Surge Suppressors and Clamps

Francois D. Martzloff
National Bureau of Standards
Gaithersburg MD 20899

Reprint from *Proceedings, EMC EXPO 1986*, Washington DC, June 1986

Significance
Part 6: Textbooks, tutorials, and reviews
Part 7: Mitigation techniques

A tutorial paper with a review of the technologies available in the mid-eighties for surge protection of low-voltage end-user equipment. The principles of operation are described for "crowbar devices" (now generally described under the standardized name of "voltage switching devices") and for "voltage clamping devices: (now generally described under the standardized name of "voltage limiting devices"). Measurement techniques are also described, in particular the parasitic effect of inducing an addition voltage into the loop formed bt a differential probe connection.

Comparisons are offered between spark gap and varistor SPDs, as well as two fusing options for what is now called a "disconector" component in a packaged SPD.

SURGE SUPPRESSORS AND CLAMPS

François D. Martzloff

National Bureau of Standards
Gaithersburg MD 20899

INTRODUCTION

Various devices have been developed for the protection of electrical and electronic equipment against transient overvoltages. They are often called "transient suppressors" although, for accuracy, they should be called "transient limiters," "clamps," or "diverters" because they cannot really suppress the transients; rather, they limit the transients to acceptable levels or make them harmless by diverting them to ground. The IEEE dictionary has selected the more generic but lengthy term of "surge protective device."

There are two categories of transient suppressors: those that block transients, preventing their propagation toward sensitive circuits, and those that divert transients, limiting residual voltages. Since many of these transients originate from a current source, blocking them may not always be possible because the current forced into the high-impedance blocking path would only result in higher voltages and breakdown. Therefore, diverting of the transient is more likely to find general application. A combination of diverting and blocking can be a very effective approach. This approach generally takes the form of a multistage circuit, where a first device diverts the transient current to ground; a second device offers a restricted path for transient propagation but an acceptable path for the signal or power, and a third device clamps the residual transient (Figure 1). Thus, we are primarily interrested in diverting devices.

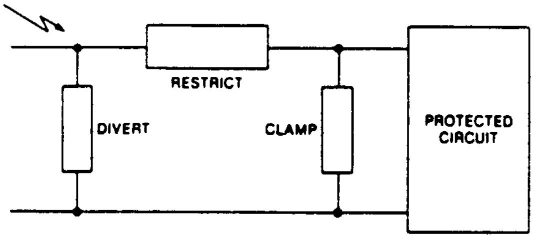

Figure 1. Multistage protection

The diverting device can be of two kinds: voltage-clamping or short-circuiting devices (the latter called "crowbars"). Both of them involve some nonlinearity, either frequency nonlinearity (as in filters) or, more usually, voltage nonlinearity. This voltage nonlinearity is the result of two different mechanisms -- a continuous change in the device conductivity as the current increases, or an abrupt switching action as the voltage increases.

Because the technical and trade literature contains many articles on these devices, we shall limit the discussions of the details but make some comparisons to point out the significant differences in the performance. Understanding the relative merits and the limitations of each technology will allow informed choice and recognition that there is room for each approach in the wide range of needs and applications.

CROWBAR DEVICES

The principle of crowbar devices is quite simple: upon occurrence of an overvoltage, the device changes from a high-impedance state to a low-impedance state, offering a low-impedance path to divert the surge to ground. This switching can be inherent to the device, as in the case of spark gaps involving the breakdown of a gas or be the recently introduced combination of multijunction solid state devices.

The major advantage of the crowbar device is that its low impedance allows the flow of substantial surge currents without the development of high energies within the device itself; the energy has to be spent elsewhere in the circuit. This "reflection" of the impinging surge can also be a disadvantage in some circuits if the transient disturbance associated with the gap firing is being considered. Where there is no power-follow (discussed below), such as in some communication circuits, the spark gap has the advantage of very simple construction with potentially low cost.

