$\int_{\Psi}$ *INTEGRATED SCIENCES*™

EMP AND HPM SUPPRESSION TECHNIQUES

This work Sponsored by

The Defense Special Weapons Agency

RDT & E Code

B 4662 D DE 63132 8010 A AF

Prepared for

Director

Defense Special Weapons Agency

6801 Telegraph Road
Alexandria, Virginia 22310-3398

6635 South New Haven Tulsa, Oklahoma 74136 (918) 493-3399

REPRODUCTION QUALITY NOTICE

This document is the best quality available. The copy furnished to DTIC contained pages that may have the following quality problems:

- Pages smaller or larger than normal.
- Pages with background color or light colored printing.
- Pages with small type or poor printing; and or
- Pages with continuous tone material or color photographs.

Due to various output media available these conditions may or may not cause poor legibility in the microfiche or hardcopy output you receive.

☑ If this block is checked, the copy furnished to DTIC contained pages with color printing, that when reproduced in Black and White, may change detail of the original copy.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	10 Mar 97	Final 29 Feb 96 10 Mar 97

4. TITLE AND SUBTITLE
EMP and HPM Suppression Techniques

5. FUNDING NUMBERS
C - DNA001-96-C-0039
PE - 62715H
PR - AF
TA - DE
WU - 63132

6. AUTHOR(S)
Clinton Dutcher, William Alston, Wayne Gindrup

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Integrated Sciences™
6635 South New Haven
Tulsa, OK 74136-2843

8. PERFORMING ORGANIZATION REPORT NUMBER
DNA001-96-C-0039

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Special Weapons Agency
6801 Telegraph Road
Alexandria, VA 22310-3398
ESE Walter Scott

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
UL

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Micron sized metal particles embedded in an insulative binder exhibit varistor action. It is believed the dominant conduction mechanism in successful devices is inter-metallic quantum tunnelling. In this report, the quantum tunnelling aspects are thoroughly investigated. Computer models are developed based on the quantum analyses along with the introduction of chaos theory. The computer models are presently limited to two dimensions, so predictions are basically qualitative. However, the qualitative results are observed empirically. The computer models and the empirical results indicate there are one or more conduction mechanisms in addition to quantum tunnelling. Devices were constructed and tested according to ESD specifications. For two classes, the failure rate was zero. Some leading edge overshoot was observed, but it is believed this is mainly an artefact of the test system configuration.

14. SUBJECT TERMS
Varistor, EMP, HPM, quantum tunnelling

15. NUMBER OF PAGES
216

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

INTRODUCTION

All advanced electronic systems technology is useless if it can not survive in the presence of its own electromagnetic interference as well as that imposed from without.

There is an ever increasing dependence on microelectronic circuitry for system control, communications, and monitoring, many in life dependency situations. Specific examples are land and space based information transmission systems, hospital monitoring systems, and aircraft fly-by-wire control systems. The associated circuits are becoming smaller, faster and operate at decreasingly lower voltages. These factors lead to increasing susceptibility to self generated and external electromagnetic interferences, particularly those that are in the class of overvoltage transients, the ones with which the present effort is concerned.

The sources of these transients are numerous. The most common source is a wide variety of electrostatic discharges (ESD) originating from human body contact and even physical circumstances encountered in electronics manufacturing processes. Another common source is overvoltage transients appearing on general power supply lines. These are most frequently induced by switching events or coupling to the power lines from external sources. Lightning is in the induced category. It is unlikely that any transient suppression device will survive or protect from a direct lightning hit; however, direct hits are extremely rare. Most damage to electronic circuits caused by lightning is the result of the associated electromagnetic pulse or associated conducted emissions. A much less likely threat is that of nuclear electromagnetic pulse or enemy generated high power microwaves, less likely, but with potentially devastating effects to both civilian and military systems.

To mitigate these threats, a number of techniques are presently used. Included are the use of special device and circuit design techniques. Although of definite advantage, these techniques generally do not provide adequate protection where substantial threats are to be encountered. Other methods include the use of peripheral devices such as gas discharge tubes, semiconductor diodes, including Zeners, and metal oxide varistors. Combined with good shielding practices, the combination of these methods can provide systems substantially hardened to transient voltages. However, these combined practices add cost in terms of the components needed and additional manufacturing processes. They also add weight, and the protective components can interfere with the normal operation of the circuits to be protected.

To obtain a clear picture of what is desired in a Transient Voltage Surge Suppression (TVSS) device, it is advantageous to think in terms of two

regimes, the normal *operational regime*, where no significant overvoltages are present and the *threat regime* where a potentially damaging overvoltage is present. In the operational regime, we want the system to behave exactly as if the TVSS device were not present, we refer to this as *system transparency*. In the threat regime, we want the TVSS device to divert deleterious currents into paths that do not compromise system operation. When the threat has passed, we want the TVSS device to return to its system transparent state and be ready to counter the next overvoltage threat.

Transient voltages can be large (on the order tens of thousands of volts) and fast (rise times less than 10^{-9} second). The high voltages lead to high currents which in turn lead to high powers. If the time duration of the transient is long, the latter leads to high energies being transmitted to the TVSS device. In many TVSS devices, when a fast transient is applied, there is a leading edge overshoot before the final clamp voltage. The energy content of this overshoot can be sufficient to destroy the circuits to be protected. We can think of the desirable characteristics of TVSS devices in the following terms: In the operational regime, system transparency. In the threat regime, fast switching times, adequately low clamping voltages with no overshoots, and adequately high energy handling capabilities. *Metal polymer varistor devices have the potential of simultaneously meeting all these requirements. None of the presently existing TVSS devices does.* The Metal Polymer Varistor devices are those to which this effort is directed.

A major advantage of the Metal Polymer Varistor (MPV) is its moldability. This allows substantial freedom in device geometry and reduced manufacturing costs. From a system standpoint, devices are more easily produced for specific system needs and can be directly incorporated into the system. A good example of this is the incorporation of MPV materials into the front end connectors of a communications or sensor system where the electrical properties of the MPV material are chosen so that no impedance mismatches are created. The connector then becomes the TVSS device and is system transparent in the operational regime. This same concept brought to the circuit level is exemplified by information processing circuits where striplines are small transmission lines. Addition of standard TVSS devices produces nontransparency in the operational regime; whereas, impedance matched MPV devices provide system transparency, and the question of added capacitances simply vanishes.

Where semiconductor junction devices are used, there is an insidious problem that may occur in information transformation systems. The junction capacitances are nonlinear, leading to the generation of modulation products causing interference in multichannel systems. MPV devices do not have this problem.

Members of the European Common Market are acutely aware of TVSS needs and are developing stringent specifications for all products imported to them. American manufacturers will ultimately need to follow suit and

are presently moving in that direction. *Development of superior TVSS devices in the United States will not only give a competitive edge to exported products but will give a competitive edge to products sold in the United states in competition with imported foreign goods.*

The advantages of MPV devices have been known to industry for a number of years. However, it appears corporate managers underestimate the need for understanding the basic underlying electrophysical mechanisms as a prerequisite for commercial success. They further fail to understand that the processing and measurement techniques needed to verify that understanding may be different from what is needed in a production operation. That has left a crucial information gap. The present project is a first step toward closing that gap. Our emphasis in Phase I is on basic principles. *Those principles will be used in Phase II to optimize materials and devices from both a manufacturability and economic viewpoint, leading to higher performance TVSS devices at lower costs.*

Part I of this report is a summary of progress to date, along with conclusions reached and recommendations for future effort. Parts II through V give the technical information on which the conclusions and recommendations are based.

Acknowledgements: *The principal author of Part III of this document is Integrated Sciences™ Executive Associate, Mr. William W. Alston.*

The principal author of Part IV of this document is Spectro Dynamic Systems V.P., R & D, Mr. Wayne L. Gindrup

Clinton Dutcher, Ph. D.
V. P. Technology Development
Principal Investigator
March 1997

CONTENTS

Section I-1

SUMMARY RESULTS AND CONCLUSIONS

OF THE PHASE I EFFORT

I-1-1 Project Objectives

With regard to quantum tunnelling varistors, the project objectives were as follows:

- Provide *Quantum, Electrical, and Thermal Analyses* as input to a finite element computer program.
- Develop *Finite Element Modelling* incorporating nonlinear quantum effects, thermodynamics, and statistical chaotic behavior to provide a basis for guidance in varistor design.
- Produce *Test Sample Preparations* for laboratory testing.
- Use *Laboratory Measurements* for empirical verification of predicted results.
- Produce varistor material formulations demonstrating the feasibility of quantum tunnelling varistors to meet one or more extant *Surge Arresting Specifications* where device electrical properties are independent of binder materials used.

These objectives have been substantially met and are further summarized in the subsections to follow. The results of the project are described in greater detail in the body of this report.

I-1-2 Quantum, Electrical, and Thermal Analyses

Another objective of the effort was to produce varistor materials where the dominant nonlinear effects are produced by pure intermetallic quantum tunnelling. It is then desirable to understand this phenomenon at a fundamental level. To that end, previous work was summarized and extended with a rigorous treatment of the quantum mechanical aspects, including temperature, barrier reflection, and Schottky effects. *It was predicted that temperature would not have a practical effect on pure quantum tunnelling devices.*

A simple expression for quantum tunnelling current densities was derived in the early 1900's by Nordheim and Fowler. Although the mathematics of quantum tunnelling are somewhat complex, *we showed the form of the simple Nordheim-Fowler result is an excellent practical approximation, even when barrier reflection and Schottky effects are included.* This

approximation was then used as the principal nonlinearity in the finite element analyses to be subsequently described.

The essential thermal aspects of varistor devices were developed. Since we were concentrating here on fast transients, no thermal diffusion was considered in the analytical treatment, representing a worst case, single pulse, threat scenario. However, *thermal diffusion is incorporated in our computer programs*, a feature needed in modelling the "Electrical Fast Transient" threat.

A charge transport formulation was developed, leading to the conclusion that mean free path calculations would be sufficiently accurate. This is followed by an analysis of quantum states and time scales which is an essential ingredient for the semiclassical treatment of electron flow in our varistor materials. These results are then applied to bulk conduction in insulators, with the indication that *building a reliable surge arresting device depending on bulk dielectric breakdown of any dielectric is highly unlikely and is particularly unlikely if the dielectric is amorphous, specifically, if it is polymeric. Thus thin inorganic insulative phases are indicated.*

Quantum tunnelling electrons are necessarily travelling at high speeds, about one hundredth the velocity of light. The relative electron densities in the metal and insulative phases were analyzed. Because of the high velocity of the tunnelling electrons, their density in the insulative phase need be only about eight orders of magnitude less than the density in the metallic phase. This significantly reduces the probability of an electron-lattice interaction in the insulative phase. *Low probability of electron-lattice interaction coupled with thin insulative phases reduces the probability of device failure.*

A number of elementary electromagnetic solutions were derived to provide test cases for the accuracy of our finite element programs.

I-1-3 Finite Element Modelling

Finite element programs were developed to model varistor performance. These programs, conceived for this project, from rigorous derivation to C/C++ computer code, address nonlinearities resulting from quantum effects, thermodynamics, regenerative phenomena, and statistical chaotic behavior. In their capability for treating non-deterministic phenomena, these programs and the underlying philosophy of computation at once extend and combine the disciplines of continuum mechanics and chaos theory. It was found that this computational strength was essential in predicting the chaotic failure mechanisms witnessed in laboratory experiments. Results show, at least in a qualitative sense, excellent agreement between computer predictions and experimental results.

Thorough testing of the finite element programs using relatively simple problems with known results showed agreement to at lease five decimal places. The varistor simulations were performed to include device geometries, applied voltages, and circuit constraints.

When the device nonlinearity was modelled using the Nordheim-Fowler approximation to define the local conduction mechanism, the models failed to exhibit the observed dominant path failure mechanism, where a single, narrow filament carries the bulk of the current flow. Statistical variability was imposed spatially on the material to reflect real-world variation in the spacing among conductive particles. This material variability in combination with the Nordheim-Fowler nonlinearity demonstrated spatial fluctuations in temperature, but was again unsuccessful in predicting dominant path failure.

It was found that when exponential regenerative effects were introduced into the material model, in combination with the Nordheim-Fowler approximation and statistical variability, models of dominant path phenomena were at last disclosed. This finding has a two-fold impact. First, *there are apparently three criteria which must be satisfied for dominant path failure to occur: extreme material nonlinearity with regard to the relationship between local voltage gradient and apparent electrical conductivity, spatial variability of conductive properties in the varistor material, and nonlinear regenerative effects.* Second, given that the regeneration criterion was not anticipated, *there is at least one undetermined conduction mechanism involved.*

Using the three criteria defined above, finite element modelling was used to find a method to quench dominant path failure. Starting with materials and conditions known through the models to exhibit dominant path, attempts were made to eliminate this failure mechanism by making subtle and easily achievable material changes. These attempts were met with apparent success. It was found that *superimposing a range of values of linear resistivity on the bulk conductive properties of the varistor material can eliminate dominant path failure.* Too little linear resistivity will not eliminate dominant path; too much will cause severe overheating.

This change in bulk property should be easily achievable through altered constructs of the conductive phase of the material, such as replacing metal spheres with metal-coated spheres of ceramic or glass. The goal is to continue to use the models to help develop a robust material having high resistance to dominant path failure, with correspondingly high energy handling capability, while still having sufficiently large tolerances reflecting ease and economy of manufacture.

There were other significant findings from the finite element models. Curvilinear shapes apparently do not alter the dominant path situation. Specifically, *coaxial and planar configurations will yield similar behavior.* Additionally, *for equivalent voltage gradients, reliability performance is*

substantially independent of interelectrode spacing, up to a relatively small threshold level.

The finite element programs have been constructed to predict the behavior of non-deterministic systems. They involve a real-world randomness. With their use it is not possible to know precisely where and how a material will fail with one computer run alone. However, with results of many runs, the information given involves the *nature* of material performance and failure modes. By contrast, creating material with no randomness whatsoever, material having perfect mathematical purity as defined in traditional models, is clearly impossible and not reasonable to attempt. We also note the programs have a much broader range of application than the one for which they were designed here. Some of these applications are briefly described in Section III-6.

Our programs have successfully modelled varistor performance and have shown an ability to predict chaotic failure mechanisms *qualitatively*. However, to achieve the goal of first modelling the behavior of a varistor before producing it in the laboratory, *quantitative* accuracy is essential. True quantitative accuracy is achievable with expansion of the finite element programs to full three-dimensional capability with higher order elements. This endeavor is achievable, but will require the further resources available during Phase II.

I-1-4 Material Fabrication

Under Integrated Sciences™ guidance, all the material fabrication was performed by Spectro Dynamic Systems using their proprietary processes. Not all of the results of the last two sections were originally fully appreciated. For the particle coatings, we started with organic materials. All the ones tested proved to be unsatisfactory. They were either too fragile at the thicknesses we thought we needed or failed by dominant path failure, displaying unacceptable energy handling capability.

The first inorganic coating tried was silver sulfide. Unfortunately, silver sulfide, though possibly classed as an insulator, is a poor one. Incorporated in these tests was our concept for mitigating dominant path failure through addition of bulk resistivity. The concept remained untried because of the poor performance of the silver sulfide.

We then sought a quick way of arriving at an inorganic highly insulative coating on any metal. Because of the known properties of aluminum oxide, this was a natural choice. Additionally, we had success with aluminum on a previous project conducted for the U. S. Army's Harry Diamond Laboratories. That success was repeated here, wherein *samples were produced and successfully independently tested by Littelfuse, Inc. in accordance with IEC 801-2 ESD test specifications*. Taking into account the relative energies associated with ESD and EMP, one would conclude *these*

samples would also be appropriate for mitigating the EMP threat, as previously demonstrated in the Harry Diamond tests.

A positive aspect of the aluminum tests is that they demonstrated the independence of the varistor action from any binder; *there was no binder in the first samples tested.* Subsequently, samples were prepared with silicone as the binder, with results similar to the ones without the binder. This gives an indication that we are producing devices whose electrical properties are independent of the binder, a major goal of this project. A drawback of the aluminum tests is they did not incorporate our concept for addition of bulk resistivity, primarily because we do not yet know how to economically coat aluminum on an insulative substrate.

The next step in the direction for a complete material was to test insulative particles coated with nickel and overcoated with nickel oxide, which we do know how to do. Relatively uniform glass spheres were used as the substrate. The original samples showed a high degree of promise. We still got failures, but the off state resistances were in the GΩ range, and the electrode failure signatures were encouraging.

For previous dominant path failures, the electrode failure signatures consisted of single pits on the electrodes, indicating failure through a single very narrow path. *For the nickel samples, the signatures looked more like clouds over a substantial area of the electrode surface.* We believe this is the result of the higher degree of morphological uniformity of the glass spheres with which we started.

Although there were some promising aspects with the nickel nickel-oxide formulation, overall results were disappointing. We then turned to silicon dioxide as the insulative phase. With glass as the substrate, coated with nickel, and overcoated with glass, consistent clamping action was obtained. There was an apparent leading edge overshoot, but we believe this is a test system artefact. Oscilloscope traces are to be found in Section V.

I-1-4 Laboratory Measurements

One of the things we really wanted to accomplish in this project was a clear demonstration of a quantum tunnelling signature. That is the measurement of a DC current versus voltage closely matching the form of the Nordheim-Fowler approximation. Instead, in *all* of the samples tested, we found an increasing current at constant voltage as a function of *time* with time constants on the order of hours. Such behavior masks any underlying quantum tunnelling signature that might be present. We have not totally given up on this matter, but it is clear the situation is more complex than originally envisaged.

Although a nuisance in terms of the quantum tunnelling signature, this drift in resistance is of no practical consequence, since the resistances are always sufficiently high and the drifts are toward higher resistances.

Section V-2 is relatively short; we refer you to it for further discussion of laboratory measurements.

I-1-5 Prototype Devices

Littelfuse, Inc. has used Integrated Sciences™ aluminum formulations and successfully incorporated them into one of their varistor devices. Testing was performed according to IEC 801-2 specifications. Definite *clamping action was observed with zero failures*. There appeared to be a leading edge overshoot with about a 2 ns width. As previously mentioned, it is believed this is mostly a test fixture artefact.

With prolonged pulsing the clamp voltage slowly increased and nearly doubled after 250 pulses. Both this effect and the question of leading edge overshoots will be thoroughly investigated in Phase II. It is to be noted that increase in clamp voltage and leading edge overshoots were not observed in previous tests performed by the Army's Harry Diamond Laboratories on a formulation similar to the one used in these Littelfuse tests. It is also to be noted that increases in clamp voltage were not observed with the glass-nickel-glass formulation previously described.

I-1-6 Conclusions

A good start has been made toward production of new varistor materials for transient voltage surge suppression. Samples were prepared and tested according to ESD specifications with promising results, and it is highly likely the same type of samples will meet EMP requirements.

A thorough analysis of pure quantum tunnelling and related phenomena is complete. At present, there appears there may be other conduction mechanisms involved. These will be investigated both theoretically and empirically in the Phase II effort.

Our computer models are not yet to a quantitative level; however, they have proved to be completely accurate at a qualitative level and have been of major benefit in directing our efforts. Further, the programs are organized in such a way that any new conduction mechanisms can be introduced as "drop in" packages without modification of the overall program structure.

An important aspect of the work performed under this contract is that we have gone from many failed formulations to ones showing definite promise. Of particular interest is that we can now coat particles with controlled amounts of coating thicknesses, a matter important to the production processes. Additionally, a formulation exhibiting consistent clamping was produced incorporating Integrated Sciences™ three phase approach.

Littelfuse, Inc. has recently voiced interest in our technical developments, the reason some of the testing was done at their facilities. It is highly likely

Integrated Sciences™ and Littelfuse will enter into an alliance agreement during Phase II, with Littelfuse supporting the Phase III effort. *Littelfuse has an exceptional materials science capability. This coupled with the talents at Spectrodynamic Systems and Integrated Sciences™ places us in a particularly strong position to positively move forward in the commercialization of this technology.*

Section I-2

RECOMMENDATIONS FOR PHASE II

I-2-1 Primary Objective

The general principles upon which quantum tunnelling varistors operate have been empirically demonstrated and their principal failure mechanism identified. *The primary objective of Phase II shall be optimization, in terms of both materials and configuration.* From a time and funding standpoint, we feel this limited goal is essential. Once the varistor materials are well categorized and performance characteristics defined, we will have a relatively clear path to commercialization in Phase III. However, we do expect that commercialization will commence during Phase II. A variety of devices can be formed from the materials we are developing. We will start with those that are easiest and work toward those that are more difficult.

In Phase I, our results, though correct, have been principally qualitative, and apparently somewhat incomplete. We now need to bring them to a complete and quantitative level, yielding reliable design information upon which sound engineering can be based. Recommendations to achieve this are given in the following subsections.

I-2-2 Quantum Analyses

So far, our emphasis has been on pure intermetallic quantum tunnelling. We still believe this is a primary mechanism in the materials under development. However, results of our computer modelling and laboratory measurements indicate there are other active conduction mechanisms. *We need to identify the other conduction mechanisms and include them in our computer programs.*

I-2-3 Electromagnetic Analyses

At our present stage of computer modelling, we are using a continuum mechanical model, where we are averaging over control volumes containing a number of particles making up the varistor material. These same programs can be used to model on the basis of individual particles. In this way, we can get a better understanding of the interparticle interfacial electrodynamics. We feel this is essential in view of the potential for multiple conduction mechanisms. *Once these mechanisms are understood, we would use this information as input back into the continuum mechanical models, in the same fashion as is currently done with the results from the analytical quantum analyses.*

So far, the electrodynamics have been modelled on a quasistatic basis, and there is no indication this is not adequate. However, at some point in time, we may need to address the complete dynamic situation.

I-2-4 Thermal Analyses

The thermal analyses are based on the continuum mechanical models. There is no apparent reason that this approach needs to be modified, and *we expect the present mode of thermal analysis, including thermal diffusion, will be used throughout the remainder of the program.*

I-2-5 Finite Element Modelling

This has been a cornerstone of our approach. Even in its apparent presently qualitative state, it has been highly useful in guiding and understanding our current efforts. Aside from modifications previously alluded to, the major improvement needed is the expansion to three dimensions.

The topology of two-space is intrinsically different from the topology of three-space. This difference has a definite impact on allowed field structures and their functional dependences. To obtain a true quantitative engineering tool for our present application, we must model in three-space. *Our goal here is to be able to reliably test varistor material concepts on the computer prior to proving them in the laboratory.*

I-2-6 Test Sample Preparation

The sample cells used in Phase I were crude. From a qualitative and feasibility standpoint they served their purpose. However, as we move to a more quantitative approach, improvements need to be made. *For giga-Herz transient responses and nano-second time frames, the sample cells will be configured as a part of a transmission line.* This reduces the parasitics of the test system to acceptable levels, as was done in the Harry Diamond tests.

For both high frequency work and measurements at DC, the containment of the varistor material must be consistent. *Sample cells meeting the consistency requirements will be designed.*

I-2-7 Laboratory Measurements

As with the sample cells, our laboratory measurement system is inadequate for reliable quantitative measurement. *For fast transient measurements, we will acquire an oscilloscope with a higher bandwith than our present one, configure the system to a transmission line environment with the sample cells as an integral part of the transmission line, and acquire a pulser with variable voltage, pulse width, and duty cycle.*

For DC measurements, we will continue to use a Keithley 6517 electrometer. *So we can monitor the time dependence of the resistance drifts previously mentioned, we will either acquire a strip chart recorder or institute a computer controlled data acquisition system.*

I-2-8 Commercialization

Although our primary objective in Phase II is the development and characterization of commercially viable varistor materials, it is not our intent to totally ignore the applications aspects. *We plan to remain in constant contact with potential users and use their guidance in determination of those characteristics needed in their applications.* Through alliances with a third party or parties, we expect that some commercialization will take place as we are pursuing the Phase II effort.

The Integrated Sciences™ goal is to remain in the research and development arena serving Government and specialty needs while spinning off high volume commercial applications to another party or parties on a deferred income basis.

Section II-1

SUMMARY AND EXTENSION OF PREVIOUS RESULTS

II-1-1 Reduction to One Dimension

Using the technology being developed here, it was often believed the devices would be those where the dominant nonlinearity was provided by quantum tunnelling. The purpose of this chapter is to more clearly delineate the consequences of pushing to the quantum tunnelling limit. However, before taking up the main thread of our discussion, there is a matter we feel is worth treating in an explicit fashion.

For a single electron, the time independent Schrödinger equation may be rendered in the form

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\boldsymbol{r})+U(\boldsymbol{r})\,\psi(\boldsymbol{r})=\varepsilon\,\psi(\boldsymbol{r})$$

where ε is the total energy of the electron and U is the potential energy. We will consider planar geometry, two pieces of metal separated by a vacuum gap across which is applied an electric field E. We are particularly interested in what happens in the vicinity of the gap and consider three regions with a potential profile as shown in Figure II-1-1.

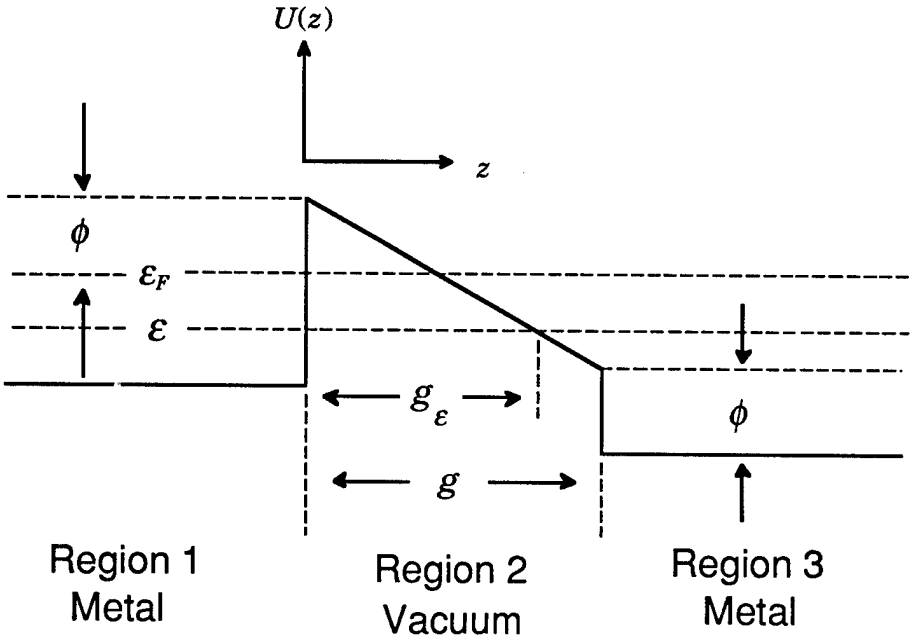


Fig. II-1-1 Potential profile for a planar interface between two pieces of metal separated by a vacuum gap g across which there is an applied electric field. The parameters ε_F and ϕ are respectively the Fermi energy and the work function. The distance g_ε is the *tunnelling gap* for an electron with energy ε. We refer to this potential as the *triangular wall barrier*.

We have defined the *tunnelling gap* here, since it will be important in subsequent developments.

If we considered only electrons incident normal to the surface of the metal, the problem would essentially reduce to a one dimensional one. However, as we shall see, that does not correspond to the situation in a real metal. This necessitates the justification of the reduction to one dimension.

We consider coordinates with the z-axis perpendicular to the metal surfaces with positive direction to the right, with x-axis perpendicularly into the page, and choose the y-axis so we have a right handed system. We take the zero of potential in Region 1. With U a function of z alone, we can separate the Schrödinger equation in the usual manner obtaining

$$\frac{1}{\psi_x}\frac{\partial^2\psi_x}{\partial x^2}+\frac{1}{\psi_y}\frac{\partial^2\psi_y}{\partial y^2}+\frac{1}{\psi_z}\frac{\partial^2\psi_z}{\partial z^2}+\frac{2m}{\hbar^2}\left[\varepsilon-U(z)\right]=0$$

With some foresight, the separated equations may be cast into the form

$$\frac{1}{\psi_x}\frac{d^2\psi_x}{dx^2}=-k_x^2,\quad \frac{1}{\psi_y}\frac{d^2\psi_y}{dy^2}=-k_y^2,\quad \frac{1}{\psi_z}\frac{d^2\psi_z}{dz^2}+\frac{2m}{\hbar^2}\left[\varepsilon-U(z)\right]=\frac{2m}{\hbar^2}\varepsilon_\rho$$

The form of the x and y solutions are the same in all three regions, complex exponentials. The boundary conditions require the wave functions and their normal derivatives be continuous at the boundaries. To satisfy these conditions, k_x and k_y must remain constant throughout all three regions. In Region 1, the total energy is given by

$$\varepsilon=\frac{\hbar^2}{2m}\left(k_x^2+k_y^2+k_1^2\right)$$

and by the conservation of energy must also be the same in all three regions. We have subscripted the z-component wave vector with its region since this is the only component that can change from Region 1 to Region 3. The z-equation may now be put in the form

$$\frac{d^2\psi_z}{dz^2}+\frac{2m}{\hbar^2}\left[\varepsilon-\varepsilon_\rho-U(z)\right]\psi_z=\frac{d^2\psi_z}{dz^2}+\frac{2m}{\hbar^2}\left[\frac{\hbar^2}{2m}k_1^2-U(z)\right]\psi_z=0$$

In polar coordinates with z as the polar axis, the wavevector is given by

$$\boldsymbol{k}=k\left(\sin\theta\cos\phi\ \hat{\boldsymbol{x}}+\sin\theta\sin\phi\ \hat{\boldsymbol{y}}+\cos\theta\ \hat{\boldsymbol{z}}\right)$$

$$\text{so}\quad k_1=k\cos\theta\quad\text{and}\quad\frac{\hbar^2k_1^2}{2m}=\varepsilon\cos^2\theta$$

The z-component Schrödinger equation then becomes

$$\frac{d^2\psi_z}{dz^2}+\frac{2m}{\hbar^2}\left[\varepsilon\cos^2\theta-U(z)\right]\psi_z=0$$

which is the appropriate equation for a particle striking the surface with an incidence angle of θ. The first term in the energy bracket is just that kinetic energy associated with the z-component of motion. One might certainly guess this intuitively; however, energy is a scalar and the wave vector is a vector. It is not immediately obvious that the energy should separate in this manner. Directional effects are important in this problem, so we felt it worthwhile to explicitly derive this result. From what is said here, in what follows, we need only be concerned with the z-component of motion. For a single electron, we have reduced the quantum mechanical problem to one dimension.

II-1-2 The Potential Step

In this section we will address an extremely simple problem. The reason for doing so is twofold. First the procedure here is archetypal with respect to those used in more complex situations. Second, and most important, it illustrates a quantum effect we will later encounter. We wish to show it in this simple context, devoid of computational complications, so there is no doubt of its validity.

From a source in Region 1, consider electrons travelling from Region 1 to Region 2 subject to a potential profile such as shown in Figure II-1-2.

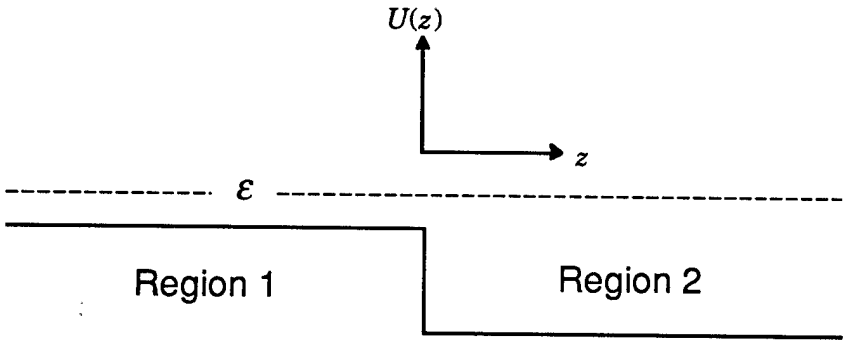


Fig. II-1-2 Step potential profile for an electron. The zero of potential is taken in Region 1. The total energy is ε, and the potential is negative in Region 2.

Classically the electron would pass from Region 1 to Region 2 simply gaining the kinetic energy imparted by the delta function force at the boundary. Quantum mechanically, the situation is rather bizarre.

In both regions, the solutions to the Schrödinger equation can be represented in terms of complex exponentials, say

$$\psi_1 = e^{ik_1 x} + Ae^{-ik_1 x}, \quad \psi_2 = Be^{ik_2 x} + Ce^{-ik_2 x}, \quad k_1 = \frac{\sqrt{2m\varepsilon}}{\hbar}, \quad k_2 = \frac{\sqrt{2m(\varepsilon + eV)}}{\hbar}$$

where V is the voltage drop across the boundary and we have arbitrarily selected the source probability current density by setting the first coefficient in ψ_1 to unity. The second coefficient in ψ_2 is zero since we assume no source in Region 2. The wave functions and their first derivatives must be continuous at the boundary between Regions 1 and 2. Taking the boundary at the origin of coordinates leads to the boundary condition matrix

$$\begin{bmatrix} k_1 & k_2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} k_1 \\ -1 \end{bmatrix} \quad \text{with solution} \quad A = -\frac{k_2 - k_1}{k_2 + k_1}, \quad B = \frac{2k_1}{k_2 + k_1}$$

The general expression for the probability current density is

$$J_p = -\frac{i\hbar}{2m}\left(\psi^* \nabla\psi - \psi \nabla\psi^*\right)$$

and the incident, reflected and transmitted probability current densities are

$$J_i = \frac{\hbar k_1}{m}, \quad J_\rho = \frac{\hbar k_1}{m}|A|^2, \quad J_\tau = \frac{\hbar k_1}{m}|B|^2$$

The reflection coefficient is the ratio of the reflected probability current density to the incident probability current density and similarly for the transmission coefficient. Respectively these are

$$\rho = \left(\frac{k_2 - k_1}{k_2 + k_1}\right)^2 \quad \text{and} \quad \tau = \frac{4k_1 k_2}{(k_2 + k_1)^2}$$

What this says is, a negative going potential drop can, with finite probability, reflect an electron. That's bizarre enough. However, one would expect that increasing the potential would decrease the reflection, but no, increasing the potential increases the reflection and correspondingly decreases the transmission.[1] For an infinite drop, the transmission is zero. Do we really need to be concerned about this effect?

Electrons with which we will subsequently be concerned will have energies on the the order of *10 eV*. Let us first ask what potential drop is required to reflect most of the electrons, say ninety percent of them. The ratio of the wave vector magnitudes in terms of the transmission coefficient may be expressed as

[1]The sum of the reflection and transmission coefficients is one.

$$\frac{k_2}{k_1}=\left(\frac{2}{\tau}-1\right)\pm\sqrt{\left(\frac{2}{\tau}-1\right)^2-1} \quad \text{with corresponding voltage} \quad V=\frac{\varepsilon}{e}\left[\left(\frac{k_2}{k_1}\right)^2-1\right]$$

For this case, the ratio of the wave vector magnitudes is *37.97* with the corresponding voltage of *48 SV*. After falling through this potential, the electron has a kinetic energy of about 2.3×10^{-8} ergs. Calculating the velocity classically, one finds $7.1 \times 10^{9}\ cm/s$, a relativistic velocity. Calculating the velocity relativistically through the equation

$$v=c\sqrt{1-\left(\frac{K}{m_o c^2}+1\right)^{-2}}$$

where K is kinetic energy, one finds 6.97×10^{9} cm/s. Thus for potentials of this magnitude, we are out of the range where the nonrelativistic Schrödinger equation applies.

However, we do need to ask if any sensible effect can fall within the domain of the Schrödinger theory. To that end, let us determine the potential drop required for a ten percent reflection. For this case, the ratio of the wave vector magnitudes is *1.925* with the corresponding voltage of 9.0×10^{-2} *SV*. After falling through this potential, the electron has a kinetic energy of about 4.3×10^{-11} ergs. Calculating the velocity classically, one finds $3.1 \times 10^{8}\ cm/s$, two orders of magnitude below the velocity of light. In fact, this is the same order of velocity for the conduction electrons in a metal. The transmission and reflection coefficients for a *10 eV* electron are shown in Figure II-1-3.

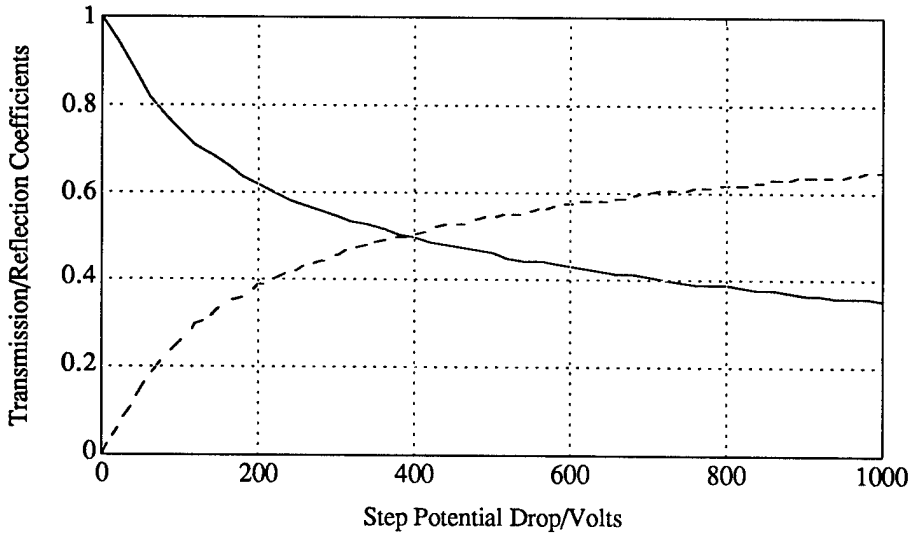


Fig. II-1-3 Linear plot of the transmission(solid)/reflection(dashed) coefficients. Calculated for a *10 eV* electron. At 1000 V, the electron velocity is about two tenths the velocity of light.

There is a substantial range where the reflection effects will be seen within the validity of the nonrelativistic Schrödinger equation. Additionally, we make note of the following. We have couched this discussion in terms of the most drastic potential step. The function and *all* of its derivatives are discontinuous. For a smoother potential drop, the effects will be smaller. At a later point we will see an example.

The main points of this whole discussion are these. In what follows, we will encounter these reflection effects, especially when we use extremely high fields to confirm probability conservation in our equations. This takes us into the relativistic domain, but what we are doing in that circumstance is checking the mathematics, not the physics. Probability conservation is a purely mathematical result following from the Schrödinger equation and the divergence theorem regardless of any physical content. In the range of practical fields, we will not be in the relativistic domain. It is expected the reflection effects will be recognizable but not necessarily of great significance. We shall see; however, forewarned is forearmed.[2]

II-1-3 The Planar Free Electron Transmission Coefficient

Considering the situation depicted in Figure II-1-1 and an electron with an incidence angle of θ with respect to the plane surface between Regions 1 and 2, the z-component Schrödinger equation takes the form

$$\frac{d^2\psi}{dz^2}+\frac{2m}{\hbar^2}\left[\varepsilon_z-\left(\varepsilon_F+\phi-eEz\right)\right]\psi=0 \qquad \text{where} \qquad \varepsilon_z=\varepsilon\cos^2\theta$$

With a change of variable given by

$$\xi=\frac{2m\zeta^2}{\hbar^2}\left[\varepsilon_z-\left(\varepsilon_F+\phi-eEz\right)\right] \qquad \text{and} \qquad \zeta=\left(\frac{\hbar^2}{2meE}\right)^{1/3}$$

the Schrödinger equation is reduced to

$$\frac{d^2\psi}{d\xi^2}+\xi\psi=0$$

which is the Airy equation except we have a plus where the Airy equation has a minus, thus if the solutions of the Airy equation are $Ai(\xi)$ and $Bi(\xi)$, our solutions will be

$$ai(\xi)=Ai(-\xi) \qquad \text{and} \qquad bi(\xi)=Bi(-\xi)$$

---

[2]When we first saw these effects in our results, we thought there must be something wrong with our calculations. The above discussion clarifies the matter. What we are seeing is real.

Satisfaction of the boundary conditions between the three regions leads to the matrix equation

$$\begin{bmatrix} 1 & -ai[\xi(0)] & -bi[\xi(0)] & 0 \\ ik_1 & \frac{1}{\zeta} ai'\xi[\xi(0)] & \frac{1}{\zeta} bi'[\xi(0)] & 0 \\ 0 & ai[\xi(g)] & bi[\xi(g)] & -e^{ik_3 g} \\ 0 & \frac{1}{\zeta} ai'[\xi(g)] & \frac{1}{\zeta} bi'[\xi(g)] & -ik_3 e^{ik_3 g} \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} -1 \\ ik_1 \\ 0 \\ 0 \end{bmatrix}$$

where

$$k_1 = \frac{\sqrt{T_1}}{\hbar} \quad \text{and} \quad k_3 = \frac{\sqrt{T_3}}{\hbar}$$

which gives

$$k_1 = \frac{\sqrt{2m\varepsilon\cos^2\theta}}{\hbar} \quad \text{and} \quad k_3 = \frac{\sqrt{2m(\varepsilon\cos^2\theta + eEg)}}{\hbar}$$

The probability current densities in Regions 1 and 3 are given by

$$J_{p1} = \frac{\hbar k_1}{m} \qquad \text{and} \qquad J_{p3} = \frac{\hbar k_3}{m}|D|^2$$

This leads to a transmission coefficient given by

$$T(\varepsilon,\theta,E) = \frac{k_3}{k_1}\left|D(\varepsilon\cos^2\theta, E)\right|^2$$

where in this last expression we have explicitly shown the independent variables. *This is the z-component transmission coefficient for an electron of total energy ε striking the surface with an incidence angle of θ in the presence of an electric field E.*

II-1-4 The Three Tunnelling Regimes

We are now in a position to qualitatively describe tunnelling behavior in varistor applications. We will recognize three regimes we designate as *cutoff*, *tunnelling*, and *reflection*. They are best described in terms of the potential functions of Figures II-1-4 to II-1-6 on the following pages.

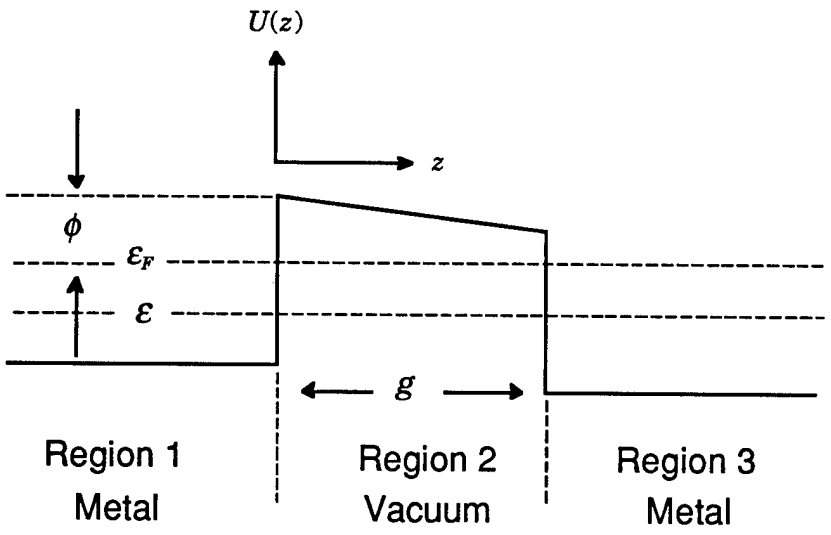


Fig. II-1-4 The Cutoff Regime. The tunneling gap for all the electrons is the full distance between the metal surfaces. The corresponding transmission coefficient is so small that negligible current flows. Tunnelling is essentially cut off. Note however that some finite current will always flow since the potential in Region 3 is everywhere less than in Region 1.

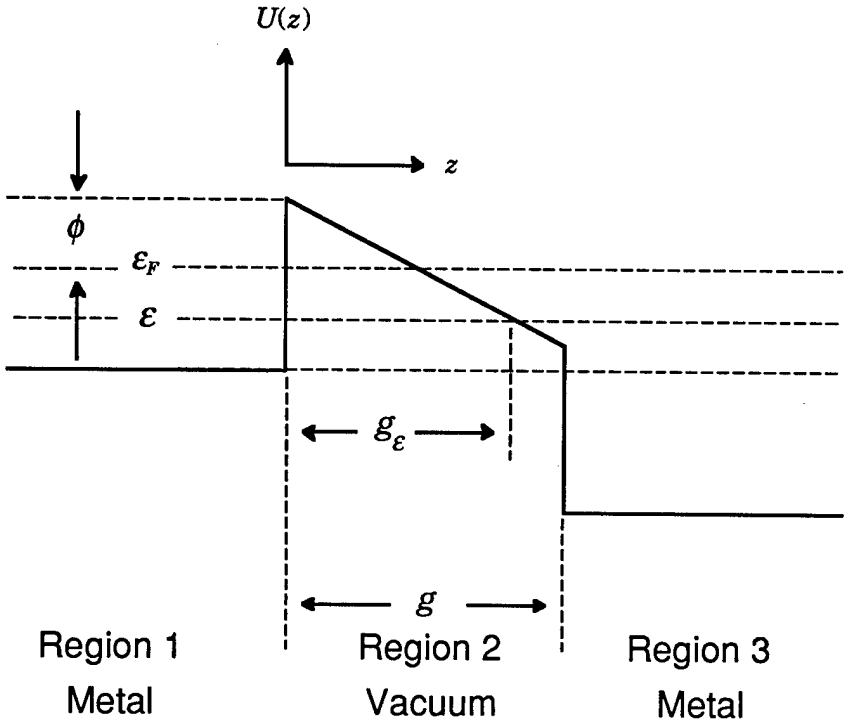


Fig. II-1-5 The Tunnelling Regime. The tunnelling gap has been reduced for some of the electrons. Under these conditions, substantial currents can flow. However in this regime the tunnelling gap is not reduced for all of the electrons; the electric potential does not cross zero, the potential of Region 1.

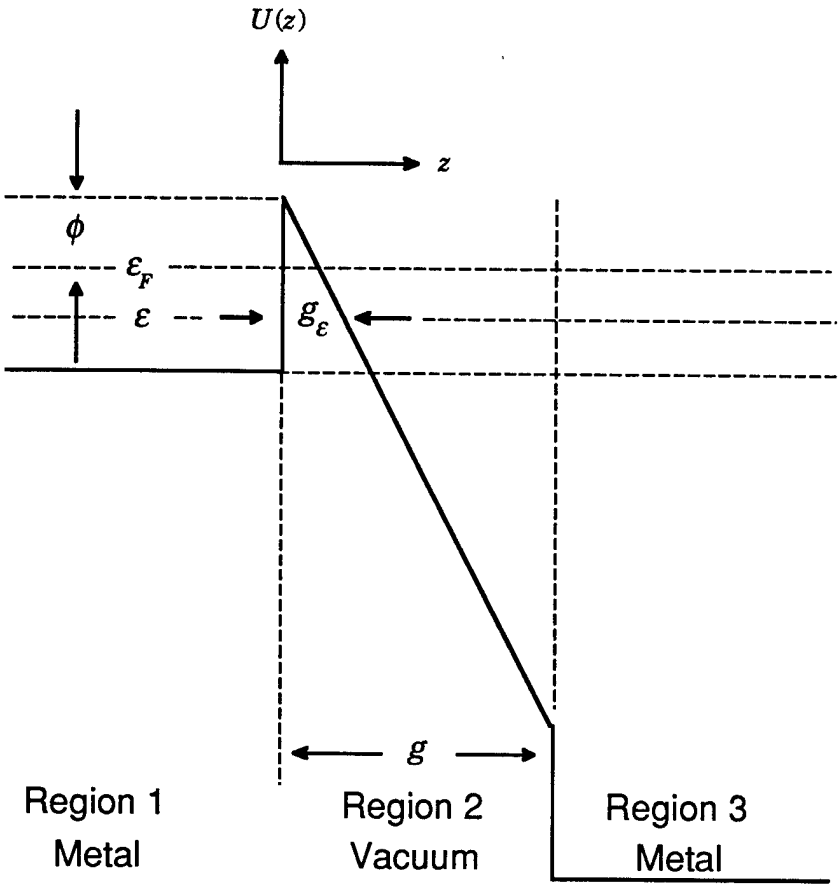


Fig. II-1-6 The Reflection Regime. The electric potential has now crossed zero. The tunnelling gap is reduced for all electrons. Decreasing the electric potential even further, the potential profile as seen by the electrons in Region 1 approaches that of the step potential with the initial peak becoming less and less significant. We would expect, and we shall find, the behavior becomes that of the step profile with its attendant reflection characteristics. This is the reason we refer to this as the reflection regime.

Summarizing the definitions, we are in the *cutoff regime* when the electric potential is always above the Fermi energy, in the *tunnelling regime* when the electric potential crosses the Fermi energy but does not cross zero, and in the *reflection regime* when the electric potential crosses zero.

II-1-5 Transmission Coefficient Models

In classical mechanics, both the free particle and the particle in a constant force field are trivial. In quantum mechanics, the situation is distinctly different. For the free particle we first have uncertainty, but then too, we have the nasty business of wave packet spreading. For the particle in a constant force field we have both of these features, and as we have seen, solution of the problem involves Bessel functions. At the beginning of our efforts, so as to not get enmired in the mathematics, we elected to look at a

number of simple models. In terms of decreasing simplicity, they were: the symmetric wall barrier, the rectangular wall barrier, and finally the triangular wall barrier. Using an intermetallic gap of *100 Å* and the parameters for copper, we will discuss the three models in turn.

Symmetric Wall Barrier

Our initial thoughts were that quantum tunnelling would be dominated by the width of the tunnelling gap, so we elected to investigate a potential function as shown in Figure II-1-7 where the tunnelling gap is adjusted to be equal to to the tunnelling gap in an applied electric field.

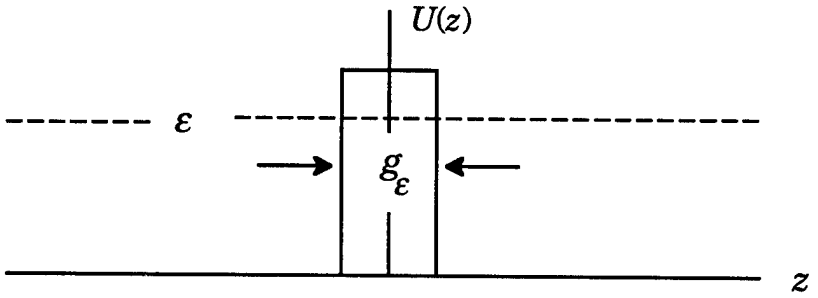


Fig. II-1-7 The symmetric wall barrier potential.

It is a simple model with an easy solution, often given as a sophomore or junior level physics exercise. Results for the transmission coefficient as a function of gap field are shown in Figure II-1-8.

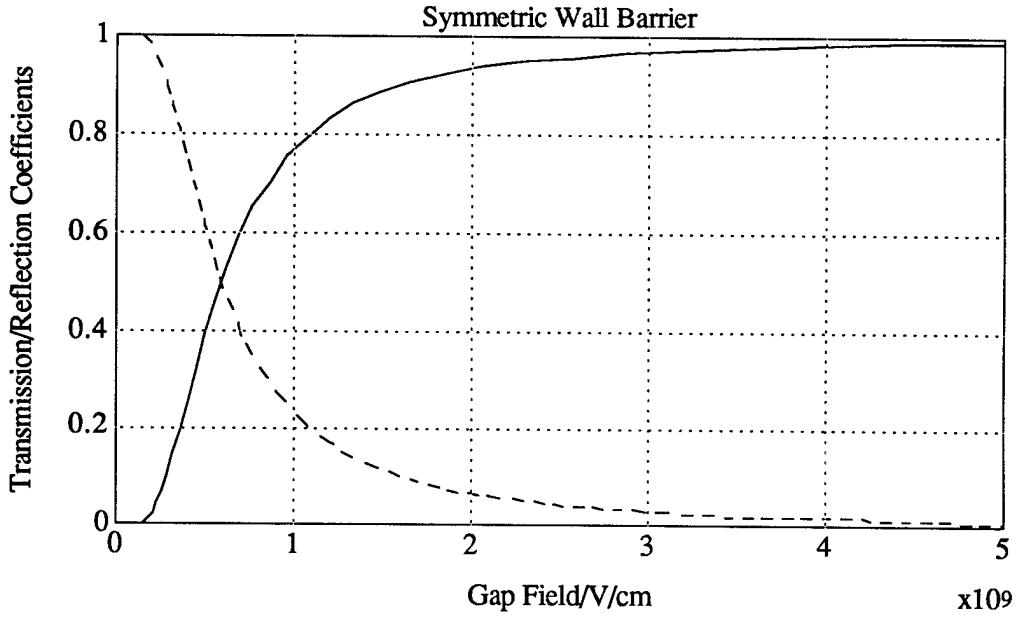


Fig. II-1-8 Linear plot of the transmission(solid)/reflection(dashed) coefficients for the symmetric wall barrier using the parameters for copper and an intermetallic gap of *100 Å*.

No real surprises here. As the field increases, the transmission coefficient monotonically increases to unity. From the linear plot, the degree of nonlinearity can not be appreciated. In Figure II-1-9 we show the same data plotted log-log.

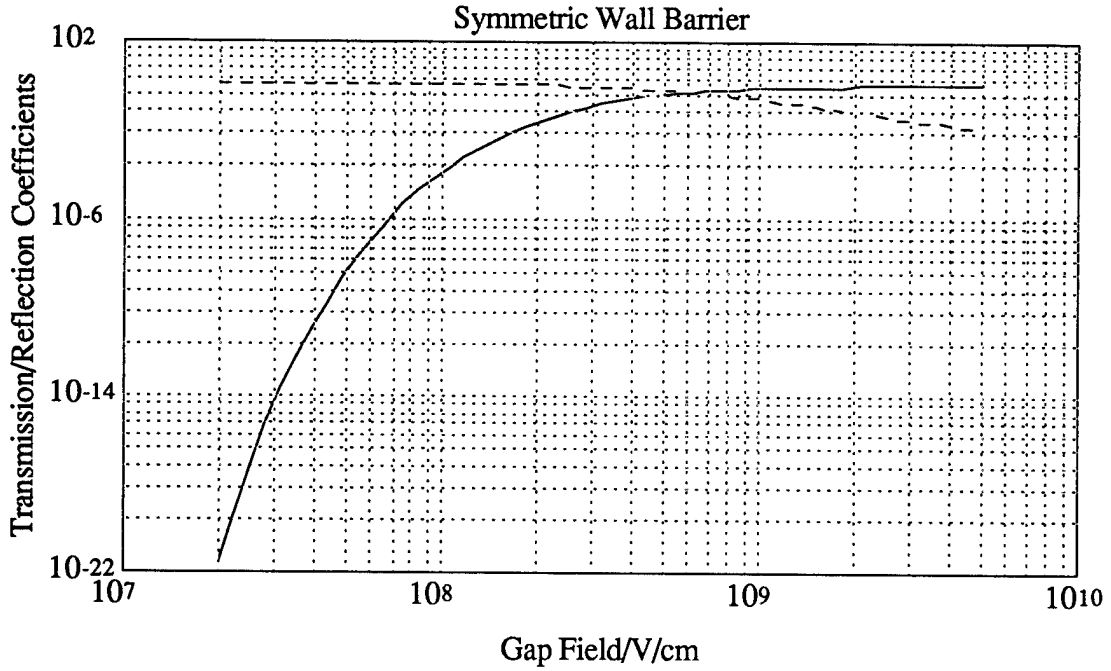


Fig. II-1-9 Log-log plot of the transmission(solid)/reflection(dashed) coefficients for the symmetric wall barrier using the parameters for copper and an intermetallic gap of *100 Å*.

Rectangular Wall Barrier

Our next approximation was the rectangular wall barrier shown in Figure II-1-10 with transmission and reflection coefficients shown in Figures II-1-11 and II-1-12.

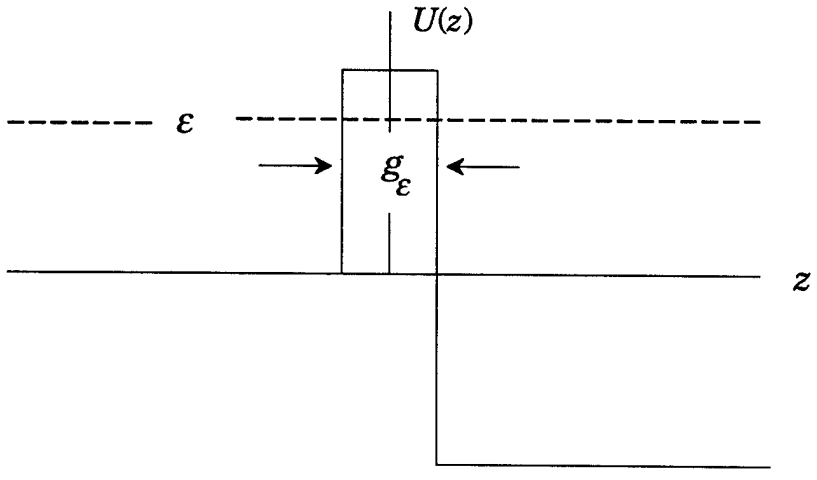


Fig. II-1-10 The rectangular wall barrier potential.

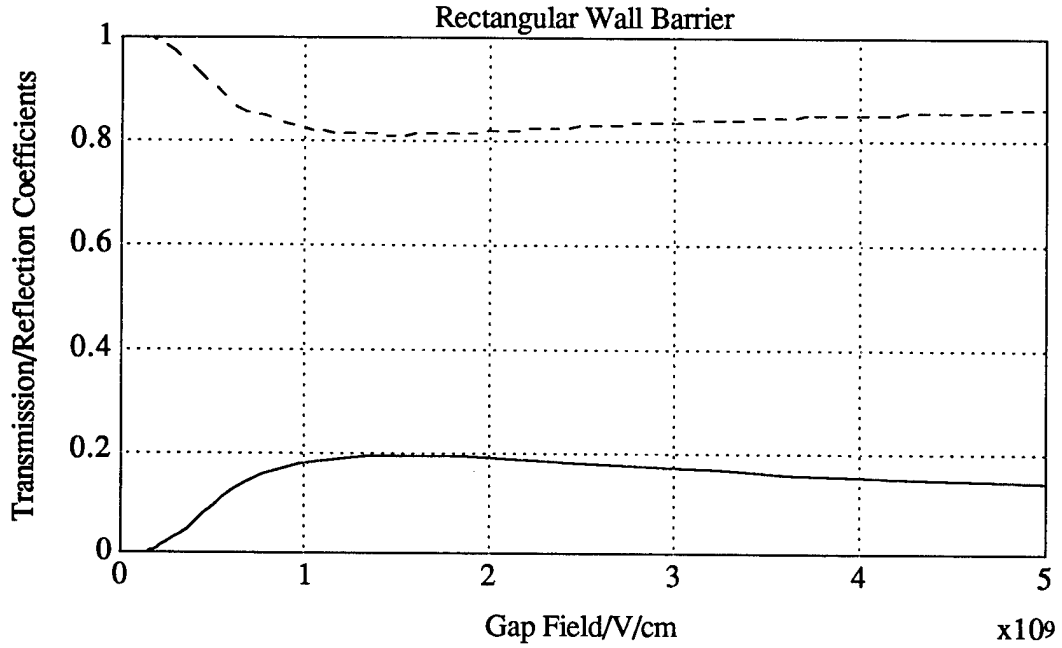


Fig. II-1-11 Linear plot of the transmission(solid)/reflection(dashed) coefficients for the rectangular wall barrier using the parameters for copper and an intermetallic gap of *100 Å*.

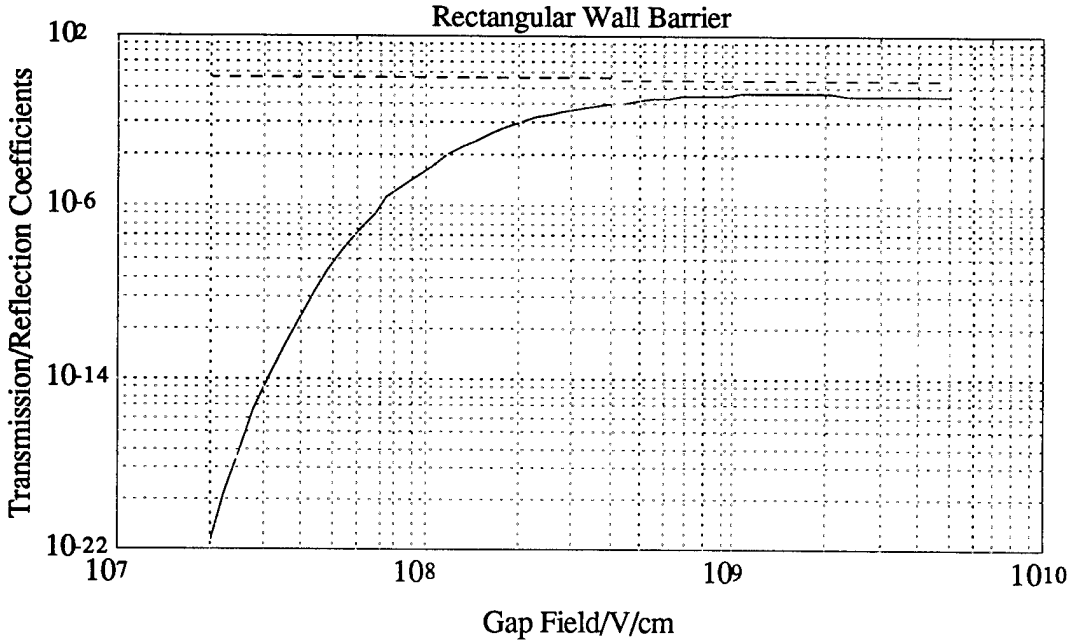


Fig. II-1-12 Log-log plot of the transmission(solid)/reflection(dashed) coefficients for the rectangular wall barrier using the parameters for copper and an intermetallic gap of *100 Å*.

It was relatively easy to modify the program for the symmetric wall barrier to accommodate this one and we still avoided the Bessel functions. A solution in terms of exponentials was obtained. Here we have our first evidence of the reflection effects. For low field values, the symmetric and rectangular results are in fairly good agreement. We were expecting an increase in the computed transmission coefficient in the high field regions by including the added potential drop. The opposite occurs; for sufficiently high fields, the transmission decreases as the field increases, and the maximum is considerably below unity.

Triangular Wall Barrier

The potential for the triangular wall barrier was previously shown as Figure II-1-1. The corresponding transmission and reflection coefficients are those of Figures II-1-13 and II-1-14.

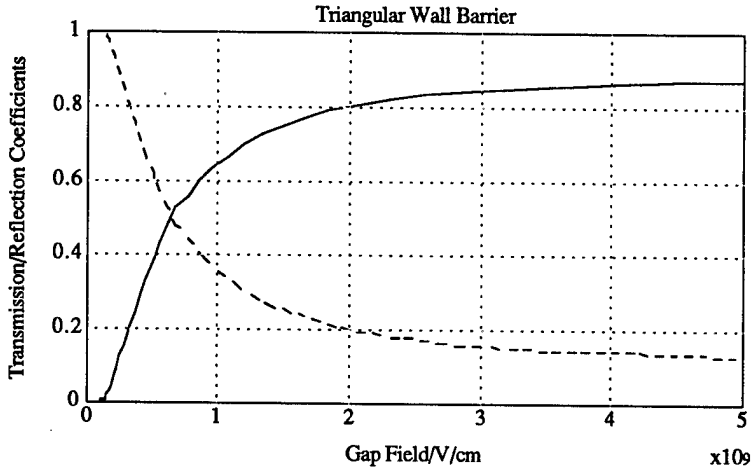


Fig. II-1-13 Linear plot of the transmission(solid)/reflection(dashed) coefficients for the triangular wall barrier using the parameters for copper and an intermetallic gap of *100 Å*.

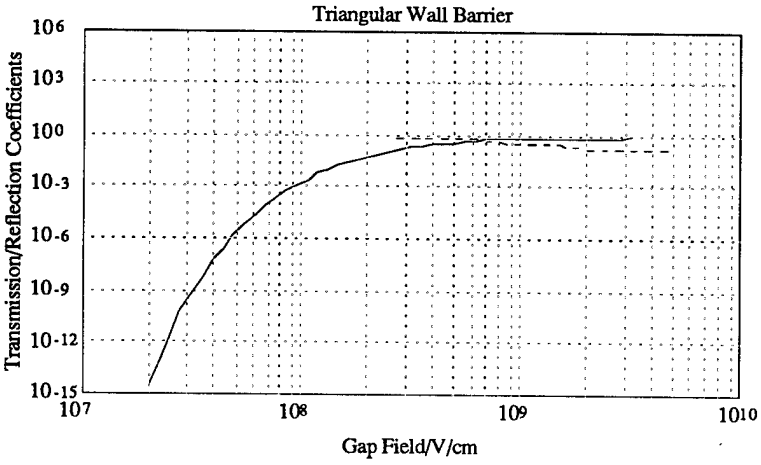


Fig. II-1-14 Log-log plot of the transmission(solid)/reflection(dashed) coefficients for the triangular wall barrier using the parameters for copper and an intermetallic gap of *100 Å*.

The first thing we observe is that in comparison to the rectangular wall barrier, the transmission has increased and the degree of nonlinearity has decreased, the effect of getting rid of one discontinuity. The potential function is now continuous although all its derivatives are still discontinuous. Removal of the functional discontinuity has also lessened the reflection effects. Where are they? They are at higher fields as shown in Figure II-1-15.

