Zener Theory and Design Considerations

Handbook

HBD854/D
Rev. 1, Dec−2017

ON Semiconductor™

Technical Information, Application Notes and Articles

ZENER DIODE THEORY

INTRODUCTION

The zener diode is a semiconductor device unique in its mode of operation and completely unreplaceable by any other electronic device. Because of its unusual properties it fills a long-standing need in electronic circuitry. It provides, among other useful functions, a constant voltage reference or voltage control element available over a wide spectrum of voltage and power levels.

The zener diode is unique among the semiconductor family of devices because its electrical properties are derived from a rectifying junction which operates in the reverse breakdown region. In the sections that follow, the reverse biased rectifying junction, some of the terms associated with it, and properties derived from it will be discussed fully.

The zener diode is fabricated from the element silicon (Si). Special techniques are applied in the fabrication of zener diodes to create the required properties.

This manual was prepared to acquaint the engineer, the equipment designer and manufacturer, and the experimenter with the fundamental principles, design characteristics, applications and advantages of this important semiconductor device.

SEMICONDUCTOR THEORY

The active portion of a zener diode is a semiconductor PN junction. PN junctions are formed in various kinds of semiconductor devices by several techniques. Among these are the widely used techniques known as alloying and diffusion which are utilized in fabricating zener PN junctions to provide excellent control over zener breakdown voltage.

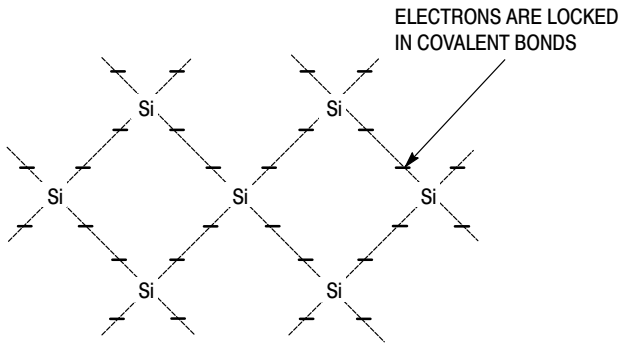
At the present time, zener diodes use silicon as the basic material in the formation of their PN junction. Silicon is in Group IV of the periodic table (tetravalent) and is classed as a “semiconductor” due to the fact that it is a poor conductor in a pure state. When controlled amounts of certain “impurities” are added to a semiconductor it becomes a better conductor of electricity. Depending on the type of impurity added to the basic semiconductor, its conductivity may take two different forms, called P- and N-type respectively.

N-type conductivity in a semiconductor is much like the conductivity due to the drift of free electrons in a metal. In pure silicon at room temperature there are too few free electrons to conduct current. However, there are ways of introducing free electrons into the crystal lattice as we shall now see. Silicon is a tetravalent element, one with four valence electrons in the outer shell; all are virtually locked into place by the covalent bonds of the crystal lattice structure, as shown schematically in Figure 1a. When controlled amounts of donor impurities (Group V elements) such as phosphorus are added, the pentavalent phosphorus atoms entering the lattice structure provide extra electrons not required by the covalent bonds. These impurities are called donor impurities since they “donate” a free electron to the lattice. These donated electrons are free to drift from negative to positive across the crystal when a field is applied, as shown in Figure 1b. The “N” nomenclature for this kind of conductivity implies “negative” charge carriers.

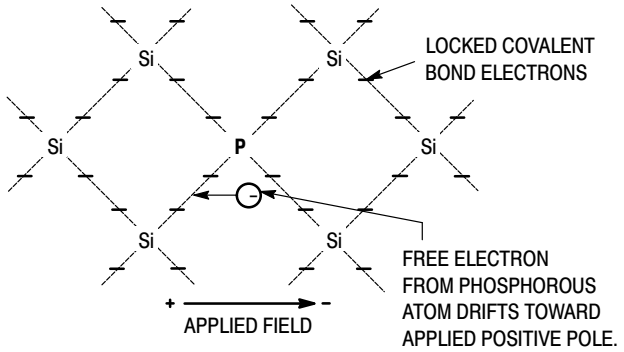
In P-type conductivity, the charges that carry electric current across the crystal act as if they were positive charges. We know that electricity is always carried by drifting electrons in any material, and that there are no mobile positively charged carriers in a solid. Positive charge carriers can exist in gases and liquids in the form of positive ions but not in solids. The positive character of the current flow in the semiconductor crystal may be thought of as the movement of vacancies (called holes) in the covalent lattice. These holes drift from positive toward negative in an electric field, behaving as if they were positive carriers.

P-type conductivity in semiconductors result from adding acceptor impurities (Group III elements) such as boron to silicon to the semiconductor crystal. In this case, boron atoms, with three valence electrons, enter the tetravalent silicon lattice. Since the covalent bonds cannot be satisfied by only three electrons, each acceptor atom leaves a hole in the lattice which is deficient by one electron. These holes readily accept electrons introduced by external sources or created by radiation or heat, as shown in Figure 1c. Hence the name acceptor ion or acceptor impurity. When an external circuit is connected, electrons from the current source “fill up” these holes from the negative end and jump from hole to hole across the crystal or one may think of this process in a slightly different but equivalent way, that is as the displacement of positive holes toward the negative terminal. It is this drift of the positively charged holes which accounts for the term P-type conductivity.

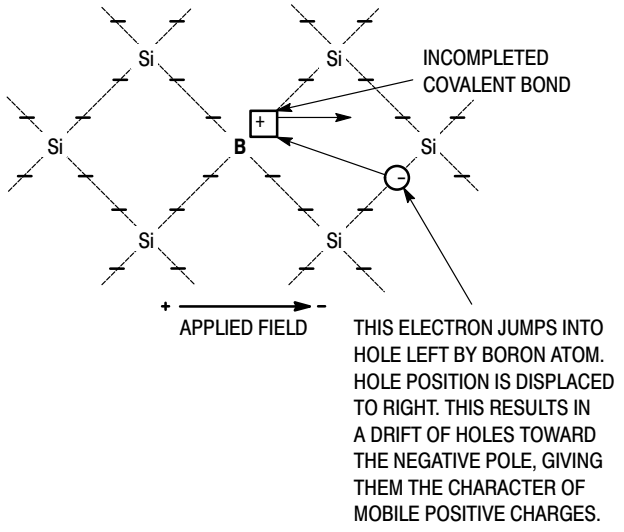
When semiconductor regions of N- and P-type conductivities are formed in a semiconductor crystal adjacent to each other, this structure is called a PN junction. Such a junction is responsible for the action of both zener diodes and rectifier devices, and will be discussed in the next section.



(a) Lattice Structures of Pure Silicon



(b) N-Type Silicon



(c) P-Type Silicon

Figure 1. Semiconductor Structure

THE SEMICONDUCTOR DIODE

In the forward-biased PN junction, Figure 2a, the P region is made more positive than the N region by an external circuit. Under these conditions there is a very low resistance to current flow in the circuit. This is because the holes in the positive P-type material are very readily attracted across the junction interface toward the negative N-type side. Conversely, electrons in the N-type are readily attracted by the positive polarity in the other direction.

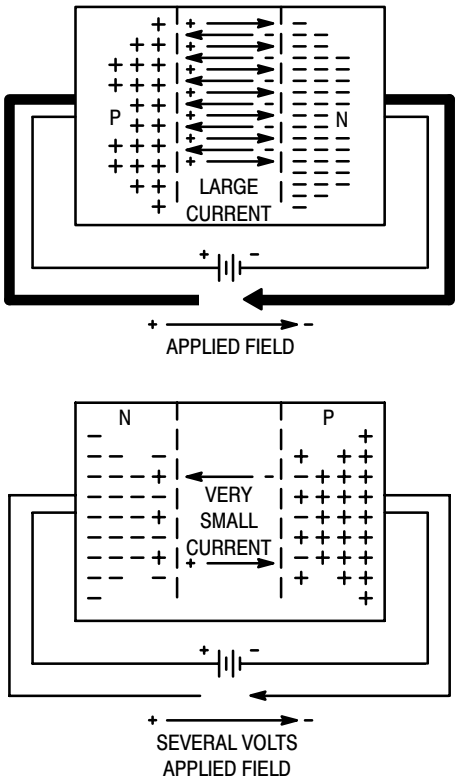
When a PN junction is reverse biased, the P-type side is made more negative than the N-type side. (See Figure 2b.) At voltages below the breakdown of the junction, there is very little current flow across the junction interface. At first thought one would expect no reverse current under reverse bias conditions, but several effects are responsible for this small current.

Under this condition the positive holes in the P-type semiconductor are repelled from the junction interface by the positive polarity applied to the N side, and conversely, the electrons in the N material are repelled from the interface by the negative polarity of the P side. This creates a region extending from the junction interface into both P- and N-type materials which is completely free of charge carriers, that is, the region is depleted of its charge carriers. Hence, this region is usually called the depletion region.

Although the region is free of charge *carriers*, the P-side of the depletion region will have an excess negative charge due to the presence of acceptor ions which are, of course, fixed in the lattice; while the N-side of the depletion region has an excess positive charge due to the presence of donor ions. These opposing regions of charged ions create a strong electric field across the PN junction responsible for the creation of reverse current.

The semiconductor regions are never perfect; there are always a few free electrons in P material and few holes in N material. A more significant factor, however, is the fact that great magnitudes of electron-hole pairs may be thermally generated at room temperatures in the semiconductor. When these electron-hole pairs are created within the depletion region, then the intense electric field mentioned in the above paragraph will cause a small current to flow. This small current is called the reverse saturation current, and tends to maintain a relatively constant value for a fixed temperature at all voltages. The reverse saturation current is usually negligible compared with the current flow when the junction is forward biased. Hence, we see that the PN junction, when not reverse biased beyond breakdown voltage, will conduct heavily in only one direction. When this property is utilized in a circuit we are employing the PN junction as a rectifier. Let us see how we can employ its reverse breakdown characteristics to an advantage.

As the reverse voltage is increased to a point called the voltage breakdown point and beyond, current conduction across the junction interface increases rapidly. The break from a low value of the reverse saturation current to heavy conductance is very sharp and well defined in most PN junctions. It is called the zener knee. When reverse voltages greater than the voltage breakdown point are applied to the PN junction, the voltage drop across the PN junction remains essentially constant at the value of the breakdown voltage for a relatively wide range of currents. This region



CHARGES FROM BOTH P AND N REGIONS DRIFT ACROSS JUNCTION AT VERY LOW APPLIED VOLTAGES.

(a) Forward-Based PN Junction

AT APPLIED VOLTAGES BELOW THE CRITICAL BREAKDOWN LEVEL ONLY A FEW CHARGES DRIFT ACROSS THE INTERFACE.

(b) Reverse-Biased PN Junction

Figure 2. Effects of Junction Bias

beyond the voltage breakdown point is called the zener control region.

ZENER CONTROL REGION: VOLTAGE BREAKDOWN MECHANISMS

Figure 3 depicts the extension of reverse biasing to the point where voltage breakdown occurs. Although all PN junctions exhibit a voltage breakdown, it is important to know that there are two distinct voltage breakdown mechanisms. One is called *zener breakdown* and the other is called *avalanche breakdown.* In zener breakdown the value of breakdown voltage decreases as the PN junction temperature increases; while in avalanche breakdown the value of the breakdown voltage increases as the PN junction temperature increases. Typical diode breakdown characteristics of each category are shown in Figure 4. The factor determining which of the two breakdown mechanisms occurs is the relative concentrations of the impurities in the materials which comprise the junction. If two different resistivity P-type materials are placed against two separate but equally doped low-resistivity pieces of N-type materials, the depletion region spread in the low resistivity P-type material will be smaller than the depletion region spread in the high resistivity P-type material. Moreover, in both situations little of the resultant depletion width lies in the N material if its resistivity is low compared to the P-type material. In other words, the depletion region always spreads principally into the material having the highest resistivity. Also, the electric field (voltage per unit

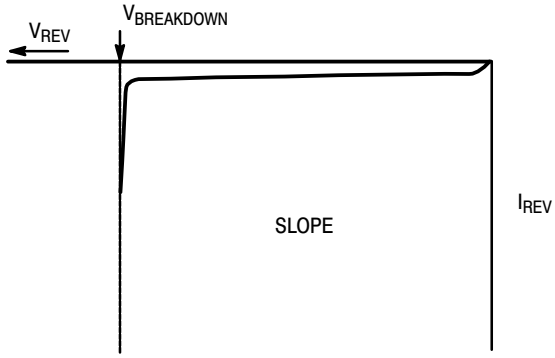


Figure 3. Reverse Characteristic Extended to Show Breakdown Effect

length) in the less resistive material is greater than the electric field in the material of greater resistivity due to the presence of more ions/unit volume in the less resistive material. A junction that results in a narrow depletion region will therefore develop a high field intensity and breakdown by the zener mechanism. A junction that results in a wider depletion region and, thus, a lower field intensity will break down by the avalanche mechanism before a zener breakdown condition can be reached.

The zener mechanism can be described qualitatively as follows: because the depletion width is very small, the application of low reverse bias (5 volts or less) will cause a field across the depletion region on the order of 3×10^5V/cm. A field of such high magnitude exerts a large force on the

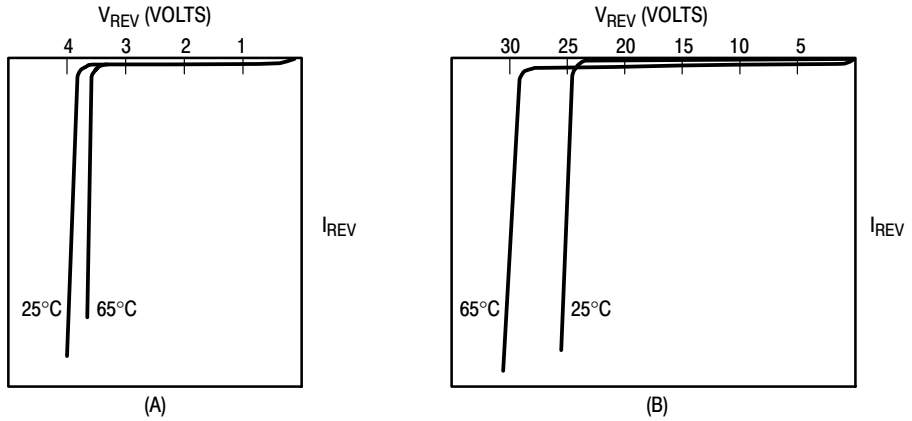


Figure 4. Typical Breakdown Diode Characteristics. Note Effects of Temperature for Each Mechanism

valence electrons of a silicon atom, tending to separate them from their respective nuclei. Actual rupture of the covalent bonds occurs when the field approaches 3 x 10^5V/cm. Thus, electron-hole pairs are generated in large numbers and a sudden increase of current is observed. Although we speak of a rupture of the atomic structure, it should be understood that this generation of electron-hole pairs may be carried on continuously as long as an external source supplies additional electrons. If a limiting resistance in the circuit external to the diode junction does not prevent the current from increasing to high values, the device may be destroyed due to overheating. The actual critical value of field causing zener breakdown is believed to be approximately 3 x 10^5V/cm. On most commercially available silicon diodes, the maximum value of voltage breakdown by the zener mechanism is 8 volts. In order to fabricate devices with higher voltage breakdown characteristics, materials with higher resistivity, and consequently, wider depletion regions are required. These wide depletion regions hold the field strength down below the zener breakdown value (3 x 10^5V/cm). Consequently, for devices with breakdown voltage lower than 5 volts the zener mechanism predominates, between 5 and 8 volts both zener and an avalanche mechanism are involved, while above 8 volts the avalanche mechanism alone takes over.

The decrease of zener breakdown voltage as junction temperature increases can be explained in terms of the energies of the valence electrons. An increase of temperature increases the energies of the valence electrons. This weakens the bonds holding the electrons and consequently, less applied voltage is necessary to pull the valence electrons from their position around the nuclei. Thus, the breakdown voltage decreases as the temperature increases.

The dependence on temperature of the avalanche breakdown mechanism is quite different. Here the depletion region is of sufficient width that the carriers (electrons or holes) can suffer collisions before traveling the region completely i.e., the depletion region is wider than one mean-free path (the average distance a carrier can travel before combining with a carrier of opposite conductivity). Therefore, when temperature is increased, the increased lattice vibration shortens the distance a carrier travels before colliding and thus requires a higher voltage to get it across the depletion region.

As established earlier, the applied reverse bias causes a small movement of intrinsic electrons from the P material to the potentially positive N material and intrinsic holes from the N material to the potentially negative P material (leakage current). As the applied voltage becomes larger, these electrons and holes increasingly accelerate. There are also collisions between these intrinsic particles and bound electrons as the intrinsic particles move through the depletion region. If the applied voltage is such that the intrinsic electrons do not have high velocity, then the collisions take some energy from the intrinsic particles, altering their velocity. If the applied voltage is increased, collision with a valence electron will give considerable energy to the electron and it will break free of its covalent bond. Thus, one electron by collision, has created an electron-hole pair. These secondary particles will also be accelerated and participate in collisions which generate new electron-hole pairs. This phenomenon is called carrier multiplication. Electron-hole pairs are generated so quickly and in such large numbers that there is an apparent avalanche or self-sustained multiplication process (depicted graphically in Figure 5). The junction is said to be in breakdown and the current is limited only by resistance external to the junction. Zener diodes above 7 to 8 volts exhibit avalanche breakdown.

As junction temperature increases, the voltage breakdown point for the avalanche mechanism increases. This effect can be explained by considering the vibration displacement of atoms in their lattice increases, and this increased displacement corresponds to an increase in the probability that intrinsic particles in the depletion region will collide with the lattice atoms. If the probability of an intrinsic particle-atom collision increases, then the probability that a given intrinsic particle will obtain high momentum

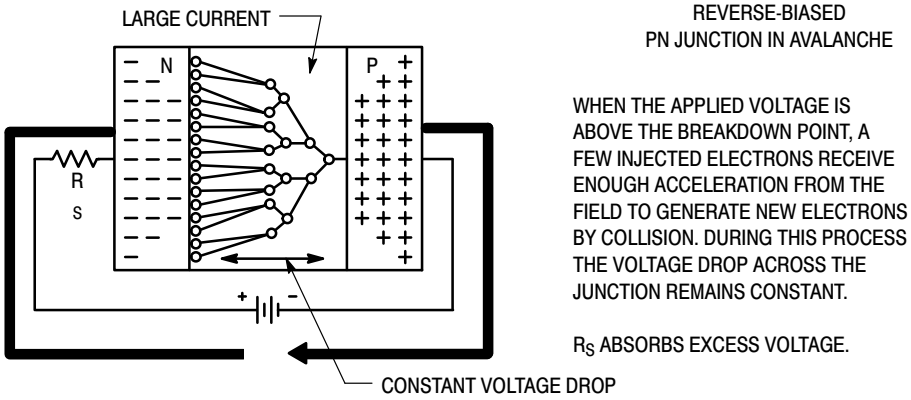


Figure 5. PN Junction in Avalanche Breakdown

decreases, and it follows that the low momentum intrinsic particles are less likely to ionize the lattice atoms. Naturally, increased voltage increases the acceleration of the intrinsic particles, providing higher mean momentum and more electron-hole pairs production. If the voltage is raised sufficiently, the mean momentum becomes great enough to create electron-hole pairs and carrier multiplication results. Hence, for increasing temperature, the value of the avalanche breakdown voltage increases.

VOLT-AMPERE CHARACTERISTICS

The zener volt-ampere characteristics for a typical 30 volt zener diode is illustrated in Figure 6. It shows that the zener diode conducts current in both directions; the forward current I_F being a function of forward voltage V_F. Note that I_F is small until $V_F \approx 0.65$ V; then I_F increases very rapidly. For $V_F > 0.65$ V I_F is limited primarily by the circuit resistance external to the diode.

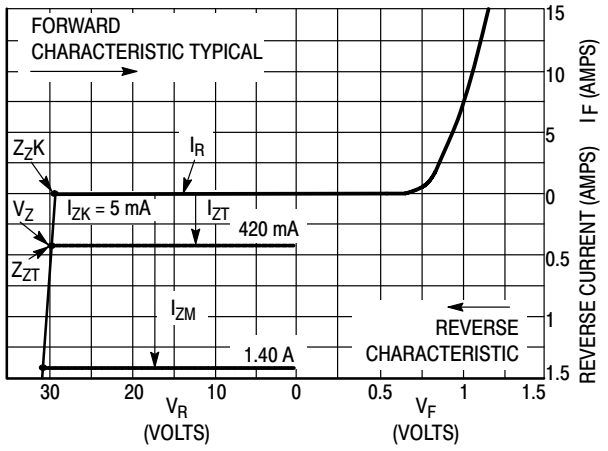


Figure 6. Zener Diode Characteristics

The reverse current is a function of the reverse voltage V_R but for most practical purposes is zero until the reverse voltage approaches V_Z, the PN junction breakdown voltage, at which time the reverse current increases very rapidly. Since the reverse current is small for $V_R < V_Z$, but great for $V_R > V_Z$ each of the current regions is specified by a different symbol. For the leakage current region, i.e. non-conducting region, between 0 volts and V_Z, the reverse current is denoted by the symbol I_R; but for the zener control region, $V_R \geq V_Z$, the reverse current is denoted by the symbol I_Z. I_R is usually specified at a reverse voltage $V_R \approx 0.8\ V_Z$.

The PN junction breakdown voltage, V_Z, is usually called the zener voltage, regardless whether the diode is of the zener or avalanche breakdown type. Commercial zener diodes are available with zener voltages from about 1.8 V − 400 V. For most applications the zener diode is operated well into the breakdown region (I_{ZT} to I_{ZM}). Most manufacturers give an additional specification of I_{ZK} (= 5 mA in Figure 6) to indicate a minimum operating current to assure reasonable regulation.

This minimum current I_{ZK} varies in the various types of zener diodes and, consequently, is given on the data sheets. The maximum zener current I_{ZM} should be considered the maximum reverse current recommended by the manufacturer. Values of I_{ZM} are usually given in the data sheets.

Between the limits of I_{ZK} and I_{ZM}, which are 5 mA and 1400 mA (1.4 Amps) in the example of Figure 6, the voltage across the diode is essentially constant, and $\approx V_Z$. This plateau region has, however, a large positive slope such that the precise value of reverse voltage will change slightly as a function of I_Z. For any point on this plateau region one may calculate an impedance using the incremental magnitudes of the voltage and current. This impedance is usually called the zener impedance Z_Z, and is specified for most zener diodes. Most manufacturers measure the maximum zener impedance at two test points on the plateau region. The first is usually near the knee of the zener plateau, Z_{ZK}, and the latter point near the midrange of the usable zener current excursion. Two such points are illustrated in Figure 6.

This section was intended to introduce the reader to a few of the major terms used with zener diodes. A complete description of these terms may be found in the zener diode characteristic section. In this section a full discussion of zener leakage, DC breakdown, zener impedance, temperature coefficients and many other topics may be found.

ZENER DIODE FABRICATION TECHNIQUES

INTRODUCTION

A brief exposure to the techniques used in the fabrication of zener diodes can provide the engineer with additional insight using zeners in their applications. That is, an understanding of zener fabrication makes the capabilities and limitations of the zener diode more meaningful. This section discusses the basic steps in the fabrication of the zener from crystal growing through final testing.

ZENER DIODE WAFER FABRICATION

The major steps in the manufacture of zeners are provided in the process flow in Figure 1. It is important to point out that the manufacturing steps vary somewhat from manufacturer to manufacturer, and also vary with the type of zener diode produced. This is driven by the type of package required as well as the electrical characteristics desired. For example, alloy diffused devices provide excellent low voltage reference with low leakage characteristics but do not have the same surge carrying capability as diffused diodes. The manufacturing process begins with the growing of high quality silicon crystals.

Crystals for ON Semiconductor zener diodes are grown using the Czochralski technique, a widely used process which begins with ultra-pure polycrystalline silicon. The polycrystalline silicon is first melted in a nonreactive crucible held at a temperature just above the melting point. A carefully controlled quantity of the desired dopant impurity, such as phosphorus or boron is added. A high quality seed crystal of the desired crystalline orientation is then lowered into the melt while rotating. A portion of this seed crystal is allowed to melt into the molten silicon. The seed is then slowly pulled and continues to rotate as it is raised from the melt. As the seed is raised, cooling takes place and material from the melt adheres to it, thus forming a single crystal ingot. With this technique, ingots with diameters of several inches can be fabricated.

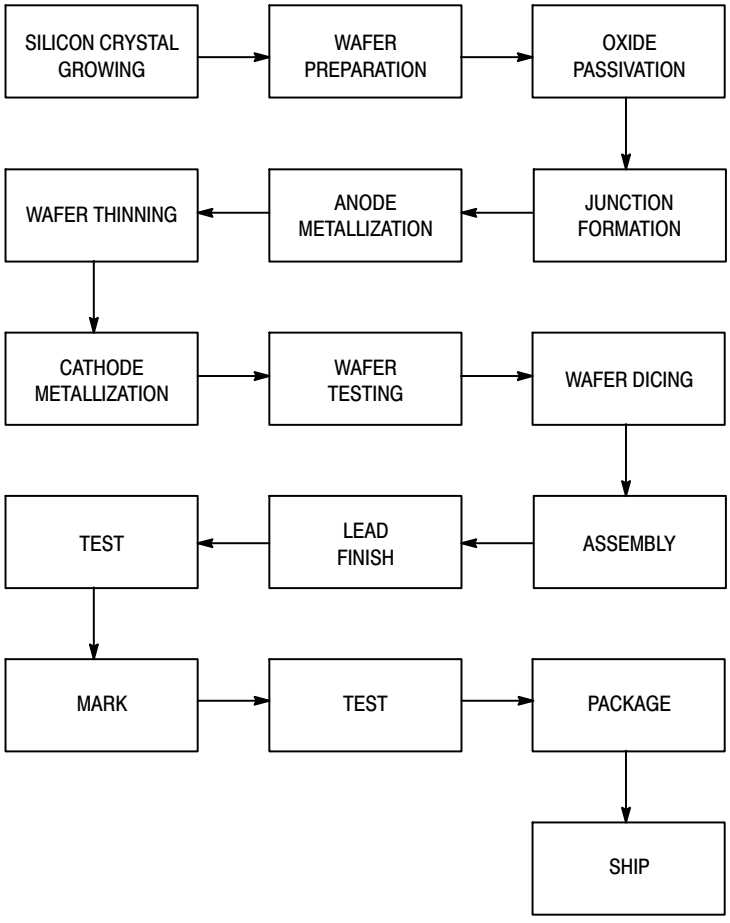


Figure 1. General Flow of the Zener Diode Process

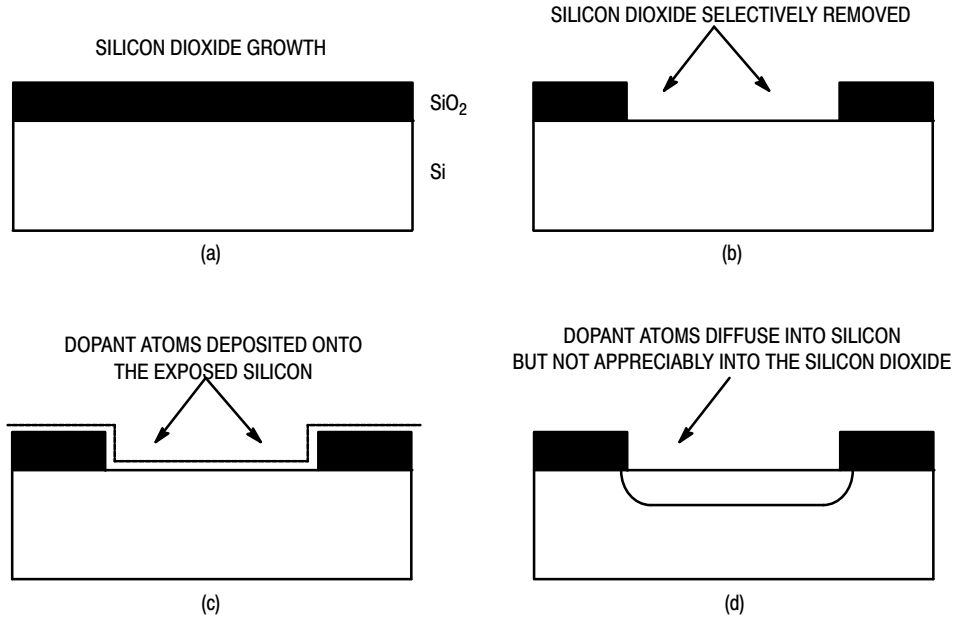


Figure 2. Basic Fabrication Steps in the Silicon Planar Process: a) oxide formation, b) selective oxide removal, c) deposition of dopant atoms, d) junction formation by diffusion of dopant atoms.

