

TVS/Zener Theory and Design Considerations

Handbook

HBD854/D
Rev. 0, Jun-2005

© SCILLC, 2005
Previous Edition © 2001 as Excerpted from DL150/D
“All Rights Reserved”



ON Semiconductor™

Technical Information, Application Notes and Articles

Zener Diode Theory 3
Zener Diode Fabrication Techniques 8
Reliability 12
Zener Diode Characteristics 18
Temperature Compensated Zeners 30
Basic Voltage Regulation Using Zener Diodes 34
Zener Protective Circuits and Techniques:
 Basic Design Considerations 44
Zener Voltage Sensing Circuits and Applications ... 54
Miscellaneous Applications of
 Zener Type Devices 61
Transient Voltage Suppression 63
AN784 82
AN843 84
Design Considerations and Performance of
 Temperature Compensated Zener Diodes 97
MOSORBs 102
AR450 106
Measurement of Zener Voltage to Thermal
 Equilibrium with Pulsed Test Current 119

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. 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.

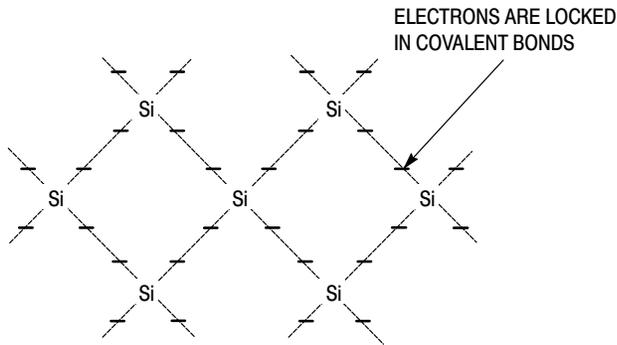
N-type conductivity in a semiconductor is much like the conductivity due to the drift of free electrons in a metal. In pure silicon at room temperature there are too few free electrons to conduct current. However, there are ways of introducing free electrons into the crystal lattice as we shall

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.

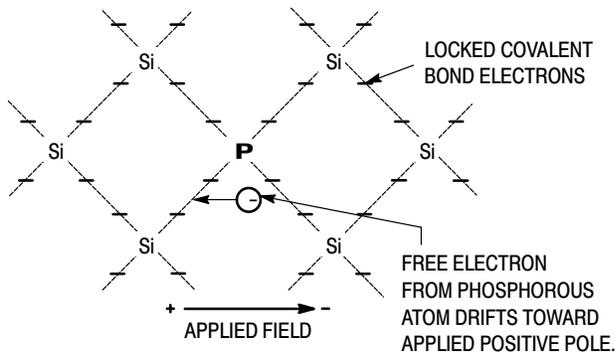
In P-type conductivity, the charges that carry electric current across the crystal act as if they were positive charges. We know that electricity is always carried by drifting electrons in any material, and that there are no mobile positively charged carriers in a solid. Positive charge carriers can 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.

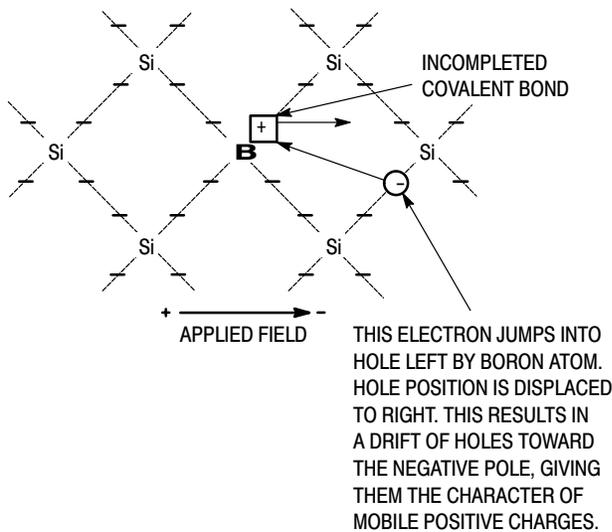
When semiconductor regions of N- and P-type conductivities are formed in a semiconductor crystal adjacent to each other, this structure is called a PN junction. Such a junction is responsible for the action of both zener diodes and rectifier devices, and will be discussed in the next section.



(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.

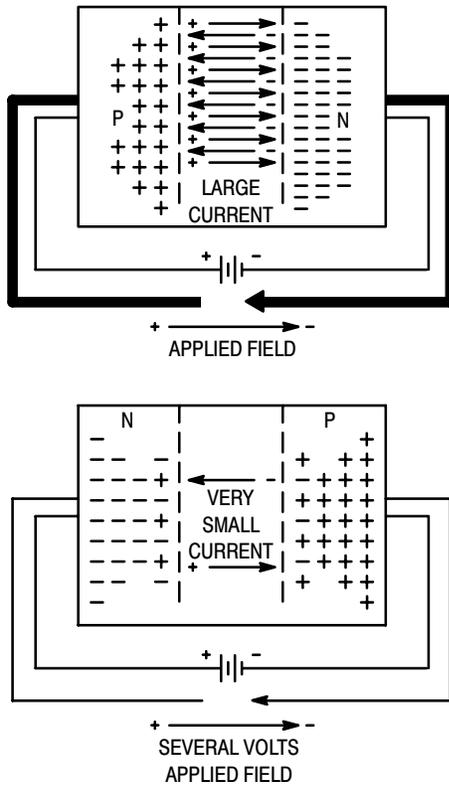
When a PN junction is reverse biased, the P-type side is made more negative than the N-type side. (See Figure 2b.) At voltages below the breakdown of the junction, there is very little current flow across the junction interface. At first thought one would expect no reverse current under reverse bias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

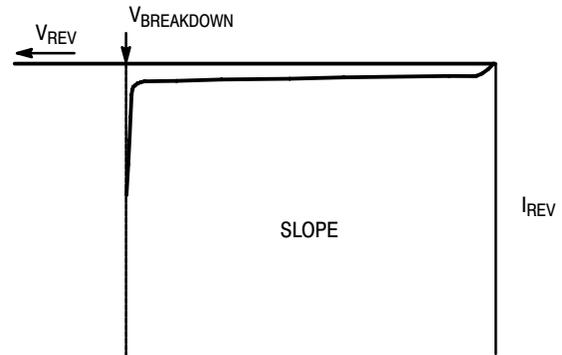


Figure 3. Reverse Characteristic Extended to Show Breakdown Effect

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 $3 \times 10^5 \text{V/cm}$. A field of such high magnitude exerts a large force on the

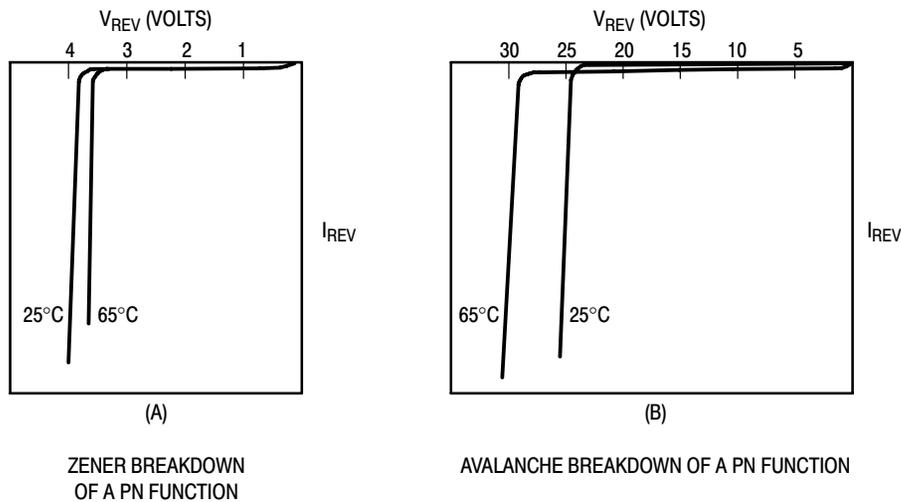


Figure 4. Typical Breakdown Diode Characteristics. Note Effects of Temperature for Each Mechanism

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 $3 \times 10^5 \text{V/cm}$. 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 $3 \times 10^5 \text{V/cm}$. 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 ($3 \times 10^5 \text{V/cm}$). 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

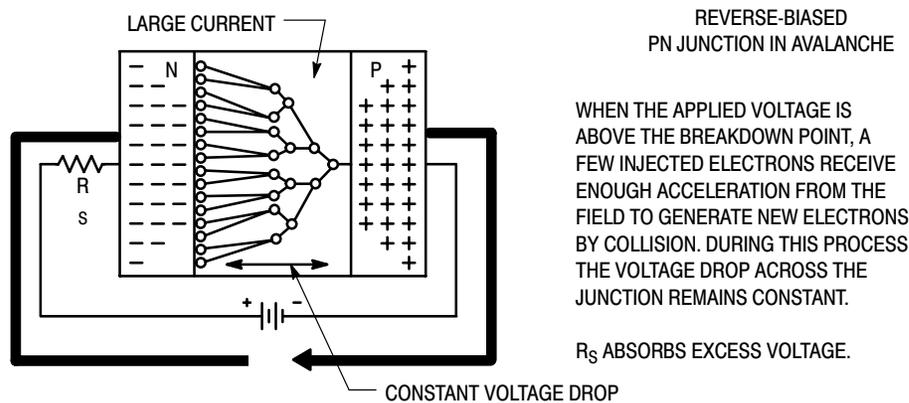


Figure 5. PN Junction in Avalanche Breakdown

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.

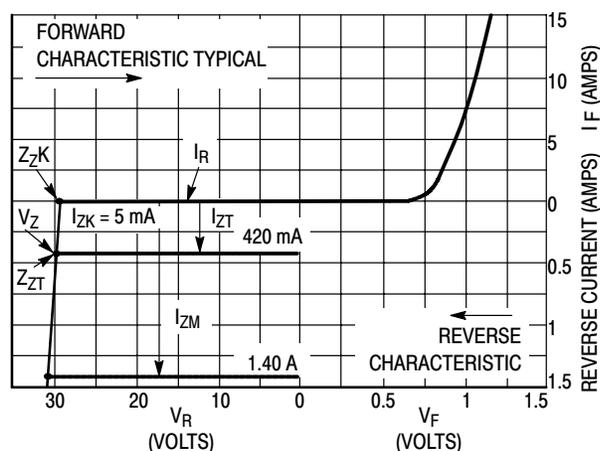


Figure 6. Zener Diode Characteristics

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 chapter four. In chapter four 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 chapter 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.

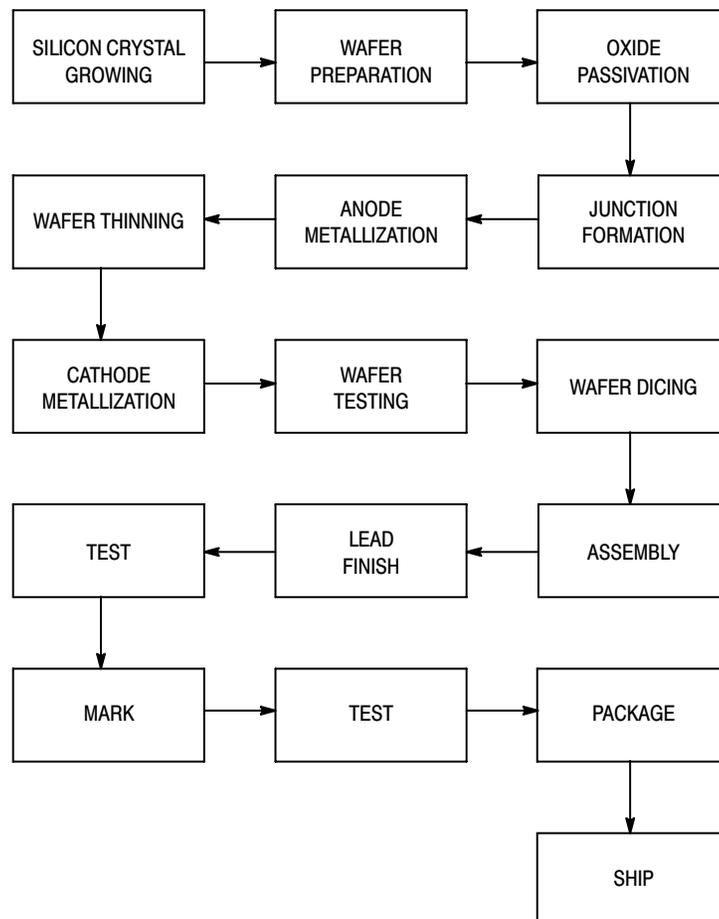


Figure 1. General Flow of the Zener Diode Process

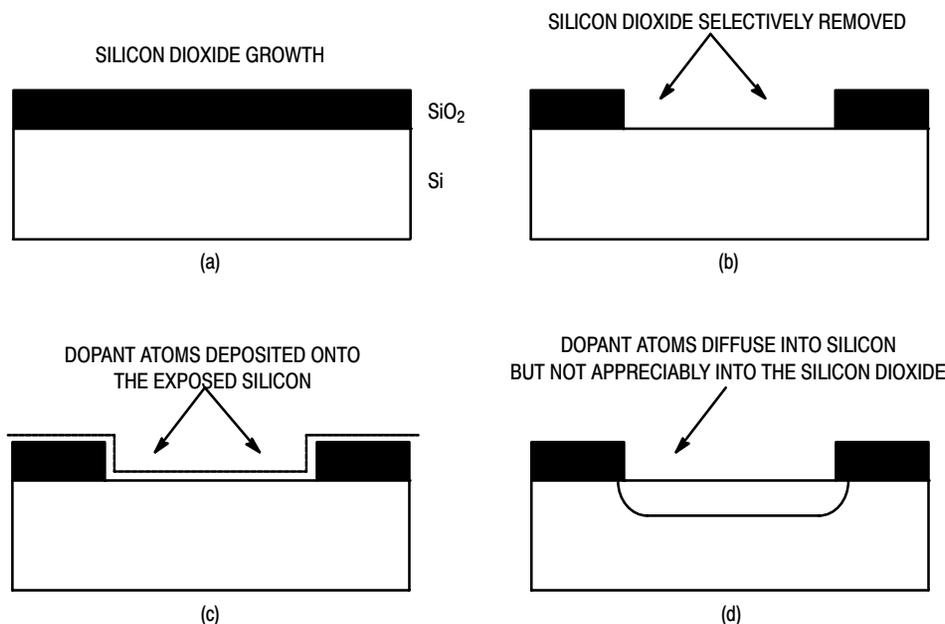


Figure 2. Basic Fabrication Steps in the Silicon Planar Process: a) oxide formation, b) selective oxide removal, c) deposition of dopant atoms, d) junction formation by diffusion of dopant atoms.

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 30, 40 and MOSORB

The plastic packages (Surmetic 30, 40 and MOSORBs) are 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.

The Surmetic 30s, 40s and MOSORBs are basically assembled in the same manner; the only difference being the MOSORBs are soldered together using a solder disc between the lead and die whereas the Surmetic 30s and Surmetic 40s utilize pre-soldered leads.

Assembly is started on the Surmetic 30 and 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. Since the MOSORBs do not use pre-soldered leads, the leads are put into the assembly boat,

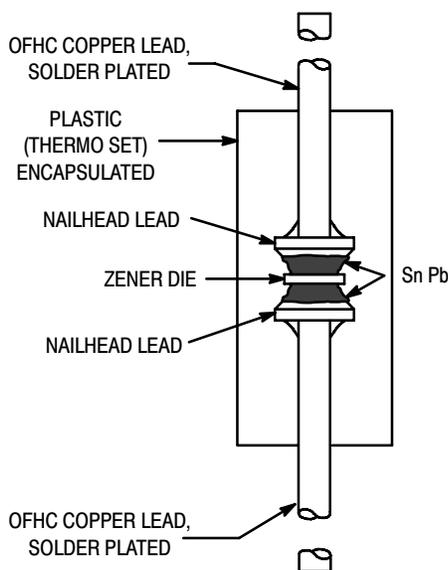


Figure 3. Double-Slug Plastic Zener Construction

a solder disc is placed into each cavity and then a die is put in on top. A solder disc is put in on top of the die. Another assembly boat containing only leads is mated to the boat containing the leads, die, and two solder discs. The boats are passed through the assembly furnace; this operation requires only one pass through the furnace.

After assembly, the leads on the Surmetic 30s, 40s and MOSORBs 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 pre-tinned 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 pre-tinned 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 tin which has wetted to each slug 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.

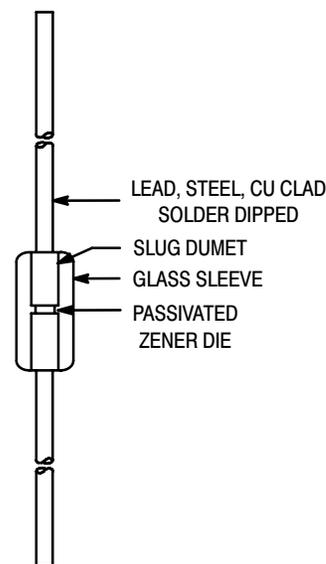


Figure 4. Double Slug Glass Zener Construction

ZENER DIODE TEST, MARK AND PACKAGING

Double Slug, Surmetic 30, 40 and MOSORB

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.

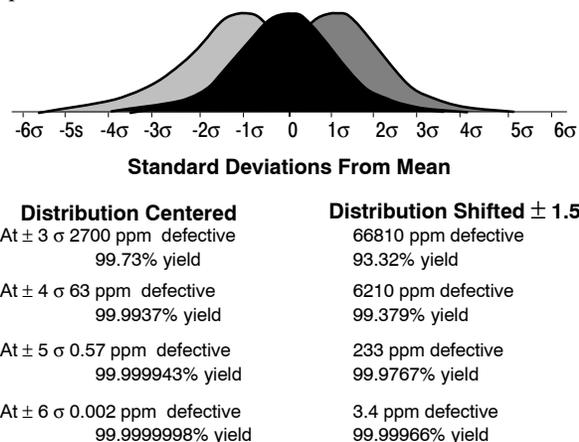


Figure 1. AOQL and Yield from a Normal Distribution of Product With 6σ Capability

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.

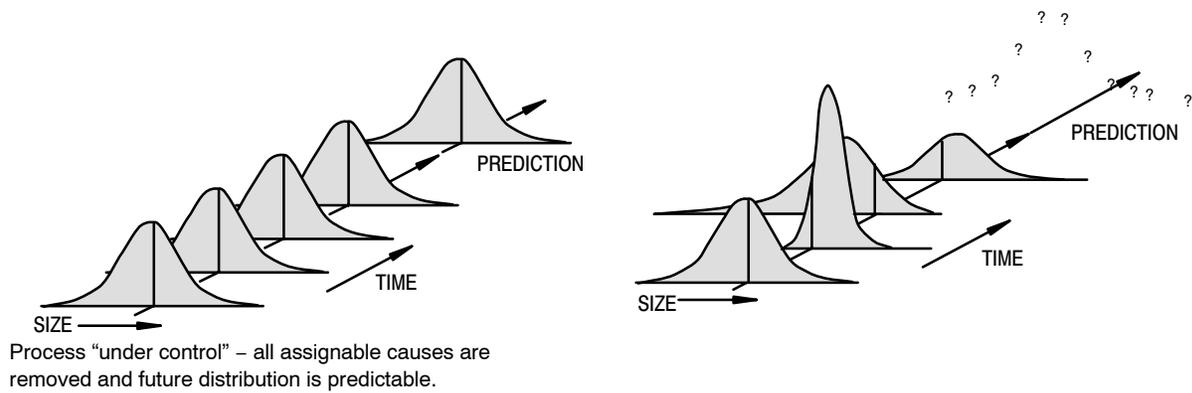


Figure 2. Impact of Assignable Causes on Process Predictable

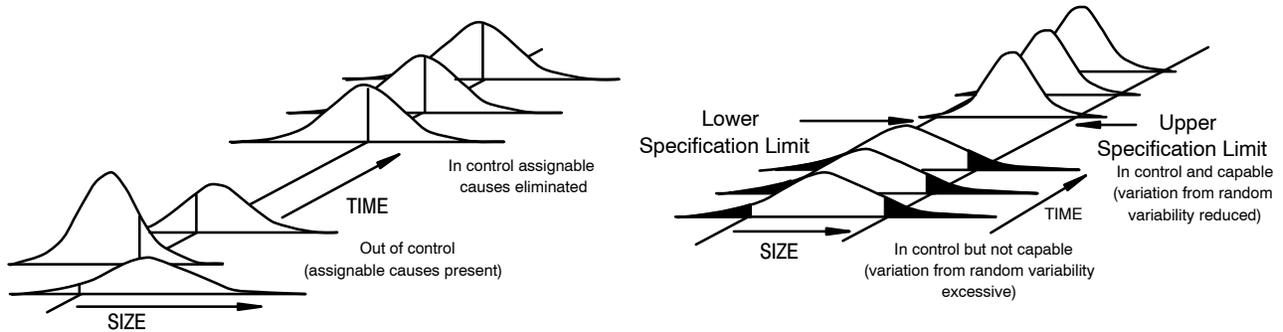


Figure 3. Difference Between Process Control and Process Capability

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 $Cp = (\text{specification width}) / 6\sigma$. Cpk is the absolute value of the closest specification value to the mean, minus the mean, divided by half the process width or $Cpk = |\text{closest specification} - \bar{X}| / 3\sigma$.

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..

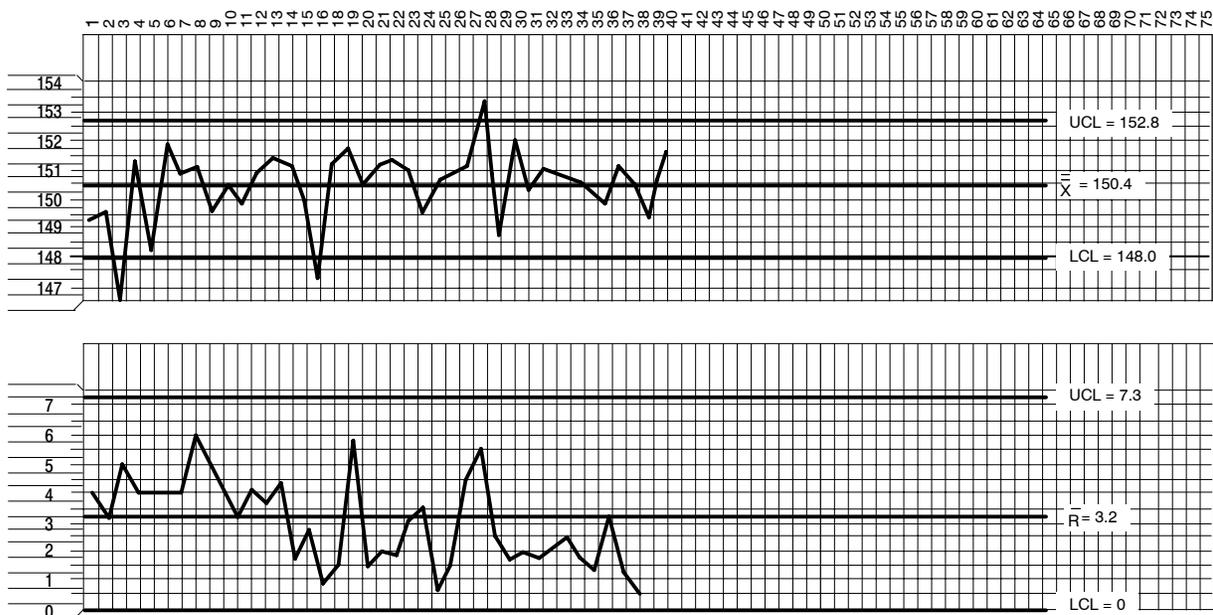


Figure 4. Example of Process Control Chart Showing Oven Temperature Data

Once the plan is established, data collection may begin. The data collected will generate \bar{X} and R values that are plotted with respect to time. \bar{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 \bar{X} and R values have been generated, the average of these values is computed as follows:

$$\bar{\bar{X}} = (\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \dots) / K$$

$$\bar{R} = (R_1 + R_2 + R_3 + \dots) / K$$

where K = the number of subgroups measured.

The values of $\bar{\bar{X}}$ and \bar{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 \bar{X} and R values with respect to time and concerning reference to upper and lower control limit values. Control limits are computed as follows:

$$R \text{ upper control limit} = UCL_R = D_4 \bar{R}$$

$$R \text{ lower control limit} = LCL_R = D_3 \bar{R}$$

$$\bar{X} \text{ upper control limit} = UCL_X = \bar{\bar{X}} + A_2 \bar{R}$$

$$\bar{X} \text{ lower control limit} = LCL_X = \bar{\bar{X}} - A_2 \bar{R}$$

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

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.

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:

$$\sigma_{\text{tot}} = \sqrt{\sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2}$$

$$\sigma_{\text{tot}} = \sqrt{5^2 + 3^2 + 2^2 + 1^2 + (0.4)^2} = 6.3$$

Now if only D is identified and eliminated then;

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

This results in less than 2% total variability improvement.

If B, C and D were eliminated, then;

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

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

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

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**.

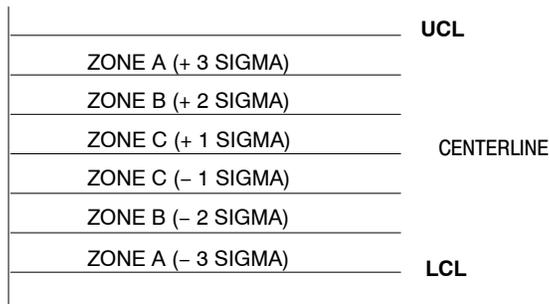


Figure 5a. Control Chart Zones

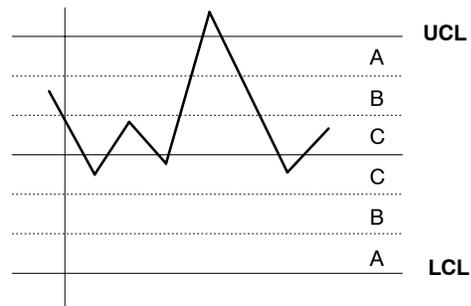


Figure 5b. One Point Outside Control Limit Indicating Excessive Variability

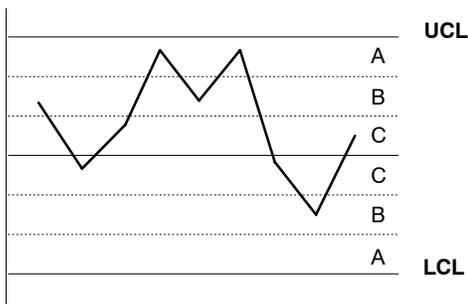


Figure 5c. Two Out of Three Points in Zone A or Beyond Indicating Excessive Variability

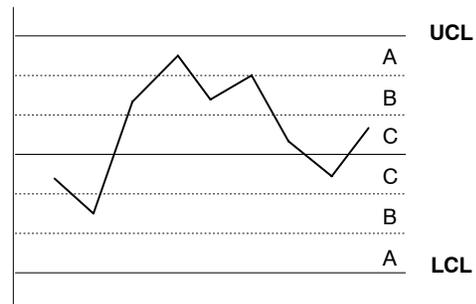


Figure 5d. Four Out of Five Points in Zone B or Beyond Indicating Excessive Variability

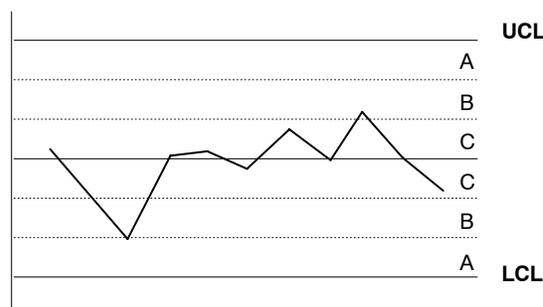


Figure 5e. Seven Out of Eight Points in Zone C or Beyond Indicating Excessive Variability

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: $T_A = 121^\circ\text{C}$, $\text{rh} = 100\%$, $p = 1$ 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: $T_A = 85^\circ\text{C}$ to 95°C , $\text{rh} = 85\%$ 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: $T_A = 85^\circ\text{C}$ 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: $T_A = 70^\circ\text{C}$ 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: $T_A = 25^\circ\text{C}$, $P_d =$ Data Book maximum rating, $T_{\text{on}} = T_{\text{off}} = \Delta$ 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: $T_A = -10^\circ\text{C}$ to 65°C , $\text{rh} = 80\%$ 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: $T_A = 25^\circ\text{C}$, $P_D =$ 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: $T_A = -65^\circ\text{C}$ 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: $T_A = 0^\circ\text{C}$ 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 chapter 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 $I_F = I_R e^{qV_F/KT}$ 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.

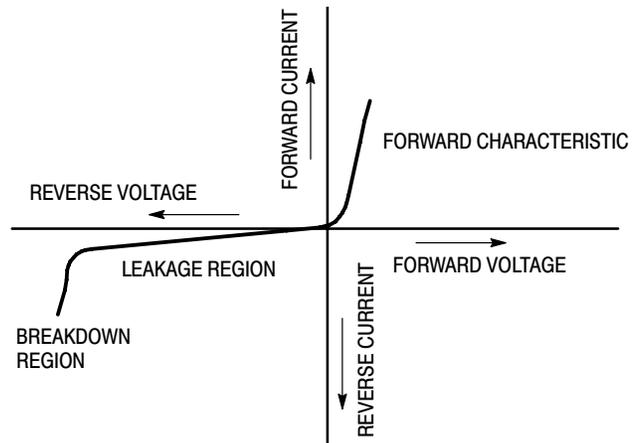


Figure 1. Typical Zener Diode DC V-I Characteristics (Not to Scale)

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.

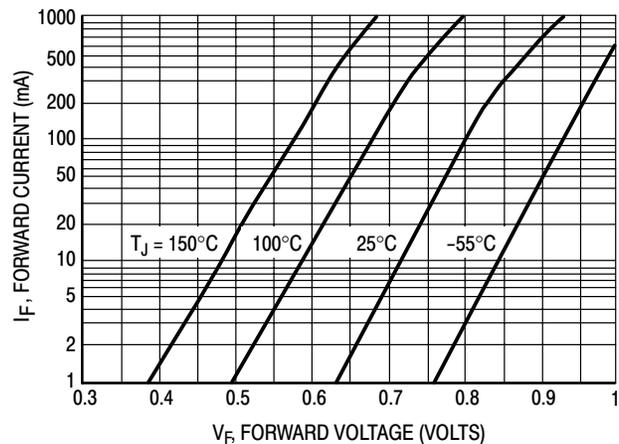


Figure 2. Typical Forward Characteristics of Zener Diodes

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.

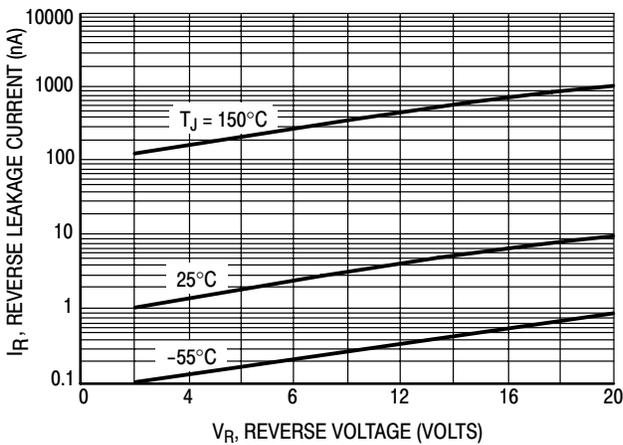


Figure 3. Typical Leakage Current versus Voltage

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.

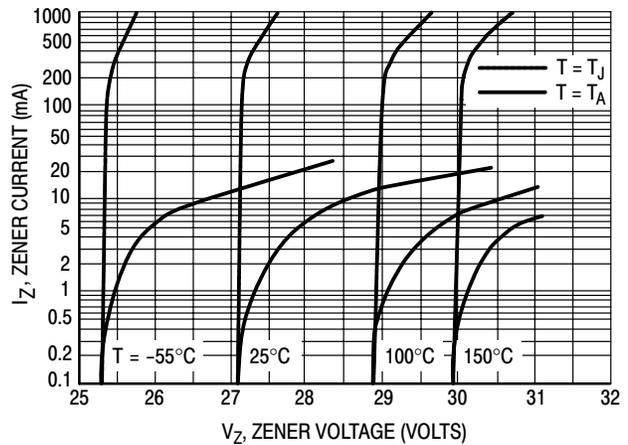


Figure 4. Typical Zener Characteristic Variation with Temperature

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.

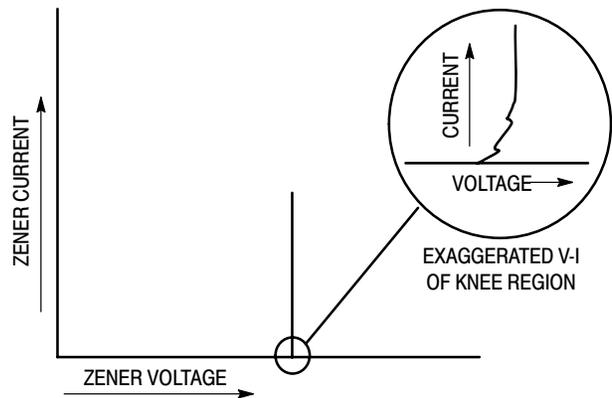


Figure 5. Exaggerated V-I Characteristics of the Knee Region

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.

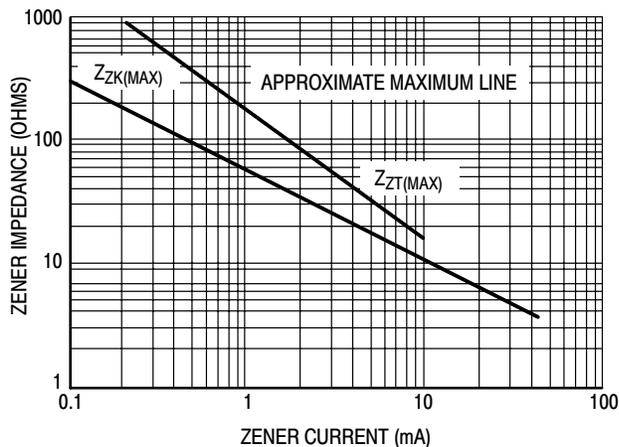


Figure 6. Zener Impedance versus Zener Current

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.

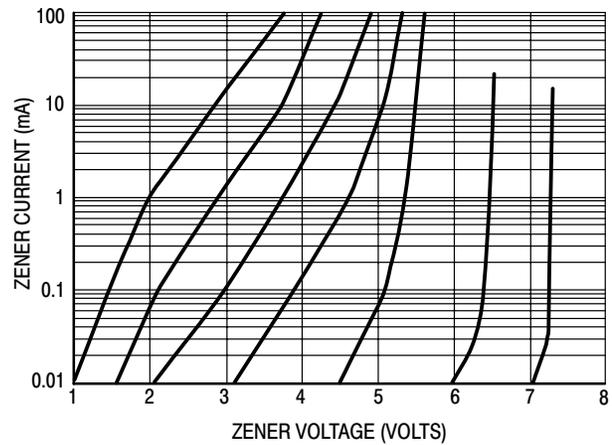


Figure 7. Zener Current versus Zener Voltage (Low Voltage Region)

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.

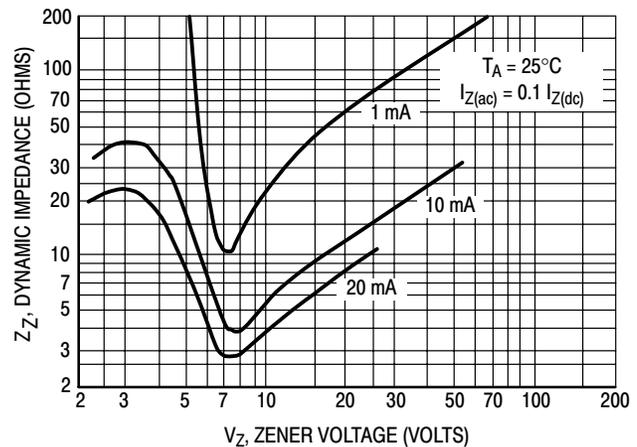
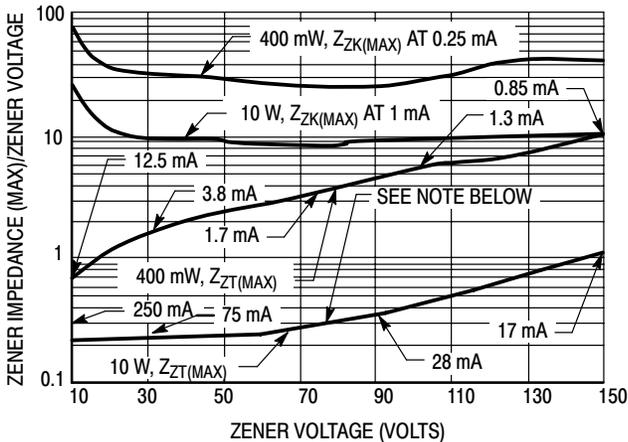


Figure 8. Dynamic Zener Impedance (Typical) versus Zener Voltage

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.

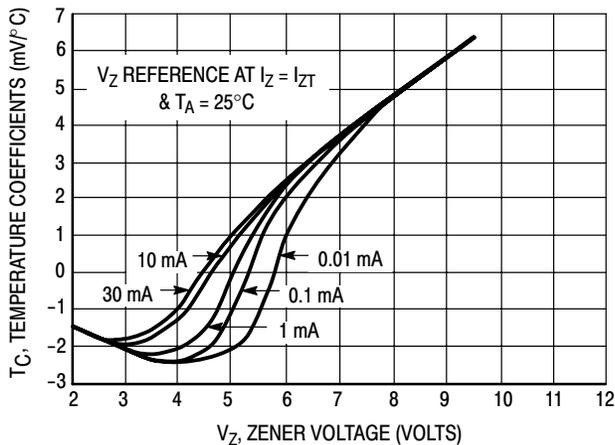


Figure 10. Temperature Coefficient versus Zener Voltage at 25°C Conditions Typical

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 $\%/^{\circ}\text{C}$) 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.

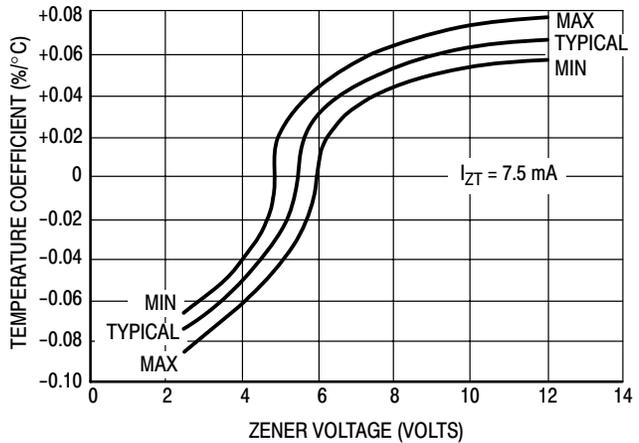


Figure 11. Temperature Coefficient Spread versus Zener Voltage

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. Information concerning the On Semiconductor temperature compensated or reference diodes is given in Chapter 4.

Typical temperature characteristics for a broad range of voltages 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).

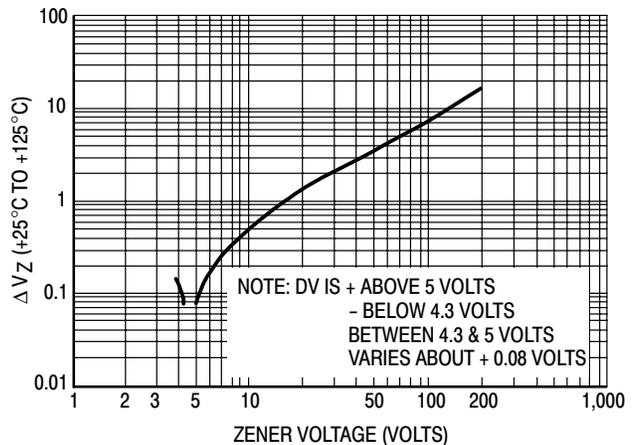


Figure 12. Typical Temperature Characteristics

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.

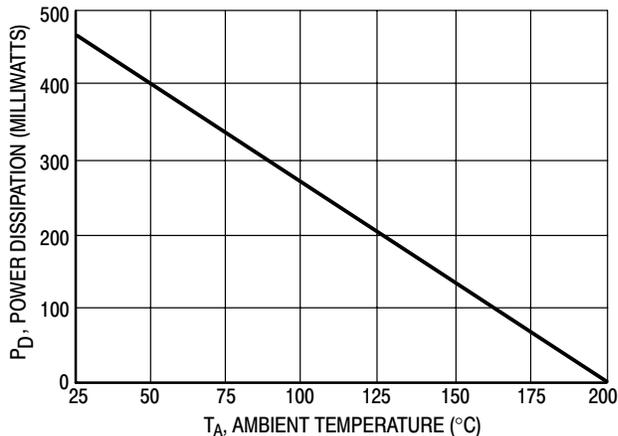


Figure 13. 400 mW Power Temperature Derating Curve

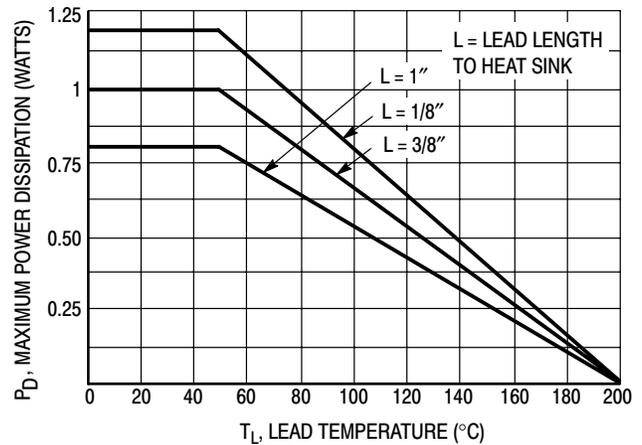


Figure 14. Power Temperature Derating Curve

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 pulse use are not so easily calculated. 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.

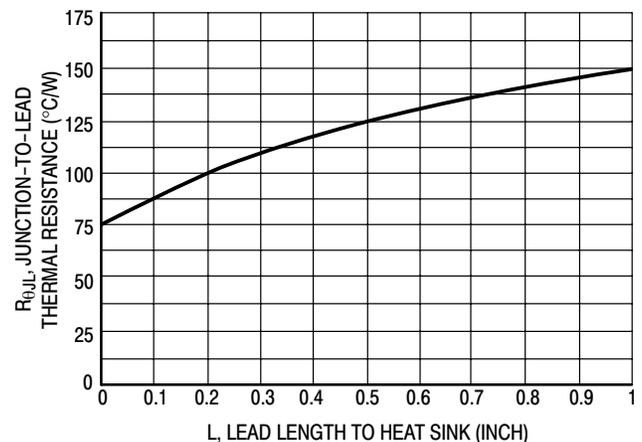


Figure 15. Typical 1N4728 Thermal Resistance versus Lead Length

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

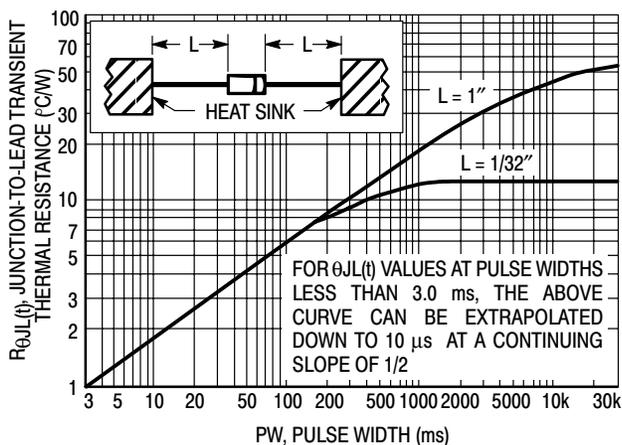


Figure 16. Typical Transient Thermal Resistance (For Axial Lead Zener)

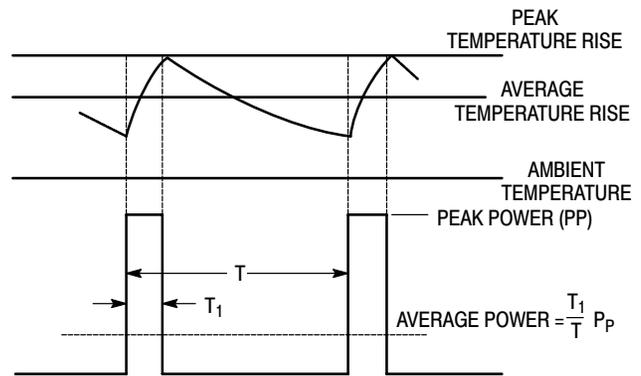


Figure 17. Relation of Junction Temperature to Power Pulses

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)

$$P_p = 5 \text{ watt (Lead length } 1/32\text{'')}$$

$$D = 0.1$$

$$T_1 = 10 \text{ ms}$$

$$T = 100 \text{ ms}$$

$$R_{\theta JA}(ss) = 12^\circ\text{C/W (for } 1/32\text{'' lead length)}$$

Then

$$R_{\theta T1} = 1.8^\circ\text{C/W}$$

$$R_{\theta T} = 5.8^\circ\text{C/W}$$

$$R_{\theta T1 + T} = 6^\circ\text{C/W}$$

And

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

$$\Delta T = 13^\circ\text{C}$$

Or at $T_A = 25^\circ$, $T_J = 38^\circ\text{C peak}$

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.

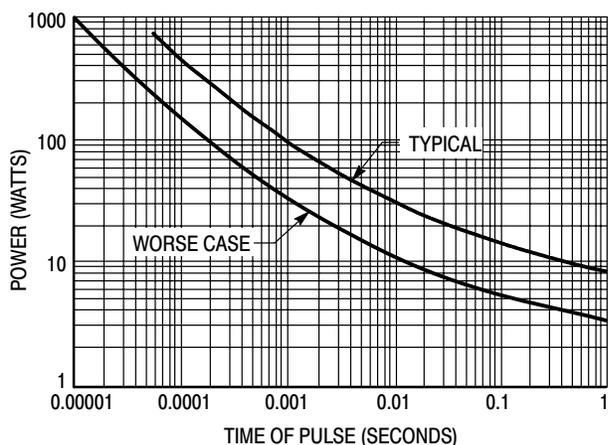


Figure 18. One Shot Power Failure Axial Lead Zener Diode

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.

Since information contained in this section may not be found on data sheets it is necessary for the designer to contact the factory when using a zener diode as a surge suppressor. Additional information on transient suppression application is presented elsewhere in this book.

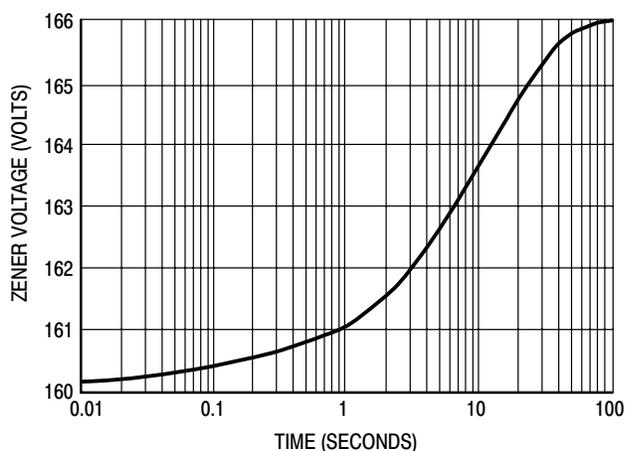


Figure 19. Zener Voltage (Typical) versus Time for Step Power Pulses (500 mW Lead Mounted Devices)

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:

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

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.

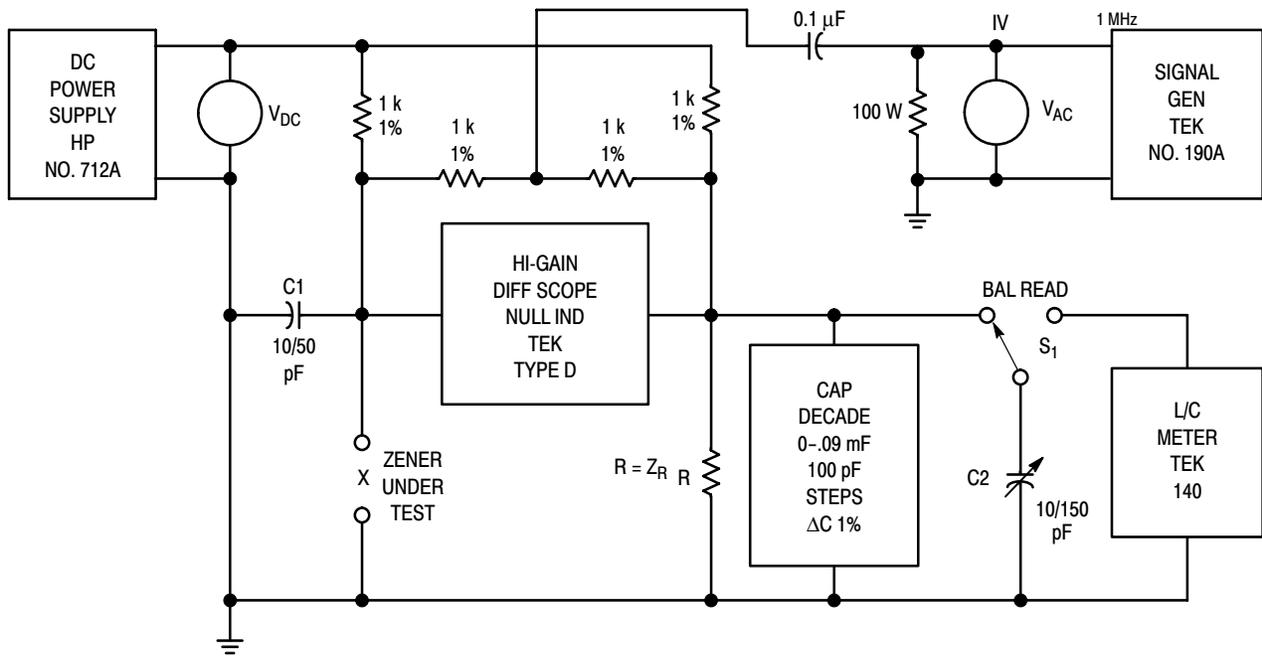


Figure 20. Capacitance Test Circuit

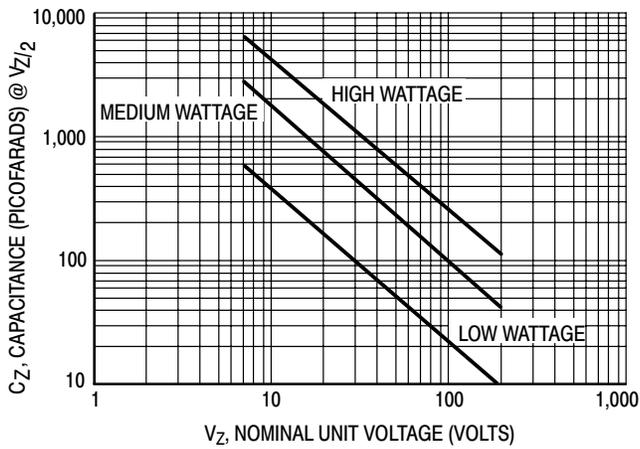


Figure 21. Capacitance versus Voltage

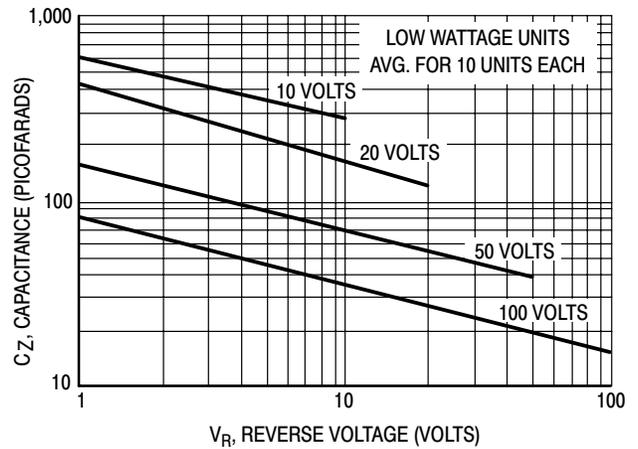


Figure 22. Capacitance versus Reverse Voltage

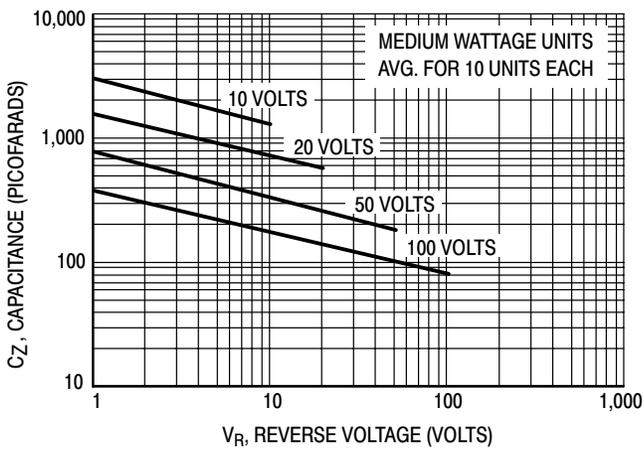


Figure 23. Capacitance versus Reverse Voltage

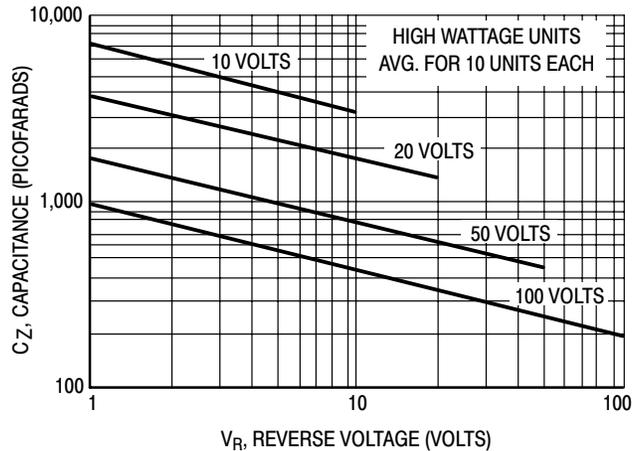


Figure 24. Capacitance versus Reverse Voltage

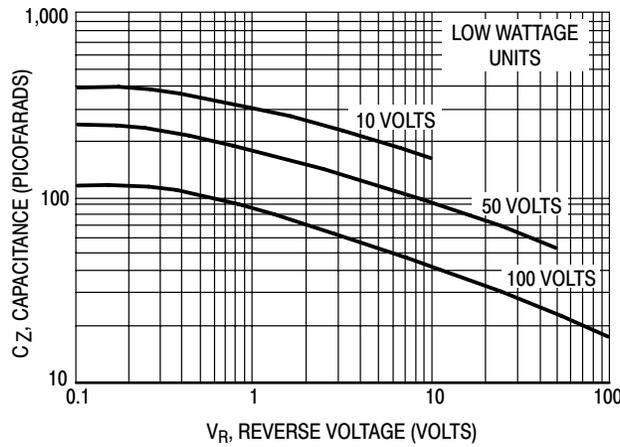


Figure 25. Flattening of Capacitance Curve at Low Voltages

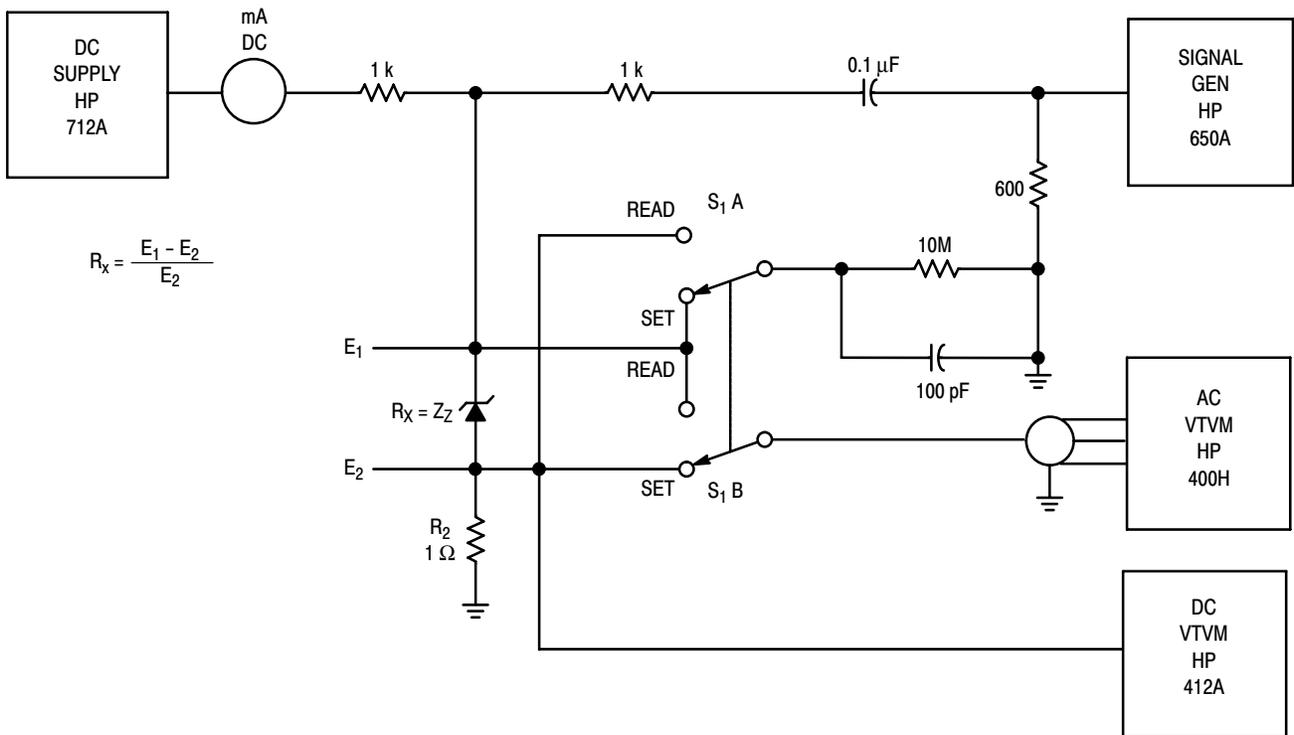


Figure 26. Impedance Test Circuit

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 R_2 .

