كلية العلـــــــــــوم
قسم الفيزيـــــــــــاء

Electronics

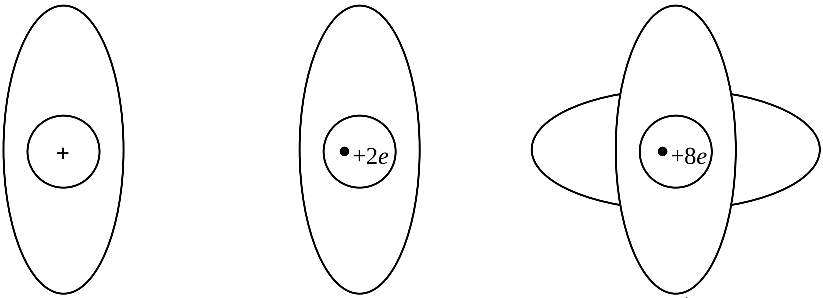
Contents

الفصل الأول
التركيب الذري والتيار الكهربي وتأثيراته

١- التركيب الذري

نعلم أن جميع العناصر أو المواد سواء كانت سائلة أو صلبة أو غازية تتكون أساساً من ذرات. وذرات العناصر تتجمع في تركيب يسمى الجزيء ويختلف تركيب الجزيء من عنصر لآخر في الشكل وعدد الذرات المكونة للجزيء الواحد. والذرة الواحدة تتكون من عدة جسيمات أولية هي:

(أ) القلب ويطلق عليها النواة: وتحتوي نواة الذرة على كل من البروتونات وهي جسيمات ذات الشحنات الموجبة والنيترونات وهي الجسيمات المتعادلة كهربياً وعددها مساوي تماماً لعدد البروتونات في كثير من الأحيان ويرتبطان معاً بقوة في النواة.



ذرة أكسجين ذرة هليوم ذرة هيدروجين

شكل (١)

(ب)يحيط بالنواة مجموعة من الشحنات السالبة تسمى الإلكترونات وهي مرتبة في عدة مدارات بيضاوية الشكل حول النواة وكل مدار له سعة معينة من

الإلكترونات. فمثلاً للمدار الأول وللثاني ٨ وللثالث ١٨ وهكذا يعتمد عدد المدارات للذرة الواحدة على عدد الإلكترونات بها وعدد الإلكترونات يساوي عدد البروتونات والشحنة السالبة للإلكترون تساوي الشحنة الموجبة للبروتون وعنى ذلك فإن الذرة كوحدة تعتبر متعادلة كهربياً.

أبسط ذرات العناصر هي ذرة الأيدروجين فهي تتألف من نواة بها بروتون واحد يتحرك حوله إليكترون واحد وذرة الهيليوم تتألف من نواة بها ٢ بروتون، ٢ نيوترون وحولها ٢ إليكترون، أما ذرة الأكسجين فإن النواة بها ٨ بروتون، ٨ نيوترون وحولها ٨ إليكترون في مدارين والشكل (١) يبين توضيحاً لهذه الذرات.

والإلكترونات سالبة التكهرب تظل دائماً في مداراتها المختلفة لا تتركها وذلك لوجود قوة ربط أو قوة جذب بينها وبين نواة الذرة الموجبة التكهرب وذلك حسب القاعدة العامة بأن الشحنات المختلفة تتجاذب والشحنات السالبة تتنافر وتبعاً لقانون كولوم بأن القوة الناشئة (سواء كانت قوة تنافر أو قوة تجاذب) بين شحنات تتناسب طردياً مع مقدار كل منها وعكسياً مع مربع المسافة بينها أي أن:

$$F = K\frac{e_1 e_2}{d^2} \qquad (1)$$

حيث أن: e_1, e_2 الشحنات (كولوم)

d المسافة بينهما (متر)

F القوة الناتجة (نيوتن)

K ثابت يتوقف على نوع الوسط الموجود به الشحنتان.

و معنى ذلك أنه كلما كانت الإلكترونات قريبة جداً من النواة كلما كان قوة الربط بينهما كبيرة بينما الإلكترونات الموجودة في المدارات البعيدة (الخارجة) تكون قوة ربطها أضعف مما يسهل عملية خروجها وتركها للمدارات الخارجية تحت أي مؤثر وانتقالها من

ذرة إلى أخرى وخاصة في العناصر أو المواد الموصلة للكهرباء وغالباً ما تكون هذه المواد مداراتها الخارجية غير مكتملة العدد من الإلكترونات مثل النحاس، الفضة، الألومنيوم وغير ذلك من العناصر الموصلة للكهرباء والإلكترونات في المدارات الخارجية الغير مكتملة تسمى إلكترونات التكافؤ أو الإليكترونات الحرة وهي أصل التيار الكهربي. بينما هناك عناصر قوة الربط بين نواة الذرة والإلكترونات في مداراتها الخارجية قوية جداً ومداراتها مكتملة العدد ولا يسهل انتقالها من ذرة إلى أخرى وتعرف هذه المواد بالمواد العازلة مثل المطاط، البلاستيك، الخشب، البوليثيلين، كلوريد البوليثيلين وهناك بعض المواد ذات الخاصية الوسط بين الموصلات و العوازل وتسمى أشباه الموصلات ويمكن أن توصل الكهرباء تحت شروط معينة مثل درجة الحرارة وإضافة شائبة للتركيب مثل الجاليوم، إنديوم، السيليكون والجرمانيوم وغير ذلك.

٢- النظرية الإلكترونية:

سبق أن عرفنا أن الإلكترونات تدور حول النواة وطاقة حركة الإلكترون في المدار تتوقف على المسافة بين المدار ومركز النواة. وسريان الكهربية بين مادتين هي نتيجة انتقال إلكترون أو أكثر من المدار الخارجي لذرة لجسم الذي يعاني زيادة في عدد الإلكترونات إلى الجسم الذي يعاني نقصاً فيها. و المادة التي تخسر إلكترون من وزنها يزداد فيها عدد البروتونات الموجبة التكهرب فتصبح ذرتها موجبة التكهرب وتسمى الذرة موجبة التكهرب (أيون موجب). والمادة التي تكتسب إلكترون يزداد فيها عدد الإلكترونات السالبة فتصبح ذرتها سالبة التكهرب وتسمى الذرة السالبة التكهرب (أيون سالب).

مما سبق نستنتج أن الشحنات الكهربية تنتقل من جسم لآخر عن طريق انتقال الإلكترونات من الجسم الذي يعاني زيادة فيها إلى الجسم الذي يعاني نقصاً فيها، وتعرف كمية الكهربية بأنها هي كمية الشحنة الكهربية التي يحملها الجسم، ومما سبق نستنتج أن

التيار الكهربي يسري من النقطة ذات الشحنة السالبة إلى النقطة ذات الشحنة الموجب.

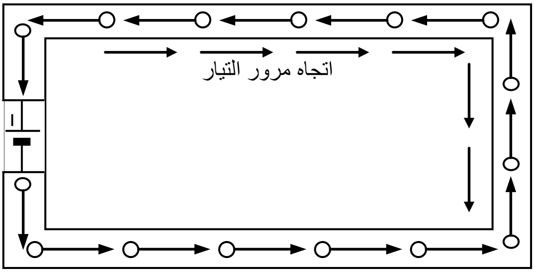
٣- التيار الكهربي:

من المعروف أن التيار الكهربي يسري من الطرف الموجب إلى الطرف السالب ولكن بدراسة النظرية الإلكترونية بأن التيار الكهربي عبارة عن مرور سيل من الإلكترونات السالبة وتمر من النقطة السالبة إلى النقطة الموجبة.

و التفسير الأخير للتيار الكهربي هو الأصح علمياً ولكن تغيير ذلك يتتبع تغيراً شاملاً في كل القوانين والقواعد المستخدمة في الهندسة الكهربية وذلك يجب أن نفرق بين:

(أ) اتجاه الإلكترونات. (ب) اتجاه التيار الكهربي.

وشكل (٢) يوضح ذلك.



شكل (٢)

٤- وحدة الشحنة (كمية الكهرباء) الكولوم:

وهي كمية الكهرباء الموجودة على الجسم أو هي الشحنة التي تتنافر مع شحنة متماثلة لها وموضوعة على بعد واحد متر بقوة قدرها واحد نيوتن.

$$Q = I\,t$$

حيث Q = الشحنة بالكولوم أو أمبير ثانية

I = التيار (أمبير)

t = الزمن (ثانية)

٥- معدل سريان التيار الكهربي:

نفرض أن هناك موصل وعدد الإلكترونات التي تمر في كل واحد متر مكعب هي N إلكترون، وأن السرعة المحورية لهذه الإلكترونات هي v بالمتر/ثانية، في زمن قدره dt يكون قد قطع مسافة قدرها

$$v.\ dt$$

وإذا كانت مساحة هذا الموصل هي A في هذا الحجم هي

$$N.v.dt.A$$

حيث

N = عدد الإلكترونات في المتر المكعب.

v = السرعة المحورية مقاسة بالمتر/ثانية.

A = مساحة مقطع الموصل مقاسة بالمتر المربع.

dt= الزمن الذي استغرقه الإلكترون مقاساً بالثواني

وإذا كانت شحنة الإلكترون هي e.

$\therefore$ الشحنة الكلية التي تمر من خلال مساحة المقطع في الزمن dt هي:

$$Q = N\,e\,A\,v\,dt \qquad (2)$$

وإذا كان التيار المار هو معدل تدفق الشحنة:

$$I = \frac{dQ}{dt} = \frac{N.A.v.dt.e}{dt}$$

$$I = N.A.v.e \qquad (3)$$

ويمكن قياس كثافة التيار J بالآتي:

$$J = \frac{I}{A} = \frac{N.A.v.e}{A} = N.v.e \ \text{Amp/m}^2$$

$$J = N.e.v \ \text{Amp/m}^2 \qquad (4)$$

مثال (١):

تيار قيمته 1.5 أمبير يمر في موصل فينتج عدد من الإلكترونات الحرة قدره 10^{30} إلكترون/متر٣ وكثافة هذا التيار تساوي 3×10^{2} أمبير/متر٢. احسب مساحة مقطع هذا الموصل والسرعة المحورية للإلكترونات إذا علمت أن شحنة الإلكترون هي 1.6×10^{-19} كولوم.

الحــل

$N = 10^{30}$ electron/m^3, $J = 300$ Amp/m^2, $I = 1.5$ Amp

$e = 1.6\times10^{-19}$C, $v = ?$ m/sec, $A = ?$ m^2

$$A = \frac{I}{J} = \frac{1.5}{300} = 0.5\times10^{-2} \ \text{m}^2$$

$$I = N.e.A.v \ \text{Amper}$$

$$1.5 = 10^{30} \times 0.5 \times 10^{-2} \times 1.6 \times 10^{-19} \times v$$

$$\therefore \ v = 1.875 \times 10^{-9} \ \text{m/sec}$$

٦- وحدة شدة التيار (الأمبير) *Amp* (*I*)

وهو التيار الذي إذا مر في سلكين متوازين ذات طول لا نهائي ومساحة مقطعهما مهملة والمساحة بينهما واحد متر في الفراغ لنتجت بينهما قوة قدرها 2×10^{-7} نيوتن/متر

من الطول أو هو عبارة عن كمية الكهرباء التي تمر في موصل كل ثانية.

وحدة قياس شدة التيار هي الأمبير ويرمز لها بالرمز Amp والوحدات العملية الأكبر هي كيلوأمبير = 1000 أمبير.

و الوحدات العملية الأصغر هي مللي أمبير = 0.001 أمبير والميكروأمبير = 10^{-6} أمبير.

٧- وحدة فرق الجهد (الفولت) ***volt (V)***

وهو فرق الجهد الذي يجب أن يوضع بين نقطتين لكي يعمل التيار شغلاً قيمته واحد إرج لتحريك شحنة قدرها واحد كولوم بين النقطتين.

$$\text{Potentialdifference}(\text{V}) = \frac{\text{Work}}{\text{Charge}} = \frac{J}{C} = \text{Volt} .$$

وبمعنى آخر فإنه لكي يظل سريان التيار الكهربي فإنه يجب أن يكون أحد طرفي التوصيل أقل جهداً من الآخر، وحدة فرق الجهد هي الفولت نسبة إلى العالم فولتا والوحدات الأكبر والأصغر من الفولت هي:

١ ك. ف= 1000 فولت   ١ ميجا. ف = 10^{6} فولت

١ مللي. ف = 10^{-3} فولت   ١ ميكرو.ف = 10^{-6} فولت

٨- وحدة المقاومة (الأوم) : ***Ohm (R)***

هي مقاومة عمود من الزئبق في درجة حرارة انصهار الجليد (صفر مئوي) ووزنه 104.452 جم وطوله 106.3 سم ومساحة مقطعه 1 مم^{٢}.

$$1\text{Ohm}(\Omega) = \frac{1\text{Volt}(\text{V})}{1\text{Amper}(\text{Amp})}$$

ويمكن إتباع نفس قواعد الأكبر والأصغر من الأوم.

٩- الدائرة الكهربية ومكوناتها:

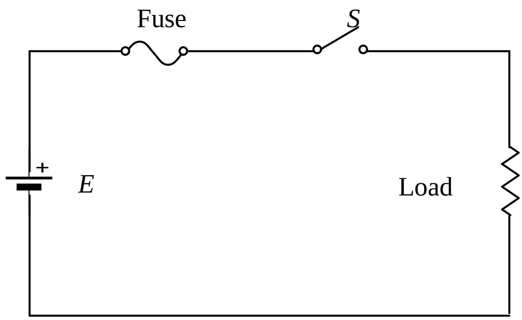
يمكن تعريف الدائرة الكهربية البسيطة بأنها مسار مغلق للتيار الكهربي، حيث يخرج التيار من النقطة ذات الجهد الأعلى ويعود إلى الدائرة من النقطة ذات الجهد الأقل، وتتكون الدائرة الكهربية في أبسط صورها من ٣ عناصر أساسية وهي:

أ- منبع كهربي.

ب- أحمال.

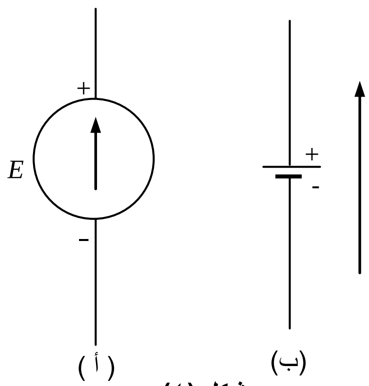
ج- أسلاك توصيل بين المنبع الكهربي والأحمال.

هذا بالإضافة لمفتاح لغلق وفتح الدائرة فلا يسري التيار إلا عند قفل المفتاح أي أنه يمر فقط في الدائرة المقفلة. أنظر شكل (٣).



شكل (٣)

كذلك قد تحتوي الدائرة على وسائل حماية (مصهرات مثلاً) وتنقسم الدائرة الكهربية عموماً إلى دائرتين، دائرة كهربية خارجية، ودائرة كهربية داخلية. والدائرة الكهربية الداخلية تتكون من المنبع الكهربي بما فيه مقاومته الداخلية أما الدائرة الكهربية الخارجية فهي تتكون من الأحمال المختلفة وأسلاك للتوصيل والأجهزة... إلخ. وسنتكلم الآن عن العناصر الأساسية والتي تكون الدائرة الكهربية.



شكل (٤)

أ- المنبع الكهربي Electric Source

إن عمل المنبع الكهربي هو تحويل أي صورة من صور الطاقة المختلفة مثل الطاقة الميكانيكية – الكيميائية – الحرارية – الضوئية عن طريق المولد الكهربي ويرمز للمولد بالرمز المبين في شكل (٤ أ)

والمولدات تعطي طاقة كهربية هائلة مع فرق جهد عالي، وتتحول الطاقة الكيميائية إلى طاقة كهربية أثناء تفريغ المركم ويرمز للبطارية بالرمز المبين بالشكل (٤ ب)، والمراكم تعطي فرق جهد صغير نسبياً، وتتحول الطاقة الحرارية إلى طاقة كهربية عن طريق الإزدواج الحراري، كما تتحول الطاقة الضوئية إلى طاقة كهربية عن طريق الخلية الكهروضوئية والتي تعطي طاقة كهربية ضعيفة مع فرق جهد صغير جداً. ويرمز للمنبع الكهربي عموماً بطرفين، الطرف السالب وهو الطرف الأقل جهداً ويميز بالعلامة (-)، الطرف الموجب وهو الطرف الأعلى جهداً ويميز بالعلامة (+)، وفي معظم الأحيان لا توضع العلامات على أطراف البطارية فالمفهوم أن الخط الطويل يمثل الطرف الموجب و الخط القصير يمثل الطرف السالب، أما المولد فيكتفى بوضع السهم أو العلامات. ويسري التيار من الطرف الموجب إلى الطرف السالب في الدائرة الخارجية، ومن الطرف السالب إلى الطرف الموجب خلال المنبع الكهربي، والقوة التي تدفع التيار الكهربي على التحرك ضد مقاومات الدائرة كلها تعرف باسم القوة الدافعة الكهربية وتكتب باختصار ق.د.ك. (e.m.f.) ويرمز لها بالرمز E وهي لا تساوي فرق الجهد بين طرفي المنبع إلا إذا كانت الدائرة مفتوحة (الدائرة المفتوحة هي الدائرة التي لا يمر بها تيار). أما فرق الجهد بين نقطتين في دائرة كهربية والذي يسبب مرور التيار من إحدى النقطتين إلى الأخرى فيسمى الجهد أو الضغط أو الفلطية، ويرمز لها بالرمز V، ووحدته الفولت أيضاً.

ب- الأحمال Loads

والأحمال متعددة بتعدد الاستخدامات، والمقصود بالحمل هو الجهاز المراد

استغلال الطاقة الكهربية من المنبع الكهربي وتحويلها إلى نوع آخر من الطاقة عن طريق هذا الجهاز، فتحويل الطاقة الكهربية إلى طاقة حرارية تتم في مصابيح الإنارة، والمدفئات الكهربية، أفران التسخين الكهربي، المصهرات، ... إلخ، وتحويل الطاقة الكهربية إلى طاقة ميكانيكية تتم عن طريق المحركات الكهربية بأنواعها المتعددة. وتحويل الطاقة الكهربية إلى طاقة كيميائية تتم عن طريق أجهزة الطلاء الكهربي ... إلخ.

ج- أسلاك التوصيل Conductors

وهي أساساً للتوصيل بين الحمل والمنبع ويجب أن تكون من الموصلات الجيدة التوصيل للكهرباء حتى لا تفقد بها جزء من الطاقة الكهربية ومعظم الموصلات المستخدمة سواء في المنازل أو المصانع أو من محطات التوليد الكهربية إلى المحطات الفرعية هي من النحاس. وتختلف أيضاً من أسلاك مفردة ومعزولة إلى جدائل النحاس للتيارات الكبيرة، ومن الأسلاك المستخدمة للأحمال البسيطة إلى الكابلات في باطن الأرض إلى الأسلاك الهوائية لنقل القدرة.

د- المفتاح Switch

للتحكم في سريان التيار الكهربي حسب الرغبة ويختلف أيضاً من المفاتيح البسيطة المستخدمة في المنازل في الإنارة إلى المفاتيح الضخمة المغمورة في الزيت للتيارات الكبيرة.

هـ- المصهر Fuse

وهو وسيلة لحماية الدائرة من القصر، لأنه في حالة القصر أو زيادة التيار زيادة كبيرة يؤدي إلى ارتفاع درجة حرارة الموصلات والأحمال مما يؤدي إلى تلفها كلياً. وأيضاً هذه الوسيلة للحماية تختلف بين المصهر البسيط والقواطع المستخدمة في المصانع

والأحمال الكبيرة.

١٠- قانون أوم *Ohm's Law*

هو النسبة بين فرق الجهد بين طرفي موصل إلى التيار المار في نفس الموصل.

$$V = I\,R \ \ (\mathrm{V})$$

$$R = \frac{V}{I} \ \ (\Omega)$$

حيث: V = فرق الجهد بالفولت. I = شدة التيار بالأمبير

R = قيمة المقاومة بالأوم.

١١- المقاومة الداخلية لمصدر التغذية:

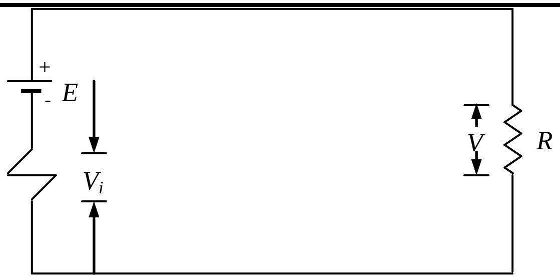
في حسابات قانون أوم يجب أن لا نتجاهل الفقد في الفولت عبر المقاومة الداخلية لمصدر الطاقة نفسه وعليه يجب حساب القوة الدافعة على المقاومة الخارجية والجزء الآخر يقع على المقاومة الداخلية للمصدر وطبقاً للشكل (٥).

$$E = V_i + V \quad \mathrm{Volt}$$

$$E = Ir_i + IR \quad \mathrm{Volt}$$

$$E = I(r_i + R) \quad \mathrm{Volt}$$

$$I = \frac{E}{r_i + R} \ \ \mathrm{Amp} \qquad (5)$$



شكل (٥)

حيث:

R = قيمة المقاومة الخارجية بالأوم. r_i = قيمة المقاومة الداخلية بالأوم.

V_i = قيمة الجهد الداخلي بالفولت. V = قيمة الجهد الخارجي بالفولت.

E = القوة الدافعة الكهربية للمصدر بالفولت.

١٢- تأثيرات التيار الكهربي:

ينتج مرور التيار الكهربي في الموصلات سواء كانت هذه الموصلات صلبة كالمعادن أو سائلة كمحاليل الأحماض والأملاح، تأثيرات مختلفة وهي:

١- تأثيرات حرارية.

٢- تأثيرات مغناطيسية.

٣- تأثيرات كيميائية.

ويجب أن نعلم أن بسبب اكتشاف هذه التأثيرات بدأ اختراع جميع الأجهزة الكهربائية.

١٣- الشغل – القدرة – الطاقة الكهربائية:

أ- قانون جول Joule's Law

ويعرف قانون جول بأنه كمية الشغل المبذول لاستمرار مرور تيار (I) ويقاس بالأمبير خلال مقاومة (R) مقاسة بالأوم في زمن قدره (t) مقاساً بالثانية، ويعبر عنه كالآتي:

$$W = I^2 R\, t \text{ Joules} \qquad (6)$$

حيث:

W = كمية الشغل المبذول مقاساً بالجول.

I = شدة التيار المار مقاساً بالأمبير.

R = قيمة المقاومة بالأوم.

t = الزمن اللازم مقاساً بالثانية.

$$I = \frac{V}{R} \text{ Amp.}$$

$$W = \frac{V}{R} IRt = V\, I\, t \text{ Joules.} \qquad (7)$$

هذا الشغل المبذول يمكن أن يفقد على هيئة طاقة حرارية إذا اعتبرنا التأثير الحراري للتيار وفي هذه الحالة يرمز لها بالرمز (H) حيث:

$$H = \frac{\text{Work}}{\text{Heat Equivallent}}$$

$$H = \frac{W}{J} = \frac{I^2 Rt}{J} \qquad (8)$$

حيث:

J = هي المكافئ الحراري مقاساً بالجول لكل سعر.

H= هي كمية الحرارة مقاساً بالسعر.

$$\therefore\ J = 4.188 \text{ Joule/Cal}$$

$$H = \frac{W}{J} = \frac{I^2 Rt \text{ Joule}}{4.188 \text{ Joule/Cal}}$$

$$H = 0.24\, I^2\, Rt = 0.24\, VIt \qquad (9)$$

ب- القدرة Watt

معدل بذل الشغل يسمى بالقدرة Power ويرمز لها بالرمز P وتقاس بوحدة الوات (Watt).

$$P = \frac{W}{t} = \frac{V\,It}{t} = VI \text{ Watt} \qquad (10)$$

والوحدات العملية المستخدمة هي الكيلووات.

K.W = 1000 Watt.

ج- الطاقة الكهربية Electric Energy

يوجد في كل منزل عداد كهربي يحسب مقدار الطاقة الكهربية المستخدمة والوحدة التجارية التي تقاس بها الطاقة الكهربية في مثل هذه الحالة هي كيلووات ساعة (K.W.h) والوات ساعة هي كمية الطاقة الكهربية المستخدمة من مرور تيار شدته واحد أمبير لمدة ساعة في سلك فرق الجهد بين طرفيه واحد فولت ومعنى ذلك أن الطاقة الكهربية تساوي الشغل الكهربي، أي أن:

$$W = V\,It = Pt \quad \text{Watt.sec} \qquad (11)$$

والوحدة العملية لقياس الطاقة هي: K.W.h.

ويمكن اعتبار العلاقات التالية:

1 W.h. = 3600 W.sec.　　1 W.sec. = 1 Joule.

1 K.W.h. $= 3600 \times 1000 = 36 \times 10^5$ Joule.

العلاقة بين الجول و السعر:

إذا كانت هناك قوة مقدارها F تحرك جسم مسافة فإن الشغل المطلوب لذلك هو:

$$W = F.d \text{ (N.m)} = \text{Joule}$$

حيث:

W = كمية الشغل بالجول. F = القوة مقاسة بالنيوتن.

D = المسافة بالمتر.

مما سبق يمكن تعريف الجول بأنه القوة التي مقدارها واحد نيوتن لكي تحرك جسم خلال مسافة قدرها واحد متر.

الحصان الميكانيكي .H.P:

1 H.P. = 75 kg.m/sec = 736 Joule/sec K.Watt = 1.36 H.P.

المكافئ الحراري

1 k.cal = 427 kg.m J= 427 kg.m/k.cal.

1 k.cal = 4188 Joule. 1 k.cal = 427×9.81 N.m.

1 cal = 4.188 Joule.

مثال (٢):

وضع في براد على هيئة سخان كهربي كمية من الماء مقدارها واحد لتر.

احسب الزمن اللازم لرفع درجة حرارة كمية الماء من 15 درجة مئوية إلى 100 درجة مئوية مع العلم بأن مقاومة السخان 125 أوم والفولت المستخدم 250 فولت، وأن الوزن المكافئ للبراد بالنسبة للماء هو 200 جرام مع فرض أن نسبة الاستفادة بالنسبة للدائرة هي 85%.

الحــل

1 Litre = 1000 gram = 1 Kg, m_w = 1000 g, t_1 = 15°C, m_k = 200 g, t_2 = 100°C, R = 125 Ω, V = 250 Volt, ξ = 85%, T = ?

$$m_t = m_w + m_k = 1000 + 200 = 1200 \text{ g}$$

$$\Delta t = 100 - 15 = 85° \text{ C}$$

$$H_a = m_t \, \Delta t = 1200 \times 85 = 102000 \text{ Cal}$$

$$I = \frac{V}{R} = \frac{250}{125} = 2 \text{ Amp}$$

$$\xi = \frac{H_a}{H_t}$$

$$\frac{85}{100} = \frac{102000}{H_t}$$

$$H_t = 120000 \text{ Cal}$$

$$H_t = 0.24 \, VIT$$

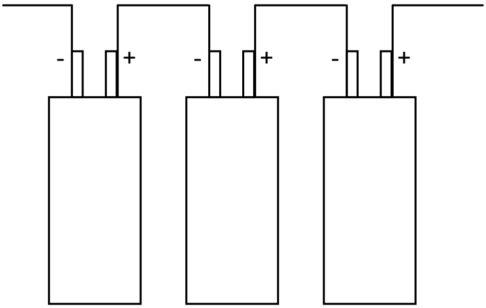
$$120000 = 0.24 \times 250 \times 2T$$

$$\therefore T = 1000 \text{ sec}$$

١٤- توصيل البطاريات:

البطاريات كمصدر طاقة سواء كانت ابتدائية أو ثانوية يمكن توصيلها على التوالي أو التوازي أو التضاعف. ويتم هذا التوصيل طبقاً للغرض المحدد للبطاريات للحصول على قسم أكبر للتيارات أو الفولت.

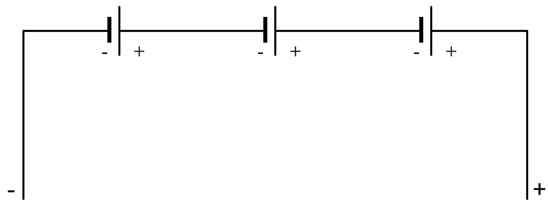
١- توصيل البطاريات على التوالي:



شكل (٦ أ)

عند توصيل مجموعة من البطاريات بحيث يوصل القطب الموجب للبطارية بالقطب السالب للبطارية الثانية والقطب الموجب بالثانية بالقطب السالب للثالثة وهكذا كما هو موضح بالشكل (٦ أ) ويبقى في النهاية القطب السالب للبطارية الأولى والقطب الموجب للبطارية الثالثة ليكونان طرفي مجموعة البطاريات لتوصيلها بالدائرة الخارجية وتكون مجموعة من القوى الدافعة الكهربية للبطاريات:

$$E_t = E_1 + E_2 + E_3 \qquad (12)$$



شكل (٦ ب)

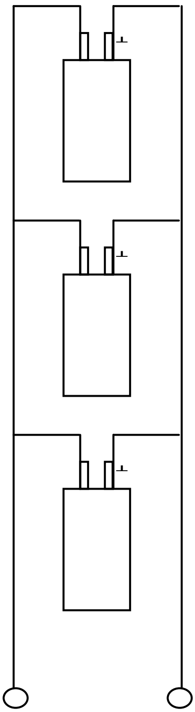
والمقاومة الداخلية الكلية:

$$R_{i\,t} = R_{i\,1} + R_{i\,2} + R_{i\,3} \qquad (13)$$

وعند توصيل البطاريات بحمل خارجي R فإن:

$$I = \frac{E_t}{R_{it} + R}$$

٢- توصيل البطاريات على التوازي:



شكل (١-٧)

يتم توصيل البطاريات على التوازي بحيث تتصل الأقطاب الموجبة بنقطة واحدة وكذلك الأقطاب السالبة كما هو موضح بشكل (٧) وتكون القوة الدافعة الكهربية الكلية:

$$E_t = E_1 = E_2 = E_3$$

وتكون المقاومة الداخلية الكلية هي:

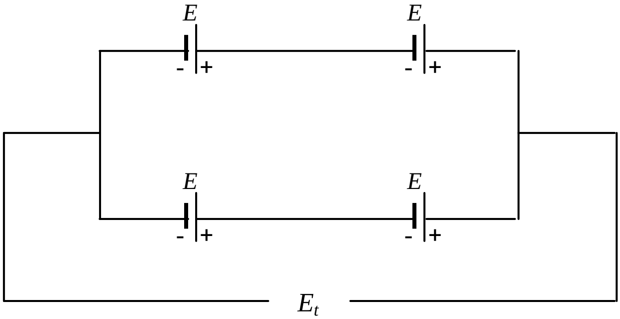
$$R_{it} = \frac{R_{i1}}{N}$$

حيث N هي عدد البطاريات في الدائرة والمقاومة الداخلية لجميع البطاريات متساوية وإذا وصلت البطاريات بحمل خارجي فإن:

$$I = \frac{E_t}{R_{it} + R} \qquad (17)$$

٣- توصيل البطاريات على التضاعف:

ترتب الأعمدة في هذه الحالة في صفوف متساوية وتوصل أعمدة كل صف على التوالي ثم توصل هذه الصفوف على التوازي كما هو موضح بالشكل (٨) والقوة الدافعة الكهربية الكلية هي:



شكل (٨)

$$E_t = E \times n \qquad (18)$$

حيث n هي عدد الوحدات الموصلة على التوالي في كل صف وتكون المقاومة الكلية للمجموعة متساوية:

$$R_{it} = \frac{R_{it} \times n}{m} \qquad (19)$$

حيث m هي عدد توصيلات التوازي والتيار المتولد عند توصيل مقاومة خارجية قدرها R هي:

$$I = \frac{E_t}{R_{it} + R} \qquad (20)$$

وتوصل البطاريات على التضاعف عندما يكون الفولت والتيار للحمل أكبر من فولت وتيار البطارية الواحدة.

مثال (٣):

وصلت أربع بطاريات على التوالي، القوة الدافعة الكهربية لكل منهما تساوي 1.2 فولت والمقاومة الداخلية 0.2 أوم. وصلت أطرافها بمقاومة 4 أوم. احسب القوة

الدافعة الكهربية الكلية وشدة التيار المار بالدائرة.

الحــل

$E_i = 1.2$ Volt, $R_i = 0.2\ \Omega$, $R = 4$

$$E_t = 1.2 \times 4 = 4.8 \text{ Volt}$$

$$R_{it} = 0.2 \times 4 = 0.8\ \Omega$$

$$I = \frac{E}{R_{it} + R} = \frac{4.8}{0.8 + 4} = 1 \text{ Amp}$$

في حالة التوازي فإن:

$$R_{it} = \frac{0.2}{4} = 0.05\,\Omega$$

$$E_t = E_1 = 1.2 \text{ Volt.}$$

$$I = \frac{E}{R_{it} + R} = \frac{1.2}{0.05 + 4} \cong 0.3 \text{ Amp.}$$

مثال (٤):

بطارية مكونة من 20 خلية موصلة على التوالي أعيد شحنها بواسطة منبع قوته 120فولت. فإذا كانت ق.د.ك. لكل عمود 2 فولت ومقاومته الداخلية 0.1 أوم، فاحسب المقاومة التي توصل على التوالي في دائرة الشحن ليكون تيار الشحن قيمته 4 أمبير.

الحــل

$E_1 = E_2 = 2$ Volt, $R_i = 0.1\ \Omega$, $V_s = 120$ V, $I_{ch} = 4$ Amp

$$I_{ch} = \frac{V_s - V_b}{R_i + R}$$

$$V_b = 20 \times 2 = 40 \text{ Volt}$$

$$R_i = 20 \times 0.1 = 2\ \Omega$$

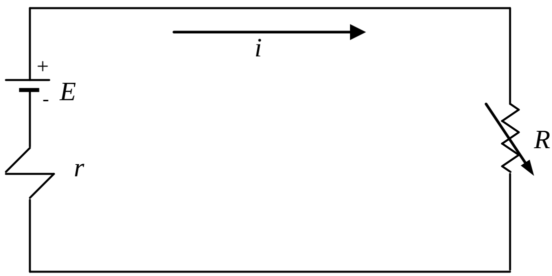
$$4 = \frac{120 - 40}{2 + R}$$

$$\therefore R = 18\ \Omega.$$

١٥- نقل أقصى قدرة من المنبع إلى الحمل:

في كثير من دوائر الاتصالات السلكية واللاسلكية والتحكم الآلي يكون من المرغوب فيه نقل أقصى قدرة ممكنة من المنبع إلى الحمل دون الأخذ في الاعتبار كفاءة النقل. اعتبرنا الدائرة شكل (٩) حيث مقاومة المنبع الداخلية r ومقاومة الحمل R وهي مقاومة متغيرة فإن التيار المار هو I.

$$I = \frac{E}{r + R} \text{ Amp}$$



شكل (٩)

القدرة التي يستهلكها الحمل هي:

$$P = I^2 R = \frac{E^2 R}{(r+R)^2}$$

ويلاحظ أن المنبع E ومقاومته الداخلية r ثابتان فإن المؤثر الوحيد في القدرة هو مقاومة الحمل المتغير R وللحصول على أقصى قدرة:

$$\frac{dP}{dR} = 0$$

$$\frac{dP}{dR} = E^2 \left[\frac{(r+R)^2 - 2R(r+R)}{(r+R)^4} \right] = 0$$

$$(r + R)^2 = 2R(r + R)$$

$$r + R = 2R$$

$$R = r$$

معنى ذلك أن شرط نقل أقصى قدرة من المنبع للحمل هو أن تكون قيمة الحمل مساوية تماماً لمقاومة المنبع الداخلية.

1- Introduction to Electronics

1–1 The Atom:

All matter is composed of atoms; all atoms consist of electrons, protons, and neutrons except normal hydrogen, which does not have a neutron. Each element in the periodic table has a unique atomic structure, and all atoms within a given element have the same number of protons. At first, the atom was thought to be a tiny indivisible sphere. Later it was shown that the atom was not a single particle but was made up of a small dense nucleus around which electrons orbit at great distances from the nucleus, similar to the way planets orbit the sun. Niels Bohr proposed that the electrons in an atom circle the nucleus in different obits, similar to the way planets orbit the sun in our solar system. The Bohr model is often referred to as the planetary model. Another view of the atom called the *quantum model* is considered a more accurate representation, but it is difficult to visualize. For most practical purposes in electronics, the Bohr model suffices and is commonly used because it is easy to visualize.

The Bohr Model:

An **atom** is the smallest particle of an element that retains the characteristics of that element. Each of the known 118 elements has atoms that are different from the atoms of all other elements. This gives each element a unique atomic structure. According to the classical Bohr model, atoms have a planetary type of structure that consists of a central nucleus surrounded by orbiting electrons, as illustrated in Figure 1–1. The **nucleus** consists of positively charged particles called **protons** and uncharged particles called **neutrons.** The basic particles of negative charge are called **electrons**. Each type of atom has a certain number of electrons and protons that distinguishes it from the atoms of all other elements. For example, the simplest atom is that of hydrogen, which has one proton and one electron, as shown in Figure 1–2(a). As another example, the helium atom, shown in Figure 1–2(b), has two protons and two neutrons in the nucleus and two electrons orbiting the nucleus.

Atomic Number

All elements are arranged in the periodic table of the elements in order according to their atomic number. The **atomic number** equals the number of protons in the nucleus, which is the same as the number of electrons in an electrically balanced (neutral) atom. For example, hydrogen has an atomic number of 1 and helium has an atomic number of 2. In their normal (or neutral) state, all atoms of a given element have the same number of electrons as protons; the positive charges cancel the negative charges, and the atom has a net charge of zero. Atomic numbers of all the elements are shown on the periodic table of the elements in Figure 1–3.

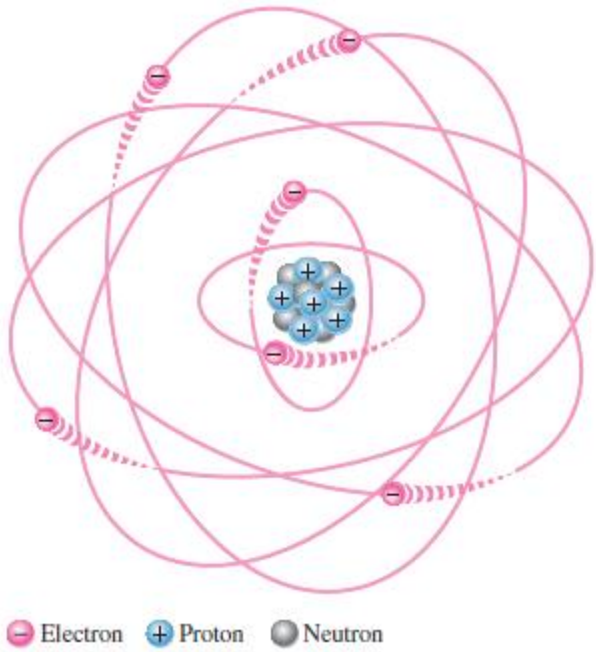


Figure (1-1): The Bohr model of an atom showing electrons in orbits around the nucleus, which consists of protons and neutrons. The "tails" on the electrons indicate motion.

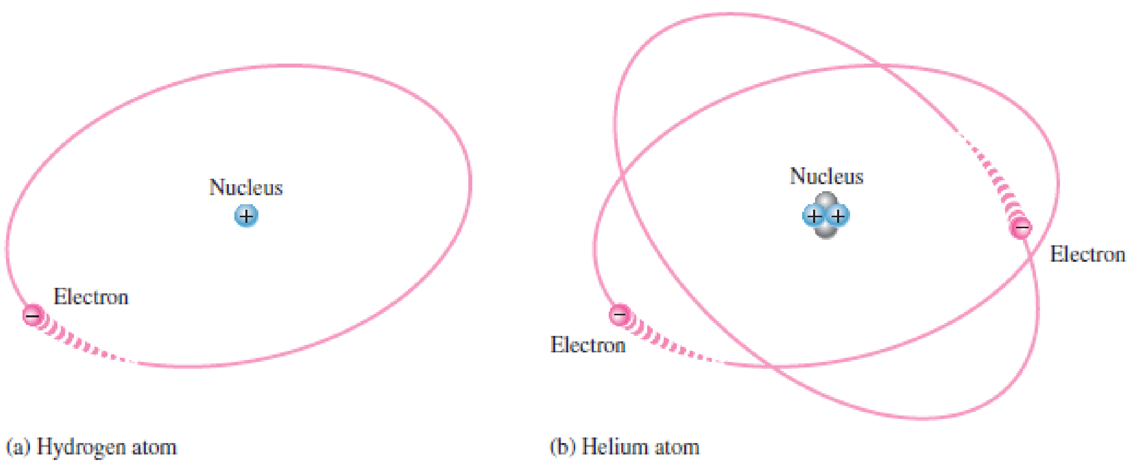


Figure (1-2): Two simple atoms, hydrogen and helium.

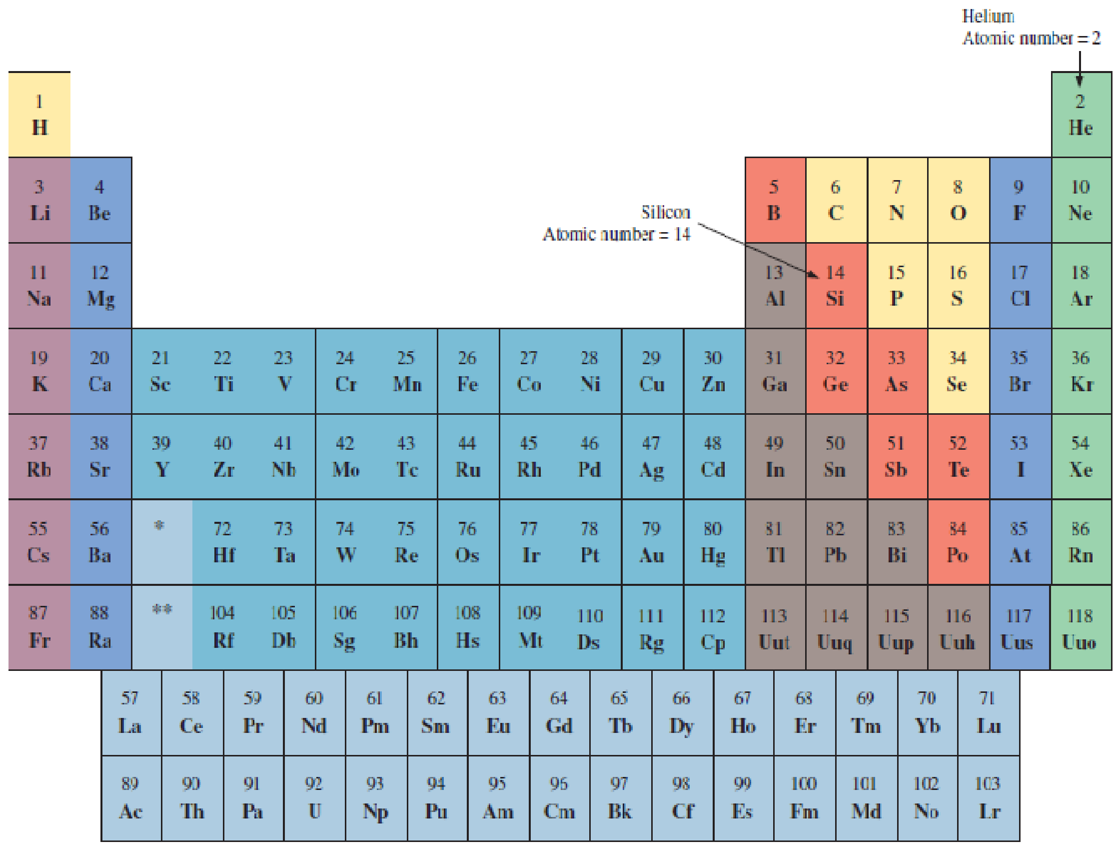


Figure (1-3): The periodic table of the elements. Some tables also show atomic mass.

Electrons and Shells

Energy Levels Electrons orbit the nucleus of an atom at certain distances from the nucleus. Electrons near the nucleus have less energy than those in more distant orbits. Only discrete (separate and distinct) values of electron energies exist within atomic structures. Therefore, electrons must orbit only at discrete distances from the nucleus. Each discrete distance (**orbit**) from the nucleus corresponds to a certain energy level. In an atom, the orbits are grouped into energy levels known as **shells**. A given atom has a fixed number of shells. Each shell has a fixed maximum number of electrons. The shells (energy levels) are designated 1, 2, 3, and so on, with 1 being closest to the nucleus. The Bohr model of the silicon atom is shown in Figure 1–4. Notice that there are 14 electrons and 14 each of protons and neutrons in the nucleus.

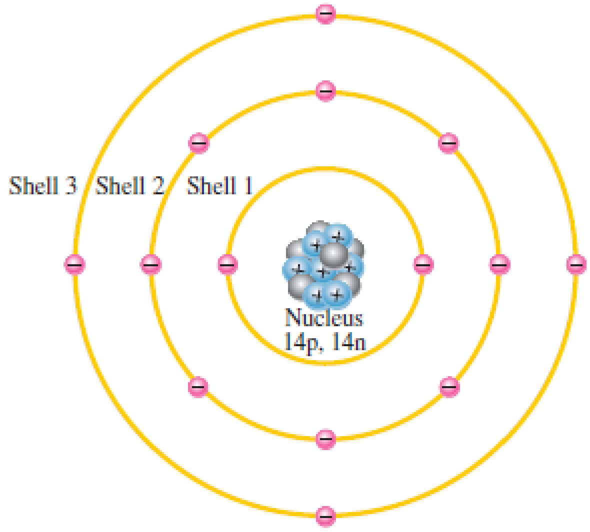


Figure (1-4): Illustration of the Bohr model of the silicon atom.

The Maximum Number of Electrons in Each Shell The maximum number of electrons (*Ne*) that can exist in each shell of an atom is a fact of nature and can be calculated by the formula,

$$\boldsymbol{Ne} = \mathbf{2}\boldsymbol{n}^2$$

where *n* is the number of the shell. The maximum number of electrons that can exist in the innermost shell (shell 1) is

$$\boldsymbol{Ne} = \mathbf{2}\boldsymbol{n}\mathbf{2 = 2(1)2 = 2}$$

The maximum number of electrons that can exist in shell 2 is

$$\boldsymbol{Ne} = \mathbf{2}\boldsymbol{n}\mathbf{2 = 2(2)2 = 2(4) = 8}$$

The maximum number of electrons that can exist in shell 3 is

$$\boldsymbol{Ne} = \mathbf{2}\boldsymbol{n}\mathbf{2 = 2(3)2 = 2(9) = 18}$$

The maximum number of electrons that can exist in shell 4 is

$$\boldsymbol{Ne} = \mathbf{2}\boldsymbol{n}\mathbf{2 = 2(4)2 = 2(16) = 32}$$

Valence Electrons

Electrons that are in orbits farther from the nucleus have higher energy and are less tightly bound to the atom than those closer to the nucleus. This is because the force of attraction between the positively charged nucleus and the negatively charged electron decreases with increasing distance from the nucleus. Electrons with the highest energy exist in the outermost shell of an atom and are relatively loosely bound to the atom. This outermost shell is known as the **valence** shell and electrons in this shell are called *valence electrons.* These valence electrons contribute to chemical reactions and bonding within the structure of a material and determine its electrical properties. When a valence electron gains sufficient energy from an external source, it can break free from its atom. This is the basis for conduction in materials.

Ionization

When an atom absorbs energy from a heat source or from light, for example, the energies of the electrons are raised. The valence electrons possess more energy and are more loosely bound to the atom than inner electrons, so they can easily jump to higher energy shells when external energy is absorbed by the atom. If a valence electron acquires a sufficient amount of energy, called *ionization energy,* it
can actually escape from the outer shell and the atom's influence. The departure of a valence electron leaves a previously neutral atom with an excess of positive charge (more protons than electrons). The process of losing a valence electron is known as **ionization**, and the resulting positively charged atom is called a *positive ion.* For example, the chemical symbol for hydrogen is H. When a neutral hydrogen atom loses its valence electron and becomes a positive ion, it is designated H^+. The escaped valence electron is called a **free electron**. The reverse process can occur in certain atoms when a free electron collides with the atom and is captured, releasing energy. The atom that has acquired the extra electron is called a *negative ion*. The ionization process is not restricted to single atoms. In many chemical reactions, a group of atoms that are bonded together can lose or acquire one or more electrons. For some nonmetallic materials such as chlorine, a free electron can be captured by the neutral atom, forming a negative ion. In the case of chlorine, the ion is more stable than the neutral atom because it has a filled outer shell. The chlorine ion is designated as Cl^-.

The Quantum Model

Although the Bohr model of an atom is widely used because of its simplicity and ease of visualization, it is not a complete model. The quantum model, a more recent model, is considered to be more accurate. The quantum model is a statistical model and very difficult to understand or visualize. Like the Bohr model, the quantum model has a nucleus of protons and neutrons surrounded by electrons. Unlike the Bohr model, the electrons in the quantum model do not exist in precise circular orbits as particles. Two important theories underlie the quantum model: the wave-particle duality and the uncertainty principle.

◆***Wave-particle duality***. Just as light can be both a wave and a particle **(photon),** electrons are thought to exhibit a dual characteristic. The velocity of an orbiting electron is considered to be its wavelength, which interferes with neighboring electron waves by amplifying or canceling each other.

◆ ***Uncertainly principle***. As you know, a wave is characterized by peaks and valleys; therefore, electrons acting as waves cannot be precisely identified in terms of their position. According to Heisenberg, it is impossible to determine simultaneously both the position and velocity of an electron with any degree of accuracy or certainty. The result of this principle produces a concept of the atom with *probability clouds,* which are mathematical descriptions of where electrons in an atom are most likely to be located. In the quantum model, each shell or energy level consists of up to four subshells called **orbitals**, which are designated *s*, *p*, *d*, and *f*. Orbital *s* can hold a maximum of two electrons, orbital *p* can hold six electrons, orbital *d* can hold ten electrons, and orbital *f* can hold fourteen electrons. Each atom can be described by an electron configuration table that shows the shells or energy levels, the orbitals, and the number of electrons in each orbital. For example, the electron configuration table for the nitrogen atom is given in Table 1–1. The first full-size number is the shell or energy level, the letter is the orbital, and the exponent is the number of electrons in the orbital.

Notation	***Explanation***
1*s*2	2 electrons in shell 1, orbital *s*
2*s*2 2*p*3	5 electrons in shell 2: 2 in orbital *s*, 3 in orbital *p*

Table 1-1: Electron configuration table for nitrogen.

Atomic orbitals do not resemble a discrete circular path for the electron as depicted in Bohr’s planetary model. In the quantum picture, each shell in the Bohr model is a three dimensional space surrounding the atom that represents the mean (average) energy of the electron cloud. The term **electron cloud** (probability cloud) is used to describe the area around an atom’s nucleus where an electron will probably be found.

EXAMPLE 1–1: Using the atomic number from the periodic table in Figure 1–3, describe a silicon (Si) atom using an electron configuration table.

Solution**:** The atomic number of silicon is 14. This means that there are 14 protons in the nucleus. Since there is always the same number of electrons as protons in a neutral atom, there are also 14 electrons. As you know, there can be up to two electrons in shell 1, eight in shell 2, and eighteen in shell 3. Therefore, in silicon there are two electrons in shell 1, eight electrons in shell 2, and four electrons in shell 3 for a total of 14 electrons. The electron configuration table for silicon is shown in Table 1–2.

Notation	***Explanation***
1*s*2	2 electrons in shell 1, orbital *s*
2*s*2 2*p*6	8 electrons in shell 2: 2 in orbital *s*, 6 in orbital *p*
3*s*2 3*p*2	4 electrons in shell 3: 2 in orbital *s*, 2 in orbital *p*

Table 1–2

Related Problem* Develop an electron configuration table for the germanium (Ge) atom in the periodic table.

In a three-dimensional representation of the quantum model of an atom, the *s*-orbitals are shaped like spheres with the nucleus in the center. For energy level 1, the sphere is "solid" but for energy levels 2 or more, each single *s*-orbital is composed of spherical surfaces that are nested shells. A *p*-orbital for shell 2 has the form of two ellipsoidal lobes with a point of tangency at the nucleus (sometimes referred to as a dumbbell shape.) The three *p*-orbitals in each energy level are oriented at right angles to each other. One is oriented on the *x*-axis, one on the *y*-axis, and one on the *z*-axis. For example, a view of the quantum model of a sodium atom (Na) that has 11 electrons is shown in Figure 1–5. The three axes are shown to give you a 3-D perspective.

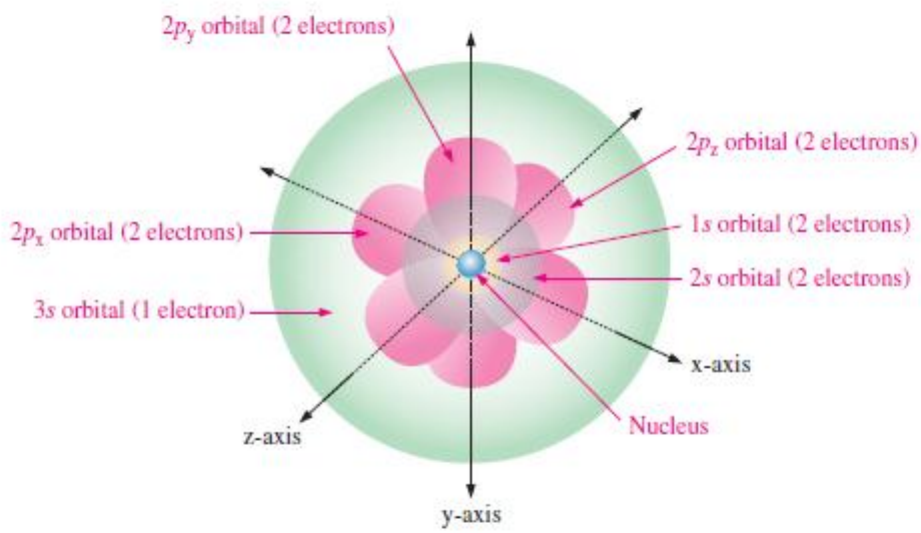


Figure (1-5): Three-dimensional quantum model of the sodium atom, showing the orbitals and number of electrons in each orbital.

SECTION 1–1 CHECKUP:

1. Describe the Bohr model of the atom.
2. Define *electron*.
3. What is the nucleus of an atom composed of? Define each component.
4. Define *atomic number*.
5. Discuss electron shells and orbits and their energy levels.
6. What is a valence electron?
7. What is a free electron?
8. Discuss the difference between positive and negative ionization.
9. Name two theories that distinguish the quantum model.

1–2 **Materials used in electronics:**

In terms of their electrical properties, materials can be classified into three groups: conductors, semiconductors, and insulators. When atoms combine to form a solid, crystalline material, they arrange themselves in a symmetrical pattern. The atoms within the crystal structure are held together by covalent bonds, which are created by the interaction of the valence electrons of the atoms. Silicon is a crystalline material.

Insulators, Conductors, and Semiconductors

All materials are made up of atoms. These atoms contribute to the electrical properties of a material, including its ability to conduct electrical current. For purposes of discussing electrical properties, an atom can be represented by the valence shell and a **core** that consists of all the inner shells and the nucleus. This concept is illustrated in Figure 1–6 for a carbon atom. Carbon is used in some types of electrical resistors. Notice that the carbon atom has four electrons in the valence shell and two electrons in the inner shell. The nucleus consists of six protons and six neutrons, so the +6 indicates the positive charge of the six protons. The core has a net charge of +4 (+6 for the nucleus and -2 for the two inner-shell electrons).

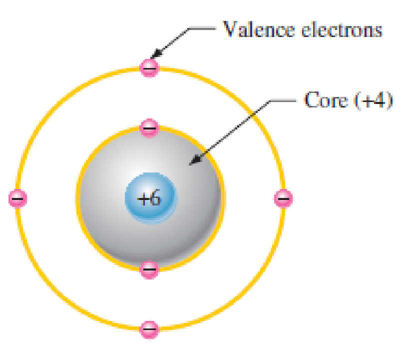


Figure (1-6): Diagram of a carbon atom.

***Insulators*:**

An **insulator** is a material that does not conduct electrical current under normal conditions. Most good insulators are compounds rather than single-element materials and have very high resistivities. Valence electrons are tightly bound to the atoms; therefore, there are very few free electrons in an insulator. Examples of insulators are rubber, plastics, glass, mica, and quartz.

***Conductors*:**

A **conductor** is a material that easily conducts electrical current. Most metals are good conductors. The best conductors are single-element materials, such as copper (Cu), silver (Ag), gold (Au), and aluminum (Al), which are characterized by atoms with only one valence electron very loosely bound to the atom. These loosely bound valence electrons become free electrons. Therefore, in a conductive material the free electrons are valence electrons.

***Semiconductors*:**

A **semiconductor** is a material that is between conductors and insulators in its ability to conduct electrical current. A semiconductor in its pure (intrinsic) state is neither a good conductor nor a good insulator. Single-element semiconductors are antimony (Sb), arsenic (As), astatine (At), boron (B), polonium (Po), tellurium (Te), silicon (Si), and germanium (Ge). Compound semiconductors such as gallium arsenide, indium phosphide, gallium nitride, silicon carbide, and silicon germanium are also commonly used. The single-element semiconductors are characterized by atoms with four valence electrons. Silicon is the most commonly used semiconductor.

Band Gap:

Recall that the valence shell of an atom represents a band of energy levels and that the valence electrons are confined to that band. When an electron acquires enough additional energy, it can leave the valence shell, become a *free electron,* and exist in what is known as the *conduction band.* The difference in energy between the valence band and the conduction band is called an *energy gap* or **band gap**. This is the amount of energy that a valence electron must have in order to jump from the valence band to the conduction band. Once in the conduction band, the electron is free to move throughout the material and is not tied to any given atom. Figure 1–7 shows energy diagrams for insulators, semiconductors, and conductors. The energy gap or band gap is the difference between two energy levels and is “not allowed” in quantum theory. It is a region in insulators and semiconductors where no electron states exist. Although an electron may not exist in this region, it can “jump” across it under certain conditions. For insulators, the gap can be crossed only when breakdown conditions occur—as when a very high voltage

is applied across the material. The band gap is illustrated in Figure 1–7(a) for insulators. In semiconductors the band gap is smaller, allowing an electron in the valence band to jump into the conduction band if it absorbs a photon. The band gap depends on the semiconductor material. This is illustrated in Figure 1–7(b). In
conductors, the conduction band and valence band overlap, so there is no gap, as shown in Figure 1–7(c). This means that electrons in the valence band move freely into the conduction band, so there are always electrons available as free electrons.

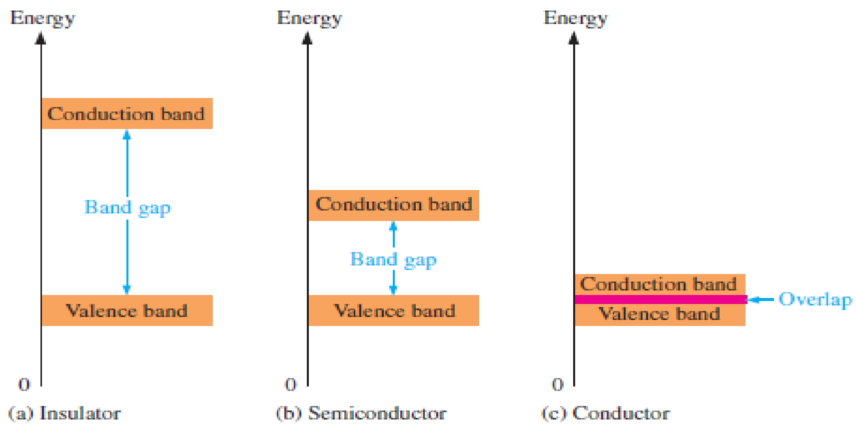


Figure (1-7): Energy diagrams for the three types of materials.

<u>Comparison of a Semiconductor Atom to a Conductor Atom:</u>

Silicon is a semiconductor and copper is a conductor. Bohr diagrams of the silicon atom and the copper atom are shown in Figure 1–8. Notice that the core of the silicon atom has a net charge of +4 (14 protons - 10 electrons) and the core of the copper atom has a net charge of +1 (29 protons - 28 electrons). The core includes everything except the valence electrons. The valence electron in the copper atom “feels” an attractive force of +1 compared to a valence electron in the silicon atom which “feels” an attractive force of +4. Therefore, there is more force trying to hold a valence electron to the atom in silicon than in copper. The copper’s valence electron is in the fourth shell, which is a greater distance from its nucleus than the silicon’s valence electron in the third shell. Recall that electrons farthest from the nucleus have the most energy. The valence electron in copper has more energy than the valence electron in silicon. This means that it is easier for valence electrons in copper to acquire enough additional energy to escape from their atoms and become free electrons than it is in silicon. In fact, large numbers of valence electrons in copper already have sufficient energy to be free electrons at normal room temperature.