Once the single-crystal silicon ingot is grown, it is tested for doping concentration (resistivity), undesired impurity levels, and minority carrier lifetime. The ingot is then sliced into thin, circular wafers. The wafers are then chemically etched to remove saw damage and polished in a sequence of successively finer polishing grits until a mirror-like defect free surface is obtained. The wafers are then cleaned and placed in vacuum sealed wafer carriers to prevent any contamination from getting on them. At this point, the wafers are ready to begin device fabrication.

Zener diodes can be manufactured using different processing techniques such as planar processing or mesa etched processing. The majority of ON Semiconductor zener diodes are manufactured using the planar technique as shown in Figure 2.

The planar process begins by growing an ultra-clean protective silicon dioxide passivation layer. The oxide is typically grown in the temperature range of 900 to 1200 degrees celcius. Once the protective layer of silicon dioxide has been formed, it must be selectively removed from those areas into which dopant atoms will be introduced. This is done using photolithographic techniques.

First a light sensitive solution called photo resist is spun onto the wafer. The resist is then dried and a photographic negative or mask is placed over the wafer. The resist is then exposed to ultraviolet light causing the molecules in it to cross link or polymerize becoming very rigid. Those areas of the wafer that are protected by opaque portions of the mask are not exposed and are developed away. The oxide is then etched forming the exposed regions in which the dopant will be introduced. The remaining resist is then removed and the wafers carefully cleaned for the doping steps.

Dopant is then introduced onto the wafer surface using various techniques such as aluminum alloy for low voltage devices, ion-implantation, spin-on dopants, or chemical vapor deposition. Once the dopant is deposited, the junctions are formed in a subsequent high temperature (1100 to 1250 degrees celcius are typical) drive-in. The resultant junction profile is determined by the background concentration of the starting substrate, the amount of dopant placed at the surface, and amount of time and temperature used during the dopant drive-in. This junction profile determines the electrical characteristics of the device. During the drive-in cycle, additional passivation oxide is grown providing additional protection for the devices.

After junction formation, the wafers are then processed through what is called a getter process. The getter step utilizes high temperature and slight stress provided by a highly doped phosphosilicate glass layer introduced into the backside of the wafers. This causes any contaminants in the area of the junction to diffuse away from the region. This serves to improve the reverse leakage characteristic and the stability of the device. Following the getter process, a second photo resist step opens the contact area in which the anode metallization is deposited.

Metal systems for ON Semiconductor's zener diodes are determined by the requirements of the package. The metal systems are deposited in ultra-clean vacuum chambers utilizing electron-beam evaporation techniques. Once the metal is deposited, photo resist processing is utilized to form the desired patterns. The wafers are then lapped to their final thickness and the cathode metallization deposited using the same e-beam process.

The quality of the wafers is closely monitored throughout the process by using statistical process control techniques and careful microscopic inspections at critical steps. Special wafer handling equipment is used throughout the manufacturing process to minimize contamination and to avoid damaging the wafers in any way. This further enhances the quality and stability of the devices.

Upon completion of the fabrication steps, the wafers are electrically probed, inspected, and packaged for shipment to the assembly operations. All ON Semiconductor zener diode product is sawn using 100% saw-through techniques stringently developed to provide high quality silicon die.

ZENER DIODE ASSEMBLY

Surmetic 40

The Surmetic 40 plastic package is assembled using oxygen free high conductivity copper leads for efficient heat transfer from the die and allowing maximum power dissipation with a minimum of external heatsinking. Figure 3 shows typical assembly. The leads are of nail head construction, soldered directly to the die, which further enhances the heat dissipating capabilities of the package.

Assembly is started on the Surmetic 40 by loading the leads into assembly boats and pre-soldering the nail heads. After pre-soldering, one die is then placed into each cavity of one assembly boat and another assembly boat is then mated to it.

After assembly, the leads on the Surmetic 40 are plated with a tin-lead alloy making them readily solderable and corrosion resistant.

Double Slug (DO-35 and DO-41)

Double slugs receive their name from the dumet slugs, one attached to one end of each lead. These slugs sandwich the die between them and are hermetically sealed to the glass envelope or body during assembly. Figure 4 shows typical assembly.

The assembly begins with the copper clad steel leads being loaded into assembly “boats.” Every other boat load of leads has a glass body set over the slug. A die is placed into each glass body and the other boat load of leads is mated to the boat holding the leads, body and die. These mated boats are then placed into the assembly furnace where the total mass is heated. Each glass body melts; and as the boat proceeds through the cooling portion of the furnace chamber, the compression solidifies forming a bond between the die and both slugs. The glass hardens, attaching itself to the sides of the two slugs forming the hermetic seal. The above illustrates how the diodes are completely assembled using a single furnace pass minimizing assembly problems.

The encapsulated devices are then processed through lead finish. This consists of dipping the leads in molten tin/lead solder alloy. The solder dipped leads produce an external finish which is tarnish-resistant and very solderable.

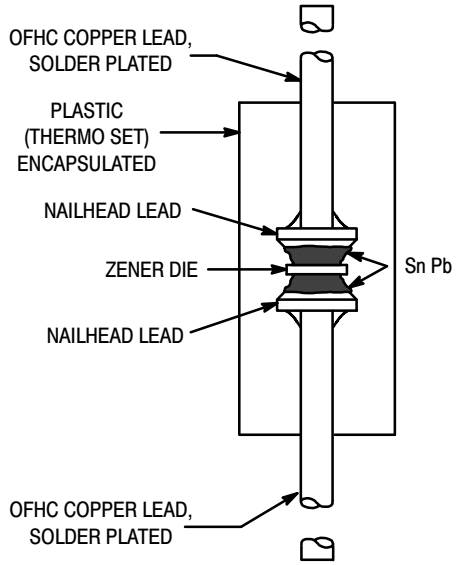


Figure 3. Double-Slug Plastic Zener Construction

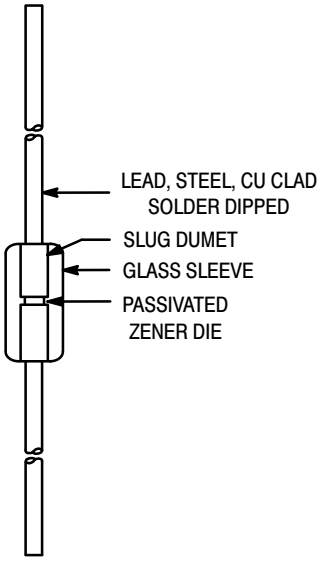


Figure 4. Double Slug Glass Zener Construction

ZENER DIODE TEST, MARK AND PACKAGING

Double Slug, Surmetic 40

After lead finish, all products are final tested, whether they are double slug or of Surmetic construction, all are 100 percent final tested for zener voltage, leakage current, impedance and forward voltage drop.

Process average testing is used which is based upon the averages of the previous lots for a given voltage line and package type. Histograms are generated for the various parameters as the units are being tested to ensure that the lot is testing well to the process average and compared against other lots of the same voltage.

After testing, the units are marked as required by the specification. The markers are equipped to polarity orient the devices as well as perform 100% redundant test prior to packaging.

After marking, the units are packaged either in “bulk” form or taped and reeled or taped and ammo packed to accommodate automatic insertion.

RELIABILITY

INTRODUCTION

ON Semiconductor's Quality System maintains "continuous product improvement" goals in all phases of the operation. Statistical process control (SPC), quality control sampling, reliability audits and accelerated stress testing techniques monitor the quality and reliability of its products. Management and engineering skills are continuously upgraded through training programs. This maintains a unified focus on Six Sigma quality and reliability from the inception of the product to final customer use.

STATISTICAL PROCESS CONTROL

ON Semiconductor's Discrete Group is continually pursuing new ways to improve product quality. Initial design improvement is one method that can be used to produce a superior product. Equally important to outgoing product quality is the ability to produce product that consistently conforms to specification. Process variability is the basic enemy of semiconductor manufacturing since it leads to product variability. Used in all phases of ON Semiconductor's product manufacturing, STATISTICAL PROCESS CONTROL (SPC) replaces variability with predictability. The traditional philosophy in the semiconductor industry has been adherence to the data sheet specification. Using SPC methods assures the product will meet specific process requirements throughout the manufacturing cycle. The emphasis is on defect prevention, not detection. Predictability through SPC methods requires the manufacturing culture to focus on constant and permanent improvements. Usually these improvements cannot be bought with state-of-the-art equipment or automated factories. With quality in design, process and material selection, coupled with manufacturing predictability, ON Semiconductor can produce world class products.

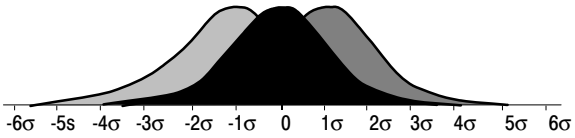
The immediate effect of SPC manufacturing is predictability through process controls. Product centered and distributed well within the product specification benefits ON Semiconductor with fewer rejects, improved yields and lower cost. The direct benefit to ON Semiconductor's customers includes better incoming quality levels, less inspection time and ship-to-stock capability. Circuit performance is often dependent on the cumulative effect of component variability. Tightly controlled component distributions give the customer greater circuit predictability. Many customers are also converting to just-in-time (JIT) delivery programs. These programs require improvements in cycle time and yield predictability achievable only through SPC techniques. The benefit derived from SPC helps the manufacturer meet the customer's expectations of higher quality and lower cost product.

Ultimately, ON Semiconductor will have Six Sigma capability on all products. This means parametric distributions will be centered within the specification limits with a product distribution of plus or minus Six Sigma about mean. Six Sigma capability, shown graphically in Figure 1, details the benefit in terms of yield and outgoing quality levels. This compares a centered distribution versus a 1.5 sigma worst case distribution shift.

New product development at ON Semiconductor requires more robust design features that make them less sensitive to minor variations in processing. These features make the implementation of SPC much easier.

A complete commitment to SPC is present throughout ON Semiconductor. All managers, engineers, production operators, supervisors and maintenance personnel have received multiple training courses on SPC techniques. Manufacturing has identified 22 wafer processing and 8 assembly steps considered critical to the processing of zener products. Processes, controlled by SPC methods, that have shown significant improvement are in the diffusion, photolithography and metallization areas.

To better understand SPC principles, brief explanations have been provided. These cover process capability, implementation and use.

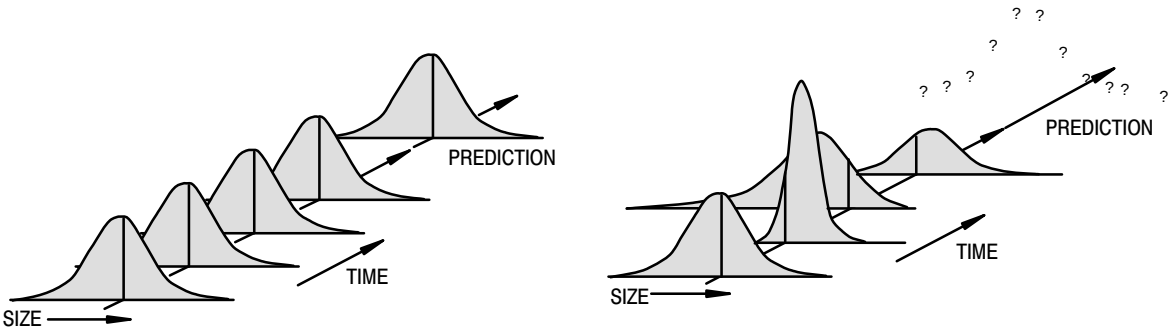


Distribution Centered	Distribution Shifted ± 1.5
At ± 3 σ 2700 ppm defective 99.73% yield	66810 ppm defective 93.32% yield
At ± 4 σ 63 ppm defective 99.9937% yield	6210 ppm defective 99.379% yield
At ± 5 σ 0.57 ppm defective 99.999943% yield	233 ppm defective 99.9767% yield
At ± 6 σ 0.002 ppm defective 99.9999998% yield	3.4 ppm defective 99.99966% yield

Figure 1. AOQL and Yield from a Normal Distribution of Product With 6σ Capability

PROCESS CAPABILITY

One goal of SPC is to ensure a process is **CAPABLE**. Process capability is the measurement of a process to produce products consistently to specification requirements. The purpose of a process capability study is to separate the inherent **RANDOM VARIABILITY** from **ASSIGNABLE CAUSES**. Once completed, steps are taken to identify and eliminate the most significant assignable causes. Random variability is generally present in the system and does not fluctuate. Sometimes, these are considered basic limitations associated with the machinery, materials, personnel skills or manufacturing methods. Assignable cause inconsistencies relate to time variations in yield, performance or reliability.



Process “under control” − all assignable causes are removed and future distribution is predictable.

Figure 2. Impact of Assignable Causes on Process Predictable

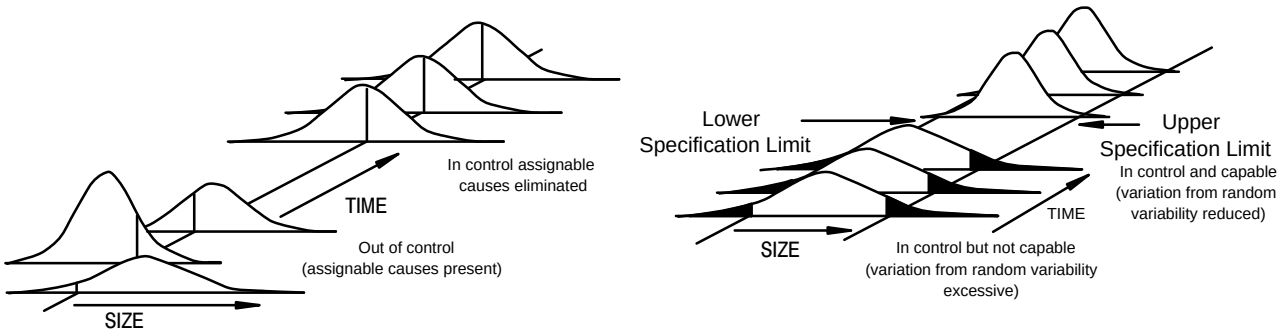


Figure 3. Difference Between Process Control and Process Capability

Traditionally, assignable causes appear to be random due to the lack of close examination or analysis. Figure 2 shows the impact on predictability that assignable cause can have. Figure 3 shows the difference between process control and process capability.

A process capability study involves taking periodic samples from the process under controlled conditions. The performance characteristics of these samples are charted against time. In time, assignable causes can be identified and engineered out. Careful documentation of the process is key to accurate diagnosis and successful removal of the assignable causes. Sometimes, the assignable causes will remain unclear requiring prolonged experimentation.

Elements which measure process variation control and capability are Cp and Cpk respectively. Cp is the specification width divided by the process width or Cp = (specification width) / 6σ. Cpk is the absolute value of the closest specification value to the mean, minus the mean, divided by half the process width or Cpk = | closest specification — $\overline{X}$ / 3σ.

At ON Semiconductor, for critical parameters, the process capability is acceptable with a Cpk = 1.33. The desired process capability is a Cpk = 2 and the ideal is a Cpk = 5. Cpk, by definition, shows where the current production process fits with relationship to the specification limits. Off center distributions or excessive process variability will result in less than optimum conditions.

SPC IMPLEMENTATION AND USE

The Discrete Group uses many parameters that show conformance to specification. Some parameters are sensitive to process variations while others remain constant for a given product line. Often, specific parameters are influenced when changes to other parameters occur. It is both impractical and unnecessary to monitor all parameters using SPC methods. Only critical parameters that are sensitive to process variability are chosen for SPC monitoring. The process steps affecting these critical parameters must be identified also. It is equally important to find a measurement in these process steps that correlates with product performance. This is called a critical process parameter.

Once the critical process parameters are selected, a sample plan must be determined. The samples used for measurement are organized into **RATIONAL SUBGROUPS** of approximately 2 to 5 pieces. The subgroup size should be such that variation among the samples within the subgroup remain small. All samples must come from the same source e.g., the same mold press operator, etc.. Subgroup data should be collected at appropriate time intervals to detect variations in the process. As the process begins to show improved stability, the interval may be increased. The data collected must be carefully documented and maintained for later correlation. Examples of common documentation entries would include operator, machine, time, settings, product type, etc..

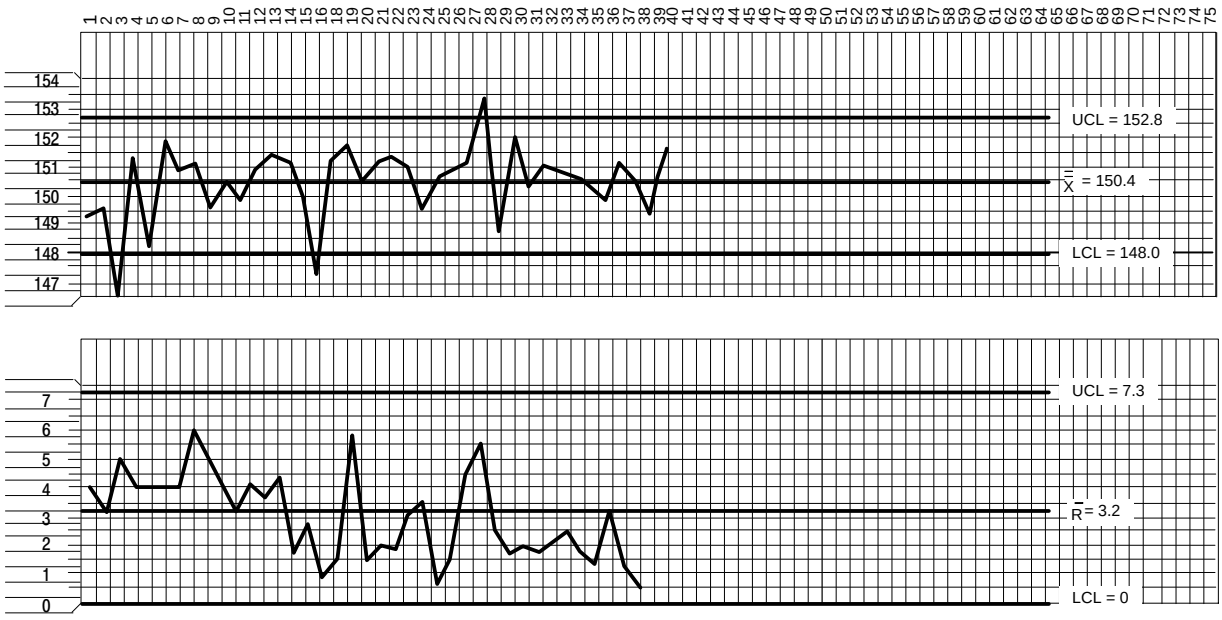


Figure 4. Example of Process Control Chart Showing Oven Temperature Data

Once the plan is established, data collection may begin. The data collected will generate $\overline{X}$ and R values that are plotted with respect to time. $\overline{X}$ refers to the mean of the values within a given subgroup, while R is the range or greatest value minus least value. When approximately 20 or more $\overline{X}$ and R values have been generated, the average of these values is computed as follows:

$$\overline{\overline{X}} = (\overline{X} + \overline{X}2 + \overline{X}3 + ...)/K$$
$$\overline{R} = (R1 + R2 + R3 + ...)/K$$

where K = the number of subgroups measured.

The values of $\overline{\overline{X}}$ and $\overline{R}$ are used to create the process control chart. Control charts are the primary SPC tool used to signal a problem. Shown in Figure 4, process control charts show $\overline{X}$ and R values with respect to time and concerning reference to upper and lower control limit values. Control limits are computed as follows:

R upper control limit = $UCL_R = D4\ \overline{R}$
R lower control limit $LCL_R = D3\ \overline{R}$
$\overline{X}$ upper control limit = $UCL_X = \overline{\overline{X}} + A2\ \overline{R}$
$\overline{X}$ lower control limit = $LCL_X = \overline{\overline{X}} - A$

Where D4, D3 and A2 are constants varying by sample size, with values for sample sizes from 2 to 10 shown in the following partial table:

Control charts are used to monitor the variability of critical process parameters. The R chart shows basic problems with piece to piece variability related to the process. The X chart can often identify changes in people, machines, methods, etc. The source of the variability can be difficult to find and may require experimental design techniques to identify assignable causes.

Some general rules have been established to help determine when a process is **OUT-OF-CONTROL**. Figure 5a shows a control chart subdivided into zones A, B, and C corresponding to 3 sigma, 2 sigma, and 1 sigma limits respectively. In Figure 5b through Figure 5e four of the tests that can be used to identify excessive variability and the presence of assignable causes are shown. As familiarity with a given process increases, more subtle tests may be employed successfully.

Once the variability is identified, the cause of the variability must be determined. Normally, only a few factors have a significant impact on the total variability of the process. The importance of correctly identifying these factors is stressed in the following example. Suppose a process variability depends on the variance of five factors A, B, C, D and E. Each has a variance of 5, 3, 2, 1 and 0.4 respectively.

Since:

$$\sigma\ tot = \sqrt{\sigma A^2 + \sigma B^2 + \sigma C^2 + \sigma D^2 + \sigma E^2}$$
$$\sigma\ tot = \sqrt{5^2 + 3^2 + 2^2 + 1^2 + (0.4)^2} = 6.3$$

n	2	3	4	5	6	7	8	9	10
D_4	3.27	2.57	2.28	2.11	2.00	1.92	1.86	1.82	1.78
D_3	*	*	*	*	*	0.08	0.14	0.18	0.22
A_2	1.88	1.02	0.73	0.58	0.48	0.42	0.37	0.34	0.31

* For sample sizes below 7, the LCL_R would technically be a negative number; in those cases there is no lower control limit; this means that for a subgroup size 6, six “identical” measurements would not be unreasonable.

Now if only D is identified and eliminated then;

$$s\ tot = \sqrt{5^2 + 3^2 + 2^2 + (0.4)^2} = 6.2$$

This results in less than 2% total variability improvement. If B, C and D were eliminated, then;

$$\sigma\ tot = \sqrt{5^2 + (0.4)^2} = 5.02$$

This gives a considerably better improvement of 23%. If only A is identified and reduced from 5 to 2, then;

$$\sigma\ tot = \sqrt{2^2 + 3^2 + 2^2 + 1^2 + (0.4)^2} = 4.3$$

Identifying and improving the variability from 5 to 2 gives us a total variability improvement of nearly 40%.

Most techniques may be employed to identify the primary assignable cause(s). Out-of-control conditions may be correlated to documented process changes. The product may be analyzed in detail using best versus worst part comparisons or Product Analysis Lab equipment. Multi-variance analysis can be used to determine the family of variation (positional, critical or temporal). Lastly, experiments may be run to test theoretical or factorial analysis. Whatever method is used, assignable causes must be identified and eliminated in the most expeditious manner possible.

After assignable causes have been eliminated, new control limits are calculated to provide a more challenging variability criteria for the process. As yields and variability improve, it may become more difficult to detect improvements because they become much smaller. When all assignable causes have been eliminated and the points remain within control limits for 25 groups, the process is said to be in a state of control.

SUMMARY

ON Semiconductor is committed to the use of STATISTICAL PROCESS CONTROLS. These principles, used throughout manufacturing, have already resulted in many significant improvements to the processes. Continued dedication to the SPC culture will allow ON Semiconductor to reach the Six Sigma and zero defect capability goals. SPC will further enhance the commitment to **TOTAL CUSTOMER SATISFACTION**.