The crowbar device, however, has three limitations. One is the volt-time sensitivity of the breakdown process in air gaps or gas tubes. As the voltage increases across a gap, significant conduction of current -- therefore voltage limitation for the surge -- cannot occur until the transition to the arc mode of conduction, by avalanche breakdown of the gas between the two electrodes. The load is unprotected during the initial rise because of this delay time (typically in microseconds). Large variations exist in sparkover voltage attained in successive operations since the process is statistical in nature. This sparkover voltage can also be substantially higher after a long period of rest than after successive discharges. Because of the physics of the process, it is difficult to produce consistent sparkover voltage for low voltage ratings. This difficultly is increased by the effect of manufacturing tolerances on very small gap distances, but can be alleviated by filling the tube with a gas having lower breakdown voltage than air. The technology developed by manufacturers of gas tubes has minimized these effects. One advantage of the solid-state crowbars over the gas breakdown tubes is the absence of any practical time delay in firing.

The second limitation is associated with the speed of the sparkover, which produces fast current rise in the circuits and, thus objectionable noise. An example is found in hybrid protective systems. Figure 2 shows the circuit of such a commercially available device. The gap does a very nice job of diverting impinging high energy surges, but the magnetic field associated with the high di/dt induces a voltage in the loop adjacent to the second suppressor, adding a substantial spike to what was expected to be a low clamping voltage. Some product literature advocates locating solid-state crowbars on the very circuit board where components to be protected are mounted -- a perfect recipe for injecting interference right into the circuits!

A third limitation occurs if power current from the steady-state voltage source can follow the surge discharge (hence the term "power-follow"). In ac circuits, this power follow current may or may not be cleared at a natural current zero. Additional means, therefore, must be provided to open the power circuit if the crowbar device is not designed to provide self-clearing action within specified limits of surge energy, system voltage, and power-follow current. This combination of a gap with a current-limiting, nonlinear varistor has been very successful in the utility industry as high voltage surge arresters.

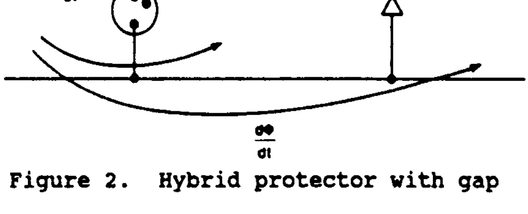

Figure 2. Hybrid protector with gap

VOLTAGE-CLAMPING DEVICES

Voltage-clamping devices have variable impedance, depending on the current flowing through the device or the voltage across its terminal. Impedance variation is monotonic and does not contain discontinuities, in contrast to crowbar devices, which show a turn-on action. As far as their volt-ampere characteristics are concerned, these devices are time-dependent to a certain degree. However, unlike the sparkover of a gap or the triggering of a thyristor, time delay is not involved.

When a voltage clamping device is applied, the circuit remains essentially unaffected by the device before and after the transient for any steady-state voltage below clamping level. Increased current drawn through the device as the surge voltage attempts to increase results in voltage-clamping action. Nonlinear impedance is the result if this current rise is greater than the voltage increase. The increased voltage drop in the source impedance (Z_S in Figure 3) due to higher current results in the apparent clamping of the voltage. It must be emphasized that the device depends on the <u>source impedance</u> to produce this clamping. The circuit behaves as a voltage divider where the source impedance (high side of the divider) is constant but the clamping device impedance (low side of the divider) is changing. If the impedance of the source is very low, the ratio is low, and eventually the suppressor could not work at all with a zero source impedance. In contrast, a crowbar-type device effectively short circuits the transient toward ground, but, once established, this short-circuit will remain until the current (the surge current as well as any power-follow current supplied by the power system) is brought to a low level.