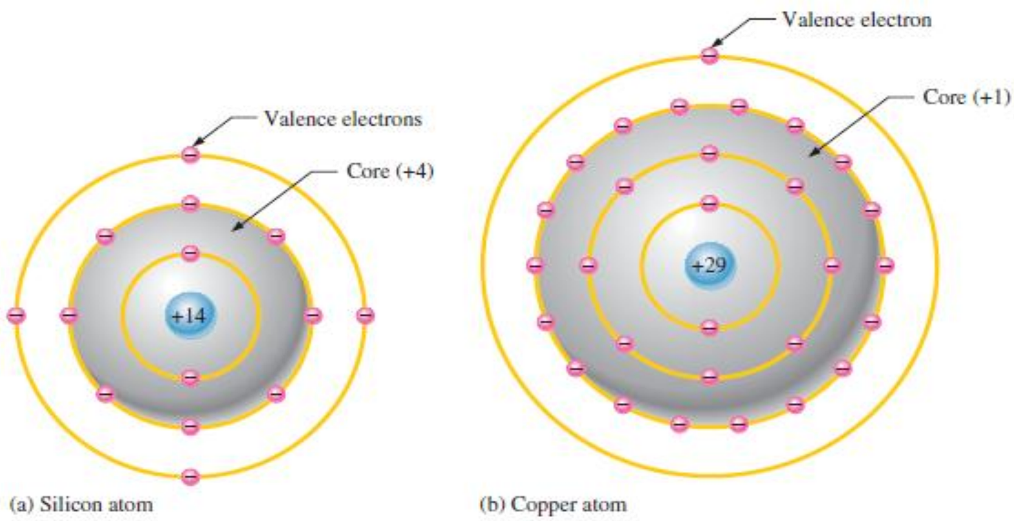


Figure (1-8): Bohr diagrams of the silicon and copper atoms.

Silicon and Germanium:

The atomic structures of silicon and germanium are compared in Figure 1–9. **Silicon** is used in diodes, transistors, integrated circuits, and other semiconductor devices. Notice that both silicon and **germanium** have the characteristic four valence electrons.

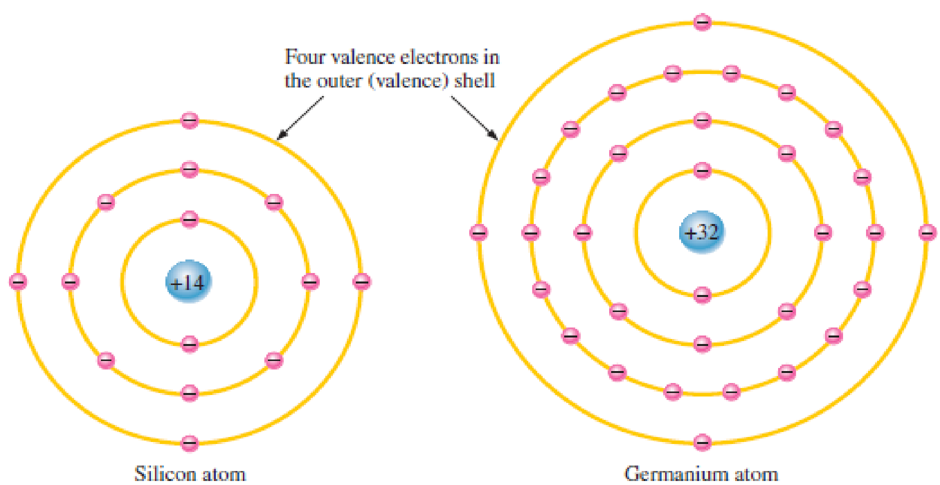
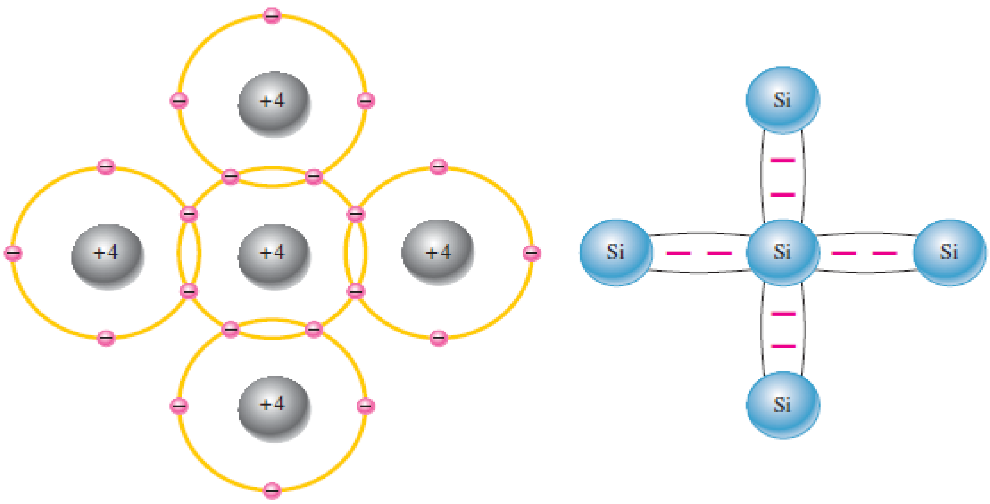


Figure (1-9): Diagrams of the silicon and germanium atoms.

The valence electrons in germanium are in the fourth shell while those in silicon are in the third shell, closer to the nucleus. This means that the germanium valence electrons are at higher energy levels than those in silicon and, therefore, require a smaller amount of additional energy to escape from the atom. This property makes germanium more unstable at high temperatures and results in excessive reverse current. This is why silicon is a more widely used semiconductive material.

***Covalent Bonds*:**

Figure 1–10 shows how each silicon atom positions itself with four adjacent silicon atoms to form a silicon **crystal**. A silicon (Si) atom with its four valence electrons shares an electron with each of its four neighbors. This effectively creates eight shared valence electrons for each atom and produces a state of chemical stability. Also, this sharing of valence electrons produces the **covalent** bonds that hold the atoms together; each valence electron is attracted equally by the two adjacent atoms which share it. Covalent bonding in an intrinsic silicon crystal is shown in Figure 1–11. An **intrinsic** crystal is one that has no impurities. Covalent bonding for germanium is similar because it also has four valence electrons.



(a) The center silicon atom shares an electron with each of the four surrounding silicon atoms, creating a covalent bond with each. The surrounding atoms are in turn bonded to other atoms, and so on.

(b) Bonding diagram. The red negative signs represent the shared valence electrons.

Figure (1-10): Illustration of covalent bonds in silicon.

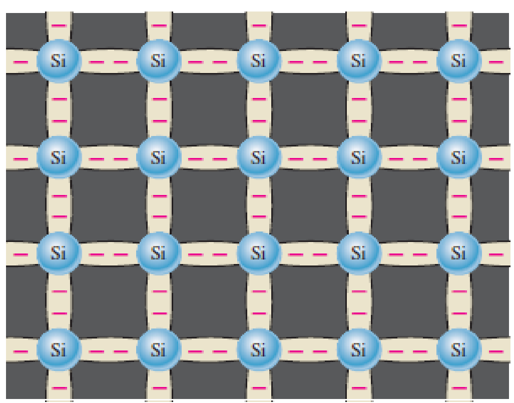


Figure (1-11): Covalent bonds in a silicon crystal.

SECTION 1–2 CHECKUP

1. What is the basic difference between conductors and insulators?
2. How do semiconductors differ from conductors and insulators?
3. How many valence electrons does a conductor such as copper have?
4. How many valence electrons does a semiconductor have?
5. Name three of the best conductive materials.
6. What is the most widely used semiconductive material?
7. Why does a semiconductor have fewer free electrons than a conductor?
8. How are covalent bonds formed?
9. What is meant by the term *intrinsic*?
10. What is a crystal?

1–3: Current in semiconductors:

The way a material conducts electrical current is important in understanding how electronic devices operate. You can't really understand the operation of a device such as a diode or transistor without knowing something about current in semiconductors.

As you have learned, the electrons of an atom can exist only within prescribed energy bands. Each shell around the nucleus corresponds to a certain energy band and is separated from adjacent shells by band gaps, in which no electrons can exist. Figure 1–12 shows the energy band diagram for an unexcited (no external energy such as heat) atom in a pure silicon crystal. This condition occurs *only* at a temperature of absolute 0 Kelvin.

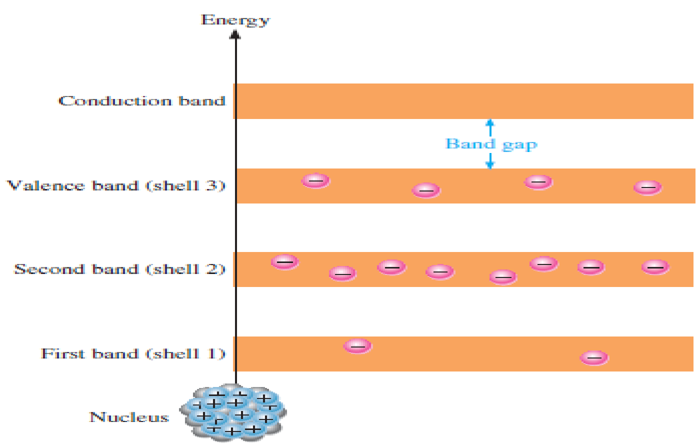


Figure (1-12): Energy band diagram for an unexcited atom in a pure (intrinsic) silicon crystal. There are no electrons in the conduction band.

Conduction Electrons and Holes:

An intrinsic (pure) silicon crystal at room temperature has sufficient heat (thermal) energy for some valence electrons to jump the gap from the valence band into the conduction

band, becoming free electrons. Free electrons are also called **conduction electrons.** This is illustrated in the energy diagram of Figure 1–13(a) and in the bonding diagram of Figure 1– 13(b). When an electron jumps to the conduction band, a vacancy is left in the valence band within the crystal. This vacancy is called a **hole**. For every electron raised to the conduction band by external energy, there is one hole left in the valence band, creating what is called an **electron-hole pair. Recombination** occurs when a conduction-band electron
loses energy and falls back into a hole in the valence band.

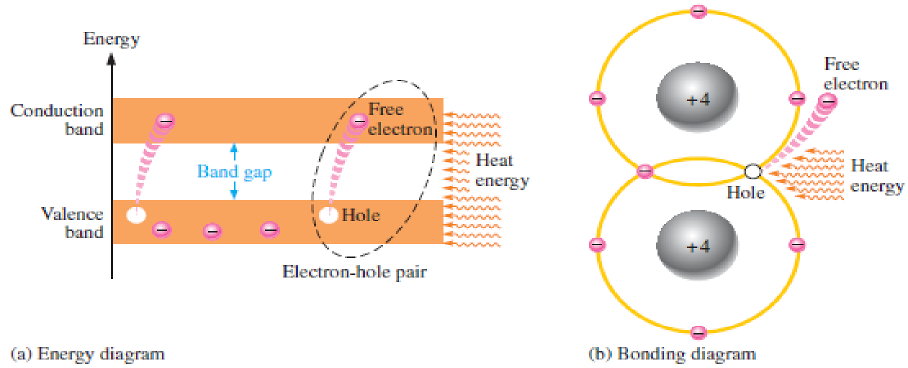


Figure (1-13): Creation of electron-hole pairs in a silicon crystal. Electrons in the conduction band are free electrons.

To summarize, a piece of intrinsic silicon at room temperature has, at any instant, a number of conduction-band (free) electrons that are unattached to any atom and are essentially drifting randomly throughout the material. There is also an equal number of holes in the valence band created when these electrons jump into the conduction band. This is illustrated in Figure 1–14.

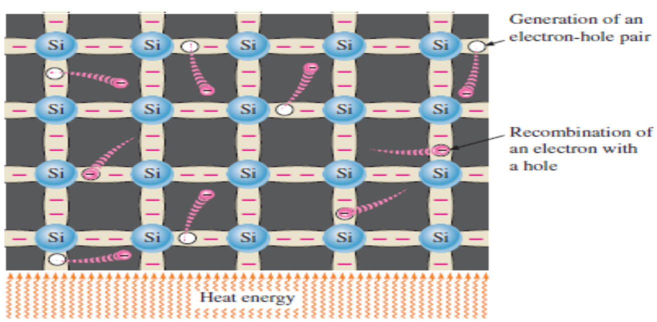


Figure (1-14): Electron-hole pairs in a silicon crystal. Free electrons are being generated continuously while some recombine with holes.

Electron and Hole Current:

When a voltage is applied across a piece of intrinsic silicon, as shown in Figure 1-15, the thermally generated free electrons in the conduction band, which are free to move randomly in the crystal structure, are now easily attracted toward the positive end. This movement of free electrons is one type of **current** in a semiconductive material and is called *electron current.*

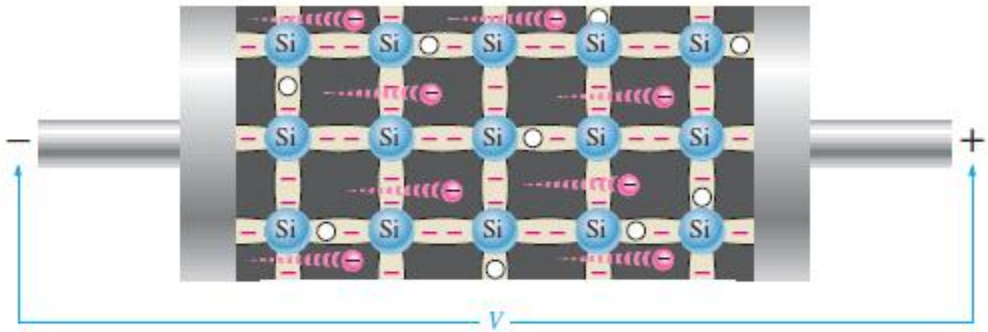


Figure (1-15): Electron current in intrinsic silicon is produced by the movement of thermally generated free electrons.

Another type of current occurs in the valence band, where the holes created by the free electrons exist. Electrons remaining in the valence band are still attached to their atoms and are not free to move randomly in the crystal structure as are the free electrons. However, a valence electron can move into a nearby hole with little change in its energy level, thus leaving another hole where it came from. Effectively the hole has moved from one place to another in the crystal structure, as illustrated in Figure 1–16. Although current in the valence band is produced by valence electrons, it is called *hole current* to distinguish it from electron current in the conduction band. As you have seen, conduction in semiconductors is considered to be either the movement of free electrons in the conduction band or the movement of holes in the valence band, which is actually the movement of valence electrons to nearby atoms, creating hole current in the opposite direction.

It is interesting to contrast the two types of charge movement in a semiconductor with the charge movement in a metallic conductor, such as copper. Copper atoms form a different type of crystal in which the atoms are not covalently bonded to each other but consist of a “sea” of positive ion cores, which are atoms stripped of their valence electrons. The valence electrons are attracted to the positive ions, keeping the positive ions together and forming the metallic bond. The valence electrons do not belong to a given atom, but to the crystal as a whole. Since the valence electrons in copper are free to move, the application of a voltage results in current. There is only one type of current—the movement of free electrons—because there are no “holes” in the metallic crystal structure.

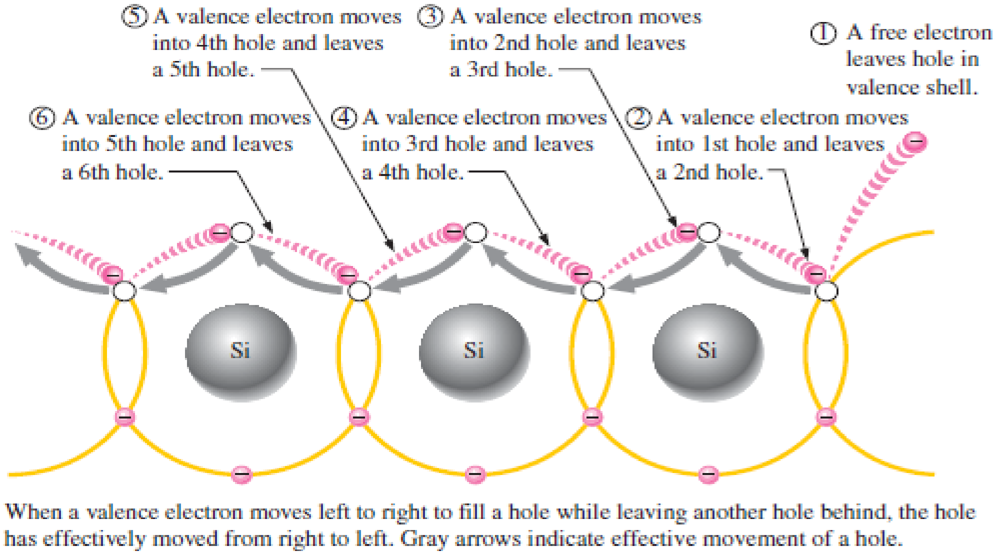


Figure (1-16): Hole current in intrinsic silicon.

SECTION 1–3 CHECKUP

1. Are free electrons in the valence band or in the conduction band?
2. Which electrons are responsible for electron current in silicon?
3. What is a hole?
4. At what energy level does hole current occur?

1–4 ***N*-type and P-type semiconductors:**

Semiconductive materials do not conduct current well and are of limited value in their intrinsic state. This is because of the limited number of free electrons in the conduction band and holes in the valence band. Intrinsic silicon (or germanium) must be modified by increasing the number of free electrons or holes to increase its conductivity and make it useful in electronic devices. This is done by adding impurities to the intrinsic material. Two types of extrinsic (impure) semiconductive materials, *n*-type and *p*-type, are the key building blocks for most types of electronic devices.

Since semiconductors are generally poor conductors, their conductivity can be drastically increased by the controlled addition of impurities to the intrinsic (pure) semiconductive material. This process, called **doping**, increases the number of current carriers (electrons or holes). The two categories of impurities are *n*-type and *p*-type.

***N*-Type Semiconductor**

To increase the number of conduction-band electrons in intrinsic silicon, **pentavalent** impurity atoms are added. These are atoms with five valence electrons such as arsenic (As), phosphorus (P), bismuth (Bi), and antimony (Sb).

As illustrated in Figure 1–17, each pentavalent atom (antimony, in this case) forms covalent bonds with four adjacent silicon atoms. Four of the antimony atom's valence electrons are used to form the covalent bonds with silicon atoms, leaving one extra electron. This extra electron becomes a conduction electron because it is not involved in bonding. Because the pentavalent atom gives up an electron, it is often called a *donor atom.* The number of conduction electrons can be carefully controlled by the number of impurity atoms added to the silicon. A conduction electron created by this doping process does not leave a hole in the valence band because it is in excess of the number required to fill the valence band.

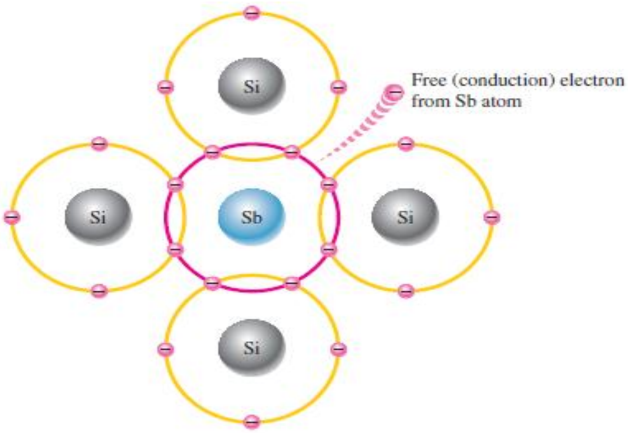


Figure (1-17): Pentavalent impurity atom in a silicon crystal structure. An antimony (Sb) impurity atom is shown in the center. The extra electron from the Sb atom becomes a free electron.

Majority and Minority Carriers Since most of the current carriers are electrons, silicon (or germanium) doped with pentavalent atoms is an *n*-type semiconductor (the *n* stands for the negative charge on an electron). The electrons are called the **majority carriers** in *n*-type material. Although the majority of current carriers in *n*-type material are electrons, there are also a few holes that are created when electron-hole pairs are thermally generated. These holes are *not* produced by the addition of the pentavalent impurity atoms. Holes in an *n*-type material are called **minority carriers.**

***P*-Type Semiconductor**

To increase the number of holes in intrinsic silicon, **trivalent** impurity atoms are added. These are atoms with three valence electrons such as boron (B), indium (In), and gallium (Ga). As illustrated in Figure 1–18, each trivalent atom (boron, in this case) forms covalent bonds with four adjacent silicon atoms. All three of the boron atom's valence electrons are used in the covalent bonds; and, since four electrons are required, a hole results when each trivalent atom is added. Because the trivalent atom can take an electron, it is often referred to as an *acceptor atom.* The number of holes can be carefully

controlled by the number of trivalent impurity atoms added to the silicon. A hole created by this doping process is *not*
accompanied by a conduction (free) electron.

Majority and Minority Carriers Since most of the current carriers are holes, silicon (or germanium) doped with trivalent atoms is called a *p*-type semiconductor. The holes are the majority carriers in *p*-type material. Although the majority of current carriers in *p*-type material are holes, there are also a few conduction-band electrons that are created when electron-hole pairs are thermally generated. These conduction-band electrons are *not* produced by the addition of the trivalent impurity atoms. Conduction-band electrons in *p*-type material are the minority carriers.

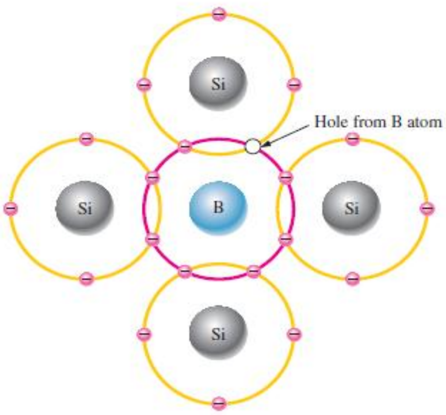


Figure (1-18): Trivalent impurity atom in a silicon crystal structure. A boron (B) impurity atom is shown in the center.

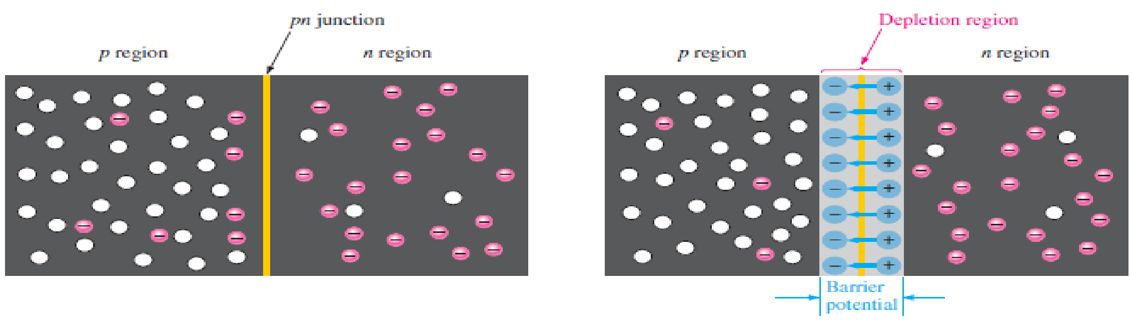
SECTION 1–4 CHECKUP

1. Define *doping*.
2. What is the difference between a pentavalent atom and a trivalent atom?
3. What are other names for the pentavalent and trivalent atoms?
4. How is an *n*-type semiconductor formed?
5. How is a *p*-type semiconductor formed?
6. What is the majority carrier in an *n*-type semiconductor?
7. What is the majority carrier in a *p*-type semiconductor?
8. By what process are the majority carriers produced?
9. By what process are the minority carriers produced?
10. What is the difference between intrinsic and extrinsic semiconductors?

1–5 **The PN junction**

When you take a block of silicon and dope part of it with a trivalent impurity and the other part with a pentavalent impurity, a boundary called the *pn* junction is formed between the resulting *p*-type and *n*-type portions. The *pn* junction is the basis for diodes, certain transistors, solar cells, and other devices, as you will learn later.

A *p*-type material consists of silicon atoms and trivalent impurity atoms such as boron. The boron atom adds a hole when it bonds with the silicon atoms. However, since the number of protons and the number of electrons are equal throughout the material, there is no net charge in the material and so it is neutral. An *n*-type silicon material consists of silicon atoms and pentavalent impurity atoms such as antimony. As you have seen, an impurity atom releases an electron when it bonds with four silicon atoms. Since there is still an equal number of protons and electrons (including the free electrons) throughout the material, there is no net charge in the material and so it is neutral. If a piece of intrinsic silicon is doped so that part is *n*-type and the other part is *p*-type, a **pn junction** forms at the boundary between the two regions and a diode is created, as indicated in Figure 1–19(a). The *p* region has many holes (majority carriers) from the impurity atoms and only a few thermally generated free electrons (minority carriers). The *n* region has many free electrons (majority carriers) from the impurity atoms and only a few thermally generated holes (minority carriers).



(a) The basic silicon structure at the instant of junction formation showing only the majority and minority carriers. Free electrons in the *n* region near the *pn* junction begin to diffuse across the junction and fall into holes near the junction in the *p* region.

(b) For every electron that diffuses across the junction and combines with a hole, a positive charge is left in the *n* region and a negative charge is created in the *p* region, forming a barrier potential. This action continues until the voltage of the barrier repels further diffusion. The blue arrows between the positive and negative charges in the depletion region represent the electric field.

Figure (1-19): Formation of the depletion region. The width of the depletion region is exaggerated for illustration purposes.

Formation of the Depletion Region

The free electrons in the *n* region are randomly drifting in all directions. At the instant of the *pn* junction formation, the free electrons near the junction in the *n* region begin to diffuse across the junction into the *p* region where they combine with holes near the junction, as shown in Figure 1–19(b). Before the *pn* junction is formed, recall that there are as many electrons as protons in the *n*-type material, making the material neutral in terms of net charge. The same is true for the *p*-type material. When the *pn* junction is formed, the *n* region loses free electrons as they diffuse across the junction. This creates a layer of positive charges (pentavalent ions) near the junction. As the electrons move across the junction, the *p* region loses holes as the electrons and holes combine. This creates a layer of negative charges (trivalent ions) near the junction. These two layers of positive and negative charges form the **depletion region,** as shown in Figure 1–19(b). The term *depletion* refers to the fact that the region near the *pn* junction is depleted of charge carriers (electrons and holes) due to diffusion across the junction. Keep in mind that the depletion region is formed very quickly and is very thin compared to the *n* region and *p* region. After the initial surge of free electrons across the *pn* junction, the depletion region has expanded to a point where equilibrium is established and there is no further diffusion of electrons across the junction. This occurs as follows. As electrons continue to diffuse across the junction, more and more positive and negative charges are created near the junction as the depletion region is formed. A point is reached where the total negative charge in the depletion region repels any further diffusion of electrons (negatively charged particles) into the *p* region (like charges repel) and the diffusion stops. In other words, the depletion region acts as a barrier to the further movement of electrons across the junction.

Barrier Potential:

Any time there is a positive charge and a negative charge near each other, there is a force acting on the charges as described by Coulomb's law. In the depletion region there are many positive charges and many negative charges on opposite sides of the *pn* junction. The forces between the opposite charges form an *electric field,* as illustrated in Figure 1–19(b) by the blue arrows between the positive charges and the negative charges. This electric field is a barrier to the free electrons in the *n* region, and energy must be expended to move an electron through the electric field. That is, external energy must be applied to get the electrons to move across the barrier of the electric field in the depletion region. The potential difference of the electric field across the depletion region is the amount of voltage required to move electrons through the electric field. This potential difference is called the **barrier potential** and is expressed in volts. Stated another way, a certain amount of voltage equal to the barrier potential and with the proper polarity must be applied across a *pn* junction before electrons will begin to flow across the junction. You will learn more about this when we discuss *biasing* in Chapter 2. The barrier potential of a *pn* junction depends on several factors, including the type of semiconductive material, the amount of doping, and the temperature. The typical barrier potential is approximately 0.7 V for silicon and 0.3 V for germanium at Because germanium devices are not widely used, silicon will be used throughout the rest of the book.

<u>Energy Diagrams of the *PN* Junction and Depletion Region</u>

The valence and conduction bands in an *n*-type material are at slightly lower energy levels than the valence and conduction bands in a *p*-type material. Recall that *p*-type material has trivalent impurities and *n*-type material has pentavalent impurities. The trivalent impurities exert lower forces on the outer-shell electrons than the pentavalent impurities. The lower forces in *p*-type materials mean that the electron orbits are slightly larger and hence have greater energy than the electron orbits in the *n*-type materials. An energy diagram for a *pn* junction at the instant of formation is shown in Figure 1–20(a). As you can see, the valence and conduction bands in the *n* region are at lower energy levels than those in the *p* region, but there is a significant amount of overlapping. The free electrons in the *n* region that occupy the upper part of the conduction band in terms of their energy can easily diffuse across the junction (they do not have to gain additional energy) and temporarily become free electrons in the lower part of the *p*-region conduction band. After crossing the junction, the electrons quickly lose energy and fall into the holes in the *p*-region valence band as indicated in Figure 1-20(a). As the diffusion continues, the depletion region begins to form and the energy level of the *n*-region conduction band decreases.

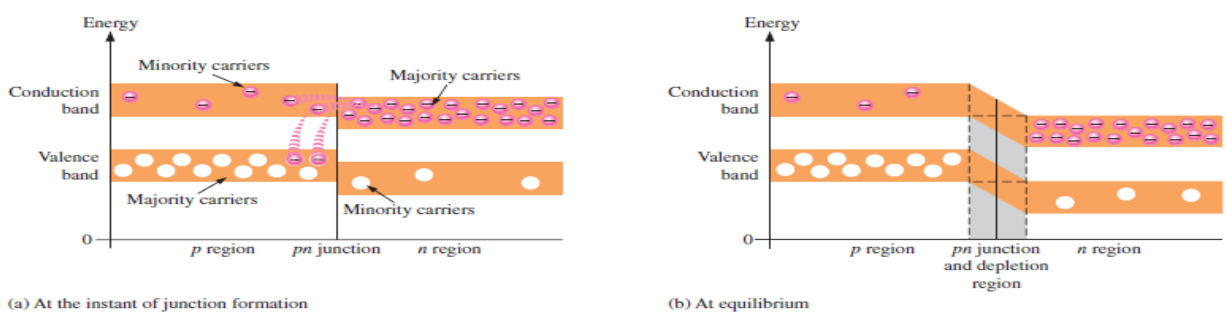


Figure (1-20): Energy diagrams illustrating the formation of the *pn* junction and depletion region.

The decrease in the energy level of the conduction band in the *n* region is due to the loss of the higher-energy electrons that have diffused across the junction to the *p* region. Soon, there are no electrons left in the *n*-region conduction band with enough energy to get across the junction to the *p*-region conduction band, as indicated by the alignment of the top of the *n*-region conduction band and the bottom of the *p*-region conduction band in Figure 1–20(b). At this point, the junction is at equilibrium; and the depletion region is complete because diffusion has ceased. There is an energy gradiant across the depletion region which acts as an "energy hill" that an *n*-region electron must climb to get to the *p* region. Notice that as the energy level of the *n*-region conduction band has shifted downward, the energy level of the valence band has also shifted downward. It still takes the same amount of energy for a valence electron to become a free electron. In other words, the energy gap between the valence band and the conduction band remains the same.

SECTION 1–5 CHECKUP

1. What is a *pn* junction?
2. Explain diffusion.
3. Describe the depletion region.
4. Explain what the barrier potential is and how it is created.
5. What is the typical value of the barrier potential for a silicon diode?
6. What is the typical value of the barrier potential for a germanium diode?

Problems

Section 1–2 Materials Used in Electronics

3. For each of the energy diagrams in Figure 1–21, determine the class of material based on relative comparisons.

4. A certain atom has four valence electrons. What type of atom is it?

5. In a silicon crystal, how many covalent bonds does a single atom form?

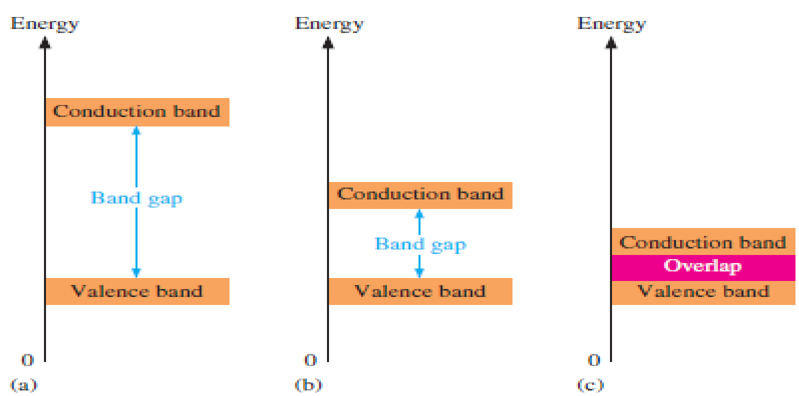


Figure (1-21)

Section 1–3 Current in Semiconductors

6. What happens when heat is added to silicon?

7. Name the two energy bands at which current is produced in silicon.

Section 1–4 *N*-Type and *P*-Type Semiconductors

8. Describe the process of doping and explain how it alters the atomic structure of silicon.

9. What is antimony? What is boron?

Section 1–5 The *PN* Junction

10. How is the electric field across the *pn* junction created?

11. Because of its barrier potential, can a diode be used as a voltage source? Explain.

2- Diodes and Applications

2–1 Diode operation

Similar to the solar cell in Chapter 1, a diode is a two-terminal semiconductor device formed by two doped regions of silicon separated by a *pn* junction. In this chapter, the most common category of diode, known as the general-purpose diode, is covered. Other names, such as rectifier diode or signal diode, depend on the particular type of application for which the diode was designed. You will learn how to use a voltage to cause the diode to conduct current in one direction and block it in the other direction. This process is called *biasing*.

The Diode

As mentioned, a **diode** is made from a small piece of semiconductor material, usually silicon, in which half is doped as a *p* region and half is doped as an *n* region with a *pn* junction and depletion region in between. The *p* region is called the **anode** and is connected to a conductive terminal. The *n* region is called the **cathode** and is connected to a second conductive terminal. The basic diode structure and schematic symbol are shown in Figure 2–1.

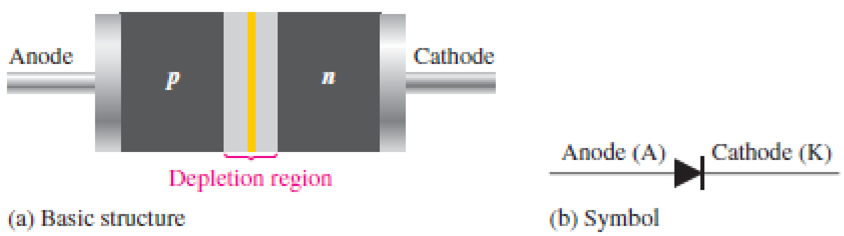


Figure 2–1: The diode.

Typical Diode Packages Several common physical configurations of through-hole mounted diodes are illustrated in Figure 2–2(a). The anode (A) and cathode (K) are indicated on a diode in several ways, depending on the type of package. The cathode is usually marked by a band, a tab, or some other feature. On those packages where one lead is connected to the case, the case is the cathode.

Surface-Mount Diode Packages Figure 2–2(b) shows typical diode packages for surface mounting on a printed circuit board. The SOD and SOT packages have gull-wing shaped leads. The SMA package has L-shaped leads that bend under the package. The SOD and SMA types have a band on one end to indicate the cathode. The SOT type is a three-terminal package in which there are either one or two diodes. In a single-diode SOT package, pin 1 is usually the anode and pin 3 is the cathode. In a dual-diode SOT package, pin 3 is the common terminal and can be either the anode or the cathode. Always check the datasheet for the particular diode to verify the pin configurations.

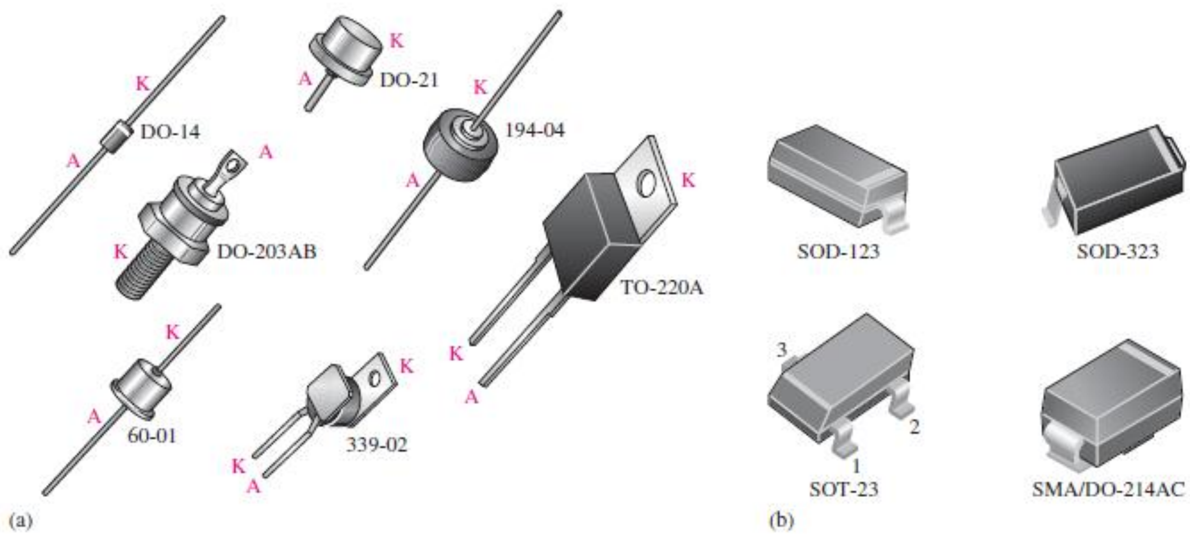


Figure 2–2: Typical diode packages with terminal identification. The letter K is used for cathode to avoid confusion with certain electrical quantities that are represented by *C*. Case type numbers are indicated for each diode.

Forward Bias

To **bias** a diode, you apply a dc voltage across it. **Forward bias** is the condition that allows current through the ***pn*** junction. Figure 2–3 shows a dc voltage source connected by conductive material (contacts and wire) across a diode in the direction to produce forward bias. This external bias voltage is designated as *V*BIAS. The resistor limits the forward current to a value that will not damage the diode. Notice that the negative side of *V*BIAS is connected to the *n* region of the diode and the positive side is connected to the *p* region. This is one requirement for forward bias. A second requirement is that the bias voltage, *V*BIAS, must be greater than the **barrier potential**.

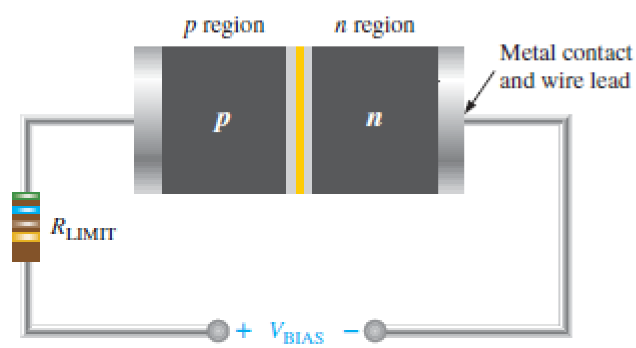


Figure 2–3: A diode connected for forward bias.

A fundamental picture of what happens when a diode is forward-biased is shown in Figure 2–4. Because like charges repel, the negative side of the bias-voltage source "pushes" the free electrons, which are the majority carriers in the *n* region, toward the ***pn*** junction. This flow of free electrons is called *electron current*. The negative side of the

source also provides a continuous flow of electrons through the external connection (conductor) and into the *n* region as shown. The bias-voltage source imparts sufficient energy to the free electrons for them to overcome the barrier potential of the depletion region and move on through into the *p* region. Once in the *p* region, these conduction electrons have lost enough energy to immediately
combine with holes in the valence band. Now, the electrons are in the valence band in the *p* region, simply because they have lost too much energy overcoming the barrier potential to remain in the conduction band. Since unlike charges attract, the positive side of the bias-voltage source attracts the valence electrons toward the left end of the *p* region. The holes in the *p* region provide the medium or "pathway" for these valence electrons to move through the *p* region. The valence electrons move from one hole to the next toward the left. The holes, which are the majority carriers in the *p* region, effectively (not actually) move to the right toward the junction, as you can see in Figure 2–4. This *effective* flow of holes is the hole current. You can also view the hole current as being created by the flow of valence electrons through the *p* region, with the holes providing the only means for
these electrons to flow.

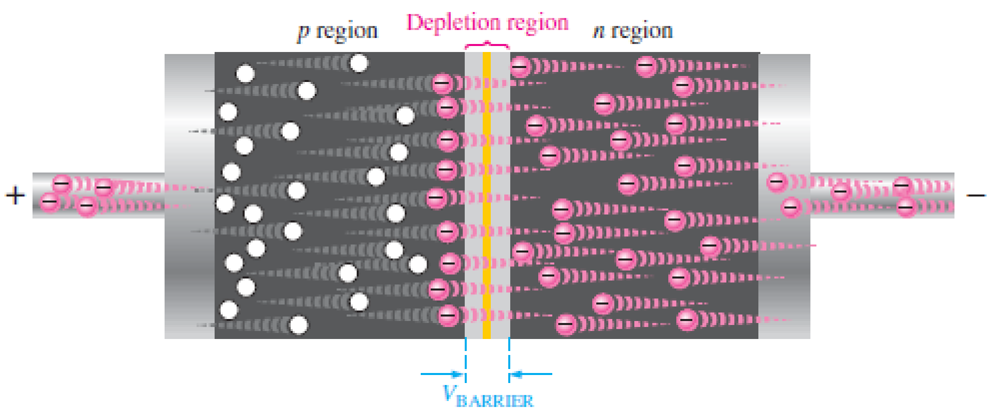


Figure 2–4: A forward-biased diode showing the flow of majority carriers and the voltage due to the barrier potential across the depletion region.

As the electrons flow out of the *p* region through the external connection (conductor) and to the positive side of the bias-voltage source, they leave holes behind in the *p* region; at the same time, these electrons become conduction electrons in the metal conductor. Recall that the conduction band in a conductor overlaps the valence band so that it takes much less energy for an electron to be a free electron in a conductor than in a semiconductor and that metallic conductors do not have holes in their structure. There is a continuous availability of holes effectively moving toward the *pn* junction to combine with the continuous stream of electrons as they come across the junction into the *p* region.

The Effect of Forward Bias on the Depletion Region As more electrons flow into the depletion region, the number of positive ions is reduced. As more holes effectively flow into the depletion region on the other side of the *pn* junction, the number of negative ions is reduced. This reduction in positive and negative ions during forward bias causes the depletion region to narrow, as indicated in Figure 2–5.

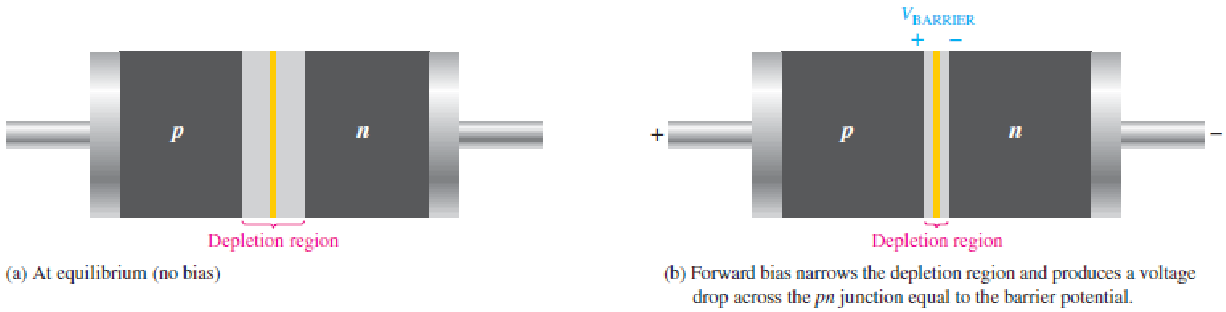


Figure 2–5: The depletion region narrows and a voltage drop is produced across the ***pn*** junction when the diode is forward-biased.

The Effect of the Barrier Potential During Forward Bias Recall that the electric field between the positive and negative ions in the depletion region on either side of the junction creates an "energy hill" that prevents free electrons from diffusing across the junction at equilibrium. This is known as the *barrier potential.* When forward bias is applied, the free electrons are provided with enough energy from the bias-voltage source to overcome the barrier potential and effectively "climb the energy hill" and cross the depletion region. The energy that the electrons require in order to pass through the depletion region is equal to the barrier potential. In other words, the electrons give up an amount of energy equivalent to the barrier potential when they cross the depletion region. This energy loss results in a voltage drop across the *pn* junction equal to the barrier potential (0.7 V), as indicated in Figure 2–5(b). An additional small voltage drop occurs across the *p* and *n* regions due to the internal resistance of the material. For doped semiconductive material, this resistance, called the **dynamic resistance,** is very small and can usually be neglected. This is discussed in more detail in Section 2–2.

Reverse Bias

Reverse bias is the condition that essentially prevents current through the diode. Figure 2–6 shows a dc voltage source connected across a diode in the direction to produce reverse bias. This external bias voltage is designated as *V*BIAS just as it was for forward bias. Notice that the positive side of *V*BIAS is connected to the *n* region of the diode and the negative side is connected to the *p* region. Also note that the depletion region is shown much wider than in forward bias or equilibrium.

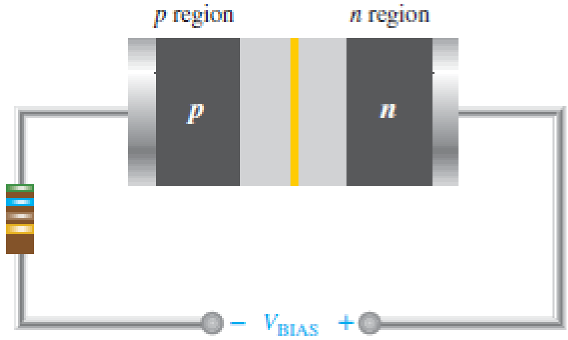


Figure 2–6: A diode connected for reverse bias. A limiting resistor is shown although it is not important in reverse bias because there is essentially no current.

An illustration of what happens when a diode is reverse-biased is shown in Figure 2–7. Because unlike charges attract, the positive side of the bias-voltage source "pulls" the free electrons, which are the majority carriers in the *n* region, away from the *pn* junction. As the electrons flow toward the positive side of the voltage source, additional positive ions are created. This results in a widening of the depletion region and a depletion of majority carriers. In the *p* region, electrons from the negative side of the voltage source enter as valence electrons and move from hole to hole toward the depletion region where they create additional negative ions. This results in a widening of the depletion region and a depletion of majority carriers. The flow of valence electrons can be viewed as holes being "pulled" toward the positive side. The initial flow of charge carriers is transitional and lasts for only a very short time after the reverse-bias voltage is applied. As the depletion region widens, the availability of majority carriers decreases.

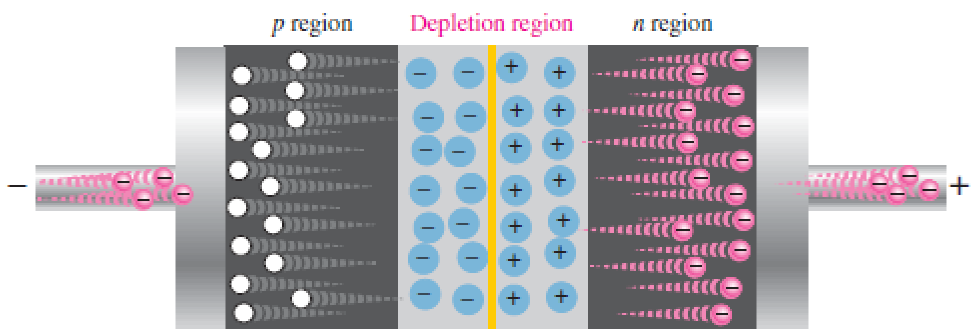


Figure 2–7: The diode during the short transition time immediately after reverse-bias voltage is applied.

As more of the *n* and *p* regions become depleted of majority carriers, the electric field between the positive and negative ions increases in strength until the potential across the

depletion region equals the bias voltage, *V*BIAS. At this point, the transition current essentially ceases except for a very small reverse current that can
usually be neglected.

Reverse Current The extremely small current that exists in reverse bias after the transition current dies out is caused by the minority carriers in the *n* and *p* regions that are produced by thermally generated electron-hole pairs. The small number of free minority electrons in the *p* region are "pushed" toward the *pn* junction by the negative bias voltage. When these electrons reach the wide depletion region, they "fall down the energy hill" and combine with the minority holes in the *n* region as valence electrons and flow toward the positive bias voltage, creating a small hole current. The conduction band in the *p* region is at a higher energy level than the conduction band in the *n* region. Therefore, the minority electrons easily pass through the depletion region because they require no additional energy. Reverse current is illustrated in Figure 2–8.

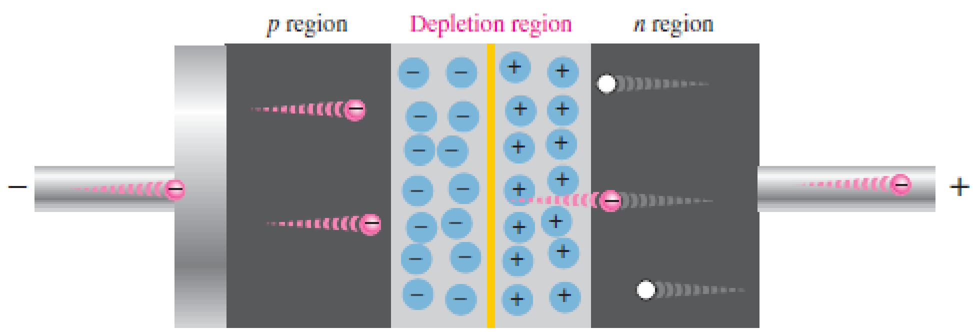


Figure 2–8: The extremely small reverse current in a reverse-biased diode is due to the minority carriers from thermally generated electron-hole pairs.

Reverse Breakdown Normally, the reverse current is so small that it can be neglected. However, if the external reverse-bias voltage is increased to a value called the *breakdown voltage*, the reverse current will drastically increase. This is what happens. The high reverse-bias voltage imparts energy to the free minority electrons so that as they speed through the *p* region, they collide with atoms with enough energy to knock valence electrons out of orbit and into the conduction band. The newly created conduction electrons are also high in energy and repeat the process. If one electron knocks only two others out of their valence orbit during its travel through the *p* region, the numbers quickly multiply. As these high-energy electrons go through the depletion region, they have enough energy to go through the *n* region as conduction electrons, rather than combining with holes. The multiplication of conduction electrons just discussed is known as the **avalanche effect,** and reverse current can increase dramatically if steps are not taken to limit the current. When the reverse current is not limited, the resulting heating will permanently damage the diode. Most diodes are not operated in reverse breakdown, but if the current is limited (by adding a series-limiting resistor for example), there is no permanent damage to the diode.

SECTION 2–1 CHECKUP

1. Describe forward bias of a diode.
2. Explain how to forward-bias a diode.
3. Describe reverse bias of a diode.
4. Explain how to reverse-bias a diode.
5. Compare the depletion regions in forward bias and reverse bias.
6. Which bias condition produces majority carrier current?
7. How is reverse current in a diode produced?
8. When does reverse breakdown occur in a diode?
9. Define *avalanche effect* as applied to diodes.

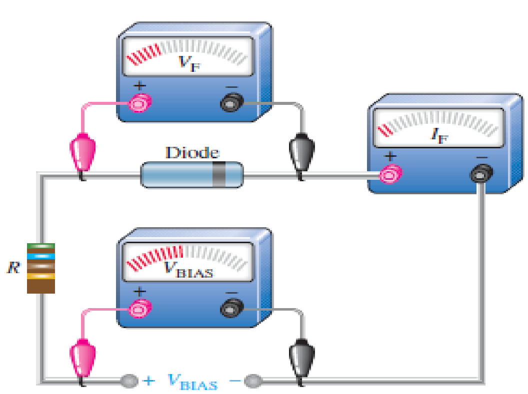
2–2 Voltage-current characteristic of a diode

As you have learned, forward bias produces current through a diode and reverse bias essentially prevents current, except for a negligible reverse current. Reverse bias prevents current as long as the reverse-bias voltage does not equal or exceed the breakdown voltage of the junction. In this section, we will examine the relationship between the voltage and the current in a diode on a graphical basis.

V-I Characteristic for Forward Bias

When a forward-bias voltage is applied across a diode, there is current. This current is called the *forward current* and is designated I_F. Figure 2–9 illustrates what happens as the forward-bias voltage is increased positively from 0 V. The resistor is used to limit the forward current to a value that will not overheat the diode and cause damage. With 0 V across the diode, there is no forward current. As you gradually increase the forward-bias voltage, the forward current *and* the voltage across the diode gradually increase, as shown in Figure 2–9(a). A portion of the forward-bias voltage is dropped across the limiting resistor. When the forward-bias voltage is increased to a value where the voltage across the diode reaches approximately 0.7 V (barrier potential), the forward current begins to increase rapidly, as illustrated in Figure 2–9(b). As you continue to increase the forward-bias voltage, the current continues to increase very rapidly, but the voltage across the diode increases only gradually above 0.7 V. This small increase in the diode voltage

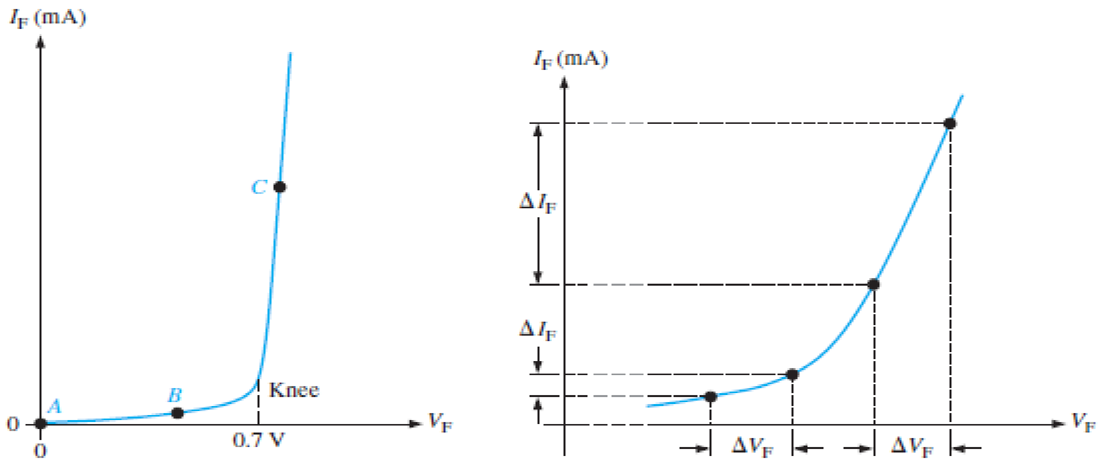
above the barrier potential is due to the voltage drop across the internal dynamic resistance of the semiconductive material. ***Graphing the V-I Curve*** If you plot the results of the type of measurements shown in Figure 2–9 on a graph, you get the **V-I characteristic** curve for a forward-biased diode, as shown in Figure 2–10(a). The diode forward voltage (*V*F) increases to the right along the
horizontal axis, and the forward current (*I*F) increases upward along the vertical axis.



(a) Small forward-bias voltage (V_F < 0.7 V), very small forward current.

(b) Forward voltage reaches and remains nearly constant at approximately 0.7 V. Forward current continues to increase as the bias voltage is increased.

Figure 2–9: Forward-bias measurements show general changes in *V*F and *I*F as *V*BIAS is increased.



(a) *V-I* characteristic curve for forward bias.

(b) Expanded view of a portion of the curve in part (a). The dynamic resistance r'_d decreases as you move up the curve, as indicated by the decrease in the value of $\Delta V_F/\Delta I_F$.

Figure 2–10: Relationship of voltage and current in a forward-biased diode.

As you can see in Figure 2–10(a), the forward current increases very little until the forward voltage across the *pn* junction reaches approximately 0.7 V at the knee of the curve. After this point, the forward voltage remains nearly constant at approximately 0.7 V, but *I*F increases rapidly. As previously mentioned, there is a slight increase in *V*F above 0.7 V as the current increases due mainly to the voltage drop across the dynamic resistance. The *I*F scale is typically in mA, as indicated.

Three points *A*, *B*, and *C* are shown on the curve in Figure 2–10(a). Point *A* corresponds to a zero-bias condition. Point *B* corresponds to Figure 2–10(a) where the forward voltage is less than the barrier potential of 0.7 V. Point *C* corresponds to Figure 2–10(a) where the forward voltage *approximately* equals the barrier potential. As the external bias voltage and forward current continue to increase above the knee, the forward voltage will increase slightly above 0.7 V. In reality, the forward voltage can be as much as approximately 1 V, depending on the forward current.

Dynamic Resistance Figure 2–10(b) is an expanded view of the *V-I* characteristic curve in part (a) and illustrates dynamic resistance. Unlike a linear resistance, the resistance of the forward-biased diode is not constant over the entire curve. Because the resistance changes as you move along the *V-I* curve, it is called *dynamic* or *ac resistance*. Internal resistances of electronic devices are usually designated by lowercase italic *r* with a prime, instead of the standard *R*. The dynamic resistance of a diode is designated r_d'. Below the knee of the curve the resistance is greatest because the current increases very little for a given change in voltage ($r_d' = \frac{\Delta V_F}{\Delta I_F}$). The resistance begins to decrease in the region of the knee of the curve and becomes smallest above the knee where there is a large change in current for a given change in voltage.

V-I Characteristic for Reverse Bias

When a reverse-bias voltage is applied across a diode, there is only an extremely small reverse current (I_R) through the *pn* junction. With 0 V across the diode, there is no reverse current. As you gradually increase the reverse-bias voltage, there is a very small reverse current and the voltage across the diode increases. When the applied bias voltage is increased to a value where the reverse voltage across the diode (V_R) reaches the breakdown value (V_{BR}), the reverse current begins to increase rapidly. As you continue to increase the bias voltage, the current continues to increase very rapidly, but the voltage across the diode increases very little above V_{BR}. Breakdown, with exceptions, is not a normal mode of operation for most ***pn*** junction devices.

Graphing the V-I Curve

If you plot the results of reverse-bias measurements on a graph, you get the *V-I* characteristic curve for a reverse-biased diode. A typical curve is shown in Figure 2–11. The diode reverse voltage (V_R) increases to the left along the horizontal axis,

and the reverse current (I_R) increases downward along the vertical axis. There is very little reverse current (usually µA or nA) until the reverse voltage across the diode reaches approximately the breakdown value (V_{BR}) at the knee of the curve. After this point, the reverse voltage remains at approximately V_{BR}, but I_R increases very rapidly, resulting in overheating and possible damage if current is not limited to a safe level. The breakdown voltage for a diode depends on the doping level, which the manufacturer sets, depending on the type of diode. A typical rectifier diode (the most widely used type) has a breakdown voltage of greater than 50 V. Some specialized diodes have a breakdown voltage that is only 5 V.

The Complete *V-I* Characteristic Curve

Combine the curves for both forward bias and reverse bias, and you have the complete *V-I* characteristic curve for a diode, as shown in Figure 2–12.

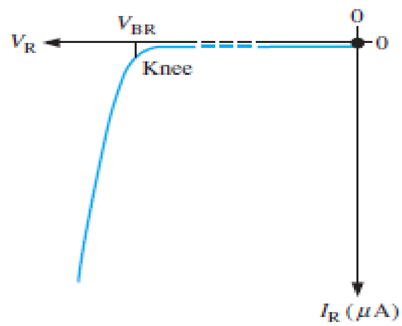


Figure 2–11: *V-I* characteristic curve for a reversebiased diode.

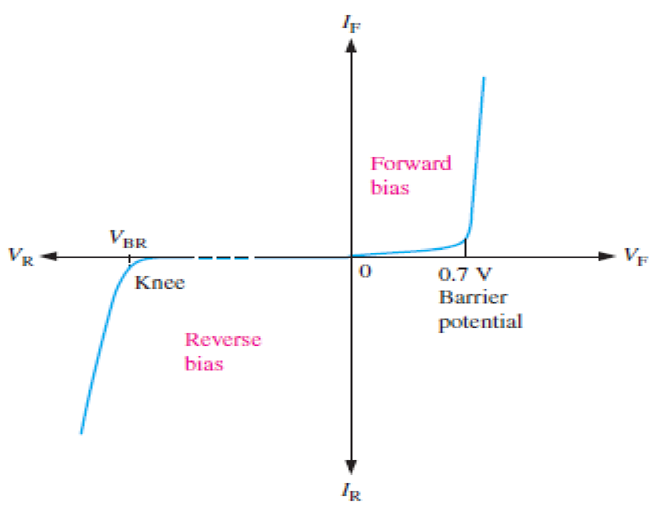


Figure 2–12: The complete *V-I* characteristic curve for a diode.

Temperature Effects

For a forward-biased diode, as temperature is increased, the forward current increases for a given value of forward voltage. Also, for a given value of forward current, the forward voltage decreases. This is shown with the *V-I* characteristic curves in Figure 2–13. The blue curve is at room temperature (25°C) and the red curve is at an elevated temperature (25°C + ΔT). The barrier potential decreases by 2 mV for each degree increase in temperature.

