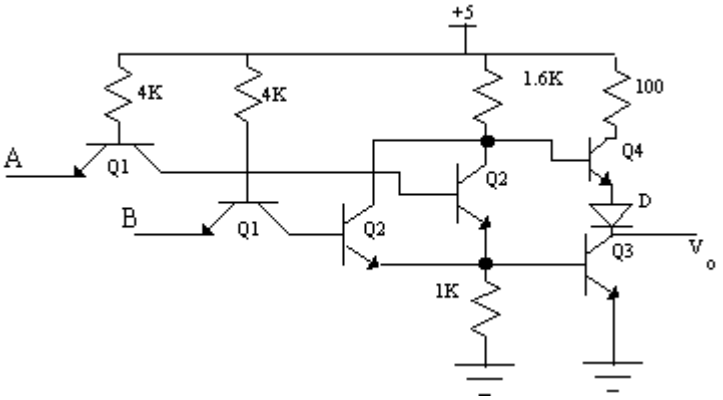


Figure 15. TTL NOR gate implementation

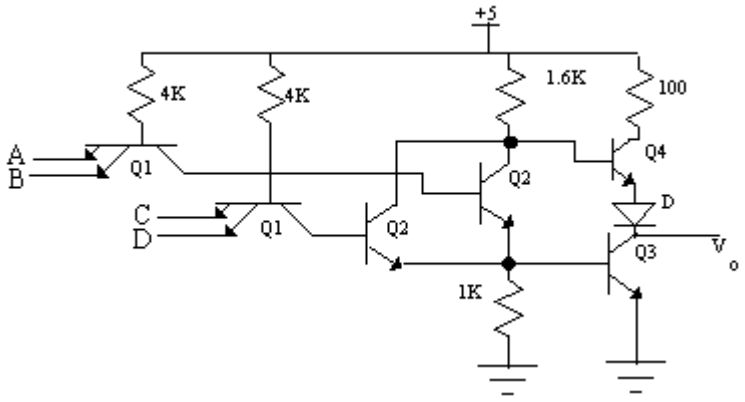


Figure 16. TTL AND-OR-INVERT (AOI) gate implementation

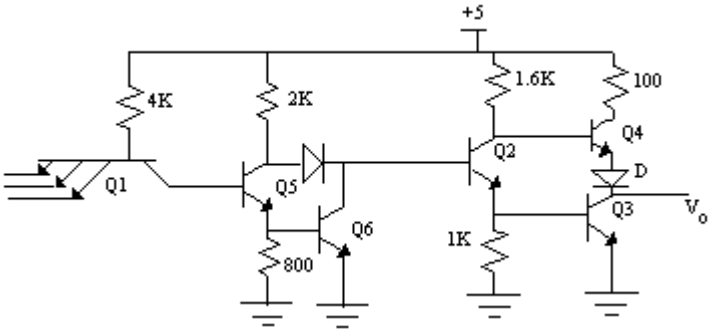


Figure 17. TTL AND gate implementation

How do these modifications in the basic TTL circuit affect propagation delays? Table 3 show the typical and maximum propagation delays for each type of gate discussed. It is apparent from the table that the addition of the extra internal inverter stage in the AND gate does have a significant effect on propagation delays while the addition of the parallel circuits for the NOR and AOI gates has very little effect. This latter effect is caused primarily by the addition of internal capacitance.

Table 3. Propagation delays for NAND, NOR, AND, and AOI gates.
Data abstracted from Texas Instruments TTL Data Book

Logic Gate	t_{PLH} (ns)		t_{PHL} (ns)	
	typ	max	typ	max
7400 (NAND)	11	22	7	15
7402 (NOR)	12	22	8	15
7408 (AND)	17.5	27	12	19
7451 (AOI)	13	22	8	15

EXERCISES

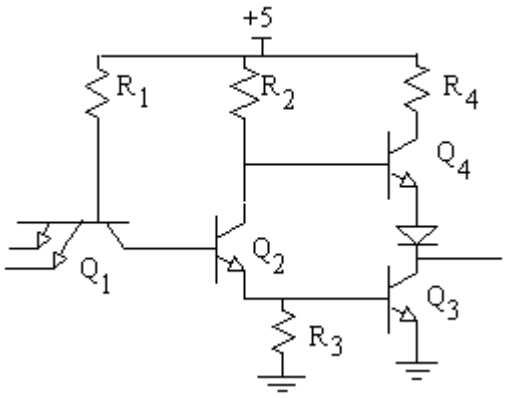
1. For the TTL NAND gate circuit shown draw a wire connecting both inputs to +5 volts. Then:

a. Indicate the output logic level: _____

b. Indicate the state of each transistor and the diode.

Q_1 _____, Q_2 _____, Q_3 _____, Q_4 _____, D _____

c. On the diagram, draw an arrow showing the direction of current in each branch.



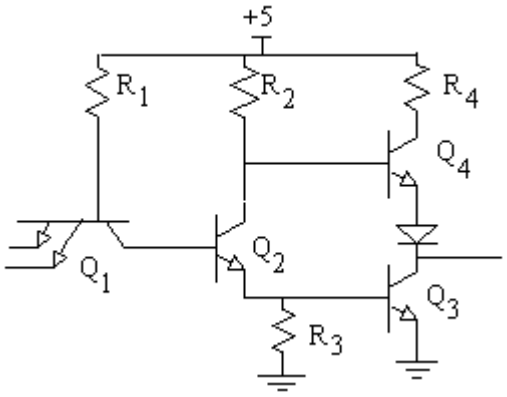
2. On the circuit below, draw wires connecting one input to ground and the other to +5 volts. Then:

a. Indicate the output logic level: _____

b. Indicate the state of each transistor and the diode.

Q_1 _____, Q_2 _____, Q_3 _____, Q_4 _____, D _____

c. On the diagram, draw an arrow showing the direction of current in each branch.



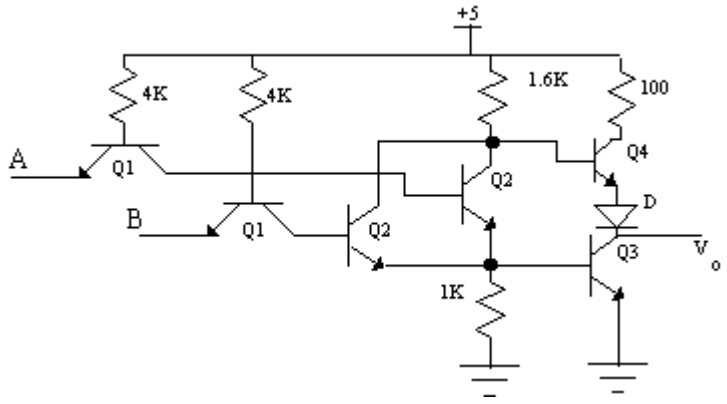
3. For the TTL NOR gate circuit shown draw a wire connecting both inputs to ground. Then:

a. Indicate the output logic level: _____

b. Indicate the state of each transistor and the diode.

Q_{1a} _____, Q_{1b}_____, Q_{2a} _____, Q_{2b}_____, Q_3 _____, Q_4 _____, D _____

c. On the diagram, draw an arrow showing the direction of current in each branch.



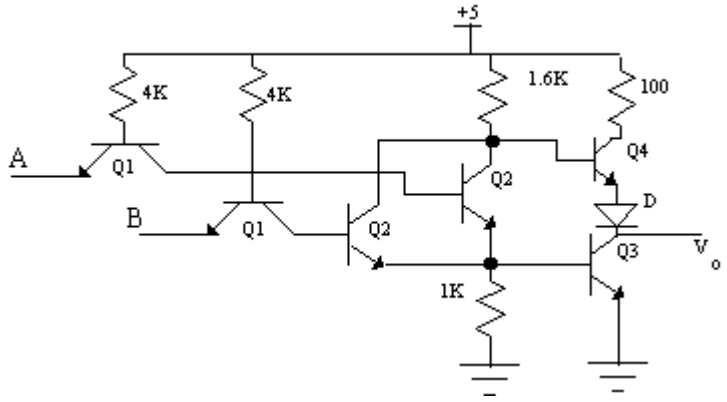
4. For the TTL NOR gate circuit shown draw a wire connecting the A input to +5 volts and the B input to ground. Then:

a. Indicate the output logic level: _____

b. Indicate the state of each transistor and the diode.

Q_{1a} _____, Q_{1b}_____, Q_{2a} _____, Q_{2b}_____, Q_3 _____, Q_4 _____, D _____

c. On the diagram, draw an arrow showing the direction of current in each branch.



5. Assume the following circuit is connected on a proto-board in the laboratory. You are troubleshooting the circuit. The gates are all 74LS00.

a. What voltage do you expect at each node, A-E?

b. What current would you expect in each branch, (beside each arrow)?

6. In most cases a resistor is used bewteen an input terminal and +5 volts to make a "1" at the input. Sometimes a resistor is used to pull-down the input to a zero. Both cases are shown below. What is the maximum value in each case?

a. $R_{\text{pull-up max}}$ _____ b. $R_{\text{pull-down max}}$ _____

7. Calculate the minimum resistor allowed in each case for the two 74LS00 gates below.

a. $R_{\text{pull-up min}}$ _____ b. $R_{\text{pull-down min}}$ _____

Problems

1. You are in the midst of designing a combination portable computer/walkman/Internet-cellphone which you intend to market and hopefully become a millionaire. To complete the design you need a 3-input NAND gate. The only store open at this time of night is the Handy-Combo Electronics/Oil Change/Hair Style/Pizza Parlor where you purchase the TTL gate shown below. Unfortunately, there are no specifications except that β_{sat} =15 and β_r =0.05. You must determine the specifications for this gate before you can use it. List the eight terminal current and voltage specifications in tabular form.

2. Make a compatibility chart for three TTL series shown in the data sheet in Figure 9 of the text. This chart is a matrix showing how many loads of a particular series can be driven by a driver of the same or another series, all of course, within specifications given by the manufacturer.

3. The two gates shown in the diagram labeled problem 3 are 74LS00. Gate A is used to drive gate B as well as the two-transistor buffer circuit. When the output of Gate A is high the load is turned off. When the output of gate A is low, the load is carrying current. Gate B should see valid logic levels and must operate properly. Determine the absolute maximum current that can be delivered through the load using the following characteristics for the buffer transistors.

$$V_{BEcutin} = 0.5 \text{ volts}, V_{BEsat} = 0.8 \text{ volts}, V_{CEsat} = 0.2 \text{ volts}, Beta_{sat} = 20$$

Problem 3 Problem 4

4. In the above logic system for problem 4, the output connection for each of the eight gates is numbered. If you were to measure the current in each of these wires, what is the maximum value you would expect to find? Use the specifications in Figure 9. For the load resistors connected to outputs 6 and 7, give the allowable range assuming valid logic output voltage levels. The symbols inside the gate indicate the family. No symbol indicates a 7400 gate.

5. Draw the figures showing the allowed operating regions, as shown in Figure 9 in the textbook for a 74S00 gate.

6. Laboratory Measurement

For a 7400 gate, make the following measurments:

a. No-load output voltage b. V_{inLmax} c. V_{inHmin}

For the last two measurement, describe the criteria you used to determine the limits.

7. A TTL gate has the following specifications:

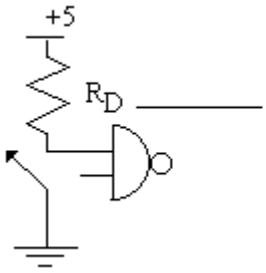
V_{inLmax} =1.00 V_{inHmin} =2.00 I_{inL} = - 0.500 I_{inH} = 1.00

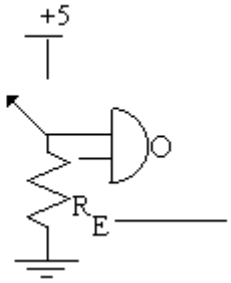
V_{oLmax} = 0.50 V_{oHmin} = 2.50 I_{oLmax} = 5.50 I_{oHmax} = -7.00

In both cases below, the gate must see a Low with the switch in one position and a High with the switch in the other position.

a. Determine the allowed values of resistor R_D in the circuit below.

b. Determine the allowed values of resistor R_E in the circuit below.





Chapter 7

EMITTER-COUPLED LOGIC

The major speed limitation of TTL is the turn-off time of saturated transistors. To be sure, TTL has come a long way from the 100 ns time of DTL to the 2-4 ns propagation delays of ASTTL. ECL is designed so that the transistors are either cutoff or in the active region, rather than cutoff or saturation for the DTL/TTL. With the transistors in the active region, the charge stored in the base region of the transistors is kept to a minimum, allowing shorter turn-off times. Typical propagation delays of "standard" ECL are about 1 ns, and down to about 0.5 ns for some of the advanced types.

There are several disadvantages associated with ECL. It uses a negative power supply so that the logic levels are not compatible with any other logic family, and makes analysis and measurement inconvenient. ECL requires large currents and the noise margins are small. On the other hand, power supply currents remain much more stable when the logic switches compared to TTL, thus reducing noise on the power leads. In practice, ECL is used only when necessary for its high speed. ECL has been around since the early 60's, being developed at about the same time as DTL.

Before we start the analysis of the circuit, we need to look at the circuit model we will use for the transistors in the ECL circuit. Because ECL is predicated on speed, the size of the transistors is as small as possible to keep stored charge to a minimum. Thus, current densities and hence, voltage drops will be slightly higher than normal. For this reason, we will use different voltages for base-emitter voltage for ECL.

$V_{BE\gamma} = 0.65$ and $V_{BEactive} = 0.75$ volts

ECL is based on the emitter coupled pair shown in Figure 1. This pair will have one transistor on and the other cutoff in each logic state. If we compare the currents through the two transistors, one with 0.75 volts between base and emitter and the other with 0.65 volts, we will find that the transistor with the higher voltage will be carrying nearly 50 times the current of the other. Thus, the lower voltage transistor will essentially be cutoff.

$$\frac{I_{B1}}{I_{B2}} = \frac{I_o e^{\frac{0.75}{V_T}}}{I_o e^{\frac{0.65}{V_T}}} = 47.5 \qquad \text{where } V_T = 25.9 \text{ mV}$$

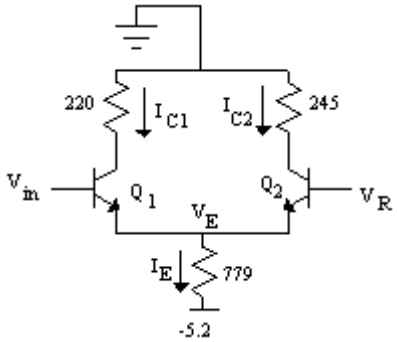


Figure 1. Emitter coupled Pair

BASIC OPERATION OF THE EMITTER-COUPLED PAIR

We start by looking at the operation of the difference amplifier shown in Figure 1. The two transistors are connected at their emitters. The base of Q_2 is connected to a reference voltage, V_R.

Consider for an example, that the input voltage is low enough to keep Q_1 cutoff. Then, Q_2 will be conducting and the emitter voltage will be

$$V_E = V_R - V_{BEon} = V_R - 0.75$$

assuming Q_2 is operating in the active region. We know Q_2 is in the active region because the collector resistors are selected to keep the transistors from saturating.

