Spring 1968

The mechanism of second breakdown in transistors

Eugene David Fabricius

New Jersey Institute of Technology

Follow this and additional works at: https://digitalcommons.njit.edu/dissertations

Part of the Electrical and Electronics Commons

Recommended Citation
https://digitalcommons.njit.edu/dissertations/1334

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Digital Commons @ NJIT. It has been accepted for inclusion in Dissertations by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.
Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be “used for any purpose other than private study, scholarship, or research.” If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “fair use” that user may be liable for copyright infringement.

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation.

Printing note: If you do not wish to print this page, then select “Pages from: first page # to: last page #” on the print dialog screen.
The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.
68-16,971

FABRICIUS, Eugene David, 1929-
THE MECHANISM OF SECOND BREAKDOWN IN TRANSISTORS.
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan
THE MECHANISM OF SECOND BREAKDOWN IN TRANSISTORS

BY

EUGENE DAVID FABRICIUS

A DISSERTATION

PRESENTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF SCIENCE IN ELECTRICAL ENGINEERING

AT

NEWARK COLLEGE OF ENGINEERING

This dissertation is to be used only with due regard to the rights of the author. Bibliographical references may be noted, but passages must not be copied without permission of the College and without credit being given in subsequent written or published work.

Newark, New Jersey
1968
The purpose of this dissertation is to examine the behavior of electrons and holes in a semiconductor or diode under conditions of high current density as a function of temperature, and to relate this behavior to the phenomenon of Second Breakdown. The approach used is that of magnetohydrodynamics, the electrons and holes being treated as a plasma "gas" embedded in the dielectric of the semiconductor.

This approach is unique in the following respects:
1. This is the first attempt to explain second breakdown in terms of magnetohydrodynamics.
2. This is the first time an explanation of pinching in a solid at room temperature has been presented which does not rely on some type of crystal imperfection to initiate the pinching.
3. This is the first time variations in the forbidden gap width have been considered as causing voltage drops, and therefore, electric fields in a semiconductor.

The author is convinced that there are really two types of second breakdown, depending upon the emitter bias. The first type of second breakdown occurs when the emitter-base junction is forward-biased, in which case current constrictions are due to pinching of
electrons and holes in the base region of the device. This is examined in Part I of the dissertation, where the theory is developed for low electric fields.

Computer calculations of electron-hole concentration and temperature versus distance from the hot-spot center are presented along with infrared data obtained for temperature versus distance for several measured hot spots. Agreement between theory and data is very good.

The theory predicts that second breakdown is due to thermal effects at or near room temperature, and due to magnetic effects at or near liquid nitrogen temperature. This leads to the definition of the transition temperature as an indication of the temperature at which the transition occurs between second breakdown due to Joule heating, and second breakdown due to magnetic pinching.

The most striking conclusion to be drawn from the computer results is that the theory predicts that units are much more susceptible to failure at lower temperatures than at higher temperatures, contrary to popular opinion.

The second type of second breakdown occurs when the emitter-base junction is reverse-biased, in which case
current constrictions are due to avalanche effects in the collector-base junction. This is examined in Part II of the dissertation, where the theory is developed for high electric fields. A critical current for the onset of second breakdown is determined as a function of electric field in the collector depletion region. Comparison of published data on the temperature dependence of second breakdown with theory is given first. Then data taken on a reverse-biased test set are presented. The temperatures investigated are 77°K, 195°K, 273°K, and 300°K.

Theory predicts that devices are much more easily driven into second breakdown at low temperatures than at high temperatures, and this is verified experimentally. Experimental agreement with the theory is excellent over the entire temperature range investigated, in complete agreement with the theoretical results obtained in Part I.

Actual devices will have flaws and defects in them, and will fail at power levels below their theoretical capabilities. This leads to the definition of a quality factor, which is a measure of the actual performance of a given device as compared to its theoretical capability.
APPROVAL OF DISSERTATION
THE MECHANISM OF SECOND BREAKDOWN IN TRANSISTORS
BY
EUGENE DAVID FABRICIUS
FOR
DEPARTMENT OF ELECTRICAL ENGINEERING
NEWARK COLLEGE OF ENGINEERING

BY
FACULTY COMMITTEE

APPROVED: __________________________, Chairman

______________________________

______________________________

______________________________

______________________________

NEWARK, NEW JERSEY
JUNE, 1968
ACKNOWLEDGMENTS

The author wishes to thank Mr. William Kerns, Executive Vice-President of Solitron Devices, and Mr. Richard Trivison, Engineering Vice-President of Solitron Devices, for providing facilities to develop the transistor test set and for donating it to the college after construction, and also for donating several high-power transistors.

The author also wishes to thank the Newark College of Engineering Research Foundation and the National Science Foundation for granting him fellowships which made this research possible, and the various members of the Departments of Physics and Electrical Engineering who have helped with their comments and criticisms of this work.

Special thanks are due to Dr. Phyllis Fox and Mr. Victor Miller of the Computer Center for their assistance with the programming of various equations, to Dean Alex Bedrosian and Mr. James Earle for their continuous administrative assistance, and to Mr. André Buser of Solitron Devices, under whose supervision the
test set was built.

The author wishes to express his sincere appreciation and gratitude to his academic advisor, Dr. Raj P. Misra, for accepting the advisorship of this project and for providing constant criticism, encouragement, and leadership not only on this dissertation but throughout the entire doctoral program, and to Dr. Kenneth S. Sohn, Dr. Joseph J. Padalino, and Dr. Paul O. Hoffmann for their most helpful comments and criticisms of this work.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>History of Second Breakdown</td>
<td>1</td>
</tr>
<tr>
<td>The Characteristics of Second Breakdown</td>
<td>3</td>
</tr>
<tr>
<td>Early Theories of Second Breakdown</td>
<td>10</td>
</tr>
<tr>
<td>Lateral Thermal Stability</td>
<td>12</td>
</tr>
<tr>
<td>Turnover Phenomena and Second Breakdown</td>
<td>13</td>
</tr>
<tr>
<td>Mesoplasmas and Second Breakdown</td>
<td>15</td>
</tr>
<tr>
<td>Thermal Breakdown and Second Breakdown</td>
<td>15</td>
</tr>
<tr>
<td>Avalanche Breakdown and Second Breakdown</td>
<td>16</td>
</tr>
<tr>
<td>Fine Structure in Second Breakdown</td>
<td>17</td>
</tr>
<tr>
<td>Beta Dependence of Second Breakdown</td>
<td>18</td>
</tr>
<tr>
<td>Miscellaneous Theories of Second Breakdown</td>
<td>19</td>
</tr>
<tr>
<td>Miscellaneous Comments on Second Breakdown</td>
<td>19</td>
</tr>
<tr>
<td><strong>PART I: CURRENT CONSTRICTIONS AND SECOND BREAKDOWN</strong></td>
<td></td>
</tr>
<tr>
<td><strong>II. THEORY</strong></td>
<td>22</td>
</tr>
<tr>
<td>The Boltzmann Transport Equation</td>
<td>22</td>
</tr>
<tr>
<td>Ambipolar Diffusion and the Conservation Equations</td>
<td>27</td>
</tr>
<tr>
<td>Introduction</td>
<td>27</td>
</tr>
<tr>
<td>Conservation of Charge</td>
<td>33</td>
</tr>
<tr>
<td>Conservation of Momentum</td>
<td>35</td>
</tr>
<tr>
<td>Conservation of Energy</td>
<td>37</td>
</tr>
<tr>
<td>Transient Analysis</td>
<td>38</td>
</tr>
<tr>
<td>Magnetic Pinching Force</td>
<td>42</td>
</tr>
<tr>
<td>Electrical Pinching Force</td>
<td>43</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Steady-State Analysis</td>
<td>49</td>
</tr>
<tr>
<td>Special Cases</td>
<td>51</td>
</tr>
<tr>
<td>Power Delivered to the Plasma</td>
<td>51</td>
</tr>
<tr>
<td>III. CALCULATIONS</td>
<td>53</td>
</tr>
<tr>
<td>IV. RESULTS</td>
<td>58</td>
</tr>
<tr>
<td>Theoretical Results</td>
<td>58</td>
</tr>
<tr>
<td>Experimental Results</td>
<td>68</td>
</tr>
<tr>
<td>V. CONCLUSIONS</td>
<td>77</td>
</tr>
<tr>
<td>PART II: CRITICAL CURRENT TO INITIATE SECOND BREAKDOWN</td>
<td></td>
</tr>
<tr>
<td>VI. THEORY</td>
<td>83</td>
</tr>
<tr>
<td>Critical Current for Pinching</td>
<td>83</td>
</tr>
<tr>
<td>Critical Current as a Function of Temperature and Field</td>
<td>88</td>
</tr>
<tr>
<td>VII. RESULTS</td>
<td>96</td>
</tr>
<tr>
<td>VIII. CONCLUSIONS</td>
<td>105</td>
</tr>
<tr>
<td>IX. RECOMMENDATIONS</td>
<td>108</td>
</tr>
<tr>
<td>Theoretical Investigations</td>
<td>108</td>
</tr>
<tr>
<td>Experimental Investigations</td>
<td>109</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>110</td>
</tr>
<tr>
<td>APPENDICES</td>
<td></td>
</tr>
<tr>
<td>A. THE GENERAL TRANSPORT EQUATIONS</td>
<td>124</td>
</tr>
<tr>
<td>The Continuity Equation</td>
<td>124</td>
</tr>
<tr>
<td>The Momentum Transport Equation</td>
<td>126</td>
</tr>
<tr>
<td>The Energy Transport Equation</td>
<td>129</td>
</tr>
<tr>
<td>B. THE SPACE-CHARGE NEUTRALITY CONDITION</td>
<td>134</td>
</tr>
<tr>
<td>C. GENERATION AND RECOMBINATION OF CARRIERS</td>
<td>138</td>
</tr>
</tbody>
</table>
## APPENDICES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special Cases</td>
<td>139</td>
</tr>
<tr>
<td>Space-Charge Neutrality</td>
<td>139</td>
</tr>
<tr>
<td>N-Type Material</td>
<td>139</td>
</tr>
<tr>
<td>Low-Level Injection</td>
<td>139</td>
</tr>
<tr>
<td><strong>D. MAXWELLIAN VELOCITY DISTRIBUTION</strong></td>
<td>142</td>
</tr>
<tr>
<td>Electron Temperature</td>
<td>144</td>
</tr>
<tr>
<td>Electron-Hole Pressure</td>
<td>145</td>
</tr>
<tr>
<td><strong>E. FIELDS OF A CHARGE IN UNIFORM MOTION</strong></td>
<td>149</td>
</tr>
<tr>
<td><strong>F. TRANSISTOR SECOND-BREAKDOWN TEST DEVICE</strong></td>
<td>153</td>
</tr>
<tr>
<td><strong>G. RESISTIVITY VERSUS INTRINSIC TEMPERATURE</strong></td>
<td>160</td>
</tr>
<tr>
<td><strong>BIBLIOGRAPHY ON SECOND BREAKDOWN</strong></td>
<td>170</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Second Breakdown in the Reverse-Biased Emitter Configuration</td>
<td>4</td>
</tr>
<tr>
<td>2.</td>
<td>The Locus of Trigger Points for Second Breakdown</td>
<td>6</td>
</tr>
<tr>
<td>3.</td>
<td>Common-Emitter Characteristics of a Transistor</td>
<td>7</td>
</tr>
<tr>
<td>4.</td>
<td>Effect of Emitter Bias on Second Breakdown</td>
<td>8</td>
</tr>
<tr>
<td>5.</td>
<td>Variation in Energy Bands Due to Localized Heating</td>
<td>45</td>
</tr>
<tr>
<td>6.</td>
<td>Decrease in the Forbidden-Gap Width Due to Localized Heating</td>
<td>47</td>
</tr>
<tr>
<td>7.</td>
<td>Silicon Power Transistor Operated at 300°K with High Input Power</td>
<td>59</td>
</tr>
<tr>
<td>8.</td>
<td>Germanium Power Transistor Operated at 300°K with High Input Power</td>
<td>62</td>
</tr>
<tr>
<td>9.</td>
<td>Silicon Unit Operated at 77°K with Medium Input Power</td>
<td>64</td>
</tr>
<tr>
<td>10.</td>
<td>Germanium Unit Operated at 77°K with Medium Input Power</td>
<td>66</td>
</tr>
<tr>
<td>11.</td>
<td>Temperature versus Distance for a Typical Hot Spot</td>
<td>70</td>
</tr>
<tr>
<td>12.</td>
<td>Temperature versus Distance for Two Hot Spots in Close Proximity</td>
<td>71</td>
</tr>
<tr>
<td>13.</td>
<td>Comparison of Theory and Data for a Typical Hot Spot</td>
<td>73</td>
</tr>
<tr>
<td>14.</td>
<td>Comparison of Theory and Data for Three Additional Hot Spots</td>
<td>76</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>15.</td>
<td>$R/\xi^2$ versus Temperature for Silicon</td>
<td>90</td>
</tr>
<tr>
<td>16.</td>
<td>Critical Current versus Temperature in Germanium</td>
<td>92</td>
</tr>
<tr>
<td>17.</td>
<td>Critical Current versus Temperature in Silicon</td>
<td>93</td>
</tr>
<tr>
<td>18.</td>
<td>Comparison of Schiff's Data with the Theoretical Curve for Silicon</td>
<td>95</td>
</tr>
<tr>
<td>20.</td>
<td>Critical Current versus Temperature for Several 2N4128 Devices</td>
<td>103</td>
</tr>
<tr>
<td>21.</td>
<td>Electric and Magnetic Fields Acting on a Hole and Electron Moving in Opposite Directions</td>
<td>150</td>
</tr>
<tr>
<td>22.</td>
<td>Block Diagram of Second-Breakdown Test Circuit</td>
<td>156</td>
</tr>
<tr>
<td>23.</td>
<td>Typical Current and Voltage Pulses Obtained with the Test Set</td>
<td>157</td>
</tr>
<tr>
<td>24.</td>
<td>Duty Cycle Utilized by Test Set</td>
<td>159</td>
</tr>
<tr>
<td>25.</td>
<td>Initial Resistivity versus Intrinsic Temperature for Germanium</td>
<td>164</td>
</tr>
<tr>
<td>26.</td>
<td>Initial Resistivity versus Intrinsic Temperature for Silicon</td>
<td>165</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Data for Calculations</td>
<td>54</td>
</tr>
<tr>
<td>II.</td>
<td>Constants Used in Computer Program for Results Shown in Figures 7 through 10</td>
<td>56</td>
</tr>
<tr>
<td>III.</td>
<td>Data Used to Compute R</td>
<td>87</td>
</tr>
<tr>
<td>IV.</td>
<td>Data on Four 2N3599 Devices</td>
<td>98</td>
</tr>
<tr>
<td>V.</td>
<td>Data on Eight 2N4128 Devices</td>
<td>102</td>
</tr>
</tbody>
</table>
**LIST OF UNUSUAL SYMBOLS**

**USED IN THIS PAPER**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\triangleq$</td>
<td>is equal to, by definition.</td>
</tr>
<tr>
<td>$\equiv$</td>
<td>is identical to.</td>
</tr>
<tr>
<td>$\vec{a}$</td>
<td>vector symbol.</td>
</tr>
<tr>
<td>$\hat{a}$</td>
<td>unit vector symbol.</td>
</tr>
<tr>
<td>$\mathbf{H}$</td>
<td>tensor of rank two.</td>
</tr>
<tr>
<td>$\iiint_\mathcal{V}$</td>
<td>the integral over the entire volume bounded by a surface.</td>
</tr>
<tr>
<td>$\iint_\mathcal{S}$</td>
<td>the integral over the entire closed surface bounding a volume, with the direction of the outward normal taken as positive.</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

History of Second Breakdown

Second breakdown* was first reported in 1958 by Thornton and Simmons,\(^1\) who suggested that this effect was the cause of certain types of failures occurring in some devices. Schafft and French\(^2\) found that every type of transistor having leads capable of passing the required current could be driven into second breakdown, and concluded that second breakdown is a very fundamental property of transistors.

Thermal runaway is a well-known problem in large power transistors, and involves the entire transistor base.\(^3\) However, the internal current distribution can change significantly with changes in operation conditions.\(^4\)\(^-\)\(^7\) Second breakdown in transistors and diodes is associated with a sudden concentration of current in a small area.\(^1\),\(^2\),\(^8\)\(^-\)\(^24\),\(^99\)

At a certain critical internal temperature, the

current density and temperature tend to build up in one small region, resulting in what is referred to as a "hot spot" or "microplasma." This sudden increase in current dissipates power in a very small volume, which may cause the temperature of the hot spot to be much higher than the assumed uniform temperature of the junction.

Ford\textsuperscript{25} suggested that second breakdown would occur when the temperature of the collector junction rises to 200°C. Tauc\textsuperscript{10} observed hot spots at temperatures of 600°C to 780°C associated with second breakdown in silicon p-n junctions. Haitz\textsuperscript{26} observed that the temperature at the surface of a second breakdown region is above 830°C, and decided that the maximum temperature in the bulk may be over 1000°C.

If constant voltage is maintained across the device, the current builds up rapidly and local melting and burnout occur.\textsuperscript{11,27,28} Also, more drastic current constriction occurs in reverse-biased emitter than in forward-biased emitter operation.\textsuperscript{22}

The current constrictions are thought to be due to the same effect as encountered in the parallel operation of power transistors\textsuperscript{29} and silicon-controlled-rectifiers.\textsuperscript{30} This implies nonuniformities in the base region of the device to initiate current density buildup.
Collector and emitter junctions are never perfectly uniform\textsuperscript{31,32} and current nonuniformities can be caused by nonuniformities in the base width\textsuperscript{33} and in heat sinking,\textsuperscript{34} as well as by localized regions of reduced collector-base avalanche breakdown voltage.\textsuperscript{35}

However, in guard-ring diodes, current concentrations usually take place near the center of the active area\textsuperscript{11,26,36,37} while all devices with reasonably uniform pre-second-breakdown current distributions generally go into second breakdown near the center of the emitter.\textsuperscript{38}

The catastrophic failure mechanism in second breakdown is a collector-to-emitter short-circuit, localized in a region of at most a few mils in diameter.\textsuperscript{1}

The Characteristics of Second Breakdown

Second breakdown was first reported under swept conditions where an oscilloscope was used to observe the $V_{CE}$-$I_C$ characteristics using a 120 cps swept voltage.\textsuperscript{1} Figure 1 shows the breakdown characteristic of $V_{CE}$ for a unit exhibiting second breakdown in the reverse-biased emitter configuration. When the device begins to break down, the voltage first drops slightly with increasing current (First Breakdown), then drops drastically and the unit goes into Second Breakdown.
Figure 1.--Breakdown characteristics of device operated in common emitter configuration with emitter junction reverse biased.
Second breakdown is associated with the collector-base junction, but is controlled by the emitter-base junction. Second breakdown is triggered at lower collector voltages as the base drive voltage is changed from reverse bias of the emitter-base junction to forward bias, during which time the collector current increases. Figure 2 shows the empirically determined locus of triggering points for second breakdown which is often observed and is now attributed to a thermal cause. Figure 3 shows the effect of base current on a transistor operated into second breakdown.

The swept $V_{CE}-I_C$ characteristics are sketched in Figure 4 for three different constant base current drive conditions; the letters R, F, and O signify, respectively, reverse base drive operation, forward base drive operation, and operation with the base open-circuited. For sufficiently large collector currents, each of the three curves drawn in Figure 4 shows an extremely abrupt decrease in $V_{CE}$. This drop in voltage is the most obvious indication of the initiation of second breakdown.