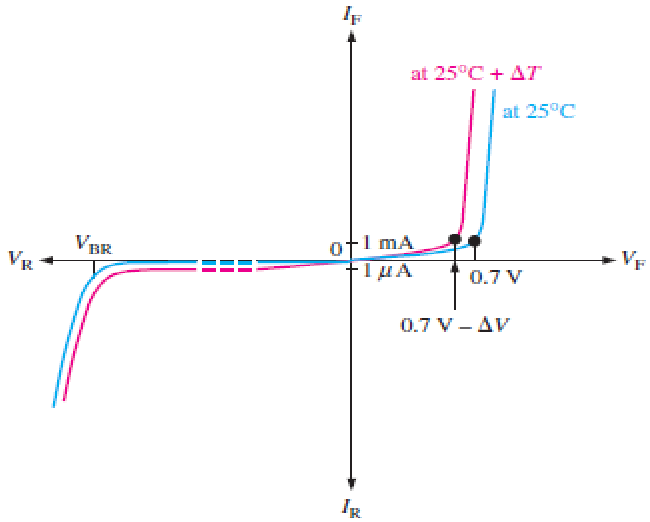


Figure 2–13: Temperature effect on the diode *V-I* characteristic. The 1 mA and 1 μA marks on the vertical axis are given as a basis for a relative comparison of the current scales.

For a reverse-biased diode, as temperature is increased, the reverse current increases. The difference in the two curves is exaggerated on the graph in Figure 2–13 for illustration. Keep in mind that the reverse current below breakdown remains extremely small and can usually be neglected.

SECTION 2–2 CHECKUP

1. Discuss the significance of the knee of the characteristic curve in forward bias.
2. On what part of the curve is a forward-biased diode normally operated?
3. Which is greater, the breakdown voltage or the barrier potential?
4. On what part of the curve is a reverse-biased diode normally operated?
5. What happens to the barrier potential when the temperature increases?

2–3 Diode models

You have learned that a diode is a *pn* junction device. In this section, you will learn the electrical symbol for a diode and how a diode can be modeled for circuit analysis using any one of three levels of complexity. Also, diode packaging and terminal identification are introduced.

Bias Connections

Forward-Bias**:** Recall that a diode is forward-biased when a voltage source is connected as shown in Figure 2–14(a). The positive terminal of the source is connected to the anode through a current-limiting resistor. The negative terminal of the source is connected to the cathode. The forward current (I_F) is from cathode to anode as indicated. The forward voltage drop (V_F) due to the barrier potential is from positive at the anode to negative at the cathode.

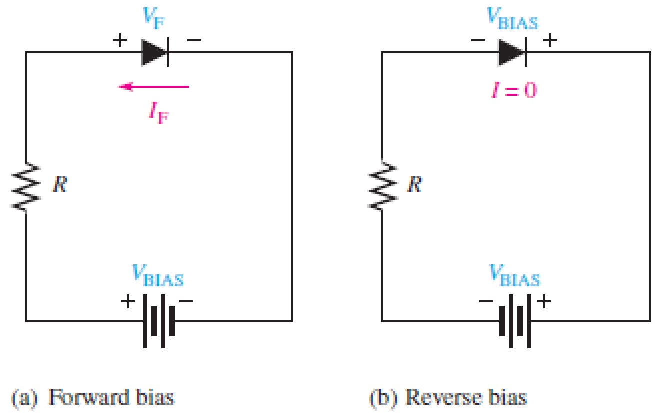


Figure 2–14: Forward-bias and reverse-bias connections showing the diode symbol.

***Reverse-Bias Connection*:** A diode is reverse-biased when a voltage source is connected as shown in Figure 2–14(b). The negative terminal of the source is connected to the anode side of the circuit, and the positive terminal is connected to the cathode side. A resistor is not necessary in reverse bias but it is shown for circuit consistency. The reverse current is extremely small and can be considered to be zero. Notice that the entire bias voltage (V_{BIAS}) appears across the diode.

Diode Approximations

The Ideal Diode Model The ideal model of a diode is the least accurate approximation and can be represented by a simple switch. When the diode is forward-biased, it ideally acts like a closed (on) switch, as shown in Figure 2–15(a). When the diode is reverse-biased, it ideally acts like an open (off) switch, as shown in part (b). Although the barrier potential, the forward dynamic resistance, and the reverse current are all neglected, this model is adequate for most troubleshooting when you are trying to determine if the diode is working properly.

In Figure 2–15(c), the ideal *V-I* characteristic curve graphically depicts the ideal diode operation. Since the barrier potential and the forward dynamic resistance are neglected, the diode is assumed to have a zero voltage across it when forward-biased, as indicated by the portion of the curve on the positive vertical axis.

$$V_F = 0 \text{ V}$$

The forward current is determined by the bias voltage and the limiting resistor using Ohm's law.

$$I_F = \frac{V_{Bais}}{R_{limit}} \tag{2-1}$$

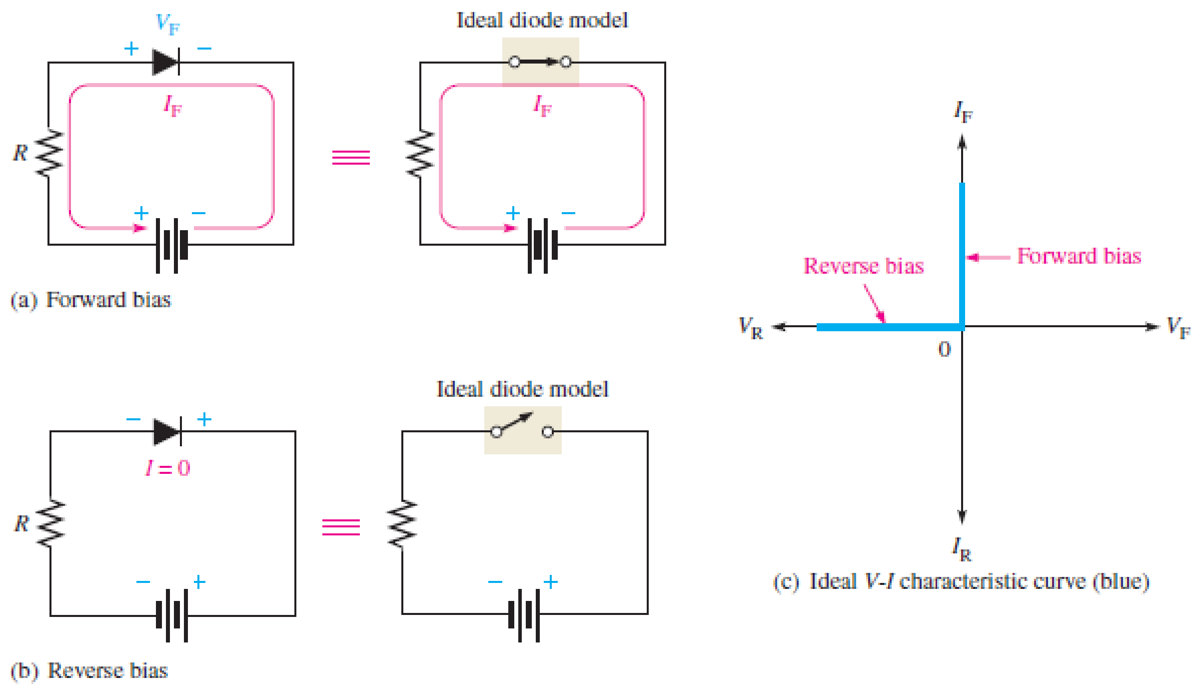


Figure 2–15: The ideal model of a diode.

Since the reverse current is neglected, its value is assumed to be zero, as indicated in Figure 2–15(c) by the portion of the curve on the negative horizontal axis.

$$I_R = 0 \text{ A}$$

The reverse voltage equals the bias voltage.

$$V_R = V_{BIAS}$$

You may want to use the ideal model when you are troubleshooting or trying to figure out the operation of a circuit and are not concerned with more exact values of voltage or current.

***The Practical Diode Model*:** The practical model includes the barrier potential. When the diode is forward-biased, it is equivalent to a closed switch in series with a small equivalent voltage source (V_F) equal to the barrier potential (0.7 V) with the positive side toward the anode, as indicated in Figure 2–16(a). This equivalent voltage source represents the barrier potential that must be exceeded by the bias voltage before the diode will conduct and is not an active source of voltage. When conducting, a voltage drop of 0.7 V appears across the diode.

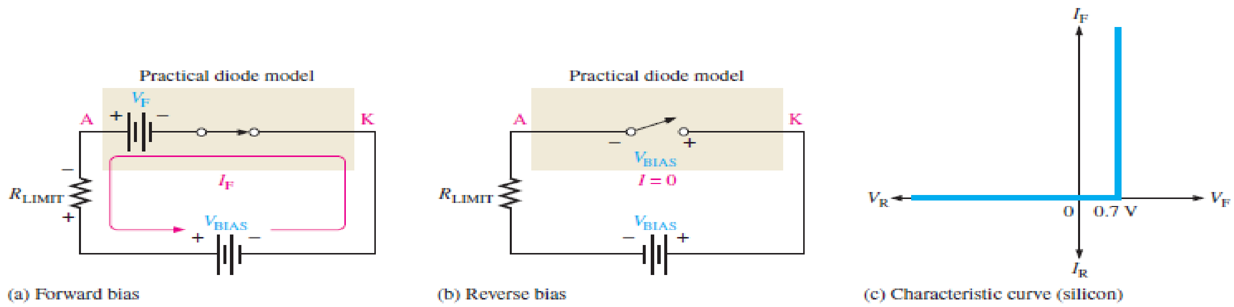


Figure 2–16: The practical model of a diode.

When the diode is reverse-biased, it is equivalent to an open switch just as in the ideal model, as shown in Figure 2–16(b). The barrier potential does not affect reverse bias, so it is not a factor. The characteristic curve for the practical diode model is shown in Figure 2–16(c). Since the barrier potential is included and the dynamic resistance is neglected, the diode is assumed to have a voltage across it when forward-biased, as indicated by the portion of the curve to the right of the origin.

$$V_F = 0.7\ \text{V}$$

The forward current is determined as follows by first applying Kirchhoff's voltage law to Figure 2–16(a):

$$V_{BIAS} - V_F - V_{R(LIMIT)} = 0$$

Substituting and solving for I_F,

$$I_F = \frac{V_{Bais} - V_F}{R_{limit}} \qquad (2\text{-}2)$$

The diode is assumed to have zero reverse current, as indicated by the portion of the curve on the negative horizontal axis.

$$I_R = 0\ \text{A}$$
$$V_R = V_{BIAS}$$

The practical model is useful when you are troubleshooting in lower-voltage circuits. In these cases, the 0.7 V drop across the diode may be significant and should be taken into account. The practical model is also useful when you are designing basic diode circuits.

***The Complete Diode Model*:** The complete model of a diode is the most accurate approximation and includes the barrier potential, the small forward dynamic resistance r'_d and the large internal reverse resistance r'_R. The reverse resistance is taken into account because it provides a path for the reverse current, which is included in this diode model. When the diode is forward-biased, it acts as a closed switch in series with the equivalent barrier potential voltage (V_B) and the small forward dynamic resistance r'_d as indicated in Figure 2–17(a). When the diode is reverse-biased, it acts as an open switch in parallel with the large internal reverse resistance r'_R as shown in Figure 2–17(b). The barrier potential does not affect reverse bias, so it is not a factor.

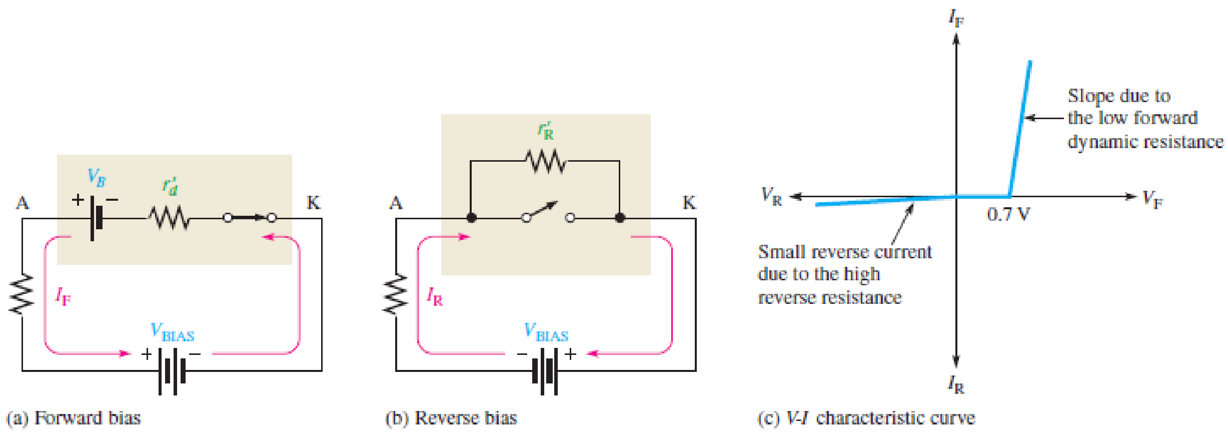


Figure 2–17: The complete model of a diode.

The characteristic curve for the complete diode model is shown in Figure 2–17(c).
Since the barrier potential and the forward dynamic resistance are included, the diode is assumed to have a voltage across it when forward-biased. This voltage (V_F) consists of the barrier potential voltage plus the small voltage drop across the dynamic resistance, as indicated by the portion of the curve to the right of the origin. The curve slopes because the voltage drop due to dynamic resistance increases as the current increases. For the complete model of a silicon diode, the following formulas apply:

$$V_F = 0.7\ V + I_R\ r'_R$$

$$I_F = \frac{V_{Bais} - 0.7\ V}{R_{limit} + r'_R}$$

The reverse current is taken into account with the parallel resistance and is indicated by the portion of the curve to the left of the origin. The breakdown portion of the curve is not shown because breakdown is not a normal mode of operation for most diodes. For troubleshooting work, it is unnecessary to use the complete model, as it involves complicated calculations. This model is generally suited to design problems using a computer for simulation. The ideal and practical models are used for circuits in this text, except in the following example, which illustrates the differences in the three models.

Example 2–1

(a) Determine the forward voltage and forward current for the diode in Figure 2–18(a) for each of the diode models. Also find the voltage across the limiting resistor in each case. Assume r'_R = 10 Ω at the determined value of forward current.

(b) Determine the reverse voltage and reverse current for the diode in Figure 2–18(b) for each of the diode models. Also find the voltage across the limiting resistor in each case. Assume $I_R = 1\ \mu A$.

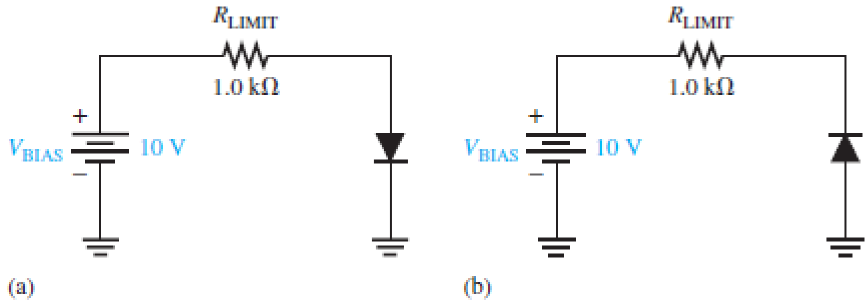


Figure 2–18

<u>Solution</u>

(a) Ideal model:

$V_F = \mathbf{0\ V}$

$$I_F = \frac{V_{Bais}}{R_{limit}} = \frac{10\ V}{1.0\ k\Omega} = 10\ mA$$

$$V_{R_{Limit}} = I_F\ R_{Limit} = (10\ mA)(1.0\ kA) = 10\ V$$

Practical model:

$V_F = \mathbf{0.7\ V}$

$$I_F = \frac{V_{Bais} - V_F}{R_{limit}} = \frac{10\ V - 0.7\ V}{1.0\ k\Omega} = \frac{9.3\ V}{1.0\ k\Omega} = 9.3\ mA$$

$$V_{R_{Limit}} = I_F\ R_{Limit} = (9.3\ mA)(1.0\ kA) = 9.3\ V$$

Complete model:

$$I_F = \frac{V_{Bais} - V_F}{R_{limit} + r_d^{'}} = \frac{10\ V - 0.7\ V}{1.0\ k\Omega + 10\ \Omega} = \frac{9.3\ V}{1010\ \Omega} = 9.21\ mA$$

$$V_F = 0.7\ V + I_F\ r_d^{'} = 0.7\ V + (9.21\ mA)(10\ \Omega) = 792\ mV$$

$$V_{R_{Limit}} = I_F\ R_{Limit} = (9.21\ mA)(1.0\ k\Omega) = 9.21\ V$$

(b) Ideal model:

$I_R = \mathbf{0\ A}$

$V_R = V_{BIAS} = \mathbf{10\ V}$

$$V_{R_{Limit}} = 0\ V$$

Practical model:

$I_R = \mathbf{0\ A}$

$V_R = V_{BIAS} = \mathbf{10\ V}$

$$V_{R_{Limit}} = 0\ V$$

Complete model:

I_R = **1** μ**A**

$$V_{R_{Limit}} = I_R\, R_{Limit} = (1\ \mu A)(1.0\ k\Omega) = 1\ mV$$

$$V_R = V_{Bias} - V_{R_{Limit}} = 10\ V - 1\ mV = 9.999\ V$$

Related Problem: Assume that the diode in Figure 2–18(a) fails open. What is the voltage across the diode and the voltage across the limiting resistor?

SECTION 2–3 CHECKUP

1. What are the two conditions under which a diode is operated?
2. Under what condition is a diode never intentionally operated?
3. What is the simplest way to visualize a diode?
4. To more accurately represent a diode, what factors must be included?
5. Which diode model represents the most accurate approximation?

2–4 Half wave rectifiers

Because of their ability to conduct current in one direction and block current in the other direction, diodes are used in circuits called rectifiers that convert ac voltage into dc voltage. Rectifiers are found in all dc power supplies that operate from an ac voltage source. A power supply is an essential part of each electronic system from the simplest to the most complex.

The Basic DC Power Supply

All active electronic devices require a source of constant dc that can be supplied by a battery or a dc power supply. The **dc power supply** converts the standard 120 V, 60 Hz ac voltage available at wall outlets into a constant dc voltage. The dc power supply is one of the most common circuits you will find, so it is important to understand how it works. The voltage produced is used to power all types of electronic circuits including consumer electronics (televisions, DVDs, etc.), computers, industrial controllers, and most laboratory instrumentation systems and equipment. The dc voltage level required depends on the application, but most applications require relatively low voltages.

A basic block diagram of the complete power supply is shown in Figure 2–19(a).

Generally the ac input line voltage is stepped down to a lower ac voltage with a transformer (although it may be stepped up when higher voltages are needed or there may be no transformer at all in rare instances). As you learned in your dc/ac course, a **transformer** changes ac voltages based on the turns ratio between the primary and secondary. If the secondary has more turns than the primary, the output voltage across the secondary will be higher and the current will be smaller. If the secondary has fewer turns than the primary, the output voltage across the secondary will be lower and the current will be higher. The rectifier can be either a half-wave rectifier or a full-wave rectifier (covered in Section 2–5). The **rectifier** converts the ac input voltage to a pulsating dc voltage, called a half-wave rectified voltage, as shown in Figure 2–19(b). The **filter** eliminates the fluctuations in the rectified voltage and produces a relatively smooth dc voltage. The power supply filter is covered in Section 2–6. The **regulator** is a circuit that maintains a constant dc voltage for variations in the input line voltage or in the load. Regulators vary from a single semiconductor device to more complex integrated circuits. The load is a circuit or device connected to the output of the power supply and operates from the power supply voltage and current.

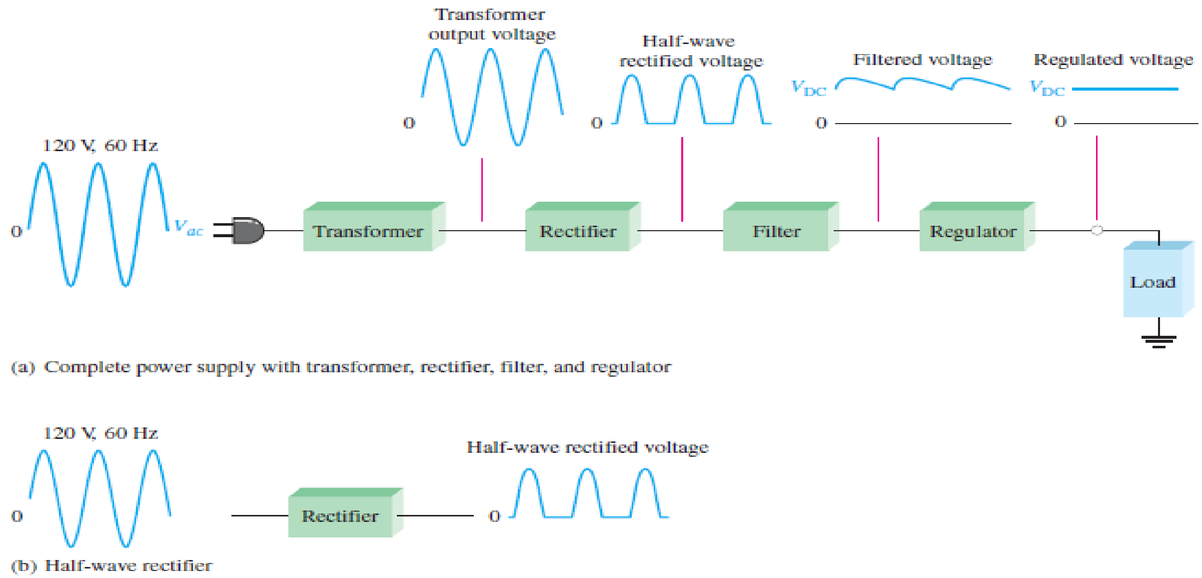
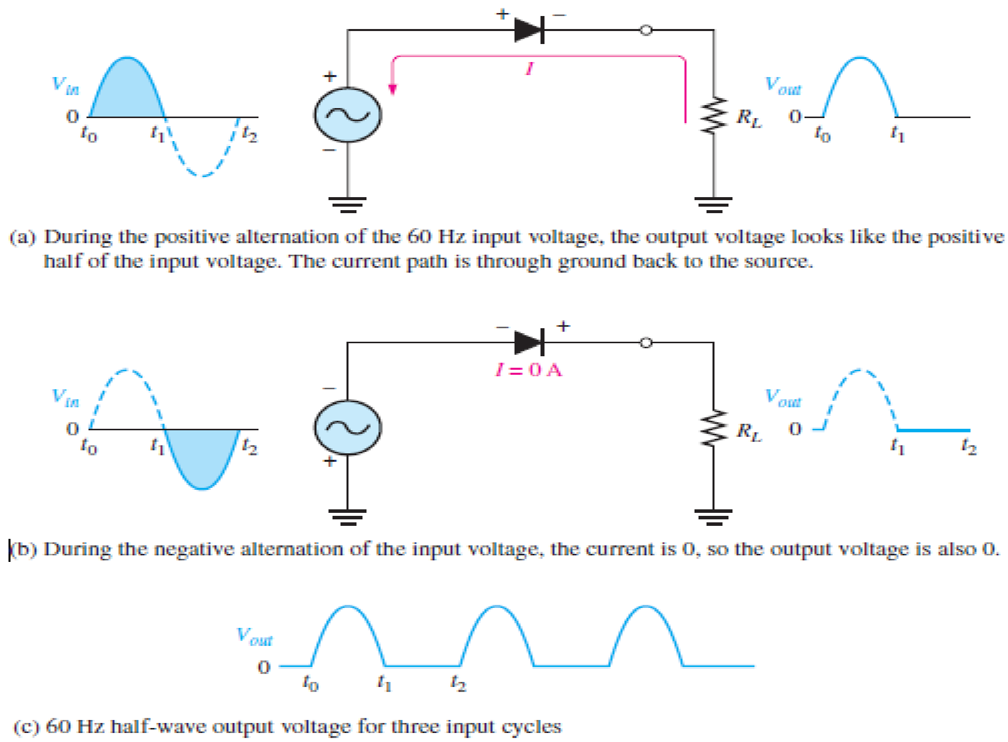


Figure 2–19: Block diagram of a dc power supply with a load and a rectifier.

Half-Wave Rectifier Operation

Figure 2–20 illustrates the process called *half-wave rectification.* A diode is connected to an ac source and to a load resistor, R_L, forming a **half-wave rectifier**. Keep in mind that all ground symbols represent the same point electrically. Let's examine what happens during one cycle of the input voltage using the ideal model for the diode. When the sinusoidal input voltage (V_{in}) goes positive, the diode is forward-biased and conducts current through the load resistor, as shown in part (a). The current produces an output voltage across the load R_L, which has the same shape as the positive half-cycle of the input voltage.



(a) During the positive alternation of the 60 Hz input voltage, the output voltage looks like the positive half of the input voltage. The current path is through ground back to the source.

(b) During the negative alternation of the input voltage, the current is 0, so the output voltage is also 0.

(c) 60 Hz half-wave output voltage for three input cycles

Figure 2–20: Half-wave rectifier operation. The diode is considered to be ideal.

When the input voltage goes negative during the second half of its cycle, the diode is reverse-biased. There is no current, so the voltage across the load resistor is 0 V, as shown in Figure 2–20(b). The net result is that only the positive half-cycles of the ac input voltage appear across the load. Since the output does not change polarity, it is a pulsating dc voltage with a frequency of 60 Hz, as shown in part (c).

Average Value of the Half-Wave Output Voltage: The average value of the half-wave rectified output voltage is the value you would measure on a dc voltmeter. Mathematically, it is determined by finding the area under the curve over a full cycle, as illustrated in Figure 2–21, and then dividing by 2π the number of radians in a full cycle. The result of this is expressed in Equation 2–3, where V_p is the peak value of the voltage. This equation shows that V_{AVG} is approximately 31.8% of V_p for a half-wave rectified voltage.

$$V_{AVG} = \frac{V_P}{\pi} \tag{2-3}$$

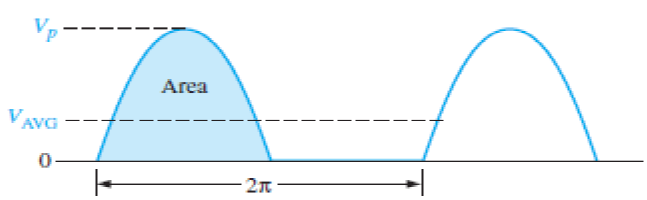


Figure 2–21: Average value of the half-wave rectified signal.

Example 2–2

What is the average value of the half-wave rectified voltage in Figure 2–22?

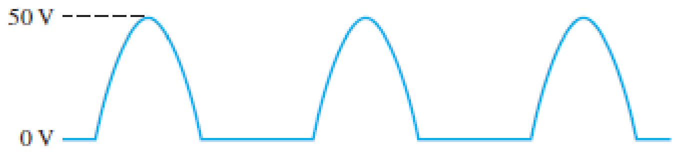


Figure 2–22:

Solution:

$$V_{AVG} = \frac{V_p}{\pi} = \frac{50\ V}{\pi} = 15.9\ V$$

Notice that *V*AVG is 31.8% of V_p.

***Related Problem*:** Determine the average value of the half-wave voltage if its peak amplitude is 12 V.

Effect of the Barrier Potential on the Half-Wave Rectifier Output

In the previous discussion, the diode was considered ideal. When the practical diode model is used with the barrier potential of 0.7 V taken into account, this is what happens. During the positive half-cycle, the input voltage must overcome the barrier potential before the diode becomes forward-biased. This results in a half-wave output with a peak value that is 0.7 V less than the peak value of the input, as shown in Figure 2–23. The expression for the peak output voltage is

$$V_{p(out)} = V_{p(in)} - 0.7\ V \qquad (2\text{-}4)$$

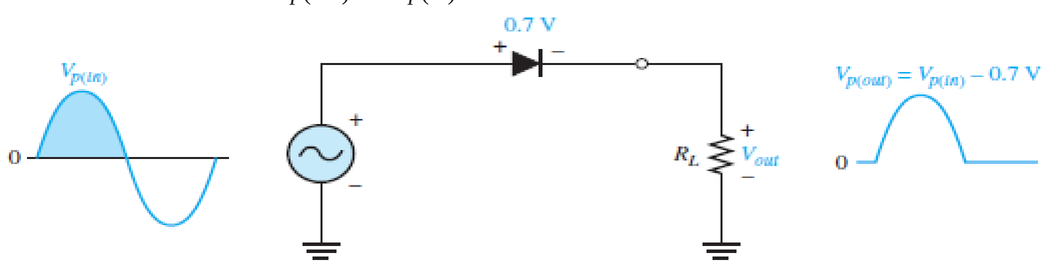


Figure 2–23: The effect of the barrier potential on the half-wave rectified output voltage is to reduce the peak value of the input by about 0.7 V.

It is usually acceptable to use the ideal diode model, which neglects the effect of the barrier potential, when the peak value of the applied voltage is much greater than the barrier potential (at least 10 V, as a rule of thumb). However, we will use the practical model of a diode, taking the 0.7 V barrier potential into account unless stated otherwise.

Example 2–3:

Draw the output voltages of each rectifier for the indicated input voltages, as shown in Figure 2–24. The 1N4001 and 1N4003 are specific rectifier diodes.

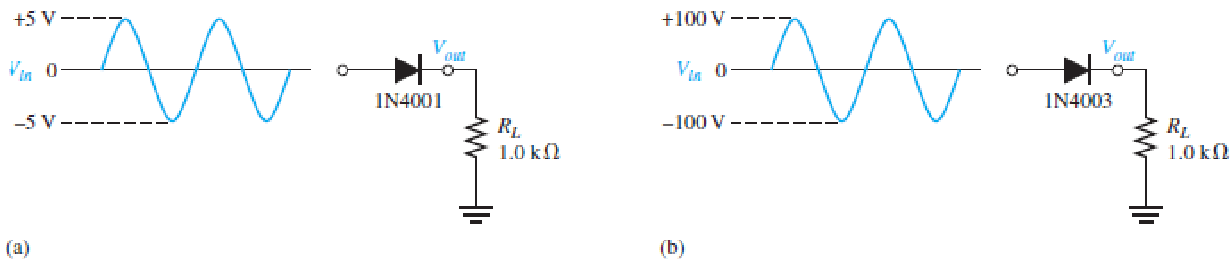


Figure 2–24

Solution

The peak output voltage for circuit (a) is

$$V_{p(out)} = V_{p(in)} - 0.7\text{ V} = 5\text{ V} - 0.7\text{ V} = \mathbf{4.30\ V}$$

The peak output voltage for circuit (b) is

$$V_{p(out)} = V_{p(in)} - 0.7\text{ V} = 100\text{ V} - 0.7\text{ V} = \mathbf{99.3\ V}$$

The output voltage waveforms are shown in Figure 2–25. Note that the barrier potential could have been neglected in circuit (b) with very little error (0.7 percent); but, if it is neglected in circuit (a), a significant error results (14 percent).

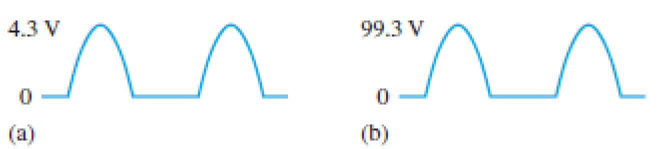


Figure 2–25: Output voltages for the circuits in Figure 2–24. They are not shown on the same scale.

***Related Problem*:**

Determine the peak output voltages for the rectifiers in Figure 2–24 if the peak input in part (a) is 3 V and the peak input in part (b) is 50 V.

Peak Inverse Voltage (PIV)

The **peak inverse voltage (PIV)** equals the peak value of the input voltage, and the diode must be capable of withstanding this amount of repetitive reverse voltage. For the diode in Figure 2–26, the maximum value of reverse voltage, designated as PIV, occurs at the peak of each negative alternation of the input voltage when the diode is reverse-biased. A diode should be rated at least 20% higher than the PIV.

$$\mathbf{PIV} = V_{p(in)} \tag{2-5}$$

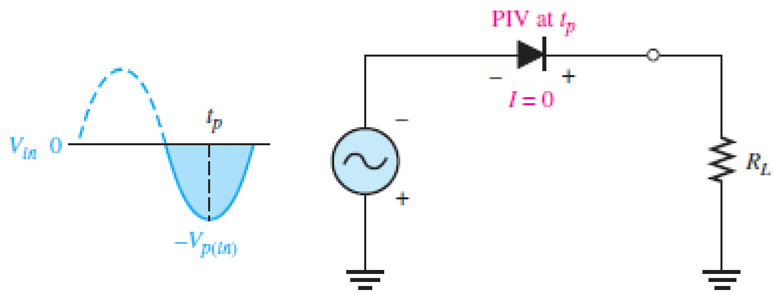


Figure 2–26: The PIV occurs at the peak of each half-cycle of the input voltage when the diode is reverse-biased. In this circuit, the PIV occurs at the peak of each negative half-cycle.

Transformer Coupling

As you have seen, a transformer is often used to couple the ac input voltage from the source to the rectifier, as shown in Figure 2–27. Transformer coupling provides two advantages. First, it allows the source voltage to be stepped down as needed. Second, the ac source is electrically isolated from the rectifier, thus preventing a shock hazard in the secondary circuit.

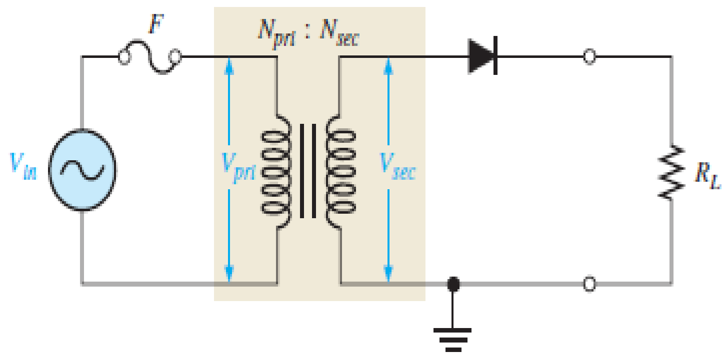


Figure 2–27: Half-wave rectifier with transformer- coupled input voltage.

The amount that the voltage is stepped down is determined by the **turns ratio** of the transformer. Unfortunately, the definition of turns ratio for transformers is not consistent between various sources and disciplines. In this text, we use the definition given by the IEEE for electronic power transformers, which is "the number of turns in the secondary (N_{sec}) divided by the number of turns in the primary (N_{pri})." Thus, a transformer with a turns ratio less than 1 is a step-down type and one with a turns ratio greater than 1 is a step-up type. To show the turns ratio on a schematic, it is common practice to show the numerical ratio directly above the windings.

The secondary voltage of a transformer equals the turns ratio, n, times the primary voltage.

$$V_{sec} = n\,V_{pri}$$

If $n > 1$, the secondary voltage is greater than the primary voltage. If $n < 1$, the secondary voltage is less than the primary voltage. If $n = 1$, then $V_{sec} = V_{pri}$.

The peak secondary voltage, $V_{p(sec)}$, in a transformer-coupled half-wave rectifier is the same as $V_{p(in)}$ in Equation 2–4. Therefore, Equation 2–4 written in terms of $V_{p(sec)}$ is

$$V_{p(out)} = V_{p(sec)} - 0.7\text{ V}$$

and Equation 2–5 in terms of $V_{p(sec)}$ is

$$\text{PIV} = V_{p(sec)}$$

Turns ratio is useful for understanding the voltage transfer from primary to secondary. However, transformer datasheets rarely show the turns ratio. A transformer is generally specified based on the secondary voltage rather than the turns ratio.

Example 2–4:

Determine the peak value of the output voltage for Figure 2–28 if the turns ratio is 0.5.

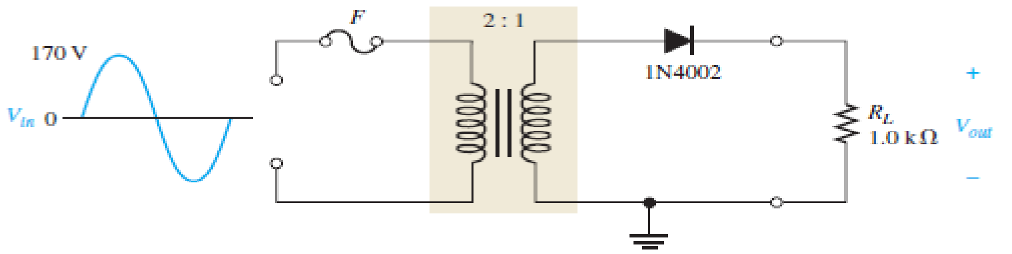


Figure 2–28

Solution

$$V_{p(pri)} = V_{p(in)} = 170\text{ V}$$

The peak secondary voltage is

$$V_{p(sec)} = n\,V_{p(pri)} = 0.5\,(170\text{ V}) = 85\text{ V}$$

The rectified peak output voltage is

$$V_{p(out)} = V_{p(sec)} - 0.7\text{ V} = 85\text{ V} - 0.7\text{ V} = \mathbf{84.3\text{ V}}$$

where $V_{p(sec)}$ is the input to the rectifier.

Related Problem

(a) Determine the peak value of the output voltage for Figure 2–28 if $n = 2$ and $V_{p(in)} =$ 312 V.

(b) What is the PIV across the diode?

(c) Describe the output voltage if the diode is turned around.

SECTION 2–4 CHECKUP

1. At what point on the input cycle does the PIV occur?
2. For a half-wave rectifier, there is current through the load for approximately what percentage of the input cycle?
3. What is the average of a half-wave rectified voltage with a peak value of 10 V?
4. What is the peak value of the output voltage of a half-wave rectifier with a peak sine

wave input of 25 V?

5. What PIV rating must a diode have to be used in a rectifier with a peak output voltage of 50 V?

2–5 Full-wave rectifiers

Although half-wave rectifiers have some applications, the full-wave rectifier is the most commonly used type in dc power supplies. In this section, you will use what you learned about half-wave rectification and expand it to full-wave rectifiers. You will learn about two types of full-wave rectifiers: center-tapped and bridge.

A **full-wave rectifier** allows unidirectional (one-way) current through the load during the entire 360^{o} of the input cycle, whereas a half-wave rectifier allows current through the load only during one-half of the cycle. The result of full-wave rectification is an output voltage with a frequency twice the input frequency and that pulsates every half-cycle of the input, as shown in Figure 2–29.

Figure 2–29: Full-wave rectification.

The number of positive alternations that make up the full-wave rectified voltage is twice that of the half-wave voltage for the same time interval. The average value, which is the value measured on a dc voltmeter, for a full-wave rectified sinusoidal voltage is twice that of the half-wave, as shown in the following formula:

$$V_{AVG} = \frac{2V_p}{\pi} \quad (2\text{-}6)$$

V_{AVG} is approximately 63.7% of V_p for a full-wave rectified voltage.

Example 2–5:

Find the average value of the full-wave rectified voltage in Figure 2–30.

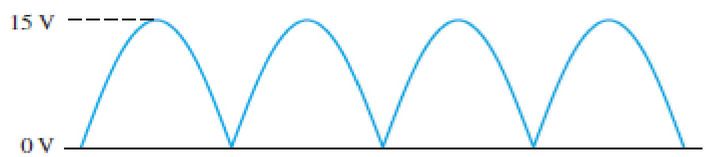


Figure 2–30

Solution

$$V_{AVG} = \frac{2V_p}{\pi} = \frac{2(15)}{\pi} = 9.55\ V$$

V_{AVG} is 63.7% of Vp.

***Related Problem*:**

Find the average value of the full-wave rectified voltage if its peak is 155 V.

Center-Tapped Full-Wave Rectifier Operation

A **center-tapped rectifier** is a type of full-wave rectifier that uses two diodes connected to the secondary of a center-tapped transformer, as shown in Figure 2–31. The input voltage is coupled through the transformer to the center-tapped secondary. Half of the total secondary voltage appears between the center tap and each end of the secondary winding as shown.

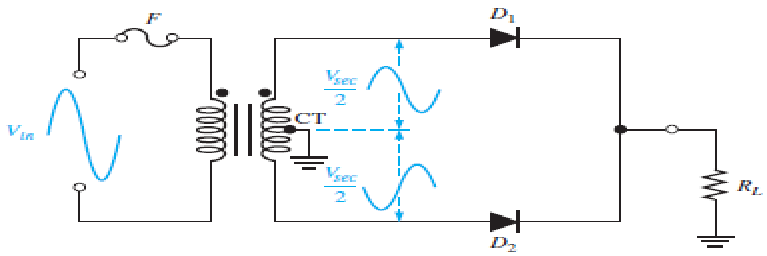


Figure 2–31: A center-tapped full-wave rectifier.

For a positive half-cycle of the input voltage, the polarities of the secondary voltages are as shown in Figure 2–32(a). This condition forward-biases diode D_1 and reverse-biases diode D_2. The current path is through D_1 and the load resistor R_L, as indicated. For a negative half-cycle of the input voltage, the voltage polarities on the secondary are as shown in Figure 2–32(b). This condition reverse-biases D_1 and forward-biases D_2. The current path is through D_2 and R_L, as indicated. Because the output current during both the positive and negative portions of the input cycle is in the same direction through the load, the output voltage developed across the load resistor is a full-wave rectified dc voltage, as shown.

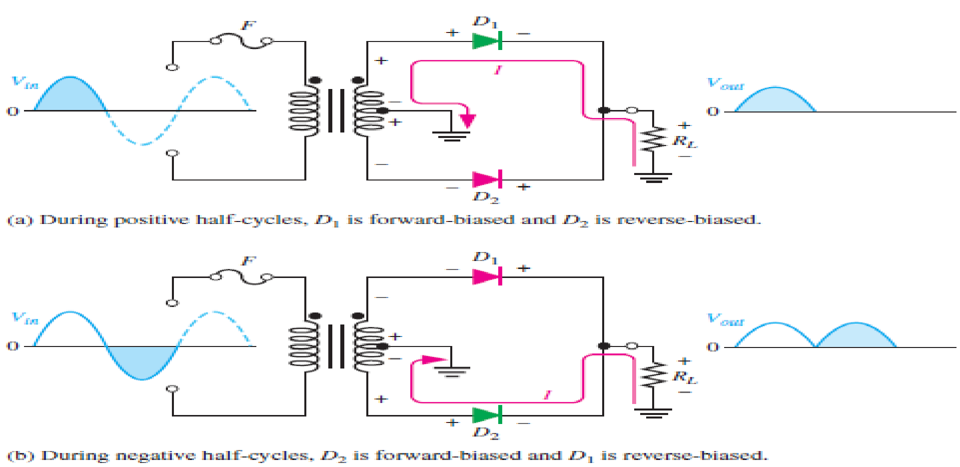


Figure 2–32: Basic operation of a center-tapped full-wave rectifier. Note that the current through the load resistor is in the same direction during the entire input cycle, so the output voltage always has the same polarity.

Effect of the Turns Ratio on the Output Voltage

If the transformer's turns ratio is 1, the peak value of the rectified output voltage equals half the peak value of the primary input voltage less the barrier potential, as illustrated in Figure 2–33. Half of the primary voltage appears across each half of the secondary winding ($V_{p(sec)} = V_{p(pri)}$). We will begin referring to the forward voltage due to the barrier potential as the **diode drop.**

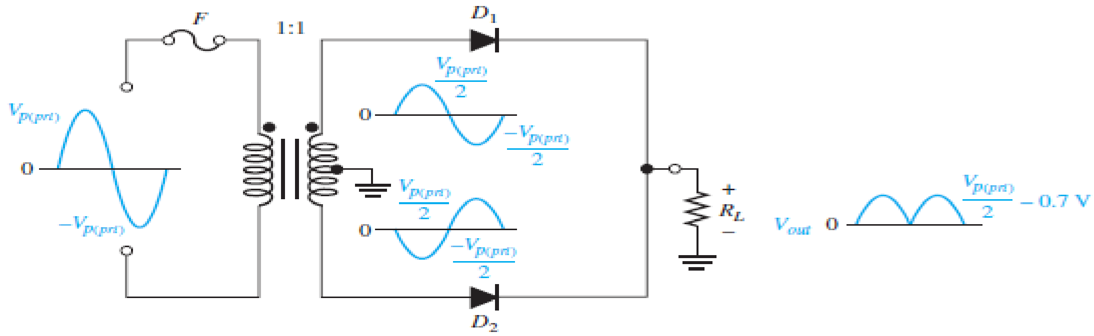


Figure 2–33: Center-tapped full-wave rectifier with a transformer turns ratio of 1. $V_{p(pri)}$ is the peak value of the primary voltage.

In order to obtain an output voltage with a peak equal to the input peak (less the diode drop), a step-up transformer with a turns ratio of $n = 2$ must be used, as shown in Figure 2–34. In this case, the total secondary voltage (V_{sec}) is twice the primary voltage ($2V_{pri}$), so the voltage across each half of the secondary is equal to V_{pri}.

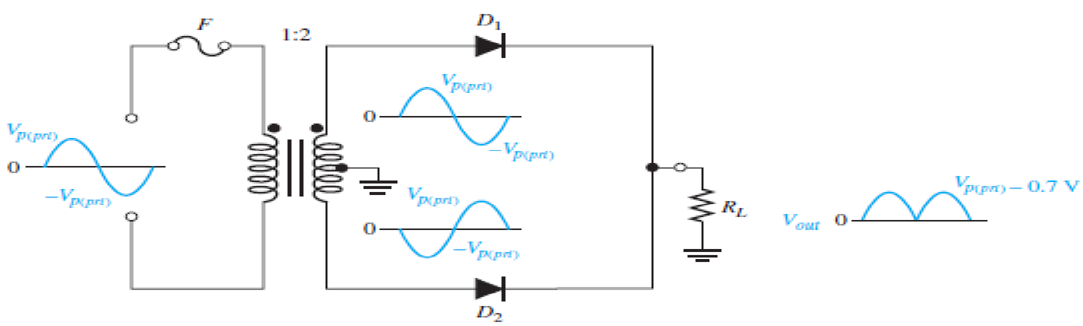


Figure 2–34: Cent er-tapped full-wave rectifier with a transformer turns ratio of 2.

In any case, the output voltage of a center-tapped full-wave rectifier is always one-half of the total secondary voltage less the diode drop, no matter what the turns ratio.

$$V_{out} = \frac{V_{sec}}{2} - 0.7 \qquad (2\text{-}7)$$

***<u>Peak Inverse Voltage</u>*:** Each diode in the full-wave rectifier is alternately forward-biased and then reverse-biased. The maximum reverse voltage that each diode must withstand is the peak secondary voltage $V_{p(sec)}$. This is shown in Figure 2–35 where D_2 is assumed to be reverse-biased (red) and D_1 is assumed to be forward-biased (green) to illustrate the concept.

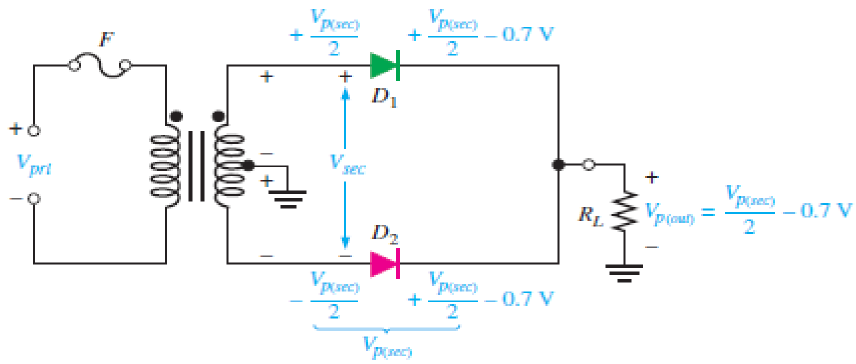


Figure 2–35: Diode reverse voltage (D_2 shown reverse-biased and D_1 shown forward-biased).

When the total secondary voltage V_{sec} has the polarity shown, the maximum anode voltage of D_1 is $+V_{p(sec)}/2$ and the maximum anode voltage of D_2 is $-V_{p(sec)}/2$. Since D_1 is assumed to be forward-biased, its cathode is at the same voltage as its anode minus the diode drop; this is also the voltage on the cathode of D_2.

The peak inverse voltage across D_2 is

$$PIV = \left(\frac{(V_{P(\text{sec})})}{2} - 0.7\right) - \left(-\frac{(V_{P(\text{sec})})}{2}\right) = \frac{(V_{P(\text{sec})})}{2} + \frac{(V_{P(\text{sec})})}{2} - 0.7 = (V_{P(\text{sec})}) - 0.7$$

Since $V_{p(out)} = (V_{p(sec)}/2) - 0.7$ V, then by multiplying each term by 2 and transposing,

$$V_{p(sec)} = 2V_{p(out)} + 1.4\text{ V}$$

Therefore, by substitution, the peak inverse voltage across either diode in a full-wave centertapped rectifier is

$$\text{PIV} = 2V_{p(out)} - 0.7\text{ V} \qquad (2\text{-}8)$$

Example 2–6

(a) Show the voltage waveforms across each half of the secondary winding and across R_L when a 100 V peak sine wave is applied to the primary winding in Figure 2–36.

(b) What minimum PIV rating must the diodes have?

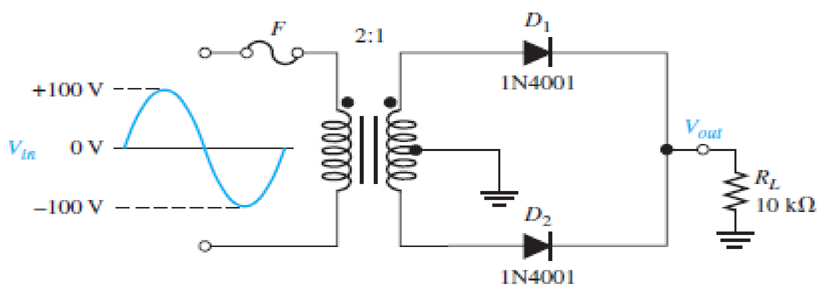


Figure 2–36

Solution

(a) The transformer turns ratio n = 0.5. The total peak secondary voltage is

$$V_{p(sec)} = nV_{p(pri)} = 0.5(100\text{ V}) = 50\text{ V}$$

There is a 25 V peak across each half of the secondary with respect to ground. The output load voltage has a peak value of 25 V, less the 0.7 V drop across the diode. The waveforms are shown in Figure 2–37.

(b) Each diode must have a minimum PIV rating of

$$\text{PIV} = 2V_{p(out)} + 0.7\text{ V} = 2(24.3\text{ V}) + 0.7\text{ V} = \mathbf{49.3\ V}$$

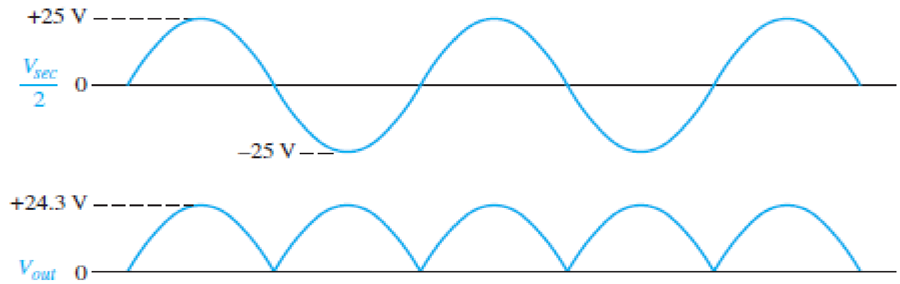


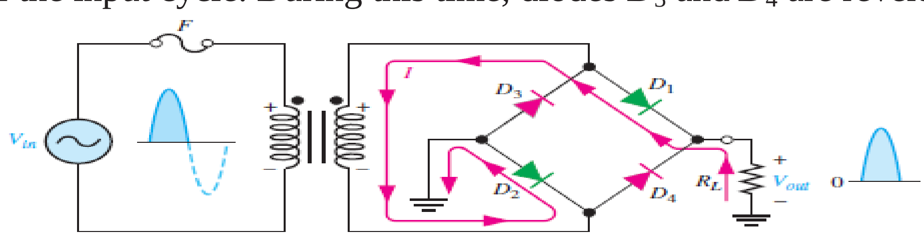
Figure 2–37

Related Problem

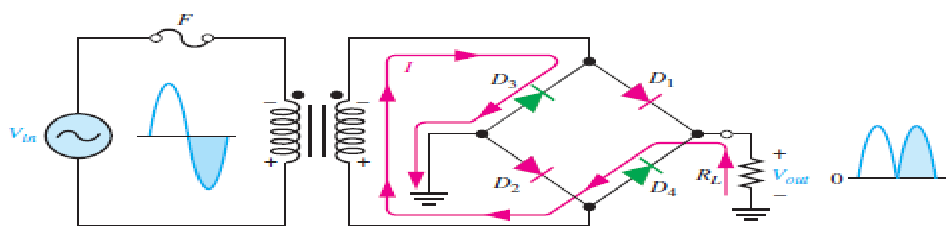
What diode PIV rating is required to handle a peak input of 160 V in Figure 2–36?

Bridge Full-Wave Rectifier Operation:

The **bridge rectifier** uses four diodes connected as shown in Figure 2–38. When the input cycle is positive as in part (a), diodes D_1 and D_2 are forward-biased and conduct current in the direction shown. A voltage is developed across R_L that looks like the positive half of the input cycle. During this time, diodes D_3 and D_4 are reverse-biased.



(a) During the positive half-cycle of the input, D_1 and D_2 are forward-biased and conduct current. D_3 and D_4 are reverse-biased.



(b) During the negative half-cycle of the input, D_3 and D_4 are forward-biased and conduct current. D_1 and D_2 are reverse-biased.

Figure 2–38: Operation of a bridge rectifier.

When the input cycle is negative as in Figure 2–38(b), diodes D_3 and D_4 are forward-biased and conduct current in the same direction through R_L as during the positive half-cycle. During the negative half-cycle, D_1 and D_2 are reverse-biased. A full-wave rectified output voltage appears across R_L as a result of this action.

***Bridge Output Voltage*:**

A bridge rectifier with a transformer-coupled input is shown in Figure 2–39(a). During the positive half-cycle of the total secondary voltage, diodes D_1 and D_2 are forward-biased. Neglecting the diode drops, the secondary voltage appears across the load resistor. The same is true when D_3 and D_4 are forward-biased during the negative half-cycle.

$$V_{p(out)} = V_{p(sec)}$$

As you can see in Figure 2–39(b), two diodes are always in series with the load resistor during both the positive and negative half-cycles. If these diode drops are taken into account, the output voltage is

$$V_{p(out)} = V_{p(sec)} - 1.4\text{ V} \qquad (2\text{-}9)$$

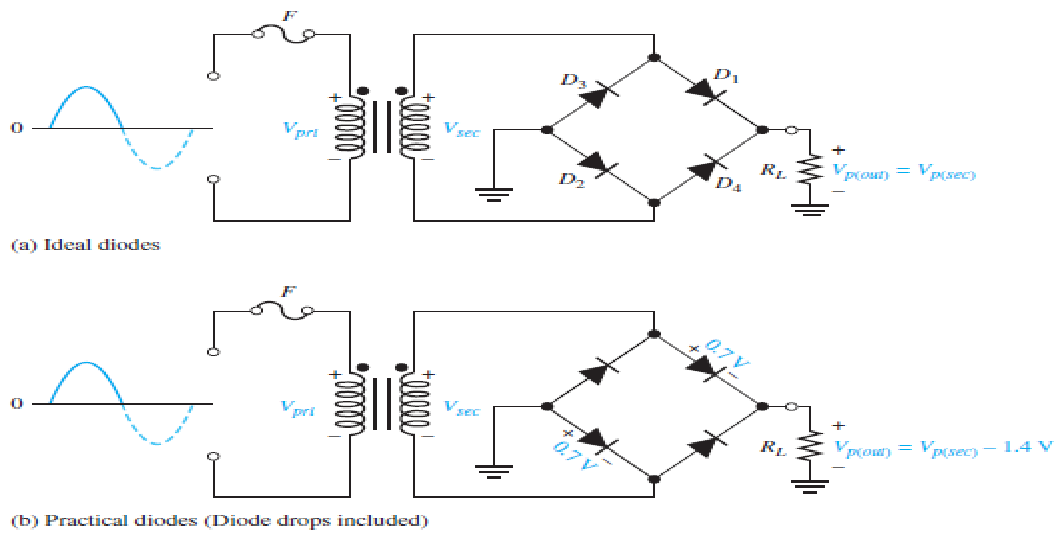


Figure 2–39: Bridge operation during a positive half-cycle of the primary and secondary voltages.

***Peak Inverse Voltage*:**

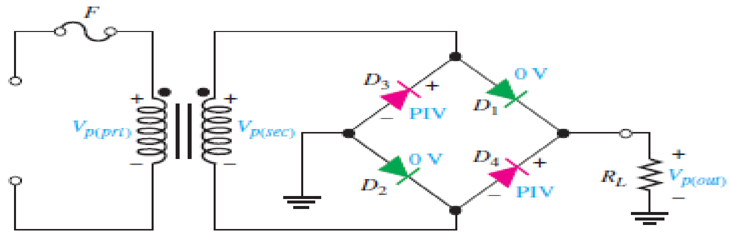
Let's assume that D_1 and D_2 are forward-biased and examine the reverse voltage across D_3 and D_4. Visualizing D_1 and D_2 as shorts (ideal model), as in Figure 2–40(a), you can see that D_3 and D_4 have a peak inverse voltage equal to the peak secondary voltage. Since the output voltage is *ideally* equal to the secondary voltage,

$$\text{PIV} = V_{p(out)}$$

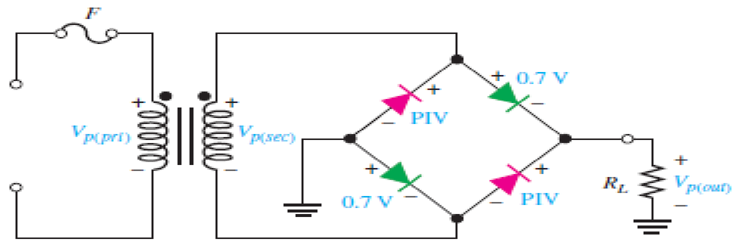
If the diode drops of the forward-biased diodes are included as shown in Figure 2–40(b), the peak inverse voltage across each reverse-biased diode in terms of $V_{p(out)}$ is

$$\text{PIV} = Vp(out) + 0.7\text{ V} \qquad (2\text{-}10)$$

The PIV rating of the bridge diodes is less than that required for the center-tapped configuration. If the diode drop is neglected, the bridge rectifier requires diodes with half the PIV rating of those in a center-tapped rectifier for the same output voltage.



(a) For the ideal diode model (forward-biased diodes D_1 and D_2 are shown in green). PIV = $V_{p(out)}$.



(b) For the practical diode model (forward-biased diodes D_1 and D_2 are shown in green). PIV = $V_{p(out)}$ + 0.7 V.

Figure 2–40: Peak inverse voltages across diodes D_3 and D_4 in a bridge rectifier during the positive half-cycle of the secondary voltage.

Example 2–7

Determine the peak output voltage for the bridge rectifier in Figure 2–41. Assuming the practical model, what PIV rating is required for the diodes? The transformer is specified to have a 12 V rms secondary voltage for the standard 120 V across the primary.

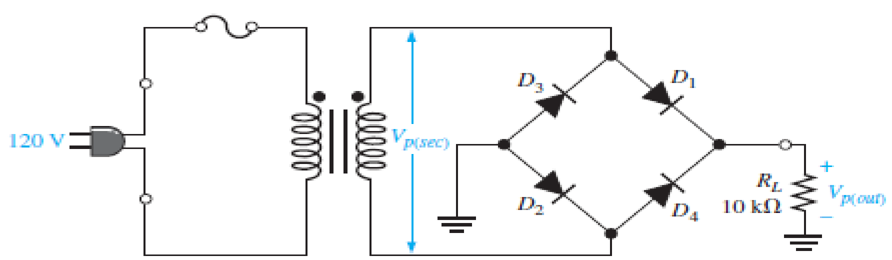


Figure 2–41

Solution

The peak output voltage (taking into account the two diode drops) is

$$V_{p(sec)} = 1.414\ V_{rms} = 1.414(12\text{ V}) = 17\text{ V}$$

$$V_{p(out)} = V_{p(sec)} - 1.4\text{ V} = 17\text{ V} - 1.4\text{ V} = 15.6\text{ V}$$

The PIV rating for each diode is

$$\text{PIV} = V_{p(out)} + 0.7\text{ V} = 15.6\text{ V} + 0.7\text{ V} = 16.3\text{ V}$$

***Related Problem*:**

Determine the peak output voltage for the bridge rectifier in Figure 2–41 if the transformer produces an rms secondary voltage of 30 V. What is the PIV rating for the diodes?

SECTION 2–5 CHECKUP

1. How does a full-wave voltage differ from a half-wave voltage?
2. What is the average value of a full-wave rectified voltage with a peak value of 60 V?
3. Which type of full-wave rectifier has the greater output voltage for the same input voltage and transformer turns ratio?
4. For a peak output voltage of 45 V, in which type of rectifier would you use diodes with a PIV rating of 50 V?
5. What PIV rating is required for diodes used in the type of rectifier that was not selected in Question 4?