How high can we raise the input voltage and still keep Q_1 off?

$$V_{inLmax} = V_E + V_{BE\gamma} = V_R - 0.75 + 0.65 = V_R - 0.10 \text{ volts}$$

Now, let the input voltage start rising above this voltage. Then Q_1 begins to turn on. This increases the current through R_E, raising the voltage at the emitters. This voltage rise causes Q_2 to begin to turn off. The input voltage rises only a little bit before Q_2 turns off entirely. The minimum high level input voltage occurs when $V_E = V_R - 0.65$ and

$$V_{inHmin} = V_E + 0.75 = V_R + 0.10 \text{ volts}$$

Thus, a 200 mV swing at the input causes a complete reversal in which transistor is turned on and which is turned off.

ANALYSIS OF THE ECL GATE

We now turn our attention to Figure 2, the complete ECL gate circuit. In this circuit we have added two emitter follower output stages. As we shall see later, these output stages serve the purposes of buffering the gate output, providing added fanout, as well as shifting the voltage levels.

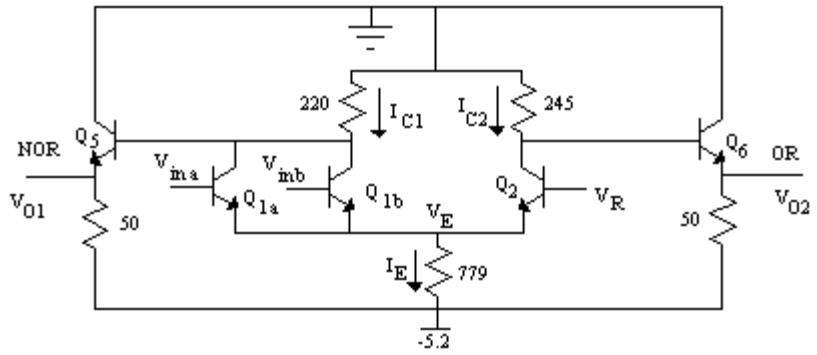


Figure 2. Complete ECL Gate

OR OUTPUT --- Output high, Q_2 off
With Q_2 off, the output circuit looks like that shown in Figure 3.

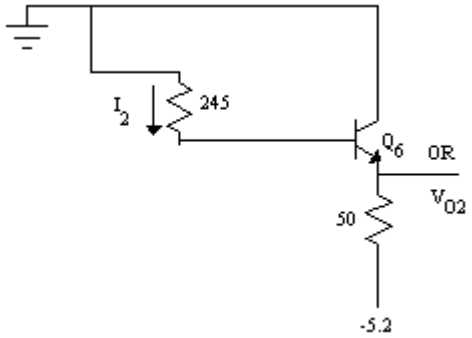


Figure 3. Equivalent Circuit With Q_2 OFF.

If we assume a very high β, we can ignore base current ($I_2 \cong 0$) but the transistor Q_6 is in the active region so

$$V_{oH} = V_{o2} \cong -0.75 \text{ volts}$$

(Note that since the collector is at 0V, and the emitter is at -0.75V, the transistor is in the active region. This type of circuit is called an emitter follower. The output voltage is always $V_{BEactive}$ below the base voltage and follows the base voltage, even when the output goes low as we shall see next.)

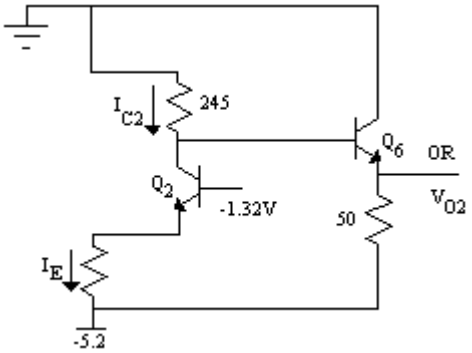


Figure 4. Equivalent Circuit With Q_2 ON.

OR OUTPUT LOW - Q_2 ON, Q_1 OFF,

The circuit for this case is shown in Figure 4. The emitter voltage is:

$V_E = V_R - 0.75 = -2.07$ volts

Thus, the current down through the emitter resistor is

$$I_E = \frac{-2.07 - (-5.2)}{779} = \frac{3.13}{779} = 4.018mA$$

If we assume β is very large so base currents can be neglected, the collector current in Q_2 will be the same as the emitter current. Thus, the voltage at the collector of Q_2 (and the base of Q_6) will be

$V_{C2} = V_{B6} = -I_{C2}*245 = -0.98$ (Note that $V_{C2} - V_E = 1.09$v. Q_2 is in active region.)

Thus, the voltage at the emitter of Q_6 will be

$V_{oL} = V_{o2} = -0.98 - 0.75 = -1.73$ volts

We can now plot the output at V_{o2} versus the input voltage. The result is shown in Figure 5.

NOR OUTPUT, Q_1 OFF, Output High

This case is essentially the same as for the OR output when it is high, except that we have a 220 Ω resistor instead of a 245 Ω resistor at the collector. Again, because the base current is so small, the output voltage will be -0.75 volts, the same as for the OR output.

NOR OUTPUT, Q_1 ON, Q_2 OFF, Output Low.

This case is similar to the OR output high case, except that the input voltage can directly affect the collector current through Q_1 causing the voltage at the base of Q_1 to change. As V_{in} increases, the current increases causing the voltage at the base of Q_5 to decrease. Because the output follows the base voltage with a 0.75 volt difference, the output drops as V_{in} increases. This scenario holds until Q_1 saturates. Then as V_{in} increases further, the base-collector junction is forward biased, and current flows from the input toward the collector, raising the voltage at the base of Q_5, hence raising the output voltage. The resulting plot of V_{o1} vs V_{in} is also shown in Figure 5.

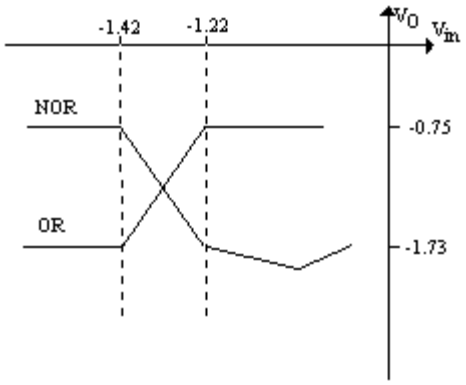
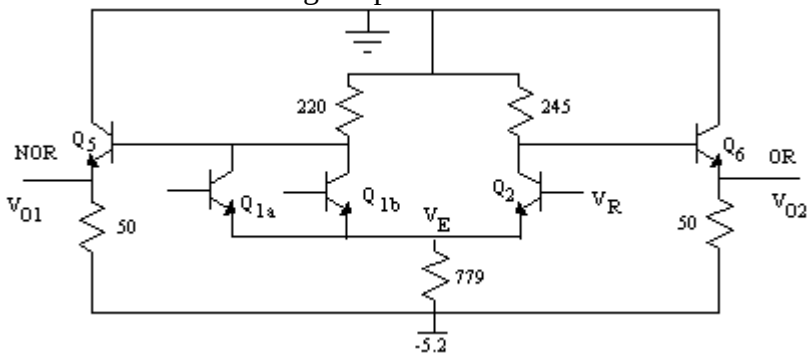


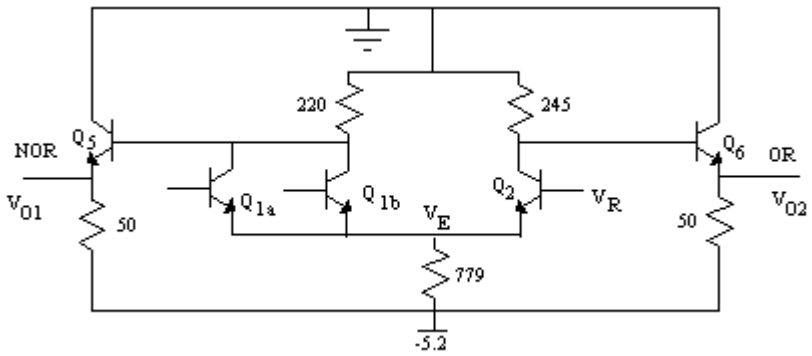
Figure 5. Voltage Transfer Characteristic For The ECL Gate

EXERCISES

1. For the 2-input ECL gate below, write a logic LOW at both inputs.
   a. Beside each transistor, write the state of the transistor.
   b. At the two outputs, write the logic level.
   c. Draw arrows showing the paths of current in the circuit.



2. For the 2-input ECL gate below, write a logic LOW at input a, and logic HIGH at input b.
   a. Beside each transistor, write the state of the transistor.
   b. At the two outputs, write the logic level.
   c. Draw arrows showing the paths of current in the circuit.



For the following exercise, assume $V_{BE\gamma} = 0.65$ volts, $V_{BEact} = 0.75$ volts, and β is very large (assume base currents are negligible).

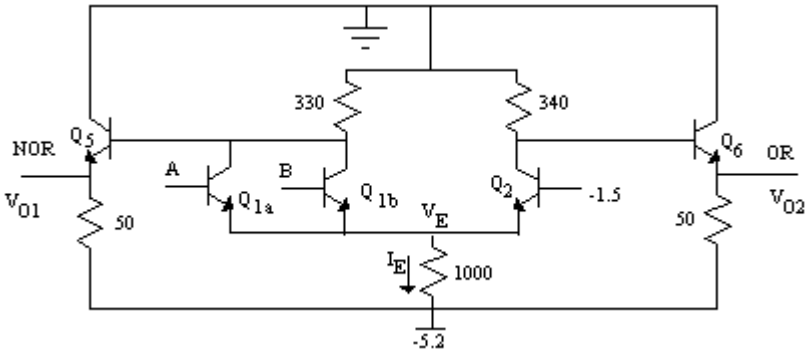
3. For the ECL gate below:

   Assume Both A and B inputs are low find:

   a. V_E = _____ b. V_{inLmax} = _____ c. I_E = _____

   d. I_{330} = _____ e. I_{340} = _____ f. V_{C2} = _____

   g. V_{C1} = _____ h. V_{O1} = _____ i. V_{O2} = _____

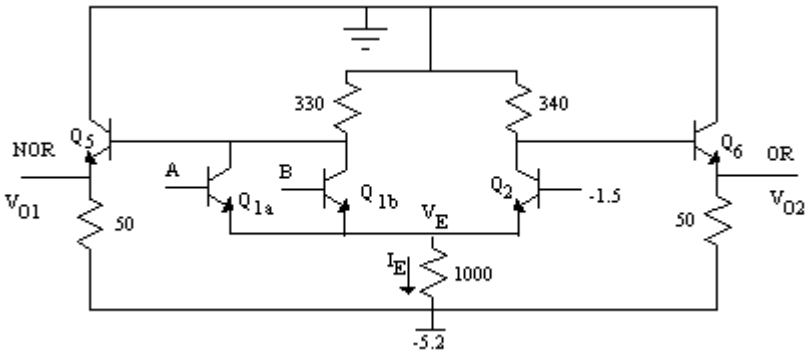


4. For the ECL gate, assume input A is high and input B is low:

   a. V_E = _____ b. V_{inHmin} = _____ c. I_E = _____

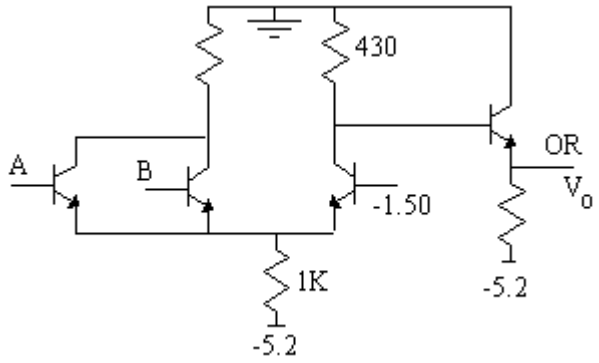
   d. I_{330} = _____ e. I_{340} = _____ f. V_{C2} = _____

   g. V_{C1} = _____ h. V_{O1} = _____ i. V_{O2} = _____



Problems

1. An ECL gate is shown below. Sketch the V_o vs V_{in} characteristic for the OR output. Show the values at all breakpoints. Assume V_{BEact} = 0.75 and the transistor is effectively non-conducting at V_{BE}=0.65V. Also assume beta is very high so that base current can be neglected.



Chapter 10

Advanced CMOS Circuits

Transmission Gates

NMOS Transmission Gate

The active pull-up inverter circuit leads one to thinking about alternate uses of NMOS devices. Consider the circuit shown in Figure 1. Here we have an NMOS transistor with its substrate connected to ground rather than the source terminal. In effect, the transistor is symmetric and can conduct in either direction. In order to turn it on, the gate voltage must be higher than either terminal by the threshold voltage. To turn the transistor off, the gate voltage must be lower than both terminal voltages plus the threshold voltage.

On: $V_G > V_{in} + V_T$ or $V_G > V_{out} + V_T$ ($V_G - V_T > V_{in}$ or $V_G - V_T > V_{out}$)

Off: $V_G < V_{in} + V_T$ and $V_G < V_{out} + V_T$ ($V_G - V_T < V_{in}$ or $V_G - V_T < V_{out}$)

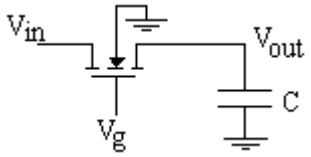


Figure 1. NMOS Transmission Gate

Consider the following three cases

Case I:

$V_G = 0$, $V_{in} = 0$ or 5 volts, $V_{out} = 0$ or 5 volts

In this case, the transistor will be cutoff because the gate voltage is not higher than either terminal. Whatever voltage is on the capacitor will stay there. In other words, the capacitor will hold its charge.

Case II:

$V_G = 5$ volts, $V_{in} = 5$ volts, $V_{out} = 0$ volts (initially)

In this case, the gate voltage is higher than V_{out} by more than the threshold. Thus, the transistor will conduct and the capacitor will be charged up. The output voltage will rise until it reaches $V_o = V_G - V_T$. Note that this is the same as the output voltage of the active pull-up NMOS inverter.

Case III:

$V_G = 5$ volts, $V_{in} = 0$ volts, $V_{out} = 5 - V_T$ volts (initially)

Here the gate voltage is higher than V_{in} so that the transistor is turned on and the capacitor will discharge. It will discharge all the way down to zero because $V_G > V_{in} + V_T$.

Generalizing these cases, we can see that with the gate voltage low, the output voltage will not change. In other words, the gate is off. If the gate voltage is high, the transmission gate will be on and the output will follow the input voltage. In this case, the output voltage is constrained to be less than the gate voltage by the threshold voltage. It should be noted that all terminal voltages are assumed to be more positive than the substrate voltage. We normally assume that the gate voltage will switch between the power supply rails.

PMOS Transmission gate

the NMOS transmission gate has the disadvantage that the output voltage is constrained to be less than the gate voltage. We will look to see how the PMOS gate in Figure 2 compares.

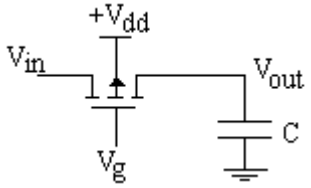


Figure 2. PMOS Transmission Gate

In this case, the substrate is connected to V_{DD} so that the gate will turn on if the gate voltage is low and off when the gate voltage is high. The conditions for the PMOS transmission gate are:

On: $V_G < V_{in} - V_T$ or $V_G < V_{out} - V_T$

Off: $V_G > V_{in} - V_T$ and $V_G > V_{out} - V_T$

We look at the same three cases

Case I:

$V_G = 5$, $V_{in} = 5$ or 0, $V_{out} = 0$ or 5 volts

In this case, the transistor will be cutoff because the gate voltage is not lower than either terminal

Case II:

$V_G = 0$ volts, $V_{in} = 5$ volts, $V_{out} = 0$ volts (initially)

In this case, the gate voltage is lower than V_{in} by more than the threshold. Thus, the transistor will conduct and the capacitor will be charged all the way up to 5 volts. The transistor always stays on because the gate voltage always remains lower than the input voltage by more than the threshold voltage.