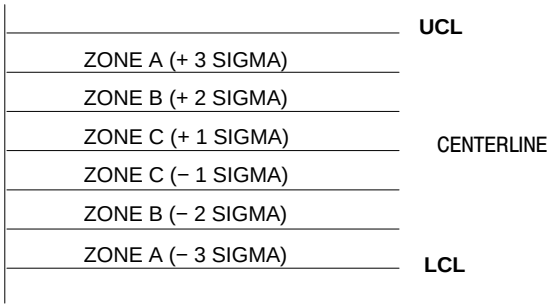


Figure 5a. Control Chart Zones

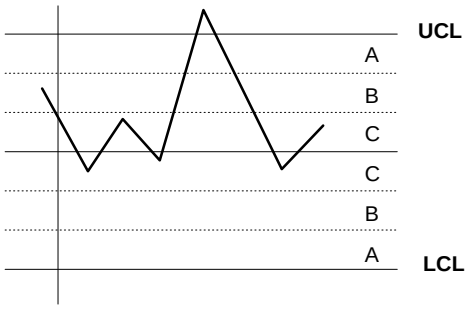


Figure 5b. One Point Outside Control Limit Indicating Excessive Variability

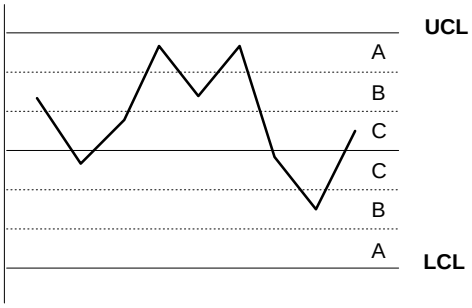


Figure 5c. Two Out of Three Points in Zone A or Beyond Indicating Excessive Variability

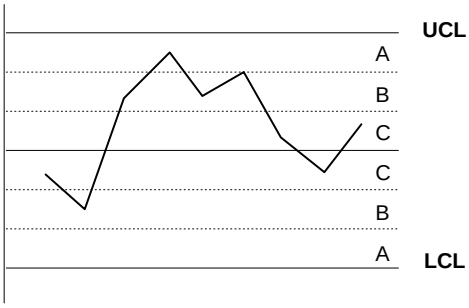


Figure 5d. Four Out of Five Points in Zone B or Beyond Indicating Excessive Variability

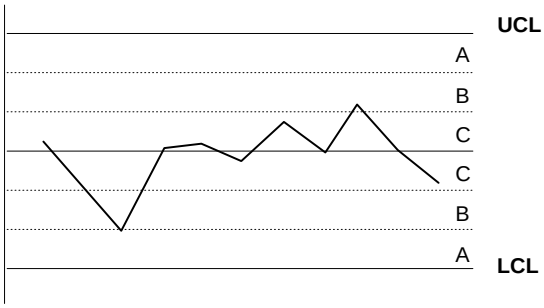


Figure 5e. Seven Out of Eight Points in Zone C or Beyond Indicating Excessive Variability

RELIABILITY STRESS TESTS

The following gives brief descriptions of the reliability tests commonly used in the reliability monitoring program. Not all of the tests listed are performed on each product. Other tests may be performed when appropriate. In addition some form of preconditioning may be used in conjunction with the following tests.

AUTOCLAVE (aka, PRESSURE COOKER)

Autoclave is an environmental test which measures device resistance to moisture penetration and the resultant effects of galvanic corrosion. Autoclave is a highly accelerated and destructive test.

Typical Test Conditions: T_A = 121°C, rh = 100%, p = 1 atmosphere (15 psig), t = 24 to 96 hours

Common Failure Modes: Parametric shifts, high leakage and/or catastrophic

Common Failure Mechanisms: Die corrosion or contaminants such as foreign material on or within the package materials. Poor package sealing

HIGH HUMIDITY HIGH TEMPERATURE BIAS (H3TB or H3TRB)

This is an environmental test designed to measure the moisture resistance of plastic encapsulated devices. A bias is applied to create an electrolytic cell necessary to accelerate corrosion of the die metallization. With time, this is a catastrophically destructive test.

Typical Test Conditions: T_A = 85°C to 95°C, rh = 85% to 95%, Bias = 80% to 100% of Data Book max. rating, t = 96 to 1750 hours

Common Failure Modes: Parametric shifts, high leakage and/or catastrophic

Common Failure Mechanisms: Die corrosion or contaminants such as foreign material on or within the package materials. Poor package sealing

Military Reference: MIL-STD-750, Method 1042

HIGH TEMPERATURE REVERSE BIAS (HTRB)

The purpose of this test is to align mobile ions by means of temperature and voltage stress to form a high-current leakage path between two or more junctions.

Typical Test Conditions: T_A = 85°C to 150°C, Bias = 80% to 100% of Data Book max. rating, t = 120 to 1000 hours

Common Failure Modes: Parametric shifts in leakage

Common Failure Mechanisms: Ionic contamination on the surface or under the metallization of the die

Military Reference: MIL-STD-750, Method 1039

HIGH TEMPERATURE STORAGE LIFE (HTSL)

High temperature storage life testing is performed to accelerate failure mechanisms which are thermally activated through the application of extreme temperatures.

Typical Test Conditions: T_A = 70°C to 200°C, no bias, t = 24 to 2500 hours

Common Failure Modes: Parametric shifts in leakage

Common Failure Mechanisms: Bulk die and diffusion defects

Military Reference: MIL-STD-750, Method 1032

INTERMITTENT OPERATING LIFE (IOL)

The purpose of this test is the same as SSOL in addition to checking the integrity of both wire and die bonds by means of thermal stressing.

Typical Test Conditions: T_A = 25°C, Pd = Data Book maximum rating, $T_{on} = T_{off} = \Delta$ of 50°C to 100°C, t = 42 to 30000 cycles

Common Failure Modes: Parametric shifts and catastrophic

Common Failure Mechanisms: Foreign material, crack and bulk die defects, metallization, wire and die bond defects

Military Reference: MIL-STD-750, Method 1037

MECHANICAL SHOCK

This test is used to determine the ability of the device to withstand a sudden change in mechanical stress due to abrupt changes in motion as seen in handling, transportation, or actual use.

Typical Test Conditions: Acceleration = 1500 g's, Orientation = X_1, Y_1, Y_2 plane, t = 0.5 msec, Blows = 5

Common Failure Modes: Open, short, excessive leakage, mechanical failure

Common Failure Mechanisms: Die and wire bonds, cracked die, package defects

Military Reference: MIL-STD-750, Method 2015

MOISTURE RESISTANCE

The purpose of this test is to evaluate the moisture resistance of components under temperature/humidity conditions typical of tropical environments.

Typical Test Conditions: T_A = −10°C to 65°C, rh = 80% to 98%, t = 24 hours/cycle, cycle = 10

Common Failure Modes: Parametric shifts in leakage and mechanical failure

Common Failure Mechanisms: Corrosion or contaminants on or within the package materials. Poor package sealing

Military Reference: MIL-STD-750, Method 1021

SOLDERABILITY

The purpose of this test is to measure the ability of device leads/terminals to be soldered after an extended period of storage (shelf life).

Typical Test Conditions: Steam aging = 8 hours, Flux = R, Solder = Sn60, Sn63

Common Failure Modes: Pin holes, dewetting, non-wetting

Common Failure Mechanisms: Poor plating, contaminated leads

Military Reference: MIL-STD-750, Method 2026

SOLDER HEAT

This test is used to measure the ability of a device to withstand the temperatures as may be seen in wave soldering operations. Electrical testing is the endpoint criterion for this stress.

Typical Test Conditions: Solder Temperature = 260°C, t = 10 seconds

Common Failure Modes: Parameter shifts, mechanical failure

Common Failure Mechanisms: Poor package design

Military Reference: MIL-STD-750, Method 2031

STEADY STATE OPERATING LIFE (SSOL)

The purpose of this test is to evaluate the bulk stability of the die and to generate defects resulting from manufacturing aberrations that are manifested as time and stress-dependent failures.

Typical Test Conditions: T_A = 25°C, P_D = Data Book maximum rating, t = 16 to 1000 hours

Common Failure Modes: Parametric shifts and catastrophic

Common Failure Mechanisms: Foreign material, crack die, bulk die, metallization, wire and die bond defects

Military Reference: MIL-STD-750, Method 1026

TEMPERATURE CYCLING (AIR TO AIR)

The purpose of this test is to evaluate the ability of the device to withstand both exposure to extreme temperatures and transitions between temperature extremes. This testing will also expose excessive thermal mismatch between materials.

Typical Test Conditions: T_A = −65°C to 200°C, cycle = 10 to 1000

Common Failure Modes: Parametric shifts and catastrophic

Common Failure Mechanisms: Wire bond, cracked or lifted die and package failure

Military Reference: MIL-STD-750, Method 1051

THERMAL SHOCK (LIQUID TO LIQUID)

The purpose of this test is to evaluate the ability of the device to withstand both exposure to extreme temperatures and sudden transitions between temperature extremes. This testing will also expose excessive thermal mismatch between materials.

Typical Test Conditions: T_A = 0°C to 100°C, cycles = 10 to 1000

Common Failure Modes: Parametric shifts and catastrophic

Common Failure Mechanisms: Wire bond, cracked or lifted die and package failure

Military Reference: MIL-STD-750, Method 1056

VARIABLE FREQUENCY VIBRATION

This test is used to examine the ability of the device to withstand deterioration due to mechanical resonance.

Typical Test Conditions: Peak acceleration = 20 g's, Frequency range = 20 Hz to 20 kHz, t = 48 minutes.

Common Failure Modes: Open, short, excessive leakage, mechanical failure

Common Failure Mechanisms: Die and wire bonds, cracked die, package defects

Military Reference: MIL-STD-750, Method 2056

ZENER DIODE CHARACTERISTICS

INTRODUCTION

At first glance the zener diode is a simple device consisting of one P-N junction with controlled breakdown voltage properties. However, when considerations are given to the variations of temperature coefficient, zener impedance, thermal time response, and capacitance, all of which are a function of the breakdown voltage (from 1.8 to 400 V), a much more complicated picture arises. In addition to the voltage spectrum, a variety of power packages are on the market with a variation of dice area inside the encapsulation.

This section is devoted to sorting out the important considerations in a “typical” fashion. For exact details, the data sheets must be consulted. However, much of the information contained herein is supplemental to the data sheet curves and will broaden your understanding of zener diode behavior.

Specifically, the following main subjects will be detailed:

Basic DC Volt-Ampere Characteristics

Impedance versus Voltage and Current

Temperature Coefficient versus Voltage and Current

Power Derating

Mounting

Thermal Time Response − Effective Thermal Impedance

Surge Capabilities

Frequency Response − Capacitance and Switching Effects

BASIC ZENER DIODE DC VOLT-AMPERE CHARACTERISTICS

Reverse and forward volt-ampere curves are represented in Figure 1 for a typical zener diode. The three areas − forward, leakage, and breakdown − will each be examined.

FORWARD DC CHARACTERISTICS

The forward characteristics of a zener diode are essentially identical with an “ordinary” rectifier and is shown in Figure 2. The volt-ampere curve follows the basic diode equation of $I_F = I_R e^{qV/KT}$ where KT/q equals about 0.026 volts at room temperature and I_R (reverse leakage current) is dependent upon the doping levels of the P-N junction as well as the area. The actual plot of V_F versus I_F deviates from the theoretical due to slightly “fixed” series resistance of the lead wire, bonding contacts and some bulk effects in the silicon.

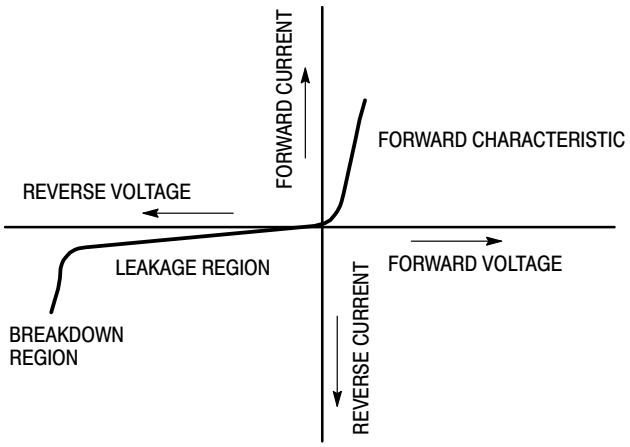


Figure 1. Typical Zener Diode DC V-I Characteristics (Not to Scale)

While the common form of the diode equation suggests that I_R is constant, in fact I_R is itself strongly temperature dependent. The rapid increase in I_R with increasing temperature dominates the decrease contributed by the exponential term in the diode equation. As a result, the forward current increases with increasing temperature. Figure 2 shows a forward characteristic temperature dependence for a typical zener. These curves indicate that for a constant current, an increase in temperature causes a decrease in forward voltage. The voltage temperature coefficient values are typically in the range of −1.4 to −2 mV/°C.

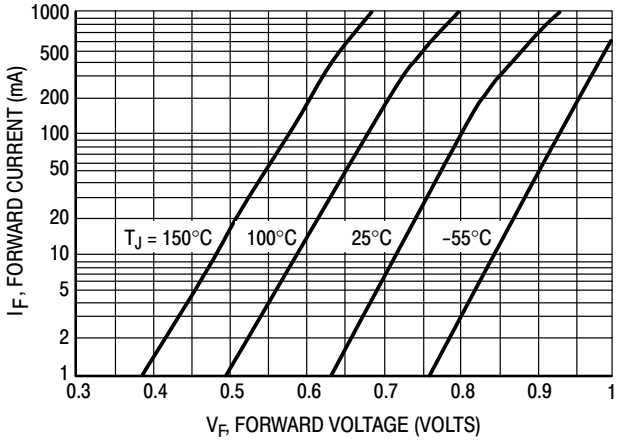


Figure 2. Typical Forward Characteristics of Zener Diodes

LEAKAGE DC CHARACTERISTICS

When reverse voltage less than the breakdown is applied to a zener diode, the behavior of current is similar to any back-biased silicon P-N junction. Ideally, the reverse current would reach a level at about one volt reverse voltage and remain constant until breakdown is reached. There are both theoretical and practical reasons why the typical V-I curve will have a definite slope to it as seen in Figure 3. Multiplication effects and charge generation sites are present in a zener diode which dictate that reverse current (even at low voltages) will increase with voltage. In addition, surface charges are ever present across P-N junctions which appear to be resistive in nature.

The leakage currents are generally less than one microampere at 150°C except with some large area devices. Quite often a leakage specification at 80% or so of breakdown voltage is used to assure low reverse currents.

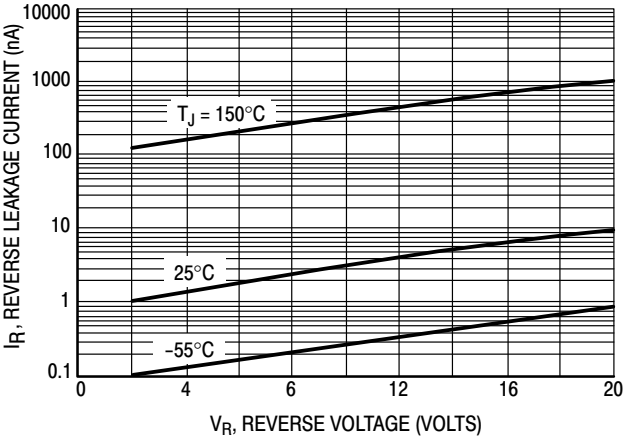


Figure 3. Typical Leakage Current versus Voltage

VOLTAGE BREAKDOWN

At some definite reverse voltage, depending on the doping levels (resistivity) of the P-N junction, the current will begin to avalanche. This is the so-called “zener” or “breakdown” area and is where the device is usually biased during use. A typical family of breakdown curves showing the effect of temperature is illustrated in Figure 4.

Between the minimum currents shown in Figure 4 and the leakage currents, there is the “knee” region. The avalanche mechanism may not occur simultaneously across the entire area of the P-N junction, but first at one microscopic site, then at an increasing number of sites as further voltage is applied. This action can be accounted for by the “microplasma discharge” theory and correlates with several breakdown characteristics.

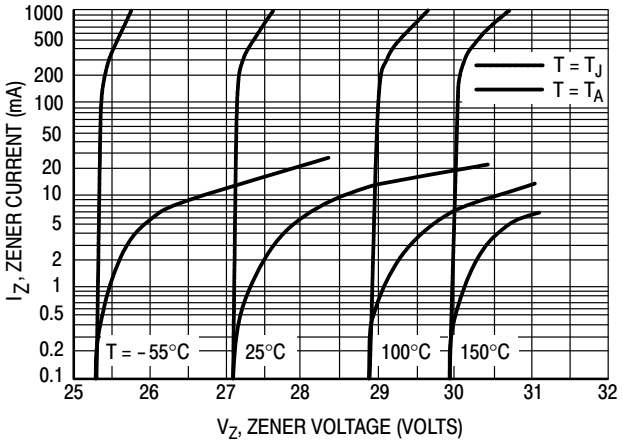


Figure 4. Typical Zener Characteristic Variation with Temperature

An exaggerated view of the knee region is shown in Figure 5. As can be seen, the breakdown or avalanche current does not increase suddenly, but consists of a series of smoothly rising current versus voltage increments each with a sudden break point.

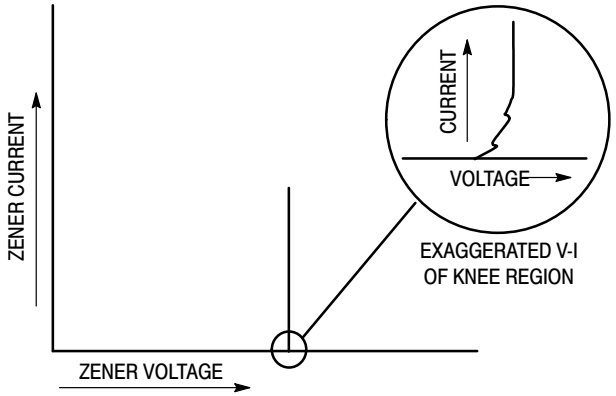


Figure 5. Exaggerated V-I Characteristics of the Knee Region

At the lowest point, the zener resistance (slope of the curve) would test high, but as current continues to climb, the resistance decreases. It is as though each discharge site has high resistance with each succeeding site being in parallel until the total resistance is very small.

In addition to the resistive effects, the micro plasmas may act as noise generators. The exact process of manufacturing affects how high the noise will be, but in any event there will be some noise at the knee, and it will diminish considerably as current is allowed to increase.

Since the zener impedance and the temperature coefficient are of prime importance when using the zener diode as a reference device, the next two sections will expand on these points.

ZENER IMPEDANCE

The slope of the $V_Z - I_Z$ curve (in breakdown) is defined as zener impedance or resistance. The measurement is generally done with a 60 Hz (on modern, computerized equipment this test is being done at 1 kHz) current variation whose value is 10% in rms of the dc value of the current. (That is, ΔI_Z peak to peak = 0.282 I_Z.) This is really not a small signal measurement but is convenient to use and gives repeatable results.

The zener impedance always decreases as current increases, although at very high currents (usually beyond I_Z max) the impedance will approach a constant. In contrast, the zener impedance decreases very rapidly with increasing current in the knee region. On Semiconductor specifies most zener diode impedances at two points: I_{ZT} and I_{ZK}. The term I_{ZT} usually is at the quarter power point, and I_{ZK} is an arbitrary low value in the knee region. Between these two points a plot of impedance versus current on a log-log scale is close to a straight line. Figure 6 shows a typical plot of Z_Z versus I_Z for a 20 volt−500 mW zener. The worst case impedance between I_{ZT} and I_{ZK} could be approximated by assuming a straight line function on a log-log plot; however, at currents above I_{ZT} or below I_{ZK} a projection of this line may give erroneous values.

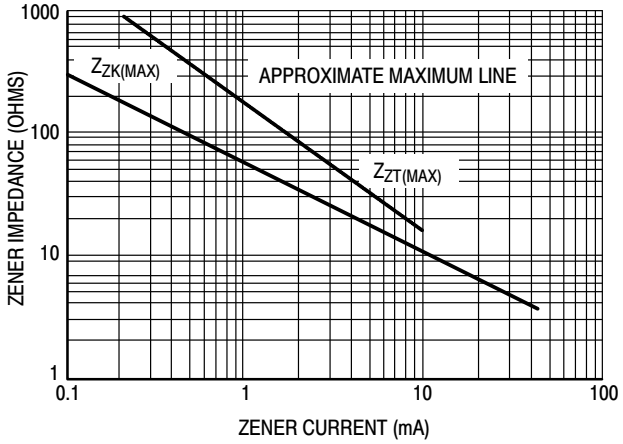


Figure 6. Zener Impedance versus Zener Current

The impedance variation with voltage is much more complex. First of all, zeners below 6 volts or so exhibit “field emission” breakdown converting to “avalanche” at higher currents. The two breakdowns behave somewhat differently with “field emission” associated with high impedance and negative temperature coefficients and “avalanche” with lower impedance and positive temperature coefficients.

A V-I plot of several low voltage 500 mW zener diodes is shown in Figure 7. It is seen that at some given current (higher for the lower voltage types) there is a fairly sudden decrease in the slope of $\Delta V/\Delta I$. Apparently, this current is the transition from one type of breakdown to the other. Above 6 volts the curves would show a gradual decrease of $\Delta V/\Delta I$ rather than an abrupt change, as current is increased.

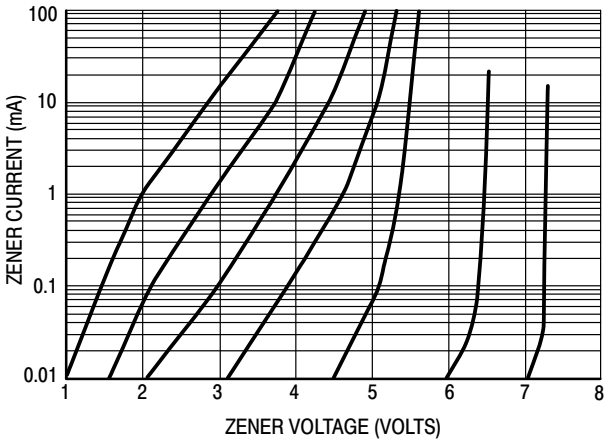


Figure 7. Zener Current versus Zener Voltage (Low Voltage Region)

Possibly the plots shown in Figure 8 of zener impedance versus voltage at several constant I_Z’s more clearly points out this effect. It is obvious that zener diodes whose breakdowns are about 7 volts will have remarkably low impedance.

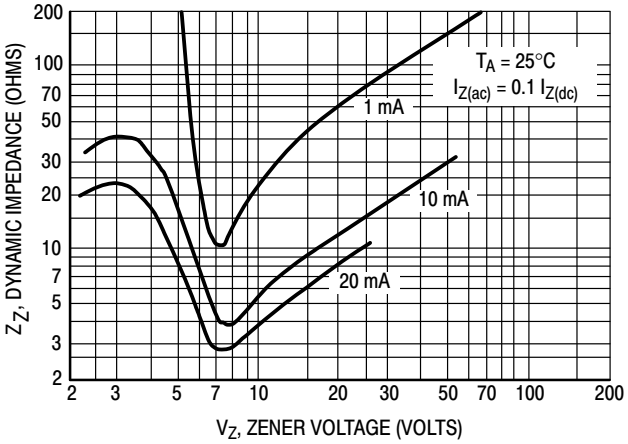
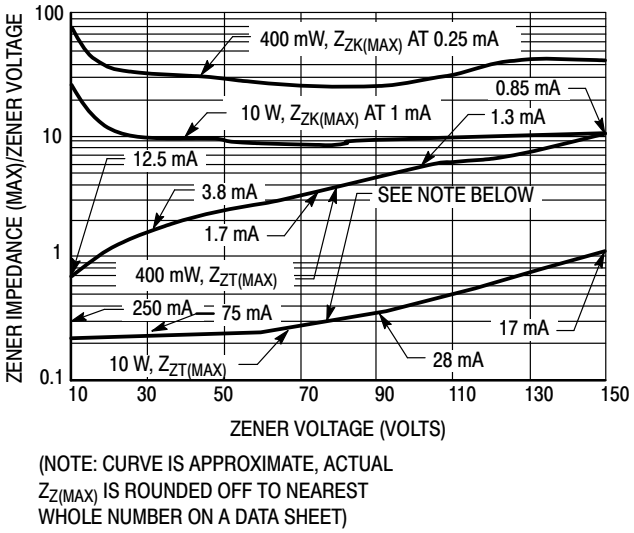


Figure 8. Dynamic Zener Impedance (Typical) versus Zener Voltage

However, this is not the whole picture. A zener diode figure of merit as a regulator could be Z_Z/V_Z. This would give some idea of what percentage change of voltage could be expected for some given change in current. Of course, a low Z_Z/V_Z is desirable. Generally zener current must be decreased as voltage is increased to prevent excessive power dissipation; hence zener impedance will rise even higher and the “figure of merit” will become higher as voltage increases. This is the case with I_{ZT} taken as the test point. However, if I_{ZK} is used as a comparison level in those devices which keep a constant I_{ZK} over a large range of voltage, the “figure of merit” will exhibit a bowl-shaped curve − first decreasing and then increasing as voltage is increased. Typical plots are shown in Figure 9. The conclusion can be reached that for uses where wide swings of current may occur and the quiescent bias current must be high, the lower voltage zener will provide best regulation,

but for low power applications, the best performance could be obtained between 50 and 100 volts.



(NOTE: CURVE IS APPROXIMATE, ACTUAL $Z_{Z(MAX)}$ IS ROUNDED OFF TO NEAREST WHOLE NUMBER ON A DATA SHEET)

Figure 9. Figure of Merit: $Z_{Z(Max)}/V_Z$ versus V_Z (400 mW & 10 W Zeners)

TEMPERATURE COEFFICIENT

Below three volts and above eight volts the zener voltage change due to temperature is nearly a straight line function and is almost independent of current (disregarding self-heating effects). However, between three and eight volts the temperature coefficients are not a simple affair. A typical plot of T_C versus V_Z is shown in Figure 10.

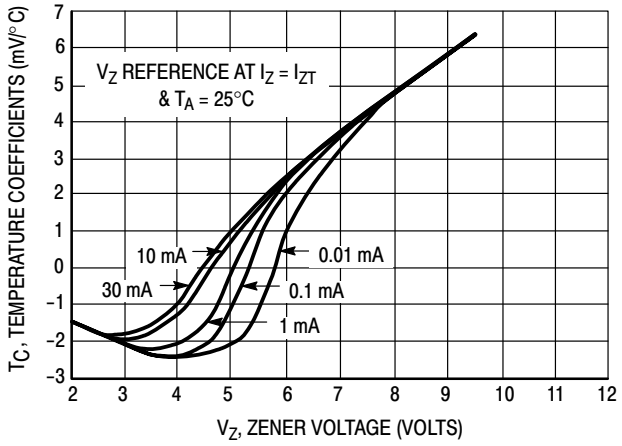


Figure 10. Temperature Coefficient versus Zener Voltage at 25°C Conditions Typical

Any attempt to predict voltage changes as temperature changes would be very difficult on a “typical” basis. (This, of course, is true to a lesser degree below three volts and above eight volts since the curve shown is a typical one and slight deviations will exist with a particular zener diode.) For example, a zener which is 5 volts at 25°C could be from 4.9 to 5.05 volts at 75°C depending on the current level. Whereas, a zener which is 9 volts at 25°C would be close to 9.3 volts at 75°C for all useful current levels (disregarding impedance effects).

As was mentioned, the situation is further complicated by the normal deviation of T_C at a given current. For example, for 7.5 mA the normal spread of T_C (expressed in %/°C) is shown in Figure 11. This is based on limited samples and in no manner implies that all On Semiconductor zeners between 2 and 12 volts will exhibit this behavior. At other current levels similar deviations would occur.