2–6 Power supply filters and regulators

A power supply filter ideally eliminates the fluctuations in the output voltage of a halfwave or full-wave rectifier and produces a constant-level dc voltage. Filtering is necessary because electronic circuits require a constant source of dc voltage and current to provide power and biasing for proper operation. Filters are implemented with capacitors, as you will see in this section. Voltage regulation in power supplies is usually done with integrated circuit voltage regulators. A voltage regulator prevents changes in the filtered dc voltage due to variations in input voltage or load.

In most power supply applications, the standard 60 Hz ac power line voltage must be converted to an approximately constant dc voltage. The 60 Hz pulsating dc output of a half-wave rectifier or the 120 Hz pulsating output of a full-wave rectifier must be filtered to reduce the large voltage variations. Figure 2–42 illustrates the filtering concept showing a nearly smooth dc output voltage from the filter. The small amount of fluctuation in the filter output voltage is called *ripple.*

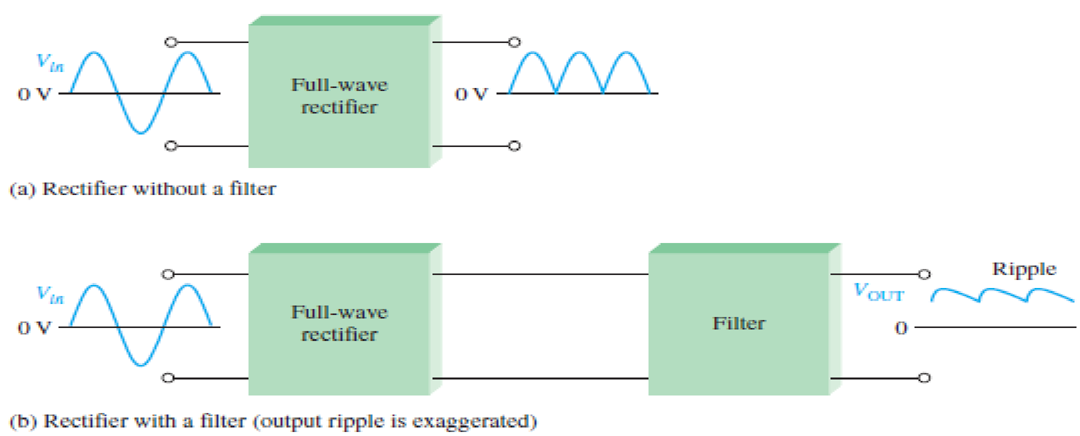
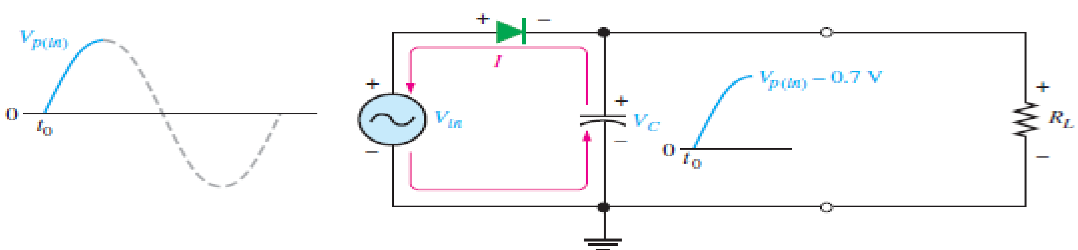


Figure 2–41: Power supply filtering.

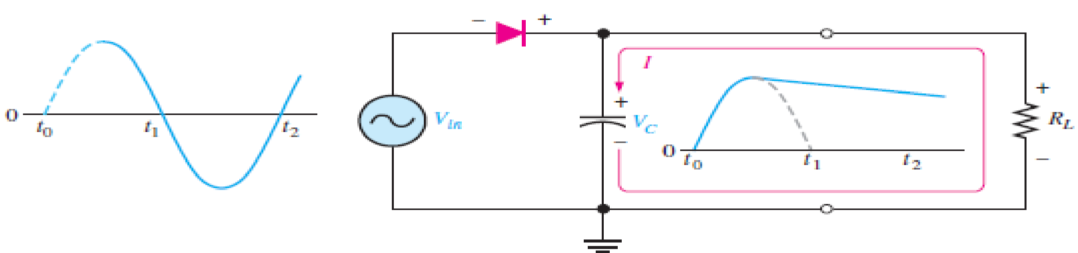
Capacitor-Input Filter

A half-wave rectifier with a capacitor-input filter is shown in Figure 2–43. The filter is simply a capacitor connected from the rectifier output to ground. *RL* represents the equivalent resistance of a load. We will use the half-wave rectifier to illustrate the basic principle and then expand the concept to full-wave rectification.

During the positive first quarter-cycle of the input, the diode is forward-biased, allowing the capacitor to charge to within 0.7 V of the input peak, as illustrated in Figure 2–43(a). When the input begins to decrease below its peak, as shown in part (b), the capacitor retains its charge and the diode becomes reverse-biased because the cathode is more positive than the anode. During the remaining part of the cycle, the capacitor can discharge only through the load resistance at a rate determined by the R_LC time constant, which is normally long compared to the period of the input. The larger the time constant, the less the capacitor will discharge. During the first quarter of the next cycle, as illustrated in part (c), the diode will again become forward-biased when the input voltage exceeds the capacitor voltage by approximately 0.7 V.



(a) Initial charging of the capacitor (diode is forward-biased) happens only once when power is turned on.



(b) The capacitor discharges through R_L after peak of positive alternation when the diode is reverse-biased. This discharging occurs during the portion of the input voltage indicated by the solid dark blue curve.

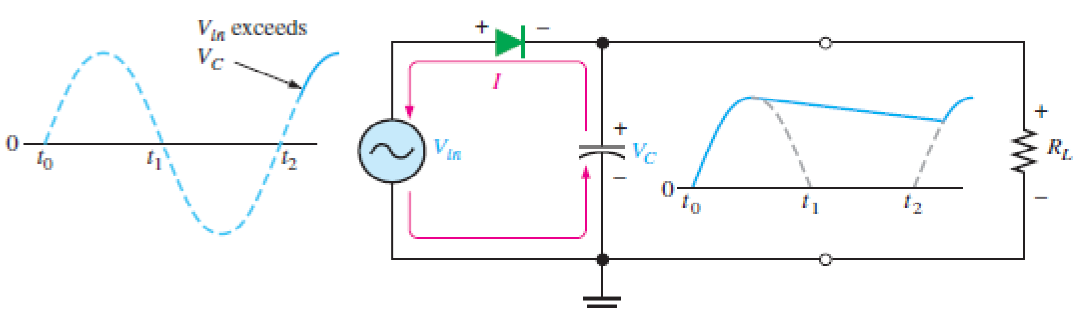


Figure 2–43: Operation of a half-wave rectifier with a capacitor-input filter. The current indicates charging or discharging of the capacitor.

***Ripple Voltage*:**

As you have seen, the capacitor quickly charges at the beginning of a cycle and slowly discharges through R_L after the positive peak of the input voltage (when the diode is reverse-biased). The variation in the capacitor voltage due to the charging and discharging is called the **ripple voltage**. Generally, ripple is undesirable; thus, the smaller the ripple, the better the filtering action, as illustrated in Figure 2–44.

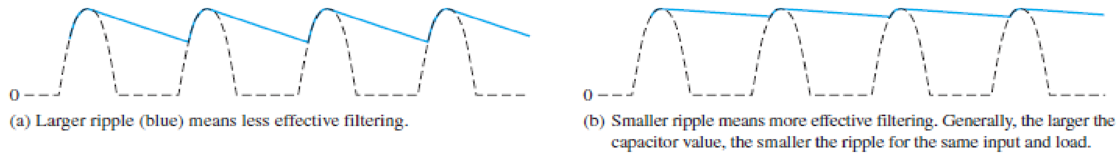


Figure 2–44: Half-wave ripple voltage (blue line).

For a given input frequency, the output frequency of a full-wave rectifier is twice that of a half-wave rectifier, as illustrated in Figure 2–45. This makes a full-wave rectifier easier to filter because of the shorter time between peaks. When filtered, the full-wave rectified voltage has a smaller ripple than does a half-wave voltage for the same load resistance and capacitor values. The capacitor discharges less during the shorter interval between fullwave pulses, as shown in Figure 2–46.

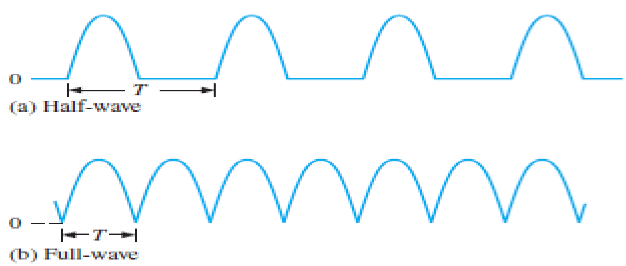


Figure 2–45: The period of a full-wave rectified voltage is half that of a half-wave rectified voltage. The output frequency of a full-wave rectifier is twice that of a half-wave rectifier.

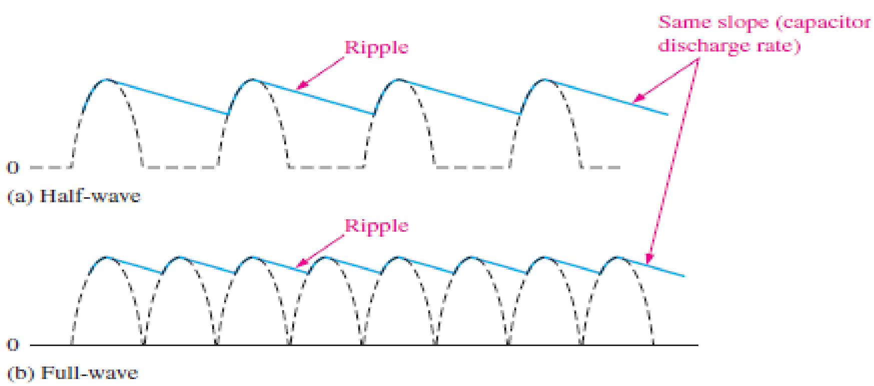


Figure 2–46: Comparison of ripple voltages for half-wave and full-wave rectified voltages with the same filter capacitor and load and derived from the same sinusoidal input voltage.

Ripple Factor

The **ripple factor (r)** is an indication of the effectiveness of the filter and is defined as

$$r = \frac{V_{r(pp)}}{V_{DC}} \qquad (2\text{-}11)$$

where $V_{r(pp)}$ is the peak-to-peak ripple voltage and V_{DC} is the dc (average) value of the filter's output voltage, as illustrated in Figure 2–47. The lower the ripple factor, the better the filter. The ripple factor can be lowered by increasing the value of the filter capacitor or increasing the load resistance.

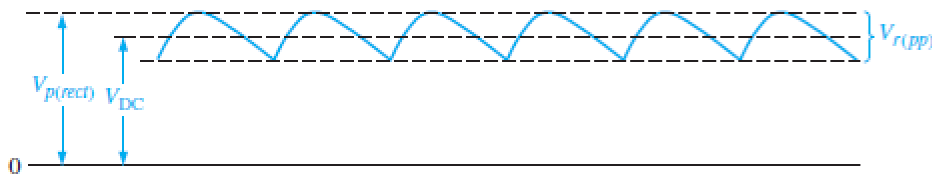


Figure 2–47: V_r and V_{DC} determine the ripple factor.

For a full-wave rectifier with a capacitor-input filter, approximations for the peak-topeak ripple voltage, $V_{r(pp)}$, and the dc value of the filter output voltage, V_{DC}, are given in the following equations. The variable $V_{p(rect)}$ is the unfiltered peak rectified voltage. Notice that if R_L or C increases, the ripple voltage decreases and the dc voltage increases.

$$V_{r(pp)} = \left(\frac{1}{f\,R_L C}\right) V_{p(rect)} \qquad (2\text{-}12)$$

$$V_{DC} = \left(1 - \frac{1}{2\,f\,R_L C}\right) V_{p(rect)} \qquad (2\text{-}13)$$

Example 2–8:

Determine the ripple factor for the filtered bridge rectifier with a load as indicated in Figure 2–48.

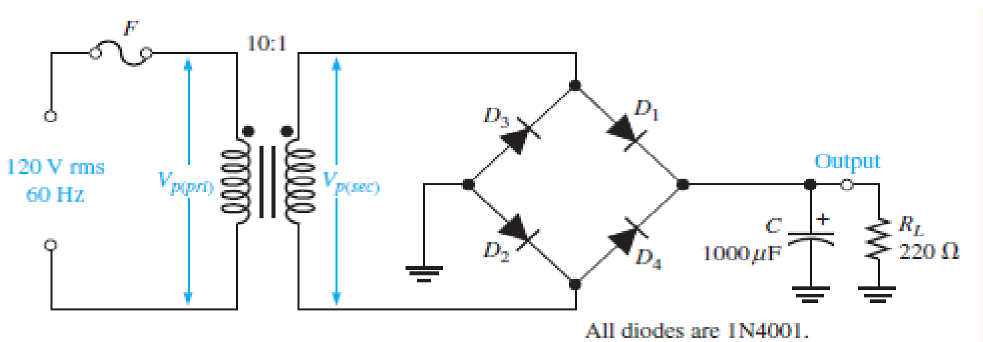


Figure 2–48

Solution

The transformer turns ratio is $n = 0.1$. The peak primary voltage is

$$V_{p(pri)} = 1.414\ V_{rms} = 1.414(120\text{ V}) = 170\text{ V}$$

The peak secondary voltage is

$$V_{p(sec)} = nV_{p(pri)} = 0.1(170\text{ V}) = 17.0\text{ V}$$

The unfiltered peak full-wave rectified voltage is

$$V_{p(rect)} = V_{p(sec)} - 1.4\text{ V} = 17.0\text{ V} - 1.4\text{ V} = 15.6\text{ V}$$

The frequency of a full-wave rectified voltage is 120 Hz. The approximate peak-topeak ripple voltage at the output is

$$V_{r(pp)} = \left(\frac{1}{f\,R_L C}\right) V_{p(rect)} = \left(\frac{1}{(120\ Hz)(220\Omega)(1000\ \mu f)}\right) 15.6\ V = 0.591\ V$$

The approximate dc value of the output voltage is determined as follows:

$$V_{DC} = \left(1 - \frac{1}{2f\,R_L C}\right) V_{p(rect)} = \left(1 - \frac{1}{(240\ Hz)(220\Omega)(1000\ \mu f)}\right) 15.6\ V = 15.3\ V$$

The resulting ripple factor is

$$r = \frac{V_{r(pp)}}{V_{DC}} = \frac{0.591\ V}{15.3\ V} = 0.039$$

The percent ripple is 3.9%.

Related Problem:

Determine the peak-to-peak ripple voltage if the filter capacitor in Figure 2–48 is increased to 2200 μF and the load resistance changes to 2.2 kΩ.

Surge Current in the Capacitor-Input Filter

Before the switch in Figure 2–49 is closed, the filter capacitor is uncharged. At the instant the switch is closed, voltage is connected to the bridge and the uncharged capacitor appears as a short, as shown. This produces an initial surge of current, I_{surge}, through the two forward-biased diodes D_1 and D_2. The worst-case situation occurs when the switch is closed at a peak of the secondary voltage and a maximum surge current, $I_{surge(max)}$, is produced, as illustrated in the figure.

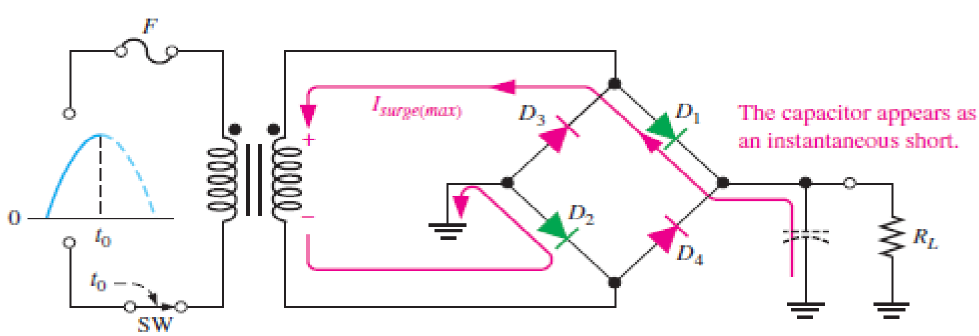


Figure 2–49: Surge current in a capacitor-input filter.

In dc power supplies, a **fuse** is always placed in the primary circuit of the transformer, as shown in Figure 2–49. A slow-blow type fuse is generally used because of the surge current that initially occurs when power is first turned on. The fuse rating is determined by calculating the power in the power supply load, which is the output power. Since P_{in} = P_{out} in an ideal transformer, the primary current can be calculated as

$$I_{pri} = \frac{P_{in}}{120\ V}$$

The fuse rating should be at least 20% larger than the calculated value of I_{pri}.

<u>Voltage Regulators:</u>

While filters can reduce the ripple from power supplies to a low value, the most effective approach is a combination of a capacitor-input filter used with a voltage regulator. A voltage regulator is connected to the output of a filtered rectifier and maintains a constant output voltage (or current) despite changes in the input, the load current, or the temperature. The capacitor- input filter reduces the input ripple to the regulator to an acceptable level. The combination of a large capacitor and a voltage regulator helps produce an excellent power supply. Most regulators are integrated circuits and have three terminals—an input terminal, an output terminal, and a reference (or adjust) terminal. The input to the regulator is first filtered with a capacitor to reduce the ripple to <10%. The regulator reduces the ripple to a negligible amount. In addition, most regulators have an internal voltage reference, shortcircuit protection, and thermal shutdown circuitry. They are available in a variety of voltages, including positive and negative outputs, and can be designed for variable outputs with a minimum of external components. Typically, voltage regulators can furnish a constant output of one or more amps of current with high ripple rejection. Three-terminal regulators designed for fixed output voltages require only external capacitors to complete the regulation portion of the power supply, as shown in Figure 2–50. Filtering is accomplished by a large-value capacitor between the input voltage and ground. An output capacitor (typically 0.1 μF to 1.0 μF) is connected from the output to ground to improve the transient response.

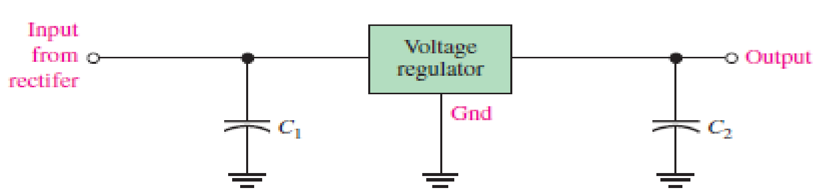


Figure 2–50: A voltage regulator with input and output capacitors.

A basic fixed power supply with a +5 V voltage regulator is shown in Figure 2–51.

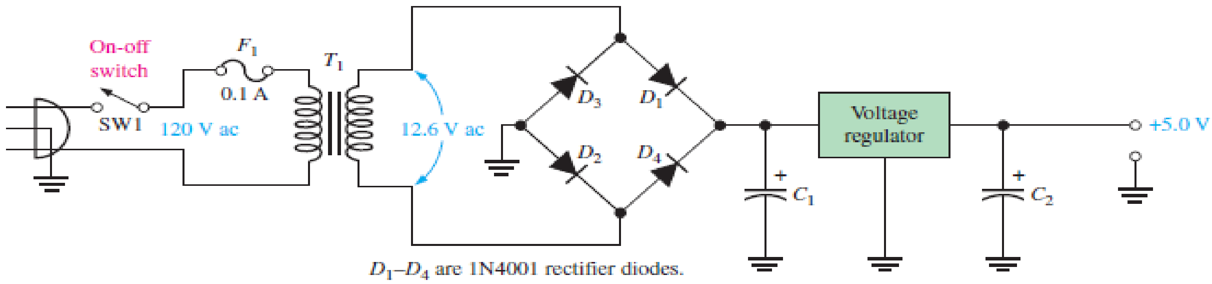


Figure 2–51: A basic +5.0 V regulated power supply.

Percent Regulation

The regulation expressed as a percentage is a figure of merit used to specify the performance of a voltage regulator. It can be in terms of input (line) regulation or load regulation.

Line Regulation

The **line regulation** specifies how much change occurs in the output voltage for a given change in the input voltage. It is typically defined as a ratio of a change in output voltage for a corresponding change in the input voltage expressed as a percentage.

$$\text{line regulation} = \left(\frac{\Delta V_{OUT}}{\Delta V_{IN}}\right)100\% \qquad (2\text{-}14)$$

Load Regulation:

The **load regulation** specifies how much change occurs in the output voltage over a certain range of load current values, usually from minimum current (no load, NL) to maximum current (full load, FL). It is normally expressed as a percentage and can be calculated with the following formula:

$$\text{load regulation} = \left(\frac{V_L - V_{FL}}{V_{FL}}\right)100\% \qquad (2\text{-}15)$$

where V_{NL} is the output voltage with no load and V_{FL} is the output voltage with full (maximum) load.

Example 2–9:

A certain 7805 regulator has a measured no-load output voltage of 5.18 V and a full load output of 5.15 V. What is the load regulation expressed as a percentage?

Solution:

$$\text{load regulation} = \left(\frac{V_L - V_{FL}}{V_{FL}}\right)100\% = \left(\frac{5.18V - 5.15V}{5.15V}\right)100\% = 0.58\%$$

***Related Problem*:**

If the no-load output voltage of a regulator is 24.8 V and the full-load output is 23.9 V, what is the load regulation expressed as a percentage?

<u>SECTION 2–6 CHECKUP</u>

1. When a 60 Hz sinusoidal voltage is applied to the input of a half-wave rectifier, what is the output frequency?
2. When a 60 Hz sinusoidal voltage is applied to the input of a full-wave rectifier, what is the output frequency?
3. What causes the ripple voltage on the output of a capacitor-input filter?
4. If the load resistance connected to a filtered power supply is decreased, what happens to the ripple voltage?
5. Define *ripple factor*.
6. What is the difference between input (line) regulation and load regulation?

2–7 Diode limiters and clampers

Diode circuits, called limiters or clippers, are sometimes used to clip off portions of signal voltages above or below certain levels. Another type of diode circuit, called a clamper, is used to add or restore a dc level to an electrical signal. Both limiter and clamper diode circuits will be examined in this section.

Diode Limiters

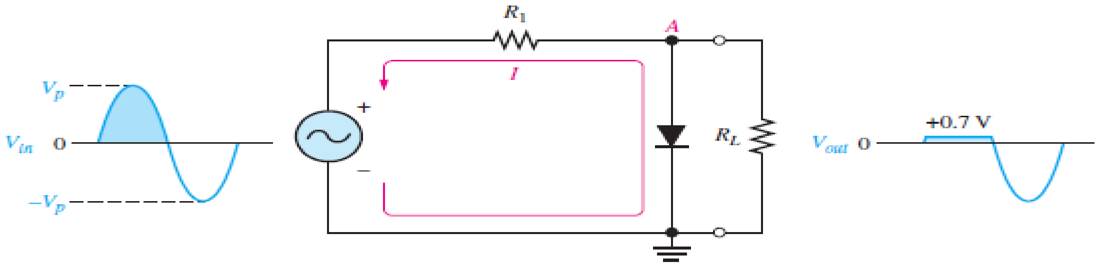
Figure 2–52(a) shows a diode positive **limiter** (also called **clipper**) that limits or clips the positive part of the input voltage. As the input voltage goes positive, the diode becomes forward-biased and conducts current. Point *A* is limited to +0.7 V when the input voltage exceeds this value. When the input voltage goes back below 0.7 V, the diode is reverse-biased and appears as an open. The output voltage looks like the negative part of the input voltage, but with a magnitude determined by the voltage divider formed by R_1 and the load resistor, R_L, as follows:

$$V_{out} = \left(\frac{R_L}{R_1 + R_L} \right) V_{in}$$

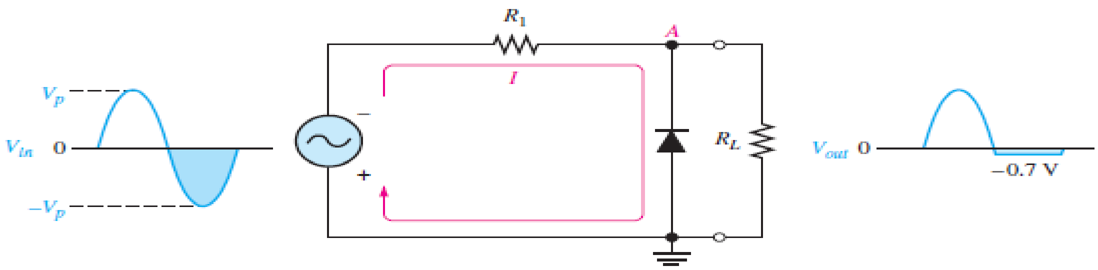
If R_1 is small compared to R_L, then If the diode is turned around, as in Figure 2–52(b), the negative part of the input voltage is clipped off. When the diode is forward-biased during the negative part of the input voltage, point *A* is held at -0.7 V by the diode drop. When the input voltage goes above -0.7 V the diode is no longer forward-biased; and a voltage appears across R_L proportional to the input voltage.

<u>Example 2–10:</u>

What would you expect to see displayed on an oscilloscope connected across R_L in the limiter shown in Figure 2–53?



(a) Limiting of the positive alternation. The diode is forward-biased during the positive alternation (above 0.7 V) and reverse-biased during the negative alternation.



(b) Limiting of the negative alternation. The diode is forward-biased during the negative alternation (below −0.7 V) and reverse-biased during the positive alternation.

Figure 2–52: Examples of diode limiters (clippers).

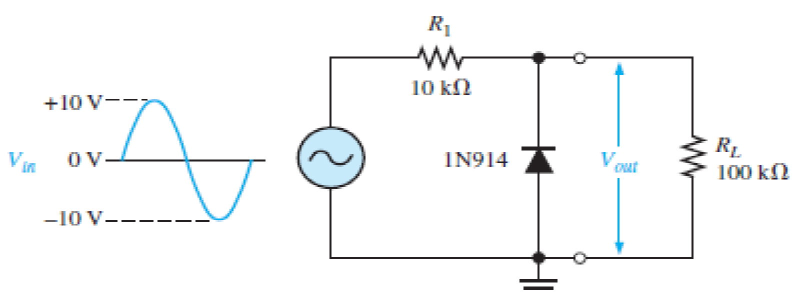


Figure 2–53

***Solution*:**

The diode is forward-biased and conducts when the input voltage goes below -0.7

So, for the negative limiter, determine the peak output voltage across R_L by the following equation:

$$V_{p(out)} = \left(\frac{R_L}{R_1 + R_L}\right) V_{p(in)} = \left(\frac{100\ k\Omega}{110\ k\Omega}\right) 10V = 9.09\ V$$

The scope will display an output waveform as shown in Figure 2–54.

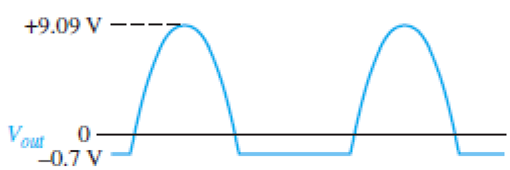


Figure 2–54: Output voltage waveform for Figure 2–53.

Related Problem:
Describe the output waveform for Figure 2–53 if R_1 is changed to 1 kΩ.

Biased Limiters: The level to which an ac voltage is limited can be adjusted by adding a bias voltage, V_{BIAS}, in series with the diode, as shown in Figure 2–55. The voltage at point *A* must equal V_{BIAS} + 0.7 V before the diode will become forward-biased and conduct. Once the diode begins to conduct, the voltage at point *A* is limited to *V*BIAS + 0.7 V so that all input voltage above this level is clipped off.

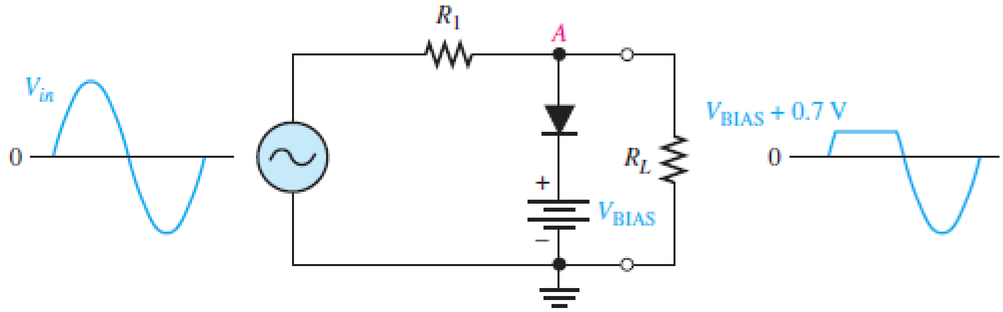


Figure 2–55: A positive limiter.

To limit a voltage to a specified negative level, the diode and bias voltage must be connected as in Figure 2–56. In this case, the voltage at point *A* must go below -V_{BIAS} - 0.7 V to forward-bias the diode and initiate limiting action as shown.

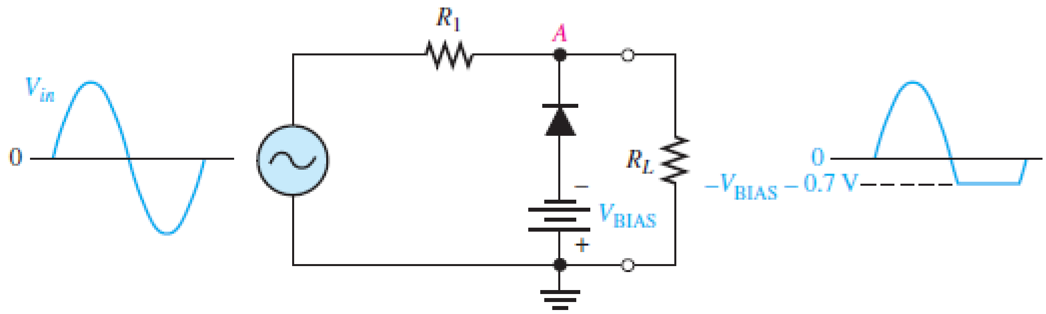


Figure 2–56: A negative limiter.

By turning the diode around, the positive limiter can be modified to limit the output voltage to the portion of the input voltage waveform above V_{BIAS} - 0.7 V, as shown by the output waveform in Figure 2–57(a). Similarly, the negative limiter can be modified to limit the output voltage to the portion of the input voltage waveform below -V_{BIAS} + 0.7 V, as shown by the output waveform in part (b).

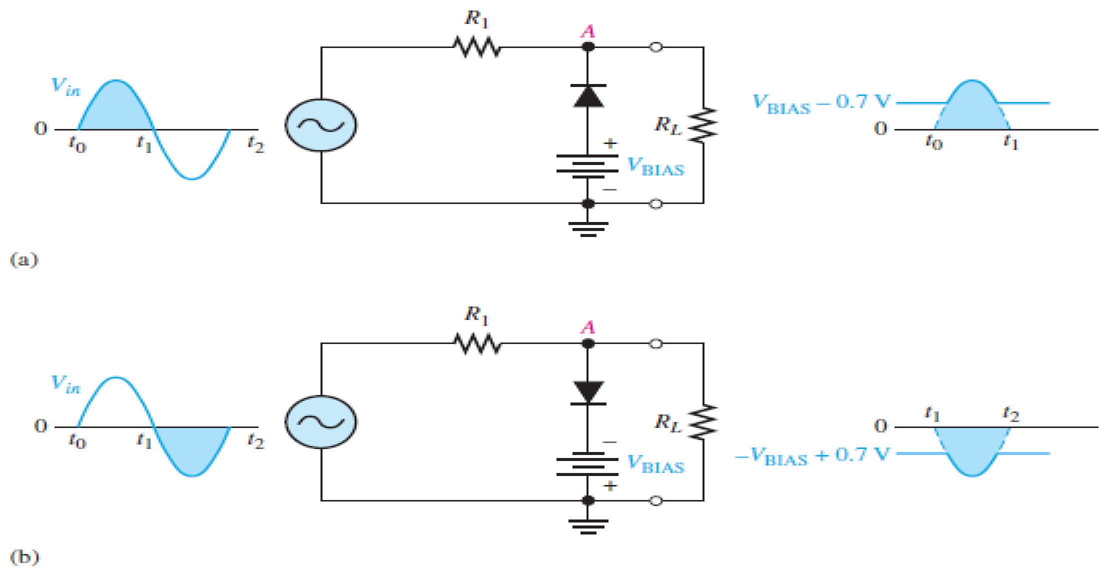


Figure 2–57

Example 2–11:

Figure 2–58 shows a circuit combining a positive limiter with a negative limiter. Determine the output voltage waveform.

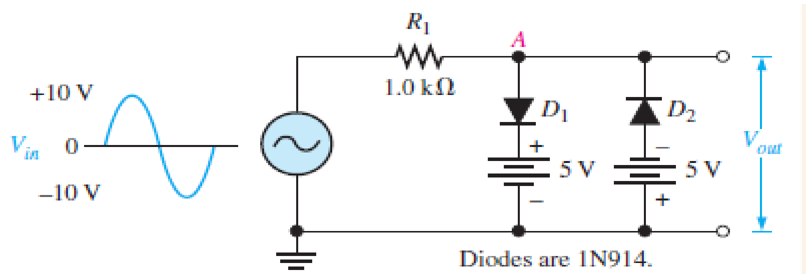


Figure 2–58

***Solution*:**

When the voltage at point A reaches +5.7 V, diode D_1 conducts and limits the waveform to +5.7 V. Diode D_2 does not conduct until the voltage reaches -5.7 V. Therefore, positive voltages above +5.7 V and negative voltages below -5.7 V are clipped off. The resulting output voltage waveform is shown in Figure 2–59.

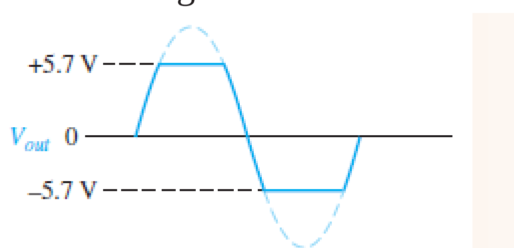


Figure 2–59: Output voltage waveform for Figure 2–58.

***Related Problem*:**

Determine the output voltage waveform in Figure 2–58 if both dc sources are 10 V and the input voltage has a peak value of 20 V.

Voltage-Divider Bias**:** The bias voltage sources that have been used to illustrate the basic operation of diode limiters can be replaced by a resistive voltage divider that derives the desired bias voltage from the dc supply voltage, as shown in Figure 2–60. The bias voltage is set by the resistor values according to the voltage-divider formula.

$$V_{BIAS} = \left(\frac{R_3}{R_2 + R_3}\right) V_{SUPPLY}$$

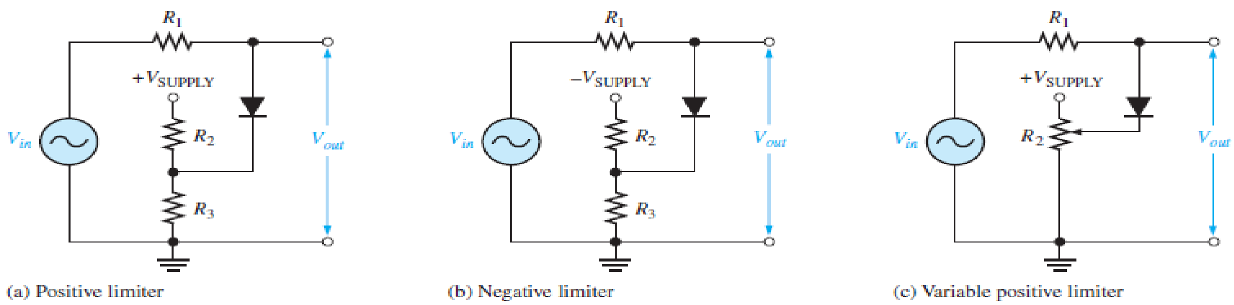


Figure 2–60: Diode limiters implemented with voltage-divider bias.

A positively biased limiter is shown in Figure 2–60(a), a negatively biased limiter is shown in part (b), and a variable positive bias circuit using a potentiometer voltage divider is shown in part (c). The bias resistors must be small compared to R_1 so that the forward current through the diode will not affect the bias voltage.

A Limiter Application**:** Many circuits have certain restrictions on the input level to avoid damaging the circuit. For example, almost all digital circuits should not have an input level that exceeds the power supply voltage. An input of a few volts more than this could damage the circuit. To prevent the input from exceeding a specific level, you may see a diode limiter across the input signal path in many digital circuits.

Example 2–12:

Describe the output voltage waveform for the diode limiter in Figure 2–61.

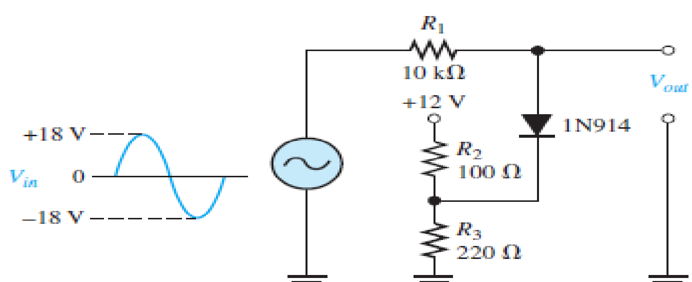


Figure 2–61

Solution:

The circuit is a positive limiter. Use the voltage-divider formula to determine the bias voltage.

$$V_{BIAS} = \left(\frac{R_3}{R_2 + R_3}\right) V_{SUPPLY} = \left(\frac{220\ \Omega}{100\ \Omega + 220\ \Omega}\right) 12\ V = 8.25\ V$$

The output voltage waveform is shown in Figure 2–62. The positive part of the output voltage waveform is limited to V_{BIAS} + 0.7 V.

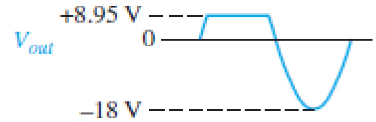


Figure 2–62

Related Problem:
How would you change the voltage divider in Figure 2–61 to limit the output voltage to +6.7 V?

Diode Clampers:

A clamper adds a dc level to an ac voltage. **Clampers** are sometimes known as *dc restorers*. Figure 2–63 shows a diode clamper that inserts a positive dc level in the output waveform. The operation of this circuit can be seen by considering the first negative half-cycle of the input voltage. When the input voltage initially goes negative, the diode is forward biased, allowing the capacitor to charge to near the peak of the input ($V_{p(in)}$ - 0.7 V), as shown in Figure 2–63(a). Just after the negative peak, the diode is reverse-biased. This is because the cathode is held near $V_{p(in)}$ - 0.7 V,by the charge on the capacitor. The capacitor can only discharge through the high resistance of R_L. So, from the peak of one negative half-cycle to the next, the capacitor discharges very little. The amount that is discharged, of course, depends on the value of R_L.

If the capacitor discharges during the period of the input wave, clamping action is affected. If the *RC* time constant is 100 times the period, the clamping action is excellent. An *RC* time constant of ten times the period will have a small amount of distortion at the ground level due to the charging current. The net effect of the clamping action is that the capacitor retains a charge approximately equal to the peak value of the input less the diode drop. The capacitor voltage acts essentially as a battery in series with the input voltage. The dc voltage of the capacitor adds
to the input voltage by superposition, as in Figure 2–63(b).

If the diode is turned around, a negative dc voltage is added to the input voltage to produce the output voltage as shown in Figure 2–64.

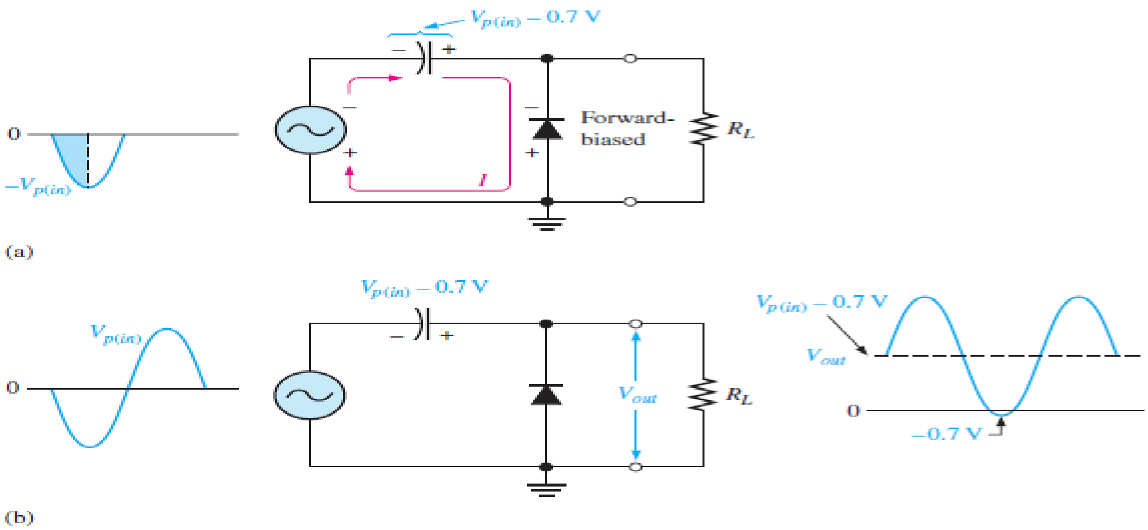


Figure 2–63: Positive clamper operation.

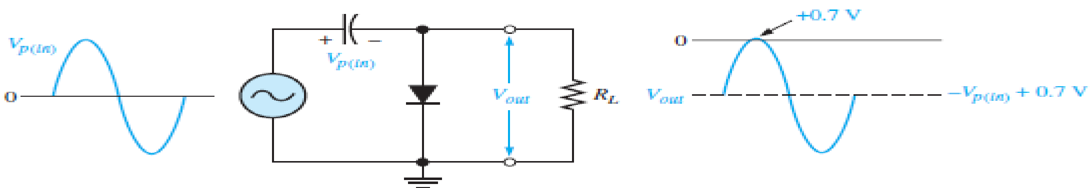


Figure 2–64: Negative clamper.

<u>Example 2–13</u>:

What is the output voltage that you would expect to observe across R_L in the clamping circuit of Figure 2–65? Assume that *RC* is large enough to prevent significant capacitor discharge.

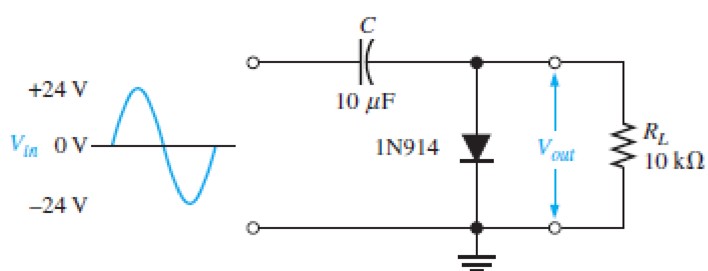


Figure 2–65

***<u>Solution</u>*:**

Ideally, a negative dc value equal to the input peak less the diode drop is inserted by the clamping circuit.

$$V_{DC} \approx -(V_{p(in)} - 0.7\text{ V}) = -(24\text{ V} - 0.7\text{ V}) = \mathbf{-23.3\ V}$$

Actually, the capacitor will discharge slightly between peaks, and, as a result, the output voltage will have an average value of slightly less than that calculated above. The output waveform goes to approximately +0.7 V, as shown in Figure 2–66.

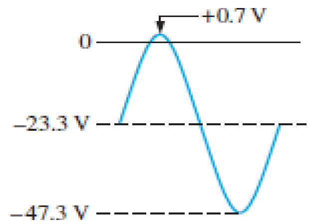


Figure 2–66: Output waveform across R_L for Figure 2–65.

***Related Problem*:**
What is the output voltage that you would observe across R_L in Figure 2–65 for C = 22 μF and R_L = 18 kΩ?

SECTION 2–7 CHECKUP:

1. Discuss how diode limiters and diode clampers differ in terms of their function.
2. What is the difference between a positive limiter and a negative limiter?
3. What is the maximum voltage across an unbiased positive silicon diode limiter during the positive alternation of the input voltage?
4. To limit the output voltage of a positive limiter to 5 V when a 10 V peak input is applied, what value must the bias voltage be?
5. What component in a clamping circuit effectively acts as a battery?

2–8 Voltage multipliers

Voltage multipliers use clamping action to increase peak rectified voltages without the necessity of increasing the transformer's voltage rating. Multiplication factors of two, three, and four are common. Voltage multipliers are used in high-voltage, low-current applications such as cathode-ray tubes (CRTs) and particle accelerators.

Voltage Doubler

***Half-Wave Voltage Doubler*:** A voltage doubler is a **voltage multiplier** with a multiplication factor of two. A half-wave voltage doubler is shown in Figure 2–67. During the positive half-cycle of the secondary voltage, diode D_1 is forward-biased and D_2 is reverse-biased. Capacitor C_1 is charged to the peak of the secondary voltage (V_p) less the diode drop with the polarity shown in part (a). During the negative half-cycle, diode D_2 is forward-biased and D_1 is reverse-biased, as shown in part (b). Since C_1 can't discharge, the peak voltage on C_1 adds to the secondary voltage to charge C_2 to approximately $2V_p$. Applying Kirchhoff's law around the loop as shown in part (b), the voltage across C_2 is

$$V_{C1} - V_{C2} + V_p = 0$$

$$V_{C2} = V_p + V_{C1}$$

Neglecting the diode drop of D_2, $V_{C1} = V_p$. Therefore,

$$V_{C2} = V_p + V_p = 2Vp$$

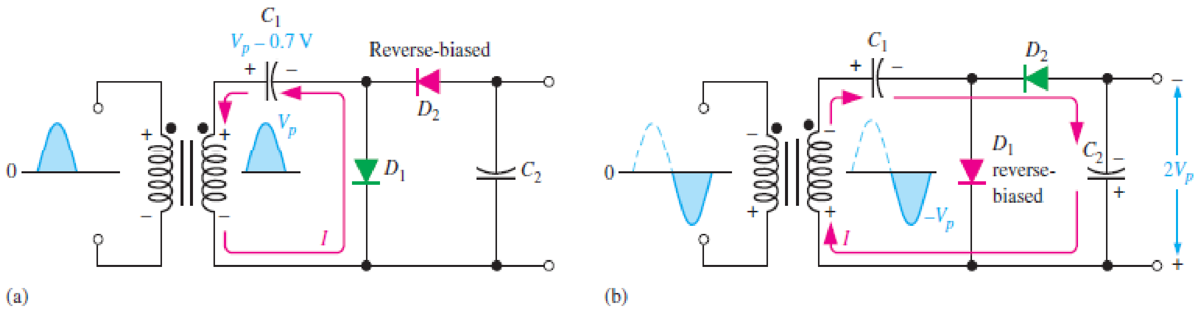


Figure 2–67: Half-wave voltage doubler operation. V_p is the peak secondary voltage.

Under a no-load condition, C_2 remains charged to approximately $2V_p$. If a load resistance is connected across the output, C_2 discharges slightly through the load on the next positive half-cycle and is again recharged to $2V_p$ on the following negative half-cycle. The resulting output is a half-wave, capacitor-filtered voltage. The peak inverse voltage across each diode is $2V_p$. If the diode were reversed, the output voltage across C_2 would have the opposite polarity.

<u>Full-Wave Voltage Doubler</u>: A full-wave doubler is shown in Figure 2–68. When the secondary voltage is positive, D_1 is forward-biased and C_1 charges to approximately V_p, as shown in part (a). During the negative half-cycle, D_2 is forward-biased and C_2 charges to approximately V_p, as shown in part (b). The output voltage, $2V_p$, is taken across the two capacitors in series.

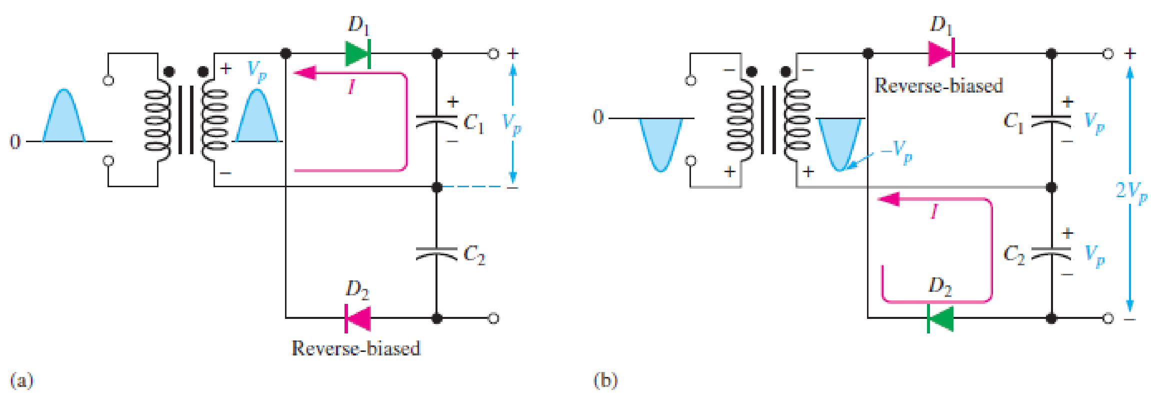


Figure 2–68: Full-wave voltage doubler operation.

<u>Voltage Tripler</u>

The addition of another diode-capacitor section to the half-wave voltage doubler creates a voltage tripler, as shown in Figure 2–69. The operation is as follows: On the positive half-cycle of the secondary voltage, C_1 charges to V_p through D_1. During the negative halfcycle, C_2 charges to $2V_p$ through D_2, as described for the doubler. During the next positive half-cycle, C_3 charges to $2V_p$ through D_3. The tripler output is taken across C_1 and C_3, as shown in the figure.

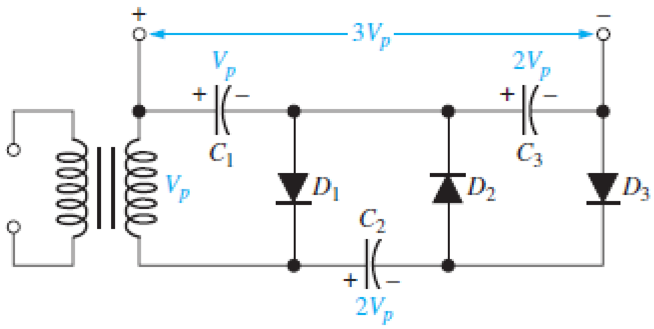


Figure 2–69: Voltage tripler.

Voltage Quadrupler

The addition of still another diode-capacitor section, as shown in Figure 2–70, produces an output four times the peak secondary voltage. C_4 charges to $2V_p$ through D_4 on a negative half-cycle. The $4V_p$ output is taken across $C2$ and $C4$, as shown. In both the tripler and quadrupler circuits, the PIV of each diode is $2V_p$.

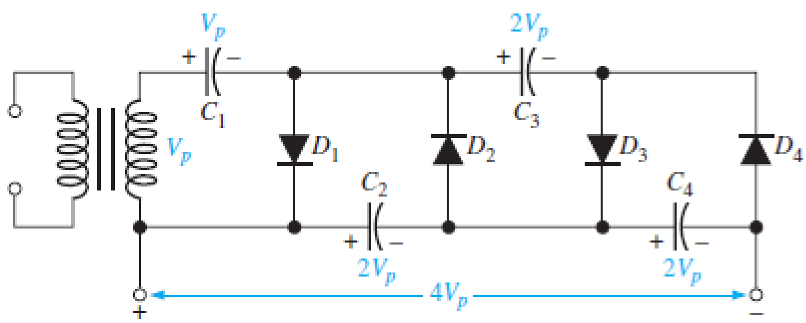


Figure 2–70: Voltage quadrupler.

SECTION 2–8 CHECKUP:

1. What must be the peak voltage rating of the transformer secondary for a voltage doubler that produces an output of 200 V?
2. The output voltage of a quadrupler is 620 V. What minimum PIV rating must each diode have?

PROBLEMS:

Section 2–1

1. To forward-bias a diode, to which region must the positive terminal of a voltage source be connected?

2. Explain why a series resistor is necessary when a diode is forward-biased.

Section 2–2

3. Explain how to generate the forward-bias portion of the characteristic curve.

4. What would cause the barrier potential of a silicon diode to decrease from 0.7 V to 0.6 V?

Section 2–3

5. Determine whether each silicon diode in Figure 2–85 is forward-biased or reverse-biased.

6. Determine the voltage across each diode in Figure 2–85, assuming the practical model.
7. Determine the voltage across each diode in Figure 2–85, assuming an ideal diode.
8. Determine the voltage across each diode in Figure 2–85, using the complete diode model with $r'_d = 10\ \Omega$ and $r'_R = 100\ \text{M}\Omega$.

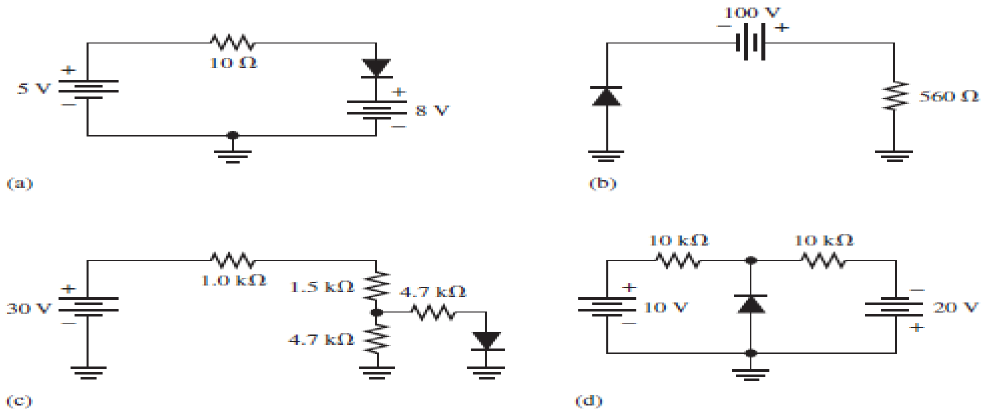


Figure 2–85

Section 2–4

9. Draw the output voltage waveform for each circuit in Figure 2–86 and include the voltage values.

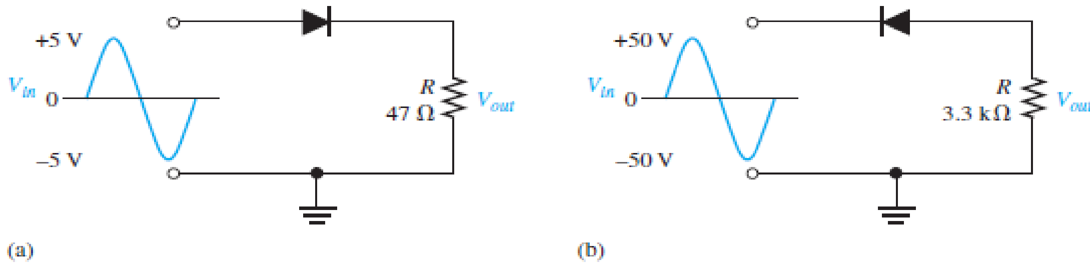


Figure 2–86

10. What is the peak inverse voltage across each diode in Figure 2–86?
11. Calculate the average value of a half-wave rectified voltage with a peak value of 200 V.
12. What is the peak forward current through each diode in Figure 2–86?
13. A power-supply transformer has a turns ratio of 5:1. What is the secondary voltage if the primary is connected to a 120 V rms source?
14. Determine the peak and average power delivered to R_L in Figure 2–87.

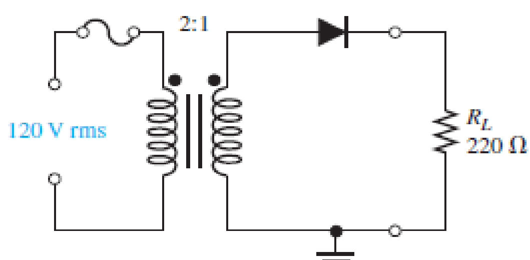


Figure 2–87

Section 2–5

15. Find the average value of each voltage in Figure 2–88.

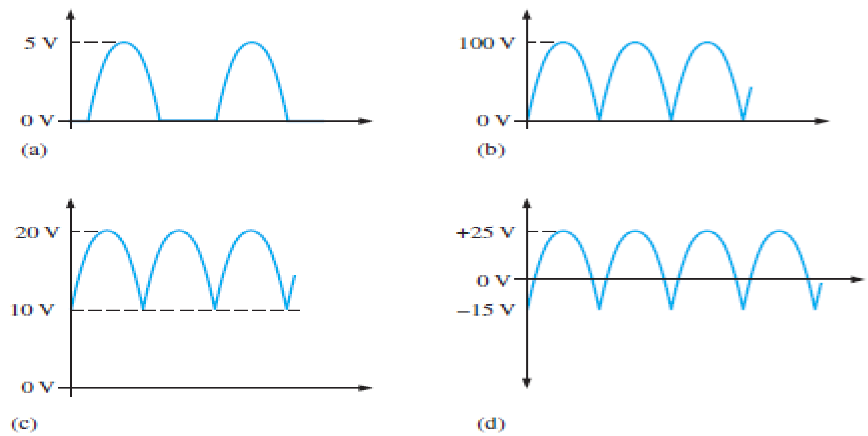


Figure 2–88

16. Consider the circuit in Figure 2–89.

(a) What type of circuit is this?

(b) What is the total peak secondary voltage?

(c) Find the peak voltage across each half of the secondary.

(d) Sketch the voltage waveform across R_L.

(e) What is the peak current through each diode?

(f) What is the PIV for each diode?

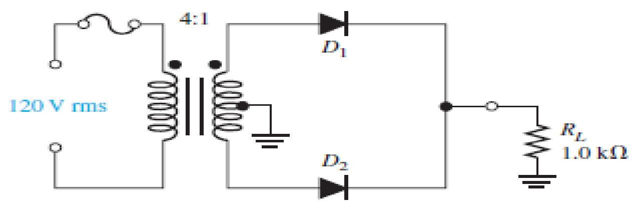


Figure 2–89

17. Calculate the peak voltage across each half of a center-tapped transformer used in a full-wave rectifier that has an average output voltage of 120 V.

18. Show how to connect the diodes in a center-tapped rectifier in order to produce a negative- going full-wave voltage across the load resistor.

19. What PIV rating is required for the diodes in a bridge rectifier that produces an average output voltage of 50 V?

20. The rms output voltage of a bridge rectifier is 20 V. What is the peak inverse voltage across the diodes?

21. Draw the output voltage waveform for the bridge rectifier in Figure 2–90. Notice that all the diodes are reversed from circuits shown earlier in the chapter.

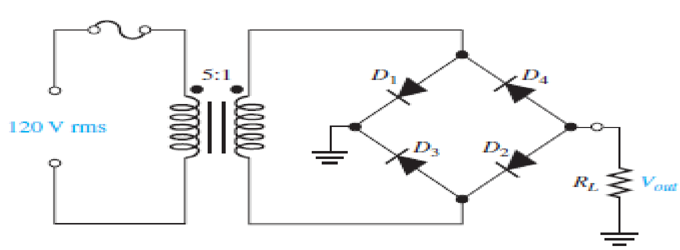


Figure 2–90

Section 2–6

22. A certain rectifier filter produces a dc output voltage of 75 V with a peak-to-peak ripple voltage of 0.5 V. Calculate the ripple factor.

23. A certain full-wave rectifier has a peak output voltage of 30 V. A 50 F capacitor-input filter is connected to the rectifier. Calculate the peak-to-peak ripple and the dc output voltage developed across 600 Ω a load resistance.

24. What is the percentage of ripple for the rectifier filter in Problem 23?

25. What value of filter capacitor is required to produce a 1% ripple factor for a full-wave rectifier having a load resistance of 1.5 kΩ? Assume the rectifier produces a peak output of 18 V.

26. A full-wave rectifier produces an 80 V peak rectified voltage from a 60 Hz ac source. If a 10 µF filter capacitor is used, determine the ripple factor for a load resistance of 10 kΩ.

27. Determine the peak-to-peak ripple and dc output voltages in Figure 2–91. The transformer has a 36 V rms secondary voltage rating, and the line voltage has a frequency of 60 Hz.

28. Refer to Figure 2–91 and draw the following voltage waveforms in relationship to the input waveforms: V_{AB}, V_{AD}, and V_{CD}. A double letter subscript indicates a voltage from one point to another.

29. If the no-load output voltage of a regulator is 15.5 V and the full-load output is 14.9 V, what is the percent load regulation?

30. Assume a regulator has a percent load regulation of 0.5%. What is the output voltage at fullload if the unloaded output is 12.0 V?