Case III:

V_G = 0 volts, V_{in} = 0 volts, V_{out} = 5 volts (initially)
Here the gate voltage is lower than V_{out} so that the transistor is turned on and the capacitor will discharge. It will only discharge down to V_T because at that point, $V_{out} = V_G + V_T$, which turns the transistor off.

The PMOS has a similar, but symmetric characteristic to the NMOS when used as a transmission gate. Next we combine the two to develop an ideal transmission gate.

CMOS Transmission Gate

A CMOS transmission gate is shown in Figure 3. Included in the gate is an inverter (CMOS, of course) so that the NMOS and PMOS gates get complementary signals. When the control voltage, V_C is low, the NMOS gate gets a low voltage and is turned off. The PMOS gate gets a high voltge and is likewise turned off. When the control voltage is high, the gate is turned on with both transistors able to conduct. As discussed above, each cannot conduct all the way to both power supply rails, but they do so at either end so they complement each other and together they can conduct to the rails.

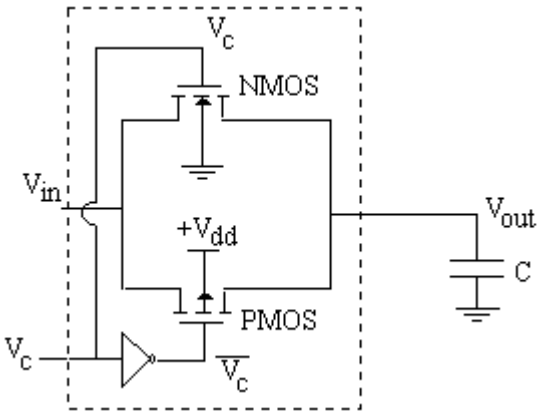


Figure 3. CMOS Transmission Gate

Transmission Gate Resistance

When the gate is off, the resistance of the gate is theoretically infinite. However, there is always some conduction so in reality, the resistance is finite. The off resistance is given in the specifications for the device and is typically in the megohms. When the gate is on, however, we would like the resistance to be low. To examine this case, consider the circuit in Figure 4. Here we look at the case where the input voltage is V_{DD}, and the gate is turned on with the output voltage starting at zero. We look at the resistance as the capacitor is charged and the output voltage rises to V_{DD}.

The voltage across both devices is $V_{DD} - V_o$, and the resistance of each is simply the voltage across it divided by the current through it, I_{DS} for the NMOS of I_{SD} for the PMOS.

$$R = \frac{V_{DD} - V_o}{I}$$

The total equivalent resistance is the parallel combination of the the two.

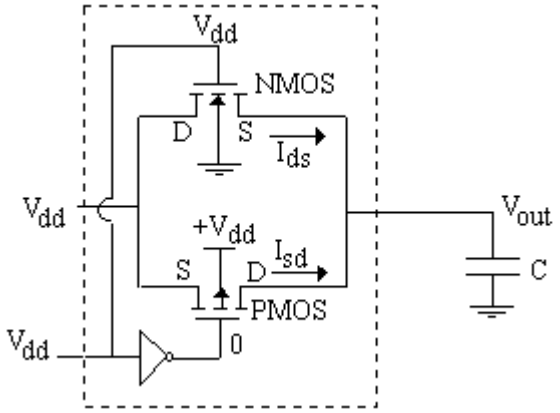


Figure 4. CMOS transmission gate used to calculate resistance.

We start with $V_{out} = 0$. Here, the NMOS is in pinchoff

$$I_{DS} = k_n(V_{DD} - V_T)^2$$

$$R_n = \frac{V_{DD} - V_o}{k_n(V_{DD} - V_o - V_T)^2}$$

The PMOS is also in pinchoff

$$I_{SD} = k_p(V_{DD} - V_{Tp} - 0)^2 = k_p(V_{DD} - V_{Tp})^2$$

$$R_p = \frac{V_{DD} - V_o}{k_p(V_{DD} - V_{Tp})^2}$$

The NMOS stays in pinchoff up until $V_o = V_{DD} - V_{Tn}$ at which time it cuts off. The PMOS, however, switches to triode when the output voltage rises to V_{Tp}. At this point, V_{SD} is $V_{DD} - V_{Tp}$ and $V_{SG} = V_{DD} - 0 = V_{DD}$. The criterion for triode region is

$$V_{SG} - V_{Tp} < V_{SD}$$

Thus, when V_o rises to V_{Tp}, the PMOS switches to triode and

$$I_{SD} = k_p(2(V_{DD} - V_{Tp})(V_{DD} - V_o) - (V_{DD} - V_o)^2)$$

$$R_p = \frac{V_{DD} - V_o}{k_p(2(V_{DD} - V_{Tp})(V_{DD} - V_o) - (V_{DD} - V_o)^2)}$$

$$R_P = \frac{1}{k_P(2(V_{DD} - V_{Tp}) - (V_{DD} - V_o))}$$

We have defined three distinct regions. If we assume $V_{DD} > V_{Tp} + V_{Tn}$, the regions are as shown in Figure 5.

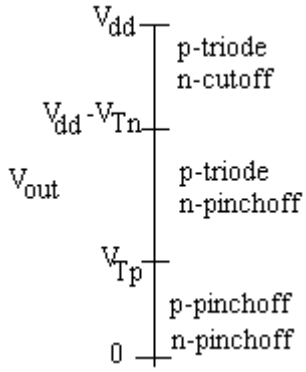


Figure 5. The three regions of operation
for the CMOS transmission gate in positive transition

Combining the two resistances in parallel for each region, we can plot the equivalent resistance over the transition at the output from 0 to V_{DD}. The results in Figure 6 are ploted for the device parameters shown. The three curves are for the NMOS device, the PMOS device, and the parallel combination. The combined resistance is nearly constant over the entire region, a characteristic very helpful when designing. In this case, the designer usually assumes the device can be modelled by a fixed resistance when on and a very high resistance when off.

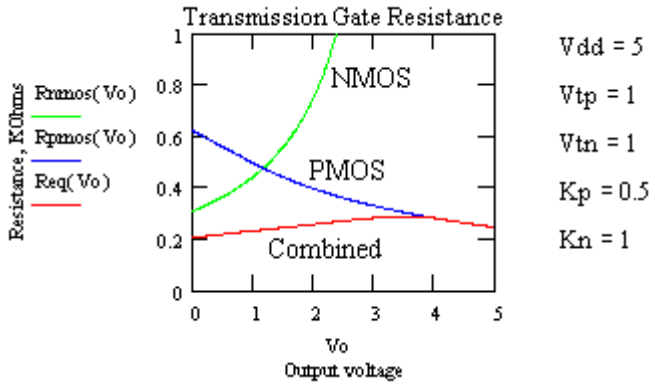


Figure 6. Transmission gate resistance as a function of output voltage.

Logic Using Transmission Gates

For the purposes of simplicity, we will represent a CMOS transmission gate with the symbol in Figure 7. This device uses 2 transistors

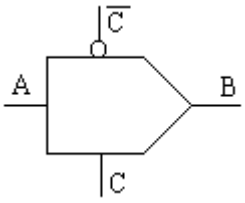


Figure 7. Symbol for Transmission Gate

Many times, logic can be performed using transmission gates more efficiently than using standard CMOS logic gates. Efficiency in this case is the number of transistors. For example, the two input multiplexer using 2-input CMOS gates uses 10 transistors while it uses only 6 transistors using CMOS transmission gates as shown in Figure 8.

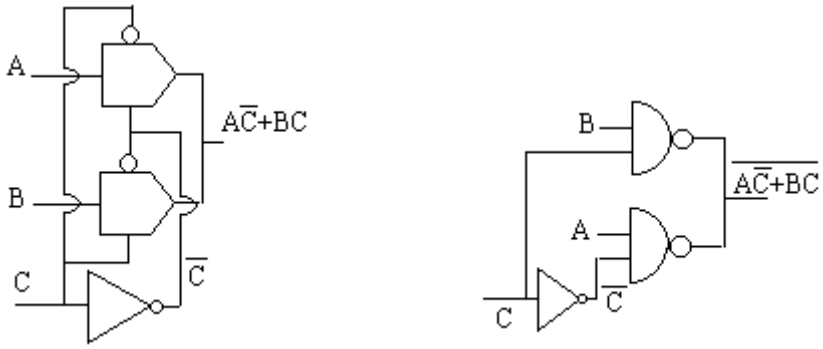


Figure 8. Two implementations of a 2-input multiplexer

In Figure 8, We assume the CMOS 2-input NAND gates each take 4 transistors and the CMOS inverter takes two transistors. In the transmission gate implementation, we assume the transmission gate takes two transistors for the gate and two transistors for the inverter. We are assuming that the inverters in the transmission gate implementation can be shared on a single integrated circuit.

The implementation of an exclusive or function is shown in Figure 9. The CMOS logic gate implementation takes 16 transistors while the transmission gate implementation takes 8 transistors.

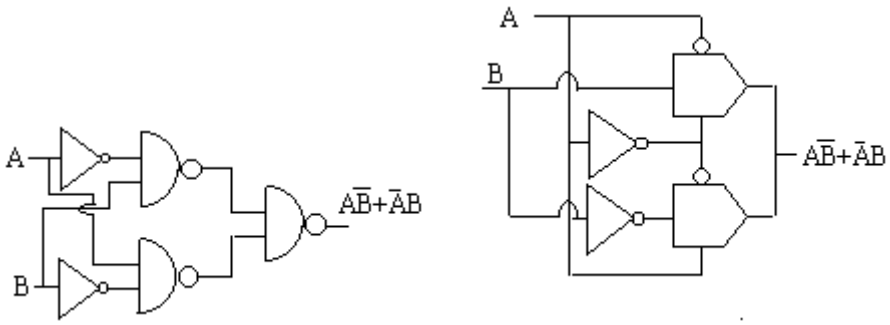


Figure 9. Logic gate implementation and transmission gate implementation of an XOR.

Not all functions can be made more efficient using transmission gates. For example, a logic AND and a logic NAND made with transmission gates are shown in Figure 10. A CMOS AND requires 6 transistors while a CMOS NAND requires only 4 transistors. The transmission gate implementations require 4 and 6 respectively.

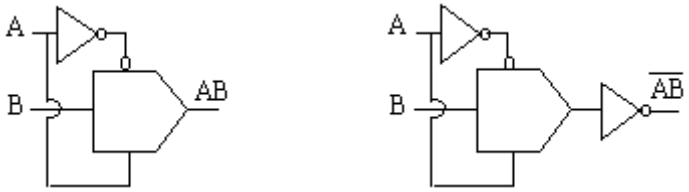


Figure 10. Transmission gate implementation of logic AND and NAND.

BiCMOS

The major advantages of CMOS logic is its low power dissipation, simple fabrication process, and their abiltity to be made extremely small in size. Because CMOS gates are made as small as possible, their current drive capability is quite low and thus their greatest limitation is their limited ability to drive capacitive loads. This limitation becomes serious when implementing large scale memories or when driving loads external to the integrated circuit chip.

The BiCMOS circuit was devised to take advantage of the bipolar transistor's high current drive while still minimizing static power dissipation. A typical BiCMOS driver circuit is shown in Figure 11 with a capacitive load.

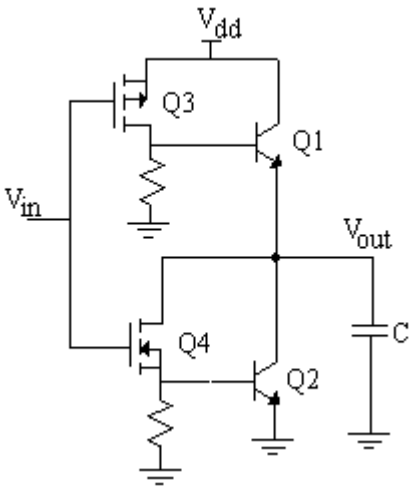


Figure 11. Simple BiCMOS inverter.

In the simple inverter in Figure 11, Q_1 and Q_2 are NPN bipolar transistors. Q_3 is a p-channel MOSFET, while Q_4 is an n-channel MOSFET. When the input voltage is low, the n-channel is turned off and Q2 is turned off. The p-channel is turned on. The base of Q_1 is pulled up toward V_{DD} and thus, turns Q_1 on which in turn charges the load capacitor and drives any other high-level loads. When the output voltage rises high enough (V_{DD} - $V_{BE\gamma}$), Q_1 will be turned off.

When the input is high, Q_3 and Q_1 are both turned off, while Q_4 is turned on. A high voltage at the load will cause current through Q_4 which causes the voltage at the base of Q_2 to rise up and turn on Q_2. Thus, the load capacitance can be quickly discharged through Q_2. When the load voltage drops low enough, Q_2 will turn off and the load continues to discharge through Q_4.

In both cases, when the load capacitance has been charged or discharged sufficiently, the bipolar transistors are turned off and there is no steady state current, preserving the low static power dissipation of CMOS logic.

When the input is high, Q_3 and Q_1 are turned off. The resistance to ground allows any excess base charge to leak off to ground, allowing the transistor Q_1 to turn off quickly. Similarly, the lower resistor can turn off Q_2 quickly when the input is low. These turn-off times are important when there is a static load, and are not normally a problem with fully charged or discharged capacitive loads. The circuit shown in Figure 12 uses n-channel MOSFETs instead of resistors to provide even quicker discharge of excess base charge and thus, faster switching.

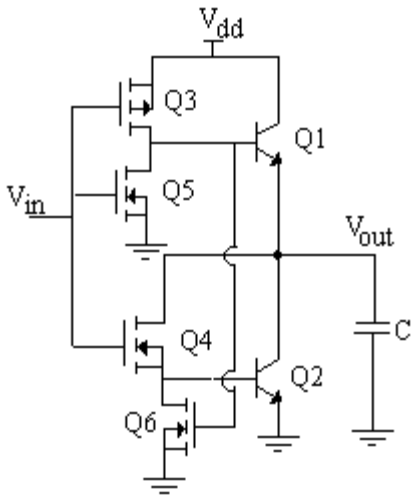


Figure 12. BiCMOS inverter circuit.

Reference:
CMOS Digital Integrated Circuits, second Edition by Kang and Leblebici, McGraw-Hill, 1999

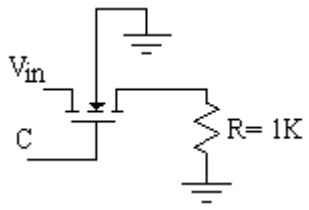
Exercises

1. A simple NMOS transmission gate is shown for Problem 1.

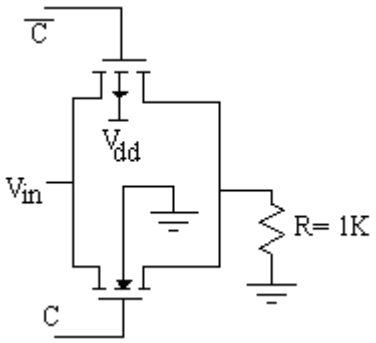
   V_{Tn} = 1 volt, k_n = 0.4 mA/V^2, V_{DD} = 5 volts

   To make the transmission gate conduct (turn it on) the control voltage should be:
   (High, Low)

   With the gate on, what is the output voltage when the input voltage is V_{DD}?
   V_{out} = _____



Exercise Problem 1.



Exercise Problem 2.