A behavior that Schafft and French believe to be characteristic of second breakdown is the positional order reversal of the usual dependence of $V_{CE}$ on the polarity of the base current. This reversal is shown in Figure 4 where the order of voltage levels for curves
Figure 2.--The locus of trigger points for second breakdown as a function of $V_{EB}$. Back-biasing the emitter junction increases the collector voltage at which second breakdown can be triggered, while forward-biasing the emitter junction reduces the collector voltage at which second breakdown can be initiated.
Figure 3.--Common-emitter characteristics of a typical device, with \( I_B \) in arbitrary units. The device will enter second breakdown where the I-V curves cross the locus of trigger points for second breakdown.
Figure 4.--Swept $I_C$-$I_{CE}$ characteristics of a transistor with constant forward (F), zero (0), and constant reverse (R) emitter-to-base current. The characteristics are drawn for only the first half of the sweep cycle. The initiation of second breakdown is indicated by the extremely abrupt drop in $V_{CE}$, indicated by the dotted lines. Note the reversal of the order of F, 0, and R, after the device enters second breakdown.
R, F, and O is inverted upon entering second-breakdown operation.

Second breakdown can also be observed under dc conditions where the transistor is placed in series with a load resistor and a dc voltage supply, and set at a particular operating point in the $V_{CE}-I_C$ plane.

Schafft and French\textsuperscript{2} found that when the transistor current and voltage were such that second breakdown is possible, $V_{CE}$ will remain constant for a time called the "delay time," which is generally of the order of milliseconds. At the end of this delay time, the voltage will begin to drop to the low voltage level. The transition time for this drop to occur was found to be of the order of microseconds or more for some transistors.

The observation that there is always a delay time before second breakdown occurs led Schafft and French\textsuperscript{2} to conclude that if conditions are suitable for second breakdown, then a certain energy, which they called the "triggering energy," must be dissipated in the transistor before second breakdown will occur.\textsuperscript{16,40} This energy dissipation concept can explain the characteristics of second breakdown under swept conditions.\textsuperscript{41} Also, this energy dissipation concept serves as the foundation for many of the methods used by the
transistor industry to specify operating conditions free of second breakdown.\textsuperscript{8}

The delay time, or equivalently, the triggering energy, associated with a particular set of operating conditions can be used as a measure of the susceptibility of the device to the development of second breakdown. The longer the delay time and the greater the triggering energy, the more resistant the device is to second breakdown.

The initiation of second breakdown is characterized by a very rapid and drastic decrease in voltage, accompanied by a sudden and marked constriction of the current in the base region. This constriction of current and accompanying hot spot has been seen with the use of temperature-sensitive paints,\textsuperscript{9} phosphors,\textsuperscript{39,42} electrical probes for devices with interdigitated emitters,\textsuperscript{9} infrared scanning techniques\textsuperscript{43,100} and the scanning electron microscope.\textsuperscript{44-48}

\textbf{Early Theories of Second Breakdown}

The idea that second breakdown is a surface phenomenon was rejected by Thornton and Simmons,\textsuperscript{1} and their conclusion has been verified by later work.\textsuperscript{2} No great difference in second breakdown characteristics has been observed between devices with well-passivated surfaces
and those with poor surfaces.\textsuperscript{2} Also, the high current density sites are now known to be internal.

Thornton and Simmons\textsuperscript{1} suggested the "pinch-in" effect, which is a localization of high emitter current density due to potential gradients developed in the base region by the flow of reverse base current, as a tentative explanation for second breakdown. They proposed that, at sufficiently high currents, some type of regenerative condition would be reached which would lead to second breakdown.

For forward base drive, the constriction would occur near the rim of a disk-emitter; while, for reverse base drive, the constriction would occur near the center of the emitter.\textsuperscript{1} The pinch-in effect should not be important near open base operation where potential gradients across the emitter are at a minimum, yet second breakdown does occur under these conditions.

Lin et al.\textsuperscript{49} suggested an explanation in terms of p-n-p-n action. However, the dynamic resistance in second breakdown has a negative slope which is opposite to that predicted by p-n-p-n theory, and the theory is unable to explain the long delay time observed for second breakdown.\textsuperscript{2,39}

Avalanche injection can produce a large drop in
voltage and a negative resistance characteristic similar to that observed in second breakdown. However, the onset of avalanche injection is too fast, being of the order of nanoseconds.

**Lateral Thermal Stability**

A study of the thermal stability of transistor structures was first described by Scarlett and Shockley in 1962,\(^\text{14}\) and later enlarged by them.\(^\text{9,12,15,34}\) A similar study was made independently by Bergmann and Gerstner in Germany.\(^\text{50}\) This instability is a lateral instability in which current builds up in one region of the device and decreases in the remaining area so that a region of high local current density and high temperature is produced. The theory was developed for the case where avalanche multiplication could be ignored.

Both groups studied a transistor model with a constant total emitter current, a constant applied voltage across the emitter-base junction, and an initially uniform current distribution. Thermal instability can occur if a perturbation of this distribution at one place leads to a buildup of current there at the expense of other regions of the transistor. If the associated localized temperature rise results in an increase in injection current which is larger than the original current perturbation, the current distribution becomes
unstable, and this small part of the device bears almost all of the current. This instability can occur independently of possible diffusion defects in the transistor, and although the device is loaded below the theoretical maximum power dissipation, local overheating can occur at these hot spots. A stability index can be derived to determine when the device will go into second breakdown.\textsuperscript{9,12}

Other investigators also believe that second breakdown is thermal in nature.\textsuperscript{11,13,27,28,51-53} Tauc\textsuperscript{10} observed a V-I characteristic which is nearly hyperbolic, which lends support to the thermal model for second breakdown. While the theory of thermal instability agrees with the features observed in forward-biased second breakdown, it does not predict correctly the behavior of devices operated with reverse biased emitter junctions.\textsuperscript{54,55}

**Turnover Phenomena and Second Breakdown**

Melchior and Strutt\textsuperscript{56,57} believe that second breakdown is due to the disappearance of the space charge region when the density of mobile carriers in the junction becomes equal to the density of dopant atoms, and an intrinsic zone is formed which short circuits the space-charge region of the p-n junction, and causes a voltage reduction across the transistor.
Melchior\textsuperscript{56} found that the temperature at which second breakdown occurs coincides with the intrinsic temperature of the semiconducting material used. Fujinuma\textsuperscript{58} has shown that second breakdown damage can occur in a few microseconds, and that a temperature rise in such a short time is possible. This model is supported by experimental results of Ford\textsuperscript{25} and Weitzsch.\textsuperscript{59}

Agatsuma observed the above effect in n\textsubscript{v}n devices, and labeled it the "turnover" phenomenon.\textsuperscript{60-63} He plotted the conductivity of various v layers at the surface temperature of turnover,\textsuperscript{63} and found them to correspond to the intrinsic conductivities of silicon as given by Putley and Mitchel.\textsuperscript{160} He concluded that turnover occurs when the surface temperature of the v layer approaches the intrinsic temperature of the layer. The field strength at which turnover occurs is much lower than the avalanche field strength of the material.\textsuperscript{5}

Agatsuma concluded that the onset of turnover is due to an increase in carrier concentration and current constriction; and that second breakdown in transistors results from a combined state of turnover breakdown in the emitter, base, and collector regions, with the added complication of a high field in the collector-base junction.
Mesoplasmas and Second Breakdown

According to a theory by English, second breakdown is initiated by a microscopic melt at the collector junction. He labeled this melt a mesoplasma. English and others, originally investigated this phenomenon in diodes, and later extended the theory to transistors.

Weitzsch decided that a melt at a junction might not be stable, and Schafft and French think that second breakdown cannot occur in diodes at all.

Thermal Breakdown and Second Breakdown

Takagi and Mano consider second breakdown to be due to a runaway of the collector current due to the temperature rise of the collector junction. Diebold has calculated the junction temperature rise, and Strickland has derived a thermal equivalent circuit for the transistor. Mortenson determined temperature rises by measuring the decay response of the junction temperature and reported that his results agreed with the calculated results for grown transistors.

Takagi and Mano determined the relationship between the applied power and the runaway time for a device by considering the thermal resistance of the device under the application of power. They reported that the
calculated and measured values are in agreement, and concluded that second breakdown in the open-base condition is triggered by the runaway of collector saturation current. This is in agreement with the view that the temperature of second breakdown should decrease with increasing base resistivity.60

Khurana et al.70 observed that the I-V characteristics in second breakdown lie within the range predicted by the thermal model, and that the calculated thermal sustaining voltage and radius of constriction compared favorably with the experimentally observed values.

**Avalanche Breakdown and Second Breakdown**

Avalanche breakdown in a reverse-biased p-n junction generally occurs at small localized spots known as microplasmas,71,72 but uniform breakdown has also been observed.73 Egawa23 has shown theoretically that an increase in current density at a spot will lower its breakdown voltage. Therefore, increasing the applied voltage will further concentrate the current at weak points. This concentration of current will raise the temperature of the weak spots above the temperature of the remainder of the junction, and if the device is in or near breakdown, this will produce a further current constriction. When the temperature of any spot reaches
the intrinsic or turnover temperature, the junction barrier at that spot vanishes, and the high electric field required to produce avalanche ionization is no longer required. This is a transition from avalanche to "thermal" breakdown, and its occurrence will cause the terminal voltage of the device to be reduced. This reduction in voltage will quench the avalanche breakdown in the remainder of the device, and the entire current will now pinch through this weakest spot. The temperature at this spot might rise sufficiently to produce localized melting.

**Fine Structure in Second Breakdown**

Multiple voltage levels have been observed in transistors, and have been interpreted as second breakdown, third breakdown, etc. Portnoy and Gamble\textsuperscript{74} were the first to report the observation of such multiple voltage levels. Goryunov \textit{et al.}\textsuperscript{75} reported similar observations in Russian transistors.

Portnoy and Gamble\textsuperscript{74} interpreted the presence of fine structure as due to the interaction between different possible second breakdown current constriction sites. This conclusion is supported by the work of Schafft and French\textsuperscript{39} in which it was demonstrated that second breakdown could be made to occur at a number of locations. Morrison and Billette\textsuperscript{35} have also presented
evidence to indicate the existence of discrete breakdown regions.

Although the above results do not negate the thermal theory, it appears that heat alone cannot be the sole cause of second breakdown in transistors, and other phenomena appear to be involved.

Schafft and French found that only the last of the multiple levels exhibited the characteristic polarity reversal of the dependence of collector-emitter voltage upon the base current. Their definition of second breakdown includes this reversal of polarity, hence they conclude that only the last breakdown level is a true second breakdown. They also claim that the apparent dependence of the delay time of the other levels on the current distribution was contrary to that for second breakdown.

Ferry and Dougall found that transistors which exhibited fine structure also showed several melt channels clustered in the area of large damage. English has also reported observing several mesoplasmas in junction regions of devices.

**Beta Dependence of Second Breakdown**

Agatsuma correlated second breakdown with common emitter current gain, beta, or $h_{FE}$. However, beta is a
function of the common base current gain, \( h_{FB} \),
and \( \alpha \) is a complicated function of emitter current density due to current constriction to the emitter periphery\(^79-87\), collector voltage due to the Early effect\(^88\) and temperature\(^89\).

**Miscellaneous Theories of Second Breakdown**

According to a theory by Ebers and Moll\(^90\), the transistor current is increased by 10\% when the junction temperature is increased by 1\(^\circ\)C. This theoretical current dependence on temperature is too weak to explain the large differences in current stability observed empirically\(^12,54\).

Josephs has devised many models to simplify and explain second breakdown\(^91-96\). Many other authors have devised ingenious models to simplify the problem and explain at least one aspect of second breakdown\(^18,22,56,57,59,61,79,86,92,95,97\).

**Miscellaneous Comments on Second Breakdown**

Devices have been reported to be less susceptible to second breakdown at low temperatures, particularly at liquid nitrogen temperatures\(^52\). Also, a distributed resistance, having a positive temperature coefficient, placed over the emitter junction has been observed to enhance thermal stability\(^15,20,43\). Gerstner\(^98\) believes
that strong thermal coupling and a low temperature coefficient of current for the active parts of the transistor are very important design concepts.
PART I

CURRENT CONSTRUCTIONS AND SECOND BREAKDOWN

EMITTER JUNCTION FORWARD BIASED
CHAPTER II

THEORY

The Boltzmann Transport Equation

The Boltzmann transport equation is a mathematical relation which specifies the phase-space distribution function for a particular system of particles.\textsuperscript{101-105} It can be applied to solids, liquids, gases and plasmas; and, as used in the following pages, it applies to holes and electrons in a semiconductor. All the properties of the system under investigation can be determined from the solution to the Boltzmann transport equation, subject to the applicable initial conditions and boundary conditions.

The solution to the Boltzmann equation gives the distribution in velocity and position of all the particles of the system. The equation takes into account the statistical nature of the particles, and incorporates the time variations of the transport properties. From this equation, one can derive the continuity equation (the conservation of particles, mass, and/or charge), the momentum transport equation (conservation of momentum), and the energy transport equation...
Essentially, the Boltzmann transport equation is a "continuity equation" for representative points in phase space. Let $f(\vec{r}, \vec{u}, t)$ be a distribution function in the six-dimensional Cartesian phase space, $x, y, z, u_x, u_y, u_z$, at time $t$, such that the number of particles $dN$ having position between $\vec{r}$ and $\vec{r} + d\vec{r}$ and velocity between $\vec{u}$ and $\vec{u} + d\vec{u}$ is given by:

$$dN = f(\vec{r}, \vec{u}, t) d\vec{r} d\vec{u} = f(\vec{r}, \vec{u}, t) d^3r d^3u,$$

where $d\vec{r} = d^3r = dx dy dz$, and $d\vec{u} = d^3u = du_x du_y du_z$.

The number of particles having a given position and velocity may change with time due to the following processes:

1. The drift of points in position coordinates due to inherent velocity, $-\vec{u} \cdot \text{grad}_r f$.
2. The drift of points in velocity coordinates due to applied forces, $\vec{F}$, generally electric forces and magnetic forces, $-(\vec{F}/m) \cdot \text{grad}_u f$, where $m$ is the mass associated with the points.
3. The spontaneous velocity change without position coordinate change due to non-ionizing collisions, $\partial f/\partial t|_C$.
4. The spontaneous change in the number of particles due to generation and recombination, $\partial f/\partial t|_{gr}$. 
5. The spontaneous change in the velocity and number of particles due to impact ionization collisions, $\partial f/\partial t$.

The first two processes are referred to as "drift processes," the third process is referred to as a "conservative collision process," processes 4 and 5 are "nonconservative processes" which create electrons and holes. Thus, the change in $f$ with time is given by:

$$\frac{\partial f}{\partial t} = \left. \frac{\partial f}{\partial t} \right|_{\text{drift}} + \left. \frac{\partial f}{\partial t} \right|_{\text{conservative}} + \left. \frac{\partial f}{\partial t} \right|_{\text{nonconservative}}.$$ (2)

Upon substitution of the standard expressions for the two drift terms, one arrives at the Boltzmann Transport Equation, viz:

$$\frac{\partial f}{\partial t} + \vec{u} \cdot \vec{v}_f + \frac{\vec{F}}{m^*} \cdot \vec{v}_f = \left. \frac{\partial f}{\partial t} \right|_{\text{conservative}} + \left. \frac{\partial f}{\partial t} \right|_{\text{nonconservative}},$$ (3)

where $\vec{v}_r = \frac{\partial}{\partial x} \vec{i} + \frac{\partial}{\partial y} \vec{j} + \frac{\partial}{\partial z} \vec{k}$, $\vec{v}_u = \frac{\partial}{\partial u_x} \vec{i} + \frac{\partial}{\partial u_y} \vec{j} + \frac{\partial}{\partial u_z} \vec{k}$,

$m^*$ is the effective mass of the electron or hole; and for electrons, $\vec{a} = \vec{F}/m^* = -q/m^*(\vec{E} + \vec{u} \times \vec{B})$, where $q$ is the charge on the electron and $\vec{E}$ and $\vec{B}$ are the electric and magnetic fields acting on the electrons. (The force term for holes is the same except for the sign of the charge.)

There will be one Boltzmann equation for electrons
and one for holes. In general, these two equations are coupled through their collision terms, and generation-recombination terms.

Note that the velocity, $\mathbf{u}$, used above is the total particle velocity, not just the drift velocity. Also, a collision process, and an impact ionization process change velocities only, not particle positions. Gravitational forces and interparticle nuclear forces are neglected in the force term. Gravitational forces are negligible always, and nuclear forces are only significant under high compression. When nuclear forces are significant, quantum mechanics must be used as classical analysis fails. It will be assumed in this paper that classical analysis is sufficient. The classical Boltzmann Transport Equation can be applied to electrons and holes in a solid provided an effective mass is used.$^{107}$ Lastly, the term $\mathbf{u} \cdot \nabla \rho f$ is the normal particle diffusion current in position space.

For a Boltzmann or nondegenerate electron-hole distribution, the thermal equilibrium value of $f$ is given by the Maxwell-Boltzmann distribution (see Appendix D):

$$f = f_o = n(m/2\pi kT)^{3/2} \exp(-mu^2/2kT),$$  \hspace{1cm} (4)

where $n$ is the density of electrons or holes. Under thermal equilibrium, as many particles enter a given
volume, $d\mathbf{d}\mathbf{u}$, as leave this volume. Therefore, the rate of change of $f$ with time due to collisions is zero. In the presence of externally applied disturbances, this is not true. The routine approach is to introduce a collision relaxation time, $\tau_c$, such that:

$$\left. \frac{\partial f}{\partial t} \right|_c = - \frac{(f-f_0)}{\tau_c}, \quad (5)$$

i.e., if

$$\frac{\partial f}{\partial t} = \frac{\partial (f-f_0)}{\partial t} = - \frac{(f-f_0)}{\tau_c},$$

then:

$$(f-f_0)_t = (f-f_0)_{t=0} e^{-t/\tau_c}.$$ 

If the introduction of a relaxation time is not justified, one must investigate the collision term in detail, introducing the transition probabilities for processes which take particles into and out of $d\mathbf{d}\mathbf{u}$. This leads one to an integrodifferential equation.

The general continuity equation, momentum transport equation, and energy transport equation have been derived and the derivations presented in Appendix A. The results are presented below for convenience. (The double vector symbol above a term signifies that it is a tensor of rank 2.) These are:
The Continuity Equation:

$$\frac{\partial n}{\partial t} + \nabla \cdot (n \vec{v}) = g - r + \alpha_i \mu_n n \vec{E} + \beta_i \mu_p P \vec{E}. \quad (6)$$

The Momentum Transport Equation:

$$\frac{\partial (n \vec{v})}{\partial t} + \vec{v} \cdot (n \nabla \vec{v}) + \frac{1}{m} \nabla \cdot \vec{P} + \frac{nq}{m} (\vec{E} + \vec{v} \times \vec{B})$$

$$= \iiint \nabla \left[ \frac{\partial f}{\partial t} \right]_{c} + \frac{\partial f}{\partial t} \bigg|_{gr} + \frac{\partial f}{\partial t} \bigg|_{i-} \right] d^3 \vec{u}. \quad (7)$$

The Energy Transport Equation:

$$\frac{\partial}{\partial t} \left[ 3nkT + mnv^2 \right] + \vec{v} \cdot \left[ (3nkT + mnv^2) \vec{v} \right]$$

$$+ 2v_r \cdot \vec{P} \cdot \vec{v} + 2v_r \cdot \vec{Q} - 2nq (\vec{E} + \vec{v} \times \vec{B}) \cdot \vec{v}$$

$$= m \iiint \left[ \frac{\partial f}{\partial t} \right]_{c} + \frac{\partial f}{\partial t} \bigg|_{gr} + \frac{\partial f}{\partial t} \bigg|_{i-} \right] u^2 d^3 \vec{u}. \quad (8)$$

Due to the complexity of these equations they will be used only for comparison purposes in order to determine what approximations and assumptions are being made in the derivations to follow.