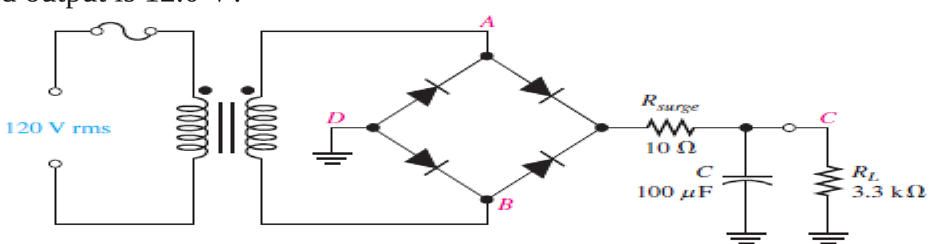


Figure 2–91

Section 2–7

31. Determine the output waveform for the circuit of Figure 2–92.

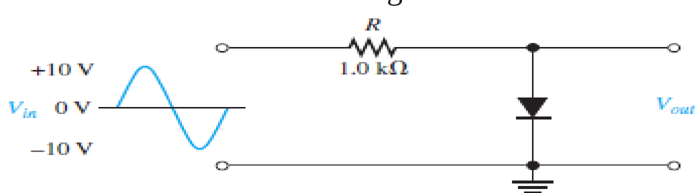


Figure 2–92

32. Determine the output voltage for the circuit in Figure 2–93(a) for each input voltage in (b), (c), and (d).

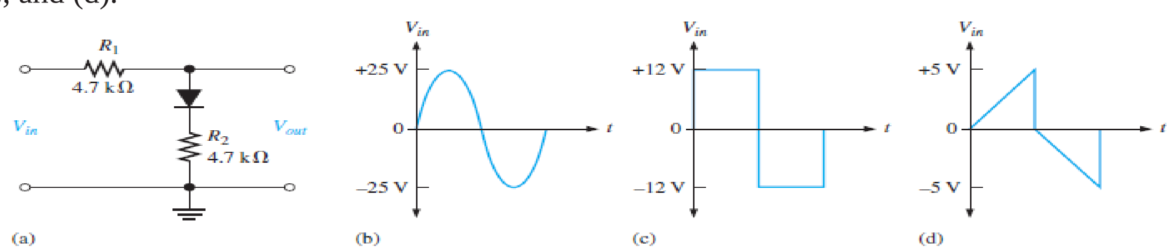


Figure 2–93

33. Determine the output voltage waveform for each circuit in Figure 2–94.

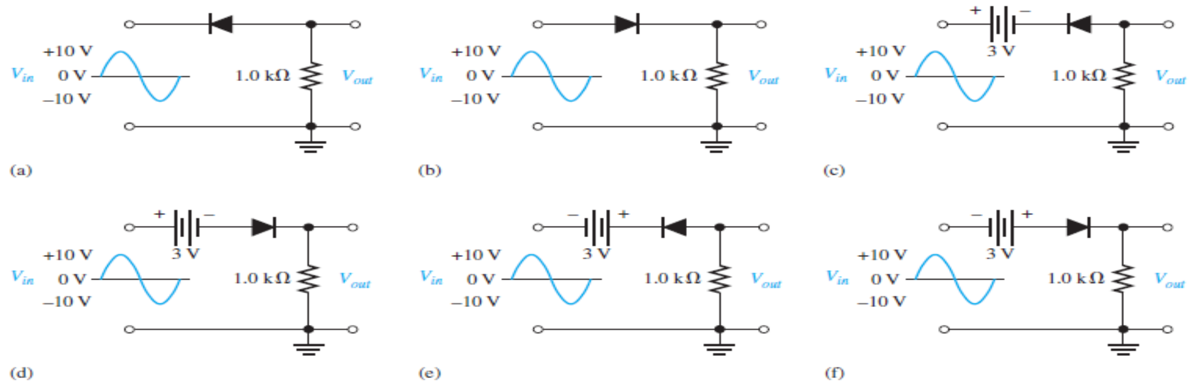


Figure 2–94

34. Determine the R_L voltage waveform for each circuit in Figure 2–95.

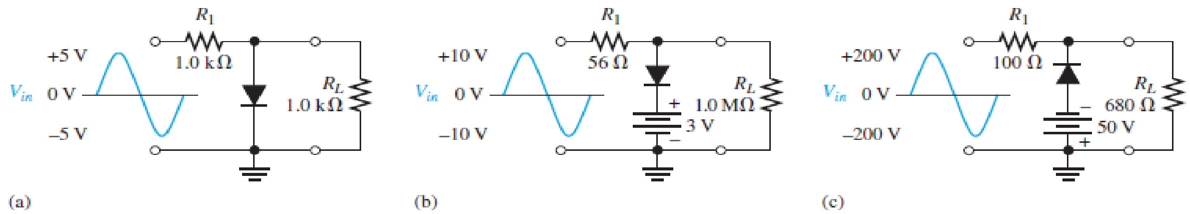


Figure 2–95

35. Draw the output voltage waveform for each circuit in Figure 2–96.

36. Determine the peak forward current through each diode in Figure 2–96.

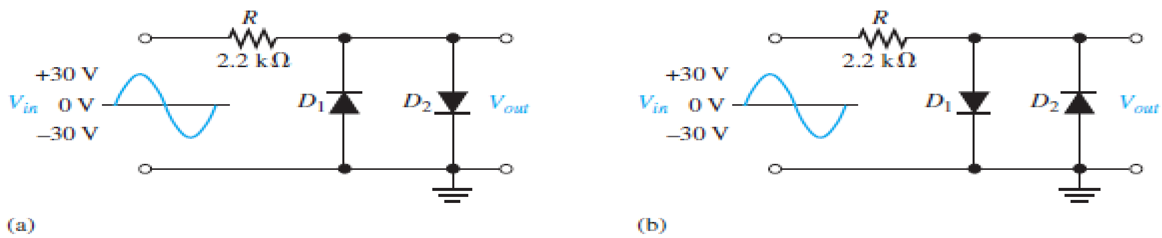


Figure 2–96

37. Determine the peak forward current through each diode in Figure 2–97.

38. Determine the output voltage waveform for each circuit in Figure 2–97.

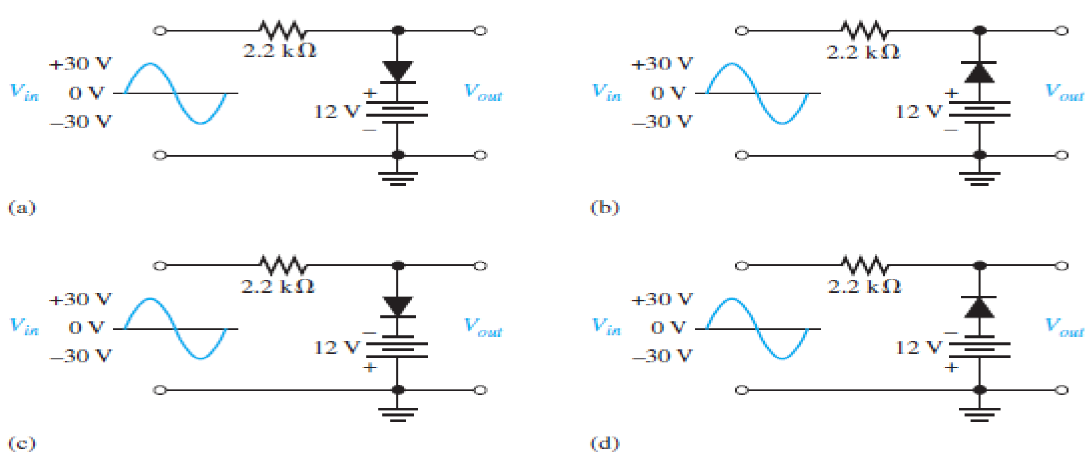


Figure 2–97

39. Describe the output waveform of each circuit in Figure 2–98. Assume the *RC* time constant is much greater than the period of the input.

40. Repeat Problem 39 with the diodes turned around.

Section 2–8

41. A certain voltage doubler has 20 V rms on its input. What is the output voltage? Draw the circuit, indicating the output terminals and PIV rating for the diode.

42. Repeat Problem 41 for a voltage tripler and quadrupler.

Section 2–9

43. From the datasheet in Figure 2–71, determine how much peak inverse voltage that a 1N4002 diode can withstand.

44. Repeat Problem 43 for a 1N4007.

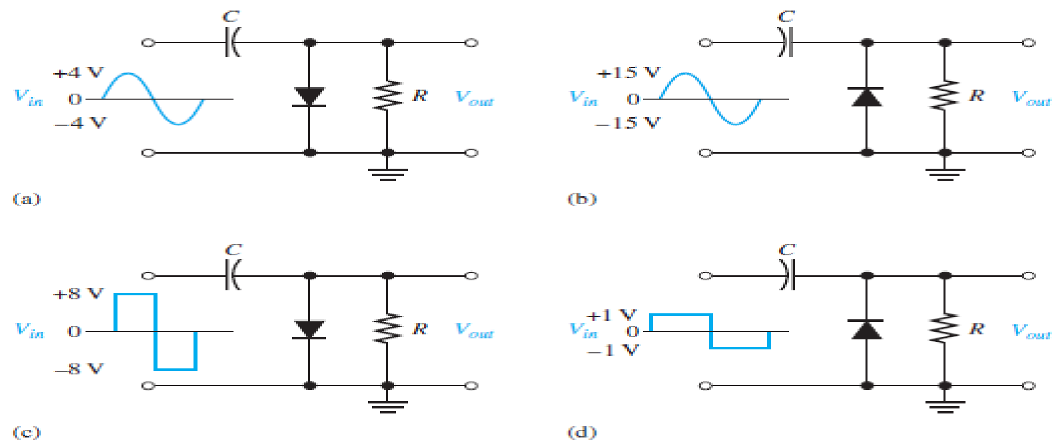


Figure 2–98

45. If the peak output voltage of a bridge full-wave rectifier is 50 V, determine the minimum value of the load resistance that can be used when 1N4002 diodes are used.

Section 2–10

46. Consider the meter indications in each circuit of Figure 2–99, and determine whether the diode is functioning properly, or whether it is open or shorted. Assume the ideal model.

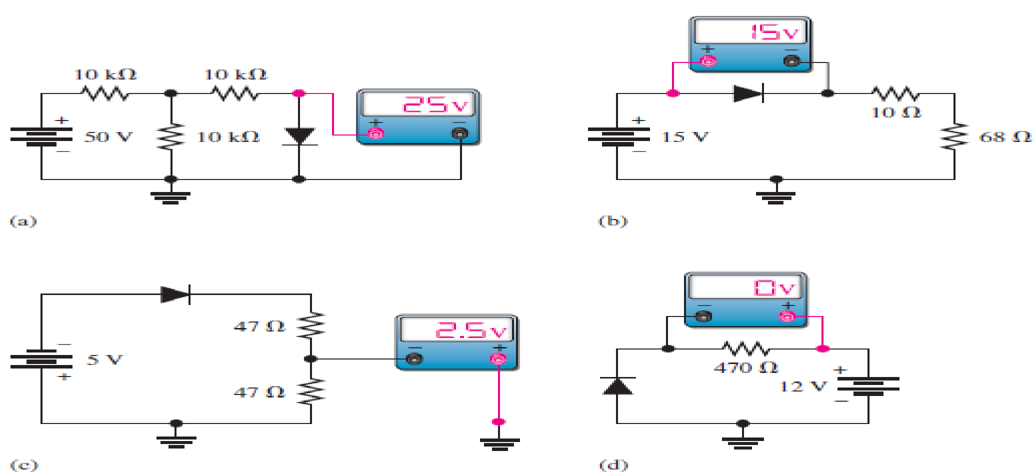


Figure 2–99

47. Determine the voltage with respect to ground at each point in Figure 2–100. Assume the practical model.

48. If one of the diodes in a bridge rectifier opens, what happens to the output?

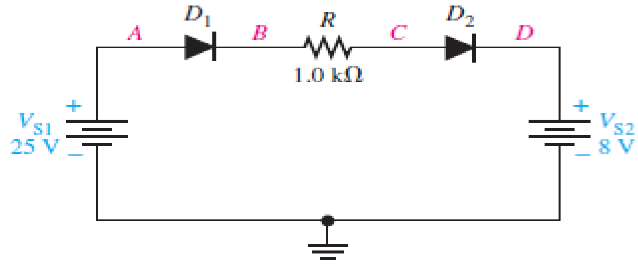


Figure 2–100

49. From the meter readings in Figure 2–101, determine if the rectifier is functioning properly. If it is not, determine the most likely failure(s).

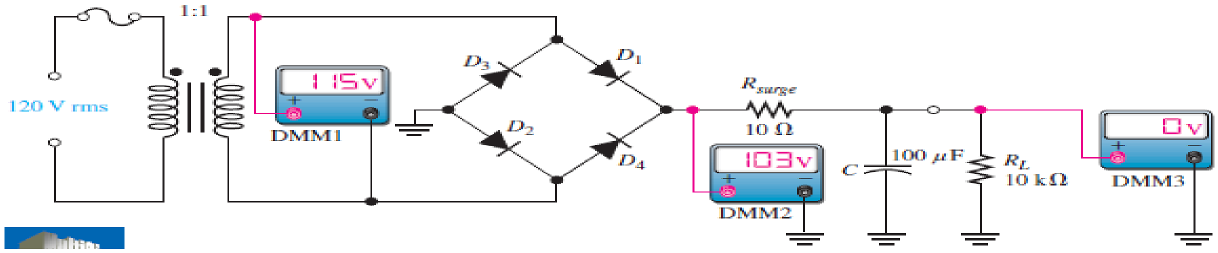
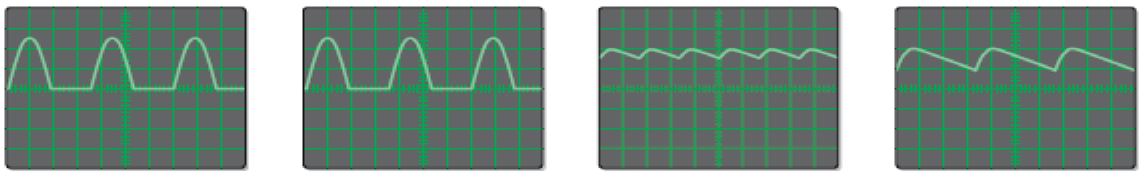


Figure 2–101

50. Each part of Figure 2–102 shows oscilloscope displays of various rectifier output voltages. In each case, determine whether or not the rectifier is functioning properly and if it is not, determine the most likely failure(s).



(a) Output of a half-wave unfiltered rectifier

(b) Output of a full-wave unfiltered rectifier

(c) Output of a full-wave filter

(d) Output of same full-wave filter as part (c)

Figure 2–102

51. Based on the values given, would you expect the circuit in Figure 2–103 to fail? If so, why?

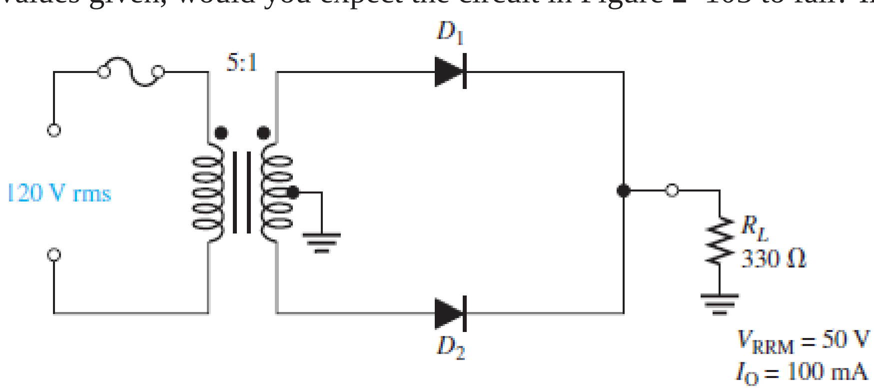


Figure 2–103

3–1 The zener diode

A major application for zener diodes is as a type of voltage regulator for providing stable reference voltages for use in power supplies, voltmeters, and other instruments. In this section, you will see how the zener diode maintains a nearly constant dc voltage under the proper operating conditions. You will learn the conditions and limitations for properly using the zener diode and the factors that affect its performance.

The symbol for a zener diode is shown in Figure 3–1. Instead of a straight line representing the cathode, the zener diode has a bent line that reminds you of the letter Z (for zener). A **zener diode** is a silicon ***pn*** junction device that is designed for operation in the reverse-breakdown region. The breakdown voltage of a zener diode is set by carefully controlling the doping level during manufacture. Recall, from the discussion of the diode characteristic curve in Chapter 2, that when a diode reaches reverse breakdown, its voltage remains almost constant even though the current changes drastically, and this is the key to zener diode operation. This volt-ampere characteristic is shown again in Figure 3–2 with the normal operating region for zener diodes shown as a shaded area.

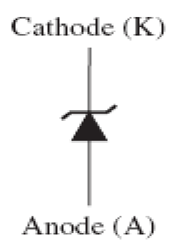


Figure (3-1): Zener diode symbol.

Zener Breakdown

Zener diodes are designed to operate in reverse breakdown. Two types of reverse breakdown in a zener diode are *avalanche* and *zener*. The avalanche effect, discussed in Chapter 2, occurs in both rectifier and zener diodes at a sufficiently high reverse voltage. **Zener breakdown** occurs in a zener diode at low reverse voltages.

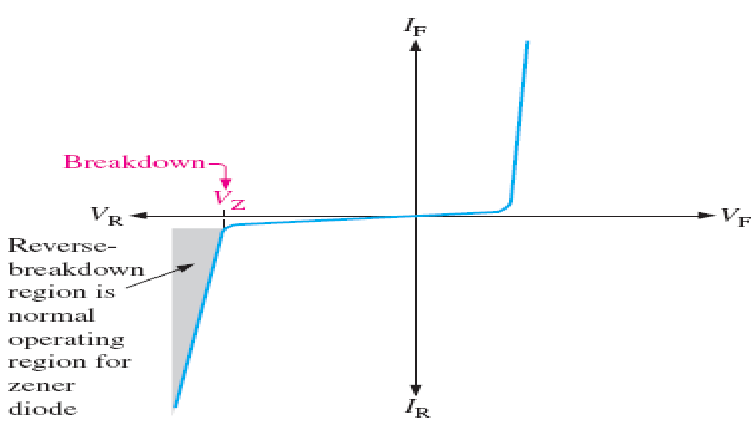


Figure (3-2): General zener diode *V-I* characteristic.

A zener diode is heavily doped to reduce the breakdown voltage. This causes a very thin depletion region. As a result, an intense electric field exists within the depletion region. Near the zener breakdown voltage (*V*Z), the field is intense enough to pull electrons from their valence bands and create current. Zener diodes with breakdown voltages of less than approximately 5 V operate predominately in zener breakdown. Those with breakdown voltages greater than approximately 5 V operate predominately in **avalanche breakdown.** Both types, however, are called *zener diodes*. Zeners are commercially available with breakdown voltages from less than 1 V to more than 250 V with specified tolerances from 1% to 20%.

Breakdown Characteristics

Figure 3–3 shows the reverse portion of a zener diode's characteristic curve. Notice that as the reverse voltage (V_R) is increased, the reverse current (I_R) remains extremely small up to the "knee" of the curve. The reverse current is also called the zener current, I_Z. At this point, the breakdown effect begins; the internal zener resistance, also called zener impedance (Z_Z), begins to decrease as the reverse current increases rapidly. From the bottom of the knee, the zener breakdown voltage (V_Z) remains essentially constant although it increases slightly as the zener current, I_Z, increases.

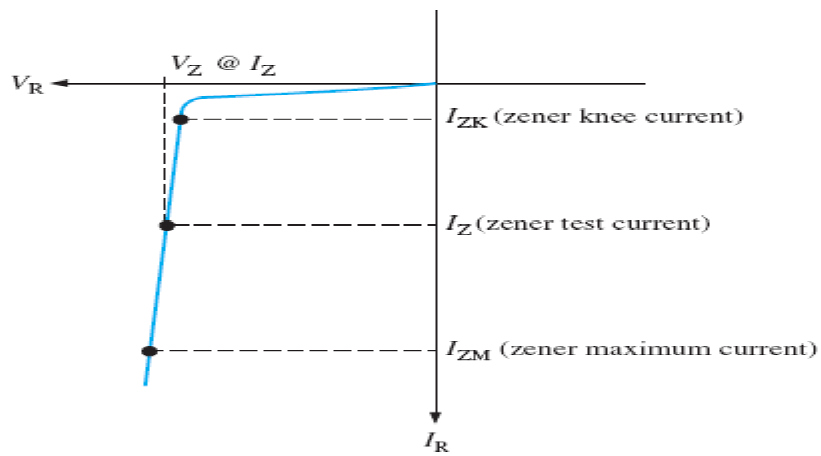


Figure (3-3): Reverse characteristic of a zener diode. *V*Z is usually specified at a value of the zener current known as the test current.

Zener Regulation

The ability to keep the reverse voltage across its terminals essentially constant is the key feature of the zener diode. A zener diode operating in breakdown acts as a voltage regulator because it maintains a nearly constant voltage across its terminals over a specified range of reverse-current values.

A minimum value of reverse current, I_{ZK}, must be maintained in order to keep the diode in breakdown for voltage regulation. You can see on the curve in Figure 3–3 that when the reverse current is reduced below the knee of the curve, the voltage decreases drastically and regulation is lost. Also, there is a maximum current, I_{ZM}, above which the diode may be damaged due to excessive power dissipation. So, basically, the zener diode maintains a nearly constant voltage across its terminals for values of reverse current ranging from I_{ZK} to I_{ZM}. A nominal zener voltage, V_Z, is usually specified on a datasheet at a value of reverse current called the *zener test current.*

Zener Equivalent Circuits

Figure 3–4 shows the ideal model (first approximation) of a zener diode in reverse breakdown and its ideal characteristic curve. It has a constant voltage drop equal to the nominal zener voltage. This constant voltage drop across the zener diode produced by reverse breakdown is represented by a dc voltage symbol even though the zener diode does not produce a voltage.

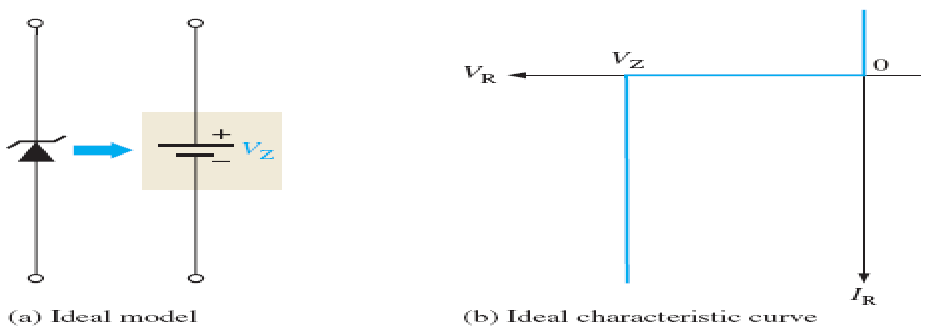


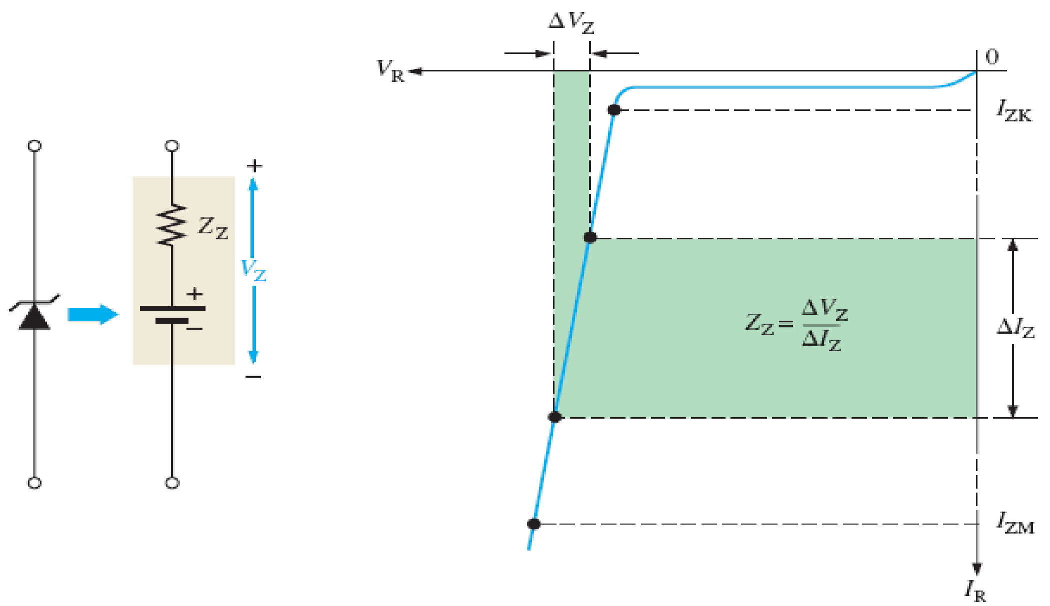
Figure (3-4): Ideal zener diode equivalent circuit model and the characteristic curve.

Figure 3–5(a) represents the practical model (second approximation) of a zener diode, where the zener impedance (resistance), Z_Z, is included. Since the actual voltage curve is not ideally vertical, a change in zener current (ΔI_Z) produces a small change in zener voltage (ΔV_Z) as illustrated in Figure 3–5(b). By Ohm's law, the ratio of is ΔV_Z to ΔI_Z the impedance, as expressed in the following equation:

$$Z_Z = \frac{\Delta V_Z}{\Delta I_Z}, \qquad (3\text{-}1)$$

Normally, *ZZ* is specified at the zener test current. In most cases, you can assume that *ZZ* is a small constant over the full range of zener current values and is purely resistive. It is best to avoid operating a zener diode near the knee of the curve because the impedance changes dramatically in that area.

For most circuit analysis and troubleshooting work, the ideal model will give very good results and is much easier to use than more complicated models. When a zener diode is operating normally, it will be in reverse breakdown and you should observe the nominal breakdown voltage across it. Most **schematics** will indicate on the drawing what this voltage should be.



(a) Practical model (b) Characteristic curve. The slope is exaggerated for illustration.

Figure (3-5): Practical zener diode equivalent circuit and the characteristic curve illustrating Z_Z.

<u>Example 3–1</u>

A zener diode exhibits a certain change in VZ for a certain change in I_Z on a portion of the linear characteristic curve between I_{ZK} and I_{ZM} as illustrated in Figure 3–6. What is the zener impedance?

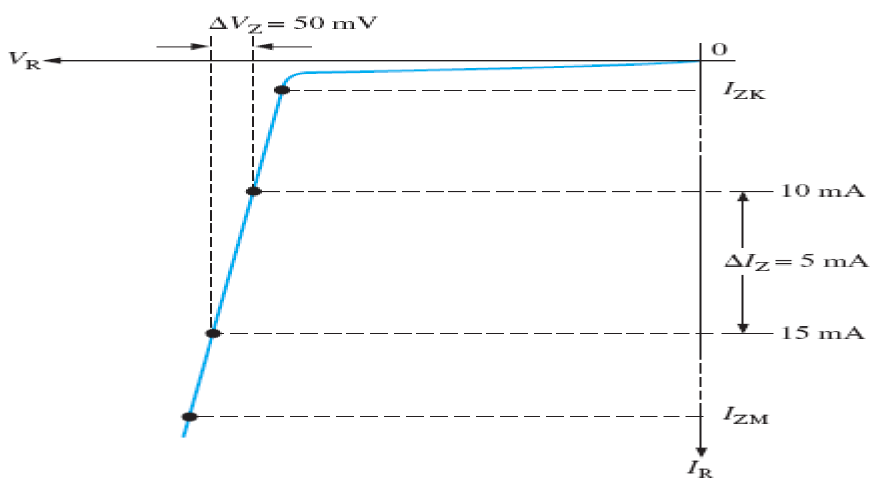


Figure (3-6)

<u>Solution</u>

$$Z_Z = \frac{\Delta V_Z}{\Delta I_Z} = \frac{50\ mV}{5\ mA} = 5\,\Omega$$

***<u>Related Problem</u>*: ** Calculate the zener impedance if the change in zener voltage is 100 mV for a 20 mA change in zener current on the linear portion of the characteristic curve.

Temperature Coefficient

The temperature coefficient specifies the percent change in zener voltage for each degree Celsius change in temperature. For example, a 12 V zener diode with a positive temperature coefficient of 0.01% /°C will exhibit a 1.2 mV increase in V_Z when the junction temperature increases one degree Celsius. The formula for calculating the change in zener voltage for a given junction temperature change, for a specified temperature coefficient, is

$$\Delta V_z = V_z \times TC \times \Delta T \quad (3\text{-}2)$$

where V_Z is the nominal zener voltage at the reference temperature of 25 °C, TC is the temperature coefficient, and ΔT is the change in temperature from the reference temperature. A positive TC means that the zener voltage increases with an increase in temperature or decreases with a decrease in temperature. A negative TC means that the zener voltage decreases with an increase in temperature or increases with a decrease in temperature. In some cases, the temperature coefficient is expressed mV/°C in rather than %/ °C as For these cases, ΔV_Z is calculated as

$$\Delta V_z = TC \times \Delta T \quad (3\text{-}3)$$

<u>**Example 3–2**</u>

An 8.2 V zener diode (8.2 V at 25 °C) has a positive temperature coefficient of 0.05%/°C. What is the zener voltage at 60 °C?

<u>***Solution***</u>

The change in zener voltage is

$$\Delta V_z = V_z \times TC \times \Delta T = 8.2V \times TC\, 0.05\%/^{\circ}C \times (60\,^{\circ}C - 25\,^{\circ}C)$$

$$= 8.2V \times 0.0005/^{\circ}C \times (35\,^{\circ}C) = 144\ mV$$

Notice that 0.05%/°C was converted to 0.0005/°C. The zener voltage at 60 °C is

$$V_Z + \Delta V_Z = 8.2\text{ V} + 144\text{ mV} = 8.34\text{ V}$$

<u>***Related Problem*****:**</u> A 12 V zener has a positive temperature coefficient of 0.075%/°C. How much will the zener voltage change when the junction temperature decreases 50 degrees Celsius?

Zener Power Dissipation and Derating

Zener diodes are specified to operate at a maximum power called the maximum dc power dissipation, P_D(max). For example, the 1N746 zener is rated at a P_D(max) of 500 mW and the 1N3305A is rated at a P_D(max) of 50 W. The dc power dissipation is determined by the formula,

$$P_D = V_Z I_Z$$

Power Derating

The maximum power dissipation of a zener diode is typically specified for temperatures at or below a certain value (for example). Above the specified temperature, the maximum power dissipation is reduced according to a derating factor. The derating factor is expressed in The maximum derated power can be determined with the following formula:

$$P_{D(derated)} = P_{D(max)} - (mW/°C)\ \Delta T$$

Example 3–3

A certain zener diode has a maximum power rating of 400 mW at 50 ^{o}C and a derating factor of 3.2 *mW*/°C. Determine the maximum power the zener can dissipate at a temperature of 90 ^{o}C.

Solution

$$P_{D(derated)} = P_{D(max)} - (mW/°C)\ \Delta T$$
$$= 400\text{ mW} - (3.2\ mW/°C)(90\ ^{o}C - 50\ ^{o}C)$$
$$= 400\text{ mW} - 128\text{ mW} = 272\text{ mW}$$

Related Problem**:** A certain 50 W zener diode must be derated with a derating factor of 0.5 *W*/°C above 75°C. Determine the maximum power it can dissipate at 160°C.

Electrical Characteristics**:** The first column in the datasheet lists the zener type numbers, 1N4728A through 1N4764A.

Zener voltage, *V*Z, and zener test current, I_Z: For each device type, the minimum, typical, and maximum zener voltages are listed. V_Z is measured at the specified zener test current, I_Z. For example, the zener voltage for a 1N4728A can range from 3.315 V to 3.465 V with a typical value of 3.3 V at a test current of 76 mA.

Maximum zener impedance Z_Z is the maximum zener impedance at the specified test current, I_Z. For example, for a 1N4728A, Z_Z is 10 Ώ at 76 mA. The maximum zener impedance **C**, at the knee of the characteristic curve is specified at I_{ZK} which is the current at the knee of the curve. For example, Z_{ZK} is 400 Ώ at 1 mA for a 1N4728A.

Leakage current Reverse leakage current is specified for a reverse voltage that is less than the knee voltage. This means that the zener is not in reverse breakdown for these measurements. For example I_R is 100 µA for a reverse voltage of 1 V in a 1N4728A.

Example 3–4

From the datasheet in Figure 3–7, a 1N4736A zener diode has a Z_Z of 3.5 Ώ. The datasheet gives V_Z = 6.8 V at a test current, I_Z, of 37 mA. What is the voltage across the zener terminals when the current is 50 mA? When the current is 25 mA? Figure 3–8 represents the zener diode.

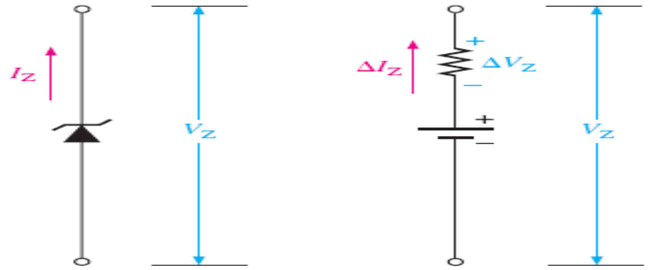


Figure (3-8):

Solution

For I_Z = 50 mA: The 50 mA current is a 13 mA increase above the test current, I_Z, of 37 mA.

$$\Delta I_Z = I_Z - 37\text{ mA} = 50\text{ mA} - 37\text{ mA} = +13\text{ mA}$$
$$\Delta V_Z = \Delta I_Z Z_Z = (13\text{ mA})(3.5\,\Omega) = +45.5\text{ mV}$$

The change in voltage due to the increase in current above the *IZ* value causes the zener terminal voltage to increase. The zener voltage for I_Z = 50 mA is

$$V_Z = 6.8\text{ V} + \Delta V_Z = 6.8\text{ V} + 45.5\text{ mV} = \mathbf{6.85\text{ V}}$$

For I_Z = 25 mA: The 25 mA current is a 12 mA decrease below the test current, I_Z, of 37 mA.

$$\Delta I_Z = -12\text{ mA}$$
$$\Delta V_Z = \Delta I_Z Z_Z = (-12\text{ mA})(3.5\,\Omega) = -42\text{ mV}$$

The change in voltage due to the decrease in current below the test current causes the zener terminal voltage to decrease. The zener voltage for I_Z = 25 mA is

$$V_Z = 6.8\text{ V} - \Delta V_Z = 6.8\text{ V} - 42\text{ mV} = \mathbf{6.76\text{ V}}$$

Related Problem: Repeat the analysis for I_Z = 10 mA and for I_Z = 30 mA using a 1N4742A zener with V_Z = 12 V at I_Z = 21 mA and $Z_Z = 9\,\Omega$.

SECTION 3–1 CHECKUP

1. In what region of their characteristic curve are zener diodes operated?
2. At what value of zener current is the zener voltage normally specified?
3. How does the zener impedance affect the voltage across the terminals of the device?
4. What does a positive temperature coefficient of **0.05%/°C** mean?
5. Explain power derating.

3–2 Zener diode applications

The zener diode can be used as a type of voltage regulator for providing stable reference voltages. In this section, you will see how zeners can be used as voltage references, regulators, and as simple limiters or clippers.

Zener Regulation with a Variable Input Voltage

Zener diode regulators can provide a reasonably constant dc level at the output, but they are not particularly efficient. For this reason, they are limited to applications that require only low current to the load. Figure 3–9 illustrates how a zener diode can be used to regulate a dc voltage.

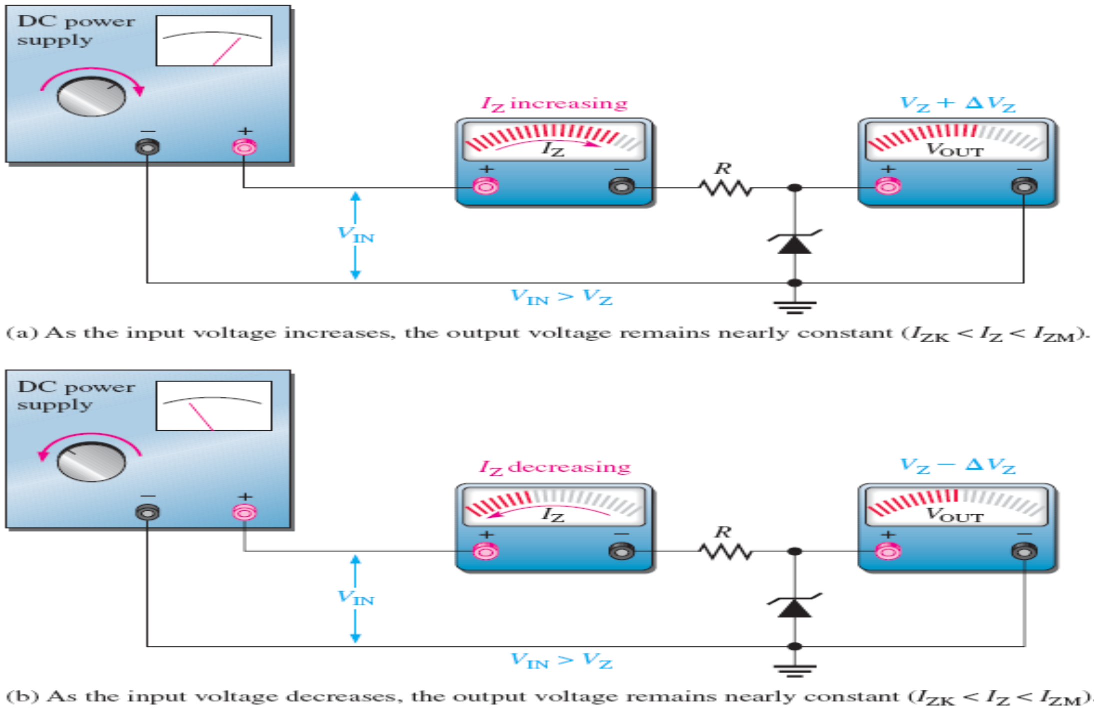


Figure (3-9): Zener regulation of a varying input voltage.

As the input voltage varies (within limits), the zener diode maintains a nearly constant output voltage across its terminals. However, as V_{IN} changes, I_Z will change proportionally so that the limitations on the input voltage variation are set by the minimum and maximum current values (I_{ZK} and I_{ZM}) with which the zener can operate. Resistor R is the series current-limiting resistor. The meters indicate the relative values and trends.

To illustrate regulation, let's use the ideal model of the 1N4740A zener diode (ignoring the zener resistance) in the circuit of Figure 3–10. The absolute lowest current that will maintain regulation is specified at which for the 1N4740A is 0.25 mA and represents the no-load current. The maximum current is not given on the datasheet but can be calculated from the power specification of 1 W, which is given on the datasheet. Keep in mind that both the minimum and maximum values are at the operating extremes and represent worst-case operation.

$$I_{ZM} = \frac{P_{D(\max)}}{V_Z} = \frac{1\ W}{10V} = 100\ mA$$

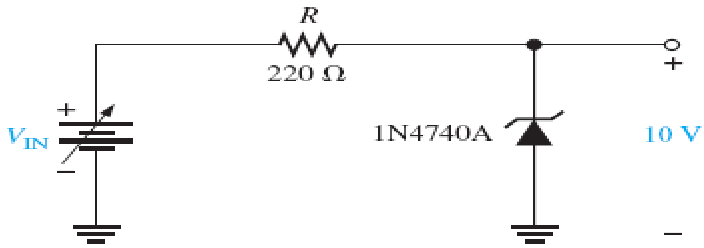


Figure (3-10)

For the minimum zener current, the voltage across the 220 Ω resistor is

$$V_R = I_{ZK}R = (0.25\text{ mA})(220\ \Omega) = 55\text{ mV}$$

Since $V_R = V_{IN} - V_Z$,

$$V_{IN(min)} = V_R + V_Z = 55\text{ mV} + 10\text{ V} = 10.055\text{ V}$$

For the maximum zener current, the voltage across the 220 Ω resistor is

$$V_R = I_{ZM}R = (100\text{ mA})(220\ \Omega) = 22\text{ V}$$

Therefore,

$$V_{IN(max)} = 22\text{ V} + 10\text{ V} = 32\text{ V}$$

This shows that this zener diode can ideally regulate an input voltage from 10.055 V to 32 V and maintain an approximate 10 V output. The output will vary slightly because of the zener impedance, which has been neglected in these calculations.

Example 3–5

Determine the minimum and the maximum input voltages that can be regulated by the zener diode in Figure 3–11.

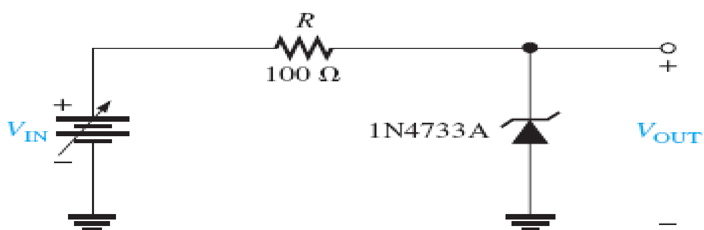


Figure (3-11)

Solution

From the datasheet in Figure 3–7 for the 1N4733A: V_Z = 5.1 V at I_Z = 49 mA, I_{ZK} = 1 mA, and Z_Z = 7 Ω at I_Z. For simplicity, assume this value of Z_Z over the range of current values. The equivalent circuit is shown in Figure 3–12.

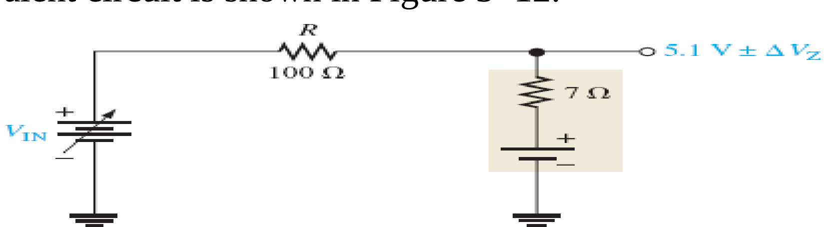


Figure (3-12): Equivalent of circuit in Figure 3–11.

At I_{ZK} =1 mA, the output voltage is

$$V_{OUT} = 5.1\text{ V} - \Delta V_Z = 5.1\text{ V} - (I_Z - I_{ZK})Z_Z = 5.1\text{ V} - (49\text{ mA} - 1\text{ mA})(7\ \Omega)$$
$$= 5.1\text{ V} - (48\text{ mA})(7\ \Omega) = 5.1\text{ V} - 0.336\text{ V} = 4.76\text{ V}$$

Therefore,

$$V_{IN(min)} = I_{ZK}R + V_{OUT} = (1\text{ mA})(100\ \Omega) + 4.76\text{ V} = \mathbf{4.86\ V}$$

To find the maximum input voltage, first calculate the maximum zener current. Assume the temperature is 50 °C or below; so from Figure 3–7, the power dissipation is 1 W.

$$I_{ZM} = \frac{P_{D(max)}}{V_Z} = \frac{1\ W}{5.1V} = 196\ mA$$

At I_{ZM}, the output voltage is

$$V_{OUT} = 5.1\text{ V} + \Delta VZ = 5.1\text{ V} + (I_{ZM} - I_Z)Z_Z$$
$$= 5.1\text{ V} + (147\text{ mA})(7\ \Omega) = 5.1\text{ V} + 1.03\text{ V} = 6.13\text{ V}$$

Therefore,

$$V_{IN(max)} = I_{ZM}R + V_{OUT} = (196\text{ mA})(100\ \Omega) + 6.13\text{ V} = \mathbf{25.7\ V}$$

***Related Problem*:** Determine the minimum and maximum input voltages that can be regulated if a 1N4736A zener diode is used in Figure 3–11.

Zener Regulation with a Variable Load:

Figure 3–13 shows a zener voltage regulator with a variable load resistor across the terminals. The zener diode maintains a nearly constant voltage across R_L as long as the zener current is greater than I_{ZK} and less than I_{ZM}.

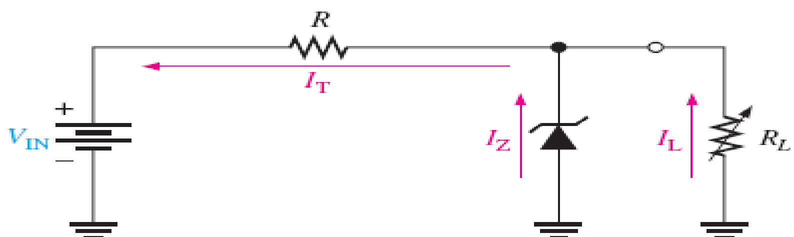


Figure (3-13): Zener regulation with a variable load.

From No Load to Full Load

When the output terminals of the zener regulator are open ($R_L = \infty$) the load current is zero and *all* of the current is through the zener; this is a no-load condition. When a load resistor R_L is connected, part of the total current is through the zener and part through R_L. The total current through R remains essentially constant as long as the zener is regulating. As R_L is decreased, the load current, I_L, increases and I_Z decreases. The zener diode continues to regulate the voltage I_Z until reaches its minimum value, I_{ZK}. At this point the load current is maximum, and a full-load condition exists. The following example will illustrate this.

Example 3–6

Determine the minimum and the maximum load currents for which the zener diode in Figure 3–14 will maintain regulation. What is the minimum value of R_L that can be used? V_Z = 12 V, I_{ZK} = 1 mA, and I_{ZM} = 50 mA. Assume an ideal zener diode where Z_Z = 0 Ω and V_Z remains a constant 12 V over the range of current values, for simplicity.

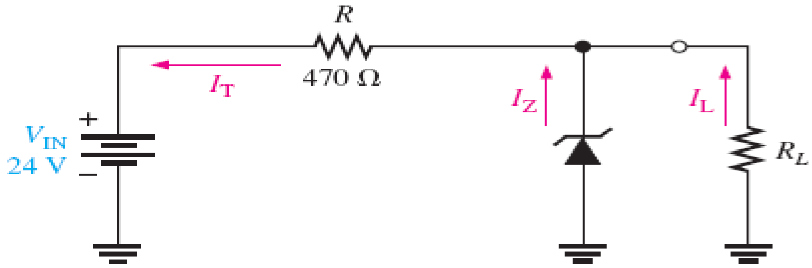


Figure (3-14)

***Solution*:**

When I_L = 0 A ($R_L = \infty$), I_Z is maximum and equal to the total circuit current I_T.

$$I_{Z\,(\max)} = I_T = \frac{V_{IN} - V_Z}{R} = \frac{24V - 12V}{470\,\Omega} = 25.5\,mA$$

If R_L is removed from the circuit, the load current is 0 A. Since $I_{Z(\max)}$ is less than I_{ZM}, 0 A is an acceptable minimum value for I_L because the zener can handle all of the 25.5 mA.

$$I_{L(\min)} = \mathbf{0\ A}$$

The maximum value of I_L occurs when I_Z is minimum ($I_Z = I_{ZK}$), so

$$I_{L(\max)} = I_T - I_{ZK} = 25.5\text{ mA} - 1\text{ mA} = \mathbf{24.5\ mA}$$

The minimum value of R_L is

$$R_{L\,(\min)} = \frac{V_Z}{I_{L\,(\max)}} = \frac{12V}{24.5\,mA} = 490\,\Omega$$

Therefore, if R_L is less than 490 Ω, R_L will draw more of the total current away from the zener and I_Z will be reduced below I_{ZK}. This will cause the zener to lose regulation. Regulation is maintained for any value of R_L between 490 Ω and infinity.

***Related Problem*:**

Find the minimum and maximum load currents for which the circuit in Figure 3–14 will maintain regulation. Determine the minimum value of R_L that can be used. V_Z = 3.3 V (constant), I_{ZK} = 1 mA, and I_{ZM} = 150 mA. Assume an ideal zener.

In the last example, we assumed that Z_Z was zero and, therefore, the zener voltage remained constant over the range of currents. We made this assumption to demonstrate the concept of how the regulator works with a varying load. Such an assumption is often acceptable and in many cases produces results that are reasonably accurate. In Example 3–7, we will take the zener impedance into account.

Example 3–7

For the circuit in Figure 3–15:

(a) Determine V_{OUT} at I_{ZK} and at I_{ZM}.

(b) Calculate the value of R that should be used.

(c) Determine the minimum value of R_L that can be used.

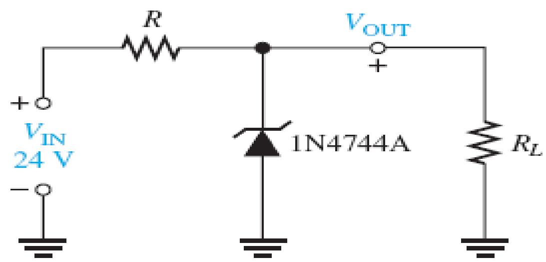


Figure 3–15

Solution:

The 1N4744A zener used in the regulator circuit of Figure 3–15 is a 15 V diode. The datasheet in Figure 3–7 gives the following information:

$V_Z = 15$ V @ $I_Z = 17$ mA, $I_{ZK} = 0.25$ mA, and $Z_Z = 14\ \Omega$.

(a) For I_{ZK}:

$V_{OUT} = V_Z - \Delta I_Z Z_Z = 15\text{ V} - \Delta I_Z Z_Z = 15\text{ V} - (I_Z - I_{ZK})Z_Z$

$= 15\text{ V} - (16.75\text{ mA})(14\ \Omega) = 15\text{ V} - 0.235\text{ V} = \mathbf{14.76\ V}$

Calculate the zener maximum current. The maximum power dissipation is 1 W.

$$I_{ZM} = \frac{P_{D(\max)}}{V_Z} = \frac{1\ W}{15V} = 66.7\ mA$$

For I_{ZM}:

$V_{OUT} = V_Z + \Delta I_Z Z_Z = 15\text{ V} + \Delta I_Z Z_Z$

$= 15\text{ V} + (I_{ZM} - I_Z)Z_Z = 15\text{ V} + (49.7\text{ mA})(14\ \Omega) = \mathbf{15.7\ V}$

(b) Calculate the value of R for the maximum zener current that occurs when there is no load as shown in Figure 3–16(a).

$$R = \frac{V_{IN} - V_{OUT}}{I_{ZK}} = \frac{24V - 15.7V}{66.7\ mA} = 124\ \Omega$$

$R = \mathbf{130}\ \Omega$ (nearest larger standard value).

(c) For the minimum load resistance (maximum load current), the zener current is minimum (*I*ZK = 0.25 mA) as shown in Figure 3–16(b).

$$I_T = \frac{V_{IN} - V_{OUT}}{R} = \frac{24V - 14.76V}{130\ \Omega} = 71.0\ mA$$

$$I_L = I_T - I_{ZK} = 71.0 \text{ mA} - 0.25 \text{ mA} = 70.75 \text{ mA}$$

$$R_L = \frac{V_{OUT}}{I_L} = \frac{14.76V}{70.75\,mA} = 209\ \Omega$$

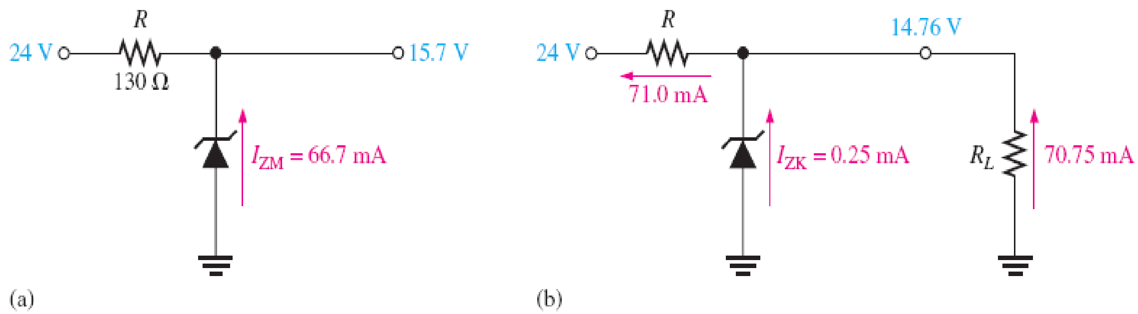


Figure 3–16

***Related Problem*:**

Repeat each part of the preceding analysis if the zener is changed to a 1N4742A 12 V device.

You have seen how the zener diode regulates voltage. Its regulating ability is somewhat limited by the change in zener voltage over a range of current values, which restricts the load current that it can handle. To achieve better regulation and provide for greater variations in load current, the zener diode is combined as a key element with other circuit components to create a 3-terminal linear voltage regulator. Three-terminal voltage regulators that were introduced in Chapter 2 are IC devices that use the zener to provide a reference voltage for an internal amplifier. For a given dc input voltage, the 3-terminal regulator maintains an essentially constant dc voltage over a range of input voltages and load currents. The dc output voltage is always less than the input voltage. Figure 3–17 illustrates a basic 3-terminal regulator showing where the zener diode is used.

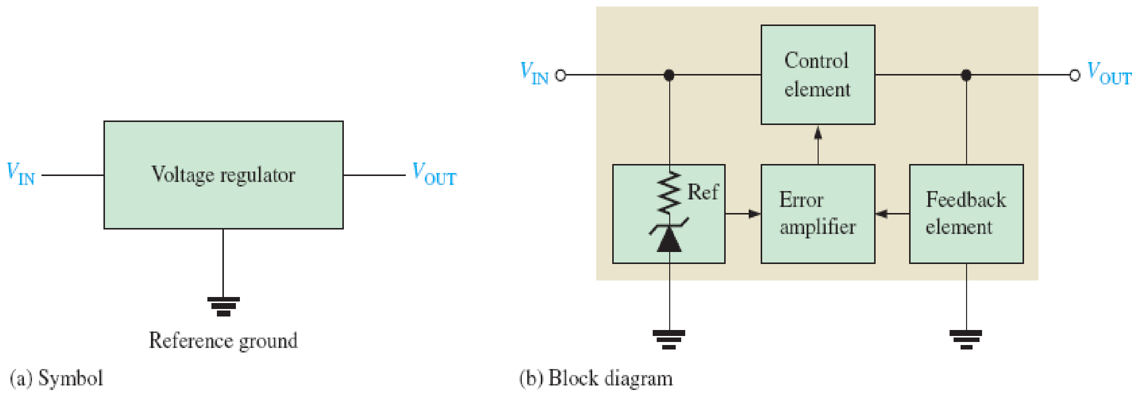


Figure 3–17: Three-terminal voltage regulators.

Zener Limiter

In addition to voltage regulation applications, zener diodes can be used in ac applications to limit voltage swings to desired levels. Figure 3–18 shows three basic ways the limiting action of a zener diode can be used. Part (a) shows a zener used to limit the positive peak of a signal voltage to the selected zener voltage. During the negative alternation, the zener acts as a forward-biased diode and limits the negative voltage to -0.7 V. When the zener is turned around, as in part (b), the negative peak is limited by zener action and the positive voltage is limited to +0.7 V. Two back-to-back zeners limit both peaks to the zener voltage ±0.7 V, as shown in part (c). During the positive alternation, D_2 is functioning as the zener limiter and D_1 is functioning as a forward-biased diode. During the negative alternation, the roles are reversed.

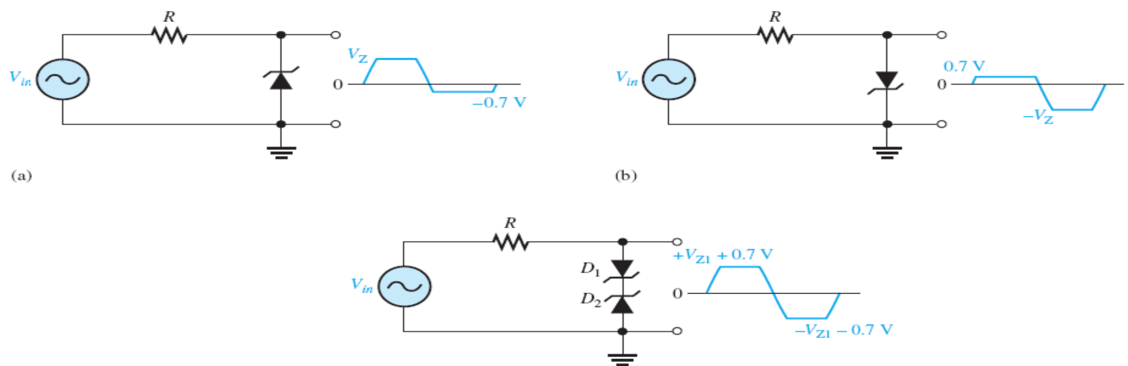


Figure 3–18: Basic zener limiting action with a sinusoidal input voltage.

Example 3–8:

Determine the output voltage for each zener limiting circuit in Figure 3–19.

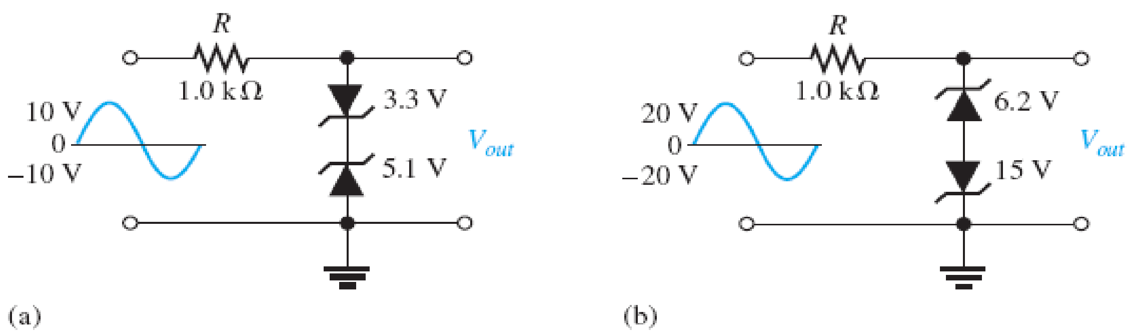


Figure 3–19

Solution :

See Figure 3–20 for the resulting output voltages. Remember, when one zener is operating in breakdown, the other one is forward-biased with approximately 0.7 V across it.

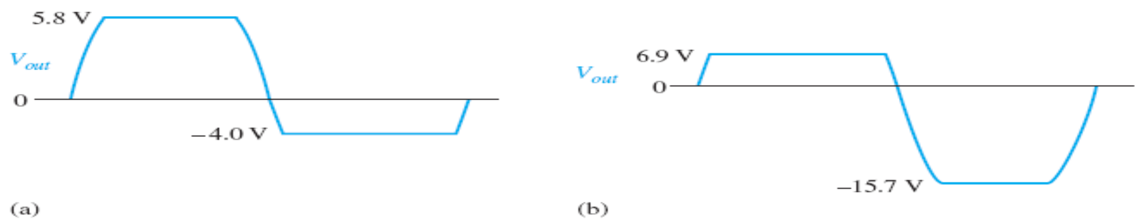


Figure 3–20

***Related Problem*:**

(a) What is the output in Figure 3–19(a) if the input voltage is increased to a peak value of 20 V?

(b) What is the output in Figure 3–19(b) if the input voltage is decreased to a peak value of 5 V?

SECTION 3–2 CHECKUP

1. In a zener diode regulator, what value of load resistance results in the maximum zener current?
2. Explain the terms *no load* and *full load*.
3. How much voltage appears across a zener diode when it is forward-biased?

3–3 The varactor diode

The junction capacitance of diodes varies with the amount of reverse bias. Varactor diodes are specially designed to take advantage of this characteristic and are used as voltage-controlled capacitors rather than traditional diodes. These devices are commonly used in communication systems. Varactor diodes are also referred to as *varicaps* or *tuning diodes*.

A **varactor** is a diode that always operates in reverse bias and is doped to maximize the inherent capacitance of the depletion region. The depletion region acts as a capacitor dielectric because of its nonconductive characteristic. The *p* and *n* regions are conductive and act as the capacitor plates, as illustrated in Figure 3–21.

***Basic Operation*:**

Recall that capacitance is determined by the parameters of plate area (A), dielectric constant (ε), and plate separation (d), as expressed in the following formula:

$$C = \frac{A\,\varepsilon}{d}$$

As the reverse-bias voltage increases, the depletion region widens, effectively increasing the plate separation, thus decreasing the capacitance. When the reverse-bias voltage decreases, the depletion region narrows, thus increasing the capacitance. This action is shown in Figure 3–22(a) and (b). A graph of diode capacitance (C_T) versus reverse voltage for a certain varactor is shown in Figure 3–22(c). For this particular device, C_T varies from 30 pF to slightly less than 4 pF as V_R varies from 1 V to 30 V.

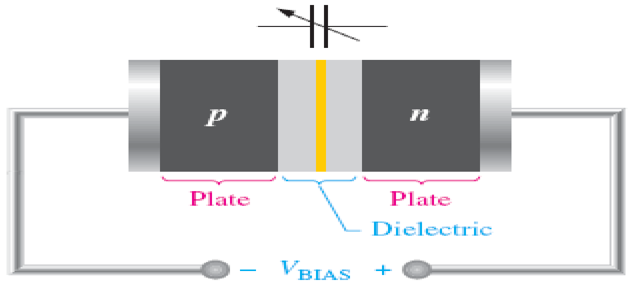


Figure 3–21: The reverse-biased varactor diode acts as a variable capacitor.

In a varactor diode, these capacitance parameters are controlled by the method of doping near the *pn* junction and the size and geometry of the diode's construction. Nominal varactor capacitances are typically available from a few picofarads to several hundred picofarads. Figure 3–23 shows a common symbol for a varactor.