2. A CMOS transmission gate is shown for Problem 2.

$V_{Tp} = V_{Tn} = 1$ volt, $k_p = k_n = 0.4$ mA/V^2, $V_{DD} = 5$ volts

To make the transmission gate conduct (turn it on) the gate voltage for the n-channel should be: (High, Low)

To make the transmission gate conduct (turn it on) the gate voltage for the p-channel should be: (High, Low)

With the gate on, what is the output voltage when the input voltage is V_{DD}?

V_{out} = _____

What is the current through the p-channel?

I_p = _____

What is the current through the n-channel?

I_n = _____

3. The BiCMOS circuit shown as problem 3 below has a resisive load and the input is at 0 volts. What is the state of each device in the circuit?

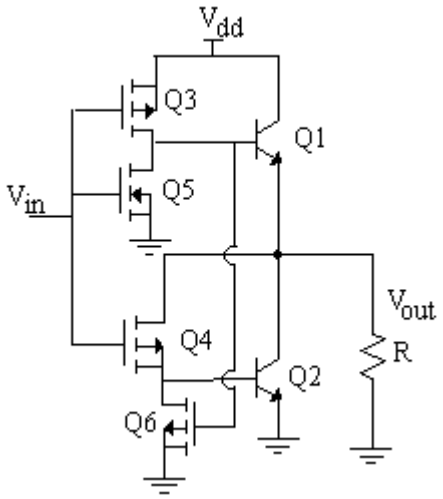
Q_1 _____, Q_2 _____, Q_3 _____, Q_4 _____, Q_5 _____, Q_6 _____

Draw all currents in the circuit.

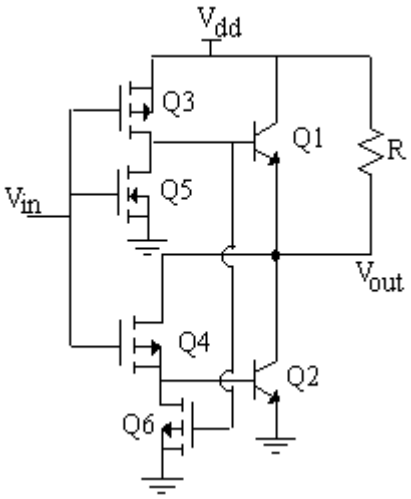
Write an equation for each current in the circuit in terms of the device parameters: β, V_{BEact}, V_{BEsat}, k_n, k_p, V_{Tn}, V_{Tp}, V_{DD}
All equations should be written in terms of these parameters and V_o.

Write the nodal equation necessary to solve this circuit.

If the load was capacitive instead of resistive, write the nodal equation.



Problem 3



Problem 4

4. The BiCMOS circuit shown as problem 4 above has a resisive load and the input is at V_{DD}. What is the state of each device in the circuit?

   Q_1 _____, Q_2 _____, Q_3 _____, Q_4 _____, Q_5 _____, Q_6 _____

   Draw all currents in the circuit.

   Write an equation for each current in the circuit in terms of the device parameters: β, V_{BEact}, V_{BEsat}, k_n, k_p, V_{Tn}, V_{Tp}, V_{DD}
   All equations should be written in terms of these parameters and V_o.

   Write the nodal equation necessary to solve this circuit.

5. For the BiCMOS gate shown for Problem 4, the resistor is replaced by a capacitor to ground. The initial voltage on the capacitor is $V_C(0)$. Rewrite the nodal equation to needed to solve the circuit.

Chapter 11
Analog-Digital Conversion

One of the common functions that are performed on signals is to convert the voltage into a digital representation. The converse function, digital-analog is also common. We will start with the digital-analog, D/A conversion.

A simple D/A conversion circuit is shown in Figure 1. This circuit is essentially an operational amplifier inverting, summing circuit followed by a unity gain inverter. Each switch connects to zero volts, or $+V_R$. The output voltage is

$$V_{out} = R_f(V_0/R_0 + V_1/R_1 + V_2/R_2 + V_3/R_3 \ldots + V_{N-1}/R_{N-1})$$

where V_0 is the voltage at the switch 0, which is either $+V_R$ or 0, etc..

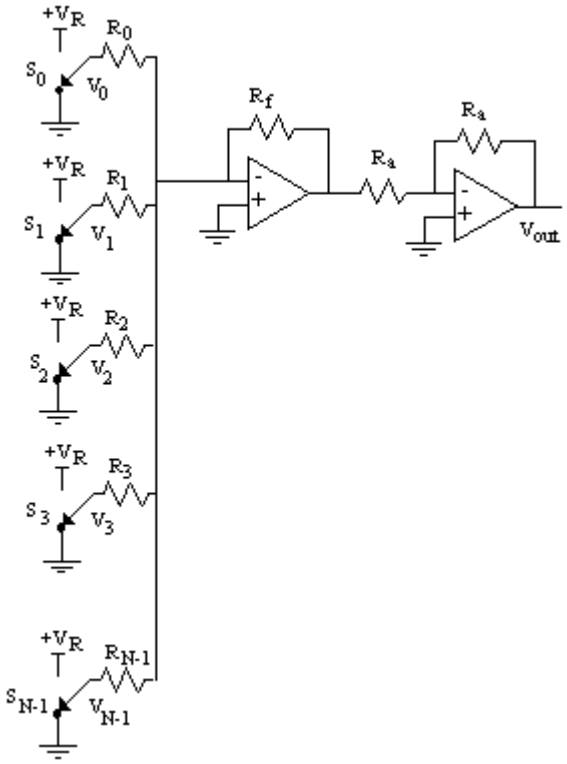


Figure 1. Simple digital to analog converter

This representation is compared to a binary representation of a number with N bits.

$$\text{Binary Value} = 2^{N-1}a_{N-1} + \ldots + 2^3a_3 + 2^2a_2 + 2^1a_1 + 2^0a_0$$

where a_0 can take on either 0 or 1, etc.

If N = 8, then the binary number can take on any value between 0 and 255. In the D/A circuit, there would be 8 switches and input resistors. We want 256 discrete steps, 0-255. The smallest or least significant step is controlled by switch 0. The output voltage from

switch 0 is $V_R(R_f/R_0)$, while the maximum output voltage is 255 times as high, when all switches are connected to V_R. Thus, $R_0 = 2R_1$, $R_1 = 2R_2$, etc., so that each switch contributes exactly twice as much to the output voltage as the previous switch, and $R_0 = 128R_7$.

The total range is 0 - $255V_R(R_f/R_0)$. We know that operational amplifier outputs cannot exceed the rails, or power supply voltages, so $255V_R(R_f/R_0)$ must be less than the power supply voltage. This limitation determines the relationship between the resistors and V_R. For example, if we wish to limit the range to 10 volts with $V_R = 5$ volts, then $R_f/R_0 = 2/255$, or $R_0 = 127.5R_f$. To make numbers convenient in the binary system, we usually make $R_0 = 128R_f$. The resistor ratios in the input string are related by factors of two so that most significant resistor, $R_7 = 1R_f$.

Resistor values used in operational amplifier circuits are typically in the range of 1-100 KOhms. Lower values require large currents while large resistor values may cause errors due to bias currents and offset voltages. With high quality operational amplifiers, it is possible to use this range of 1-128 without serious problems, but extending the number of digits may cause significant errors which may exceed the least significant values.

One way to overcome the large range in resistor values is to use a ladder network or the R/2R network shown in Figure 2. In the ladder, only two values are used, R and 2R. The overall gain is controlled by R_a and R_b.

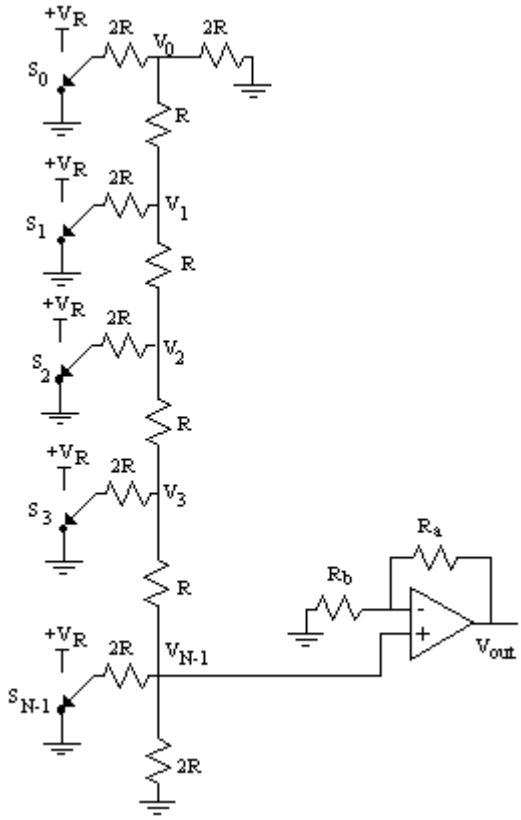


Figure 2. Ladder type D/A converter

We will assume the op-amp has infinite input impedance and all switches are connected to ground so the resistance seen looking any direction from any node is 2R as seen in Figure 3. Here we see that the two resistors looking left and right from node V_0, are 2R. The parallel combination is R, so that the resistance looking up from node V_1 is also 2R. This process can be carried out down the entire string. Similarly, the same process is used to see that the resistance looking down from each node is also 2R.

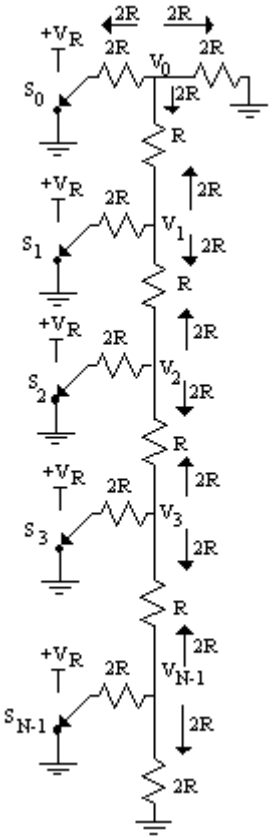


Figure 3. Ladder network showing resistances seen looking away from each node

If switch S_{N-1} is connected to $+V_R$, the voltage at node V_{N-1} is $1/3V_R$. This voltage is amplified by the op-amp as seen in Figure 2. Similarly, the voltage $V_3 = 1/3V_R$ when switch V_3 is up. However, the voltage $V_{N-1} = 1/2V_3$. Thus, the switch S_3 contributes only half the voltage to the op-amp. Similarly, each switch above, contributes successively less voltage by factors of two at each step.

Up to now, we have simply shown switches in the A/D converters. However, solid state switches are universally used. These switches are most commonly CMOS switches as shown in Figure 4. The control voltage ussed is the same as the switched voltages to guarantee low resistance connections with the MOSFETS in the triode region when on.

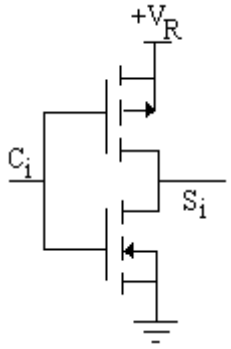


Figure 4. CMOS switch used in A/D converters

Exercises

1. For a 4-bit D/A converter as shown in Figure 1, assume V_R is 5 volts, and we want the range to be approximately 0-5 volts. the op-amp feedback resistor is R_f = 2.2 kOhm. Determine the necessary values of the input resistors. What is the resolution, or the smallest step? What is the highest output voltage?

2. A four-bit ladder type D/A converter is to be designed. Assume V_R is 5 volts, and we want the range to be approximately 0-5 volts. What is the necessary gain of the op-amp?

Analog - Digital Conversion

Analog - digital conversion is the opposite to the process described above. In this case we want to convert an analog voltage to a digital representation. We see this function performed in digital voltmeters where a voltage is converted to digital and then shown in LED or LCD numerical displays.

If we assume the input voltage can take on any value in the allowed range, such as 0-5 volts, the A/D converter will interpret that range in discrete steps. For example, if we are using an 8-bit converter, there are 256 steps within the 0-5 volt range. Each step represents a voltage range of 5/256 = 0.0195 volts, or about 20 mV. This discreteness means that any voltage between 0 and 19.5 mV will be represented as 00000000, while between 19.5 and 39 mV will be represented as 00000001, etc.

The most straightforward method of doing an A/D conversion is to use a series of voltage comparators as shown in Figure 5. Here a string of resistors generates a string of voltage references for the comparators. The output of each comparator will switch to a high when the input analog voltage is above the reference voltage. There is a comparator for each step in the digital output above the zero step. Thus, in the case of an 8-bit conversion, there would be 255 comparators. These 255 signals would normally be encoded into the 8-bit digital representation. This method is very fast and is called a flash converter, but is impractical for all but the most critical applications where more than just a few steps are needed.

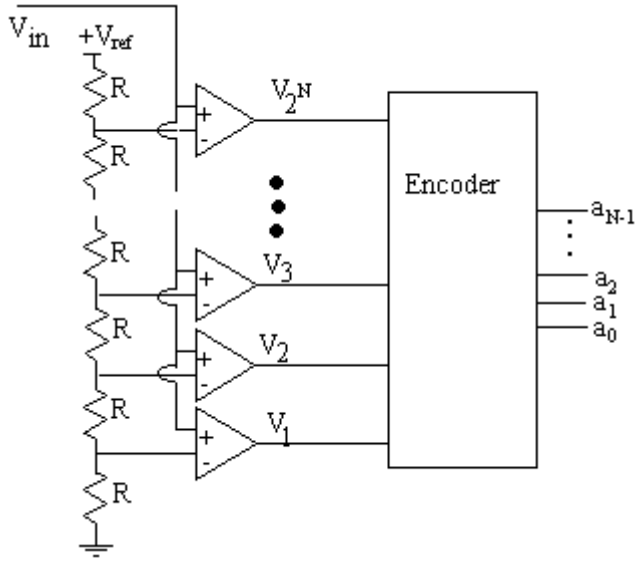


Figure 5. Flash A/D converter

A more practical, and more common, A/D converter using a counter and a D/A converter is shown in Figure 6. In this circuit, the counter starts at zero. The output of the counter goes to the D/A converter which presents its output to the voltage comparator. If the input voltage is above the output of the D/A, the counter counts up. As the conter counts up, the output of the D/A goes higher. When the D/A output goes above the input signal, the counter stops. At that point, the output of the counter is the digital representation of the input signal. The output of the voltage comparator can be used to tell the outside world when the conversion is completed. To take the next sample, the counter must be reset and the process starts over. The resolution is determined by the counter-D/A resolution and the range is determined by the range of output from the D/A converter. Of course, there must be a few more controls added such as stopping the counter when it reaches its maximum and an overflow signal enabled.

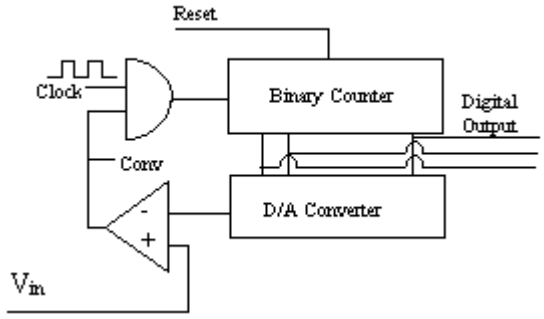


Figure 6. Counting A/D converter

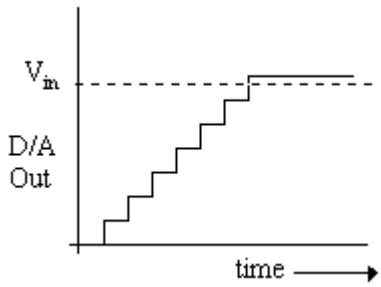


Figure 7. Output of the D/A module

The output signal of the D/A converter module would look like Figure 7. The counter starts at zero which gives zero output from the D/A. The counter counts until the D/A output rises above the input voltage.