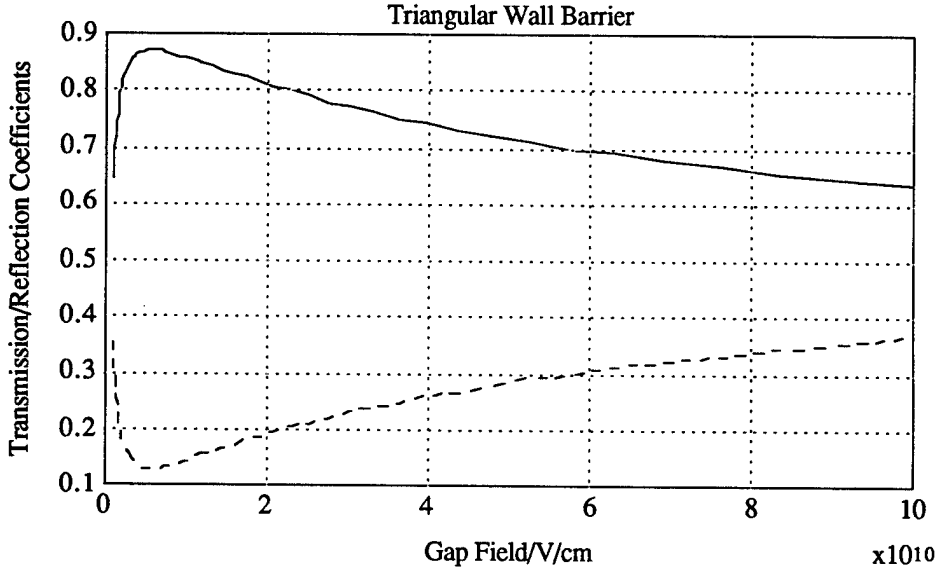


Fig. II-1-15 Linear plot of the transmission(solid)/reflection(dashed) coefficients for the triangular wall barrier using the parameters for copper and an intermetallic gap of *100 Å*. The solid curve is the transmission coefficient. The lowest field shown is $10^9\ V/cm$.

It should be noted that in all the plots displayed, the transmission and reflection coefficients are independently calculated. Their sum being unity serves as a check on the correctness of the calculations but, of course says nothing about the validity of the model. Of the three models we have just described, the triangular is the most realistic, but because of the discontinuities still remaining in the derivatives of the potential function, we expect predicted transmission coefficients will generally be underestimated. However, we also expect the general trends will be adequately represented by this model and will serve as a guide in our further efforts. With that said, it is now time to turn our attention to the electrical current densities.

II-1-6 The Fundamental Tunnelling Integral, Quasiclassical Derivation

For the present, we will continue to use planar geometry for our analysis. It does not correspond to the geometry of our materials, but will give insight as to the important parameters involved and their interrelations. A number of approximations were carried out in a previous effort. We now wish to refine them and their associated numerical results and develop a

design procedure for quantum tunnelling devices. The first step is to give an accurate derivation of the fundamental tunnelling equation. In the present section, we will give a quasiclassical derivation. The advantage of the quasiclassical approach is that it gives a more pictorial view.

The tunnelling current density is dependent upon the tunnelling transmission coefficient, the number of electrons per unit time striking the tunnelling barrier, and their energies and directions. We will use the free electron model of electrons in a metal. In that case, the energy and speed of the electrons are related through the classical expression for kinetic energy. We will start with the directional characteristics and consider a unit volume in the physical space of the electrons. We next translate all the velocity vectors of N electrons to a common origin, inscribe them in a unit radius sphere, and extend the vectors to the surface of the sphere. The directions of the velocities are uniformly distributed so the intersections of the vector extensions with the unit sphere have a surface density of $N/4\pi$. We use spherical coordinates to define direction. The number of electrons in a $\theta\phi$-cone is then given by

$$dN_{\theta\phi} = \frac{N}{4\pi} \sin\theta \, d\theta \, d\phi$$

Now considering a parallepiped of electrons with base dA on the tunnelling barrier and axis making an angle θ with the z-axis, if $v \times dt$ is the length of the axis, the volume of the parallepiped is $dA \times v \times dt$. All the electrons contained in the volume will strike the surface in time dt. The number of $\theta\phi v$-electrons striking the surface per unit time and per unit area will be

$$\left(\frac{dN}{dA\,dt}\right)_{\theta\phi v} = \frac{dN_v}{4\pi} v \cos\theta \sin\theta \, d\theta \, d\phi$$

where dN_v is the number of electrons with velocities between v and $v + dv$. In terms of the velocity density n, we can write

$$dN_v = n(v)dv = \mathcal{D}(\varepsilon) f(\varepsilon) d\varepsilon$$

where $f(\varepsilon)$ is the Fermi-Dirac distribution function and $\mathcal{D}$ is the energy density of states which for the free electron model is given by [13]

$$\mathcal{D}(\varepsilon) = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \sqrt{\varepsilon} \quad \text{and with} \quad v = \sqrt{\frac{2\varepsilon}{m}}$$

For the present, we will take the low temperature approximation where $f(\varepsilon)$ is unity up to the Fermi energy and zero beyond.[3] We then have

[3]Even at room temperature, this is not a bad approximation. We will explore this further in Section II-2.

$$\left(\frac{dN}{dA\,dt}\right)_{\theta\phi\varepsilon} = \frac{m}{2\pi^3\hbar^3}\,\varepsilon\cos\theta\sin\theta\,d\theta\,d\phi$$

Multiplying this by the electronic charge and the transmission coefficient gives the differential of current density. Completing the integration over ϕ, we have the *fundamental tunnelling integral*

$$J(E) = \frac{em}{\pi^2\hbar^3}\int_0^{\varepsilon_F} d\varepsilon \int_0^{\pi/2} d\theta\;\varepsilon\,T\left(\varepsilon\cos^2\theta, E\right)\cos\theta\sin\theta$$

II-1-7 The Fundamental Tunnelling Integral, Quantum Derivation
In this section, all variables will be those associated with a purely quantum mechanical approach. Let us return to our expression for the tunnelling probability current density

$$J_{p3} = \frac{\hbar k_3}{m}|D|^2$$

In Region 3, the kinetic energy is given by

$$\frac{\hbar^2}{2m}\left(k_x^2 + k_y^2 + k_3^2\right) = \varepsilon + eEg = \frac{\hbar^2}{2m}\left(k_x^2 + k_y^2 + k_1^2\right) + eEg$$

$$\frac{\hbar^2 k_3^2}{2m} = \frac{\hbar^2 k_1^2}{2m} + eEg = \varepsilon\cos^2\theta + eEg \quad \text{or} \quad k_3 = \frac{\sqrt{2m\left(\varepsilon\cos^2\theta + eEg\right)}}{\hbar}$$

Substituting for k_3, multiplying by the electronic charge, and showing the independent variables explicitly we have for the z-component electric current density in Region 3

$$J_{3z}(\varepsilon,\theta,E) = \frac{e}{m}\sqrt{2m\left(\varepsilon\cos^2\theta + eEg\right)}\;\left|D(\varepsilon,\theta,E)\right|^2$$

that is, the current density for a single electron with an incidence angle θ.

Let us now consider dN_ε electrons in a unit volume of the metal with energies in a $d\varepsilon$ interval. The directions of their wave vectors will be randomly distributed. In a fashion similar to what was done in the last section, we can visualize this situation by translating all the wave vectors to a common origin and inscribing them within a sphere of unit radius. The vector ray intersections with the surface of the sphere will be uniformly distributed with a surface density equal to $1/4\pi$. The number of vectors within a $\theta\phi$ solid angle will be

$$dN_{\theta\phi\varepsilon} = \frac{dN_\varepsilon}{4\pi} \sin\theta \, d\theta \, d\phi$$

Integrating this with respect to ϕ gives the number of electrons in the $d\varepsilon$ interval with incidence angle θ.

$$dN_{\theta\varepsilon} = \frac{dN_\varepsilon}{2} \sin\theta \, d\theta$$

If we now multiply this number by the current density for a single electron, we will have the differential current density for all the $\theta\varepsilon$ electrons.

$$d^2J(\varepsilon,\theta,E) = \frac{e}{m}\frac{dN_\varepsilon}{2} \sin\theta \, d\theta \sqrt{2m\left(\varepsilon\cos^2\theta + eEg\right)} \left|D(\varepsilon,\theta,E)\right|^2$$

For dN_ε we take the free electron energy density of states previously given by

$$\mathcal{D}(\varepsilon) = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{\varepsilon}$$

We then obtain

$$d^2J(\varepsilon,\theta,E) = \frac{em}{\pi^2\hbar^3} \sqrt{\varepsilon\left(\varepsilon\cos^2\theta + eEg\right)} \sin\theta \left|D(\varepsilon,\theta,E)\right|^2 d\theta \, d\varepsilon$$

Integrating this with respect to angle and energy gives the total current density in Region 3.

$$J(E) = \frac{em}{\pi^2\hbar^3} \int_0^{\varepsilon_F} d\varepsilon \int_0^{\pi/2} d\theta \sin\theta \sqrt{\varepsilon\left(\varepsilon\cos^2\theta + eEg\right)} \left|D\left(\varepsilon\cos^2\theta, E\right)\right|^2$$

This is *the fundamental tunnelling integral.* The constant in front of the integral has the value

$$K_f = 3.174 \times 10^{10} \frac{A}{cm^2}\frac{1}{(eV)^2}$$

The interested reader may readily verify that this result is identical to the one obtained from the quasiclassical derivation, a cross check on both derivations.

As a further check we have evaluated these integrals at field strengths where the transmitted and reflected currents are of the same order of magnitude and confirmed that charge conservation is maintained in our formulation. In the form of the last section, if we set the transmission coefficient to unity, the transmitted current will be equal to the incident current. In this case the integrations are readily performed giving a value for the incident current density of

$$J_i = \frac{em\varepsilon_F^2}{4\pi^2\hbar^3}$$

For copper, the incident current density is $3.989 \times 10^{11} A/cm^2$. It is the magnitude of this current density that allows significant currents to flow even in the face of small transmission coefficients. This magnitude results from the fact that there are a large number of electrons, and those at the Fermi energy are moving at about one hundredth the velocity of light.

II-1-8 Evaluation of the Tunnelling Integral

Since we wish to cover a rather large range of various parameters and to include all significant physical effects, our approach will be numerical. It is then desirable to know the general character of the integrand. Unscaled mesh plots are shown in Figures 16 and 17 with quantitative data in Figures 18 and 19.

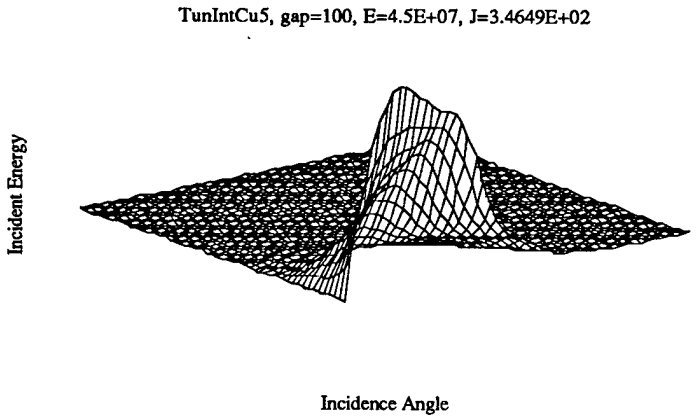


Fig. II-1-16 Unscaled mesh plot of the tunnelling integral integrand. The origin of coordinates is at the upper left corner with energy and angle increasing from that point. Parameters used are those for copper.

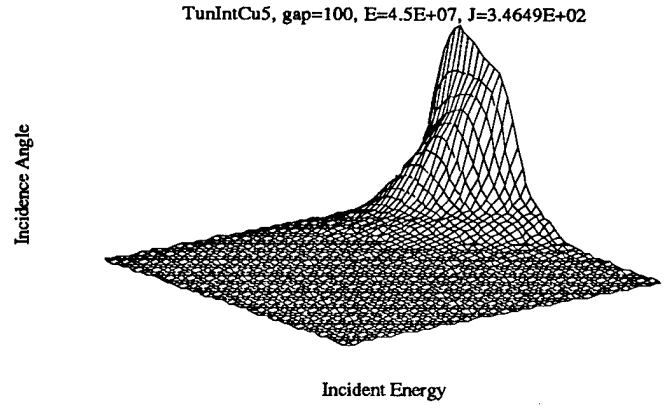


Fig. II-1-17 Unscaled mesh plot of the tunnelling integral integrand. The origin of coordinates is at the upper left corner with energy and angle increasing from that point. Parameters used are those for copper.

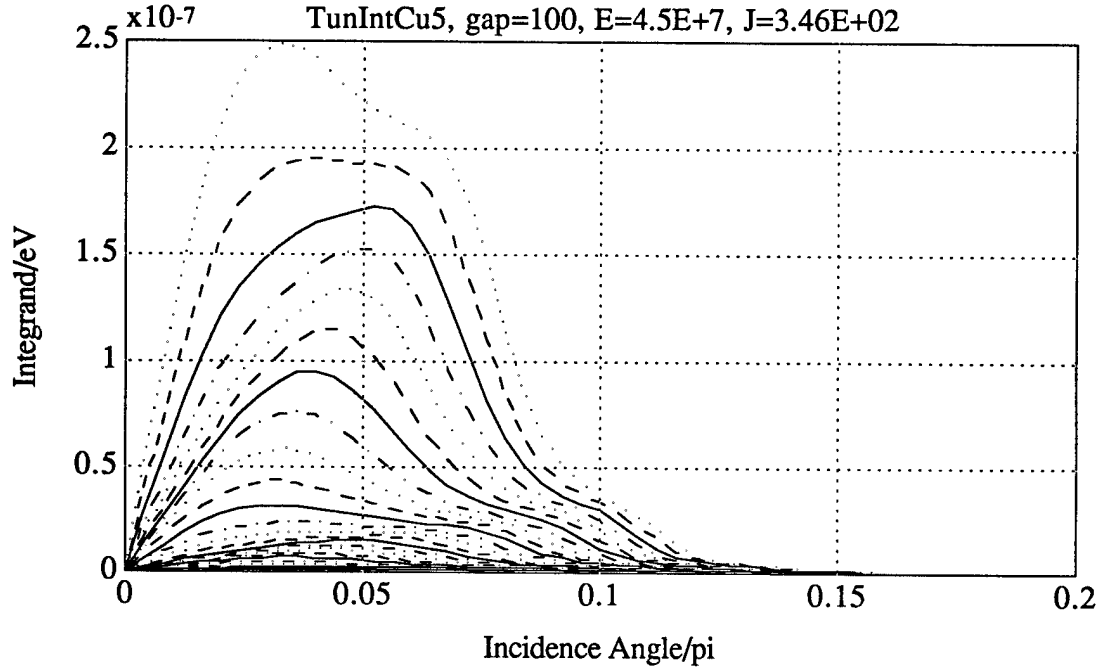


Fig. II-1-18 Plots of the integrand as a function of incidence angle for various energies. The largest values are for the Fermi energy. Parameters used are those for copper.

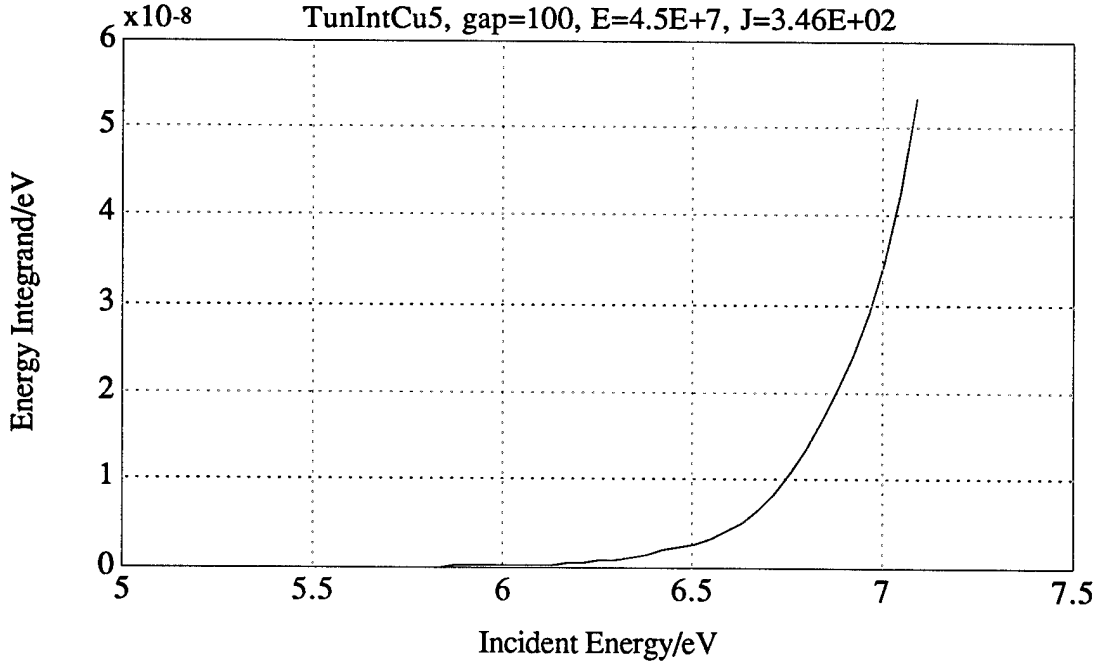


Fig. II-1-19 Integrand remaining after the angle integral is complete.

We observe a fair amount of structure for the total integrand, but the curves are relatively smooth; thus, we did not institute a sophisticated integration routine and simply used the trapezoidal rule over a 51×51 grid. A number of checks were performed to test the accuracy of the calculations. As previously described, we checked the transmission/reflection balance. To check the accuracy of the Airy functions, we had recourse to an identity [15]

$$ai(\xi)\frac{d}{d\xi}bi(\xi)-bi(\xi)\frac{d}{d\xi}ai(\xi)=-\frac{1}{\pi}$$

The identity was always satisfied to within five place accuracy throughout the domain of the variables we are using. For the complete tunnelling integral, we checked conservation of charge and found the ratio of the sum of the transmitted and reflected current densities divided by the incident current density to be 1.0005. This represents sufficient accuracy for our present purposes, general trends and semiquantitative results. In the final analysis, it is empirical data that will be used.

II-1-9 The Nordheim-Fowler Approximation

An approximating function for the current density is afforded in an early paper [12] of Nordheim and Fowler (NF). They find an expression in the form

$$I=\frac{\varepsilon}{2\pi h}\frac{\mu^{1/2}}{(\chi+\mu)\chi^{1/2}}F^2e^{-4\kappa\chi^{3/2}/3F} \quad \text{with} \quad \kappa=\frac{8\pi^2m}{h^2}=\frac{2m}{\hbar^2}$$

ε is the electronic charge, μ is the Fermi energy and χ is the work function. Evaluating the aggregates of fundamental constants they write

$$I=6.2\times10^{-6}\frac{\mu^{1/2}}{(\chi+\mu)\chi^{1/2}}F^2e^{-2.1\times10^{7}\chi^{3/2}/F}$$

or somewhat more legibly

$$I=6.2\times10^{-6}\frac{\mu^{1/2}}{(\chi+\mu)\chi^{1/2}}F^2exp\left(-2.1\times10^{?}\chi^{3/2}/F\right)$$

where the current density is expressed in terms of A/cm^2 and the Fermi energy and work function are in eV. Unfortunately, in the copy of the paper we have, we can not read what is supposed to take the place of the question mark in the exponential above. In addition, NF tends to switch the meaning of F between force and electric field.

If we assume in the first of the above equations that F is force, then the equation is dimensionally correct. In the second two equations, if we assume F is electric field, for the first constant we find $6.118\times10^{-6}\,A/V$, close enough. However using the same procedure, for the constant in the exponent, we find $6.792\times10^{7}\ 1/(cm\ V^{1/2})$, not so close, and this occurs in the

most sensitive part of the relation.[4] Evaluating these constants by setting the gap in the fundamental tunnelling integral to one centimeter, we find $5.55 \times 10^{-6}\, A/V$ and $6.77 \times 10^{7}\, 1/cm\, V^{3/2}$.

We believe the NF result with parameters $6.118 \times 10^{-6}\, A/V$ and $6.792 \times 10^{7}\, 1/(cm\, V^{1/2})$ is correct and the departure from our computed results stems from a difference in the assumed potential function. NF considers only one boundary, that between our Regions 1 and 2, between the first metal and the vacuum. They then assume a linear potential throughout Region 2 with Region 2 extending to infinity, with no Region 3. Since we will have reflections at the boundary between Regions 2 and 3, we would expect our results to be below those of NF, which is what we observe.

Our preliminary conclusions are:

> Reflection effects play a noticeable role in the behavior of quantum tunnelling varistor devices.
>
> The reflection effects will be most significant for small intermetallic gap widths.
>
> For our present purposes, our calculations are sufficiently accurate.
>
> Over restricted domains, a relation of the form
>
> $$J_{NF}(E) = K_{NF} \frac{\varepsilon_F^{1/2}}{(\phi + \varepsilon_F)\phi^{1/2}} E^2 exp\left(-\alpha_{NF} \phi^{3/2}/E\right)$$
>
> will serve as an adequate representation of quantum tunnelling varistor current densities where the *NF-parameters* K_{NF} and α_{NF} are determined from the fundamental tunnelling integral, which includes reflection effects.[5]

The inversion equations for the NF parameters are

$$\alpha_{NF} = \left[\phi^{3/2}\left(\frac{1}{E_1} - \frac{1}{E_2}\right)\right]^{-1} ln\left(\frac{E_1}{E_2}\right)^2 \frac{J_2}{J_1}$$

$$K_{NF} = \frac{J_1}{E_1^2} \frac{(\varepsilon_F + \phi)\phi^{1/2}}{\varepsilon_F^{1/2}} exp\left(\alpha_{NF} \phi^{3/2}/E_1\right) = \frac{J_2}{E_2^2} \frac{(\varepsilon_F + \phi)\phi^{1/2}}{\varepsilon_F^{1/2}} exp\left(\alpha_{NF} \phi^{3/2}/E_2\right)$$

---

[4]We suspect NF may have erroneously dropped the factor of three in the exponent which would have given them the factor 2.037×10^{8}.

[5]In our future work, we will continually monitor this assumption.

As previously stated, in the final analysis, the NF-parameters will be determined empirically. Our major concern here is to recognize the factors affecting them and to semiquantitatively assess the effects of their variation.

In some of our work, we will not wish to explicitly carry the work function and Fermi energy. We express the NF approximation in the form

$$J_{NF}(E) = K_E E^2 e^{-\beta_E/E}$$

and identifying the new parameters as

$$K_E = K_{NF} \frac{\varepsilon_F^{1/2}}{(\phi + \varepsilon_F)\phi^{1/2}} \quad \text{and} \quad \beta_E = \alpha_{NF} \phi^{3/2}$$

for copper, we have

$$K_E = 6.4 \times 10^{-7}\ A/V^2 \quad \text{and} \quad \beta_E = 6.8 \times 10^8\ V/cm$$

II-1-10 Evaluation of the Schottky Effect

When an electron is removed from a metallic surface, it leaves behind its image charge. In the presence of an applied electric potential, the total potential is the sum of the applied potential and the Coulomb potential of the electron and its image. The potential energy of the electron and its image is given by

$$U(x) = -\int_{-\infty}^{x} \frac{e^2}{(2x)^2} dx + \int_{\infty}^{x} \frac{e^2}{(2x)^2} dx = -\frac{e^2}{2x}$$

The total potential energy then takes the form

$$U(x) = \phi - eEx - \frac{e^2}{2x}$$

where we have taken the origin of coordinates at the metal surface and the zero of potential at the Fermi energy.

With some rather heavy handed mathematics, Nordheim addresses this problem in the August issue of the same journal previously cited. Following a lead given by Solymar and Walsh [16],[6] we can avoid the mathematics by noting that we have previously shown the symmetric wall barrier gives a very good approximation to the correct functional *form* for the current density.

---

[6]However, note there is the ubiquitous factor of two error in Solymar and Walsh's potential function which we have here corrected.

$$J(E) = J_0 e^{-qw} \quad \text{with} \quad q = \frac{\sqrt{2m\,\Delta U}}{\hbar}$$

where w is the width of the barrier and ΔU is the height of the barrier above the tunnelling electron energy. For our present case, we take the symmetric wall barrier with a height equal to the maximum of U and a width given by the coordinate where U crosses the zero of potential. with a slight approximation for w, for the field strengths of interest, these two are given by

$$w = \frac{\phi}{eE} \qquad \text{and} \qquad \Delta U = \phi - \left(2e^3 E\right)^{1/2}$$

The current density then takes the form

$$J(E) = J_0 exp\left(\frac{\phi\sqrt{2m\,\Delta U}}{e\hbar E}\right)$$

and we wish to compare this to the NF approximation. The matter of interest here is the effect on the change of the exponential function, so for the sake of comparison, we will take J_0 to be the same in both cases. The result is shown in Figure II-1-20.

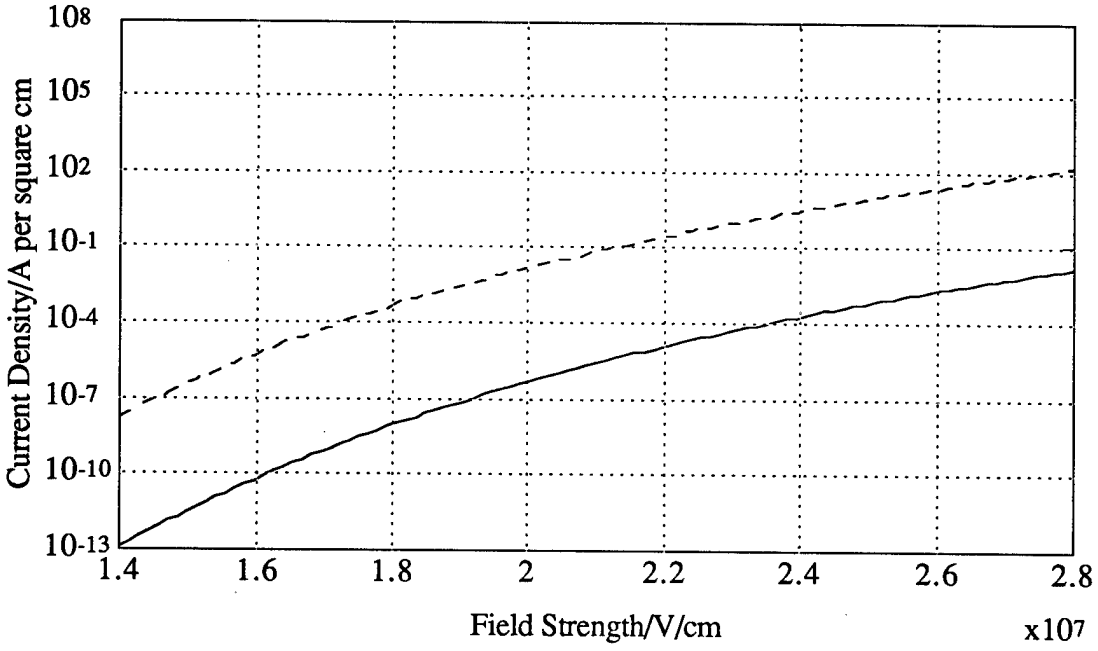


Fig. II-1-20 Comparison of results with and without the Schottky effect. The solid curve is without Schottky and the dashed with Schottky.

In their article, NF claimed the Schottky effect would be small for high field emission, one reason we have not given it much attention until the present.

The above result indicates to the contrary. However, we observe an important result. Although the magnitudes of the current densities are rather drastically altered, the *form* of the function is not. This means, with proper adjustment of parameters, we can still use the NF function over limited ranges of fields for design purposes. In general, we will usually be interested in the current density variation with a two-to-one variation of applied potential.

II-1-11 Varistor Volt-Ampere Characteristic

We now move from the physics to the engineering. Varistors are electrically described by their volt-ampere characteristic. If n is the number of intermetallic gaps from one electrode to another and g is the width of the gaps, in terms of the voltage across the electrodes, the gap field will be given by[7]

$$E = \frac{V}{ng}$$

Where A is the area of the electrodes, substituting into the NF approximation, we have for the varistor current in terms of voltage

$$I = A K_E \left(\frac{V}{ng}\right)^2 e^{-ng\beta_E/V}$$

We will also write this as

$$I = K_V\, V^2 e^{-\beta_V/V} \quad \text{with} \quad K_V = A K_E \left(\frac{1}{ng}\right)^2 \quad \text{and} \quad \beta_V = ng\beta_E$$

The work functions of the metals can be varied by the nature of the coatings we put on them. For a given metallic formulation[8], our only other degrees of freedom are the electrode area A and the ng-product.

II-1-12 Primary Electrical Varistor Specifications

We will recognize two regimes: the operational regime and the threat regime. The *operational regime* is where there are no high level spurious voltages present. The *threat regime* is where there are high level spurious voltages present. There are a number of specifications applied to varistor devices; however, in this section, we will only be concerned with the primary electrical specifications. They are: reverse standoff voltage, V_R;

---

[7]We assume here all the gaps are the same. Since we will be dealing with coated particles in contact with one another, this is not likely a bad assumption.

[8]Henceforward when we speak of a metal, we shall mean elemental metal *and* its associated coating.

maximum reverse leakage, I_R; maximum peak current, I_P; and maximum clamping voltage, V_C. We will define these in turn.

Reverse standoff voltage, V_R
Maximum voltage to appear in the circuit in the operational regime.

Reverse leakage, I_R
Maximum allowed current through the varistor at the reverse standoff voltage.

Maximum peak current, I_P
Maximum current through the varistor in the threat regime.

Maximum clamping voltage, V_C
Maximum voltage across the varistor at maximum peak current.

For a quantum tunnelling varistor, these specifications uniquely determine the bounds on K_V and β_V through the equations

$$\beta_V = \left(\frac{1}{V_R} - \frac{1}{V_C}\right)^{-1} ln \frac{I_P}{I_R}\left(\frac{V_R}{V_C}\right)^2 \quad \text{and} \quad K_V = \frac{I_R}{V_R^2} e^{\beta_V/V_R} = \frac{I_P}{V_C^2} e^{\beta_V/V_C}$$

where the last equation can serve as a check on the calculation. The *ng*-product is then given by

$$ng = \frac{\beta_V}{\beta_E} \quad \text{and the electrode area by} \quad A = (ng)^2 \frac{K_V}{K_E}$$

As an example, we consider the following set of primary specifications:

$$V_R = 180\ V,\ I_R = 10\ \mu A,\ I_P = 180\ A,\ V_C = 335\ V$$

For the metal, we will use copper with a coating that does not alter its work function. We then find

$$\beta_V = 6.02 \times 10^3\ V/cm,\ K_V = 1.00 \times 10^5\ A/V^2,\ ng = 8.8 \times 10^{-6}\ cm,\ A = 13.4\ cm^2$$

The resultant volt-ampere characteristic is shown in Figures II-1-21 and II-1-22.

The specification given here was taken from a commercial varistor catalog. It is interesting to note that the dimensions for our example are about the same as for the commercial device. If we used a material with a gap width of *100 Å*, the required number of gaps would be about nine.

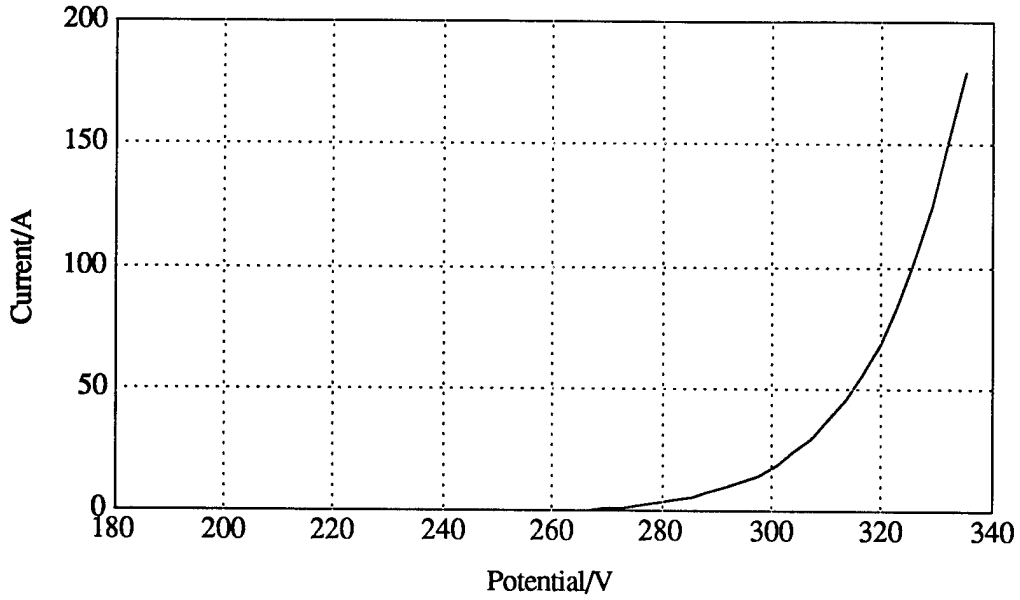


Fig. II-1-21 Linear plot of the volt-ampere characteristic for the varistor of our example.

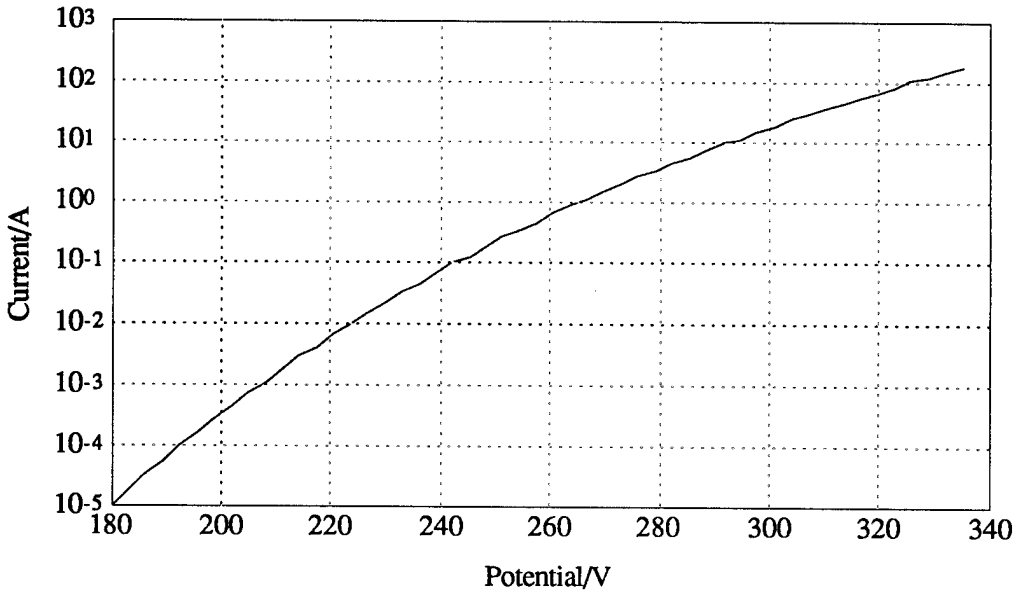


Fig. II-1-22 Semilog plot of the volt-ampere characteristic for the varistor of our example.

II-1-13 Elementary Thermal Considerations

Among other factors, the survivability of a quantum tunnelling device is dependent upon its thermal capacity and the details of the threat regime scenario. In this section, we will take a highly simplified view and consider a single pulse threat regime. We will assume the pulse is sufficiently short so there is insignificant thermal diffusion before the

varistor reaches its maximum temperature. Saying it a different way, all the pulse energy is instantaneously converted to thermal energy with a consequent rise in temperature. It is to be noted that for a single pulse, this is a worst case scenario.

In quantum tunnelling, the majority of the the energy will be delivered to the metallic phase. With this as an additional simplifying assumption, the rise in temperature will be given by

$$\Delta T = \frac{\Delta Q}{C}$$

where Q is the pulse energy and C is the heat capacity of the metallic phase. Another important varistor specification is its peak pulse power defined as follows.

Peak Pulse Power, P_P
The power delivered at the clamp voltage with peak current, numerically, the product of the clamp voltage and the peak current.

The energy in a pulse may be given by the product of its peak pulse power and the pulse equivalent time t_{eq}. The equivalent time is the duration of a rectangular pulse delivering the same energy as a given pulse [14]. Using these concepts, the temperature rise may be written as

$$\Delta T = \frac{I_P V_C t_{eq}}{c \rho V_m}$$

where c is the metallic specific heat, ρ is the density, and V_m the volume. The area was previously determined by the electrical specifications. What the last equation determines is the minimum thickness to meet the thermal requirements. For a specified temperature rise, this thickness is given by

$$t_{min} = \frac{I_P V_C t_{eq}}{A c \rho \Delta T}$$

This, in turn, determines the minimum diameter of particles to be used.

$$d_{min} = \frac{t_{min}}{n}$$

For our example, the specified equivalent time was about $40\ \mu s$ which gives a total energy of $2.41\ J$. Using the parameters for copper with a temperature rise of $100\ C$, we find $t_{min} = 340\ \mu m$ and $d_{min} = 37.8\ \mu m$. For experimental convenience, our test samples will have a thickness of $2\ mm$. For this example, this would then require particles with diameters of $222\ \mu m$.

II-1-14 Critique of the Preceding Analyses

In the preceding discussion, we have made two drastic assumptions. The first is that there is no significant interaction between the tunnelling electrons and the insulative coatings. For sufficiently thin coatings, that may well be true, but consideration of this matter will be a major concern of further efforts in this project. The questions are, what are the interactions and to what use can they be put?

The second drastic assumption is the one of planar geometry and the subsequent reduction to one dimension; whereas, the tunnelling gaps in our materials will be between spherical particles. Our presumption is that though the detailed results may vary, the overall functional dependences and orders of magnitude will be similar. For spherical particles, there will be a field enhancement by a factor not exceeding three that is not accounted for in our analyses. We have assumed a discontinuous potential function. Because of reflection effects, our estimates of the planar current densities will be lower than for a more realistic potential function.

We consistently used the free electron model for electrons in a metal for which the Fermi surface is strictly spherical. This allows us to perform our integrations in spherical coordinates with no functional dependences in the upper limit of the energy integral. This is one of the reasons we have used copper in our examples, because the Fermi surface of copper is nearly spherical. It is not likely copper will be a metal of choice.

Other metals have more complex Fermi surfaces especially those with overlapping bands, ones with an even number of electrons [16]. However, the departure from sphericity stems from the density of states function penetrating the surfaces of the first Brillouin zone. Although this leads to profound effects in phenomena such as cyclotron resonances, the penetrations are relatively small and will not greatly affect bulk transport properties, the ones with which we are dealing here. Because of this, the free electron assumption is likely to represent a small contribution to our overall departures from a more realistic situation.

The volumes, areas, and thicknesses computed in the previous two sections should be considered as *equivalent volumes, areas, and thicknesses*, since planar geometry was tacitly assumed. We also assumed the particles were stacked directly above one another and did not account for the volumetric packing fraction, about seventy-four percent for close packing.

Our present tunnelling model is certainly crude. However, we regard empirical results as an essential ingredient for guidance in its refinement, a major goal of the current effort, to be reported in subsequent sections.

One of the recurrent themes appearing during the execution of these tasks is the importance of paying attention to the relative magnitudes of various quantities that occur in our analyses. It is frequently these relative

magnitudes that allow certain approximations and dictates the direction the analysis should take. A great deal of care must be exercised, because the magnitudes with which we are dealing are so far removed from our daily experiences. What seems to us in our everyday lives to be a large quantity may be quite ordinary on an atomic scale. We would like to illustrate this point in a somewhat amusing way.

Consider half a gallon of water. Say we send all the nuclei from that sample to Mr. Alston in California and all the electrons to Mr. Scott in Virginia. We invite all who read this to calculate the force of attraction at that distance between the nuclei in California and the electrons in Virginia. You may wish to express your final answer in units of kilotons.

In the work presented so far, we have mainly dealt in terms of static and steady state concepts. The associated quantum mechanical analyses have lead to a substantial amount of insight, but our ultimate aim is the treatment of transient phenomena. Thus we must, at some point, face up to the demands of addressing the time dependent equations. Further, we have substantially ignored the insulating phases in our composite system. The following sections are a first effort to extend the previous efforts to cover these aspects.

Section II-2

ANALYSIS OF TEMPERATURE DEPENDENCE

II-2-1

A major advantage of quantum tunnelling devices is their near independence from temperature effects. To see why this is so, we consider the fundamental tunnelling integral derived in Section II-1. There, we had

$$J(E)=\frac{em}{\pi^2\hbar^3}\int_0^{\varepsilon_F}d\varepsilon\int_0^{\pi/2}d\theta\;\varepsilon\,T\left(\varepsilon\cos^2\theta,E\right)\cos\theta\sin\theta$$

where we took the low temperature form for the Fermi-Dirac distribution function which, in general, is given by

$$f(\varepsilon)=\frac{1}{1+exp\left(\dfrac{\varepsilon-\varepsilon_F}{k_BT}\right)}$$

where T is the absolute Kelvin temperature.[1] The exact form for the fundamental tunnelling integral is then

$$J(E)=\frac{em}{\pi^2\hbar^3}\int_0^{\infty}d\varepsilon\int_0^{\pi/2}d\theta\;\varepsilon\,f(\varepsilon)T\left(\varepsilon\cos^2\theta,E\right)\cos\theta\sin\theta$$

which we choose to write in the form

$$J(E)=\frac{em}{\pi^2\hbar^3}\int_0^{\infty}f(\varepsilon)I_o(\varepsilon,E)d\varepsilon\quad\text{with}\quad I_o(\varepsilon.E)=\int_0^{\pi/2}\varepsilon\,T\left(\varepsilon\cos^2\theta,E\right)\cos\theta\sin\theta\,d\theta$$

where I_o is the zero temperature energy integrand with the angles integrated out. Using the formulation for the triangular wall barrier and evaluating this function for an acceptable number of points to perform the energy integration would entail a rather large amount of computing time. To obtain an overall picture of what happens, we will resort to the symmetric wall barrier and match I_o computed from the symmetric wall barrier to one computed from the triangular wall barrier. The result is surprisingly good and is shown in Figure II-2-1.

---

[1]Not to be confused with the transmission coefficient. The distinction is made from context.

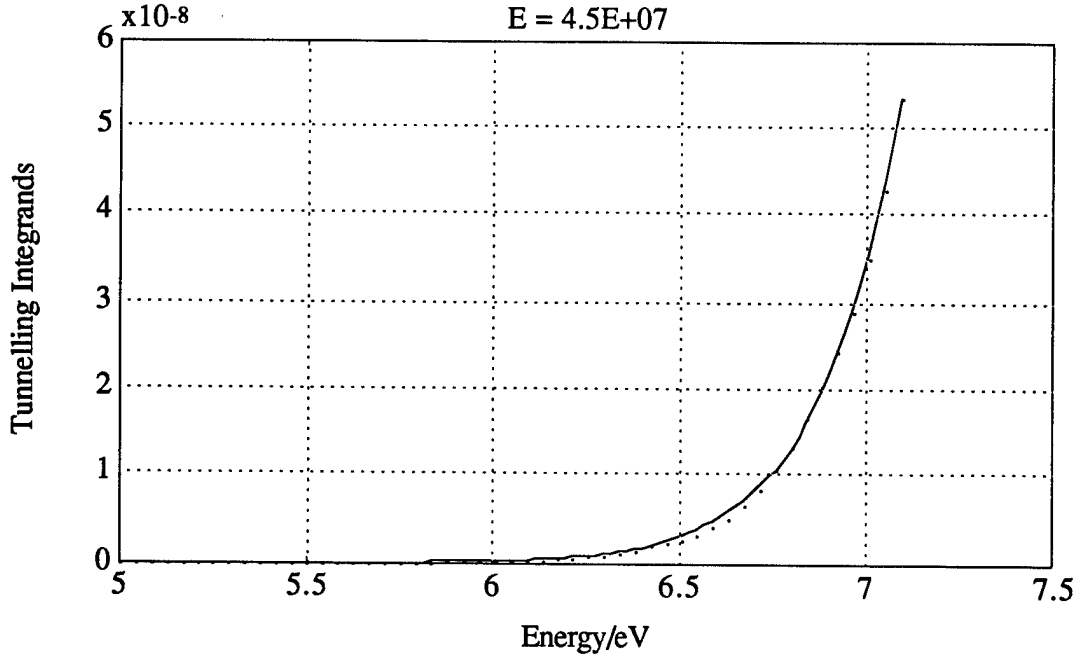


Fig. II-2-1 The solid curve represents the values of I_o as computed from the functional form of the symmetric wall barrier and matched to the points shown, computed from the triangular wall barrier.

The other factor appearing in the energy integral for J is the Fermi function shown for various temperatures in Figure II-2-2.

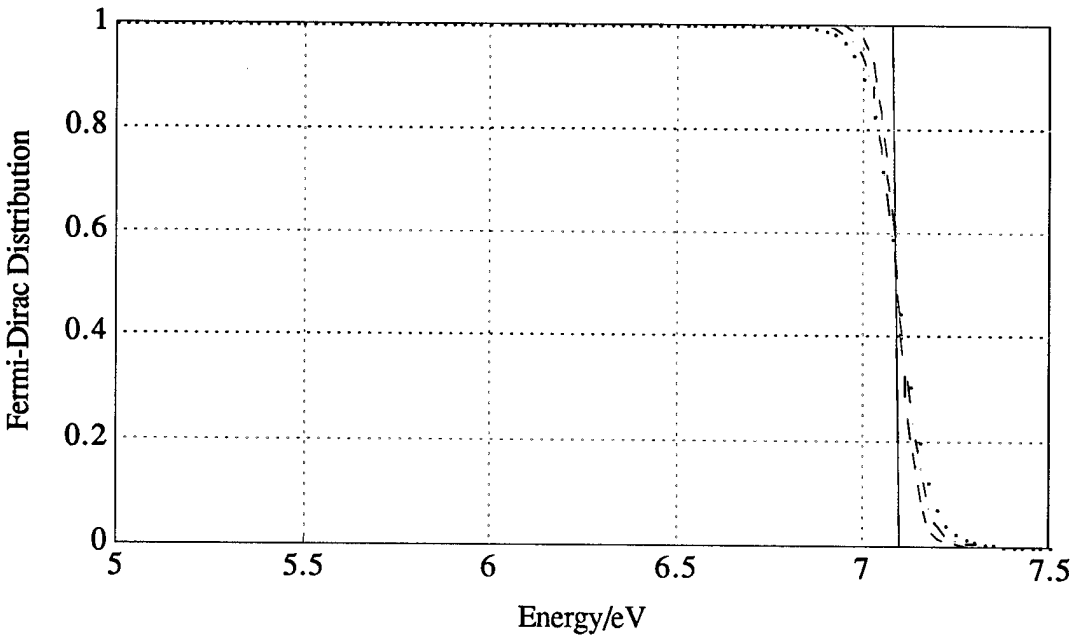


Fig. II-2-2 The temperatures in degrees Kelvin are: solid, 4; dashed, 295; dash dot, 395; dot, 495.

The rest of the story is told by the following figures and their captions.

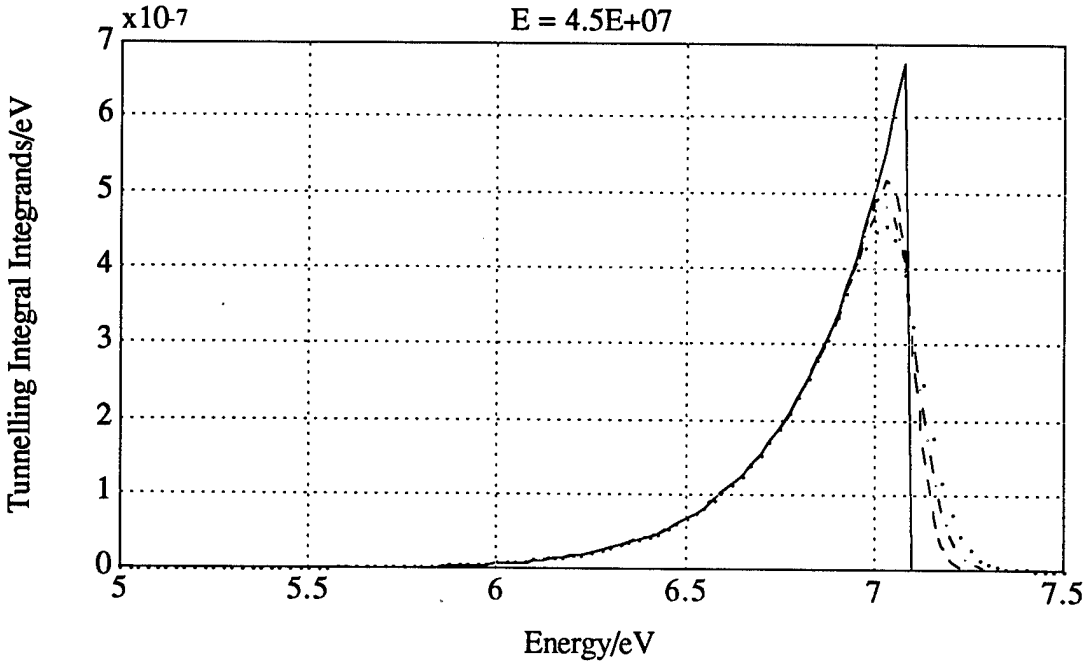


Fig. II-2-3 The effect of the Fermi function on the tunnelling integral integrand for various temperatures. The temperatures in degrees Kelvin are: solid, 4; dashed, 295; dash dot, 395; dot, 495.

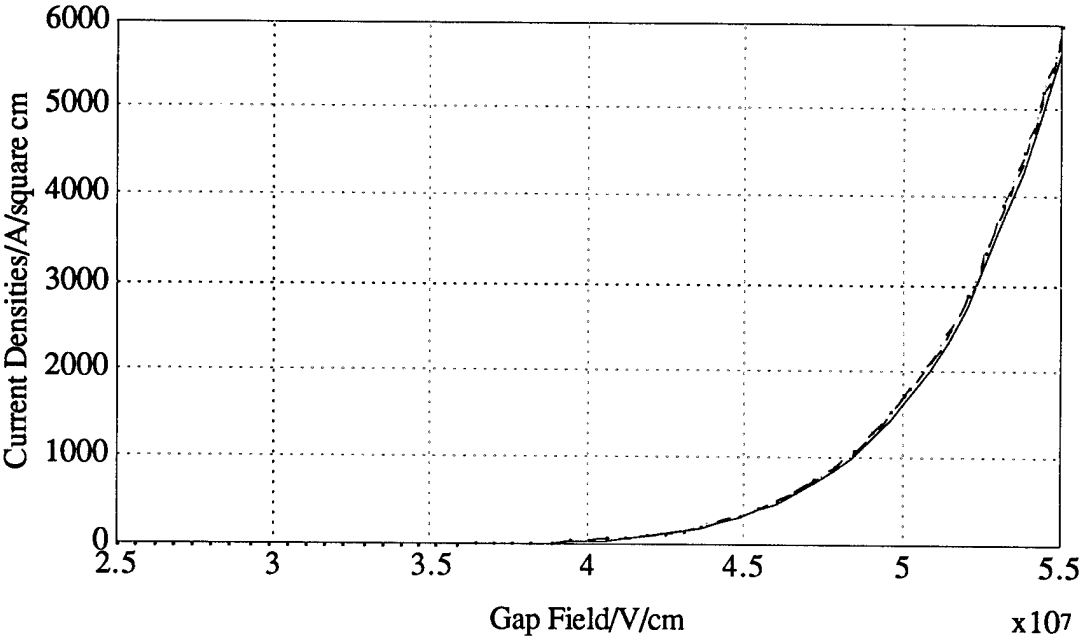


Fig. II-2-4 Resultant current density after the energy integration. On this scale, there is little observed temperature effect. The temperatures in degrees Kelvin are: solid, 4; dashed, 295; dash dot, 395; dot, 495.

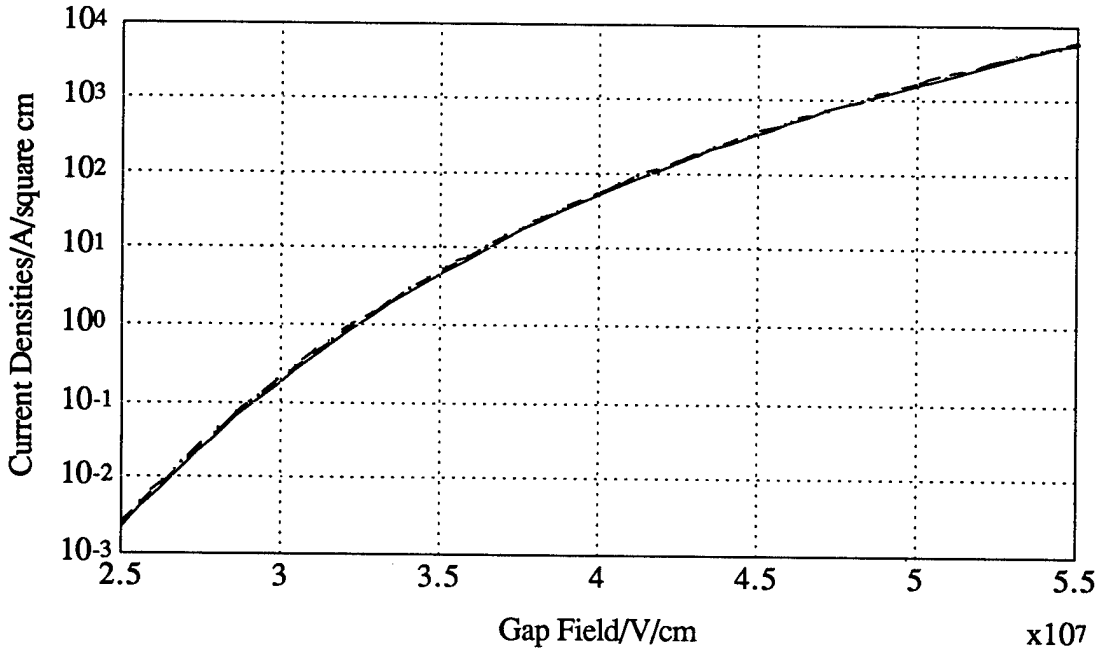


Fig. II-2-5 Semilog plot of the previous result. The temperatures in degrees Kelvin are: solid, 4; dashed, 295; dash dot, 395; dot, 495.

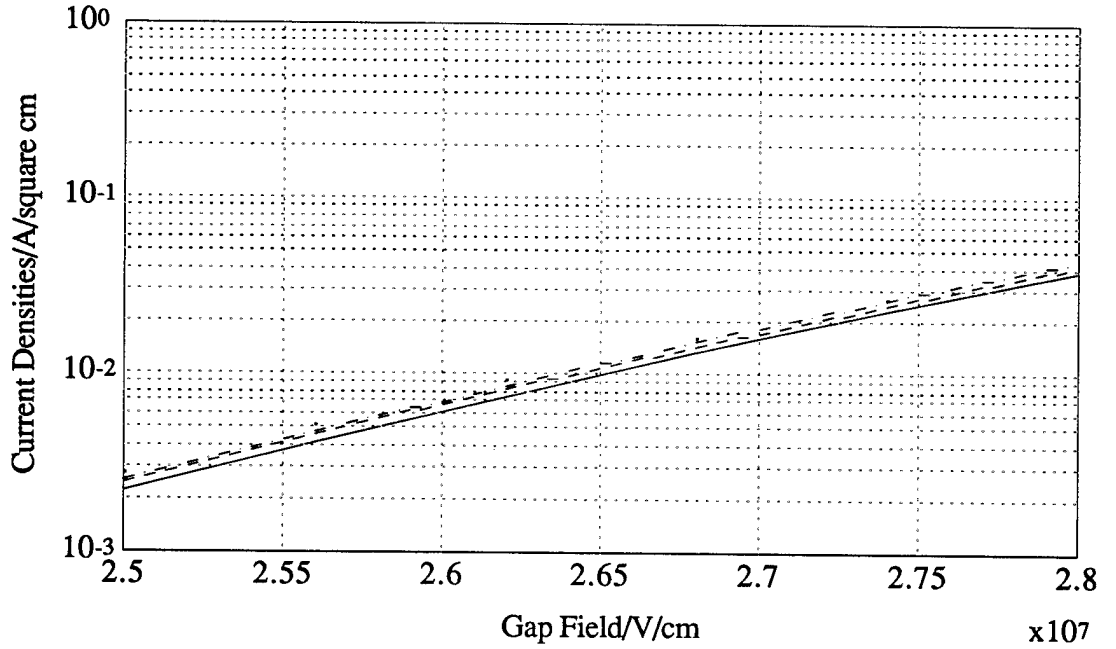


Fig. II-2-6 Expansion at low fields. The temperatures in degrees Kelvin are: solid, 4; dashed, 295; dash dot, 395; dot, 495.

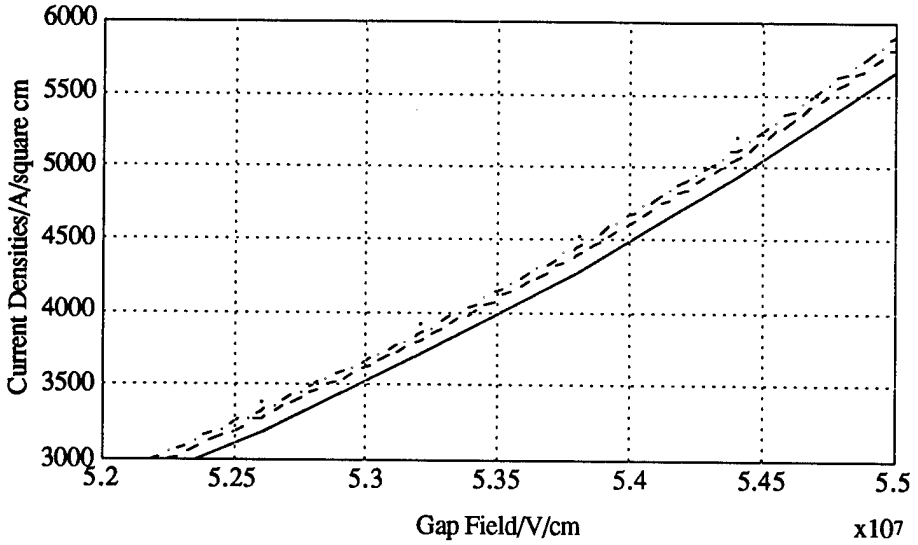


Fig. II-2-7 Expansion at high fields. The temperatures in degrees Kelvin are: solid, 4; dashed, 295; dash dot, 395; dot, 495.

We note that *295 °K* is essentially room temperature, *395 °K* is the boiling point of water, and *495 °K* is above the melting point of lead. What we observe is, there is little practical temperature dependence predicted.

Although the above presentation gives an excellent qualitative picture of what is happening, it is quantitatively inaccurate except in the field range where I_o was matched. The predicted current densities for low fields are above the exact calculations and those for high fields are below the exact calculations. However, the matter in which we are mainly interested here is the current density spreads as a function of temperature. These have been calculated using the more exact triangular wall barrier with the results shown in Table II-2.

$\downarrow E, \vec{T}$		4	195	295	395	495
2.5	$\vec{J}$	5.56×10^{-4}	6.43×10^{-4}	6.84×10^{-4}	7.42×10^{-4}	8.19×10^{-4}
	$\vec{\%}$	21	6.2	0	8.1	18
4.0	$\vec{J}$	38.0	42.4	43.4	44.8	46.4
	$\vec{\%}$	13	2.3	0	3.2	6.7
5.5	$\vec{J}$	7.64×10^{3}	8.19×10^{3}	8.27×10^{3}	8.38×10^{3}	$8.52\times10\times10^{3}$
	$\vec{\%}$	7.92	0.97	0	1.3	1.66

Table II-2 The current densities, J, are in A/cm^2, the temperatures, T, are in degrees Kelvin and the field strengths, E, in $V/cm \times 10^7$. The percentages are changes with respect to room temperature.

The percentage changes with respect to temperature are small and decrease as the fields are increased. We conclude that *this conduction mechanism does not significantly contribute to any tendency toward thermal runaway*. Further, the smallness of the changes is such that *devices with these characteristics would be virtually temperature independent from a practical standpoint.*

Section II-3

CHARGE TRANSPORT FORMULATION

II-3-1 Some Properties of the Fermi-Dirac Distribution Function

For present and later purposes, it is worthwhile to briefly review some of the properties of the Fermi-Dirac distribution function which is a central result of equilibrium quantum statistical mechanics. For a gas of weakly interacting particles obeying the Pauli exclusion principle, it states that in an ensemble of systems, the average number of particles in quantum state r with energy ε_r is given by the function

$$f(\varepsilon_r) = \frac{1}{1 + exp\left(\frac{\varepsilon_r - \mu}{k_B T}\right)}$$

Where μ is the thermodynamic chemical potential, k_B is Boltzmann's constant, and T is the absolute temperature. For electrons in a metal, the label r includes the wave number $\boldsymbol{k}$ and the spin quantum number $\pm s$. The chemical potential is implicitly defined by the requirement that the sum over all states of the distribution function must equal the total number of particles in the system, symbolically

$$\sum_r f(\varepsilon_r) = N$$

If this sum is approximated by an integral with respect to energy, the density of states must be included.

$$\int_0^\infty \mathcal{D}(\varepsilon) f(\varepsilon) d\varepsilon = N$$

For electrons in solids, the chemical potential is called the Fermi energy.

With T equal to zero, if ε_r is greater than μ, the function is unity. If ε_r is less than μ, the function is zero. In this case, the function is rectangular. Further, in this case, the derivative of the function is strictly a negative Dirac delta function.

At ordinary temperatures in the neighborhood of *300 K°*, the delta function is still a reasonably good approximation of the derivative. To see this, we observe the derivative function

$$\frac{df(\varepsilon)}{d\varepsilon} = -\left\{ \frac{1}{k_B T} exp\left(\frac{\varepsilon - \mu}{k_B T} \right) \Big/ \left[1 + exp\left(\frac{\varepsilon - \mu}{k_B T} \right) \right]^2 \right\}$$

Setting the derivative of the derivative function to zero shows the minimum of the derivative function occurs when the energy is equal to the chemical potential. This minimum value is

$$\left. \frac{df(\varepsilon)}{d\varepsilon} \right]_{\varepsilon=\mu} = -\frac{1}{4k_B T}$$

which goes to minus infinity as the temperature goes to zero. Further, the integral of the derivative function is seen to be minus one. This is an alternate demonstration that the derivative of the distribution function is a negative delta function in the limit as the temperature goes to zero.

It is further found that the half width points occur at

$$\varepsilon - \mu = k_B T ln\left(3 \pm \sqrt{8}\right)$$

At room temperature, $k_B T$ is found to be about $0.025\ eV$; thus the width of the function is small in comparison to the chemical potential ($\approx 7.09\ eV$ for copper), so the delta function approximation is certainly adequate for order of magnitude calculations.

II-3-2 A Variant of Liouville's Theorem

For a system of particles, we define the *phase space* as the 6-dimensional space whose coordinates are the canonically conjugate positions and momenta of the individual particles. A particle motion can be described by a trajectory in phase space parameterized with respect to time. The system is then described by a swarm of points in the phase space and if there are a large number of points, we can treat them as an imaginary continuous fluid.

Consider a small volume of phase space δV_p, always containing the same particle points. As the motion of the system unfolds, this volume will change but will always contain the same number of points. We define the position vector in phase space and the gradient in phase space respectively as

$$\boldsymbol{r}_{ph} = \sum_{i=1}^{3} \hat{x}_i x_i + \hat{p}_i p_i \quad \text{and} \quad \nabla_{ph} = \sum_{i=1}^{3} \hat{x}_i \frac{\partial}{\partial x_i} + \hat{p}_i \frac{\partial}{\partial p_i} \quad \text{with velocity} \quad v_{ph} = \dot{\boldsymbol{r}}_{ph}$$

Recall that the interpretation of the divergence of a vector field is that it is the flux of the field per unit volume exiting a closed surface. If the field is a velocity field, the change in the volume in a time dt is

$$\Delta\left(\delta V_{ph}\right) = \nabla_{ph} \cdot v_{ph}\, \delta V_{ph}\, dt \quad \text{from which it follows} \quad \frac{dV_{ph}}{dt} = \int \nabla_{ph} \cdot v_{ph}\, dV_{ph}$$

Writing out the integrand and applying Hamilton's equations of motion shows that

$$\nabla_{ph} \cdot v_{ph} = 0 \quad \text{which means} \quad \frac{dV_{ph}}{dt} = 0$$

which is the variant of Liouville's theorem we wanted. A volume in phase space always containing the same particles is constant. As the volume moves through phase space, its shape may change, but its volume will not. Since the number of particle points in the volume is always the same, the density of points within the volume does not change. This number per unit volume will be referred to as the *phase space density*, denoted by the *distribution function*, $f(\boldsymbol{q}, \boldsymbol{p}, t)$. This function uniquely determines all system properties, and our first task is to find the equation of motion for the distribution function, the subject of the next section.

II-3-3 The Boltzmann Equation

Consider a system of particles, and divide the forces acting on the particles of the system into external and internal forces. The external forces might be applied electric or magnetic fields and the internal forces we categorize as collisions of one sort or another. Now consider the system of particles where there are no internal forces acting and the external forces are denoted by $\boldsymbol{F}$, and also consider a number of particle points moving through phase space contained in a small volume

$$f(q,p,t)dV_{ph}$$

From the last section, we know both factors are constant, so the equation of motion for the distribution function is

$$\frac{df}{dt} = 0$$

and from the calculus of multidimensional variables and Newton's second law

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \dot{\boldsymbol{r}} \cdot \frac{\partial f}{\partial \boldsymbol{r}} + \dot{\boldsymbol{p}} \cdot \frac{\partial f}{\partial \boldsymbol{p}} = \frac{\partial f}{\partial t} + \dot{\boldsymbol{r}} \cdot \frac{\partial f}{\partial \boldsymbol{r}} + \boldsymbol{F} \cdot \frac{\partial f}{\partial \boldsymbol{p}}$$

Where we emphasize that $\boldsymbol{F}$ represents the *external* forces. If we now allow for collisions df/dt is no longer zero and may change because of the collisions.

$$\frac{df}{dt} = \left.\frac{\partial f}{\partial t}\right]_{collision}$$

which defines the meaning of the collision term. It is the time derivative of f in the presence of collisions with all other parameters held constant. From our previous relation this may be written in the form

$$\frac{\partial f}{\partial t} + \dot{\boldsymbol{r}} \cdot \frac{\partial f}{\partial \boldsymbol{r}} + \boldsymbol{F} \cdot \frac{\partial f}{\partial \boldsymbol{p}} = \frac{\partial f}{\partial t}\Bigg]_{collision}$$

This is the Boltzmann equation, a fundamental relation for the discussion of transport properties.