**Ambipolar Diffusion and the Conservation Equations**

**Introduction.** When minority-carrier current is negligible with respect to majority-carrier current,
one can solve either the diffusion equation or the continuity equation for a known electric field in order to obtain the distribution of minority carriers and the minority-carrier current. Once this is done, the majority-carrier distribution and current can be obtained.

In early intrinsic semiconductors or under conditions of high injection leading to large excess-carrier densities, the minority-carrier current is too large to neglect, but the electric field is determined by boundary conditions imposed on the majority carriers. In this case, one must solve two simultaneous nonlinear differential equations, namely the continuity equations of both types of carriers.¹⁰⁹

When both types of charge carriers must be considered, the phenomenon is referred to as "ambipolar diffusion." The term "ambipolar," as used herein, has been borrowed from gas theory, due to the similarity between this type of electron-hole flow and the flow of positive and negative ions in gases.¹¹⁰

Ambipolar diffusion arises due to the mutual attraction between electrons and holes. In the presence of external electric fields, electrons acquire a velocity opposite to that of holes, and hence, the two charge
carriers tend to drift in opposite directions. This gives rise to an excess of electron density in one region of the material, and an excess of holes in another region, thereby producing a space charge.

This space charge sets up a retarding electric field opposing the drift of electrons and holes, and the two charges will tend to diffuse together, in order to preserve space-charge neutrality. In other words, in the presence of electric fields, if there are neighboring regions in which the densities of charges of either sign exhibit a space gradient, there will be a tendency for such a gradient to diminish.\textsuperscript{111}

The ambipolar type of concentration transport is characterized by a diffusivity, an apparent mobility, and a decay time which reduce to the diffusion coefficient, the mobility, and the mean lifetime of minority carriers in the limiting case of small injection level in reasonably extrinsic semiconductor material.\textsuperscript{(110)} The diffusivity and apparent mobility are generally referred to as the "ambipolar diffusion coefficient" and the "ambipolar mobility," respectively.\textsuperscript{109}

Van Roosbroeck\textsuperscript{110,112} has derived the continuity equations for the case of ambipolar diffusion. His derivations are based upon the assumptions of
space-charge neutrality (see Appendix B), a uniformly
doped semiconductor material, equal recombination rates
for electrons and holes, and the assumption that trap­
ping effects arising from the immobilization of charge
carriers in traps for times large compared with their
mean lifetime before their recombination or release can
be neglected. The assumption that electrons and holes
always recombine at equal rates implies a recombination
mechanism via trapping at imperfections as explained by
Shockley and Read.113

The most general form of the continuity equations
for electrons and holes is the form derived from the
Boltzmann transport equation, viz:

\[ \frac{\partial n}{\partial t} + \mathbf{v}_r \cdot (n\mathbf{V}) = g - r + \frac{\partial}{\partial t} \int \int \int f |_{i} d^3u , \]

\[ \text{all } \mathbf{u} \]

with a similar equation for holes. In general the term
on the right due to impact ionization is neglected
(implying low electric fields), current density, \( \mathbf{J} \), is
substituted for \( qn\mathbf{V} \), and the generation-recombination
term, \( g - r \), is replaced by a term of the form \((n/\tau_n - g_0)\),
where \( \tau_n \) is the minority-carrier lifetime, and \( g_0 \) is the
equilibrium generation rate of carriers, which equals
the thermal equilibrium recombination rate.114

The introduction of a minority lifetime, \( \tau \), into
the continuity equation obviously limits it to an equation for minority carriers only, since the implicit assumption is that the deviation of injected carrier density from its equilibrium value will tend to relax with a characteristic lifetime independent of the concentration of the carriers\textsuperscript{110,114,115} (see Appendix C).

The derivation of the ambipolar continuity equations which follows is based upon the work of Van Roosbroeck\textsuperscript{110,112} but deviates from his derivation in the following respects:
1. To avoid loss of generality, the generation-recombination term and the ionization term will be retained in the form obtained from the Boltzmann transport equation.
2. The diffusion coefficients are treated as functions of position and time. The diffusion coefficient for any substance in a solid is an exponential function of temperature,\textsuperscript{116} and the temperature as treated herein is a function of both time and position.
3. The mobilities are treated as functions of position and time, since they are related to the diffusion coefficients through the Einstein relationship.\textsuperscript{117} Note that the derivation of the Einstein relationship is based upon the assumption that Maxwell-Boltzmann statistics apply to the material under discussion.\textsuperscript{117,118,119}
This limits the validity of any results based upon the following derivation to a plasma with a shifted Maxwell-Boltzmann velocity distribution. This is a minor restriction however, since the lack of information as to the correct distribution function to use generally forces the investigator to arbitrarily choose a distribution function, and the Maxwell-Boltzmann function is almost always chosen for its simplicity (see Appendix D).

4. Finally, when a magnetic field is present, the drift current does not usually flow in a direction parallel to the applied electric field, and the diffusion current does not flow parallel to the concentration gradient. Under these conditions, the conductivity, $\sigma$, and the diffusion coefficient, $D$, are tensors of rank 2. They can be derived from the resistivity tensor, $\bar{\rho}$, in this case.\textsuperscript{118,120}

The principle of detailed balancing, the Boltzmann distribution of velocity function, the formulation of the continuity equation and the Einstein relationship are not affected by the presence of a magnetic field, provided the above-mentioned scalar quantities are replaced by Tensor quantities.\textsuperscript{121}

The effective mass is also a tensor. Assuming elliptical energy-momentum surfaces, there are two
components of the effective mass, a transverse and a longitudinal component. The assumption made here is the simplest one, namely spherical energy-momentum surfaces, in which case the effective mass reduces to a scalar.\textsuperscript{122}

**Conservation of Charge.** The conservation of charge in an arbitrary volume, \( V \), enclosed by a surface, \( S \), is obtained as follows: The rate of decrease of charge in the volume \( V \) equals the rate of flow of charge out of the volume through \( S \) plus the rate of recombination in the volume minus the rate of generation in the volume, viz:

\[
- \int \frac{\partial \rho}{\partial t} dV = \int \rho \mathbf{u} \cdot \mathbf{n} da + \int (R - G) dV,
\]

where \( dV \) is a volume element, \( da \) is an area element, \( \mathbf{n} \) is a unit vector in the direction of the outward normal to \( da \), \( \rho \) is the charge density, \( R \) and \( G \) are the rates of recombination and generation of charge in the volume \( dV \), and \( \mathbf{u} \) is the particle velocity.

From the divergence theorem:

\[
\int \rho \mathbf{u} \cdot \mathbf{n} da = \int \nabla \cdot (\rho \mathbf{u}) dV.
\]

Therefore:

\[
\int_\Theta [R - G + \nabla \cdot (\rho \mathbf{u}) + \frac{\partial \rho}{\partial t}] dV = 0.
\]
To be true for an arbitrary volume, the term in the brackets must vanish, and:

$$\frac{\partial \rho}{\partial t} = G - R - \text{div}(\rho \mathbf{u}). \quad (10)$$

For holes, $\rho = q \rho_p$, where $\rho$ is the concentration of holes per unit volume, and $q$ is the electronic charge. From the definition of hole current density, $J_p = q \rho \mathbf{u} = \rho \mathbf{u}$, and $\text{div}(\rho \mathbf{u}) = \text{div} J_p$. Therefore:

$$\frac{\partial \rho}{\partial t} = \frac{1}{q} [G - R - \text{div} J_p].$$

Let: $g' \triangleq G/q$ be the generation rate of holes, and $r \triangleq R/q$ be the recombination rate of holes, and one obtains the general continuity equation for holes.

$$\frac{\partial \rho}{\partial t} = g' - r - \frac{1}{q} \text{div} J_p. \quad (11)$$

One can separate the generation rate, $g'$, into two terms, a thermal generation rate, $g$, and an impact ionization rate $(\alpha_i \mu_n n + \beta_i \mu_p p) \mathcal{E}$, where $\alpha_i$ is the electron ionization coefficient, and $\beta_i$ is the hole ionization coefficient.\textsuperscript{123}

Thus, the continuity equation for holes is:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}_p) = g - r + (\alpha_i \mu_n n + \beta_i \mu_p p) \mathcal{E}, \quad (12)$$

and the continuity equation for electrons is:
\[ \frac{\partial n}{\partial t} + \nabla \cdot (n\vec{v}_n) = g - r + (\alpha_i \mu_n n + \beta_i \mu_p p) \mathcal{E}, \] (13)

where \( \vec{v}_n \) and \( \vec{v}_p \) are the average electron and hole velocities, respectively. These are the same continuity equations as those derived rigorously in Appendix A from the Boltzmann Transport Equation.

Upon adding Equations (12) and (13), one obtains the "ambipolar" continuity equation for electrons and holes, viz:

\[ \frac{\partial (n + p)}{\partial t} + \nabla \cdot (n\vec{v}_n + p\vec{v}_p) = 2(g - r) + 2(\alpha_i \mu_n n + \beta_i \mu_p p) \mathcal{E}. \] (14)

**Conservation of Momentum.** Let \( \vec{u} \) be the individual velocity of a particle, and \( \vec{v} \) be the average velocity of the particles in a volume \( dV \). The total pressure force on the electron "fluid" in a volume \( dV \) is:

\[ d\mathcal{F}_p = -\text{div} \vec{F} \, dV. \]

The Lorentz force on one electron due to an electric field, \( \vec{E} \), and a magnetic field, \( \vec{B} \), is:

\[ \vec{F}_{EB} = -q(\vec{E} + \vec{v} \times \vec{B}). \]

There are \( n \) \( dV \) electrons in the volume \( dV \). Therefore, the total Lorentz force on the electron fluid in \( dV \) is:

\[ d\mathcal{F}_{EB} = -qn \, dV(\vec{E} + \vec{v} \times \vec{B}). \]
The remaining forces acting on the electron fluid are due to the interaction of electrons with the crystal and with holes. The interaction of electrons (and holes) with the crystal lattice is accounted for in the effective mass and mobility terms. Let \( \mathbf{f}_p \) be the force due to the interaction of electrons with holes. From Newton's second law:

\[
\frac{d\mathbf{F}}{dt} = \frac{\partial}{\partial t} (\mathbf{m}_n \mathbf{v}_n) = - \mathbf{v} \cdot \mathbf{F}_n - q_n \mathbf{E} dV + \mathbf{f}_p \mathbf{dV},
\]

Thus the force acting on \( n \) electrons per unit volume is:

\[
\mathbf{F}_n = \mathbf{m}_n \frac{\partial}{\partial t} (n \mathbf{v}_n) = - \mathbf{v} \cdot \mathbf{F}_n - q_n (\mathbf{E}_n + \mathbf{v}_n \times \mathbf{B}) + \mathbf{f}_p \tag{15}
\]

where \( \mathbf{F}_n \) is the electron pressure tensor, \( \mathbf{m}_n \) is the effective mass of electrons, \( \mathbf{v}_n \) is their average velocity, and \( \mathbf{E}_n \) is the electric field acting on the electrons.

Likewise, the force acting on \( p \) holes is:

\[
\mathbf{F}_p = \mathbf{m}_p \frac{\partial}{\partial t} (p \mathbf{v}_p) = - \mathbf{v} \cdot \mathbf{F}_p + q_p (\mathbf{E}_p + \mathbf{v}_p \times \mathbf{B}) + \mathbf{f}_n, \tag{16}
\]

where \( \mathbf{F}_p \) is the hole pressure tensor, \( \mathbf{f}_n \) is the force on the hole fluid due to the interaction with electrons, \( \mathbf{m}_p \) is the effective mass of holes whose average velocity is \( \mathbf{v}_p \), and \( \mathbf{E}_p \) is the electric field acting on the holes.
Upon adding Equations (15) and (16) one obtains the conservation of "ambipolar" momentum, viz:

\[
\overrightarrow{F} = \overrightarrow{F}_n + \overrightarrow{F}_p = - \nabla \cdot (\overrightarrow{F}_n + \overrightarrow{F}_p)
+ q (p \overrightarrow{e}_p - n \overrightarrow{e}_n)
+ q (p \vec{v} \times \vec{B} - n \vec{v} \times \vec{B}),
\]

(17)

where \( \overrightarrow{F}_p = - \overrightarrow{F}_n \) from Newton's third law,\(^{124}\) and the two terms cancel.

**Conservation of Energy.** The conservation of energy can be obtained from the Fourier Heat Flow Equation\(^{125}\) as follows: The rate of heat generation in an arbitrary volume \( V \) equals the rate of heat flow out of \( V \) through the enclosing surface \( S \) plus the rate of increase of heat due to sources in the volume \( V \).

\[
\int_{\Omega} p_v \, dV = \oint_{\partial \Omega} [-K \nabla T] \cdot \hat{n} \, da + \frac{\partial}{\partial t} \int_{\Omega} C \nabla T \, dV,
\]

(18)

where \(-K \nabla T\) is the heat flux density, \( T \) is the absolute temperature, \( C \) is the specific heat, \( K \) is the thermal conductivity, and \( p_v \) is the source of power density of heat in the volume.

From the divergence theorem:

\[
\oint_{\partial \Omega} [-K \nabla T] \cdot \hat{n} \, da = \int_{\Omega} \nabla \cdot [-K \nabla T] \, dV.
\]

Assume the operations of integration and partial with
respect to time are interchangeable, and one obtains:

\[
\int_0^T \left[ \nabla \cdot (K \nabla T) - \frac{\partial (CT)}{\partial t} + p_v \right] \, dV = 0.
\]

To be true for an arbitrary volume, the term in brackets must vanish, and:

\[
\frac{\partial (CT)}{\partial t} = \nabla \cdot (K \nabla T) + p_v.
\] (19)

The Lorentz force on the free charge in the volume dV is:

\[
d\vec{F} = \vec{F}_v \, dV = \rho_f (\vec{E} + \vec{u}_f \times \vec{B}) \, dV,
\]

where \( \rho_f \) is the free charge density. The instantaneous power delivered to the free charge is:

\[
dP = p_v dV = d\vec{F} \cdot \vec{u}_f = \vec{F}_v \cdot \vec{u}_f \, dV,
\]

therefore \( p_v = \rho_f \vec{u}_f \cdot \vec{E} \). But: \( \rho_f \vec{u}_f = \sigma \vec{E} \). Therefore: \( p_v = \sigma \vec{E} \cdot \vec{E} = \frac{1}{2} \sigma \vec{E}^2 \) where \( \vec{E} \) is the axial electric field across the semiconductor material, which would be the base region of a transistor. The conservation of energy expression is thus:

\[
C \frac{\partial T}{\partial t} = \frac{1}{2} \sigma \vec{E}^2 + \nabla \cdot (K \nabla T),
\] (20)

where \( C \) is assumed to be a constant.

**Transient Analysis**

To preserve space-charge neutrality, donor or acceptor atoms must remain neutralized (see Appendix B);
therefore, only the number of electrons and holes equal
to the number of minority carriers can pinch. In the
case of emitter injection, the minority carrier concen-
tration would be determined by the emitter current, and
in the case of avalanche breakdown the minority carrier
concentration would be determined by avalanche multipli-
cation.

In any case, let \( n = p \) be the minority number of car-
riers in the above equations and Equations (14), (17),
and (20) reduce, respectively, to:

The Ambipolar Continuity Equation:

\[
2 \frac{\partial n}{\partial t} + \nabla \cdot [n(\vec{v}_n + \vec{v}_p)] = 2(g - r) + 2n(\alpha_i\mu_n + \beta_i\mu_p)\vec{E}. (21)
\]

The Ambipolar Momentum Transport Equation:

\[
\vec{F} = m \frac{\partial}{\partial t} [n(\vec{v}_n + \vec{v}_p)] = - \nabla \cdot (\frac{\vec{F}_n}{\rho_n} + \frac{\vec{F}_p}{\rho_p})
+ qn(\vec{E}_p - \vec{E}_n) + qn(\vec{v}_p - \vec{v}_n) \times \vec{B}. (22)
\]

The Energy Transport Equation:

\[
C \frac{\partial T}{\partial t} = qn(\mu_n + \mu_p)\vec{E}^2 + \nabla \cdot (KnT). (23)
\]

The simultaneous solution to the above three equa-
tions is desired. Since these equations are still
rather formidable, further simplifications are in order.
In the remainder of this article, cylindrical symmetry will be assumed, and the effect of a carrier source at the emitter junction and a carrier sink at the collector junction will be ignored, as will be the effect of a thermal source at the collector junction.

The assumption of cylindrical symmetry limits the validity of the following equations to an alloyed transistor. Neglecting various sources and sinks is not easily justified, but is mandatory in view of the mathematical complications their consideration would engender. The justification of these assumptions will be determined when theory and experiment are compared.

Base-spreading resistance is also ignored, which is tantamount to assuming an open-base, common-emitter configuration. The impact-ionization term will be dropped, which restricts the remaining discussion to the base region of a transistor. The base-doping profile will be assumed uniform. Then the only position variable is the radial variable $r$.

A Maxwell-Boltzmann velocity distribution will be assumed, in which case the pressure tensors reduce to scalars, that is: $P_n = nkT_n$, and $P_p = pkT_p$, where $k$ is the Boltzmann constant, $T_n$ and $T_p$ are the electron and hole temperatures, respectively.
Finally, the electron and hole temperatures will be assumed equal. Thus, since \( n = p \): \( P_n = P_p = nkT \), where \( T \) is the common electron-hole temperature.

Under the above assumptions, Equations (21) and (22) reduce, respectively, to:

The Ambipolar Continuity Equation:

\[
2 \frac{\partial n}{\partial t} + \nabla \cdot [n(\vec{v}_n + \vec{v}_p)] = 2(g - r). \tag{24}
\]

The Ambipolar Momentum Transport Equation:

\[
\vec{F} = - 2k\nu(nT) + qn(\vec{E}_p - \vec{E}_n) + qn(\vec{V}_p - \vec{V}_n) \times \vec{B}. \tag{25}
\]

Equations (24) and (25) can be combined in the following manner. Under the assumption of space-charge neutrality, the positive and negative charges will pinch radially with the same mobility, which is taken to be the average mobility of the two particles, \( \mu_a \).

That this is consistent with the assumption of space-charge neutrality can be demonstrated by the following reasoning. Since electrons have a higher velocity (mobility) than holes, the \( \vec{J} \times \vec{B} \) force on the electrons is greater than the \( \vec{J} \times \vec{B} \) force on the holes. This causes the electrons to pinch faster than the holes are pinching, and hence a radial charge separation occurs. Since the electric field due to the space
charge is much greater than the magnetic field for non-relativistic electrons and holes (see Appendix E), the electric field will tend to retard the electrons and speed up the holes, forcing them to drift at the same speed in the radial direction.