The counting A/D converter does not resolve the conversion until the counter has counted far enough. Conceivably, this time could be 2^N-1 counts. The maximum clock speed is

limited to the time delays of the various modules; counter, D/A converter, comparator, and the AND gate. An 8-bit A/D with a 1 MHz clock could take 255 μs to resolve the output. This delay is very long compared to the flash A/D converter which would only take a few microseconds. We will look at two methods to try to speed this up; tracking A/D, and a successive approximation A/D

Tracking A/D

A tracking A/D converter uses an up/down counter and runs continuously. This type A/D converter is shown in Figure 8. The direction of the count is determined by the voltage comparator. Presumably, a change in the input analog voltage will be close to the previous voltage so the count does not take as long as a count from zero every time. This system has the added advantage of "tracking" the input voltage continuously. It also has the disadvantage that it continuously toggles around the input voltage.

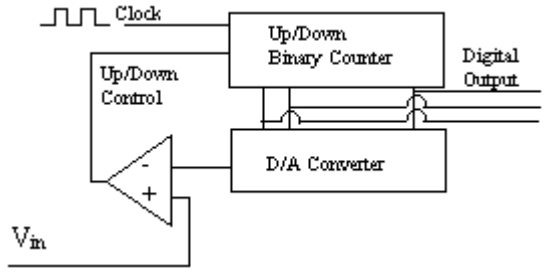


Figure 8. Tracking A/D converter

Successive Approximation A/D

The successive approximation converter uses the same diagram as the tracking A/D converter shown in Figure 8. However, the binary counter use additional logic to control the count sequence differently. The counter switches the most significant bit to a 1 first. If the output of the D/A is lower than V_{in}, the bit stays a 1, otherwise, it is switched back to a 0 on the next clock pulse. Then the counter goes through the same sequence for the second most significant bit, etc., until all bits have been switched and compared. This type of A/D takes a maximum of 2N clock pulses to resolve the output, much faster than the maximum time of the simple counting A/D.

Exercises

3. A simple counting 16-bit A/D converter uses a clock frequency of 2 MHz and has a range of 0-2 volts. What is the binary output when the input voltage is 1.5 volts? What is the settling time?

4. If the A/D converter in the previous problem is a successive approximation converter, what is the settling time?

BCD Converters

Typical digital voltmeters display the voltage in decimal representation. It would be appropriate for these voltmeters to maintain the digital representation in BCD to minimize encoding and decoding logic. The most logical way to handle BCD is to used BCD counters and BCD D/A converters in the feedback loop. BCD counters are well known, but we need to look at the coding of the D/A converters. In particular, we are interested in the resistor ladder networks that produce the BCD weighting.

A simple BCD D/A converter is shown in Figure 9. In this case, the least significant decimal digit has resistance values a factor of ten time the corresponding resistor values for the most significant digit. An alternative implementation is given in Figure 10. In this case, the ratio between the digits is accomplished by the second summing inverting amplifier.

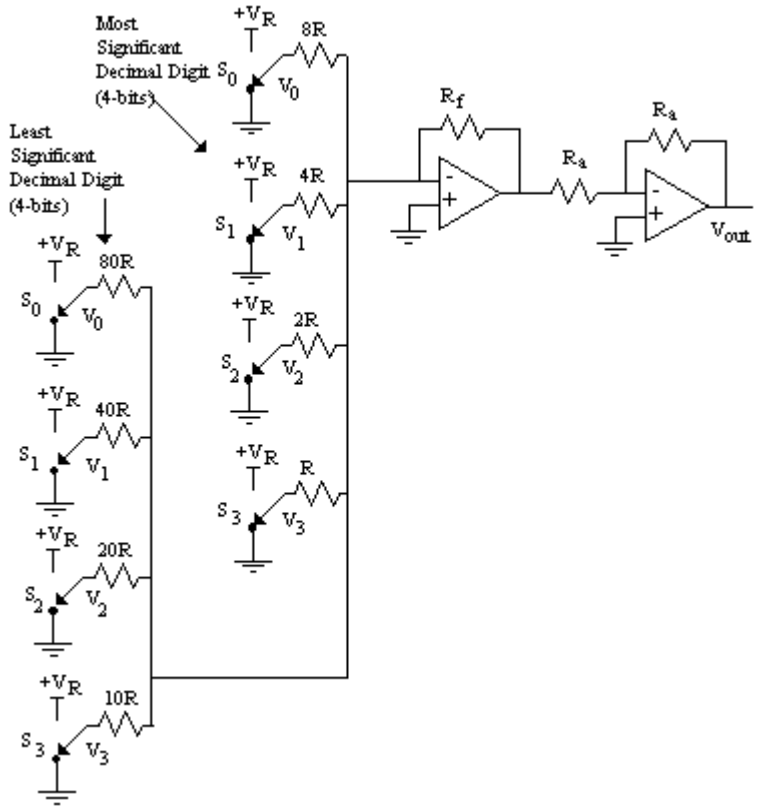


Figure 9. BCD D/A converter

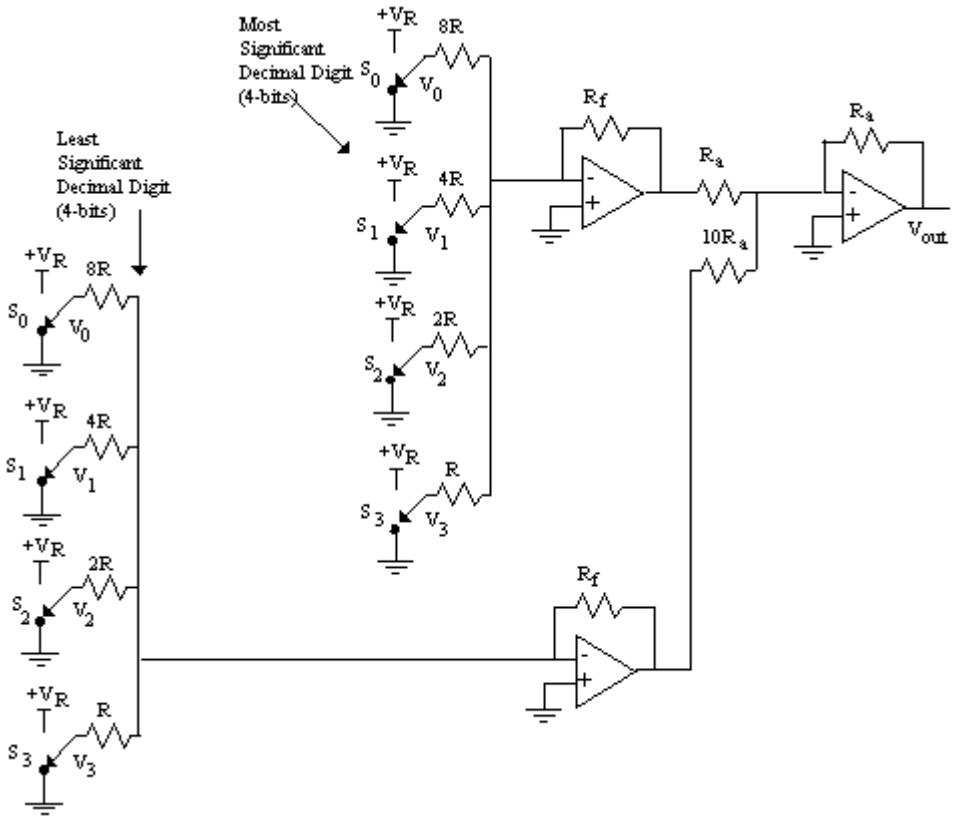


Figure 10. Alternative BCD D/A converter

Chapter 12

Common Switching Functional Blocks

Voltage Comparators

In many applications, it is necessary to cause a digital switching action when an analog voltage rises above or drops below some value. An example would be a case when we wanted a digital signal to turn on a "discharge" light when the battery voltage dropped below a specific point, say 12.5 volts for an automotive application. In this case, we would want a logic high (or low) when the battery voltage dropped below 12.5 volts.

We have seen how diodes and BJT base-emitter junctions as well as enhancement MOSFETs have thresholds where they beging to conduct. It would be possible to construct a circuit to create an output transition from low to high as the input voltage crossed the trigger point. As simple example would be an appropriate voltage divider to drop the battery voltage down to an appropriate level and feed it into the base of a transistor inverter followed by several more inverters to provide gain and make the output rail-rail transistion occur for a very small transition of the input signal.

A few years ago, discrete devices were used as discussed to create this voltage comparison function. However, integrated circuits allow the voltage comparator function to be done in a single integrated circuit chip. These chips are based on operational amplifiers circuits with a switching transistor at the output. A functional representation is shown in Figure 1.

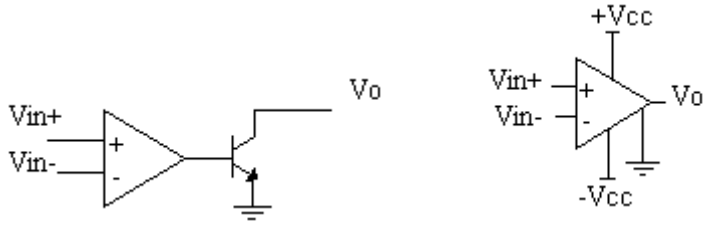


Figure 1. Functional representation of a voltage comparator and its symbol.

This circuit is simply an operational amplifier followed by a transistor whose emitter is connected to ground and whose collector is open. A typical application is shown in Figure 2. In this case a reference voltage is connected to the Vin- input or the inverting input. When the signal voltage at the Vin+ input rises above the reference, the output voltage goes high. What's really happening is that the output of the op-amp goes low, cutting off the transistor so the external pull-up resistor causes the output voltage to go high. When the input voltage goes below the reference, the output of the op-amp goes high, causing the transistor to saturate and pull the output low. If you reversed the connection of the two input signals, the output function would switch directions. All of the internal buffering and interfacing is taken care of within the op-amp part of the circuit.

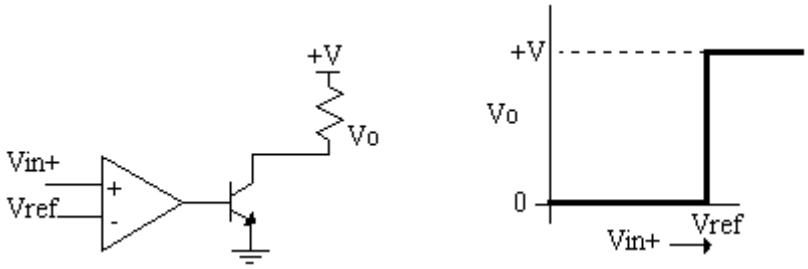


Figure 2. Typical application of a voltage comparator and its voltage transfer characteristic

The VTC shows an abrupt switch when the input voltage passes the reference voltage. The transition is not instantaneous, but occurs over a few millivolts of the input signal for most comparators. The LM311 whose data is given on page 100 of the "'Design Compendium", shows a minimum gain of 40 V/mV, or a gain of 40,000. Thus, for the example above, if +V is 5 volts, the transistion occurs with a change of 0.125 mV at the input.

The circuits shown above are typical of the LM319, and 339 voltage comparators. However, the LM311 is a little different in that the emitter of the output transistor is not connected to ground. It is left uncommitted, similar to the collector, so the user can have switching between other voltage ranges. An interface between positive voltage systems TTL or CMOS to the negative voltage ECL would be a typical application. It should be noted, however, that the emitter must be connected to a voltage within the +/- Vcc rails of the voltage comparator. Appropriate rails would be +5 and -5.2 volts of the two logic systems.

One problem using voltage comparators with very slowly varying input signals is output signal oscillation when the inputs are in close proximity of the switch point. All signals including power supplies have a smal amount of noise embedded on top of the desired signal. We saw earlier that the switching transition occurs with less than a millivolt change in the input voltage. If, for example, the input signal was just at the switch point and there was a small noise blip on it, the output would switch. This switching can, and often does, induce more noise on the system. This additional noise can cause the input signal to drop, causing another transition in the output. This second transistion can then again induce noise reversing the transisiton again. Such oscillations are frequently seen and cause many problem in the circuits.

Hysteresis

One way to prevent the spontaneouus oscillation of voltage comparator circuits is to introduce hysteresis as shown in Figure 3. In this case, two additional resistors are added to provide feedback. Typically, $R_2 >> R_3$ so only a small amount of feedback is provided. As far as the comparator is concerned, the voltage at the non-inverting input is the voltage that it sees and will control the output state.

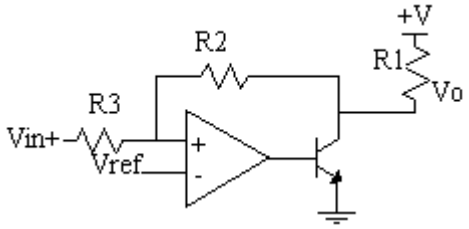


Figure 3. Voltage comparator with hysteresis

If Vin+ is just slightly below Vref, $V_o \approx 0$, the output transistor is saturated.

$$V_{noninv} = \frac{R_2}{R_2 + R_3} V_{in+}$$

As long as $R_2 >> R_3$, $V_{noninv} = V_{in+}$

If we now raise the input voltage slightly so the output switches state, then

$$V_{noninv} = \frac{R_2 + R_1}{R_3 + R_2 + R_1} V_{in+} + \frac{R_3}{R_3 + R_2 + R_1} V_{DD}$$

Again, as long as $R_1 + R_2 >> R_3$ then the first term is approximately V_{in+}. However, the second term adds to the voltage at the non-inverting terminal. Because R_3 is much smaller than the the sum of R_1 and R_2, this amount is quite small, but not zero. Typically, you want this amount to be a few millivolts to prevent oscillation.

Example:

Provide 10 mV of hysteresis with $V_{DD} = 5$ volts, $R_1 = 1$ kOhm, $V_{REF} = 2.00$ volts.

Solution:

We want $R_2 >> R_1$ so we choose $R_2 = 1$ MOhm.

Then to get 10 mV hysteresis,

$$10mV = \frac{R_3}{R_3 + R_2 + R_1} 5$$

$$\frac{R_3}{R_3 + 1.001MOhm} = \frac{0.01}{5} = .002$$

$$R_3 \approx 0.002x1MOhm = 2KOhm$$

If we use $R_3 = 2$ KOhm, the actual hysteresis is 9.97 mV.

The voltage transfer characterisitc with hysteresis looks like that shown in Figure 4. Here we see that the high output transistion occurs at 2.00 volts as would have been the case without hysteresis, but the input voltage must drop significantly lower by the amount of hysteresis than before to cause the comparator to switch back.

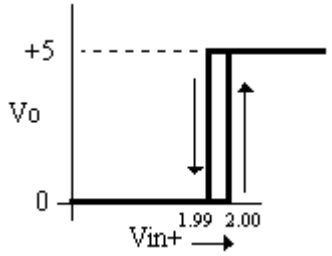


Figure 4. VTC with hysteresis for example.

Exercise

Draw the voltage transfer characterisitc for the voltage comparator circuit in Figure 5. Assume the transistor acts as a perfect switch, open or short. Note in this case, the reference is on the non-inverting input to the comparator and the input voltage is on the inverting input. The reference is supplied by a voltage divider from the power supply. (Hint: Convert the reference supply to it Thevenin equivalent.)