II-3-4 Collision Probability

If there were a partially filled band in a solid that had a perfectly periodic lattice, the electrical conductivity of the solid would be infinite. The finiteness of electrical conductivity is caused by electron collisions with lattice imperfections. At room temperature, these are dominated by electron-phonon interactions. The electron-electron interactions serve to maintain the conduction electrons in local thermal equilibrium but do not contribute to charge transport because they do not alter the net electronic momentum. Thus we will concentrate on the electron-phonon collisions and make the following assumptions:

The time between collisions is long in comparison to the time that an electron spends in the vicinity of a scattering center.

The scattering centers are sufficiently dilute so that multiple collisions need not be considered.

The electrons can be described by wave packets in a quasiclassical manner. This means that the extent of the wave packet in position space must be a reasonable fraction of a mean free path.

The probability of collision is independent of the past history of the electron. This is related to our assertion above concerning local equilibrium.

It will be part of our subsequent work to determine to what extent these assumptions are consistent with all results obtained.

We define the collision probability $\mathcal{P}(t)dt$ as the probability that after a collision no collision takes place for a time t and a collision does take place at time $t + dt$. We will take $\Delta t / \tau$ as the probability that a collision takes place in a small time interval Δt. With these definitions, the collision probability can be approximated by

$$\mathcal{P}(n\Delta t)\Delta t = \frac{\Delta t}{\tau}\left(1 - \frac{\Delta t}{\tau}\right)^n \quad \text{or} \quad \mathcal{P}(n\Delta t) = \frac{1}{\tau}\left(1 - \frac{\Delta t}{\tau}\right)^{\frac{t}{\Delta t}}$$

The exact result is found in the limit as Δt goes to zero.

$$\mathcal{P}(t) = \lim_{\Delta t \to 0} \frac{1}{\tau}\left(1 - \frac{\Delta t}{\tau}\right)^{\frac{t}{\Delta t}} = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

The normalization for this probability law is given by

$$\int_0^\infty \mathcal{P}(t)\, dt = \int_0^\infty \frac{1}{\tau} e^{-\frac{t}{\tau}}\, dt = 1$$

In the last equality, taking τ to the right side of the equation and differentiating with respect to τ gives

$$\int_0^\infty t\, \mathcal{P}(t)\, dt = \tau$$

The integral here is just the average of t which gives the interpretation for τ; it is the *mean time between collisions* or the *mean lifetime after a collision.*

II-3-5 Relaxation Time Approximation and Path Integral

It is frequently empirically observed that restoration of equilibrium occurs in an exponential fashion. Thus, our first approximation to a solution will be to assume the relaxation is uniquely exponential throughout phase space. We then have

$$f(\boldsymbol{r}, \boldsymbol{p}, t) = f^o + h e^{-\frac{t}{\tau}}$$

where f^o is the equilibrium distribution function. Our assumption that the relaxation is uniquely exponential throughout phase space implies h may be a function of $\boldsymbol{q}$ and $\boldsymbol{p}$ but is not a function of time, which implies

$$\left.\frac{\partial f}{\partial t}\right]_{collision} = -\frac{1}{\tau}\left(f - f^o\right)$$

and the Boltzmann equation in the relaxation time approximation can be written as

$$\frac{\partial f}{\partial t} + \dot{\boldsymbol{r}} \cdot \frac{\partial f}{\partial \boldsymbol{r}} + \boldsymbol{F} \cdot \frac{\partial f}{\partial \boldsymbol{p}} = -\frac{1}{\tau}\left(f - f^o\right)$$

In these relations, we refer to τ as the *relaxation time.*

Let us next consider a system in equilibrium with a group of particles in phase space located at ($\boldsymbol{r}$, $\boldsymbol{p}$) at time t. In equilibrium and in the absence of collisions, those particles would come from a point ($\boldsymbol{r}'$, $\boldsymbol{p}'$) located on a

collision free phase space trajectory at time $t - t'$. In symbols we would have

$$f^o(\boldsymbol{r}, \boldsymbol{p}, t) = f^o(\boldsymbol{r}', \boldsymbol{p}', t') = f^o\left[\boldsymbol{r}(t-t'), \boldsymbol{p}(t-t'), (t-t')\right]$$

However, when collisions take place, the probability that this actually happens is given by the result of the last section. We take the average of such events to be the distribution function when there are collisions. That is

$$f(\boldsymbol{r}, \boldsymbol{p}, t) = \int_0^\infty f^o\left[\boldsymbol{r}(t-t'), \boldsymbol{p}(t-t'), (t-t')\right] \mathcal{P}(t')\, dt'$$

$$= \int_0^\infty f^o\left[\boldsymbol{r}(t-t'), \boldsymbol{p}(t-t'), (t-t')\right] e^{-\frac{t'}{\tau}} \frac{dt'}{\tau}$$

This integral is referred to as the *path integral* [21], because the presumption is that we are integrating along a phase space trajectory, the path the particles would follow in the absence of collisions. The other implication of this is that $\boldsymbol{r}(t)$ and $\boldsymbol{p}(t)$ are to be found through the equations of motion in the absence of collisions. That is, the paths the particles would take when acted on by the *external* forces only.

By recalling that the Boltzmann equation can be written simply as

$$\frac{df}{dt} = -\frac{1}{\tau}(f - f^o)$$

by direct substitution with an integration by parts, it can be shown that the path integral is a solution of the Boltzmann equation in the relaxation time approximation if we equate the relaxation time with the mean lifetime after collision. What we see here is, in the relaxation time approximation, the path integral is entirely equivalent to the Boltzmann equation. Thus, we will cast the integral into a somewhat more convenient form and then particularize it to the type of equilibrium distribution we will use in our further work. The first step is a further integration by parts which yields

$$f(\boldsymbol{r}, \boldsymbol{p}, t) = f^o + \int_0^\infty \frac{df^o}{dt'} e^{-\frac{t'}{\tau}}\, dt'$$

II-3-6 The Short Relaxation Time Approximation

In this approximation, we assume the relaxation time is so short that the derivative in the integrand does not change significantly in a time on the order of τ. In that case, the derivative can be removed from the integral. Noting that

$$\frac{df^o}{dt'} = -\frac{df^o}{dt} \qquad \text{we can write} \qquad f = f^o + \tau \frac{df^o}{dt}$$

Writing out the total derivative gives

$$f = f^o + \tau\left[\frac{\partial f^o}{\partial t} + \frac{d\boldsymbol{r}}{dt}\cdot\frac{\partial f^o}{\partial \boldsymbol{r}} + \frac{d\boldsymbol{p}}{dt}\cdot\frac{\partial f^o}{\partial \boldsymbol{p}}\right] \qquad \text{or} \qquad f = f^o + \tau\left[\frac{\partial f^o}{\partial t} + \dot{\boldsymbol{r}}\cdot\frac{\partial f^o}{\partial \boldsymbol{r}} + \boldsymbol{F}\cdot\frac{\partial f^o}{\partial \boldsymbol{p}}\right]$$

II-3-7 Metallic Electrons

In a metal, the conduction electrons are governed by the Fermi-Dirac distribution. For purposes of illustration, the free electron Fermi sphere at zero absolute temperature will be used. Using the relaxation time approximation, the average electronic momentum per particle will be given by

$$\langle \boldsymbol{p} \rangle = \frac{1}{N}\int d^3p\ d^3r\ \boldsymbol{p}\left[f^o + \tau\left(\frac{\partial f^o}{\partial t} + \dot{\boldsymbol{r}}\cdot\frac{\partial f^o}{\partial \boldsymbol{r}} + \boldsymbol{F}\cdot\frac{\partial f^o}{\partial \boldsymbol{p}}\right)\right]$$

where N is the total number of particles. By symmetry, the first term vanishes. The Fermi-Dirac distribution has no explicit time or position dependence so the second and third terms vanish. This leaves us with

$$\langle \boldsymbol{p} \rangle = \frac{1}{N}\int d^3p\ d^3r\ \boldsymbol{p}\tau\boldsymbol{F}\cdot\frac{\partial f^o}{\partial \boldsymbol{p}}$$

In this situation, the distribution function is a function of the magnitude of $\boldsymbol{p}$; it is spherically symmetric. It follows that

$$\langle \boldsymbol{p} \rangle = \frac{1}{N}\int d^3p\ d^3r\ \boldsymbol{p}\tau\boldsymbol{F}\cdot\frac{\partial f^o}{\partial p}\frac{\partial p}{\partial \boldsymbol{p}} = \frac{1}{N}\int d^3p\ d^3r\ \boldsymbol{p}\tau\boldsymbol{F}\cdot\hat{\boldsymbol{p}}\frac{\partial f^o}{\partial p}$$

$$= \frac{1}{N}\int d^3p\ d^3r\ \boldsymbol{p}\tau\boldsymbol{F}\cdot\hat{\boldsymbol{p}}\frac{\partial f^o}{\partial \varepsilon}\frac{\partial \varepsilon}{\partial p} = \frac{1}{Nm}\int d^3p\ d^3r\ \boldsymbol{p}\tau\boldsymbol{F}\cdot\boldsymbol{p}\frac{\partial f^o}{\partial \varepsilon}$$

As shown in the first subsection, even at room temperature, a delta function is a good approximation to the derivative of the of the Fermi-Dirac distribution, giving

$$\langle \boldsymbol{p} \rangle = \frac{1}{N}\int d^3p\ d^3r\ \boldsymbol{p}\tau\boldsymbol{F}\cdot\boldsymbol{p}\,\delta\left(\frac{p^2}{2m} - \varepsilon_F\right)$$

By transforming to spherical variables and picking them so that the force is in the z-direction, and noting the normalization for our distribution function is

$$N_{ph} = \frac{3}{4\pi p_F^3}\frac{N}{V}$$

we can now perform the integration[1]

$$\langle \boldsymbol{p} \rangle = \tau_F F_z \hat{\boldsymbol{k}}$$

This is reasonable since it is just the impulse applied to the electron by the force over the mean time of application. The average velocity is given by

$$\langle \boldsymbol{v} \rangle = \frac{\langle \boldsymbol{p} \rangle}{m} = \frac{\tau_F F_z \hat{\boldsymbol{k}}}{m}$$

If the force is created by an electric field, the current density would then be given by

$$\langle \boldsymbol{J} \rangle = (ne)\langle \boldsymbol{v} \rangle = (ne)\frac{\tau_F e\boldsymbol{E}}{m} = \frac{n\,e^2\tau_F}{m}\boldsymbol{E}$$

where n is the number of electrons per unit volume. Thus we find for the conductivity[2]

$$\sigma = \frac{n\,e^2\tau_F}{m}$$

There are two aspects of this result that are important for our present considerations. The first is that the collision time is generally a function of velocity. Thus taking it outside the integral involves a rather gross assumption. However, we see that when we are working with Fermi-Dirac statistics, it is a relatively good assumption. *The importance to us here is that this means the mean free path calculations will also be relatively accurate* where we define the mean free path by the relation

$$\Lambda_F = v_F \tau_F$$

The second point is that the conductivity equation implies the electrons contributing to the conductivity are those whose energies are near the Fermi surface. That is a true statement. However, note that the n in the above relation is the *total* number of electrons per unit volume. In a sense, that is an accident of the form of the functions that appear in the formulation of the problem and how they are affected under integration.

---

[1] The integration is carried out using spherical variables with the vectors referenced to a Cartesian frame. It is instructive to write this out to see why the x and y components average to zero which, by symmetry, they must.

[2] This is a well known result and one may wonder why we went through this analysis. It was the delineation of the assumptions leading to the result we wished to uncover.

Section II-4

QUANTUM STATES AND TIME SCALES

II-4-1 Introduction

We shall make a distinction here between a dynamical variable and the value of the dynamical variable. In quantum physics, dynamical variables are not represented by numbers. They are represented by operators. We do not want to keep repeating the word operator, so when we say *dynamical variable*, we shall mean the operator associated with that dynamical variable. The results of measurements yield *values* of the dynamical variable, specifically, eigenvalues of the operator.

The state of a system is specified by giving the values of a complete set of commuting dynamical variables. If a dynamical variable commutes with the system Hamiltonian, the value of the dynamical variable is a constant of the motion. To find the general time development of a system, one frequently resorts to the time dependent Schrödinger equation. Except for the simplest cases, that is a difficult equation to solve and one must usually resort to approximate methods.

In cases where the Hamiltonian is itself a function of time, the Hamiltonian can often be separated as a sum of time dependent and time independent parts where the time independent part represents the natural motion of the system. In these circumstances, the approximations applicable are dependent on how fast the Hamiltonian is changing with respect to the natural motion of the system. However, the natural motion of a system may also be a difficult problem. Thus, we need some general rules for conveniently estimating the time scales of natural motions. Before we do this, we need to recognize the character of various states that can exist.

II-4-2 Classification of States

A classification we will use is based on a comparison of the total energy of the system to its potential energy. Although we will not go through the details, the results are based directly on what one can expect from the general solutions of the Schrödinger equation.

Bound states are those for which there is a center of attraction and there exists a distance beyond which the system total energy never exceeds the potential energy. The wave function will be exponentially damped at distances far from the center of attraction and the energy eigenvalues will be discrete.

Unbound states are those for which there are always regions where the total energy is above the potential energy, no matter how far from the center of attraction. The energy eigenvalues are always continuous.

Metastable states are those for which the potential energy is above the total energy in the vicinity of the center of attraction but drops below, and stays below, the total energy at some distance from the center of attraction. These states are unbound according to the previous definition; their energy eigenvalues are continuous.

II-4-3 Bound State Time Scales

The picture that is often invoked here is that of a particle in a box. As noted above, solution of the Schrödinger equation for this problem will lead to discrete energy levels. For the time scale, what is generally envisaged is that the particle is bouncing back and forth between the walls of the box. The time scale is defined in terms of the particle transit time which is found using equations from classical mechanics.

$$\tau_n = \frac{L}{v}$$

where L is a linear dimension of the box and v is the particle velocity. Setting the zero of potential interior to the box, the velocity is found from

$$v = \sqrt{\frac{2T}{m}} \qquad \text{yielding} \qquad \tau_n = L\sqrt{\frac{m}{2T}}$$

where m is the particle mass and T the kinetic energy. Quantum mechanically, this picture is, of course, nonsense. There is no particle trajectory here and hence, no velocity. However, it is well known that time and energy in physics are intimately related. For example, the conservation of energy can be regarded as a consequence of a time symmetry. Further, the last equation for τ_n leads to reasonably accurate results in situations where we intend to use it. Thus, one might conclude that the time energy relation is the fundamental one, and the picture is simply a mnemonic for remembering the equation. One of the advantages of this equation is that the energy is something one may know from measurements as opposed to being required to calculate the energy theoretically.

In our applications, we will be concerned with high speed transients acting on materials. The approach used for analyzing the material responses is dependent on the time rates of the natural motions of the materials as compared to the time rate of any applied perturbation. Thus we need convenient means for estimating time rates of material natural motions. Although we will couch the discussion in quantum mechanical terms, it is noted that similar considerations apply to macroscopic systems adequately described by classical mechanics.

For an unperturbed system, the Hamiltonion will be time independent. Thus for the Hamiltonion of the perturbed system we write

$$H = H_o + U$$

where the first term completely describes the unperturbed system and the second represents the time dependent perturbation.

There are two sources for nonzero electric susceptibilities. The first is caused by charge displacement which can occur in both polar and nonpolar molecules and the second is caused by the orientation of polar molecules.[1] We first consider charge displacement in oxide type insulators.

In this case, it is likely the tight binding approximation will be applicable and for these considerations the electronic energy levels will not depart significantly from those of a hydrogen like atom. Thus for illustration and as a first cut at orders of magnitude we will look at the time scales associated with the natural motion of an electron in its ground state bound to a proton. In this situation, the Bohr-Sommerfeld quantization rules can be applied, specifically

$$\oint p dq = nh$$

where p and q are canonically conjugate variables. The two variables with which we are concerned here are the angular momentum and angular displacement.

$$\oint L d\theta = nh$$

Noting that the ground state is circular, from elementary mechanics, we have

$$mr^2\omega = n\hbar$$

where ω is the angular frequency of revolution of the electron about the nucleus. The time of revolution τ, is given by $2\pi/\omega$, and using the relation between the centripetal acceleration and the interaction force between the electron and the nucleus, one finds

$$\tau = \frac{2\pi\hbar^3}{me^4} = 1.547 \times 10^{-16} \ s$$

This is clearly well below any of the characteristic transient times we will encounter. Further, there is little likelihood, even in relatively complex solids, that this order of magnitude estimate could be increased by more

[1]We note that this discussion gives us an indication from whence the term "displacement current" comes.

than one or two orders of magnitude. This does not mean that that the material response times are on the order of τ. It does mean, in calculating or visualizing quantum mechanical perturbations, that an adiabatic approach can be used so long as the perturbation is not so strong as to produce quantum transitions. The latter case is treated in detail by Bohm [19]. The condition for no quantum transitions to occur is given by

$$\frac{h}{\left(\varepsilon_s^o - \varepsilon_n^o\right)^2}\frac{\partial U}{\partial t} << 1$$

where here the ε's are the energy levels between which a transition could occur. Again using the parameters from the hydrogen atom for order of magnitude estimates, an analysis following the lines presented above shows that the energy levels are given by

$$(\Delta\varepsilon)_{mn} = \varepsilon_o\left(\frac{1}{m^2} - \frac{1}{n^2}\right) \qquad \text{where} \qquad \varepsilon_o = 13.6\ eV$$

and m and n are integers. For a transition from the ground state to the first excited state, the change in energy is about $10\ eV$. For this transition, the Bohm condition becomes

$$\frac{\partial U}{\partial t} << 3.844 \times 10^4\ ergs/s$$

If the change in potential is caused by an electric field over a region of atomic dimensions, the restriction on the time rate of change of the electric field is

$$\frac{\partial E}{\partial t} << 8.008 \times 10^{21}\ S\,V/cm/s = 2.403 \times 10^{24}\ V/cm/s$$

At first sight this seems to be excessive. However, think of it in terms of a resonant condition and think of the Fourier transform of a linearly rising pulse that flattens out at some point in time. The energy content of the pulse is spread over a continuous range of frequencies. For a small range around a resonant frequency, the energy content will be correspondingly small.

Even for a nuclear electromagnetic pulse with a rise time of $1\ ns$ and peak gap fields of $10^7\ V/cm$ the time rate of change would only be on the order of $10^{16}\ V/cm/s$. It might appear that we could never get a transition. However, in quantum mechanics, one can never say never. The real question is a matter of how small is the probability in comparison to the number of opportunities one has available. We will return to this point in a later discussion.

II-4-4 Wave Packet Spreading

We wish to treat the electrons in the conduction band of either a metal or an insulator by methods of quasiclassical dynamics. The general theory was previously described and an example in terms of a particle in a constant force field given [20]. In the latter case, it was found that the product of the standard deviations in position and momentum space are given by

$$(\Delta x)(\Delta k) = \frac{1}{2}\sqrt{1+\left(\frac{2\hbar}{m}\right)^2(\Delta k)^4 t^2}$$

and that for a constant force, there was no spreading of the momentum wave packet. It is observed that the standard deviation product starts at the minimum value allowed by the uncertainty principle. For the standard deviation in position space as a function of time, we have

$$\Delta x = \frac{1}{2}\sqrt{\frac{1}{(\Delta k)^2}+\left(\frac{2\hbar}{m}\right)^2 t^2(\Delta k)^2}$$

For a specific time, this function will be a minimum if we pick the momentum standard deviation according to

$$\Delta k_{min} = \sqrt{\frac{m}{2\hbar t}}$$

The relationship between this value and the standard deviation of the position probability at time t is

$$(\Delta x)_t(\Delta k)_{min} = \frac{1}{\sqrt{2}}$$

The significance of these results for our present considerations, is that to treat the particles quasiclassically, there must not be substantial spreading of the wave packets during the time between collisions or along a path length on the order of a mean free path. For a metal, the latter two parameters were found in terms of the conductivity in Chapter 2.

$$\sigma = \frac{ne^2\tau_F}{m} \qquad \text{and} \qquad \Lambda_F = v_F\tau_F$$

We take copper as a representative metal because it is a simple metal with well established properties that are nearly free electron like. Assuming the parameters for copper,[2] we have

$$\tau_F = 24.5\ fs \qquad \text{and} \qquad \Lambda_F = 387.3\ \text{Å}$$

[2]See Appendix II-A-1.

which leads to

$$\Delta k_{min} = 0.0419\ \text{Å}^{-1} \qquad \text{and} \qquad (\Delta x)_{\tau_F} = 16.9\ \text{Å}$$

and we see that the conditions for a quasiclassical treatment are well satisfied.

Perhaps one of the most interesting numbers in this set is the value for the mean free path. Insulative particle coatings with thicknesses of this magnitude are probably feasible. *If there were no collisions in the insulator, the insulator would then be acting very much like a vacuum barrier which would lead to almost pure metallic quantum tunnelling.*

Section II-5

BULK CONDUCTION IN INSULATORS

II-5-1 Insulator Characteristics

In our analyses we will not assume the insulators are highly ordered. If we do not make this assumption, what do we give up in terms of our previous thinking and what can we retain? The most important thing we give up is crystallinity and along with it the reciprocal lattice and the crystal momentum form of Newton's second law. However, if the material shows a linear conductivity range at low fields, a conduction band may exist. If such a band does exist, because of the low conductivity of the material, we know that there is a substantial gap between the conduction band and the valence band.

Thermodynamics and the underlying statistical mechanics are very general. What we retain is the thermodynamic principles of our previous analyses. These principles will be our primary tools. In particular, the application of the Fermi-Dirac distribution function at the valence and conduction band edges is appropriate. Although out of context, with respect to which the the term is generally used, we shall continue to refer to the chemical potential as the Fermi energy.

We next wish to enquire as to where conduction charges come in an insulator. At low fields there will be electrons thermally excited into the conduction band with an equal number of holes in the valence band. Their number and mobilities will determine the electrical conductivity. In an impure material, there will likely be available states for donors near the bottom of the conduction band and acceptors near the top of the valence band. There will also be Tamm states. Since these are associated with the surface contacts, and since we are not considering these at this time, we will ignore the Tamm states.

We do not know a great deal about the parameters related to conduction in insulators. We will need to make some "educated" guesses. Our aim is to produce a model that makes physical sense, can act as a qualitative guide for further progress, and will yield more quantitative results when the appropriate parameters are given. The educated guesses for the starting model-parameters are given in Appendix II-A-1. To keep this first analysis as simple as possible, we make the assumption that the impurity states consist only of donors in the vicinity of the conduction band. This will be amended as future data dictates.

II-5-2 Zener Tunnelling

In the past, we have considered Zener tunnelling as a possible mechanism for producing conduction charge carriers in excess of those under low field

conditions. That is, the applied field is of sufficient strength to directly lift an electron from the valence band into the conduction band. We will now argue, that is an unlikely event.

In our previous discussions, the gap fields predicted for metallic quantum tunnelling have been on the order of $10^7 V/cm$. However empirically, we know the fields are, at least, an order of magnitude lower.[1] For computational purposes we will pick $2.5 \times 10^6 V/cm$. The valence electrons will be relatively tightly bound and will be in core states with dimensions on the order of 1 Å. The corresponding electric fields to which the valence electrons will be subjected by the remainder of the electrons and the constituent nuclei will be on the order of 4.8×10^6 *statvolt/cm* or $1.44 \times 10^9 V/cm$. This is far in excess of the applied fields we expect to see in our surge arresting materials.

Another way of arriving at the same conclusion is to consider how far a valence electron must move in a given applied field to acquire the necessary energy to be moved into the conduction band. Assuming a band gap of 2.5 *eV*, the required distance is found to be about 100 Å. That's about twenty interatomic spacings, again, not very likely. *We thus conclude that Zener tunnelling from the valence band to the conduction band will not be a substantial contributor to charge transport in an insulator and that we should focus our attention on other conduction mechanisms.*[2]

II-5-3 Impurity State Parameters

We do not know anything about the nature of the impurities in our insulators. However, it turns out that we can obtain some semiquantitative results on the basis of a very simple model. We can consider the donor atom as a hydrogen like structure consisting of an ion core and a single orbiting electron. To get something related to the dynamics in our insulator, in the results for the hydrogen atom, we simply use the effective electron mass of the electron at the band edge and the dielectric constant of the insulator. For the ionization energy and radius of the first Bohr orbit we then obtain [13]

$$\varepsilon_i = \left(\frac{13.6}{\varepsilon_r^2} \frac{m_e}{m} \right) eV \qquad and \qquad a_d = \frac{0.53 \varepsilon_r}{m_e/m} \text{ Å}$$

where ε_r is the relative permittivity of the insulator. Using the parameters in Appendix 1, we find

$$\varepsilon_i = 251 \ meV \qquad and \qquad a_d = 9.58 \text{ Å}$$

---

[1] At least partly caused by the Schottky effect..

[2] However, see related comments in the next subsection.

The field strength at the first Bohr radius is found to be $5.23 \times 10^6\ V/cm$ and the distance to impart the ionization energy is $10\ Å$. Thus, although marginal, there is a definite possibility for Zener tunnelling from the donor level to the conduction band. Whether or not this actually occurs depends on factors that will be subsequently addressed. However, this type of tunnelling is a difficult problem. If it turns out that we need to address it, it will be done in a later task.

In the next two sections, we will describe the thermostatistics of insulators, first with no impurities, and then with impurities. Recognizing that the major distinction between an insulator and a conductor is the magnitude of the valence to conduction band gap, we will use silicon as an archetypal example. Even though silicon[3] is a semiconductor, we have chosen it because there is extensive empirical data available. Thus we can use this as a gage of the accuracy of the calculations we make prior to moving to the relatively unknown situation in a true insulator.

II-5-4 Thermostatistics With No Impurities

As stated above, the main distinction between a semiconductor and an insulator is that the insulator has a larger band gap than the semiconductor. The equilibrium thermal statistics of the two are essentially the same. Under normal circumstances ($\varepsilon - \varepsilon_F >> k_B T$), and the concentration of intrinsic electrons is given by [13]

$$n_e = 2\left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} exp\left[(\varepsilon_F - \varepsilon_C)/k_B T\right]$$

where ε_C is the energy of the conduction band edge. In general, the determination of a chemical potential may involve a rather difficult calculation. However, in this case, recognizing that for a pure substance, the electron and hole concentrations must be equal, leads to the result

$$\varepsilon_F = \frac{1}{2}\varepsilon_g + \frac{3}{4}k_B T\, ln(m_h/m_e)$$

where ε_g is the difference in energy between the conduction and valence band edges. For silicon, we find

$$\varepsilon_F = 0.5673\ eV \qquad \text{and} \qquad n_e = 1.223 \times 10^9\ cm^{-3}$$

The density of occupied states is given by the product of the density of states function and the Fermi function. At the conduction band edge, the density of states function is rising quadratically and the Fermi function is falling exponentially. Thus the density of occupied states will be fairly sharply

[3]Only by definition.

peaked at a value slightly above the conduction band edge. The energy at which this occurs is found to be

$$\varepsilon_m = \varepsilon_C + \frac{k_B T}{2}$$

For silicon at room temperature, this value is *1.123 eV*. The velocity of electrons at this energy is given by

$$v_m = \sqrt{\frac{2(\varepsilon_m - \varepsilon_C)}{m_e}} = \sqrt{\frac{k_B T}{m_e}}$$

For the electrons under consideration the value is $1.3 \times 10^7\ cm/s$. Using the average of the electron and hole effective masses and the measured conductivity of silicon, this implies a mean collision time of *2.126 ps* and a mean free path of *2,330 Å*. Both the collision time and the mean free path are long.

There are two comments concerning this result. First, the sample with which we are dealing here is extremely pure and the crystal structure is highly ordered. Second, the electron velocities at the peak of the occupied density of states is about an order of magnitude below those found in metals at the Fermi energy at the same temperature.

How large would the band gap need to be to make silicon an insulator? A generally accepted dividing line is a conductivity of $10^{-13}\ mho/m$. Thus we will retain other parameters constant and determine the charge concentration that will lead to this conductivity. The required charge concentration is $0.318\ cm^{-3}$ which leads to a band gap of *2.16 eV* and a Fermi energy of *1.094 eV*. Note the smallness of the electron concentration for this level of conductivity, and this is not even a really good insulator. Recall that these calculations are based on a pure material. In a quality practical insulator, there may be donor and acceptor impurities. In order not to degrade the insulator, they must be small in number and the energy gap will need to be larger than calculated above. The energy gap in diamond is *5 eV* and we have selected *3 eV* as a typical value for use in our insulator model.

For our modified silicon model, if the energy gap were set at *3 eV*, the intrinsic electron concentration would be $4.682 \times 10^{-8}\ cm^{-3}$ and the conductivity would be $1.472 \times 10^{-20}\ mho/m$. If the measured conductivity were above this value, which it generally is for most insulators, one would then conclude that the conductivity of most insulators is dominated by impurity donors and acceptors.

II-5-5 Thermostatistics With Impurities

In this subsection, we assume that donor impurities dominate. In that case, the number of electrons in the conduction band is accurately

approximated by the number of ionized donors. On the average, the number of ionized donors will be given by the probability of an electron missing from a donor level. Mathematically, we have

$$n_e = N_D^+ \qquad \text{and} \qquad N_D^+ = N_D\left[1 - f(\varepsilon_D)\right]$$

where the pluses represent the ionized donors. Combining these with the equation for n_e on the previous page gives

$$2\left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} exp\left[(\varepsilon_F - \varepsilon_C)/k_B T\right] = N_D\left[1 + exp\left[(\varepsilon_F - \varepsilon_D)/k_B T\right]\right]^{-1}$$

which is a transcendental equation from which the Fermi energy may be determined. For silicon, the donor energy level predicted by our equation on p. II-5-3 is *25.83 meV*. The measured values depend on the type of impurity and range between *39* and *49 meV*, with an average of *44 meV*. It appears that the predicted value is too low by a factor of about *1.7*. A similar result is obtained for germanium. Thus we will modify the donor ionization equation accordingly and write

$$\varepsilon_i = \left(\frac{23.1}{\varepsilon_r^2}\frac{m_e}{m}\right) eV$$

retaining the form in terms of effective mass and relative dielectric constant. For our insulator model, we then have a donor ionization energy of *426.1 meV*. This is much larger than what is encountered in a typical semiconductor and is occasioned by the relatively lower dielectric constant. Note also that this goes in the right direction for a good insulator where the low field resistivity is likely to be dominated by the impurities.

For our insulator model with a donor concentration of $10^{16}\ cm^{-3}$ (about one part per million), we find a Fermi energy of *2.723 eV*, an electron density of $2.574 \times 10^{13}\ cm^{-3}$, a mean thermal velocity of $1.626 \times 10^{7}\ cm/s$, a collision time of $2.295 \times 10^{-27}\ s$, and a mean free path of $3.732 \times 10^{-20}\ cm$.

It is not likely that the last two numbers are totally realistic. However, the smallness of those numbers delivers a message. The message is this: For a material with parameters similar to our insulator model, you have two choices, either there is no extensive conduction band in which conduction can take place or, if there is one, the collision sites are so densely packed that their density accounts for the low value of conductivity observed. A corollary to that conclusion is, *any substantial bulk conduction through the insulator will ultimately destroy the insulator*. A further discussion will be given in the following section.

II-5-6 Dielectric Breakdown

In all cases observed, dielectric breakdown leads to permanent alteration of a device that depends on the bulk insulative property of a solid dielectric. This characteristic is ubiquitously demonstrated in capacitors. A major factor in this behavior is the fact that breakdown generally occurs at one "hot spot" and once initiated, overly large currents are passed through the hot spot leading to permanent damage to the capacitor. Once breakdown occurs, the operating characteristics of the capacitor are, at best, lowered from their specified values and generally, the capacitor is simply shorted.

The same type of behavior is observed in metal particle surge arresting materials, the likely causes are as follows. Prior to breakdown, all interparticle junctions in the material operate on a relatively equivalent basis. As in the capacitor, breakdown will occur through a highly localized preferential path. *This has been repeatedly observed experimentally.* Once established, this path dominates all others and draws an inordinately high amount of current to the near exclusion of the other paths, a condition not existent prior to breakdown. The local heat generated in the high current path will damage the dielectric in that path and is likely to melt the metal particles in the path, forming a connected metallic bridge, resulting in a permanent short.

In a perfectly ordered crystalline dielectric, the mobility of an electron in the conduction band would be extremely high because there would be no electron-phonon collisions. No thermal energy would be transferred to the ion core lattice. If the order is reduced, either by impurities or phonon distortions, the mobility will also be reduced but could remain relatively high because of the basic underlying order of the crystalline state.

In an amorphous dielectric, the situation is much worse. there is no intrinsic long range order. It is unlikely that there is a significant conduction band that extends throughout a macroscopic region. There will be high electron-lattice interaction. A large amount of the energy transferred to any mobile electron will be delivered to the lattice. For high current densities, this is a prescription for the destruction of the dielectric.

Accounting for these factors, it would appear that building a reliable surge arresting device that depends on bulk dielectric breakdown of any dielectric is highly unlikely and is particularly unlikely if the dielectric is amorphous, specifically, if it is polymeric.

A key to success in this whole matter is to constitute the surge arresting materials in such a way that, throughout the entire surge arresting event, there is no path through the material that is the dominant current carrying path to the exclusion of other paths. The current density throughout the device needs to be as uniform as possible. Additionally, there is an insidious statistical effect which is described in Appendix II-A-2.

All of this indicates the material must be physically highly uniform, and avalanche conditions are to be avoided. It should also be noted that these are related to conditions that lead to fast switching times and optimal energy handling capability.[4]

II-5-7 Relative Electron Densities

In Section II-1, there were clear indications that intermetallic gaps on the order of atomic dimensions are required to enter the pure quantum tunnelling regime. In what we have seen in the last subsection, it is equally clear we want little interaction between the tunnelling electrons and the insulator in the intermetallic gaps. We have indicated destruction of the insulator is most likely the result of electron avalanche. It is observed[5] that for a mica 1 mm thick, the breakdown strength is on the order of 500 kV/cm, while for a thickness of 0.1 mm, the breakdown strength is on the order of 1500 kV/cm. We believe the increase in breakdown strength is because the distance between electrodes is shorter in the latter case. There is less probability for interaction, and avalanche can not be initiated until higher field strengths are reached.

With gaps on the order of atomic dimensions, the probability of an interaction is certainly reduced, as born out by our results concerning mean free paths, apparently supported by the last paragraph. However, there is another factor involved that could well be of equal or greater importance. There is a distinct difference between bulk current in the metal particles and the tunnelling current across the intermetallic gaps. The current density in the metal is the product of the conduction electron charge density and their *drift* velocity. The electrons emitted from the metal into the intermetallic gap are entering the gap with velocities corresponding to the Fermi energy, and the current density is this velocity times the electron charge density. The drift velocities are on the order of cm/s, whereas the tunnelling velocities are on the order of 10^8 cm/s. From conservation of charge we must have

$$\rho_B v_B = \rho_T v_T \quad \text{or} \quad \frac{\rho_T}{\rho_B} = \frac{v_B}{v_T} \approx 10^{-8}$$

The charge density in the gaps is about eight orders of magnitude below that in the metal, with a significant reduction in the probability of interaction with the insulator.[6]

II-5-8 Summary and Conclusions

To our knowledge, all previous attempts at producing quantum tunnelling varistor devices used materials consisting of two phases only, bare metal

---

[4]See Dutcher [2].

[5]Smithsonian Physical Tables, 9th revised ed.

[6]A more refined analysis would predict a ratio of the order of 10^{-6}, still a large reduction.

particles embedded in a polymeric binder. Some limited success has been attained. However, manufacturing yields and energy handling capacities have been low. The typical binder of choice has been a silicone rubber. For mixing the metal particles into the binder, the crudest method used a rubber mill. In two instances of which we are aware, sophisticated mixing machines of the Braebender and Hake-Buehler types were used, with results no better than with the rubber mills.

With what is presented in this part of our effort, it seems apparent why results have not been better. The factors involved are as follows:

The binder and mixing methods did not permit sufficient metal loadings to attain a substantially quantum tunnelling material.

There was a high likelihood for continuous metal to metal contact from one electrode to another.

The dominant insulator conduction mechanism was avalanche in the binder resulting in leading edge overshoots and low energy capacity.

The materials had a fairly high degree of nonuniformity.

In the approach proposed for the present effort, these difficulties are averted in the following ways:

Our metal particles are coated with a thin insulative layer with a thickness on the order of atomic spacings.

The particles are loaded into our test cells in the presence of an epoxy binder and allowed to settle before curing, achieving a near close packed configuration with the particle coatings setting the tunnelling gap.

Dominant path failure is averted by limiting the amount of current through any individual path.

With these conditions true quantum tunnelling can be approached with little interaction between the tunnelling electrons and the insulative layer separating the metal particles.

Section II-6

ELECTRODYNAMICS AND THERMODYNAMICS

II-6-1 Introduction

In much of the preceding, we have been concerned solely with occurrences on a microscopic level, obtaining estimates of what happens at individual interparticle gaps. Our surge arresting materials will not be entirely homogeneous. As previously indicated, there will be statistical variations enhancing the probability of dominant path failure modes. This is a matter of material properties in the bulk. What we need is a macroscopic treatment. To effect this, we will develop a continuum mechanical model using parameters from the microscopic analyses.

We will subdivide the surge arresting material into domains, called control volumes, that are large with respect to the sizes of the particles but small with respect to the total volume of material. In particular, the control volumes will be sufficiently small so that the microscopic properties of a control volume can be considered constant with the statistics applied from one control volume to another. The field equations are then applied to each control volume by finite element analysis with the appropriate boundary conditions satisfied between all control volumes. The latter aspect is described in detail in Part III. The purpose of this section is to derive the field equations, to define the initial and boundary conditions and to delineate the parameters to which the statistics will be applied.

II-6-2 Electrodynamics

The fundamental equations of electrodynamics are those of Maxwell. In the presence of material media, they may be cast in the form

$$\nabla \cdot \boldsymbol{D} = \rho \qquad \nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \qquad \nabla \cdot \boldsymbol{B} = 0 \qquad \nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t}$$

We will be concerned with conductive materials where we will take the constituitive relations to be

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \qquad \boldsymbol{B} = \mu \boldsymbol{H} \qquad \boldsymbol{J} = \sigma \boldsymbol{E}$$

and tentatively assume ε and μ are the same for all control volumes and the electrical conductivity σ varies in a statistical manner from one control volume to another.

With a step input, if it were not for changes in material properties with time, the voltage across our varistor devices would rise to a steady state value. The time over which this occurs is on the order of half a nanosecond. It is after this initial transient that the statistical effects will

manifest themselves, and the time rates of change of the fields will be relatively slow. Under these conditions, the time derivatives in the Maxwell equations may be neglected and assumed to equal zero.

We are mainly concerned with energy transfers; these will not involve the magnetic field, since magnetic forces are perpendicular to particle motions, so we will concentrate on the electric field and current density. The governing relations are the continuity equation and the first and second Maxwell equations.

$$\nabla \cdot \boldsymbol{J} = 0 \qquad \nabla \cdot \boldsymbol{D} = \rho \qquad \nabla \times \boldsymbol{E} = 0$$

Assuming the nonconducting electrons are not polarized, the boundary conditions are

$$\boldsymbol{n} \cdot (\sigma_2 \boldsymbol{E}_2 - \sigma_1 \boldsymbol{E}_1) = 0 \qquad \boldsymbol{n} \cdot (\boldsymbol{E}_2 - \boldsymbol{E}_1) = \rho_s / \varepsilon_o \qquad \boldsymbol{n} \times (\boldsymbol{E}_2 - \boldsymbol{E}_1) = 0$$

where the first and third equations are the required boundary conditions, and the first two equations can be solved, yielding the value of the surface charge density in terms of either electric field

$$\rho_s = -\varepsilon_o \frac{\sigma_2 - \sigma_1}{\sigma_2} \boldsymbol{n} \cdot \boldsymbol{E}_1 = -\varepsilon_o \frac{\sigma_2 - \sigma_1}{\sigma_1} \boldsymbol{n} \cdot \boldsymbol{E}_2$$

Since the curl of $\boldsymbol{E}$ is zero, we can express $\boldsymbol{E}$ as the gradient of a scalar potential and our problem reduces to standard solutions of Laplace's equation,

$$\boldsymbol{E} = -\nabla V \qquad \nabla^2 V = 0$$

with boundary conditions

$$\sigma_2 (\nabla V_2)_n = \sigma_1 (\nabla V_1)_n \qquad (\nabla V_2)_t = (\nabla V_1)_t$$

Laplace's equation is elliptic; this means one obtains meaningful solutions for either Dirichlet, Neumann, or mixed boundary conditions. In our present case, the things over which we have control are the potentials of the electrodes. The first boundary condition above implies the normal component of the electric field at the edges of our material is zero. Thus we have a mixed boundary condition problem, Dirichlet at the top and bottom and Neumann on the edges, with the boundary values given by

$$V_{y=0} = 0 \qquad V_{y=y_o} = V_o \qquad \boldsymbol{n} \cdot (\nabla V)_{edge} = 0$$

where y_o is the interelectrode spacing and V_o is the potential of the upper electrode.

In stark contrast to dielectrics, there are no fringe fields, a result of the surface charge accumulations. In a good insulator, there are too few charges to supply the charge accumulations, thus the fringe fields. For a nonhomogeneous conductor, the directions and magnitudes of the tangential components at the outer surfaces may vary, but there is no component normal to the surface.

II-6-3 Thermodynamics

For the present our main interest is in the qualitative aspects of statistical effects on dominant path failure, to that end, we will consider the worse case scenario where the full threat energy is delivered prior to the time any significant heat diffusion can take place. In Phase II, for other reasons, heat diffusion will be considered. For the present, we will consider the internal heat source to be the Joule heating in the form

$$s(\boldsymbol{r},E,T,t) = \boldsymbol{J}(\boldsymbol{r},E,T,t)\cdot\boldsymbol{E}(\boldsymbol{r},t) = \sigma(\boldsymbol{r},E,T,t)E^2(\boldsymbol{r},t)$$

and consider no heat sinks. This equation is then coupled with the ones of the previous section to obtain a solution in an iterative fashion.

The rise in temperature within a control volume will be given by

$$\frac{\partial T}{\partial t} = \frac{1}{c}\frac{\partial Q}{\partial t} = \frac{1}{c}\int \boldsymbol{J}\cdot\boldsymbol{E}\,d\tau$$

where c is the heat capacity of the control volume, Q is the heat input and the integral is over the control volume.

II-6-4 Control Volume Conductivity

The conductivity will vary from one control volume to another. We will assume a control volume conductivity of the form

$$\sigma_i(\boldsymbol{r},E,t,\boldsymbol{q})$$

where $\boldsymbol{q}$ is the vector containing the statistical variables from the quantum analyses. We have left the temperature out of this equation, since we have previously shown temperature will not have a significant effect, at least, not until the destruct point.

II-6-5 Circuit Model and Initial Conditions

We will assume the varistor is excited from a step voltage source with a finite internal impedance. We will divide the time into small steps of size Δt. Using finite elements, we then solve the boundary problem for the potential determining the current density and energy density in each control volume through the relations

$$\boldsymbol{E} = -\nabla V \qquad \boldsymbol{J} = \sigma \boldsymbol{E} \qquad Q_d = \boldsymbol{J} \cdot \boldsymbol{E}\, \Delta t$$

where Q_d is the heat density generated within a control volume, which, in turn, determines the increase in temperature of each control volume in time Δt. The new temperature is used in the next time step and this process is continued until a path temperature reaches the point of path destruction.

II-6-6 Test of the Field Equation Solutions

In checking our finite element programs, it is desirable to have problems to which we know the answer. The simplest problem we can think of to test our solution to the field equations, yet having enough content to act as an effective test, is a cylinder of conductivity σ_1 embedded in a medium of infinite extent with conductivity σ_2 with a uniform asymptotic field. We will take the cylinder axis parallel to the z-axis and the asymptotic field parallel to the x-axis. With the electric field in the positive x-direction, at infinity the asymptotic field will be

$$\boldsymbol{E}_{2\infty} = \hat{\boldsymbol{x}} E \quad \text{with potential function} \quad V_{2\infty} = -Ex = -Er\cos\phi$$

From the geometry of the problem and the functional form of the cylindrical harmonics, one might readily guess a solution may be of the form

$$V_1 = Br\cos\phi \qquad \text{and} \qquad V_2 = Er\cos\phi + A\frac{1}{r}\cos\phi$$

Applying the boundary conditions, we find

$$V_1 = -E\frac{2\sigma_2}{\sigma_2+\sigma_1} r\cos\phi \quad \text{and} \quad V_2 = -E\left(1+\frac{\sigma_2-\sigma_1}{\sigma_2+\sigma_1}\frac{r_c^2}{r^2}\right) r\cos\phi$$

is, in fact, the solution. For the electric field components in cylindrical coordinates we have

$$E_{1r} = E\frac{2\sigma_2}{\sigma_2+\sigma_1}\cos\phi \qquad E_{1\phi} = -E\frac{2\sigma_2}{\sigma_2+\sigma_1}\sin\phi$$

$$E_{2r} = E\left(1-\frac{\sigma_2-\sigma_1}{\sigma_2+\sigma_1}\frac{r_c^2}{r^2}\right)\cos\phi \qquad E_{2\phi} = -E\left(1+\frac{\sigma_2-\sigma_1}{\sigma_2+\sigma_1}\frac{r_c^2}{r^2}\right)\sin\phi$$

with the rectangular components given by

$$E_x = E_r\cos\phi - E_\phi\sin\phi \quad \text{and} \quad E_y = E_r\sin\phi + E_\phi\cos\phi$$

From the difference in the normal components we find the surface charge density

$$\rho_s = \varepsilon \boldsymbol{n} \cdot (\boldsymbol{E}_2 - \boldsymbol{E}_1) = -2\varepsilon E \frac{\sigma_2 - \sigma_1}{\sigma_2 + \sigma_1} \cos\phi$$

In making the finite element test, Region 2 will, of course, be finite. To approach a result corresponding to the present problem, we must center Region 1 in Region 2 and make Region 1 small with respect to Region 2 and check to verify all constraints of the analytic solution are reasonably satisfied.

II-6-7 Test of the Nonlinearities and Temperature Dependence

In this subsection, we will kill two birds with one stone. In the first place, we will give a rudimentary presentation of the steps gone through to determine varistor requirements and secondly present a device meeting the requirements that is suitable for another test case for the finite element analyses.

Suppose we have a power matched transmitter operating at *15 kW rms* into a *50 Ω* load. The peak voltage will then be about *2.5 kV*. We might use this value to set the value of the reverse standoff voltage of a protective varistor. In addition, we will specify a reverse leakage of no more than *1 mA*. The transmitter is capable of withstanding a hit of *5 kV* and the maximum threat is a pulse with peak voltage of *80 kV*. To the varistor, the threat appears as a *40 kV* source in series with a *25 Ω* resistor[1], and the voltage across the transmitter input will be equal to that of the varistor. The peak current through the varistor will be *1.4 kA* with a peak power delivered to the varistor of *7 MW*. Summarizing in terms of varistor parameters, we have

$$V_r = 2.5\,kV \qquad I_r = 1\,mA \qquad I_p = 1.4\,kA \qquad V_C = 5\,kV \qquad P_p = 7\,MW$$

In Subsection II-1-6 the device current was given by

$$I = K_V\, V^2 e^{-\beta_V/V} \quad \text{with} \quad K_V = A\,K_E \left(\frac{1}{n\,g}\right)^2 \quad \text{and} \quad \beta_V = n\,g\,\beta_E$$

Using the design equations of Subsection II-1-12 we arrive at

$$K_V = 64 \qquad \beta_V = 6.8 \times 10^4 \qquad A = 1.0\,cm^2$$

The resultant varistor characteristic and load lines are shown in Figures II-6-1 and II-6-2. The time at peak current to reach a *100 °C* temperature is found to be *25.7 μs* for a varistor with a volume of one cubic centimeter. In this example, the number of particles per unit volume is $10^{12}\ cm^{-3}$ with diameters of *100 μm*. The interparticle gaps are *100 Å*.

---

[1]Thévenin's theorem.

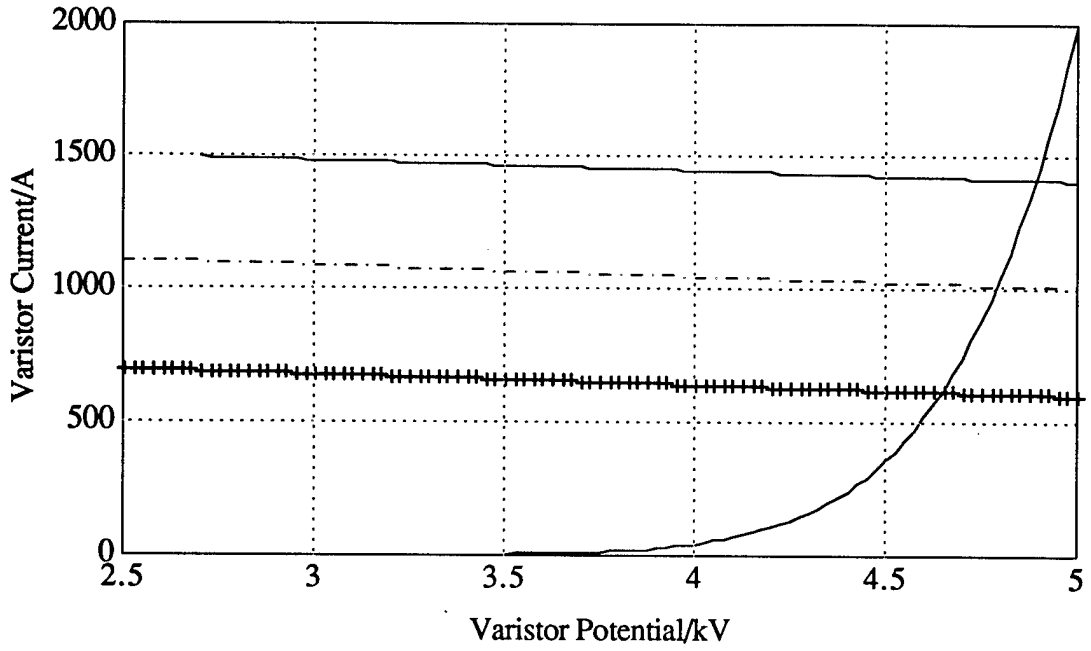


Fig. II-6-1 Linear plot of the varistor characteristic and associated load lines. The load lines shown are for threat voltages of 40, 60, and 80 kV.

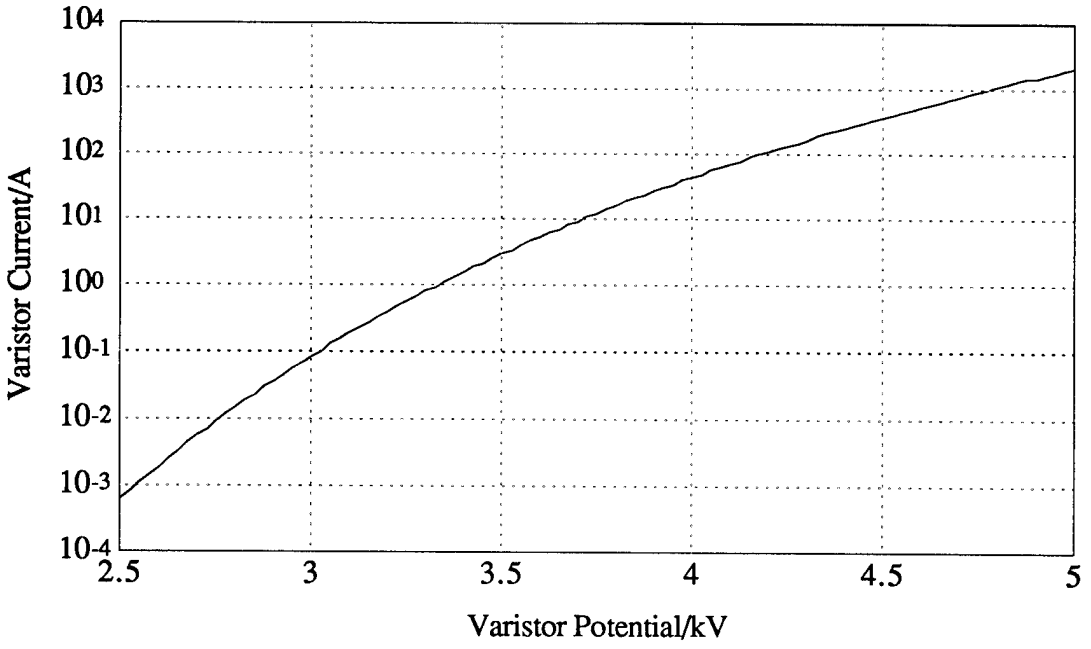


Fig. II-6-2 Log linear plot of the varistor characteristic.

On page II-1-22 we had the equation for the tunnelling current density in the Nordheim-Fowler approximation

$$J(E_g) = K_E E_g^2 e^{-\beta_E/E_g}$$

where here we have added the subscript g to the field to emphasize the field in this equation is the gap field. In moving to a finite element analysis, the field we will be computing with Laplace's equation is an apparent field which we will denote by E with no subscript.

Now consider an element of path Δl. There will be a number of particles along this path which we will denote by Δn and the number of gaps will be closely the same. If ΔV is the potential difference between the ends of the path element, the apparent field will be approximated by

$$E = -\frac{\Delta V}{\Delta l}$$

Within a finite element control volume we assume the parameters are uniform, so the potential across the path element may also be written as

$$\Delta V = -(\Delta n) g E_g$$

Combining the last two equations gives

$$E_g = \frac{E}{Ng}$$

where N *is the number of particles per unit length along the path element*, and approximate N by

$$N = (N_V)^{1/3}$$

where N_V is the number of particles per unit volume. The NF-equation is then given by

$$J(E) = K_E \left(\frac{E}{Ng}\right)^2 e^{-N g \beta_E/E} \quad \text{with conductivity} \quad \sigma = \frac{K_E}{(Ng)^2} E e^{-N g \beta_E/E}$$

The test of the finite element program is to verify that using the above values for conductivity, the temperature rises to the value previously given. Once verified, this gives reasonable assurance the finite element results will be accurate when g is statistically varied to determine dominant path effects.

In fabricating our materials, we will strive for as nearly a close packed structure as possible. Ideally all the particles would be touching one another with the gaps set by the insulative coatings. We will use a Gaussian random number generator, taking the absolute values of the numbers generated as additions to the nominal gap value.

Section II-7

DEFINITION OF STATISTICAL PARAMETERS

II-7-1 Evaluation of Bulk Conductivity

As indicated in the previous section, our continuum mechanical model will be based on approximate bulk conductivity. As usual, we are here mainly interested in the qualitative behavior of the materials, so we will use the Nordheim-Fowler approximation of Section II-1-9 and the parameters for copper. The predicted current densities will probably be higher than actual, but the general trends should be well illustrated.

Expressing the NF result in our standard notation we have

$$J = 6.1\times 10^{-6}\frac{\varepsilon_F^{1/2}}{(\phi+\varepsilon_F)\phi^{1/2}}E^2 exp\left(-6.8\times 10^7\,\phi^{3/2}/E\right)$$

with the conductivity being given by

$$\sigma = \frac{J}{E} = 6.1\times 10^{-6}\frac{\varepsilon_F^{1/2}}{(\phi+\varepsilon_F)\phi^{1/2}}E exp\left(-6.8\times 10^7\,\phi^{3/2}/E\right)$$

The E in this equation is the gap field. The number of gaps along a given path Δs is roughly given by

$$\Delta n = \frac{\Delta s}{D}$$

where D is the diameter of the particles. The gap E-fields along this path are then approximately

$$E = -\frac{D}{g}\frac{\Delta V}{\Delta s}$$

where g is the width of the gaps and ΔV is the electrostatic potential difference across the path.

II-7-2 Statistical Parameters

For each control volume, we are assuming the conductivity is constant. From the previous section we had for the i-th control volume

$$\sigma_i(\boldsymbol{r},E,t,\boldsymbol{q})$$

where $\boldsymbol{q}$ is a vector containing the statistical parameters. From the our approximate bulk conductivity equation we have

$$\boldsymbol{q} = (\varepsilon_F, \phi, D, g)$$

Each of these parameters will be given a Gaussian distribution about some nominal value, and it is expected that variations of ϕ and g will have the most dramatic effects.

II-7-3 Inclusion of Series Path Resistance

We have previously been concerned with the dominant path failure mode. A possible amelioration of this effect is to purposely introduce a resistance in all paths. One way of accomplishing this is to use metal coated insulative particles overcoated with an additional insulative layer to set the gap widths, where the metal coating is sufficiently thin so as to provide the additional path resistance. To model this we take the conductivity of the control volumes to be

$$\sigma_{li} = \frac{\sigma_{io}\sigma_i(\boldsymbol{r}, E, t, \boldsymbol{q})}{\sigma_{io} + \sigma_i(\boldsymbol{r}, E, t, \boldsymbol{q})}$$

where σ_{io} is the reciprocal of the additional path resistance and is an additional statistical parameter to be Gaussianly distributed. One of the goals of this effort is to find the approximate values of σ_{io} that quench dominant path failure.

II-7-4 Gap Statistics

Previous discussions have been couched in somewhat general form so they will be appropriate for future efforts. It is expected that variations in gap widths will have the most dramatic effects, so in Phase I we will concentrate on this parameter as far as statistical variation is concerned.

There are two contributions to the gap width, the thickness of the insulative coatings and the separation distance between the particles. The first we will denote by t and the second by s. The gap width is then given by[1]

$$g = t + s$$

We will assume the statistics on s are half Gaussian with a density function given by

$$p_s(s) = \frac{\sqrt{2}}{\sigma_s\sqrt{\pi}} exp\left[-\frac{1}{2}\left(\frac{s}{\sigma_s}\right)^2\right]$$

In fabricating our materials, we attempt to have all the particles touching one another. This density function concentrates probability at zero separation. A plot of the function with unit σ_s is shown in Figure II-7-1.

[1]The thickness here is the thickness of the coatings on two particles.

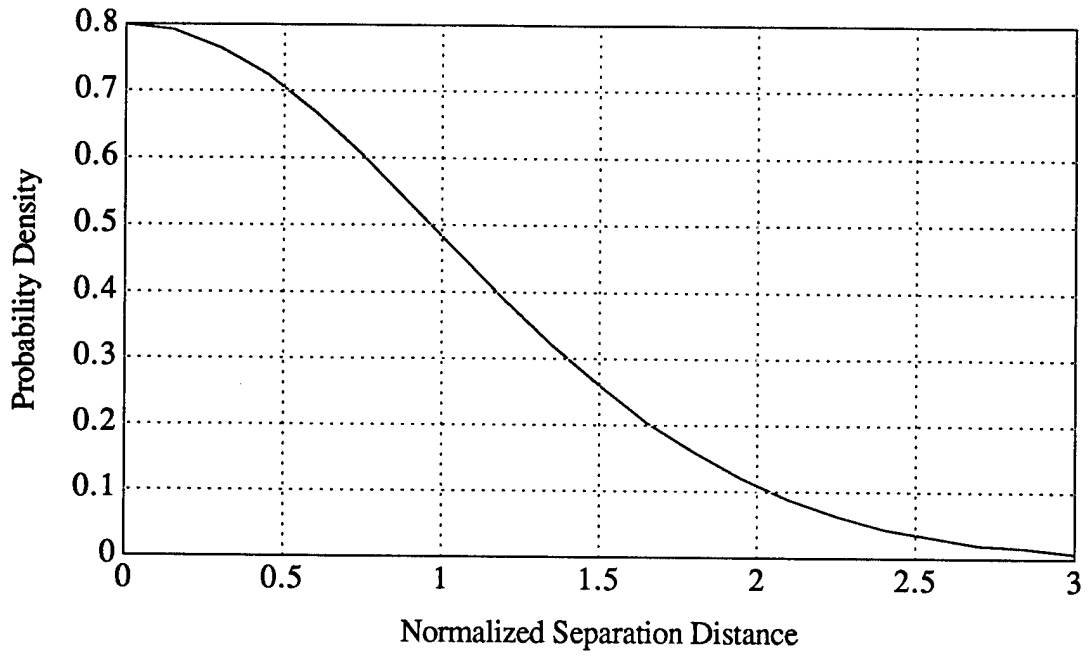


Fig. II-7-1 Half Gaussian probability density function with unit σ_s.

By straightforward integration, the mean value for this density function is found to be

$$\langle s \rangle = \sqrt{\frac{2}{\pi}} \int_0^\infty s \, exp\left[-\frac{1}{2}s^2\right] ds = \sqrt{\frac{2}{\pi}}$$

where in general we would have

$$\langle s \rangle = \sigma_s \sqrt{\frac{2}{\pi}} \approx 0.7979 \, \sigma_s$$

For the separation parameter, we again need a density function with a positive domain, but one which concentrates probability away from the origin. Such a density function is the Rayleigh density function given by

$$p_t(t) = \frac{1}{\sigma_t^2} t \, exp\left[-\frac{1}{2}\left(\frac{t}{\sigma_t}\right)^2\right]$$

This function with unit σ_t is shown in Figure II-7-2.

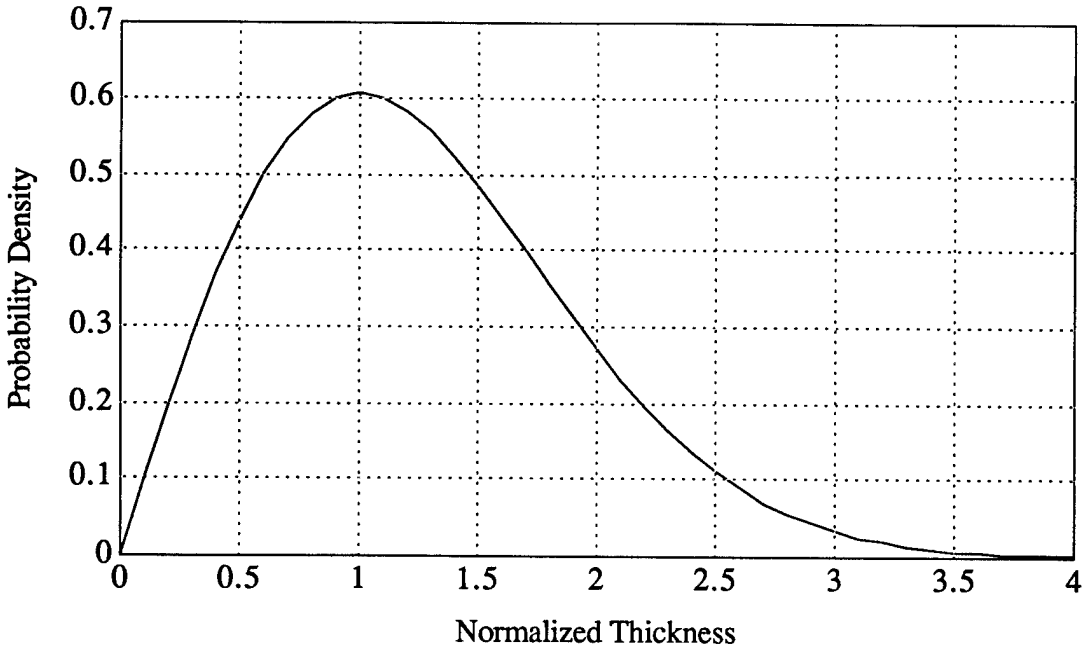


Fig. II-7-2 Rayleigh probability density function with unit σ_t.

Using the results of the integral for the standard deviation of a normal density, one finds for the mean of the Rayleigh density

$$\langle t \rangle = \frac{1}{\sigma_t^2} \int_0^\infty t^2 exp\left[-\frac{1}{2} \left(\frac{t}{\sigma_t} \right)^2 \right] dt = \sigma_t \sqrt{\frac{\pi}{2}} \equiv 1.253\, \sigma_t$$

The Rayleigh density function has a nice property, useful in our present circumstances. It can be generated from Gaussian processes. If x and y are two Gaussian random variables with zero means and standard deviations both equal to σ_t, the variable t given by

$$t = \sqrt{x^2 + y^2}$$

will be a random variable with a Rayleigh density function with parameter σ_t. Take note of the fact that the parameter σ_t in the Rayleigh density function is neither its mean nor its standard deviation. It is the standard deviation of the Gaussian density functions from which the Rayleigh variable is generated. The geometric significance of σ_t is that it is that value of t for which the Rayleigh density function is at its maximum.

Since the coating thickness and separation are statistically independent variables, the mean of the gap width will be

$$\langle g \rangle = 1.253\,\sigma_t + 0.7979\,\sigma_s$$

This expression can be used to explore the effects of the tradeoff between departures of the mean of t from its nominal value and departures of s from its nominal value under the statistical assumptions made.

II-7-5 Critique of the Preceding Procedure

Although the above procedure will give us some insight to the dominant path phenomenon, in both the half Gaussian and Rayleigh density functions described, there is only one parameter at our disposal. This means we can not independently set the mean of the functions and their standard deviations. Ultimately it will be desirable to do this, requiring a deeper analysis of the statistical situation. This must wait until Phase II. However, the deeper analysis will only affect the methods for generating the statistical parameters and will not impact the remainder of the computer programs.