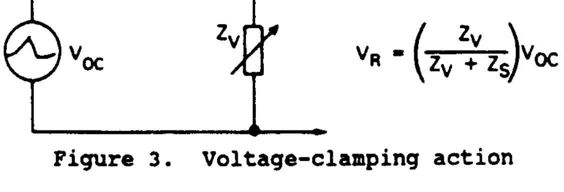

Figure 3. Voltage-clamping action

The action of voltage clamping can be performed by any device exhibiting a nonlinear impedance. Two categories of such devices, having the same effect but operating quite different physical processes, have found an acceptance in the industry: polycrystalline varistors and single-junction avalanche diodes. Another technology, using selenium rectifiers, has been practically eliminated because of the improved characteristics of modern varistors.

Avalanche diodes

Avalanche diodes were initially applied as voltage clamps in the form of Zener diodes, a natural outgrowth of their application as voltage regulators. Improved construction, specifically aimed at surge absorption, has made them very effective suppressors. Large diameter junction and low thermal impedance connections are used to deal with the problem of dissipating the heat of the surge, inherent in a thin single-layer junction.

The advantage of the avalanche diode is the possibility of obtaining quite low clamping voltages and nearly flat volt-ampere characteristics over its useful power range. Therefore, these diodes are widely used in low-voltage electronic circuits to protect 5 V or 15 V logic circuits, as an example. At higher voltages, the problem of the heat generation associated with single junctions can be overcome by stacking several lower voltage junctions.

Characteristics of Avalanche Diodes --- Silicon avalanche diodes are available with characteristics especially tailored to providing transient suppression. These special diodes must not be confused with regulator-type Zener diodes although many engineers tend to use the generic term "Zener diode." May Zeus help them if they were to misapply a regulator-type Zener, expecting effective protection!

Figure 4 shows a typical family of the V-I characteristics for one product. The parameter I is the peak of a specified current waveform, generally the 10/1000 µs double-exponential test waveform preferred by communication engineers, rather than a 8/20 µs waveform preferred by power engineers. The V-I curves are remarkably flat from 1 to 20 amperes, with clamping voltages low enough to protect sensitive electronics.

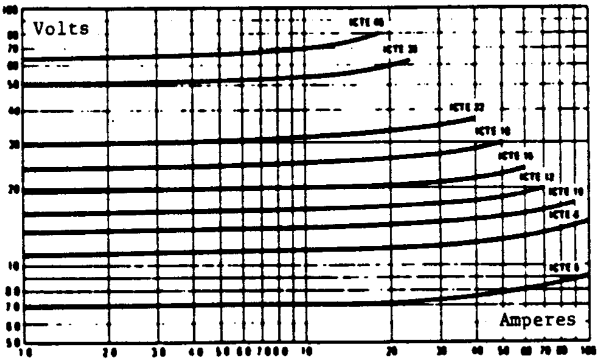

Figure 4. Typical V-I characteristic

Manufacturers of avalanche diodes generally rate their devices in terms of maximum peak power for a specified pulse duration. Even though the rated power will decrease with increasing pulse duration, it is not a case of constant total energy (see Figure 5). A two-decade increase in pulse duration will reduce the power by only about one decade, primarily because of increasing heat transfer for longer pulses.

Since the junction is very thin, the capacitance of an avalanche diode is appreciable. This capacitance can be a concern in high-frequency circuits where it would produce an undesirable insertion loss. It is possible to minimize this effect by using a combination with a low-capacitance ordinary diode in series with the avalanche diode.

Properly packaged and installed avalanche diodes exhibit a quick response to steep-front pulses, and have been used for NEMP protection. However, this quick response can be completely obliterated by improper wiring. The effect of lead length, discussed in detail in the next section, is applicable to any transient suppressor.

Varistors

The term varistor is derived from its function as a variable resistor. It is also called a voltage-dependent resistor, but that description implies that the voltage is the independent parameter in surge protection, an incorrect perception. Two very different devices have been successfully developed as varistors: 1) silicon carbide disks have been used for years in the surge arrester industry and 2) metal oxide varisristors are now widely used.