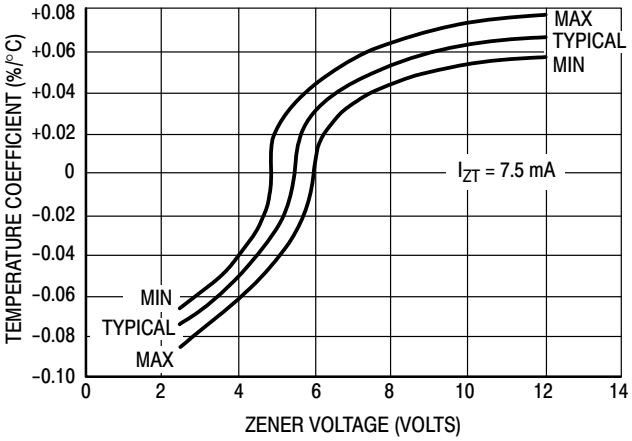


Figure 11. Temperature Coefficient Spread versus Zener Voltage

Obviously, all of these factors make it very difficult to attempt any calculation of precise voltage shift due to temperature. Except in devices with specified maximum T.C., no “worse case” design is possible.

Typical temperature characteristics is illustrated in Figure 12. This graphically shows the significant change in voltage for high voltage devices (about a 20 volt increase for a 100°C increase on a 200 volt device).

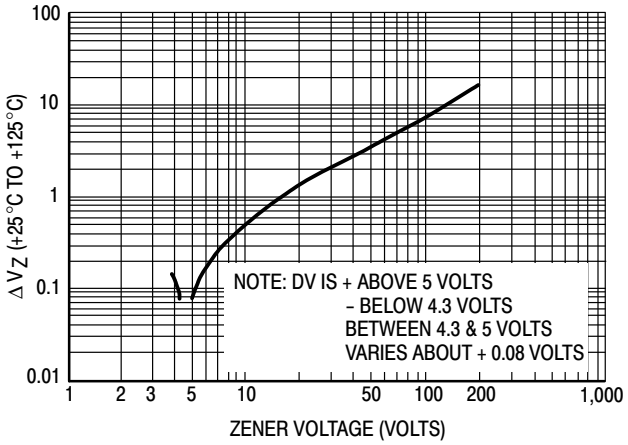


Figure 12. Typical Temperature Characteristics

POWER DERATING AND MOUNTING

The zener diode like any other semiconductor has a maximum junction temperature. This limit is somewhat arbitrary and is set from a reliability viewpoint. Most semiconductors exhibit an increasing failure rate as temperature increases. At some temperature, the solder will melt or soften and the failure rate soars. The 175°C to 200°C junction temperature rating is quite safe from solder failures and still has a very low failure rate.

In order that power dissipated in the device will never cause the junction to rise beyond 175°C or 200°C (depending on the device), the relation between temperature rise and power must be known. Of course, the thermal resistance ($R_{\theta JA}$ or $R_{\theta JL}$) is the factor which relates power and temperature in the well known "Thermal Ohm's Law'' relation:

$$\Delta T = T_J - T_A = R_{\theta JA} P_Z \quad (1)$$

and

$$\Delta T = T_J - T_L = R_{\theta JL} P_Z \quad (2)$$

where

T_J = Junction temperature
T_A = Ambient temperature
T_L = Lead temperature
$R_{\theta JA}$ = Thermal resistance junction to ambient
$R_{\theta JL}$ = Thermal resistance junction to lead
P_Z = Zener power dissipation

Obviously, if ambient or lead temperature is known and the appropriate thermal resistance for a given device is known, the junction temperature could be precisely calculated by simply measuring the zener dc current and voltage ($P_Z = I_Z V_Z$). This would be helpful to calculate voltage change versus temperature. However, only maximum and typical values of thermal resistance are given for a family of zener diodes. So only "worst case" or typical information could be obtained as to voltage changes.

The relations of equations 1 and 2 are usually expressed as a graphical derating of power versus the appropriate temperature. Maximum thermal resistance is used to generate the slope of the curve. An example of a 400 milliwatt device derated to the ambient temperature and a 1 watt device derated to the lead temperature are shown in Figures 13 and 14.

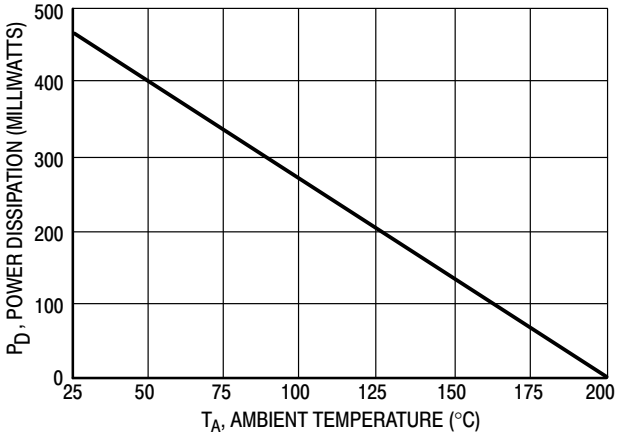


Figure 13. 400 mW Power Temperature Derating Curve

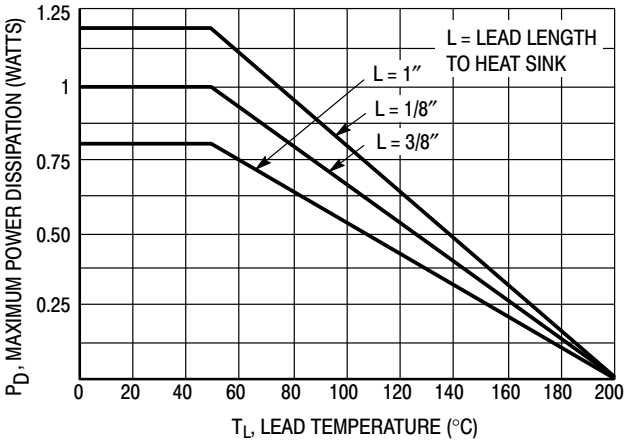


Figure 14. Power Temperature Derating Curve

A lead mounted device can have its power rating increased by shortening the lead length and "heatsinking" the ends of the leads. This effect is shown in Figure 15, for the 1N4728, 1 watt zener diode.

Each zener has a derating curve on its data sheet and steady state power can be set properly. However, temperature increases due to pulsed power are not so easily calculated and therefore the use of "Transient Thermal Resistance" would be required. The next section expounds upon transient thermal behavior as a function of time and surge power.

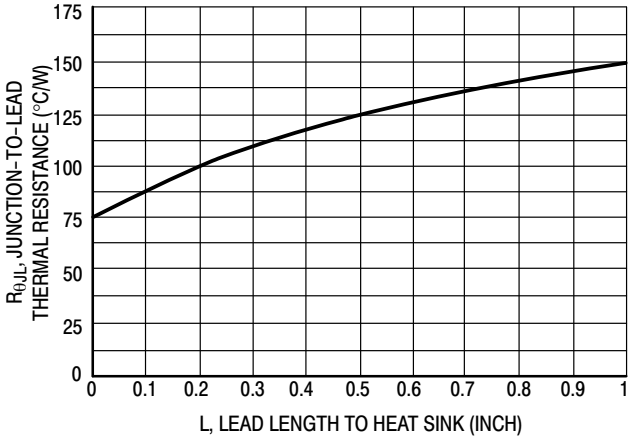


Figure 15. Typical 1N4728 Thermal Resistance versus Lead Length

THERMAL TIME RESPONSE

Early studies of zener diodes indicated that a "thermal time constant" existed which allowed calculation of temperature rise as a function of power pulse height, width, and duty cycle. More precise measurements have shown that temperature response as a function of time cannot be represented as a simple time constant. Although as shown in the preceding section, the steady state conditions are analogous in every way to an electrical resistance; a simple "thermal capacitance" placed across the resistor is not the true equivalent circuit. Probably a series of parallel R-C

networks or lumped constants representing a thermal transmission line would be more accurate.

Fortunately a concept has developed in the industry wherein the exact thermal equivalent circuit need not be found. If one simply accepts the concept of a thermal resistance which varies with time in a predictable manner, the situation becomes very practical. For each family of zener diodes, a “worst case” transient thermal resistance curve may be generated.

The main use of this transient $R_{\theta JL}$ curve is when the zener is used as a clipper or a protective device. First of all, the power wave shape must be constructed. (Note, even though the power-transient thermal resistance indicates reasonable junction temperatures, the device still may fail if the peak current exceeds certain values. Apparently a current crowding effect occurs which causes the zener to short. This is discussed further in this section.)

TRANSIENT POWER-TEMPERATURE EFFECTS

A typical transient thermal resistance curve is shown in Figure 16. This is for a lead mounted device and shows the effect of lead length to an essentially infinite heatsink.

To calculate the temperature rise, the power surge wave shape must be approximated by its rectangular equivalent as shown in Figure 17. In case of an essentially non-recurrent pulse, there would be just one pulse, and $\Delta T = R_{\theta T1} P_p$. In the general case, it can be shown that

$$\Delta T = [DR_{\theta JA}(ss) + (1 - D) R_{\theta T1 + T} + R_{\theta T1} - R_{\theta T}] P_P$$

where

- D = Duty cycle in percent
- $R_{\theta T1}$ = Transient thermal resistance at the time equal to the pulse width
- $R_{\theta T}$ = Transient thermal resistance at the time equal to pulse interval
- $R_{\theta T1 + T}$ = Transient thermal resistance at the time equal to the pulse interval plus one more pulse width.
- $R_{\theta JA}(ss)$ or $R_{\theta JL}(ss)$ = Steady state value of thermal resistance

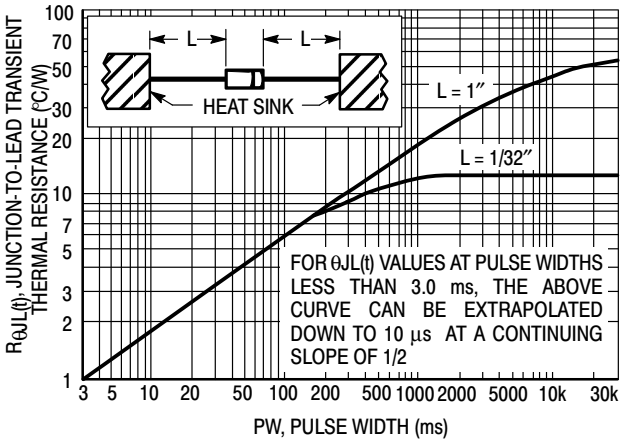


Figure 16. Typical Transient Thermal Resistance (For Axial Lead Zener)

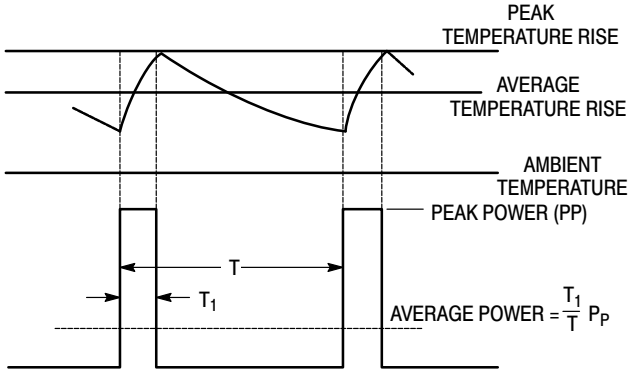


Figure 17. Relation of Junction Temperature to Power Pulses

This method will predict the temperature rise at the end of the power pulse after the chain of pulses has reached equilibrium. In other words, the average power will have caused an average temperature rise which has stabilized, but a temperature “ripple” is present.

Example: (Use curve in Figure 16)

P_P = 5 watt (Lead length 1/32″)
D = 0.1
T_1 = 10 ms
T = 100 ms
$R_{\theta JA}(ss)$ = 12°C/W (for 1/32″ lead length)

Then

$R_{\theta T1}$ = 1.8°C/W
$R_{\theta T}$ = 5.8°C/W
$R_{\theta T1 + T}$ = 6°C/W

And

$\Delta T = [0.1 \times 12 + (1 - 0.1) 6 + 1.8 - 5.8] 5$
ΔT = 13°C

Or at T_A = 25°, T_J = 38°C peak

SURGE FAILURES

If no other considerations were present, it would be a simple matter to specify a maximum junction temperature no matter what pulses are present. However, as has been noted, apparently other fault conditions prevail. The same group of devices for which the transient thermal curves were generated were tested by subjecting them to single shot power pulses. A failure was defined as a significant shift of leakage or zener voltage, or of course opens or shorts. Each device was measured before and after the applied pulse. Most failures were shifts in zener voltage. The results are shown in Figure 18.

Attempts to correlate this to the transient thermal resistance work quite well on a typical basis. For example, assuming a value for 1 ms of 90 watts and 35 watts at 10 ms, the predicted temperature rise would be 180°C and 190°C. But on a worst case basis, the temperature rises would be about one half these values or junction temperatures, on the order of 85°C to 105°C, which are obviously low. Apparently at very high power levels a current restriction occurs causing hot spots. There was no apparent correlation

of zener voltage or current on the failure points since each group of failures contained a mixture of voltages.

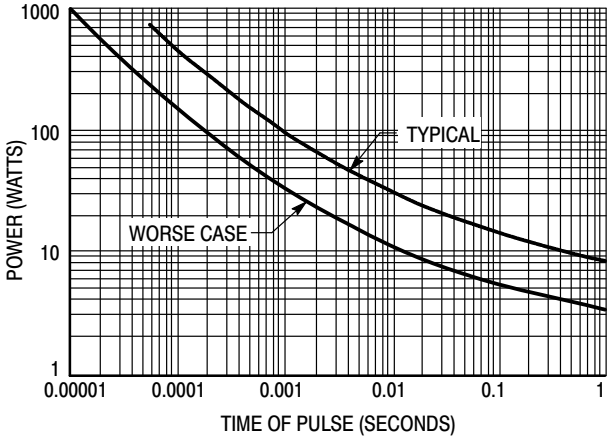


Figure 18. One Shot Power Failure Axial Lead Zener Diode

VOLTAGE VERSUS TIME

Quite often the junction temperature is only of academic interest, and the designer is more concerned with the voltage behavior versus time. By using the transient thermal resistance, the power, and the temperature coefficient, the designer could generate V_Z versus time curves. The On Semiconductor zener diode test group has observed device voltages versus time until the thermal equilibrium was reached. A typical curve is shown for a lead mounted low wattage device in Figure 19 where the ambient temperature was maintained constant. It is seen that voltage stabilizes in about 100 seconds for 1 inch leads.

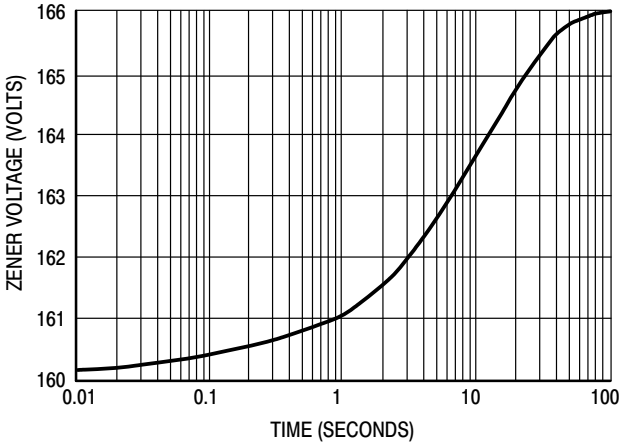


Figure 19. Zener Voltage (Typical) versus Time for Step Power Pulses (500 mW Lead Mounted Devices)

FREQUENCY AND PULSE CHARACTERISTICS

The zener diode may be used in applications which require a knowledge of the frequency response of the device. Of main concern are the zener resistance (usually specified as “impedance”) and the junction capacitance. The capacitance curves shown in this section are typical.

ZENER CAPACITANCE

Since zener diodes are basically PN junctions operated in the reverse direction, they display a capacitance that decreases with increasing reverse voltage. This is so because the effective width of the PN junction is increased by the removal of charges (holes and electrons) as reverse voltage is increased. This decrease in capacitance continues until the zener breakdown region is entered; very little further capacitance change takes place, owing to the now fixed voltage across the junction. The value of this capacitance is a function of the material resistivity, ρ, (amount of doping − which determines V_Z nominal), the diameter, D, of junction or dice size (determines amount of power dissipation), the voltage across the junction V_C, and some constant, K. This relationship can be expressed as:

$$C_C = \sqrt[n]{\frac{KD^4}{\rho V_C}}$$

After the junction enters the zener region, capacitance remains relatively fixed and the AC resistance then decreases with increasing zener current.

TEST CIRCUIT CONSIDERATIONS: A capacitive bridge was used to measure junction capacitance. In this method the zener is used as one leg of a bridge that is balanced for both DC at a given reverse voltage and for AC (the test frequency 1 MHz). After balancing, the variable capacitor used for balancing is removed and its value measured on a test instrument. The value thus indicated is the zener capacitance at reverse voltage for which bridge balance was obtained. Figure 20 shows capacitance test circuit.

Figure 21 is a plot of junction capacitance for diffused zener diode units versus their nominal operating voltage. Capacitance is the value obtained with reverse bias set at one-half the nominal V_Z. The plot of the voltage range from 6.8 V to 200 V, for three dice sizes, covers most On Semiconductor diffused-junction zeners. Consult specific data sheets for capacitance values.

Figures 22, 23, and 24 show plots of capacitance versus reverse voltage for units of various voltage ratings in each of the three dice sizes. Junction capacitance decreases as reverse voltage increases to the zener region. This change in capacitance can be expressed as a ratio which follows a one-third law, and $C_1/C_2 = (V_2/V_1)^{1/3}$. This law holds only from the zener voltage down to about 1 volt, where the curve begins to flatten out. Figure 25 shows this for a group of low wattage units.

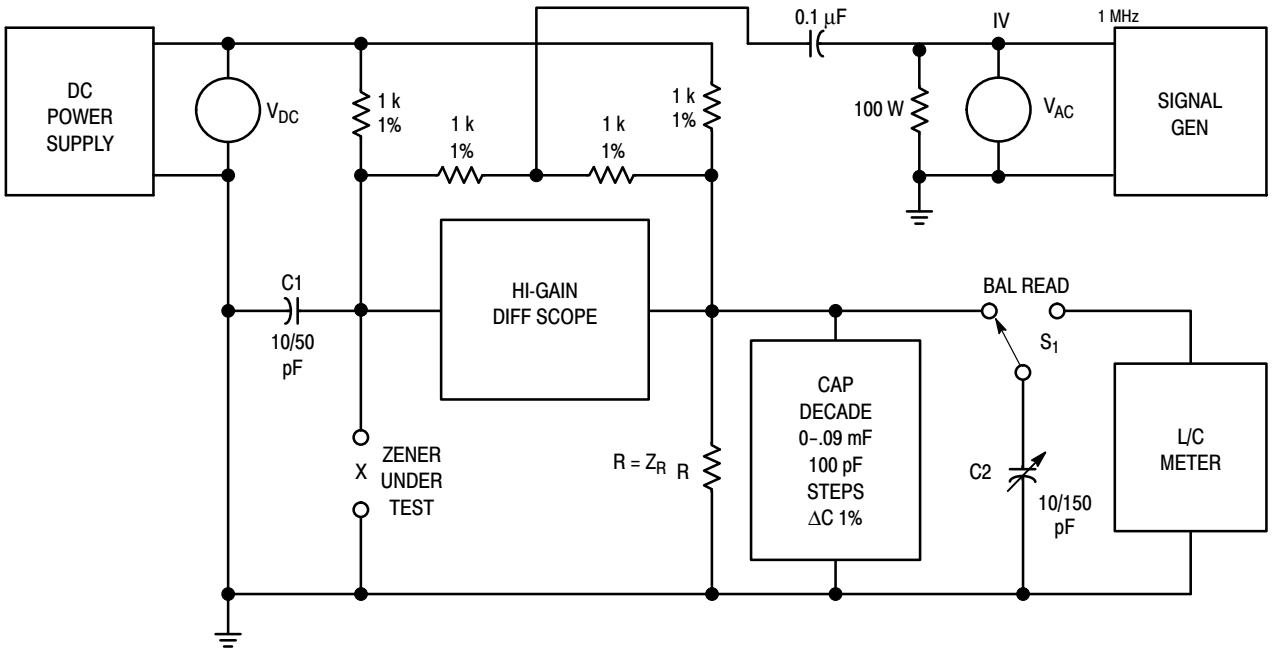


Figure 20. Capacitance Test Circuit

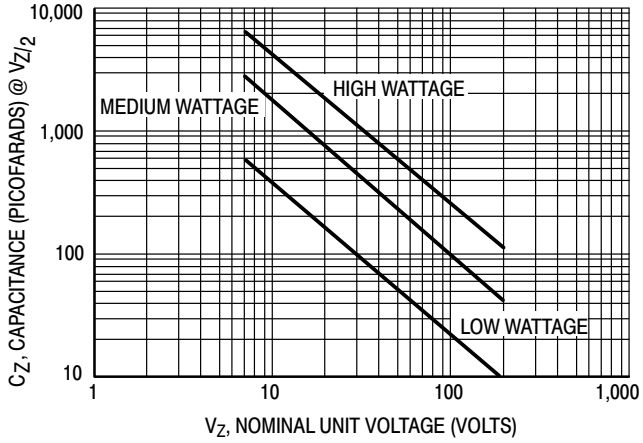


Figure 21. Capacitance versus Voltage

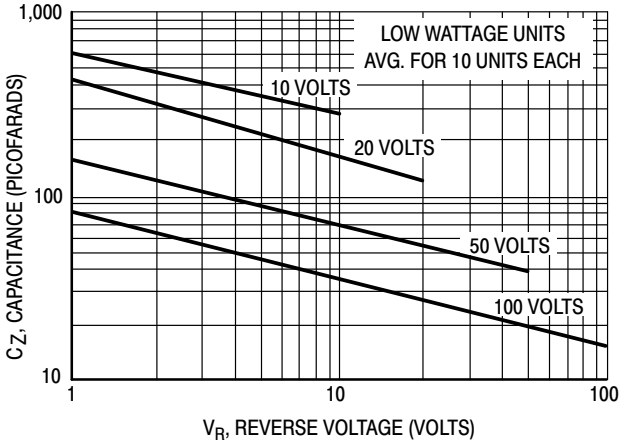


Figure 22. Capacitance versus Reverse Voltage

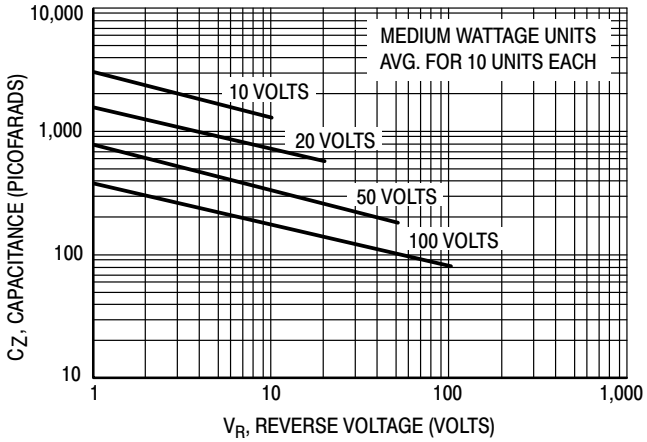


Figure 23. Capacitance versus Reverse Voltage

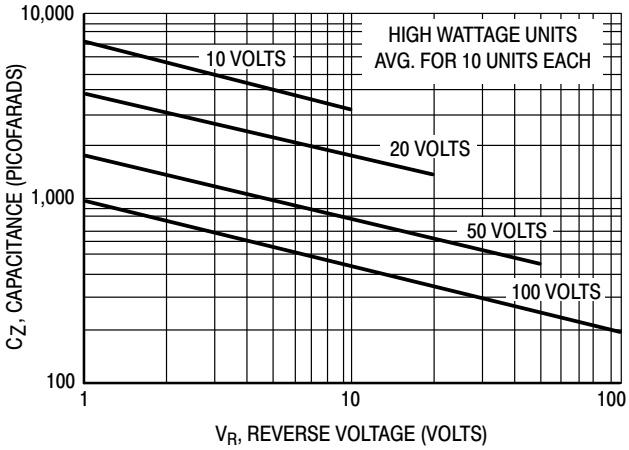


Figure 24. Capacitance versus Reverse Voltage

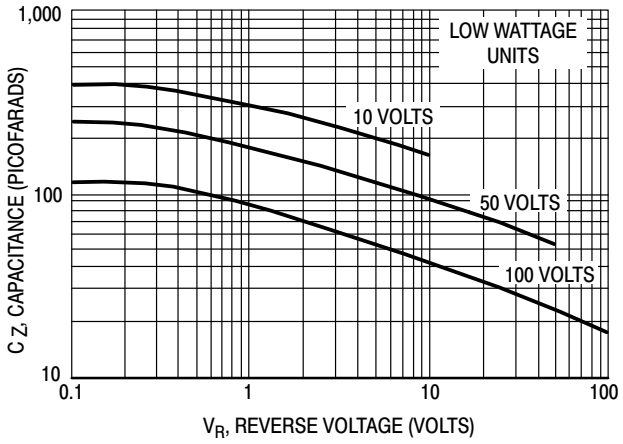


Figure 25. Flattening of Capacitance Curve at Low Voltages

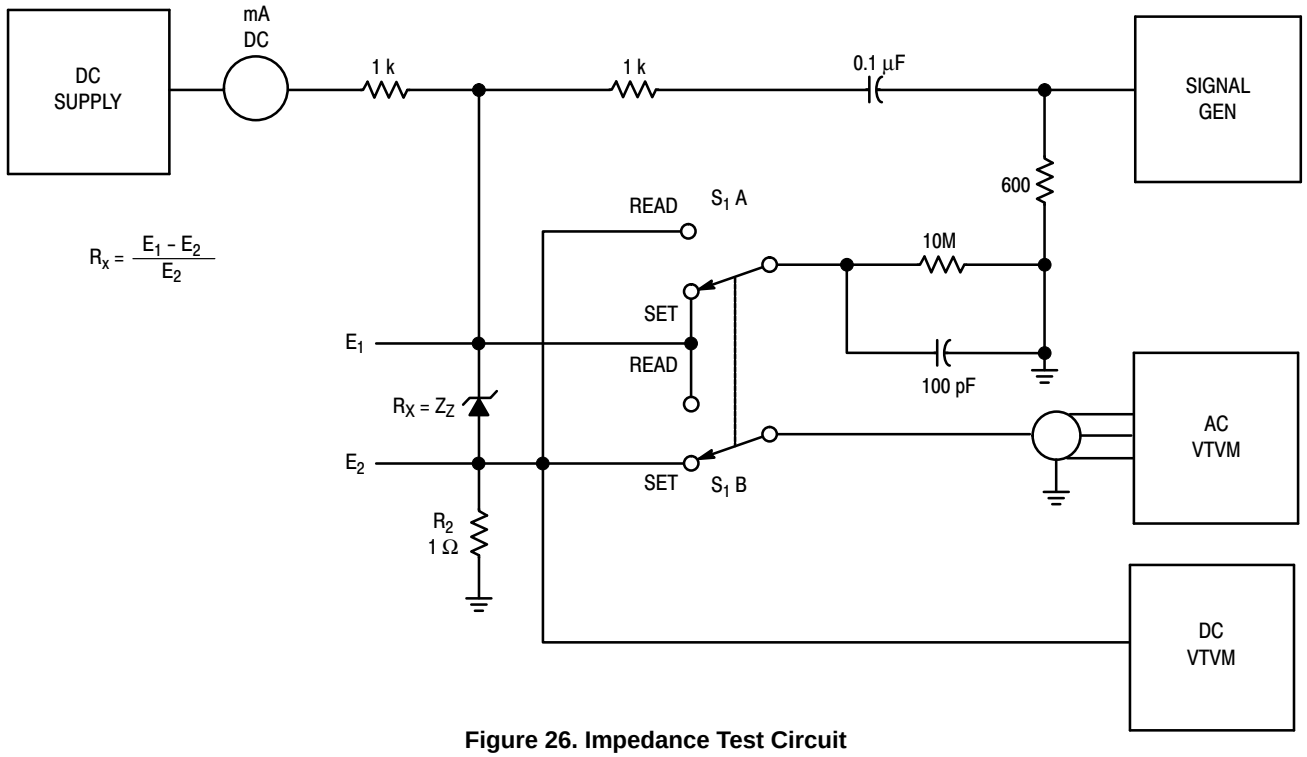


Figure 26. Impedance Test Circuit

ZENER IMPEDANCE

Zener impedance appears primarily as composed of a current-dependent resistance shunted by a voltage-dependent capacitor. Figure 26 shows the test circuit used to gather impedance data. This is a voltage-impedance ratio method of determining the unknown zener impedance. The operation is as follows:

(1) Adjust for desired zener I_{ZDC} by observing IR drop across the 1-ohm current-viewing resistor R_2.