- (3) Measure the voltage across the entire network by switching S_1 . 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 .

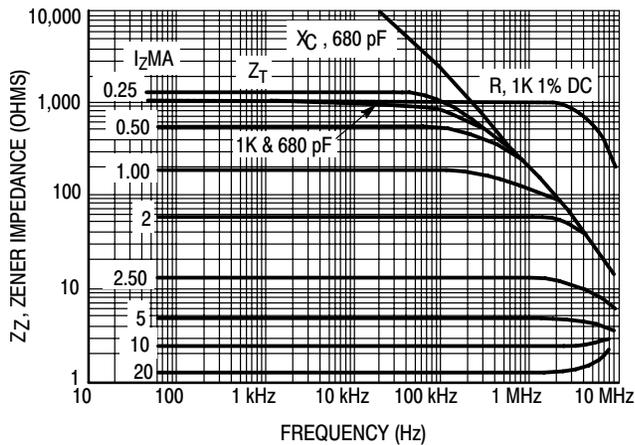


Figure 27. Zener Impedance versus Frequency

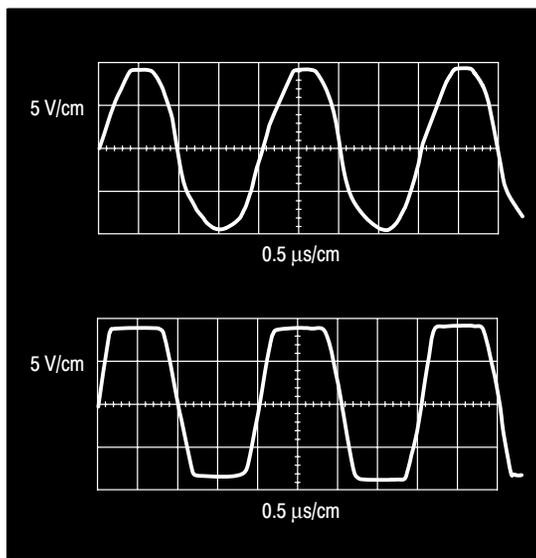
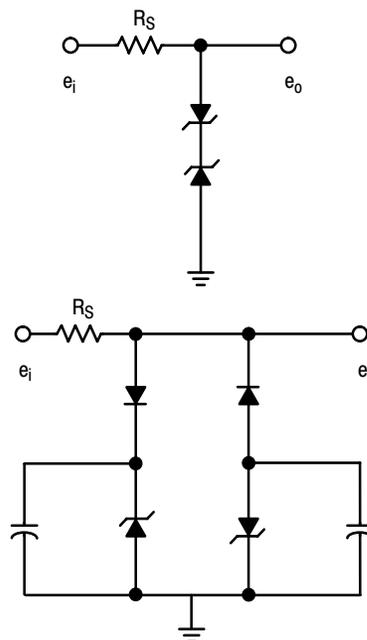


Figure 28. Symmetrical Peak Limiter

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.



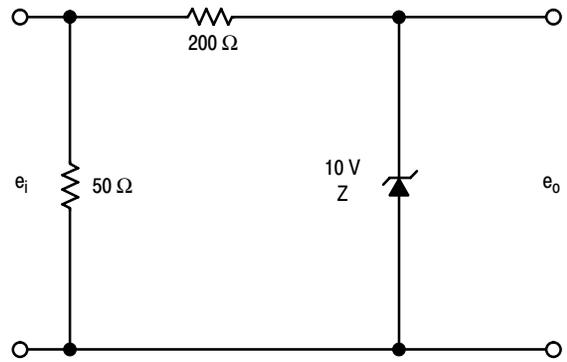
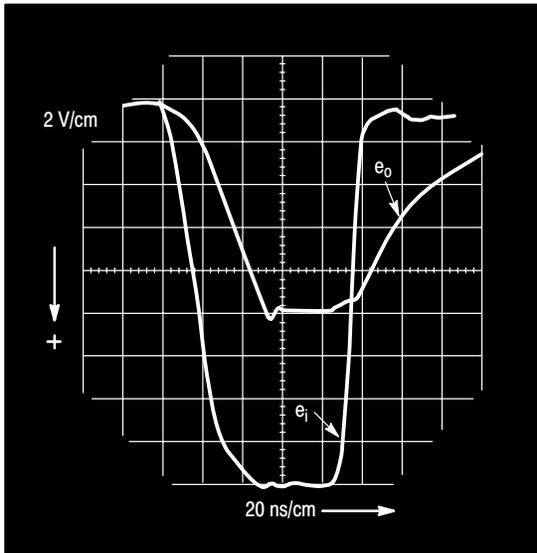


Figure 29. Shunt Clipper

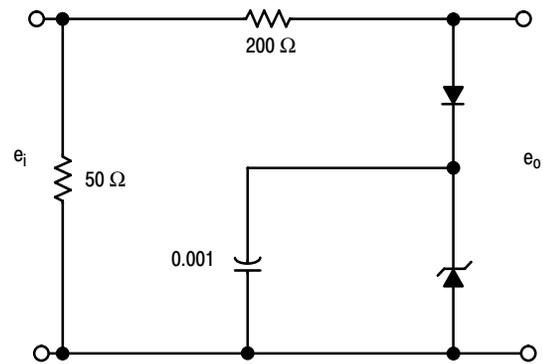
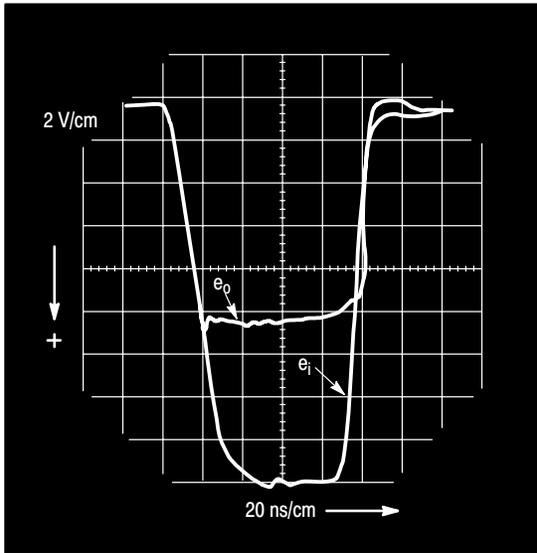


Figure 30. Shunt Clipper with Clamping Network

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.

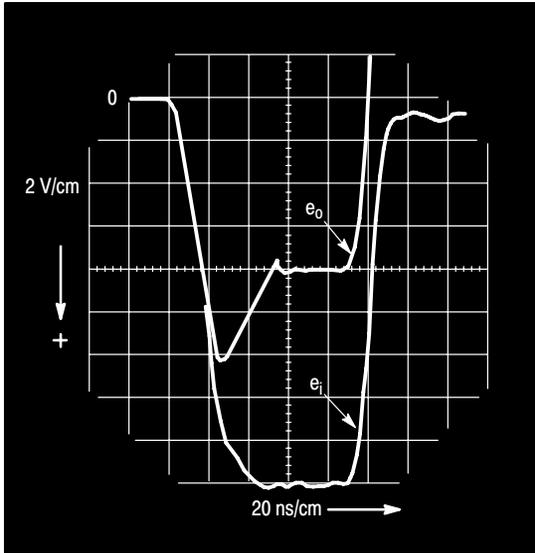


Figure 31. Basic Series Clipper

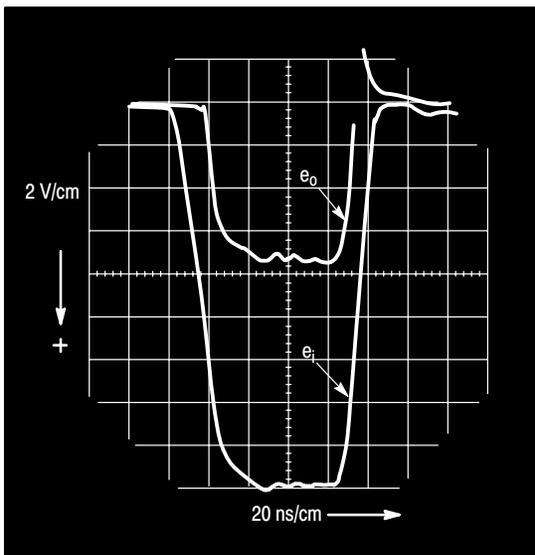
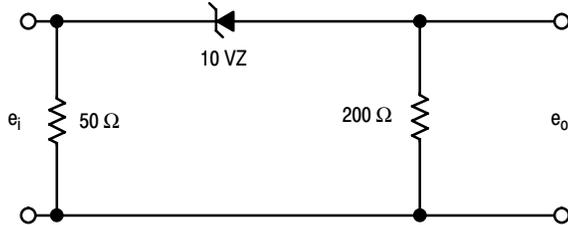
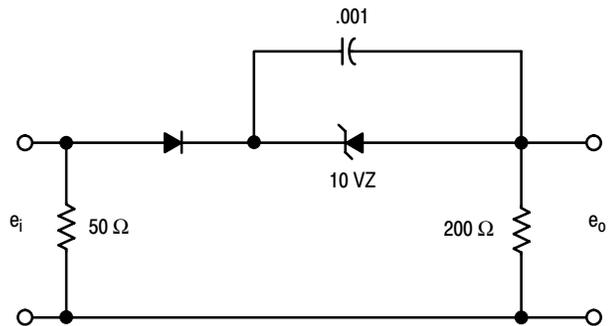


Figure 32. Series Clipper with Clamping Network



TEMPERATURE COMPENSATED ZENERS

INTRODUCTION

A device which provides reference voltages in a special manner is a reference diode.

As was discussed in the preceding chapters, the zener diode has the unique characteristic of exhibiting either a positive or a negative temperature coefficient, or both. By properly employing this phenomenon in conjunction with other semiconductor devices, it is possible to manufacture a zener reference element exhibiting a very low temperature coefficient when properly used. This type of low temperature coefficient device is referred to as a reference diode.

INTRODUCTION TO REFERENCE DIODES

The temperature characteristics of the zener diode are discussed in a previous chapter, where it was shown that change in zener voltage with temperature can be significant under severe ambient temperature changes (for example, a 100 V zener can change 12.5 volts from 0 to 125°C). The reference diode (often called the temperature compensated zener or the TC zener) is specially designed to minimize these specific temperature effects.

Design of temperature compensated zeners make possible devices with voltage changes as low as 5 mV from -55 to +100°C, consequently, the advantages of the temperature compensated zener are obvious. In critical applications, as a voltage reference in precision dc power supplies, in high stability oscillators, in digital voltmeters, in frequency meters, in analog-to-digital converters, or in other precision equipment, the temperature compensated zener is a necessity.

Conceivably temperature compensated devices can be designed for any voltage but present devices with optimum voltage temperature characteristics are limited to specific voltages. Each family of temperature compensated zeners is designed by careful selection of its integral parts with special attention to the use conditions (temperature range and current). A distinct operating current is associated with each device. Consequently, changes from the specified operating current can only degrade the voltage-temperature relationships. This will be discussed in more detail later.

The device "drift" or voltage-time stability is critical in some reference applications. Typically zeners and TC zeners offer stability of better than 500 parts per million per 1000 hours.

TEMPERATURE CHARACTERISTICS OF THE P-N JUNCTION AND COMPENSATION

The voltage of a forward biased P-N junction, at a specific current, will decrease with increasing temperature. Thus, a device so biased displays a negative temperature coefficient (Figure 1). A P-N junction in avalanche (above 5 volts breakdown) will display a positive temperature coefficient; that is, voltage will increase as temperature increases. Due to energy levels of a junction which breaks down below 5 volts, the temperature coefficient is negative.

It follows that various combinations of forward biased junctions and reverse biased junctions may be arranged to achieve temperature compensation. From Figure 2 it can be seen that if the absolute value of voltage change (ΔV) is the same for both the forward biased diode and the zener diode where the temperature has gone from 25°C to 100°C, then the total voltage across the combination will be the same at both temperatures since one ΔV is negative and the other positive. Furthermore, if the rate of increase (or decrease) is the same throughout the temperature change, voltage will remain constant. The non-linearity associated with the voltage temperature characteristics is a result of this rate of change not being a perfect match.

$$V_{REF} = V_Z + \Delta V_Z + V_D - \Delta V_D$$

THE METHODS OF TEMPERATURE COMPENSATION

The effect of temperature is shown in Figure 1. The forward characteristic does not vary significantly with reverse voltage breakdown (zener voltage) rating. A change in ambient temperature from 25° to 100°C produces a shift in the forward curve in the direction of lower voltage (a negative temperature coefficient — in this case about 150 mV change), while the same temperature change produces approximately 1.9 V increase in the zener voltage (a positive coefficient). By combining one or more silicon diodes biased in the forward direction with the P-N biased zener diode as shown in Figure 3, it is possible to compensate almost completely for the zener temperature coefficient. Obviously, with the example shown, 13 junctions would be needed. Usually reference diodes are low voltage devices, using zeners with 6 to 8 volts breakdown and one or two forward diodes.

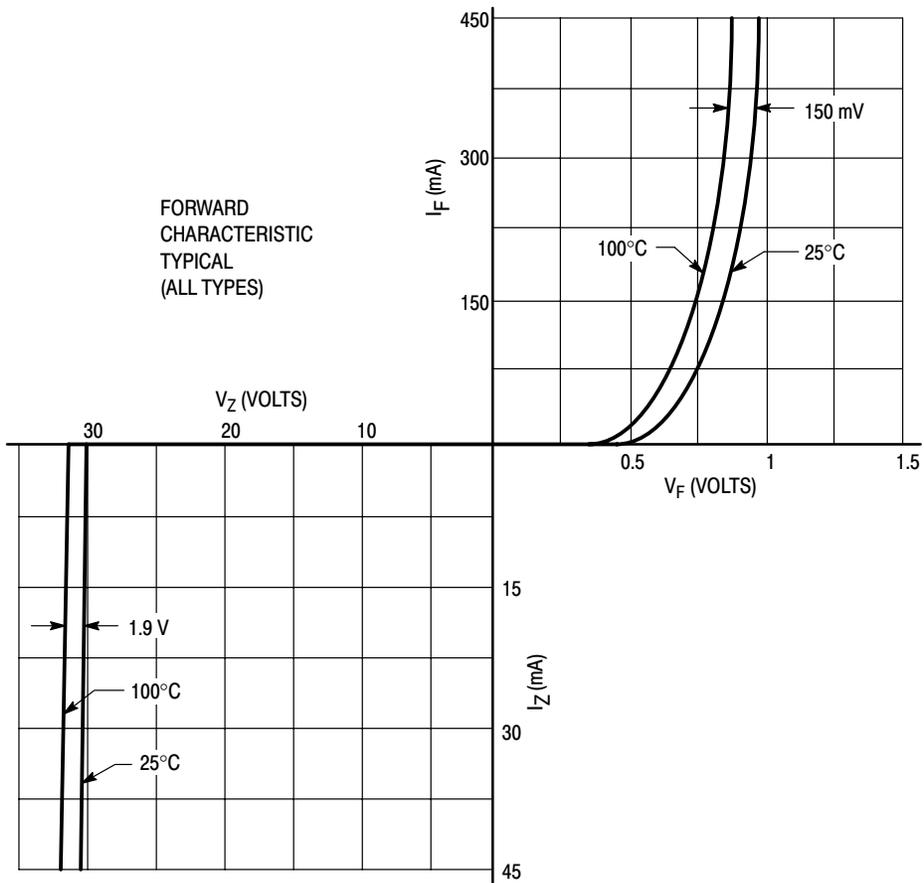


Figure 1. Effects of Temperature on Zener Diode Characteristics

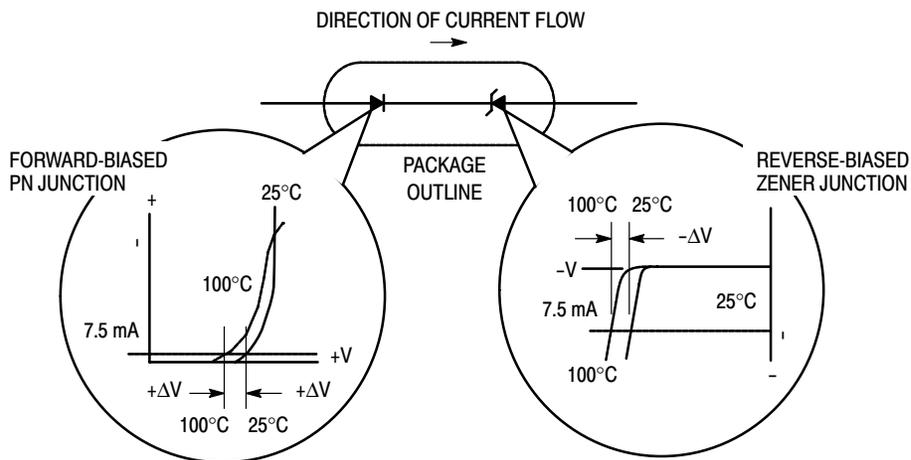


Figure 2. Principle of Temperature Compensation

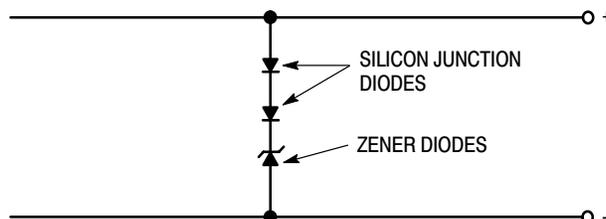


Figure 3. Zener Temperature Compensation with Silicon Forward Junctions

In ac regulator and clipper circuits where zener diodes are normally connected cathode to cathode, the forward biased diode during each half cycle can be chosen with the correct forward temperature coefficient (by stacking, etc.) to correctly compensate for the temperature coefficient of the reverse-biased zener diode. It is possible to compensate for voltage drift with temperature using this method to the extent that zener voltage stabilities on the order of 0.001%/°C are quite feasible.

This technique is sometimes employed where higher wattage devices are required or where the zener is compensated by the emitter base junction of a transistor stage. Consider the example of using discrete components, 1N4001 rectifier and ON Semiconductor 5 Watt zener, to obtain compensated voltage-temperature characteristics. Examination of the curve in Figure 4 indicates that a 10 volt zener diode exhibits a temperature coefficient of approximately +5.5 mV/°C. At a current level of 100 mA a temperature coefficient of approximately -2.0 mV/°C is characteristic of the 1N4001 rectifiers. A series connection of three silicon 1N4001 rectifiers produces a total temperature coefficient of approximately -6 mV/°C and a total forward drop of approximately 2.17 volts at 25°C. The combination of three silicon rectifiers and the 10 volt zener diode produces a device with a coefficient of approximately -0.5 mV/°C and a total breakdown voltage at 100 mA of approximately 12.2 volts. Calculation shows this to be a temperature stability of -0.004%/°C.

$$\left(\frac{-0.5 \text{ mV}/^\circ\text{C}}{12.2 \text{ V}} \right) \times 100$$

The temperature compensated zener employs the technique of specially selected dice. This provides optimum voltage temperature characteristics by close control of dice resistivities.

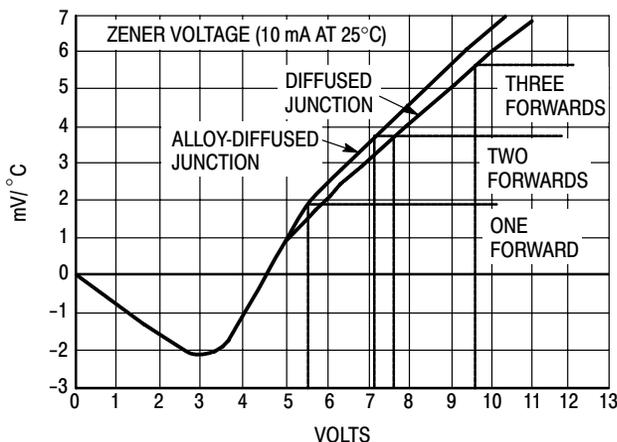


Figure 4.

TEMPERATURE COEFFICIENT STABILITY

Figure 5 shows the voltage-temperature characteristics of the TC diode. It can be seen that the voltage drops slightly with increasing temperature.

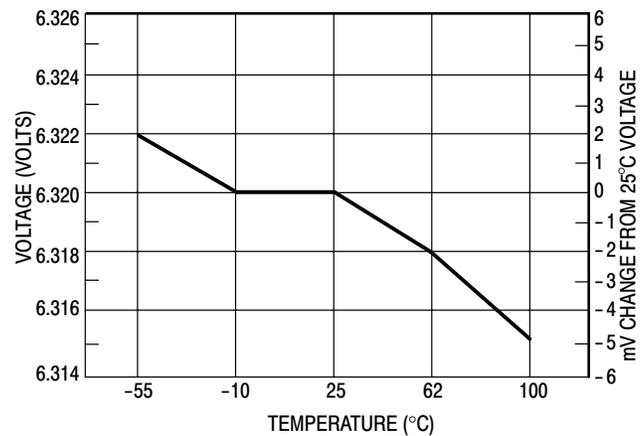


Figure 5. Voltage versus Temperature, Typical for ON Semiconductor 1N827 Temperature Compensated Zener Diode

This non-linearity of the voltage temperature characteristic leads to a definition of a representative design parameter ΔV_Z . For each device type there is a specified maximum change allowable. The voltage temperature stability measurement consists of voltage measurement at specified temperatures (for the 1N821 Series the temperatures are -55, 0, +25, +75, and +100°C). The voltage readings at each of the temperatures is compared with readings at the other temperatures and the largest voltage change between any of the specified temperatures determines the exact device type. For devices registered prior to complete definition of the voltage temperature stability measurement, the allowable maximum voltage change over the temperature range is derived from the calculation converting %/°C to mV over the temperature range. Under this standard definition, %/°C is merely a nomenclature and the meaningful allowable voltage deviation to be expected becomes the designed parameter.

CURRENT

Thus far, temperature compensated zeners have been discussed mainly with regard to temperature and voltage. However, the underlying assumption throughout the previous discussion was that current remained constant.

There is a significant change in the temperature coefficient of a unit depending on how much above or below the test current the device is operated.

A particular unit with a 0.01%/°C temperature coefficient at 7.5 mA over a temperature range of -55°C to +100°C could possibly have a 0.0005%/°C temperature coefficient at 11 mA. In fact, there is a particular current which can be determined for each individual unit that will give the lowest TC.

Manufacturing processes are designed so that the yields of low TC units are high at the test specification for current. A unit with a high TC at the test current can have a low TC at some other current. A look at the volt-ampere curves at different temperatures illustrates this point clearly (see Figure 6).

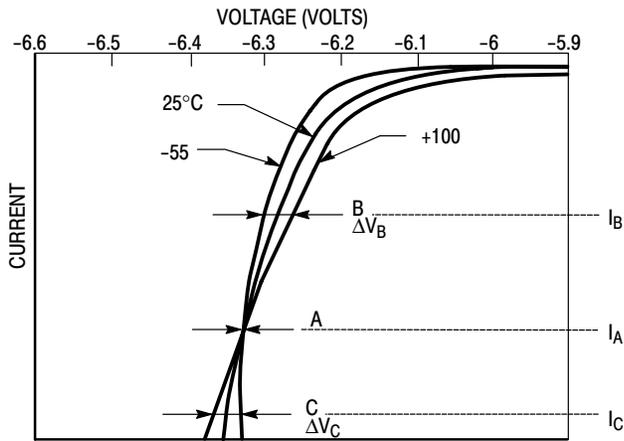


Figure 6. Voltage-Ampere Curves Showing Crossover at A

If the three curves intersect at A, then operation at I_A results in the least amount of voltage deviation due to temperature from the +25°C voltage. At I_B and I_C there are greater excursions (ΔV_B and ΔV_C) from the +25°C voltage as temperature increases or decreases.

THE EFFECTS OF POOR CURRENT REGULATION

If current shifts (randomly or as a function of temperature), then an area of operation can be defined for the temperature compensated zener.

Once again the curves are drawn, this time a shaded area is shown on the graph. The upper and lower extremities denote the maximum current values generated by the current supply while the voltage extremes at each current are shown by the left and right sides of the area, shown in Figure 7.

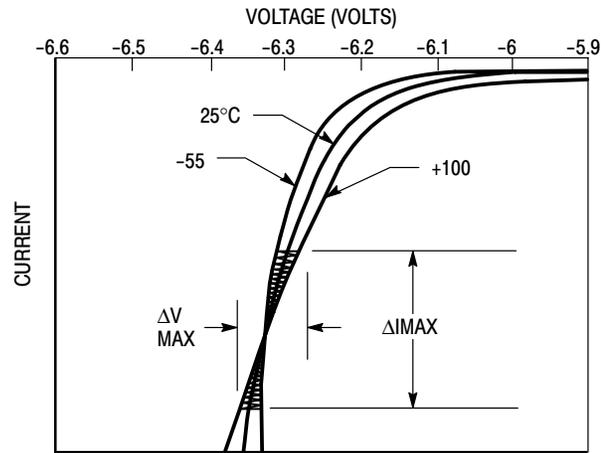


Figure 7. Effects of Poorly Regulated Current

The three volt-ampere curves do not usually cross over at exactly the same point. However, this does not take away from the argument that current regulation is probably the most critical consideration when using temperature-compensated units.

ZENER IMPEDANCE AND CURRENT REGULATION

Zener impedance is defined as the slope of the V-I curve at the test point corresponding to the test current. It is measured by superimposing a small ac current on the dc test current and then measuring the resulting ac voltage. This procedure is identical with that used for regular zeners.

Impedance changes with temperature, but the variation is usually small and it can be assumed that the amount of current regulation needed at +25°C will be the same for other temperatures.

As an example, one might want to determine the amount of current regulation necessary for the device described below when the maximum deviation in voltage due to current variation is ± 5 millivolts.

$$\begin{aligned} V_{ZT} &= 6.32 \text{ V} \\ I_{ZT} &= 7.5 \text{ mA} \\ Z_{ZT} &= 15 \Omega @ +25^\circ\text{C} \\ \Delta V &= \Delta I \cdot V Z_{ZT} \\ 0.005 &= \Delta I \cdot V 15 \\ \Delta I &= \frac{0.005}{15} = 0.33 \text{ mA} \end{aligned}$$

Therefore, the current cannot vary more than 0.33 mA. The amount of current regulation necessary is:

$$\frac{0.33}{7.5} \times 100\% = 4.5\% \text{ regulation.}$$

If the device of Figure 5 is considered to be the device used in the preceding discussion, it becomes apparent that on the average more voltage variation is due to current fluctuation than is due to temperature variation. Therefore, to obtain a truly stable reference source, the device must be driven from a constant current source.

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:

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

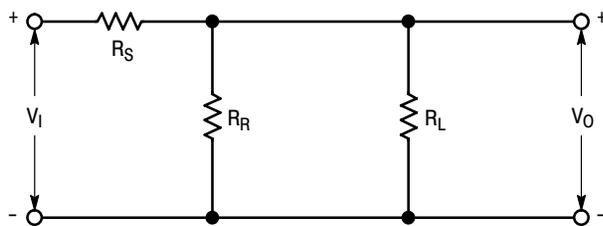


Figure 1. Shunt Resistance Regulator

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.

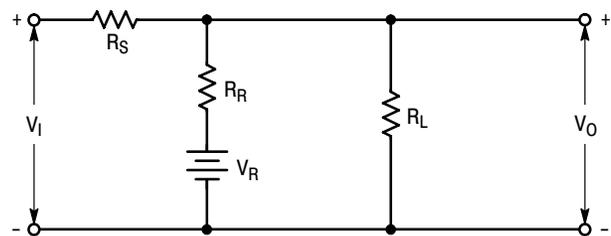


Figure 2. Regulator with Battery Reference Source

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.

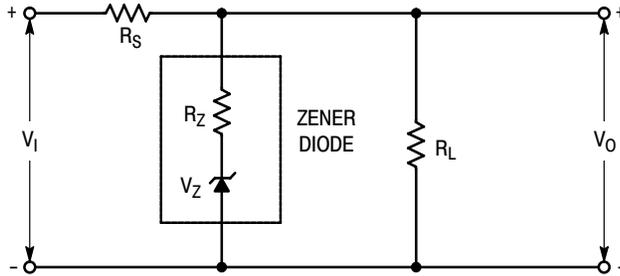


Figure 3. Basic Zener Diode Shunt Regulator

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:

$$V_I = (I_Z + I_L)R_S + V_Z \quad (5)$$

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:

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

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

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

Where:

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

Therefore:

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

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:

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

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.

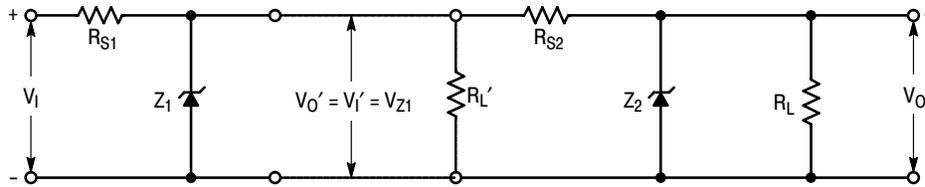


Figure 4. Cascaded Zener Shunt Regulators Reduce ΔV_O by Reducing ΔV_I to the Succeeding Stages

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:

$$\Delta V_O = \frac{\Delta V_{Z1}}{1 + \frac{R_{S2}}{R_L} + \frac{R_{S2}}{R_{Z2}}} - \frac{R_{S2}R_{Z2}}{R_{S2} + R_{Z2}} \Delta I_L + TC_2 \Delta TV_{Z2} \quad (11)$$

Where:

$$\Delta V_{Z1} = \Delta V_{O'} = \frac{\Delta V_I}{1 + \frac{R_{S1}}{R_{L'}} + \frac{R_{S1}}{R_{Z1}}} - \frac{R_{S1}R_{Z1}}{R_{S1} + R_{Z1}} \Delta I_{L'} + TC_1 \Delta TV_{Z1} \quad (12)$$

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

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

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

$$\frac{\Delta V_{O'}}{\Delta V_I} = \text{Regulation of first stage} \quad (14)$$

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

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 $\pm 20\%$ 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 $\pm 0.05\%$.

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.

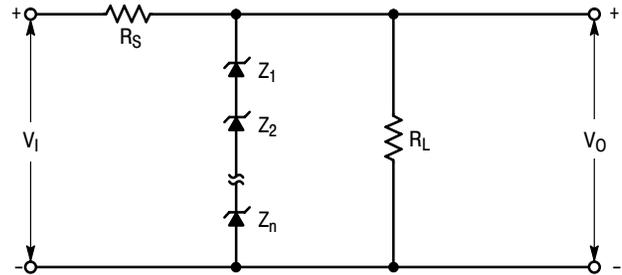


Figure 5. Series Zener Improve Dynamic Impedance and Temperature Coefficient

With the temperature coefficient TC defined as the % change per $^{\circ}\text{C}$, the change in output for a given temperature range will equal some overall TC $\times \Delta T \times \text{Total } V_Z$. Such as

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

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

$$\Delta V_{O(\Delta T)} = \Delta T (TC_1 V_{Z1} + TC_2 V_{Z2} + \dots + TC_N V_{ZN}) \quad (17)$$

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

$$TC \Delta T (V_{Z1} + V_{Z2} + \dots + V_{ZN}) = \Delta T (TC_1 V_{Z1} + TC_2 V_{Z2} + \dots + TC_N V_{ZN}) \quad (18)$$

$$TC = \frac{TC_1 V_{Z1} + TC_2 V_{Z2} + \dots + TC_N V_{ZN}}{V_{Z1} + V_{Z2} + \dots + V_{ZN}} \quad (19)$$

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 centigrade, 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.

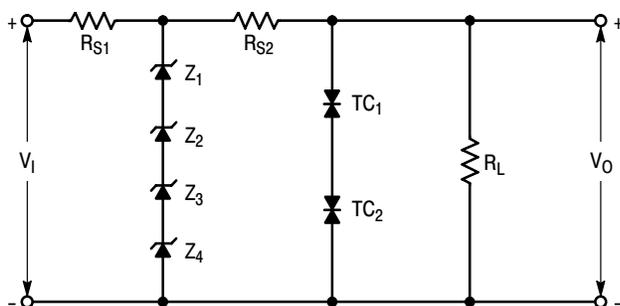


Figure 6. Series Zeners Cascaded With Series Reference Diodes for Improved Zener Shunt Regulation

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 chapter. 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.

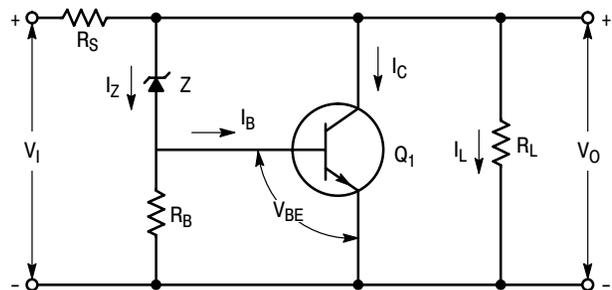


Figure 7. Basic Transistor Shunt Regulator

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

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

therefore

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

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

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

hence

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

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

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 $2 \text{ mV}/^\circ\text{C}$ TC and the transistor base to emitter being a negative $2 \text{ mV}/^\circ\text{C}$ 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.

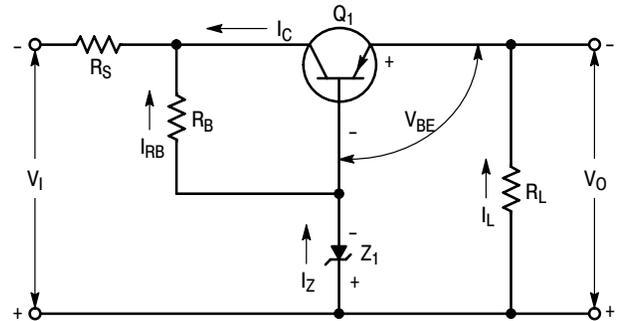


Figure 8. Emitter Follower Regulator

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:

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

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

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

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:

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

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

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

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).

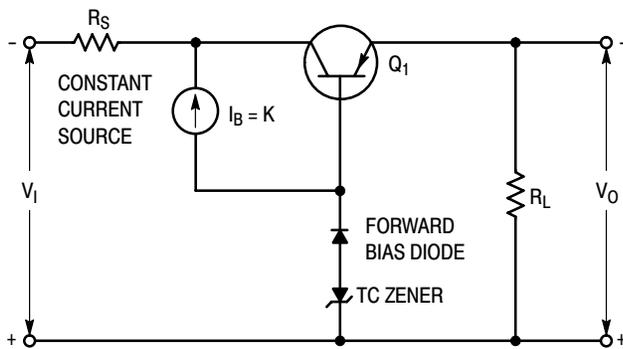


Figure 9. Improved Emitter Follower Regulator

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.

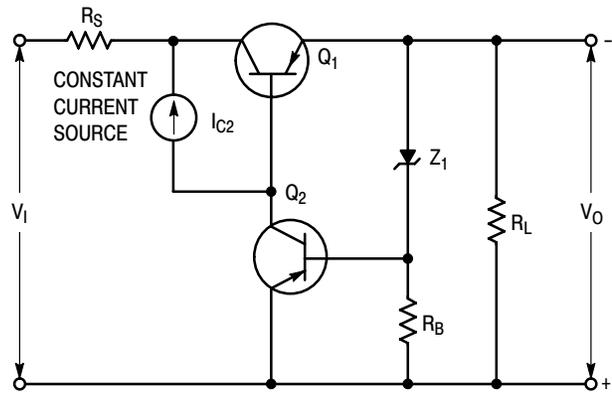


Figure 10. Series Pass Regulator

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.

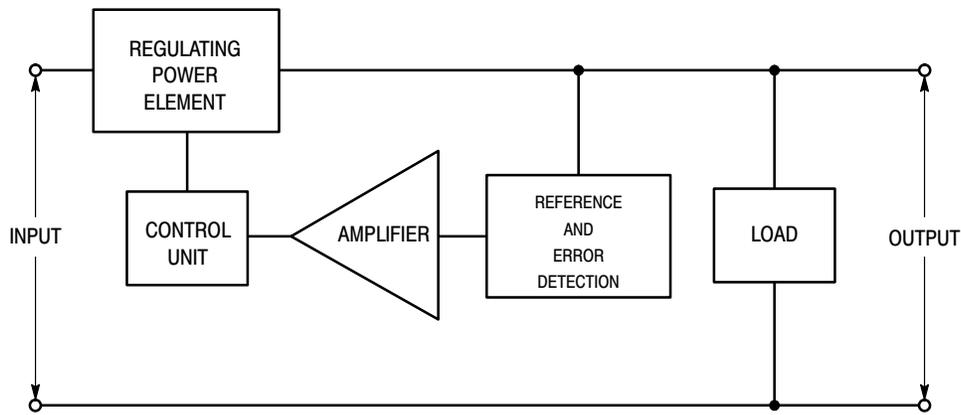


Figure 11. Block Diagram of Regulator with Feedback

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.

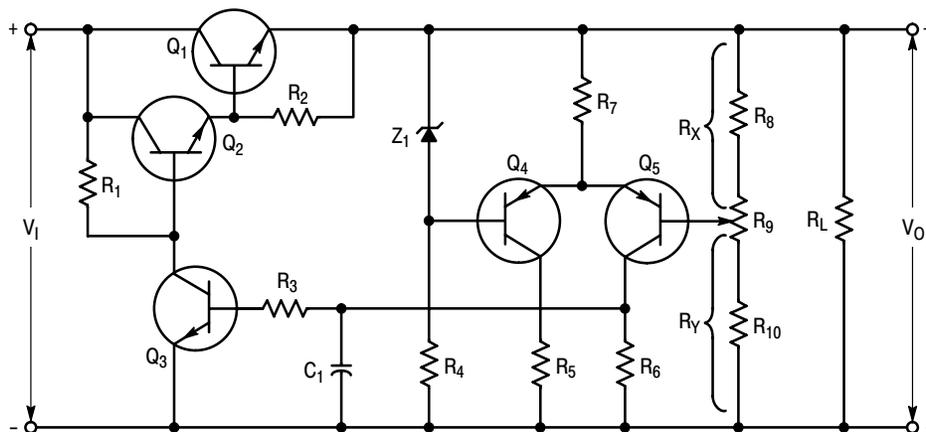


Figure 12. Series Pass Regulator with Error Detection and Feedback Amplification Derived from a Differential Amplifier

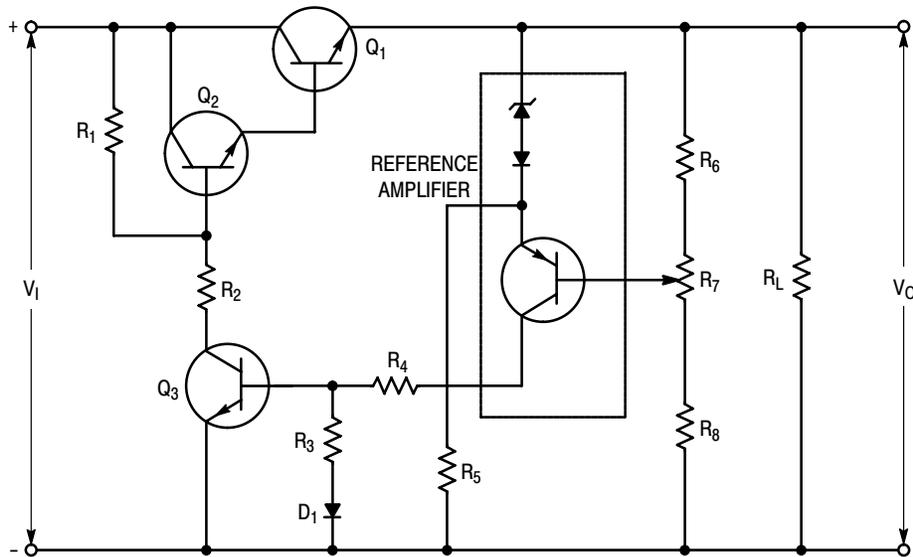


Figure 13. Series Pass Regulator with Temperature Compensated Reference Amplifier

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 chapter 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.

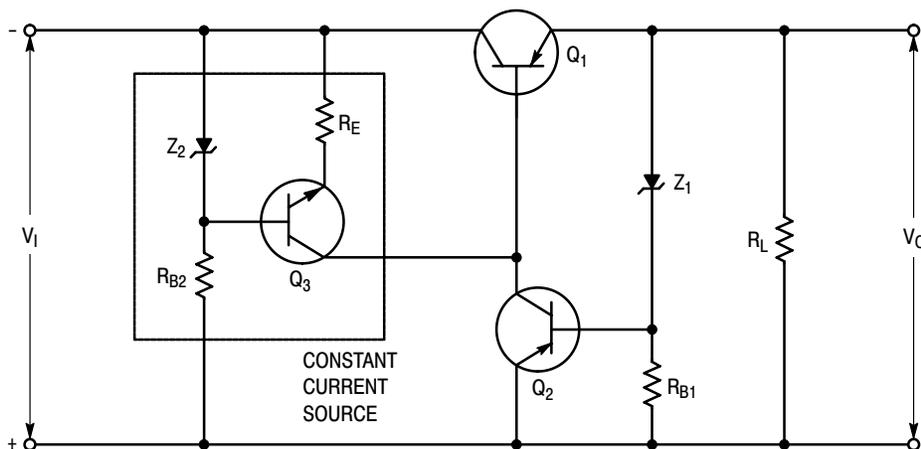


Figure 14. Constant Current Source Incorporated in a Basic Regulator Circuit

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.

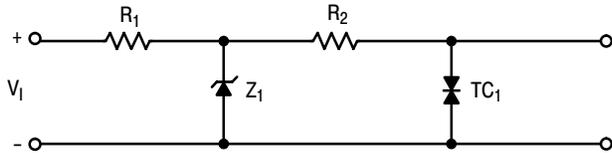


Figure 15. Impedance Cancellation with An Uncompensated Zener

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_1 in Figure 15 increase by an amount ΔV_1

$$\text{Then } \Delta I = \frac{\Delta V_1 - \Delta V_Z}{R_1} \quad (38)$$

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

$$\text{Solving } \Delta V_1 R_Z - \Delta V_Z R_Z - \Delta V_Z R_1 = 0$$

$$\text{Or } \frac{\Delta V_Z}{\Delta V_1} = \frac{R_Z}{R_1 + R_Z} \quad (40)$$

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:

$$CF_{VZ} = \frac{\Delta V_Z}{\Delta V_1} = \frac{R_Z}{R_1 + R_Z} = R_A \quad (41)$$

$$CF_{V2} = \frac{\Delta V_2}{\Delta V_1} = \frac{R_3}{R_2 + R_3} = R_B \quad (42)$$

$$CF_{V_R} = \frac{\Delta V_R}{\Delta V_1} = \frac{R_Z}{R_1 + R_Z} - \frac{R_3}{R_2 + R_3} = R_A - R_B \quad (43)$$

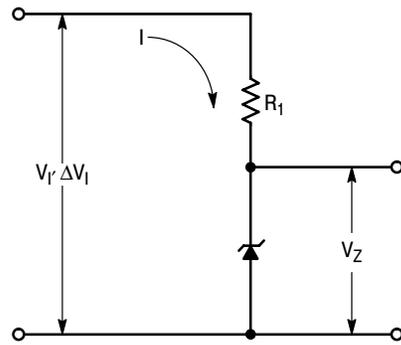


Figure 16. Standard Voltage Regulation Circuit

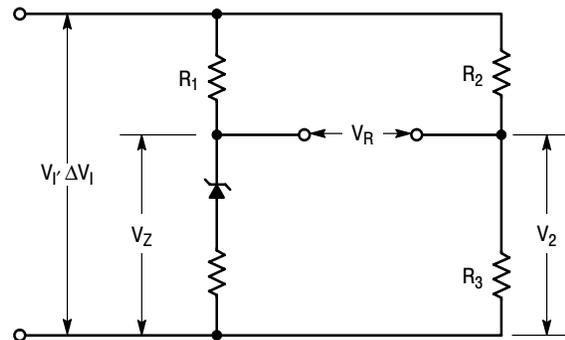


Figure 17. Impedance Cancellation Bridge

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

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

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

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

It is seen that γ_{VR} 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 PROTECTIVE CIRCUITS AND TECHNIQUES: BASIC DESIGN CONSIDERATIONS

INTRODUCTION

The reliability of any system is a function of the ability of the equipment to operate satisfactorily during moderate changes of environment, and to protect itself during otherwise damaging catastrophic changes. The silicon zener diode offers a convenient, simple but effective means of achieving this result. Its precise voltage sensitive breakdown characteristic provides an accurate limiting element in the protective circuit. The extremely high switching speed possible with the zener phenomenon allows the circuit to react faster by orders of magnitude than comparable mechanical and magnetic systems.

By shunting a component, circuit, or system with a zener diode, the applied voltage cannot exceed that of the particular device's breakdown voltage. (See Figure 1.)

A device should be chosen so that its zener voltage is somewhat higher than the nominal operating voltage but lower than the value of voltage that would be damaging if allowed to pass. In order to adequately incorporate the zener diode for circuit protection, the designer must consider several factors in addition to the required zener voltage. The first thing the designer should know is just what transient characteristics can be anticipated, such as magnitude, duration, and the rate of reoccurrence. For short duration transients, it is usually possible to suppress the voltage spike and allow the zener to shunt the transient current away from the load without a circuit shutdown. On the other hand, if the over-voltage condition is for a long duration, the protective circuit may need to be complimented with a disconnect element to protect the zener from damage created by excessive heating. In all cases, the end circuit will have to be designed around the junction temperature limits of the device.

The following sections illustrate the most common zener protective circuits, and will demonstrate the criteria to be followed for an adequate design.

BASIC PROTECTIVE CIRCUITS FOR SUPPLY TRANSIENTS

The simple zener shunt protection circuit shown in Figure 1 is widely used for supply voltage transient protection where the duration is relatively short. The circuit applies whether the load is an individual component or a complete circuit requiring protection. Whenever the input exceeds the zener voltage, the device avalanches into conduction clamping the load voltage to V_Z . The total current the diode must carry is determined by the magnitude

of the input voltage transient and the total circuit impedance minus the load current. The worst case occurs when load current is zero and may be expressed as follows:

$$I_{Z(\max)} = \frac{V_{I(\max)} - V_Z}{R_S} \quad (1)$$

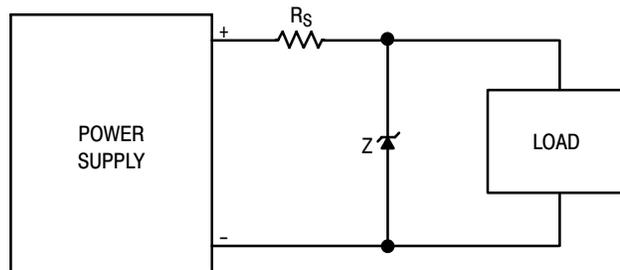


Figure 1. Basic Shunt Zener Transient Protection Circuit

The maximum power dissipated by the zener is

$$P_{Z(\max)} = I_{Z(\max)} V_{Z(\max)} = \frac{V_{I(\max)} - V_Z}{R_S} V_{Z(\max)} \quad (2)$$

Also, more than one device can be used, i.e., a series string, which will reduce the percentage of total power to be dissipated per device by a factor equal to the number of devices in series. The number of diodes required can be found from the following expression:

$$\text{Number} = \frac{P_{Z(\max)}}{P_Z \text{ (allowable per device)}} \quad (3)$$

Any fraction of a zener must be taken as the next highest whole number. This design discussion has been based upon the assumption that the transient is of a single shot, non-recurrent type. For all practical purposes it can be considered non-recurrent if the "off period" between transients is at least four times the thermal time constant of the device. If the "off period" is shorter than this, then the design calculations must include a factor for the duty cycle. This is discussed in detail in Chapter 4. In Chapter 4 there are also some typical curves relating peak power, pulse duration and duty cycle that may be appropriate for some designs.