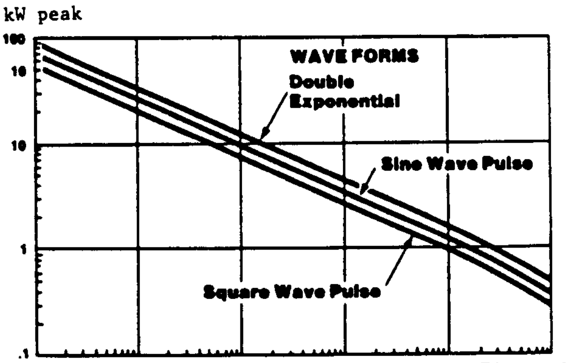

Figure 5. Peak pulse power vs. pulse time

In silicon carbide varistors, as well as in metal oxide varistors, the relationship between the current flowing in the device and the voltage appearing across its terminals can be represented approximately by a power function $I = kV^{\alpha}$. In this equation, the exponent α, can be considered as a figure of merit: the higher the value of α, the more effective the clamping. Hence, there has been a race between manufacturers and specification writers for higher and higher values of α. We will see, however, that there are practical limits to this race and that better performance can be obtained at higher current densities by departing somewhat from large values of the exponent α.

In silicon carbide varistors, the physical process of nonlinear conduction is not completely understood, and the manufacturing of the material, successful as it is, has remained an art. It appears that the process takes place at the tips of the grains of silicon carbide which are held together by a binder. The story is told that this device's action was found accidentally by having a grinding wheel, on a disorderly work bench, inadvertently connected to an experimental circuit; for many years the silicon carbide varistors indeed looked like grinding wheels, each complete with a hole in the center.

Metal oxide varistors depend on the conduction process occurring at the boundaries between the grains of oxide (typically zinc oxide) grown during a carefully controlled sintering process. The physics of the nonlinear conduction have been described in the literature [1-4]; in these application notes, we will be more concerned with the behavior of the varistors as two-terminal electrical components.

Electrical Characteristics of Varistors--- Because the prime function of a varistor is to provide the nonlinear effect, the other parameters are generally the result of a tradeoff in design and inherent characteristics. Electrical behavior of a varistor can be understood through the equivalent circuit of Figure 6. The major element is the varistor proper, R_V, whose V-I characteristic is assumed to be the perfect power law $I = kV^{\alpha}$. In parallel with this varistor, there is a capacitor, C, and a leakage resistance, R_p. In series with this three-component group, there is the bulk resistance of the zinc oxide grains, R_S, and the inductance of the leads, L.

Under dc conditions (at low current densities, because obviously no varistor could stand a high energy produced by dc currents of high density), only the varistor element and the leakage resistance are significant. Under pulse conditions at high current densities, all but the leakage resistance are significant: the varistor provides low impedance to the current flow, but eventually the series resistance produces an upturn in the V-I characteristic; the lead inductance can produce spurious overshoot problems if it is not dealt with properly; the capacitance can offer either a welcome additional path for fast transients or an obectionable loading for high-frequency signals, depending on the application.

V-I Characteristics --- When the V-I characteristics are plotted on a log-log graph, the curve of Figure 7 is obtained. It has three regions as shown, resulting from the dominance of R_p, R_V, R_S as the current in the device goes from nanoamperes to kiloamperes.

The V-I characteristic is the basic application design tool for selecting a device in order to perform a protective function. For successful application, however, other factors, which are discussed in detail in information available from manufacturers, must also be taken into consideration.