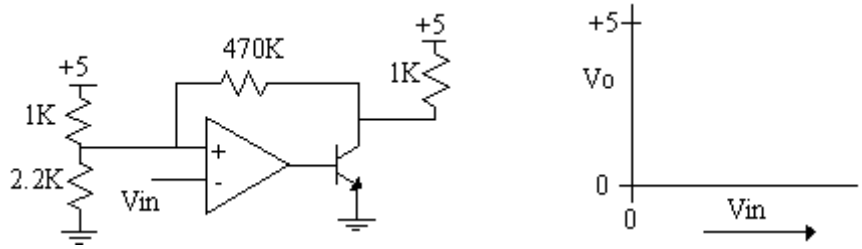


Figure 5. Voltage comparator circuit for Exercise.

The 555 Timer

The 555 timer is a functional block on a single integrated circuit chip that provides several timeing functions. The most common uses are as a square wave oscillator and as a source for a timed pulse or single-shot oscillator, astable and monostable multivibrators respectively. Timing and function are controlled by external componets and connections.

Figure 6 shows the functional block diagram of the 555's internal components. Comp A and compB are voltage comparators as discussed above. The three-5 KOhm resistor string provides two reference voltages for the voltage comparators at $2/3V_{CC}$ for comparator A and $1/3V_{CC}$ for comparator B. When the threshold input is above $1/3V_{CC}$, the flip-flop is reset, and when the trigger input is above $2/3V_{CC}$, the flip-flop is set. The ouptut comes from an interface/driver circuit which is driven by the flip-flop. The output

is capable of sourcing or sinking current. The discharge output is used to discharge external timing capacitors as we will see when we discuss specific applications.

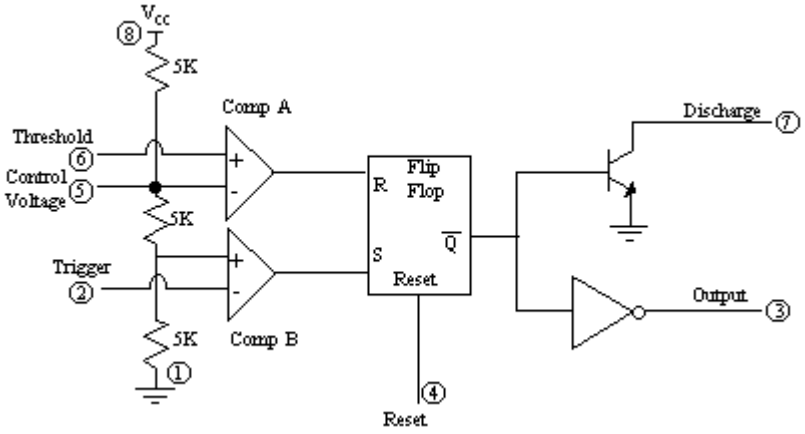


Figure 6. Functional block diagram of the 555 timer circuit

While the 555 timer is touted as being capable of sinking or sourcing 200 mA, that is with a 15 volts power supply and at that, the output voltage rises or falls substantially (typically 2.5 volts) from the rails. With a power supply of 5 volts, the consumer version of the chip can sink only 5 mA; still capable of driving TTL loads. Consult the data sheets from several manufacturers for complete specifications.

Monostable Application

In the this application, with the external connections shown in Figure 7, the timer circuit will operate as a single-shot multivibrator. Here, if the trigger gets a negative pulse, the flip-flop is set, making Q' high, truning off the discharge transistor, which then allows the capacitor to be charged up toward V_{CC}. When the capacitor voltage reaches $2/3V_{CC}$, the threshold signal causes the flip-flop to be reset, discharging the capacitor again. Typical waveforms are shown in Figure 8. It can be seen from the waveforms that the ouput remains low until a trigger signal is received. Then the output goes high while the capacitor charges and then goes back low where it remains until another trigger pulse is received. Hence, the name single-shot. Multiple triggers or continuous low voltage on the trigger input during charging have no effect, but the trigger signal must go back high again before the flip-flop can be reset by the threshold signal.

Timing is dependent on the time it takes the capacitor to charge up from a discharged state, or very near zero volts, to $2/3V_{CC}$. The charging equation is

$$v_C(t) = V_{CC}(1 - e^{-t/RC})$$

We solve this equation for $V_C = 2/3V_{CC}$.

$$0.667V_{CC} = V_{CC}(1 - e^{-t/RC})$$

and solve for t

$$1 - 0.667 = e^{-t/RC}$$

or

$$t/RC = - \ln(0.667) = 1.099$$

or the length of the pulse is

$$t = 1.099RC$$

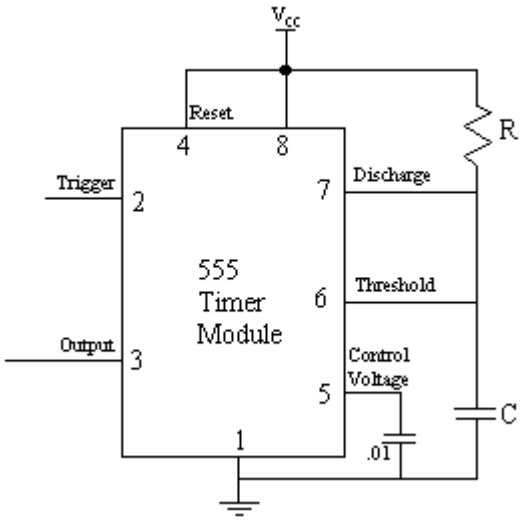


Figure 7. Connections for monostable operation.

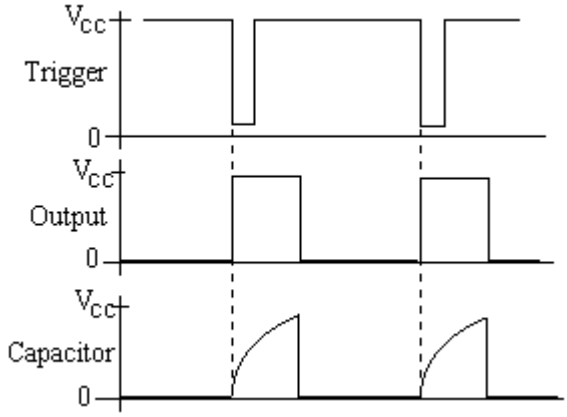


Figure 8. Wveforms for monostable operation.

The capacitor is shown as being discharged intstantaneously. However, the discharge current, and hence the rate, is limited by the β of the discharge transistor. In this mode of operation, the discharge is accomplished quickly, but it does limit how soon the cycle can be restarted. If a trigger signal is received during this discharge time, the flip-flop will be set and the capacitor recharged, this time, starting before it is completely discharged and shortening the output pulse.

Astable Application

Figure 9 shows the connections for astable or free-running operation. Both the trigger and threshold inputs are connected directly to the capacitor. There is an additional resistor, R_B, connected between the capacitor and the discharge transistor to slow the discharge. When the capacitor discharges to $1/3V_{CC}$, the trigger comparator switches and sets the flip-flop which in turn turns off the discharge transistor, allowing the capacitor to start charging up through both resistors, R_A and R_B. When the capacitor reaches $2/3V_{CC}$, the threshold input causes the flip-flop to reset which in turn turns on the discharge transistor and the capacitor discharges again. Thus, the capacitor charges and discharges back and forth between $1/3V_{CC}$ and $2/3V_{CC}$. Typical waveforms are shown in Figure 10.

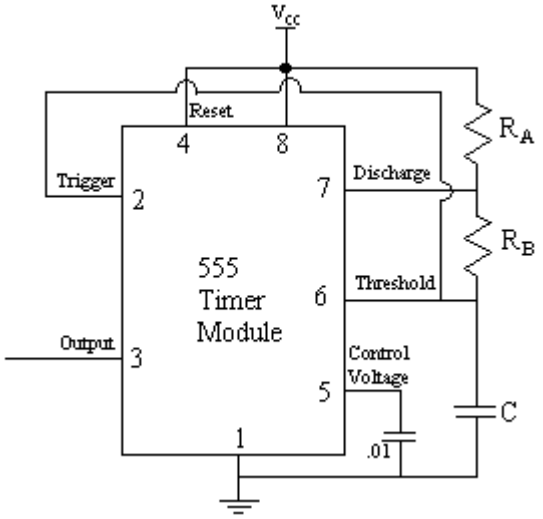


Figure 9. Astable Circuit

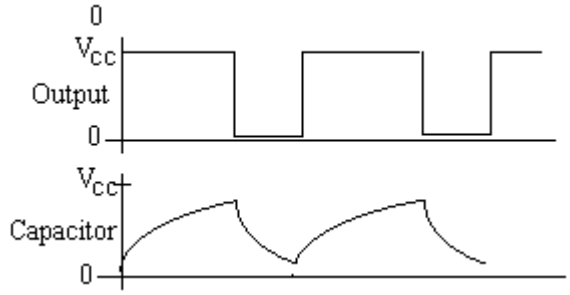


Figure 10. Astable waveforms

The discharge equation, starting at $2/3V_{CC}$ is

$$v_C(t) = \frac{2}{3}V_{CC}e^{-\frac{t}{R_BC}}$$

The capacitor discharges to $1/3V_{CC}$ at time = t_D, the time during discharge

$$\frac{1}{3}V_{CC} = \frac{2}{3}V_{CC}e^{-\frac{t_D}{R_B C}}$$

$t_D = R_B C(-\ln 0.5) = 0.693 R_B C$

During charging, the capacitor starts at $1/3V_{CC}$, and charges toward V_{CC}. The equation is

$$v_C(t) = V_{CC} - \frac{2}{3}V_{CC}e^{-\frac{t}{(R_B+R_A)C}}$$

The capacitor charges up to $2/3V_{CC}$ at time = t_C, the time during charge

$$\frac{2}{3}V_{CC} = V_{CC} - \frac{2}{3}V_{CC}e^{-\frac{t_C}{(R_B+R_A)C}}$$

$t_C = C(R_A + R_B)(-\ln 0.5) = 0.693(R_A + R_B)C$

The total period is the sum of the charge and discharge time

$t_T = 0.693(R_A + 2R_B)C$

It is important to note that the charging time will always be laronger than the discharging time with the result that the timer output will always be high longer than it will be low. From the equations, the only way to make a square wave is to make $R_A = 0$. But since R_A is connected from V_{CC} to the collector of the discharge transistor, that would result in very large currents through that transistor, which cannot be allowed.

Also note that the time is independent of the supply voltage, a very desirable result. The accuracy of the time, however, is dependent on not only the accuracy of the external resistor and capacitor values, but is dependent on the accuracy of the internal voltage divider string that set the voltage comparator reference levels for charge and discharge trigger points. This 3 resistor voltage divider string is made on the chip simultaneously so that the values tend to track. Any deviation from nominal value of one resistor tends to be seen by all resistors in the string so that the reference voltages are fairly accurate.

The control voltage input is connected directly to the reference voltage for the threshold voltage comparator. By external manipulation of this voltage, the thresholds for charge and discharge can be manipulated. The trigger point for discharge will be one-half the charge trigger point.

Does this control give the designer a means by which to make the charge and discharge times equal, producing a square wave at the timer output?

Exercises

1. You have a 0.1 μF, ±10% capacitor, all standard value 5% resistors, and a 555 timer integrated circuit. Design a monostable multivibrater circuit with a nominal pulse width of 20 ms. What are the worst case minimum and maximum pulse widths?

2. With the same components from the previous exercise, design an astable timing circuit with a nominal frequency of 1kHz.

APPENDIX A

DEFINITION OF TERMS FOR LOGIC GATE SPECIFICATIONS

V_{inLmax} The maximum input voltage that will be seen as a "LOW".

V_{inHmin} The minimum input voltage that will be seen as a "HIGH".

V_{oL} Output voltage when the output is "LOW". Spec's often give a V_{oLmax} which is the maximum value you will normally find, considering component variations.

V_{oH} Output voltage when the output is "HIGH".

Note that the output voltage specifications are given under some specific current condition, usually maximum current. The V_{oH} term is almost always associated with I_{oH}.

I_{inL} The current entering the input terminal when the input voltage is low. Normally specified at a fixed input voltage (Usually at V_{oL}). This current is usually negative; leaving the input terminal.

I_{inH} The current entering the input terminal when the input voltage is high. In the cases of DTL and TTL circuits, the input current is specified at maximum input voltage which results in maximum input circuit.

I_{oH} The current entering the output terminal when the output voltage is high. Note that this current is usually negative, the current actually leaves the output terminal. This current is usually specified when the output voltage is at its minimum, high-level value, V_{oHmin}. What this definition means is that this is the maximum current you can take out of the output and still have the output voltage stay at or above the minimum specified value.

I_{oL} The current that the gate can sink when the output is low. Since the gate usually has a transistor driver at the output, the low-level output occurs when the transistor is saturated. Thus, this current is specified at the largest current that the transistor can sink and still stay in saturation. As discussed in the book, while the transistor is in saturation, the voltage rises as the current increases. Thus, this current is usually specified at some specific V_{oLmax}, the maximum output voltage when the output is low (and sinking this large current).

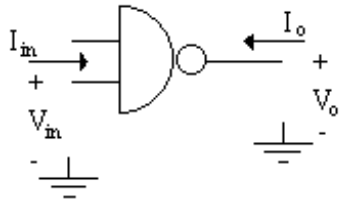


Figure A1. Terminal definitions

Current polarity is defined as positive going into a terminal. Voltages are measured from the terminal to ground. The subscript refer the terminal and its logic level. For example, I_{oL} is the output current when the output is low. and I_{inL} is the input current when the input is low.

APPENDIX B

NOISE MARGINS

NML Noise Margin Low. NML = V_{inLmax} - V_{oLmax}. This is the amount of noise voltage that can appear on the low level output signal of a gate and still have that low level signal be guaranteed to be recognized as a low at the input of a load gate.

NMH Noise Margin High. NMH = V_{oHmin} - V_{inHmin}. This is the amount of noise voltage that can appear on a high level signal at the output of a gate and still have that high level signal be guaranteed to be recognized as a high at the input of a load gate.

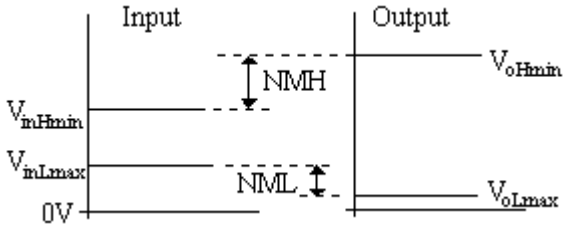


Figure B1. Definition of noise margins

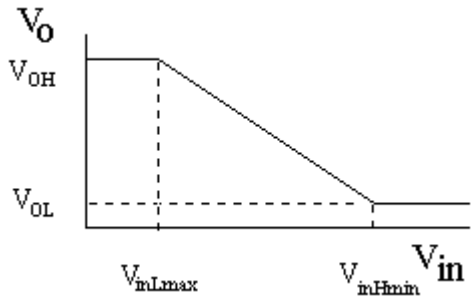


Figure B2. Typical voltage transfer characteristic for an inverting gate

APPENDIX C

PROPAGATION DELAYS

Propagation delays are measured from the 50% points of the voltage transitions bwtween the input and output waveforms. The example waveforms given below are for an inverting gate, but measuremnts are taken the same way for non-inverting gattes. The subscripts HL and LH refer to the transitions of the output waveform, low-high and high-low respectively.