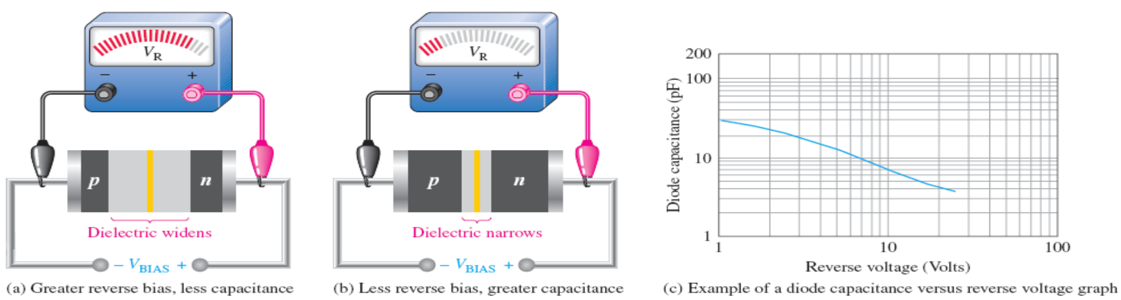


Figure 3–22: Varactor diode capacitance varies with reverse voltage.

Figure 3–23: Varactor diode symbol.

***<u>Capacitance Ratio</u>*:** The varactor **capacitance ratio** is also known as the *tuning ratio.* It is the ratio of the diode capacitance at a minimum reverse voltage to the diode capacitance at a maximum reverse voltage. For the varactor diodes represented in Figure 3–24, the capacitance ratio is the ratio of C measured at a V_R of 2 V divided by C measured at a V_R of 20 V. The capacitance ratio is designated as C_2/C_{20} in this case.

For the 832A, the minimum capacitance ratio is 5.0. This means that the capacitance value decreases by a factor of 5.0 as V_R is increased from 2 V to 20 V. The following calculation illustrates how to use the capacitance ratio (CR) to find the capacitance range for the 832A. If C_2 = 22 pF and the minimum $CR = C_2/C_{20} = 5.0$,

$$C_{20} = \frac{C_2}{CR} = \frac{22\ pF}{5} = 4.4\ pF$$

The diode capacitance varies from 22 pF to 4.4 pF when V_R is increased from 2 V to 20 V. The Zetex 830 series of varactor diodes are hyper-abrupt junction devices. The doping in the *n* and *p* regions is made uniform so that at the *pn* junction there is a very abrupt change from *n* to *p* instead of the more gradual change found in the rectifier diodes. The abruptness of the *pn* junction determines the capacitance ratio.

***Back-to-Back Configuration*:** One of the drawbacks of using just a single varactor diode in certain applications, such as rf tuning, is that if the diode is forward-biased by the rf signal during part of the ac cycle, its reverse leakage will increase momentarily. Also, a type of distortion called *harmonic distortion* is produced if the varactor is alternately biased positively and negatively. To avoid harmonic distortion, you will often see two varactor diodes back to back, as shown in Figure 3–25(a) with the reverse dc voltage applied to both devices simultaneously. The two tuning diodes will be driven alternately into high and low capacitance, and the net capacitance will remain constant and is unaffected by the rf signal amplitude. The Zetex 832A varactor diode is available in a back-to-back configuration in an SOT23 package or as a single diode in an SOD523 package, as shown in Figure 3–25(b). Although the cathodes in the back-to-back configuration are connected to a common pin, each diode can also be used individually.

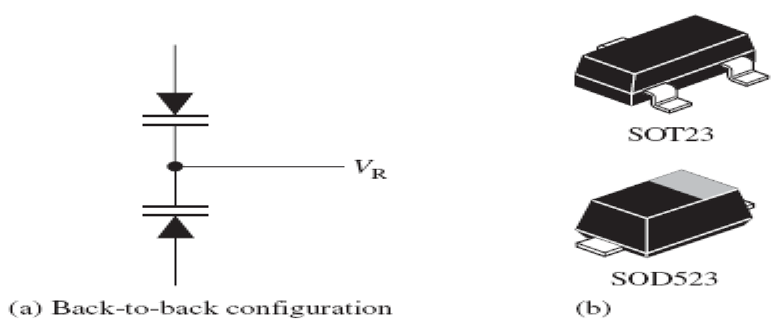


Figure 3–25: Varactor diodes and typical packages.

An Application:

A major application of varactors is in tuning circuits. For example, VHF, UHF, and satellite receivers utilize varactors. Varactors are also used in cellular communications. When used in a parallel resonant circuit, as illustrated in Figure 3–26, the varactor acts as a variable capacitor, thus allowing the resonant frequency to be adjusted by a variable voltage level. The varactor diode provides the total variable capacitance in the parallel resonant band-pass filter. The varactor diode and the inductor form a parallel resonant circuit from the output to ac ground. The capacitors C_1 and C_2 have no effect on the filter's frequency response because their reactances are negligible at the resonant frequencies. C_1 prevents a dc path from the potentiometer wiper back to the ac source through the inductor and R_1. C_2 prevents a dc path from the wiper of the potentiometer to a load on the output. The potentiometer R_2 forms a variable dc voltage for biasing the varactor. The reverse-bias voltage across the varactor can be varied with the potentiometer.

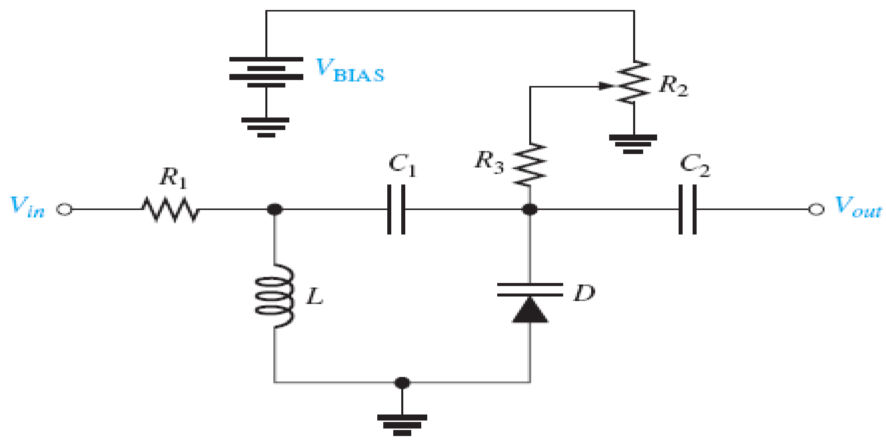


Figure 3–26: A resonant band-pass filter using a varactor diode for adjusting the resonant frequency over a specified range.

Recall that the parallel resonant frequency is

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

<u>Example 3–8</u>:

(a) Given that the capacitance of a Zetex 832A varactor is approximately 40 pF at 0 V bias and that the capacitance at a 2 V reverse bias is 22 pF, determine the capacitance at a reverse bias of 20 V using the specified minimum capacitance ratio.

(b) Using the capacitances at bias voltages of 0 V and 20 V, calculate the resonant frequencies at the bias extremes for the circuit in Figure 3–26 if L = 2 mH.

(c) Verify the frequency calculations by simulating the circuit in Figure 3–26 for the following component values: R_1 = 47 kΩ, R_2 = 10 kΩ, R_3 = 5.1 MΩ, C_1 = 10 nF, C_2 = 10 nF, L = 2 mH, and V_{BIAS} = 20 V.

***<u>Solution</u>*:**

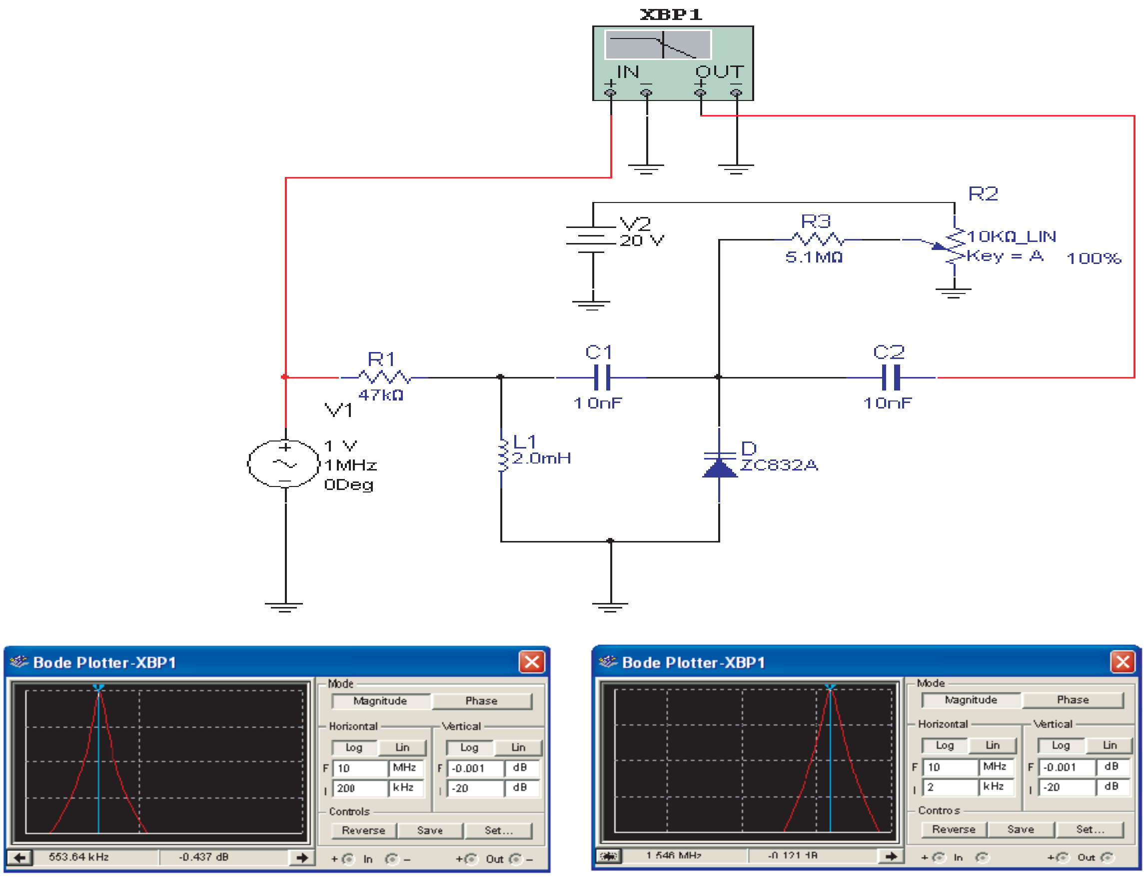
(a) $$C_{20} = \frac{C_2}{CR} = \frac{22\ pF}{5} = 4.4\ pF$$

(b) $$f_o = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(2mH)(40pF)}} = 563\ kHz$$

$$f_o = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(2mH)(4.4pF)}} = 1.7\ MHz$$

(C) The Multsim simulation of the circuit is shown in Figure 3–27. The Bode plotters show the frequency responses at 0 V and 20 V reverse bias. The center of the 0 V bias response curve is at 553.64 kHz and the center of the 20 V bias response curve is at 1.548 MHz. These results agree reasonably well with the calculated values.

These results show that this circuit can be tuned over most of the AM broadcast band.



Frequency response for 0 V varactor bias

Frequency response for 20 V reverse varactor bias

Figure 3–27: Multisim simulation.

Related Problem :

How could you increase the tuning range of the circuit?

SECTION 3–3 CHECKUP

1. What is the key feature of a varactor diode?
2. Under what bias condition is a varactor operated?
3. What part of the varactor produces the capacitance?
4. Based on the graph in Figure 3–22(c), what happens to the diode capacitance when the reverse voltage is increased?
5. Define *capacitance ratio*.

3–4 Optical diodes

In this section, three types of optoelectronic devices are introduced: the light-emitting diode, quantum dots, and the photodiode. As the name implies, the light-emitting diode is a light emitter. Quantum dots are very tiny light emitters made from silicon with great promise for various devices, including light-emitting diodes. On the other hand, the photodiode is a light detector.

The Light-Emitting Diode (LED)

The symbol for an LED is shown in Figure 3–28.

The basic operation of the **light-emitting diode (LED)** is as follows. When the device is forward-biased, electrons cross the *pn* junction from the *n*-type material and recombine with holes in the *p*-type material. Recall from Chapter 1 that these free electrons are in the conduction band and at a higher energy than the holes in the valence band. The difference in energy between the electrons and the holes corresponds to the energy of visible light. When recombination takes place, the recombining electrons release energy in the form of **photons.** The emitted light tends to be monochromatic (one color) that depends on the band gap (and other factors). A large exposed surface area on one layer of the semiconductive material permits the photons to be emitted as visible light. This process, called **electroluminescence**, is illustrated in Figure 3–29. Various impurities are added during the doping process to establish the **wavelength** of the emitted light. The wavelength determines the color of visible light. Some LEDs emit photons that are not part of the visible spectrum but have longer wavelengths and are in the **infrared** (IR) portion of the spectrum.

LED Semiconductor Materials: The semiconductor gallium arsenide (GaAs) was used in early LEDs and emits IR radiation, which is invisible. The first visible red LEDs were produced using gallium arsenide phosphide (GaAsP) on a GaAs substrate. The efficiency was increased using a gallium phosphide (GaP) substrate, resulting in brighter red LEDs and also allowing orange LEDs. Later, GaP was used as the light-emitter to achieve pale green light. By using a red and a green chip, LEDs were able to produce yellow light. The first super-bright red, yellow, and green LEDs were produced using gallium aluminum arsenide phosphide (GaAlAsP). By the early 1990s ultrabright LEDs using indium gallium aluminum phosphide (InGaAlP) were available in red, orange, yellow, and green.

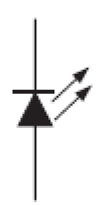


Figure 3–28: Symbol for an LED. When forwardbiased, it emits light.

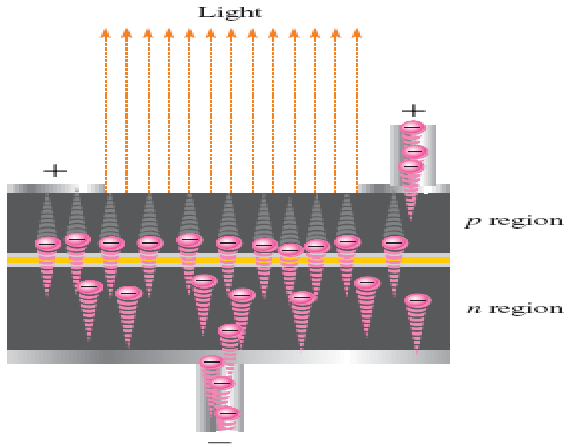


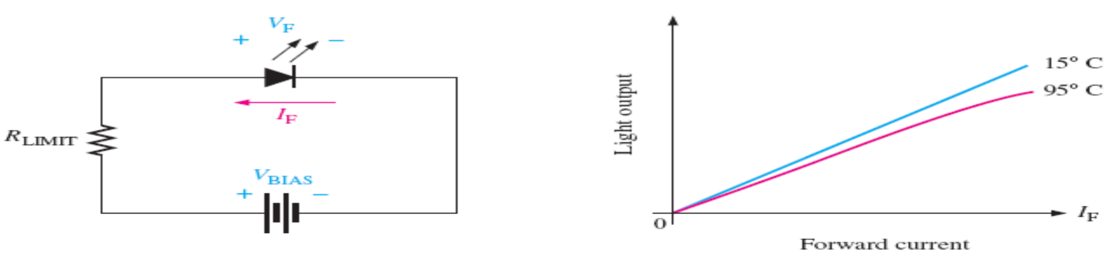
Figure 3–29: Electroluminescence in a forwardbiased LED.

Blue LEDs using silicon carbide (SiC) and ultrabright blue LEDs made of gallium nitride (GaN) became available. High intensity LEDs that produce green and blue are also made using indium gallium nitride (InGaN). High-intensity white LEDs are formed using ultrabright blue GaN coated with fluorescent phosphors that absorb the blue light and reemit it as white light.

<u>LED Biasing</u>**:** The forward voltage across an LED is considerably greater than for a silicon diode. Typically, the maximum V_F for LEDs is between 1.2 V and 3.2 V, depending on the material. Reverse breakdown for an LED is much less than for a silicon rectifier diode (3 V to 10 V is typical).

The LED emits light in response to a sufficient forward current, as shown in Figure 3–30(a).

The amount of power output translated into light is directly proportional to the forward current, as indicated in Figure 3–30(b). An increase in I_F corresponds proportionally to an increase in light output. The light output (both intensity and color) is also dependent on temperature. Light intensity goes down with higher temperature as indicated in the figure.



(a) Forward-biased operation

(b) General light output versus forward current for two temperatures

Figure 3–30: Basic operation of an LED.

<u>Light Emission</u>**:** An LED emits light over a specified range of wavelengths as indicated by the **spectral** output curves in Figure 3–31. The curves in part (a) represent the light output versus wavelength for typical visible LEDs, and the curve in part (b) is for a typical infrared LED. The wavelength is expressed in nanometers (nm). The normalized output of the visible red LED peaks at 660 nm, the yellow at 590 nm, green at 540 nm, and blue at 460 nm. The output for the infrared LED peaks at 940 nm.

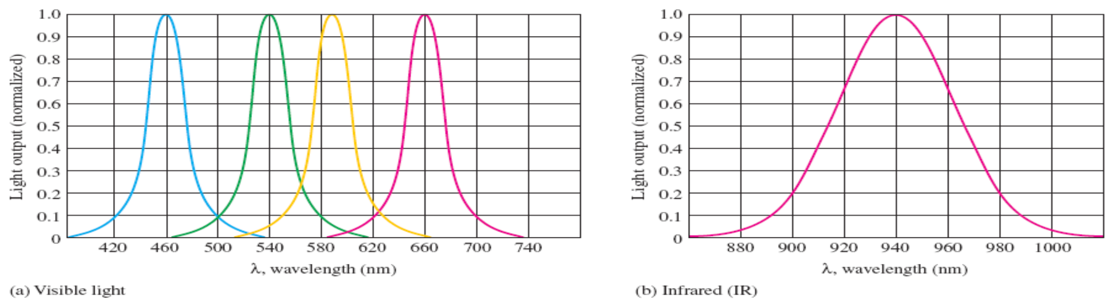


Figure 3–31: Examples of typical spectral output curves for LEDs.

The graphs in Figure 3–32 show typical **radiation** patterns for small LEDs. LEDs are directional light sources (unlike filament or fluorescent bulbs). The radiation pattern is generally perpendicular to the emitting surface; however, it can be altered by the shape of the emitter surface and by lenses and diffusion films to favor a specific direction. Directional patterns can be an advantage for certain applications, such as traffic lights, where the light is intended to be seen only by certain drivers. Figure 3–32(a) shows the pattern

for a forward-directed LED such as used in small panel indicators. Figure 3–32(b) shows the pattern for a wider viewing angle such as found in many super-bright LEDs. A wide variety of patterns are available from manufacturers; one variation is to design the LED to emit nearly all the light to the side in two lobes. Typical small LEDs for indicators are shown in Figure 3–33(a). In addition to small LEDs for indicators, bright LEDs are becoming popular for lighting because of their superior efficiency and long life. A typical LED for lighting can deliver 50–60 lumens per watt, which is approximately five times greater efficiency than a standard incandescent bulb. LEDs for lighting are available in a variety of configurations, including even flexible tubes for decorative lighting and low-wattage bulbs for outdoor walkways and gardens.

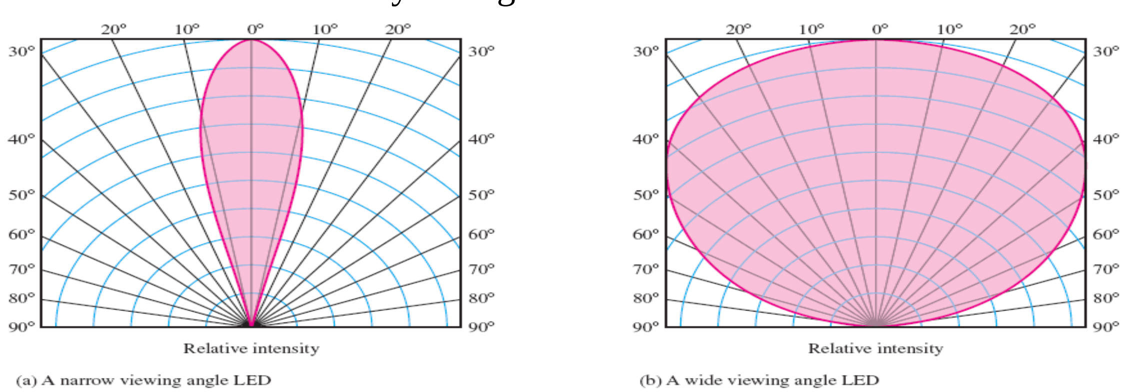


Figure 3–32: Radiation patterns for two different LEDs.

Applications

Standard LEDs are used for indicator lamps and readout displays on a wide variety of instruments, ranging from consumer appliances to scientific apparatus. A common type of display device using LEDs is the seven-segment display. Combinations of the segments form the ten decimal digits as illustrated in Figure 3–34. Each segment in the display is an LED. By forward-biasing selected combinations of segments, any decimal digit and a decimal point can be formed. Two types of LED circuit arrangements are the common anode and common cathode as shown.

One common application of an infrared LED is in remote control units for TV, DVD, gate openers, etc. The IR LED sends out a beam of invisible light that is sensed by the receiver in your TV, for example. For each button on the remote control unit, there is a unique code. When a specific button is pressed, a coded electrical signal is generated that goes to the LED, which

converts the electrical signal to a coded infrared light signal. The TV receiver recognizes the code and takes appropriate action, such as changing the channel or increasing the volume.

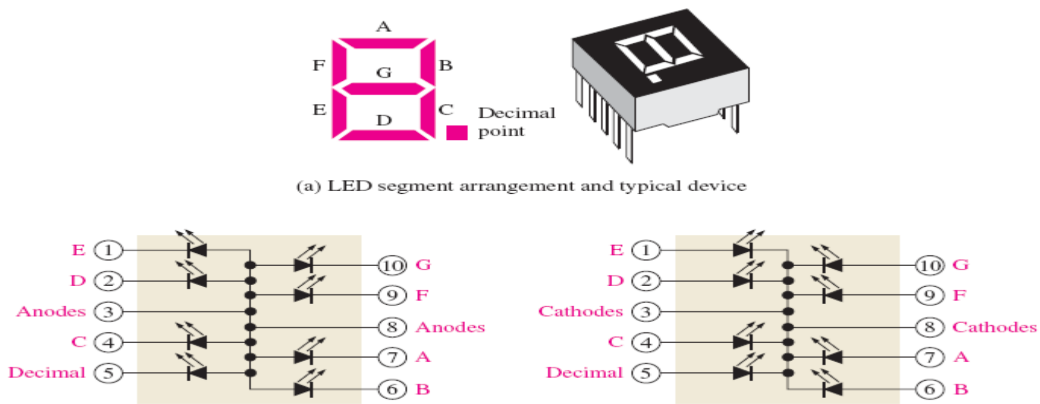


Figure 3–34: The 7-segment LED display.

Also, IR light-emitting diodes are used in optical coupling applications, often in conjunction with fiber optics. Areas of application include industrial processing and control, position encoders, bar graph readers, and optical switching. An example of how an IR LED could be used in an industrial application is illustrated in Figure 3–35. This particular system is used to count baseballs as they are fed down a chute into a box for shipping. As each ball passes through the chute, the IR beam emitted by the LED is interrupted. This is detected by the photodiode (discussed later) and the resulting change in current is sensed by a detector circuit. An electronic circuit counts each time that the beam is interrupted; and when a preset number of balls pass through the chute, the “stop” mechanism is activated to stop the flow of balls until the next empty box is automatically moved into place on the conveyor. When the next box is in place, the "stop" mechanism is deactivated and the balls begin to roll again. This idea can also be applied to inventory and packing control for many other types of products.

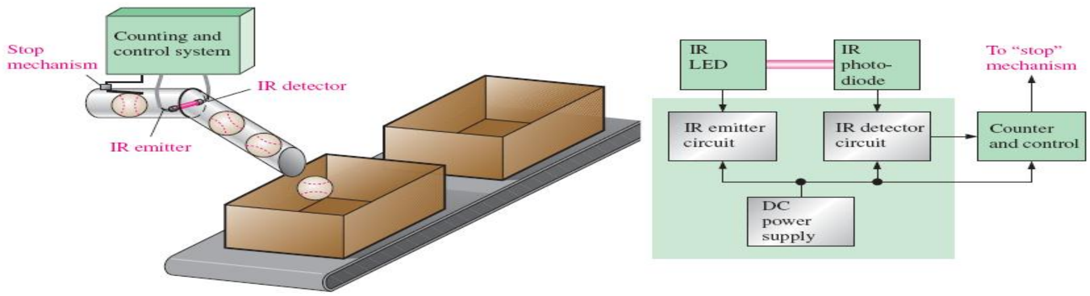


Figure 3–35: Basic concept and block diagram of a counting and control system.

__Traffic Lights__**:** LEDs are quickly replacing the traditional incandescent bulbs in traffic signal applications. Arrays of tiny LEDs form the red, yellow, and green lights in a traffic light unit. An LED array has three major advantages over the incandescent bulb: brighter light, longer lifetime (years vs. months), and less energy consumption (about 90% less). LED traffic lights are constructed in arrays with lenses that optimize and direct the light output. Figure 3–36(a) illustrates the concept of a traffic light array using red LEDs. A relatively low density of LEDs is shown for illustration. The actual number and spacing of the LEDs in a traffic light unit depends on the diameter of the unit, the type of lens, the color, and the required light intensity. With an appropriate LED density and a lens, an 8- or 12- inch traffic light will appear essentially as a solid-color circle. LEDs in an array are usually connected either in a series-parallel or a parallel arrangement.

A series connection is not practical because if one LED fails open, then all the LEDs are disabled. For a parallel connection, each LED requires a limiting resistor. To reduce the number of limiting resistors, a series-parallel connection can be used, as shown in Figure 3–36(b).

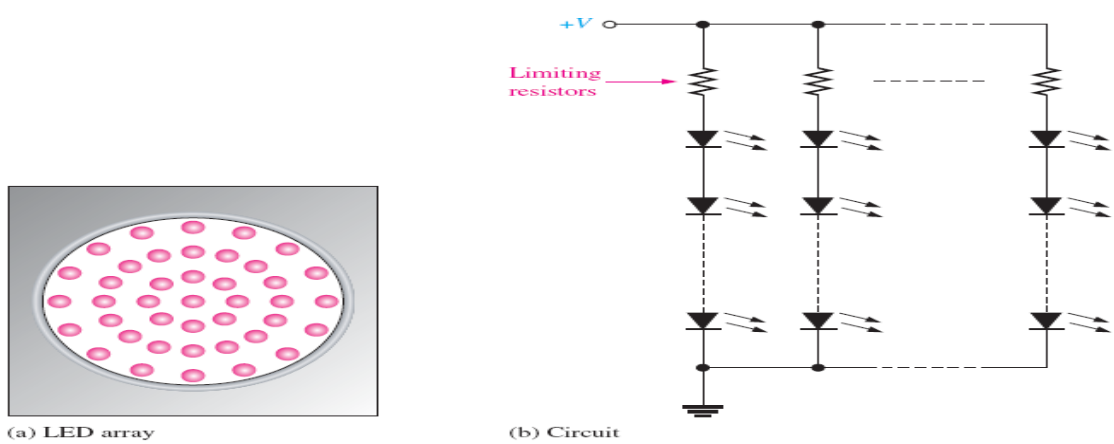


Figure 3–36: LED traffic light.

***LED Displays*:** LEDs are widely used in large and small signs and message boards for both indoor and outdoor uses, including large-screen television. Signs can be single-color, multicolor, or full-color. Full-color screens use a tiny grouping of high-intensity red, green, and blue LEDs to form a **pixel**. A typical screen is made of thousands of RGB pixels with the exact number determined by the sizes of the screen and the pixel. Red, green, and blue (RGB) are primary colors and when mixed together in varying amounts, can be used to produce any color in the visible spectrum. A basic pixel formed by three LEDs is shown in Figure 3–37. The light emission from each of the three diodes can be varied independently by varying the amount of forward current. Yellow is added to the three primary colors (RGBY) in some TV screen applications.

***Other Applications*:** High-intensity LEDs are becoming more widely used in automotive lighting for taillights, brakelights, turn signals, back-up lights, and interior applications. LED arrays are expected to replace most incandescent bulbs in automotive lighting. Eventually, headlights may also be replaced by white LED arrays. LEDs can be seen better in poor weather and can last 100 times longer than an incandescent bulb. LEDs are also finding their way into interior home and business lighting applications.

Arrays of white LEDs may eventually replace incandescent light bulbs and fluorescent lighting in interior living and work areas. As previously mentioned, most white LEDs use a blue GaN (gallium nitride) LED covered by a yellowish phosphor coating made of a certain type of crystals that have been powdered and bound in a type of viscous adhesive. Since yellow light stimulates the red and green receptors of the eye, the resulting mix of blue and yellow light gives the appearance of white.

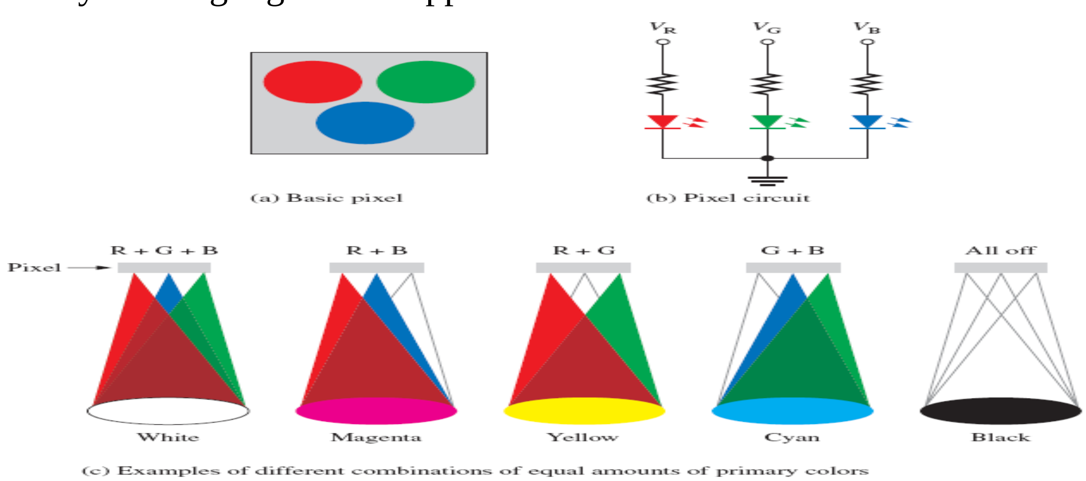


Figure 3–37: The concept of an RGB pixel used in LED display screens.

The Organic LED (OLED)

An **OLED** is a device that consists of two or three layers of materials composed of organic molecules or polymers that emit light with the application of voltage. OLEDs produce light through the process of electrophosphorescence. The color of the light depends on the type of organic molecule in the emissive layer. The basic structure of a 2-layer OLED is shown in Figure 3–38.

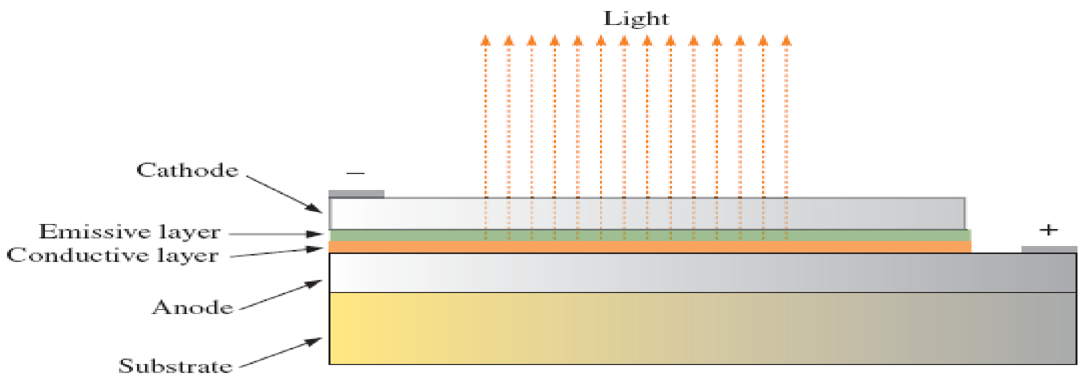


Figure 3–38: Basic structure of a top-emitting 2-layer OLED.

Electrons are provided to the emissive layer and removed from the conductive layer when there is current between the cathode and anode. This removal of electrons from the conductive layer leaves holes. The electrons from the emissive layer recombine with the holes from the conductive layer near the junction of the two layers. When this recombination occurs, energy is released in the form of light that passes through the transparent cathode material. If the anode and substrate are also made from transparent materials, light is emitted in both directions, making the OLED useful in applications such as heads-up displays. OLEDs can be sprayed onto substrates just like inks are sprayed onto paper during printing. Inkjet technology greatly reduces the cost of OLED manufacturing and allows OLEDs to be printed onto very large films for large displays like 80-inch TV screens or electronic billboards.

Quantum Dots:

Quantum dots are a form of nanocrystals that are made from semiconductor material such as silicon, germanium, cadmium sulfide, cadmium selenide, and indium phosphide. Quantum dots are only 1 nm to 12 nm in diameter (a nm is one billionth of a meter). Billions of dots could fit on the head of a pin! Because of their small size, quantum effects arise due to the confinement of electrons and holes; as a result, material properties are very different than the normal material. One important property is that the band gap is dependent on the size of the dots. When excited from an external source, dots formed from semiconductors emit light in the visible range as well as infrared and ultraviolet, depending on their size. The higher-frequency blue light is emitted by smaller dots suspended in solution (larger band gap); red light is emitted from solutions with larger dots (smaller band gap). Solutions containing the quantum dots glow eerily with specific colors as shown in the photograph in Figure 3–39.

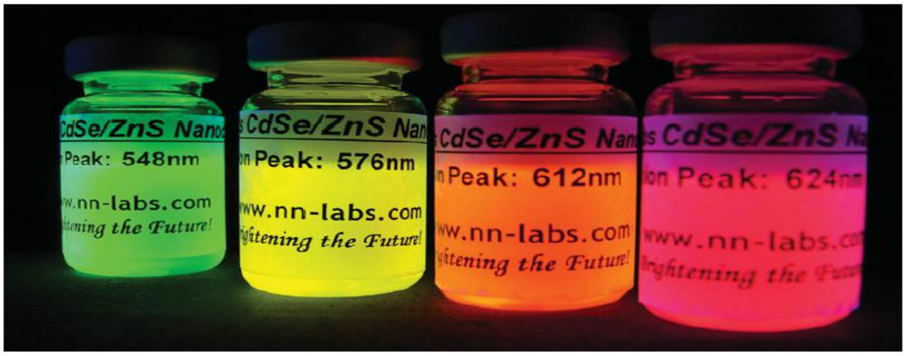


Figure 3–39: Solutions containing quantum dots glow with specific colors that depend on the size of the dots. Courtesy of NN-Labs.

Although quantum dots are not diodes themselves, they can be used in construction of light-emitting diodes as well as display devices and a variety of other applications. As you know, LEDs work by generating a specific frequency (color) of light, which is determined by the band gap. To produce white light, blue LEDs are coated with a phosphor that adds yellow light to the blue, forming white. The result is not a pure white, but tends to be harsh

and makes colors appear unnatural. While this is satisfactory for displays and signs, many people do not like it for home lighting.

Quantum dots can be used to modify the basic color of LEDs by converting higher energy photons (blue) to photons of lower energy. The result is a color that more closely approximates an incandescent bulb. Quantum dot filters can be designed to contain combinations of colors, giving designers control of the spectrum. The important advantage of quantum dot technology

is that it does not lose the incoming light; it merely absorbs the light and reradiates it at a different frequency. This enables control of color without giving up efficiency. By placing a quantum dot filter in front of a white LED, the spectrum can be made to look like that of an incandescent bulb. The resulting light is more satisfactory for general illumination, while retaining the advantages of LEDs. There are other promising applications, particularly in medical applications. Water-soluble quantum dots are used as a biochemical luminescent marker for cellular imaging and medical research. Research is also being done on quantum dots as the basic device units for information processing by manipulating two energy levels within the quantum dot.

<u>The Photodiode:</u>

The **photodiode** is a device that operates in reverse bias, as shown in Figure 3–40(a), where is the reverse light current. The photodiode has a small transparent window that allows light to strike the *pn* junction. Some typical photodiodes are shown in Figure 3–40(b). An alternate photodiode symbol is shown in Figure 3–40(c).

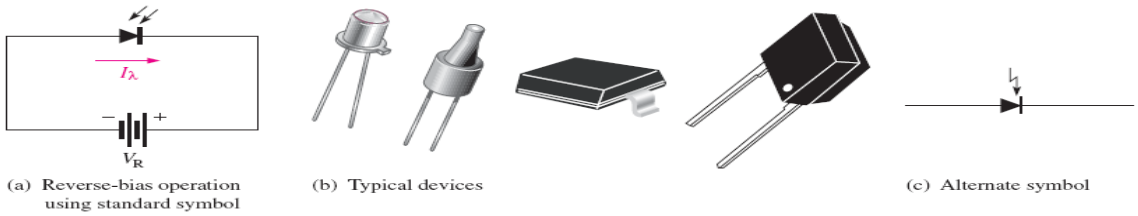


Figure 3–40: Photodiode.

Recall that when reverse-biased, a rectifier diode has a very small reverse leakage current. The same is true for a photodiode. The reverse-biased current is produced by thermally generated electron-hole pairs in the depletion region, which are swept across the *pn* junction by the electric field created by the reverse voltage. In a rectifier diode, the reverse leakage current increases with temperature due to an increase in the number of electron-hole pairs. A photodiode differs from a rectifier diode in that when its *pn* junction is exposed to light, the reverse current increases with the light intensity. When there is no incident light, the reverse current, I_λ, is almost negligible and is called the **dark current.** An increase in the amount of light intensity, expressed as irradiance (mW/cm2), produces an increase in the reverse current, as shown by the graph in Figure 3–41(a). From the graph in Figure 3–41(b), you can see that the reverse current for this particular device is approximately 1.4 µA at a reverse-bias voltage of 10 V with an irradiance of 0.5 mW/cm2. Therefore, the resistance of the device is

$$R_R = \frac{V_R}{I_\lambda} = \frac{10V}{1.4\ \mu A} = 7.14\ M\Omega$$

At 20 mW/cm2, the current is approximately The resistance under this condition is

$$R_R = \frac{V_R}{I_\lambda} = \frac{10V}{55\ \mu A} = 182\ k\Omega$$

These calculations show that the photodiode can be used as a variable-resistance device controlled by light intensity.

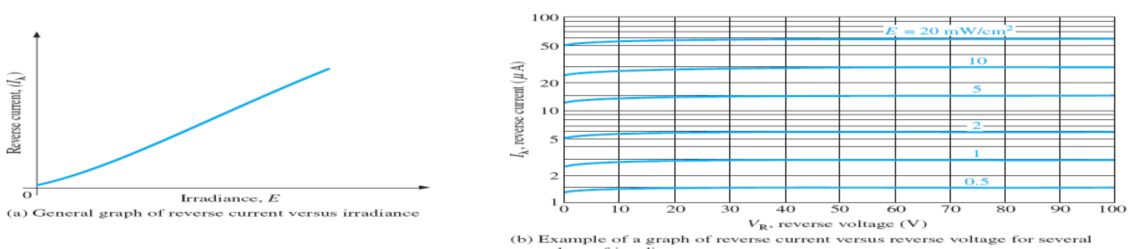


Figure 3–41: Typical photodiode characteristics.

Figure 3–42 illustrates that the photodiode allows essentially no reverse current (except for a very small dark current) when there is no incident light. When a light beam strikes the photodiode, it conducts an amount of reverse current that is proportional to the light intensity (irradiance).

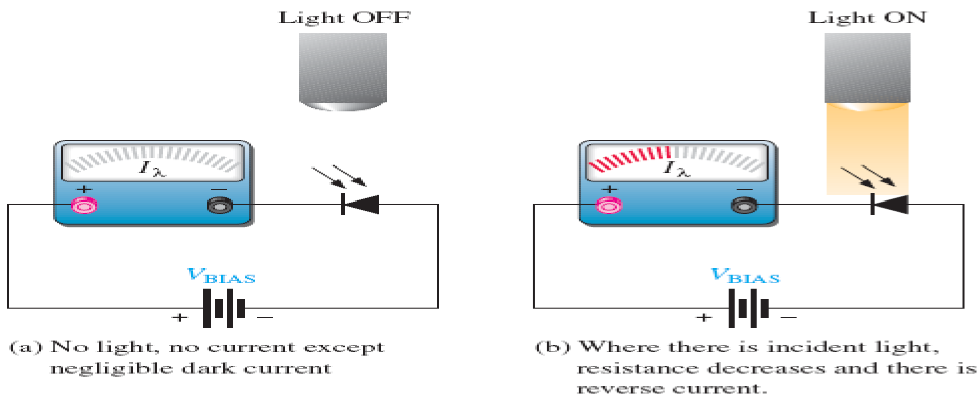


Figure 3–42: Operation of a photodiode.

Example 3–12:

For a TEMD1000 photodiode,

(a) Determine the maximum dark current for $V_R = 10$ V.

(b) Determine the reverse light current for an irradiance of 1 mW/cm^2 at a wavelength of 850 nm if the device angle is oriented at 10°with respect to the maximum irradiance and the reverse voltage is 5 V.

***Solution*:**

(a) From Figure 3–43(a), the maximum dark current I_{ro} = **10 nA.**

(b) From the graph in Figure 3–43(d), the reverse light current is 12 µA at 950 nm. From Figure 3–43(b), the relative sensitivity is 0.6 at 850 nm. Therefore, the reverse light current is

$$I_\lambda = I_{ra} = 0.6(12\ \mu A) = 72\ \mu A$$

For an angle of the relative sensitivity is reduced to 0.92 of its value at 0°.

$$I_\lambda = I_{ra} = 0.92\ (7.2\ \mu A) = \mathbf{6.62}\ \mu\mathbf{A}$$

***Related Problem*:**

What is the reverse current if the wavelength is 1050 nm and the angle is 0°?

SECTION 3–4 CHECKUP

1. Name two types of LEDs in terms of their light-emission spectrum.
2. Which has the greater wavelength, visible light or infrared?
3. In what bias condition is an LED normally operated?
4. What happens to the light emission of an LED as the forward current increases?
5. The forward voltage drop of an LED is 0.7 V. (true or false)
6. What is a pixel?
7. In what bias condition is a photodiode normally operated?
8. When the intensity of the incident light (irradiance) on a photodiode increases, what happens to its internal reverse resistance?
9. What is dark current?

3–5 Other types of diodes

In this section, several types of diodes that you are less likely to encounter as a technician but are nevertheless important are introduced. Among these are the laser diode, the Schottky diode, the *pin* diode, the step-recovery diode, the tunnel diode, and the current regulator diode.

The Laser Diode

The term **laser** stands for *l*ight *a*mplification by *s*timulated *e*mission of *r*adiation. Laser light is **monochromatic,** which means that it consists of a single color and not a mixture of colors. Laser light is also called **coherent light,** a single wavelength, as compared to incoherent light, which consists of a wide band of wavelengths. The laser diode normally emits coherent light, whereas the LED emits incoherent light. The symbols are the same as

shown in Figure 3–44(a).

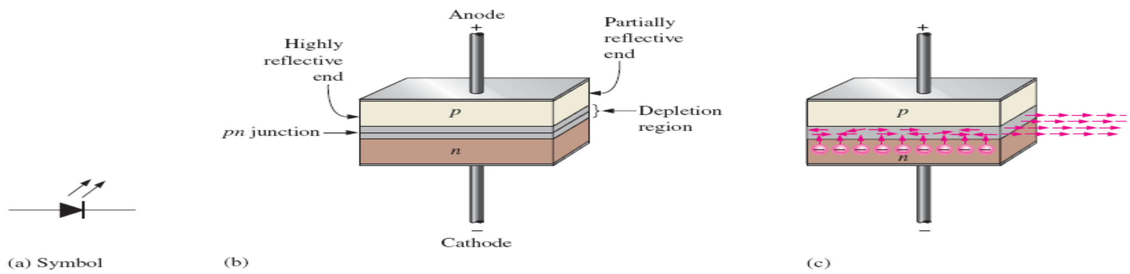


Figure 3–44: Basic laser diode construction and operation.

The basic construction of a laser diode is shown in Figure 3–44(b). A *pn* junction is formed by two layers of doped gallium arsenide. The length of the *pn* junction bears a precise relationship with the wavelength of the light to be emitted. There is a highly reflective surface at one end of the *pn* junction and a partially reflective surface at the other end, forming a resonant cavity for the photons. External leads provide the anode and cathode connections. The basic operation is as follows. The laser diode is forward-biased by an external voltage source. As electrons move through the junction, recombination occurs just as in an ordinary diode. As electrons fall into holes to recombine, photons are released. A released photon can strike an atom, causing another photon to be released. As the forward current is increased, more electrons enter the depletion region and cause more photons to be emitted. Eventually some of the photons that are randomly drifting within the depletion region strike the reflected surfaces perpendicularly. These reflected photons move along the depletion region, striking atoms and releasing additional photons due to the avalanche effect. This back-and-forth movement of photons increases as the generation of photons "snowballs" until a very intense beam of laser light is formed by the photons that pass through the partially reflective end of the *pn* junction. Each photon produced in this process is identical to the other photons in energy level, phase relationship, and frequency. So a single wavelength of intense light emerges from the laser diode, as indicated in Figure 3–44(c). Laser diodes have a threshold level of current above which the laser action occurs and below which the diode behaves essentially as an LED, emitting incoherent light.

***An Application*:** Laser diodes and photodiodes are used in the pick-up system of compact disk (CD) players. Audio information (sound) is digitally recorded in stereo on the surface of a compact disk in the form of microscopic "pits" and "flats." A lens arrangement focuses the laser beam from the diode onto the CD surface. As the CD rotates, the lens and beam follow the track under control of a servomotor. The laser light, which is altered by the pits and flats along the recorded track, is reflected back from the track through a lens and optical system to infrared photodiodes. The signal from the photodiodes is then used to reproduce the digitally recorded sound. Laser diodes are also used in laser printers and fiber-optic systems.

The Schottky Diode:

Schottky diodes are high-current diodes used primarily in high-frequency and fast-switching applications. They are also known as *hot-carrier diodes*. The term *hot-carrier* is derived from the higher energy level of electrons in the *n* region compared to those in the metal region. A Schottky diode symbol is shown in Figure 3–45. A Schottky diode is formed by joining a doped semiconductor region (usually *n*-type) with a metal such as gold, silver, or platinum. Rather than a *pn* junction, there is a metal-to-semiconductor junction, as shown in Figure 3–46. The forward voltage drop is typically around 0.3 V because there is no depletion region as in a *pn* junction diode.

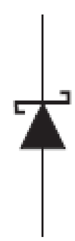


Figure 3–45: Schottky diode symbol.

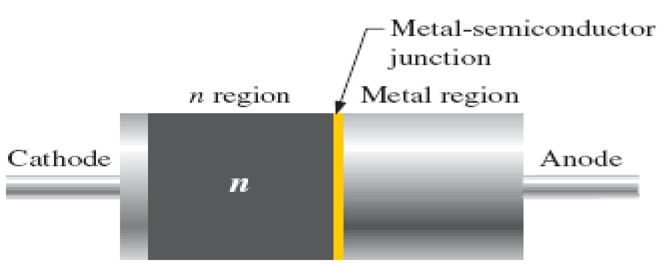


Figure 3–46: Basic internal construction of a Schottky diode.

The Schottky diode operates only with majority carriers. There are no minority carriers and thus no reverse leakage current as in other types of diodes. The metal region is heavily occupied with conduction-band electrons, and the *n*-type semiconductor region is lightly doped. When forward-biased, the higher energy electrons in the *n* region are injected into the metal region where they give up their excess energy very rapidly. Since there are no minority carriers, as in a conventional rectifier diode, there is a very rapid response to a change in bias. The Schottky is a fast-switching diode, and most of its applications make use of this property. It can be used in high-frequency applications and in many digital circuits to decrease switching times. The LS family of TTL logic (LS stands for low-power Schottky) is one type of digital integrated circuit that uses the Schottky diode.

The *PIN* Diode:

The *pin* diode consists of heavily doped *p* and *n* regions separated by an intrinsic (*i*) region, as shown in Figure 3–47(a). When reverse-biased, the *pin* diode acts like a nearly constant capacitance. When forward-biased, it acts like a current-controlled variable resistance. This is shown in Figure 3–47(b) and (c). The low forward resistance of the intrinsic region decreases with increasing current.

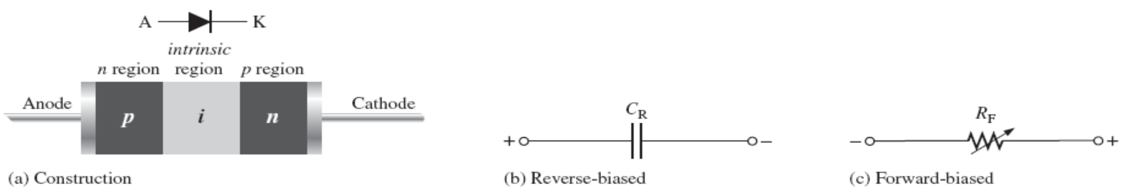


Figure 3–47: *PIN* diode.

The forward series resistance characteristic and the reverse capacitance characteristic are shown graphically in Figure 3–48 for a typical *pin* diode.

The *pin* diode is used as a dc-controlled microwave switch operated by rapid changes in bias or as a modulating device that takes advantage of the variable forward-resistance characteristic. Since no rectification occurs at the *pn* junction, a high-frequency signal can be modulated (varied) by a lower-frequency bias variation. A *pin* diode can also be used in attenuator applications because its resistance can be controlled by the amount of current. Certain types of *pin* diodes are used as photodetectors in fiber-optic systems.

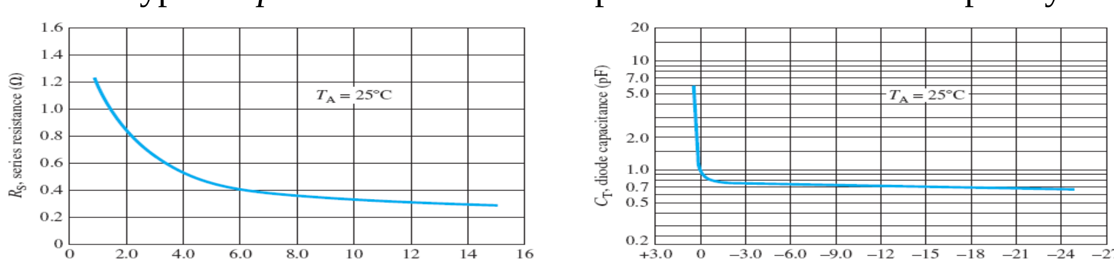


Figure 3–48: *PIN* diode characteristics.

The Step-Recovery Diode:

The step-recovery diode uses graded doping where the doping level of the semiconductive materials is reduced as the *pn* junction is approached. This produces an abrupt turn-off time by allowing a fast release of stored charge when switching from forward to reverse bias. It also allows a rapid re-establishment of forward current when switching from reverse to forward bias. This diode is used in very high frequency (VHF) and fast-switching applications.

The Tunnel Diode:

The tunnel diode exhibits a special characteristic known as *negative resistance.* This feature makes it useful in oscillator and microwave amplifier applications. Two alternate symbols are shown in Figure 3–49. Tunnel diodes are constructed with germanium or gallium arsenide by doping the *p* and *n* regions much more heavily than in a conventional rectifier diode. This heavy doping results in an extremely narrow depletion region. The heavy doping allows conduction for all reverse voltages so that there is no breakdown effect as with the conventional rectifier diode. This is shown in Figure 3–50. Also, the extremely narrow depletion region permits electrons to "tunnel" through the *pn* junction at very low forward-bias voltages, and the diode acts as a conductor. This is shown in Figure 3–50 between points *A* and *B.* At point *B*, the forward voltage begins to develop a barrier, and the current begins to decrease as the forward voltage continues to increase.

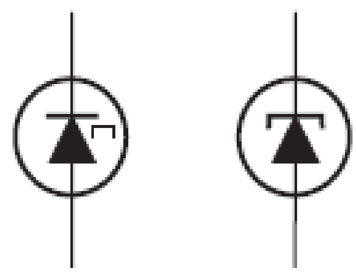


Figure 3–49: Tunnel diode symbols.

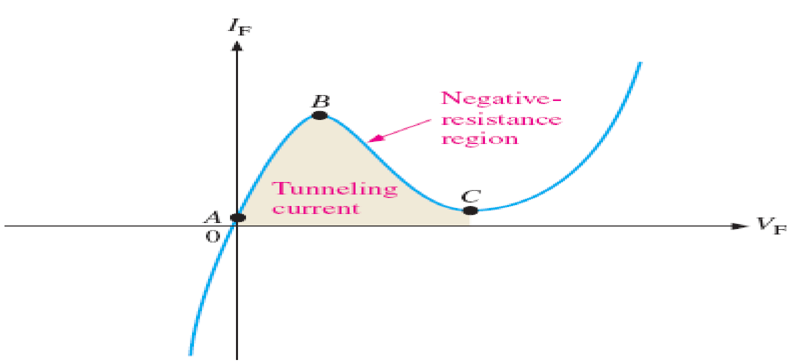


Figure 3–50: Tunnel diode characteristic curve.

This is the *negative-resistance region.*

$$R_F = \frac{\Delta V_F}{\Delta I_F}$$

This effect is opposite to that described in Ohm's law, where an increase in voltage results in an increase in current. At point *C*, the diode begins to act as a conventional forward-biased diode.

An Application: A parallel resonant circuit can be represented by a capacitance, inductance, and resistance in parallel, as in Figure 3–51(a).

R_P is the parallel equivalent of the series winding resistance of the coil. When the tank circuit is "shocked" into oscillation by an application of voltage as in Figure 3–51(b), a damped sinusoidal output results. The damping is due to the resistance of the tank, which prevents a sustained oscillation because energy is lost when there is current through the resistance.

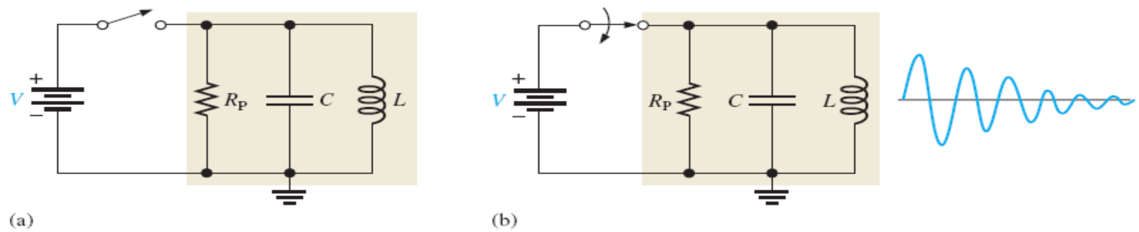


Figure 3–51: Parallel resonant circuit.

If a tunnel diode is placed in series with the tank circuit and biased at the center of the negative-resistance portion of its characteristic curve, as shown in Figure 3–52, a sustained oscillation (constant sinusoidal voltage) will result on the output. This is because the negative-resistance characteristic of the tunnel diode counteracts the positiveresistance characteristic of the tank resistance. The tunnel diode is only used at very high frequencies.

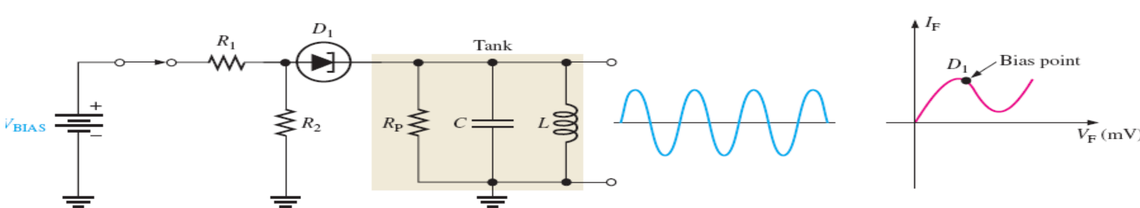


Figure 3–52: Basic tunnel diode oscillator.

Current Regulator Diode:

The current regulator diode is often referred to as a constant-current diode. Rather than maintaining a constant voltage, as the zener diode does, this diode maintains a constant current. The symbol is shown in Figure 3–53.

Figure 3–53: Symbol for a current regulator diode.

Figure 3–54 shows a typical characteristic curve. The current regulator diode operates in forward bias (shaded region), and the forward current becomes a specified constant value at forward voltages ranging from about 1.5 V to about 6 V, depending on the diode type. The constant forward current is called the *regulator current* and is designated I_P. For example, the 1N5283–1N5314 series of diodes have nominal regulator currents ranging from These diodes may be used in parallel to obtain higher currents 220 µA to 4.7 mA. This diode does not have a sharply defined reverse breakdown, so the reverse current begins to increase for V_{AK} values of less than 0 V (unshaded region of the figure). This device should never be operated in reverse bias.

In forward bias, the diode regulation begins at the limiting voltage, V_L, and extends up to the POV (peak operating voltage). Notice that between V_K and POV, the current is essentially constant. V_T is the test voltage at which I_P and the diode impedance, Z_T, are specified on a datasheet. The impedance Z_T has very high values ranging from 235 kΩ to 25 MΩ for the diode series mentioned before.

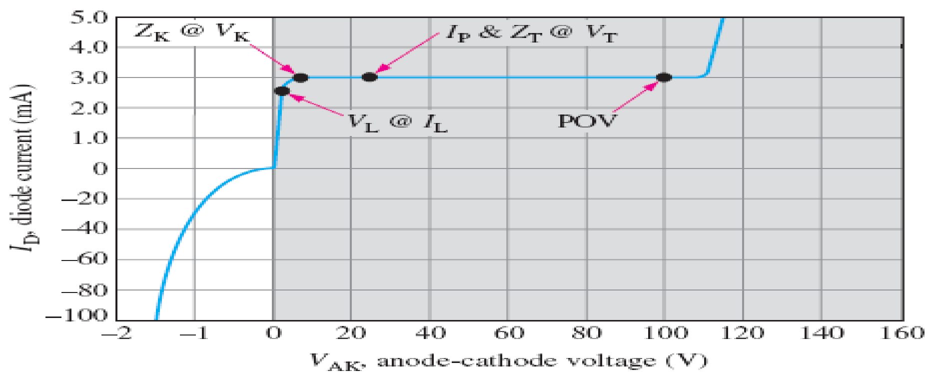


Figure 3–54: Typical characteristic curve for a current regulator diode.

SECTION 3–5 CHECKUP:

1. What does *laser* mean?
2. What is the difference between incoherent and coherent light and which is produced by a laser diode?
3. What are the primary application areas for Schottky diodes?
4. What is a hot-carrier diode?
5. What is the key characteristic of a tunnel diode?
6. What is one application for a tunnel diode?
7. Name the three regions of a *pin* diode.
8. Between what two voltages does a current regulator diode operate?

PROBLEMS:

Section 3–1: The Zener Diode

1. A certain zener diode has a V_Z = 7.5 V and an Z_Z = 5 Ω at a certain current. Draw the equivalent circuit.
2. From the characteristic curve in Figure 3–67, what is the approximate minimum zener current (I_{ZK}) and the approximate zener voltage at I_{ZK}?

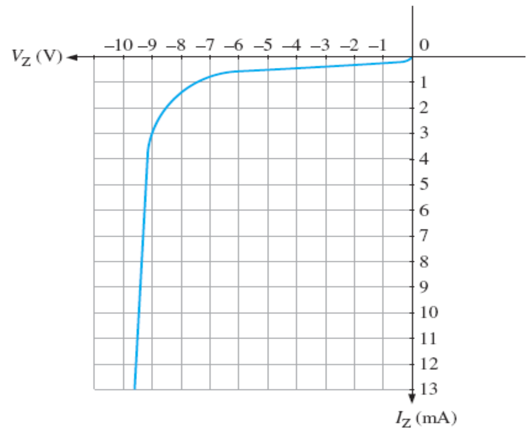


Figure 3–58

3. When the reverse current in a particular zener diode increases from 20 mA to 30 mA, the zener voltage changes from 5.6 V to 5.65 V. What is the impedance of this device?
4. A zener has an impedance of 15 Ω. What is its terminal voltage at 50 mA if V_Z = 4.7 V at I_Z = 25 mA?
5. A certain zener diode has the following specifications: V_Z = 6.8 V at 25°C and T_C = +0.04%/°C. Determine the zener voltage at 70°C.

Section 3–2 Zener Diode Applications

6. Determine the minimum input voltage required for regulation to be established in Figure 3–59. Assume an ideal zener diode with I_{ZK} = 1.5 mA and V_Z = 14 V.

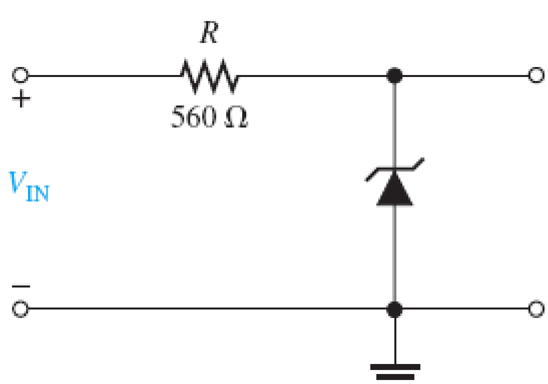


Figure 3–59

7. Repeat Problem 6 with $Z_Z = 20\ \Omega$ and $V_Z = 14$ V at 30 mA.

8. To what value must R be adjusted in Figure 3–60 to make $I_Z = 40$ mA? Assume $V_Z = 12$ V at 30 mA and $Z_Z = 30\ \Omega$.