(2) Adjust I_{ZAC} to 100 μA by observing AC IR drop across R2.

(3) Measure the voltage across the entire network by switching S1. The ratio of these two AC voltages is then a measure of the impedance ratio. This can be expressed simply as $R_X = [(E_1 - E_2)/E_2] R_2$.

Section A of S_1 provides a dummy load consisting of a 10-M resistor and a 100 pF capacitor. This network is required to simulate the input impedance of the AC VTVM while it is being used to measure the AC IR drop across R_2.

This method has been found accurate up to about three megahertz; above this frequency, lead inductances and strap capacitance become the dominant factors.

Figure 27 shows typical impedance versus frequency relationships of 6.8 volt 500 mW zener diodes at various DC zener currents. Before the zener breakdown region is entered, the impedance is almost all reactive, being provided by a voltage-dependent capacitor shunted by a very high resistance. When the zener breakdown region is entered, the capacitance is fixed and now is shunted by current-dependent resistance. For comparison, Figure 27 also shows the plot for a 680 pF capacitor X_C, a 1K 1% nonreactive resistor, R, and the parallel combination of these two passive elements, Z_T.

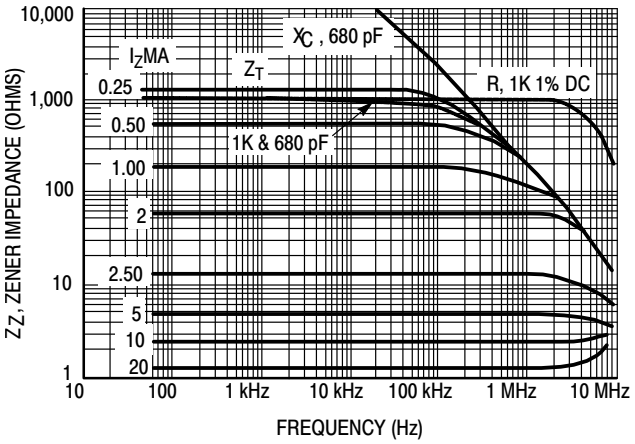


Figure 27. Zener Impedance versus Frequency

HIGH FREQUENCY AND SWITCHING CONSIDERATIONS

At frequencies about 100 kHz or so and switching speeds above 10 microseconds, shunt capacitance of zener diodes begins to seriously effect their usefulness. The upper photo of Figure 28 shows the output waveform of a symmetrical peak limiter using two zener diodes back-to-back. The capacitive effects are obvious here. In any application where the signal is recurrent, the shunt capacitance limitations can be overcome, as lower photo of Figure 28 shows. This is done by operating fast diodes in series with the zener. Upon application of a signal, the fast diode conducts in the forward direction charging the shunt zener capacitance to the level where the zener conducts and limits the peak. When the signal swings the opposite direction, the fast diode becomes back-biased and prevents fast discharge of shunt capacitance. The fast diode remains back-biased when the signal reverses again to the forward direction and remains off until the input signal rises and exceeds the charge level of the capacitor. When the signal exceeds this level, the fast diode conducts as does the zener. Thus, between successive cycles or pulses the charge in the shunt capacitor holds off the fast diode, preventing capacitive loading of the signal until zener breakdown is reached. Figures 29 and 30 show this method applied to fast-pulse peak limiting.

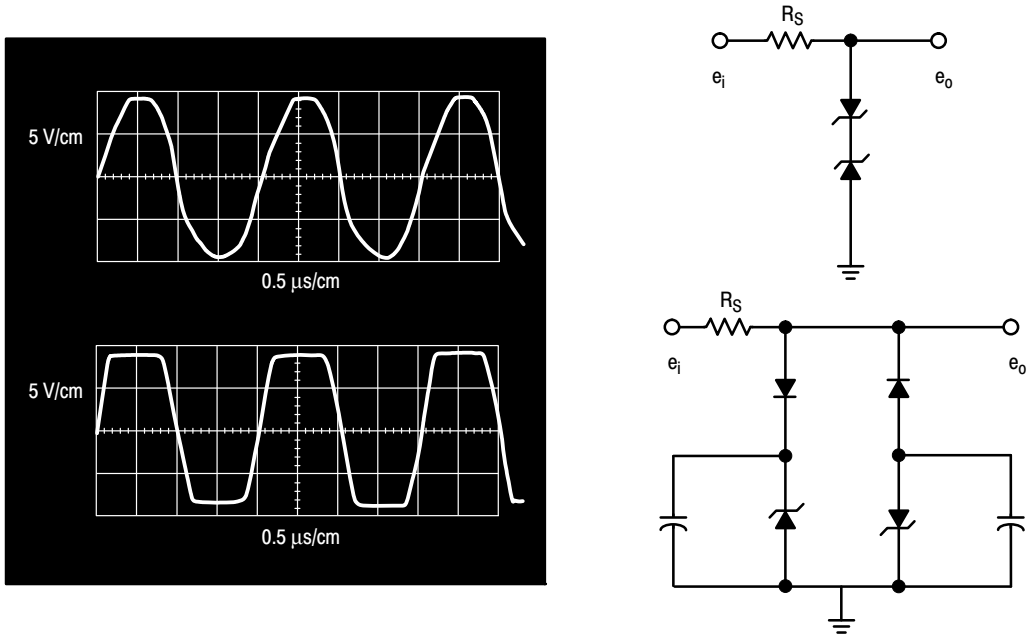


Figure 28. Symmetrical Peak Limiter

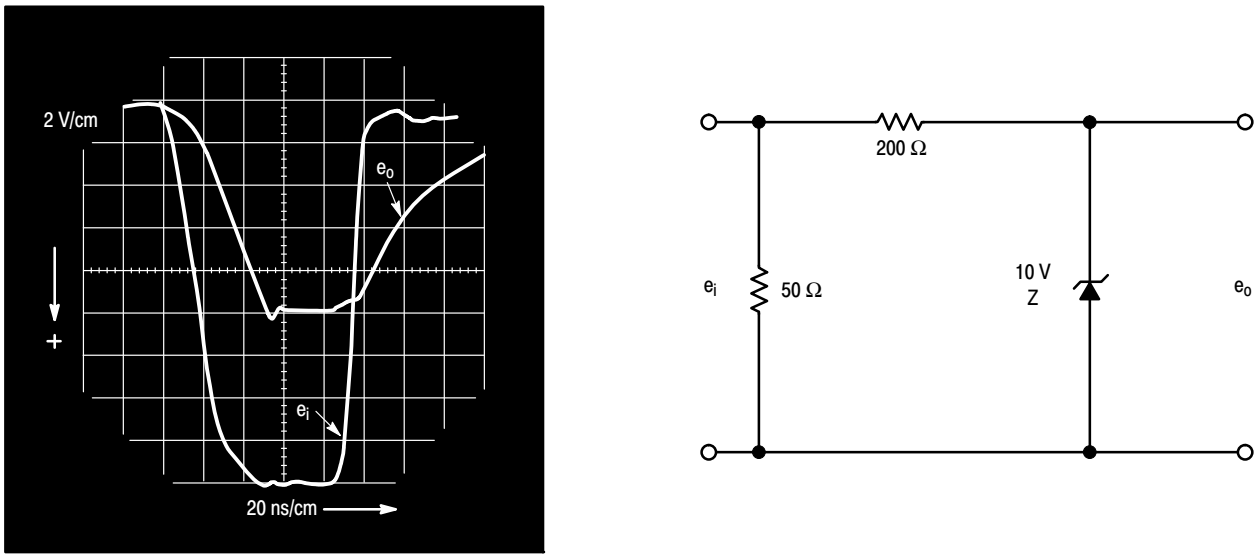


Figure 29. Shunt Clipper

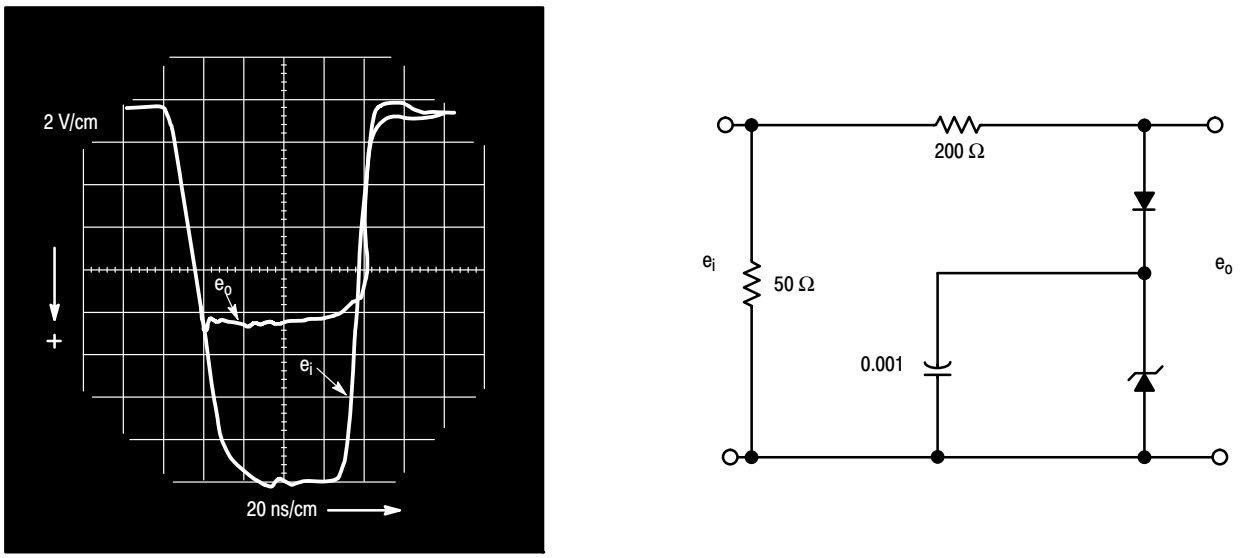
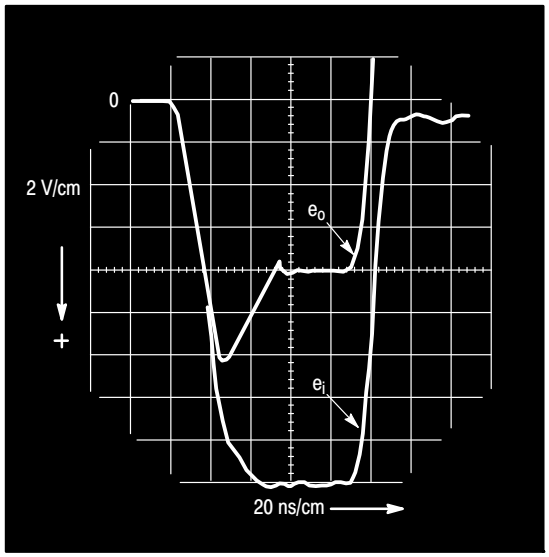


Figure 30. Shunt Clipper with Clamping Network

Figure 31 is a photo of input-output pulse waveforms using a zener alone as a series peak clipper. The smaller output waveform shows the capacitive spike on the leading edge. Figure 32 clearly points out the advantage of the clamping network.



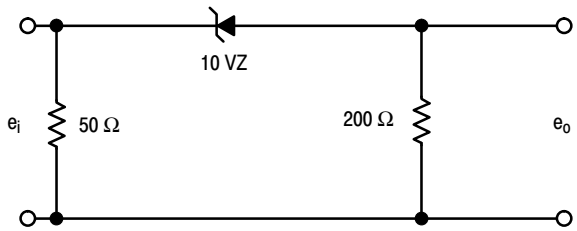
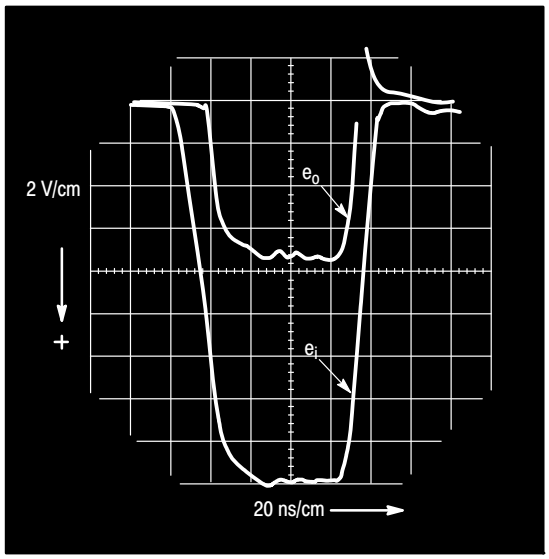


Figure 31. Basic Series Clipper



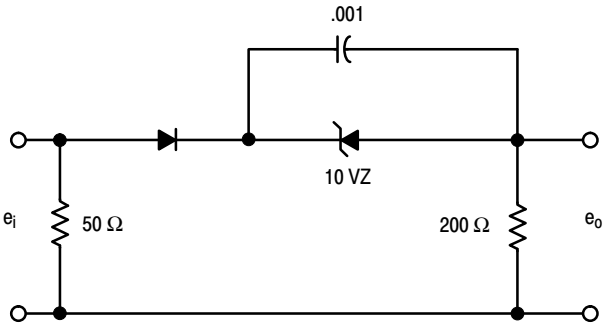


Figure 32. Series Clipper with Clamping Network

BASIC VOLTAGE REGULATION USING ZENER DIODES

BASIC CONCEPTS OF REGULATION

The purpose of any regulator circuit is to minimize output variations with respect to variations in input, temperature, and load requirements. The most obvious use of a regulator is in the design of a power supply, but any circuit that incorporates regulatory technique to give a controlled output or function can be considered as a regulator. In general, to provide a regulated output voltage, electronic circuitry will be used to pass an output voltage that is significantly lower than the input voltage and block all voltage in excess of the desired output. Allocations should also be made in the regulation circuitry to maintain this output voltage for variation in load current demand.

There are some basic rules of thumb for the electrical requirements of the electronic circuitry in order for it to provide regulation. Number one, the output impedance should be kept as low as possible. Number two, a controlling reference needs to be established that is relatively insensitive to the prevailing variables. In order to illustrate the importance of these rules, an analysis of some simple regulator circuits will point out the validity of the statements. The circuit of Figure 1 can be considered a regulator. This circuit will serve to illustrate the importance of a low output impedance.

The resistors R_S and R_R can be considered as the source and regulator impedances, respectively.

The output of the circuit is:

$$V_O = V_I \times \frac{R_R R_L}{R_R + R_L} \Bigg/ \left(R_S + \frac{R_R R_L}{R_R + R_L} \right) = \frac{V_I}{\frac{R_S}{R_L} + \frac{R_S}{R_R} + 1} \quad (1)$$

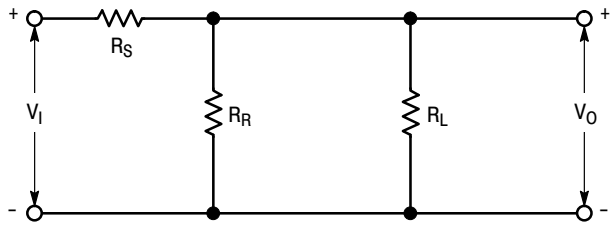


Figure 1. Shunt Resistance Regulator

For a given incremental change in V_I, the changes in V_O will be:

$$\Delta V_O = \Delta V_I \left(\frac{1}{\frac{R_S}{R_L} + \frac{R_S}{R_R} + 1} \right) \quad (2)$$

Assuming R_L fixed at some constant value, it is obvious from equation (2) that in order to minimize changes in V_O for variations in V_I, the shunt resistor R_R should be made as small as possible with respect to the source resistor R_S. Obviously, the better this relation becomes, the larger V_I is going to have to be for the same V_O, and not until the ratio of R_S to R_R reaches infinity will the output be held entirely constant for variation in V_I. This, of course, is an impossibility, but it does stress the fact that the regulation improves as the output impedance becomes lower and lower. Where the output impedance of Figure 1 is given by

$$R_O = \frac{R_S R_R}{R_S + R_R} \quad (3)$$

It is apparent from this relation that as regulation is improving with R_S increasing and R_R decreasing the output impedance R_O is decreasing, and is approximately equal to R_R as the ratio is 10 times or greater. The regulation of this circuit can be greatly improved by inserting a reference source of voltage in series with R_R such as Figure 2.

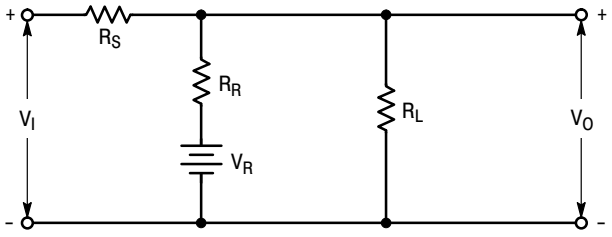


Figure 2. Regulator with Battery Reference Source

The resistance R_R represents the internal impedance of the battery. For this circuit, the output is

$$V_O = V_R + V_I \frac{V}{\frac{R_S}{R_L} + \frac{R_S}{R_R} + 1} \quad (4)$$

Then for incremental changes in the input V_I, the changes in V_O will be dependent on the second term of equation (4), which again makes the regulation dependent on the ratio of R_S to R_R. Where changes in the output voltage or the regulation of the circuit in Figure 1 were directly and solely dependent upon the input voltage and output impedance, the regulation of circuit 2 will have an output that varies about the reference source V_R in accordance with the magnitude of battery resistance R_R and its fluctuations for changes in V_I. Theoretically, if a perfect battery were used, that is, V_R is constant and R_R is zero, the circuit would be a perfect regulator. In other words, in line with the basic rules of thumb the circuit exhibits optimum regulation with an output impedance of zero, and a constant reference source.

For regulator application, a zener diode can be used instead of a battery with a number of advantages. A battery's resistance and nominal voltage will change with age and load demand; the ON Semiconductor zener diode characteristics remain unchanged when operating within its

specified limits. Any voltage value from a couple of volts to hundreds of volts is available with zener diodes, where conventional batteries are limited in the nominal values available. Also, the zener presents a definite size advantage, and is less expensive than a battery because it is permanent and need not be regularly replaced. The basic zener diode shunt regulator circuit is shown in Figure 3.

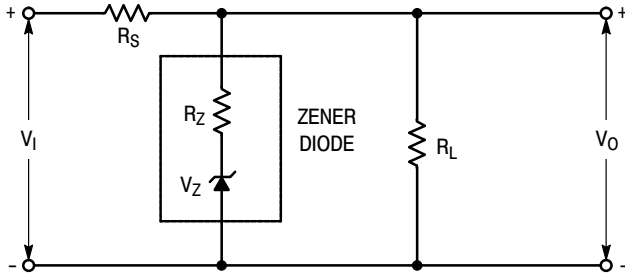


Figure 3. Basic Zener Diode Shunt Regulator

Depending upon the operating conditions of the device, a zener diode will exhibit some relatively low zener impedance R_Z and have a specified breakover voltage of V_Z that is essentially constant. These inherent characteristics make the zener diode suited for voltage regulator applications.

DESIGNING THE ZENER SHUNT REGULATOR

For any given application of a zener diode shunt regulator, it will be required to know the input voltage variations and output load requirements. The calculation of component values will be directly dependent upon the circuit requirements. The input may be constant or have maximum and minimum values depending upon the natural regulation or waveform of the supply source. The output voltage will be determined by the designer's choice of V_Z and the circuit requirements. The actual value of V_Z will be dependent upon the manufacturer's tolerance and some small variation for different zener currents and operating temperatures.

For all practical purposes, the value of V_Z as specified on the manufacturer's data sheet can be used to approximate V_O in computing component values. The requirement for load current will be known and will vary within some given range of $I_{L(min)}$ to $I_{L(max)}$.

The design objective of Figure 3 is to determine the proper values of the series resistance, R_S, and zener power dissipation, P_Z. A general solution for these values can be developed as follows, when the following conditions are known:

V_I (input voltage) from $V_{I(min)}$ to $V_{I(max)}$

V_O (output voltage) from $V_{Z(min)}$ to $V_{Z(max)}$

I_L (load current) from $I_{L(min)}$ to $I_{L(max)}$

The value of R_S must be of such a value so that the zener current will not drop below a minimum value of $I_{Z(min)}$. This minimum zener current is mandatory to keep the device in the breakover region in order to maintain the zener voltage reference. The minimum current can be either chosen at some point beyond the knee or found on the manufacturer's data sheet (I_{ZK}). The basic voltage loop equation for this circuit is:

$$V_I = (I_Z + I_L)R_S + V_Z \tag{5}$$

The minimum zener current will occur when V_I is minimum, V_Z is maximum, and I_L is maximum, then solving for R_S, we have:

$$R_S = \frac{V_{I(min)} - V_{Z(max)}}{I_{Z(min)} + I_{L(max)}} \tag{6}$$

Having found R_S, we can determine the maximum power dissipation P_Z for the zener diode.

$$P_{Z(max)} = I_{Z(max)}\, V_{Z(max)} \tag{7}$$

Where:

$$I_Z(max) = \frac{V_{I(max)} - V_{Z(min)}}{R_S} - I_{L(min)} \tag{8}$$

Therefore:

$$P_{Z(max)} = \left[\frac{V_{I(max)} - V_{Z(min)}}{R_S} - I_{L(min)}\right] V_{Z(max)} \tag{9}$$

Once the basic regulator components values have been determined, adequate considerations will have to be given to the variation in V_O. The changes in V_O are a function of four different factors; namely, changes in V_I, I_L, temperature, and the value of zener impedance, R_Z. These changes in V_O can be expressed as:

$$\Delta V_O = \frac{\Delta V_I}{1 + \frac{R_S}{R_Z} + \frac{R_S}{R_L}} - \frac{R_S R_Z}{R_S + R_Z}\ \Delta I_L + TC\Delta T V_Z \tag{10}$$

The value of ΔV_O as calculated with equation (10) will quite probably be slightly different from the actual value when measured empirically. For all practical purposes though, this difference will be insignificant for regulator designs utilizing the conventional commercial line of zener diodes.

Obviously to precisely predict ΔV_O with a given zener diode, exact information would be needed about the zener impedance and temperature coefficient throughout the variation of zener current. The "worst case" change can only be approximated by using maximum zener impedance and with typical temperature coefficient.

The basic zener shunt regulator can be modified to minimize the effects of each term in the regulation equation (10). Taking one term at a time, it is apparent that the regulation or changes in output ΔV_O will be improved if the magnitude of ΔV_I is reduced. A practical and widely used technique to reduce input variation is to cascade zener shunt regulators such as shown in Figure 4.

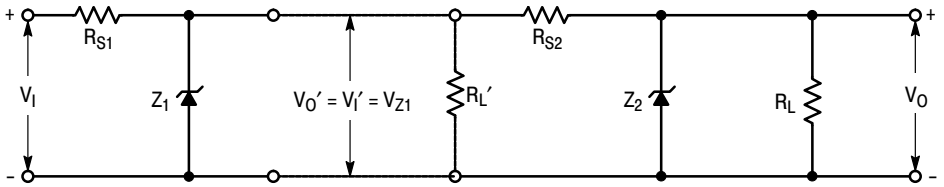


Figure 4. Cascaded Zener Shunt Regulators Reduce ΔV_O by Reducing ΔV_I to the Succeeding Stages

This, in essence, is a regulator driven with a pre-regulator so that the over all regulation is the product of both. The regulation or changes in output voltage is determined by:

$$\Delta V_O = \frac{\Delta V_{ZI}}{1 + \frac{R_{S2}}{R_L} + \frac{R_{S2}}{R_{Z2}}} - \frac{R_{S2}R_{Z2}}{R_{S2} + R_{Z2}}$$
$$\Delta I_L + TC_2\,\Delta TV_{Z2} \qquad (11)$$

Where:

$$\Delta V_{Z1} = \Delta V_O' = \frac{\Delta V_I}{1 + \frac{R_{S1}}{R_L'} + \frac{R_{S1}}{R_{Z1}}} - \frac{R_{S1}R_{Z1}}{R_{S1} + R_{Z1}}$$
$$\Delta I_L' + TC_1\,\Delta TV_{Z1} \qquad (12)$$

$$R_L' = R_{S2} + \frac{R_L R_{Z2}}{R_L + R_{Z2}} \text{ and } I_L' = I_L + I_{Z2}$$

The changes in output with respect to changes in input for both stages assuming the temperature and load are constant is

$$\frac{\Delta V_O}{\Delta V_{Z1}} = \frac{\Delta V_O}{\Delta V_O'} = \text{Regulation of second stage} \qquad (13)$$

$$\frac{\Delta V_O'}{\Delta V_I} = \text{Regulation of first stage} \qquad (14)$$

$$\frac{\Delta V_O}{\Delta V_I} = \frac{\Delta V_O}{\Delta V_O'} \times \frac{\Delta V_O'}{\Delta V_I} = \text{Combined regulation} \qquad (15)$$

Obviously, this technique will vastly improve overall regulation where the input fluctuates over a relatively wide range. As an example, let's say the input varies by ±20% and the regulation of each individual stage reduces the variation by a factor of 1/20. This then gives an overall output variation of $\pm 20\% \times (1/20)^2$ or ±0.05%.