Therefore:

\[ n\vec{v}_{nr} = \frac{\mu_n}{q} \vec{f}_{nr}, \quad \text{and} \quad n\vec{v}_{pr} = \frac{\mu_p}{q} \vec{f}_{pr}, \]  

(26)

where \( \vec{f}_{nr} \) and \( \vec{f}_{pr} \) are the forces in the radial direction on the n electrons and the n holes, respectively. Add the above expressions and:

\[ n(\vec{v}_{nr} + \vec{v}_{pr}) = \frac{\mu}{q} \vec{F}_r, \]  

(27)

where \( \vec{F}_r \) is the combined radial force \( \vec{F}_{nr} + \vec{F}_{pr} \). Under the assumption of cylindrical symmetry, the pressure term in Equation (25) is a function of the position variable, \( \vec{r} \), and time, \( t \), only. Let \( \vec{E}_p, \vec{E}_n \), and \( \vec{B} \) be radial forces only, and \( \vec{F}_r \) is the force given by Equation (25).

**Magnetic Pinching Force.** Under the assumption of space-charge neutrality, the electrons and holes will pinch radially with the same mobility, and the net current in the radial direction will vanish. Thus, the current density which causes pinching is the longitudinal current density, \( \vec{J} = \sigma \vec{E} \), where \( \vec{E} \) is the
longitudinal electric field across the base region.

To evaluate the magnetic pinching force, consider a cylindrical current shell of radius $r$, and thickness $dr$. Apply Ampère's law and:

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 I = 2\pi r B,$$

where $I$ is the current enclosed by the shell. Now:

$$dI = J \cdot \hat{n} da = J (2\pi r \, dr) = \sigma E (2\pi r \, dr).$$

Therefore,

$$I = 2\pi \sigma \int_0^r \sigma r \, dr, \text{ and:}$$

$$B = \frac{\mu_0 \sigma}{r} \int_0^r \sigma r \, dr. \quad (28)$$

The magnetic force on $n$ electrons and $p = n$ holes is: $qn(\nabla_p - \nabla_n) \times \mathbf{B} = J \times \mathbf{B} = \sigma E \times \mathbf{B}$. Therefore, the radial magnetic force on the $n$ electrons and $n$ holes is:

$$J \times \mathbf{B} = -\frac{\mu_0 \sigma}{r} \mathbf{E} \int_0^r \sigma r \, dr \hat{r}, \quad (29)$$

where $\hat{r}$ is a unit vector in the radial direction.

**Electrical Pinching Force.** Thermal effects give rise to two radial electric fields. The first is due to conductivity modulation. As the pinched region heats up, the temperature increases, the $n$-$p$ product increases, and the Fermi level, $E_F$, approaches the intrinsic value, $E_i$.

Assuming thermal equilibrium exists, the potential
difference seen by electrons and holes can be obtained from Figure 5 as follows:

\[ \Delta E = (E_{cp} - E_{fp}) - (E_{ci} - E_{fi}) = \Delta (E_c - E_f) \]

\[ n = N_c \exp\left[-(E_c - E_f)/kT\right], \]

where \( N_c = \frac{2}{(2\pi mkT/h^2)^{3/2}} \) is the density of states in the conduction band. Therefore,

\[ E_f - E_c = kT \ln(n/N_c), \]

and:

\[ qV = \Delta (E_f - E_c) = kT \ln(n/N_c) - kT_0 \ln(n_0/N_{c0}). \]

The second term is a constant. Therefore,

\[ |\vec{E}| = |qV| = \frac{k}{q} \left( T \ln \frac{n}{N_c} \right). \]

From Figure 5, \( \vec{E} = E \hat{r} \) is pointing radially outward from the hot spot for p-type material. For n-type material the magnitude of \( \vec{E} \) is given by a similar expression and the direction is radially inward. However, in either case, the force on the electrons is opposite to the force on the holes and the resultant force on both cancels. Thus, these terms drop out of the ambipolar equation.

The above forces do produce a radial space-charge which will be ignored in keeping with the assumption of charge neutrality.

The second thermally induced electric field is due
Figure 5.--Variation in energy bands due to localized heating of a p-type material. The electric fields, acting on both electrons, $\vec{\mathcal{E}}_n$, and holes, $\vec{\mathcal{E}}_p$, point to the cooler, less intrinsic region.
to the decrease in the forbidden gap, $E_g$, with increasing temperature. Assuming space-charge neutrality, one-half of the change in $E_g$ will contribute to lowering the conduction band and attracting electrons to the hot spot, and the other half will contribute to raising the valence band and attracting holes to the hot spot (see Figure 6). Thus, the potential difference seen by both electrons and holes due to this effect is: $V_g(T) = \frac{1}{2} \gamma (T - T_0)$ volts, and the magnitude of the radial electric field is $\frac{1}{2} \gamma \nu T$ volts/meter, where $\gamma$ is the thermal coefficient of contraction of the forbidden gap, and $T_0$ is the operating temperature of the ambient base material.

The resultant electrical force in the radial direction on $n$ electron-hole pairs due to thermal effects is:

$$qn(\vec{E}_p - \vec{E}_n) =qn \gamma \nu T.$$  \hspace{1cm} (30)$$

Substitute Equations (29) and (30) into Equation (25), and one obtains for the total force acting in the radial direction on the pinching "fluid":

$$\vec{F}_r = -2k\nu(nT) - \frac{\mu_0 e^2}{r} \int_{o}^{r} \sigma r dr \hat{r} + nq \gamma \nu T.$$  \hspace{1cm} (31)$$

Combine Equations (24), (27), and (31) and one obtains:
Figure 6.--Decrease in the forbidden-gap width due to localized heating of a semiconductor. The electric field acting on electrons, $\vec{\mathcal{E}}_n$, is directed to the cooler region and the electric field acting on holes, $\vec{\mathcal{E}}_p$, is directed to the hot spot.
Upon utilizing the assumed symmetry, one obtains the following equations:

Conservation of Electron-hole Pairs and Momentum:

\[
\frac{\partial n}{\partial t} = \frac{1}{2q} \nabla \cdot \left[ \mu_a \left( 2k \nabla (nT) + \frac{\mu_0 \sigma E^2}{r} \int_0^r \sigma r \, dr \, \hat{r} - nq \, \gamma \nabla T \right) \right] \\
+ (g - r). \tag{32}
\]

Conservation of Energy:

\[
\sigma \frac{\partial T}{\partial t} = \sigma E^2 + \frac{1}{r} \frac{\partial}{\partial r} \left( Kr \frac{\partial T}{\partial r} \right). \tag{34}
\]

Equations (33) and (34) are the general magneto-hydrodynamic equations governing the behavior of electrons and holes in solids, under the influence of the magnetic and thermal pinching forces discussed previously. A solution to these equations would determine the dynamics of pinching in the material, and therefore the dynamics of second breakdown.

An exhaustive attempt to solve the above two equations on the digital computer was made, but every program devised was unsuccessful.
On the other hand, it would be extremely difficult, if not impossible, to obtain experimental data on the transient behavior of any transistors in order to verify the theoretical results. Infrared techniques and equipment make it relatively easy to observe steady-state behavior, however. Therefore, let us determine the steady-state equations, and then proceed to compare the results with experimental data.

**Steady-State Analysis**

Under steady-state conditions, the radial force in Equation (31) and the time derivative in Equation (34) both vanish, and Equations (31) and (34) reduce to:

\[
2k \frac{d(nT)}{dr} + \frac{\mu_o \sigma E^2 q}{r} \int_{0}^{R} \mu n r \, dr - nq \gamma \frac{dT}{dr} = 0 \tag{35}
\]

and

\[
q \mu n E^2 = - \frac{1}{r} \frac{d}{dr} \left( Kr \frac{dT}{dr} \right), \tag{36}
\]

where \( \mu = \mu_n + \mu_p \).

Integrate Equation (36) to obtain:

\[
Kr \frac{dT}{dr} = - qE^2 \int_{0}^{R} \mu n r \, dr. \tag{37}
\]

Substitute Equation (37) into Equation (35) and:

\[
2k \frac{d(nT)}{dr} - \mu_o \sigma K \frac{dT}{dr} - nq \gamma \frac{dT}{dr} = 0. \tag{38}
\]
Divide Equation (38) by $2kT$, multiply by $dr$, and rearrange to obtain:

$$
\frac{d(nT)}{nT} - \frac{q\gamma}{2k} \frac{dT}{T} = \frac{q\mu_0}{2k} \frac{\mu K}{T} \frac{dT}{T} .
$$

(39)

Both $K$ and $\mu$ are functions of temperature, and in order to integrate the above, $K$ and $\mu$ must be expressed as functions of temperature. Let $K = aT^{-b}$, and $\mu = BT^{-d}$. Then Equation (39) becomes:

$$
\frac{d(nT)}{nT} - \frac{q\gamma}{2k} \frac{dT}{T} = \frac{q\mu_0aB}{2k} \frac{dT}{T(b+d+1)} .
$$

(40)

Integrate Equation (40) to obtain:

$$
\ln(nT) - \ln(T^{q\gamma/2k}) = - \frac{q\mu_0aB}{2k(b+d)} \frac{1}{T(b+d)} + \ln n_0
$$

$$
= - \frac{q\mu_0\mu K}{2k(b+d)} + \ln n_0
$$
or:

$$
n = n_0 \frac{T^{q\gamma/2k}}{T} \exp \left[ - \frac{q\mu_0\mu K}{2k(b+d)} \right] ,
$$

(41)

where $n_0$ is a constant proportional to the value of the minority carrier concentration at the initial operating temperature, $T_0$. 
Special Cases. To ignore thermal effects let $\gamma$ be zero and Equation (41) reduces to:

$$n = \frac{n_0}{T} e^{-q\mu_0 K/2k(b+d)}. \quad (42)$$

To ignore magnetic effects let $\mu_0$ be zero and Equation (41) reduces to:

$$n = n_0 T \left( \frac{qY}{2k} - 1 \right). \quad (43)$$

Equation (37) must be solved simultaneously with either Equation (41), (42), or (43).

Power Delivered to the Plasma. The power, $P$, delivered to the pinched plasma is given by the Poynting vector.\(^\text{27}\)

$$P = - \int \left( \vec{E} \times \vec{H} \right) \cdot \hat{n} \, da = \int \vec{E} \hat{r} \cdot \hat{n} \, da. \quad (44)$$

Integrate over the cylinder and:

$$P = \int_0^W \vec{E} \cdot \hat{r} 2\pi r_m \, dz = \vec{E} \cdot \hat{r} 2\pi r_m W,$$

where $r_m$ is the maximum radius of the pinch and $W$ is the base width, or length of the cylinder. The power per unit length, $P_L$, delivered to the plasma is:

$$P_L = 2\pi r_m \vec{E} \cdot \hat{r}.$$  \quad (45)

But $\vec{H} = B/\mu_0$, and $B$ is given by Equation (28) with $r$
replaced by $r_m$. Therefore:

\[ P_L = 2\pi q \epsilon^2 \int_{0}^{r_m} \mu nr \, dr. \tag{46} \]

Equation (46) can also be derived from the power per unit volume, $\sigma \epsilon^2$, as follows:

\[ P_L = \int \sigma \epsilon^2 \, dA = \int_{0}^{r} \int_{0}^{2\pi} \sigma \epsilon^2 \, r \, dr \, d\theta. \]

Therefore, \[ P_L = 2\pi q \epsilon^2 \int_{0}^{r_m} \mu nr \, dr. \]
CHAPTER III

CALCULATIONS

The following results were obtained by using the numerical values listed in Table I:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Silicon</th>
<th>Germanium</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{q \mu_0 \mu K}{2(b+d)k} )</td>
<td>(5.70 \times 10^5 \ T^{-2.83} )</td>
<td>(6.20 \times 10^5 \ T^{-2.76} )</td>
</tr>
<tr>
<td></td>
<td>= ((108/T)^{2.83})</td>
<td>= ((126/T)^{2.76})</td>
</tr>
</tbody>
</table>

\( \frac{q \gamma}{2k} \) | 2.08 | 2.05 |

Calculations were made for both silicon and germanium NPN power units. A typical base resistivity and emitter radius were chosen, and various emitter injection current levels were chosen. The thermally generated minority carrier concentration is negligible for this base-resistivity, and the initial minority carrier concentration at the onset of pinching was taken to be that due to injection, viz:

\[
n = \frac{J}{q \mu_n C}.
\]

Using the above data, Equations (41), (42), and (43) were programmed for the IBM 1620 computer in the form:

\[
n = G \mu T e^{-B T C}.
\]
TABLE I
DATA FOR CALCULATIONS

<table>
<thead>
<tr>
<th>Item</th>
<th>Value for Germanium</th>
<th>Value for Silicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>K = aT⁻ᵇ</td>
<td>4.5x10⁴ (MKS) (128)</td>
<td>3.0x10⁵ (MKS) (128)</td>
</tr>
<tr>
<td></td>
<td>b 1.16</td>
<td>1.33</td>
</tr>
<tr>
<td>μ = BT⁻ᵈ</td>
<td>5.2x10³ (MKS) (129)</td>
<td>7.34x10² (MKS) (130)</td>
</tr>
<tr>
<td></td>
<td>d 1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Y</td>
<td>3.5x10⁻⁴ ev/K⁰ (131)</td>
<td>3.6x10⁻⁴ ev/K⁰ (132)</td>
</tr>
</tbody>
</table>

K = aT⁻ᵇ = thermal conductivity
μ = BT⁻ᵈ = electron plus hole mobility.²
γ = coefficient of thermal contraction of the forbidden gap.

²μ calculated from μ = μ_n (1 + μ_p/μ_n) where it was assumed that μ_n/μ_p is the value at room temperature. These values were taken to be 2.05 for germanium,²²⁹ and 3.0 for silicon.²³⁰
Equation (37) was programmed in the form:

\[
\frac{dT}{dr} = D \frac{T^E}{r} \int_0^r \frac{T-A}{T} e^{BT} r \, dr.
\]  

(48)

Equation (46) was programmed in the form:

\[
P_L = F \int_0^r T-A e^{BT} r \, dr.
\]  

(49)

The computer was programmed to accept a guess for the initial temperature at \( r = 0 \), and calculate a temperature at a distance from the hot spot equal to the emitter radius, where the temperature was assumed to be the normal ambient base temperature. The computer then compared this value with a given ambient temperature (either 300°K or 77°K), and corrected its original guess. The calculation was reiterated until the final temperature was within one degree of the assumed ambient temperature. At this point the computer tabulated the temperature and electron density as a function of distance, for a chosen increment of distance.

The values of the various constants substituted into Equations (47) and (48) are listed in Table II. The corresponding results obtained are shown in Figures 7 through 10.

Three cases were programmed on each run. Case 1
### TABLE II

CONSTANTS USED IN COMPUTER PROGRAM FOR RESULTS SHOWN IN FIGURES 7 THROUGH 10

<table>
<thead>
<tr>
<th>Case</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(-5.17 \times 10^5)</td>
<td></td>
<td>(-2.83 \times 10^6)</td>
<td></td>
<td>(5.89 \times 10^{12})</td>
<td></td>
<td>(1.56 \times 10^{20})</td>
</tr>
<tr>
<td>1</td>
<td>(-0.42)</td>
<td></td>
<td>(-2.83)</td>
<td>(-3.12 \times 10^6)</td>
<td>(1.33)</td>
<td>(5.66 \times 10^{12})</td>
<td>(1.50 \times 10^{20})</td>
<td>(1.08)</td>
</tr>
<tr>
<td>2</td>
<td>(-0.42)</td>
<td></td>
<td>(0)</td>
<td>(-3.00 \times 10^6)</td>
<td>(1.33)</td>
<td>(5.66 \times 10^{12})</td>
<td>(1.50 \times 10^{20})</td>
<td>(1.08)</td>
</tr>
<tr>
<td>3</td>
<td>(-2.50)</td>
<td>(-5.17 \times 10^5)</td>
<td>(-2.83)</td>
<td>(-4.40 \times 10^{11})</td>
<td>(1.33)</td>
<td>(8.30 \times 10^{17})</td>
<td>(2.20 \times 10^{25})</td>
<td>(-1.00)</td>
</tr>
</tbody>
</table>

Constants for Figure 7, silicon unit at an ambient temperature of 300\(^0\)K

<table>
<thead>
<tr>
<th>Case</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(-0.52)</td>
<td>(-6.20 \times 10^5)</td>
<td>(-2.76)</td>
<td>(-1.70 \times 10^7)</td>
<td>(1.16)</td>
<td>(4.82 \times 10^{12})</td>
<td>(1.64 \times 10^{20})</td>
<td>(1.05)</td>
</tr>
<tr>
<td>1</td>
<td>(-0.52)</td>
<td>(-6.20 \times 10^5)</td>
<td>(0)</td>
<td>(-1.56 \times 10^7)</td>
<td>(1.16)</td>
<td>(4.42 \times 10^{12})</td>
<td>(1.50 \times 10^{20})</td>
<td>(1.05)</td>
</tr>
<tr>
<td>2</td>
<td>(-2.60)</td>
<td>(-6.20 \times 10^5)</td>
<td>(-2.76)</td>
<td>(-2.42 \times 10^{12})</td>
<td>(1.16)</td>
<td>(6.85 \times 10^{17})</td>
<td>(2.32 \times 10^{25})</td>
<td>(-1.00)</td>
</tr>
</tbody>
</table>

Constants for Figure 8, germanium unit at an ambient temperature of 300\(^0\)K

<table>
<thead>
<tr>
<th>Case</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(-0.42)</td>
<td>(-5.17 \times 10^5)</td>
<td>(-2.83)</td>
<td>(-2.62 \times 10^6)</td>
<td>(1.33)</td>
<td>(4.92 \times 10^{12})</td>
<td>(6.54 \times 10^{20})</td>
<td>(1.08)</td>
</tr>
<tr>
<td>1</td>
<td>(-0.42)</td>
<td>(-5.17 \times 10^5)</td>
<td>(0)</td>
<td>(-2.59 \times 10^5)</td>
<td>(1.33)</td>
<td>(4.88 \times 10^{11})</td>
<td>(6.47 \times 10^{19})</td>
<td>(1.08)</td>
</tr>
<tr>
<td>2</td>
<td>(-2.50)</td>
<td>(-5.17 \times 10^5)</td>
<td>(-2.83)</td>
<td>(-2.23 \times 10^{10})</td>
<td>(1.33)</td>
<td>(4.20 \times 10^{16})</td>
<td>(5.57 \times 10^{24})</td>
<td>(-1.00)</td>
</tr>
</tbody>
</table>

Constants for Figure 9, silicon unit at an ambient temperature of 77\(^0\)K

<table>
<thead>
<tr>
<th>Case</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(-0.52)</td>
<td>(-6.20 \times 10^5)</td>
<td>(-2.76)</td>
<td>(-1.50 \times 10^7)</td>
<td>(1.16)</td>
<td>(4.23 \times 10^{13})</td>
<td>(2.87 \times 10^{20})</td>
<td>(+1.05)</td>
</tr>
<tr>
<td>1</td>
<td>(-0.52)</td>
<td>(-6.20 \times 10^5)</td>
<td>(0)</td>
<td>(-3.37 \times 10^6)</td>
<td>(1.16)</td>
<td>(9.60 \times 10^{11})</td>
<td>(6.47 \times 10^{18})</td>
<td>(+1.05)</td>
</tr>
<tr>
<td>2</td>
<td>(-2.60)</td>
<td>(-6.20 \times 10^5)</td>
<td>(-2.76)</td>
<td>(-1.26 \times 10^{12})</td>
<td>(1.16)</td>
<td>(3.59 \times 10^{17})</td>
<td>(2.43 \times 10^{24})</td>
<td>(-1.00)</td>
</tr>
</tbody>
</table>

Constants for Figure 10, germanium unit at an ambient temperature of 77\(^0\)K
refers to the combined thermal plus magnetic pinch, using Equation (41), which is the only one physically realizable. Case 2 refers to thermal pinching only, using Equation (43). Case 3 refers to magnetic pinching only, using Equation (42).