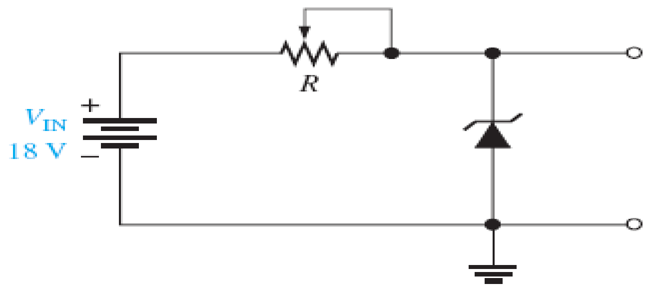


Figure 3–60

9. A 20 V peak sinusoidal voltage is applied to the circuit in Figure 3–60 in place of the dc source. Draw the output waveform. Use the parameter values established in Problem 8.

10. A loaded zener regulator is shown in Figure 3–61. $V_Z = 5.1$ V at $I_Z = 49$ mA, $I_{ZK} = 1$ mA, $Z_Z = 7\ \Omega$, and $I_{ZM} = 70$ mA. Determine the minimum and maximum permissible load currents.

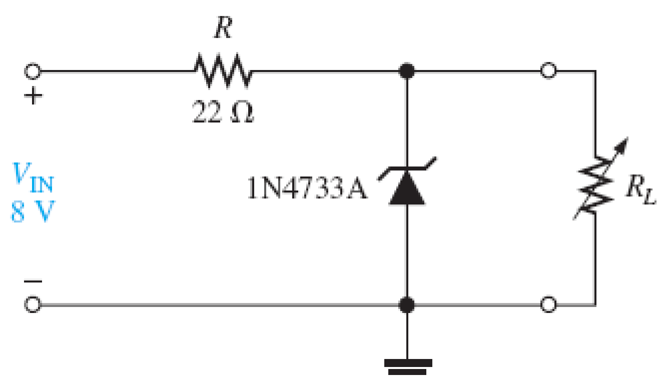


Figure 3–61

11. Find the load regulation expressed as a percentage in Problem 10. Refer to Chapter 2, Equation 2–15.

12. Analyze the circuit in Figure 3–61 for percent line regulation using an input voltage from 6 V to 12 V with no load. Refer to Chapter 2, Equation 2–14.

13. The no-load output voltage of a certain zener regulator is 8.23 V, and the full-load output is 7.98 V. Calculate the load regulation expressed as a percentage. Refer to Chapter 2, Equation 2–15.

14. In a certain zener regulator, the output voltage changes 0.2 V when the input voltage goes from 5 V to 10 V. What is the input regulation expressed as a percentage? Refer to Chapter 2, Equation 2–14.

15. The output voltage of a zener regulator is 3.6 V at no load and 3.4 V at full load. Determine the load regulation expressed as a percentage. Refer to Chapter 2, Equation 2–15.

Section 3–3: The Varactor Diode

16. Figure 3–62 is a curve of reverse voltage versus capacitance for a certain varactor. Determine the change in capacitance if V_R varies from 5 V to 20 V.

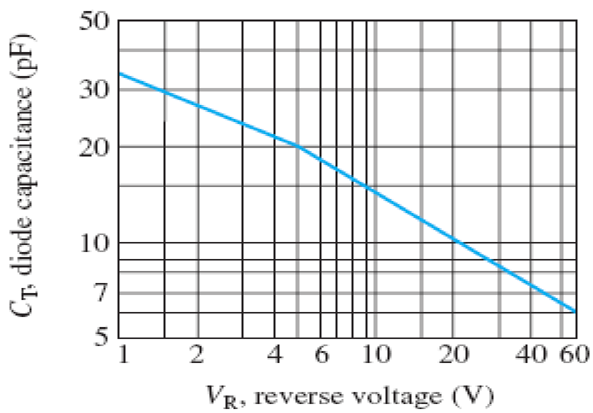


Figure 3–62

17. Refer to Figure 3–62 and determine the approximate value of *V*R that produces 25 pF.

18. What capacitance value is required for each of the varactors in Figure 3–63 to produce a resonant frequency of 1 MHz?

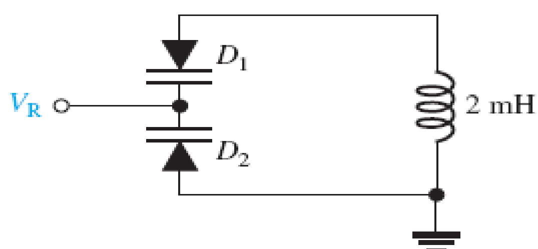


Figure 3–63

19. At what value must the voltage V_R be set in Problem 18 if the varactors have the characteristic curve in Figure 3–63?

Section 3–4: Optical Diodes

20. The LED in Figure 3–64(a) has a light-producing characteristic as shown in part (b). Neglecting the forward voltage drop of the LED, determine the amount of radiant (light) power produced in mW.

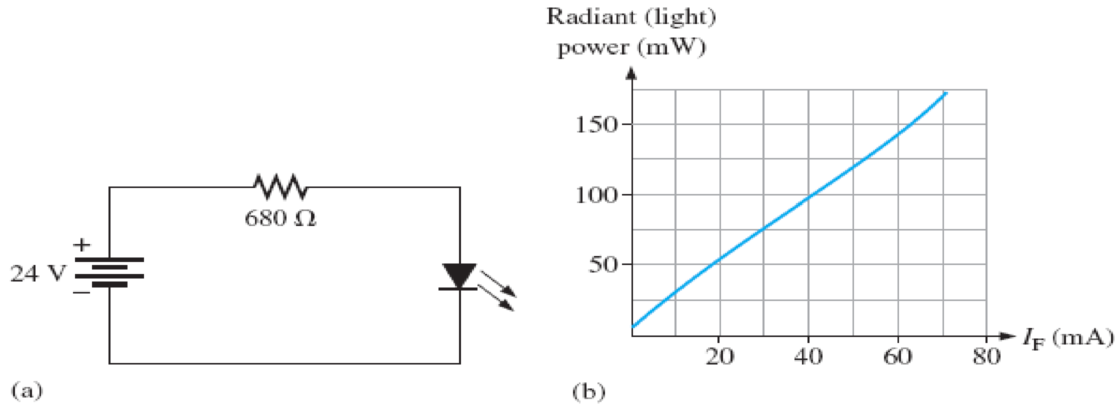


Figure 3–64(a)

21. Determine how to connect the seven-segment display in Figure 3–65 to display "5." The maximum continuous forward current for each LED is 30 mA and a +5 V dc source is to be used.

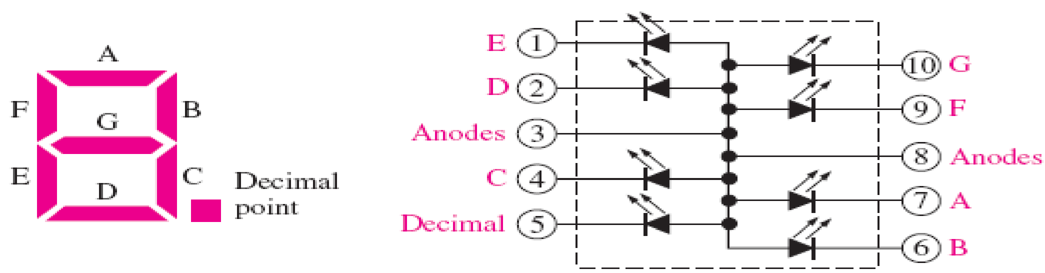


Figure 3–65

22. Specify the number of limiting resistors and their value for a series-parallel array of 48 red LEDs using a 9 V dc source for a forward current of 20 mA.

23. Develop a yellow LED traffic-light array using a minimum number of limiting resistors that operates from a 24 V supply and consists of 100 LEDs with I_F = 30 mA and an equal number of LEDs in each parallel branch. Show the circuit and the resistor values.

24. For a certain photodiode at a given irradiance, the reverse resistance is 200 kΩ and the reverse voltage is 10 V. What is the current through the device?

25. What is the resistance of each photodiode in Figure 3–66?

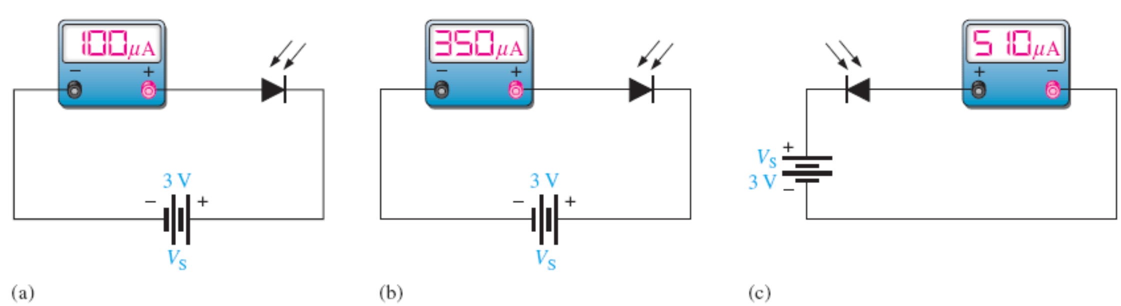


Figure 3–66

26. When the switch in Figure 3–67 is closed, will the microammeter reading increase or decrease? Assume D_1 and D_2 are optically coupled.

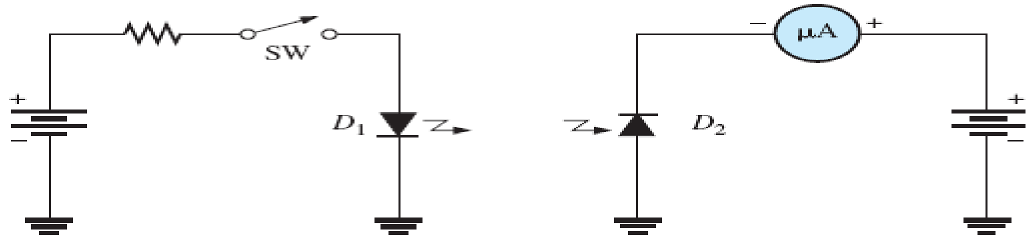


Figure 3–67

Section 3–5: Other Types of Diodes

27. The *V-I* characteristic of a certain tunnel diode shows that the current changes from 0.25 mA to 0.15 mA when the voltage changes from 125 mV to 200 mV. What is the resistance?

28. In what type of circuit are tunnel diodes commonly used?

29. What purpose do the reflective surfaces in the laser diode serve? Why is one end only partially reflective?

4–1 Bipolar junction transistor structure (BJT)

The basic structure of the bipolar junction transistor (BJT) determines its operating characteristics. In this section, you will see how semiconductive materials are used to form a BJT, and you will learn the standard BJT symbols.

The **BJT** is constructed with three doped semiconductor regions separated by two ***pn*** junctions, as shown in the epitaxial planar structure in Figure 4–1(a). The three regions are called **emitter**, **base**, and **collector**. Physical representations of the two types of BJTs are shown in Figure 4–1(b) and (c). One type consists of two ***n*** regions separated by a ***p*** region (***npn***), and the other type consists of two ***p*** regions separated by an ***n*** region (*pnp*). The term **bipolar** refers to the use of both holes and electrons as current carriers in the transistor structure.

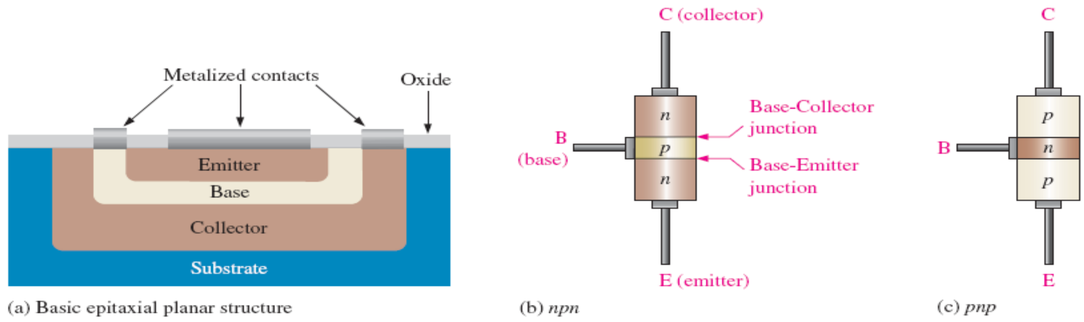


Figure 4–1: Basic BJT construction.

The ***pn*** junction joining the base region and the emitter region is called the *base-emitter junction*. The ***pn*** junction joining the base region and the collector region is called the *base-collector junction*, as indicated in Figure 4–1(b). A wire lead connects to each of the three regions, as shown. These leads are labeled **E**, **B**, and **C** for emitter, base, and collector, respectively. The base region is lightly doped and very thin compared to the heavily doped emitter and the moderately doped collector regions. (The reason for this is discussed in the next section.) Figure 4–2 shows the schematic symbols for the ***npn*** and ***pnp*** bipolar junction transistors.

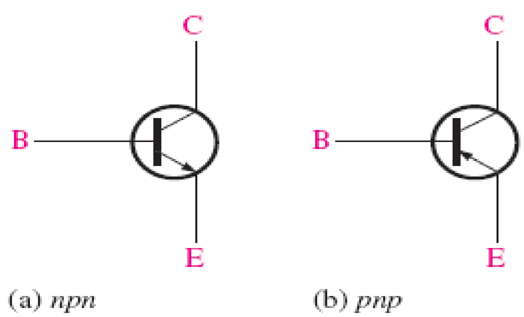


Figure 4–1: Standard BJT (bipolar junction transistor) symbols.

SECTION 4–1 CHECKUP

1. Name the two types of BJTs according to their structure.
2. The BJT is a three-terminal device. Name the three terminals.
3. What separates the three regions in a BJT?

4–2 Basic BJT operation

In order for a BJT to operate properly as an amplifier, the two ***pn*** junctions must be correctly biased with external dc voltages. In this section, we mainly use the ***npn*** transistor for illustration. The operation of the ***pnp*** is the same as for the ***npn*** except that the roles of the electrons and holes, the bias voltage polarities, and the current directions are all reversed.

Biasing:

Figure 4–3 shows a bias arrangement for both ***npn*** and ***pnp*** BJTs for operation as an **amplifier.** Notice that in both cases the base-emitter (BE) junction is forward-biased and the base-collector (BC) junction is reverse-biased. This condition is called *forward-reverse bias.*

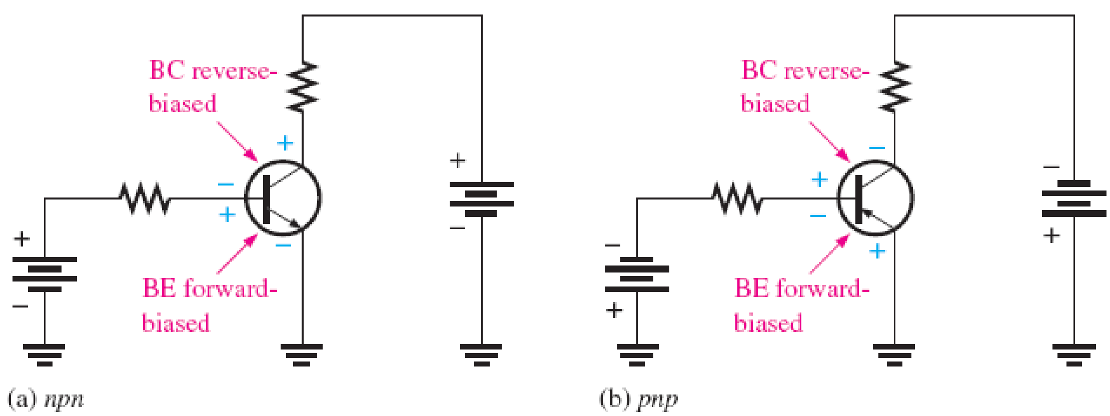


Figure 4–3: Forward-reverse bias of a BJT.

Operation:

To understand how a transistor operates, let's examine what happens inside the ***npn*** structure. The heavily doped ***n*-type** emitter region has a very high density of conduction-band (free) electrons, as indicated in Figure 4–4. These free electrons easily diffuse through the forwardbased BE junction into the lightly doped and very thin ***p*-type** base region, as indicated by the wide arrow. The base has a low density of holes, which are the majority carriers, as represented by the white circles. A small percentage of the total number of free electrons injected into the base region recombine with holes and move as valence electrons through the base region and into the emitter region as hole current, indicated by the red arrows.

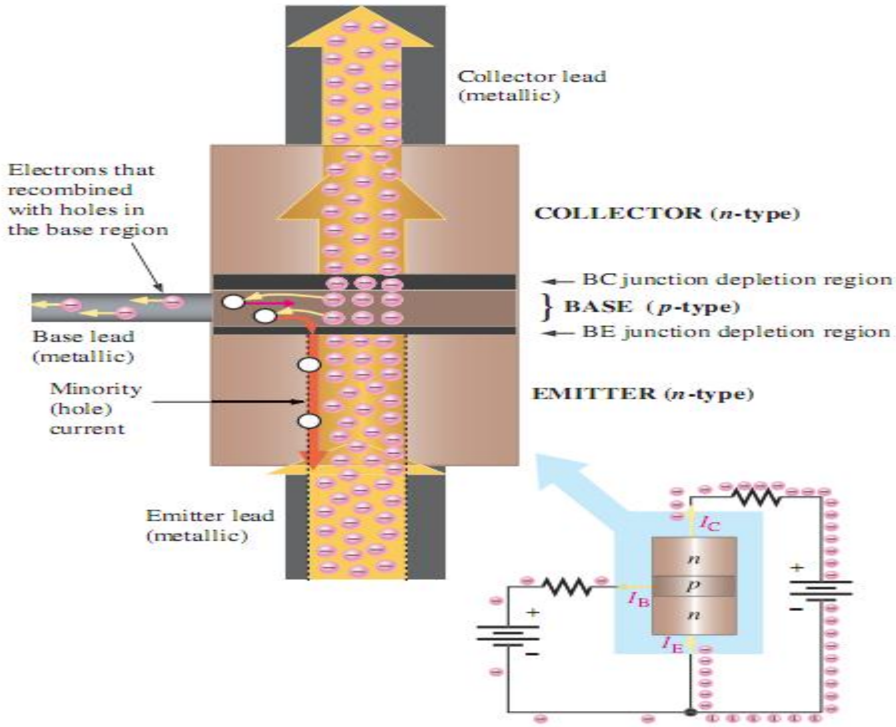


Figure 4–4: BJT operation showing electron flow.

When the electrons that have recombined with holes as valence electrons leave the crystalline structure of the base, they become free electrons in the metallic base lead and produce the external base current. Most of the free electrons that have entered the base do not recombine with holes because the base is very thin. As the free electrons move toward the reverse-biased BC junction, they are swept across into the collector region by the attraction of the positive collector supply voltage. The free electrons move through the collector region, into the external circuit, and then return into the emitter region along with the base current, as indicated. The emitter current is slightly greater than the collector current because of the small base current that splits off from the total current injected into the base region from the emitter.

Transistor Currents:

The directions of the currents in an npn transistor and its schematic symbol are as shown in Figure 4–5(a); those for a pnp transistor are shown in Figure 4–5(b). Notice that the arrow on the emitter inside the transistor symbols points in the direction of conventional current. These diagrams show that the emitter current (I_E) is the sum of the collector current (I_C) and the base current (I_B), expressed as follows

$$I_E = I_C + I_B \qquad (4\text{-}1)$$

As mentioned before, I_B is very small compared to I_E or I_C.

The capital-letter subscripts indicate dc values.

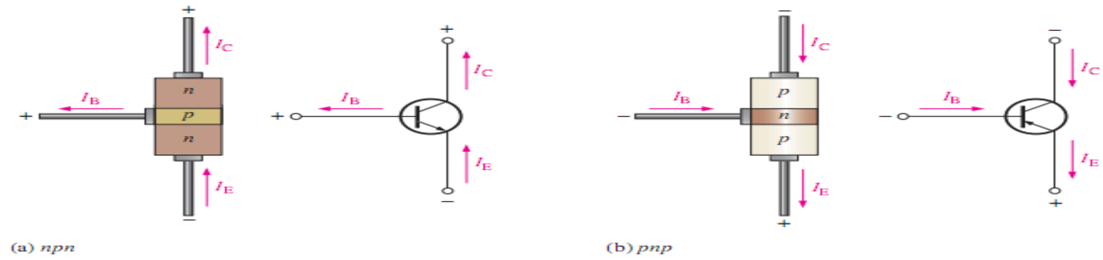


Figure 4–5: Transistor currents.

<u>SECTION 4–2 CHECKUP</u>:

1. What are the bias conditions of the base-emitter and base-collector junctions for a transistor to operate as an amplifier?
2. Which is the largest of the three transistor currents?
3. Is the base current smaller or larger than the emitter current?
4. Is the base region much thinner or much wider than the collector and emitter regions?
5. If the collector current is 1 mA and the base current is 10 μA, what is the emitter current?

4–3 BJT characteristics and parameters:

Two important parameters, β_{DC} (dc current gain) and α_{DC} are introduced and used to analyze a BJT circuit. Also, transistor characteristic curves are covered, and you will learn how a BJT's operation can be determined from these curves. Finally, maximum ratings of a BJT are discussed.

When a transistor is connected to dc bias voltages, as shown in Figure 4–6 for both ***npn*** and ***pnp*** types, V_{BB} forward-biases the base-emitter junction, and V_{CC} reverse-biases the base-collector junction. Although in this chapter we are using separate battery symbols to represent the bias voltages, in practice the voltages are often derived from a single dc power supply.

For example, V_{CC} is normally taken directly from the power supply output and V_{BB} (which is smaller) can be produced with a voltage divider.

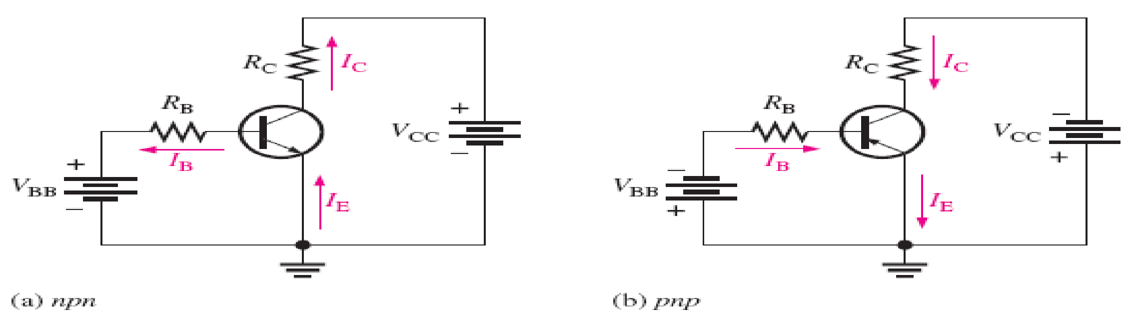


Figure 4–6: Transistor dc bias circuits.

DC Beta (β_{DC}) and DC Alpha (α_{DC})

The dc current **gain** of a transistor is the ratio of the dc collector current (I_C) to the dc base current (I_B) and is designated dc **beta** (β_{DC}).

$$\beta_{DC} = \frac{I_C}{I_B} \tag{4.2}$$

Typical values of β_{DC} range from less than 20 to 200 or higher. β_{DC} is usually designated as an equivalent hybrid (*h*) parameter, h_{FE}, on transistor datasheets. *h*-parameters are covered in Chapter 6. All you need to know now is that

$$h_{FE} = \beta_{DC}$$

The ratio of the dc collector current (I_C) to the dc emitter current (I_E) is the dc **alpha** (α_{DC}). The alpha is a less-used parameter than beta in transistor circuits.

$$\alpha_{DC} = \frac{I_C}{I_E}$$

Typically, values of α_{DC} range from 0.95 to 0.99 or greater, but α_{DC} is always less than 1. The reason is that I_C is always slightly less than I_E by the amount of I_B. For example, if I_E = 100 mA and I_B = 1 mA, then I_C = 99 mA and α_{DC} = 0.99.

Example 4–1:

Determine the dc current gain β_{DC} and the emitter current I_E for a transistor where I_B = 50 µA and I_C = 3.65 mA.

***Solution*:**

$$\beta_{DC} = \frac{I_C}{I_B} = \frac{3.65\ mA}{50\ \mu A} = 73$$

$I_E = I_C + I_B$ = 3.65 mA + 50 µA = **3.70 mA**

***Related Problem*:**

A certain transistor has a β_{DC} of 200. When the base current is 50 µA, determine the collector current.

Transistor DC Model:

You can view the unsaturated BJT as a device with a current input and a dependent current source in the output circuit, as shown in Figure 4–7 for an ***npn***.The input circuit is a forward-biased diode through which there is base current. The output circuit is a dependent current source (diamond-shaped element) with a value that is dependent on the base current, I_B, and equal to $\beta_{DC}I_B$. Recall that independent current source symbols have a circular shape.

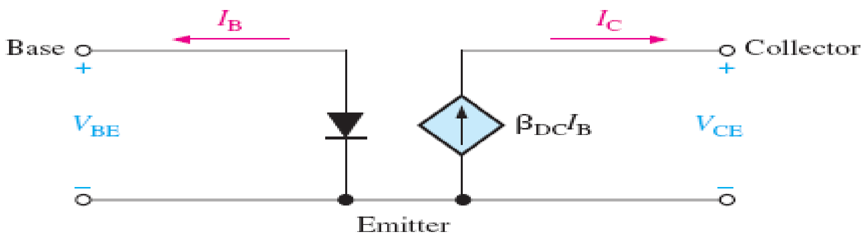


Figure 4–7: Ideal dc model of an ***npn*** transistor.

<u>BJT Circuit Analysis</u>:

Consider the basic transistor bias circuit configuration in Figure 4–8. Three transistor dc currents and three dc voltages can be identified.

I_B: dc base current

I_E: dc emitter current

I_C: dc collector current

V_{BE}: dc voltage at base with respect to emitter

V_{CB}: dc voltage at collector with respect to base

V_{CE}: dc voltage at collector with respect to emitter

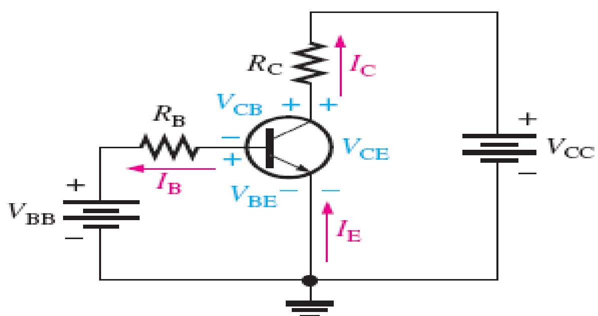


Figure 4–8: Transistor currents and voltages.

The base-bias voltage source, V_{BB}, forward-biases the base-emitter junction, and the collector-bias voltage source, V_{CC}, reverse-biases the base-collector junction. When the base-emitter junction is forward-biased, it is like a forward-biased diode and has a nominal forward voltage drop of

$$V_{BE} \approx 0.7 \text{ V} \qquad (4.3)$$

Although in an actual transistor V_{BE} can be as high as 0.9 V and is dependent on current, we will use 0.7 V throughout this text in order to simplify the analysis of the basic concepts. Keep in mind that the characteristic of the base-emitter junction is the same as a normal diode curve like the one in Figure 2-12. Since the emitter is at ground (0 V), by Kirchhoff's voltage law, the voltage across R_B is

$$V_{R_B} = V_{BB} - V_{BE}$$

Also, by Ohm's law,

$$V_{R_B} = I_B R_B$$

Substituting for V_{R_B} yields

$$I_B R_B = V_{BB} - V_{BE}$$

Solving for I_B,

$$I_B = \frac{V_{BB} - V_{BE}}{R_B} \qquad (4.4)$$

The voltage at the collector with respect to the grounded emitter is

$$V_{CE} = V_{CC} - V_{R_C}$$

Since the drop across R_C is

$$V_{R_C} = I_C R_C$$

the voltage at the collector with respect to the emitter can be written as

$$V_{CE} = V_{CC} - I_C R_C \qquad (4.5)$$

where $I_C = \beta_{DC} I_B$.

The voltage across the reverse-biased collector-base junction is

$$V_{CB} = V_{CE} - V_{BE} \qquad (4.6)$$

Example 4–2:

Determine I_B, I_C, I_E, V_{BE}, V_{CE}, and V_{CB} in the circuit of Figure 4–9. The transistor has a β_{DC} = 150.

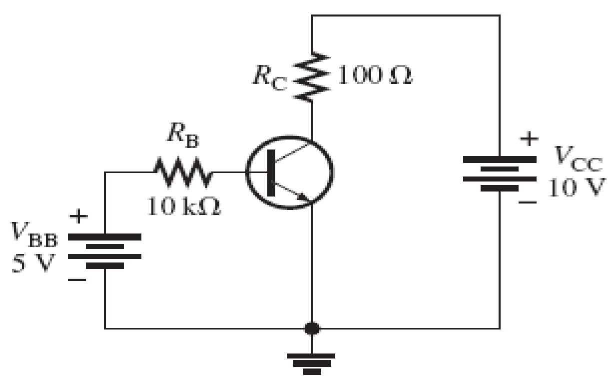


Figure 4–9

***Solution*:**

From Equation 4–3 , $V_{BE} \approx 0.7$ V, Calculate the base, collector, and emitter currents, as follows:

$$I_B = \frac{V_{BB} - V_{BE}}{R_B} = \frac{5V - 0.7V}{10\,k\Omega} = 430\ \mu A$$

$$I_C = \beta_{DC} I_B = (150)(430\text{ mA}) = \mathbf{64.5\text{ mA}}$$

$$I_E = I_C + I_B = 64.5\text{ mA} + 430\ \mu\text{A} = \mathbf{64.9\text{ mA}}$$

Solve for V_{CE} and V_{CB}.

$$V_{CE} = V_{CC} - I_C R_C = 10\text{ V} - (64.5\text{ mA})(100\ \Omega) = 10\text{ V} - 6.45\text{ V} = \mathbf{3.55\text{ V}}$$

$V_{CB} = V_{CE} - V_{BE} = 3.55\text{ V} - 0.7\text{ V} = \mathbf{2.85\text{ V}}$

Since the collector is at a higher voltage than the base, the collector-base junction is reverse-biased.

Related Problem: Determine I_B, I_C, I_E, V_{CE}, and V_{CB} in Figure 4–9 for the following values: R_B = 22 kΩ, R_C = 220 Ω, V_{BB} = 6 V, V_{CC} = 9 V, and β_{DC} = 90.

Collector Characteristic Curves

Using a circuit like that shown in Figure 4–10(a), a set of *collector characteristic curves* can be generated that show how the collector current, I_C, varies with the collector-toemitter voltage, V_{CE}, for specified values of base current, I_B. Notice in the circuit diagram that both V_{BB} and V_{CC} are variable sources of voltage. Assume that V_{BB} is set to produce a certain value of I_B and V_{CC} is zero. For this condition, both the base-emitter junction and the base-collector junction are forward-biased because the base is at approximately 0.7 V while the emitter and the collector are at 0 V. The base current is through the base-emitter junction because of the low impedance path to ground and, therefore, I_C is zero.

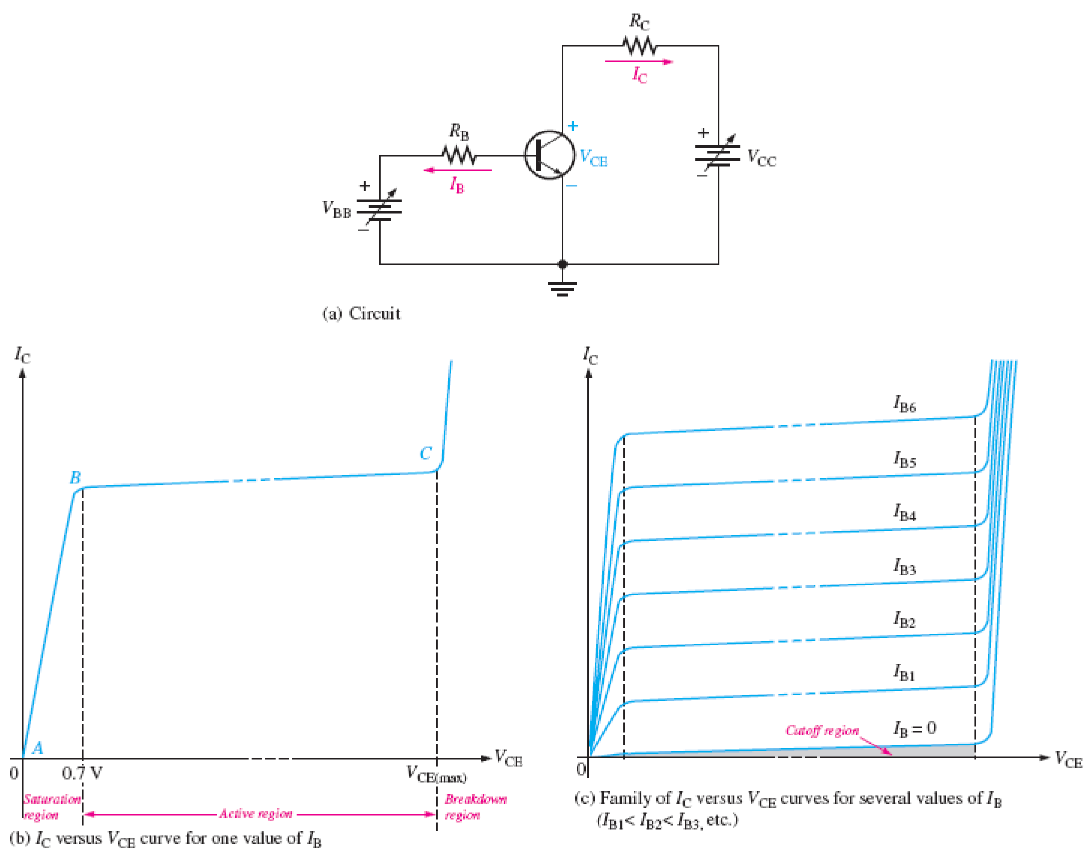


Figure 4–10: Collector characteristic curves.

When both junctions are forward-biased, the transistor is in the saturation region of its operation. **Saturation** is the state of a BJT in which the collector current has reached a maximum and is independent of the base current. As V_{CC} is increased, V_{CE} increases as the collector current increases. This is indicated by the portion of the characteristic curve between points *A* and *B* in Figure 4–10(b). I_C increases as V_{CC} is increased because V_{CE} remains less than 0.7 V due to the forward-biased base-collector junction. Ideally, when V_{CE} exceeds 0.7 V, the base-collector junction becomes reverse-biased and the transistor goes into the *active*, or **linear**, *region* of its operation. Once the basecollector junction is reverse-biased, I_C levels off and remains essentially constant for a given value of I_B as V_{CE} continues to increase. Actually, I_C increases very slightly as V_{CE} increases due to widening of the base-collector depletion region. This results in fewer holes for recombination in the base region which effectively causes a slight increase in β_{DC}. This is shown by the portion of the characteristic curve between points *B* and *C* in Figure 4–10(b). For this portion of the characteristic curve, the value of I_C is determined only by the relationship expressed as $I_C = \beta_{DC} I_B$. When V_{CE} reaches a sufficiently high voltage, the reverse-biased base-collector junction goes into breakdown; and the collector current increases rapidly as indicated by the part of the curve to the right of point *C* in Figure 4–10(b). A transistor should never be operated in this breakdown region. A family of collector characteristic curves is produced when I_C versus V_{CE} is plotted for several values of I_B, as illustrated in Figure 4–10(c). When $I_B = 0$, the transistor is in the cutoff region although there is a very small collector leakage current as indicated. **Cutoff** is the nonconducting state of a transistor. The amount of collector leakage current for $I_B = 0$ is exaggerated on the graph for illustration.

Example 4–3:

Sketch an ideal family of collector curves for the circuit in Figure 4–11 for $I_B = 5$ µA to 25 µA in 5 µA increments. Assume $\beta_{DC} = 100$ and that V_{CE} does not exceed breakdown.

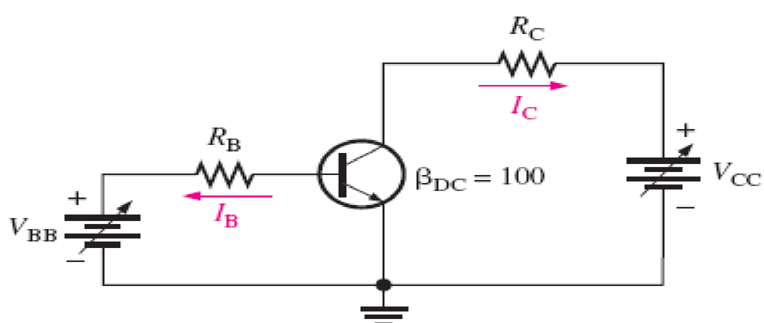


Figure 4–11

***Solution*:**

Using the relationship $I_C = \beta_{DC} I_B$, values of I_C are calculated and tabulated in Table 4–1. The resulting curves are plotted in Figure 4–12.

I_B	I_C
5 μA	0.5 mA
10 μA	1.0 mA
15 μA	1.5 mA
20 μA	2.0 mA
25 μA	2.5 mA

Table 4–1

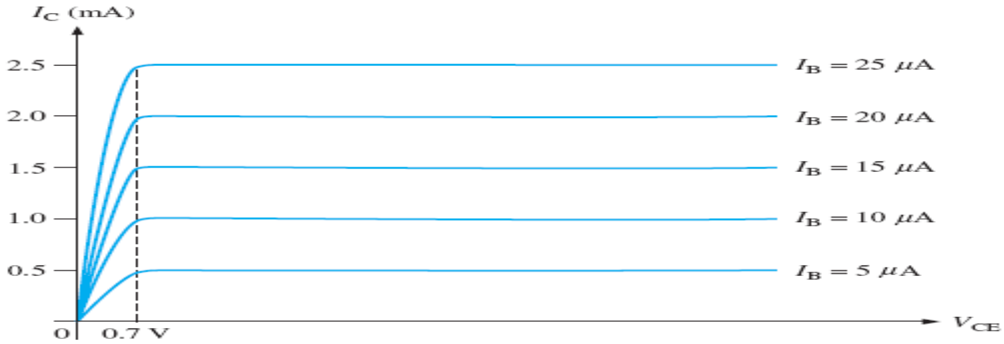


Figure 4–12

Related Problem:

Where would the curve for I_B = 0 appear on the graph in Figure 4–12, neglecting collector leakage current?

Cutoff:

As previously mentioned, when I_B = 0, the transistor is in the cutoff region of its operation. This is shown in Figure 4–13 with the base lead open, resulting in a base current of zero. Under this condition, there is a very small amount of collector leakage current, I_{CEO}, due mainly to thermally produced carriers. Because I_{CEO} is extremely small, it will usually be neglected in circuit analysis so that $V_{CE} = V_{CC}$. In cutoff, neither the base-emitter nor the base-collector junctions are forward-biased. The subscript CEO represents collectorto- emitter with the base open.

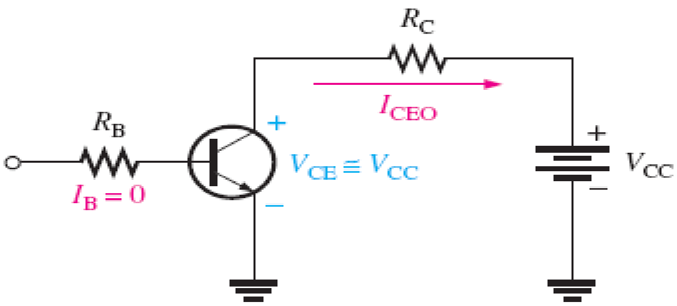


Figure 4–13: Cutoff: Collector leakage current (I_{CEO}) is extremely small and is usually neglected. Base-emitter and base-collector junctions are reverse-biased.

Saturation

When the base-emitter junction becomes forward-biased and the base current is increased, the collector current also increases ($I_C = \beta_{DC} I_B$) and V_{CE} decreases as a result of more drop across the collector resistor ($V_{CE} = V_{CC} - I_C R_C$). This is illustrated in Figure 4–14. When V_{CE} reaches its saturation value, $V_{CE(sat)}$, the base-collector junction becomes forward-biased and I_C can increase no further even with a continued increase in I_B.

At the point of saturation, the relation $I_C = \beta_{DC} I_B$ is no longer valid. $V_{CE(sat)}$ for a transistor occurs somewhere below the knee of the collector curves, and it is usually only a few tenths of a volt.

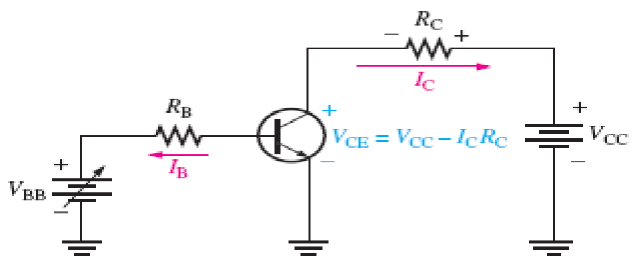


Figure 4–14: Saturation: As I_B increases due to increasing V_{BB}, I_C also increases and V_{CE} decreases due to the increased voltage drop across R_C. When the transistor reaches saturation, I_C can increase no further regardless of further increase in I_B. Base-emitter and base-collector junctions are forward-biased.

<u>DC Load Line</u>:

Cutoff and saturation can be illustrated in relation to the collector characteristic curves by the use of a load line. Figure 4–15 shows a dc load line drawn on a family of curves connecting the cutoff point and the saturation point.

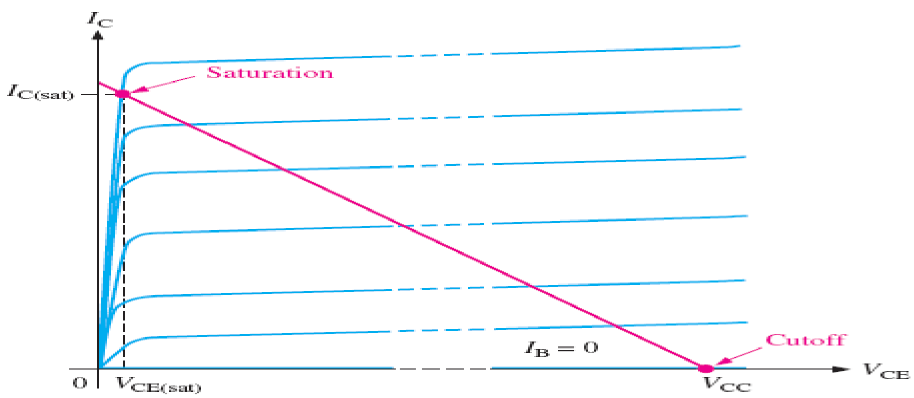


Figure 4–15: DC load line on a family of collector characteristic curves illustrating the cutoff and saturation conditions.

The bottom of the load line is at ideal cutoff where $I_C = 0$ and $V_{CE} = V_{CC}$. The top of the load line is at saturation where $I_C = I_{C(sat)}$ and $V_{CE} = V_{CE(sat)}$. In between cutoff and saturation along the load line is the *active region* of the transistor's operation.

Example 4–4:

Determine whether or not the transistor in Figure 4–16 is in saturation. Assume $V_{CE(sat)}$ = 0.2 V.

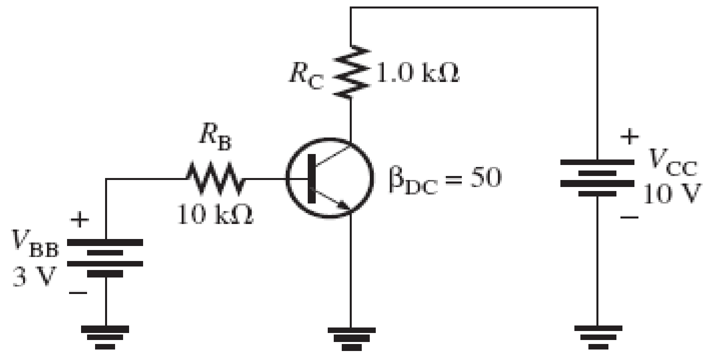


Figure 4–16

Solution:

First, determine $I_{C(sat)}$.

$$I_{C(sat)} = \frac{V_{CC} - V_{CE(sat)}}{R_C} = \frac{10V - 0.2V}{1.0\,k\Omega} = \frac{9.8V}{1.0\,k\Omega} = 9.8\,mA$$

Now, see if I_B is large enough to produce $I_{C(sat)}$.

$$I_B = \frac{V_{BB} - V_{BE}}{R_B} = \frac{3V - 0.7V}{10\,k\Omega} = \frac{2.3V}{10\,k\Omega} = 0.23\,mA$$

$$I_C = \beta_{DC} I_B = (50)(0.23 \text{ mA}) = 11.5 \text{ mA}$$

This shows that with the specified β_{DC}, this base current is capable of producing an I_C greater than $I_{C(sat)}$. Therefore, the **transistor is saturated,** and the collector current value of 11.5 mA is never reached. If you further increase I_B, the collector current remains at its saturation value of 9.8 mA.

***Related Problem*:**

Determine whether or not the transistor in Figure 4–16 is saturated for the following values: β_{DC} = 125, V_{BB} = 1.5 V, R_B =6.8 kΩ, R_C =180 Ω, and V_{CC} =12 V.

More About DC

The β_{DC} or h_{FE} is an important BJT parameter that we need to examine further. β_{DC} is not truly constant but varies with both collector current and with temperature. Keeping the junction temperature constant and increasing I_C causes β_{DC} to increase to a maximum. A further increase in I_C beyond this maximum point causes β_{DC} to decrease. If I_C is held constant and the temperature is varied, β_{DC} changes directly with the temperature. If the temperature goes up, β_{DC} goes up and vice versa. Figure 4–17 shows the variation of β_{DC} with I_C and junction temperature (T_J) for a typical BJT.

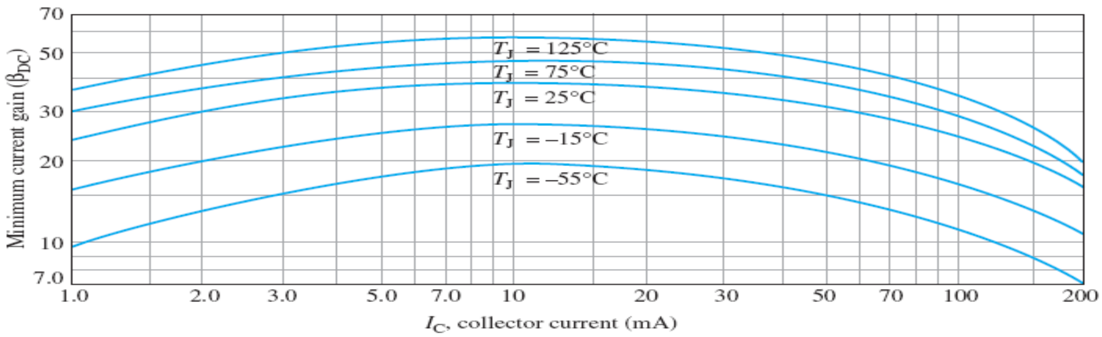


Figure 4–17: Variation of β_{DC} with I_C for several temperatures.

A transistor datasheet usually specifies β_{DC} (h_{FE}) at specific I_C values. Even at fixed values of I_C and temperature, β_{DC} varies from one device to another for a given type of transistor due to inconsistencies in the manufacturing process that are unavoidable. The β_{DC} specified at a certain value of I_C is usually the minimum value, $\beta_{DC(min)}$, although the maximum and typical values are also sometimes specified.

Maximum Transistor Ratings:

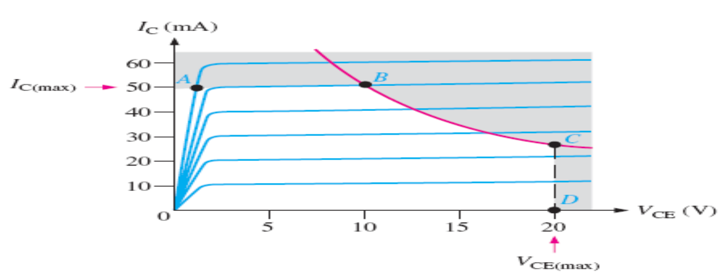
A BJT, like any other electronic device, has limitations on its operation. These limitations are stated in the form of maximum ratings and are normally specified on the manufacturer's datasheet. Typically, maximum ratings are given for collector-to-base voltage, collector-to-emitter voltage, emitter-to-base voltage, collector current, and power dissipation. The product of V_{CE} and I_C must not exceed the maximum power dissipation. Both V_{CE} and I_C cannot be maximum at the same time. If V_{CE} is maximum, I_C can be calculated as

$$I_C = \frac{P_{D\,(\max)}}{V_{CE}}$$

If I_C is maximum, V_{CE} can be calculated by rearranging the previous equation as follows:

$$V_{CE} = \frac{P_{D\,(\max)}}{I_C}$$

For any given transistor, a maximum power dissipation curve can be plotted on the collector characteristic curves, as shown in Figure 4–18(a).



$P_{D(max)}$	V_{CE}	I_C
500 mW	5 V	100 mA
500 mW	10 V	50 mA
500 mW	15 V	33 mA
500 mW	20 V	25 mA

Figure 4–18: Maximum power dissipation curve and tabulated values.

These values are tabulated in Figure 4–18(b). Assume $P_{D(max)}$ is 500 mW, $V_{CE(max)}$ is 20 V, and $I_{C(max)}$ is 50 mA. The curve shows that this particular transistor cannot be operated in the shaded portion of the graph. $I_{C(max)}$ is the limiting rating between points *A* and *B*, $P_{D(max)}$ is the limiting rating between points *B* and *C*, and $V_{CE(max)}$ is the limiting rating between points *C* and *D*.

Example 4–5:

A certain transistor is to be operated with V_{CE} = 6 V. If its maximum power rating is 250 mW, what is the most collector current that it can handle?

***Solution*:**

$$I_C = \frac{P_{D(max)}}{V_{CE}} = \frac{250\ mW}{6} = 41.7\ mA$$

This is the maximum current for this particular value of V_{CE}. The transistor can handle more collector current if V_{CE} is reduced, as long as $P_{D(max)}$ and $I_{C(max)}$ are not exceeded.

***Related Problem*:**

If $P_{D(max)}$ = 1 W, how much voltage is allowed from collector to emitter if the transistor is operating with I_C = 100 mA?

Example 4–6:

The transistor in Figure 4–19 has the following maximum ratings: $P_{D(max)}$ = 800 mW, $V_{CE(max)}$ = 15 V, and $I_{C(max)}$ = 100 mA. Determine the maximum value to which V_{CC} can be adjusted without exceeding a rating. Which rating would be exceeded first?

***Solution*:**

First, find I_B so that you can determine I_C.

$$I_B = \frac{V_{BB} - V_{BE}}{R_B} = \frac{5V - 0.7V}{22\ k\Omega} = \frac{4.3V}{22\ k\Omega} = 195\ \mu A$$

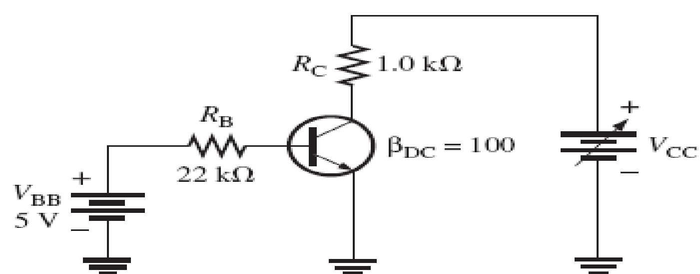


Figure 4–19

$$I_C = \beta_{DC} I_B = (100)(195\ \mu A) = 19.5\ mA$$

I_C is much less than $I_{C(max)}$ and ideally will not change with V_{CC}. It is determined only by I_B and β_{DC}.

The voltage drop across R_C is

$$V_{R_C} = I_C R_C = (1.0\ k\Omega)(19.5\ mA) = 19.5V$$

Now you can determine the value of V_{CC} when $V_{CE} = V_{CE(max)}$ = 15 V.

$$V_{R_C} = V_{CC} - V_{CE}$$

So

$$V_{R_C} = V_{CE\,(\max)} + (V_{R_C} = 15V + 19.5V) = 34.5V$$

V_{CC} can be increased to 34.5 V, under the existing conditions, before $V_{CE(max)}$ is exceeded. However, at this point it is not known whether or not $P_{D(max)}$ has been exceeded.

$$P_D = V_{CE(max)} I_C = (15\text{ V})(19.5\text{ mA}) = 293\text{ mW}$$

Since $P_{D(max)}$ is 800 mW, it is *not* exceeded when V_{CC} = 34.5 V. So, *V*CE(max) = 15 V is the limiting rating in this case. If the base current is removed causing the transistor to turn off, $\boldsymbol{V_{CE(max)}}$ **will be exceeded first** because the entire supply voltage, V_{CC}, will be dropped across the transistor.

***Related Problem*: The transistor in Figure 4–19 has the following maximum ratings: $P_{D(max)}$ = 500 mW, $V_{CE(max)}$ = 25 V, and $I_{C(max)}$ = 200 mA. Determine the maximum value to which V_{CC} can be adjusted without exceeding a rating. Which rating would be exceeded first?

Derating $P_{D(max)}$:

$P_{D(max)}$ is usually specified at 25°C. For higher temperatures, $P_{D(max)}$ is less. Datasheets often give derating factors for determining $P_{D(max)}$ at any temperature above 25°C. For example, a derating factor of 2 mW/°C indicates that the maximum power dissipation is reduced 2 mW for each degree Celsius increase in temperature.

Example 4–7:

A certain transistor has a $P_{D(max)}$ of 1 W at 25°C. The derating factor is 5 mW/°C. What is the $P_{D(max)}$ at a temperature of 70°C?

***Solution*:**

The change (reduction) in *P*D(max) is

$$\Delta P_{D(max)} = (5\text{ mW/°C})(70°\text{C} - 25°\text{C}) = (5\text{ mW/°C})(45°\text{C}) = 225\text{ mW}$$

Therefore, the $P_{D(max)}$ at 70°C is

1 W - 225 mW = **775 mW**

***Related Problem*:** A transistor has a $P_{D(max)}$ = 5 W at 25°C. The derating factor is 10 mW/°C. What is the $P_{D(max)}$ at 70°C?

***Related Problem*:**

Use the datasheet in Figure 4–20 to find the maximum *P*D at 50°C.

SECTION 4–3 CHECKUP

1. Define β_{DC} and α_{DC}. What is h_{FE}?
2. If the dc current gain of a transistor is 100, determine β_{DC} and α_{DC}.

3. What two variables are plotted on a collector characteristic curve?
4. What bias conditions must exist for a transistor to operate as an amplifier?
5. Does β_{DC} increase or decrease with temperature?
6. For a given type of transistor, can β_{DC} be considered to be a constant?

4–4 The BJT as an amplifier

Amplification is the process of linearly increasing the amplitude of an electrical signal and is one of the major properties of a transistor. As you learned, a BJT exhibits current gain (called b). When a BJT is biased in the active (or linear) region, as previously described, the BE junction has a low resistance due to forward bias and the BC junction has a high resistance due to reverse bias.

DC and AC Quantities:

Before discussing the concept of transistor amplification, the designations that we will use for the circuit quantities of current, voltage, and resistance must be explained because amplifier circuits have both dc and ac quantities.

In this text, italic capital letters are used for both dc and ac currents (I) and voltages (V). This rule applies to rms, average, peak, and peak-to-peak ac values. AC current and voltage values are always rms unless stated therwise. Although some texts use lowercase I and v for ac current and voltage, we reserve the use of lowercase i and v only for instantaneous values. In this text, the distinction between a dc current or voltage and an ac current or voltage is in the subscript. DC quantities always carry an uppercase roman (nonitalic) subscript. For example, I_{B}, I_{C}, and I_{E} are the dc transistor currents. V_{BE}, V_{CB}, and V_{CE} are the dc voltages from one transistor terminal to another. Single subscripted voltages such as V_{B}, V_{C}, and V_{E} are dc voltages from the transistor terminals to ground. AC and all time-varying quantities always carry a lowercase italic subscript. For example, I_b, I_c, and I_e are the ac transistor currents. V_{be}, V_{cb}, and V_{ce} are the ac voltages from one transistor terminal to another. Single subscripted voltages such as V_b, V_c, and V_e are ac voltages from the transistor terminals to ground. The rule is different for *internal* transistor resistances. As you will see later, transistors have internal ac resistances that are designated by lowercase r' with an appropriate subscript. For example, the internal ac emitter resistance is designated as r'_e. Circuit resistances external to the transistor itself use the standard italic capital R with a subscript that identifies the resistance as dc or ac (when applicable), just as for current and voltage. For example R_{E} is an external dc emitter resistance and R_e is an external ac emitter resistance.

Voltage Amplification:

As you have learned, a transistor amplifies current because the collector current is equal to the base current multiplied by the current gain, β. The base current in a transistor is very small compared to the collector and emitter currents. Because of this, the collector current is approximately equal to the emitter current. With this in mind, let's look at the circuit in Figure 4–21. An ac voltage, V_s, is superimposed on the dc bias voltage V_{BB} by capacitive coupling as shown. The dc bias voltage V_{CC} is connected to the collector through the collector resistor, R_C.

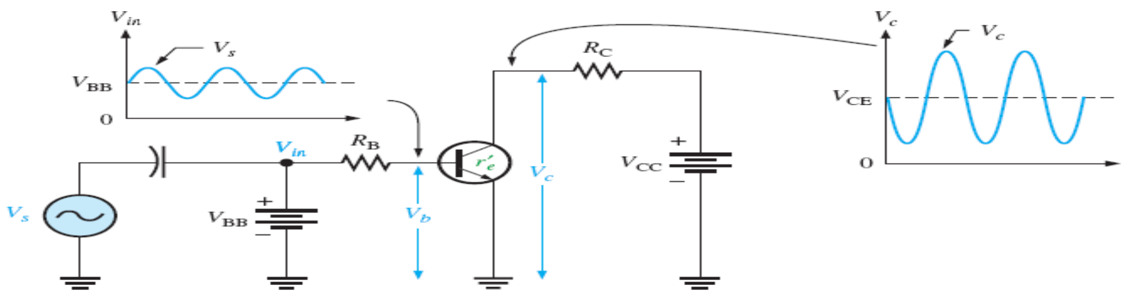


Figure 4–21: Basic transistor amplifier circuit with ac source voltage V_s and dc bias voltage V_{BB} superimposed.

The ac input voltage produces an ac base current, which results in a much larger ac collector current. The ac collector current produces an ac voltage across R_C, thus producing an amplified, but inverted, reproduction of the ac input voltage in the active region of operation, as illustrated in Figure 4–21.

The forward-biased base-emitter junction presents a very low resistance to the ac signal. This internal ac emitter resistance is designated r'_e in Figure 4–21 and appears in series with R_B. The ac base voltage is

$$V_b = I_e r'_e$$

The ac collector voltage, V_c, equals the ac voltage drop across R_C.

$$V_c = I_c R_C$$

Since $I_c \approx I_e$, the ac collector voltage is

$$V_c \approx I_e R_C$$

V_b can be considered the transistor ac input voltage where $Vb = Vs - I_b R_B$.

V_c can be considered the transistor ac output voltage. Since *voltage gain* is defined as the ratio of the output voltage to the input voltage, the ratio of V_c to V_b is the ac voltage gain, A_v, of the transistor.

$$A_v = \frac{V_c}{V_b}$$

Substituting $I_e R_C$ for V_c and $I_e r'_e$ for V_b yields

$$A_v = \frac{V_c}{V_b} \cong \frac{I_e R_C}{I_e r_e'}$$

The I_e terms cancel; therefore,

$$A_v \cong \frac{R_C}{r_e'} \qquad (4\text{-}7)$$

Equation 4–7 shows that the transistor in Figure 4–21 provides amplification in the form of voltage gain, which is dependent on the values of R_C and r'_e.

Since R_C is always considerably larger in value than, the output voltage for this configuration is greater than the input voltage. Various types of amplifiers are covered in detail in later chapters.

Example 4–9:

Determine the voltage gain and the ac output voltage in Figure 4–22 if $r'_e = 50\ \Omega$.

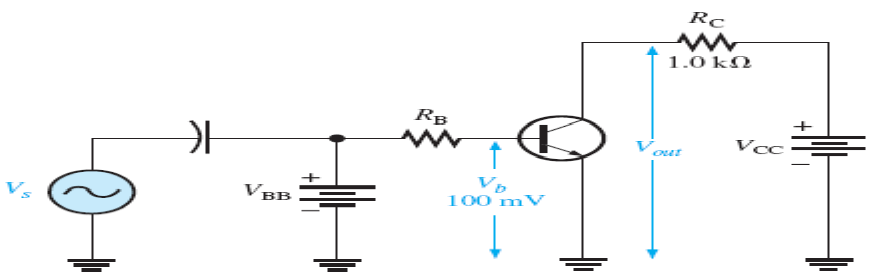


Figure 4–22

Solution:

The voltage gain is

$$A_v \cong \frac{R_C}{r_e'} = \frac{1.0\ k\Omega}{50\ \Omega} = 20$$

Therefore, the ac output voltage is

$$V_{out} = A_v V_b = (20)(100\ \text{mV}) = \mathbf{2\ V\ rms}$$

Related Problem:

What value of R_C in Figure 4–22 will it take to have a voltage gain of 50?