APPENDIX II-A-1

Frequently Used Relations, Constants,

and Conversion Factors

FREQUENTLY USED RELATIONS, CONSTANTS, AND CONVERSION FACTORS

Relations

Free Electron Fermi Sphere	$\varepsilon_F = \frac{\hbar^2}{2m}\left(\frac{3\pi^2 N}{V}\right)^{2/3}$
Hydrogen atom energy levels	$E_n = 13.6\frac{1}{n^2}\ eV$
Density of states in k - space	$\frac{1}{\pi}$ one dimensional
	$\frac{1}{4\pi^3}$ three dimensional
Free electron density of states	$\mathcal{D}(\varepsilon) = \frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{\varepsilon}$

Physical Constants

Electronic charge	$e = 4.8\times10^{-10}\ esu$ $= 1.6\times10^{-19} C$
Electronic mass	$m = 9.11\times10^{-28}\ gm$
Planck's constant divided by two pi	$\hbar = 1.06\times10^{-27}\ erg\ s$
Boltzmann's constant	$k_B = 1.38\times10^{-16}\ erg / K^{-1}$
Avogadro's number	$N_o = 6.02\times10^{23}\ molecules / mole$
Nordheim – Fowler parameters	$K_{NF} = 6.118\times10^{-8}\ A/V$
	$\alpha_{NF} = 6.792\times10^{7}\ 1/cm\ V^{1/2}$

Conversion Factors

$1\ eV = 1.6\times10^{-12}\ erg$	$1\ C = 3\times10^{9}\ esu$
$1\ stat\ V = 300\ V$	$1\ mho / m = 9\times10^{9}\ s^{-1}$

$1\,J = 10^7\,ergs$ $1\,cal = 4.19\,J$

$1eV/molecule = 96.4\,kJ/mole$

Parameters for Copper

Density, ρ	$8.92\,gm/cm^3$
Electric conductivity, σ	$5.82 \times 10^7\,S/m$
Atomic number	29
Atomic mass	63.55
Crystal structure	Face Centered Cubic
Interatomic spacing	3.62 Å
Oxidation states	1, 2
Conduction electron density	$8.450 \times 10^{22}\,cm^{-3}$
Average work function, ϕ	4.64 eV
Free electron Fermi energy, ε_F	7.09 eV
Fermi sphere wave number, k_F	$1.36\,Å^{-1}$
Fermi sphere wavelength, λ_F	4.63 Å
Work function wavenumber, q_F	$1.10\,Å^{-1}$
Fermi sphere velocity, v_F	$1.58 \times 10^8\,cm/s$
Electronic collision time, τ_F	24.5 fs
Electronic mean free path, Λ_F	387.3 Å
Thermal conductivity, γ	$4.01\,W/cm\,°C$
Specific heat, c_p	$0.385\,J/gm\,°C$
Enthalpy of fusion, ΔH	208.7 J/gm
Melting point, t_m	1085 °C

Parameters for Aluminum

Density, ρ	$2.70\ gm/cm^3$
Conductivity, σ	$3.55 \times 10^7\ S/m$
Atomic number	13
Atomic mass	26.98
Crystal structure	Face Centered Cubic
Interatomic spacing	4.05 Å
Oxidation states	1, 3
Electronic density	$1.812 \times 10^{23}\ cm^{-3}$
Average work function, ϕ	4.28 eV
Free electron Fermi energy, ε_F	11.79 eV
Fermi sphere wave number, k_F	$1.75 Å^{-1}$
Fermi sphere wavelength, λ_F	3.59 Å
Work function wavenumber, q_F	$1.05\ Å^{-1}$
Fermi sphere velocity, v_F	$2.04 \times 10^8\ cm/s$
Thermal conductivity, γ	$2.37\ W/cm\ °C$
Specific heat, c_p	$0.897\ J/gm\ °C$
Enthalpy of fusion, ΔH	$397\ J/gm$
Melting point, t_m	660 °C

Parameters for Aluminum Nitride (AlN)

Density, ρ	$3.23\ gm/cm^3$
Thermal conductivity, γ	$165\ W/m^2\ °C$
Dielectric constant	8.8
Dielectric loss	0.001

Coefficient of expansion	*4.45 ppm/°C*

Parameters for Nickel

Density, ρ	$8.90\ gm/cm^3$
Electrical conductivity, σ	$1.16 \times 10^7\ S/m$
Atomic number	*28*
Atomic mass	*58.7*
Crystal structure	Face Centered Cubic
Interatomic spacing	*3.52 Å*
Oxidation states	*2, 3*
Average work function, ϕ	*5.19 eV*
Free electron Fermi energy, ε_F	*11.7 eV*
Fermi sphere wave number, k_F	$1.74 Å^{-1}$
Fermi sphere wavelength, λ_F	*3.61* Å
Work function wavenumber, q_F	$1.16\ Å^{-1}$
Thermal conductivity, γ	*0.907 W/cm °C*
Specific heat, c_p	*0.444 J/gm °C*
Enthalpy of fusion, ΔH	*298 J/gm*
Melting point, t_m	*1455 °C*

Silicon Dioxide

Energy gap	*8.0 eV*

Parameters for Intrinsic Silicon

Electrical conductivity, σ	$3.846 \times 10^{-4}\ S/m$ (@ *300 K*)
Band gap, ε_g	*1.11 eV*

Electron effective mass, m_e	$0.26\ m$
Hole effective mass, m_h	$0.5\ m$
Electron mobility, μ_e	$1350\ cm^2/V\,s$
Hole mobility, μ_h	$480\ cm^2/V\,s$
Donar ionization energy	$49\ meV$ (Arsenic)
Relative dielectric constant, ε_r	11.7

Parameters for Insulator Model

Density, ρ	$1.2\ gm/cm^3$
Electrical conductivity, σ	$10^{-14}\ S/m$
Direct band gap, ε_g	$3\ eV$
Electron effective mass, m_e	$0.166\ m$
Hole effective mass, m_h	$0.264\ m$
Relative dielectric constant, ε_r	3
Thermal conductivity, γ	$3.79\ mW/cm\,°C$
Specific heat, c_p	$1.34\ J/gm\,°C$

Parameters for Alumina (Al_2O_3)

Density, ρ	$3.7\ gm/cm^3$
Electrical conductivity, σ	$10^{-10}\ S/m$
Relative dielectric constant, ε_r	9.34
Thermal conductivity, γ	$0.30\ W/cm^2\,C°$
Specific heat, c_p	$0.7789\ J/gm\,°C$

APPENDIX II-A-2

Statistical Effects on the Failure Mechanisms of Quantum Tunnelling Varistor Devices

$\int\Psi$ *INTEGRATED SCIENCES*™
Technical Memorandum

STATISTICAL EFFECTS ON THE FAILURE MECHANISMS OF QUANTUM TUNNELLING VARISTOR DEVICES

Prepared for

Oryx Technology Corporation

Prepared by

Dr. Clinton Dutcher, NCE

September 1993

Introduction

Quantum tunnelling varistor devices depend on the highly nonlinear characteristic of current versus voltage at metallic interparticle gaps. The general theory for the gap characteristic has been previously described.[1] William Alston has voiced concern that variations in material parameters will lead to a large disparity in the magnitudes of the currents flowing through various paths in the material, providing "hot spots"that contribute to device failure.

An exact analysis of this situation is extremely difficult because of its geometry and physics. The purpose of this memorandum is to suggest a simplified approach that will give insight as to the nature of the problem and serve as an archetypal statistical model for future work. Results indicate that Mr. Alston's concerns are valid, and that control of material parameters that affect the standard deviations of gap betas is essential for reliable device performance.

Comments Concerning Aluminum

Theoretical quantum tunnelling calculations performed to date have been based on the free electron model of metals. For that reason, copper has been used as the archetypal example, since its Fermi surface is nearly

---

[1] Dutcher, Clinton, "Metallic Quantum Tunnelling Varistor Devices," United States Army Harry Diamond Laboratories, Contract No. DAAL02-91-C-0024 (1991).

spherical. Near term varistor materials are using aluminum as the metallic phase. Thus the parameters used in the following calculations will be those for aluminum and the following comments should be noted.

The Fermi surface of aluminum departs substantially from a sphere and is fairly complex. However the Fermi energy calculated on the basis of the free electron model is in good agreement with that calculated by more sophisticated means. In the current density equation, the Fermi energy appears only in the factor J_o.[2] Small variances in the Fermi energy will not lead to large variances in current.

The major source of variances in device current will come from variances in β_d, which will be influenced by variances in the interparticle gaps and the work function. The work function is an empirically derived datum and will be dependent upon particle coating and polymer interfacial chemistries. The value given is one for a chemically pure aluminum surface. Nonetheless, the parameters given here are the only available estimates and will be used until empirical evidence indicates otherwise.

The Independent Path Model

We will assume that current paths can be identified from one electrode to another such that no current in one path enters another. *That is the independent path assumption.* In that case, the current paths consist of metallic particles and their associated interparticle gaps where each carries current associated with that path only. For any path i the voltage across a device is then given by

$$V = \sum_{j=1}^{n_i} V_{ij}$$

where V_{ij} is the voltage across the j-th gap in path i, and n_i is the number of gaps in path i. The gap voltages are given by

$$V_{ij} = \frac{\beta_{ij}}{ln\left(I_g^{ij}/I_{ij}\right)}$$

From the conservation of charge and our independent path assumption, the current at any point in the path must be the same as all others. Thus the current is independent of j. The parameter in the numerator of the logarithm is dependent only on the gap effective area and J_o. Variations in this parameter will not have strong effects since it is contained in a logarithmic function. We will make the further assumption that it is equal for all gaps. We may then write.

---

[2] For notational definitions, refer to the summary of quantum tunnelling design equations.

$$V = \sum_{j=1}^{n_i} \frac{\beta_{ij}}{ln\left(I_g^i/I_i\right)} = \frac{1}{ln\left(I_g^i/I_i\right)} \sum_{j=1}^{n_i} \beta_{ij}$$

Defining the path β as

$$\beta_i = \sum_{j=1}^{n_i} \beta_{ij} \qquad \text{we have} \qquad V = \frac{\beta_i}{ln\left(I_g^i/I_i\right)} \qquad \text{and} \qquad I_i = I_g^i e^{-\frac{\beta_i}{V}}$$

Through our assumptions, we have now confined all of the statistics to β_i and those statistics can be handled by elementary means. Considering that β_i is the parameter that controls the major portion of the nonlinearity, one might expect that there is a reasonable degree of validity to the assumptions made.

The Statistical Model

We will treat β_i as a random variable. It is proportional to the product of g_{ij} and $(\phi_{ij})^{3/2}$ which are the fundamental underlying random variables. Without further evidence, the most logical assumption to make is that g_{ij} and ϕ_{ij} are normally distributed. However, that would considerably complicate our analysis. Instead, we will assume β_i is normally distributed and return to consideration of the g_{ij} - $(\phi_{ij})^{3/2}$ product at a later time.

The current through a device will be given by

$$I = \sum_{i=1}^{N} I_i$$

where N is the total number of paths. The current through any one path is statistically independent of the others, so the average current through the device will be the sum of the means of the individual paths.

$$\langle I \rangle = N \langle I_i \rangle$$

We have thus reduced the problem to finding the mean and standard deviation for I_i. To do that we need the probability density function for I_i.

Probability Density for I_i

The probability density for the current in terms of that for β_i is given by[3]

$$p_I(I_i) = p_\beta(\beta_i)\left|\frac{d\beta_i}{dI_i}\right| = \frac{V}{I_i} p_\beta\left(V ln \frac{I_g}{I_i}\right)$$

[3] A result that may be found in any intermediate statistics text.

where the density function for β_i is given by the Gaussian function

$$p_\beta(\beta_i) = \frac{1}{\sigma_\beta \sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{\beta_i - \beta_o}{\sigma_\beta} \right)^2 \right]$$

where β_o is the mean of β_i.

Numerical Example

Using the summary design equations, the properties of aluminum, and *ignoring* all that has been said above,[4] a device was designed with an electrode area of one square millimeter, a thickness of 4 mils, 10 micron metallic particles, and the following parameters

$$\beta = 7.25 \times 10^8 \ V/cm \qquad k_F/q_F = 1.667 \qquad J_o = 7.935 \times 10^{12}$$

$$g = 12.5 \ \text{Å} \qquad n_g = 10 \qquad \beta_d = 906.3$$

No attempt was made at optimization. Typical voltages versus current are

$$[(V/volts),(I/amps)] = [20, 1.66 \times 10^{-9}] \ [30, 6 \times 10^{-3}] \ [35, 0.45]$$

$$[40, 11.47] \ [45, 142.1] \ [50, 1.065 \times 10^3]$$

For such a device, we might say the operational voltage was about *20 V*, and the clamp voltage about *40 V*.

Our assumption in the past has been that β_i was constant and given by

$$\beta_i = g \, n_g \beta$$

For our statistical analysis, we will assume this is the mean value of β_i. We will assume there are one hundred particles along each edge of the device. Under our other assumptions, this then gives 10^4 paths and we will take the effective area as $10^{-6} \ cm^2$. All of this gives

$$I_g^i = 7.935 \times 10^6 \ A \qquad \text{and} \qquad \beta_o = 906.3 \ V$$

With a standard deviation 10% of the mean, or *90.63 V*, the probability density for β_i is shown in Fig. 1. With an applied voltage equal to *42.5 V*, the corresponding probability density for the current is shown in Fig. 2. It has a mean value of *0.0341 A*. The average current current through devices of this type would be *341 A* at *42.5 V*.

---

[4] What has always been done in the past.

This is a most unusual probability density requiring a plot on a logarithmic scale. The unusual character stems from the unusual nature of the quantum tunnelling IV characteristic. We will see that this has significant effects that *can not be ignored.*

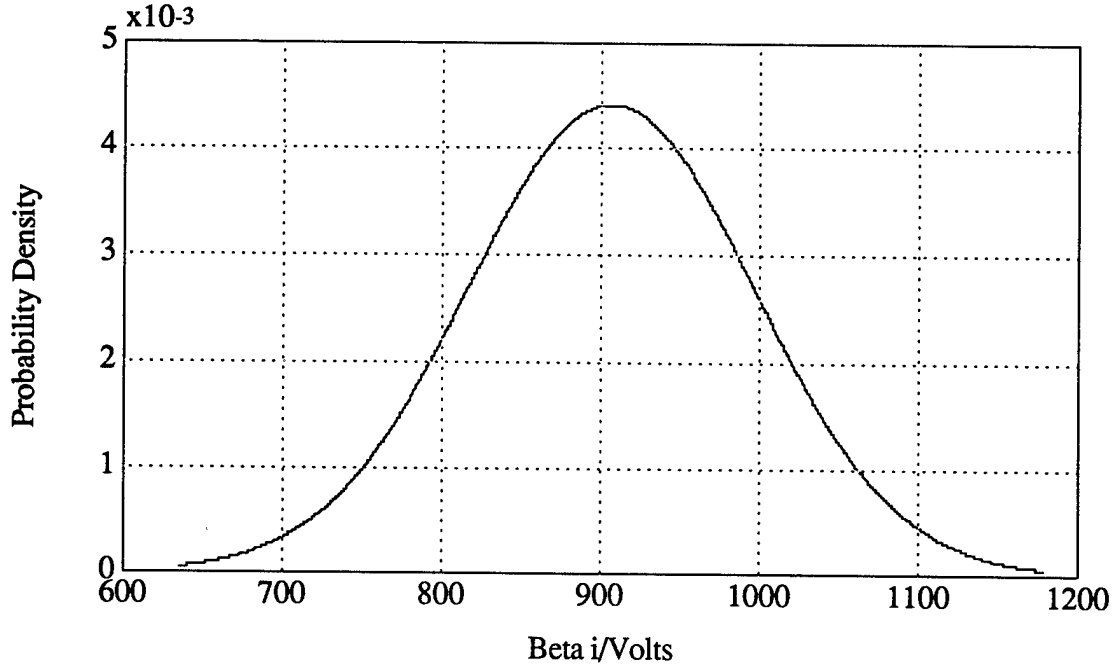


Fig. 1 Probability density assumed for path β with a mean of *906.3 V* and a standard deviation of *90.63 V.*

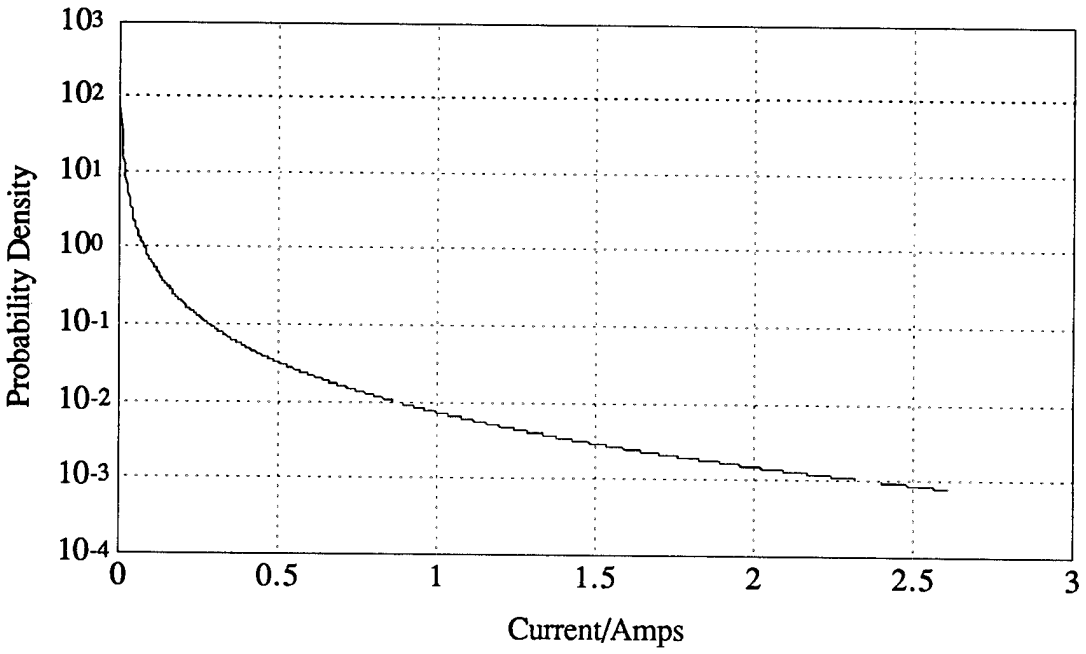


Fig. 2 Derived probability density for path current with *42.5 V* applied.

Thermal Considerations

In previous work,[5] an upper limit on energy handling capabilities was derived based on a simplistic model that ignored time frames and lumped the thermal properties of the metal and binder together. Here, we will make a first refinement of that model under the assumption that the surges with which we deal are sufficiently fast that no thermal diffusion takes place during the threat event and all of the surge energy is delivered uniformly to the metal particles. The end of the surge event then marks time zero for thermal diffusion.

In the previous analysis, it was assumed that thermal destruction of the binder was the major contributor to device failure and only a 100 C rise from room temperature was allowed. It is reported[6] that the silicone rubbers being used can withstand much higher temperatures for short times such as we are considering here.

Another potential failure mechanism is the melting of the metal particles, and this has been observed.[7] The melting point of aluminum is relatively low, *633 C*. For purposes of further illustration, we will set the maximum allowed temperature rise of the metal particles at the aluminum melting point but will limit the total energy so the heat of fusion is not supplied. Under these conditions, the particles would not actually melt. We will assume the path is destroyed if more energy is supplied.

As an example, we will use the device previously described in this paper and assume the volume fraction of metal is 70%. Since the particle diameters are 10 microns, 10^{-3} *cm*, we will assume the area of each path is 100 square microns, 10^{-6} cm^2. Each path is about 4 mils in length; we will take the path length as 10^{-2} *cm*, which gives the path volume as 10^{-8} cm^3. The energy absorbed by the metal for a given rise in temperature is

$$\Delta Q = f_m V_p \rho_A c_A \Delta T$$

where f_m is the volume fraction of metal, V_p is the path volume, ρ_A is the density of aluminum, c_A is the specific heat of aluminum, and ΔT the temperature rise. For a room temperature of *23 C* and a final temperature of *633 C*, the path energy is *15.3 μJ*. This is the maximum energy the path will sustain without being destroyed.

We now want to relate this energy to various types of pulses. To do that we use the equivalent time relation[8]

---

[5] Dutcher, Clinton, "Quantum Tunnelling Varistor Devices and Their Applications," IEEE International Electromagnetic Compatibility Symposium, Anaheim, California (1993).

[6] Jerry Behling.

[7] Karen Shrier.

[8] Dutcher, Clinton, "Equivalent Time in Varistor Applications," Integrated Sciences Technical Memorandum No. 26 (1993).

$$I_{max} = \frac{\Delta Q}{V_C t_{eq}}$$

where I_{max} is the current at the maximum energy, E_{max} allowed, V_C is the clamp voltage, and t_{eq} is the pulse equivalent time. We will assume the pulse energy is the *15.3 μJ* calculated above and a clamp voltage of *40 V*. For an IEC ESD pulse we have a maximum allowed current of *11.25 A* and for an *8* x *20*, *18.21 mA*. We also want to consider the Harry Diamond test system which has an equivalent time of *400 ns*. The corresponding maximum current is *956.3 mA*.

For a given device not to fail, *all* of its path currents must be below these values. That is where the path current statistics enter.

Statistical Interpretations

We are ultimately interested in predicting the failure rate of devices produced. That may be done in terms of the probability that a path will burn out in a single device and Bernoulli distributions. The Bernoulli distribution relates to an event that occurs with probability p and does not occur with probability $q = 1 - p$. Where n is the number of trials and x is the number of occurrences of the event, the probability of x occurrences in n trials is

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

The probability of failure of a given path is given by

$$q_B = \int_{I_B}^{\infty} p_I(I)\,dI$$

where I_B is the path burnout current. The probability that a given device will survive is equal to the probability that there are zero bath burnouts. From the Bernoulli distribution this is given by

$$p_d = (1 - q_B)^{n_p}$$

where n_p is the number of paths in the device. Out of N devices, the average number that will survive is given by the Bernoulli average

$$N_s = N p_d$$

These are interpretational relations to be used with the calculations that follow.

Computational Procedure

For all the computations that follow, we will assume a device with the characteristics previously described. We will then determine its power handling capability and survivability in terms of the standard deviation of β_i and three test waveforms. The waveforms chosen are the IEC ESD at *15 kV*, the Harry Diamond at *3.7 kV*, and the *8* x *20* at *500 W*. For each of these the required device peak currents are: IEC, *56.25 A*; Harry Diamond, *74 A*; and *8* x *20*, *11.9 A*. For each of these, three different standard deviations were used. As a percentage of the mean of *906.3 V*, they are: *10*, *5* and *1*.

The computational procedure is as follows:

Pick a test waveform. That determines the required device current at the clamp voltage.

Pick a standard deviation. That determines the probability density for β_i and and the form of the probability density for a path current as a function of applied voltage.

Calculate the average device current as a function of applied voltage and pick the voltage that gives the average current appropriate for the test waveform being considered. That determines the device clamp voltage and the probability density function for path current at the clamp voltage.

Calculate the path burnout current for this test waveform.[9]

Calculate the probability of burnout for a single path.

Calculate the probability of device failure.

For all of the computational steps, computer programs were written. In the next section we will give our observations on the probability densities in terms of graphs and the IEC waveform. Final results for all of the test waveforms are found in the conclusions section.

In the following observations on probability densities, all of the currents shown are single path currents. All average currents are at the IEC value of 56.25 A. Changes in standard deviation lead to the observed changes in clamp voltages.

[9] For the set of test wave forms under consideration here, we did this on the previous page.

Observations on Probability Densities

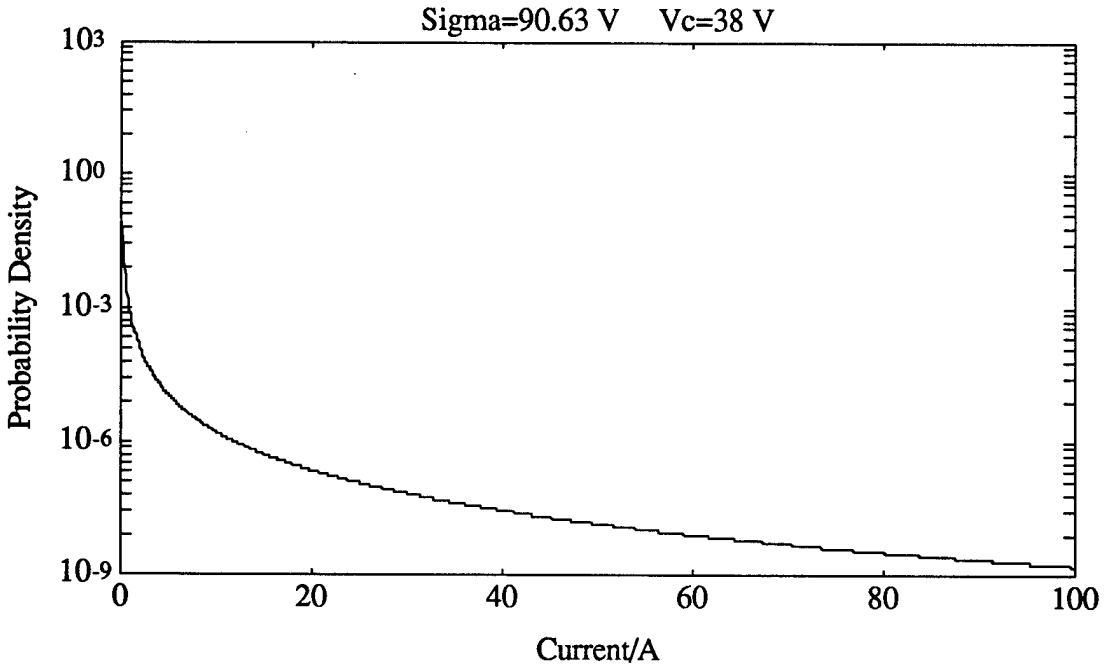


Fig. 3 On this scale, the true character of the function is not evident. It is valid for all currents to the right of the origin where the graph is visible, but on this scale, what is happening in the vicinity of the origin is obscured.

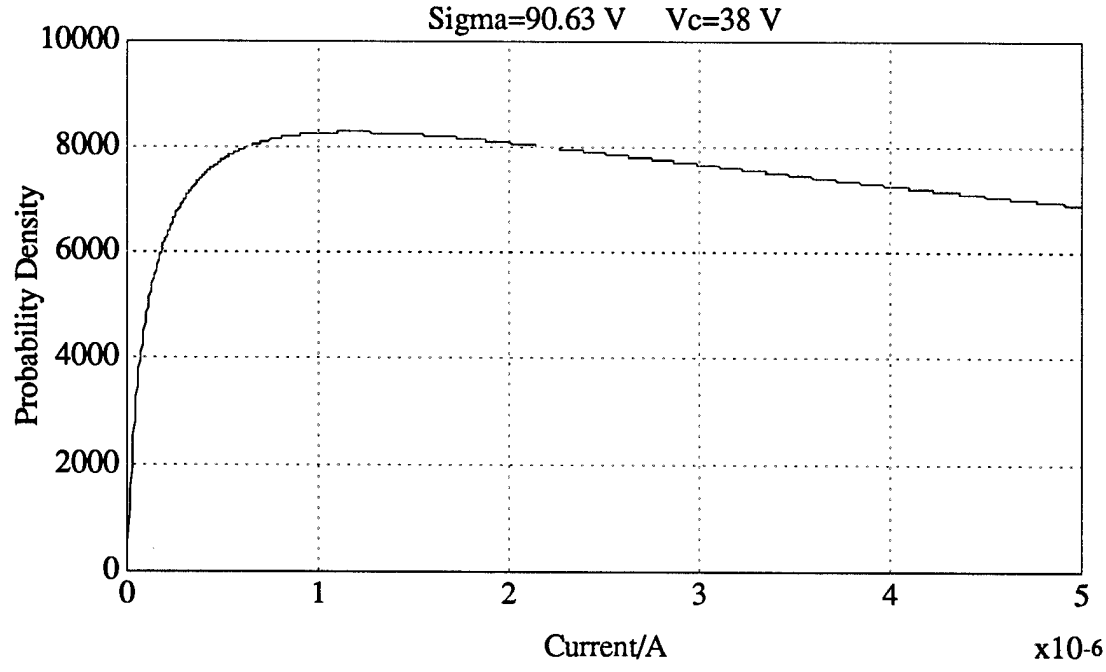


Fig. 4 An expanded view of the origin. It is now evident that there is a high concentration of probability in the vicinity of the origin. The function is quite broad and the standard deviation will be correspondingly large.

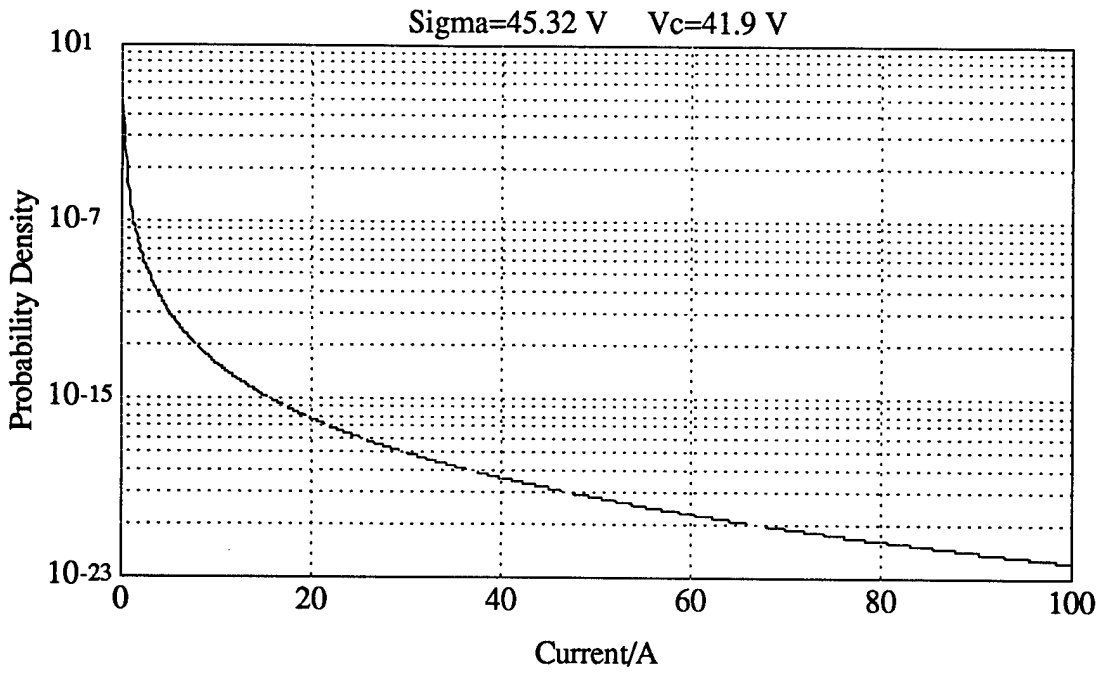


Fig. 5 Note the increase in the fall of the probability function for this standard deviation. More of the probability is now concentrated near the peak value.

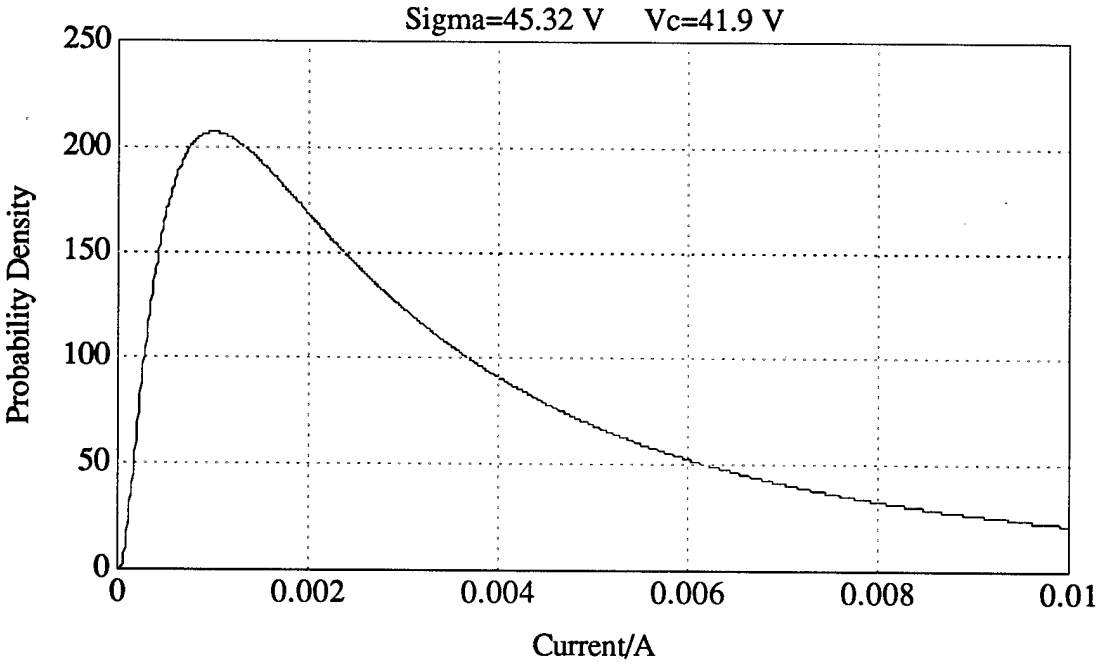


Fig. 6 The effect is even more evident here.

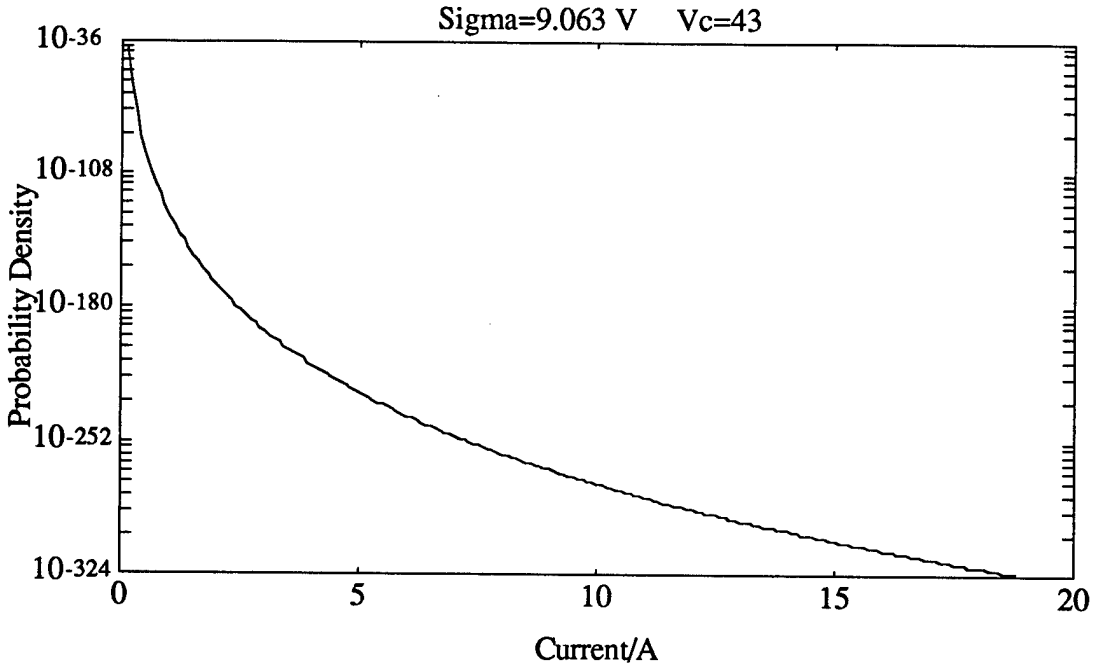


Fig. 8 It is now clear that as we decrease the standard deviation, the skewness of the probability density is reduced as well as its standard deviation.

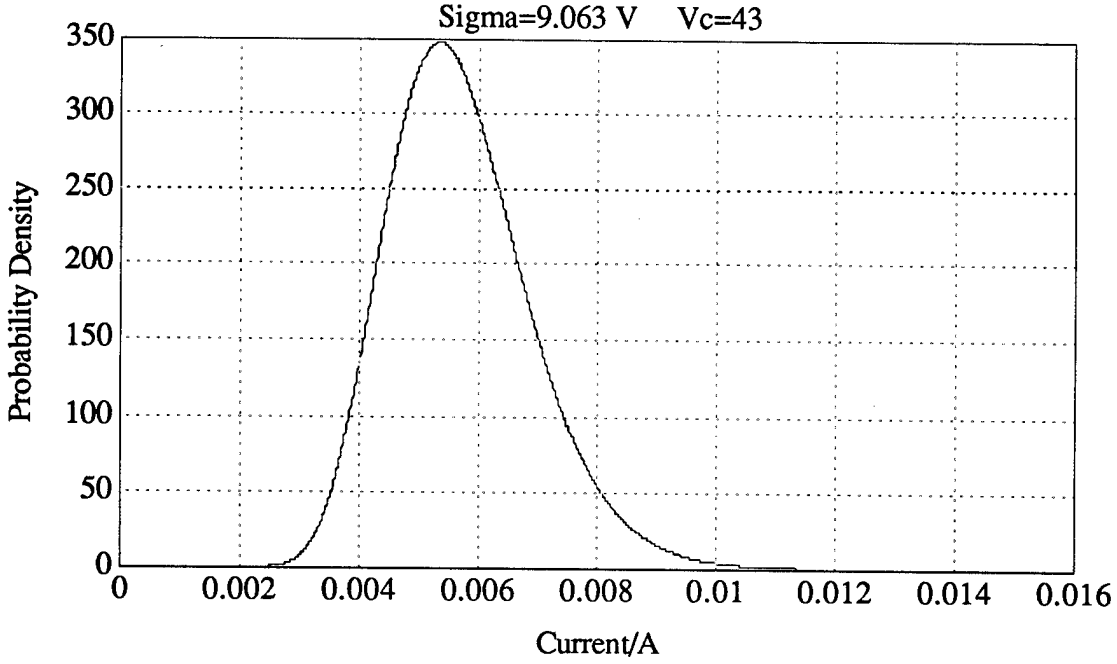


Fig. 9 This is more like what we are accustomed to seeing. Further, this is the first graph that discloses the full nature of the probability density, including the zero at the origin.

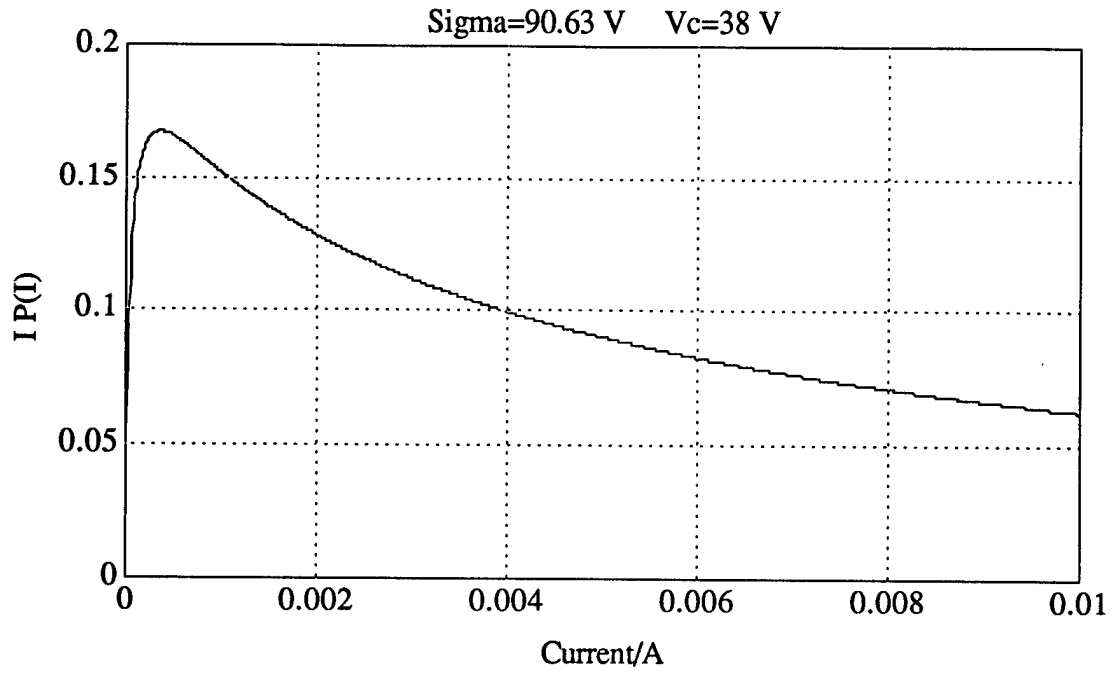


Fig. 10 Other curves of interest are the current-probability products, since their integrals give the average current. In this case, the peak is far from the average of 5.98 mA. The tail is long and rather flat, skewing the distribution toward high currents.

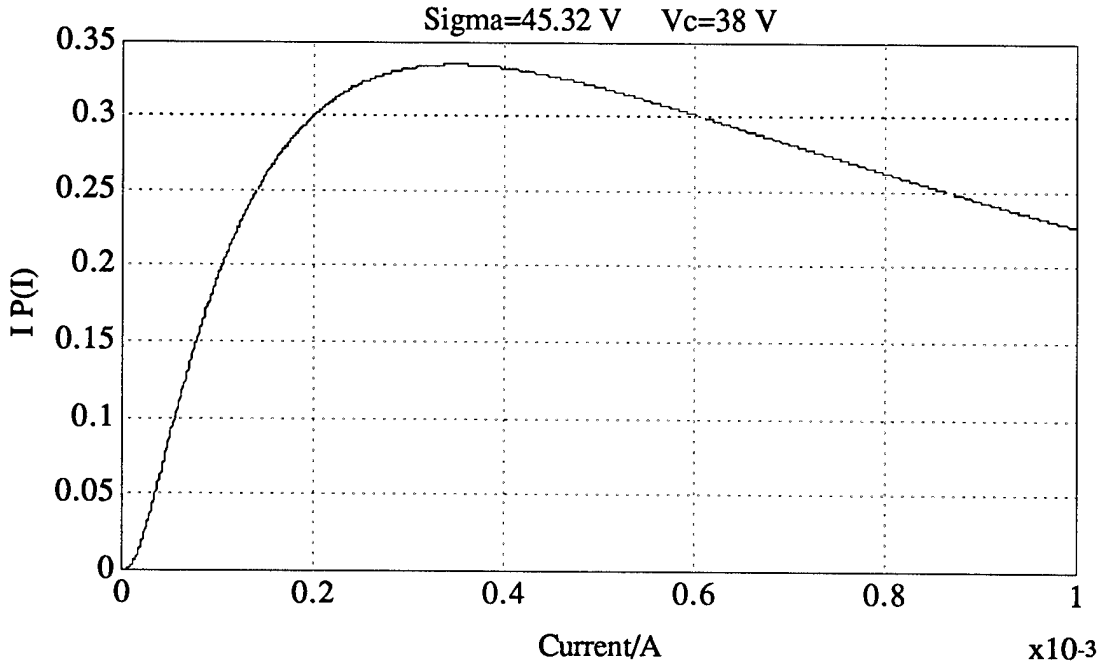


Fig. 11 Here we see the situation improving somewhat.

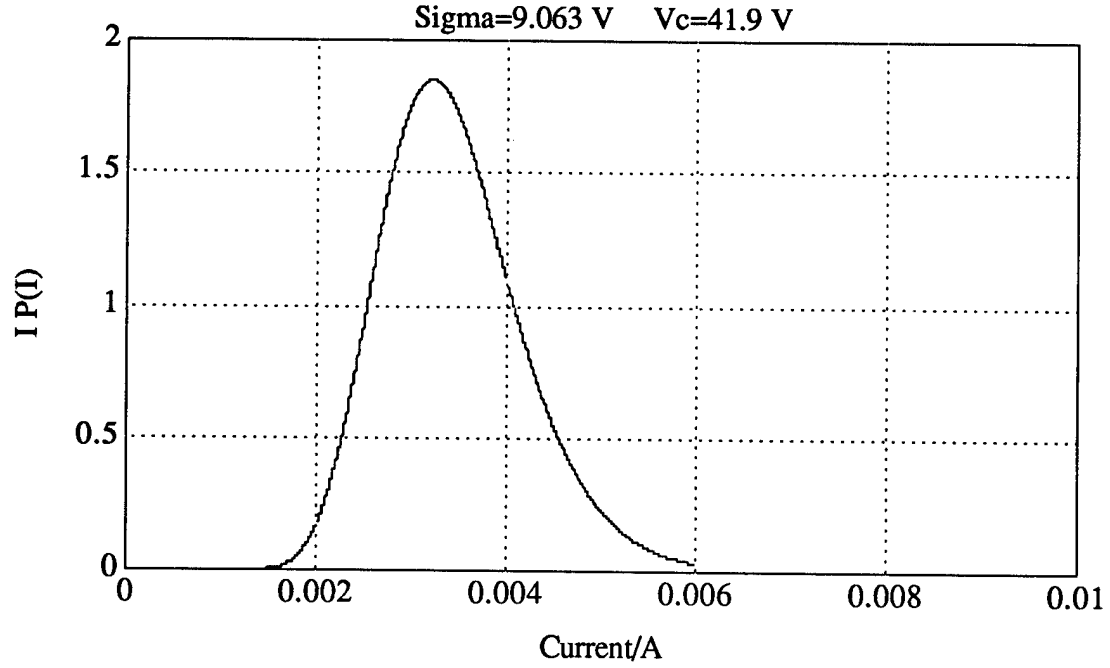


Fig. 12 This is still better; however we are still skewed with substantial probability beyond the peak and the peak is below the average current of 5.98 mA, indicating the skew is still toward the higher currents.

Conclusions

What we are observing here will be characteristic of any nonlinear multipath device that is suitable for surge protection.[10] It will even be true for junction diodes. However, the fabrication techniques of semiconductor devices has been so refined that it is likely the associated standard deviations are relatively small.

The results of our calculations are summarized in the table on the following page.

[10] In particular, MOV's, and may contribute to their unfavorable failure rates.

Test Waveform	Standard Deviation of β_i	Required Device Peak Current	Path Burnout Current	Device Clamp Voltage	Device Average Current at Clamp	Device Peak Power at Clamp	Predicted Proba-bility of Failure
IEC	10% 5% 1%	56.25 mA	11.25 A	38 V 41.9 V 43 V	59.8 A 57.5 A 56.9 A	2.3 kW 2.4 kW 2.4 kW	6×10^{-2} 3×10^{-10} 0
Harry Diamond	10% 5% 1%	74 A	956.3 mA	38.46 V 42.43 V 43.55 V	74.4 A 74.3 A 74.3 A	2.9 kW 3.2 kW 3.2 kW	.996 3×10^{-3} 0
8 x 20	5% 1%	11.9 A	18.21 mA	39.05 V 40.11 V	12.9 A 12.5 A	504 W 501 W	1 0

Table 1 Calculated results for hypothetical quantum tunnelling varistor device.

In all cases, the predicted probability of failure for standard deviations of 10% are unacceptable.

At 5% standard deviation and an IEC test waveform, the device looks good, an indication of why our materials play well in ESD applications. For the Harry Diamond test, the predicted probability of failure is marginal,[11] and for the *8 x 20 500 W*, the device is certain to fail.

With a 1% standard deviation, the device survives all of the test waveforms.

Multipath devices will survive best in circumstances where the equivalent times are short. That follows from the fact that the required device burnout current is inversely proportional to the equivalent time. The effect is reflected in the above table. This is also true of the clamp voltage. Multipath devices will be favored in applications with high clamp voltages and short equivalent times.

In the past, it was observed that clamp voltages were generally lower than those predicted by theory. This was considered as a microscopic parameter effect and was referred to as beta reduction. Since the effects of the statistics tend to decrease clamp voltages with respect to those based on a nonstatistical model, it is now apparent that the effect may also be

[11] Was I lucky?

influenced by the material statistics both on a macroscopic as well as a microscopic level.

In the past, it has been assumed that simply increasing the volume of material would increase the device power handling capabilities. The clamping voltage is related to the thickness of material between the electrodes and the size of the metallic particles. If we leave those fixed, the only way we can increase the volume is to increase the electrode area. This reduces the mean of the path probability density, which moves in the direction of reducing the probability of path burnout. However, it also increases the number of paths, which moves in the direction of reducing the probability of device survival. Increasing the area is likely to increase the power handling capability. If it does, it will not be as effective as previously thought.

We have brought the analysis to the point where a number of significant matters are now apparent. It was not carried further because the statistical model is critically dependent upon the gap IV characteristic, for which we do not have a good model in the absence of any estimates of avalanche effects.

Recommendations

A statistical model such as described here should be incorporated in the computer aided design package.

An analysis of the avalanche effects needs to be completed and incorporated with our knowledge of quantum tunnelling.

We need to complete our materials test system and use it to establish parameters that enter into the analytical model.

For test waveforms with long equivalent times, we need a macroscopic thermal analysis. It may be that there are some ameliorating effects on the path burnout current produced by heat flow. We need to quantify this.

In the light of results from the above, refine the present statistical analysis and bring it to the point that it is a useful predictive tool.

Once we have a better handle on the true gap IV characteristic, we also need to consider performance statistics as well as failure statistics.

The effects of increasing electrode area to obtain higher power handling capabilities needs investigation.

APPENDIX II-A-3

Original Problem Definition

Appendix II-A-3

ORIGINAL PROBLEM DEFINITION

Introduction

In the following table we compare some of characteristics of metallic tunnelling devices to those of conventional devices.

Table 1

METALLIC TUNNELLING DEVICES COMPARED TO CONVENTIONAL DEVICES

Metallic Tunnelling Devices	Conventional Devices
• Response times are dominated by metallic relaxation times (≈ 100 fs). Device time constants are independent of the gap width. There is no substantial heating in the gaps, so device reliability is relatively independent of gap width.	• Response times are dominated by semiconductor junction diffusion times. Device time constants are dependent on junction widths. To accommodate fast rise times, the junctions must be made small which compromises device reliability.
• Quantum tunnelling conduction is a monotonically increasing function of the applied voltage. There are no leading edge over-shoots for either fast or slow pulses.	• In some circumstances, conventional devices require a leading edge initiation. The slower the pulse rise time, the higher the energy content of the leading edge spike.
• Electron-phonon interactions in the gaps are minimal. The gaps are not the dominant heat sources.	• Electron-phonon interactions are dominant in the semiconductor junctions. The semiconductor junctions are the dominant heat sources.
• The dominant failure mechanism is melting of the metallic particles which requires a higher temperature than the failure of a semiconductor junction.	• The dominant failure mechanism is semiconductor junction burnout which requires a lower temperature than the melting of metallic particles.
• Heat generated in the metallic particles is generated in that constituent with the highest thermal conductivity. Heat dissipation is enhanced.	• Heat generated in the semiconductor junctions is generated in that constituent with the least thermal conductivity. Heat dissipation is impeded.
• Quantum tunnelling has no first order temperature dependence. Temperature stability is enhanced.	• Diffusion is exponentially dependent on temperature. Temperature stability is degraded.

Recently, materials have been developed ostensibly based on metallic tunnelling phenomena. As can be seen from the table and will be further illustrated in the following subsections, materials operating in the metallic tunnelling regime have characteristics that are entirely different from those of conventional type materials. These characteristics are well suited for applications involving surge suppression of high power pulses with subnanosecond risetimes.

However, no information has been found that implies that any of the presently available materials are truly operating in the metallic tunnelling regime. Although there are strong indications in the very early literature concerning this matter [1-11], it appears that the material manufacturers have not had a means to determine whether or not the tunnelling regime has been realized.

In the discussion of the application of varistors to the protection of electronic equipment, there are two important aspects that need to be considered. The first is how the material affects system operation under non-threat conditions, and the second is how the material behaves under threat conditions. For this proposal, it will be assumed that the varistor element is to be an integral part of the transmission line leading from a receiving antenna to the receiver electronics. Thus the varistor material will be interior to the transmission line and will act as a propagating medium with a given complex permittivity and complex permeability in the non-threat condition.

For production of suitable surge arresting devices, it is necessary to know the values of the complex permittivity and complex permeability and how they affect device performance. During the execution of the proposed project, guidelines will be given for the measurement of these parameters and their use in device design. The samples supplied will be configured so that these parameters are easily measured with currently available laboratory instrumentation. In this section, we will concentrate on those aspects that affect performance in the threat condition.

In the proposed project, we will either verify, or modify as necessary, an empirical quantum tunnelling signature indicated in the earlier literature. This activity will simultaneously provide guidelines for selecting material parameters that will lead to materials operating in the metallic tunnelling regime.

Using these guidelines we will fabricate prototype samples of varistor devices, subject them to laboratory tests, and compare the results to theoretical predictions, specifically, to the metallic tunnelling empirical signature. The following paragraphs summarize previous theoretical work that is part of the foundation on which we base our conclusions.

Qualitative Description and Assumptions

The classes of materials with which we will be concerned here are those that consist of small metal and semiconductor particles embedded in a polymeric matrix. In our analyses, we will always take the simplest physical model that is consistent with empirical data. For the particular class of materials in which we are specifically interested, if semiconducting particles are introduced, it is likely that their purpose will only be to provide some form of current limiting. The function of the polymeric binder is simply to hold the material together. The effects of the semiconductors, if needed, and the binder will be considered in the proposed effort, but in what follows, we wish to concentrate on the metallic tunnelling phenomenon *per se*. Thus, to simplify the discussion and

illustrate the general concepts and results, we will consider the quantum tunnelling of electrons through a vacuum gap, in planar geometry, from one piece of metal to another.

First Conduction Model

Our model will consider planar geometry with layers of metal separated by layers of dielectric, as shown in Fig. 1.

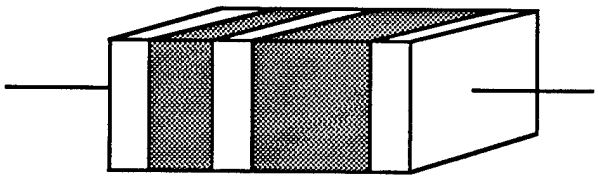


Fig. 1 Surge arresting device configured in planar geometry. The white layers represent metal and the shaded layers dielectric. Note that the thicknesses of the metal and dielectric layers are variable. In an actual device, there would be many more layers than depicted here.

With no applied potential difference across the end faces, a first approximation to the potential profile seen by a free electron interior to the device is that shown in Fig. 2.

In solid state physics, this model corresponds to the free electron theory of metals [12] where ε_F is the Fermi energy and ϕ is the metal work function. When there is an applied voltage the electrical stress across each of the metal layers is the same as for each of the other layers. This allows us to consider the situation associated with only one of the metal layers, leading to the picture shown in Fig. 3.

From Fig. 3, it appears that the electric field inside the metal is zero. This is not strictly true; however, the interior field is so small in comparison to the exterior field that it is not discernible on the scale depicted here.

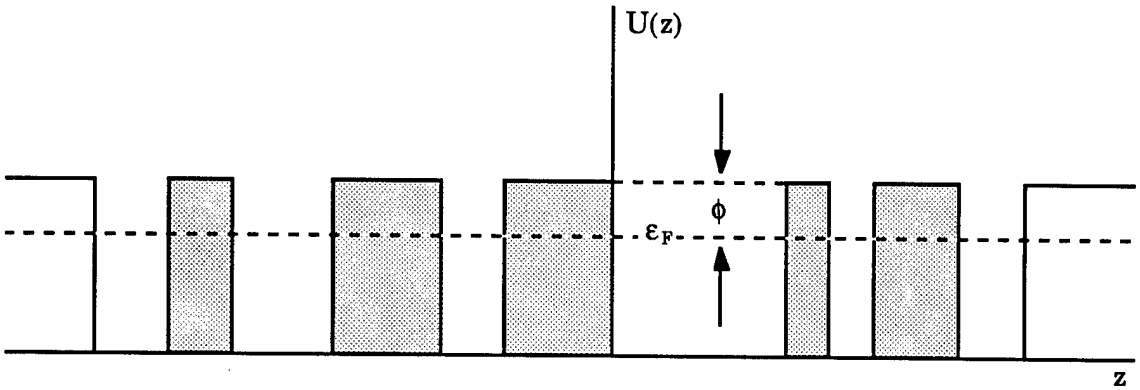


Fig. 2 Potential energy profile for a surge arresting device in a layered geometry with no applied field. The shaded areas indicate the presence of dielectric while the unshaded areas represent the presence of metal.

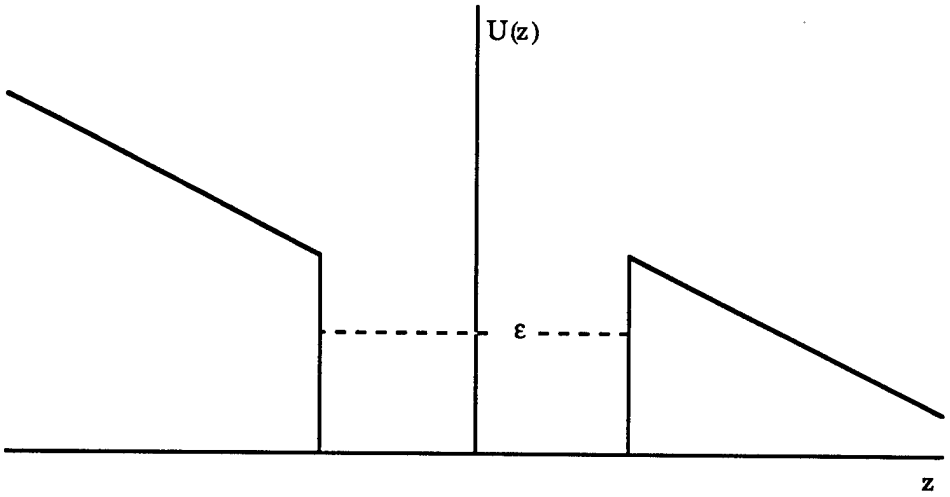


Fig. 3 Potential in the vicinity of a single metallic layer as modified by an applied electric field. The energy ε represents the energy of any specific electron under consideration.

In this first conduction model, we assume that the dielectric layers are so thin that they essentially act only as separators between the metallic layers and otherwise do not impede electron flow. We further assume that when an electron enters from the left, that it comes to thermal equilibrium with the rest of the electrons in the metal before it arrives at the right hand boundary. This implies that the right hand boundary is the major factor in impeding electron flow. Thus, it is at this boundary that we need to analyze tunnelling effects.

Stating the assumptions somewhat differently, the electron is accelerated in the dielectric regions and moves through the metal according to normal conduction transport. Since we are concerned about very short transients, the assumption concerning relaxation to electronic thermal equilibrium may come into question. Although there is some controversy concerning the exact electronic relaxation mechanisms in metals [18], it is generally accepted that the relaxation times are on the order of 100 fs. *This is a major reason for expecting metallic tunnelling varistors to have exceptionally fast risetimes.*

As we shall see, even for kA currents, the drift velocities of the current carrying electrons is extremely slow so that they spend times much longer than the electronic relaxation time interior to a metallic constituent. We also note that the electrons with which we are concerned must thermalize with respect to the other electrons, not the ionic lattice where the relaxations times may be much longer than the electronic relaxation times. The lattice relaxation times may become important when we investigate the thermal aspects that need consideration when analyzing failure mechanisms.

The other thing we note here is that since a large current density can be transported by electrons in metals moving at slow velocities, even for large current densities, the thermal equilibrium of the conduction electrons is not much disturbed. This is a key point in our treatment, and we will strengthen the argument when we later investigate the typical electron energies at the Fermi surface.

Since we have assumed ourselves into the position that the right hand boundary is the major impediment to conduction through our layered structure, we need to determine how to treat conduction at that boundary. First we note that even for the small particle sizes used in varistor materials, the particles are large in terms of things the size of electrons. Thus we will look at the details of matters close to the boundary and ignore what is going on far from the boundary. This is in keeping with our assumption of near thermal equilibrium and the fact that conduction electron correlation lengths in metals at normal temperatures are typically measured in terms of atomic spacings.

Near the boundary, the potential profile may be depicted as shown in Fig. 4. The width of the barrier at energy ε is defined to be 2a, where the factor of two is included to match later notation. The effect of increasing the field is to decrease the width but does not alter the height of the barrier. In the barrier region, the potential of an electron of charge e and a gap field E_g is given by

$$U(z) = U_b - eE_g z$$

Thus, the barrier width is given by

$$2a = \frac{U_b - \varepsilon}{eE_g}$$

A potential such as this is the one we would like to work with at this point. The resultant differential equation would have non-constant coefficients. However, the most important features of this potential profile are the barrier height and width. Thus we will replace the potential above with one that has the same height and width but is in the form of a pulse as shown in Fig. 5.

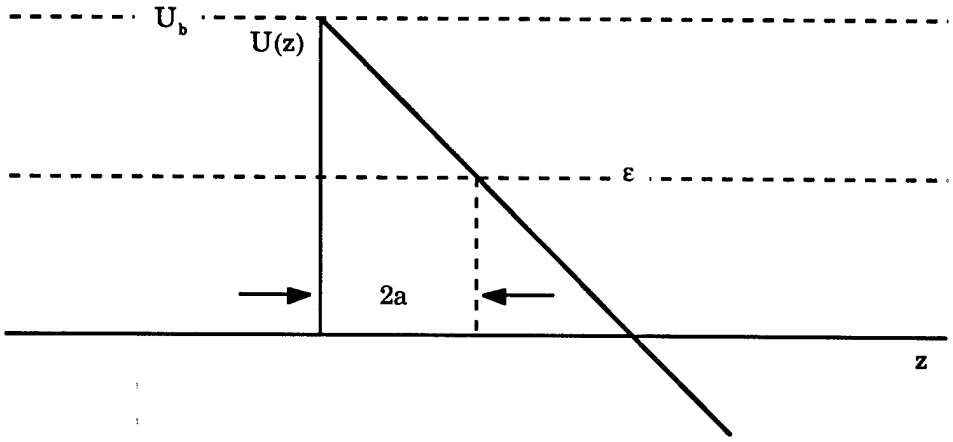


Fig. 4 Potential as seen at the right hand boundary of one of the metallic slabs. The width 2a will be related to our subsequent simplified model and the factor of two is included here for later notational convenience. U_b is the potential barrier height.

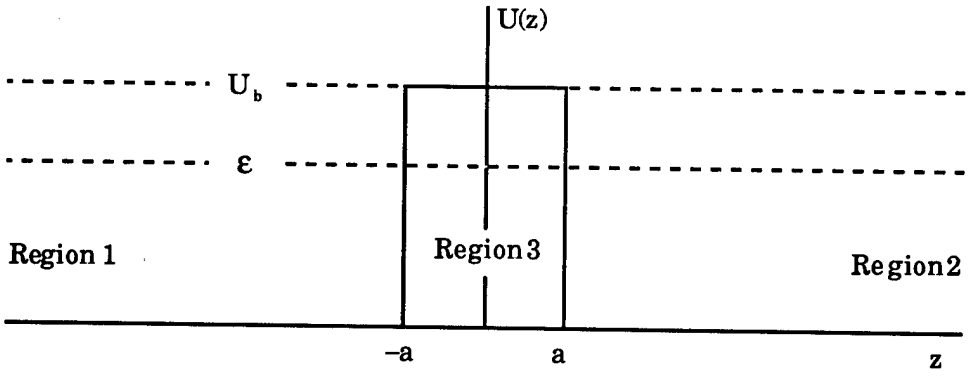


Fig. 5 Potential barrier with simpler analytic properties than those of Fig. 4. This potential profile will be referred to as the symmetric wall barrier (SWB).

It is expected that the qualitative aspects of the tunnelling phenomenon are relatively independent of the exact details of the potential barrier. In particular, we are concerned here with the gross functional dependence and intend to match parameters to empirical data. Thus for preliminary analysis, we will use the potential profile of Fig. 5.

Summarizing, our conduction model will be viewed in the following way. In zero applied field, the potential profile is that of Fig. 2. The conduction electrons in the metal are at thermal equilibrium and there are as many electrons travelling to the left as to the right. The net current flow is zero.

In the presence of an applied field, the potential profile is that of Fig. 3. If it is a modest field, so that no appreciable tunnelling takes place, the metal simply becomes polarized in such a way that the net field in the metal is zero. The polarization field that reduces the external applied field to zero in the interior of the metal is produced by the surface polarization charges that appear on the metal left and right surfaces. The electrons in the interior of the metal are again found in thermal equilibrium with no net current flow.

Next let us suppose that we increase the field strength so that some tunnelling takes place at the right hand boundary. We will assume here (and we will justify the assumption later) that the amount of tunnelling is not sufficient to appreciably disrupt the thermal equilibrium of the electrons interior to the metal.

Now here are the key points. Under the conditions described above, the number of electrons striking either the right or left boundary of the metal is given by the equilibrium statistics of the electrons interior to the metal. There is no net transmission to the left because the potential is rising there and, it only takes a flat potential to guarantee zero transmission (c.f., Wichmann [15, p. 284]. Since the slope of the potential profile on the left hand side is always negative, electrons readily enter the region of the metal where they come to thermal equilibrium with the rest of the conduction electrons. On the right hand side, they tunnel through the potential barrier. The net current is given by the product of the barrier transmission coefficient and the number of electrons per unit time striking the barrier,

this latter number being given by the equilibrium statistics of the electrons interior to the metal. To simplify the mathematics of the situation, we replace the potential profile of Fig. 4 with the one of Fig. 5 in such a way that it has the same width at energy ε and the same height as the profile of Fig. 4.

We shall refer to the profile of Fig. 5 as the symmetric wall barrier (SWB). In the next section, we will calculate the transmission coefficient for the SWB. This will subsequently be combined with the free electron theory of metals to obtain our first estimate of tunnelling conduction.

Particle in a Layered Potential

We first consider the case of a particle whose potential energy consists of layers where the potential in each layer is constant as described in the previous Subsection. The Schroedinger equation for such a particle is given by

$$-\frac{\hbar^2}{2m}\nabla^2\phi + U(\mathbf{r})\phi = \varepsilon\phi \quad \text{where}$$

$\hbar$ = Planck's constant divided by 2π, m = particle mass,

ϕ = wave function, ε = particle total energy, $\mathbf{r}$ = particle position vector

We will take the z-direction as perpendicular to the potential layers. Thus ϕ is a function of x, y, and z and we take U as a function of z alone. Assuming ϕ is a product of three factors, each of which is dependent on only one of the coordinates, the equation may be separated, the z-dependence being given by

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial z^2} + U(z)\psi = \varepsilon_z\psi \quad \text{where}$$

ε_z = the energy associated with the z-component of motion

ψ = the z-factor of the wave function product

It is this component of motion with which we will be concerned. What we have done here is to reduce the three dimensional problem for this layered geometry to a one dimensional case.

Before proceeding, we need to think a little more about the assumptions we are making. Assuming that the potential is a function of z alone is equivalent to assuming our sample is infinite in extent in the x and y directions. By separating the original equation the way we have, we have tacitly assumed that the particle energy is composed of three parts, one associated with the z-direction and two with the other degrees of freedom. With the potential assumed, the energies associated with each degree of freedom are noninteracting. Thus we have assumed that there is no scattering.