Physically, the magnetic and thermal terms are not separable. However, mathematically they are, and by comparing the results of case 2 and case 3, one can see the relative importance of each in producing current constrictions and hot spots.
CHAPTER IV

RESULTS

Theoretical Results

Some typical results for a silicon power unit operating at an ambient temperature of 300°K are shown in Figure 7. It is seen that at room temperature thermal pinching produces almost as much rise in temperature as the combined thermal plus magnetic pinching, while the magnetic pinching alone contributes much less to increasing the temperature or current constriction. Perhaps even more significant is the fact that a magnetic pinch acting alone produces sufficient rise in temperature to cause the thermal pressure, nkT, to exceed the magnetic pressure and result in an "anti-pinch" (i.e., when the thermal pressure nkT exceeds the magnetic pressure, the electron concentration actually becomes lower in the hot spot).

Other computer runs show that increasing the power supplied to the base of the transistor, by increasing the emitter injection current through the base region, enhances each effect. The temperature rises in all three cases, the thermal force becomes more dominant
Figure 7.—Silicon power transistor operated at 300°K with high input power. A: Temperature versus relative distance from center of hot spot, and B: electron concentration versus relative distance from center of hot spot.
Combined Effect

Thermal Pinch Only

Magnetic Pinch Only

Electron Concentration ($10^{22} m^{-3}$)

Relative Distance
as the temperature rises, while the "anti-pinch" becomes more pronounced in the case of magnetic pinching only. The pinch radius is of the same order of magnitude as the emitter radius.

Figure 8 shows the results for a germanium unit having the same base resistivity, emitter radius, base width, etc., as the silicon unit of Figure 7 and operating under the same input power to the base. The higher mobility of germanium enhances the magnetic pinch effect slightly over that of the silicon unit.

Figure 9 is for a silicon unit operated at a lower power level at an ambient temperature of 77°K. At this low temperature the magnetic effect is more important than the thermal effect, and both effects pinch the electron-hole concentration.

Figure 10 is for a germanium unit operated at 77°K with an input power five times that of the silicon unit of Figure 9. At this power level and ambient temperature, the temperature curve for the magnetic pinch only, crosses the curve for both effects combined. Therefore, the "hot spot" is more localized due to the presence of both effects. The electron-hole concentration for magnetic pinching only shows a hollow effect similar to that reported by Drummond and Ancker-Johnson for InSb.
Figure 8.—Germanium power transistor of construction identical to the silicon unit in Figure 7, operated at 300°K with identical input power. A: Temperature versus relative distance from center of hot spot, and B: electron concentration versus relative distance from center of hot spot.
Figure 9.--Silicon unit of Figure 7 operated at 77°C with greatly reduced input power. A: Temperature versus relative distance from center of hot spot, and B: electron concentration versus relative distance from center of hot spot.
Figure 10.--Germanium unit of Figure 8 operated at 77°K with an input power five times that of the silicon unit in Figure 9. A: Temperature versus relative distance from center of hot spot, and B: electron concentration versus relative distance from center of hot spot.
at 77°K. Note that at room temperature where thermal effects are predominant, the silicon unit is slightly more susceptible than the germanium unit to hot-spot formation for a given input power.

At 77°K, where magnetic effects dominate, the germanium unit showed an almost negligible thermal rise when operated with an input power equal to that of the identical silicon unit of Figure 9. In fact, from Figure 10 it is evident that the germanium unit can handle almost five times the input power of an identical silicon unit for the same amount of thermal rise at the hot spot.

Thus, in both cases, the silicon unit is more susceptible to hot-spot formation than an identical germanium unit for a given input power.

It should be noted that the equations programmed are extremely nonlinear and the results shown in Figures 7 through 10 reflect this nonlinearity.

Experimental Results

Experimental data on hot-spot temperatures were obtained for some 2N3599 devices. These data were taken with an infrared scanner which reads an average IR flux
over an area of five square mils. The data were taken at 10-mil intervals in four directions from the center of the hot spot, and were obtained from devices operating at low emitter current levels, where base spreading resistance is not important.

Figure 11 shows the variation of temperature with distance in two opposite directions for a typical hot spot, having a maximum temperature of 212°C, an ambient temperature of 70°C, and a hot-spot radius of about 60 mils (about 1.5 mm). The unit was operating at $V_{CB} = 40$ volts and $I_C = 2.5$ amperes. It can be seen from Figure 11 that the hot spot is reasonably symmetrical, since the data points in both directions are in close agreement.

Not all hot spots are so symmetrical. Figure 12 shows two hot spots located in close proximity to each other. For comparison purposes, the temperature profile in the opposite direction is also shown. The ambient operating temperature is 75°C. These hot spots are obviously due to defects in the base or collector junction.

A computer run was made for the symmetrical hot spot shown in Figure 11. By choosing 60 mils for the hot-spot radius and 70°C for the ambient temperature,
Figure 11.--Temperature versus distance for a typical hot spot. Two diametrically opposite directions are shown to point out the symmetry of the hot spot.
Figure 12.--Temperature versus distance for two hot spots in close proximity. The temperature profile in the opposite direction from the major hot spot is also shown, and is seen to be similar to the curves in Figure 11.
one obtains a curve which matches the empirical points at both ends of the curve only. However, by adjusting the radius of the computed curve, very good agreement with theory was obtained over the first 60% of the total radius.

Figure 13 shows the experimentally obtained points (taken in all four directions from the center of the hot spot), the empirical curve drawn through the average value of the four sets of points, and the computed curve normalized to fit. The "theoretical" radius of the hot spot appears to be about 54 mils or about 1.37 mm.

The deviation from theory at the edge of the hot spot is easily explained. In the theory, it was assumed that the current went to zero at the edge of the hot spot. There were emitter leads on both sides of the hot spot measured, and hence the current to the hot spot was flowing radially through the emitter. Thus, some heating of the device occurred outside the hot-spot region, and the radius of the hot spot was not well defined.

As a further check on the theory, three other hot spots were examined, and the temperatures averaged over the four directions. The hot-spot temperatures were all similar, the maxima being 210°C, 182°C, and 177°C for the three hot spots. These data were normalized with
Figure 13.--Comparison of theory and data for a typical hot spot. Experimental points are shown for all four directions from the center of the hot spot. The empirical curve is drawn through the average values of these points.
respect to maximum temperature and with respect to hot-spot radius, and then compared with the previous computer run, also normalized. The results are shown in Figure 14. Again, it is seen that the empirical data and the theoretical curve agree over about 60% of the hot-spot radius.
Figure 14.--Normalized plot of data for three hot spots of similar radius and maximum temperature. The theoretical curve of Figure 13 is also normalized and shown for comparison purposes.
Theoretical Curve
\(T_{\text{max}} = 212^\circ C\)

<table>
<thead>
<tr>
<th>Hot spot</th>
<th>(T_{\text{max}})</th>
<th>(r_{\text{max}})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>210^\circ C</td>
<td>50 mils</td>
</tr>
<tr>
<td></td>
<td>182^\circ C</td>
<td>41 mils</td>
</tr>
<tr>
<td></td>
<td>177^\circ C</td>
<td>38 mils</td>
</tr>
</tbody>
</table>
CHAPTER V

CONCLUSIONS

The results indicate that current constriction (pinching) will occur at any ambient temperature due to thermal and magnetic effects inherent in the device operation, as shown in Figures 7 through 10. Note that no nonuniformities in initial base resistivity are necessary to cause pinching. In the absence of inhomogeneities in base resistivity, the pinching will tend to occur near the center of the emitter. The pinching produces a significant temperature rise at a "hot spot" only at high input power levels. This is obvious from the curves shown in Figures 7 through 10, which were all obtained at relatively high input power levels.

At, or near, room temperature the electric field produced by Joule heating is the principal cause of the hot-spot formation, as shown in Figures 7 and 8; while at temperatures near 77°K the magnetic field associated with the carriers is the principal cause of hot spots, as shown in Figures 9 and 10.
As an indication of the temperature at which the transition occurs from second breakdown due primarily to Joule heating to second breakdown due primarily to magnetic pinching, define the transition temperature as that temperature at which the power of the exponential term in Equation (48) is unity. From Table I, one obtains 108°K and 126°K as the transition temperatures for silicon and germanium, respectively. The difference in the two temperatures is due to the higher mobilities of carriers in germanium, which leads to larger magnetic fields in germanium than in silicon at any temperature.

Base spreading resistance was ignored in the above calculations, which is tantamount to assuming an open-base common-emitter configuration. It is well known that under forward-biased base drive conditions, the emitter current tends to crowd to the periphery of the emitter. This is an "anti-pinch" effect which will oppose current constriction to the center of the device, and hence more input power will be required to obtain a given current density and a given temperature at the hot spot. That forward-biased base drive does minimize current constrictions has been reported.¹³⁶ The result is that as the emitter current increases due to increased forward bias of the emitter-base junction, second
breakdown will be entered at higher collector current levels.

If the device is operated in the reverse-biased emitter junction mode, it has to be driven into avalanche breakdown in order to produce sufficient minority carriers to cause appreciable pinching. Microplasmas generally control the breakdown and current is already being funneled through the small constrictions.\textsuperscript{137-139} The pinch radius should be determined by the microplasma in this case, and at the onset of second breakdown, the hot spot should have a radius of the order of magnitude of a microplasma radius, which is about 500Å.\textsuperscript{140} Hot spots as small as 10^{-3} cm have been observed and reported.\textsuperscript{141} In this case, second breakdown would be entered at lower collector current since the current density is the determining factor of the hot-spot temperature.

The above reasoning would explain the dependence of second breakdown upon the base drive of the device as reported by Schafft and French.\textsuperscript{134,136}

The above calculations show that even a perfect device of a given geometry will exhibit current pinching effects at some calculable current level. Second breakdown is usually associated with defects in the base
These defects cause the pinching to occur at a fraction of the theoretically computed value. This fraction could be utilized as a quality measure, Q, for a particular production unit, thus specifying the ratio of its actual open-base second breakdown to its theoretical capability. The theoretical calculation would have to be based on the assumption of an "equivalent" circular emitter area.

The approach taken by the author in the derivation of the theory and the results obtained from the theory are unique in the following respects:
1. This is the first attempt to explain second breakdown in terms of magnetohydrodynamics.
2. This is the first time an explanation of pinching in a solid at room temperature has been presented which does not rely on some type of crystal imperfection to initiate the pinching.
3. This is the first time variations in the forbidden gap width have been considered as causing voltage drops and, therefore, electric fields in a semiconductor.
4. This is the first theory of second breakdown which predicts that units are much more susceptible to failure at lower temperatures than at higher temperatures, contrary to popular belief.

In fact, item 4 above was so startling that it was
felt necessary to explore the possibilities in detail. The outcome of this investigation was the determination of the critical current at which second breakdown is initiated. The theory and results of this investigation are presented in Part II.
PART II

CRITICAL CURRENT TO INITIATE SECOND BREAKDOWN

EMITTER JUNCTION REVERSE BIASED
CHAPTER VI

THEORY

Critical Current for Pinching

Bennett\textsuperscript{143} derived the equation for the critical current to initiate pinching, under the assumption that the magnetic pressure must balance the thermal pressure for pinching to start.

If there is an electric field, $\vec{E}$, as well as a magnetic field, $\vec{B}$, acting to pinch the plasma, then the total pressure (energy density) on the plasma due to the two fields is:\textsuperscript{144}

$$\frac{1}{2} \left[ \varepsilon \langle \vec{E}^2 \rangle + \mu_0 \langle \vec{H}^2 \rangle \right], \quad (50)$$

where $\varepsilon$ is the permittivity of the material, $\mu_0$ is the permeability of the material, $\vec{E}$ and $\vec{H}$ are the electric and magnetic fields acting to pinch the plasma, and $\langle \rangle$ denotes the average value.

The magnetic field, $B$, is obtained by applying Ampère's law to a cylinder of radius $r$, enclosing a current, $I$:

$$B = \mu_0 \frac{I}{2\pi r}. \quad (51)$$
The radial electric field acting on the plasma due to narrowing of the forbidden gap with increasing temperature is: \(^{145}\)

\[ E_r = \gamma \nu T, \tag{52} \]

where \( \gamma \) is the coefficient of thermal contraction of the forbidden gap, and \( T \) is the temperature of the material. To determine \( \nabla T \), consider the conservation of energy as given by the Fourier heat-flow equation. \(^{125}\)

\[ \frac{\partial (CT)}{\partial t} = \nabla \cdot (K \nu T) + \sigma \vec{E}^2, \tag{53} \]

where \( C \) is the specific heat, \( K \) is the thermal conductivity, \( \sigma = q \mu n \), is the conductivity, \( q \) is the electronic charge, \( \mu \) is the sum of the electron and hole mobilities, \( n \) is the number of electrons or holes that are free to pinch, and \( \vec{E} \) is the longitudinal electric field across the pinching material.

Assume that prior to the onset of pinching, a steady-state operating temperature has been reached. Then:

\[ \frac{1}{r} \frac{d}{dr} (rK \nu T) = - \sigma \vec{E}^2 \]

\[ \nu T = - \frac{q \vec{E}^2}{rK} \int_0^r \mu n r \, dr. \tag{54} \]

From the definition of mobility, \( \mu = v/\vec{E} \), where \( v \) is the sum of the electron and hole velocities. Therefore:
\[ \nabla T = - \frac{qE_r}{rK} \int_0^r vnr \, dr. \quad (55) \]

The current through the cylinder of radius \( r \) is:

\[ I = 2\pi q \int_0^r vnr \, dr. \quad (56) \]

Substitute Equation (56) into Equation (55), and:

\[ \nabla T = - \frac{I E_r}{2\pi rK}, \quad (57) \]
\[ E_r = - \frac{vI E_r}{2\pi rK}. \quad (58) \]

Equate the Lorentz pressure of Equation (50) to the thermal pressure, and:

\[ B^2 \frac{\mathcal{E}_r^2}{2\mu_0} + \frac{\varepsilon \mathcal{E}_r^2 \gamma^2}{2} = nkT, \quad (59) \]

where \( k \) is Boltzmann's constant, and \( T \) is the sum of the electron and hole temperatures. Substitute the value of \( B \) from Equation (51), and the value of \( \mathcal{E}_r \) from Equation (58), and:

\[ I^2 = \frac{8\pi^2 r^2 K^2 nkT}{\mu_0 K^2 + \varepsilon \gamma^2 \mathcal{E}_r^2}. \quad (60) \]

Prior to the onset of pinching, the current is uniform through the material, and Equation (56) reduces to:

\[ I = \pi q nvr^2. \quad (61) \]

Divide both sides of Equation (60) by Equation (61) and
the critical current for pinching is given by:

\[
I_c = \frac{8\pi k T}{q \nu_0} \frac{1}{1 + \frac{e \gamma^2 \xi^2}{\mu_0 K^2}}.
\]  

(62)

Note that the first quantity on the right in Equation (62) is the Bennett critical current for pinching.\(^\text{143}\)

The presence of an electric field has added a correction term to the Bennett current, given by:

\[
R = \frac{e \gamma^2 \xi^2}{\mu_0 K^2}.
\]  

(63)

The data from Table III were substituted into Equation (63) with the following results at room temperature:

For germanium: \( R = 1.9 \times 10^{-14} \frac{\xi^2}{K^2} \)  

(64)

For silicon: \( R = 4.7 \times 10^{-16} \frac{\xi^2}{K^2} \).

For both silicon and germanium the correction term is very small for normal values of the electric field, \( \xi^2 \). Thus, in the base region of a transistor, \( R \) is negligible, and magnetic effects are responsible for initiating the pinch. In the depletion region of an avalanching junction the electric field is very large, and \( R \) may be larger than unity, in which case electrical effects are predominant in initiating the pinch. In this case the critical current to pinch is much lower than predicted by Bennett.

One could define a critical field for pinching as
<table>
<thead>
<tr>
<th>Item</th>
<th>Germanium</th>
<th>(Ref)</th>
<th>Silicon</th>
<th>(Ref)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_e$</td>
<td>16</td>
<td>(146)</td>
<td>12</td>
<td>(146)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$3.5\times10^{-4}$ ev/K°</td>
<td>(147)</td>
<td>$3.6\times10^{-4}$ ev/K°</td>
<td>(148)</td>
</tr>
<tr>
<td>$K(300^\circ K)$</td>
<td>24 $\text{watts \over mK}^0$</td>
<td>(149)</td>
<td>152 $\text{watts \over mK}^0$</td>
<td>(149)</td>
</tr>
<tr>
<td>$K=aT^{-b}$ $a$</td>
<td>$4.5\times10^4$ (MKS)</td>
<td>(149)</td>
<td>$3.0\times10^5$</td>
<td>(149)</td>
</tr>
<tr>
<td>$b$</td>
<td>1.16</td>
<td>(149)</td>
<td>1.33</td>
<td>(149)</td>
</tr>
</tbody>
</table>

$k_e$ is the dielectric constant ($\varepsilon = k_e\varepsilon_0$).

$\gamma$ is the coefficient of thermal contraction of the forbidden gap.

$K$ is the thermal conductivity.
that value of $\xi$ which makes $R$ equal to unity. From Equation (64), the critical fields, at room temperature are:

For germanium: $\xi_{GC} = 7.26 \times 10^6$ volts/m.

For silicon: $\xi_{RC} = 4.62 \times 10^7$ volts/m. \hspace{1cm} (65)

In the derivation of Equation (52) for the radial electric field, base-spreading resistance was neglected. Base-spreading resistance gives rise to an electric field in the radial direction also. Under reverse bias, this field would tend to pinch the plasma also.

Thus the critical current at which second breakdown is initiated will be influenced by the base current also. Since the base current is determined by $V_{BE}$ and $R_{BE}$, the onset of second breakdown will depend on these parameters. This dependence for several types of devices has been determined by Schiff.\textsuperscript{150}

Critical Current as a Function of Temperature and Field

$R$ may be expressed as a function of electric field and temperature in the following manner. Substitute data from Table I, and one obtains for silicon:

$$R = 1.06 \times 10^{-11} \xi^2 / K^2,$$ \hspace{1cm} (66)

where $K$ is given as a function of temperature in Table I.
Under reverse-biased emitter conditions, $\mathcal{E}_A$ is the electric field in the depletion region under avalanche conditions, which is in the range of 200 kv/cm to 500 kv/cm.\textsuperscript{151} However, breakdown voltage, and hence the critical electric field for breakdown, varies with temperature also. McKay\textsuperscript{151} has measured this variation and found that:

$$\mathcal{E}_A(T) = \mathcal{E}_A(T_0) [1 + 4.4 \times 10^{-4}(T - T_0)].$$ \hspace{1cm} (67)

Substitute $K(T)$ from Table I and $\mathcal{E}_2(T)$ from Equation (67) into Equation (66) and one obtains $R$ as a function of $T$. The thermal variation of $R/\mathcal{E}_2^2$ is shown in Figure 15. Note that $\mathcal{E}_A$ in Figure 15 is the value of field for avalanche at room temperature.

The Bennett current is a function of electron and hole temperatures and velocities. It was empirically determined that the duty cycle of the second breakdown test set (see Appendix F) is low enough to neglect heating of the device under test. Therefore, the electron and hole temperatures at the onset of breakdown under test conditions may be assumed to be related to the ambient temperature of the device under test. Since the exact electron and hole temperatures are not known, and since the electron and hole velocities depend upon the doping concentration of the device,\textsuperscript{152} the Bennett critical current to pinch will be taken to be 100% at room
Figure 15. -- $R/\xi^2$ versus temperature for silicon.
temperature.