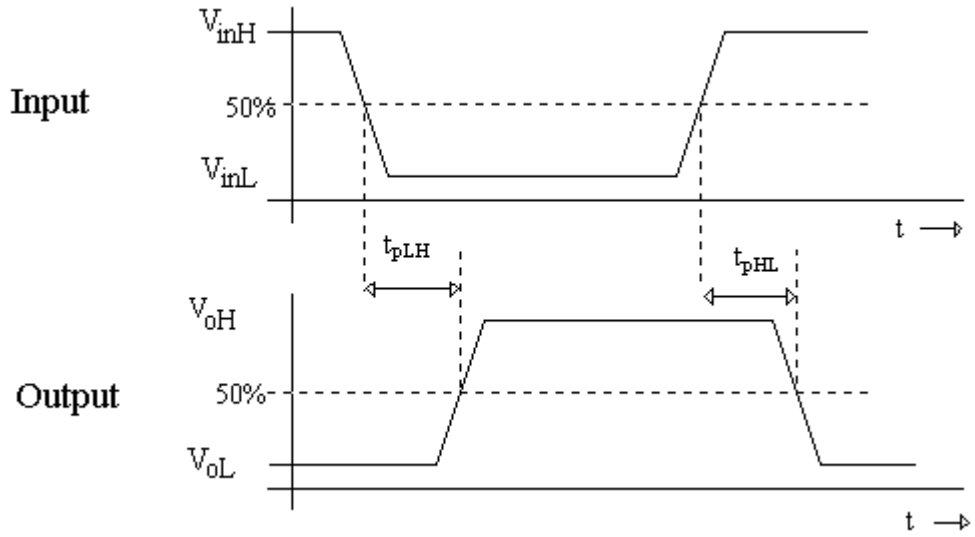


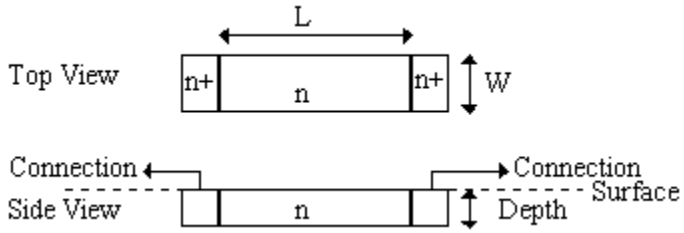
Figure C1. Propagation delay measurements

APPENDIX D

DESIGN EXERCISES

D1. Integrated Circuit Resistor

Mountaineer Electronics is developing a line of logic integrated circuits and your job is to design a 5000 Ohm resistor to be made simultaneously with the IC transistors. The resistor will be made by diffusing donors into a p-type substrate. Connections to the resistor will be made by attaching conducting material to the ends of the resistor with heavily doped ohmic contacts.



The length of the resistor is L and the cross-sectional area is WxDepth. Due to process limitations, the minimum dimension on the surface, W and L, is 0.01 mm.

Several process are available with 0.25, 0.5, 1.0, 2.0, and 4.0 x 10^{15} donors/cm^3. For each of these processes, the time required in the high temperature oven is proportional to the doping level. For example, for the lowest doping level, 0.25x10^{15}, the oven time is 30 minutes, and for 0.5x10^{15} doping level, the oven time is 60 minutes.

Diffusion depth is proportional to the time in the oven at 0.01mm/hr.

Cost is a combination of

- silicon surface area at \$2/cm^2
- oven time at \$1/hr-cm^2

Design a 5000 Ohm resistor for minimum cost.

The format required for submission of design exercises is given at the end of the listing of design exercises.

D2. LED POWER INDICATOR

McMee Toys Inc. is developing a talking teddy bear toy, the “Hippy-Dippy Teddy Bear”. The toy is to use two ME56 LEDs for the eyes that light up when the toy is active which are turned on by a simple switch and are powered from four 1.5 volt batteries, for a total of 6V. Testing on the LEDs in the lab indicate that the current must be kept above 15 mA to be bright enough. You must design a circuit that will keep the current above 15 mA but also must not exceed the ratings given in the specifications.

Design for minimum cost while guaranteeing operation within specifications over +/- 5% resistor tolerance and specified range of electrical characteristics of the LEDs.

This type of design is called worst-case design. It is intended to be used to guarantee the circuit will operate over all possible values of components. To do worst-case design, you must select the extreme possible values of all components that will present the worst cases. In this case, you must use the minimum LED forward voltage and minimum resistance values to determine maximum current and vice-versa.

The design engineer has substantial control over the quality of the finished device by how well it is designed. For example, if only typical or nominal values are used in the design calculations, then as components are selected from the bin for the manufacturing operation, many combinations of components would cause the circuit to operate outside its specifications and would likely fail prematurely if it operated at all. Rework is extremely expensive and is often much more expensive than the selling price of the product. The implications for failure of marketed devices is substantial. One dissatisfied customer has much greater effect on the profitability of the business than the purchase price of the original product. The implications go even further than company profits in that profits are necessary to keep employees working and even have effects on the economic health of the community.

D3. SOLENOID DRIVER

The Hippy-Dippy Teddy Bear from McMee Toys will wiggle its ears when the child rubs its tummy. Actually, a sensitive switch in the bear's tummy is activated with a slight pressure on its tummy. When the switch is depressed, a solenoid pulls the ears backward slightly, and when the switch is released, the solenoid is deactivated and a spring pulls the ears back to the normal position. The effect is (supposed to be anyway) that the ears wiggle when the tummy is rubbed.

Your job is to design the circuit to activate the solenoid when the switch button is depressed. Again, you must use worst-case design using the components from the list in the appendix.

D4. Logic Interface

The Hippy-Dippy Teddy Bear from McMee Toys will alternately blink its eyes (LEDs) when its left hand is squeezed. The left hand has a sensitive switch installed which is depressed when the hand is squeezed. Internally, there is a 1 Hz clock signal which runs the logic to make the appropriate logic signals for blinking the lights. Assume a 6 volt (from batteries) and a 5-volt power supply are available. The ME56 LEDs must have a minimum of 20 mA when on. Operating temperature range is 0-70 ^{0}C.

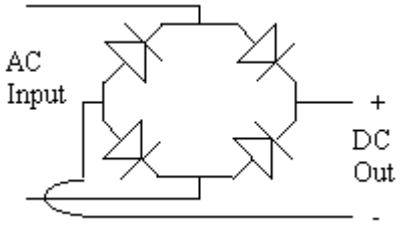
Design the system using 74LS00 and 74LS04 gates which take the 1 Hz signal and the switch input to alternately blink the LEDs. When the switch is not depressed, both LEDs

must be on continuously. The switch interface circuit must provide two logic levels in its two states. You may use either high or low logic when the switch is depressed, but you must specify which and you must specify how to wire the switch. The 74LS00 and 74LS04 ICs cost $1.25 each, including insertion and PCB space.

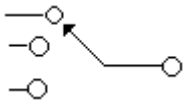
D5. RMS VOLTMETER

D'Arsonval meters move only in response to average voltage. Since the average of a sine wave is zero, these meters cannot directly measure the RMS value of a sine wave. Instead, a diode bridge is used to full-wave rectify the sine wave which will have a non-zero average value. Then resistors are used to scale the meter to provide a reading from the average of the full-wave rectified signal to the appropriate RMS value. Such a meter provides accurate scaling only for sine waves.

Fifty-Six Electronics Company is designing an old-fashioned multimeter using a 50 millivolt full-scale D'Arsonval meter with a 5000 Ohm coil. Your job is to design the meter to have three scales: 2, 20, and 200 volts RMS full-scale deflection for sine waves. You may use a diode model $I_F = 0$ for $V_F<0.60$ volts, $V_F = 0.60$ for $I_F > 0$. In addition, you may use standard 1% resistors. You have a rotary switch to select scales.



Diode Bridge Circuit


Rotary Switch

Meter Equivalent Circuit

D6. Transducer - Logic Interface

The Hippy-Dippy Teddy Bear from McMee Toys, Inc., responds to temperatures around 70 ^{0}F. When the temperature drops below 70, a voice circuit asks for a blanket. When the temperature goes back up, the voice circuit responds by saying it is too warm.

The temperature transducer provides a 1.00 volt output signal when the temperature is above 70 degrees and zero volts when below 70 degrees. The transducer Thevenin equivalent resistance is 1000 Ohms. Design an interface that will read the transducer outputs and provide appropriate logic levels to drive two 7400 gates. Note that when the transducer output is low, the gates must see a valid logic level, and when the transducer output is high, the gates must see the other logic level.

D7. MOSFET INVERTER

Mountaineer Electronics is developing an integrated circuit using NMOS logic with a 3 volt power supply. To accomodate TTL loads, the output inverter stage will use an integrated circuit resistor pull-up (to 3 volts) and an n-channel MOSFET.

Design the output inverter devices, the resistor and n-channel MOSFET to drive two 74LS00 gates. Noise margins must be at least 0.5 volts. The process for MOSFETs gives a $k' = 0.025\ mA/V^2$ and $V_T = 1.0$ Volts. You must determine the W and L for the channel of the MOSFET. The process parameters for the resistor are given in problem D1. In both devices, the minimum feature size (W and L) is 3.0 microns and can only be made in increments of 1 micron.

D8. CMOS Nand Gate

Instead of the resistor inverter in problem D7, design a 2-input Nand gate output stage to drive four 74LS00 loads with a noise margin of at least 0.4 volts. The CMOS power supply is 3 volts and the process provides $V_T = 1.0$ volt, and $k' = 0.040\ mA/V^2$. The minimum feature size is 2.0 microns with increments of 1.0 micron.

D9. DTL Nand Gate

Mountaineer Electronics is developing a DTL logic family for use in automotive applications. The automotive environment requires the system to operate in a temperature range of $-40\ ^0C$ to $+150\ ^0C$ and a power supply voltage of 9-15 volts.

Design a DTL Nand gate to operate in this environment. Because the ignition system on the automobile creates a lot of noise, the noise margins must be at least 2 volts.

DESIGN EXERCISE FORMAT

The following is the format required when submitting design exercises. Please follow it. It has been developed to minimize the amount of writing for the student and reading for the grader.

EE 56

D1, Solid State Resistor Design Date_____, Name_____

DESIGN OBJECTIVE:

Design a solid state resistor, n-type material on a p-type substrate.

Performance Requirements:

R = 5 KΩ

Design Constraints:

Doping Levels N_D = 0.25, 0.5, 1.0, 2.0, and 4.0 x 10^{16} donors/cm^3

Silicon Area Charge \$2/cm^2

Oven Time Charge \$500/hr/m^2 , $t = kN_D^2$

Diffusion Depth 0.01 mm/hr

Minimum surface dimension 0.01 mm

Design Criteria:

Minimum Cost

DESIGN COMPOSITION

Make a drawing here showing dimensions

Provide all specifications so a technician can fabricate the device

W = _____, L = _____, Doping = _____, (Oven time = _____),
(Diffusion Depth = ___)

Predicted Performance (*Exactly how does your device perform?*)

Resistance = _____ Cost = _____

PROOF OF PERFORMANCE

Resistance Calculation

Cost Calculation

Proof of peformance is not a description of how you arrived at the component values, but ***is a proof that your values will cause the device to perform as you predict****. The proof may be calculation, simulation etc.. In most cases you will have to show the models used for devices or other assumptions made.*

Copper Wire Tables

Standard Annealed Copper, American Wire Gauge

$20\ ^{0}C$

Gauge AWG	Diameter (mm)	Ohms per Kilometer	Kg per Kilometer
0000	11.68	0.1608	953.2
000	10.40	0.2028	755.8
00	9.266	0.2557	599.5
0	8.252	0.3223	475.5
1	7.348	0.4065	377.0
2	6.543	0.5128	298.9
3	5.827	0.6466	237.1
4	5.189	0.8152	188.0
5	4.620	1.028	149.0
6	4.115	1.297	118.2
7	3.665	1.634	93.80
8	3.264	2.061	74.38
9	2.906	2.600	58.95
10	2.588	3.277	46.77
11	2.30	4.14	37.1
12	2.05	5.21	29.4
13	1.83	6.56	23.4
14	1.63	8.28	18.5
15	1.45	10.4	14.7
16	1.29	13.2	11.6
17	1.15	16.6	9.24
18	1.02	21.0	7.32
19	0.912	26.4	5.81
20	0.813	33.2	4.61
21	.724	41.9	3.66
22	.643	53.2	2.88
23	.574	66.6	2.30
24	.511	84.2	1.82
25	.455	106	1.44
26	.404	135	1.14
27	.361	169	0.908
28	.320	214	.715
29	.287	266	.575
30	.254	340	.450
31	.226	430	.357
32	.203	532	.288
33	.180	675	.227
34	.160	857	.179
35	.142	1090	.141
36	.127	1360	.113
37	.114	1680	.0912
37	.102	2130	.0721
39	.089	2780	.0552
40	.079	3540	.0433

RESISTORS

5% Standard Sequence Cost = $0.25 Including Insertion and PCB space

10	16	27	43	68
11	18	30	47	75
12	20	33	51	82
13	22	36	56	91
15	24	39	62	

Power ratings: 1/8, 1/4, 1/2, 1, and 2 watts

Obtain standard values by multiplying by a power of 10 (10^{-1}, 10^{0}, 10^{1}, 10^{2}, etc.)
Values available: $1 \leq R \leq 10$ M Ohms
Temperature Coefficient: $-0.025 \leq \%/^{0}C \leq +0.075$

1% Standard Sequence Cost = $0.70 including insertion and PCB space

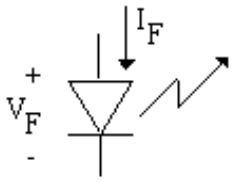
10.0	15.4	23.7	36.5	56.2	86.6
10.2	15.8	24.3	37.4	57.6	88.7
10.5	16.2	24.9	38.3	59.0	90.9
10.7	16.5	25.5	39.2	60.4	93.1
11.0	16.9	26.1	40.2	61.9	95.3
11.3	17.4	26.7	41.2	63.4	97.6
11.5	17.8	27.4	42.2	64.9	
11.8	18.2	28.0	43.2	66.5	
12.1	18.7	28.7	44.2	68.1	
12.4	19.1	29.4	45.3	69.8	
12.7	19.6	30.1	46.4	71.5	
13.0	20.0	30.9	47.5	73.2	
13.3	20.5	31.6	48.7	75.0	
13.7	21.0	32.4	49.9	76.8	
14.0	21.5	33.2	51.1	78.7	
14.3	22.1	34.0	52.3	80.6	
14.7	22.6	34.8	53.6	82.5	
15.0	23.2	35.7	54.9	84.5	

Obtain standard values by multiplying by a power of 10 (10^{-1}, 10^{0}, 10^{1}, 10^{2}, etc.)
Available Values: $10 \leq R \leq 10$ M Ohms
Temperature Coefficient: $\pm$ 100 ppm/^{0}C

ME56 LED Specifications **Mountaineer Electronics Inc.**
Visible Red LED Cost $1.75 Released 12/27/00

Maximum Ratings:

Continuous Forward Current, I_F	35 mA
Reverse Voltage, V_R	2.5 Volts
Operating Temperature Range	-40 - +100 ^{0}C



Electrical Characteristics

Characteristic	*Min*	*Typ*	*Max*	*Unit*
Reverse Leakage Current (V_R = 2.5V)		100		nA
Forward Voltage (I_F = 35 mA),V_F	0.8	1.8	2.2	V
Temperature Coefficient, Forward Voltage		-2.2		mV/^{0}K
Capacitance (f = 1 MHz)		50		pF

ME57 Silicon Diode Specifications **Mountaineer Electronics Inc.**
General Purpose Diode Cost $0.75 Released 12/27/00

Maximum Ratings:

Continuous Forward Current	100 mA
Reverse Voltage	100 Volts
Operating Temperature Range	-40 - +150 ^{0}C