Obviously, the factor that limits the feasibility of the basic zener shunt protective circuit is the pulse durations "t". As the duration increases, the allowable peak power for a given configuration decreases and will approach a steady state condition.

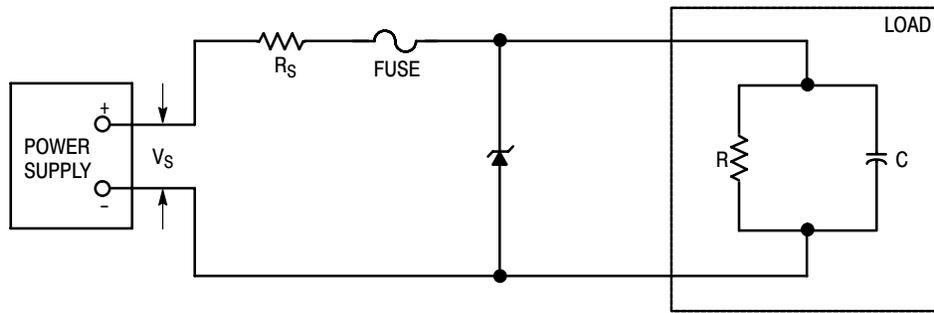


Figure 2. Overvoltage Protection with Zener Diodes and Fuses

When the anticipated transients expected to prevail for a specific situation are of long duration, a basic zener shunt becomes impractical, in such a case the circuit can be improved by using a complementary disconnect element. The most common overload protective element is without a doubt the standard fuse. The common fuse adequately protects circuit components from over-voltage surges, but at the same time must be chosen to eliminate “nuisance fusing” which results when the maximum current rating of the fuse is too close to the normal operational current of the circuit.

AN EXAMPLE PROBLEM: SELECTING A FUSE-ZENER COMBINATION

Consider the case illustrated in Figure 2. Here the load components are represented by a parallel combination of R and C, equivalent to many loads found in practice. The maximum capacitor voltage rating is usually the circuit-voltage limiting factor due to the cost of high voltage capacitors. Consequently, a protective circuit must be designed to prevent voltage surges greater than 1.5 times normal working voltage of the capacitor. It is common, however, for the supply voltage to increase to 135% normal for long periods. Examination of fuse manufacturers’ melting time-current curves shows the difficulty of trying to select a fuse which will melt rapidly at overload (within one or two cycles of the supply frequency to prevent capacitor damage), and will not melt when subjected to voltages close to overload for prolonged periods.

By connecting a zener diode of correct voltage ratings across the load as shown, a fuse large enough to withstand normal current increases for long periods may be chosen. The sudden current increase when zener breakdown occurs melts the fuse rapidly and protects the load from large surges. In Figure 3, fuse current was plotted against supply voltage to illustrate the improvement in load protection obtained with zener-fuse combinations. Fuse current “A” would be selected to limit current resulting from voltage surges above 112 V to 90 mA, which would melt the fuse in 100 ms. It is a simple matter, however, to select a fuse which melts in 30 ms at 200 mA but is unaffected by 100 mA currents. The zener connection allows fuse current “B” to be selected, eliminating this design problem and providing a faster, more reliable protective circuit. If the same fuse was used without the zener diode, a supply voltage of 210 volts

would be reached before the fuse would begin to protect the load.

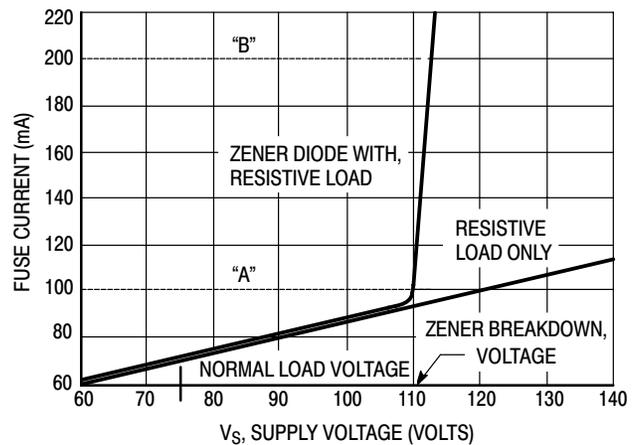


Figure 3. Fuse Current versus Supply Voltage

Selection of the correct power rating of zener diodes to be used for surge protection depends upon the magnitude and duration of anticipated surges. Often in circuits employing both fuses and zener diodes, the limiting surge duration will be the melting time of the fuse. This, in turn, depends on the nature of the load protected and the length of time it will tolerate an overload.

As a first solution to the example problem, consider a zener diode with a nominal breakdown voltage of 110 volts measured at a test current (I_{ZT}) of 110 mA. Since the fuse requires about 200 mA to melt and 100 mA are drawn through the load at this voltage, the load voltage will never exceed the zener breakdown voltage on slowly rising inputs. Transients producing currents of approximately 200 mA but of shorter duration than 30 ms will simply be clipped by zener action and diverted from the load. Transients of very high voltage will produce larger currents and, hence, will melt the fuse more rapidly. In the limiting case where transient power might eventually destroy the zener diode, the fuse always melts first because of the slower thermal time constant inherent in the zener diode’s larger geometry.

The curves in Figure 4 illustrate the change in zener voltage as a function of changing current for a typical device type.

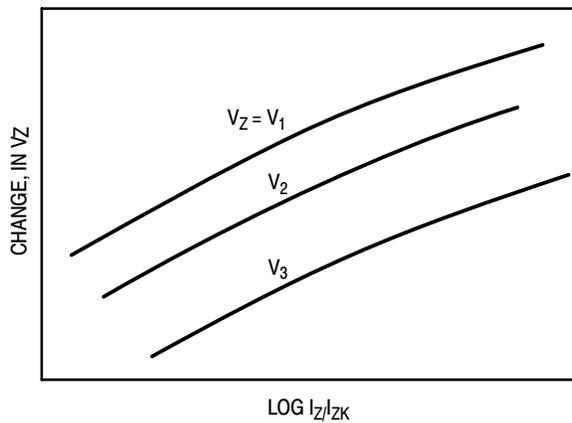


Figure 4. Change in V_Z for Changes in I_Z

If an actual curve for the device being used is not available, the zener voltage at a specific current above or below the test current may be approximated by equation 4.

Where: $V = V_Z + Z_{ZT}(I - I_{ZT})$ (4)
 V_Z = zener voltage at test current I_{ZT}
 Z_{ZT} = zener impedance at test current I_{ZT}
 I_{ZT} = test current
 V = zener voltage at current I

For a given design, the maximum zener voltage to expect for the higher zener current should be determined to make sure the limits of the circuit are met. If the maximum limit is excessive for the original device selection, the next lower voltage rating should be used.

The previous discussion on design consideration for protective circuits incorporating fuses is applicable to any protective element that permanently disconnects the supply when actuated. Rather than a fuse, a non-resetting magnetic circuit breaker could have been used, and the same reasoning would have applied.

LOAD CURRENT SURGES

In many actual problems the designer must choose a protective circuit to perform still another task. Not only must the equipment be protected from the voltage surges in the supply, but the supply itself often requires protection from shorts or partial shorts in the load. A direct short in the load is fairly easy to handle, as the drastic current change permits the use of fuses with ratings high enough to avoid problems with supply surges. More common is the partial short, as

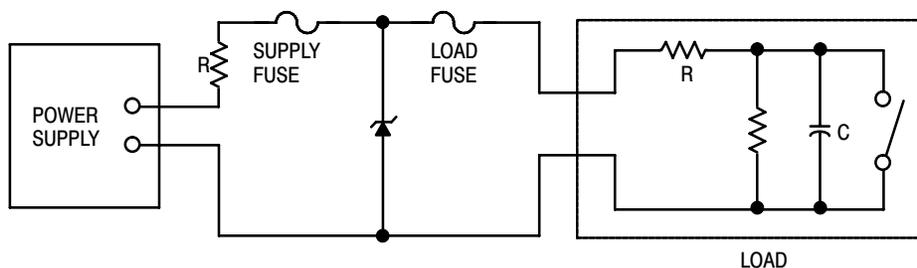


Figure 5. Supply and Load with Zener Diode; Fuse Circuitry

illustrated in Figure 5. If a short circuit occurs in the capacitive section of the load (represented by C) the resulting fault current is limited by the resistive section (represented by R) to a value which may not be great enough to melt the fuse. The fault current could be sufficient, however, to damage the supply and other components in the load.

The problem is resolved by employing a zener diode to protect against supply surges as described in the previous section, and by selecting a separate fuse to protect from load faults. The load fuse in Figure 5 is chosen close to the normal operating current. Abnormal supply surges do not affect it and equipment operates reliably but with ample protection for the supply against load changes.

ZENER DIODES AND RECLOSING DISCONNECT ELEMENTS

An interesting application of zener diodes as overvoltage protectors, which offers the possibility of designing for both long and short duration surges, is shown in Figure 6.

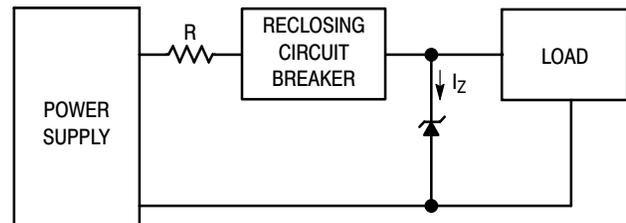


Figure 6. Zener Diode Reclosing Circuit Breaker Protective Circuit

In the event of a voltage overload exceeding a chosen zener voltage, a large current will be drawn through the diode. The reclosing disconnect element opens after an interval determined by its time constant, and the supply is disconnected. After another interval, again depending on the switch characteristics, the supply is reconnected and the voltage “sampled” by the zener diode. This leads to an “on-off” action which continues until the supply voltage drops below the predetermined limit. At no time can the load voltage or current exceed that set by the zener. The chief advantage in this type of circuit is the elimination of fuse replacement in similar fusing circuits, while providing essentially the same load protection.

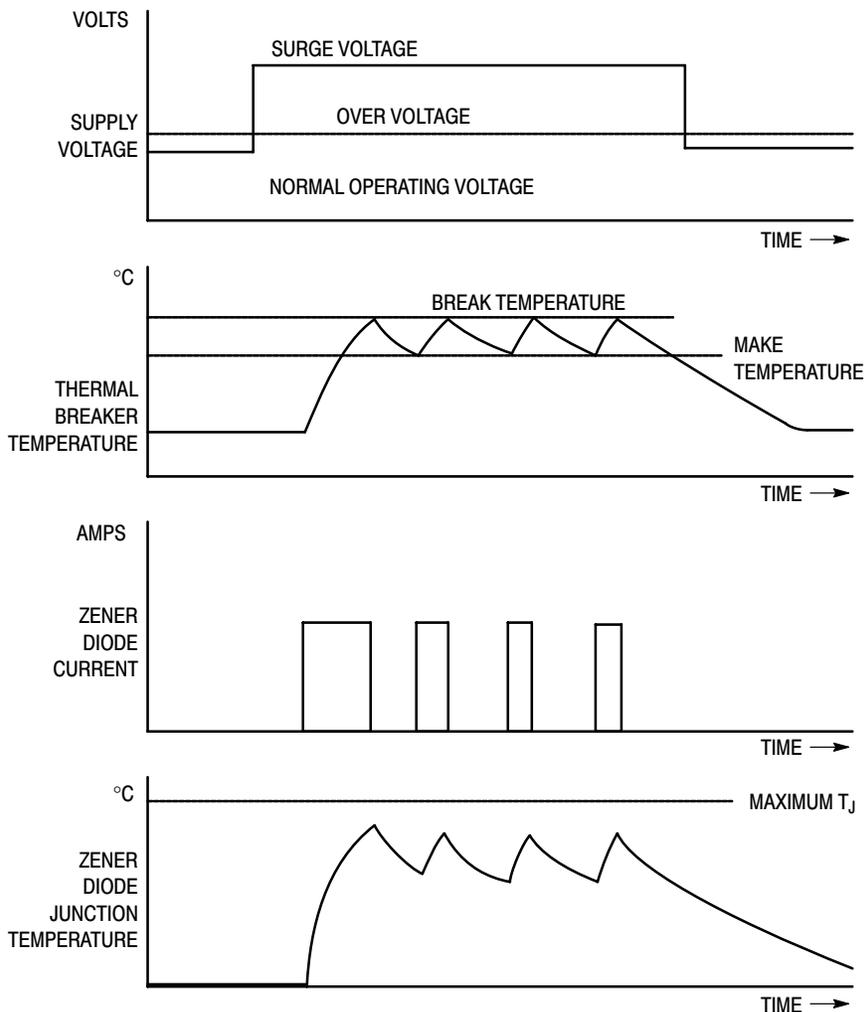


Figure 7. (Typical) Voltage, Current and Temperature Waveforms for a Thermal Breaker

It is difficult to define a set design procedure in this case, because of the wide variety of reclosing, magnetic and thermal circuit breakers available. Care should be taken to ensure that the power dissipated in the zener diode during the conduction time of the disconnect element does not exceed its rating. As an example, assume the disconnect element was a thermal breaker switch. The waveforms for a typical over-voltage situation are shown in Figure 7.

It is apparent that the highest zener diode junction temperature is reached during the first conduction period. At this time the thermal breaker is cold and requires the greatest time to reach its break temperature. The breaker then cycles thermally between the make and break temperatures as long as the supply voltage is greater than the zener voltage, as shown in Figure 7.

The zener diode current and junction temperature variation are shown in the last two waveforms of Figure 7. Overvoltage durations longer than the trip time of the thermal breaker do not affect the diode as the supply is disconnected. An overvoltage of much higher level simply

causes the thermal breaker to open sooner. In effect, the zener diode rating must be high enough to ensure that maximum junction temperature is not reached during the longest interval that the thermal switch will be closed.

Manufacturers of thermally operated circuit breakers publish current-time curves for their devices similar to that shown in Figure 8. By estimating the peak supply overvoltage and determining the maximum overvoltage tolerated by the load, an estimation of peak zener current can be made. The maximum breaker trip time may then be read from Figure 8. (After the initial current surge, the duration of “of” time is determined entirely by the breaker characteristics and will vary widely with manufacture.) The zener diode junction temperature rise during conduction may be calculated now from the thermal time constant of the device and the heatsink used.

Because the reclosing circuit breaker is continually cycling on and off, the zener current takes on the characteristics of a repetitive surge, as can be seen in Figure 7.

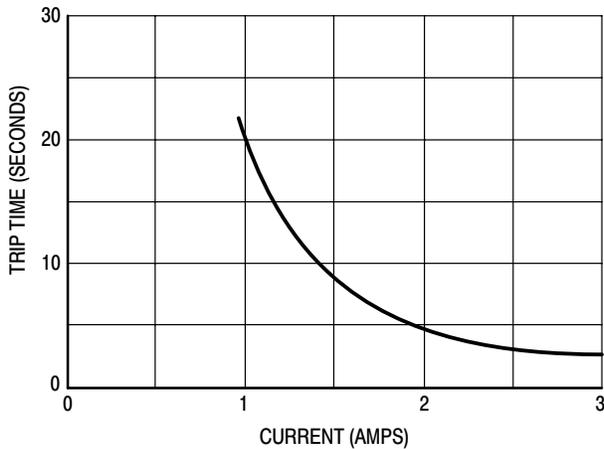


Figure 8. Trip Time versus Current for Thermal Breaker

TRANSISTOR OVERVOLTAGE PROTECTION

In many electronic circuits employing transistors, high internal voltages can be developed and, if applied to the transistors, will destroy them. This situation is quite common in transistor circuits that are switching highly inductive loads. A prime example of this would be in transistorized electronic ignition systems such as shown in Figures 9a and 9b.

The zener diode is an important component to assure solid state ignition system reliability. There are two basic methods

of using a zener diode to protect an ignition transistor. These are shown in Figures 9a and 9b. In Figure 9b the transistor is protected by a zener diode connected between base and collector and in Figure 9a, the zener is connected between emitter and collector. In both cases the voltage level of the zener must be selected carefully so that the voltage stress on the transistor is in a region where the safe operating area is adequate for reliable circuit operation.

Figure 10 illustrates "safe" and "unsafe" selection of a zener diode for collector-base protection of a transistor in the ignition coil circuit. It can be seen that the safe operating area of a transistor must be known if an adequate protective zener is to be selected.

The zener diode must be able to take the stress of peak pulse current necessary to clamp the voltage rise across the transistor to a safe value. In a typical case, a 5 watt, 100 volt zener transient suppressor diode is required to operate with an 80 μ s peak pulse current of 8 amperes when connected between the collector-emitter of the transistor. The waveform of this pulse current approaches a sine wave in shape (Figure 11). The voltage rise across a typical transient suppressor diode due to this current pulse is shown in Figure 12. This voltage rise of approximately 8 volts indicates an effective zener impedance of approximately 1 ohm. However, a good share of this voltage rise is due to the temperature coefficient and thermal time constant of the zener. The temperature rise of the zener diode junction is indicated by the voltage difference between the rise and fall of the current pulse.

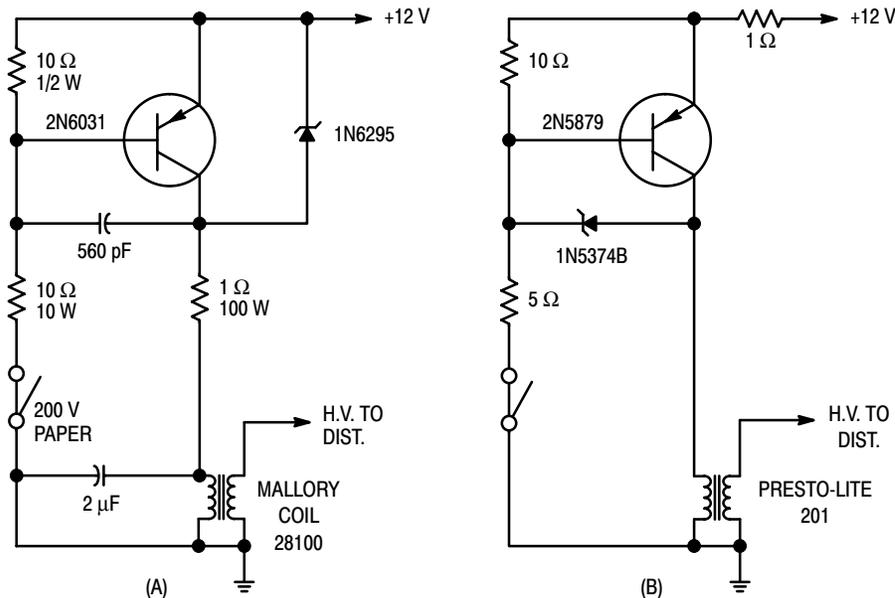


Figure 9. Transistor Ignition Systems with Zener Overvoltage Surge Protection

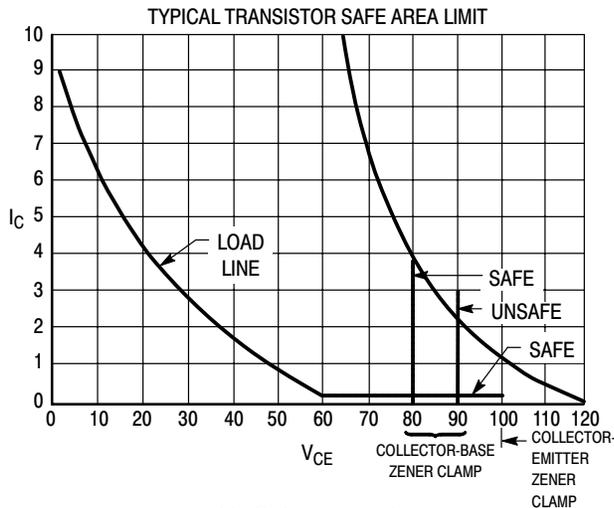


Figure 10. Safe Zener Protection

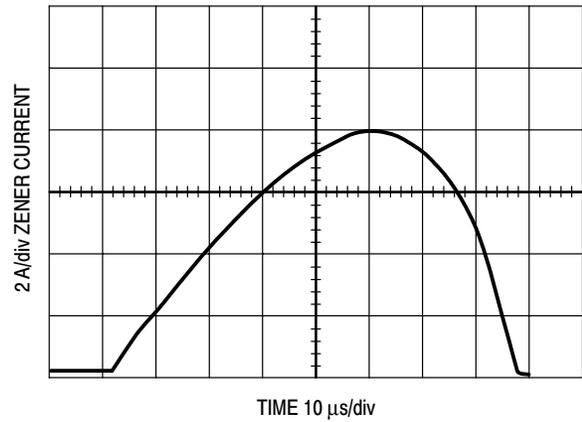


Figure 11. Zener Diode Current Pulse

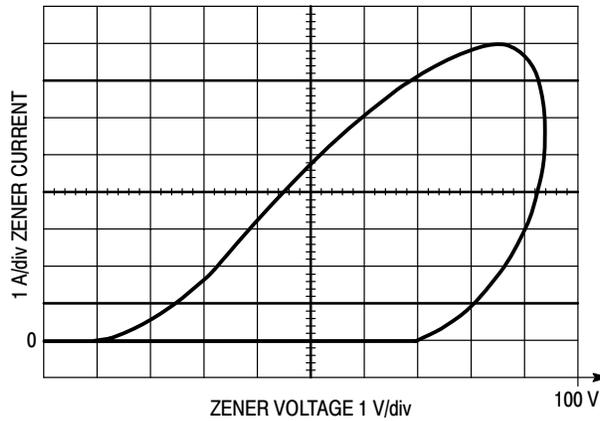


Figure 12. Voltage-Current Representation on 100 V Zener

In order to assure safe operation, the change in zener junction temperature for the peak pulse conditions must be analyzed. In making the calculation, the method described in Chapter 3 should be used, taking into account duty cycle, pulse duration, and pulse magnitude.

When the zener diode is connected between the collector and emitter of the transistor, additional power dissipation will result from the clipping of the ringing voltage of the ignition coil by the forward conduction of the zener diode. This power dissipation by the forward diode current will

result in additional zener voltage rise. It is not uncommon to observe a 15-volt rise above the zener device voltage rating due to temperature coefficient and impedance under these pulse current conditions.

The zener diode should be connected as close as possible to the terminals of the transistor the zener is intended to protect. This insures that induced voltage transients, caused by current changes in long lead lengths, are clamped by the zener and do not appear across the transistor.

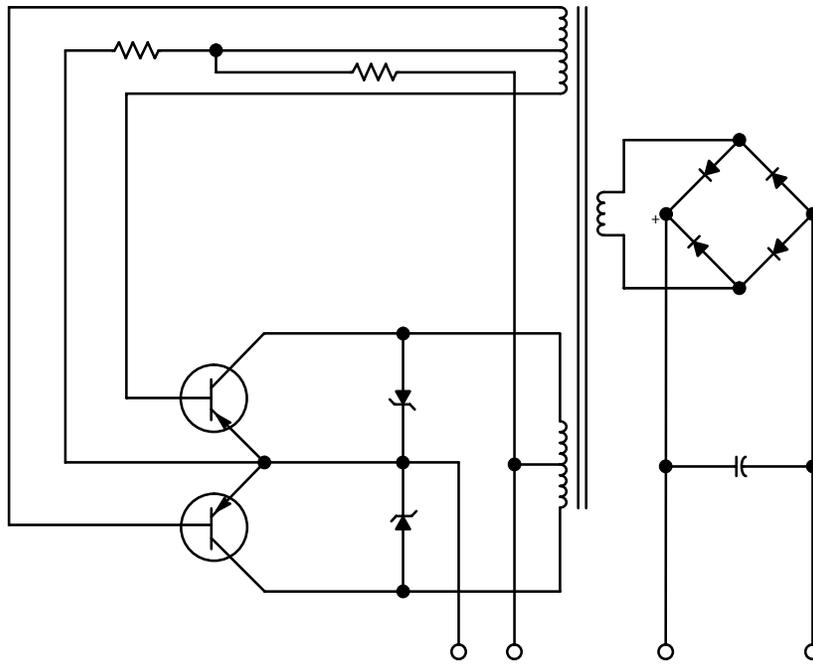


Figure 13. DC-DC Converter with Surge Protecting Diodes

Another example of overvoltage protection of transistor operating in an inductive load switch capacity is illustrated in Figure 13. The DC-DC converter circuit shows a connection from collector to emitter of two zener diodes as collector overvoltage protectors. Without some type of limiting device, large voltage spikes may appear at the collectors, due to the switching transients produced with normal circuit operation. When this spike exceeds the collector breakdown rating of the transistor, transistor life is considerably shortened. The zener diodes shown are chosen with zener breakdowns slightly below transistor breakdown voltage to provide the necessary clipping action. Since the spikes are normally of short duration (0.5 to 5 μ s) and duty cycle is low, normal chassis mounting provides adequate heatsinking.

METER PROTECTION

The silicon zener diode can be employed to prevent overloading sensitive meter movements used in low range DC and AC voltmeters, without adversely affecting the meter linearity. The zener diode has the advantage over thermal protective devices of instantaneous action and, of course, will function repeatedly for an indefinite time (as compared to the reset time necessary with thermal devices). While zener protection is presently available for voltages as low as 2.4 volts, forward diode operation can be used for meter protection where the voltage drop is much smaller. A typical protective circuit is illustrated in Figure 14. Here the meter movement requires 100 μ Amps for full scale deflection and has 940 ohms resistance. For use in a

voltmeter to measure 25 V, approximately 249 thousand ohms are required in series.

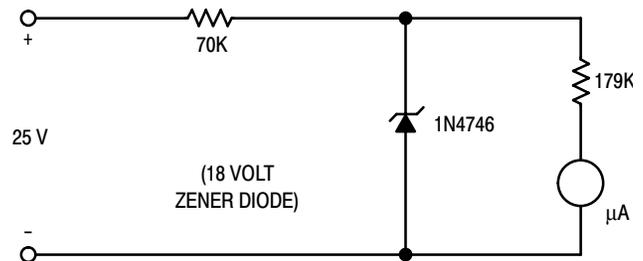


Figure 14. Meter Protection with Zener Diode

The protection provided by the addition of an 18 volt zener is illustrated in Figure 15. With an applied voltage of 25 volts, the 100 μ Amps current in the circuit produces a drop of 17.9 volts across the series resistance of 179 thousand ohms. A further increase in voltage causes the zener diode to conduct, and the overload current is shunted away from the meter. Since ON Semiconductor zener diodes have zener voltages specified within 5 and 10%, a safe design may always be made with little sacrifice in meter linearity by assuming the lowest breakdown voltage within the tolerance. The shunting effect on the meter of the reverse biased diode is generally negligible below breakdown voltage (on the order of 0.5° full scale). For very precise work, the zener diode breakdown voltage must be accurately known and the design equations solved for the correct resistance values.

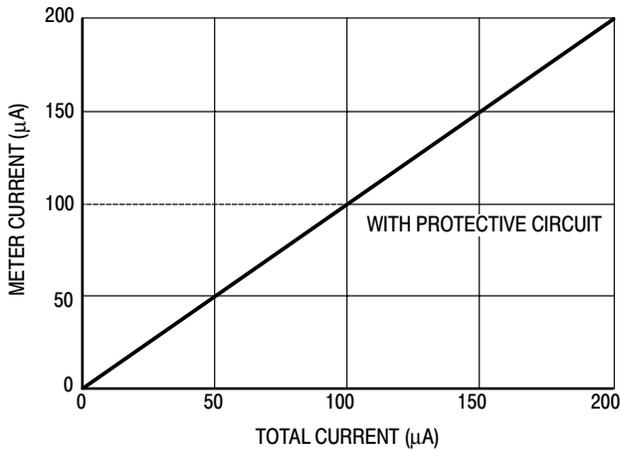


Figure 15. Meter Protection with Zener Diodes

ZENER DIODES USED WITH SCRS FOR CIRCUIT PROTECTION

An interesting aspect of circuit protection incorporating the reliable zener diode is the protective circuits shown in Figures 16 and 17.

In a system that is handling large amounts of power, it may become impractical to employ standard zener shunt protection because of the large current it would be required to carry. The SCR crowbar technique shown in Figure 16 can be effectively used in these situations. The zener diode is still the transient detection component, but it is only required to carry the gate current for SCR turn on, and the SCR will carry the bulk of the shunt current. Whenever the incoming voltage exceeds the zener voltage, it avalanches, supplying gate drive to the SCR which, when fired, causes a current demand that will trip the circuit breaker. The resistors shown are for current limiting so that the SCR and zener ratings are not exceeded.

The circuit of Figure 17 is designed to disconnect the supply in the event a specified load current is exceeded. This is done by means of a series sense resistor and a compatible zener to turn the shunt SCR on. When the voltage across the series resistor, which is a function of the load current, becomes sufficient to break over the zener, the SCR is fired, causing the circuit breaker to trip.

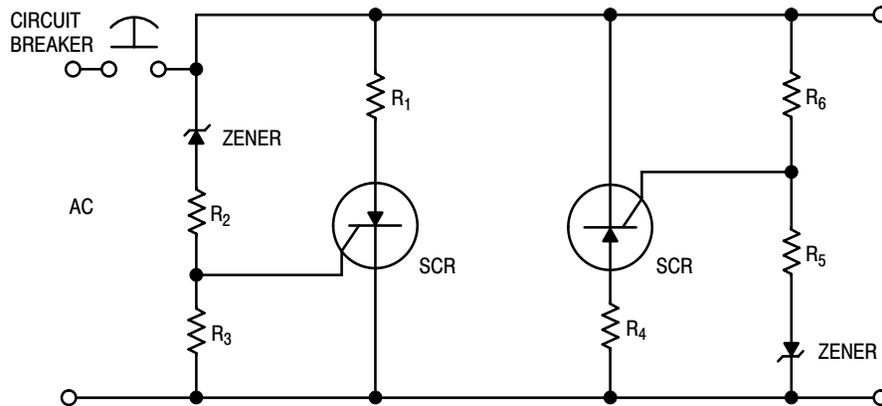


Figure 16. SCR Crowbar Over-Voltage Protection Circuit for AC Circuit Operation

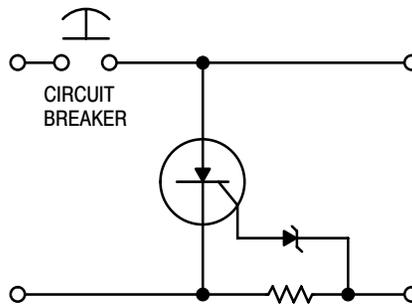


Figure 17. SCR Longterm Current Overload Protection

ZENER TRANSIENT SUPPRESSORS

The transient suppressor is used as a shunt element in exactly the same manner as a conventional zener. It offers the same advantages such as low insertion loss, immediate recovery after operation, a clamping factor approaching unity, protection against fast rising transients, and simple circuitry. The primary difference is that the transient suppressor extends these advantages to higher power levels.

Even in the event of transients with power contents far in excess of the capacity of the zeners, protection is still provided the load. When overloaded to failure, the zener will approximate a short. The resulting heavy drain will aid in opening the fuse or circuit breaker protecting the load against excess current. Thus, even if the suppressor is destroyed, it still protects the load.

The design of the suppressor-fuse combination for the required level of protection follows the techniques for conventional zeners discussed earlier in this chapter.

TRANSIENT SUPPRESSION CHARACTERISTICS

Zener diodes, being nearly ideal clippers (that is, they exhibit close to an infinite impedance below the clipping level and close to a short circuit above the clipping level), are often used to suppress transients. In this type of application, it is important to know the power capability of the zener for short pulse durations, since they are intolerant of excessive stress.

Some ON Semiconductor data sheets such as the ones for devices shown in Table 1 contain short pulse surge capability. However, there are many data sheets that do not contain this data and Figure 18 is presented here to supplement this information.

Table 1. Transient Suppressor Diodes

Series Numbers	Steady State Power	Package	Description
1N4728A	1 W	DO-41	Double Slug Glass
1N6267A	5 W	Case 41A	Axial Lead Plastic
1N5333B	5 W	Case 102	Surmetic 40
1N746A/957B /4370A	500 mW	DO-35	Double Slug Glass
1N5221B	500 mW	DO-35	Double Slug Glass

Some data sheets have surge information which differs slightly from the data shown in Figure 18. A variety of reasons exist for this:

1. The surge data may be presented in terms of actual surge power instead of nominal power.
2. Product improvements have occurred since the data sheet was published.

3. Large dice are used, or special tests are imposed on the product to guarantee higher ratings than those shown in Figure 18.
4. The specifications may be based on a JEDEC registration or part number of another manufacturer.

The data of Figure 18 applies for non-repetitive conditions and at a lead temperature of 25°C. If the duty cycle increases, the peak power must be reduced as indicated by the curves of Figure 19. Average power must be derated as the lead or ambient temperature rises above 25°C. The average power derating curve normally given on data sheets may be normalized and used for this purpose.

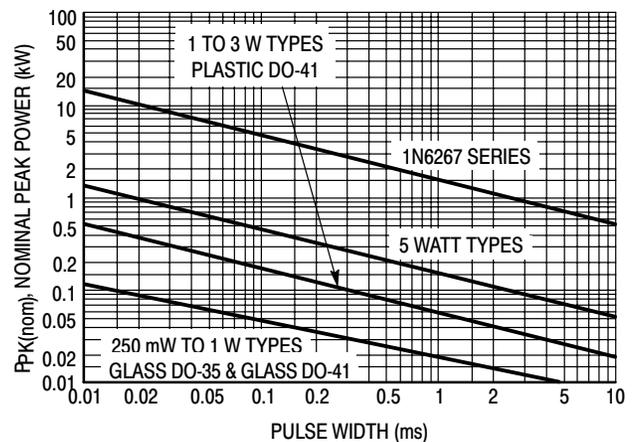


Figure 18. Peak Power Ratings of Zener Diodes

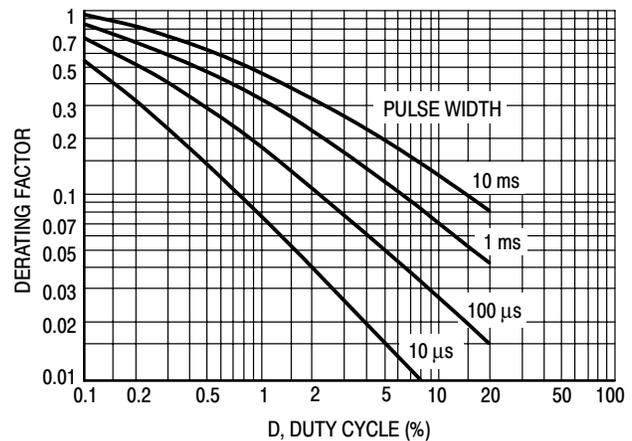


Figure 19. Typical Derating Factor for Duty Cycle

When it is necessary to use a zener close to surge ratings, and a standard part having guaranteed surge limits is not suitable, a special part number may be created having a surge limit as part of the specification. Contact your nearest ON Semiconductor OEM sales office for capability, price, delivery, and minimum order quantities.

MATHEMATICAL MODEL

Since the power shown on the curves is not the actual transient power measured, but is the product of the peak current measured and the nominal zener voltage measured at the current used for voltage classification, the peak current can be calculated from:

$$I_{Z(PK)} = \frac{P_{(PK)}}{V_{Z(nom)}} \quad (5)$$

The peak voltage at peak current can be calculated from:

$$V_{Z(PK)} = F_C \times V_{Z(nom)} \quad (6)$$

where F_C is the clamping factor. The clamping factor is approximately 1.20 for all zener diodes when operated at their pulse power limits. For example, a 5 watt, 20 volt zener can be expected to show a peak voltage of 24 volts regardless of whether it is handling 450 watts for 0.1 ms or 50 watts for 10 ms. This occurs because the voltage is a function of junction temperature and IR drop. Heating of the junction is more severe at the longer pulse width, causing a higher voltage component due to temperature which is roughly offset by the smaller IR voltage component.

For modeling purposes, an approximation of the zener resistance is needed. It is obtained from:

$$R_{Z(nom)} = \frac{V_{Z(nom)}(F_C - 1)}{P_{PK(nom)} / V_{Z(nom)}} \quad (7)$$

The value is approximate because both the clamping factor and the actual resistance are a function of temperature.

CIRCUIT CONSIDERATIONS

It is important that as much impedance as circuit constraints allow be placed in series with the zener diode and the components to be protected. The result will be a lower clipping voltage and less zener stress. A capacitor in parallel

with the zener is also effective in reducing the stress imposed by very short duration transients.

To illustrate use of the data, a common application will be analyzed. The transistor in Figure 20 drives a 50 mH solenoid which requires 5 amperes of current. Without some means of clamping the voltage from the inductor when the transistor turns off, it could be destroyed.

The means most often used to solve the problem is to connect an ordinary rectifier diode across the coil; however, this technique may keep the current circulating through the coil for too long a time. Faster switching is achieved by allowing the voltage to rise to a level above the supply before being clamped. The voltage rating of the transistor is 60 V, indicating that approximately a 50 volt zener will be required.

The peak current will equal the on-state transistor current (5 amperes) and will decay exponentially as determined by the coil L/R time constant (neglecting the zener impedance). A rectangular pulse of width L/R (0.01 s) and amplitude of I_{PK} (5 A) contains the same energy and may be used to select a zener diode. The nominal zener power rating therefore must exceed $(5 A \times 50) = 250$ watts at 10 ms and a duty cycle of $0.01/2 = 0.5\%$. From Figure 19, the duty cycle factor is 0.62 making the single pulse power rating required equal to $250/0.62 = 403$ watts. From Figure 18, one of the 1N6267 series zeners has the required capability. The 1N6287 is specified nominally at 47 volts and should prove satisfactory.

Although this series has specified maximum voltage limits, equation 7 will be used to determine the maximum zener voltage in order to demonstrate its use.

$$R_Z = \frac{47(1.20 - 1)}{500/47} = \frac{9.4}{10.64} = 0.9 \Omega$$

At 5 amperes, the peak voltage will be 4.5 volts above nominal or 51.5 volts total which is safely below the 60 volt transistor rating.

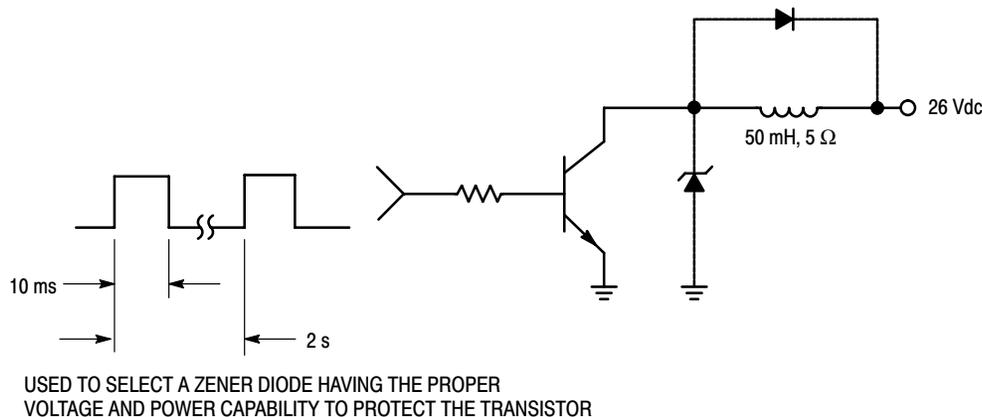


Figure 20. Circuit Example

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, function control, or circuit protection. 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.

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.

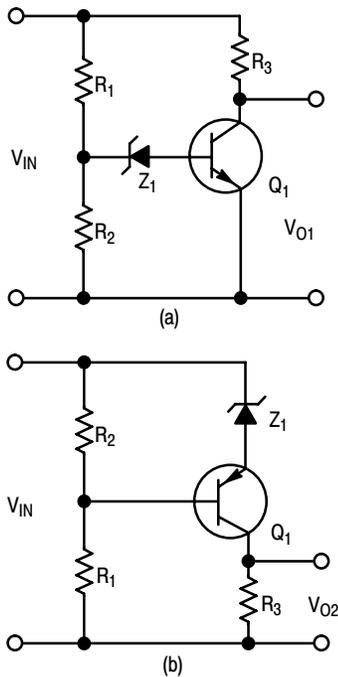


Figure 1. Basic Transistor-Zener Diode Sensing Circuits

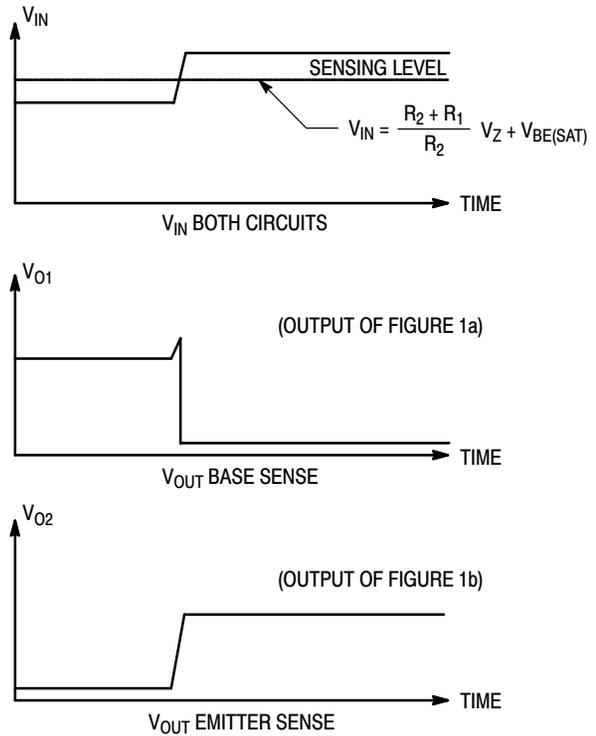


Figure 2. Outputs of Transistor-Zener Voltage Sensing Circuits

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	$\begin{cases} \text{High} \\ V_{OUT} = V_{IN} - I_{CO}R_3 \cong V_{IN} \\ \text{Low} \\ V_{OUT} = V_{IN} - I_C R_3 = V_{CE(sat)} \end{cases}$
1b	$\begin{cases} \text{Low} \\ V_{OUT} = V_{IN} - V_Z - V_{CE(off)} = I_{CO}R_3 \\ \text{High} \\ V_{OUT} = V_{IN} - V_{CE(sat)} = I_C R_3 \end{cases}$

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.

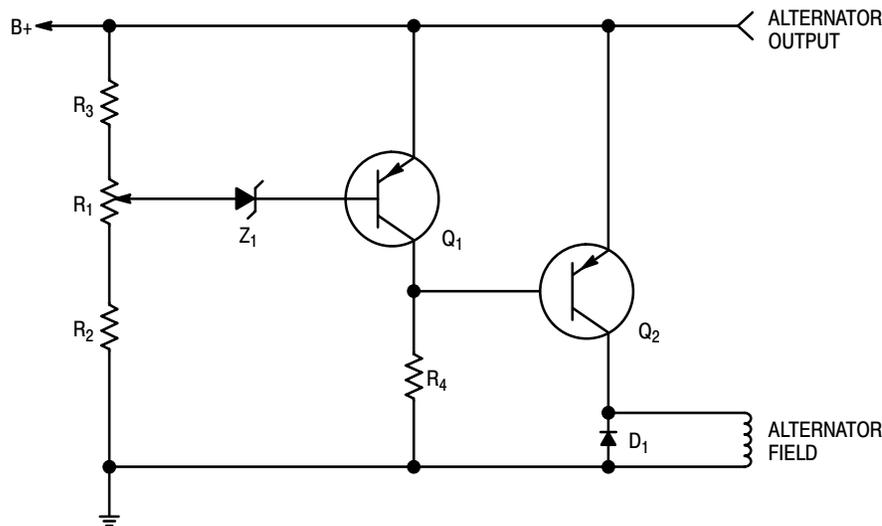


Figure 3. Simplified Solid State Voltage Regulator

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.

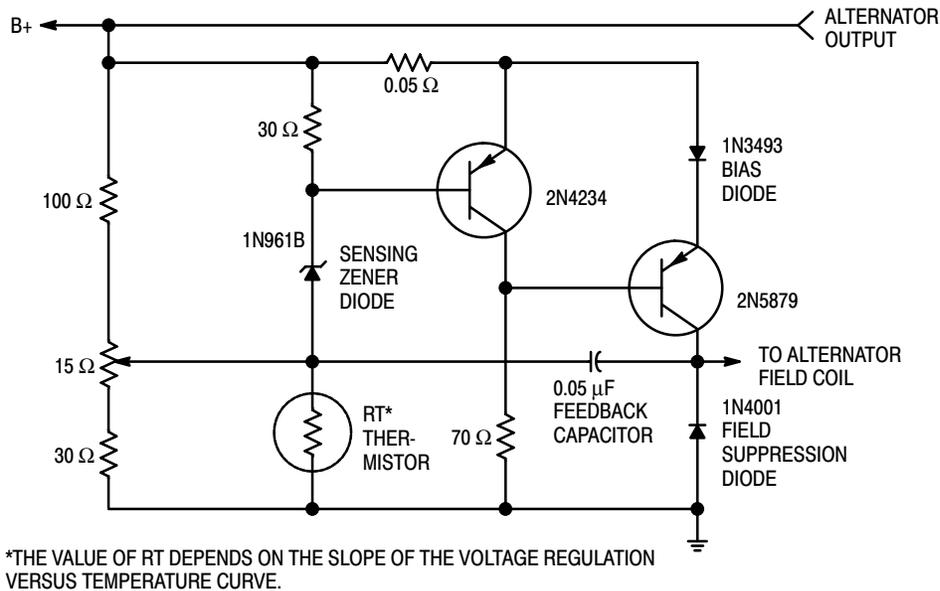


Figure 4. Complete Solid State Alternator Voltage Regulator

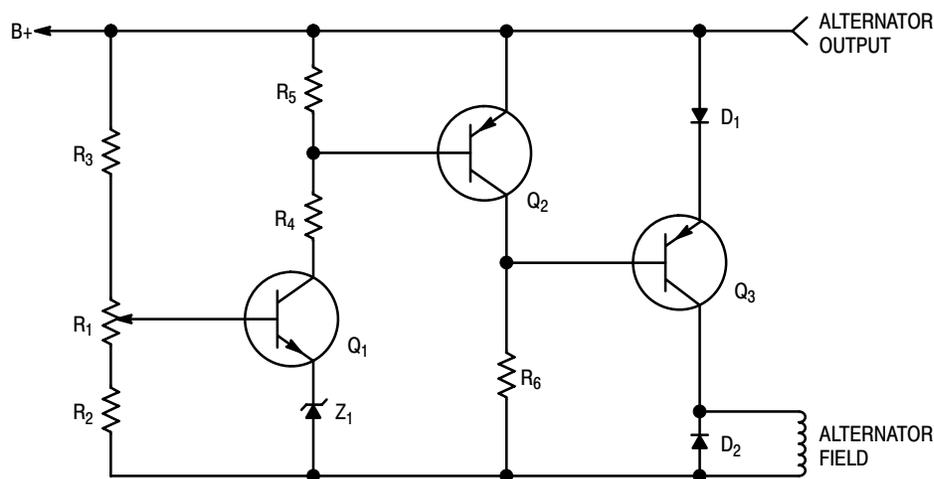


Figure 5. Alternator Regulator with Emitter Sensor

UNIUNCTION-ZENER SENSE CIRCUITS

Unijunction transistor oscillator circuits can be made GO-NO GO voltage sensitive by incorporating a zener diode clamp. The UJT operates on the criterion: under proper biasing conditions the emitter-base one junction will breakover when the emitter voltage reaches a specific value given by the equation:

$$V_p = \eta V_{BB} + V_D \quad (1)$$

where:

- V_p = peak point emitter voltage
- η = intrinsic stand-off ratio for the device
- V_{BB} = interbase voltage, from base two to base one
- V_D = emitter to base one diode forward junction drop

Obviously, if we provide a voltage clamp in the circuit such that the conditions of equation 1 are met only with restriction on the input, the circuit becomes voltage sensitive. There are two basic techniques used in clamping UJT relaxation oscillators. They are shown in Figure 6 and Figure 7.

The circuit in Figure 6 is that of a clamped emitter type. As long as the input voltage V_{IN} is low enough so that V_p

does not exceed the Zener voltage V_Z , the circuit will generate output pulses. At some given point, the required V_p for triggering will exceed V_Z , the circuit will not oscillate. This, in essence, means the circuit is GO as long as V_{IN} is below a certain level, and NO GO above the critical clamp point.

The circuit of Figure 7, is a clamped base UJT oscillator. In this circuit V_{BB} is clamped at a voltage V_Z and the emitter tied to a voltage dividing network by a diode D_1 . When the input voltage is low, the voltage drop across R_2 is less than V_p . The forward biased diode holds the emitter below the trigger level. As the input increases, the R_2 voltage drop approaches V_p . The diode D_1 becomes reversed biased and, the UJT triggers. This phenomenon establishes the operating criterion that the circuit is NO GO at a low input and GO at an input higher than the clamp voltage. Therefore, the circuits in Figures 6 and 7 are both input voltage sensitive, but have opposite input requirements for a GO condition. To illustrate the usefulness of the clamped UJT relaxation oscillators, the following two sections show them being used in practical applications.

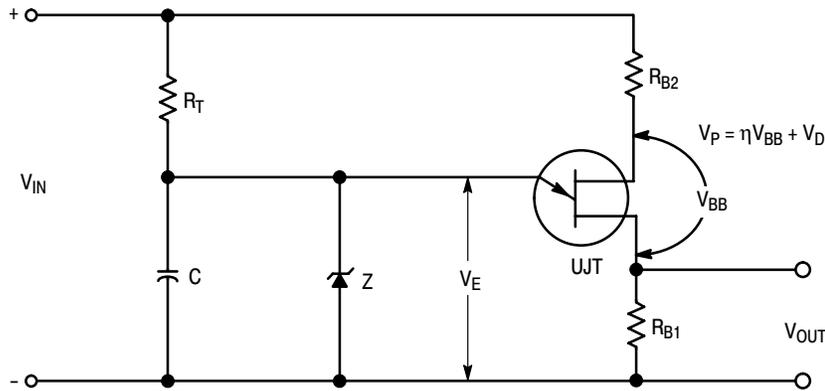


Figure 6. UJT Oscillator, GO — NO GO Output, GO for Low V_{IN} — NO GO for High V_{IN}

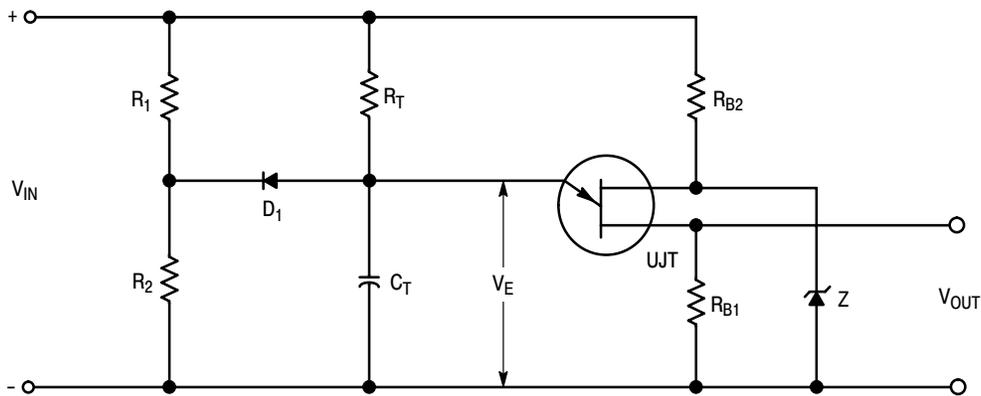


Figure 7. UJT — NO GO Output, NO GO for Low V_{IN} — GO for High V_{IN}

When the peak point voltage (switching voltage) of the unijunction transistor exceeds the breakdown voltage of the Zener diode, Z_1 , connected across the delay circuit capacitor, C_1 , the unijunction transistor ceases to oscillate. If the relaxation oscillator does not operate, the controlled rectifier will not receive trigger pulses and will not conduct. This indicates that the battery has attained its desired charge as set by R_2 .

The unijunction cannot oscillate unless a voltage somewhere between 3 volts and the cutoff setting is present at the output terminals with polarity as indicated. Therefore, the SCR cannot conduct under conditions of a short circuit, an open circuit, or a reverse polarity connection to the battery.

ALTERNATOR REGULATOR FOR PERMANENT MAGNET FIELD

In alternator circuits such as those of an outboard engine, the field may be composed of a permanent magnet. This increases the problem of regulating the output by limiting the control function to opening or shorting the output. Because of the high reactance source of most alternators, opening the output circuit will generally stress the bridge rectifiers to a very high voltage level. It is, therefore, apparent that the best control function would be shorting the output of the alternator for regulation of the charge to the battery.

Figure 10 shows a permanent magnet alternator regulator designed to regulate a 15 ampere output. The two SCRs are connected on the ac side of the bridge, and short out the alternator when triggered by the unijunction voltage sensitive triggering circuit. The sensing circuit is of the type shown in Figure 7. The shorted output does not appreciably increase the maximum output current level.

A single SCR could be designed into the dc side of the bridge. However, the rapid turn-off requirement of this type of circuit at high alternator speeds makes this circuit impractical.

The unijunction circuit in Figure 10 will not oscillate until the input voltage level reaches the voltage determined by the intrinsic standoff ratio. The adjustable voltage divider will calibrate the circuit. The series diode in the voltage divider circuit will compensate for the emitter-base-one diode temperature change, consequently, temperature compensation is necessary only for the zener diode temperature changes.