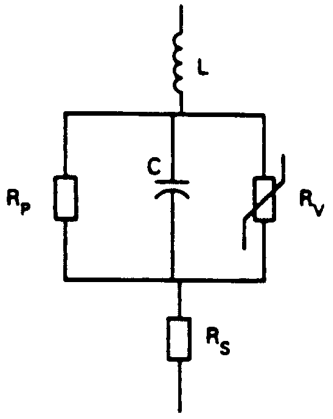

Figure 6. Varistor equivalent circuit

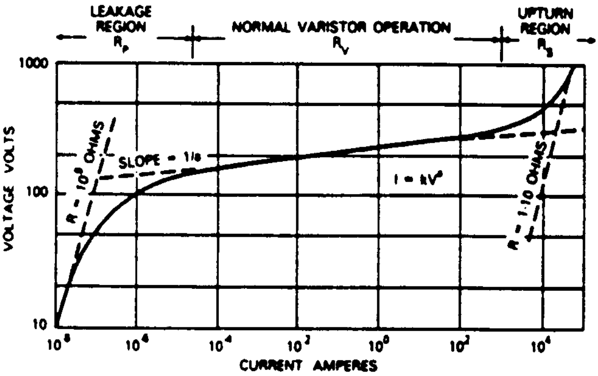

Figure 7. Theoretical V-I characteristic

Some of these factors are:

- Selection of the appropriate nominal voltage for the line voltage (Avoid unnecessarily low clamping.)
- Selection of energy-handling capability (Consideration of the source impedance, waveshape, and number of occurrences of the transient [5].)
- Heat dissipation (Ambient temperature, steady state and transient energy.)
- Proper installation in the circuit (Lead length).

In fact, enough instances of poor installation practices have been observed that a brief discussion of lead effects is quite in order.

When making voltage measurements across a clamping device for evaluating its performance, one must recognize possible instrumentation artifacts which require special precautions to avoid errors:

1. Use two probes in a differential mode to make a measurement directly at device terminals, and not a single-ended system.

2. Avoid contaminating the true device voltage by the additional voltage caused by magnetic coupling into the probes loop.

Commercial oscilloscope preamplifiers offer a wide choice of differential mode operation, either through an [add + invert] mode of two-channel preamplifiers, or through a differential amplifier built specifically for high common-mode rejection (sometimes at the expense of bandwidth). Thus, careful attention must be given to these two aspects of measurements.

The voltage measured by the two probes (or the voltage that would be applied to a protected load downstream from the device) is the sum of the actual clamping voltage at the device terminals and a spurious voltage caused by magnetic coupling. This spurious voltage is induced into the loop formed by the clamping device leads and the probes (or the downstream wiring in the case of an actual protective scheme) by the changing magnetic field of the surge current flowing in the device.

To illustrate this situation, the measurement circuit shown in Figure 8 was set up in the output circuit of a generator producing a 8/20 µs impulse. The "device" was a hollow conductor, with a hole at the center through which a twisted pair was fed, one wire of the pair branching out to each end of the conductor, separated by 10 cm. At the same 10-cm separation, but outside of the hollow conductor, two thin wires were also soldered, brought to the midpoint of the hollow conductor and in close contact with the conductor; from the midpoint outward, they were twisted in the same manner as the inside pair. A third pair of wires was soldered at the end points of the hollow conductor, and arranged to form a rectangle, the hollow conductor being one side of that rectangle. Various widths were set up for the rectangle, and each time the measured voltage was recorded. Figure 9 shows measured voltage vs. radial distance of the opposite side of the rectangle. The effect is present even for close proximity, and reaches a saturation beyond 10 cm.

This example shows that one must not only connect the probes as close as possible to the clamping devices terminals, but also strive to minimize the area established by the probes close to the devices.

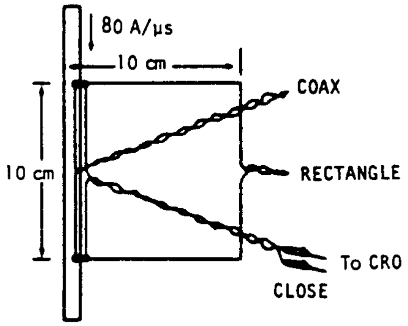

Figure 8. Circuit configuration for measurement of voltage drop along conductor with various probe geometries