SECTION 4–4 CHECKUP:

1. What is amplification?
2. How is voltage gain defined?
3. Name two factors that determine the voltage gain of an amplifier.
4. What is the voltage gain of a transistor amplifier that has an output of 5 V rms and an input of 250 mV rms?
5. A transistor connected as in Figure 4–22 has an $r'_e = 20\ \Omega$. If R_C is 1200 Ω, what is the voltage gain?

4–5 The BJT as a switch

In the previous section, you saw how a BJT can be used as a linear amplifier. The second major application area is switching applications. When used as an electronic switch, a BJT is normally operated alternately in cutoff and saturation. Many digital circuits use the BJT as a switch.

Switching Operation

Figure 4–23 illustrates the basic operation of a BJT as a switching device. In part (a), the transistor is in the cutoff region because the base-emitter junction is not forward-biased. In this condition, there is, ideally, an *open* between collector and emitter, as indicated by the switch equivalent. In part (b), the transistor is in the saturation region because the base-emitter junction and the base-collector junction are forward-biased and the base current is made large enough to cause the collector current to reach its saturation value. In this condition, there is, ideally, a *short* between collector and emitter, as indicated by the switch equivalent. Actually, a small voltage drop across the transister of up to a few tenths of a volt normally occurs, which is the saturation voltage, $V_{CE(sat)}$.

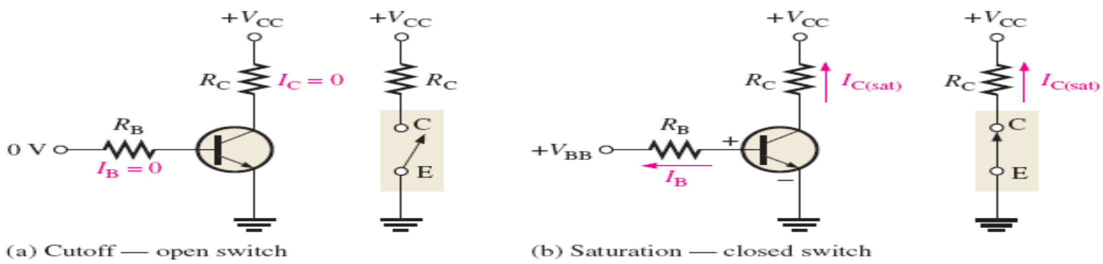


Figure 4–23: Switching action of an ideal transistor.

<u>Conditions in Cutoff</u>: As mentioned before, a transistor is in the cutoff region when the base-emitter junction is not forward-biased. Neglecting leakage current, all of the currents are zero, and V_{CE} is equal to V_{CC}.

$$V_{CE(cutoff)} = V_{CC} \tag{4-8}$$

<u>Conditions in Saturation</u>: As you have learned, when the base-emitter junction is forward-biased and there is enough base current to produce a maximum collector current, the transistor is saturated. The formula for collector saturation current is

$$I_{C(sat)} = \frac{V_{CC} - V_{CE(sat)}}{R_C} \tag{4-9}$$

Since $V_{CE(sat)}$ is very small compared to V_{CC}, it can usually be neglected.

The minimum value of base current needed to produce saturation is

$$I_{B(\min)} = \frac{I_{C(sat)}}{\beta_{DC}} \tag{4-10}$$

Normally, I_B should be significantly greater than $I_{B(min)}$ to ensure that the transistor is saturated.

Example 4–10:

(a) For the transistor circuit in Figure 4–24, what is V_{CE} when $V_{IN} = 0$ V?

(b) What minimum value of I_B is required to saturate this transistor if β_{DC} is 200? Neglect $V_{CE(sat)}$.

(c) Calculate the maximum value of R_B when $V_{IN} = 5$ V.

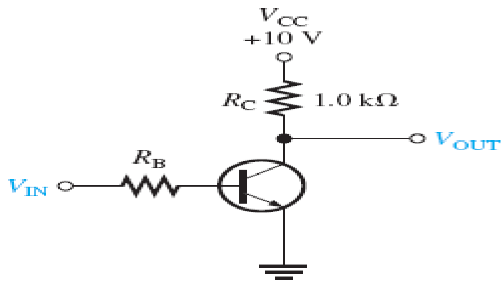


Figure 4–24

***Solution*:**

(a) When $V_{IN} = 0$ V, the transistor is in cutoff (acts like an open switch) and

$$V_{CE} = V_{CC} = \mathbf{10\ V}$$

(b) Since $V_{CE(sat)}$ is neglected (assumed to be 0 V),

$$I_{C(sat)} = \frac{V_{CC}}{R_C} = \frac{10V}{1.0\ k\Omega} = 10\ mA$$

$$I_{B(\min)} = \frac{I_{C(sat)}}{\beta_{DC}} = \frac{10\ mA}{200} = 50\ \mu A$$

This is the value of I_B necessary to drive the transistor to the point of saturation. Any further increase in I_B will ensure the transistor remains in saturation but there cannot be any further increase in I_C.

(c) When the transistor is on, $V_{BE} = 0.7$ V. The voltage across R_B is

$$V_{R_B} = V_{IN} - V_{BE} \cong 5V - 0.7V = 4.3V$$

Calculate the maximum value of R_B needed to allow a minimum I_B of 50 mA using Ohm's law as follows:

$$R_{B(\max)} = \frac{V_{R_B}}{I_{B(\min)}} = \frac{4.3V}{50\ \mu A} = 86\ k\Omega$$

***Related Problem*:**

Determine the minimum value of I_B required to saturate the transistor in Figure 4–24 if β_{DC} is 125 and $V_{CE(sat)}$ is 0.2 V.

A Simple Application of a Transistor Switch:

The transistor in Figure 4–25 is used as a switch to turn the LED on and off. For example, a square wave input voltage with a period of 2 s is applied to the input as indicated. When the square wave is at 0 V, the transistor is in cutoff; and since there is no collector current, the LED does not emit light. When the square wave goes to its high level, the transistor saturates. This forward-biases the LED, and the resulting collector current through the LED causes it to emit light. Thus, the LED is on for 1 second and off for 1 second.

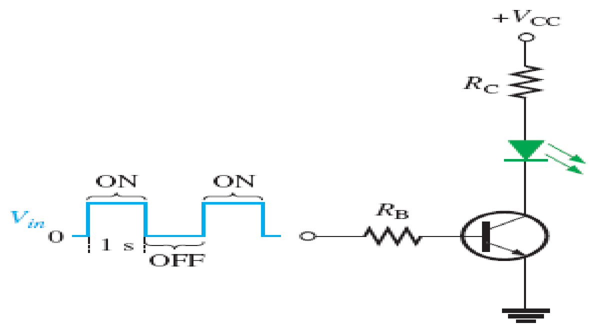


Figure 4–25: A transistor used to switch an LED on and off.

Example 4–11:

The LED in Figure 4–25 requires 30 mA to emit a sufficient level of light. Therefore, the collector current should be approximately 30 mA. For the following circuit values, determine the amplitude of the square wave input voltage necessary to make sure that the transistor saturates. Use double the minimum value of base current as a safety margin to ensure saturation. V_{CC} = 9 V, $V_{CE(sat)}$ = 0.3 V, R_C = 220 Ω, R_B = 3.3 kΩ, β_{DC} = 50, and V_{LED} = 1.6 V.

Solution:

$$I_{C(sat)} = \frac{V_{CC} - V_{LED} - V_{CE(sat)}}{R_C} = \frac{9V - 1.6V - 0.3V}{220\,\Omega} = 32.3\,mA$$

$$I_{B(\min)} = \frac{I_{C(sat)}}{\beta_{DC}} = \frac{32.3\,mA}{50} = 646\,\mu A$$

To ensure saturation, use twice the value of $I_{B(min)}$, which is 1.29 mA. Use Ohm's law to solve for V_{in}.

$$I_B = \frac{V_{R_B}}{R_B} = \frac{V_{in} - V_{BE}}{R_B} = \frac{V_{in} - 0.7V}{303\,k\Omega}$$

V_{in} - 0.7 V = $2I_{B(min)} R_B$ = (1.29 mA)(3.3 kΩ)

V_{in} = (1.29 mA)(3.3 kΩ) + 0.7 V = **4.96 V**

Related Problem:

If you change the LED in Figure 4–25 to one that requires 50 mA for a specified light emission and you can't increase the input amplitude above 5 V or V_{CC} above 9 V, how would you modify the circuit? Specify the component(s) to be changed and the value(s).

SECTION 4–5 CHECKUP:

1. When a transistor is used as a switch, in what two states is it operated?
2. When is the collector current maximum?
3. When is the collector current approximately zero?
4. Under what condition is $V_{CE} = V_{CC}$?
5. When is V_{CE} minimum?

4-6 The phototransistor

A phototransistor is similar to a regular BJT except that the base current is produced and controlled by light instead of a voltage source. The phototransistor effectively converts light energy to an electrical signal.

In a phototransistor the base current is produced when light strikes the photosensitive semiconductor base region. The collector-base pn junction is exposed to incident light through a lens opening in the transistor package. When there is no incident light, there is only a small thermally generated collector-to-emitter leakage current, I_{CEO}; this dark current is typically in the nA range. When light strikes the collector-base pn junction, a base current, I_λ, is produced that is directly proportional to the light intensity. This action produces a collector current that increases with I_λ. Except for the way base current is generated, the phototransistor behaves as a conventional BJT. In many cases, there is no electrical connection to the base.

The relationship between the collector current and the light-generated base current in a phototransistor is

$$I_C = \beta_{DC} I_\lambda \qquad (4\text{-}11)$$

The schematic symbol and some typical phototransistors are shown in Figure 4–26. Since the actual photogeneration of base current occurs in the collector-base region, the larger the physical area of this region, the more base current is generated. Thus, a typical phototransistor is designed to offer a large area to the incident light, as the simplified structure diagram in Figure 4–27 illustrates. A phototransistor can be either a two-lead or a three-lead device. In the three-lead configuration, the base lead is brought out so that the device can be used as a conventional BJT with or without the additional light-sensitivity feature. In the two-lead configuration, the base is not electrically available, and the device can be used only with light as the input. In many applications, the phototransistor is used in the two-lead version.

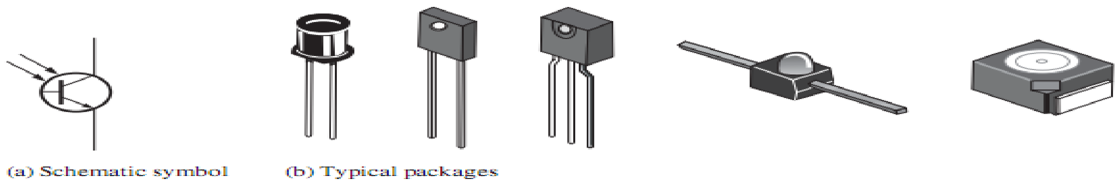


Figure 4–26: Phototransistor.

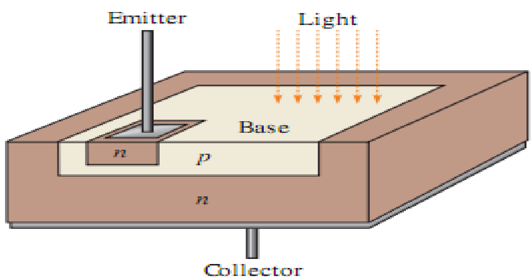


Figure 4–27: Typical phototransistor structure.

Figure 4–28 shows a phototransistor with a biasing circuit and typical collector characteristic curves. Notice that each individual curve on the graph corresponds to a certain value of light intensity (in this case, the units are mW/cm2) and that the collector current in creases with light intensity.

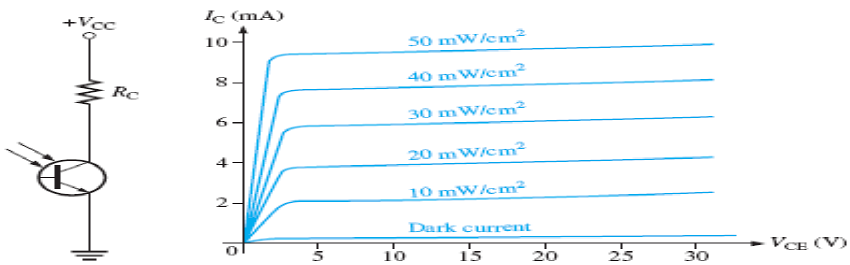


Figure 4–28: Phototransistor circuit and typical collector characteristic curves.

Phototransistors are not sensitive to all light but only to light within a certain range of wavelengths. They are most sensitive to particular wavelengths in the red and infrared part of the spectrum, as shown by the peak of the infrared spectral response curve in Figure 4–29.

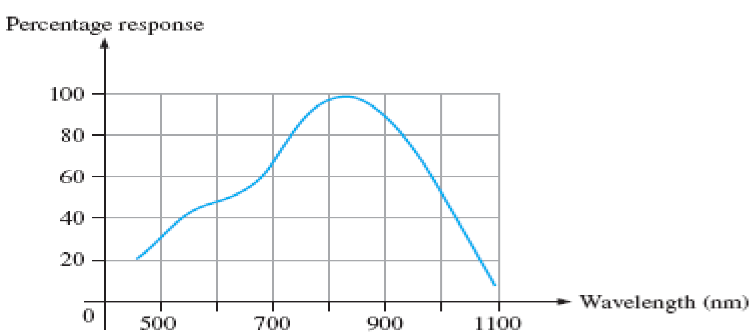


Figure 4–29: Typical phototransistor spectral response.

Applications

Phototransistors are used in a variety of applications. A light-operated relay circuit is shown in Figure 4–30(a). The phototransistor Q_1 drives the BJT Q_2. When there is sufficient incident light on Q_1, transistor Q_2 is driven into saturation, and collector current through the relay coil energizes the relay. The diode across the relay coil prevents, by its limiting action, a large voltage transient from occurring at the collector of Q_2 when the transistor turns off. Figure 4–30(b) shows a circuit in which a relay is deactivated by incident light on the phototransistor. When there is insufficient light, transistor Q_2 is biased on, keeping the relay energized. When there is sufficient light, phototransistor Q_1 turns on; this pulls the base of Q_2 low, thus turning Q_2 off and de-energizing the relay.

Optocouplers An optocoupler uses an LED optically coupled to a photodiode or a phototransistor in a single package.

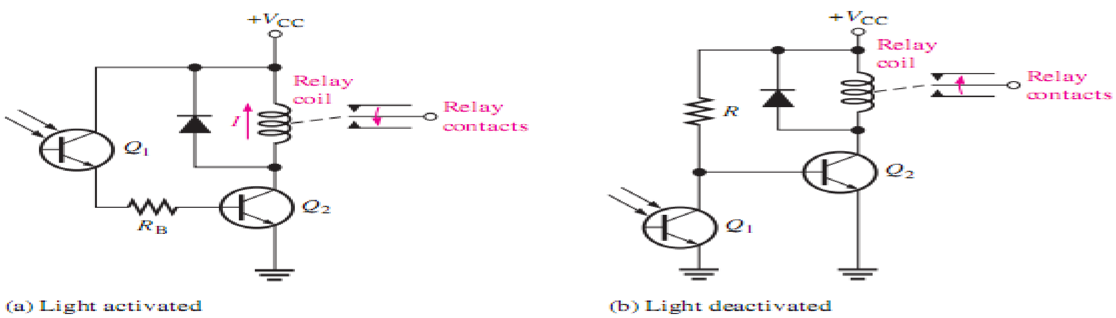


Figure 4–30: Relay circuits driven by a phototransistor.

Two basic types are LED-to-photodiode and LED-to-phototransistor, as shown in Figure 4–31. Examples of typical packages are shown in Figure 4–32.

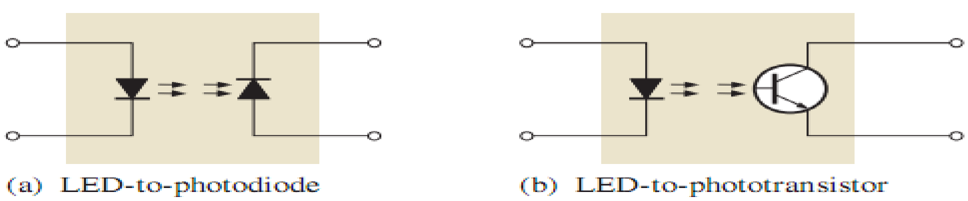


Figure 4–31: Basic optocouplers.

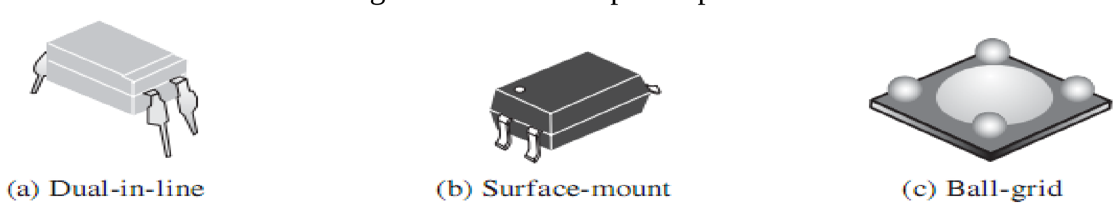


Figure 4–32: Examples of optocoupler packages.

A key parameter in optocouplers is the CTR (current transfer ratio). The CTR is an indication of how efficiently a signal is coupled from input to output and is expressed as the ratio of a change in the LED current to the corresponding change in the photodiode or phototransistor current. It is usually expressed as a percentage. Figure 4–33 shows a

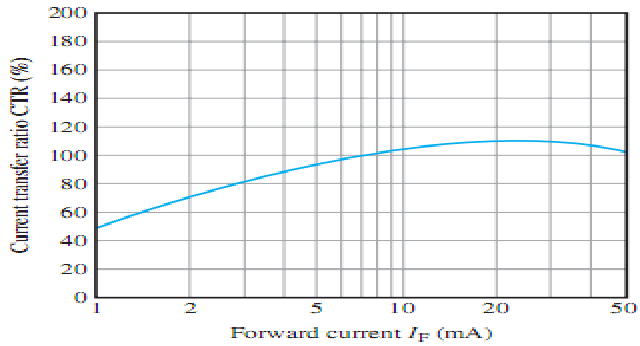


Figure 4–33: CTR versus IF for a typical optocoupler.

typical graph of CTR versus forward LED current. For this case, it varies from about 50% to about 110%. Optocouplers are used to isolate sections of a circuit that are incompatible in terms of the voltage levels or currents required. For example, they are used to protect hospital patients from shock when they are connected to monitoring instruments or other devices. They are also used to isolate low-current control or signal circuits from noisy power supply circuits or higher-current motor and machine circuits.

SECTION 4–6 CHECKUP

1. How does a phototransistor differ from a conventional BJT?
2. A three-lead phototransistor has an external (emitter, base, collector) lead.
3. The collector current in a phototransistor circuit depends on what two factors?
4. What is the optocoupler parameter, OTR?

Problems

Section 4–1

1. What are the majority carriers in the base region of an *npn* transistor called?

2. Explain the purpose of a thin, lightly doped base region.

Section 4–2

3. Why is the base current in a transistor so much less than the collector current?

4. In a certain transistor circuit, the base current is 2 percent of the 30 mA emitter current. Determine the collector current.

5. For normal operation of a *pnp* transistor, the base must be (+ or -) with respect to the emitter, and (+ or -) with respect to the collector.

6. What is the value of I_C for $I_E = 5.34$ mA and $I_B = 475$ µA?

Section 4–3

7. What is the α_{DC} when $I_C = 8.23$ mA and $I_E = 8.69$ mA?

8. A certain transistor has an *I*C 25 mA and an *I*B = 200 µA. Determine the β_{DC}.

9. What is the β_{DC} of a transistor if I_C = 20.3 mA and I_E = 20.5 mA?

10. What is the α_{DC} if I_C = 5.35 mA and I_B = 50 μA?

11. A certain transistor exhibits an α_{DC} of 0.96. Determine I_C when I_E = 9.35 mA.

12. A base current of 50 mA is applied to the transistor in Figure 4–53, and a voltage of 5 V is dropped across R_C. Determine the β_{DC} of the transistor.

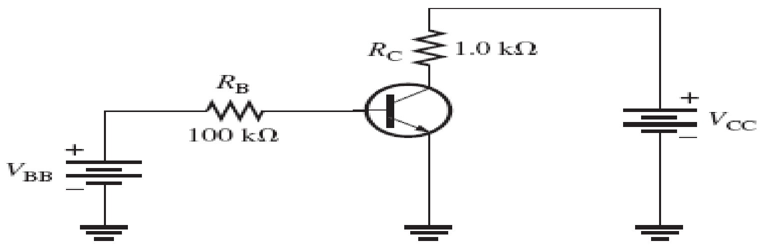


Figure 4–53

13. Calculate α_{DC} for the transistor in Problem 12.

14. Assume that the transistor in the circuit of Figure 4–53 is replaced with one having a β_{dc} of 200. Determine I_B, I_C, I_E, and V_{CE} given that V_{CC} = 10 V and V_{BB} = 3 V.

15. If V_{CC} is increased to 15 V in Figure 4–53, how much do the currents and V_{CE} change?

16. Determine each current in Figure 4–54. What is the β_{DC}?

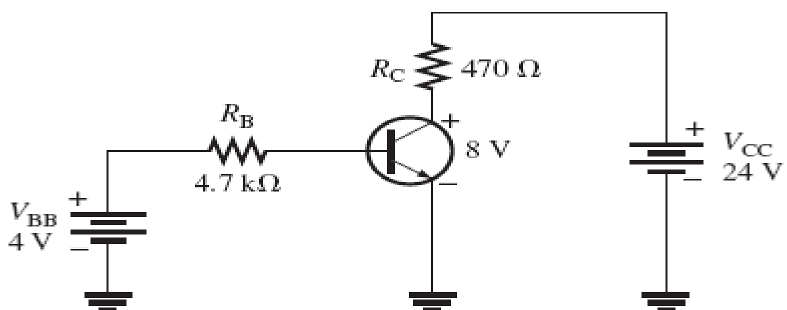


Figure 4–54

17. Find V_{CE}, V_{BE}, and V_{CB} in both circuits of Figure 4–55.

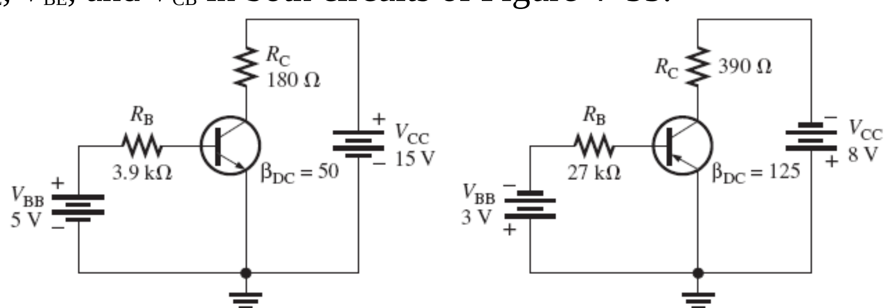


Figure 4–55 (a) (b)

18. Determine whether or not the transistors in Figure 4–55 are saturated.

19. Find I_B, I_E, and I_C in Figure 4–56. α_{DC} = 0.98.

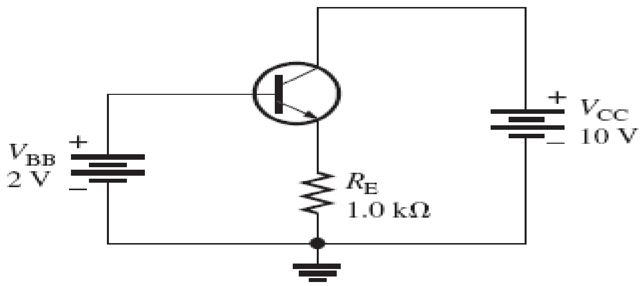


Figure 4–56

20. Determine the terminal voltages of each transistor with respect to ground for each circuit in Figure 4–57. Also determine V_{CE}, V_{BE}, and V_{CB}.

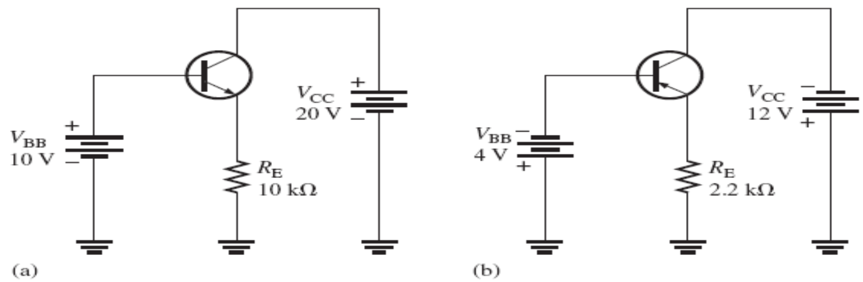


Figure 4–57

21. If the β_{DC} in Figure 4–57(a) changes from 100 to 150 due to a temperature increase, what is the change in collector current?

22. A certain transistor is to be operated at a collector current of 50 mA. How high can V_{CE} go without exceeding a $P_{D(max)}$ of 1.2 W?

23. The power dissipation derating factor for a certain transistor is 1 mW/°C. The $P_{D(max)}$ is 0.5 W at 25°C. What is $P_{D(max)}$ at 100°C?

Section 4–4

24. A transistor amplifier has a voltage gain of 50. What is the output voltage when the input voltage is 100 mV?

25. To achieve an output of 10 V with an input of 300 mV, what voltage gain is required?

26. A 50 mV signal is applied to the base of a properly biased transistor with $r'_e = 10\ \Omega$ and $R_C = 560\ \Omega$. Determine the signal voltage at the collector.

27. Determine the value of the collector resistor in an *npn* transistor amplifier with $\beta_{DC} = 250$, $V_{BB} = 2.5$ V, $V_{CC} = 9$ V, $V_{CE} = 4$ V, and $R_B = 100$ kΩ.

28. What is the dc current gain of each circuit in Figure 4–55?

Section 4–5

29. Determine $I_{C(sat)}$ for the transistor in Figure 4–58. What is the value of I_B necessary to produce saturation? What minimum value of V_{IN} is necessary for saturation? Assume $V_{CE(sat)} = 0$ V.

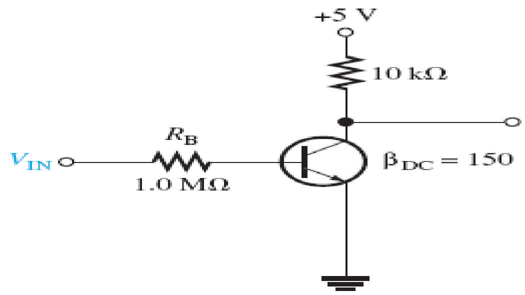


Figure 4–58

30. The transistor in Figure 4–59 has a β_{DC} of 50. Determine the value of R_B required to ensure saturation when V_{IN} is 5 V. What must V_{IN} be to cut off the transistor? Assume $V_{CE(sat)} = 0$ V.

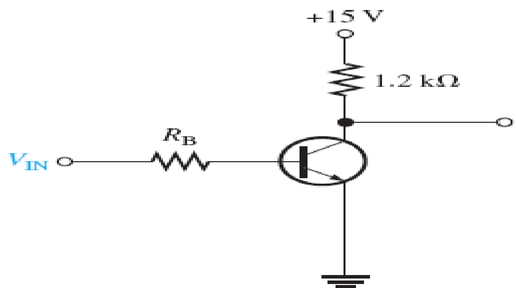


Figure 4–59

<u>Section 4–6</u>

31. A certain phototransistor in a circuit has a β_{DC} = 200. If I_λ = 100 mA, what is the collector current?

32. Determine the emitter current in the phototransistor circuit in Figure 4–60 if, for each lm/m^2 of light intensity, 1 μA of base current is produced in the phototransistor.

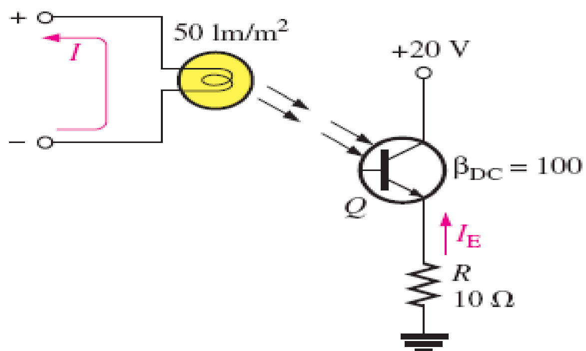


Figure 4–60

33. A particular optical coupler has a current transfer ratio of 30 percent. If the input current is 100 mA, what is the output current?

34. The optical coupler shown in Figure 4–61 is required to deliver at least 10 mA to the external load. If the current transfer ratio is 60 percent, how much current must be supplied to the input?

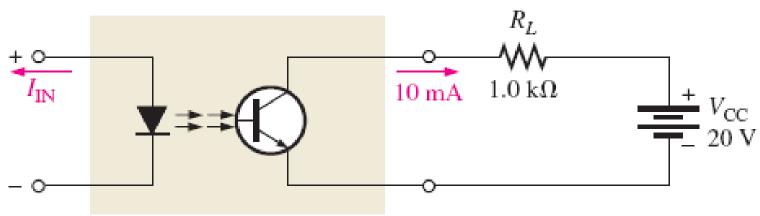


Figure 4–61

Section 4–7

35. Identify the leads on the transistors in Figure 4–62. Bottom views are shown.

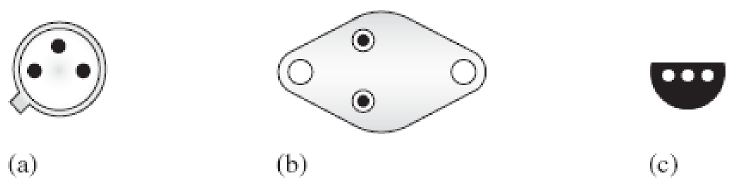


Figure 4–62

36. What is the most probable category of each transistor in Figure 4–63?

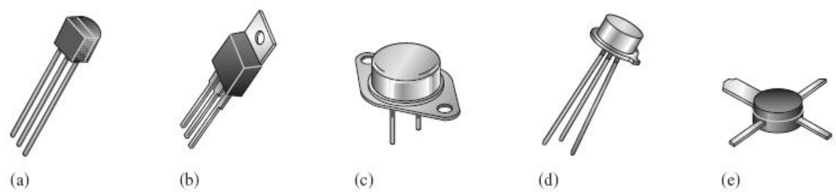


Figure 4–63

Section 4–8

37. In an out-of-circuit test of a good *npn* transistor, what should an analog ohmmeter indicate when its positive probe is touching the emitter and the negative probe is touching the base? When its positive probe is touching the base and the negative probe is touching the collector?

38. What is the most likely problem, if any, in each circuit of Figure 4–64? Assume a β_{DC} of 75.

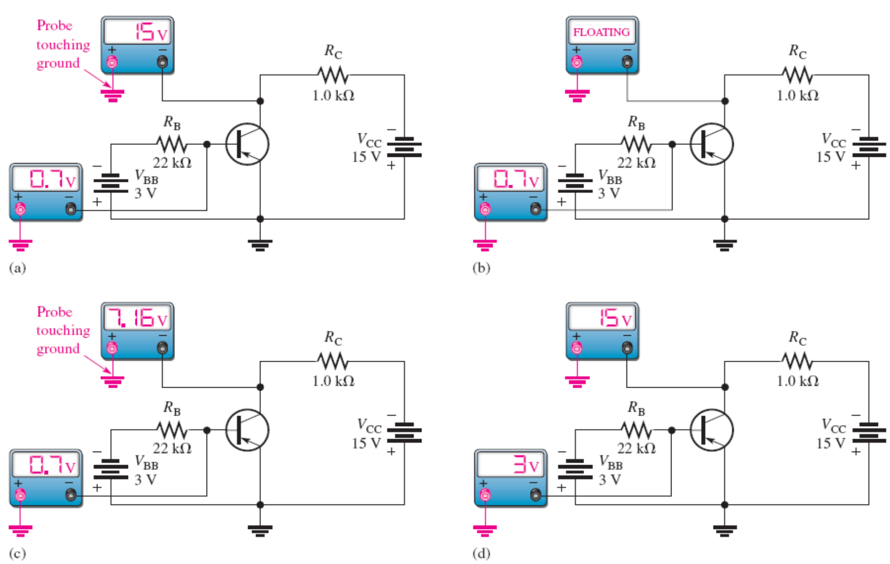


Figure 4–64

39. What is the value of the β_{DC} of each transistor in Figure 4–65?

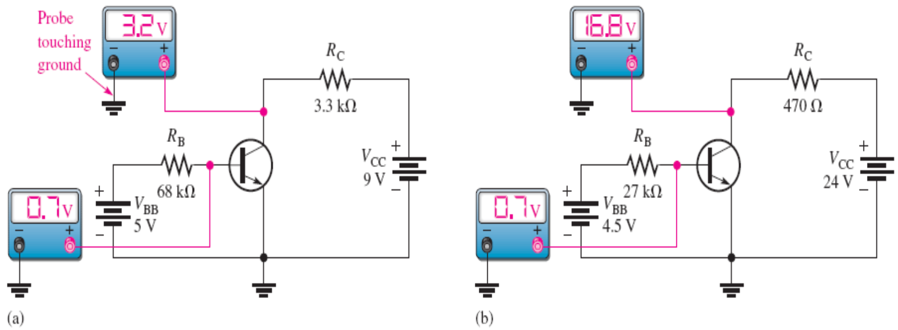


Figure 4–65

Application activity problems

40. Calculate the power dissipation in each resistor in Figure 4–51 for both states of the circuit.

41. Determine the minimum value of load resistance that Q_2 can drive without exceeding the maximum collector current specified on the datasheet.

42. Develop a wiring diagram for the printed circuit board in Figure 4–52 for connecting it in the security alarm system. The input/output pins are numbered from 1 to 10 starting at the top.

Advanced problems

50. Derive a formula for α_{DC} in terms of β_{DC}.

51. A certain 2N3904 dc bias circuit with the following values is in saturation. I_B = 500 µA, V_{CC} = 10 V, and R_C = 180 Ω, h_{FE} = 150. If you increase V_{CC} to 15 V, does the transistor come out of saturation? If so, what is the collector-to-emitter voltage and the collector current?

52. Design a dc bias circuit for a 2N3904 operating from a collector supply voltage of 9 V and a base-bias voltage of 3 V that will supply 150 mA to a resistive load that acts as the collector resistor. The circuit must not be in saturation. Assume the minimum specified β_{DC} from the datasheet.

53. Modify the design in Problem 52 to use a single 9 V dc source rather than two different sources. Other requirements remain the same.

54. Design a dc bias circuit for an amplifier in which the voltage gain is to be a minimum of 50 and the output signal voltage is to be "riding" on a dc level of 5 V. The maximum input signal voltage at the base is 10 mV rms. V_{CC} = 12 V, and V_{BB} = 4 V. Assume $r'_e = 8\ \Omega$.

1- THE JFET

Figure (1-a) shows the basic structure of an n-channel JFET (junction field-effect transistor). Wire leads are connected to each end of then-channel; the drain is at the upper end, and the source is at the lower end. Two p-type regions are diffused in then-type material to form a channel, and both p-type regions are connected to the gate lead. For simplicity, the gate lead is shown connected to only one of the p regions. A p-channel JFET is shown in Figure (1-b).

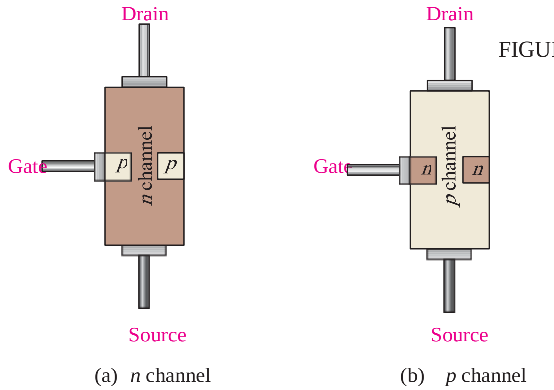


FIGURE (1): A representation of the basic structure of the two types of JFET.

Basic Operation

To illustrate the operation of a JFET, Figure (2) shows dc bias voltages applied to an *n*-channel device. V_{DD} provides a drain-to-source voltage and supplies current from drain to source. V_{GG} sets the reverse-bias voltage between the gate and the source, as shown.

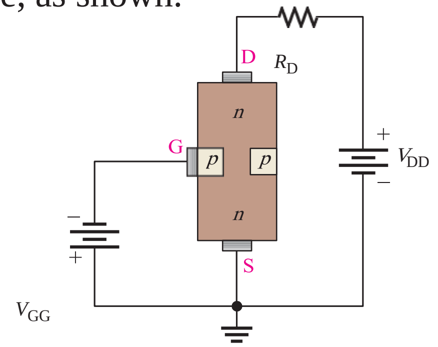
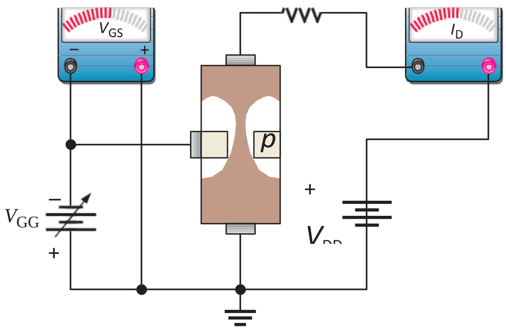
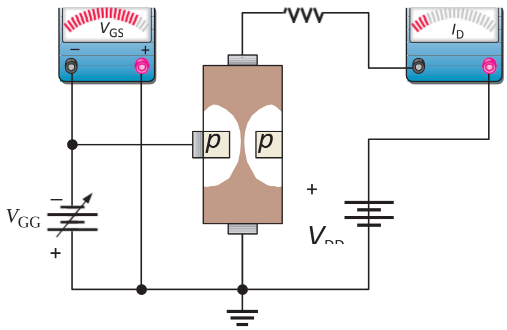


FIGURE (2): A biased n-channel JFET.

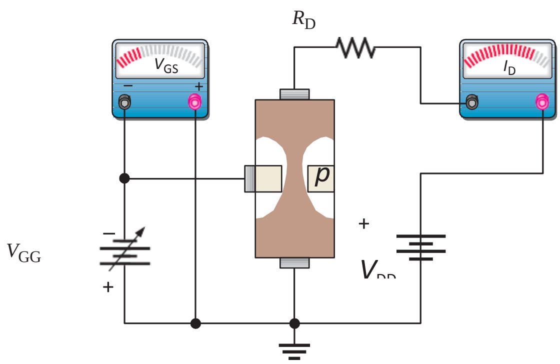
The JFET is always operated with the gate-source pn junction reverse-biased. Reverse biasing of the gate-source junction with a negative gate voltage produces a depletion region along the *pn* junction, which extends into the *n* channel and thus increases its resistance by restricting the channel width. The channel width and thus the channel resistance can be controlled by varying the gate voltage, thereby controlling the amount of drain current, I_D. Figure (3) illustrates this concept. The white areas represent the depletion region created by the reverse bias. It is wider toward the drain end of the channel because the reverse-bias voltage between the gate and the drain is greater than that between the gate and the source. We will discuss JFET characteristic curves and some parameters in Section (2).



(b) JFET biased for conduction



(c) Greater V_{GG} narrows the channel (between the white areas) which increases the resistance of the channel and decreases I_D.



(d) Less V_{GG} widens the channel (between the white areas) which decreases the resistance of the channel and increases I_D.

FIGURE (3): Effects of V_{GS} on channel width, resistance, and drain current (V_{GG} V_{GS}).

JFET Symbols

The schematic symbols for both *n*-channel and *p*-channel JFETs are shown in Figure (4). Notice that the arrow on the gate points "in" for *n* channel and "out" for *p* channel.

FIGURE (4): JFET schematic symbols.

2- **JFET CHARACTERISTICS AND PARAMETERS**

Drain Characteristic Curve

Consider the case when the gate-to-source voltage is zero (V_{GS} 0 V). This is produced by shorting the gate to the source, as in Figure (5-a) where both are grounded. As V_{DD} (and thus V_{DS}) is increased from 0 V, I_D will increase proportionally, as shown in the graph of Figure (5-b) between points *A* and *B*. In this area, the channel resistance is essentially constant because the depletion region is not large enough to have significant effect. This is called the *ohmic region* because V_{DS} and I_D are related by Ohm's law. At point *B* in Figure (5-b), the curve levels off and enters the active region where I_D becomes essentially constant. As V_{DS} increases from point *B* to point *C*, the reverse-bias voltage from gate to drain (V_{GD}) produces a depletion region large enough to offset the increase in V_{DS}, thus keeping I_D relatively constant.

Pinch-Off Voltage For V_{GS} 0 V, the value of V_{DS} at which I_D becomes essentially constant (point *B* on the curve in Figure (5-b) is the **pinch-off voltage**, V_P. For a given JFET, V_P has a fixed value. As you can see, a continued increase in V_{DS}

above the pinchoff voltage produces an almost constant drain current. This value of drain current is I_{DSS} (*D*rain to *S*ource current with gate *S*horted) and is always specified on JFET datasheets. I_{DSS} is the *maximum* drain current that a specific JFET can produce regardless of the external circuit, and it is always specified for the condition, V_{GS} 0 V.

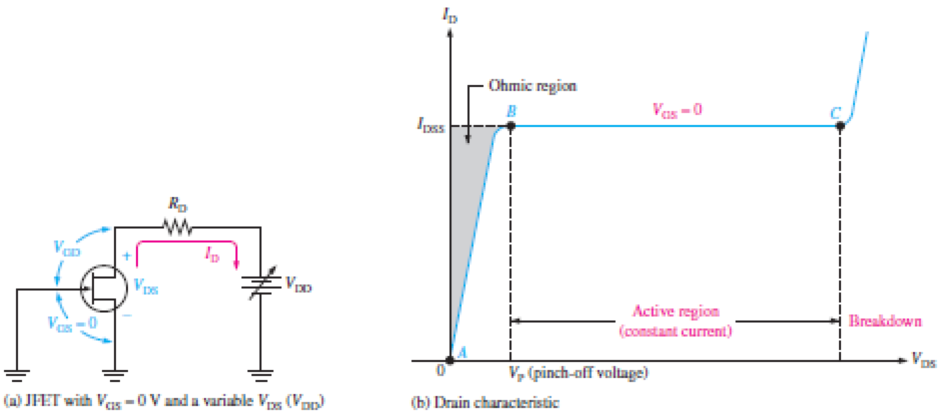


FIGURE (5): The drain characteristic curve of a JFET for V_{GS} 0 showing pinch-off voltage.

Breakdown

As shown in the graph in Figure (5-b), **breakdown** occurs at point *C* when I_D begins to increase very rapidly with any further increase in V_{DS}. Breakdown can result in irreversible damage to the device, so JFETs are always operated below breakdown and within the active region (constant current) (between points *B* and *C* on the graph). The JFET action that produces the drain characteristic curve to the point of breakdown for V_{GS} 0 V is illustrated in Figure (6).

V_{GS} Controls I_D

Let's connect a bias voltage, V_{GG}, from gate to source as shown in Figure (7-a). As V_{GS} is set to increasingly more negative values by adjusting V_{GG}, a family of drain characteristic curves is produced, as shown in Figure (7-b). Notice that I_D decreases as the magnitude of V_{GS} is increased to larger negative values because of the narrowing of the channel. Also notice that, for each increase in V_{GS}, the

JFET reaches pinch-off (where constant current begins) at values of V_{DS} less than V_P. The term *pinch-off* is not the same as pinchoff voltage, Vp. Therefore, the amount of drain current is controlled by V_{GS}, as illustrated in Figure (8).

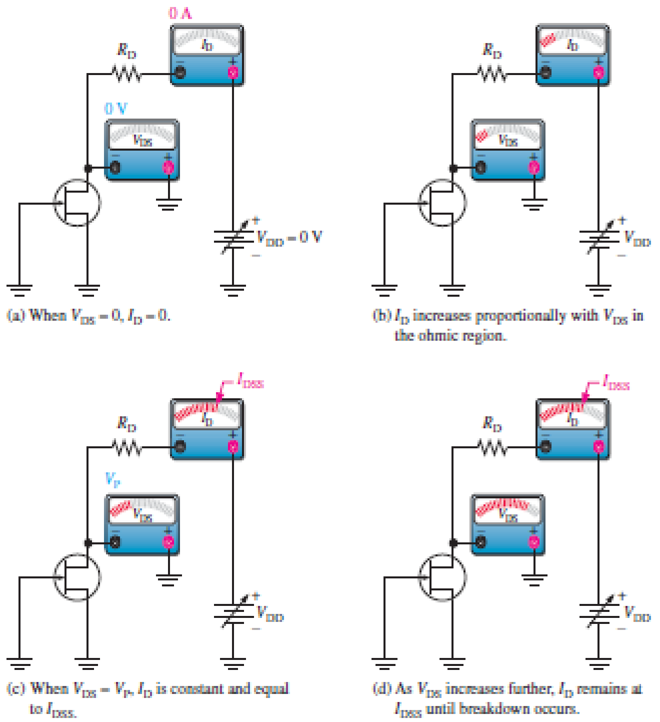


FIGURE (6): JFET action that produces the characteristic curve for V_{GS} 0 V.

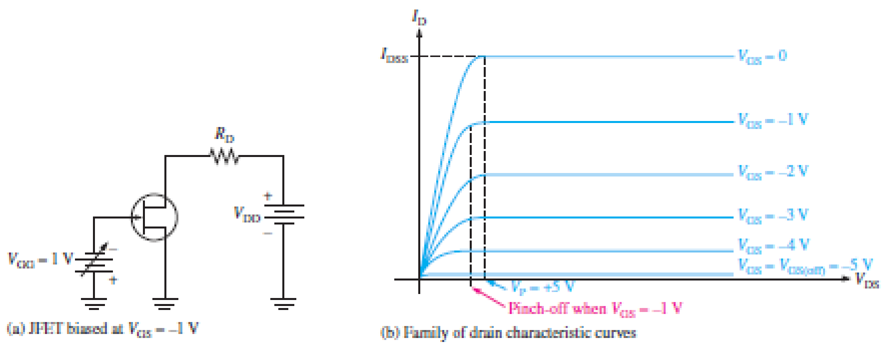


FIGURE (7): Pinch-off occurs at a lower V_{DS} as V_{GS} is increased to more negative values.

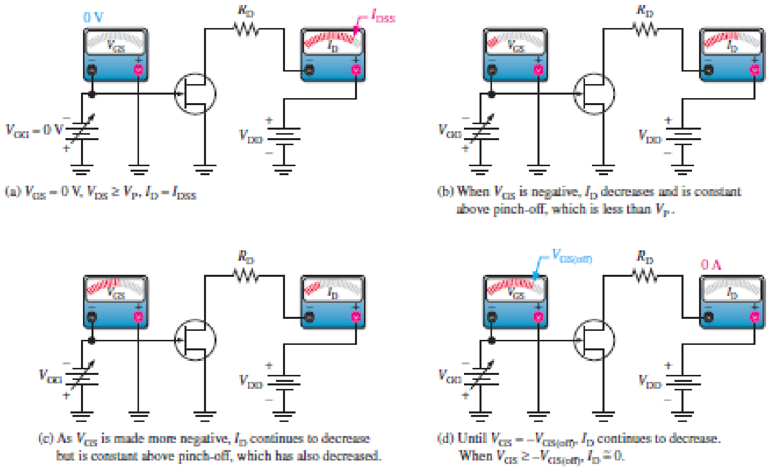


FIGURE (8):V_{GS} controls I_D.

As you have seen, for an *n*-channel JFET, the more negative V_{GS} is, the smaller I_D becomes in the active region. When V_{GS} has a sufficiently large negative value, I_D is reduced to zero. This cutoff effect is caused by the widening of the depletion region to a point where it completely closes the channel, as shown in Figure (9).

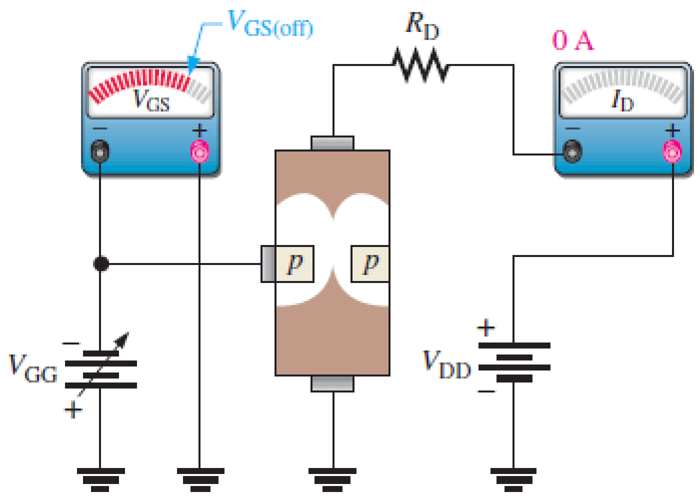


FIGURE (9): JFET at cutoff.

The basic operation of a *p*-channel JFET is the same as for an *n*-channel device except that a *p*-channel JFET requires a negative V_{DD} and a positive V_{GS}, as illustrated in Figure (10).

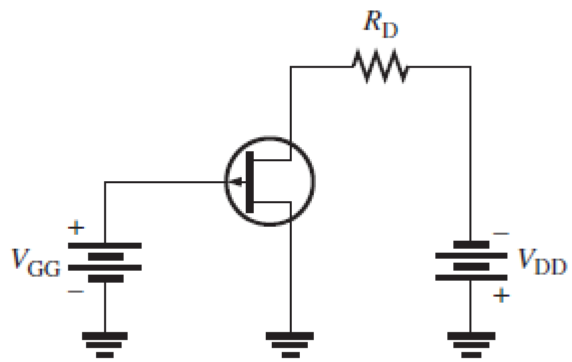


FIGURE (10): A biased *p*-channel JFET.

Comparison of Pinch-Off Voltage and Cutoff Voltage

As you have seen, there is a difference between pinch-off and cutoff voltages. There is also a connection. The pinch-off voltage V_P is the value of V_{DS} at which the drain currentbecomes constant and equal to I_{DSS} and is always measured at $V_{GS} = 0$ V. However, pinch-off occurs for V_{DS} values less than V_P when V_{GS} is nonzero. So, although V_P is aconstant, the minimum value of V_{DS} at which I_D becomes constant varies with V_{GS}, $V_{GS(off)}$ and V_P are always equal in magnitude but opposite in sign. A datasheet usually will give either $V_{GS(off)}$ or V_P, but not both. However, when you know one, you have the other. For example, if then $V_{GS(off)} = -5$ V, $V_P = +5$ V, as shown in Figure (7-b).

<u>**Example -1**</u>: For the JFET in Figure (11) $V_{GS(off)}$ = - 4 V and I_{DSS} = 12 mA. Determine the minimum value of V_{DD} required to put the device in the constant current region of operation when V_{GS} = 0 V.

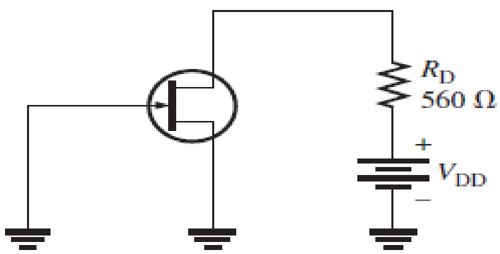


FIGURE (11)

Solution: Since $V_{GS(off)} = -4$ V, $V_P = 4$ V. The minimum value of V_{DS} for the JFET to be in its constant-current region is

$$V_{DS} = V_P = 4\text{ V}$$

In the constant current region with $V_{GS} = 0$ V

$$I_D = I_{DSS} = 12\text{ mA}$$

The drop across the drain resistor is

$$V_{RD} = I_D R_D = (12\text{ mA})(560\ \Omega) = 6.72\text{ V}$$

Apply Kirchhoff's law around the drain circuit

$$V_{DD} = V_{DS} + V_{RD} = 4\text{ V} + 6.72\text{ V} = 10.7\text{ V}$$

This is the value of V_{DD} to make $V_{DS} = V_P$ and put the device in the constant current region.

Related Problem* If V_{DD} is increased to 15 V, what is the drain current?

JFET Universal Transfer Characteristic

You have learned that a range of V_{GS} values from zero to $V_{GS(off)}$ controls the amount of drain current. For an *n*-channel JFET, $V_{GS(off)}$ is negative, and for a *p*-channel JFET, $V_{GS(off)}$ is positive. Because V_{GS} does control I_D, the relationship between these two quantities is very important. Figure (12) is a general transfer characteristic curve that illustrates graphically the relationship between V_{GS} and I_D. This curve is also known as a transconductance curve.

Notice that the bottom end of the curve is at a point on the *V*GS axis equal to *V*GS(off), and the top end of the curve is at a point on the *I*D axis equal to *I*DSS. This curve shows that

$$I_D = 0 \qquad \text{when } V_{GS} = V_{GS(off)}$$

$$I_D = I_{DSS}/4 \qquad \text{when } V_{GS} = 0.5V_{GS(off)}$$

$$I_D = I_{DSS}/2 \qquad \text{when } V_{GS} = 0.3V_{GS(off)}$$

and

$$I_D = I_{DSS} \qquad \text{when } V_{GS} = 0$$

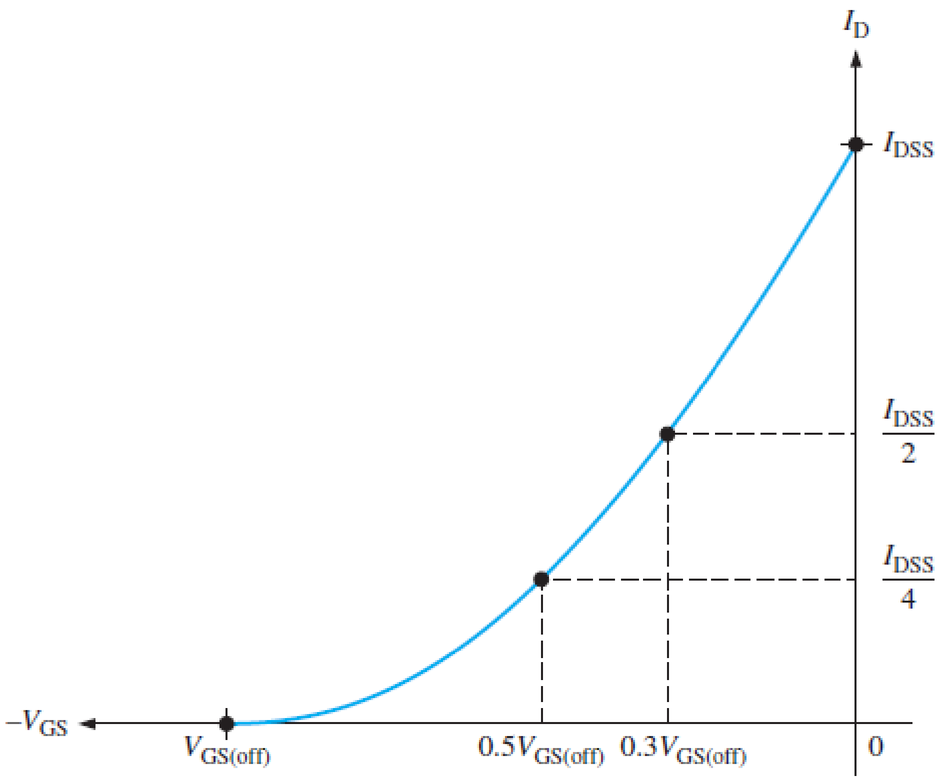


FIGURE (12): JFET universal transfer characteristic curve (*n*-channel)

The transfer characteristic curve can also be developed from the drain characteristic curves by plotting values of I_D for the values of V_{GS} taken from the family of drain curves at pinch-off, as illustrated in Figure (13) for a specific set of curves. Each point on the transfer characteristic curve corresponds to specific values of V_{GS} and I_D on the drain curves. For example, when Also, for this specific JFET, $V_{GS(off)} = -5$ V and $I_{DSS} = 12$ mA.

A JFET transfer characteristic curve is expressed approximately as

$$I_D \cong I_{Dss}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2 \qquad (1)$$

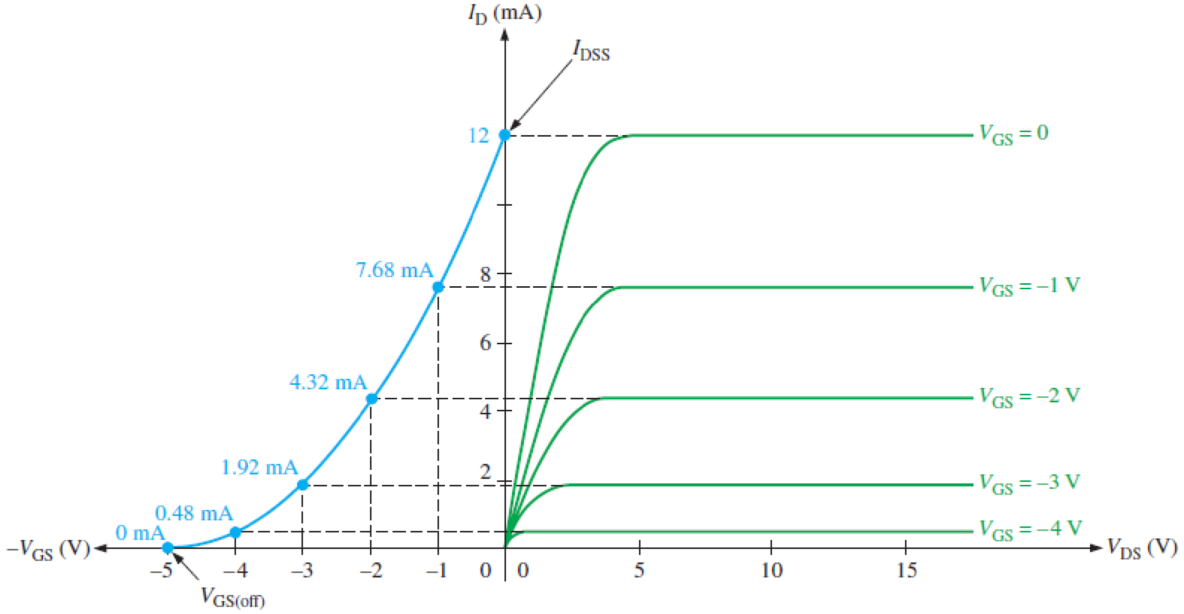


FIGURE (13): Example of the development of an *n*-channel JFET transfer characteristic curve (blue) from the JFET drain characteristic curves (green).

With Equation (1), I_D can be determined for any V_{GS} if $V_{GS(off)}$ and I_{DSS} are known. These quantities are usually available from the datasheet for a given JFET. Notice the squared term in the equation. Because of its form, a parabolic relationship is known as a *square law*, and therefore, JFETs and MOSFETs are often referred to as *square-law devices.*

<u>Example -1</u>:

For JFET indicates that typically $I_{DSS} = 9$ mA and $V_{GS(off)} = -8$ V (maximum). Using these values, determine the drain current for $V_{GS} = 0$ V, −1 V, and −4 V.

<u>Solution:</u>

For $V_{GS} = 0$ V,

$$I_D = I_{DSS} = \mathbf{9\ mA}$$

For $V_{GS} = -1$ V, use Equation 8–1.

$$I_D \cong I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2 = (9\text{ mA})\left(1 - \frac{-1\text{ V}}{-8\text{ V}}\right)^2$$

$$= (9\text{ mA})(1 - 0.125)^2 = (9\text{ mA})(0.766) = \mathbf{6.89\ mA}$$

For $V_{GS} = -4$ V,

$$I_D \cong (9\text{ mA})\left(1 - \frac{-4\text{ V}}{-8\text{ V}}\right)^2 = (9\text{ mA})(1 - 0.5)^2 = (9\text{ mA})(0.25) = \mathbf{2.25\ mA}$$

JFET Forward Transconductance

The forward **transconductance** (transfer conductance), g_m, is the change in drain current (ΔI_D) for a given change in gate-to-source voltage (ΔV_{GS}) with the drain-to-source voltage constant. It is expressed as a ratio and has the unit of siemens (S).

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}$$

Because the transfer characteristic curve for a JFET is nonlinear, g_m varies in value depending on the location on the curve as set by VGS. The value for g_m is greater near the top of the curve (near V_{GS} 0) than it is near the bottom (near $V_{GS(off)}$), as illustrated in Figure (14)

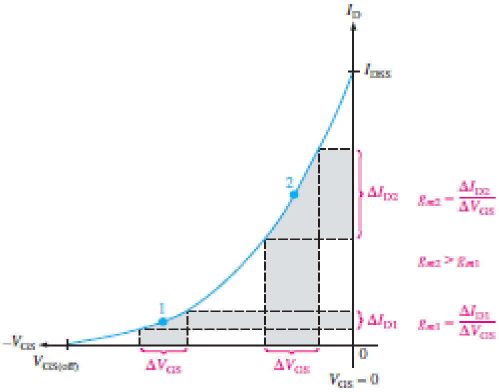


FIGURE (14): g_m varies depending on the bias point (V_{GS}).