It is also known that electron and hole velocities in silicon and germanium saturate at electric fields an order of magnitude lower than breakdown fields.\textsuperscript{153-157} The temperature dependence of electron saturation velocity in germanium is given by Ryder\textsuperscript{153} and Gunn,\textsuperscript{154} and the temperature variation of electron saturation velocity in silicon is given by Duh and Moll.\textsuperscript{157}

Using the data for germanium as given by Ryder\textsuperscript{153} and Gunn,\textsuperscript{154} one obtains the curve for the Bennett critical current to pinch, \( I_{\text{CB}} \), as a function of temperature shown in Figure 16. The critical field for avalanche breakdown is between about 200 kv/cm and 500 kv/cm in germanium.\textsuperscript{158} From the values of \( R/E^2 \) versus temperature shown in Figure 15, one can calculate the correction to \( I_{\text{CB}} \) for a given avalanche electric field. This has been done and is also shown in Figure 16.

Using the data for silicon as given by Duh and Moll,\textsuperscript{157} one obtains the curve for the Bennett critical current to pinch in silicon as a function of temperature as shown in Figure 17. The critical field for avalanche breakdown is between about 200 kv/cm and 500 kv/cm in silicon also.\textsuperscript{151} The correction to the Bennett critical current for silicon is shown in Figure 17.
Figure 16.--The critical current to initiate pinching versus temperature for $E_L$ negligible, $E_L = 2 \times 10^7$ v/m and $E_L = 5 \times 10^7$ v/m. $I_{GB}$ calculated assuming electron and hole limiting velocities vary with temperature as electrons in germanium.
Figure 17.--The critical current to initiate pinching versus temperature for $\xi_\lambda$ negligible, $\xi_\lambda = 2 \times 10^7$ v/m and $\xi_\lambda = 5 \times 10^7$ v/m. IGB calculated assuming electron and hole limiting velocities vary with temperature as electrons in silicon.
Data on the critical current to initiate second breakdown over the temperature range of 300°K to 400°K have been published by Schiff. Schiff's data for the TA2110 devices and the 2N3265 devices tested are shown in Figure 18, along with the theoretical curves for silicon for $I_{CB}$ and for $I_C$ with $\mathcal{E}_2 = 10^7$ v/m. The agreement is very good, indicating that the electric field at the onset of second breakdown is about 100 kv/cm for these devices, and also that the limiting velocities vary with temperature in about the same manner as the electron velocities vary in silicon.

It should be noted also that Ancker-Johnson et al. observed that, in InSb at 77°K, the critical current at which pinching occurs agreed with Bennett's theory, and that the pinch times were consistent with Bennett's theory.  

An upper limit to this theory is determined by the temperature at which the device becomes intrinsic (see Appendix G).
Figure 18.--Comparison of Schiff's data with the theoretical curve for ICB for silicon. Data and theoretical curve adjusted to 100% at 300°K.
RESULTS

To verify the theoretical variation of critical current with temperature, it was decided to measure the second-breakdown current of some 2N3599 devices at various known temperatures. For this purpose, an insulated container was made in which liquid could be contained, with the transistor and its heat sink immersed in the liquid.

Three known temperatures below room temperature were chosen for the ease in experimentally obtaining them, and for their spacing on the thermal scale. These temperatures were:

1. 0°C or 273°K, obtained by immersing the heat sink of the device under test in a mixture of crushed ice and water.

2. -78°C or 195°K, obtained by immersing the heat sink of the device under test in a mixture of crushed dry ice (CO$_2$) and acetone (CH$_3$COCH$_3$).

3. -196°C or 77°K, obtained by immersing the heat sink of the device under test in liquid nitrogen.
The data were obtained under the following conditions: $L = 40$ microhenrys, $V_{BE} = -6$ volts, $I_B = 2$ amperes, and $R_{BE} = 10$ ohms.

All the devices tested which were susceptible to second breakdown (the test set is limited to a maximum collector current of 20 amperes) showed a trend to lower $I_Q$ with lower temperature. The results of four 2N3599 devices are listed in Table IV, and the normalized average values for the four devices as a function of temperature are shown in Figure 19, along with the theoretical curve for silicon.

Through repeated testing, the units began to show deterioration and entered second breakdown at lower values of $I_Q$. It generally required about two hundred tests before deterioration was appreciable.

Unit number 32 was severely damaged in testing and its room-temperature second-breakdown current changed from 9.5 A to 3.9 A. This would indicate that permanent damage was done to the unit. If so, second breakdown should now be controlled by a defect and should be independent of temperature. Measurements were carried out and these results are also shown in Table IV, for unit 32 after being damaged.

Units which exhibited low $I_Q$ (as compared to other
## TABLE IV

DATA ON FOUR 2N3599 DEVICES

<table>
<thead>
<tr>
<th>Unit number</th>
<th>Collector current (amps) at ambient temperature of</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300°K</td>
</tr>
<tr>
<td>9</td>
<td>7.0</td>
</tr>
<tr>
<td>11</td>
<td>4.6</td>
</tr>
<tr>
<td>23</td>
<td>6.5</td>
</tr>
<tr>
<td>32</td>
<td>9.5</td>
</tr>
<tr>
<td>32**</td>
<td>3.9</td>
</tr>
<tr>
<td>Total</td>
<td>27.6</td>
</tr>
<tr>
<td>Total %</td>
<td>100</td>
</tr>
</tbody>
</table>

*Unit destroyed at onset of test.

**After being severely damaged.
Figure 19. Critical current versus temperature for 2N3599 devices. The solid line is the theoretical curve for silicon. The points represent the average values of four devices.
devices of the same family) when first tested, also showed a behavior similar to the damaged unit 32. Again this is probably due to defects controlling the onset of second breakdown, instead of magnetic pinching controlling it.

The damaged devices were examined on the Tektronix 575 Transistor Curve Tracer. Slightly damaged units exhibited betas that were lower than those measured prior to the unit's being damaged, or erratic base control as exhibited by poor I-V characteristics. Severely damaged units showed an almost complete loss of base control. They were almost short-circuited from emitter to collector, with the low resistance slightly dependent upon base drive. The most severely damaged units were short-circuited from emitter to collector. All damaged units showed good diode characteristics at both the emitter-base junction and at the collector-base junction.

To further substantiate the theoretical dependence of critical collector current for second breakdown, the preceding experiments were repeated on eight 2N4128 devices. The data taken on these devices were obtained under the following conditions: \( L = 40 \) microhenrys, \( V_{BE} = -3 \) volts, \( I_B = 2 \) amperes, and \( R_{BE} = 50 \) ohms and 100 ohms.
One of the added precautions taken in repeating the experiments was to repeat the current reading at room temperature prior to each additional low-temperature reading. It was found that, while a given reading is completely reproduceable at any given time, readings vary slightly from day to day for a given device. The reason for this is unknown at present. However, it was found through repeated testing that the percentage change in critical current from room temperature to any given lower temperature remained almost constant. For the above reason, the results shown in Table V for the eight 2N4128 devices show both the critical current at room temperature prior to cooling the device, and the low-temperature critical current of the device.

The average relative change in $I_C$ versus temperature is shown in Figure 20, as well as the theoretical curve. As can be seen from an inspection of Table V, increasing $R_{BE}$ generally increases the critical current at a given temperature. To determine the cause of this, the collector-to-emitter voltage was monitored for all eight devices. In all cases it was found that increasing $R_{BE}$ decreases the collector voltage pulse height and simultaneously increases the pulse width. It was therefore concluded that changing $R_{BE}$ changes the switching time of the device, since $V_{CE} = L I_C/\tau$, and increasing $\tau$
### Table V

**Data on Eight 2N4128 Devices**

<table>
<thead>
<tr>
<th>Unit number</th>
<th>$R_{BE}$</th>
<th>$300^\circ K$</th>
<th>$273^\circ K$</th>
<th>$300^\circ K$</th>
<th>$195^\circ K$</th>
<th>$300^\circ K$</th>
<th>$77^\circ K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>5.4</td>
<td>5.4</td>
<td>5.5</td>
<td>3.0</td>
<td>5.4</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>3.7</td>
<td>6.0</td>
<td>1.8</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>3.4</td>
<td>3.0</td>
<td>3.4</td>
<td>1.8</td>
<td>3.3</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4.0</td>
<td>3.6</td>
<td>4.1</td>
<td>2.3</td>
<td>3.6</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>3.1</td>
<td>2.6</td>
<td>3.1</td>
<td>1.4</td>
<td>3.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4.0</td>
<td>3.4</td>
<td>4.0</td>
<td>2.0</td>
<td>3.4</td>
<td>1.1</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>3.6</td>
<td>3.1</td>
<td>3.6</td>
<td>1.7</td>
<td>3.3</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4.4</td>
<td>3.9</td>
<td>4.4</td>
<td>2.2</td>
<td>3.8</td>
<td>1.4</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>3.6</td>
<td>3.2</td>
<td>3.7</td>
<td>1.8</td>
<td>3.5</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4.7</td>
<td>4.1</td>
<td>4.7</td>
<td>2.4</td>
<td>4.0</td>
<td>1.2</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>4.0</td>
<td>3.6</td>
<td>4.0</td>
<td>2.0</td>
<td>3.7</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4.8</td>
<td>4.4</td>
<td>4.8</td>
<td>2.6</td>
<td>4.2</td>
<td>1.7</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>3.8</td>
<td>3.4</td>
<td>4.2</td>
<td>1.8</td>
<td>3.6</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4.8</td>
<td>4.2</td>
<td>4.8</td>
<td>2.4</td>
<td>4.0</td>
<td>1.2</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>3.4</td>
<td>3.1</td>
<td>3.4</td>
<td>1.6</td>
<td>3.1</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4.2</td>
<td>3.9</td>
<td>4.2</td>
<td>2.2</td>
<td>3.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>67.2</td>
<td>60.9</td>
<td>67.9</td>
<td>34.9</td>
<td>61.4</td>
<td>19.8</td>
</tr>
<tr>
<td>Total %</td>
<td></td>
<td>100</td>
<td>90.5</td>
<td>100</td>
<td>51.4</td>
<td>100</td>
<td>32.3</td>
</tr>
</tbody>
</table>
Figure 20.--Critical current versus temperature for 2N4128 devices. The solid line is the theoretical curve for silicon. The points represent the average values of sixteen readings taken on eight devices.
will both lower $V_{CE}$ and increase the pulse width.

It was found that repeated testing of the device, particularly at liquid nitrogen temperature, was producing sufficient heating to change the value of $I_C$ read. This was not observed with the much larger 2N3599 devices. To overcome this heating effect, the units were tested just above second breakdown, the test current being gradually lowered until the unit could sustain the current pulses without breaking down. By testing in this manner, the device would break down after the first current pulse as long as the current was above the critical value, giving a duty cycle of less than one pulse per second or less than 0.5% duty cycle, as compared to a duty cycle of 5% when the pulsing is continuous. This was found to be quite satisfactory.

In the course of testing the devices in the above manner, it was found that some devices exhibited regions of collector current in which they went into second breakdown, with regions above and below this range of current where the device performance was stable. This is the "fine structure" reported by Portnoy and Gamble. This phenomenon, when it occurred, was found to occur at all temperatures tested, and had the same temperature dependence as the initial critical current of the device.
CHAPTER VIII

CONCLUSIONS

The results shown in the previous section indicate a very close agreement between theory and experimental data for the devices tested which had good second-breakdown properties. Devices which broke down at very low current levels compared to other devices of the same family when first tested showed only a very slight temperature dependence. Devices which were damaged in testing also showed a small temperature dependence of second breakdown after being damaged. The more severe the damage, the more temperature independent the resulting unit became. This is understandable when one considers that, if the current is being constricted because of a weak spot in the device, heating will occur leading to second breakdown before the current ever reaches the value necessary to produce magnetic pinching at the ambient temperature of the device.

The fact that the observed temperature dependence of second breakdown agrees with the theoretical curve based on the temperature dependence of the limiting velocities of electrons and holes, and not on the
mobility variation with temperature which is much more severe, is fairly strong proof that second breakdown in the reverse-biased emitter configuration is determined by the behavior of electrons and holes in the collector depletion region and not in the base region of the device.

The fact that the data points follow the curve for negligible electric field follows from the observation that the pulses were not heating the devices, and therefore second breakdown had to be due to magnetic pinch- ing. In a device operating on a longer duty cycle, the current would have time to produce a thermal hot spot and the device would follow one of the curves for large electric fields. In this case, which is more typical of normal device operation, the temperature dependence of second breakdown would be much less severe, as shown by the curves in Figures 16 and 17.

The theory presented herein predicts a drastic lowering of the current level at which second breakdown will occur with lowering of the ambient temperature, contrary to the generally accepted ideas of second breakdown as a thermal effect only.

The excellent agreement between theory and empirical data, both in Part I and in Part II of this
dissertation, leads the author to conclude that a fundamental understanding of second breakdown has been achieved.
CHAPTER IX

RECOMMENDATIONS

Theoretical Investigations

The next logical step regarding the theory would be to devise a way of programming the general equations in order to determine the transient behavior of second breakdown.

After that, the equations could be generalized to include first, base current in cylindrical devices, and second, more complicated and realistic geometries. The inclusion of base current effects will also involve the dependence of second breakdown upon beta ($hFE$), which in turn leads to an investigation of the dependence of second breakdown upon beta.

Also, the theory does not explain fine structure. The possibility of multiple breakdown sites and fine structure would be a logical extension of the theory.

Moving breakdown sites have been reported in the literature too. This phenomenon is probably due to the nonsymmetric geometry encountered in most power devices, and might possibly be explained in terms of the effect
of lopsided hot spots on the stability of the hot spots. This should be investigated theoretically also.

Experimental Investigations

Experimentally, the first thing to do is to determine the intrinsic temperature of various devices and see if the critical current to initiate second breakdown saturates at this temperature. Most devices of commercial manufacture probably have too high an intrinsic temperature, and alloying and diffusing would occur under test temperatures. To avoid this difficulty, devices with high base resistivity could be constructed specifically for the purpose of this investigation.

Following that, a test set could perhaps be devised to nondestructively test devices in the forward-biased emitter configuration.

The theory developed herein also applies to diodes. Possibly a test set for testing diodes nondestructively could also be designed to determine the temperature dependence of second breakdown in diodes.

Finally, investigations could be conducted on other devices to determine if they are susceptible to second breakdown, and if so, to determine whether the theory developed herein applies to these devices. Under this heading would come such devices as silicon-controlled-rectifiers.
REFERENCES


34 W. Shockley, R. M. Scarlett, R. Haitz, and W. Hooper, "Failure mechanisms in silicon semiconductors


H. B. Grutchfield and T. J. Moutoux, "Current mode second breakdown in epitaxial planar transistors,"


84L. J. Giacoletto, "Variation of junction-transistor current-amplification factor with emitter current," Proc. IRE (Correspondence), vol. 43, October, 1955, p. 1529.

85N. H. Fletcher, "Note on the variation of junction transistor current amplification factor with emitter current," Proc. IRE (Correspondence), vol. 44, October, 1956, pp. 1475-76.


100 M. F. Nowakowski and F. A. Laracuente, "Infrared pinpoints second breakdown before failure," delivered at


106 For gases, the Boltzmann Transport Equation has only a collision term. The generation-recombination term is peculiar to the solid state, and the impact-ionization term only applies to the depletion region of an avalanching semiconductor diode or transistor.

107 *op. cit.*, S. Wang, p. 209.

108 *op. cit.*, C. Kittel, p. 194.


114 *op. cit.*, L. B. Valdes, p. 121.


143 W. H. Bennett, "Magnetically self-focussing


145See Equation (30).


156v. Rodriguez, H. Ruegg, and M. A. Nicolet,


APPENDIX A
THE GENERAL TRANSPORT EQUATIONS

The Continuity Equation

To obtain the continuity equation, integrate the Boltzmann Transport Equation over all velocity, \( \bar{\mathbf{u}} \).
Assume derivative and integral operations are interchangeable.

\[
\begin{align*}
(1) \quad & \iiint \frac{\partial f}{\partial t} \, d^3u = \frac{\partial}{\partial t} \iiint f \, d^3u = \frac{\partial n}{\partial t} \\
& \quad \bar{\mathbf{u}} \cdot \mathbf{v}_r f = \mathbf{v}_r \cdot (\bar{\mathbf{u}}f) - f \mathbf{v}_r \cdot \bar{\mathbf{u}} \\
& \quad \text{and } \mathbf{v}_r \cdot \bar{\mathbf{u}} = \sum_i \frac{\partial u_i}{\partial x_i} = 0. \\
(2) \quad & \iiint \bar{\mathbf{u}} \cdot \mathbf{v}_r f d^3u = \iiint \mathbf{v}_r \cdot (\bar{\mathbf{u}}f) d^3u = \mathbf{v}_r \cdot \iiint \bar{\mathbf{u}}f d^3u \\
& \quad = \mathbf{v}_r \cdot (n\bar{\mathbf{v}}),
\end{align*}
\]

where \( \bar{\mathbf{v}} = \langle \bar{\mathbf{u}} \rangle = \iiint \bar{\mathbf{u}}f d^3u / \iiint f d^3u. \)

\[
(3) \quad \iiint \mathbf{a} \cdot \mathbf{v}_u f d^3u = \iiint \left[ a_x \frac{\partial f}{\partial u_x} + a_y \frac{\partial f}{\partial u_y} + a_z \frac{\partial f}{\partial u_z} \right] d^3u.
\]
Integrate by parts over \( u_x \), and:

\[
\iiint a_x \frac{\partial f}{\partial u_x} \, du_x \, du_y \, du_z = \int \int \left[ a_x f \right]_{-\infty}^{\infty} \, du_y \, du_z \\
- \iiint f \frac{\partial a_x}{\partial u_x} \, du_x \, du_y \, du_z
\]

\( \frac{\partial a_x}{\partial u_x} = 0 \), since acceleration in the \( x \)-direction is independent of velocity in the \( x \)-direction. Physical distribution functions (Boltzmann, Fermi-Dirac, or Bose-Einstein) go to zero at infinity faster than acceleration becomes infinite. Therefore: \( a_x f \big|_{-\infty}^{\infty} = 0 \)

and \( \iiint \vec{a} \cdot \vec{v}_r f \, d^3u = 0 \).

(4) \( \iiint \frac{\partial f}{\partial t} \bigg|_c \, d^3u = 0 \)

since the same number of representative points must be scattered into some parts of phase-space as are scattered out of other parts. (Generation and recombination are not involved in this term.)