The next two factors in equation (10) affecting regulation are changes in load current and temperature excursions. In order to minimize changes for load current variation, the output impedance $R_Z R_S/(R_Z + R_S)$ will have to be reduced. This can only be done by having a lower zener impedance because the value of R_S is fixed by circuit requirements. There are basically two ways that a lower zener impedance can be achieved. One, a higher wattage device can be used which allows for an increase in zener current of which will reduce the impedance. The other technique is to series lower voltage devices to obtain the desired equivalent voltage, so that the sum of the impedance is less than that for a single high voltage device. So to speak, this technique will kill two birds with one stone, as it can also be used to minimize temperature induced variations of the regulator.

In most regulator applications, the single most detrimental factor affecting regulation is that of variation in junction temperature. The junction temperature is a function of both the ambient temperature and that of self heating. In order to illustrate how the overall temperature coefficient is improved with series lower voltage zener, a mathematical relationship can be developed. Consider the diagram of Figure 5.

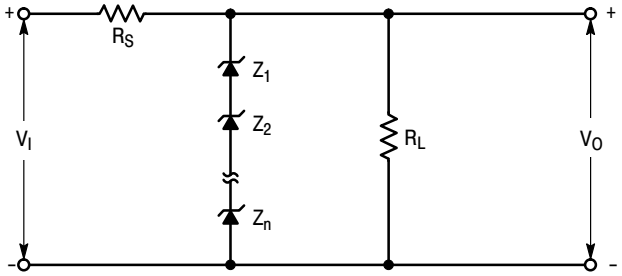


Figure 5. Series Zener Improve Dynamic Impedance and Temperature Coefficient

With the temperature coefficient TC defined as the % change per °C, the change in output for a given temperature range will equal some overall TC x ΔT x Total V_Z. Such as

$$\Delta V_{O(\Delta T)} = TC\ \Delta T\ (V_{Z1} + V_{Z2} + \ldots + V_{ZN}) \qquad (16)$$

Obviously, the change in output will also be equal to the sum of the changes as attributed from each zener.

$$\Delta V_{O(\Delta T)} = \Delta T(TC_1V_{Z1} + TC_2V_{Z2} + \ldots + TC_NV_{ZN}) \qquad (17)$$

Setting the two equations equal to each other and solving for the overall TC, we get

$$TC\Delta T(V_{Z1} + V_{Z2} + \ldots + V_{ZN}) = \Delta T(TC_1V_{Z1} + TC_2V_{Z2} + \ldots + TC_NV_{ZN}) \qquad (18)$$

$$TC = \frac{TC_1\,V_{Z1} + TC_2V_{Z2} + \ldots + TC_NV_{ZN}}{V_{Z1} + V_{Z2} + \ldots + V_{ZN}} \qquad (19)$$

For equation (19) the overall temperature coefficient for any combination of series zeners can be calculated. Say for instance several identical zeners in series replace a single higher voltage zener. The new overall temperature coefficient will now be that of one of the low voltage devices. This allows the designer to go to the manufacturer's

data sheet and select a combination of low TC zener diodes in place of the single higher TC devices. Generally speaking, the technique of using multiple devices will also yield a lower dynamic impedance. Advantages of this technique are best demonstrated by example. Consider a 5 watt diode with a nominal zener voltage of 10 volts exhibits approximately 0.055% change in voltage per degree centrigrade, a 20 volt unit approximately 0.075%/°C, and a 100 volt unit approximately 0.1%/°C. In the case of the 100 volt diode, five 20 volt diodes could be connected together to provide the correct voltage reference, but the overall temperature coefficient would remain that of the low voltage units, i.e. 0.075%/°C. It should also be noted that the same series combination improves the overall zener impedance in addition to the temperature coefficient. A 20 volt, 5 watt ON Semiconductor zener diode has a maximum zener impedance of 3 ohms, compared to the 90 ohms impedance which is maximum for a 100 volt unit. Although these impedances are measured at different current levels, the series impedance of five 20 volt zener diodes is still much lower than that of a single 100 volt zener diode at the test current specified on the data sheet.

For the ultimate in zener shunt regulator performance, the aforementioned techniques can be combined with the proper selection of devices to yield an overall improvement in regulation. For instance, a multiple string of low voltage zener diodes can be used as a preregulator, with a series combination of zero TC reference diodes in the final stage such as Figure 6.

The first stage will reduce the large variation in V_I to some relatively low level, i.e. ΔV_Z. This ΔV_Z is optimized by utilizing a series combination of zeners to reduce the overall TC and ΔV_Z. Because of this small fluctuation of input to the second stage, and if R_L is constant, the biasing current of the TC units can be maintained at their specified level. This will give an output that is very precise and not significantly affected by changes in input voltage or junction temperature.

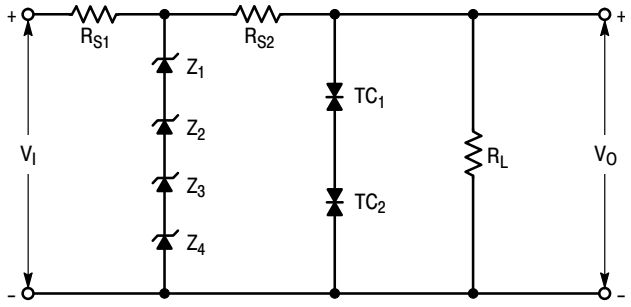


Figure 6. Series Zeners Cascaded With Series Reference Diodes for Improved Zener Shunt Regulation

The basic zener shunt regulator exhibits some inherent limitations to the designer. First of all, the zener is limited to its particular power dissipating rating which may be less than the required amount for a particular situation. The total magnitude of dissipation can be increased to some degree by utilizing series or parallel units. Zeners in series present few problems because individual voltages are additive and the devices all carry the same current and the extent that this technique can be used is only restricted by the feasibility of circuit parameters and cost. On the other hand, caution must be taken when attempting to parallel zener diodes. If the devices are not closely matched so that they all break over at the same voltage, the low voltage device will go into conduction first and ultimately carry all the current. In order to avoid this situation, the diodes should be matched for equal current sharing.

EXTENDING POWER AND CURRENT RANGE

The most common practice for extending the power handling capabilities of a regulator is to incorporate transistors in the design. This technique is discussed in detail in the following sections of this section. The second disadvantage to the basic zener shunt regulator is that because the device does not have a gain function, a feedback system is not possible with just the zener resistor combination. For very precise regulators, the design will normally be an electronic circuit consisting of transistor devices for control, probably a closed loop feedback system with a zener device as the basic referencing element.

The concept of regulation can be further extended and improved with the addition of transistors as the power absorbing elements to the zener diodes establishing a reference. There are three basic techniques used that combine zener diodes and transistors for voltage regulation. The shunt transistor type shown in Figure 7 will extend the power handling capabilities of the basic shunt regulator, and exhibit marked improvement in regulation.

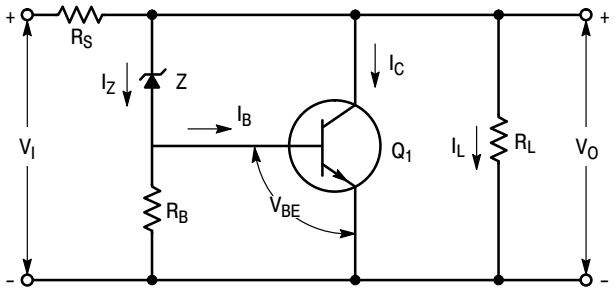


Figure 7. Basic Transistor Shunt Regulator

In this configuration the source resistance must be large enough to absorb the overvoltage in the same manner as in the conventional zener shunt regulator. Most of the shunt regulating current in this circuit will pass through the transistor reducing the current requirements of the zener diode by essentially the dc current gain of the transistor h_{FE}. Where the total regulating shunt current is:

$$I_S = I_Z + I_C = I_Z + I_B\, h_{FE}$$

where

$$I_Z = I_B + I_{RB} \text{ and } I_B >> I_{RB}$$

therefore

$$I_S \approx I_Z + I_Z\, h_{FE} = I_Z\,(1 + h_{FE}) \qquad (20)$$

The output voltage is the reference voltage V_Z plus the forward junction drop from base to emitter V_{BE} of the transistor.

$$V_O = V_Z + V_{BE} \quad (21)$$

The values of components and their operating condition is dictated by the specific input and output requirements and the characteristics of the designer's chosen devices, as shown in the following relations:

$$R_S = \frac{V_{I(min)} - V_{O(max)}}{I_{Z(min)}\,[1+h_{FE(min)}] + I_{L(max)}} \quad (22)$$

$$R_B = \frac{V_{I(min)} - V_{Z(max)}}{I_{Z(min)}} \quad (23)$$

$$P_{DZ} = I_{Z(max)}\,V_{Z(max)} \quad (24)$$

when

$$I_{Z(max)} = \left[\frac{V_{I(max)} - V_{O(min)}}{R_S} - I_{L(min)}\left(\frac{1}{1 + h_{FE(min)}}\right)\right] \quad (25)$$

hence

$$P_{DZ} = \left[\frac{V_{I(max)} - V_{O(min)}}{R_S} - I_{L(min)}\left(\frac{V_{Z(max)}}{1 + h_{FE(min)}}\right)\right] \quad (26)$$

$$P_{DQ} = \left[\frac{V_{I(max)} - V_{O(min)}}{R_S} - I_{L(min)}\left(V_{O(max)}\right)\right] \quad (27)$$

Regulation with this circuit is derived in essentially the same manner as in the shunt zener circuit, where the output impedance is low and the output voltage is a function of the reference voltage. The regulation is improved with this configuration because the small signal output impedance is reduced by the gain of Q_1 by $1/h_{FE}$.

One other highly desirable feature of this type of regulator is that the output is somewhat self compensating for temperature changes by the opposing changes in V_Z and V_{BE} for $V_Z \approx 10$ volts. With the zener having a positive 2 mV/°C TC and the transistor base to emitter being a negative 2 mV/°C TC, therefore, a change in one is cancelled by the change in the other. Even though this circuit is a very effective regulator it is somewhat undesirable from an efficiency standpoint. Because the magnitude of R_S is required to be large, and it must carry the entire input current, a large percentage of power is lost from input to output.

EMITTER FOLLOWER REGULATOR

Another basic technique of transistor-zener regulation is that of the emitter follower type shown in Figure 8.

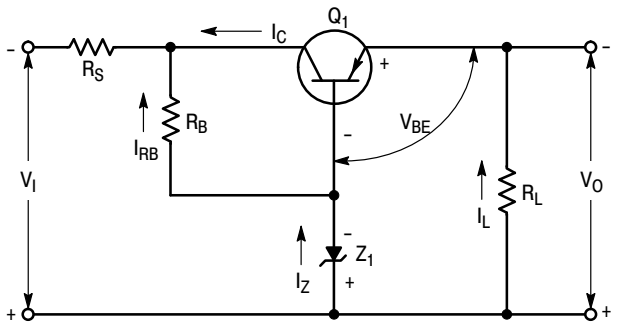


Figure 8. Emitter Follower Regulator

This circuit has the desirable feature of using a series transistor to absorb overvoltages instead of a large fixed resistor, thereby giving a significant improvement in efficiency over the shunt type regulator. The transistor must be capable of carrying the entire load current and withstanding voltages equal to the input voltage minus the load voltage. This, of course, imposes a much more stringent power handling requirement upon the transistor than was required in the shunt regulator. The output voltage is a function of the zener reference voltage and the base to emitter drop of Q_1 as expressed by the equation (28).

$$V_O = V_Z - V_{BE} \quad (28)$$

The load current is approximately equal to the transistor collector current, such as shown in equation (29).

$$I_{L(max)} \approx I_{C(max)} \quad (29)$$

The designer must select a transistor that will meet the following basic requirements:

$$P_D \cong (V_{I(max)} - V_O)I_{L(max)}$$

$$I_{C(max)} \approx I_{L(max)}$$

$$BV_{CES} \geq (V_{I(max)} - V_O) \quad (30)$$

Depending upon the designer's choice of a transistor and the imposed circuit requirements, the operation conditions of the circuit are expressed by the following equations:

$$V_Z = V_O + V_{BE}$$

$$= V_O + I_{L(max)}/g_{FE(min)} \text{ @ } I_{L(max)}$$

$$R_S = \frac{V_{I(min)} - V_Z - V_{CE(min)} \text{ @ } I_{L(max)}}{I_{L(max)}} \quad (31)$$

Where $V_{CE(min)}$ is an arbitrary value of minimum collector to emitter voltage and g_{FE} is the transconductance.

This is sufficient to keep the transistor out of saturation, which is usually about 2 volts.

$$R_B = \frac{V_{CE(min)} @ I_{L(max)}}{I_{L(max)}/h_{FE(min)} @ I_{L(max)} + I_{Z(min)}} \quad (32)$$

$$I_{Z(max)} = \frac{V_{I(max)} - V_Z}{R_B + R_Z} \quad (33)$$

$$P_{DZ} = I_{Z(max)} V_Z \quad (34)$$

$$\text{Actual } P_{DQ} = (V_{I(max)} - V_O)\, I_{L(max)} \quad (35)$$

There are two primary factors that effect the regulation most in a circuit of this type. First of all, the zener current may vary over a considerable range as the input changes from minimum to maximum and this, of course, may have a significant effect on the value of V_Z and therefore V_O. Secondly, V_Z and V_{BE} will both be effected by temperature changes which are additive on their effect of output voltage. This can be seen by altering equation (28) to show changes in V_O as dependent on temperature, see equation (36).

$$V_{O(\Delta T)} = \Delta T[(+TC)\, V_Z - (-TC)\, V_{BE}] \quad (36)$$

The effects of these detrimental factors can be minimized by replacing the bleeder resistor R_B with a constant current source and the zener with a reference diode in series with a forward biased diode (see Figure 9).

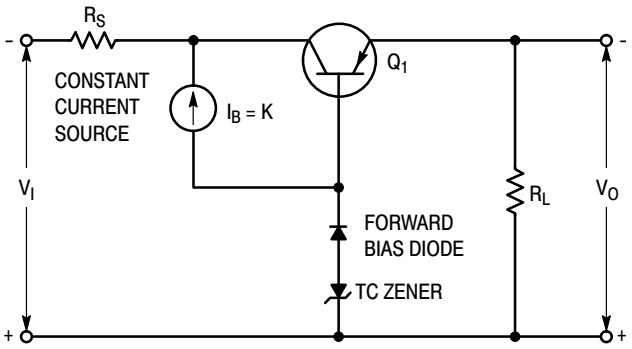


Figure 9. Improved Emitter Follower Regulator

The constant current source can be either a current limiter diode or a transistor source. The current limiter diode is ideally suited for applications of this type, because it will supply the same biasing current irregardless of collector to base voltage swing as long as it is within the voltage limits of the device. This technique will overcome changes in V_Z for changes in I_Z and temperature, but changes in V_{BE} due to load current changes are still directly reflected upon the output. This can be reduced somewhat by combining a transistor with the zener for the shunt control element as illustrated in Figure 10.

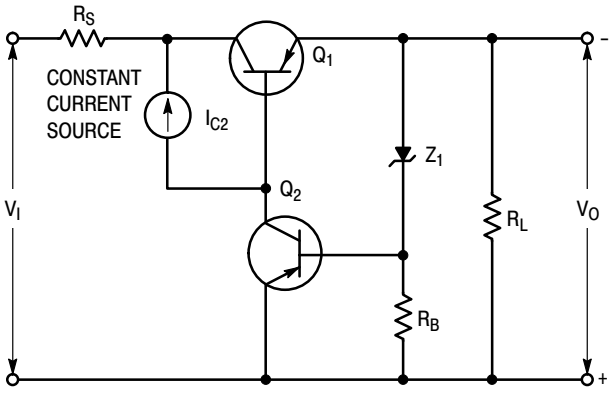


Figure 10. Series Pass Regulator

This is the third basic technique used for transistor-zener regulators. This technique or at least a variation of it, finds the widest use in practical applications. In this circuit the transistor Q_1 is still the series control device operating as an emitter follower. The output voltage is now established by the transistor Q_2 base to emitter voltage and the zener voltage. Because the zener is only supplying base drive to Q_2, and it derives its bias from the output, the zener current remains essential constant, which minimizes changes in V_Z due to I_Z excursions. Also, it may be possible ($V_Z \approx 10$ V) to match the zener to the base-emitter junction of Q_2 for an output that is insensitive to temperature changes. The constant current source looks like a very high load impedance to the collector of Q_2 thus assuming a very high voltage gain. There are three primary advantages gained with this configuration over the basic emitter follower:

1. The increased voltage gain of the circuit with the addition of Q_2 will improve regulation for changes in both load and input.
2. The zener current excursions are reduced, thereby improving regulation.
3. For certain voltages the configuration allows good temperature compensation by matching the temperature characteristics of the zener to the base-emitter junction of Q_2.

The series pass regulator is superior to the other transistor regulators thus far discussed. It has good efficiency, better stability and regulation, and is simple enough to be economically practical for a large percentage of applications.

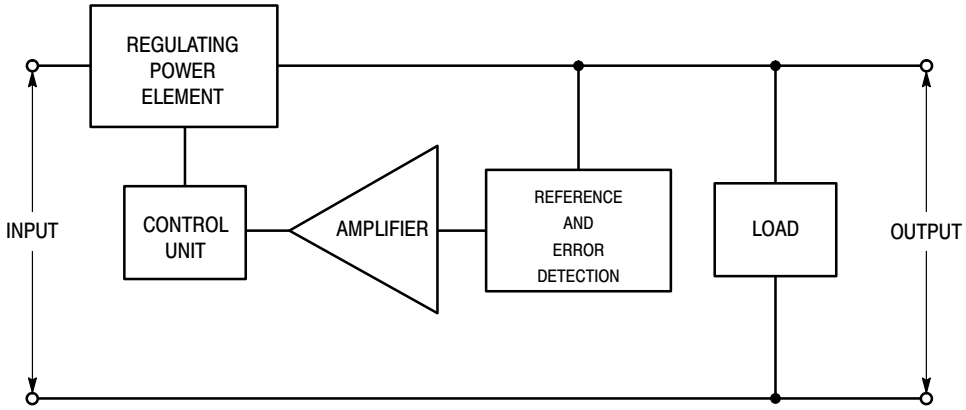


Figure 11. Block Diagram of Regulator with Feedback

EMPLOYING FEEDBACK FOR OPTIMUM REGULATION

The regulators discussed thus far do not employ any feedback techniques for precise control and compensation and, therefore, find limited use where an ultra precise regulator is required. In the more sophisticated regulators some form of error detection is incorporated and amplified through a feedback network to closely control the power elements as illustrated in the block diagram of Figure 11.

Regulating circuits of this type will vary in complexity and configuration from application to application. This technique can best be illustrated with a couple of actual circuits of this type. The feedback regulators will generally be some form of series pass regulator, for optimum performance and efficiency. A practical circuit of this type that is extensively utilized is shown in Figure 12.

In this circuit, the zener establishes a reference level for the differential amplifier composed of Q_4 and Q_5 which will set the base drive for the control transistor Q_3 to regulate the series high gain transistor combination of Q_1 and Q_2. The differential amplifier samples the output at the voltage dividing network of R_8, R_9, and R_{10}. This is compared to the reference voltage provided by the zener Z_1. The difference, if any, is amplified and fed back to the control elements. By adjusting the potentiometer, R_9, the output level can be set to any desired value within the range of the supply. (The output voltage is set by the relation $V_O = V_Z[(R_X + R_Y)/R_X]$.) By matching the transistor Q_4 and Q_5 for variations in V_{BE} and gain with temperature changes and incorporating a temperature compensated diode as the reference, the circuit will be ultra stable to temperature effects. The regulation and stability of this circuit is very good, and for this reason is used in a large percentage of commercial power supplies.

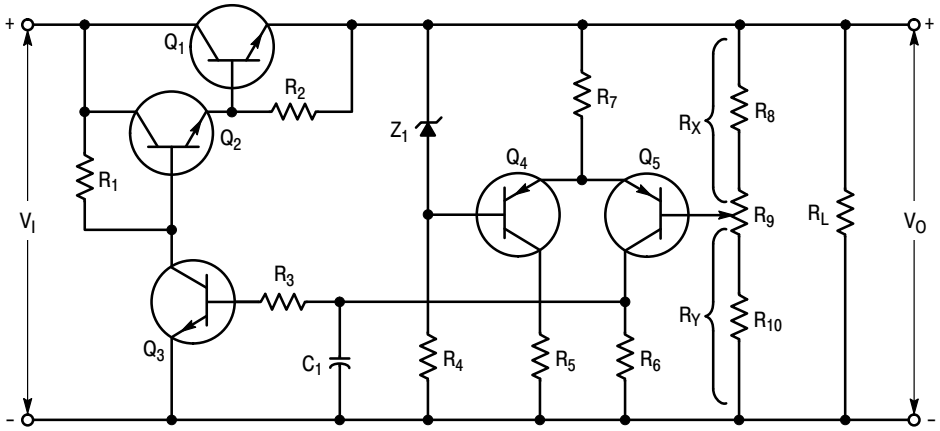


Figure 12. Series Pass Regulator with Error Detection and Feedback Amplification Derived from a Differential Amplifier

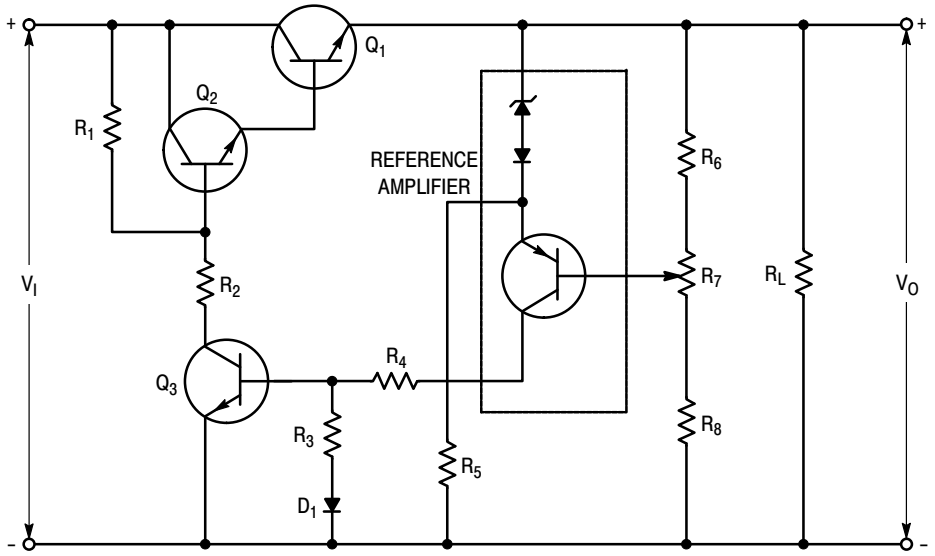


Figure 13. Series Pass Regulator with Temperature Compensated Reference Amplifier

Another variation of the feedback series pass regulator is shown in Figure 13. This circuit incorporates a stable temperature compensated reference amplifier as the primary control element.

This circuit also employs error detection and amplified feedback compensation. It is an improved version over the basic series pass regulator shown in Figure 10. The series element is composed of a Darlington high gain configuration formed by Q_1 and Q_2 for an improved regulation factor. The combined gain of the reference amplifier and Q_3 is incorporated to control the series unit. This reduced the required collector current change of the reference amplifier to control the regulator so that the bias current remains close to the specified current for low temperature coefficient. Also the germanium diode D_1 will compensate for the base to emitter change in Q_3 and keep the reference amplifier collector biasing current fairly constant with temperature changes. Proper biasing of the zener and transistor in the reference amplifier must be adhered to if the output voltage changes are to be minimized.

CONSTANT CURRENT SOURCES FOR REGULATOR APPLICATIONS

Several places throughout this section emphasize the need for maintaining a constant current level in the various biasing circuits for optimum regulation. As was mentioned previously in the discussion on the basic series pass regulator, the current limiter diode can be effectively used for the purpose.

Aside from the current limiter diode a transistorized source can be used. A widely used technique is shown incorporated in a basic series pass regulator in Figure 14.

The circuit is used as a preregulated current source to supply the biasing current to the transistor Q_2. The constant current circuit is seldom used alone, but does find wide use in conjunction with voltage regulators to supply biasing current to transistors or reference diodes for stable operation. The Zener Z_2 establishes a fixed voltage across R_E and the base to emitter of Q_3. This gives an emitter current of $I_E = (V_Z - V_{BE})/R_E$ which will vary only slightly for changes in input voltage and temperature.

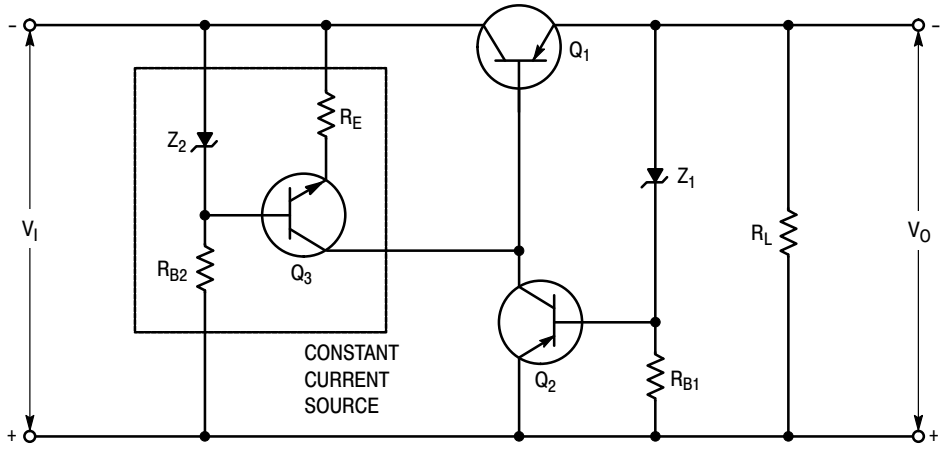


Figure 14. Constant Current Source Incorporated in a Basic Regulator Circuit

IMPEDANCE CANCELLATION

One of the most common applications of zener diodes is in the general category of reference voltage supplies. The function of the zener diode in such applications is to provide a stable reference voltage during input voltage variations. This function is complicated by the zener diode impedance, which effectively causes an incremental change in zener breakdown voltage with changing zener current.