Due to the delay in charging the unijunction capacitor, when the battery is disconnected the alternator voltage will produce high stress voltage on all components before the SCRs will be fired. The 1N971B Zener was included in the circuit to provide a trigger pulse to the SCRs as soon as the alternator output voltage level approaches 30 volts.

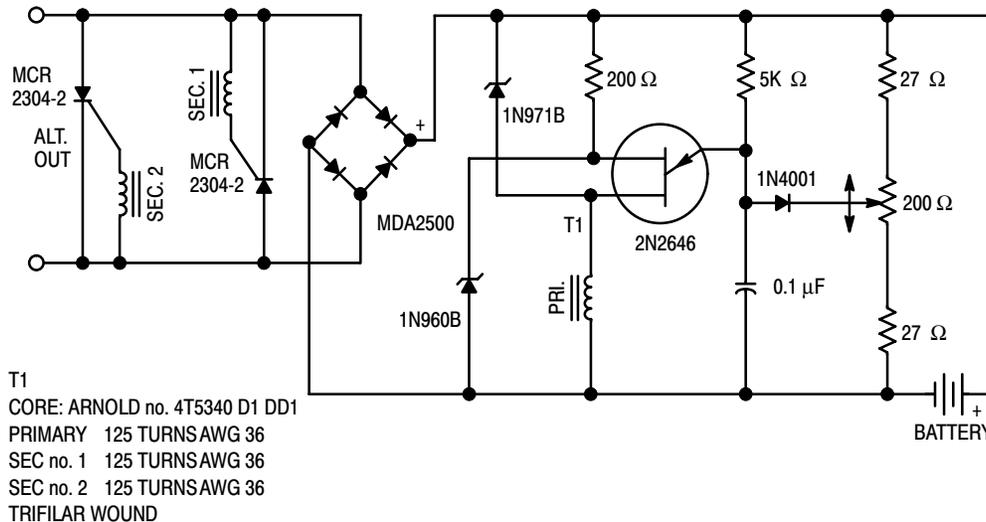


Figure 10. Permanent Magnet Field Alternator Regulator

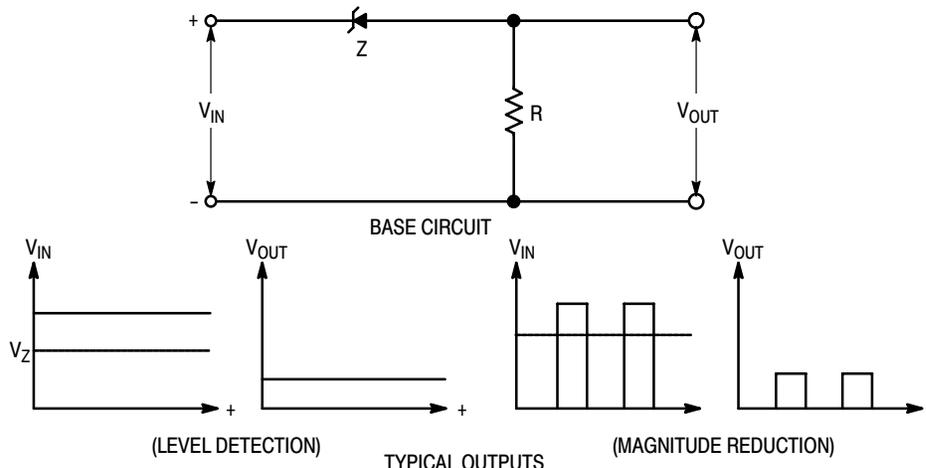


Figure 11. Zener-Resistor Voltage Sensitive Circuit

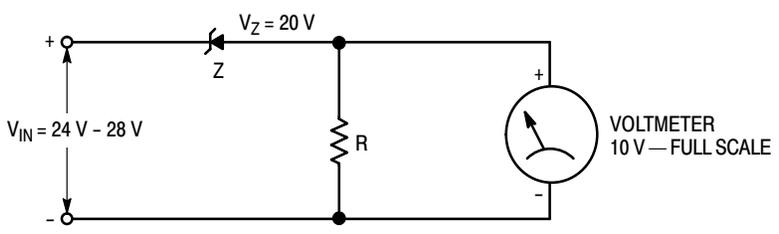


Figure 12. Improving Meter Resolution

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 11.

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 12, 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 13, 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.

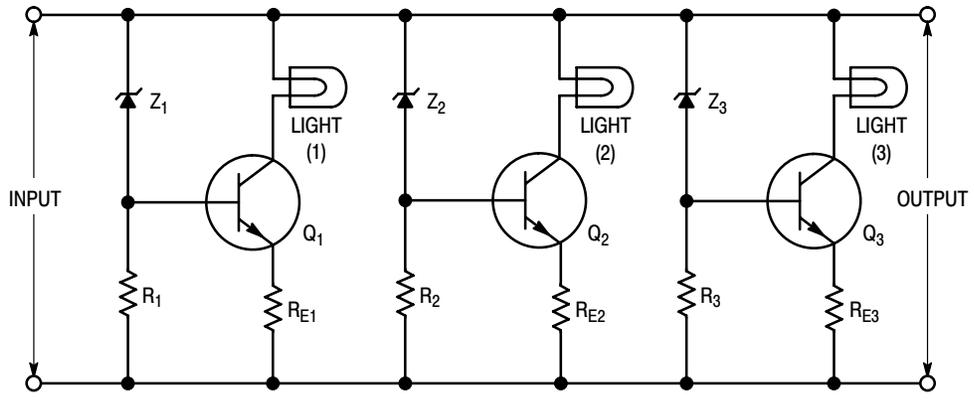


Figure 13. Sequential Voltage Level Indicator

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 chapters. This chapter 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 chapter 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.

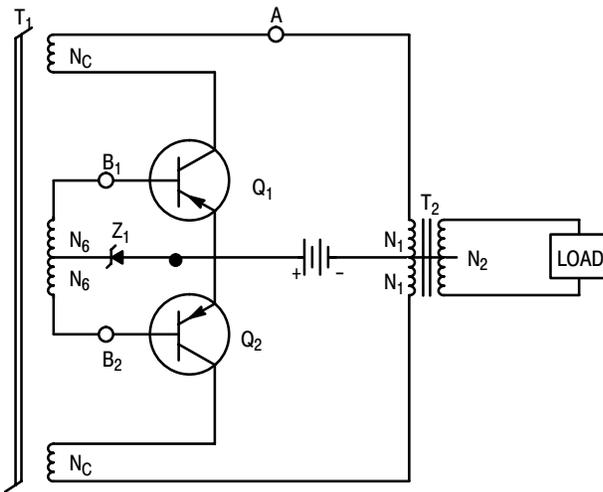


Figure 1. Frequency Controlled Current Feedback Inverter

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 2N3791 transistors (operating from a 12 volt supply), exhibited frequency regulation of $\pm 2\%$ with $\pm 25\%$ 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.

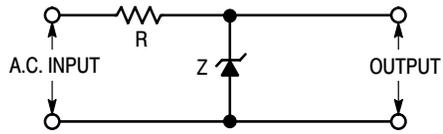
Transformer T_2 should not saturate at the lowest inverter frequency.

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

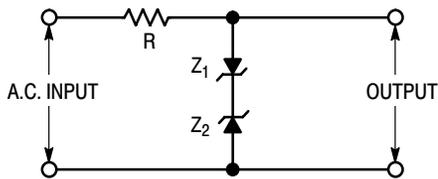
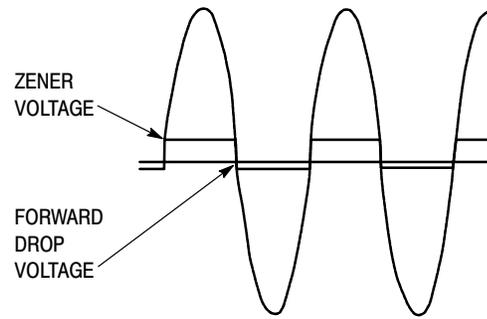
SIMPLE SQUARE WAVE GENERATOR

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

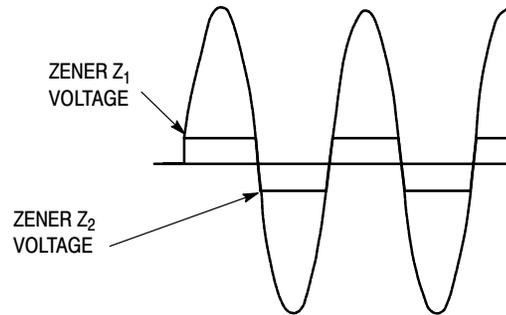


Figure 2. Zener Diode Square Wave Generator

TRANSIENT VOLTAGE SUPPRESSION

INTRODUCTION

Electrical transients in the form of voltage surges have always existed in electrical distribution systems, and prior to the implementation of semiconductor devices, they were of minor concern. The vulnerability of semiconductors to lightning strikes was first studied by Bell Laboratories in 1961.¹ A later report tried to quantify the amount of energy certain semiconductors could absorb before they suffered latent or catastrophic damage from electrostatic discharge.² Despite these early warnings, industry did not begin to address the issue satisfactorily until the late 1970s. Listed below are the seven major sources of overvoltage.

- Lightning
- Sunspots
- Switching of Loads in Power Circuits
- Electrostatic Discharge
- Nuclear Electromagnetic Pulses
- Microwave Radiation
- Power Cross

Most electrical and electronic devices can be damaged by voltage transients. The difference between them is the amount of energy they can absorb before damage occurs. Because many modern semiconductor devices, such as small signal transistors and integrated circuits can be damaged by disturbances that exceed the voltage ratings at only 20 volts or so, their survivability is poor in unprotected environments.

In many cases, as semiconductors have evolved their ruggedness has diminished. The trend to produce smaller and faster devices, and the advent of MOSFET and gallium arsenide FET technologies has led to an increased vulnerability. High impedance inputs and small junction sizes limit the ability of these devices to absorb energy and to conduct large currents. It is necessary, therefore, to supplement vulnerable electronic components with devices specially designed to cope with these hazards. Listed below are the four primary philosophies for protecting against transients.

- Clamping, or “clipping” is a method of limiting the amplitude of the transient.
- Shunting provides a harmless path for the transient, usually to ground by way of an avalanche or a crowbar mechanism.
- Interrupting opens the circuit for the duration of the transient.
- Isolating provides a transient barrier between hostile environments and vulnerable circuits through the use of transformers or optoisolators.

Selection of the proper protective method should be made based upon a thorough investigation of the potential sources of the overvoltage hazard. Different applications and environments present different sources of overvoltage.

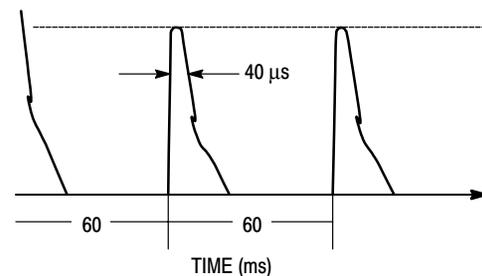
LIGHTNING

At any given time there are about 1800 thunderstorms in progress around the world, with lightning striking about 100 times each second. In the U.S., lightning kills about 150 people each year and injures another 250. In flat terrain with an average lightning frequency, each 300 foot structure will be hit, on average, once per year. Each 1200 foot structure, such as a radio or TV tower, will be hit 20 times each year, with strikes typically generating 600 million volts.

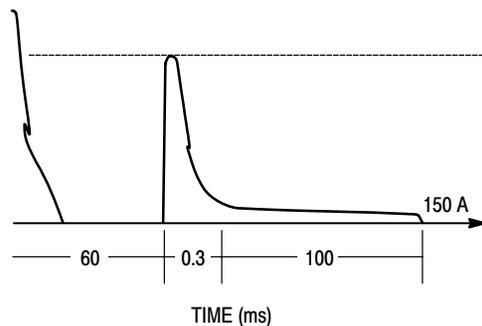
Each cloud-to-ground lightning flash really contains from three to five distinct strokes occurring at 60 ms intervals, with a peak current of some 20,000 amps for the first stroke and about half that for subsequent strokes. The final stroke may be followed by a continuing current of around 150 amps lasting for 100 ms.

The rise time of these strokes has been measured at around 200 nanoseconds or faster. It is easy to see that the combination of 20,000 amps and 200 ns calculates to a value of dI/dt of 10^{11} amps per second! This large value means that transient protection circuits must use RF design techniques, particularly considerate of parasitic inductance and capacitance of conductors.

While this peak energy is certainly impressive, it is really the longer-term continuing current which carries the bulk of the charge transferred between the cloud and ground. From various field measurements, a typical lightning model has been constructed, as shown in Figure 1.



Flash with No Continuing Current



Flash with Continuing Current

Figure 1. Typical Lightning Model, with and without Continuing Current

Depending on various conditions, continuing current may or may not be present in a lightning strike. A severe lightning model has also been created, which gives an indication of the strength which can be expected during worst case conditions at a point very near the strike location. Figure 2 shows this model. Note that continuing current is present at more than one interval, greatly exacerbating the damage which can be expected. A severe strike can be expected to ignite combustible materials.

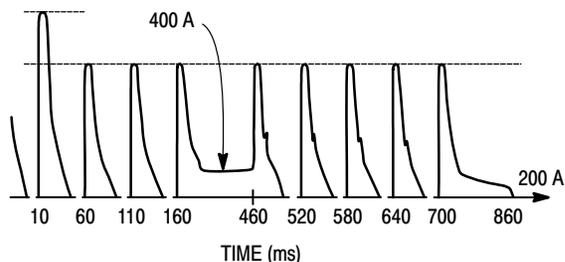


Figure 2. Severe Lightning Model

A direct hit by lightning is, of course, a dramatic event. In fact, the electric field strength of a lightning strike nearby may be enough to cause catastrophic or latent damage to semiconductor equipment. It is a more realistic venture to try to protect equipment from these nearby strikes than to expect survival from a direct hit.

With this in mind, it is important to be able to quantify the induced voltage as a function of distance from the strike. Figure 3 shows that these induced voltages can be quite high, explaining the destruction of equipment from relatively distant lightning flashes.

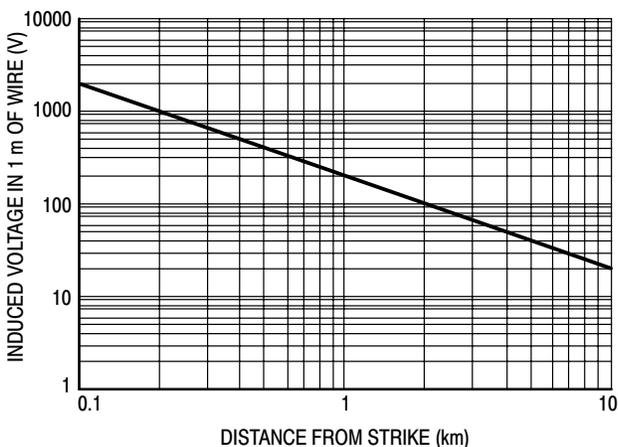


Figure 3. Voltage Induced by Nearby Lightning Strike

Burying cables does not provide appreciable protection as the earth is almost transparent to lightning radiated fields. In fact, underground wiring has a higher incidence of strikes than aerial cables.³

SUNSPOTS

The sun generates electromagnetic waves which can disrupt radio signals and increase disturbances on residential and business power lines. Solar flares, which run in cycles of 11 years (1989 was a peak year) send out electromagnetic waves which disrupt sensitive equipment.

Although not quantified, the effects of sunspot activity should be considered. Sunspots may be the cause of sporadic, and otherwise unexplainable problems in such sensitive areas.

SWITCHING OF LOADS IN POWER CIRCUITS

Inductive switching transients occur when a reactive load, such as a motor or a solenoid coil, is switched off. The rapidly collapsing magnetic field induces a voltage across the load's winding which can be expressed by the formula:

$$V = -L (di/dt)$$

where L is inductance in henrys and di/dt is the rate of change of current in amps per second.

Such transients can occur from a power failure or the normal opening of a switch. The energy associated with the transient is stored within the inductance at power interruption and is equal to:

$$W = 1/2 L i^2$$

where W is energy in joules and i is instantaneous current in amps at the time of interruption.

As an example, a 1.4 to 2.5 kV peak transient can be injected into a 120 vac power line when the ignition system of an oil furnace is fired. It has also been shown that there are transients present on these lines which can reach as high as 6 kV. In locations without transient protection devices, the maximum transient voltage is limited to about 6 kV by the insulation breakdown of the wiring.

Inductive switching transients are the silent killers of semiconductors as they often occur with no outward indication. A graphic example is the report of a large elevator company indicating the failure of 1000 volt rectifiers during a power interruption. In another area, power interruption to a 20 HP pump motor in a remote area was directly related to failure of sensitive monitoring equipment at that same site.⁴

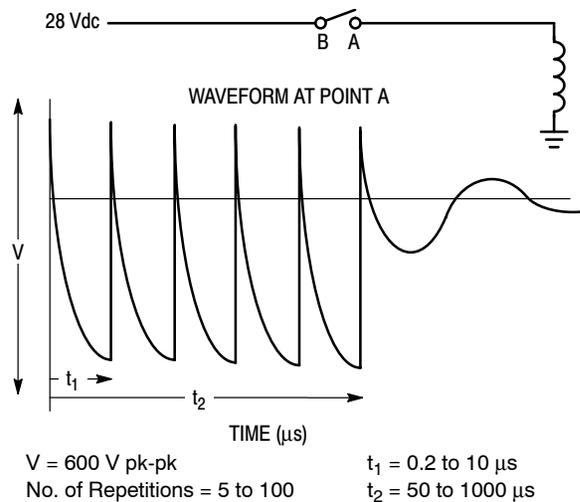


Figure 4. Switching Transient Definition for Aircraft and Military Buses, per Boeing Document D6-16050

After characterizing electrical overstress on aircraft power buses, Boeing published **Document D6-16050** as shown in Figure 4.

The military has developed switching transient definitions within several specifications including:

DOD-STD-1399 for shipboard

MIL-STD-704 for aircraft

MIL-STD-1275 for ground vehicles

The International Electrotechnical Commission (IEC) is now promoting their specification **IEC 801-4** throughout the European community. This describes an inductive switching transient voltage threat having 50 ns wide spikes with amplitudes from 2 kV to 4 kV occurring in 300 ms wide bursts.⁵

Besides these particular military specifications, many are application specific and functional tests exist. A supplier of transient voltage suppressor components will be expected to perform to a wide variety of them.

ELECTROSTATIC DISCHARGE (ESD)

ESD is a widely recognized hazard during shipping and handling of many semiconductor devices, especially those that contain unprotected MOSFETs, semiconductors for use at microwave frequencies and very high speed logic with switching times of 2 ns or less. In response to this threat, most semiconductors are routinely shipped in containers made of conductive material.

In addition to various shipping precautions, electronic assembly line workers should be grounded, use grounded-tip soldering irons, ionized air blowers and other techniques to prevent large voltage potentials to be generated and possibly discharged into the semiconductors they are handling.

Once the assembled device is in normal operation, ESD damage can still occur. Any person shuffling his feet on a carpet and then touching a computer keyboard can possibly cause a software crash or, even worse, damage the keyboard electronics.

The electrical waveform involved in ESD is a brief pulse, with a rise time of about 1 ns, and a duration of 100–300 μs . The peak voltage can be as large as 30 kV in dry weather, but is more commonly 0.5–5.0 kV.⁶ The fastest rise times occur from discharges originating at the tip of a hand-held tool, while discharges from the finger tip and the side of the hand are slightly slower.⁷ A typical human with a body capacitance of 150 pF, charged to 3 microcoulombs, will develop a voltage potential of 20 kV, according to the formula:

$$V = Q / C$$

where V is voltage, Q is charge and C is capacitance. The energy delivered upon discharge is:

$$W = 1/2 CV^2$$

where W is energy in joules, C is capacitance and V is voltage.

It is interesting to note that most microcircuits can be destroyed by a 2500 volt pulse, but a person cannot feel a static spark of less than 3500 volts!

NUCLEAR ELECTROMAGNETIC PULSES (NEMP)

When a nuclear weapon is detonated, a very large flux of photons (gamma rays) is produced. These rays act to produce an electromagnetic field known as a nuclear electromagnetic pulse or NEMP. When a nuclear detonation occurs above the atmosphere, a particularly intense pulse illuminates all objects on the surface of the earth, and all objects in the lower atmosphere within line of sight of the burst. A burst 300–500 km above Kansas would illuminate the entire continental U.S.

A typical NEMP waveform is a pulse with a rise time of about 5 ns and a duration of about 1 μs . Its peak electric field is 50–100 kV/m at ground level. After such a pulse is coupled into spacecraft, aircraft and ground support equipment, it produces a waveform as described in **MIL-STD-461C**. The insidious effect of NEMP is its broad coverage and its potential for disabling military defense systems.

MICROWAVE RADIATION

Microwaves can be generated with such high power that they can disable electronic hardware upon which many military systems depend. A single pulse flux of 10^{-8} MJ/cm^2 burns out receiver diodes, and a flux of 10^{-4} MJ/cm^2 causes bit errors in unshielded computers.⁸ With automobiles utilizing MPU controls in more applications, it is important to protect against the effects of driving by a microwave transmitter. Likewise, a nearby lightning strike could also have detrimental effects to these systems.

POWER CROSS

Yet another source of electrical overstress is the accidental connection of signal lines, such as telephone or cable television, to an ac or dc power line. Strictly speaking, this phenomenon, known as a power cross, is a continuous state, not a transient. However, the techniques for ensuring the

survival of signal electronics after a power cross are similar to techniques used for protection against transient overvoltages.

STANDARDIZED WAVEFORMS

Fortunately, measurements of these hazards have been studied and documented in several industry specifications. For example, Bellcore **Technical Advisory TA-TSY-000974** defines the generic measurement waveform for any double exponential waveform, which is the basis for most of the specific applications norms.

The predominant waveform for induced lightning transients, set down by Rural Electrification Administration **Document PE-60**, is shown in Figure 5. This pulse test, performed at the conditions of 100 V/ μ s rise, 10/1000 μ s, $I_p = 1$ kV, is one of the two most commonly specified in the industry. The other is the 8/20 μ s waveform, shown in Figure 6.

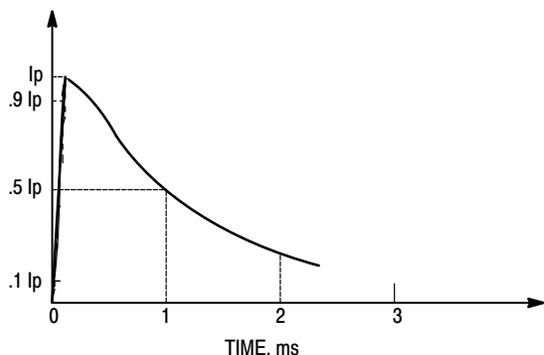


Figure 5. Pulse Waveform (10/1000 μ s)

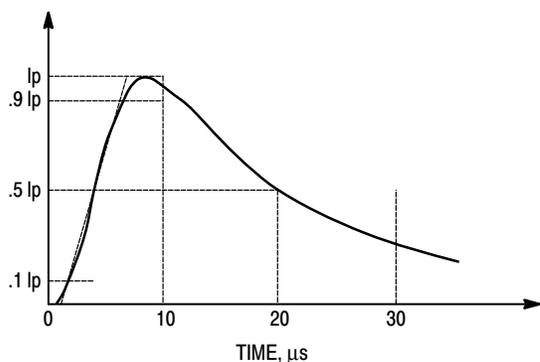


Figure 6. Pulse Waveform (8/20 μ s)

TRANSIENT VOLTAGE SUPPRESSION AND TELECOM

TRANSIENT VOLTAGE SURGE SUPPRESSION COMPONENTS ON DATA AND TELEPHONE LINES

Lines carrying data and telephone signals are subject to a number of unwanted and potentially damaging transients primarily from two sources: lightning and “power crosses.” A power cross is an accidental connection of a signal line to a powerline. Transients from lightning can impress voltages well above a kilovolt on the line but are of short duration, usually under a millisecond. Lightning transients are suppressed by using Transient Voltage Surge Suppressor

(TVS) devices. TVS devices handle high peak currents while holding peak voltage below damaging levels, but have relatively low energy capability and cannot protect against a power cross fault. The first TVS used by telephone companies is the carbon block, but its peak let-through voltage was too high for modern equipment using unprotected solid state circuitry. A number of other components fill today’s needs.

The power cross condition causes a problem with telephone lines. Fast acting fuses, high speed circuit breakers and positive temperature coefficient thermistors have been successfully used to limit or interrupt current surges exceeding a millisecond.

Over the years, telecommunications switching equipment has been transitioning from electromechanical relays to integrated circuits and MOSFET technology. The newer equipment operates at minimal electrical currents and voltages, which make it very efficient. It is therefore quite sensitive to electrical overloads caused by lightning strikes and other transient voltage sources, and by power crosses.

Because of the deployment of new technology, both in new installations and in the refurbishment of older systems, the need for transient protection has grown rapidly. It is widely recognized that any new equipment must include protection devices for reasons of safety, reliability and long term economy.

The major telecom companies, in their never ending quest for the elimination of electromechanical technology have been looking at a number of novel methods and implementations of protection. These methods provide for solutions to both the primary and secondary protection categories.

A number of studies have been conducted to determine the transient environment on telephone lines. Very little has been done with data lines because a typical situation does not exist. However, information gathered from telephone line studies can serve as a guide for data lines.

Past studies on telephone lines coupled with the high current capability of arc type arrestors and the conservative nature of engineers seem to have produced specifications which far exceed the real need. A recent study by Bell South Services⁹ reported that the highest level of transient energy encountered was well below standards and specifications in common use. Now, solid state devices perform adequately for many applications. However the stringent specifications of some regulatory agencies promote arc-type arrestors, though solid state devices would be a better choice.

PRIMARY PROTECTION

Primary protection is necessary to protect against high voltage transients which occur in the outdoor environment. These transients include induced lightning surges and arc cross conditions.

As such, primary protection is located at the point where wiring enters the building or terminal box. It is the first line defense against outside hazards. TVS devices located where lines enter a building are called primary protectors.

Protectors connected to indoor lines are referred to as secondary protectors. Both primary and secondary protectors are required to provide complete equipment protection.

Today, primary protection is most generally accomplished through the use of surge protector modules. For telecom, these are designed specifically for the environment and the standards dictated by the telecom applications. They typically contain a two or three element gas arrestor tube and a mechanically-triggered heat coil. Some also include air gap carbon block arrestors which break over at voltages above about 1500 volts.

Some modules contain high speed diodes for clamp response in the low nanoseconds. This provides protection until the gas tube fires, generally in about one microsecond. The diodes may be connected between the tip, ring and ground conductors in various combinations. The 5ESS electronic switching system norms dictate design and performance requirements of TVS modules in use today. Test methods are spelled out in **REA PE-80**, a publication of the Rural Electrification Administration.

In the U.S. alone, 58 million primary protection modules are sold annually, about 40 million for central offices and 18 million for station locations, such as building entrances. Eighty percent of these use gas tubes, 16% use air-gap carbon blocks, and only 2% (so far) are solid state.

SECONDARY PROTECTION

Secondary protection is necessary for the equipment inputs, and as such, is located between the primary protector and the equipment. Secondary protection is generally accomplished with one or more TVS components, as opposed to the modules used for primary. It is often placed on a circuit board along with other components handling other duties, such as switching. Secondary protection is applied to lines associated with long branch circuits which have primary protection a significant distance away, to internal data lines, and to other locations requiring additional local hazard-proofing.

While not as open to external transients as the primary, secondary can still see peak open circuit voltages in excess of 1000 volts and short circuit currents of hundreds of amps. These transients may be locally generated, or they may be residuals from the primary protectors upstream.

STANDARDS

Transient voltage waveforms are commonly described in terms of a dual exponential wave as defined in Figure 7. The standard chosen for power lines is a 1.2/50 μ s voltage wave which causes an 8/20 μ s current wave. Although the source of the most severe transients on telecom lines is the same as for power lines and lightning, the higher impedance per unit length of the telephone line stretches the waves as they propagate through the lines.

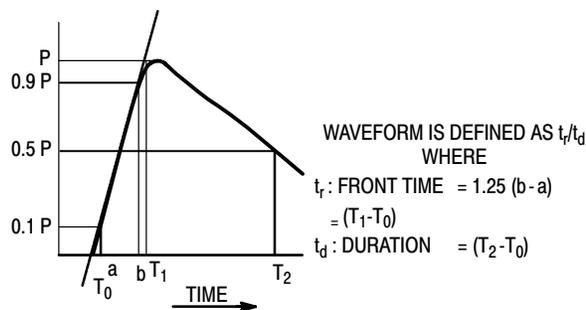


Figure 7. Definition of Double Exponential Impulse Waveform

The 10/1000 μ s wave approximates the worst case waveform observed on data and telecom lines. TVS devices intended for this service are usually rated and characterized using a 10/1000 waveform. The Bell South study revealed that the worst transient energy handled by primary protectors on lines entering a central office was equivalent to only 27 A peak of a 10/1000 wave. This level is considerably less than that required by secondary protectors in most of the standards in use today. This finding is particularly significant because the Bell South service area includes Central Florida, the region experiencing the highest lightning activity in the U.S.

The United States Federal Communications Commission (FCC) has defined mandatory requirements for equipment which is to be connected to the U.S. telephone network. In some cases, U.S. equipment must meet standards developed by the Rural Electrification Agency (REA). Many nations demand compliance to standards imposed by the Consultative Committee, International Telegraph and Telephone (CCITT). In addition, most equipment users demand safety certification from U.L., which has its own standards.

The FCC Standards are based on a worst case residue from a carbon block primary protector installed where the phone line enters the building. The CCITT standard is applicable for situations lacking primary protection, other than wiring flashover. Companies entering the telephone equipment or protector market will need to obtain and become familiar with the appropriate governing standards.

TRANSIENT VOLTAGE PROTECTION COMPONENTS

GENERAL TVS CHARACTERISTICS

A number of transient voltage suppressor (TVS) devices are available. Each finds use in various applications based upon performance and cost. All types are essentially transparent to the line until a transient occurs; however, some devices have significant capacitance which loads the line for ac signals. A few of these are described in Table 1.

Based upon their response to an overvoltage, TVS devices fit into two main categories, clamps and crowbars. A clamp conducts when its breakdown voltage is exceeded and reverts back to an open circuit when the applied voltage

drops below breakdown. A crowbar switches into a low voltage, low impedance state when its breakover voltage is exceeded and restores only when the current flowing through it drops below a “holding” level.

CLAMP DEVICES

All clamp devices exhibit the general V-I characteristic of Figure 8. There are variations; however, some clamps are asymmetric. In the non clamping direction, some devices such as the zener TVS exhibit the forward characteristic of a diode while others exhibit a very high breakdown voltage and are not intended to handle energy of “reverse” polarity. Under normal operating conditions, clamp devices appear virtually as an open circuit, although a small amount of leakage current is usually present. With increasing voltage a point is reached where current increases rapidly with voltage as shown by the curved portion of Figure 8. The rapidly changing curved portion is called the “knee region.” Further increases in current places operation in the “breakdown” region.

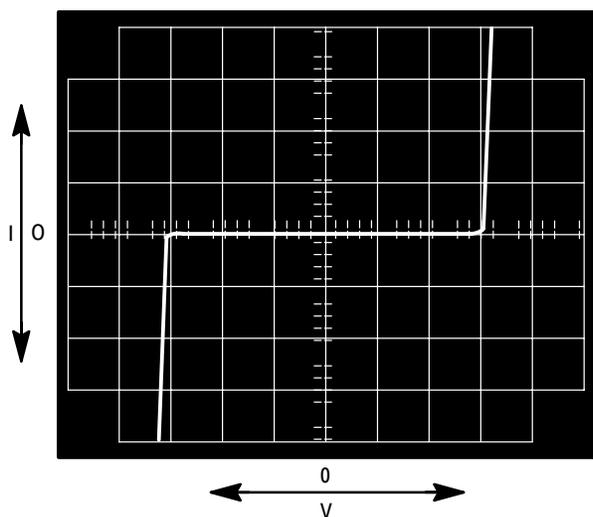


Figure 8. Static Characteristics of a Clamp Device

In the knee region the V-I characteristic of clamping devices can be approximated by the equation:

$$I = K V^s \quad (1)$$

where K is a constant of proportionality and s is an exponent which defines the “sharpness factor” of the knee. The exponent s is 1 for a resistor and varies from 5 to over 100 for the clamping devices being used in TVS applications. A high value of s i.e., a sharp knee, is beneficial. A TVS device can be chosen whose breakdown voltage is just above the worst case signal amplitude on the line without concern of loading the line or causing excessive dissipation in the TVS.

As the current density in the clamp becomes high, the incremental resistance as described by Equation 1 becomes very small in comparison to the bulk resistance of the material. The incremental resistance is therefore ohmic in the high current region.

Unfortunately, a uniform terminology for all TVS devices has not been developed; rather, the terms were developed in conjunction with the appearance of each device in the marketplace. The key characteristics normally specified define operation at voltages below the knee and at currents above the knee.

Leakage current is normally specified below the knee at a voltage variously referred to as the stand-off voltage, peak working voltage or rated dc voltage. Some devices are rated in terms of an RMS voltage, if they are bidirectional. Normal signal levels must not exceed this working voltage if the device is to be transparent.

Breakdown voltage is normally specified at a fairly low current, typically 1 mA, which places operation past the knee region. Worst case signal levels should not exceed the breakdown voltage to avoid the possibility of circuit malfunction or TVS destruction.

The voltage in the high current region is called the clamping voltage, V_C . It is usually specified at the maximum current rating for the device. To keep V_C close to the breakdown voltage, s must be high and the bulk resistance low. A term called clamping factor, (F_C) is sometimes used to describe the sharpness of the breakdown characteristic. F_C is the ratio of clamping voltage to the breakdown voltage. As the V-I characteristic curve of the TVS approaches a right angle, the clamping factor approaches unity. Clamping factor is not often specified, but it is useful to describe clamp device behavior in general terms.

Table 1. Comparison of TVS Components

Type	Protection Time	Protection Voltage	Power Dissipation	Reliable Performance	Expected Life	Other Considerations
GAS TUBE	> 1 μ s	60–100 V	Nil	No	Limited	Only 50–2500 surges. Can short power line.
MOV	10–20 ns	> 300 V	Nil	No	Degrades	Fusing required. Degrades. Voltage level too high.
AVALANCHE TVS	50 ps	3–400 V	Low	Yes	Long	Low power dissipation. Bidirectional requires dual.
THYRISTOR TVS	< 3 ns	30–400 V	Nil	Yes	Long	High capacitance. Temperature sensitive.

Clamp devices generally react with high speed and as a result find applications over a wide frequency spectrum. No delay is associated with restoration to the off state after operation in the breakdown region.

CROWBAR DEVICES

Crowbar TVS devices have the general characteristics shown in Figure 9. As with clamp devices, asymmetric crowbars are available which may show a diode forward characteristic or a high voltage breakdown in one direction.

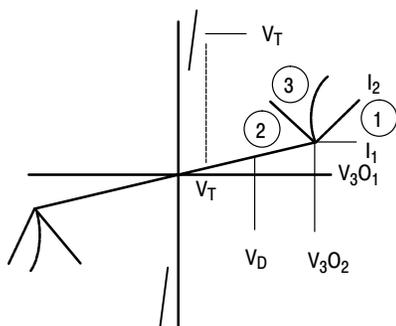


Figure 9. Static Characteristics of a Bidirectional Crowbar Device

The major difference between a crowbar and a clamp is that, at some current in the breakdown region, the device switches to a low voltage on-state. In the clamping region from I_1 to I_2 , the slope of the curve may be positive as shown by segment 1, negative (segment 2) or exhibit both characteristics as shown by segment 3. A slightly positive slope is more desirable than the other two curves because a negative resistance usually causes a burst of high frequency oscillation which may cause malfunction in associated circuitry. However, a number of performance and manufacturing trade-offs affect the shape of the slope in the clamping region.

A crowbar TVS has an important advantage over a clamp TVS in that it can handle much larger transient surge current densities because the voltage during the surge is considerably lower. In a telephone line application, for example, the clamping level must exceed the ring voltage peak and will typically be in the vicinity of 300 volts during a high current surge. The on-state level of the crowbar may be as low as 3 volts for some types which allows about two orders of magnitude increase in current density for the same peak power dissipation.

However, a crowbar TVS becomes “latched” in the on-state. In order to turn off its current flow the driving voltage must be reduced below a critical level called the holding or extinguishing level. Consequently, in any application where the on-state level is below the normal system voltage, a follow-on current occurs. In a dc circuit crowbars will not turn off unless some means is provided to interrupt the current. In an ac application crowbars will turn off near the zero crossing of the ac signal, but a time delay is associated with turn-off which limits crowbars to

relatively low frequency applications. In a data line or telecom application the turn-off delay causes a loss of intelligence after the transient surge has subsided.

A telephone line has both ac and dc signals present. Crowbars can be successfully used to protect telecom lines from high current surges. They must be carefully chosen to ensure that the minimum holding current is safely above the maximum dc current available from the lines.

TVS DEVICES

A description of the various types of TVS devices follows in the chronological order in which they became available. Used appropriately, sometimes in combination, any transient protection problem can be suitably resolved. Their symbols are shown in Figure 10.

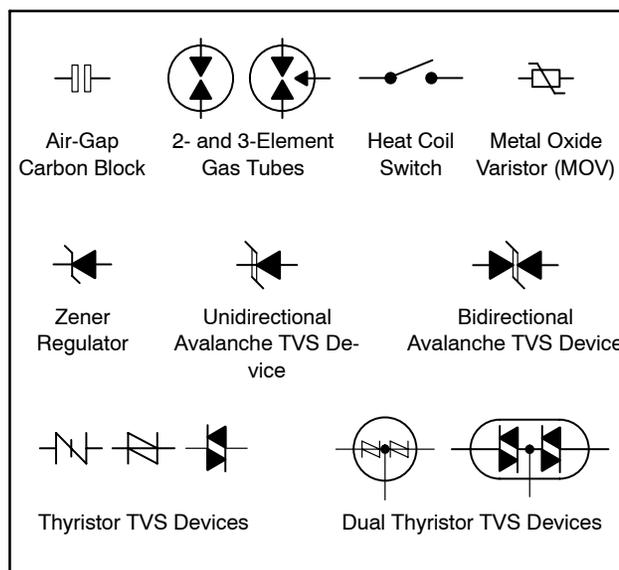


Figure 10. TVS Devices and Their Symbols

AIR GAP ARRESTORS

The air gap is formed by a pair of metal points rigidly fixed at a precise distance. The air ionizes at a particular voltage depending upon the gap width between the points. As the air ionizes breakover occurs and the ionized air provides a low impedance conductive path between the points.

The breakover threshold voltage is a function of the air’s relative humidity; consequently, open air gaps are used mainly on high voltage power lines where precise performance is not necessary. For more predictable behavior, air gaps sealed in glass and metal packages are also available.

Because a finite time is required to ionize the air, the actual breakover voltage of the gap depends upon the rate of rise of the transient overvoltage. Typically, an arrestor designed for a 120 V ac line breaks over at 2200 volts.

Air gaps handle high currents in the range of 10,000 amperes. Unfortunately, the arc current pits the tips which causes the breakover voltage and on-state resistance to increase with usage.

CARBON BLOCK ARRESTORS

The carbon block arrestor, developed around the turn of the century to protect telephone circuits, is still in place in many older installations. The arrestor consists of two carbon block electrodes separated by a 3 to 4 mil air gap. The gap breaks over at a fairly high level – approximately 1 kV – and cannot be used as a sole protection element for modern telecom equipment. The voltage breakdown level is irregular. With use, the surface of the carbon block is burned which increases the unit's resistance. In addition, the burned material forms carbon tracks between the blocks causing a leakage current path which generates noise. Consequently, many of the carbon blocks in service are being replaced by gas tubes and are seldom used in new installations.

SILICON CARBIDE VARISTORS

The first non-linear resistor to be developed was called a “varistor.” It was made from specially processed silicon carbide and found wide use in high power, high voltage TVS applications. It is not used on telecom lines because its clamping factor is too high: s is only about 5.

GAS SURGE ARRESTORS

Gas surge arrestors are a sophisticated modification of the air gap more suited to telecom circuit protection. Most often used is the “communication” type gap which measures about 3/8 inch in diameter and 1/4 inch thick. A cross section is shown in Figure 11. They consist of a glass or ceramic envelope filled with a low pressure inert gas with specialized electrodes at each end. Most types contain a minute quantity of radioactive material to stabilize breakover voltage. Otherwise, breakover is sensitive to the level of ambient light.

Because of their small size and fairly wide gap, capacitance is very low, only a few picofarads. When not activated, their off-state impedance or insulation resistance is virtually infinite.

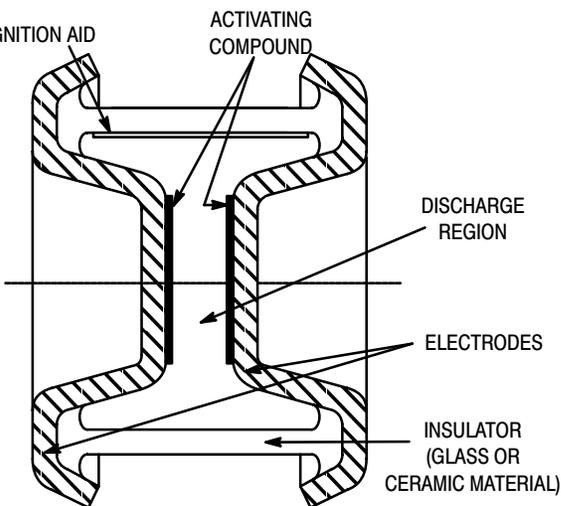


Figure 11. Gas Arrestor Cross Section

Key electrical specifications for this TVS type include breakover voltage (dc & pulse), maximum holdover voltage, arc voltage, and maximum surge current.

The breakover voltage is rated at a slow rate of rise, 5000 V/s, essentially dc to a gas arrestor. Typical dc voltage ratings range from 75 V through 300 V to provide for most communication systems protection requirements. The maximum pulse voltage rating is that level at which the device fires and goes into conduction when subjected to a fairly rapid rate of voltage rise, (dv/dt) usually 100 V/ μ s. Maximum rated pulse voltages typically range from 400 V to 600 V, depending on device type.

A typical waveform of a gas surge arrestor responding to a high voltage pulse is shown in Figure 12. From the waveform, it can be seen that the dv/dt of the wave is 100 V/ μ s and the peak voltage (the breakover voltage) is 520 V.

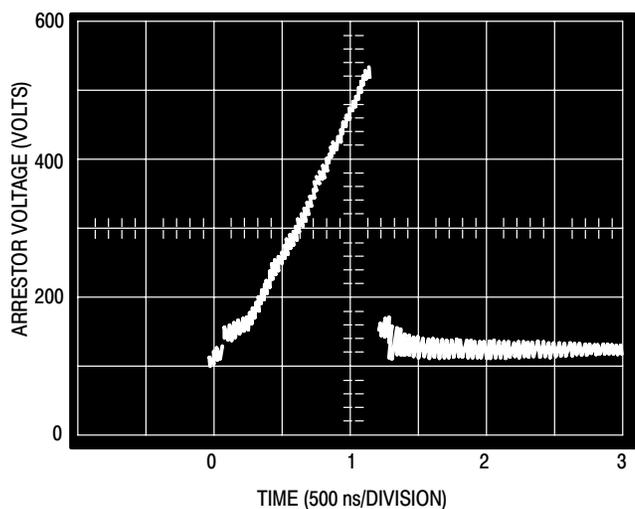


Figure 12. Voltage Waveform of Gas Surge Arrestor Responding to a Surge Voltage

Gas surge arrestors fire faster but firing voltage increases as the transient wave fronts increase in slope as illustrated in Figure 13. The near vertical lines represent the incident transient rise time. Note that the response time is greater than 0.1s at slow rise times but decreases to less than 0.1 μ s for risetimes of 20 kV/ μ s. However, the firing voltage has increased to greater than 1000 V for the gas tube which breaks over at 250 Vdc.

The driving circuit voltage must be below the holdover voltage for the gap to extinguish after the transient voltage has passed. Holdover voltage levels are typically 60% to 70% of the rated dc breakdown voltage.

Arc voltage is the voltage across the device during conduction. It is typically specified at 5 to 10 V under a low current condition, but can exceed 30 V under maximum rated pulse current.

The maximum surge current rating for a 8/20 μ s waveform is typically in the 10 kA to 20 kA range for communication type devices. For repetitive surges with a 10/1000 wave, current ratings are typically 100 A, comfortably above the typical exposure levels in a telephone subscriber loop.

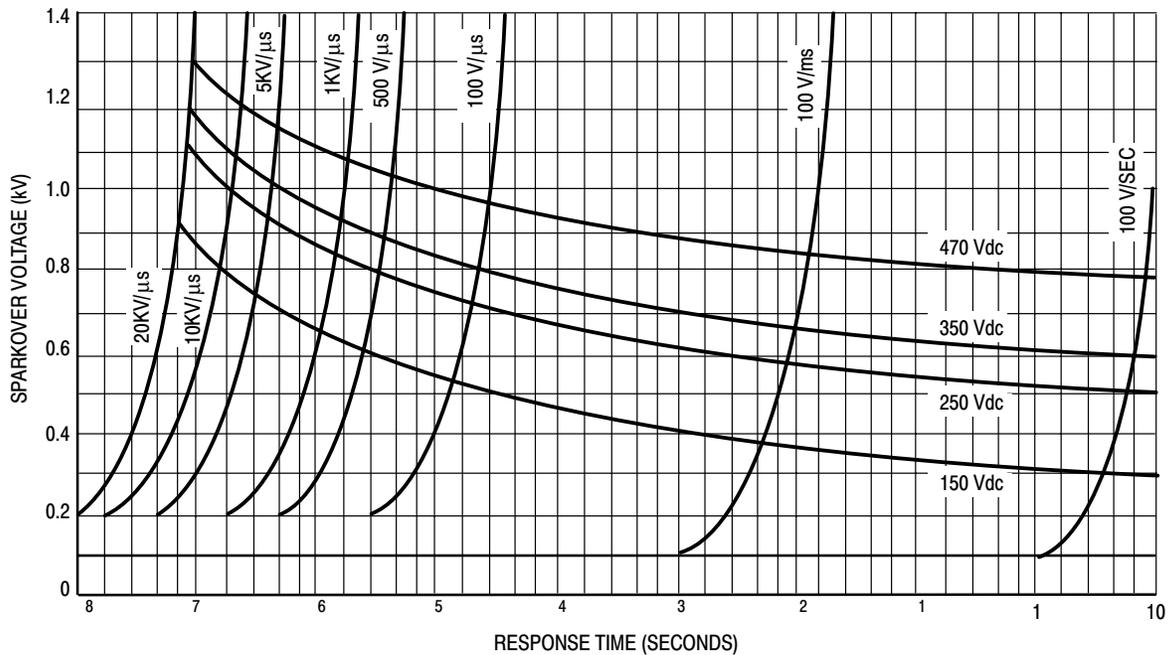


Figure 13. Typical Response Time of a Gas Surge Arrestor

Gas tubes normally provide long life under typical operating conditions, however; wear-out does occur. Wear-out is characterized by increased leakage current and firing voltage. An examination of gas tubes in service for six to eight years revealed that 15% were firing outside of their specified voltage limits.⁹ Because firing voltage increases with use, protectors often use an air gap backup in parallel with the gas tube. End-of-life is often specified by manufacturers as an increase of greater than 50% of breakover or firing voltage. Other limits include a decrease in leakage resistance to less than 1 mW.

The features and limitations of gas tube surge arrestors are listed below.

Advantages:

- High current capability
- Low capacitance
- Very high off-state impedance

Disadvantages:

- Slow response time
- Limited life
- High let-through voltage
- Open circuit failure mode

Principally because of their high firing voltage, gas surge arrestors are not suitable for use as the sole element to protect modern equipment connected to a data or telecom line. However, they are often part of a protection network where they are used as the primary protector at the building interface with the outside world.

SELENIUM CELLS

Polycrystalline diodes formed from a combination of selenium and iron were the forerunners of monocrystalline

semiconductor diodes. The TVS cells are built by depositing the polycrystalline material on a metal plate to increase their thermal mass thereby raising energy dissipation. The cells exhibit typical diode characteristics and a non-linear reverse breakdown which is useful for transient suppression. Cells can be made which are “self-healing”; that is, the damage which occurs when subjecting them to excessive transient current is repaired with time.

Selenium cells are still used in high power ac line protection applications because of their self-healing characteristic; however, their high capacitance and poor clamping factor ($s \gg 8$) rule them out for data or telecom line applications.

METAL OXIDE VARISTORS

The metal oxide varistor (MOV) is composed of zinc oxide granules in a matrix of bismuth and other metal oxides. The interface between the zinc oxide and the matrix material exhibits characteristics similar to that of a p-n junction having a voltage breakdown of about 2.6 V. With this structure the electrical equivalent is that of groups of diodes in parallel which are stacked in series with similar parallel groups to provide the desired electrical parameters. The taller the stack, the higher the breakdown and operating voltage. Larger cross-sections provide higher current capability. The structure of an MOV is shown in Figure 14.

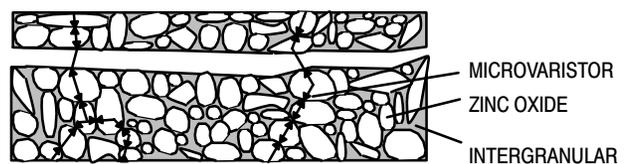
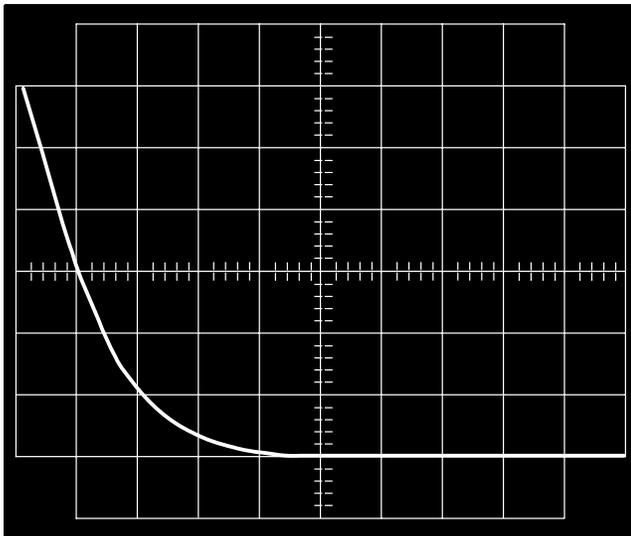


Figure 14. MOV Cross Section



MOV (27 V)
 Vert: 10 V/div
 Horiz: 0.5 ms/div
 Transient Source Impedance: 0.55 Ω
 $V_{\text{peak}} = 62.5 \text{ V}$

Figure 15. MOV Clamping Voltage Waveform

MOVs, formed from a ceramic-like material, are usually produced in the shape of discs with most widely used MOVs having diameters of 7 mm, 14 mm, and 20 mm. The disc surfaces are coated with a highly conductive metal such as silver to assure uniform conduction through the cross sectional area of the device. After terminal attachment the parts are coated with a durable plastic material.

The typical voltage spectrum of MOVs ranges from 8 V through 1000 V for individual elements. Pulse current capability (8/20 μs) ranges from a few amperes to several thousands of amperes depending on the element's size. The V-I characteristic of MOVs is similar to Figure 8. Their clamping factor is fairly good; s is in the vicinity of 25.

Key electrical specifications include: operating voltage, breakdown voltage, peak current maximum clamping voltage, and leakage current.

The maximum operating voltage specified is chosen to be below the breakdown voltage by a margin sufficient to produce negligible heating under normal operating conditions. Breakdown voltage is the transition point at which a small increase in voltage results in a significant increase in current producing a clamping action. Maximum limits for breakdown voltage are typically specified at 1 mA with upper end limits ranging from 20% to 40% greater than the minimum breakdown voltage.

Maximum peak current is a function of element area and ranges from tens of amperes to tens of thousands of amperes. MOVs are typically pulse rated with an 8/20 μs waveform since they are intended primarily for use across power lines.

The clamping characteristics of a 27 V ac rated MOV, with a 4 joule maximum pulse capability is shown in Figure 15. The transient energy is derived from an exponentially decreasing pulse having a peak amplitude of 90 V. The pulse generator source impedance is 0.55 Ω. Peak clamping voltage is 62.5 V while the developed current is 50 A. The clamping factor calculates to be 2.3.

Leakage currents are listed for MOVs intended for use in sensitive protection applications but are not normally listed

for devices most often used on power lines. Leakage current behavior is similar to that of a p-n junction. It roughly doubles for every 10°C increase in temperature and also shows an exponential dependence upon applied voltage. At a voltage of 80% of breakdown, leakage currents are several microamperes at a temperature of 50°C.

Although the theory of MOV operation is not fully developed, behavior is similar to a bidirectional avalanche diode. Consequently its response time is very fast.