(5) \( \iiint \frac{\partial f}{\partial t} \bigg|_{gr} \, d^3u = \frac{\partial}{\partial t} \iiint f \bigg|_{gr} \, d^3u = \frac{\partial n}{\partial t} \bigg|_{gr} \Delta g - r \),

where \( g \) is the generation rate and \( r \) is the recombination rate.
where $\alpha_i$ is the electron ionization coefficient, $\beta_i$ is the hole ionization coefficient, $\mu_n$ and $\mu_p$ are the electron and hole mobilities, respectively, $n$ and $p$ are the electron and hole densities, and $E$ is the applied electric field. The Continuity Equation is thus:

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\vec{v}) = g - r + \alpha_i \mu_n n E + \beta_i \mu_p p E.$$

The Momentum Transport Equation

To obtain the momentum transport equation, multiply each term in the Boltzmann transport equation by the particle momentum, $m\vec{u}$, and integrate over all velocity, $\vec{u}$.

\begin{equation}
\iint \iint m\vec{u} \frac{\partial f}{\partial t} d^3u = m \frac{\partial}{\partial t} \iiint f d^3u - m \iiint f \frac{\partial m\vec{u}}{\partial t} d^3u
\end{equation}

\begin{equation}
\iiint \frac{\partial m\vec{u}}{\partial t} f d^3u = \left\langle \frac{\partial \vec{u}}{\partial t} \right\rangle n = 0 \text{ as } \left\langle \frac{\partial \vec{u}}{\partial t} \right\rangle = \frac{\partial}{\partial t} \left\langle \vec{u} \right\rangle = 0.
\end{equation}

Therefore:

\begin{equation}
\iint \iint m\vec{u} \frac{\partial f}{\partial t} d^3u = m \frac{\partial}{\partial t} \iiint f d^3u = m \frac{\partial}{\partial t} (n\vec{v})
\end{equation}
(2) \( \mathbf{u} (\mathbf{u} \cdot \nabla f) = \nabla \cdot (\mathbf{u} f) \).

Therefore:

\[
\iiint \mathbf{m} \mathbf{u} (\mathbf{u} \cdot \nabla f) d^3u = m \iiint \nabla \cdot (\mathbf{u} f) d^3u
\]

\[
= m \nabla \cdot \iiint (\mathbf{u} f) d^3u = m \nabla \cdot (n \langle \mathbf{u} f \rangle).
\]

(3) \( \iiint \mathbf{m} (\mathbf{a} \cdot \nabla u) f d^3u = m \iiint \mathbf{u} (\mathbf{a} \cdot \nabla u) d^3u \)

\[
\mathbf{u} (\mathbf{a} \cdot \nabla u) = \mathbf{u} \left( a_x \frac{\partial f}{\partial u_x} + a_y \frac{\partial f}{\partial u_y} + a_z \frac{\partial f}{\partial u_z} \right).
\]

Integrate by parts over \( u_x \), and:

\[
\iiint \mathbf{u} a_x \frac{\partial f}{\partial u_x} d^3u = \int \int \left[ \mathbf{u} a_x f \right] d u_y d u_z
\]

\[
- \iiint a_x d f d^3u.
\]

The first term is zero since \( f \) goes to zero at infinity faster than \( \mathbf{u} a_x \) goes to infinity. The integration by parts proceeds as follows:

\[
\mathbf{a}^* = a_x \mathbf{u}
\]

\[
d\mathbf{a}^* = (a_x \frac{du}{du_x} + \mathbf{u} \frac{da_x}{du_x}) d u_x = a_x \mathbf{d} u_x
\]
\[ dt = \frac{\partial f}{\partial u_x} du_x \]

\[ t = f \]

Therefore:

\[ m \iiint \hat{u}(\hat{a} \cdot \nabla f) d^3u = - m \iiint [a_x \hat{i} + a_y \hat{j} + a_z \hat{k}] f d^3u \]

\[ = - m \iiint d^3u = - mn \langle \hat{a} \rangle. \]

The Momentum Transport Equation is thus:

\[ m \frac{\partial}{\partial t} (n \vec{v}) + mn \cdot (n \langle \hat{u} \hat{u} \rangle) - mn \langle \hat{a} \rangle \]

\[ = \iiint \mu \frac{\partial f}{\partial t} \left| \frac{\partial f}{\partial t} \right|^c + \frac{\partial f}{\partial t} \left| \frac{\partial f}{\partial t} \right|^r + \frac{\partial f}{\partial t} \left| \frac{\partial f}{\partial t} \right|^l d^3u. \]

Let \( \hat{u} = \vec{v} + \vec{w} \), where as before, \( \vec{v} \) is the local average velocity, and \( \vec{w} \) is the excess or deficit velocity for a given particle.* Then:

\[ \vec{v}_r \cdot (n \langle \hat{u} \hat{u} \rangle) = \vec{v}_r \cdot (n \langle \vec{v} \hat{v} \rangle) + \vec{v}_r \cdot (n \langle \vec{w} \hat{w} \rangle) \]

\[ + \vec{v}_r \cdot (n \langle \vec{w} \hat{v} \rangle) + \vec{v}_r \cdot (n \vec{v} \langle \hat{w} \rangle) \]

\[ \langle \vec{w} \rangle = \frac{\iiint (\hat{u} - \vec{v}) f d^3u}{\iint f d^3u} = \frac{\iint \hat{u} f d^3u - \vec{v} \iint f d^3u}{\iint f d^3u} \]

*\( \vec{w} \) is the peculiar velocity of a molecule; its average velocity relative to the local average velocity \( \vec{v}(\hat{r}, t) \) at \( \hat{r} \) and \( t \). (See: Fundamental Problems in Statistical Mechanics, E. G. D. Cohen, Ed., North-Holland Publishing Co., Amsterdam, 1962, p. 112.)
\[ \nabla - \nabla = 0. \]

Therefore:

\[ \nabla_r \cdot (n \langle \vec{u} \vec{u} \rangle) = \nabla_r \cdot (n \vec{v} \vec{v}) + \nabla_r \cdot (n \langle \vec{w} \vec{w} \rangle). \quad (1) \]

Also:

\[ \langle \vec{a} \rangle = - \frac{q}{m} (\vec{E} + \nabla \times \vec{B}) \quad (2) \]

and:

\[ \vec{p} = mn \langle \vec{w} \vec{w} \rangle \text{ is the pressure tensor.} \quad (3) \]

Substitute Equations (1), (2), and (3) into the momentum transport equation, divide by \( m \), and the momentum transport equation becomes:

\[
\frac{\partial}{\partial t} (n \vec{v}) + \vec{v}_r \cdot (n \vec{v} \vec{v}) + \frac{1}{m} \vec{v}_r \cdot \vec{p}^r \\
+ \frac{m q}{m} (\vec{E} + \nabla \times \vec{B}) = \iiint_{all \ \vec{u}} \left[ \frac{\partial f}{\partial t} \left|_c + \frac{\partial f}{\partial t} \left|_r + \frac{\partial f}{\partial t} \left|_l \right. \right. \right] d^3u.
\]

The Energy Transport Equation

To obtain the energy transport equation, multiply the Boltzmann transport equation by \( 1/2 \mu v^2 \), and integrate over all velocity, \( \vec{u} \). \( 1/2 \mu \) is a constant and may be divided out of the equation before performing the integration.

\[
(1) \quad \iiint_{all \ \vec{u}} u^2 \frac{\partial f}{\partial t} d^3u = \frac{\partial}{\partial t} \iiint_{all \ \vec{u}} u^2 f d^3u - \iiint_{all \ \vec{u}} \int_{all \ \vec{u}} f \frac{\partial}{\partial t} (u^2) d^3u
\]

\[ \quad * \langle \vec{a} \rangle \text{ is positive for holes.} \]
\[
\iiint f \frac{\partial u^2}{\partial t} \, d^3u = n \frac{\partial}{\partial t} \langle u^2 \rangle = 0.
\]
Therefore:

\[
\iiint u^2 \frac{\partial f}{\partial t} \, d^3u = \frac{\partial}{\partial t} (n \langle u^2 \rangle).
\]

(2) \[
\iiint u^2 \cdot v_r \, d^3u = v_r \cdot \iiint u^2 \cdot f \, d^3u
\]

since \(u^2 \cdot v_r = u^2 v_r \cdot (df) - u^2 f v_r \cdot \vec{u}\)
and \(v_r \cdot \vec{u} = 0\); \(u^2 v_r \cdot (df) = v_r \cdot (u^2 df)\).

Therefore:

\[
\iiint u^2 \cdot v_r \, d^3u = v_r \cdot (n \langle u^2 \rangle).
\]

(3) \[
\iiint u^2 \cdot v_{uf} \, d^3u = \iiint \left[ a_x \frac{\partial f}{\partial u_x} + a_y \frac{\partial f}{\partial u_y} + a_z \frac{\partial f}{\partial u_z} \right] u^2 \, d^3u.
\]

Integrate by parts over \(u_x\), using the same arguments as those involved in deriving the third term of the continuity equation, and:

\[
\iiint a_x \frac{\partial f}{\partial u_x} u^2 \, d^3u = \int \int \left[ a_x u^2 f \right]_{-\infty}^{\infty} du_y du_z
\]

all \(u\)

all \(u\)
\[ - \iiint \mathbf{f} \frac{\partial (u^2 a_x)}{\partial u_x} \, d^3u \]

\[ \frac{\partial}{\partial u_x} (u^2 a_x) = 2u a_x \frac{\partial u}{\partial u_x} + u^2 \frac{\partial a_x}{\partial u_x} \]

Now, \( \frac{\partial u}{\partial u_x} = \frac{u_x}{u} \) and \( \frac{\partial a_x}{\partial u_x} = 0 \).

Therefore:

\[ \frac{\partial}{\partial u_x} (u^2 a_x) = 2u a_x \frac{u_x}{u} = 2a_x u_x. \]

Repeat for the other two terms and add terms, and:

\[ \iiint u^2 \mathbf{a} \cdot \mathbf{v} u d^3u = - \iiint 2(a_x u_x + a_y u_y + a_z u_z) \mathbf{v} d^3u \]

\[ = - 2 \iiint (\mathbf{a} \cdot \mathbf{n}) \mathbf{v} d^3u = -2 \langle \mathbf{a} \cdot \mathbf{n} \rangle |\mathbf{n}|. \]

Therefore, the energy transport equation can be written as:

\[ \frac{\partial}{\partial t} (n \langle u^2 \rangle) + \mathbf{v} \cdot (n \langle w^2 \mathbf{u} \rangle) - 2n \langle \mathbf{a} \cdot \mathbf{n} \rangle \]

\[ = \iiint u^2 \left[ \frac{\partial f}{\partial t} |_c + \frac{\partial f}{\partial t} |_{gr} + \frac{\partial f}{\partial t} |_{i} \right] d^3u. \]

As before: \( \mathbf{u} = \mathbf{v} + \mathbf{w} \).

Therefore:

\[ u^2 = \mathbf{u} \cdot \mathbf{u} = (\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) = v^2 + 2w \cdot \mathbf{v} + w^2 \]
and, \( n \langle u^2 \rangle = n v^2 + n \langle w^2 \rangle \)

since \( \langle \vec{w} \cdot \vec{v} \rangle = \langle \vec{w} \rangle \cdot \vec{v} = 0 \).

Therefore:

\[
\frac{\partial}{\partial t} (mn \langle u^2 \rangle) = \frac{\partial}{\partial t} [mnv^2 + mn \langle w^2 \rangle]
\]

\(3kT = m \langle w^2 \rangle \) defines the temperature, \( T \), of the carriers.

Therefore:

\[
\frac{\partial}{\partial t} (mn \langle u^2 \rangle) = \frac{\partial}{\partial t} [3nkT + mnv^2]
\]

(4)

\( \langle u^2 \vec{u} \rangle = \langle (\vec{v} + \vec{w}) \cdot (\vec{v} + \vec{w})(\vec{v} + \vec{w}) \rangle 
\]

\[
= \langle v^2 \vec{v} + \vec{v}^2 \vec{w} + (2\vec{w} \cdot \vec{v})(\vec{v}) + 2\vec{w}(\vec{w} \cdot \vec{v}) 
+ \vec{w}^2 \vec{v} + \vec{w}^2 \vec{w} \rangle 
\]

\[
= v^2 \vec{v} + v^2 \langle \vec{w} \rangle + 2 \langle \vec{w} \rangle \cdot \vec{v} + 2 \langle \vec{w} \vec{w} \rangle \cdot \vec{v} 
+ \langle \vec{w}^2 \rangle \vec{v} + \langle \vec{w}^2 \rangle
\]

\( \langle u^2 \vec{u} \rangle = v^2 \vec{v} + 2 \langle \vec{w} \vec{w} \rangle \cdot \vec{v} + \langle \vec{w}^2 \rangle \vec{v} + \langle \vec{w}^2 \rangle
\).

Therefore:

\[
v_r \cdot (mn \langle u^2 \vec{u} \rangle) 
= v_r \cdot [mnv^2 \vec{v} + 2mn \langle \vec{w} \vec{w} \rangle \cdot \vec{v} + mn \langle \vec{w}^2 \rangle \vec{v} 
+ mn \langle \vec{w}^2 \rangle]
\]
2mn \left< w \hat{w} \right> \cdot \hat{v} = 2 \vec{P} \cdot \hat{v} \\
mn \left< w^2 \right> \hat{v} = 3nkT \hat{v} \\
mn \left< w^2 v \right> = 2 \vec{Q} \\
where \vec{Q} is the heat flow vector.

Therefore:

\begin{align*}
\nabla_r \cdot (mn \left< u^2 u \right>) &= \nabla_r \cdot (mn v^2 \hat{v}) + 2 \nabla_r \cdot \vec{P} \cdot \hat{v} \\
&+ \nabla_r \cdot (3nkT \hat{v}) + 2 \nabla_r \cdot \vec{Q} \\
&= (a \cdot u) = (a \cdot (v + \hat{w})) = (a \cdot v) + (a \cdot \hat{w}) \\
&= (a \cdot \hat{v}). \quad (5)
\end{align*}

Upon substitution of Equations (4), (5), and (6) into the energy transport equation, one may express the energy transport equation as follows:

\begin{align*}
\frac{\partial}{\partial t} [3nkT + mn v^2] + \nabla_r \cdot [(3nkT + mn v^2) \hat{v}] \\
&+ 2 \nabla_r \cdot \vec{P} \cdot \hat{v} + 2 \nabla_r \cdot \vec{Q} - 2nq (\hat{E} + \hat{v} \times \vec{B}) \cdot \hat{v} \\
&= m \iiint \left[ \frac{\partial f}{\partial t} \right]_c + \frac{\partial f}{\partial t} \left|_g r \right. + \frac{\partial f}{\partial t} \left|_i \right] u^2 d^3u.
\end{align*}
APPENDIX B

THE SPACE-CHARGE NEUTRALITY CONDITION

The ambipolar transport equations were derived under the assumption of space-charge neutrality, i.e., $\Delta p = \Delta n = \text{excess over the thermal equilibrium hole (electron) concentration}$. For low-level injection or heavily extrinsic semiconductors, space-charge neutrality is known to be a good assumption. However, for intrinsic or nearly intrinsic materials (such as the base region of a device under high-level injection) the spatial distribution of excess carriers is determined exclusively by ambipolar diffusion processes, and space-charge neutrality is violated.\textsuperscript{1,2} Therefore, it was deemed necessary to reexamine the validity of this assumption.

Local electrical neutrality is tantamount to having infinite Coulomb forces or zero dielectric relaxation time. Hence, effects associated with space charge are more pronounced in high resistivity material with large dielectric relaxation times. Consider an increased concentration of holes (positive space charge) in the $x$-direction. From the continuity
equation: \( \frac{\partial J}{\partial x} + \frac{\partial \rho}{\partial t} = 0 \). Combine with Ohm's law, \( \vec{J} = \sigma \vec{E} \), and Poisson's equation and:

\[
\nabla^2 \phi = - \nabla \cdot \vec{E} = - \frac{d\vec{E}}{dx} = - \frac{1}{\sigma} \frac{dJ}{dx} = - \frac{\rho}{\varepsilon}
\]

or: \( \frac{dJ}{dx} = \frac{\sigma}{\varepsilon} \rho = \frac{d\rho}{dt} \)

and: \( \rho = \rho_0 e^{-(\sigma/\varepsilon)t} \). (1)

Thus, any deviation from space-charge neutrality will decay exponentially with a time constant of \( \varepsilon/\sigma \). As typical examples, one ohm-cm germanium has a time constant of about one micro-micro second, while for intrinsic silicon the time constant is of the order of a micro-second.

Furthermore, to preserve space-charge neutrality, donors (acceptors) must remain essentially compensated by an electron (hole) cloud.\(^3,^4\) This implies that injection of electrons and holes due to an emitter or due to avalanche multiplication is required in order to have a large concentration of pinching electrons and holes. In the absence of injected carriers, only the number of electrons and holes equal to the minority carrier concentration can pinch, since for n-type material; \( N_d = n - p \) is the number of electrons needed to preserve space-charge neutrality. This leaves \( n - N_d = p \).
electrons free to pinch with the minority carriers \( p \).

Sikorski\(^5\) has calculated the deviation from space-charge neutrality for current densities from 1 to 1000 amperes/cm\(^2\). Space-charge neutrality was shown to be a very good assumption for current densities below 10 amps/cm\(^2\).

The assumption of space-charge neutrality is not essential, as methods of considering it are available,\(^6\),\(^7\),\(^8\) but the added complexity of the equations is not justified in view of the other assumptions made in this paper.
REFERENCES

APPENDIX B


APPENDIX C

GENERATION AND RECOMBINATION OF CARRIERS

The introduction of a minority lifetime into the continuity equation obviously limits the equation to an equation for minority carriers only, since the implicit assumption is that any deviation of injected carrier density from its equilibrium value will tend to relax with a characteristic lifetime independent of the concentration of the carriers. Following Nussbaum, one can define a lifetime for ambipolar transport of carriers. Assume direct conduction-to-valence-band recombination only. Then the generation rate \( g \) is proportional to the number of empty levels in each band, \( g = K_1(N_v - p)(N_c - n) \), where \( N_v \) and \( N_c \) are the total energy levels in the valence and conduction bands respectively. For nondegenerate materials, \( p \ll N_v \) and \( n \ll N_c \), and: \( g = K_1 N_v N_c \).

The recombination rate is proportional to the product \( np \); \( r = K_2 np \). At equilibrium: \( g_0 = K_1 N_v N_c \) and \( r_0 = K_2 n_0 p_0 \), where \( n_0 \) and \( p_0 \) are the electron and hole concentrations at thermodynamic equilibrium; and, from the principle of detailed balance: \( g_0 = r_0 \).
Define a lifetime, $\tau$, by:

$$
\tau = \frac{n_0p_0}{r_0(n_0 + p_0)}
$$

Then: $K_2 = 1/(n_0 + p_0)\tau$, and:

$$
g - r = - \frac{np - n_0p_0}{\tau(n_0 + p_0)}
$$

or:

$$
g - r = - \frac{(n_0\Delta p + p_0\Delta n + \Delta n \Delta p)}{\tau(n_0 + p_0)}
$$

where $n = n_0 + \Delta n$, and $p = p_0 + \Delta p$.

**Special Cases**

**Space-Charge Neutrality.** For space-charge neutrality, $\Delta n = \Delta p$, and Equation (1) reduces to:

$$
g - r = \frac{\Delta p(n_0 + p_0 + \Delta p)}{\tau(n_0 + p_0)}
$$

**N-Type Material.** If the material is also extrinsic, either $n_0$ or $p_0$ will be negligible compared to the other carrier concentration. Assume n-type material with $n_0 \gg p_0$, and:

$$
g - r = - \frac{\Delta p(n_0 + \Delta p)}{\tau n_0}
$$

**Low-Level Injection.** If one further assumes low-level injection, Equation (3) reduces to:

$$
g - r = - \frac{\Delta p}{\tau p}
$$

Equation (4) is the usual low-level injection
Thus, in theory, the generation-recombination terms in the ambipolar continuity equations could be replaced by expressions of the form of Equation (1). However, one must still know the value of lifetime, $\tau$, to use in Equation (1) for computational purposes; and, at the present, the entire problem is one of academic interest only.
REFERENCES

APPENDIX C


APPENDIX D

MAXWELLIAN VELOCITY DISTRIBUTION

The ambipolar transport equations were derived under the assumption of a Maxwellian or displaced Maxwellian velocity distribution. This limits the validity of the resulting equations to a plasma with a shifted Maxwell-Boltzmann velocity distribution, which may be a very poor assumption for a pinching plasma. Hence, it was deemed necessary to examine this assumption in more detail.