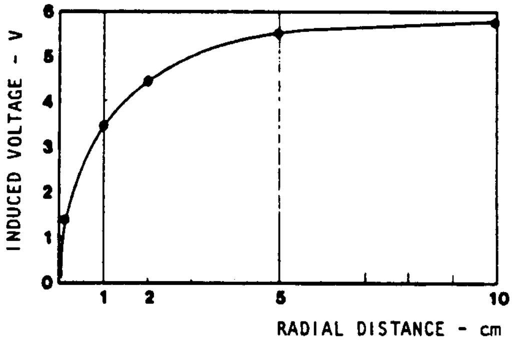

Figure 9. Voltage observed by probes versus radial distance from conductor to opposite side of rectangular loop

In this case of a low-voltage suppressor, it would be better to solder short leads to the device terminals, bring them together while hugging the device, then twist them into a pair and connect the oscilloscope probes some distance away from the device. The same conclusion is applicable to the wiring layout in actual hardware: Creating a loop near the protective device is an invitation to induce additional voltages in the output of the protective device, thus losing effectiveness (Figure 3).

Hence, when one is making measurements as well as when one is designing a circuit for a protection scheme, it is essential to be alert to the effects of lead length (more accurately of loop area) for connecting the protective devices. This warning is especially important when the currents are in excess of a few amperes with rise times of less than 1 μs.

COMPARISONS OF PROTECTIVE DEVICES

Linear Versus Nonlinear Devices

When a protection scheme is designed for an electronic system operating in an environment which is not completely defined, it is often necessary to make an assumption about the parameters of the transient expected to occur. In particular, if an error is made in assuming the source impedance of this transient, the consequences are dramatically different between a linear device for which the protective level will be affected in proportion to the error, and a nonlinear device where the protective level will be affected approximately in inverse ratio to the exponent of the V-I characteristic.

Spark Gap Versus Varistor

The choice between these two devices will be influenced by the inherent characteristics of the application. Where power-follow may be a problem, there is little opportunity to apply a simple gap. Where very steep-front transients occur, the gap alone may let an excessive voltage go by the "protected" circuit until the voltage is limited by sparkover. Where the capacitance of a varistor is objectionable, the low inherent capacitance of a gap will seem attractive. If very high energy levels can be deposited in a varistor (the power dissipated remains as heat energy trapped in the bulk material), compared to the lower levels inherent to the crowbar action of a gap, then a surge arrester of any type with high capacity connected at the service entrance may be combined with a varistor of lower clamping voltage installed further in the circuit. This combined protection, however, requires proper coordination between the two suppressors [6].

Avalanche Diode Versus Varistor

The basic performance characteristics of these two devices are similar, and therefore the choice may be dictated by clamping voltage requirements (avalanche diodes are available at lower clamping voltages), by energy-handling capabilities (avalanche diodes are generally lower in capability per unit of cost), or by packaging requirements (varistor material is more flexible and does not require hermetic packaging).

FAILURE MODES

Failure of electrical components can occur because their capability was exceeded by the applied stress or because some latent defect in the component went by unnoticed in the quality control processes. This situation is well recognized for ordinary components but a surge protective device, which is no exception to these limitations, tends to be expected to perform miracles, or at least to fail graciously in a "fail-safe" mode.

The term "fail-safe," however, means different failure modes to different users, therefore it should not be used. To some users, fail-safe means that the protected hardware must never be exposed to an overvoltage, so that failure of the protective device must be in the fail-short mode, even if it puts the system out of operation. To other users, fail-safe means that the function must be maintained, even if the hardware is left temporarily unprotected, so that failure of the protective device must be in the open-circuit mode. It is more accurate and less misleading to describe a failure mode "fail-short" or "fail-open," as the case may be.

When the diverting path is a crowbar-type device, little energy is dissipated in the crowbar, as noted earlier. In a voltage-clamping device, because more energy is deposited in the device, the energy handling capability of a candidate protective device is an important parameter to consider in the design of a protection scheme. With nonlinear devices, an error made in the assumed value of the current surge produces little error on the voltage developed across the protective device and applied to the protected circuit (the error, however, affects directly the amount of energy which the protective device has to absorb. At worst, when surge currents in excess of the protective device capability are imposed by the environment, such as an error made in the assumptions, a human error in the use of the device, or because nature tends to support Murphy's law, the circuit in need of protection can generally be protected at the price of failure in the short-circuit mode of the protective device. However, if substantial power-frequency currents can be supplied by the power system, a fail-short protective device generally terminates as fail-open when the power system fault in the failed device is not quickly cleared by a series overcurrent protective device (a fuse or a breaker).