Life expectancy is an important characteristic generated under pulse conditions. A typical example is shown in Figure 16. The data applies to 20 mm diameter disc types having rated rms voltage from 130 V to 320 V. Lifetime rating curves are usually given for each device family.

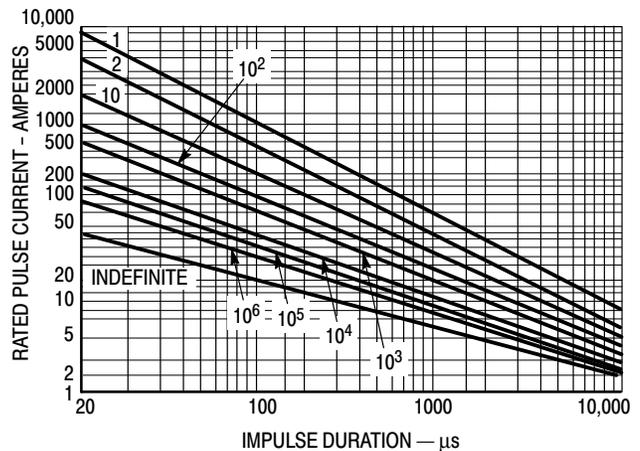


Figure 16. MOV Pulse Life Curve

For a single 8/20 μs pulse, the device described in Figure 16 is rated at 6500 A; however, it must be derated by more than two orders of magnitude for large numbers of pulses. Longer duration pulses also require further derating. For example, for a 10/1000 μs duration pulse, this family of devices has a maximum pulse rating of about 100 A on a single shot basis and devices must be derated to less than 10A for long lifetimes in excess of 100,000 pulses.

End-of-life for an MOV is defined as the voltage breakdown degrading beyond the limits of $\pm 10\%$. As MOVs are pulsed, they degrade incrementally as granular interfaces are overheated and changed to a highly conductive state. Failure occurs in power line applications when the breakdown voltage has degraded to the point where the MOV attempts to clip the powerline peaks. In telecom applications, their breakdown must be above the peaks of the impressed ac line during a ring cycle or a power cross; otherwise an immediate catastrophic failure will occur.

When MOVs fail catastrophically they initially fail short. However, if a source of high energy is present as might occur with a power cross, the follow-on current may cause the part to rupture resulting in an open circuit.

The advantages and shortcomings of using an MOV for general purpose protection in microprocessor based circuitry include the following:

Advantages:

- High current capability
- Broad voltage spectrum
- Broad current spectrum
- Fast response
- Short circuit failure mode

Disadvantages:

- Gradual decrease of breakdown voltage
- High capacitance

The capacitance of MOVs is fairly high because a large device is required in order to achieve a low clamping factor; consequently, they are seldom used across telecom lines.

ZENER TVS

Zener TVS devices are constructed with large area silicon p-n junctions designed to operate in avalanche and handle much higher currents than their cousins, zener voltage regulator diodes. Some manufacturers use small area mesa chips with metal heatsinks to achieve high peak power capability. However, ON Semiconductor has determined that large area planar die produce lower leakage current and clamping factor. The planar construction cross section is shown in Figure 17 and several packages are shown in Figure 18.

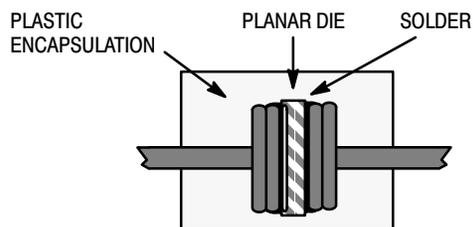


Figure 17. Zener TVS Cross-Section

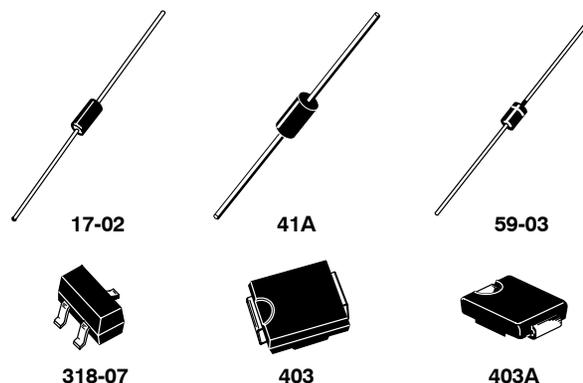


Figure 18. Typical Insertion and Surface Mount Silicon TVS Packages — Zeners and Thyristors

Key electrical parameters include maximum operating voltage, maximum reverse breakdown voltage, peak pulse current, peak clamping voltage, peak pulse power, and leakage current.

The normal operating or working voltage is usually called the reverse standoff voltage in specification sheets. Devices are generally available over the range of 5 V through 250 V. Standoff voltage defines the maximum peak ac or dc voltage which the device can handle. Standoff voltage is typically 10% to 15% below minimum reverse breakdown voltage. A listing of TVS products available from ON Semiconductor is shown in Table 2.

Table 2. ON Semiconductor Zener TVS Series

DEVICE SERIES	V _Z RANGE	PULSE POWER RATING (100/1000 PULSE)	PACKAGE
*SA5.0A-SA170A	6.8-200	500 W	Axial
*P6KE6.8A - P6KE200A	6.8-200	600 W	Axial
*1.5KE6.8A - 1.5KE250A	6.8-250	1500 W	Axial
1SMB5.0AT3 - 1SMB170AT3	6.8-200	600 W	SMB
P6SMB6.8AT3 - P6SMB200T3	6.8-200	600 W	SMB
1SMC5.0AT3 - 1SMC78AT3	6.8-91	1500 W	SMC
1.5SMC6.8AT3 1.5SMC91AT3	6.8-91	1500 W	SMC
MMBZ15VDLT1	15	ESD Protection >15 kV	SOT-23

* Available in bidirectional configurations

The reverse breakdown voltage is specified at a bias level at which the device begins to conduct in the avalanche mode. Test current levels typically are 1 mA for diodes which breakdown above 10 V and 10 mA for lower voltage devices. Softening of the breakdown knee, that is, lower s , for lower voltage p-n junction devices requires a higher test current for accurate measurements of reverse breakdown voltage. Diodes that break down above 10 V display a very sharp knee; s is over 100.

Peak pulse current is the maximum upper limit at which the device is expected to survive. Silicon p-n junctions are rated for constant power using a particular transient waveform; consequently, current is a function of the peak clamping voltage. For example, a 6.8 V device handles about 28 times the pulse current that a 220 V device will withstand; however, both the 6.8 V and 220 V types dissipate the same peak power under the same pulse waveform conditions. Most Zener TVS diodes are rated for 10/1000 μ s waveform pulses which are common in the telecom industry.

The clamping voltage waveform of a 27 V Zener TVS having a 1.5 joule capability is illustrated in Figure 19. Its peak voltage is 30.2 V. The transient energy source is the same as applied to the MOV whose response is shown in Figure 10. However, the current through the Zener TVS is over 100 A, much higher than occurs with the MOV because the clamping voltage is significantly lower. Despite the higher pulse current, the Zener displays much better clamping action; its clamping factor is 1.1.

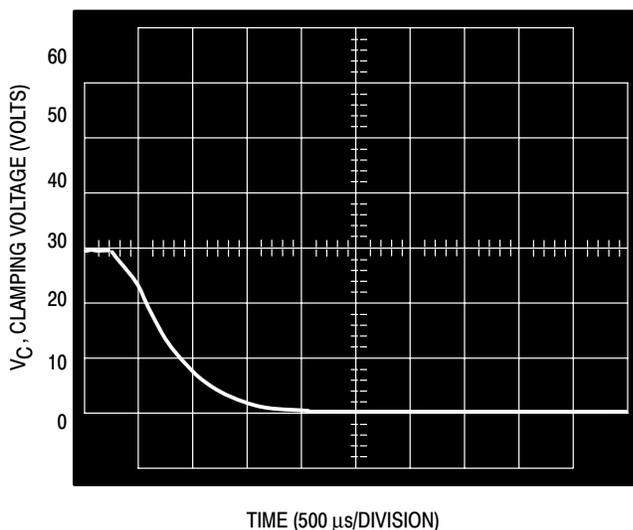


Figure 19. Zener TVS Clamping Voltage Waveform

Peak pulse power is the instantaneous power dissipated at the rated pulse condition. Common peak pulse power ratings are 500 W, 600 W, and 1500 W for 10/1000 μ s waveforms. As the pulse width decreases, the peak power capability increases in a logarithmic relationship. An example of a curve depicting peak pulse power versus pulse width is shown in Figure 20. The graph applies to the 1.5 kW series (10/1000 pulse) of TVS diodes and can be interpolated to determine power ratings over a broad range of pulse widths. At 50 μ s, the maximum rated power shown in the curve is

6 kW, which is four times greater than the rating at 1ms. The current handling capability is also increased roughly by this same factor of four.

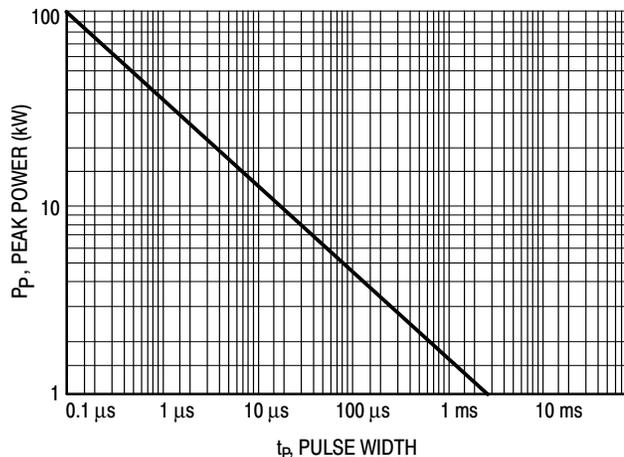


Figure 20. Peak Pulse Power Rating for a Popular Zener TVS Family

To increase power capability devices are stacked in series. For example, doubling the power capability requirement for a 100 V, 1.5 kW Zener TVS is easily done by placing two 50 V devices in series. Clamping factor is not significantly affected by this arrangement.

Although leakage current limits are relatively high for the industry low voltage types (500 μ A to 1000 μ A), dropping off to 5 μ A or less for voltages above 10 V, the planar die in use by ON Semiconductor exhibit considerably less leakage than the specified limits of the industry types.

Capacitance for the popular 1500 W family exceeds 10,000 pF at zero bias for a 6.8 V part, dropping exponentially to less than 100 pF for a 200 V device. Capacitance drops exponentially with a linear increase in bias. The capacitance of a 6.8 V device is 7000 pF, while the 200 V part measures under 60 pF, at their respective standoff voltages.

Capacitance loads the signal line at high frequencies. For high speed data transmission circuits, low capacitance is achieved by placing two diodes in a series stack as shown in Figure 21. Under normal operation the top diode (D_S) operates at essentially zero bias current. Since its power dissipation requirement is small, its area can be much smaller than that of the TVS diode (D_Z) in order to provide low capacitance. The top diode normally is not intended to be used in avalanche. Consequently if a negative voltage exceeding the reverse rating of the stack could occur, the low capacitance diode must be protected by another diode (D_P) shown connected by dotted lines on Figure 21. The arrangement of Figure 21 is satisfactory for situations where the signal on the line is always positive. When the signal is ac, diode D_P is replaced by another low capacitance stack, connected in anti-parallel.

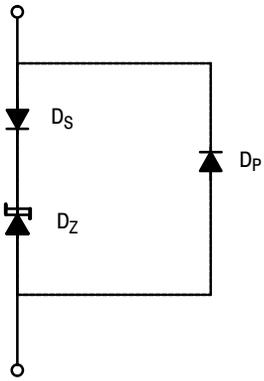


Figure 21. A Series Stack to Achieve Low Capacitance with Zener TVS Diodes

Switching speed is a prime attribute of the zener TVS. Avalanche action occurs in picoseconds but performing tests to substantiate the theory is extremely difficult. As a practical matter, the device may be regarded as responding instantaneously. Voltage overshoots which may appear on protected lines are the result of poor layout and packaging or faulty measurement techniques.

The p-n junction diode is a unidirectional device. For use on ac signal lines, bidirectional devices are available which are based upon stacking two diodes back to back. Most manufacturers use monolithic NPN and PNP structures. The center region is made relatively wide compared to a transistor base to minimize transistor action which can cause increased leakage current.

No wearout mechanism exists for properly manufactured Zener diode chips. They are normally in one of two states; good, or shorted out from over-stress. Long-term life studies show no evidence of degradation of any electrical parameters prior to failure. Failures result from stress which causes separation of the metal heat sink from the silicon chip with subsequent overheating and then failure. Like MOVs, silicon chips quickly fail short under steady state or long duration pulses which exceed their capabilities.

The strengths and weaknesses of Zener TVS devices are listed below.

Advantages:

- High repetitive pulse power ratings
- Low clamping factor
- Sub-nanosecond turn-on
- No wearout
- Broad voltage spectrum
- Short circuit failure mode

Disadvantages:

- Low non-repetitive pulse current
- High capacitance for low voltage types

Because of their fast response and low clamping factor, silicon devices are used extensively for protecting microprocessor based equipment from voltage surges on dc power buses and I/O ports.

THYRISTOR DIODES

The most recent addition to the TVS arsenal is the thyristor surge suppressor (TSS). The device has the low clamping factor and virtually instantaneous response characteristic of a silicon avalanche (Zener) diode but, in addition, it switches to a low voltage on-state when sufficient avalanche current flows. Because the on-state voltage is only a few volts, the TSS can handle much higher currents than a silicon diode TVS having the same chip area and breakdown voltage. Furthermore, the TSS does not exhibit the large overshoot voltage of the gas tube.

Thyristor TVS diodes are available with unidirectional or bidirectional characteristics. The unidirectional type behaves somewhat like an SCR with a Zener diode connected from anode to gate. The bidirectional type behaves similarly to a triac having a bidirectional diode (Diac) from main terminal to gate.

Packaged TSS chips are shown in Figure 18. Figure 22a shows a typical positive switching resistance bidirectional TSS chip. Construction of the device starts with an n material wafer into which the p-bases and n-emitters are diffused. There are four layers from top to bottom on each side of the chip, forming an equivalent SCR. Only half the device conducts for a particular transient polarity.

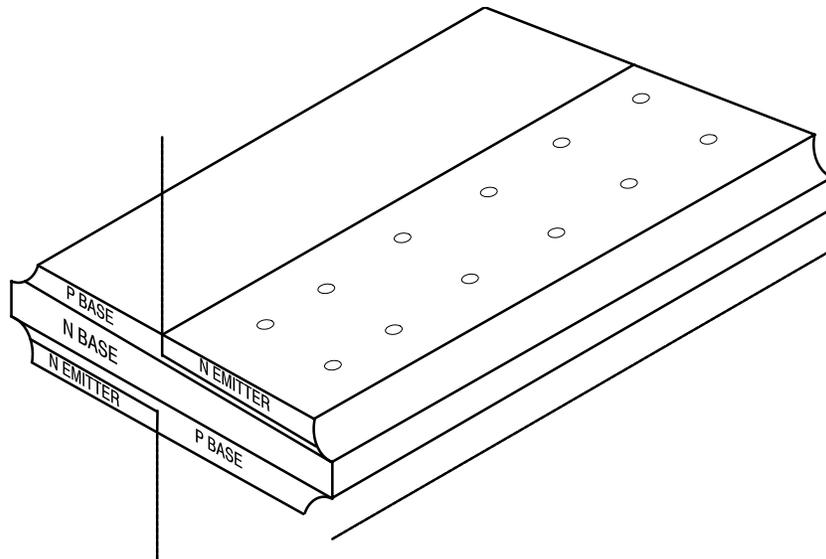


Figure 22a. Chip Construction

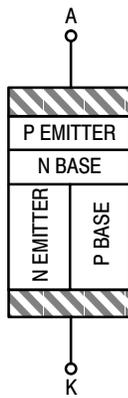


Figure 22b. Cross Section of Left Side

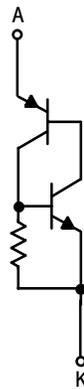


Figure 22c. Circuit Model of Left Side



Figure 22d. Circuit Symbol

The “gate” does not trigger the SCR, instead, operation in the Zener mode begins when the collector junction avalanches. Note that the p-bases pass through the n-emitters in a dot pattern and connect to the contact metal covering both halves of the chip. This construction technique provides a low resistance path for current flow and prevents it from turning on the NPN transistor. Therefore, at relatively low currents, the device acts like a low gain PNP transistor in breakdown. The Zener diode is the collector base junction of the PNP transistor. Negative resistance TSS devices are similarly constructed but start with p-type material wafer, allowing the fabrication of a high-gain NPN transistor. The switchback in voltage with increasing current is caused by the gain of the NPN.

Both device types switch on completely when the current flow through the base emitter shunt resistance causes enough voltage drop to turn on the emitter and begin four layer action. Now the device acts like an SCR. The collector current of the PNP transistor (Figure 22c) provides the base drive for the NPN transistor. Likewise, the collector current

of the NPN transistor drives the base of the PNP causing the two devices to hold one another on. Both the p and n emitters flood the chip with carriers resulting in high electrical conductivity and surge current capability.

When driven with high voltage ac, which occurs during a power cross, positive resistance TSS devices act like a Zener diode until the ac voltage drives the load line through the point where regeneration occurs. Then it abruptly switches to a low voltage. When the peak ac current is just below the current required for breakover, the device operates mainly as a Zener and power dissipation is high although the current is low. When the ac current peak is well above breakover, (>10 A), the device operates mainly as an SCR, and the low on-state voltage causes power dissipation to be relatively low.

Negative resistance devices operate in a similar fashion. However, their behavior is dependent upon the “load line,” that is, the equivalent resistance which the device “sees.” When the load resistance is high (>1000 W) behavior is similar to that of a positive resistance TSS in that high

instantaneous power dissipation occurs as the load line is driven along the high voltage region of the TSS prior to switching.

When a TSS with a negative resistance characteristic is driven with a low “load” resistance, switching occurs when the load line is tangent to the peak of the negative resistance curve. Thus, complete turn-on can occur at a very low current if the load resistance is low and the device has a “sharp” switchback characteristic.

Leakage and the Zener knee voltage increase with temperature at eight percent and 0.11% per °C respectively for 200 V positive resistance types. But the current required to cause regeneration falls with temperature, causing less Zener impedance contribution to the breakover voltage, resulting in a large reduction in the breakover voltage temperature coefficient to as little as 0.05%/°C. Negative resistance types can show positive or negative breakover voltage coefficients depending on temperature and the sharpness of the negative switchback.

The response of both positive and negative switching resistance units to fast transients involves a race between their Zener and regenerative attributes. At first the device conducts only in the small chip area where breakdown is occurring. Time is required for conduction to spread across the chip and to establish the currents and temperatures leading to complete turn-on. The net result is that both types exhibit increasing breakover voltage with fast transients. However, this effect is very small compared to gas discharge tubes, being less than 25% of the breakover voltage.

Negative resistance types are more sensitive to unwanted turn-on by voltage rates (dv/dt) at peak voltages below the avalanche value. The transient current that flows to charge the self-capacitance of the device sets up an operating point on the negative resistance slope leading to turn-on. Reduction of dv/dt capability becomes significant when the signal voltage exceeds 80% of the avalanche value.

Complete turn-off following a transient requires the load line to intersect the device leakage characteristic at a point below the avalanche knee. During turn-off the load line must not meet an intermediate conducting state which can occur with a negative resistance device. Positive resistance types are free of states causing turn-off “sticking.” Both types have temperature sensitive holding currents that lie between 1 and 4 mA/°C.

Recent product developments and published studies have generated much interest. Based on a study sponsored by Bell South,⁹ the authors concluded that these new devices offered the highest level of surge protection available.

Key electrical parameters for the Thyristor TVS include operating voltage, clamping voltage, pulse current, on-state voltage, capacitance, and holding current.

Operating voltage is defined as the maximum normal voltage which the device should experience. Operating voltages from 60 V to 200 V are available.

Clamping voltage is the maximum voltage level attained before thyristor turn-on and subsequent transition to the

on-state conduction mode. The transition stage to conduction may have any of the slopes shown in Figure 9.

The important voltages which define the thyristor operating characteristics are also shown in Figure 9. V_D is operating voltage, V_C is the clamping voltage and V_T is on-state voltage.

On-state voltage for most devices is approximately 3 V. Consequently, transient power dissipation is much lower for the thyristor TVS than for other TVS devices because of its low on-state voltage. For example, under power cross conditions Bell South Services reported their tests showed that the thyristor TVS devices handled short bursts of commercial power with far less heating than arc type surge arrestors.⁹

Capacitance is also a key parameter since in many cases the TSS is a replacement for gas surge arrestors which have low capacitance. Values for the TSS range from 100 pF to 200 pF at zero volts, but drop to about half of these values at a 50 Vdc bias.

Holding current (I_H) is defined as the current required to maintain the on-state condition. Device thru-current must drop below I_H before it will restore to the non-conducting state. Turn-off time is usually not specified but it can be expected to be several milliseconds in a telecom application where the dc follow-on current is just slightly below the holding current.

The major advantages and limitations of the thyristor are:

Advantages:

- Fast response
- No wearout
- Produces no noise
- Short circuit failure mode

Disadvantages:

- Narrow voltage spectrum
- Non-restoring in dc circuits unless current is below I_H
- Turn-off delay time

The thyristor TVS is finding wide acceptance in telecom applications because its characteristics uniquely match telecom requirements. It handles the difficult “power cross” requirements with less stress than other TVS devices while providing the total protection needed.

SURGE PROTECTOR MODULES

TYPES OF SURGE PROTECTOR MODULES

Several component technologies have been implemented either singly or in combination in surge protector modules and devices. The simplest surge protectors contain nothing more than a single transient voltage suppression (TVS) component in a larger package. Others contain two or more in a series, parallel or series-parallel arrangement. Still others contain two or more varieties of TVS elements in combination, providing multiple levels of protection.

Many surge protectors contain non-semiconductor elements such as carbon blocks and varistors. If required,

other modes of protection components may be incorporated, such as circuit breakers or EMI noise filters.

Surge protector modules are one solution to the overvoltage problem. Alternatives include:

- Uninterruptible power systems (UPS), whose main duty is to provide power during a blackout, but secondarily provide protection from surges, sags and spikes.
- Power line conditioners, which are designed to isolate equipment from raw utility power and regulated voltages within narrow limits.

Both UPS and power line conditioners are far more expensive than surge protector modules.

THE 6 MAJOR CATEGORIES OF SURGE PROTECTION MODULES

Plug-in
Hardwired
Utility
Datacom
Telecom
RF and microwave

PLUG-IN MODULES

Plug-in modules come in a variety of sizes and shapes, and are intended for general purpose use. They permit the protection of vulnerable electronic equipment, such as home computers, from overvoltage transients on the 115 vac line. These products are sold in retail outlets, computer stores and via mail order. Most models incorporate a circuit breaker or fuse, and an on/off switch with a neon indicator. The module may have any number of receptacles, with common models having from two to six. These products comply with **UL 1449**, and are generally rated to withstand the application of multiple transients, as specified in **IEEE 587**. Plug-in modules generally provide their protection through the use of these devices which are typically connected between line and neutral, and between neutral and ground.

HARDWIRED

Hardwired modules take on a wide variety of styles, depending upon their designed application. They provide protection for instrumentation, computers, automatic test equipment, industrial controls, motor controls, and for certain telecom situations.

Many of these modules provide snubbing networks employing resistors and capacitors to produce an RC time constant. Snubbers provide common mode and differential mode low-pass filtering to reduce interference from line to equipment, and are effective in reducing equipment generated noise from being propagated onto the line. Snubbers leak current however, and many of these modules are designed with heat sinks and require mounting to a chassis. The surge protection is performed in a similar manner to the plug-in modules mentioned earlier. Hardwired products, therefore, present a prime opportunity for avalanche TVS components.

UTILITY

The power transmission and distribution equipment industry has an obvious need for heavy duty protection against overvoltage transients. Many utility situations require a combination of techniques to provide the necessary solution to their particular problems. This industry utilizes many forms of transient suppression outside the realm of semiconductors.

DATAKOM

Local area networks and other computer links require protection against high energy transients originating on their data lines. In addition, transients on adjacent power lines produce electromagnetic fields that can be coupled onto unprotected signal lines. Datacom protectors have a ground terminal or pigtail which must be tied to the local equipment ground with as short a lead as possible. Datacom protectors should be installed on both ends of a data link, or at all nodes in a network. This protection is in addition to the ac line transient protection, which is served by the plug-in or hardwired protection modules. Some datacom protector modules contain multi-stage hybrid circuits, specially tailored for specific applications, such as 4–20 mA analog current loops.

TELECOM

Included here are devices used to protect central office and station telecommunications (telecom) equipment against voltage surges. None of these devices are grounded through an ac power receptacle. Those that are grounded through an ac power receptacle are categorized as plug-in modules. Not only can overvoltages cause disruptions of telecom service, but they can destroy the sophisticated equipment connected to the network. Also, users or technicians working on the equipment can be injured should lightning strike nearby. It is estimated that 10 to 15 people are killed in the U.S. each year while talking on the telephone during lightning storms. For these reasons, surge protectors are used both in central offices and in customer premises.

There are three types of telecom surge protectors now in service: air-gap carbon block, gas tube, and solid state. The desire of the telecom market is to convert as many of the non-solid state implementations into solid state as cost will permit.

SELECTING TVS COMPONENTS

From the foregoing discussion it should be clear that the silicon junction avalanche diode offers more desirable characteristics than any other TVS component. Its ability to clamp fast rising transients without overshoot, low clamping factor, non-latching behavior, and lack of a wearout mechanism are the overriding considerations. Its one-shot surge capability is lower than most other TVS devices but is normally adequate for the application. Should an unusually severe event occur, it will short yet still protect the equipment.

For example, an RS-232 data line is specified to operate with a maximum signal level of ± 25 V. Failure analysis studies¹¹ have shown that the transmitters and receivers used on RS-232 links tolerate 40 V transients. A 1.5KE27CA diode will handle the maximum signal level while holding the peak transient voltage to less than 40 V with a 40 A 10/1000 pulse which is adequate for all indoor and most outdoor data line runs. As a practical matter, few data links use 25 V signals; 5 V is most common. Consequently, much lower voltage silicon diodes may be used which will allow a corresponding increase in surge current capability. For example, a 10 V breakdown device from the same 1500 W family will clamp to under 15 V (typically 12 V) when subjected to a 100 A pulse.

Telecommunications lines which must accommodate the ring voltage have much more severe requirements. For example, one specification¹¹ from Bellcore suggests that leakage current be under 20 mA over the temperature range from -40°C to $+65^{\circ}\text{C}$ with 265 V peak ac applied. To meet this specification using Zener TVS parts, devices must be stacked. Devices which breakdown at 160 V are chosen to accommodate tolerances and the temperature coefficient. A part number with a 10% tolerance on breakdown could supply a unit which breaks down at 144 V. At -40°C breakdown could be 133 V. The breakdown of two devices stacked just barely exceeds the worst case ring peak of 265 V. A 1500 W unit has a surge capability of 6.5 A (10/1000) which is too low to be satisfactory while higher power units are too expensive as a rule.

Another problem which telephone line service presents to a surge suppressor is survival during a power cross. An avalanche diode is impractical to use because the energy delivered by a power cross will produce diode failure before any overcurrent protective element can react.

As indicated by the Bell South studies, the thyristor TVS is ideal for telephone line applications. Suppliers offer unidirectional and bidirectional units which meet the FCC impulse wave requirements as shown in Table 1. In addition, the thyristor can handle several cycles of 50/60 Hz power before failure. The ON Semiconductor MKT1V200 series, for example, can handle 10 A for 4 cycles, which is enough time for a low current fuse or other current activated protective device to react.

APPLICATION CONSIDERATIONS

In most cases, it is not advisable to place a zener TVS directly across a data line because of its relatively high capacitance. The arrangement previously discussed and shown in Figure 21 works well for an unbalanced line such as RS-232. When using discrete steering diodes, they should have low capacitance and low turn-on impedance to avoid causing an overshoot on the clamped voltage level.

Most noise and transient surge voltages occur on lines with respect to ground. A signal line such as RS-232 which uses ground as a signal reference is thus very vulnerable to noise and transients. It is, however, easy to protect using a single TVS at each end of the line.

Telephone and RS-422 lines are called balanced lines because the signal is placed between two lines which are “floating” with respect to ground. A signal appears between each signal line and ground but is rejected by the receiver; only the difference in potential between the two signal lines is recognized as the transmitted signal. This system has been in use for decades as a means of providing improved noise immunity, but protection from transient surge voltages is more complex.

A cost-effective means of protecting a balanced line is shown in Figure 23. The bridge diode arrangement allows protection against both positive and negative transients to be achieved, an essential requirement but the TVS devices Z_1 and Z_2 need only be unidirectional. The diodes are chosen to have low capacitance to reduce loading on the line and low turn-on impedance to avoid causing an overshoot on the clamped voltage level. Although a zener TVS is shown, a TSS is more appropriate when a telephone line having ring voltages is to be protected.

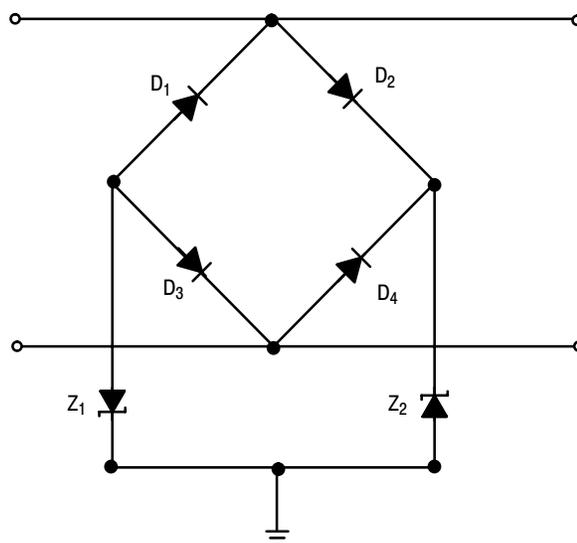


Figure 23. A Method of Reducing Capacitance and Protecting a Balanced Line

Since transients are usually common-mode, it is important that the TVS circuit operate in a balanced fashion; otherwise, common mode transients can cause differential mode disturbances which can be devastating to the line receiver. For example, suppose that an identical positive common mode surge voltage appears from each line-to-ground. Diodes D_2 and D_4 will conduct the transients to Z_2 . However, if one of these diodes has a slower turn-on or higher dynamic impedance than the other, the voltage difference caused by the differing diode response appears across the signal lines. Consequently, the bridge diodes must be chosen to be as nearly identical as possible.

Should a differential mode transient appear on the signal lines, it will be held to twice that of the line-to-ground clamping level. In many cases a lower clamping level is needed which can be achieved by placing another TVS across the signal lines. It must be a bidirectional low

capacitance device. With a line-to-line TVS in the circuit diode matching is not required.

Other schemes appearing in the literature use two bidirectional TVS devices from each line-to-ground as shown in Figure 24 that of the line-to-ground voltage. To avoid generating a differential mode signal, the TVS must be closely matched or a third TVS must be placed line-to-line. By using a third TVS differential mode transients can be held to a low level.

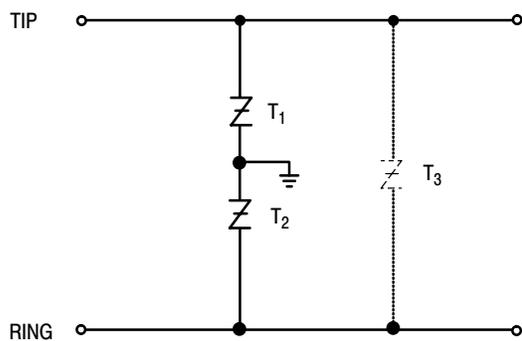


Figure 24. Protecting a Balanced Line with Bidirectional TSS Devices

The arrangement of Figure 25 offers the advantage of lower capacitance when differential mode transient protection is required. If all three TSS devices have the same capacitance (C), the line-to-line and line-to-ground capacitance of Figure 24 is $1.5C$. However, the arrangement of Figure 25 exhibits a capacitance of only $C/2$. To design the circuit to handle the same simultaneous common mode energy as the circuit of Figure 24, T_3 must be twice as large as T_1 and T_2 . In this case the capacitance of T_3 is doubled which causes the line-to-ground capacitance to be $2C/3$, still a considerable improvement over the arrangement of Figure 24.

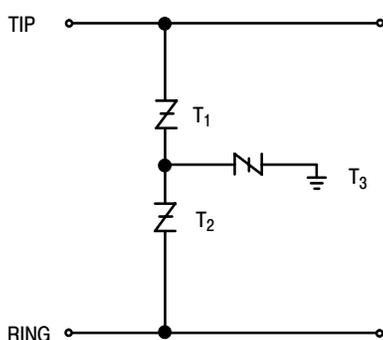


Figure 25. Preferred Method of Protecting a Balanced Line Using Bidirectional TSS Devices

Protectors are usually designed to be “fail safe” if their energy ratings are exceeded, but the definition of “fail safe” is often dependent upon the application. The most common requirement is that the surge voltage protective element should fail short and remain shorted regardless of the

resulting current flow. To insure that this occurs, semiconductor devices use heavy gauge clips or bonding wires between the chip and terminals. In addition, parts are available in plastic packages having a spring type shorting bar which shorts the terminals when the package softens at the very high temperatures generated during a severe overload.

The shorted TVS protects the equipment, but the line feeding it could be destroyed if the source of energy which shorted the TVS is from a power cross. Therefore, it is wise and necessary for a UL listing to provide a series element such as a fuse or PTC device to either open the circuit or restrict its value to a safe level.

The design of circuit boards is critical and layout must be done to minimize any lead or wiring inductance in series with the TVS. Significant voltage is developed in any loop subject to transients because of their high current amplitudes and fast risetimes.

REFERENCES

1. D. W. Bodle and P. A. Gresh, “Lightning Surges in Paired Telephone Cable Facilities,” *The Bell System Technical Journal*, Vol. 4, March 1961, pp. 547–576.
2. D. G. Stroh, “Static Electricity Can Kill Transistors,” *Electronics*, Vol. 35, 1962, pp. 90–91.
3. J. D. Norgard and C. L. Chen, “Lightning-Induced Transients on Buried Shielded Transmission Lines,” *IEEE Transactions on EMC*, Vol. EMC28, No. 3, August 1986, pp. 168–171.
4. O. M. Clark, “Transient Voltage Suppression (TVS),” 1989, pp. 6–7.
5. Clark, p. 7.
6. World Information Technologies, “U. S. Electrical and Electronic Surge Protection Markets,” 1989, p. 5.
7. T. J. Tucker, “Spark Initiation Requirements of a Secondary Explosive,” *Annals of the New York Academy of Sciences*, Vol. 152, Article I, 1968, pp. 643–653.
8. H. K. Florig, “The Future Battlefield: A Blast of Gigawatts?,” *IEEE Spectrum*, Vol. 25, No. 3, March 1988, pp. 50–54.
9. Mel Thrasher, “A Solid State Solution,” *Telephony*, June 1989, pp. 48–52.
10. A. Urbietta, “Sensitivity Study to EOS/SSD of Bipolar Integrated Circuits,” *EOS-8*, 1986.
11. M. Tetreault and F.D. Martzloff, “Characterization of Disturbing Transient Waveforms on Computer Data Communication Lines,” *EMC Proceedings, Zurich*, March 1985, pp. 423–428.
12. F. Martzloff, “Coupling, Propagation, and Side Effects of Surges in an Industrial Building Wiring System,”

Conference Record of IEEE IAS Meeting, 1988, pp. 1467–1475.

IMPORTANT REGULATORY REQUIREMENTS AND GUIDELINES

GENERAL

DOD-STD-1399, MIL-STD-704, MIL-STD-1275, MIL-STD-461C. These military specifications are important, if we intend to target devices for military or commercial aviation markets.

IEEE 587. This specification describes multiple transient waveforms.

UL1449. This is a compulsory test which demonstrates performance to criteria establishing the maximum voltage that can pass through a device after clamping has taken place. It is important that we comply, and say so on our data sheets.

INDUSTRIAL

ANSI/IEEE C62.41. Established by the American National Standards Institute (ANSI) and the Institute of Electrical and Electronic Engineers (IEEE), this guideline tests the effectiveness of devices to typical power disturbances. To meet the most rigorous category of this spec, a device or module must be capable of withstanding a maximum repetitive surge current pulse of 3000 amps with a 8/20 μ s waveform.

IEC TC 102 D. Requirements for remote control receivers for industrial applications are detailed in this International Electrotechnical Commission (IEC) document.

IEC 255-4 and IEC TC 41. These documents describe testing for static relays for industrial use.

IEC 801-1 thru -3. These are specifications for various industrial control applications.

IEC 801-4. The IEC specifies transient voltage impulses which occur from the switching of inductive loads. We must be aware of the importance of this specification, especially in Europe, and characterize our devices' performance to it.

IEEE 472/ANSI C 37.90.1. Requirements for protective relays, including 10/1000 nS waveform testing is described.

UL943. This requirement defines a 0.5 μ s/100 kHz waveform for ground fault and other switching applications.

VDE 0420. Industrial remote control receivers are detailed, and test procedures defined.

VDE 0860/Part 1/II. This includes a description of 0.2/200 μ s, 10kV test requirements.

TELECOM

CCITT IX K.17, K.20, K.15. These documents relate to repeaters.

EIA PM-1361. This document covers requirements for telephone terminals and data processing equipment.

FTZ 4391 TV1. This is a general German specification for telecom equipment.

FCC Part 68. The Federal Communications Commission (FCC) requirements for communications equipment is defined. Of special note is **§68.302**, dealing with telecom power lines.

NT/DAS/PRL/003. Telephone instrument, subscriber equipment and line requirements are documented.

PTT 692.01. This is a general Swiss specification for telecom exchange equipment.

REA PE-60. The Rural Electrification Administration (REA) has documented the predominant waveform for induced lightning transients. This test is now commonly known as the 10/1000 μ s pulse test.

REA PE-80. The REA defines requirements for gas tubes and similar devices for telecom applications.

TA-TSY-000974. This technical advisory by Bellcore defines double exponential waveforms, which are the basis for many telecom applications norms.

UL 1459 and UL 4978. These document detail tests for standard telephone equipment and data transmission.

TA-TSY-000974. This technical advisory by Bellcore defines double exponential waveforms, which are the basis for many telecom applications norms.

UL 1459 and UL 4978. These document detail tests for standard telephone equipment and data transmission.

Transient Power Capability of Zener Diodes

Prepared by
Applications Engineering and
Jerry Wilhardt, Product Engineer —
Industrial and Hi-Rel Zener Diodes



ON Semiconductor™

<http://onsemi.com>

APPLICATION NOTE

INTRODUCTION

Because of the sensitivity of semiconductor components to voltage transients in excess of their ratings, circuits are often designed to inhibit voltage surges in order to protect equipment from catastrophic failure. External voltage transients are imposed on power lines as a result of lightning strikes, motors, solenoids, relays or SCR switching circuits, which share the same ac source with other equipment. Internal transients can be generated within a piece of equipment by rectifier reverse recovery transients, switching of loads or transformer primaries, fuse blowing, solenoids, etc. The basic relation, $v = L di/dt$, describes most equipment developed transients.

ZENER DIODE CHARACTERISTICS

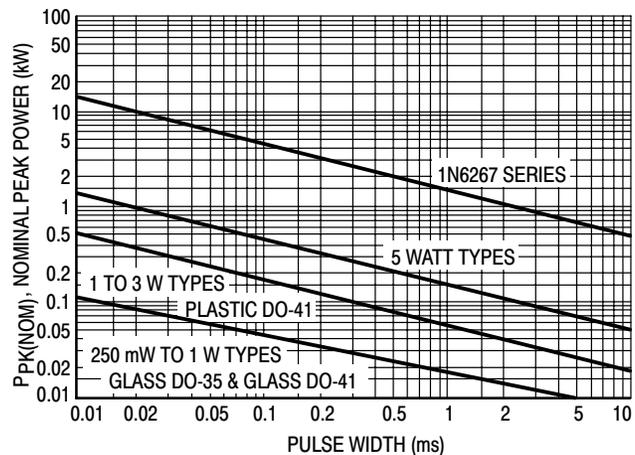
Zener diodes, being nearly ideal clippers (that is, they exhibit close to an infinite impedance below the clipping level and close to a short circuit above the clipping level), are often used to suppress transients. In this type of application, it is important to know the power capability of the zener for short pulse durations, since they are intolerant of excessive stress.

Some ON Semiconductor data sheets such as the ones for devices shown in Table 1 contain short pulse surge capability. However, there are many data sheets that do not contain this data and Figure 1 is presented here to supplement this information.

Table 1. Transient Suppressor Diodes			
Series Numbers	Steady State Power	Package	Description
1N4728	1 W	DO-41	Double Slug Glass
1N6267	5 W	Case 41A	Axial Lead Plastic
1N5333A	5 W	Case 17	Surmetic 40
1N746/957 A/4371	400 mW	DO-35	Double Slug Glass
1N5221A	500 mW	DO-35	Double Slug Glass

Some data sheets have surge information which differs slightly from the data shown in Figure 1. A variety of reasons exist for this:

1. The surge data may be presented in terms of actual surge power instead of nominal power.
2. Product improvements have occurred since the data sheet was published.



Power is defined as $V_{Z(NOM)} \times I_{Z(PK)}$ where $V_{Z(NOM)}$ is the nominal zener voltage measured at the low test current used for voltage classification.

Figure 1. Peak Power Ratings of Zener Diodes

3. Larger dice are used, or special tests are imposed on the product to guarantee higher ratings than those shown on Figure 1.

4. The specifications may be based on a JEDEC registration or part number of another manufacturer.

The data of Figure 1 applies for non-repetitive conditions and at a lead temperature of 25°C. If the duty cycle increases, the peak power must be reduced as indicated by the curves of Figure 2. Average power must be derated as the lead or ambient temperature rises above 25°C. The average power derating curve normally given on data sheets may be normalized and used for this purpose.

At first glance the derating curves of Figure 2 appear to be in error as the 10 ms pulse has a higher derating factor than the 10 μs pulse. However, when the derating factor for a given pulse of Figure 2 is multiplied by the peak power value of Figure 1 for the same pulse, the results follow the expected trend.

When it is necessary to use a zener close to surge ratings, and a standard part having guaranteed surge limits is not suitable, a special part number may be created having a surge limit as part of the specification. Contact your nearest ON Semiconductor OEM sales office for capability, price, delivery, and minimum order criteria.

MATHEMATICAL MODEL

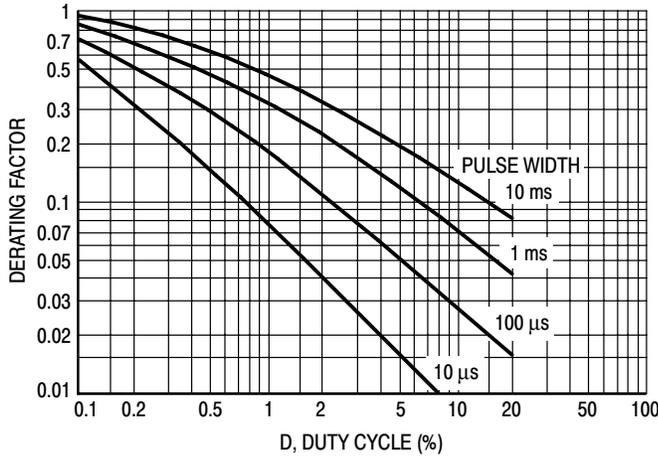


Figure 2. Typical Derating Factor for Duty Cycle

Since the power shown on the curves is not the actual transient power measured, but is the product of the peak current measured and the nominal zener voltage measured at the current used for voltage classification, the peak current can be calculated from:

$$I_{Z(PK)} = \frac{P_{(PK)}}{V_{Z(NOM)}} \quad (1)$$

The peak voltage at peak current can be calculated from:

$$V_{Z(PK)} = F_C \times V_{Z(NOM)} \quad (2)$$

where F_C is the clamping factor. The clamping factor is approximately 1.20 for all zener diodes when operated at their pulse power limits. For example, a 5 watt, 20 volt zener can be expected to show a peak voltage of 24 volts regardless of whether it is handling 450 watts for 0.1 ms or 50 watts for 10 ms. This occurs because the voltage is a function of junction temperature and IR drop. Heating of the junction is more severe at the longer pulse width, causing a higher voltage component due to temperature which is roughly offset by the smaller IR voltage component.

For modeling purposes, an approximation of the zener resistance is needed. It is obtained from:

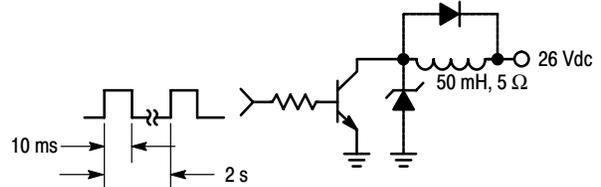
$$R_{Z(NOM)} = \frac{V_{Z(NOM)}(F_C - 1)}{P_{PK(NOM)}/V_{Z(NOM)}} \quad (3)$$

The value is approximate because both the clamping factor and the actual resistance are a function of temperature.

CIRCUIT CONSIDERATIONS

It is important that as much impedance as circuit constraints allow be placed in series with the zener diode and the components to be protected. The result will be a lower clipping voltage and less zener stress. A capacitor in parallel with the zener is also effective in reducing the stress imposed by very short duration transients.

To illustrate use of the data, a common application will be analyzed. The transistor in Figure 3 drives a 50 mH solenoid which requires 5 amperes of current. Without some means of clamping the voltage from the inductor when the transistor turns off, it could be destroyed.



Used to select a zener diode having the proper voltage and power capability to protect the transistor.

Figure 3. Circuit Example

The means most often used to solve the problem is to connect an ordinary rectifier diode across the coil; however, this technique may keep the current circulating through the coil for too long a time. Faster switching is achieved by allowing the voltage to rise to a level above the supply before being clamped. The voltage rating of the transistor is 60 V, indicating that approximately a 50 volt zener will be required.

The peak current will equal the on-state transistor current (5 amperes) and will decay exponentially as determined by the coil L/R time constant (neglecting the zener impedance). A rectangular pulse of width L/R (0.01 sec) and amplitude of I_{PK} (5 A) contains the same energy and may be used to select a zener diode. The nominal zener power rating therefore must exceed $(5 A \times 50) = 250$ watts at 10 ms and a duty cycle of $0.01/2 = 0.5\%$. From Figure 2, the duty cycle factor is 0.62 making the single pulse power rating required equal to $250/0.62 = 403$ watts. From Figure 1, one of the 1N6267 series zeners has the required capability. The 1N6287 is specified nominally at 47 volts and should prove satisfactory.

Although this series has specified maximum voltage limits, equation 3 will be used to determine the maximum zener voltage in order to demonstrate its use.

$$R_Z = \frac{47(1.20 - 1)}{500/47} = \frac{9.4}{10.64} = 0.9\Omega$$

At 5 amperes, the peak voltage will be 4.5 volts above nominal or 51.5 volts total which is safely below the 60 volt transistor rating.

A Review of Transients and Their Means of Suppression

Prepared by
Steve Cherniak
Applications Engineering



ON Semiconductor™

<http://onsemi.com>

APPLICATION NOTE

INTRODUCTION

One problem that most, if not all electronic equipment designers must deal with, is transient overvoltages. Transients in electrical circuits result from the sudden release of previously stored energy. Some transients may be voluntary and created in the circuit due to inductive switching, commutation voltage spikes, etc. and may be easily suppressed since their energy content is known and predictable. Other transients may be created outside the circuit and then coupled into it. These can be caused by lightning, substation problems, or other such phenomena. These transients, unlike switching transients, are beyond the control of the circuit designer and are more difficult to identify, measure and suppress.

Effective transient suppression requires that the impulse energy is dissipated in the added suppressor at a low enough voltage so the capabilities of the circuit or device will not be exceeded.

REOCCURRING TRANSIENTS

Transients may be formed from energy stored in circuit inductance and capacitance when electrical conditions in the circuit are abruptly changed.

Switching induced transients are a good example of this; the change in current ($\frac{di}{dt}$) in an inductor (L) will generate

a voltage equal to $L \frac{di}{dt}$. The energy (J) in the transient is equal to $1/2Li^2$ and usually exists as a high power impulse for a relatively short time ($J = Pt$).

If load 2 is shorted (Figure 1), devices parallel to it may be destroyed. When the fuse opens and interrupts the fault current, the slightly inductive power supply produces a transient voltage spike of $V = L \frac{di}{dt}$ with an energy content of $J = 1/2Li^2$. This transient might be beyond the voltage limitations of the rectifiers and/or load 1. Switching out a high current load will have a similar effect.

TRANSFORMER PRIMARY BEING ENERGIZED

If a transformer is energized at the peak of the line voltage (Figure 2), this voltage step function can couple to the stray capacitance and inductance of the secondary winding and generate an oscillating transient voltage whose oscillations depend on circuit inductance and capacitance. This transient's peak voltage can be up to twice the peak amplitude of the normal secondary voltage.

In addition to the above phenomena the capacitively coupled (C_S) voltage spike has no direct relationship with the turns ratio, so it is possible for the secondary circuit to see the peak applied primary voltage.

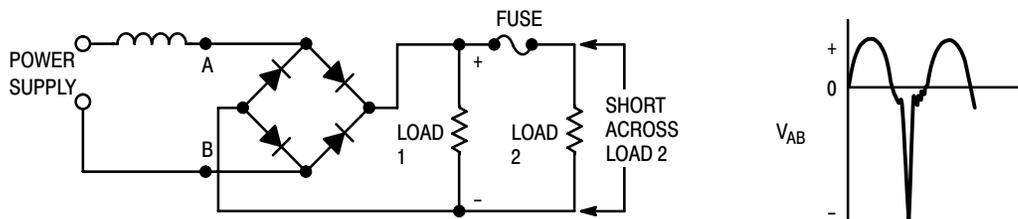


Figure 1. Load Dump with Inductive Power Supply

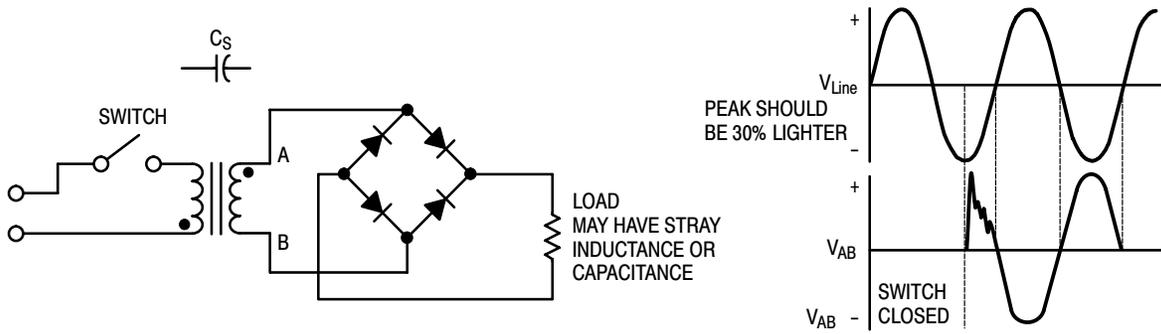


Figure 2. Situation Where Transformer Capacitance Causes a Transient

TRANSFORMER PRIMARY BEING DE-ENERGIZED

If the transformer is driving a high impedance load, transients of more than ten times normal voltage can be created at the secondary when the primary circuit of the transformer is opened during zero-voltage crossing of the ac line. This is due to the interruption of the transformer magnetizing current which causes a rapid collapse of the magnetic flux in the core. This, in turn, causes a high voltage transient to be coupled into the transformer's secondary winding (Figure 3).

Transients produced by interrupting transformers magnetizing current can be severe. These transients can destroy a rectifier diode or filter capacitor if a low impedance discharge path is not provided.

SWITCH "ARCING"

When a contact type switch opens and tries to interrupt current in an inductive circuit, the inductance tries to keep current flowing by charging stray capacitances. (See Figure 4.)