Obviously, when the base of a transistor is non-uniformly heated, thermal equilibrium can only exist locally. A local shifted Maxwellian velocity distribution is defined as:

$$f(r, u) = n \left( \frac{m}{2\pi kT} \right)^{3/2} \exp \left[ -\frac{m (u - \bar{v})^2}{2kT} \right] , \quad \text{all } r$$

where $\bar{v} = \langle \bar{u} \rangle$ is the local velocity, $\bar{u}$ is the particle velocity, and $T$ is the local temperature at the point $r$. The above is known to be inexact, but attempts to improve upon it generally lead to severe complications.
Alder and Wainwright\textsuperscript{3} have shown that for several types of initial conditions, the time to approach a Boltzmann velocity distribution is very rapid, being of the order of 3 or 4 mean collision times for hard spheres. Hence, the assumption of a Maxwellian velocity distribution is valid for non-equilibrium provided the excess carriers have time to distribute themselves as though they were in internal equilibrium.\textsuperscript{4}

Reik and Risken have calculated the energy distribution of "hot" electrons in germanium in the limit of high electric fields, when impurity scattering and electron-electron scattering are negligible.\textsuperscript{5,6} They found that the hot-electron distributions are Maxwellian within the different valleys, provided the intervalley scattering rate is small compared to the intravalley scattering rate. Also, when optical scattering is important, intervalley scattering is at least one or two orders of magnitude weaker than the optical intravalley scattering.\textsuperscript{5}

In view of the above two paragraphs it seems quite reasonable to assume a local Maxwellian velocity distribution in the steady-state when the excess carriers have had time to reach internal equilibrium; and also under avalanche conditions when the electrons are "hot." In general, the assumption of a local Maxwellian velocity
distribution may not be valid; but at least it is consistent with the simplifying assumption of spherical energy-momentum surfaces, in which case, valleys are neglected. Spherical energy surfaces are implied when the effective mass and the resistivity are assumed scalars, and not tensors.\textsuperscript{7,8}

**Electron Temperature**

It should also be mentioned that the introduction of an electron temperature (\(T\) in Equation 1 above) implies that the rate of energy loss of a fast electron to the other electrons must be greater than the rate of energy loss of an electron to other scattering mechanisms. Thus, a high density of electrons is required in high electric fields in order for the rate of loss to other electrons to match that to the polar optical modes.\textsuperscript{9} This condition may not exist initially, but after the electron-hole "plasma" has pinched significantly, this condition will be met. Fortunately, the more the electron-hole plasma pinches the more interesting the case becomes, and the more nearly this condition is met.

Finally, it is obvious that the electron temperature and hole temperature will not in general be equal; and both the local electron and hole temperatures must exceed the local crystal temperature since the electrons
and holes will transfer energy to the crystal lattice when undergoing phonon collisions with the lattice. Due to the complete lack of information regarding the electron and hole temperatures, it is assumed in this paper that the electron, hole, and crystal lattice temperatures are all equal on the local level. This is really tantamount to assuming that the electron and hole temperatures are higher than the lattice temperature by a negligible amount. The higher the local temperature of the "hot spot," the more accurate this assumption becomes.

**Electron-Hole Pressure**

Electron-electron and hole-hole collision effects are the source of a bulk behavior included in the term "pressure," and plasmas, especially when pinched, introduce a tensor-type pressure.\textsuperscript{10} The presence of a magnetic field does not affect the equilibrium in space of charged carriers, nor their Maxwellian distribution of velocity in spite of the resulting curvature of their paths.\textsuperscript{11} For a Maxwellian velocity distribution, the equipartition of energy gives:\textsuperscript{12}

\[ \frac{1}{2} m \langle v^2 \rangle = \frac{3}{2} kT. \]  \hspace{1cm} (2)

From the definition of pressure, $P$, for a gas consisting of $n$ molecules, all having the same mass,
m, one obtains:

\[ P = \frac{1}{3} n m \langle v^2 \rangle. \]  

(3)

Combining Equations (2) and (3), one obtains for a local shifted Maxwellian gas:

\[ P = nkT. \]  

(4)

Equation (4) is the equation of state assumed in this article.

Also, the heat flow vector, \( \mathbf{Q} \), vanishes for a Maxwellian velocity distribution. This has been tacitly assumed in the present paper.

Finally, it should be mentioned that the Boltzmann Transport Equation includes the influence of binary collisions only, and is therefore applicable only to gases having a pressure less than one atmosphere: and the entire approach of magnetohydrodynamics rests on the assumption that collisions are so strong that the pressure always remains a scalar, but still so weak that the conductivity may be considered infinite. These assumptions should be good approximations throughout most of the pinching considered herein.
REFERENCES

APPENDIX D


12Ibid., p. 5 and p. 365.
13Ibid., p. 8.


APPENDIX E

FIELDS OF A CHARGE IN UNIFORM MOTION

The fields due to an electron moving with uniform velocity are governed by the laws of Special Relativity Theory, and may be derived from the Lienard-Wiechert retarded potentials of a point charge. The relationship between the magnetic field, $\vec{B}$, and the electric field, $\vec{E}$, of the electron moving at constant velocity $\vec{u}$, is given by:

$$\vec{B} = \frac{1}{c^2} (\vec{u} \times \vec{E}),$$

(1)

where $c$ is the speed of light, and $\vec{E}$ and $\vec{B}$ reduce to the Coulomb and Biot-Savart fields for low values of $\vec{u}$.

In particular, the magnetic force exerted on a hole moving with velocity $\vec{v}$, due to the presence of an electron moving with velocity $\vec{u}$ in the opposite direction is given by:

$$\vec{F}_m = q(\vec{v} \times \vec{B}) = q[\vec{v} \times \frac{1}{c^2} (\vec{u} \times \vec{E})],$$

(2)

where $q$ is the charge on the hole.

Consider Figure 21 on the following page.
Figure 21.—Electric and magnetic fields acting on a hole moving to the right with velocity \( \vec{v} \), due to an electron moving to the left with velocity \( \vec{u} \).

From Figure 21, it can be seen that the electric field felt by a hole, due to the presence of an electron, makes an angle \( \theta \) with the hole velocity, \( \vec{v} \). Therefore, \( |\vec{u} \times \vec{E}| = uE \sin \theta \), and the direction of \( \vec{u} \times \vec{E} \) is into the page. \( |\vec{v} \times (\vec{u} \times \vec{E})| = uvE \sin \theta \), and the direction is downward, as shown in Figure 21. Thus, the magnitude of \( \vec{v} \times (\vec{u} \times \vec{E}) \) in the direction of the electric field, \( \vec{E} \), is \( uvE \). For an infinite cylinder, in summing forces on the hole due to all the electrons, the components \( uvE \cos \theta \) will cancel due to symmetry, and the resultant force on the hole will be given by \( uvE/c^2 \).

Therefore, the Lorentz force between an electron and a hole is given by:
\[ \vec{F}_{np} = q(\vec{E} + \vec{v} \times \vec{B}) = q\vec{E}_{np} (1 + uv/c^2). \] (3)

In a similar manner, the Lorentz force between two electrons, moving at velocities \( \vec{u} \), is:
\[ \vec{F}_{nn} = q\vec{E}_{nn} (1 - u^2/c^2), \] (4)
and the force between two holes, moving at velocities \( \vec{v} \), is:
\[ \vec{F}_{pp} = q\vec{E}_{pp} (1 - v^2/c^2). \] (5)

The above expressions show that in each case, the magnetic effects are negligible in comparison to the electric effects, when \( u << c \), and \( v << c \). Thus, magnetic effects are only important when the densities of electrons and holes are equal, in which case the electric attractions and repulsions cancel over a finite volume. (I.e., space-charge neutrality exists.)

The magnetic forces are in every case attractive, and will still add in this particular case. (Under the initial assumption that electrons and holes are moving in directly opposite directions.)

Hence, pinching can only occur when space-charge neutrality is approximately valid, and the magnetic effects due to high-speed electrons and holes produces a large resultant force.
REFERENCES

APPENDIX G


APPENDIX F

TRANSISTOR SECOND-BREAKDOWN TEST DEVICE

A transistor second-breakdown non-destructive test device patterned after that of Schiff was constructed under the supervision of Mr. André Buser of Solitron. This device will test units non-destructively under conditions of reverse-biased emitter junction only, and is limited to NPN devices.

Generally, the only load line in which reverse-biased second breakdown occurs is associated with an inductive circuit. In this circuit, the transistor under test is driven into saturation by a base current pulse long enough to establish a steady current in the collector inductance. When the base pulse is terminated, the device is abruptly driven into cutoff due to the reverse-biased emitter supply voltage. The current in the inductance induces a voltage across the inductor which is of such a magnitude and direction as to maintain the collector current. Thus, the device is driven into $BV_{CEX}$, and the current then decays at a rate of $BV_{CEX}/L$. (I.e., the induced voltage is given by $V = L \frac{di}{dt}$, therefore, $\frac{di}{dt} = \frac{V}{L}$.)
In high-speed transistors, the fall time (dt) is relatively small, thus permitting high voltages to be induced across the transistor under test for suitable inductances and collector currents. When these voltages are large enough across the device, the collector-base junction avalanches, clamping the output voltage of the inductor.

Due to the high base currents with accompanying large reverse currents possible with the test set, severe ringing can occur which will trigger the second-breakdown detector. This was eliminated with circuitry allowing the detector to trigger only when the device was in cutoff mode, with a high $V_{CE}$ voltage.

Once a unit goes into second breakdown the device must be disconnected as quickly as possible, since permanent second breakdown damage can occur in a fraction of a microsecond. Because a protective device, such as a transistor, can be turned on faster than it can be turned off, the test set was designed to short the emitter-to-collector of the device under test rather than disconnect it.

Pressing the test button releases the clamp on the pulse generator, which generates a 5.5-millisecond current pulse at a 10-pulse/second rate. The output of the
pulse generator turns on the variable calibrated current source and drives the transistor under test into saturation. After deciding on a value of inductance, L, base resistance, $R_{BE}$, reverse base bias, $V_{BE}$, and temperature $T$, the calibrated current source control is progressively increased until second-breakdown oscillations are detected at the base of the device under test. The detector triggers the monostable, clamping the device and pulse generator, and illuminates the second-breakdown indicator. The Zener diode in the collector circuit cancels the resistance of diode $D_3$ and the clamp.

After a device is driven into second breakdown, the collector current, $I_C$, should be reduced to a relatively low value before the transistor is again subjected to second breakdown. The results are completely repeatable. Feedback is minimized by short leads, and shielding is used to avoid mistriggering of the second-breakdown detector.

Figure 22 shows a block diagram of the test set.

In Figure 23 are shown typical current and voltage pulses of a device not susceptible to second breakdown at this current level. When a device enters second breakdown it usually does so at about the time the device is entering $BV_{CEX}(sus)$. By plotting current
Figure 22.--Block diagram of second-breakdown test circuit.
Figure 23a.--Typical current pulse of the test set.

Figure 23b.--Typical voltage pulse of a device not susceptible to second breakdown.
versus voltage from Figure 23, one obtains the duty-cycle of the device. This is shown in Figure 24.
a = $V_{SAT}$

b = $BV_{CEX}(sus)$

c = $BV_{CEX}$

Figure 24.--Duty cycle obtained by eliminating time in Figure 23.
APPENDIX G

RESISTIVITY VERSUS INTRINSIC TEMPERATURE

As the temperature of the hot spot rises, the resistivity approaches the intrinsic value for the given temperature. In order to obtain a reference level for the temperature required to make material of a given starting resistivity intrinsic, define the "Intrinsic Temperature" as that temperature at which the intrinsic number of electrons, \( n_i \), is equal to the number of ionized donor atoms, \( N_d \). For p-type material, the intrinsic temperature would be that temperature at which the concentration of holes, \( p_i \), is equal to the number of ionized acceptor atoms, \( N_a \).

Donor and acceptor atoms are almost completely ionized at room temperature. In the remainder of this discussion, it will be assumed that they are 100% ionized, and that \( N_a \) and \( N_d \) are the total impurity concentrations.

From the conservation of charge one obtains:

\[
\begin{align*}
n - p &= N_d - N_a, \\
\end{align*}
\]
where \( n - p \) is the net concentration of free electrons, and \( N_d - N_a \) is the net number of positively ionized donors. For intrinsic material, \( N_d = N_a = 0 \), and Equation (1) reduces to \( n = p = n_i \), where \( n_i \) is the intrinsic number of electrons or holes. This leads to:

\[
np = n_i^2.
\]  

Equation (2) is true under any doping conditions, provided that thermal equilibrium is maintained.\(^1\)

For n-type material (uncompensated), \( N_a = 0 \), and Equation (1) reduces to:

\[
n - p - N_d = n - n_i^2/n - N_d = 0.
\]  

For \( n_i = N_d \), Equation (3) can be solved for \( n \) to yield:

\[
n = 1.618 \, N_d.
\]  

In this case one obtains from Equation (2), \( p = 0.618 \, N_d \), and the ratio of \( n \) to \( p \) is:

\[
n/p = 2.62.
\]  

To obtain the equivalent equations for p-type material, replace \( n \) by \( p \), and \( N_d \) by \( N_a \).

One next needs an equation relating the concentration of impurities to the resistivity. From the definition of resistivity, \( \rho \):
\[ 1/\rho = q(\mu_n n + \mu_p p), \]  
\[ \text{where } \mu_n \text{ and } \mu_p \text{ are the electron and hole mobilities, respectively, and } q \text{ is the electronic charge.} \]

For heavily doped n-type material, Equation (6) reduces to:
\[ 1/\rho = q\mu_n n \approx q \mu_n N_d, \]  
while, for heavily doped p-type material, Equation (6) reduces to:
\[ 1/\rho = q\mu_p p \approx q \mu_p N_a. \]

At low resistivities, the electron and hole mobilities are functions of the impurity concentration, due to impurity scattering of the charge carriers, and they are therefore functions of the resistivity. For germanium\textsuperscript{2} of resistivity less than 1 ohm-cm:
\[ \mu_n = 3800 + 1000 \log \rho, \]
\[ \mu_p = 1800 + 535 \log \rho, \]
\[ \text{and} \]
while for silicon\textsuperscript{2} of resistivity less than 10 ohm-cm:
\[ \mu_n = 1280 + 470 \log \rho, \]
\[ \mu_p = 300 + 100 \log \rho. \]

Upon substituting the correct mobility from Equation (9) or (10) into either Equation (7) or (8), and solving for the impurity concentration, one obtains the
following results:

(a) Germanium:

\[
N_a = \frac{1.17 \times 10^{16}}{3.37 + \log \rho} \quad (11)
\]

\[
N_d = \frac{6.25 \times 10^{15}}{3.80 + \log \rho} \quad (12)
\]

(b) Silicon:

\[
N_a = \frac{6.25 \times 10^{16}}{3.00 + \log \rho} \quad (13)
\]

\[
N_d = \frac{1.33 \times 10^{16}}{2.72 + \log \rho} \quad (14)
\]

The temperature dependences of the intrinsic electron concentrations for germanium and silicon have been measured by Morin and Maita.\(^3,4\) Upon combining Equations (11) and (12), or Equations (13) and (14) with the data of Morin and Maita, one can obtain a graph of initial resistivity versus the intrinsic temperature, where the intrinsic temperature is defined as that temperature at which \(n_i = N_d\) for n-type material, and \(p_i = N_a\) for p-type material.

Figures 25 and 26 are graphs of initial resistivity versus intrinsic temperature for germanium and silicon, respectively.

The intrinsic temperature, as given by Figure 25
Figure 25.--Initial resistivity versus intrinsic temperature for germanium.
Figure 26.--Initial resistivity versus intrinsic temperature for silicon.
or 26, gives an upper limit to the thermal variation of the critical current to initiate second breakdown. Once a region of the device becomes intrinsic it is no longer an NPN (or PNP) transistor, and the above analysis fails.

The theory of junctions between p-type and intrinsic material, designated as p-i junctions, and between n-type and intrinsic material, designated as n-i junctions, differs from the theory of p-n junctions. Several authors have considered these types of junctions in detail.5-12 Prim7 gives an exact solution for a p-i junction. His results show that practically all the electrostatic potential drop appears in the intrinsic material, and that under reverse bias, these junctions exhibit a high electric field everywhere in the intrinsic region, and have no true saturation. In particular, if the base region is the lightly doped region, as it is in alloy devices, the space-charge region extends throughout the base once it has become intrinsic. In this case the effect is the same as when "punch-through" occurs, and the collector is shorted to the emitter.

While all devices driven into destructive second breakdown must reach temperatures above the intrinsic temperature of all three regions of the device, the theory will fail, since once the hot spot has exceeded
the intrinsic temperature of the base material, a short circuit across the base region exists at the hot spot. Beyond this temperature, the device would be expected to behave in a manner similar to devices whose second breakdown is controlled by defects, and the critical current to initiate second breakdown should be relatively insensitive to further increases in temperature. Thus, the device should follow the theoretical curve for temperatures below the intrinsic temperature, and the critical current should saturate above the intrinsic temperature. Such saturation effects have been observed at high temperatures.¹³
REFERENCES

APPENDIX G


13A. Buser, Solitron Devices, Inc., Transistor Division, Riviera Beach, Fla., private communication.
BIBLIOGRAPHY ON SECOND BREAKDOWN

1967


Nienhuis, R. J. "Second Breakdown in the forward and


1965


Lucovsky, G. and R. B. Emmons. "Avalanche multiplication


Tager, A. S. "Current fluctuations in a semiconductor (dielectric) under the conditions of impact


1964


1963


1962


1961


1960


1959


1958


Kaiser, W., H. L. Frisch, and H. Reiss. "Mechanism of the formation of donor states in heat-treated

Matz, A. W. "Variation of junction transistor current amplification factor with emitter current," Proc. IRE (Correspondence), vol. 46 (March, 1958), pp. 616-17.


1957


Sclar, N. and E. Burstein. "Impact ionization of


1956


Fletcher, N. H. "Note on 'the variation of junction transistor current amplification factor with emitter current,'" Proc. IRE (Correspondence), vol. 44 (October, 1956), pp. 1475-76.


1955


1954


1953


McKay, K. G. and K. B. McAfee. "Electron multiplication


1952


1951


1950


1949


1944


1934


1903

Townsend, J. S. "The genesis of ions by the motion of positive ions in a gas; and a theory of the sparking potential," Phil. Mag., vol. 6 (November, 1903), pp. 598-618.

1901

VITA

Eugene D. Fabricius was born on [date]. He received his elementary education in St. Louis, through the first year of high school. He then served five years in the U.S. Army, including one year of combat with the Second Infantry Division in Korea.

He received his B.S. degree in Physics from Missouri School of Mines and Metallurgy in June 1956, and then taught mathematics there until he received his M.S. degree in Physics in June 1958. He had a summer appointment as Resident Research Associate at Argonne National Laboratories in 1957.

He taught mathematics and physics at Texas A. and M. for one year. Following that he was a Design and Development Engineer at Texas Instruments, and later, Senior Physicist at Giannini Controls Corporation, both positions being held for a year and one half.

In 1962 and 1963 he was awarded research assistantships from Newark College of Engineering where he taught mathematics for one year; and physics, mechanics, and electrical engineering for three years. He received a National Science Foundation Fellowship for the academic year of 1966-67.