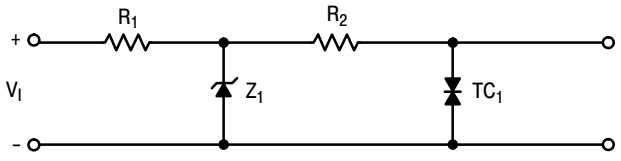


Figure 15. Impedance Cancellation with An Uncompensated Zener

It is possible, however, by employing a bridge type circuit which includes the zener diode and current regulating resistance in its branch legs, to effectively cancel the effect of the zener impedance. Consider the circuit of Figure 15 as an example. This is the common configuration for a zener diode voltage regulating system. The zener impedance at 20 mA of a 1N4740 diode is typically 2 ohms. If the supply voltage now changes from 30 V to 40 V, the diode current determined by R_1 changes from 20 to 30 mA; the average zener impedance becomes 1.9 ohms; and the reference voltage shifts by 19 mV. This represents a reference change of .19%, an amount far too large for an input change of 30% in most reference supplies.

The effect of zener impedance change with current is relatively small for most input changes and will be neglected for this analysis. Assuming constant zener impedance, the zener voltage is approximated by

$$V'_Z = V_Z + Z(I'_Z - I_Z) \quad (37)$$

where V'_Z is the new zener voltage
V_Z is the former zener voltage
I'_Z is the new zener current
I_Z is the new zener current flowing at V_Z
R_Z is the zener impedance

Then $\Delta V_Z = Z\Delta I_Z$

Let the input voltage V_I in Figure 15 increase by an amount ΔV_I

$$\text{Then } \Delta I = \frac{\Delta V_I - \Delta V_Z}{R_1} \quad (38)$$

$$\text{Also } \Delta I = \frac{\Delta V_Z}{R_Z} \quad (39)$$

Solving $\Delta V_I R_Z - \Delta V_Z R_Z - \Delta V_Z R_1 = 0$

$$\text{Or} \quad \frac{\Delta V_Z}{\Delta V_I} = \frac{R_Z}{R_1 + R_Z} \quad (40)$$

Equation 40 merely states that the change in reference voltage with input tends to zero when the zener impedance tends also to zero, as expected.

The figure of merit equation can be applied to the circuits of Figure 16 and 17 to explain impedance cancellation. The Change Factor equations for each leg and the reference voltage V_R are:

$$CF_{VZ} = \frac{\Delta V_Z}{\Delta V_I} = \frac{R_Z}{R_1 + R_Z} = R_A \quad (41)$$

$$CF_{V2} = \frac{\Delta V_2}{\Delta V_I} = \frac{R_3}{R_2 + R_3} = R_B \quad (42)$$

$$CF_{VR} = \frac{\Delta V_R}{\Delta V_I} = \frac{R_Z}{R_1 + R_Z} = \frac{R_3}{R_2 + R_3} = R_A - R_B \quad (43)$$

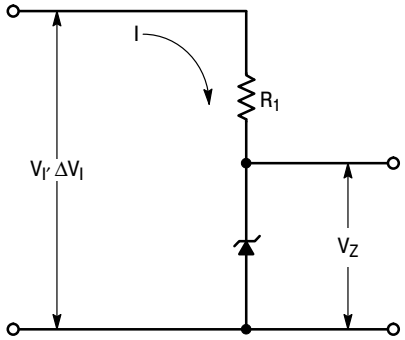


Figure 16. Standard Voltage Regulation Circuit

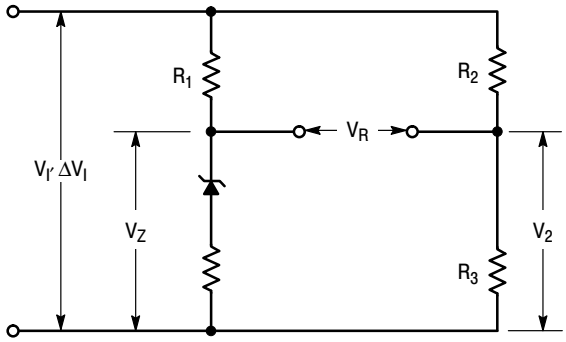


Figure 17. Impedance Cancellation Bridge

Since the design is to minimize CF_{VR}, R_B can be set equal to R_A. The Input Regulation Factors are:

$$\gamma VZ = \frac{\Delta V_Z}{\Delta V_I}\left(\frac{V_I}{V_Z}\right) = \frac{1}{1 + \frac{V_Z}{V_I}\left(\frac{R_1}{RZ}\right)} \quad (44)$$

$$\gamma V2 = \frac{\Delta V_2}{\Delta V_I}\left(\frac{V_I}{V_2}\right) = 1 \quad (45)$$

$$\gamma VR = \frac{\Delta V_R}{\Delta V_I}\left(\frac{V_I}{V_R}\right) = \frac{1}{1 + \left(\frac{V_Z}{V_I}\right)\left(\frac{R_1}{RZ}\right)\left(\frac{1}{1 - \frac{R_B}{R_A}}\right)} \quad (46)$$

It is seen that γVR can be minimized by setting $R_B = R_A$.

Note that it is not necessary to match R_3 to R_Z and R_2 to R_1. Thus R_3 and R_2 can be large and hence dissipate low power. This discussion is assuming very light load currents.

ZENER VOLTAGE SENSING CIRCUITS AND APPLICATIONS

BASIC CONCEPTS OF VOLTAGE SENSING

Numerous electronic circuits require a signal or voltage level to be sensed for circuit actuation or function control. The circuit may alter its mode of operation whenever an interdependent signal reaches a particular magnitude (either higher or lower than a specified value). These sensing functions may be accomplished by incorporating a voltage dependent device in the system creating a switching action that controls the overall operation of the circuit.

The zener diode is ideally suited for most sensing applications because of its voltage dependent characteristics. The following sections are some of the more common applications and techniques that utilize the zener in a voltage sensing capacity.

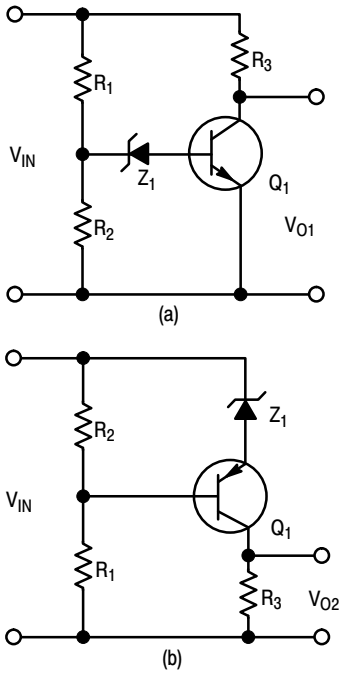


Figure 1. Basic Transistor-Zener Diode Sensing Circuits

TRANSISTOR-ZENER SENSING CIRCUITS

The zener diode probably finds its greatest use in sensing applications in conjunction with other semiconductor devices. Two basic widely used techniques are illustrated in Figures 1a and 1b.

In both of these circuits the output is a function of the input voltage level. As the input goes from low to high, the output will switch from either high to low (base sense circuit) or low to high (emitter sense circuit), (see Figure 2).

The base sense circuit of Figure 1a operates as follows: When the input voltage is low, the voltage dropped across R_2 is not sufficient to bias the zener diode and base emitter junction into conduction, therefore, the transistor will not conduct. This causes a high voltage from collector to emitter. When the input becomes high, the zener is biased into conduction, the transistor turns on, and the collector to emitter voltage, which is the output, drops to a low value.

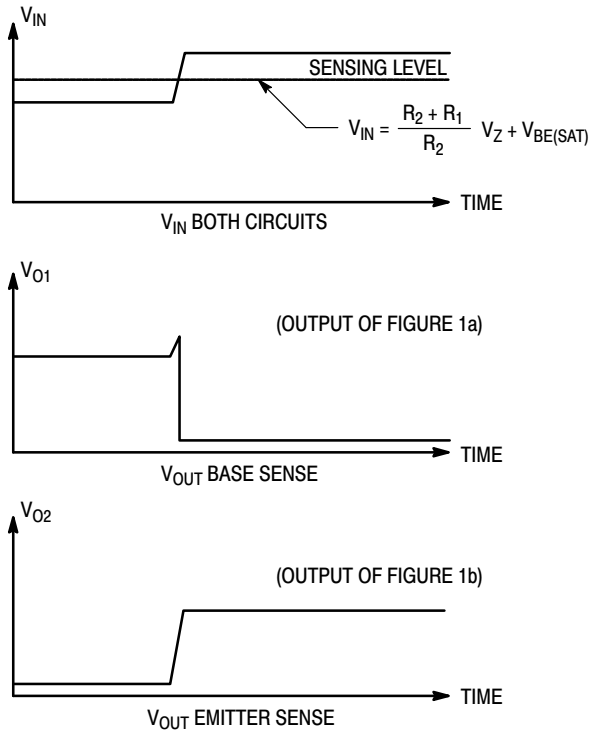


Figure 2. Outputs of Transistor-Zener Voltage Sensing Circuits

The emitter sense circuit of Figure 1b operates as follows: When the input is low the voltage drop across R_3 (the output) is negligible. As the input voltage increases the voltage drop across R_2 biases the zener into conduction and forward biases the base-emitter junction. A large voltage drop across R_3 (the output voltage) is equal to the product of the collector current times the resistance, R_3. The following relationships indicate the basic operating conditions for the circuits in Figure 1.

Circuit	Output
1a	High $V_{OUT} = V_{IN} - I_{CO}R_3 \cong V_{IN}$ Low $V_{OUT} = V_{IN} - I_CR_3 = V_{CE(sat)}$
1b	Low $V_{OUT} = V_{IN} - V_Z - V_{CE(off)} = I_{CO}R_3$ High $V_{OUT} = V_{IN} - V_{CE(sat)} = I_CR_3$

In addition, the basic circuits of Figure 1 can be rearranged to provide inverse output.

AUTOMOTIVE ALTERNATOR VOLTAGE REGULATOR

Electromechanical devices have been employed for many years as voltage regulators, however, the regulation setting of these devices tend to change and have mechanical contact problems. A solid state regulator that controls the charge rate by sensing the battery voltage is inherently more accurate and reliable. A schematic of a simplified solid state voltage regulator is shown in Figure 3.

The purpose of an alternator regulator is to control the battery charging current from the alternator. The charge level of the battery is proportional to the battery voltage level. Consequently, the regulator must monitor the battery voltage level allowing charging current to pass when the battery voltage is low. When the battery has attained the proper charge the charging current is switched off. In the case of the solid state regulator of Figure 3, the charging current is controlled by switching the alternator field current on and off with a series transistor switch, Q_2. The switching action of Q_2 is controlled by a voltage sensing circuit that is identical to the base sense circuit of Figure 1a. When under-charged, the zener Z_1 does not conduct keeping Q_1 off. The collector-emitter voltage of Q_1 supplies a forward bias to the base-emitter of Q_2, turning it on. With Q_2 turned on, the alternator field is energized allowing a charging current to be delivered to the battery. When the battery attains a proper charge level, the zener conducts causing Q_1 to turn on, and effectively shorting out the base-emitter junction of Q_2. This short circuit cuts off Q_2, turns off the current flowing in the field coil which consequently, reduces the output of the alternator. Diode D_1 acts as a field suppressor preventing the build up of a high induced voltage across the coil when the coil current is interrupted.

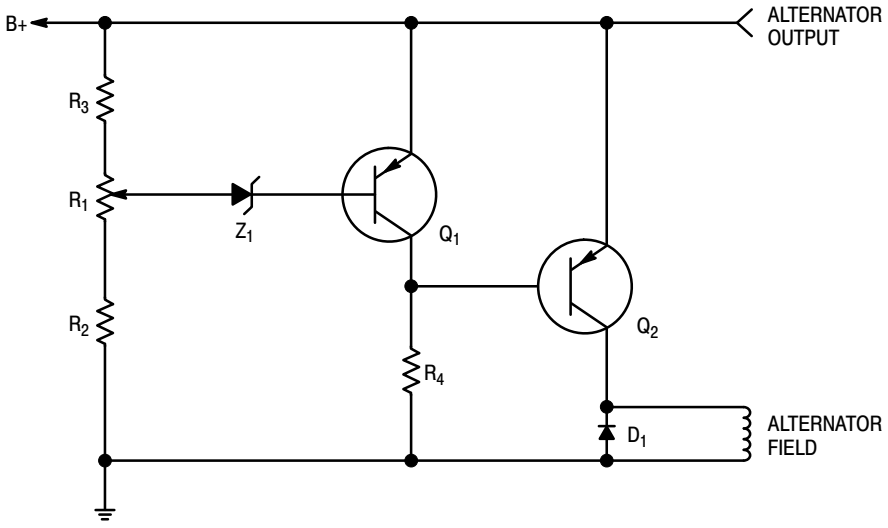


Figure 3. Simplified Solid State Voltage Regulator

In actual operation, this switching action occurs many times each second, depending upon the current drain from the battery. The battery charge, therefore, remains essentially constant and at the maximum value for optimum operation.

A schematic of a complete alternator voltage regulator is shown in Figure 4.

It is also possible to perform the alternator regulation function with the sensing element in the emitter of the control transistor as shown in Figure 5.

In this configuration, the sensing circuit is composed of Z_1 and Q_1 with biasing components. It is similar to the sensing circuit shown in Figure 1b. The potentiometer R_1 adjusts the conduction point of Q_1 establishing the proper charge level. When the battery has reached the desired level, Q_1 begins to conduct. This draws Q_2 into conduction, and therefore shorts off Q_3 which is supplying power to the alternator field. This type of regulator offers greater sensitivity with an increase in cost.

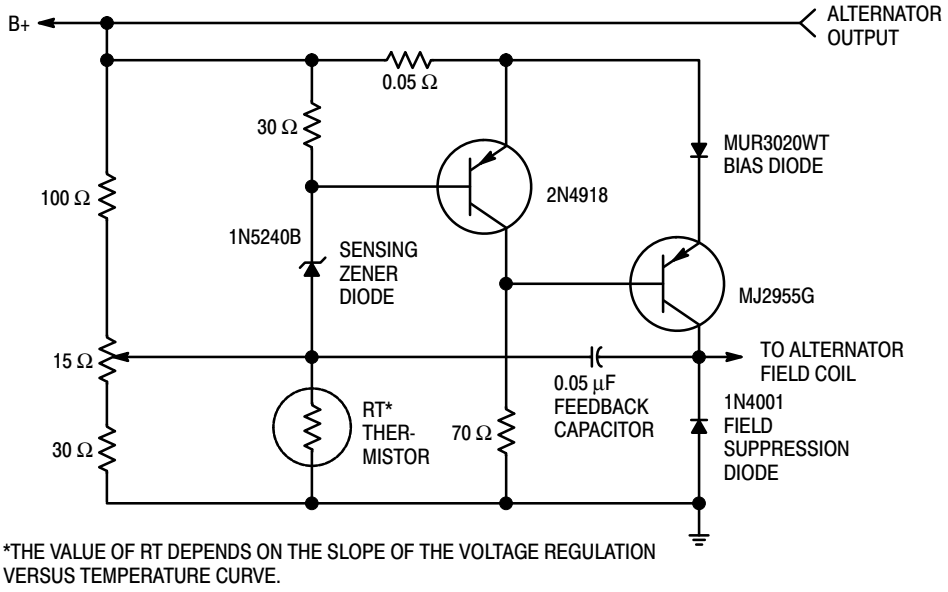


Figure 4. Complete Solid State Alternator Voltage Regulator

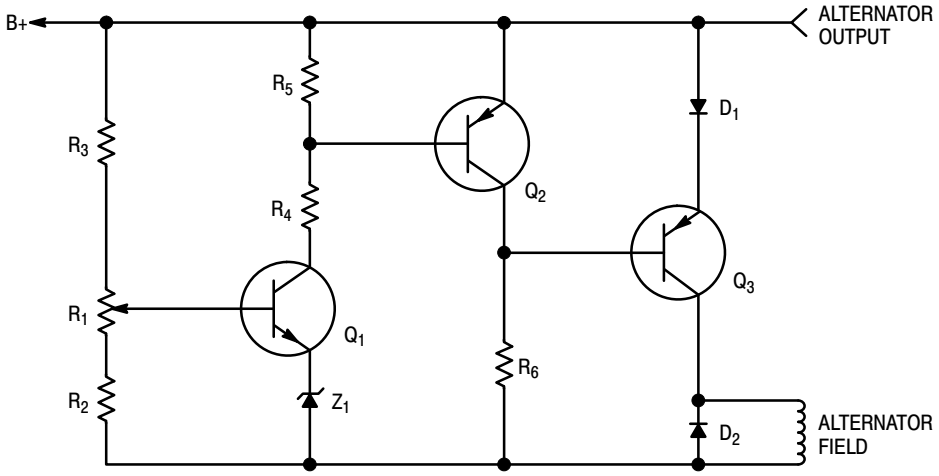


Figure 5. Alternator Regulator with Emitter Sensor

ZENER-RESISTOR VOLTAGE SENSING

A simple but useful sense circuit can be made from just a Zener diode and resistor such as shown in Figure 6.

Whenever the applied signal exceeds the specific Zener voltage V_Z, the difference appears across the dropping resistor R. This level dependent differential voltage can be used for level detection, magnitude reduction, wave shaping, etc. An illustrative application of the simple series Zener sensor is shown in Figure 7, where the resistor drop is monitored with a voltmeter.

If, for example, the input is variable from 24 to 28 volts, a 30 voltmeter would normally be required. Unfortunately, a 4 volt range of values on a 30 volt scale utilizes only 13.3% of the meter movement — greatly limiting the accuracy with which the meter can be read. By employing a 20 volt zener, one can use a 10 voltmeter instead of the 30 volt unit, thereby utilizing 40% of the meter movement instead of 13.3% with a corresponding increase in accuracy and readability. For ultimate accuracy a 24 volt zener could be combined with a 5 voltmeter. This combination would have the disadvantage of providing little room for voltage fluctuations, however.

In Figure 8, a number of sequentially higher-voltage Zener sense circuits are cascaded to actuate transistor switches. As each goes into avalanche its respective switching transistor is turned on, actuating the indicator light for that particular voltage level. This technique can be expanded and modified to use the zener sensors to actuate some form of logic system.

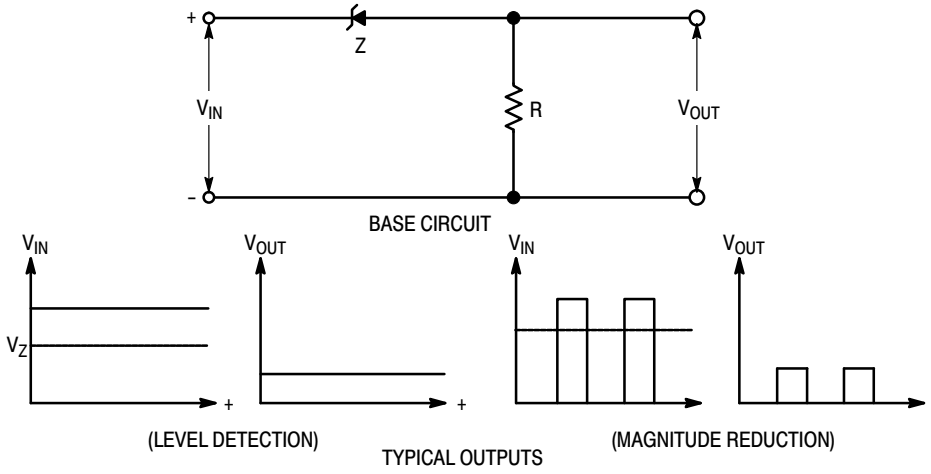


Figure 6. Zener-Resistor Voltage Sensitive Circuit

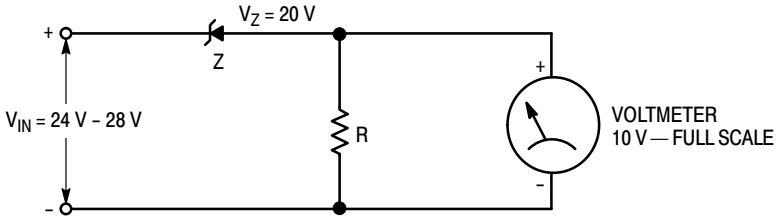


Figure 7. Improving Meter Resolution

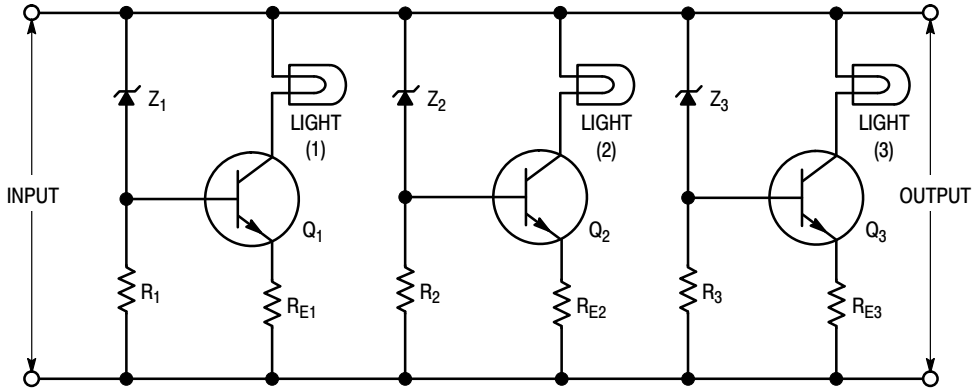


Figure 8. Sequential Voltage Level Indicator

MISCELLANEOUS APPLICATIONS OF ZENER TYPE DEVICES

INTRODUCTION

Many of the commonly used applications of zener diodes have been illustrated in some depth in the preceding sections. This section shows how a zener diode may be used in some rarer applications such as voltage translators, to provide constant current, wave shaping, frequency control and synchronized SCR triggers.

The circuits used in this section are not intended as finished designs since only a few component values are given. The intent is to show some general broad ideas and not specific designs aimed at a narrow use.

FREQUENCY REGULATION OF A DC TO AC INVERTER

Zener diodes are often used in control circuits, usually to control the magnitude of the output voltage or current. In this unusual application, however, the zener is used to control the output frequency of a current feedback inverter. The circuit is shown in Figure 1.

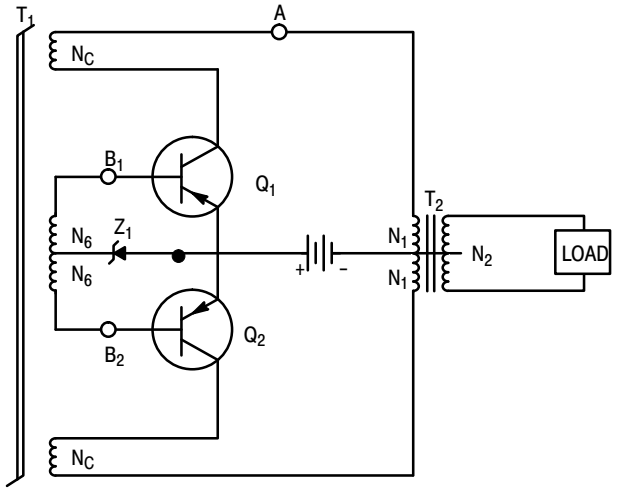


Figure 1. Frequency Controlled Current Feedback Inverter

The transformer T_1 functions as a current transformer providing base current $I_B = (N_C/N_B)I_C$. Without the zener diode, the voltage across N_B windings of the timing transformer T_1 is clamped to V_{BE} of the ON device, giving an inverter frequency of

$$f = \frac{V_{BE} \times 10^8}{4B_{S1}A_1N_B}$$

where $B_{S1}A_1$ is the flux capacity of T_1 transformer core. The effect on output frequency of V_{BE} variations due to changing load or temperature can be reduced by using a zener diode in series with V_{BE} as shown in Figure 1. For this circuit, the output frequency is given by

$$f = \frac{(V_{BE} + V_Z) \times 10^8}{4B_{S1}A_1N_B}$$

If V_{BE} is small compared to the zener voltage V_Z, good frequency accuracy is possible. For example, with V_Z = 9.1 volts, a 40 Watt inverter using MJD2955 transistors (operating from a 12 volt supply), exhibited frequency regulation of ±2% with ±25% load variation.

Care should be taken not to exceed $V_{(BR)EBO}$ of the non-conducting transistor, since the reverse emitter-base voltage will be twice the introduced series voltage, plus V_{BE} of the conducting device.

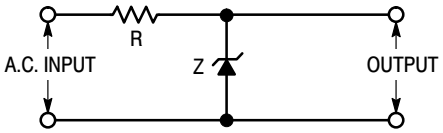
Transformer T_2 should not saturate at the lowest inverter frequency.

Inverter starting is facilitated by placing a resistor from point A to B_1 or a capacitor from A to B_2.

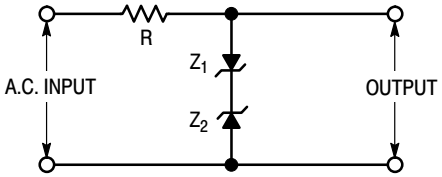
SIMPLE SQUARE WAVE GENERATOR

The zener diode is widely used in wave shaping circuits; one of its best known applications is a simple square wave generator. In this application, the zener clips sinusoidal waves producing a square wave such as shown in Figure 2a. In order to generate a wave with reasonably vertical sides, the ac voltage must be considerably higher than the zener voltage.

Clipper diodes with opposing junctions built into the device are ideal for applications of the type shown in Figure 2b.