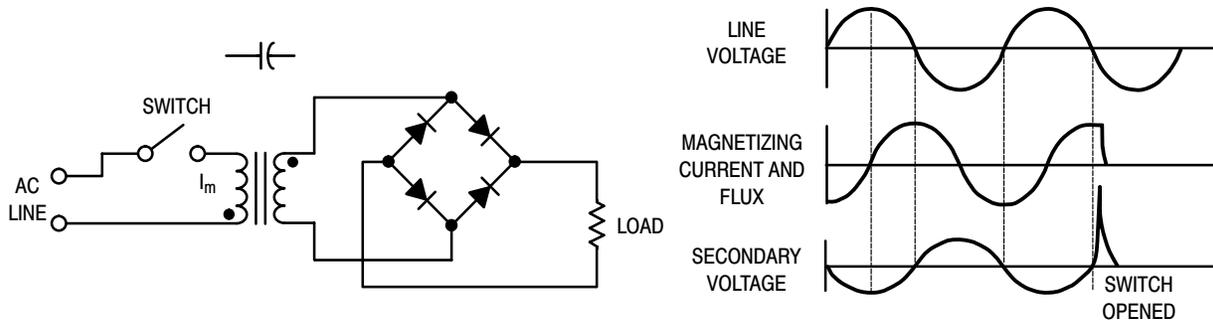


Figure 3. Typical Situation Showing Possible Transient When Interrupting Transformer Magnetizing Current

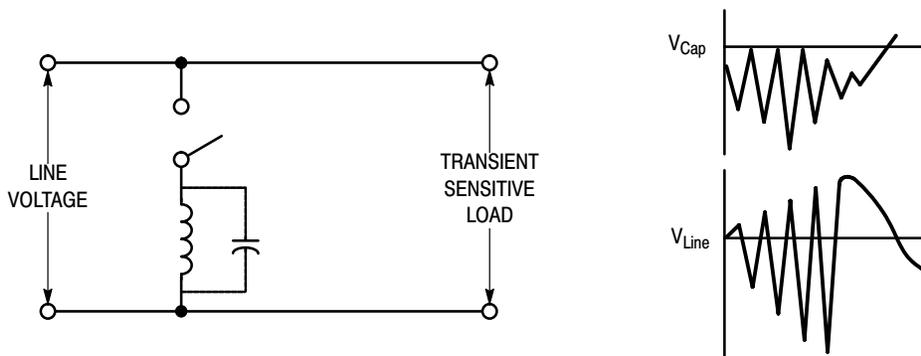


Figure 4. Transients Caused by Switch Opening

This can also happen when the switch contacts bounce open after its initial closing. When the switch is opened (or bounces open momentarily) the current that the inductance wants to keep flowing will oscillate between the stray capacitance and the inductance. When the voltage due to the oscillation rises at the contacts, breakdown of the contact gap is possible, since the switch opens (or bounces open) relatively slowly compared to the oscillation frequency, and the distance may be small enough to permit “arcing.” The arc will discontinue at the zero current point of the oscillation, but as the oscillatory voltage builds up again and the contacts move further apart, each arc will occur at a higher voltage until the contacts are far enough apart to interrupt the current.

WAVESHAPES OF SURGE VOLTAGES

Indoor Waveshapes

Measurements in the field, laboratory, and theoretical calculations indicate that the majority of surge voltages in indoor low-voltage power systems have an oscillatory waveshape. This is because the voltage surge excites the natural resonant frequency of the indoor wiring system. In addition to being typically oscillatory, the surges can also have different amplitudes and waveshapes in the various places of the wiring system. The resonant frequency can range from about 5 kHz to over 500 kHz. A 100 kHz frequency is a realistic value for a typical surge voltage for most residential and light industrial ac wire systems.

The waveshape shown in Figure 5 is known as an “0.5 μ s – 100 kHz ring wave.” This waveshape is reasonably representative of indoor low-voltage (120 V – 240 V) wiring system transients based on measurements conducted by several independent organizations. The waveshape is defined as rising from 10% to 90% of its final amplitude in 0.5 μ s, then decays while oscillating at 100 kHz, each peak being 60% of the preceding one.

The fast rise portion of the waveform can induce the effects associated with non-linear voltage distribution in windings or cause dv/dt problems in semiconductors. Shorter rise times can be found in transients but they are lengthened as they propagate into the wiring system or reflected from wiring discontinuities.

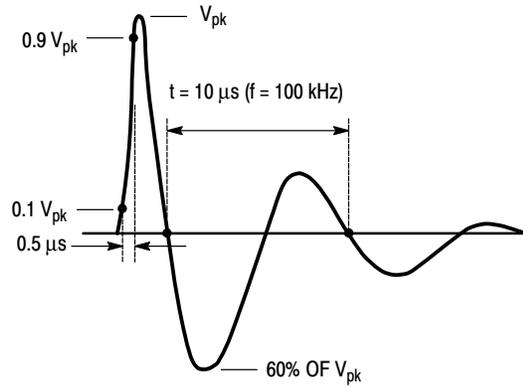


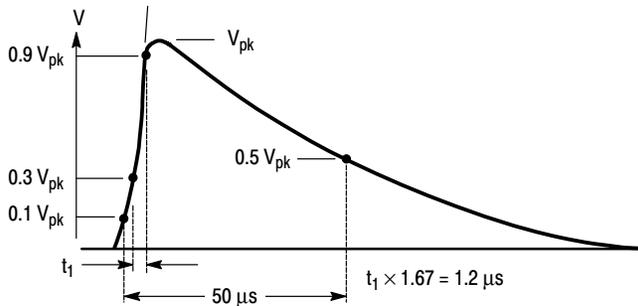
Figure 5. 0.5 μ s 100 kHz Ring Wave

The oscillating portion of the waveform produces voltage polarity reversal effects. Some semiconductors are sensitive to polarity changes or can be damaged when forced into or out of conduction (i.e. reverse recovery of rectifier devices). The sensitivity of some semiconductors to the timing and polarity of a surge is one of the reasons for selecting this oscillatory waveform to represent actual conditions.

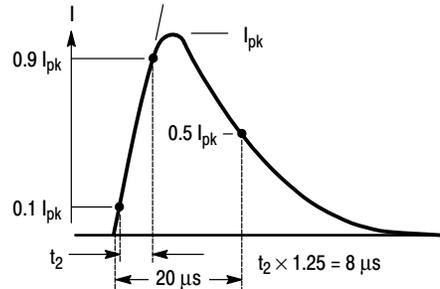
Outdoor Locations

Both oscillating and unidirectional transients have been recorded in outdoor environments (service entrances and other places nearby). In these locations substantial energy is still available in the transient, so the waveform used to model transient conditions outside buildings must contain greater energy than one used to model indoor transient surges.

Properly selected surge suppressors have a good reputation of successful performance when chosen in conjunction with the waveforms described in Figure 6. The recommended waveshape of $1.2 \times 50 \mu$ s (1.2 μ s is associated with the transients rise time and the 50 μ s is the time it takes for the voltage to drop to $1/2V_{pk}$) for the open circuit voltage and $8 \times 20 \mu$ s for the short circuit current are as defined in IEEE standard 28-ANSI Standard C62.1 and can be considered a realistic representation of an outdoor transients waveshape.



(a) Open-Circuit Voltage Waveform



(b) Discharge Current Waveform

Figure 6. Unidirectional Wave Shapes

The type of device under test determines which waveshape in Figure 6 is more appropriate. The voltage waveform is normally used for insulation voltage withstand tests and the current waveform is usually used for discharge current tests.

RANDOM TRANSIENTS

The source powering the circuit or system is frequently the cause of transient induced problems or failures. These transients are difficult to deal with due to their nature; they are totally random and it is difficult to define their amplitude, duration and energy content. These transients are generally caused by switching parallel loads on the same

branch of a power distribution system and can also be caused by lightning.

AC POWER LINE TRANSIENTS

Transients on the ac power line range from just above normal voltage to several kV. The rate of occurrence of transients varies widely from one branch of a power distribution system to the next, although low-level transients occur more often than high-level surges.

Data from surge counters and other sources is the basis for the curves shown in Figure 7. This data was taken from unprotected (no voltage limiting devices) circuits meaning that the transient voltage is limited only by the sparkover distance of the wires in the distribution system.

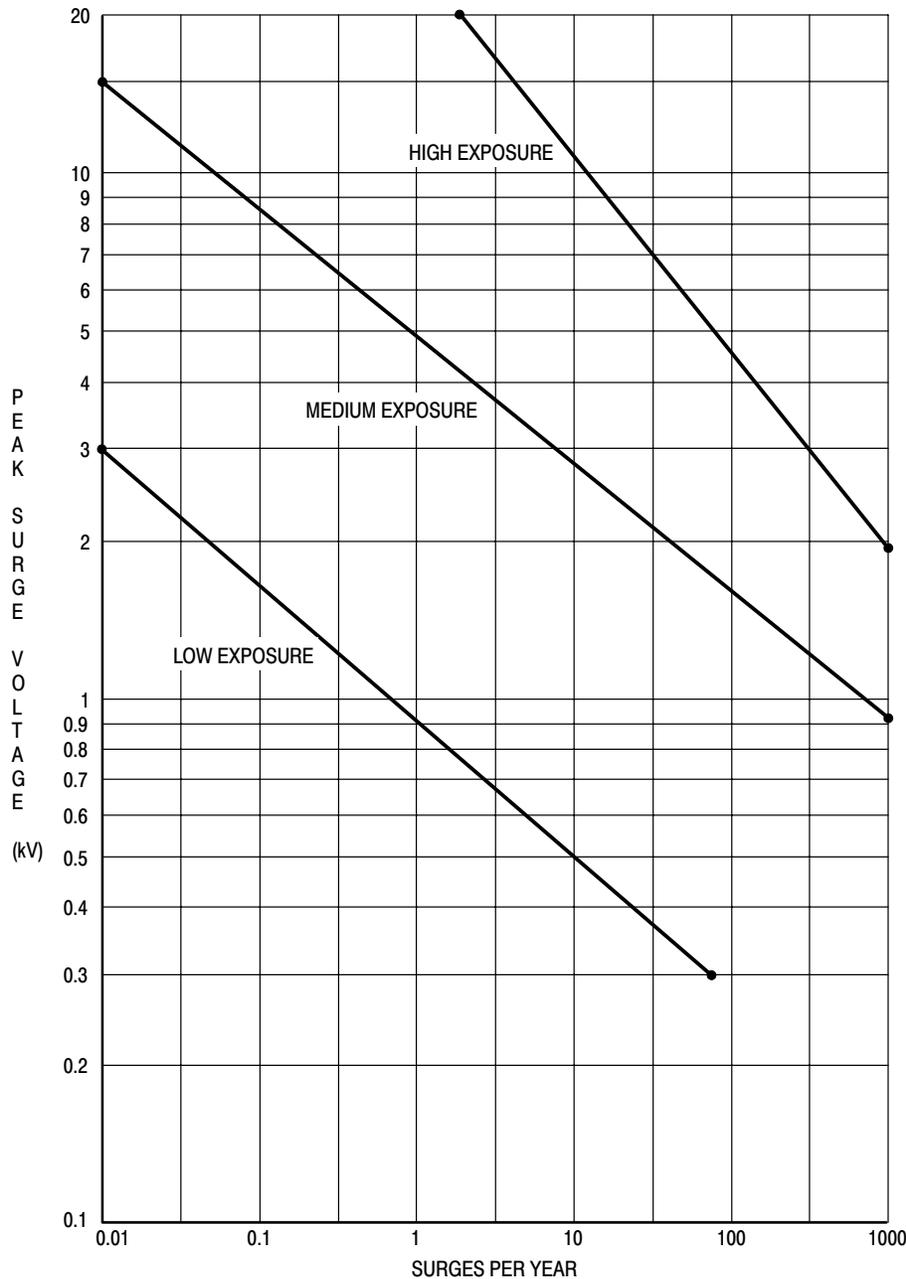


Figure 7. Peak Surge Voltage versus Surges per Year*

*EIA paper, P587.1/F, May, 1979, Page 10

The low exposure portion of the set of curves is data collected from systems with little load-switching activity that are located in areas of light lightning activity.

Medium exposure systems are in areas of frequent lightning activity with a severe switching transient problem.

High exposure systems are rare systems supplied by long overhead lines which supply installations that have high sparkover clearances and may be subject to reflections at power line ends.

When using Figure 7 it is helpful to remember that peak transient voltages will be limited to approximately 6 kV in indoor locations due to the spacing between conductors using standard wiring practices.

TRANSIENT ENERGY LEVELS AND SOURCE IMPEDANCE

The energy contained in a transient will be divided between the transient suppressor and the source impedance of the transient in a way that is determined by the two impedances. With a spark-gap type suppressor, the low impedance of the Arc after breakdown forces most of the transient’s energy to be dissipated elsewhere, e.g. in a current limiting resistor in series with the spark-gap and/or in the transient’s source impedance. Voltage clamping suppressors (e.g. zeners, mov’s, rectifiers operating in the breakdown region) on the other hand absorb a large portion of the transient’s surge energy. So it is necessary that a realistic assumption of the transient’s source impedance be made in order to be able to select a device with an adequate surge capability.

The 100 kHz “Ring Wave” shown in Figure 5 is intended to represent a transient’s waveshape across an open circuit. The waveshape will change when a load is connected and the amount of change will depend on the transient’s source impedance. The surge suppressor must be able to withstand the current passed through it from the surge source. An assumption of too high a surge impedance (when testing the suppressor) will not subject the device under test to sufficient stresses, while an assumption of too low a surge impedance may subject it to an unrealistically large stress; there is a trade-off between the size (cost) of the suppressor and the amount of protection obtained.

In a building, the transient’s source impedance increases with the distance from the electrical service entrance, but

open circuit voltages do not change very much throughout the structure since the wiring does not provide much attenuation. There are three categories of service locations that can represent the majority of locations from the electrical service entrance to the most remote wall outlet. These are listed below. Table 1 is intended as an aid for the preliminary selection of surge suppression devices, since it is very difficult to select a specific value of source impedance.

Category I: Outlets and circuits a “long distance” from electrical service entrance. Outlets more than 10 meters from Category II or more than 20 meters from Category III (wire gauge #14 – #10)

Category II: Major bus lines and circuits a “short distance” from electrical service entrance. Bus system in industrial plants. Outlets for heavy duty appliances that are “close” to the service entrance.

- Distribution panel devices.
- Commercial building lighting systems.

Category III. Electrical service entrance and outdoor locations.

- Power line between pole and electrical service entrance.
- Power line between distribution panel and meter.
- Power line connection to additional near-by buildings.
- Underground power lines leading to pumps, filters, etc.

Categories I and II in Table 1 correspond to the extreme range of the “medium exposure” curve in Figure 7. The surge voltage is limited to approximately 6 kV due to the sparkover spacing of indoor wiring.

The discharge currents of Category II were obtained from simulated lightning tests and field experience with suppressor performance.

The surge currents in Category I are less than in Category II because of the increase in surge impedance due to the fact that Category I is further away from the service entrance.

Category III can be compared to the “High Exposure” situation in Figure 7. The limiting effect of sparkover is not available here so the transient voltage can be quite large.

Table 1. Surge Voltages and Currents Deemed to Represent the Indoor Environment Depending Upon Location						
Category	Waveform	Surge Voltage ⁽¹⁾	Surge Current ⁽²⁾	Energy (Joules) Dissipated in a Suppressor with a Clamping Voltage of ⁽³⁾		
				250 V	500 V	1000 V
I	0.5 μs 100 kHz Ring Wave	6 kV	200 A	0.4	0.8	1.6
II	0.5 μs 100 kHz Ring Wave	6 kV	500 A	1	2	4
	1.2 × 50 μs 8 × 20 μs	6 kV	3 kA	20	40	80
III	1.2 × 50 μs 8 × 20 μs	10 kV or more	10 kA or more			

Notes: 1. Open Circuit voltage
 2. Discharge current of the surge (not the short circuit current of the power system)
 3. The energy a suppressor will dissipate varies in proportion with the suppressor’s clamping voltage, which can be different with different system voltages (assuming the same discharge current).

LIGHTNING TRANSIENTS

There are several mechanisms in which lightning can produce surge voltages on power distribution lines. One of them is a direct lightning strike to a primary (before the substation) circuit. When this high current, that is injected into the power line, flows through ground resistance and the surge impedance of the conductors, very large transient voltages will be produced. If the lightning misses the primary power line but hits a nearby object the lightning discharge may also induce large voltage transients on the line. When a primary circuit surge arrester operates and limits the primary voltage the rapid dv/dt produced will effectively couple transients to the secondary circuit through the capacitance of the transformer (substation) windings in addition to those coupled into the secondary circuit by normal transformer action. If lightning struck the secondary circuit directly, very high currents may be involved which would exceed the capability of conventional surge suppressors. Lightning ground current flow resulting from nearby direct to ground discharges can couple onto the common ground impedance paths of the grounding networks also causing transients.

AUTOMOTIVE TRANSIENTS

Transients in the automotive environment can range from the noise generated by the ignition system and the various accessories (radio, power window, etc.) to the potentially destructive high energy transients caused by the charging (alternator/regulator) system. The automotive “Load

Dump” can cause the most destructive transients; it is when the battery becomes disconnected from the charging system during high charging rates. This is not unlikely when one considers bad battery connections due to corrosion or other wiring problems. Other problems can exist such as steady state overvoltages caused by regulator failure or 24 V battery jump starts. There is even the possibility of incorrect battery connection (reverse polarity).

Capacitive and/or inductive coupling in wire harnesses as well as conductive coupling (common ground) can transmit these transients to the inputs and outputs of automotive electronics.

The Society of Automotive Engineers (SAE) documented a table describing automotive transients (see Table 2) which is useful when trying to provide transient protection.

Considerable variation has been observed while gathering data on automobile transients. All automobiles have their electrical systems set up differently and it is not the intent of this paper to suggest a protection level that is required. There will always be a trade-off between cost of the suppressor and the level of protection obtained. The concept of one master suppressor placed on the main power lines is the most cost-effective scheme possible since individual suppressors at the various electronic devices will each have to suppress the largest transient that is likely to appear (Load Dump), hence each individual suppressor would have to be the same size as the one master suppressor since it is unlikely for several suppressors to share the transient discharge.

Table 2. Typical Transients Encountered in the Automotive Environment

Length of Transient	Cause	Energy Capability	Possible Frequency of Application
		Voltage Amplitude	
Steady State	Failed Voltage Regulator	∞	Infrequent
		+18 V	
5 Minutes	Booster starts with 24 V battery	∞	Infrequent
		± 24 V	
4.5–100 ms	Load Dump — i.e., disconnection of battery during high charging rates	≥ 10 J	Infrequent
		≤ 125 V	
≤ 0.32 s	Inductive Load Switching Transient	< 1 J	Often
		–300 V to +80 V	
≤ 0.2 s	Alternator Field Decay	< 1 J	Each Turn-Off
		–100 V to –40 V	
90 ms	Ignition Pulse Disconnected Battery	< 0.5 J	≤ 500 Hz Several Times in vehicle life
		≤ 75 V	
1 ms	Mutual Coupling in Harness	< 1 J	Often
		≤ 200 V	
15 μ s	Ignition Pulse Normal	< 0.001 J	3 500 Hz Continuous
		3 V	
	Accessory Noise	≤ 1.5 V	50 Hz to 10 kHz
	Transceiver Feedback	≈ 20 mV	R.F.

There will, of course, be instances where a need for individual suppressors at the individual accessories will be required, depending on the particular wiring system or situation.

TRANSIENT SUPPRESSOR TYPES

Carbon Block Spark Gap

This is the oldest and most commonly used transient suppressor in power distribution and telecommunication systems. The device consists of two carbon block electrodes separated by an air gap, usually 3 to 4 mils apart. One electrode is connected to the system ground and the other to the signal cable conductor. When a transient over-voltage appears, its energy is dissipated in the arc that forms between the two electrodes, a resistor in series with the gap, and also in the transient's source impedance, which depends on conductor length, material and other parameters.

The carbon block gap is a fairly inexpensive suppressor but it has some serious problems. One is that it has a relatively short service life and the other is that there are large variations in its arcing voltage. This is the major problem since a nominal 3 mil gap will arc anywhere from 300 to 1000 volts. This arcing voltage variation limits its use mainly to primary transient suppression with more accurate suppressors to keep transient voltages below an acceptable level.

Gas Tubes

The gas tube is another common transient suppressor, especially in telecommunication systems. It is made of two metallic conductors usually separated by 10 to 15 mils encapsulated in a glass envelope which is filled with several gases at low pressure. Gas tubes have a higher current carrying capability and longer life than carbon block gaps. The possibility of seal leakage and the resultant loss of protection has limited the use of these devices.

Selenium Rectifiers

Selenium transient suppressors are selenium rectifiers used in the reverse breakdown mode to clamp voltage transients. Some of these devices have self-healing properties which allows the device to survive energy discharges greater than their maximum capability for a limited number of surges. Selenium rectifiers do not have the voltage clamping capability of zener diodes. This is causing their usage to become more and more limited.

METAL OXIDE VARISTORS (MOV'S)

An MOV is a non-linear resistor which is voltage dependent and has electrical characteristics similar to back-to-back zener diodes. As its name implies it is made up of metal oxides, mostly zinc oxide with other oxides added to control electrical characteristics. MOV characteristics are compared to back-to-back zeners in Photos 2 through 7.

When constructing MOV's the metal oxides are sintered at high temperatures to produce a polycrystalline structure of conductive zinc oxide separated by highly resistive

intergranular boundaries. These boundaries are the source of the MOV's non-linear electrical behavior.

MOV electrical characteristics are mainly controlled by the physical dimensions of the polycrystalline structure since conduction occurs between the zinc oxide grains and the intergranular boundaries which are distributed throughout the bulk of the device.

The MOV polycrystalline body is usually formed into the shape of a disc. The energy rating is determined by the device's volume, voltage rating by its thickness, and current handling capability by its area. Since the energy dissipated in the device is spread throughout its entire metal oxide volume, MOV's are well suited for single shot high power transient suppression applications where good clamping capability is not required.

The major disadvantages with using MOV's are that they can only dissipate relatively small amounts of average power and are not suitable for many repetitive applications. Another drawback with MOV's is that their voltage clamping capability is not as good as zeners, and is insufficient in many applications.

Perhaps the major difficulty with MOV's is that they have a limited life time even when used below their maximum ratings. For example, a particular MOV with a peak current handling capability of 1000 A has a lifetime of about 1 surge at 1000 A_{pk} , 100 surges at 100 A_{pk} and approximately 1000 surges at 65 A_{pk} .

TRANSIENT SUPPRESSION USING ZENERS

Zener diodes exhibit a very high impedance below the zener voltage (V_Z), and a very low impedance above V_Z . Because of these excellent clipping characteristics, the zener diode is often used to suppress transients. Zeners are intolerant of excessive stress so it is important to know the power handling capability for short pulse durations.

Most zeners handle less than their rated power during normal applications and are designed to operate most effectively at this low level. Zener transient suppressors such as the ON Semiconductor 1N6267 Mosorb series are designed to take large, short duration power pulses.

This is accomplished by enlarging the chip and the effective junction area to withstand the high energy surges. The package size is usually kept as small as possible to provide space efficiency in the circuit layout, and since the package does not differ greatly from other standard zener packages, the steady state power dissipation does not differ greatly.

Some data sheets contain information on short pulse surge capability. When this information is not available for ON Semiconductor devices, Figure 8 can be used. This data applies for non-repetitive conditions with a lead temperature of 25°C.

It is necessary to determine the pulse width and peak power of the transient being suppressed when using Figure 8. This can be done by taking whatever waveform

the transient is and approximating it with a rectangular pulse with the same peak power. For example, an exponential discharge with a 1 ms time constant can be approximated by a rectangular pulse 1 ms wide that has the same peak power as the transient. This would be a better approximation than a rectangular pulse 10 ms wide with a correspondingly lower amplitude. This is because the heating effects of different pulse width lengths affect the power handling capability, as can be seen by Figure 8. This also represents a conservative approach because the energy of an exponential discharge will contain $\approx 1/2$ the energy of a rectangular pulse with the same pulse width and amplitude.

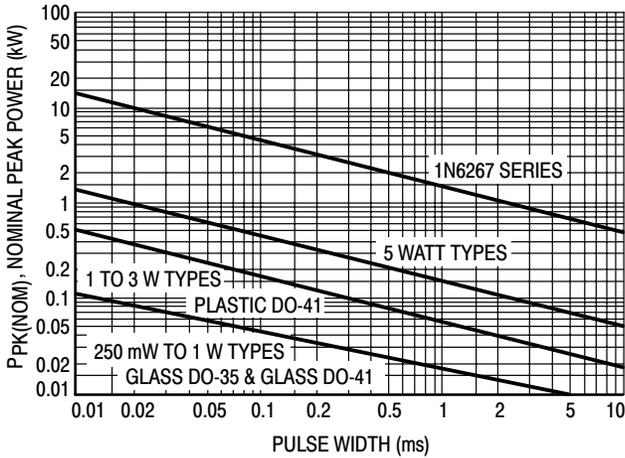


Figure 8. Peak Power Ratings of Zener Diodes

When used in repetitive applications, the peak power must be reduced as indicated by the curves of Figure 9. Average power must be derated as the lead or ambient temperature exceeds 25°C. The power derating curve normally given on data sheets can be normalized and used for this purpose.

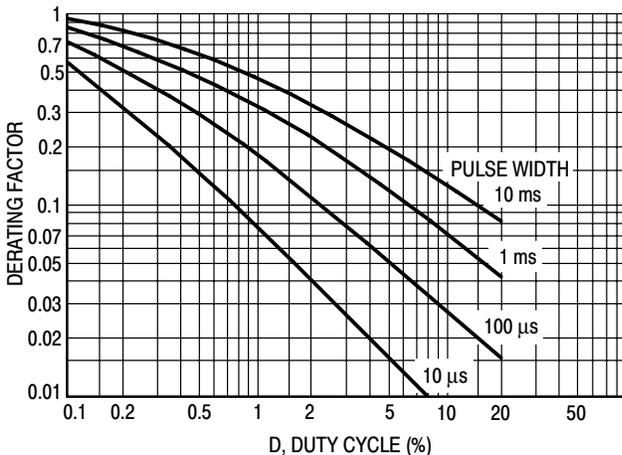


Figure 9. Typical Derating Factor for Duty Cycle

The peak zener voltage during the peak current of the transient being suppressed can be related to the nominal zener voltage (Eqtn 1) by the clamping factor (F_C).

$$\text{Eqtn 1: } V_{Z(pK)} = F_C (V_{Z(nom)})$$

Unless otherwise specified F_C is approximately 1.20 for zener diodes when operated at their pulse power limits.

For example, a 5 watt, 20 volt zener can be expected to show a peak voltage of 24 volts regardless of whether it is handling 450 watts for 0.1 ms or 50 watts for 10 ms. (See Figure 8.)

This occurs because the zener voltage is a function of both junction temperature and IR drop. Longer pulse widths cause a greater junction temperature rise than short ones; the increase in junction temperature slightly increases the zener voltage. This increase in zener voltage due to heating is roughly offset by the fact that longer pulse widths of identical energy content have lower peak currents. This results in a lower IR drop (zener voltage drop) keeping the clamping factor relatively constant with various pulse widths of identical energy content.

An approximation of zener impedance is also helpful in the design of transient protection circuits. The value of $R_{Z(nom)}$ (Eqtn 2) is approximate because both the clamping factor and the actual resistance is a function of temperature.

$$\text{Eqtn 2: } R_{Z(nom)} = \frac{V_{Z(nom)}^2 (F_C - 1)}{P_{pK(nom)}}$$

$V_{Z(nom)}$ = Nominal Zener Voltage

$P_{pK(nom)}$ = Found from Figure 8 when device type and pulse width are known. For example, from Figure 8 a 1N6267 zener suppressor has a $P_{pK(nom)}$ of 1.5 kW at a pulse width of 1 ms.

As seen from equation 2, zeners with a larger $P_{pK(nom)}$ capability will have a lower $R_{Z(nom)}$.

ZENER versus MOV TRADEOFFS

The clamping characteristics of Zeners and MOV's are best compared by measuring their voltages under transient conditions. Photos 1 through 9 are the result of an experiment that was done to compare the clamping characteristics of a Zener (ON Semiconductor 1N6281, approximately 1.5J capability) with those of an MOV (G.E. V27ZA4, 4J capability); both are 27 V devices.

Photo 1 shows the pulse generator output voltage. This generator synthesizes a transient pulse that is characteristic of those that may appear in the real world.

Photos 2 and 3 are clamping voltages of the MOV and Zener, respectively with a surge source impedance of 500 Ω.

Photos 4 and 5 are the clamping voltages with a surge source impedance of 50 Ω.

Photos 6 and 7 simulate a condition where the surge source impedance is 5 Ω.

Photos 8 and 9 show a surge source impedance of 0.55 Ω, which is at the limits of the Zener suppressor's capability.

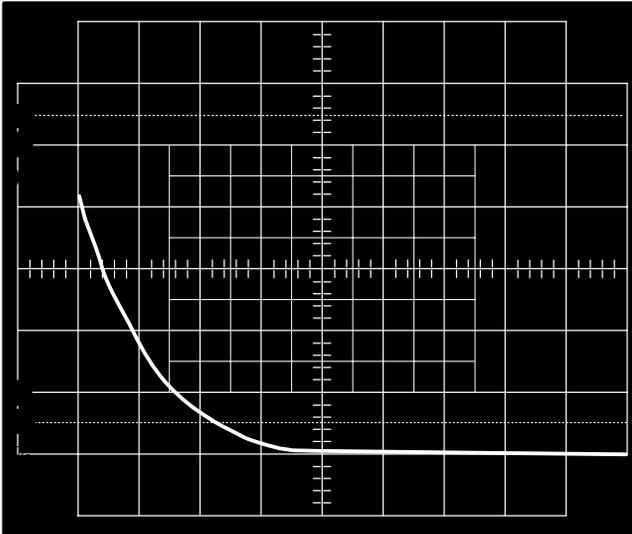


PHOTO 1
Open Circuit Transient Pulse
Vert: 20 V/div
Horiz: 0.5 ms/div
 $V_{\text{peak}} = 90 \text{ V}$

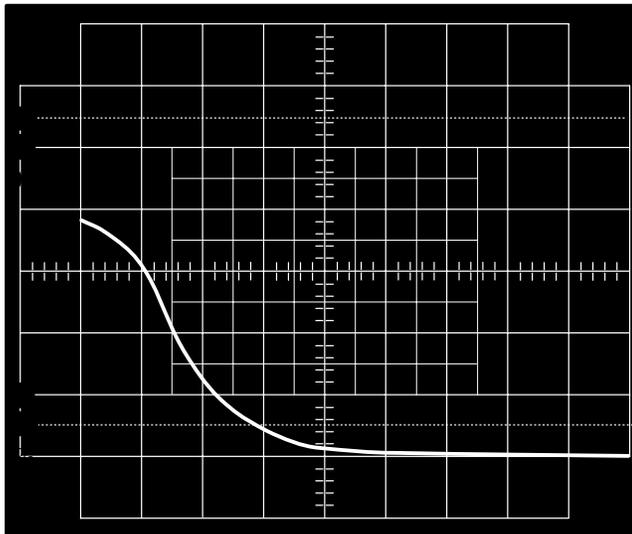


PHOTO 2
MOV (27 V)
Vert: 10 V/div
Horiz: 0.5 ms/div
Transient Source Impedance: 500 Ω
 $V_{\text{peak}} = 39.9 \text{ V}$

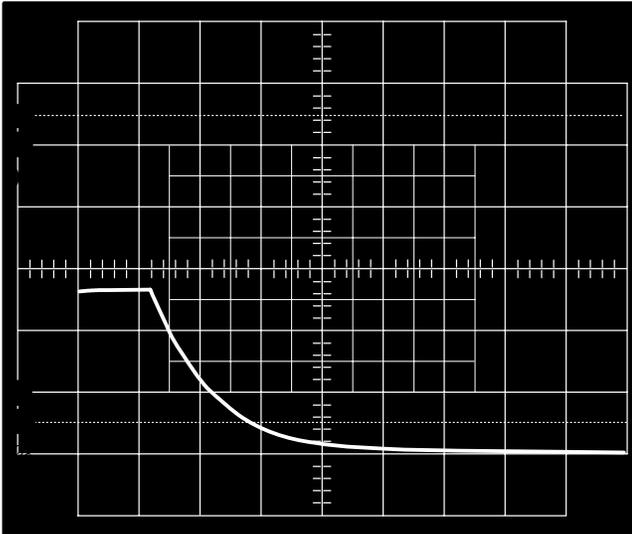


PHOTO 3
Zener (27 V)
Vert: 10 V/div
Horiz: 0.5 ms/div
Transient Source Impedance: 500 Ω
 $V_{\text{peak}} = 27 \text{ V}$

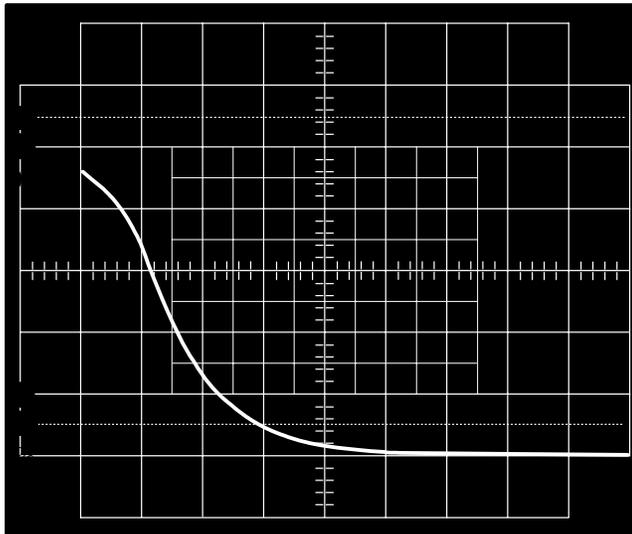


PHOTO 4
MOV (27 V)
Vert: 10 V/div
Horiz: 0.5 ms/div
Transient Source Impedance: 50 Ω
 $V_{\text{peak}} = 44.7 \text{ V}$

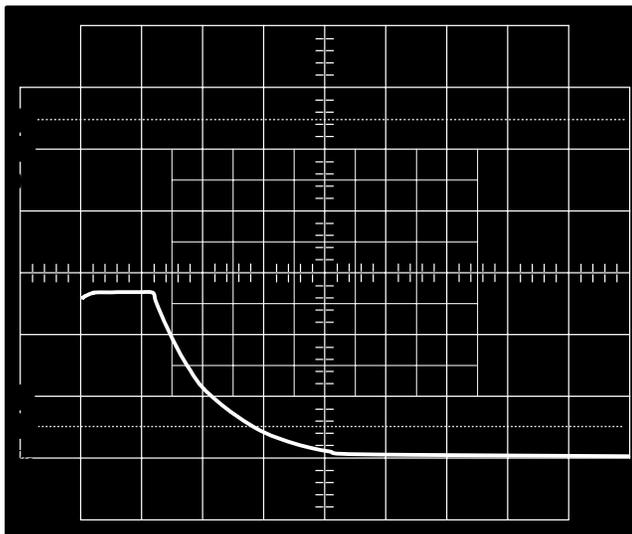


PHOTO 5
Zener (27 V)
Vert: 10 V/div
Horiz: 0.5 ms/div
Transient Source Impedance: 50 Ω
 $V_{\text{peak}} = 27 \text{ V}$

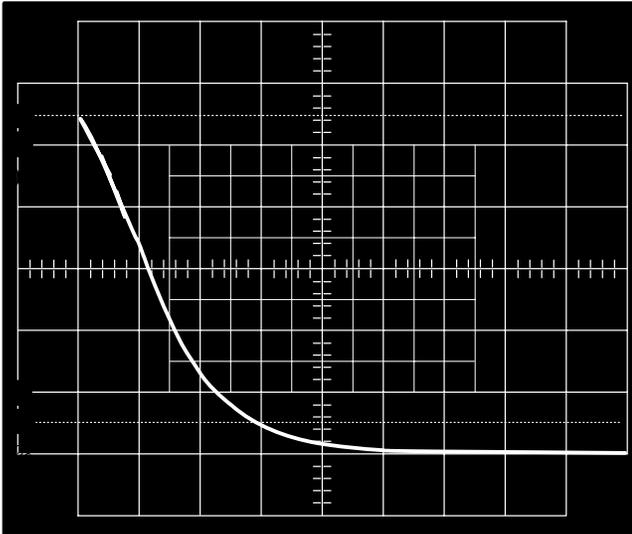


PHOTO 6
 MOV (27 V)
 Vert: 10 V/div
 Horiz: 0.5 ms/div
 Transient Source Impedance: 5 Ω
 $V_{\text{peak}} = 52 \text{ V}$

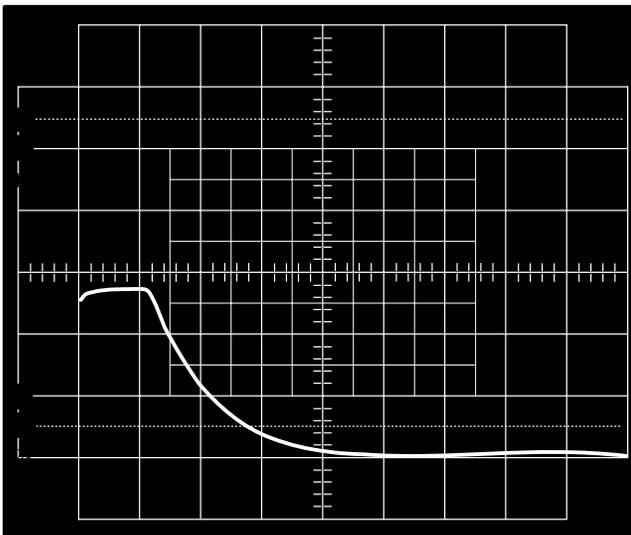


PHOTO 7
 Zener (27 V)
 Vert: 10 V/div
 Horiz: 0.5 ms/div
 Transient Source Impedance: 5 Ω
 $V_{\text{peak}} = 28 \text{ V}$

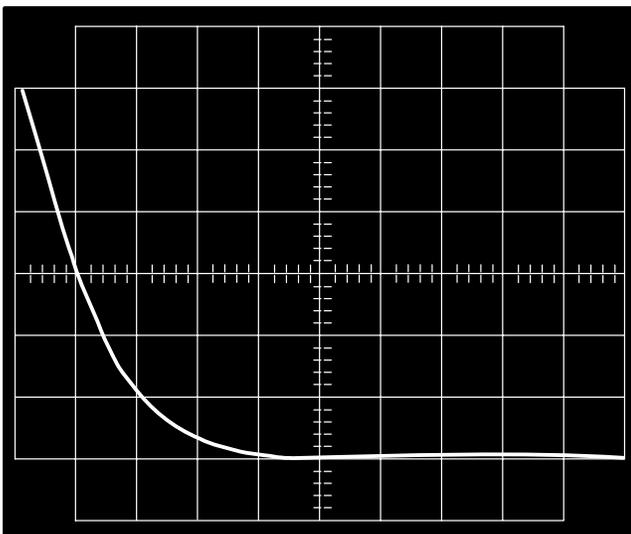


PHOTO 8
 MOV (27 V)
 Vert: 10 V/div
 Horiz: 0.5 ms/div
 Transient Source Impedance: 0.55 Ω
 $V_{\text{peak}} = 62.5 \text{ V}$

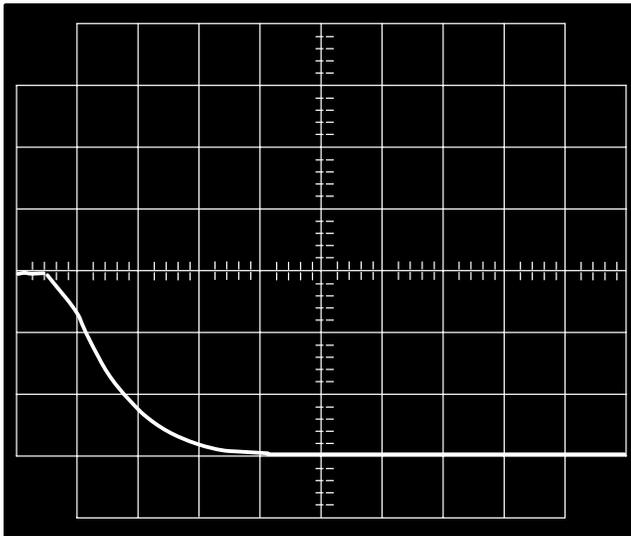


PHOTO 9
 Zener (27 V)
 Vert: 10 V/div
 Horiz: 0.5 ms/div
 Transient Source Impedance: 0.55 Ω
 V_{peak} : 30.2 V
 Peak Power: Approx 2000 W_{peak}
 (The limit of this device's capability)

As can be seen by the photographs, the Zener suppressor has significantly better voltage clamping characteristics than the MOV even though that particular Zener has less than one-fourth the energy capability of the MOV it was compared with. However, the energy rating can be misleading because it is based on the clamp voltage times the surge current, and when using an MOV, the high impedance results in a fairly high clamp voltage. The major tradeoff with using a zener type suppressor is its cost versus power handling capability, but since it would take an “oversized” MOV to clamp voltages (suppress transients) as well as the zener, the MOV begins to lose its cost advantage.

If a transient should come along that exceeds the capabilities of the particular Zener, or MOV, suppressor that was chosen, the load will still be protected, since they both fail short.

The theoretical reaction time for Zeners is in the picosecond range, but this is slowed down somewhat with lead and package inductance. The 1N6267 Mosorb series of transient suppressors have a typical response time of less than one nanosecond. For very fast rising transients it is important to minimize external inductances (due to wiring, etc.) which will minimize overshoot.

Connecting Zeners in a back-to-back arrangement will enable bidirectional voltage clamping characteristics. (See Figure 10.)

If Zeners A and B are the same voltage, a transient of either polarity will be clamped at approximately that voltage since one Zener will be in reverse bias mode while the other will be in the forward bias mode. When clamping low voltage it may be necessary to consider the forward drop of the forward biased Zener.

The typical protection circuit is shown in Figure 11a. In almost every application, the transient suppression device is placed in parallel with the load, or component to be protected. Since the main purpose of the circuit is to clamp the voltage appearing across the load, the suppressor should

be placed as close to the load as possible to minimize overshoot due to wiring (or any inductive) effect. (See Figure 11b.)

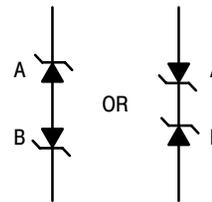


Figure 10. Zener Arrangement for Bidirectional Clamping

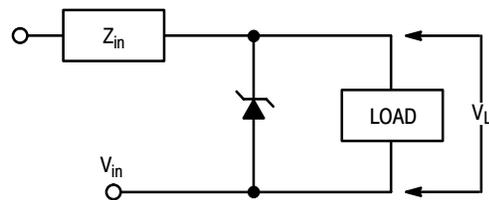


Figure 11a. Using Zener to Protect Load Against Transients

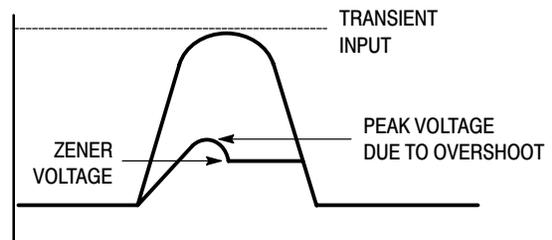


Figure 11b. Overshoot Due to Inductive Effect

Zener capacitance prior to breakdown is quite small (for example, the 1N6281 27 Volt Mosorb has a typical capacitance of 800 pF). Capacitance this small is desirable in the off-state since it will not attenuate wide-band signals.

When the Zener is in the breakdown mode of operation (e.g. when suppressing a transient) its effective capacitance increases drastically from what it was in the off-state. This makes the Zener ideal for parallel protection schemes since, during transient suppression, its large effective capacitance will tend to hold the voltage across the protected element constant; while in the off-state (normal conditions, no transient present), its low off-state capacitance will not attenuate high frequency signals.

Input impedance (Z_{in}) always exists due to wiring and transient source impedance, but Z_{in} should be increased as much as possible with an external resistor, if circuit constraints allow. This will minimize Zener stress.

CONCLUSION

The reliable use of semiconductor devices requires that the circuit designer consider the possibility of transient overvoltages destroying these transient-sensitive components.

These transients may be generated by normal circuit operations such as inductive switching circuits, energizing and deenergizing transformer primaries, etc. They do not present much of a problem since their energy content, duration and effect may easily be obtained and dealt with.

Random transients found on power lines, or lightning transients, present a greater threat to electronic components since there is no way to be sure when or how severe they will be. General guidelines were discussed to aid the circuit designer in deciding what size (capability and cost) suppressor to choose for a certain level of protection. There will always be a tradeoff between suppressor price and protection obtained.

Several different suppression devices were discussed with emphasis on Zeners and MOV's, since these are the most popular devices to use in most applications.

REFERENCES

1. GE Transient Voltage Suppression Manual, 2nd edition.
2. ON Semiconductor Zener Diode Manual.

DESIGN CONSIDERATIONS AND PERFORMANCE OF ON SEMICONDUCTOR TEMPERATURE-COMPENSATED ZENER (REFERENCE) DIODES

Prepared by
Zener Diode Engineering
and
Ronald N. Racino
Reliability and Quality Assurance

INTRODUCTION

This application note defines ON Semiconductor temperature-compensated zener (reference) diodes, explains the device characteristics, describes electrical testing, and discusses the advanced concepts of device reliability and quality assurance. It is a valuable aid to those who contemplate designing circuits requiring the use of these devices.

Zener diodes fall into three general classifications: Regulator diodes, reference diodes and transient voltage suppressors. Regulator diodes are normally employed in power supplies where a nearly constant dc output voltage is required despite relatively large changes in input voltage or load resistance. Such devices are available with a wide range of voltage and power ratings, making them suitable for a wide variety of electronic equipments.

Regulator diodes, however, have one limitation: They are temperature-sensitive. Therefore, in applications in which the output voltage must remain within narrow limits during input-voltage, load-current, and temperature changes, a temperature-compensated regulator diode, called a reference diode, is required.

The reference diode is made possible by taking advantage of the differing thermal characteristics of forward- and reverse-biased silicon p-n junctions. A forward-biased junction has a negative temperature coefficient of approximately 2 mV/°C, while reverse-biased junctions have positive temperature coefficients ranging from about 2 mV/°C at 5.5 V to 6 mV/°C at 10 V. Therefore it is possible, by judicious combination of forward- and reverse-biased junctions, to fabricate a device with a very low overall temperature coefficient (Figure 1).

The principle of temperature compensation is further illustrated in Figure 2, which shows the voltage-current characteristics at two temperature points (25 and 100°C) for both a forward- and a reverse-biased junction. The diagram shows that, at the specified test current (I_{ZT}), the absolute value of voltage change (ΔV) for the temperature change between 25 and 100°C is the same for both junctions. Therefore, the total voltage across the combination of these two junctions is also the same at these temperature points, since one ΔV is negative and the other is positive. However, the rate of voltage change with temperature over the

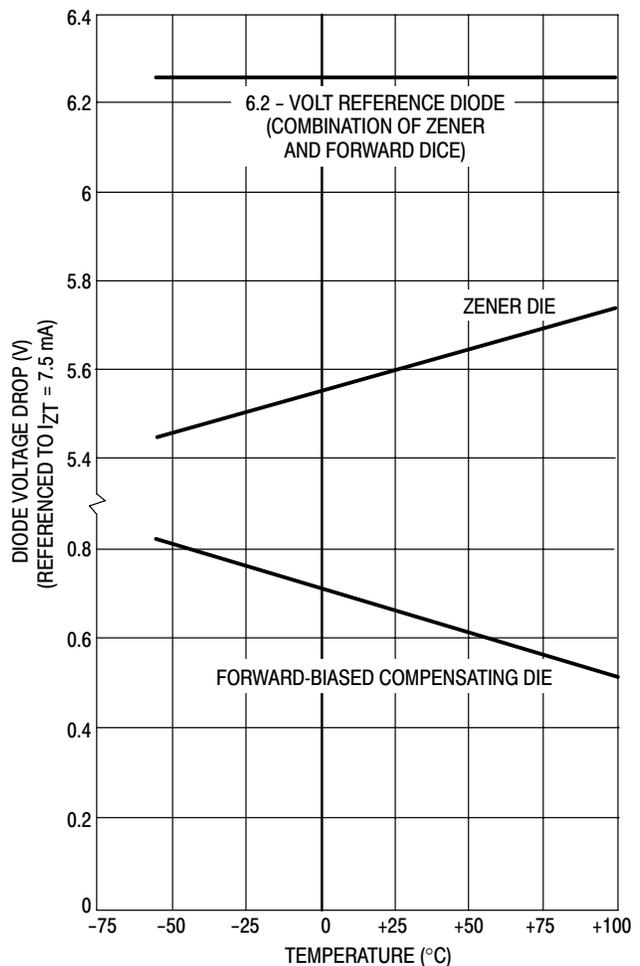


Figure 1. Temperature Compensation of a 6.2 Volt Reference Diode (1N821 Series)

temperature range defined by these points is not necessarily the same for both junctions, thus the temperature compensation may not be linear over the entire range.

Figure 2 also indicates that the voltage changes of the two junctions are equal and opposite only at the specified test current. For any other value of current, the temperature compensation may not be complete.

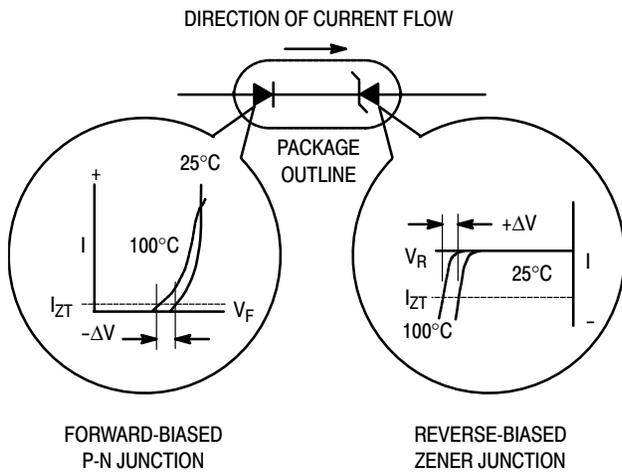


Figure 2. Temperature Compensation of P-N Junctions

IMPORTANT ELECTRICAL CHARACTERISTICS OF REFERENCE DIODES

The three most important characteristics of reference diodes are 1) reference voltage, 2) voltage-temperature stability, and 3) voltage-time stability.

1. Reference Voltage. This characteristic is defined as the voltage drop measured across the diode when the specified test current passes through it in the zener direction. It is also called the zener voltage (V_Z , Figure 3). On the data sheets, the reference voltage is given as a nominal voltage for each family of reference diodes.

The nominal voltages are normally specified to a tolerance of $\pm 5\%$, but devices with tighter tolerances, such as $\pm 2\%$ and $\pm 1\%$, are available on special order.

2. Voltage-Temperature Stability. The temperature stability of zener voltage is sometimes expressed by means of the temperature coefficient. This parameter is usually defined as the percent voltage change across the device per degree centigrade. This method of indicating voltage stability accurately reflects the voltage deviation at the test temperature extremes but not necessarily at other points within the specified temperature range. This fact is due to variations in the rate of voltage change with temperature for the forward- and reverse-biased dice of the reference diode. Therefore, the temperature coefficient is given in ON Semiconductor data sheets only as a quick reference, for designers who are accustomed to this method of specification.

A more meaningful way of defining temperature stability is the “box method.” This method, used by ON Semiconductor, guarantees that the zener voltage will not vary by more than a specified amount over a specified temperature range at the indicated test current, as verified by tests at several temperatures within this range.

Some devices are accurately compensated over a wide temperature range (-55°C to 100°C), others over a narrower range (0 to 75°C). The wide-range devices are, as a rule, more expensive. Therefore, it would be economically wasteful for the designer to specify devices with a temperature range much wider than actually required for the specific device application.

During actual production of reference diodes, it is difficult to predict the compensation accuracy. In the interest of maximum economy, it is common practice to test all devices coming off the production line, and to divide the production lot into groups, each with a specified maximum ΔV_Z . Each group, then, is given a different device type number.

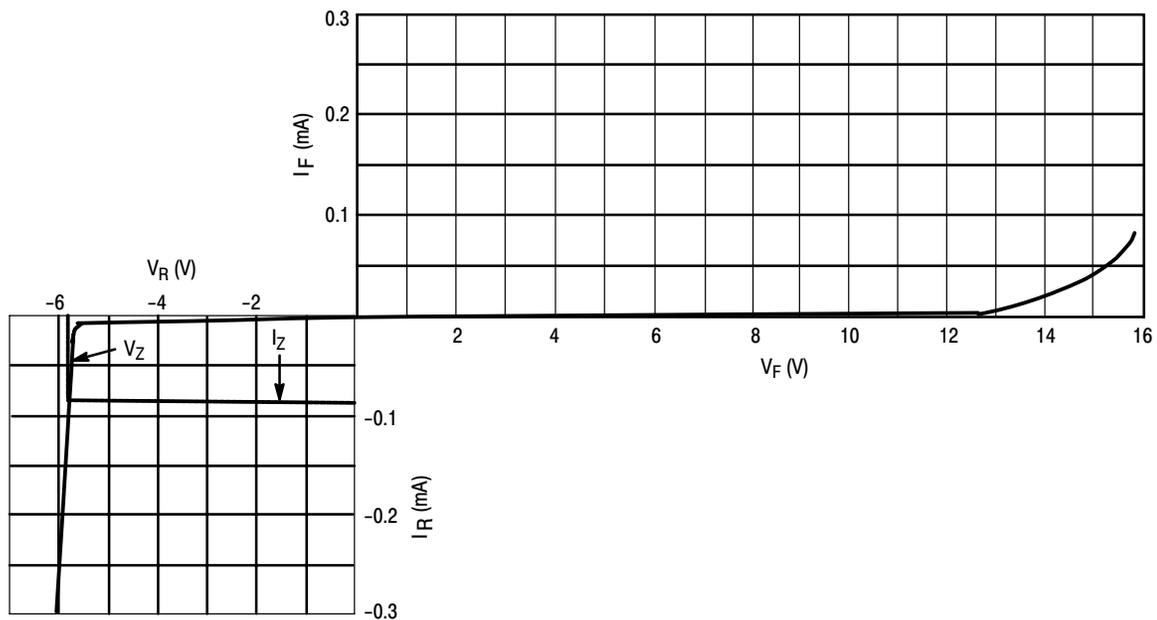
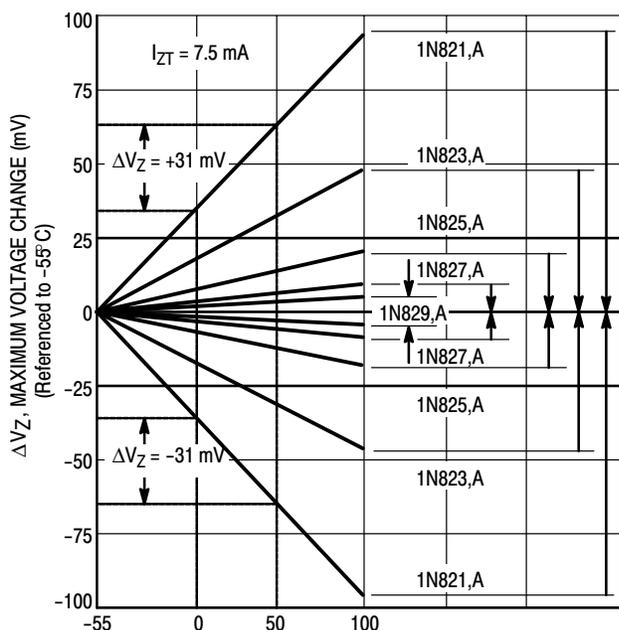


Figure 3. Typical Voltage – Current Characteristic of Reference Diodes

On the data sheet, the voltage-temperature characteristics of the most widely used device types are illustrated in a graph similar to the one shown in Figure 4. The particular production line represented in this figure produces 6.2 volt devices, but the line yields five different device type numbers (1N821 through 1N829), each with a different temperature coefficient. The 1N829, for example, has a maximum voltage change of less than 5 mV over a temperature range of -55 to $+100^{\circ}\text{C}$, while the 1N821 may have a voltage change of up to 96 mV over the same temperature range.