For the z-component of motion, the solution is a standard boundary value problem. We take the potential of Fig. 5. For each region shown, we have free particle wave functions. Assuming unity flux of particles from the left, the wave functions can be written in the forms

$$\psi_1 = e^{ikz} + Ae^{-ikz} \qquad \psi_2 = Be^{qz} + Ce^{-qz} \qquad \psi_3 = De^{ikz} \quad \text{with}$$

$$k = \frac{\sqrt{2m\varepsilon_z}}{\hbar} \qquad q = \frac{\sqrt{2m(U_b - \varepsilon_z)}}{\hbar}$$

The boundary conditions on the wave function are that it must be continuous and its derivative must be continuous. This leads to four simultaneous equations for the constants A, B, C, and D. In matrix form these are

$$\begin{bmatrix} e^{ika} & -e^{-qa} & -e^{qa} & 0 \\ 0 & e^{qa} & e^{-qa} & -e^{ika} \\ -ike^{ika} & -qe^{-qa} & qe^{qa} & 0 \\ 0 & qe^{qa} & -qe^{-qa} & -ike^{ika} \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} -e^{-ika} \\ 0 \\ -ike^{-ika} \\ 0 \end{bmatrix}$$

This system may be solved by determinants. The easiest way to proceed is to use row and column operations [16]. The transmission coefficient is the ratio of the particle flux in Region 3 to the incoming flux and is given by

$$T = \frac{(2kq)^2}{\left[(k^2 - q^2)\sinh(2qa)\right]^2 + \left[2kq\cosh(2qa)\right]^2}$$

First Transmission Estimate

As we shall see later, to get accurate results concerning the current flows in tunnelling devices, an integration of the product of the transmission function and the incident current densities is required. An estimate for the incident current density will be obtained in the following subsection. At this point we would like to obtain a first estimate for the transmission function.

For this purpose, we assume all of the incident electrons have energies equivalent to the nickel free electron Fermi sphere. For a gap field of 10^7 V/cm, the transmission function has a value of 1.7×10^{-52}. For 10^8 V/cm, the transmission function has a value of 2.0×10^{-5}. The major point to note here is that for a one order of magnitude change in applied voltage the transmission function has changed by forty seven orders of magnitude. We also discern that if any substantial current is to flow through the device, the current density incident upon the potential barrier must be correspondingly high. The fact that it is, is an important result of the following subsection.

Implications of the Free Electron Model

Our purpose in this subsection is to summarize some of the results of the free electron model of metals and to make a number of order of magnitude

calculations that support the assumptions we have made. Our main reference here is the text by Kittel [12].

The basic premise of the free electron model of metals is that for each electron in the metal, the potential that it sees is essentially of the type shown in Fig. 2. The corresponding wave functions can be represented as standing waves in terms of sines and cosines or, if periodic boundary conditions are used, in terms of free particle exponentials.

In either case, the energy levels are given by

$$\varepsilon = \frac{\hbar^2 k^2}{2m} \quad \text{where} \quad k = \text{the electron wave number, } m = \text{the electron mass}$$

The electron is a spin 1/2 particle and therefore obeys Fermi Dirac statistics and the Pauli exclusion principle applies. Thus, when filling the available energy levels with electrons, two electrons are permitted for each allowed value of k. At the absolute zero of temperature, the lower energy levels will be occupied first, with successive electrons being added at higher energies until all of the electrons are accommodated. The energy level at which the last electrons are placed is referred to as the Fermi energy which we have denoted by ε_F.

From the equation for ε above, we note that in the space of k, the various energy levels are represented by spheres and the sphere that corresponds to the Fermi energy is often called the Fermi sphere or Fermi surface. It is well known [12] that when the temperature is raised to the vicinity of room temperature, the distribution of the electron states in k-space is not much altered from the absolute zero form.

The density of states, the number of allowed states in the energy range between ε and $\varepsilon + d\varepsilon$, is given by [12]

$$dN(\varepsilon) = \frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\sqrt{\varepsilon}\, d\varepsilon \quad \text{where } V = \text{the sample volume}$$

This is basically all we need to know about the free electron theory of metals to pursue our present goals in regard to tunnelling conduction.

Electron Velocities

For the examples that follow, parameters appropriate for the metal nickel will be used. The parameters of other metals are of the same order of magnitude and for the present, we are mainly interested in order of magnitude estimates. The relation between the kinetic energy of a particle and its velocity is

$$v(\varepsilon) = \sqrt{\frac{2\varepsilon}{m}}$$

The Fermi energy of nickel is 11.7 eV. For an electron at the Fermi energy, this corresponds to a velocity of 2.02 Mm/s. We note that this is only two orders of magnitude below the velocity of light.

Conduction Electron Drift Velocities

It is an elementary physics problem to calculate the drift velocity required to yield a given current in a wire. For a one quarter inch diameter wire carrying 1 kA of current, the velocities are found to be on the order of 1 mm/s. Thus, even for a 1 kA current, the required drift velocities are miniscule in comparison to the velocities of the electrons at the Fermi surface.

This proves our contention that the equilibrium distribution of the velocities is not much disturbed by the presence of applied fields of the order of magnitude with which we are concerned.

Free Electron Current Densities

Let us consider a cube of metal and find the current density of the internal electrons that impinge on a face of the cube. If dJ_i is the current density caused by incident electrons with energies between ε and $\varepsilon+d\varepsilon$, we have

$$dJ_i = \frac{dN(\varepsilon)}{V} e\, v(\varepsilon) = \frac{2me}{\pi^2\hbar^3}\varepsilon\, d\varepsilon \quad \text{where} \quad e = \text{the electronic charge}$$

Since, at normal temperatures, there are few electrons outside the Fermi sphere, we have to good approximation,

$$J_i = \frac{2me}{\pi^2\hbar^3}\int_0^{\varepsilon_F} \varepsilon\, d\varepsilon = \frac{me\varepsilon_F^2}{\pi^2\hbar^3}$$

However, recall that the velocity that appears in the expression for kinetic energy has three components. Thus for this reason, the above expression should be reduced by the square root of three. In addition, for any velocity component, half of the particles are travelling away from the corresponding cube surface. Thus, the expression should be reduced by another factor of two. As a final result we find

$$J_i = \frac{me\varepsilon_F^2}{\pi^2\hbar^3 2\sqrt{3}}$$

When the parameters for nickel are substituted into this expression, we find J_i = 1.25 TA/cm^2. That is indeed a substantial current density!

When combined with the estimate for the transmission function found in the previous section, for a gap field of 10^7 V/cm, we obtain a current density through the device of 2.13 x 10^{-40} A, essentially zero. For a gap field of 10^8 V/cm, the current is 25 MA.

The Fundamental Conduction Equation

We have now come far enough with our model that we can very simply write the expression for the current density in the following form.

$$J = \int_0^{\varepsilon_F} J_i(\varepsilon) T(\varepsilon) d\varepsilon \quad \text{where}$$

$J_i(\varepsilon)$ = the metallic electron current density at a metal dielectric interface where the incident electrons have energy ε.

$T(\varepsilon)$ = interface potential barrier transmission coefficient.

This is our fundamental equation for the static case.

We next want to write this equation for the case where we assume the symmetric wall barrier and the free electron model. We then have

$$J = \frac{em}{\pi^2 \hbar^3 \sqrt{3}} \int_0^{\varepsilon_F} \varepsilon T(\varepsilon) d\varepsilon$$

In this type of application, the free electron model of metals generally serves as an excellent approximation. If future modifications of the theory are to take place, it is likely that they will be associated with the transmission coefficient $T(\varepsilon)$ or the inclusion of some other factor not presently taken into account. Thus, we will refer to the above equation as the fundamental conduction equation.

The estimates previously obtained for the expected current densities appear quite reasonable. However, no attempt has yet been made to carry out the indicated integration of the fundamental conduction equation. That will be the topic of the next subsection.

Reduction of the Conduction Equation to Integrable Form

It does not appear that an analytic integration is possible. For purposes of numerical integration, we recast the equations in the following manner.

$$J = K_0 \int_0^{\varepsilon_F} \varepsilon T(\varepsilon) d\varepsilon \quad \text{with} \quad K_0 = 1.833 \times 10^{10} \, A/cm^2 \, (eV)^2$$

$$T(\varepsilon) = \frac{(2k/q)^2}{\left\{\left[(k/q)^2 - 1\right]\sinh(2qa)\right\}^2 + \left\{(2k/q)\cosh(2qa)\right\}^2}$$

$$\frac{k}{q} = \left(\frac{\varepsilon}{U_b - \varepsilon}\right)^{1/2} \qquad 2qa = \frac{K_1}{E}(U_b - \varepsilon)^{3/2} \qquad K_1 = 5.094 \times 10^7 \, V/cm \, (eV)^{3/2}$$

Stating the units explicitly, the energies in the above equations are expressed in eV and the field strength in V/cm. The current density is in A/cm^2.

To obtain some insight into what the results will indicate, before performing the integration, we explored the properties of the integrand with the characteristic result that is typical of physical effects depending on

the electron dynamics in metals. That is, the major effects are contributed by electrons whose energies are close to the Fermi energy.

We found by direct calculation that there is no significant contribution from electrons with energies below about 10 eV. In future integrations, we will set the lower limit of integration to 8 eV. Further, at gap fields at and above 100 MV/cm, the current densities are so high that no material could sustain them. The latter means that the lowest significant value of the 2qa product is about 6.02. These are the keys to simplifying the expression for the current density.

Simplified Expression for the Current Density

At and above the value of 2qa equal to 6.023, we have the approximate relation

$$\sinh(2qa) \cong \cosh(2qa) \cong \frac{1}{2} e^{2qa}$$

With this given, the transmission coefficient can be written in the form

$$T(\varepsilon) = \left[\frac{4(k/q)}{(k/q)^2 + 1} \right]^2 e^{-4qa}$$

In the conduction integral, the variable of integration is the electron energy and the value of the integral is a function of the gap field E. Writing this explicitly, we have

$$J(E) = K_0 \int_0^{\varepsilon_F} \varepsilon \left\{ \frac{4[k(\varepsilon)/q(\varepsilon)]}{[k(\varepsilon)/q(\varepsilon)]^2 + 1} \right\}^2 e^{-4q(\varepsilon)a} \, d\varepsilon$$

Even with this simplification, it still does not appear that an analytic integration is possible. Thus, to get a global picture of the general behavior of the current density as a function of gap field, a computer program was written that evaluates the above integral for twenty points over the gap-field domain of 10 to 100 MV/cm. The result is shown in Fig. 6.

Because of the current levels involved, for surge arresting devices, the interesting gap-field domain is between 50 and 80 MV/cm. This domain is shown in both semi-log and linear forms in Fig. 7. For a change in the gap field by a factor of 1.6, the current density changes by five orders of magnitude.

Approximation to the Conduction Integral

For any given value of E, the mean value theorem of integral calculus [19] states that there is a value ε_m of ε such that

$$J(E) = K_0 \varepsilon_F \varepsilon_m T(\varepsilon_m)$$

In general, there will be a different value for ε_m for each value of E. However, recall that the main contributions to the conduction integral

occur over rather restricted domains of both ε and E. Thus, we might expect that the last equation would provide a practical approximation for the field dependence of the current density over the domain of E necessary for characterizing tunneling devices. In this case, the current density would be given by

$$J_a(E) = J_o e^{-\frac{\beta}{E}} \quad \text{where} \quad J_o = K_0 \varepsilon_m \varepsilon_F \left[\frac{4(k/q)}{(k/q)^2 + 1}\right]^2 \qquad \beta = 2K_1(U_b - \varepsilon_m)^{3/2}$$

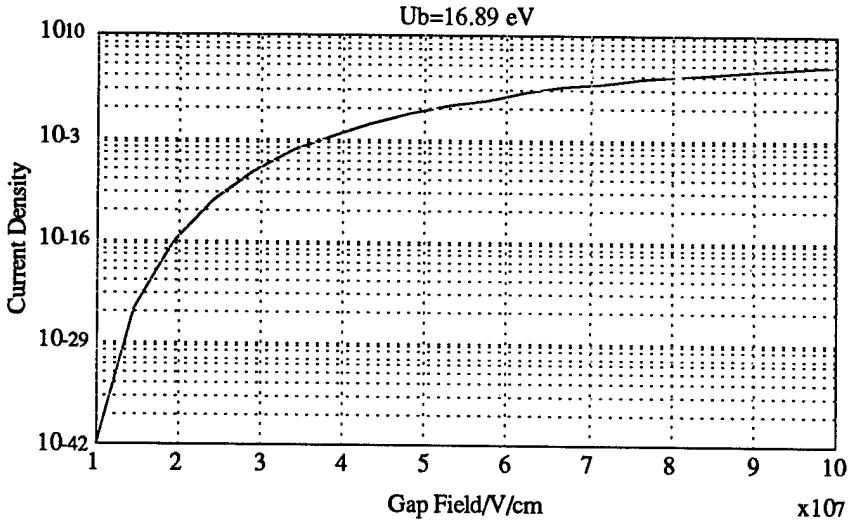


Fig. 6 Global picture of the behavior of the current density as a function of gap field.

and the k and q in J_o are evaluated for $\varepsilon = \varepsilon_m$. The subscript "a" indicates that this is an approximation.

Thus in characterizing tunnelling devices, we can regard J_o and β as parameters to be empirically determined, with the above expressions, at least, giving the correct orders of magnitude. In general, it is expected that the value of ε_m would be fairly close to the Fermi energy. A comparison of J_a and the values of J previously obtained by integrating the conduction integral is shown in Fig. 8. The value of ε_m, 11.0 eV, used in producing this figure was obtained by matching the values of J and J_a, approximately in the center of the domain of interest.

Since U_b is the sum of the work function and the Fermi energy, the order of β is given by

$$\beta \approx 2K_1\phi^{3/2} \approx 10^9$$

From the above, we see a major difference between conventional solid state diode devices and metallic tunnelling devices. For the current in a conventional device, an approximation often used is

$$I = I_o\left(e^{\alpha V} - 1\right) \quad \text{where} \quad \alpha = \frac{e}{kT}$$

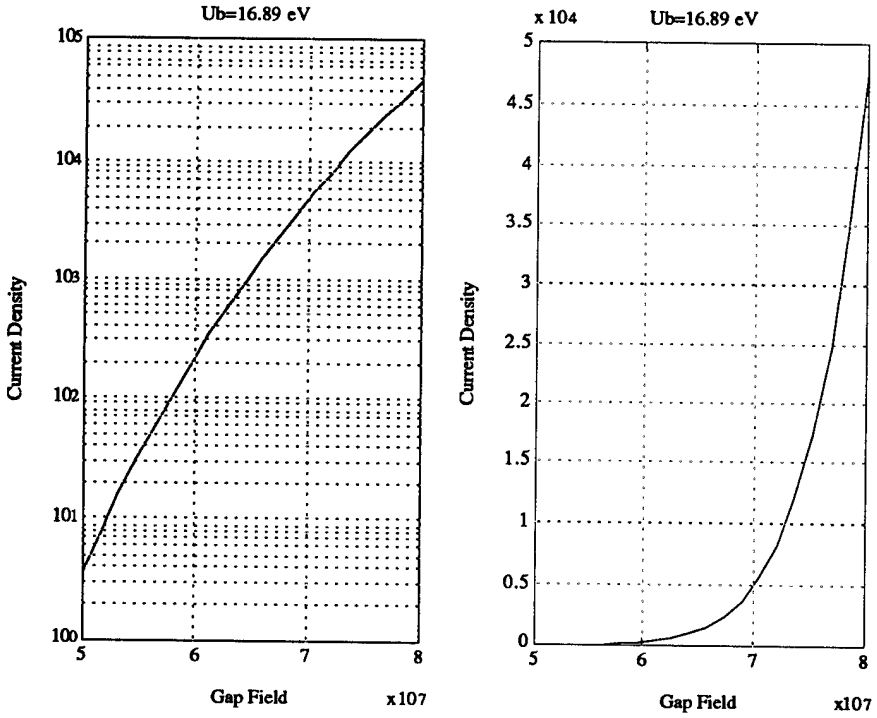


Fig. 7 Semi-log and linear plots of the current density over the range of interest for surge arresting devices.

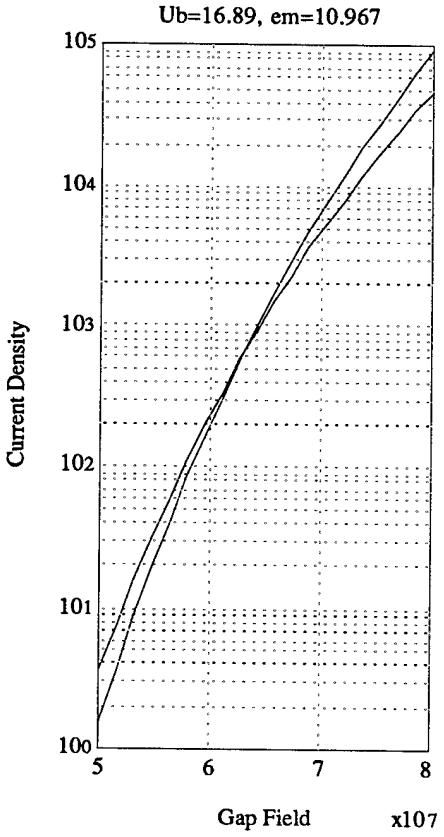


Fig. 8 Comparison of J obtained by integration of the conduction integral to the approximation J_a. J_a is the lower curve at the left of the graph and the higher curve at the right of the graph.

I_o is the reverse bias current, k is Boltzmann's constant, and T is the absolute temperature. Like our results, the last equation is derived on the basis of a rather simple first order model. The first thing we note is that the functional dependence for conventional devices is entirely different than that for tunnelling devices.

The second thing we note is the explicit appearance of the temperature in the equation for conventional devices. In a more sophisticated model, the temperature might play a role in metallic tunnelling devices, but it appears that it is, at most, a second order effect rather than the first order effect of conventional devices. The implication here is that it is likely that metallic tunnelling devices will be more stable with respect to temperature than conventional devices.

The final thing we note is that for all types of devices, on a linear plot of current versus voltage in the application range, the curve must be concave up. On a log-linear plot the curves for conventional devices are also concave up. However, for tunnelling devices, the plots are concave down, as displayed in Fig. 7. This is the tunnelling signature alluded to in our previous comments.

Concluding Remarks

Our approach has been elementary with many simplifying assumptions. However, the failure of material manufacturers to meet specific goals indicates that some manufacturers may have a lack of knowledge of the fundamental physics of the problem, not only in regard to materials, but also with regard to how the materials interact with the structures required for incorporation into practical devices. For example, there is a tendency to think in terms of dc without adequate attention being given to the fact that the electromagnetic interactions are taking place in a transmission line environment.

Of mechanisms available for the production of nonlinear effects in varistor materials, metallic quantum tunnelling offers potential advantages not present in other mechanisms. However, there is no evidence that a metallic tunnelling varistor device has yet been produced. Thus, further theoretical and experimental work is needed in regard to metallic quantum tunnelling.

The fundamental theory presented here gives a qualitative picture of what may be expected and forms the basis for further work. It needs to be expanded to more correctly account for factors ignored through the simplifying assumptions. In addition, analysis is needed in regard to how the materials interact with device structures. These two topics constitute a major portion of the proposed effort.

References

1. C. Davisson and L. H. Germer, Phys. Rev., **20**, 300 (1922); **30**, 634 (1927).
2. G. M. Fleming and J. E. Henderson, Phys. Rev., **58**, 887 (1940).
3. W. Schottky, Physik Z., **15**, 872, (1914).
4. J. Bardeen, Phys. Rev., **49**, 653 (1936); 58, 727 (1940).
5. W. B. Nottingham, Phys. Rev., **47**, 806 (1935); **58**, 927 (1940).
6. R. H. Fowler and L. Nordheim, Proc. Roy. Soc. (London), **119A**, 173 (1928).

7. R. Haefer, Z. Physik, **116**, 604 (1940).
8. L. W. Nordheim, Proc. Roy. Soc. (London), **121**, 626 (1928).
9. L. A. MacColl, Phys. Rev., **56**, 699 (1939).
10. C. Herring and M. H. Nichols, Revs. Mod. Phys., **21**, 185, (1949).
11. J. B. Taylor and I. Langmuir, Phys. Rev., **44**, 423 (1933).
12. Kittel, Charles, "Introduction to Solid State Physics," Sixth Edition, John Wiley & Sons, New York (1986).
13. Jackson, J. D., "Classical Electrodynamics," Second Edition, John Wiley & Sons, New York (1975).
14. Dirac, P. A. M., "The Principals of Quantum Mechanics," Oxford University Press, New York (1958).
15. Wichmann, E. H., "Quantum Physics," Berkeley Physics Course, Vol. 4, McGraw Hill Book Company, New York (1971).
16. Strang, Gilbert, "Linear Algebra and Its Applications," Third Edition, Harcourt Brace Jovanovich, San Diego (1988).
17. Chester, Marvin, "Primer of Quantum Mechanics," John Wiley & Sons, New York (1987).
18. Bochove, E. J., and Walkup, J. F., "A Communication on Electrical Charge Relaxation in Metals," American Journal of Physics, Vol. 58, No. 2 (1990).
19. Thomas, G. B, "Calculus and Analytic Geometry," Alternate Edition, John Wiley & Sons, New York (1972).

II-R

REFERENCES

SELECTED BIBLIOGRAPHY

[1] Hyatt, H. M.; Shrier, K. P., "Electrical Overstress Pulse Protection Material and Process," U. S. Patent 4, 726, 991 (1988).

[2] Dutcher, Clinton, "Metallic Quantum Tunnelling Varistor Devices, Phase I," United States Army Harry Diamond Laboratories, Contract No. DAAL02-91-C-0024 (August 1991).

[3] O'Dwyer, J. J., IEEE Trans. Elec. Ins., 1084, **EI-9**, pp. 1-9.

[4] Lampert, M. A., "Injection Currents in Solids," Academic Press, New York (1965).

[5] Simmons, J. G., "Schottky Injection," Phys. Rev., 1967, 155(3), pp. 657-659.

[6] Hikita, M.; Nagao, M.; Sawa, G., Ieda, M., "Thermal Breakdown," J. Phys. D., Appl. Phys. 1980, **EI-15**, pp 206-224.

[7] Klein, N., "Avalanche," Adv. Phys., 1972, 21, pp. 605-645.

[8] Klein, N., "Avalanche," J. Appl. Phys., 1982, **53**, pp. 5828-5839.

[9] Mayoux, C. J., IEEE Trans. Elec. Insul., 1976, **EI-11**, pp.153-158.

[10] Joncher, A. K; Lacoste, R., IEEE Trans. Elec. Insul., 1984, **EI-19**, pp. 567-577.

[11] Dutcher, Clinton, "Statistical Effects on the Failure Mechanisms of Quantum Tunnelling Varistor Devices," Technical Memorandum, Oryx Technology Corporation, Fremont, CA. (1993).

[12] Nordheim, L.; Fowler, R. H., "Electron Emission in Intense Electric Fields, "Royal Society Proceedings," Volume 119 (1928).

[13] Kittel, Charles, "Introduction to Solid State Physics," Sixth Edition, John Wiley & Sons, New York (1986).

[14] Dutcher, Clinton, "Equivalent Time in Varistor Applications," ITEM, The International Journal of EMC, Robar Enterprises, Conshohocken, PA. (1994).

[15] Spanier, Jerome; Oldham, Keith B., "An Atlas of Functions," Hemisphere Publishing Company, New York (1987).

[16] Solymar, L. and Walsh, D., "Lectures on the Electrical Properties of Materials," Oxford University Press, New York (1993).

[17] Dissado, L. A., and Fothergill, J. C., "Electrical Degradation and Breakdown in Polymers," Peter Perigrinus, Ltd., London (1992).

[18] Fischer, P., and Nissan, K. W., IEEE Transactions on Electrical Insulation, EI-11(2) (1976).

[19] Bohm, David, "Quantum Theory," Prentice-Hall, New Jersey (1951).

[20] Dutcher, Clinton, "Vector Space Quantum Formalism," Integrated Sciences™, in preparation for publication (1996).

[21] Chambers, R. G., Proc. Phys. Soc. London, vol. 65A, (1952).

[22] Stratton, R.: *Progr. Dielectrics*, 1957, **3**, pp. 235-292.

[23] Goodman, B., Lawson, A. W., and Schiff, L. I.: *Phys. Rev.* 1947, **71(3)**, pp. 191-194.

[24] Frölich, H.,: *Proc. Roy. Soc., London*, 1937, **A160**, pp. 230-241.

SECTION III-1

APPLICATION OF FINITE ELEMENT ANALYSIS IN QUANTUM PHYSICAL MODELS

III-1-1 Background

The use of finite element modelling in these electronic solid state analyses is an unusual, and probably a unique, application of an engineering discipline developed for continuum mechanics. It would be quite correct to suggest that such an application, if taking a traditional finite element approach, would have limited value. However, the distinct departure from tradition introduced below and developed in some detail in the following sections will be shown for the first time to accurately characterize chaotic failure mechanisms of solid state materials.

The quantum, electrical and thermal analyses presented in Part II are based on local calculations. That is, the mathematics necessarily apply to a small, idealized *control volume*. The control volume may be represented by a single gap between conductive particles, across which tunnelling or other conducting phenomena may occur. From such simple constructs, overall behavior of a macroscopic system may be inferred. Solutions involving single-electron tunnelling across a gap can provide insight regarding material constituents to be used, idealized gap spacing, and total material thickness between electrodes.

While the above local approach is essential, it is also incomplete. For instance, it was observed that with these voltage gradient sensitive materials the primary mode of failure is of a chaotic, dominant path nature. Current is not conducted uniformly between electrodes but instead forms filaments of extremely high current density which in turn cause overheating and material degradation. These filaments cannot be predicted from inferences made from local calculations alone.

Continuum mechanical solutions based on local relationships likewise cannot predict dominant path failure. One technique already attempted is to apply the local expressions regarding current density vs. temperature and voltage gradient in a traditional application of the finite element method. Each subdomain, or element, within the continuum is allowed to conduct current according to those expressions. In such analyses the predicted current density in material subjected to high voltage transients between parallel electrodes is uniform, thereby departing from observation. If predictions do not agree with measurement, then there is a significant mechanism for which the predicting methods do not account. This failure to predict requires a careful review of those mechanisms.

Dominant path phenomena are in part caused by the unstable nature of the materials. Through a combination of quantum tunnelling and regenerative avalanche, small changes in local voltage gradients result in very large changes in local current density. However, it can be shown that chaotic, dominant path phenomena cannot be brought about by material instability alone.

The other condition necessary for dominant path failure in materials addressed by high voltage transients across parallel electrodes is material non-uniformity. Even minute variability in material properties from control volume to control volume throughout the material space can result in variability among possible conduction paths. If higher conductivity leads to higher current density in a particular path, regenerative effects will tend to accelerate this current concentration many fold. The macroscopic effect will be that nearly all current between electrodes will flow through one or a very few such paths, or filaments.

III-1-2 Predicting Dominant Path Failure Modes

This second condition necessary for dominant path failure, material nonuniformity, is not within the scope of traditional continuum mechanics. However, a mathematical technique is developed below whereby Gaussian-distributed non-uniformity is randomized and imposed spatially in the context of finite elements. Just as each control volume of material is slightly different from other control volumes of the material in space, the model analogy allows for element-to-element material properties to vary both randomly and within a logical Gaussian structure. Chaotic phenomena such as dominant path failure are accurately predicted. An example of such a finite element model is shown in Figure III-1 at right. This model shows dominant path overheating in voltage-sensitive material sandwiched between parallel electrodes and subjected to an overvoltage transient.

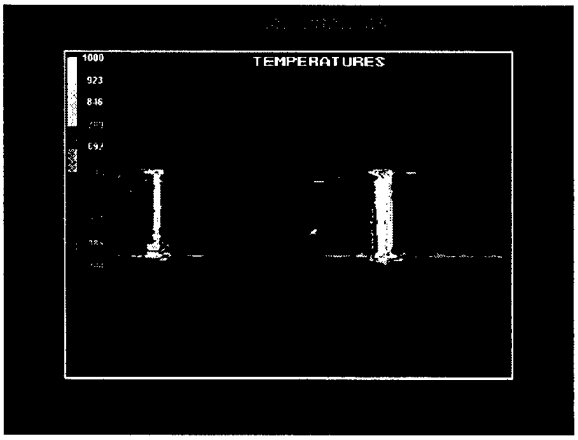


Figure III-1
Model of chaotic dominant path failure in voltage-sensitive material subjected to overvoltage transient.

Models using this technique cannot predict the exact location of dominant path formation. However, the overall *behavior* of the assemblage and the *character* of the performance of the material will be far better represented when compared to traditional techniques.

III-1-3 Developing New Materials

Predicting failure modes is only part of the benefit of modelling. The goal is to develop new, robust materials which do not fail when subjected to overvoltage transients. Using models to develop such robust materials requires three steps. Step one is to model the failure mechanisms successfully such that the performance of the modelled material matches laboratory measurement. Step two is to manipulate the bulk properties within the model until the desired performance is achieved (i.e., chaotic failure mechanisms disappear), being careful to select properties to manipulate that can be realistically altered in the real world. Step three is to find a material formulation that can actually achieve the modelled change in bulk properties that achieved success in model space.

SECTION III-2

FINITE ELEMENT EXPRESSION OF THE LOCAL QUANTUM MECHANICAL PROBLEM

III-2-1 Control Volumes As Finite Elements

Extending the relationships developed in Part II to macroscopic geometries can theoretically be done by expressing the local quantum mechanical equations in continuum mechanical form. Control volumes are chosen to be large enough to average the effects of individual quanta. The properties expressed within any given control volume are consistent with the quantum mechanical relationships, and those relationships are applied uniformly within that control volume. The assumption is that if one chooses the control volume to be small enough, the predictions within that control volume will closely match measured behavior. The difficulty is in attempting to use a single set of expressions relating to a small control volume and to apply these expressions seamlessly to a macroscopic structure. Material variations among control volumes within the structure are thereby ignored. Chaotic failure mechanisms such as the dominant path phenomenon described in Part III-1 cannot be predicted.

The finite element method lends itself to macroscopic extensions of such control volume calculations.[1] This divide-and-conquer approach to continuum mechanics seeks to divide a large solution domain into a large number of subdomains, called *elements*. The element shapes are typically chosen to be polygons, having straight or deformable sides, such that they can be fit together to approximate the shape of the solution domain. See Figure III-2. Interpolation functions are used to describe the field variables within each element and across the solution domain in piecewise-continuous form. With regard to the instant problem, each element may be treated as an individual control volume, within which the governing equations are satisfied.

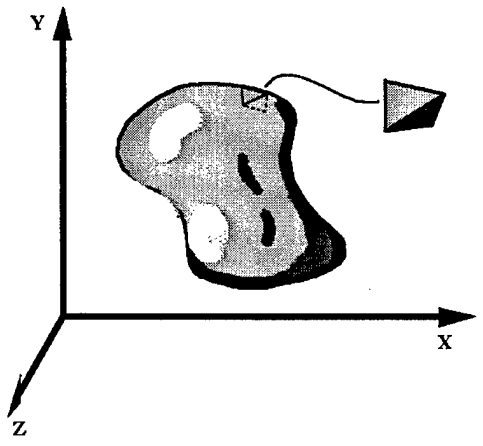


Figure III-2
3-D solution domain Ω showing a single tetrahedral element as part of discretization.

III-2-2 The Problem of Material Variability

Although the finite element method can be used to alleviate the problems of geometric complexity in a continuum mechanical problem, material variability is ignored in traditional applications of this method. Perhaps as a leftover from closed-form solution techniques, materials of a given type within the solution domain are

---

[1] Zienkiewicz, O.C., *The Finite Element Method*, McGraw-Hill, London, 1977

assumed to possess certain average properties that are imposed throughout with mathematical perfection. Such perfection does not exist in the real world. What is introduced below is a formal method for introducing random material variation in a form that mimics the variation in real materials. The method follows with means to impose that variation on an element-by-element basis for finite element solution.

The description proceeds in this sequence. First, the generalized continuum mechanical field equation is introduced. This equation is then reduced separately to the thermal and electrical conduction field equations, along with detailed finite element formulations for each. Equations for coupling between the thermal and electrical phenomena, developed in Part II, are then reintroduced. Iterative schemes are developed for handling the coupling terms. Finally, a Gaussian distributed and spatially imposed (i.e.;element-by-element) material randomization scheme is presented.

III-2-3 Continuum Mechanical Field Equation

Transient field problems in domain Ω can be expressed in the general form

$$\nabla \cdot \boldsymbol{k} \nabla \varphi = f(x,y,z,t) + k_t \dot{\varphi} + k_{tt} \ddot{\varphi}$$

where

$$\dot{\varphi} \equiv \frac{\partial \varphi}{\partial t}; \qquad \ddot{\varphi} \equiv \frac{\partial^2 \varphi}{\partial \varphi^2}$$

with boundary conditions

$$\varphi = \Phi(x,y,z,t) \qquad \text{on surface} \quad S_1, \; t > 0$$

$$k_{xx} \frac{\partial \varphi}{\partial x} n_x + k_{yy} \frac{\partial \varphi}{\partial y} n_y + k_{zz} \frac{\partial \varphi}{\partial z} n_z + q(x,y,z,t) + h(x,y,z,t)\varphi = 0$$

$$\text{on surface } S_2, \; t > 0$$

and initial conditions

$$\varphi = \varphi_0(x,y,z) \text{ in } \Omega \text{ at } t = 0$$

$$\dot{\varphi} = \xi_0(x,y,z) \text{ in } \Omega \text{ at } t = 0.$$

Expression of field problems in such a generalized form is useful in developing a finite element solution to coupled field problems. By using a general form the solution technique, and therefore entire sections of implementing computer code, can shared by each of the coupled phenomena. In the instant case, the problem is thermoelectrical. The solution must solve the temperature and voltage field problems, providing for flow of thermal energy and flow of electrical current. The analogies are shown in Table III-1. The solution must also provide for coupling terms that affect two-way interdependence: local fluctuations in thermal energy as a function of

nonlinear joule heating, and local changes in apparent resistivity as a function of local thermal energy. Finite element formulation of the general problem is developed below, followed by specifics for the thermal and electrical problems, coupling, and material variability.

Temperature $\varphi = T$	General Term	Voltage $\varphi = V$
Temperature T	φ	Voltage V
Thermal conductivity k	k	Electrical conductivity σ
Heat flow across surface	q	Current flow across surface
Volumetric heat generation	q*	Charge generation
Temp. of convective fluid	φ_∞	Input voltage
Convective heat flow	$h(\varphi-\varphi_\infty)$	Current input across resistor, where R = 1/hA

Table III-1
Thermoelectrical Analogies

III-2-4 Finite Element Formulation of the Continuum Mechanical Field Equation

As described above, the finite element method requires the solution domain to be divided into subdomains, or elements. *Interpolation functions* are chosen to describe the field variable(s) everywhere within the element. These functions express the value of field variable(s) relative to spatial coordinates and are formulated in such a way that the field variable is specified within the element at distinct points, or *nodes.* The nodes typically lie along the sides where the elements connect to each other. The interpolation functions (N_i) have undetermined parameters, or multipliers. These undetermined parameters may be, for instance, coefficients of a linear polynomial when such a function is used for interpolation. The function has the following general form for a time dependent problem where for any instant of time the time derivatives are assumed to be functions of the spatial coordinates only:[2]

$$\varphi^{(e)} = \sum_{i=1}^{r} N_i(x,y,z)\, a_i(t) = \lfloor N \rfloor \{a(t)\}^{(e)}$$

where $a_1(t), a_2(t), \ldots, a_r(t)$ are the undetermined time-dependent nodal parameters for an element of r degrees of freedom. $\lfloor N \rfloor$ is a row vector of dimension $1 \times r$, and $\{a(t)\}^{(e)}$ is a column vector of dimension $r \times 1$.

---

2 Huebner, K.H., *The Finite Element Method for Engineers*, John Wiley & Sons, New York, 1982

The number of undetermined parameters in the interpolation function equals the number of nodes at which the field variable is to be calculated within the element. There are three implications. First, equations describing the nodal values in terms of the element properties can be collected to form *element stiffness matrices*. These are assembled in turn to form the global stiffness matrix that, with the addition of boundary and initial conditions, fully describes the problem. The interim result is a system of linear algebraic equations which is solved simultaneously for all nodal values of the field variable.

Second, the sequence of repetitive operations, or algorithm, needed to solve such a system of equations is a task that is particularly suited to computer implementation. Nonlinear problems typically require repeated iteration of the equation solving process.

Third, once the nodal values are known, values of the field variable everywhere within each element can be solved for by introducing these nodal values back into the interpolation function. This exercise is especially useful in displaying results.

What sets the finite element method apart from other discretization schemes, such as finite differences, is that with proper choice of interpolation function it provides true piecewise continuity for field variable(s) within the solution domain. One way of picturing this piecewise continuity is to imagine that the interpolation functions "zip together" to provide continuous expression of the field variable(s) from element to element.

If the element stiffness matrices are expressed in the form

$$\varphi^{(e)}(x, y, z) = \lfloor N^{(e)} \rfloor \{a^{(e)}\}$$

then for M elements within the solution domain the complete representation of the field variable is given by

$$\varphi(x,y,z) = \sum_{e=1}^{M} \varphi^{(e)}(x,y,z) = \sum_{e=1}^{M} \lfloor n^{(e)} \rfloor \{a^{(e)}\}$$

No variational formulation exists for the complete parabolic expression of the transient heat transfer equation, but Galerkin's method leads to the following general finite element solution:

$$[k_\varphi]^{(e)} \{\varphi\}^{(e)} + [k_t]^{(e)} \{\dot{\varphi}\}^{(e)} + [k_{tt}]^{(e)} \{\ddot{\varphi}\}^{(e)} + [k_h]^{(e)} \{\varphi\}^{(e)} + \{F(t)\}^{(e)} = \{0\}$$

where

$$K_{\varphi_{ij}} = \int_{\Omega^{(e)}} \left(k_{xx} \frac{\partial N_i}{\partial x} \frac{\partial N_j}{\partial x} + k_{yy} \frac{\partial N_i}{\partial y} \frac{\partial N_j}{\partial y} + k_{zz} \frac{\partial N_i}{\partial z} \frac{\partial N_j}{\partial z} \right) d\Omega$$

$$K_{t_{ij}} = \int_{\Omega^{(e)}} K_t N_i N_j \, d\Omega$$

$$K_{tt_{ij}} = \int_{\Omega^{(e)}} K_{tt} N_i N_j \, d\Omega$$

$$K_{h_{ij}} = \int_{S_2^{(e)}} h N_i N_j \, dS_2$$

$$F_i = \int_{\Omega^{(e)}} f N_i \, d\Omega + \int_{S_2^{(e)}} q_i \, dS_2$$

III-2-5 Transient Thermal Equation

Consider transient heat transfer in a three-dimensional anisotropic solid Ω bounded by a surface Γ as shown in Figure III-3. The generalized field equation takes on the parabolic form[3]

$$\nabla \cdot k \nabla T + q^* = \rho C_p \frac{\partial T}{\partial t}$$

where

k represents the anisotropic conductivity matrix,

$q^* = q^*(x,y,z,t)$ is the internal heat generation rate per unit volume,

$\rho = \rho(x,y,z,t)$ is the material density,

$C_p = C_p(x,y,z,t)$ is the specific heat.

---

[3] Chapman, A.J., *Heat Transfer*, Third Edition, Macmillan, New York, 1974

Initial conditions are

$$T(x,y,z,0) = T_0(x,y,z)$$

Boundary conditions include specified surface temperature

$$T_s = T_1(x,y,z,t) \qquad \text{on } S_1$$

specified surface heat flow

$$q_x n_x + q_y n_y + q_z n_z = -q_s \qquad \text{on } S_2$$

convective heat exchange

$$q_x n_x + q_y n_y + q_z n_z = h(T_s - T_\infty) \qquad \text{on } S_3$$

Note that radiation heat exchange expected to be second order in the instant problem and is therefore not considered in these analyses.

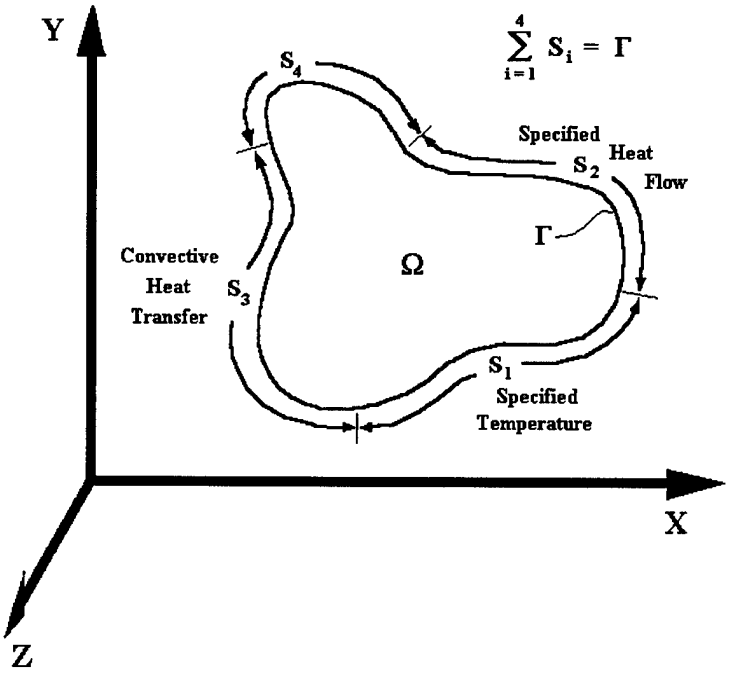


Figure III-3

3-D solution domain Ω for general heat conduction, with surface Γ.

III-2-6 Finite Element Formulation of the Transient Thermal Equation

The finite element formulation for the generalized field equation reduces to the parabolic, thermal form. $\lfloor N \rfloor$ becomes the *temperature* interpolation matrix. For economy of expression, we choose $[B]$ to represent the temperature gradient interpolation matrix

$$[B(x,y,z)] = \begin{bmatrix} \frac{\partial N_1}{\partial x} & \frac{\partial N_2}{\partial x} & \cdots & \frac{\partial N_r}{\partial x} \\ \frac{\partial N_1}{\partial y} & \frac{\partial N_2}{\partial y} & \cdots & \frac{\partial N_r}{\partial y} \\ \frac{\partial N_1}{\partial z} & \frac{\partial N_2}{\partial z} & \cdots & \frac{\partial N_r}{\partial z} \end{bmatrix}$$

The transient thermal finite element formulation becomes

$$[C]\left\{\frac{dT}{dt}\right\} + [[K_C] + [K_h]]\{T\} = \{R_T\} + \{R_Q\} + \{R_q\} + \{R_h\}$$

where

$$[C] = \int_{\Omega^{(e)}} \rho c \{N\}\lfloor N \rfloor d\Omega$$

is the thermal capacitance matrix. The coefficient matrices

$$[K_c] = \int_{\Omega^{(e)}} [B]^T [k] [B] \, d\Omega$$

$$[K_h] = \int_{S_3} h \{N\}\lfloor N \rfloor d\Gamma$$

relate to conduction and convection, respectively, and the heat load vectors

$$\{R_T\} = -\int_{S_1} (q \cdot \hat{n}) \{N\} d\Gamma$$

$$\{R_Q\} = \int_{\Omega} Q \{N\} d\Omega$$

$$\{R_q\} = \int_{S_2} q_S \{N\} d\Gamma$$

$$\{R_h\} = \int_{S_3} h T_e \{N\} d\Gamma$$

define specified nodal temperatures along a surface, internal heat generation, specified surface heating, and surface convection, in the above order.

III-2-7 Two-Dimensional Implementation: The Linear Triangular Element

A three-node triangular element is shown at right in Figure III-4. The element is assumed to have thickness τ and to have temperature defined throughout as a first-order linear equation in x and y as shown. Although higher-order elements could be used, this element is chosen for its simplicity while still expressing temperature (although not the *derivatives* of temperature) with piecewise continuity from elem ent to element.[4]

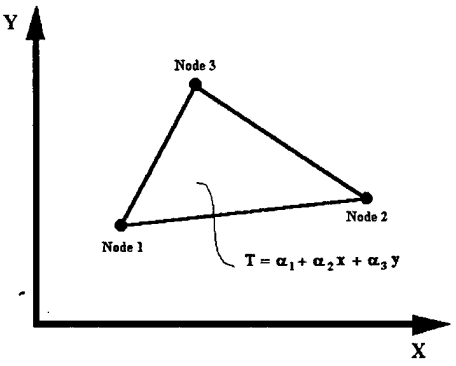


Figure III-4
Linear triangular element

The first-order expression for temperature can be rewritten in terms of *natural coordinates* L_i where

$$L_i(x,y) = \frac{1}{2\Delta}(a_i + b_i x + c_i y) \quad , \qquad i = 1, 2, 3$$

for the triangle of area Δ, and where

$$2\Delta = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix} = 2\,(\textit{area of triangle } 1\text{-}2\text{-}3)$$

$$\{a\} = \begin{Bmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{Bmatrix}$$

$$\{b\} = \begin{Bmatrix} y_2 - y_3 \\ y_3 - y_1 \\ y_1 - y_2 \end{Bmatrix}$$

$$\{c\} = \begin{Bmatrix} x_3 - x_2 \\ x_1 - x_3 \\ x_2 - x_1 \end{Bmatrix}$$

---

[4] Segerlind, L.J., *Applied Finite Element Analysis*, John Wiley & Sons, New York, 1976

For the three-noded linear triangle the element interpolation functions equal the natural coordinates, or

$$N_i = L_i$$

Since

$$\frac{\partial N_i}{\partial x} = \frac{b_i}{2\Delta}, \qquad \frac{\partial N_i}{\partial y} = \frac{c_i}{2\Delta}$$

The temperature gradient interpolation matrix becomes

$$[B] = \frac{1}{2\Delta}\begin{bmatrix} b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

which leads to the conduction matrix

$$[K_c] = \Delta\tau\,[B]^T\,[k]\,[B]$$

capacitance matrix

$$[C] = \frac{\rho C_p \Delta\tau}{12}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

internal heating vector

$$\{R_Q\} = \frac{q\Delta\tau}{3}\begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix}$$

$$R_q = \frac{q_s \tau l_{12}}{2}\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$$

the vector for heat flow across one element edge l_{12}

and components for convection across one element edge l_{12} to fluid at temperature T_∞ with convection constant h

$$[K_h] = \frac{h\tau l_{12}}{6}\begin{Bmatrix} 2 & 1 \\ 1 & 2 \end{Bmatrix}$$

$$\{R_h\} = \frac{h T_\infty \tau l_{12}}{2}\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$$

The gradient across this first order element is constant and given by

$$\nabla T = \frac{\partial T}{\partial x}\vec{i} + \frac{\partial T}{\partial y}\vec{j} = a_1\vec{i} + a_2\vec{j}$$

Constant gradients within elements imply discontinuities in the gradients at the boundaries. Heat (and electric current in the electrical problem) are not rigorously conserved. However, with sufficient element packing in ares of expected large gradient change, the resulting solution will not suffer significantly from this constraint.

III-2-8 Solution in the Time Domain

Solution in the time domain is accomplished explicitly with difference terms substituted for differential operators in time. For instance, a central difference approximation of the derivative of generalized field variable φ with respect to time would be, for time interval t_s,

$$\dot{\varphi} = \frac{d\varphi}{dt} \approx \frac{\varphi_1 - \varphi_0}{t_s}$$

Note that the best approximation for $\dot{\varphi}$ can be found at time

$$t = \frac{t_0 + t_1}{2}$$

requiring that all parts of the solution be evaluated at this time.

Time dependency in the instant thermoelectrical problem is solved explicitly for several reasons. First, the coupling procedure is itself iterative. Second, extreme nonlinearities possible in local expressions for

$$\sigma = \sigma(\xi, T)$$

require that they be evaluated at small time intervals anyway. Third, a structure is implied for adding subsequent nonlinearities and additional field variables with a minimum of effort. The central difference formulation makes those additions less cumbersome.

Implementing central difference time stepping in vector form gives

$$\{\dot{\varphi}\} \approx \frac{1}{t_s}(\{\varphi_1\} - \{\varphi_0\})$$

Likewise, the next update for $\{\varphi\}$, designated $\{\varphi_1\}$, is

$$\{\varphi_1\} = \frac{1}{2}(\{\varphi_0\} + \{\varphi_1\})$$

and the forcing function update is

$$\{F_t\}^* = \frac{1}{2}(\{F_0\} + \{F_1\})$$

III-2-9 Electrical Equation

As a first approximation to the electrical equation, all derivatives with respect to time are assumed to be zero over a time interval represented by the small time step used in the calculations (as small as 10^{-9} second). Therefore

$$\frac{\partial^n V}{\partial t^n} = 0 \qquad \textit{for all } n > 0$$

Although this may seem to be a gross oversimplification, it is believed justified on the basis of the speed of electrical response vs. thermal response. Since the electrical response is much faster, during each time step it is reasonable to assume that in relative terms the electrical problem occurrs instantaneously. With the availability of a faster computer and more time for analysis, the secondary effects of electrical capacitance and inductance may eventually be addressed.

With these assumptions, the electrical equation reduces to the elliptical form

$$\nabla \cdot \sigma \nabla V + I^* = 0$$

where I^* represents a current source and σ is the anisotropic conductivity matrix.

Boundary conditions include specified voltage at a surface

$$V_s = V_1(x,y,z,t) \qquad \text{on } S_1$$

specified surface current flow

$$I_x n_x + I_y n_y + I_z n_z = -I_s \qquad \text{on } S_2$$

and time-dependant voltage input with constraints imposed by external electronic circuit behavior

$$V = V(t, R_B, \ldots)$$

where R_B is an external ballast resistor as an example of circuit constraint.

III-2-10 Finite Element Formulation of the Electrical Conduction Equation

The finite element formulation for the generalized field equation, given the extreme simplifying assumptions, reduces to the elliptic form for electrical conduction. $[N]$ becomes the *voltage* interpolation matrix; $[B]$ is the voltage gradient interpolation

matrix. Other development proceeds as in the thermal formulation, but in quasi steady-state form:

$$[K_{c_{elec}}]\{V\} = \{R_V\} + \{R_I\}$$

with coefficient matrix

$$[K_{c_{elec}}] = \int_{\Omega^{(e)}} [B]^T [\sigma] [B] \, d\Omega$$

and load vectors

$$\{R_V\} = -\int_{S_1} (I \cdot \hat{n}) \{N\} \, d\Gamma$$

$$\{R_I\} = \int_{\Omega} I \{N\} \, d\Omega$$

define specified voltage at a surface and charge generation, respectively.

III-2-11 Thermoelectrical Coupling Equations

Local thermal input from electrical conduction relates to the local voltage field ξ and to the local electrical conductivity σ. Local current density is[5]

$$J = \sigma \xi$$

and local heat generation rate (per unit volume) is

$$q^* = \frac{J^2}{\sigma} = \sigma \xi^2$$

The elemental heat generation rate is then

$$q^*_{(e)} = \sigma_{(e)} \xi^2_{(e)}$$

---

[5] Solymar, L., and Walsh, D., *Lectures on the Electrical Properties of Materials*, Fourth Edition, Oxford University Press, Oxford, 1988

For a two-dimensional element of area Δ and depth τ, the total heat generated over time step t_s is

$$Q_{(e)} = \int_0^{t_s} \tau \int_\Omega \sigma_{(e)} \xi^2_{(e)} \, d\Omega \, dt = \tau \Delta \, \sigma_{(e)} \, \xi^2_{(e)} \, t_s$$

For the above equation the element electrical conductivity is assumed to be constant. However, with voltage gradient-sensitive materials, the conductivity is considered to be a highly nonlinear function of voltage field and temperature, or

$$\sigma = \sigma(\xi, T)$$

which becomes for a single element having constant gradient in voltage and average temperature $T_{ave(e)}$

$$\sigma_{(e)} = \sigma_{(e)} (\xi_{(e)}, T_{ave_{(e)}})$$

There is an implied complexity which may have escaped the reader: the above equation is for *local* σ , assumed to be constant within a single element. However, changes in local σ may profoundly affect the global description of ξ , which in turn affects ξ within each element ... which affects local σ , etc., etc. Implied in the construction of the last sentence is the iterative scheme for solving the electrical equation involving materials where local voltage gradient governs, in part, local electrical conductivity. That iterative scheme within the electrical formulation can be summarized as

$$\sigma_{(e)_1} = \sigma_{(e)} (\xi_{(e)_0}, T_{ave_{(e)}})$$

where the subscript **1** refers to the updated value and subscript **0** refers to the last calculated value. $T_{ave(e)}$ refers to the average element temperature calculated in the last time step and assumed to be constant within the iteration solving the electrical equations. Since this iteration is often unstable, a dampening term p is employed such that

$$\sigma_{(e)_1} = (1 - p) \, \sigma_{(e)_0} + p \, \sigma_{(e)} (\xi_{(e)_0}, T_{ave_{(e)}})$$

where $p = 0.3$ is a typical value used to slow convergence.[6]

The updated σ and ξ are used to calculate heat generation in the next time step. A flow chart of the complete solution scheme is shown in Figure III-5.

---

[6] Arden, B.W., Astill, K.N., *Numerical Algorithms: Origins and Applications*, Addison-Wesley, Reading, MA, 1970

III-2-12 Spatial Variability of Materials

The finite element formulation developed here is based on the assumption that real-world material properties vary spatially. To address such variability, properties must vary from control volume to control volume within a given solution domain. The hypothesis is that the spatial variability is Gaussian. That is, if one chooses to divide a given volume of material into control volumes of a given size, then the resulting probability density function for a given property over that volume would have the appearance of that shown in Figure III-6. Choosing a smaller size for each control volume, thereby increasing the total number filling the volume, results in a flatter distribution, i.e., larger variance.

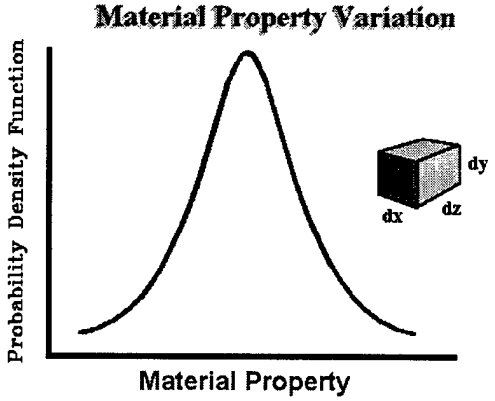


Figure III-6
Material property distribution for all control volumes dx by dy by dz.

Implementing this Gaussian-distributed and spatially-imposed randomness within a finite element framework comes naturally from the discretization of the solution domain. The elements become the control volumes referred to above. Many smaller elements result in a flatter probability distribution when compared to a few larger elements covering the solution domain, the limit being *one* element.

The probability density function for a Gaussian distribution can be expressed

$$f_x(y) = \frac{1}{\sqrt{2\pi}\,\sigma_p} e^{-(y-u)^2/2\sigma_p^2}, \qquad for \; -\infty < y < \infty$$

where the standard deviation σ_p is subscripted to avoid confusion with electrical conductivity, σ.

Using the above density function in a random way to material properties could be computationally intensive. However, good approximations of that distribution can be obtained easily from an application of the Central Limit Theorem. For example, let the random variables $X_1, X_2, \ldots, X_n$ be independent with means $\mu_1, \mu_2, \ldots, \mu_n$ and variance $\sigma^2_{p_1}, \sigma^2_{p_2}, \ldots, \sigma^2_{p_n}$. Consider the random variable Z_n,

$$Z_n = \frac{\left[\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right]}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

Z_n is approximately Gaussian distributed with zero mean and unit variance in the sense that[7]

$$\lim_{n \to 0} P\{Z_n \le b\} = \int_b^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$

Note that if the X_i form a random sample, with each X_i having mean μ and variance σ_p^2, then

$$Z_n = (\bar{X} - \mu)\frac{\sqrt{n}}{\sigma_p}$$

Therefore sample means from random samples tend to form Gaussion distributions as described by the Central Limit Theorem even if the X_i are not Gaussian, or *normally*, distributed. Since a random real number has a uniform distribution from 0 to 1, it has mean 0.5 and a standard deviation $1/\sqrt{12}$. The theorem then implies that the sum of n random numbers has approximately a Gaussian distribution with mean $n/2$ and standard deviation $\sqrt{n/12}$. Given a sample of random decimal numbers $r_1, r_2, \ldots, r_n$, then a random observation from an apparent Gaussian distribution with mean μ and standard deviation σ_p can be constructed from

$$x = \frac{\sigma_p}{\sqrt{n/12}} \sum_{i=1}^{n} r_i + \left(\mu - \frac{n}{2} \frac{\sigma_p}{\sqrt{n/12}} \right)$$

One may choose n such that the square roots need not be calculated. In the practical implementation of the above expression, n is chosen to be 48, giving

$$x = \frac{\sigma_p}{2} \sum_{i=1}^{48} r_i + (\mu - 12\,\sigma_p)$$

which requires no more than five statements of FORTRAN or C code to execute.

A sample Gaussian distribution calculated from the above expression is shown in Figure III-7. The sample of 1000 x's was calculated by a program written in Borland C where the 48 r's for each x were calculated from a pseudorandom number generator using the time of day in milliseconds as a seed. The input μ and σ_p were 10.000 and 0.100, respectively. The resulting values calculated from the generated numbers were 9.9994 and 0.1011 as plotted.

---

[7] Hillier, F.S., Lieberman, G.J., *Operations Research*, Second Edition, Holden-Day, San Francisco, 1974

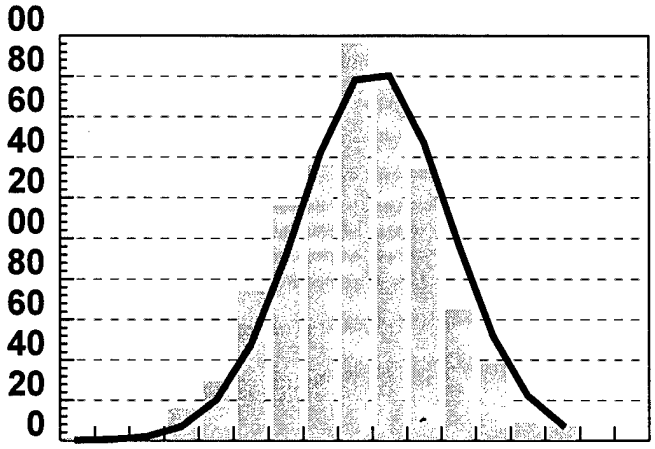


Figure III-7
Sample Gaussian distribution calculated from 1000 sums of evenly distributed random numbers.

Practical values for μ and σ_p come from laboratory tests of material samples of volume V_T, yielding statistics from measurement. For a given finite element analysis having all element volumes of exactly V_T, the variable material property could be calculated from the above expression, where the r_i come from a pseudo-random number generator. If the element sizes vary, a condition which is far more likely, then the material property would need to be adjusted for an element of volume $V_{(e)}$ according to

$$x = \frac{\sigma_{p_{(e)}}}{2} \sum_{i=1}^{48} r_i + (\mu - 12\,\sigma_{p_{(e)}})$$

where

$$\sigma_{p_{(e)}} = \sigma_p \sqrt[3]{\frac{V_T}{V_{(e)}}}$$

There is a major implication of using the above Gaussian-distributed and spatially-imposed randomization technique. Provided the random number seeds are varied, no two runs of the same problem will yield exactly the same results. The collective behavior of a set of such otherwise identical problems is non-deterministic, just as with real physical systems.

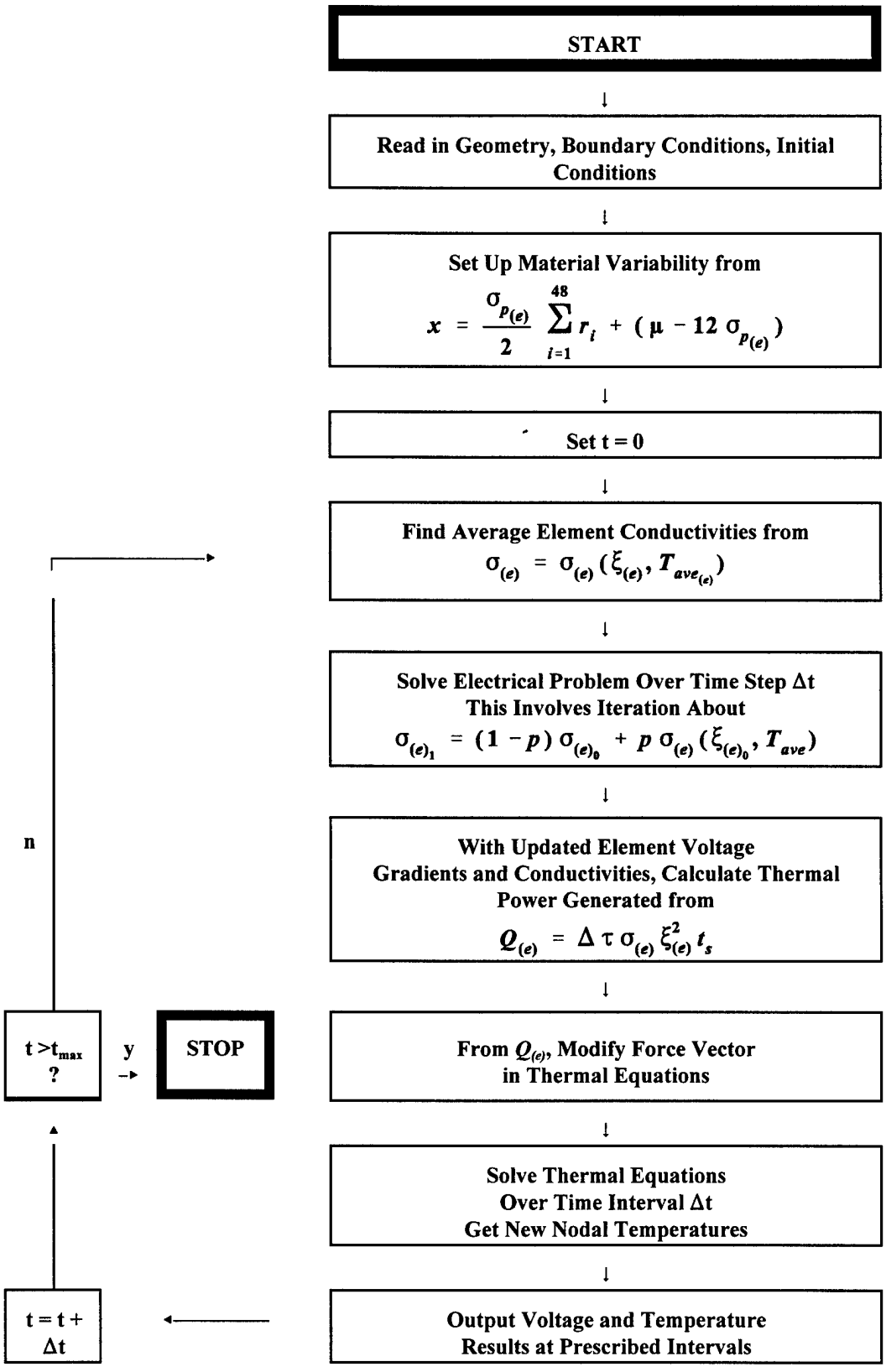


Figure III-5 Solution Scheme Flow Chart

III-2-13 Axisymmetry

In solution domains having an axis of rotation **z** it is convenient to express the problem axisymmetrically, where **x** and **y** are replaced by **r** and **z**. This formulation is equivalent to expression in cylindrical coordinates, where the generalized Laplacian operator becomes

$$\nabla^2 \varphi = \frac{\partial^2 \varphi}{\partial r^2} + \frac{1}{r}\frac{\partial \varphi}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \varphi}{\partial \theta^2} + \frac{\partial^2 \varphi}{\partial z^2}$$

but with the assumption that the effect of θ is negligible, or

$$\frac{\partial}{\partial \theta} \rightarrow 0 \quad .$$

giving

$$\nabla^2 \varphi = \frac{\partial^2 \varphi}{\partial r^2} + \frac{1}{r}\frac{\partial \varphi}{\partial r} + \frac{\partial^2 \varphi}{\partial z^2}$$

such that

$$\boldsymbol{K}_{\varphi_{ij}} = \int_{\Omega^{(e)}} 2\,\pi\,\bar{\boldsymbol{R}} \left(k_{rr}\frac{\partial N_i}{\partial r}\frac{\partial N_j}{\partial r} + k_{zz}\frac{\partial N_i}{\partial z}\frac{\partial N_j}{\partial z} \right) d\Omega$$

where, for the three-noded axisymmetric triangle,

$$\bar{\boldsymbol{R}} = \frac{1}{12} \lfloor r \rfloor \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \{r\}$$

the capacitance matrix is

$$[\boldsymbol{C}] = \int_{V^{(e)}} r\,\rho c_p \{N\} \lfloor N \rfloor dV$$

for element volume V. These and other terms similarly reflect the fact that axisymmetric elements represent cross-sections of area Δ swept 2π radians around the **z** axis to form *rings.* See Figure III-8.