(a) Single Zener Diode Square Wave Generator



(b) Opposed Zener Diodes Square Wave Generator

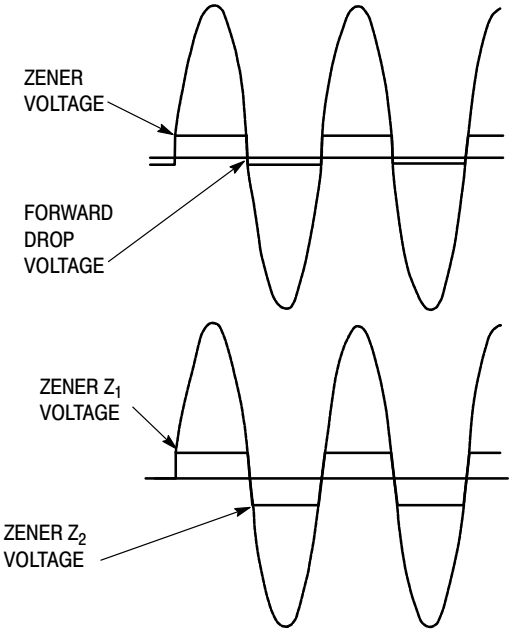


Figure 2. Zener Diode Square Wave Generator

MEASUREMENT OF ZENER VOLTAGE TO THERMAL EQUILIBRIUM WITH PULSED TEST CURRENT

INTRODUCTION

This paper discusses the zener voltage correlation problem which sometimes exists between the manufacturer and the customer's incoming inspection. A method is shown to aid in the correlation of zener voltage between thermal equilibrium and pulse testing. A unique double pulsed sample and hold test circuit is presented which improves the accuracy of correlation.[1]

Several zener voltages versus zener pulsed test current curves are shown for four package styles. An appendix is attached for incoming inspection groups giving detailed information on tolerances involved in correlation.

For many years the major difficulty with zener diode testing seemed to be correlation of tight tolerance voltage specifications where accuracy between different test setups was the main problem. The industry standard and the EIA Registration system adopted thermal equilibrium testing of zener diodes as the basic test condition unless otherwise specified. Thermal equilibrium was chosen because it was the most common condition in the final circuit design and it was the condition that the design engineers needed for their circuit design and device selection. Thermal equilibrium testing was also fairly simple to set-up for sample testing at incoming inspection of standard tolerance zeners.

In recent years with the advent of economical computerized test systems many incoming inspection areas have implemented computer testing of zener diodes which has been generating a new wave of correlation problems between customers and suppliers of zener diodes.

The computerized test system uses short duration pulse test techniques for testing zener diodes which does not directly match the industry standard thermal equilibrium test specifications.

This paper was prepared in an attempt to clarify the differences between thermal equilibrium and short duration pulse testing of zener diodes, to provide a test circuit that allows evaluation at various pulse widths and a suggested procedure for incoming inspection areas that will allow meaningful correlation between thermal equilibrium and pulse testing.

In the measurement of zener voltage (V_Z), the temperature coefficient effect combined with test current heating can present a problem if one is attempting to correlate V_Z measurements made by another party (Final Test, Quality Assurance or Incoming Inspection).[2] This paper is intended as an aid in determining V_Z at some test current (I_{ZT}) pulse width other than the pulse width used by the manufacturer.

Thermal equilibrium (TE) is reached when diode junction temperature has stabilized and no further change will occur in V_Z if the I_{ZT} time is increased.[2] This absolute value can vary depending on the mounting method and amount of heatsinking. Therefore, thermal equilibrium conditions have to be defined before meaningful correlation can exist.

Normalized V_Z curves are shown for four package styles and for three to five voltage ratings per package. Pulse widths from 1 ms up to 100 seconds were used to arrive at or near thermal equilibrium for all packages with a given method of mounting.

Mounting

There are five conditions that can affect the correlation of V_Z measurements and are: 1) instrumentation, 2) T_A, 3) I_{ZT} time, 4) P_D and 5) mounting. The importance of the first four conditions is obvious but the last one, mounting, can make the difference between good and poor correlation. The mounting can have a very important part in V_Z correlation as it controls the amount of heat and rate of heat removal from the diode by the mass and material in contact with the diode package.

Two glass axial lead packages (DO-35 and DO-41), curves (Figures 5 and 6) were measured with standard Grayhill clips and a modified version of the Grayhill clips to permit lead length adjustment.

Test Circuit

The test circuit (Figure 8) consists of standard CMOS logic for pulse generation, inverting and delaying. The logic drives three bipolar transistors for generation of the power pulse for I_{ZT}. V_Z is fed into an unique sample and hold (S/H) circuit consisting of two high input impedance operational amplifiers and a field effect transistor switch.

For greater accuracy in V_Z measurements using a single pulse test current, the FET switch is double pulsed. Double pulsing the FET switch for charging the S/H capacitor increases accuracy of the charge on the capacitor as the second pulse permits charging the capacitor closer to the final value of V_Z.

The timing required for the two pulse system is shown in waveform G-3C whereby the initial sample pulse is delayed from time zero by a fixed 100 μs to allow settling time and the second pulse is variable in time to measure the analog input at that particular point. The power pulse (waveform G-2D) must also encompass the second sample pulse.

To generate these waveforms, four time delay monostable multivibrators (MV) are required. Also, an astable MV, is required for free-running operation; single pulsing is simply initiated by a push-button switch S1. All of the pulse generators are fashioned from two input, CMOS NOR gates; thus three quad gate packages (MC14001) are required. Gates 1A and 1B form a classical CMOS astable MV clock and the other gates (with the exception of Gate 2D) comprise the two input NOR gate configured monostable MV's. The Pulse Width variable delay output (Gate 1D) positions the second sample pulse and also triggers the 100 μs Delay MV and the 200 μs Extended Power Pulse MV, The respective positive going outputs from gates 3A and 2C are diode NOR'ed to trigger the Sample Gate MV whose output will consequently be the two sample pulses. These pulses then turn on the PNP transistor Q1 level translator and the following S/H N-channel FET series switch Q2. Op amps U4 and U5, configured as voltage followers, respectively provide the buffered low output impedance drive for the input and output of the S/H. Finally, the pulse extended Power Gate is derived by NORing (Gate 2D) the Pulse Width Output (Gate 1D) with the 200 μs MV output (Gate 2C). This negative aging gate then drives the Power Amplifier, which, in turn, powers the D.U.T. The power amplifier configuration consists of cascaded transistors Q3−Q5, scaled for test currents up to 2 A.

Push button switch (S4) is used to discharge the S/H capacitor. To adjust the zero control potentiometer, ground the non-inverting input (Pin 3) of U4 and discharge the S/H capacitor.

Testing

The voltage V_{CC}, should be about 50 volts higher than the D.U.T. and with R_C selected to limit the I_{ZT} pulse to a value making $V_{ZT}\ I_{ZT} = 1/4\ P_D$ (max), thus insuring a good current source. All testing was performed at a normal room temperature of 25°C. A single pulse (manual) was used and at a low enough rate that very little heat remained from the previous pulse.

The pulse width MV (1C and 1D) controls the width of the test pulse with a selector switch S3 (see Table 1 for capacitor values). Fixed widths in steps of 1, 3 and 5 from 1 ms to 10 seconds in either a repetitive mode or single pulse is available. For pulse widths greater than 10 seconds, a stop watch was used with push button switch (S1) and with the mode switch (S2) in the > 10 seconds position.

For all diodes with V_Z greater than about 6 volts a resistor voltage divider is used to maintain an input of about 6 V to the first op amp (U4) so as not to overload or saturate this device. The divider consists of R5 and R6 with R6 being 10 kΩ and R5 is selected for about a 6 V input to U4. Precision resistors or accurate known values are required for accurate voltage readout.

Table 1. S3 — Pulse Width

Switch Position	*C(μF)	t(ms)
1	0.001	1
2	0.004	3
3	0.006	5
4	0.01	10
5	0.04	30
6	0.06	50
7	0.1	100
8	0.4	300
9	0.6	500
10	1.0	1K
11	1.2	3K
12	6.0	5K
13	10	10K

*Approximate Values

Using Curves

Normalized V_Z versus I_{ZT} pulse width curves are shown in Figure 1 through 6. The type of heatsink used is shown or specified for each device package type. Obviously, it is beyond the scope of this paper to show curves for every voltage rating available for each package type. The object was to have a representative showing of voltages including when available, one diode with a negative temperature coefficient (TC).

These curves are actually a plot of thermal response versus time at one quarter of the rated power dissipation. With a given heatsink mounting, V_Z can be calculated at some pulse width other than the pulse width used to specify V_Z.

For example, refer to Figure 5 which shows normalized V_Z curves for the axial lead DO-35 glass package. Three mounting methods are shown to show how the mounting effects device heating and thus V_Z.

In Figure 5, the two curves generated using the Grayhill mountings are normalized to V_Z at TE using the ON Semiconductor fixture. There is very little difference in V_Z at pulse widths up to about 10 seconds and mounting only causes a very small error in V_Z. The maximum error occurs at TE between mountings and can be excessive if V_Z is specified at TE and a customer measures V_Z at some narrow pulse width and does not use a correction factor.

Using the curves of Figure 5, V_Z can be calculated at any pulse width based upon the value of V_Z at TE which is represented by 1 on the normalized V_Z scale. If the 1N52xx diode is specified at 12 V ± 1.0% at 90 seconds which is at TE, V_Z at 100 ms using either of the Grayhill clips curves would be 0.984 of the V_Z value at TE or 1 using the ON Semiconductor fixture curve. If the negative TC diode is specified at 3.9 V ± 1.0% at TE (90 seconds), V_Z at 100 ms would be 1.011 of V_Z at TE (using ON Semiconductor fixture curve) when using the Grayhill Clips curves.

In using the curves of Figure 5 and 6, it should be kept in mind that V_Z can be different at TE for the three mountings because diode junction temperature can be different for each mounting at TE which is represented by 1 on the V_Z normalized scale. Therefore, when the correlation of V_Z between parties is attempted, they must use the same type of mounting or know what the delta V_Z is between the two mountings involved.

The Grayhill clips curves in Figure 6 are normalized to the ON Semiconductor fixture at TE as in Figure 5. Figures 1 through 4 are normalized to V_Z at TE for each diode and would be used as Figures 5 and 6.

Measurement accuracy can be affected by test equipment, power dissipation of the D.U.T., ambient temperature and accuracy of the voltage divider if used on the input of the first op-amp (U4). The curves of Figures 1 through 6 are for an ambient temperature of 25°C, at other ambients, θV_Z has to be considered and is shown on the data sheet for the 1N52xx series of diodes. θV_Z is expressed in mV/°C and is multiplied by the difference in T_A from the 25°C value and either subtracted or added to the calculated V_Z depending upon whether the diode has a negative or positive TC.

General Discussion

The TC of zener diodes can be either negative or positive, depending upon die processing. Generally, devices with a breakdown voltage greater than about 5 V have a positive TC and diodes under about 5 V have a negative TC.

Conclusion

Curves showing V_Z versus I_{ZT} pulse width can be used to calculate V_Z at a pulse width other than the one used to specify V_Z. A test circuit and method is presented to obtain V_Z with a single pulse of test current to generate V_Z curves of interest.

References

1. Al Pshaenich, “Double Pulsing S/H Increases System Accuracy”; Electronics, June 16, 1983.
2. ON Semiconductor Zener Diode Manual, Series A, 1980.

FIGURES 1 thru 8 — Conditions: Single Pulse, T_A = 25°C, V_Z I_{ZT} = 1/4 P_D (Max) Each device normalized to V_Z at TE.

AXIAL LEAD PACKAGES: MOUNTING STANDARD GRAYHILL CLIPS

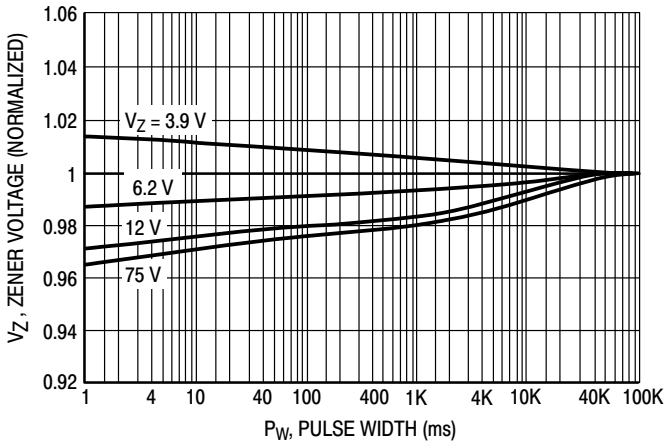


Figure 1. DO-35 (Glass) 500 mW Device

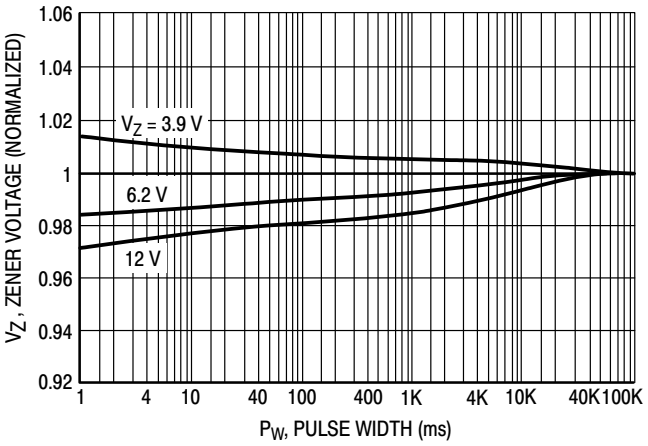


Figure 2. DO-41 (Glass) 1 Watt Device

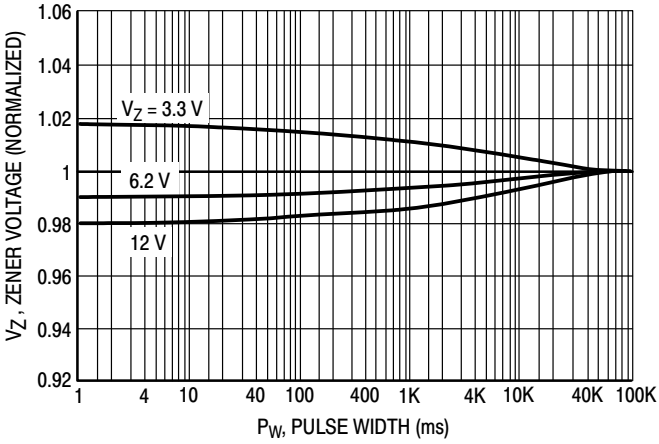


Figure 3. DO-41 (Plastic) 1.5 Watt Device

THREE MOUNTING METHODS: DO-35 AND DO-41

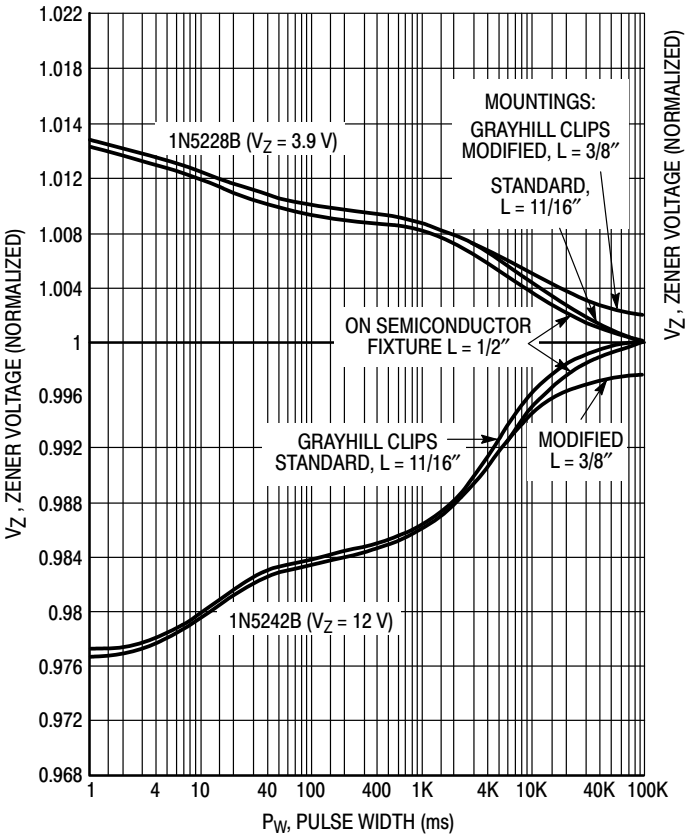


Figure 4. DO-35 (Glass) 500 mW Device

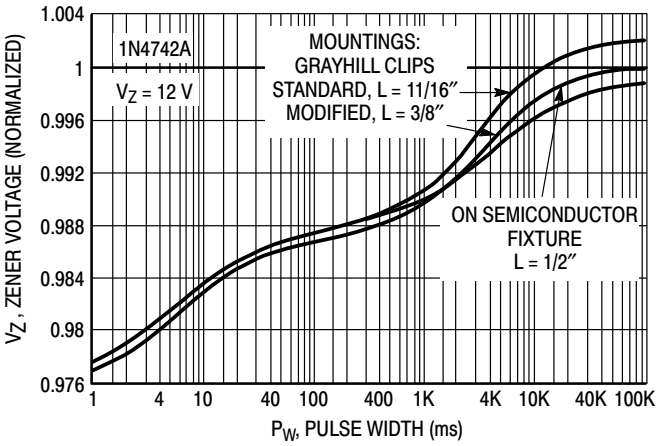


Figure 5. DO-41 (Glass) 1 Watt Device

MOUNTING FIXTURE

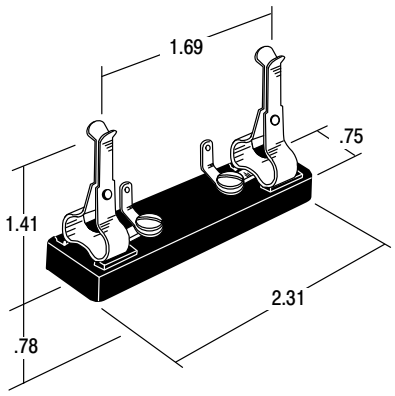


Figure 6. Standard Grayhill Clips

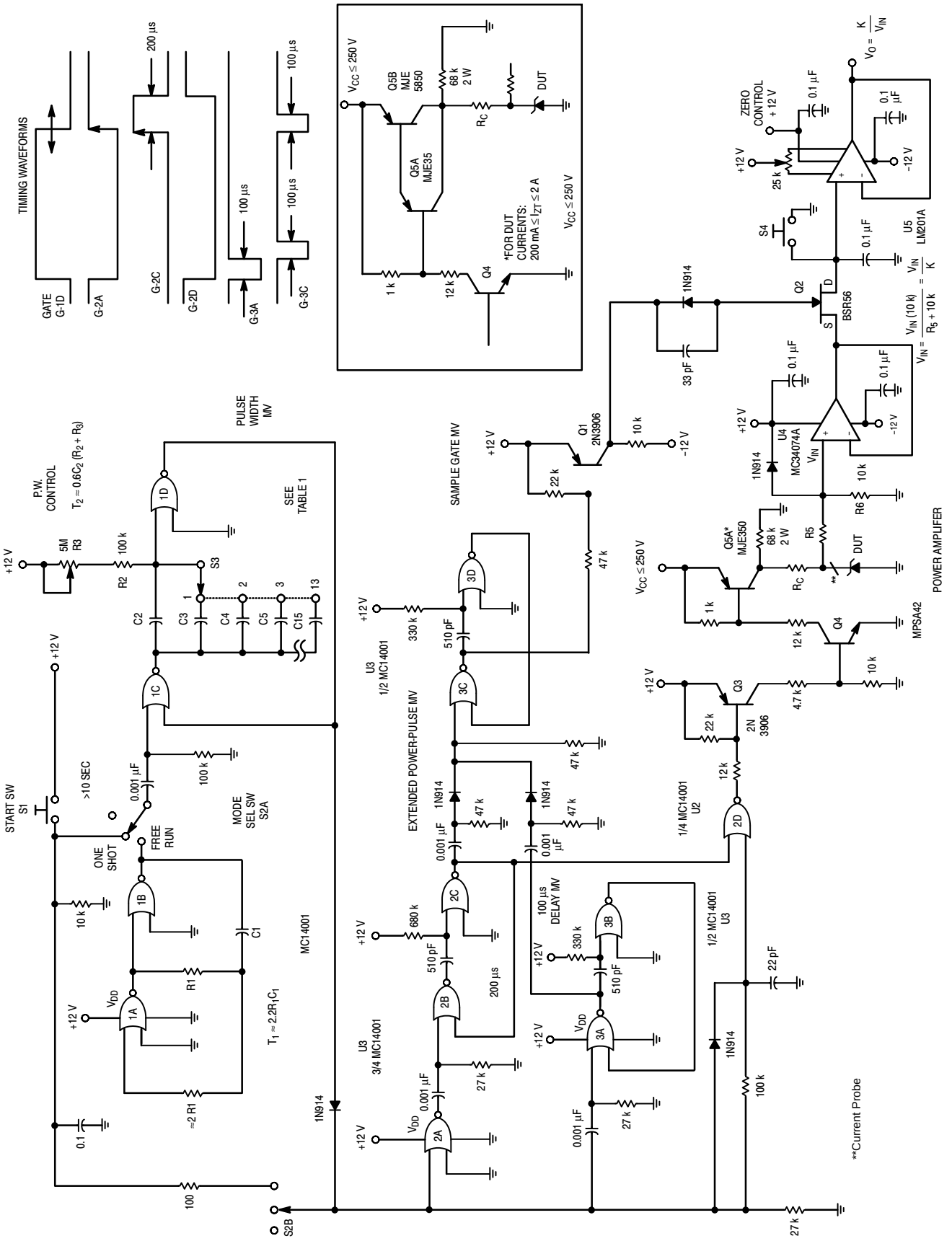


Figure 8. Zener Voltage Double Pulsing S/H Test Circuit

APPENDIX A
Recommended Incoming Inspection Procedures
Zener Voltage Testing
Pulsed versus Thermal Equilibrium

This section is primarily for use of incoming inspection groups. The subject covered is the measurement of zener voltage (V_Z) and the inherent difficulty of establishing correlation between supplier and buyer when using pulsed test techniques. This difficulty, in part, is due to the interpretation of the data taken from the variety of available testers and in some cases even from the same model types. It is therefore, our intent to define and reestablish a standardized method of measurement to achieve correlation no matter what test techniques are being used. This standardization will guarantee your acceptance of good product while maintaining reliable correlation.

DEFINITION OF TERMS

Temperature Coefficient (TC):

The temperature stability of zener voltages is sometimes expressed by means of the temperature coefficient (TC). This parameter is usually defined as the percent voltage change across the device per degree centigrade, or as a specific voltage change per degree centigrade. Temperature changes during test are due to the self heating effects caused by the dissipation of power in the zener junction. The V_Z will change due to this temperature change and will exhibit a positive or negative TC, depending on the zener voltage. Generally, devices with a zener voltage below five volts will have a negative TC and devices above five volts will exhibit a positive TC.

Thermal Equilibrium (TE)

Thermal equilibrium (TE) is reached when the diode junction temperature has stabilized and no further change will occur. In thermal equilibrium, the heat generated at the junction is removed as rapidly as it is created, hence, no further temperature changes.

MEASURING ZENER VOLTAGE

The zener voltage, being a temperature dependent parameter, needs to be controlled for valid V_Z correlation. Therefore, so that a common base of comparison can be established, a reliable measure of V_Z can only occur when all possible variables are held constant. This common base is achieved when the device under test has had sufficient time to reach thermal equilibrium (heatsinking is required to stabilize the lead or case temperature to a specified value for stable junction temperatures). The device should also be powered from a constant current source to limit changes of power dissipated and impedance.

All of the above leads us to an understanding of why various pulse testers will give differing V_Z readings; these differences are, in part, due to the time duration of test (pulse width), duty cycle when data logging, contact resistance, tolerance, temperature, etc. To resolve all of this, one only needs a reference standard to compare their pulsed results against and then adjust their limits to reflect those differences. It should be noted that in a large percentage of applications the zener diode is used in thermal equilibrium.

ON Semiconductor guarantees all of it's axial leaded zener products (unless otherwise specified) to be within specification ninety (90) seconds after the application of power while holding the lead temperatures at $30 \pm 1°C$, 3/8 of an inch from the device body, any fixture that will meet that criteria will correlate. 30°C was selected over the normally specified 25°C because of its ease of maintenance (no environmental chambers required) in a normal room ambient. A few degrees variation should have negligible effect in most cases. Hence, a moderate to large heatsink in most room ambients should suffice.

Also, it is advisable to limit extraneous air movements across the device under test as this could change thermal equilibrium enough to affect correlation.

SETTING PULSED TESTER LIMITS

Pulsed test techniques do not allow a sufficient time for zener junctions to reach TE. Hence, the limits need to be set at different values to reflect the V_Z at lower junction temperatures. Since there are many varieties of test systems and possible heatsinks, the way to establish these limits is to actually measure both TE and pulsed V_Z on a serialized sample for correlation.

The following examples show typical delta changes in pulsed versus TE readings. The actual values you use for pulsed conditions will depend on your tester. Note, that there are examples for both positive and negative temperature coefficients. When setting the computer limits for a positive TC device, the largest difference is subtracted from the upper limit and the smallest difference is subtracted from the lower limit. In the negative coefficient example the largest change is added to the lower limit and the smallest change is added to the upper limit.

ON Semiconductor Zeners

- Thermal equilibrium specifications:
  V_Z at 10 mA, 9 V minimum, 11 V maximum:
  (Positive TC)

TE	Pulsed	Difference
9.53 V	9.45 V	−0.08 V
9.35 V	9.38 V	−0.07 V
9.46 V	9.83 V	−0.08 V
9.56 V	9.49 V	−0.07 V
9.50 V	9.40 V	−0.10 V

Computer test limits:

Set V_Z max. limit at 11 V − 0.10 V = 10.9 V

Set V_Z min. limit at 9 V − 0.07 V = 8.93 V

- Thermal equilibrium specifications:
  V_Z at 10 mA, 2.7 V minimum, 3.3 V maximum:
  (Negative TC)

TE	Pulsed	Difference
2.78 V	2.83 V	+0.05 V
2.84 V	2.91 V	+0.07 V
2.78 V	2.84 V	+0.05 V
2.86 V	2.93 V	+0.07 V
2.82 V	2.87 V	+0.05 V

Computer test limits:

Set V_Z min. limit at 2.7 V + 0.07 V = 2.77 V

Set V_Z max. limit at 3.3 V + 0.05 V = 3.35 V

PUBLICATION ORDERING INFORMATION

LITERATURE FULFILLMENT:
Literature Distribution Center for ON Semiconductor
19521 E. 32nd Pkwy, Aurora, Colorado 80011 USA
Phone: 303−675−2175 or 800−344−3860 Toll Free USA/Canada
Fax: 303−675−2176 or 800−344−3867 Toll Free USA/Canada
Email: orderlit@onsemi.com

N. American Technical Support: 800−282−9855 Toll Free USA/Canada
Europe, Middle East and Africa Technical Support:
Phone: 421 33 790 2910
Japan Customer Focus Center
Phone: 81−3−5817−1050

ON Semiconductor Website: **www.onsemi.com**

Order Literature: http://www.onsemi.com/orderlit

For additional information, please contact your loca
Sales Representative