In the past, design data and characteristic curves on data sheets for reference diodes have been somewhat limited: The devices have been characterized principally at the recommended operating point. ON Semiconductor has introduced a data sheet, providing device data previously not available, and showing limit curves that permit worst-case circuit design without the need for associated tests required in conjunction with the conventional data sheets.

Figure 4. Temperature Dependence of Zener Voltage (1N821 Series)

Graphs such as these permit the selection of the lowest-cost device that meets a particular requirement. They also permit the designer to determine the maximum voltage change of a particular reference diode for a relatively small change in temperature. This is done by drawing vertical lines from the desired temperature points at the abscissa of the graph to intersect with each the positive- and negative-going curves of the particular device of interest. Horizontal lines are then drawn from these intersects to the ordinate of the graph. The difference between the intersections of these horizontal lines with the ordinate yields the maximum voltage change over the temperature increment. For example, for the 1N821, a change in ambient temperature from 0 to 50°C results in a voltage change of no more than about ± 31 mV.

The reason that the device reference voltage may change in either the negative or positive direction is that after assembly, some of the devices within a lot may be overcompensated while others may be undercompensated. In any design, the “worst-case” condition must be considered. Therefore, in the above example, it can be assumed that the maximum voltage change will not exceed 31 mV.

It should be understood, however, that the above calculations give the maximum possible voltage change for the device type, and by no means the actual voltage change for the individual unit.

3. Voltage-Time Stability. The voltage-time stability of a reference diode is defined by the voltage change during operating time at the standard test current (I_{ZT}) and test temperature (T_A). In general, the voltage stability of a reference diode is better than 100 ppm per 1000 hours of operation.

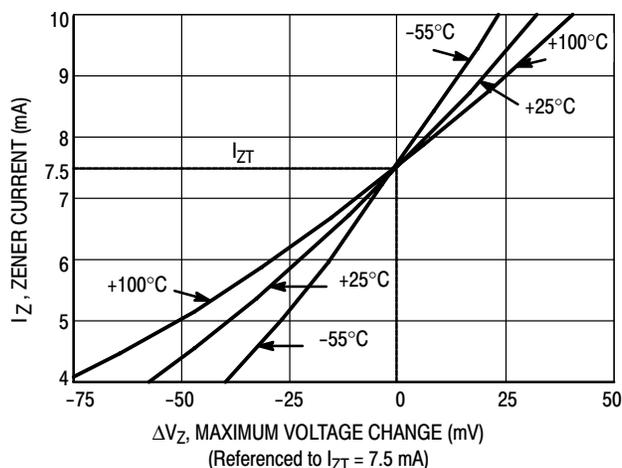


Figure 5. Current Dependence of Zener Voltage at Various Temperatures (1N821 Series)

THE EFFECT OF CURRENT VARIATION ON ZENER VOLTAGE

The nominal zener voltage of a reference diode is specified at a particular value of current, called the zener test current (I_{ZT}). All measurements of voltage change with temperature are referenced to this test current. If the operating current is varied, all these specifications will change.

The effect of current variation on zener voltage, at various temperatures, is graphically illustrated on the 1N821 data sheet as “Zener Current versus Maximum Voltage Change.” A typical example of such a graph is shown for the 1N821 series in Figure 5. The voltage change shown is due entirely to the impedance of the device at the fixed temperature. It does not reflect the change in reference voltage due to the change in temperature since each curve is referenced to $I_{ZT} = 7.5$ mA at the indicated temperature. As shown, the

greatest voltage change occurs at the highest temperature represented in the diagram. (See “Dynamic Impedance” under the next section).

Figure 5 shows that, at 25°C, a change in zener current from 4 to 10 mA causes a voltage shift of about 90 mV. Comparing this value with the voltage-change example in Figure 4 (31 mV), it is apparent that, in general, a greater voltage variation may be due to current fluctuations than to temperature change. Therefore, good current regulation of the source should be a major consideration when using reference diodes in critical applications.

It is not essential, however, that a reference diode be operated at the specified test current. The new voltage-temperature characteristics for a change in current can be obtained by superimposing the data of Figure 5 on that of Figure 4. A new set of characteristics, at a test current of 4 mA, is shown for the 1N823 in Figure 6, together with the original characteristics at 7.5 mA.

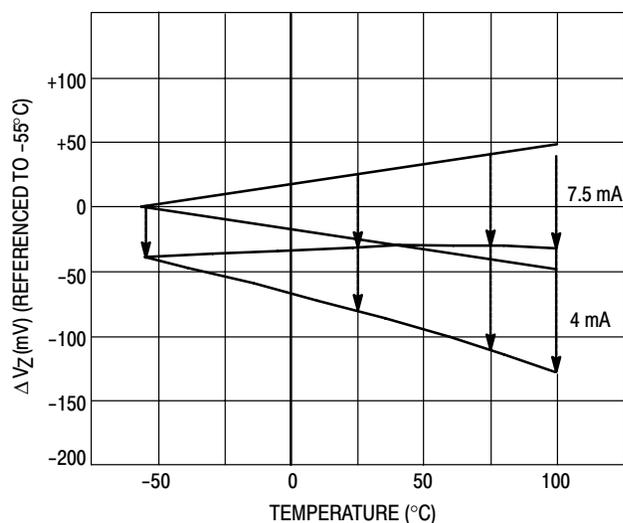


Figure 6. Voltage Change with Temperature for 1N823 at Two Different Current Levels

From these characteristics, it is evident that the voltage change with temperature for the new curves is different from that for the original ones. It is also apparent that if the test current varies between 7.5 and 4 mA, the voltage changes would lie along the dashed lines belonging to the given temperature points. This clearly shows the need for a well-regulated current source.

It should be noted, however, that even when a well-regulated current supply is available, other factors might influence the current flowing through a reference diode. For example, to minimize the effects of temperature-sensitive passive elements in the load circuit on current regulation, it is desirable that the load in parallel with the reference diode have an impedance much higher than the dynamic impedance of the reference diode.

OTHER CHARACTERISTICS

In addition to the three major characteristics discussed earlier, the following parameters and ratings of reference diodes may be considered in some applications.

Power Dissipation

The maximum dc power dissipation indicates the power level which, if exceeded, may result in the destruction of the device. Normally a device will be operated near the specified test current for which the data-sheet specifications are applicable. This test current is usually much below the current level associated with the maximum power dissipation.

Dynamic Impedance

Zener impedance may be construed as composed of a current-dependent resistance shunted by a voltage-dependent capacitance. Figure 7 indicates the typical variations of dynamic zener impedance (Z_Z) with current and temperature for the 1N821 reference diode series. These diagrams are given in the 1N821 data sheet. As shown, the zener impedance decreases with current but increases with ambient temperature.

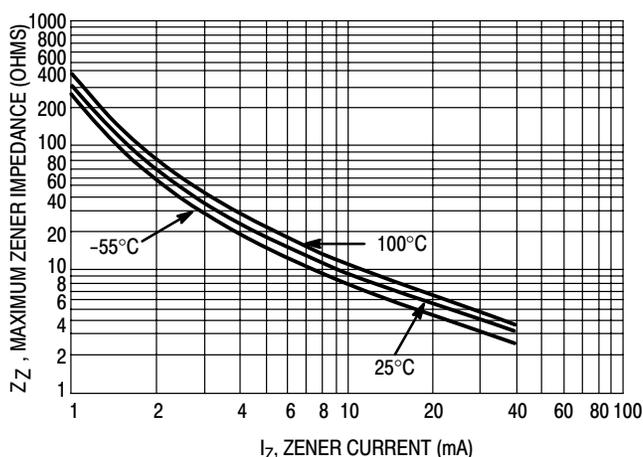


Figure 7. Variation of Zener Impedance With Current and Temperature (1N821 Series)

The impedance of a reference diode is normally specified at the test current (I_{ZT}). It is determined by measuring the ac voltage drop across the device when a 60 Hz ac current with an rms value equal to 10% of the dc zener current is superimposed on the zener current (I_{ZT}). Figure 8 shows the block diagram of a circuit used for testing zener impedance.

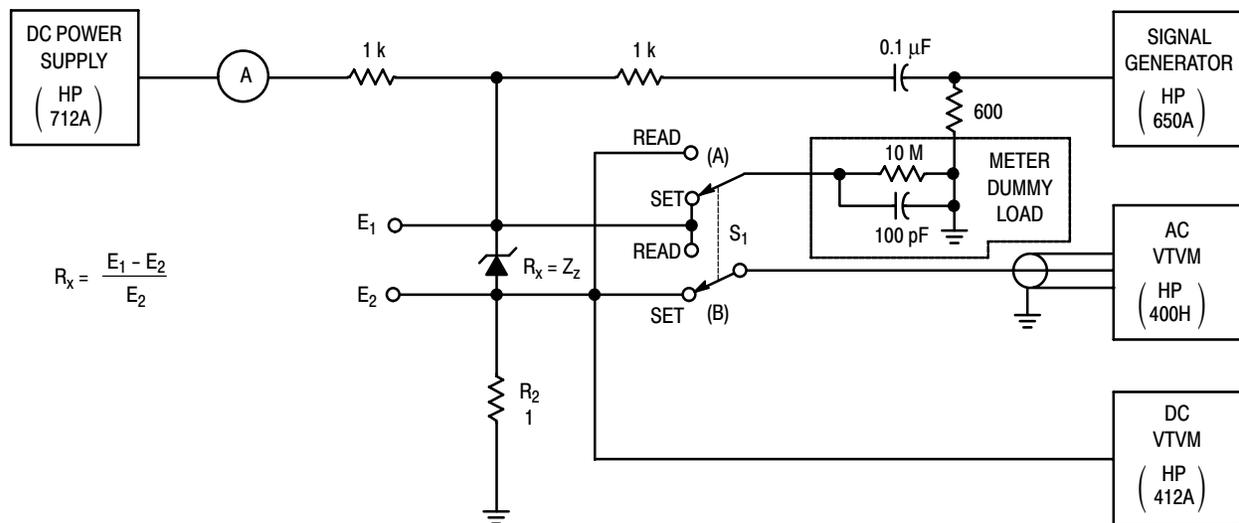


Figure 8. Block Diagram of Test Circuit for Measuring Dynamic Zener Impedance

ELECTRICAL TESTING

All devices are tested electrically as a last step in the manufacturing process.

The subsequent final test procedures represent an automated and accurate method of electrically classifying reference diodes. First, an electrical test is performed on all devices to insure the correct voltage-breakdown and stability characteristics. Next, the breakdown voltage and dynamic impedance are measured. Finally, the devices are placed in an automatic data acquisition system that automatically cycles them through the complete temperature range specified. The actual voltage measurements at the various temperature points are retained in the system computer memory until completion of the full temperature excursion. The computer then calculates the changes in voltage for each device at each test temperature and classifies all units on test into the proper category. The system provides a printed readout for every device, including the voltage changes to five digits during temperature cycling, and the corresponding EIA type number, as well as the data referring to test conditions such as device position, lot number, and date.

DEVICE RELIABILITY AND QUALITY ASSURANCE

Insuring a very low failure rate requires maximum performance in all areas effecting device reliability: Device design, manufacturing processes, quality control, and reliability testing. ON Semiconductor's basic reliability concept is based on the belief that reference diode reliability is a complex yet controllable function of all these variables.

Under this "total reliability" concept, ON Semiconductor can mass-produce high-reliability reference diodes.

The reliability of a reference diode fundamentally depends upon the device design, regardless of the degree of effort put into device screening and circuit designing. Therefore, reliability measures must be incorporated at the

device design and process development stages to establish a firm foundation for a comprehensive reliability program. The design is then evaluated by thorough reliability testing, and the results are supplied to the Design Engineering department. This closed-loop feedback procedure provides valuable information necessary to improve important design features such as electrical instability due to surface effects, mechanical strength, and uniformly low thermal resistance between the die and ambient environment.

Process Control

There are more than 2000 variables that must be kept under control to fabricate a reliable reference diode. The in-process quality control group controls most of these variables. It places a strict controls on all aspects of manufacturing from materials procurement to the finished product. Included in this broad spectrum of controls are:

- **Materials Control.** All materials purchased or fabricated in-plant are checked against rigid specifications. A quality check on vendors' products is kept up to date to insure that only materials of a proven quality level will be purchased.

- **In-Process Inspection and Control.** Numerous on-line inspection stations maintain a statistical process control program on specific manufacturing processes. If any of these processes are found to be out of control, the discrepant material is diverted from the normal production flow and the cognizant design engineer notified. Corrective action is initiated to remedy the cause of the discrepancy.

Reliability Testing

The Reliability Engineering group evaluates all new products and gives final conclusions and recommendations to the device design engineer. The Reliability Engineering group also performs independent testing of all products and includes, as part of this testing program, step-stress-to-failure testing to determine the maximum capabilities of the product.

SOME STRAIGHT TALK ABOUT MOSORBs TRANSIENT VOLTAGE SUPPRESSORS

INTRODUCTION

Distinction is sometimes made between devices trademarked Mosorb (by ON Semiconductor Inc.), and standard zener/avalanche diodes used for reference, low-level regulation and low-level protection purposes. It must be emphasized from the beginning that Mosorb devices are, in fact, zener diodes. The basic semiconductor technology and processing are identical. The primary difference is in the applications for which they are designed. Mosorb devices are intended specifically for transient protection purposes and are designed, therefore, with a large effective junction area that provides high pulse power capability while minimizing the total silicon use. Thus, Mosorb pulse power ratings begin at 500 watts – well in excess of low power conventional zener diodes which in many cases do not even include pulse power ratings among their specifications.

MOVs, like Mosorbs, do have the pulse power capabilities for transient suppression. They are metal oxide varistors (not semiconductors) that exhibit bidirectional avalanche characteristics, similar to those of back-to-back connected zeners. The main attributes of such devices are low manufacturing cost, the ability to absorb high energy surges (up to 600 joules) and symmetrical bidirectional “breakdown” characteristics. Major disadvantages are: high clamping factor, an internal wear-out mechanism and an absence of low-end voltage capability. These limitations restrict the use of MOVs primarily to the protection of insensitive electronic components against high energy transients in application above 20 volts, whereas, Mosorbs are best suited for precise protection of sensitive equipment even in the low voltage range – the same range covered by conventional zener diodes. The relative features of the two device types are covered in Table 1.

IMPORTANT SPECIFICATIONS FOR MOSORB PROTECTIVE DEVICES

Typically, a Mosorb suppressor is used in parallel with the components or circuits being protected (Figure 1), in order to shunt the destructive energy spike, or surge, around the more sensitive components. It does this by avalanching at its “breakdown” level, ideally representing an infinite impedance at voltages below its rated breakdown voltage, and essentially zero impedance at voltages above this level.

In the more practical case, there are three voltage specifications of significance, as shown in Figure 1a.

- a) V_{RWM} is the maximum reverse stand-off voltage at which the Mosorb is cut off and its impedance is at its highest value – that is, the current through the device is essentially the leakage current of a back-biased diode.
- b) $V_{(BR)}$ is the breakdown voltage – a voltage at which the device is entering the avalanche region, as indicated by a slight (specified) rise in current beyond the leakage current.
- c) V_{RSM} is the maximum reverse voltage (clamping voltage) which is defined and specified in conjunction with the maximum reverse surge current so as not to exceed the maximum peak power rating at a pulse width (t_p) of 1 ms (industry std time for measuring surge capability).

RELATIVE FEATURES OF MOVs and MOSORBs

Table 1.	
MOV	Mosorb/Zener Transient Suppressor
<ul style="list-style-type: none"> • High clamping factor. • Symmetrically bidirectional. • Energy capability per dollar usually higher than a silicon device. However, if good clamping is required the energy capability would have to be grossly overspecified resulting in higher cost. • Inherent wear out mechanism clamp voltage degrades after every pulse, even when pulsed below rated value. • Ideally suited for crude ac line protection. • High single-pulse current capability. • Degrades with overstress. • Good high voltage capability. • Limited low voltage capability. 	<ul style="list-style-type: none"> • Very good clamping close to the operating voltage. • Standard parts perform like standard zeners. Symmetrical bidirectional devices available for many voltages. • Good clamping characteristic could reduce overall system cost. • No inherent wear out mechanism. • Ideally suited for precise DC protection. • Medium multiple-pulse current capability. • Fails short with overstress. • Limited high voltage capability unless series devices are used. • Good low voltage capability.

In practice, the Mosorb is selected so that its V_{RWM} is equal to or somewhat higher than the highest operating voltage required by the load (the circuits or components to be protected). Under normal conditions, the Mosorb is inoperative and dissipates very little power. When a transient occurs, the Mosorb converts to a very low dynamic impedance and the voltage across the Mosorb becomes the clamping voltage at some level above $V_{(BR)}$. The actual clamping level will depend on the surge current through the Mosorb. The maximum reverse surge current (I_{RSM}) is specified on the Mosorb data sheets at 1 ms and for a logarithmically decaying pulse waveform. The data sheet also contains curves to determine the maximum surge current rating at other time intervals.

Typically, Mosorb devices have a built-in safety margin at the maximum rated surge current because the clamp voltage, V_{RSM} , is itself, guardbanded. Thus, the parts will be operating below their maximum pulse-power (P_{pk}) rating even when operated at maximum reverse surge current).

If the transients are random in nature (and in many cases they are), determining the surge-current level can be a problem. The circuit designer must make a reasonable estimate of the proper device to be used, based on his knowledge of the system and the possible transients to be encountered. (e.g., transient voltage, source impedance and time, or transient energy and time are some characteristics that must be estimated). Because of the very low dynamic impedance of Mosorb devices in the region between $V_{(BR)}$ and V_{RSM} , the maximum surge current is dependent on, and limited by the external circuitry.

In cases where this surge current is relatively low, a conventional zener diode could be used in place of a Mosorb or other dedicated protective device with some possible savings in cost. The surge capabilities most of ON Semiconductor's zener diode lines are discussed in ON Semiconductor's Application Note AN784.

In the data sheets of some protective devices, the parameter for response time is emphasized. Response time on these data sheets is defined as the time required for the voltage across the protective device to rise from 0 to $V_{(BR)}$, and relates primarily to the effective series impedance associated with the device. This effective impedance is somewhat complex and changes drastically from the blocking mode to the avalanche mode. In most applications (where the protective device shunts the load) this response time parameter becomes virtually meaningless as indicated by the waveforms in Figures 1b and 1c. If the response time as defined is very long, it still would not affect the performance of the surge suppressor.

However, if the series inductance becomes appreciable, it could result in "overshoot" as shown in Figure 1d that would be detrimental to circuit protection. In Mosorb devices, series inductance is negligible compared to the inductive effects of the external circuitry (primarily lead lengths). Hence, Mosorbs contribute little or nothing to overshoot and, in essence, the parameter of response time has very little significance. However, care must be exercised in the design of the external circuitry to minimize overshoot.

SUMMARY

In selecting a protective device, it is important to know as much as possible about the transient conditions to be encountered. The most important device parameters are reverse working voltage (V_{RWM}), surge current (I_{RSM}), and clamp voltage (V_{RSM}). the product of V_{RSM} and I_{RSM} yields the peak power dissipation, which is one of the prime categories for device selection.

The selector guide, in this book, gives a broad overview of the Mosorb transient suppressors now available from ON Semiconductor. For more detailed information, please contact your ON Semiconductor sales representative or distributor.

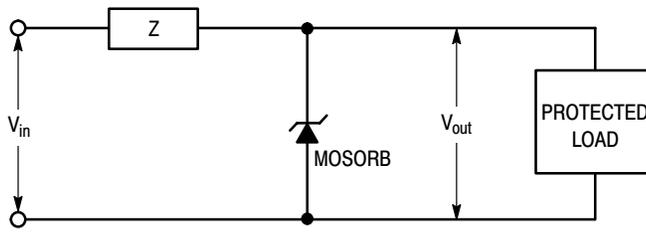


Figure 1.

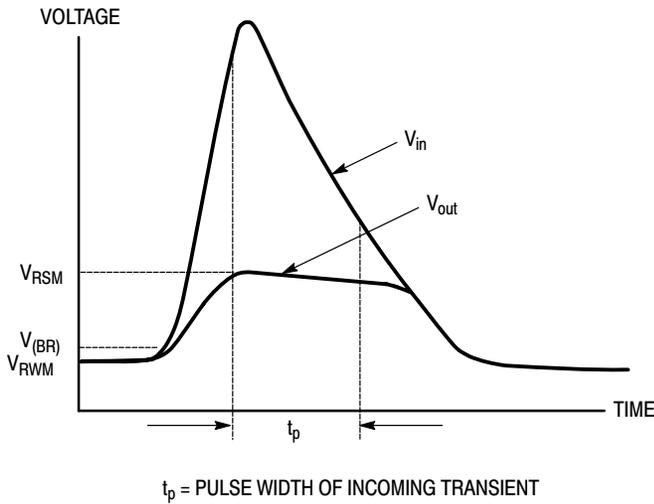


Figure 1a.

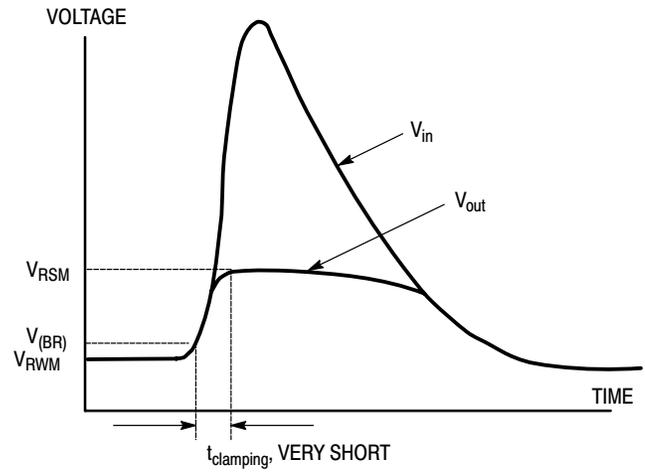


Figure 1b.

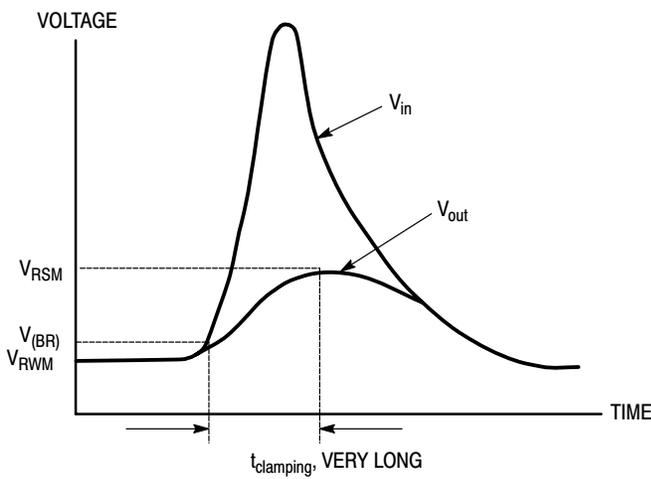


Figure 1c.

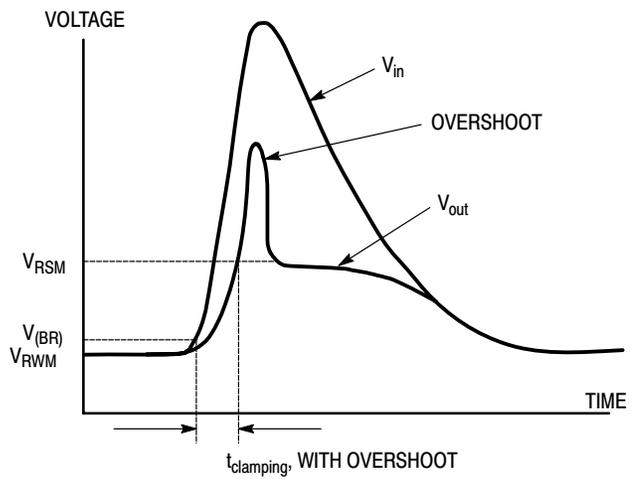
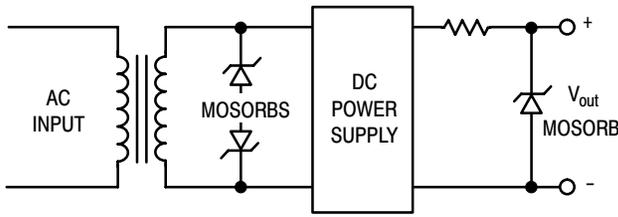


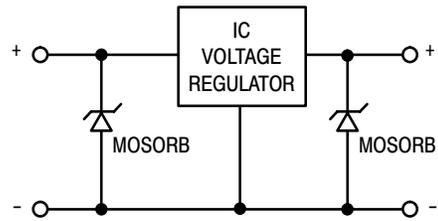
Figure 1d.

TYPICAL MOSORB APPLICATIONS

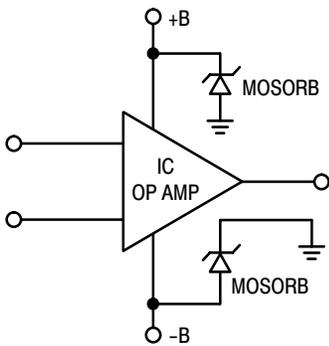
DC Power Supplies



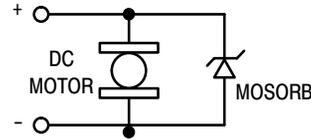
Input/Output Regulator Protection



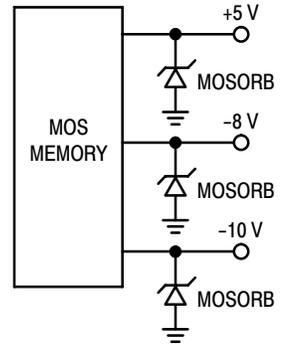
Op Amp Protection



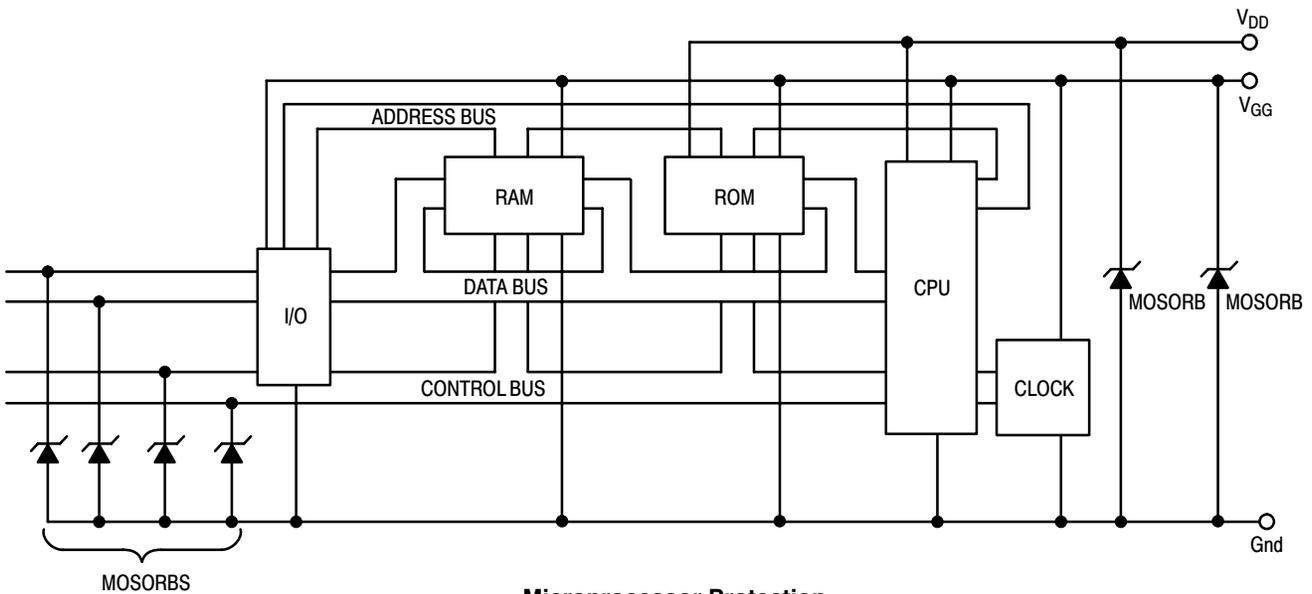
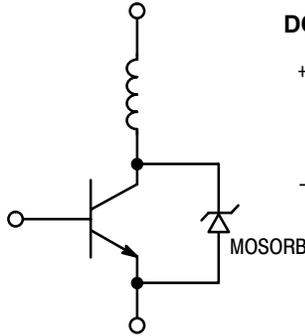
DC Motors - Reduces EMI



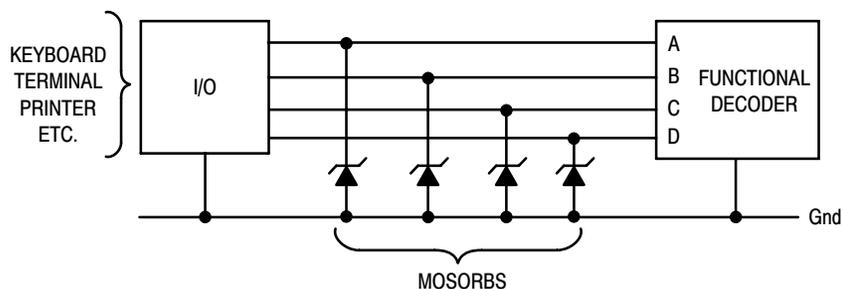
Memory Protection



Inductive Switching Transistor Protection



Microprocessor Protection



Computer Interface Protection

AR450 – CHARACTERIZING OVERVOLTAGE TRANSIENT SUPPRESSORS

Prepared by
AI Pshaenich
ON Semiconductor Power Products Division

The use of overvoltage transient suppressors for protecting electronic equipment is prudent and economically justified. For relatively low cost, expensive circuits can be safely protected by one or even several of the transient suppressors on the market today. Dictated by the type and energy of the transient, these suppressors can take on several forms.

For example, in the telecommunication field, where lightning induced transients are a problem, such primary suppressors as gas tubes are often used followed by secondary, lower energy suppressors. In an industrial or automotive environment, where transients are systematically generated by inductive switching, the transient energy is more well-defined and thus adequately suppressed by relatively low energy suppressors. These lower energy suppressors can be zener diodes, rectifiers with defined reverse voltage ratings, metal oxide varistors (MOVs), thyristors, and trigger devices, among others. Each device has its own niche: some offer better clamping factors than others, some have tighter voltage tolerances, some are higher voltage devices, others can sustain more energy and still others, like the thyristor family, have low on-voltages. The designer's problem is selecting the best device for the application.

Thus, the intent of this article is twofold:

1. To describe the operation of the surge current test circuits used in characterizing lower energy transient suppressors.
2. To define the attributes of the various suppressors, allowing the circuit designer to make the cost/performance tradeoffs.

Surge suppressors are generally specified with exponentially decaying and/or rectangular current pulses. The exponential surge more nearly simulates actual surge current conditions – capacitor discharges, line and switching transients, lightning induced transients, etc., whereas rectangular surge currents are usually easier to implement and control.

To generate an exponential rating, a charged capacitor is simply dumped into the device under test (DUT) and the energy of each successive pulse increased until the device ultimately fails. The simplified circuit of Figure 1a describes the circuit. By varying the size of the capacitor C, the limiting resistor R2, and the voltage to which C is charged to, various peak currents and pulse widths (defined to the 10% discharge point in this paper) can be obtained. To

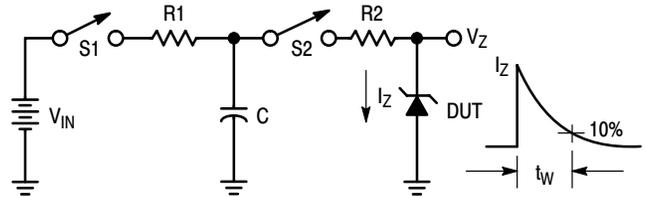


Figure 1a. Simplified Exponential Tester

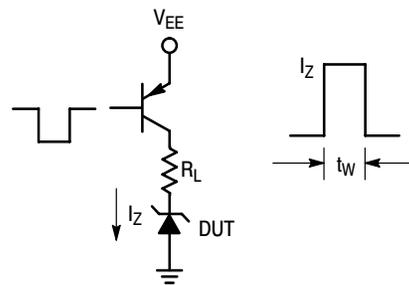


Figure 1b. Simplified Rectangular Tester Using PNP Switch

Figure 1. Basic Surge Current Testers

automate this circuit, the series switches S1 and S2 can be replaced with appropriately controlled transistors or SCRs.

One method of easily implementing a rectangular surge current pulse is shown in the simplified schematic of Figure 1b. A PNP transistor switch connected to the positive supply V_{EE} applies power to the DUT. The current is obviously set by varying either V_{EE} and/or R_L . If however, the transistor switch were replaced with a variable, constant current source, measurement procedures are simplified as how the limiting resistor need not be selected for various current conditions.

As in most surge current evaluations, the DUT is ultimately subjected to destructive energy (current, voltage, pulse width), the failure points noted, and the derated points plotted to produce the energy limitation curve. Of particular interest is the junction temperature at which the DUTs are operated, be it near failure or at the specified derated point. This measurement relates to the overall reliability of the suppressor, i.e., can the suppressor sustain one surge current pulse or a thousand, and will it be degraded when operated above the specified maximum operating temperature?

The Rectangular Current Surge Suppressor Test Circuit to be described addresses these questions by implementing and measuring the rectangular current capability of the

suppressor and determining the device junction temperature T_J shortly after the end of the surge current pulse. Knowing T_J , the energy to the DUT can be limited just short of failure and thus a complete surge curve generated with only one, or a few DUTs (Figure 6). Second, with the junction temperature known, a reliability factor can be determined for a practical application.

CIRCUIT OPERATION FOR THE RECTANGULAR CURRENT TESTER

The Surge Suppressor Test Circuit block diagram is shown in Figure 2 with the main blocks being the Constant Current Amplifier supplying I_Z to the DUT (a zener diode in this instance) during the power pulse and the Diode Forward Current Switch supplying I_F during the temperature sensing time. These two pulses are applied sequentially, first the much larger I_Z , and then the very small sense current I_F . During the I_F time, the forward voltage V_F of the diode is measured from which the junction temperature of the zener diode can be determined. This is simply done by calibrating the forward biased DUT with a specified low value of I_F in a temperature chamber, one point at 25°C and a second point at some elevated temperature. The result is the familiar diode forward voltage versus temperature linear plot with a slope of about $-2 \text{ mV}/^\circ\text{C}$ for typical diodes (Figure 7a). Comparing the plot with the test circuit measured V_F yields the DUT junction temperature for that particular pulse width and I_Z (Figure 7b).

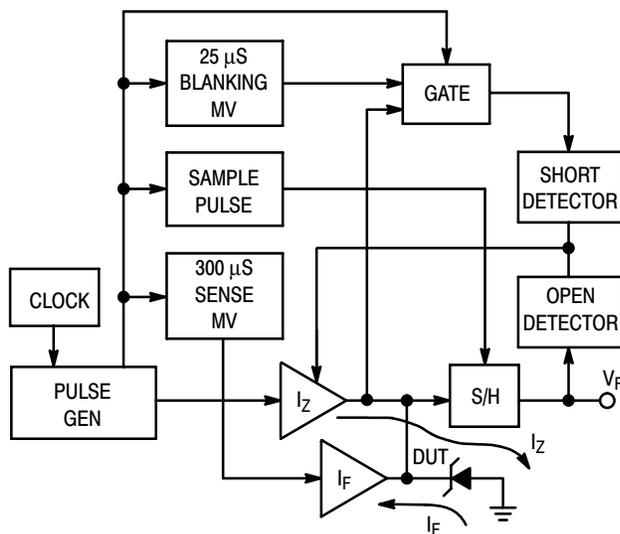


Figure 2. Surge Suppressor Test Circuit Block Diagram

The System Clock, Pulse Generator, the several monostable multivibrators (25 μs Blanking, Sample Pulse and 300 μs Sense MVs) and Gate are fashioned from three

CMOS gate ICs. The remaining blocks are the Sample and Hold (S/H) circuit and two detectors for determining the status of failed DUTs, either shorted devices or open.

Shown in Figures 3 and 4 respectively, are the complete circuit and significant waveforms. Clocking for the system is derived from a CMOS, two inverters, astable MV (gates 1A and 1B) whose output triggers the two input NOR gate configured monostable MV (gates 1C and 1D) to produce the Pulse Generator output pulse (Figure 4b). Alternatively, a single pulse can be obtained by setting switch S2 to the One Shot position and depressing the pushbutton Start switch S1. Contact bounce is suppressed by the 100 ms MV (gates 4C and D). Frequency of the astable MV, set by potentiometer R1, can vary from about 200 Hz to 0.9 Hz and the pulse width, controlled by R2 and the capacitor timing selector switch S3, from about 300 μs to 1.3 s.

The positive going Pulse Generator output feeds the Constant Current Amplifier I_Z and turns on, in order, NPN transistor Q1, PNP transistor Q3, NPN Darlington Q4, PNP Power Darlington Q6 and parallel connected PNP Power Transistors Q8 and Q9. Transistor Q4 is configured as a constant current source whose current is set by the variable base voltage potentiometer R3. Thus, the voltage to the bases of Q6, Q8 and Q9 are also accordingly varied. Transistors Q8 and Q9 (MJ14003, I_C continuous of 60 A), also connected as constant current sources with their 0.1 Ω emitter ballasting resistors, consequently can produce a rectangular current pulse from a minimum of about 0.5 A and still have adequate gain for 1 ms pulses of 150A peak. Due to propagation delays of this amplifier, the I_Z current waveform is as shown in Figure 4f. Since Q8 and Q9 must be in the linear region for constant current operation, these transistors are power dissipation limited at high currents to the externally connected power supply V_+ of 60 V. Thus the maximum DUT voltage, taking into account the clamping factor of the device, should be limited to about 50 V. At wider pulse widths and consequently lower currents before the DUT fails, the V_+ supply should be proportionally reduced to minimize Q8, Q9 dissipation. As an example, a 28 V surge suppressor operating at 100 ms pulse widths can be tested to destructive limits with V_+ of about 40 V. Although a zener diode is shown as the DUT in the schematic, the test devices can be any rectifier with defined reverse voltage, e.g., surge suppressors.

Immediately after the power pulse is applied to the DUT, the negative going sense pulse from the 300 μs MV (Gate 2A, Figure 4e) turns on series connected PNP transistor Q10 and NPN transistor Q11 of the Diode Forward Current Switch I_F . Sense current, set by current limiting resistor selector switch S4, thus flows up from ground through the forward biased DUT, the limiting resistor, and Q11 to the -15 V supply. The result, by monitoring the cathode of the DUT, is a 300 μs wide, approximately -0.6 V pulse.

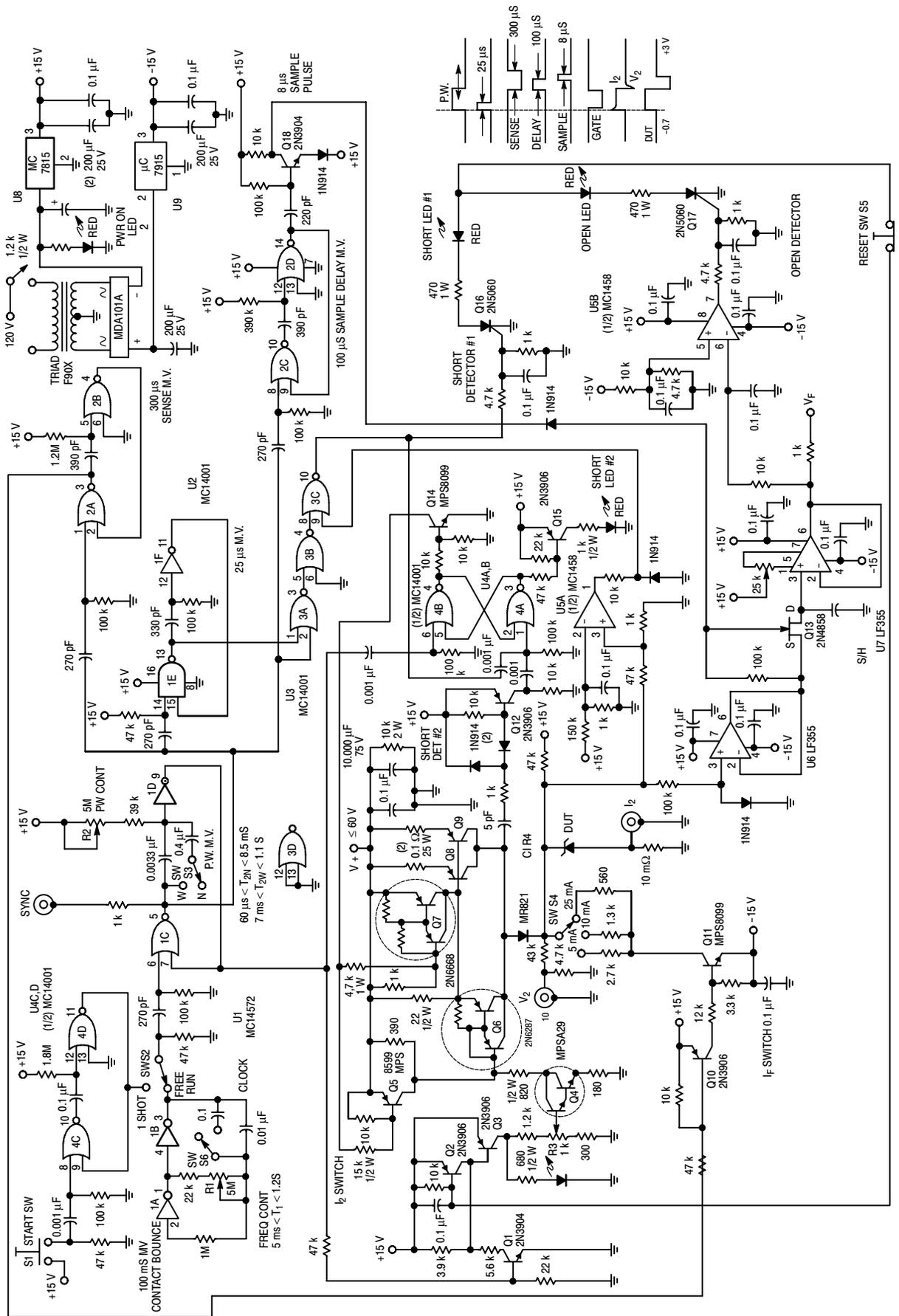


Figure 3. Surge Suppressors Surge Current Fixture

For accurate measurements of this pulse amplitude, sample and hold circuitry is employed. This consists of unity gain buffer amp U6, series FET switch Q13 and capacitor hold buffer amp U7. The sample pulse (Figure 4H) to the gate of the FET is delayed about 100 μ s (by monostable MV G-2C and G-2D) to allow for switching and thermal transients to settle down. This pulse is derived from the negative going, trailing edge output pulse of Gate 2D cutting off transistor Q18 for the RC time constant in its base circuit. The result is an approximate 8 μ s wide sample pulse. Consequently, the DC output voltage from hold amplifier U7 is a measure of the DUT junction temperature.

Invariably, most DUTs will fail short. When the surge suppressor tester is in the Free-run Mode, the power pulse subsequent to the DUT shorting could excessively stress the constant current drivers Q8 and Q9. To prevent this occurrence, the Short Detector circuit was implemented. This circuit consists of comparator U5A, 2 input NOR gate configured 25 μ s monostable MV (G1E and G1F), Gate Circuit G3A, 3B and 3C, and SCR Q16. The 25 μ s MV

(Figure 4D) is required to blank out turn-on switching transients to produce the waveform shown in Figure 4I. During the power pulse, U5A is normally high for a good DUT (Figure 4J). This waveform is NOR'd with gate 3B (inverted waveform of Figure 4I) to produce a low level (0 V) gate 3C output (Figure 4K).

If, however, the DUT is shorted, U5A output switches low resulting in a positive pulse output from G3C. This pulse triggers the SCR on, lighting the LED in its anode circuit and turning on the PNP transistor Q2 across the emitter-base of Q3, thus clamping off the I_z power pulse. The circuit (Q16) can be reset by opening switch S5.

By and large, this Short Detector circuit was found adequate to protect transistors Q8 and Q9. However, for some wide pulse widths, relatively high current conditions, the propagation delay through the Short Detector was too great, resulting in excessive FBSOA (Forward Bias Safe Operating Area) stress on Q8 and Q9. Consequently, a faster Short Detector #2 was implemented.

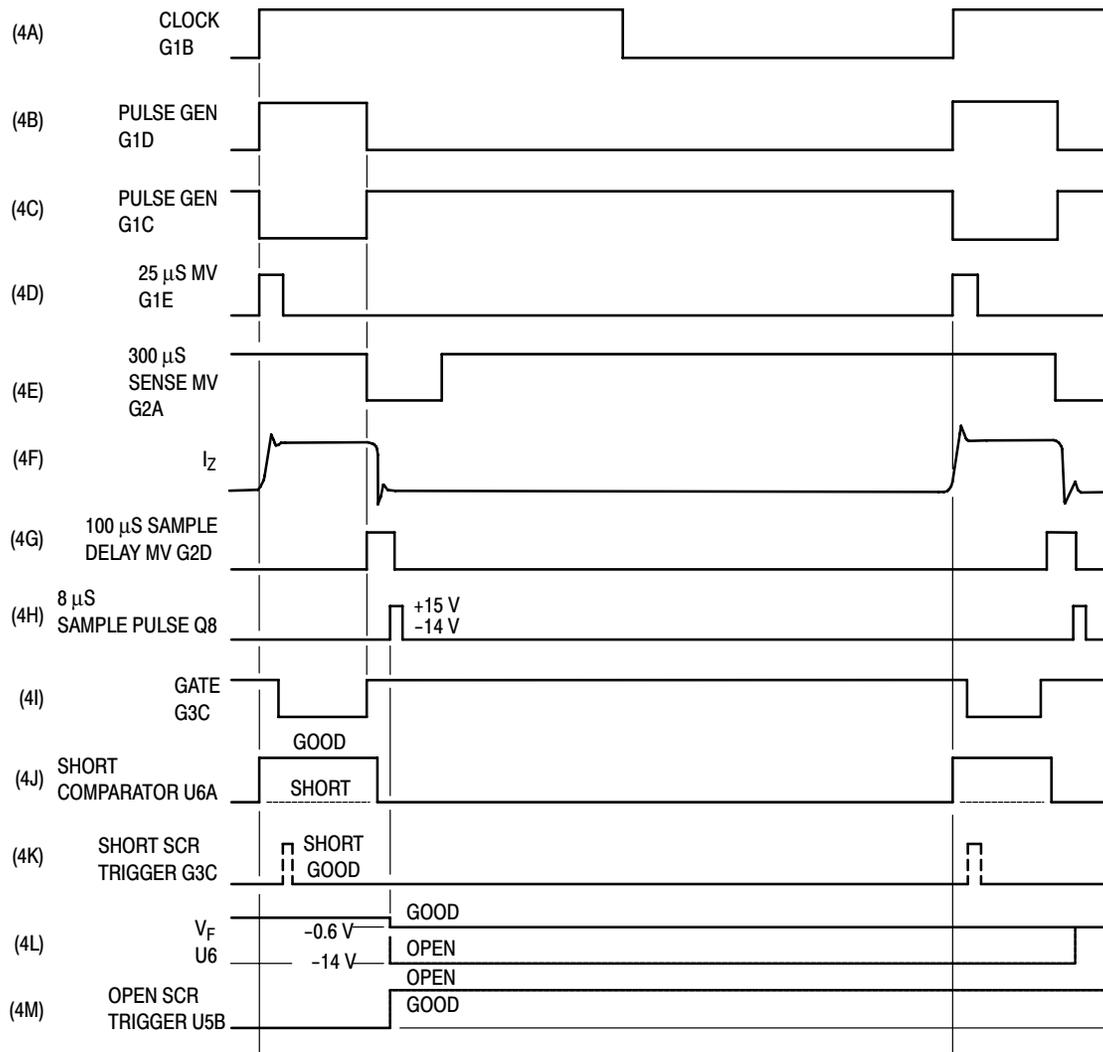


Figure 4. Surge Suppressor Test Circuit Waveforms

This circuit, connected to the collectors of Q8 and Q9, uses a differentiating network (R4, C1) to discriminate between the normally relatively slow fall time of the voltage pulse on the DUT, and the exceedingly fast fall time when the device fails. Thus, the R4-C1 time constant (5 ns) will only generate a negative going trigger to PNP transistor Q12 when the DUT voltage collapses during device failure. The positive going output from Q12 resets the flip-flop (gates 4A and 4B), which turns on the NPN transistor Q14. This transistor supplies drive to the two PNP clamp transistors (Q5 and Q7) placed respectively across the emitter-bases of the high, constant current stages Q6 and Q8 and Q9. Propagation delay is thus minimized, providing greater protection to the power stages of the tester. As an added safety feature, the positive going output from Gate 3C when Short Detector #1 is activated is also used to trigger the flip-flop.

On the few occasions when the DUT fails open, then the Open Detector consisting of comparator U5B and SCR Q17 comes into play. This circuit measures the DUT integrity during the sense time. For a good DUT ($V_F < -1$ V), U5B output remains low (see Figures 4L and 4M). However for an open DUT, V_F switches to the negative rail and U5B goes high, turning on Q17. As in the Short Detector, Q2 clamps off the I_Z power amplifier.

All of the circuitry including the +15 V and -15 V regulated power supplies are self-contained, with the exception of the V_+ supply. For high current, narrow pulse width testing, this external supply should have 10 to 15 A capability. If not, additional energy storing capacitors across the supply output may be required.

CIRCUIT OPERATION FOR THE EXPONENTIAL SURGE CURRENT TESTER

To generate the surge current curve of peak current versus exponential discharge pulse width, the test circuit of Figure 5 was designed. This tester is an implementation of the simplified capacitor discharge circuit shown in Figure 1A, with the PNP high voltage transistor Q2 allowing the capacitor C to charge through limiting resistor R1 and a triggered SCR discharging the capacitor. As shown in Figure 5, the DUTs can be of any technology, although the device connected to the capacitor and discharge limiting resistor R_S is shown as an MOS SCR. It could just as well have been an SCR as the DUT or as the switch for the zener diode, rectifier, SIDAC, etc., DUTs.

System timing for this Exponential Surge Current Tester is derived from a CMOS quad 2 input NOR gate with gates 1A and 1B comprising a non-symmetrical astable MV of about 13 seconds on and about one second off (switch S3 open). The positive On pulse from gate 1B turns on the 500 V power MOSFET Q1 and the following PNP transistor Q2. The extremely high current gain FET allows for the large base current variation of Q2 with varying supply voltage (V_+). This capacitor charging circuit has a 400 V blocking capability (limited by the V_{CEO} of Q2) and thus the capacitor C1 used should be comparably rated. When operating with

high voltage ($V_+ = 200$ to 350 V) and large capacitors (>3000 μ F), the power dissipated in the current limiting resistor R1 can be substantial, thus necessitating the illustrated 20 W rating. For longer charging times, switch S3 is closed, doubling the timing capacitor and the astable MV on time.