When there is a need to eliminate a failed protector at the specific equipment level, insertion of the fuse in the line provides protection of the equipment (Figure 10a), while insertion of the fuse in series with the shunt-connected protector provides protection of the function, albeit with loss of overvoltage protection (Figure 10b). A fuse must be selected with suitable i^2t to withstand the effect of repetitive surges expected to flow in the application.

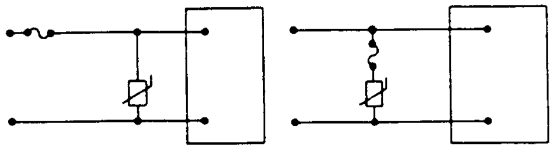

Figure 10. Fusing options for suppressors

CONCLUSIONS

1. Surge protective devices are available for protecting low-voltage electronics. Two basic types offer different advantages:

- Crowbar devices have a high current capability but they would produce a power-follow, if applied in a power system; their fast switching action can be a cause of problems if it is not recognized.
- Voltage clamping devices, either the avalanche junctions or the varistors, are free from the problems of power-follow and interference caused by a fast switching, but the energy which is deposited in these must be recognized for long-term reliability.

2. Avalanche diodes offer low clamping voltage, which makes them most suitable for low-voltage, low-power electronics.

3. Metal-oxide varistors are now available in a wide range of forms, clamping voltages and energy-handling capacities.

4. Each of these devices has its own best field of application for better reliability of the circuits in the not-quite-defined electromagnetic environment of power and communication systems.

REFERENCES

1. Matsuoka, M., Masa Yama, T. and Lida.,Y. "Nonlinear Electrical Properties of Zinc Oxide Ceramics," Proc. First Conf. Solid State Devices, Tokyo, 1969, J. Japan Soc. Appl. Phys., 39 1970, Suppl. p.94.

2. Harnden, J.D., Martzloff, F.D., Morris, W.G. and Golden, F.B., "The GE-MOV varistor -- The Super Alpha Varistor," Electronics, 45, no.21, 1972, p.91.

3. Sakshaug, E.C., Kresge, J.S. and Miske, S.A., "A New Concept in Station Arrester Design," IEEE PAS-96 No.2, March-April 1977, p.647.

4. Mahan, G.D., Levinson, L.M., and Philipp, H.R.,"Theory of Conduction in ZnO Varistors," Report 78CRD205, General Electric Company, Schenectady, NY, 1978.

5. Martzloff, F.D., "Matching Surge Protective Devices to Their Environment," IEEE IA-21 No.1, Jan/Feb 1985, p.99.

6. Martzloff, F.D., "Coordination of Surge Protectors in Low-Voltage Power Circuits," IEEE PAS-99 No.1, Jan/Feb 1980, p.129.

ABOUT THE AUTHOR - François D. Martzloff is a graduate from French engineering schools and continued his studies at Georgia Tech for an MS, EE and at Union College for an MS in Industrial Administration.

His 29-year career at General Electric Company included design of high voltage bushings, fuses, and 20 years in the area of transient overvoltages measurements and suppression, especially with varistors.

In 1985, he joined the National Bureau of Standards, Electrosystems Division, with a charter of measurements and correction of Conducted Electromagnetic Interference.

As a Fellow of the IEEE, he is active in the Surge Protective Devices Committee, as the Chairman of the Working Group in Surge Characterization in Low-Voltage Circuits. He is also a member of the Surge Protection and Wave Distortion Subcommittee.

He is serving in advisory capacity to two International Electrotechnical Commission Technical Committees and to the Industry Advisory Group of the Underwriters Laboratories on Transient Suppressors.