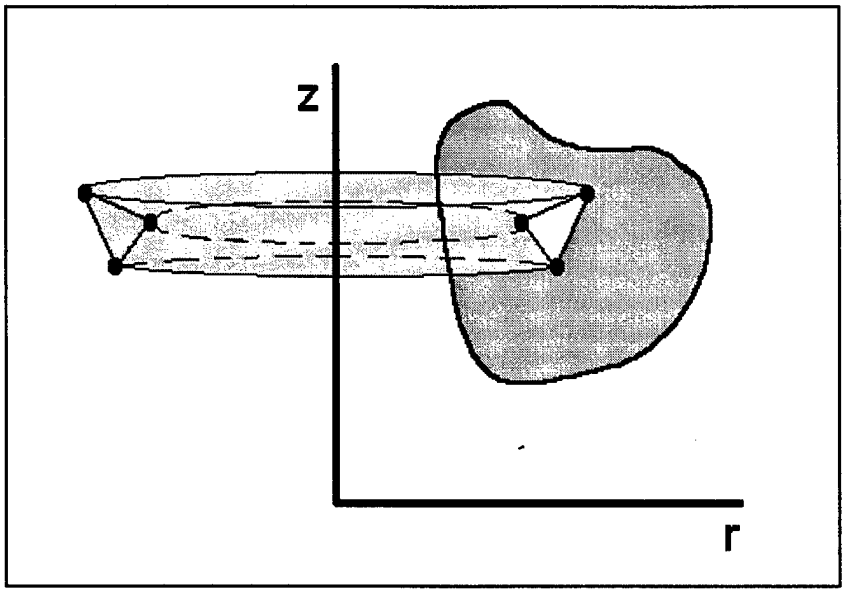


Figure III-8
Axisymmetric three-noded triangular element

The axisymmetric transient thermal equation becomes

$$\nabla \cdot k\, r\, \nabla T + r\, q^{*} = r\, \rho C_p \frac{\partial T}{\partial t}$$

where the finite element formulation for the three-noded triangular element gives natural coordinates

$$L_i(x,y) = \frac{1}{2\Delta}(a_i + b_i r + c_i z) \quad , \qquad i = 1,\ 2,\ 3$$

Vectors $\{a\}$, $\{b\}$, $\{c\}$, and Δ are likewise calculated as shown in the two dimensional formulation in III-2-6 by replacing x and y with r and z, respectively.

The capacitance matrix, expanded from the above general expression, is

$$[C] = \frac{2\pi\rho C_p \Delta}{12}\begin{bmatrix} 6r_1+2r_2+2r_3 & 2r_1+2r_2+r_3 & 2r_1+r_2+2r_3 \\ sym & 2r_1+6r_2+2r_3 & r_1+2r_2+2r_3 \\ sym & sym & 2r_1+2r_2+6r_3 \end{bmatrix}$$

and heat generation matrix becomes

$$[R_Q] = \frac{2\pi q^* \Delta}{12} \begin{bmatrix} 6r_1+2r_2+2r_3 & 2r_1+2r_2+r_3 & 2r_1+r_2+2r_3 \\ sym & 2r_1+6r_2+2r_3 & r_1+2r_2+2r_3 \\ sym & sym & 2r_1+2r_2+6r_3 \end{bmatrix}$$

also reflect the increasing volume implied by increasing r.

SECTION III-3

ITERATIVE SOLUTION TECHNIQUES FOR LARGE SPARSE SYSTEMS OF SIMULTANEOUS NONLINEAR EQUATIONS

III-3-1 Background

The global stiffness matrix and force vector in a finite element problem have certain characteristic properties. First, for any problem of substance, they represent a system of equations with a large number of unknowns. Second, particularly after modifications for the boundary conditions, the global stiffness matrix tends to be quite sparse, with a low percentage of nonzero components. Third, wherever possible, symmetry is preserved to reduce storage and number of arithmetic operations in solution. Fourth, nodal renumbering methods can be used to reduce the bandwidth, outside of which all terms are zero. See Figure III-8 at right. Storage can then be reduced to the diagonal and one side of the symmetric matrix, as shown in Figure III-9.

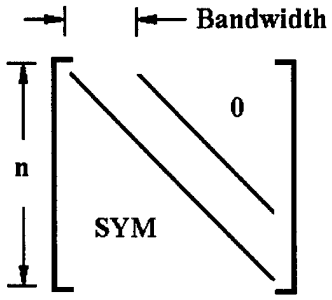


Figure 1
Global stiffness matrix for n unknowns showing band width.

Such storage schemes are useful for implicit solution methods such as Gaussian elimination, where the number of arithmetic operations is known (n^2 times bandwidth). Explicit iterative methods can benefit as well. However, for three dimensional problems bandwidth cannot be reduced as much by renumbering as is possible in two dimensional problems. Storage therefore becomes a severe limitation with large three dimensional problems even on the largest mainframes. Even modest two dimensional problems can be a challenge for a small desktop computer.

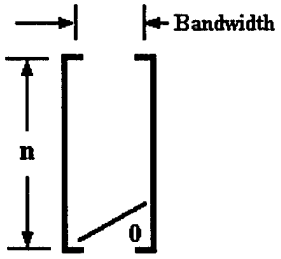


Figure 2
Storage of a banded symmetric matrix.

The number of unknowns in the problems addressed in this work are expected to exceed 20,000, representing a full matrix of that number squared times four bytes precision, or 1.6 *billion* bytes. The available computer is a Pentium-based PC with 32 *mega*bytes of RAM. In that context, the need for this Section and discussion of solution techniques can be understood.

III-3-2 Iterative Methods

For a system of n equations

$$[k]\{\varphi\} = \{f\}$$

each equation has the form

$$k_{i,1}\varphi_1 + k_{i,2}\varphi_2 + \ldots + k_{i,i}\varphi_i + \ldots + k_{i,n-1}\varphi_{n-1} + k_{i,n}\varphi_n = f_i$$

Choosing the notation such that the superscript **0** represents the last iteration and the superscript **1** represents the new, Gauss-Seidel can be expressed as

$$\varphi_i^1 = \varphi_i^0 + \frac{\left[f_i - \sum_{j=1}^{i-1} k_{i,j}\varphi_j^1 - \sum_{j=i}^{n} k_{i,j}\varphi_j^0 \right]}{k_{i,i}}$$

Selective over-relaxation (SOR) uses an acceleration multiplier to speed convergence, resulting in the following refinement in Gauss-Seidel:

$$\varphi_i^1 = \varphi_i^0 + \omega \frac{\left[f_i - \sum_{j=1}^{i-1} k_{i,j}\varphi_j^1 - \sum_{j=i}^{n} k_{i,j}\varphi_j^0 \right]}{k_{i,i}}$$

where ω is chosen to maximize the rate of convergence. It can be shown that the method can only converge for $0 < \omega < 2$, and that diagonal dominance is required for stability and proper convergence in these methods.

III-3-3 Sparse Iterative Methods

Sparse iterative methods attempt to make ultimate use of the "space" within a typical finite element stiffness matrix, where that space is made up of zeros. The advantage in saving of space may be a factor of 100 or more over banded matrices. Typically one large vector contains the diagonal plus only the nonzero components. The locations of those nonzeros are preserved in a two column integer matrix containing the pointers to the stiffness matrix row and column positions. Iterations using Gauss-Seidel or other methods requires careful attention to the pointer array. Pointer arrays are developed more fully below.

III-3-4 Introducing Vectorized Multipliers in Iterative Methods

The problem of using a single acceleration multiplier as in SSOR is that not all equations have the same conditioning. That is, some equations are always more diagonally dominant than others. Those that are not diagonally dominant will tend

to diverge, while those that are diagonally dominant can be induced to converge more quickly. It has been found that by using a *vector* of accelerator multipliers, one term per equation instead of one term for the whole set of equations, convergence can be greatly accelerated while preserving stability.

In the method described below, Vectorized Sparse Selective Over Relaxation (VSSOR), an accelerator vector is combined with certain coding techniques that sacrifice a small amount of storage in the interests of speed. In test problems VSSOR ran 10 to 20 times faster than Gaussian elimination with bandwidth optimization.

In VSSOR, the diagonal is stored in vector ***diag***

$$\uparrow \atop n \atop \downarrow \left\{ \begin{bmatrix} \\ diag \\ \\ \end{bmatrix} \right\}$$

where ***diag[i]*** contains the floating point value of the diagonal on the ***i***th row of the matrix **[*k*]**. The matrix **[*a*]**, having dimensions

$$\begin{matrix} & \leftarrow nnrp \rightarrow \\ \uparrow & \\ n & \\ \downarrow & \end{matrix} \begin{bmatrix} \\ a \\ \\ \end{bmatrix}$$

such that ***a[i][j]*** contains the actual value of the off-diagonal matrix components where ***i*** is the row number and ***j*** is the arbitrary value of the position where the value is stored. There are ***nrp[i]*** such positions for each row ***i*** such that

$$nrp[\,i\,] \leq nnrp, \quad 1 < i < n$$

where ***nnrp*** is the upper limit of possible off-diagonal nonzeroes in a given equation, as determined by the stiffness matrix assembly overlay.

Also necessary is the integer matrix

$$\begin{matrix} & \leftarrow nnrp \rightarrow \\ \uparrow & \\ n & \\ \downarrow & \end{matrix} \begin{bmatrix} \\ nz \\ \\ \end{bmatrix}$$

where ***nz[i][j]*** contains the column position of ***a[i][j]*** in the original matrix. Also useful for limiting the number of operations is ***nbw***, the bandwidth of the original matrix.

The vector

$$\begin{matrix}\uparrow \\ n \\ \downarrow\end{matrix}\left\{\begin{matrix} \\ acc \\ \\ \end{matrix}\right\}$$

contains the floating point multipliers for the accelerating convergence such that each equation is iterated according to

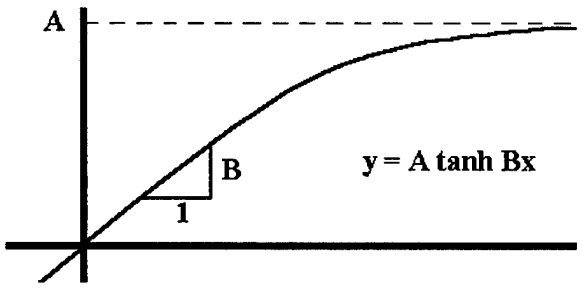
$$\varphi_i^1 = (1 - acc_i)\,\varphi_i^0 + acc_i \frac{\left[f_i - \sum_{j=1}^{i-1} k_{i,j}\,\varphi_j^1 - \sum_{j=i+1}^{n} k_{i,j}\,\varphi_j^0 \right]}{k_{i,i}}$$

where an effective expression for $\{acc\}$, keeping the value between 0 and 2.0 is

$$acc_i = A \tanh\left[\frac{|\,B\,a_{i,i}\,|}{\left[\sum_{j=1}^{i-1} |\,a_{i,j}\,| + \sum_{j=i+1}^{n} |\,a_{i,j}\,| \right]} \right]$$

with constants A and B having typical values of 2.0 and 0.5, respectively.

It can be seen from Figure 10 that the tanh function can be effective in controlling the accelerator multipliers, exploiting the relationship between the diagonal and off-diagonal terms while maintaining the range of acceleration multipliers between the prescribed values of 0 and 2.



VSSOR is expected to have additional benefits in speed on two fronts. First, in the full three dimensional problems to be faced in Phase II, bandwidth optimization has limited value in reducing the size of the global stiffness matrix. Sparse iterative methods then represent the only known techniques of limiting both storage requirements and number of arithmetic operations to manageable levels.

Second, the iterative coupling described in Section III-2 can combine with the iterative VSSOR methods to allow immediate updates of the entire system – not just an interim set of equations – as global iteration proceeds.

SECTION III-4

TEST CASES FOR THE TWO DIMENSIONAL AND TWO DIMENSIONAL AXISYMMETRIC FINITE ELEMENT PROGRAMS

III-4-1 Purpose of Testing

The computer programs based on the above derivations and algorithms are demonstrated below. The purpose is twofold. First, the programs are tested for accuracy in problems that have analytical results against which to compare. These problems are intended to test one feature at a time in a simple, unambiguous way. The obvious assumption is that the features which test successfully by themselves will also work together and produce accurate results in concert.

The second purpose of the program testing is to demonstrate multiple features simultaneously. Finding analytical results are difficult. In some cases the finite element model is presented as a qualitative example where no known solution by another technique is available.

III-4-2 JOULE HEATING IN A RECTANGULAR PIPED

This problem tests transient temperature evolution as a result of electrical conduction. A block of linearly conducting material is powered by applying a voltage across parallel surfaces, as shown in Figure III-13.

These values define the problem:

length (x)

L = 0.01 m

height (y)

H = 0.005 m

width (z)

W = 0.01 m

electrical conductivity

$\sigma = 20\ \Omega^{-1}\ m^{-1}$

thermal capacitance

$\rho C_p = 2.0\ e^6\ J/(m^3\ °C)$

applied voltage

V = 10 volts

time

t = 1.0 second

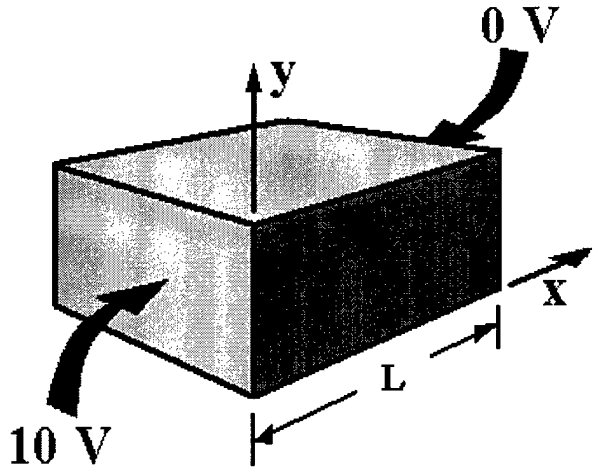


Figure III-13
Conductive block with applied voltage for Joule heating

As defined in Section III-2-11, local thermal input from electrical conduction relates to the local voltage field ξ and to the local electrical conductivity σ. Local current density is[1]

$$J = \sigma \xi$$

and local heat generation rate (per unit volume) is

$$q^* = \frac{J^2}{\sigma} = \sigma \xi^2$$

for the case where σ is a constant. Since the voltage in the test problem is applied uniformly across one face at 10V and across the parallel face at 0V, then the **y** and **z** components of ξ are zero and

$$\xi = 1000 \frac{V}{m}$$

then

$$q^* = \sigma \xi^2 = 2.0\,e^7 \frac{W}{m^3}$$

The rate of temperature rise everywhere within the material volume is inversely proportional to the thermal capacitance

$$\frac{dT}{dt} = \frac{q^*}{\rho C_p}$$

or, solving for constant temperature rise in linear material

$$\Delta T = \frac{q^* \Delta t}{\rho C_p}$$

For the instant problem this gives

$$\Delta T = \frac{(2.0\,e^7)(1.0)}{2.0\,e^6} \,^\circ C$$

Solymar, L., and Walsh, D., *Lectures on the Electrical Properties of Materials*, Fourth Edition, Oxford University Press, Oxford, 1988

or

$$\Delta T = 10\ ^{\circ}C$$

The total power generated in the block is

$$P = \int_{x=0}^{.01} \int_{y=0}^{.005} \int_{z=0}^{.01} q^{*}\, dx\, dy\, dz = 10\, W$$

Program *TVtranR* was used in conjunction with the finite element mesh shown at right in Figure III-14, employing triangular elements, to arrive at these same answers to *eight decimal places.*

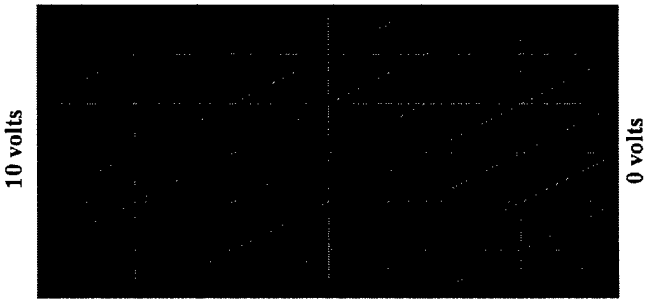


Figure III-14
Finite element grid for 2-D Joule heating problem

The temperature at the end of one second is a uniform 10°C throughout the block. The voltages attain a constant gradient from 0 to 10 volts as shown in Figure III-15.

This is a simple problem. It does serve to demonstrate the program capability with regard to basic fully coupled linear thermoelectrical phenomena.

This procedure will now be repeated with a similar example having axisymmetry.

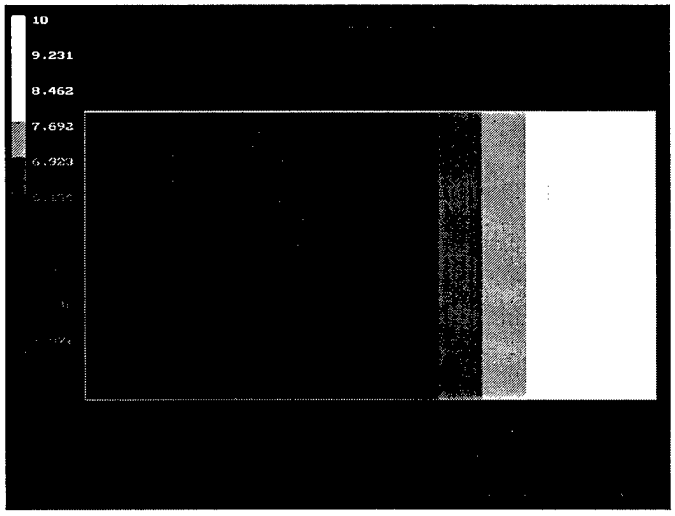


Figure III-15
Voltage distribution

III-4-3 JOULE HEATING IN A CYLINDER

In this problem, 10 volts are applied across opposing faces of a cylinder having a radius of 0.005m and a length of 0.01m, as shown in Figure III-16. Material properties are the same as the previous problem. Because of the similar geometry, the finite element grids could be shared by both sets of calculations. However, the axisymmetric code interprets the **x** direction as **z**, and the **y** direction as **r**.

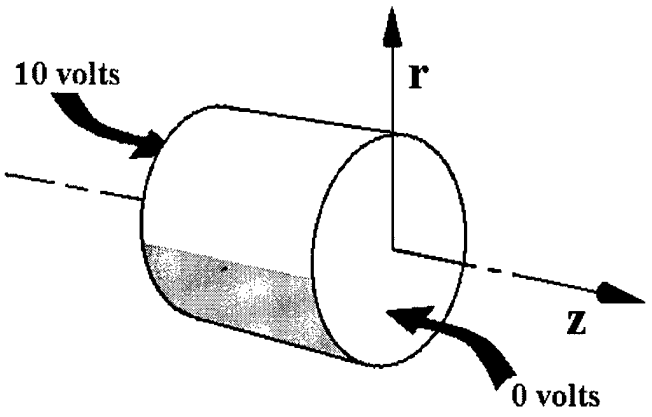


Figure III-16
Electrical conduction in a cylinder

As in the previous problem, the voltage applied across the same distance in the same material results in the same voltage gradient

$$\xi = 1000\frac{V}{m}$$

since $\sigma = 20\ \Omega^{-1}\mathbf{m}^{-1}$ as before, then

$$q^* = \sigma \xi^2 = 2.0\, e^7 \frac{W}{m^3}$$

Since the thermal properties also are unchanged, the one-second temperature rise is

$$\Delta T = 10\ ^\circ C$$

as before. Program *axiTVTR* supplies these answers accurate to eight decimal place precision. See Figure III-17.

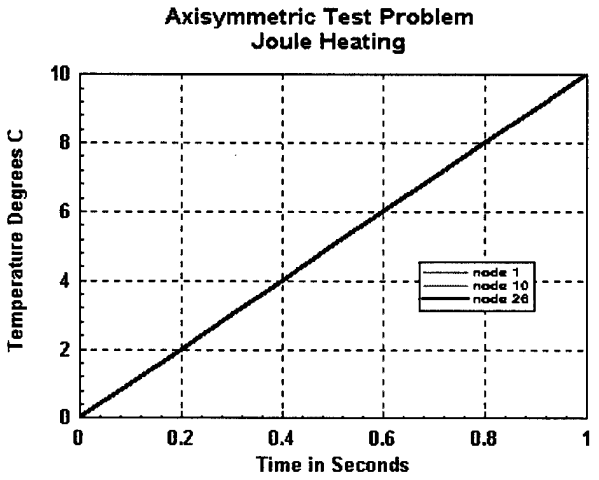


Figure III-17

Although the power density for the two problems is the same, the total power, being generated over two different geometries, differs. For the cylinder,

$$P = \int_{\theta=0}^{2\pi} \int_{r=0}^{.005} \int_{z=0}^{.01} q^* \, d\theta \, dr \, dz = 15.708 \, W$$

is also accurately calculated by program *axiTVTR*.

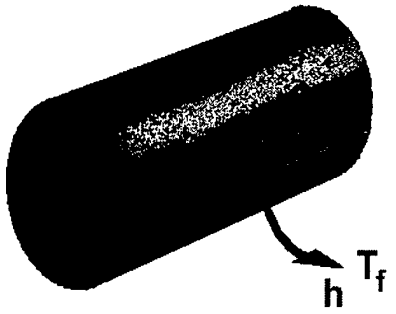


Figure III-18

III-4-4 CONVECTIVE COOLING IN A CYLINDER

This problem tests program *axiTVTR* capability in transient, convective gain/loss in an axisymmetric domain. Here, a long stainless steel cylinder at an initial uniform temperature T_i is subjected to contact with a fluid at temperature T_∞ with linear convection constant h. See Figure III-18.

Heat flux at the surface of the cylinder is

$$q = h A (T_s - T_\infty)$$

where T_s is the local surface temperature and A is the area of convective contact. The relevant parameters are:

h	=	100 BTU/(hr ft^2 °F)
T_∞	=	100 °F
R	=	4.0 in
Ti	=	1800 °F
k	=	9.4 BTU/(hr ft °F)
ρ	=	488 lb/ft^3
C_p	=	0.11 BTU/(lb°F)

The analytical solution for the transient case is

$$T = T_\infty + (T_i - T_\infty) \, 2 \sum_{n=1}^{\infty} \frac{1}{\lambda_n R} e^{-\lambda_n^2 \alpha t} \frac{J_0(\lambda_n r) \, J_1(\lambda_n R)}{J_0^2(\lambda_n R) + J_1^2(\lambda_n R)}$$

where:

α is the thermal diffusivity

$$\alpha = \frac{k}{\rho C_p}$$

J_i is the Bessel function of order **I**

λ_n is the nth root of

$$\lambda_n R \frac{J_1(\lambda_n R)}{J_0(\lambda_n R)} - \frac{hR}{k} = 0$$

Solutions to problems defined by the above set of equations are often found graphically from dimensionless plots of these expressions. Comparisons of four temperatures calculated by this quite approximate approach nevertheless achieves good agreement with the finite element results using program *axiTVTR,* as shown in Table III-2.

Time (hr)	Surface Temperature (°F)		Center Temperature (°F)	
	analytical	*axiTVTR*	analytical	*axiTVTR*
0.123	500	502.7	1324	1326.5
0.332	218	217.9	500	504.2

Table III-2
Comparison of Analytical and Finite Element Solutions

The finite element solution is shown graphically as a surface in Figure III-19.

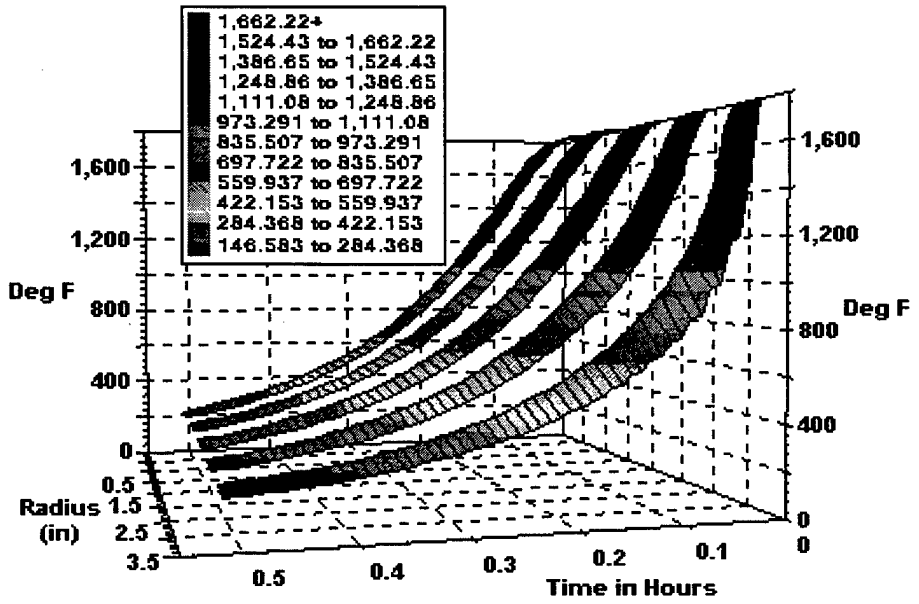


Figure III-19
Axisymmetric Convection Problem: Finite Element Results

III-4-5 JOULE HEATING, VOLTAGE APPLIED THROUGH RESISTOR

This problem is similar to the Joule heating example of Section III-4-2, except that copper electrodes are employed at each end of the block of material, and voltage is applied to one of the electrodes through a 10Ω ballast resistor. The opposite electrode is grounded. See Figure III-20.

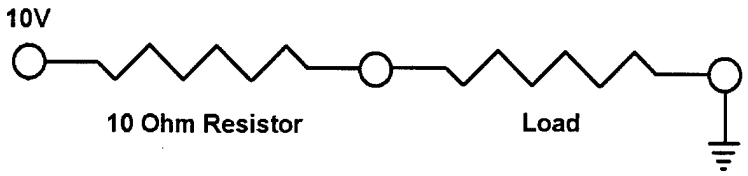


Figure III-20

The finite element grid is shown in Figure III-21. Note that a special feature of program *TvtranR* is employed that permits voltage to be applied uniformly to a group

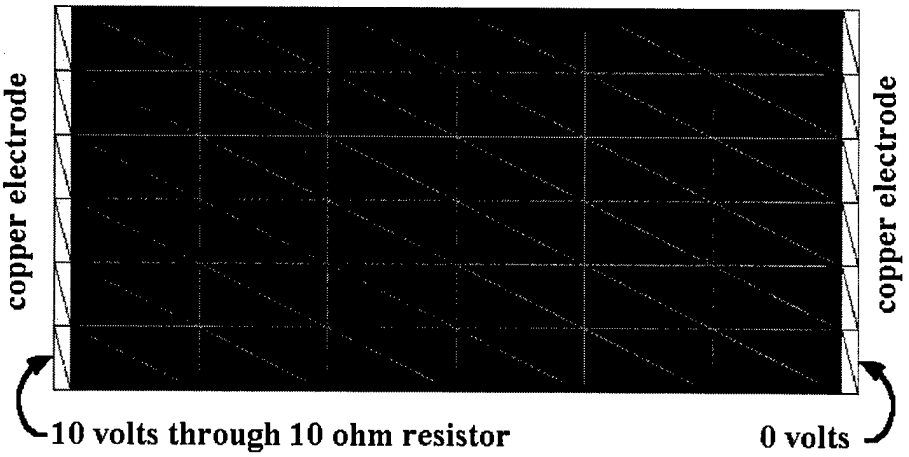


Figure III-21
Finite element grid showing copper electrodes

of elements with a specified in-line ballast resistor.

The effect of the copper electrodes on voltage drop should be negligible, but the ballast resistor should halve the voltages in the block of material and should therefore reduce the power dissipation and rate of temperature rise by a factor of four. The results are: P = 2.500 watts,
uniform temperature rise over 1 sec = 2.500 °C,
voltage varies uniformly from 0 to 5.000 volts.

All are correct to eight decimal places.

III-4-6 JOULE HEATING, BALLAST RESISTOR, GAUSSIAN-DISTRIBUTED SPATIAL RANDOMNESS IN MATERIAL

This problem is identical to the example of III-4-5, except that material randomness is introduced. The element-to-element electrical conductivity is made to vary according to a Gaussian distribution with the mean of that distribution equal to the nominal conductivity for that material. That distribution is then imposed spatially in a random manner employing the technique discussed in Section III-2.

For this problem, the standard deviation of electrical conductivity multiplier is set to $0.001\Omega^{-1}m^{-1}$, and the hypothetical sample size used to discover that standard deviation is set to $1.0e^{-6}$ m^3, equivalent to $1.0e^{-4}$ m^2 area in this two dimensional problem.

The resulting temperature patterns are not the uniform 2.500°C of the non-random model as shown in Figure III-22. Instead these temperatures vary from approximately 2.4°C to 2.6°C, the range used to get the spectacular variation in Figures III-23 through III-26. All are actual results from the same boundary and initial conditions, but the random numbers for the material randomness in each case uses as a seed the time of day in milliseconds. Using this technique, getting identical results from any two runs is extremely unlikely. The seeding can be forced to a specific value at the user's discretion, however.

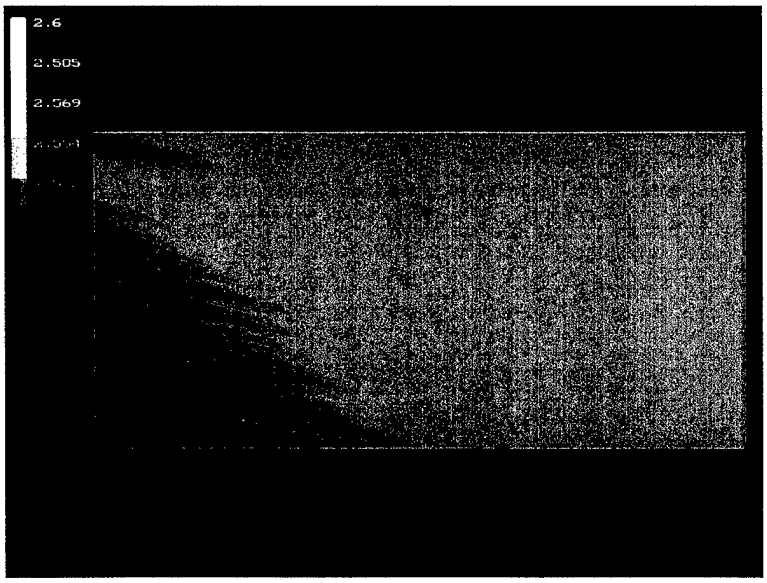


Figure III-22
Finite element solution of Joule heating in block
with voltage applied through ballast resistor.
No randomness in material.

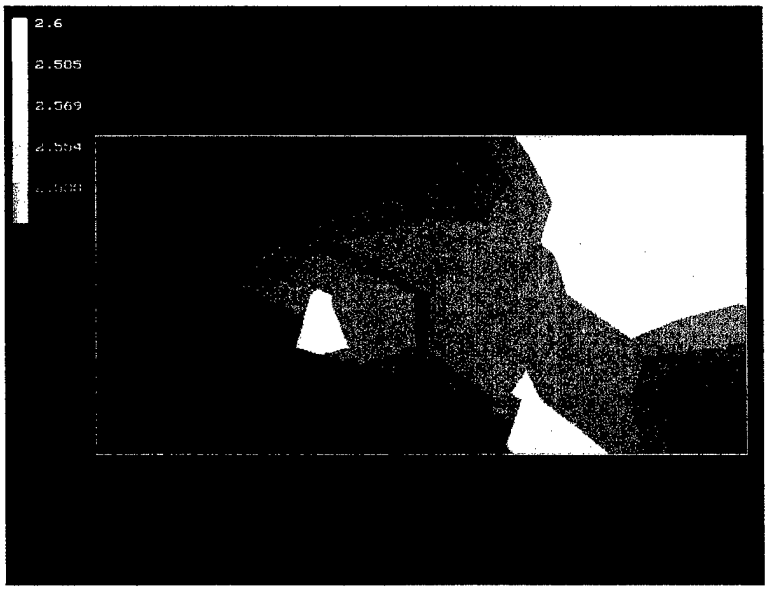


Figure III-23
Finite element solution of Joule heating in block
with voltage applied through ballast resistor.
Randomness in material, case 1

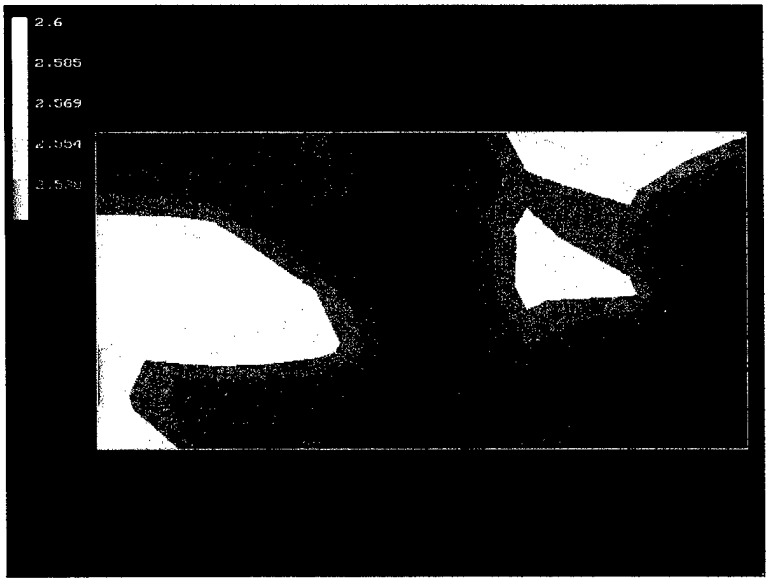


Figure III-24
Finite element solution of Joule heating in block
with voltage applied through ballast resistor.
Randomness in material, case 2

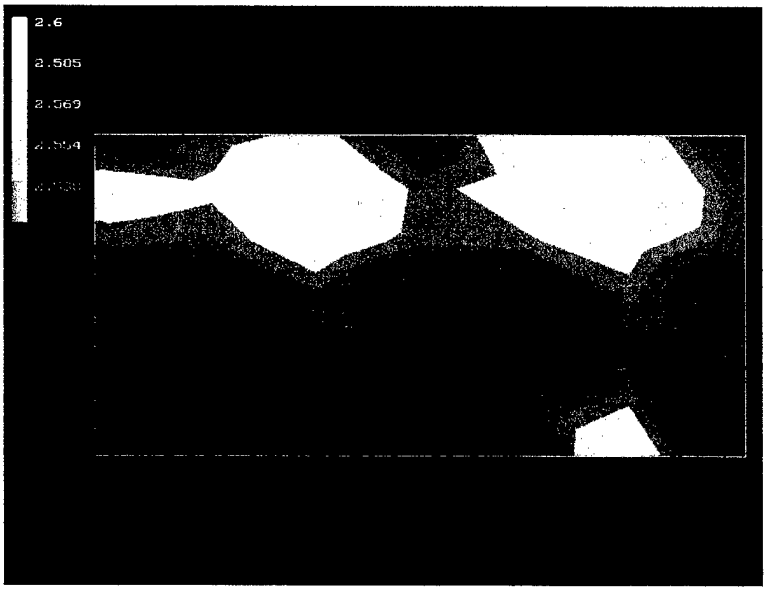


Figure III-25
Finite element solution of Joule heating in block
with voltage applied through ballast resistor.
Randomness in material, case 3

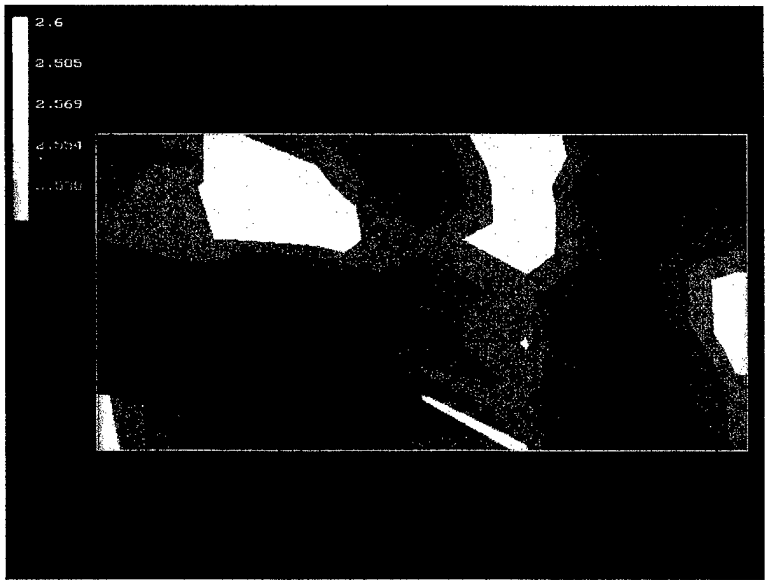


Figure III-26
Finite element solution of Joule heating in block
with voltage applied through ballast resistor.
Randomness in material, case 4

III-4-7 CYLINDER EMBEDDED IN MEDIUM WITH UNIFORM ASYMPTOTIC FIELD

The following is a solution to the test problem proposed in Section II-6-6, where a cylinder of conductivity σ_1 is embedded in a medium having conductivity σ_2 with a uniform asymptotic field applied at infinity. We choose the cylinder axis to be parallel with the z axis and the asymptotic field parallel to the x axis. Addressing the problem in two dimensions and choosing axes such that in both Cartesian and cylindrical coordinates the origin is at the center of the embedded cylinder ($\mathbf{x} = \mathbf{y} = 0$ and $\mathbf{r} = 0$ respectively), then, as reported in Section II-6-6, the asymptotic field at infinity is

$$\boldsymbol{E}_{2_\infty} = \hat{\boldsymbol{x}}\,\boldsymbol{E}$$

with potential function

$$\boldsymbol{V}_{2_\infty} = -\boldsymbol{E}\boldsymbol{x} = -\boldsymbol{E}\,\boldsymbol{r}\cos\phi$$

Results should give a field within the cylinder of

$$\boldsymbol{E}_{1_r} = \boldsymbol{E}_{2_\infty}\frac{2\,\sigma_2}{\sigma_2 + \sigma_1}\cos\phi$$

which gives in the x-direction ($\cos\phi = 1$)

$$\boldsymbol{E}_{1_x} = \boldsymbol{E}_{2_\infty}\frac{2\,\sigma_2}{\sigma_2 + \sigma_1}$$

which is a constant value within the cylinder, and in the y direction the field is also a constant within the cylinder

$$\boldsymbol{E}_{1_y} = 0$$

In the finite element solution, good approximation of the above result is obtained by choosing the solution domain to extend to a distance of ***5r***, or five times the radius of the cylinder. See Figure III-27.

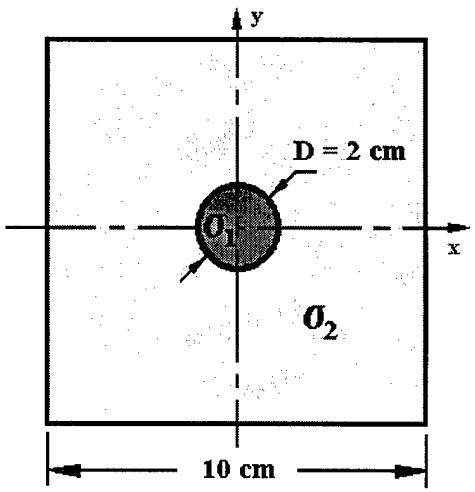


FIGURE III-27

The finite element solution is further simplified by taking advantage of the two planes of symmetry in the problem. The solution domain then appears as shown in Figure III-28. Dirichlet conditions are then imposed as voltages:

$$V = 10.0\,V \quad ; \quad x = -5\,r$$

and

$$V = 0.0\,V \quad ; \quad x = 0$$

giving a field

$$E_{2_\infty} \approx \frac{10.0V}{5\,cm} = 2.0\,\frac{V}{cm}$$

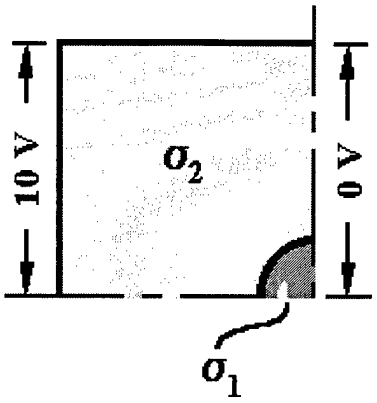


Figure III-28

The Neumann conditions are

$$\frac{\partial V}{\partial y} = 0 \quad ; \quad y = 5r$$

and

$$\frac{\partial V}{\partial y} = 0 \quad ; \quad y = 0$$

also implied from symmetry is

$$\frac{\partial^2 V}{\partial^2 x} = 0 \quad ; \quad x = 0$$

The finite element grid for this problem is shown in Figure III-29.

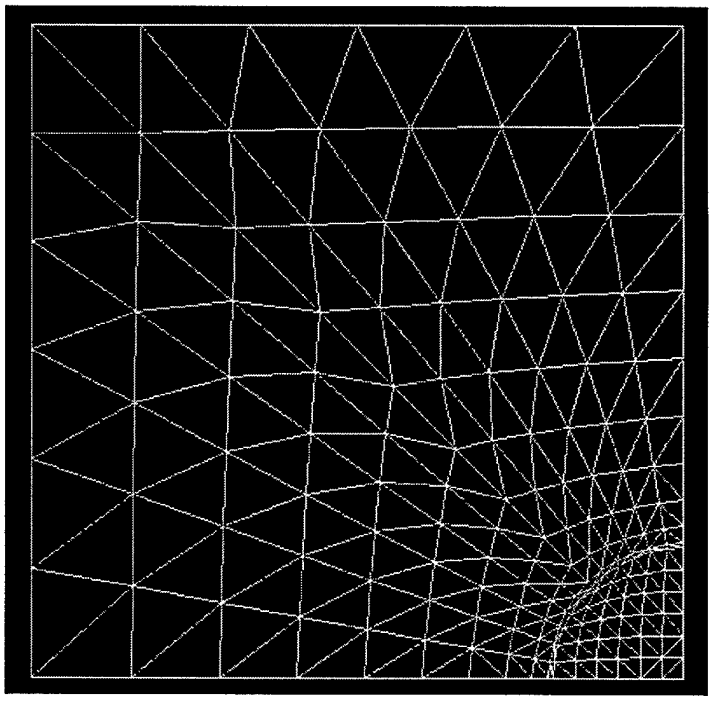


Figure III-29
Finite Element Grid

Finite element results are shown graphically in Figure III-30 for the case where $\sigma_1 = 1.0\ \Omega^{-1}cm^{-1}$ and $\sigma_2 = 10.0\ \Omega^{-1}cm^{-1}$. A close-up of the general vicinity of the cylinder is shown in figure III-31, where voltage resolution is also enhanced.

Finite element results for the case where $\sigma_1 = 10.0\ \Omega^{-1}cm^{-1}$ and $\sigma_2 = 1.0\ \Omega^{-1}cm^{-1}$ are shown in Figures III-32 and III-33. Note that the voltages within the cylinder do not vary in y, and that the gradient is constant in x, in complete agreement with the analytical expressions.

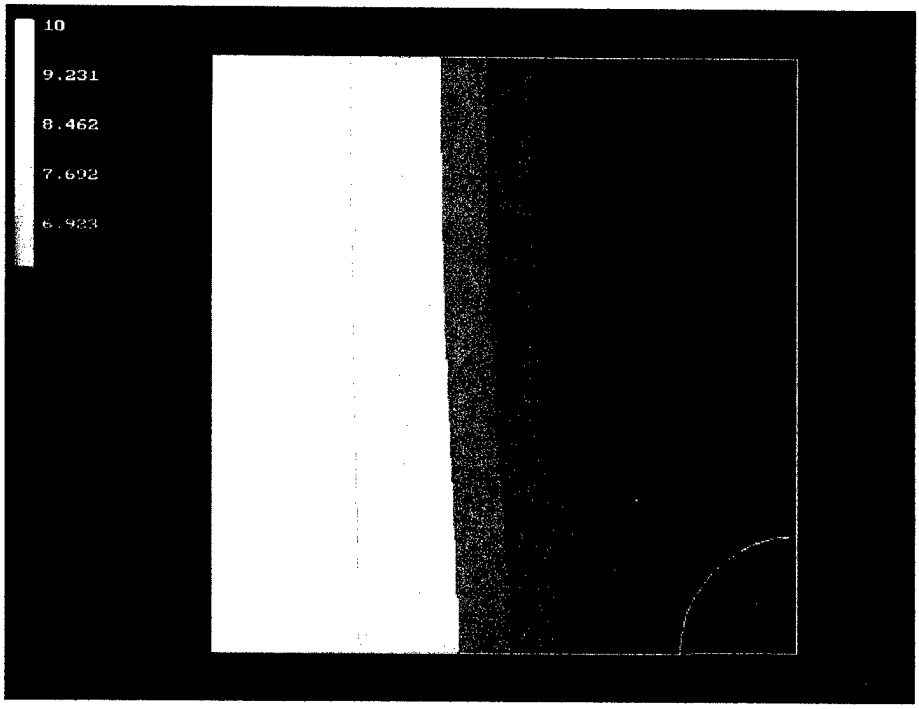


Figure III-30

Test case, cylinder in infinite medium: $\sigma_1 = 1.0\ \Omega^{-1}cm^{-1}$, $\sigma_2 = 10.0\ \Omega^{-1}cm^{-1}$.

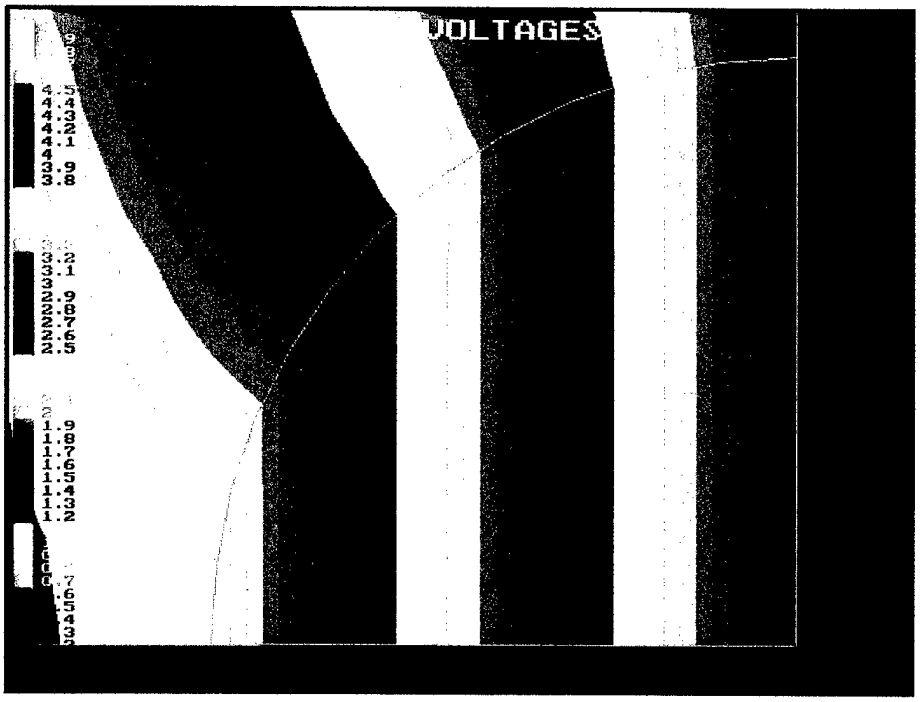


Figure III-31

Test case, cylinder in infinite medium: $\sigma_1 = 1.0\ \Omega^{-1}cm^{-1}$, $\sigma_2 = 10.0\ \Omega^{-1}cm^{-1}$.
Zoomed view.

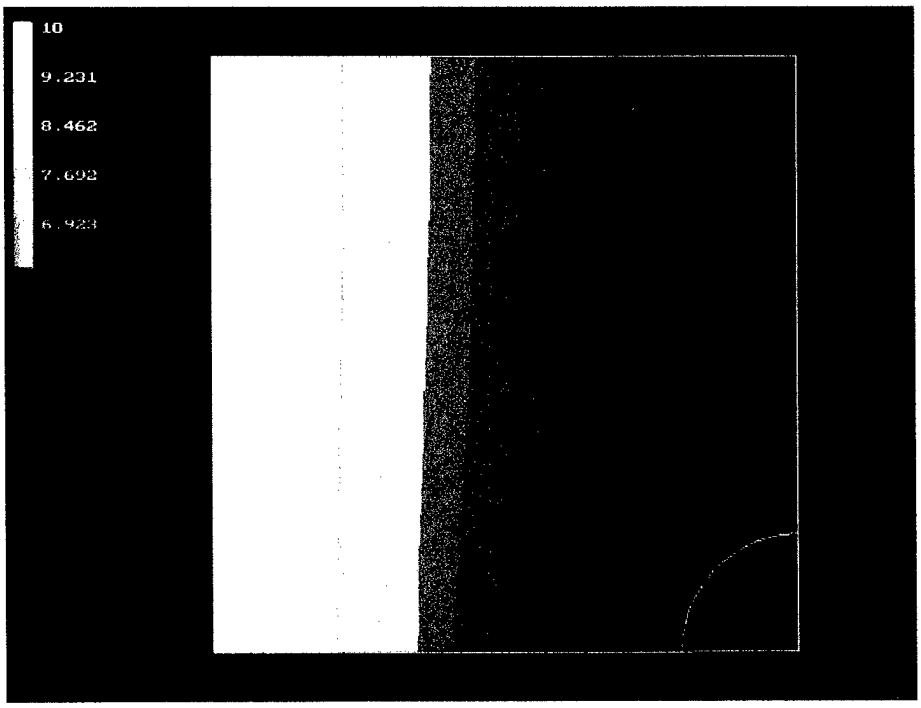


Figure III-32

Test case, cylinder in infinite medium: $\sigma_1 = 10.0\ \Omega^{-1}cm^{-1}$, $\sigma_2 = 1.0\ \Omega^{-1}cm^{-1}$.

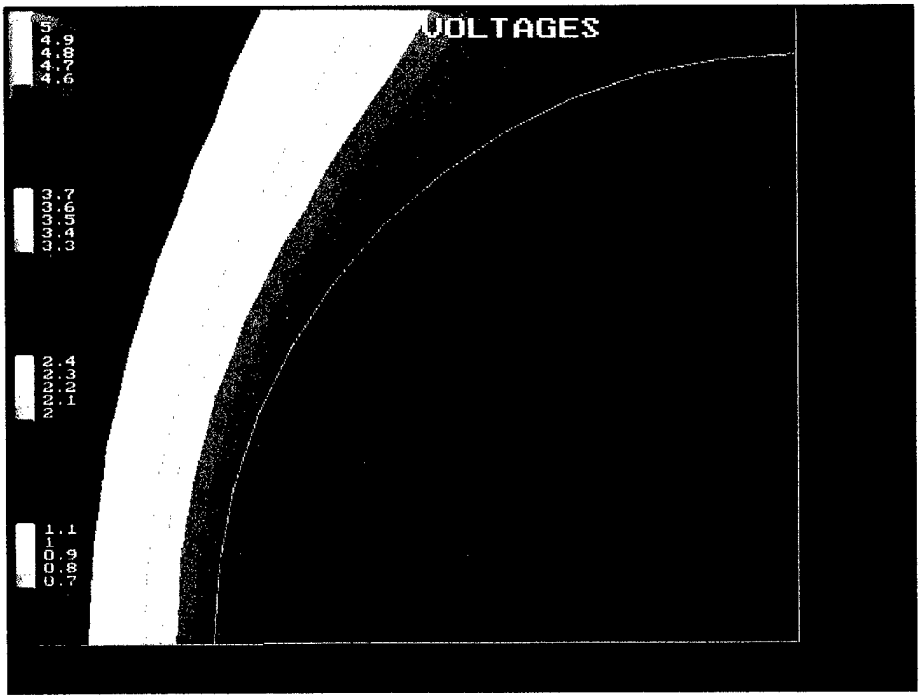


Figure III-33

Test case, cylinder in infinite medium: $\sigma_1 = 10.0\ \Omega^{-1}cm^{-1}$, $\sigma_2 = 1.0\ \Omega^{-1}cm^{-1}$.
Zoomed View.

Comparison of the analytical and finite element results is shown in Table III-3. Slight departure from the analytical results can be attributed to the departure from "infinite" domain: the finite element solution domain extends only to $r = 5.0$ in the x and y directions. This example should be viewed as a strength of the finite element technique, since arbitrary shape changes and departures from classical solutions can be accommodated easily.

Table III-3
Analytical and Finite Element Results Comparison
Voltage Values

Location	σ_1	σ_2	Analytical	Finite Element
Center of Cylinder	1.0	10.0	0.0000	0.0000
x = -1.0, y = 0.	1.0	10.0	3.6363	3.5158
x = -1.0, y = 0.	10.0	1.0	0.3636	0.3766
x = -1.64, y = 0.	1.0	10.0	4.2778	4.1503
x = -1.64, y = 0.	10.0	1.0	2.2822	2.3478

The above voltage results represent only half of the solution to the embedded cylinder problem provided by program *TVtranR*. Coincident with the development of voltage patterns is the coupled transient thermal problem. These thermal properties were applied to materials 1 and 2:

$$\rho_1 = \rho_2 = 20.0 \text{ kg/m}^3 \quad \mathbf{C}_{p1} = \mathbf{C}_{p2} = 10.0 \text{ J/kg°C} \quad \mathbf{k}_1 = \mathbf{k}_1 = 1.0 \text{ W/m°C}$$

with initial conditions:

temperature $\mathbf{T} = 0.0$ everywhere at time $\mathbf{t} = 0$.

The resulting temperature patterns for the above two cases after one second of heating are shown in Figures III-34 and III-35. Note that the heating patterns for the two cases are almost perfect mirror images of each other.

In areas of the cylinder not subject to thermal influence from material 2, temperature rise calculated from the simple expression for Joule heating

$$\Delta T = \frac{P}{\rho C_p} = \frac{E^2 \sigma}{\rho C_p}$$

agree with temperature output from the finite element calculations to four decimal places.

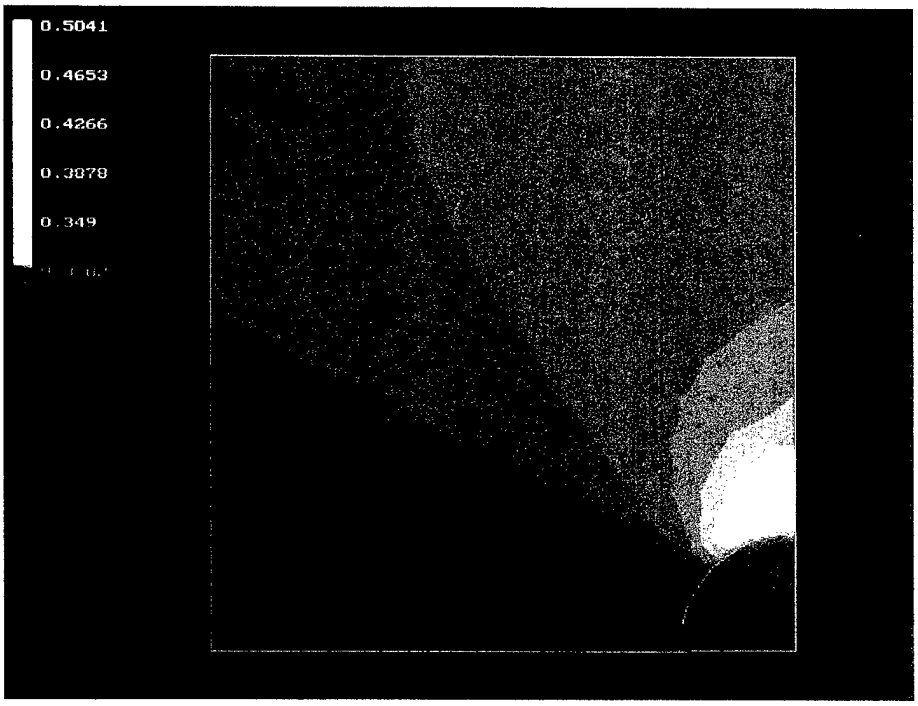


Figure III-34

Test case, cylinder in infinite medium: $\sigma_1 = 1.0\ \Omega^{-1}cm^{-1}$, $\sigma_2 = 10.0\ \Omega^{-1}cm^{-1}$.

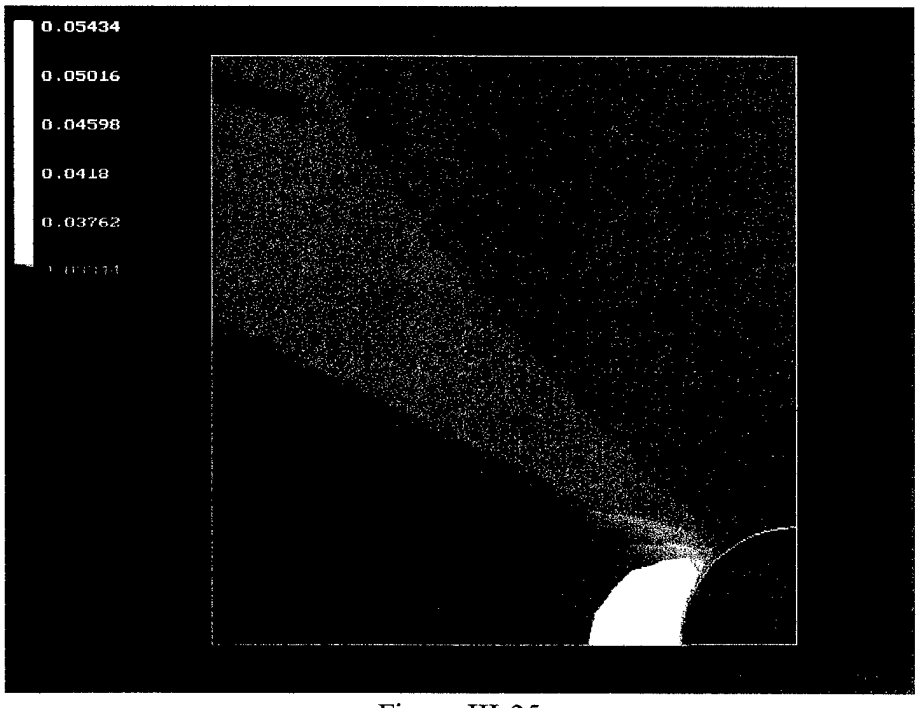


Figure III-35

Test case, cylinder in infinite medium: $\sigma_1 = 10.0\ \Omega^{-1}cm^{-1}$, $\sigma_2 = 1.0\ \Omega^{-1}cm^{-1}$.

III-4-8 TRANSIENT OVERVOLTAGE APPLIED TO CUBE HAVING QUANTUM TUNNELLING NONLINEARITIES

This test was proposed in section II-6-7 as confirmation of the nonlinear capabilities of the finite element programs. The modelled object is a 1.0 cm cube of nonlinear varistor material having 0.1 cm copper electrodes on top and bottom. The bottom electrode is grounded. At time zero 40KV is applied to the top electrode through a 25Ω resistor. See Figure III-36. Temperature at time zero is 0°C. The conductive medium is assumed to have the thermal properties of copper. Because the time frame of the model is so short (<50μs), it is expected that the diffusion of elevated temperatures into the binders will be minimum. The heat capacity ρC_p of the varistor material is therefore assumed to be that of copper multiplied by the volume loading factor of 0.524, giving $\rho C_p = (0.524)(8.92\ g/cm^3)(0.385\ J/g\text{-}°C)$.

The material conductivity is given locally (within each control volume as defined by individual finite elements) by the Nordheim-Fowler approximation derived in Sections II-1-9 and II-6-7:

$$\sigma = \frac{K_E}{(N g)^2} E\, e^{-N g \beta_E / E}$$

where $K_E = 6.4 \times 10^{-7}\ A/V^2$
$\beta_E = 6.3 \times 10^6\ V/cm$
$N = 100\ cm^{-1}$
$g = 100\ Å = 10^{-6}\ cm$

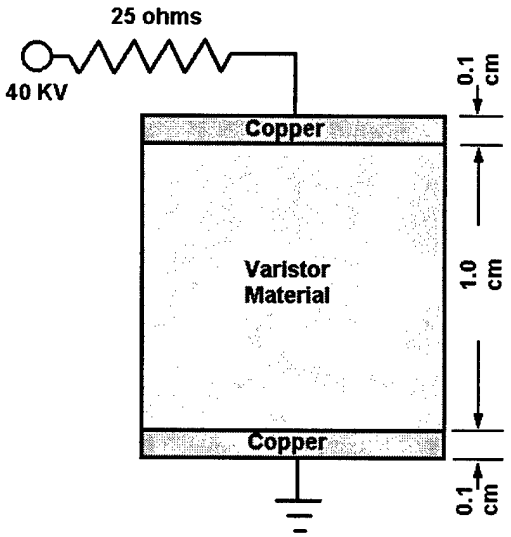


Figure III-36

and E is determined locally in space (within each element) and time (each time step) by the finite element program solving the modified Laplacian. Time steps were chosen to be 100 ns.

The overall solution scheme is shown as a flow chart in Figure III-5. Note that in this preliminary example there are two simplifications. First, there is no direct relationship between electrical conductivity and temperature in the above expression for σ.

Second, the statistical variability in the material is assumed to be zero, which can be expressed as a zero value for the standard deviation

$$\sigma_{P(e)} = 0$$

The finite element discretization for this problem has 3111 nodes and 6000 elements, as shown in Figure III-37. Five rows of elements were used for copper electrodes on both top and bottom.

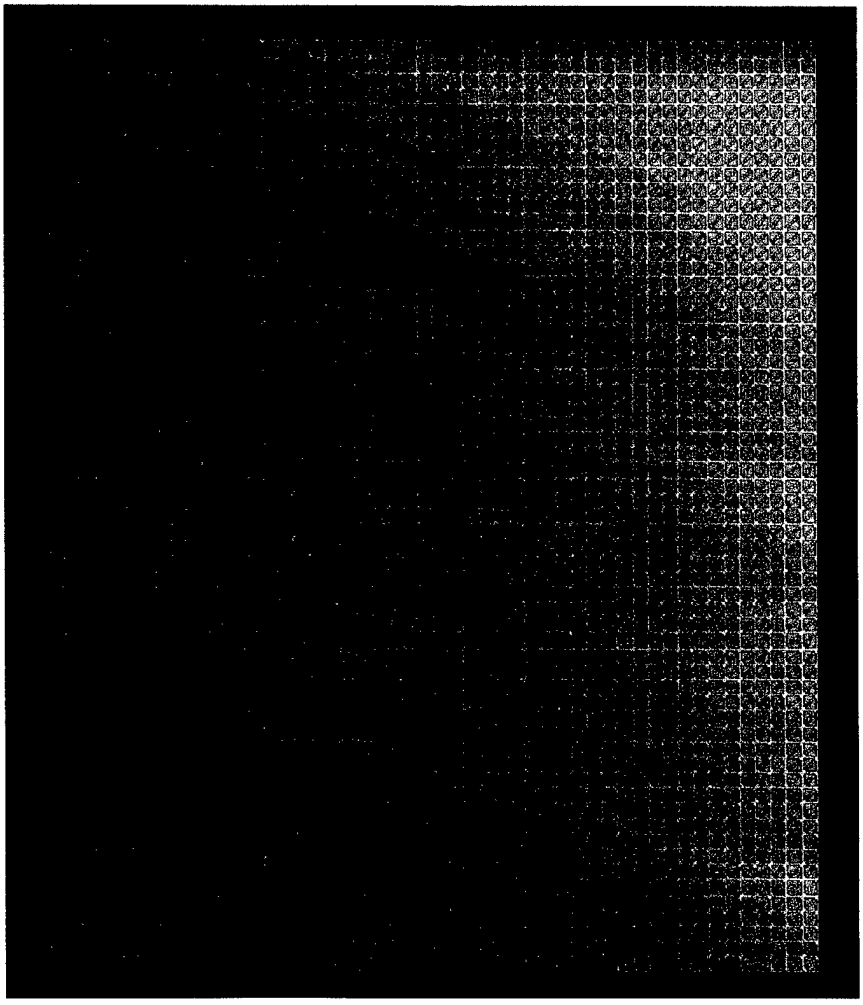


Figure III-37

The results show a uniform voltage drop across the varistor material of 4891 volts shown graphically in Figure III-38, a uniform temperature rise to 100°C in 26.18 μs, and total power generated of 6.869 MW. All are in agreement with the analytical solution.

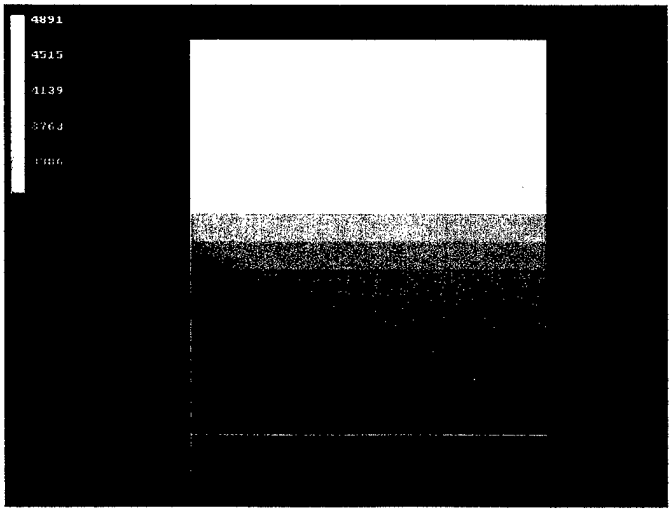


Figure III-38

Resulting temperature pattern at time of 24μs is shown in Figure III-39. Note that the higher thermal conductivity of the copper electrodes results in a slightly lower temperature of the varistor material just at the interface with the copper.

Although this graphic representation of temperature is relatively uninteresting in itself, it is included here for contrast with four figures to follow.

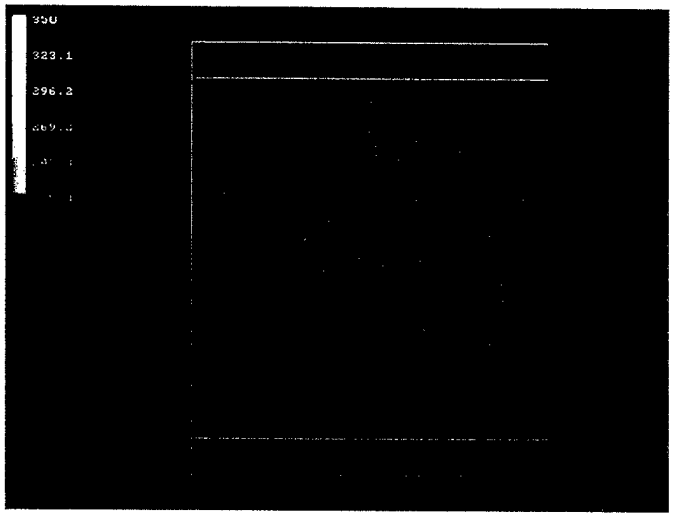


Figure III-39
Temperature pattern at 24μs with no statistical variation in particle gaps.