To discharge capacitor C1 and thereby generate the exponential surge current, the SCR must be fired. This trigger is generated by the positive going one second pulse from gate 1A being integrated by the R2C2 network, and then shaped by gates 1C and 1D. The net result of about 100 μ s time delay from gate 1D ensures noncoincident timing conditions. This pulse output is then differentiated by C3-R3 with the positive going leading edge turning on Q3, Q4 and finally the SCR with about a 4 ms wide, 15 mA gate pulse. Consequently, the DUT is subjected to a surge current pulse whose magnitude is dictated by the voltage on the capacitor C1 and value of resistor R_S , and also whose pulse width to the 10% point is $2.3 R_S C1$. For a fixed pulse width, the DUT is then stressed with increasing charge (by increasing V_+) until failure occurs, usually a shorted device.

If the DUT is the SCR (or MOS SCR), the failed condition is obvious as the capacitor C1 will not be allowed to charge for subsequent timing cycles. However, when the DUT is the zener, rectifier, SIDAC or even an MOV, and the SCR is an adequately rated switch, the circuit will still discharge through the shorted DUT, but now the SCR alone will be stressed by the surge current. A shorted DUT can be detected by noting the voltage across the device during testing.

One problem encountered when stressing SCRs with high voltage is when the DUT fails short. The limiting resistor R1, which is only rated for 20 W, would now experience continuous power dissipation for the full On time – as much as 123 W ($[350 \text{ V}]^2/1\text{K}$). To prevent this occurrence, the P_{R1} Short Protection Circuit is incorporated. Since this is only a problem when high V_+ 's (>100 V) are used, the circuit can be switched in or out by means of switch S2. When activated, this circuit monitors the voltage on capacitor C1 some time after the charging cycle begins. If the capacitor is charging, normal operation occurs. However, if the SCR DUT is shorted, the absence of voltage on the capacitor is detected and the system is disabled.

The circuit consists of one CMOS IC with NAND gates 2A and 2B comprising a one second monostable time delay MV and gates 2C and 2D forming a comparator and NAND gate, respectively.

The negative going, trailing edge of gate 2A is differentiated by R4-C4, and amplified by Q5 to form a positive, 10 ms wide pulse (delayed by 1 sec) to gate 2D input. If the capacitor C1 is shorted, gate 2C output is high, allowing the now negative pulse from gate 2D to turn on PNP transistor Q6 and SCR Q7. This latches the input to the astable MV gate 1A low, disabling the timing and consequently removing the power from R1. Resetting the tester for a new device is accomplished by depressing the pushbutton switch S1.

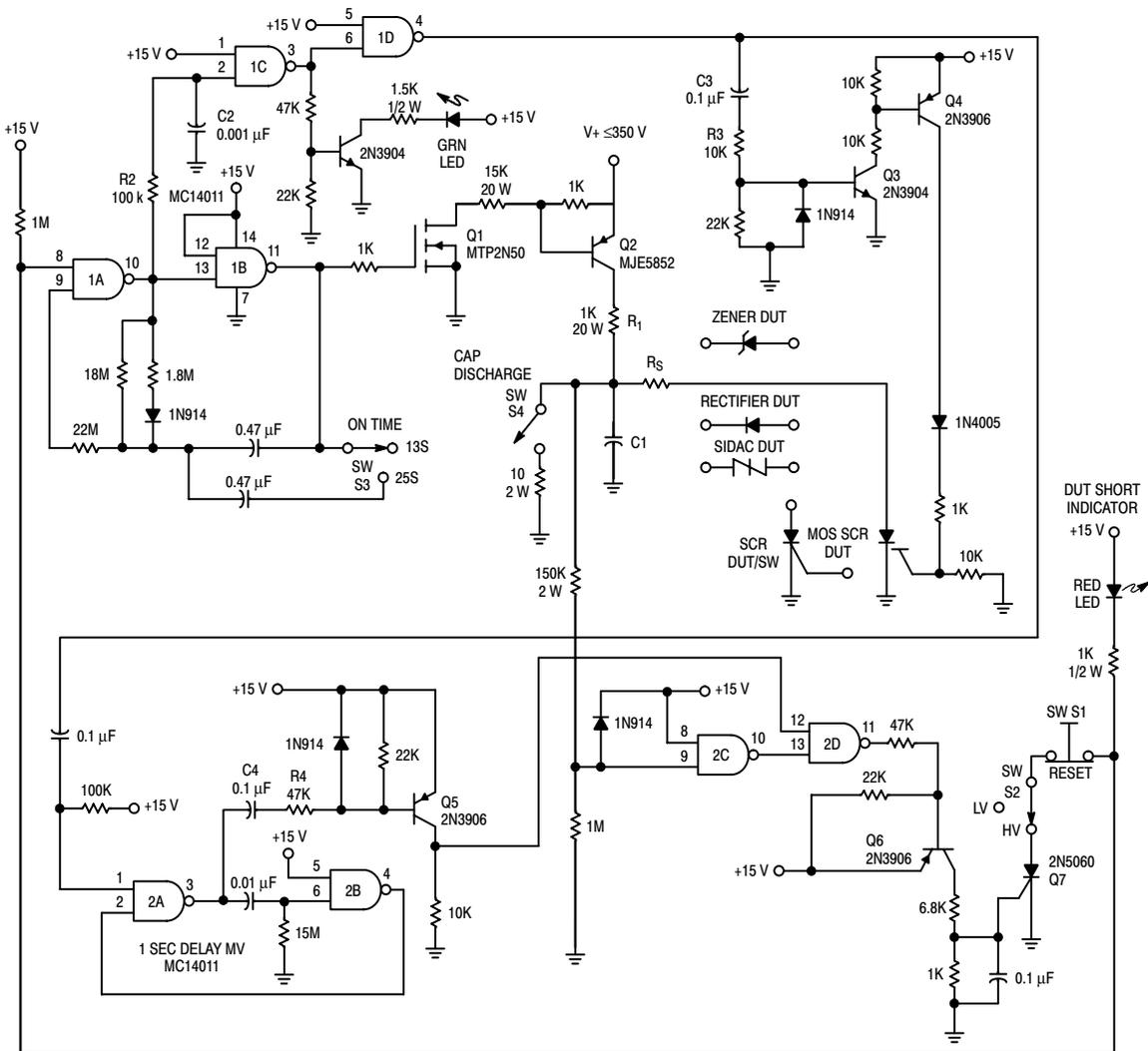


Figure 5. Exponential Surge Current Tester

Exponential surge current curves, as well as rectangular, are generated by destructive testing of at least several DUTs at various pulse widths and derating the final curve by perhaps 20–30%. These tests were conducted at low duty cycles (<2%). To ensure multicycle operation, the DUTs are then tested for about 1000 surges at a derated point on the curve.

TEST RESULTS

In trying to make a comparison of the several different technologies of transient suppressors, some common denominator has to be chosen, otherwise, the amount of testing and data reduction becomes unwieldy. For this exercise, voltage was used, generally in the 20 V to 30 V range, although some of the more unique suppressors (SIDACs, MOS SCRs, SCRs) were tested at their operating voltage. As an example, the SIDAC trigger families of devices were tested with voltages greater than their breakover voltage (104 V to 280 V) and the SCRs were subjected to exponential surge currents derived from

voltages generally greater than 30 V. Also, since energy capability is related to die size, this parameter is also listed.

For several devices, both rectangular and exponential surge current pulses are listed. Other devices were tested with only rectangular pulses (where the junction temperature can be determined) and still others, whose applications include crowbars, with exponential current only.

AVALANCHE RECTIFIER

The Rectangular Surge Current Tester was originally designed for characterizing rectifier surge suppressors used in automotive applications. For this operation, where temperatures under the hood can reach well over 125°C, it is important to know the device junction temperature at elevated ambient temperature. Figures 6 and 7 describe the results of such testing on a typical suppressor, the 24 V–32 V MR2520L. It should be noted that these axial lead suppressors, as well as all other axial lead devices tested, were mounted between two spring loaded clips spaced 1 inch apart.

As shown in Figure 6 of the actual current failure points of the DUTs, at least four devices were tested at the various pulse widths, t_w (in this example from 0.5 ms to 100 ms).

Also shown in Figure 6 is the curve derived with a single DUT at an energy level just short of failure. This measurement was obtained by maintaining a constant rectifier forward voltage drop, V_F (0.25 V) for all pulse widths (junction temperature, T_J of 230°C) by varying the avalanche current. Thus, one device can be used, non-destructively, to generate the complete rectangular surge current curve.

It should also be pointed out that the definition for the exponential t_w in this article is the current discharge point to the 10% value of the peak test current I_{ZM} . Expressed in time constant τ , this would be 2.3 RC. Some data sheets describe t_w to the 50% point of I_{ZM} (0.69 τ) and others to 5 τ . To normalized these time scales (abscissa of curves) simply change the scales accordingly; i.e., $I_{ZM}/2$ pulse widths would be multiplied by 2.3/0.69 = 3.33 for t_w at 10% current pulses.

Figure 7a describes the actual temperature calibration curve (measured in a temperature chamber) of the MR2520L and Figure 7b, the junction temperature of the DUT at various 10 ms rectangular pulse current amplitudes. These temperatures are taken from the calibration curve (in actuality, an extremely linear curve), knowing the rectifier forward voltage drop immediately (within 100 μ s) after cessation of power. Note that the junction temperature just prior to device failure is about 290°C.

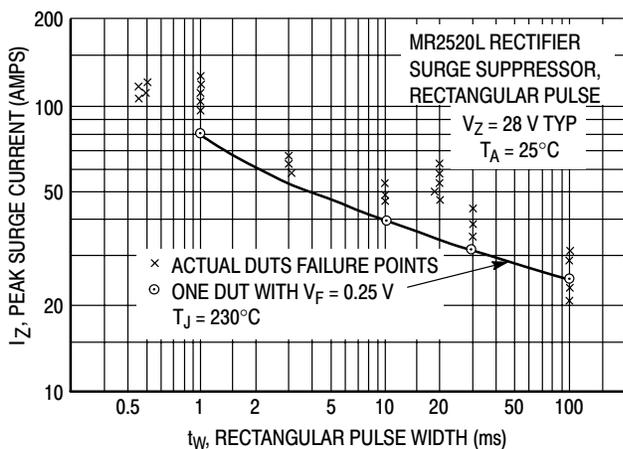


Figure 6. Experimental Rectangular Surge Current Capability Of The MR2520L Rectifier Surge Suppressor

ZENER OVERVOLTAGE TRANSIENT SUPPRESSOR

Illustrated in Figure 8 are the actual rectangular and exponential surge current curves of the P6KE30 overvoltage transient suppressor, an axial lead, Case 17, 30 V zener diode characterized and specified for surge currents. This device is specified for 600 W peak for a 1 ms exponential pulse measured at $I_{ZM}/2$. From the exponential curve, it is

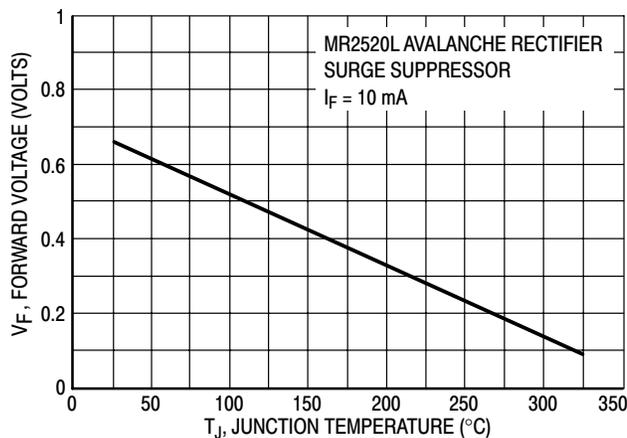


Figure 7a. Temperature Calibration Curve Of The MR2520L

RECTANGULAR PULSE
 $t_w = 10$ ms
 $I_F = 10$ mA

I_Z (A)	V_F (V)	T_J (°C)
1	0.64	25
10	0.57	75
20	0.48	120
30	0.36	180
40	0.25	230
50	0.15	290
55	0.10	DUT FAILED

Figure 7b. Measured Forward Voltage

Figure 7. Calculated Junction Temperature Of The MR2520L Surge Suppressor At Various Avalanche Currents

apparent that the device is very conservatively specified. Also, the relative magnitudes of the two curves reflect the differences in the rms values of the two respective pulses.

SIDAC

SIDACs are increasingly being used as overvoltage transient suppressors, particularly in telephone applications. Being a high voltage bilateral trigger device with relatively high current capabilities, they serve as a costeffective overvoltage protection device. As in other trigger devices, when the SIDACs breakover voltage is exceeded, the device switches to a low voltage conduction state, allowing an inordinate amount of surge current to be passed. This is well illustrated by the surge current curves of Figure 9 which describe the small die size ([37]²mil) axial lead, Case 59-04, MKP9V240 SIDAC. The curves show that this 240 V device was able to handle, to failure, as much as 31 A and 15 A, respectively, for 1 ms exponential and rectangular current pulses. Under the same pulse conditions, the large die ([78]²mil) MK1V270 SIDAC handled 170 A and 60 A, respectively, as shown in Table 2.

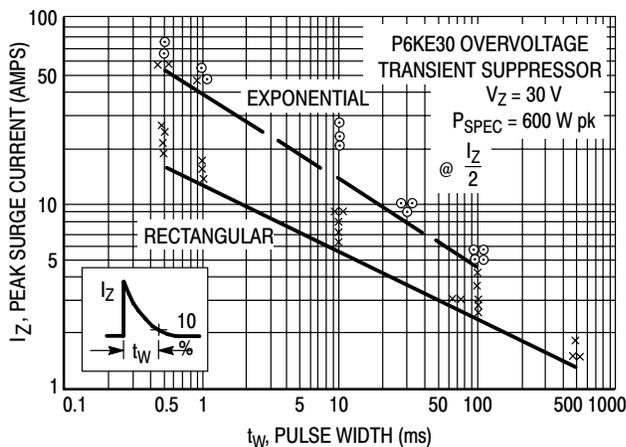


Figure 8. Surge Current Capability Of The P6KE30 Overvoltage Transient Suppressor As A Function Of Exponential & Rectangular Pulse Widths

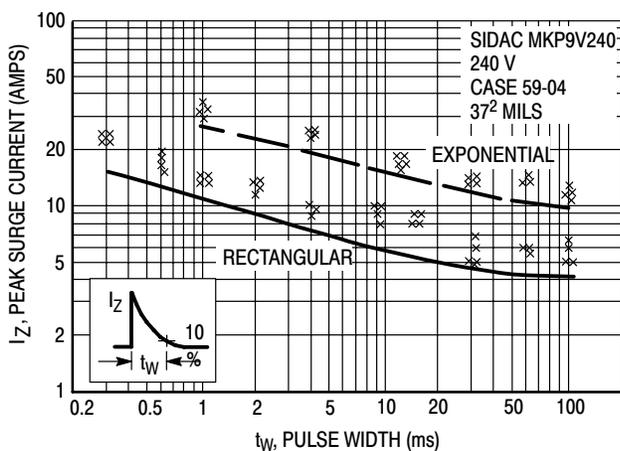


Figure 9. Measured Surge Current To Failure Of A SIDAC MKP9V240

OVERALL RATINGS

The compilation of all of the testing to date on the various transient suppressors is shown in Tables 1 and 2. Table 1 describes the zener suppressors, avalanche rectifiers and MOVs, comparing the die size and normalized costs (referenced to the MOV V39MA2A). From this data, the designer can make a cost/performance judgment.

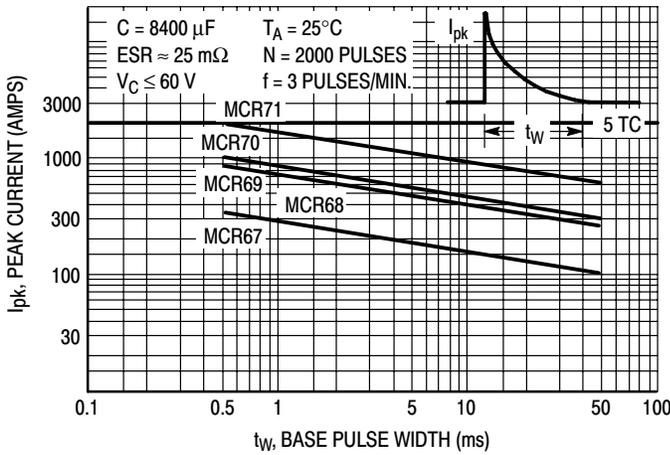
Of interest is that the small pellet MOV is not the least expensive device. The P6KE30 overvoltage transient suppressor costs about 85% of the MOV, yet it can handle about three times the current (2.5 A to 0.7 A) for a 100 ms rectangular pulse. Under these conditions, the resultant clamping voltages for the zener and MOV were 32 V and 60 V respectively.

Also shown in the table is a 1.5 W zener diode specified for zener applications. This low surge current device costs three times the MOV, illustrating that tight tolerance zener diodes are not cost effective and that the user should use devices designed and priced specifically for the suppressor application.

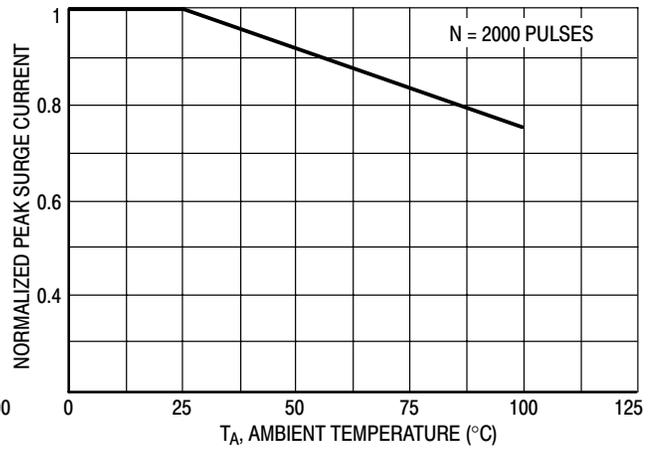
Thyristor type surge suppressors are shown in Table 2. They include four SIDAC series, two SCRs designed and characterized specifically for crowbar applications and also the MOS SCR MCR1000. The MOS SCR, a process variation of the vertical structure power MOSFET, combines the input characteristics of the FET with the latching action of an SCR.

All devices were surge current tested with the resultant peak currents being impressively high. The TO-220 (150)² mil SCR MCR69 for example, reached peak current levels approaching 700 A for a 1 ms exponential pulse. The guaranteed, derated, time base translated curves for the crowbar SCR family of devices are shown in Figure 10, as is the MK1V SIDAC in Figure 11.

Figures 12A–C describe the guaranteed, reverse surge design limits for the avalanche rectifier devices. These three figures illustrate, respectively, the peak current, power and energy capabilities of these overvoltage transient suppressors derived from exponential testing. The peak power, P_{pk} , ordinate of the curve is simply the product of the derated I_Z and V_Z and the energy curve, the product of P_{pk} and t_w .



a. Peak Surge Current versus Pulse Width



b. Peak Surge Current versus Ambient Temperature

Figure 10. SCR Crowbar Derating Curves

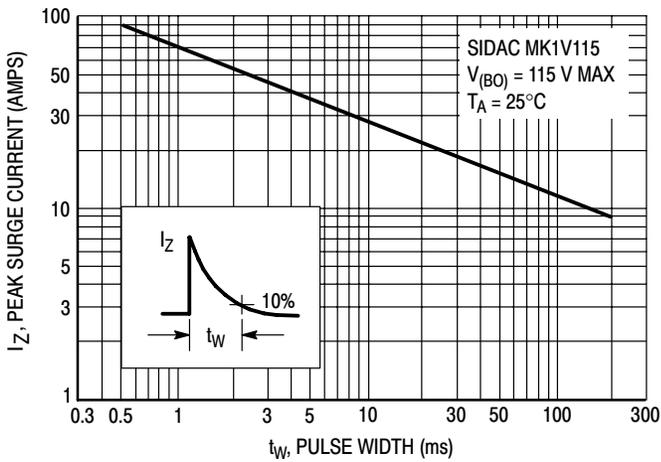


Figure 11. Exponential Surge Current Capability Of The MK1V SIDAC, Pulse Width versus Peak Current

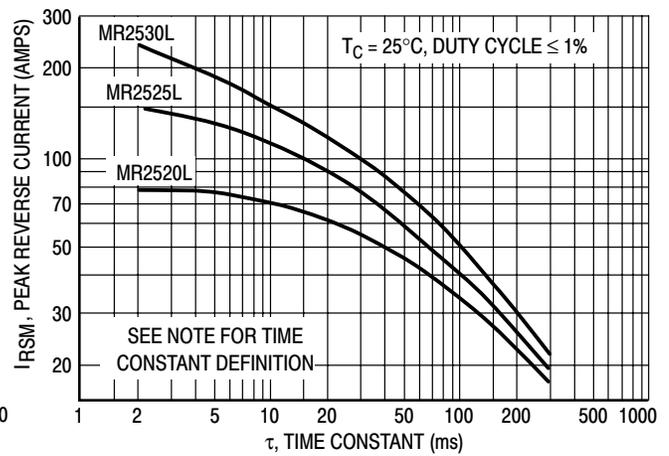


Figure 12a. Peak Current

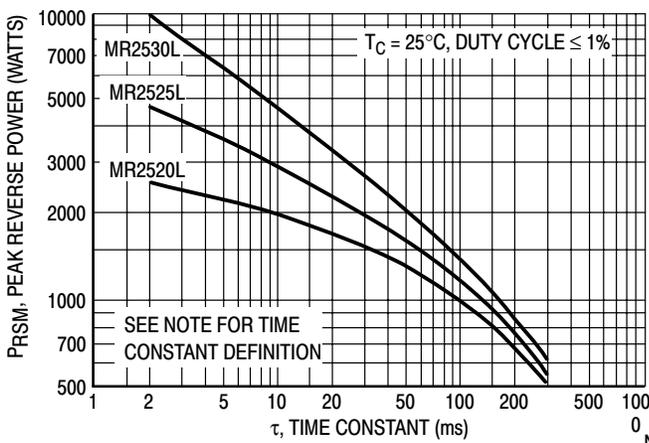


Figure 12b. Peak Power

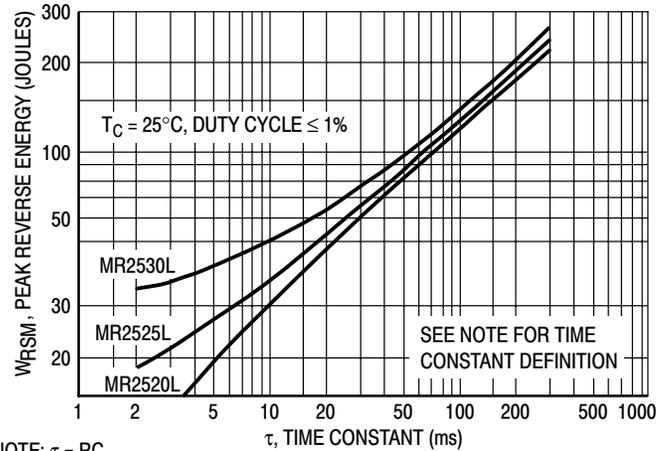


Figure 12c. Energy

Figure 12. Guaranteed Reverse Surge Design Limits for the MR2525L & MR2530L Overload Transient Suppressors

Table 1. Measured Surge Current Capability of Transient Suppressors

Dev Type	Title	Part	Case	Spec.		Die Size	Rated Currents (Amps)				Clamp V _{1ms}	Norm Cost *				
				Voltage	Power (Energy)		1s	10s	20s	100s						
							Exponent	Rectang.	Exponent	Rectang.			Exponent	Rectang.		
Avalanche Rectifier	Surge Supp., Overvoltage Transient Suppressor	MR2520L	194-05	24-32 V	2.5 KW Peak	150 ² mil		85 A	40		30	18	$\frac{27 V}{22 V} = 1.2$	4.0		
		MR2525L		24-32 V	10 KW Peak	196 ² mil		150 A	70		54	37	$\frac{31}{23} = 1.3$			
Zener	1.5 W Zener Diode	1N5936A	DO-41	30 V	1.5 W Cont.	37 ² mil	12 A	5	6	2.5	5	2	3	1.3	$\frac{41}{30} = 1.4$	3.2
		1N5932A		20 V			23 A	6	10	2.8	7	2.3	5	1.4	$\frac{28}{23} = 1.2$	
	Overvoltage Transient Suppressor	P6KE30	17	30 V	600 W Peak	60 ² mil	43 A	14	14	5	10	4.5	5	2.5	$\frac{41}{32} = 1.3$	0.85
		P6KE10		10 V			24 A		12		9		5.5	$\frac{16}{13} = 1.2$		
	MOSORB	1.5KE30	41A-02	30 V	1500 W Peak	104 ² mil	35 A		10				4	$\frac{35}{33} = 1.1$	1.8	
		1.5KE24		24 V			45 A		14			6	$\frac{30 V}{28 V} = 1.1$			
MOV**	Metal Oxide Varistor	V39MA2A	Axial Lead	28 V	(0.16 Joules)	3 mm	9 A		5				0.7	80 V 60 V	6 A 0.7 A	1.0
		V33ZA1	Radial Lead	26 V	(1.0 Joules)	7 mm			35				4 A	105 V 80 V	35 A 4 A	1.4

**G.E.

Table 2. Measured Surge Current Of Thyristor Type Devices

Technology	Device	Voltage Ratings	Case	Die Size	I _{pk} @ t _w				Norm Cost *
					1 ms		10 ms		
					Exponent.	Rectang.	Exponent.	Rectang.	
SIDAC	MKP9V130 Series	104 V-135 V	59-04	37 ² mil	40 A	13 A	16 A	8 A	0.87
	MKP9V240 Series	220 V-280 V			31 A	15 A	20 A	8 A	
	MK1V135 Series	120 V-135 V	267-01	78 ² mil	140 A	80 A	55 A	30 A	1.1
	MK1V270 Series	220 V-280 V			170 A	60 A	90 A	28 A	
SCR	MCR68 Series	25 V-400 V	TO-220	92 ² mil	300 A		170 A		1.2
	MCR69 Series			150 ² mil	700 A		400 A		1.9
MOS SCR	MCR1000 Series	200 V-600 V		127 mil x 183 mil	250 A		170 A		9.3

*Normalized to G.E. MOV V39MA2A, Qty 1-99, 1984 Price

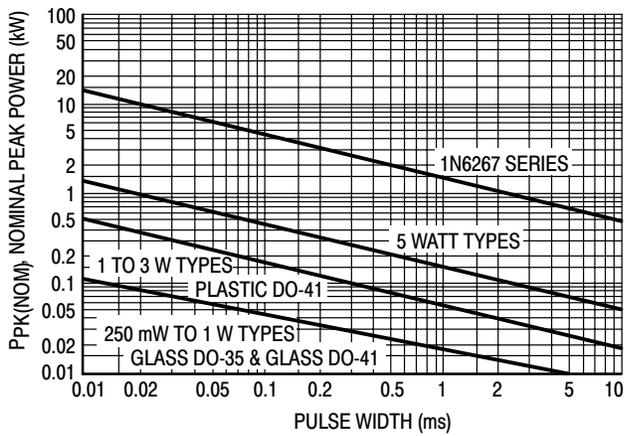
Additionally, the published non-repetitive peak power ratings of the various zener diode packages are illustrated in Figure 13. Figure 14 describes the typical derating factor for repetitive conditions of duty cycles up to 20%. Using these two empirically derived curves, the designer can then determine the proper zener for the repetitive peak current conditions.

At first glance the derating of curves of Figure 14 appear to be in error as the 10 ms pulse has a higher derating factor than the 10 μs pulse. However, when the mathematics of multiplying the derating factor of Figure 14 by the peak power value of Figure 13 is performed, the resultant respective power and current capability of the device follows the expected trend. For example, for a 5 W, 20 V

zener operating at a 1.0% duty cycle, the respective derating factors for 10 μs and 10 ms pulses are 0.08 and 0.47. The non-repetitive peak power capabilities for these two pulses (10 μs and 10 ms) are about 1300 W and 50 W respectively, resulting in repetitive power and current capabilities of about 104 W and 24 W and consequently 5.2 A and 1.2 A.

MOV

All of the surge suppressors tested with the exception of the MOV are semiconductors. The MOV is fabricated from a ceramic (ZnO), non-linear resistor. This device has wide acceptance for a number of reasons, but for many applications, particularly those requiring good clamping



Power is defined as $V_{Z(NOM)} \times I_{Z(PK)}$ where $V_{Z(NOM)}$ is the nominal zener voltage measured at the low test current used for voltage classification.

Figure 13. Peak Power Ratings of Zener Diodes

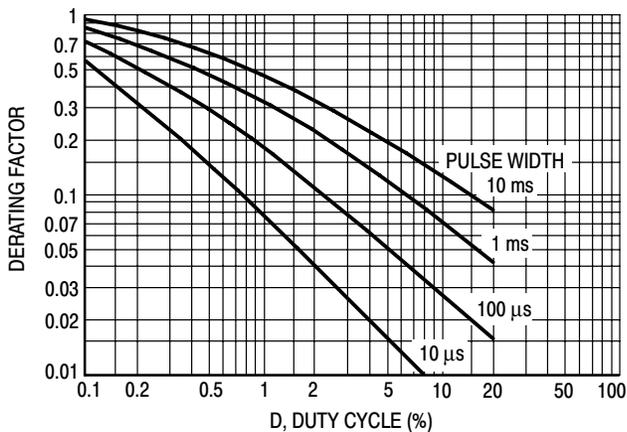
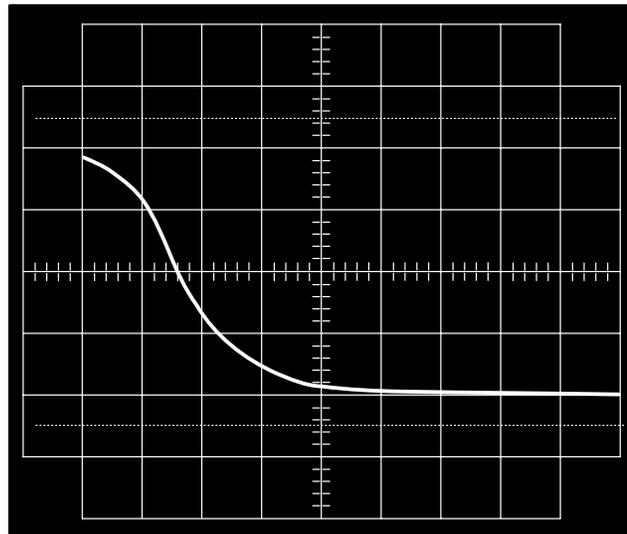


Figure 14. Typical Derating Factor for Duty Cycle

factors, the MOV is found lacking; (clamping factor is defined as the ratio of V_Z at the test current to that at 1.0 mA). This is photographically illustrated in Figure 15 which compares a 27 V zener (1N6281) with a 27 V MOV (V27ZA4). The input waveform, through a source impedance resistance to the DUTs, was an exponentially decaying voltage waveform of 90 V peak. Figures 15A and B compare the output waveforms (across the DUTs) when the source impedance was 500 Ω and Figures 15C and D for a 50 Ω condition. The zener clamped at about 27 V for both impedances whereas the MOV was about 40 V and 45 V respectively.

Surge current capabilities of a comparably powered MOV were also determined, as shown in the curve of Figure 16. Although the MOV, a V39MA2A, is specified as a 28 V

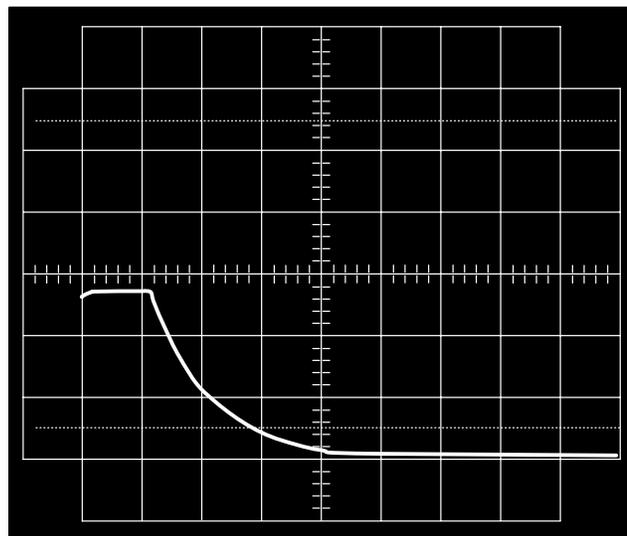
SOURCE IMPEDANCE $R_S = 500 \Omega$



27 V MOV
G.E. V27ZA4, 4 JOULES CAPABILITY

Figure 15a.

SOURCE IMPEDANCE $R_S = 500 \Omega$



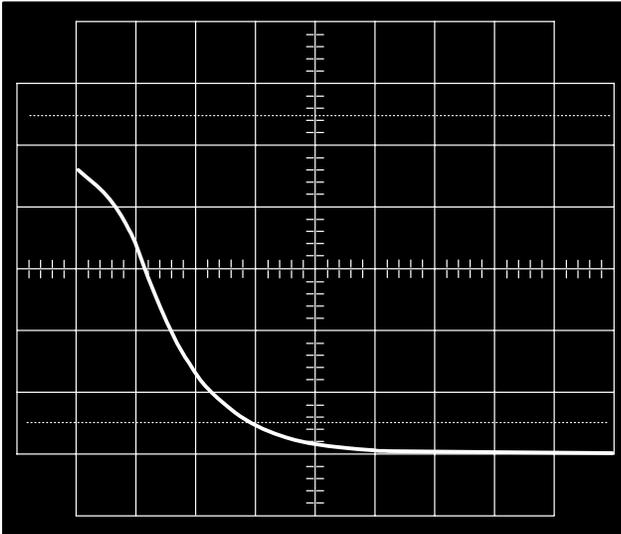
27 V ZENER DIODE
ON Semiconductor 1N6281, APPROX. 1.5 JOULES

Figure 15b.

continuous device ($39 V \pm 10\%$ at 1 mA) at the pulse widths and currents tested, the resultant voltage V_Z across the MOV – 80 V at about 6 A – necessitated a high voltage fixture. This was accomplished with a circuit similar to that of Figure 1B.

But MOVs do have their own niche in the marketplace, as described in Table 3, the Relative Features of MOVs and MOSORBs.

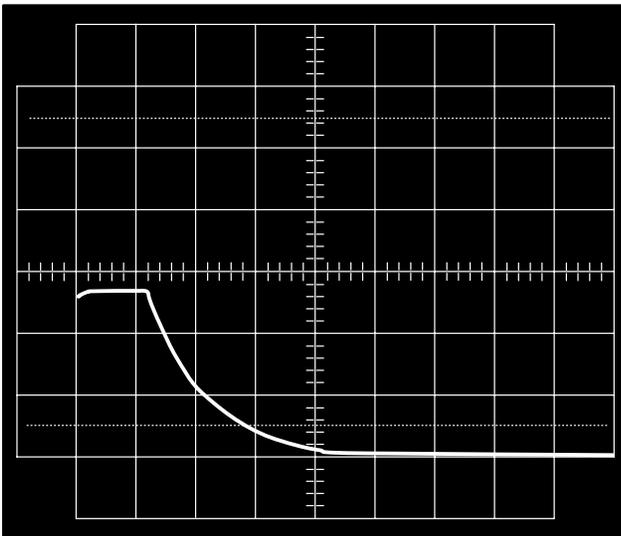
SOURCE IMPEDANCE $R_S = 50 \Omega$



27 V MOV

Figure 15c.

SOURCE IMPEDANCE $R_S = 50 \Omega$



27 V ZENER DIODE

Figure 15d.

Figure 15. Clamping Characteristics of a 27 V Zener Diode and 27 V MOV

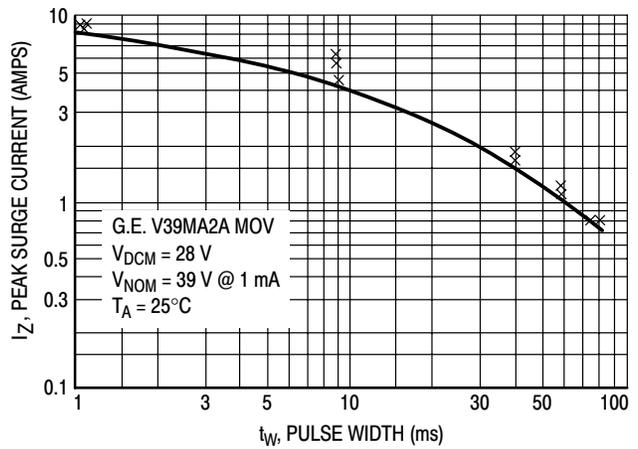


Figure 16. Rectangular Surge Current Capability Of The V39MA2A MOV

Table 3. Relative Features of MOVs and MOSORBs

MOV	MOSORB/Zener Transient Suppressor
High Clamping Factor	Very good clamping close to the operating voltage.
Symmetrically bidirectional	Standard parts perform like standard zeners. Symmetrical bidirectional devices available for many voltages.
Energy capability per dollar usually much greater than a silicon device. However, if good clamping is required a higher energy device would be needed, resulting in higher cost.	Good clamping characteristics could reduce overall cost.
Inherent wear out mechanism, clamp voltage degrades after every pulse, even when pulsed below rated value.	No inherent wear out mechanism.
Ideally suited for crude AC line protection.	Ideally suited for precise DC protection.
High single-pulse current capability.	Medium multiple-pulse current capability.
Degrades with overstress.	Fails short with overstress.
Good high voltage capability.	Limited high voltage capability unless series devices are used.
Limited low voltage capability.	Good low voltage capability.

SUMMARY

The surge current capabilities of low energy overvoltage transient suppressors have been demonstrated, including cost/performance comparison of rectifiers, zeners, thyristor type suppressors, and MOVs. Both rectangular and exponential testing have been performed with the described testers. Additionally, the Rectangular Current Surge Tester has the capability of measuring the diode junction temperature of zeners and rectifiers at various power levels, thus establishing safe operating limits.

REFERENCES

1. Cherniak, S., *A Review of Transients and Their Means of Suppression*, ON Semiconductor Application Note AN843.
2. Wilhardt, J., *Transient Power Capability of Zener Diodes*, ON Semiconductor Application Note AN784.
3. Pshanenich, A., *Characterizing the SCR for Crowbar Applications*, ON Semiconductor Application Note AN879.
4. Pshaenich, A., *The SIDAC, A New High Voltage Trigger that Replaces Circuit Complexity and Cost*, ON Semiconductor Engineering Bulletin EB-106.
5. General Electric, *Transient Voltage Suppression Manual*, Second Edition.

MEASUREMENT OF ZENER VOLTAGE TO THERMAL EQUILIBRIUM WITH PULSED TEST CURRENT

Prepared by
Herb Saladin
Discrete Power Application Engineering

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 a 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 μ s 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 μ s Delay MV and the 200 μ s 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 μ s 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 k Ω 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(μ F)	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 . Curves are shown for a 3.9 V diode (1N5228B) which has a negative TC and a 12 V diode (1N5242B) having a positive TC.

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 1N5242B diode is specified at 12 V \pm 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 \pm 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 1N5221B series of diodes. θV_Z is expressed in mV/°C and for the 1N5228B diode is about -2 mV/°C and for the 1N5242B, about 1.6 mV/°C. These values are 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, $T_A = 25^\circ\text{C}$, $V_Z |_{I_{ZT}} = 1/4 P_D$ (Max) Each device normalized to V_Z at TE.

AXIAL LEAD PACKAGES: MOUNTING STANDARD GRAYHILL CLIPS

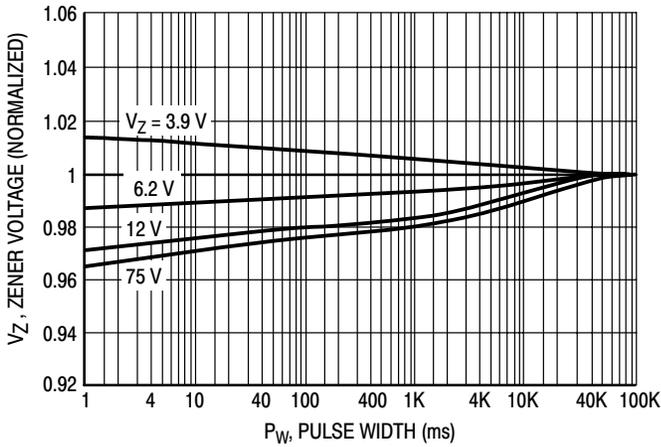


Figure 1. DO-35 (Glass) 500 mW Device

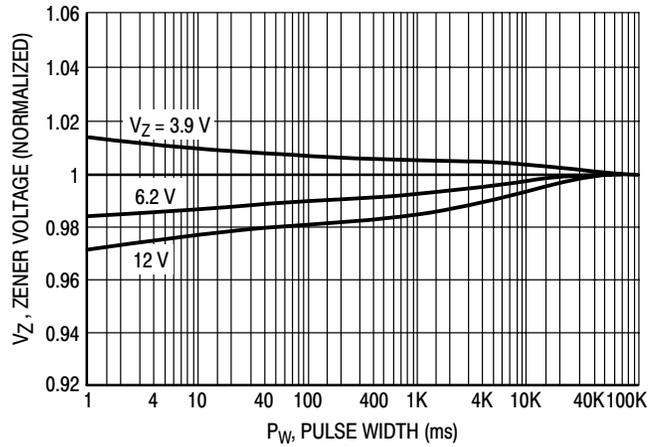


Figure 2. DO-41 (Glass) 1 Watt Device

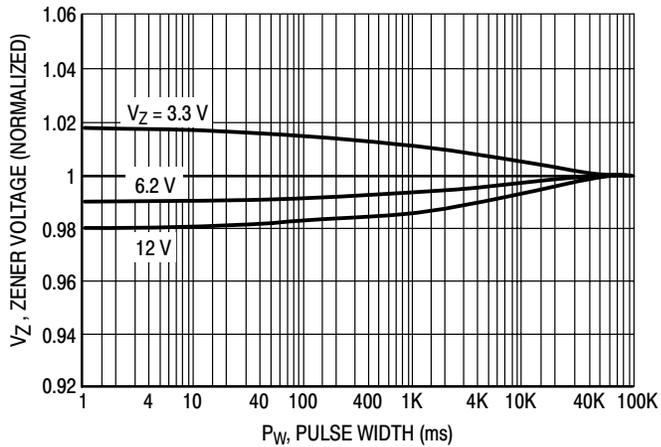


Figure 3. DO-41 (Plastic) 1.5 Watt Device

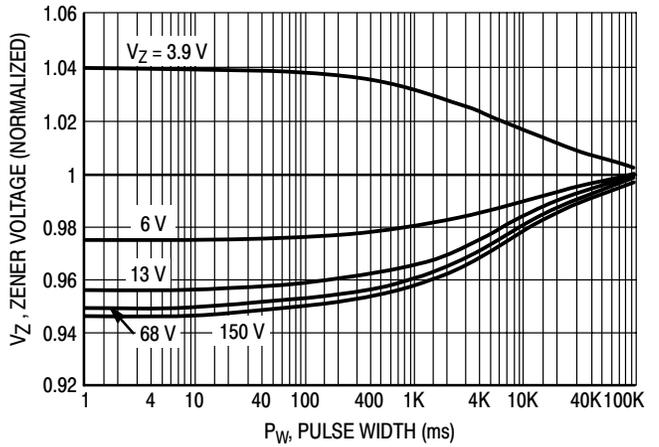


Figure 4. Case 17 (Plastic) 5 Watt Device

THREE MOUNTING METHODS: DO-35 AND DO-41

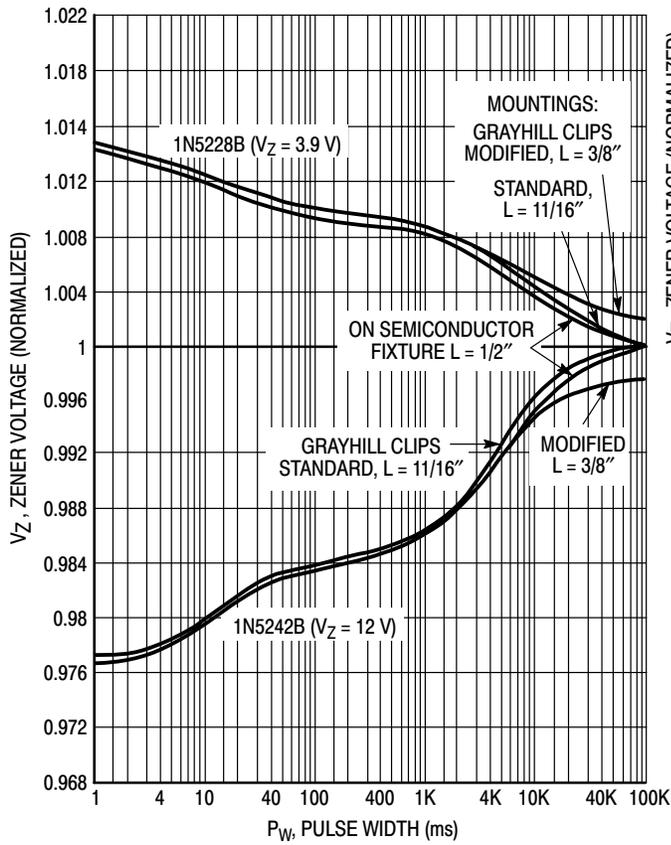


Figure 5. DO-35 (Glass) 500 mW Device

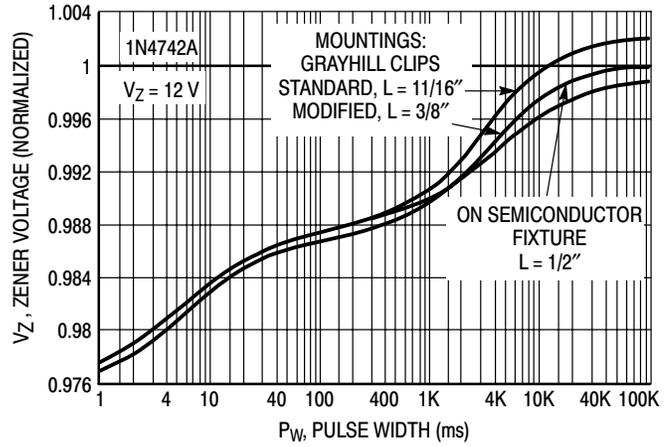


Figure 6. DO-41 (Glass) 1 Watt Device

MOUNTING FIXTURE

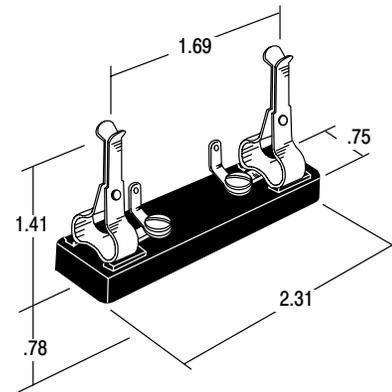


Figure 7. Standard Grayhill Clips

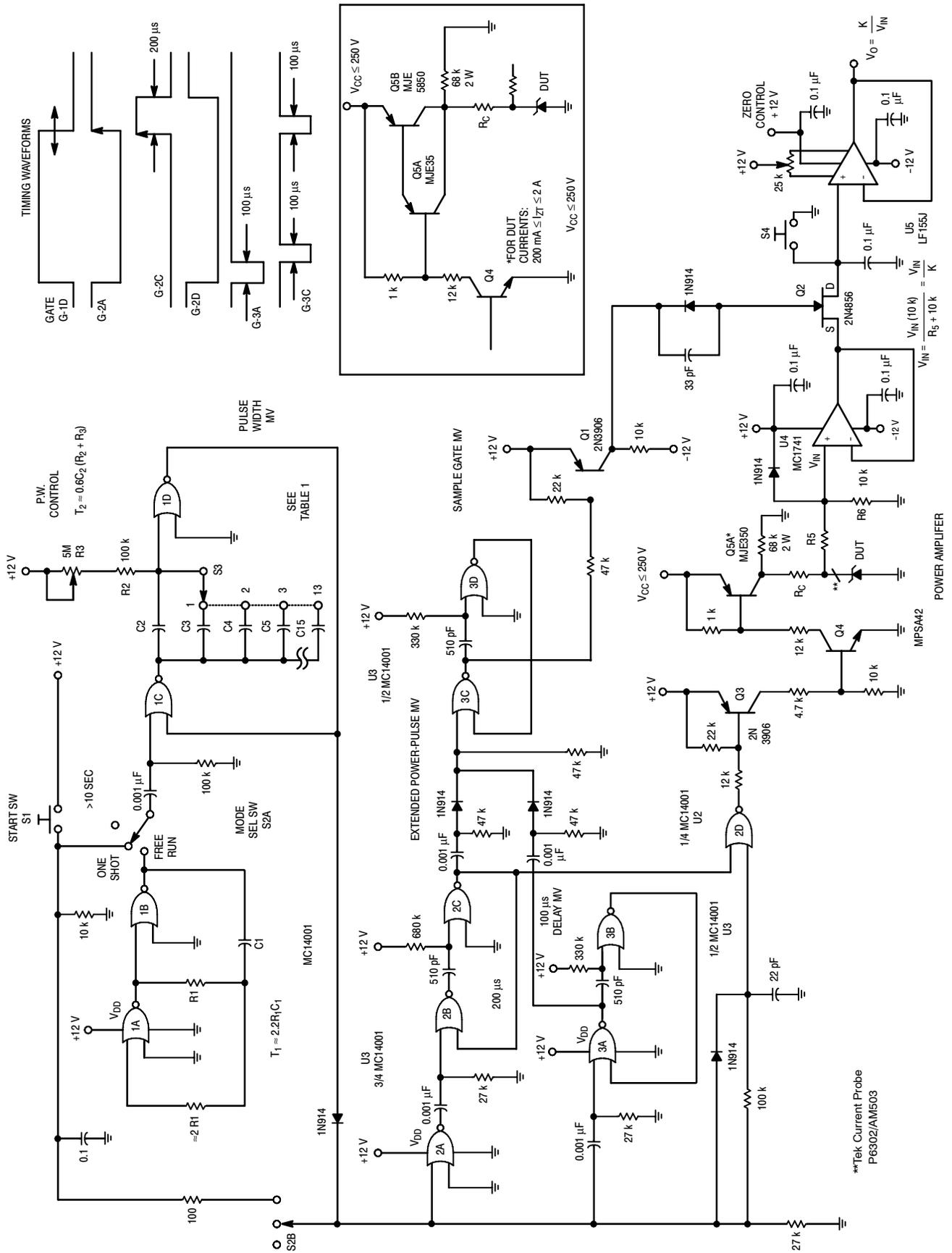


Figure 8. Zener Voltage Double Pulsing S/H Test Circuit

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 its 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 $30 \pm 1^\circ\text{C}$, 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 $11 \text{ V} - 0.10 \text{ V} = 10.9 \text{ V}$

Set V_Z min. limit at $9 \text{ V} - 0.07 \text{ V} = 8.93 \text{ V}$

- 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 $2.7 \text{ V} + 0.07 \text{ V} = 2.77 \text{ V}$

Set V_Z max. limit at $3.3 \text{ V} + 0.05 \text{ V} = 3.35 \text{ V}$

ON Semiconductor and  are registered trademarks of Semiconductor Components Industries, LLC (SCILLC). SCILLC reserves the right to make changes without further notice to any products herein. SCILLC makes no warranty, representation or guarantee regarding the suitability of its products for any particular purpose, nor does SCILLC assume any liability arising out of the application or use of any product or circuit, and specifically disclaims any and all liability, including without limitation special, consequential or incidental damages. "Typical" parameters which may be provided in SCILLC data sheets and/or specifications can and do vary in different applications and actual performance may vary over time. All operating parameters, including "Typicals" must be validated for each customer application by customer's technical experts. SCILLC does not convey any license under its patent rights nor the rights of others. SCILLC products are not designed, intended, or authorized for use as components in systems intended for surgical implant into the body, or other applications intended to support or sustain life, or for any other application in which the failure of the SCILLC product could create a situation where personal injury or death may occur. Should Buyer purchase or use SCILLC products for any such unintended or unauthorized application, Buyer shall indemnify and hold SCILLC and its officers, employees, subsidiaries, affiliates, and distributors harmless against all claims, costs, damages, and expenses, and reasonable attorney fees arising out of, directly or indirectly, any claim of personal injury or death associated with such unintended or unauthorized use, even if such claim alleges that SCILLC was negligent regarding the design or manufacture of the part. SCILLC is an Equal Opportunity/Affirmative Action Employer. This literature is subject to all applicable copyright laws and is not for resale in any manner.

PUBLICATION ORDERING INFORMATION

LITERATURE FULFILLMENT:

Literature Distribution Center for ON Semiconductor
P.O. Box 61312, Phoenix, Arizona 85082-1312 USA
Phone: 480-829-7710 or 800-344-3860 Toll Free USA/Canada
Fax: 480-829-7709 or 800-344-3867 Toll Free USA/Canada
Email: orderlit@onsemi.com

N. American Technical Support: 800-282-9855 Toll Free
USA/Canada

Japan: ON Semiconductor, Japan Customer Focus Center
2-9-1 Kamimeguro, Meguro-ku, Tokyo, Japan 153-0051
Phone: 81-3-5773-3850

ON Semiconductor Website: <http://onsemi.com>

Order Literature: <http://www.onsemi.com/litorder>

For additional information, please contact your
local Sales Representative.

HBD854/D