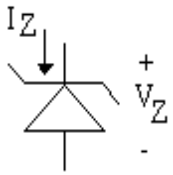
Electrical Characteristics

Characteristic	*Min*	*Typ*	*Max*	*Unit*
Reverse Leakage Current @V_R = 100V		1	100	nA
Forward Voltage (I_F = 100 mA)	0.50	0.60	0.75	V
Temperature Coefficient, forward voltage		-2.0		mV/^{0}K
Capacitance (f = 1 MHz)		25		pF

ME58 Silicon Zener Diode Specifications **Mountaineer Electronics Inc.**
Zener Diode Cost $1.20 Released 12/30/00

Maximum Ratings:

Continuous Forward Current, I_F	50	mA
Continuous Zener Current, I_Z	10	mA
Maximum Power Dissipation	25	mW
Operating Temperature Range	-40 - +150	^{0}C



Electrical Characteristics (@25 ^{0}C)

Characteristic	*Min*	*Typ*	*Max*	*Unit*
Forward Voltage (I_F = 50 mA)	0.40		1.00	V
Zener Voltage (I_Z = 10 mA), V_Z	2.05	2.20	2.35	V
Temperature Coefficient, Zener Voltage		-2.0		mV/^{0}K
Capacitance (f = 1 MHz)		50		pF

ME256 NPN BJT Specifications **Mountaineer Electronics Inc.**
NPN General Purpose Transistor Cost $1.50 Released 12/29/00

Maximum Ratings:

Collector-Emitter Voltage	40 V
Collector-Base Voltage	60 V
Emitter-Base Voltage	5.0 V
Continuous Collector Current	200 mA
Total Device Dissipation @T_A = 25 ^{0}C	625 mW
Derate Above 25 ^{0}C	5 mW/^{0}C
Operating Temperature Range (T_J)	-40 - +150 ^{0}C

Thermal Characteristics

	Max
Thermal Resistance, Junction-Case $R_{\theta JC}$	83.3 ^{0}C/W
Thermal Resistance, Junction-Ambient $R_{\theta JA}$	200 ^{0}C/W

Electrical Switching Characteristics

Off Characteristics	*Min*	*Typ*	*Max*	*Unit*
Collector-Emitter Breakdown Voltage (I_C= 1.0 mA, I_B = 0)	40			V
Collector-Base Breakdown Voltage (I_C = 10μA, I_E = 0)	60			V
Base Cutoff Current (V_{CE} = 30 V, V_{EB} = 3 V)		1	50	nA
Collector Cutoff Current (V_{CE} = 30 V, V_{EB} = 3 V)		1	50	nA
Base-Emitter Cut-In Voltage, $V_{BE\gamma}$ (V_{CE} = 30 V, I_C = 0.01 mA)	0.45	0.50		V
Saturation Characteristics				
Collector-Emitter Saturation Voltage, V_{CEsat}				
I_C = 1.0 mA, I_B = 0.10 mA			0.20	V
I_C = 50.0 mA, I_B = 5.0 mA			0.30	V
Base-Emitter Saturation Voltage, V_{BEsat}				
I_C = 1.0 mA, I_B = 0.10 mA	0.55	0.65		V
I_C = 50 mA, I_B = 5.0 mA		0.75	0.95	V

ME257 PNP BJT Specifications **Mountaineer Electronics Inc.**
PNP General Purpose Transistor Cost $1.60 Released 12/29/00

Maximum Ratings:

Collector-Emitter Voltage	-40 V
Collector-Base Voltage	-60 V
Emitter-Base Voltage	-5.0 V
Continuous Collector Current	-200 mA
Total Device Dissipation @T_A = 25 ^{0}C	625 mW
Derate Above 25 ^{0}C	5 mW/^{0}C
Operating Temperature Range (T_J)	-40 - +150 ^{0}C

Thermal Characteristics	Max
Thermal Resistance, Junction-Case $R_{\theta JC}$	83.3 ^{0}C/W
Thermal Resistance, Junction-Ambient $R_{\theta JA}$	200 ^{0}C/W

Electrical Switching Characteristics

Off Characteristics	*Min*	*Typ*	*Max*	*Unit*
Collector-Emitter Breakdown Voltage (I_C= -1.0 mA, I_B = 0)	-40			V
Collector-Base Breakdown Voltage (I_C = -10μA, I_E = 0)	-60			V
Base Cutoff Current (V_{CE} = -30 V, V_{EB} = -3 V)		-1	-50	nA
Collector Cutoff Current (V_{CE} = -30 V, V_{EB} = -3 V)		-1	-50	nA
Base-Emitter Cut-In Voltage, $V_{BE\gamma}$ (V_{CE} = -30 V, I_C = -0.01 mA)	-0.45	-0.55		V
Saturation Characteristics				
Collector-Emitter Saturation Voltage, V_{CEsat}				
I_C = -1.0 mA, I_B = -0.10 mA			-0.20	V
I_C = -50.0 mA, I_B = -5.0 mA			-0.35	V
Base-Emitter Saturation Voltage, V_{BEsat}				
I_C = -1.0 mA, I_B = -0.10 mA	-0.55	-0.65		V
I_C = -50 mA, I_B = -5.0 mA		-0.75	-0.95	V

ME181 N-Channel Enhancement Mode MOSFET

Specifications **Mountaineer Electronics Inc.**

General Purpose NMOS Transistor Cost $0.75 Released 12/29/01

Maximum Ratings:

Drain-Source Voltage	50 V
Gate-Source Voltage	+/- 20 V
Continuous Drain Current	200 mA
Total Device Dissipation @T_A = 70 ^{0}C	240 mW
Derate Above 70 ^{0}C	3 mW/^{0}C
Operating Temperature Range (T_J)	-65 - +150 ^{0}C

Thermal Characteristics Max

Thermal Resistance, Junction-Case $R_{\theta JC}$	40 ^{0}C/W
Thermal Resistance, Junction-Ambient $R_{\theta JA}$	333 ^{0}C/W

Electrical Switching Characteristics

Off Characteristics	*Min*	*Typ*	*Max*	*Unit*
Drain-Source Breakdown Voltage, (V_{GS} = 0V, I_D = 50μA)	50			V
Zero Gate Voltage Drain Current, I_{DDS} (V_{DS} = 50V, V_{GS} = 0V)			1	μA
Gate-Source Leakage Current, I_{GSS} (V_{GS} = 20 V, V_{DS} = 0 V)		1	μA	
On Characteristics				
Gate Threshold Voltage, $V_{GS(th)}$ V_{DS} = 10 V, I_D = 5 μA	1.0	1.5	2.0	V
Gate Threshold Temperature Coefficient		-4		mV/^{o}K
Drain-Source On Voltage, $V_{DS(on)}$ V_{GS} = 10 V, I_D = 50 mA	0.3	0.66	0.8	V
Gate Quiescent Voltage, $V_{GS(q)}$ V_{DS} = 28 V, I_D = 2.00 mA	3.5		5.5	V
Forward Transconductance, g_m V_{GS} = 10V, V_{DS} = 5V	5.0	7.0		mhos
Static Drain-Source On Resistance, R_{DSon} V_{GS} = 10V, I_D = 100 mA		50	80	Ω

Dynamic Characteristics

V_{DS} = 28Vdc, V_{GS} = 0V, f=1.0MHz

Input Capacitance, C_{iss}	13	pF
Output Capacitance, C_{oss}	7	pF
Reverse Transfer Capacitance, C_{rss}	0.7	pF

ME182 P-Channel Enhancement Mode MOSFET

Specifications **Mountaineer Electronics Inc.**

General Purpose PMOS Transistor Cost $0.75 Released 12/29/01

Maximum Ratings:

Drain-Source Voltage	-50 V
Gate-Source Voltage	+/- 20 V
Continuous Drain Current	-100 mA
Total Device Dissipation @T_A = 70 ^{0}C	240 mW
Derate Above 70 ^{0}C	3 mW/^{0}C
Operating Temperature Range (T_J)	-65 - +150 ^{0}C

Thermal Characteristics Max

Thermal Resistance, Junction-Case $R_{\theta JC}$	40 ^{0}C/W
Thermal Resistance, Junction-Ambient $R_{\theta JA}$	333 ^{0}C/W

Electrical Switching Characteristics

Off Characteristics	*Min*	*Typ*	*Max*	*Unit*
Drain-Source Breakdown Voltage, (V_{GS} = 0V, I_D = - 50μA)	-50			V
Zero Gate Voltage Drain Current, I_{DDS} (V_{DS} = -50V, V_{GS} = 0V)			-1	μA
Gate-Source Leakage Current, I_{GSS} (V_{GS} = -20 V, V_{DS} = 0 V)		-1	μA	
On Characteristics				
Gate Threshold Voltage, $V_{GS(th)}$ V_{DS} = -10 V, I_D = -5 μA	-1.0	-1.5	-2.0	V
Gate Threshold Temperature Coefficient		3.2		mV/^{o}K
Drain-Source On Voltage, $V_{DS(on)}$ V_{GS} = -10 V, I_D = -50 mA	-0.3	-0.66	-0.8	V
Gate Quiescent Voltage, $V_{GS(q)}$ V_{DS} = -28 V, I_D = -200.0 mA	-3.5		-5.5	V
Forward Transconductance, g_m V_{GS} = -10V, V_{DS} = -5V	5.0	7.0		mhos
Static Drain-Source On Resistance, R_{DSon} V_{GS} = -10V, I_D = -100 mA		50	80	Ω

Dynamic Characteristics

V_{DS} = -28Vdc, V_{GS} = 0V, f=1.0MHz

Input Capacitance, C_{iss}	14	pF
Output Capacitance, C_{oss}	8	pF
Reverse Transfer Capacitance, C_{rss}	0.7	pF

ME183 N-Channel Enhancement Mode MOSFET

Specifications **Mountaineer Electronics Inc.**

General Purpose NMOS Transistor Cost $1.75 Released 12/29/00

Maximum Ratings:

Drain-Source Voltage	65 V
Gate-Source Voltage	+/- 20 V
Continuous Drain Current	2 A
Total Device Dissipation @T_A = 70 ^{0}C	35 W
Derate Above 70 ^{0}C	275 mW/^{0}C
Operating Temperature Range (T_J)	-65 - +150 ^{0}C

Thermal Characteristics Max

Thermal Resistance, Junction-Case $R_{\theta JC}$	5.4 ^{0}C/W
Thermal Resistance, Junction-Ambient $R_{\theta JA}$	40 ^{0}C/W

Electrical Switching Characteristics

Off Characteristics	*Min*	*Typ*	*Max*	*Unit*
Drain-Source Breakdown Voltage, (V_{GS} = 0V, I_D = 50μA)	65			V
Zero Gate Voltage Drain Current, I_{DDS} (V_{DS} = 28V, V_{GS} = 0V)			1	mA
Gate-Source Leakage Current, I_{GSS} (V_{GS} = 20 V, V_{DS} = 0 V)			1	mA
On Characteristics				
Gate Threshold Voltage, $V_{GS(th)}$ V_{DS} = 10 V, I_D = 50 μA	2.0	3.6	4.0	V
Gate Threshold Temperature Coefficient		-4		mV/^{o}K
Drain-Source On Voltage, $V_{DS(on)}$ V_{GS} = 10 V, I_D = 0.50 A	0.3	0.66	0.8	V
Gate Quiescent Voltage, $V_{GS(q)}$ V_{DS} = 28 V, I_D = 200.0 mA	3.5		5.5	V
Forward Transconductance, g_m V_{GS} = 10V, V_{DS} = 5V	5.0	7.0		mhos
Static Drain-Source On Resistance, R_{DSon} V_{GS} = 10V, I_D = 2A		1.5	2.0	Ω

Dynamic Characteristics

V_{DS} = 28Vdc, V_{GS} = 0V, f=1.0MHz

Input Capacitance, C_{iss}	13	pF
Output Capacitance, C_{oss}	7	pF
Reverse Transfer Capacitance, C_{rss}	0.7	pF

ME184 P-Channel Enhancement Mode MOSFET

Specifications **Mountaineer Electronics Inc.**

General Purpose PMOS Transistor Cost $1.75 Released 12/29/00

Maximum Ratings:

Drain-Source Voltage	-65 V
Gate-Source Voltage	+/- 20 V
Continuous Drain Current	-2 A
Total Device Dissipation @T_A = 70 ^{0}C	40 W
Derate Above 70 ^{0}C	300 mW/^{0}C
Operating Temperature Range (T_J)	-65 - +150 ^{0}C

Thermal Characteristics Max

Thermal Resistance, Junction-Case $R_{\theta JC}$	7.4 ^{0}C/W
Thermal Resistance, Junction-Ambient $R_{\theta JA}$	45 ^{0}C/W

Electrical Switching Characteristics

Off Characteristics	*Min*	*Typ*	*Max*	*Unit*
Drain-Source Breakdown Voltage, (V_{GS} = 0V, I_D = - 50μA)	-65			V
Zero Gate Voltage Drain Current, I_{DDS} (V_{DS} = -28V, V_{GS} = 0V)			-1	mA
Gate-Source Leakage Current, I_{GSS} (V_{GS} = -20 V, V_{DS} = 0 V)		-1	mA	
On Characteristics				
Gate Threshold Voltage, $V_{GS(th)}$ V_{DS} = -10 V, I_D = -50 μA	-2.0	-3.4	-4.0	V
Gate Threshold Temperature Coefficient		3.2		mV/^{o}K
Drain-Source On Voltage, $V_{DS(on)}$ V_{GS} = -10 V, I_D = -0.50 A	-0.3	-0.66	-0.8	V
Gate Quiescent Voltage, $V_{GS(q)}$ V_{DS} = -28 V, I_D = -200.0 mA	-3.5		-5.5	V
Forward Transconductance, g_m V_{GS} = -10V, V_{DS} = -5V	5.0	7.0		mhos
Static Drain-Source On Resistance, R_{DSon} V_{GS} = -10V, I_D = -2A		1.75	2.5	Ω

Dynamic Characteristics

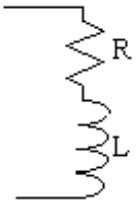
V_{DS} = -28Vdc, V_{GS} = 0V, f=1.0MHz

Input Capacitance, C_{iss}	14	pF
Output Capacitance, C_{oss}	8	pF
Reverse Transfer Capacitance, C_{rss}	0.7	pF

SOL-1 **Mountaineer Electronics**
Electro-Mechanical Solenoid Cost $2.65 Released 12/29/00

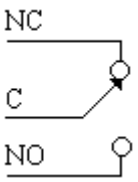
Coil Resistance	850 +/- 10%	Ohms
Inductance		
1 Inch Extension	50	mH
1/2 Inch Extension	120	mH
Closed	200	mH
Pull-in Characteristics:	Min	
1 Inch Extension		
Pull-in force @12 mA	8.0	Oz
@ 30 mA	30	Oz
1/2 Inch Extension		
Pull-in Force @ 12 mA	30	Oz
@ 30 mA	75	Oz
Holding Force@ 12 mA	120	Oz
@ 30 mA	300	Oz
Maximum Release Current@ 4 Oz	1.5 mA	
Maximum Continuous Current	20	mA
Pulse Mode Current (2 second max) (< 10% duty cycle)	40	mA

Equivalent Circuit



SW-1 **Mountaineer Electronics**
Sensitive Switch Cost $2.25 Released 12/29/00

Electrical Characteristics	
Maximum Voltage (non-inductive)	25 V
Maximum Current	6.0 mA
Mechanical Characteristics	
Button Travel	1.0 mm +/- 0.1 mm
Maximum Force to Fully Depress Button	
SPDT	2.0 Oz
DPDT	2.5 Oz



Electrical diagram for SPDT