In these next four problems, statistical variability is introduced in the form of nonuniform spacing between conductors, represented by parameter ***g***. The assumption in these models is that the conductive spheres have been coated with an insulative material having a thickness of 50Å. This means that the minimum spacing between spheres is 100Å, and that the distribution of spacings among spheres is likely to be a mirrored Gaussian distribution as shown in Figure III-40.

Four different choices are made for the standard deviation of the Gaussian distribution: 10Å, 20Å, 50Å, and 100Å. The resulting calculated power increases with increasing gap variability, yielding 7.1 MW, 7.3 MW, 7.9 MW, and 8.6 MW, respectively. The resulting temperature patterns are shown in Figures III-41 through III-44. Note that the patterns of higher temperature appear to be increasingly organized in channels perpendicular to the electrodes as gap variability increases, a phenomenon which suggests increasing tendency to form dominant current paths.

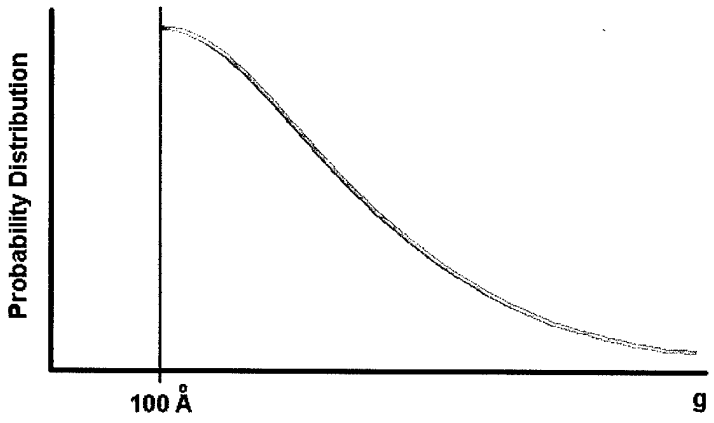


Figure III-40
Mirrored Gaussian distribution of particle gaps

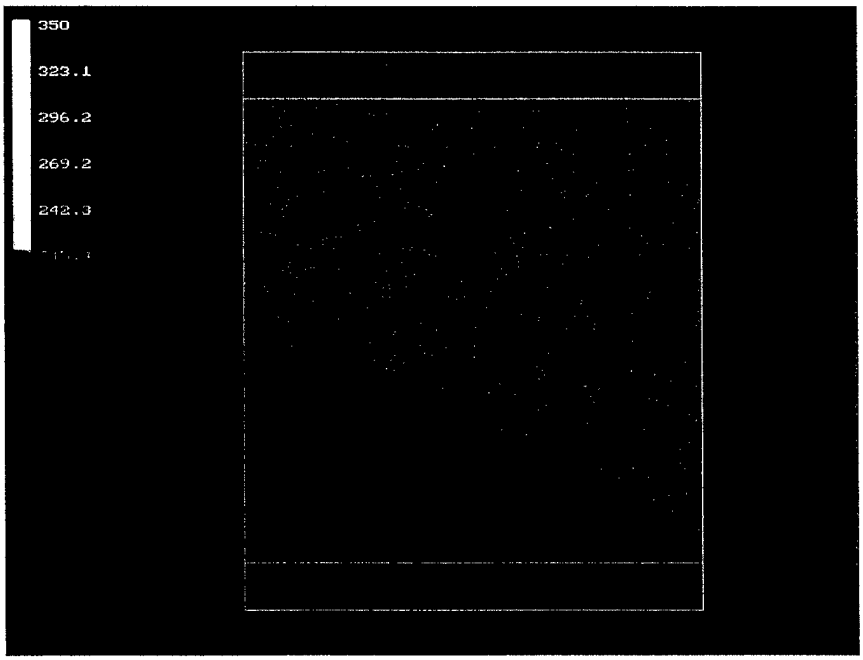


Figure III-41 Standard deviation = 10 Å

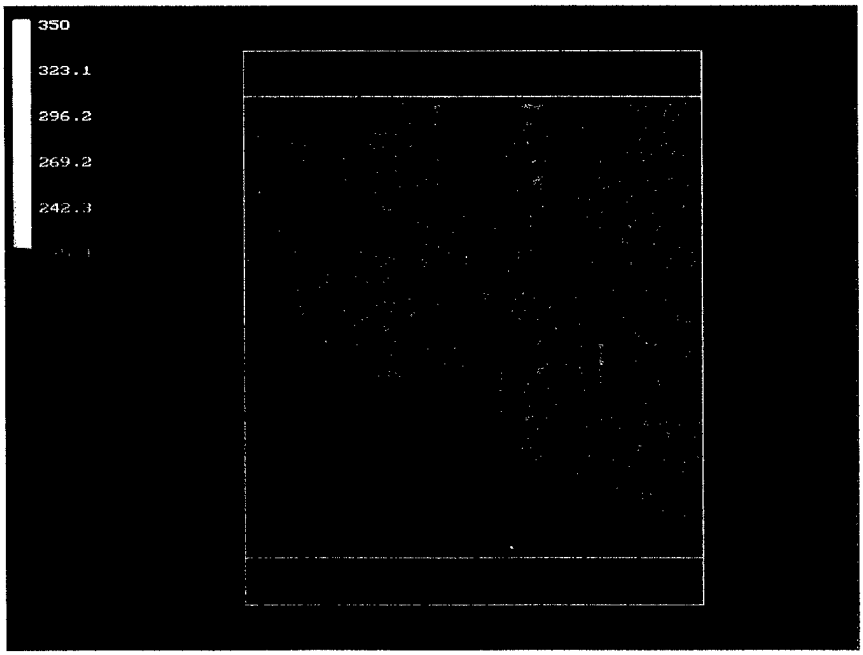


Figure III-42 Standard deviation = 20 Å

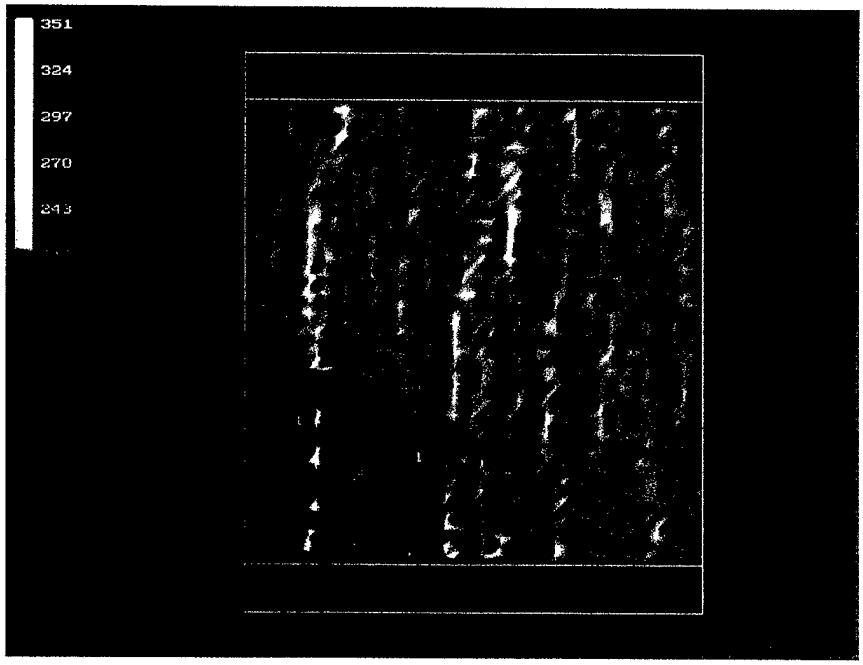


Figure III-43 Standard deviation = 50 Å

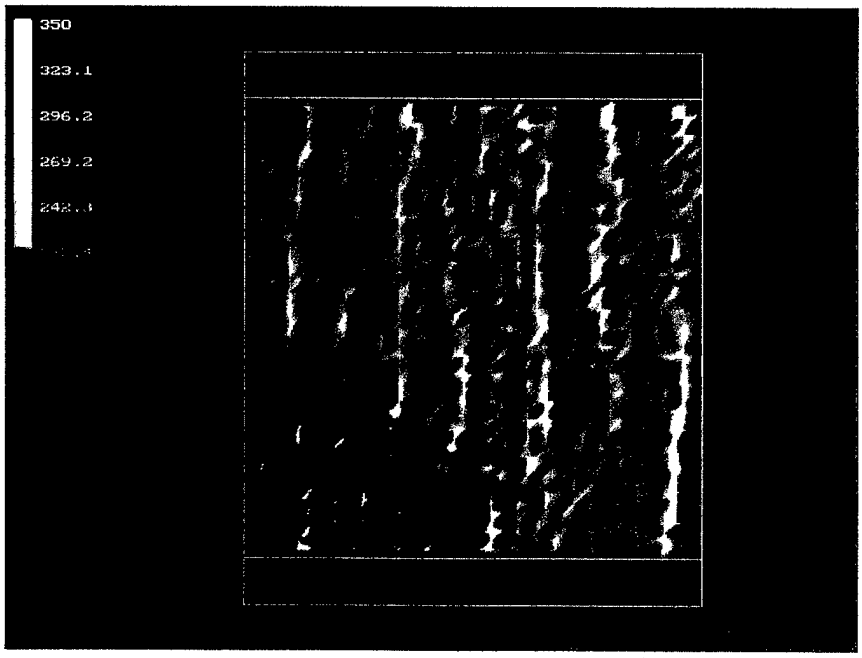


Figure III-44 Standard deviation = 100 Å

SECTION III-5

MODELLING DOMINANT PATH RESPONSE TO TRANSIENT OVERVOLTAGE

III-5-1: NECESSARY CONDITIONS FOR DOMINANT PATH

The last four test cases presented in Section III-4-8 (Figures III-41 to III-44) exhibit nonuniform heating as a result of statistical variability of the material. However, even though those examples seem to visually demonstrate both warmer regions and an alignment of these warmer regions normal to the electrodes, *none of those examples exhibit dominant path failure*. This is an important finding in that experimental test samples appear to fail in dominant path.

The model requirements for dominant path seem to be satisfied. First, the material is unstable, with marked voltage gradient sensitivity as defined by the Nordheim-Fowler approximation derived in Sections II-1-9 and II-6-7 and repeated here:

$$\sigma = \frac{K_E}{(N g)^2} E e^{-N g \beta_E / E}$$

where, for the test cases, these parameters are used:

$K_E = 6.4 \times 10^{-7}$ A/V^2
$\beta_E = 6.3 \times 10^6$ V/cm
$N = 100$ cm^{-1}
$g = 100$ Å $= 10^{-6}$ cm

Second, the material is spatially varied with randomness approximating a mirrored Gaussian distribution of gaps between conducive particles. Four values of standard deviation for this distribution are modelled to simulate increasing variability among gaps, all without exhibiting dominant path.

A third condition for dominant appears to be required. This assumption of a necessary third condition has led to an hypothesis: the related third condition for dominant path formation is a *regenerative effect.* That is, some mechanism must exist to impose higher local conductivity σ as a result of higher current density J. Higher local conductivity in turn leads to higher current density which leads again to even higher conductivity, etc. The modified Nordheim-Fowler approximation might then be expressed

$$\sigma = \frac{K_E}{(N g)^2} E e^{-N g \beta_E / E} f(J)$$

where

$$J = \sigma E$$

Candidates for the regenerative function $f(J)$ could be inductive effects or other mechanisms related to the electromagnetic problem alone, such as regenerative avalanche. Other candidates for regenerative function are temperature-related phenomena emerging from thermoelectric coupling. In the latter case, where some aspect of thermoelectric coupling provides the mechanism for regenerative effect, the Nordheim-Fowler approximation with would take the form

$$\sigma = \frac{K_E}{(N g)^2} E e^{-N g \beta_E / E} f(T)$$

For the models that follow in this section, the regenerative effect is chosen to be temperature-dependent as in this last expression. Note that the two options are closely related, since temperature T relates to current density J by

$$\Delta T = \frac{P}{\rho C_p} = \frac{E J}{\rho C_p}$$

This relationship suggests that with temperature as the basis for regenerative effect the speed of response is gated by the values of ρC_p, thereby slowing the development of dominant path phenomena when compared to a similar function of J alone.

The following models are chosen to test the hypothesis that regeneration is a requirement for dominant path. An overvoltage transient is simulated with 40KV applied at time $t = 0$ to one electrode through a 25 ohm resistor as shown in Figure III-45. The other electrode is grounded. These models also employ the same finite element discretization used in previous models, having 3111 nodes and 6000 elements as shown above in Figure III-37.

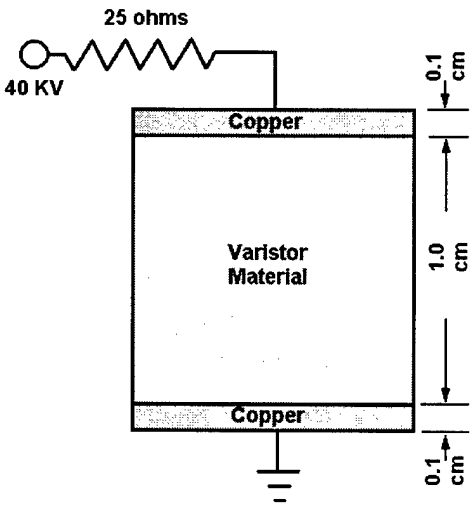


Figure III-45

All models are calculated for the first 20μs of the overvoltage transient only, using time steps of 100ns. This means that each model contains 200 discrete solutions in time. Each time step in turn typically requires between 200 and 500 iterations depending on the severity of the nonlinear event occurring during a particular time step. This means that a typical 20μs model can involve up to 10,000 solutions of the system of equations containing 3111 nodes in voltage and temperature, or 6222 total unknowns.

III-5-2: LINEAR REGENERATIVE MODELS

In the first set of three models, the regenerative function is chosen to be

$$f(T) = 1.0 + \alpha (T - T_i)$$

where T = local instantaneous temperature
T_i = local starting temperature
α = linear slope of temperature effect

for each element in the finite element domain, such that

$$\sigma_{(e)} = \frac{K_E}{(Ng)^2} E_{(e)} e^{-Ng\beta_E/E} \left(1.0 + \alpha (T_{(e)} - T_i) \right)$$

where the subscript (e) refers to a constant or average (in the case of temperature) value across a single element. The values for K_E, β_E, N, and g retain the values listed above. T_i equals 0 and the standard deviation for spatial randomness equals 40Å in all models.

Temperature results are shown in Figures III-46, III-47, and III-48 for the α values of 0.001, 0.010, and 0.100, respectively. The scale used is 13 colors over a range of 0°C to 500°C. Time step shown is at 20μs for all.

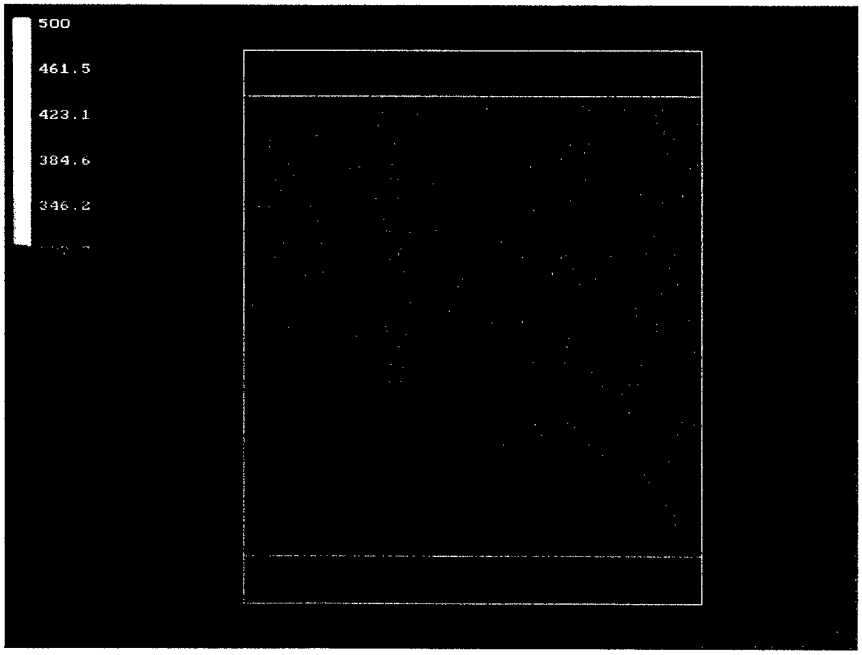


Figure III-46
Linear Regenerative Model at 20μs, α = 0.001

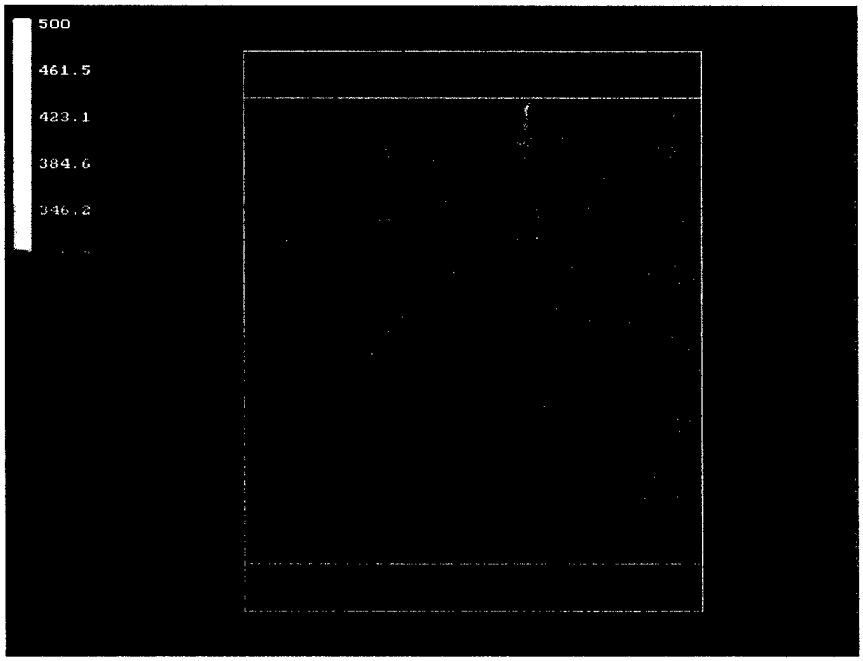


Figure III-47
Linear Regenerative Model at 20μs, $\alpha = 0.010$

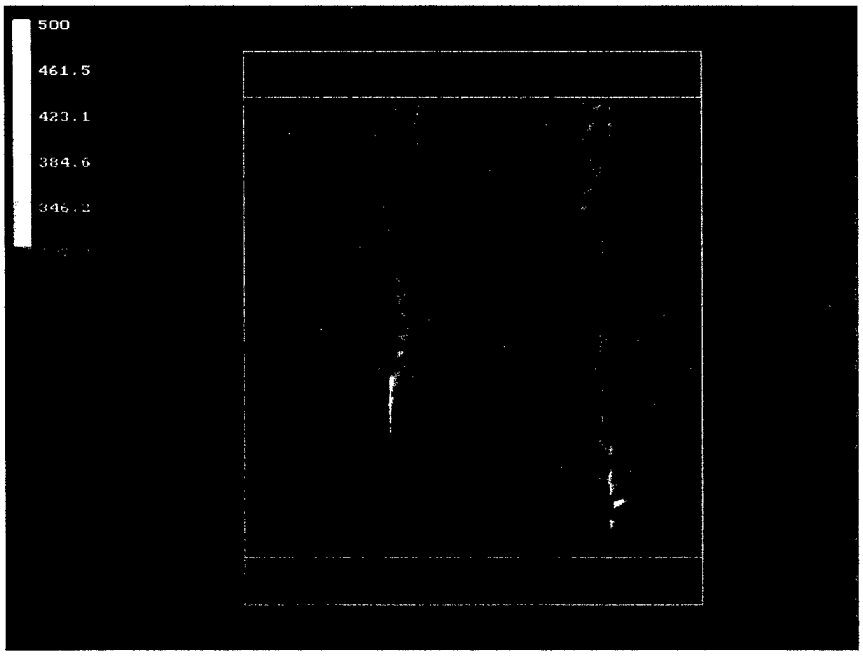


Figure III-48
Linear Regenerative Model at 20μs, $\alpha = 0.100$

Although there seems to be a clear tendency to organize warm areas into traces normal to the electrode surfaces, no dominant path phenomena are apparent. A typical voltage pattern at time 20μs for the linear regenerative model, though relatively uneventful, is shown below in Figure III-49 for comparison to models to follow. Voltages are plotted with 104 colors (8 repetitions of 13) over a range of 0 volts to 6000 volts. Slight unevenness in voltage patterns can be attributed to material randomness.

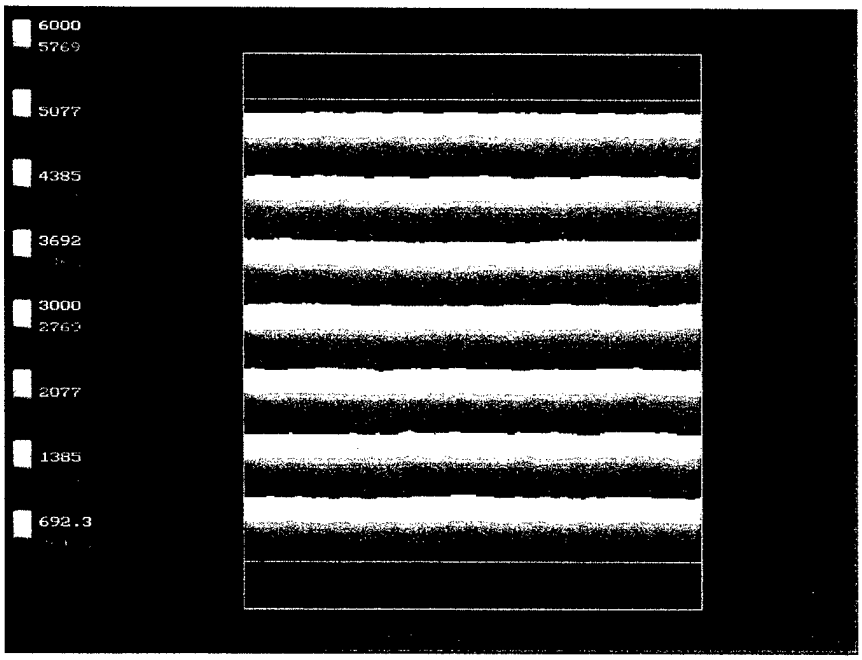


Figure III-49
Linear Regenerative Model at 20μs, $\alpha = 0.010$

Figure III-50 is a plot of transient potential across electrodes for all three linear regenerative models. There are several features of note. First, the voltages in all three cases start out somewhat higher in these models, given the randomness of the materials, than in materials that are mathematically pure. Second, increasing values of α produce increasing downward slopes of potential, as expected, because of the increased conductivity as the material temperature rises. Third, the anomalies in the curve for $\alpha = 0.100$ are believed to be convergence problems in the calculations, to be discussed later in Subsection III-5-7.

Power generation curves for these models are shown in Figure III-51.

The most important result from these three models is the apparent total absence of dominant path phenomena. These three sets of conditions therefore represent cases of non-chaotic behavior.

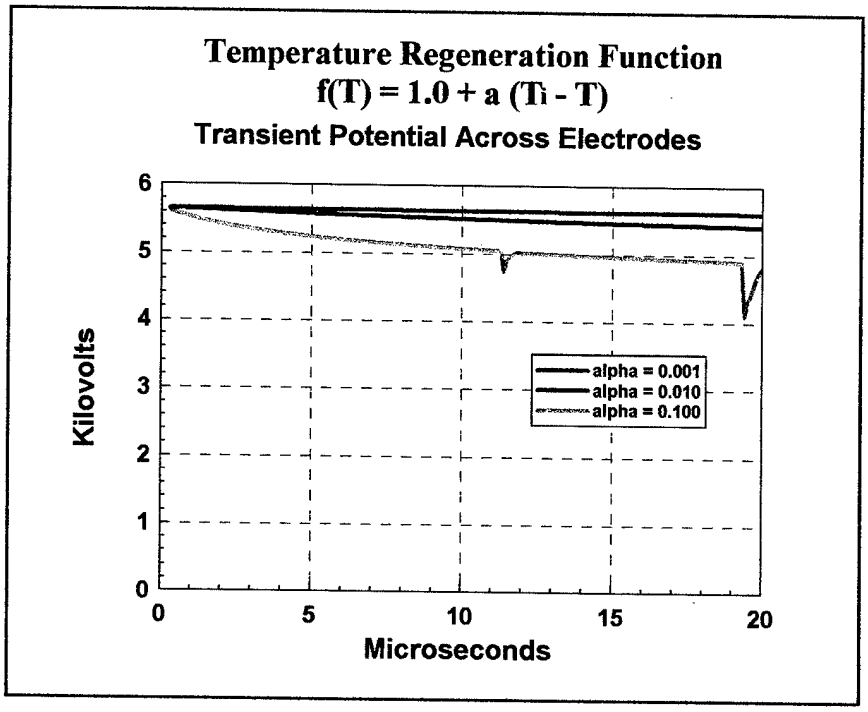


Figure III-50
Linear Regenerative Models: Voltages

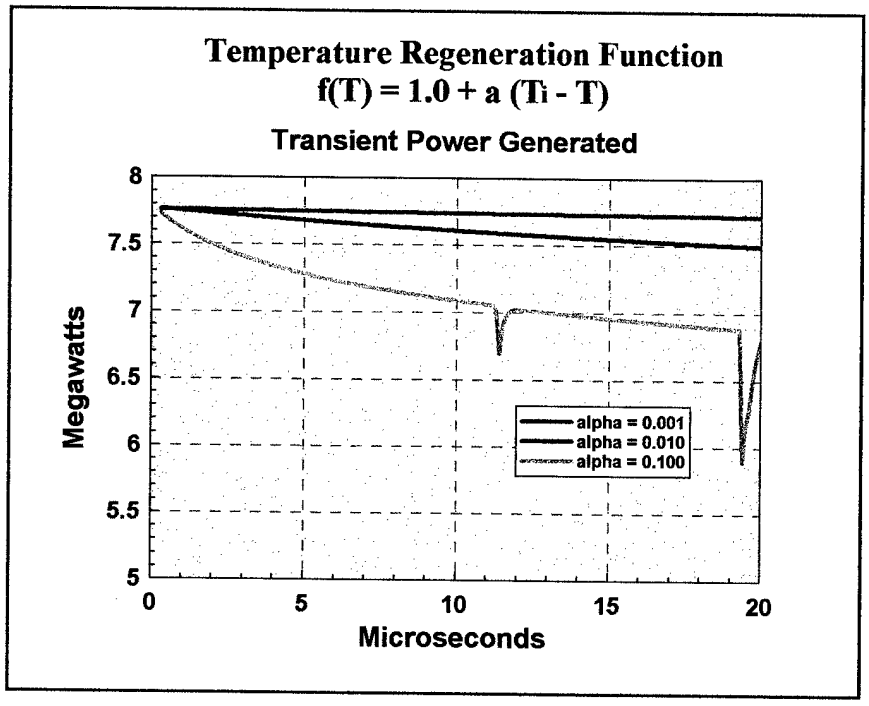


Figure III-51
Linear Regression Models: Power

III-5-3: EXPONENTIAL REGENERATIVE MODELS

In the second set of models, the regenerative function is chosen to be

$$f(T) = 1.0 + e^{\alpha (T - T_i)}$$

where T = local instantaneous temperature
T_i = local starting temperature
α = exponential multiplier of temperature effect

for each element in the finite element domain, such that

$$\sigma_{(e)} = \frac{K_E}{(Ng)^2} E_{(e)} e^{-Ng\beta_E/E} \left(1.0 + e^{\alpha (T_{(e)} - T_i)} \right)$$

where the subscript (e) refers to a constant or average (in the case of temperature) value across a single element. The values for K_E, β_E, N, and g retain the values listed above. T_i equals 0 and the standard deviation for spatial randomness equals 40Å in all models.

Temperature results are shown in Figures III-52 for the case of α = 0.010. The temperature scale used is 13 colors over a range of 0°C to 500°C in order to provide consistency among temperature displays. Once again, evidence of dominant path failure is absent.

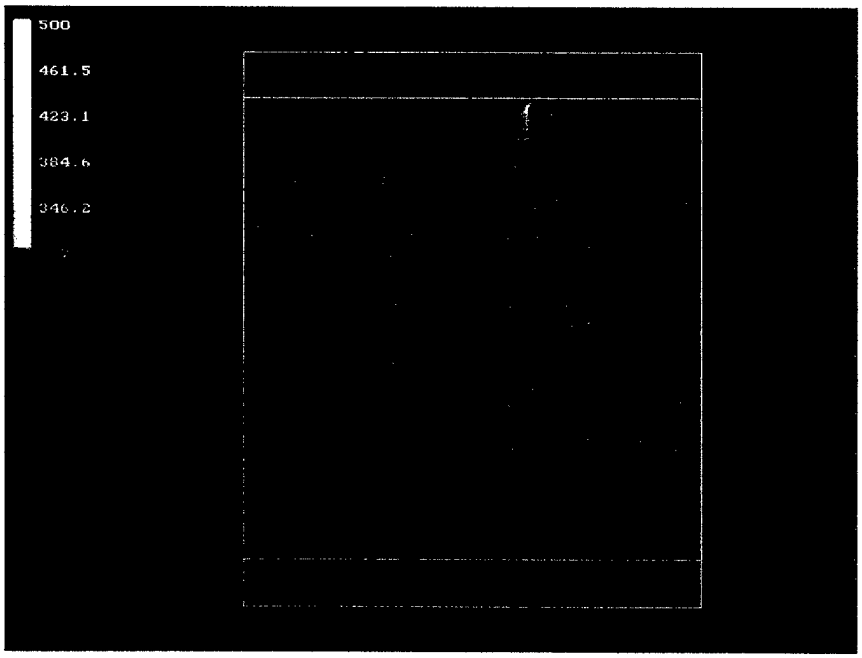


Figure III-52
Exponential Regenerative Model at 20μs, α = 0.010

In exponential regenerative function cases where α = 0.050 and α = 0.100, however, the character of the modelled results depart remarkably from those in the previous models. Dominant path conduction is clearly evident in the temperature plots of Figures III-53 and III-54 using the same temperature scale as before.

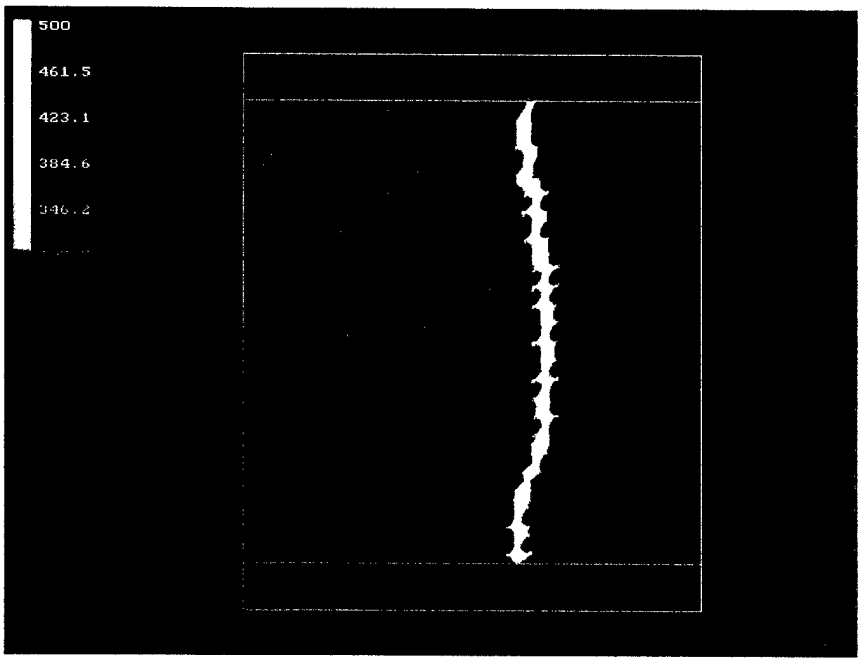


Figure III-53
Exponential Regenerative Model at 20μs, α = 0.050

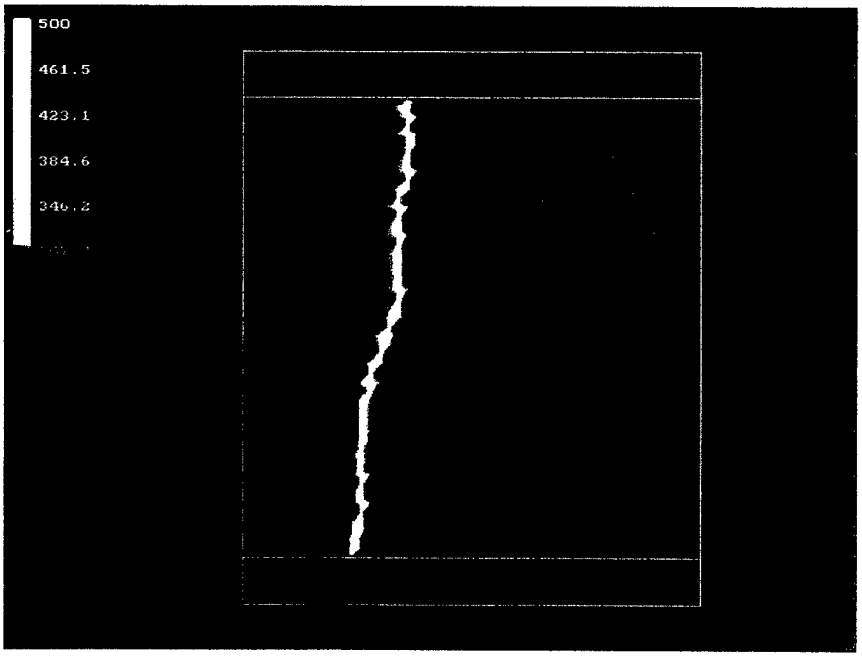


Figure III-54
Exponential Regenerative Model at 20μs, α = 0.100

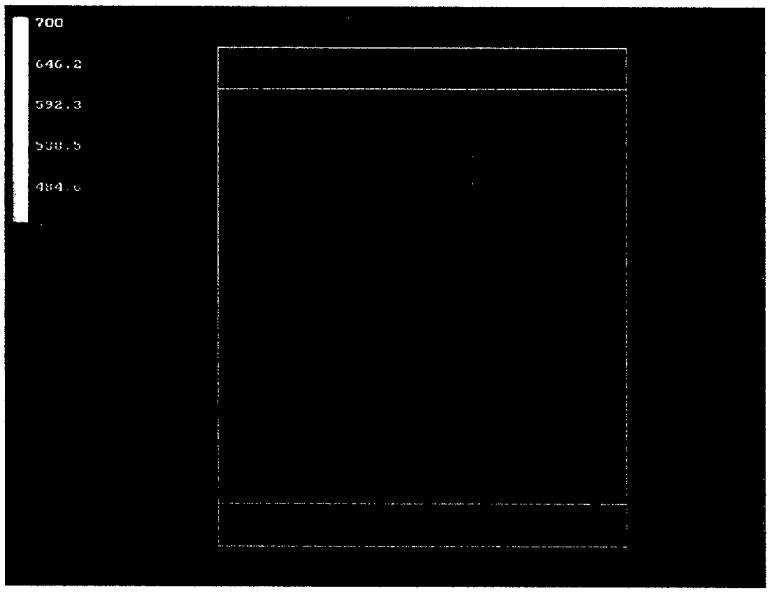


Figure III-55 t = 6μs

An intriguing aspect of dominant path as modelled is the manner in which the phenomenon develops over time. The following Figures III-55 through III-57 comprise a sequence of temperature "snapshots" at 2μs intervals to show the transient formation of chaotic dominant path failure.

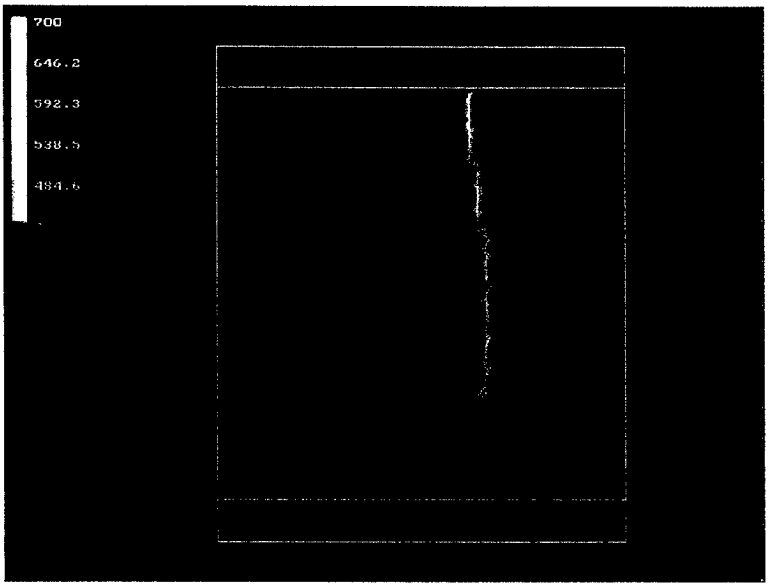


Figure III-56 t = 8μs

This sequence models the case where $\alpha = 0.05$. Note that 13 colors are used to depict temperatures ranging from 0°C to 700°C, a range chosen to show this time sequence to full effect. Maximum temperature predicted by the model at 20μs exceeds 1000°C.

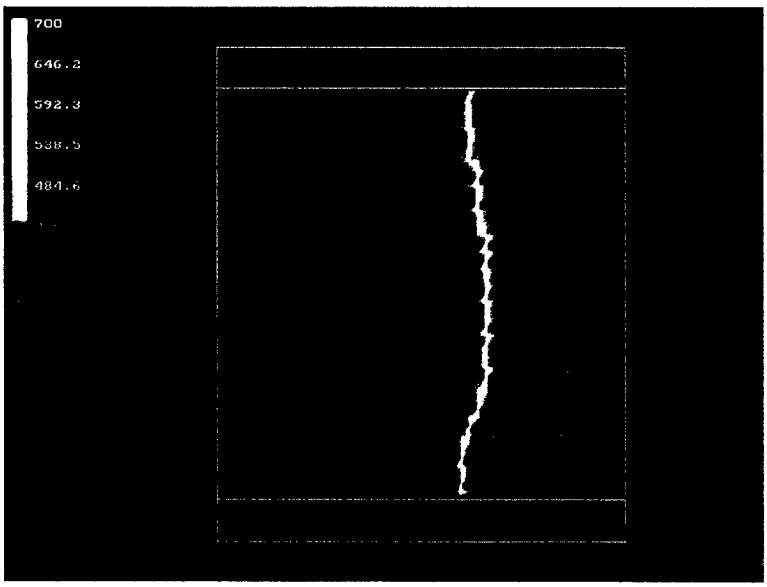


Figure III-57 t = 10μs

The transient voltage patterns for this model change abruptly as dominant path is achieved. Note the difference between the 8μs voltage pattern shown in Figure III-58, at the beginning of dominant path formation, and the 18μs voltage pattern shown in Figure III-59, when dominant path is more fully developed. The latter pattern may be considered true evidence of chaos. Voltages are once again plotted with 104 colors (8 repetitions of 13) over a range of 0 volts to 6000 volts.

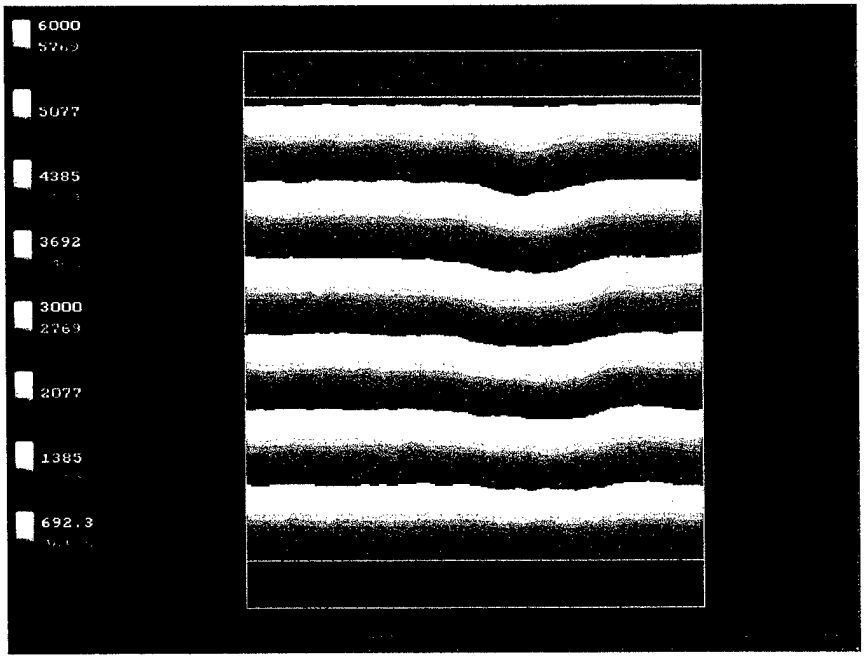


Figure III-58
Exponential Regenerative Model at 8μs, $\alpha = 0.050$

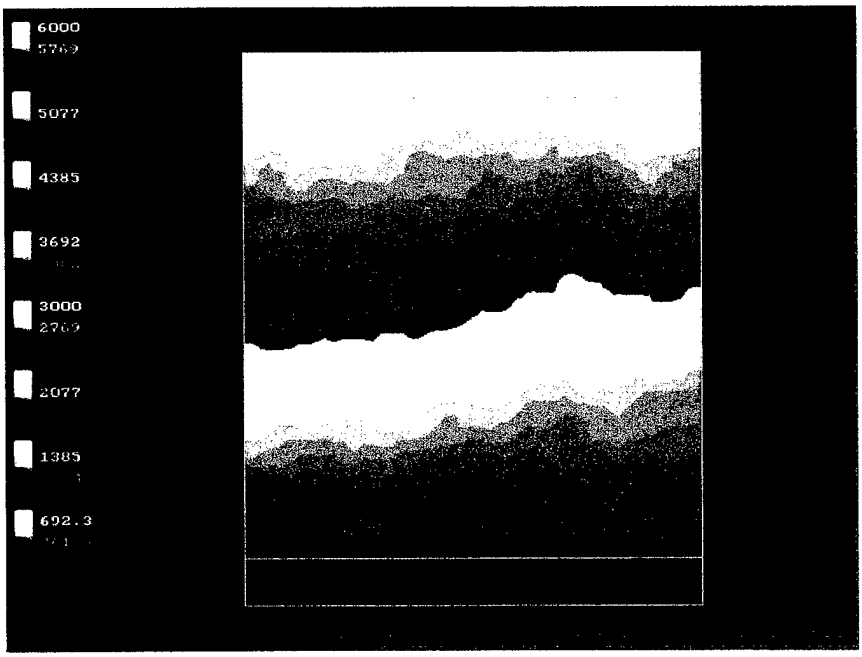


Figure III-59
Exponential Regenerative Model at 18μs, $\alpha = 0.050$

Figure III-60 shows graphs of transient voltage response for the three exponential regenerative models shown above, with α equal to 0.01, 0.05, and 0.10. The gentle downward slope in the first curve is a relatively quiet response to the material Joule heating in combination with the exponential regenerative function linking electrical conductivity to temperature. The much sharper sudden drop in voltage in each of the other two models is coincident with the dominant path phenomena shown in the temperature patterns of the above figures.

Note that the time necessary for dominant path to form correlates negatively with the value of α.

This graph invites the interpretation that for dominant path to occur, it is insufficient for the material to be extremely sensitive to voltage gradient, possess spatial randomness, and exhibit regenerative behavior as an exponential function. That function, if it is in the form shown, must also have an exponential multiplier α greater than 0.010.

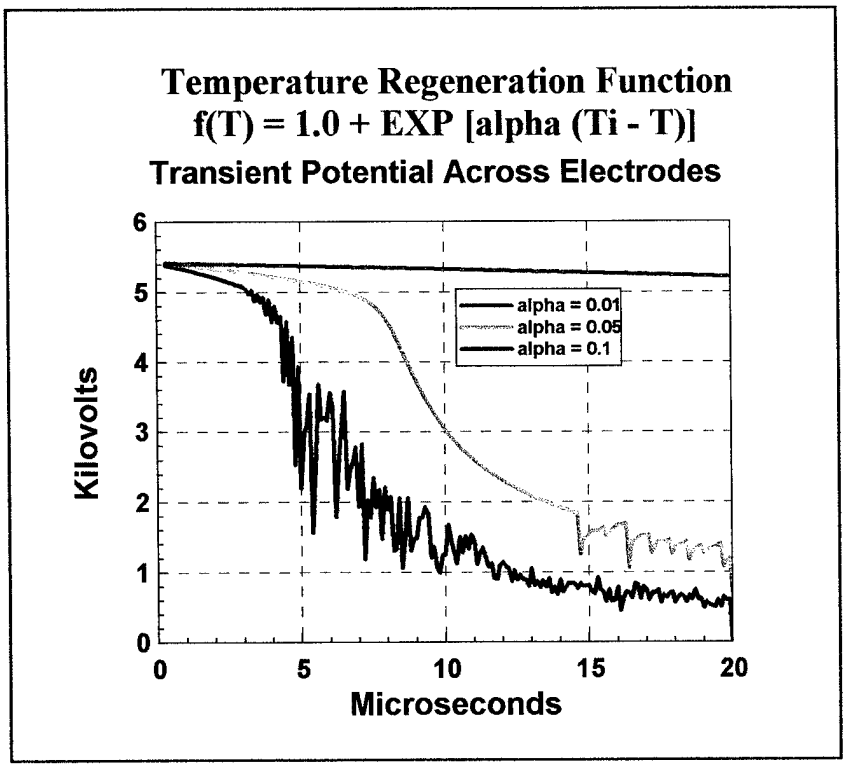


Figure III-60
Exponential Regenerative Models: Voltage

As a check to be sure that material randomness is required for dominant path failure, the case of $\alpha = 0.05$ was modeled without randomness. Temperature results are shown in Figure III-61. A comparison with the model having randomness is shown graphically in Figure 62.

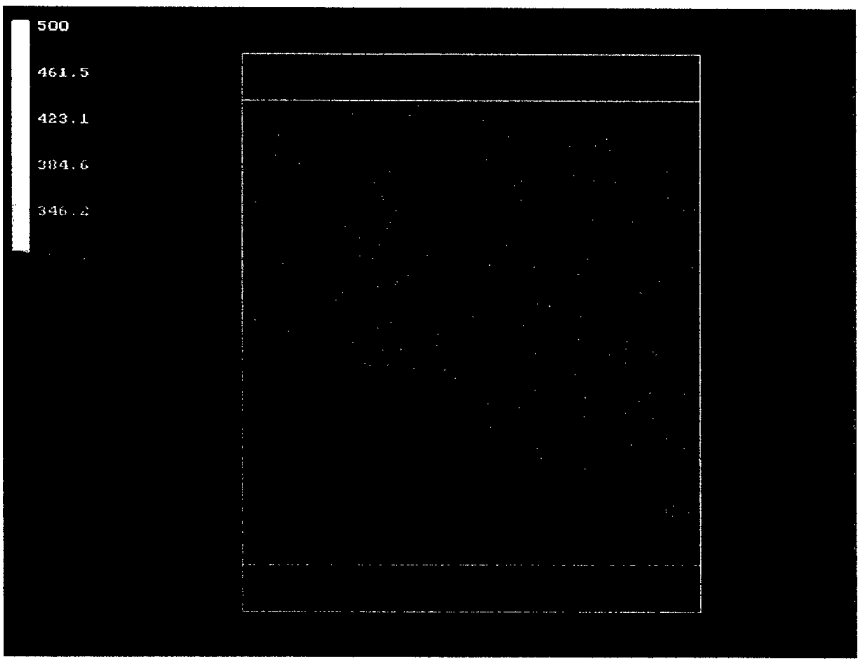


Figure III-61

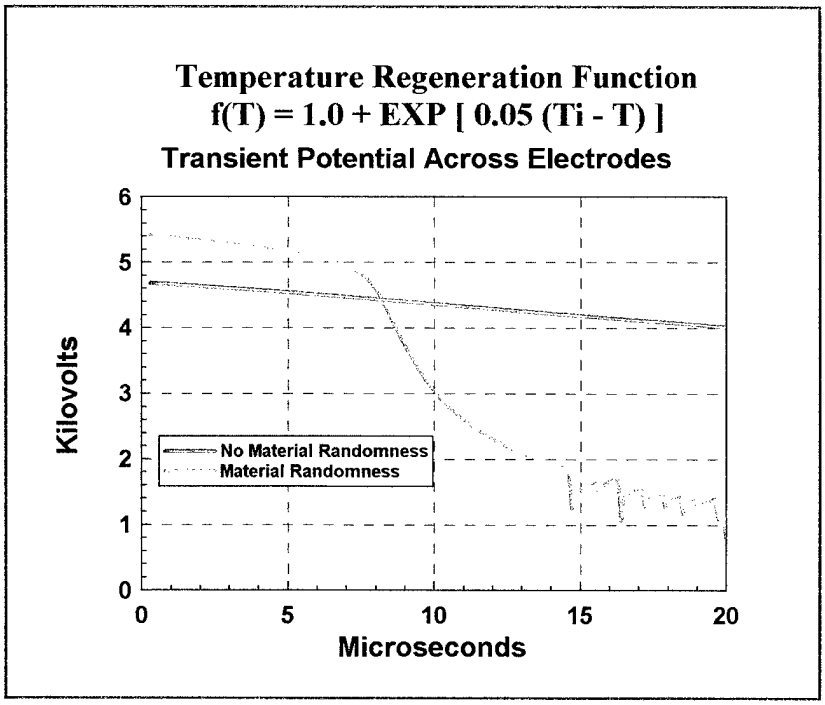


Figure III-62

III-5-4: QUENCHING DOMINANT PATH PHENOMENA

Once it has been shown that modelling of dominant path phenomena is possible, the resulting models beg an immediate question: how might the dominant path failure be prevented? What property can be introduced into the material to provide robustness in the face of voltage gradient sensitivity, spatial randomness, extreme regenerative effects, and overvoltage threats?

Any such "magic property" must allow for voltage gradient sensitivity such that the resulting material is effective in a varistor device. However, while permitting voltage sensitivity, the material must also resist local current concentration (high local J). Given the speed of the events, it is unlikely that a mechanism that responds to an impending high J could react in time to "quench" the phenomenon. This means that the mechanism would have to be an inherent property of the material itself.

One hypothesis is that a simple linear bulk resistivity, if it could be introduced into the material, could serve to limit local current concentrations. There are at least two major tasks required to implement this potential solution. First, it is necessary to discover the proper value of this hypothetical linear resistivity. If the value is too high, the device constructed from it will behave as a simple linear resistor. If the value is too low, dominant path will not be denied. Second, it is necessary to find a mechanism for implementing this linear resistivity. Since binder materials and organic or inorganic resistive coatings are designed for purposes of quantum tunnelling and can be vulnerable to avalanche or other regenerative phenomena, manipulating their properties is unlikely to be effective in attaining a stable linear bulk resistivity. The conductive phase, the conductive particles themselves, must be the focus of development with this regard. Departing from metal spheres and substituting metal-coated ceramics appears to have the most promise for constructs of the conductive phase. Details are addressed elsewhere in this report.

Given that means can be found to impart linear bulk resistivity for thwarting dominant path phenomena, it remains to find a proper resistive value. The following models test the bulk resistivity hypothesis and examine a range of these values.

When adding a bulk resistivity ρ_b to the material construct, the expression for electrical conductivity takes the form

$$\sigma_{(e)} = \frac{1}{\dfrac{1}{\left[\dfrac{K_E}{(Ng)^2} E_{(e)} e^{-Ng\beta_E/E}\left(1.0 + e^{\alpha(T_{(e)} - T_i)}\right)\right]}} + \rho_b$$

Three values of ρ_b are chosen for the models that follow: 0.001Ω-cm, 10.0Ω-cm, and 0.100Ω-cm. Temperature results are shown in Figures III-63 through III-65. It is interesting to note that these three values bracket the desired value for ρ_b.

A value of 0.001Ω-cm for ρ_b does not prevent dominant path. Figure III-63 thoroughly underscores this statement in dramatic fashion.

A value of 10.0Ω-cm does prevent dominant path, as shown clearly in Figure III-64, but the modelled device behaves as a simple linear resistor, resulting in a nearly constant voltage across the electrodes of nearly 12,000 volts! Clearly neither choice is acceptable as a material for a varistor device.

However, the value of 0.1Ω-cm appears to stop dominant path from forming while not interfering with the intended sensitivity to voltage gradient. The Temperature patterns shown in Figure III-65 help to convey an important finding. At this value of linear bulk resistance, temperature anomalies arising from the material spatial randomness are clearly visible. The implication is that the proper value of linear resistivity does not prevent temperature anomalies from forming, but it does prevent these anomalies from coalescing into continuous channels of dominant path.

In these models, two principles have been demonstrated: that dominant path can be avoided with linear bulk resistivity, and that an ideal range exists for linear bulk resistivity that provides optimum performance.

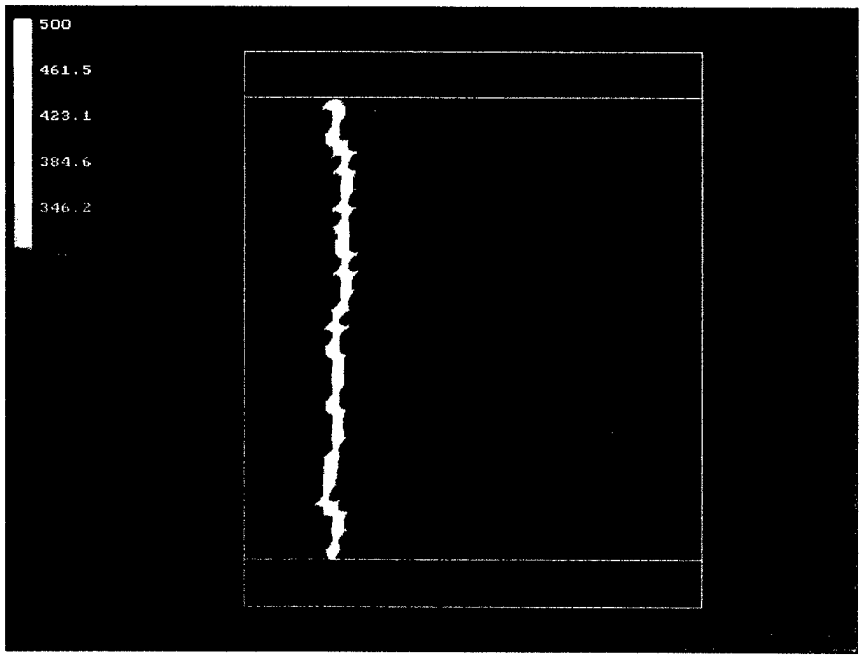


Figure III-63
Exponential Regenerative Model at 20μs, $\alpha = 0.050$, $\rho_b = .001\Omega$-cm

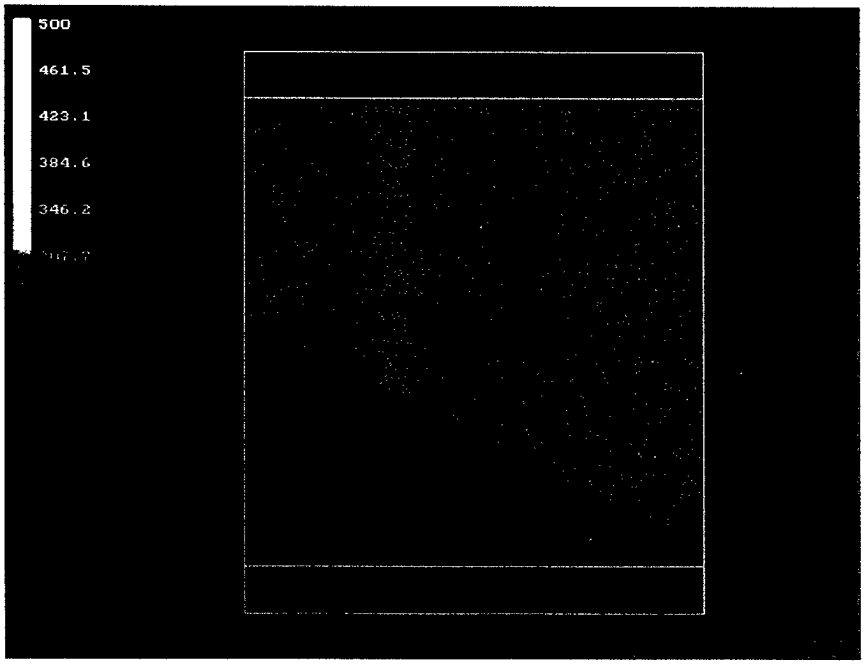


Figure III-64
Exponential Regenerative model at 20μs, $\alpha = 0.050$, $\rho_b = 10.00\ \Omega$-cm

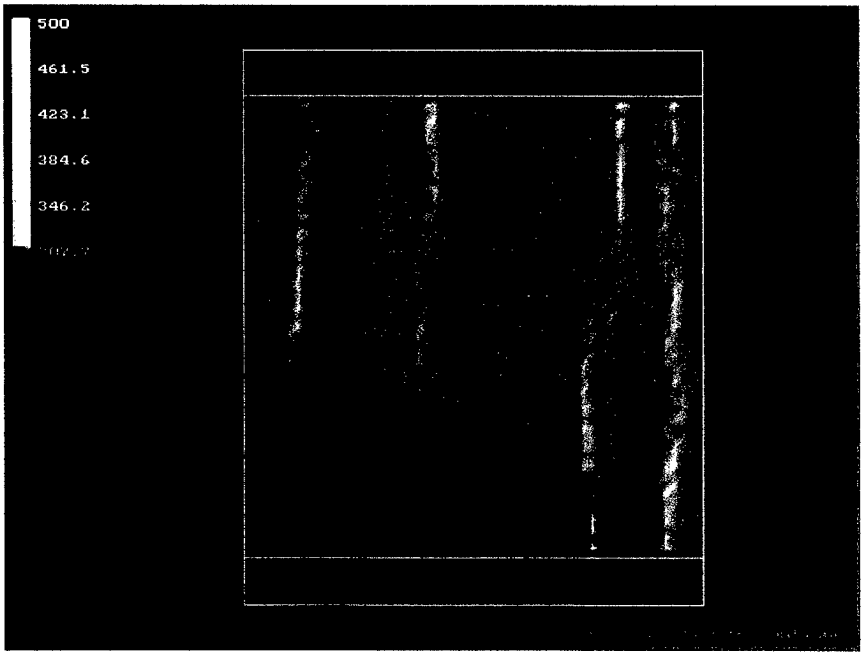


Figure III-65
Exponential Regenerative model at 20μs, $\alpha = 0.050$, $\rho_b = 0.100\ \Omega$-cm

Figure III-66 shows graphs of transient voltages across electrodes for these three models. Figure III-67 shows power generated in the modelled device vs. time. Note that at the highest linear resistivity the power exceeds 13 megawatts.

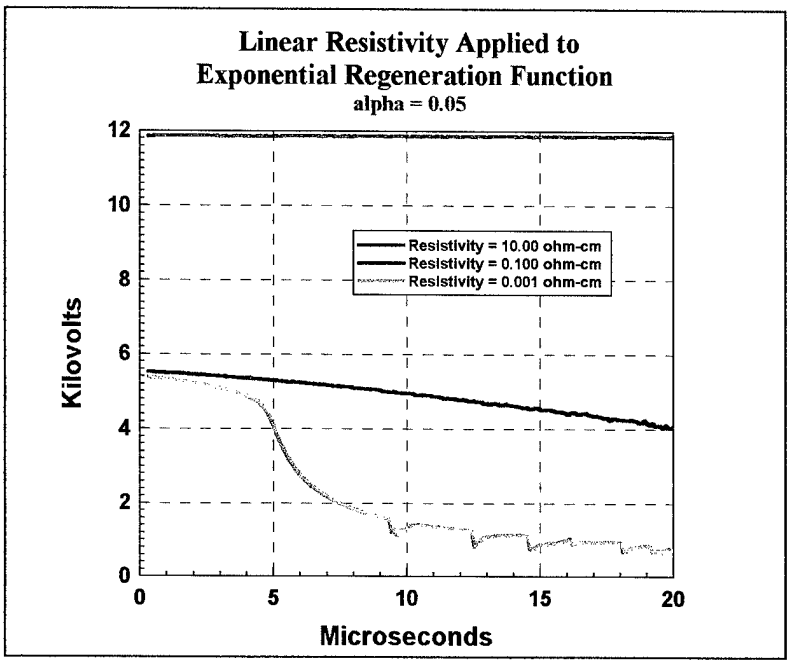


Figure III-66
Exponential regenerative models using three values of linear bulk resistivity to quench dominant path formation, $\alpha = 0.050$: Voltages

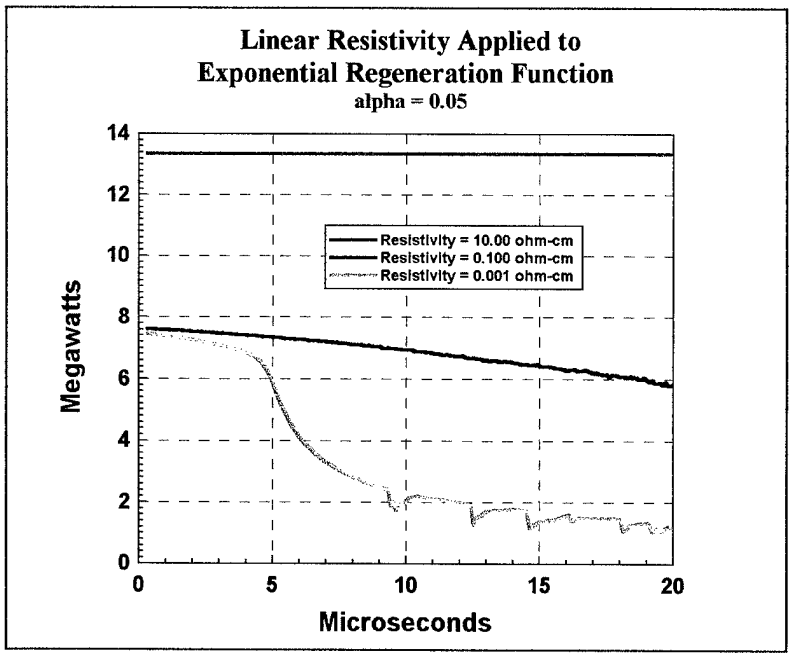


Figure III-67
Exponential regenerative models with linear bulk resistivity : Power

III-5-5: EFFECT OF GEOMETRY ON DOMINANT PATH PHENOMENA

Open questions relate to curvilinear geometries and their effect on dominant path: would a cylindrical construction for a varistor device, with its inherent logarithmic relationship of voltage vs. radius, give superior performance by increasing the voltage gradient threshold for dominant path failure? Conversely, would the higher current density closer to the center of a cylinder exacerbate the problem? Answers to these questions have obvious impact on varistor design.

Several models were run to compare flat and cylindrical constructs in order to provide preliminary answers to these questions. The general form of the cylindrical models is shown in Figure III-68. Two cylindrical models employ the voltage and ballast resistance shown, but with different inner and outer radii of the varistor material. Cylinder No. 1 has varistor material inner and outer radii of 4.0mm and 5.0mm, while cylinder No. 2 has radii of 2.02mm and 4.0mm, respectively. These radii are chosen to simulate the volume of a 1.0mm flat construction, also modelled for comparison. Depth of all models is 1.0cm. In all cases the electrode thicknesses are consistent at 1.0 mm. 7.5KV is applied at time zero to one electrode through a ballast resistor of 400 ohms; the other electrode is grounded.

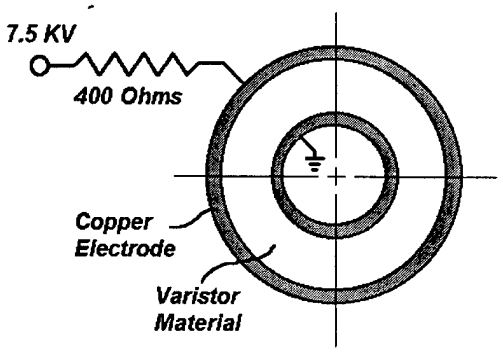


Figure III-68
Cylindrical Models

The material used in these models is identical to that used in Section III-5-3, with statistical material variability, and where alpha in the exponential regenerative function is chosen to have a value of 0.05. This material has already demonstrated dominant path failure in the models depicted in that section. The models here are expected to demonstrate the influence of cylindrical geometry on the onset of dominant path.

Results of these models are shown in Figures III-69 through III-71 as one-quarter constructs, which reduced computation time. Note that all three exhibit dominant path failure in these temperature maps at 10.0μs after voltage application.

Plots of voltage response are shown overlapped in Figure III-72. This figure demonstrates the most important preliminary finding: judging by the time to onset of dominant path , when voltage across electrodes begins to drop, one may conclude that cylindrical designs do not help to alleviate the dominant path problem. Cylindrical constructs may even exacerbate the problem as curvature increases. Note that these findings are extremely preliminary and need to be verified with extensive models and laboratory trials.

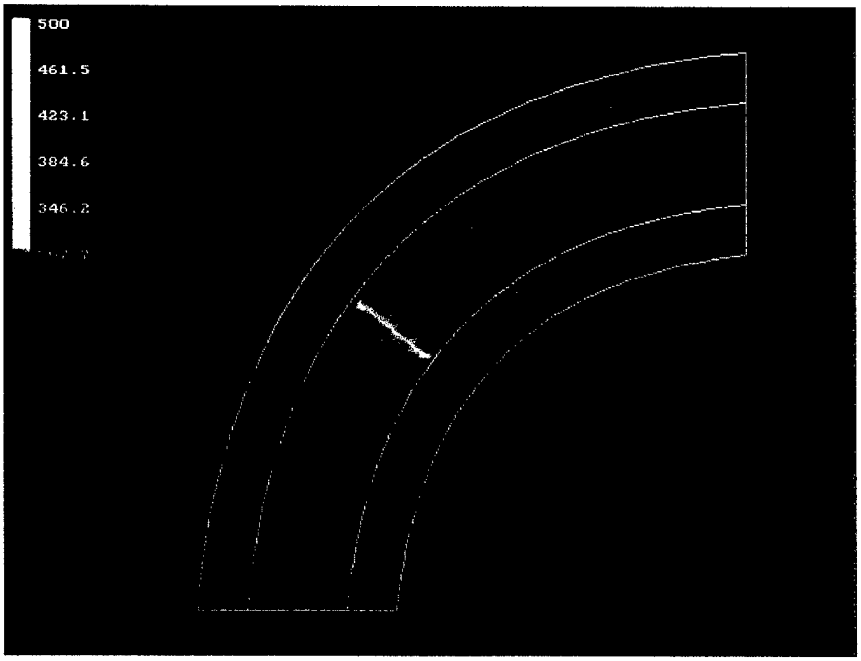


Figure III-69
Cylindrical Model No. 1 at 10μs, $\mathbf{r}_i$ = 4.0 mm, $\mathbf{r}_o$ = 5.0 mm

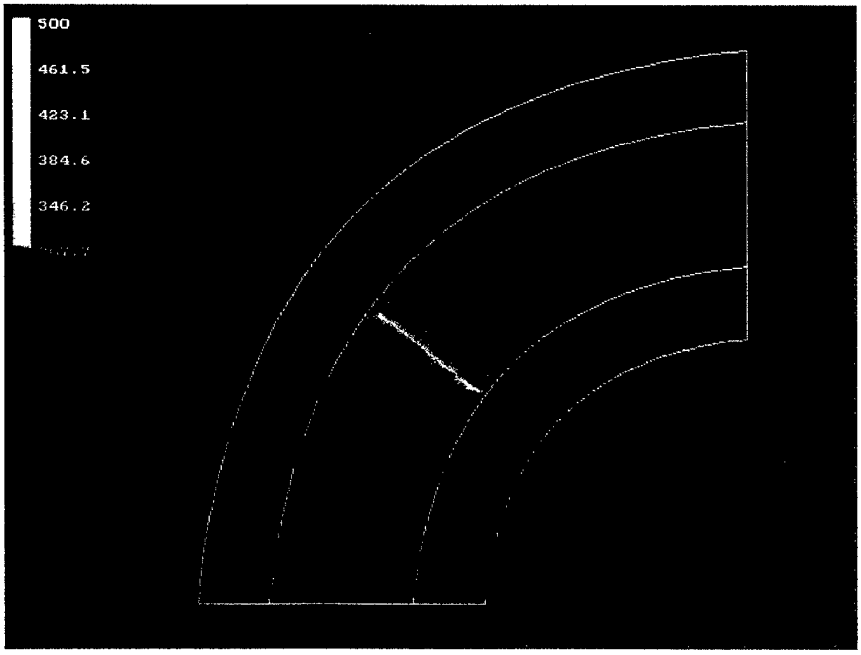


Figure III-70
Cylindrical Model No. 2 at 10μs, $\mathbf{r}_i$ = 2.02 mm, $\mathbf{r}_o$ = 4.0 mm

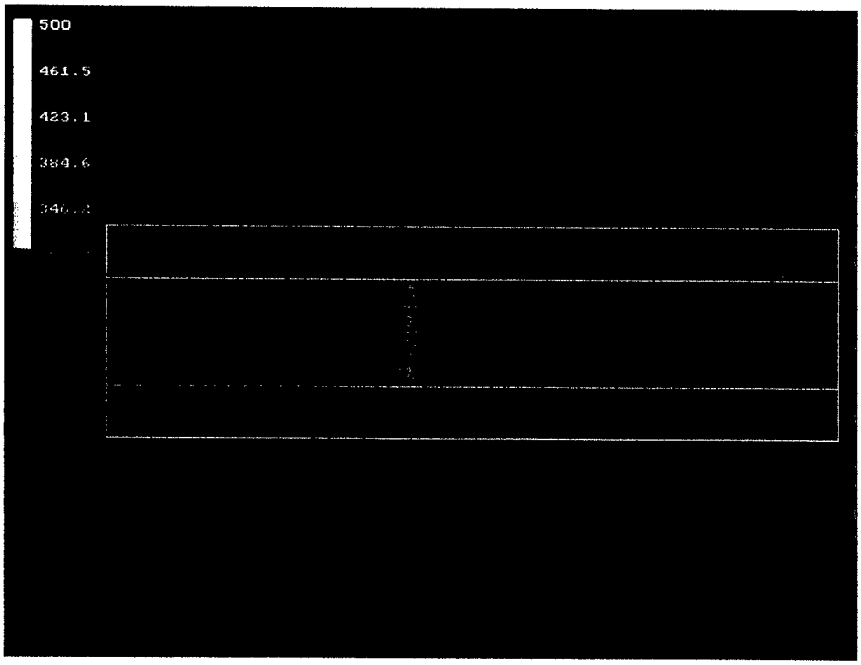


Figure III-71
Flat Model at 10μs, Thickness = 1.0 mm

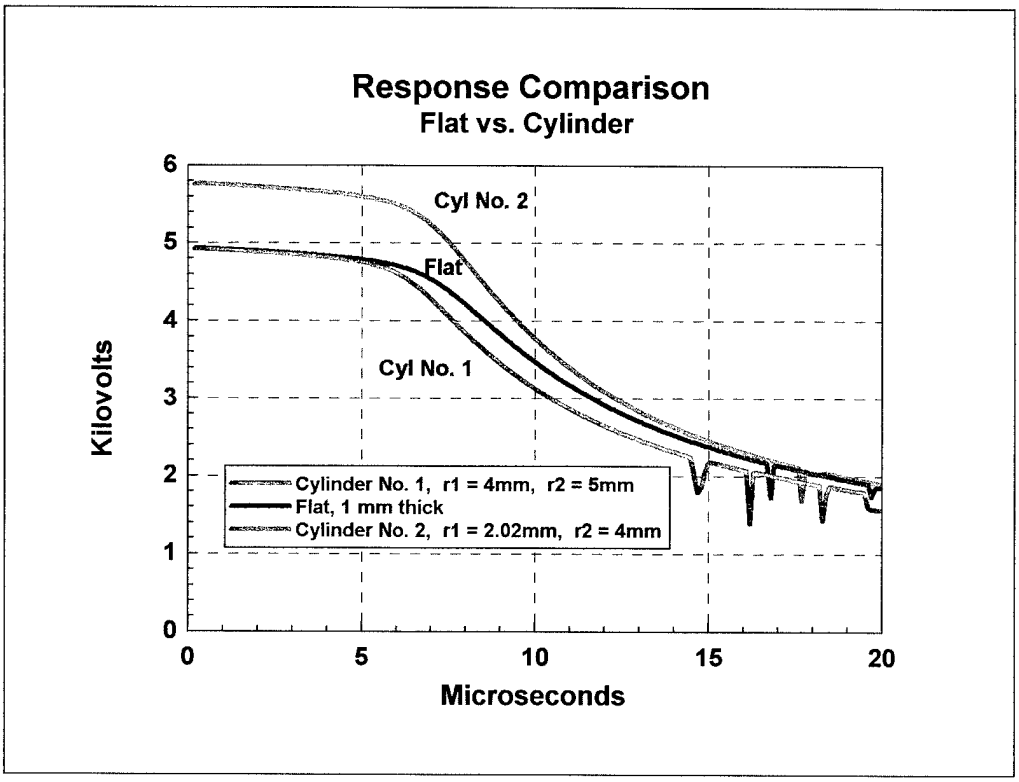


Figure III-72

III-5-6: EFFECT OF DEVICE THICKNESS ON DOMINANT PATH

Another question relates to device thickness in the plane normal to the electrodes: are thinner devices more susceptible to dominant path failure than thicker devices?

Models were run to simulate the effect of thickness changes. The three thicknesses employed are 5.0mm, 2.0mm, and 1.0mm. Applied voltages and ballast resistances are altered to keep the applied field constant. The applied field is also consistent with the 40KV applied through 25 ohms as used in the 1cm cube models reported in previous sections. Again the varistor material is chosen to be identical to the one used in Section III-5-5, a material which has already demonstrated a tendency to exhibit dominant path failure at this applied field strength. The models here are expected to demonstrate the influence of thickness on the onset of dominant path.

Results are shown in Figure III-73 as superimposed plots of voltage across electrodes vs. Time. Differences between results for models having the same input values are a result of the subtle differences imparted from different random number seeds used in generating material randomness.

These models demonstrate another important preliminary result, evident in this figure: judging by the time to onset of dominant path , one may conclude that thickness has little bearing on performance down to a threshold thickness of about one millimeter. Below this threshold, device thickness is an important design consideration. This result is consistent with intuition regarding the lower number of conductive particles between electrodes and the implied statistical influence in thinner devices.

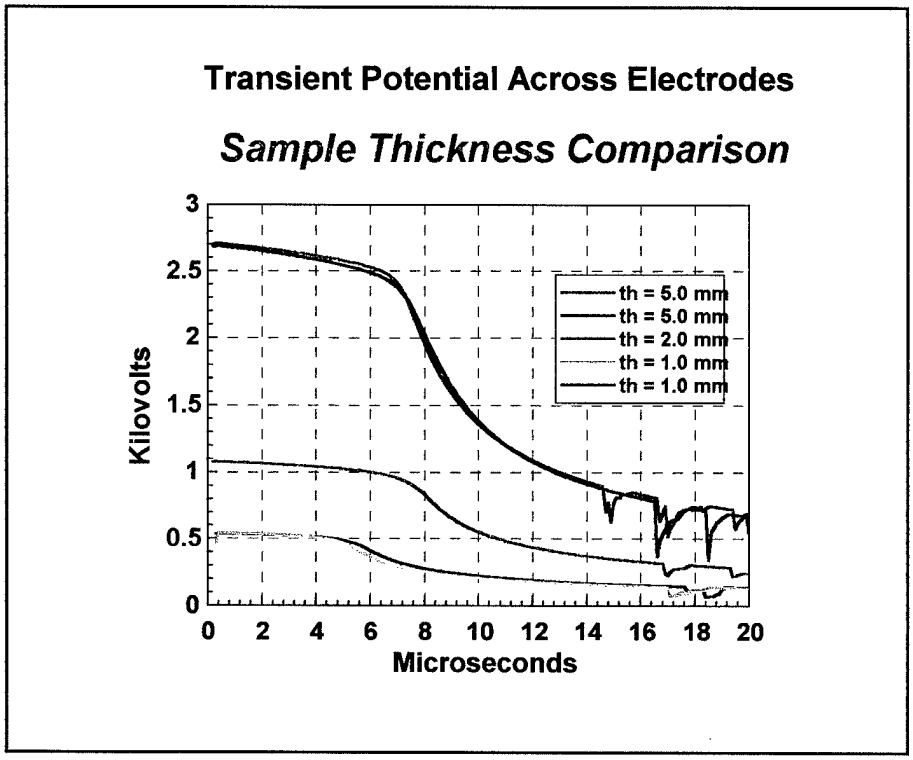


Figure III-73

III-5-7: SOURCES OF ERROR AND PROPOSED SOLUTIONS

The noise in the voltage plots of Figure III-60 and elsewhere are likely to be caused by difficulties of convergence in the program performing the calculations. At least part of the convergence difficulty stems from the size of the time step. Although the time step of 100ns might appear at first glance to be sufficiently small, it may still be too large to handle the rapid transient behavior occurring throughout the solution domain.

Too broad a time step in the face of rapid transient phenomena can have a negative impact on the finite element method itself, especially when the functions chosen to represent the field variables within the element are of low order. In the case of the linear first-order elements used in these models, the mathematical difficulty can be defined graphically in a simplistic way, as shown in the following two figures. Figure III-74 shows a hypothetical finite element first-order triangle on an X - Y plane with temperature T used for the third dimension. All nodes are at 0 °C.

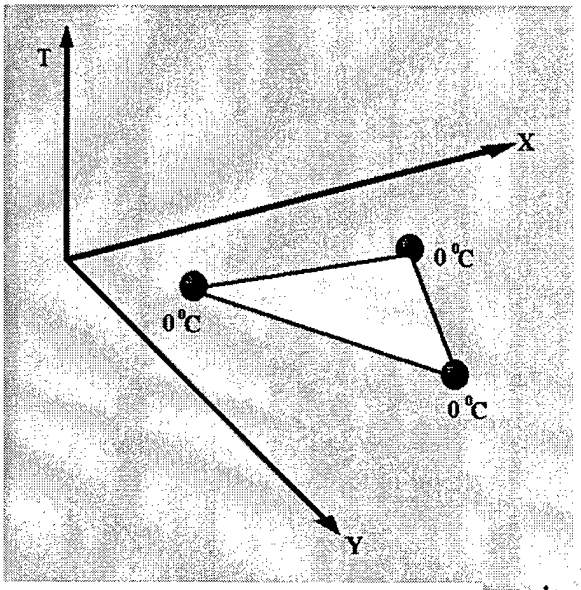


Figure III-74

Figure III-75 shows the triangle after a sudden transient event. In this instance it is possible that the sudden rise in temperatures in two of the nodes, in combination with the inherent thermal capacitance inertia, can drive the third node to a negative value, even though only positive thermal energy has been applied.

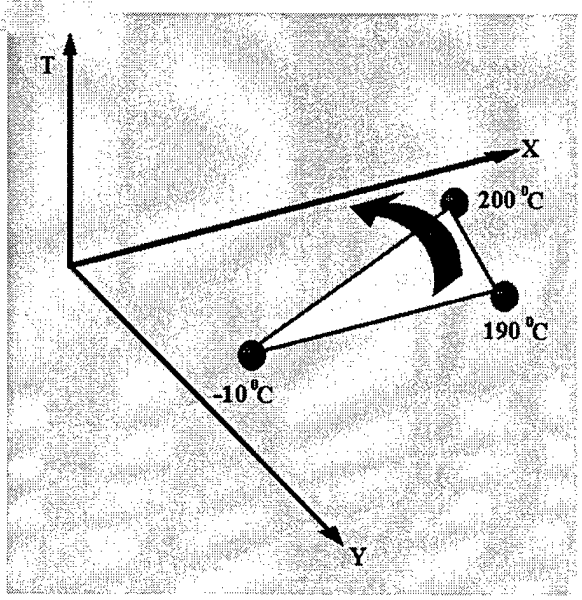


Figure III-75

The solution to this calculation problem is twofold: to decrease the time step and to reformulate the software with higher-order elements, which will effectively "bend" in the situation shown in Figure III-75.

Decreasing the time step by itself will result in prohibitively long run times, which already exceed 12 hours. Changing to higher order elements is appropriate as full 3-D problems are addressed in Phase II of the project.

SECTION III-6

IMPACT OF THE NEW MODELLING TECHNIQUES

III-6-1: DESIGNING VARISTOR MATERIALS THROUGH MODELLING

The models shown above clearly demonstrate the ability to predict chaotic dominant path phenomena and, much more importantly, how to use these modelling techniques to create robust materials that maintain their desired voltage gradient sensitivity while avoiding dominant path formation entirely.

The search for new material constructs using these techniques will depart significantly from tradition in both material science and finite elements. The models contain randomness. Therefore it is not possible to know precisely where and how a material will fail with one model alone. Many models of the very same conditions need to be run, with only different seeds for the pseudo-random number generators.[1] The information to be discovered involves the *nature* of material performance and failure modes. Such an endeavor will keep computers busy. By contrast, creating material with no randomness whatsoever, material having perfect mathematical purity as defined in traditional models, is clearly impossible and foolish to attempt.

III-6-2: BEYOND VARISTOR DEVICES

The impact of this modelling technique extends far beyond varistor devices. There are quite possibly immediate applications in these fields:

- Superconductivity
- Nuclear hardening
- Crack propagation and crack initiation
- Transdermal iontophoresis drug delivery systems
- Semiconductors and integrated circuits

The above list is not intended to be complete. The techniques apply to any system containing materials which are not perfect. Since that set of materials includes virtually all those known, it is better to describe the systems for which these modelling techniques are essential: those in which material imperfection results in significant and otherwise unpredictable modes of behavior. The list of such systems is still very large indeed.

---

[1] In the above models, random number seeds came from the computer time of day clock in milliseconds.

Section IV-1

MATERIAL FABRICATION

IV-1-1 General Considerations

The object of this work was to prepare conductive particles in a resin matrix between electrodes supplied by Integrated Sciences. The goal was to keep the conductive particles separated by coating the particles with a nonconductor. The target separation between the spheres was originally expected to be on the order of 100 angstra. At this separation it was anticipated that the dominant electrical conductance would be by tunneling through the insulative layers between the conductive particles.

At the start of this work, exploratory experiments were carried out with 50 micron hollow ceramic spheres (cenospheres) which were coated with 600 angstra of pure silver metal and then further coated with approximately 2000 angstroms of highly cross-lined acrylic polymer. Mixtures of these insulated conductive spheres were mixed with various binders at volume loadings between 35 and 50 percent.

The electrodes supplied by Integrated Sciences have a conductive surface on each face with a diameter of 15 mm. Therefore each electrode face has a surface area of 176.7 mm^2. If the electrodes are spaced at a 1 mm distance, the volume between the electrodes is 176.7 mm^3. For monodisperse spheres having diameters of 50 microns packed in a cubic lattice, one cubic millimeter will contain 8,000 spheres. Therefore, each electrode at a spacing of 1 mm will contain approximately 1.4 million spheres between the faces.

Since the spheres we are utilizing are not monodisperse, but have a Gaussian size distribution peaking near 50 microns, and the packing is not cubic, but is somewhat random, these calculations only give an approximation of the number of spheres between the faces. The theoretical packing for monodisperse spheres is near 73 volume percent. This percentage cannot be reached practically. The theoretical maximum volume loading for non-monodisperse spheres is near 55 volume percent.

The actual volume loading that can be obtained in a trowelable paste is dependent on the particle size. The binder must completely wet the surface of each particle in a trowelable paste. As the particle size decreases, the surface area of a given weight of powder increases tremendously. The volume loading limit of material that can be put in a trowelable past is limited by the surface area of the particles. If insufficient binder is available to completely wet the particle surfaces, the material becomes crumbly and cannot flow between the electrode faces.

IV-1-2 Binder Trials With Organic Coatings

The first binder tested was 100% solid epoxy resin. The resultant material consistently gave a DC resistivity between the electrodes of less than 0.5 ohms. This measurement indicated that the acrylic film was being distorted sufficiently to allow the metallized spheres to make ohmic contact. It was known that these acrylic coated spheres in a solvent and binder could be sprayed like a conventional paint to yield nonconductive films which were 10 to 30 mils thick when dry.

To determine if the epoxy were somehow responsible for the degradation of the acrylic film, other binders were tested including water soluble silicates, thermoplastic rubber, molten wax, and other acrylic polymers. All of these binder/metallized sphere matrices consistently give high DC conductivity with resistance measurements between the electrodes of below 0.5 ohms.

By connecting the digital multimeter to the electrodes before the binder/sphere mixture was pressed between the electrodes, it was possible to observe that the original mixture had no apparent DC conductivity until it was pressed between the faces of the electrodes. With many of the mixtures, it was possible to pull the electrodes apart slightly and achieve a very high resistance, and then with very light pressure to watch the conductivity increase until the resistance was only a few tenths of an ohm.

These experiments indicated that at the loadings utilized in the binder. the spheres would form a bridge structure between the electrodes as they were pressed together which would"punch through" the thin acrylic layer on the surface of the spheres. Various binders tested demonstrated that those which tended to soften the acrylic layer were much more prone to form a conductive bridge between the electrodes. As an example, an acrylic resin in a solution of MEK/toluene would form a conductive path with much less pressure than was observed in the case of the water based sodium silicate solutions. In both cases, it was practically impossible to squeeze the mixture between the electrodes without creating a conductive path.

In an attempt to avoid the softening of the acrylic, which is generally caused by low molecular weight polymer or solvents slowly diffusing into the acrylic film, a very fast curing cyanoacrylate was used as a binder. It was hoped that this fast curing system would not have sufficient time to attack the acrylic before curing to a solid. Gel times could be varied from a few seconds to several minutes. It was found that this faster curing material did yield nonconductive cured matrices between the electrodes. Initial tests of these electrode/matrices at Integrated Sciences indicated that these mixtures showed some promise, but it was felt that the 2000 angstrom acrylic coating on the spheres was too thick to allow tunneling between the spheres.

In an attempt to generate an insulative layer on the order of 100 angstroms, a thin film of monomer was deposited on the beads as a near monomolecular layer by Langmuir techniques. This material was tested in

the cyanoacrylate binders with variable results. At thicknesses on the order of 1 molecule, the physical resistance of such thin films to "punch through" is very low. Many attempts were made to adjust the technique of placing the mixtures between the electrodes which would yield no DC conductive paths. The results of these attempts were poor in general, and this direction was abandoned.

Experiments were also performed using 50 micron cenospheres coated with nickel and Permalloy 60/40. These materials consistently showed high conductivities between the electrode faces and were, in general, more sensitive to slight pressure between the electrodes. The surface of these materials was known to be rougher than spheres coated with silver. It appears that this surface roughness enhances the ability to form ohmic contacts between the spherical particles as the requirement to squeeze binder from the interface is decreased by the projecting metal from the rough surface. These spheres appear to the eye to be quite smooth, and even under a microscope at 40X appear fairly smooth, but SEM photographs demonstrate that the surface is quite rough at the 100 angstrom level.

The previous experiments were carried out with conductive particles having diameters near 50 microns. If the electrodes were at a 1 mm spacing, only about 20 spheres were needed to bridge the electrode faces and create a conductive DC path. Smaller diameter bead would be expected to have a lower probability of creating a conductive path as the number that must be in a continuous chain is greater. Silver spheres with diameters of 12 and 15 microns were tried with epoxy, acrylic, and cyanoacrylates. These tests resulted in performance as poor as was obtained with the larger spheres. It appears that the advantage of smaller spheres is offset by having more potential paths between the electrode faces.

IV-1-3 Inorganic Coating/Silver Sulfide

The results of the above experiments indicated that the insulative layer needs to have a strong physical integrity to avoid "punch through" when the conductive sphere/resin mix is placed between the electrodes and pressed to the desired thickness. It was felt that there were no organic polymers which would have this degree of physical strength. Techniques were available to simply react a silver coated sphere with other materials to generate a very thin silver compound on the surface of the spheres which would perform as an insulator. These inorganic films can be controlled with acceptable accuracy and also have an extremely good bond to the surface. It was felt that as two spheres were squeezed together that these inorganic films would not be squeezed out of the interface on the sphere surfaces as readily as an organic polymer.

Spherical particles were coated with silver metal at a nominal thickness of 600 angstra. This material was subsequently reacted with sulfur to produce different thickness layers of silver sulfide as insulative layers over the metallic silver. All of the samples prepared by this technique gave

unexpectedly high DC conductivity when placed between the electrodes. This high conductivity was ultimately traced to the fact that silver sulfide is a poor insulator and allows appreciable DC conductivity through the silver sulfide layer.

IV-1-4 Inorganic Coating: Aluminum/Aluminum Oxide

Aluminum oxide was known to be a very good insulator, and further work was directed toward particles of aluminum having an aluminum oxide coating of various thicknesses on their surfaces. "Spherical" aluminum particles were obtained and techniques were devised to oxidize these in a controllable manner. Samples prepared by these methods demonstrated that aluminum oxide is indeed a very good insulator and ohmic conductivity for these materials between the electrodes was in the giga-ohm range.

These materials gave variable results; however, clamping was observed in samples with the thicker oxide coatings. IS determined that the oxide layer was more than 3000 angstroms thick. This presented a puzzle, as this is much thicker than predicted for electron tunneling. Examination of the "spherical" aluminum particles indicated that a large proportion were not spherical and can best be described as "acicular bloblets". It was reasoned that the more pointed areas of these bloblets created a concentration of electrical field which could cause electron tunneling through the thicker oxide layers.

The acicular nature of these powders tended to give nonreproducible results when a sample of the same material was prepared and retested. This was thought to be caused by the random nature of the distribution of the acicular particles. However, three samples were prepared from this powder with similar, if not exact, characteristics. *The samples were subsequently tested according to ESD specifications and all three performed well.*[1]

IV-1-5 Glass/Nickel/Nickel Oxide

In order to obtain more spherical particles, samples of 55 micron solid glass spheres were obtained. These were coated with approximately 3337 angstra of nickel metal. A portion of the nickel on the surface was oxidized to yield approximately 100 angstra of nickel oxide. This material consistently failed by a dominant path mechanism. Examination of the glass demonstrated that a fair amount of acicular glass particles were present. These acicular particles when coated with nickel and subsequently oxidized, demonstrated dominant path failure consistently. A method was developed to remove some of the acicular particles from the material with 100 angstroms of nickel oxide on the surface. These samples also failed by a dominant path mode.

---

[1]C.f., Subsection V-2-6.

A second set of nickel coated glass spheres also having approximately 3337 angstroms nickel were prepared and oxidized more vigorously to yield a thicker oxide coating. This sample was also divided and a portion of the acicular particles were removed from half of the sample. IS reported that the material with the acicular particles partially removed was showing interesting clamping behavior. As noted in the earlier samples, the non-spherical nature of some of the glass starting material would cause field hot spots requiring a thicker nickel oxide layer to avoid immediate dominant path failure. SDS is currently attempting to obtain samples of glass spheres with fewer nonspherical particles and perhaps all near-spherical particles.

There are two positive aspects of the tests made on the last powder. *Definite clamping action was observed, but more importantly, examination of the electrodes showed that the failure paths were spread over a much larger area of the electrode surfaces, probably attributable to the higher uniformity of this powder compared to others.*[2]

IV-1-6 Glass/Nickel/Glass

We next coated glass spheres with nickel and then overcoated with glass. The results are given in Section V and are quite encouraging. Consistent clamping action is obtained with no failures. It is anticipated that glass/metal/glass will be one of the mainstays of our materials.

IV-1-7 Cenospheres/Silver Islands

The difficulty in obtaining near spherical substrates prompted a different approach. SDS prepared hollow ceramic microballoons (cenospheres) with silver islands of various sizes and distributions on their surface. This tends to avoid the high field concentrations that can occur on acicular particles. Tunnelling in these particles would be between the islands on the surface of the sphere.

A range of various silver amounts on the surface of the particles was supplied to IS along with some uncoated cenospheres. The uncoated cenospheres may be mixed in a small amount with the silver coated cenospheres to disrupt any electrically conductive chains between the electrode faces. SDS experience with these chain disrupters in other areas demonstrated that this is an effective method of destroying conductivity in these mixtures.

This particular approach is still in the speculative stage, but should be given further consideration in Phase II.

---

[2]C.f., Subsection V-2-7.

IV-1-8 Closing Comments

Particularly in the face of the apparent unavailability of truly spherical and uniform metal powders, *it appears the path for future success will be in the direction of metal coated insulators overcoated with an insulative layer.*

SECTION IV-2

TEST CELL DESIGN

IV-2-1 Design Criteria

The main criterion for our test cells was simplicity of geometry so that no extraneous factors would be introduced into the physics of our observations. Beyond that, ease of manufacture and handling were considerations. This led to the simple design depicted in Figure IV-2-1.

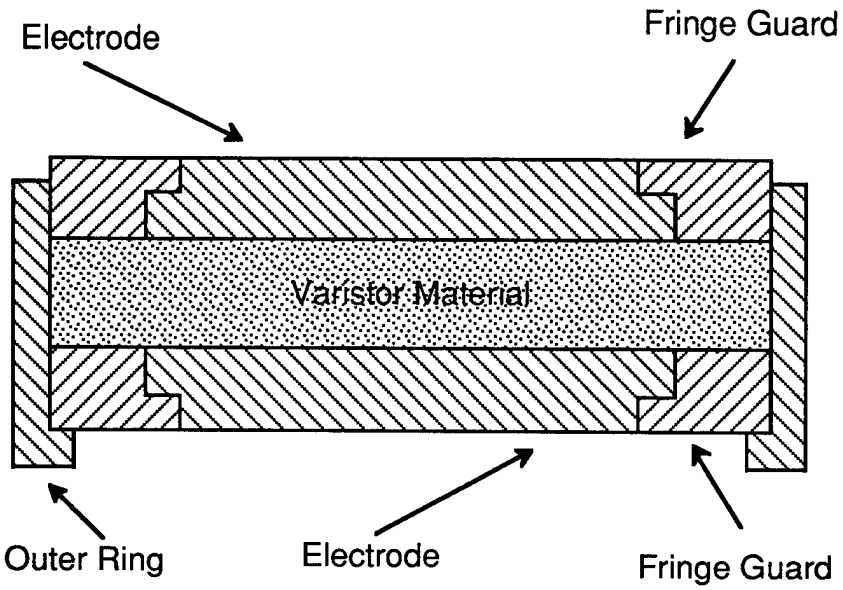


Fig. IV-2-1 Test Cell crossectional view. The Test Cell is cylindrical in shape. The outer ring and the fringe guard are made of the same insulative material. The electrodes are copper. All contacting components are press fits with respect to one another. The function of the fringe guard is the prevention of arcing at the outer edges of the electrodes.

The samples are labelled and the thickness from lower to upper electrode calipered. The electrode thicknesses are measured previously, so the latter measurement gives the interelectrode spacing.

SECTION V-1

TEST METHODOLOGY

V-1-1 Critique of Previous Measurement Methods

A number of companies have made attempts at producing quantum tunnelling varistor devices. For various reasons, some quite mysterious, corporate managers have elected not to take a sound scientific approach. This is particularly evident with regard to the testing methods used. The devices are intended to protect against high voltage transient threats and it immediately occurs to one that high voltage pulsing would be the appropriate method of testing. Several companies have bought expensive high voltage pulsers meeting a number of current waveform specifications.

What they learned from this was that they could destroy their samples and not meet the specifications. This type of testing tells you nothing concerning device physics, does not lead to an understanding of the mechanisms involved, nor indicate a path for improvement. Finding the path to improvement is required for success, and we believe careful analysis of the physical mechanisms involved, coupled with appropriate measurements, is the key to finding the path.

V-1-2 Previous Efforts by Integrated Sciences™

In the work performed by Integrated Sciences™ for the U. S. Army cited in the foregoing sections, a start was made toward the careful analysis. Although all contract obligations were met, the funding level did not allow for the completion of the analyses nor a definitive testing program.

Major Army concerns were varistor leading edge responses and the ability to absorb repeated EMP pulses. These were successfully demonstrated by the Army at their Harry Diamond Laboratory. Thus the leading edge response is not a major concern of the present effort. Our focus will be on testing that reveals the device physics, particularly as it relates to device reliability and energy handling capabilities. In Phase II, it is planned to acquire sufficient instrumentation to both observe the leading edge responses and device microwave properties.

The planned parameters to be measured in Phase I were low current I-V characteristic, reverse standoff voltage, clamp voltage, and device destruct energy. The planned measurement methods will be described in the following subsections. *We say "planned" because experience dictated a different path as will be described in the section on test results.*

V-1-3 Low Current I-V Characteristic and Reverse Standoff Voltage

In a varistor device such as a Zener, the transition from a low conducting to a high conducting state is accompanied by a change in the conduction mechanism. This gives rise to an apparent discontinuity in the Zener I-V characteristic. In pure quantum tunnelling varistors, the conduction mechanism is always the same. The low conducting to high conducting transition is strictly continuous; although, in a circuit application, the transition may appear discontinuous because of the high degree of quantum tunnelling nonlinearity.

The quantum tunnelling nonlinearity is most pronounced in the region of the low conducting to high conducting transition. It is this low current region that most clearly delineates the conduction mechanisms involved. In some past devices that were presumed to be quantum tunnelling devices, it is not at all clear that the dominant conduction mechanism was pure quantum tunnelling. The low current I-V measurements serve to resolve this. Measurements will be made with an electrometer (Keithley 6517). The reverse standoff voltage is read directly from the low voltage I-V characteristic.

V-1-4 Clamp Voltage and Energy to Destruct

The clamp voltage is the voltage across a varistor at maximum peak current. A direct current of this magnitude would certainly destroy the varistor. Thus it is necessary to determine clamp voltage with a pulse method. We use a simple capacitive discharge to accomplish this. A schematic of the circuit is shown in Figure V-1-1.

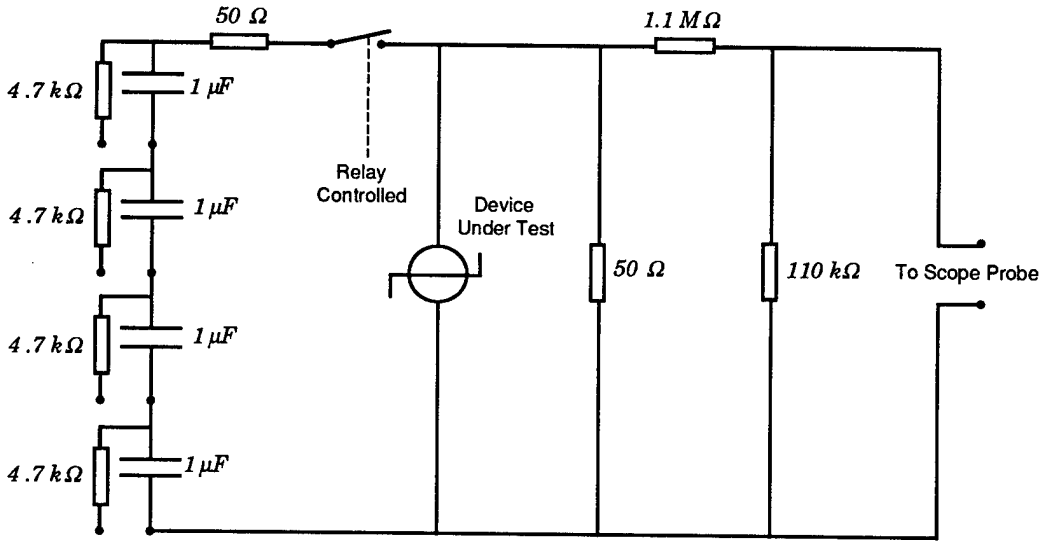


Fig. V-1-1 Schematic of circuit used to determine clamp voltage. The resistors on the far left are used for current limiting during the capacitor charging process. The two resistors on the far right constitute a voltage divider for protection of the scope input. The vertical *50 Ω* resistor is used to determine the fixture characteristics and is removed for clamp voltage determination. The maximum peak output voltage is *1.6 kV*.

An output voltage of *1.6 kV* is sufficient to determine the clamp voltages, but we do not have the capability of supplying anywhere near the theoretical destruct energies. To this end, we use the pulse capabilities of Hyger Physics, Inc. located in Camarillo, California. Their capability is up to *1 kJ, 10 kV, 10 kA,* with an 8×20 pulse. With four *1 μF* capacitors in series, each charged to *400 V*, our system is capable of delivering only *320 mJ*.

A schematic representation of the test fixture is shown in Figure V-1-2.

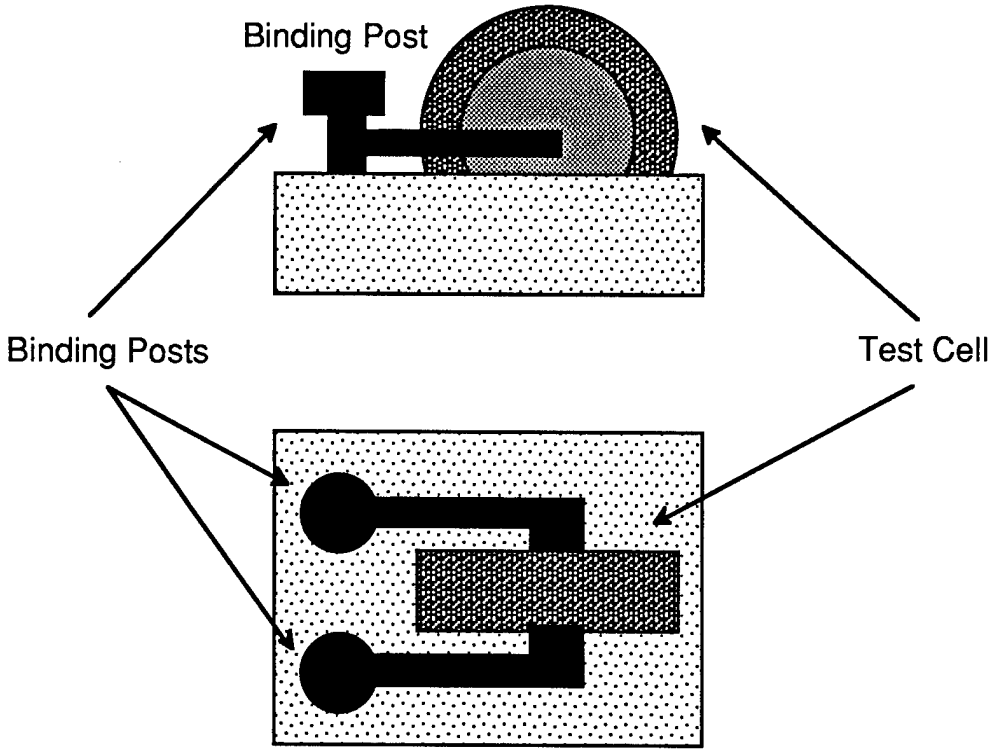


Fig. V-1-2 Test Fixture Designed to accommodate the test cells. The base is insulator. The all black parts are metallic. The test cell is electrically contacted and it is easy to change from one test cell to another.

V-1-5 Quantum Tunnelling Signature

In the results of the testing, we will be looking for characteristics indicating quantum tunnelling. In Subsection II-1-11 we had for device current the expression

$$I = K_V \, V^2 e^{-\beta_V / V}$$

for the resistance of the device this may be written as

$$R = \frac{e^{\beta_V / V}}{K_V V}$$

Matching this function at two different voltages gives the inversion relations

$$\ln K_v = \frac{1}{\Delta}\left[\frac{1}{V_1}\ln\left(\frac{1}{R_2 V_2}\right) - \frac{1}{V_2}\ln\left(\frac{1}{R_1 V_1}\right)\right] \qquad \beta_V = \frac{1}{\Delta}\ln\left(\frac{R_1 V_1}{R_2 V_2}\right)$$

where

$$\Delta = \frac{V_2 - V_1}{V_2 V_1}$$

For comparisons, we will usually match laboratory data with these relations at the end points of the data.

V-1-6 Test Procedure

The parallel resistance of the test fixture was first measured at 1 kV and found to be greater than 200 TΩ, thus the effect of test fixture leakage is expected to be negligible.

The following test procedure was planned:

1. Determine resistance at 1.5 V with an ohm meter. Any sample with a resistance below 1 MΩ was rejected and no further tests made.

2. Using a Keithley 6517 electrometer, the power delivered to the devices was measured in increments of 10 V. When the power level reached 0.1 W, this step of the testing process was terminated. Any devices shorting during this test were rejected.

2. Over the range of voltages determined in the previous test, resistances of surviving devices are measured as a function of applied voltage.

3. Samples surviving the above two tests are then pulse tested as previously described to determine clamp voltage.

4. Surving samples are then sent to Hyger Physics, Inc. for determination of energy to destruct.

V-1-7 Test Procedure Verification

Before testing of our samples, steps 1 - 3 of the above procedure were verified using a commercially available MOV as the device under test (AVX TVS

056 042). The results of the electrometer tests are plotted in Figure V-1-4 along with the quantum tunnelling curve as derived in the last section.

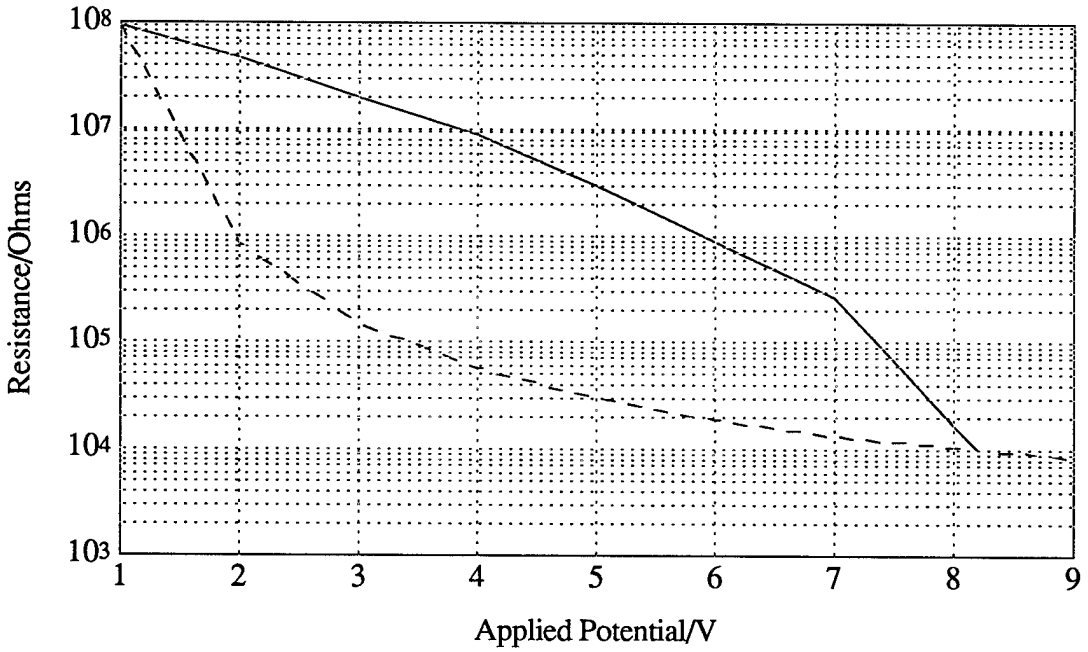


Fig. V-1-4 The solid curve is the measured resistance of a commercial MOV varistor. The dashed curve is the quantum tunnelling function matched at end points of the measured data.

From this plot it is evident that MOV's are not quantum tunnelling devices.[1]

The clamp voltage of this device was tested in 200 V increments from 400 V to 1.6 kV with a 50 Ω source and 50 Ω load. The clamp voltage was steady between 14 and 15 V. A repeat of the electrometer readings showed no appreciable change from the curve of Figure V-1-4.

[1]We do envy their the downward concavity.

Section V-2

SUMMARY OF TEST RESULTS

V-2-1 Revised Test Procedure

In the previous section, we described what we thought would be the ideal testing scenario prior to making the first measurement. Another title for this subsection might be, "Welcome to the Real World." The current procedure is relatively simple, and we believe it gives us the most information for time and samples spent. We outline it as follows:

1. Determine the clamp voltage starting at a low applied pulse voltage and moving up.

2. Determine the reverse leakage current with the electrometer with the reverse standoff voltage defined as half the clamp voltage.

3. Determine the destruct energy by continuing the applied pulse tests beyond the clamp voltage.

This sequence gives us all the essential electrical information for a given sample prior to the time the sample fails.

V-2-2 The Quantum Tunnelling Signature

Many of the test results are described in Part IV and will not be repeated here. We will concentrate on those aspects having the most likely impact on future development One of our major goals was to produce materials clearly exhibiting the quantum tunnelling signature. Throughout all of our electrometer measurements, a disturbing factor emerged. At constant applied voltage, the resistance readings would start at some lower value and increase as a function of time, thus masking any quantum tunnelling signature that might be underlying the observed data. The time constants for these resistance increases were quite long, on the order of hours. To study these effects further requires a strip chart recorder or an automatic data acquisition system, one of which will be implemented in Phase II. We suspect the rising resistances are the result of space charge effects.

At first, we thought these effects might portend a potential death knell. Upon further reflection, we realized that from a practical standpoint, our devices need only be of sufficiently high resistance throughout the operational mode and exhibit clamping action in the threat mode. Subsequently, devices with these characteristics were produced.

V-2-3 Conduction Mechanisms

The working materials produced to date appear to require insulative barrier thicknesses outside the range expected for pure quantum tunnelling. Part of this may be caused by morphological irregularities of the base powders with an attendant field enhancement at positions with high radii of curvature. Alternatively, there may be other conduction mechanisms present under high field conditions, such as von Hippel or Fröhlich. These possibilities will be considered in Phase II.

V-2-4 Dominant Path Failure

For the most part, samples that fail, do so through the dominant path mechanism. That is, a very small conductive path is created through the material. This is predicted by our computer simulations and is confirmed by observing the electrodes of failed samples. The electrodes show spots about the size of a period on this page and the spots on opposing electrodes are directly above one another, indicating a path straight through the material perpendicular to the electrodes. Computer modelling indicates two criteria for this situation to exist, positive feedback and statistical variation. Reducing either or both of these will aid in ameliorating the dominant path phenomenon.

The statistical variations may not be as damaging as one might first expect. An aluminum powder, WG 20,[1] has a high degree of statistical variation and does fail dominant path. However, the powder failure point is well within the bounds to make it a successful ESD arrestor.

We have produced samples with greatly reduced statistical variation. When examining the electrodes of these samples, the pattern is in the form of clouds rather than a single point, backing up our contention that reduction in statistical variation reduces dominant path; this is also supported by computer simulation.

As indicated in Part II, the addition of bulk resistance to our materials will reduce the positive feedback effects. This is accomplished by using insulative spheres with a conductive coating and overcoating with an insulative barrier.

V-2-5 Binder Effects

Another objective of this effort was the production of materials whose electrical properties were independent of the binder. A first step in the direction of demonstrating the feasibility of this approach was to show that powders with no binder exhibit the desirable properties, and this has been accomplished. Two associated electrical properties of desirable binders are either high dielectric strength or high electron transparency in ultra thin

[1]To be subsequently described.

films. The other binder properties are associated with the mixing process and mechanical integrity of the finnished product.

In some of the organic binders tested, dominant path failure was evident, with the dominant path being formed by a carbon chain extracted from the binder. This effect wuld be ameliorated by high dielectric strengths or high electron transparency.

V-2-6 Aluminum/Aluminum Oxide

This powder consists of aluminum with nominal diameters of 20 μm. It was coated with aluminum oxide to a depth of between 500 and 1000 Å. Morphologically, this powder is a potential disaster. It has high statistical variability and a wide variety of shapes, some with points. With a 1 mm interelectrode spacing samples fail dominant path at about 100 mJ. With a 2 mm spacing, no failure occurred at the highest level of our test system (211 mJ).

Three test cells containing this powder were independently tested at Littelfuse, Inc. to determine if it would meet their ESD requirements. The tests were performed to IEC 801-2 specifications and were completely successful. The threat voltages ranged from 1 to 11 kV. Many pulses were applied and as measured by a DMM, the samples remained open. The clamp voltages were at 200 V. Upon returning to Tulsa, the off state resistances were measured with an electrometer at 100 V reverse standoff voltage with all three samples well in the GΩ range.

The resolution of the Littelfuse test system was 500 ps. Although less than other materials tested by Littelfuse, there appeared to be some leading edge overshoot. However, it was demonstrated that this same overshoot is present with a blank test cell, so is likely an artefact of the test set-up. This contention is backed up by measurements shown in Figures V-2-1 and V-2-2.

The first is a sample labelled WG 22; the powder is WG 20. The applied voltage is 11 kV, with an apparent clamp at about 200 V. There appears to be a leading edge overshoot and the waveform is not particularly clean.

The second is a plot of the measurement system response to a blank test cell, that is, one containing no varistor material with the electrodes pushed together forming a dead short. The applied voltage is the same as that of the previous plot. It is apparent that there is a high level of system parasitics, and there is a high degree of correlation between the two plots.

If these two plots are superimposed and the parasitics visually subtracted, one would be led to believe the varistor material response is relatively clean, with no overshoot, and a clamp voltage of about 150 V. The likely reason for the lower initial spike with the varistor material present is that the digital sampling did not occur at the peak value.

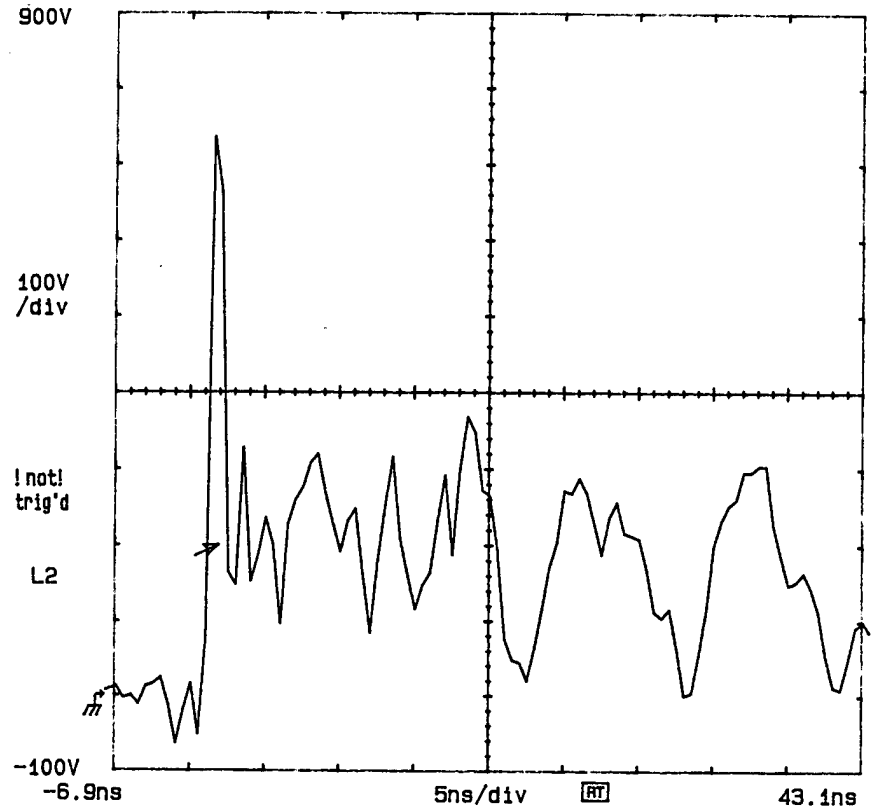


Fig. V-2-1 Measurement with varistor material present.

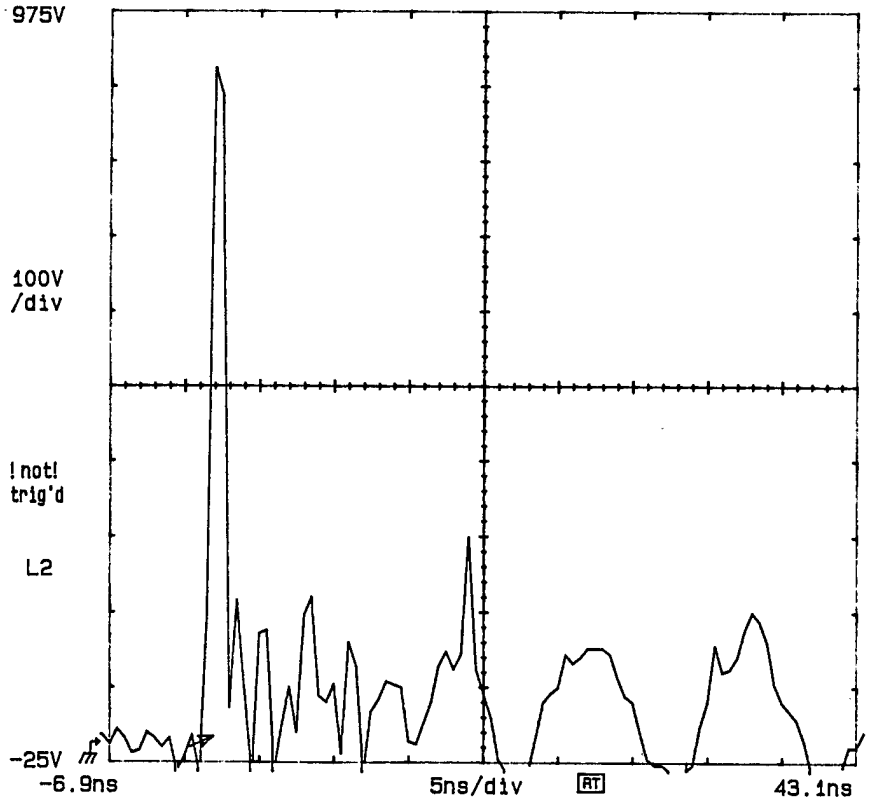


Fig. V-2-2 Test system response to a blank test cell.

These tests are particularly interesting since there was no binder present, demonstrating the desired clamping characteristics independently of a binder. Subsequent to these tests, Littelfuse incorporated a powder of this type into a binder and produced a prototype device with essentially the same results previously obtained, demonstrating the results are still independent of the binder. This was a major goal of this effort, *production of devices whose electrical properties are independent of the binder used.*

V-2-7 Glass/Nickel/Glass

In this sample, glass spheres were coated with nickel and overcoated with glass. A sequence of oscilloscope traces are shown in Figures V-2-3 through V-2-7. Of particular note is the absolute consistency of the clamp voltage at 200 V with an applied pulse of 8 kV. There is still an apparent leading edge overshoot and we still believe this is an artefact of the test system. This is supported by Figure V-2-5 which exhibits the characteristic of an LCR ringdown.

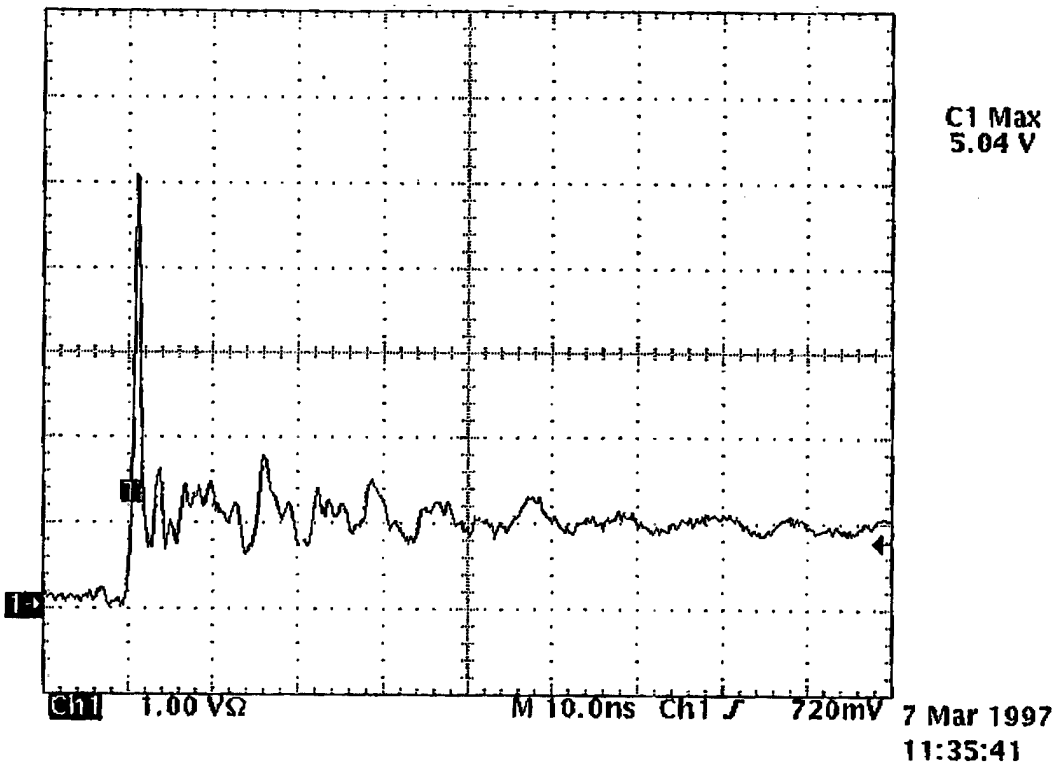


Fig. V-2-3 First pulse, 200 V/div., 8 kV applied.

This sample exemplifies the Integrated Sciences™ three phase concept and demonstrates the concept has potential for production of useful devices. Additional traces will be found on the following two pages.

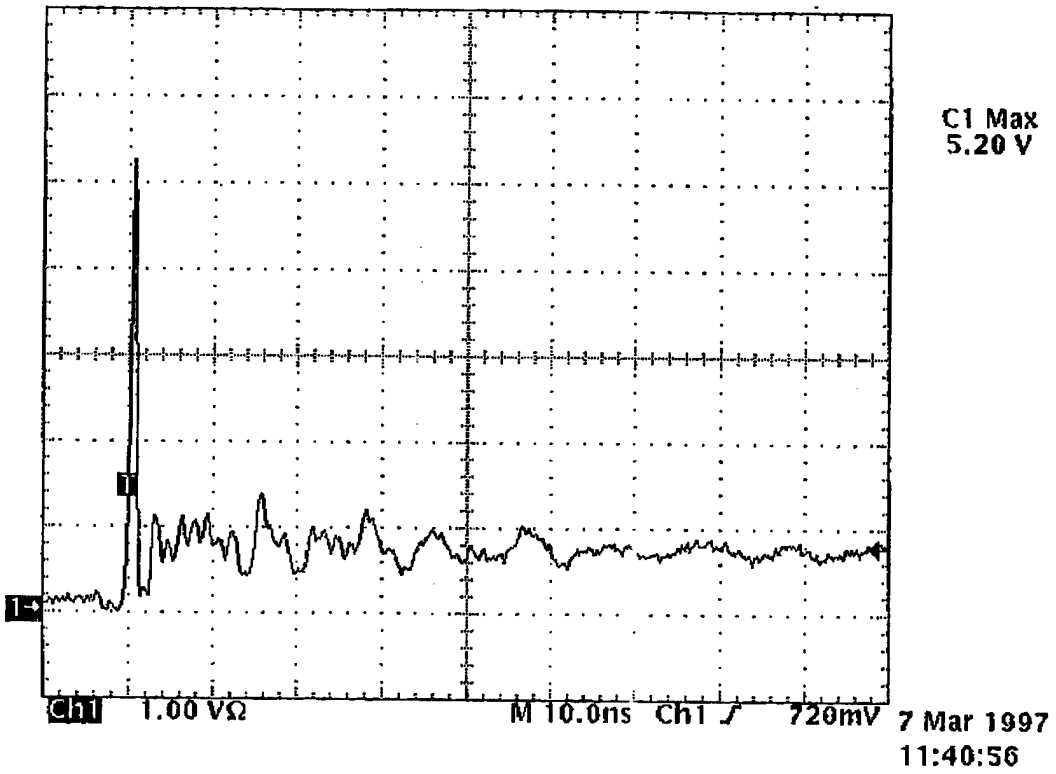


Fig. V-2-4 Second pulse, 200 V/div., 8 kV applied.

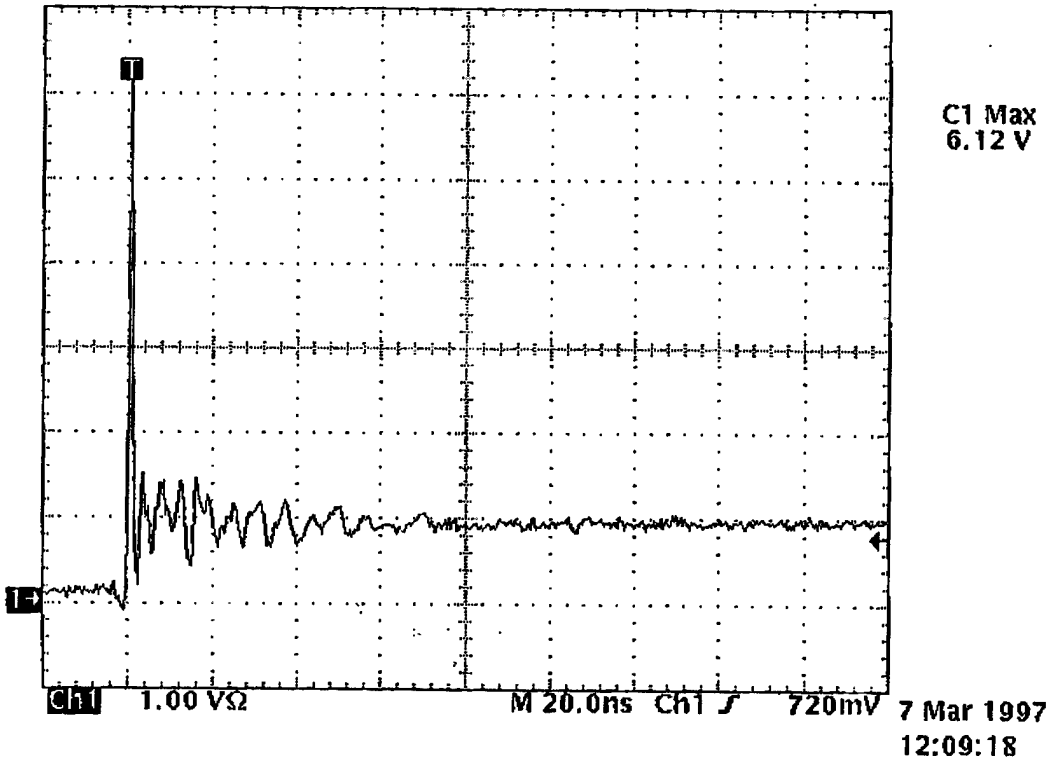


Fig. V-2-5 Tenth pulse, 200 V/div., 8 kV applied.

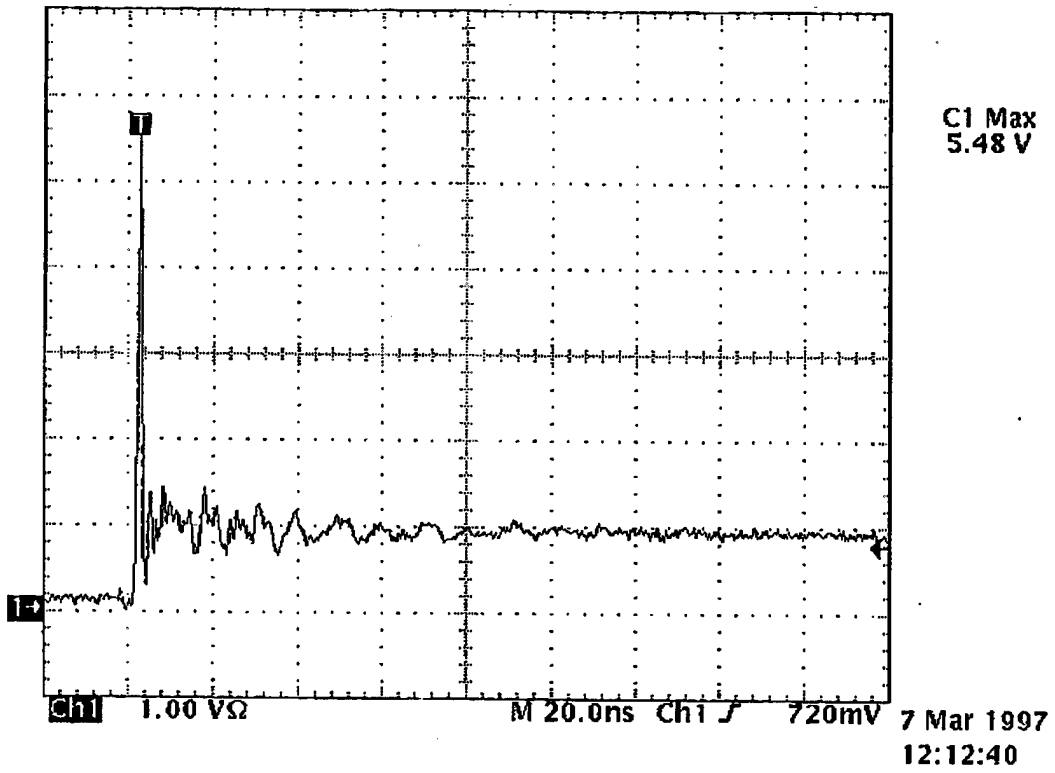


Fig. V-2-6 Fifteenth pulse, 200 V/div., 8 kV applied.

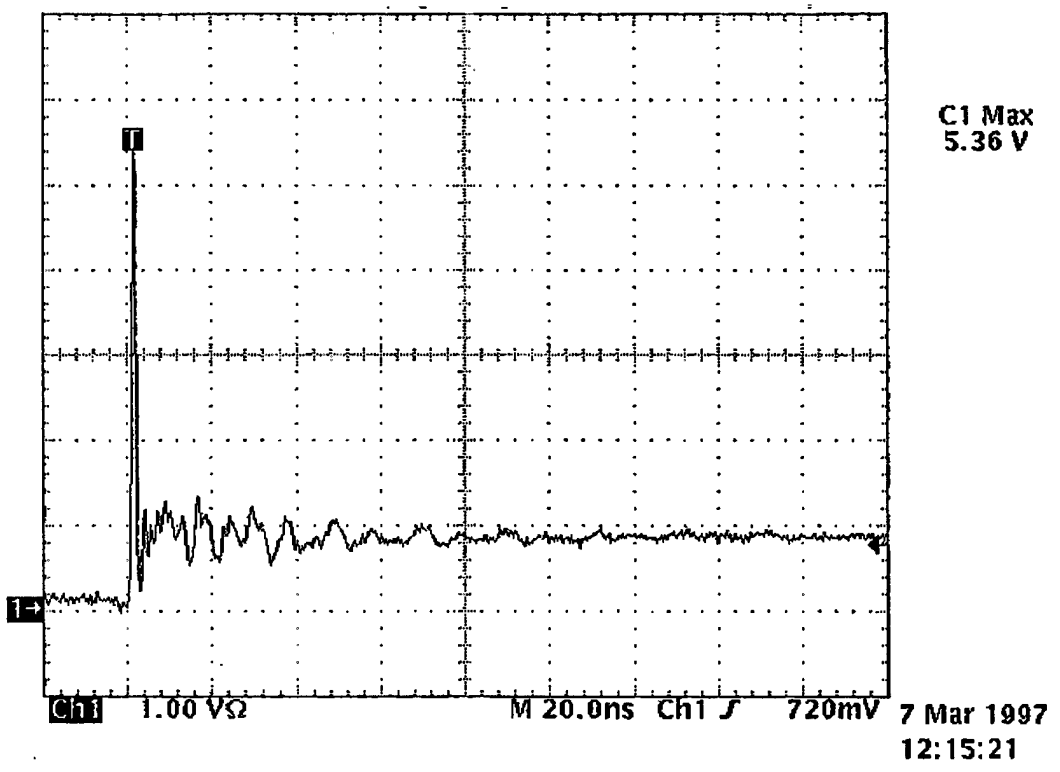


Fig. V-2-7 Twentyfifth pulse, 200 V/div., 8 kV applied.